DTSA 5304 Fundamentals of Data Visualizations

Final Project (October-December 2022)

Data:

Link: https://www.kaggle.com/datasets/deepcontractor/marvel-comic-books

The data I used for this project is titled “Marvel Comic Books Dataset”. It includes every comic book released by Marvel since the company’s origin in 1939 up until mid-2022. It contains the name, issue title, publish date, writer, penciler (aka main artist), and some other minor data for every issue released. Unfortunately, much of the data is “null” or “none” so some other fields such as rating, price, or format couldn’t be reliably used as most of those values are empty. That being said, there is still a wealth of data here with over 30,000 individual comics being accounted for. This was the most comprehensive dataset I was able to find on this subject.

Goals:

My goals for this project were fairly simple. My plan was to create some simple visualizations that showed the top Marvel writers, pencilers, and series as well as breaking it down over time. I was able to accomplish this goal but as I will discuss later, I wasn’t able to do more complicated things to my data that may have made the visualizations more interesting.

Tasks:

- User should be able view bar chart of top 20 Marvel writers. User should be able to hover over each bar to get exact number of issues written.
- User should be able to view bar chart of top 20 Marvel pencilers. User should be able to hover over each bar to get exact number of issues penciled.
- User should be able to view bar chart of top 20 Marvel series by issue count. User should be able to hover over each bar to get exact number of issues released.
- User should be able to view line chart of the number of issues Marvel has released per year. User should be able to hover over the chart to see the number of issues released per year.

Visualization:

Link: https://www.kaggle.com/code/hamzasclone2/dtsa-5304-final-project

Please go to the link above to view the visualization. This is the same one that the people I recruited for the evaluation section viewed as well.

The visualization was made in Kaggle, which hosts a Jupyter Notebook. I used python, pandas, and altair to interact with the data and create the visualizations.

I have included some screenshots of the visualization at the bottom of this report in case the link does not load.

Design:

As detailed above in the Tasks sections, my ideas for my design were fairly simple. I would have liked to include more but many of the other fields in the dataset were mostly empty. Thus the primary data I had was the titles of comics, when they were released, and who worked on them. I love Marvel comics so I still found this data very interesting and it was entertaining to make these designs and find out who the most common writers were and so on.

The first chart was over the Top 20 Marvel Writers by the number of issues they wrote. I think a bar chart is what makes the most sense and I made it so the writers are on the y-axis as it would make their names easier to read. I created a tooltip that allows the user to hover over each bar and get the exact number of issues they wrote. I also made the ability to select each bar, but this doesn't really serve any function at the moment.

The second chart was over the Top 20 Marvel Pencilers by the number of issues they penciled. It was designed exactly the same as the first chart. The third chart was over the Top 20 Longest Running Marvel Comics by Issue Count. It was also designed the same as the other charts. All three had the same justification.

The last chart was over the number of comic book issues Marvel has put out every year. I made this one as a line chart as I felt that was the simplest way to show this information. I also made it so that you can hover over the chart to see the exact amount at a specific year.

Evaluation:

I did my evaluation using a Summative Evaluation method. I asked three people to view the visualizations and interact with them. I then asked them to answer a series of questions and give me their general thoughts. I don't think this fits in neatly with any of the specific methods we discussed in class, but it is sort of a combination of an interview and journaling. The questions I asked were:

- Are the visualizations understandable?
- What did you notice first for each visualizations?
- What did you learn that you didn't know before?
- Was there anything confusing/unwieldy about any of the visualizations?
- What do you think about the color choice?
- What about the font and font size?
- What about the bar size and line size?
- Did the tooltips (hovering over the charts) help in understanding?
- Is there anything else you would have liked to have seen?
- Do you think there was a better way to visualize any of this data?
- Is there any other data related to this topic that you think would make for a good visualization?

Here are the results from each person:

#1: (A friend)

1. The visualizations are very clear
2. I was looking though the list of names first

3. Basically everything was new information. It was interesting to see the spike in issues in the mid 2000s
4. No, not confusing at all
5. Color was fine! Nothing amazing, but easily seen/not distracting.
6. Same thoughts as color. It's fine, probably best for it to be nothing too crazy
7. Bar graph length was good! I think having it stretched out as far as it went really showed the differences between authors, etc
8. I'm on mobile so I couldn't hover over anything
9. I'd like to see the stats for the overall characters (like how many issues of all x men versions, etc)
10. I think this was the best way to show the data. For example I think changing to the line graph was a really good idea for issues released
11. If data is available, digging into the demographics of who buys comics/which specifically would be interesting

#2: (My youngest brother)

"Both the bar charts and line charts were very understandable. It was super easy to read and all the data was put together in an orderly fashion. The tooltips were extremely helpful and showed the exact number. I don't think there was an easier way to display the data that was displayed."

#3: (My other brother)

"I think that the bar charts and the line charts both look simple and easy to understand. The axes are properly labeled so I know what each chart is depicting and when I hover over them it gives me more specific detail about each of the lines/bars, which is helpful. I think the color and font choices all complement each other and are pleasing to the eye, although I would have personally made the text a little bit larger so it is easier to read. As for the actual data on the charts, I learned a little about Marvel comic books that I didn't know. The bar graphs were an appropriate way to visualize the data, and it is pleasing to the eye and easy to understand. The line graph was easy to understand as well, and if I were doing a similar project, I would have made the same graph. For improvements, I would have tried to make it so that you can see the whole chart rather than having to scroll to the right to see the rest of them, but this is a minor detail. Other than that, I believe the charts were well made and easy to navigate."

Overall, all the people I recruited thought the visualizations were simple to read and informative. They only had minor comments on improvements such as having a larger font size and making the line graph fit on the screen.

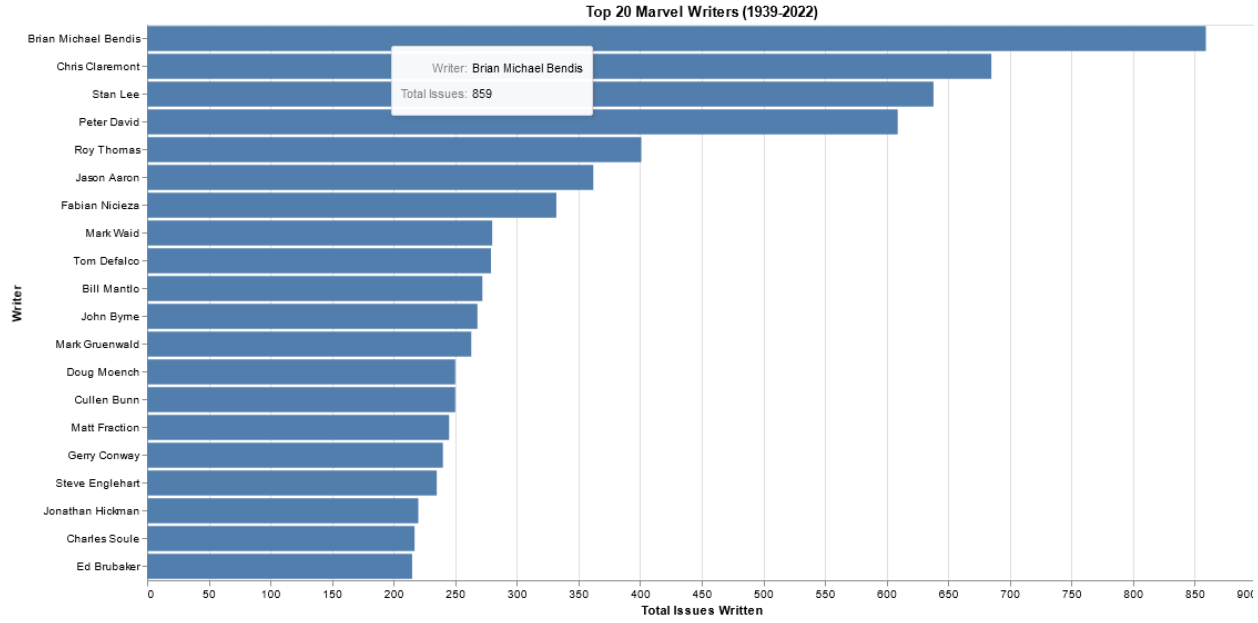
Conclusion:

For a first draft, I think these visualizations were successful, if a bit uninteresting. For all of the charts, the information that is shown is fairly simple and there isn't much of a reason to "drill" deeper. That is something that could be amended in further iterations. The main reason this wasn't done at the outset is because of my own limited knowledge of working with data. While I was able to learn a lot about Altair during this class, I still don't have much experience actually working with datasets in python (or other languages). Thus, more complicated ways of looking at the data is currently out of my reach. I wanted to further break down some of this data by decade, so that the user could also see the

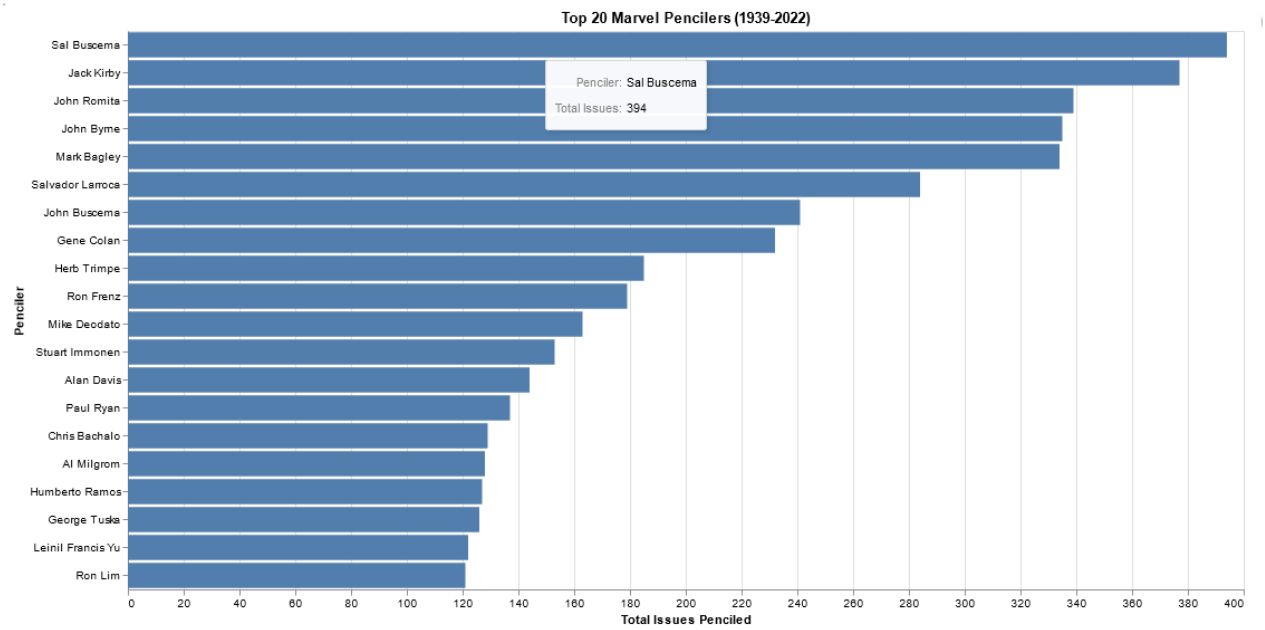
top writers/pencilers/series for each decade, but I had trouble dealing with dates and had to scrap that idea. I think it also would be interesting to see which writers/artists worked on what series and other relations like that. Also, it would be interesting to include data from other comic book companies such as DC, Image, or Dark Horse. In conclusion, I think this is a good start but would need to be worked on further until it was as visualization people would be interested in viewing.

Extra Screenshots:

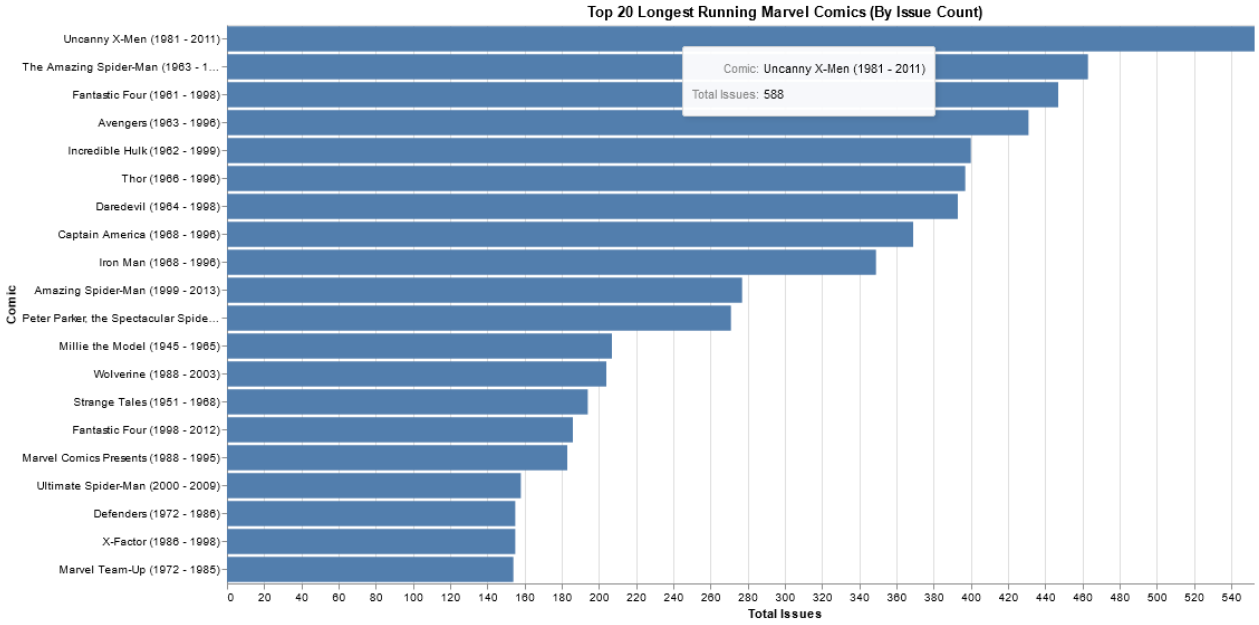
Out[2]:



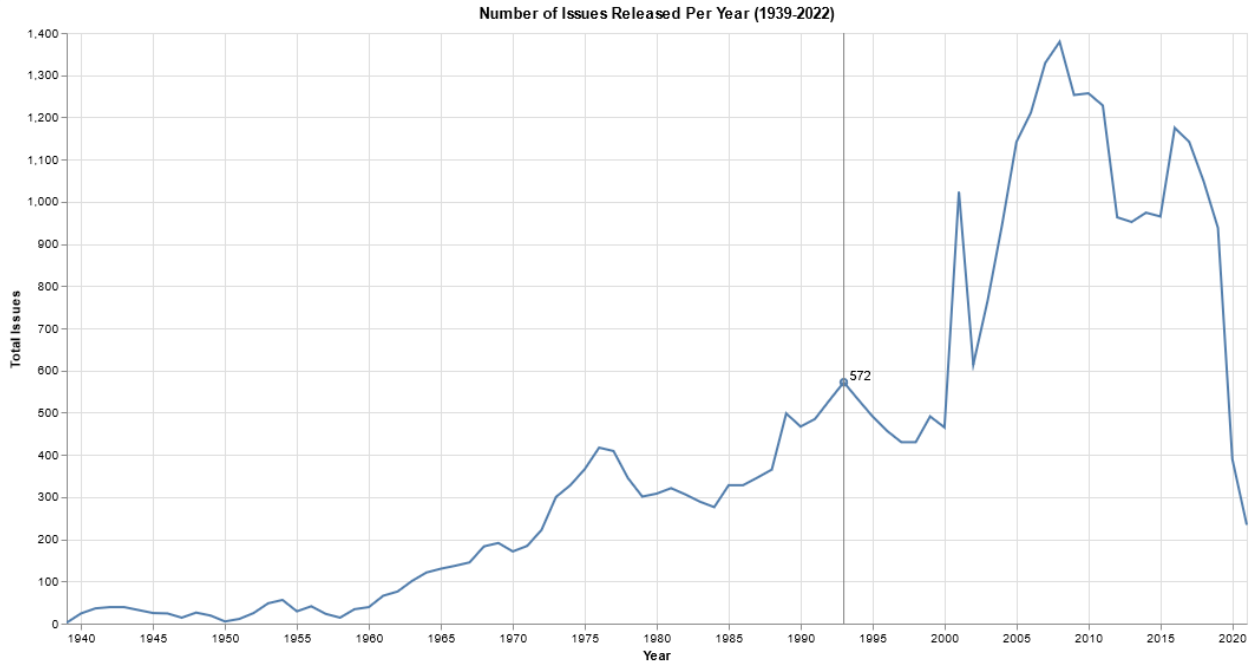
Out[3]:



Out[4]:


Top 20 Longest Running Marvel Comics (By Issue Count)
Uncanny X-Men (1981 - 2011)
The Amazing Spider-Man (1963 - 1...
Fantastic Four (1961 - 1998)
Avengers (1963 - 1996)
Incredible Hulk (1962 - 1999)
Thor (1966 - 1996)
Daredevil (1964 - 1998)
Captain America (1968 - 1996)
Iron Man (1968 - 1996)
Amazing Spider-Man (1999 - 2013)
Peter Parker, the Spectacular Spide...
Millie the Model (1945 - 1965)
Wolverine (1988 - 2003)
Strange Tales (1951 - 1968)
Fantastic Four (1998 - 2012)
Marvel Comics Presents (1988 - 1995)
Ultimate Spider-Man (2000 - 2009)
Defenders (1972 - 1986)
X-Factor (1986 - 1998)
Marvel Team-Up (1972 - 1985)
Comic
Comic: Uncanny X-Men (1981 - 2011)
Total Issues: 588
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540
Total Issues

Out[6]:


Number of Issues Released Per Year (1939-2022)
1,400
1,300
1,200
1,100
1,000
900
800
700
600
500
400
300
200
100
0
Total Issues
572
1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020
Year