**Session 1: 1\_1 Getting your data.R**  
(1) Import the excel sheet Error\_data from Homophone\_ExpData

Store it in an object called Homophone\_Errors

(2) What is the name of the 5th column of Homophone\_Errors?  
  
  
  
  
  
(3) What is the value of the 15th row in the 5th column?

(4) Rename the first column of Homophone Error “Subject”

(5) Rename the first column of Homophone Error “Statistics”

(6) Check what is in columns 164 and 165

(7) Check for any junk columns in Homophone\_Errors

(8) Get the Mean and Standard Deviation for reaction time data from the 5th column in Homophone\_RT

(9) Complete an internet search to find out how to create a histogram in R.

When you have found the code, run it to create a histogram of the data in   
Homophone\_RT$e\_10.homophone

(10) In the Homophone\_RT data set, each participant is on one row. How would you calculate the average RT for one participant? That is, how would get the average/mean for a row?  
Write out that code and try it for participant 6. You can try an internet search to help!

(11) Generate a histogram, density plot and qq/quantile plots for the Twitches column in Mydata, Make sure they all display on the same window

**Session 2: 1\_2 Manipulating Data.R**

1. Rename columns 5 and 6 in the Homophone\_RT\_Long data so that column 5 is labelled “Item” and column 6 is labelled “WordCondition”
2. Can you do the above in ONE line of code?

1. Use dcast to generate ‘wide format’ data for each participant, with their average reaction time to each Item (collapsing across condition)
2. Use melt to create a long form data set for the Homophone\_Errors data.   
   You do not need to make the Item/Word Condition labels.
3. From Mydata, select the columns SubNo, Trial, ACC, RT, Congruence and Twitches. Store this in a new object called ‘Mydata\_1’.

1. Using Mydata\_1, use melt and dcast to create a file with average RT for each Participant for the Congruent and Incongruent conditions in the ‘Congruence’ column.

1. Create a boxplot of the data above, showing average RT for the Congruent and Incongruent conditions.

**Session 3: 1\_3 Descriptive Statistics TTests Correlation**

1. Generate the average RT for each Item from the Homophone\_RT\_Long data set – use ddply and summarise
2. Generate the average RT for each Item from the Homophone\_RT\_Long data set – use the psych package and describeBy
3. Generate a dataset that has the average Participant RT for the two conditions in WordCondition.   
   Hint: you can use ddply and summarise.   
   Store it in ‘temp’

1. From Mydata, Generate a dataset that has the average Participant RT for the two different experiments in Task.  
   Store it in ‘temp’.

1. Create a boxplot for the above data, with labelled axes and a title

1. Read in the following data / use the following line of code:

movies <- read.csv(url("http://s3.amazonaws.com/dcwoods2717/movies.csv"))

1. Create a new columns called ‘profit’. This is the column ‘gross’ minus the ‘budget’. Hint: you can use mutate from dplyr to do this
2. Correlate the ‘ratings’ with the ‘profit’. What do you find?

1. How could you visualize/create a graph for this data?

**Session 4: 1\_4 Chi Square**

1. Complete a chi square analysis for the association between sex and smoking

**Session 5: 2\_1 regression / linear models**

(1) Write some code for a regression that predicts the movie’s profits from its budget, year and cast facebook likes. Save in lm.mov.2

(2) Now write code for a regression that predicts the movie’s profits from its budget, year, cast facebook likes and genre. Save in lm.mov.3

(3) Now compare lm.mov.2 with lm.mov.3  
Does including the genre improve the prediction of the movies profit?

Why?

(4) You should now have four models (lm.mov.1, 2 3 and 4). Look at the plots for each one – do the residuals look OK / normally distributed?   
Make sure you use plot and hist(resid()).

(6) Go back to the code we used for correlation. Look at the correlation between the predictor variables we are using for our movie regression.  
budget, year, cast\_facebook\_likes, genre  
Hint: use corrplot to look at many correlations at the same time

(7) Are any of the predictors correlated? Why?

1. Plot a line graph just for the predictor Year. What do you see?