

....



My AP Stats class has started their first R programming assignment this week. I gave them the code for them to type in and play with. This will give them some experience with RStudio and basic function commands.

I have a total of six assignments for them to complete over the next few months. All my students have a laptop to use so there should be fewer issues getting up and running. The first thing we did was to download R and RStudio. Everyone was able to get RStudio up and running. I then went over some rules for coding and how the assignments will be submitted.

Some of the rules I went over were:

- Comment out your code with the # sign. Every function should have a comment.
- Create white space. It's easier for me to read and debug.
- If you are stuck: a) ask a classmate b) copy/paste error code into Google, c) ask Mr. Smith
- Make sure your name and all relevant information are at the top of your code.
- Expect to get stuck sometimes. Google and Stackoverflow are your friends. You will learn a lot.

After we all had RStudio working, I showed them how to install the tideverse package. This is the greatest package ever and allows me to teach the students on all kinds of larger data. I'll go into more detail the next lesson on using dplyr to filter and select from a data frame.

For this first assignment, I'm using the data from our book on page 35.

Here is the code for the first assignment and the output.

```
# My name is _____
# This is my first programming assignment for AP Stats and I will copy and see
# if everything runs properly
# November 11, 2019

# I need to comment out everything using the "#"
# This lesson is from my website at saturnscience.com

# web link here to see details
# http://www.saturnscience.com/category/r/

##### Assignment 1---students type in the data
#####
# Everything here works for the latest version of R and RStudio

## The general form of a command will look like this:
## note to myself
## myGraph <- ggplot(myData, aes(variable for x axis, variable for y axis)) +
## geom()
## You can also use =, its the same as -<
## NOTE: DO NOT make variables names with a space, use one word or two
## connected with a period "."
```

```

## Here I enter the data from page 35
## The "c" function combines the data into a vector

#### Please load dplyr and ggplot2 now. ####

foreign.born=c(2.8,7.0,15.1,3.8,27.2,10.3,12.9,8.1,18.9,9.2,
16.3,5.6,13.8,4.2,3.8,
               6.3,2.7,2.9,3.2,12.2,14.1,5.9,6.6,1.8,3.3,1.9,5.6,19.1,5.4,20.1,
               10.1,21.6,6.9,2.1,3.6,4.9,9.7,5.1,12.6,4.1,2.2,3.9,15.9,8.3,
               3.9,10.1,12.4,1.2,4.4,2.7)

summary(foreign.born) # Gives the five number summary.

str(foreign.born) # the str function shows me the type of structure of the data.

fivenum(foreign.born) # gives the five number summary

mean(foreign.born) # just shows the mean

head(foreign.born, n=12) # shows the first 12, pick n. Used with large data
files.

tail(foreign.born) # shows the end of the data. You can pick n or leave it
alone.

plot(foreign.born) # this is R's generic scatter plot function and only shows
basic information.
# we will use this more later.

hist(foreign.born) # This is base R basic histogram function.

# Below is ggplot's better graphing abilities
ggplot() + aes(foreign.born)+
  geom_histogram(binwidth = 2.5)

# I change the variable name so I don't confuse with the prior graphs
foreign.born3=c(2.8,7.0,15.1,3.8,27.2,10.3,12.9,8.1,18.9,9.
2,16.3,5.6,13.8,4.2,3.8,
               6.3,2.7,2.9,3.2,12.2,14.1,5.9,6.6,1.8,3.3,1.9,5.6,19.1,5.4,20.1,
               10.1,21.6,6.9,2.1,3.6,4.9,9.7,5.1,12.6,4.1,2.2,3.9,15.9,8.3,
               3.9,10.1,12.4,1.2,4.4,2.7)

# This is a histogram with base R
hist(foreign.born3, breaks = 10,
     main = "Histogram with Base Graphics",
     ylim = c(0,15))

# check the structure
str(foreign.born3)

# make sure it's a data frame by changing to a data.frame.
fb3=as.data.frame(foreign.born3)

# I check to see the structur of fb3
str(fb3)

```

```

# I use ggplot to make a histogram similar to the book's histogram
ggplot(fb3,aes(x=foreign.born3))+
  geom_histogram(color="black",fill="orange",binwidth = 3)+
  labs(x="Percent of foreign born residents",y="Number of States")+
  geom_density()

# I can add a density curve to the histogtam
ggplot(fb3,aes(x=foreign.born3))+
  geom_histogram(aes(y=..density..),color="black",fill="orange",binwidth = 3)+
  labs(x="Percent of foreign born residents",y="Density of States")+
  geom_density(alpha=0.2,fill="#FF6666")

# Same histogram but I just change the colors a bit.
ggplot(fb3, aes(x=foreign.born3)) +
  geom_histogram(aes(y=..density..),
                binwidth=3,
                colour="black", fill="white") +
  geom_density(alpha=.2, fill="#FF6666")

# use control-l to clear the console

```

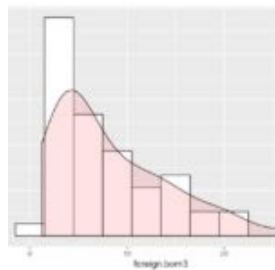
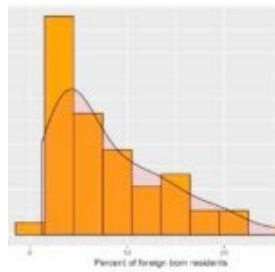
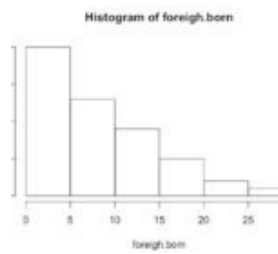
### Some of the output:

```

> ##### Please load dplyr and ggplot2 now. #####
>
> foreign.born=c(2.8,7.0,15.1,3.8,27.2,10.3,12.9,8.1,18.9,9.2,
16.3,5.6,13.8,4.2,3.8,
+               6.3,2.7,2.9,3.2,12.2,14.1,5.9,6.6,1.8,3.3,1.9,5.6,19.1,5.4,
20.1,
+               10.1,21.6,6.9,2.1,3.6,4.9,9.7,5.1,12.6,4.1,2.2,3.9,15.9,8.3,
+               3.9,10.1,12.4,1.2,4.4,2.7)
>
> summary(foreign.born) # Gives the five number summary.
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.200  3.800   6.100   8.316 12.350  27.200
>
> str(foreign.born) # the str function shows me the type of structure of the
data.
 num [1:50] 2.8 7 15.1 3.8 27.2 10.3 12.9 8.1 18.9 9.2 ...
>
> fivenum(foreign.born) # gives the five number summary
[1] 1.2 3.8 6.1 12.4 27.2
>
> mean(foreign.born) # just shows the mean
[1] 8.316
>
> head(foreign.born, n=12) # shows the first 12, pick n. Used with large data
files.
[1] 2.8 7.0 15.1 3.8 27.2 10.3 12.9 8.1 18.9 9.2 16.3 5.6
>
> tail(foreign.born) # shows the end of the data. You can pick n or leave it
alone.
[1] 3.9 10.1 12.4 1.2 4.4 2.7

```

Here are some of the plots using ggplot2



We have completed Unit 4 and will start Unit 5 next week. We are where we need to be at this time of the year. At this rate, we'll finish the class on time and have a few weeks to review for the exam in May 2020.

---