Methods

- It’s definitely important to be clear about the data collection process. You may want to comment on this with vocabulary from class. Is this data population or sample? Can it be used to generalize to a broader population?

- You discuss this a bit when you claim that because you have all games from Steam, you can generalize to all games for PC (prior to 2016). This is an interesting claim and requires support. Not all PC games are on Steam. Just because Steam is the “number one game distribution service” doesn’t mean that this data can be used to generalize about non-Steam games. That is generally an indication that your sample is representative of the entire population. Is that the case here? It may be, but if so, you need to argue it with better evidence.

- This ties in to your proposed methods of data analysis. Explain why you chose to focus on descriptive and visual measures.

- This is where it is also helpful to draw on your background reading results. Based on your background reading (not just on your own knowledge!), do you expect to encounter any confounding or lurking variables? If so, how will you deal with that? How will you visualize the impact of lurking variables, if you find any?

Data analysis

- What happened to question 1?

- for Question 2:

+ Speaking of regression, you need to assess whether the conditions for regression are met before trying to interpret the p-values. Is there a linear association? Are the residuals normally distributed? Is there constant variance around the regression line at each x-location? Are the data independent (random)?

For final paper & other:

- You have great info on metacritic score; this is a nice level of detail, and there are important implications for your research findings. You will likely want to expand/reflect on this info in your conclusion, as you discuss how your results tie in to the background information if you find an association or not

Question 2&3

+ You should consider transforming one or both of your variables to deal with the outliers. This can be as simple as taking the log or square root of the variable. Transformations can help linearize relationships, though they do affect the interpretability of any output from regression.