

Dear Sir/ Madam:

Hello, we are group 18. Our members are :

- HE Qiu 57642787
- HO Yin Shan 57487297
- LYU Rui 57585819
- SI Wen 57322699
- SUN Weichen 57706020

As our presentation contains some interactive plots and animations,
The file needs to be rendered with some CSS settings and output in html.

Please view our file from Github Repository.

<https://github.com/ShaineHo/6335-presentation>

Thanks for your attention!

Yours truly,
Group 18



Studies of Bechdel Test

HE Qiu, HO Yin Shan, LYU Rui, SI Wen, SUN Weichen

2022-11-04

Presentation Contents

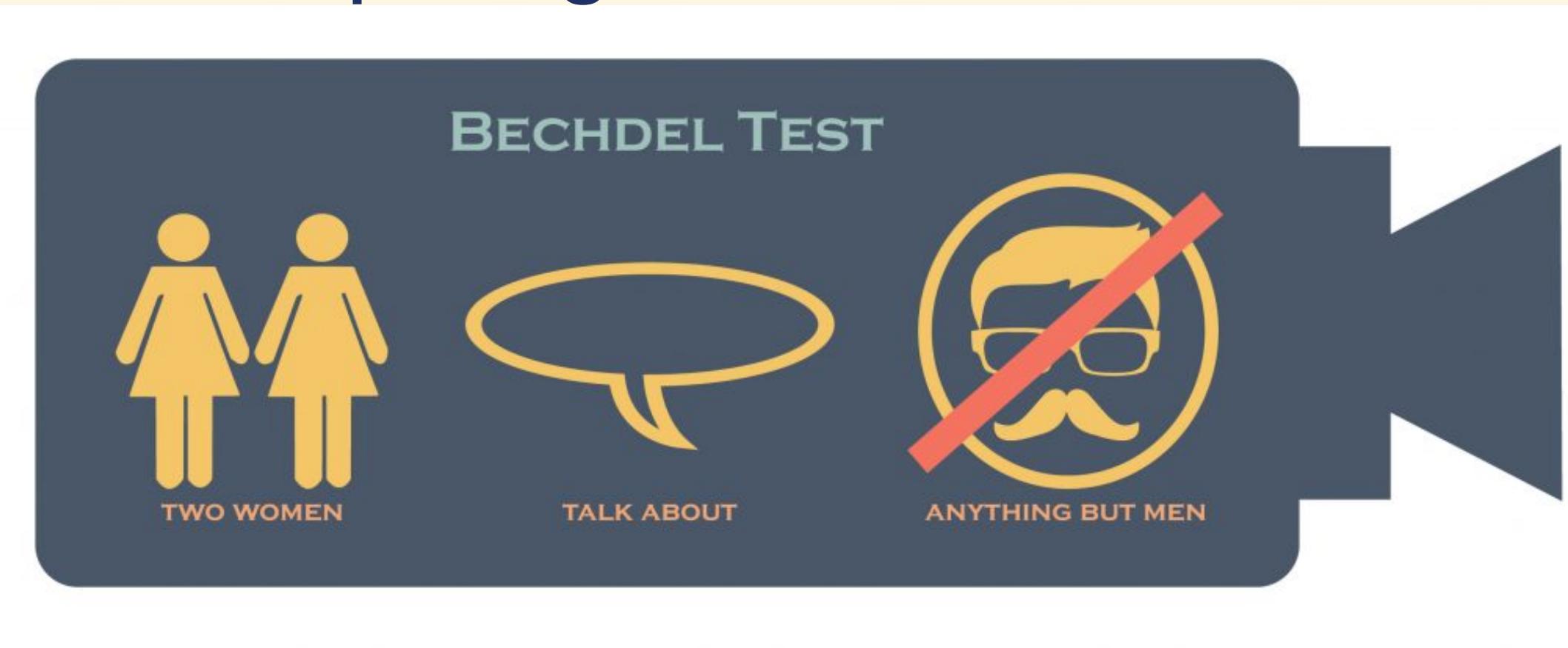
- Introduction
- Data Description
- Bechdel Results Analysis:
 1. IMDB Ratings
 2. Time Change
 3. Region
 4. Genre
 5. Runtime
 6. Budget & ROI
- Conclusion & Limitations
- References



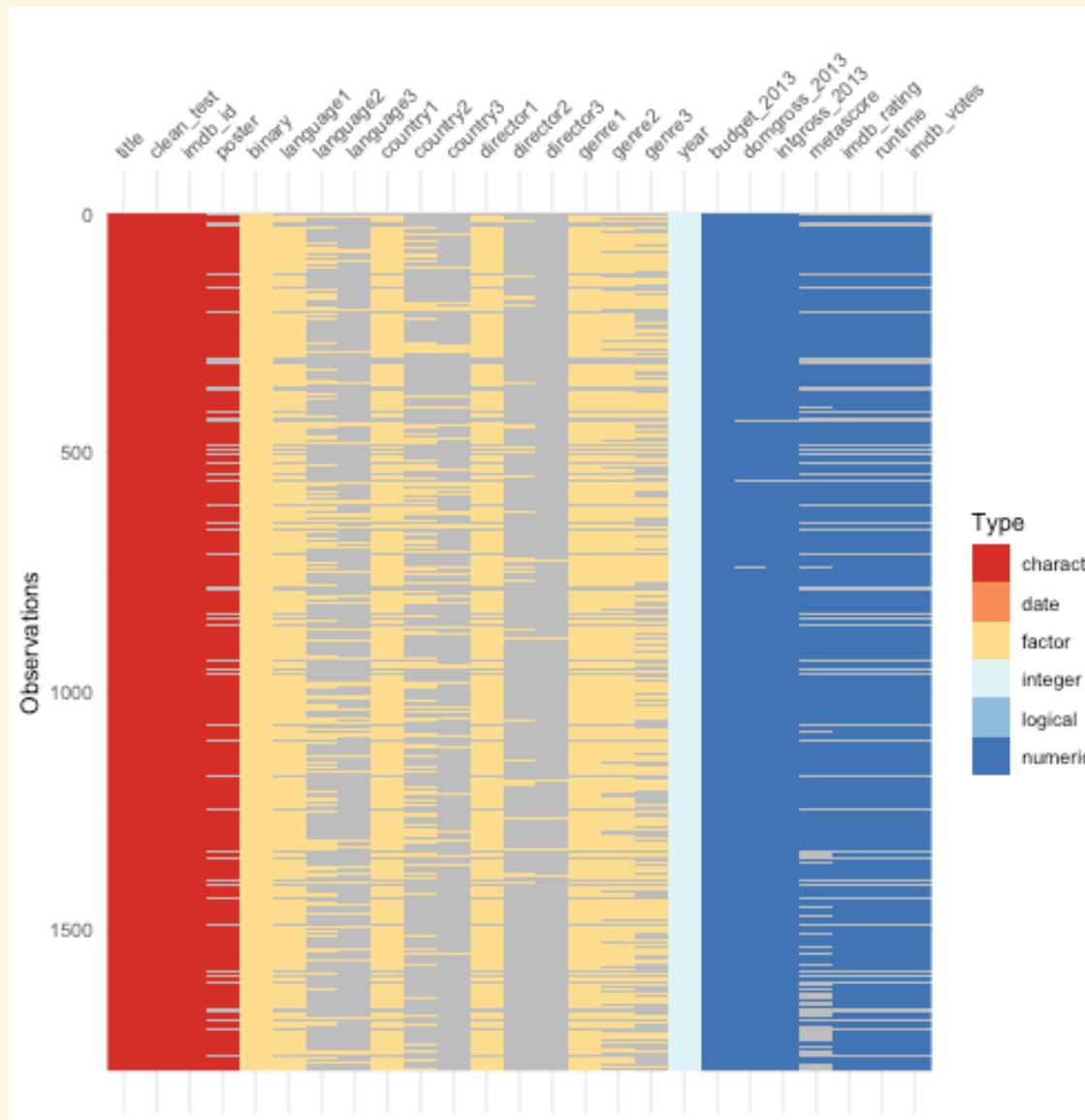
Introduction

- A technique to evaluate female participation in a movie
- Origin from a comic 'Dykes to Watch Out For' (1985)
- Named after the cartoonist Alison Bechdel
- Reflect gender bias in a movie

Criteria of passing the test



Data Overview



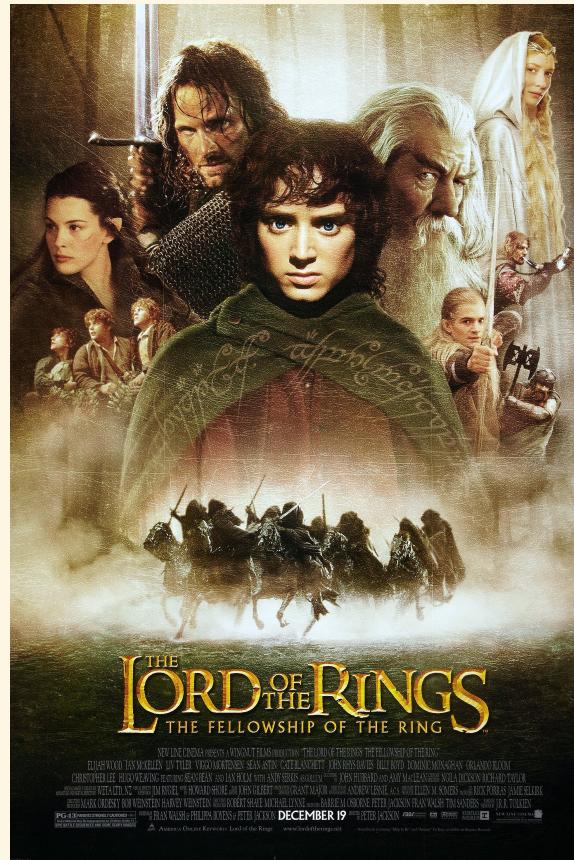
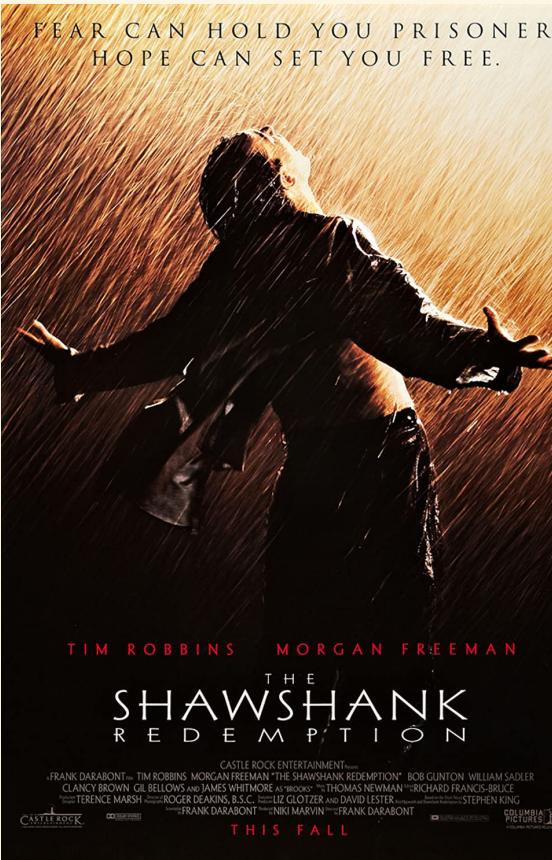
- Topic from [TidyTuesday](#)
- Data Source: [GitHub](#)
- **movies.csv** : 8839 x 5
- **raw_bechdel.csv** : 1794 x 34

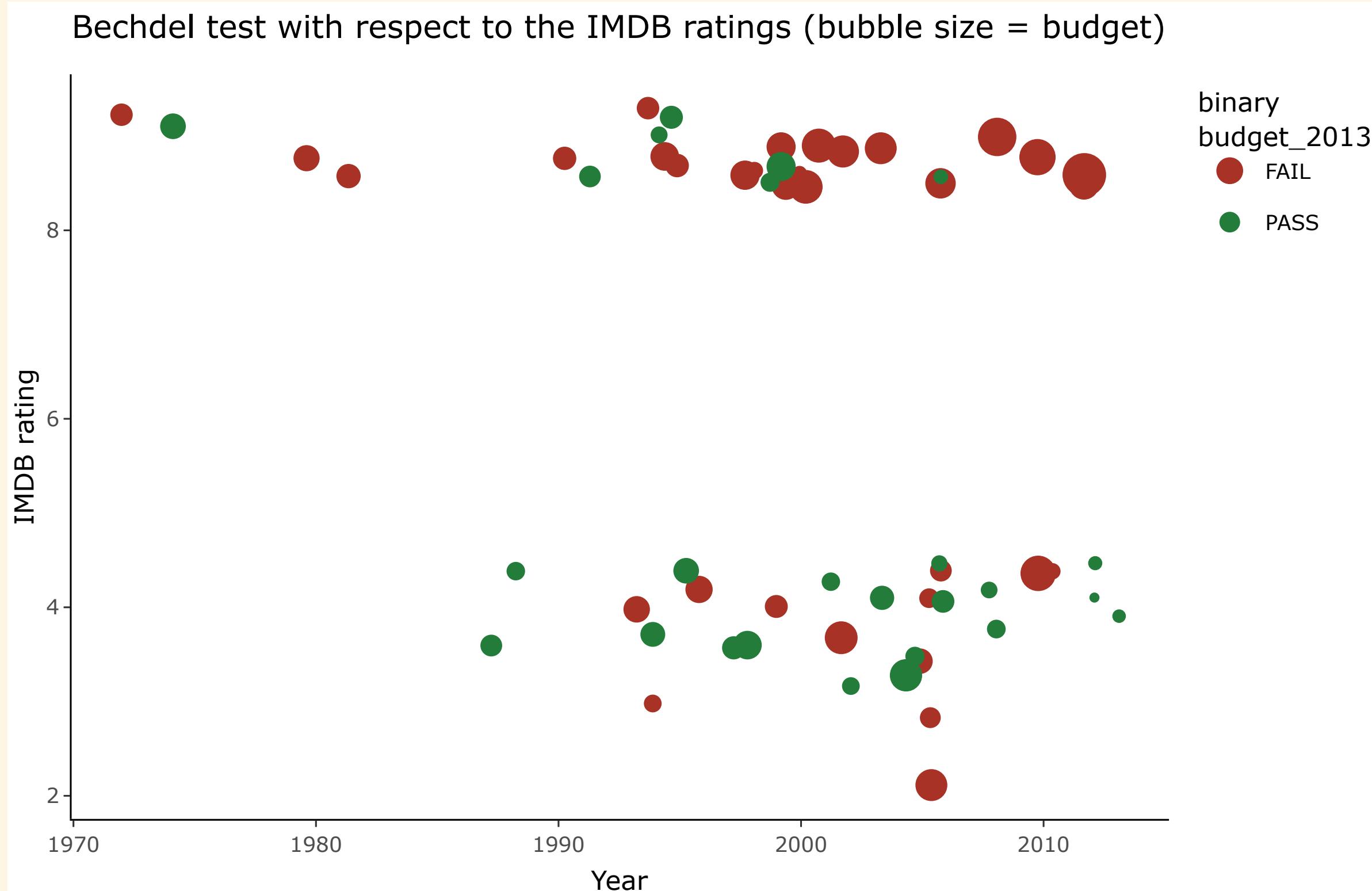
Data Dictionary

Variables	Description	Type
year	The year movie was released	Integer
title	The title of the movie.	Character
clean_test	Result of the Bechdel test with 5 levels	Factor
binary	Binary result of the Bechdel test	Factor
budget_2013	Inflation adjusted budget of the movie in 2013 dollars	Numeric
domgross_2013	Inflation adjusted domestic gross of the movie in 2013 dollars	Numeric
intgross_2013	Inflation adjusted international gross of the movie in 2013 dollars	Numeric
imdb_id	Unique ID for each	Character

Overview of the test

Analysis





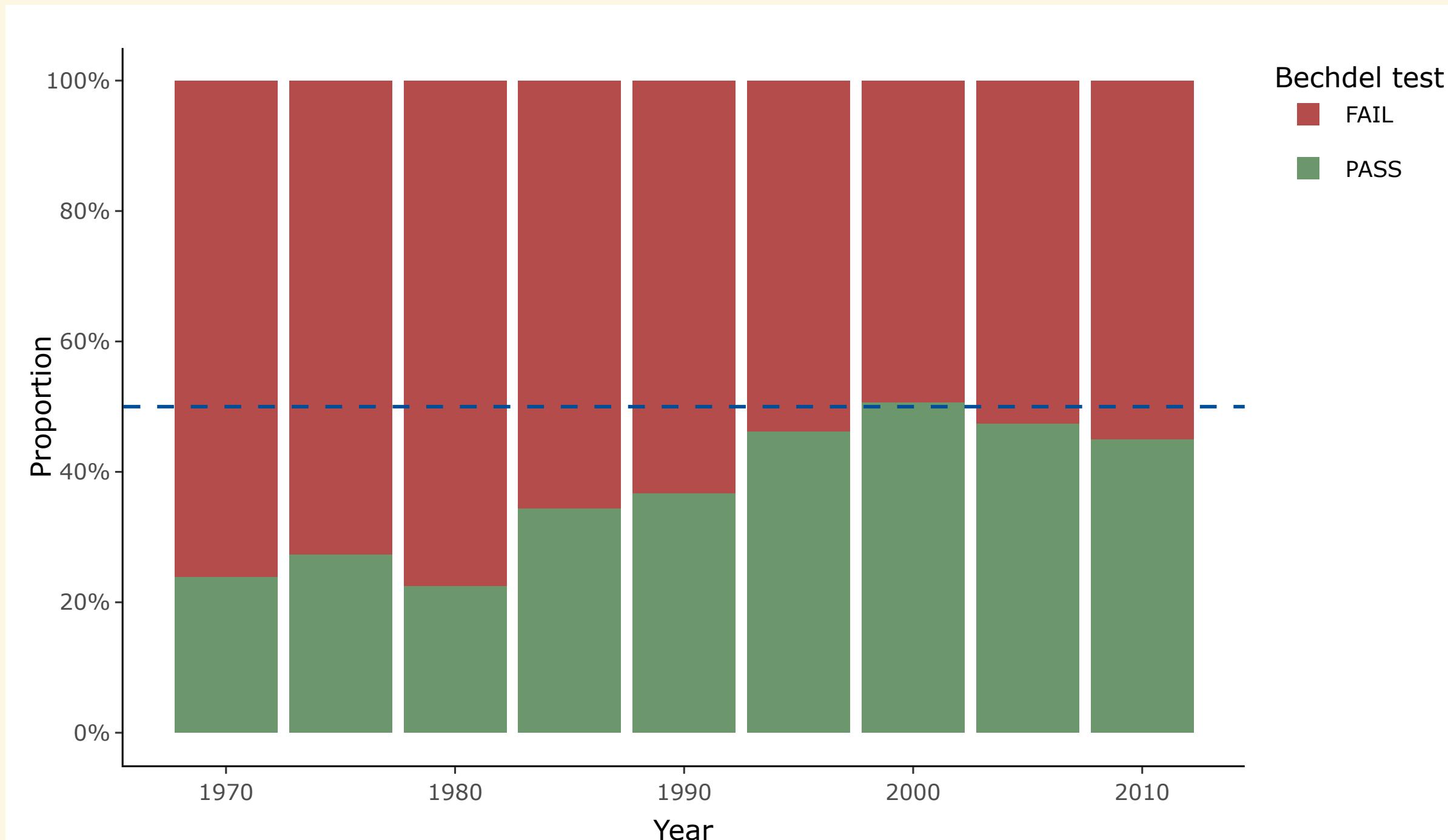
- First 30 and Last 30 in IMDB ratings were chosen

The trend of Bechdel Test Results

By percentage

By Number

Explanations

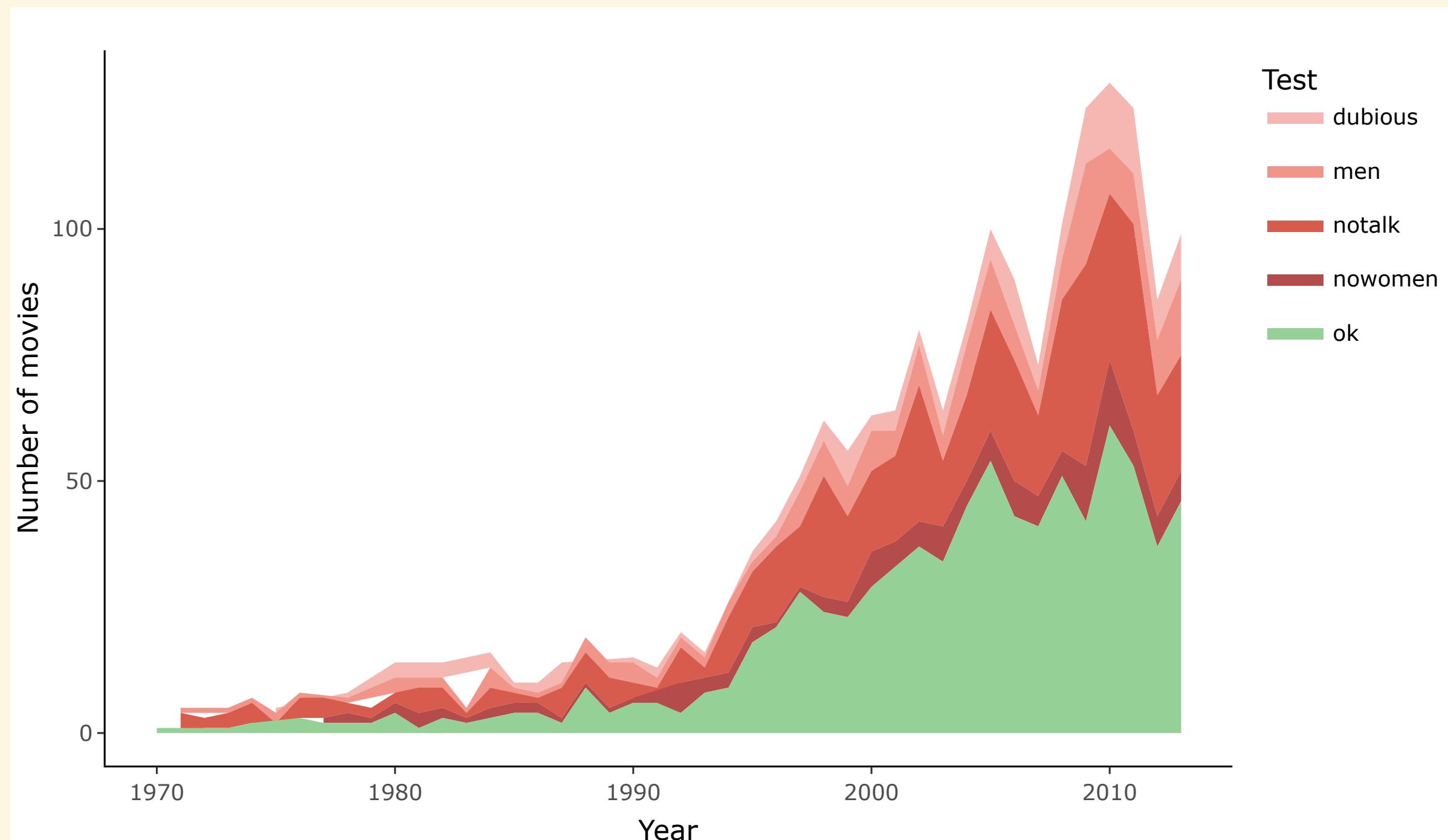


The trend of Bechdel Test Results

By percentage

By Number

Explanations



The trend of Bechdel Test Results

[By percentage](#)[By Number](#)[Explanations](#)

Result of the Bechdel test with 5 levels

- ***dubious***

the topic the conversation between the women in the picture was dubious

- ***men***

the women in the picture only talked about men

- ***notalk***

the women in the picture did not talk to each other

- ***nowomen***

there was less than 2 named women in the picture

- ***ok***

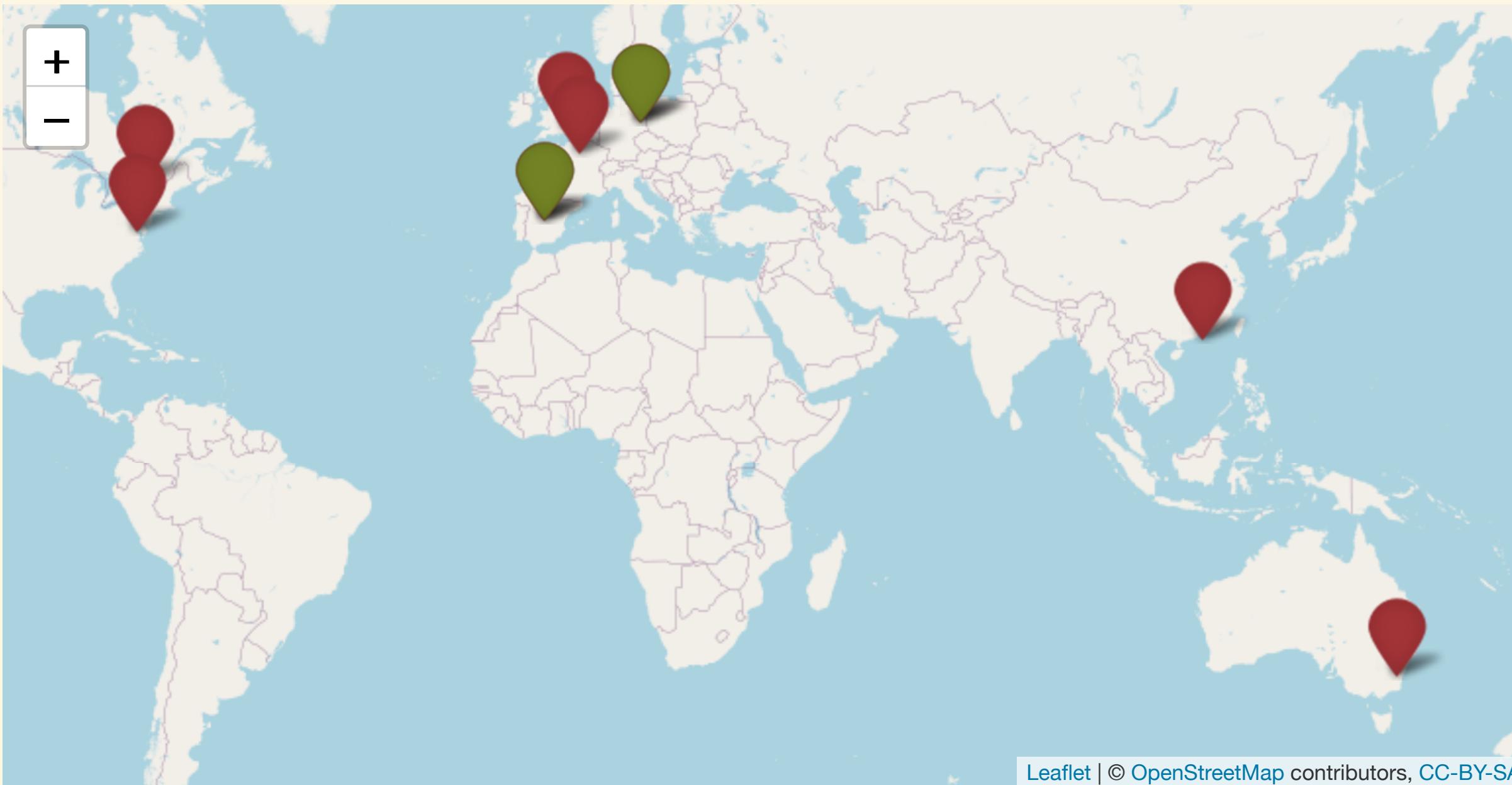
passed the test

Regions

Leaflet

Leaflet Analysis

Trend in USA



Regions

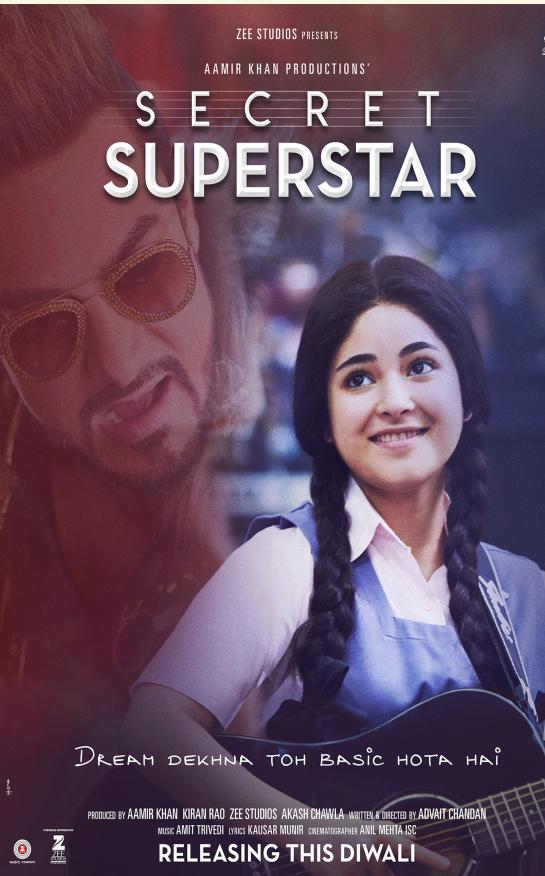
Leaflet

Leaflet Analysis

Trend in USA

- Include Regions with >5 data only

- Most of the countries failed the test
- 2 out of 3 Indian movies passed



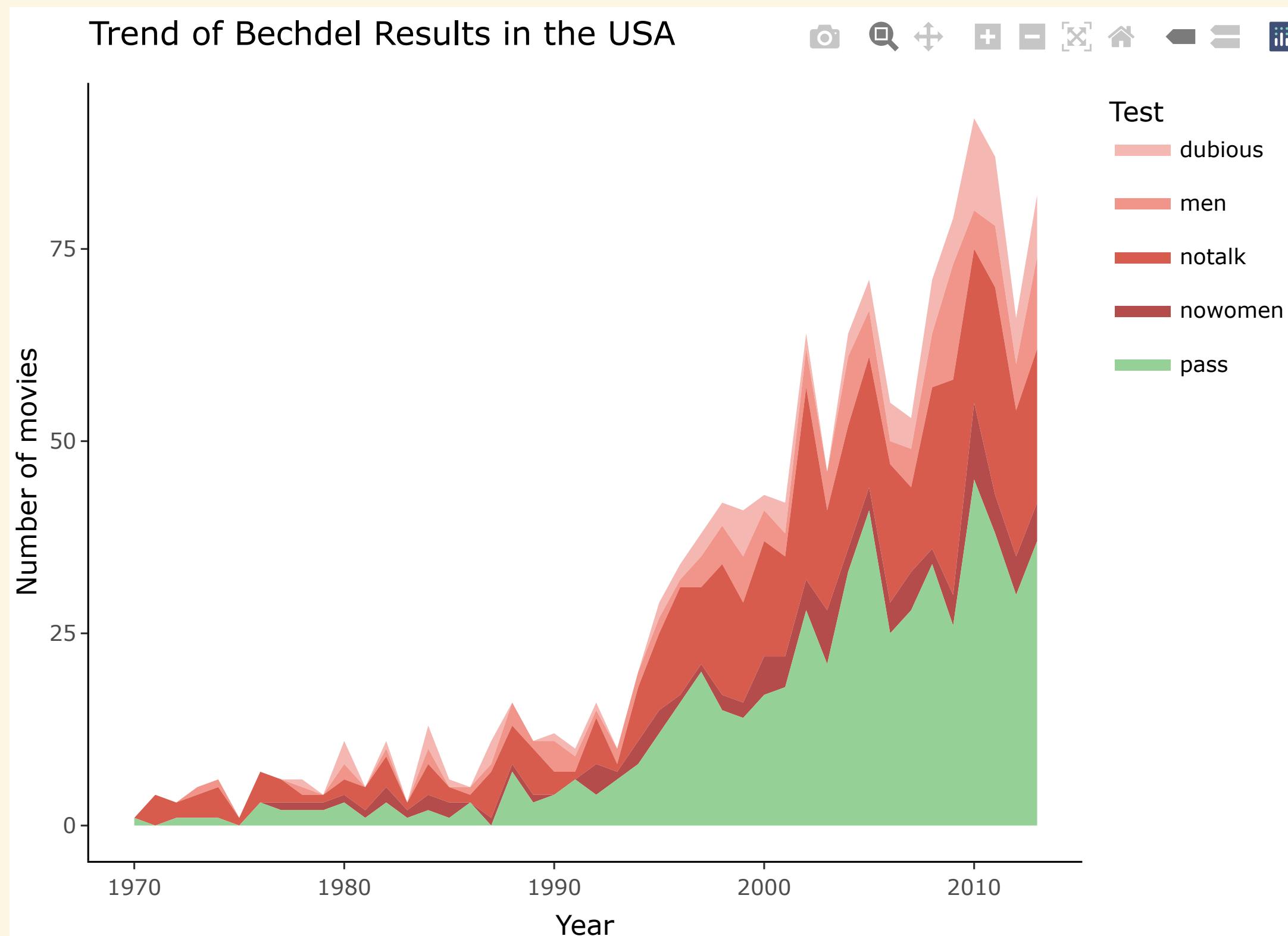
Location	FAILURE	PASS	PASSING RATE (%)
India	1	2	66.67
Spain	4	6	60.00
Germany	12	14	53.85
Finland	1	1	50.00
UK	76	71	48.30
Canada	12	11	47.83
France	17	14	45.16
USA	739	563	43.24
Australia	8	6	42.86
New Zealand	2	1	33.33
China	3	1	25.00
Ireland	3	1	25.00
Hong Kong	4	1	20.00

Regions

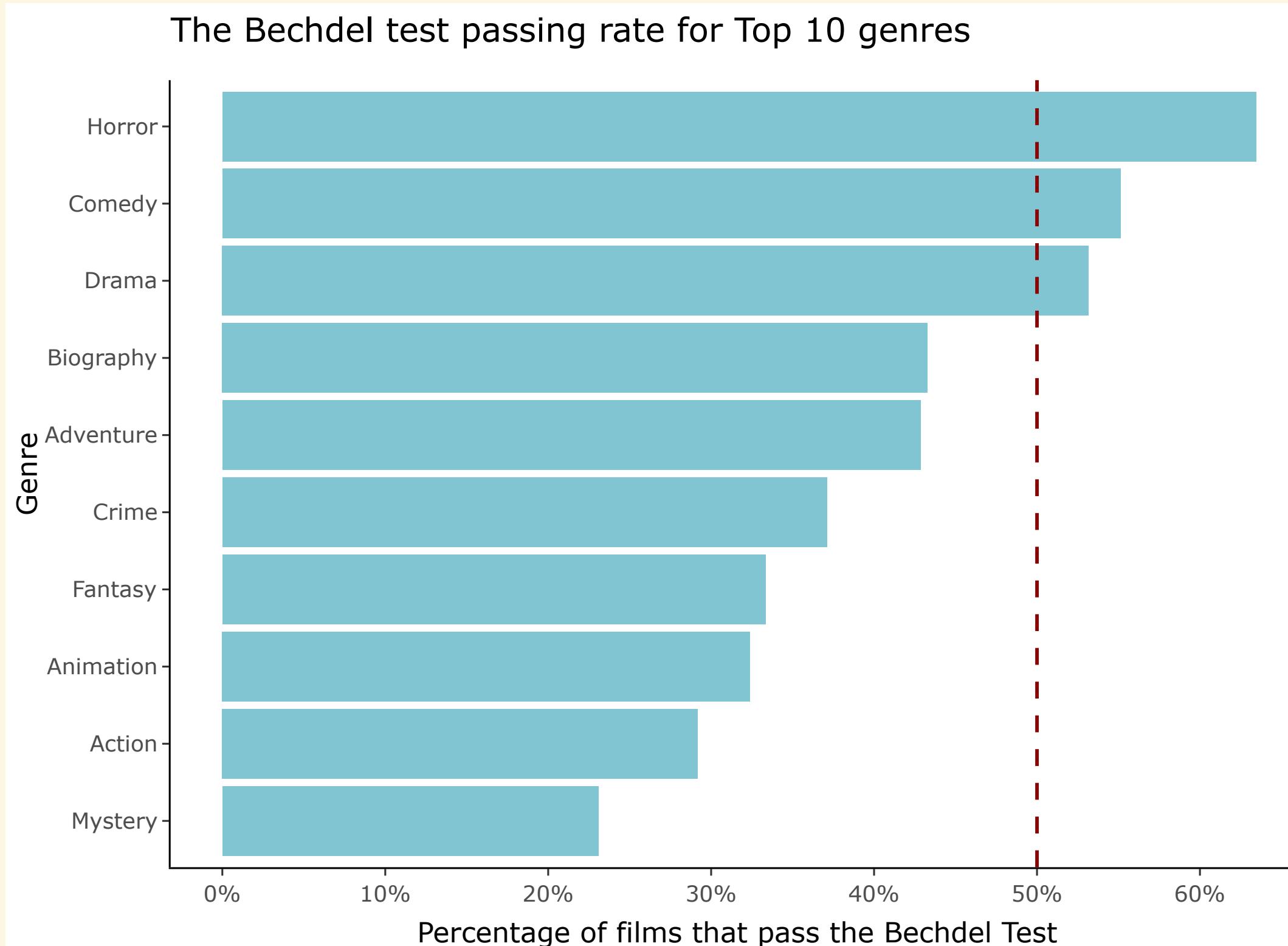
Leaflet

Leaflet Analysis

Trend in USA

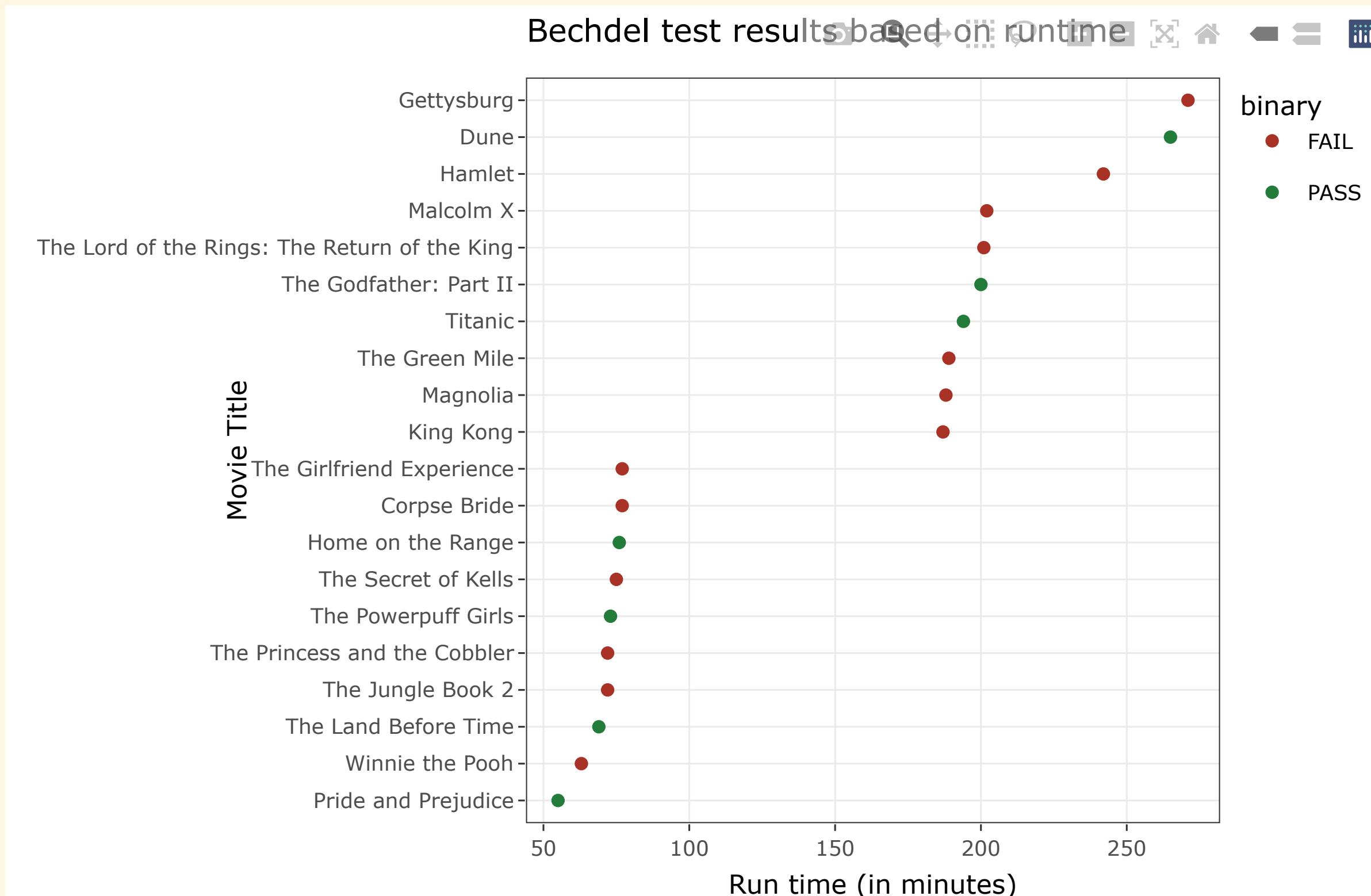


Genre

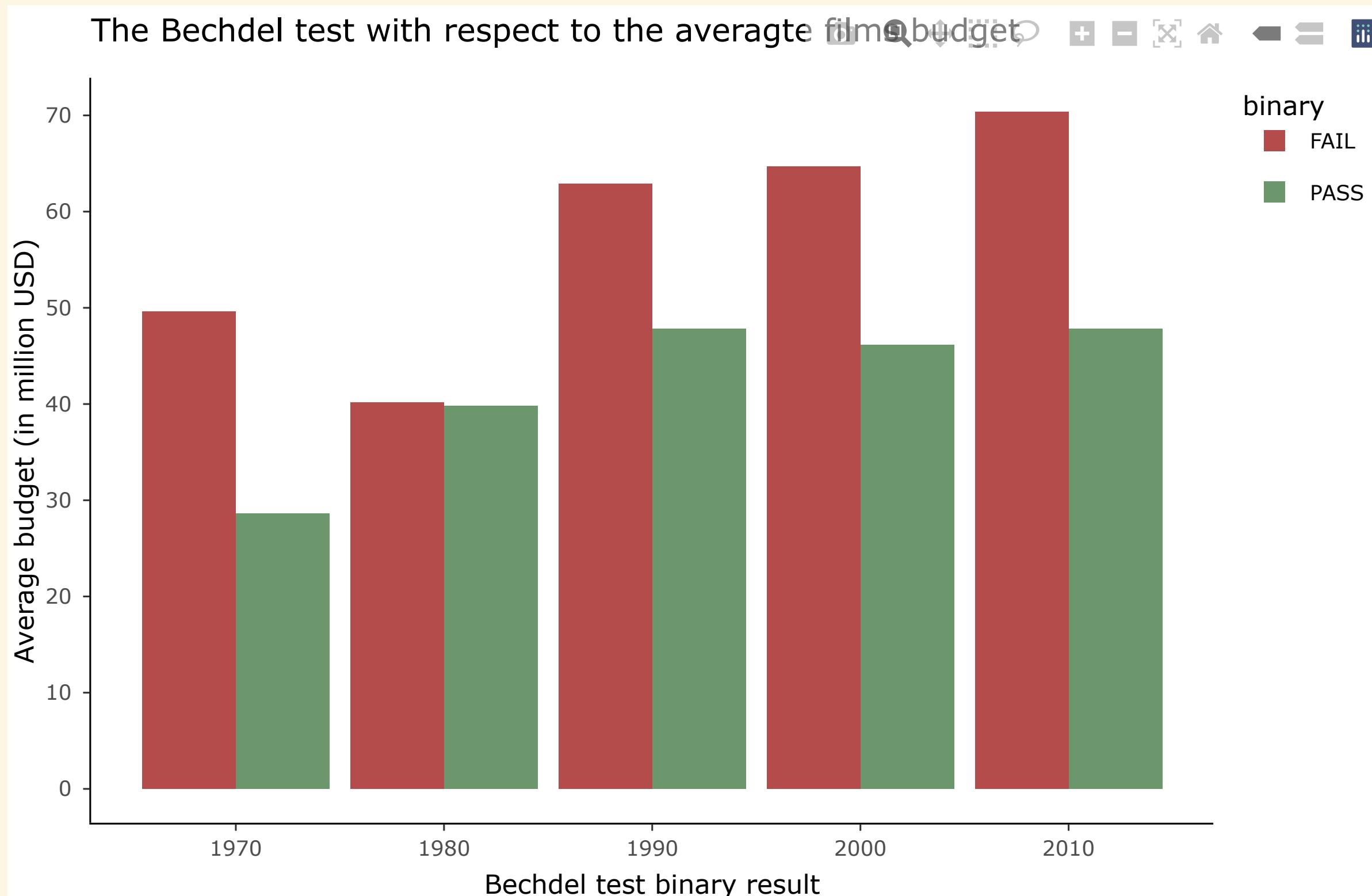


- Highest passing rate in **Horror movies**
- Only **horror, drama** and **comedy** genre pass

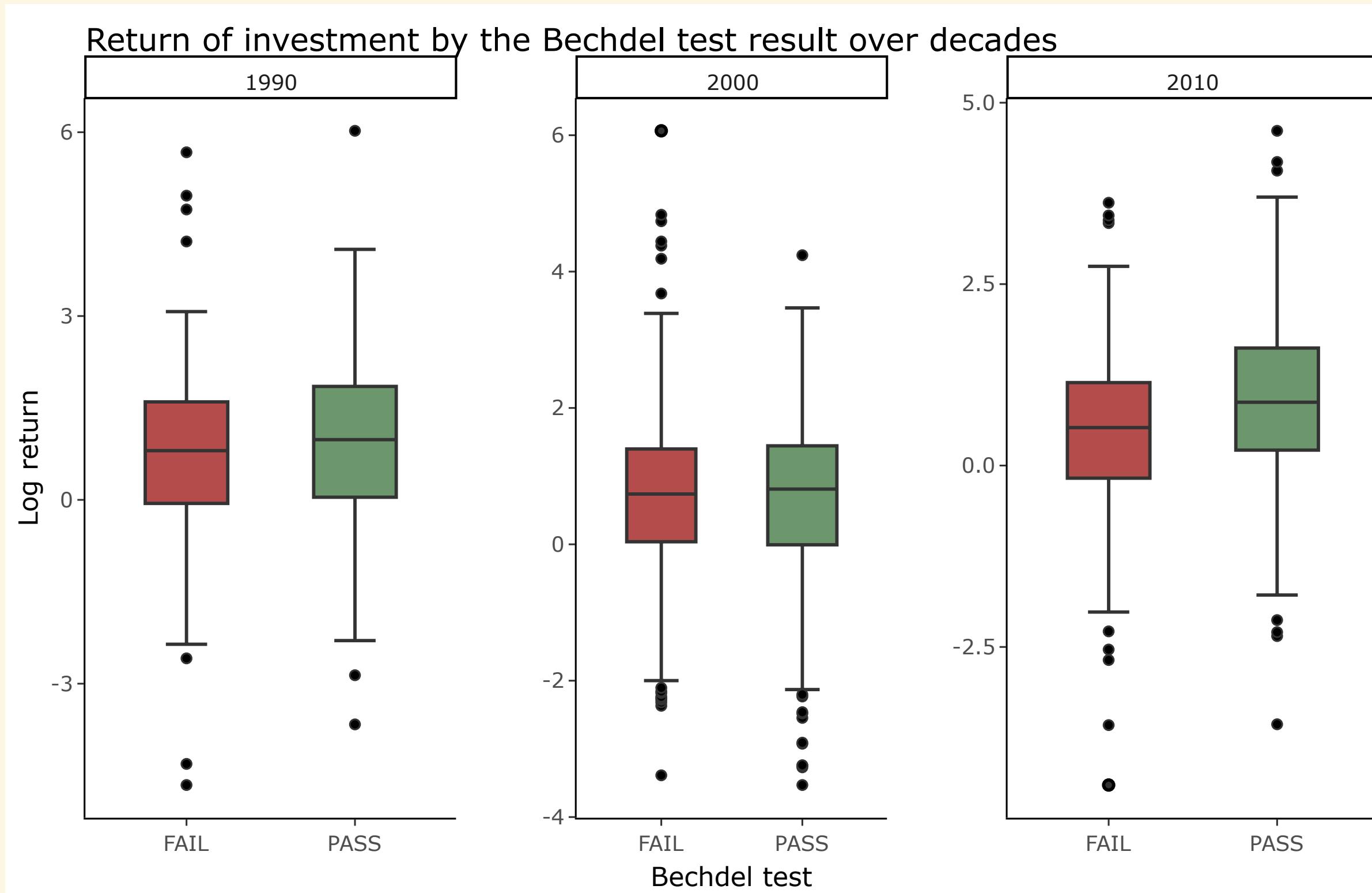
Runtime



- The test results **not likely affected by runtime**



- More investments in **male dominant movies**



- Movies **passed** slightly outperformed the **failed**

Conclusion & Limitations

1. More female participation in the movie but still low passing rate
2. Less gender biased in German and Spanish movies
3. Relatively higher female participation in the USA, but only supporting role(No conversations between them)
4. Female plays relatively important role in horror, comedy and drama genre
5. Results of the test are not affected by runtime
6. Investors of Movie industry invested more in the male dominant movies
7. Movies with female participation tend to have relatively higher profit
8. Data coverage is not enough



References

- Thomas Mock (2021). Tidy Tuesday: A weekly data project aimed at the R ecosystem. <https://github.com/rfordatascience/tidytuesday>.
- Walt Hickey (Apr 1, 2014). FiveThirtyEight: The Dollar-And-Cents Case Against Hollywood's Exclusion of Women. <https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/>
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- Antoine Fabri (2020). unglue: Extract Matched Substrings Using a Pattern. R package version 0.1.0. <https://CRAN.R-project.org/package=unglue>
- Garrett Grolemund, Hadley Wickham (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1-25. URL <https://www.jstatsoft.org/v40/i03/>.
- Hao Zhu (2021). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4. <https://CRAN.R-project.org/package=kableExtra>
- Elin Waring, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis (2021). skimr: Compact and Flexible Summaries of Data. R package version 2.1.3. <https://CRAN.R-project.org/package=skimr>

- Nicholas Tierney, Di Cook, Miles McBain and Colin Fay (2021). naniar: Data Structures, Summaries, and Visualisations for Missing Data. R package version 0.6.1. <https://CRAN.R-project.org/package=naniar>
- Thomas Lin Pedersen (2020). patchwork: The Composer of Plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>
- Stefan Milton Bache and Hadley Wickham (2020). magrittr: A Forward-Pipe Operator for R. R package version 2.0.1. <https://CRAN.R-project.org/package=magrittr>
- Winston Chang, (2014). extrafont: Tools for using fonts. R package version 0.17. <https://CRAN.R-project.org/package=extrafont>
- Hadley Wickham (2021). tidyr: Tidy Messy Data. R package version 1.1.3. <https://CRAN.R-project.org/package=tidyr>
- Claus O. Wilke (2020). ggtext: Improved Text Rendering Support for 'ggplot2'. R package version 0.1.1. <https://CRAN.R-project.org/package=ggtext>
- Thomas Lin Pedersen (2021). ggforce: Accelerating 'ggplot2'. R package version 0.3.3. <https://CRAN.R-project.org/package=ggforce>
- C. Sievert. Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida, 2020.
- Thomas Lin Pedersen (2020). patchwork: The Composer of Plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>
- Data downloaded from <https://bechdeltest.com>

