Assignment

2023-01-29

Assignment 1 - Introduction to R

Source of dataset - Kaggle.com

https://www.kaggle.com/datasets/ritesh2000/trending-tv-shows-on-netflix (https://www.kaggle.com/datasets/ritesh2000/trending-tv-shows-on-netflix)

Title of the dataset: Trending TV shows on Netflix

#Data Content: #Title: Name of the TV show #Year: Year in which the show was released #Rating: Ratings given by Netflix #IMDB_Ratings: Ratings by IMDB #Netflix: If the show is currently streaming on Netflix or not

```
library(readxl)
Netflix_data <- read_excel("Netflix_data.xls")
View(Netflix_data)
```

#Question 1: #Print out descriptive statistics for quantitative and qualitative variables

```
summary(Netflix_data)
```

```
##     Titles               Year         Rating           IMDB_Rating
##  Length:50          Min.   :1989   Length:50          Min.   :8.000
##  Class :character   1st Qu.:2005   Class :character   1st Qu.:8.500
##  Mode  :character   Median :2011   Mode  :character   Median :8.700
##                     Mean   :2010                      Mean   :8.722
##                     3rd Qu.:2015                      3rd Qu.:8.900
##                     Max.   :2019                      Max.   :9.500
##     Netflix
##  Min.   :0.00
##  1st Qu.:0.00
##  Median :0.00
##  Mean   :0.38
##  3rd Qu.:1.00
##  Max.   :1.00
```

```
table(Netflix_data$Year)
```

```
##
## 1989 1994 1997 1999 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011 2013 2014
##    1    1    1    1    1    2    2    4    1    2    2    4    2    3    6    2
## 2015 2016 2017 2019
##    4    3    4    4
```

```
table(Netflix_data$Titles)
```

```
##
##                  Attack on Titan        Avatar: The Last Airbender
##                                1                                 1
##                Band of Brothers                Better Call Saul
##                                1                                 1
##                    Black Mirror                    Breaking Bad
##                                1                                 1
##               Brooklyn Nine–Nine                       Chernobyl
##                                1                                 1
##                       Community                            Dark
##                                1                                 1
##                          Dexter                           Fargo
##                                1                                 1
##                         Firefly                         Friends
##                                1                                 1
##                          Fringe  Fullmetal Alchemist: Brotherhood
##                                1                                 1
##                 Game of Thrones                         Hannibal
##                                1                                 1
##                           House It's Always Sunny in Philadelphia
##                                1                                 1
##                            Lost                           Lucifer
##                                1                                 1
##                         Mad Men                Marvel's Daredevil
##                                1                                 1
##                      Mindhunter                   Modern Family
##                                1                                 1
##                     Money Heist                       Mr. Robot
##                                1                                 1
##                          Narcos                           Ozark
##                                1                                 1
##              Parks and Recreation                  Peaky Blinders
##                                1                                 1
##                   Rick and Morty                        Sherlock
##                                1                                 1
##                      South Park                  Stranger Things
##                                1                                 1
##                           Suits                     Supernatural
##                                1                                 1
##               The Big Bang Theory                 The Mandalorian
##                                1                                 1
##                      The Office                    The Simpsons
##                                1                                 1
##                    The Sopranos             The Umbrella Academy
##                                1                                 1
##                 The Walking Dead                         The Wire
##                                1                                 1
##                     The Witcher                   True Detective
##                                1                                 1
##                         Vikings                        Westworld
##                                1                                 1
```

```
library(psych)
describe(Netflix_data$Year)
```

	vars	n	mean	sd	median	trimmed	mad	min	max
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
X1	1	50	2009.9	6.746881	2011	2010.55	6.6717	1989	2019

1 row | 1-10 of 14 columns

#Question2: #Transforming categorical variable into numeric variable

```
Netflix_data$Titles=as.numeric(Netflix_data$Titles)
```

```
## Warning: NAs introduced by coercion
```

```
as.numeric(Netflix_data$Titles)
```

```
##  [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
```

#Categorical variable was converted to N/A as the data was non-numeric. #Tranforming numeric variable into categorical varaible

```
attach(Netflix_data)
names(Netflix_data)
```

```
## [1] "Titles"      "Year"        "Rating"      "IMDB_Rating" "Netflix"
```

```
Year[1:10]
```

```
##  [1] 2008 2011 2013 2017 2016 2005 2010 2019 1994 2005
```

```
CatYear <- cut(Year, breaks = c(1990,1995,2000,2005,2010,2015,2020), labels = c("A",
"B","C","D","E","F"))
CatYear[1:5]
```

```
## [1] D E E F F
## Levels: A B C D E F
```

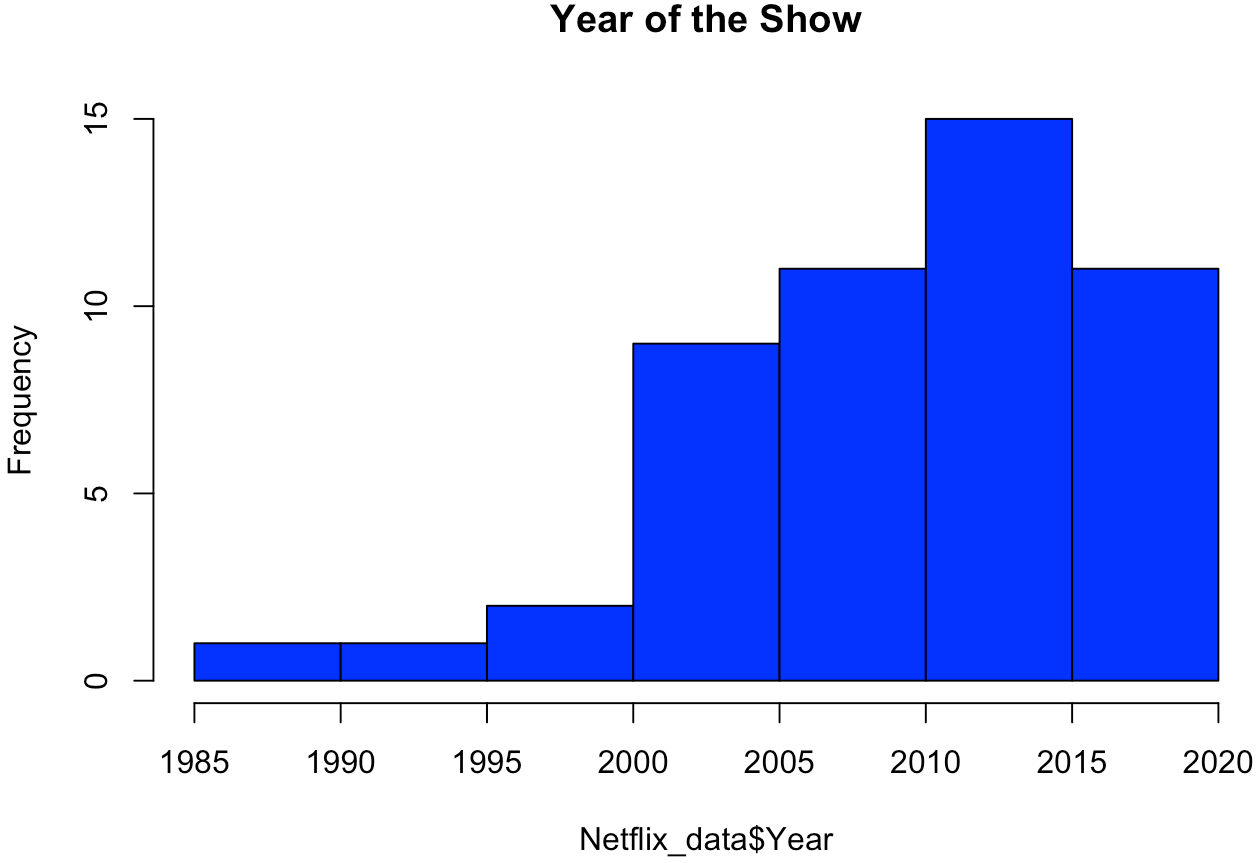
#Tranforming categorical variable into factor (numeric) variable

```
Titles=c("Breaking Bad","Game of Thrones","Breaking Bad","Game of Thrones","Rick and
Morty","Dark","Stranger Things","Avatar: The Last Airbender","Sherlock","Chernobyl","
Friends","The Office","Better Call Saul","The Wire","Supernatural","Fargo","Black Mir
ror","Band of Brothers","Attack on Titan","House","Money Heist","Peaky Blinders","Mr.
Robot","Vikings","The Sopranos","The Simpsons","The Mandalorian","Lost","Westworld","
The Umbrella Academy","Narcos","True Detective","Marvel's Daredevil","Brooklyn Nine-N
ine","The Walking Dead","Parks and Recreation","Mindhunter","South Park","Suits","Han
nibal","Dexter","Community","Fringe","Modern Family","The Big Bang Theory","Firefl
y","Mad Men","Fullmetal Alchemist: Brotherhood","Ozark","The Witcher","Lucifer","It's
Always Sunny in Philadelphia")
Titles_factor<-factor(Titles)
Titles_factor
```

```
##  [1] Breaking Bad                      Game of Thrones
##  [3] Breaking Bad                      Game of Thrones
##  [5] Rick and Morty                    Dark
##  [7] Stranger Things                   Avatar: The Last Airbender
##  [9] Sherlock                          Chernobyl
## [11] Friends                           The Office
## [13] Better Call Saul                  The Wire
## [15] Supernatural                      Fargo
## [17] Black Mirror                      Band of Brothers
## [19] Attack on Titan                   House
## [21] Money Heist                       Peaky Blinders
## [23] Mr. Robot                         Vikings
## [25] The Sopranos                      The Simpsons
## [27] The Mandalorian                   Lost
## [29] Westworld                         The Umbrella Academy
## [31] Narcos                            True Detective
## [33] Marvel's Daredevil                Brooklyn Nine-Nine
## [35] The Walking Dead                  Parks and Recreation
## [37] Mindhunter                        South Park
## [39] Suits                             Hannibal
## [41] Dexter                            Community
## [43] Fringe                            Modern Family
## [45] The Big Bang Theory               Firefly
## [47] Mad Men                           Fullmetal Alchemist: Brotherhood
## [49] Ozark                             The Witcher
## [51] Lucifer                           It's Always Sunny in Philadelphia
## 50 Levels: Attack on Titan Avatar: The Last Airbender ... Westworld
```

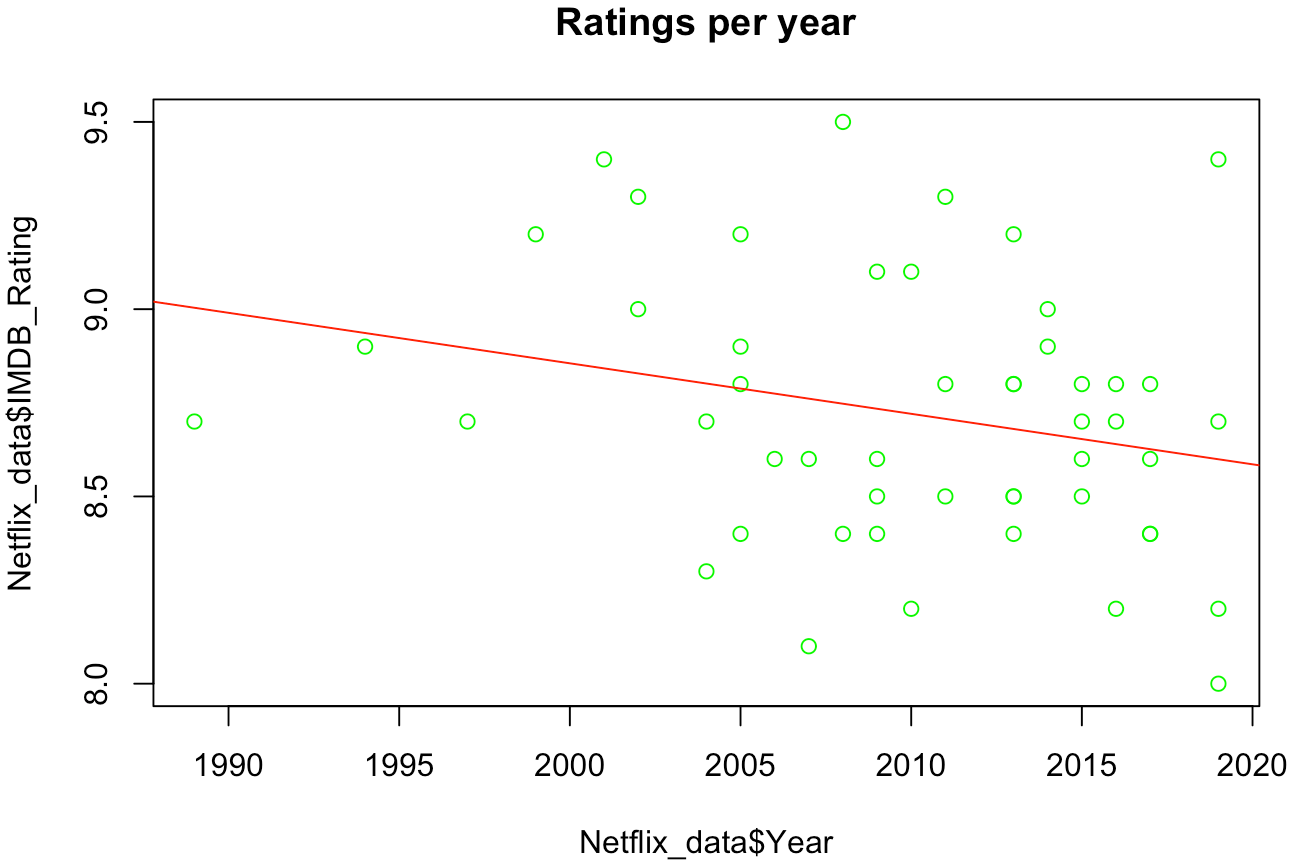
#Question 3: #Plot at least one quantitative variable, and one scatter plot #Histogram

```
variable=Netflix_data$Year
hist(Netflix_data$Year, main = "Year of the Show", col = ("blue"))
```



#Scatter Plot

```
plot(Netflix_data$Year,Netflix_data$IMDB_Rating, main = "Ratings per year", col="gree
n")
abline(lm(IMDB_Rating~Year), col=("red"))
```


Ratings per year
Netflix_data$IMDB_Rating
9.5
9.0
8.5
8.0
1990
1995
2000
2005
2010
2015
2020
Netflix_data$Year