M6_Lab

Emily Shepherd

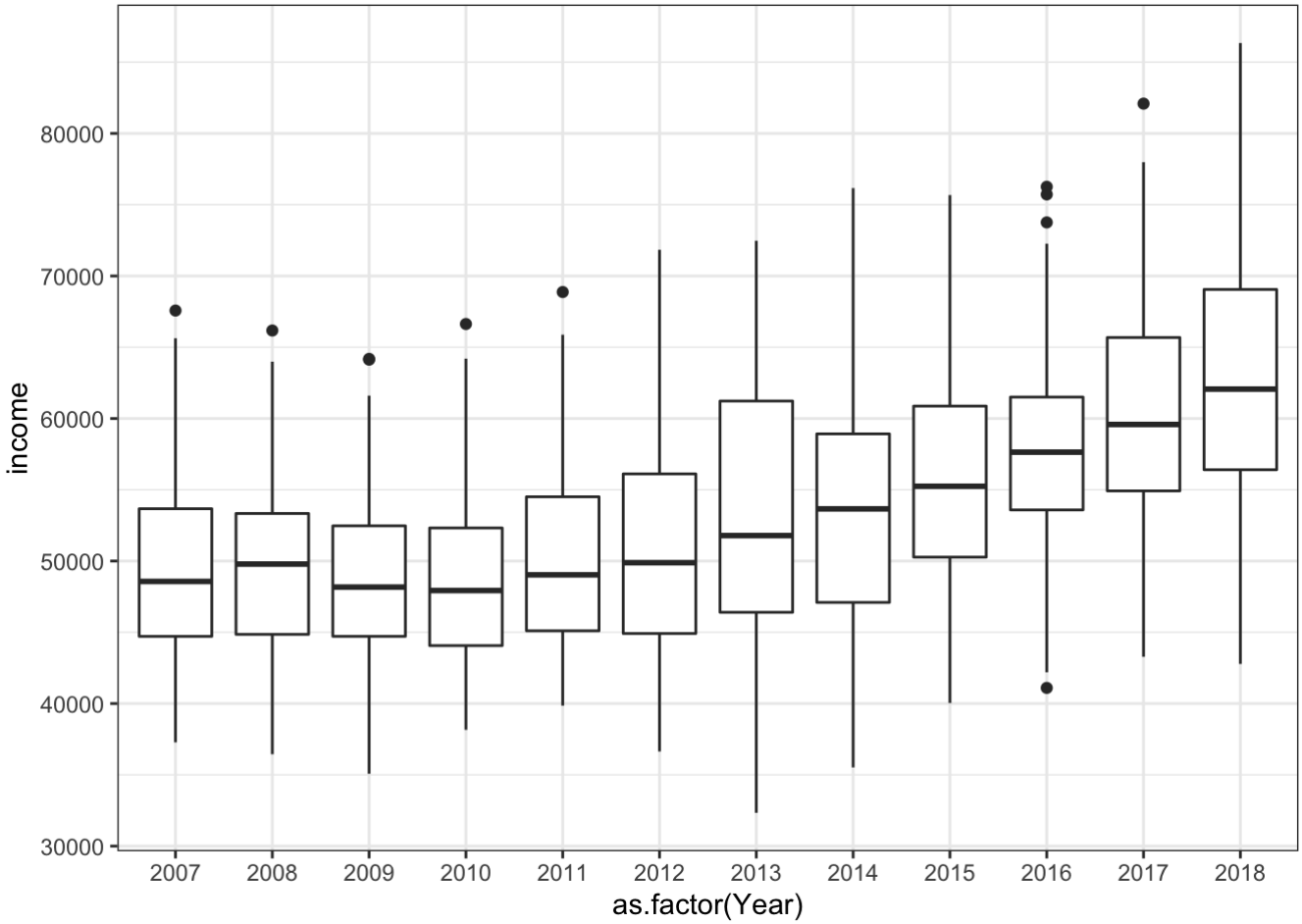
2023-02-18

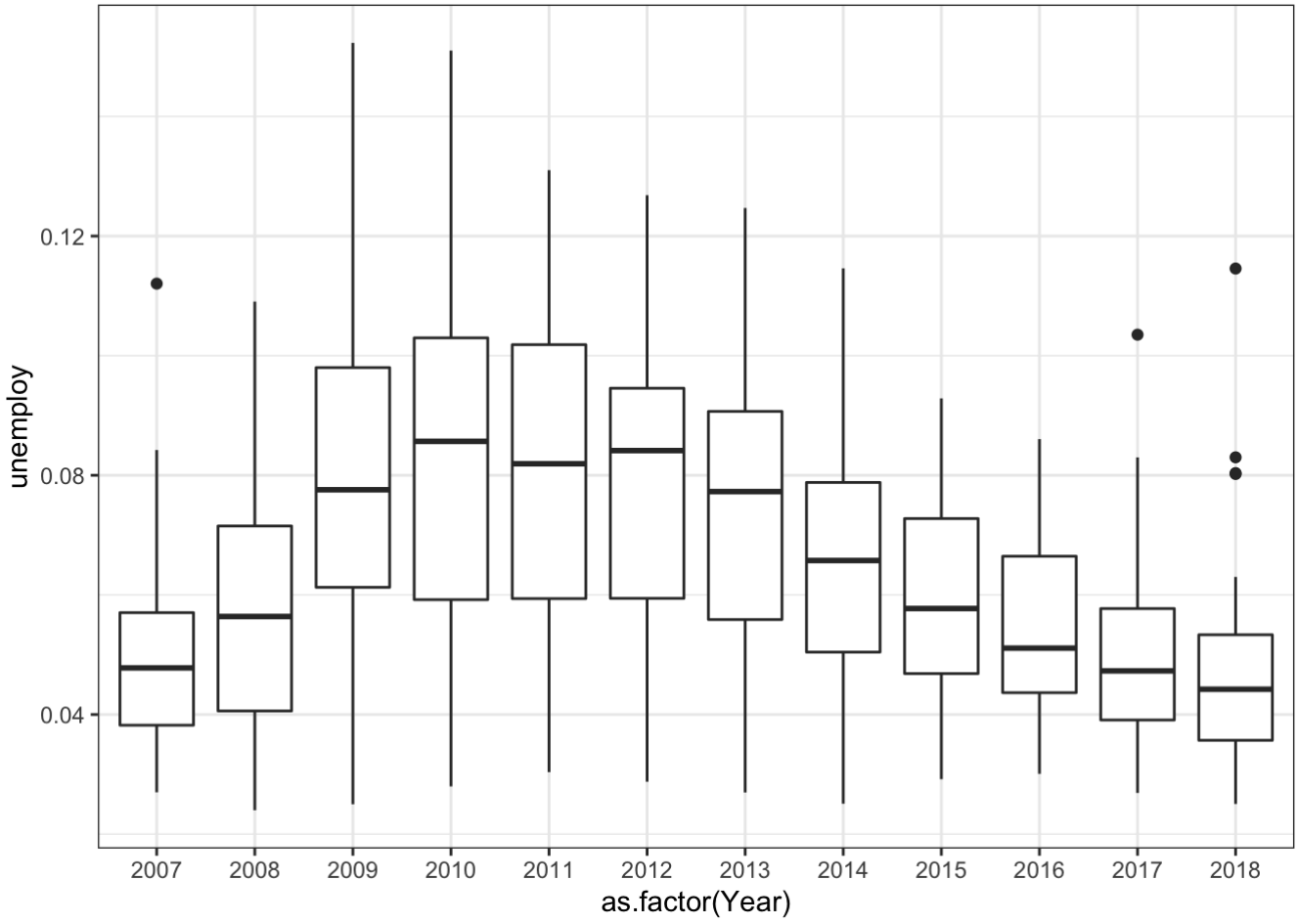
Part 1: The Data

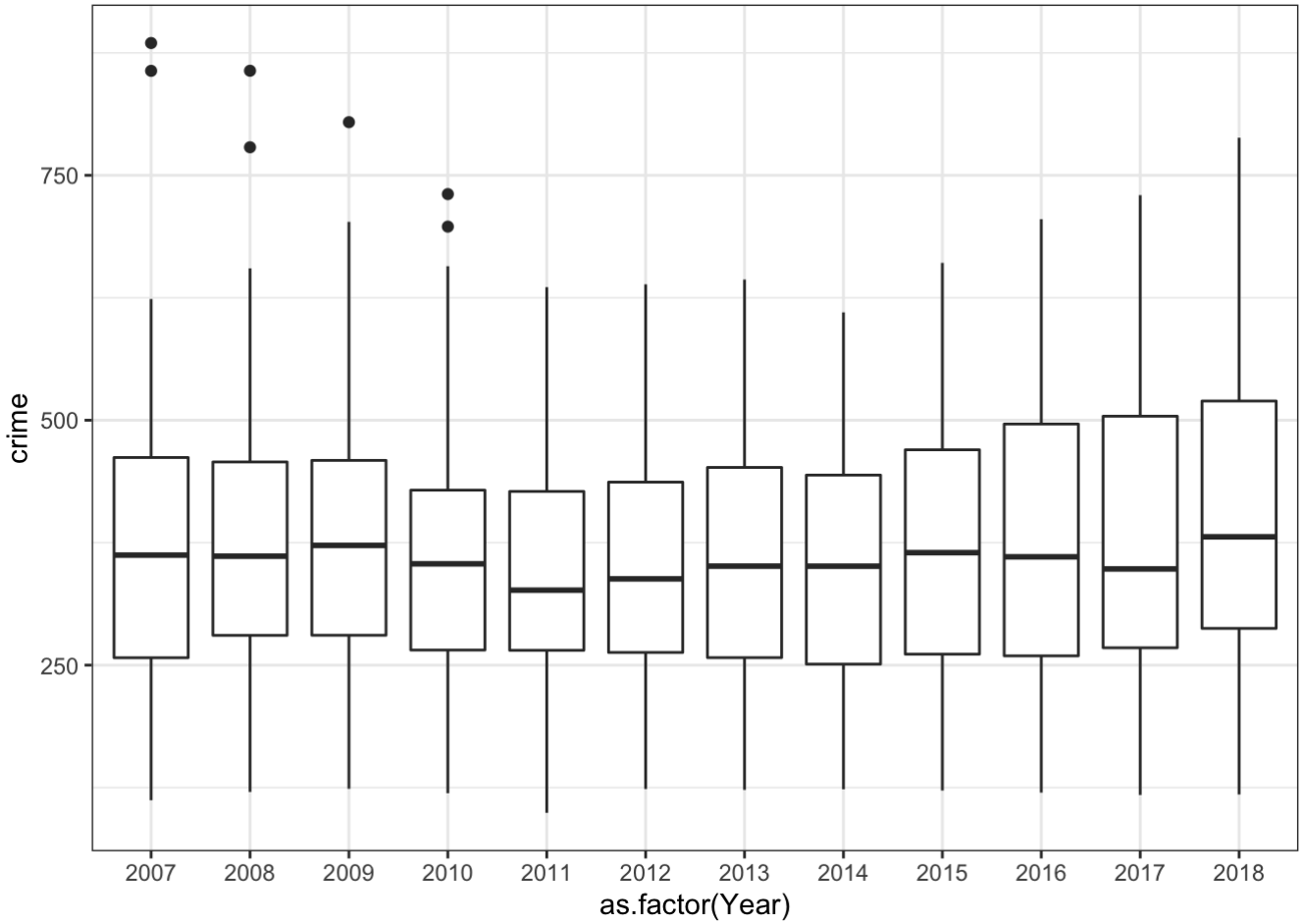
The data consists of average annual income, unemployment rate, and the crime rate (instances of crime per 100,000 people) for each state from the years 2007 to 2018. While I was cleaning the data, I noticed that there were some state data missing particularly from the states New Jersey, Delaware, and Connecticut. I created a few variables to make the data easier to use. First, I calculated the unemployment rate by dividing the number of people unemployed by the population. Second, I created two categorical variables. One ranked the average income as high, medium, or low. The other variable I created categorized the unemployment rate as high, greater than or equal to 5%, or ideal, less than 5%.

Part 2: Exploratory Data Analysis

I explored data using rudimentary plots for the data to determine interesting relationships to explore in the data. Below are the plots, I used. Based on the below plots, I decided to focus on the relationships among crime rate, income, and unemployment, focusing on the year before the financial crisis in 2008 and then three years afterward, 2009, 2014, and 2018.

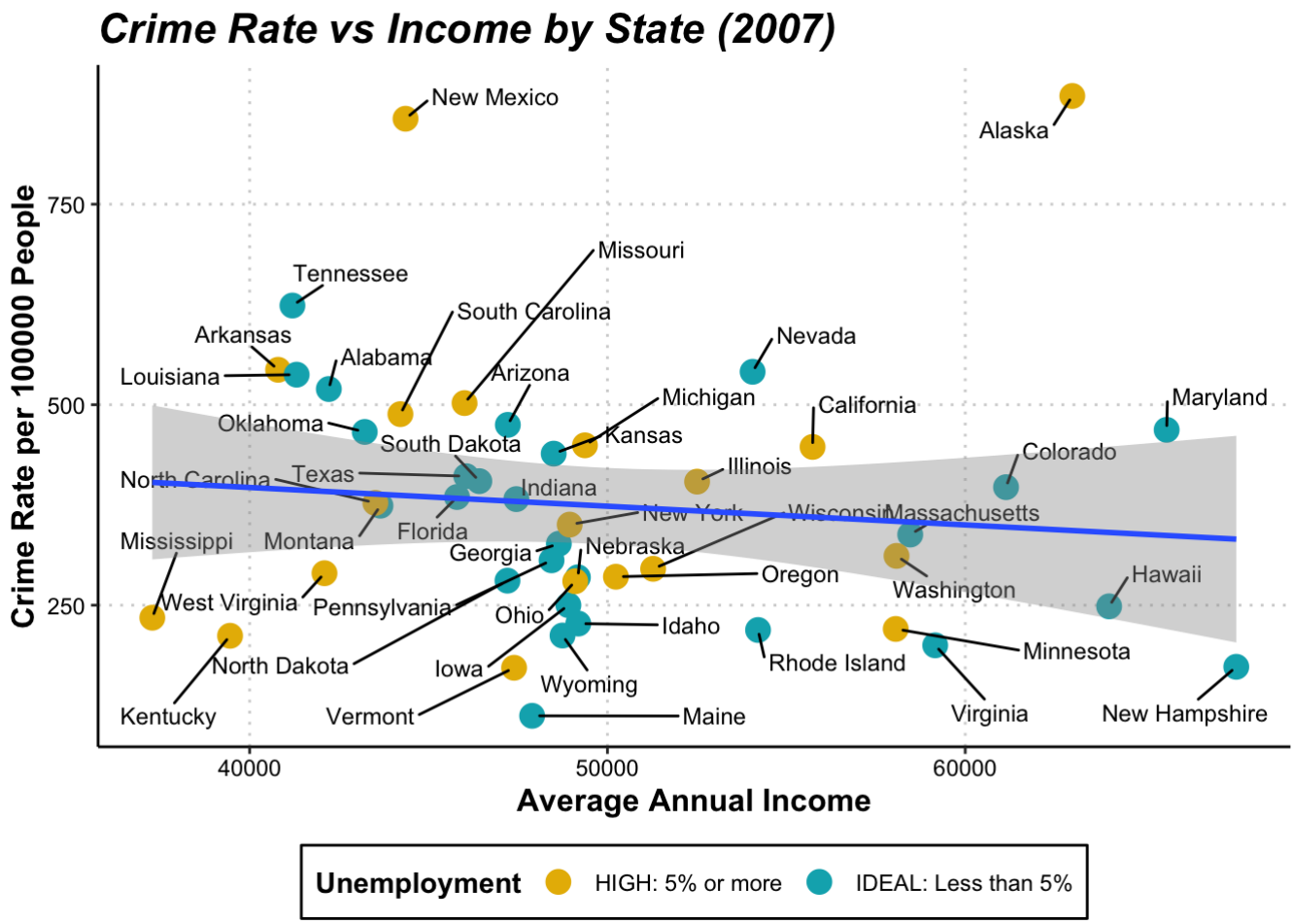

income
80000
70000
60000
50000
40000
30000
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
as.factor(Year)


unemploy
0.12
0.08
0.04
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
as.factor(Year)

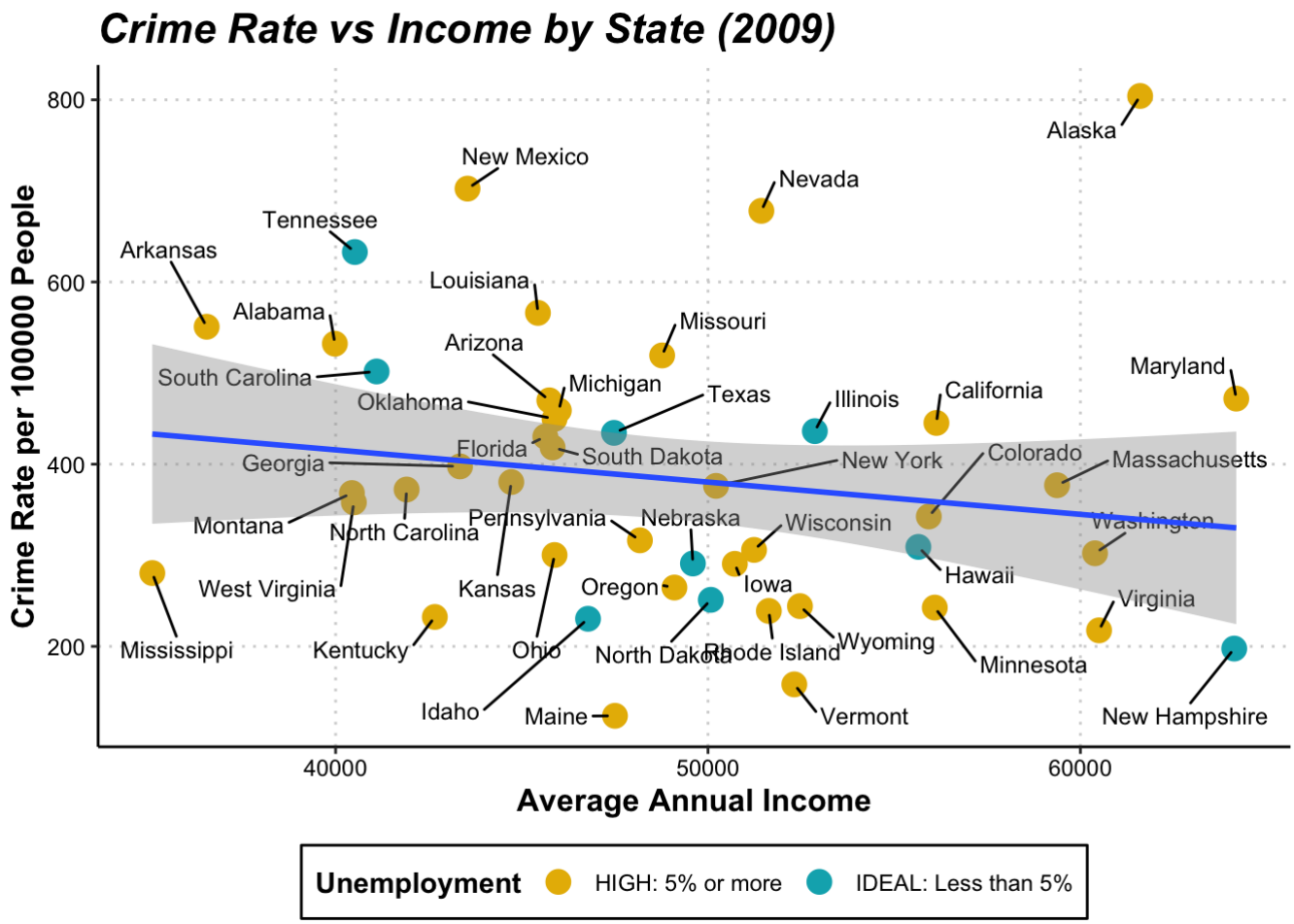


Part 3: The Graphs

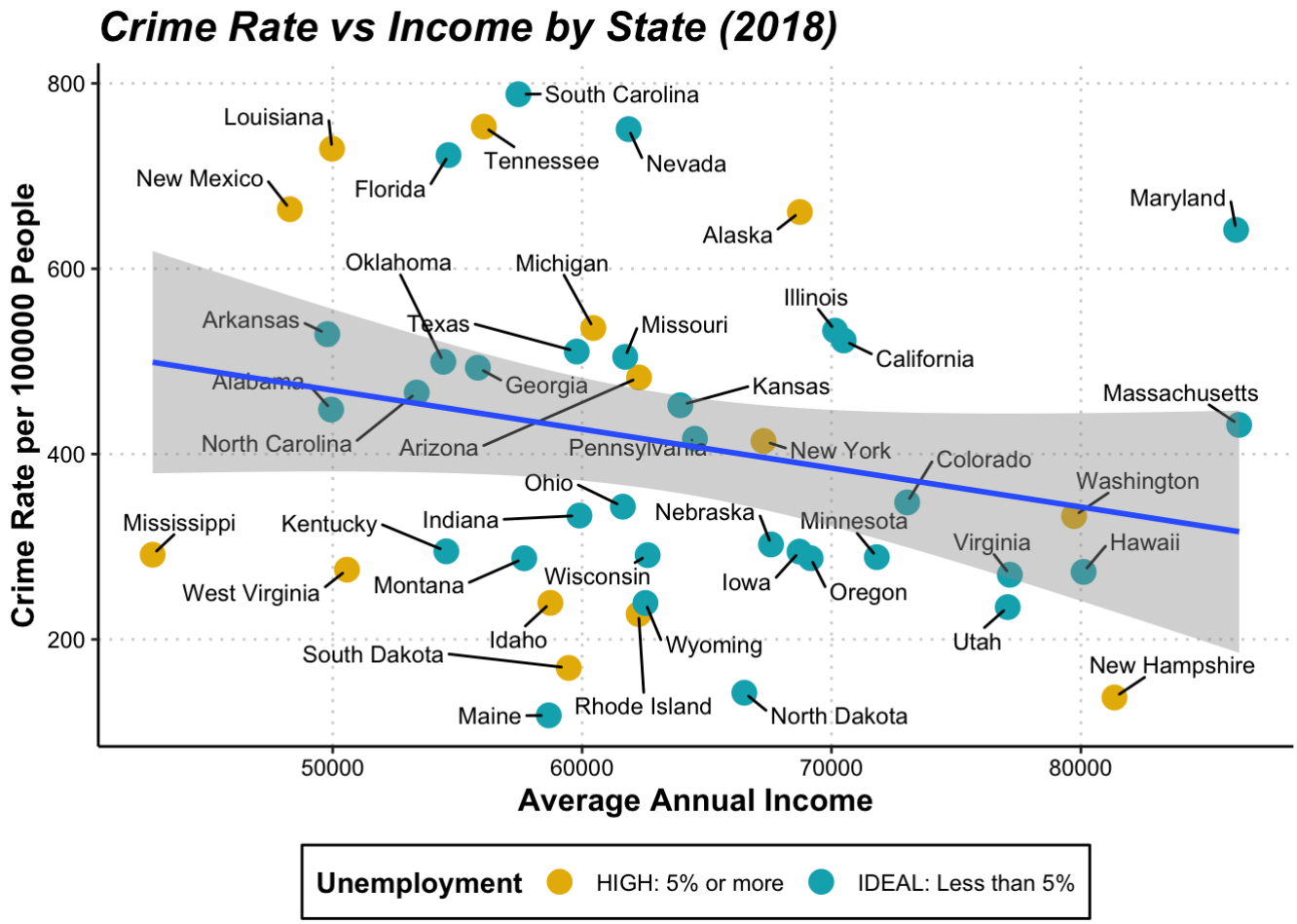
The first graph displays the crime rate vs average annual income for the year 2007. The color of the points is determined by whether the unemployment is rated as high or ideal.



The second graph displays the data for 2009. One year after the financial crisis hit.



The third graph displays the data for 2018, ten years after the initial financial crisis.



Part 4: Conclusion

I prepared this for an academic audience interested in the effects of the 2008 financial crisis on state level economic indicators. I selected colors for for my scatterplots which would be friendly for people who have difficulties discerning color. I chose to not to use color in my panels and background to keep the graphs clean and crisp and also, to save on ink if they were to be printed.