

CS 6630: Process Book - Formula 1 Stats

Team Members: Harshini Keerthi Vasan, Vinu Sreenivasan, Sanjana Aravindan

Basic Info

Project Title

Formula 1 Stats

Team Members

1. Sanjana Aravindan, u1135274, u1135274@utah.edu
2. Harshini Keerthi Vasan, u1136693, u1136693@utah.edu
3. Vinu Sreenivasan, u1138972, vinu.sreenivasan@utah.edu

Project Repository

<https://github.com/HarshiniVasu/Formula-1-Stats>

Overview

Formula 1 Stats is a data visualization project that would allow the user to visualize the most successful teams (car brands) over a period of time and top circuits in the world. It would also show a comparison between various drivers who were active over a period of time. This project would include three main visualizations:

1. Showing the most successful teams over a period of time and the respective drivers who represented those teams in those years.
2. A comparative visualization between driver statistics who were active over a period of years.
3. Showing a map visualization of all the regions in the world that have formula 1 circuits (race tracks). This part would also show the top circuits in the world on map.

Background and Motivation

We, as a team are ardent fans of Formula 1 racing. This racing is a perfect blend of technology and the automobile industry. The mere adrenaline rush that the viewers get watching this sport, is something that gets us excited about working on this project. We are inspired enough to visualize various interesting correlations between all the drivers of the league and eminent teams like Ferrari, Mercedes and McLaren.

Though there are many visualizations on this data online, we couldn't find a place which shows the most successful teams over a selected period of years. Also, most of the visualizations did not have a comparative analysis of drivers over a period of time (for example from 2010-2015). This idea motivated us to take up the project.

Related Work

Formula 1 racing is a very well renowned sport. While we see a lot of visualizations on FIFA world cup and Olympics, we don't get to see that much for Formula 1 racing. That was one of the reasons that we chose to work on an unconventional sport data and build visualizations on it.

Project Objectives

We aim to show the highly sought after mainstream Formula 1 teams that are/were in the league over a period of time that the user wishes to visualize. The main questions that we are trying to answer are:

1. What are the top most circuits in the world? This would be on the basis of the number of races that took place on those circuits between 1950 and 2018.
2. What are the teams that scored well on those circuits respectively?
3. What are the most successful teams in a sorted order over a specific period of time?
4. Who are the drivers who drove for those successful teams in that specific period of time?
5. What are the basic details and stats of the driver such as points earned by the driver on a specific period of time?
6. Comparison of two or more driver's performance in terms of points and average lap time over a specific period of team.

We would like to use the knowledge we have in d3 to visualize the formula 1 data set. We will use concepts covered in class, assignments and also a few other d3 based visualizations to build an effective user interface with interesting visualizations of formula 1's records.

Data

We did a significant amount of research and found a set of multiple csv files related to formula 1 racing from Ergast Developer API. These csv files provide mapping of various sets of data using primary keys. There is a circuit (race track) data file which comprises of a unique key id, the track location of the races with its latitude and longitude. The constructors file (f1 teams) has fields like id, nationality of the team and their standings game-wise and year-wise (1950-2018). The driver's data file includes the history of their games (wins), the team they raced for, their standings in global level year wise as well as match wise. There is another interesting csv file which provides the status of the drivers with respect to every game denoting the number of laps, damages, collisions, etc.

Data Processing

We have done a bit of data cleansing and re-organized the data according to our visualization specification. We have multiple csv files like circuits (race tracks), constructors (teams), drivers, results, status, races. Using the primary keys (ids) present in every csv file, we combined the data that were only necessary from all of these files like driver statistics (nationality, DOB, points, position, laps), team statistics (nationality, results, laps), status (the issues or status tracking during the racing - oil leak, collision, engine, accident, axle, count of laps successfully). We also split the raw data year-wise so that retrieval of the data year-wise for team statistics and driver statistics visualization becomes easier.

```

In [5]: races_results_circuit_drivers = pd.merge(races_results_circuit, drivers_data[['driverId', 'driverRef', 'number', 'forename', 'surname', 'dob', 'nationality']], on='driverId')

In [6]: races_results_circuit_drivers.rename(columns={'year': 'season',
                                                    'name_x': 'raceName',
                                                    'circuitRef': 'circuitId_x', 'name_y': 'circuitName'},
                                           inplace=True)

In [7]: races_results_circuit_drivers.rename(columns={'dob': 'dateOfBirth', 'forename': 'givenName', 'surname': 'familyName', 'grid': 'results.grid', 'laps': 'results.laps'},
                                           inplace=True)

In [8]: races_results_circuit_drivers_cons = pd.merge(races_results_circuit_drivers, constructors_data[['constructorId', 'constructorRef', 'name', 'nationality']], on="constructorId")

In [9]: races_results_circuit_drivers_cons.rename(columns={'nationality_x': 'nationality', 'nationality_y': 'nationality.1', 'name': 'name_x'},
                                           inplace=True)

In [10]: races_results_circuit_drivers_cons_status = pd.merge(races_results_circuit_drivers_cons, status_data, on='statusId')

In [11]: races_results_circuit_drivers_cons_status.rename(columns={'status': 'results.status'}, inplace=True)

In [12]: races_results_circuit_drivers_cons_status.columns
Out[12]: Index([u'raceId', u'season', u'round', u'circuitId', u'raceName', u'date',
                u'driverId', u'constructorId', u'number_x', u'results.grid',
                u'position', u'positionText', u'results.laps', u'points', u'statusId',
                u'circuitId_x', u'circuitName', u'location', u'country', u'lat', u'lng',
                u'driverRef', u'number_y', u'givenName', u'familyName', u'dateOfBirth',
                u'nationality', u'constructorRef', u'name_x', u'nationality.1',
                u'results.status'],
                dtype='object')

In [13]: races_results_circuit_drivers_cons_status = races_results_circuit_drivers_cons_status.drop(['number_y', 'circuitId', 'constructorId', 'driverId'], axis=1)

In [14]: races_results_circuit_drivers_cons_status.rename(columns={'number_x': 'number'}, inplace=True)

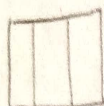
In [15]: races_results_circuit_drivers_cons_status = races_results_circuit_drivers_cons_status[['season', 'round', 'raceName', 'date', 'circuitId_x', 'circuitName', 'number', 'position', 'positionText', 'points', 'driverRef', 'givenName', 'familyName', 'dateOfBirth', 'nationality', 'constructorRef', 'name_x', 'nationality.1', 'results.grid', 'results.laps', 'results.status', 'location', 'country', 'lat', 'lng']]
races_results_circuit_drivers_cons_status.columns
Out[15]: Index([u'season', u'round', u'raceName', u'date', u'circuitId_x',
                u'circuitName', u'number', u'position', u'positionText', u'points',
                u'driverRef', u'givenName', u'familyName', u'dateOfBirth',
                u'nationality', u'constructorRef', u'name_x', u'nationality.1',
                u'results.grid', u'results.laps', u'results.status', u'location',
                u'country', u'lat', u'lng'],
                dtype='object')

```

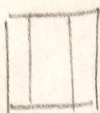
Exploratory Data Analysis and Design Evolution

The images below are the initial set of visualizations that we considered doing as a part of the project during our project proposal phase.

Driver Champions:



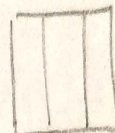
1990



1991



1992



1993



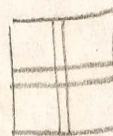
1994



1995



1996



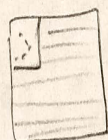
1997



1998



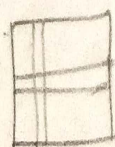
1999



2000



2001



2002



2003

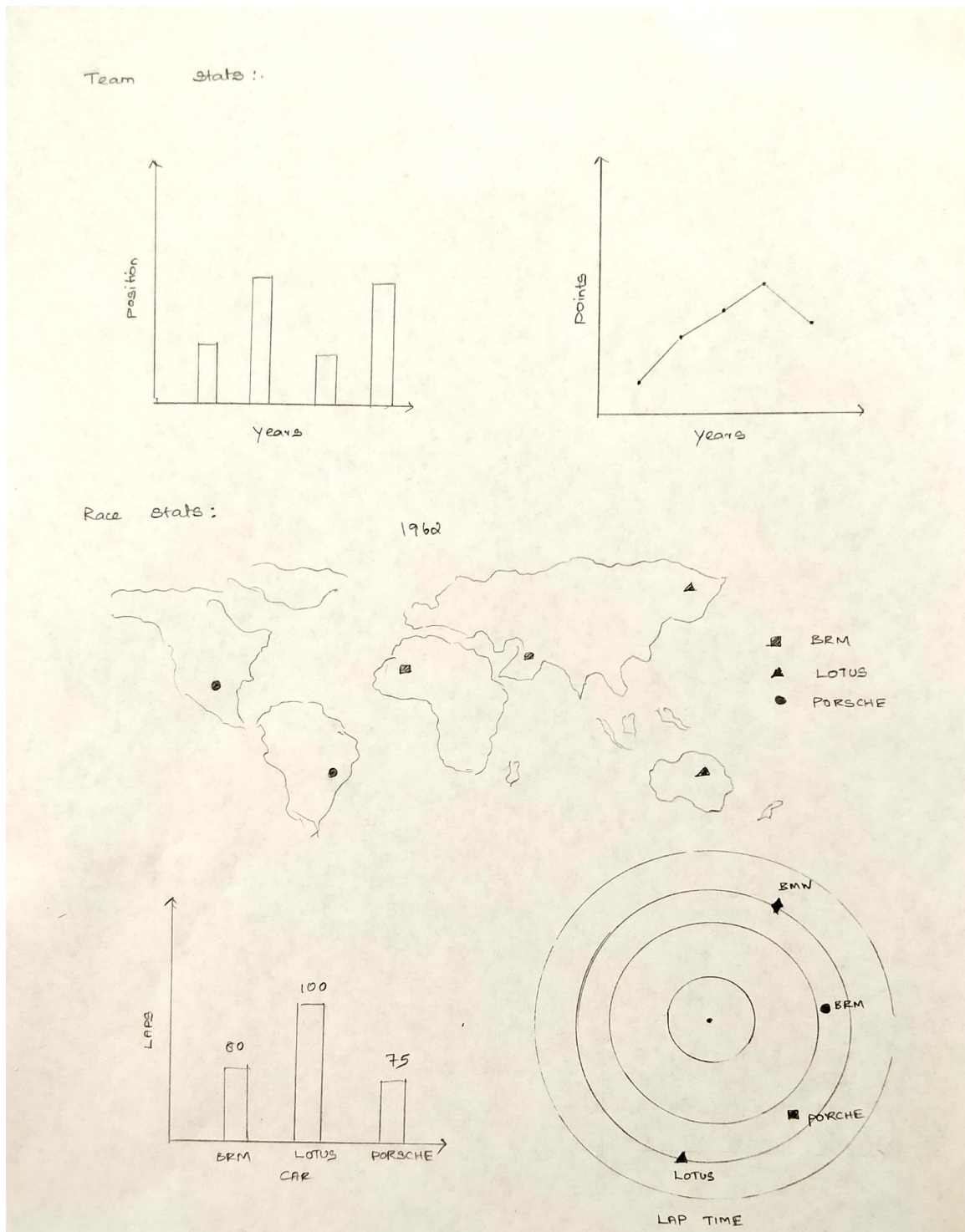


2004

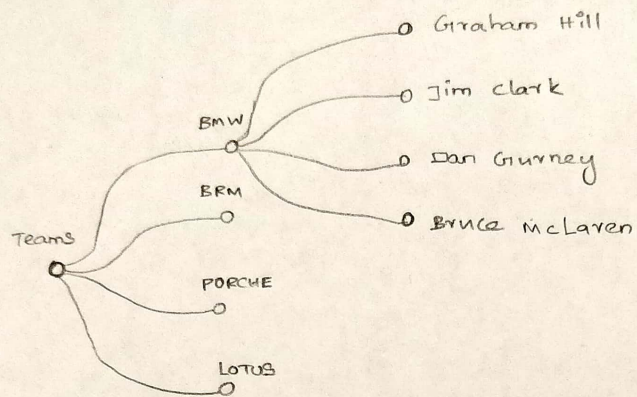


2005

Team Statistics Images

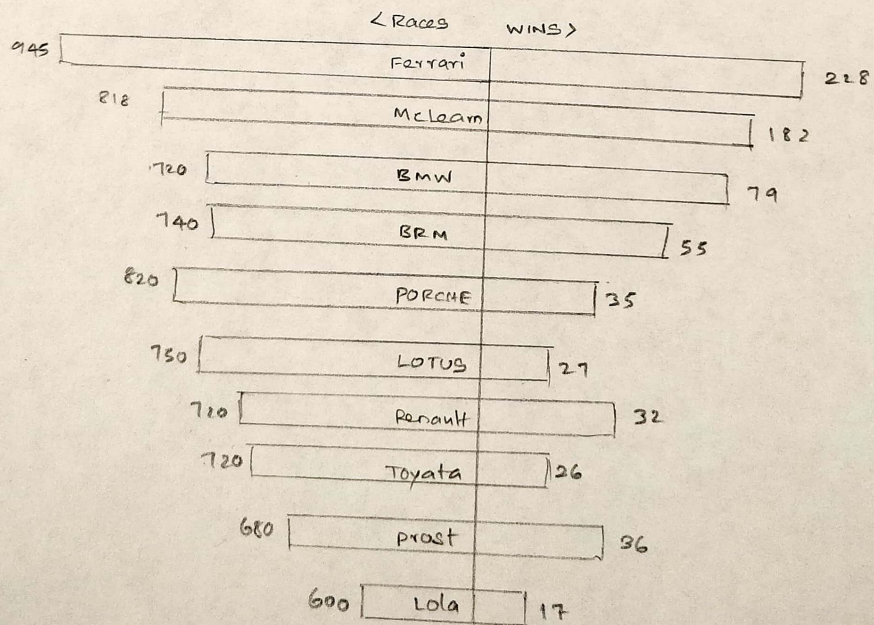


Team info:



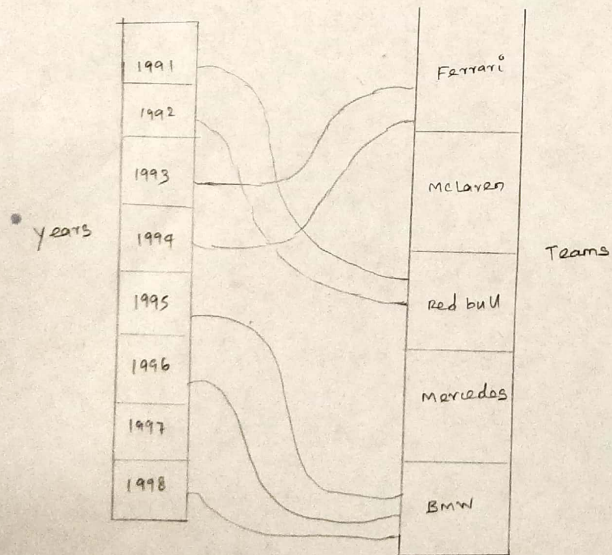
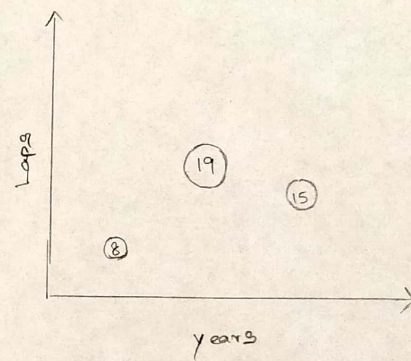
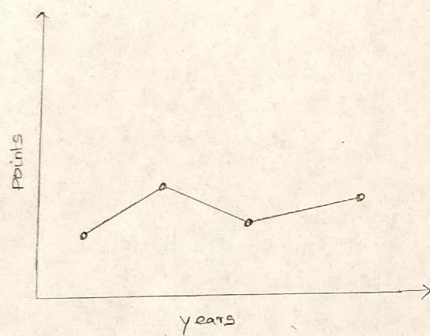
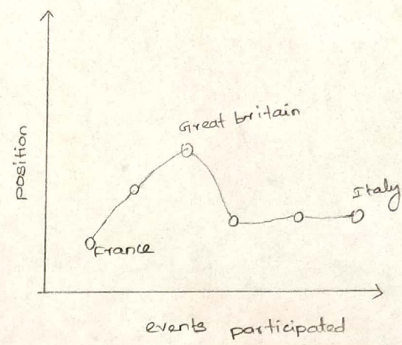
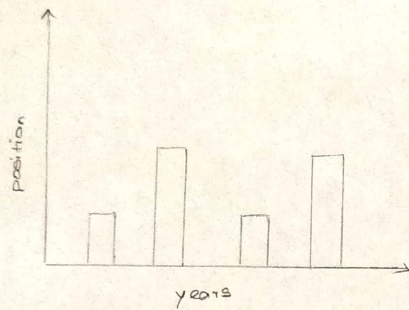
Treemap for teams

Team Stats: Races/Wins

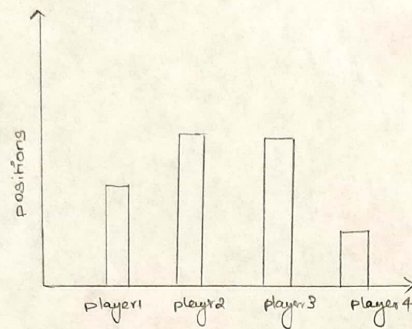


Individual Driver Statistics Images

Individual driver stats:

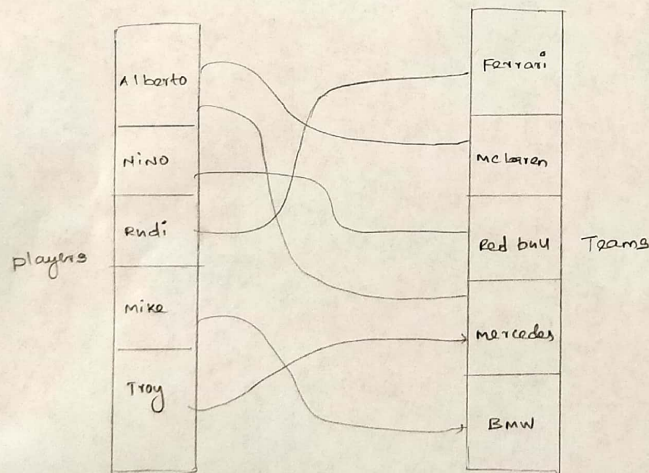
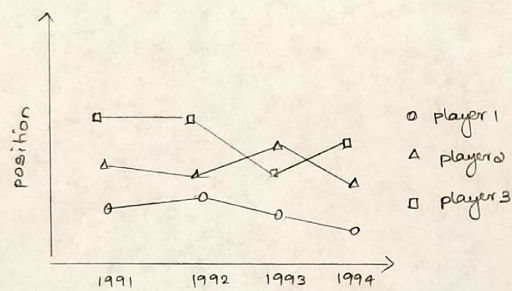


comparison of drivers per year



events around the world

- player 1
- △ player 2
- player 3



While the visualizations above depict a lot of interesting correlations, we could figure out that these visualizations are very generic and does not answer any specific question or give any distinct insight to the data. After a feedback session with a peer team, and with the TA, we came up with a modified version of our visualizations. The visualizations shown below are well organized and provides a directional flow to the project.

Modified Design

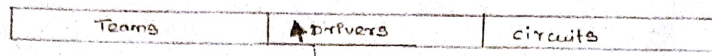
One of the images below shows the home page of our project. This would contain three tabs - Teams, Drivers, Circuits.

The first tab - "Teams" would have a year chart from 1950 - 2018 . Selecting a particular year would give us a stacked horizontal bar chart with races/wins of the top most car teams in that year. There would also be a year slider functionality over the year chart that would help the user visualize the top teams over a specific period of time(for example from 1950 -1985). In order to make it interactive, we have decided to display the names and other details such as nationality and points of the drivers who represented a particular team on clicking the bar corresponding to the team.

The second tab - "Drivers" will have a drop down box of years with to and from labels. On selecting a range of years, we would get a list of drivers who drove during those years. Clicking on a particular driver would give his statistics . Below the names of the drivers and the description of the selected driver(teams he represented, nationality, date of birth), we would give the user an option to compare the drivers. The user can select one or more drivers to compare(using a drop-down box) with the default selected driver. This would display two line charts with the average lap time and the position between drivers compared.

The third tab - "Circuits" will represent a map that highlights all the regions in the world that have formula 1 circuits. We would further have a button "Top Circuits". Clicking on that would modify the map and highlight all the regions of the world that have the top most circuits. There would also be a table below showing the top circuits of the world with it statistics. We would rank the circuits based on the number of races that happened in that circuit. The more the races, the higher the rank. To make the map interactive, we have decided to provided an info box. Clicking on a region would give us the name of the circuit, the number of races that happened on that circuit and the name of the team that was the most successful on that circuit with respect to wins and the number of wins that the successful team had on that circuit.

Main Page:



Drivers

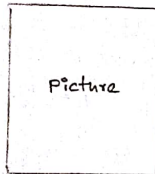
Dropdown

From TO

Display
all drivers
names

• John wall	• Andrew Bynum	• Brandon Roy
• Yao ming	• Brad miller	• Paul Pierce
• Jason Kidd	• Ryan Millsap	• Tony Parker
• Theo Ratiff	• Tim Duncan	• Dale Davis

Driver 1 stats:



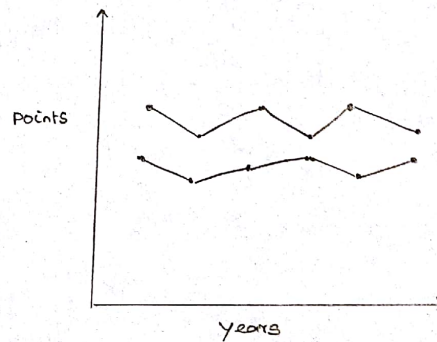
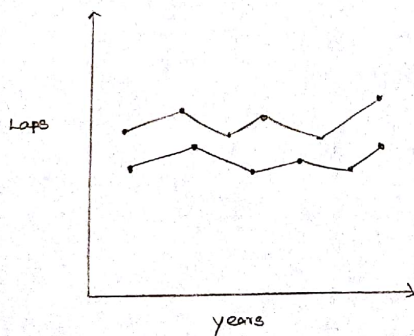
Name: Michael Schumacher

Nationality: Germany

Birthday: 1969-01-03

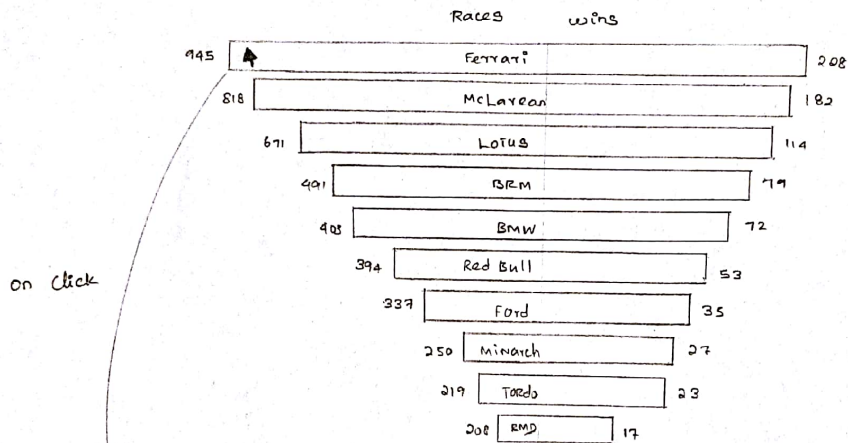
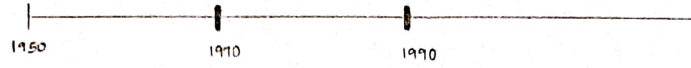
Teams: Jordan Grand Prix, Benetton, Ferrari, Mercedes.

Comparison → dropdown



Team - (constructors)

year slider



Drivers (Ferrari)

	Laps	nationality	points
Sebastian Vettel	58	Canada	90
Kimi Raikkonen	56	France	89
Lewis Hamilton	43	Germany	88
Fernando Alonso	42	Hungary	84
Felipe Massa	40	Spain	82
Michael Schumacher	39	Malaysia	80

circuit (Map)

All

Top circuits



Circuit : Las Vegas
Races : 65
Most wins: Ferrari
No. wins: 31

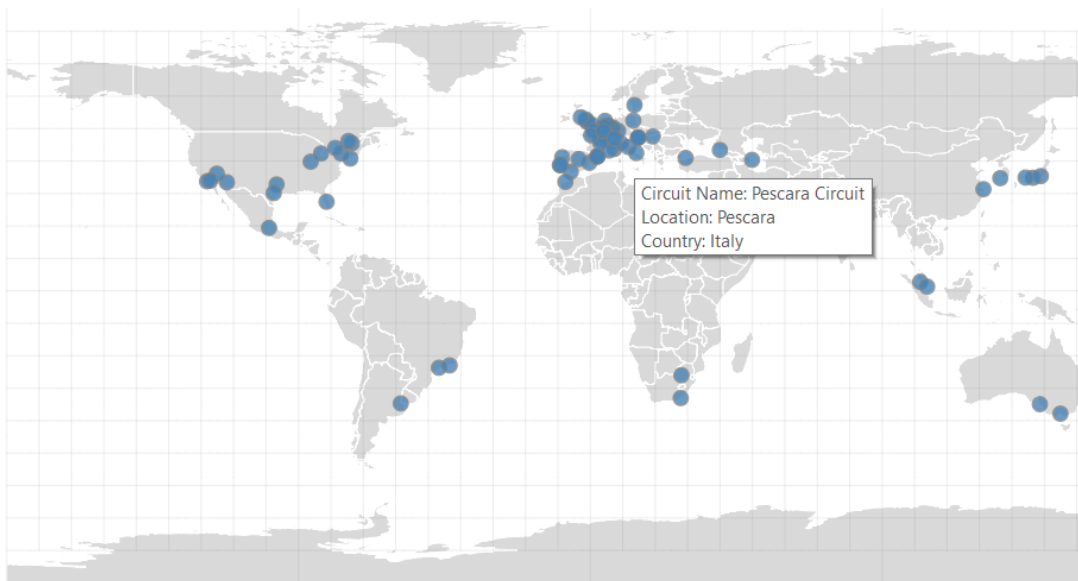
Top circuits:

Las Vegas	Nevada	USA	65
Gilles Villeneuve	Montreal	Canada	64
Hungaroring	Budapest	Hungary	51
Sepang	Kuala Lumpur	Malaysia	48
Bahrain	Sakhir	Bahrain	40
Hockenheimring	Hockenheim	Germany	38
Shanghai	Shanghai	China	35
Albert Park	Melbourne	Australia	34

Implementation

We decided to start with the implementation of the objectives defined for the circuits tab of the project. The visualization below denotes a map representation of the regions that contain the formula 1 circuits all over the world. This would be the default view of this tab. We would further be adding a button to this page which would generate a map representation of the regions that contain the top most circuits of the world(top most would be evaluated by the number of races that happened on a particular circuit). This visualization would answer one of our project objectives.

Formula 1 Stats



References

- [1] Ergast Developer API : <https://ergast.com/mrd/db/#csv>
- [2] Formula 1 : <https://www.formula1.com/>