Megan Castrataro

To begin my final project, I found a dataset that interested me. After looking through Kaggle, I selected a movie dataset that encompasses information on the top-grossing movies of all time. For my visualization, I knew I wanted to use a scatterplot, so I looked through the different attributes to see which ones I could use. I knew I needed quantitative variables, so I initially used domestic sales and global sales. However, I ended up changing the variables to domestic sales and international sales. A key element of my visualization is its interactivity. As the user looks through the scatterplot, they can hover over the points to receive additional information on the specific movie (the title, description, distributor, and license). As seen in my sketches, I also wanted the user to be able to filter this information by certain variables (distributor and license). This allows the user to explore the data rather than just look at a static visualization.

Throughout the creation of this visualization, my design choices were made intentionally. This dataset had thousands of records, so I decided to only include the first 100 records to not overwhelm the user with data. I set the dimensions of the chart to take up most of the page so that the user could easily view the chart but not have to scroll. When sorting the data by distributor/license, I chose to do this by color. I wanted the user to quickly detect any patterns after sorting and I figured this way would be easiest. I picked colors that would allow for sufficient color contrast. I put the hover tooltip to the right of the mouse and made it a different color for the user to easily view the details. For the styling of the page, I chose an easy-to-read font and made sure the size was large enough.

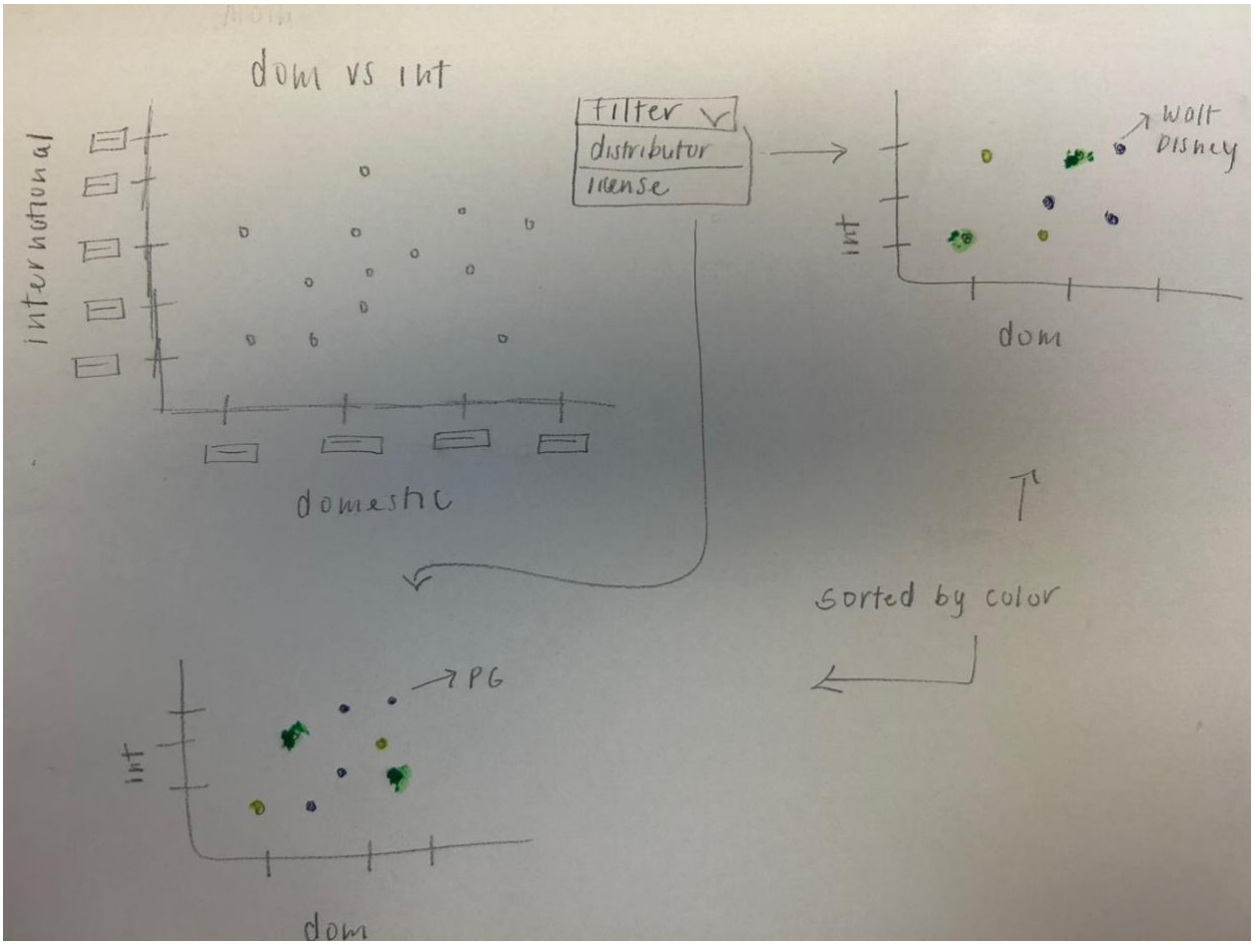
When creating this visualization, I wanted to answer the following questions:

- *Is there a correlation between domestic sales and international sales?*
- *Which distributors have the highest/lowest domestic and international sales?*
- *Which licenses have the highest/lowest domestic and international sales?*

Looking at the initial scatterplot, it's evident that there is a positive correlation between domestic sales and international sales (as seen in screenshot 1). There are a few outliers outside of this correlation, such as Furious 7 (low domestic, high international) and Star Wars Episode VII - The Force Awakens (high domestic, low international). These outliers can be seen in screenshots 2 and 3. When filtering the data by the distributor, we can see that Walt Disney distributed a lot more of the higher-grossing movies domestically and internationally, while Twentieth Century Fox distributed more of the lower-grossing movies (except for Avatar - as seen in screenshot 4). When filtering the data by license, we can see that PG-13 and movies without a license (NA) hold the higher grossing spots, while R and G movies hold the lower grossing spots.

Megan Castrataro

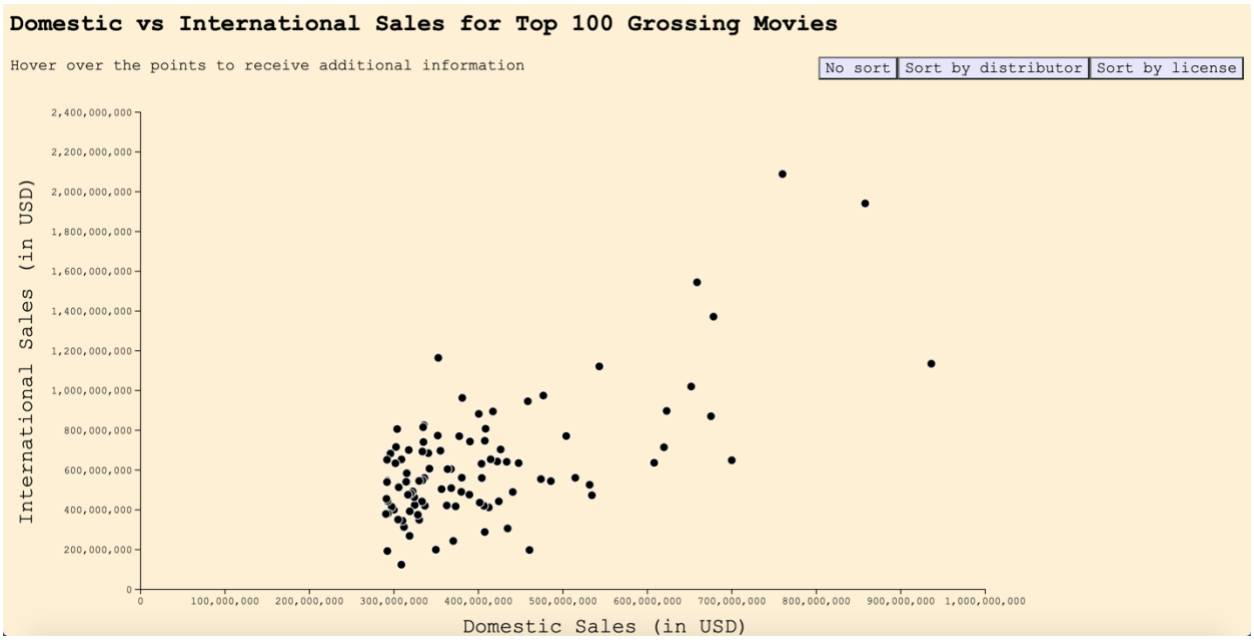
Sketches:



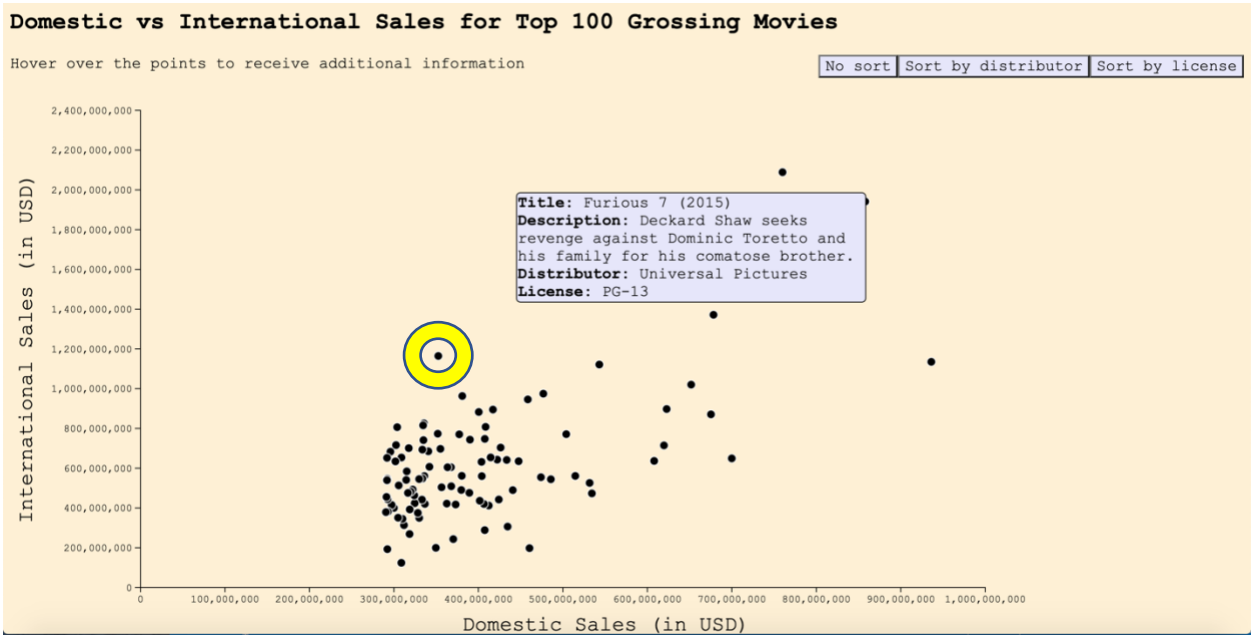
Megan Castrataro

Screenshots:

Screenshot 1

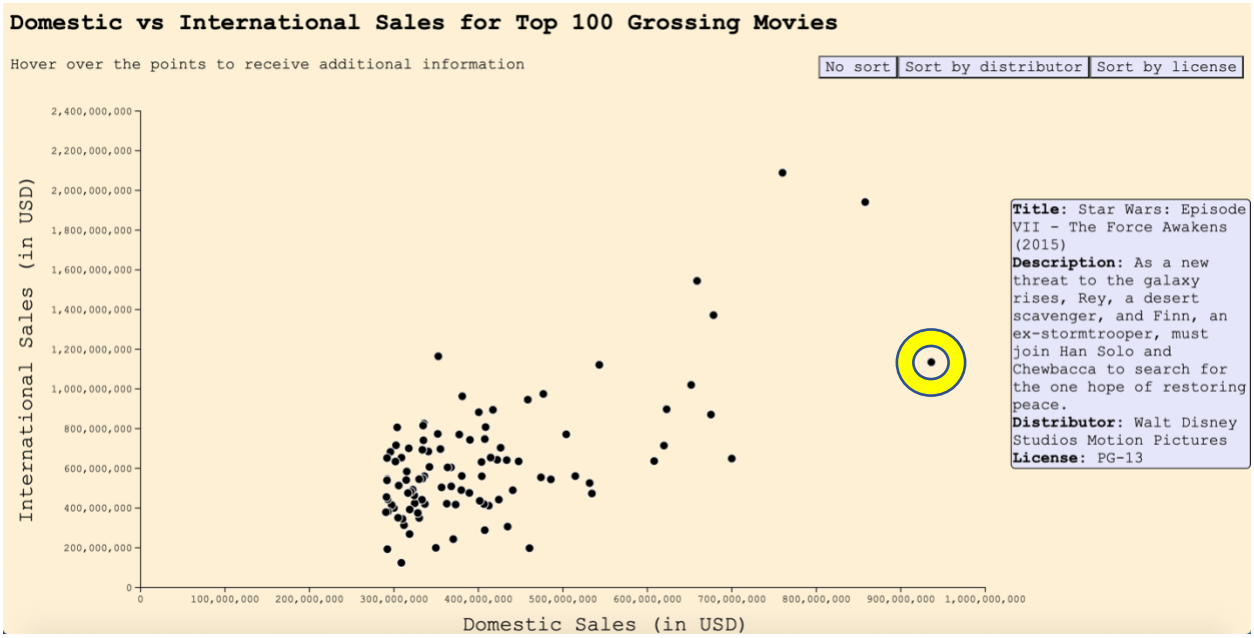


Screenshot 2



Megan Castrataro

Screenshot 3



Screenshot 4