Practice ggplot2

Math 241, Week 1

```
# it's good practice to check that all the packages required are loaded and installed
libs <- c('tidyverse','dplyr','ggplot2','knitr','viridis','mdsr', 'macleish','babynames')
for(l in libs){
  if(!require(l,character.only = TRUE, quietly = TRUE)){
    message( sprintf('Did not have the required package << %s >> installed. Downloading now ... ',l))
    install.packages(l)
  }
  library(l, character.only = TRUE, quietly = TRUE)
}
```

Goals of this in-class activity:

- Practice creating and refining graphs with `ggplot2`.
- Consider the strengths and weaknesses of various `geom`s and `aes`thetics for telling a data story.

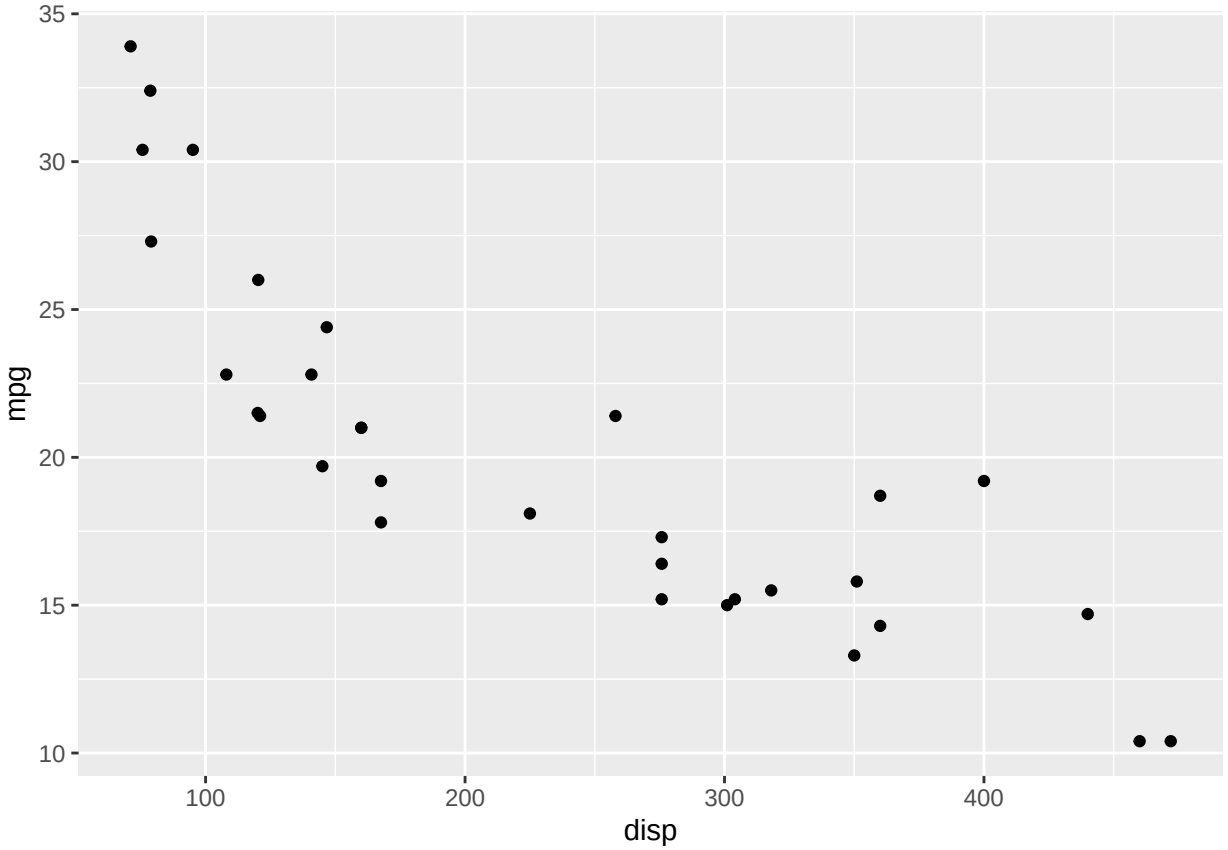
Notes:

- When creating your graphs, consider context (i.e. axis labels, title, . . .)!
- If I provide partially completed code, I will put `eval = FALSE` in the chunk. Make sure to change that to `eval = TRUE` once you have completed the code in the chunk.
- Be prepared to ask for help from me, Tory, and your classmates! We scratch the surface of `ggplot2` in class. But I encourage you to really dig in and make your graphs your own (i.e. don't rely on defaults).

Problem 1 (Easy):

Consider the following data graphic.

```
ggplot(mtcars, aes(x = disp, y = mpg)) +
geom_point()
```



The am variable takes the value 0 if the car has automatic transmission and 1 if the car has manual transmission. How could you differentiate the cars in the graphic based on their transmission type?

Problem 2 (Easy):

Angelica Schuyler Church (1756–1814) was the daughter of New York Governer Philip Schuyler and sister of Elizabeth Schuyler Hamilton. Angelica, New York was named after her.

Using the `babynames` package generate a plot of the reported proportion of babies born with the name Angelica over time and interpret the figure.

```
data(babynames) # this will explicitly ask R to load the babynames dataset to your environment
```

Problem 3 (Medium):

The `macleish` package contains weather data collected every 10 minutes in 2015 from two weather stations in Whately, MA.

Using ggplot2, create a data graphic that displays the average temperature over each 10-minute interval (temperature) as a function of time (when).

Problem 4 (Medium):

The data set `MLB_teams` in the `mdsr` package contains information about Major League Baseball teams from 2008–2014. There are several quantitative and a few categorical variables present. See how many variables

you can illustrate on a single plot in R. The current record is 7. (Note: **This is not good graphical practice**—it is merely an exercise to help you understand how to use visual cues and aesthetics!)

Problem 5 (Medium):

Use the `MLB_teams` data in the `mdsr` package again to create an informative data graphic that illustrates the relationship between winning percentage and payroll in context.

Problem 6 (Hard):

Use the function `make_babynames_dist()` in the `mdsr` package to recreate the "Deadest Names" graphic from FiveThirtyEight.