Exploring Ted Talks

March 2, 2022

The Dataset

the dataset contains 6 different features of each talk available on TED's website which you can find below

- title - Title of the Talk
- author - Author of Talk
- date - Date when the talk took place
- views - Number of views of the Talk
- likes - Number of likes of the Talk
- link - Link of the talk from ted.com

The data has been scraped from the official TED Website and is available under the Creative Commons License.

For this challenge you should submit a printed report to Mr. Merrick. All questions should make use of the tidyverse and ggplot2 packages.

1. How many observations are in this dataset?
2. How many variables are in this dataset?
3. Visualize the distribution for Ted Talk Video Views and provide a full description of the shape, center, outliers/unusual points, and spread of the distribution.
4. Visualize the distribution for Ted Talk Video Likes and provide a full description of the shape, center, outliers/unusual points, and spread of the distribution.
5. You may notice that the majority of Ted Talk Videos have less than 15 million views. Filter the dataset to only select these videos, and visualize the distribution using a histogram.
6. Create a scatter plot of Ted Talk Video likes against Ted Talk video views. Do these variables appear to be correlated?
7. Filter the Ted Videos to select the top 5 with the most views. Be sure to arrange the titles in descending order. To you which of these videos sounds the most interesting and why?

Solution:

```
## Load Packages
library(ggplot2)
library(tidyverse)

# Load the dataset
ted <- read.csv('ted.csv')
glimpse(ted)

```

```r
# Question 1/2
glimpse(ted) # 5440 observations (rows) and 6 variables (columns)

# Question 3
ted %>% ggplot(aes(x=views))+geom_histogram()+xlab('Views')+
  theme_classic()

# Question 4
ted %>% ggplot(aes(x=likes))+geom_histogram()+xlab('Likes')+
  theme_classic()

# Question 5
ted %>% filter(views<15000000) %>% ggplot(aes(x=views))+geom_histogram()+
  theme_classic()

# Question 6
ted %>% ggplot(aes(x=views, y=likes)) + geom_point(alpha=0.1)

# Question 6
ted %>% arrange(desc(views)) %>% select(title) %>% head(5)
```