

# Importing Data and Constructing Graphs in R Commander

Statistics 1, Section 2,6,7

## 1. Starting R Commander

Open R by clicking Start | Applications | R

Open the R Commander GUI (Graphical User Interface) by typing `library(Rcmdr)` at the screen prompt.

**Note:** R Commander uses multiple windows.. Sometimes you will “lose” the screen you need to view. If this happens, go to the taskbar at the bottom of your computer screen and click on the screen you need to see. Graphs will often be in the R screen instead of R Commander.

## 2. Entering and Importing Data

In order to analyze data in R, you must first get the data into R. You can either enter the data directly or import data that is already stored in an existing text file (.txt) or Excel spreadsheet (.xls).

- We will enter the following data directly. The exam scores for 5 statistics students are: 76, 80, 60, 90, and 82.
  - From the menu bar across the top of the R Commander window, select Data | New data set...
  - A dialog box will pop up, choose a name for your dataset like “testscores” and enter it in the dialog box, then click OK
  - A spreadsheet will appear (you may need to click on the R Gui button along the bottom of your computer screen to see the spreadsheet). Click on the cell beneath “var1” and type 76.
  - You can hit the enter key or click with the mouse to move to the next cell down in the same column. Then type 80.
  - Continue the process until all 5 values have been entered
  - Click on “var1” at the top of the column, and enter the variable name in the dialog box. You can use “score” for these data.
  - Click the X in the upper right corner when you are done entering the data. The data you just entered is now the active dataset in R Commander.
  - Back in the main R Commander window you can click the View data set button on the top of the screen to see the active data set (which you just entered).
  
- We now import an existing dataset giving quarterly sales for different market segments (auto, furniture, etc.) from 1979-1989. This dataset may be downloaded from the course webpage.
  - From the menu bar, select Data | import data | from text file, clipboard or URL... A dialog box with a lot of choices will pop up. Next to “enter name for data set” type a name such as sales or quarterlysals. Click on the open circle next to “Internet URL” so it fills in with a black dot. This means you are getting the data from an internet website. Click the OK button.
  - A dialog box pops up asking for the internet URL. Type or copy and paste <http://www.csus.edu/indiv/n/norrisa/stat1f09/quarterlysals.txt> then click the OK button.
  - Return to the R Commander window. The Messages box at the bottom of the screen should say “NOTE: The dataset sales has 44 rows and 7 columns.” if you succeeded in importing the data.
  - Click the View data set button on the top of the R commander window if you’d like to see the data.

### 3. Time Series Plots

- Time series plots help us understand the historic behavior of a variable. To construct a time series plot, the time is plotted on the x-axis and the variable of interest along the y-axis. So if you want to plot annual tuition at CSUS (for residents) by year, the x-axis would be labeled “years” and the y-axis would be “fees in dollars.”
- We will construct a time series plot of retail sales which is in the data set we just imported. At the top of the R commander screen in the box next to “Data set”, you should see “sales” or whatever you named the quarterly sales data set. If not, you need to make “sales” your active data set, by selecting Data | active data set | select active data set...
- Select Graphs | Line graph . A dialog box will pop up. Beneath “x variable (pick one)”, click on TIME. Next, beneath “y variable (pick one or more)” click on AUTO, then hold down the Ctrl key on your computer keyboard while clicking on GMER. Then click the OK button. AUTO stands for automotive sales in millions of dollars. GMER stands for sales of general merchandise in millions of dollars.
- If you don’t see the graph, go to the taskbar and click on R Gui. It is likely in the “hidden” R Gui window.

Interpreting the Data: Recall the 0 on the x-axis represents the first quarter of 1979, 1 is the second quarter of 1979 and so on.

- Consider the GMER data. There are at least two very strong features that emerge when the data are viewed graphically. What are they?
- For the GMER data, why do the spikes appear in a fairly regular pattern?
- What differences do you notice in the auto sales data compared to general merchandise?

### 4. Histograms

- Histograms are useful for understanding the shape of a data set. We now construct a histogram of data for overall scores of Stat 1 students.
- First, import the data from <http://www.csus.edu/indiv/n/norrisa/stat1f09/stat1scores.txt> .Go back to Step 2 in this handout if you don’t remember how.
- Select Graphs | Histogram and a dialog box will pop up. Underneath “Variable (pick one)” click on “score”. Click the OK button.
- Examine the histogram. Note the test score have been grouped into 9 bins. Specifically, along the x-axis, the base the first rectangle extends from 20-30, the next base goes from 30-40, the next from 40-50 and so on. These ranges into which the data are grouped are called “bins.”

- R Commander allows you to specify the number of bins if you want. Go back and redo the histogram, *but change the number of bins to 15*. Often, it is necessary to play with the number of bins before you get a suitable histogram. You don't want too few or too many bins. The more data you have the more bins you can use. Actually, since this data set only has 51 values, 15 bins is a bit much, but I want you to see the difference. Getting a really nice graph often involves some trial-and-error.
- What does the height of each bar in the histogram represent?
- Does the histogram appear *symmetric* or *skewed*?
- Outliers are values that are far from the bulk of the other values in the data set. Are there any outliers?
- Is Professor Norris a tough grader or an easy grader?

## 5. Bar and Pie Charts

- Important question: What type of data are histograms and time series graphs use for?
- Bar and pie charts are used for qualitative or categorical data.
- Import the data from the Popularity study.
- Construct a bar chart for the "Goals" variable. Select Graphs | Bar chart. In the dialog box, choose "Goals" and click the OK button.
- Now, construct a pie chart for "Goals."

## 6. Frequency and relative frequency tables.

- Select Statistics | Summaries | Frequency distributions... Then a dialog box pops up. Select the "Goals" variable and click the OK button. The relative frequency distribution below appears in the Output Window within the R Commander window.

```
> 100*.Table/sum(.Table) # percentages for Goals
```

```
Grades Popular Sports
51.67364 29.49791 18.82845
```

- Construct a relative frequency table to determine the percent of students from each school.
- Note: R Commander will only construct a relative frequency table for qualitative variable. Other software will allow you to bin a numeric variable and construct a relative frequency table based on the bins.