Exercise 8: Graphics (Answer Key)

Marcy Shieh

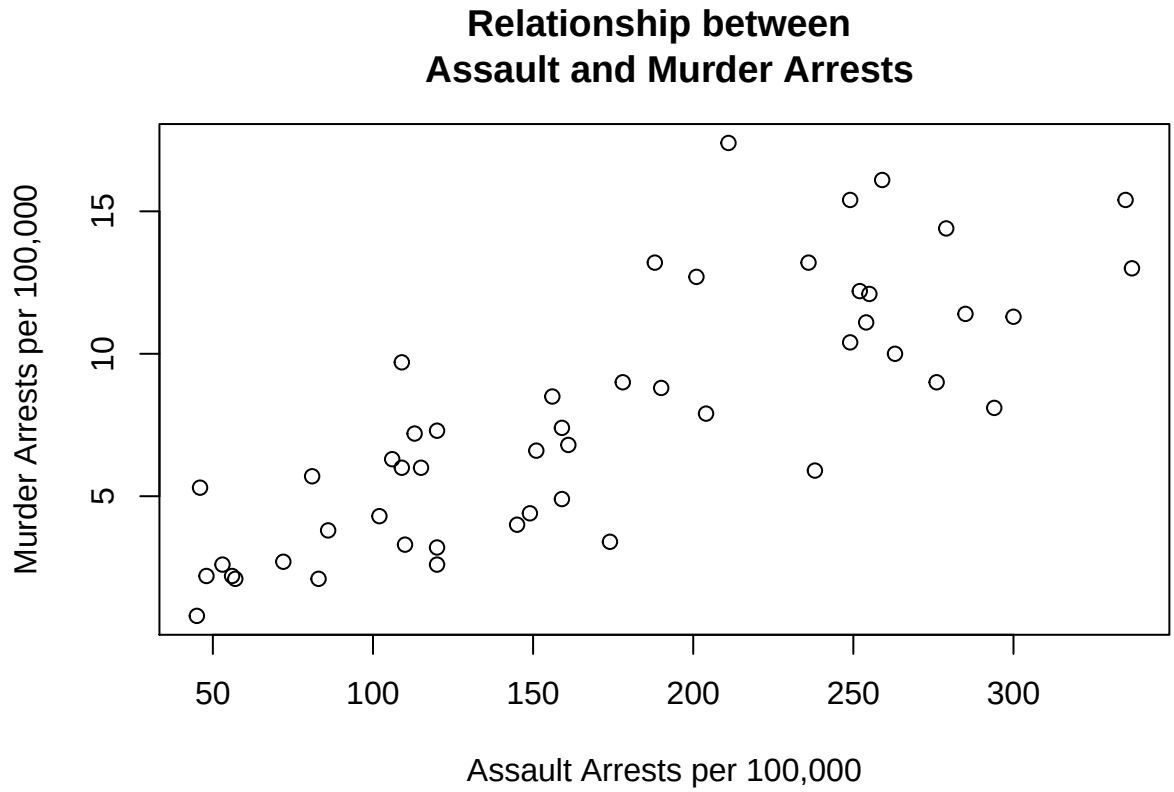
10/22/2020

```
library(ggplot2)
```

Create the following graphs in both base R and `ggplot2`.

1. Check out the base R built-in dataset, `data("USArrests")`.
2. Create a scatterplot that looks at the correlation between murder and assault arrests. Label the x and y axes and title the graph.

```
# base R
plot(USArrests$Assault, USArrests$Murder,
     main = "Relationship between \n Assault and Murder Arrests",
     xlab = "Assault Arrests per 100,000", ylab = "Murder Arrests per 100,000")
```

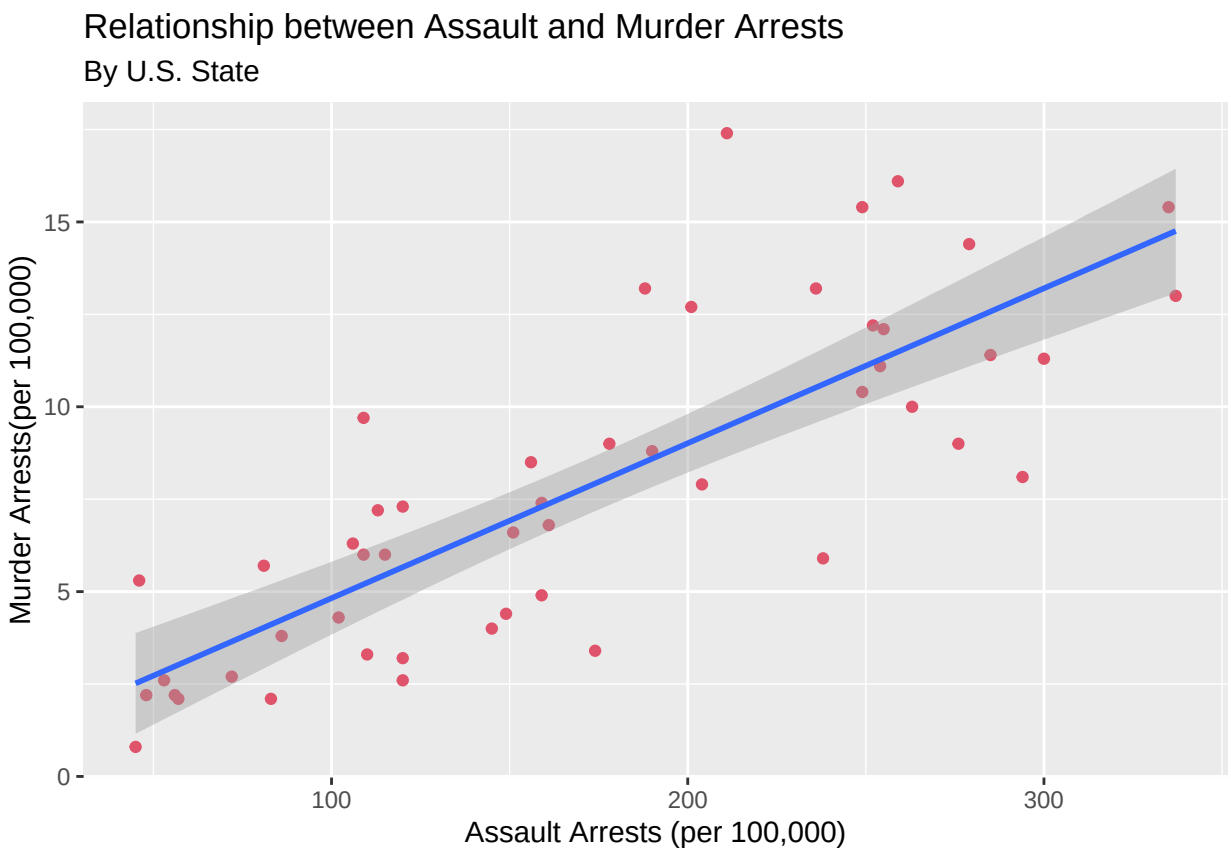



```
names(USArrests)
```

```
## [1] "Murder"   "Assault"  "UrbanPop" "Rape"
```

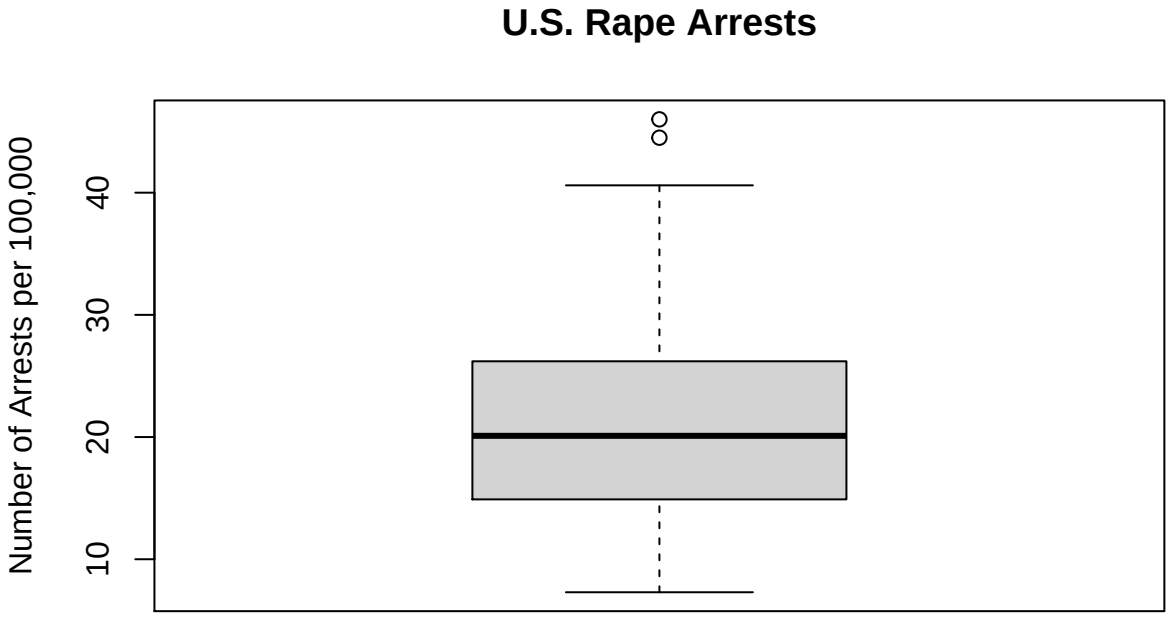
```r
# ggplot2
ggplot(data = USArrests, aes(x = Assault, y = Murder)) +
  geom_point(col = 10) + geom_smooth(method = "lm") +
  labs(title = "Relationship between Assault and Murder Arrests",
       y = "Murder Arrests(per 100,000)",
       x= "Assault Arrests (per 100,000)",
       subtitle = "By U.S. State")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

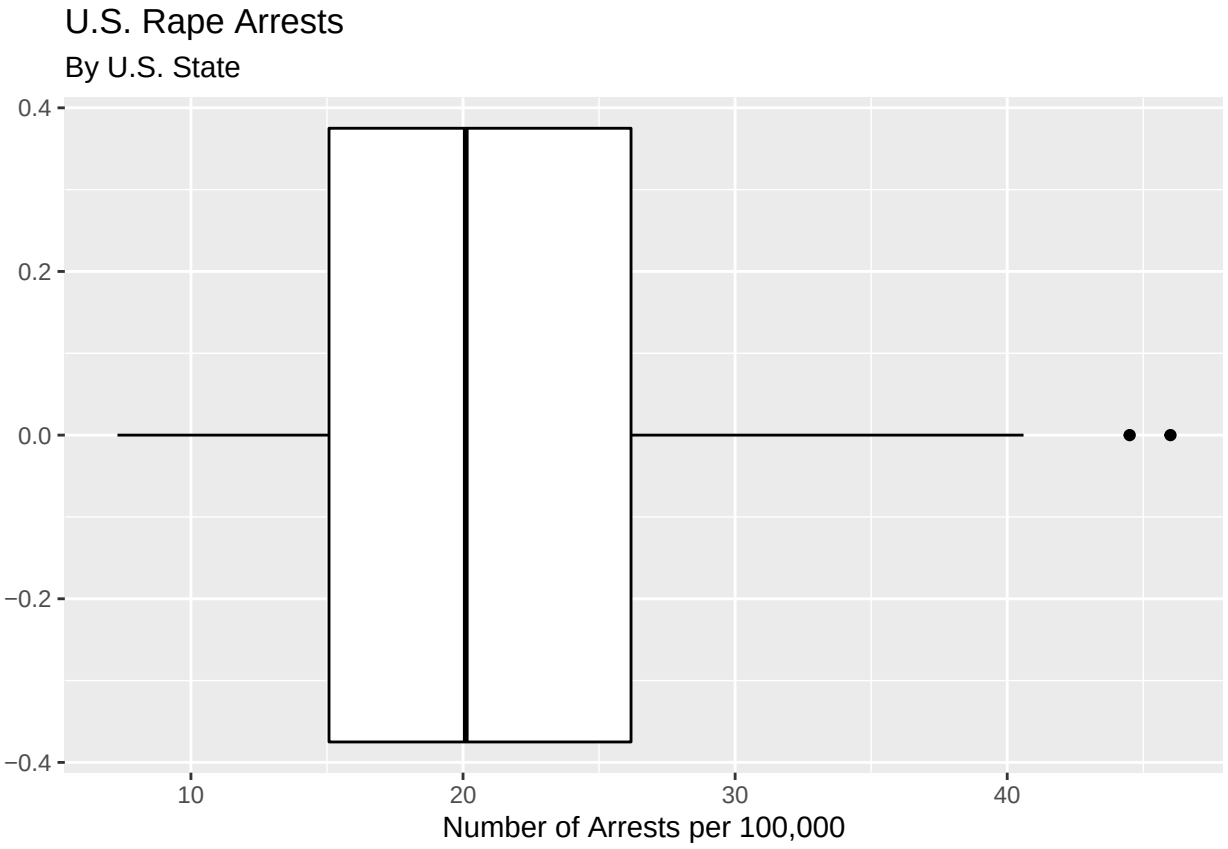


3. Create a boxplot of rape arrests. Label the plot.

```r
# base R
boxplot(USArrests$Rape, main = "U.S. Rape Arrests",
        ylab = "Number of Arrests per 100,000")
```

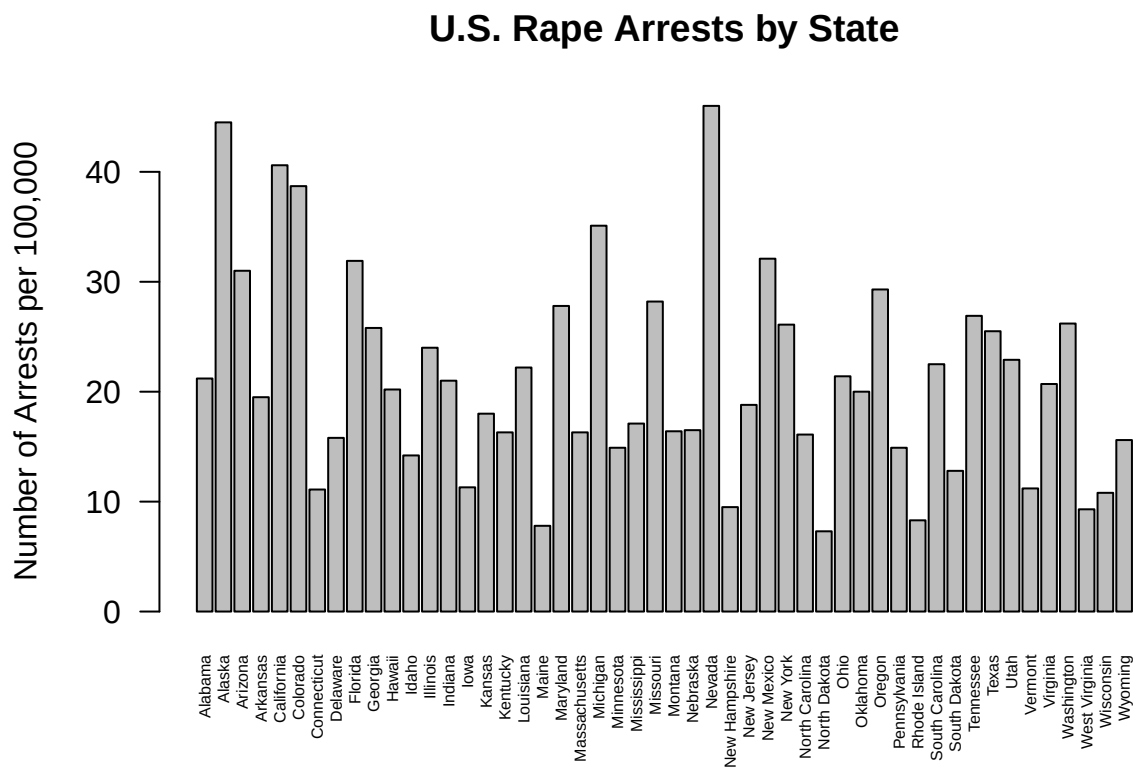



```
# ggplot2
ggplot(data = USArrests, aes(x = Rape))+
  geom_boxplot(col = 1) +
  labs(title = "U.S. Rape Arrests",
       subtitle = "By U.S. State",
       x= "Number of Arrests per 100,000")
```

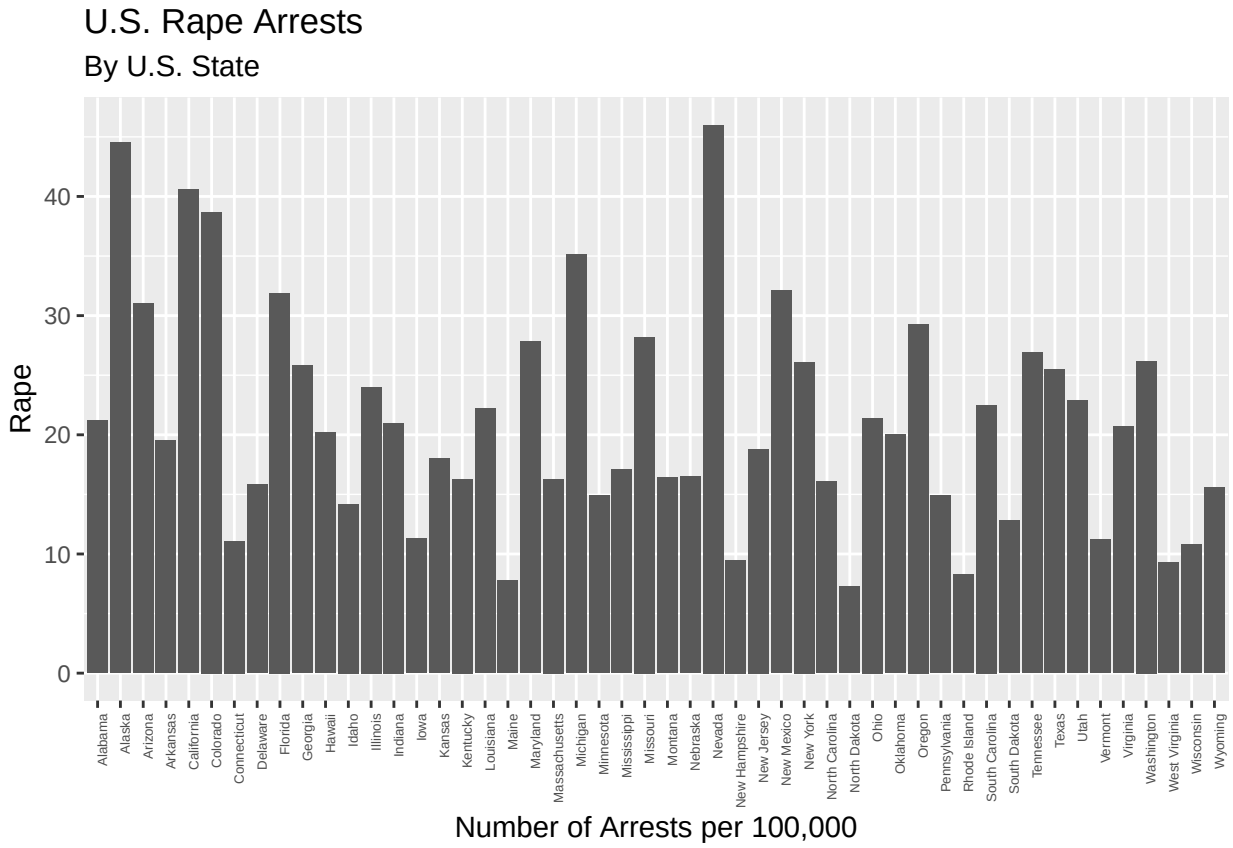


4. Create a barplot of the number of rape arrests per state.

```r
# base R
barplot(USArrests$Rape,
        names = row.names(USArrests),
        main = "U.S. Rape Arrests by State",
        ylab = "Number of Arrests per 100,000",
        las = 2,
        cex.names = 0.5
        )
```

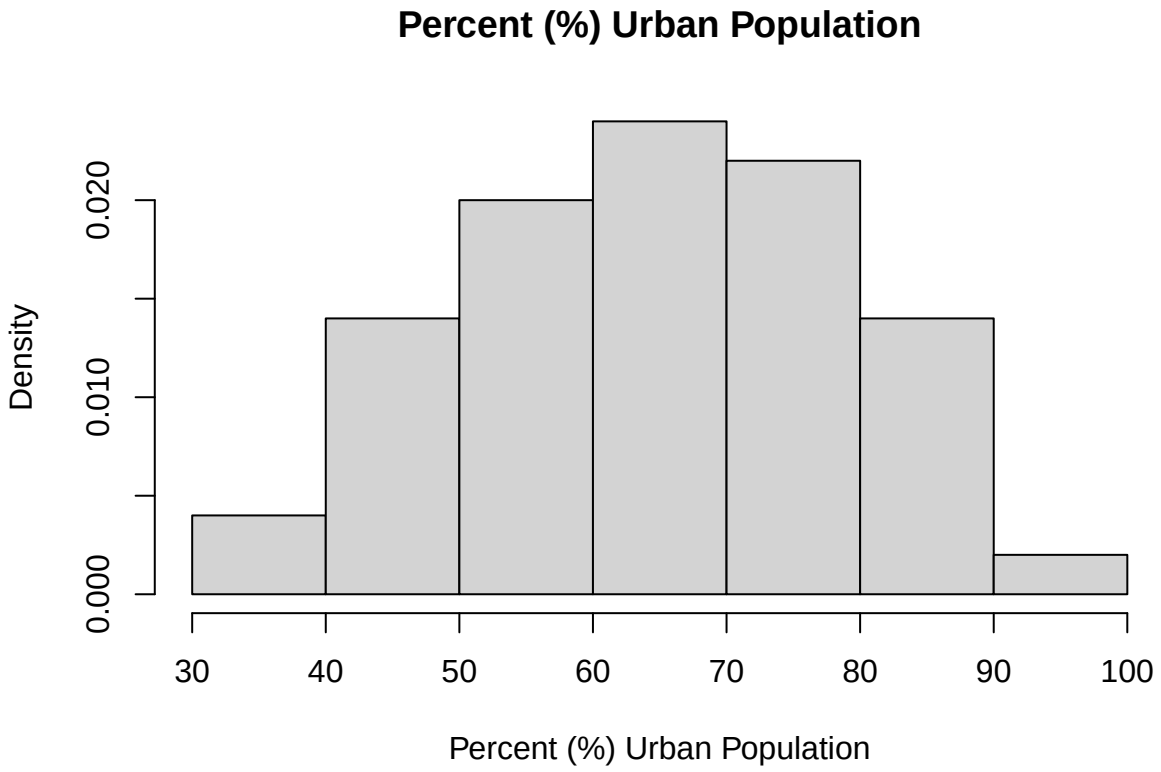



```r
# ggplot2
ggplot(data=USArrests, aes(x = row.names(USArrests), y = Rape)) +
  geom_bar(stat= "Identity") +
  theme(axis.text.x= element_text(angle = 90, hjust = 1, size = 5)) +
  labs(title = "U.S. Rape Arrests",
       subtitle = "By U.S. State",
       x= "Number of Arrests per 100,000")
```

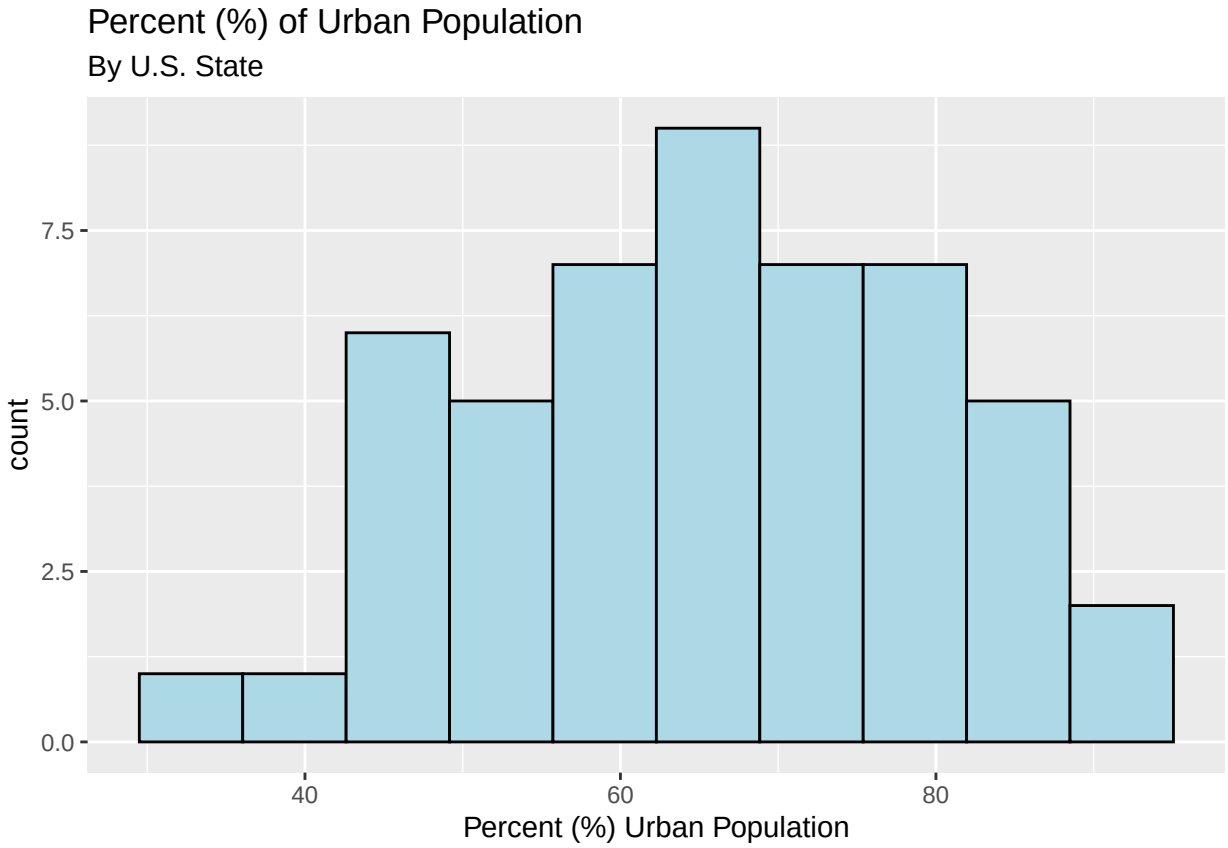


5. Create a histogram for the percent of urban population.

```
# base R
hist(USArrests$UrbanPop, main = "Percent (%) Urban Population",
     xlab = "Percent (%) Urban Population", prob = TRUE)
```




```
# ggplot2
ggplot(data=USArrests, aes(x=UrbanPop)) +
  geom_histogram(color = "Black", fill = "lightblue", bins = 10)+
  labs(title = "Percent (%) of Urban Population",
       subtitle = "By U.S. State",
       x= "Percent (%) Urban Population")
```



Your project

Now it's your turn. Use the `ggplot2` tools you used today to conduct data analysis for one of your final seminar papers.

1. Create a Git repository for your project.
2. Upload the dataset you are planning to use. If are you planning to collect original data, please provide all the variables that will be in that original dataseet.
3. How will these variables help answer your question?
4. Consider the variables. Which variables would you want to highlight? How would you visually represent them? Plan to create at least three descriptive graphs. Some suggestions:
    - If you have time-related variable, create a line graph showing changes over time!
    - If your observations can be separated by certain groups, create bar graphs or facets~
5. Given what you brainstormed in Question 4, create plots using the `ggplot2` package. Label all axes and title each graph. Provide descriptions for each graph.

Submit

Email me (mshieh2@wisc.edu) the link to your `ps811-exercises` repository when you are done.