Homework 3

Matthew Runyon

9/14/2020

Problem #1: Write a function to calculate the mean:

I am going to call this new function that I am creating `average()`. To write this function, I use the `function()` function, with `x` as the input. The mean is simply calculated as the sum of all data values divided by the number of observed values, so I will define this function using the `sum()` and `length()` functions. Finally, I will use the `return()` function to give the output of this calculation.

```
average<-function(x){
  average<-(sum(x)/length(x))
  return(average)
}
```

As a note, this function will not work with missing data values, as we will see in Problem #3, so additional data manipulations are required in that circumstance.

Problem #2: Find the range of the variable Year from the "Key Crop Yields" data set.

First, I downloaded the file and saved it as a .csv file in my homework repository. Then I imported it to R using the `read.csv()` function and created the data frame `Yields` using this data.

```
Yields<-read.csv("key_crop_yields_MJR.csv", head=T)
```

To calculate the range, we can use the function `range()`, which will return the largest and smallest value for a given set of data. To limit this to the variable `Year`, we need to specify this in the primary argument of this function using the `$`.

```
range(Yields$Year)
```

```
## [1] 1961 2018
```

Problem #3: Using the function created in Problem #1, calculate the mean of the variable "Maize (tonnes per hectare)":

Completing this problem is very similar to Problem #2. We simply need to use the created `average()` function, here specifying the variable to use (Maize (tonnes per hectare)) with the `$` in the primary argument of this function. However, my function doesn't account for `NA` values, so I am going to remove these from this data set using the `na.omit()` function, demonstrated as follows.

```
Yields<-read.csv("key_crop_yields_MJR.csv", head=T)
maize_yields_omit<-na.omit((c(Yields$Maize..tonnes.per.hectare.)))
```

Now we can use the `maize_yields_omit` as the data for the `average()` function that I created to act upon.

```
average(maize_yields_omit)
```

```
## [1] 3.024496
```

If we double check this using the actual `mean()` function on the original data set, setting the `na.rm` option equal to `TRUE`, we can see that the mean was calculated correctly with my personally created function.

```
mean(Yields$Maize..tonnes.per.hectare., na.rm=T)
```

```
## [1] 3.024496
```

Problem #4: Save a file with the name of the variable Entity:

Saving a file can be done with the `write.table()` function. Since I only want to save the variable Entity in this file, I will specify that column with the `$`. Finally, I will name the file `"entity.txt"`. Since I am working in my homework repository and this is where I want to save this file as well, I do not need to explicitly specify a file path.

```
write.table(Yields$Entity, file="entity.txt")
```