- Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

    The project will be a visualization tool of how one steam game is related to others through common buyers. Using two different data files, the customer data and the game description data, we are aimed at giving a new insight of the game dataset by connecting games through user behaviors.
    There will be a zoomable view of connected graph, and a list of related games with their attributes shown in detail. The user can see, select, filter result and examine the list of games given as potential games to purchase.

- Related Work: Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.

    Steam has its own hardware & software survey to present some of its user information, but as the title suggested the focus is on devices and operating systems only. There are also many third-party websites designed to provide insights into steam games data, such as Steamspy and SteamDB. These, however, often just displays and sort the data and thus serves mostly for ranking purpose.
    There is also an embedded "Recommended > similar items" section in Steam, but the evaluation method is unclear and it is designed for commercial use, and thus might be biased.

- Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

    Given the two datasets, we discover that they are connected through game name. Purchase data provides individual steam customer behavior, while Steam Games data provides detailed information on each game sold on Steam. By connecting these, we can pose the following question: what other games that a user is more likely to purchase if they have purchased one game? How does each game connect to other games, and how would this connection affect a user's behavior? With this visualization, we want to answer these questions, and offer advertisement tips for game developers and market sales based on our findings.

    While developing this project, there are several questions we need to answer:

    1. The given dataset has too many attributes. What should be visualized to best answer the questions above?
    2. Two dataset have only one attribute in common, which is the game name. What visualization design can best associate the two datasets?

- Data: Source, scraping method, cleanup, etc.

Data Source

     Both dataset comes from dataworld. The game-features.csv is adopted from a sample data project on github, and their original data is from public Steam datas on Steam's API and steamspy.com. The steam-200k-csv is from a kaggle page where they list 200k Steam user interaction.

Data Processing

     The game-features dataset is a very large dataset and contains very extensive information about games on steam. It will slow the visualization significantly or even cause the browser not responding. Therefore we need to reduce columns in the dataset. Information such as minimal system requirement and supported languages is irrelevant to the question we are trying to answer. These columns will be deleted in data preprocessing. No extra quantities is expected to be derived from the data. The stream-200k-csv data only has four columns and are all essential to the visualization. So no cleanup will be done on this dataset.

     The network uses its own node and links dataset, which we use python to preprocess it. Using the steam-200k dataset, we find the links in the dataset, by looking at the same user id that games share, and build a link list to populate the network.

- Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
       We use Excel spreadsheet to explore the data and remove some unnecessary columns. We discovered the steam-200k-csv contains game titles are all included in the game-features dataset. So we think making corresponding highlight between views will make it much easier to navigate. We also realize that the DLC content is not specifically labeled in game-features dataset, so there is no simple method to remove the DLC titles in the visualization for now.

- Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?

     We received good feedback from the peer review session. We received several suggestions as add-on features to implement in the visualization. After considering all the visualization from our proposal, including different types of table visualizations and

networks, we decide to have a network of games connected by player purchasing behaviors and a table with game information. This makes good use of the data we have, and is easy to navigate for the user.

- Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

    As in the final design we first visualized the table. As a first step the price bar is drawn as an attribute visualization example.

    We use an experimental datasets of 23 games and 50 rows of user respectively to generate this initial view.
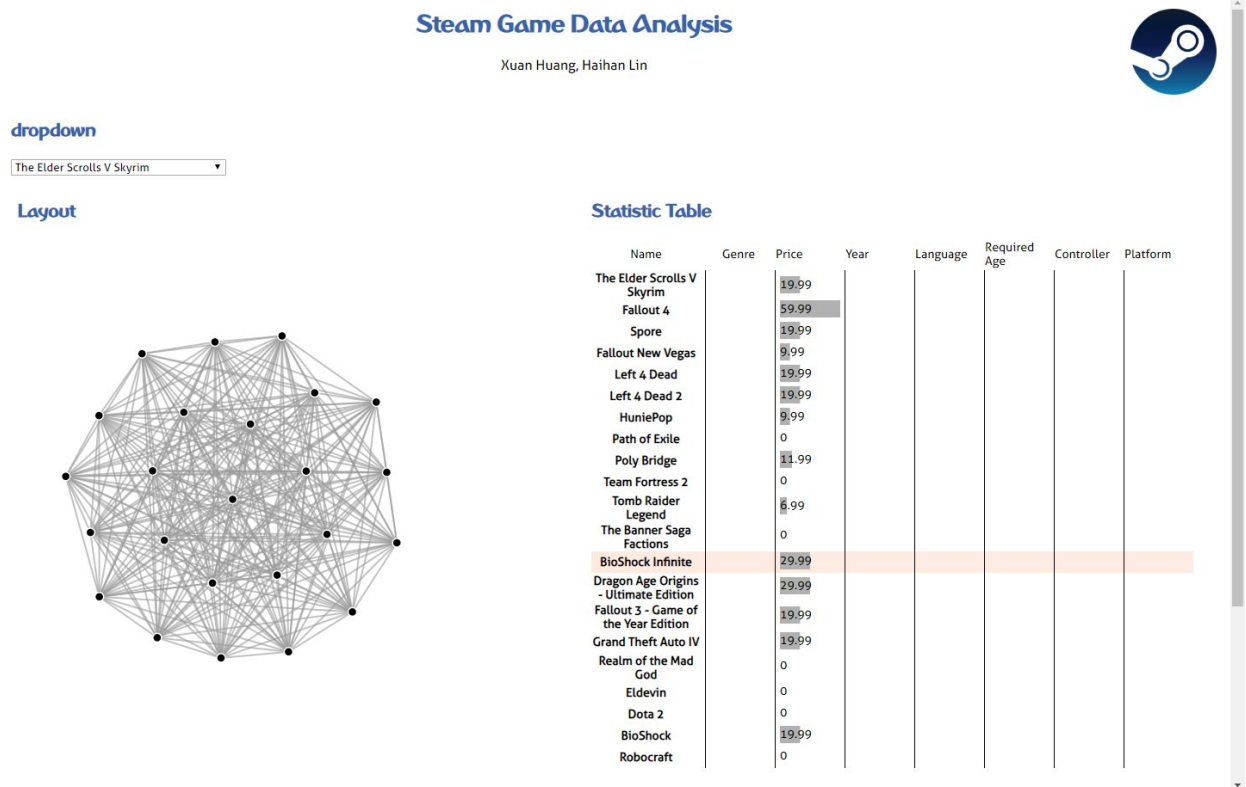
- Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

    We have found that the game are highly connected in the network. Therefore, some level of hiding will be required when we put the entire dataset in the network. And when zooming in, the user will be able to see more neighbors. Currently the two views are separated, which we want to fix as the next step.