Intro to R

Contents

Introduction to R

```
a_thing <- 4
another_thing <- 1
another_Thing <- 7

both_things <- a_thing + another_thing
```

Then we created a tiny data set.

```
a_data_thing <- data.frame(x = 2, y = 8)

a_data_thing$x
```

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

How would we print the variable y? Type your answer in the chunk below

Write notes for yourself in the white space. Maybe explain to your future self what dollar signs do.

ERASE THIS AND TYPE SOME NOTES HERE

Enough playing around, let's load some data!

```
acitelli <- read.csv("/Users/randigarcia/Desktop/acitelli.csv", header=TRUE)
```

Next, you want to look at your data.

```
head(acitelli)
```

```
##   cuplid  Yearsmar gender self_pos other_pos satisfaction tension simhob
## 1      3  8.202667     -1      4.8       4.6     4.000000     1.5      0
## 2      3  8.202667      1      3.8       4.0     3.666667     2.5      1
## 3     10 10.452667     -1      4.6       3.8     3.166667     4.0      0
## 4     10 10.452667      1      4.2       4.0     3.666667     2.0      0
```

```
## 5     11 -8.297333     -1      5.0       4.4     3.833333     2.5      0
## 6     11 -8.297333      1      4.2       4.8     3.833333     2.5      0
```

```
str(acitelli)
```

```
## 'data.frame':    296 obs. of  8 variables:
##  $ cuplid      : int  3 3 10 10 11 11 17 17 21 21 ...
##  $ Yearsmar    : num  8.2 8.2 10.5 10.5 -8.3 ...
##  $ gender      : int  -1 1 -1 1 -1 1 -1 1 -1 1 ...
##  $ self_pos    : num  4.8 3.8 4.6 4.2 5 4.2 4 4 4.2 4.4 ...
##  $ other_pos   : num  4.6 4 3.8 4 4.4 4.8 3.6 4.4 3.8 4.8 ...
##  $ satisfaction: num  4 3.67 3.17 3.67 3.83 ...
##  $ tension     : num  1.5 2.5 4 2 2.5 2.5 3 2 3.5 2.5 ...
##  $ simhob      : int  0 1 0 0 0 0 -1 0 0 0 ...
```

```
names(acitelli)
```

```
## [1] "cuplid"       "Yearsmar"     "gender"       "self_pos"
## [5] "other_pos"    "satisfaction" "tension"      "simhob"
```

There is also documentation about functions.

```
?head
```

You probably also want descriptive statistics.

```
summary(acitelli)
```

```
##      cuplid         Yearsmar              gender      self_pos
##  Min.   :  3.0   Min.   :-11.214000   Min.   :-1   Min.   :2.600
##  1st Qu.:165.2   1st Qu.: -7.089000   1st Qu.:-1   1st Qu.:4.000
##  Median :313.5   Median : -1.089000   Median : 0   Median :4.200
##  Mean   :282.6   Mean   : -0.000036   Mean   : 0   Mean   :4.186
##  3rd Qu.:401.2   3rd Qu.:  6.077667   3rd Qu.: 1   3rd Qu.:4.400
##  Max.   :485.0   Max.   : 15.036000   Max.   : 1   Max.   :5.000
##    other_pos      satisfaction      tension          simhob
##  Min.   :2.600   Min.   :1.167   Min.   :1.000   Min.   :-1.0000
##  1st Qu.:4.000   1st Qu.:3.333   1st Qu.:2.000   1st Qu.: 0.0000
##  Median :4.200   Median :3.833   Median :2.500   Median : 0.0000
##  Mean   :4.264   Mean   :3.605   Mean   :2.431   Mean   : 0.0777
##  3rd Qu.:4.600   3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.: 0.2500
##  Max.   :5.000   Max.   :4.000   Max.   :4.000   Max.   : 1.0000
```

We can also select pieces of a data frame. That first number is the row, the second is the column.

```
acitelli[2, 6]
```

```
## [1] 3.666667
```

```
#You try it! Find a numder you want to pull from the dataset.
#riggsi[ ?, ?]
```

If it is instead a single variable, you can also select a piece.

```
acitelli$satisfaction[2]
```

```
## [1] 3.666667
```

In the chunk below, pick out the gender of the person in the 50th case.

```
#try it by referring to the row and column of the data frame.

#try it by referring to the variable, using the dollar sign notation.
```

Installing Packages

You might want to get descriptive stats or frequencies for specific variables. There are `base R` functions, but I like to use the package `mosaic`. You can find more information and a cheat sheet for `mosaic` at this website.

First we need to install the `mosaic` package using the `install.packages()` function. The package name goes inside of the paratheses in double quotes: ‘“mosaic”’. This is something we do only once in the console, you wouldn't want to save it in your .Rmd file.

```
#install.packages("mosaic")
```

Once a package is installed, any time we start a new R session and we want to use functions inside of that package, we will need to load the package with the `library()` function.

```
library(mosaic)
```

Basic Descriptive Statistics with `mosaic`

The function `favstats()` will give descriptive statistics for a numerical vairable, and the function `tally()` will give you frequencies for a categorical variable (or a numberical variable...if you want it). Functions in `mosaic` use the formula syntax, where `y ~ x`, or for a single variable, `~x`. The `~` key can be found just below your `esc` key. The first argument is the formula, and the second argument is the data frame, e.g., `data = acitelli`.

```
favstats(~satisfaction, data = acitelli)
```

```
##       min       Q1   median Q3 max    mean        sd   n missing
##  1.166667 3.333333 3.833333  4   4 3.60473 0.4964205 296       0
```

```r
tally(~gender, data = acitelli)
```

```
## gender
##  -1   1
## 148 148
```

```r
#tally() can also give you percentages
tally(~gender, data = acitelli, format = "percent")
```

```
## gender
## -1  1
## 50 50
```

Descriptives split by gender.

```r
favstats(satisfaction ~ gender, data = acitelli)
```

```
##   gender      min       Q1   median Q3 max     mean        sd   n missing
## 1     -1 1.500000 3.333333 3.833333  4   4 3.591216 0.5300260 148       0
## 2      1 1.166667 3.500000 3.833333  4   4 3.618243 0.4617875 148       0
```

What are the standard deviations of perceived `tension` by gender?

What is(are) the mode(s) of the `self_pos` variable?

The `mosaic` package also has a function for getting the correlation coefficent, it's called `cor()`. Using the same format (i.e., formula then data), how would you get the correlation of satisfaction and tension?