An analysis of complementary products associated with unhealthy food purchases using household grocery sales data in Montréal, Canada

Kody Crowell
Department of Epidemiology, Biostatistics, and Occupational Health
McGill University, Montréal, Québec, H3A 0G4, Canada
August 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of MSc in Epidemiology.

Table of contents

Abstract

Background

Consumption of soft drinks and snack food contributes to the increasing global incidence of chronic illnesses such as cardiovascular diseases and type II diabetes. Previous studies addressing the purchasing patterns of such foods have emphasized the importance of complementary and co-occurring products, as both may undermine interventions seeking to limit the intake of unhealthy foods. However, research using household-level data to analyze such patterns in purchasing has traditionally been limited in terms of volume, objectivity, and representativeness. Moreover, few studies have searched explicitly for co-purchasing of soft drinks and snack foods in the same basket.

Research objective

The objective of this study is to identify patterns in food categories purchased by households together with, or complementary to, soft drinks and snack foods as well as fresh fruits and vegetables.

Methods

We used longitudinal, household-level transaction data from 14,999 loyalty card members of a large grocery retailer in Montréal, Canada between February 2015 and September 2017 (1,522,501 transactions). Association rule mining was used to identify frequently co-purchased item categories for soft drinks, snack foods, juice, fruits, and vegetables.

Results

Transactions (baskets) containing snack foods and soft drinks were also likely to contain canned or highly-processed foods. For example, soft drinks were highly associated with salty snacks (confidence: 17%; odds ratio: 1.82 ± 0.02), bottled water (confidence: 16%, odds ratio: 1.77 ± 0.02), and frozen meals and sides (confidence: 16%; odds ratio: 1.78 ± 0.03). Conversely, purchases with qualitatively healthier foods were found to be associated with purchases of fruits and vegetables: purchases with vegetables were highly associated with fresh herbs (confidence: 84%; odds ratio: 1.90 ± 0.03) and packaged salads (confidence: 73%; odds ratio: 1.61 ± 0.01).

Conclusions

These empirical results quantify the extent to which healthy and unhealthy food-purchasing behaviours cluster within baskets. Public health practitioners seeking to design interventions that decrease the frequency of soft drink and snack food purchases in the grocery retail environment should consider the tendency for multiple unhealthy foods to be purchased concurrently. While loyalty card data do not capture the entirety of a household's food purchasing behaviour, they represent objective and proximal outcomes to dietary patterns and should therefore be used alongside more traditional means of dietary assessment.

Résumé

Contexte

La consommation de boissons gazeuses et de collations contribue à l'augmentation de l'incidence globale des maladies chroniques comme les maladies cardiovasculaires et le diabète de type II. Plusieurs études sur les habitudes d'achats de ces types d'aliments ont insisté sur l'importance des produits complémentaires et des produits coexistant, puisque ceux-ci peuvent diminuer l'impact des interventions qui visent à réduire l'apport en malbouffe. Cependant, les études qui font l'analyse de ce type d'achat en utilisant des données agrégées au niveau des ménages sont d'ordinaire limitées en termes de volume, d'objectivité et de représentativité. De plus, peu d'études se sont intéressées aux co-achats de boissons gazeuses et de collations dans le même panier de consommation.

Objectif de recherche

L'objectif de cette étude est d'identifier les catégories alimentaires que les familles achètent conjointement avec les boissons gazeuses et les collations, ainsi qu'avec les fruits et les légumes frais.

Méthodes

Nous avons utilisé les données longitudinales de transactions agrégées au niveau des ménages de 14 999 membres d'un programme de carte de fidélité d'une chaîne d'épicerie de grande surface à Montréal, Canada, entre février 2015 et septembre 2017 (1 522 501 transactions). Nous avons utilisé les règles d'association pour identifier les catégories qui étaient fréquemment co-achetées avec les boissons gazeuses, les collations, le jus, les fruits et les légumes.

Résultats

Les transactions (paniers) comportant des collations et des boissons gazeuses étaient aussi susceptibles de comporter aussi des aliments traités ou en conserve. Par exemple, les boissons gazeuses avaient une forte association avec les collations salées (confiance: 17%; rapport des cotes: 1.82 ± 0.02), l'eau en bouteille (confiance: 16%, rapport des cotes: 1.77 ± 0.02) et les repas surgelés (confiance: 16%; rapport des cotes: 1.78 ± 0.03). À l'inverse, les achats de fruits et de légumes étaient très associés avec les aliments sains comme les herbes fraiches (confiance: 84%; rapport des cotes: 1.90 ± 0.03) et les salades emballées (confiance: 73%; rapports des cotes: 1.61 ± 0.01).

Conclusions

Ces résultats empiriques quantifient jusqu'à quel point les achats d'aliments sain et malsain se rassemblent dans les paniers. Les professionnels en santé publique qui cherchent à développer des interventions pour réduire la fréquence d'achats de boissons gazeuses et de collations doivent prendre en considération la tendance à acheter d'autres aliments malsains en parallèle. Même si les données de carte de fidélité ne saisissent pas la totalité des achats d'un ménage, elles représentent un résultat objectif et rapproché des habitudes alimentaires. Ainsi, ces données devraient être utilisées conjointement aux méthodes d'évaluation alimentaire plus traditionnelles.

Acknowledgements

I begin by extending my gratitude to my supervisor, Dr. David Buckeridge, without whom this project would not be possible. His guidance over the course of my degree granted me possibilities I would not have otherwise received. Eternal praise to the members of the McGill Clinical and Health Informatics research group for their wisdom, patience, and advice. To Anya and Nikita; to Guido, with whom I shared many lunches and conversations (and who always seemed to have the right words to say to quell a graduate student's anxiety); to Hiroshi, who reminded me why I love research; to Deepa, whose experience in this field I relied heavily upon; and to Aman, my Virgil, who fielded every question I asked and navigated every intricate detail of the data with me, who kept me on the true and steady path and who remains a true friend to this day. To my other colleagues of 1140 Pine, Bettina, Daniala, and Siyana: thank you for sharing laughs with me. And of course, we would all be nothing but for the toil of Patricia and Rosalba — the group's foundations rest on the shoulders of these two magnificent women. I am indebted to both of you. I thank Dr. Yu Ma as well, for the indispensable input.

Thank you to the Department of Epidemiology, Biostatistics, and Occupational Health, and all the kind souls who guided me over the last few years (Deirdre, Katherine, Dolores, Gilles, and Theresa). Thank you to Mathieu, for breaking me from the mundanity of my work. The lessons given by all of the professors of this department (both good and bad) are ones I will keep for the rest of my life.

Of course, a master's degree is about more than just your own primary research. It is about the connections you make and the windows you open (for every door that may be closed). My career interests and content knowledge has been both widened and refined by this degree, a result of the labor of my teachers. Indeed, there are some educators you meet that are able to subtly change the way you think about things, be them systems of oppression, systems of justice, or systems of care. Dr. Amélie Quesnel-Vallée is one that comes to mind. Thank you for letting me be a part of your class and your research project. Dr. Catherine Mah: I am so grateful for having met you. It was as if I had stumbled upon a fountain of knowledge, an oasis in a desert of my own confusion. Your patience and your support have been instrumental in lifting the sandstorm from my eyes. To my committee members, I extend my warmest gratitude.

To my friends, the ones who helped carry my upwards into the clouds: thank you. To Arisha: I share only a fraction of the passion that drives your heart, but I like to think that you have inflamed my own slumbering appetites for social change. May your ship sail long and far. To Sarie: I cannot stress enough the value your companionship meant to me. My relationship with writing and editing ebbs and flows seasonally, but my relationship to my writers and my editors – my relationship to you – never falters. To Luca: your friendship means so much to me. I wish you the best in Copenhagen! Many wishes of luck to all of my colleagues at the McGill International Review (Gracie, Alec, Helena, Phoebe, Marissa, and Pauline, Zoë, Shaista). To my classmates (Berson, Jenkin, Christina, Arnaud, and Vaidehi), your persistence has inspired me to be my better self. I hope the summer wines continues to flow for you until well after we have emptied our decanter and the sun has sunk into the horizon. Lilia, I wish you the best of luck in your upcoming MScPH degree with the department. I hope to be the one you lean on when times are rough. To Sara: ما بين شوقين قلبي فيك يرتحل

I am eternally grateful to the queer women in my life, whose sapphic wisdom I carry in my heart wherever I travel. To my roommates: thank you for doing the dishes in a timely manner. To Jon, Kirsty, Robin, and Mia: thank you for carrying me from Acadia into the new world. You always have a place in my home. To Keith: thank you for bringing my heart to your neighbourhood. Thank you to Eileen, for helping me rise to the occasion of my master's degree, and to Andrew, for helping come down. To Kat: thank you for sharing precious moments with me in this city. I hope to relive them again. To Delana: thank you for the endless banter and discourse. I can think of no other person I'd rather spend an evening of burlesque with. To Omar, thank you for opening my eyes. To Andrea: lasciarti non è possibile. To Mahmoud: it is hard to imagine how different my life would be if you were not in it. I'd like to ignore the counterfactual, and be content that you are a part of it still. To Dawn: I hope your obsession with McDonald's does not kill you and that you make it to this city before it does. To Justina, to Jed, to Celine, and to Sally: I love you dearly. Where your footsteps land, my heart skips two beats.

Finally, to Montréal, this damned city. Thank you for being my muse and my curse.

Preface

Author contributions

I am the primary author of this thesis and the manuscript included within. Input and revisions were made by Dr. David L. Buckeridge, Aman Verma, Dr. Amélie Quesnel-Vallée, and Dr. Catherine Mah. From McGill, with the technological support granted by the McGill Clinical and Health Informatics research group, I cleaned and processed the database from its PostgreSQL server and conducted all analysis. Dr. Yu Ma and his research assistants were responsible for linking UPC to category. Aman Verma was responsible for transferring and uploading the data to the server from which I worked. Through Dr. David L. Buckeridge was this entire project conceived. He directed the communication between the retailer before the data was transferred.

Thesis organization

This thesis explores complementary food purchases associated with soft drinks and snack foods using household-level, grocery sales data. The document has been prepared according the guidelines of for a manuscript-based thesis. The rationale and objectives for this research are outlined in Chapter 1. Following this, the topic and its overall relevance is introduced in Chapter 2, where I provide background information on the global expansion of retail food environments and their role in shaping consumer habits. Research surrounding the growing prevalence of unhealthy food products in these spaces and the efforts to combat them is presented. In this regard, a summary of previous uses of sales data is also discussed in relation to my own work. Chapter 3 contains the submitted manuscript, the results of this research. Chapter 4 interprets these results and reflects on the limitations that may have impeded their realization, concluding by situating them in the broader discourse on healthy eating and suggesting potential future work. Chapter 4 briefly concludes the document and references are provided in Chapter 5. Appendix A gives a detailed description of the data and the loyalty card member population and study methodology used. Lastly, Appendix B provides full tabulations for the results of the study, which are discussed at length in Chapter 4.

List of abbreviations

ARM	Association Rule Mining
BMI	Body Mass Index
CCHS	Canadian Community Health Survey
CI	Confidence Interval
DALY	Disability-Adjusted Life Year
ERD	Entity Relationship Diagram
F&V	Fruits and Vegetables
FM	Frequency, Monetary value
GfK	Growth From Knowledge
HFCS	High-Fructose Corn Syrup
KWP	Kantar WorldPanel
LOESS	LOcally-Estimated Scatterplot Smoothing
NCD	Non-Communicable Disease
NCP	National Consumer Panel
OR	Odds Ratio
PCA	Principal Component Analysis
PCCF	Postal Code Conversion File
PK	Primary Key
QC	Québec
RFM	Recency, Frequency, Monetary value
RTE	Ready-To-Eat
SNAP	Supplemental Nutrition Assistance Program
SSB	Sugar-Sweetened Beverages
UPC	Universal Product Code
UPFP	Ultra-Processed Food Product
US	United States
WC	Waist Circumference

List of tables and figures

I. Introduction

Rationale

The global proliferation of the retail food environment has brought about an unprecedented dietary shift towards the consumption of nutrient-poor and energy-dense ultra-processed food products (UPFP). These products have have entrenched themselves within the global diet. The increased intake of these foods — often high in added sugar, sodium, or trans and saturated fats — coincides with the growing burden of non-communicable diseases (NCD) worldwide. Indeed the burgeoning displacement of traditional diets in favour of highly-processed, nutrient-deficient foods has become so manifest in terms of their deleterious outcomes, some researchers have called for a recognition of these products as "edible pathogens" and "pseudo-foods."[1,2] Meanwhile, the availability of unhealthy food products across almost all food outlets, especially large supermarket chains, shows no signs of regressing.[3]

Interventions have sought to curb the intake of these foods and promote the consumption of healthier foods through various policies and programs: fiscal measures such as taxation or subsidization, improving geographic access to healthy food outlets, mandatory nutritional labelling, or reducing marketing activities towards children are among such programs. To date, these efforts show weakly positive or neutral outcomes.[4–6] Evaluations of the interventions in these retail spaces rest largely on the purchasing patterns of a single food category (such as soda) and do not consider the comprehensive composition of a food basket. This lack of knowledge surrounding complementary products translates into a sizable shortfall: targeted regulation against a single set of items may not be effective in improving the diet of the population if equally unhealthy foods are not taken into account.[7] Understanding the broader range of purchases made by a consumer in a retail supermarket environment is therefore an important step in crafting more effective interventions.

Research objectives

In order to fill this knowledge gap, we explored co-purchasing patterns of soft drinks, juices, and snack foods across households by using grocery transaction records from a large retail grocery chain in Montréal, Québec. The specific aim of this thesis is to illuminate co-purchasing patterns among loyalty card member transactions that contained soft drinks, sweet or salty snacks, and fruit juice as well as those that contained fresh vegetables and fruits. To achieve these goals, we collected a sample of 14,999 anonymized card members (over 1.5 million transactions) from a database of loyalty cards provided by a large grocery retailer in Montréal and subsequently applied an unsupervised machine learning technique to identify associations between purchase categories.

II. Review: unhealthy food purchases in unhealthy food environments

The increasing global health burden of non-communicable diseases

NCDs including cancers, diabetes, and cardiovascular diseases accounted for 73.4% of deaths worldwide in 2017, a drastic 22.7% increase over the previous decade.[8] Historically, the burden of such chronic diseases were carried by developed countries, a result of an increasingly sedentary lifestyle and diets with a higher energy density.[9] However, the skyrocketing global incidence of NCDs has brought into question the traditional idea of NCDs as "diseases of affluence," as many developing countries have now converged or surpassed the burden experienced by developed countries. The fearful prospect of deadly epidemics and outbreaks is now compounded by the growing need to adapt to a 'double burden' of disease, one that deepens inequities and multiplies healthcare utilization costs. Indeed, by 2030, it is estimated that 75% of deaths globally will be attributable to NCDs.[10] Although Canada experiences a lower risk of NCD mortality compared to other countries, NCDs still pose a grave problem.[11] In 2016, cardiovascular diseases were found to be among the top NCDs driving mortality in the country, claiming 81,352 all-age deaths.[12] This growing trend in Canada has prompted public health researchers to shift their focus towards preventative risk factors such as diet.

Links between non-communicable diseases and dietary risk factors

The relationship between NCDs and diet is well-documented.[13,14] Indeed, a great deal of NCD mortality and morbidity may be preventable vis-à-vis improvements in dietary habits. It has been estimated that dietary risk factors — including high intakes of sodium and low intakes of whole grains, fruits, and vegetables — claimed 11 million deaths worldwide in 2017.[15] Epidemiological evidence has supported causal relationships between the intake of fruits and vegetables (F&V), whole grains, and meats with a reduction in a number of NCDs.[13,14] In particular, fruit and vegetable consumption has been significantly associated with lower risks of coronary heart disease (CHD), cardiovaascular diseases (CVD), and ischemic and hemorrhagic stroke.[16–18] Similarly, the intake of fibre-rich foods has been linked to a lower risk of all four outcomes, along with diabetes, hypertension, and some gastrointestinal disorders.[19,20] Whole grain consumption follows suit, demonstrating a protective effect on type II diabetes and CVD.[21,22]

Meanwhile, the jury on red meat is still out, as some studies have identified conflicting effects on the risk of type II diabetes, hypertension, and CHD; however, the effects of processed meats appear to be much stronger.[23,24] Red and processed meat consumption has been associated with a higher incidence of colorectal, colon, and rectal cancers.[25] Fish and seafood consumption show a weakly protective effect on CHD mortality, while nuts and legumes similarly show inverse associations with CHD and diabetes, but not stroke.[26,27] Systematic reviews of dairy consumption have shown either favourable or neutral associations with non-communicable diseases, although some studies suggest that low-fat dairy and milk could prevent hypertension and CVD.[28–31]

Higher intakes of added sugar and non-communicable diseases

The effect of added sugars (in contrast to naturally occurring sugar) is currently at the center of a number of debates within epidemiological and medical communities. There is a consensus that a high intake of added sugars is generally associated with a higher CVD risk.[32,33] Added fructose, in particular, has come under considerable scrutiny for its role in de novo lipogenesis and increased risk to cardiometabolic health.[34–36] However, since fructose is rarely consumed in isolation, any putative mechanisms are frequently confounded: glucose and fructose intake tend to co-vary, and as such, it becomes difficult to differentiate the effects of the two.[36] This does not, however, warrant a dismissal of the large-scale shifts in diet towards the increased consumption of added sugar.

High-fructose corn syrup (HFCS), a processed sweetener derived from corn starch, has largely replaced sucrose and other simple sugars due to its low cost of production; it now makes up roughly 40% of all consumed sweeteners and is a popular substitute in almost all foods which contain caloric sweeteners, especially processed foods.[34] The intake of HFCS has increased dramatically over the past several decades, representing more than two fifths of the intake of caloric sweeteners among the general population in the United States (US), a trend driven largely by sugar-sweetened beverages (SSB) such as soft drinks, energy drinks, sugar-sweetened coffee and tea, flavoured water, chocolate milk, and fruit juices.[34] The high intake of HFCS and other added sugars in the form of SSB coincides with increasing obesity rates.[37,38]

The growing demand for sugar-sweetened beverages and snack foods

Trends in sale and consumption of soft drinks and snack foods and their impacts

The evidence relating diet quality to the prevalence of NCDs has done little to encourage manufacturers to reformulate their products, promote healthier foods, or limit the availability of unhealthy foods high in added sugar and sodium. Indeed, despite some reductions, many food and beverage purchases among households in the US still contain excessive quantities of sodium.[39] Food-at-home purchases (as opposed to food-away-from-home) of soft drinks and other refined sugary products rose steadily from the 1960's into the early 21st century.[40] As of 2010, the average global SSB consumption in adults aged 20 or more was 0.58 8-oz. servings per day.[41]

In Canada, 35% of the daily sugar intake came from food categories other than F&V

such as SSB or candy, with beverages representing 44% and 35% of sugar intake in children and adults, respectively.[42] According to the 2004 Canadian Community Health Survey (CCHS), nearly 50% of all calories consumed by Canadians came from nutritionally-inferior processed foods, which include, inter alia, sweet and salty snacks, SSB, desserts, and confectionery.[43] This is a sizable departure from recommended dietary guidelines, as frequent intake of vegetables and fruits remains consistently low.[44] Moreover, there is a marked sociodemographic gradient in the dietary quality of Canadian populations, with marginalized groups consuming F&V less frequently, due in part to a lack of access and affordability.[45,46]

Although sales of soft drinks and fruit juices have decreased since 2004, the per capita sales volume for other SSB has increased with sugary drink consumption being the highest among youth.[47,48] This increased intake of SSB and other unhealthy UPFPs pose serious health risks for Canadians. Consumption of SSB is positively associated with weight gain, risk of coronary heart disease and a greater incidence of type II diabetes.[49–51] It is estimated that 63,000 deaths and 2.2 million disability-adjusted life years (DALYs) are attributable to Canadians' sugary-drink consumption.[47] Particularly concerning is the effect of these nutrient-deficient products on children, as studies have linked consumption of SSB with childhood obesity.[52] This, in turn, has spurred activists to politically mobilize against the presence of SSB and snack foods in schools, bringing them in direct conflict with multinational beverage and snack food corporations that form industry partnerships with the educational institutions in order to continue marketing their products.[1,53]

Industrialization, capitalism, and the normalization of ultra-processed foods

The existence of nutrient-deficient UPFPs in the global, and more specifically, North American diet (for changes in the latter have heavily influenced the former) is a rather recent phenomenon in the history of humankind, and it is not one that arose from mere chance.[54–56] The economic and technological developments of the 20th century provided new processing and manufacturing techniques for commercial use: the invention of new canning technologies; new bleaching and fortification processes for flour; the hydrogenation of oil leading to the use of mono- and polyunsaturated fats; the invention of mechanization and automation technologies; the explosive post-war growth of the chemical industry that saw the advent of preservatives and additives; advances in plastic packaging; the replacement of natural sugars with substitutes such as HFCS; and the globalization of retailing and processing spurred on by the capitalist logic of accumulation and concentration.[57]

While these developments were greatly economical for manufacturers and producers, they proved to be detrimental to consumers. From the beginning of the 1900s, there has been a remarkable growth in the consumption of refined sugars and fats and an overall increase in total caloric consumption — "overnutrition," as Nestle (2007) would call it.[53,57] Today, it is estimated that 77% of all calories purchased in the US contain caloric sweeteners.[58] This alteration has, in part, been driven by food environments across the globe, who continually exploit the psychological and socio-economic vulnerabilities of populations towards unhealthy dietary patterns by capitalizing on time and financial constraints, impulsivity, and preference for foods that may trigger overeating.[59,60]

This degradation of the global diet, accompanied by a shift to an ever more sedentary lifestyle, has also been heavily impacted by the emergence of mass marketing media, which has served to normalize these manifestly unhealthy UPFPs as sine qua non of one's diet. Children and adolescents, especially, are frequent targets for intense advertising campaigns promoting the consumption of fast foods, candy, snack foods, and sugary beverages.[61–63] These targeted campaigns occur within and outside of schools.[53,64] This exposure has been correlated with a heavier intake of foods high in added sugar and fats and a lower intake of F&V.[65,66] Food advertising has also been heavily directed towards adults, similarly demonstrating a strong effect on individual food choice.[67,68]

In response to these practices, many have called for stronger regulatory restrictions against the commercial marketing of unhealthy foods and beverages, particularly towards children.[61,63] These calls have been countered by the lobbying efforts of the food industry, as manufacturers and lobby groups seek to protect their differential profits and maintain market monopoly.[69,70] The influence these multinational food industries have on the economic and political landscape cannot to be understated: roughly 75% of the US beverage market share is dominated by Coca-cola and PepsiCo; Frito-lay (PepsiCo) controls 56% of the salty snack food industry, while 50% of the cookie and cracker industry is dominated by Nabisco (Kraft) and Kellogg.[71] Globally, just ten packaged food companies account for over 15% of all packaged food sales, while the top ten soft drink companies account for 52% of sales worldwide.[72] These large organizations frequently seek to undermine existing dietary science, co-opt academics and government officials, shift the rhetoric around diet and health, and deregulate the standards for food safety (all notably similar tactics used by the tobacco industry in the latter half of the 20th century when it sought to limit public health efforts to curb smoking).[69,70,73,74]

The spatial colonization of the supermarket

The omnipresence of mega-brands in the retail food environment is particularly noticeable in the grocery supermarket, an example of what Glanz (2005) calls the 'consumer nutrition environment,' where in-store characteristics actively influence individual purchasing patterns.[75] This process by which these highly-processed, nutrient-deficient food brands come to dominate such retail grocery environments is referred to as "spatial colonization" by Winson (2004).[1] This spatial colonization involves the gradual takeover of shelf space in supermarkets by foods high in sugar and fats, all while pushing basic food items to the store's margins. Such a phenomenon is one of many tactics used by supermarkets to influence the shopping patterns of consumers and increase profit. How such environments came to be, however, is another question entirely.

The dominance of large supermarket chains in urban food environments

Supermarkets qua food outlets remain an important source of nutrition for many households. In many developing countries across Latin America, Asia, and Africa, supermarket chains are expanding at unprecedented rates, usurping traditional food economies and replacing local fresh markets as the de facto source of nutrition.[55,76,77] In the developed world, the supermarket has already embedded itself within the food economy: in Canada, chain food stores accounted for an estimated 59 billion Canadian dollars in sales in 2017.[78] Although only one food location among many from which consumers may choose to shop, supermarkets maintain a heavy presence in urban landscapes and play a major role in food purchasing.[79]

A great deal of research has been devoted to improving equitable accessibility to supermarkets and healthy retail food environments while eliminating food deserts.[80–83] Studies typically focused on associations between food environments and health suggest that physical access may be associated with a decreased risk of obesity.[84,85] Neighbourhood supermarket density has been associated with an increased intake in F&V,[86] as well as a lower body mass index (BMI) and a lower prevalence of obesity.[87,88] However, some research has suggested that the outcomes associated with such efforts is weak, and that opening a supermarket in a food desert does little to improve the availability of healthy foods.[89–91] Moreover, evidence of inequitable access to healthy food environments is not apparent in developed countries outside of the US.[82] For example, while Montréal demonstrates heavy inequalities in terms of neighbourhood income,[92] there are very few identified food deserts.[93]

In this light, it is thus important to consider as well the items sold within such outlets, rather than just the type of outlet and their accessibility.[94,95] Many have pointed out the plenitude and variety of healthful eating options that supermarkets stock: compared to smaller, discount supermarkets and convenience stores, large supermarkets offer the most healthful shopping environments in terms of variety, price, and quality, among other things.[79,96,97] This, however, does not guarantee healthy eating behaviours. Gustafson et. al. (2013) found that consumers who shopped at supermarkets reported a higher odds of consuming SSB compared to those who shopped at farmers' markets or specialty grocery stores, who had higher odds of purchasing F&V.[98]

Recognizing the different dietary preferences of shopper segments is also important in identifying potential health inequities: Chaix et. al. (2012) found that individuals from low-education neighbourhoods who shopped at discount supermarkets tended to have a higher BMI and waist circumference (WC).[99] A systematic review conducted by Giskes et. al. (2011) discovered that living in an economically deprived neighbourhood was the only consistent environmental factor associated with obesogenic dietary behaviours.[100] In this sense, supermarkets may actually exacerbate health inequities by reducing marginalised populations' ability to purchase healthful foods by encouraging the overconsumption of UPFPs.[96]

The display of unhealthy food products and their role in impulsive purchasing

Alongside an assortment of F&V, supermarkets are also responsible for carrying and heavily promoting a vast assortment of UPFPs all around the world. This may include the use of discount flyers and coupons or product placement in prominent aisle and checkout displays, activities that play a massive role in shaping both regular and impulse purchasing patterns. Acton et. al. (2017) estimated that in Canada, roughly two-thirds of displayed packaged foods and beverages available in grocery retail environments contained added sugar.[3] The sheer amount of shelf space occupied by UPFPs alone is enough for concern. In Montréal, UPFPs are disproportionately promoted inside supermarkets: one study found that UPFPs occupied up to 26 displays *outside* of their normal shelf location within the store, while F&V only appeared in 1 to 2 displays outside of their normal locations.[101] Farley et. al. (2009) found that the shelf space in supermarkets devoted to unhealthy snack foods is typically much greater than the shelf space devoted to F&V.[102] Shelf space availability of such energy-dense snack foods has been weakly correlated with BMI.[103]

Beyond this, many end-of-aisle and checkout displays are also frequented by unhealthy sweet and salty snacks as well as soft drinks: one study found that across 35 supermarkets, nearly 40% of all front-of-store end-of-aisle displays carried snack foods and that only five stores did not have snack foods at 100% of their checkouts.[104] Exposure to unhealthy foods is therefore largely unavoidable, as snack foods and soft drinks are prominently displayed in supermarkets across the globe.[105] One observational study found that a remarkable 30% of supermarket sales come from aisle-ends, and that placing beverages in end-of-aisle locations may inflate sales for beverages by up to 114%.[106] Supermarket chains also heavily promote unhealthy foods in their marketing catalogues and circulars.[107]

Between price and preference

Alongside brand and outlet loyalty, such in- and out-of-store stimuli play a huge role in precipitating unplanned buying.[108] However, not all checkout purchases are entirely impulsive — they may simply be a function of a shoppers' store-visit frequency.[109] Indeed, there are clear differences in spending habits and store loyalty between what the literature has called "routine" or "frequent" shoppers, who tend to shop at relatively fixed intervals and spend more per trip, and "random" or "less frequent" shoppers.[110,111]

In addition to frequency, it is also important to consider the affordability and social needs of consumers. With respect to the former, supermarkets may be more costly than other outlets: one study found that, on average, staple foods were generally more expensive in smaller stores than in supermarkets, with the exception of white bread.[112] Price is extremely important to consumers, who are generally highly adaptable when it comes to accommodating their financial needs.[113] Consumers also use other various criteria to assess the healthfulness of their groceries while shopping, which may include the explicit inclusion of healthy foods, the restriction of unhealthy foods, and balance between the two.[114] Furthermore, consumers often have to juggle the needs of their family while satisfying time requirements outside of work.[113] In seeking to meet these social needs, whether they be finance, security, or convenience-related, consumers will selectively shop at a range of stores. A lack of cooking skills has also been identified as a barrier to healthy shopping.[114] All of these practices, alongside the aforementioned activities of retail food outlets, materialize in consumers' dietary patterns, and for this reason are supermarkets a prime target of food policy intervention.

Interventions on food environments and their evaluation

Being an important risk factor for many NCDs, several interventions have sought to improve dietary quality through a number of fiscal, environmental, and informational interventions. The literature on these policies is too vast to cite in its entirety. Studies in this domain typically attempt to isolate the effects of a single policy on one or more groups of foods: for example, taxation and its effects on soft drink purchases or the impact of mandatory food labelling. In this sense, investigations into the purchasing trends of consumers or households are generally focused on a comparison of pre-intervention (baseline) purchasing trends to post-intervention trends. In general, interventions focused on restricting marketing of unhealthy food products have shown weak effects in terms of reduced consumption,[4–6] while other studies have linked either taxation or subsidies to reduced consumption or improved diet, respectively.[115–121] Hartmann-Boyce et. al. (2018) synthesized the results of 35 RCTs of retail grocery interventions, finding those that targeted price in some form or another seem to showed the most promise, compared to those that altered the store environment or simply provided consumer education without economic incentives.[122]

Whether or not this translates into an observable population shift in weight outcomes or diet-related NCDs is still debated, as many studies are still conducted in virtual supermarkets or a controlled laboratory setting.[116,121,123–126] This could result in findings that are not generalizable. Others rely on modeling; that is, simulation of effects based on reported expenditure surveys, dietary intake surveys using memory recall, or sales data from scanned or annotated receipts.[115,126] Collection of the data tends to be costly and time-consuming. Furthermore, many of the empirical studies that have been conducted may suffer from poor generalizability due to their omission of substitute and complementary purchases.[7,123,127]

This blind spot could result in an overestimation of the effects of these policies, as consumer loyalty to certain foods or strategic pricing by retailers and manufacturers may limit the effect of fiscal policies on the retail food environment. While such policies should continue to be evaluated as potential policy instruments, there has been a growing call to pursue more integrated, multi-component approaches to addressing unhealthy eating.[6,128–130] Sales data have shown promise in addressing these considerations.[124]

Surveillance and evaluation using grocery sales data

Substitutes and complements associated with soft drinks and snack foods

Much of the literature concerning patterns of healthy and unhealthy eating pertain specifically to the interventions in supermarkets. As previously mentioned, this literature is far too vast to cite in full. Of the sparse empirical research that has been published, substitute or complementary products are rarely considered. The study of purchasing patterns tends to be restricted to a small number of food categories. Few studies *directly* examine substitutes and complements of soft drinks and snack foods.[7,127] Those that do often only consider them in relation to the product of focus, such as soft drinks. For example, Andreyeva et. al. (2011) found a 24% reduction in SSB consumption after implementing a penny-per-ounce taxation plan, assuming no substitution to other beverages or food occurred.[115] Others focus on direct relationships between food categories vis-à-vis cross-price elasticity. For example, Colchero et. al. (2016) found that the implementation of the excise tax on SSB in Mexico was associated with a reduction in purchases of SSB and increases in the purchases of untaxed beverages.[131]

Studies in public health which consider the role complementary products are only quite recent. The phenomena is still not well-understood and many conflicting conclusions have been presented. Moreover, the relationship between “complementary” and “substitute” products do not depend on strict demarcations, especially within large, heterogeneous urban populations. Any conclusions predicated on such assumptions tend to be reductive, as food products may occupy the role of complements and substitutes from situation to situation. However, by identifying potential associated products within a consumer’s purchases, public health practitioners may be better-equipped to craft interventions in retail food environments that promote more holistic changes in diet. Unfortunately, however, the research to this end tends to be more heavily focused on substitutions and their limiting effects on intervention outcomes, with little discussion given to complementary foods. The little research that has been accomplished has achieved these goals by using large collections of sales data.

Leveraging sales data to investigate composition of purchased foods

Traditionally, supermarket sales data have been used to refine pricing and promotional strategies within stores.[132–135] However, several recent studies have also demonstrated the potential of sales data as a means of assessing overall household eating patterns for public health surveillance or the evaluation of interventions on the retail food environment.[136] These data are accessible from commercial, third-party companies — such as Kantar or Nielsen — that collect consumer purchase data in order to generate marketing insights, or they might be obtained directly from the retailer itself. These collection efforts have historically been limited in terms of size, cost, or methodology; however, given the growing availability of high-powered computing and the ease with which data may now be collected and stored, such concerns are no longer quite the barrier they were.[137] Moreover, in a departure from the traditional practice of nutritional epidemiology, there is a growing recognition that understanding dietary patterns and consumer behaviour, as opposed to examining individual foods or specific nutrients, may be more useful in terms of predicting disease risk insofar as a broader picture of consumption and health consciousness is provided.[124,138,139]

Assessment of food purchasing, either in the form of receipts, scanned food inventories, or sales data, may provide a reasonably accurate indicator of diet quality and dietary practices compared to food diaries and recalls.[140,141] The utility of such data comes not just from its novelty in public health applications, but also in its ability to move beyond the traditional scope of size and coverage in order to provide objective measures that may be used to evaluate 'real-world' interventions in an observational or quasi-experimental setting.[142] Linking supermarket sales data to nutritional information has been found to be feasible; however, there is question as to whether it is appropriate or useful to do so given the limitations of using purchasing as a proxy for diet.[141,143] Such data may suffer from questions pertaining to ownership, representativeness, validity, cost, lack of control over data acquisition, a lack of transparency, or the complexities of data management and processing.[142]

Sales data have been remarkably useful in characterizing trends of purchasing and analysing responses to fiscal or informational policies of large groups of consumers over long periods of time. Poti et. al. (2015) found that UPFPs were responsible for a sizable proportion of energy in purchases by US households, and higher levels of saturated fats, sugar, and sodium.[144] The authors used the Nielsen National Consumer Panel (NCP), formerly known as the Homescan Panel (or simply Homescan), linking processed and Ready-To-Eat (RTE) products to nutritional content. Using itemized purchase data from a representative sample of households in Great Britain, Smith et. al. (2017) found that increases in the price of chocolate, confectionery, and biscuits may lead to greater health gains than similar increases in SSB prices.[145] They found that reductions in one category had strong associations with reductions in other categories of food and beverages, thus suggesting a greater beneficial impact of price increases.

With Homescan, Piernas et. al. (2015) used a dynamic panel model with instrumental variables to examine associates of sweetened beverages, finding that every additional serving per day of calorie-sweetened beverages was associated with significantly higher purchases of caloric-sweetened desserts and sweeteners, with no difference in plain F&V.[146] By examining KWP data in Scotland, Whybrow et. al. (2018) identified six dietary patterns via principal component analysis (PCA): factor 1 showed a high correlation between purchases of convenience foods such as ready-made meals, chips, canned baked beans, savoury snacks, and diet soft drinks; higher amounts purchased of F&V, pasta, rice, sauce, and fish was also characterized by lower amounts of purchased sugary snacks, biscuits, and cake in factor 2.[147] Binkley and Golub (2007) found that consumers who frequently purchased diet soda also tended to purchase lower calorie versions of other foods and beverages, opting for products like low-fat milk, fruit, and yogurt over fruit juices and frozen fries.[148] However, diet soda prone consumers did not spend more or less on snack foods cookies, and candy compared to consumers who purchased regular soda. Moreover, consumers who purchased no soft drinks at all (few as they were), spent a significantly smaller portion of their budget on fruit juices, candies, and snack foods.

Others used such data to look at general "healthfulness" of shopping baskets, normally in response to some exposure: Pechey and Monsivais (2015) used Kantar WorldPanel (KWP) household data from to assess how choice of supermarket contributes to the healthfulness of food purchasing.[149] They discovered that the effect of supermarket choice was significant for purchases of F&V and less-healthy foods, and that healthier outcomes were observed for both frequent trips and fewer, small trips. In contrast, Stern et. al. (2016) used NCP data to categorize the food purchasing patterns of households and found that shopping primarily at supermarket grocery chains was not associated with a better nutrient profile of packaged food products for any racial or ethnic group.[150] Griffith et. al. (2015) used KWP to show that while preferences for vegetables, ready-made meals, and processed sweets remained high before and after the shock to world commodity prices in late 2007 to 2009, the demand for sugar and sugary foods declined drastically between 2006-07 to 2010-11, while preference for fruits increased.[151] However, the authors note that any changes in the nutritional profile of shoppers remain ambiguous.

One Danish study aligned Growth from Knowledge panel data (GfK) with registry employment data and discovered that unemployment led to substantial changes in diet composition.[152] Long term employment led to the substitution of fats and proteins with carbohydrates and added sugars. By using Homescan data, Grummon and Taillie (2017) found that US households participating in the federal Supplemental Nutrition Assistance Program (SNAP) generally exhibited less healthy purchases than groups of non-participating households.[153] The former purchased more calories from SSBs and processed meats and fewer calories from fruits and (non-starchy) vegetables. Quite recently, Aiello et. al. (2019) regressed food purchases in London down to the nutrient level against publiclyavailable prescription records used to derive health outcomes in the city.[154] They found that the amount of calories and nutrient variation moderately correlated with hypertension, diabetes, and high cholesterol. However, these authors were limited in terms of their choice of methodology, opting for a linear regression of area-level prescription data (as a measurefor chronic disease prevalence) against the nutrient diversity of purchased food products across London grocery stores.

Conclusions

To date, there is ample evidence that consumers' purchasing behaviour is clustered; that is, among individuals healthy or unhealthy purchases tends to cluster within baskets. Quantitative factors have previously investigated this phenomenon using proprietary sales data. Indeed, almost all previously mentioned studies relied on some form of household scanner data, which suffers from the fact that it is self-recorded — discrepancies in their reported expenditures have been well-noted.[155] These data typically use a randomly selected sample of households in which a household members scan the receipts of all purchases over one week. Items without a UPC are asked to be weighed and recorded manually. All of this recording places a tremendous burden on respondents, which increases risks of attrition and low response rates. Issues in coverage and selection bias arise from the sample of households selected by these panels, as there has been noted difficulty in recruiting specific groups.[137]

Other sources of data include store-based, point-of-purchase scanner data.[156] Unlike household scanner data, this retail data is collected from checkout scanners in a random sample of stores. However, purchases are only reported weekly: such temporal resolution means analysis must occur with respect to aggregate market trends, as information regarding individual transactions is not available. Objective monitoring of consumer purchasing patterns for individual transactions is therefore limited. Studies seeking to address this have typically suffered in terms of methodology, poor temporal or spatial resolution, selection bias, attrition bias, and small volume.

The knowledge gap these studies leave is therefore prima facie grounds for further evaluation. Given the high burden of diet-related non-communicable diseases and the priority of retail food environment regulation among governmental actors, there is a pressing need for understanding household purchasing behaviour using objective, empirical data, specifically where complementary products may play a role.

III. Manuscript: Complementary and co-occurring products associated with purchases of soft drinks and other snack foods: An analysis using household grocery purchasing data in Montréal, Canada

Preface

The results of this thesis are presented in the following manuscript:

Kody Crowell, Aman Verma, Hiroshi Mamiya, Amélie Quesnel-Vallée, Catherine Mah, David L. Buckeridge. *Complementary and co-occurring products associated with purchases of soft drinks and other snack foods: An analysis using household grocery purchasing data in Montréal, Canada.*

This manuscript has been targeted for publication in a public health journal. It addresses the objective of this thesis by analysing complementary and co-purchased products associated with soft drinks and snack foods using sales data from a grocery retailer in Montréal using a machine learning method known as association rule mining (ARM). Further information on ARM, the data, and the cohort of loyalty card members sampled is found in Appendix A while full tabulations of the ARM are provided in Appendix B.

Complementary and co-occurring products associated with purchases of soft drinks and other snack foods: An analysis using household grocery purchasing data in Montréal, Canada

Co-purchases of snack foods in Montréal, CA

Kody Crowell[1], Hiroshi Mamiya[1], Amélie Quesnel-Vallée[1], Catherine L. Mah[2], Aman Verma[1], David L. Buckeridge[1]*

[1] McGill University. Montréal, Canada
[2] Dalhousie University, Halifax, Canada
* Corresponding author,
Mailing Address: Clinical and Health Informatics Research
1140 Pine Avenue West, 3rd Floor
Montreal, Quebec, Canada H3A 1A3;
Email: david.buckeridge@mgcill.ca;
Phone: 1 (514) 934-1934 ext. 32991
Fax: 1 (514) 843-1551

Abstract

Background

Consumption of soft drinks and snack food contributes to the increasing global incidence of chronic illnesses such as cardiovascular diseases and type II diabetes. Previous studies addressing the purchasing patterns of highly-processed, nutrient-deficient foods have emphasized the importance of co-purchased products, as they may undermine interventions seeking to limit the intake of unhealthy foods. However, research using household-level transaction data to analyze such patterns in purchasing has traditionally been limited in terms of volume and representativeness. The objective of this study is to identify trends in food groups purchased by households together with, or complementary to, soft drinks and snack foods in comparison to fresh fruits and vegetables.

Methods

We used longitudinal, household-level transaction data from 14,999 loyalty card members of a large grocery retailer in Montréal, Canada between February 2015 and September 2017 (1,522,501 transactions). Association rule mining was used to identify frequently co-purchased item categories for soft drinks, snack foods, juice, fruits, and vegetables.

Results

Transactions containing snack foods and soft drinks were also likely to contain canned or highly-processed, nutrient-deficient foods. For example, soft drinks were highly associated with salty snacks (confidence: 17%; Odds Ratio [OR]: 1.82 ± 0.02), bottled water (confidence: 16%, OR: 1.77 ± 0.02), and frozen meals and sides (confidence: 16%; OR: 1.78 ± 0.03). Conversely, purchases with qualitatively healthier foods were found to be associated with purchases of fruits and vegetables: purchases with vegetables were highly associated with fresh herbs (confidence: 84%; OR: 1.90 ± 0.03) and packaged salads (confidence: 73%; OR: 1.61 ± 0.01).

Conclusions

These empirical results quantify the extent to which food-purchasing behaviours cluster within baskets among households. Public health researchers and practitioners seeking to design interventions that decrease the frequency of unhealthy food purchases in the supermarket environment should consider the tendency for unhealthier or healthier foods to be purchased concurrently.

Background

Studies have linked dietary patterns to non-communicable diseases (NCD) such as obesity, type II diabetes, and cardiovascular diseases.[1,2] Yet in Canada, the consumption of nutrient-poor and energy dense ultra-processed foods (those high in refined carbohydrates, sodium, and trans and saturated fats) supplies an estimated 48% of the population's caloric intake.[3] In 2011, approximately 35% of Canadians' sugar intake came from sources other than fruits and vegetables, including soft drinks (carbonated beverages) and candy.[4] Particularly concerning are the trends among children and adolescents, for whom the consumption of beverages (milk, fruit juices, energy and sports drinks, soft drinks) comprised 44% of their daily sugar intake. Meanwhile, the consumption of fruits and vegetables among Canadians dropped 13% between 2004 and 2015.[5]

Recent studies have considered how the retail food environment influences individual behaviour related to food purchasing.[6,7] Supermarkets, in particular, maintain a heavy presence in urban landscapes and play a major role in food purchasing.[6] In Canada, research in this domain is quite new: for the most part, studies tend to be observational and cross-sectional, with inferences made at the population level.[7] Moreover, supermarket venues are frequently presented as 'healthy' because they carry a greater variety of foods such as unprocessed fruits and vegetables compared to other outlets.[8] However, supermarkets also devote large amounts of shelf space and advertising to promoting soft drinks and snack foods, often alongside each other to promote co-purchasing as complementary products.[8–10] These foods have become the principal targets of many public health interventions.[11]

Evaluations of such interventions that look at purchasing behaviour typically focus on changes in purchasing patterns of a single food category (such as soft drinks); the comprehensive set of products in a food basket is rarely monitored.[12] Yet many food products are frequently bought together or in substitution to one another.[13] As a result, any intervention targeting a single food category such as sugar-sweetened beverages may have a limited ability to improve the overall diet of the population. To craft broader interventions aimed at reducing purchases of snack foods and soft drinks in the retail grocery environment, public health agencies and researchers must understand how consumers interact with products that are substitutes for, or complementary to, the targeted foods. Indeed, the co-occurrence of products within a shopping basket reflects the multicategory decisions made by consumers in specific contexts.[14]

The knowledge gap regarding patterns in household purchasing behaviour, particularly the role of complementary products, limits inferences regarding how these behaviours influence household food consumption. Studies addressing this knowledge gap have traditionally relied on surveys, annotated receipts, or household scanner data such as Nielsen Homescan or Kantar WorldPanel (KWP), all of which may suffer from attrition, selection bias, recall bias, poor coverage, or poor spatial resolution.[15,16] The cost of accessing such data and the ensuing analytical complexity have also limited the contribution of studies' use these data sources.[16] On the other hand, point-of-purchase (i.e. store-level), commercial scanner data provides an objective measure of purchasing yet aggregates sales on a weekly basis, thereby masking the rich variation of longitudinal purchasing patterns across and within consumers.[16,17] Observational sales data obtained directly from the retailer addresses these limitations, providing objective measures of food purchasing at the level of individual transactions, thus illuminating mechanisms of association between the retail food environment and consumer purchasing behaviors.[7] In this regard, supermarket sales data and the loyalty cards that link household purchases over time can be invaluable sources of information for public health practitioners to measure and monitor patterns of food purchasing.

The objective of this study is to identify patterns in grocery item co-purchasing among households. Specifically, we seek to identify associated product categories within purchases by applying unsupervised machine learning to household grocery transaction records for members of a loyalty program in a large retail grocery chain in Montréal. We resist the call to explain the association of products causally and instead turn our focus toward the empirical distribution of these associations across loyalty card members and their many transactions. The results provide insight into the co-occurring purchasing of food products among consumers visiting in a large supermarket chain.

Methods

We acquired loyalty card and transaction records from a large chain (one that operates across multiple provinces) of grocery supermarkets in Montréal, Canada between February 1, 2015, and September 30, 2017. We selected all 14,999 anonymized loyalty card members who made at least one purchase during the two-year period. Each included member was thereafter linked to their respective transactions, which awee time-stamped for each purchasing occasion and contained a set of purchased grocery products as uniquely defined by Universal Product Codes (UPC). A transaction — the collection of these products — is hereafter simultaneously referred to as a "basket." Note that baskets linked by a loyalty card may reflect the purchases of more than one individual. Indeed, card members holding multiple cards were previously linked by the retailer; this linkage was used for analysis of transactions. Since transactions involving these cards may entail purchases by more than one individual, we refer to our members as "households."

To avoid redundancies in our analysis, the 65,270 UPCs in the database of transactions were linked to 68 retail product categories collected from the store website (e.g. a red apple and a green apple, both present in the disaggregated data, are assigned to the same category). Note that these product categories were retailer-defined, and not based on food or nutrient content of products. Associations between purchasing of these product categories were mined using an unsupervised machine learning technique known as association rule learning.

Association rules are generated using a nonparametric learning algorithm that identifies the conditional relationships between items in a dataset.[18] In the context of a grocery basket, these associations are identified empirically through the co-purchasing of products among transactions. The relevance of these rules is determined by metrics of probability and proportion for different sets of items (called itemsets): if a basket contains an itemset X (called the antecedent), then it will also contain item Y (called the consequent) with certain probabilities *C%* and *S%*. *Support* (called *S%* in the example above) is the proportion of baskets that contain both X and Y; that is, $P(X \text{ AND } Y)$. It is a measure of the prevalence of an itemset. The *confidence* of a rule $X \rightarrow Y$ (called *C%* above) is the proportion of baskets containing X that also contain Y; that is, the conditional probability $P(Y \mid X)$. It indicates how often a rule holds true among baskets that contain the antecedent. Lastly, the *lift* (or *interest*) of a rule is the ratio of the confidence to the expected confidence, the proportion of all baskets that contain just Y; that is, $P(Y \mid X) / P(Y)$. Lift is a measure of the importance of a rule: a lift greater than 1 implies that the antecedent and consequent co-occur more often than expected, indicating a stronger association. Note that lift is bidirectional; that is, it will be the same for the rules $X \rightarrow Y$ and $Y \rightarrow X$. "Strong" rules are typically defined using these metrics of probability. We also computed the odds ratio (OR) and variance for each association (an OR of 1 indicates that the two itemsets X and Y are not associated).

To implement association rule learning, we applied the *apriori algorithm* as implemented in the arules package (version 1.6-3) in the statistical software R (version 3.4.4).[19] Association rules are discovered for a user-defined minimum support and confidence. If the support threshold is too low, rare itemsets may be picked up and given an extraordinarily high lift, even if the itemsets occur by chance. To filter out association rules with high variance, the computation was repeated with different thresholds for minimum support and confidence. Results shown are with a minimum support of 0.01 and a minimum confidence equal to the support of the consequents in all transactions. Note that association rules are not causal, merely associative. Although the rules underscore correlations in purchasing using conditional probabilities, a relationship $X \rightarrow Y$ does not imply that X is a precedent of Y, which would be a necessary criterion for establishing causality. We focused specifically on co-purchases of highly processed food, namely soft drinks, salty snacks, sweet snacks, and juices. These categories include products such as chips, crackers, chocolates, cookies, candies, confectionery, fruit juices, and other snack foods that are often high in refined carbohydrates or sodium. These products are of interest given their high priority in advertising and promotion in retail spaces. For comparison, vegetables and fruits were also considered.

Results

1,522,501 transactions were linked to the 14,999 loyalty card members. Only 38% of all product UPCs within these transactions were successfully matched to retail categories (see Table 1 for a list of all categories); however, these UPCs covered over 96% of purchased items in the database. Of the 1,522,501 transactions, 22,337 contained only non-food items. In general, households spent a median $24.22 (95% CI: $2.93 - $157.55) per visit, purchasing a median of 7 (95% CI: 1 - 40) items. The top fifteen most frequently purchased product categories across all transactions, given in Table 2, are vegetables (42% of transactions), fruits (42%), milk and cream (30%), sweet snacks (22%), packaged bread (20%), juice (19%), packaged cheese (19%), salty snacks (19%), desserts and pastries (17%), ready-to-eat meals (17%), beef (17%), yogurt (17%), deli meats (16%), spices (14%), and condiments and toppings (13%). Table 2 also shows 338 UPCs related to soft drinks that are found in 135,308 transactions (9%).

We began with a 0.01 support threshold and a confidence threshold equal to that of the expected confidence of the consequent. Omitting non-food items and unmatched UPCs (as they would otherwise dominate the association rules) left 1,500,164 transactions with only food items. Across all categories used as consequents, 272 rules were found. Of all transactions, Table 3 shows the top ten results of the algorithm on household baskets for 1-1 associations in order of descending lift, where the consequents are soft drinks, salty snacks, sweet snacks, juice, fruits, and vegetables. The number of households that supported these associations are also given.

Salty snacks and soft drinks were likely to co-occur in a basket (support: 3.3%; lift: 1.9), as were sweet snacks and soft drinks (support: 3.3%; lift: 1.6), and sweet snacks and salty snacks (support: 7.3%; lift: 1.7). In general, transactions with soft drinks, salty snacks, and sweet snacks were likely to contain item categories high in fat, sodium, or sugar: canned or highly-processed, ready-to-eat meals or sides, and nutrient-deficient frozen foods. Soft drinks were associated with water (support: 1.5%, lift: 1.8); frozen meals and sides (support: 1.2%, lift: 1.8); sausages and bacon (support: 1.4%, lift: 1.8); and ice cream (support: 1.2%, lift: 1.7).

Meanwhile, snack foods were correlated with other “convenience” goods: salty snacks were heavily associated with antipasto/dips (support: 2.4%, lift: 2.0), nuts/seeds/dried fruits (support: 2.3%, lift: 1.9) and canned soups (support: 1.90%, lift: 1.9), while sweet snacks shows association with nuts/seeds/dried fruits (support: 2.8%, lift: 2.0) and canned soups (support: 2.1%, lift: 1.8). Fruit juices were strongly correlated with canned foods.

Using more prevalent food categories as consequents, such as vegetables and fruits, yields higher confidence values with similar lifts. Indeed, vegetables were associated with more minimally processed foods such as fresh herbs (support: 3.4%, lift: 2.0) and packaged salads (support: 6.8%, lift: 1.7); while fruit was heavily associated with fresh herbs (support: 2.8%, lift: 1.7), frozen fruits (support: 1.2%, lift: 1.6), and packaged salads (support: 6.1%, lift: 1.6). The strengths of the associations for fruits and vegetables were generally lower than those found for the other categories.

Discussion

We used supermarket transaction data in Montréal, Canada to measure household co-purchasing patterns within supermarket environments. Using association rule learning with many transactions, we found that sweet snacks, salty snacks, and soft drinks were more likely to be purchased together with foods that were qualitatively unhealthy compared to transactions with fruits and vegetables. Foods associated with sweet and salty snacks could be considered other convenience snack foods. Meanwhile, associations with fruits and vegetables tended to reflect more minimally processed products that were qualitatively healthier. These findings suggest that food purchasing behaviour is clustered within baskets in households, who tend to buy multiple unhealthy or healthy products at the same time. Such results are important insofar as they reveal broad patterns of purchasing of multiple categories in a diverse urban setting using data from real supermarket baskets.

Without overstating the significance of the discovered rules in Table 3, it is important to recognize that these associations do not necessarily reflect the attitudes, intentions, or health-consciousness of the households themselves. Moreover, associations cannot distinguish between which products are complementary and which products are merely coincidental.[14] For example, associations between snack foods and convenience foods like frozen and canned meals *could* imply that unhealthier eating habits such as consuming snack foods and favoring processed foods over cooking "from scratch" cluster together, maybe due to a lack of time, energy, or knowledge.[20] Alternatively, they could simply be the result of many smaller purchases meant to fill in missing ingredients.

Sweet snacks are one of the most frequently purchased items in the dataset in terms of the number of unique UPCs, yet the associations with sweet snacks do not immediately reveal clustering with other food groups. This may be due to the size and variety of the category: sweet snacks cover candies, chocolates, cookies, bulk sweets, packaged sweets, etc. Where the purchase of more unprocessed vegetables in a basket may be reflective of more routine, planned shopping trips, individual products within sweet snacks may refer to candies and chocolate bars featured at the checkout, impulsively purchased with larger and otherwise healthier baskets.[10,21] Both types of shopping (planned and unplanned) manifest within individuals in different situations and are not necessarily indicative of overall health behaviours: unplanned purchases, either unhealthy or healthy, may signal impulsive purchases or simple "out-of-stock" inventory additions.[21] In any case, such unplanned purchases will impact the strength of the association rules.

Few studies have explicitly searched for co-purchases to soft drinks and snack foods, especially where it concerns the entire basket composition. Ranjit et. al. used a cross-sectional survey to show that the consumption of soft drinks and flavoured sports beverages were associated with unhealthy dietary practices among youth.[22] Piernas et. al. used Nielsen Homescan data to identify associates of sugar-sweetened beverages, finding that every additional serving per day of calorie-sweetened beverages was associated with significantly higher purchases of desserts and other sweeteners.[23] More recently, Whybrow et. al. used KWP data to show correlations between purchases of convenience foods such as ready-to-eat meals, chips, and salty snacks while similarly showing that a higher amount of purchased fruits and vegetables was associated with a lower purchased amount of sugary snacks.[24] Other studies have sought to identify a clustering of both unhealthy and healthy behaviours pertaining to both diet and physical activity, thereby suggesting that health promotion should target multiple health risk behaviours.[25]

Although marketing researchers have long understood the value of finding associations between products in order to better target promotions,[26] the use of supermarket sales data in public health to characterize consumer purchasing patterns for a broad range of products in supermarkets is novel. With current computing resources, it is feasible to evaluate large retail data, assuming public health researchers can access them. Studies such as those by Piernas et. al. (2015) and Whybrow et. al. (2018) rely on commercial panel data, while others yet use simulated data from virtual supermarkets.[16,23,24]

The data used in this study thus carries great potential to elucidate the patterns of purchasing that would be unavailable from other data sources. Moreover, retail transactions represent an objective measure of purchasing since they does not rely on recall or self-reporting, and the linking of card numbers across several transactions allows for longitudinal analyses that would otherwise require more expensive and invasive methods such as the receipt scanning in many commercial data (which itself does not have the same level of spatial and temporal resolution).[7,27] The use of loyalty card numbers also separates frequent from non-frequent shoppers in order to identify nutritional habits among different segments of the population.

There are limitations to this study and to the use of supermarket sales data that must be acknowledged. The strength of association rule mining is that it obtains these relationships from large databases that would otherwise not be analyzed exhaustively using traditional methods. However, this is also a drawback of unsupervised learning: the optimal support and confidence thresholds are not truly known. Further studies should verify these purchasing patterns in other supermarkets.

Moreover, any change in the categorization of the UPC codes would result in a different set of association rules. Although a hierarchy of categories is needed to discover interpretable rules between purchases, there are many ways to classify foods, and within each classification, there is a great deal of ambiguity in their perceived and real "healthiness." More consideration should be given to the role of in-store features with respect to their characterization of "healthy" and "unhealthy" foods and subsequent consumer purchasing practices within these spaces. Greater attention should also be given to the other valuations of food beyond mere "healthfulness." Co-purchases of soft drinks and salty snacks, for example, do not necessarily reflect an individual's diet, but rather a social practice with material and symbolic elements that transcend simple functional understandings of food. Future work should also investigate the situational behaviours of consumers, looking at how timing factors into decision-making.

Note that transactions linked by a loyalty card may reflect the purchases of more than one individual, as no information on the individuals making the purchases is given. For this reason, we assume that our unit of analysis is localized to a "household" and refer to it in this way. With no way of distinguishing loyalty card use between multiple individuals, discovered association rules do not apply to one consumer. Furthermore, this study does not consider purchases from other retail environments. Therefore, our supermarket transaction data only represents one part of a household's purchasing behaviour — and even then, purchasing can only ever be a proxy for actual diet.[28] It is quite possible that some food categories are more likely to be purchased outside of supermarkets (soft drinks and snack foods are widely available outside of supermarkets, for example in convenience stores). Future research should thus seek to use supermarket transaction data alongside transactions from other retail formats (or store types) to improve evaluations of retail interventions. One need not stop at single transactions, either: the temporal component of these transactions can be further studied to reveal substitution and complementarity responses to promotions and price.

One implication of this finding is that regulations targeting single elements of grocery food environments, such as soda taxation, may result in unintended consequences in the purchasing of target food categories and their correlates. When it comes to reducing the burden of dietary risk factors, co-occurring or complementary products may undermine these public health interventions. Although association rules do not imply causality, a correlation in purchasing could also lead to an additional benefit vis-à-vis a decrease in the purchasing of co-occurring food products if they are truly complementary. On the other hand, reduction of the purchasing of target food categories, such as soft drinks, maybe interfered without policies to discourage associated food categories. Needless to say, the findings thus motivate further investigation of causal association across multiple food categories that determine healthiness of overall basket composition. In addition, out results stress the importance of monitoring comprehensive set of food categories as part of evaluation for nutrition-related public health intervention.

In general, more attention should be given to understanding the larger marketing practices that guide individual decision-making in retail environments. There is a need for a better definition of healthiness that is more inclusive and that considers food practices outside of strict nutritional reductionism. A better solution may yet be found in transformative food movements that seek to undermine the hegemony of supermarkets in today's industrialized and commoditized society. In any case, understanding individual purchasing behaviour in such spaces is crucial. As mentioned, previous studies have often focused on the purchases within food categories in isolation, and as a result have not considered the different kinds of transactions in which these purchases occur (whether they be convenience purchases, routine trips, fill-in inventory additions, or impulse purchases). It is for these reasons that retail grocery data hold a great potential for research and surveillance of population nutrition. Given the growing use of supermarket transcation data in the public health realm, these data show great promise in identifying co-purchases of products to unhealthy and healthy foods and could, therefore, be used to craft more effective interventions in the retail grocery environment.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data are confidential as the agreement to access was only given to the investigators. Information regarding the data is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DB conceived the project and collected the data. KC performed all data processing and analysis with assistance from AV. DB and CM encouraged KC to investigate complementarity and supervised the findings of the work. KC wrote the manuscript with the input from all authors.

Acknowledgements

We would like to thank Yu Ma for his input into the project and work matching the unique product codes in the dataset to retail categories category matching.

Funding

KC was funded by a McGill University research stipend and a Tri-Council Canadian Graduate Scholarship from the Canadian Institutes for Health Research. DB was funded by the Public Health Agency of Canada (arrangement number 1516-HQ-000069). These sources of funding played no part in the design, analysis, or interpretation of the data nor in

the writing of the manuscript.

References

1. Micha R, Shulkin ML, Peñalvo JL, et al. Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the Nutrition and Chronic Diseases Expert Group (NutriCoDE). *PloS One*. 2017;12(4):e0175149. doi:10.1371/journal.pone.0175149

2. Willett WC, Stampfer MJ. Current evidence on healthy eating. *Annu Rev Public Health*. 2013;34:77-95. doi:10.1146/annurev-publhealth-031811-124646

3. Moubarac J-C, Batal M, Louzada ML, Martinez Steele E, Monteiro CA. Consumption of ultra-processed foods predicts diet quality in Canada. *Appetite*. 2017;108:512-520. doi:10.1016/j.appet.2016.11.006

4. Langlois K, Garriguet D. Sugar consumption among Canadians of all ages. *Health Rep*. 2011;22(3):23-27.

5. Tugault-Lafleur CN, Black JL. Differences in the Quantity and Types of Foods and Beverages Consumed by Canadians between 2004 and 2015. *Nutrients*. 2019;11(3):526. doi:10.3390/nu11030526

6. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating Healthy Food and Eating Environments: Policy and Environmental Approaches. *Annu Rev Public Health*. 2008;29(1):253-272. doi:10.1146/annurev.publhealth.29.020907.090926

7. Minaker LM, Shuh A, Olstad DL, Engler-Stringer R, Black JL, Mah CL. Retail food environments research in Canada: A scoping review. *Can J Public Health*. 2016;107:4. doi:10.17269/cjph.107.5344

8. Farley TA, Rice J, Bodor JN, Cohen DA, Bluthenthal RN, Rose D. Measuring the Food Environment: Shelf Space of Fruits, Vegetables, and Snack Foods in Stores. *J Urban Health Bull N Y Acad Med*. 2009;86(5):672-682. doi:10.1007/s11524-009-9390-3

9. Charlton EL, Kähkönen LA, Sacks G, Cameron AJ. Supermarkets and unhealthy food marketing: An international comparison of the content of supermarket catalogues/circulars. *Prev Med*. 2015:6.

10. Thornton LE, Cameron AJ, McNaughton SA, Worsley A, Crawford DA. The availability of snack food displays that may trigger impulse purchases in Melbourne supermarkets. *BMC Public Health*. 2012;12(1):194. doi:10.1186/1471-2458-12-194

11. Niebylski ML, Redburn KA, Duhaney T, Campbell NR. Healthy food subsidies and unhealthy food taxation: A systematic review of the evidence. *Nutrition*. 2015;31(6):787-795. doi:10.1016/j.nut.2014.12.010

12. Cameron AJ, Charlton E, Ngan WW, Sacks G. A Systematic Review o the Effectiveness of Supermarket-Based Interventions Involving Product, Promotion, or Place on the Healthiness of Consumer Purchases. *Current Nutrition Reports*. 2016;5(3):129-138. doi:10.1007/s13668-016-0172-8

13. Fletcher J, Frisvold D, Tefft N. Substitution Patterns Can Limit the Effects of Sugar-

Sweetened Beverage Taxes on Obesity. *Prev Chronic Dis*. 2013;10. doi:10.5888/pcd10.120195
14. Manchanda P, Ansari A, Gupta S. The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions. *Mark Sci*. 1999;18(2):95-114. doi:10.1287/mksc.18.2.95
15. French SA, Wall M, Mitchell NR, Shimotsu ST, Welsh E. Annotated receipts capture household food purchases from a broad range of sources. *Int J Behav Nutr Phys Act*. 2009:11.
16. Bandy L, Adhikari V, Jebb S, Rayner M. The use of commercial food purchase data for public health nutrition research: A systematic review. *PLOS ONE*. 2019;14(1):e0210192. doi:10.1371/journal.pone.0210192
17. Mamiya H, Moodie EEM, Buckeridge DL. A novel application of point-of-sales grocery transaction data to enhance community nutrition monitoring. *AMIA Annu Symp Proc*. 2018;2017:1253-1261.
18. Agrawal R, Imieliński T, Swami A. Mining Association Rules Between Sets of Items in Large Databases. In: *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*. SIGMOD '93. New York, NY, USA: ACM; 1993:207–216. doi:10.1145/170035.170072
19. Hahsler M, Grün B, Hornik K. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. *J Stat Softw*. 2005;14(1):1-25. doi:10.18637/jss.v014.i15
20. Nestle M, Wing R, Birch L, et al. Behavioral and Social Influences on Food Choice. *Nutr Rev*. 1998;56(5):50-64. doi:10.1111/j.1753-4887.1998.tb01732.x
21. Kollat DT, Willett RP. Customer Impulse Purchasing Behavior. *J Mark Res*. 1967;4(1):21-31. doi:10.1177/002224376700400102
22. Ranjit N, Evans MH, Byrd-Williams C, Evans AE, Hoelscher DM. Dietary and Activity Correlates of Sugar-Sweetened Beverage Consumption Among Adolescents. *Pediatrics*. 2010;126(4):e754-e761. doi:10.1542/peds.2010-1229
23. Piernas C, Ng SW, Mendez MA, Gordon-Larsen P, Popkin BM. A Dynamic Panel Model of the Associations of Sweetened Beverage Purchases With Dietary Quality and Food-Purchasing Patterns. *Am J Epidemiol*. 2015;181(9):661-671. doi:10.1093/aje/kwu317
24. Whybrow S, Craig L, Macdiarmid J. Dietary patterns of households in Scotland: Differences by level of deprivation and associations with dietary goals. *Nutr Health*. 2018;24(1):29-35. doi:10.1177/0260106017745389
25. Leech RM, McNaughton SA, Timperio A. The clustering of diet, physical activity and sedentary behavior in children and adolescents: a review. *Int J Behav Nutr Phys Act*. 2014;11(1):4. doi:10.1186/1479-5868-11-4
26. Lattin JM, McAlister L. Using a Variety-Seeking Model to Identify Substitute and Complementary Relationships among Competing Products. *J Mark Res*. 1985;22(3):330-339. doi:10.2307/3151429
27. Lytle LA, Sokol RL. Measures of the food environment: A systematic review of the

field, 2007–2015. *Health Place*. 2017;44:18-34. doi:10.1016/j.healthplace.2016.12.007
28. Appelhans BM, French SA, Tangney CC, Powell LM, Wang Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? *Int J Behav Nutr Phys Act*. 2017;14(1):46. doi:10.1186/s12966-017-0502-2

Tables

Table 1. Available categories for food products and the number of UPCs associated with each. Categories are presented in alphabetical order according to retail grocery department. These categories are used in the ARM algorithm. Note that the number of UPC products per category does not necessarily correlate with their prevalence among transactions: packaged bread, for example, is not particularly diverse in terms of number of UPCs per category; however it is one of the most frequently purchased items in the sampled member's transactions.

Department	Category	No. UPCs
Snacks	Gluten-free snacks	107
	Nuts, seeds, and dried fruits	1025
	Salty snacks	1278
	Sweet snacks and candy	3888
Pantry	Baking ingredients	900
	Breakfast foods	39
	Canned fruits	76
	Canned meats	29
	Canned soups	440
	Canned vegetables	440
	Canned seafood	405
	Cereal	174
	Cereal bars	122
	Condiments and toppings	1344
	Oils and vinegars	689
	Pasta, rice, and beans	1129
	Spices, herbs, and sauces	2193
	Spreads and syrups	1001
Bread and bakery	Baked bread and baguettes	157
	Buns and rolls	165
	Desserts, pastries, baked goods	1513
	Packaged bread	436
	Tortillas and flatbreads	229
Frozen	Frozen baked goods	582
	Frozen beverages	65
	Frozen fish and seafood	523

	Frozen fruits	75
	Frozen meals and sides	723
	Frozen meat and poultry	333
	Frozen pizza and pasta	392
	Frozen vegetables	127
	Ice cream treats	615
Meat and poultry	Beef and veal	348
	Chicken and turkey	316
	Lamb, horse, and game	66
	Pork	245
	Rabbit and fowl	46
	Sausages and bacon	439
Beverages	Beer and cider	1312
	Drink mixes	142
	Fruit juices and drinks	1125
	Soft drinks	433
	Soy, rice, and nut beverages	109
	Sports and energy drinks	214
	Tea and hot drinks	1187
	Water	366
	Wines, cocktails, and coolers	850
Dairy and eggs	Butter and margarine	109
	Cheese counter	406
	Eggs and substitutes	103
	Milk and cream	492
	Packaged cheese	1146
	Sour cream dips	160
	Yogurt	605
Deli and prepared meals	Antipasto, dips, and pâtés	270
	Deli meats	699
	Ready meals and sides	1678
Vegan and vegetarian	Vegan and vegetarian foods	122
World cuisine	World cuisine products	275
Fish and seafood	Fish and seafood	498
Fruits and vegetables	Fresh herbs	202
	Fruits	1193
	Packaged salad mixes	180
	Vegetables	1490

Table 2. 15 most frequently purchased item categories as well as soft drinks showing the number of unique UPCs that comprise the category, the total number of transactions in which the category appears, and the number of loyalty card members (households) that purchased in this category.

Category	**UPCs (%)**	**Transactions (%)** Total: 1,522,501	**Households** Total: 14,999
Vegetables	1,224 (1.9)	635,862 (42)	13,972
Fruits	911 (1.4)	637,243 (42)	13,926
Milk/cream	402 (0.6)	460,569 (30)	13,199
Sweet snacks	3,183 (4.9)	329,791 (22)	13,023
Packaged bread	379 (0.6)	304,881 (20)	13,275
Juice	981 (1.5)	294,054 (19)	13,733
Packaged cheese	1,031 (1.6)	292,627 (19)	13,761
Salty snacks	1,195 (1.8)	287,194 (19)	13,543
Desserts/pastries	1,334 (2.0)	257,792 (17)	13,478
Ready meals	1,509 (2.3)	257,293 (17)	12,015
Beef/veal	308 (0.5)	256,339 (17)	11,918
Yogurt	566 (0.9)	255,287 (17)	12,120
Deli meats	653 (1.0)	239,804 (16)	11,709
Spices	1,798 (2.8)	209,688 (14)	12,296
Condiments	1,077 (1.7)	193,225 (13)	12,332
Soft drinks	388 (0.6)	135,308 (9)	9,702

Table 3. The top twelve results of a 1-1 association rule mining applied to six different food categories as consequents according to descending lift.

Support	Confidence	Lift	Antecedent	OR (95% CI)	Households
1-1 association rules for soft drinks					
3.3	17	1.88	Salty snacks	1.82 (1.80, 1.84)	6,556
1.5	16	1.80	Water	1.77 (1.75, 1.80)	4,279
1.2	16	1.80	Frozen meals and sides	1.78 (1.75, 1.81)	3,704
1.4	16	1.79	Sausages/bacon	1.76 (1.74, 1.79)	4,240
1.2	15	1.70	Butter/margarine	1.68 (1.65, 1.71)	4,029
1.0	15	1.69	Buns/rolls	1.67 (1.64, 1.70)	3,602
1.2	15	1.67	Ice cream	1.65 (1.62, 1.67)	4,114
3.3	15	1.64	Sweet snacks	1.59 (1.57, 1.60)	6,220
2.9	15	1.62	Juice	1.58 (1.56, 1.60)	6,261
1.9	15	1.61	Condiments/toppings	1.58 (1.56, 1.61)	5,418
1.0	14	1.55	Tea/hot drinks	1.53 (1.51, 1.56)	3,406
2.4	14	1.53	Desserts/pastries	1.50 (1.48, 1.52)	5,373
1-1 association rules for salty snacks					
2.4	39	2.03	Antipasto/dips	1.98 (1.96, 2.01)	6,522
2.3	37	1.94	Nuts/seeds/dried fruits	1.89 (1.87, 1.92)	6,322
1.9	37	1.92	Canned soup	1.88 (1.86, 1.91)	5,690
3.3	36	1.88	Soft drinks	1.82 (1.80, 1.84)	6,556
1.4	36	1.88	Canned seafood	1.85 (1.82, 1.89)	4,842
4.6	35	1.85	Condiments/toppings	1.77 (1.75, 1.79)	8,938
2.9	34	1.79	Cheese	1.74 (1.72, 1.76)	6,863
1.6	33	1.74	Tortillas/flat breads	1.71 (1.69, 1.74)	5,155
7.3	33	1.73	Sweet snacks	1.60 (1.59, 1.61)	9,638
2.8	33	1.72	Sausages/bacon	1.67 (1.65, 1.69)	7,062
2.8	33	1.72	Cereals	1.67 (1.65, 1.69)	6,767
2.7	32	1.67	Canned vegetables	1.62 (1.60, 1.64)	6,957
1-1 association rules for sweet snacks					
2.8	45	2.05	Nuts/seeds/dried fruits	1.99 (1.97, 2.02)	6,827
3.5	41	1.87	Cereals	1.80 (1.78, 1.82)	7,366
2.1	40	1.83	Canned soup	1.79 (1.77, 1.81)	5,750
3.7	39	1.77	Spreads/syrups	1.71 (1.69, 1.73)	8,195
1.5	39	1.76	Canned seafood	1.73 (1.70, 1.76)	4,860
7.3	38	1.73	Salty snacks	1.60 (1.59, 1.61)	9,638
2.8	37	1.69	Tea/hot drinks	1.64 (1.63, 1.66)	6,632
3.1	37	1.67	Canned vegetables	1.62 (1.60, 1.64)	7,265
2.8	37	1.67	Butter/margarine	1.62 (1.60, 1.64)	7,376
2.2	37	1.67	Antipasto/dips	1.63 (1.61, 1.65)	6,154
2.7	37	1.67	Frozen meals	1.62 (1.60, 1.64)	6,488
2.9	37	1.66	Baking ingredients	1.61 (1.60, 1.63)	7,569

1-1 association rules for juice					
2.0	37	1.92	Canned soup	1.88 (1.85, 1.91)	5,585
1.4	37	1.90	Canned seafood	1.87 (1.84, 1.90)	4,633
3.2	37	1.90	Cereals	1.83 (1.81, 1.85)	6,883
3.1	37	1.89	Canned vegetables	1.82 (1.80, 1.85)	7,253
2.3	36	1.88	Nuts/seeds/dried fruits	1.82 (1.80, 1.84)	6,071
2.7	35	1.86	Butter/margarine	1.72 (1.70, 1.74)	7,158
3.3	35	1.77	Spreads/syrups	1.71 (1.69, 1.73)	7,742
5.9	34	1.77	Yogurt	1.66 (1.64, 1.67)	8,820
4.3	34	1.76	Pasta/rice/beans	1.67 (1.65, 1.68)	8,417
2.1	34	1.74	Antipasto/dips	1.69 (1.66, 1.71)	5,856
4.2	33	1.72	Eggs	1.63 (1.62, 1.65)	8,696
4.3	33	1.71	Condiments/toppings	1.62 (1.61, 1.64)	8,588
1-1 association rules for vegetables					
3.4	84	1.97	Fresh herbs	1.90 (1.88, 1.93)	7,117
6.8	73	1.73	Packaged salads	1.61 (1.60, 1.62)	9,374
1.5	71	1.68	Vegan/vegetarian foods	1.65 (1.62, 1.68)	3,742
5.8	69	1.63	Canned vegetables	1.54 (1.52, 1.55)	9,469
3.3	68	1.60	Tortillas/flat breads	1.55 (1.53, 1.57)	7,095
2.0	68	1.59	'World cuisine' foods	1.56 (1.54, 1.59)	6,401
5.6	67	1.59	Cheese	1.50 (1.49, 1.51)	8,672
2.6	67	1.59	Canned seafood	1.55 (1.53, 1.57)	6,464
1.9	67	1.59	Sour cream dips	1.56 (1.53, 1.58)	6,201
9.3	67	1.57	Spices	1.43 (1.41, 1.44)	10,827
2.0	66	1.56	Oils/vinegar	1.52 (1.50, 1.54)	7,787
3.0	66	1.56	Frozen fish/seafood	1.51 (1.49, 1.53)	7,324
1-1 association rules for fruits					
2.8	70	1.67	Fresh herbs	1.62 (1.60, 1.64)	6,360
1.2	68	1.63	Frozen fruits	1.61 (1.58, 1.65)	3,959
4.2	67	1.61	Nut/seeds/dried fruits	1.54 (1.53, 1.56)	7,974
6.1	66	1.59	Packaged salads	1.49 (1.48, 1.51)	8,854
1.1	65	1.56	Yogurt	1.38 (1.37, 1.39)	10,608
5.5	65	1.54	Cereals	1.46 (1.44, 1.47)	8,747
2.5	64	1.54	Canned seafood	1.50 (1.48, 1.52)	6,173
1.3	64	1.53	Vegan/vegetarian foods	1.51 (1.49, 1.54)	3,555
2.4	63	1.52	Soy/rice/nut beverages	1.48 (1.46, 1.50)	4,598
5.3	63	1.51	Cheese	1.43 (1.42, 1.45)	8,455
3.0	62	1.49	Tortillas/flat breads	1.45 (1.43, 1.46)	6,802
3.8	62	1.49	Antipasto/dips	1.43 (1.42, 1.45)	7,729

IV. Discussion and conclusions

Interpretation of results

Using a rule-based machine learning method, associations present between different product categories were found. This thesis suggests that purchases of soft drinks, salty snacks, sweet snacks and candy among a large grocery chain in Montreal are often accompanied by other qualitatively unhealthy or RTE foods while purchases of fruits and vegetables are correlated with qualitatively healthier foods. These results are timely, given the growing interest in larger, more non-traditional sources of observational data that complement existing means of evaluating nutritional intake.[136,142] As researchers and public health practitioners continue to call for more holistic understandings of food consumption and its contribution to generating disease risk, the value of objective, longitudinal grocery sales data in assessing population-level dietary patterns is quite singular.[138] Indeed, in addition to Chapter 3, few studies to our knowledge have previously used sales data to such an end in the domain of public health. While traditional surveys which rely on self-report by customers and proprietary sales data from stores continue to show promise in their respective applications — as presented in Chapter 2 — their shortcomings demonstrate the need for novel approaches in evaluating population patterns of food intake.

The primary results of this analysis reveal that food purchases are clustered within baskets, a finding consistent with previous research.[146–149] In particular, soft drinks were found to be associated with salty snacks (support: 3.3%, lift: 1.9); water (support: 1.5%, lift: 1.8); frozen meals and sides (support: 1.2%, lift: 1.8); sausages and bacon (support: 1.4%, lift: 1.8); ice cream (support: 1.2%, lift: 1.7); and sweet snacks (support: 3.3%, lift: 1.6). All of the product categories in this list of rules are convenience foods, somewhat known to be typically high in sodium, carbohydrates, and trans and saturated fats. Even the category water contains many bottled products that are flavoured or have added sugar. Although the support of these rules is small, they must be understood in relation to the support of the rule's antecedent and consequent. In this case, soft drinks appear in just 9% of transactions (135,308 transactions). As previously mentioned, there are no 'best' thresholds when filtering association rules. Itemsets that are very prevalent will dominate if the minimum support is set too high, which risks lowering the association between products, and potentially not finding any rules at all. For example, as seen in Table 6, vegetables co-occur in purchases with soft drinks with a support of 4.3%, the highest of the displayed rules. However, the lift is 1.07, a very weak association. Similarly, if the support is too low, then the mined rules will show item combinations that are almost never purchased, but appear with an extremely high lift. The algorithm was therefore set with a minimum 1.0% support and a confidence equal to or greater than the support of the antecedent.

Sweet snacks were heavily associated with nuts, seeds, and dried fruits (support:

2.8%, lift: 2.1); cereals (support: 3.2%, lift: 1.9); canned soups (2.1%, lift: 1.8); salty snacks (support: 7.3%, lift: 1.7); and canned seafood (support: 1.5%, lift: 1.8); while salty snacks were associated with antipasto, dips, and pâtés (support: 2.4%, lift: 2.0); nuts, seeds, and dried fruits (support: 2.3%, lift: 1.9); canned soups (support: 1.9%, lift: 1.9); and canned seafood (support: 1.4%, lift: 1.9). These foods can be characterized as convenience foods and rather high in sodium or added sugar. In contrast, vegetables and fruits tended to appear in purchases with more minimally processed foods: vegetables were highly associated with fresh herbs (support: 3.4%, lift: 2.0); packaged salads (support: 6.7%, lift: 1.7); and vegan/vegetarian foods (support: 1.5%, lift: 1.7). Similarly, fruits demonstrated high association with fresh herbs (support: 2.8%, lift: 1.7) and frozen fruits (support: 1.2%, lift: 1.6). Interestingly, vegetables and fruits both showed a weak association with ready-to-eat meals and soft drinks. Fruit juices and drinks were associated with a much less consistent group of foods: convenience foods like canned soups, canned seafood, cereals, and canned vegetables all figured prominently (lift ≥ 1.9), as did yogurt (support: 5.9% lift: 1.8) and pasta, rice, and beans (support: 4.3%, lift: 1.7).

Contextual perspectives

Understanding population behaviours with respect to food is of utmost importance if we are to craft policies which lead to tangible shifts in diet. These findings, of course, must be contextualized within the broader portrait of food purchasing. Chapter 2 highlighted the loss of traditional diets and the historical and economic construction of the modern diet, one replete with nutrient-deficient UPFPs. This development has been shaped by unhealthy food advertising, whose scale and penetration is unprecedented in human history; the massive presence of UPFPs in supermarkets, who occupies a unique place as an intermediary between an individual and the production of their food; and the zealous efforts of large, agribusiness multinationals to undermine public health efforts which seek to address these issues.

Access to healthy foods in food environments has been extensively researched, but is primarily focused on productivist concerns as opposed to the socio-economic factors at play.[157] As a result, interventions predicated on nutritionism and other reductive assumptions on human behaviour and food risk being rather myopic in practice, focusing on increasing healthy eating in unhealthy food environments. Similarly, research and interventions which focus solely on the purchase of one product category or one nutrient may not be helpful if agribusiness actors and supermarkets are not simultaneously held to a certain critical standard for their role in promoting unhealthy food purchases (efforts in the latter may be just as important: one study has shown that the reformulation of products sold in retail spaces led to a change in the nutritional quality of food purchases, especially among potato chips and breakfast cereals).[158]

The individual shopper is hardly culpable nor are they in any real capacity to change these circumstances (at least individually). Yet research and interventions have been largely structured around assumptions of individual behaviour, thereby propagating a pervasive "ethic of responsibility" that moves the discussion away from systemic issues of food.[159] Simultaneous research that does concern the structural factors shaping unhealthy behaviours have failed to show how they manifest into daily lives of different populations. Studies that marry individual preferences and class using concepts such as Bordieu's theory of habitus, for instance (where everyday behaviours are bounded by distinctions of taste according to social class), are needed.[160] That is not to say that there is no hope: grassroots organizing that seek to re-localize food production, promote organic foods and fair trade practices, build sustainable food systems, and advocate for food sovereignty have demonstrated considerable promise in reinventing our relationship with food.

Limitations and future avenues for research

There are a number of limitations to the study at hand. To begin, the data come from a single grocery chain in Montréal. Transactions from other retailers that households may shop at, including smaller, independent grocers, are not considered. Furthermore, purchase data from restaurants and from convenience stores, or from outlets other than supermarkets and grocers (but still selling food products) are not included. This sales data is therefore not indicative of the entirety of a household's food expenditure. This may pose problems for the discovered rules, as individuals may be more likely to purchase certain products, such as convenience foods, outside of supermarkets. Indeed, as a large portion of the sampled members shopped with very low frequency (see A.3.3), there is no guarantee that the purchases they made are actually denotative of their general diet. Differences in stock may also influence this bias. Indeed, one study found that the volume of nutrient-poor, packaged food products from convenience stores and club stores increased from 2000 to 2012, and were higher compared to grocery stores.[161] Other studies have confirmed that the nutritional profile of packaged food products from retail food chains (such as the one in this study) have improved over time.[162]

Product information in the data is limited. UPCs did not always identify items clearly. A number of UPCs were unsuccessfully matched, and even among those that were matched, the possibility for misclassification remained. It is difficult to quantify this misclassification without some gold standard of UPC-category mapping. Moreover, random-weight data (certain bulk goods, deli products, produce, or bakery goods) may be systematically missing for specific stores. Loyalty card transactions from a store without a fresh bakery, for example, will systematically decrease the support with which those product categories appear in the data. Any rules with deli products will therefore be weighed less favourably than more general products that appear in many stores. However, as the food categories of this study are quite universal (F&V, soft drinks, juice, snacks), it is unlikely that the prevalence of these products varies greatly from store to store. Given that large retail grocery chains tend to carry a large assortment of similar brands, it is unlikely to have great variability between stores of the same retail name. However, nothing conclusive can be stated without a full inventory of products for each store, something not accessible during the course of this study. Futhermore, it is important to recognize the *proportionality* of the products for which these rules are observed: the most commonly purchased items in any baskets include fruits and vegetables, while soft drinks are only found in 9% of purchases.

Retail-provided sales data are not a panacea to other commercial sales data. As observational data that is collected primarily for marketing purposes and not public health research, there arise several issues that limit the type of questions we can ask. For example, all data regarding loyalty cards is anonymized and does not contain personal demographic characteristics. Although postal codes may be used to link to area-level census statistics, this is quite distinct from individual measures of age, gender, ethnicity, and other predictors that may be important, such as physical activity and sedentary behaviour. Researchers are therefore limited in the range of the inferences they are able to make accurately. Like sales data, loyalty card transactions only contain purchases, and the purchase of food does not always imply the consumption of food, since waste is not recorded: individual dietary habits are therefore outside of the scope of this data, and such data has limited explanatory power with respect to health outcomes.

Moreover, the data concerns only those customers that are enrolled in a loyalty card program of one retail store. This could limit the study in terms of sample bias, as enrollees may systematically differ from other segments of the urban population of study. Although area-level variables did not vary too much from averages for the island of Montréal, the magnitude of these differences is difficult to accurately evaluate without individual characteristics. With consent, time, and effort, more granular demographic could theoretically be collected from the loyalty card population. This is itself is difficult to conceptualize, for it is conceivable that a card may be used by many people, both within and outside of a specific household. As previously mentioned, this thesis only considers the purchasing patterns of a "household," insofar as one loyalty card is assumed to be used among the members of that household. Discovered rules should therefore not be interpreted as applicable to one individual, but rather, to a population of households that shops at a particular retail store.

An alternative analysis would focus on restricting stores in Montréal, irrespective of card holder location, and compare the discovered rules to those generated from a similar sample of transactions that don't have loyalty cards attached to them. The difference, however, is that inference would be directed towards store sales history, rather than shopper history, as only partial loyalty card member histories would be captured. Another option would be to instead segment shoppers based on RFM criteria or likewise (See A.3.3). RFM segmentation typically permits stronger association rules, especially when segmented customers share demographic attributes.[163] The rules generated would illuminate differences in the purchasing patterns of frequent shoppers who spend a lot per trip from casual shoppers who spend very little per trip. The longitudinal histories of these shoppers should also be leveraged to their full potential through sequential pattern mining, which yields association rules derived from transactions occurring in a particular sequence (this may be difficult given the size of the data).

All of these limitations should be acknowledged and contextualized within the question the study is seeking to answer. As mentioned, this thesis does not capture causal relationships, as the lack of directionality within association rules occludes causal inference. Individual dietary behaviours, though important to understand, are not considered. Motivations for such behaviours are similarly outside of the scope of this work. Thus, it is critical that future studies recognize this gap when pursuing population patterns of dietary intake, as relying solely on sales data ignores the conceptual relationships with food that influence these behaviours in the first place. In general, there is a need for more qualitative studies that look "under the hood" of food purchasing behaviour. Health consciousness, deeply ingrained in our social fabric, is one such example that should be understood in relation to household food purchases.[164] Other different cultural and social valuations of food should also be included.

Moreover, although there are a great deal of nutrient-deficient, processed and packaged foods which permeate society and may be said to be "unhealthy," no food is inherently good or bad. Indeed, the demarcation between processed and unprocessed foods is not the sole predictor of whether a food is considered healthy or unhealthy: there is a wide variability in the amount of fat, sodium, or sugar in processed products, and even minimally processed foods may contribute to high levels of sugar and sodium intake in one's diet.[165] However, highly processed, packaged foods are distinguished from minimally processed foods and are still likely to be important determinants of nutrient intake.[144]

Conclusions

This thesis demonstrates the utility of grocery sales data as a means for analysing population patterns of food purchasing. We processed and analysed a large collection of transactions obtained from a sample of loyalty card members of a large retail grocery chain in Montréal and discovered complementary products purchased alongside soft drinks, sweet and salty snacks, juices, and F&V using a rule-based machine learning method. The interpretation of these rules was discussed in the context of the food environments in which they occur. To the best of this author's knowledge, very few studies have done such an analysis on such a scale using retail-provided sales data. The strengths of this data — its volume, its granularity — and its capacity to be used in generating transaction-level associations between product categories speak to the importance of improved measures of purchasing that capture basket composition among households. This thesis therefore joins a chorus of other studies using traditional recall surveys and commercial retail data sets in calling for the adoption of more novel data sources for evaluating population-level dietary patterns and responses to food interventions.

V. Bibliography

1. Winson A. Bringing political economy into the debate on the obesity epidemic. *Agric Hum Values*. 2004;21(4):299-312. doi:10.1007/s10460-003-1206-6
2. Campbell NRC, Raine KD, McLaren L. "Junk Foods," "Treats," or "Pathogenic Foods"? A Call for Changing Nomenclature to Fit the Risk of Today's Diets. *Can J Cardiol*. 2012;28(4):403-404. doi:10.1016/j.cjca.2011.11.019
3. Acton RB, Vanderlee L, Hobin EP, Hammond D. Added sugar in the packaged foods and beverages available at a major Canadian retailer in 2015: a descriptive analysis. *CMAJ Open*. 2017;5(1):E1-E6. doi:10.9778/cmajo.20160076
4. Brambila-Macias J, Shankar B, Capacci S, et al. Policy Interventions to Promote Healthy Eating: A Review of What Works, What Does Not, and What is Promising. *Food Nutr Bull*. 2011;32(4):365-375. doi:10.1177/156482651103200408
5. Afshin A, Penalvo J, Del Gobbo L, et al. CVD Prevention Through Policy: a Review of Mass Media, Food/Menu Labeling, Taxation/Subsidies, Built Environment, School Procurement, Worksite Wellness, and Marketing Standards to Improve Diet. *Curr Cardiol Rep*. 2015;17(11):98. doi:10.1007/s11886-015-0658-9
6. Hyseni L, Atkinson M, Bromley H, et al. The effects of policy actions to improve population dietary patterns and prevent diet-related non-communicable diseases: scoping review. *Eur J Clin Nutr*. 2017;71(6):694-711. doi:10.1038/ejcn.2016.234
7. Fletcher J, Frisvold D, Tefft N. Substitution Patterns Can Limit the Effects of Sugar-Sweetened Beverage Taxes on Obesity. *Prev Chronic Dis*. 2013;10. doi:10.5888/pcd10.120195
8. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392(10159):1736-1788. doi:10.1016/S0140-6736(18)32203-7
9. Ezzati M, Vander Hoorn S, Lawes CMM, et al. Rethinking the "Diseases of Affluence" Paradigm: Global Patterns of Nutritional Risks in Relation to Economic Development. *PLoS Med*. 2005;2(5). doi:10.1371/journal.pmed.0020133
10. Terzic A, Waldman S. Chronic Diseases: The Emerging Pandemic. *Clin Transl Sci*. 2011;4(3):225-226. doi:10.1111/j.1752-8062.2011.00295.x
11. Bennett JE, Stevens GA, Mathers CD, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *The Lancet*. 2018;392(10152):1072-1088. doi:10.1016/S0140-6736(18)31992-5
12. Lang JJ, Alam S, Cahill LE, et al. Global Burden of Disease Study trends for Canada from 1990 to 2016. *CMAJ*. 2018;190(44):E1296-E1304. doi:10.1503/cmaj.180698
13. Willett WC, Stampfer MJ. Current evidence on healthy eating. *Annu Rev Public Health*. 2013;34:77-95. doi:10.1146/annurev-publhealth-031811-124646
14. Micha R, Shulkin ML, Peñalvo JL, et al. Etiologic effects and optimal intakes of foods and nutrients for risk of cardiovascular diseases and diabetes: Systematic reviews and meta-analyses from the Nutrition and Chronic Diseases Expert Group (NutriCoDE). *PloS One*. 2017;12(4):e0175149. doi:10.1371/journal.pone.0175149
15. Afshin A, Sur PJ, Fay KA, et al. Health effects of dietary risks in 195 countries, 1990–

2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2019;0(0). doi:10.1016/S0140-6736(19)30041-8

16. Aune D, Giovannucci E, Boffetta P, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality—a systematic review and dose-response meta-analysis of prospective studies. *Int J Epidemiol*. 2017;46(3):1029-1056. doi:10.1093/ije/dyw319
17. Dauchet L, Amouyel P, Hercberg S, Dallongeville J. Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies. *J Nutr*. 2006;136(10):2588-2593. doi:10.1093/jn/136.10.2588
18. He FJ, Nowson CA, MacGregor GA. Fruit and vegetable consumption and stroke: meta-analysis of cohort studies. *Lancet Lond Engl*. 2006;367(9507):320-326. doi:10.1016/S0140-6736(06)68069-0
19. Anderson JW, Baird P, Davis RH, et al. Health benefits of dietary fiber. *Nutr Rev*. 2009;67(4):188-205. doi:10.1111/j.1753-4887.2009.00189.x
20. Threapleton DE, Greenwood DC, Evans CEL, et al. Dietary fibre intake and risk of cardiovascular disease: systematic review and meta-analysis. *BMJ*. 2013;347:f6879. doi:10.1136/bmj.f6879
21. Mellen PB, Walsh TF, Herrington DM. Whole grain intake and cardiovascular disease: a meta-analysis. *Nutr Metab Cardiovasc Dis NMCD*. 2008;18(4):283-290. doi:10.1016/j.numecd.2006.12.008
22. Aune D, Norat T, Romundstad P, Vatten LJ. Whole grain and refined grain consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of cohort studies. *Eur J Epidemiol*. 2013;28(11):845-858. doi:10.1007/s10654-013-9852-5
23. Micha R, Wallace SK, Mozaffarian D. Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. *Circulation*. 2010;121(21):2271-2283. doi:10.1161/CIRCULATIONAHA.109.924977
24. Cui K, Liu Y, Zhu L, Mei X, Jin P, Luo Y. Association between intake of red and processed meat and the risk of heart failure: a meta-analysis. *BMC Public Health*. 2019;19(1):354. doi:10.1186/s12889-019-6653-0
25. Chan DSM, Lau R, Aune D, et al. Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PloS One*. 2011;6(6):e20456. doi:10.1371/journal.pone.0020456
26. Zheng J, Huang T, Yu Y, Hu X, Yang B, Li D. Fish consumption and CHD mortality: an updated meta-analysis of seventeen cohort studies. *Public Health Nutr*. 2012;15(4):725-737. doi:10.1017/S1368980011002254
27. Afshin A, Micha R, Khatibzadeh S, Mozaffarian D. Consumption of nuts and legumes and risk of incident ischemic heart disease, stroke, and diabetes: a systematic review and meta-analysis. *Am J Clin Nutr*. 2014;100(1):278-288. doi:10.3945/ajcn.113.076901
28. Chen M, Sun Q, Giovannucci E, et al. Dairy consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *BMC Med*. 2014;12:215. doi:10.1186/s12916-014-0215-1
29. Drouin-Chartier J-P, Brassard D, Tessier-Grenier M, et al. Systematic Review of the Association between Dairy Product Consumption and Risk of Cardiovascular-Related

Clinical Outcomes. *Adv Nutr Bethesda Md*. 2016;7(6):1026-1040. doi:10.3945/an.115.011403
30. Soedamah-Muthu SS, Verberne LDM, Ding EL, Engberink MF, Geleijnse JM. Dairy consumption and incidence of hypertension: a dose-response meta-analysis of prospective cohort studies. *Hypertens Dallas Tex 1979*. 2012;60(5):1131-1137. doi:10.1161/HYPERTENSIONAHA.112.195206
31. Astrup A. Yogurt and dairy product consumption to prevent cardiometabolic diseases: epidemiologic and experimental studies. *Am J Clin Nutr*. 2014;99(5):1235S-1242S. doi:10.3945/ajcn.113.073015
32. Coutinho M, Gerstein HC, Wang Y, Yusuf S. The relationship between glucose and incident cardiovascular events. A metaregression analysis of published data from 20 studies of 95,783 individuals followed for 12.4 years. *Diabetes Care*. 1999;22(2):233-240. doi:10.2337/diacare.22.2.233
33. Johnson Rachel K., Appel Lawrence J., Brands Michael, et al. Dietary Sugars Intake and Cardiovascular Health. *Circulation*. 2009;120(11):1011-1020. doi:10.1161/CIRCULATIONAHA.109.192627
34. Bray GA, Nielsen SJ, Popkin BM. Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. *Am J Clin Nutr*. 2004;79(4):537-543. doi:10.1093/ajcn/79.4.537
35. Hu FB, Malik VS. Sugar-sweetened beverages and risk of obesity and type 2 diabetes: Epidemiologic evidence. *Physiol Behav*. 2010;100(1):47-54. doi:10.1016/j.physbeh.2010.01.036
36. Malik VS, Hu FB. Fructose and Cardiometabolic Health: What the Evidence From Sugar-Sweetened Beverages Tells Us. *J Am Coll Cardiol*. 2015;66(14):1615-1624. doi:10.1016/j.jacc.2015.08.025
37. Woodward-Lopez G, Kao J, Ritchie L. To what extent have sweetened beverages contributed to the obesity epidemic? *Public Health Nutr*. 2011;14(3):499-509. doi:10.1017/S1368980010002375
38. Tappy L, Lê K-A. Metabolic Effects of Fructose and the Worldwide Increase in Obesity. *Physiol Rev*. 2010;90(1):23-46. doi:10.1152/physrev.00019.2009
39. Poti JM, Dunford EK, Popkin BM. Sodium Reduction in US Households' Packaged Food and Beverage Purchases, 2000 to 2014. *JAMA Intern Med*. 2017;177(7):986-994. doi:10.1001/jamainternmed.2017.1407
40. Caillavet F, Darmon N, Létoile F, Nichèle V. Is nutritional quality of food-at-home purchases improving? 1969–2010: 40 years of household consumption surveys in France. *Eur J Clin Nutr*. 2018;72(2):220. doi:10.1038/s41430-017-0041-6
41. Singh GM, Micha R, Khatibzadeh S, et al. Global, Regional, and National Consumption of Sugar-Sweetened Beverages, Fruit Juices, and Milk: A Systematic Assessment of Beverage Intake in 187 Countries. *PLOS ONE*. 2015;10(8):e0124845. doi:10.1371/journal.pone.0124845
42. Langlois K, Garriguet D. Sugar consumption among Canadians of all ages. *Health Rep*. 2011;22(3):23-27.
43. Moubarac J-C, Batal M, Louzada ML, Martinez Steele E, Monteiro CA. Consumption of ultra-processed foods predicts diet quality in Canada. *Appetite*. 2017;108:512-520. doi:10.1016/j.appet.2016.11.006
44. Government of Canada SC. Trends and correlates of frequency of fruit and vegetable

consumption, 2007 to 2014. https://www.statcan.gc.ca/pub/82-003-x/2018001/article/54901-eng.htm. January 17, 2018. Accessed May 16, 2018.
45. Azagba S, Sharaf MF. Disparities in the frequency of fruit and vegetable consumption by socio-demographic and lifestyle characteristics in Canada. *Nutr J*. 2011;10:118. doi:10.1186/1475-2891-10-118
46. Kristin B, Kelly S, Travis H, Joseph L, Lori C. Retail food environments, shopping experiences, First Nations and the provincial Norths. *Health Promot Chronic Dis Prev Can Res Policy Pract*. 2017;37(10):333-341.
47. Jones AC, Veerman JL, Hammond D. THE HEALTH AND ECONOMIC IMPACT OF A TAX ON SUGARY DRINKS IN CANADA. March 2017:77.
48. Vercammen KA, Koma JW, Bleich SN. Trends in Energy Drink Consumption Among U.S. Adolescents and Adults, 2003–2016. *Am J Prev Med*. April 2019. doi:10.1016/j.amepre.2018.12.007
49. Malik VS, Popkin BM, Bray GA, Després J-P, Willett WC, Hu FB. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. *Diabetes Care*. 2010;33(11):2477-2483. doi:10.2337/dc10-1079
50. Malik Vasanti S., Popkin Barry M., Bray George A., Després Jean-Pierre, Hu Frank B. Sugar-Sweetened Beverages, Obesity, Type 2 Diabetes Mellitus, and Cardiovascular Disease Risk. *Circulation*. 2010;121(11):1356-1364. doi:10.1161/CIRCULATIONAHA.109.876185
51. Malik VS, Schulze MB, Hu FB. Intake of sugar-sweetened beverages and weight gain: a systematic review. *Am J Clin Nutr*. 2006;84(2):274-288. doi:10.1093/ajcn/84.1.274
52. Ludwig DS, Peterson KE, Gortmaker SL. Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis. *Lancet Lond Engl*. 2001;357(9255):505-508. doi:10.1016/S0140-6736(00)04041-1
53. Nestle M. *Food Politics: How the Food Industry Influences Nutrition and Health*. Berkeley. University of California Press.; 2013. https://www.ucpress.edu/book/9780520275966/food-politics. Accessed April 16, 2019.
54. Popkin BM. Global nutrition dynamics: the world is shifting rapidly toward a diet linked with noncommunicable diseases. *Am J Clin Nutr*. 2006;84(2):289-298. doi:10.1093/ajcn/84.2.289
55. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev*. 2012;70(1):3-21. doi:10.1111/j.1753-4887.2011.00456.x
56. Baker P, Friel S. Food systems transformations, ultra-processed food markets and the nutrition transition in Asia. *Glob Health*. 2016;12(1):80. doi:10.1186/s12992-016-0223-3
57. Welch RW, Mitchell PCS. Food processing: a century of change. *Br Med Bull*. 2000;56(1):1-17. doi:10.1258/0007142001902923
58. Ng SW, Slining MM, Popkin BM. Use of caloric and noncaloric sweeteners in US consumer packaged foods, 2005-2009. *J Acad Nutr Diet*. 2012;112(11):1828-1834.e1-6. doi:10.1016/j.jand.2012.07.009
59. Lennerz BS, Alsop DC, Holsen LM, et al. Effects of dietary glycemic index on brain regions related to reward and craving in men1234. *Am J Clin Nutr*. 2013;98(3):641-647. doi:10.3945/ajcn.113.064113

60. Gorski MT, Roberto CA. Public health policies to encourage healthy eating habits: recent perspectives. *J Healthc Leadersh*. 2015;7:81-90. doi:10.2147/JHL.S69188
61. Nestle M, Jacobson MF. Halting the obesity epidemic: a public health policy approach. *Public Health Rep*. 2000;115(1):12-24.
62. Ebbeling CB, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. *The Lancet*. 2002;360(9331):473-482. doi:10.1016/S0140-6736(02)09678-2
63. Story M, French S. Food Advertising and Marketing Directed at Children and Adolescents in the US. *Int J Behav Nutr Phys Act*. 2004;1(1):3. doi:10.1186/1479-5868-1-3
64. Fried EJ, Nestle M. The Growing Political Movement Against Soft Drinks in Schools. *JAMA*. 2002;288(17):2181-2181. doi:10.1001/jama.288.17.2181-JMS1106-7-1
65. Coon KA, Tucker KL. Television and children's consumption patterns. *Minerva Pediatr*. 2001;53:15.
66. Dixon HG, Scully ML, Wakefield MA, White VM, Crawford DA. The effects of television advertisements for junk food versus nutritious food on children's food attitudes and preferences. *Soc Sci Med*. 2007;65(7):1311-1323. doi:10.1016/j.socscimed.2007.05.011
67. Zimmerman FJ, Shimoga SV. The effects of food advertising and cognitive load on food choices. *BMC Public Health*. 2014;14:342. doi:10.1186/1471-2458-14-342
68. Zheng Y, Kaiser HM. Advertising and U. S. Nonalcoholic Beverage Demand. *Agric Resour Econ Rev*. 2008;37(2):147-159. doi:10.1017/S1068280500002963
69. Vartanian LR, Schwartz MB, Brownell KD. Effects of Soft Drink Consumption on Nutrition and Health: A Systematic Review and Meta-Analysis. *Am J Public Health*. 2007;97(4):667-675. doi:10.2105/AJPH.2005.083782
70. Mialon M, Mialon J, Andrade GC, Jean-Claude M. 'We must have a sufficient level of profitability': food industry submissions to the French parliamentary inquiry on industrial food. *Crit Public Health*. 2019;0(0):1-11. doi:10.1080/09581596.2019.1606418
71. Tillotson JE. The Mega-Brands That Rule Our Diet, Part 1. *Nutr Today*. 2005;40(6):257.
72. Alexander E, Yach D, Mensah GA. Major multinational food and beverage companies and informal sector contributions to global food consumption: implications for nutrition policy. *Glob Health*. 2011;7(1):26. doi:10.1186/1744-8603-7-26
73. Sacks G, Swinburn BA, Cameron AJ, Ruskin G. How food companies influence evidence and opinion – straight from the horse's mouth. *Crit Public Health*. 2018;28(2):253-256. doi:10.1080/09581596.2017.1371844
74. Clapp J, Scrinis G. Big Food, Nutritionism, and Corporate Power. *Globalizations*. 2017;14(4):578-595. doi:10.1080/14747731.2016.1239806
75. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy Nutrition Environments: Concepts and Measures. *Am J Health Promot*. 2005;19(5):330-333. doi:10.4278/0890-1171-19.5.330
76. Reardon T, Timmer CP, Barrett CB, Berdegué J. The Rise of Supermarkets in Africa, Asia, and Latin America. *Am J Agric Econ*. 2003;85(5):1140-1146. doi:10.1111/j.0092-5853.2003.00520.x
77. Reardon T, Timmer CP, Berdegue JA. The Rapid Rise of Supermarkets in Developing

Countries: Induced Organizational, Institutional, and Technological Change in Agrifood Systems. *EJADE Electron J Agric Dev Econ*. 2004;01(2):1-16.
78. Canadian Grocer. Chain food store sales in Canada from 2004 to 2019 (in billion Canadian dollars). Statista - The Statistics Portal. https://www.statista.com/statistics/480291/sales-of-chain-food-stores-canada/. Accessed April 18, 2019.
79. Story M, Kaphingst KM, Robinson-O'Brien R, Glanz K. Creating Healthy Food and Eating Environments: Policy and Environmental Approaches. *Annu Rev Public Health*. 2008;29(1):253-272. doi:10.1146/annurev.publhealth.29.020907.090926
80. Gould AC, Apparicio P, Cloutier M-S. Classifying Neighbourhoods by Level of Access to Stores Selling Fresh Fruit and Vegetables and Groceries: Identifying Problematic Areas in the City of Gatineau, Quebec. *Can J Public Health*. 2012;103(6):433-437. doi:10.17269/cjph.103.3228
81. Lamichhane AP, Warren J, Puett R, et al. Spatial patterning of supermarkets and fast food outlets with respect to neighborhood characteristics. *Health Place*. 2013;23. doi:10.1016/j.healthplace.2013.07.002
82. Black C, Moon G, Baird J. Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment? *Health Place*. 2014;27:229-242. doi:10.1016/j.healthplace.2013.09.015
83. Luan H, Minaker LM, Law J. Do marginalized neighbourhoods have less healthy retail food environments? An analysis using Bayesian spatial latent factor and hurdle models. *Int J Health Geogr*. 2016;15(1). doi:10.1186/s12942-016-0060-x
84. Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV. Obesity and Supermarket Access: Proximity or Price? *Am J Public Health*. 2012;102(8):e74-e80. doi:10.2105/AJPH.2012.300660
85. Minaker LM, Shuh A, Olstad DL, Engler-Stringer R, Black JL, Mah CL. Retail food environments research in Canada: A scoping review. *Can J Public Health*. 2016;107:4. doi:10.17269/cjph.107.5344
86. Morland K, Wing S, Roux AD. The Contextual Effect of the Local Food Environment on Residents' Diets: The Atherosclerosis Risk in Communities Study. *Am J Public Health*. 2002;92(11):1761-1768.
87. Bodor JN, Rice JC, Farley TA, Swalm CM, Rose D. The Association between Obesity and Urban Food Environments. *J Urban Health Bull N Y Acad Med*. 2010;87(5):771-781. doi:10.1007/s11524-010-9460-6
88. Morland K, Diez Roux AV, Wing S. Supermarkets, Other Food Stores, and Obesity: The Atherosclerosis Risk in Communities Study. *Am J Prev Med*. 2006;30(4):333-339. doi:10.1016/j.amepre.2005.11.003
89. Block JP, Subramanian SV. Moving Beyond "Food Deserts": Reorienting United States Policies to Reduce Disparities in Diet Quality. *PLOS Med*. 2015;12(12):e1001914. doi:10.1371/journal.pmed.1001914
90. Dubowitz T, Ghosh-Dastidar M, Cohen DA, et al. Diet And Perceptions Change With Supermarket Introduction In A Food Desert, But Not Because Of Supermarket Use. *Health Aff (Millwood)*. 2015;34(11):1858-1868. doi:10.1377/hlthaff.2015.0667
91. Ghosh-Dastidar M, Hunter G, Collins RL, et al. Does opening a supermarket in a food desert change the food environment? *Health Place*. 2017;46:249-256. doi:10.1016/j.healthplace.2017.06.002

92. Ross NA, Tremblay S, Graham K. Neighbourhood influences on health in Montréal, Canada. *Soc Sci Med*. 2004;59(7):1485-1494. doi:10.1016/j.socscimed.2004.01.016
93. Apparicio P, Cloutier M-S, Shearmur R. The case of Montréal's missing food deserts: Evaluation of accessibility to food supermarkets. *Int J Health Geogr*. 2007:13.
94. Kelly B, Flood VM, Yeatman H. Measuring local food environments: An overview of available methods and measures. *Health Place*. 2011;17(6):1284-1293. doi:10.1016/j.healthplace.2011.08.014
95. Ni Mhurchu C, Vandevijvere S, Waterlander W, et al. Monitoring the availability of healthy and unhealthy foods and non-alcoholic beverages in community and consumer retail food environments globally. *Obes Rev Off J Int Assoc Study Obes*. 2013;14 Suppl 1:108-119. doi:10.1111/obr.12080
96. Hawkes C. Dietary Implications of Supermarket Development: A Global Perspective. *Dev Policy Rev*. 2008;26(6):657-692. doi:10.1111/j.1467-7679.2008.00428.x
97. Black C, Ntani G, Inskip H, et al. Measuring the healthfulness of food retail stores: variations by store type and neighbourhood deprivation. *Int J Behav Nutr Phys Act*. 2014;11:69. doi:10.1186/1479-5868-11-69
98. Gustafson A, Christian JW, Lewis S, Moore K, Jilcott S. Food venue choice, consumer food environment, but not food venue availability within daily travel patterns are associated with dietary intake among adults, Lexington Kentucky 2011. *Nutr J*. 2013;12:17. doi:10.1186/1475-2891-12-17
99. Chaix B, Bean K, Daniel M, et al. Associations of Supermarket Characteristics with Weight Status and Body Fat: A Multilevel Analysis of Individuals within Supermarkets (RECORD Study). *PLOS ONE*. 2012;7(4):e32908. doi:10.1371/journal.pone.0032908
100. Giskes K, Lenthe F van, Avendano-Pabon M, Brug J. A systematic review of environmental factors and obesogenic dietary intakes among adults: are we getting closer to understanding obesogenic environments? *Obes Rev*. 2011;12(5):e95-e106. doi:10.1111/j.1467-789X.2010.00769.x
101. Jalbert-Arsenault É, Robitaille E, Paquette M-C. Development, reliability and use of a food environment assessment tool in supermarkets of four neighbourhoods in Montréal, Canada. *Health Promot Chronic Dis Prev Can*. 2017;37:293-302. doi:10.24095/hpcdp.37.9.04
102. Farley TA, Rice J, Bodor JN, Cohen DA, Bluthenthal RN, Rose D. Measuring the Food Environment: Shelf Space of Fruits, Vegetables, and Snack Foods in Stores. *J Urban Health Bull N Y Acad Med*. 2009;86(5):672-682. doi:10.1007/s11524-009-9390-3
103. Rose D, Hutchinson PL, Bodor JN, et al. Neighborhood Food Environments and Body Mass Index The Importance of In-Store Contents. *Am J Prev Med*. 2009;37(3):214-219. doi:10.1016/j.amepre.2009.04.024
104. Thornton LE, Cameron AJ, McNaughton SA, Worsley A, Crawford DA. The availability of snack food displays that may trigger impulse purchases in Melbourne supermarkets. *BMC Public Health*. 2012;12(1):194. doi:10.1186/1471-2458-12-194
105. Thornton LE, Cameron AJ, McNaughton SA, et al. Does the availability of snack foods in supermarkets vary internationally? *Int J Behav Nutr Phys Act*. 2013;10(1):56. doi:10.1186/1479-5868-10-56
106. Nakamura R, Pechey R, Suhrcke M, Jebb SA, Marteau TM. Sales impact of displaying alcoholic and non-alcoholic beverages in end-of-aisle locations: An observational study. *Soc Sci Med*. 2014;108:68-73. doi:10.1016/j.socscimed.2014.02.032

107. Charlton EL, Kähkönen LA, Sacks G, Cameron AJ. Supermarkets and unhealthy food marketing: An international comparison of the content of supermarket catalogues/circulars. *Prev Med*. 2015:6. doi:10.1016/j.ypmed.2015.08.023
108. Abratt R, Goodey SD. Unplanned buying and in-store stimuli in supermarkets. *Manag Decis Econ*. 1990;11(2):111-121. doi:10.1002/mde.4090110204
109. Miranda MJ. Determinants of shoppers' checkout behaviour at supermarkets. *J Target Meas Anal Mark*. 2008;16(4):312-321. doi:10.1057/jt.2008.23
110. Kim B-D, Park K. Studying patterns of consumer's grocery shopping trip. *J Retail*. 1997;73(4):501-517. doi:10.1016/S0022-4359(97)90032-4
111. Rhee H, Bell DR. The inter-store mobility of supermarket shoppers. *J Retail*. 2002;78(4):225-237. doi:10.1016/S0022-4359(02)00099-4
112. Caspi CE, Pelletier JE, Harnack LJ, Erickson DJ, Lenk K, Laska MN. Pricing of Staple Foods at Supermarkets versus Small Food Stores. *Int J Environ Res Public Health*. 2017;14(8). doi:10.3390/ijerph14080915
113. Cannuscio CC, Hillier A, Karpyn A, Glanz K. The social dynamics of healthy food shopping and store choice in an urban environment. *Soc Sci Med*. 2014;122:13-20. doi:10.1016/j.socscimed.2014.10.005
114. Hollywood LE. Healthful grocery shopping. Perceptions and barriers. 2013:8.
115. Andreyeva T, Chaloupka FJ, Brownell KD. Estimating the potential of taxes on sugar-sweetened beverages to reduce consumption and generate revenue. *Prev Med*. 2011;52(6):413-416. doi:10.1016/j.ypmed.2011.03.013
116. Powell LM, Chriqui JF, Khan T, Wada R, Chaloupka FJ. Assessing the Potential Effectiveness of Food and Beverage Taxes and Subsidies for Improving Public Health: A Systematic Review of Prices, Demand and Body Weight Outcomes. *Obes Rev Off J Int Assoc Study Obes*. 2013;14(2):110-128. doi:10.1111/obr.12002
117. Niebylski ML, Redburn KA, Duhaney T, Campbell NR. Healthy food subsidies and unhealthy food taxation: A systematic review of the evidence. *Nutrition*. 2015;31(6):787-795. doi:10.1016/j.nut.2014.12.010
118. Roberts KE, Ells LJ, McGowan VJ, et al. A rapid review examining purchasing changes resulting from fiscal measures targeted at high sugar foods and sugar-sweetened drinks. *Nutr Diabetes*. 2017;7(12). doi:10.1038/s41387-017-0001-1
119. Afshin A, Peñalvo JL, Del Gobbo L, et al. The prospective impact of food pricing on improving dietary consumption: A systematic review and meta-analysis. *PLoS ONE*. 2017;12(3). doi:10.1371/journal.pone.0172277
120. Gittelsohn J, Trude ACB, Kim H. Pricing Strategies to Encourage Availability, Purchase, and Consumption of Healthy Foods and Beverages: A Systematic Review. *Prev Chronic Dis*. 2017;14. doi:10.5888/pcd14.170213
121. Thow AM, Downs SM, Mayes C, Trevena H, Waqanivalu T, Cawley J. Fiscal policy to improve diets and prevent noncommunicable diseases: from recommendations to action. *Bull World Health Organ*. 2018;96(3):201-210. doi:10.2471/BLT.17.195982
122. Hartmann-Boyce J, Bianchi F, Piernas C, et al. Grocery store interventions to change food purchasing behaviors: a systematic review of randomized controlled trials. *Am J Clin Nutr*. 2018;107(6):1004-1016. doi:10.1093/ajcn/nqy045
123. Edwards RD. Commentary: Soda taxes, obesity, and the shifty behavior of consumers. *Prev Med*. 2011;52(6):417-418. doi:10.1016/j.ypmed.2011.04.011
124. Glanz K, Bader MDM, Iyer S. Retail Grocery Store Marketing Strategies and Obesity.

Am J Prev Med. 2012:10. doi:10.1016/j.amepre.2012.01.013
125. Goryakin Y, Monsivais P, Suhrcke M. Soft drink prices, sales, body mass index and diabetes: Evidence from a panel of low-, middle- and high-income countries. *Food Policy*. 2017;73:88-94. doi:10.1016/j.foodpol.2017.09.002
126. Thow AM, Downs S, Jan S. A systematic review of the effectiveness of food taxes and subsidies to improve diets: Understanding the recent evidence. *Nutr Rev*. 2014;72(9):551-565. doi:10.1111/nure.12123
127. Pomeranz JL. Advanced policy options to regulate sugar-sweetened beverages to support public health. *J Public Health Policy*. 2012;33(1):75-88. doi:10.1057/jphp.2011.46
128. Cummins S, Petticrew M, Sparks L, Findlay A. Large scale food retail interventions and diet. *BMJ*. 2005;330(7493):683-684. doi:10.1136/bmj.330.7493.683
129. Mah CL, Cook B, Rideout K, Minaker LM. Policy options for healthier retail food environments in city-regions. *Can J Public Health*. 2016;107(1):eS64-eS67. doi:10.17269/CJPH.107.5343
130. Sisnowski J, Street JM, Merlin T. Improving food environments and tackling obesity: A realist systematic review of the policy success of regulatory interventions targeting population nutrition. *PLoS ONE*. 2017;12(8). doi:10.1371/journal.pone.0182581
131. Colchero MA, Popkin BM, Rivera JA, Ng SW. Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. *BMJ*. 2016;352:h6704. doi:10.1136/bmj.h6704
132. Agrawal R, Imieliński T, Swami A. Mining Association Rules Between Sets of Items in Large Databases. In: *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*. SIGMOD '93. New York, NY, USA: ACM; 1993:207–216. doi:10.1145/170035.170072
133. Brin S, Motwani R, Silverstein C. Beyond Market Baskets: Generalizing Association Rules to Correlations. In: *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*. SIGMOD '97. New York, NY, USA: ACM; 1997:265–276. doi:10.1145/253260.253327
134. Montgomery AL. Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data. *Mark Sci*. 1997;16(4):315-337. doi:10.1287/mksc.16.4.315
135. Mladenic D, Eddy WF, Ziolko S. Exploratory Analysis of Retail Sales of Billions of Items. In: ; 2002.
136. Bandy L, Adhikari V, Jebb S, Rayner M. The use of commercial food purchase data for public health nutrition research: A systematic review. *PLOS ONE*. 2019;14(1):e0210192. doi:10.1371/journal.pone.0210192
137. Council NR, Education D of B and SS and, Statistics C on N, Making P on E the DI in S of F and NP Research, and Decision. *Improving Data to Analyze Food and Nutrition Policies*. National Academies Press; 2005.
138. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13(1):3-9.
139. Prasad A, Strijnev A, Zhang Q. What can grocery basket data tell us about health consciousness? *Int J Res Mark*. 2008;25(4):301-309. doi:10.1016/j.ijresmar.2008.05.001
140. French SA, Shimotsu ST, Wall M, Gerlach AF. Capturing the Spectrum of Household Food and Beverage Purchasing Behavior: A Review. *J Am Diet Assoc*.

2008;108(12):2051-2058. doi:10.1016/j.jada.2008.09.001
141. Appelhans BM, French SA, Tangney CC, Powell LM, Wang Y. To what extent do food purchases reflect shoppers' diet quality and nutrient intake? *Int J Behav Nutr Phys Act*. 2017;14(1):46. doi:10.1186/s12966-017-0502-2
142. Timmins KA, Green MA, Radley D, Morris MA, Pearce J. How has big data contributed to obesity research? A review of the literature. *Int J Obes*. July 2018. doi:10.1038/s41366-018-0153-7
143. Brinkerhoff KM, Brewster PJ, Clark EB, Jordan KC, Cummins MR, Hurdle JF. Linking Supermarket Sales Data To Nutritional Information: An Informatics Feasibility Study. *AMIA Annu Symp Proc*. 2011;2011:598-606.
144. Poti JM, Mendez MA, Ng SW, Popkin BM. Is the degree of food processing and convenience linked with the nutritional quality of foods purchased by US households? 1234. *Am J Clin Nutr*. 2015;101(6):1251-1262. doi:10.3945/ajcn.114.100925
145. Smith RD, Cornelsen L, Quirmbach D, Jebb SA, Marteau TM. Are sweet snacks more sensitive to price increases than sugar-sweetened beverages: analysis of British food purchase data. *BMJ Open*. 2018;8(4):e019788. doi:10.1136/bmjopen-2017-019788
146. Piernas C, Ng SW, Mendez MA, Gordon-Larsen P, Popkin BM. A Dynamic Panel Model of the Associations of Sweetened Beverage Purchases With Dietary Quality and Food-Purchasing Patterns. *Am J Epidemiol*. 2015;181(9):661-671. doi:10.1093/aje/kwu317
147. Whybrow S, Craig L, Macdiarmid J. Dietary patterns of households in Scotland: Differences by level of deprivation and associations with dietary goals. *Nutr Health*. 2018;24(1):29-35. doi:10.1177/0260106017745389
148. Binkley J, Golub A. Comparison of grocery purchase patterns of diet soda buyers to those of regular soda buyers. *Appetite*. 2007;49(3):561-571. doi:10.1016/j.appet.2007.03.225
149. Pechey R, Monsivais P. Supermarket Choice, Shopping Behavior, Socioeconomic Status, and Food Purchases. *Am J Prev Med*. 2015;49(6):868. doi:10.1016/j.amepre.2015.04.020
150. Stern D, Poti JM, Ng SW, Robinson WR, Gordon-Larsen P, Popkin BM. Where people shop is not associated with the nutrient quality of packaged foods for any racial-ethnic group in the United States. *Am J Clin Nutr*. 2016;103(4):1125-1134. doi:10.3945/ajcn.115.121806
151. Griffith R, O'Connell M, Smith K. Relative prices, consumer preferences, and the demand for food. *Oxf Rev Econ Policy*. 2015;31(1):116-130. doi:10.1093/oxrep/grv004
152. Smed S, Tetens I, Lund TB, Holm L, Nielsen AL. The consequences of unemployment on diet composition and purchase behaviour: a longitudinal study from Denmark. *Public Health Nutr*. 2018;21(3):580-592. doi:10.1017/S136898001700266X
153. Grummon AH, Taillie LS. Nutritional profile of Supplemental Nutrition Assistance Program household food and beverage purchases12. *Am J Clin Nutr*. 2017;105(6):1433-1442. doi:10.3945/ajcn.116.147173
154. Aiello LM, Schifanella R, Quercia D, Prete LD. Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci*. 2019;8(1):14. doi:10.1140/epjds/s13688-019-0191-y
155. Einav L, Leibtag E, Nevo A. Recording discrepancies in Nielsen Homescan data: Are

they present and do they matter? *QME*. 2010;8(2):207-239. doi:10.1007/s11129-009-9073-0
156. Mamiya H, Moodie EEM, Buckeridge DL. A novel application of point-of-sales grocery transaction data to enhance community nutrition monitoring. *AMIA Annu Symp Proc*. 2018;2017:1253-1261.
157. Nisbett N. Understanding the nourishment of bodies at the centre of food and health systems – systemic, bodily and new materialist perspectives on nutritional inequity. *Soc Sci Med*. March 2019. doi:10.1016/j.socscimed.2019.02.041
158. Spiteri M, Soler L-G. Food reformulation and nutritional quality of food consumption: an analysis based on households panel data in France. *Eur J Clin Nutr*. 2018;72(2):228-235. doi:10.1038/s41430-017-0044-3
159. Blue S, Shove E, Carmona C, Kelly MP. Theories of practice and public health: understanding (un)healthy practices. *Crit Public Health*. 2016;26(1):36-50. doi:10.1080/09581596.2014.980396
160. Wills W, Backett-Milburn K, Roberts M-L, Lawton J. The framing of social class distinctions through family food and eating practices. *Sociol Rev*. 2011; 59(4). doi:10.1111/j.1467-954X.2011.02035.x
161. Stern D, Ng SW, Popkin BM. The Nutrient Content of U.S. Household Food Purchases by Store Type. *Am J Prev Med*. 2016;50(2):180-190. doi:10.1016/j.amepre.2015.07.025
162. Taillie LS, Ng SW, Popkin BM. Walmart and Other Food Retail Chains: Trends and Disparities in the Nutritional Profile of Packaged Food Purchases. *Am J Prev Med*. 2016;50(2):171-179. doi:10.1016/j.amepre.2015.07.015
163. Takci H, Sarvari PA, Ustundag A. Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes*. 2016;45(7):1129-1157. doi:10.1108/K-07-2015-0180
164. Ayo N. Understanding health promotion in a neoliberal climate and the making of health conscious citizens. *Crit Public Health*. 2012;22(1):99-105. doi:10.1080/09581596.2010.520692
165. Eicher-Miller HA, Fulgoni VL, Keast DR. Contributions of processed foods to dietary intake in the US from 2003-2008: a report of the Food and Nutrition Science Solutions Joint Task Force of the Academy of Nutrition and Dietetics, American Society for Nutrition, Institute of Food Technologists, and International Food Information Council. *J Nutr*. 2012;142(11):2065S-2072S. doi:10.3945/jn.112.164442
166. Suchacka G, Chodak G. Using association rules to assess purchase probability in online stores. *Inf Syst E-Bus Manage*. 2017; 15:751-780. doi:10.1007/s10257-016-0329-4
167. Chang H-J, Hung L-P, Ho C-L. An anticipation model of potential customers' purchasing behaviour based on clustering analysis and association rules analysis. *Expert Syst Appl*. 2007; 32(3):753-764. doi:10.1016/j.eswa.2006.01.049
168. Gupta N, Mangal N, Tiwari K, Mitra P. Mining Quantitative Association Rules in Protein Sequences. In: Williams GJ, Simoff SJ, editors. Data Mining: Theory, Methodology, Techniques, and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 273-81. (Lecture Notes in Computer Science). doi:10.1007/11677437_21
169. Doddi S, Marathe A, Ravi SS, Torney DC. Discovery of association rules in medical data. *Med Inform Internet Med.* 2001; 26(1):25-33.

Appendix A: Materials and methods

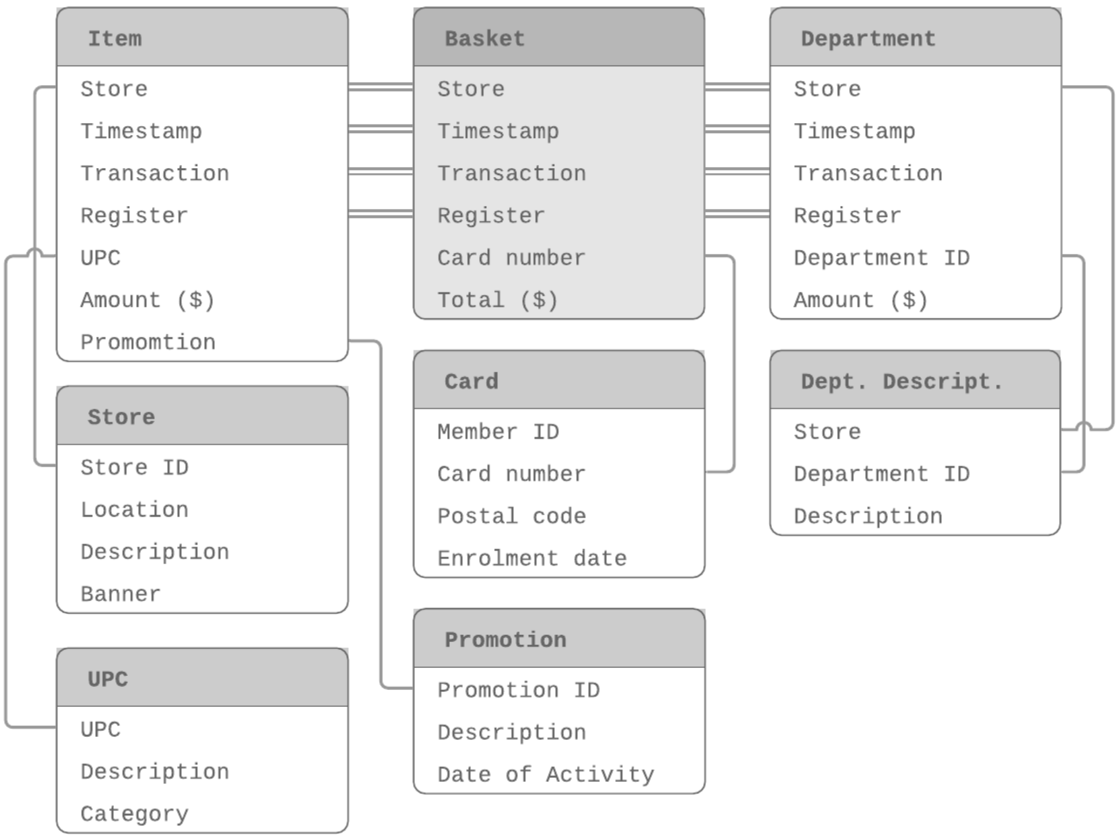
Structure and content of the data

Transaction-level grocery data provided by a large retailer in QC was hierarchically managed within a PostgreSQL database which consisted of a series of tables containing information about individual transactions, card numbers, products, and stores from February 1, 2015 to September 31, 2017. The total size of this database was roughly 650 GB. The data tables were structured in such a way so that information could be recovered using primary keys. Information for each transaction, for example, was recovered from three tables, namely, the basket table, the department table, and the item table. These three tables were linked by a unique, four-column primary key (PK) made up of a store ID, a transaction ID, a timestamp, and a register lane number. This PK uniquely identifies one transaction: a single customer interaction with a checkout.

Individual items purchased within this transaction were found in the department and item tables, respectively. In the department table, this information is given in the form of a department number, quantity, and a spending total. Department numbers are store-specific, and all information regarding departments is found in the department description table. Just over 32% of all purchases in the department table were due to bottle deposit fees, while the rest were devoted to individual department-in-store purchases (making up 1-8% of the total each), for example: pharmacy products, bulk goods, lottery tickets, plants, and tobacco. The department table accounted for $170 million total. Conversely, the item table contained individual retail brand items identified through a unique product code (UPC), and their spending value. The table also contained sparse information on applied promotions or discounts. At this level, there are 145,602 unique UPCs that accounted for a $9.66 trillion total. UPCs were linked to the product table which, although incomplete, contained a short description for each product. An additional table was used to link UPCs to retail categories (UPC categorization is discussed later in the chapter). The store ID was connected to the store table, which included information on the store's location, size, and chain banner.

The basket table was the central table in this trio: each row corresponded to one transaction (or "basket"). There were just over 348 million entries in the basket table, each storing the receipt total of the store transaction and an anonymized card number attached to it. This card number was, in turn, linked to the card table through a unique customer ID which contained information on each loyalty card member, including their date of registration into the loyalty card program, their membership status, and their postal code. Note that one loyalty card member (defined by a member number) may possess multiple card IDs and not all transactions contained a card number. Indeed, only 44% of the basket table had a card number attached to it. A simplified entity relationship diagram (ERD) between tables in the database and their constituents are provided in Figure 1.

Figure 1. ***Simplified entity relationship diagram (ERD) outlining linkages in the database.*** *The central table, basket, links individually purchased items in the Item and department tables to a specific transaction via a four column PK (store, timestamp, transaction, and register), which is thereafter linked to the card table via the card number. As the card number contains a postal code for each household, loyalty card members are linked to census geography and demographic variables. The store table yields location and store banner information on individual stores. Individual items may be further expounded in the UPC and department description table, respectively.*

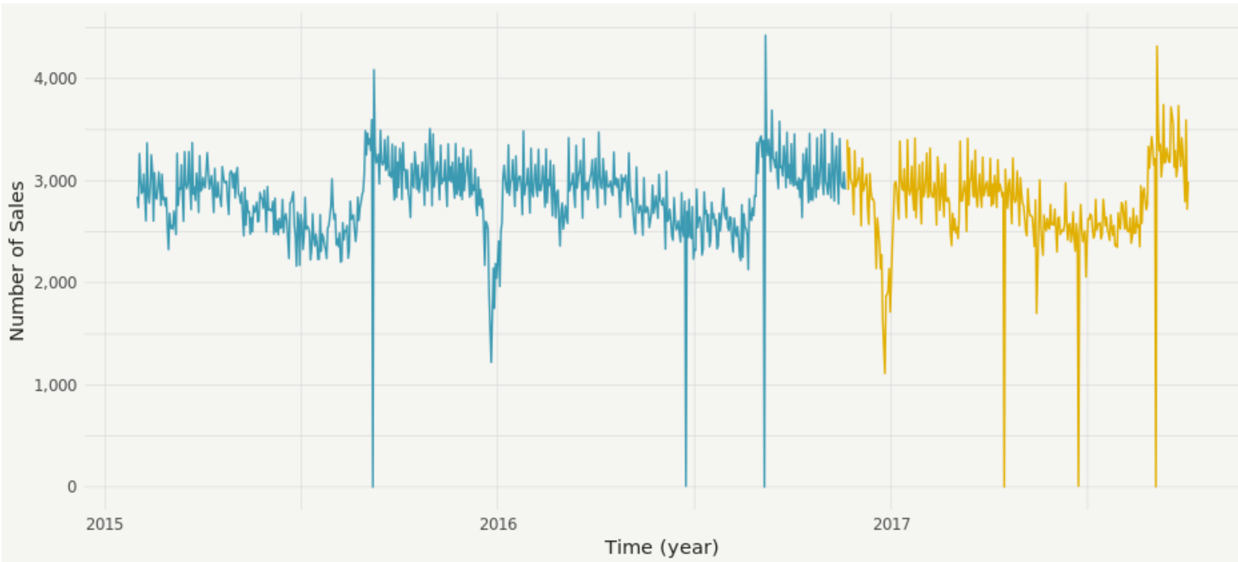


Data processing and preparation

Cleaning the store table

The store table initially consisted of a number of duplicates, albeit with different store IDs. This problem was resolved by retrieving store information from the retailer's website and matching entries based on their postal code. Duplicates were thereafter confirmed by visually inspecting time series plots for each store, showing a continuation in the number of transactions over time across multiple store IDs which were actually the same store (different entries in the store table were created each time a store underwent a banner change, a name change, or a change in ownership). Figure 2 provides an example of a store's transition at a particular date and how this is reflected in their number of transactions over time plot. These cases were double-confirmed using Google Maps. Stores that were unmatched were confirmed to be closed. After processing, there were 208 unique stores in the data. All stores were located in the province of Québec, Canada.

Figure 2. ***Example of the continuity of the number of sales over a store's transition in time.*** *In this particular case, the store underwent a change in ownership in late 2016. The large troughs in number of sales represent statutory holidays in Québec.*



Restricting loyalty card members

Figure 3 provides a flowchart showing the retention of loyalty card members following a number of processing steps. Initially, there were an identified 1,372,546 members in the basket table, making up just over 153 million transactions. 19,076 loyalty card members that moved within the two-and-a-half year period (i.e., member numbers with a change in postal code between card numbers) were omitted. Postal codes were then matched to a concatenated Postal Code Conversion File (PCCF) which allowed linkage between card members and census geography through a point-in-polygon process involving postal code coordinates and dissemination areas (DA) from 2010, 2013, and 2017 versions of the file. 2,506 card members that were not found in any PCCF were dropped. 5,155 card members were with postal codes that did not belong to Québec, and were subsequently dropped. Other criteria for omission included: members with postal codes matched to a DA that belonged to a reserve or another census region whose data was suppressed (1,604), members with transaction timestamps outside of the two year period (1,207), and members whose total expenditure over the two years was zero (4,130). This left 1,338,868 loyalty card members, 97.55% of the initial count. Given the area of interest, the data was then restricted to those card members whose postal code was located on the island of Montréal, leaving a total of 290,261 cards over 30,190 distinct postal codes. 94.3% of loyalty card members in this subset had an active membership status. The principal reason for restricting the database this way was to avoid any bias that might be introduced through a rural-urban divide in purchasing behaviour and the fact that the census geography of rural areas is typically much larger in size and the postal code geo-coordinates are much less accurate in position than those found in urban centers, issues which could introduce errors into the PCCF matching.

Figure 3. ***Retention of card members before sampling.***
Card members were omitted based on whether or not they experienced a postal code change in the two and a half year period, whether or not their postal code was valid and/or matched a PCCF entry, whether or not their postal codes were from QC, whether or not their postal codes corresponded to suppressed census regions, whether they had transactions outside of the study period, and whether or not they had a total expenditure of zero. 1,338,868 loyalty card members remained (97.5% of the original member count). From this subset were the 14,999 sampled.

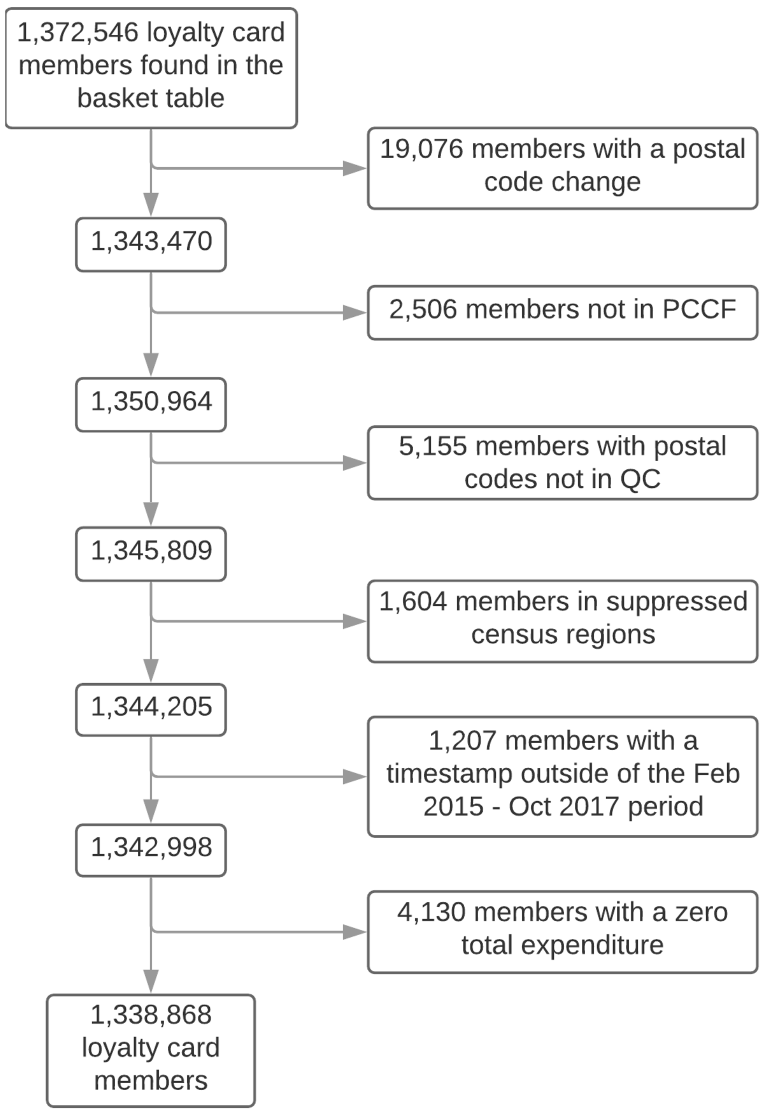
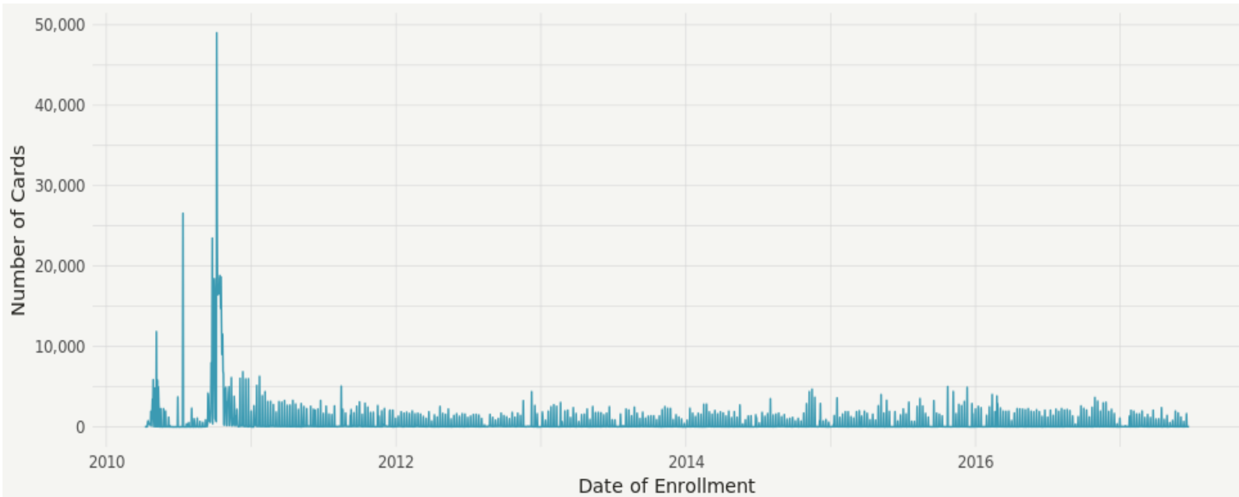


Figure 4. ***Enrolment date for loyalty card members.***

Most loyalty cards appear to have been activated well before the beginning of the study period. The large peaks in 2010 may reflect a new upload or data transfer in the retailer's database.



UPC classification and price extraction

After extracting retail categories from the store's website, UPCs were manually classified with some guidance from the product table. Although only 36% of all UPCs were successfully matched, these UPCs covered 96% of all purchases in the data. Additional category, price, weight, and mean value per unit (MVU) information was also collected from the website and linked through product numbers; however, only 38% of unique UPCs in the collection of sample transactions was successfully matched. Table 1 shows all categories and the number of UPCs associated with each, while Table 2 shows the top five purchased food categories and some example constituents in terms of quantity in the sample. Note that there is no "true" categorization of food products. Conceivably, there are a good many different demarcations that could be created based on the needs of the user. The decision to use pre-existing retail categories was based on convenience and the assumption that these distinctions were already deemed useful enough for the retailer to use in its own business operations. Although a great deal of effort was placed into correctly categorizing products, there is still a possibility that some remain misclassified.

Table 1. Available food product categories and the number of UPCs associated with each. *Categories are presented according to retail department. These categories will be used in the ARM algorithm. Note that the number of UPC products per category does not necessarily correlate with their prevalence among transactions: packaged bread, for example, is not particularly diverse in terms of number of UPCs per category; however it is one of the most frequently purchased items in the sampled member's transactions (see Table 2). This table is a duplicate of that in the manuscript.*

Department	Category	No. UPCs
Snacks	Gluten-free snacks	107
	Nuts, seeds, and dried fruits	1025
	Salty snacks	1278
	Sweet snacks and candy	3888
Pantry	Baking ingredients	900
	Breakfast foods	39
	Canned fruits	76
	Canned meats	29
	Canned soups	440
	Canned vegetables	440
	Canned seafood	405
	Cereal	174
	Cereal bars	122
	Condiments and toppings	1344
	Oils and vinegars	689
	Pasta, rice, and beans	1129
	Spices, herbs, and sauces	2193
	Spreads and syrups	1001
Bread and bakery	Baked bread and baguettes	157
	Buns and rolls	165
	Desserts, pastries, baked goods	1513
	Packaged bread	436
	Tortillas and flatbreads	229
Frozen	Frozen baked goods	582
	Frozen beverages	65
	Frozen fish and seafood	523
	Frozen fruits	75
	Frozen meals and sides	723
	Frozen meat and poultry	333
	Frozen pizza and pasta	392
	Frozen vegetables	127
	Ice cream treats	615

Meat and poultry	Beef and veal	348
	Chicken and turkey	316
	Lamb, horse, and game	66
	Pork	245
	Rabbit and fowl	46
	Sausages and bacon	439
Beverages	Beer and cider	1312
	Drink mixes	142
	Fruit juices and drinks	1125
	Soft drinks	433
	Soy, rice, and nut beverages	109
	Sports and energy drinks	214
	Tea and hot drinks	1187
	Water	366
	Wines, cocktails, and coolers	850
Dairy and eggs	Butter and margarine	109
	Cheese counter	406
	Eggs and substitutes	103
	Milk and cream	492
	Packaged cheese	1146
	Sour cream dips	160
	Yogurt	605
Deli and prepared meals	Antipasto, dips, and pâtés	270
	Deli meats	699
	Ready meals and sides	1678
Vegan and vegetarian	Vegan and vegetarian foods	122
World cuisine	World cuisine products	275
Fish and seafood	Fish and seafood	498
Fruits and vegetables	Fresh herbs	202
	Fruits	1193
	Packaged salad mixes	180
	Vegetables	1490

Table 2. ***Top five most frequently purchased categories and some example constituents.*** *Purchase frequency is based on transactions from the sampled loyalty card members. The following five categories were: packaged cheese, salty snacks, fruit juices/drinks, beef and veal, and yogurt.*

Category	No. transactions (%)	Example items
Vegetables	635,862 (42)	Cucumber, broccoli, shallot, greenhouse tomato, field pepper, celery, lettuce, and spinach.
Fruits	627,243 (42)	Banana, red grape, pineapple, local apple, fig, apricot, honeydew melon, and organic fruits.
Milk and cream	460,569 (30)	Milk, cream, specialty milk, and flavoured milk.
Sweet snacks and candy	329,791 (22)	Fruit snack, cookies, chocolates, pudding, bulk confectionery, bulk chocolates, and candy.
Packaged bread	304,881 (20)	White wheat bread, organic bread, specialty and commercial bread.

Descriptive statistics

Linking postal code to census demography

Figure 5 displays a map of the card members distributed around the island of Montréal. Colors represent the proportion of cards in a particular DA mapped via a hybrid PCCF. The hybrid PCCF was constructed using the 2017, 2013, and 2010 file versions provided by Statistics Canada. The rationale for concatenating the three files was to capture older, inactive postal codes in the database that would not otherwise be linked to a newer PCCF and link them to updated census geographies. As some postal codes may be mapped to multiple unique DAs in the PCCF, card counts within the proportions were weighted according to the populations corresponding to the associated DAs in order to reduce the possibility of overcounting cards in regions. Clusters of loyalty card members correspond roughly to the locations of the supermarkets (not shown). More clusters appear on the Eastern half of the island. Note that some DAs are either uninhabited or have very few census-recorded inhabitants. The distribution of card members according to the weighting scheme across the constituent DAs for each postal code means that some regions may be colored when they would otherwise be grey.

Table 3 provides summary statistics for the sampled cohort of card members alongside island statistics using the 2016 Canadian census. Sample statistics for all 1,338,868 members are also presented. As can be seen, there is very little variation between the three that would give cause for major concern. When differences do occur, they are not significant. In general; however, the card member DA population shows a slightly higher median education status than the island median, a slightly higher population of unmarried individuals living alone, and a lower percentage of immigrants.

Figure 5. ***Distribution of loyalty card members on the island of Montréal.***

Colors correspond to the weighted proportion of the DA population with a loyalty card. As expected, card postal codes are clustered heavily around the store locations (not shown). A heavy concentration of card members is seen on the Eastern half of the island, a predominantly renting population with a lower median household income. Grey areas within the island represent uninhabited DAs, such as municipal parks. The large dark blue regions in the middle of the island correspond roughly to the city airport, the train yards, and their surrounding industrial parks.

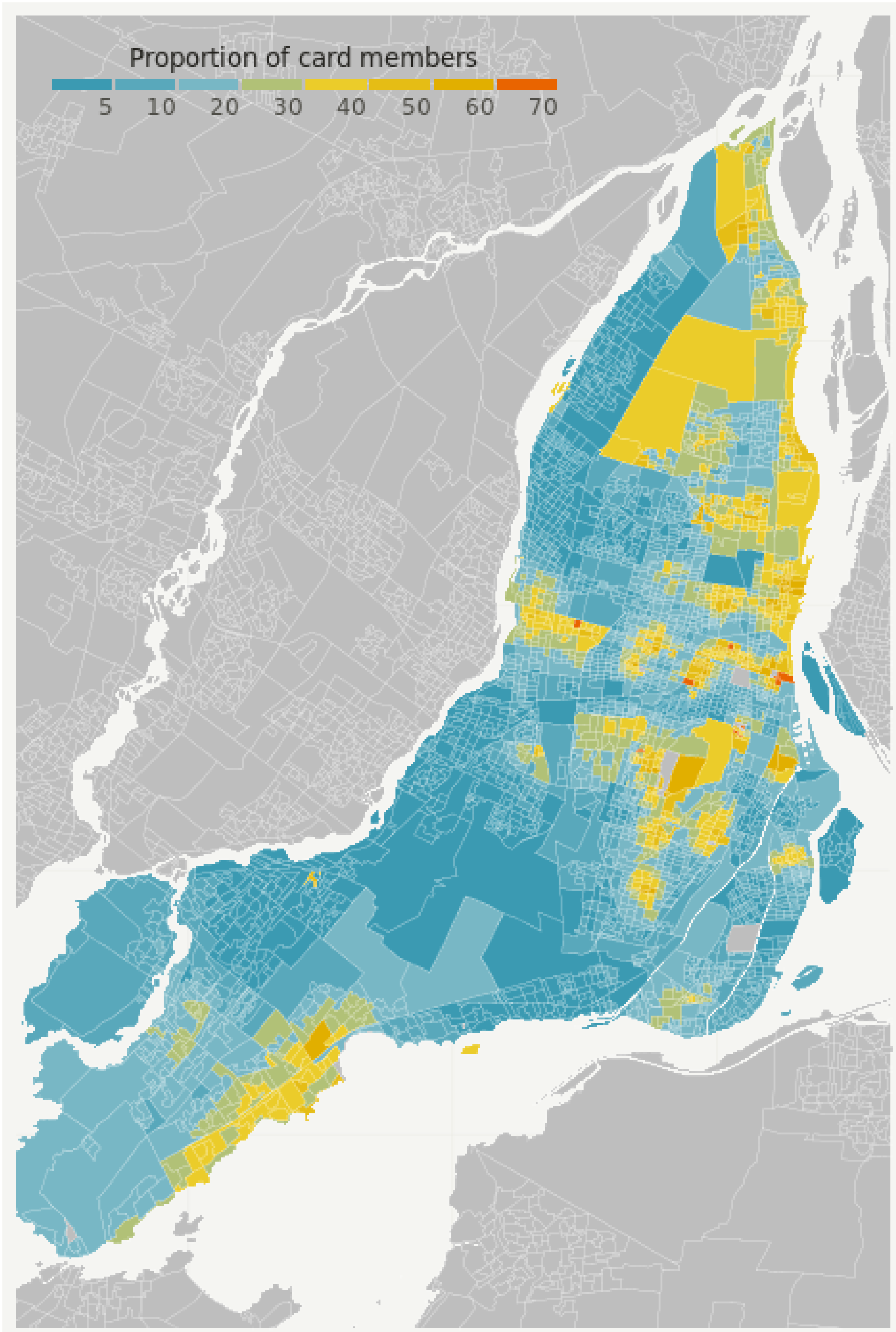


Table 3. Census demographic variables for sampled loyalty card DAs.
Census demography was adopted from the 2016 Canadian Census. Selected demographic variables show the range of social and material deprivation across the sampled card members in relation to the island of Montréal. Sample and all-card statistics are weighted according to the number of card members per DA, while island statistics represent the demography across the entire island without consideration of card members. Medians are given alongside 95% confidence intervals (CI).

Demographic variable	**All card median (95% CI)** *Sample median (95% CI)* Island median
Median household income in 1,000 dollars (CAD)	**51.2 (24.8 - 145.1)** *51.4 (23.9 - 151.5)* 52.3
Population without a certificate, diploma, or degree (%)	**13.6 (2.2 - 36.5)** *13.6 (2.2 - 36.5)* 16.1
Employment rate (%)	**59.6 (33.3 - 78.8)** *59.6 (33.9 - 78.8)* 58.5
Population not married or not in common law (%)	**52.9 (32.9 - 71.9)** *52.6 (32.9 - 72.3)* 50.7
Lone parent families in private households (%)	**19.3 (7.0 - 38.3)** *19.3 (7.0 - 38.3)* 20.0
Total households with one person (%)	**41.9 (9.7 - 66.0)** *41.2 (9.4 - 67.2)* 36.7
Immigrant status (%)	**27.0 (7.7 - 61.4)** *27.0 (7.6 - 61.4)* 31.2
Population (in 1,000s)	**594 (344 - 2,333)** *590 (344 - 2,217)* 546

Temporal patterns of the transaction data of the sample

The hourly distribution of the total number of baskets over a given day and the hourly median expenditure per time of day are presented in Figure 6 and Figure 7, respectively. More transactions occur on weekdays than on weekends. In particular, more weekday shopping occurs in the afternoon between the hours of 15h00 and 18h00. However, there does not seem to be any significant changes in the basket total throughout the working hours of the day, save for the small decreases in the early mornings and late evenings. Note that store hours vary between location and between owner changes of the same store.

Figure 8 shows the number of transactions, denoted here as the number of store visits, per day throughout the study period. Once again, no significant gaps are present. The drops in the time series reflect national and provincial statutory holidays, namely, Easter, Jean-Baptiste Day, Labour Day, Christmas, and New Year's Day. Interestingly, spikes in the data are observed in the days preceding these many of these holidays.

Figure 6. ***Distribution of shopping times across sampled card members.***

Bins represent the number of transactions per hour at any given store. The colors blue and gold denote the number of transactions for weekends (Saturday, Sunday) and weekdays (Monday to Friday), respectively. On weekdays, more transactions occur in the early afternoon, between 15h00 to 18h00.

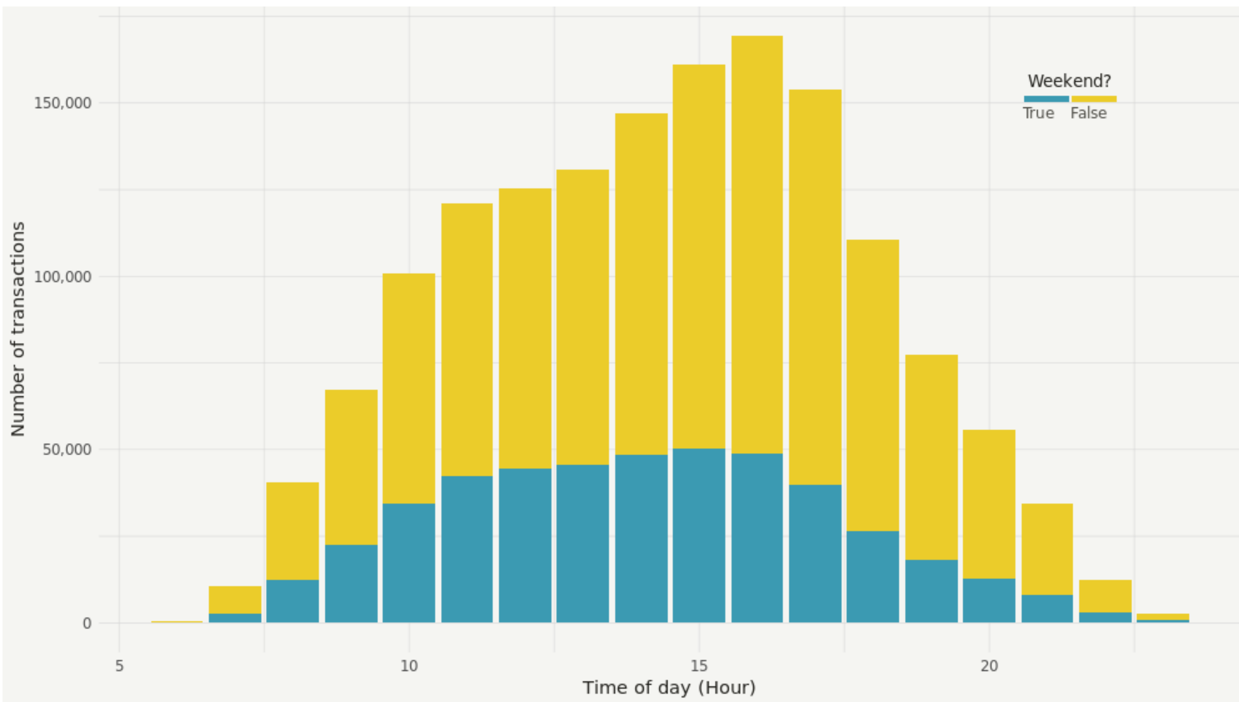


Figure 7. ***Median total expenditure per basket of sampled card members per hour.***
The lower and upper bounds correspond to the first and third quartiles, respectively. The amount spent per transaction is roughly similar throughout the day, with slight decreases in the late evenings and early mornings.

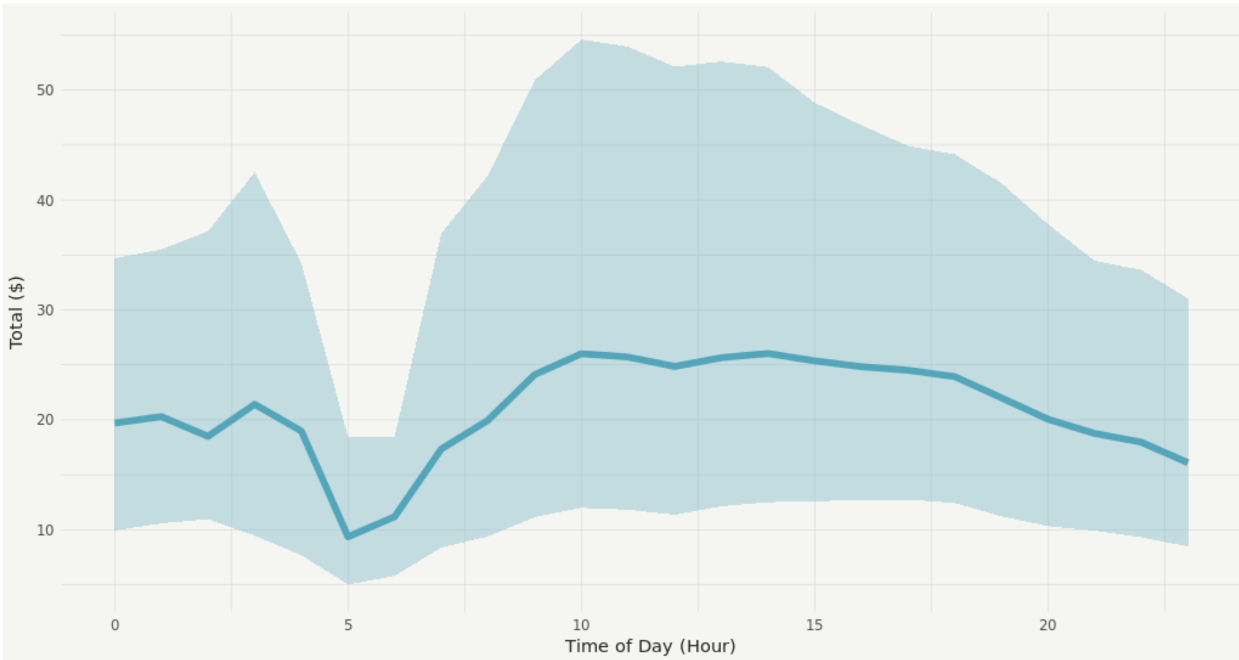
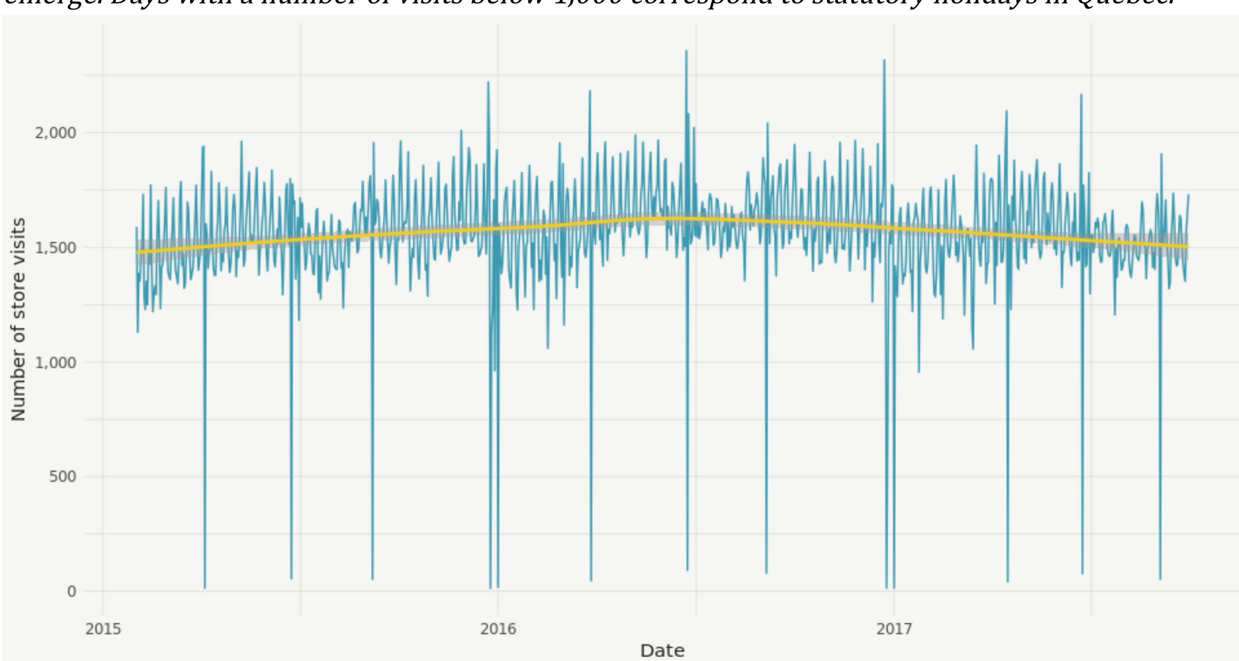


Figure 8. ***Daily visit frequency for all stores and customers.***
A locally-estimated scatterplot smoothing (LOESS) curve is shown in gold. Apart from the variation between days and a slight seasonal effect between months, no major departures or variations emerge. Days with a number of visits below 1,000 correspond to statutory holidays in Québec.



Recency, frequency, and monetary value segmentation

A matrix displaying the frequency and monetary value of the sampled shoppers is presented in Figure 9. The five bins per criterion are calculated according to RFM segmentation, which creates scores for each customer based on the date of the customers' most recent transaction (Recency), the frequency with which the customer shops (Frequency), and total expenditure of the customer (Monetary value). As recency is less important for the scope of this paper, only frequency and monetary value are shown. As shown in Figure 9, most sampled members are either low-frequency, low-expenditure, or high-frequency, high-expenditure. As expected, the number of times a member shops is correlated with their total expenditure. Shop statistics for the size of baskets across sampled members' transactions is presented in Table 4. In general, most baskets contain relatively few items. In general, households spent a median \$24.22 (95% CI: \$2.93 - \$157.55) per visit, purchasing a median of 7 (95% CI: 1 - 40) items, with both distributions heavily right-skewed. This is apparent in the RFM plot as well.

Figure 9. ***Distribution of the sampled members by frequency and monetary value.***
Monetary value, in this case, means total expenditure. Each cell corresponds to an FM score calculated according to RFM segmentation into five bins per criterion. As expected, there is a significant correlation between the number of visits and the total amount spent.

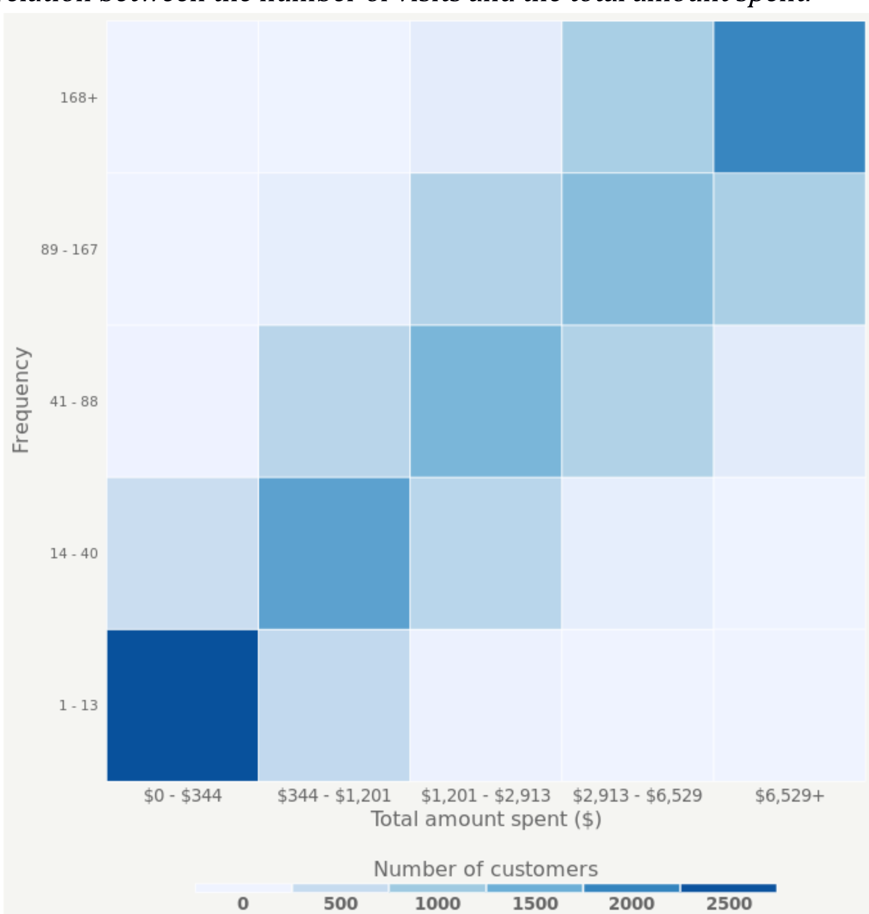


Table 4. Statistics for basket details among sampled loyalty card members.
Median and means are presented alongside the 95% confidence interval (CI). The distribution of basket totals is heavily skewed. The number of items per basket remains modest. On average, transactions are expected to have at least one unmatched UPC.

	Median / mean (95% CI)
Number of items per basket	**7** / 9.4 (1 - 37)
Number of null (unmatched) UPCs per basket	**1** / 1.3 (1 - 2)
Basket total ($)	**24.22** / 38.11 (2.98 - 157.55)

Association rule mining

As a rule-based machine learning technique, association rule mining (ARM) discovers relations between products in a database of transactions in the form of conditionals; that is, if a transaction contains an item X, then it also contains an item Y, and this relationship is codified as $X \rightarrow Y$, where X is the antecedent and Y is the consequent. Initially developed in the early 1990s, association rule mining has typically been used in a number of retail environments in order to analyse customers' purchasing behaviour, either online or in-store.[166,167] Other applications include mining spatial relationships in census data, uncovering the co-occurrence of certain amino acids in a protein, and identifying relationships between procedures performed on a patient and the reported diagnoses in large medical record databases.[168,169]

In applying ARM to the database of transactions for sampled card members, rules were pruned using a minimum support threshold of 0.01 (1%) and a minimum confidence threshold equal to that of the support of the products on the right hand side (consequent). The rationale for such threshold is two-fold: to avoid any rare co-occurrences that may have a high lift and confidence due to its small support, and to capture enough rules that differences in co-purchases may be observed in detail. Moreover, the decision to define products of interest as the consequent rather than the antecedent stems from the fact that discovered rules are heavily conditional on the support of the LHS. A minimum of 1% was used: if any greater, the discovered rules would simply mirror all products seen in Table 2, for those products are the most frequently purchased. Moreover, they compose the most frequently purchased *itemsets* (see Table 5).

All discovered rules for soft drinks, sweet snacks, salty snacks, fruit juices, vegetables, and fruits are presented in Tables 6 through 11 in Appendix B. Lift, a bi-directional measure that compares the confidence of a rule to its expected confidence (i.e., the support of the consequent), is highly susceptible to rare transactions. The odds ratio (OR) of each rule was therefore computed. As can be seen, the OR is highly correlated with lift. The number of card members in which the rule occurs is also presented.

Appendix B: Frequent itemsets and association rules

Frequent itemsets

Table 5. Frequent itemsets for transactions of sampled loyalty card members.

Itemset	No. transactions (%)
Fruits + vegetables	385,137 (25.7)
Fruits + milk and cream	251,257 (16.7)
Vegetables + milk and cream	246,787 (16.4)
Fruits + sweet snacks and candy	183,197 (12.2)
Vegetables + sweet snacks and candy	178,171 (11.9)
Vegetables + packaged cheese	177,468 (11.8)
Fruits + packaged bread	175,009 (11.7)
Vegetables + packaged bread	174,958 (11.7)
Fruits + fruit juices and drinks	169,428 (11.3)
Fruits + packaged cheese	168,213 (11.1)
Vegetables + fruit juices and drinks	166,634 (11.1)
Fruits + yogurt	166,176 (11.1)
Vegetables + beef and veal	161,103 (10.7)
Vegetables + salty snacks	159,718 (10.6)
Fruits + salty snacks	156,120 (10.4)
Vegetables + yogurt	156,035 (10.4)

Association rules

Table 6. Association rules for soft drinks among sampled loyalty card members.

Supp	Conf	Lift	Antecedent	OR (95% CI)	Cards
3.25	17	1.88	Salty snacks	1.82 (1.80 - 1.84)	6,556
1.49	16	1.80	Water	1.77 (1.75 - 1.80)	4,279
1.21	16	1.80	Frozen meals and sides	1.78 (1.75 - 1.81)	3,704
1.38	16	1.79	Sausages and bacon	1.76 (1.74 - 1.79)	4,240
1.18	15	1.70	Butter and margarine	1.68 (1.65 - 1.71)	4,029
1.02	15	1.69	Buns and rolls	1.67 (1.64 - 1.70)	3,602
1.19	15	1.67	Ice cream treats	1.65 (1.62 - 1.67)	4,114
3.25	15	1.64	Sweet snacks and candy	1.59 (1.57 - 1.60)	6,220
2.87	15	1.62	Fruit juices and drinks	1.58 (1.56 - 1.60)	6,261
1.88	15	1.61	Condiments and toppings	1.58 (1.56 - 1.61)	5,418
1.03	14	1.55	Tea and hot drinks	1.53 (1.51 - 1.56)	3,406
2.38	14	1.53	Desserts, pastries, baked goods	1.50 (1.48 - 1.52)	5,373
2.21	14	1.53	Deli meats	1.50 (1.48 - 1.52)	5,071
1.15	14	1.51	Canned vegetables	1.49 (1.47 - 1.52)	3,879
2.34	14	1.51	Ready meals and sides	1.47 (1.46 - 1.49)	5,159
2.73	13	1.49	Packaged bread	1.45 (1.43 - 1.47)	5,487
1.25	13	1.47	Spreads and syrups	1.45 (1.43 - 1.48)	4,266
2.24	13	1.45	Beef and veal	1.42 (1.40 - 1.44)	5,044
2.52	13	1.43	Packaged cheese	1.40 (1.38 - 1.41)	5,785
1.23	13	1.42	Pork	1.41 (1.38 - 1.43)	3,914
1.79	13	1.42	Spices, herbs, and sauces	1.39 (1.37 - 1.41)	5,038
1.60	13	1.40	Pasta, rice, and beans	1.38 (1.36 - 1.40)	4,806
1.04	12	1.34	Cereal	1.33 (1.31 - 1.35)	3,688
3.71	12	1.34	Milk and cream	1.29 (1.28 - 1.30)	6,554
1.01	12	1.33	Cheese counter	1.32 (1.30 - 1.34)	3,690
1.50	12	1.33	Eggs and substitutes	1.31 (1.29 - 1.33)	4,948
1.34	11	1.26	Chicken and turkey	1.24 (1.22 - 1.26)	4,367
1.16	11	1.21	Baked bread and baguettes	1.19 (1.18 - 1.21)	3,938
1.85	11	1.20	Yogurt	1.18 (1.16 - 1.20)	4,989
4.29	10	1.12	Vegetables	1.07 (1.06 - 1.09)	7,314
4.19	10	1.11	Fruits	1.07 (1.06 - 1.08)	7,307

Table 7. Association rules for sweet snacks/candy among sampled loyalty card members.

Supp	Conf	Lift	Antecedent	OR (95% CI)	Cards
2.80	45	2.05	Nuts, seeds, and dried fruits	1.99 (1.97 - 2.02)	6,827
3.52	41	1.87	Cereal	1.80 (1.78 - 1.82)	7,366
2.13	40	1.83	Canned soups	1.79 (1.77 - 1.81)	5,750
3.68	39	1.77	Spreads and syrups	1.71 (1.69 - 1.73)	8,195
1.48	39	1.76	Canned seafood	1.73 (1.70 - 1.76)	4,860
7.27	38	1.73	Salty snacks	1.60 (1.59 - 1.61)	9,638
2.75	37	1.69	Tea and hot drinks	1.64 (1.63 - 1.66)	6,632
3.10	37	1.67	Canned vegetables	1.62 (1.60 - 1.64)	7,265
2.83	37	1.67	Butter and margarine	1.62 (1.60 - 1.64)	7,376
2.24	37	1.67	Antipasto, dips, and pâtés	1.63 (1.61 - 1.65)	6,154
2.73	37	1.66	Frozen meals and sides	1.62 (1.60 - 1.64)	6,488
2.91	37	1.66	Baking ingredients	1.61 (1.60 - 1.63)	7,569
7.14	36	1.66	Fruit juices and drinks	1.54 (1.53 - 1.55)	9,508
6.19	36	1.65	Yogurt	1.55 (1.54 - 1.56)	8,988
3.25	36	1.64	Soft drinks	1.59 (1.57 - 1.60)	6,220
1.04	36	1.64	World cuisine products	1.62 (1.59 - 1.65)	4,474
1.97	36	1.64	Frozen pizza and pasta	1.60 (1.58 - 1.63)	5,869
1.72	36	1.63	Tortillas and flatbreads	1.60 (1.58 - 1.63)	5,222
4.53	36	1.63	Pasta, rice, and beans	1.55 (1.54 - 1.57)	8,619
2.82	35	1.61	Ice cream treats	1.57 (1.55 - 1.58)	7,001
7.13	35	1.60	Packaged bread	1.48 (1.47 - 1.49)	9,079
6.01	35	1.59	Desserts, pastries, baked goods	1.49 (1.48 - 1.51)	9,055
2.98	35	1.59	Sausages and bacon	1.54 (1.52 - 1.56)	7,056
4.48	35	1.58	Condiments and toppings	1.51 (1.50 - 1.53)	8,852
2.89	35	1.57	Cheese counter	1.53 (1.51 - 1.54)	6,798
1.53	34	1.55	Frozen fish and seafood	1.53 (1.50 - 1.55)	5,250
1.33	34	1.55	Oils and vinegars	1.53 (1.50 - 1.55)	5,615
6.62	34	1.54	Packaged cheese	1.44 (1.43 - 1.45)	9,490
4.69	34	1.53	Spices, herbs, and sauces	1.45 (1.44 - 1.47)	8,780
3.04	33	1.51	Water	1.47 (1.45 - 1.48)	6,335
5.29	33	1.51	Deli meats	1.43 (1.41 - 1.44)	8,343
1.28	33	1.50	Soy, rice, and nut beverages	1.48 (1.46 - 1.51)	3,435
3.16	33	1.50	Pork	1.45 (1.44 - 1.47)	7,033
2.22	33	1.50	Buns and rolls	1.47 (1.45 - 1.49)	6,051
4.10	33	1.49	Eggs and substitutes	1.43 (1.41 - 1.44)	8,641
5.58	33	1.48	Ready meals and sides	1.40 (1.38 - 1.41)	8,507
3.69	31	1.42	Chicken and turkey	1.37 (1.35 - 1.38)	7,836
9.58	31	1.42	Milk and cream	1.28 (1.27 - 1.29)	10,218
2.87	31	1.41	Packaged salads	1.37 (1.36 - 1.39)	6,816
5.29	31	1.41	Beef and veal	1.33 (1.32 - 1.35)	8,371
2.71	30	1.34	Fish and seafood	1.31 (1.29 - 1.32)	6,196
12.21	29	1.33	Fruits	1.17 (1.16 - 1.17)	11,146
1.14	28	1.28	Fresh herbs	1.26 (1.24 - 1.29)	4,219
11.88	28	1.27	Vegetables	1.12 (1.12 - 1.13)	11,070
2.95	28	1.26	Baked bread and baguettes	1.22 (1.21 - 1.24)	6,771
1.42	24	1.07	Wines, cocktails, and coolers	1.06 (1.04 - 1.07)	4,250
1.55	23	1.03	Beer and cider	1.01 (1.00 - 1.03)	4,655

Table 8. Association rules for salty snacks among sampled loyalty card members.

Supp	Conf	Lift	Antecedent	OR (95% CI)	Cards
2.37	39	2.03	Antipasto, dips, and pâtés	1.98 (1.96 - 2.01)	6,522
2.30	37	1.94	Nuts, seeds, and dried fruits	1.89 (1.87 - 1.92)	6,322
1.94	37	1.92	Canned soups	1.88 (1.86 - 1.91)	5,690
3.25	36	1.88	Soft drinks	1.82 (1.80 - 1.84)	6,556
1.38	36	1.88	Canned seafood	1.85 (1.82 - 1.89)	4,842
4.57	35	1.85	Condiments and toppings	1.77 (1.75 - 1.79)	8,938
2.87	34	1.79	Cheese counter	1.74 (1.72 - 1.76)	6,863
1.60	33	1.74	Tortillas and flatbreads	1.71 (1.69 - 1.74)	5,155
7.27	33	1.73	Sweet snacks and candy	1.60 (1.59 - 1.61)	9,638
2.80	33	1.72	Sausages and bacon	1.67 (1.65 - 1.69)	7,062
2.82	33	1.71	Cereal	1.67 (1.65 - 1.69)	6,767
2.69	32	1.67	Canned vegetables	1.62 (1.60 - 1.64)	6,957
1.75	32	1.67	Frozen pizza and pasta	1.64 (1.62 - 1.66)	5,632
2.92	32	1.67	Water	1.62 (1.60 - 1.64)	6,342
3.01	32	1.67	Spreads and syrups	1.62 (1.60 - 1.63)	7,683
6.22	32	1.67	Packaged cheese	1.56 (1.55 - 1.57)	9,344
4.01	32	1.65	Pasta, rice, and beans	1.59 (1.57 - 1.60)	8,325
2.35	32	1.65	Frozen meals and sides	1.61 (1.59 - 1.63)	6,264
2.12	31	1.64	Buns and rolls	1.61 (1.59 - 1.63)	6,114
2.39	31	1.62	Butter and margarine	1.58 (1.56 - 1.60)	6,981
4.32	31	1.61	Spices, herbs, and sauces	1.54 (1.53 - 1.56)	8,504
1.21	31	1.61	Oils and vinegars	1.59 (1.57 - 1.62)	5,426
6.04	31	1.61	Fruit juices and drinks	1.51 (1.50 - 1.53)	9,150
2.43	31	1.59	Ice cream treats	1.56 (1.54 - 1.57)	6,689
2.24	30	1.58	Tea and hot drinks	1.55 (1.53 - 1.57)	6,088
1.35	30	1.57	Frozen fish and seafood	1.55 (1.53 - 1.58)	5,058
4.74	30	1.55	Deli meats	1.48 (1.46 - 1.49)	8,112
2.36	30	1.55	Baking ingredients	1.51 (1.49 - 1.53)	6,924
5.01	29	1.54	Yogurt	1.46 (1.45 - 1.47)	8,407
5.93	29	1.52	Packaged bread	1.43 (1.42 - 1.45)	8,727
2.67	29	1.51	Packaged salads	1.47 (1.45 - 1.49)	6,588
4.94	29	1.50	Desserts, pastries, baked goods	1.43 (1.41 - 1.44)	8,684
3.58	29	1.49	Eggs and substitutes	1.44 (1.43 - 1.46)	8,150
2.70	28	1.48	Pork	1.44 (1.42 - 1.45)	6,676
1.09	28	1.47	Soy, rice, and nut beverages	1.46 (1.43 - 1.48)	3,206
1.13	28	1.46	Fresh herbs	1.44 (1.42 - 1.47)	4,216
4.78	28	1.46	Ready meals and sides	1.39 (1.37 - 1.40)	8,299
1.85	27	1.40	Beer and cider	1.38 (1.36 - 1.40)	5,158
4.58	27	1.40	Beef and veal	1.34 (1.32 - 1.35)	8,103
3.16	27	1.40	Chicken and turkey	1.35 (1.34 - 1.37)	7,428
2.41	26	1.37	Fish and seafood	1.34 (1.33 - 1.36)	5,925
7.98	26	1.36	Milk and cream	1.25 (1.24 - 1.26)	9,765
2.72	26	1.33	Baked bread and baguettes	1.30 (1.28 - 1.31)	6,703
10.65	25	1.31	Vegetables	1.17 (1.16 - 1.18)	10,682
10.41	25	1.30	Fruits	1.16 (1.16 - 1.17)	10,667
1.47	24	1.27	Wines, cocktails, and coolers	1.25 (1.24 - 1.27)	4,356

Table 9. Association rules for juices/other drinks among sampled loyalty card members.

Supp	Conf	Lift	Antecedent	OR (95% CI)	Cards
1.99	38	1.92	Canned soups	1.88 (1.85 - 1.91)	5,585
1.43	37	1.90	Canned seafood	1.87 (1.84 - 1.90)	4,633
3.18	37	1.89	Cereal	1.83 (1.81 - 1.85)	6,883
3.11	37	1.88	Canned vegetables	1.82 (1.80 - 1.85)	7,253
2.27	36	1.86	Nuts, seeds, and dried fruits	1.82 (1.80 - 1.84)	6,071
2.68	35	1.77	Butter and margarine	1.72 (1.70 - 1.74)	7,158
3.27	35	1.77	Spreads and syrups	1.71 (1.69 - 1.73)	7,742
5.87	35	1.76	Yogurt	1.66 (1.64 - 1.67)	8,820
4.33	34	1.74	Pasta, rice, and beans	1.67 (1.65 - 1.68)	8,417
2.06	34	1.72	Antipasto, dips, and pâtés	1.69 (1.66 - 1.71)	5,856
4.19	33	1.71	Eggs and substitutes	1.63 (1.62 - 1.65)	8,696
4.27	33	1.69	Condiments and toppings	1.62 (1.61 - 1.64)	8,588
1.81	33	1.69	Frozen pizza and pasta	1.66 (1.64 - 1.68)	5,595
2.81	33	1.68	Sausages and bacon	1.63 (1.61 - 1.65)	6,946
1.57	33	1.67	Tortillas and flatbreads	1.64 (1.62 - 1.67)	4,915
2.73	33	1.67	Cheese counter	1.62 (1.60 - 1.64)	6,610
1.28	33	1.67	Oils and vinegars	1.64 (1.62 - 1.67)	5,488
2.98	33	1.66	Water	1.61 (1.60 - 1.63)	6,341
7.14	32	1.66	Sweet snacks and candy	1.54 (1.53 - 1.55)	9,508
4.53	32	1.65	Spices, herbs, and sauces	1.58 (1.56 - 1.59)	8,674
2.41	32	1.65	Frozen meals and sides	1.61 (1.59 - 1.63)	6,113
2.39	32	1.65	Tea and hot drinks	1.61 (1.59 - 1.63)	6,201
6.25	32	1.63	Packaged cheese	1.53 (1.52 - 1.54)	9,294
2.87	32	1.62	Soft drinks	1.58 (1.56 - 1.60)	6,261
2.52	32	1.62	Baking ingredients	1.58 (1.56 - 1.60)	7,091
1.43	32	1.62	Frozen fish and seafood	1.60 (1.57 - 1.62)	5,052
6.04	32	1.61	Salty snacks	1.51 (1.50 - 1.53)	9,150
6.39	31	1.61	Packaged bread	1.50 (1.49 - 1.51)	8,911
1.20	31	1.59	Soy, rice, and nut beverages	1.57 (1.54 - 1.60)	3,330
2.94	31	1.57	Pork	1.52 (1.51 - 1.54)	6,846
2.42	30	1.55	Ice cream treats	1.51 (1.50 - 1.53)	6,660
5.17	30	1.54	Desserts, pastries, baked goods	1.46 (1.44 - 1.47)	8,662
9.16	30	1.52	Milk and cream	1.38 (1.37 - 1.39)	10,127
1.21	30	1.52	Fresh herbs	1.50 (1.48 - 1.53)	4,298
2.75	30	1.52	Packaged salads	1.48 (1.46 - 1.50)	6,718
1.99	30	1.51	Buns and rolls	1.48 (1.46 - 1.50)	5,807
4.72	30	1.51	Deli meats	1.44 (1.42 - 1.45)	8,124
3.45	29	1.49	Chicken and turkey	1.44 (1.42 - 1.45)	7,611
4.94	29	1.47	Beef and veal	1.40 (1.39 - 1.41)	8,254
4.82	28	1.43	Ready meals and sides	1.36 (1.35 - 1.38)	8,206
2.54	28	1.41	Fish and seafood	1.38 (1.36 - 1.40)	6,027
11.29	27	1.38	Fruits	1.22 (1.21 - 1.23)	11,024
11.11	26	1.34	Vegetables	1.19 (1.18 - 1.20)	10,937
2.68	25	1.28	Baked bread and baguettes	1.25 (1.24 - 1.26)	6,579
1.56	23	1.16	Beer and cider	1.14 (1.12 - 1.16)	4,674
1.31	22	1.11	Wines, cocktails, and coolers	1.10 (1.08 - 1.12)	4,088

Table 10. Association rules for vegetables among sampled loyalty card members.

Supp	Conf	Lift	Antecedent	OR (95% CI)	Cards
3.39	84	1.97	Fresh herbs	1.90 (1.88 - 1.93)	7,117
6.75	73	1.73	Packaged salads	1.61 (1.60 - 1.62)	9,374
1.48	71	1.68	Vegan and vegetarian foods	1.65 (1.62 - 1.68)	3,742
5.83	69	1.63	Canned vegetables	1.54 (1.52 - 1.55)	9,469
3.27	68	1.60	Tortillas and flatbreads	1.55 (1.53 - 1.57)	7,095
1.96	68	1.59	World cuisine products	1.56 (1.54 - 1.59)	6,401
5.64	67	1.59	Cheese counter	1.50 (1.49 - 1.51)	8,672
2.58	67	1.59	Canned seafood	1.55 (1.53 - 1.57)	6,464
1.87	67	1.59	Sour cream dips	1.56 (1.53 - 1.58)	6,201
9.31	67	1.57	Spices, herbs, and sauces	1.43 (1.41 - 1.44)	10,827
2.59	66	1.56	Oils and vinegars	1.52 (1.50 - 1.54)	7,787
2.96	66	1.56	Frozen fish and seafood	1.51 (1.49 - 1.53)	7,324
4.01	66	1.55	Antipasto, dips, and pâtés	1.49 (1.47 - 1.51)	7,955
1.19	66	1.55	Frozen fruits	1.53 (1.50 - 1.56)	3,903
8.28	65	1.54	Pasta, rice, and beans	1.41 (1.40 - 1.42)	10,552
4.05	65	1.54	Nuts, seeds, and dried fruits	1.48 (1.46 - 1.49)	7,891
8.37	65	1.53	Condiments and toppings	1.41 (1.39 - 1.42)	10,772
6.17	65	1.52	Pork	1.43 (1.42 - 1.44)	9,154
7.61	64	1.52	Chicken and turkey	1.40 (1.39 - 1.41)	10,215
3.38	64	1.50	Canned soups	1.45 (1.44 - 1.47)	7,367
1.68	64	1.50	Frozen vegetables	1.48 (1.46 - 1.50)	5,701
5.78	63	1.49	Fish and seafood	1.40 (1.39 - 1.42)	8,212
10.74	63	1.48	Beef and veal	1.32 (1.31 - 1.33)	10,708
7.83	62	1.47	Eggs and substitutes	1.36 (1.35 - 1.37)	10,806
4.75	62	1.45	Butter and margarine	1.39 (1.37 - 1.40)	9,285
25.67	61	1.45	Fruits	1.08 (1.07 - 1.08)	12,873
5.23	61	1.45	Sausages and bacon	1.37 (1.36 - 1.38)	9,063
10.40	61	1.44	Yogurt	1.29 (1.28 - 1.30)	10,471
5.23	61	1.44	Cereal	1.36 (1.35 - 1.38)	8,623
11.83	61	1.43	Packaged cheese	1.26 (1.25 - 1.27)	11,204
4.80	60	1.42	Baking ingredients	1.36 (1.34 - 1.37)	9,093
5.67	60	1.42	Spreads and syrups	1.33 (1.32 - 1.35)	9,691
2.32	60	1.42	Soy, rice, and nut beverages	1.38 (1.36 - 1.40)	4,574
9.59	60	1.42	Deli meats	1.28 (1.27 - 1.29)	10,202
1.20	59	1.38	Frozen meat and poultry	1.37 (1.34 - 1.39)	5,273
3.92	58	1.38	Buns and rolls	1.32 (1.31 - 1.34)	7,957
4.29	58	1.37	Tea and hot drinks	1.31 (1.30 - 1.32)	8,001
11.66	57	1.35	Packaged bread	1.20 (1.19 - 1.20)	10,704
11.11	57	1.34	Fruit juices and drinks	1.19 (1.18 - 1.20)	10,937
10.65	56	1.31	Salty snacks	1.17 (1.16 - 1.18)	10,682
5.86	55	1.30	Baked bread and baguettes	1.22 (1.21 - 1.23)	8,707
11.88	54	1.27	Sweet snacks and candy	1.12 (1.12 - 1.13)	11,070
4.94	54	1.27	Water	1.21 (1.20 - 1.22)	7,555
16.45	54	1.26	Milk and cream	1.06 (1.05 - 1.06)	11,701
2.92	53	1.26	Frozen pizza and pasta	1.22 (1.21 - 1.24)	7,289
3.94	53	1.25	Frozen meals and sides	1.20 (1.19 - 1.21)	8,038
4.20	53	1.25	Ice cream treats	1.19 (1.18 - 1.21)	8,424
9.06	53	1.24	Desserts, pastries, baked goods	1.13 (1.12 - 1.14)	10,548
8.42	49	1.16	Ready meals and sides	1.06 (1.05 - 1.07)	10,053
4.29	48	1.12	Soft drinks	1.07 (1.06 - 1.09)	7,314

2.73	45	1.07	Wines, cocktails, and coolers	1.04 (1.03 - 1.05)	5,793
2.94	43	1.01	Beer and cider	0.98 (0.97 - 0.99)	6,096

Table 11. Association rules for fruits among sampled loyalty card members.

Supp	Conf	Lift	Antecedent	OR (95% CI)	Cards
1-1 association rules for fruits					
2.83	70	1.67	Fresh herbs	1.62 (1.60 - 1.64)	6,360
1.23	68	1.63	Frozen fruits	1.61 (1.58 - 1.65)	3,959
4.18	67	1.61	Nuts, seeds, and dried fruits	1.54 (1.53 - 1.56)	7,974
6.14	66	1.59	Packaged salads	1.49 (1.48 - 1.51)	8,854
11.08	65	1.56	Yogurt	1.38 (1.37 - 1.39)	10,608
5.54	65	1.54	Cereal	1.46 (1.44 - 1.47)	8,747
2.46	64	1.54	Canned seafood	1.50 (1.48 - 1.52)	6,173
1.34	64	1.53	Vegan and vegetarian foods	1.51 (1.49 - 1.54)	3,555
2.45	63	1.52	Soy, rice, and nut beverages	1.48 (1.46 - 1.50)	4,598
5.30	63	1.51	Cheese counter	1.43 (1.42 - 1.45)	8,455
3.00	62	1.49	Tortillas and flatbreads	1.45 (1.43 - 1.46)	6,802
3.79	62	1.49	Antipasto, dips, and pâtés	1.43 (1.42 - 1.45)	7,729
5.62	61	1.47	Fish and seafood	1.39 (1.37 - 1.40)	8,094
2.74	61	1.46	Frozen fish and seafood	1.42 (1.40 - 1.44)	7,032
2.39	61	1.46	Oils and vinegars	1.42 (1.40 - 1.44)	7,452
5.76	61	1.46	Spreads and syrups	1.37 (1.36 - 1.39)	9,663
25.67	61	1.45	Vegetables	1.08 (1.07 - 1.08)	12,873
7.59	61	1.45	Eggs and substitutes	1.34 (1.33 - 1.35)	10,491
1.66	60	1.43	Sour cream dips	1.40 (1.38 - 1.43)	5,770
3.16	60	1.43	Canned soups	1.38 (1.37 - 1.40)	7,021
1.72	59	1.42	World cuisine products	1.40 (1.37 - 1.42)	5,937
4.98	59	1.41	Canned vegetables	1.34 (1.33 - 1.36)	8,935
4.70	59	1.41	Baking ingredients	1.35 (1.33 - 1.36)	9,041
7.49	59	1.41	Pasta, rice, and beans	1.31 (1.30 - 1.32)	10,208
1.54	58	1.39	Frozen vegetables	1.37 (1.35 - 1.40)	5,278
4.49	58	1.39	Butter and margarine	1.33 (1.32 - 1.34)	9,044
6.84	58	1.38	Chicken and turkey	1.29 (1.28 - 1.30)	9,824
4.27	58	1.38	Tea and hot drinks	1.32 (1.31 - 1.33)	7,974
7.42	58	1.38	Condiments and toppings	1.28 (1.27 - 1.29)	10,416
11.29	58	1.38	Fruit juices and drinks	1.22 (1.21 - 1.23)	11,024
5.50	57	1.37	Pork	1.30 (1.29 - 1.31)	8,767
11.21	57	1.37	Packaged cheese	1.22 (1.21 - 1.23)	11,032
11.67	57	1.37	Packaged bread	1.21 (1.20 - 1.22)	10,680
7.95	57	1.36	Spices, herbs, and sauces	1.25 (1.24 - 1.26)	10,440
12.21	56	1.33	Sweet snacks and candy	1.17 (1.16 - 1.17)	11,146
5.04	55	1.32	Water	1.25 (1.24 - 1.26)	7,695
8.79	55	1.32	Deli meats	1.20 (1.19 - 1.21)	9,976
4.36	55	1.31	Ice cream treats	1.25 (1.24 - 1.27)	8,419
4.66	55	1.31	Sausages and bacon	1.24 (1.23 - 1.26)	8,681
16.75	55	1.30	Milk and cream	1.09 (1.08 - 1.09)	11,654
10.41	54	1.30	Salty snacks	1.16 (1.16 - 1.17)	10,667

9.32	54	1.30	Desserts, pastries, baked goods	1.18 (1.17 - 1.18)	10,592
9.12	53	1.28	Beef and veal	1.16 (1.15 - 1.17)	10,221
2.88	53	1.26	Frozen pizza and pasta	1.22 (1.21 - 1.24)	7,214
1.07	52	1.25	Frozen meat and poultry	1.24 (1.21 - 1.26)	4,821
5.57	52	1.25	Baked bread and baguettes	1.18 (1.17 - 1.19)	8,474
3.88	52	1.24	Frozen meals and sides	1.20 (1.18 - 1.21)	7,917
3.48	52	1.24	Buns and rolls	1.19 (1.18 - 1.21)	7,564
8.58	50	1.20	Ready meals and sides	1.09 (1.09 - 1.10)	10,031
4.19	47	1.11	Soft drinks	1.07 (1.06 - 1.08)	7,307