

Formula 1 VRD

Process Book

Alex Liebig and Sharad Bhat

Table of Contents

1. Overview and Motivation	3
2. Related Work	3
3. Questions	5
4. Data	5
4.1 Additional Data	5
5. Exploratory Data Analysis	5
6. Design Evolution	6
6.1 Initial Sketches	6
6.2 Visualization Sketches	9
6.3 Experiments	12
7. Implementation	15
7.1 World Championship Points Line Chart	15
7.2 Points Heatmap	16
7.3 Track Locations World Map	17
7.4 Race Position Changes Line Chart	17
7.5 Lap Times Scatterplot	18
7.6 Race Position List	19
7.7 Additional Components	20
8. Evaluation	22
9. Links	23

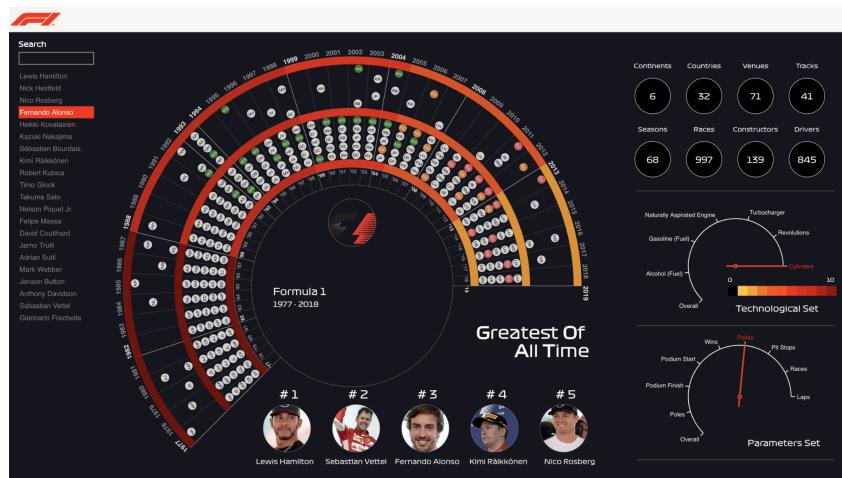
Overview and Motivation

A big part of the reason why we chose to visualize Formula 1 was due to our love for cars and passion for racing. Formula 1 has always been a sport of numbers, where milliseconds determine who's the champion. Visualizing these numbers in a graphical user interface will truly help with understanding the different trends in Formula 1. With this project we hope to spark interest in more Americans, and encourage them to start watching Formula 1.

Related Work

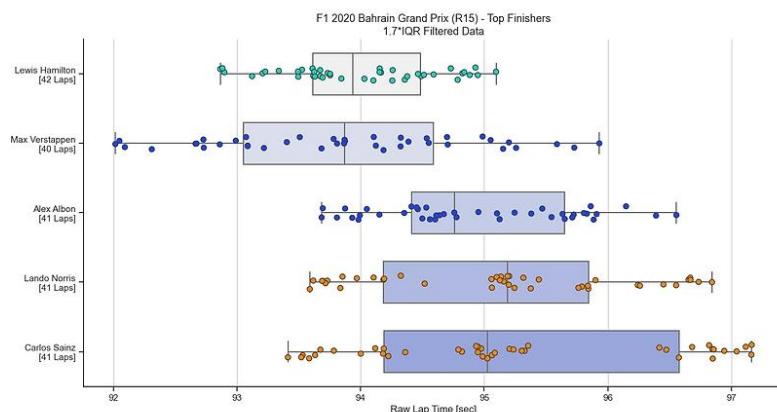
1. This one was by far the best visualization we found on the topic of Formula 1. The visualization shows the performance of drivers across multiple eras of Formula 1 and helps find which driver is the Greatest of All Time. The authors chose to stylize the visualization as gauges.

Link - <https://jasonjpaul.squarespace.com/formula-1-data-vis>



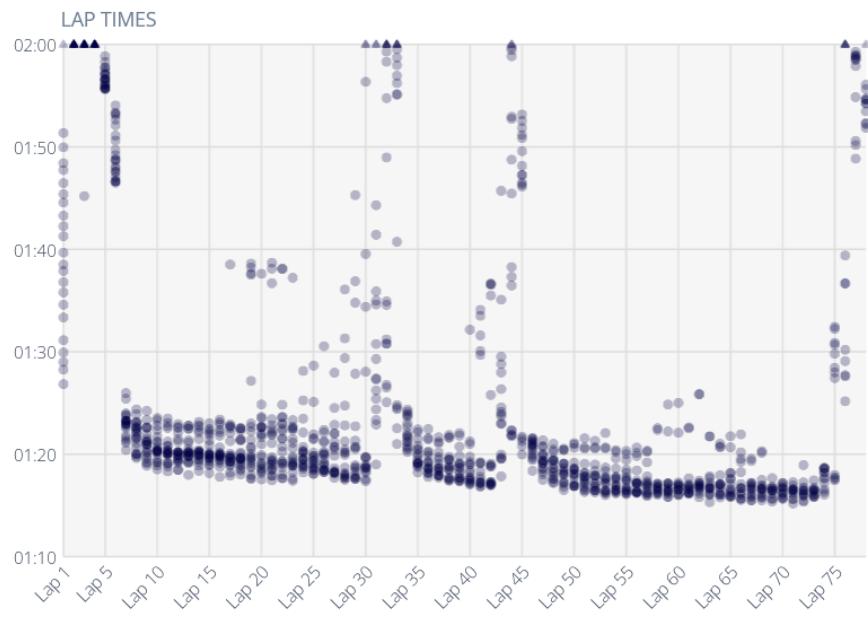
2. This blog contains multiple visualizations on the topic of Formula 1. In this particular visualization, the author chose to display each driver's lap time using a boxplot and a scatterplot. However, we thought that this can quickly get hard to read with 20+ drivers.

Link - <https://www.f1trends.com/blog>



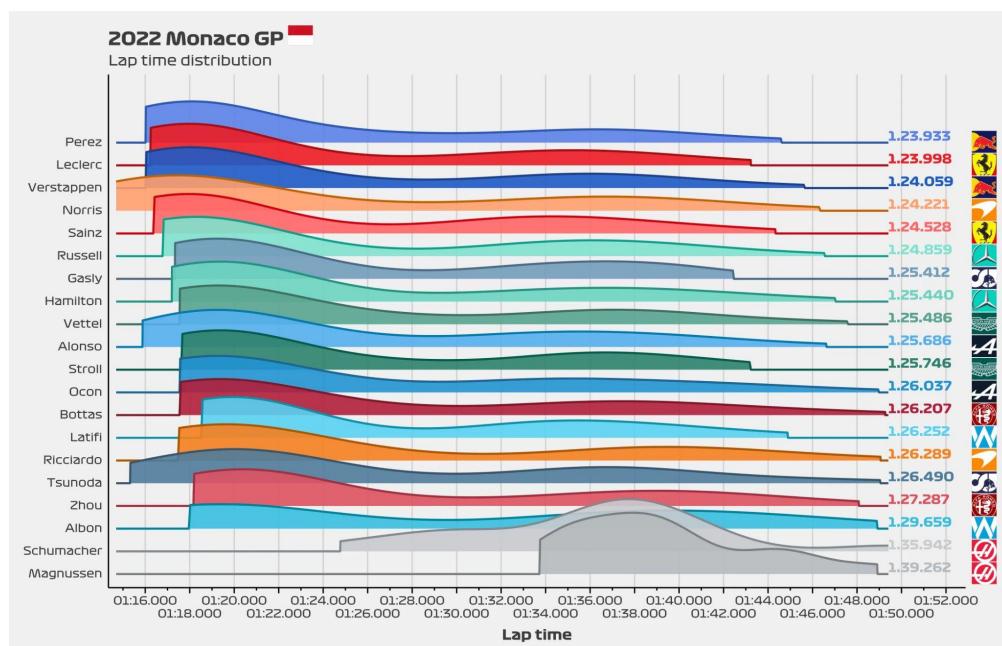
3. Similar to a trend line, this visualization highlights where the drivers lap times are clustering in this scatter plot. It does so by increasing the opacity towards the median lap time, and decreasing the opacity the further away it gets.

Link - <https://kwjames.com/f1times.html>



4. This plot below highlights how each racer is doing in a single race in terms of lap times, shown as a ridge line graph. Essentially the curves represent the density of lap times. The more dense a curve is the more times a driver finishes a lap around that specific time.

Link - <https://f1pace.com/p/2022-monaco-gp-lap-time-distribution/>



Questions

With this project, we aimed to answer the following questions.

- Which drivers and teams are scoring the majority of points in a season?
- How do drivers and teams perform as the season progresses?
- How do drivers perform in a given race?
- How do lap times compare in a given race between drivers?
- How do position changes take place in a race?
- Where do the overtakes happen in a race?

Data

The data is from an API, called the Ergast Developer API. However, since this API is community funded and can get rate limited, we opted to use a Kaggle Dataset of the same data, stored in CSV format.

API Link - [Ergast Developer API](#)

Kaggle Dataset Link - [Formula 1 World Championship \(1950 - 2022\) | Kaggle](#)

Additional Data

We wrote Python scripts to scrape descriptions and image links of track and drivers from their Wikipedia pages. The script used BeautifulSoup to extract data from the webpage. This data was stored in a CSV file which is then read and displayed using D3.

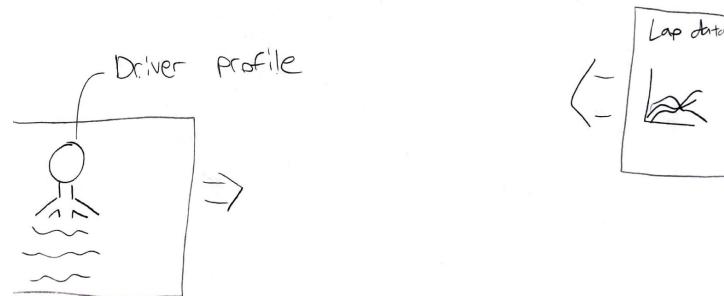
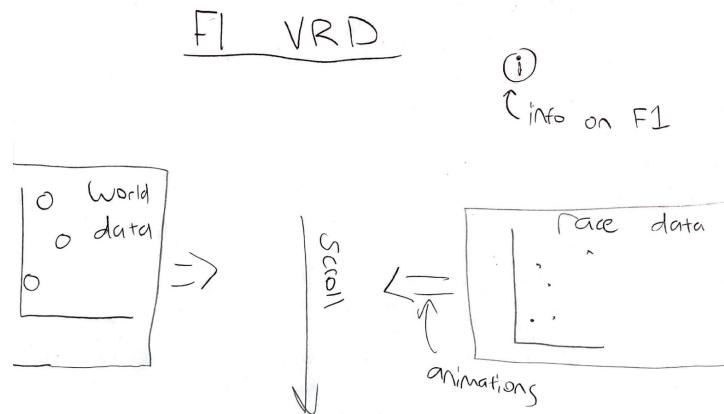
Exploratory Data Analysis

The dataset was available as CSV files. We combed through the files and looked for connecting primary and foreign keys across different CSV files. Once we had a good idea of the file structure and data distribution, we used our prior knowledge of Formula 1 to look for data that would be interesting to visualize and explore.

Design Evolution

Initially, we sketched out a few designs. Improving on our designs in each iteration, we believe we converged at the best possible layout and visualizations for displaying the data.

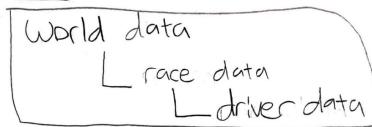
Sketch 1



Sketch 2

Scroll down & visualize more race data

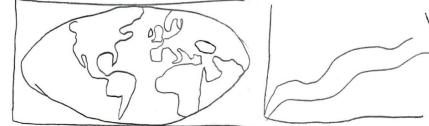
Structure:



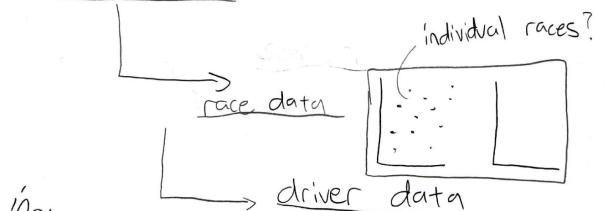
Possible structure?

World data

Season, stats



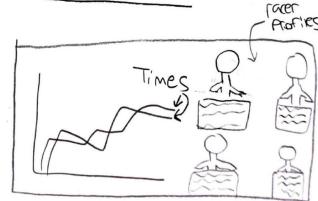
↓



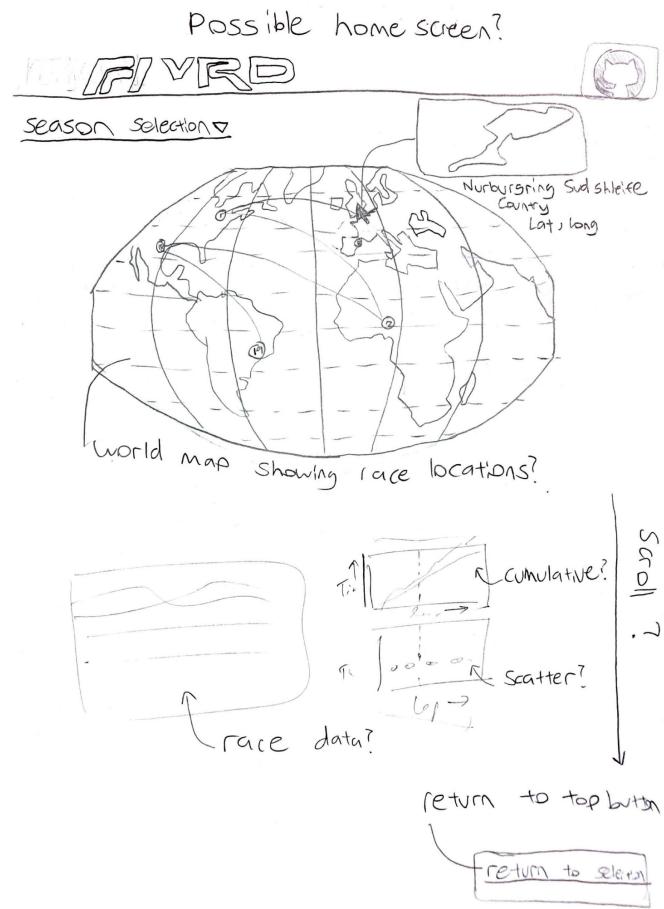
race data

driver data

include styling? F1 related

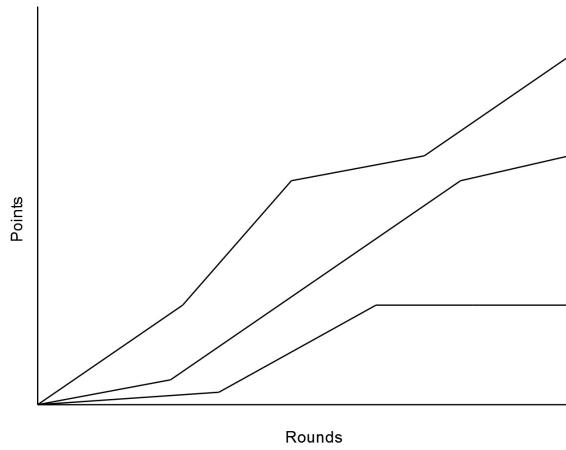


Sketch 3



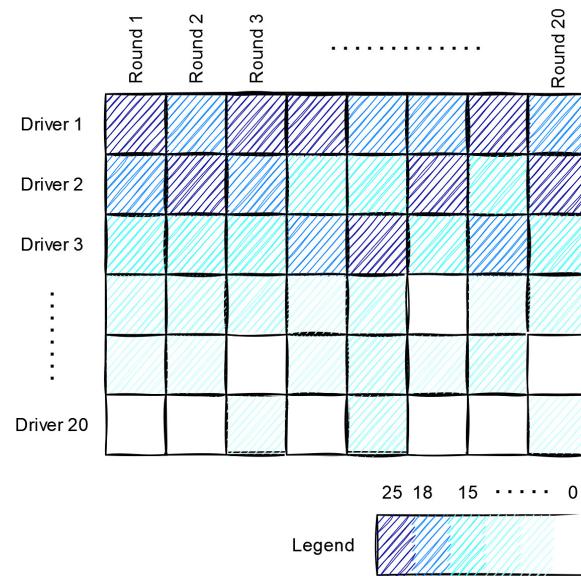
Once we had a general idea of our layout, we began sketching out our visualizations using [draw.io](#).

Visualization Sketch 1



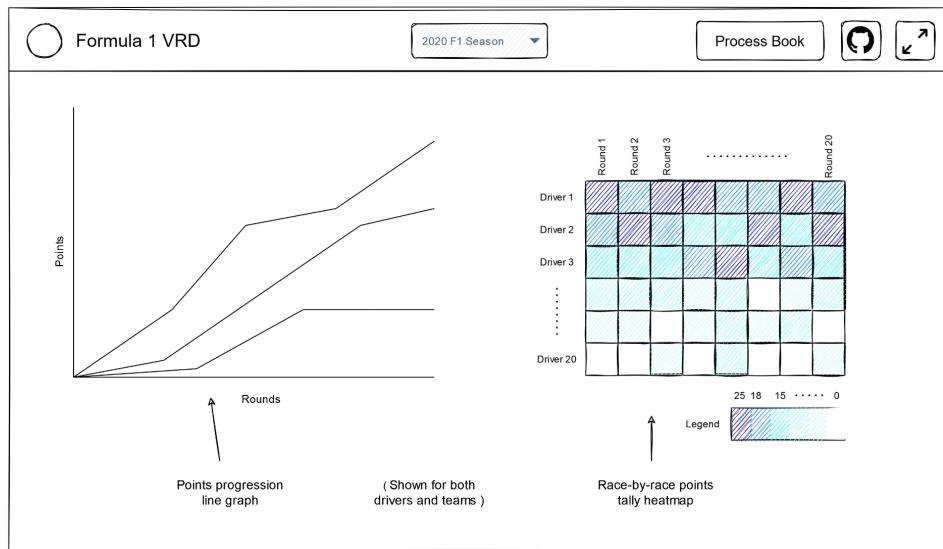
World Championship Points Line Chart

Visualization Sketch 2



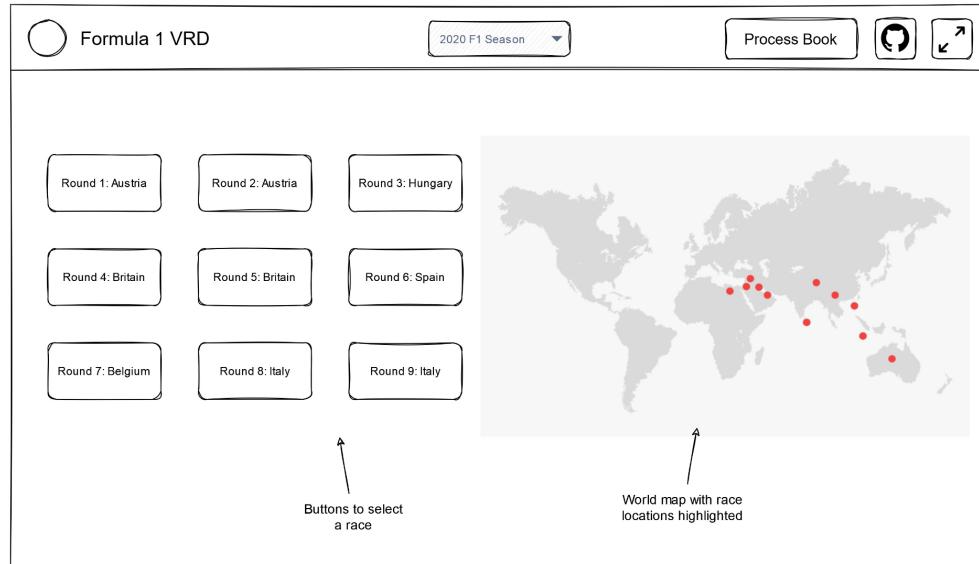
Points Heatmap

Visualization Sketch 3



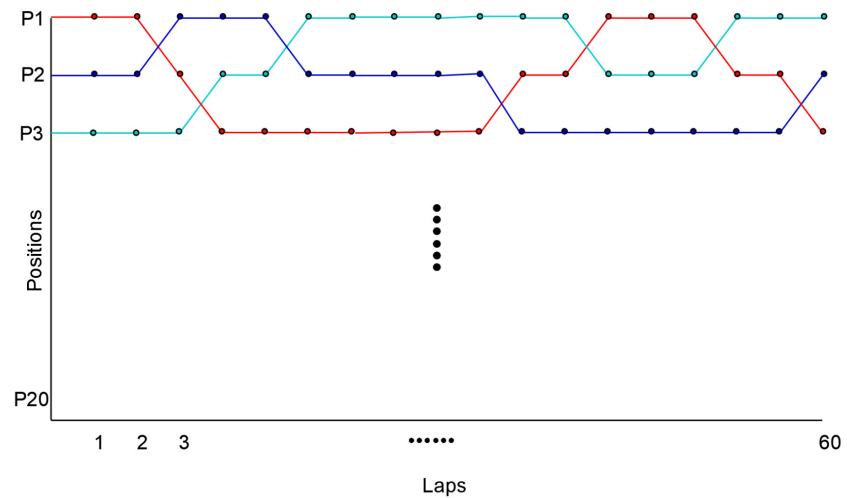
Layout 1

Visualization Sketch 4



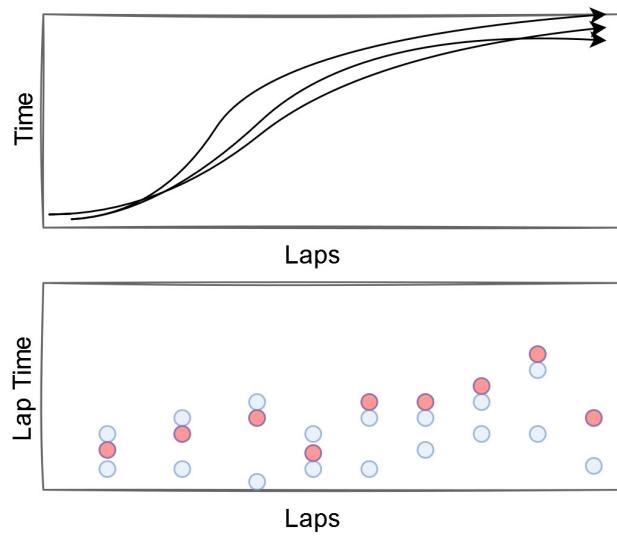
World Map with Track Selector

Visualization Sketch 5



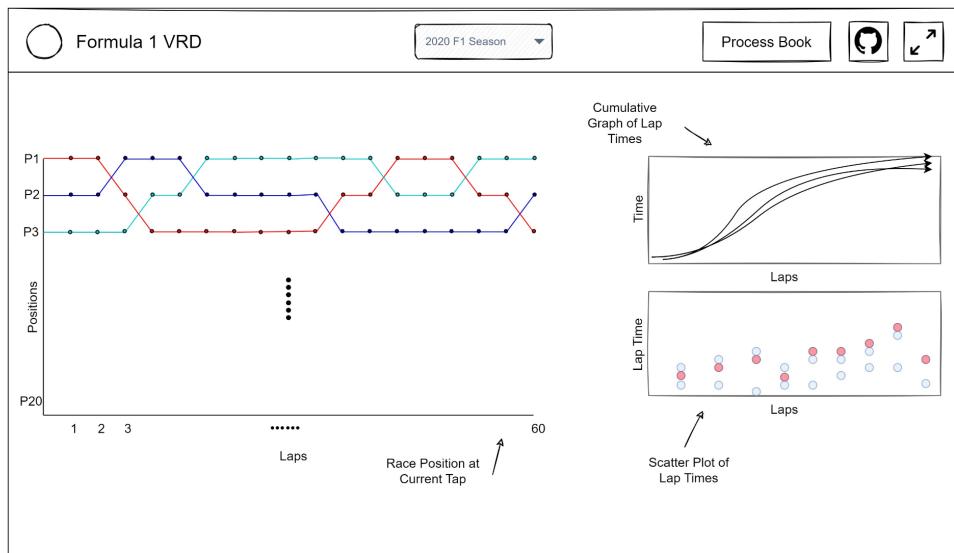
Race Position Line Chart

Visualization Sketch 6



Cumulative Lap Time Line Chart and Lap Time Scatterplot

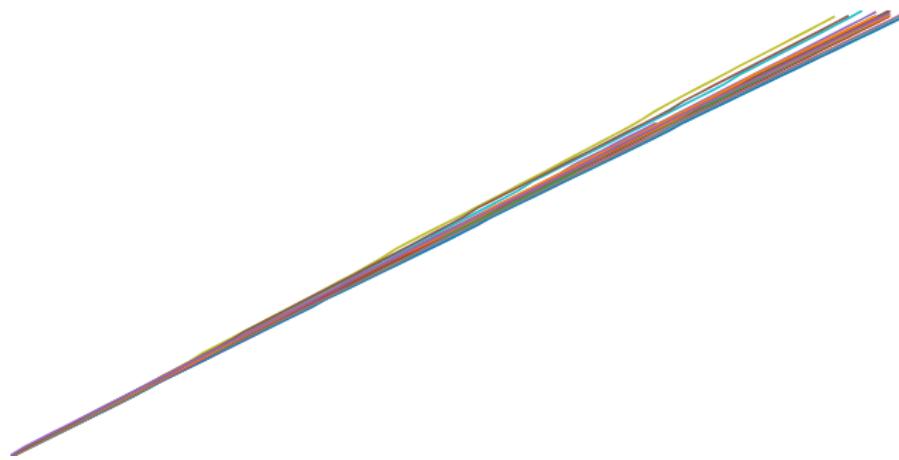
Visualization Sketch 7



Layout 2

Experiments

One thing we thought might have been a good idea at the time was to use a graph that displays the lap times cumulatively. However, after implementing it this is what it looked like.



As seen, the data is very difficult to read compared to the individual lap data (Shown below)



Overall, we believe this is the case since it not only highlights how competitive Formula 1 races are, but it also highlights how the spread between racers is fairly close. Also, the cumulative graph gives the false impression that the top line is the best time, which it clearly isn't since the lowest time is always the best. Since it is misleading to display the data this way, we decided to do away with using a cumulative graph and use only the lap-time graph above.

We also experimented with using a grid layout to display the country/name of each race. Although this was what we originally wanted to use, during the peer feedback, we were informed that this can be improved by using a list. Below is what that grid looked like:

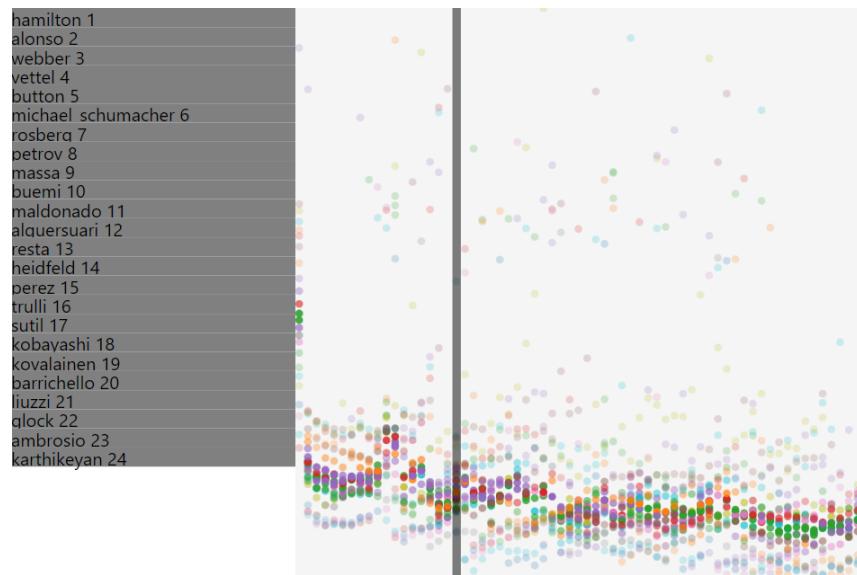
Melbourne	São Paulo	Imola	Silverstone	Montmeló
Nürburg	Monte-Carlo	Montreal	Magny Cours	Spielberg
Hockenheim	Budapest	Spa	Monza	Indianapolis
Suzuka	Kuala Lumpur			

For what it is, the grid takes up too much space and does not convey enough information. The list we ended up implementing fixed these issues.

Our initial implementation of the world map looked like the image below. However, we decided to improve the colors, add more information in the tooltip, and use a wider projection.

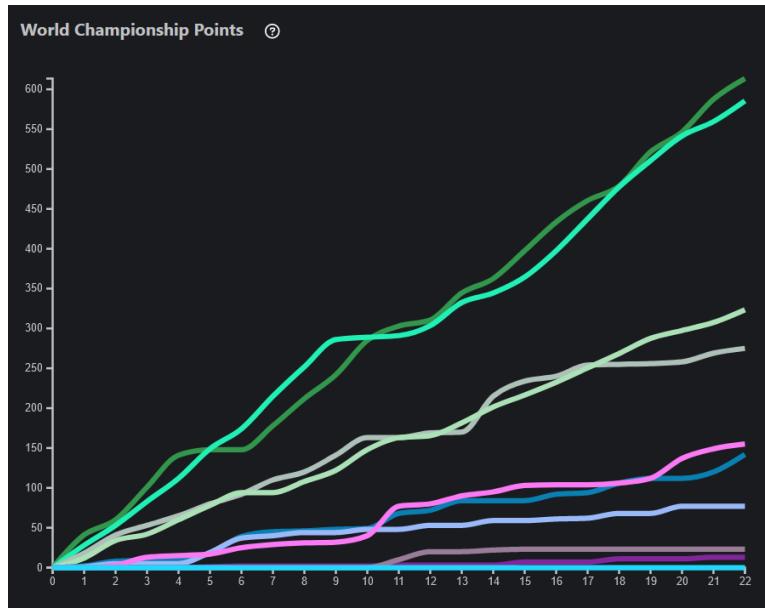


The image below is what the lap times scatterplot looked like initially. However, the colors can be hard to read. We decided to remove the colors and only show it when the user selects a driver.

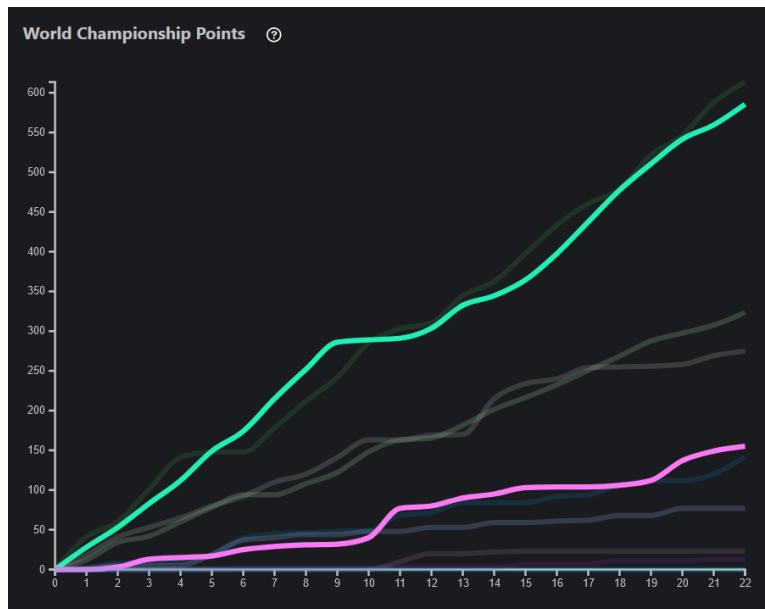


Implementation

World Championship Points Line Chart

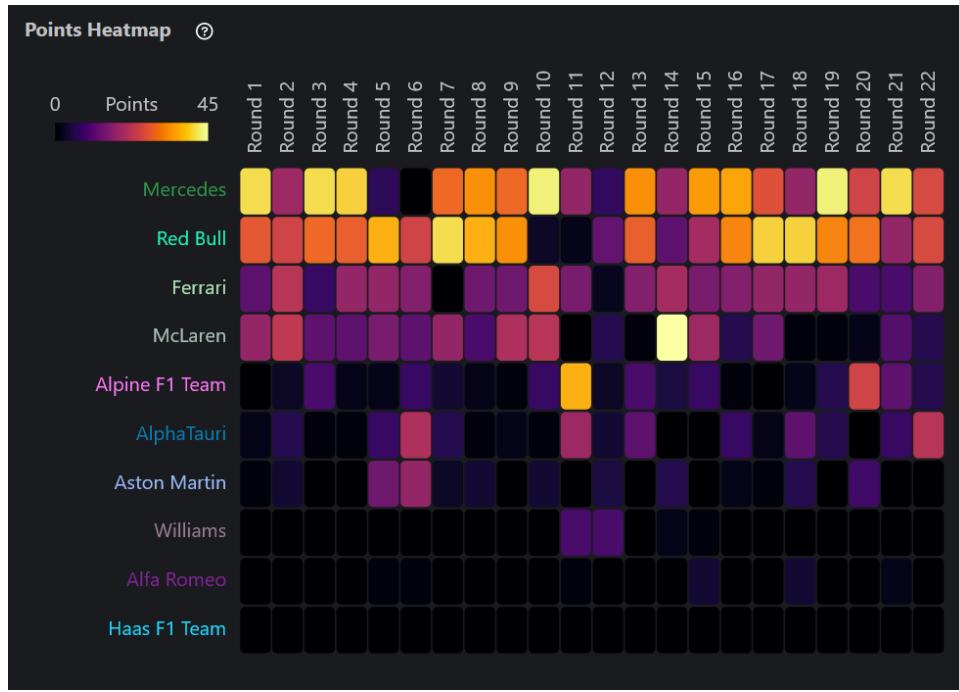


We show a line chart to show how teams and drivers progress as the season goes on. Each line corresponds to an individual driver or a team. The purpose of this was to show the points race that takes place throughout the season between drivers and teams to be crowned the world champions.

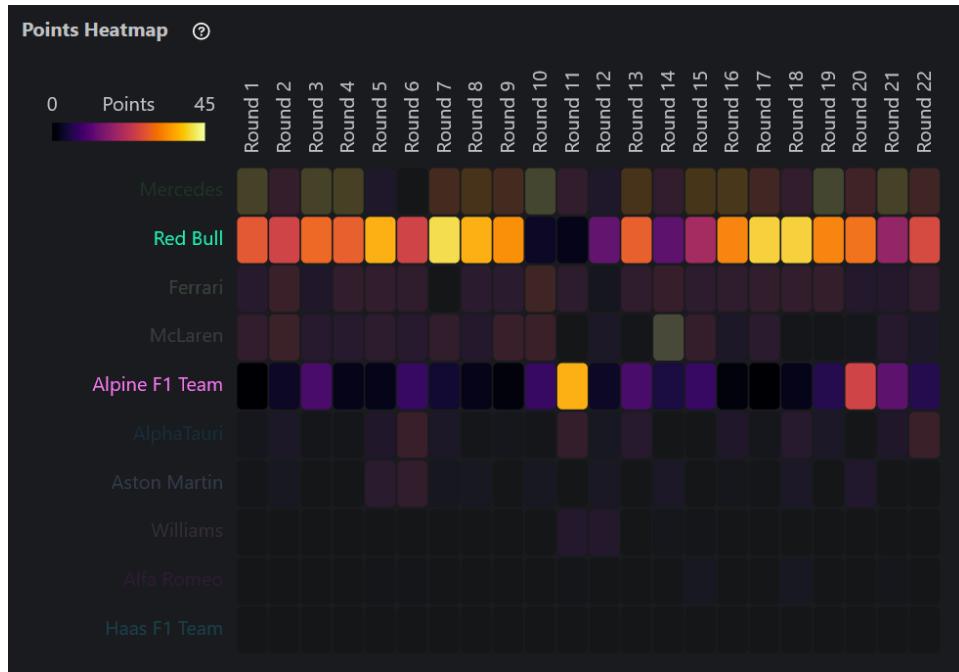


When the user selects drivers or teams, we bring attention to these lines by lowering the opacity of all other lines. The user can also hover over a line to see details in a tooltip such as the driver name, the current round and the points accumulated up until the round.

Points Heatmap

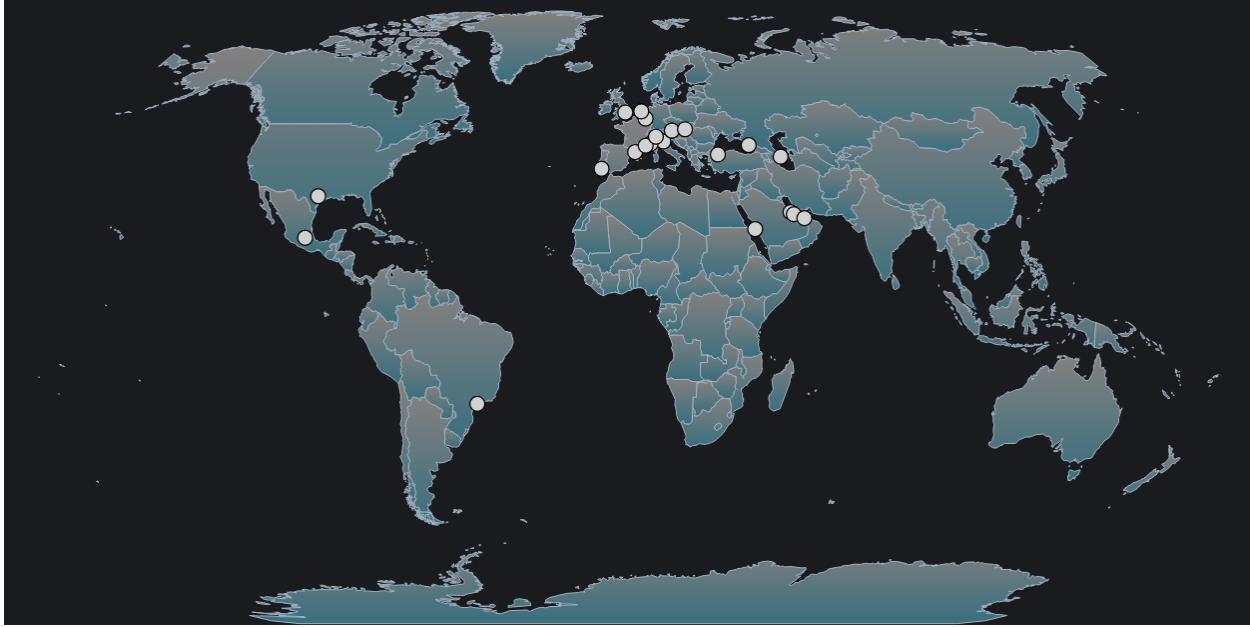


A heatmap to show which drivers and teams are scoring the majority of the points at each race in a season. The user can hover over a rectangle to see details on a tooltip such as points scored at that race, the round number and the name of the race.



When the user selects different teams or drivers, we bring attention to those rows by lowering the opacity of the other rows.

Track Locations World Map



In the world map view, we display the different races in a season and where it takes place. The user can hover over a particular circle to view the race track's name and which round it was. The purpose of this was to make it easier for users to find the name and location of particular races. To reduce the amount of data being loaded in the world map, we used [mapshaper](#) to lower the space used from 3,543 KB to 152 KB.

Race Position Changes Line Chart

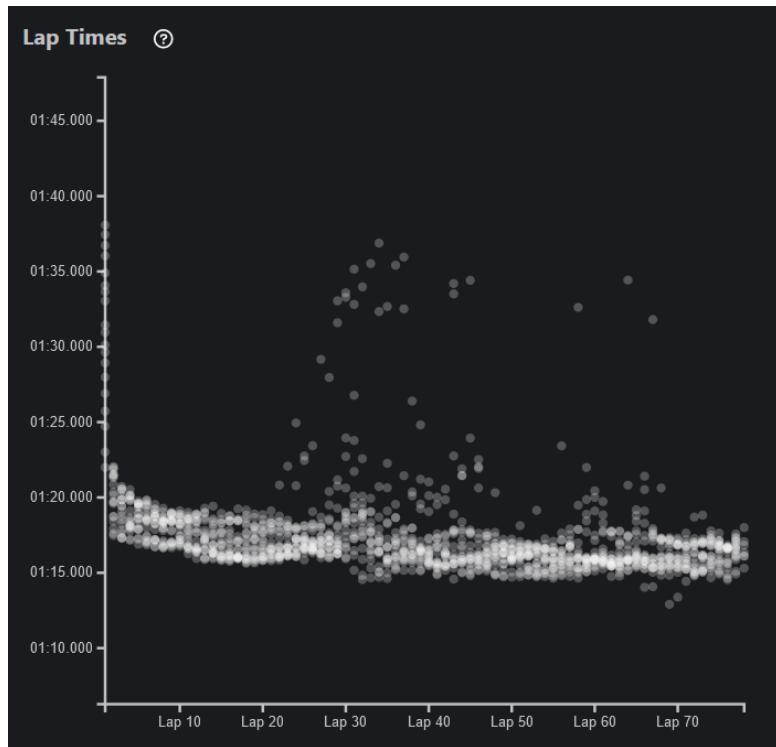


The multi-line chart to visualize position changes between drivers in a race. Each line corresponds to a single driver. The line shows the position changes of a particular driver and which lap each change took place in.

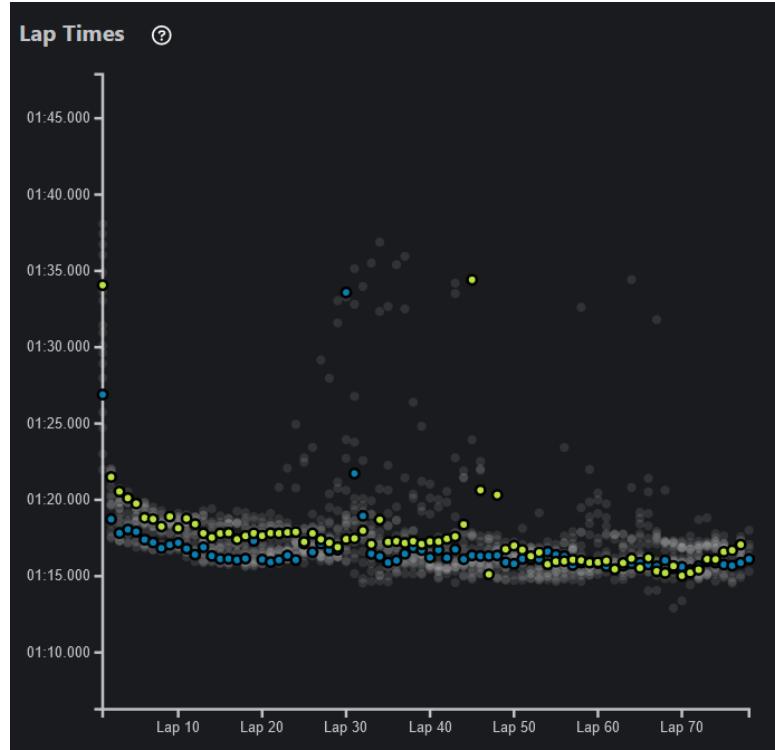


When the user selects a driver, we lower the opacity of the other lines to draw attention to the selected drivers.

Lap Times Scatter Plot



A scatter plot of the lap times of all drivers throughout the course of a given race. Each circle corresponds to the time taken by a driver to complete a lap. The purpose of this is to show how close the drivers are, with respect to the lap times.



Similar to the previous visualization, when the user selects a driver, we lower the opacity of the other circles to draw attention to the lap times of the selected drivers.

The user can select multiple drivers to compare their lap times.

Race Position List

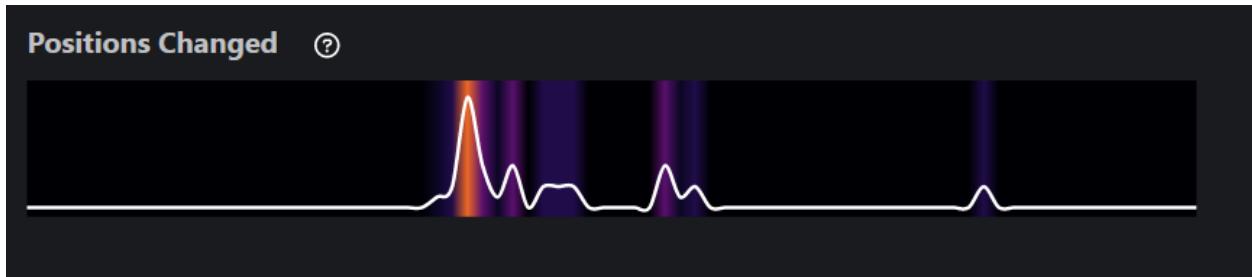
Final Results	
P1	Max Verstappen
P2	Carlos Sainz
P3	Lando Norris
P4	Sergio Pérez
P5	Sebastian Vettel
P6	Pierre Gasly
P7	Lewis Hamilton
P8	Lance Stroll
P9	Esteban Ocon
P10	Antonio Giovinazzi
P11	Kimi Räikkönen
P12	Daniel Ricciardo
P13	Fernando Alonso
P14	George Russell
P15	Nicholas Latifi
P16	Yuki Tsunoda
P17	Nikita Mazepin
P18	Mick Schumacher
DNF	Valtteri Bottas
DNS	Charles Leclerc

Lap 17	
P1	Max Verstappen
P2	Valtteri Bottas
P3	Carlos Sainz
P4	Lando Norris
P5	Pierre Gasly
P6	Lewis Hamilton
P7	Sebastian Vettel
P8	Sergio Pérez
P9	Antonio Giovinazzi
P10	Esteban Ocon
P11	Lance Stroll
P12	Kimi Räikkönen
P13	Daniel Ricciardo
P14	Fernando Alonso
P15	George Russell
P16	Nicholas Latifi
P17	Yuki Tsunoda
P18	Mick Schumacher
P19	Nikita Mazepin

This is a list of the positions of each driver at the end of the race. It shows what position a driver finished at, and whether a driver did not complete a race due to some factor.

When the user hovers over a lap on the race position line chart or the scatterplot, this list updates in real time to show the position changes. The user can also select drivers to highlight their positions.

Positions Changed Kernel Density Map

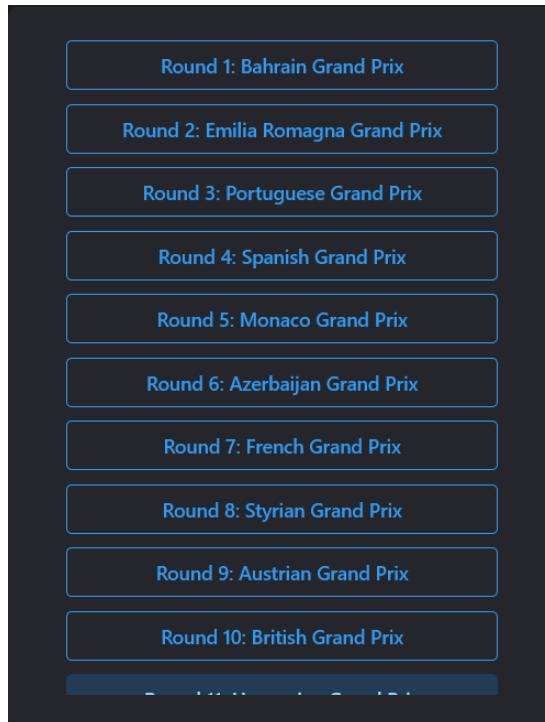


A kernel density chart to show where the majority of the position changes happened during a race.

We also use a line chart on top of it to emphasize the fact. The user can hover on the chart to view the exact number of position changes that happened in a particular lap.

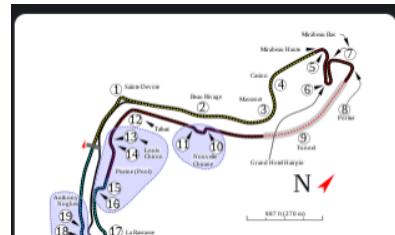
Additional Components

1. Track Selector



This list allows the user to look through all races and select whichever they want to explore further.

2. Track Details Card



Circuit de Monaco

Location	Monte-Carlo
Country	Monaco

[Hide Description](#)

Circuit de Monaco is a 3.337 km (2.074 mi) street circuit laid out on the city streets of Monte Carlo and La Condamine around the harbour of the Principality of

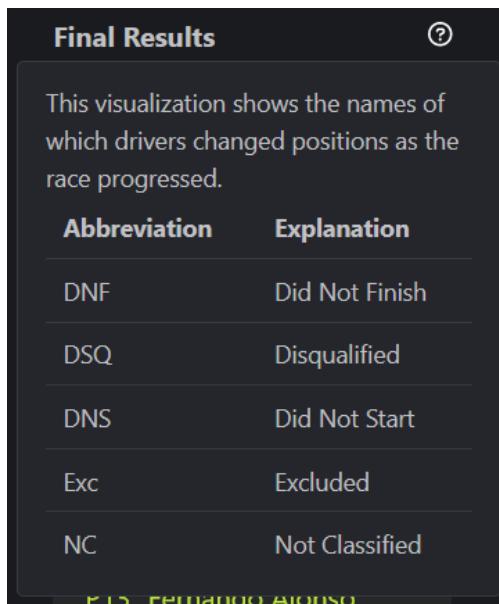
This card shows some details about the selected race track. It shows an image of the track outline and some details like the location, country and track name. It also contains a link to the track's Wikipedia article. The scraped data for tracks is used here.

3. Driver Details Card

Driver Profile(s)	
	
Sebastian Vettel	Lewis Hamilton
Nationality	German
Driver Number	5
Driver Code	VET
DOB	July 1987
Nationality	British
Driver Number	44
Driver Code	HAM
DOB	January 1985

This component shows a card for each selected driver. Each card shows an image of the driver, the nationality, driver number, driver code and the date of birth. Each card also contains a link to the driver's Wikipedia article. The scraped data for drivers is used here.

4. Visualization Description Hover Cards



Every chart is accompanied by a “question” icon which, when hovered over, provides the user with relevant information to better understand the chart.

Evaluation

We believe what's true for the race visualizations is how close each driver's times are to one another and how competitive each race is. Unlike sports such as College Football where it's usually two teams competing at a time in a game, the outcome of Formula 1 races are much more variable as more than ten teams and twenty drivers are competing in the same race at the same time. With the exception of top drivers such as Lewis Hamilton and Michael Schumacher, Formula 1 races are hard to predict.

We focused primarily on having great interactions. Every chart is connected to another chart in some way. The user is provided with multiple options for selecting drivers and teams. Some examples of these are, hovering over a round in the World Championship line chart updates the heatmap, the world map and the track list. And, selecting a driver in the line chart updates the Lap Times scatterplot and the list. We thought these improvements greatly affect how a user interacts with a visualization.

We ran into many hurdles while developing this project. One of the recurring problems was getting ReactJS and D3JS to work together. Scouring through many forums and blog posts, we think we arrived at a stable solution. On the design front, one aspect we constantly went back and forth on was the use of colors to represent drivers. With there being more than 20 drivers in some races, we decided on using a categorical colormap with as many colors as drivers. Since each driver's points were connected in line charts, we thought that this could be easy to follow and understand. For this reason, we allow the user to select drivers and highlight only their points and lower the opacity of other points to reduce distractions.

Links

Project Website - <https://sharadbhat.github.io/Formula-1-VRD/>

Github Repository - <https://github.com/sharadbhat/Formula-1-VRD>

Demonstration Video - <https://youtu.be/HZR8Wibx924>