In my visualization, I chose to use the movies.csv dataset.

Analytical tasks:

Various different analytical tasks can be completed with this visualization. Information about a specific movie's year, genre, and profit are displayed. In addition, filtering allows the ability to compare certain genres, and how profitable they were. Here are a few analytical tasks that could be potentially completed with this visualization:

- In what year were comedies most popular?
- What was the highest grossing sci-fi movie in 2012?
- Are dramas or romantic movies more popular?
- Has the popularity of horror movies declined over time?
- How many sports movies were made each year?
- What genre was the most popular film in 2014?

Design overview:

My visualization consists of a scatterplot of the various movies included in the dataset. The main genre of the film is encoded with a color, which the user can identify with a key on the

right-hand side of the chart.

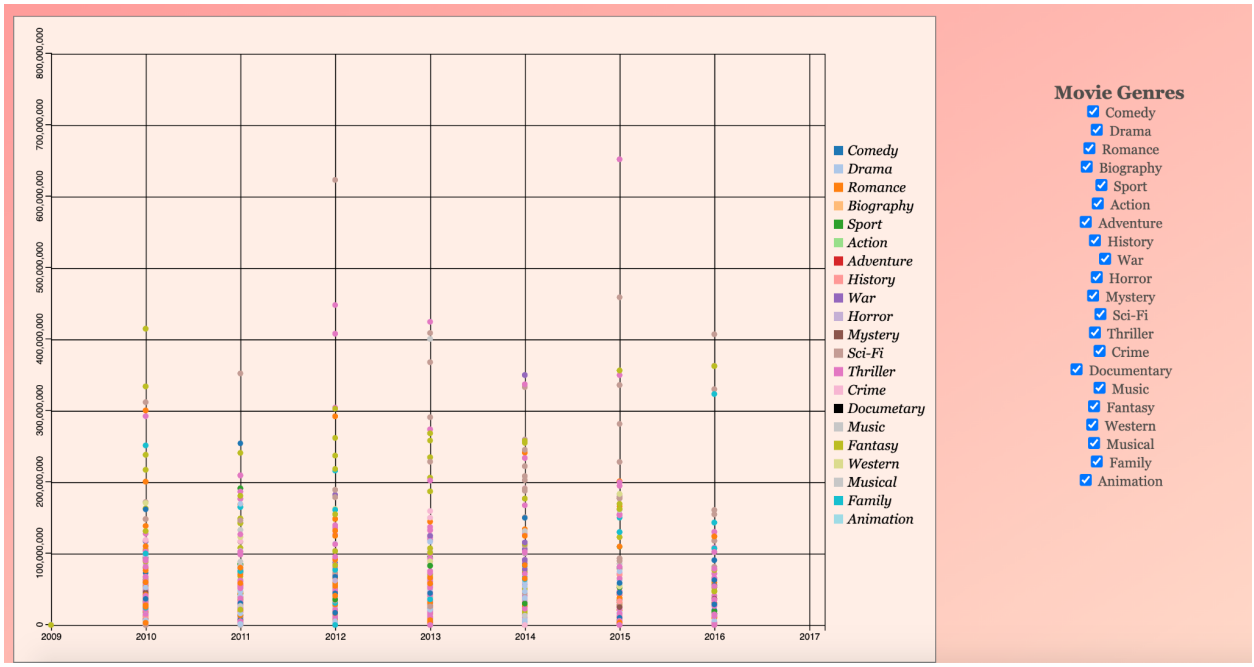

Image 1: The visualization when it is first loaded

The user can choose to filter by a particular genre of choice by utilizing the checkboxes on the right-hand side of the screen. Selecting a checkbox displays films with that genre as its main genre on the scatterplot.

As there are a great number of points, the user also has the option to zoom in and pan on the graph in order to see the points in greater detail. However, as the dataset did not come with specific release dates, and instead only the release year, zooming and panning in the X direction would not work (when zoomed in, dots will only display on January of that year, instead of the actual release date of the film). I decided to only let the user zoom/pan in the y direction as opposed to in both the X and the Y direction. This will eliminate confusion, as if zoom was enabled in both directions, it would be difficult to draw conclusions. Since zooming and panning is only enabled in the Y direction, it is much clearer to see which films had the highest profit in a

certain year.

Image 2: The visualization, filtered for only comedies and sport movies, and zoomed in

In addition to these, enabling a tooltip on hover provides the user with more information about a specific movie. When a user hovers over a particular dot, the movie's title and total box office revenue are displayed. This helps the user answer particular questions about revenue, and also gives specifics about the particular movie.

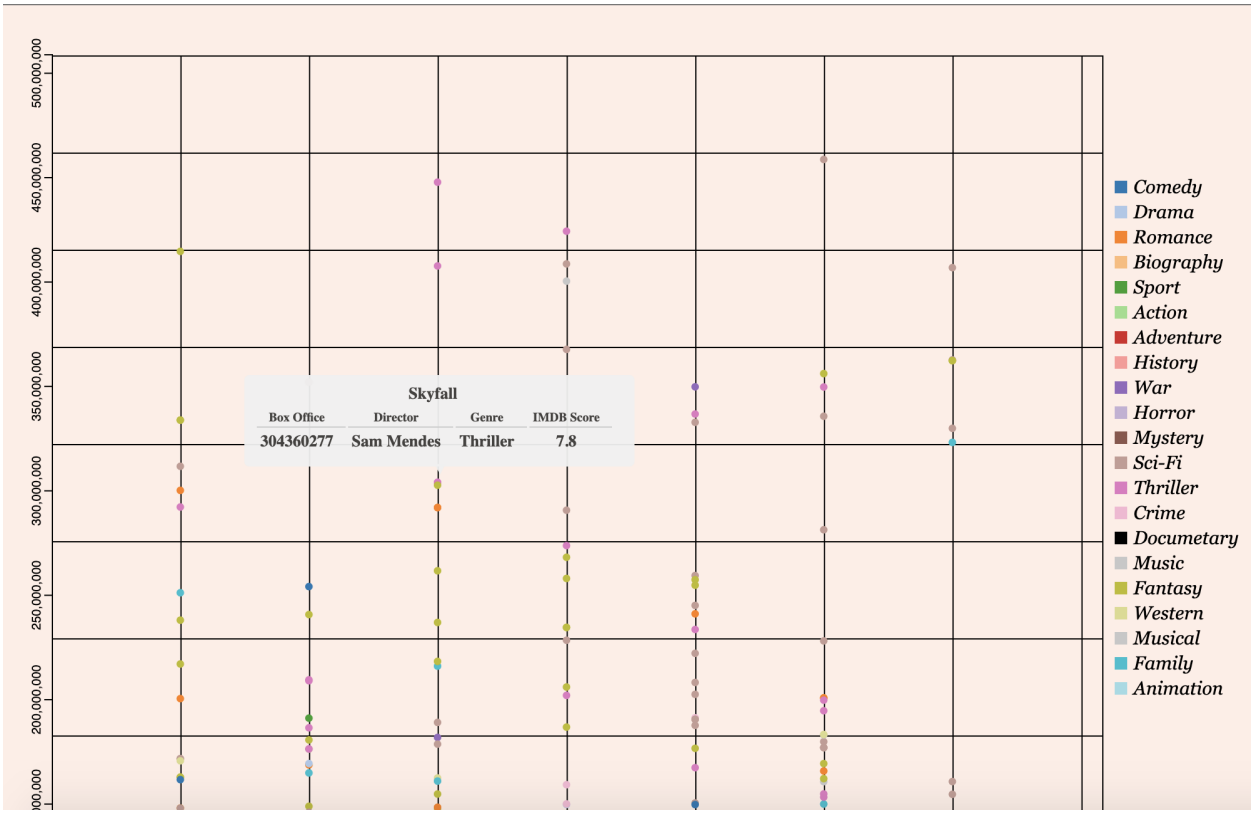

Image 3: Cursor hovering on a particular dot, revealing information about the movie

Some challenges I encountered when building this visualization was determining how much data to show on the chart. As taught in class, overloading the chart with too much information can overwhelm the user. Although there are many genres displayed on this chart, allowing the user to zoom and filter lets them see overall trends at first (ex: are movies not generating as much revenue anymore?), and then gather more data about particular instances (ex: what was the highest grossing movie in 2012?). In addition, choosing a color scheme for the genres was difficult, as I wanted to choose a scheme that had enough contrast. I ended up utilizing d3SchemeCategory20, as it seemed to have the most contrast between the colors.