

Intro to R Continued

Qualitative vs. Quantitative

- Qualitative: classify individuals into categories
- Quantitative: tell how much or how many of something there is

- Which are qualitative and which are quantitative?
 - Person's Age
 - Person's Gender
 - Mileage (in miles per gallon) of a car
 - Color of a car

Qualitative: Ordinal vs. Nominal

- Ordinal variables:
 - One whose categories have a natural ordering
 - Example: grades
- Nominal variables:
 - One whose categories have no natural ordering
 - Example: state of residence

Create Ordinal Values

<http://www.statmethods.net/input/valuelabels.html>

```
classRank=c(1, 1, 2, 1, 3)
```

```
classRankOrdinal = ordered(classRank,  
levels=c(1,2,3,4),  
labels=c("Fr", "So", "Jr", "Sr") )
```

```
print(classRankOrdinal)
```

```
barplot(summary(classRankOrdinal))
```

Why do we want ordinal values?

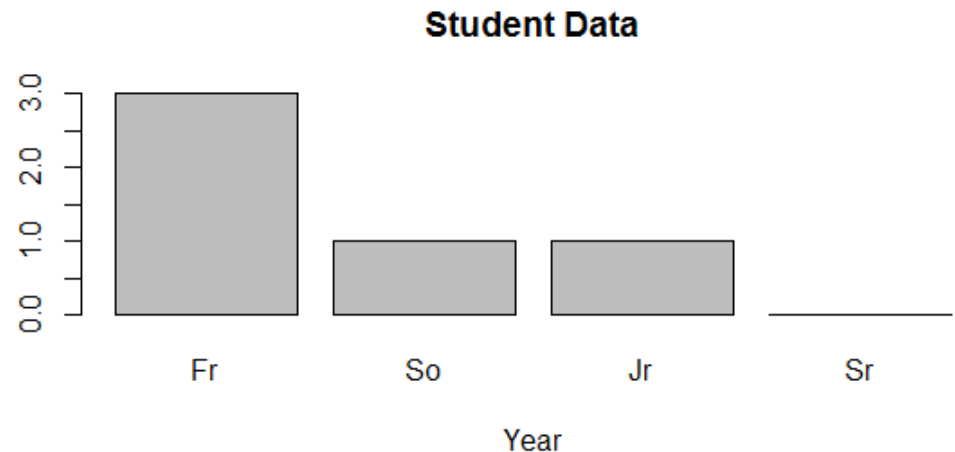
```
classRankNotOrdinal=c("Fr", "Fr", "So", "Fr", "Jr")  
barplot(table(classRankNotOrdinal))
```

Bar Chart

<http://statmethods.net/graphs/bar.html>

- A **bar chart** or **bar graph** is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent.
- function `table` returns a vector of frequency data

```
> barplot(table(classRankOrdinal),  
main = "Student Data",  
xlab = "Year")
```



Quantitative

- Discrete variables: Variables whose possible values can be listed
 - Example: number of children
- Continuous variables: Variables that can take any value in an interval
 - Example: height of a person

Problem

- Using the command `str(dogData)`, identify:
 - variable name
 - quantitative or qualitative
 - discrete, continuous, neither
 - nominal, ordinal, neither
- A specific variable can be selected and passed to the class function. Pass the variable `age` of `dogData` to `class`. What does the result tell us?

Importing Data into R

- `getwd()`
 - `data = read.table("filename.txt", header=FALSE)`
- NOTE: filename.txt can be a URL
- Copy testData.txt from CS130 Public to the location provided by `getwd()`
 - Open testData.txt in a text editor
 - `testData = read.table("testData.txt", header=TRUE)`
 - `print(testData)`
 - `str(testData)`

Candy Dataset Example

<http://zeus.cs.pacificu.edu/ryand/cs130/fall17/candy.txt>

This file contains a header

Brand	Name	ServingPerPkg	OzPerPkg	Calories	TotalFatInGrams	SatFatInGrams
M&M/Mars	Snickers Peanut Butter	1.0	2.00	310	20.0	7.0
Hershey	Cookies 'n Mint	1.0	1.55	230	12.0	6.0
Hershey	Cadbury Dairy Milk	3.5	5.00	220	12.0	8.0
M&M/Mars	Snickers	3.0	3.70	170	8.0	3.0
Charms	Sugar Daddy	1.0	1.70	200	2.5	2.5

Write dataframe to file

```
write.table( dataframe, "file.txt")  
getwd()
```

```
write.table(candy, "candy.txt")
```

Go to Documents and open candy.txt in a text editor

Problem

- Identify each of the following for Total Fat in Grams:
 - Minimum:
 - Maximum:
 - Mean:
 - Standard Deviation:

Use the help feature!

More Dataset Questions

- For the given dataset, for each variable identify Categorical or Quantitative and Nominal/Ordinal/Discrete/Continuous? Why?
 - Brand
 - Name
 - ServingPerPkg
 - OzPerPkg
 - Calories
 - TotalFatInGrams
 - SatFatInGrams