Getting Started with R: Basics

1. Create a numeric vector named **n1** containing integers from 1 to 20.

   Create a numeric vector named **n2** containing integers from 20 down to 1

   Create a numeric vector named **n3** containing integers from 1 to 20 with step size 2. (For this use the *seq* function, which works much like the range function in Python.)

   Create a character vector named **colors** containing the colors "red", "blue", "green" "yellow" and "black".

   Create a logical vector named **claims** containing the values *TRUE* and *FALSE*.

   Display the length of each vector above.

   Look in the environment and notice how the information on type, data structure and beginning values is displayed. What is the index of the first element of a vector?

2. Add, subtract, multiply and divide the first element of **n1** by the first element of **n2**

   Use the ordinary operators (+,-,*,/) to add, subtract, multiply and divide **n1** by **n2**.

   Use the ordinary operators (+,-,*,/) to add, subtract, multiply and divide **n1[1]** by **n2**.

   Use comparison operators <,==,> to compare the first element of **n1** to the first element of **n2**.

   Use comparison operators <,==,> to compare **n1** to **n2**.

   Use comparison operators <,==,> to compare **n1[1]** to **n2**.

   Type in the following: all(n1<n2)

   Type in the following: any(n1<n2)

   What governs whether the result of vector operations is a vector or a single value?

3. Combine the vectors **n1, n2** and **n3** using the *c( )* function and store in a variable named **numbers**.

   Display **numbers**. What is its type? (Use the *class()* function). What are its dimensions?

   Use *data.frame()* to create a data frame from **n1, n2** and **n3** named **numbers2**.

   Display **numbers2**. What is its type? (Use the *class()* function). What are its dimensions? (You can get this with *nrow, ncol*).

   Describe how the combining process of *data.frame()* differs from *c()*. Was data modified to fit? What does this tell you about a constraint on data frames?

4. Combine the vectors **n1, colors** and **claims** into a variable named **combined** using the *c( )* function. Was data modified to fit? What does this tell you about a constraint on vectors?

   Combine the vectors **n1, colors** and **claims** into a data frame named **my.data.** Briefly describe what resulted when R combined the data.

   In the environment, expand the entry for **my.data** by clicking on its triangle. What is the name and type of the middle column?

   Use the names() function to display the names of the rows and columns.

5. Create a new vector named **colors.v** to hold the contents of the column color of **my.data**. (Use a slice [,2] to select all rows and second column). Display type of **colors.v**.

   Now create a new data frame named **my.data2** with the same code you used in #4 to create **my.data**, except this time use *stringsAsFactors=F*.

   Create a new vector named **colors2.v** to hold the contents of the column color of my.data2. Print its type. What did you learn?

   Create a new data frame named **colors.df** to hold the contents of the column **color** of **my.data**. Print its type to verify it is a data frame. Hint: use a slice with column name ["colors"] instead of with indices.

   Execute the following command:

   colors3 <- my.data$colors

   How is **colors3** related to **colors.df?** Is it more like **colors2.v** or **colors.df** ?

6. Create a new vector named **d** to hold 2 dates: the current date (this is *Sys.Date()*) and May 20th, 2016

   Optional: Add the number of days in the year up to May 20th, 2016 to the current date. (You can use the lubridate package, the yday function)

   Subtract the May date from the current date

   Subtract the current date from the May date

7. Execute the following two commands to draw a histogram:

   hist(numbers2["n1"])

   hist(numbers2[,1])

   Explain what happened.

   Redraw the plot, this time set the argument: *breaks=10*

   Was the plot produced that of a uniform distribution or a normal distribution?

8. Execute the following three commands:

```
plot(numbers)

plot(my.data)

plot(my.data2)
```

Briefly summarize how *plot()* handles character data and plots vector as opposed to data frame. Which one gives an error?

9. Use ggplot2 to execute the following commands (If you have not already installed, first use *install.packages()*).

```
set.seed(123)
num <- rnorm(20)
hist(num, breaks=10)
my.data2$num <- num
library("ggplot2")
g<- ggplot(data=my.data, aes(x=num))
g + geom_histogram(aes(fill=colors)) #default position and binwidth
g + geom_histogram(aes(fill=colors), binwidth=1) #default position is stack
g + geom_histogram(aes(fill=colors),position = "dodge", binwidth=1)
```

From the last two plots (ie: stack and dodge) determine the count of items of each color group in the bin from zero to one. (Your answer should be the same). In your opinion, is "stack" or "dodge" easier for obtaining the count of different groups?

Resources:

R Language Fundamentals (Steven Buechler)
http://www3.nd.edu/~steve/Rcourse/Lecture1v2.pdf

The Ultimate R Cheat Sheet (Andreas Hamann)
https://sites.ualberta.ca/~ahamann/teaching/renr690/R_Cheat_Data.pdf

Google R Cheat Sheets
https://drive.google.com/drive/u/0/folders/0ByIrJAE4KMTtcVBmdm1BOEZoeEk

GGPlot2 Cheat Sheet – Rstudio (www.rstudio.com)
https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf