BEYOND THE LAP

Novel Perspectives on Formula 1

A report by

Arundhati Balasubramaniam,
Hans Kristian Bjørgo Kværum, and
RIccardo Lionetto

as part of the final deliverable of

COM480 - Data Visualisation

INTRODUCTION

This project embarked on an ambitious journey to demystify the world of Formula 1 racing through an interactive visualization platform. With a rich history dating back to 1950, Formula 1 is celebrated for its prestige, global fanbase, and technological sophistication. Modern F1 cars, equipped with over 300 sensors, generate a staggering 100,000 data points per second, accumulating approximately 1.5 terabytes of data over the course of a race.

Our project aimed to bridge the knowledge-gap for fans wanting to get into the sport, as well as make novel insights accessible to enthusiasts. Through interactive visualisations, the user can familiarize themselves with F1s history and highly relevant topics currently being discussed.

This process book details our journey in the creation of this platform, from initial concept to the final product. We delve into the challenges we encountered, the design decisions we made, and the contributions of each team member. Hence, we aim to provide a thorough understanding of our project, highlighting the novelty and innovativeness of our approaches in efforts to make Formula 1 insights engaging and informative to all users.

DATASET & PREPROCESSING

Most of our project is based on the comprehensive Kaggle dataset "Formula 1 World Championship (1950 - 2023)". Among other things, this dataset contains information about races, circuits, drivers, teams, pit stops and dnfs. While our original intention was to supplement this dataset with the Fast F1 Ergast API, which is also the original source of the dataset, to augment the dataset with detailed telemetry and lap-time data, we conceptually moved away from ideas leveraging this information during milestone 2. It also proved difficult to effectively merge data from the two sources, leading us to focus solely on the Kaggle dataset for milestone 3.

Despite the dataset being fairly high quality, a lot of efforts were put into data processing. Initially, this involved correctly parsing data, merging tables and create proper aggregations that could be utilized for our visualizations. Given the specificity of the analysis we wanted to do, a lot of processing also went into transforming the data to its desired formats, leading to custom json-files for each of the visualizations. We also augmented the dataset with crucial continent-information using pycountry-convert.

VISUALISATIONS

Racing Bar Chart

Considering the topic, we thought a cool way to visualize F1s history from a statistical point of view, was to create an interactive animation showing the top 10 drivers or teams “race” for the most impressive cumulative numbers in terms of select criteria. Using two dropdowns, the user can select what data (teams or drivers) and metric (races, points, poles, wins, podiums and dnfs and championships) to show, leading to an auto-play animation of horizontal bars sorted in descending order. The transition between years takes 1s, and we include cumulative counts from 1950 to 2023.

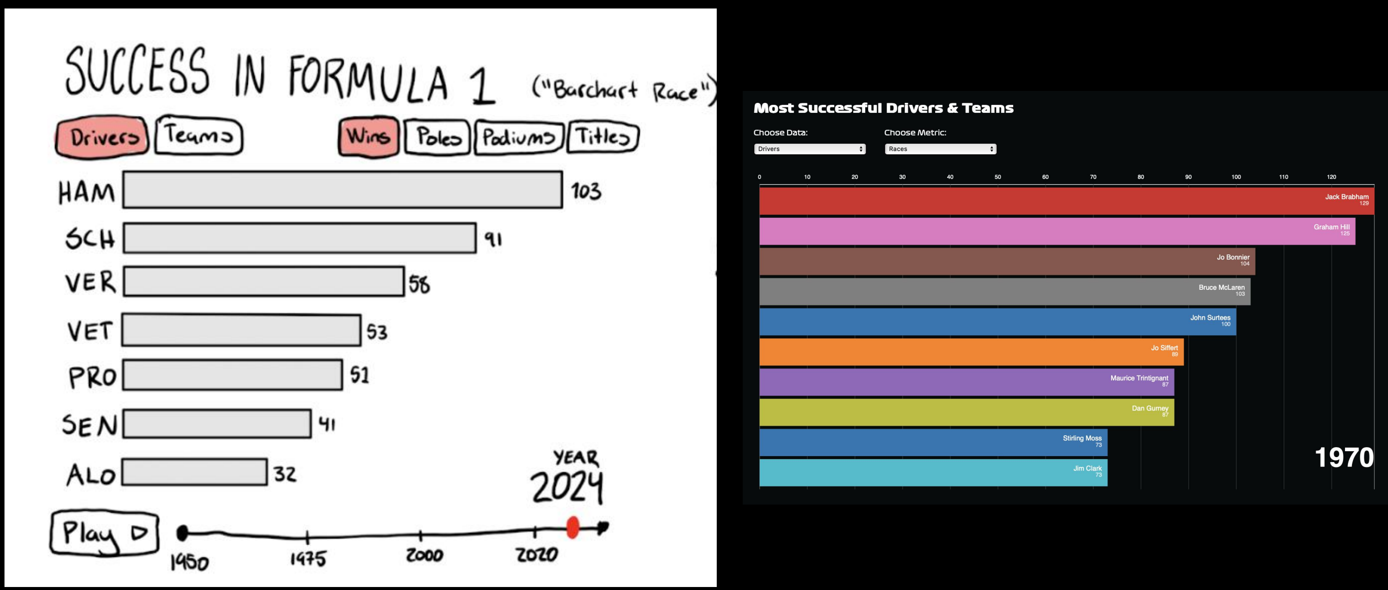


Figure 1: Sketch vs. final implementation of facing bar chart

While the overall concept remains the same, our final design deviates a bit from the sketch and the prototype. We opted for dropdowns instead of buttons both because of visual simplicity, but also scalability as it is easier to expand with an additional option than to create a new button for a different metric. We also modified text placement to clean it up a bit, as well as use ordinal colours instead of our discrete red-gradient theme to prevent changing the colour of the bars throughout the animation. This makes it easier for the user to associate a driver or a team with a colour.

Technically, we did not run into many issues, and this visualisation is by far the simplest one on the site. However, to further improve it, we would have liked to add a start/stop button to allow the user to pause the animation, and a timeline to allow the user to skip to the the year of interest, allowing for more control of the animation.

Global Presence

While F1 is often considered a global racing series, having raced on all continents, its strongest presence has historically been in Europe. Post Netflix' "Drive-to-Survive", we wanted to explore whether there are any notable shifts or trends hidden in the race calendars, such as efforts to expand to new markets. Thus, we wanted to combine three charts:

1. **A bubble map** showing the location of circuits F1 has raced on. The size of the bubbles represents the number of races, and upon hover, the user can see when F1 raced there (first and most recent year). This map showcases the aggregated global presence since 1950. We also wanted to color the bubbles based on the continent.
2. **A stacked horizontal barchart** giving context to the bubble map, aggregating the number of races per continent. This chart assists the map by putting into numbers what the bubbles gives an impression of.
3. **A stacked area chart** representing number of races per continent as a time series. This chart has two main functions: provide insights on historic developments both in terms of overall number of races, but also number of races per continent.

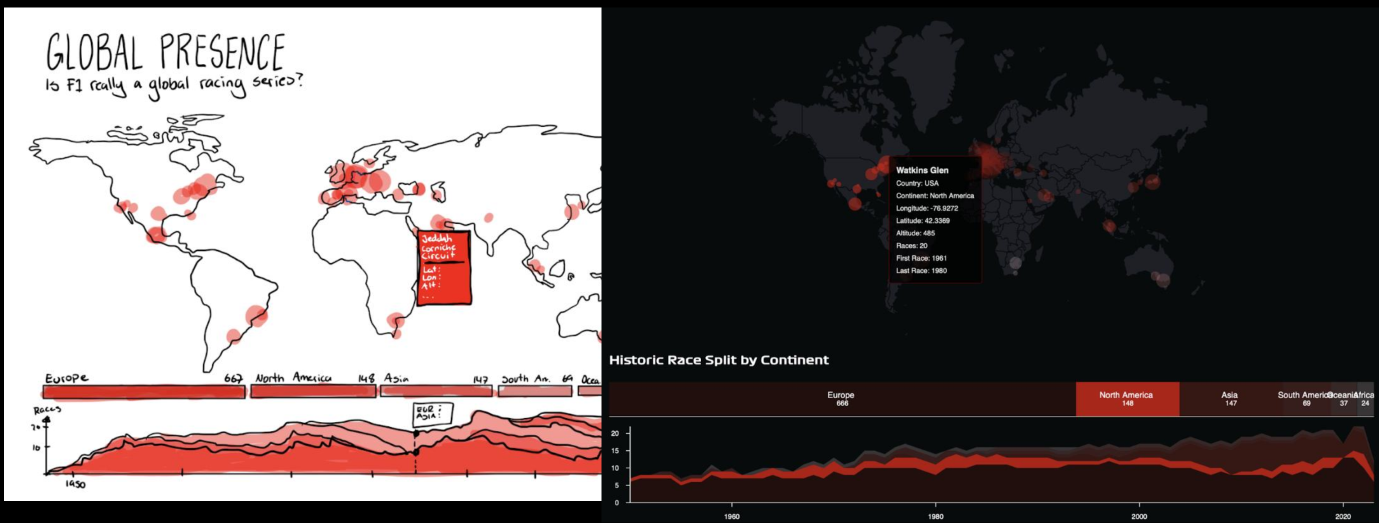


Figure 2: Sketch vs. final implementation of global presence visualization.

While the overall concept remains the same, our final design deviates a bit from the sketch and the prototype (see Milestone 2). We opted for dropdowns instead of buttons both because of visual simplicity, but also scalability as it is easier to expand with an additional option than to create a new button for a different metric. We also modified text placement to clean it up a bit, as well as use ordinal colours instead of our discrete red-gradient theme to prevent changing the colour of the bars throughout the animation. This makes it easier for the user to associate a driver or a team with a colour.

Technically, we did not run into many issues, and this visualisation is by far the simplest one on the site. However, to further improve it, we would have liked to add a start/stop button to allow the user to pause the animation, and a timeline to allow the user to skip to the the year of interest, allowing for more control of the animation.

Impact of F1s Point System on Championship Outcomes

A hot-topic that is debated on a seasonal basis is the point-system of Formula 1. Currently, only the finishing top 10 drivers get awarded with points, thus making it especially hard to accurately and fairly rank teams that are not consistent contenders for points. The financial impact(s) and prestige associated with final championship order, has prompted fans, teams and drivers to lobby for a new system, resulting in an ongoing investigation into a revision of the point system. To come the F1 Commission in advance, we wanted to construct a visualisation that allow users to explore championship outcomes under different point systems (old, current and new propositions). This visualisation would consist of three parts:

1. **A header showcasing the top 3 drivers** or teams and their total points scoring, effectively highlighting the most significant results.
2. **A line chart showing seasonal point accumulation** per championship participants. This demonstrates how teams and drivers performed worse/better throughout the season, implicitly giving an impression of in-season developments in performance on track.
3. A heatmap detailing point scoring per championship participant per round/race. This plot provides granular insights into both the results of each round, also highlighting dnfs.

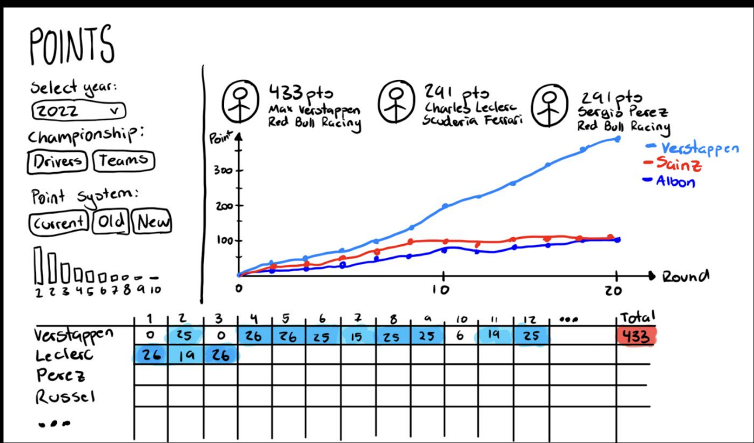


Figure 3: Sketch vs. final implementation of the Revised F1 point systems.

While the final product meets our expectations, some key shortcomings include the lack of a legend to assist the line chart, and colouring of the labels on light background for the "total column". Both of these things should be fairly straight forward changes that we unfortunately did not have time to make.

Additionally, we would have liked to add a parenthesis showing the change in position (compared to the current system) to further highlight the impact of the point system. We experimented with this feature initially, but postponed the implementation after a failed attempt, in order to make all the other plots work. Unfortunately, we did not have time to revisit it, but admit that it would significantly enhance the accessibility of a key insight.

To combat this, we ensured that the feature is well explained for our novice audiences with the use of an info button.

Revising F1s Race Calendar to Minimize Travel & CO2 Emission

As our project establishes quite early in the final data story, F1 is a global racing series. Consequently, a lot of travelling is necessary to move staff, parts and cars to each race. Given Formula 1's commitment to be Net-Zero Carbon by 2030 as part of their wider Sustainability Strategy, we wanted to explore how simply reorganizing the ordering of the races would impact travel from race to race.

We acknowledge that our analysis is based on fairly strong assumptions, given that the teams' travel schedules are a lot more complex than simply moving staff to the correct location (including travel back and forth from the factory and headquarters between races, along with all logistics associated with manufacturing parts, assembling them at a different location and bringing them to races). Additionally, seasonality plays a crucial role in race planning as weather conditions in some parts of the world would prevent safe racing on certain dates or periods. However, by ignoring these constraints, we were able to construct race calendars that significantly reduced the travel between races, over 50% in some cases. We hypothesize that these revised calendars also would reduce the actual travel of teams.

Method: after testing different approaches of estimating the Hamiltonian path, we ended up with a nearest-neighbour. The connectedness of our graph allowed it to perform surprisingly well despite its simplicity, often finding shorter paths than more sophisticated optimization-models. Additionally, we constrained the problem by defining a start node equal to the first race of the original calendar.

The visualisation consists of two key elements:

1. A connection map showing the travel paths taken from one circuit to the next. to differentiate between the original and revised paths, we used a dashed line for the original connections.
2. A table displaying the ordering of races in the original and the revised calendar. This adds clarification to a map that could easily become messy with a lot of connections crossing each other.

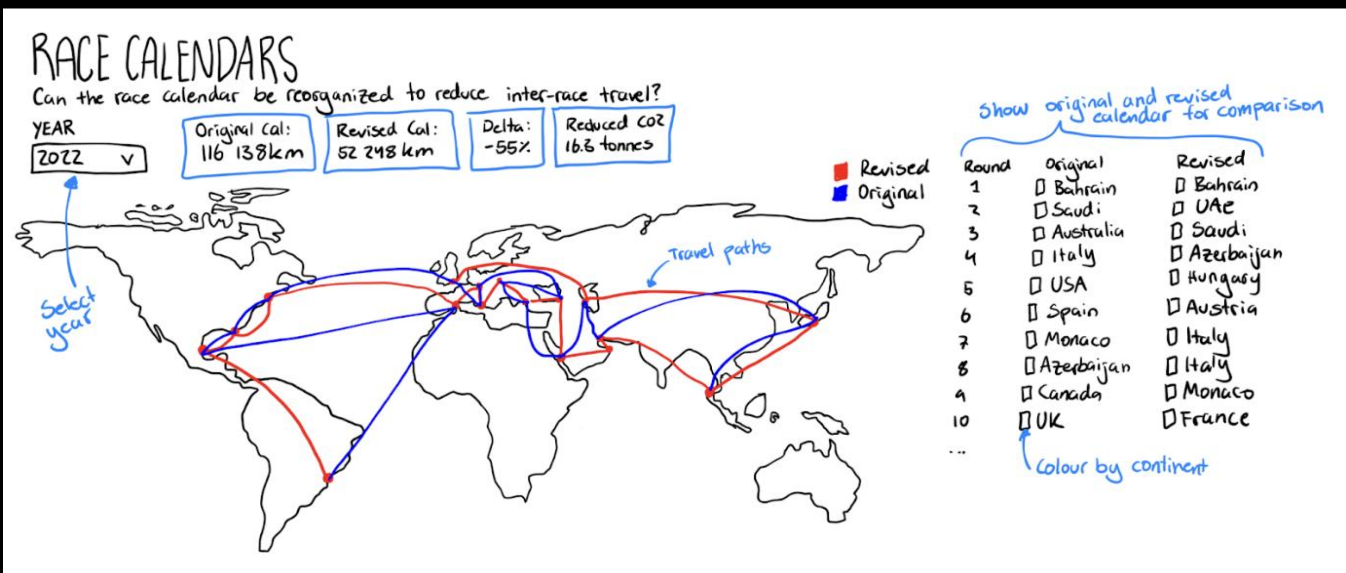
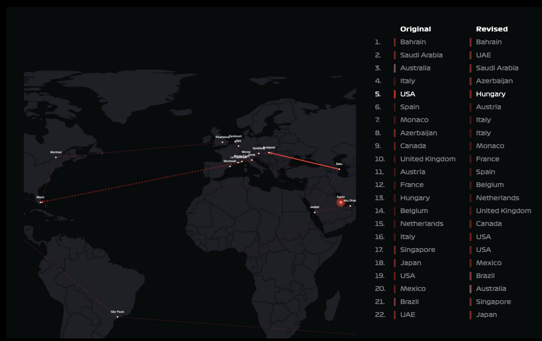


Figure 4: Sketch of race calendar visualization

This visualization was by far the most complex one, both in terms of preparing data, performing the analysis and putting it together. As the paths crossing each other tend to make for a messy visualization, we added route highlighting with labels for distance travelled upon hover in the map. Additionally, the user can hover over the calendar to see a comparison of the corresponding edges taken by both calendars, with dynamic repositioning and scaling applied to ensure the user actually can see both edges (regardless of previous interaction with the map). The start node (circuit) is also highlighted on the map.

However, we were unable to implement all the features we would have liked. The zoom interaction is not perfect (as all components does not scale with user-applied zoom, and hovering over the circuits could yield an informative tooltip.



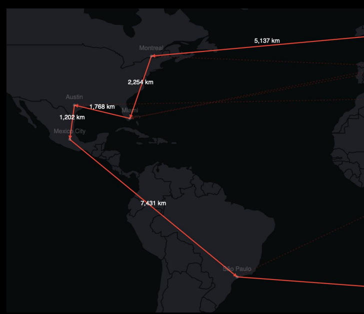


Figure 5: Screenshots of race calendar module in use with different hover effects.

CONSIDERATIONS

For the observant, our final product does not include any other plots than the ones we defined as our MVP in Milestone 2. Nevertheless, we could easily have produced a greater number of visualizations as there were no shortcomings of ideas (nor data to analyse). Instead, we spent a lot of time creating interactions, adjust styling, and make everything coherent. However, the following list summarizes things we would have liked to do:

1. **Implement more visual cues** to make certain insights more accessible. This includes legends, more advanced hovers and the positional changes caused by a different point system.
2. **Additional controls to visualisations**, such as a start/stop button and a timeline for the racing bars, and +/- buttons on the maps
3. **Additional analysis** exploring DNF causes and associations between drivers, teams and circuits using a parallel set, exploring pit stops and tire strategies on different circuits
4. **Bug fixes and responsiveness** to further improve the user experience of our site

This project was a joint effort between the three team members. More specifically, our contributions are outlined below:

Arundhati: Responsible for creating the main D3 skeleton of the visualisations (sans major styling) as well designing a major part of the high-fidelity prototype on Figma. She also worked on the report's content and styling and helped developed the design system we used.

Hans Kristian: Responsible for the final implementation of the visualisations on the website as well as the final layout for the Figma prototype. He contributed the content for the final report and textual explanations for the graphs. He also contributed the visual language of the website.

Riccardo: Responsible for plot styling and adding UX nuances to the visualisations to make them more accessible. He set up the deployment server, cleaned up the landing page to make it more aesthetically pleasing and adding custom CSS to match the design system we designed.

Overall, we were happy with how the work was divided among us and it was difficult to isolate the effect that each individual member had without overlap with another.