Homework 2: One-way ANOVA

Due Friday, March 13 (spooky)

Car Insurance Rate Scenario

About the data: Consumer Reports surveyed the top 9 largest banks in six cities (Atlanta, Chicago, Houston, Memphis, New York, and Philadelphia) and reported their annual average interest rates (%) for new-car loans. Data are found in **Insurance** in the HW 2 folder.

About the problem: A new (and frankly horrible) federal law has been passed that requires all city banks to have the same interest rates as other city banks. Any cities that are statistically significantly different from each other are both fined $5 million each time they are different from another city and an additional $2 million for each time they are different from another city. *Read: Both cities are being fined, not just the city with the higher insurance rate.* However, the overall process of determining which cities are fined has not yet been developed. Your goal as a government contracted statistician is to:

1) (2 pts) Visually show the differences in the cities however you feel is most appropriate.
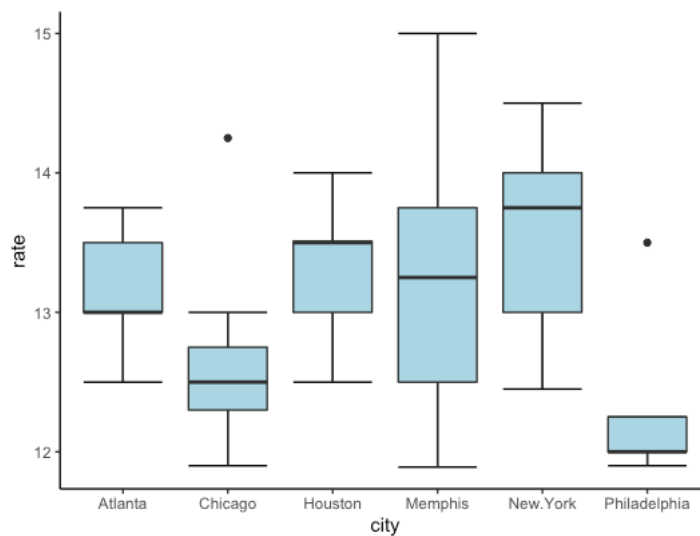
2) (2pts) Determine which cities are statistically significant from one another. **Important note**: feel free to base your analysis and methods on your own opinions on the relationships between banks, the government, and the potential effect on the city population (i.e. Tukey vs Bonferroni vs False Discovery Rate, changing alpha, creating arbitrary thresholds based on the data or based on attributes of the city, etc). Of course, you won't be graded on your

opinion, but do not roll through this assignment without giving any thought to your process. Also, please **state which post-hoc correction method you decide to use.**

After running an ANOVA on the data, the result was there was significant difference among the cities. We ran Tukey's Honest Significant Difference correction method to determine which pairs of cities had a statistically significant difference. Tukey's HSD showed that Philadelphia is statistically significantly different from Atlanta, Houston and Memphis.

> diff   lwr   upr   p adj

Philadelphia-Atlanta -0.99444444 -1.93640440 -0.05248449 0.0329734

Philadelphia-Houston -1.10666667 -2.04862662 -0.16470671 0.0127365

Philadelphia-Memphis -1.04444444 -1.98640440 -0.10248449 0.0217928

Philadelphia-New.York -1.28333333 -2.22529329 -0.34137338 0.0024794

3) (2pts) Create a table with the number of times a city is fined, how much they are fined and any other information you feel is necessary to report.

City	Number of times fined	Total fine
Atlanta	1	$5,000,000
Chicago	0	$0
Houston	1	$5,000,000
Memphis	1	$5,000,000
New York	1	$5,000,000
Philadelphia	4	$5,000,000 +$2,000,000 x 3 = $11,000,000

4) (4 pts) Write 1 to 2 paragraphs on your process, why you chose the test you chose, and interpret the results from the tests in layman's terms. Again, be sure to reference which post-hoc correction method you decided to use.

We first trimmed the data to remove empty entries, since the original dataset had numerous empty cells and only the first nine entries had data. We then translated the insurance data from a wide format to a long format, pairing the city to the nine rates from that particular city. Changing the format allowed us to display the data with boxplots. Using the long format, we were also able to run an ANOVA on the data because we could tell the AOV function in R that the "rates" column was a function of the "city" column to appropriately conduct the analysis. Once looking at the summary of the ANOVA, we saw the p-value was 0.00117, which is less than 0.05, indicating that there were significant differences in the rates of the different cities, however, it was unclear which cities were different from one another. We used Tukey's Honest Significant Difference to see which cities were different from other, because it is a moderate post-hoc correction method that is not too liberal or too strict, and it also provides useful additional information like the numerical differences between the cities. In a real life, the analysis may have used a

Name: Emily Lukshin and Cole Akin

Bonferroni Correction, because it is a very conservative comparison method that would better ensure that there really is a statistically significant difference between two cities, warranting a large fine. When we ran the Bonferroni Correction, the pairing of cities that Tukey's HSD reported as different were also reported as different (p-value < 0.05) using Bonferroni. However, it is worth noting that the Philadelphia-Atlanta pairing had a p-value of 0.0442 using Bonferroni, indicating that the difference is slight and possibly with a larger sample of rates, there would not be a difference and therefore no fine for the cities.

Our conclusion is that we are 95% confident that Philadelphia's bank interest rates are statistically significantly different from the bank interest rates in Atlanta, Houston, Memphis, and New York. According to the federal law, these differences mean that Philadelphia would be fined $11,000,000 and Atlanta, Houston, Memphis, and New York would each receive a $5,000,000 fine.

Please answer the following questions for your homework to be graded:

A) How did you and your partner(s) split the work on this assignment?

We worked together through the problems, both contributed to the answers we provided, and talked through the problem, but answered in our own words.

B) What did you find difficult in this assignment?

I think the hardest part was trying to accurately describe what was going on in the math/code in a way that made sense and was appropriate to what was happening.