Process Book

Group 1: Mackenzie B., Jake H., Tony L., Alicia A.

**Overview and Motivation**

Our data analysis topic focuses on crimes that were reported in Durham County. The dataset is an excel file that was retrieved from the Durham County website. The excel file has a list of all reported crimes that happened since 2013. Various information about the crime was recorded such as: The longitude/latitude that the crime occured, time/date of occurence, type of crime, and zone. Our goal is to construct a shiny app that displays where crimes are most frequently committed in Durham County, what hour crimes most frequently occur, and types of crimes that are more likely to happen during certain times of the day. The shiny app will have interactive features that allows users to select different months of the year and see a map of crimes that occured in that month.

Our goal for this project is to inform people about locations in Durham that are most prone to crimes and least prone. Durham has many popular restaurants that attract Chapel Hill and Duke students, so informing the public about potential crimes that could happen where they are is a good idea.

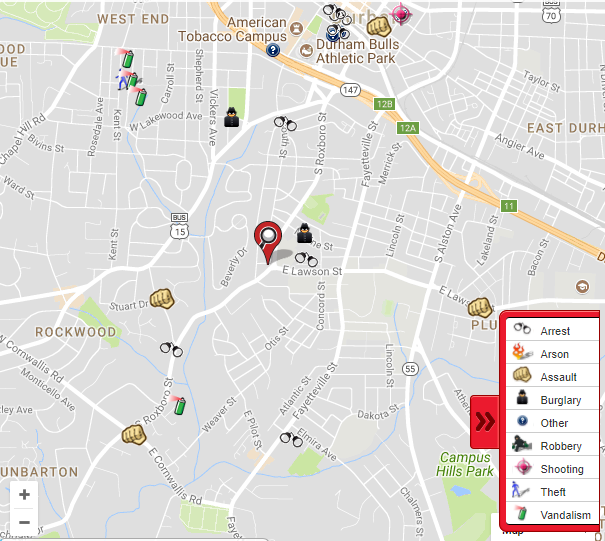
Some of the other projects we considered were:

* Using past trends of lottery numbers to predict future drawings
* Cost of living for the 100 counties of North Carolina: Which counties have the highest wage to housing cost ratio, and which counties have the lowest?

**Related Work**

Interactive crimes map released by WRAL: <<http://www.wral.com/news/local/page/5561108/>>

* On a map of Durham, users can click and drag to certain streets and click on points of crime. Different symbols represent different crimes. Some of the crimes included in the map are burglary, theft, and assault. Clicking on these symbols returns information on the date and time that the crime was committed.



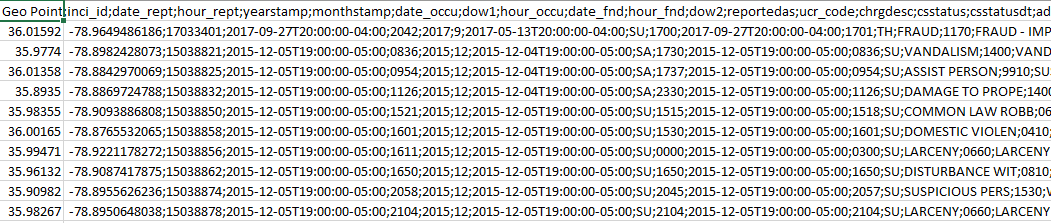
Due to time constraints, we wanted to create a simplified variation of this project. One thing that the app by WRAL does not include is a total monthly view of all the crimes that happened in Durham. Also, as data scientists, it would be useful to look at whether certain crimes are more prone to being committed at certain times of the day.

**Initial Questions**

Original questions that we started off with:

1. At what time of day do most crimes occur?
2. Are there specific types of crime that occur most at certain types of the day? (ex. Do robberies mostly happen late at night?)
3. Where in Durham, location-wise, do crimes most frequently occur?
4. Are some areas more prone to specific types of crimes, and if so which areas, and which crimes?
5. How we can use our analysis to make durham law enforcement and citizens more aware of potential crime?
6. Are certain crimes seasonal?

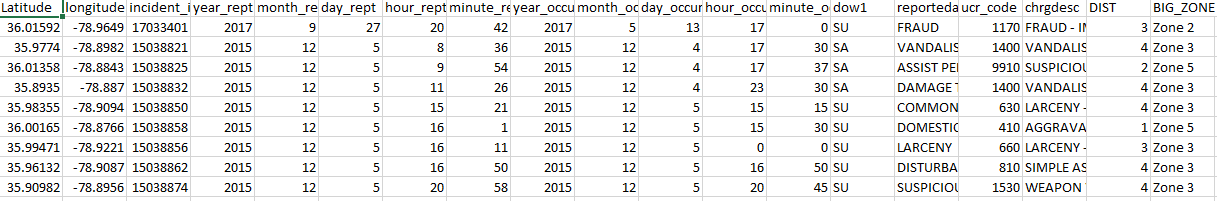
The original dataset that we used was very messy. Here is a screenshot:



Using the tidyverse and tidyr package, we cleaned the dataset by doing the following:

* *Separated* the columns that were combined
* *Selected* relevant columns that we want to use to describe our dataset: latitude, longitude, incident ID, year/month/day/hour/minute reported, year/month/day/hour/minute occured, day of the week, crime, crime code, charged as, district, and zone
* *Filtered* out crimes that did not occur in 2016
* *Filtered* out crimes that occurred outside of Durham County
* Exported the cleaned dataset as a .csv file

Here is a screenshot of the tidy dataset:



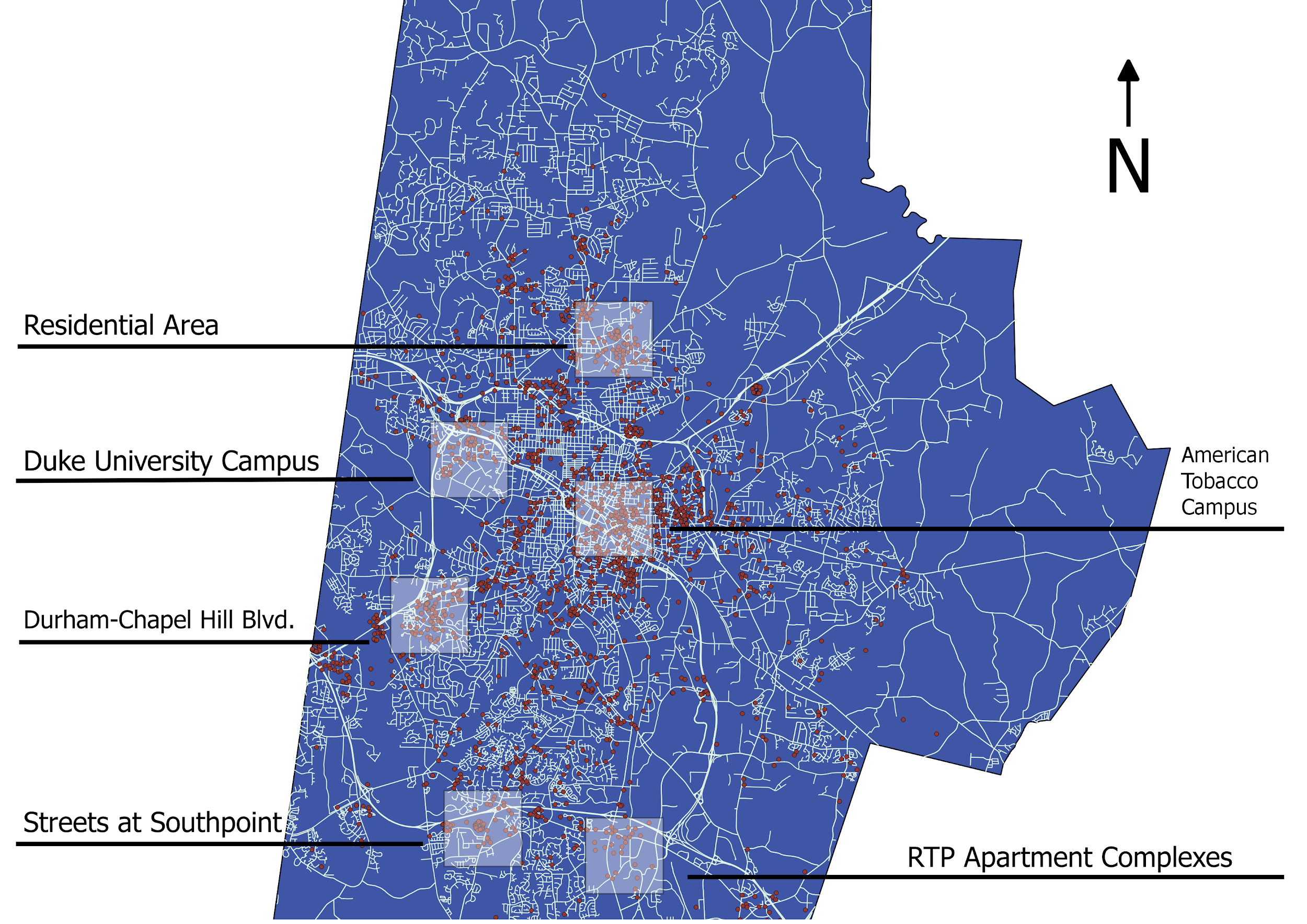
From looking at basic plots, we found that question 4 was infeasible. Since the latitude and longitude area is so small, it is difficult to analyze if certain areas are more prone to certain crimes. Instead, we altered the question to whether certain areas have more crimes, in general, than other areas. Also, for question 6 we found that there was no seasonal variation in number of crimes committed, but rather, they varied by months.

So, the final questions that we are answering are as follows:

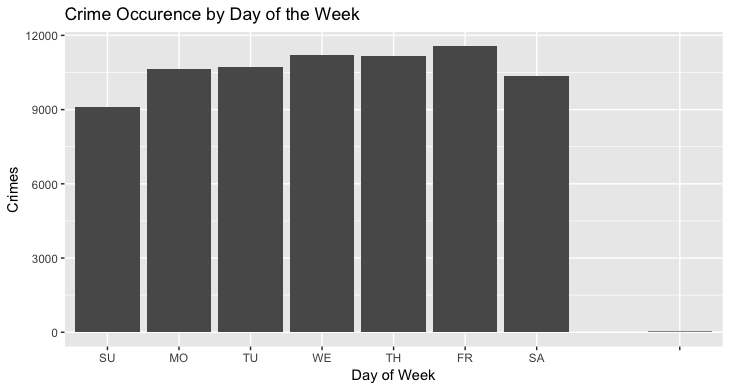
1. What parts of Durham are more prone to high crime rates? Which areas are the “safest” parts of Durham?
   1. In the shiny app, there is a map of Durham that shows all the points of crime for 2016, and users can use a slider bar to choose different months that they want to view. Areas of interest (Duke University, American Tobacco Center, Streets of Southpoint, etc.) are highlighted and labeled on the map.
2. At what time of the day do crimes frequently occur?
   1. In the shiny app, there is an interactive section where you can select different crimes and see a graph for up to 6 different crimes for each hour of the day. When you click on different hours of the graph, you can see counts for that crime for 2016.
3. Are there certain types of crimes that occur most frequently at certain times of the day?
   1. In the shiny app, there are panels of different crimes that you can choose. In each panel, you can view the graph of total counts for each hour of the day.

**Exploratory Data Analysis**

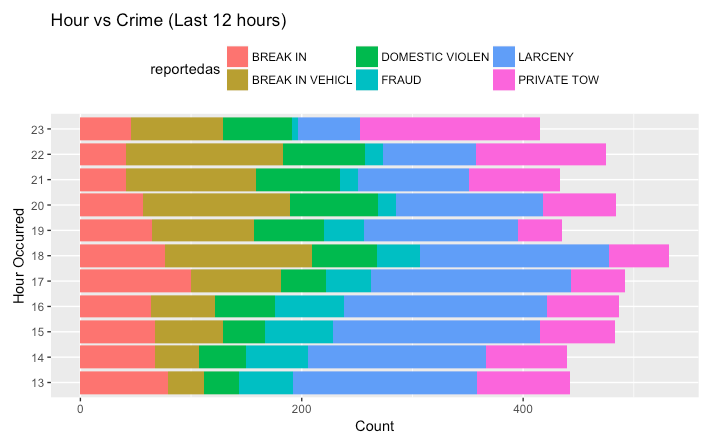
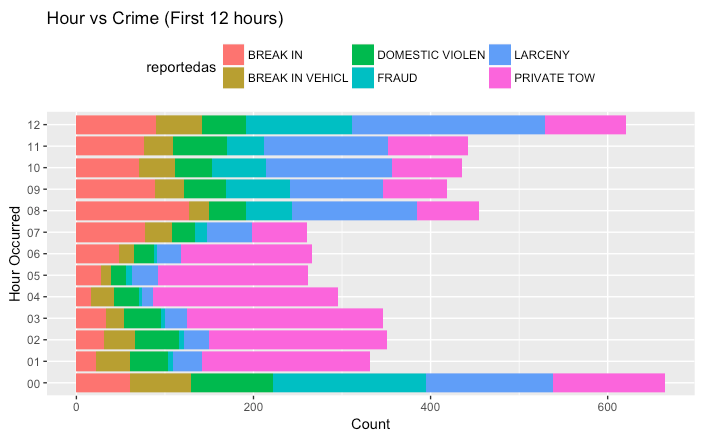
*Figure 1:* Using the software application QGIS, 12 different maps were created for each month of the year. It was clear that March, April, and November had the most crimes. So, there isn’t really a seasonal trend for number of crimes committed. From the map, you can see that crimes occur mostly around the Downtown/American Tobacco Campus area. The Duke University area has a little to medium amount of crime. The Streets at Southpoint are relatively safe, and RTP has very few crimes reported.



*Figure 2*: The plot shows crime counts for each day of the week, for all of 2016. While Friday is when most crimes are committed, Saturday has less crimes compared to weekdays and Sunday has significantly less crimes committed compared to all the other days. This is what you would expect: On Fridays, after work, more people go out to bars and restaurants. The fact that there are a lot of people walking around public areas at night is good reason to assume that there might be more crimes. Saturdays and Sundays are generally “family days”, which might be a reason for why these days have the least crimes, even though there’s a lot of people walking around town.

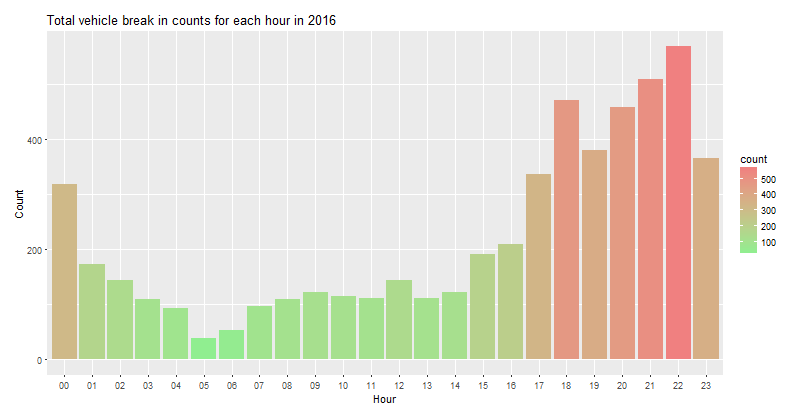


*Figure 3*: A summary of the top 6 most frequent crimes are shown for each hour of the day. The top 6 most frequent crimes are: Break-in home/business, break-in vehicles, domestic violence, fraud, larceny, and towing. Significantly less crimes are committed between 1am and 7am. Apart from those hours, the number of crimes are committed are relatively consistent throughout the day. The most crimes happen at midnight. However, it is important to note that for crimes where the time committed was unknown, the system might assign it to midnight by default.

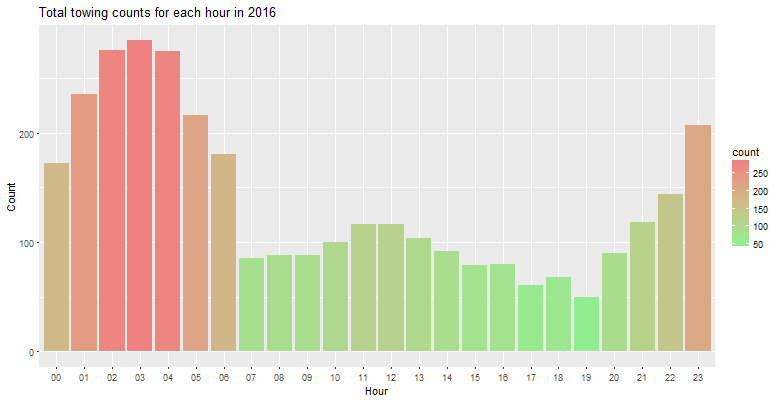


From figure 3, we can see that some crimes have trends that deviate from the total crimes trend. So, we decided to take a deeper look at individual crimes.

*Figure 3a.* Vehicle break ins occur disproportionately in the evening/night hours. Between 6pm and 10pm, the most break ins occur. This is usually the time when people are at home eating dinner, so it’s surprising that robbers would trespass someone’s property while they have a high risk of being caught. An explanation could be that cars are robbed when the owners are out somewhere public, and robbers take advantage of them being at a restaurant, grocery, theater, etc. for a long period of time. There are not many vehicle break-ins between 1am and 2pm.

**

*Figure 3b*: It is clear that towing happens much more frequently when it’s dark than when it’s bright outside. From 7am-8pm, the number of towings for each hour is almost half of that from 11pm-5am. When cars are towed during the daytime, people who are close to their illegally parked vehicles might see that they are getting towed and then quickly go retrieve it before the tower drives away. During the night, towing companies can take advantage of the fact that people who are illegally parked are likely to be asleep, so they can tow cars without anybody getting angry.



Other important figures are included in the technical document attached.

Important takeaway from graphs: In the analysis code, it was very clear that different types of crimes occur at different parts of the day. Vehicle break-ins occur mostly during dinner hours, towing happens mostly in the middle of the night, larceny happens during the middle of the day, and domestic violence cases happen at about the same rate throughout the day. Therefore, it would be difficult to do a regression analysis and create a model that predicts when and where crimes will occur next, since there are so many different crimes as variables. This is why we decided to create a shiny app. Rather than try to predict when and where the next crime occurs, the shiny app provides an interactive feature that allows users to select different types of crime and see the number of crimes committed throughout the day for that specific crime. The app also allows users to see what areas of Durham, overall, are relatively safe and relatively dangerous.

**Final Analysis**

What we learned from our data:

* Location
  + A large number of crimes happen in the downtown area/American Tobacco Campus. This might be attributed to the fact that it is an urban area, and many people travel by foot to get around the area. So, there is a lot of people-to-people interaction, which can explain why there are so many crimes.
  + Durham-Chapel Hill Blvd. (the road that continues off East Franklin St.) has quite a few reports of crime as well. This area has a lot of sit-down restaurants.
  + Duke University and the Streets at Southpoint have low to moderate crime rates.
  + The residential area west of RTP has a few crimes. There were very few crimes that were reported in the RTP area. This might be of interest to graduating seniors who have accepted full-time job offers in the area upon graduation.
* Day of the week
  + In 2016, there were about 2000 more crimes reported on Friday than Sunday. While most crimes occurred on Friday, fewer crimes happened on Saturday and Sunday than on weekdays. Sunday had the least amount of crimes.
* Hour
  + Different crimes tend to happen during different parts of the day. Here are some interesting ones:
  + Residential/business break-ins tend to happen during working hours, from 9-6. Vehicle break-ins tend to happen during the dinner hours, from 6-10.
  + Domestic violence happens pretty consistently at every hour of the day, with drops from 4-7am. A surprising amount happens during late night hours, from 1-3. Hours with the highest frequency of domestic violence are from 9-12.
  + Towing happens mostly in the middle of the night.
  + A majority of fraud cases are reported at midnight and noon. This trend might be because people don’t know the exact time that the fraud happened, so they report it at midnight or noon.
  + Most larceny happens during daylight hours, between noon and 6pm.

**Presentation**

* To present all of our information, we created a shiny app. Here’s a breakdown of the app:
* There are 7 different side panels to choose from
  + *Interactive crimes map*: Users can select months from 1-12 (1 representing January and 12 representing December) and view all the crimes for that month, in 2016. Since not everybody is familiar with the area, notable features are labeled such as: Duke University campus, the Streets at Southpoint, American Tobacco Campus, and RTP.
  + *Crimes per Hour*: Users can select different types of crimes, and whether they want to view the first 12 hours of the day (midnight-11am) or the last 12 hours of the day (noon-11pm). Upon selecting a crime, a graphical display is shown of the total number of crimes for each hour of the day, in 2016. A plot is also shown for six of the most frequent crimes.
  + *Break-ins*: Total number of break-ins for each hour of the day in 2016
  + *Domestic violence*: Total number of domestic violence for each hour of the day in 2016
  + *Towing*: Total number of towing for each hour of the day in 2016
  + *Fraud*: Total number of fraud for each hour of the day in 2016
  + *Larceny*: Total number of larceny for each hour of the day in 2016
* The last five categories are the highest frequency of crimes, so we thought it might be important to highlight trends for these crimes.