Introduction to R

Session 1 – Introduction

Statistical Consulting Centre

19 July, 2017

1. Using R as a calculator

1. Find the values of:

   (a) $1 + 4$

```
1+4
```

```
## [1] 5
```

   (b) $2^3 + \frac{4}{\sqrt{34}}$

```
2^3 + 4/sqrt(34)
```

```
## [1] 8.685994
```

   (c) $\log 30$

```
log(30)
```

```
## [1] 3.401197
```

   (d) $\log_{10} 30$

```
log(30)
```

```
## [1] 3.401197
```

   (e) $|-2|$ (Hint: $|x|$ denotes the *absolute value* of x. Search on Google if you're unsure.)

```
abs(-2)
```

```
## [1] 2
```

2. Now open Rstudio, open a `R` script clicking `File` → `New` → `R` script.
3. Save this script by clicking `File` → `Save As...`.
4. Select a directory/location and save the script. Note: the saved script should have `.r` as extension. For example, if you call your file `exercise one`, then you should save it as `exercise one.r`
5. Copy and paste the code you typed (*not the output, not the* > *symbol, just the code you typed*) at the console for into the `R` script opened in Rstudio.
6. Submit your entire script at once to the `R` Console by highlighting all codes and pressing Ctrl + R.
7. From now on, type all of your code in your `R` script and submit it to the `R` Console using Ctrl + R.

2. Reading data into R

1. `lake.csv` contains data on mercury contamination in 53 different lakes in Florida. The variable names and what has been measured are presented below.

- `ID`: ID number of the lake
- `Lake`: Name of the lake
- `pH`: pH value
- `Calcium`: concentration of Calcium
- `Chlorophyll`: concentration of Chlorophyll (mg/L)

2. Read the data into `R`, saving it in object named `lake.df`.

```
lake.df <- read.csv("location of your folder/Lake.csv",
                    stringsAsFactors = FALSE)
```

3. Use `dim()` and `head()` to look at some of the properties of the dataset you have just read into `R`. *Always* perform this important step to check that your dataset is as it should be.

```
dim(lake.df)
```

```
## [1] 53  5
```

```
head(lake.df)
```

```
##   ID         Lake  pH Calcium Chlorophyll
## 1  1    Alligator 6.1     Low         0.7
## 2  2        Annie 5.1     Low         3.2
## 3  3       Apopka 9.1    High       128.3
## 4  4 Blue Cypress 6.9  Medium         3.5
## 5  5        Brick 4.6     Low         1.8
## 6  6       Bryant 7.3     Low        44.1
```

4. Calculate the mean and standard deviation of both `pH` and `Chlorophyll`.

```
mean(lake.df$pH, na.rm = TRUE)
```

```
## [1] 6.590566
```

```
mean(lake.df$Chlorophyll, na.rm = TRUE)
```

```
## [1] 23.11698
```

```
sd(lake.df$pH, na.rm = TRUE)
```

```
## [1] 1.288449
```

```
sd(lake.df$Chlorophyll, na.rm = TRUE)
```

```
## [1] 30.81632
```

5. Check out what `summary()` does by running `summary(lake.df$pH)`.

```
summary(lake.df$pH)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.600   5.800   6.800   6.591   7.400   9.100
```

6. Check the ferquency of `Calcium` concentration

```
table(lake.df$Calcium)
```

```
##
##   High    Low Medium
##     16     18     19
```

7. Turn the freuqency table in 2.6 into proportion, keep only 2 decimal places.

```r
round(prop.table(table(lake.df$Calcium)) * 100, 1)
```

```
##
##   High    Low Medium
##   30.2   34.0   35.8
```