FANDANGO LOVES MOVIES: Cleaning The Data

<https://fivethirtyeight.com/features/fandango-movies-ratings/>

# Deceptive Data

* Only displays movies that have been rated 30+ times by users on Fandango, this restriction is not put on the other five movie review sites.
* This creates discrepancy where Fandango ratings are more statistically reliable with more data points than the other movie review services.
* This creates biased or skewed distributions if movies are based on just a handful of reviews.
* Can disproportionately influence the shape of the graph, creating the illusion of uniformity or difference that isn't fair.

Why it’s an Issue?

This introduces selection bias, making Fandango’s ratings appear more stable or positive simply because low-rated or less-reviewed movies are excluded. As a result, the data can create a misleading impression of consistency or quality across platforms, distorting the overall analysis and making the visual comparisons unreliable.

Solution

Only use movies that have been reviewed 30+ times by all movie review sites.

# Overplotting

* Six-line plots overlapping each other.
* Annotations creating more cluster on the visualization.
* Dark grey highlight covers the line plots.

Why it’s an Issue?

Because of the overplotting, and clustering of data, it becomes difficult to read the visualisation and understand the message it’s trying to convey. This noisy visualization makes it less accessible for the vision impaired.

Solution

Facet the data by each movie review site, remove grey highlight, and add headings above each graph facet instead in the line graph.

# Lack of Insight

* The visualization shows general trends where Fandango rates higher overall. But it struggles to show the audience how much more it favours movies compared to other sites.

Why it’s an Issue?

Without clear patterns, context, or comparisons, the visualization becomes just a display of numbers rather than a tool for understanding. This can lead to confusion, misinterpretation, or missed opportunities to uncover trends, relationships, or important findings in the data.

Solution

Annotate the visualization with black dots that represent a movie and how each of the movie review sites rate it depending on its x axis position. Allow the audience to choose the movie to compare to introduce an interactive data visualisation using Shiny instead of choosing a movie by default avoiding bias.

# Accessibility

* The visualisation used red and green colours. This becomes an issue for colour blind audiences.

Why is it an issue?

The visualisation becomes unreadable for the colour blind, and they will not be able to understand the message the visualisation is trying to convey.

Solution

Use a colour-blind friendly palette that has no reds or greens. The colour-blind palette used was collected here:

<https://cambiocteach.com/accessibility/colourchoice/>