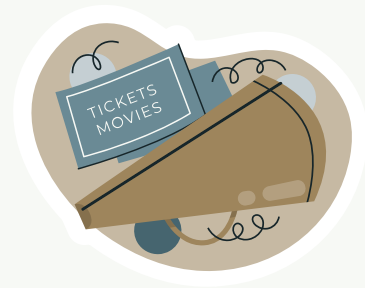# Oscar analysis

4ST626 - Exploratory Data Analysis

Table of contents

01

Data introduction

02

Data manipulation
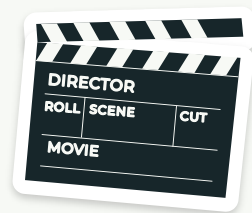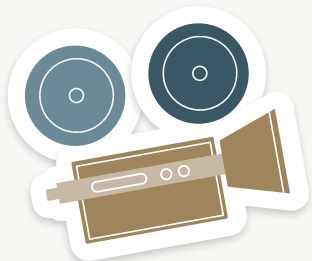
03

Dataset Movies

04

Dataset Oscars

05
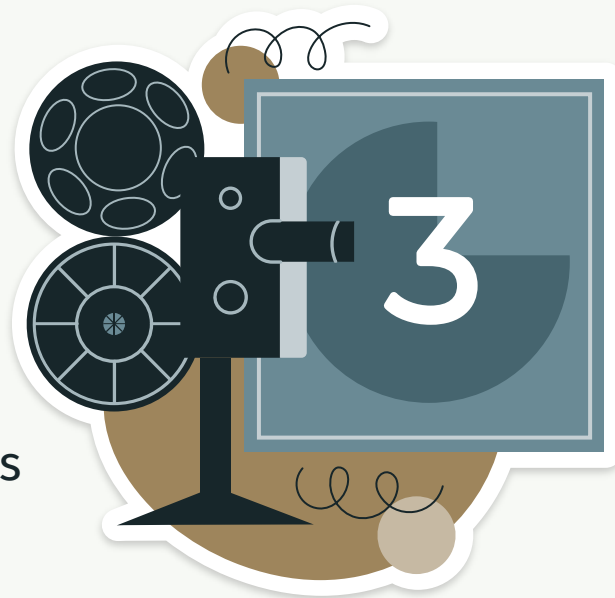
Final Summary

About the data

Dataset: Movies
   - 7 668 rows, 15 columns
   - movie characteristics
      - name, rating, genre, year, budget...

Dataset: Oscars
   - 9 964 rows, 6 columns
   - data about oscar nominations and winners
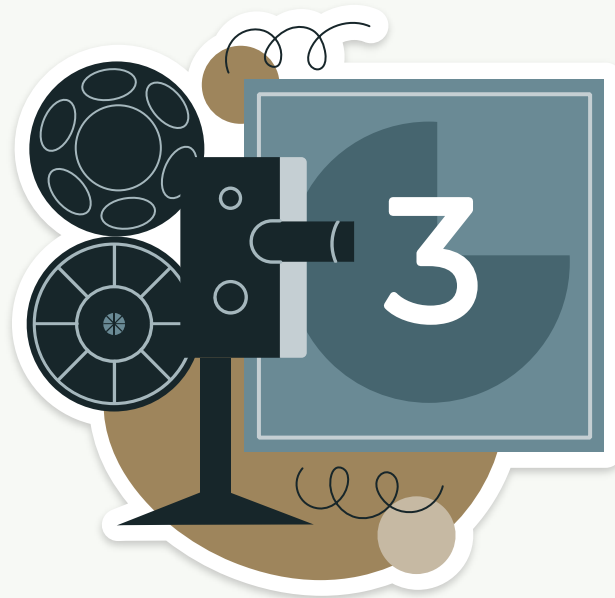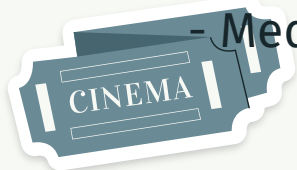   - year, award, name, movie, winner...

Data filtering

Dataset: Movies
 - Listwise deletion of missing values from columns: released country, released day of week, writer, rating, star, country, company

Dataset: Oscars
 - Movies after 1980 (incl.)
 - Column „Award" does not contain words:
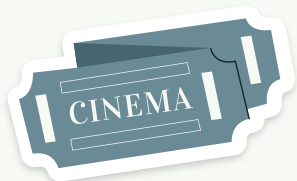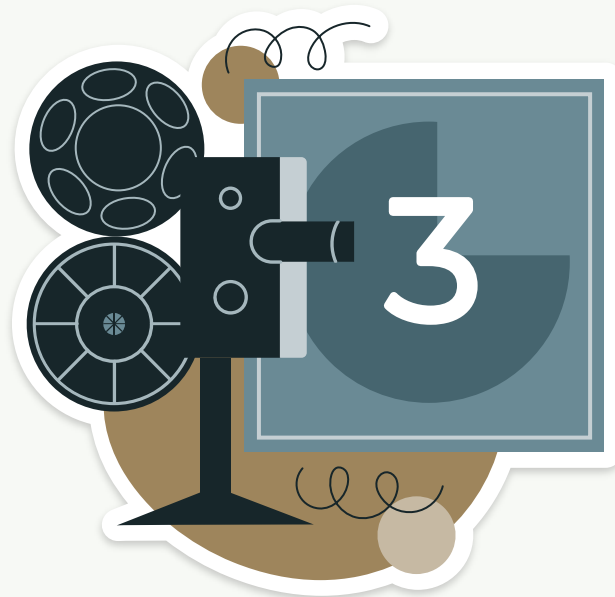   - Award,
   - Memorial, or
   - Medal

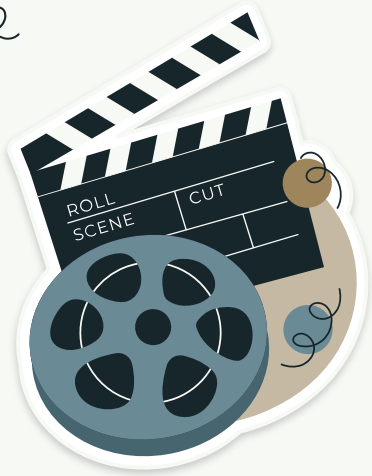# Missing values

Dataset: Movies
- Empty strings across various columns -> NA
- Missing continuous values -> Median
- Inconsistent date (February 2018, 1997...)

Dataset: Oscars
- Nominees (NA -> 0)
- Film ("" -> removed)
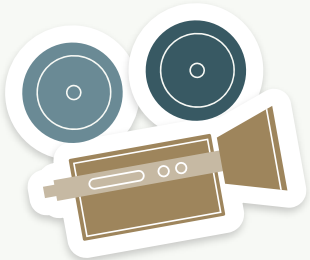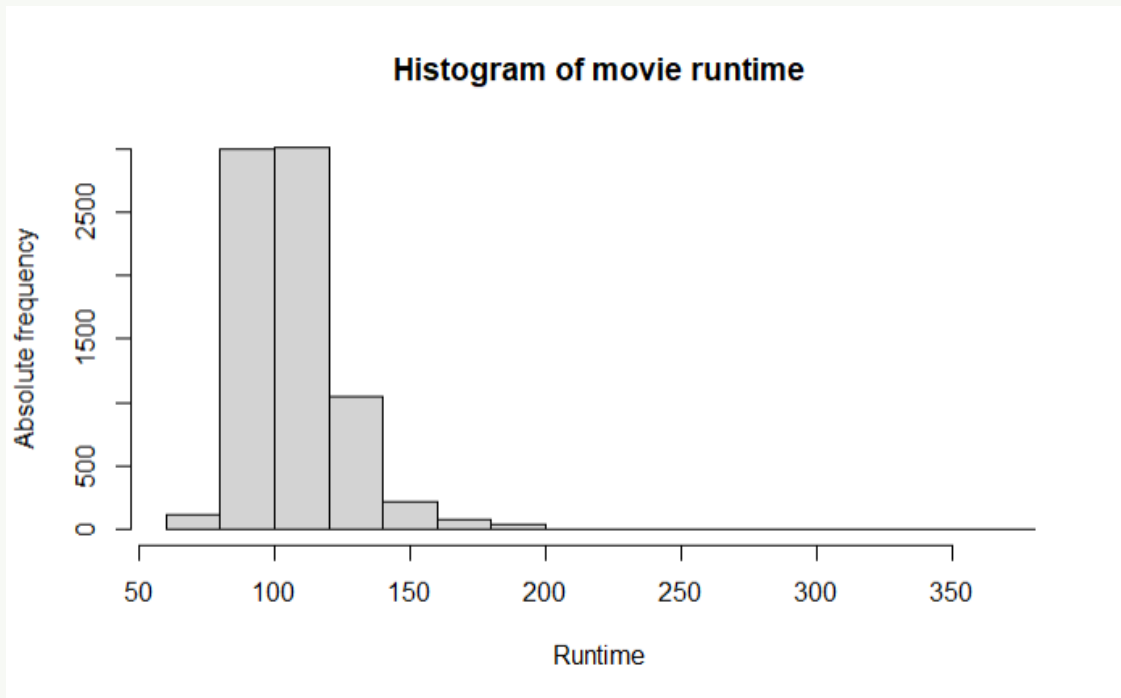
Movie join

Datasets were joined using columns
Name (Movie name)
and
Year (Release year).

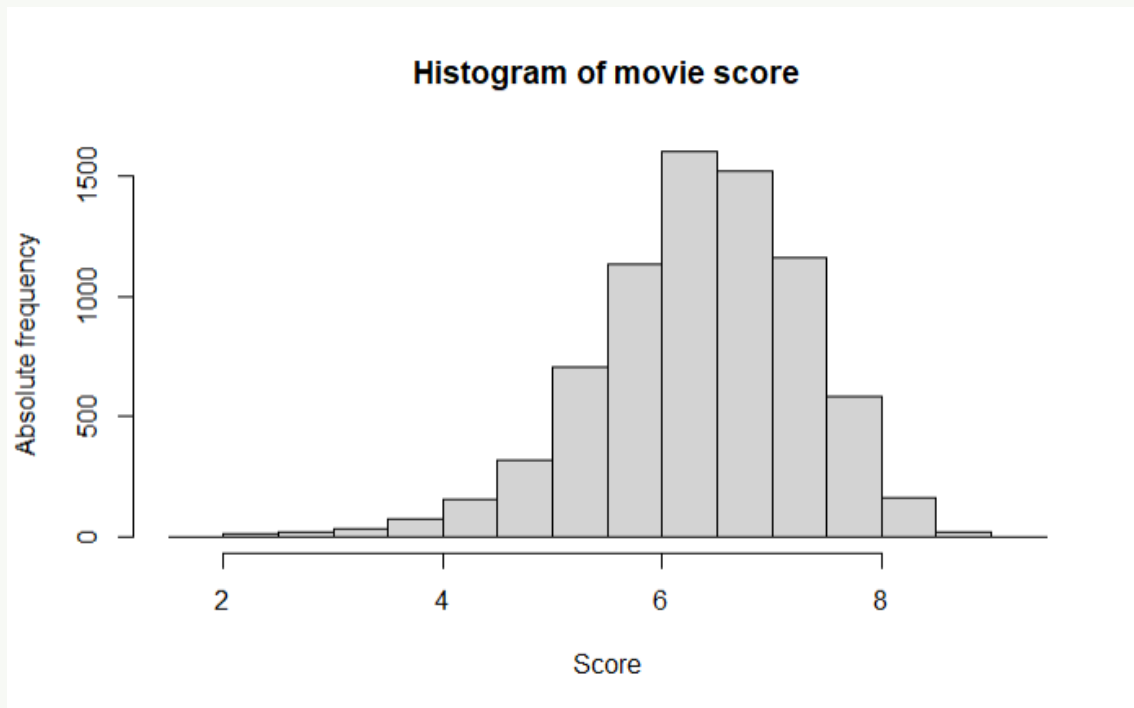2 008 movies were joined together.

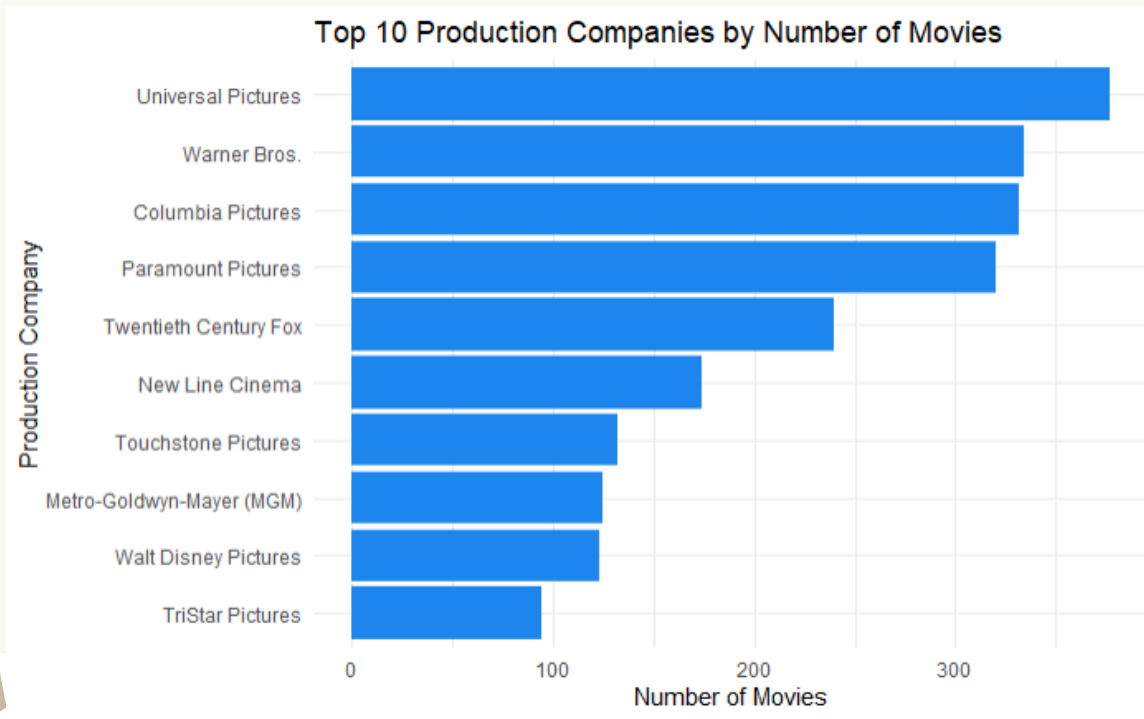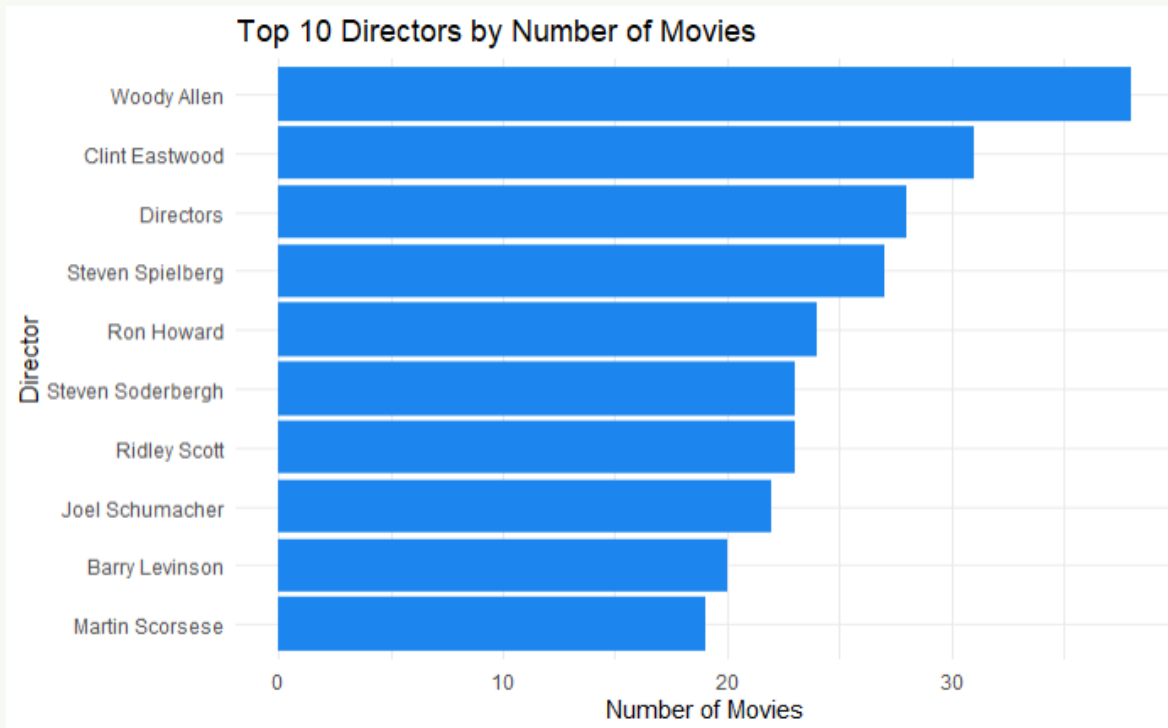# Movie runtime histogram

Histogram of movie runtime

Movie score histogram

Histogram of movie score

Production companies **in movies**

Top 10 Production Companies by Number of Movies

MOVIE TONIGHT

CINEMA

03

Directors in movies

Top 10 Directors by Number of Movies

Director	Number of Movies
Woody Allen	
Clint Eastwood	
Directors	
Steven Spielberg	
Ron Howard	
Steven Soderbergh	
Ridley Scott	
Joel Schumacher	
Barry Levinson	
Martin Scorsese	

MOVIE TONIGHT

CINEMA

03

Best picture by country

Top 6 countries by relative frequency of oscar wins for Best Picture

Winrate: 41 %

Winrate: 20,1 %

Relative frequency of wins

France   New Zealand   United Kingdom   United States

MOVIE TONIGHT

CINEMA
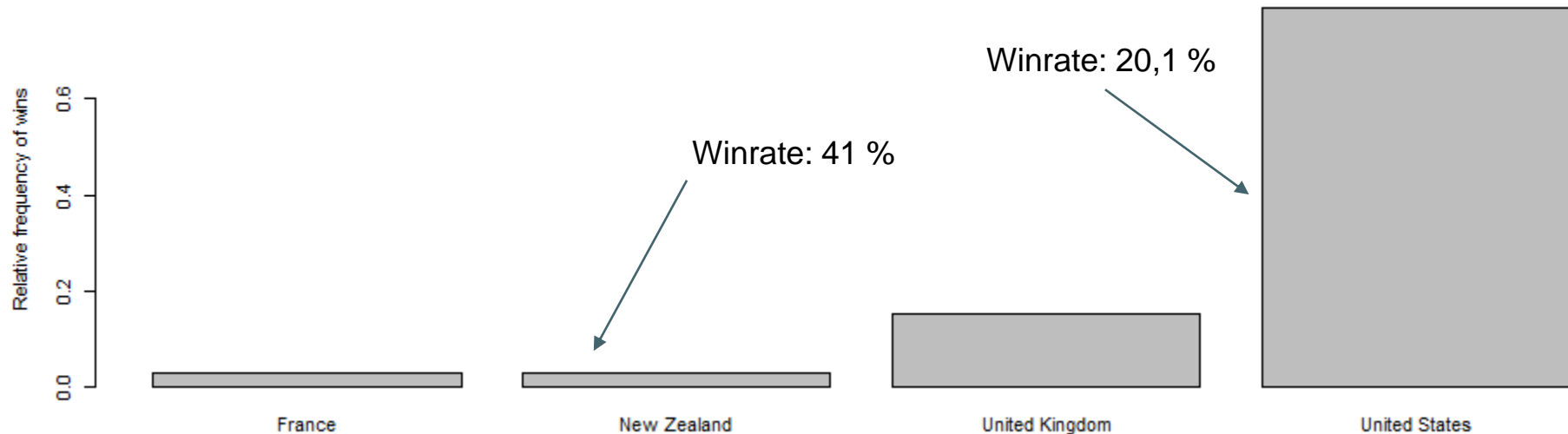
Best picture by company

Top 6 production companies by relative frequency of oscar wins for Best Picture

Inception, The Matrix, American Sniper
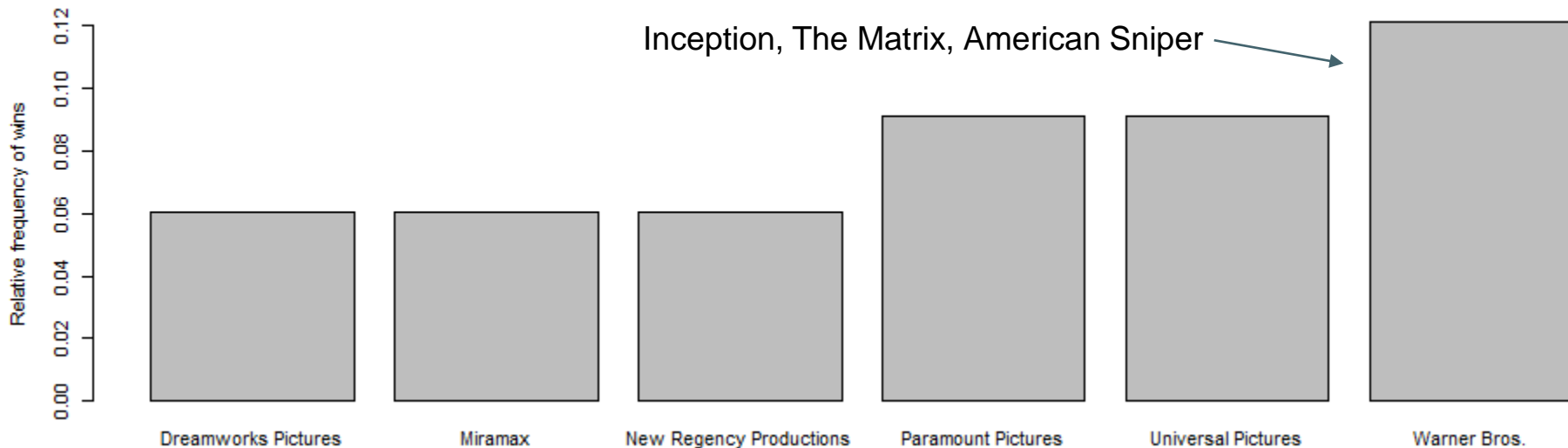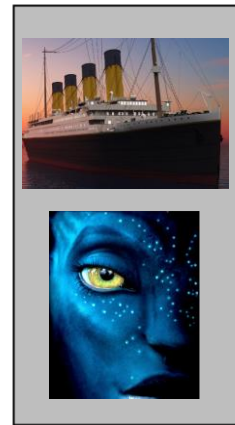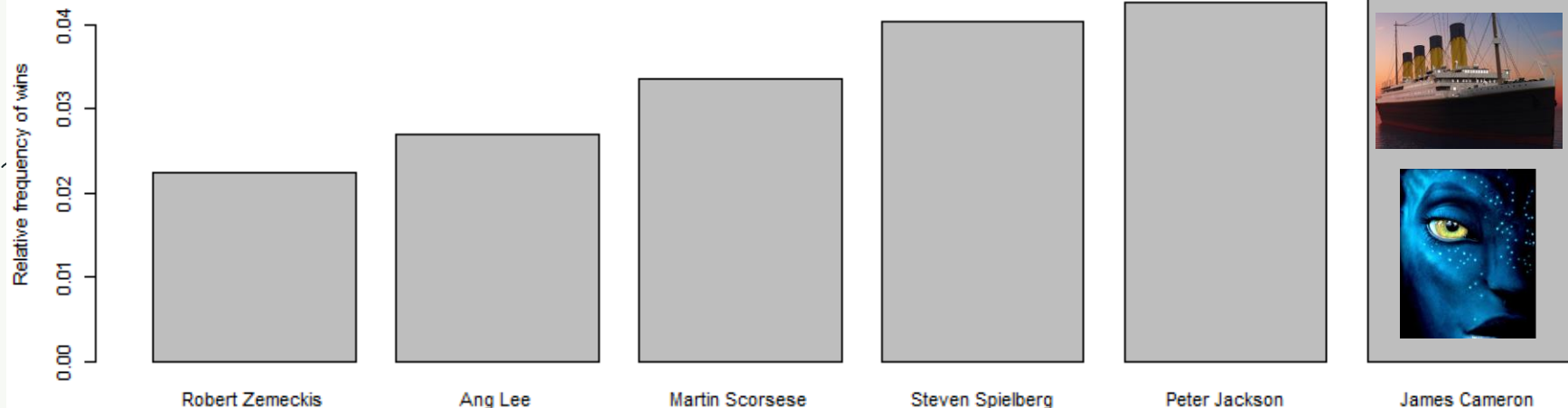
Directing by directors

Top 6 directors by relative frequency of oscar wins for Directing

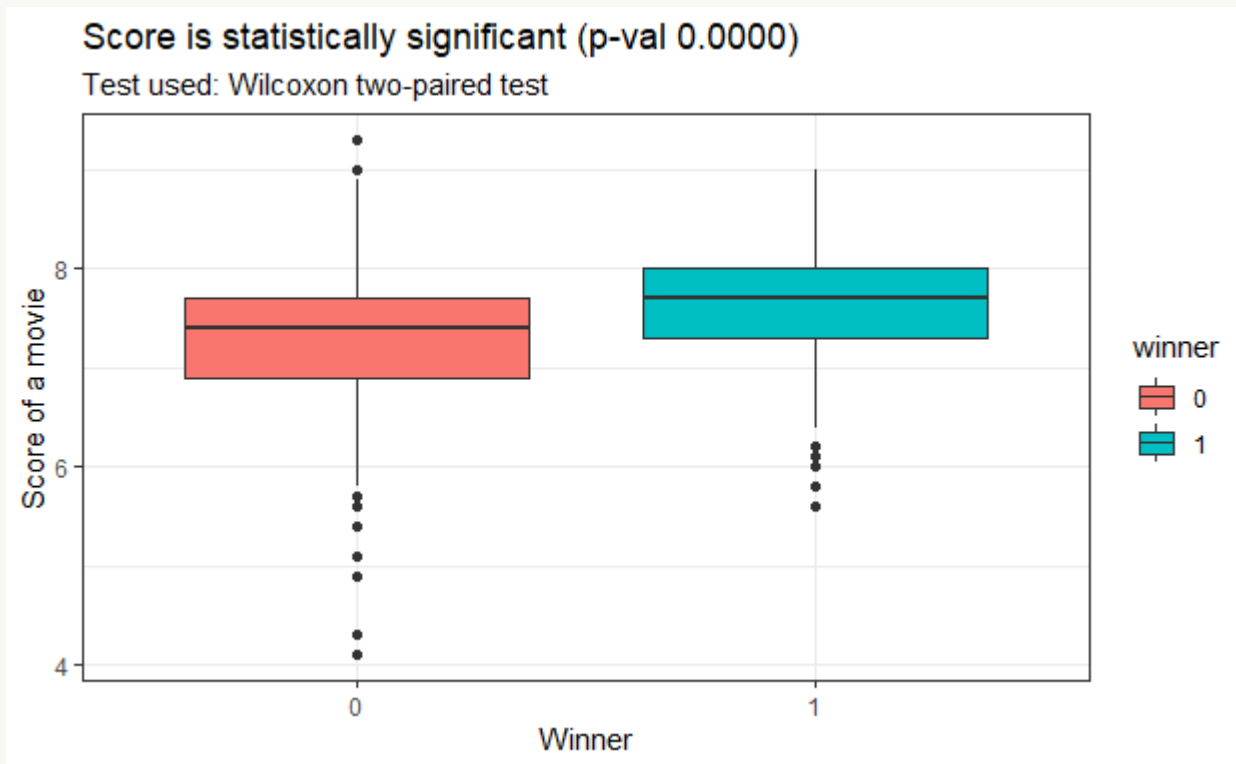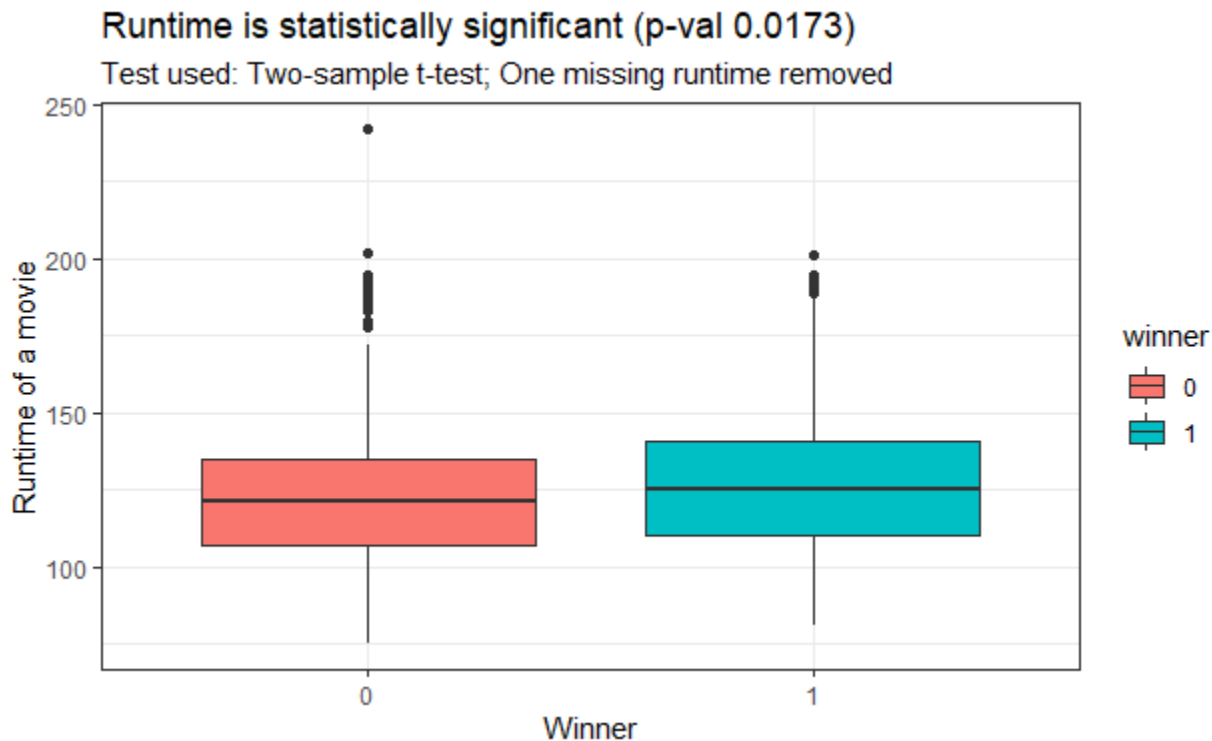# All categories by directors

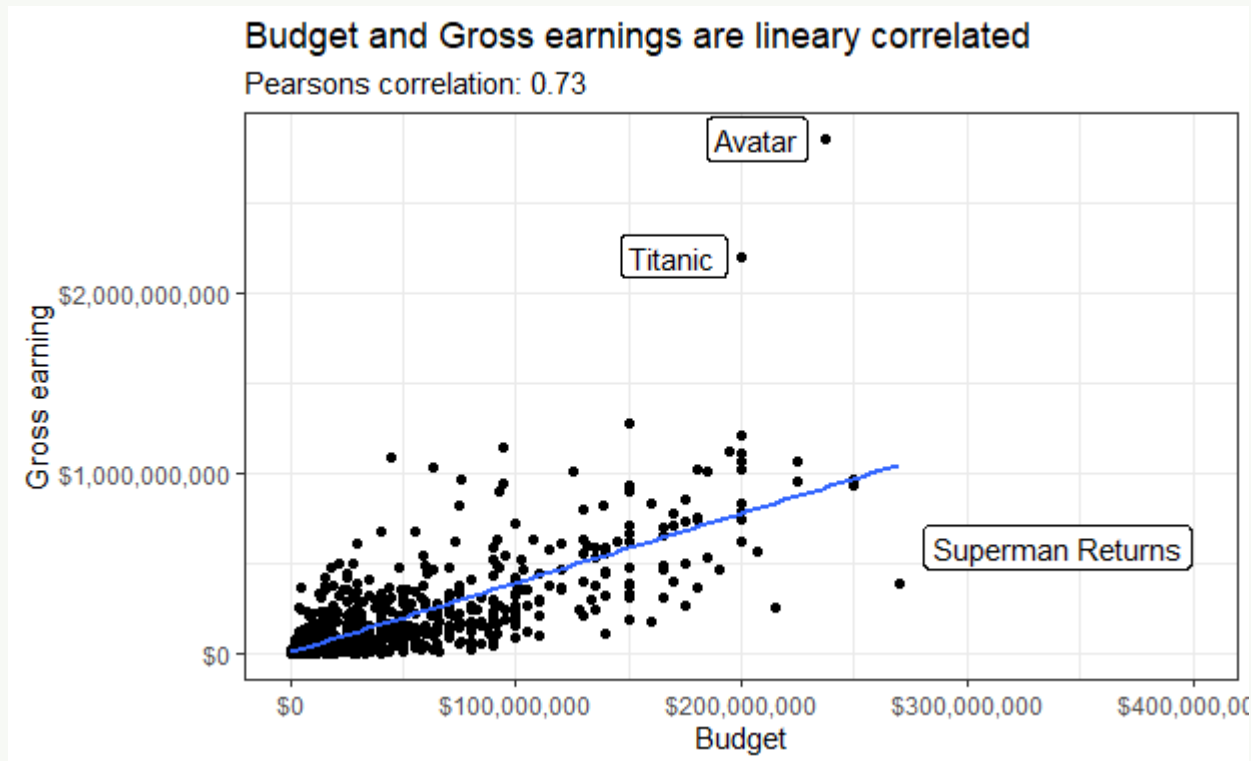Top 6 directors by relative frequency of oscar wins for All

Movie score by result

Score is statistically significant (p-val 0.0000)

Test used: Wilcoxon two-paired test

Movie runtime by result

Runtime is statistically significant (p-val 0.0173)

Test used: Two-sample t-test; One missing runtime removed

04

Budget and gross earnings

Budget and Gross earnings are lineary correlated

Pearsons correlation: 0.73

Four main conclusions

Country

United states

Directors

James Cameron

Movie company

Warner Bros

Budget vs. Gross

High positive correlation

Thanks!