Design & Data final project
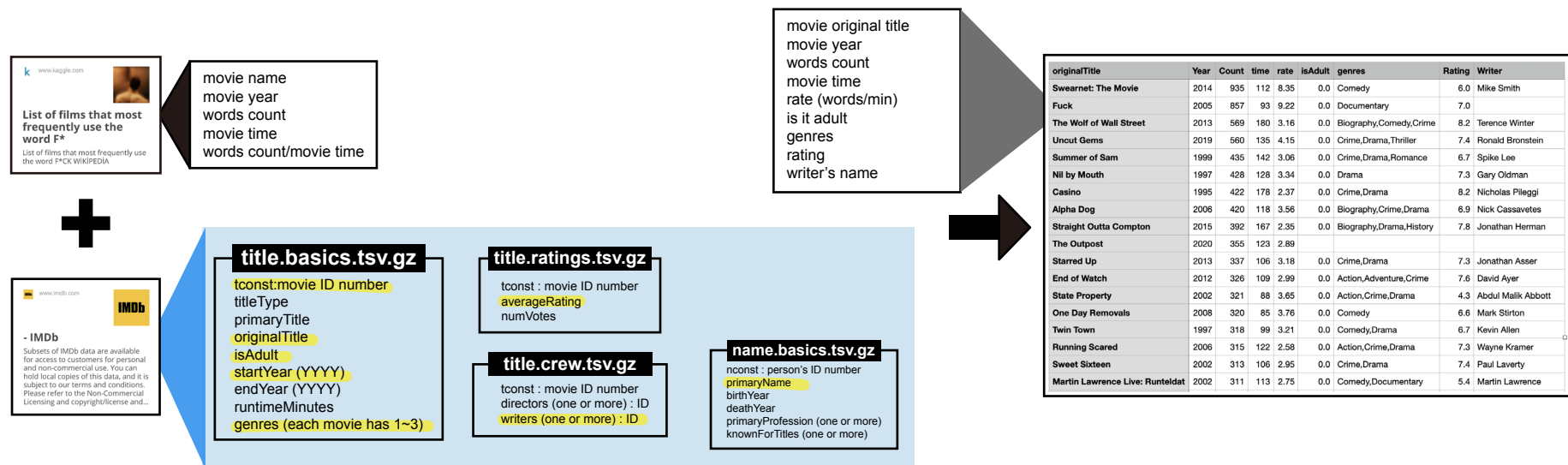
Yuhan Tseng

How Many Fxcks in a Movie?

The main purpose of this project is to practice the techniques I have learned in this course.

About data

At first, I found a dataset that records around 100+ movies and how many FxCK words are used in the movie. I decided to use this as the main data for my final project. To find the relation between word count and the movie's other information, I also use some datasets from IMDb. Because the datasets from IMDb are very big and Python has some convenient libraries for data organizing, I combined all these datasets into one final dataset in Python before making the visualization part on the website.
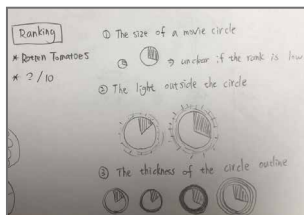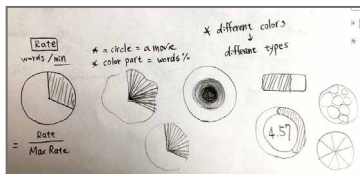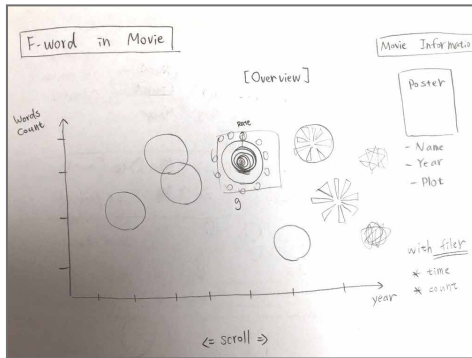

www.kaggle.com

List of films that most frequently use the word F*

List of films that most frequently use the word F*CK WIKIPEDIA

movie name
movie year
words count
movie time
words count/movie time

+

www.imdb.com   IMDb

- IMDb

Subsets of IMDb data are available for access to customers for personal and non-commercial use. You can hold local copies of this data, and it is subject to our terms and conditions. Please refer to the Non-Commercial Licensing and copyright/license and...

title.basics.tsv.gz
tconst:movie ID number
titleType
primaryTitle
originalTitle
isAdult
startYear (YYYY)
endYear (YYYY)
runtimeMinutes
genres (each movie has 1~3)

title.ratings.tsv.gz
tconst : movie ID number
averageRating
numVotes

title.crew.tsv.gz
tconst : movie ID number
directors (one or more) : ID
writers (one or more) : ID

name.basics.tsv.gz
nconst : person's ID number
primaryName
birthYear
deathYear
primaryProfession (one or more)
knownForTitles (one or more)

movie original title
movie year
words count
movie time
rate (words/min)
is it adult
genres
rating
writer's name

originalTitle	Year	Count	time	rate	isAdult	genres	Rating	Writer
Swearnet: The Movie	2014	935	112	8.35	0.0	Comedy	6.0	Mike Smith
Fuck	2005	857	93	9.22	0.0	Documentary	7.0	
The Wolf of Wall Street	2013	569	180	3.16	0.0	Biography,Comedy,Crime	8.2	Terence Winter
Uncut Gems	2019	560	135	4.15	0.0	Crime,Drama,Thriller	7.4	Ronald Bronstein
Summer of Sam	1999	435	142	3.06	0.0	Crime,Drama,Romance	6.7	Spike Lee
Nil by Mouth	1997	428	128	3.34	0.0	Drama	7.3	Gary Oldman
Casino	1995	422	178	2.37	0.0	Crime,Drama	8.2	Nicholas Pileggi
Alpha Dog	2006	420	118	3.56	0.0	Biography,Crime,Drama	6.9	Nick Cassavetes
Straight Outta Compton	2015	392	167	2.35	0.0	Biography,Drama,History	7.8	Jonathan Herman
The Outpost	2020	355	123	2.89				
Starred Up	2013	337	106	3.18	0.0	Crime,Drama	7.3	Jonathan Asser
End of Watch	2012	326	109	2.99	0.0	Action,Adventure,Crime	7.6	David Ayer
State Property	2002	321	88	3.65	0.0	Action,Crime,Drama	4.3	Abdul Malik Abbott
One Day Removals	2008	320	85	3.76	0.0	Comedy	6.6	Mark Stirton
Twin Town	1997	318	99	3.21	0.0	Comedy,Drama	6.7	Kevin Allen
Running Scared	2006	315	122	2.58	0.0	Action,Crime,Drama	7.3	Wayne Kramer
Sweet Sixteen	2002	313	106	2.95	0.0	Crime,Drama	7.4	Paul Laverty
Martin Lawrence Live: Runteldat	2002	311	113	2.75	0.0	Comedy,Documentary	5.4	Martin Lawrence

p.s. Some movies have more than one writer, I am not familiar with these writers so I only keep the first one.
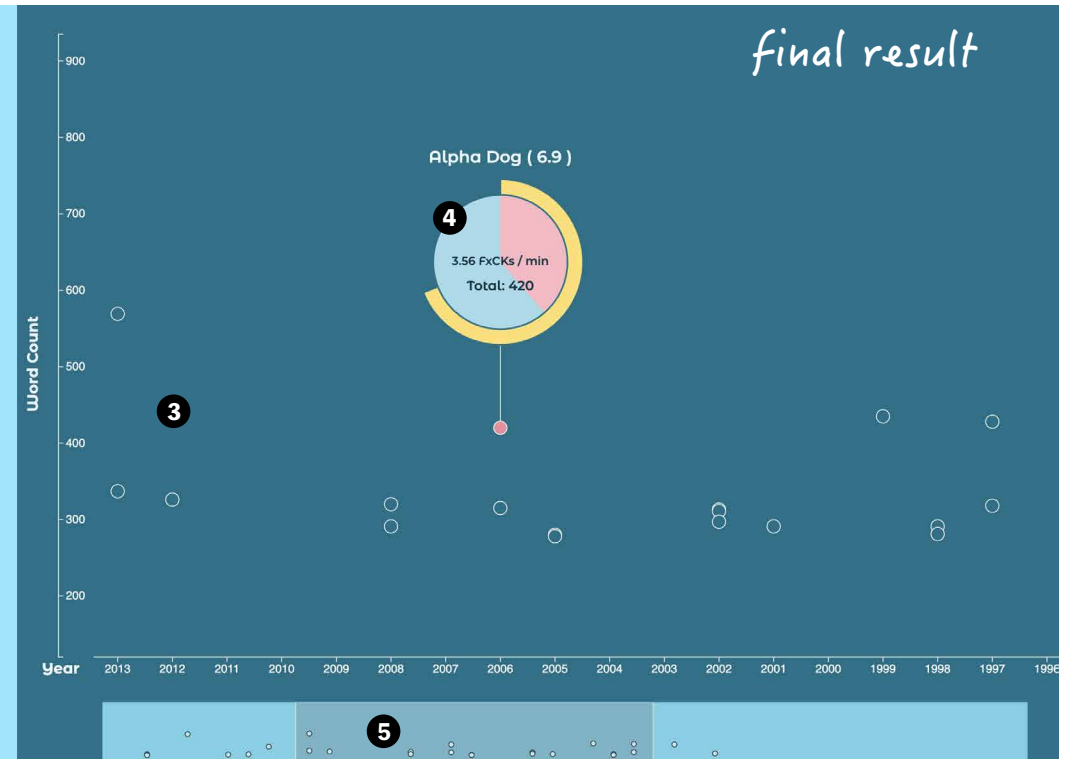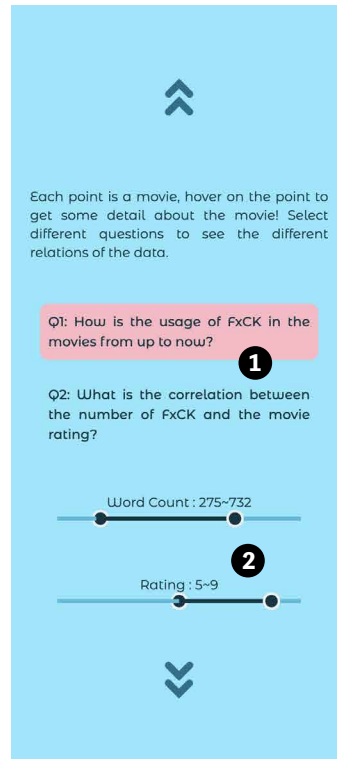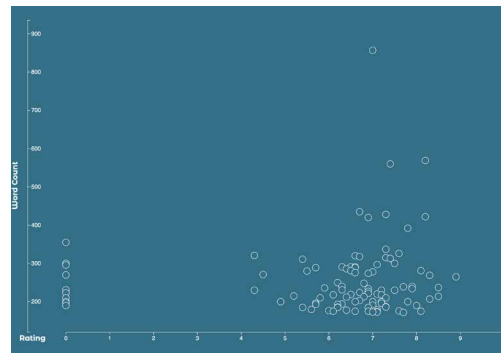
About Visual

This part is an overall view of all the movies up to now, trying to show the distribution in the timeline and how many words they used. Also, including some information about each movie.

I am also curious about whether is there any correlation between word count and the movie rating, so I used questions as buttons to switch the x-aix.

initial sketches

the chart for 2nd question

final result

Each point is a movie, hover on the point to get some detail about the movie! Select different questions to see the different relations of the data.

Q1: How is the usage of FxCK in the movies from up to now?

Q2: What is the correlation between the number of FxCK and the movie rating?

Word Count : 275~732

Rating : 5~9

Alpha Dog (6.9)
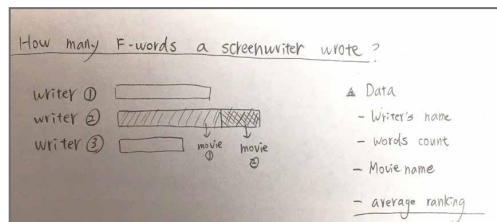
3.56 FxCKs / min

Total: 420

1. Two questions are buttons, select one and change the x-axis and show the corresponding scatter plot.
2. Slider to filter the showing data.
3. The scatter plot of the movies. Each point is a movie. Hover it to will show 4.
4. The inner circle represents one minute and the pink part is the percentage of the word counts. The outer yellow ring is its rating (from 0 - 10).
5. A preview window to see the whole scatter plot, helping users to know how long the chart is. Dragging the gray handle or scrolling on the chart can move left and right.

About Visual

Another part I am curious about is how many FxCK words a writer has written. I assume that all the lines spoken by the actors are the words that the writer wrote in the script.

Also, I use genres as a filter to see what type of genres uses more FxCK.

initial sketches

Each bar is a movie, hover on the bar to get some detail about the movie! Use the checkboxes below to filter out movies by different genres.

How many FxCKs a writer has written?

☐ Comedy ☐ Biography ☐ Crime
☑ Drama ☐ Thriller ☑ Romance
☑ History ☑ Action ☑ Adventure
☑ Documentary ☑ Mystery ☑ War
☑ Music ☑ Horror ☑ Western
☑ Fantasy ☑ Sci-Fi ☑ Musical
☑ Animation ☑ All Genres

133

Original Title : Summer of Sam
Genere : Crime,Drama,Romance
Runing Time : 142 min
Year : 1999
Rating : 6.7

final result

Chart data:

Writer	Count
Mike Smith	935
Terence Winter	185
Ronald Bronstein	720
Spike Lee	675
Gary Oldman	428
Nicholas Pileggi	422
Nick Cassavetes	420
Jonathan Herman	392
Jonathan Asser	337
David Ayer	326
Abdul Malik Abbott	321
Mark Stirton	320
Kevin Allen	318
Wayne Kramer	315
Paul Laverty	763
Martin Lawrence	311
Allen Hughes	457
Joe Carnahan	588
Eddie Little	291
Jon Favreau	291
Sam Levinson	449
David Gordon Green	289

1. Writer's name
2. The total amount of the words
3. Bar chart of the word count. Each single bar (rectangle) is one movie.
4. A preview window to see the whole bar chart, helping users to know how long the chart is. Dragging the gray handle or scrolling on the chart can move up and down.
5. Filter the chart by genres. The number below are the total number of the movies after filtering.
6. Hovering on each bar can see the movie name and its word count on the bar, other additional information will show here.