Game of Thrones kaggle

2022-11-17

import library

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(janitor)
```

```
## Warning: package 'janitor' was built under R version 4.2.2
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

view data

```
  Game_of_Thrones <- read_csv("Game_of_Thrones.csv")
```

```
## Rows: 73 Columns: 19
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (11): Title of the Episode, Directed by, Written by, Original Air Date, ...
## dbl  (8): Season, No. of Episode (Season), No. of Episode (Overall), Running...
##
## 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.
```

```
colnames(Game_of_Thrones)
```

```
##  [1] "Season"                                "No. of Episode (Season)"
##  [3] "No. of Episode (Overall)"              "Title of the Episode"
##  [5] "Running Time (Minutes)"                "Directed by"
##  [7] "Written by"                            "Original Air Date"
##  [9] "U.S. Viewers (Millions)"               "Music by"
## [11] "Cinematography by"                     "Editing by"
## [13] "IMDb Rating"                           "Rotten Tomatoes Rating (Percentage)"
## [15] "Metacritic Ratings"                    "Ordered"
## [17] "Filming Duration"                      "Novel(s) Adapted"
## [19] "Synopsis"
```

```
glimpse(Game_of_Thrones)
```

```
## Rows: 73
## Columns: 19
## $ Season                                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2,~
## $ `No. of Episode (Season)`             <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1~
## $ `No. of Episode (Overall)`            <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1~
## $ `Title of the Episode`                <chr> "Winter Is Coming", "The Kingsro~
## $ `Running Time (Minutes)`              <dbl> 61, 55, 57, 55, 54, 52, 57, 58, ~
## $ `Directed by`                         <chr> "Tim Van Patten", "Tim Van Patte~
## $ `Written by`                          <chr> "David Benioff, D. B. Weiss", "D~
## $ `Original Air Date`                   <chr> "17-Apr-2011", "24-Apr-2011", "1~
## $ `U.S. Viewers (Millions)`             <dbl> 2.22, 2.20, 2.44, 2.45, 2.58, 2.~
## $ `Music by`                            <chr> "Ramin Djawadi", "Ramin Djawadi"~
## $ `Cinematography by`                   <chr> "Alik Sakharov", "Alik Sakharov"~
## $ `Editing by`                          <chr> "Oral Norrie Ottey", "Oral Norri~
## $ `IMDb Rating`                         <dbl> 8.9, 8.6, 8.5, 8.6, 9.0, 9.1, 9.~
## $ `Rotten Tomatoes Rating (Percentage)` <dbl> 100, 100, 81, 100, 95, 100, 100,~
## $ `Metacritic Ratings`                  <dbl> 9.1, 8.9, 8.7, 9.1, 9.0, 9.2, 9.~
## $ Ordered                               <chr> "March 2, 2010", "March 2, 2010"~
## $ `Filming Duration`                    <chr> "Second half of 2010", "Second h~
## $ `Novel(s) Adapted`                    <chr> "A Game of Thrones", "A Game of ~
## $ Synopsis                              <chr> "North of the Seven Kingdoms of ~
```

column name contain spaces use janitor for clean

```
Game_of_Thrones <- clean_names(Game_of_Thrones)
colnames(Game_of_Thrones)
```

```
##  [1] "season"                          "no_of_episode_season"
##  [3] "no_of_episode_overall"           "title_of_the_episode"
##  [5] "running_time_minutes"            "directed_by"
##  [7] "written_by"                      "original_air_date"
##  [9] "u_s_viewers_millions"            "music_by"
## [11] "cinematography_by"               "editing_by"
## [13] "im_db_rating"                    "rotten_tomatoes_rating_percentage"
## [15] "metacritic_ratings"              "ordered"
## [17] "filming_duration"                "novel_s_adapted"
## [19] "synopsis"
```

question

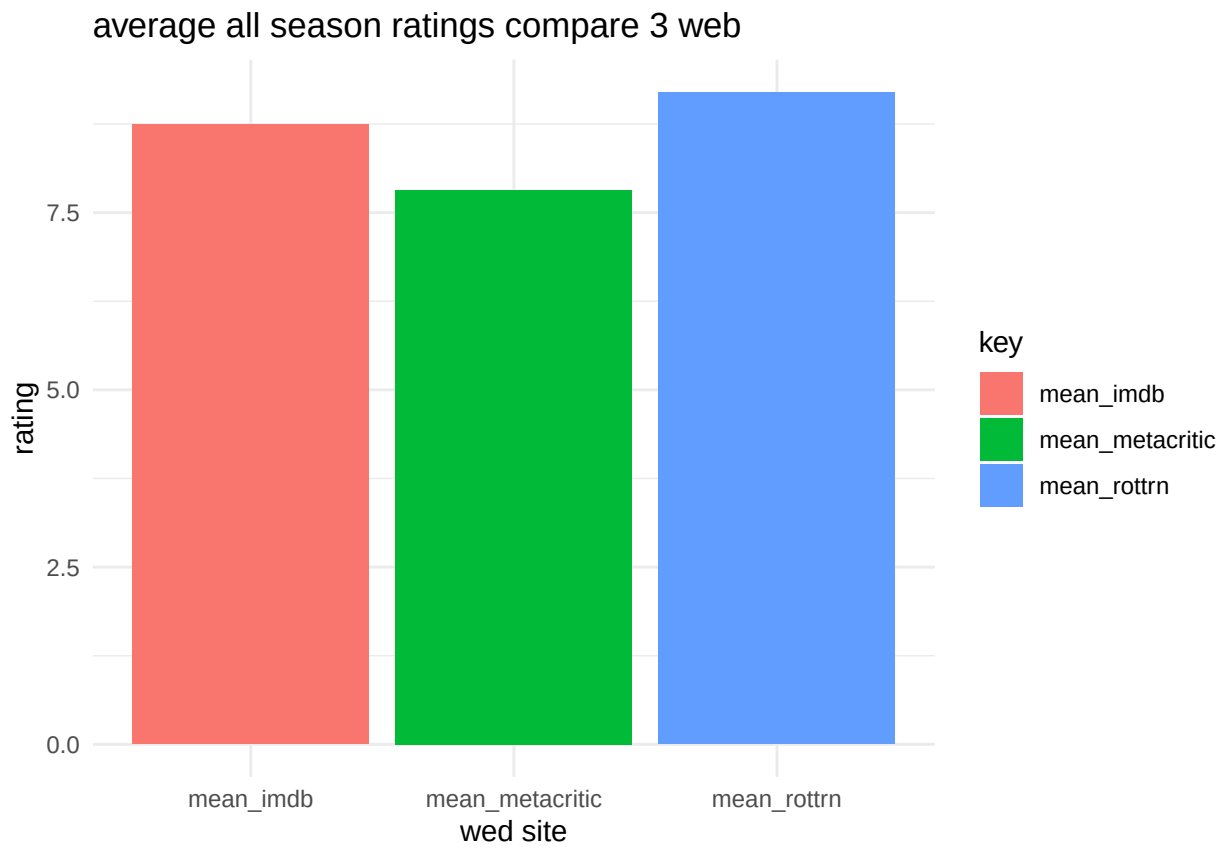
1.All average ratings compare 3 web

```
rating_of_tree <- Game_of_Thrones %>%
 select(im_db_rating,
        rotten_tomatoes_rating_percentage,
        metacritic_ratings) %>%
  summarise(mean_imdb = mean(im_db_rating),
            mean_rottrn = mean(rotten_tomatoes_rating_percentage)/10,
            mean_metacritic = mean(metacritic_ratings))
```

```
rating_of_tree
```

```
## # A tibble: 1 x 3
##   mean_imdb mean_rottrn mean_metacritic
##       <dbl>       <dbl>           <dbl>
## 1      8.74        9.20            7.82
```

```
rating_of_tree %>%
  gather() %>%
  ggplot(aes(key,value,fill=key)) +
  geom_col()+
  theme_minimal()+
  labs(title = "average all season ratings compare 3 web",
      x = "wed site",
      y= "rating")
```



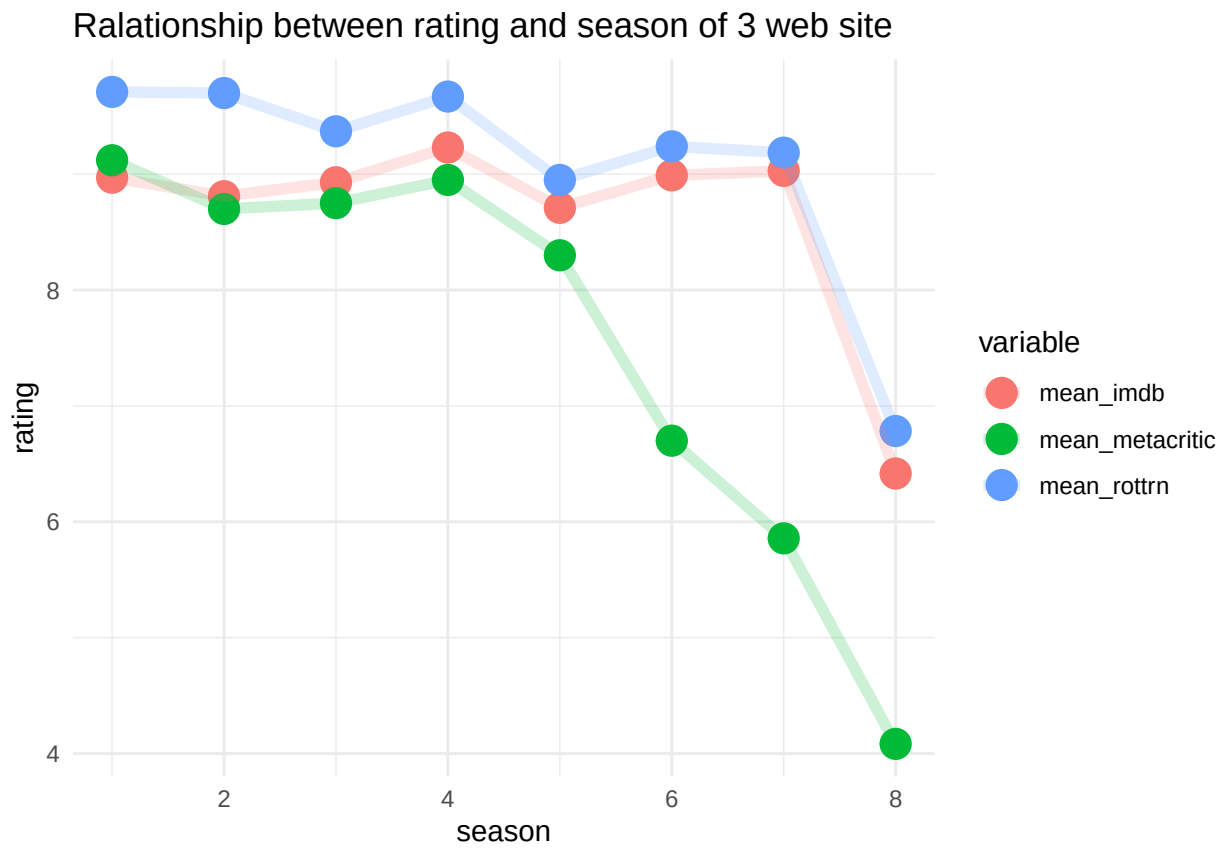
2.Average ratings by season compare 3 web

```
rating_of_season <- Game_of_Thrones %>%
  select(season,
         im_db_rating,
         rotten_tomatoes_rating_percentage,
         metacritic_ratings) %>%
  group_by(season) %>%
  summarise(mean_imdb = mean(im_db_rating),
            mean_rottrn = mean(rotten_tomatoes_rating_percentage)/10,
            mean_metacritic = mean(metacritic_ratings))
```

```
rating_of_season
```

```
## # A tibble: 8 x 4
##   season mean_imdb mean_rottrn mean_metacritic
##    <dbl>     <dbl>       <dbl>           <dbl>
## 1      1      8.97        9.71            9.12
## 2      2      8.81        9.7             8.7
## 3      3      8.93        9.37            8.75
## 4      4      9.23        9.67            8.95
## 5      5      8.71        8.95            8.3
## 6      6      8.99        9.24            6.7
## 7      7      9.03        9.19            5.86
## 8      8      6.42        6.78            4.08
```

```
rating_of_season %>%
  pivot_longer(-season,
               names_to = "variable",
               values_to = "value") %>%
  ggplot(aes(season, value, colour= variable))+
  geom_point(size=5)+
  geom_line(alpha=0.2,size=2)+
  theme_minimal()+
  labs(title = "Ralationship between rating and season of 3 web site",
      x = "season",
      y= "rating")
```



season 4 is best of rating

3.Average ratings of each episode by season4

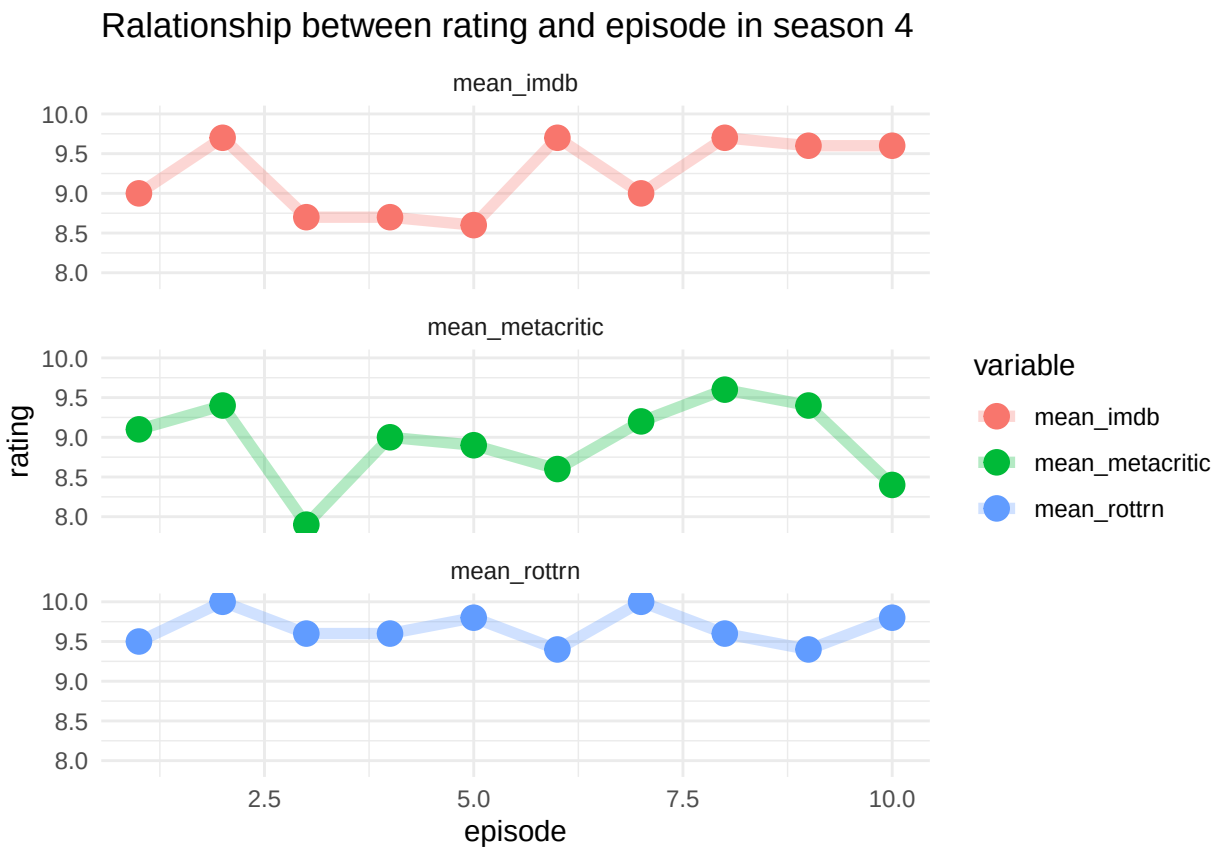
```
rating_of_season4 <- Game_of_Thrones %>%
  filter(season == 4) %>%
  select(no_of_episode_season,
         im_db_rating,
         rotten_tomatoes_rating_percentage,
         metacritic_ratings) %>%
  group_by(no_of_episode_season) %>%
```

```
  summarise(mean_imdb = mean(im_db_rating),
            mean_rottrn = mean(rotten_tomatoes_rating_percentage)/10,
            mean_metacritic = mean(metacritic_ratings))
```

```
rating_of_season4
```

```
## # A tibble: 10 x 4
##    no_of_episode_season mean_imdb mean_rottrn mean_metacritic
##                   <dbl>     <dbl>       <dbl>           <dbl>
##  1                    1       9           9.5             9.1
##  2                    2       9.7        10               9.4
##  3                    3       8.7         9.6             7.9
##  4                    4       8.7         9.6             9
##  5                    5       8.6         9.8             8.9
##  6                    6       9.7         9.4             8.6
##  7                    7       9          10               9.2
##  8                    8       9.7         9.6             9.6
##  9                    9       9.6         9.4             9.4
## 10                   10       9.6         9.8             8.4
```

```
rating_of_season4 %>%
  pivot_longer(-no_of_episode_season,
               names_to = "variable",
               values_to = "value") %>%
  ggplot(aes(no_of_episode_season,
             value,
             colour= variable))+
  geom_point(size=4)+
  geom_line(alpha=0.3,size=2)+
  theme_minimal()+
  labs(title = "Ralationship between rating and episode in season 4",
      x = "episode",
      y= "rating")+
  facet_wrap(~variable,ncol=1)
```



4.Top 5 directors who have worked with the most

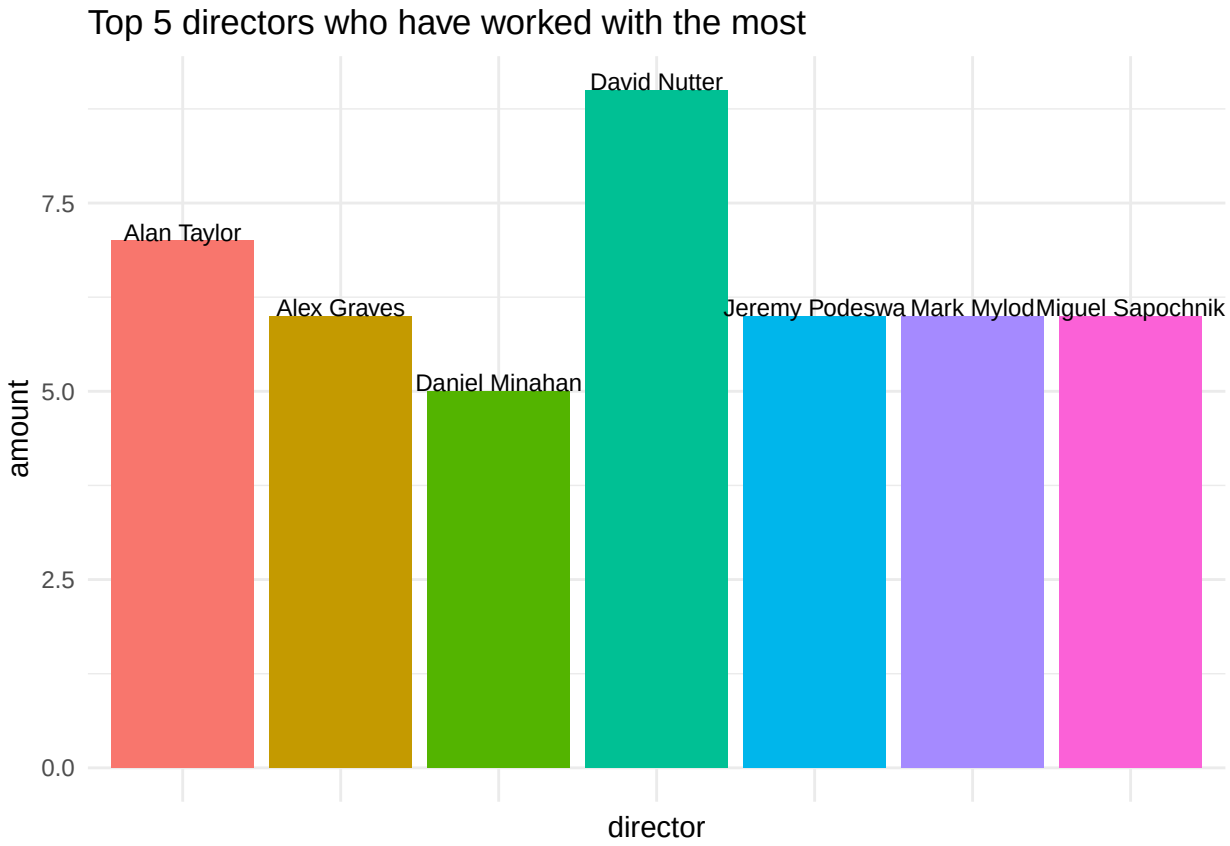
```
diredtors <- Game_of_Thrones %>%
  select(directed_by) %>%
  group_by(directed_by) %>%
  count() %>%
  arrange(desc(n)) %>%
  filter(n >= 5)
```

```
diredtors
```

```
## # A tibble: 7 x 2
## # Groups:   directed_by [7]
##   directed_by          n
##   <chr>            <int>
## 1 David Nutter         9
## 2 Alan Taylor          7
## 3 Alex Graves          6
## 4 Jeremy Podeswa       6
## 5 Mark Mylod           6
## 6 Miguel Sapochnik     6
## 7 Daniel Minahan       5
```

```
diredtors %>%
  ggplot(aes(directed_by,
             n,
```

```
            label = diredtors$directed_by,
            fill=directed_by))+
  geom_col()+
  geom_text(size=3,
            position = position_dodge(0.9),
            vjust=0)+
  theme_minimal()+
  labs(title = "Top 5 directors who have worked with the most",
      x = "director",
      y= "amount")+
  theme(axis.text.x = element_blank(),
        legend.position = "none")
```



David Nutter and Alan Taylor is worked with the most

5.Which episodes are the top 5 directors taking care of?

David Nutter

```
Game_of_Thrones %>%
  filter(directed_by == 'David Nutter') %>%
  select(1,2,4,5,6,7,8,9,11,12,13,14,15,19)
```

```
## # A tibble: 9 x 14
##   season no_of~1 title~2 runni~3 direc~4 writt~5 origi~6 u_s_v~7 cinem~8 editi~9
##    <dbl>   <dbl> <chr>     <dbl> <chr>   <chr>   <chr>     <dbl> <chr>   <chr>
## 1      2       6 The Ol~      54 David ~ Vaness~ 6-May-~    3.88 Martin~ Oral N~
## 2      2       7 A Man ~      55 David ~ David ~ 13-May~    3.69 Martin~ Oral N~
## 3      3       9 The Ra~      50 David ~ David ~ 2-Jun-~    5.22 Robert~ Oral N~
## 4      3      10 Mhysa        62 David ~ David ~ 9-Jun-~    5.39 Robert~ Oral N~
## 5      5       9 The Da~      52 David ~ David ~ 7-Jun-~    7.14 Robert~ Katie ~
## 6      5      10 Mother~      60 David ~ David ~ 14-Jun~    8.11 Robert~ Tim Po~
## 7      8       1 Winter~      53 David ~ Dave H~ 14-Apr~   11.8  David ~ Crispi~
## 8      8       2 A Knig~      57 David ~ Bryan ~ 21-Apr~   10.3  David ~ Crispi~
## 9      8       4 The La~      77 David ~ David ~ 5-May-~   11.8  David ~ Katie ~
## # ... with 4 more variables: im_db_rating <dbl>,
## #   rotten_tomatoes_rating_percentage <dbl>, metacritic_ratings <dbl>,
## #   synopsis <chr>, and abbreviated variable names 1: no_of_episode_season,
## #   2: title_of_the_episode, 3: running_time_minutes, 4: directed_by,
## #   5: written_by, 6: original_air_date, 7: u_s_viewers_millions,
## #   8: cinematography_by, 9: editing_by
```

Alan Taylor

```
Game_of_Thrones %>%
  filter(directed_by == 'Alan Taylor') %>%
  select(1,2,4,5,6,7,8,9,11,12,13,14,15,19)
```

```
## # A tibble: 7 x 14
##   season no_of~1 title~2 runni~3 direc~4 writt~5 origi~6 u_s_v~7 cinem~8 editi~9
##    <dbl>   <dbl> <chr>     <dbl> <chr>   <chr>   <chr>     <dbl> <chr>   <chr>
## 1      1       9 Baelor       56 Alan T~ David ~ 12-Jun~    2.66 Alik S~ France~
## 2      1      10 Fire a~      52 Alan T~ David ~ 19-Jun~    3.04 Alik S~ France~
## 3      2       1 The No~      52 Alan T~ David ~ 1-Apr-~    3.86 Kramer~ France~
## 4      2       2 The Ni~      53 Alan T~ David ~ 8-Apr-~    3.76 Kramer~ France~
## 5      2       8 The Pr~      53 Alan T~ David ~ 20-May~    3.86 Jonath~ France~
## 6      2      10 Valar ~      63 Alan T~ David ~ 3-Jun-~    4.2  Jonath~ France~
## 7      7       6 Beyond~      70 Alan T~ David ~ 20-Aug~   10.2  Jonath~ Tim Po~
## # ... with 4 more variables: im_db_rating <dbl>,
## #   rotten_tomatoes_rating_percentage <dbl>, metacritic_ratings <dbl>,
## #   synopsis <chr>, and abbreviated variable names 1: no_of_episode_season,
## #   2: title_of_the_episode, 3: running_time_minutes, 4: directed_by,
## #   5: written_by, 6: original_air_date, 7: u_s_viewers_millions,
## #   8: cinematography_by, 9: editing_by
```

6.Top 5 most viewed people

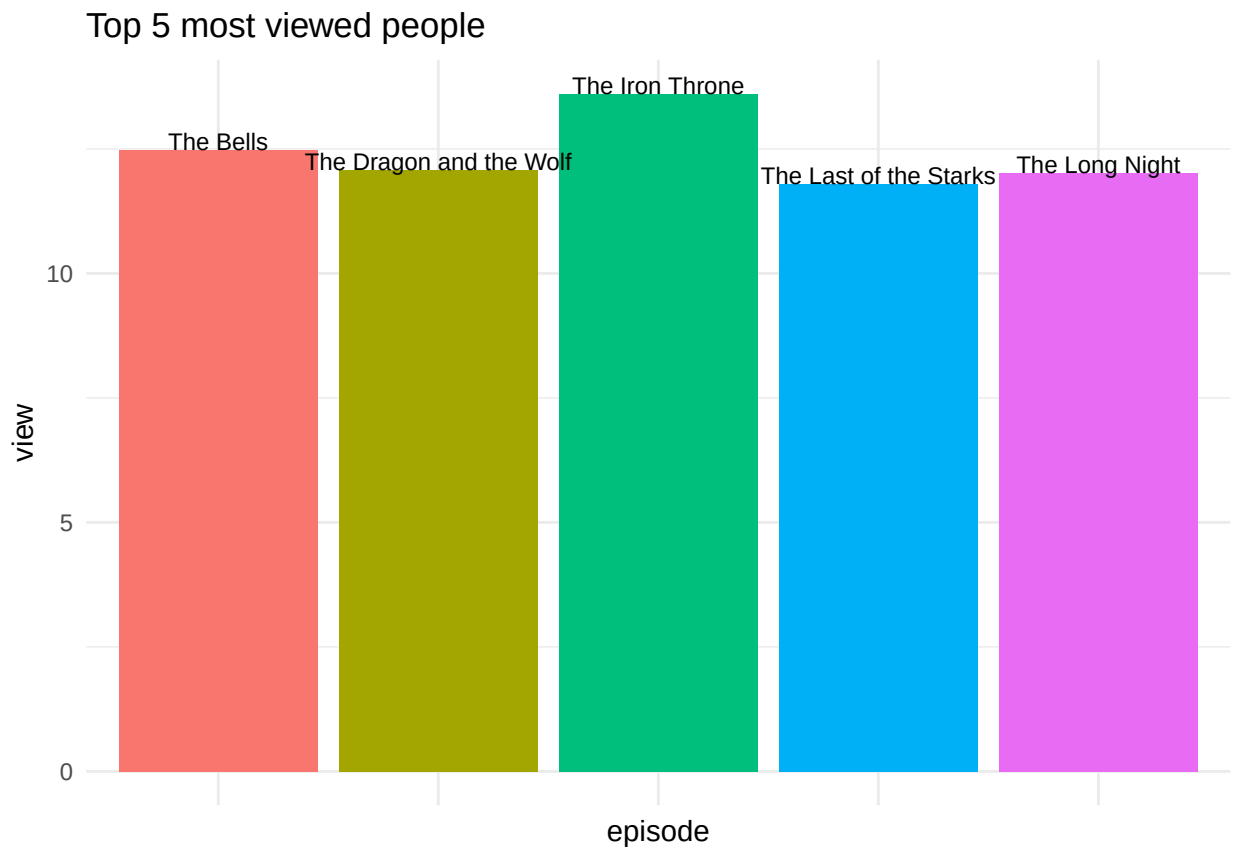
```
top_5_viewed <- Game_of_Thrones %>%
  select(1,2,4,9) %>%
  arrange(desc(u_s_viewers_millions)) %>%
  head(5)
```

```
top_5_viewed
```

```
## # A tibble: 5 x 4
##   season no_of_episode_season title_of_the_episode    u_s_viewers_millions
##    <dbl>                <dbl> <chr>                                 <dbl>
## 1      8                    6 The Iron Throne                        13.6
## 2      8                    5 The Bells                              12.5
## 3      7                    7 The Dragon and the Wolf                12.1
## 4      8                    3 The Long Night                         12.0
## 5      8                    4 The Last of the Starks                 11.8
```

season 8 ep6 and ep5 is popular

```
top_5_viewed %>%
  ggplot(aes(title_of_the_episode,
             u_s_viewers_millions,
             label = title_of_the_episode,
             fill=title_of_the_episode))+
  geom_col()+
  geom_text(size=3,
            position = position_dodge(0.9),
            vjust=0)+
  theme_minimal()+
  labs(title = "Top 5 most viewed people",
      x = "episode",
      y= "view")+
  theme(axis.text.x = element_blank(),
        legend.position = "none")
```



7.The top 5 cinematography who has worked together the most

```
cinematography <- Game_of_Thrones %>%
    select(cinematography_by) %>%
    group_by(cinematography_by) %>%
    count() %>%
    arrange(desc(n)) %>%
    filter(n >= 5)
```

```
cinematography
```

```
## # A tibble: 6 x 2
## # Groups:   cinematography_by [6]
##   cinematography_by     n
##   <chr>             <int>
## 1 Anette Haellmigk     10
## 2 Jonathan Freeman      9
## 3 Fabian Wagner         8
## 4 Robert McLachlan      8
## 5 David Franco          6
## 6 Gregory Middleton     6
```

Anette Haellmigk and Jonathan Freeman

```r
cinematography %>%
  ggplot(aes(cinematography_by,
             n,
             label = cinematography_by,
             fill=cinematography_by))+
  geom_col()+
  geom_text(size=3,
            position = position_dodge(0.9),
            vjust=0)+
  theme_minimal()+
  labs(title = "The top 5 cinematography who has worked together the most",
       x = "cinematography",
       y= "amount")+
  theme(axis.text.x = element_blank(),
        legend.position = "none")
```