Abstract

Question:
Our project explores the intersection of culture and digital trends, specifically examining how significant cultural events can reignite interest in traditional activities. We've chosen to explore this phenomenon through the lens of chess and its resurgence in popularity in the digital age. Our primary focus is on understanding the impact of "The Queen's Gambit," a popular Netflix series and prominent content creators like GothamChess. These modern cultural influencers have brought chess to the forefront of public attention, especially among younger audiences. By analysing Google search trends, our project aims to uncover the extent to which such media influences have altered the public's engagement with chess, transforming it from a classic pastime to a trending topic in the digital era.

Conclusion:
Our analysis has led us to a conclusion: There is a marked increase in chess-related online searches post 2020, a trend closely aligned with the release of "The Queen's Gambit" and a rise in chess related content creation on platforms like Twitch and YouTube. This correlation suggests a significant impact of media representation on public interest. The spike in searches not only highlights the power of modern media in shaping public hobbies and interests but also illustrates a renewed enthusiasm for chess, a game often perceived as traditional and static. Our findings underscore the dynamic relationship between cultural phenomena and digital engagement, showing how media can breathe new life into classic games, connecting them with a broader, more diverse audience in the digital world.

Data description

After clean data, the dataset's size is 4299 bytes(8KB) including 263 rows and there are three attributes/columns: 'Week', 'The Queen's Gambit: (Worldwide)', and 'chess: (Worldwide)'.
Data Types Present: The 'Week' column contains date information, while 'The Queen's Gambit: (Worldwide)' and 'chess: (Worldwide)' contain numerical data, which may include integers and "<1" notations indicating values less than one.

After the data was collected it was imported as a csv file into google sheets. The first step was cleaning the data to find any figures misrepresented or redundant that could hinder the visualisation of the data. The data ranged from the year 2018 until current day, 2023, and went week by week showing google hits on both searches. 2018 felt too far away from the trends and correlations we were seeking so the data was divided to only show searches from 2020 till 2023. This gave a more focused window on relevant information for the data visualisation, as the show, Queen's Gambit, did not exist prior to 2020, it was redundant to our data. Though the show Queen's Gambit did not exist until November of 2020, the whole year was kept as it allowed us to see the comparison pre and post show. How drastically the numbers changed and see how chess had a constant number of search hits even before the spike in hits. Throughout the information table there were figures that showed up less than zero, i.e <1, which is not fully understood what was meant by such a figure but needed to be

pruned from the data. Keeping that value changed the column from an integer column to a string column which caused issues when importing it to Tableau. This hindered the way the information appeared so it had to be changed to be more fitting. In mathematical terms, <1, denotes a value less than 1, however since negative searches cannot occur the values were replaced to 0. This gave a better fit and allowed the information to appear better in the graph without creating issues with importing it. Once the data was cleaned, by making sure there were no blanks in the data it was ready to be processed into Tableau to be created into a graph fitting for the data. There was information from a youtuber named GothamChess who has a contribution to another spike in chess in the year 2023. However not enough information was available to help find a clear correlation between the two. The later spike in popularity is due to Twitch,Youtube and content creators.

Visualisation

Initially there were a few ideas for the data visualisation, such as scatterplot and bubble chart, in the end we ended up with a simple line chart. A scatter plot would be beneficial to see how many hits occured in that week and be able to spot the relationships in trends but could be visually unappealing with a lot of points on the graph. The same idea with the bubble chart, it could help to denote areas of high volume where the numbers peaked for chess but could become very crowded on the chart making it difficult to read. By using a line chart, it allowed us to be able to compare the differences in their trends. By using a line graph we can easily follow the trends by allowing the x-axis to be the timeline, a week by week timeframe, and the y-axis to be the two trends we want to observe. The line chart in tableau also allows us to be able to hover over each week and see the amount of hits that occurred week. .

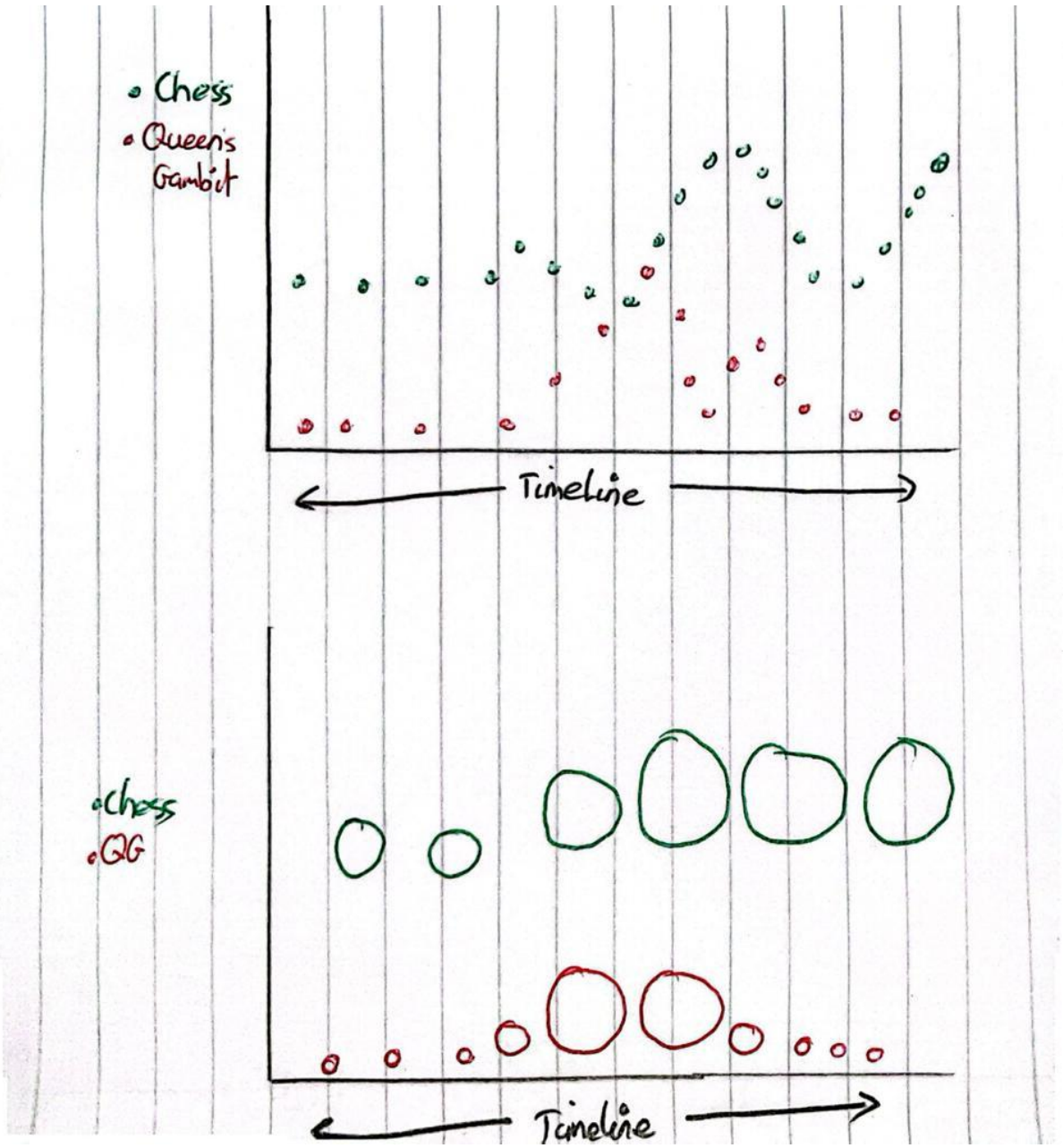

Figure 1

By using a line graph we can easily inspect the two different trends and their correlation, we can overlap the trendlines and see how the chess trends change in relation with the insurge of Queen's Gambit viewership. This gives us an easy way to simultaneously see the difference and similarities in the graph to analyse the direct behaviour they have on each other.

With this graph we aim to show the impact the Queen's Gambit show run had on the popularity of chess. The data used in our graph was from Google Trends which consisted of how popular a search hit is. It is on a scale of 0 - 100, the higher the more searched it is. The time frame for the searches are done in week increments, it shows how many hits occur in a week. This allows us to have an accurate way to see how the hits change week by week.

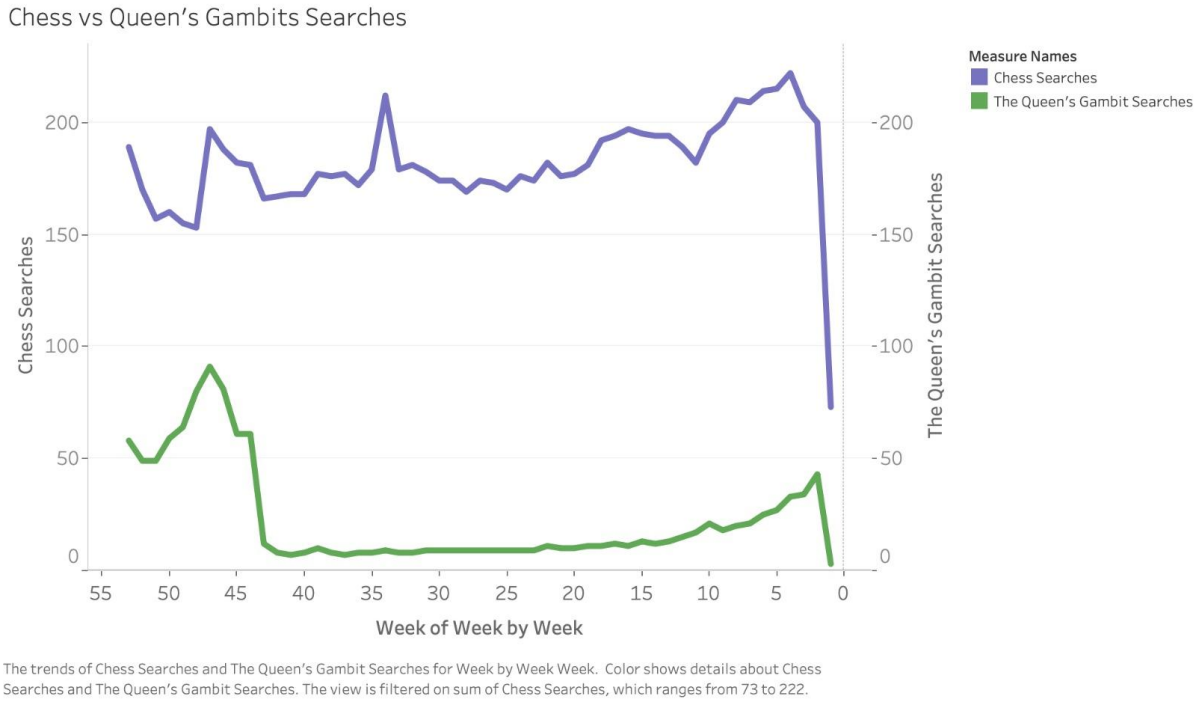

Figure 2

From this we can observe that during 2020 that the number of chess searches increased majorly because of the TV the Queen's Gambit entry into the scene. It creates this drastic increase in the interest of chess and continues to grow even after the show's run has ended. In its entirety of its run the show had 62 million, even chess.com saw an increase of users into the site. From October 2020 to April 2022 their numbers have doubled from 8 million to 17 million users.

For colour changes and some of the layout. Since our project is a line chart, purple and green were used to complete the differentiation, and the lines were chosen to be medium thickness so that they would not be too thin and difficult to read. For the three axes, I edited them to Auto Range to accommodate different devices for reading the chart, and set the purple line for Chess Searches and the green line for The Queen's Gambit Searches to synchronise the axis to better contrast the changes in the lines. Moreover, I set the Scale in Axis' format to a dark grey solid line to highlight the data. I tried setting the Ticks Scale of the Periodic Axis to Fixed, but our project is on a weekly cycle rather than a yearly cycle so this didn't work. For the borders on either side of the icons, I set the Row divider and Column divider on the left side of the chart to 'None' to make the edges of the chart look neater.

Conclusion

List of Tools or Libraries Used:

- Tableau Desktop: The primary tool for our data visualisation project, offering a platform for creating a line chart.

- Microsoft Excel: Used for the crosstab export of the data from Tableau which allows for further manipulation or detailed analysis if needed.
- Google Sheets: Initially used for data cleaning and preparation before importing into Tableau.

Critical Analysis

The visualisation highlights the trends in Google search volume for chess related terms, especially in relation to the release of "The Queen's Gambit". The use of line charts enables a temporal tracking of interest over time which directly supports the narrative of the project.
The juxtaposition of the search volumes for 'chess' against 'The Queen's Gambit' allows for a direct comparison, illustrating the impact of the series on the popularity of chess. The use of contrasting colors in the lines helps in distinguishing the two different datasets.
The decision to narrow the timeframe to post 2020 data was judicious, focusing on the most relevant period for the story being told.

Aspects for Improvement

- Including interactive elements such as filters or highlighters could allow viewers to engage with the data more meaningfully. For example, viewers could be given the option to select specific time frames or events to see their direct impact on search volumes.
- Adding annotations to significant data points, like the release date of "The Queen's Gambit" or major chess events, could provide context and enhance the narrative.
- Comparing the chess search trends with other similar trends triggered by cultural events could provide a broader context on the power of media influence.

In the context of a data visualisation project, there may be several desired effects or functionalities that could not be achieved due to technical constraints or limitations of the tools used. Here are some limitations for our project and how they could potentially improve a visualisation if addressed:

- Tools like Tableau we used for our project offer a wide range of interactivity features, but they might still be limited compared to custom code visualisations. For instance, implementing more complex interactive elements such as dynamic sliders that adjust the granularity of the time scale could offer deeper insights.
- Moreover, it will be better if we can do real time data streaming for our project, but the ability to incorporate real time data streaming can be a challenge. Real time updates would allow the visualisation to be continually current, reflecting the latest search trends without manual refreshing.
- The part that was tried but ultimately rejected:
  When we chose the topic we were going to analyse four groups of activities such as chess,volleyball, baseball and football. In the case of baseball, we chose the correlation between Japan's most typical Koshien summer tournament and the sales of Japanese baseball comics, but we found that the amount of data collected was not

enough to support us to complete the data visualisation of the project, so we finally decided to focus on one cultural activity only: Chess.

The following dataset, which we failed to collect midway through the data collection

	Japan Koshien Collegiate Baseball Conference Summer Tournament	Opening ceremony Viewing rate (%)	Finals Viewing rate (%)	Number of participating schools	Total number of visitors
2023	105th National High School Baseball Championship - Final	6.7	19.3	3486	10000
2022	104th National High School Baseball Championship - Final	7.4	13	3547	566500
2021	103rd National High School Baseball Championship - Final	8.5	9.4	3603	10000
2020	102nd National High School Baseball Championship - Final				
2019	101st National High School Baseball Championship - Final	10.3	15.5	3730	841000
2018	100th National High School Baseball Championship - Final	13.5	20.3	3781	1015000
2017	99th National High School Baseball Championship - Final	8.4	17.1	3839	827000
2016	98th National High School Baseball Championship - Final	12.2	19.6	3906	837000
2015	97th National High School Baseball Championship - Final	11.6	20.2	3906	862000

Figure 3

One of the hottest selling baseball manga in Japanダイヤの A			
num of volumes	Number of initial sales(within 7 days)	Final number of copies sold(within 7 days)	release date
1	1210	1542	21/05/2014
2	4964	5675	18-Jun
3	4671	5366	16-Jul
4	4872	5596	20-Aug
5	4905	5737	17-Sep
6	5136	5740	24-Oct
7	5690	4237	17-Dec
8	2967	3291	18/02/2015
9	3100	3730	24-Apr
10	3508	3925	20-May
11	3001	3307	17-Jun
12	3058	3379	15-Jul
13	3292	3610	19-Aug
14	3477	4293	16-Sep
15	2798	3376	21-Oct
16	2850	3780	18-Nov
17	2816	3127	16-Dec
18	2853	2684	20/01/2016
19	2902	2180	17-Feb
20	2406	2453	21-Mar
21	2715	3121	20-Jun

Figure 4

- Lastly, if there's a need for custom statistical analysis or computations, Tableau's built in functions might not be sufficient, and integrating with a more flexible computing environment like R or Python could be necessary.

Collaboration

We collected the dataset together, and the final dataset used was found and finished data clean by Rufus.

In the tableau data visualisation project, Rufus did the main visualisation tasks, drawing the line graphs and Jiaying assisted in modifying the colour and layout of the line graphs and adding some labels.

References

Below are the data sources used for data visualisation in the chess project:

https://trends.google.com/trends/explore?date=today%205-y&q=%2Fg%2F11h8yyk4xy,chess

https://www.nytimes.com/2022/06/17/crosswords/chess/chess-is-booming.html#:~:text=Chess%2C%20the%20granddaddy%20of%20all,million%20to%20nearly%2017%20million.

The part that was tried but ultimately rejected source
https://www.videor.co.jp/tvrating/past_tvrating/

https://note.com/entamenodendo/n/n8e5a6009c400

https://www.tsp21.com/sports/mlb/feature/tvrating.html

https://www.kaggle.com/code/darkaxe/starter-myanimelist-cleaned-dataset-24aba8aa-a/notebook

https://sakidori.co/article/1577915

https://www.lanovelibrary.com/diaace/