Reading Guide

Dan Maxwell

May 16, 2017

Section Notes for *R in Action*

5.2.1 Mathematical functions
The natural logarithm, usually denoted as e in equations, is calculated as follows:

```
exp(1)  # Calculate the natural logarithm, a constant = 2.718282
```

```
## [1] 2.718282
```

```
pi      # Pi is another built-in constant = 3.141593
```

```
## [1] 3.141593
```

5.2.2 Statistical functions
Interestingly, R does not provide a built-in function to calculate `mode`.

5.4.1 Repetition and looping
Kabacoff does not provide much detail about R's three looping mechanisms. Please see the loops_clinic.Rmd file for a working example of `for`, `while`, and `repeat` loops.

5.5 User-written functions
Custom functions are not covered in this course. However, they are presented in my R functions workshop. This section is therefore optional, though it is worth reading in order to get a sense of how functions work.

5.6.3 The reshape2 package
This is an important section as many students run into trouble when they present data to a statistical function, and it is in the wrong format. Table 5.8 presents a hypothetical dataset in *wide* format whereas Table 5.9 presents that same dataset in *long* format. Remember: you only need as many columns as there are variables.

Let's illustrate what these formats look like with some real data. The file home_ownership.txt contains "homeownership rates, in percentages, for a subset of states for the years 1985, 1996, and 2002. These values represent the proportion of homes owned by the occupant to the total number of occupied homes" (Ott & Longnecker, 2016, p. 129).

```
library(reshape2)  # Load the reshape2 library.

# Read in the file -- values are separated by tabs -- do not convert to factors.

home_own_wide <- read.delim("home_ownership.txt", sep = "\t", stringsAsFactors = FALSE)

home_own_wide      # Take a look at the data in wide format.
```

```
##          state pct_1985 pct_1996 pct_2002
## 1      Alabama     70.4     71.0     73.5
## 2       Alaska     61.2     62.9     67.3
## 3      Arizona     64.7     62.0     65.9
## 4     Arkansas     66.6     66.6     70.2
## 5   California     54.2     55.0     58.0
## 6     Colorado     63.6     64.5     69.1
## 7  Connecticut     69.0     69.0     71.6
```

```
## 8     Delaware     70.3     71.5     75.6
## 9      Florida     67.2     67.1     68.7
## 10     Georgia     62.7     69.3     71.7
## 11      Hawaii     51.0     50.6     57.4
```

```
home_own_long <- melt(home_own_wide, id = "state")

home_own_long      # Now take a look at the data in long format.
```

```
##          state variable value
## 1      Alabama pct_1985  70.4
## 2       Alaska pct_1985  61.2
## 3      Arizona pct_1985  64.7
## 4     Arkansas pct_1985  66.6
## 5   California pct_1985  54.2
## 6     Colorado pct_1985  63.6
## 7  Connecticut pct_1985  69.0
## 8     Delaware pct_1985  70.3
## 9      Florida pct_1985  67.2
## 10     Georgia pct_1985  62.7
## 11      Hawaii pct_1985  51.0
## 12     Alabama pct_1996  71.0
## 13      Alaska pct_1996  62.9
## 14     Arizona pct_1996  62.0
## 15    Arkansas pct_1996  66.6
## 16  California pct_1996  55.0
## 17    Colorado pct_1996  64.5
## 18 Connecticut pct_1996  69.0
## 19    Delaware pct_1996  71.5
## 20     Florida pct_1996  67.1
## 21     Georgia pct_1996  69.3
## 22      Hawaii pct_1996  50.6
## 23     Alabama pct_2002  73.5
## 24      Alaska pct_2002  67.3
## 25     Arizona pct_2002  65.9
## 26    Arkansas pct_2002  70.2
## 27  California pct_2002  58.0
## 28    Colorado pct_2002  69.1
## 29 Connecticut pct_2002  71.6
## 30    Delaware pct_2002  75.6
## 31     Florida pct_2002  68.7
## 32     Georgia pct_2002  71.7
## 33      Hawaii pct_2002  57.4
```

Ott, L., & Longnecker, M. (2016). *An introduction to statistical methods & data analysis* (7th ed.). Boston, MA: Cengage Learning.