

RHandout#2

What to do?

1. Start RStudio by clicking on the RStudio icon or by choosing it from the start menu.
2. It will help to get the online version of the R in a Nutshell book for this exercise.
3. Type in the following commands and explore what is going on. Please understand the output of each command and what is the command does. (Possible exam question)

- a. Creating a simple data set for “exploration”; let us use our data for K-means

```
mydata <- c(23,25,24,23,21,31,32,30,31,30, 35,37, 38,37,39,42,43,45,43,45)  
kmeans(mydata,3)
```

We are asking R to cluster the data into 3 clusters. Check the output and understand. Change the cluster size.

Obtain a large data set of your choice and try it for different cluster sizes.

- b. Explore the packages on the right bottom panel of RStudio.

Make sure the package is included.

Browse to the examples given under the specific packages and try them out.

Try ggplot where you map all your points on top of a specific “map” and using google’s geocodes. Try this after loading ggplot.

```
p <- ggplot(mtcars, aes(mpg, wt)) + geom_point()  
# With one variable  
p + facet_grid(. ~ cyl)  
p + facet_grid(cyl ~ .)
```

- c. Try this:

```
require(graphics)  
pairs(mtcars, main="mtcars data")  
coplot(mpg ~ disp | as.factor(cyl), data = mtcars, panel = panel.smooth, rows = 1)
```

4. Another exercise is about loading. (R is NOT great for storing data as a data aggregator, keep that on mind.

- a. For reading data from a .csv file

```
spambase <- read.csv(file="~/Documents/book/data/spam/spambase.data.txt", header=FALSE)  
create your own .csv file and check out how it reads and work with the data using the R  
commands you know.
```