SDSHomework9

Karthikeya Kasa

2025-04-21

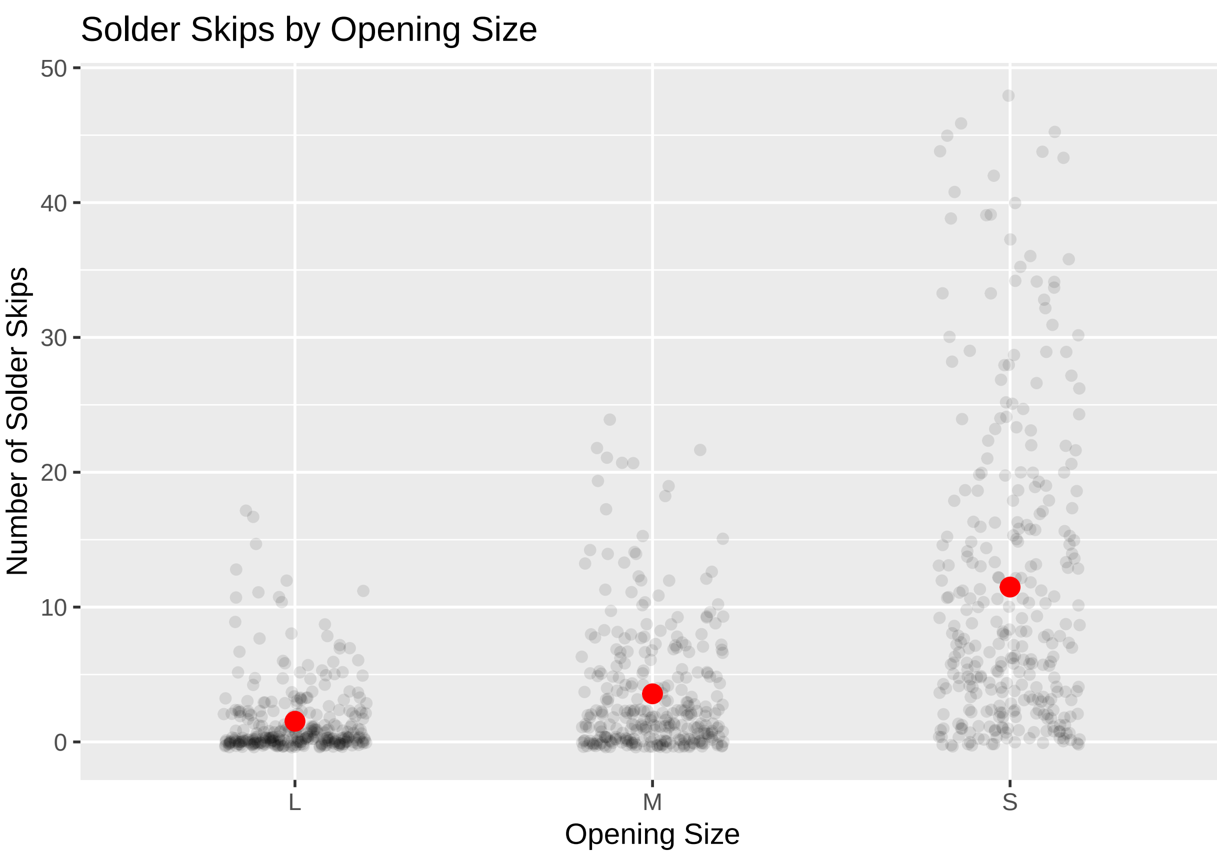
UTEID: kk38378

Github Link:

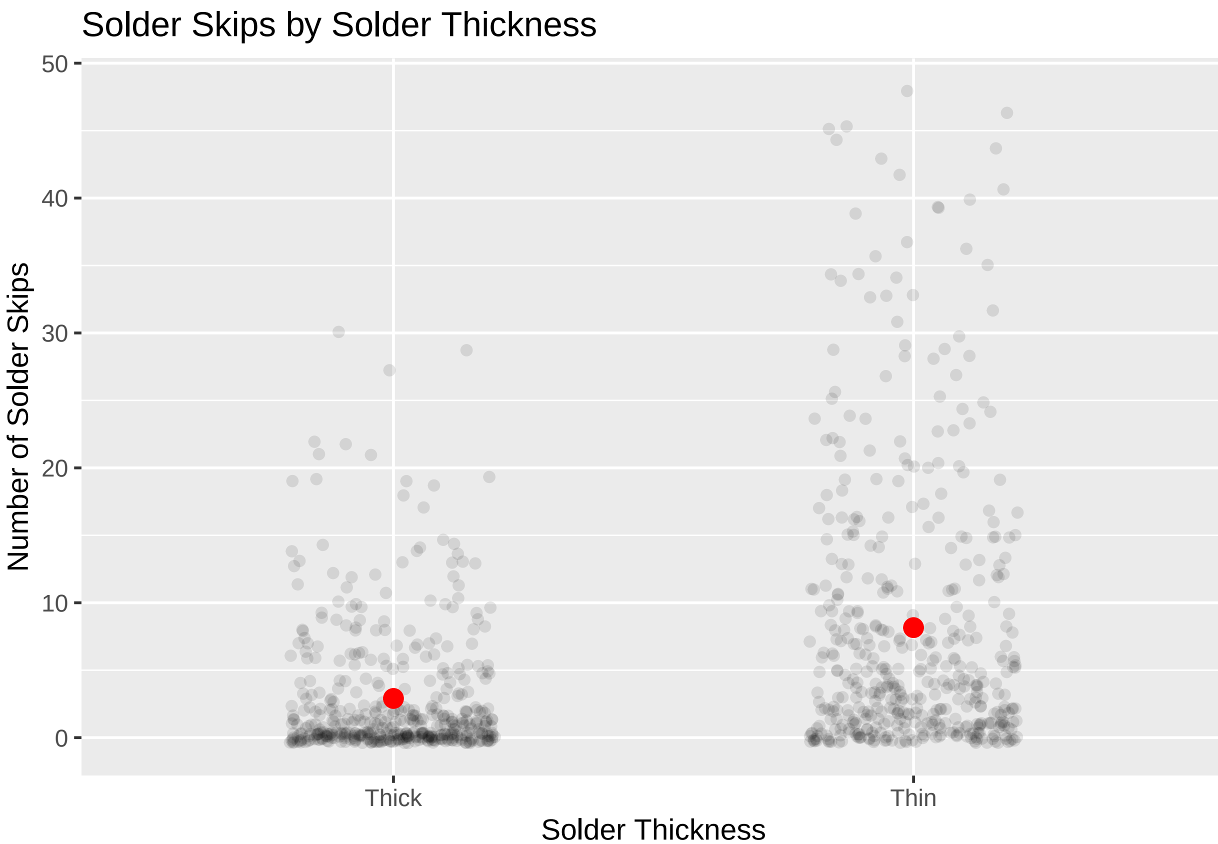
https://github.com/karthikeyakasa/SDS313Homework8

Problem 1

Part A



This plot shows the number of solder skips by opening size, with red points indicating the average for each group; smaller openings are associated with more skips on average.



This plot shows the number of solder skips by solder thickness, with red points marking group averages; thin solder is associated with more skips on average.

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

y =0.393+2.407*OpeningM+5.127*OpeningS+2.280*SolderThin-0.740*(OpeningM*SolderThin)+9.653*(OpeningS*SolderThin)

The baseline number of skips is 0.393, which corresponds to boards manufactured using a large opening and thick solder.

The main effect for OpeningM is 2.407. This means that, when thick solder is used, boards made with a medium opening have, on average, 2.41 more skips than those made with a large opening.

The main effect for OpeningS is 5.127, so boards with a small opening have 5.13 more skips, again assuming thick solder.

The main effect for SolderThin is 2.280, indicating that thin solder adds 2.28 skips, when the opening is large.

The interaction effect for OpeningM and SolderThin is -0.740. This means that when both medium opening and thin solder are used, the total effect is 0.74 skips fewer than what you would expect by simply adding their individual effects. However, this interaction is not statistically significant.

The interaction effect for OpeningS and SolderThin is 9.653, meaning that combining a small opening and thin solder results in nearly 10 additional skips beyond the additive effect of each factor alone. This interaction is large and statistically significant.

Part D

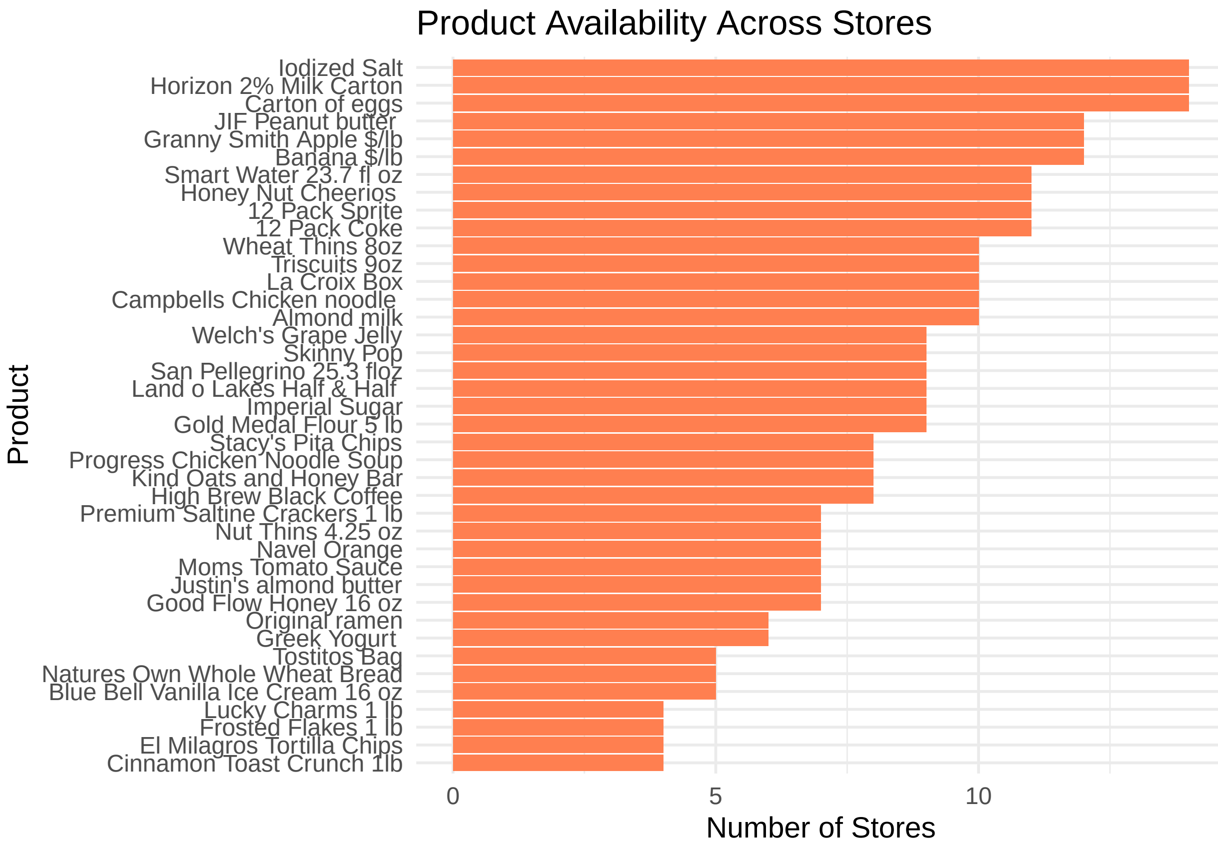
Based on the regression model, the combination of a large opening size and thick solder results in the lowest average number of solder skips—approximately 0.393 skips per board. This configuration serves as the baseline in the model and performs significantly better than any other combination. Using a medium or small opening increases the number of skips to around 2.80 and 5.52, respectively, even when thick solder is used. Switching to thin solder also raises the number of skips, with the large opening and thin solder combination yielding 2.67 skips. The interaction effects become especially pronounced when combining small openings with thin solder, resulting in a dramatic increase to 17.45 predicted skips. This suggests a compounding effect where these two design choices exacerbate each other's impact on manufacturing flaws. Therefore, to ensure the most reliable production process, AT&T should use a large solder gun opening with thick solder and avoid combinations—particularly small openings with thin solder—that significantly increase the likelihood of defects.

Problem 2

Part A

This graph shows that average product prices vary across stores, with some—like Whole Foods and Wheatsville—charging more on average than others like H-E-B and Walmart.

Part B



This graph shows how widely each product is available across the 16 stores, with staples like eggs and milk sold in nearly all stores, while others appear in far fewer locations. ## Part C

Compared with ordinary grocery stores (like Albertsons, HEB, or Krogers), convenience stores charge somewhere between \$0.41 and \$0.92 more for the same product, based on a 95% confidence interval.

Part D

The two stores that seem to charge the lowest prices when comparing the same product are Walmart and Kroger Fresh Fare, with average prices around \$0.90–\$1.00 less than the baseline.

The two stores that seem to charge the highest prices are Whole Foods and Wheatsville Food Co-Op, with average prices around \$0.30–\$0.36 higher than the baseline.

Part E

The data suggest that Central Market does charge slightly more than H-E-B for the same product, by about 7 cents on average. While this difference is statistically significant, it is quite small in the context of the broader price differences across stores. For instance, Walmart charges nearly \$1 less than the baseline, and Whole Foods charges over 30 cents more. So although Central Market has a premium image, its prices are only modestly higher than H-E-B's for the same items.

Part F

```
grocery_data <- grocery_data %>%
  mutate(
```

```
    Income10K = Income / 10000,
    Income10K_std = scale(Income10K),
    Price_std = scale(Price)
  )
model_income <- lm(Price_std ~ Product + Income10K_std, data = grocery_data)
```

The coefficient on Income10K_std is –0.032, which means that consumers in higher-income ZIP codes tend to pay slightly less for the same product, on average. However, the effect is not statistically significant at the 5% level ($p = 0.144$), so we can't rule out that this is due to chance.

In terms of effect size: A one-standard deviation increase in ZIP code income is associated with a 0.032 standard-deviation decrease in the price of the same product. This is a very small effect, and given the p-value, it likely isn't meaningful in practice.

Problem 3

Part A

True because there is a strong indication of a positive correlation between minority percentage and FAIR policy take-up. Figure A1 has an obvious upward trend, which means ZIP codes that have a higher minority percentage also have more FAIR policies per 100 housing units. This graphical confirmation is corroborated by Model A, in which the minority coefficient is 0.014 with a p-value of 0.000, indicating the result is statistically significant. In addition, R^2 is 0.516, indicating over 50% of FAIR policy variation is described by minority percentage alone. This supports the hypothesis that ZIP codes with larger minorities have worse access to private coverage, as revealed by greater use of the FAIR plan.

Part B

False because there is no evidence presented of an interaction effect between housing age and minority percentage in predicting FAIR policy take-up. Figure B1 simply shows the correlation between housing age and minority percentage, and not both of their combined effect on policies. Model B regression confirms that while the coefficient on the age of housing is positive (0.398), it is not statistically significant ($p = 0.125$), and R^2 is very low (0.061). Moreover, there is no model that includes an interaction term between minority and age that would be necessary to test this effect. The statement is therefore not substantiated by the data and must be marked false.

Part C

True since this is supported both graphically and statistically. Figure C1 shows that the positive relationship between minority percentage and FAIR policy take-up is visibly steeper for high-fire-risk ZIP codes compared to low-fire-risk ZIP codes. Model C, which includes an interaction term between minority and fire_risk, suggests the baseline slope for high-fire-risk ZIP codes is 0.01 ($p = 0.015$). Although the interaction term for low fire risk (minority:fire_riskLow) is not statistically significant ($p = 0.839$), graphical evidence is supportive of the argument that the relationship is more evident in high-risk neighborhoods. This suggests the effect of race on access to private insurance is potentially even more marked in riskier neighborhoods.

Part D

False because adding income to the model reduces the effect of minority percentage but does not eliminate it. The minority coefficient is 0.014 ($p < 0.001$) in Model D1 when it is the only predictor. When income is controlled in Model D2, the coefficient reduces to 0.010 but remains statistically significant ($p = 0.002$). This means that income partially mediate the effect, yet there remains a distinct effect of minority percentage on FAIR policy take-up. Income doesn't therefore "explain away" the entire relationship — the assertion is false.

Part E

True since Model E adjusts for all the covariates — income, fire risk, housing age, and minority percentage — and still reports a statistically significant relationship between minority and FAIR policy take-up. The minority coefficient is 0.008, p-value 0.006, showing the effect remains after adjustment for potential confounders. This means ZIP codes with higher minority proportion remain less likely to adopt private insurance, irrespective of other known risk factors. The $R^2 = 0.662$ also shows that this multivariable model explains a great deal of variation in FAIR policy take-up. So, this statement is true.