

R Markdown

Dit is het begin van ons RMD bestand voor de capstone!

Including libraries

```
if (!require("dplyr")) install.packages("dplyr")

## Loading required package: dplyr
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
if (!require("RMySQL")) install.packages("RMySQL")

## Loading required package: RMySQL
## Loading required package: DBI
library(dplyr)
library(RMySQL)
```

Importing data using mySQL

Waar bestaat de dataset uit?

```
my_db <- dbConnect(MySQL(),
  user='klaasberbee',
  password='1190KVU',
  dbname='Batting',
  host='localhost')

print(dbListTables(my_db))

## [1] "AllstarFull"          "Appearances"          "AwardsManagers"
## [4] "AwardsPlayers"       "AwardsShareManagers"  "AwardsSharePlayers"
## [7] "Batting"             "BattingPost"          "CollegePlaying"
## [10] "Fielding"            "FieldingOF"           "FieldingOFsplit"
## [13] "FieldingPost"        "HallOfFame"           "HomeGames"
## [16] "Managers"            "ManagersHalf"         "Master"
## [19] "Parks"               "Pitching"             "PitchingPost"
## [22] "Salaries"            "Schools"              "SeriesPost"
## [25] "Teams"               "TeamsFranchises"      "TeamsHalf"

dbTables <- dbListTables(my_db)

allTables <-
  lapply(dbTables, function(table) {
```

```

    dbGetQuery(my_db, paste("select * from", table))
  })
names(allTables) <- paste("db", dbTables, sep = "_")
list2env(allTables, envir = .GlobalEnv)

```

```
## <environment: R_GlobalEnv>
```

waar bestaat de dataset uit?

De totale data-set bestaat uit zevenentwintig tabels: [1] "AllstarFull" "Appearances" "AwardsManagers" "AwardsPlayers"

[5] "AwardsShareManagers" "AwardsSharePlayers" "Batting" "BattingPost"

[9] "CollegePlaying" "Fielding" "FieldingOF" "FieldingOFsplit"

[13] "FieldingPost" "HallOfFame" "HomeGames" "Managers"

[17] "ManagersHalf" "Master" "Parks" "Pitching"

[21] "PitchingPost" "Salaries" "Schools" "SeriesPost"

[25] "Teams" "TeamsFranchises" "TeamsHalf"

In de dataset is - mlb_batting.dat, hierin wordt er per rij de voor en achternaam van een slagman weergegeven met de spelers ID. Ook worden de stats weergegeven per slagman met bijvoorbeeld het aantal home runs, strikeouts en stolen bases. - playbyplay2008.dat, hierin wordt er per rij een gebeurtenis weergegeven in het spel. hierbij wordt aangegeven wie die slagman is en wie de pitcher is. Ook wordt er gedetailleerde informatie weergegeven over hoe het spel verloopt, zoals het aantal slagen en de hoeveelheid spelers er op een honk staan. - roster2008.dat, hierin wordt aangegeven wat de beste pichers zijn? - pitchfx.dat, hierin wordt er per rij de gebeurtenissen weergegeven per pitch met bijvoorbeeld de soort pitch, de snelheid, de beweging en de locatie van de pitch in de strike zone. Ook wordt er informatie gegeven over de pitcher en de slag man.

Om een duidelijker beeld te krijgen over wat er precies in de vier tabels staat kan bijlage 1 geraatpleegd worden.