Visualizing my video game dataset

2022-09-29

loading packages

```
library(knitr)
library(readr)
library(rmarkdown)
library(skimr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v dplyr   1.0.10
## v tibble  3.1.8      v stringr 1.4.1
## v tidyr   1.2.1      v forcats 0.5.2
## v purrr   0.3.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

importing csv through readr

```
library(readr)
dataset <- read_csv("C:\\Users\\brawn\\Documents\\repositories\\work\\GAME_DATA.csv")
```

```
## Rows: 157 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): title, date_released, developer, publisher, platform
## dbl (1): file_size_gb
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

data preview, summary

```
summary(dataset)
```

```
##     title            file_size_gb    date_released       developer
##  Length:157         Min.   :  0.00   Length:157         Length:157
##  Class :character   1st Qu.:  1.48   Class :character   Class :character
##  Mode  :character   Median :  7.20   Mode  :character   Mode  :character
##                     Mean   : 18.14
##                     3rd Qu.: 19.82
##                     Max.   :225.00
##   publisher           platform
##  Length:157         Length:157
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

```
head(dataset)
```

```
## # A tibble: 6 x 6
##   title                         file_size_gb date_relea~1 devel~2 publi~3 platf~4
##   <chr>                                <dbl> <chr>        <chr>   <chr>   <chr>
## 1 Call of Duty: Modern Warfare         225   10/25/2019   Infini~ Activi~ Battle~
## 2 ARK Survival Evolved                 163.  8/29/2017    Studio~ Studio~ Steam
## 3 Battlefield V                        108   11/12/2018   Dice    EA      Playst~
## 4 Grand Theft Auto V                   105.  4/14/2015    Rockst~ Rockst~ Steam
## 5 Call of Duty: Black Ops 4            100   10/12/2018   Treyar~ Activi~ Battle~
## 6 Call of Duty: Black Ops 3             91.8 11/6/2015    Treyar~ Activi~ Steam
## # ... with abbreviated variable names 1: date_released, 2: developer,
## #   3: publisher, 4: platform
```

```
skim_without_charts(dataset)
```

Table 1: Data summary

Name	dataset
Number of rows	157
Number of columns	6
Column type frequency:	
character	5
numeric	1
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
title	0	1.00	4	42	0	154	0
date_released	0	1.00	8	10	0	153	0
developer	1	0.99	2	28	0	98	0
publisher	1	0.99	2	28	0	82	0

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
platform	0	1.00	5	25	0	6	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
file_size_gb	0	1	18.14	30.72	0	1.48	7.2	19.82	225

adding column

```
mutate(dataset, file_size_gb_2=file_size_gb*100)
```

```
## # A tibble: 157 x 7
##    title                        file_s~1 date_~2 devel~3 publi~4 platf~5 file_~6
##    <chr>                           <dbl> <chr>   <chr>   <chr>   <chr>     <dbl>
##  1 Call of Duty: Modern Warfare    225   10/25/~ Infini~ Activi~ Battle~   22500
##  2 ARK Survival Evolved            163.  8/29/2~ Studio~ Studio~ Steam     16312
##  3 Battlefield V                   108   11/12/~ Dice    EA      Playst~   10800
##  4 Grand Theft Auto V              105.  4/14/2~ Rockst~ Rockst~ Steam     10516
##  5 Call of Duty: Black Ops 4       100   10/12/~ Treyar~ Activi~ Battle~   10000
##  6 Call of Duty: Black Ops 3        91.8 11/6/2~ Treyar~ Activi~ Steam      9183
##  7 Call of Duty: WWII               90.8 9/3/20~ Sledge~ Activi~ Steam      9082
##  8 Killing Floor 2                  82.1 11/18/~ Tripwi~ Tripwi~ Steam      8209
##  9 Payday 2                         80.9 8/13/2~ Overki~ Starbr~ Steam      8092
## 10 Borderlands 3                    75   9/13/2~ Gearbo~ 2K      Epic G~    7500
## # ... with 147 more rows, and abbreviated variable names 1: file_size_gb,
## #   2: date_released, 3: developer, 4: publisher, 5: platform,
## #   6: file_size_gb_2
```

looking at specific column

```
dataset %>%
select(developer)
```

```
## # A tibble: 157 x 1
##    developer
##    <chr>
##  1 Infinity Ward
##  2 Studio Wildcard
##  3 Dice
##  4 Rockstar
##  5 Treyarch
##  6 Treyarch
##  7 Sledgehammer Games
##  8 Tripwire Interactive
```

```
##  9 Overkill Software
## 10 Gearbox Software
## # ... with 147 more rows
```

looking at every column but one named in function

```
dataset %>%
  select(-developer)
```

```
## # A tibble: 157 x 5
##    title                          file_size_gb date_released publisher     platf~1
##    <chr>                                 <dbl> <chr>         <chr>         <chr>
##  1 Call of Duty: Modern Warfare        225   10/25/2019    Activision    Battle~
##  2 ARK Survival Evolved                163.  8/29/2017     Studio Wildc~ Steam
##  3 Battlefield V                       108   11/12/2018    EA            Playst~
##  4 Grand Theft Auto V                  105.  4/14/2015     Rockstar      Steam
##  5 Call of Duty: Black Ops 4           100   10/12/2018    Activision    Battle~
##  6 Call of Duty: Black Ops 3            91.8 11/6/2015     Activision    Steam
##  7 Call of Duty: WWII                   90.8 9/3/2017      Activision    Steam
##  8 Killing Floor 2                      82.1 11/18/2016    Tripwire Int~ Steam
##  9 Payday 2                             80.9 8/13/2013     Starbreeze P~ Steam
## 10 Borderlands 3                        75   9/13/2019     2K            Epic G~
## # ... with 147 more rows, and abbreviated variable name 1: platform
```

renaming column

```
dataset %>%
  rename(file_size_gigabytes=file_size_gb)
```

```
## # A tibble: 157 x 6
##    title                          file_size_giga~1 date_~2 devel~3 publi~4 platf~5
##    <chr>                                     <dbl> <chr>   <chr>   <chr>   <chr>
##  1 Call of Duty: Modern Warfare            225   10/25/~ Infini~ Activi~ Battle~
##  2 ARK Survival Evolved                    163.  8/29/2~ Studio~ Studio~ Steam
##  3 Battlefield V                           108   11/12/~ Dice    EA      Playst~
##  4 Grand Theft Auto V                      105.  4/14/2~ Rockst~ Rockst~ Steam
##  5 Call of Duty: Black Ops 4               100   10/12/~ Treyar~ Activi~ Battle~
##  6 Call of Duty: Black Ops 3                91.8 11/6/2~ Treyar~ Activi~ Steam
##  7 Call of Duty: WWII                       90.8 9/3/20~ Sledge~ Activi~ Steam
##  8 Killing Floor 2                          82.1 11/18/~ Tripwi~ Tripwi~ Steam
##  9 Payday 2                                 80.9 8/13/2~ Overki~ Starbr~ Steam
## 10 Borderlands 3                            75   9/13/2~ Gearbo~ 2K      Epic G~
## # ... with 147 more rows, and abbreviated variable names
## #   1: file_size_gigabytes, 2: date_released, 3: developer, 4: publisher,
## #   5: platform
```

sorting data

```
dataset %>%
  arrange(file_size_gb)
```

```
## # A tibble: 157 x 6
##    title                                file_~1 date_~2 devel~3 publi~4 platf~5
##    <chr>                                  <dbl> <chr>   <chr>   <chr>   <chr>
##  1 Legend of Zelda: Breath of the Wild      0    3/13/2~ Ninten~ Ninten~ Ninten~
##  2 Rayman Legends                           0    9/12/2~ Ubisoft Ubisoft Ninten~
##  3 Wolfenstein 3D                           0.01 8/3/19~ Id Sof~ Bethes~ Steam
##  4 Proteus                                  0.01 1/30/2~ Ed Key  Twiste~ Steam
##  5 Commander Keen                           0.03 12/14/~ Id Sof~ Id Sof~ Steam
##  6 Mount Your Friends                       0.04 7/29/2~ Steger~ Steger~ Steam
##  7 Timberman                                0.07 9/18/2~ Digita~ Foreve~ Steam
##  8 Super Mario 3D World and Bowsers Fury    0.08 2/12/2~ Ninten~ Ninten~ Ninten~
##  9 Risk of Rain                             0.09 11/8/2~ Hopoo ~ Chuckl~ Steam
## 10 SteamWorld Dig                           0.09 12/5/2~ Image ~ Image ~ Steam
## # ... with 147 more rows, and abbreviated variable names 1: file_size_gb,
## #   2: date_released, 3: developer, 4: publisher, 5: platform
```

```
dataset %>%
  arrange(-file_size_gb)
```

```
## # A tibble: 157 x 6
##    title                        file_size_gb date_rele~1 devel~2 publi~3 platf~4
##    <chr>                               <dbl> <chr>       <chr>   <chr>   <chr>
##  1 Call of Duty: Modern Warfare          225   10/25/2019  Infini~ Activi~ Battle~
##  2 ARK Survival Evolved                  163.  8/29/2017   Studio~ Studio~ Steam
##  3 Battlefield V                         108   11/12/2018  Dice    EA      Playst~
##  4 Grand Theft Auto V                    105.  4/14/2015   Rockst~ Rockst~ Steam
##  5 Call of Duty: Black Ops 4             100   10/12/2018  Treyar~ Activi~ Battle~
##  6 Call of Duty: Black Ops 3              91.8 11/6/2015   Treyar~ Activi~ Steam
##  7 Call of Duty: WWII                     90.8 9/3/2017    Sledge~ Activi~ Steam
##  8 Killing Floor 2                        82.1 11/18/2016  Tripwi~ Tripwi~ Steam
##  9 Payday 2                               80.9 8/13/2013   Overki~ Starbr~ Steam
## 10 Borderlands 3                          75   9/13/2019   Gearbo~ 2K      Epic G~
## # ... with 147 more rows, and abbreviated variable names 1: date_released,
## #   2: developer, 3: publisher, 4: platform
```

new data frame

```
dataset_2 <- dataset %>% arrange(-file_size_gb)

dataset_3 <- dataset %>% mutate(file_size_gb_2=file_size_gb)
```

average file size of on each platform

```
dataset %>%
  group_by(platform) %>%
    summarize(mean_file_size = mean(file_size_gb))
```

```
## # A tibble: 6 x 2
##   platform                  mean_file_size
##   <chr>                              <dbl>
## 1 BattleNet                          162.
## 2 Epic Games                          16.9
## 3 Nintendo Switch Cartridge            0.936
## 4 Nintendo Switch Online               7
## 5 Playstation Network                 39.4
## 6 Steam                               15.6
```

max file size of on each platform

```
dataset %>%
  group_by(platform) %>%
  summarize(mean_file_size = max(file_size_gb))
```

```
## # A tibble: 6 x 2
##   platform                  mean_file_size
##   <chr>                              <dbl>
## 1 BattleNet                          225
## 2 Epic Games                          75
## 3 Nintendo Switch Cartridge            4
## 4 Nintendo Switch Online               8.4
## 5 Playstation Network                108
## 6 Steam                              163.
```

filter

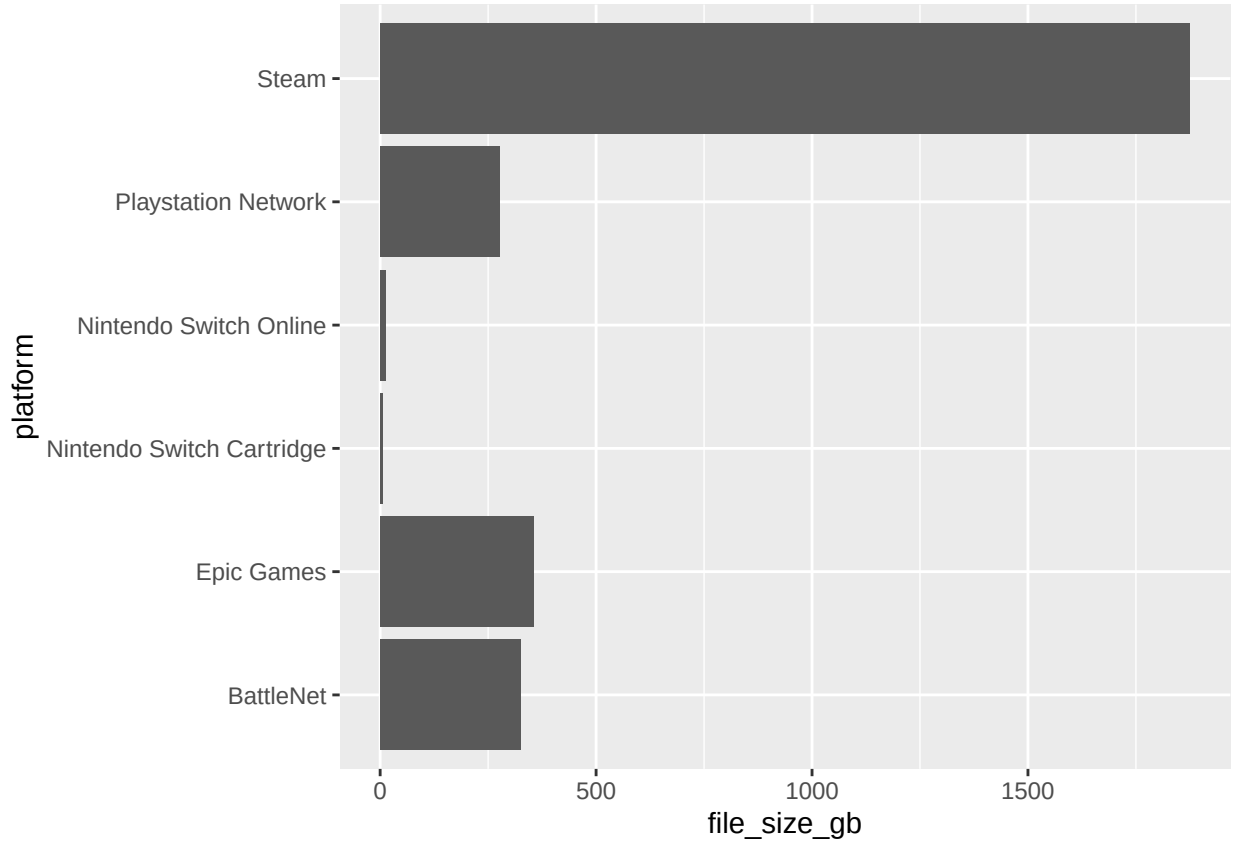
```
dataset %>%
  filter(file_size_gb < 1) %>%
    arrange(-file_size_gb)
```

```
## # A tibble: 35 x 6
##    title                        file_size_gb date_rele~1 devel~2 publi~3 platf~4
##    <chr>                               <dbl> <chr>       <chr>   <chr>   <chr>
##  1 Damned                               0.89 10/6/2014   9heads~ 9heads~ Steam
##  2 Quake III: Team Arena                0.83 12/19/2000  Id Sof~ Id Sof~ Steam
##  3 Dead Cells                           0.8  10/22/2020  Motion~ Motion~ Epic G~
##  4 Return to Castle Wolfenstein         0.77 11/20/2001  Gray M~ Bethes~ Steam
##  5 Velvet Sundown                       0.75 7/22/2014   Tribe ~ Tribe ~ Steam
```

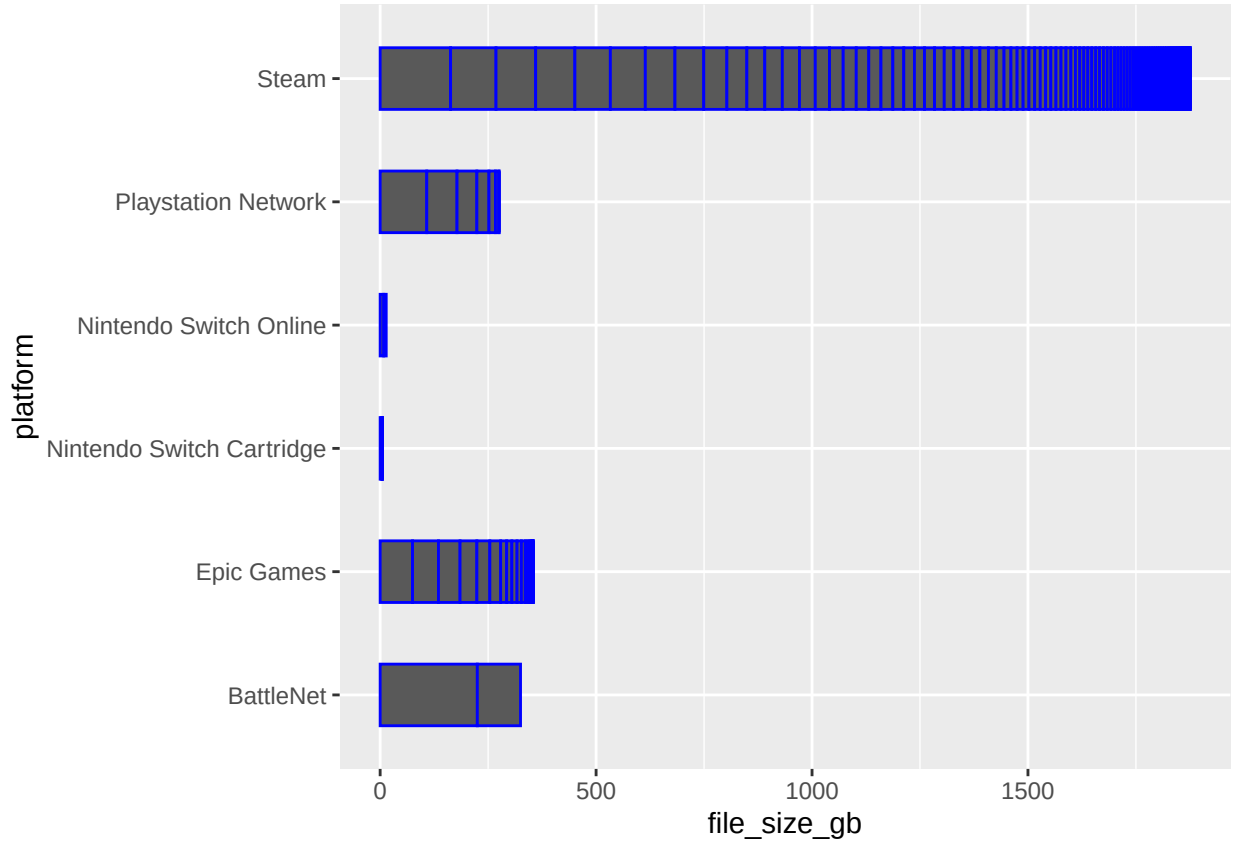
```
##  6 Sunless Sea                          0.7  2/25/2021   Failbe~ Failbe~ Epic G~
##  7 Arma 2: DayZ Mod                     0.65 2/21/2013   <NA>    <NA>    Steam
##  8 Stardew Valley                       0.62 2/26/2016   Concer~ Concer~ Steam
##  9 Final Fantasy III                    0.6  4/27/1990   Square~ Square~ Steam
## 10 Nintendo Switch Sports               0.6  4/29/2022   Ninten~ Ninten~ Ninten~
## # ... with 25 more rows, and abbreviated variable names 1: date_released,
## #   2: developer, 3: publisher, 4: platform
```

simple bar chart, with ggplot

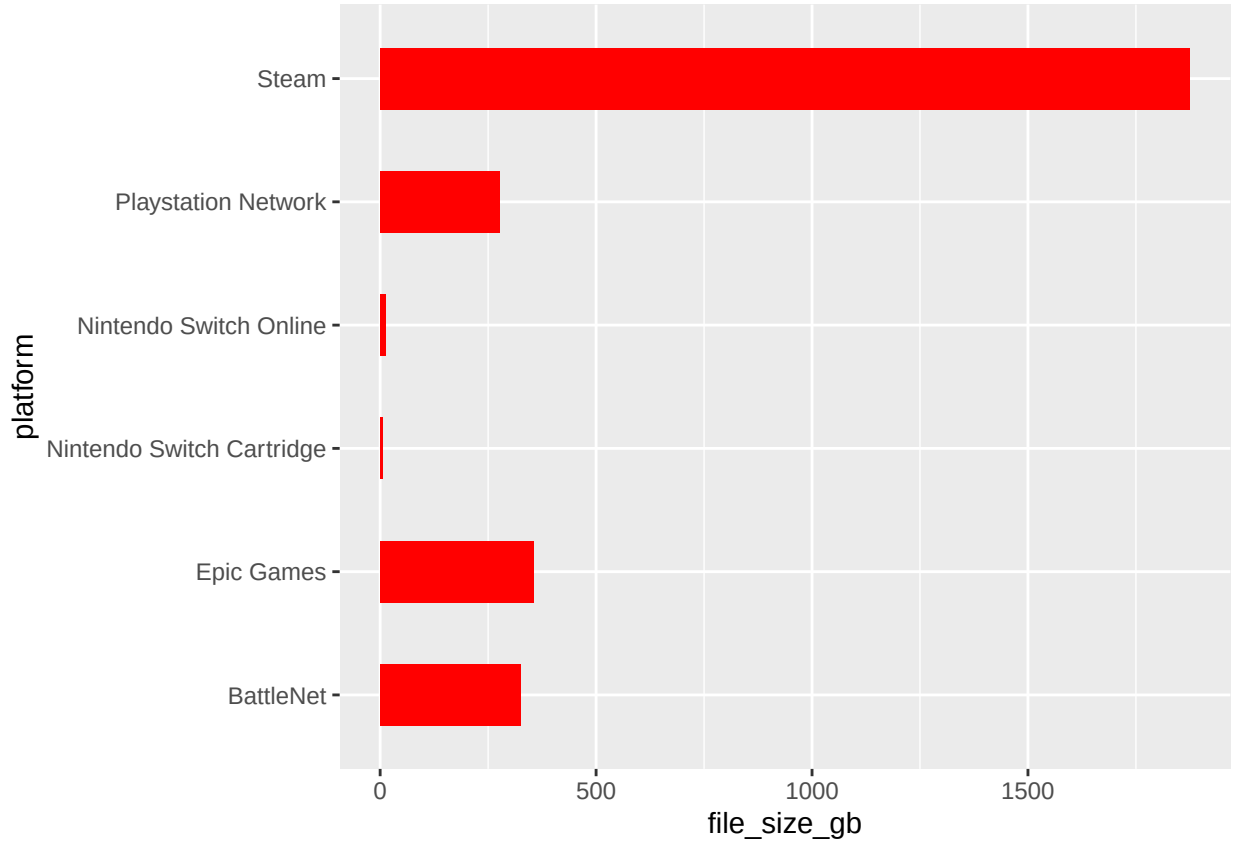
```
ggplot(data=dataset, aes(x=platform, y=file_size_gb)) +
  geom_bar(stat="identity") +
    coord_flip()
```



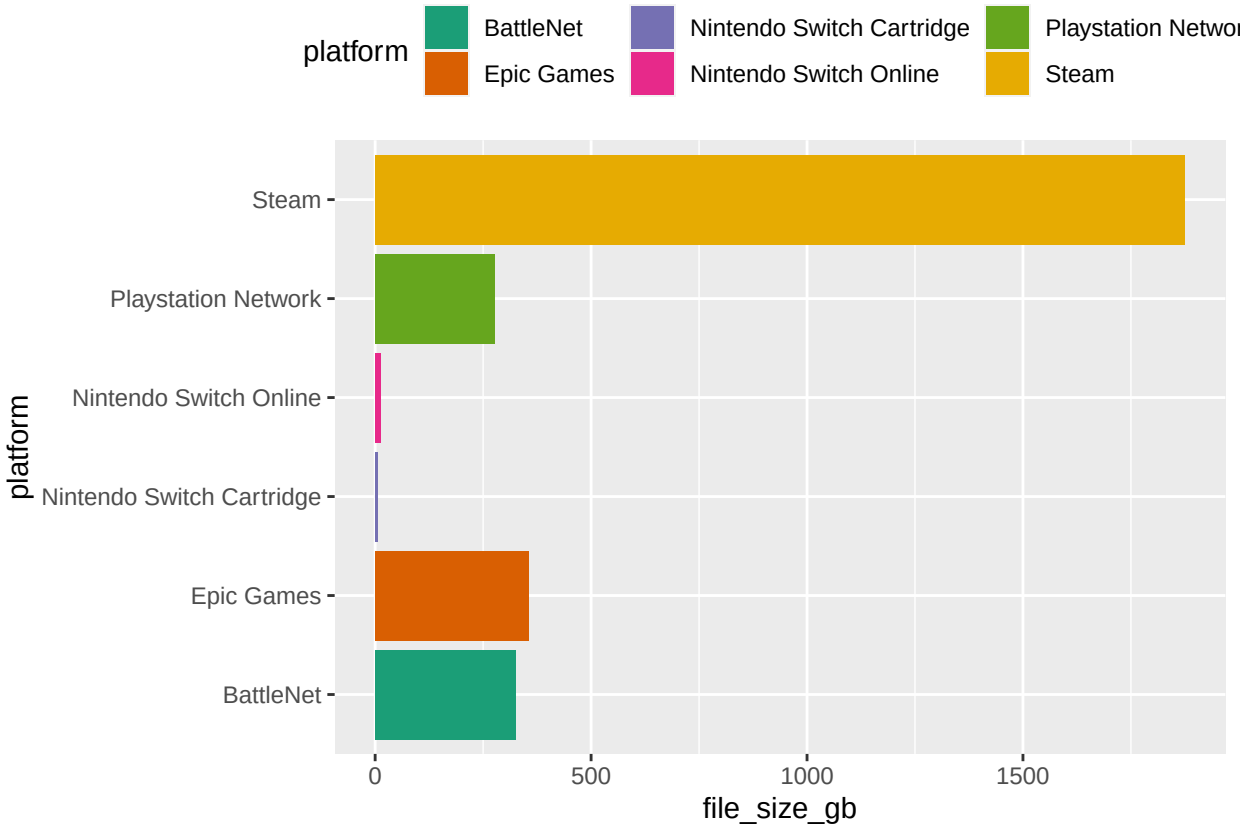

```
ggplot(data=dataset, aes(x=platform, y=file_size_gb)) +
  geom_bar(stat="identity", color="blue", width=.5) +
   coord_flip()
```




```
ggplot(data=dataset, aes(x=platform, y=file_size_gb)) +
  geom_bar(stat="identity", fill="red", width=.5) +
    coord_flip()
```




```
ggplot(data=dataset, aes(x=platform, y=file_size_gb,fill=platform))+
  geom_bar(stat="identity", width = .9) +
    scale_fill_brewer(palette="Dark2") + coord_flip() + theme(legend.position="top")
```


platform
BattleNet
Epic Games
Nintendo Switch Cartridge
Nintendo Switch Online
Playstation Networ
Steam
platform
Steam
Playstation Network
Nintendo Switch Online
Nintendo Switch Cartridge
Epic Games
BattleNet
0
500
1000
1500
file_size_gb