Data munging

The term *data munging* refers to reworking raw data into a more usable format.

The basic command to read data into R is `read.csv()`.

This can be used to retrieve data from a file or from the internet.

By default, this command assumes a header row.
Look at the help page for `read.csv()` (using `?read.csv` or following the tool tips)

The command `read.delim()` is used to read tab-separated data files

Both `read.csv()` and `read.delim()` are specific versions of `read.table()`

`read.table()` is a versatile command with many handy options

To specify that the first row has the column names, use `header=TRUE`
Other useful options for \texttt{read.table()}: 

\texttt{sep} can be used to specify how the records are separated, such as by ",", " or "\\t" (tab) or " " (any whitespace)

\texttt{row.names} gives a column number that contains rows, or a vector of row names

\texttt{col.names} gives a vector of column names

\texttt{na.strings} specifies how missing values are encoded
nrows gives the number of rows to be read in
skip gives the number lines at the top to omit
fill allows columns of different length to be filled in with blanks as needed
blank.lines.skip allows blank rows to be omitted
colClasses allows you to specify which columns are numerical, character, factor, logical, etc.
The `read.fwf()` command allows R to import data that is in a fixed width format, with the column widths specified by the `width` argument.

Excel does this easily also, with its `Text to Columns` option under the `Data` tab.
You can save an R data frame with `write.csv()`. For example:

```r
write.csv(myDataFrame, 
          file="dataFilename.txt", 
          row.names=FALSE)
```

More generally, you can use `write.table()`