CART451 Exercise 1 Report

by Arnaud Jalbert

The Dataset

I decided to choose the dataset of all the drivers that raced in Formula 1. It also contains all the statistics about their performance in the sports. I chose it because I really like Formula 1 and spend a lot of time looking at drivers' statistics and thought it was the perfect opportunity to dive more into it. The dataset was in a CSV file so I first had to convert it with a converter to JSON to easily read it in Node. To import it in Mongo, I first tried to copy the data directly into the import box but there were too many lines of data and the website was crashing. Instead, I read the JSON in Node and imported each entry from Node into Mongo, it was fast and efficient.

Queries

Championship Winning Drivers

I first made a query to get all the championship winning drivers. I was able to quickly look at the stats and was surprised by the amount of drivers that had won a championship. Seeing their other stats, it was also surprising to see that some drivers had very low wins and pole positions even if they were champions(Jody Scheckter for example).

Pole Sitting Drivers + Race Winners(2 queries)

See the definition for pole position here. I then made two queries to get the drivers that have had a pole position at least once and the drivers that won at least one race. I noticed that most pole sitters seem to have won multiple pole positions, unlike race wins that seem to have more one-time race winners. I also noticed that there are less pole sitters than there are race winners. Even if race wins are more prestigious, pole positions seem to be harder to achieve. It does make sense since there are a lot less variables in qualifying so the drivers with better cars and skills tend to grab pole positions more often since you simply have to be as fast as possible. In races, there are many more variables, including car failures, accidents, strategic decisions, and

many more. Hence, it looks like a driver with a less performant car has more chances to win a race than get a pole position

Pole Position but no Race Wins

In this query, I retrieved only the drivers with at least 1 pole position but no race wins. There aren't a lot of drivers who fit in this category and their pole positions amount is very low(5 at max). It looks like when a driver is consistent with its pole position, a race win eventually comes.

Race Wins

In this query, I retrieved only the drivers with at least 1 race win but no pole positions. Like the previous query, it does not happen really often. Most drivers in this category only have 1 win and it is likely due to other car failures, race incidents that gave them advantages on the strategy or other circumstances. Generally, drivers who win races tend to do well when qualifying and vice-versa.

Drivers active at least *n* years

I wanted to see and visualize drivers that were active for a certain amount of years and see how their stats looked with that. I made a query that took a number as a parameter and retrieved drivers who were active for that amount of years in Formula 1. The maximum years of activity is 19 years and there are 4 drivers that fit in this category, and only Rubens Barrichello is not a World Champion in this category.

Visualization

To visualize the data, I used the "Drivers active at least n years" query. I implemented a simple slider to dynamically change the number of years. Anytime the slider is moved, the canvas updates itself to show the amount of drivers that fit in the category. The canvas has one small round particle for each driver that fits the criterias. I also have a simple color code to identify certain characteristics:

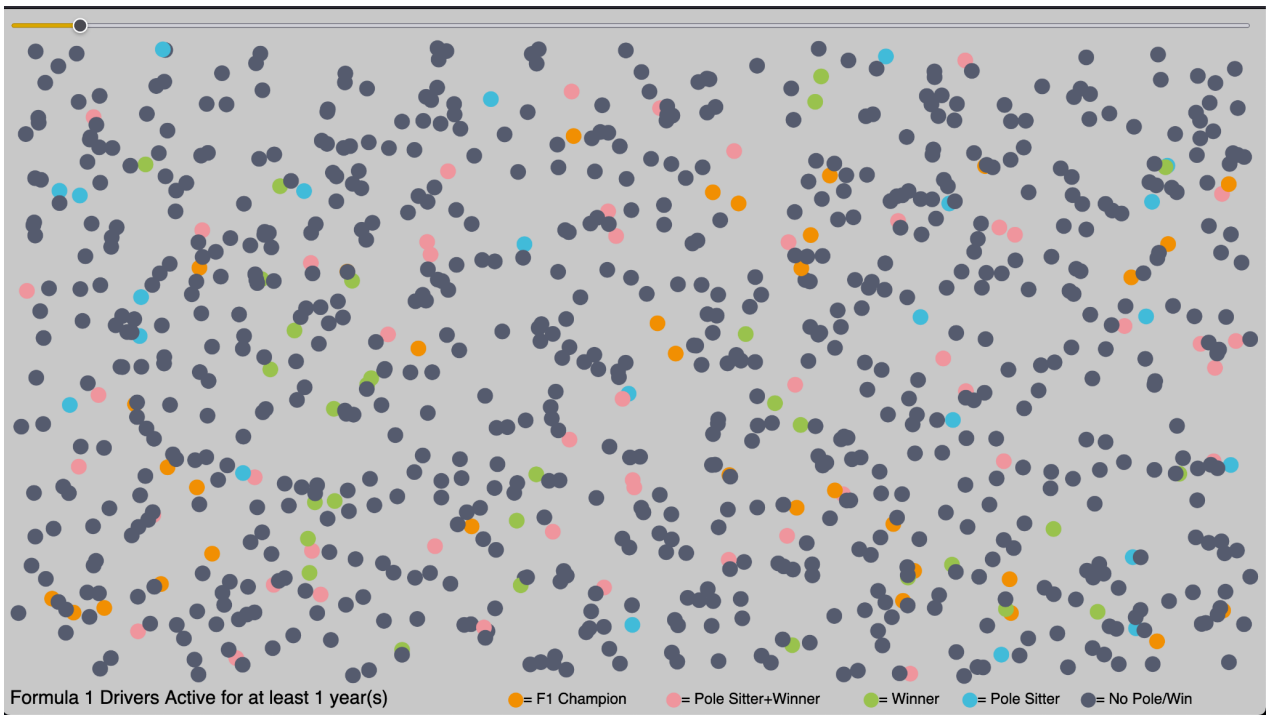
- Orange: World Champion(note that all world champions have at least one pole position and one win)
- Pink: Drivers that have at least one pole position and one win.

- Green: Drivers that have at least one win but no pole positions.
- Blue: Drivers that have at least one pole position but no wins
- Grey: All the other drivers.

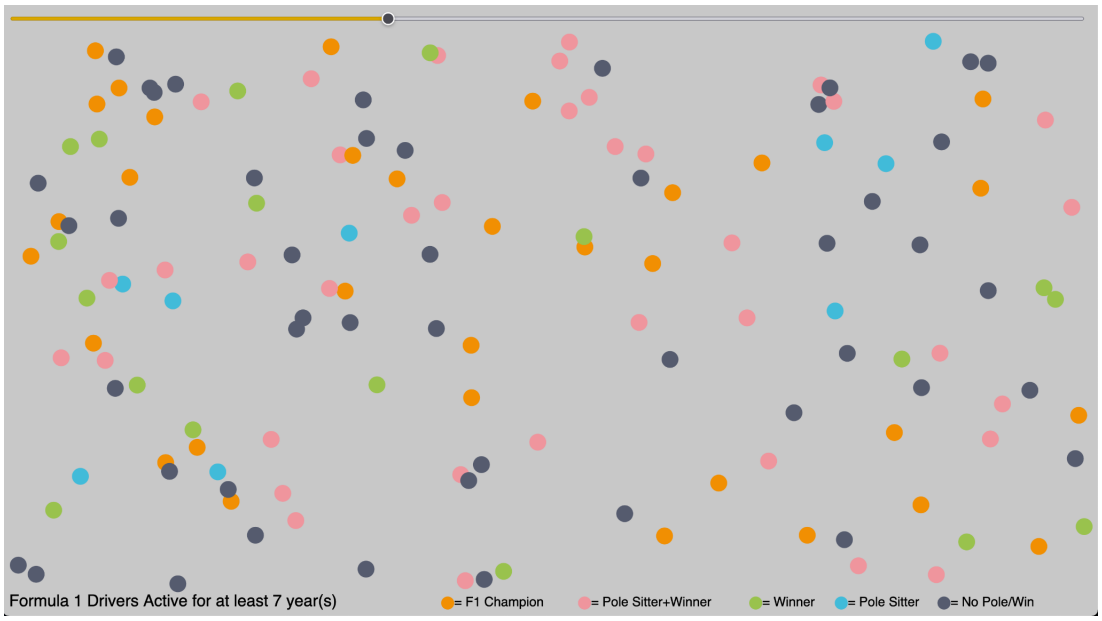
By playing around with the slider, we see that there are a lot of gray drivers when the number of years is low, but the more the number of years is high, the less gray drivers there are. Starting at 13 years, there are no more gray drivers. It makes sense, only drivers who perform well are able to stay in Formula 1 for that long.

Even in lower years(7 years for example) there are significantly more “colored” drivers than gray drivers. This goes to show how ruthless Formula 1 can be, if a driver does not perform, it is not long that they are replaced by someone else that could perform better. Knowing that only 20 persons can practice this sport every year, it makes sense that they want the best of the best.

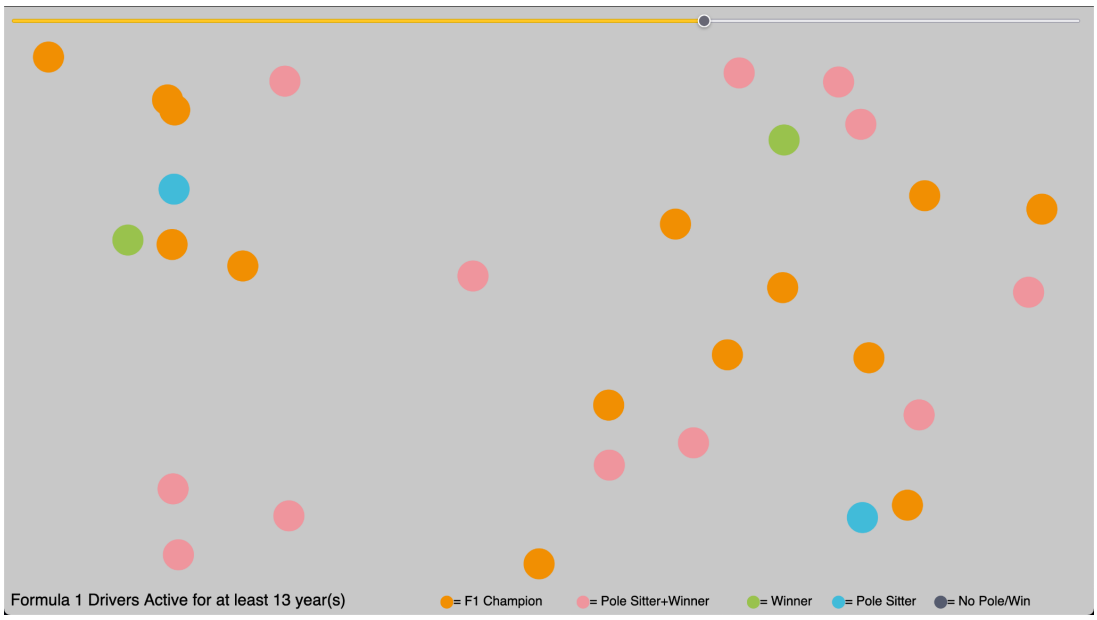
1 Year



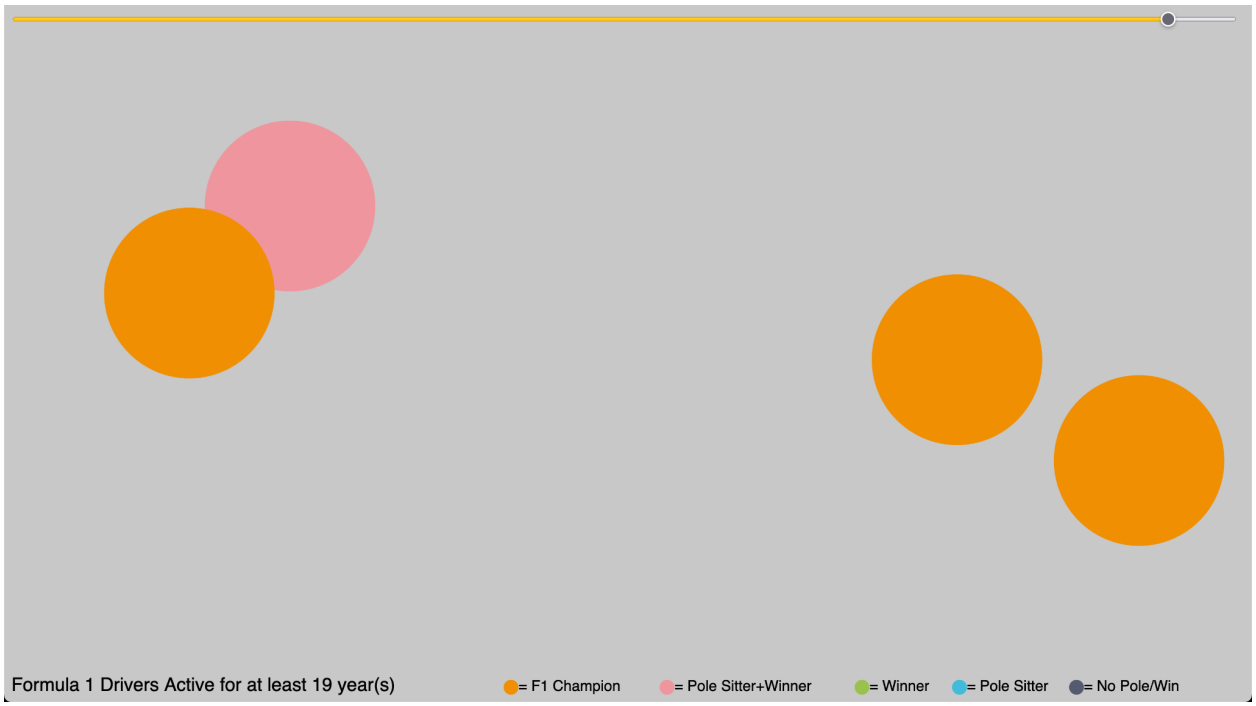
7 years



13 years



19 years


Formula 1 Drivers Active for at least 19 year(s)
= F1 Champion
= Pole Sitter+Winner
= Winner
= Pole Sitter
= No Pole/Win