TED Talks

Sanjay chauhan

3/1/2022

What is “TED Talks”?

TED Conferences, LLC (Technology, Entertainment, Design) is an American media organization that posts talks online for free distribution under the slogan “ideas worth spreading”. TED’s early emphasis was on technology and design, consistent with its Silicon Valley origins. It has since broadened its perspective to include talks on many scientific, cultural, political, humanitarian, and academic topics. For more information visit TED Talks

Introduction

This project is based upon the data of TED talks and it is regarding the popularity of program throughout its life span. We have shown some of the key data and interesting facts such as list of it’s most popular shows, Most popular authors, popularity of program from 1970 to 2022, views and videos relation are to name some. To check the raw data click

Exploring data-

```
df <- read.csv("ted_talks.csv")
glimpse(df)
## Rows: 5,440
## Columns: 6
## $ title  <chr> "Climate action needs new frontline leadership", "The dark
hist~
## $ author <chr> "Ozawa Bineshi Albert", "Sydney Iaukea", "Martin Reeves",
"Jame~
## $ date   <chr> "December 2021", "February 2022", "September 2021",
"October 20~
## $ views  <int> 404000, 214000, 412000, 427000, 2400, 422000, 412000,
455000, 6~
## $ likes  <int> 12000, 6400, 12000, 12000, 72, 12000, 12000, 13000, 1900,
17000~
## $ link   <chr>
"https://ted.com/talks/ozawa_bineshi_albert_climate_action_need~
```

Data cleaning

Now, if we look at the data, first thing to notice is the datatype of column “date”. And we need to rectify that and it would be better for analysis to separate month and year and form new columns based upon respective values of month and year.

second thing to notice is that the “link” column is of no use for analysis perspective so we would drop that.

```
df <- df %>% separate(date,c("month","year")," ") %>% select(-link)
glimpse(df)
## Rows: 5,440
## Columns: 6
## $ title  <chr> "Climate action needs new frontline leadership", "The dark
hist~
## $ author <chr> "Ozawa Bineshi Albert", "Sydney Iaukea", "Martin Reeves",
"Jame~
## $ month  <chr> "December", "February", "September", "October", "October",
"Oct~
## $ year   <chr> "2021", "2022", "2021", "2021", "2021", "2021", "2022",
"2021",~
## $ views  <int> 404000, 214000, 412000, 427000, 2400, 422000, 412000,
455000, 6~
## $ likes  <int> 12000, 6400, 12000, 12000, 72, 12000, 12000, 13000, 1900,
17000~
```

Month order-

```
df$month <-
ordered(df$month,level=c("January","February","March","April","May","June","J
uly","August","September","October","November","December"))
```

Datatype of year is still character, let’s rectify it-

```
df$year <- as.integer(df$year)
class(df$year)
## [1] "integer"
```

Empty cells and NAs if there any-

```
df %>% map(~sum(is.na(.)))
## $title
## [1] 0
##
## $author
## [1] 0
##
## $month
## [1] 0
##
## $year
## [1] 0
##
## $views
## [1] 0
##
## $likes
## [1] 0
df %>% filter(title==""|author=="")
##                 title author    month year views likes
```

```
## 1 Year In Ideas 2015        December 2015   532    15
```

There is one empty cell in the “author” column and the title of the respective row is “Year In Ideas 2015”. Researching the title online we come to know that this particular title is delivered by multiple authors that’s why a particular author’s name cannot be mentioned. Instead of leaving it blank it should be replaced by “Multiple” so that it does not create chaos further.

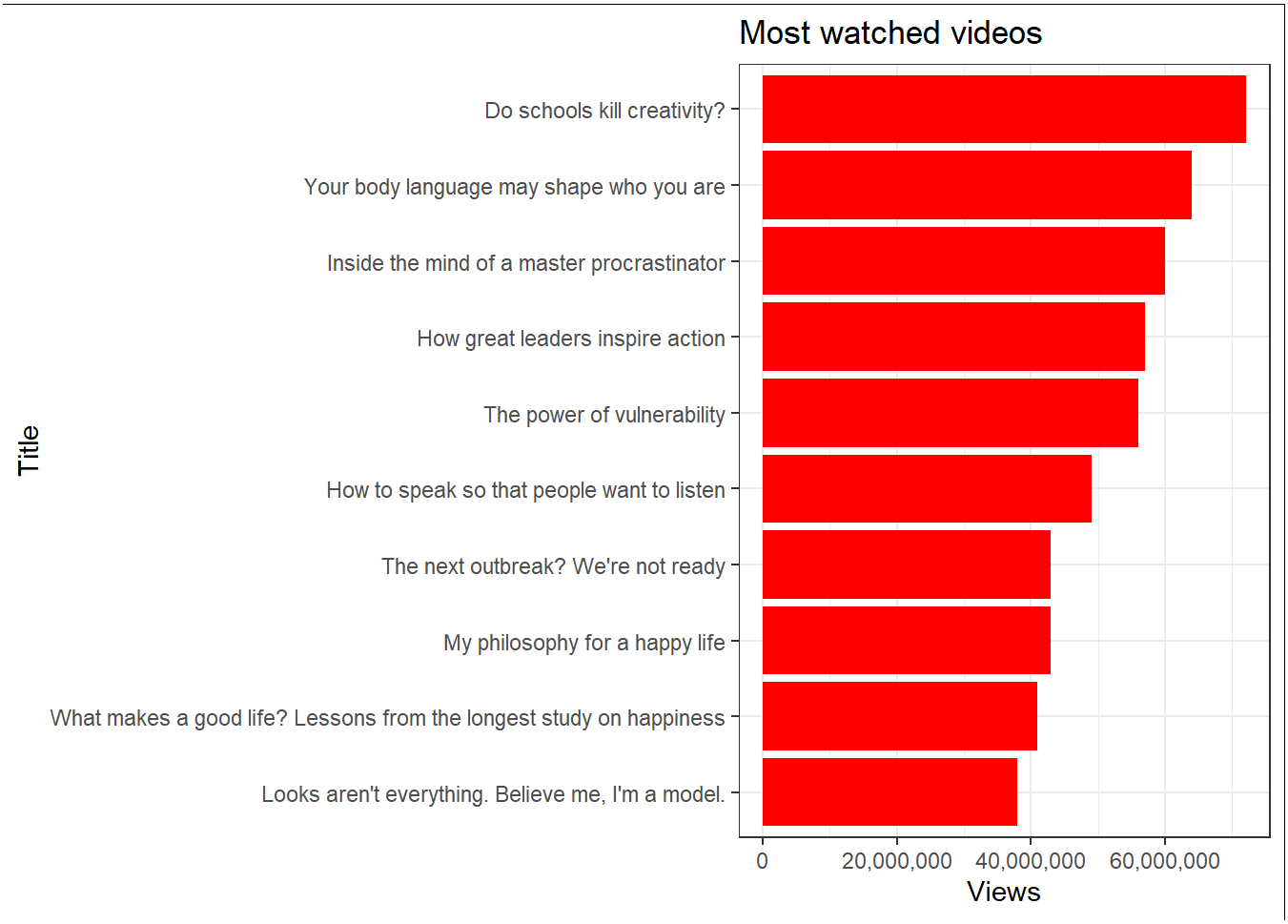
```
df$author[df$author==""]<-"Multiple"
df %>% filter(author=="Multiple")
##                title   author    month year views likes
## 1 Year In Ideas 2015 Multiple December 2015   532    15
```

Range of years-

```
range(df$year)
## [1] 1970 2022
```

Most popular programs-

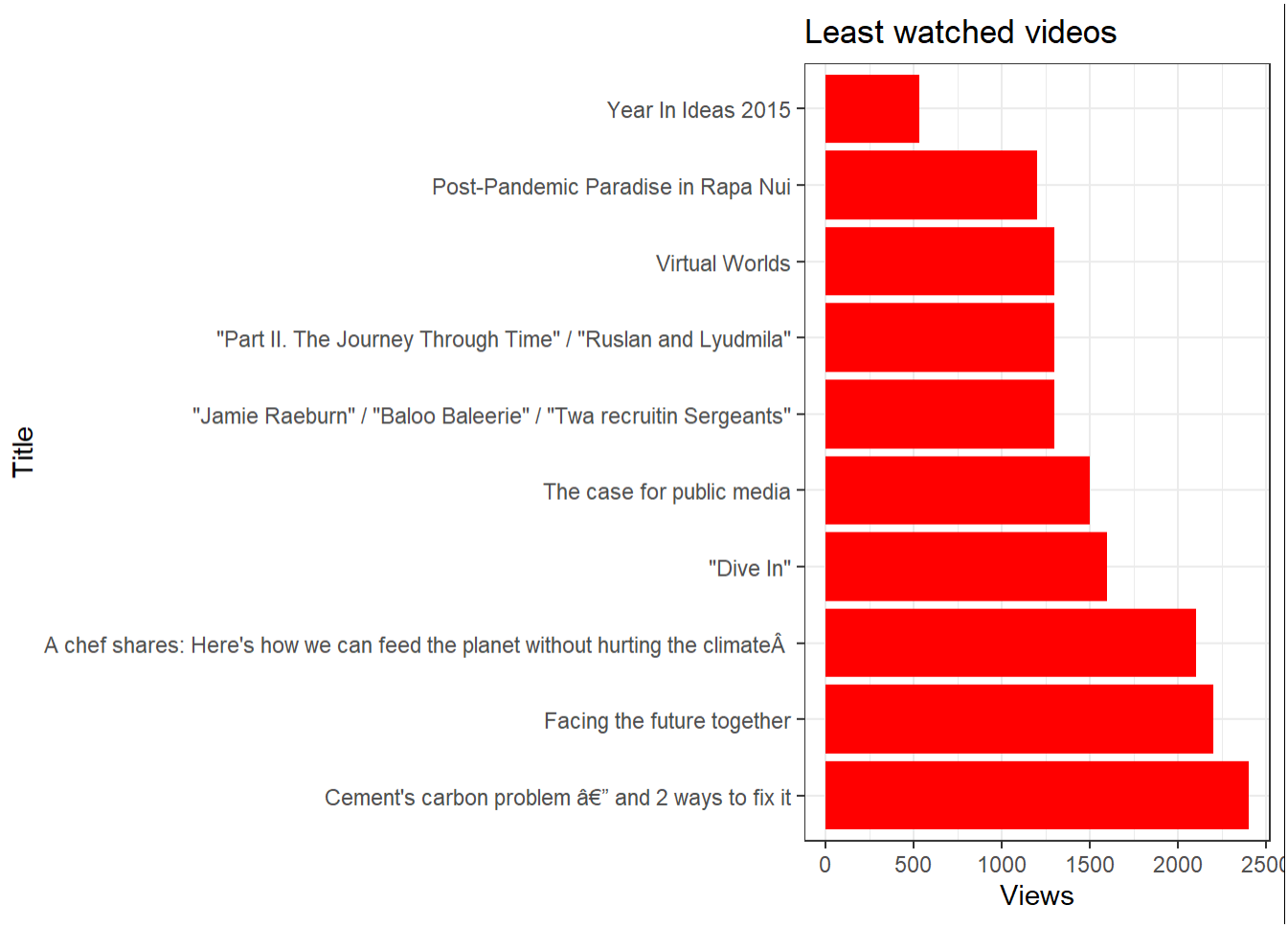
```
views_likes <-df %>% select(title,author,year,views) %>%
arrange(desc(views)) %>% head(10)
head(views_likes,10)
##                                                                  title
## 1                                          Do schools kill creativity?
## 2                            Your body language may shape who you are
## 3                           Inside the mind of a master procrastinator
## 4                                     How great leaders inspire action
## 5                                           The power of vulnerability
## 6                           How to speak so that people want to listen
## 7                                      My philosophy for a happy life
## 8                                The next outbreak? We're not ready
## 9  What makes a good life? Lessons from the longest study on happiness
## 10                  Looks aren't everything. Believe me, I'm a model.
##              author year    views
## 1  Sir Ken Robinson 2006 72000000
## 2         Amy Cuddy 2012 64000000
## 3         Tim Urban 2016 60000000
## 4       Simon Sinek 2009 57000000
## 5      BrenÃ© Brown 2010 56000000
## 6   Julian Treasure 2013 49000000
## 7         Sam Berns 2013 43000000
## 8        Bill Gates 2015 43000000
## 9  Robert Waldinger 2015 41000000
## 10  Cameron Russell 2012 38000000
views_likes %>% ggplot(aes(reorder(title,views),views,fill=I('red')))+
  geom_col(position = "dodge")+coord_flip()+theme_bw()+
  scale_y_continuous(labels = scales::comma)+
  labs(title = "Most watched videos",x="Title",y="Views")
```



Least watched videos

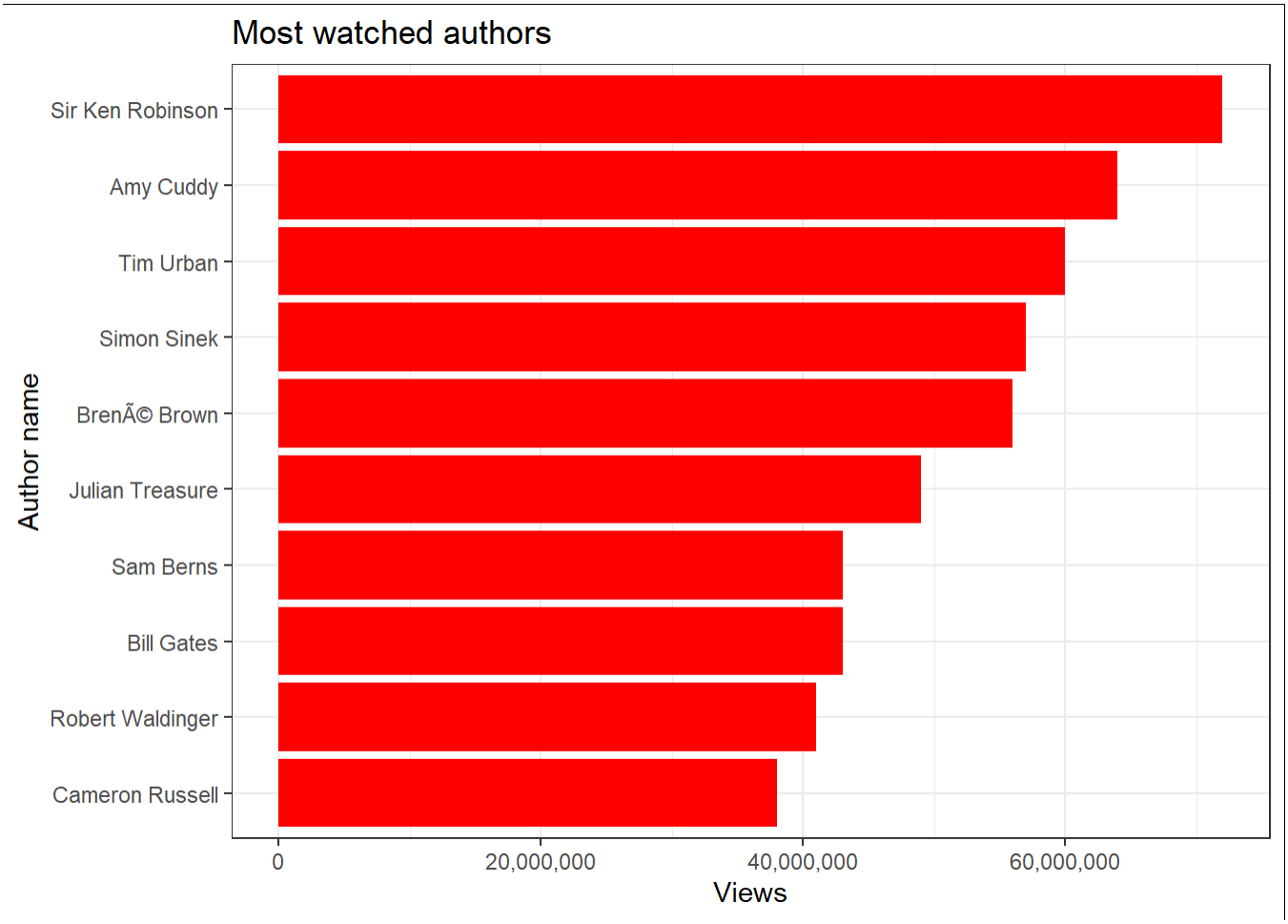
```
least_viewed <-df %>% select(title, author, views,year) %>% arrange(views)
%>% head(10)
head(least_viewed,10)
##
title
## 1                                                                  Year In
Ideas 2015
## 2                                                Post-Pandemic Paradise in
Rapa Nui
## 3                                                                     Virtual
Worlds
## 4                  "Jamie Raeburn" / "Baloo Baleerie" / "Twa recruitin
Sergeants"
## 5                     "Part II. The Journey Through Time" / "Ruslan and
Lyudmila"
## 6                                                    The case for
public media
## 7
"Dive In"
## 8  A chef shares: Here's how we can feed the planet without hurting the
climateÂ
```

```
## 9                                                        Facing the future
together
## 10                                  Cement's carbon problem â\200” and 2 ways
to fix it
##                                      author views year
## 1                                  Multiple   532 2015
## 2                                 Far Flung  1200 2020
## 3                                 Far Flung  1300 2020
## 4     Findlay Napier and Gillian Frame  1300 2019
## 5          Deutsche Philharmonie Merck  1300 2018
## 6                            Qi Wu å\220´ç\220ª  1500 2020
## 7                           Dave Matthews  1600 2021
## 8                              Peggy Chan  2100 2020
## 9  Rowan Fitzpatrick and Heart of Mind  2200 2017
## 10                         Mahendra Singhi  2400 2021
least_viewed %>% ggplot(aes(reorder(title,-views),views,fill=I("red")))+
  geom_col(position = "dodge")+
  labs(title = "Least watched videos",x="Title",y="Views")+
  coord_flip()+
  theme_bw()
```



Most watched authors

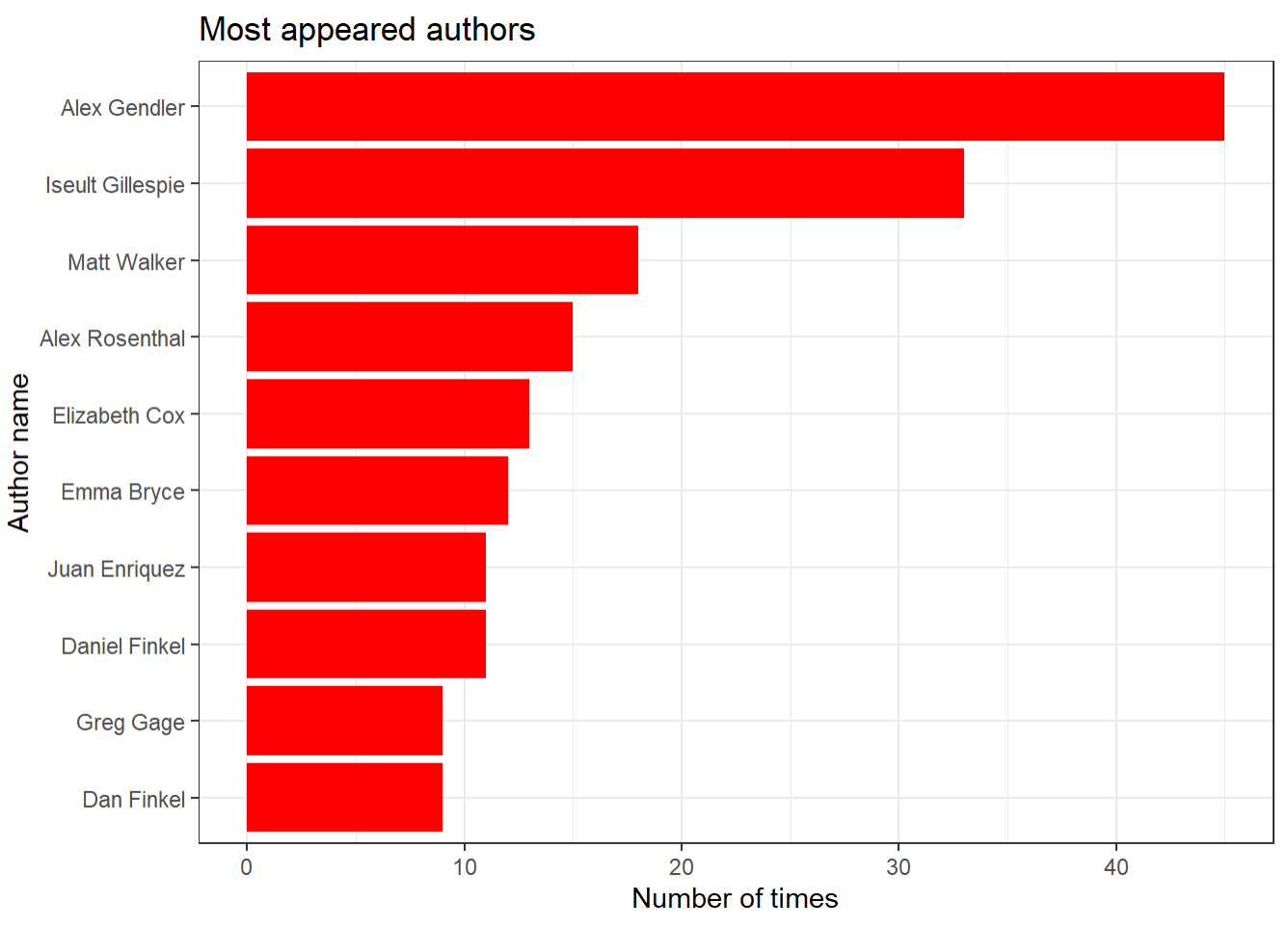
```
author_m<- df %>% select(author,views) %>%
  arrange(desc(views)) %>% head(10)
head(author_m,10)
##                author    views
## 1  Sir Ken Robinson 72000000
## 2          Amy Cuddy 64000000
## 3          Tim Urban 60000000
## 4        Simon Sinek 57000000
## 5       BrenÃ© Brown 56000000
## 6   Julian Treasure 49000000
## 7          Sam Berns 43000000
## 8         Bill Gates 43000000
## 9  Robert Waldinger 41000000
## 10  Cameron Russell 38000000
author_m %>% ggplot(aes(reorder(author,views),views,fill=I("red")))+
  geom_col(position = "dodge")+coord_flip()+
  scale_y_continuous(labels = scales::comma)+theme_bw()+
  labs(title = "Most watched authors",x="Author name",y="Views")
```



Most appeared authors-

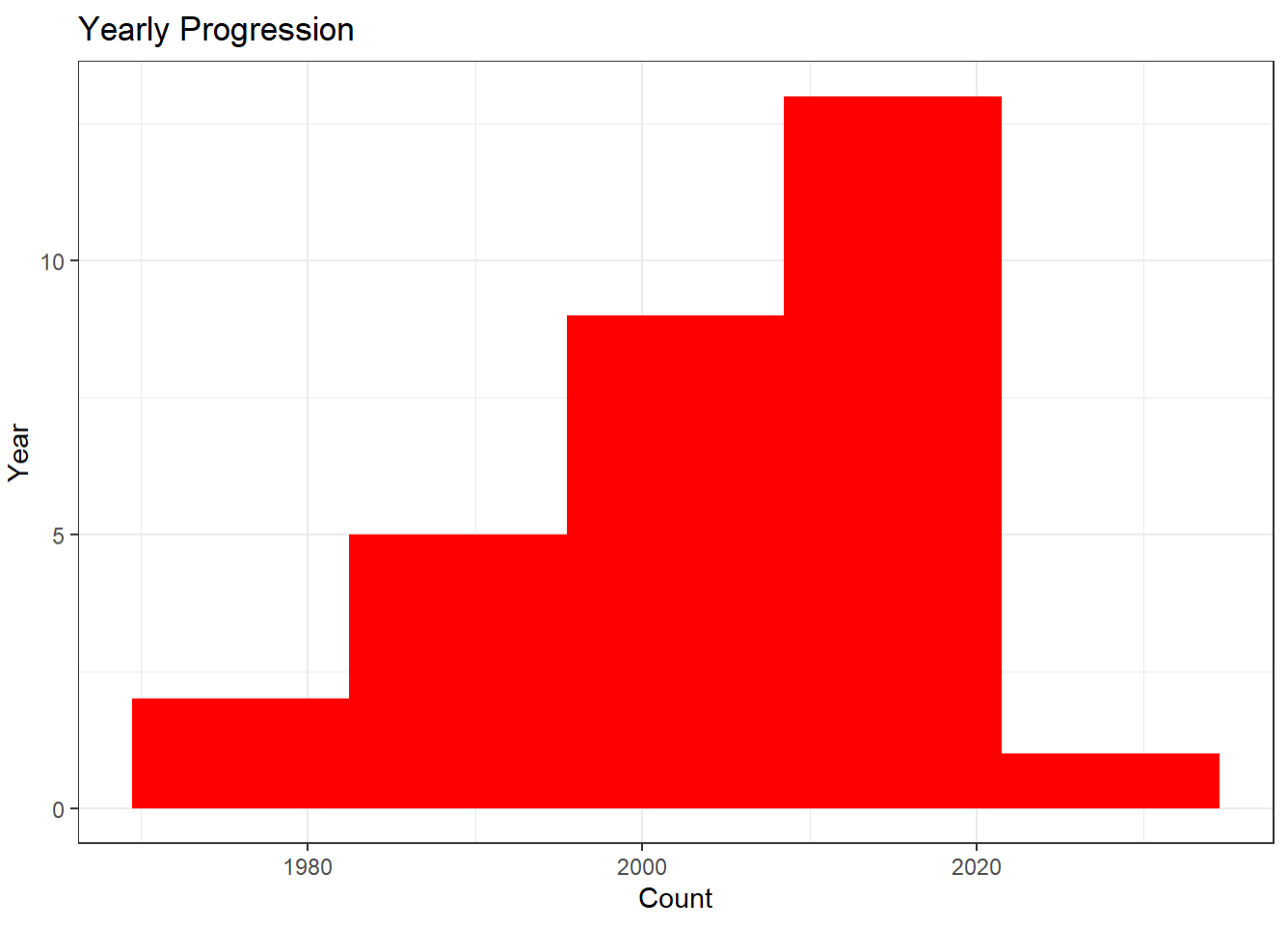
```
author_app <-df %>% group_by(author) %>%
```

```
  summarise(count=n()) %>% arrange(desc(count)) %>% head(10)
head(author_app,10)
## # A tibble: 10 x 2
##    author           count
##    <chr>            <int>
##  1 Alex Gendler        45
##  2 Iseult Gillespie    33
##  3 Matt Walker         18
##  4 Alex Rosenthal      15
##  5 Elizabeth Cox       13
##  6 Emma Bryce          12
##  7 Daniel Finkel       11
##  8 Juan Enriquez       11
##  9 Dan Finkel           9
## 10 Greg Gage            9
author_app %>%
ggplot(aes(reorder(author,count),count,fill=I('red')))+geom_col(position =
"dodge")+coord_flip()+
  labs(title = "Most appeared authors",x="Author name",y="Number of
times")+theme_bw()
```



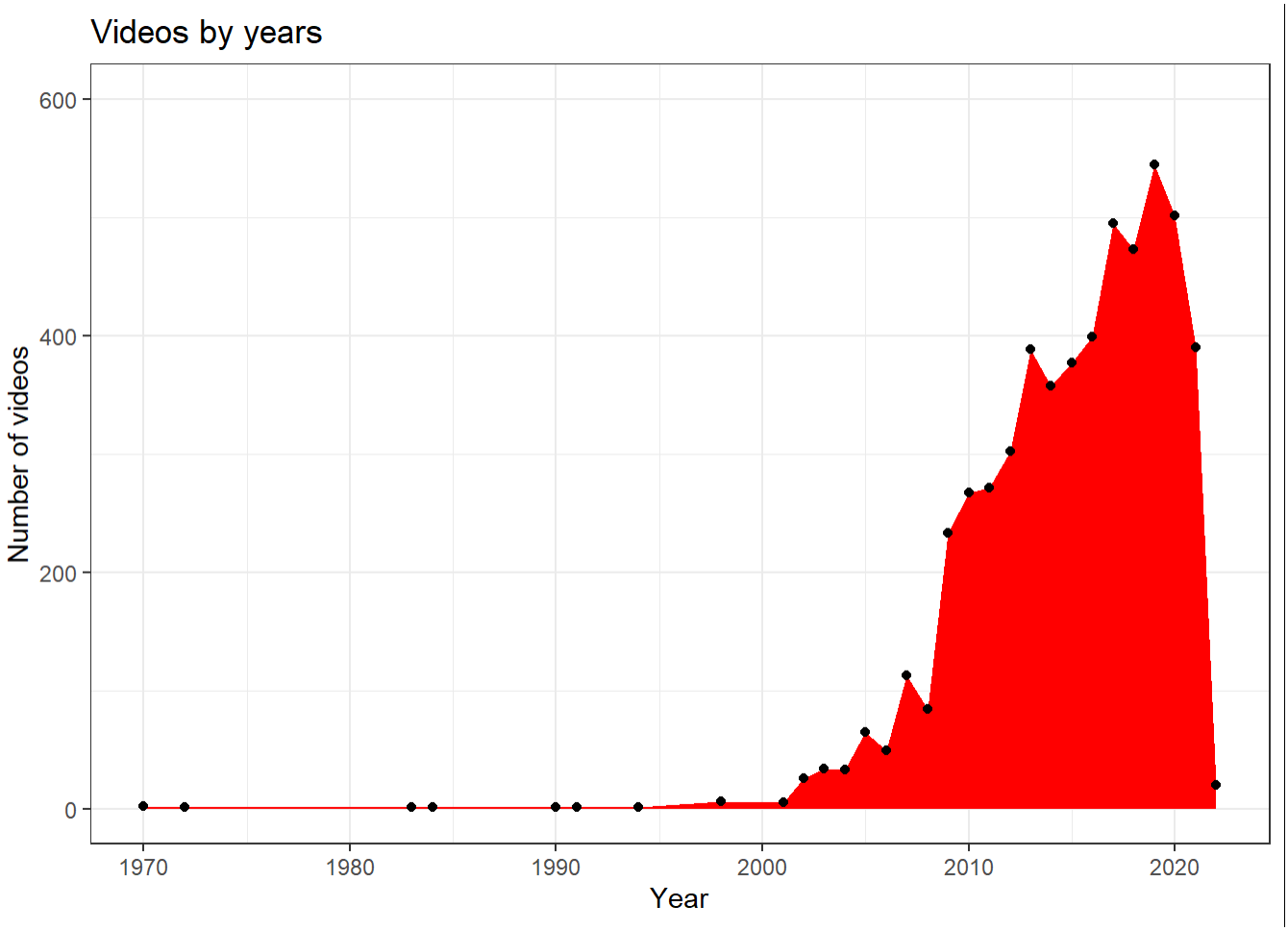
Year wise progression-

```
year_analysis <- df %>% group_by(year) %>%
  summarise(count=n()) %>% arrange(desc(count))
head(year_analysis,10)
## # A tibble: 10 x 2
##     year count
##    <int> <int>
##  1  2019   544
##  2  2020   501
##  3  2017   495
##  4  2018   473
##  5  2016   399
##  6  2021   390
##  7  2013   388
##  8  2015   377
##  9  2014   357
## 10  2012   302
year_analysis %>% ggplot(aes(year,fill=I('red')))+geom_histogram(bins
=5)+theme_bw()+
  labs(x="Count",y="Year",title = "Yearly Progression")
```



Year with most videos-

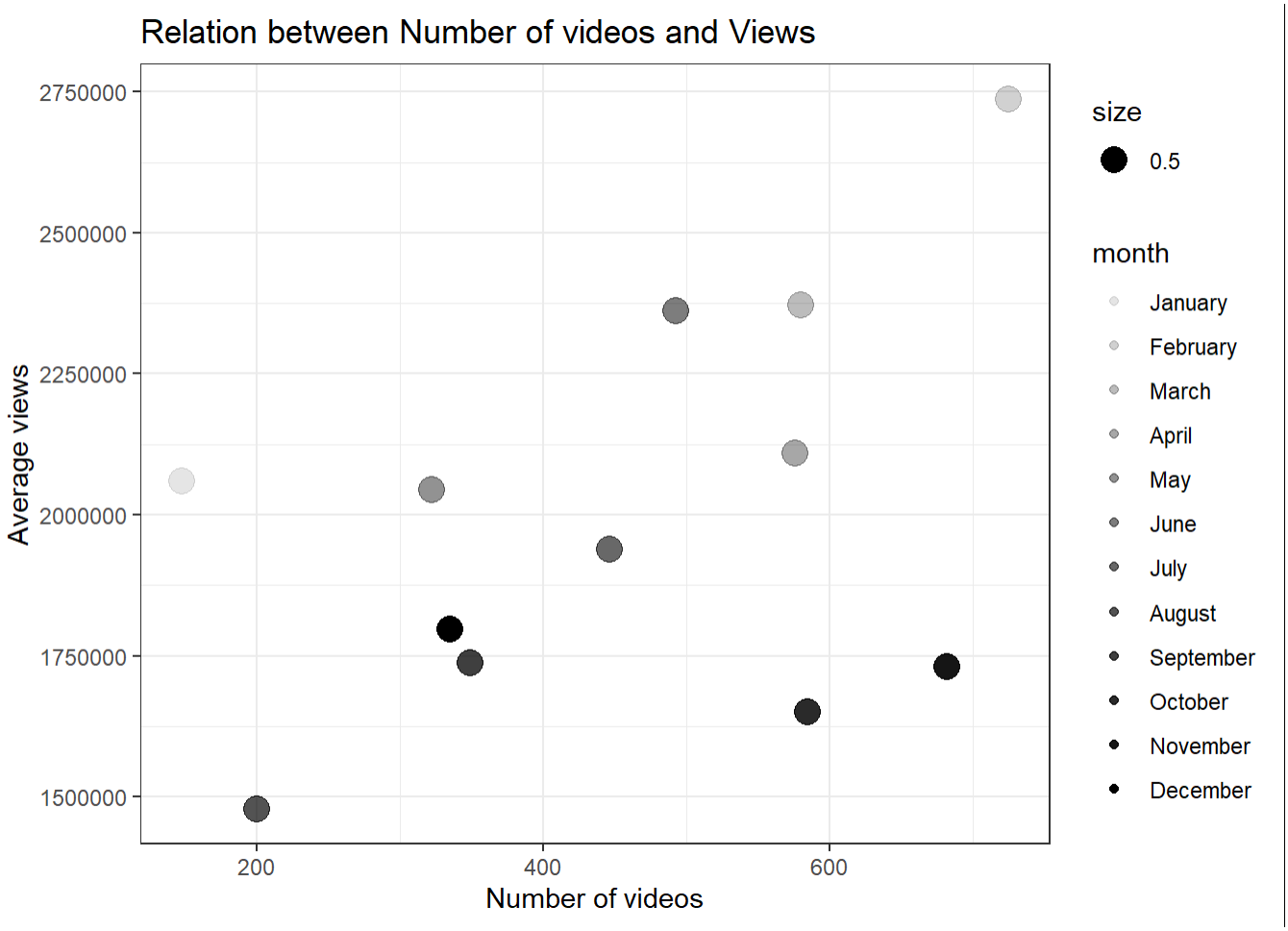
```
year_analysis %>%
ggplot(aes(year,count,fill=I("red")))+geom_area()+theme_bw()+
  geom_point()+labs(x="Year",y="Number of videos",title = "Videos by years")+
  ylim(0,600)
```



Monthly analysis

```
month_analysis <- df %>% group_by(month) %>%
  summarise(count=n(),
            views=sum(views))
month_analysis <- mutate(month_analysis,avg_views=views/count) %>%
arrange(desc(avg_views))
head(month_analysis,12)
## # A tibble: 12 x 4
##    month     count      views avg_views
##    <ord>     <int>      <int>     <dbl>
##  1 February    725 1984575996  2737346.
##  2 March       580 1375276896  2371167.
##  3 June        493 1164034699  2361125.
##  4 April       576 1214604196  2108688.
##  5 January     147  302850399  2060207.
##  6 May         322  658259600  2044284.
```

```
##  7 July         446  864732799  1938863.
##  8 December     335  602291832  1797886.
##  9 September    349  606392100  1737513.
## 10 November     682 1180802199  1731382.
## 11 October      585  965331498  1650139.
## 12 August       200  295820997  1479105.
month_analysis %>% ggplot(aes(count,avg_views,alpha=month,size=0.5))+
 geom_point()+
  labs(x="Number of videos",y="Average views",title = "Relation between
Number of videos and Views ")+
  theme_bw()+scale_x_continuous(labels = scales::comma)
```



Conclusion

One key difference between “Most watched” and “Least watched” is their release years- **Top ranked content falls in the year(2006-2016) While the least popular falls in the year(2015-2021)** which means with the time especially post-pandemic, the popularity of the show is falling compared to what it was in back years.

Before the year 2010, TED Talks had given around 923 videos, and from 2010 onwards they have already posted around 4517 videos which shows TED Talks became **proactive after the year 2010**, the reason could be due to the mass growth in popularity of Youtube after the year 2010.

We see that there is a direct relationship between average views and the number of videos with some exceptional cases such as October, November, and January, however, January shows a good sign that irrespective of least videos among all, numbers are much better.