

Exercise 1

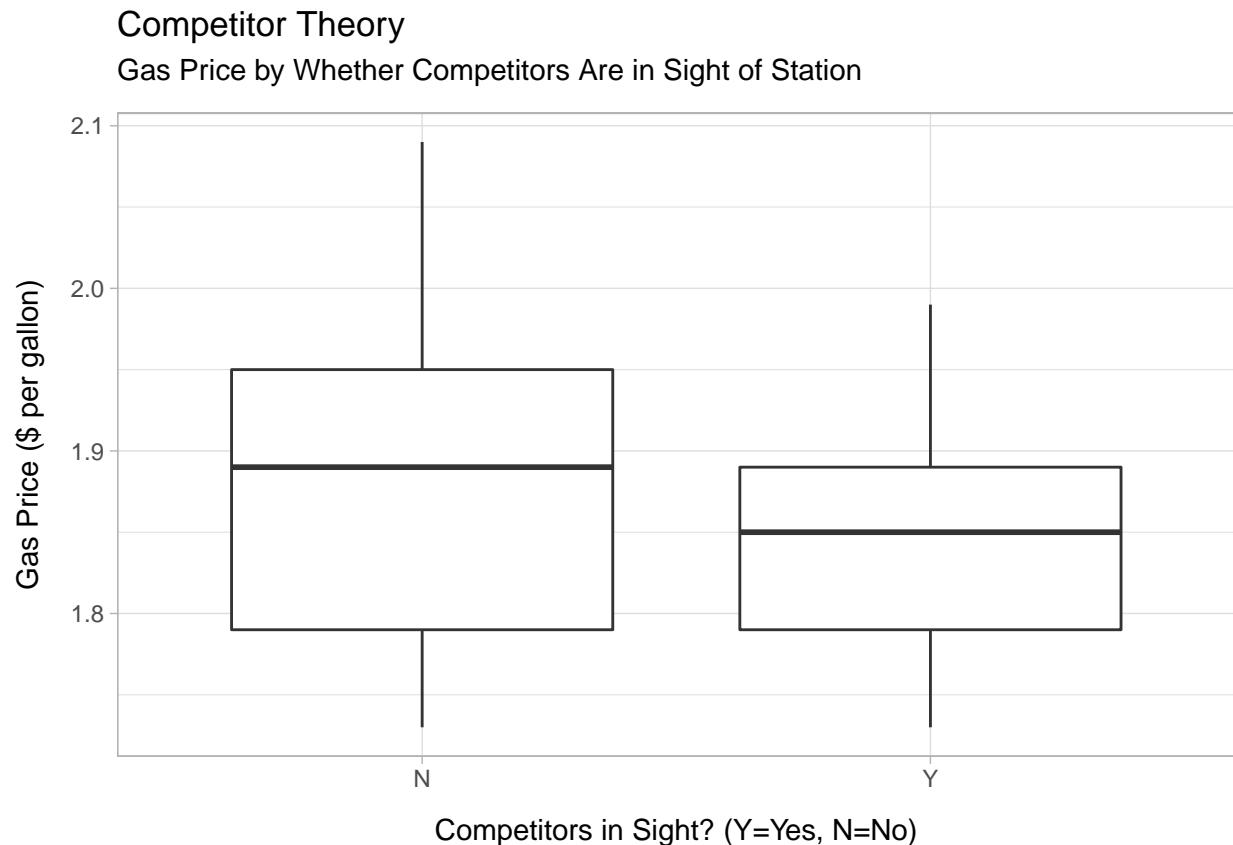
Robert Toto

2/08/2021

1) Gas Prices

In this problem, we assess five possible theories about consumer gas prices in Austin, TX using 2016 price data from 101 gas stations in the city. The theories consider five different factors that may drive gas prices: proximity of competitors, income of local neighborhood, brand name, proximity of a stop light, and proximity to a highway.

A) Competitor Theory: Gas stations charge more if there is no competition in sight

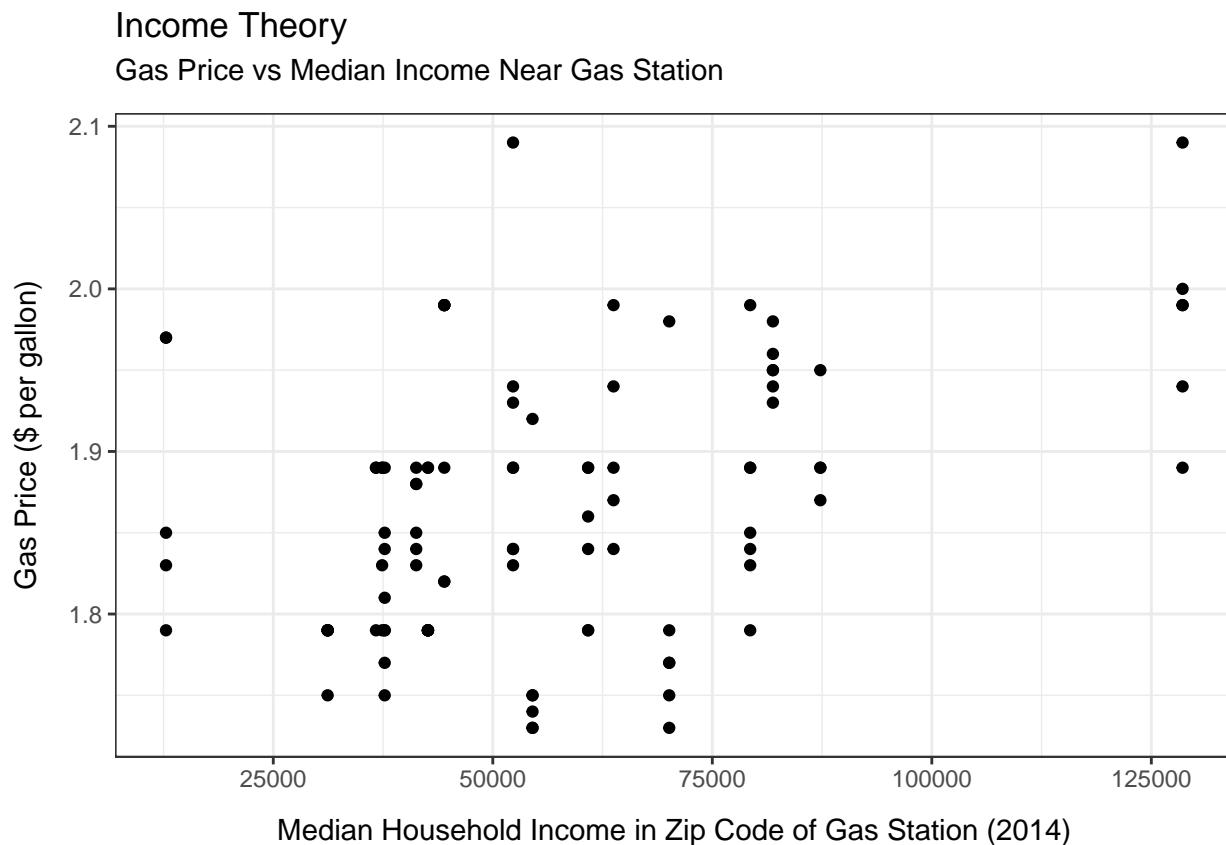


Claim: If a driver has more than one option for a gas purchase, she will minimize her expenses by choosing a station with lower prices, all else held equal. If the driver can physically view more than one gas station, the cost of optimizing (minimizing) expenses drops essentially to zero because she does not need to research alternatives to a single station. She has full information on more than one from physically viewing them.

Gas stations themselves would understand this and lower their prices to compete with nearby (viewable) competitors. However, if no competitor is nearby, a station has a localized monopoly and will charge a greater mark-up. The boxplot below attempts to validate this monopoly pricing theory by comparing the distributions of prices set by stations when a competitor is within view to the distribution when no competitor is viewable.

Conclusion: The boxplot shows a clear visual difference in the two distributions. When a competitor is **not** in sight, the median price a gas station charges is \$0.04 higher per. The boxplot further demonstrates that the upper limit of pricing is much greater for the local monopolies (no competitors in sight). For these stations, the top 25% of prices fall between \$1.95 and \$2.09, whereas the top 25% of prices for stations with visible competition is entirely beneath \$2.00. One can see that monopoly pricing has a longer upper range, higher maximum, and overall wider distribution, demonstrating that local monopolies have much greater pricing power and price-setting flexibility. The boxplot certainly lends visual credibility to the competitor theory. One can expect to pay lower prices when there is direct competition.

B) Income Theory: The richer the area, the higher the gas price



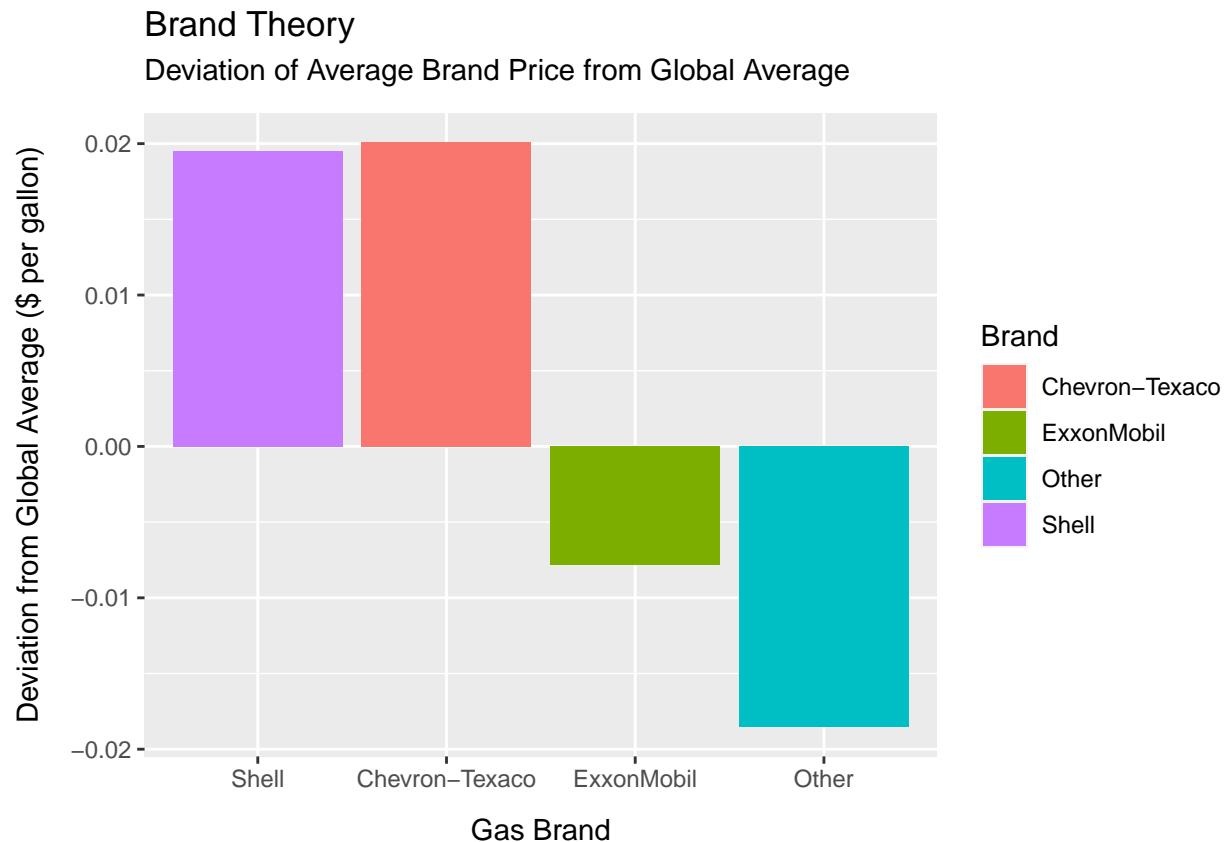
In theory, a neighborhood with higher incomes would have higher local gas station prices because people with greater expendable income have a higher market willingness-to-pay for basic goods, like gasoline, since these goods constitute such a small portion of their overall income. Conversely, low-income gas stations would have lower mark-ups and overall prices because people with lower incomes are more sensitive to the prices of basic goods, like gasoline, since these goods constitute a large portion of their overall income and since they may substitute away from some of them. For instance, if a gas station charges a very high price in a low-income neighborhood, local residents are likely to substitute away from personal transportation to public transportation—a cheaper option.

The scatterplot validates the theory because it demonstrates that, in general, as the income of a neighborhood

rises, so do the gas prices. The relationship is not perfectly linear, but there is a noticeable trend. If one were to draw an OLS line through the plot, it would have a positive slope. The data at the very high range of incomes is especially interesting. For neighborhoods with median incomes greater than \$125,000, the average gas price is never below \$1.85 per gallon, which is about the median price for all average prices from neighborhoods with median incomes under \$100,000. Interestingly, at incomes below \$25,000, there is no trend. Of the four data-points in this range, average gas prices can rise above \$1.95 per gallon, well above the median of the overall data. It would be interesting to gather more data on gas prices exclusively in lowest-income neighborhoods to see if this lack of trend holds.

C) Brand Theory: Shell charges more than other brands

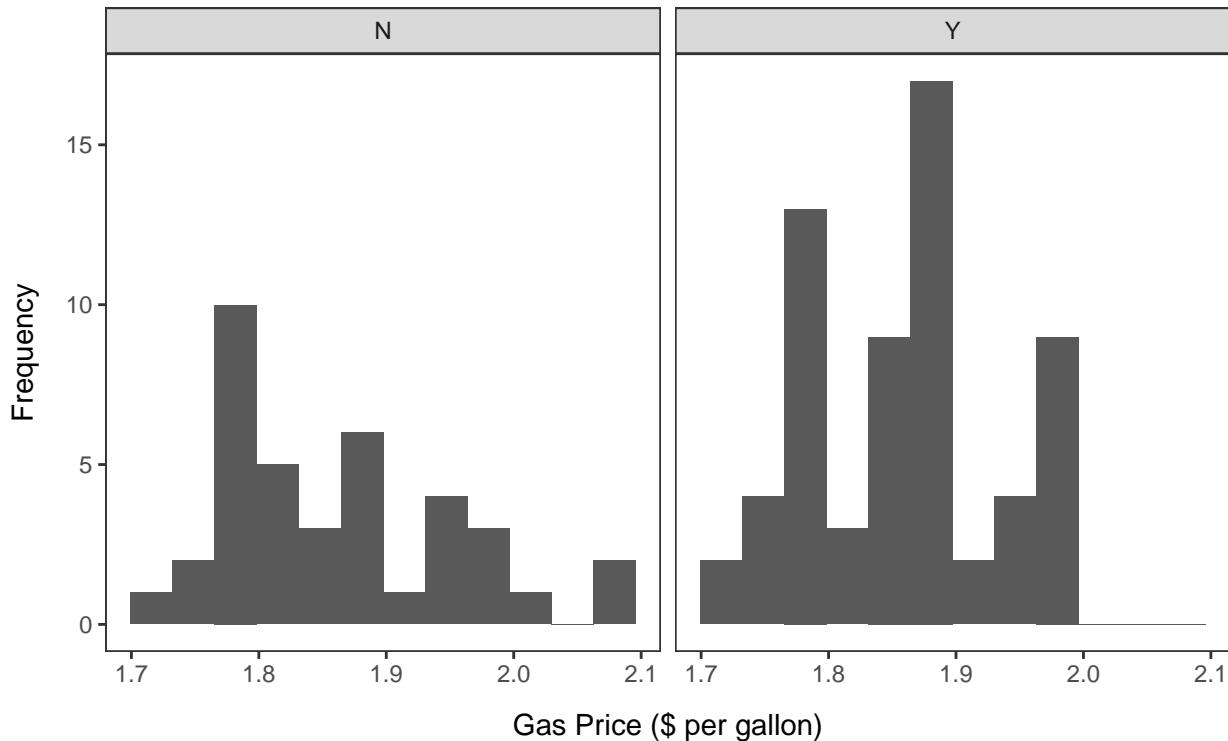
```
## `summarise()` has grouped output by 'Brand'. You can override using the '.groups' argument.
```



D) Stoplight Theory: Gas stations at stoplights charge more

Stoplight Theory

Gas Prices by Whether a Stoplight Is in Front of Station

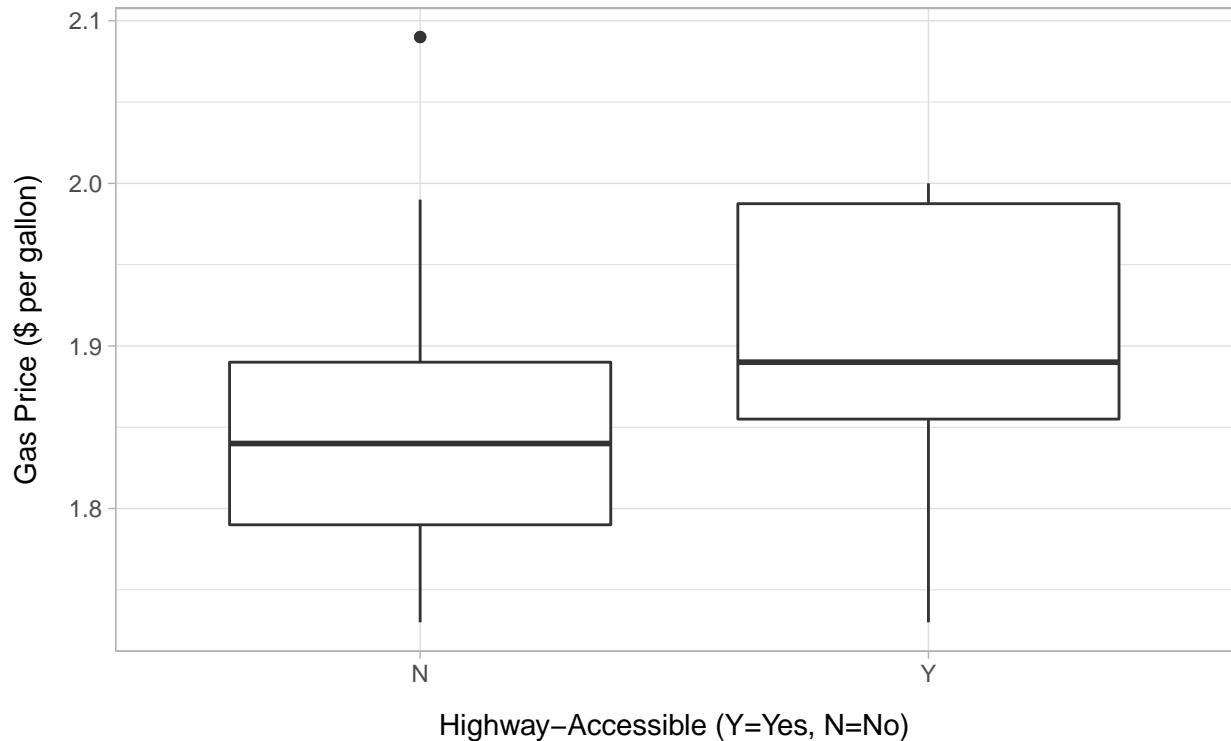


```
## # A tibble: 2 x 2
##   Stoplight mean_price2
## * <chr>          <dbl>
## 1 N              1.87
## 2 Y              1.86
```

E) Highway Theory: Gas stations with direct highway access charge more

Highway Theory

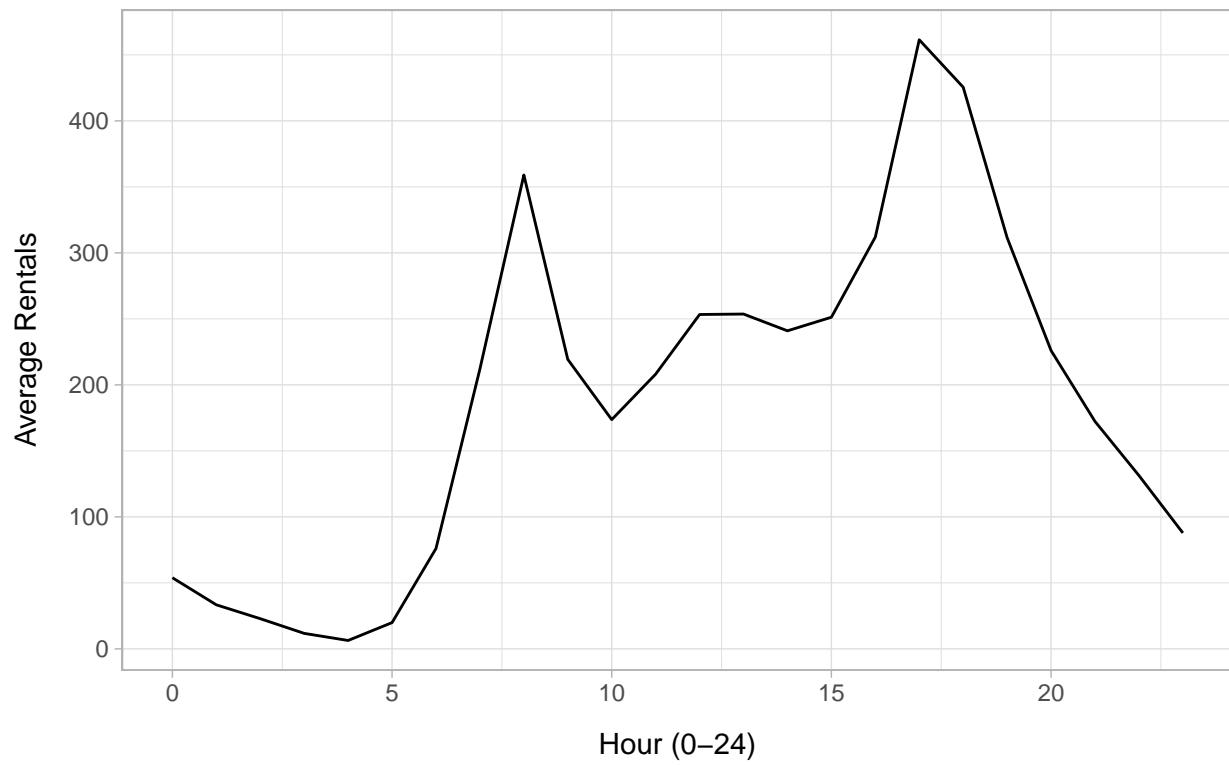
Gas Prices by Whether a Station is Accessible from Highway



2) Bike Share Network

Text...

Average Bike Rentals by Hour

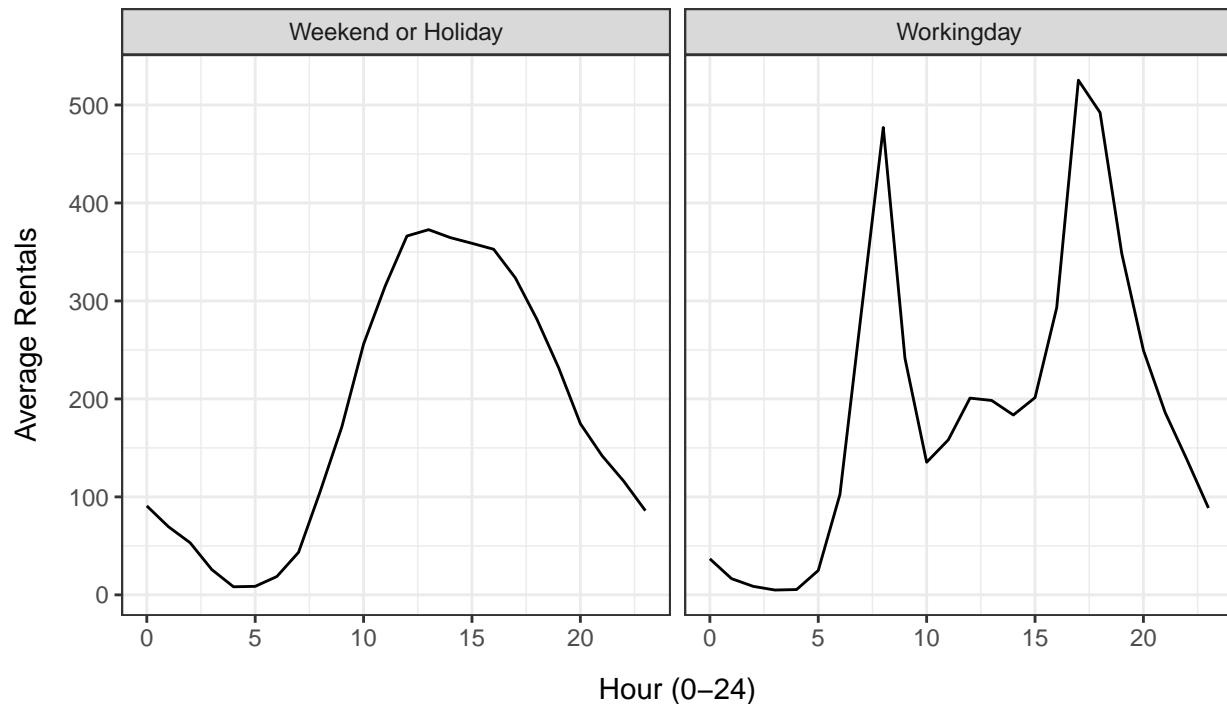


Source: Capital Bike Share

```
## `summarise()` has grouped output by 'hr'. You can override using the '.groups' argument.
```

Rentals by Type of Day

Average Hourly Rentals by Whether it is a Workday



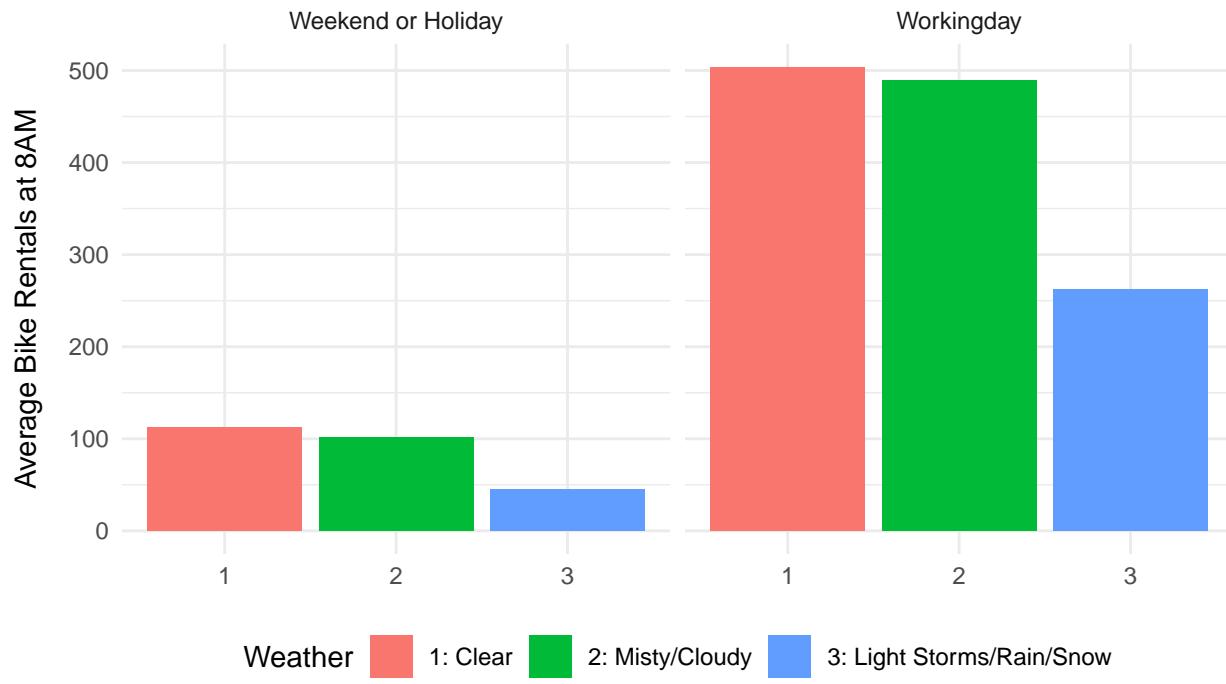
Source: Capital Bike Share

```
## `summarise()` has grouped output by 'weathersit'. You can override using the '.groups' argument.
```

```
## # A tibble: 6 x 3
## # Groups:   weathersit [3]
##   weathersit workingday mean_rentals
##       <int>      <int>      <dbl>
## 1         1          0     112.
## 2         1          1     504.
## 3         2          0     102.
## 4         2          1     489.
## 5         3          0     45.1
## 6         3          1     263.
```

Morning Bike Rentals

Average Bike Rentals at 8AM by Weather and Working Day



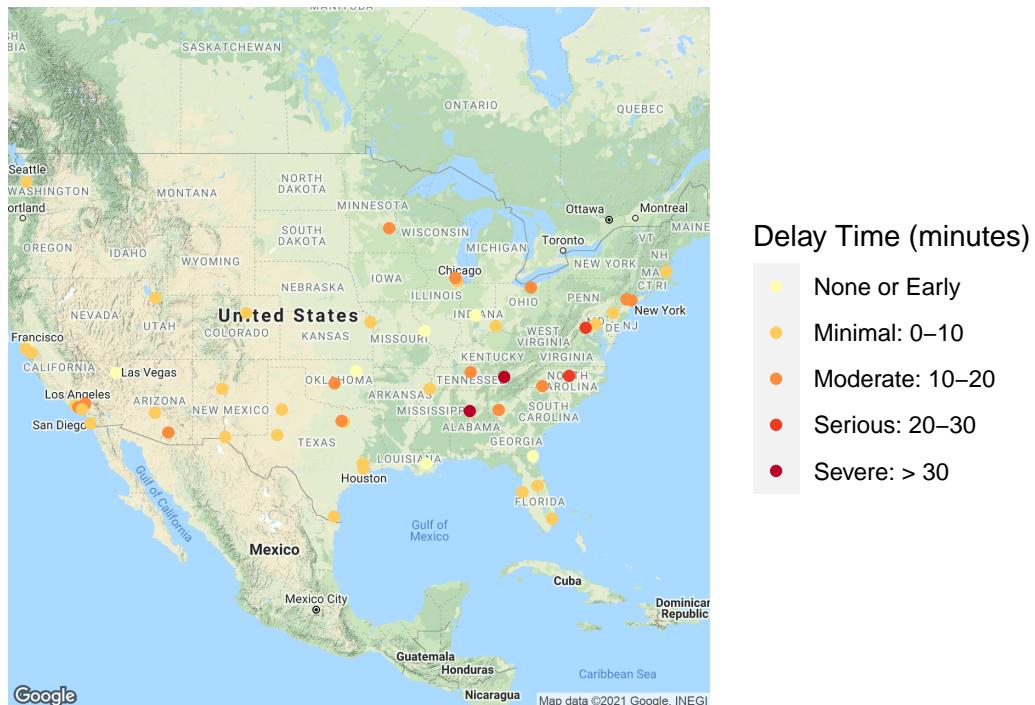
Source: Capital Bike Share

The bike rentals are much higher for workingdays since the data only shows rentals occurring at 8AM. This reflects the fact that on workingday, people use bike rentals to go to work. However, on weekends or holidays, while overall rentals may be more active due greater recreation, people are not renting bikes so early in the morning. As shown in the previous line graph, on weekends and holidays, people mainly rent bikes at or after noon.

3) Flights at ABIA

Geographic Delays

Average Delay for Inbound Flights to ABIA from US Airports

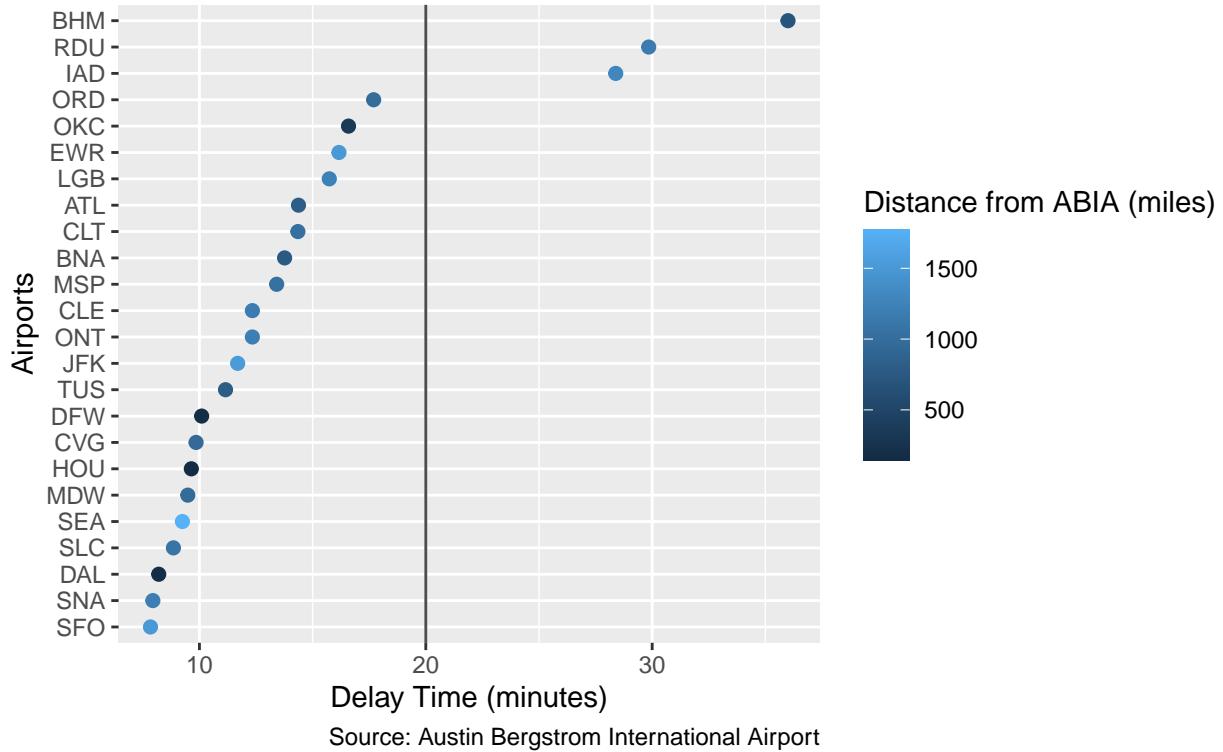


Source: Austin Bergstrom International Airport

Text...

Airports with Longest Delays

Rank of Airports with Longest Delays for Flight Arriving to ABIA



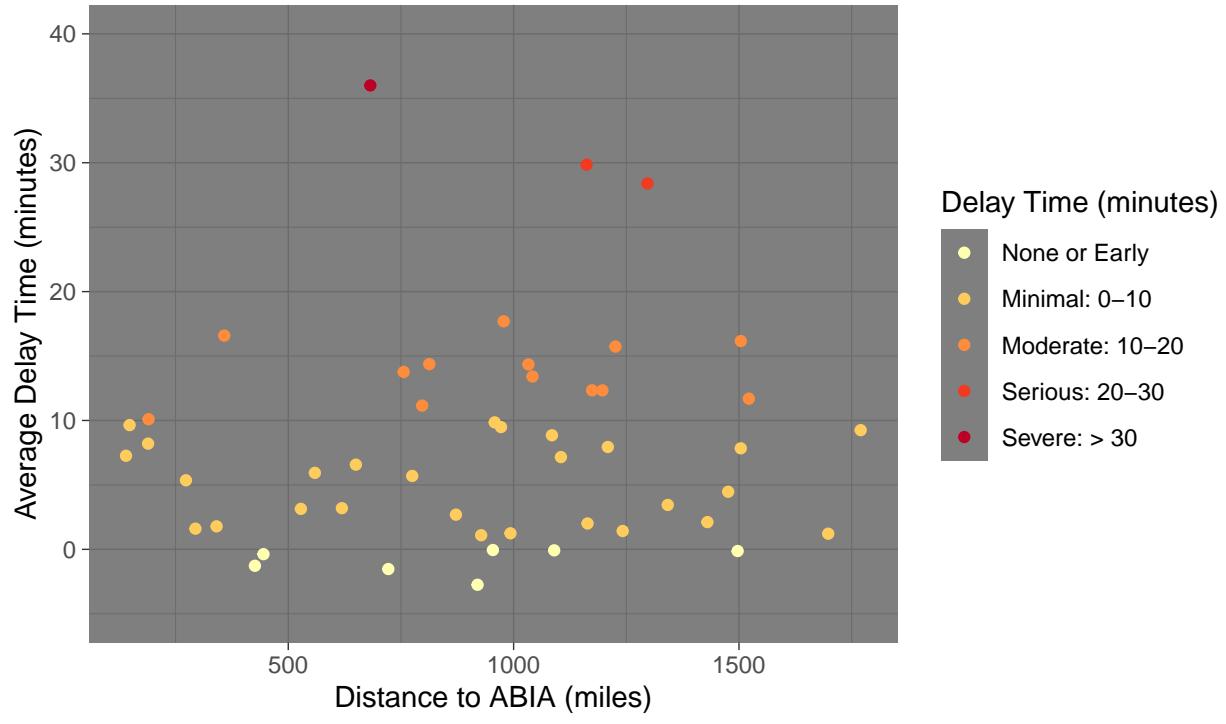
Source: Austin Bergstrom International Airport

This graphic gives the average arrival delay by Airport for flights inbound to Austin Bergstrom International Airport. The data show that most average arrival delay is below 20 minutes for most airports, which shows that most of the flights incoming to Austin are well-organized and without complication. Interestingly, there does not appear to be a relationship between arrival time and distance. For instance, the average delay for flights from Dallas (DFW) to Austin is greater than for flights coming from San Francisco (SFO) even though Dallas is only 189 miles away, while San Francisco is over 1,000 miles away. Furthermore, there appear to be outlier airports that suffer from much greater arrival delays—the three airports at the top of the list for which the average delay is well beyond 20 minutes. These airports are Dulles International Airport in Virginia (IAD), Raleigh-Durham International Airport in North Carolina (RDU), and Birmingham-Shuttlesworth Airport in Alabama (BHM). Notably, BHM is by far the closest of these three outliers but has the greatest average delay (about 37 minutes per flight). Overall, the driver of delays seem to have far more to do with airport management or weather than distance. One could surmise that SFO is very well-managed, while BHM is poorly-managed. Alternatively, it is possible that the airline is responsible. Most nonstop flights from SFO to AUS are carried by United Airlines, whereas most nonstop flights from BHM are carried by American Airlines. Looking just at Dallas, there are two airports, DFW and DAL. Flights from DFW, which are all American Airlines flights have a 10 minute average delay, whereas flights from DAL, which are all Southwest Airlines flights have half the delay (5 minutes). So while the distance is effectively the same, the main difference driving the delay variation seems to be airline. Some airlines may be better at optimizing flight schedules than others, more effectively minimizing delays.

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Does Distance Affect Arrival Delay?

Average Arrival Delay to IBIA by Distance from Departure Airport



This scatterplot confirms that there is no relationship between flight distance and average delay, discussed in the previous graph. Most airports have only a minimal delay, and these airports are homogenously scattered by distance from 0 to 1,500 miles from Austin. This is strong evidence that distance is not a major driver of delay. The same homogenous characterization holds for airports with moderate delays, which are scattered evenly across the distance axis. In fact, the only airport with severe delays on average, Birmingham-Shuttlesworth Airport in Alabama (BHM), is below the median distance for all airports. Along with the previous graph, this is good confirmation that factors other than distance are the most important for delays, including airport and airline management and optimization. Since different airports are hubs for different airlines, carriers are not homogenously distributed across all flight paths incoming to Austin. As we saw by looking at Dallas, American Airlines seems to be worse at minimizing delays than Southwest Airlines for flights from Austin to Dallas. This is likely because the incoming flights to Dallas that will carry the flight to Austin are coming from different original cities in both cases. Possibly the origin city has more delays into Dallas for the American Airlines schedule than the Southwest Airlines schedule.

4) K-Nearest Neighbors

Text...