UNIVERSITÉ CÔTE D'AZUR

Processing Large Datasets with R

Exam

Joris LIMONIER

December 7, 2021

Table of Contents

1. Exercise 1 - Shiny (Movies dataset)

1. Exercise 1 - Shiny (Movies dataset)

1.1 Question 1

Exercise 1 - Shiny (Movies dataset)

Exercise 1

Question 1

Columns

Rank Movie Release_Date Distributor Genre MPAA Gross_Sales Tickets_Sold

Dimension of the data:

50 rows and 8 columns

Import and show the data

Show 10 entries Search:

	Rank	Movie	Release_Date	Distributor	Genre	MPAA	Gross_Sales	Tickets_Sold
1	1	The Lego Movie	2/7/14	Warner Bros.	Adventure	PG	248303720	30429377
2	2	Ride Along	1/17/14	Universal	Comedy	PG-13	133659265	16379811
3	3	Lone Survivor	1/10/14	Universal	Action	R	124722648	15284638
4	4	Frozen	11/27/13	Walt Disney	Adventure	PG	121285671	14863440
5	5	300: Rise of an Empire	3/7/14	Warner Bros.	Action	R	101145414	12395271
6	6	Divergent	3/21/14	Lionsgate	Adventure	PG-13	95260008	11674020
7	7	Mr. Peabody & Sherman	3/7/14	20th Century Fox	Adventure	PG	94479448	11578363
8	8	Non-Stop	2/28/14	Universal	Action	PG-13	85091060	10427825
9	9	The Monuments Men	2/7/14	Sony Pictures	Drama	PG-13	76599461	9387188
10	10	American Hustle	12/13/13	Sony Pictures	Black Comedy	R	74500902	9130012

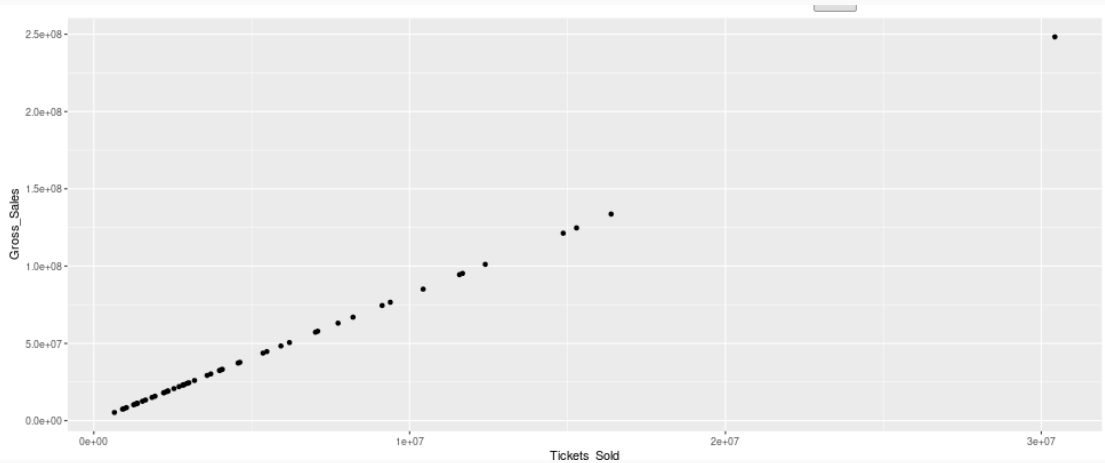
Showing 1 to 10 of 50 entries Previous 1 2 3 4 5 Next

Exercise 1 - Shiny (Movies dataset)

Question 2

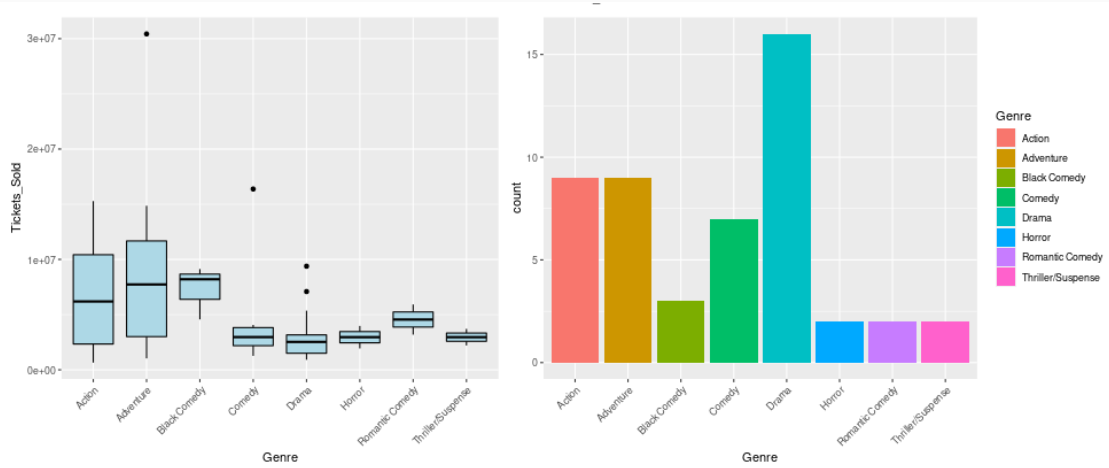
Correlation between tickets sold and sales

0.999999999999999
This is expected since there is a direct relationship between the number of tickets sold and the money gained.



Exercise 1 - Shiny (Movies dataset)

Question 3



Exercise 1 - Shiny (Movies dataset)

Watch video
Backup link: https://youtu.be/NTgGG7UvRRU

Exercise 1 - Shiny (Movies dataset)

Watch video
Backup link: https://youtu.be/w_QQVsRoOpA

2. Exercise 2 - RMarkdown (Winter dataset)

Exercise 1 - Shiny (Movies dataset)

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

Exercise 1 - Shiny (Movies dataset)

Part 1

Question 1a

```
winter <- read.csv("datasets_exam/winter_olympic.csv")
```

Question 1b

```
head(winter)
```

```
##   Rank                  NOC Gold Silver Bronze Total  Region
## 1    1       Russia (RUS)*   13     11      9    33 EURASIA
## 2    2         Norway (NOR)   11      5     10    26  EUROPE
## 3    3         Canada (CAN)   10     10      5    25 NORTH_A
## 4    4 United States (USA)    9      7     12    28 NORTH_A
## 5    5   Netherlands (NED)    8      7      9    24  EUROPE
## 6    6       Germany (GER)    8      6      5    19  EUROPE
```

Question 1c

```
colnames(winter)
```

```
## [1] "Rank"   "NOC"    "Gold"   "Silver" "Bronze" "Total"  "Region"
```

Exercise 1 - Shiny (Movies dataset)

Question 1d

```
dim(winter)
```

```
## [1] 26  7
```

```
nrow(winter)
```

```
## [1] 26
```

```
ncol(winter)
```

```
## [1] 7
```

3. Exercise 3 - Data Analysis (Summer-Winter dataset)