

Kuis IF-C

123200024_REGITA AMELIA ASNAWI PUTRI_KUIS_IF-C

2022-10-05

##Intro

1. Boleh menggunakan PC lab / Laptop pribadi
2. Pengumpulan berupa hasil knit Rmd ke pdf dengan nama NIM_NAMA_KUIS_IF-C.pdf.
3. Durasi 2 Jam + 5 menit submit, > tidak bisa mengumpulkan.
4. Pengerjaan offline, pengumpulan di Spada (online).
5. Tidak boleh buka modul.
6. Tidak boleh membuka internet (googling, WhatsApp, ig, sosmed, dan media komunikasi lain).
7. Boleh bawa catatan 1 lembar A4.
8. Izin keluar maks. 1x dengan durasi 2 menit.
9. Tas dan HP diletakkan di depan.

Selamat mengerjakan, semoga sukses!

1. Load library apa saja yang kira-kira digunakan! Lalu gunakan data 'movielens'!
point 2

```
library(dslabs)
library(tidyverse)

## — Attaching packages ————— tidyverse
1.3.2 —
## ✓ ggplot2 3.3.6      ✓ purrr  0.3.4
## ✓ tibble  3.1.8      ✓ dplyr  1.0.10
## ✓ tidyr   1.2.0      ✓ stringr 1.4.1
## ✓ readr   2.1.2      ✓ forcats 0.5.2
## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()

library(dplyr)
data(movielens)
```

2. Tampilkan struktur dari data Movielens! **point 5**

```
str(movielens)

## 'data.frame':  100004 obs. of  7 variables:
## $ movieId : int  31 1029 1061 1129 1172 1263 1287 1293 1339 1343 ...
## $ title   : chr  "Dangerous Minds" "Dumbo" "Sleepers" "Escape from New
```

```

York" ...
## $ year      : int   1995 1941 1996 1981 1989 1978 1959 1982 1992 1991 ...
## $ genres    : Factor w/ 901 levels "(no genres listed)",...: 762 510 899
120 762 836 81 762 844 899 ...
## $ userId    : int    1 1 1 1 1 1 1 1 1 1 ...
## $ rating    : num    2.5 3 3 2 4 2 2 2 3.5 2 ...
## $ timestamp: int    1260759144 1260759179 1260759182 1260759185 1260759205
1260759151 1260759187 1260759148 1260759125 1260759131 ...

```

Jawablah pertanyaan berikut a. Class dari object movielens:

```

class(movielens)
## [1] "data.frame"

```

b. Jumlah baris:

c. Jumlah kolom:

3. Implementasikan fungsi yang dapat digunakan jika hanya ingin menampilkan nama-nama variabel dari object movielens! **point 5**

```

names(movielens)
## [1] "movieId" "title" "year" "genres" "userId" "rating"
## [7] "timestamp"

```

4. Tuliskan fungsi untuk mengetahui tipe data dari variabel genres, dan jelaskan pada bawah chunk kenapa variabel genres menggunakan tipe data tersebut! **point 5**

```

class(movielens$genres)
## [1] "factor"

```

Alasan : karena variabel genres menyimpan data kategorial atau jenis sehingga tipe datanya adalah factor.

5. Tampilkan 5 observasi terakhir dari data Movielens! **point 5**

```

tiny_df <- movielens%>%
select(movieId, title, year, genres, userId, rating, timestamp)%>%
head(tiny_df, n=5)
tiny_df
##   movieId          title year
## 1      31      Dangerous Minds 1995
## 2    1029          Dumbo 1941
## 3    1061          Sleepers 1996
## 4    1129  Escape from New York 1981
## 5    1172 Cinema Paradiso (Nuovo cinema Paradiso) 1989
##           genres userId rating timestamp
## 1          Drama      1    2.5 1260759144
## 2 Animation|Children|Drama|Musical      1    3.0 1260759179
## 3          Thriller      1    3.0 1260759182

```

```
## 4 Action|Adventure|Sci-Fi|Thriller      1      2.0 1260759185
## 5                                     Drama      1      4.0 1260759205
```

6. Tampilkan 10 observasi teratas movielens tanpa variabel movieId & userId, simpan ke dalam object bernama "tiny_df"! **point 15**

```
tiny_df <- movielens%>%
select(title, year, genres, rating, timestamp)%>%
head(tiny_df, n=10)
tiny_df
```

##		title	year	
## 1		Dangerous Minds	1995	
## 2		Dumbo	1941	
## 3		Sleepers	1996	
## 4		Escape from New York	1981	
## 5	Cinema Paradiso (Nuovo cinema Paradiso)		1989	
## 6		Deer Hunter, The	1978	
## 7		Ben-Hur	1959	
## 8		Gandhi	1982	
## 9	Dracula (Bram Stoker's Dracula)		1992	
## 10		Cape Fear	1991	
##		genres	rating	timestamp
## 1		Drama	2.5	1260759144
## 2	Animation Children Drama Musical		3.0	1260759179
## 3		Thriller	3.0	1260759182
## 4	Action Adventure Sci-Fi Thriller		2.0	1260759185
## 5		Drama	4.0	1260759205
## 6		Drama War	2.0	1260759151
## 7	Action Adventure Drama		2.0	1260759187
## 8		Drama	2.0	1260759148
## 9	Fantasy Horror Romance Thriller		3.5	1260759125
## 10		Thriller	2.0	1260759131

7. Tampilkan 20 data teratas dengan genre hanya romance dengan diurutkan rating dari yang tertinggi. Tampilkan data dengan variabel title, year, genres, dan rating saja! **point 15** Hilangkan data yang dengan judul berulang(Poin plus) **point 5**

```
a <- movielens%>%
select(title, year, genres, rating)%>%
filter(rating>=5.0)
head(a, 20)
```

##		title	year
## 1		Sense and Sensibility	1995
## 2		Clueless	1995
## 3		Apollo 13	1995
## 4		Circle of Friends	1995
## 5	Like Water for Chocolate (Como agua para chocolate)		1992
## 6		Legends of the Fall	1994
## 7	Nightmare Before Christmas, The		1993
## 8		Brady Bunch Movie, The	1995

```

## 9 Terminator 2: Judgment Day 1991
## 10 Dances with Wolves 1990
## 11 Batman 1989
## 12 Shawshank Redemption, The 1994
## 13 Forrest Gump 1994
## 14 Princess Bride, The 1987
## 15 Fight Club 1999
## 16 Requiem for a Dream 2000
## 17 Babe 1995
## 18 Rumble in the Bronx (Hont faan kui) 1995
## 19 Birdcage, The 1996
## 20 Star Wars: Episode IV - A New Hope 1977
## genres rating
## 1 Drama|Romance 5
## 2 Comedy|Romance 5
## 3 Adventure|Drama|IMAX 5
## 4 Drama|Romance 5
## 5 Drama|Fantasy|Romance 5
## 6 Drama|Romance|War|Western 5
## 7 Animation|Children|Fantasy|Musical 5
## 8 Comedy 5
## 9 Action|Sci-Fi 5
## 10 Adventure|Drama|Western 5
## 11 Action|Crime|Thriller 5
## 12 Crime|Drama 5
## 13 Comedy|Drama|Romance|War 5
## 14 Action|Adventure|Comedy|Fantasy|Romance 5
## 15 Action|Crime|Drama|Thriller 5
## 16 Drama 5
## 17 Children|Drama 5
## 18 Action|Adventure|Comedy|Crime 5
## 19 Comedy 5
## 20 Action|Adventure|Sci-Fi 5
max(movielens$rating)
## [1] 5

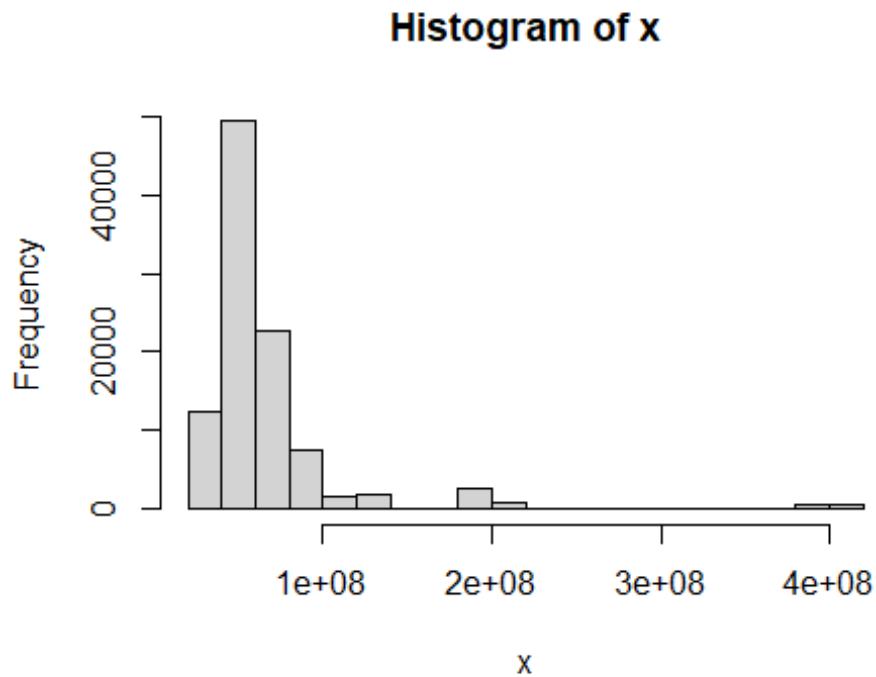
```

8. Buatlah histogram dengan sumbu x tahun pembuatan film dan sumbu y frekuensi banyaknya film yang ada di tahun tersebut. Berikan warna kuning dan garis batas biru pada histogram dan beri judul yang sesuai. Di bawah chunk tuliskan kesimpulan yang dapat diperoleh dari histogram yang anda buat! **point 13**

```

x <- with(movielens, (year / rating)* 100000)
hist(x)

```



Kesimpulan: Tahun dengan jumlah film terbanyak dalam data movielens adalah sekitar tahun...

9. Klasifikasikan data dari object "tiny_df" dari soal no 4 berdasarkan rating dengan kondisi : -Rating kurang dari 2 dikategorikan sebagai "Negative" -Rating lebih dari 3 dikategorikan sebagai "Positive" -Rating antara kedua kondisi diatas dikategorikan sebagai "Mixed" **point 15**

##rating kurang dari 2

```
tiny_df<-movielens
Negative = tiny_df%>%
filter(rating<2)
head(Negative, 5)

##      movieId          title year
##      1405 Beavis and Butt-Head Do America 1996
##      2968          Time Bandits 1981
##      223          Clerks 1994
##      319          Shallow Grave 1994
##      435          Coneheads 1993

##      userId rating  timestamp
##      1      1      1 1260759203
##      2      1      1 1260759200
```

```
## 3      2      1 835355749
## 4      2      1 835355918
## 5      4      1 949920135
```

##rating lebih dari 3

```
tiny_df<-movielens
Positive = tiny_df%>%
filter(rating>3)
head(Positive, 5)
```

```
##  movieId      title year
## 1    1172 Cinema Paradiso (Nuovo cinema Paradiso) 1989
## 2    1339      Dracula (Bram Stoker's Dracula) 1992
## 3    1953      French Connection, The 1971
## 4    2105      Tron 1982
## 5      10      GoldenEye 1995
##              genres userId rating  timestamp
## 1              Drama      1    4.0 1260759205
## 2 Fantasy|Horror|Romance|Thriller      1    3.5 1260759125
## 3      Action|Crime|Thriller      1    4.0 1260759191
## 4      Action|Adventure|Sci-Fi      1    4.0 1260759139
## 5      Action|Adventure|Thriller      2    4.0 835355493
```

##rating antara kedua kondisi

```
tiny_df<-movielens
Mixed = tiny_df%>%
filter(rating<=2, rating>=3)
head(Mixed, 5)
```

```
## [1] movieId  title    year    genres    userId    rating    timestamp
## <0 rows> (or 0-length row.names)
```

10. Tambahkan variabel baru dalam data “tiny_df” dengan tambahan kolom baru dengan nama “sentiment” yang isinya merupakan implementasi nomor 9! **point 20**

```
## datax <- tiny_df
## x <- datax
## mutate(datax, sentiment = ifelse(rating<2, "Negative"),
##              ## ifelse(rating>3, "Positive"), ifelse("Mixed",
##              "Negative"))
```