Selling Fresh Processed Produce Behind Glass: A Closed Case?
The Effect of Doors on Fresh Processed Produce Sales

By Travis Burkel

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL and RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: ____________________

Travis Burkel

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

____________________ May 13, 2015

Dr. Gary Thompson Date

Acknowledgement

I would like to acknowledge and thank all of my professors for their help and support through this journey. Specifically, I would like to thank my thesis advisor Dr. Thompson for his guidance, long hours and continuous words of encouragement. Also, my committee members Dr. Wilson and Dr. Aradhyula for their support.

Thank you to my cohort for all of their help and support. Together we faced adversity, we grew as a team and we conquered some of the most difficult academic challenges we have ever faced. We entered as strangers, but we will graduate as a family.

Thank you to my family and friends for their unconditional love throughout this process. My Mom, Celia, for always being there for me and making sure my refrigerator was well stocked. My Dad, Richard, for always having a witty remark to take the edge off of any situation. My Sister, Amy, for always keeping me grounded. My girlfriend, Katie, for being my biggest fan and always making me smile. Finally, my cycling family for all of their love and support.

Thank you all, this would not have been possible without your love and encouragement.

"As they say onward and upward" – Dr. Gary Thompson

This research was supported by USDA-NIFA (Specialty Crops Research Initiative Award No.2010-51181-21230) and USDA-ARS (Project No. 1275-43440-004-00D).

Table of Contents

List of Figures

List of Tables

Abstract

Literature shows that doors can significantly increase the energy efficiency of refrigerated cases and decrease freeze damage and spoilage of fresh processed produce products. A potential drawback to energy and shrink savings is a potential loss in sales and increased labor costs associated with doors on refrigerated display cases. This thesis examines the effect of installing doors on display cases containing fresh processed produce on weekly sales. Through the use of a difference-in-difference model, matching treatment stores to control groups and using statistical and geographic matching, the weekly effect of doors on sales is estimated.

Previous research suggested that cases containing other products experienced no change in sales after the installation of doors. The estimated effects of doors on refrigerated cases containing fresh processed produce are statistically insignificant, negative and represent a 1 to 9 percent loss in sales.

Chapter 1. Introduction

The fresh processed produce industry emerged in the late 1980's. The concept of prepackaged chopped salads was pioneered by Fresh Express and their innovation was quickly followed by Dole (Dole 2015) (Fresh Express 2015). Fresh processed produce provided customers with an entirely new eating experience. They could now have a restaurant quality salad at home with little to no preparation time. This concept was further solidified through the introduction of "salad kits." Salad kits combined leafy greens with other vegetables and included extras such as dressing, cheeses, and even meat products for a true hassle-free, ready-to-eat experience. Since the creation of the fresh processed produce, there has been continuous effort put forth to find new and innovative ways to preserve the freshness of the product and improve the safety of the product (Dole 2015) (Fresh Express 2015).

For more than two decades there have been major advances in the technology surrounding fresh processed produce, from advancements in harvesting, cleaning, and processing all the way to the atmosphere inside the product packaging. A notable advancement in product safety, quality and freshness has occurred in refrigeration of the product. Within hours of harvesting, the lettuce is cleaned and cooled to around 36 degrees Fahrenheit. This minimizes the chances that harmful pathogens are present and grow on the product. The immediate cooling serves to prevent pathogen growth and also works to preserve the product's freshness. From the moment the product is cooled begins the "cold chain." From this moment on, the product must be kept at around 36 degrees Fahrenheit. Any break in the cold chain can be catastrophic for the product. If the product gets any colder it will suffer from freeze damage. If it becomes warmer pathogens on the lettuce will begin to grow exponentially causing a potentially major food safety risk. In both cases, the freshness and quality of the product are compromised. Maintaining the cold chain is an immensely difficult and costly process. The product must be kept cold while at the processing plant, during transport to retailers, and while on display for sale.

Typically, breaks in the cold chain occur at the retailer level while product is waiting to be put on display. The process of loading and unloading the product into trucks for transport from the producer is among the most closely monitored processes with regards to temperature. Shrouds are used to prevent ambient air from entering loading docks while forklifts load the product onto refrigerated tractor trailers. Upon arrival at the retailer facility before the product is stocked in retailer’s cases is when the product is most vulnerable. While there has been substantial investment and attention to detail by producers to increase both efficiency and effectiveness of their methods of refrigeration, there are many areas controlled by retailers that need attention. One of these areas of improvement are retailer display cases.

Fresh processed products are largely displayed using open refrigerated cases, the standard method of display in produce sections of retail stores. While there have been improvements to the energy and cooling efficiency of open cases through the use of air shields and night curtains, these open refrigerators continue to be the standard method of display in the supermarket industry. Even with the latest technological improvements, open cases are still inefficient compared to their closed “doored” counterparts. Open cases consume more energy, have uneven temperature distribution, and pump cold air into the store isles causing potential consumer discomfort and putting extra strain on grocery store’s heating ventilating and air conditioning systems (HVAC). Despite these disadvantages, open refrigerated cases are still the prominent form of display for refrigerated fresh produce in grocery stores. In contrast to most dairy products, almost all refrigerated fresh produce is displayed in open cases. Some dairy products, especially milk, are now more commonly displayed in cases with doors due to high temperature sensitivity. Grocers are reluctant to install enclosed refrigerated cases for fresh processed produce because they fear losing sales. In an industry periodical, a retailer commented "Depending on the margins of the product studied… a 2% decline in sales would "offset all energy savings"" (Garry 2010) Other potential reasons for not installing doors on refrigerated cases include the initial cost of investment, lack of capital, or other investment opportunities with faster and higher rates of return. Many retail grocers believe that when something is placed behind a door two things happen: The product is not perceived as "fresh"; and the presence of the door requires customers take an extra action to get the product into their cart. It can no longer "fall off the shelves" into their cart or basket. There has been little research into the effect of doors on refrigerated cases with regard to sales. The evidence for the energy efficiency of doors is undeniable (Evans 2007) (Fricke and Becker 2010b). Why then, is it that doors are still predominately found solely on refrigerated cases that hold dairy products? It is believed that grocers perceive the estimated loss in sales from doors will be greater than the cost savings from enhanced energy efficiency gains.

This thesis measures whether these concerns about losses in fresh processed produce sales are warranted. The literature supports the energy, temperature, shrink, and food safety benefits of closed refrigerated cases. Until now, there has not been any thorough research on the effect doors have on sales of fresh processed produce (Fricke and Becker 2010a). The outcome of this research has the potential to open up a world of possibilities to grocers. If doors have no effect on sales of fresh processed produce, grocers could install closed cases in their stores, collect on energy savings, experience a reduction in product shrink, increased consumer comfort in store isles, and increase product food safety and shelf life. The potential downsides to closed cases could be more glass surfaces to clean, possible increases in stocking time, and possibly a decrease in sales if consumers perceive a lack of product freshness. This thesis will estimate an econometric model analyzing the effects of closed refrigerated cases on fresh processed produce sales.

Chapter 2. Literature Review

The energy efficiency of incorporating doors on refrigerated cases has been explored extensively. Refrigerated cases account for approximately 60% of a grocers' total energy costs (Evans 2007). Open refrigerated display cabinets consume approximately 1.3 times as much energy as closed cabinets (Fricke and Becker 2010b). Because open cabinets are exposed to ambient air, they must constantly be working to keep their contents at the proper temperature. Open cabinets are highly susceptible to warm air infiltration compared to closed cabinets. To minimize this infiltration, air-jet curtains have been designed for open cabinets to reduce the intrusion of warm ambient air from the outside. These air jets, however, are ineffective and still lead to significant temperature differentials within the case. Curtains are only lowered during the hours the store is closed so energy savings are limited. It has been documented that 80% of the heat load for open cases was caused by ambient air infiltration, compared to 10% infiltration for cases with doors (Kou, et al. 2014). When examining the energy consumption of open and closed cases, it has been found that closed cases consume less energy and energy consumption fluctuates very little. By contrast, open case energy usage fluctuates depending upon the season and the relative humidity of the store (Fricke and Becker 2010a). The energy savings of closed cabinets has been estimated as high as 80% compared to open cabinets (Islam 2012). This savings can fluctuate depending on the lighting, anti-fogging, and defrost technologies that are incorporated.

The largest cause of inefficiency in open cases is exposure to outside air. This is not just an energy efficiency problem but also a consumer comfort problem. The air from open cases makes isles in refrigerated sections three to four degrees Celsius cooler than the rest of the store, causing customers to spend less time in refrigerated isles due to discomfort. The cold air blowing from refrigerated cases also increases inefficiency by putting a heavier load of stores HVAC systems (Islam 2012). This extra load on HVAC systems is typically found in colder climates and during winters when heating is required. In warmer climates, the spillage of air from open cabinets aids in keeping the store cool. It is less efficient than if the area were cooled using the HVAC system and adds humidity to the stores atmosphere putting a heavier load on locations with dehumidifying HVAC systems (Islam 2012).

Outside air infiltration also causes temperature to not be uniform within cases. In open cases, the top most front shelf has been found to be the warmest position in an open case, and is commonly above the five degrees Celsius (41 degrees Fahrenheit) standard set by the Food and Drug Administration (De Frias and Kou 2014). To compensate for the top front being too warm, open cases are constantly running to bring down the temperature resulting in the bottom most rear position becoming too cold resulting in freeze damage (Kou, et al. 2014). The installation of night shades and the use of air curtains help minimize temperature variation, but not sufficiently to prevent this temperature variation (De Frias and Kou 2014). Closed refrigerated cases have virtually no temperature variation within the case (Evans 2007). The temperature uniformity of closed cases reduces energy costs, increases food safety, reduces shrink, maintains food quality longer and has potential to reduce labor costs of rotating product (Evans 2007).

Only two studies have examined the effect of doors on sales of refrigerated products. Neither of these studies examined the effect doors have on fresh processed produce sales. A Swedish study found that doors had no effect on sales and increased the product quality and shopping environment for consumers (Lindberg, et al. 2008). This study examined the sales of meat and dairy products in a Swedish supermarket. The sample size in this study was limited to a single store and the sales history was limited to 6 weeks during the winter. Sales comparisons were made using weeks 3 and 6 of the study. A questionnaire filled out by consumers at the store revealed that those surveyed were not “hindered” in their shopping experience by the addition of doors on the display cases (Lindberg, et al. 2008).

The other study also found that closed display cases had no impact on sales (Fricke and Becker 2010a). The authors preformed a before-and-after comparison of two stores, one with refrigerated cases in dairy and beer aisles and the other with cases in the beer isle (Fricke and Becker 2010a). The first store started with open cases in separate sections containing dairy and beer products. Part way through the 5 month sales study, open cases were replaced with closed cases. The second store, during the same time period, began with open cases containing beer that were replaced with new open cases. The second store sold its dairy products in an open case throughout the study and did not receive a new display case. Dairy sales in the second store were used as a control to be compared to sales in the first store. Sales data before and after the installation of the new cases was collected. Before and after comparisons were performed on weekly beer and dairy sales of the two stores sales. The analysis consisted of 13 weeks in the before period and 9 weeks in the after period when new cases were installed. Fricke and Becker (2010a) found that the installation of new cases with doors had no significant impact on dairy and beer sales in these store locations. They did discover that there was significant energy cost savings from the closed cases (Fricke and Becker 2010a). To date, no academic study has attempted to measure the impacts, on sales of fresh processed produce, of introducing doors on refrigerated display cases.

This thesis examines whether sales of fresh processed produce are affected by the installation of closed refrigerated cases. The data set used is bigger than any in the literature (115 stores) and contains weekly sales data for five years. The data allow for removing the effects of seasonality of sales, and permit regional market comparisons rather than a simple before-and-after comparison for two stores. This thesis will contribute to the literature by supplying research needed to create a complete cost-benefit analysis of closed refrigerated cases. The potential energy savings benefits of doors on refrigerated cases has been examined extensively in the literature, and grocery retailers are aware of the immediate installation costs of doors on refrigerated display cases. This thesis estimates the effects of doors on the weekly sales of fresh processed produce and whether this effect combined with the energy savings of doors is enough to warrant the investment.

Chapter 3. Data and Descriptive Statistics

The retailer data for each store location for this research were collected from Information Resources, Inc. (IRI) through the Economic Research Service (ERS) as well as from TDLinx (ACNielsen) for the years 2008 to 2012. The data are from a single supermarket chain. For reasons stipulated in the contracts with ERS, IRI and ACNielsen do not allow the name or location of the chain to be made public. The total number of stores in the data is 127. Twelve of these locations were removed from the data because they did not have sales for the entire five-year time period, suggesting that the store was either shut down, or newly opened during the time of this study. The remaining 115 store locations have weekly sales for the entire time period. Locations, store attributes and weekly sales data of fresh processed produce were included for this grocery chain. Location data included the store's address, zip code, county name, city name and state. The store's address was converted into latitude and longitudinal values to be used in distance calculations.

Store attributes consist of store size in square feet, number of full-time employees, average weekly total store sales in dollars, annual total store sales in dollars, number of checkouts, presence of a pharmacy and whether or not the location sells gasoline, beer, wine, and/or liquor. These attributes obtained from TDLinx are annual store-level observations. Data obtained from the IRI contained weekly sales and number of units sold at the Universal Product Code (UPC) level for fresh processed produce. This included total weekly dollar and unit sales for each individual fresh processed produce product available at each store location. Individual UPC sales in dollars were summed yielding weekly total sales at each store. Examples of fresh processed produce includes bagged salads, mixed bagged salads, mixed salad kits, organic bagged salads and coleslaw products. For simplicity, all these different types of salads will be referred to as packaged salads, even though other types of packages such as bowls, clam shells and others are used.

In this thesis, stores from the same retail chain will be separated into treatment and control groups. A treatment store is defined as a store in which doors were installed on display cases containing packaged salad between 2008 and 2012. Doors can be retrofitted on existing open display cases or new enclosed display cases can be installed. A control store is defined as a store in which there are no doors on display cases containing packaged salad during the same period. A problem with this store-level data is not knowing which stores have doors installed on their packaged salad cases and which stores do not. Efforts were made to contact the retailer but the retailer did not respond to requests to identify locations and dates of installation of enclosed cases. This retailer was chosen because of a press release covering the remodeling of a store that had doors recently installed on packaged salad cases. In the press release, dated October 17, 2011 it is stated the store remodeling work began in May 2011 and was conducted at night so as not to interrupt customers during normal store hours.

Due to the difficulties faced in obtaining this additional information regarding possible retrofits in other stores, a single treatment store with an estimated treatment date of May 2011 as indicated in the press release is used. The precise installation date is unknown but according to the grocer's press release the installation occurred sometime between May and October 2011. In figure 1, the dates for the installation are represented with circles on the potential installation dates. May 1 was chosen as the installation date as it is the earliest that doors would have been installed. Not knowing of other potential treatment stores may cause control stores to be contaminated with potential treatment stores. Currently, there is no way of isolating those locations and removing them from the control sample. The single treatment store might be the only store with doors but, given the data there is no clear answer. Ideally, the data would contain multiple treatment stores and a clean control group, but this is not possible with the current information made publically available. It is speculated that, the lack of response could be due to the potential sensitivity of the data. In an industry as competitive as retail grocery, installation dates and locations could have strategic implications when paired with sales data.

Figure 1 shows total weekly packaged salad sales for the store with doors installed (treatment location) from 2008 to 2012. It should be noted that seasonality is present and that sales fluctuate in a cyclical pattern. The two circles on figure 1 represent the time frame when doors were installed on refrigerated cases containing packaged salads. The sharp spike in sales in October may have been caused by increased store traffic due to a grand re-opening to celebrate the completion of renovations that were taking place between May and October 2011.

To account for the seasonality of packaged salad sales (Thompson and Wilson 1999), weather data was collected from the National Oceanic and Atmospheric Administration (NOAA). The NOAA dataset included all of the active weather stations that were in the grocer's territory during the 2008-2012 time period. Data included daily maximum temperatures and the latitude and longitudinal coordinates of the weather stations. Stations were then matched to store locations using these latitude and longitudinal coordinates. The geo-distance function in SAS matched the stores to weather stations based upon closest distance.

Using the maximum daily temperatures from 2008 to 2012 average weekly temperatures were computed as the average of the temperatures that were available for each week. If there were fewer than seven temperature values available for the week, the average temperature for that week is the average of the available number of daily temperatures. If a missing temperature variable for the weekly average exists, then this is due to the entire weeks' worth of temperature data missing from that station. Through a process of elimination, stations that yielded a high number of missing weekly averages were deleted from the data set. The process was repeated until an acceptable number of missing weeks was reached. There are 63 weeks missing out of 28,340 total week observations and the largest number of weeks missing for any one store is 5 out of 261 weeks. In some cases, there are multiple stores to a single weather station. The shortest distance between a station and a store is 0.27 miles and the furthest distance between a store and its matched station is 13.39 miles. The average distance between a store and its matched weather station is 4.88 miles. Figure 2 shows the weekly average maximum temperature for the treatment store from 2008 to 2012. The seasonality of temperature should be noted, as well as the plateau during the summer months where the average maximum temperature for the week is consistently between 90 and 100 degrees Fahrenheit.

The matching of weather stations to store locations is used to account for seasonality to packaged salad sales. This connection between weather and salad sales was first explored by Thompson and Wilson (1999), who found a positive correlation between the average temperature salad sales. This relationship can be seen in the figure 3 where the weekly maximum temperature is overlaid on weekly packaged salads sales for the treatment store. In figure 4, the correlation between temperature and packaged salad sales appears strong as temperatures increase during June and July, but as temperatures continue to remain high through August, there is a weakening in the correlation as packaged salad sales begin to fall off. The weakening of the relationship of processed salad consumption and temperature during this time period could be caused by the beginning of the school year for many families with children and students. The shift from summer to the start of school may result in the reduction in processed salad consumption as seen in the data. The correlation between temperature and processed salad consumption is examined in figure 4. On the y-axis is weekly packaged salad sales and on the x-axis is the average weekly maximum temperature. When plotted for the treatment location and a trend line drawn the positive correlation between temperature and packaged salad sales is evident. When calculated, the correlation between temperature and packaged salad sales is .403.

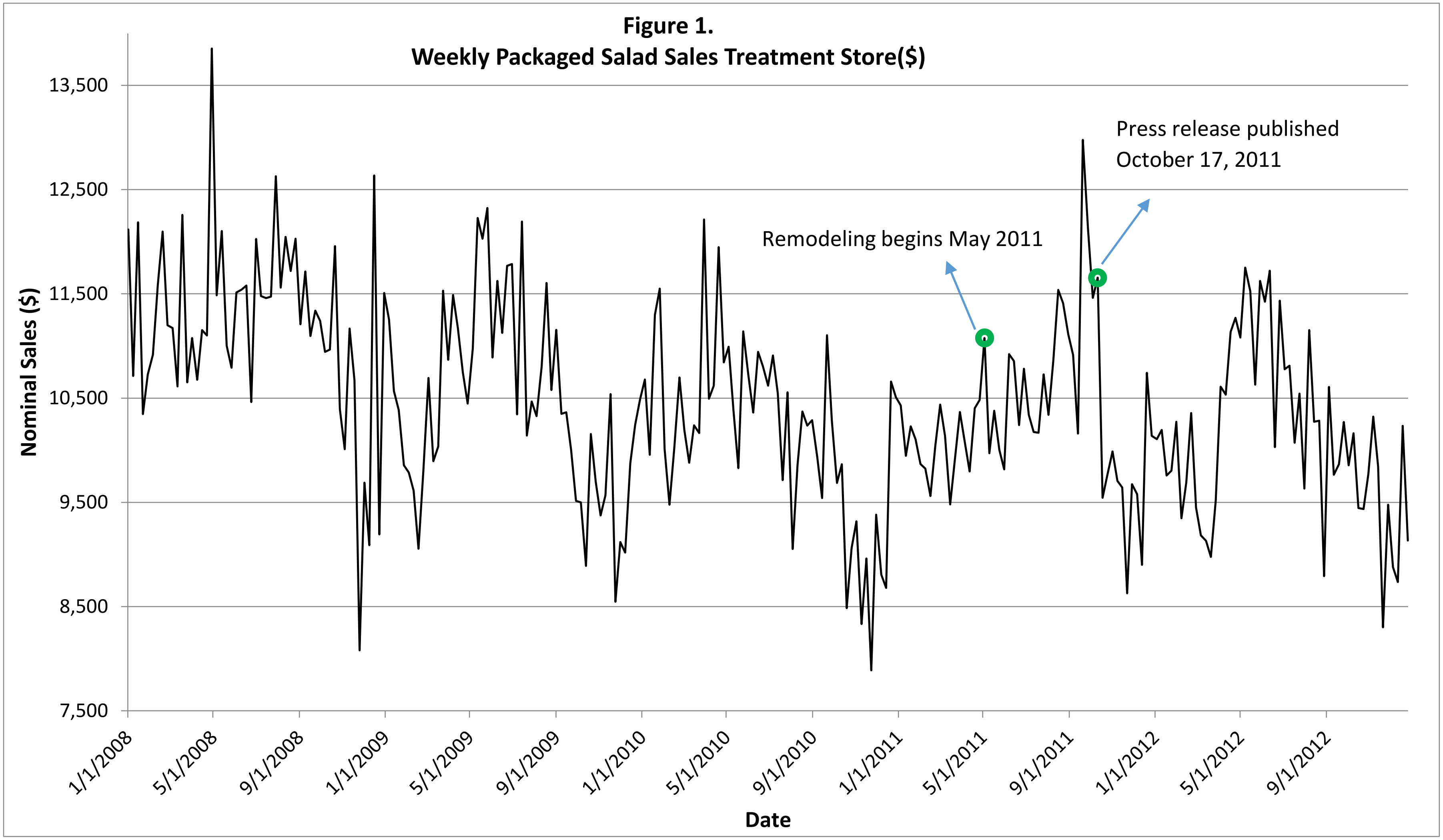


Figure 1.
Weekly Packaged Salad Sales Treatment Store($)
Press release published October 17, 2011
Remodeling begins May 2011
Nominal Sales ($)
13,500
12,500
11,500
10,500
9,500
8,500
7,500
1/1/2008
5/1/2008
9/1/2008
1/1/2009
5/1/2009
9/1/2009
1/1/2010
5/1/2010
9/1/2010
1/1/2011
5/1/2011
9/1/2011
1/1/2012
5/1/2012
9/1/2012
Date

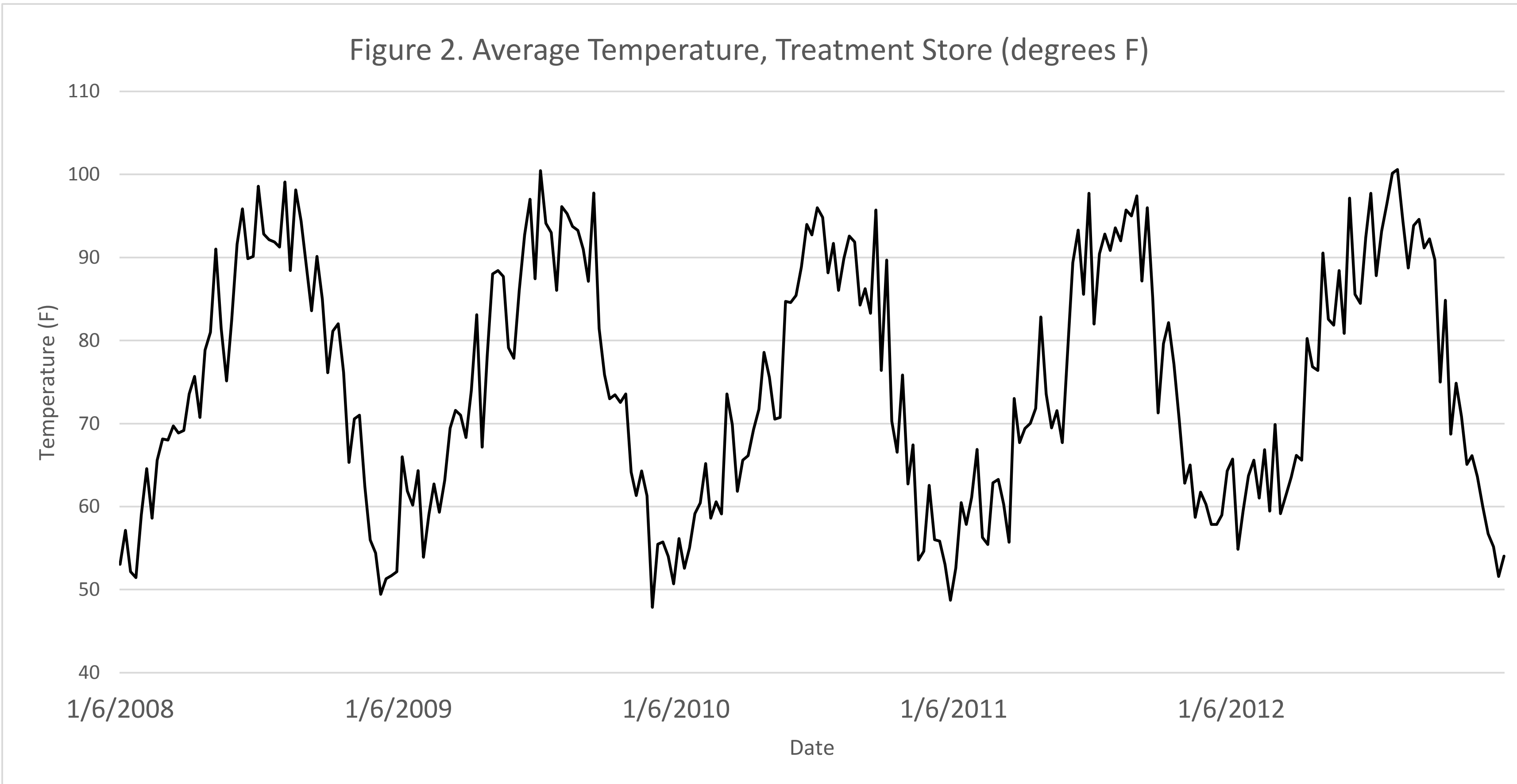


Figure 2. Average Temperature, Treatment Store (degrees F)
110
100
90
80
70
60
50
40
Temperature (F)
1/6/2008
1/6/2009
1/6/2010
1/6/2011
1/6/2012
Date

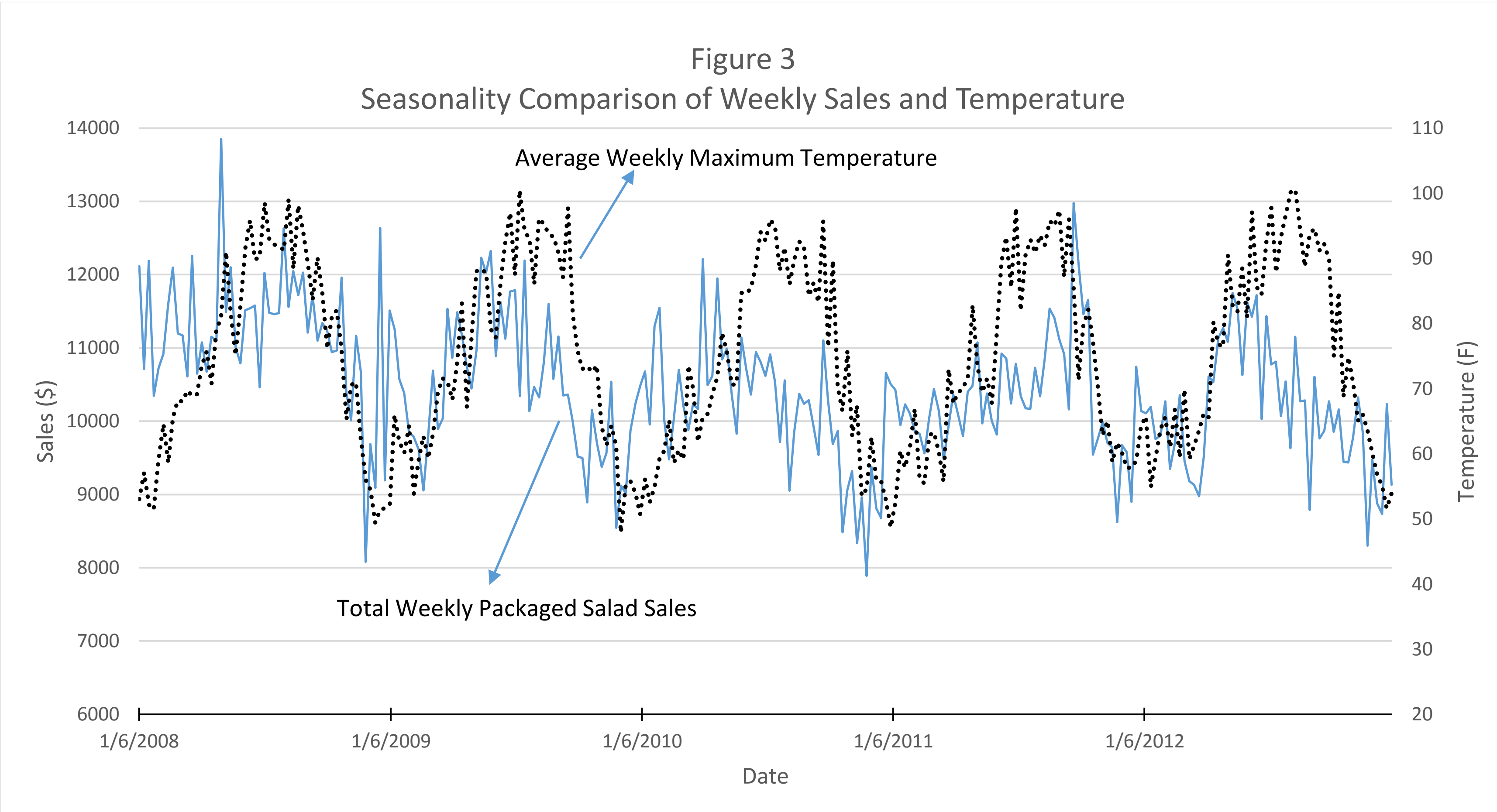


Figure 3
Seasonality Comparison of Weekly Sales and Temperature
Average Weekly Maximum Temperature
Total Weekly Packaged Salad Sales
Sales ($)
Temperature (F)
Date
14000
13000
12000
11000
10000
9000
8000
7000
6000
110
100
90
80
70
60
50
40
30
20
1/6/2008
1/6/2009
1/6/2010
1/6/2011
1/6/2012

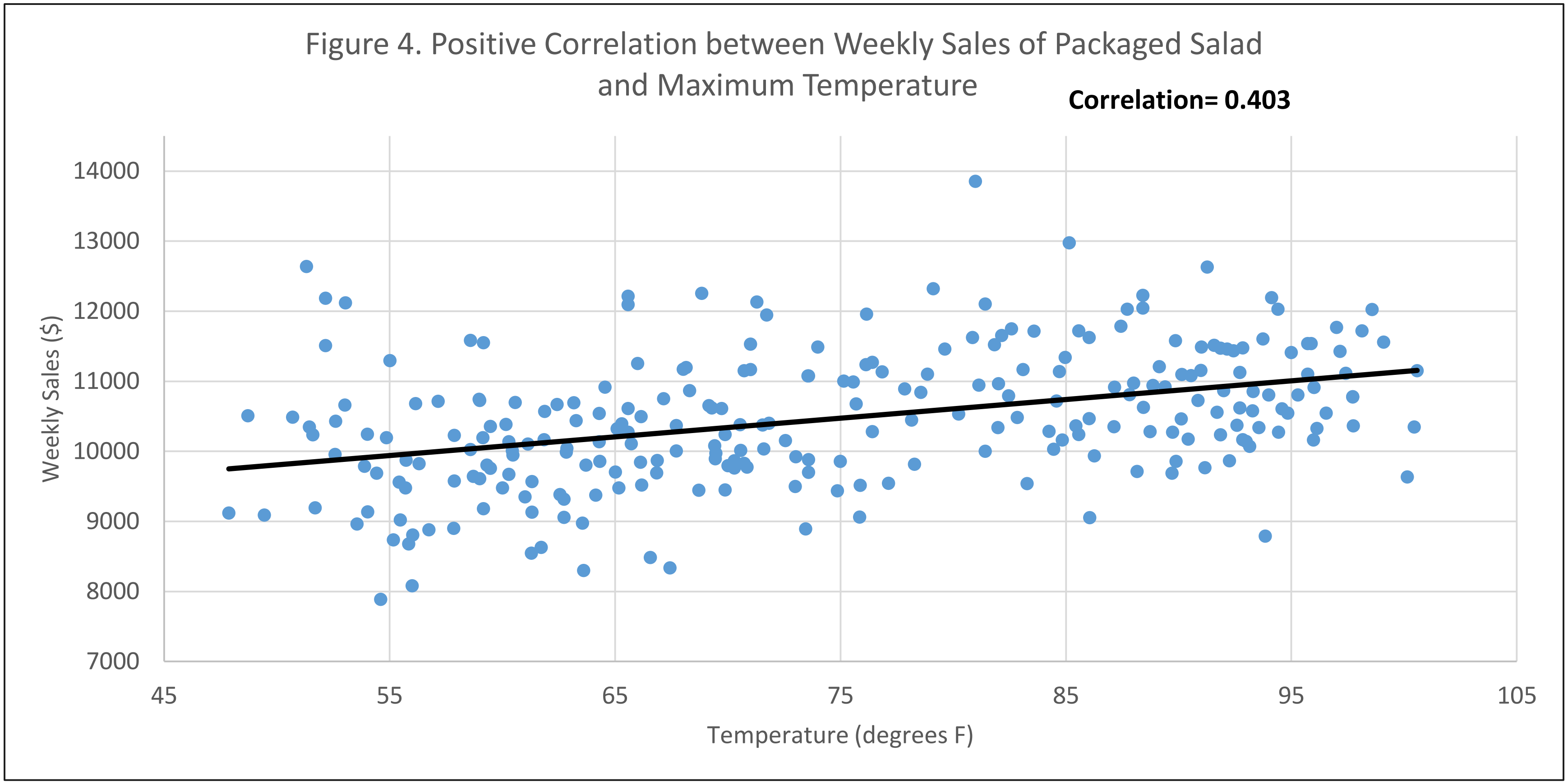


Figure 4. Positive Correlation between Weekly Sales of Packaged Salad and Maximum Temperature
Correlation= 0.403
Weekly Sales ($)
14000
13000
12000
11000
10000
9000
8000
7000
45
55
65
75
85
95
105
Temperature (degrees F)

Income data were collected in order to match the treatment store with control stores situated in areas with similar purchasing power. Data collected from the Internal Revenue Service (IRS) contained the annual gross income for residents living within the zip code of store locations. The average income for each year and zip code in the sample was calculated by dividing the total gross income of the area by the total number of tax returns filed.

$$Average\ Income_{it} = \frac{Annual\ Gross\ Income_{it}}{Total\ Number\ of\ Returns_{it}}\ Where\ i = zipcode, t = year$$

The average income of the residents living in close proximity to the store location controls for any effects income may have on a particular store. Income of the surrounding area may serve as a proxy for the composition of the surrounding population which may have an effect on the atmosphere of the store. Effects of the atmosphere could include the perception of store cleanliness, feeling of safety, perceived customer service and overall shopping experience.

Physical store attributes, sales, weather and income data will be used to match the treatment store with control stores with similar characteristics as a part of the analysis. A clustering method is used to match similar stores; this technique is introduced in the Methods chapter and further described in the appendix. Table 1 contains all variables used in the analysis and their definitions.

Table 1. Variable Definitions				
Variable Name	**Definition**	**Units of Measure**	**Temporal Variability**	**Data Source**
Store Size	Floor space of the store	Square Feet	Annual	TDLinx
Average Total Weekly Sales	Average weekly sales of all products sold in the store	Nominal Dollars	Annual	TDLinx
Annual Total Store Sales	Range of annual total sales of all products in the store	Categorical[1]	Annual	TDLinx
Average Weekly Packaged Salad Sales	Average weekly sales of all packaged salad	Nominal Dollars	Weekly	IRI
Average Maximum Temperature	Average of daily maximum temperatures in each week	Degrees Fahrenheit	Weekly	NOAA
Number of Employees	Number of full-time employees and full-time employee equivalents.[2]	Count	Annual	TDLinx
Number of Checkouts	Number of checkout registers in the store	Count	Annual	TDLinx
Zip Code Average Income	Average income for all residents within the zip code of the store location	Nominal Dollars	Annual	IRS

[1] Categories give ranges of total annual store sales, e.g. 1= $1 to $500,000. Table 14 in the appendix has a complete breakdown of all categories.
[2] 1 part-time employees= ½ full time employee

Table 2 contains the descriptive statistics of store attributes. In the data there is a large difference between the largest and smallest stores. Based on total square footage the largest store is almost three times the size of the smallest. Size differences can also be found in the number of employees. The largest store employing 122 workers while the smallest only employing 18. The mayor differences are not limited to store size or the number of employees. The average income for the zip code in which the store is located is where the largest difference can be found. The wealthiest zip code that contains a store has an average income of 885,803 dollars per year, while the least affluent zip code has an average income of 27,369 dollars per year.

Table 2: Descriptive Statistics

Variable	**Mean**	**Median**	**Max**	**Min**
Store Size (Sq. Ft.)	42,351.3	45,000	59,000	19,000
Number of Employees	68	68	122	18
Number of Checkouts	10	10	19	6
Average Weekly Total Store Sales ($)	449,812	450,000	850,000	150,000
Annual Total Store Sales ($)*	20-25 million	20-25 million	40-45 million	6-8 million
Weekly Packaged Salad Sales ($)	5,842	4,709	31,396	2
Weekly Average Maximum Temperature (F)	71.7	70.3	107.5	11.9
Zip Code Average Income ($)	69,126	59,807	885,803	27,369
Distance to Weather Station (miles)	4.9	4	13.4	0.3

*Annual Total Store Sales is a categorical output see appendix

Figures 5-7 contain the distributions of store size, number of checkouts and number of employees respectively. The distributions of store size, number of employees and number of checkouts are relatively normal. The distribution of store employees is slightly skewed to the left, store size and number of checkouts are slightly skewed to the right. Figures 8 through 11 and their corresponding tables, show the yearly averages of income and sales between 2008 and 2012.

Figure 5. Store Size Distribution, 2010

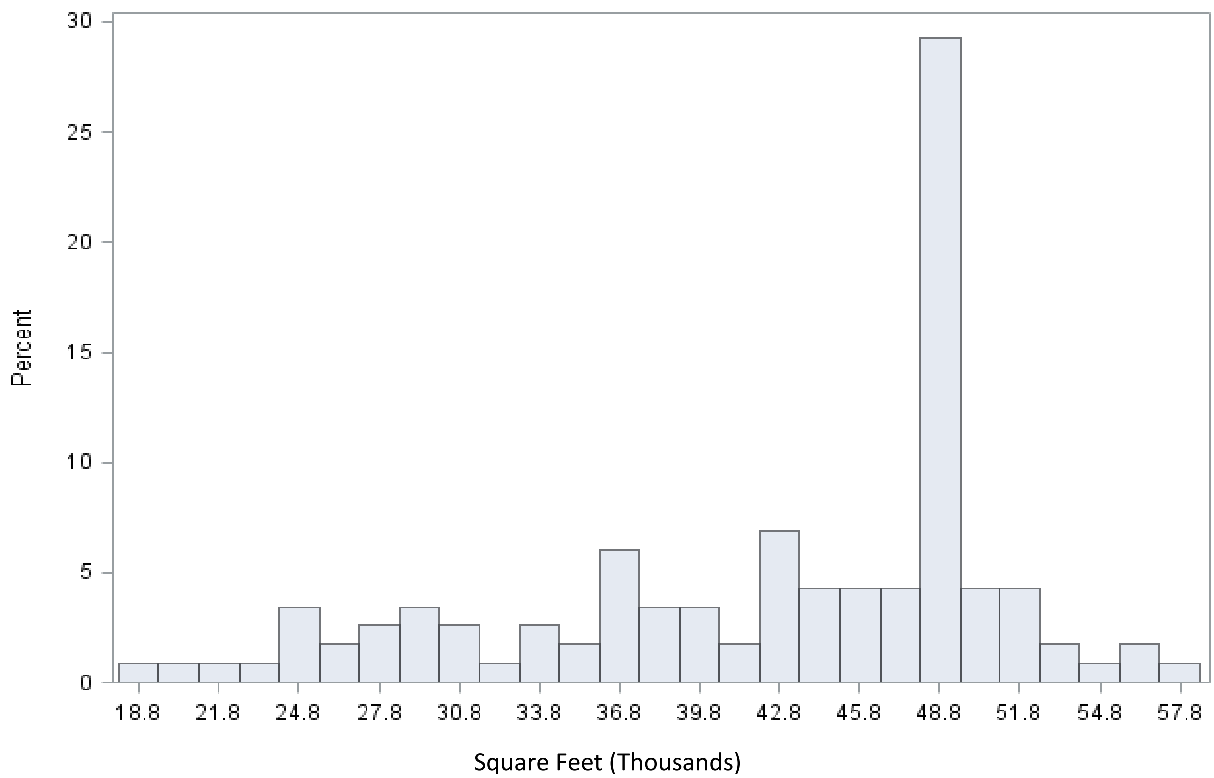


Figure 6. Number of Checkouts Distribution, 2010

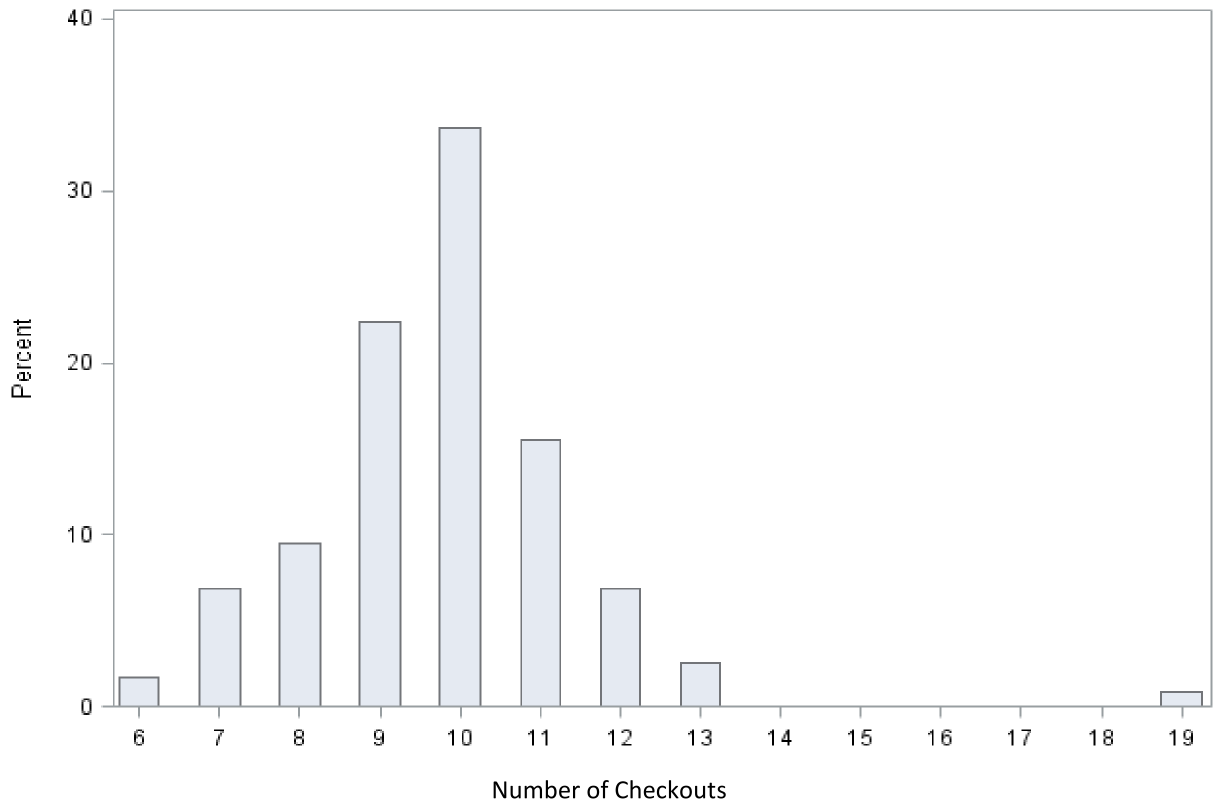
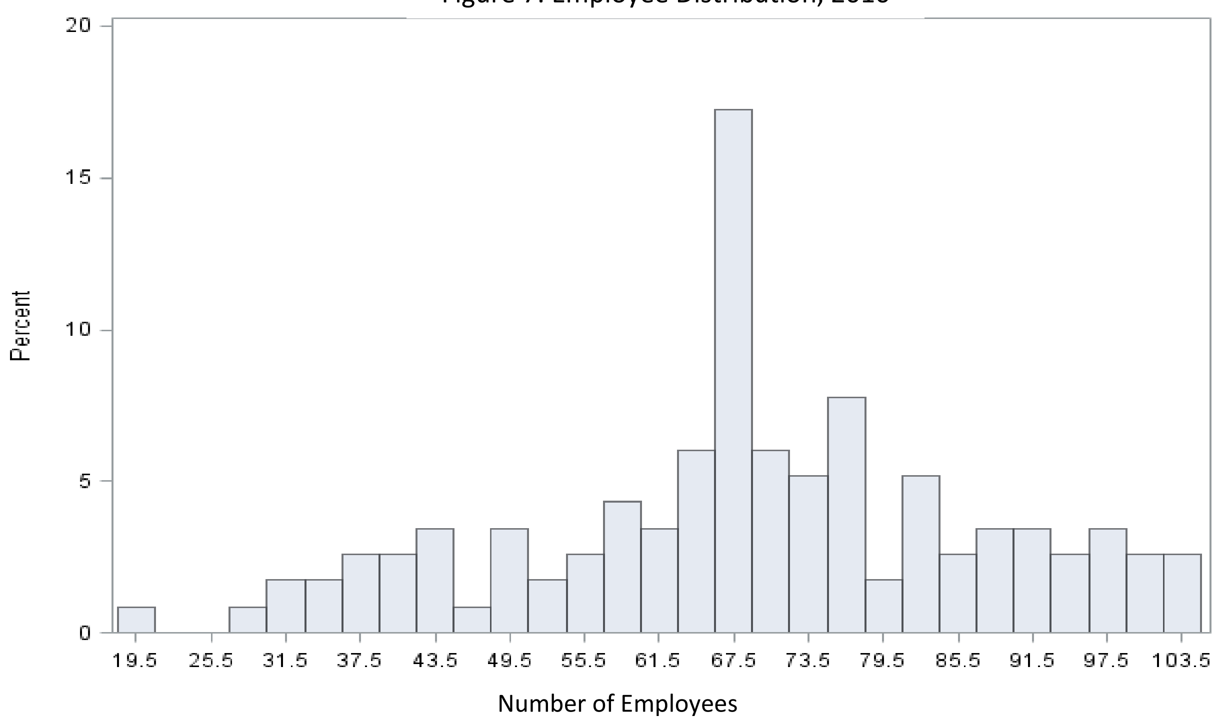


Figure 7. Employee Distribution, 2010



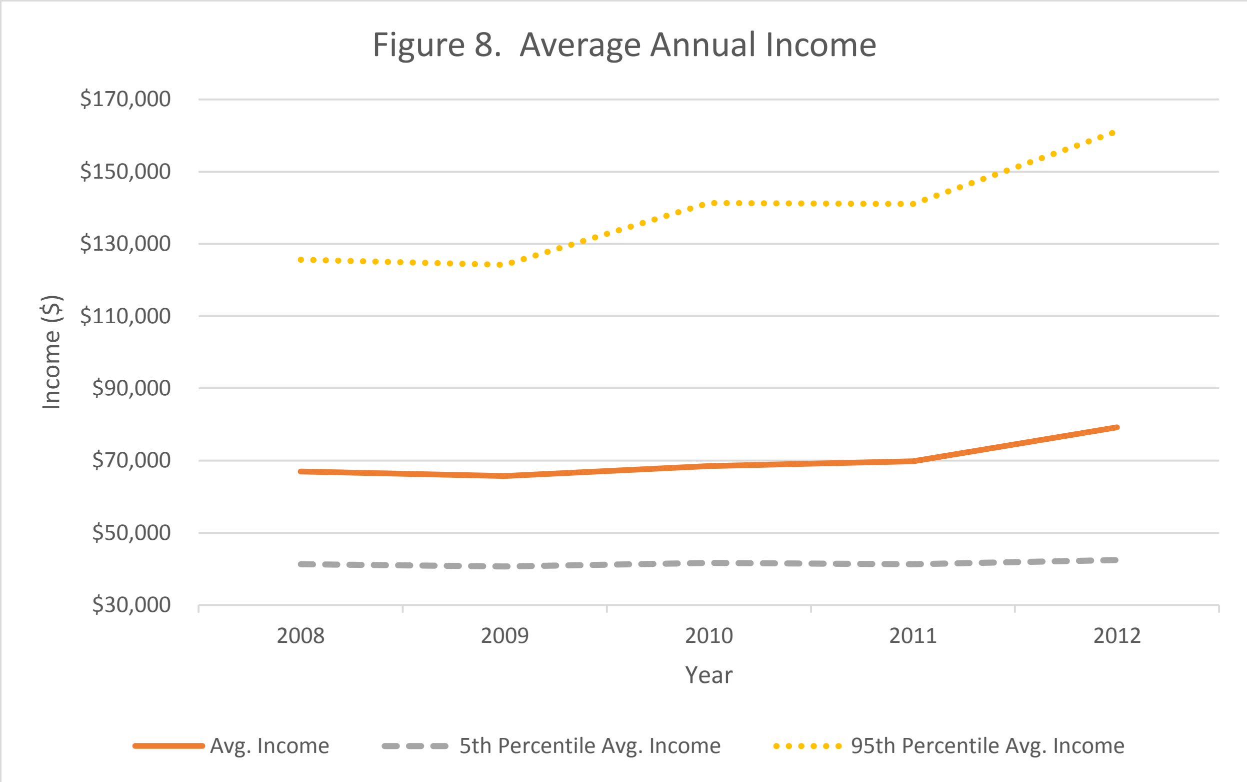


Table 3. Average Annual Income

Year	Avg. Income	5th Percentile Avg. Income	95th Percentile Avg. Income
2008	$ 67,003	$ 41,338	$ 125,592
2009	$ 65,715	$ 40,718	$ 124,193
2010	$ 68,456	$ 41,668	$ 141,326
2011	$ 69,787	$ 41,307	$ 141,092
2012	$ 79,202	$ 42,452	$ 161,177

[3]

[3] All income figures are nominal. Sample size is 115 stores over 5 years (575). Median income cannot be calculated at zip code level from IRS data due to data collection methods.

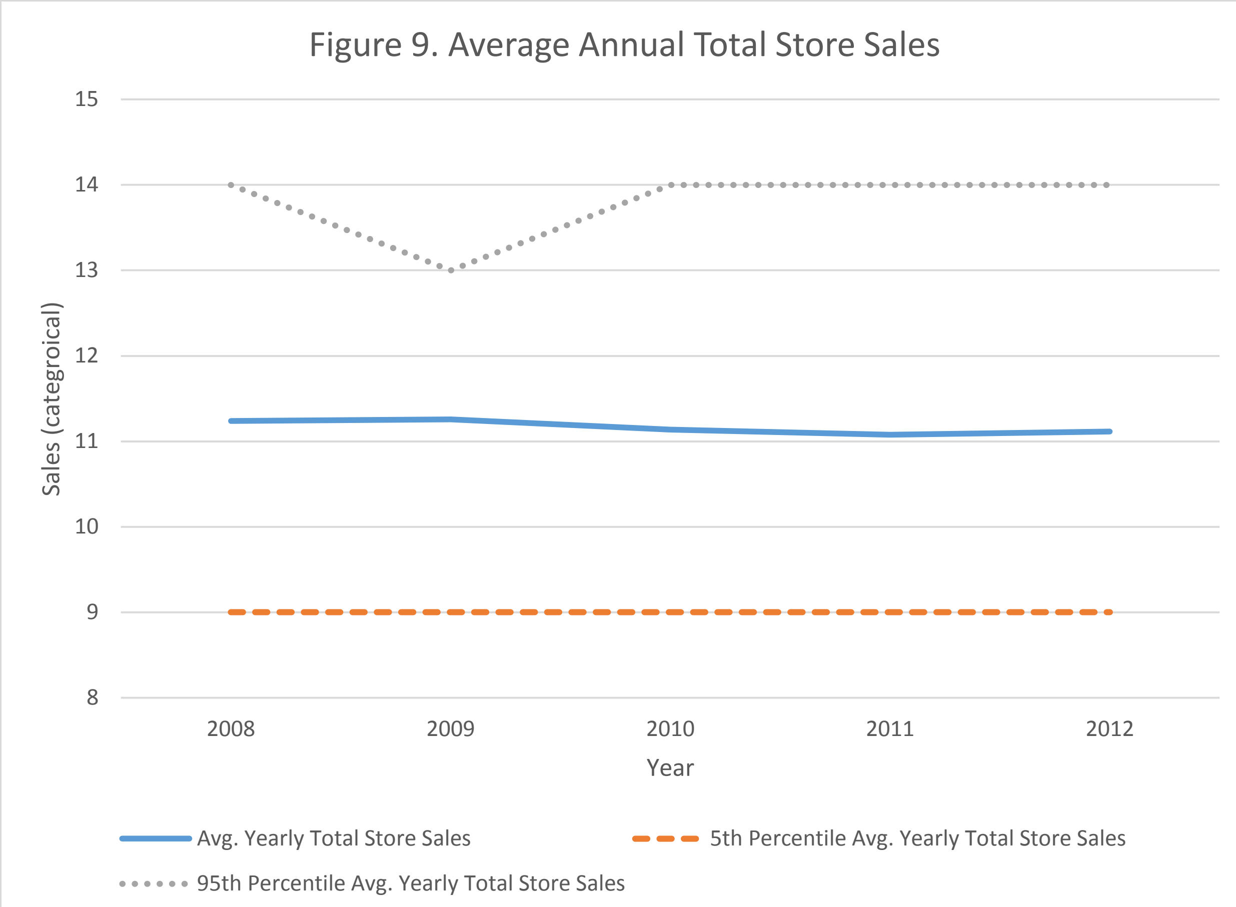


Table 4. Average Annual Total Store Sales[4]

Year	Avg. Yearly Total Sales	5th Percentile Avg. Yearly Total Sales	95th Percentile Avg. Yearly Total Sales
2008	11.2	9	14
2009	11.3	9	13
2010	11.1	9	14
2011	11.1	9	14
2012	11.1	9	14

[4] Note: 9= $12-16 Million, 11=$20-25 Million, and 14=35-40 Million. See Table 14 for complete categorical breakdown

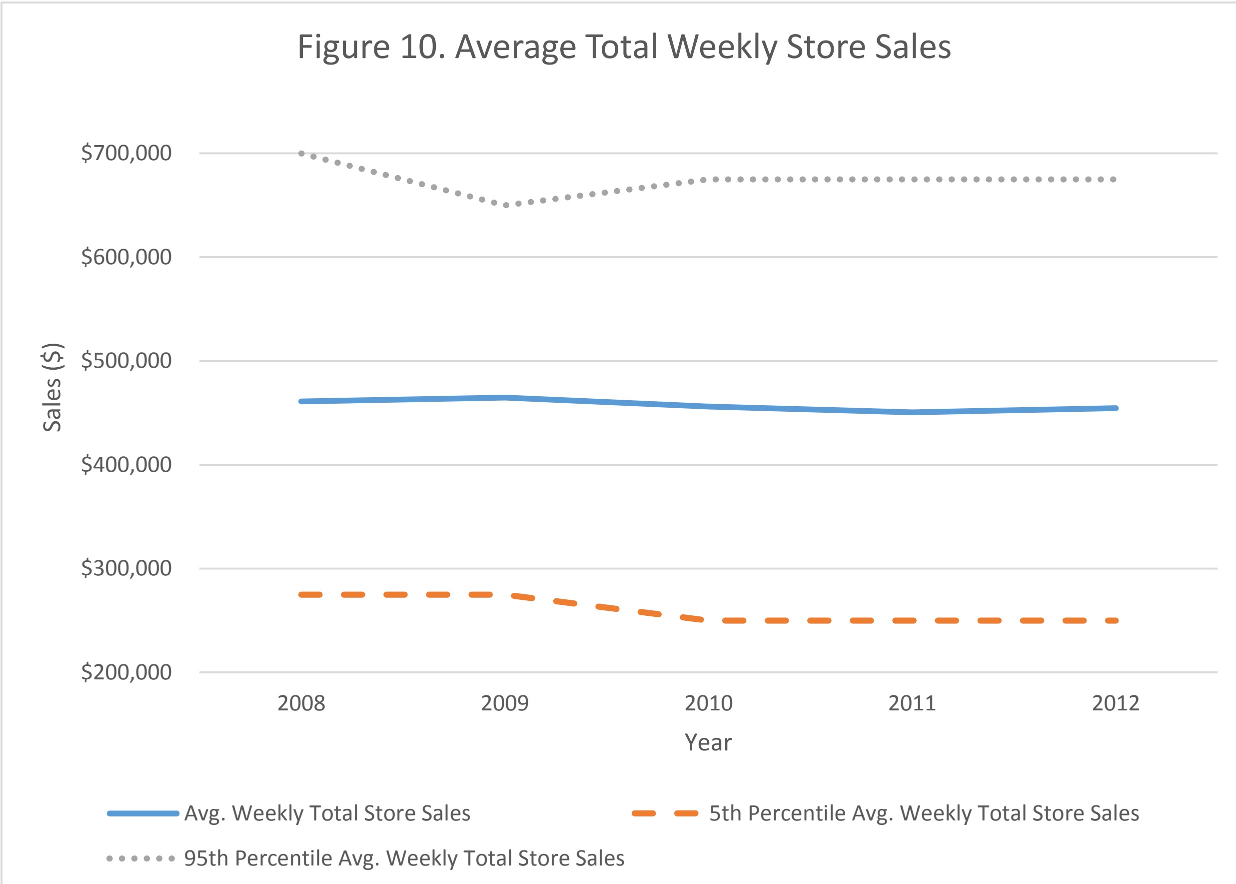


Table 5. Average Weekly Total Store Sales

Year	Avg. Weekly Total Sales	5th Percentile Avg. Weekly Total Sales	95th Percentile Avg. Weekly Total Sales
2008	$ 461,030	$ 275,000	$ 700,000
2009	$ 464,655	$ 275,000	$ 650,000
2010	$ 456,034	$ 250,000	$ 675,000
2011	$ 450,647	$ 250,000	$ 675,000
2012	$ 454,561	$ 250,000	$ 675,000

Note: all sales nominal

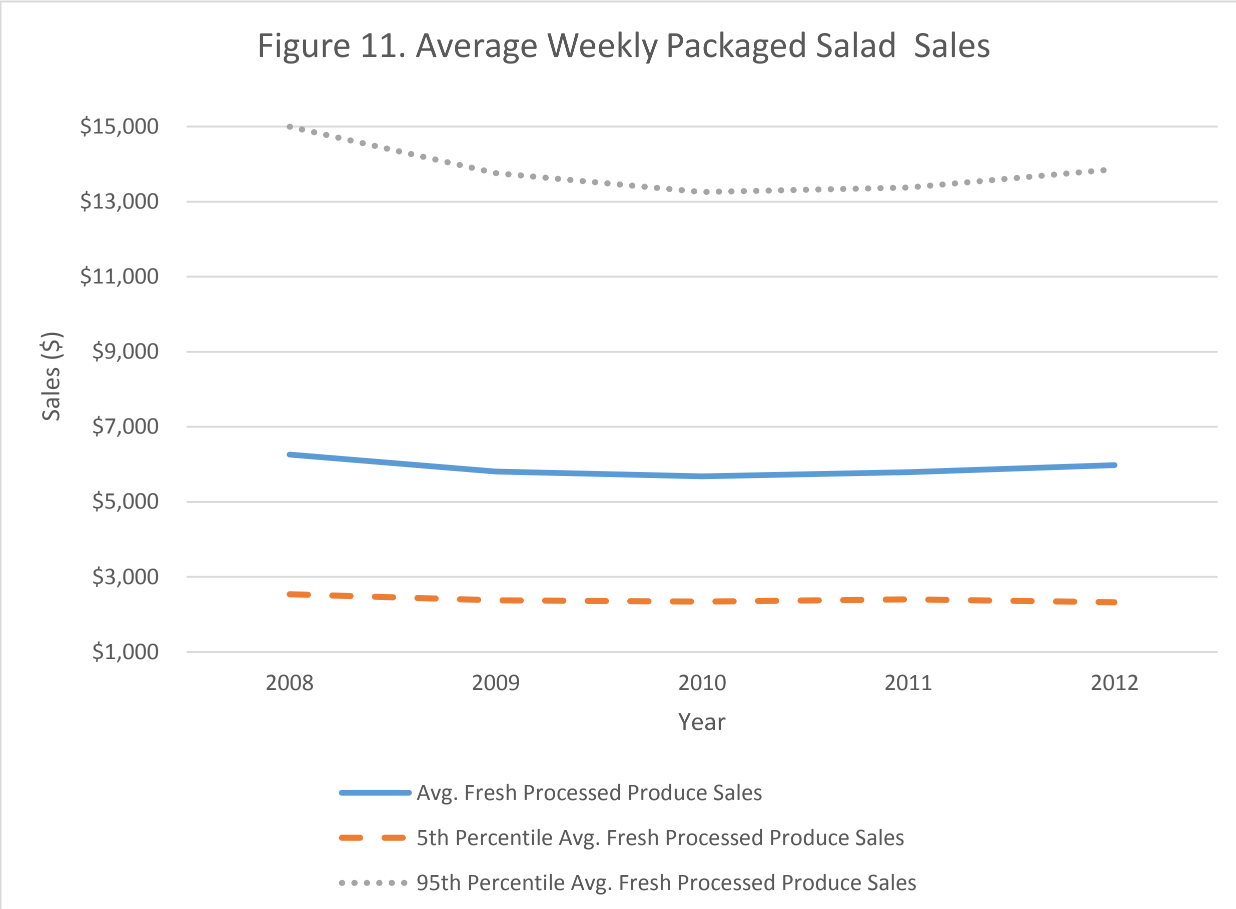


Table 6. Average Weekly Packaged Salad Sales

Year	Avg. Packaged Salad Sales	5th Percentile Avg. Packaged Salad Sales	95th Percentile Avg. Packaged Salad Sales
2008	$ 6,265	$ 2,536	$ 14,995
2009	$ 5,808	$ 2,373	$ 13,758
2010	$ 5,684	$ 2,343	$ 13,261
2011	$ 5,795	$ 2,404	$ 13,376
2012	$ 5,980	$ 2,330	$ 13,860

Note: all sales nominal

Average income and average sales have relatively little variation over the sample period. The 95th percentiles of the sales data show the most fluctuation, especially between 2008 and 2009. Total sales, annual and average weekly, as well as, average weekly packaged salad sales drop from 2008 to 2009 and never recover to 2008 levels. Not so for income. Average income recovered back up to 2008 levels by 2010 and continued to increase in nominal terms. The 95th percentile of income stands out as it has the largest increase over the sample period, while the sample average and 5th percentiles change little. Trends in packaged salad sales closely resemble total weekly sales, and average annual total store sales.

The descriptive statistics show the variation in store size, weather and income. The intervals of variation are weekly and yearly. Store size, number of checkouts, number of employees, distance to weather stations, average annual zip code income and annual total sales are all yearly measures. Average weekly total store sales, weekly fresh produce sales, and weekly average maximum temperature are calculated weekly and vary through time in seasonal cycles. The most surprising result is the average income for a zip code. The maximum value for a single zip code income is almost nine hundred thousand dollars, while the mean, median and minimum values range between $60,000 and $20,000. The large maximum income is potentially influenced by a small contingent of wealthy individuals living in a cluster (e.g. a gated community). Having a cluster of high income individuals is not concerning as these wealthy individuals appear to have little influence on the overall average income of the zip codes where the stores are located. The difference in overall store magnitude can be seen in the store size, number of employee, number of checkouts, and average total weekly/annual sales.

Chapter 4. Methods

To calculate the influences of displays with doors on packaged salad sales, a difference-in-difference model is used. Difference-in-difference models use treatment and control groups and determine the difference between the treatment and control group during two periods, before and after implementation of the treatment. The difference-in-difference method is stronger than a simple before and after comparison because it controls for changes in sales in the control group. The difference-in-difference model used in this study includes a dummy for time (before/after), a treatment dummy whether the store received the treatment and an interaction term between the time and treatment dummies. Weekly temperatures are included as control variables for seasonality. If the interaction term in the after period is statistically significant then treatment has an effect. Including average temperatures controls for seasonality that may differ depending on local weather conditions. The dependent variable is the total weekly salad sales.

$$\boldsymbol{TotalWeekSales_{it} = \beta_1 + \beta_2 Dafter_t + \beta_3 Dtreatment_i + \beta_4 (Dtreatmet * Dafter)_{it} + \gamma Control\,Variables + \epsilon_{it}}$$

$$\boldsymbol{i = 1, \dots, 115}$$

$$\boldsymbol{t = 1, \dots, 261}$$

In this model, $TotalWeekSales_{it}$ denotes the dollar value of sales of packaged salads in store *i* in week *t*. Dafter is a dummy variable that denotes time in weeks, it takes a value of 0 before the installation of doors and a value of 1 after the installation of doors. Dtreatment is a dummy variable that denotes whether the store is a treatment or control location. Dtreatment takes a value of 0 for control stores and a value of 1 for treatment locations. A treated store is defined as a store where packaged salad is displayed in a refrigerated case with doors. A control store is a store that uses open display cases to display packaged salad. Before the model is estimated the single treatment store is matched with one or more control stores. If the treatment store is matched with all control stores then i=115, but if only one store is matched then i=2. The model is limited to a single treatment store due to information constraints. Having limited treatment locations is helpful as it simplifies the "after" dummy. The "after" dummy is defined by the date that doors were installed in a location, taking a value of 0 before the installation of doors and a value of 1 after. There is a single treatment location and a single treatment date resulting in a single "after" dummy. If multiple treatment locations existed, an "after" dummy for each location would be necessary, owing to differences in the installation dates across stores. Control variables will include weekly average maximum temperature and yearly average income for the store's zip code. The model will be estimated using both controls individually and combined. Using weather controls will help smooth and sharpen estimations as weather is correlated with packaged salad sales. The use of average income is to test the robustness of the mode, but as the model is not a demand equation, results using income as a control variable will not be heavily considered.

The difference-in-difference model will be used to estimate potential changes in weekly packaged salad sales. The model is not intended to serve as a demand equation. The goal is not to estimate the quantity of packaged salad sold but simply weekly sales. Average income by zip code is not included as a control in the model because the model is not a demand equation. Income is used in matching the treatment to control stores. Average annual maximum temperature serves as a geographic matching variable, matching stores with similar average temperatures. To check robustness, the model will be estimated without the weather seasonality control, with income as a control and with weather and income as control variables.

Table 7. Yearly Descriptive Statistics of Treatment Store									
Store	Year	Average Annual Maximum Temperature (F)	Zip Code Average Income ($)	Annual Total Sales (Catagorical)[5]	Number of Checkouts	Average Weekly Total Sales (Thousands of $)	Size of Store (Thousands of Sq.ft.)	Number Of Employees	Weekly Average Packaged Salad Sales ($)
T	2008	75	69,002.9	12	11	575	43	99	11,206.4
T	2009	75	63,866.9	12	11	525	43	99	10,496.6
T	2010	73	65,310.8	12	11	500	43	99	10,155.4
T	2011	73	65,347.4	11	11	475	43	99	10,342.5
T	2012	76	73,480.2	11	11	450	43	99	10,134.6

Had the treatment and control groups been randomly assigned, there would be no need for the use of controls in the model. Due to non-random assignment of treatment and control groups, the model is estimated using different matched control groups. Matching treatment and control groups on observable variables produces matches as if they were randomly assigned. Matching is used to gather control stores that are similar to the treatment store. Using matched control stores compares the treatment store to control stores that are similar in all observable characteristics. Matched control locations are used as an estimation of what sales would have been, had the treatment store not received the treatment. Multiple control groups will show the relationship between the effects of the treatment when compared to different control situations as a check of robustness. Having different sized control groups will play a role in the statistical significance of the parameter estimates. Typically, larger samples produce smaller standard errors it is expected that statistical significance will change with sample size. The use of both

[5] 11= $20,000,001 to $25,000,000
12= $25,000,001 to $30,000,000
See table 14 in Appendix for full category table

large and small control groups will aid in determining the true significance of parameter estimates.

Using a clustering method as outlined in the appendix, seven store locations were identified as the closest matches to the treatment location based upon yearly average maximum temperature, average income for the surrounding zip code, retailer name[6], store size, number of checkouts, total annual volume in sales, average weekly sales of all products and number of employees. Variables were chosen as they demonstrate store attributes and the observable environment of the location. Store locations were matched using the 2010 annual observations for these variables because it is the middle year of this study.

The model was estimated using the top seven and the top two best matched stores as the control groups. Using closely matched control stores demonstrates how an almost perfect match in observable characteristics affects the results. The control stores used as the top seven best matches can be found below in Table 8. These locations were chosen through the use of a clustering method. The clustering was performed using annual average temperature, annual average income, annual total store sales, number of checkouts, average weekly store sales, store size, number of employees and store name. The clustering method then grouped the stores that were the most similar to one another based on calculated density scores. In this case, there were seven other stores that were grouped into the same cluster as the treatment store. Stores C1 and C2 are the closest matches to the treatment store. This was achieved through reducing the radius of the cluster containing the treatment store until only two stores, then a single store remained. The treatment store and store C1 have higher average weekly packaged salad sales than the other

[6] The anonymous grocer has stores with several different names. Each name is a distinct attribute of an individual store. Different names represent different types of stores presumably targeting different market segments.

stores, except C5, along with a relatively high number of employees and total annual sales. Estimations of the model using all seven best matched control stores and the best matched control stores C1 and C2[7], will show how stores similar in all observable characteristics compare in weekly packaged salad sales in the after period. The best match estimation is a true "apple-to-apple" comparison of similar stores, one with doors and one without.

Table 8. Yearly Descriptive Statistics of Best Matched Control Stores, 2010

Store	Year	Average Weekly Packaged Salad Sales ($)	Annual Average Maximum Temperature (F)	Annual Average Income ($)	Total Annual Store Sales (Categroical)	Number of Checkouts	Average Weekly Total Sales (Thousands of $)	Size of Store (Thousands of Sq.ft.)	Number Of Employees	Name of Store
T	2010	10,155	73.2	65,311	12	11	500	43	99	2
C1	2010	11,012	73.2	65,311	12	9	525	38	61	1
C2	2010	5,862	74.1	65,381	12	11	525	56	40	1
C3	2010	6,297	72.6	65,888	13	10	650	52	42	1
C4	2010	8,456	72.8	66,500	11	9	425	40	72	2
C5	2010	14,658	72.8	66,500	13	10	600	35	82	2
C6	2010	5,226	72.3	66,605	11	7	475	44	48	3
C7	2010	5,212	66.2	66,368	12	10	500	47	66	3

[8]

Stores were also matched based on geographic proximity, in order to assess how the treatment might have affected the consumers' choice in which location they shopped. Three geographic matches were conducted. For comparison, the model was estimated using the entire control data set (114 locations). Estimations were then preformed using controls that included all stores in the same county as the treatment store (29 locations) and all stores that were in the same suburb as the treatment store (1 location). It is assumed that consumers have the ability to choose between stores they shop whether they consciously travel to a specific location or have options on their daily commute. By matching on geographic proximity, consumer preferences might be captured. Depending how consumers felt about the construction and renovations that occurred at

[7] Store C1 was the best matched store but it is also the only store located in the same suburb as the treatment store. For analysis purposes store C2 will be used in the category "Best Matched"
[8] Store names are coded 1,2 and 3 for the 3 store names within the grocery chain

the treatment location, they may choose to shop at a neighboring location. Table 9 compares the descriptive statistics of stores in the same county as the treatment store with the treatment store. Store C1[9] is the only store located in the same suburb as the treatment store and is used as a single control in the suburb estimation of the model. The single suburb store, just three and a half miles away, will yield insight into consumer shopping preferences when given a choice between treatment store and a control store within close proximity to one another.

[9] Store C1 is the best overall match based on both store and geographic characteristics. Store C1 has a different retailer name than the treatment store.

Table 9. Geographic Location Matching										
Store	Year	Average Weekly Packaged Salad Sales ($)	Average Annual Maximum Temperature (F)	Average Annual Income ($)	Total Annual Store Sales	Number of Checkouts	Average Weekly Total Sales (Thousands of $)	Size of Store (Thousands of Sq.ft.)	Number Of Employees	Name of Store
T	2010	10,155	73.2	65,311	12	11	500	43	99	2
C1	2010	11,012	73.2	65,311	12	9	525	38	61	1
C3	2010	6,297	72.6	65,888	13	10	650	52	42	1
C4	2010	8,456	72.8	66,500	11	9	425	40	72	2
C5	2010	14,658	72.8	66,500	13	10	600	35	82	2
C8	2010	3,202	73.2	42,859	10	9	375	45	66	1
C9	2010	4,633	74.2	48,837	12	11	575	49	68	1
C10	2010	2,186	73.2	31,399	10	9	325	49	63	1
C11	2010	7,483	72.8	43,888	14	13	750	33	83	1
C12	2010	11,077	72.5	86,031	11	10	450	48	99	1
C13	2010	17,156	70.9	86,031	12	13	500	52	57	1
C14	2010	5,664	73.2	45,834	13	12	625	47	66	1
C15	2010	9,165	73.2	70,704	12	12	550	37	87	1
C16	2010	4,479	73.2	53,318	11	10	450	44	67	1
C17	2010	5,998	73.2	46,847	14	12	675	53	66	1
C18	2010	14,130	73.2	70,704	14	13	750	58	98	1
C19	2010	2,648	72.8	31,991	11	10	400	43	76	1
C20	2010	11,231	72.8	59,996	12	12	575	51	57	1
C21	2010	9,698	73.0	62,266	9	9	275	49	122	1
C22	2010	2,353	73.2	34,056	9	7	250	26	68	2
C23	2010	5,557	73.2	108,702	11	10	475	29	68	2
C24	2010	12,714	72.8	62,266	9	7	300	45	82	2
C25	2010	4,317	73.2	44,087	12	10	500	33	83	2
C26	2010	9,156	72.8	59,996	11	10	400	36	104	2
C27	2010	4,096	73.2	48,515	11	10	400	36	70	2
C28	2010	13,363	73.2	53,318	13	10	650	42	83	2
C29	2010	20,007	73.2	62,266	11	10	475	45	64	2
C30	2010	11,699	72.5	86,031	10	11	350	46	66	2
C31	2010	20,428	72.8	62,266	12	7	500	38	69	2
C32	2010	4,632	72.6	49,339	11	12	425	55	71	2

[10]

[10] *stores C1, C3, C4 and C5 are among the best matched locations

Chapter 5. Econometric Results

Tables 10-13 contain the results of the parameter estimations. Table 10 contains estimations without control variables, Table 11 contains estimations with weekly average maximum temperature as a control variable, Table 12 contains estimations with annual income as a control variable and Table 13 contains estimations with both weekly average maximum temperature and annual income as control variables. It should be noted that two different methods for matching the treatment store to control stores were used, statistical clustering and geographic. When geographic matching was used, there were no cases in which the effect on sales was significant.

The focus of the results is the coefficient on the interaction between the after and the treatment dummies "t_times_a." The size and significance of this coefficient is key in the potential decision-making process for grocers considering investing in door on their refrigerated cases containing packaged salad. The variable "t_times_a" demonstrates how sales changed for the treatment store in the during the post treatment period. The variable "after" shows how weekly sales changed for stores that did not receive treatment during the post-treatment period. If the interaction coefficient is not significant then this would suggest that grocers need not be concerned with the effect on sales, while significant values could lead to more questions.

When the model is estimated using no variable controls, only the Best Matched control group yielded a slightly significant result for treatment in the after period. The estimated result of -244.6 is only about a 4 percent reduction in sales and with a p-value of .08 making it statistically significant at the 90% level. The estimation using the two best matched stores as controls yields a change of sales of -195 but is highly insignificant with a p-value of .594. The addition of a single store to the estimation and the insignificance of the treatment effect continues when carried out to the Top 7 Best Matches. The pattern of insignificance that occurs when multiple stores are used in the control group suggest that there is no significant impact on sales when the model is estimated using no control variables.

The estimation results using the model with weather seasonality as a control variable are almost identical to the results of the model estimated without controls. The Best Match control group is significant, but when the second best match is added to the estimation this significance disappears. The addition of the weekly average maximum temperature control variable sharpens the estimations observed in the non-controlled model. When controlling for weather, parameter estimates increase and the p-value decreases. The p-value of the Best Matched store drops from .08 to .04 making it significant at the 95% level, decreasing p-values can be seen across both the cluster and geographic control groups, although not enough to bring any other estimations to significant levels. The results for both the treatment and control groups in the post-treatment period were statistically insignificant. Across all control groups the seasonality measure, weekly average maximum temperature, was positive and statistically significant. Using the weekly average maximum temperature refines the estimations through its ability to represent short run, within-year, temperature patterns that effects sales.

Significant changes in the results occur when weather is removed from the model and replaced with income. In all cases, the parameter estimate for income are statistically significant. Income being included in the model has a significant effect on the results of the Top 7 Best Matches. In the Top 7 Best Matches control group all of the parameter estimates are significant and the treatment store in the after period has an estimated reduction in sales is around 35%. The significance of the treatment store in the after period dissipates when looking at the Top 2 Best Matches to a 10% loss in sales that is not significant. Similar the estimation with no variable controls, the Best Matched location has around a 4% loss in sales and is marginally significant. The County and All Stores groups in the income model yielded no significant effects in the after period for both treatment and control groups. Income is used to check for robustness of the results, however, it is not a good control measure as it is used as a parameter to match control stores to the treatment store. The use of weather as a control is beneficial as stores are matched based on their yearly average temperature and the control used in the model is based on weekly average temperature.

The final model estimated using both weather and income produced results similar to those of the model using only income. The addition of the weekly average maximum temperature control variable sharpened the parameter estimations by lowering the p-values for all of the estimations. The lowering of the p-values resulted in the Best Matched control store to have a statistically significant reduction in sales of around 6% in addition to lowering the p-value of the already significant Top 7 Best Matches. The model was also estimated using a different treatment date. The results of this estimation are almost identical to that of the original model. Tables 15 through 18 containing the results using December 2011 as the treatment date can be found in the Appendix. Figure 12 compares the sales of packaged salad of the Top Two Best matches with the treatment store. Although the scale of sales may be different, figure 12 illustrates how the pattern of sales throughout the year is closely matched between the treatment store and the top two matched control stores.

Table 10. Parameter Estimation with No Controls												
	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**8249.9**	<.0001	**8560.6**	<.0001	**5943.9**	<.0001	**9391.7**	<.0001	**11177.0**	<.0001	**5861.3**	<.0001
after	150.4	0.3486	-101.6	0.6305	-52.0	0.6033	-0.4	0.9979	-151.2	0.2061	15.8	0.7362
treatment	**2314.8**	<.0001	**2004.1**	<.0001	**4620.7**	<.0001	**1173.0**	0.006	**-612.6**	<.0001	**4703.4**	<.0001
t_times_a	-447.0	0.3246	-195.0	0.594	-244.6	0.0842	-296.2	0.6886	-145.4	0.3898	-312.5	0.5361
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

Table 11. Parameter Estimation with Weather Control												
	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**5787.5**	<.0001	**7222.9**	<.0001	**4390.9**	<.0001	**8029.6**	<.0001	**9181.2**	<.0001	**3884.0**	<.0001
after	46.0	0.7732	-162.0	0.442	-112.2	0.2207	-59.8	0.6775	-254.3	0.0194	-67.0	0.154
treatment	**2273.3**	<.0001	**2016.6**	<.0001	**4649.5**	<.0001	**1166.1**	0.0062	**-612.6**	<.0001	**4622.4**	<.0001
t_times_a	-472.0	0.2934	-203.1	0.5764	-263.2	0.042	-307.5	0.6771	-145.4	0.3418	-336.0	0.5036
weekly avgtemp	**34.1**	<.0001	**18.0**	0.001	**20.8**	<.0001	**18.6**	<.0001	**27.2**	<.0001	**28.0**	<.0001
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

Table 12. Parameter Estimation with Income Control												
	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**-10773.0**	<.0001	-1920.6	0.3599	**4280.3**	<.0001	**1525.6**	<.0001	**9404.6**	<.0001	**5723.8**	<.0001
after	**-580.7**	0.0004	**-779.9**	0.0017	-158.5	0.1763	**-335.6**	0.0097	-267.1	0.046	-0.9	0.9846
treatment	**2625.2**	<.0001	**1928.5**	<.0001	**4596.6**	<.0001	421.9	0.2749	**-612.6**	<.0001	**4706.2**	<.0001
t_times_a	**-939.7**	0.0311	-207.2	0.5653	-248.5	0.0789	-524.6	0.4327	-145.4	0.3886	-304.5	0.5464
annual average_income	**0.3**	<.0001	**0.2**	<.0001	0.0	0.0823	**0.1**	<.0001	0.0	0.0551	**0.0**	<.0001
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

Table 13. Parameter Estimation with Weather and Income Control												
	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**-13334.0**	<.0001	-3736.8	0.0813	**2167.1**	0.0158	-62.0	0.8675	**7178.5**	<.0001	**3664.0**	<.0001
after	**-688.6**	<.0001	**-869.2**	0.0005	**-254.0**	0.0179	**-408.2**	0.0017	**-385.2**	0.0016	-90.5	0.0549
treatment	**2584.5**	<.0001	**1939.3**	<.0001	**4618.0**	<.0001	412.4	0.285	**-612.6**	<.0001	**4624.3**	<.0001
t_times_a	**-967.0**	0.0248	-216.3	0.5453	**-268.6**	0.0371	-534.8	0.423	-145.4	0.3396	-326.2	0.5159
weekly avgtemp	**34.6**	<.0001	**19.4**	0.0003	**21.1**	<.0001	**21.5**	<.0001	**27.3**	<.0001	**28.6**	<.0001
annual average_income	**0.3**	<.0001	**0.2**	<.0001	**0.0**	0.0118	**0.1**	<.0001	**0.0**	0.0171	**0.0**	<.0001
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

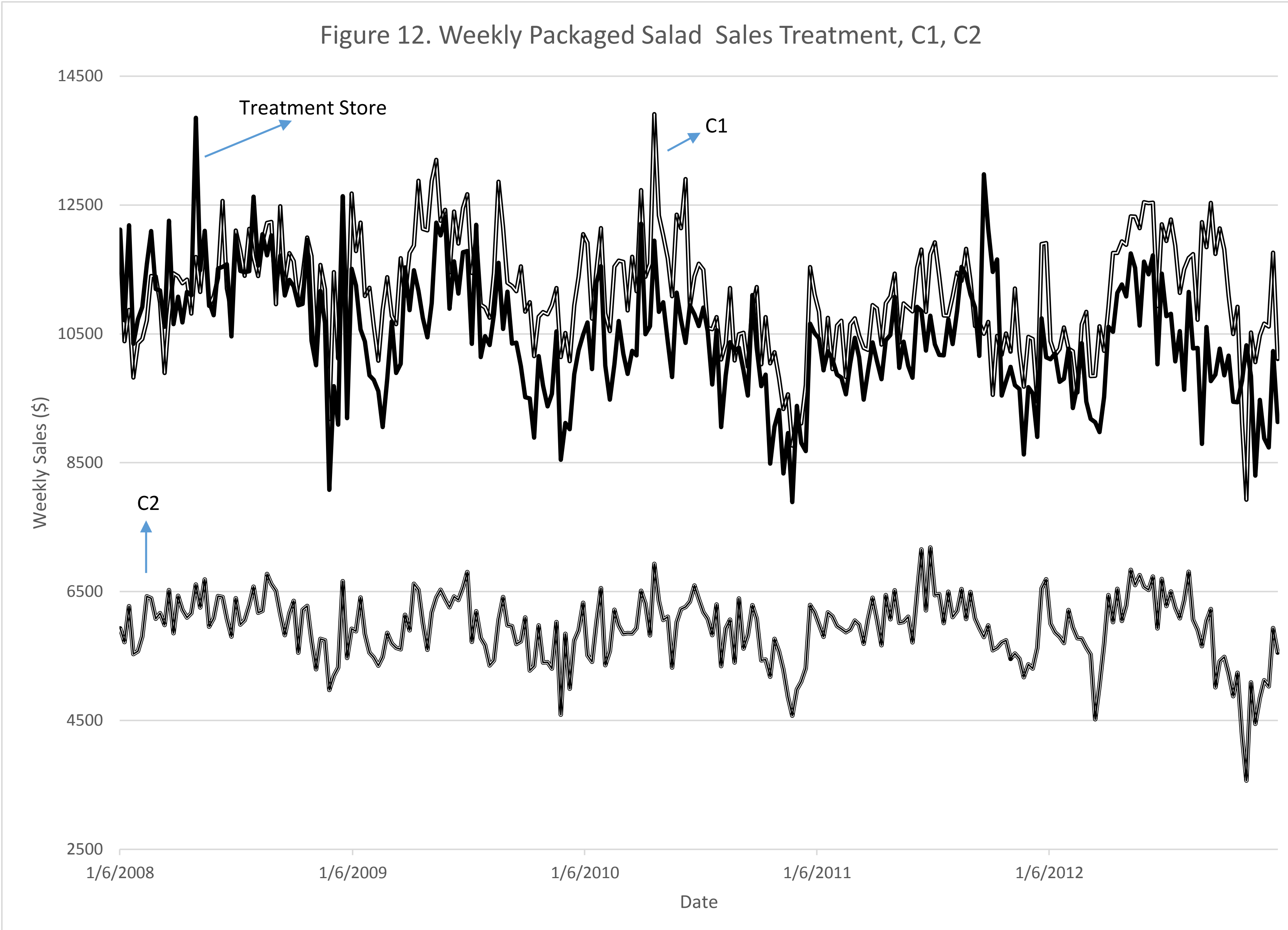


Figure 12. Weekly Packaged Salad Sales Treatment, C1, C2
Treatment Store
C1
C2
14500
12500
10500
8500
6500
4500
2500
Weekly Sales ($)
1/6/2008
1/6/2009
1/6/2010
1/6/2011
1/6/2012
Date

When estimating weekly packaged salad sales using a weather controlled model, the treatment store experienced a decline in sales regardless of the control group used. In the geographic suburb control group, which is also the best matched control, the decrease in sales for the treatment store in the after period is statistically significant. It was only in this single control group in the weather model that exhibited a statistically significant decrease in sales for the treatment store in the after period. All other control groups in the weather controlled model presented negative sales for the treatment store in the after period, none were significant. Having a single significant decrease in sales and several insignificant decreases in sales suggests that the implementation of doors for this particular location had a statistically insignificant negative effect on sales.

The limited differences between the weather controlled model and the uncontrolled model suggest that including the weather control had no significant impact on parameter estimates. The limited impact of the weather control variable demonstrates how well the seasonality of the weather matches the seasonality of sales. Matching temperature seasonality and sales seasonality was the purpose of including the weather control in the model. The weather control in the model is designed to smooth out variations in sales because of the positive relationship between temperature and packaged salad sales. When the weather control is included in the model, parameter estimates increase in value and in statistical significance.

The model where income is used as a control variable yields statistically significant decrease in sales for the best seven matches and a slightly significant decrease in sales for the best matched control group. The significance of the decrease in sales is carried over and intensified when both weather and income are used as control variables. The results of these models suggest that overall the installation of doors caused a significant decrease in sales in the after period. The inclusion of income as a control in the model gives the model the appearance of a demand equation. This is not the intended use of the model. The model is strictly designed to measure potential changes in sales between treatment and control groups in the before and after periods. Income was also used in the matching of control stores to the treatment store. Because it is used twice in matching and in parameter estimation, results using income as a control must be interpreted with caution.

When controlling for weather and when having no controls in the model, the results suggest there is no impact on sales in the after period. When controlling for the average income of the stores surrounding zip code and when combining zip code and weather for controls the results show that there is a significant negative impact on sales in some control groups. The message found in the results for this particular treatment store is the change in sales, while statistically insignificant, is negative.

Chapter 6. Conclusions

The results of this study suggest that the installation of doors on refrigerated cases containing packaged salad has little impact on sales. When compared to the controls, the treatment location experienced negative changes in sales of 1 to 9 percent per week with varying degrees of insignificance. The small treatment group (a single store) and a potentially contaminated control group diminish the trustworthiness of the results. The framework of this research has the potential to yield conclusive results if an uncontaminated group of control stores can be obtained.

There are significant hurdles to the implementation of doors on refrigerated cases containing packaged salad. The budget for grocers to improve their stores is limited and often has competing interests. Retailers allocate scarce investment capital to improvements that will give them a return on their investment without negatively affecting customers or sales. This thesis takes a step forward in affirming that enclosed display cases do not have a significant effect on sales. While the loss in weekly sales is insignificant in a majority of the cases, the few models where the treatment effect is significantly negative can be concerning to retailers, especially when operating under slim profit margins (Garry 2010). The overall scope of implementing doors must also be considered. While installing doors on refrigerated cases can reduce the energy consumption of the cases between 40 to 60%, it could potentially reduce the total energy consumption of the store by around 6% (Lindberg, et al. 2008). In some cases a loss of only 2% in sales can negate the potential energy savings benefits (Garry 2010). Retailers must also consider the potential return on the investment of installing doors on packaged salad displays. It is estimated that the time needed to recoup the costs of door installation can be as little as two years and a much as a decade (Garry 2010). The potentially slow returns on investment could play a major part in why there is not more widespread adoption of doored refrigerated cases.

Aside from cost constraints some retailers have entrenched opinions that doors will reduce sales of fresh processed produce. Many retailers feel that doors close off the products from the customer and that they will limit impulse purchasing. This opinion runs deep and may be difficult to change. However, new construction and fully remodeled stores have, in some cases, embraced doored displays on all refrigerated products. Some retailers are installing new doored display cases and pairing them with adaptive LED lighting to not only improve energy efficiency, but also the visual appeal of the product (Alaimo 2013). More and more new stores and fully remodeled stores are building their designs around doors on refrigerated cases.

Finally, the retailers in particular must be examined. Grocery retailers operate under thin margins, in a hyper-competitive market. In order for doored display cases to be adopted the benefits must outweigh the costs. Some retailers have begun doing their own internal studies to determine how implementing doors on their refrigerated cases affects their bottom line. Ongoing research by grocers may explain why obtaining specific data from the anonymous retailer in this study was not possible. Preliminary results from other grocers as mentioned in industry periodicals suggest that there have been mixed impacts of doors on sales. Some grocers are finding that implementing doors is highly effective at saving energy (up to 30% energy savings), while others are finding that the loss in sales is significant enough to negate all energy savings (Garry 2010). Retailers must also consider how doors might affect their day-to-day activities such as stocking, rotation and cleaning the shelves and product. The benefits must outweigh the costs for widespread implementation of doored display cases to occur.

Further research should be focused on improving upon treatment and control data sets, and looking at the effects of sales over a longer period of time. Improvements to the control data set would include determining store locations where doors were installed and the dates of the installation. Improved data could provide an uncontaminated control data set, providing more reliable estimates of how doors affect sales across different types of locations. Looking at a longer time span would allow for potential changes in consumer preference and understanding to appear in response to doors on packaged salad. Improved data could yield results that may be strong enough to conclude with more certainty whether doors' effect on sales is enough to outweigh the energy savings benefits and other benefits.

References

De Frias, J. A., and Liping Kou. 2014. "Modification of retail open refrigerated display cabinets result in significant energy savings, compliance with Food Code, and improved product quality." *Food and Drug Administration.* College Park.

Dole. 2015. *Dole History Timeline.* http://dole.com/Company%20Info/Timeline.

Evans, J.A. 2007. "Temperature and Energy performance of refrigerated retail display and commercial catering cabinets under test conditions." *International Journal of Refrigeration* 398-408.

Fresh Express. 2015. *Fresh Facts.* http://www.freshexpress.com/about/fresh-facts.

Fricke, Brian, and Bryan Becker. 2010a. "Doored Display Cases They save Energy, Dont Lose Sales." *American Sociery of Heating, Refrigeration, and Air Conditioning Engineers* 18-26.

Fricke, Brian, and Bryan Becker. 2010b. "Energy Use of Doored and Open Vertical Refrigerated Display Cases." *International Refrigeration and and Air Conditioning Conference* Paper 1154.

Garry, Michael. 2010. "New Tests Support Doors on Cases." *Supermarket News*, October 4.

Islam, Rafiqul. 2012. "Taking a Sensible Choice of Sustainable Super Market Refrigeration Equipment." *International Refrigeration and Air Conditioning Conference* Paper 1346.

Kou, Liping, Yaguang Lou, David Ingram, Shoulei Yan, and Wayne Jurick. 2014. "Open-refrigerated retail display case temperature profile and its impact on product quality and microbiota of stored baby spinach." *Food Control* 1-7.

Lindberg, Ulla, Monica Axell, Per Fahlaen, and Niklas Fransson. 2008. "Supermarkets, Indoor Climate and Energy Efficency- Field Measurements Before and After Installation of Doors on Refrigerated Cases." *International Refrigeration and Air Conditioning Conference* Paper 879.

Thompson, Gary, and Paul Wilson. 1999. "Market Demands for Bagged, Refrigerated Salads." *Journal of Agricultural and Rerource Economics* 463-481.

Appendix

Data Cleaning Process

The data from IRI consisted of weekly observations from 127 store locations from 2008 to 2012. Variables for packaged salad include UPC, product type, type of packaging, brand name, item number, salad parent company, UPC description, store address, the retailer store number, units sold per week, dollars sold per week, IRI week number, and the first day of the observation week among others. There are 261 weeks in the data set beginning with January 1, 2008 and ending December 31, 2012.

To check for missing values a year variable was created. Each stores was checked to verify if there were weekly observations for all five years. Twelve locations were deleted from the data because they did not have observations for all five years. This was attributed to stores opening, closing or being sold at some point between 2008 and 2012. The remaining 115 stores were then combined with data from TDLinx (ACNeilsen). TDLinx data contains store attributes not present in the IRI data. These characteristics include store size, total sales both annually and weekly averages, number of checkout stations, number of employees, the presence of a pharmacy, and whether or not the store sells beer, wine, liquor or gas.

Weather data collected from the NOAA website http://www.ncdc.noaa.gov/cdo-web/search was merged with the complete store data. NOAA data contained the maximum daily temperature, latitude and longitude coordinates for all weather stations present in the grocer's territory. Store addresses were converted to latitude and longitude to match the weather stations. Using the geo-distance function in SAS the store locations were matched to the geographically closest weather station. Daily maximum temperature values were then converted into weekly averages for each store location. These weekly averages were checked to ensure a minimal amount of missing temperature observations. The checking of weekly averages for missing values turned up several locations that had over one hundred missing weekly averages. These weather stations were identified and removed from the data set and the geo-distance function was repeated. This process was continued until the total number of missing weekly averages was brought down to 63, with no more than 5 total missing weekly averages from any single store location.

IRS data were obtained through their website at http://www.irs.gov/uac/SOI-Tax-Stats-Individual-Income-Tax-Statistics-2011-ZIP-Code-Data-(SOI).

Annual data by zipcode for 2008 through 2012 were downloaded. IRS data included total annual gross income and total number of tax returns in each zip code. Total annual gross income was divided by the total number of returns filed that year to obtain average annual income in each zip code.

Clustering Method Overview

The SAS procedure MODECLUS is the clustering method employed for this thesis. The MODECLUS procedure is a non-parametric clustering method. The procedure uses non-parametric density estimation to group the stores into clusters. Using probability density estimates based on the parameters provided, the MODECLUS procedure identifies clusters based upon the radius specified in the code. The number of clusters is not specified, instead the number of clusters is determined by the specified radius and by the density estimation. The data output contains density estimates of each store location and cluster identification numbers.

Below is an example of the code used to implement the clustering of store locations.

```
proc modeclus data=matching.matching_year2010 method=1 r=600 out=out all;
id retailer_store_number ;
```

```
var  year_avgtemp average_income retailer ssqft snmchkout annvol_n swklyvol
sftemploy;
run;
```

This method was chosen over other methods due to difficulties experienced in collecting treatment data. The single treatment location of the data set does not permit for the use of a logistic model to estimate propensity scores to be used as a matching method. More information on the clustering method can be found at http://support.sas.com/documentation/cdl/en/statug/67523/HTML/default/viewer.htm#statug_modeclus_syntax.htm

Table 14. Annual Total Store Sales Categories	
Category	Store Sales ($)
19	$100,000,001 and above
18	75,000,001 to100,000,000
17	50,000,001 to 75,000,000
16	45,000,001 to 50,000,000
15	40,000,001 to 45,000,000
14	35,000,001 to 40,000,000
13	30,000,001 to 35,000,000
12	25,000,001 to 30,000,000
11	20,000,001 to 25,000,000
10	16,000,001 to 20,000,000
9	12,000,001 to 16,000,000
8	8,000,001 to 12,000,000
7	6,000,001 to 8,000,000
6	4,000,001 to 6,000,000
5	2,000,001 to 4,000,000
4	1,500,001 to 2,000,000
3	1,000,001 to 1,500,000
2	500,001 to 1,000,000
1	1 to 500,000

Table 15. Parameter Estimation with No Controls, December 2011 Treatment Date

	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**8,254.5**	<.0001	**8,541.7**	<.0001	**5,958.6**	<.0001	**9,338.5**	<.0001	**11,125.0**	<.0001	**5,846.9**	<.0001
after	223.9	0.2335	-73.4	0.7667	-157.5	0.1772	267.3	0.1128	10.7	0.9392	97.0	0.0786
treatment	**2,295.7**	<.0001	**2,008.5**	<.0001	**4,591.6**	<.0001	**1,211.8**	0.0019	**-574.6**	<.0001	**4,703.3**	<.0001
t_times_a	-639.5	0.229	-342.2	0.4247	-258.1	0.118	-682.8	0.4306	**-426.3**	0.0316	-512.6	0.3864
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

Table 16. Parameter Estimation with Weather Control, December 2011 Treatment Date

	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter	P Value	Parameter	P Value	Parameter	P Value	Parameter	P Value	Parameter	P Value	Parameter	P Value
Intercept	**5,776.4**	<.0001	**7,232.5**	<.0001	**4,429.6**	<.0001	**7,996.0**	<.0001	**9,178.3**	<.0001	**3,866.9**	<.0001
after	186.6	0.3162	-93.2	0.7049	-169.0	0.1143	242.4	0.1502	-33.8	0.7909	64.9	0.2374
treatment	**2,250.2**	<.0001	**2,020.0**	<.0001	**4,618.3**	<.0001	**1,202.4**	0.002	**-574.6**	<.0001	**4,618.2**	<.0001
t_times_a	-659.8	0.2099	-352.1	0.4087	-281.0	0.0634	-688.9	0.4261	**-426.3**	0.0183	-527.7	0.3701
weekly avgtemp	**33.9**	<.0001	**17.5**	0.0014	**20.2**	<.0001	**18.2**	<.0001	**26.2**	<.0001	**27.8**	<.0001
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

Table 17. Parameter Estimation with Income Control, December 2011 Treatment Date

	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**-14,698.0**	<.0001	**-16,531.0**	<.0001	-2,296.4	0.1003	**1,464.4**	<.0001	**5,362.1**	0.0005	**5,712.3**	<.0001
after	**-1,246.0**	<.0001	**-2,516.6**	<.0001	**-813.1**	<.0001	-254.9	0.0958	**-653.9**	0.0036	74.1	0.1804
treatment	**2,701.6**	<.0001	**1,911.8**	<.0001	**4,527.7**	<.0001	480.9	0.1737	**-574.6**	<.0001	**4,707.0**	<.0001
t_times_a	**-1,769.8**	0.0005	**-801.8**	0.0538	**-561.9**	0.0009	-1,153.1	0.1413	**-426.3**	0.0295	-504.8	0.3935
annual average_income	**0.3**	<.0001	**0.4**	<.0001	**0.1**	<.0001	**0.1**	<.0001	**0.1**	0.0002	**0.0**	<.0001
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	

Table 18. Parameter Estimation with Weather and Income Control, December 2011 Treatment Date

	Cluster Match						Geographic Match					
	Top 7 Best Matches		Top 2 Best Matches		Best Match (C2)		County		Suburb (C1)		All Stores	
Variable	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value	Parameter Estimate	P Value
Intercept	**-16,893.0**	<.0001	**-17,844.0**	<.0001	**-3,909.1**	0.0023	-42.3	0.9095	**3,770.4**	0.0073	**3,656.5**	<.0001
after	**-1,269.5**	<.0001	**-2,536.7**	<.0001	**-830.8**	<.0001	-282.2	0.065	**-659.1**	0.0013	35.2	0.5227
treatment	**2,654.9**	<.0001	**1,923.3**	<.0001	**4,553.8**	<.0001	470.7	0.1824	**-574.6**	<.0001	**4,621.4**	<.0001
t_times_a	**-1,780.3**	0.0004	**-811.8**	0.0494	**-587.8**	0.0001	-1,161.5	0.1378	**-426.3**	0.0168	-518.3	0.3784
weekly avgtemp	**32.6**	<.0001	**17.5**	0.0009	**20.3**	<.0001	**20.3**	<.0001	**25.9**	<.0001	**28.3**	<.0001
annual average_income	**0.3**	<.0001	**0.4**	<.0001	**0.1**	<.0001	**0.1**	<.0001	**0.1**	0.0001	**0.0**	<.0001
Number of Control Stores	7		2		1		26		1		114	
Sample Size	2,088		783		522		7,009		522		30,198	