Cynthia Cho
DSC 465
Due: June 8, 2019

Homework 4

Problem 1a: (graphs are provided on the next page, so they are stacked on top for comparison purposes)

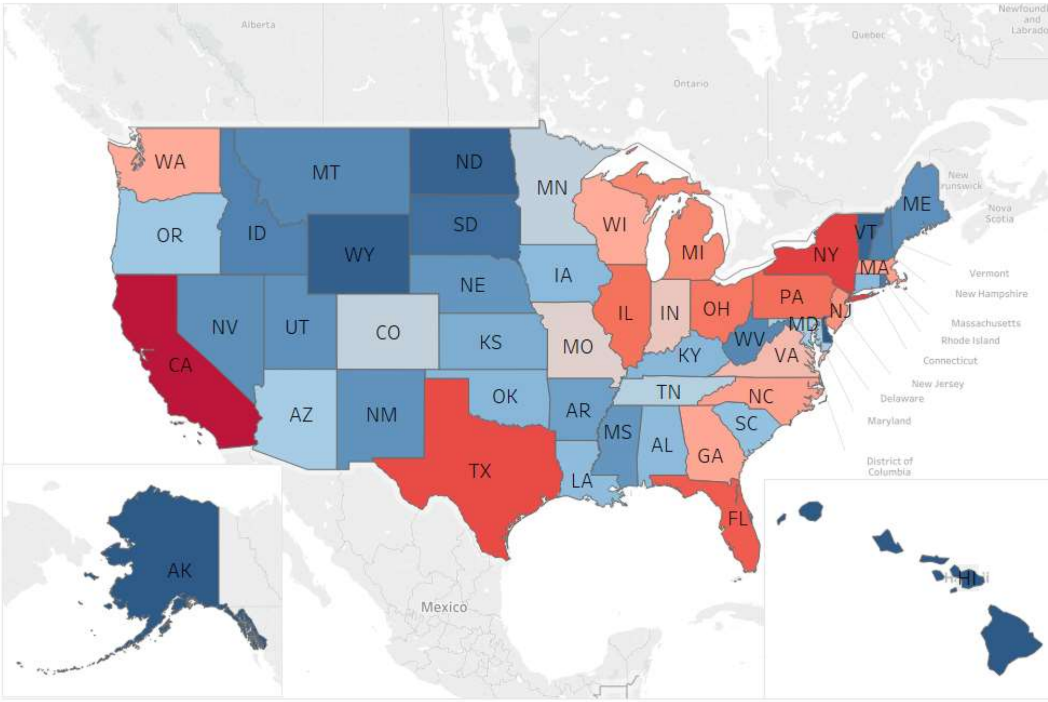
The graphs use the Mercator projection. For the lower 48 states, while the Mercator still has some minor distortions, the stretch isn't too bad to use. However, because there is large distortion when including the full-scale graph that includes Alaska and Hawaii, a dashboard was created and the two states were added to the bottom of the main graph that shows the 48 states.

The years 1997 and 2007 was chosen as the two years to compare based on the assumption that there would be more growth/decline over a 10 year period than a small interval of years. *The graphs in 1a was created using a red-blue diverging color scheme to put emphasis on the high and low values of food service*. The states that are blue indicates low food services and the red indicates a high level of food services; the measure used for the color function was scaled using log10 because it narrowed down the range and provided discernibility between color ranges by scaling the data in order of magnitude (prior to scaling, the values for each state was summed). The color scheme has the same data range for comparison as well for both graphs in 1a and 1b as well.

What's interesting about the choropleth is that is functions as a heat map with geographic boundaries where the color ramp shows the data values. And in the choropleth of the U.S., shape plays an important role in that larger areas can portray to have a bigger impact and show more importance. The five states with the greatest change between 1997 to 2007 were Nevada, Georgia, Arizona, Virginia, and North Carolina and the changes can be seen from one map to the next. While the District of Columbia (D.C.) is not a state, the capital also had a noticeable change in food service number from 1997 to 2007. However, because of the size of D.C., it goes virtually unnoticed on the choropleth - the focus is drawn to the larger states where size catches the eye. Because the larger states have a bigger impact than the smaller states, we need to be mindful when looking at choropleths and the context they are providing.

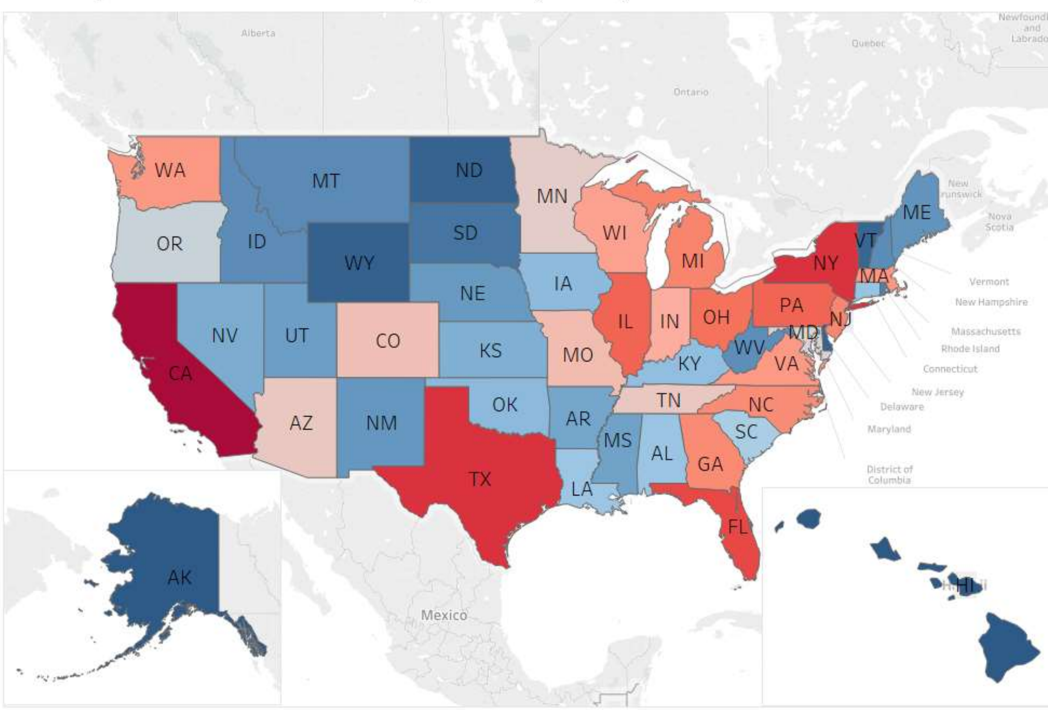
Log FS 1997
3.205 4.881

Choropleth: Food Service by State (1997)



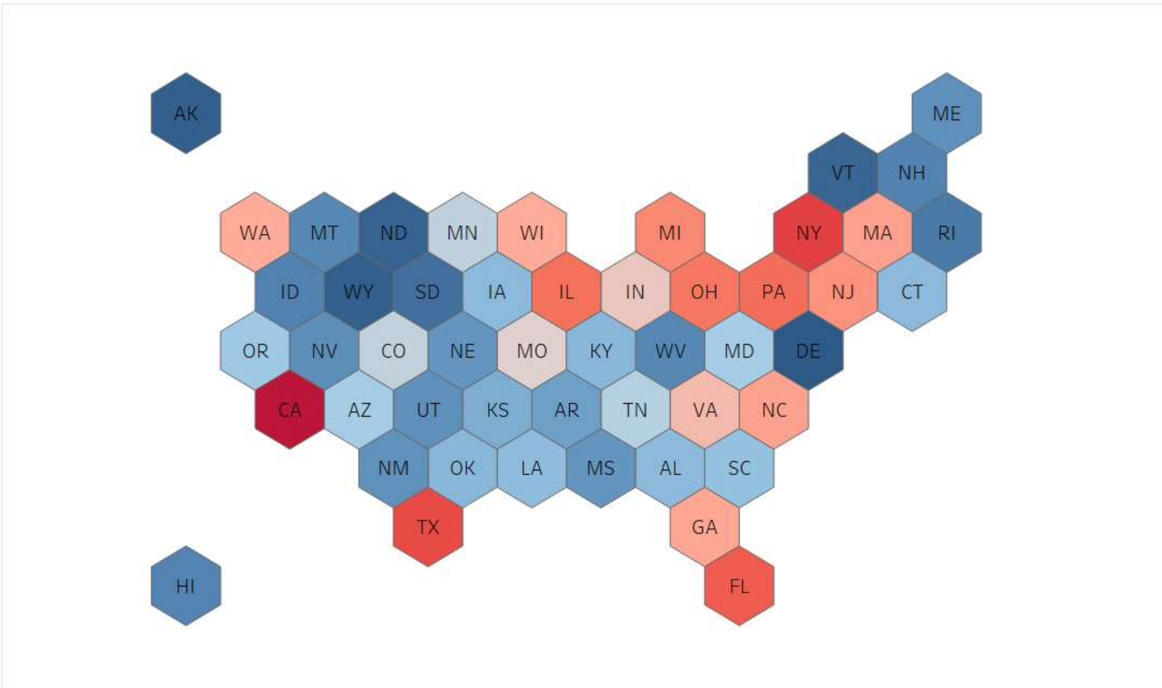
Log FS 2007
3.205 4.881

Choropleth: Food Service by State (2007)



Problem1b:

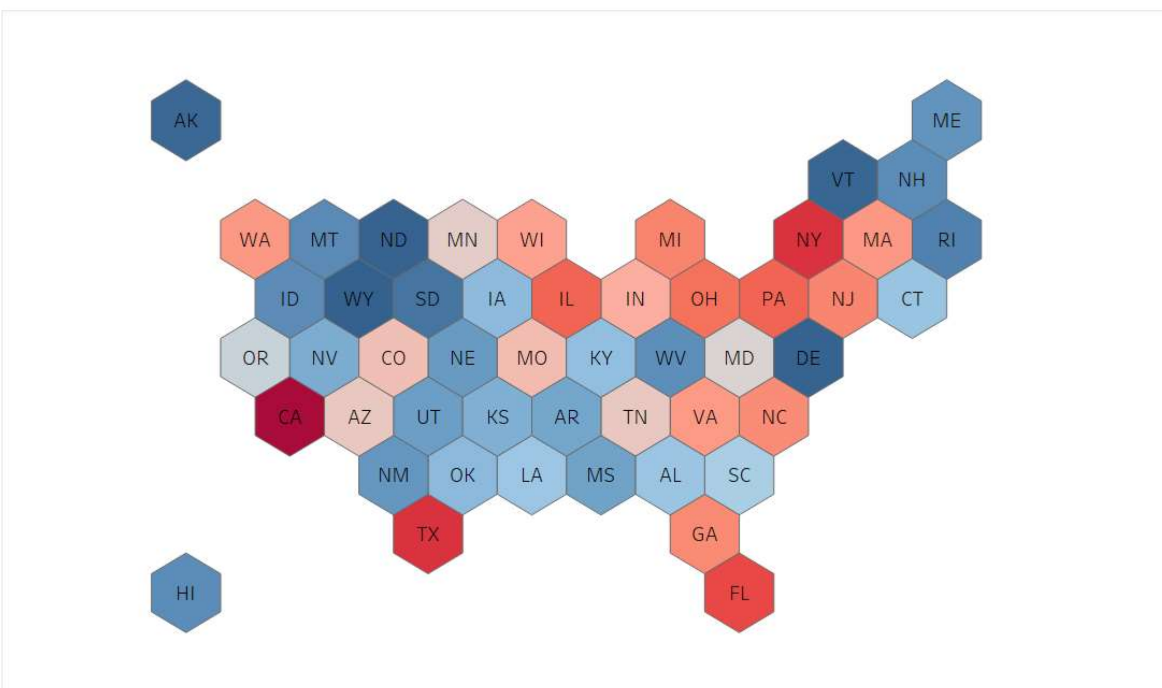
Tile Cartogram: Food Service by State (1997)



Log FS 1997

3.205 4.881

Tile Cartogram: Food Service by State (2007)

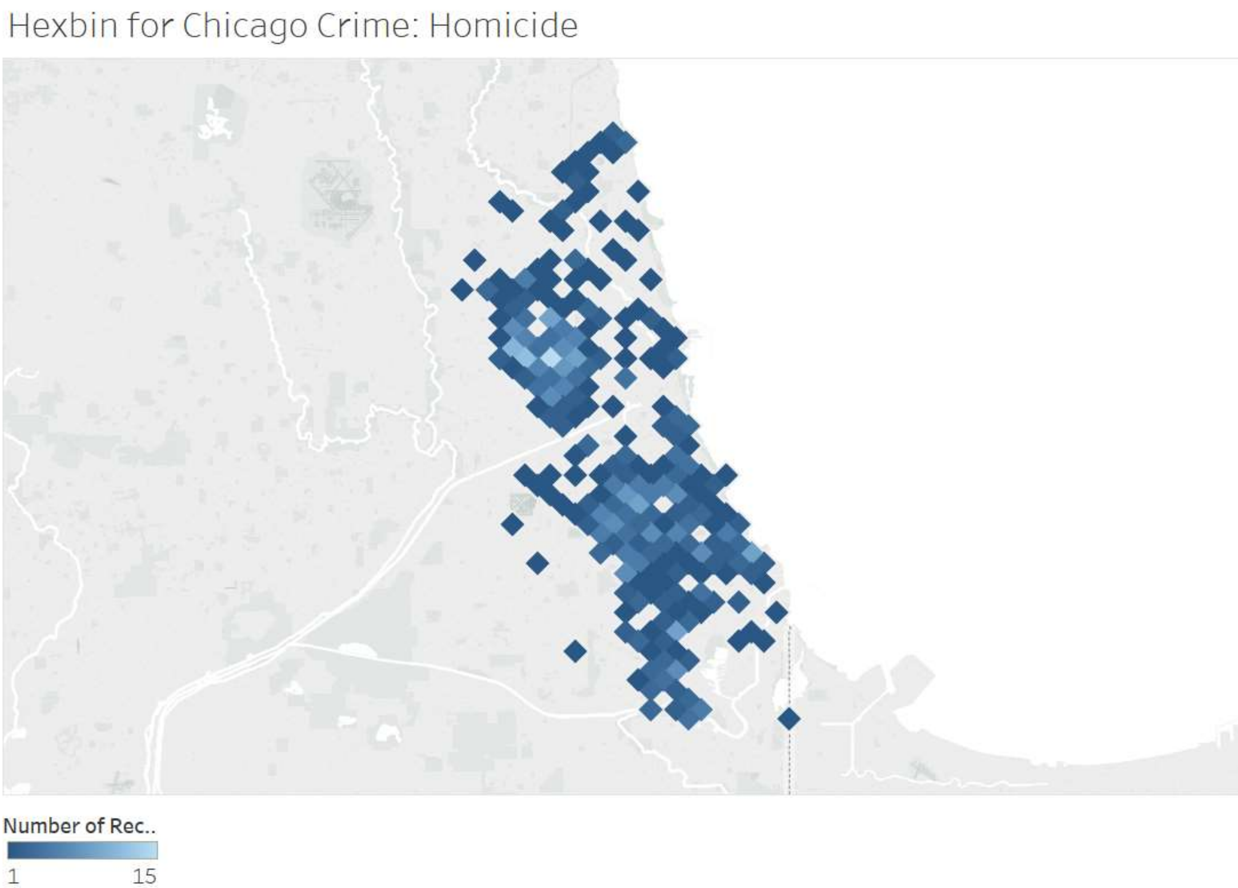


Log FS 2007

3.205 4.881

The benefit of the choropleth is that it preserves the states location on the map (geographic boundaries). The disadvantage is that while they do provide detail and preservation of its location on the map, there is some inherent distortion that occurs with the graph itself where the eyes are drawn to the bigger geographic areas even if the measurement value is higher for a smaller state (similar to the example given in 1a). The drawback of the choropleth is the strength of the tilegram because regardless of the size of the state, each state's tile color for the food service value (by year) is visible. The tilegram is the same as the choropleth but removes the effect of areas as a distortion source so that the eyes are not drawn to one state more than the others due to its size. In the case of the east coast states, the smaller states are just as easy to read as the bigger states. However, the small deficiency is positioning is approximate in the tilegram and therefore cannot map the states perfectly, but if the general layout of the states are understood, the tilegram does a good job of providing the context of the data.
** NOTE: D.C. was not on the hexagonal tile cartogram because it was not a part of the shape file and a join couldn't be made on missing data.

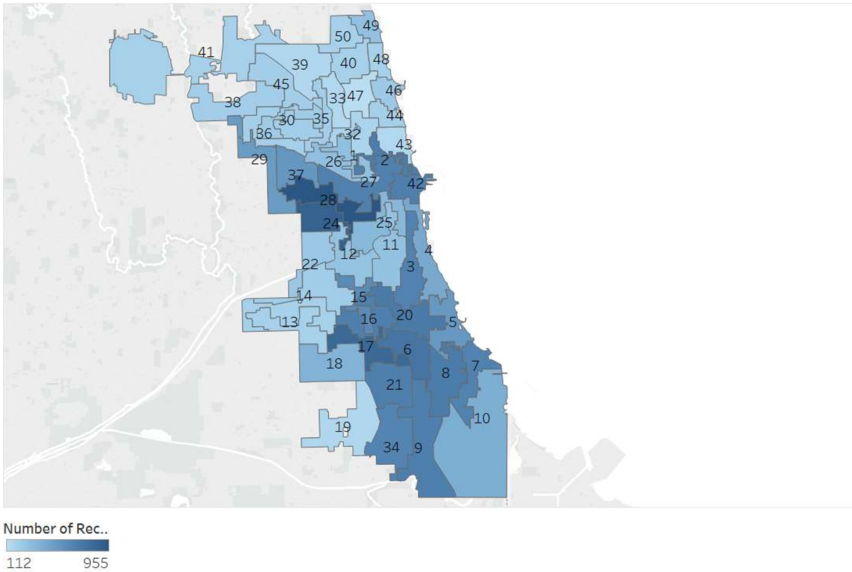
Problem 2a:



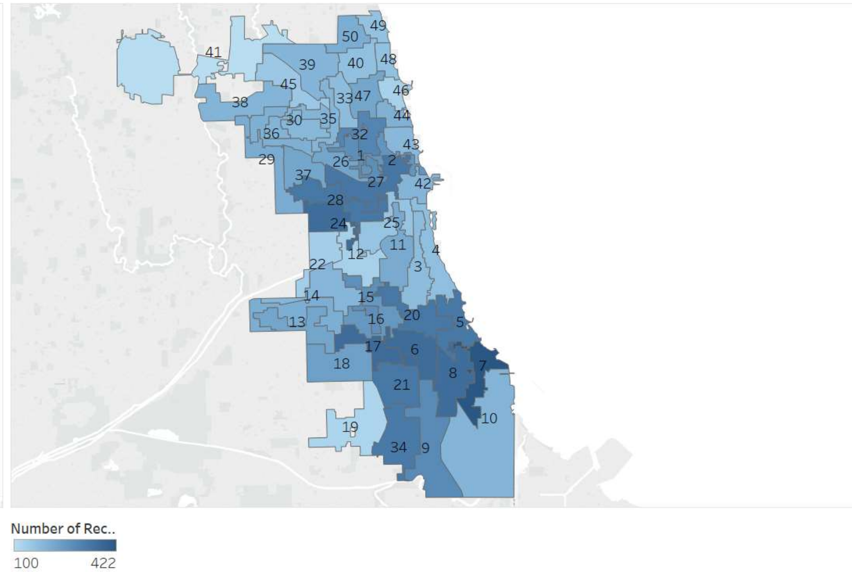
The hexbin was created as an overlay on top of the Chicago map in Tableau to reflect homicide. This was done by creating two measure variables 'HexbinX' and 'HexbinY' with ratios for adjusting their placement on the map. Afterwards, in the data source they were changed to a geographic assignment. The change in assignment allowed for the hexbin to plot over Chicago. What you see is that there are two locations with higher density than its surrounding areas, and the two locations are separated by a river. However, the higher concentration of homicide is located on the upper side of the plot. By adding the map underneath plot, it provides an approximate idea of where homicides take place and the rate. **The hexagon shape was not available, so the square shape had to be used.

Problem 2b:

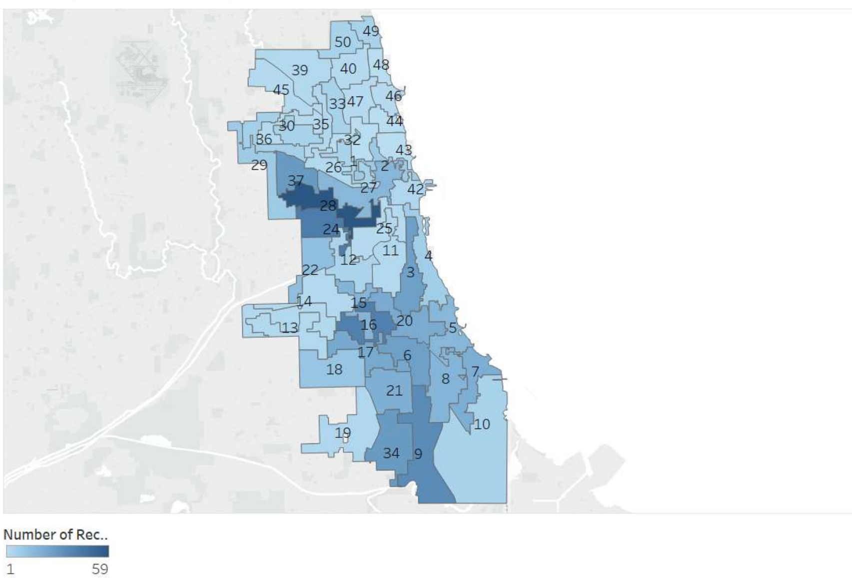
Choropleth: Assault



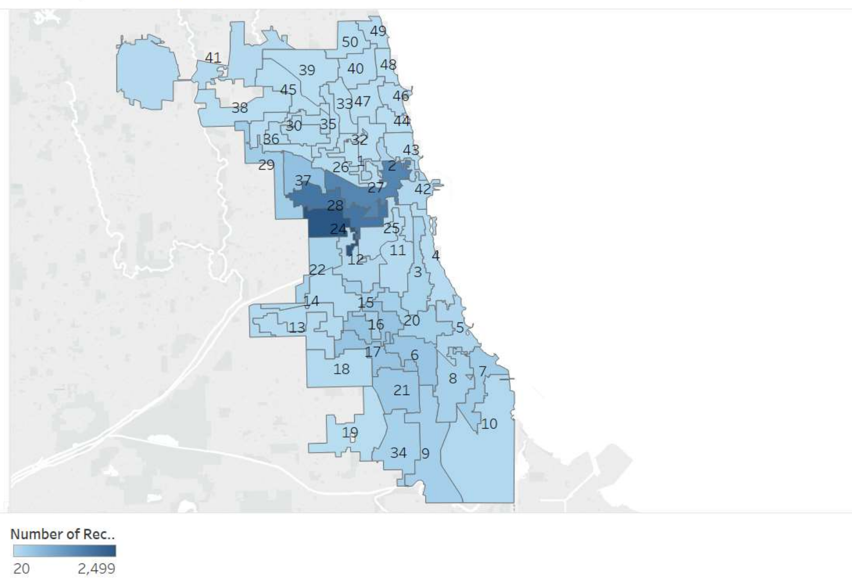
Choropleth: Burglary



Choropleth: Homicide



Choropleth: Narcotics



Four choropleths were created for Assault, Burglary, Homicide, and Narcotics using the same blue color scheme. The ward numbers were added as text labels to easily identify each community area. What is noticeable is the different wards and their concentration respective to each crime. For instance, in the Assault choropleth, the number of assaults is highest in ward 28, which happens to be the West Side of Chicago. Ward 28 also happens to be the highest in Homicide as well.

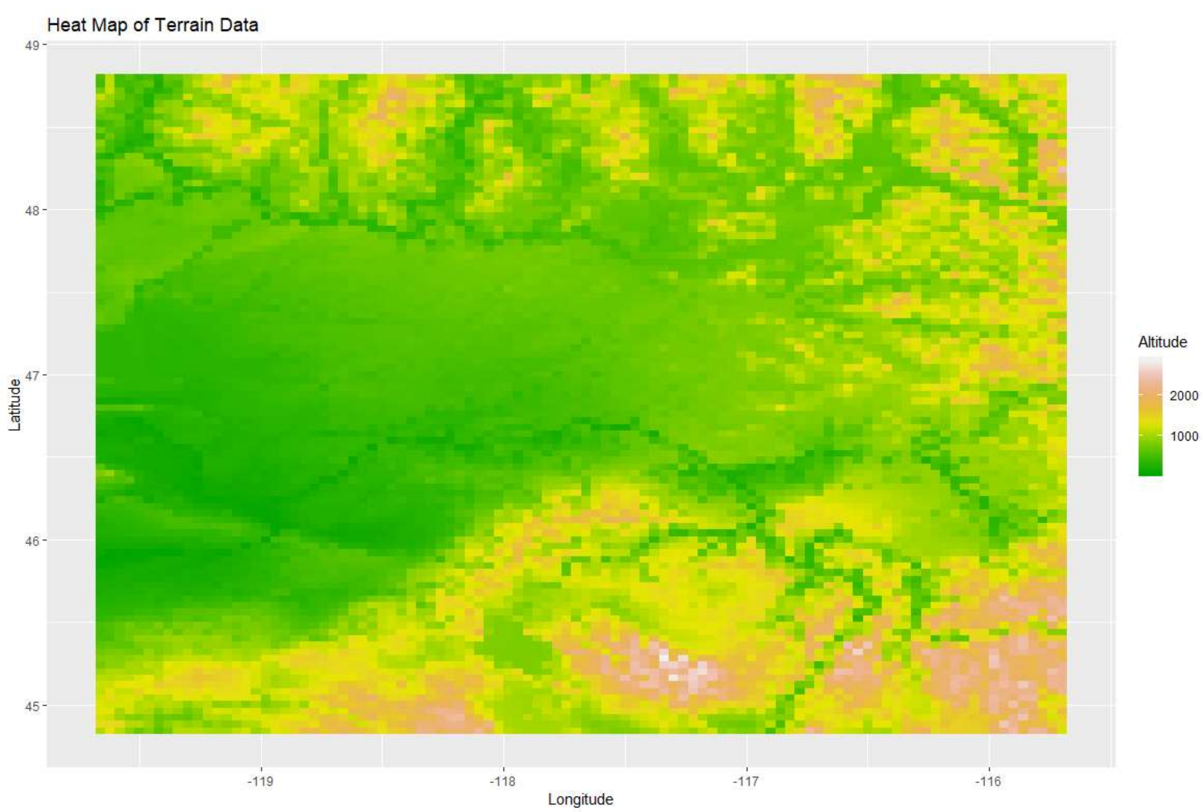
To compare the Homicide hexbin overlap in problem 2a, a Homicide choropleth was created. After looking at the two, the hexbin is provides a general level of detail by showing density through the use of color. The choropleth is able to do the same as the hexbin plot, but it also provides more granularity by preserving the geographic boundaries. And by preserving the boundaries, text labels can be added to provide additional detail/clarity.

** For Problem 3a and 3b, the R Code is attached in the appendix of this homework

Problem 3a:

R Code:

```
ggplot(data=terrain_df, aes(x=x, y=y, fill=z)) +
  scale_fill_gradientn(colors=terrain.colors(10)) +
  labs(title="Heat Map of Terrain Data", x="Longitude", y = "Latitude",  fill="Altitude") + geom_tile()
```

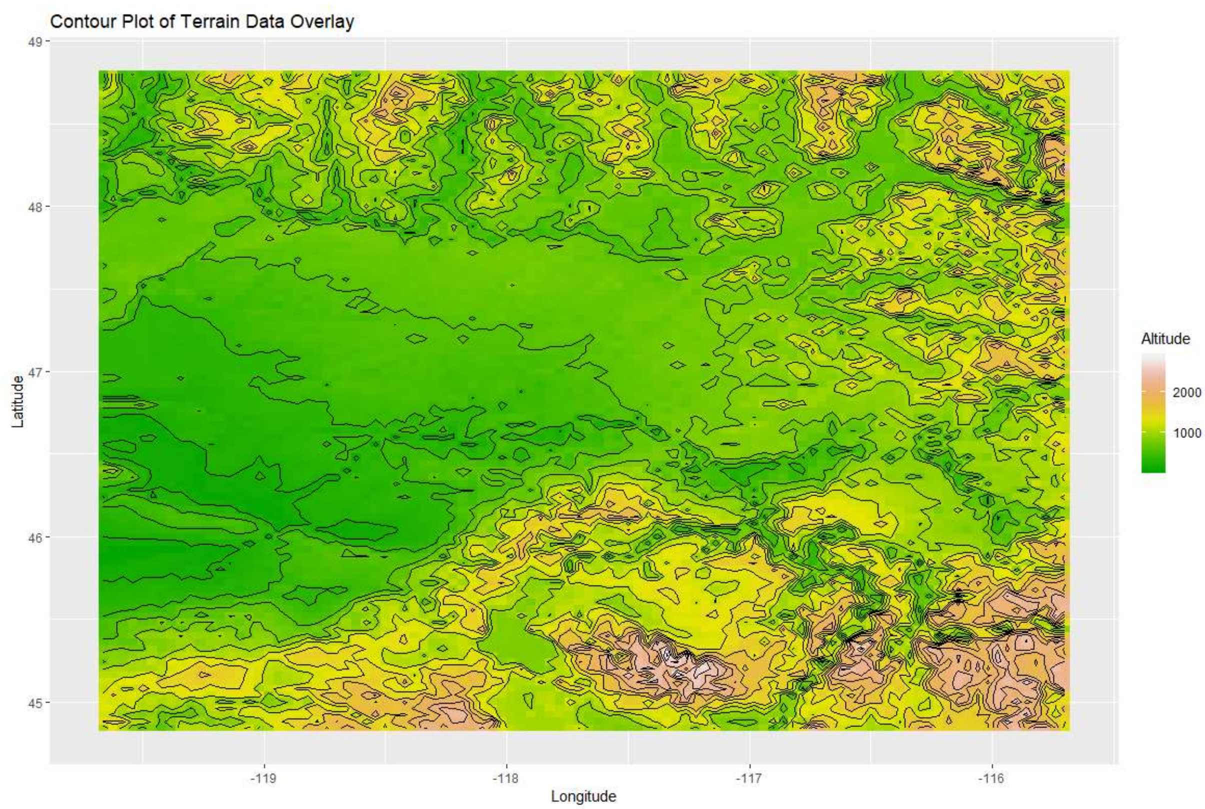


The heat map of the terrain data shows the altitude changes using color going from green to yellow to brown/pink to white through their actual value. A large portion of the heat map shows altitude/height that is less than 1000ft indicating that this is flatter surface with the use of the color green, whereas the parts that are yellow/brown and pink indicate higher altitude/height taking up less space on the terrain.

Problem 3b:

R Code:

```
ggplot(data=terrain_df, aes(x=x, y=y, fill=z)) +
  geom_tile() + scale_fill_gradientn(colors=terrain.colors(10)) +
  labs(title="Contour Plot of Terrain Data Overlay", x="Longitude", y = "Latitude", fill="Altitude") +
  geom_contour(aes(x = x, y = y, z = z), color="black")
```



With the contour plot, we overlay this on top of the heat map, and this allows us to see the actual height values and the change in as well through the closeness of the contour plots. From the details of both plots combined into one, we can judge the steepness of the terrain as well as the values through the color.

Problem 3c:

Overlaying the contour plot over the heat map allows the plot to provide richer context to the data than each plot as a standalone data visualization. The space between the curves determine how fast the actual height is rising as we move across the terrain. By adding the heat map underneath, the use of the color scheme communicates additional context with height values. On the contour plot overlay above, due to the large spacing of the contour lines on the left side of the graph and its green color, we can see that the height/altitude change is slow, which is indicated by the legend as less than 1000ft. However, as we go to the perimeter of the graph, the closeness of the lines indicate a quicker change/steepness as the geographic area also increases in altitude, which again is supported by the use of color and the legend proving that the altitude is quickly rising approaching 2000ft and higher. Towards the bottom of the graph, we can see an area on the plot that shows some white areas with close contour lines indicating that this location is the highest point on the graph reaching close to 3000ft. This segues to the next point: the contour plot is also able to help identify shape of the geographic area. For example, on the plot, the concentration of the contour line and the use of color indicating height/altitude value shows that the pink/brown areas are mountains. By overlaying both plots, the new plot provides richer context of the terrain and quicker interpretability by complementing each other.

Appendix: R Code for Problem 3

```
library(ggplot2)
library(stringr)
library(hexbin)


setwd("C:/Users/Cindy/Documents/DSC 465/Homework/Homework 4")


#Problem 3:
require(grDevices) # for colors

terrain_df = read.csv("terrain1.csv")
head(terrain_df)
help(package=colorspace)
## Heat Map"
ggplot(data=terrain_df, aes(x=x, y=y, fill=z)) +
  scale_fill_gradientn(colors=terrain.colors(10))  +
  labs(title="Heat Map of Terrain Data", x="Longitude", y = "Latitude", fill="Altitude") +
  geom_tile()

## Contour Plot
#ggplot(terrain_df, aes(x = x, y = y, z = z, fill=z)) +
#  geom_tile() + scale_fill_gradientn(colors=terrain.colors(10)) +
#  geom_contour(color="black") +
#  labs(title="Contour Plot of Terrain Data", x="Longitude", y = "Latitude", fill="Altitude")


ggplot(data=terrain_df, aes(x=x, y=y, fill=z))  +
  geom_tile() + scale_fill_gradientn(colors=terrain.colors(10)) +
  labs(title="Contour Plot of Terrain Data Overlay", x="Longitude", y = "Latitude", fill="Altitude") +
  geom_contour(aes(x = x, y = y, z = z), color="black")
```