

# HW9

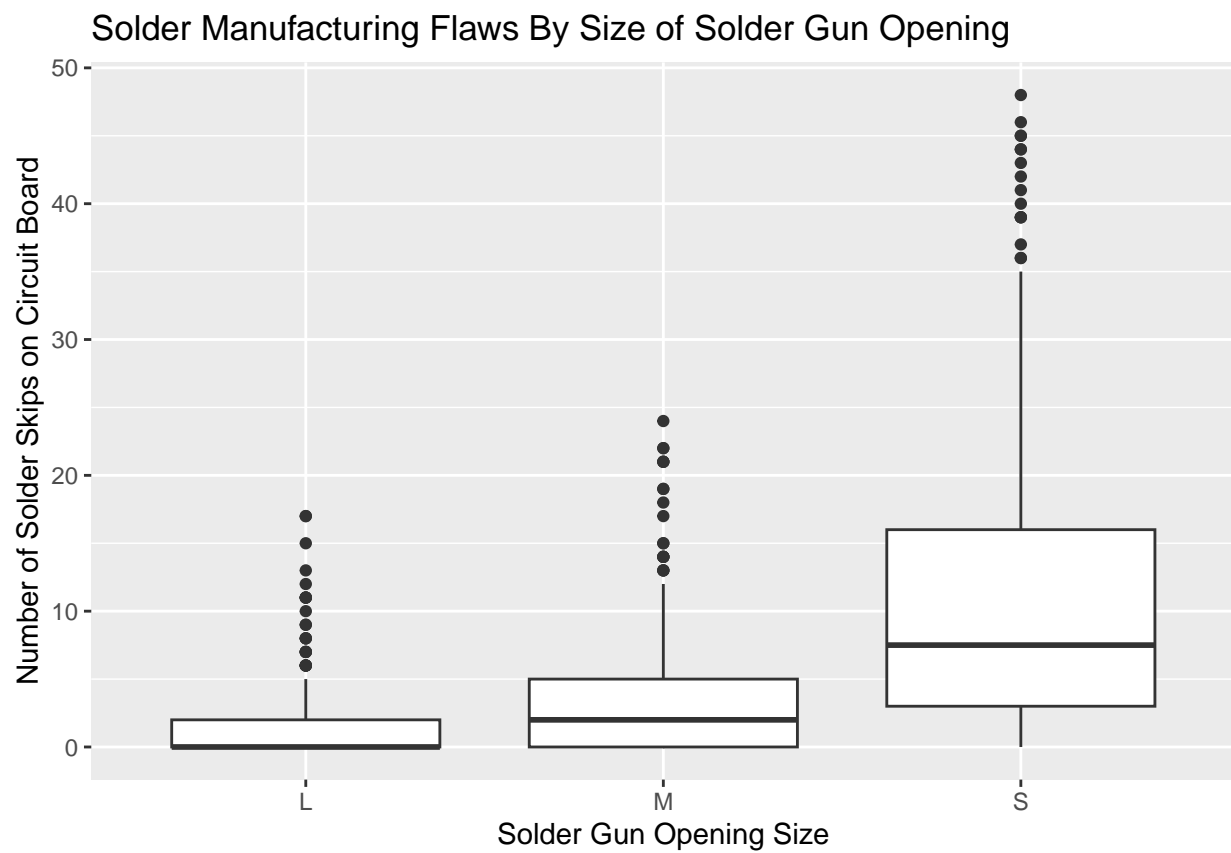
Surafel Geleta

ssg2775

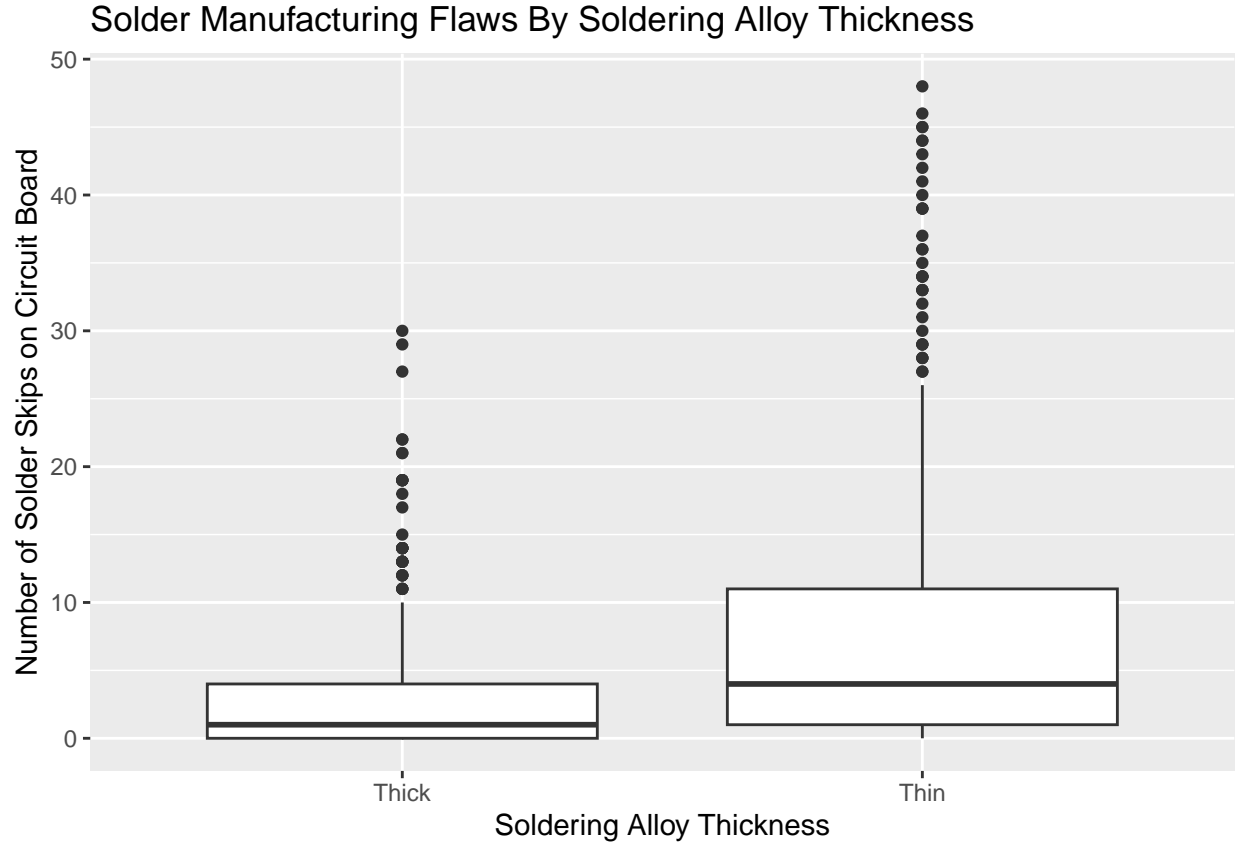
[https://github.com/surafelgeleta/SDS315\\_HW9](https://github.com/surafelgeleta/SDS315_HW9)

## Problem 1

### Part A



Above is a factored boxplot showing the relationship between the number of solder skips on a circuit board and the solder gun's opening size. Circuit boards tend to have more solder skips the smaller the solder gun's opening size is. Moreover, the distribution of solder skips for each opening size appears to be right skewed.



The boxplot above shows the relationship between solder skips on circuit boards and thickness of the soldering alloy. Thin alloys tend to have more manufacturing flaws than thick alloys. As with the previous graph, the distribution of solder skips is right skewed for both groups.

## Part B

Table 1: Interaction Regression Estimates and Confidence Intervals

| Term                | Estimate | Lower Bound | Upper Bound |
|---------------------|----------|-------------|-------------|
| Intercept           | 0.39     | -0.63       | 1.41        |
| OpeningM            | 2.41     | 0.96        | 3.85        |
| OpeningS            | 5.13     | 3.68        | 6.57        |
| SolderThin          | 2.28     | 0.84        | 3.72        |
| OpeningM:SolderThin | -0.74    | -2.78       | 1.30        |
| OpeningS:SolderThin | 9.65     | 7.61        | 11.70       |

## Part C

It can be stated with 95% confidence that the true baseline number of solder skips on AT&T's circuit boards when the soldering gun opening is large and the soldering alloy is thick lies between -0.63 and 1.41.

It can be stated with 95% confidence that the true main effect of the soldering gun's opening being medium-sized lies between 0.96 and 3.85 additional solder skips.

It can be stated with 95% confidence that the true main effect of the soldering gun's opening being small-sized lies between 3.68 and 6.57 additional solder skips.

It can be stated with 95% confidence that the true main effect of the soldering alloy being thin lies between 0.84 and 3.72 additional solder skips.

It can be stated with 95% confidence that the true interaction effect of the soldering gun opening being medium-sized and the soldering alloy being thin lies between -2.78 and 1.30 additional solder skips.

It can be stated with 95% confidence that the true interaction effect of the soldering gun opening being small-sized and the soldering alloy being thin lies between 7.61 and 11.70 additional solder skips.

## Part D

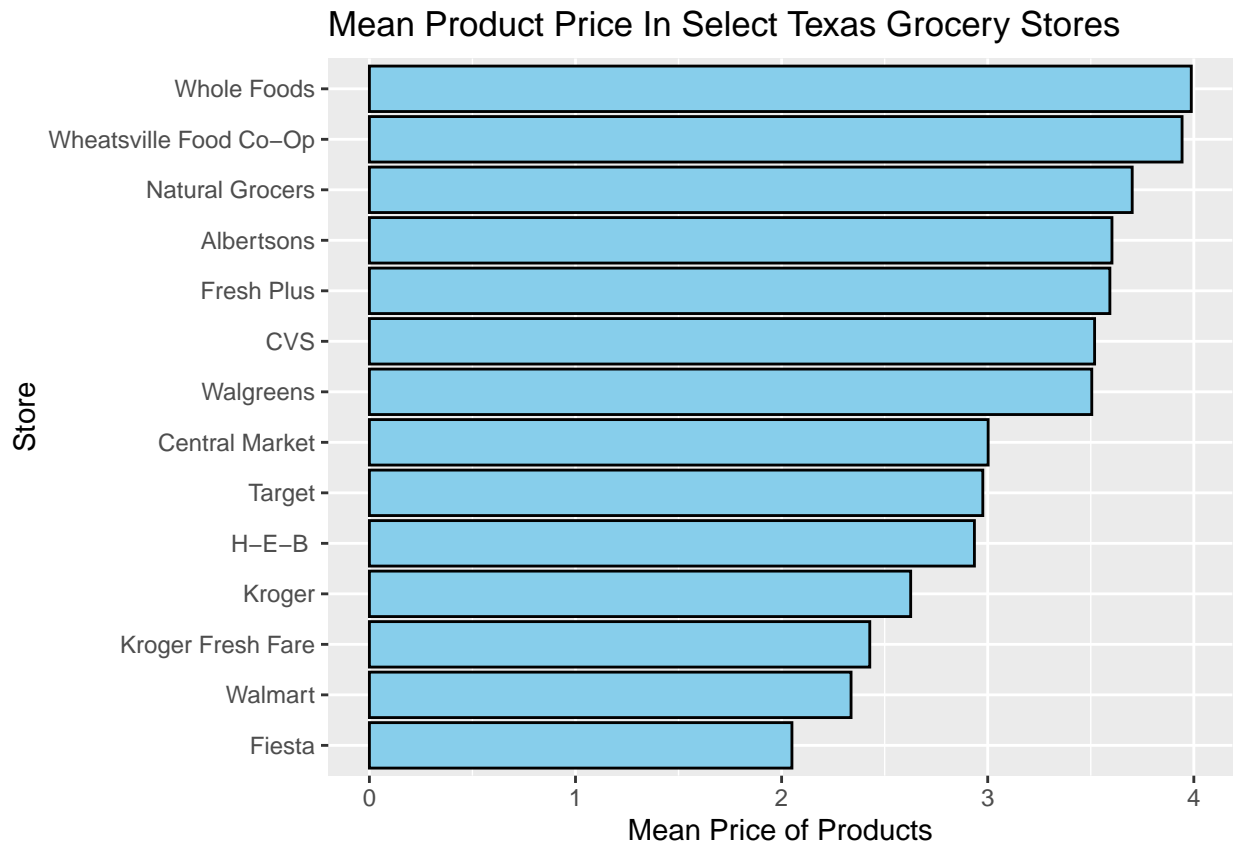
Table 2: Estimated Solder Skips by Solder Gun Opening Size and Alloy Thickness

|       | Small | Medium | Large |
|-------|-------|--------|-------|
| Thin  | 17.45 | 4.34   | 2.67  |
| Thick | 5.52  | 2.80   | 0.39  |

The table above shows predicted solder skips on AT&T circuit boards according to the soldering gun opening size and whether the soldering alloy was thick or thin. I would most likely recommend a large solder gun opening size and a thick alloy thickness because that combination has the lowest predicted solder skips on circuit boards.

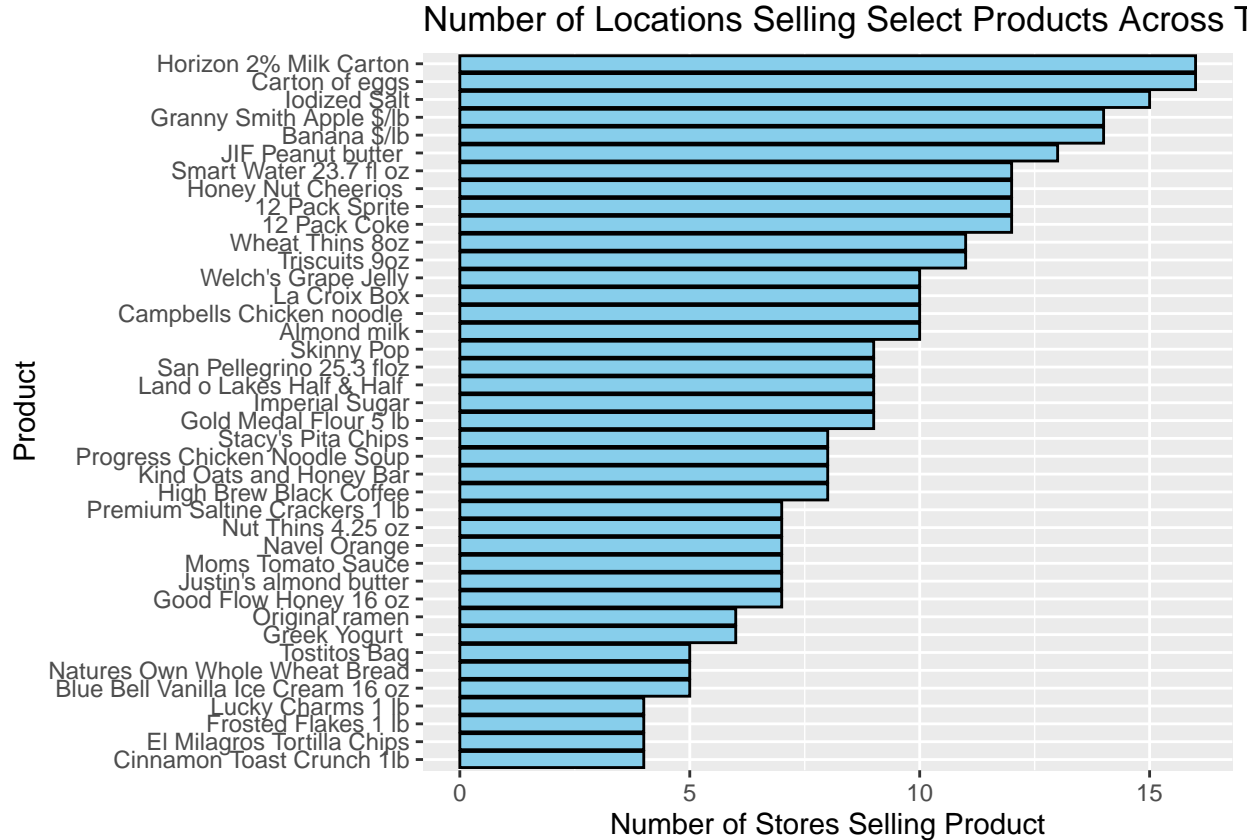
## Problem 2

### Part A



The barplot above shows the distribution of mean prices of products sold in a particular store across a number of Texas stores. Whole Foods stores, followed closely by Wheatsville Food Co-Op, have the highest mean price for the products they sell, while Fiesta stores have the lowest mean price for the products they sell.

## Part B



The barplot above displays the number of stores selling a select product across the dataset of Texas stores. Horizon 2% Milk Cartons and egg cartons are the most frequently sold, with 16 stores selling the products. On the other hand, Lucky Charms, Frosted Flakes, El Milagros Tortilla Chips, and Cinnamon Toast Crunch are the least sold products, with only 4 stores selling the goods.

## Part C

Compared with ordinary grocery stores (like Albertsons, HEB, or Krogers), convenience stores charge somewhere between 0.41 and 0.92 dollars more for the same product.

## Part D

Walmart and Kroger Fresh Fare seem to charge the lowest prices when comparing the same product. Whole Foods and Wheatsville Food Co-Op seem to charge the highest prices when comparing the same product.

## Part E

The coefficient on the storeCentralMarket variable in the regression model is -0.57 while the coefficient on the StoreH-E-B variable is -0.65. The 95% confidence intervals for the two coefficients are [-0.92, -0.23] for storeCentralMarket and [-0.95, -0.35] for storeH-E-B.

## Part F

The sign of the Income10K coefficient is negative, which indicates that consumers in poorer ZIP codes seem to pay more for the same product on average because holding product constant, for every 10,000 dollar increase in income, consumers pay 0.014 dollars less. 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: “ZIP codes with a higher percentage of minority residents tend to have more FAIR policies per 100 housing units”**

True. In Table 1, a confidence interval of the coefficient in a linear model regressing FAIR policies on minority residents shows with 95% confidence that the true change in the number of FAIR policies per 100 housing units for a percentage point increase in the minority population in a zip code lies between 0.009 and 0.018. Since this interval does not contain 0, there is some evidence to suggest that there is a positive relationship between FAIR policies and the minority percentage in a zip code.

**B: “The evidence suggests an 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.”**

Undecidable based on the current evidence. While the results of a linear model regressing minority percentage on housing age shown in Table 2 and Figure B1 do not appear to show a statistically significant relationship between minority percentage and percent housing built before WWII, this does not prove or disprove an interaction between the two variables because association is not indicative of interaction. This question may be answered if a linear model was created regressing FAIR policies on minority percentage, age of housing stock, and an interaction term between minority percentage and age of housing stock. A 95% confidence interval should be generated from the model, and the interval for the coefficient of the interaction term should be analyzed to see if an interaction is plausible.

**C: “The relationship between minority percentage and number of FAIR policies per 100 housing units is stronger in high-fire-risk ZIP codes than in low-fire-risk ZIP codes.”**

False. This statement would mean that the relationship between minority percentage and FAIR policies depends on the fire risk of the ZIP code i.e. that there is an interaction between minority percentage and fire risk in how they relate to FAIR policies. However, as shown in Table 3, a confidence interval shows with 95% confidence that the additional effect (on FAIR policies per 100 housing units) of a low fire risk on a percentage point increase in the minority percentage of a ZIP code lies between -0.012 and 0.01. This interval is not suggestive of an interaction between minority percentage and fire risk, and there is no precision in estimating the interaction term itself. Visually, in Figure C1, the slopes between high and low fire risk when comparing minority percentage and FAIR policies appear to be parallel; this also suggests that the relationship between minority percentage and FAIR policies does not seem to depend on the fire risk.

**D: Even without controlling for any other variables, income “explains away” all the association between minority percentage and FAIR policy uptake.**

False. In Table 5, a confidence interval on the linear model predicting FAIR policies, with minority percentage and income as predictor variables, still shows with 95% confidence that while adjusting for income, a percentage point increase in minority population of ZIP code is associated with a change in FAIR policies per 100 housing units between 0.004 and 0.015. Comparing this to the confidence interval on the linear model with only income and FAIR policies, this confidence interval may suggest a weaker relationship between minority percentage and FAIR policies when adjusting for income, but we cannot say that income “explains away” the association because there is still evidence to reject the null hypothesis that there is no association between minority percentage and FAIR policies.

**E: Minority percentage and number of FAIR policies are still associated at the ZIP code level, even after controlling for income, fire risk, and housing age.**

True. In Table 6, a confidence interval on a linear model adjusting for income, fire risk, and housing age (i.e. adding these variables to the model) found with 95% confidence that the true change in FAIR policies per 100 housing units for every one percentage point increase in the minority population of a ZIP code lies between 0.003 and 0.014. Compared to the first linear model regressing FAIR policies on minority percentage that estimated with 95% confidence that a minority percentage-point increase on FAIR policies lied in [0.009, 0.018], the adjusted coefficient on minority is likely smaller, but still nonzero suggesting that there is still a relationship between minority percentage and FAIR policies per 100 housing units after controlling for income, fire risk, and housing age.