Homework9

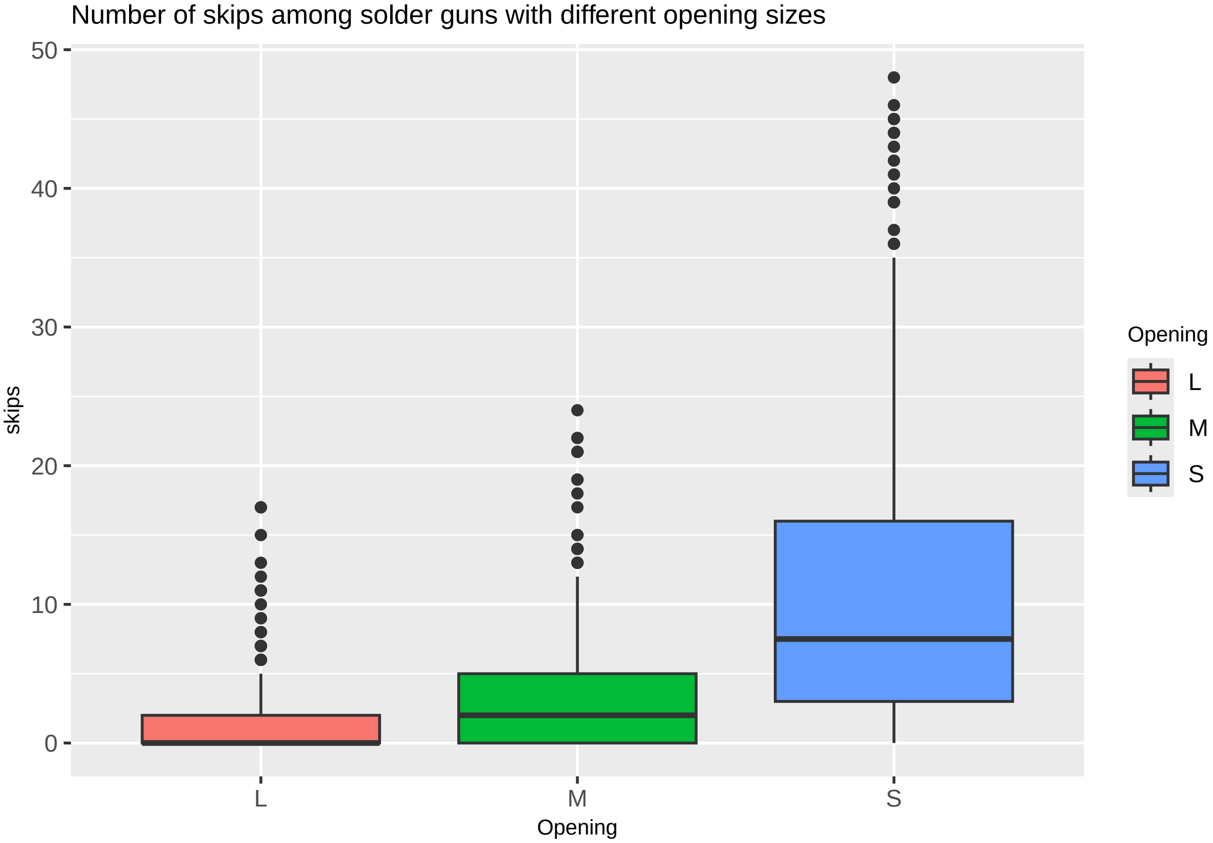
Daniel Wu (EID: djw3627)

2025-04-07

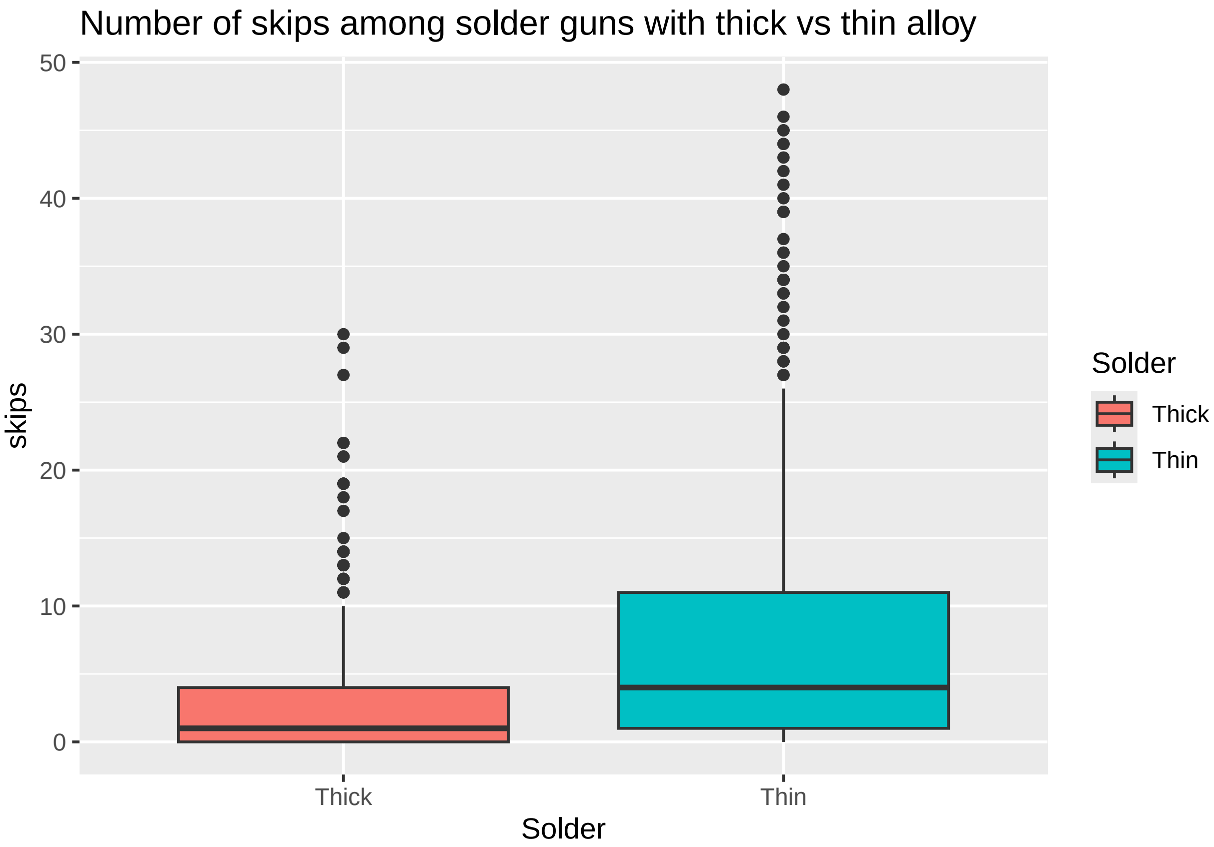
To access my GitHub repository, click here: https://github.com/DanielWu3627/SDS315. Please check the file named **Homework9.Rmd**.

Problem 1

Part A



According to the box plot, the plot showing the guns with large openings has a very low median, and most values are tightly clustered near 0-3 skips. In the plot showing the guns with medium-sized openings, the median is slightly higher than L, and there is more variation, shown by a wider spread. There are more outliers than L, going up to about 25 skips. The plot showing guns with small openings has the highest median (around 7-8). Additionally the spread is very wide, with many outliers, with some as hgih as around 47-48 steps.This suggest that guns with larger opening have fewer skips with less variation and those with smaller openings have more skips with greater variability.



According to the box plot, the plot showing the guns with thick alloy has a very small median (around 1) and most values are clustered between 0 and 4 skips. In the plot showing solder guns with thin alloy, the number of skips are more spread out, with the median also greater (around 4). Additionally, there are more outliers, with some as high as 47-48 skips. Therefore, there is a clear pattern that thin guns produce more skips and have a wide spread while thick guns keep skips low and tightly clustered around 0.

Part B

```
## # A tibble: 6 x 7
##   term                  estimate std_error statistic p_value lower_ci upper_ci
##   <chr>                    <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
## 1 intercept                0.393     0.52      0.756   0.45    -0.628     1.42
## 2 Opening: M               2.41      0.736     3.27    0.001    0.962     3.85
## 3 Opening: S               5.13      0.736     6.97    0        3.68      6.57
## 4 Solder: Thin             2.28      0.736     3.10    0.002    0.836     3.72
## 5 Opening: M:SolderThin   -0.74      1.04     -0.711   0.477   -2.78      1.30
## 6 Opening: S:SolderThin    9.65      1.04      9.28    0        7.61     11.7
```

Part C

The baseline number of skips of guns with large openings and thick alloy is 0.39.

The main effect for the OpeningM variable is 2.41. This is the effect of OpeningM in isolation, meaning the number of skips for medium-sized opening is 2.41 more than large openings. The 95% CI is [0.96, 3.85], which does not include 0, so the effect is statistically significant.

The main effect for the OpeningS variable is 5.13. This is the effect of OpeningS in isolation, meaning the number of skips for small-sized opening is 5.13 more than large openings. The 95% CI is [3.68, 6.57], which does not include 0, so the effect is statistically significant.

The main effect for the SolderThin variable is 2.28. This is the effect of SolderThin in isolation, meaning the number of skips for guns with thin alloy is 2.28 more than thick alloy. The 95% CI is [0.84, 3.72], which does not include 0, so the effect is statistically significant.

The interaction effect for OpeningM and SolderThin is -0.74. In other words, guns with a medium-sized opening and thin alloy yield 0.74 less skips when compared to the summation of the “isolated” effects of the two variables (medium-sized opening and thin alloy). The 95% CI is [-2.78, 1.30], which includes 0, so the effect is not statistically significant.

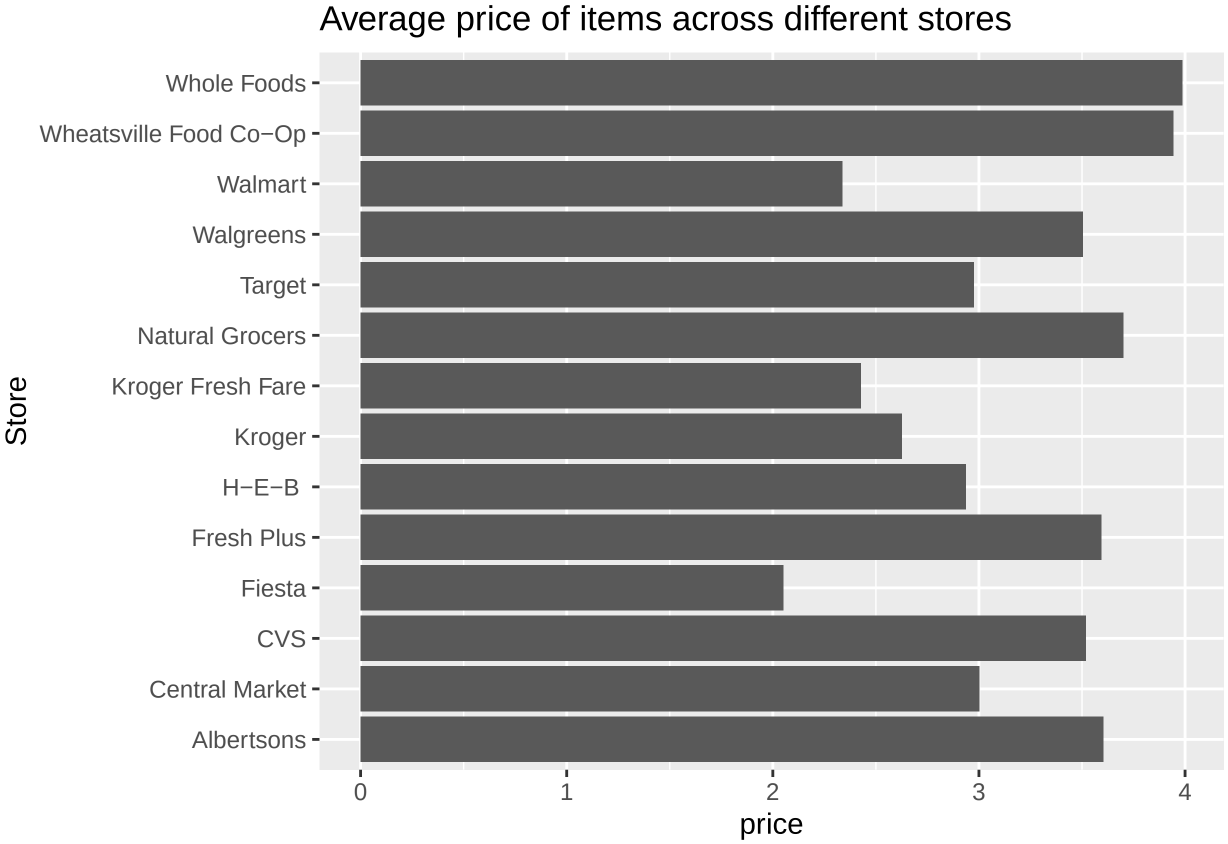
The interaction effect for OpeningS and SolderThin is 9.65. In other words, guns with a small-sized opening and thin alloy yield 9.65 more skips when compared to the summation of the “isolated” effects of the two variables (small-sized opening and thin alloy). The 95% CI is [7.61, 11.70], which does not include 0, so the effect is statistically significant.

Part D

I would recommend a combination of a large opening size and thick alloy to AT&T. Based on the regression analysis, the baseline (large opening + thick solder) told me that there were 0.39 skips for large openings and thick solders. Every other combination is associated with more skips.

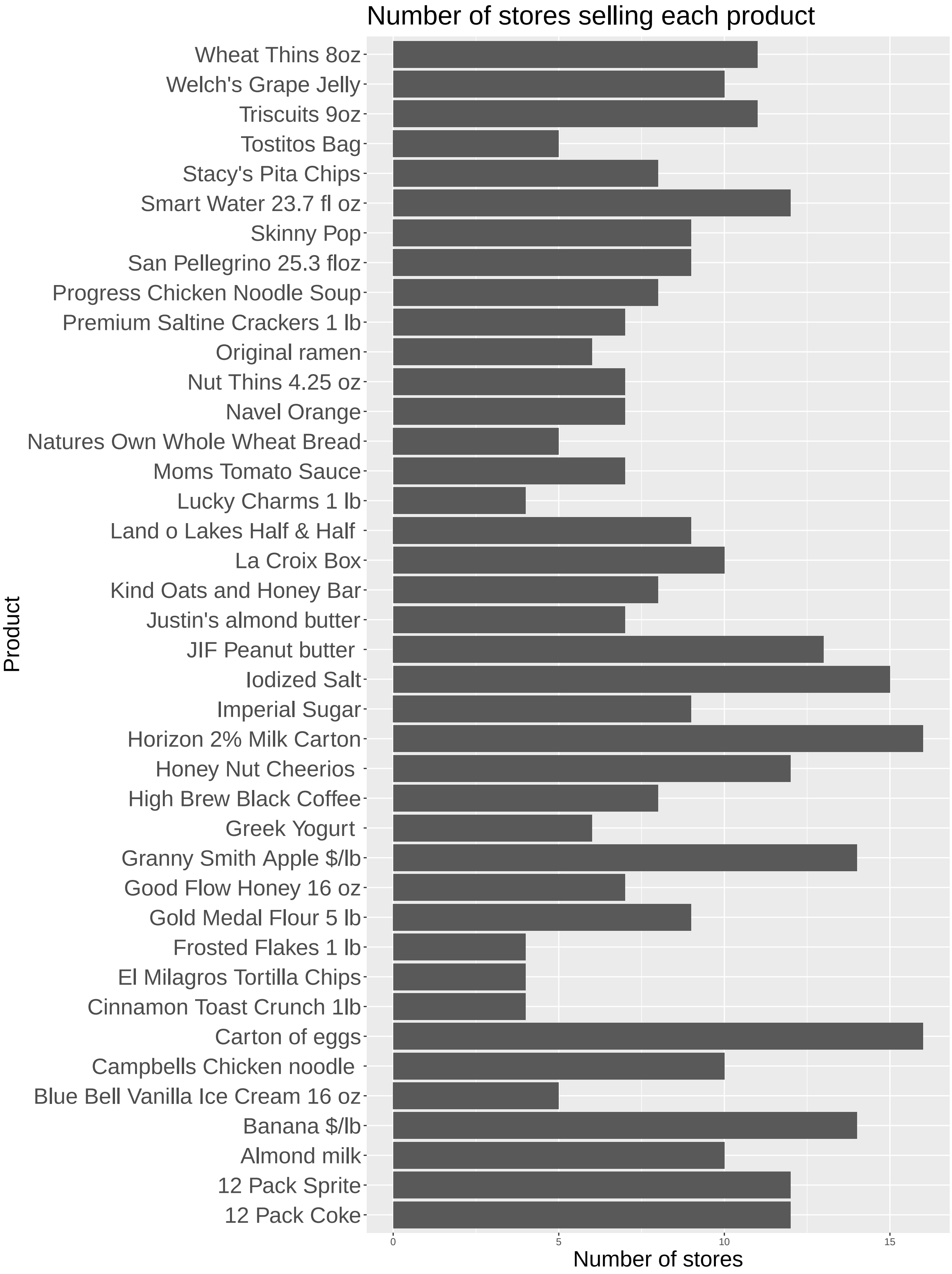
Problem 2

Part A



Based on the plot, we can see that Fiesta has the lowest average price of items, while Whole Foods has the highest average price of items.

Part B


Number of stores selling each product
Product
Wheat Thins 8oz
Welch's Grape Jelly
Triscuits 9oz
Tostitos Bag
Stacy's Pita Chips
Smart Water 23.7 fl oz
Skinny Pop
San Pellegrino 25.3 floz
Progress Chicken Noodle Soup
Premium Saltine Crackers 1 lb
Original ramen
Nut Thins 4.25 oz
Navel Orange
Natures Own Whole Wheat Bread
Moms Tomato Sauce
Lucky Charms 1 lb
Land o Lakes Half & Half
La Croix Box
Kind Oats and Honey Bar
Justin's almond butter
JIF Peanut butter
Iodized Salt
Imperial Sugar
Horizon 2% Milk Carton
Honey Nut Cheerios
High Brew Black Coffee
Greek Yogurt
Granny Smith Apple $/lb
Good Flow Honey 16 oz
Gold Medal Flour 5 lb
Frosted Flakes 1 lb
El Milagros Tortilla Chips
Cinnamon Toast Crunch 1lb
Carton of eggs
Campbells Chicken noodle
Blue Bell Vanilla Ice Cream 16 oz
Banana $/lb
Almond milk
12 Pack Sprite
12 Pack Coke
0
5
10
15
Number of stores

Part C

```
## # A tibble: 44 x 7
##    term                   estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                     <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 "intercept"                5.92      0.24      24.4     0         5.45     6.4
##  2 "Product: 12 Pack Spr~    -0.02      0.31     -0.06    0.95    -0.63     0.6
##  3 "Product: Almond milk"    -2.2       0.33      -6.7     0       -2.85    -1.56
##  4 "Product: Banana $/lb"    -4.88      0.3     -16.2     0       -5.48    -4.29
##  5 "Product: Blue Bell V~    -2.95      0.41      -7.2     0       -3.75    -2.14
##  6 "Product: Campbells C~    -3.46      0.33    -10.5     0       -4.11    -2.82
##  7 "Product: Carton of e~    -3         0.29    -10.2     0       -3.58    -2.42
##  8 "Product: Cinnamon To~    -1         0.44     -2.25    0.03    -1.87    -0.12
##  9 "Product: El Milagros~    -2.04      0.44     -4.59    0       -2.91    -1.17
## 10 "Product: Frosted Fla~    -1.25      0.44     -2.82    0.01    -2.12    -0.38
## # i 34 more rows
```

Compared with ordinary grocery stores, at the 95% confidence level, convenience stores charge somewhere betwen [0.41, 0.92] dollars more for the same product. The estimated difference between the two types of stores is 0.66 dollars for the same product.

Part D

```
## # A tibble: 53 x 7
##    term                   estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                     <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 "Product: Original ra~    -5.2       0.35    -14.7        0    -5.9     -4.5
##  2 "Product: Banana $/lb"    -4.86      0.28    -17.5        0    -5.4     -4.31
##  3 "Product: Navel Orang~    -4.06      0.34    -12          0    -4.72    -3.39
##  4 "Product: Greek Yogur~    -3.93      0.35    -11.1        0    -4.63    -3.23
##  5 "Product: Iodized Sal~    -3.86      0.27    -14.1        0    -4.39    -3.32
##  6 "Product: San Pellegr~    -3.73      0.31    -11.9        0    -4.35    -3.12
##  7 "Product: Granny Smit~    -3.7       0.28    -13.3        0    -4.24    -3.15
##  8 "Product: Smart Water~    -3.66      0.29    -12.7        0    -4.22    -3.09
##  9 "Product: Campbells C~    -3.55      0.3     -11.7        0    -4.15    -2.96
## 10 "Product: Land o Lake~    -3.38      0.31    -10.8        0    -4       -2.77
## # i 43 more rows
```

The two stores that charge the lowest prices when comparing the same product are Walmart (estimate is 0.99 dollars less than Albertsons) and Kroger Fresh Fare (estimate is 0.90 dollars less than Albertsons). The two stores that charge the highest prices are Wheatsville Food Co-Op (estimate is 0.29 dollars more than Albertsons) and Whole Foods (estimate is 0.36 dollars more than Albertsons).

Part E

Central Market charges more than HEB for the same product. Central Market has an estimate of -0.57, and HEB has an estimate of -0.65. -0.57 - (-0.65) = 0.08. This means Central Market charges about 0.08 more than HEB. The upper bound for the 95% confidence interval is -0.23 - (-0.35) = 0.12 and the lower bound is -0.92 - (-0.95) = 0.03. This means that Central Market charges somewhere between 0.03 and 0.12 dollars more than HEB for the same product.

Part F

```
## # A tibble: 41 x 7
##    term                   estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                     <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 "intercept"              5.62      0.249    22.6     0         5.13     6.11
##  2 "Product: 12 Pack Spr~   -0.018     0.328    -0.056   0.955   -0.664    0.627
##  3 "Product: Almond milk"   -2.11      0.345    -6.13    0        -2.79    -1.44
##  4 "Product: Banana $/lb"   -4.91      0.316   -15.5     0        -5.53    -4.29
##  5 "Product: Blue Bell V~   -2.91      0.429    -6.78    0        -3.75    -2.06
##  6 "Product: Campbells C~   -3.37      0.345    -9.78    0        -4.05    -2.70
##  7 "Product: Carton of e~   -2.97      0.307    -9.68    0        -3.58    -2.37
##  8 "Product: Cinnamon To~   -1.20      0.465    -2.57    0.011   -2.11    -0.281
##  9 "Product: El Milagros~   -2         0.464    -4.31    0        -2.91    -1.09
## 10 "Product: Frosted Fla~   -1.45      0.465    -3.12    0.002   -2.36    -0.536
## # i 31 more rows
```

```
## # Standardization method: refit
##
## Parameter                                    | Std. Coef. |         95% CI
## ----------------------------------------------------------------------------
## (Intercept)                                  |       1.08 | [ 0.86,  1.31]
## Product [12 Pack Sprite]                     |  -9.03e-03 | [-0.33,  0.31]
## Product [Almond milk]                        |      -1.04 | [-1.37, -0.71]
## Product [Banana $/lb]                        |      -2.42 | [-2.72, -2.11]
## Product [Blue Bell Vanilla Ice Cream 16 oz]  |      -1.43 | [-1.85, -1.02]
## ProductCampbells Chicken noodle              |      -1.66 | [-1.99, -1.33]
## Product [Carton of eggs]                     |      -1.46 | [-1.76, -1.17]
## Product [Cinnamon Toast Crunch 1lb]          |      -0.59 | [-1.04, -0.14]
## Product [El Milagros Tortilla Chips]         |      -0.98 | [-1.43, -0.53]
## Product [Frosted Flakes 1 lb]                |      -0.71 | [-1.16, -0.26]
## Product [Gold Medal Flour 5 lb]              |      -1.03 | [-1.38, -0.69]
## Product [Good Flow Honey 16 oz]              |       0.52 | [ 0.15,  0.89]
## Product [Granny Smith Apple $/lb]            |      -1.85 | [-2.15, -1.54]
## ProductGreek Yogurt                          |      -1.93 | [-2.32, -1.54]
## Product [High Brew Black Coffee]             |      -1.39 | [-1.75, -1.03]
## ProductHoney Nut Cheerios                    |      -0.83 | [-1.15, -0.52]
## Product [Horizon 2% Milk Carton]             |      -0.53 | [-0.83, -0.23]
## Product [Imperial Sugar]                     |      -1.19 | [-1.53, -0.85]
## Product [Iodized Salt]                       |      -1.89 | [-2.19, -1.59]
## ProductJIF Peanut butter                     |      -1.35 | [-1.67, -1.04]
## Product [Justin's almond butter]             |       3.38 | [ 3.01,  3.75]
## Product [Kind Oats and Honey Bar]            |      -0.83 | [-1.19, -0.47]
## Product [La Croix Box]                       |      -0.48 | [-0.82, -0.15]
## ProductLand o Lakes Half & Half              |      -1.56 | [-1.91, -1.22]
## Product [Lucky Charms 1 lb]                  |      -0.84 | [-1.29, -0.39]
## Product [Moms Tomato Sauce]                  |       0.74 | [ 0.37,  1.11]
## Product [Natures Own Whole Wheat Bread]      |      -1.22 | [-1.63, -0.80]
## Product [Navel Orange]                       |      -1.92 | [-2.29, -1.55]
## Product [Nut Thins 4.25 oz]                  |      -1.19 | [-1.56, -0.82]
## Product [Original ramen]                     |      -2.45 | [-2.84, -2.06]
## Product [Premium Saltine Crackers 1 lb]      |      -1.11 | [-1.48, -0.74]
## Product [Progress Chicken Noodle Soup]       |      -1.49 | [-1.85, -1.13]
## Product [San Pellegrino 25.3 floz]           |      -1.74 | [-2.08, -1.39]
## Product [Skinny Pop]                         |      -0.96 | [-1.31, -0.62]
## Product [Smart Water 23.7 fl oz]             |      -1.84 | [-2.16, -1.52]
## Product [Stacy's Pita Chips]                 |      -0.83 | [-1.18, -0.47]
## Product [Tostitos Bag]                         |       -0.81 | [-1.22, -0.39]
## Product [Triscuits 9oz]                        |       -1.13 | [-1.45, -0.80]
## Product [Welch's Grape Jelly]                  |       -1.48 | [-1.82, -1.15]
## Product [Wheat Thins 8oz]                      |       -1.13 | [-1.46, -0.81]
## Income10K                                      |       -0.03 | [-0.07,  0.01]
```

Based on the sign of the 10K coefficient, consumers in poorer zip codes seem to pay more for the same product on average. Based on the coefficient estimate of -0.014, for every 10,000 dollar increase in income the average price paid seemed to decrease by approximately 0.01 dollars. However, the 95% confidence interval is [-0.033, 0.005], which passes 0, so if we were to generalize to a different sample, there might be no statistically significant relationship.

A one-standard deviation increase in the income of a ZIP code seems to be associated with a -0.03 standard-deviation change in the price that consumers in that ZIP code expect to pay for the same product.

Problem 3

A. True. Accordign to Figure A1 and Model A, there is a significant relationship between % minority and FAIR policies per 100 housing units ($p<0.001$; Rˆ2 = 0.516). For every 1% increase in minority population, there is a 0.014 increease in FAIR policies.

B. Undecidable. There is no model explicitely showing any interaction effect between minority percentage and the age of the housing stock in the way that these two variables are related to the number of FAIR policies in a ZIP code. Figure B1 shows a weak (Rˆ2=0.06) and not statistically significant relationshp between housing age and % minority (p=0.125). If we have a model that includes both minority and age and their interaction, we can determine whether there is an interaction effect between between age and miniority based on the 95% confidence interval of the coefficient for the interaction term is not statistically significant (p=0.143).

C. False. Figure C1 shows that the red line (high fire risk) has a slightly steeper slope than the blue line (low fire risk). However, the p-value of the association between minority % and FAIR policies from Model C is 0.839. Therefore, the interaction effect is not statistically significant because of the high p-value. Therefore, we can only say that there is no significant difference in the relationship between minority % and number of FAIR policies among high vs. low fire risk ZIP codes.

D. False. Model D1 shows the relationship between FAIR policies and % minority alone (coefficient is 0.014, $p<0.001$). Even after controlling for income in Model D2, the p-value of the association between minority percentage and FAIR policy uptake is still statistically significant (p=0.002). Income does not explain away the relationship between minority % and FAIR policies. The association persists with statistical significance.

E. True. In Model E, the relationship between minority and FAIR policies is statistically significant (coefficient is 0.008, p=0.006) even after controlling for income, fire, and age. Rˆ2 is 0.662, meanign the model has good explanatory power.