Name(s): _____

DATA 101 Assignment 3: Visualization

Work in teams of 2

Instructions. Answer the following questions using the linked datasets and activities. Always note the source of your information when applicable. Write your responses in the spaces provided.

---

1. Go to https://ernbilen.github.io/moviesd3. Find the information below.

(a) What is the type of the graph you see on this Movie Data explorer? Scatterplot.

(b) What is the relationship between `Tomatometer` (Rotten Tomatoes) and `imdbRating`? Positive.

(c) According to the graph, does a higher score for `Meter` cause `imdbRating` to go up? Explain. No. The cause of an increase in `imdbRating` is not an increase in `Meter` score, but the fact that it may be well received by the audiences causes both scores to go up.

(d) Does the relationship you see hold across different genres? Use the dropdown selector for Genre and browse 3 different genres and explain. Yes, the positive correlation roughly holds accross genres.

(e) On the X axis dropdown selector, select `imdbRating`, on the Y axis select `BoxOffice`. Do higher rated movies achieve higher box office sales? Yes.

(f) On the X axis dropdown selector, keep `imdbRating`, on the Y axis select `Length` (runtime). Are higher rated movies tend to be longer? Explain. Yes, there are more points towards the top right hand side of the graph.

2. Go to https://www.reddit.com/r/dataisbeautiful/comments/1ncsti1/oc_apple_event_2025_most_frequently_used.

(a) What is being measured in this visualization? Identify the variable(s) on the axes. What does "frequency" refer to in this context? These are most frequently used adjectives from the Apple Event 2025. Frequency is a measure that counts the occurences of each adjective.

(b) Which adjectives are the most frequent? Which are less frequent? What might that tell you about the tone or messaging strategy of Apple during this event? Adjectives like "new", "more" were the most frequent. Notice we can't really say adjectives such as "best", "powerful" are less frequent. The fac that they made it to this graph makes them somewhat frequent. We can't really tell from this graph, but words such as, "old", "bad" etc. could be the least used words at this event. The graph shows that Apple uses positive and innovative adjectives to match the hype surrounding their event.

(c) Find one comment in the comments section that you think is interesting. Write it down and explain why you found it interesting. By Praglik: *I would love to see a comparatuve graph with a bunch of older keynotes from them!* Interesting, because we could track how Apple has been evolving as a company.

(d) Describe one other way the same data could be visualized. Instead of a single bar-chart, we could plot when certain words appear during the event (early vs middle vs end). We would have minutes since the event started and the frequency of certain words being used across time.

(e) Suppose you were asked to compare this Apple Event with a past event (say 5 years ago) in terms of adjectives used. Outline a strategy you might use (what to collect, what to plot) and one obstacle you might encounter. You would collect transcripts from both events, extract all the adjectives, and count how often each one appeared. Then you could make a side-by-side bar chart showing the most frequent adjectives in 2025 versus 2020. One obstacle is that the two transcripts might differ in length or focus, so you would need to normalize counts (for example, per 1,000 words) or else the comparison might be misleading.

3. Go to https://www.reddit.com/r/dataisbeautiful/comments/1hf04x9/2024_in_search_trends_oc.

(a) Describe what this chart is showing. Distributions of select group of Google search keywords in 2024.

(b) Pick one topic from the chart and describe when its search interest was highest and lowest in 2024. Mike Tyson was searched very little throughout 2024 except in November due to him getting back to the ring on November 15, 2024.

(c) What might a sudden spike in search interest suggest, compared to a flatter or more stable trend? A sudden burst would indicate high interest that fades quickly. A flatter distribution indicates the interest remained lower but more consistent.

(d) Pick a search topic from the figure that surprised you. What do you think explains its pattern? I found it interesting that US Election keyword only appeared as a burst near November and did not show up to be searched as much in the months leading to November.

4. Go to https://www.reddit.com/r/dataisbeautiful/comments/1ftmkwt/oc_foods_cost_vs_caloric_density.

(a) What kind of chart is being used? Scatterplot.

(b) What are the variables in this chart? Which is on the x-axis and which is on the y-axis? Calories per gram of food on the x axis; Cost (USD) per 100 calories on they y axis.

(c) Describe the relationship between cost and caloric density: do more calorie-dense foods tend to cost more, less, or is there no clear pattern? Negative. For the food included on this graph, more calorie dense food seem to be cheaper.

(d) Can you claim the relationship you see is causal? Why or why not? No. The chart shows correlation, not causation. Both cost and calorie density could be influenced

by other factors such as shelf life, processing level, or market demand. To claim causality we would need an experiment or controlled study.

(e) Are there outliers (foods that cost a lot but are low calorie, or vice versa)? What might explain those outliers? Yes. Cultivated mushrooms and ribeye steak are outliers — they cost much more per 100 calories despite not being especially calorie-dense.

(f) Name three foods that are the cheapest while providing the most calories per cost. Any of the three from: peanut, almond, sunflower seed, cashew, wheat spaghetti, brown rice, chickpea, lentil.