Getting Started with R

INFO 4100 Learning Analytics

[[ADD YOUR NAME, CORNELL ID]]

R Markdown Notebooks

This is a R Markdown (http://rmarkdown.rstudio.com) Notebook. It allows you to combine text and code in one file that can be shared with ease (like a Jupyter notebook). When you execute code within the notebook, the results appear beneath the code. You need to wrap code as shown below using ```{r}.

To view the resulting notebook that combines text with R output, you can knit it. Click the *Knit to Word* button or press *Cmd+Shift+K* to generate a Word Document (doc, docx). A knitted Word file is how you will save your homework and upload it for grading on EdX. The system only accepts .doc or .docx files.

Install useful packages

Base R has many great functions. But the real power of R is unlocked by loading other packages. In R, you first need to install a package using the `install.packages("packageName")` command, but you only need to do that one time. Once it is installed, you can simply load the package with the `library(packageName)` command. Note that when you install it, you need quotation marks, but not when you load it.

Check out this overview of data mining packages (https://cran.r-project.org/doc/contrib/YanchangZhao-refcard-data-mining.pdf) available in R.

Below you see a code snippet inserted in this notebook. It has the `eval=FALSE` argument so it won't run when you compile this notebook. Loading the following packages will take 2-4 minutes.

```
# This is just a comment to show you how to add comments to your code.
# To run the next line, click on it and press Command+Return (Mac) / Control+Return (PC)
install.packages(c("tidyverse", "class", "randomForest", "data.table", "glmnet", "party"))
```

Load a package

Now that you have installed several packages you can try loading one of them. You need to load (but not install) relevant packages each time you restart R:

```
library(tidyverse)
```

Make up some data

```
var1 = 1:20 # the colon is a quick way to get sequences of numbers
var1
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

```
var2 = rnorm(n = 20, mean = 0, sd = 1) # random sample from Normal distribution
var2
```

```
##  [1]  0.93766817 -1.43432665 -1.31814098  1.11232742 -0.23888521  1.54637342
##  [7] -1.13896303 -1.18950192  0.27919568 -0.97318053 -0.01430232 -0.52478887
## [13]  0.36429725 -0.83862762 -0.75887565  0.40872221 -0.30559376  0.40878750
## [19] -1.37693314 -0.67047711
```

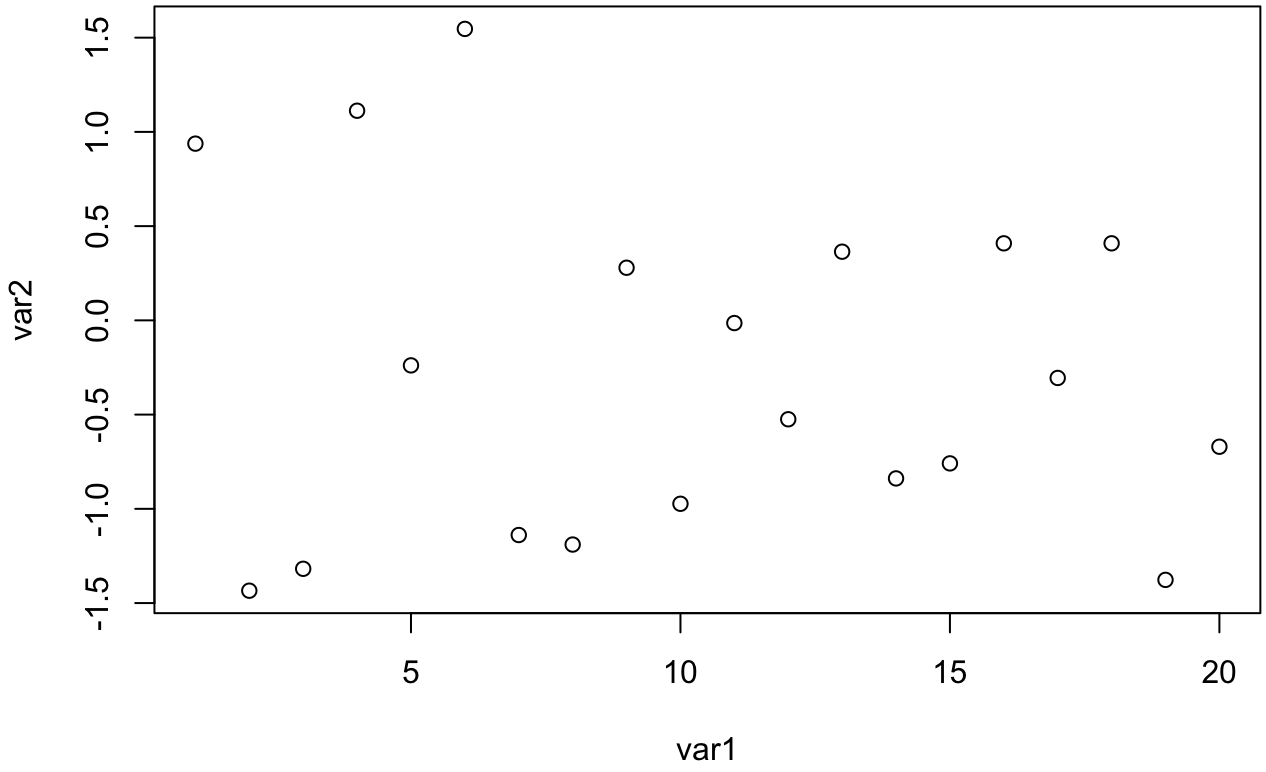
```
var3 = sample(x = c("blue", "green", "red"), size = 20, replace = T) # sampling words
var3
```

```
##  [1] "red"   "green" "red"   "red"   "red"   "red"   "red"   "blue"  "blue"
## [10] "green" "green" "green" "red"   "green" "blue"  "red"   "red"   "green"
## [19] "red"   "green"
```

Make simple plots

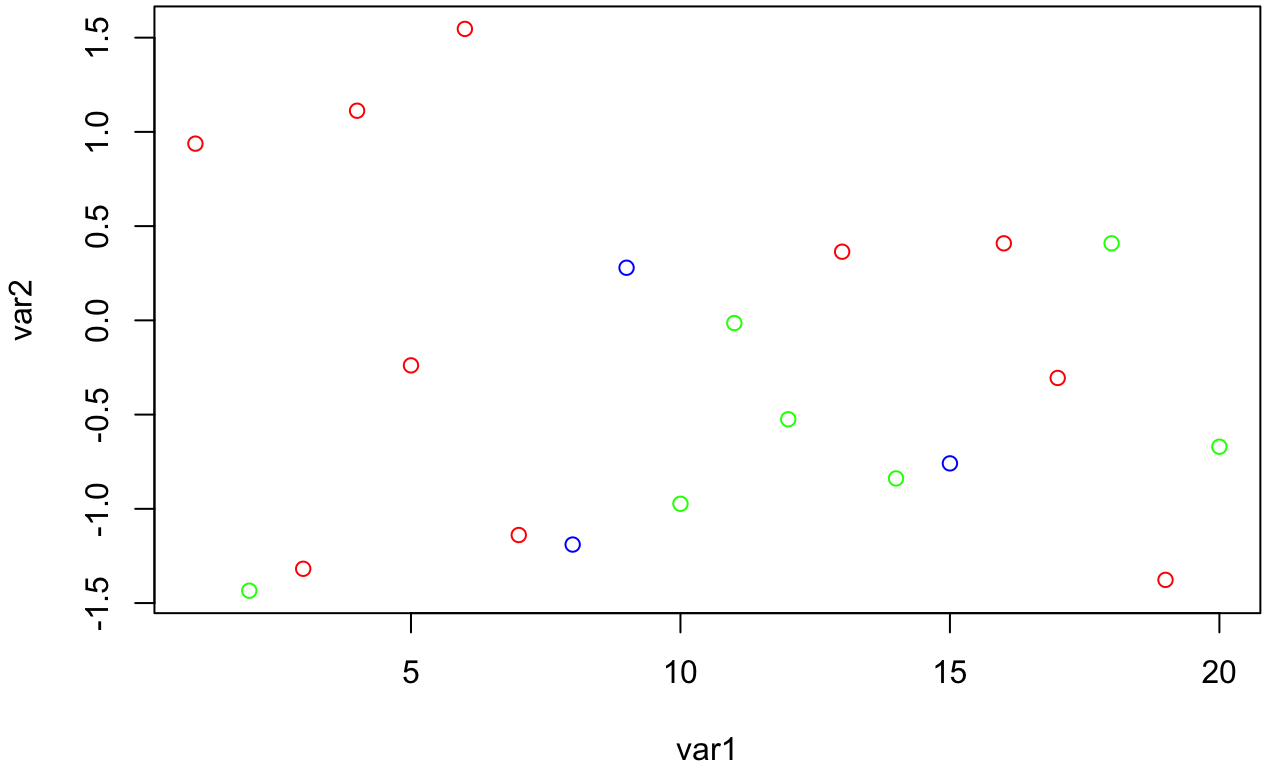
A scatterplot

```
plot(var1, var2)
```



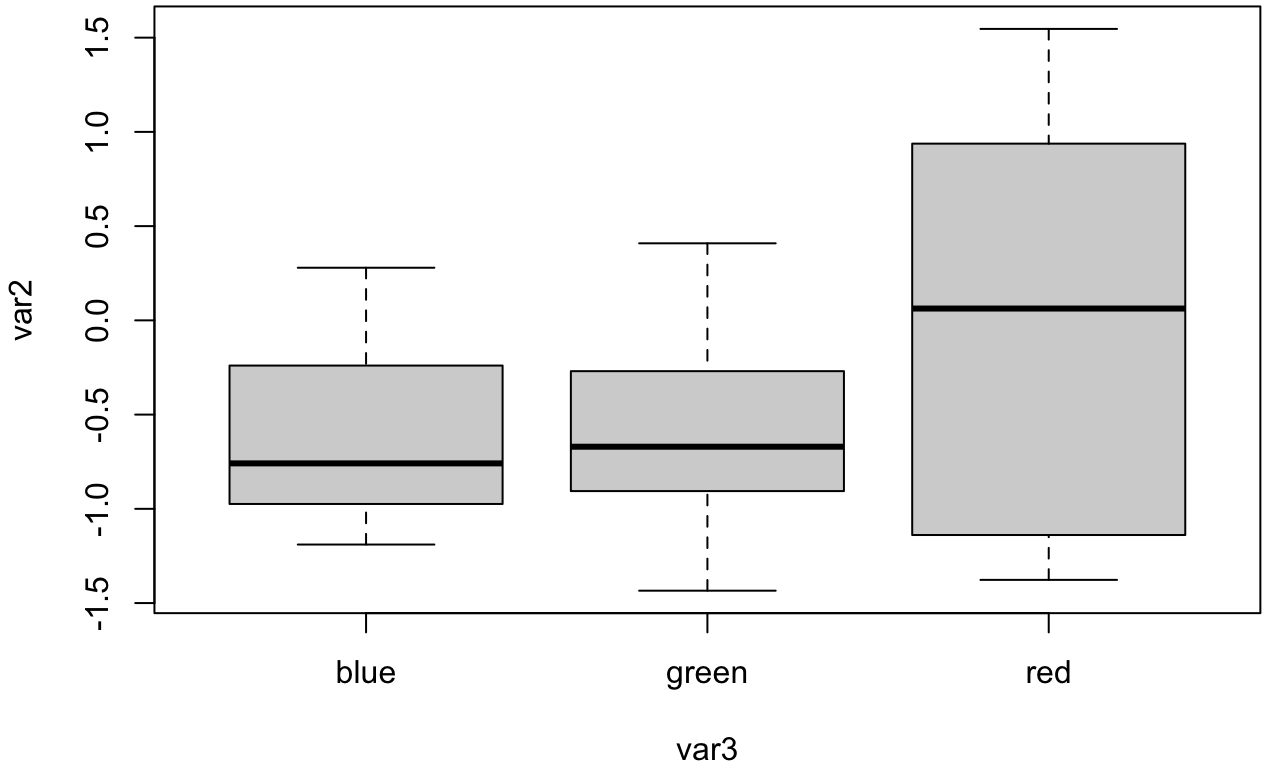
A scatterplot with colors

```
plot(var1, var2, col = var3)
```



A boxplot

```
boxplot(var2 ~ var3)
```


var2
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
blue
green
red
var3