UE20CS312 - Data Analyrics - Worksheet 1a - Part 1 - Exploring data with R

PES University

Reshmi Pradeep, Dept. of CSE - PES2UG20CS270

2022-08-24

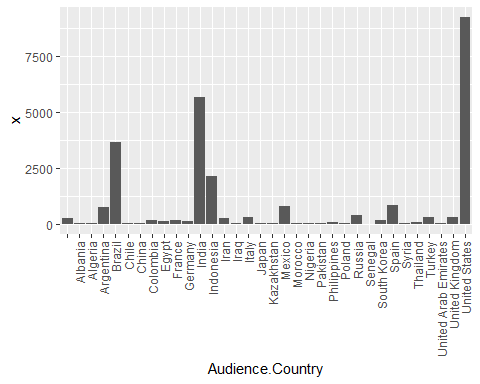
### Solutions  
  
### Problem 1  
df <- read.csv("top\_1000\_instagrammers.csv", header=TRUE)  
  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.1  
## ✔ readr 2.1.2 ✔ forcats 0.5.2  
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

df$Followers <- substr(df$Followers, 1, nchar(df$Followers) - 1)  
df$Authentic.Engagement <- substr(df$Authentic.Engagement, 1, nchar(df$Authentic.Engagement) - 1)  
df$Engagement.Avg. <- substr(df$Engagement.Avg., 1, nchar(df$Engagement.Avg.) - 1)  
  
df$Followers <- as.numeric(as.character(df$Followers))  
df$Authentic.Engagement <- as.numeric(as.character(df$Authentic.Engagement))  
df$Engagement.Avg. <- as.numeric(as.character(df$Engagement.Avg.))  
  
print(summary(df))

## Name Rank Category Followers   
## Length:1000 Min. : 1.0 Length:1000 Min. : 1.60   
## Class :character 1st Qu.: 250.8 Class :character 1st Qu.: 8.60   
## Mode :character Median : 500.5 Mode :character Median : 14.10   
## Mean : 500.5 Mean : 26.04   
## 3rd Qu.: 750.2 3rd Qu.: 25.43   
## Max. :1000.0 Max. :528.40   
##   
## Audience.Country Authentic.Engagement Engagement.Avg.  
## Length:1000 Min. : 1.0 Min. : 1.0   
## Class :character 1st Qu.:126.6 1st Qu.:128.1   
## Mode :character Median :247.8 Median :283.1   
## Mean :308.3 Mean :335.5   
## 3rd Qu.:453.2 3rd Qu.:529.2   
## Max. :990.9 Max. :998.2   
## NA's :20

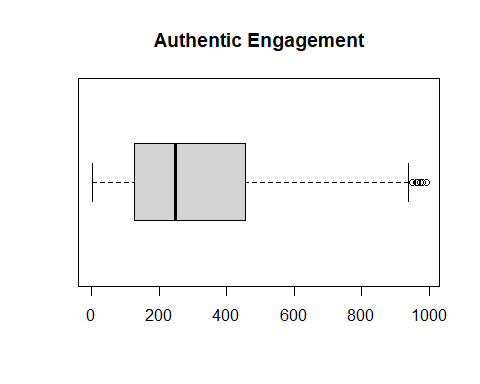
#Analysis:   
  
  
### Problem 2  
library(ggplot2)  
  
total <- aggregate(df$Followers, by=df[c('Audience.Country')], FUN=sum)  
  
ggplot(total, aes(x=Audience.Country,y=x))+ geom\_bar(stat='identity') + theme(axis.text.x = element\_text(angle=90, vjust=.5, hjust=1))



ind\_follower <- total[which(total$Audience.Country == "India"),'x']  
sprintf("Total number of followers for India -is %s", ind\_follower)

## [1] "Total number of followers for India -is 5684.3"

#United states has the most amount of followers  
  
  
### Problem3  
library(tidyverse)  
boxplot(df$Authentic.Engagement, main="Authentic Engagement", horizontal=TRUE)



### Conclusion