

Homework 1

Cory Costello

October 1, 2017

List of Steps in HW 1

- Read in dataset
 - The first chunk of code shows how to read in a .csv file. It uses the `seda` dataset from Lab 1 for the demonstration, and uses the `read.csv()` function.
- Subset data using `select`
 - The second chunk of code shows how to subset a dataset using `select` from the `dplyr` library. We subset the data such that it contains only math scores and state abbreviation.
- Making a table using `Kable`
 - The third chunk of code shows how to make a table using the `kable()` function from the `knitr` library.
- Getting a Histogram
 - The fourth chunk of code uses `hist()` from base graphics to produce a histogram of mean math scores.

First we need to read in the data. We'll be working with the `seda` dataset, which is currently in a csv format. The following chunk of code uses the `read.csv()` function to read in the `seda` dataset. The data is being saved into an object called `seda_data`

```
seda_data <- read.csv("seda.csv")
```

Next, we want to work with only a few columns of that dataset. SO, we can use the `select()` function from `dplyr` to select the columns we want to work with and save that subsetted dataset into a new column. We will select out the variables corresponding to math scores (means and standard errors) and the state abbreviation. The chunk of code below does this, by first calling the `dplyr` library, and then using the `select` function to select the state abbreviation variable (`stateabb`) and any variable that contains `math` in its name. Finally, it shows the first 6 rows of the new subsetted dataset in order to visually inspect that it worked correctly.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

math_states <- select(seda_data, stateabb, contains("math"))
head(math_states)
```

```
##   stateabb mean_link_math se_link_math
## 1      AL      220.6698      7.064998
## 2      AL           NA           NA
## 3      AL      233.1264      6.975586
## 4      AL      216.3612      2.008145
```

```
## 5      AL      223.0344      2.339266
## 6      AL      230.4161      2.290136
```

Next, we may want to put those first six rows into a formatted table. We can do so using the `kable()` function from the knitr library. The chunk of code below does this, by first calling the knitr library, and then using the `head()` function within the `kable()` function to produce a table of the first 6 rows.

```
library(knitr)

kable(head(math_states))
```

stateabb	mean_link_math	se_link_math
AL	220.6698	7.064998
AL	NA	NA
AL	233.1264	6.975586
AL	216.3612	2.008145
AL	223.0344	2.339266
AL	230.4161	2.290136

Next, we'll make a histogram of math scores using the base `hist()` function.

```
hist(math_states$mean_link_math)
```

