**Qualitative data vs. Quantitative data**

1. What is the difference between qualitative data and quantitative data?

Quantitative data uses numerical values to quantify a data set like the average annual GDP for each country on a continent, while qualitative data consists of non-numerical info that can be used to classify things such as which countries are on said continent.

Today we will explore graphs useful for displaying qualitative data.  The aphasia.csv data that you have loaded is a data set I got from **A First Course in Statistics**, by McClave and Sincich, the authors got the data from an article in the Journal of Communication Disorders. Aphasia is a disorder in which the patient shows some loss of communication ability, whether written or spoken.  The researchers wanted to see if the three different types of aphasia happened in about equal frequencies.

**Exploring the data**

2. *# to make objects within dataframes accessible*

*attach(aphasia) #now you won’t need “aphasia$Type\_of\_Aphasia”, just “Type\_of\_Aphasia”*

*names(aphasia)*

*#summarize qualitative data*

*table(Type\_of\_Aphasia)*

Then paste your code and the result into the blank below.

aphasia<-read.csv(file.choose(), header=TRUE)

attach(aphasia)

names(aphasia)

table(Type\_of\_Aphasia)

> attach(aphasia)

>

> names(aphasia)

[1] "Subject" "Type\_of\_Aphasia"

>

> table(Type\_of\_Aphasia)

Type\_of\_Aphasia

Anomic Broca's Conduction

10 5 7

**Pie charts**

3. There is a command to make pie charts in the base R version. However, the command to make pie charts requires counts, but the data list Type\_of\_Aphasia is a list of values.

*Type\_of\_Aphasia #just typing the name of the list gives a list of values*

*table(Type\_of\_Aphasia) #the command “table” turns the list into a table of values with corresponding labels*

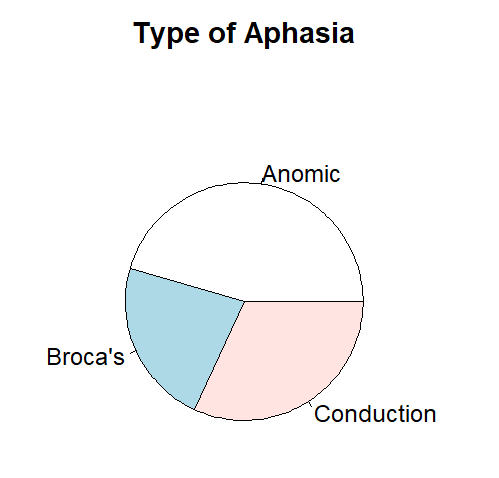
*pie(table(Type\_of\_Aphasia)) #takes the table and makes a pie chart with it*

*pie(table(Type\_of\_Aphasia), main=”Type of Aphasia”) #add a caption*

*#Note, when I copy and paste double quotes from Word, they paste wrong in R, you may have*

*#to type the command to avoid an error.*

See a pie chart displaying the proportions of each type of aphasia.  Insert the graph into this file:



What happens if you try this without making the table first?

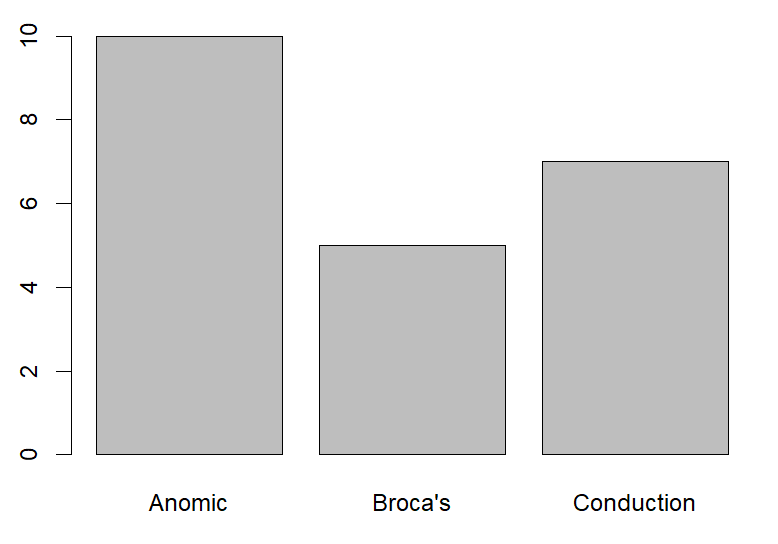
*pie(Type\_of\_Aphasia) #trying the command without the table() part*

Error in pie(Type\_of\_Aphasia) : ‘x’ values must be positive

**Bar Graphs**

4. A similar command will make a bar graph.

*barplot(table(Type\_of\_Aphasia)) #takes the table and makes a bar graph with it*



**Comparing the two graphs**

5. In above code, you created both a bar graph and a pie chart.  When is it better to use a bar graph, and when is it better to use a pie chart?

Pie charts are useful when you’re wanting to visualize parts of a whole while bar graphs are useful when trying to display/compare the numerical values of different categories.

**A more complicated file**

Look at the file titanic3.csv.  This file contains information on all the passengers who were on the Titanic. We will examine the genders of the passengers who survived vs. the passengers who died. As a first step, move the titanic3.csv to your work directory. Load the file and attach the data.

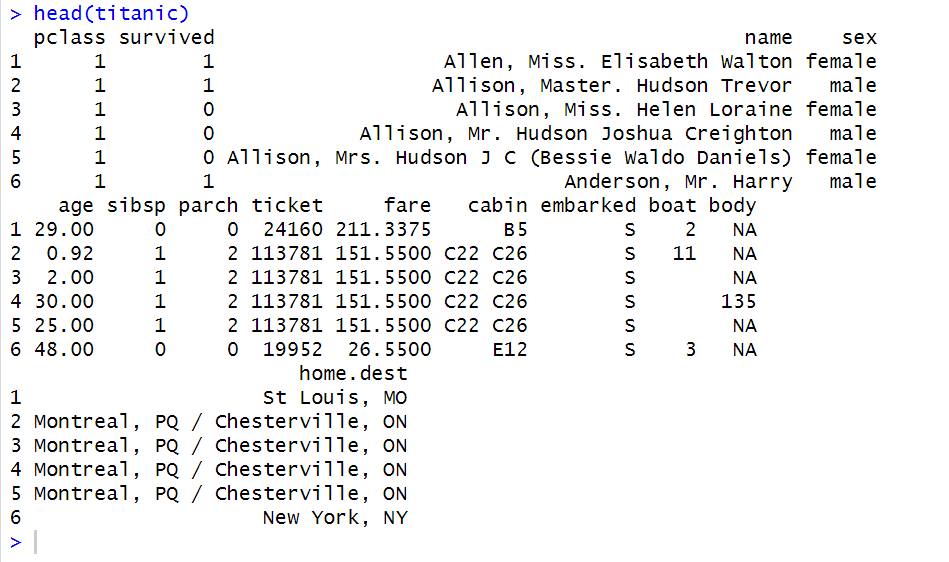
6. List the variable names. Name one variable that is qualitative, and one that is quantitative.

"pclass","survived","name","sex","age","sibsp","parch","ticket","fare","cabin","embarked","boat","body","home.dest"

Qualitative: name

Quantitative: fare

7. Type *head(titanic).* You will see the first few entries of the titanic file. Enter them here:



8. Examine the *survived* column. Notice that it contains numbers: 1 and 0. When dealing with qualitative data in which there are only two outcomes possible, it is common practice to use 1 to represent a “yes,” and 0 to represent a “no.” In this case, the question being asked is if the passenger survived the disaster. Did Miss Elizabeth Walton Allen survive?

She did indeed

9. Summarize sex vs. survival:

*table(survived,sex) #makes a 2x2 table of the frequencies of survival status vs. sex*

Now switch the order of the two variables:

*table(sex,survived)*

What is the effect of changing the order of the variables?

It swapped the x and y axes

*10.* Create a bar graph of survival by sex:

*table1 <- table(survived,sex) #name the table so that we can call it again later*

*table1 #see the result of the previous command*

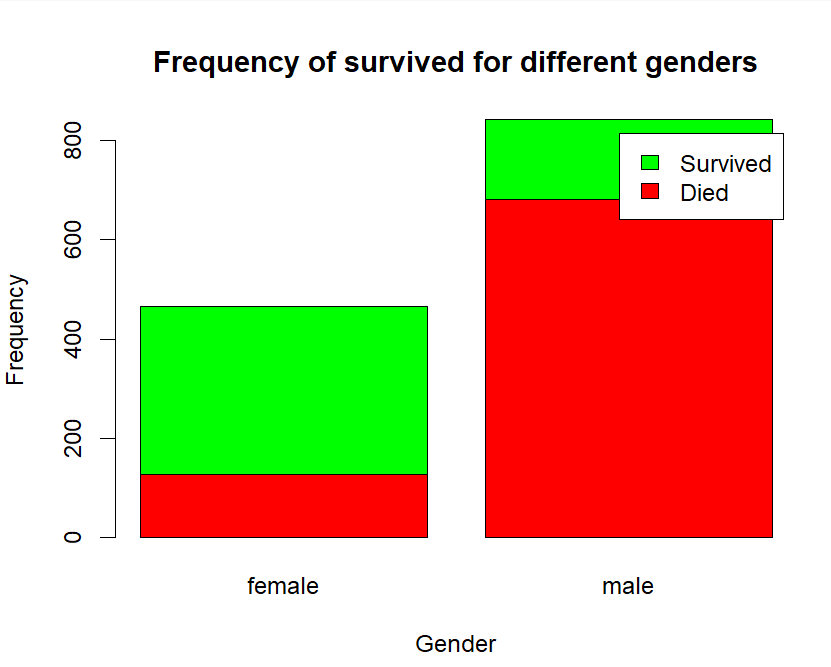
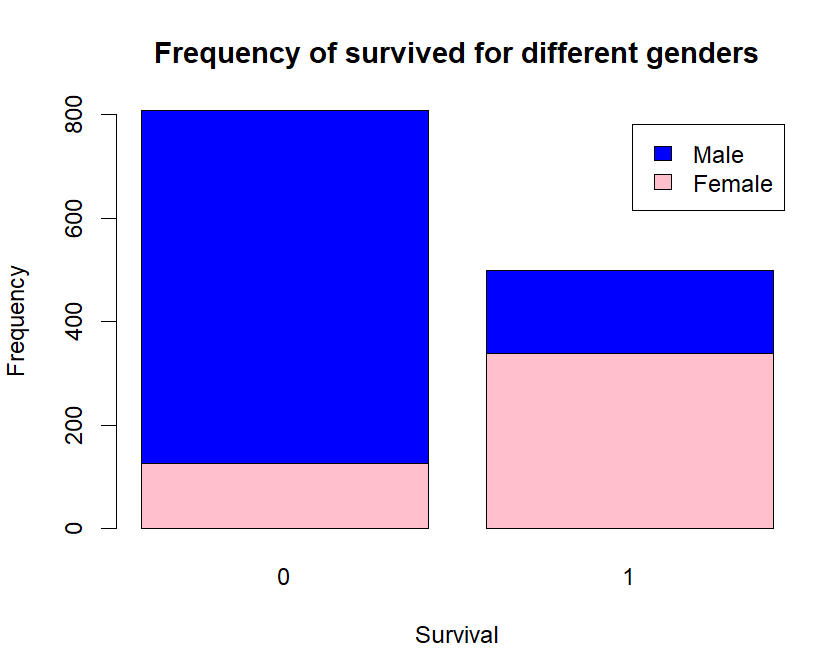
*barplot(table1) #uses the data in the table to make a bar chart*

*barplot(table1), legend=rownames(table(survived,sex))) #Adds a legend*

*barplot(table1, legend=c("Died","Survived"))* *#Changes legend to be more descriptive*

*barplot(table1,main="Frequency of survived for different genders", xlab="Gender", ylab="Frequency", col=c("red","green"), legend=c("Died","Survived")) #Add labels and colors*

Insert your final graph here, then create a graph that shows gender by survival and enter it also.

11. Create pie charts for the same data. Recall that a pie chart needs a different format of table.

*table1[1,] #extracts the first row*

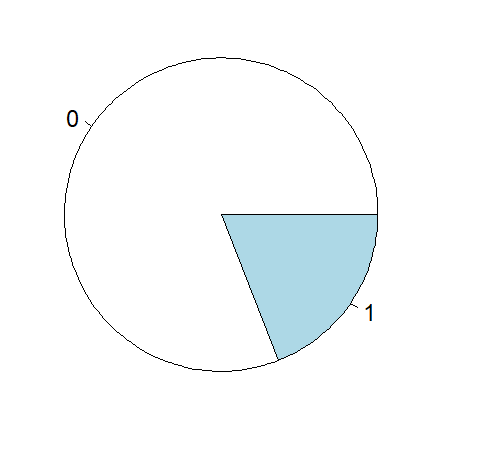
*table1[2,] #second row*

*table1[,1] #first column*

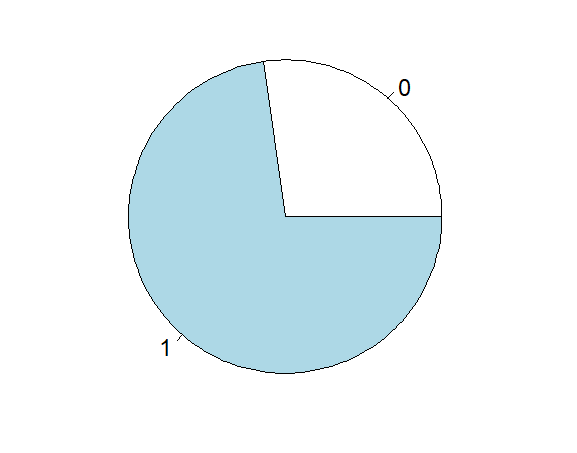
*table1[,2] #second column*

*pie(table1[1,]) #pie chart of the first row*  
From the commands above, figure out how to modify them create two pie charts. One of the charts shows survivors vs those who died for the men. The other shows survivors vs those who died for the women. Make sure to label your graphs. Paste them in this box:

MALE:



FEMALE:



The above sequences of commands use base R for creating graphs. Next time we will talk about using a package to do this, ggplot2.