---

title: "DATA607_Assg2"

author: "Alexis Mekueko"

date: "9/10/2020"

Github link to Repo: https://github.com/asmozo24/DATA607_Survey_R_SQL.git

output:

  pdf_document: default

  html_document: default

```{r setup, include=FALSE}

knitr::opts_chunk$set(echo = TRUE)

```

R Packages

```{r load-packages, message=FALSE}

library(tidyverse) #loading all library needed for this assignment

library(readxl)

library(plyr)

library(dplyr)

library(DBI)

library(dbplyr)

library(data.table)

library(rstudioapi)

library(RJDBC)

library(odbc)

library(RSQLite)

```

R Markdown

```{r connection to sql server}
```

```
# Something not working with this connection:
# require(RJDBC)
# drv <- JDBC("com.microsoft.sqlserver.jdbc.SQLServerDriver",
#   "C:\Program Files\Java\jdk1.8.0_31\")
#   conn <- dbConnect(drv, "jdbc:sqlserver://serverName", "userID", "password")
```

Connecting to SQL server

```{r connection to sql}
# ## this connection work but requires password, I am going to comment this block because preventing the knit process
 con <- dbConnect(odbc(),
                  Driver = "SQL Server",
                  Server = "ATM\\ATMSERVER", #server name
                  Database = "Data607_surveyDB", #Db that contain the field I am looking for
                  UID = "Alex", # user name for login
                  PWD = rstudioapi::askForPassword("Database password"), # login password for authentication
                  Port = 1433)

# ## Trying different ways of looking at table from the database
# ##install.packages(c("dbplyr", "RSQLite"))            #installing package to able to use library dbplyrn, sqlite and use dbi call.

# ##-- !preview conn=con
# ##SELECT * FROM "Data607_surveyDB"."dbo"."Survey_Top06Movies"  ###This string does not work well

 dbListFields(con, "Survey_Top06Movies") # list all the variables (fields) in the Survey_TopMovies
 Mydata <- dbReadTable(con, "Survey_Top06Movies") ## Read Survey_Top06Movies full table into Mydata
#Mydata <- dbWriteTable(con, "Survey_Top06Movies", Survey_Top06Movies)  ##Writing a data fram , Mydata to an SQL table, not sure if that can write to a new table...I will try this string later...
 Mydata ## Now we succeeded in reading Survey_Top06Movies full table into R data fram Mydata.....
# ##src_dbi(Data607_surveyDB) #(Survey_Top06Movies) ## this does not work , still cannot figure out
```

Survey_Id <int>	Tenet <chr>	Mulan <chr>	Black_Panther <chr>	Bill_Ted_Face_The_Music <chr>	The_Karate_Kid <chr>	No_Time_to_Die <chr>
1	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
2	4	5	5	4	5	5
3	2	4	5	2	1	4
4	3	3	5	3	3	4
5	4	3	5	2	2	5
6	5	5	5	5	2	5
7	2	4	5	5	4	5
8	5	5	5	4	4	4
9	4	5	5	5	5	5
10	5	4	5	5	4	3

1-10 of 15 rows Previous 1 2 Next

```{sql connection=con}

/*

#this connection works, but I am going to comment the block out because the knit is erroring out here

select * from Survey_Top06Movies --display Survey_Top06Movies

--Let see how we can handle missing data from Survey_Top06Movies

--delete from Survey_Top06Movies where Tenet=NA --- and (Mulan=NA) and (Black_Panther=NA) and (Bill_Ted_Face_The_Music=NA) and (The_Karate_Kid=NA) and (No_Time_to_Die =NA) --this query should delete all NA( IS NA does not work... "IS NULL usually work or ='' will delete row with empty space ) field in the Survey_Top06Movies, wondering if there is another way of deleting empty field (NA) from any colunm without calling each colunm

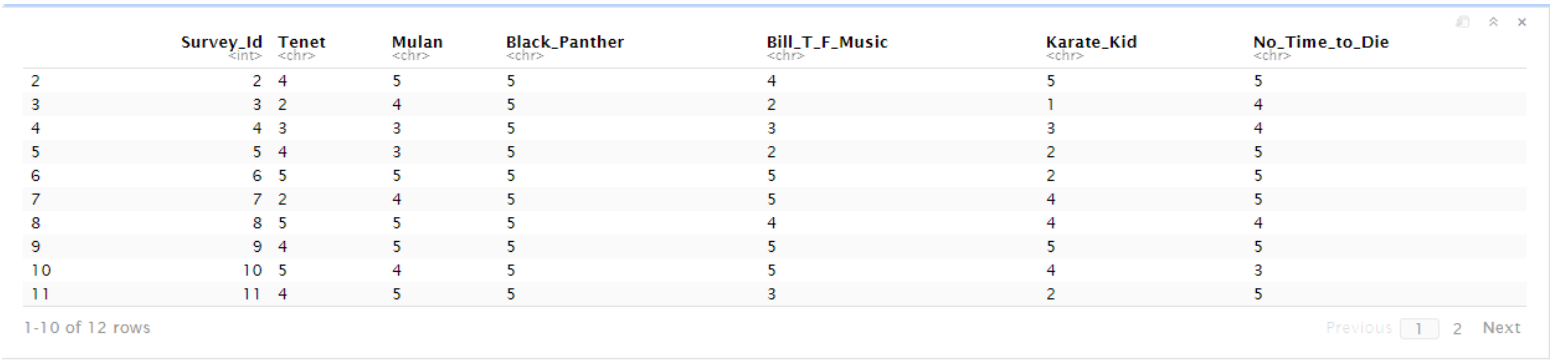
--delete from Survey_Top06Movies where coalesce (Tenet, Mulan, Black_Panther, Bill_Ted_Face_The_Music, The_Karate_Kid,No_Time_to_Die) IS NULL; --using COALESCE

*/

##```

##Handling missings data

```{r Survey_Top06Movies}

Mydata2<- rename(Mydata, c("Bill_Ted_Face_The_Music" = "Bill_T_F_Music", "The_Karate_Kid" = "Karate_Kid")) # Renaming variables

##Mydata2 <- Mydata

Mydata2 #checking new table with rename, let's remove the NA

is.na(Mydata2) # checking if there is a missing data in the dataset, return is yes

sum(is.na(Mydata2)) # file to big, checking the sum of all missing data (return is 09 missing data)

Mydata2 <- na.omit(Mydata2) # delete/remove the missings data because it is an imcomplete observation

Mydata2

## Mydata %>% rename(Bill_Ted_Face_The_Music = Bill_T_F_Music) # not working

```



	Survey_Id <int>	Tenet <chr>	Mulan <chr>	Black_Panther <chr>	Bill_T_F_Music <chr>	Karate_Kid <chr>	No_Time_to_Die <chr>
2	2	4	5	5	4	5	5
3	3	2	4	5	2	1	4
4	4	3	3	5	3	3	4
5	5	4	3	5	2	2	5
6	6	5	5	5	5	2	5
7	7	2	4	5	5	4	5
8	8	5	5	5	4	4	4
9	9	4	5	5	5	5	5
10	10	5	4	5	5	4	3
11	11	4	5	5	3	2	5

1-10 of 12 rows Previous 1 2 Next

Data Analysis

```{r }

Mydata2 # loading the data, this is the data that generate the plot

## Mydata2 <- keep.No_need <- names(Mydata2) %in% c("")

## Mydata2

## clean.crs <- mtcars [! keep.cols]

## dataframe$column_to_remove <- NULL # delete a column

##dataset$first_column <- dataset$next_column <- dataset$another_column <- NULL # delete multiple column

summary(Mydata2) # What kind of data I have in Mydata2

str(Mydata2) # what is the structure of the data, just to have a better look

Mydata2

#length(Mydata2)

#length(Mydata2$Tenet)

#length(Mydata2$Tenet == 1) # not working should return 0

#nrow(subset(Mydata2, Tenet = 5))

#MyPlot <- count(Mydata2$Tenet = '5') not good

##MyPlot not good

## count(Mydata2, 'Tenet') not good

##w = table(Mydata2$Mulan) ##Cannot figure out

##w

##group_by(Mydata2, Tenet) # not good change all the values to 5 in column = Tenet

## filter(Mydata2, Mulan =="5")

##as.data.frame(table(Mydata2)) # cray respone

##filter(Mydata2, Tenet == 4) ##, Mulan == 5, Black_Panther == 5, Bill_T_F_Music == 5, Karate_Kid == 5, No_Time_to_Die == 5) not working

# Filter call seem to work, at least I can find the the best movie, by looking how many rows  received 5 as a rating

# it is teadous but I don't know how to sum all them up

# I want to be able to use frequency by each rating and plot them as bar

# with a result of filter by 5, I could have a plot where Y = Tenet # rows, Mulan #rows,...to last variable, then y = Tenet, Mulan, Black_Panther...etc ) and do a barplot but all of this is teadeous...

filter(select(Mydata2,Survey_Id, Tenet), Tenet > 5) # return 03 rows...

filter(select(Mydata2,Survey_Id, Mulan), Mulan > 5) # return 05 rows

filter(select(Mydata2,Survey_Id, Black_Panther), Black_Panther > 5) # return 11 rows

filter(select(Mydata2,Survey_Id, Bill_T_F_Music), Bill_T_F_Music > 5) # return 4 rows

filter(select(Mydata2,Survey_Id, Karate_Kid), Karate_Kid > 5) # return 2 rows

filter(select(Mydata2,Survey_Id, No_Time_to_Die), No_Time_to_Die > 5) # return 7 rows

# ;So out of this filter we can see that Black Panther is the best movie as it receive 5 as rating 11 times

# another way could be to do a count of each recurrence and assign the result at the last column

#select(Mydata2, Tenet,Mulan, Black_Panther, Bill_T_F_Music, Karate_Kid, No_Time_to_Die ) # this work, to select columns you want

# view(Mydata2) # view table in a sheet

#filter(select(Mydata2, Tenet), Tenet > 5) # show rows that have value equal or greater than 5

#filter(select(Mydata2, Survey_Id, Tenet), Tenet > 5) # bring row where value  equal or great than 5

#filter(select(Mydata2, Survey_Id, Tenet, Mulan), Tenet > 5, Mulan > 5) # this bring row of same value

#count(filter(select(Mydata2, Survey_Id, Tenet), Tenet > 5))

#Mydata3 <- aggregate(No_Time_to_Die ~ Karate_Kid ~ Bill_T_F_Music ~ Black_Panther ~ Mulan ~ Tenet, data=Mydata2, FUN=table)

#data.frame(Tenet=Mydata3$Tenet, Mydata3$No_Time_to_Die, Mydata3$Karate_Kid, Mydata3$Bill_T_F_Music, Mydata3$Black_Panther, Mydata3$Mulan , Mydata3$Tenet)

```

Survey_Id <int>	Tenet <chr>
6	5
8	5
10	5

3 rows

Including Plots

```{r BestMovie, echo=FALSE}

BestMovie <- data.table(Movies = c("Tenet" , "Mulan", "Black_Panther", "Bill_T_F_Music", "Karate_Kid", "No_Time_to_Die"), Number_5Rating = c(3, 5, 11, 4, 2, 7))

BestMovie

#Movies <- c("Tenet" , "Mulan", "Black_Panther", "Bill_T_F_Music", "Karate_Kid", "No_Time_to_Die")

#Rating <- c(3, 5, 11, 4, 2, 7)

#png(file = "barchart_Best_Movies.png")

# Plot the bar chart

#barplot(Rating,names.arg=Movies,xlab="Movie Title",ylab="Movie Name",col="yellow",

#main="Best Movie",border="black")

```

Movies <chr>	Number_5Rating <dbl>
Tenet	3
Mulan	5
Black_Panther	11
Bill_T_F_Music	4
Karate_Kid	2
No_Time_to_Die	7

6 rows

Including Plots

```{r}

##This block work

#BestMovie <- data.table(Movies = c("Tenet" , "Mulan", "Black_Panther", "Bill_T_F_Music", "Karate_Kid", "No_Time_to_Die"),
#Number_5Rating = c(3, 5, 11, 4, 2, 7))

##BestMovie

Movies <- c("Tenet" , "Mulan", "Black_Panther", "Bill_T_F_Music", "Karate_Kid", "No_Time_to_Die")

Rating <- c(3, 5, 11, 4, 2, 7)

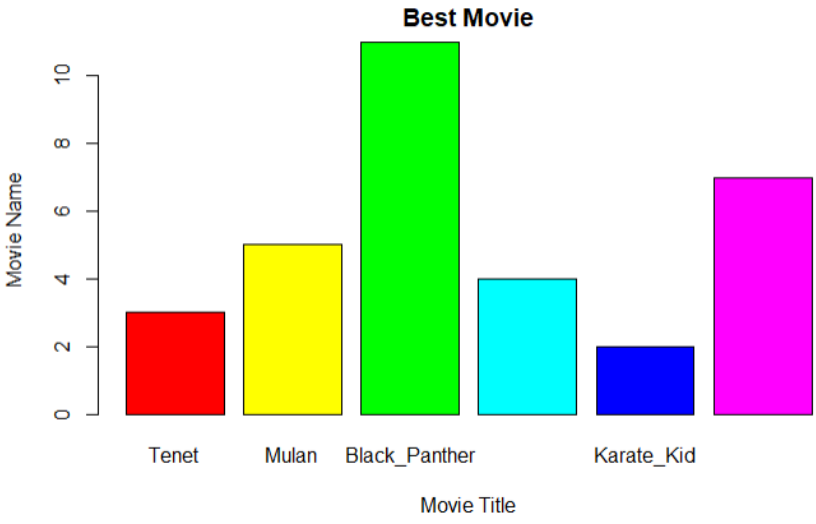
png(file = "barchart_Best_Movies.png")

# Plot the bar chart

barplot(Rating,names.arg=Movies,xlab="Movie Title",ylab="Movie Name",col= rainbow(6),

main="Best Movie",border="black")

```


```