Data Wrangling with dplyr

Math 241, Week 4

```
# it's good practice to check that all the packages required are loaded and installed
libs <- c('tidyverse','knitr','viridis','mosaicData','babynames','mdsr','Lahman','nycflights13')
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 data wrangling.

Notes:

- Be prepared to ask for help from me, Tory, and your classmates!

Problem 1 (Medium):

The `Violations` data set in the `mdsr` package contains information regarding the outcome of health inspections of restaurants in New York City. Use these data to calculate the median violation score by zip code for zip codes in Manhattan with 50 or more inspections. What pattern do you see between the number of inspections and the median score?

```
zips <- Violations %>%
  select(boro, score, zipcode) %>%
  na.omit() %>%
  filter(boro == "MANHATTAN") %>%
```

Problem 2 (Medium):

The Major League Baseball Angels have at times been called the California Angels (CAL), the Anaheim Angels (ANA), and the Los Angeles Angels of Anaheim (LAA). Using the Teams data frame in the `Lahman` package:

1. Find the 10 most successful seasons in Angels history, defining "successful" as the fraction of regular-season games won in the year. In the table you create, include the `yearID`, `teamID`, `lgID`, `W`, `L`, and `WSWin`. See the documentation for Teams for the definition of these variables.
2. Have the Angels ever won the World Series? If so, when?

Problem 3 (Medium):

Use the `nycflights13` package and the `flights` data frame to answer the following question:

- What plane (specified by the tailnum variable) traveled the most times from New York City airports in 2013?
- Plot the number of trips per week over the year.

```
library(nycflights13)
library(lubridate)
```