HUDM 5026 - Introduction to Data Analysis and Graphics in R
POTD 04 – Exploratory Data Analysis

General instructions for POTDs:

- Write up your solutions and examples in a .Rmd file and knit it and submit an html or pdf file.
- Clearly label each part by number and letter, if applicable.
- Include plenty of comments in your code.
- The file should run without any errors from top to bottom.
- The write-up is due before the next class meeting.
- Although you may work collaboratively with others in class, each individual will turn in their own assignment.

Run the following code to create a tibble using the `state.x77` data.

```
library(tidyverse)

# Create a tibble called dat using state.x77
dat <- data.frame(state.x77)
dat <- tibble(dat)

# Add the region and division and abbreviation information
dat$Region <- state.region
dat$Division <- state.division
dat$abbrev <- state.abb
print(dat, n = 50, width = 100)
```

Task 1 *Create a histogram of life expectancy (i.e., including data from all 50 states). Did you use* `binwidth =` *or* `bins =` *and how did you decide on the value?*

Task 2 *Use* `group_by()` *and* `summary()` *to determine mean life expectancies and HS graduation rates for each region.*

Task 3 *Use frequency polygons to create histograms of life expectancy by color based on state region, both by count and using density. Which do you prefer for these data and why?*

Task 4 *Create conditional boxplots of life expectancy by state region (i.e., using the same information as in the previous task). What are the pros/cons to using boxplots vs frequency polygons here and in general?*

Task 5 *Replicate the last plot but also order the boxplots by their medians.*

Task 6 *Create a scatterplot of HS graduation rate on the horizontal axis and life expectancy on the vertical axis.*

Task 7 *Replicate the scatterplot directly above using two-letter state abbreviations instead of points for plotting.*

Task 8 *Note that AK and NV are outlying. Use* `filter()`*, as in lines 163–169 of '04_ Code.R', to color those two points blue.*

Task 9 *Import your project data into R and make a plot of some kind.*