Data Visualization - Assignment 2 - Critique

Alireza Mostafizi

May 6, 2018

1 Critique

For this assignment, I decided to improve on one of the plots I created in the the first assignment using **IMDB Movies** database, loaded from the library *ggplot2movies*. Figure 1 shows the original plot, that is meant to communicate which genres are made more or less over the years. However, this plot violates a few key principles, explained hereunder.

1. Drawing the comparisons of interest from this plot is rather hard. To understand the changes in the proportion of the movies that are labeled with a specific genre, one should divide the number of those movies in a specific year, read from colored curves, by the total number of movies, read from the blue bars. The same process has to be done for each genre/type of movie and for many years, and the entire results should be compared at the same time. Therefore, it is very demanding to say, for instance, if the proportion of short movies are increased or decreased over time. We can conclude that for the comparisons of interest, the best **perceptual task** has not been employed.

2. Important comparisons are **not close** either. For example, to compare the proportion of Dramas in 2000 and 1940, the eye has to travel, after the division explained above, from middle of the graph to the far right.

3. Color has been chosen and employed as **Preattentive property** to distinguish between different genres and types. Also, the shape has been chosen as another preattentive property to show the groupings. The total number of movies are shown with blue bars, and each genre/type has been represented by the lines.

4. This plot also has **7 different colors** which is exactly on the boundary of number of colors. So, decreasing the number of colors would help. In addition, this plot won't be helpful at all if printed in **gray-scale**, and it does not have a **color-blind friendly** design.

5. The legends and the axis labels are not easily readable.

6. Axis ticks and the breaks are not helping and more or less are distracting. In addition, half of the plot is empty of data and filled with the major and minor axis breaks which decreases the "data-ink" ratio.

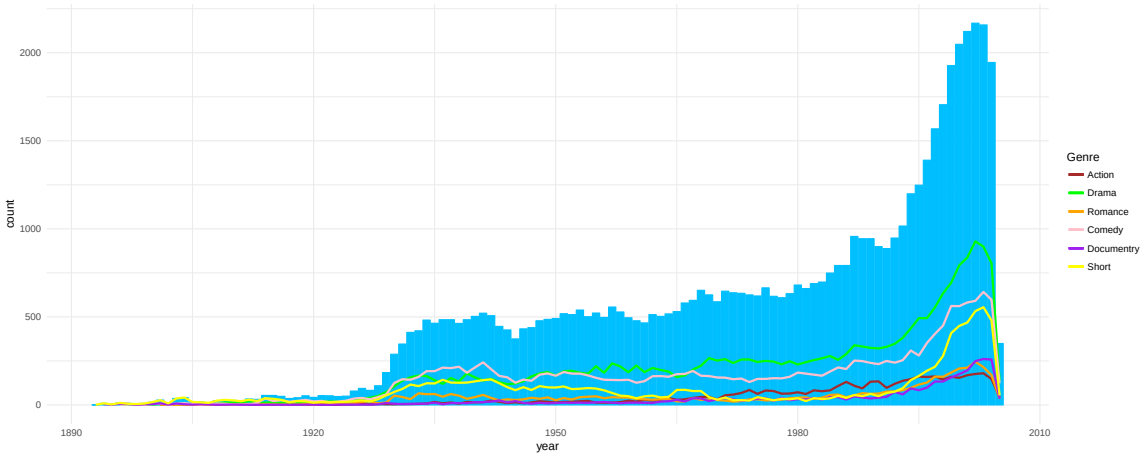

Figure 1: Movie Counts.

2 Improved Alternative

Figure 2 shows the improved plot. Here are the changes and improvements compared to the original plot.

1. The most important improvement is the change in presenting the data, and as a result, change in the **Perceptual Task** employed to make the comparisons of interest. The task to see the changes in shares, and the variation in proportions over the years is to read the *position on common scale* which is the highest task.

2. Moreover, the comparisons are kept close by limiting the data shown in the plot to only from 1940s to 2000s. This has removed almost 25% of useless plot that cannot be read due to relatively small numbers compared to higher ends of the plot. This problem has also been alleviated by the alternative value of proportion, as opposed to counts in the original plot.

3. Another improvement is the use of a **Color blind Friendly** color scheme using *viridis* package which covers a blue-yellow spectrum. In addition, the legend has been included in the plot as annotations to help color-blind audience to recognize which line represents which genre. Moreover, this makes the plot readable in grey-scale. The number of **Colors** has also been decreased to 6, as opposed to 7 in the original plot. But the use of legend annotation, alleviates the large number of colors used.

4. To avoid having two scales on the plot and complicating the main take-away, the number of movies made in each year has been shown on a secondary axis on the top, which clearly emphasizes the huge increase from 1980s to 2000s. I would have used color as a **preattentive property** to highlight this extreme change, however, I was not able to do this in *ggplot2*.

5. The legends and the axis titles are **legible**, and the use of **caption** has cleared the point of this plot. The breaks and the y axis lines are removed to put focus on the data and to increase the **data-ink** ratio.

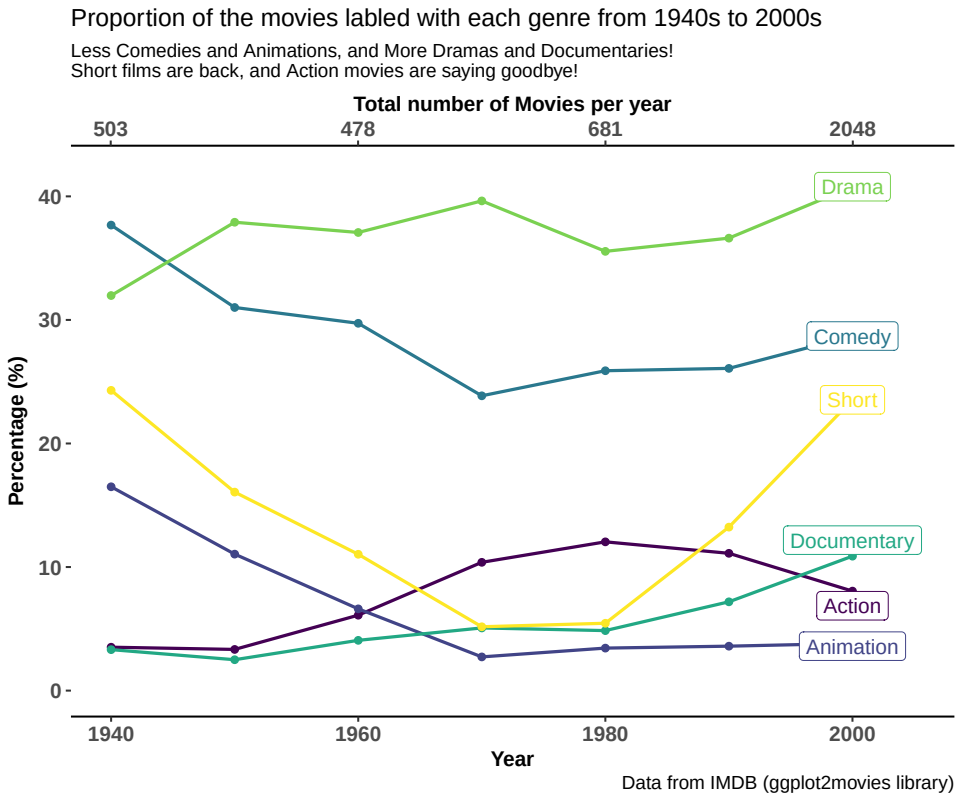

Figure 2: The improved alternative plot. Figure 2 shows the changes in the proportion of the movies made in a decade and labeled with each genre/types from 1940s to 2000s. It can be seen that more share of movies are labeled with Drama and Documentary; and less share of the movies are labeled with Comedy and Animation. The plot also shows a huge drop in short movies in 70s and 80s, and the revival of this genre in recent decades. In addition, the proportion of Action movies has been on decrease over the last couple decades.