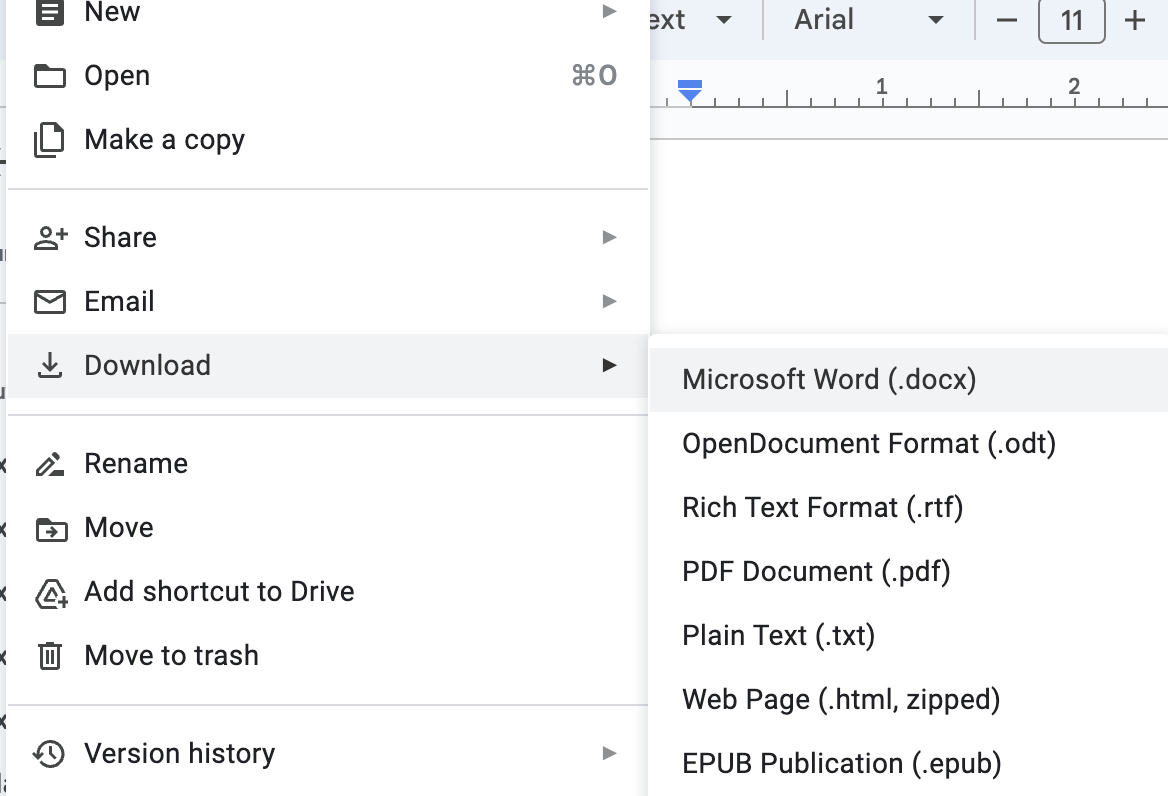
For this in-class activity, we will be working in groups of two or three.

Download a copy of this document by clicking on “File” → “Download” → “Microsoft Word (.docx)” in the menu above. You will use this doc to keep track of your work. Feel free to share the doc with teammates using Microsoft 365. You can also make a copy through google docs if you like.



**Note**: text in italics should be edited by your group!

**Note**: Make sure exactly one team member submits your in-class activity document [here](https://forms.gle/LpWDTjEBqdhFQ2yk8) *before* you leave lecture today!

In these in-class activities, we will practice designing a visualization to tell a story about a dataset, similar to what you are doing for A1. We will go through the following exercises:

1. **Review the Dataset (8 mins)**
2. **Brainstorm Analysis Questions (7 mins)**
3. **Choose a Question to Answer and Create Visualizations (25 mins)**
4. **Evaluate your Designs (10 mins)**
5. **Reflection (5 mins)**

## Exercise 1: Review the Dataset (8 mins)

For this in-class activity, we will use a small [movies dataset](https://docs.google.com/spreadsheets/d/1NrYHXJtdVq245zzBSsuIcjKWOJ1yS6lwVfj2-j-xQJg/). A CSV version of the dataset is available [here](https://drive.google.com/file/d/1ntxwW2XhvkLulactxDNJRELTnbwReMoH/).

Here are the attributes of the movies dataset. Pick a few of the attributes below and think about the data types (N, O, Q) and whether you might use them as dimensions or measures when answering potential analysis questions. You do *not* have to fill this out for every attribute.

|  |  |  |  |
| --- | --- | --- | --- |
| **Column / Variable / Attribute Name** | **Description** | **Data Type (Nominal, Ordinal, Quantitative Interval / Ratio)** | **Dimension / Measure** |
| Title | Movie Title | *Nominal* | *Dimension* |
| Release\_Date | The full release date | *Quantitative Interval* | *Both* |
| Release\_Year | Year of the release date | *Quantitative Interval* | *Both* |
| Release\_Month | Month of the release date | *Ordinal* | *Dimension* |
| Movie\_Length | Duration (in minutes) | *Quantitative Ratio* | *Measure* |
| Genre | Movie genre | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |
| Director | Director’s name, first and last | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |
| IMDB\_Rating | The IMDB rating (out of 10) | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |
| Rotten\_Tomatoes\_Rating | The rotten tomatoes rating (out of 100%) | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |
| Oscars\_Won | Total Academy Awards (Oscars) won for this movie | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |
| Production\_Budget | Estimated production budget for this movie | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |
| 2022\_Inflation\_Adjusted\_Budget | Estimated production budget, adjusted for inflation compared to 2022 dollars | *[N/O/QI/QR]* | *[Dimension/Measure/Both]* |

Is there anything confusing about this dataset? Anything that doesn’t make sense? Please feel free to ask questions and get help during the in-class activity!

**Note**: Make sure exactly one team member submits your document [here](https://forms.gle/LpWDTjEBqdhFQ2yk8) by the end of lecture!

## Exercise 2: Brainstorm Analysis Questions (7 mins)

Brainstorm at least 3 questions you could answer using this dataset and list them here:

* *Out of the sampled movies, do horror or thriller movies cost more to produce?*
* *Out of the sampled movies, do average ratings increase over time (critic v general audience)?*
* *Out of the sampled movies, do movies with longer titles score fewer oscars?*

Questions should be something you might ask another person and expect them to answer. They should not be a summary. For example this is not a question: “relationship between movie length and Oscars won”. Instead, a relevant question might be “How does movie length influence Oscar winnings?” or “Do longer movies win more oscars?”

**Note**: Make sure exactly one team member submits your document [here](https://forms.gle/LpWDTjEBqdhFQ2yk8) by the end of lecture!

## Exercise 3: Choose a Question to Answer and Create Visualizations (25 mins)

Each group should pick *one* question they want to answer with the visualizations you create. Write the question here:

*Out of the sampled movies, do movies with longer titles score fewer oscars?*

Now, the group should use whatever tools they want to create at least two visualizations to answer this question! For the sake of time and given how early it is in the quarter, you are free to draw/sketch by hand or use google sheets to generate charts for this particular exercise. Focus on generating a few quick designs rather than making a single, “perfect” design.

However, if you feel confident in using some of the following tools already, you can create visualizations using: Observable (examples [here](https://observablehq.com/collection/@uwdata/visualization-curriculum)), Tableau Desktop (assuming [you already have Tableau installed](https://www.tableau.com/academic/students)), the Vega-Lite live editor (examples [here](https://vega.github.io/vega-lite/examples/)), or any other software tool.

Paste images of the visualizations you create below. Hand-drawn visualizations can be included by taking a picture on your phone first, then transferring the photo to your computer.

**Visualization 1**

A graph with blue dots

Description automatically generated

**Visualization 2**

A diagram of a movie title

Description automatically generated

**Note**: Make sure exactly one team member submits your document [here](https://forms.gle/LpWDTjEBqdhFQ2yk8) by the end of lecture!

## Exercise 4: Evaluate your Designs (10 mins)

Now, each group should pick a single visualization they want to evaluate. Answer the following questions for this visualization:

1. Going back to the original movies data table, what data attributes (i.e., data columns) are most important to answering your question? List them here in order of importance
   1. *Oscars won*
   2. *Length of movie title*
2. Which attributes (columns) did you use?
   * *Oscars won*
   * *Title -> length*
3. What encoding channels (position, color, size, etc.) did you choose for each attribute?
   * *Oscars won → hue (and text)*
   * *Title → position*
4. How does your choice of encodings compare with Mackinlay’s effectiveness rankings discussed in lecture? We have pasted [Mackinlay’s rankings](#w83qkq5miv7y) at the end of this doc for convenience.
   * *For oscars won, we used hue which is low on the effectiveness rankings. For title length which is quantitative, we use position which is the most effective encoding.*

## Exercise 5: Reflection (5 min)

Based on your evaluation, what do you think of the visualization you picked? Were there specific design decisions you made today that you would change for A1 (and later, A2)?

*We think it was pretty funny but not the most effective. The hue and label both encoding number of oscars won was redundant.*

Are there any other general takeaways that you learned from these exercises today?:

* *Redundant encodings not always necessary*
* *We should spend more brainstorming time about how to encode data before we start plotting*

**Note**: Make sure exactly one team member submits your document [here](https://forms.gle/LpWDTjEBqdhFQ2yk8) by the end of lecture!

## Mackinlay’s Effectiveness Rankings for Encoding Channels

Below we’ve pasted Mackinlay’s ranked lists of visual encodings, ordered from top to bottom by their perceptual effectiveness. For example, position encodings are considered most effective for all three data types, according to these rankings. Similarly, a shape encoding is considered the least effective for quantitative and ordinal data.

