Crime Data- The data set was tidy when we initially downloaded it. It measured the total number of violent crimes per 100,000 people from

the states we are working with from the years 1990=2012. When gathering our unemployment and immigration data, we noticed that the scale was

off, since the data was being measured per 100,000. In R, we used the mutate command to create new columns with the crime data divided by

100. In doing so, the data was easily comparable to our other two data sets and made for more readable visualizations.

Immigration Data- Our immigration data initially measured the number of immigrants in each state from the years 1990-2012. Since these numbers

were large and difficult to compare to our other data, we used the mutate command in R to create new columns with our Immigration data divided

by the population of the state we were working with, during the particular year we had data from. This resulted in very small percentages

(less than 1%), so in Tableau, we multiplied this data by 100, which allowed us for better visual comparisons.

Unemployment Data- Our original unemployment dataset contained a good amount of unnecessary rows and columns. Instead of giving an unemployment

rate year by year, it gave a rate month by month, which made our data set 12 times larger than we wanted to work with. Some coding in R

sucessfully aggregated the data using a function which summed the data from Jan-Dec of each year, and then found an average unemployment

rate, but was unable to be represented in a table due to issues with row length. We then decided to go into excel and use the 'AVERAGE'

function to highlight 12 rows, average them and create a new column that showed an average unemployment rate for each year. We then deleted

our old data and created a shorter, tidier csv.

We then joined all 3 datasets using the command 'inner\_join" in R, joining by 'Year', a common variable that we made sure each table included.

The end result was a tidy, readable data set.