

Name:

R Studio Example for Introductory Statistics
Boshes, Broatch

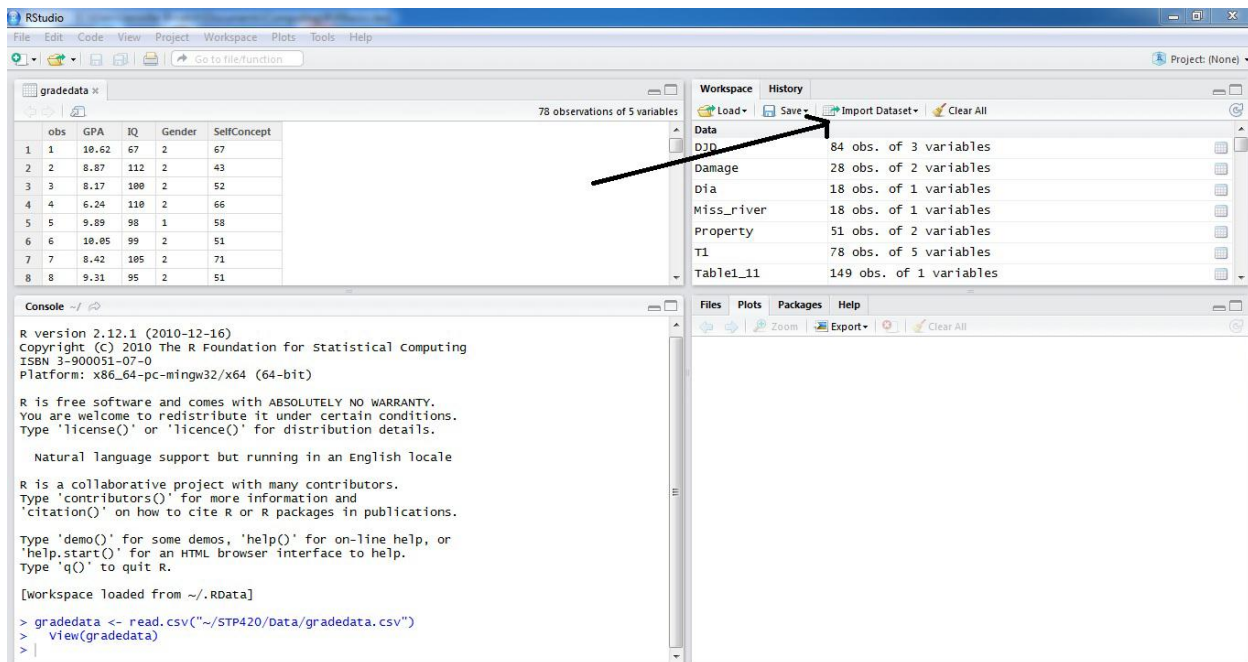
Rock & Roll Marathon 2012

In 2012, 2215 males and 1648 females ran the P.F. Chang's Rock 'n' Roll Marathon in Arizona. In this lab we will be using Rstudio to examine ways of displaying data as well as learning to calculate summary statistics based on a sample of running times.

To begin you will need to download the data set which consists of the fastest 100 male/female running times: RockNRoll.csv. Data sets can be found in Blackboard in the Data folder. The R Script that contains the commands is available at RnRlab.R.

PART I: Displaying Data

1. Import the RockNRoll.csv data set into RStudio using the Import Data Set tab in the Workspace Window. You can rename the data set if you wish from the Import Dataset screen, but you must make the necessary adjustments.



2. Attach the data set to the workspace:

```
> attach(RockNRoll)
```

****Please note that RStudio is case sensitive. All commands are given after the > symbol.**

Name:

R Studio Example for Introductory Statistics

Boshes, Broatch

3. What are the variables of the data set?

4. What type of variables are they?

5. Find the overall mean; `> mean(Time)`

6. Find the median; `> median(Time)`

7. Using RStudio create the following:

a. Histogram;

`> hist(Time, xlab='Time in Minutes', ylab='Frequency', main='Histogram of Running Times')`

b. Stem & Leaf plot;

`> stem(Time)`

c. Boxplot;

`> boxplot(Time, ylab='Time in Minutes', main='Boxplot of Overall Running Times')`

8. Comment on the distribution of the graphs. Do your answers to (5) and (6) support your answer?

Name:

R Studio Example for Introductory Statistics

Boshes, Broatch

Rock & Roll Marathon 2012

Part II: Summary Statistics and Boxplots

1. Find the following summary statistics for each gender:

a. Mean;

```
> tapply(Time, Gender, mean)
```

Women _____

Men _____

b. Median;

```
> tapply(Time, Gender, median)
```

Women _____

Men _____

c. Standard Deviation;

```
> tapply(Time, Gender, sd)
```

Women _____

Men _____

d. Five Number Summary;

```
> tapply(Time, Gender, fivenum)
```

Women _____

Men _____

Name:

R Studio Example for Introductory Statistics

Boshes, Broatch

3. Create a Side-by-Side Boxplot for Gender.

```
> boxplot(Time~Gender, xlab='Gender', ylab='Time – in Minutes')
```

4. Sketch the boxplots.

5. Compare the shapes of the distributions based on the boxplots.

6. Compare the spread of the data based on the boxplots. Do your answers in (1) support this?

7. Discuss other interesting features about these graphs.