

# Assignment #2 - SQL and R

Stefano Biguzzi

9/1/2020

## Create Connection to Local PostgreSQL database and getting tables

Creating connection

```
con <- dbConnect(  
  RPostgres::Postgres(),  
  dbname = "MovieRatings",  
  host="localhost",  
  port="5432",  
  user="postgres",  
  password=params$pwd)
```

Setting tables to dataframes

Loading tblPersons and removing any whitespaces

```
person_df <- dbGetQuery(con, 'Select * From public."tblPersons"')  
  
person_df$PersonID <- trimws(  
  person_df$PersonID,  
  which = c("both", "left", "right"),  
  whitespace = "[ \\t\\r\\n]")
```

Loading tblMovies and removing any whitespaces

```
movie_df <- dbGetQuery(con, 'Select * From public."tblMovies"')  
  
movie_df$MovieID <- trimws(  
  movie_df$MovieID,  
  which = c("both", "left", "right"),  
  whitespace = "[ \\t\\r\\n]")
```

Loading tblRatings and removing any whitespaces

```
ratings_df <- dbGetQuery(con, 'Select * From public."tblRatings"')  
  
ratings_df$MovieID <- trimws(  
  ratings_df$MovieID,
```

```

which = c("both", "left", "right"),
whitespace = "[ \\t\\r\\n]")

ratings_df$PersonID <- trimws(
  ratings_df$PersonID,
  which = c("both", "left", "right"),
  whitespace = "[ \\t\\r\\n]")

```

## Dealing with missing data

In the ratings dataframe for the MovieRating column, I decided to create two dataframes, one for seen movies with ratings and one for non seen movies. This can allow us to run different statistical analyses on the different types of movies, seen and unseen.

Figure out how many nulls there are in the MovieRatings column

```
table(is.na(ratings_df$MovieRating))
```

```
##
## FALSE  TRUE
##    37    23
```

Create dataframe for movie ratings that were seen

```

seen_ratings_df <-
  subset(ratings_df, !is.na(ratings_df$MovieRating))
rownames(seen_ratings_df) <- NULL

```

Create dataframe for movies that were not seen

```

not_seen_df <-
  subset(ratings_df, is.na(ratings_df$MovieRating))
rownames(not_seen_df) <- NULL
not_seen_df <-
  subset(not_seen_df, select = -MovieRating)

```