# R Notebook

Loading the data. I choose to analyze the soccer datasets.

```
my_path<-"C:/Users/Caio Laptop/OneDrive - The University of Kansas/Documents/PhD/11. Courses/19. EECS 73
setwd(my_path)

spi_matches<-read.csv(paste(my_path,"/Datasets/spi_matches.csv", sep=""),header=T)
spi_global_rankings_intl<-read.csv(paste(my_path,"/Datasets/spi_global_rankings_intl.csv", sep=""),heade
spi_global_rankings<-read.csv(paste(my_path,"/Datasets/spi_global_rankings.csv", sep=""),header=T)
```

Now I will analyze my 3 datasets.

```
names(spi_matches)
```

```
##  [1] "date"        "league_id"   "league"      "team1"       "team2"
##  [6] "spi1"        "spi2"        "prob1"       "prob2"       "probtie"
## [11] "proj_score1" "proj_score2" "importance1" "importance2" "score1"
## [16] "score2"      "xg1"         "xg2"         "nsxg1"       "nsxg2"
## [21] "adj_score1"  "adj_score2"
```

```
head(spi_matches)
```

```
##         date league_id                  league         team1
## 1 2016-08-12      1843          French Ligue 1        Bastia
## 2 2016-08-12      1843          French Ligue 1     AS Monaco
## 3 2016-08-13      2411 Barclays Premier League     Hull City
## 4 2016-08-13      2411 Barclays Premier League       Burnley
## 5 2016-08-13      2411 Barclays Premier League Middlesbrough
## 6 2016-08-13      2411 Barclays Premier League   Southampton
##                 team2  spi1  spi2  prob1  prob2 probtie proj_score1
## 1 Paris Saint-Germain 51.16 85.68 0.0463 0.8380  0.1157        0.91
## 2           Guingamp 68.85 56.48 0.5714 0.1669  0.2617        1.82
## 3     Leicester City 53.57 66.81 0.3459 0.3621  0.2921        1.16
## 4       Swansea City 58.98 59.74 0.4482 0.2663  0.2854        1.37
## 5         Stoke City 56.32 60.35 0.4380 0.2692  0.2927        1.30
## 6            Watford 69.49 59.33 0.5759 0.1874  0.2367        1.91
##   proj_score2 importance1 importance2 score1 score2  xg1  xg2 nsxg1 nsxg2
## 1        2.36        32.4        67.7      0      1 0.97 0.63  0.43  0.45
## 2        0.86        53.7        22.9      2      2 2.45 0.77  1.75  0.42
## 3        1.24        38.1        22.2      2      1 0.85 2.77  0.17  1.25
## 4        1.05        36.5        29.1      0      1 1.24 1.84  1.71  1.56
## 5        1.01        33.9        32.5      1      1 1.40 0.55  1.13  1.06
## 6        1.05        34.1        30.7      1      1 1.05 0.22  1.52  0.41
##   adj_score1 adj_score2
## 1       0.00       1.05
## 2       2.10       2.10
## 3       2.10       1.05
## 4       0.00       1.05
## 5       1.05       1.05
## 6       1.05       1.05
```

```
dim(spi_matches)
```

```
## [1] 20879    22
```

```
typeof(spi_matches)
```

```
## [1] "list"
```

```
str(spi_matches)
```

```
## 'data.frame':    20879 obs. of  22 variables:
##  $ date       : Factor w/ 839 levels "2016-08-12","2016-08-13",..: 1 1 2 2 2 2 2 2 2 2 ...
##  $ league_id  : int  1843 1843 2411 2411 2411 2411 2411 2411 1843 2411 ...
##  $ league     : Factor w/ 37 levels "Argentina Primera Division",..: 13 13 4 4 4 4 4 4 13 4 ...
##  $ team1      : Factor w/ 698 levels "1. FC Heidenheim 1846",..: 78 50 319 120 406 581 213 180 103 39
##  $ team2      : Factor w/ 698 levels "1. FC Heidenheim 1846",..: 473 295 369 613 604 682 634 685 593
##  $ spi1       : num  51.2 68.8 53.6 59 56.3 ...
##  $ spi2       : num  85.7 56.5 66.8 59.7 60.4 ...
##  $ prob1      : num  0.0463 0.5714 0.3459 0.4482 0.438 ...
##  $ prob2      : num  0.838 0.167 0.362 0.266 0.269 ...
##  $ probtie    : num  0.116 0.262 0.292 0.285 0.293 ...
##  $ proj_score1: num  0.91 1.82 1.16 1.37 1.3 1.91 1.47 1.35 1.39 2.69 ...
##  $ proj_score2: num  2.36 0.86 1.24 1.05 1.01 1.05 1.38 1.14 1.14 0.48 ...
##  $ importance1: num  32.4 53.7 38.1 36.5 33.9 34.1 31.9 43.6 37.9 73 ...
##  $ importance2: num  67.7 22.9 22.2 29.1 32.5 30.7 48 34.6 44.2 27 ...
##  $ score1     : int  0 2 2 0 1 1 1 0 3 2 ...
##  $ score2     : int  1 2 1 1 1 1 1 1 2 1 ...
##  $ xg1        : num  0.97 2.45 0.85 1.24 1.4 1.05 0.73 1.11 1.03 2.14 ...
##  $ xg2        : num  0.63 0.77 2.77 1.84 0.55 0.22 1.11 0.68 1.84 1.25 ...
##  $ nsxg1      : num  0.43 1.75 0.17 1.71 1.13 1.52 0.88 0.84 1.1 1.81 ...
##  $ nsxg2      : num  0.45 0.42 1.25 1.56 1.06 0.41 1.81 1.6 2.26 0.92 ...
##  $ adj_score1 : num  0 2.1 2.1 0 1.05 1.05 1.05 0 3.12 2.1 ...
##  $ adj_score2 : num  1.05 2.1 1.05 1.05 1.05 1.05 1.05 1.05 2.1 1.05 ...
```

```
names(spi_global_rankings_intl)
```

```
## [1] "rank"   "name"   "confed" "off"    "def"    "spi"
```

```
head(spi_global_rankings_intl)
```

```
##   rank      name   confed  off  def   spi
## 1    1    Brazil CONMEBOL 3.11 0.29 92.96
## 2    2     Spain     UEFA 3.46 0.48 92.54
## 3    3   Belgium     UEFA 3.06 0.54 89.10
## 4    4    France     UEFA 2.84 0.46 88.57
## 5    5   Germany     UEFA 2.96 0.56 87.93
## 6    6 Argentina CONMEBOL 2.57 0.49 85.53
```

```
dim(spi_global_rankings_intl)
```

```
## [1] 213   6
```

```
typeof(spi_global_rankings_intl)
```

```
## [1] "list"
```

```
str(spi_global_rankings_intl)
```

```
## 'data.frame':    213 obs. of  6 variables:
##  $ rank  : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ name  : Factor w/ 213 levels "Afghanistan",..: 28 175 19 68 74 8 61 150 133 203 ...
##  $ confed: Factor w/ 6 levels "AFC","CAF","CONCACAF",..: 4 6 6 6 6 4 6 6 6 4 ...
```

```
##  $ off   : num  3.11 3.46 3.06 2.84 2.96 2.57 2.32 2.38 2.55 2.3 ...
##  $ def   : num  0.29 0.48 0.54 0.46 0.56 0.49 0.51 0.56 0.68 0.54 ...
##  $ spi   : num  93 92.5 89.1 88.6 87.9 ...
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.