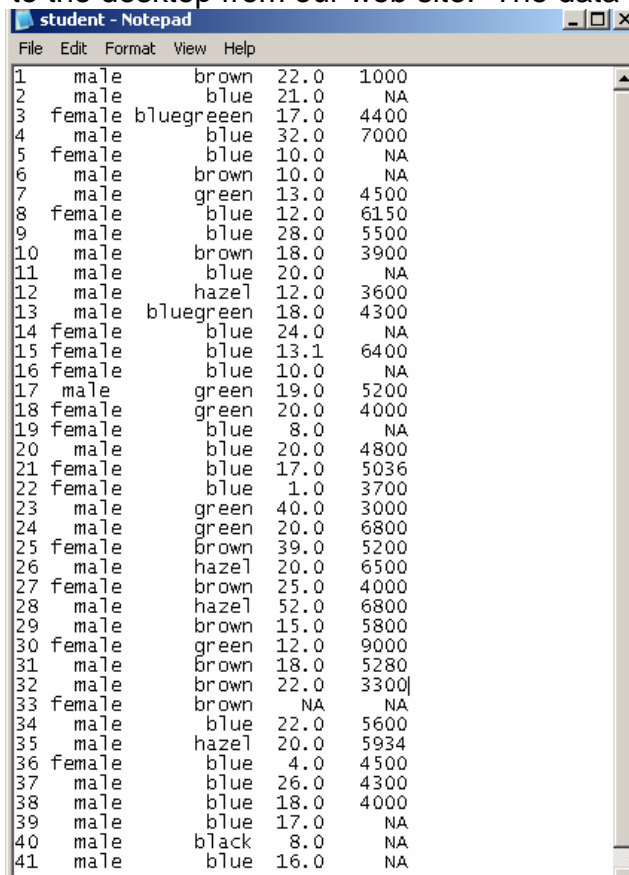


## STAT 457 Fall, 2010 Lab #3

### Topics:

- Creating graphs and tables of quantitative variables in StatCrunch
- Creating mosaic segmented bar graphs in R
- Introducing R GUI
- Using/inspecting data gathered in STAT 451 recently

=====  
[1] Early in your STAT 451 class, Brian gathered data regarding the class gender, eye color, miles traveled from school, and height of the “M”. He put the data together in a NOTEPAD, space delimited document, called `student.txt`. Download that file to the desktop from our web site. The data is shown below.



ID	gender	eye color	miles traveled	height
1	male	brown	22.0	1000
2	male	blue	21.0	NA
3	female	bluegreen	17.0	4400
4	male	blue	32.0	7000
5	female	blue	10.0	NA
6	male	brown	10.0	NA
7	male	green	13.0	4500
8	female	blue	12.0	6150
9	male	blue	28.0	5500
10	male	brown	18.0	3900
11	male	blue	20.0	NA
12	male	hazel	12.0	3600
13	male	bluegreen	18.0	4300
14	female	blue	24.0	NA
15	female	blue	13.1	6400
16	female	blue	10.0	NA
17	male	green	19.0	5200
18	female	green	20.0	4000
19	female	blue	8.0	NA
20	male	blue	20.0	4800
21	female	blue	17.0	5036
22	female	blue	1.0	3700
23	male	green	40.0	3000
24	male	green	20.0	6800
25	female	brown	39.0	5200
26	male	hazel	20.0	6500
27	female	brown	25.0	4000
28	male	hazel	52.0	6800
29	male	brown	15.0	5800
30	female	green	12.0	9000
31	male	brown	18.0	5280
32	male	brown	22.0	3300
33	female	brown	NA	NA
34	male	blue	22.0	5600
35	male	hazel	20.0	5934
36	female	blue	4.0	4500
37	male	blue	26.0	4300
38	male	blue	18.0	4000
39	male	blue	17.0	NA
40	male	black	8.0	NA
41	male	blue	16.0	NA

[1a] Open StatCrunch, then open up the `student.txt` file, remembering to make sure you clicked on **SPACE DELIMITING** before you upload the file to StatCrunch from your computer.

[1b] Copy and paste a histogram of *miles traveled* and a table of all available summary statistics to your WORD document.

[1c] Comment regarding the distribution of variable *miles traveled*:  
shape,

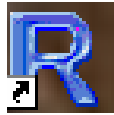
appropriate center,  
appropriate spread,  
modality (unimodal, bimodal, or multimodal)  
outliers or unusual graphical features (if any)

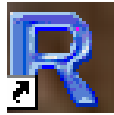
Remember that if a distribution is symmetrical, without outliers, the **appropriate** measure of center is the mean, and appropriate measure of spread is the standard deviation. Otherwise, appropriate center is the median and spread is either/both the IQR (or interquartile range which is  $Q3 - Q1$ ) and the range (which is max value – min value).

[2a] Repeat [1b] for the quantitative variable **height**.

[2b] Repeat [1c] for **height**.

-----  
[3] R is an object oriented programming language, structured similar to C+. It is a very powerful yet freely distributed program, offered to everyone from the CRAN web site (among others). Read the R installation and information documents posted on our class web site for particulars. I will be posting various information on R, including tutorials and useful “scripts” as the semester progresses, so watch out for new entries continuously. Most departments at U of M, as well as other institutions, are converting to R as their statistical “weapon of choice”.



[3a] Open up R by clicking on the icon . Two screens will then open up. Minimize the R Commander screen; we won't be using it now, but will use it in later labs. What remains is the R Console screen, which is the active input/output window for R. Input is typed in red letters, output comes in blue.

You will notice that before any input line is the prompt  $\>$ . This prompt lets you know that R is ready to receive input; if you do not see the red prompt, then R is working on something else and is not ready for your input. Read the information, then right click on the Console screen, followed by **CLEAR WINDOW**. I always like to have a “clean whiteboard” to work on, although you do not have to do this step if you don't want to.

[3b] R can be used as a calculator (Brian mentioned in class that he uses it frequently for this purpose). Type each of the commands below, followed by **ENTER**.

```
4 + 6
26 / 3^2
(7-5)*2 + 5^3
sqrt(9)
log(16)
```

```
e
log(e)
log10(10^4)
pi
Pi
sin(pi)
tan(Pi)
```

Copy this input/output to your WORD document.

Note the following about R:

- R follows order of operations
- `log` is the natural base log
- `e` is the symbol for the base of the natural logs
- R is case sensitive—it knows that `pi` stands for the irrational number  $\pi$ , but thinks `Pi` is a variable you want to specify

-----  
[4] Right click the R Console window and **CLEAR WINDOW** again. If you do not still have Brian's file `student.txt` already located on the desktop, do so now. Let's read this data set into R, calling it `studentdata`, by typing the command shown below, followed by **ENTER**.

```
studentdata <- read.table("C:/Documents and
Settings/math/Desktop/student.txt", header=FALSE, sep="",
na.strings="NA", dec=".", strip.white=TRUE)
```

It is not so important that you get it all on one line, only that it is typed as shown. This command tells R to follow the path `C:/ etc...` to the MA 306 computer lab desktop, to find the file `student.txt`. The file will be read in as a table, with (in this case) 41 rows and 5 columns, where the columns have no title (titles are called "headers"), and where separations between fields are white spaces (called "space delimited" in file jargon), blank fields are given NA designation, and where decimal points look like decimal points in the fields!! There is a better/cleaner/easier way to read in this file, which we will learn in later labs.

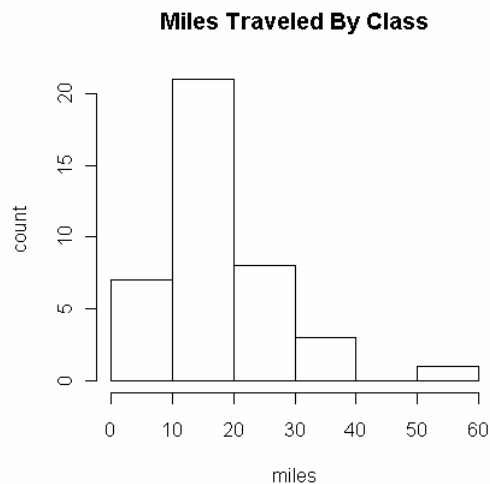
If you are reading in this file from your own computer, instead of our computers in the MA 306 lab, you would have to substitute your own path to the location of `student.txt`, which is different from the path typed above in our example. When I did this lab from my computer at home, I had the file `student.txt` on my flash drive (which showed up as drive D), so had to type "`G:/student.txt`" for my path in the `read.table` command.

[4a] Type `studentdata` **ENTER** to view the data “table” `studentdata` in R Console. Notice that the column headers are automatically given the names `v1`, `v2`, ..., `v5`. See below. Copy and paste your output to the WORD document.

```
RGui
File Edit View Misc Packages Windows Help
R Console
> studentdata <- read.table("G:/student.txt", header=FALSE, sep=" ", na.strings=$
> studentdata
  V1  V2      V3  V4  V5
1  1  male    brown 22.0 1000
2  2  male     blue 21.0  NA
3  3 female bluegreen 17.0 4400
4  4  male     blue 32.0 7000
5  5 female     blue 10.0  NA
6  6  male    brown 10.0  NA
7  7  male    green 13.0 4500
8  8 female     blue 12.0 6150
9  9  male     blue 28.0 5500
10 10 male    brown 18.0 3900
11 11 male     blue 20.0  NA
12 12 male    hazel 12.0 3600
13 13 male bluegreen 18.0 4300
14 14 female    blue 24.0  NA
15 15 female    blue 13.1 6400
16 16 female    blue 10.0  NA
17 17 male    green 19.0 5200
18 18 female   green 20.0 4000
19 19 female    blue  8.0  NA
20 20 male     blue 20.0 4800
21 21 female    blue 17.0 5036
22 22 female    blue  1.0 3700
```

[4b] Type in the following command in the R Console screen

```
hist(studentdata[,4], main="Miles Traveled By Class",
      xlab="miles", ylab="count")
```



to get the following graphical display.

The R command says to make a histogram of student data, all rows, 4<sup>th</sup> column, with main title “Miles Traveled By Class”, and x label “miles”, y label “count”. Copy and paste your histogram into the WORD document. You do this by

putting the mouse over the R graph window, right click, **COPY AS BITMAP**, then paste to WORD.

[4c] Make a histogram of **Height of M** (column 5) and paste into WORD. Add x and y labels as appropriate.

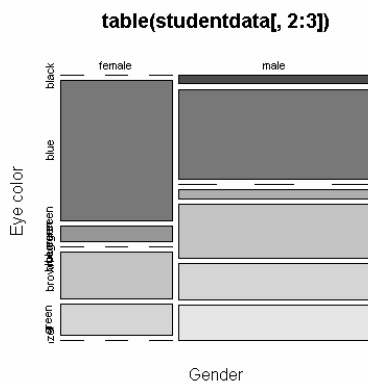
-----  
[5] Brian wrote a “program” in R to make a mosaic segmented bar graph of the eye color data of our data set `studentdata.txt`. He saved this program in an R script called `studentdata.R`, which I want you to download onto the desktop from our web site now.

First, clear the R Console window again by right clicking **CLEAR WINDOW** . Then, click on **FILE OPEN SCRIPT** then find the file `studentdata.R` on the desktop and enter it. When done, you should get a script window open next to R Console which looks like that shown below.

```
colnames(studentdata) <- c("Row", "Gender", "Eye color",  
"Miles traveled", "Height")  
head(studentdata)  
library(lattice)  
mosaicplot(table(studentdata[, 2:3]), col=T)
```

This script has column labels Row, Gender, etc. placed at the top of the table, then calls upon the R library to find the lattice package, which contains the command `mosaicplot()` in it. Notice that you can shorten the command `col=TRUE` to `col=T` for R.

Now, click on **EDIT RUN ALL** to get the graphics screen of the mosaic segmented bar graph. See below. Copy your bar graph mosaic to WORD.



[5a] Copy and paste your graph to WORD.

[5b] According to your graph, is gender and eye color independent? Remember that Brian commented in class that it should be independent, but this was not showing in the data he collected.