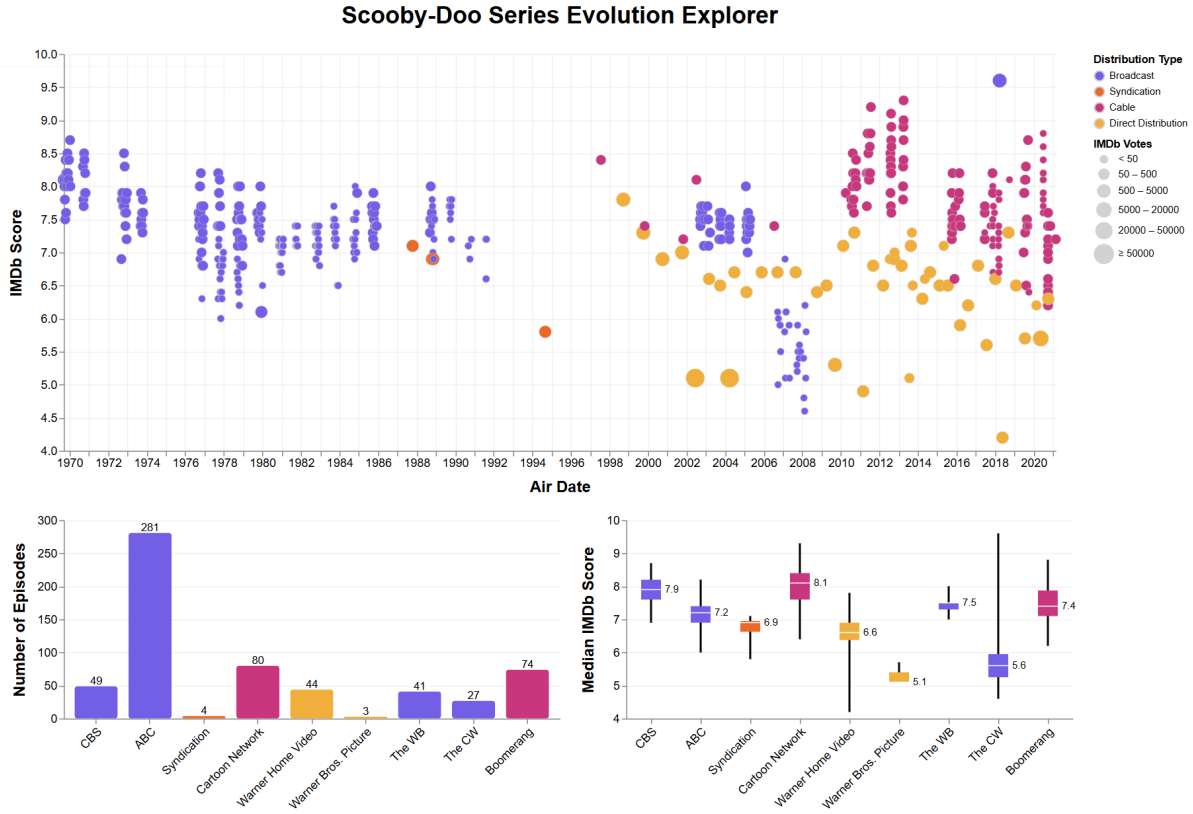
Design Document



Dataset Overview

This dataset has been processed from Tidy Tuesday Week 29, which chronicles Scooby-Doo episodes aired from 1969 to 2021. Relevant features include episode titles, series names, air dates, broadcasting networks, format types (TV series, movies, specials, crossover), IMDb scores, and IMDb vote counts. Data cleaning involved handling 15 Null values (2.5%) in IMDb scores and engagement, which were filled with current values obtained from IMDb rather than using median imputation. Additionally, several entries for specials and movies produced by Warner Home Studio were incorrectly labeled as TV series broadcast by Cartoon Network, which were identified and corrected using conditional logic in Python. As for data transformation, a new 'Distribution Type' feature was engineered to address inconsistencies in the network categorization. This transformation was necessary since the original 'network' feature included 'Syndication' – a distribution method that did not align with the standard network categories. The transformation involved researching each network and mapping it to its appropriate distribution type, providing a standardized way to analyze the quality of distribution channels and their airing networks of Scooby-Doo content.

Design Considertaions

Overall Goal

This tool aims to explore the relationship between distribution channels, networks, audience reception (IMDb scores and votes), and production volume (episode counts) of Scooby-Doo's evolution from 1969-2021. Designed for both longtime fans with basic data literacy and media professionals examining content distribution patterns, the tool helps uncover how distribution strategies have influenced the series's development and reception: how distribution channel/network shifts influenced overall quality; which networks achieved the best balance of quantity and quality; how audience engagement patterns have shifted over different eras of the show.

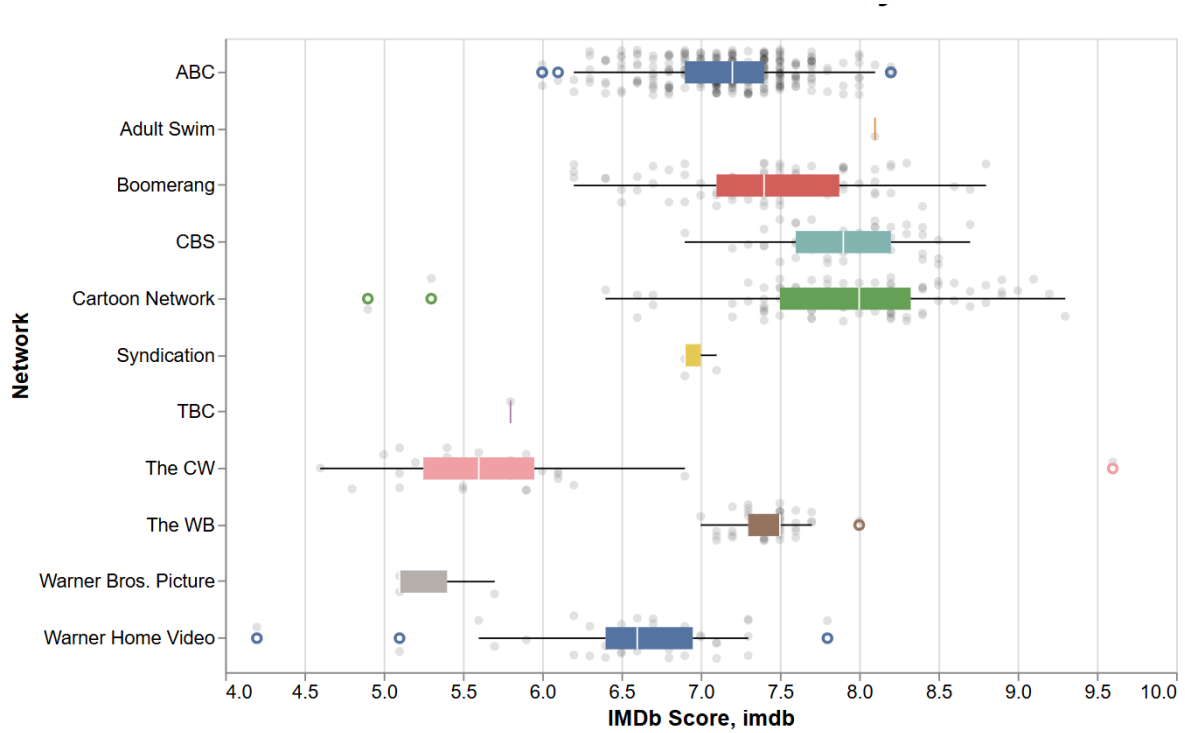
Layout

The three-panel layout is designed with an emphasis on information hierarchy and analytical flow; from the top temporal overview and legend to the below two complementary detailed views. The main bubble plot serves as the primary temporal overview and interaction point. The supporting views, including the bar chart and the box plot, provide statistical summaries. The author initially considered placing the overview alongside a supporting view, but canceled it due to its reducing the temporal plot's width, making pattern recognition harder.

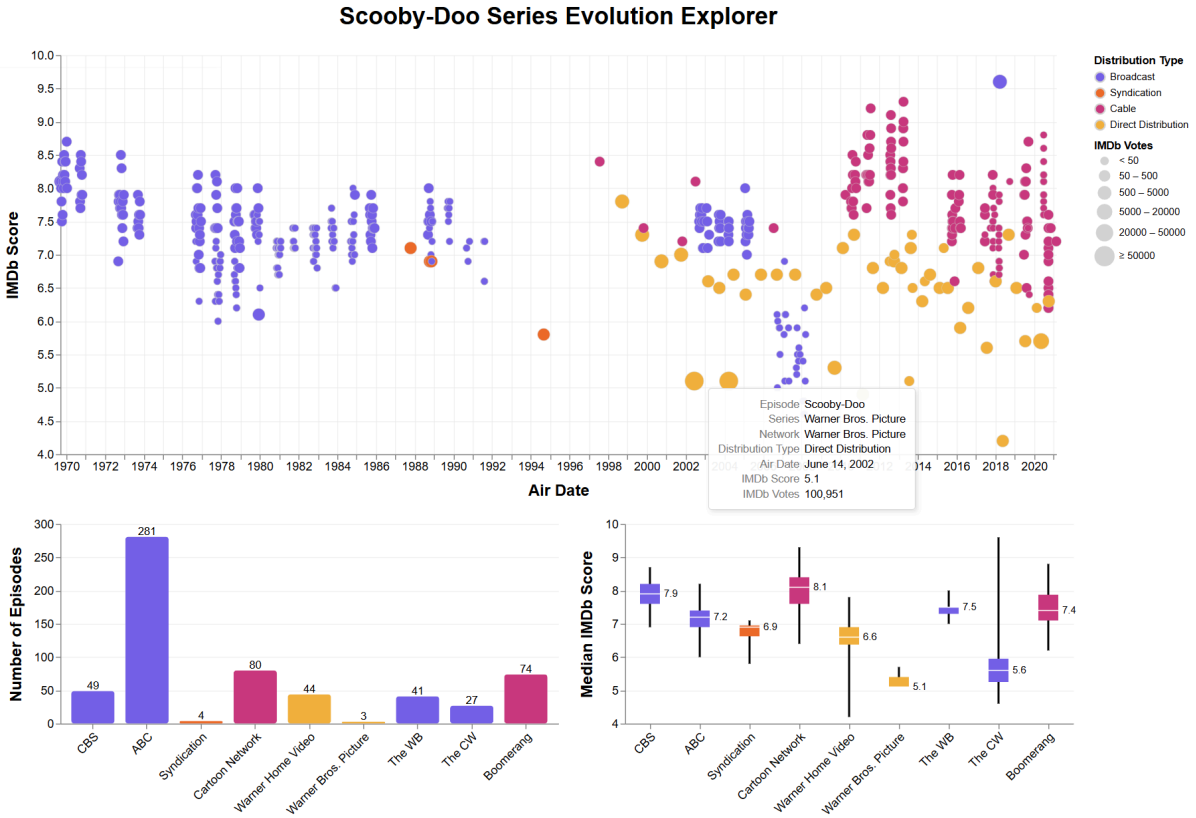
Bubble Plot & Legend

The bubble plot visualizes IMDb scores (y-axis) of Scooby-Doo productions (TV series, movies, specials, crossovers) against their air dates (x-axis). Both temporal patterns and scores are encoded through position - the most effective channel for precise reading of quantitative data. Each production is represented by a bubble, with colors mapped to distribution types (and their networks in supporting charts) and sizes reflecting IMDb vote ranges. The plot uses a colorblind-friendly IBM palette with four distinct hues for optimal visibility. The legend is arranged chronologically to maintain temporal progression and reduce cognitive load by avoiding adjacent similar colors. Color encoding was chosen here as it effectively distinguishes between categorical values and can be consistently applied across all three views. To address skewed vote differences (75th percentile: 123.5; maximum: 100,951), bubble size encoding is non-linearly split into six bins (<50 to 50k+), using a logarithmic-like approach rather than continuous scaling to compress high value ranges and emphasize low-value details for better comparison.

The bubble plot was chosen over scatter and strip plots overlaid by box plot due its capacity to encode three quantitative variables.

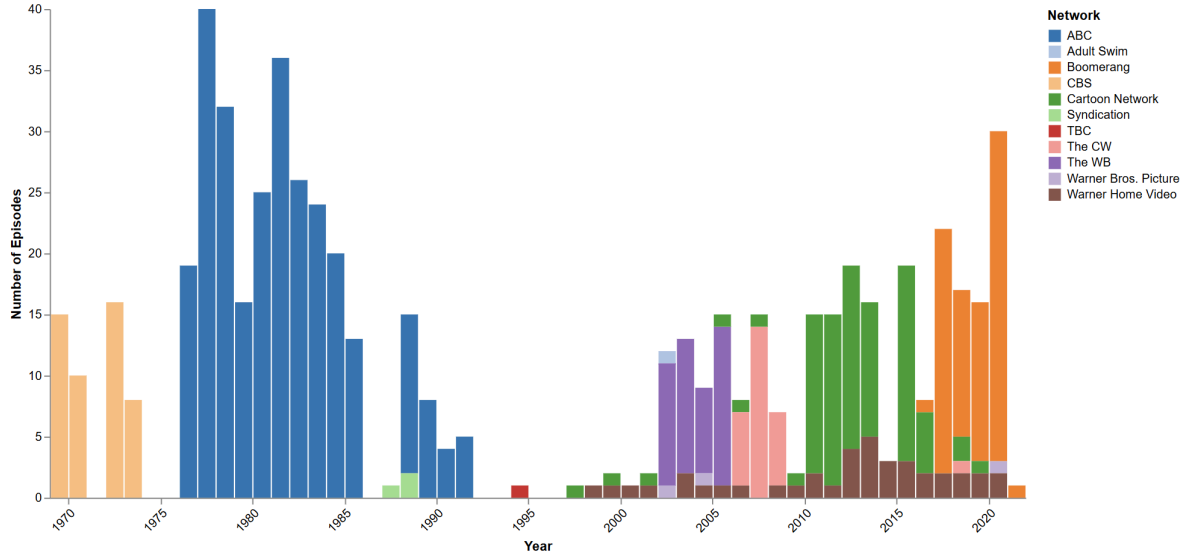


The bubble plot not only allows individual episode visibility with tooltips for inspection like its alternatives but also enables viewers to observe how fans' engagement (votes) relates to fans' perceived quality (score) and time without requiring a separate visualization.



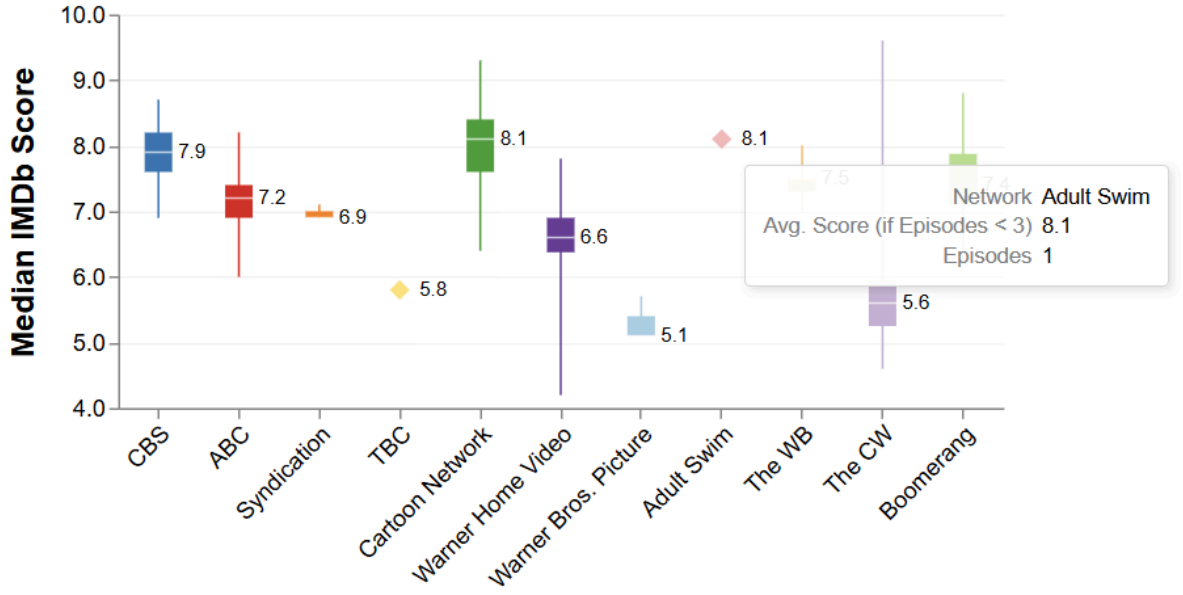
Bar Chart

This bar chart uses position and length - the most effective channels for quantitative comparison - to show production quantity. It aggregates the total episode counts (y-axis) aired by each network (x-axis) color-encoded to match distribution types. The chart is horizontally placed and sorted by chronological order to preserve historical context, aligning with the overview and legend. Compared to descending order, the main disadvantage is network comparison being not immediately intuitive to viewers. To alleviate this issue, the author added data labels to help reduce cognitive load and balance data-ink ratio. A stacked bar chart showing temporal distribution of episodes serves as an alternative but got discarded since the current option is easier to interpret and temporal patterns are better captured by the bubble plot.

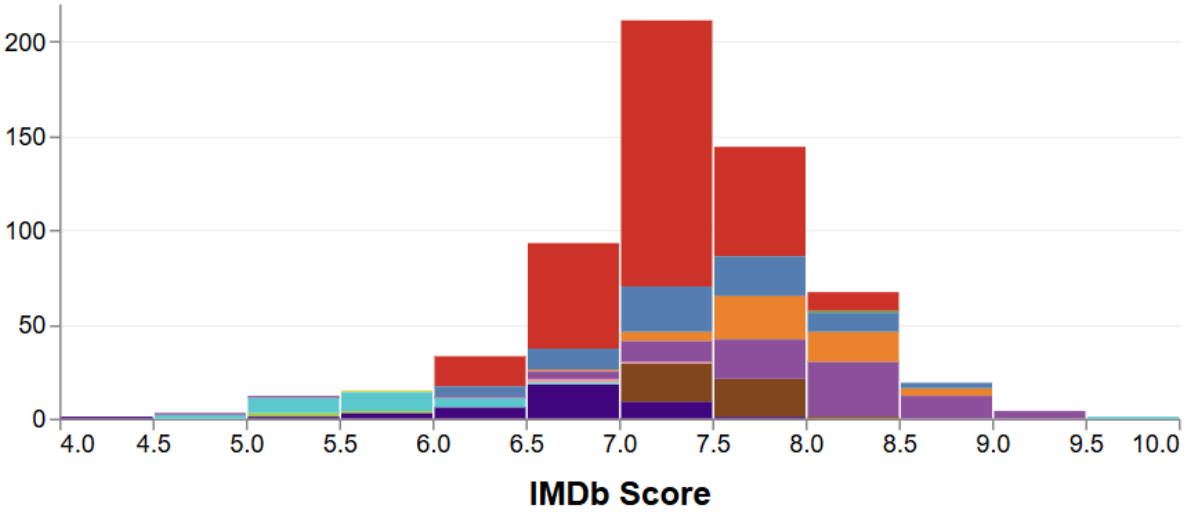


Box Plot

The box plot illustrates the IMDb score distribution (y-axis) for networks (x-axis), following both chronological sorting and distribution types' color encoding for consistency. Its whiskers shows full distribution shape while the white median lines with data labels ensure ease of interpretation. If certain networks have only one or two episode counts over time and cannot calculate median, they are encoded as diamond markers with average score instead. This approach provides clear statistical data while working well with brush filter as it dynamically re-calculates small sample size instead of showing Null values. Detailed calculations for quartiles, mean, median, and minimum-maximum scores can be inspected in boxes' tooltips.



A stacked histogram is a decent alternative if the focus is on common rating ranges of episodes by networks, which is not visible in the box plot. Nonetheless, it is more complex to read and compare networks, while the simple box plot provides clear networks' performance summary to complement the main overview.



Interaction Design

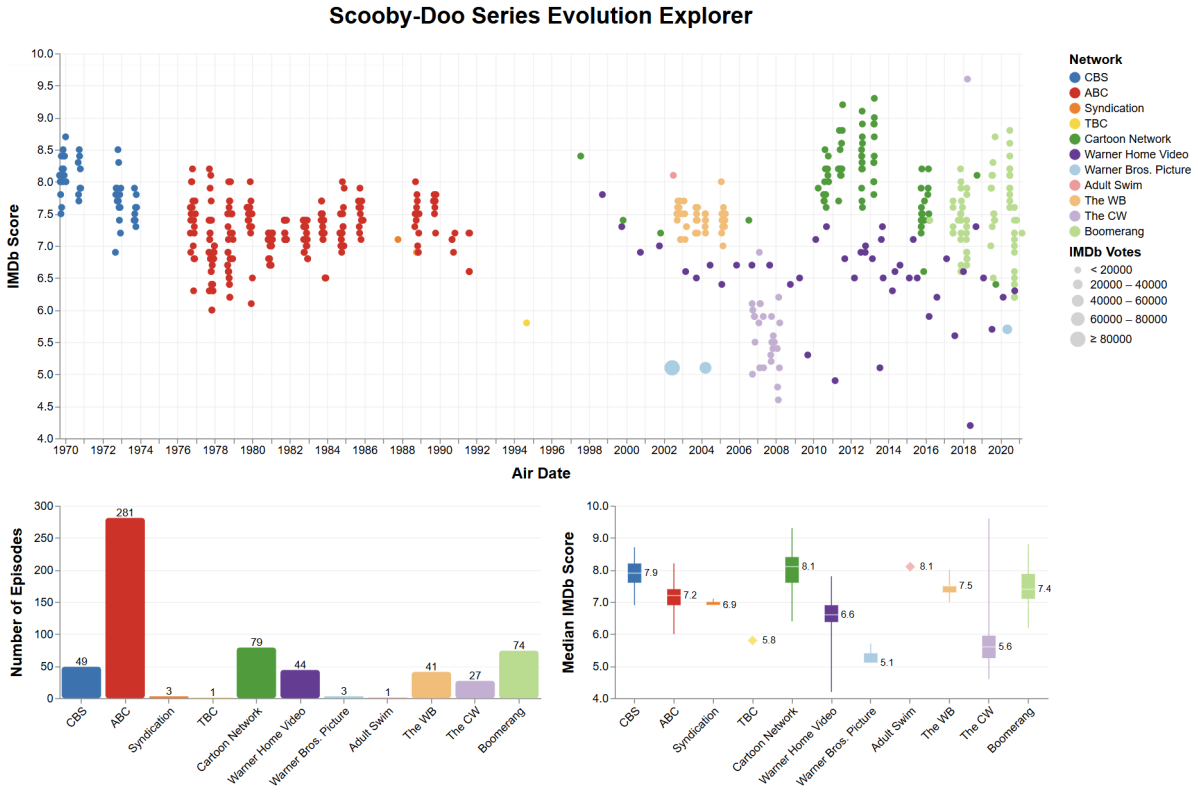
The dashboard follows a linked-view structure with bidirectional cross-filtering with grey and opacity encoding for unselected data to preserve context: Time range selection (bubble plot's brush), distribution filtering (legend), network filtering (bar click). When using brush selection to choose time periods, both bottom views are synchronized to reflect all episodes in the chosen periods. The distribution and network filterings can work simultaneously with the brush to apply data transfomation across all three views. This design supports both broad pattern discovery and detailed investigation as each view's strengths compensate for others' limitations. For instance, temporal analysis is possible by brushing time periods to compare network performance. Viewers can also perform distribution/network comparisons

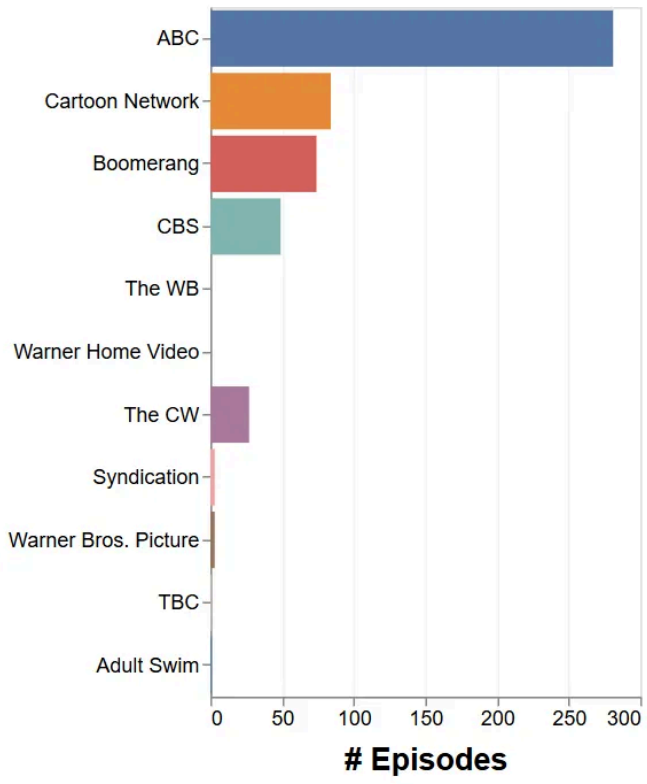
by selecting specific distribution channels/networks to analyze their temporal patterns and distribution analysis by examining score rating patterns within filtered subsets.

(Word count: 999)

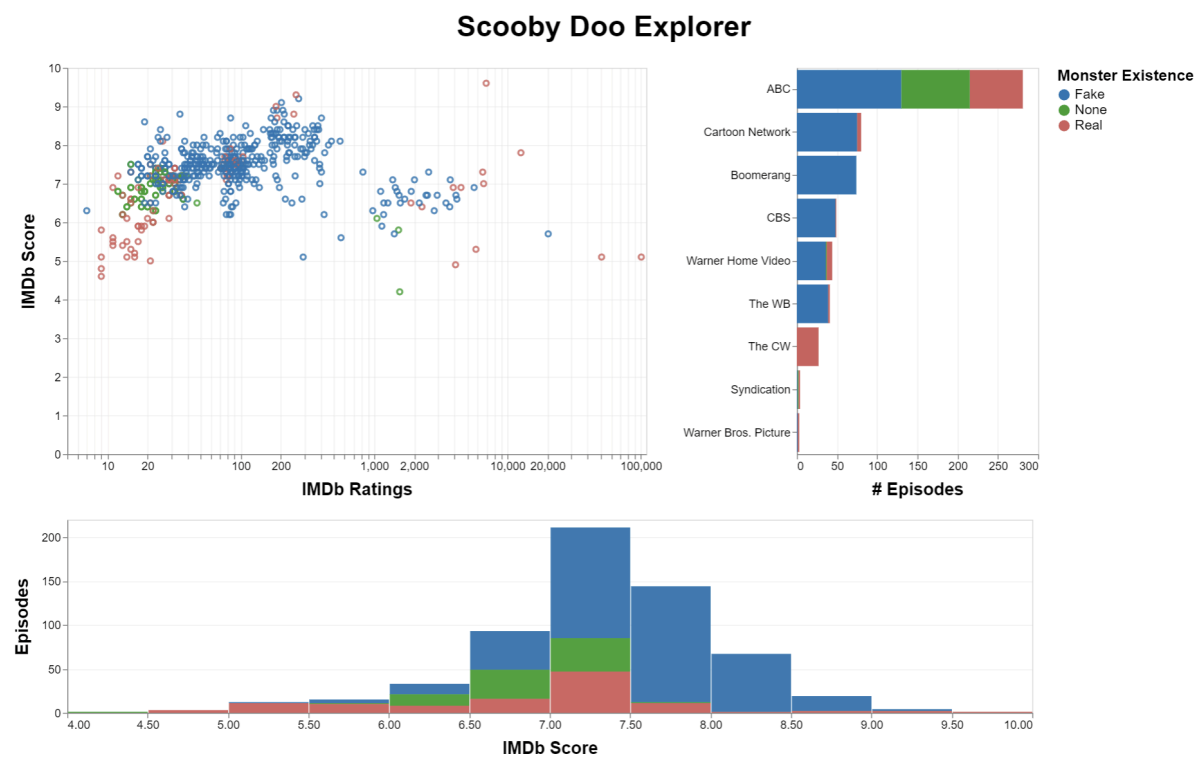
Appendix

A collection of additional alternative chart designs that did not survive natural selection






Median IMDb Score
Year
Network
ABC
Adult Swim
Boomerang
CBS
Cartoon Network
Syndication
TBC
The CW
The WB
Warner Bros. Picture
Warner Home Video


Scooby Doo Explorer
IMDb Score
IMDb Ratings
ABC
Cartoon Network
Boomerang
CBS
Warner Home Video
The WB
The CW
Syndication
Warner Bros. Picture
Episodes
Monster Existence
Fake
None
Real
Episodes
IMDb Score