



Data Visualization: Marvel & DC Live-Action Films released in Movie Theaters



Playing Now





| Team Member



Haiva Qurrota A

06211840000045



Umar Sodi Adhi

06211840000067



Andi Fahira Alsar

06211840000004

I TABLE OF CONTENTS

01 Data Source

Sumber Data

02 Visualization

Menggunakan R,
Ms. Excel, Python.





01 Data Source



I Sumber Data

Marvel & DC Live-Action Films released in Movie Theaters

Data yang digunakan merupakan data sekunder yang diakses melalui website [Kaggle.com](https://www.kaggle.com) tentang 90 film *live-action* produksi Marvel dan DC.





I Variabel Data

No.	Variabel	Skala	Keterangan
1.	Title	Nominal	Nama film
2.	Year	Nominal	Tahun rilis
3.	Runtime	Interval	Durasi film (menit)
4.	MPA Rating	Ordinal	Sistem Kategori film MPA
5.	IMDB Rating	Ordinal	Peringkat IMBD
6.	TOM_Ratings	Ordinal	Peringkat Tomato
7.	Gross	Rasio	Penghasilan kotor film
8.	Entity	Nominal	Entitas untuk premis film

[Home](#)

[Team Member](#)

[Source](#)

[Visualization](#)

[Reference](#)



02

Data Visualization



ITS
Institut
Teknologi
Sepuluh Nopember

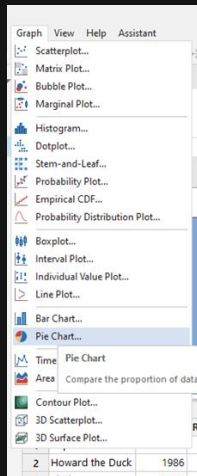


| Minitab

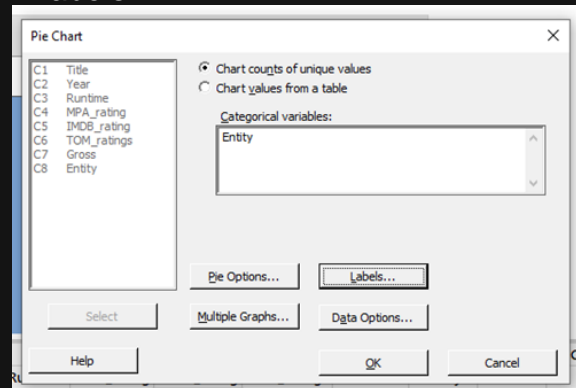
Piechart

Langkah-langkah (Minitab)

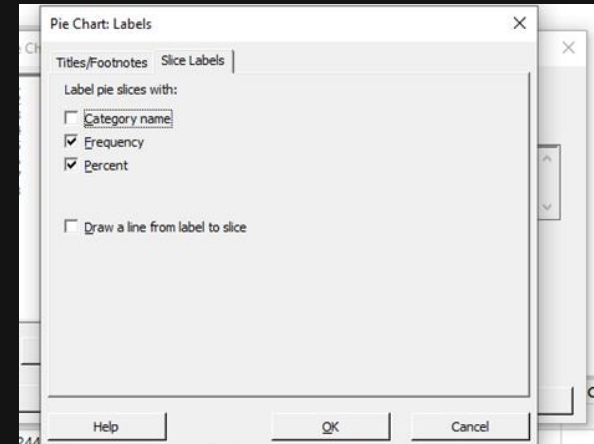
1. Klik Graph > Piechart

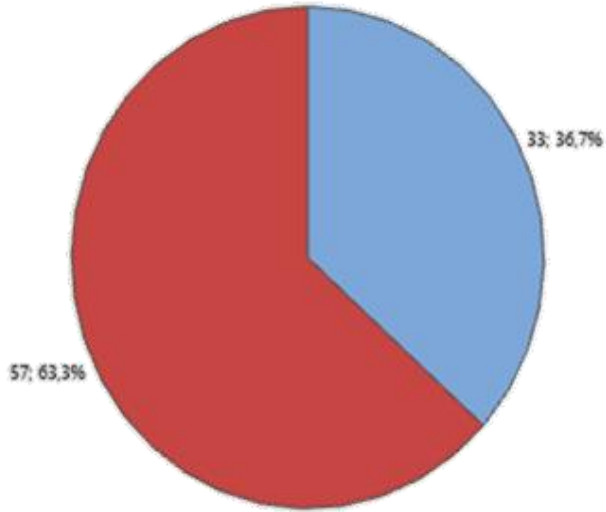
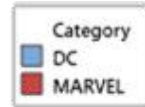


2. Masukkan variabel berikut, Klik Labels



3. Klik Frequency dan percent > Klik OK





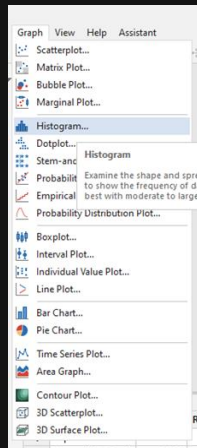
Persentase film *live-action* Marvel lebih banyak dari DC dengan angka persentase Marvel sebanyak 57,63% dan DC sebanyak 33,37%.



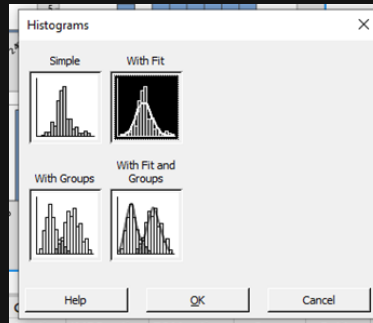
I Minitab

Histogram

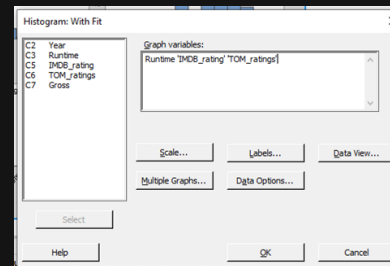
1. Klik Graph > Histogram



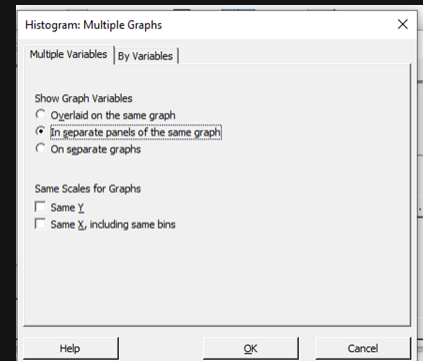
2. Pilih With Fit > OK



3. Masukkan Variabel berikut,
Klik Multiple Graph

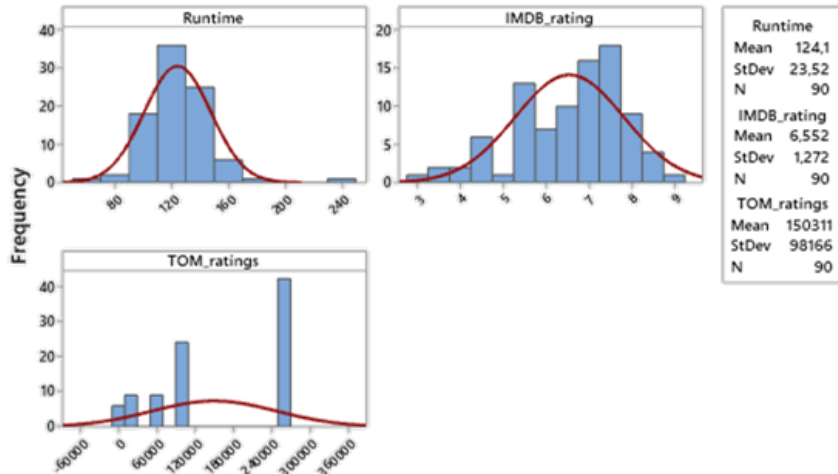


4. Pilih in separate panels > OK





Histogram of Runtime; IMDB_rating; TOM_ratings
Normal



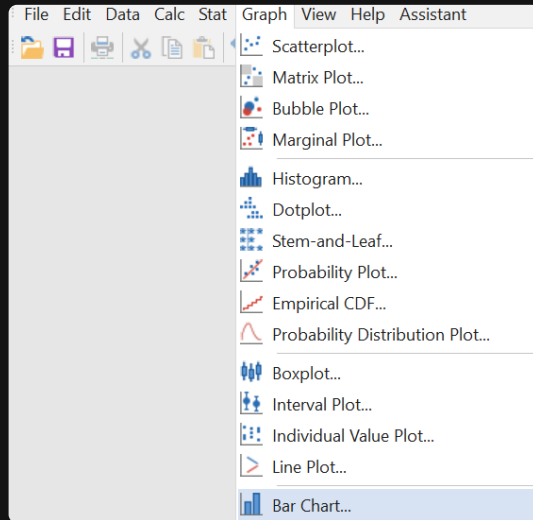
Secara umum, peringkat IMDB memiliki *Skewness* normal karena frekuensi tertinggi berada ditengah dan merata. Peringkat Tomato memiliki *Skewness* positif karena frekuensi tertinggi berada disebelah kanan. Pada variabel Runtime didominasi 90-150 menit. Pada Peringkat IMBD, mayoritas film memiliki rating 5,25-8,25. Pada Peringkat Tomato, fil mendominasi rating dari 90000-110000 dan 250000-270000.



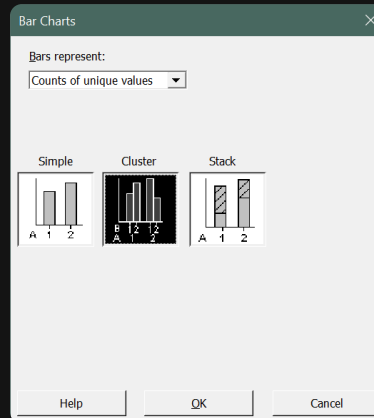
| Minitab

Barchart

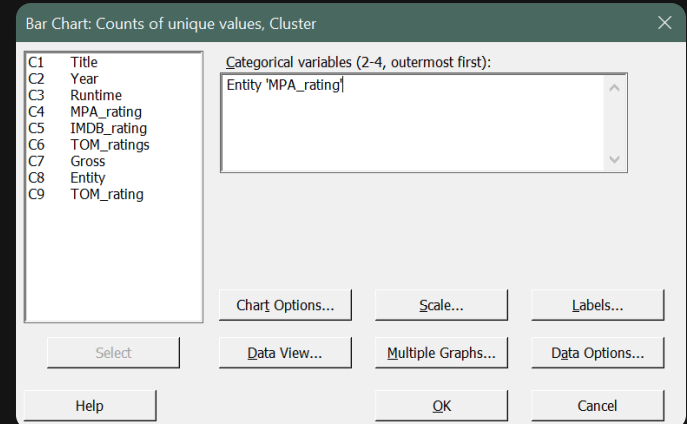
1. Klik Graph > Barchart

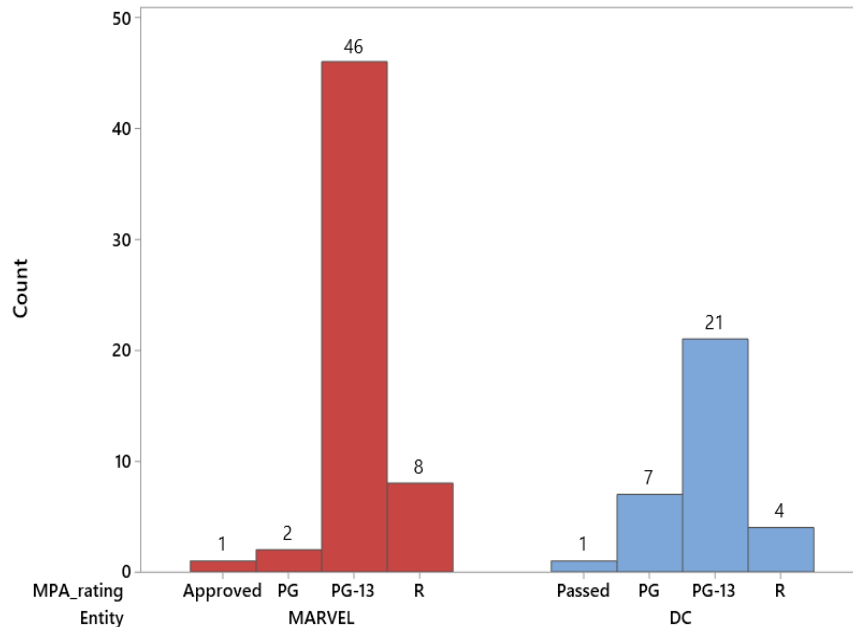


2. Pilih Cluster > OK



3. Masukkan Variabel berikut,
Klik OK





- Untuk Film *live action* kedua perusahaan mayoritas memiliki kategori film PG-13 dengan Marvel sebanyak 46 film dan DC sebanyak 21 film.
- Marvel memiliki 1 film yang mendapat rate Appored sedangkan DC tidak memiliki film yang mendapatkan rate tersebut.
- DC memiliki 1 film yang mendapat rate Paseed sedangkan Marvel tidak memiliki film yang mendapatkan rate tersebut.

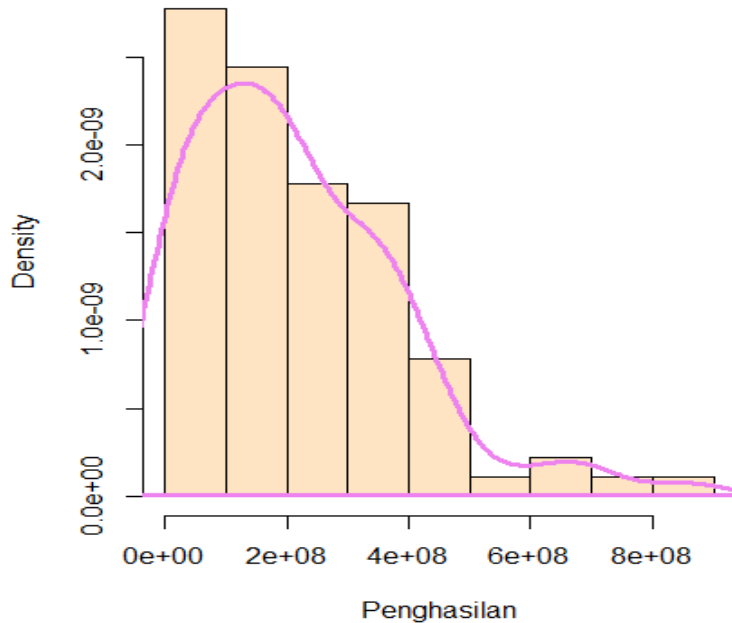


I Rstudio/ R

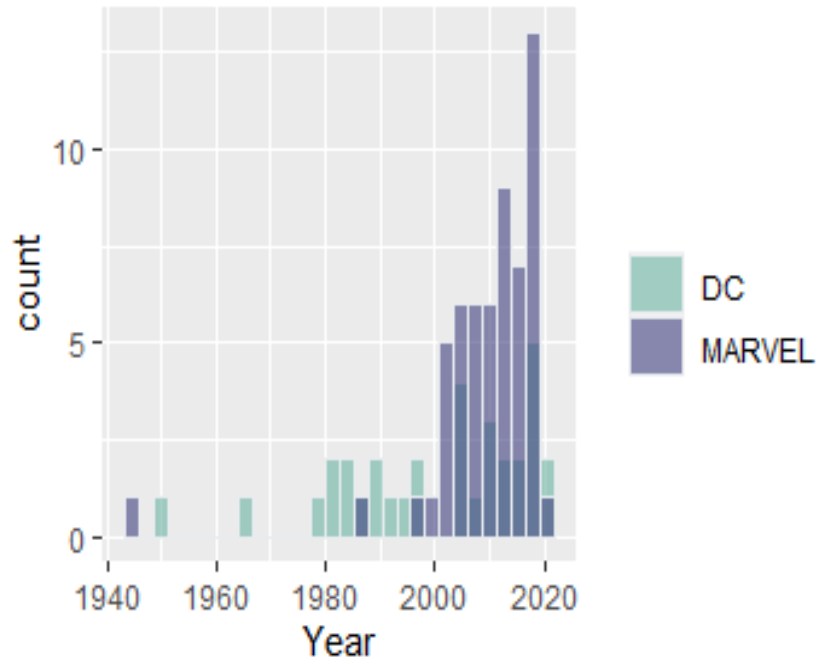
```
library(data.table)
library(ggplot2)
data = read.csv("D:/KULIAH/SEMESTER 7/ANALISA DATA/PRAKTIKUM/P2/DATA.csv", sep
  = ";")
#histogram
plot = ggplot(data,aes(x=Year, fill=Entity)) + geom_histogram( color="#e9ecef",
  alpha=0.6, position = 'identity') + scale_fill_manual(values=c("#69b3a2",
  "#404080"))+ labs(fill="")
plot
#histogram dengan density
dens=density(data$Gross)
hist(data$Gross,freq=FALSE,col="bisque",main="Penghasilan Film",xlab =
  "Penghasilan")
polygon(dens,border = "golden rod",lwd=3)
```



Penghasilan Film



- Data tidak berdistribusi normal
- Sebaran data cenderung kekiri, hal ini menunjukkan bahwa data cenderung memiliki skewness negatif.
- Mayoritas Film memiliki penghasilan yang tidak begitu tinggi yakni hingga 225 milyar US Dollar.



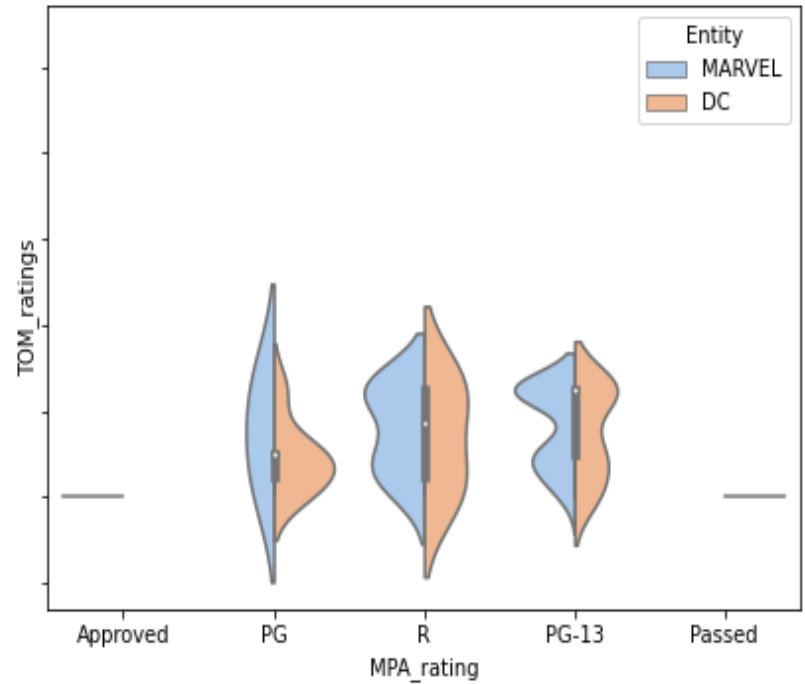
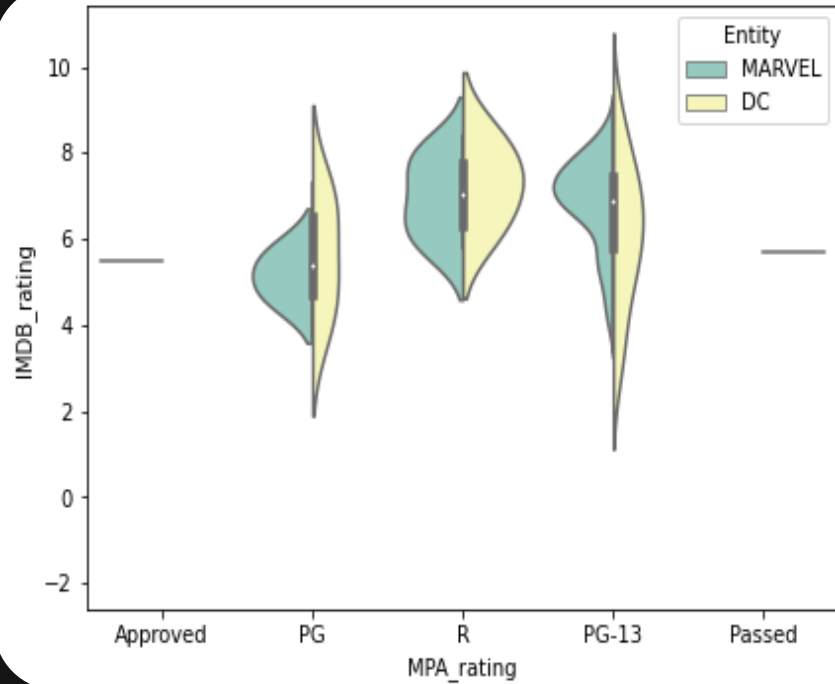
- Berdasarkan tahun rilisnya mayoritas film dirilis antara tahun 2000 sampai 2020.
- Terjadi peningkatan produksi film yang cukup signifikan sejak tahun 2000
- Antara tahun 1940-2000 DC lebih banyak merilis film dibandingkan MARVEL
- Antara tahun 2000-2020 MARVEL lebih banyak merilis film dibandingkan DC



I Phyton

```
#Plot Rating
fig, axes = plt.subplots(1, 2, figsize=(15, 5), sharey=True)
sns.violinplot(ax=axes[0], x=data["MPA_rating"],
y=data["IMDB_rating"],hue=data["Entity"], palette="Set3",split=True)
sns.violinplot(ax=axes[1], x=data["MPA_rating"],
y=data["TOM_ratings"],hue=data["Entity"], palette="pastel", split=True)
#Plot Runtime
hist= sns.displot(data,x="Runtime",hue="Entity",bins=10)
hist
```

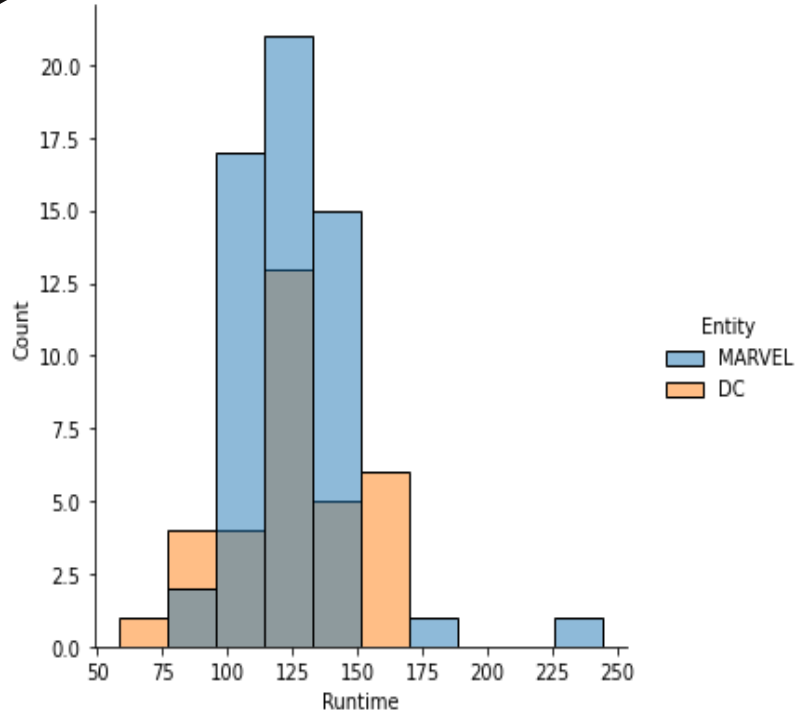






| Interpretasi

- Untuk Kategori PG, film-film produksi Marvel memiliki peringkat IMDB yang lebih homogen disekitar 5-6 dibanding DC yang memiliki peringkat disekitar 2-9. Sedangkan untuk peringkat Tomato, film-film produksi DC lebih homogen di sekitar 0-1 dibanding Marvel yang berada antara -2 hingga 4.
- Untuk Kategori R, film-film produksi Marvel mayoritas memiliki rating IMDB disekitar 5,8-7, juga disekitar 7,5-8 sedangkan DC memiliki rating mayoritas disekitar 7-8. Pada Tomato film produksi Marvel memiliki peringkat diangka 0 dan 2, sedangkan DC memiliki peringkat disekitaran-2 hingga 3.
- Untuk Kategori PG-13, film-film produksi Marvel memiliki peringkat IMDB yang lebih homogen disekitar angka 7 dibanding DC yang lebih bervariasi disekitar 2-10. Sedangkan untuk peringkat Tomato, film-film produksi marvel mayoritas memiliki rating disekitar angka 1 dan 3 dan film produksi DC mayoritas memiliki rating disekitar 3.



- Film dengan durasi terpanjang berasal dari produksi Marvel, yakni film Captain America (1944) dengan 244 menit.
- Film terpendek berasal dari produksi DC yakni Superman and the Mole-Men (1951) dengan durasi 58 menit.
- Mayoritas durasi film dari kedua perusahaan berdurasi antara 120-130 menit.



I Kesimpulan

Skala Nominal

Skala data nominal dapat divisualisasikan menggunakan histogram dan Pie chart. Pada variabel Year, Antara tahun 1940-2000 DC lebih banyak merilis film dibandingkan MARVEL. Antara tahun 2000-2020 MARVEL lebih banyak merilis film dibandingkan DC. Piechart menunjukkan bahwa film live-action diproduksi lebih banyak oleh perusahaan MARVEL dibanding DC.

Skala Ordinal

Skala data ordinal dapat divisualisasikan menggunakan bar chart dan violin plot. Untuk Film live action kedua perusahaan mayoritas memiliki kategori film PG-13 dengan Marvel sebanyak 46 film dan DC sebanyak 21 film.

Skala Interval

Skala data interval dapat divisualisasikan menggunakan histogram. Pada variabel Runtime, mayoritas durasi film dari kedua perusahaan berdurasi antara 120-130 menit

Skala Rasio

Skala data rasio dapat divisualisasikan menggunakan histogram. Pada variable gross, mayoritas film memiliki penghasilan yang tidak begitu tinggi yakni hingga 225 milyar US Dollar.



THANKS!



Praktikum Analisis Data C
Program Sarjana Jurusan Statistika
Tahun Ajaran 2021/2022



I References

1. <https://www.kaggle.com/jcraggy/marvel-vs-dc-imdb-rotten-tomatoes>