

James Scruggs

M01255052

Qualitative data vs. Quantitative data

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

Qualitative data is descriptive and cannot be measured by numbers (Example: color, gender).

Quantitative data can be measured with numbers (Example: count, measure).

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.

Bar graphs

2. *# to make objects within dataframes accessible*

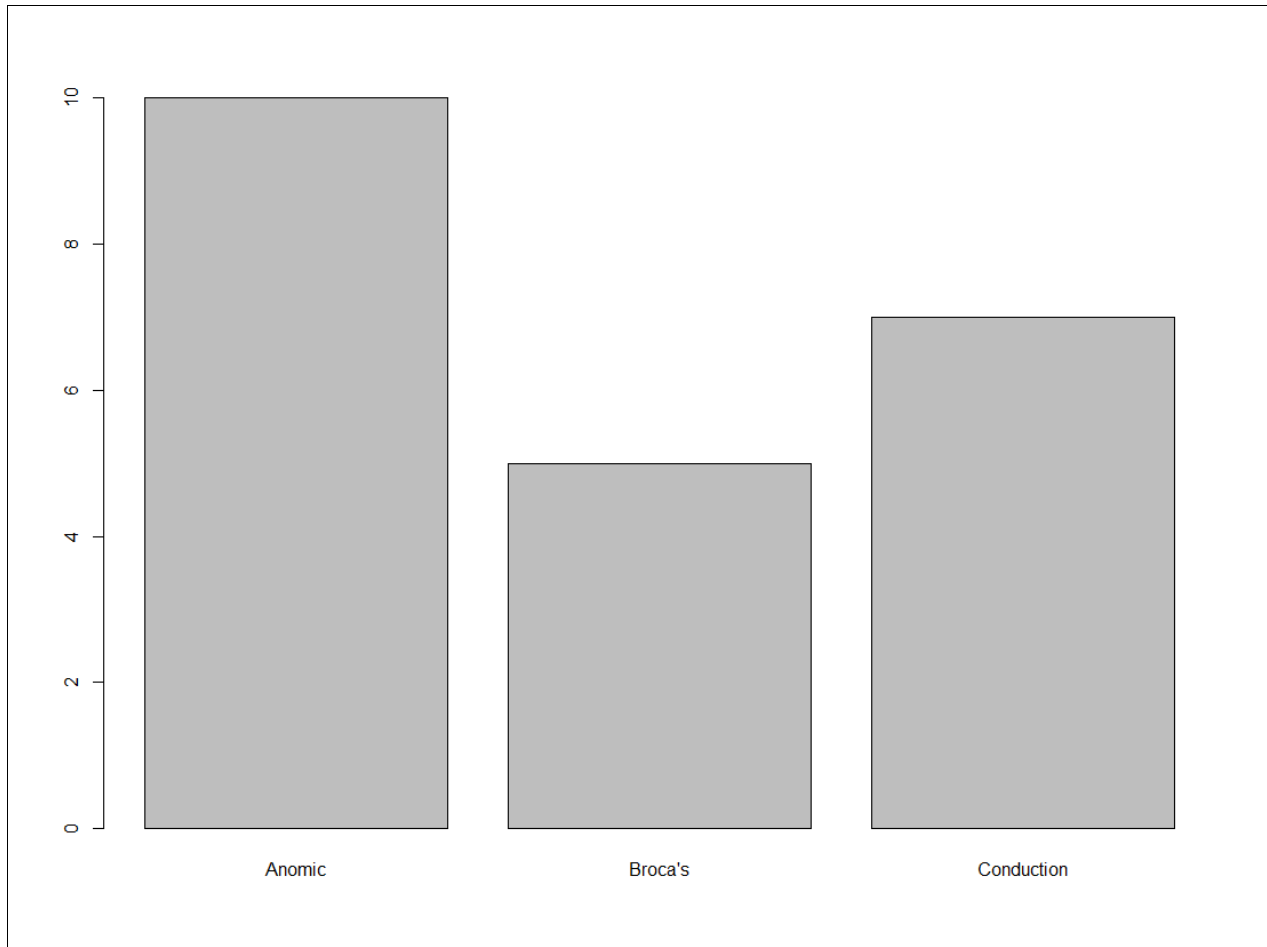
```
attach(aphasia)
```

```
names(aphasia)
```

```
#bar plot for qualitative data
```

```
plot(Type_of_Aphasia)
```

Then click *Export -> Save as*. Insert the plot into the blank below.



Pie charts

3. The command to make pie charts requires counts, but the data list `Type_of_Aphasia` is a list of values.

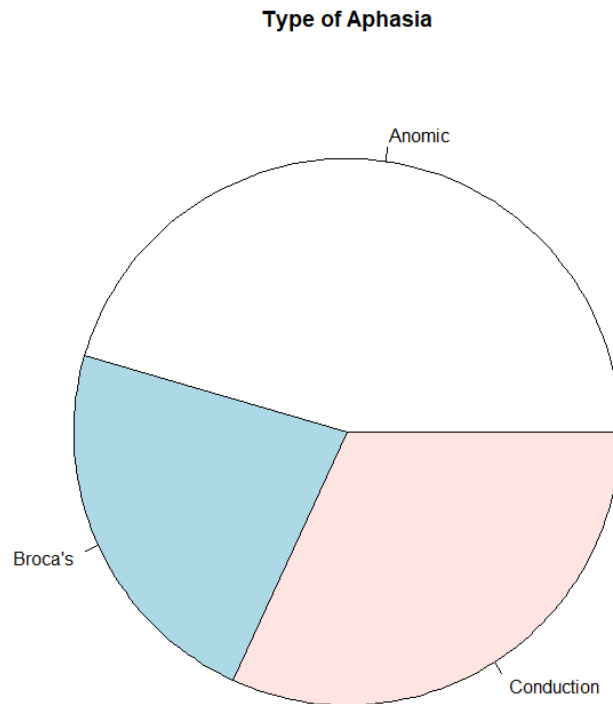
`Type_of_Aphasia` #just typing the name of the list gives a list of values

`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

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



Comparing the two graphs

4. Clearly there are situations in which both bar graphs and pie charts can be used. When is it better to use a bar graph, and when is it better to use a pie chart?

Pie charts are good when needing to find the percentages of certain values and how they make up the whole.

Bar graphs are good for finding frequency distribution with non-numerical data.

A more complicated file

Look at the file `titanic3.csv`. This file contains information on all the passengers who were on the Titanic. More detailed information about this file can be found on <http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>. 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.

5. 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? (note: The easiest way to see the value of `survived` for this passenger is to look at the file using Excel. You could also use the command `survived[name=="Allen, Miss. Elisabeth Walton"]`.)

Yes, “Allen, Miss. Elisabeth Walton”, survived.

6. Create a bar plot for sex and survival

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

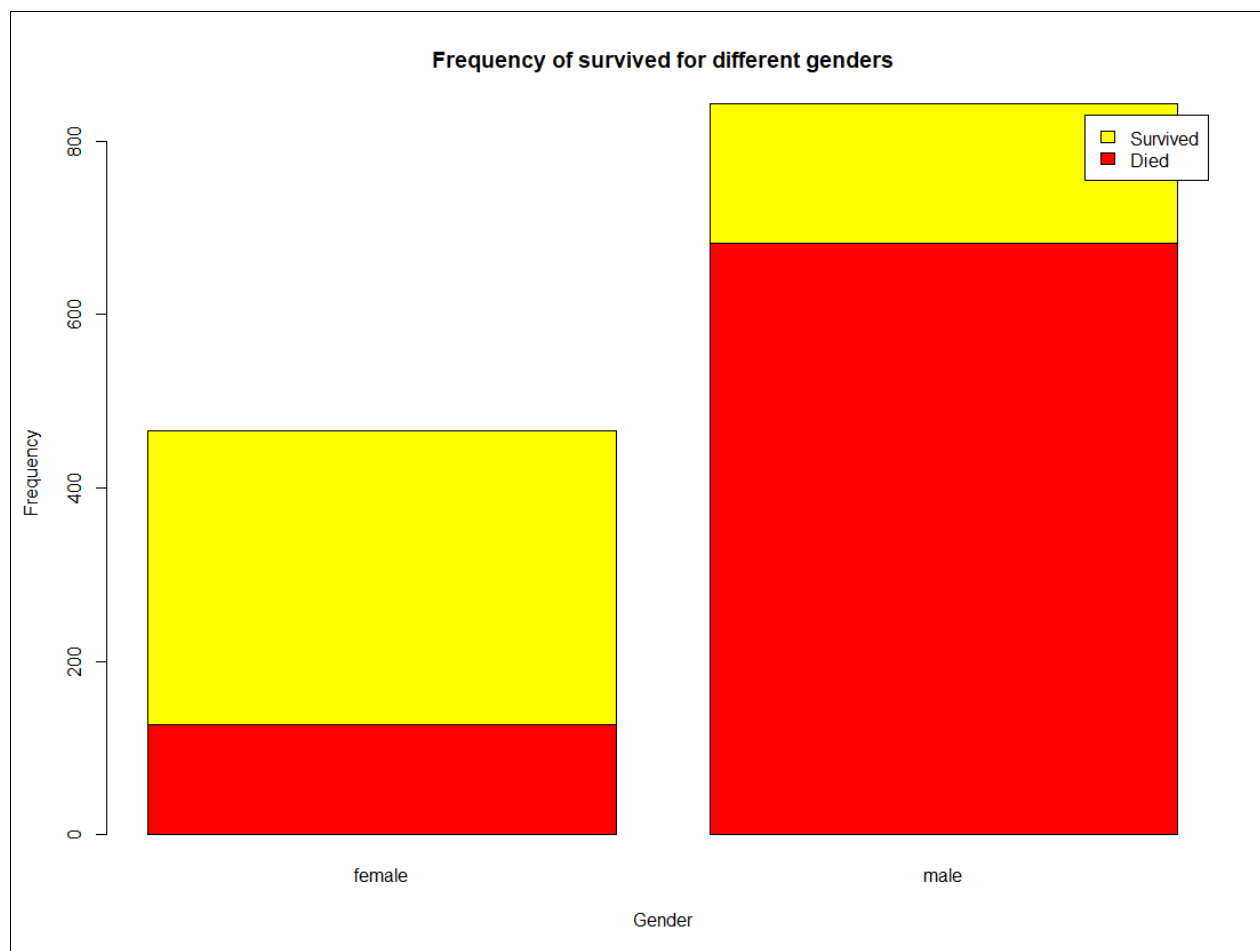
`barplot(table(survived,sex))` #uses the data in the table to make a bar chart

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

`barplot(table(survived,sex), legend=c("Died","Survived"))` #Changes legend to be more descriptive

`barplot(table(survived,sex),main="Frequency of survived for different genders",
xlab="Gender", ylab="Frequency", col=c("red","yellow"), legend=c("Died","Survived"))` #Add labels and colors

Insert your graph here:



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

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

```
table1 #see the result of the previous command
```

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

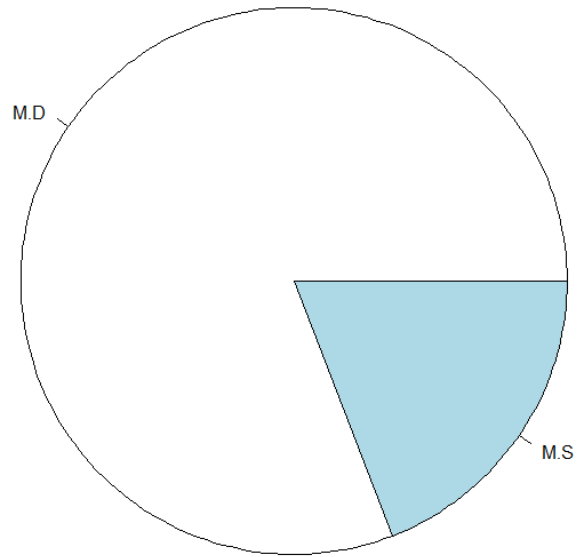
```
table1[2,] #second row
```

```
table1[,1] #first column
```

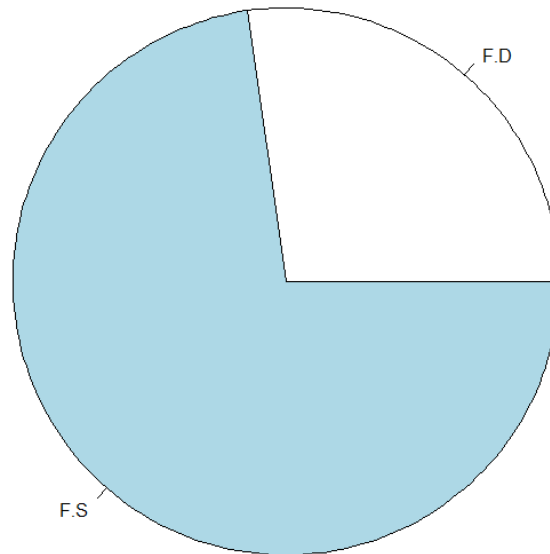
```
table1[,2] #second column
```

Use the above and the pie chart command that you learned in number 3 to create two pie charts, one for the men and one for the women, showing survivors vs those who died for each. Make sure to label your graphs. Paste them in this box:

Men Survived vs Men Dead



Women Survived vs Women Dead



8. Which graph(s) (from numbers 6, and 7) do you think most effectively communicates the data in this file? Why?

I think the bar graph better communicates the data since it shows the frequency, and I believe that better illustrates the magnitude of the event. Also, I think being able to represent the data in one graph is clearer than two charts.

10. The titanic file can be considered a census, since it has all the information available on this ship. What is the difference between a census and a sample? (Feel free to search for the answer on the internet, but if you do, cite your source.)

A **census** is a record of data for an entire population; A **sample** is a record of data for a portion of a population. (**source:** <https://keydifferences.com/difference-between-census-and-sampling.html>)

