More data Wrangling with tidyR

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 and joins with tidyR

Notes:

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

Problem 1 (Medium):

Consider the number of home runs hit (HR) and home runs allowed (HRA) for the Chicago Cubs (CHN) baseball team. Reshape the Teams data from the Lahman package into "long" format and plot a time series conditioned on whether the HRs that involved the Cubs were hit by them or allowed by them.

```
Teams %>%
  filter(teamID == "CHN") %>%
  select(yearID, HR, HRA) %>%
  pivot_longer()
```

Problem 2 (Medium):

Use the `nycflights13` package and the `flights` and `planes` tables to answer the following questions:

a. How many planes have a missing date of manufacture?

```
library(nycflights13)
planes2 <- select(planes, tailnum, year, manufacturer)
flights2 <- select(flights, tailnum)
nyc_flights <- left_join()
```

b. What are the five most common manufacturers?

```
nyc_flights %>%
```

Problem 3 (Medium):

Use the `nycflights13` package and the `flights` and `planes` tables to answer the following questions:

a. What is the oldest plane (specified by the `tailnum` variable) that flew from New York City airports in 2013?

b. How many airplanes that flew from New York City are included in the planes table?

Problem 4 (Medium):

The `knitr` package allows the analyst to display nicely formatted tables and results when outputting to pdf files. Use the following code chunk as an example to create a similar display for the `penguins` dataset, in the `palmerpenguins` package, instead (you can model penguins' `body_mass_g` as a function of their `flipper_length_mm` and `sex`):

```
library(palmerpenguins)
mod <- broom::tidy(lm(cesd ~ mcs + sex, data = HELPrct))
knitr::kable(
  mod,
  digits = c(0, 2, 2, 2, 4),
  caption = "Regression model from HELP clinical trial.",
  longtable = TRUE
)
```

Table 1: Regression model from HELP clinical trial.

term	estimate	std.error	statistic	p.value
(Intercept)	55.79	1.31	42.62	0.0000
mcs	-0.65	0.03	-19.48	0.0000
sexmale	-2.95	1.01	-2.91	0.0038