EMPRESA
CHALLENGE SOLUTION

Name: Matheus Nobre Gomes
E-mail: matt-gomes@live.com
Phone: (11) 2471-7235
Cell phone: (11) 99566-0126 / (11) 96952-5592 (WhatsApp)

SQL test – Solutions

Observations: You can see script.sql and other files that I used to more details

1 How many products does the company have?

The company has 9994 products registered.

Result Grid:

QTD_PRODUTOS_CADASTRADOS
9994

Command:

```
# 1. How many products does the company have?
select count(PRODUCT_COD) as QTD_PRODUTOS_CADASTRADOS from data_product;
-- There are 9994 products registered.
```

2 What are the 10 most expensive products in the company?

Result Grid:

	CÓDIGO	NOME	PRECO_UNITARIO_R$
▶	301409	Whisky Escoces THE MACALLAN Ruby Garrafa 7...	741.99
	176185	Whisky Escoces JOHNNIE WALKER Blue Label G...	735.90
	315481	Cafeteira Expresso 3 CORACOES Tres Modo Ve...	499.00
	100280	Vinho Portugues Tinto Vintage QUINTA DO CRA...	445.90
	320046	Escova Dental Eletrica ORAL B D34 Professional...	399.90
	190817	Champagne Rose VEUVE CLICQUOT PONSARDI...	366.90
	153795	Champagne Frances Brut Imperial MOET Rose G...	359.90
	311397	Conjunto de Panelas Allegra em Inox TRAMONT...	359.00
	147706	Whisky Escoces CHIVAS REGAL 18 Anos Garraf...	329.90
	154431	Champagne Frances Brut Imperial MOET & CHA...	315.90

Command:

```
# 2. What are the 10 most expensive products in the company?
select PRODUCT_COD as CÓDIGO, PRODUCT_NAME as NOME, PRODUCT_VAL as PRECO_UNITARIO_R$ from data_product order by PRODUCT_VAL desc limit 10;
-- COD: 301409, 176185, 315481, 100280, 320046, 190817, 153795, 311397, 147706, 154431
```

3 What sections do the 'BEBIDAS' and 'PADARIA' departments have?

Result Grid:

	CÓDIGO	SESSÃO
▶	4	BEBIDAS
	29	CERVEJAS
	8	DOCES-E-SOBREMESAS
	27	GESTANTE
	19	PADARIA
	22	QUEIJOS-E-FRIOS
	31	REFRESCOS
	30	VINHOS

Command:

```
# 3. What sections do the 'BEBIDAS' and 'PADARIA' departments have?
select min(SECTION_COD) as CÓDIGO, SECTION_NAME as SESSÃO from data_product where DEP_NAME = "BEBIDAS" or DEP_NAME = "PADARIA" group by SECTION_NAME;
-- Bebidas, cervejas, doces e sobremesas, gestante, padaria, queijos e frios, refrescos, vinhos
```

4 When were the most products sold? In which store?

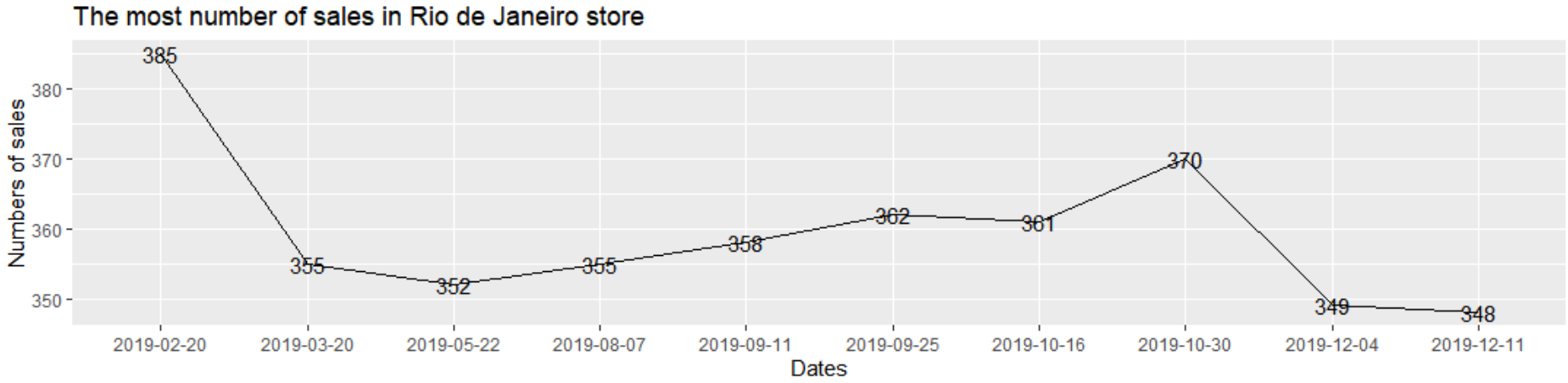
Result Grid:

	DATA	QTD_VENDAS	CÓD_DA_LOJA	NOME_DA_LOJA
▶	2019-02-20	385	11	Rio de Janeiro
	2019-10-30	370	11	Rio de Janeiro
	2019-09-25	362	11	Rio de Janeiro
	2019-10-16	361	11	Rio de Janeiro
	2019-09-11	358	11	Rio de Janeiro
	2019-03-20	355	11	Rio de Janeiro
	2019-08-07	355	11	Rio de Janeiro
	2019-05-22	352	11	Rio de Janeiro
	2019-12-04	349	11	Rio de Janeiro
	2019-12-11	348	11	Rio de Janeiro

Command:

```
# 4. When were the most products sold? In which store?
select  DATE as DATA,
        SALES_QTY as QTD_VENDAS,
        data_product_sales.STORE_CODE as CÓD_DA_LOJA,
        data_store_cad.STORE_NAME as NOME_DA_LOJA from data_product_sales
        inner join data_store_cad on (data_product_sales.STORE_CODE = data_store_cad.STORE_CODE) order by SALES_QTY DESC limit 10;
-- All dates are from Rio de Janeiro, cod. 11 at 2019.
```

Chart:



5 What was the total sale of the products of each business area in the first quarter of 2019?

Total: 20.196.315 sales considering only sales that have a department registered.

Result Grid (1):

DEPARTAMENTO	QTD_VENDAS
BEBES	2665759
BEBIDAS	2452576
CARNES	1962900
FLV	1266415
FRIOS	2161495
MEDICAMENTOS GENÉRICOS	515280
MEDICAMENTOS REFERÊNCIA	463752
MERCEARIA	2658270
PADARIA	2436955
PERFUMARIA	2913088
PET-SHOP	699825

Command:

```
# 5. Bonus!! What was the total sale of the products of each business area in the first quarter of 2019?
SELECT
    data_product.DEP_NAME AS DEPARTAMENTO,
    SUM(data_product_sales.SALES_QTY) AS QTD_VENDAS
FROM
    data_product_sales
        INNER JOIN
    data_product ON (data_product.PRODUCT_COD = data_product_sales.PRODUCT_CODE)
WHERE
    DATE BETWEEN '2019-01-01' AND '2019-03-31'
GROUP BY DEP_NAME;
```

Total: 20.884.965 sales if considering sales that haven't a department registered yet.

Result Grid (2):

DEPARTAMENTO	QTD_VENDAS
NULL	688650
BEBES	2665759
BEBIDAS	2452576
CARNES	1962900
FLV	1266415
FRIOS	2161495
MEDICAMENTOS GENÉRICOS	515280
MEDICAMENTOS REFERÊNCIA	463752
MERCEARIA	2658270
PADARIA	2436955
PERFUMARIA	2913088
PET-SHOP	699825

Command:

```
# 5. Bonus!! What was the total sale of the products of each business area in the first quarter of 2019?
SELECT
    data_product.DEP_NAME AS DEPARTAMENTO,
    SUM(data_product_sales.SALES_QTY) AS QTD_VENDAS
FROM
    data_product_sales
        LEFT JOIN
    data_product ON (data_product.PRODUCT_COD = data_product_sales.PRODUCT_CODE)
WHERE
    DATE BETWEEN '2019-01-01' AND '2019-03-31'
GROUP BY DEP_NAME;
```

Chart (1):

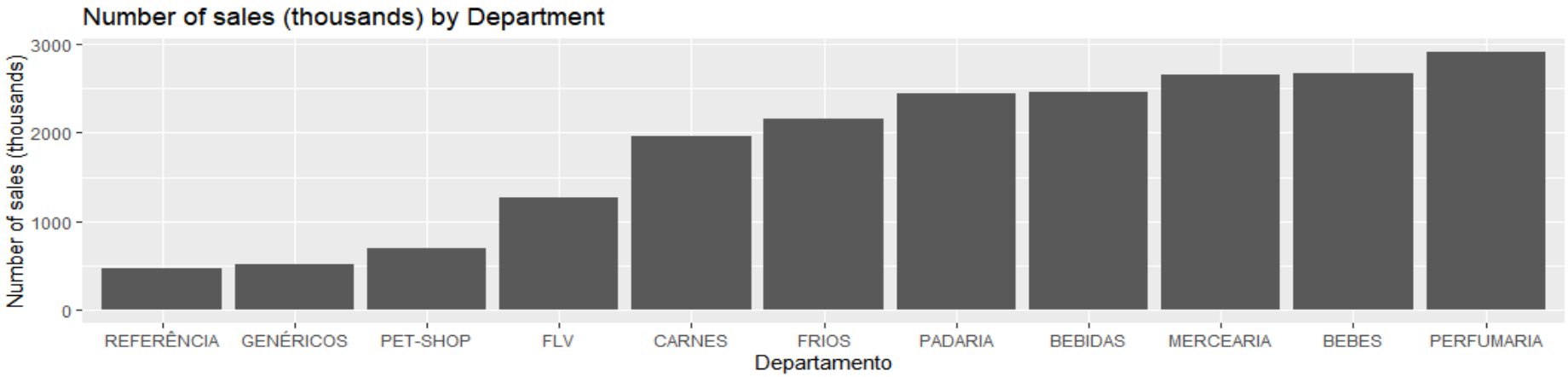
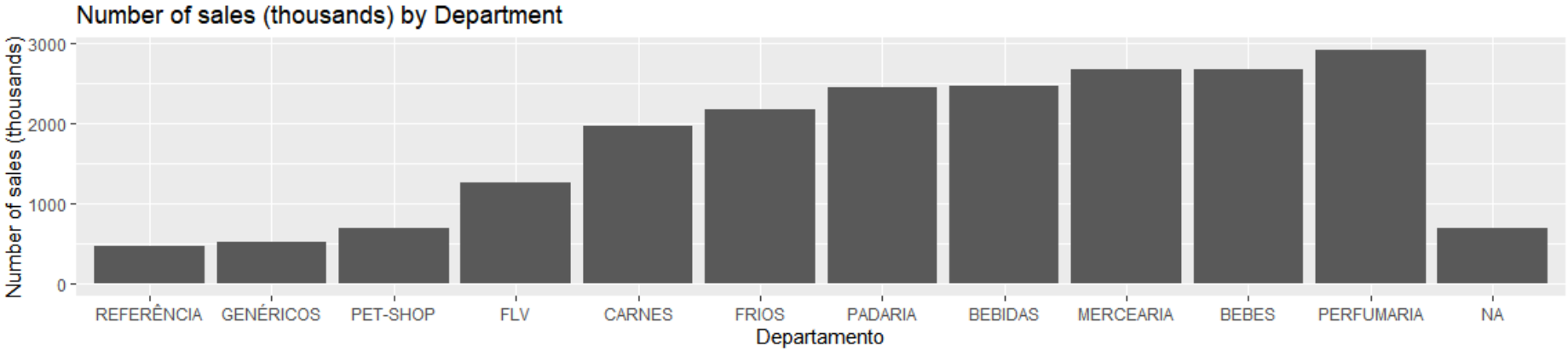


Chart (2):



Building your own visualization

1 Create at least one chart using the table IMDB_movies. The code must be in R or Python, and you are free to use any libraries, data in the table and graphic format. Explain why you chose the visualization you are submitting

---

Introduction

I love watch movies in my leisure and my favorite genre is horror but the most of my friends don't like it. Always when we're looking for a horror movie, usually I already watched them and when I read this question and saw all the information you passed I thought:

- Now is the time to figure out why they don't produce as many horror movies as the other genres.

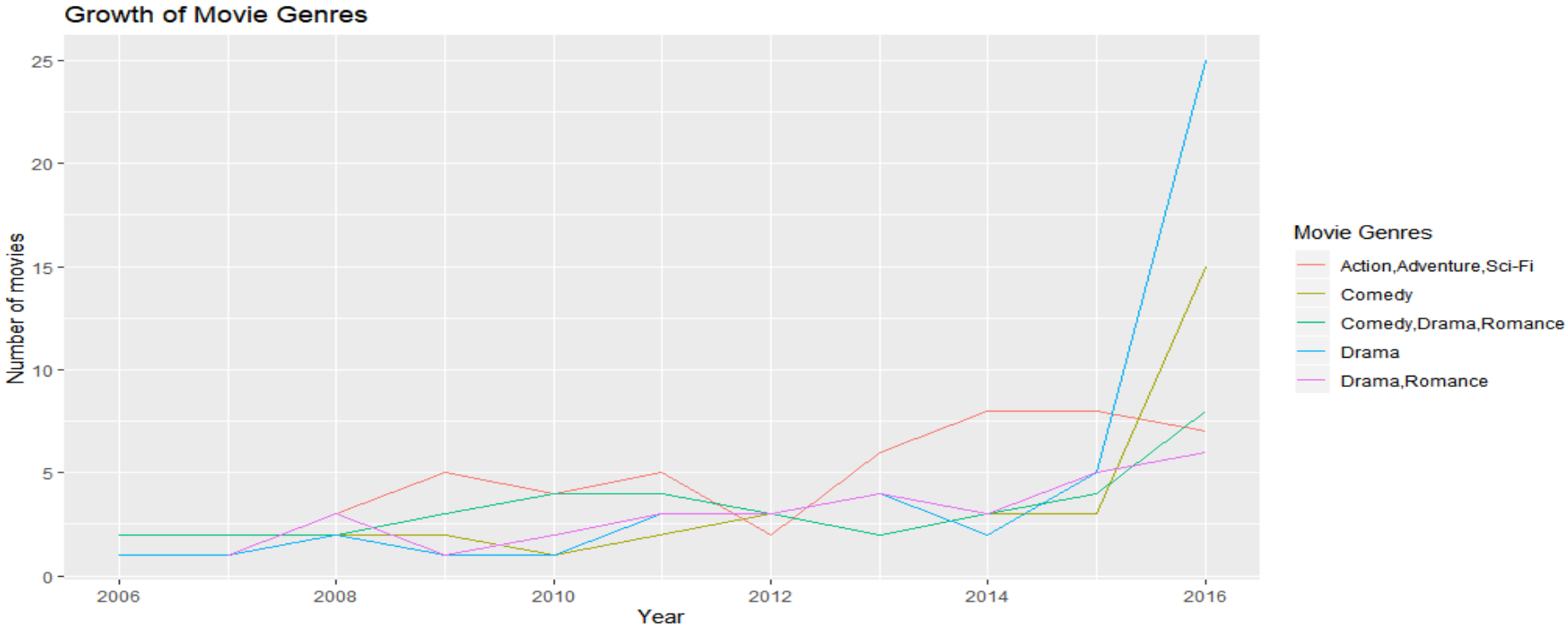
So, the objective of my visualization is understanding what the directors and the companies see when they are choosing a movie to produce.

Notes:

- ***The distribution of movies per year is not proportional, so the analysis will not be so assertive.***
- ***There are 1000 samples among 2006 and 2016 and it's a good quantity but if we add more samples will be better to my analysis.***

Development

I figured out which genres have more movies for unpolluted chart and plotted one chart by year to see how these genres are behaved.

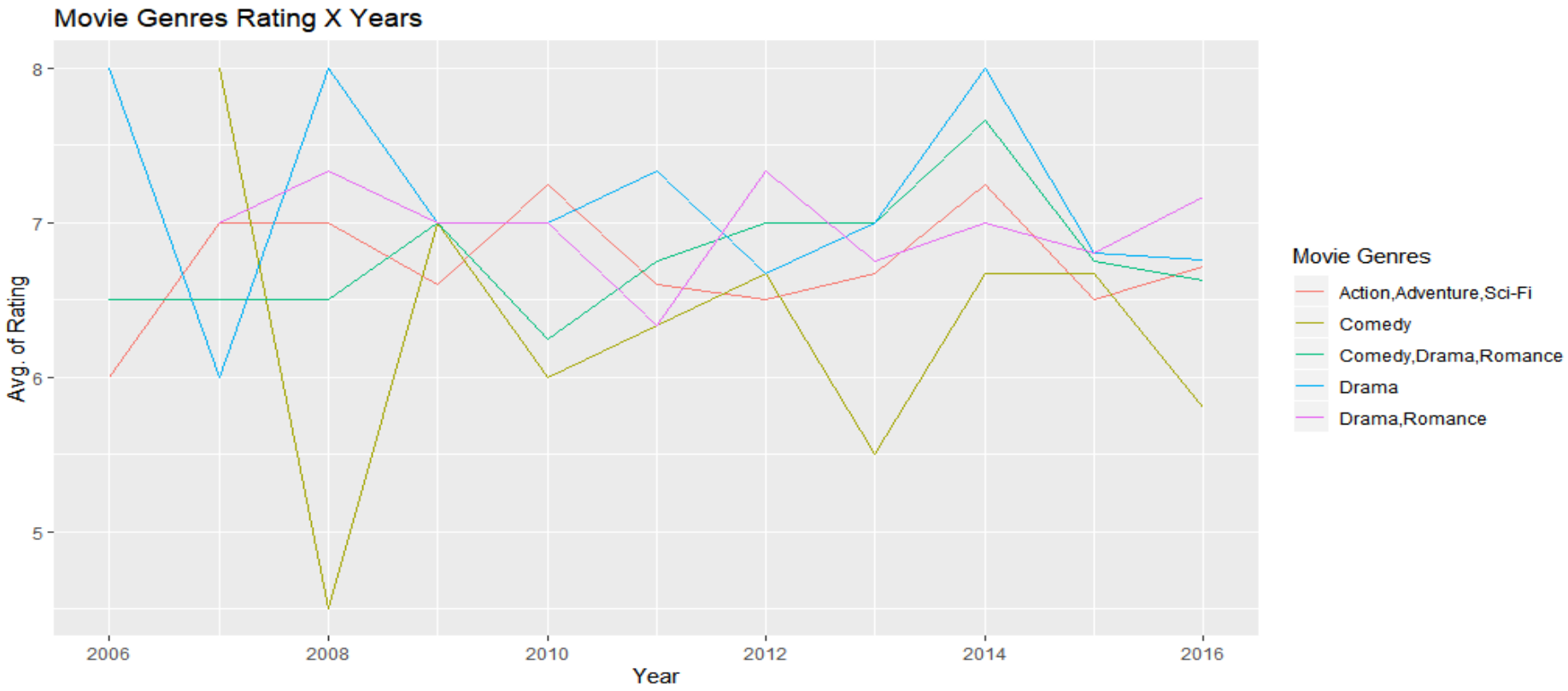


Results:

- The blue and Yellow has been growing exponentially since 2015 but in the past few years have maintained their average.
- The others genres have been keeping their average.

Genre	Number_Movies
Action,Adventure,Sci-Fi	50
Drama	48
Comedy,Drama,Romance	35
Comedy	32
Drama,Romance	31

The next step was comparing their ratings and see if they have been behaving the same.



Results:

- The blue and Yellow didn't grow like their productions
- The others genres have been keeping their average.

Conclusion:

Analyzing this information, we can deduce the genre's rating isn't what decide if more movies will be produce because the rating doesn't up or down like his productions. I need to get more information about these movies to compare, however I used multiple linear regression and backward elimination to see which the information is more significant and they are: runtime, votes and revenue.

I could analyze other information like which genres get more money or something like this but is more interesting for movie lovers like me understand this information as what is really important when you will produce a movie? What is the ideal runtime?

The first movie produced was at December 28th, 1895 and has since evolved. Today we have wonderful effects, 3d movies, 4d theaters and it makes me curious what theaters will look like in a few years.

```
> summary(regressor)

Call:
lm(formula = Rating ~ Runtime + Rating + Votes + RevenueMillions,
    data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max
-4.3168 -0.4121  0.1909  0.5473  1.8626

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.176e+00  4.213e-01  12.285  < 2e-16 ***
Runtime          1.106e-02  3.752e-03   2.948 0.003658 **
Votes            2.792e-06  4.185e-07   6.672  3.6e-10 ***
RevenueMillions -2.654e-03  7.807e-04  -3.400 0.000845 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8357 on 166 degrees of freedom
  (26 observations deleted due to missingness)
Multiple R-squared:  0.2771,    Adjusted R-squared:  0.264
F-statistic: 21.21 on 3 and 166 DF,  p-value: 1.108e-11
```