

1. Datafiles

One of the most useful ways to store data for use in R is in a CSV file. The CSV format has lines of data with the individual entries in the line separated by commas. For R dataframes, each line will be a case and each column a variable. The first line of the data file will be the variable names. For example, the first few lines of one of the data files used in class (`actgpa.csv`) looks like this:

```
ACT,GPA
30,3.992
,2.533
25,3.377
20,3.009
.....
```

CSV format is supported by Excel and Excel is one of the most convenient ways to create data files.

2. Reading and writing datafiles

The R commands `read.csv` and `write.csv` will read and write CSV files. These files can be read directly from the internet or from a local file on your own computer. To read a local file, you can first set the “working directory” so that R will know where to find it. An example of reading the `actgpa.csv` file is

```
> x=read.csv("http://www.calvin.edu/~stob/data/actgpa.csv") # from internet
> y=read.csv("actgpa.csv") # from local file
> z=read.csv(filechoose()) # choose from dialog box
```

3. Data for Mathematics 243

Data from the textbook is available at <http://www.calvin.edu/~stob/data/DVB>. Almost all of the other datasets used in class are available in the directory <http://www.calvin.edu/~stob/data>. You can get the data by directly reading it from the internet or first saving the file to a convenient place on your own computer.

4. R packages

There are R packages available for all textbook datasets and for most other datasets. To install these packages in your own version of R, you need first to download the packages from the course website. To do this, save the following files to your computer:

```
Windows version: http://www.calvin.edu/~stob/data/M243S10_1.0.zip
                  http://www.calvin.edu/~stob/data/DVB_1.0.zip
Mac and linux version: http://www.calvin.edu/~stob/data/M243S10_1.0.tar.gz
                      http://www.calvin.edu/~stob/data/DVB_1.0.tar.gz
```

Then install a package by choosing the “install package from local file” option on the R menu.

To use a package you will have to load it using the “load package” option on the R menu. The packages are called `M243S10` and `DVB`.

The `M243S10` package also makes available a function `getdata()` that will load other data sets from this website that are not in the package. Usage is:

```
> x=getdata("actgpa")
```

5. DVB data

All datasets from the DVB book have names that correspond to the name of the problem that come from. Data from a problem in Chapter 9 will have as its name `Ch09` followed by the first word of the name of the problem in the book. For example, `Ch09Law` is the name of a DVB data set. As a convenience, the function `Chapter()` takes a numerical argument and returns the names of all the data sets from the chapter of DVB.