Homework 1

BIOF 339

Installation and setup

1. Install R and RStudio on your computer following the instructions in class. If you have trouble, reach out to me on Slack
2. Install the **pacman** package by running the following R command in the RStudio console pane: `install.packages("pacman")`. Make sure of spelling and case.
3. Now install a few more packages using the **pacman** package. This simplifies some aspects of package installation and loading. Open a new script window and type the following code in that window:

```r
library(pacman)
p_load(char = c("tidyverse", 'broom', 'janitor', 'readxl'))
```

This installs and loads the meta-package **tidyverse**, and the packages **broom**, **janitor** and **readxl** into R.

> *`p_load` is a function within the **pacman** package. You can think of functions as recipes and packages as recipe books, if that helps.*
>
> *The `p_load` function installs a package if you don't have it on your computer, and then loads it. It just loads the package if you already have it installed. Note that you only need to install a package **once** on a computer.*

> *Note that I've interchanged single and double quotes in the code snippet. Please feel free to use either single- or double-quotes or a mixture, as long as the quotes are properly paired by type.*

R Markdown practice (10 pts)

We will continue with the `airquality` dataset we worked with in class. Learn more about this dataset by typing either `help(airquality)` or `?airquality` at the console prompt.

In this homework you will create a report and a presentation on this dataset using RMarkdown. You will do this in the same RStudio project you created this week.

Your documents will incorporate the following 3 code snippets (in order) as R chunks and one piece of code inline. If I'm calling a package you do not have installed, use the `p_load` function above to install it.

Snippet 1

```r
library(tidyverse)
library(knitr)
avg_temp_by_month <- airquality %>%
  group_by(Month) %>%
  summarize(avgTemp = mean(Temp, na.rm=T))
kable(avg_temp_by_month)
```

Snippet 2

```
ggplot(avg_temp_by_month, aes(x = Month, y = avgTemp)) +
  geom_point() +
  geom_line(color = 'blue') +
  labs(y = 'Average Temperature (F)')
```

Snippet 3

```
ggplot(airquality, aes(x = Wind, y = Ozone)) +
  geom_point() +
  geom_smooth(color = 'blue', se = FALSE)
```

Inline snippet

`max(airquality$Ozone, na.rm = T)`, which is the maximum recorded Ozone level. Incorporate into a sentence.

Markdown

There is a Markdown Quick Reference available in under the Help menu to get you started.

Assignment 1

Background reading

- Read the first 3 sections of the document at https://rmarkdown.rstudio.com/lesson-1.html (https://rmarkdown.rstudio.com/lesson-1.html).
- To go into more detail on each style below see:
  - HTML document: https://bookdown.org/yihui/rmarkdown/html-document.html (https://bookdown.org/yihui/rmarkdown/html-document.html)
  - HTML ioslide presentation: https://bookdown.org/yihui/rmarkdown/ioslides-presentation.html (https://bookdown.org/yihui/rmarkdown/ioslides-presentation.html)

Work (10 pts)

1. Write a RMarkdown report incorporating these code snippets and create a story around them. You will set the output to HTML. Call this file `<your name>_HW1_report.Rmd`
2. Write a RMarkdown presentation incorporating these same code snippets. The output format should be HTML(ioslides). Call this file `<your name>_HW1_slides.Rmd`

Submit both the Rmds and the corresponding HTML files to Canvas.

Assignment 2 (10 points)

The following section contains a templated R Markdown file that you can copy into a fresh R Markdown document in RStudio. You are expected to use the help system in R/RStudio as well as Google, if need be, to fill in the blanks below.

1. The symbol `==` checks for equality between two objects, and returns `TRUE` or `FALSE`
2. Just as we saw `is.character` and `is.numeric`, you can check missing values with `is.na`, which gives a `TRUE` everytime it encounters a missing value in a data array, and `FALSE` otherwise. Internally in R, `TRUE = 1`.

```
---
title: Homework 1, Part 2
author: _____  _____
date: "BIOF 339"
---


```{r, echo = FALSE, eval=TRUE}
knitr::opts_chunk$set(message=FALSE, warning=FALSE)
```


Descriptive statistics

We'll start with the `airquality` data set that is in-built in R.

1. The average temperature in June was `r _____(airquality$_____[airquality$Month==6],
 na.rm=TRUE)`.
2. Solar radiation data is missing on `r sum(is._____(airquality[,"Solar.R"]))` days, or
in `r 100 * sum(is._____(airquality[,"Solar.R"]))/_____(airquality)` percent of all th
e days collected.

We can also visualize the missing data patterns in this data set.

```{r, echo = TRUE, eval=TRUE}
library(pacman)
p_load('naniar') # This is a package for missing data
vis_miss(_____) # see the documentation for vis_miss
```


Let's grab a more interesting data set. We will download and use the [Palmer Station pen
guins data set](doi:10.1371/journal.pone.0090081), which is in the form
of an R package on GitHub.

```{r, echo = TRUE, eval=TRUE}
library(pacman)
p_____('visdat') # Install and load visdat
p_install_gh('allisonhorst/palmerpenguins')
p_load('palmerpenguins')
vis_dat(penguins)
```
```

Icons made by Freepik from www.flaticon.com (https://www.flaticon.com/)