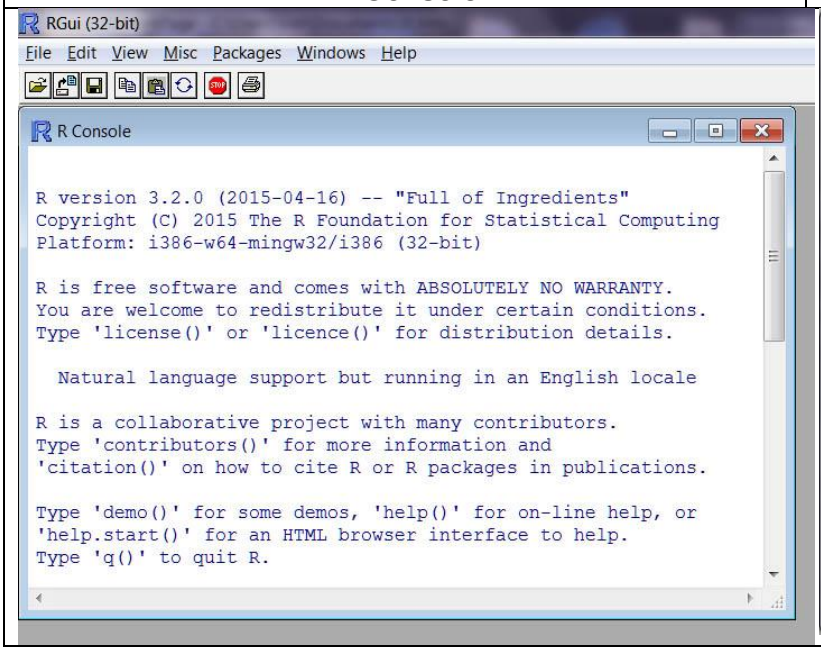
Getting Started with R

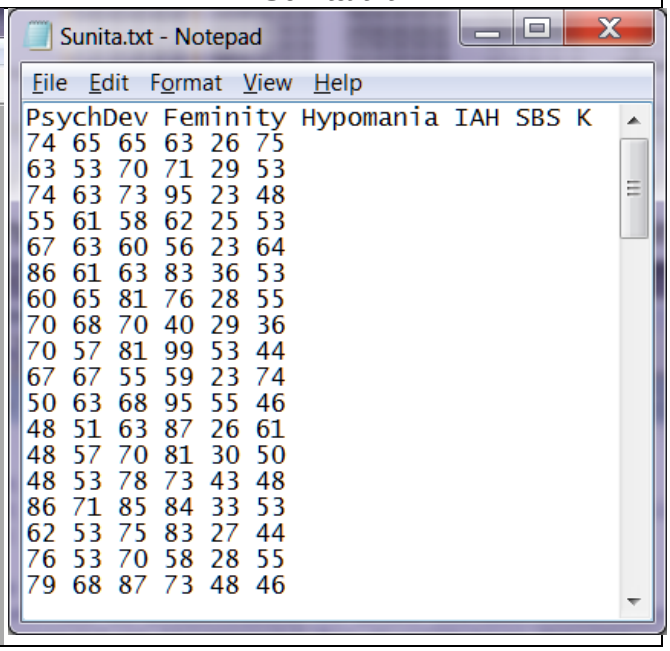
Start out by purchasing or downloading, for free, a pdf of the book "Learning Statistics with R."

Download R – select the correct version for your operating system. I use Windows, and the Windows download included both 32 bit and 64 bit versions. You might run into problems with the interface between R and Excel if one is 64 bit and the other 32 bit.

Install R and boot it up. You will get a window like that below, left with a prompt, >, to enter a command:

R Console	Sunita.txt





We shall work with the "Sunita.txt" data file. Above, right, is a snapshot of the top of the file. The first row contains the names of the six variables. Starting with the second row, each row contains six scores from one subject. Subjects were male students at ECU. Between each name or score and the next name or score is a blank space. The character that separated one name or score from another name or score is called a "delimiter." I usually use blank spaces as delimiters, but commas and tabs are also popular.

The data we are using are those from the research described here: Patel, S., Long, T. E., McCammon, S. L., & Wuensch, K. L. (1995). Personality and emotional correlates of self-reported antigay behaviors. *Journal of Interpersonal Violence, 10,* 354-366.

We had two sets of data. The one set was personality variables from the MMPI. One of these was from the PsychDev (psychopathically deviant) scale, Scale 4, on which high scores are associated with general social maladjustment, rebelliousness, antisocial behavior, criminal behavior, impulsive acting out, insensitivity, hostility, and difficulties with interpersonal relationships (family, school, and authority figures). The second was the Femininity (masculinity/femininity) scale, Scale 5, on which low scores are associated with traditional masculinity - being easy-going, cheerful, practical, coarse, adventurous, lacking insight into own motives, preferring action to thought, overemphasizing strength and physical prowess, having a narrow range of interests, and harboring doubts about one's own masculinity and identity[i]. The third was the Hypomania scale, Scale 9, on which high scores are associated with overactivity, emotional lability, flight of ideas, being easily bored, having low frustration tolerance, narcissism, difficulty inhibiting impulses, thrill-seeking, irritability, restlessness, and aggressiveness. The fourth MMPI variable was Scale K, which is a validity scale on which high scores indicate that the subject is "clinically defensive," attempting to present himself in a favorable

light, and low scores indicate that the subject is unusually frank. The second set of variables was a pair of homonegativity variables. One was the IAH (Index of Attitudes Towards Homosexuals), designed to measure affective components of homophobia. The second was the SBS, (Self-Report of Behavior Scale), designed to measure past aggressive behavior towards homosexuals, an instrument specifically developed for this study.

The first thing we need to do is read the data into R. Type or copy and paste this command into R: **sunita <- read.table("C:/Users/Vati/Documents/StatData/Sunita.txt", header=TRUE)**

You will need to substitute for "C:/Users/Vati/Documents/StatData/" the proper path to the data file on the computer you are using (the place you put it during the download). The path is easy found by right-clicking on the data file, in Windows Explorer, and selecting "Properties." WARNING: You will need to replace all of the "\" slashes with "/" slashes.

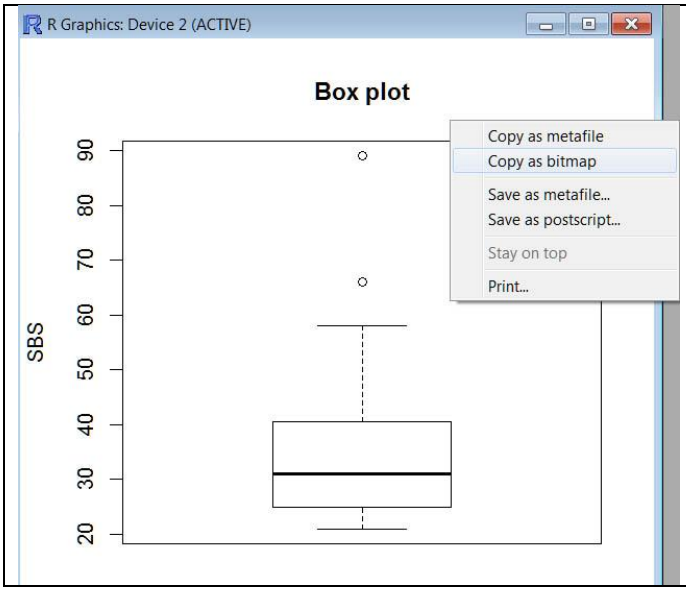
Nothing visible will happen when you hit Enter, other than a new command prompt appearing, but the data are now ready to analyze. To see the "dataframe," just type its name (sunita) and hit enter. All 80 subjects' scores will be displayed:

	PsychDev	Femininity	Hypomania	IAH	SBS	K
1	74	65	65	63	26	75
2	63	53	70	71	29	53
3	74	63	73	95	23	48
4	55	61	58	62	25	53

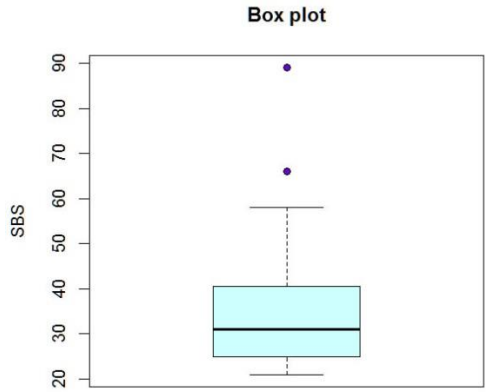
and so on.

To name a variable in R the convention is dataframe-name#variable-name, for example, "sunita$SBS. To avoid needing to specify the dataframe name each time you refer to a particular variable, you can enter this command and hit Enter: **attach(sunita)** – however, our local R guru has advised not to use the attach function. My understanding, or misunderstanding, is that when you modify an attached variable it is modified in the copy of the dataframe that is active in the current session but not in the source file.

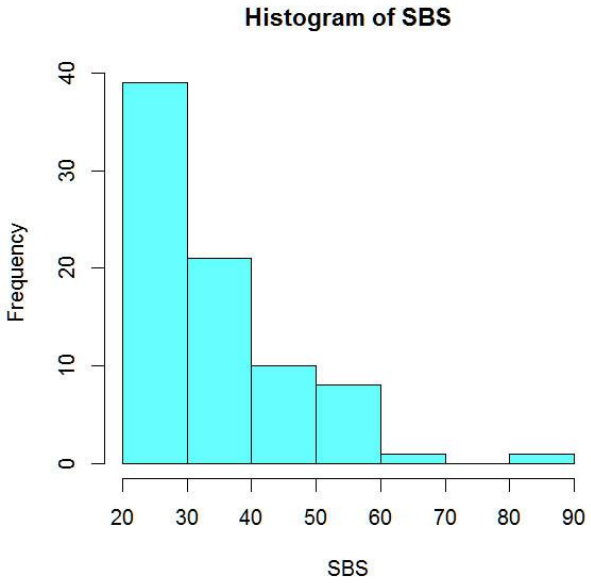
Now let's make a box and whiskers plot for the SBS variable. Enter this command and hit Enter: **boxplot(sunita$SBS, main="Box plot", ylab="SBS")**



The box and whiskers plot will appear in a new frame. You can right click in that frame and select "Copy as bitmap" to copy the graphic to the clipboard. You can then paste into another document, such as a Word document. I often prefer to paste it into a photo editor and then pretty it up a bit. Like this:



Make a histogram too: **hist(sunita$SBS, main="Histogram of SBS")**



R also has commands that redirect graphic output to a file on your hard drive, but I don't like these, since you cannot see the graphic until you exit R and if you created two or more graphics you get only the last one – it overwrites rather than appends.

Now let's get some descriptive statistics:

> **summary(sunita$SBS)**

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  21.00   25.00   31.00   34.31   40.25   89.00
```

> **sd(sunita$SBS)**

[1] 12.69466

Notice that the mean SBS is distinctly above the median SBS. This should not surprise you, as you have already seen, from the plots, that SBS is positively skewed, aka skewed to the right. Base R does not have a function to estimate skewness, but I'll show you later how to get that function from an add-on "package.)

Wuensch's R Lessons

---

[i] A high Scale 5 score indicates that the individual is more like members of the other gender than are most people. A man with a high Scale 5 score lacks stereotypical masculine interests, and a woman with a high Scale 5 score has interests that are stereotypically masculine. Low Scale 5 scores indicate stereotypical masculinity in men and stereotypical femininity in women. MMPI Scale scores are "T-scores" – that is, they have been standardized to mean 50, standard deviation 10. The normative group was residents of Minnesota in the 1930's. The MMPI-2 was normed on what should be a group more representative of US residents.