Project Progress Report 2

(due May 24th 11:59p.m)

You can start working on the project once your report is accepted and graded by your TA. The entire final project is worth **35%** of your final grade and this report accounts for **10%**. This project is done individually.

Submission Guideline

Download this google doc, fill the table. **Type** your answers, no handwritten answers will be accepted (except for the very last question). Submit it in **PDF** format on Gradescope.

If you need some inspirations please feel free to take a look at:
Showcase of Information is Beautiful Awards
Bloomberg Year In Graphics Review
The Pudding
The New York Times

Project Guidelines

Note: The guideline has been further clarified from Progress Report 1, so double-check whether your dataset choice still satisfies the updated guideline below.

1. You may use more than one dataset, however, regardless if you use one or multiple datasets, your visualizations must make use of at least three following data types - **link, position, and attribute.**
2. You cannot use any dataset from the class (Labs, Assignments, Lecture Exercises)
3. You can make your own dataset (Web scrape etc.) provided point 1. is satisfied.

Part 1 - Story and Narrative

Link to the dataset	https://www.kaggle.com/datasets/jrobischon/wikipedia-movie-plots https://www.kaggle.com/datasets/thedevastator/rotten-tomatoes-top-movies-ratings-and-technical
Example item from the dataset	title: Mad Max: Fury Road year: 2015 critic_score: 97 people_score: 86.0 type: Action & Adventure genre: adventure, action rating: R (Intense Sequences of Violence|Disturbing Im... original_language: English box_office_(gross_usa): $153.6M runtime: 2h director: George Miller producer: Doug Mitchell, George Miller, P.J. Voeten writer: George Miller, Brendan McCarthy, Nick Lathouris crew: Tom Hardy, Charlize Theron, Nicholas Hoult, Hu... original_language: English Origin/Ethnicity: American
Story you want to deliver	(a story should be in a form of a list of facts, insights, and messages - refer to the lecture slide) Facts: People are estimated to watch 3,600+ movies in their lifetime. Hollywood contributes $504 billion to US GDP. (reference) Insight: It's important to understand what elements make a good movie. Messages: Critics and common audiences in general have common taste. People tend to give movies of type "Classic", "Drama", "Mystery and Suspense" higher rotten tomatoes score. Movies from "Japan", "Hong Kong", and "Bollywood" have the highest scores. Some directors like to collaborate with some actors, and their movies have higher scores.
Describe your target audience.	My target is the general audience who wants to know what elements indicate a movie might be attractive. This can be producers who want to understand more about the movie market, and can also be common movie audiences who wanna pick a good movie. Since the data is about movies, most of the audience should be familiar with it and attracted by the nature of this dataset. I wish the audience can take away what elements like genre and crew make a movie attractive(high-score and high box office). I don't expect my audience to have any background in visualization or

	mathematics, so the plots should be intuitive, entertaining and with detailed explanations. The visualization should be convenient to view on mobile phones and computers, so I can apply some transitions compared to printed media.
The goal of your project outcome. And why?	Explanatory. Because this project is mainly made for general audiences, I want my visualization to clearly convey my findings in an explanatory way.
Narrative structure you plan to use	Interactive Slideshow
Elaborate your choice of narrative structure.	I want to clearly present my findings, so the project would be more author driven. And since I will go over many elements regarding movies, the project will be heavy messaging. Considering these two characters, an interactive slideshow is a more appropriate narrative structure.
Narrative genre you plan to use	Magazine Style
Elaborate your choice of narrative genre.	My visualization will go over several elements that affect the score and box office of movies, and I will need to provide a detailed explanation with the graphs. Since my targets are common audiences, I would like to make the explanation as clearly and as detailed as possible. In this case, the magazine style is a good fit for my project.

Part 2 - Outline

Story you want to deliver	(you can copy/paste from Part 1) Facts: People are estimated to watch 3,600+ movies in their lifetime. Hollywood contributes $504 billion to US GDP. (reference) Insight: It's important to understand what elements make a good movie. Messages: People tend to give movies of type "Classic",

	“Drama”, “Mystery and Suspense” higher rotten tomatoes score. Movies from “Japan”, “Hong Kong”, and “Bollywood” have the highest scores. Some directors like to collaborate with some actors, and their movies have higher scores.
Specifications on each plot **in the order of how you lay out on your project**	(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story) 1) Scatter plot: Scores given by common audiences and critics 1. Task: This scatter plot a) analyzes the trend between Critics Score and People’s Score and b) located outliers(Critics give high scores but people give low scores). 2. Attribute: People Score, Critic Score 3. Marks: points(tomato/splash) 4. Channels: a. Aligned vertical position channel for People Score b. Aligned horizontal channel for Critic Score c. Color and shape channel for both 5. My visualization aims to deliver elements of a good movie. This plot gives an in-depth understanding of a “good” movie: whether a movie is loved by critics or the general audience. ~~2) Bar plot: Mean Scores of movies from different types~~ ~~1. Task: This bar plot analyzes the score among movies of different types~~ ~~2. Attribute: Movie Type, Difference between Mean Score and Overall Mean~~ ~~3. Marks: area~~ ~~4. Channels:~~ ~~a. Size of bars as channel for Score~~ ~~b. Vertical position channel for score (indicate below or above overall mean)~~ ~~5. My visualization aims to deliver elements of a good movie. This plot gives specific insights on the effect of movie types.~~ 3) Geometric plot: Mean Scores of movies from various Origin/Ethnicity 1. Task: This plot analyzes the score among movies of different origins/ethnicities 2. Attribute: Origin/Ethnicity, Mean Score 3. Marks: Area 4. Channels: Color indicates the mean score of the movie from the origin. 5. My visualization aims to deliver elements of a good movie. This plot gives specific insights on the effect of origin. 4) Network plot: Network of Directors and Actors 1. Task: This plot a) analyzes the mean score of directors and crews, and b) display the network between them 2. Attribute: Director, Crew, Mean Score 3. Marks: Point(Photo of director/actor/actress), Line

	4. Channels: a. Vertical position of director/actor/actress is channel for mean score b. Width of line connection them is channel for how many times they collaborated 5. My visualization aims to deliver elements of a good movie. This plot gives specific insights on the effect of directors and crews.
Elaborate the choice of their marks and channels for each vis	Plot 1: Since our score is collected from rotten tomatoes, the mark would be tomatoes when the score is high and splashes when the score is low. Using tomatoes and splashes add in channels, color and shape, to increase readability, and the use of vertical and horizontal position is common in visualization. Thus makes the visualization friendly to general audiences. ~~Plot 2: It is intuitive to use area as mark and size as channel for visualization of categorical variables. Also, adding in position to help indicate the value is below or above the mean makes the plot more readable.~~ Plot 3: Since we are showing quantitative value of categorical items, it's a good idea to use area as mark. In this case, as items are regions/ethnicity, it's natural to use colormap, so I would take color as a channel. Plot 4: One of the aims of this visualization is to show the network between directors and actors/actress, then points and lines are natural choices of marks. To increase the readability, we substitute points with photos of corresponding directors/ actors/ actresses. To indicate the strength of connection, the width of the line is more readable than other channels like color and shape. Besides, to indicate the score earned by directors/actors/actresses and maintain the connection of the network, vertical position is a good choice of channel.

Following sample answer about a single plot shows how detailed your answers to part 2 should be.

1. Plot 1
   1) Task: This chart a) analyzes trend between Height and Weight of patients with heart diseases and b) locates outliers within the patients
   2) Attributes: Height, Weight
   3) Marks: point mark
   4) Channels:
      - aligned vertical position channel for Height

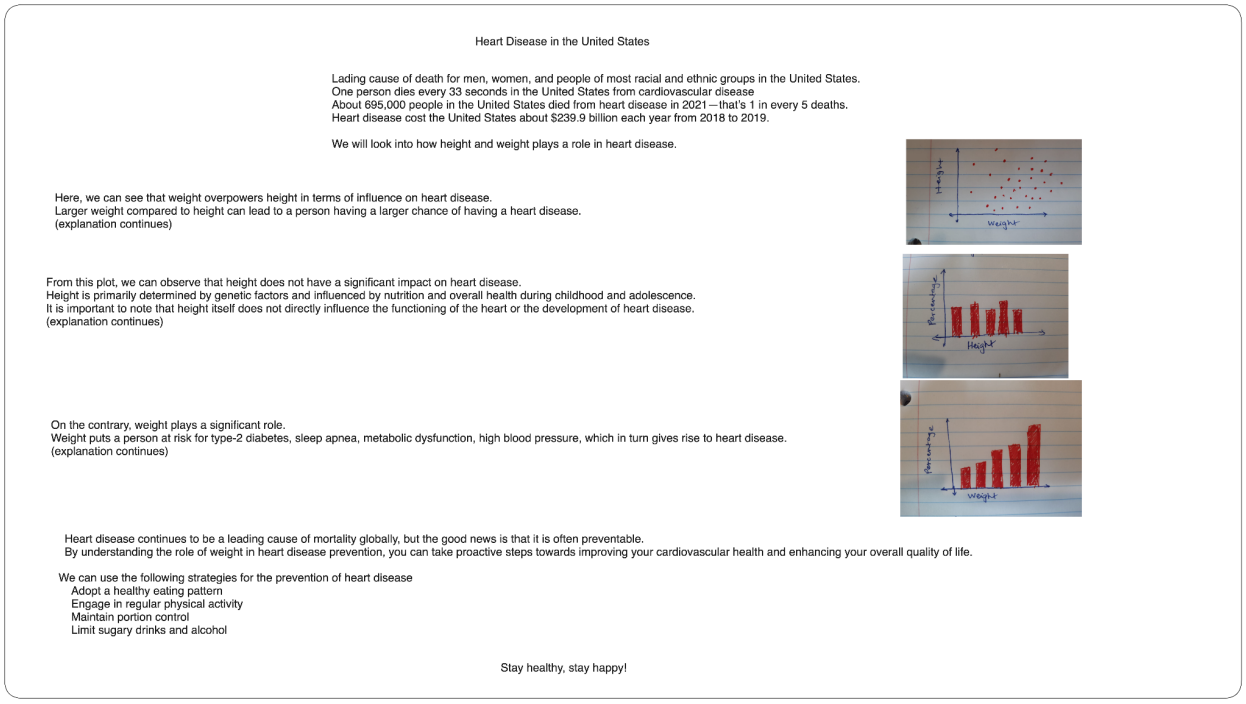
- aligned horizontal channel for Weight

5) How this plot adds to the story:
   My visualizations aim to deliver health characteristics of patients with heart disease. This plot will provide more specific insights on Height and Weight.

Part 3 - Prototype

Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.

A basic, barebones sample prototype for this project



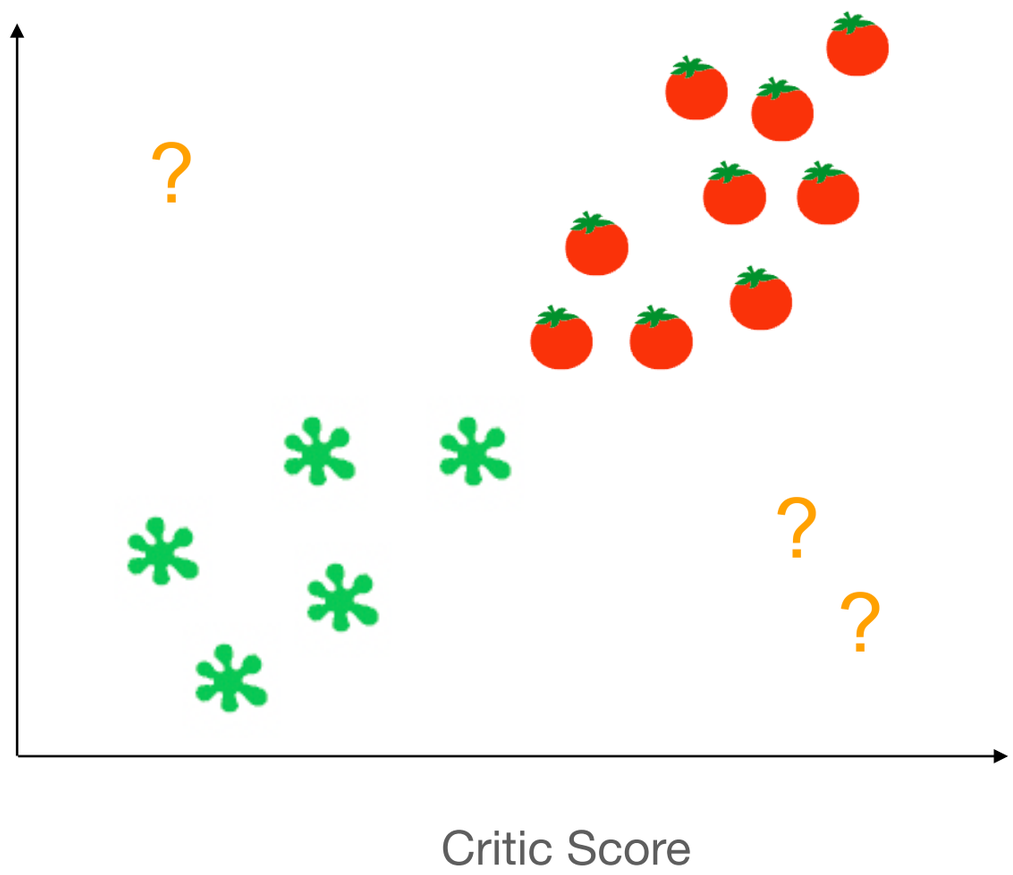
Good Movie? Bad Movie?

People are estimated to watch 3,600+ movies in their lifetime. Hollywood contributes $504 billion to US GDP.

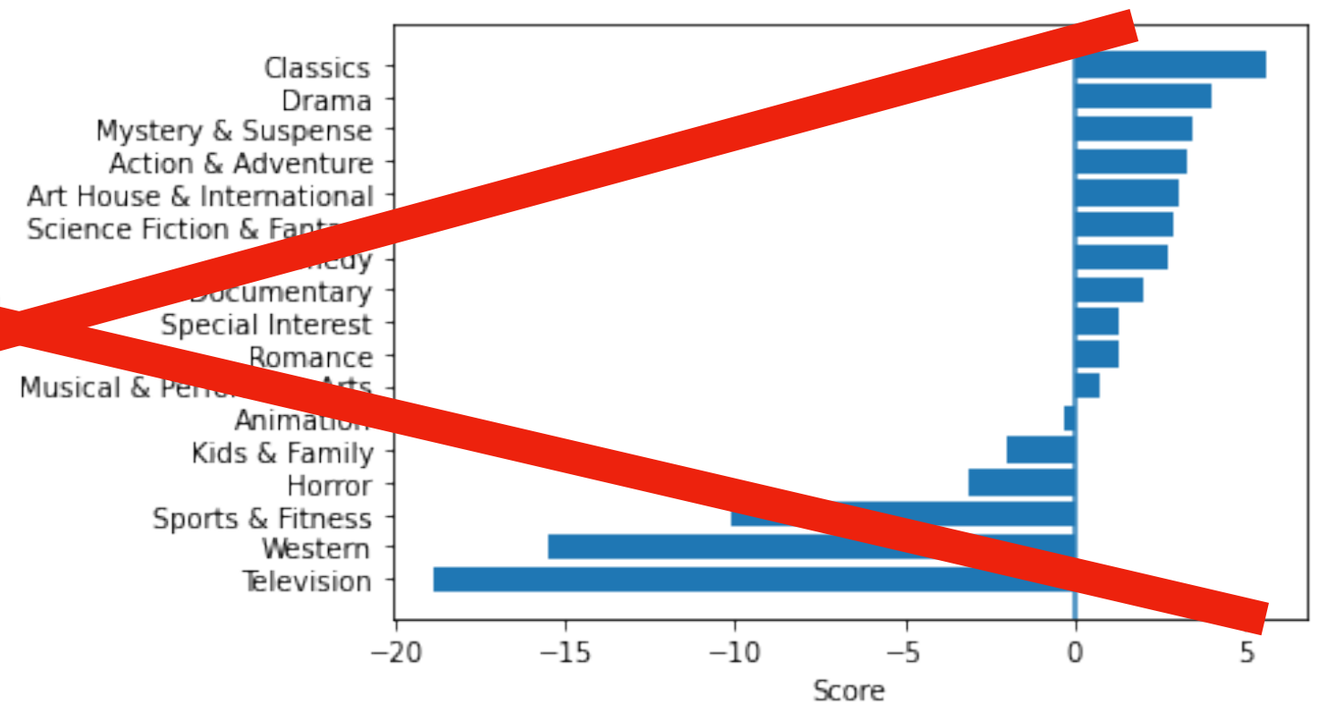
It's important to understand what elements make a good movie.

Critics and common audiences in general have common taste.

It is also notable that people's view divides from critics' for some movies. "It comes at night" is one them. Reason XXX.



People tend to give movies of type "Classic", "Drama", "Mystery and Suspense" higher rotten tomatoes score, while "Television", "Western" and "Sports & Fitness" have lower scores.



Movies from "Japan", "Hong Kong", and "Bollywood" have the highest scores.

Director A and Actress B have the highest score. Director C collaborated x time with Actor C.