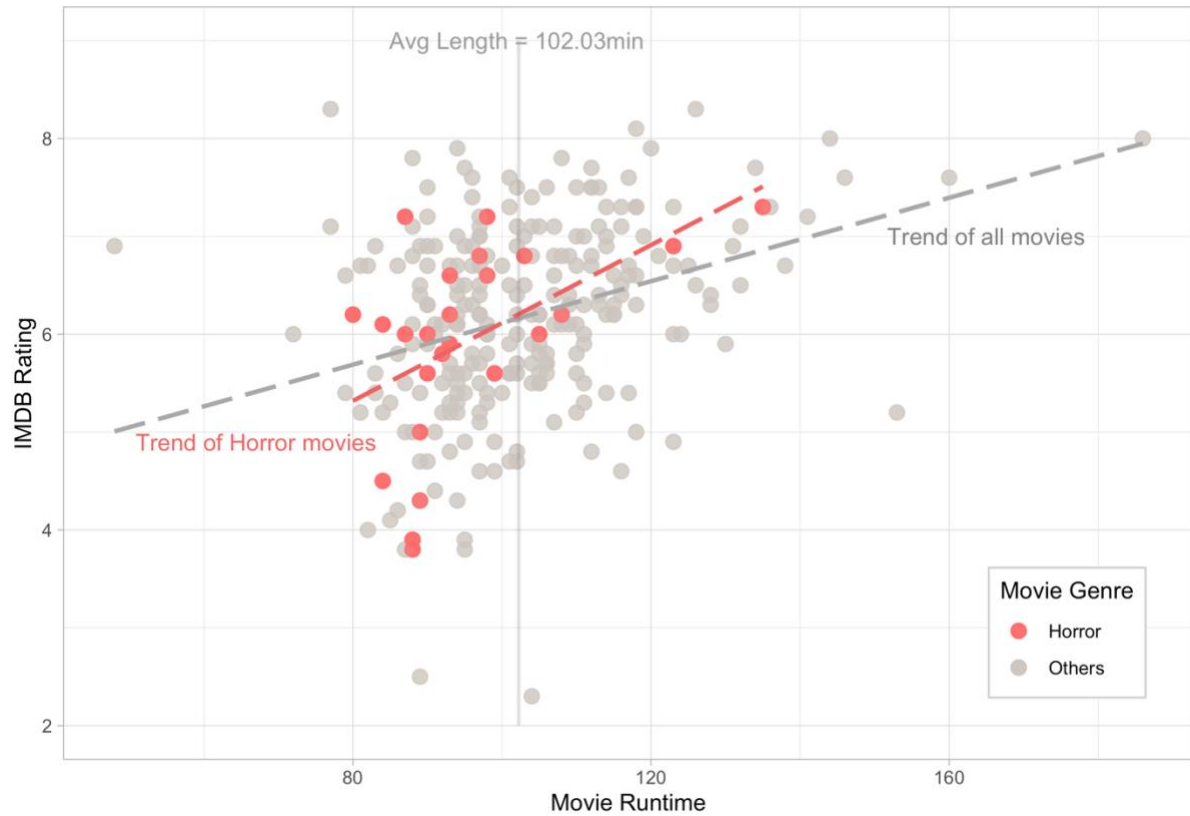


## Assignment 1: Visualization Design

Do longer movies have higher ratings? What about horror movies specifically?

Movies shot in Toronto from 1933 to 2019



My visualization aims to illustrate the relationship between movies' runtime and their ratings, especially for horror movies. The data comes from films shot in Toronto from 1933 to 2019. As clearly depicted in the graph, there is a general trend of rising ratings for longer movies. For horror movies, the message is better incorporated with a further slanted line, implying longer runtime tends to generate more popular horror films.

I decide to analyze the runtime and rating topic because I find it hard to balance movie length and plot completeness. Movies would be excruciating to watch if lengthy, while they could also be too short to explain the stories fully. For horror movies, they usually have more straightforward plots. The visualization exhibits that the audiences rate long-runtime films higher. It implies that a complete reveal of stories weighs more from viewers' perspectives, where the horror genre works in the same way. Directors should emphasize this factor when producing films.

To discover the correlation between movie ratings and runtime, I choose to present it in a scatterplot because both data attributes are quantitative and sequential. The two data variables were categorical originally. I change their characteristics to numerical to better reflect the information they contain. Additionally, I remove film observations without the IMDB rating or the runtime info since they do not fit the question. Since one of the visualization goals is to display the horror genre, I sort the data by setting all movies that contain the "Horror" element as horror movies. Otherwise, only two "pure" horror films belong to the genre, and the analysis would be meaningless with too few observations. To achieve effective communication, I draw two dashed lines to show the relationship trend for all movies in general and horror movies specifically. Since the grey one doesn't represent other genres, I include texts to explain the lines so that they will not obfuscate viewers. The lines are designed as dashed instead of full lines to reduce attractions since they are intended to help reveal the trends to answer the question. In like manner, I add a vertical-line segment to represent the average length of all movies. The line separates the graph into two sides, making it more visible that the right-side points locate higher as a whole. It indirectly facilitates better comprehension of the answers to the question.

Besides, I plant some ingenuity when deciding on colors. Since my questions focus on the horror genre, I believe the communication would be enhanced by popping out all the relevant data points. Hence, I choose to use red for the horror genre and grey for other movies. I also bring all the horror data points to the front so that they won't be covered under the grey ones. As shown, red is a conspicuous color compared to grey, which will attract the viewer's attention directly to the message. In terms of data point sizes, I enlarge them from the default setting so that it would be apparent to see how they are plotted and corresponded to what values. The scales are in default settings.

I make some tradeoffs when designing to avoid misleading communication. For instance, I thought about pointing out some outliers with short lengths and high ratings or the opposite. However, I found it may contradict my original message and confuse viewers about my intention to do that. Another alternative design is to include a horizontal line to infer the average rating of all films. I tried but then found the line could be deceptive due to the distribution of the data. The horizontal line may indicate irrelevance between runtime and ratings, which comes from an unavoidable limitation of my visualization. The data size is not large enough to keep my conclusion well-grounded. There are more films with length below average, where they may convey erroneous information like movies could be better with less runtime. Despite that, my visualization may conduct an obscured message that all genres will correspond to my findings, which they would require further analysis to testify.