Software Design Mini Project 3

Sander Miller

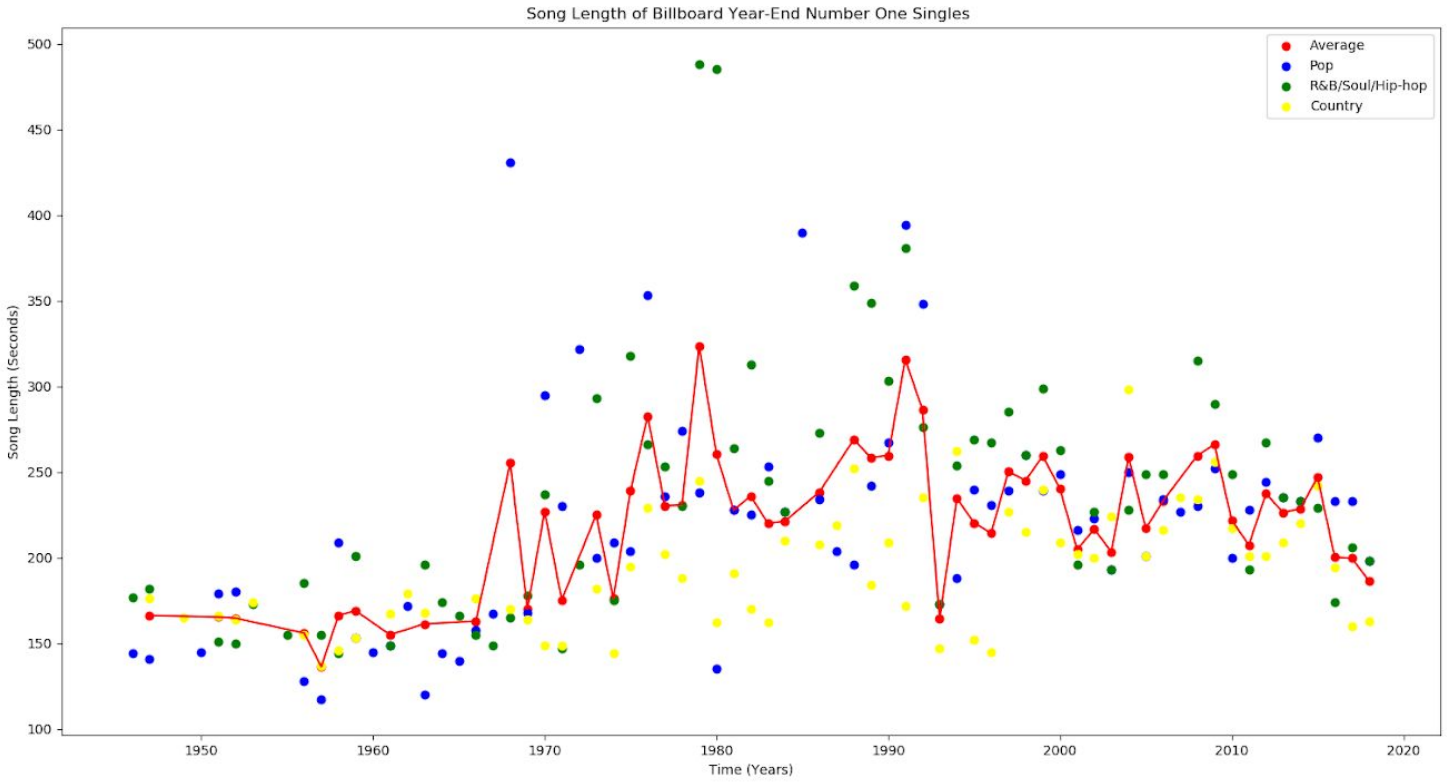
Project Overview:
For my project three assignment I examined how the length of popular songs has changed over time. My project involved scraping Wikipedia for links, and then following those links and scraping data from each of those sites. I then pickled, interpreted, and visually represented the data I collected. I hoped to be able to see a measurable change in song length over time.

Implementation:
First I pulled out a large list of links from a table. Next I filtered down the list of links to just the ones I wanted, the links to each song's Wikipedia page. I then scraped the song's Wikipedia page for the song length. Next I converted the song lengths to seconds, and found the average song length for each year. I stored the list of songs as well as the lengths in a dataframe indexed by year and genre, and pickled the data I collected.

When filtering out the links I wanted versus the ones I didn't it was very difficult to differentiate between a link to a song's Wikipedia and a link to an artist's Wikipedia. I considered following all the links and scraping for information such as birthdate to differentiate between the two, but decided to hard code the ones I wanted for simplicity and to minimize run time. If I had more time I would have liked to examine the html code a little more in-depth and see if there was an easier, more efficient way of doing things.

Results:
I was very pleased with my results. I was able to plot my data and see a noticeable increase in song length over time. Unfortunately there were a few parts that I had to hard code in workarounds. If the Wikipedia page were to be changed I would most likely need to make some tweaks to my code to account for it. However, currently it works perfectly.

Looking at my data you can see a clear increase in song length at about 1965. After some research I realize this increase is due to improvements in the technology. Until the late 50's music was primarily recorded on '78s' that could only store 3-4 minutes of music. After LP's were introduced song length increased because longer songs could be stored.

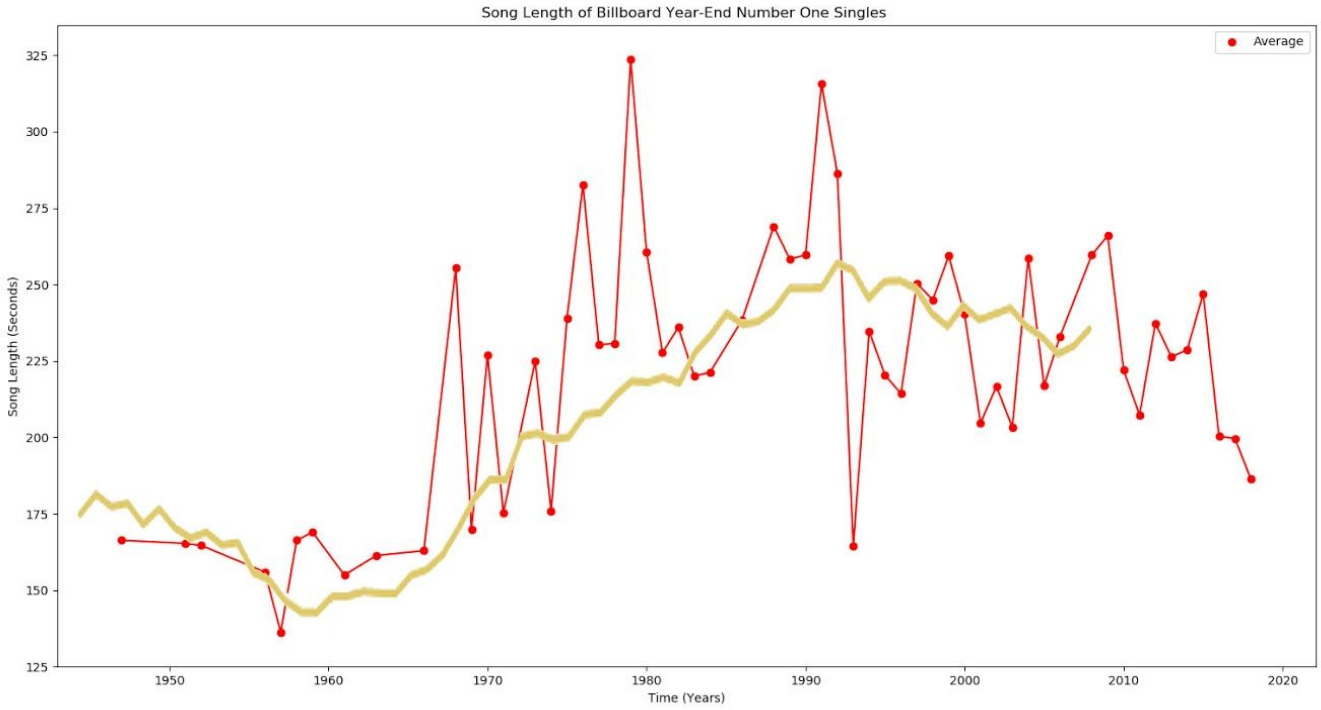


Alignment:

I ended up with very close to what I originally planned. I planned on using more songs per year to get more accurate data for average song length per year. I had expected to use a database of many different songs with there song lengths included, rather than scraping Wikipedia. I had heard that there was a magic length for songs to become popular, and my data was representative of that. Most songs were within a minute or two of each other with the occasional outliers.

I had also read about news clips becoming shorter and shorter as people's attention spans decreased I wondered whether the same would apply to music. There is a slight decrease in length from the 80's to present day, but not significantly. Unfortunately the Wikipedia was missing many song lengths, and only provided the #1 most popular song from each year for three genres.

I did some research after finishing my project about what research had been done on song lengths over time. I found the database Musicbrainz had done a study in which they too graphed the average song length over time. I overlaid the two graphs, lining up the years and length in seconds and to my delight the general trend of both graphs were fairly similar. My data is kind of grainy and I should have calculated the line of best fit, but here are the two graphs, mine in red and the Musicbrainz in yellow.



Reflection:

I think my project was well scoped. It was challenging, but feasible. Scraping Wikipedia was pretty straight forward. I had some prior experience using the pandas library to scrape tables from a web page which was very helpful. There were quite a few instances of hard coding to make my code functional which I wish to improve. Also I would like to have found a line of best fit to match my data. I could have used doc tests to do unit testing rather than testing my code manually. This project was my first look into html, and I am eager to explore it more.