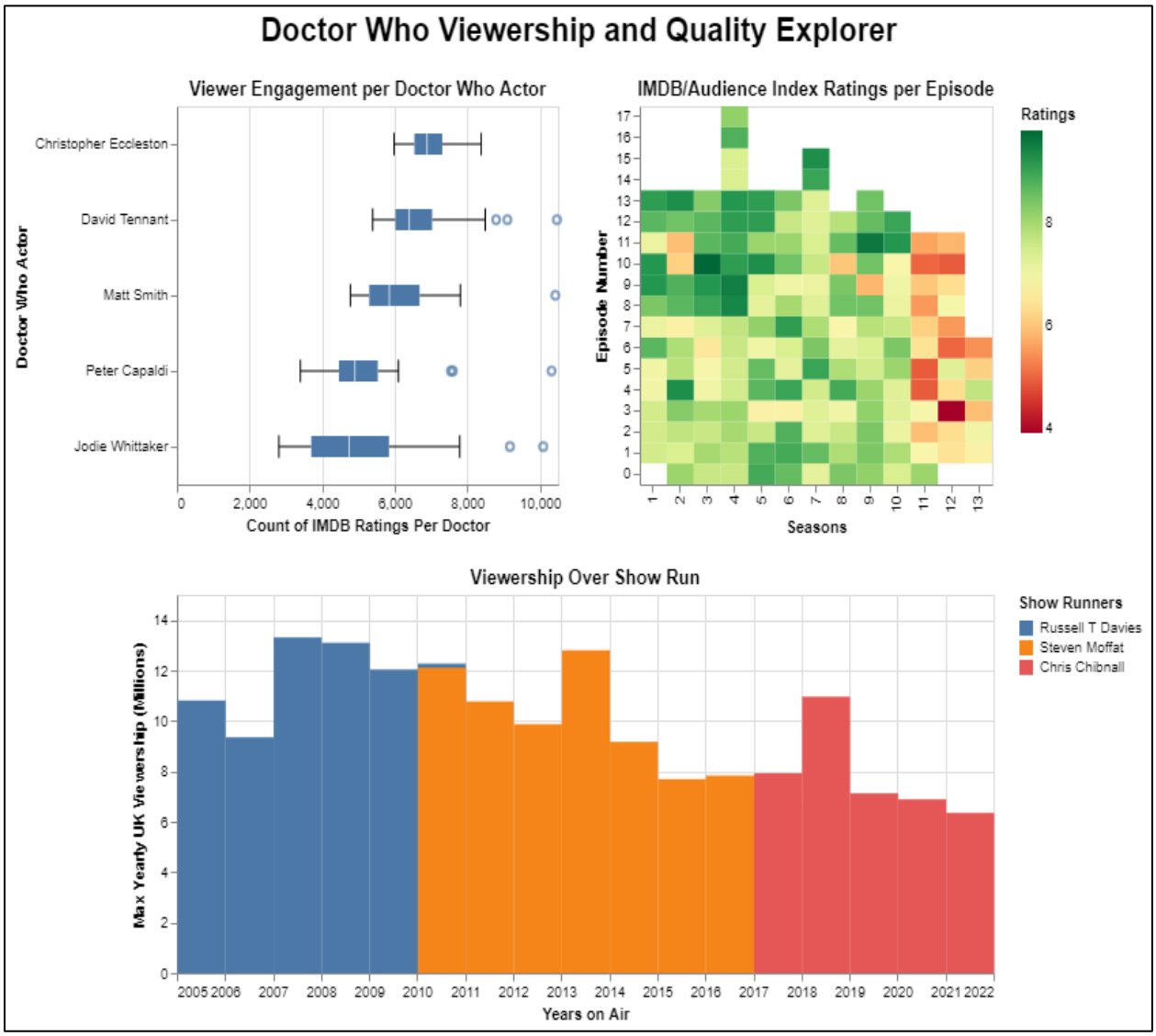
COMP40610 Visual Exploration Tool Design Document

Student Name: Thomas Pollock

Student Number: 17483996

Title: Doctor Who Viewership and Quality Explorer

Screenshot:



Data Overview:

This data was taken and derived from the Tidy Tuesday 2021 data project, specifically the "Dr Who" section from the 23rd November. This data was originally split into 4 csv files, (imdb, episodes, directors and writers) but were joined together into a new master csv file "doctor_who_master_dataset.csv" and uploaded to a personal GitHub repo after a data cleaning and preprocessing process using Python. Data from this master dataset used in the dashboard includes episode titles, air dates, UK viewership from those air dates, IMDB average ratings, rating counts and episode descriptions. The features for the doctor who actor and current show runner/head writer were added manually using Python as they were necessary to create the boxplot and histogram for the dashboard.

Data Considerations:

Overall goal:

My goal for this tool was to allow users the ability to explore the relationship between the shows viewership and several key factors, including the current head writer (Showrunner), the perceived quality of episodes (IMDB ratings), the current actor playing Doctor Who (Doctor Who Actor) and time.

Boxplot chart:

This chart displays the median count of IMDB ratings per episode for each actor who's played Doctor Who. As a boxplot, it also includes the minimum, maximum and the interquartile range of these counts to give more precise insight into the fanbase's impression of each actor's interpretation. From top to bottom, the actors are ordered from the first iteration of the character in the revived series to the newest one. The main objective of this chart is to demonstrate fan engagement with the show through extrapolating that those who give an episode a rating online will tend to be more generally engaged with the show compared to the general audience. Less ratings on an episode will indicate that fans are more apathetic to the episode in question regardless of the quality and more likely to drop off. A simple horizontal bar chart was considered in lieu of the boxplot; however, the precise insights gained from the boxplot like the extremely high outlier for the female actor Jodie Whittaker's first episode "The Woman Who Fell to Earth" indicating excitement for the first female doctor would have been lost. A vertical version of this chart was considered to emphasise the decreasing engagement with each proceeding doctor, however the long actor names on the x-axis would have misaligned the chart from the other two and vertical names would have been more difficult to read.

Heatmap chart:

This chart displays the average IMDB and Audience Index (AI) ratings for each episode of the show. Each cell on the heatmap represents a single episode, with a green/yellow/red gradient colour encoding to represent high/medium/low ratings respectively. This gradient was chosen

as it is the most culturally common method to denote good/mediocre/bad scores or ratings. A parameter for the heatmap is set up so a user can pick between an IMDB rating colour encoding or an AI one. The AI rating was included to give a direct insight into the general audience feeling of episodes at the time of airing. Episodes in the 0 row of the heatmap indicate Christmas specials. Since these are part of their respective seasons but not technically included in a seasonal run order, indicating these episodes with a 0 seemed the most appropriate; episodes released prior to the start of a season. A scatter plot with the IMDB/AI ratings on the y-axis with episode air dates on the x-axis was considered. However, as episodes were aired within a similar 3–4-month yearly timeframe, episodes would cluster together making it difficult to discriminate between them. The colour encoding of the heatmap for the ratings was also more visually intuitive to understand compared to the positional encoding due to the cultural green/red, good/bad association and the use of a heatmap allowed to better interactivity pairing between itself and the time-series histogram used for viewership.

Histogram chart:

This chart maps the highest viewership of an episode within a yearly timeframe. Colour encoding is also used to differentiate between different showrunners' time on the show. This colour encoding was included to compare and correlate the IMDB/AI ratings of the heatmap to any viewership gains or losses with each proceeding showrunner. The highest viewership of an episode is used in lieu of the mean or median UK viewership as Christmas specials have a much higher viewership on average compared to the rest of the show. The year-on-year difference between specials also gives a better view of the viewership numbers over time, as use of a mean or median of the high viewership of Christmas specials with the lower viewership of the main season would cancel each other out, giving the impression that the show is maintaining its viewership when it is actively losing it over time. A line plot was considered but time gaps between seasons and clustered episodes within 3–4-month timeframes produced erratic and eligible graphs.

Interaction consideration:

For interactivity, the main approach was to implement cross filtering between charts. Interval selection boxes were implemented for both the histogram and heatmap charts to filter out unselected data from the remaining two charts. The heatmap allows you to select a range of episodes to explore their viewership or online engagement concurrently. Selecting only one episode allows you to see its viewership number on the histogram and rating count on the boxplot. The histogram is broader, allowing you to see the IMDB ratings and the number of them within a year timeframe highlighted on the heatmap and boxplot respectively.