UCD School of Medicine
Scoil an Leighis UCD

Assignment 2: Developing a Novel InfoVis

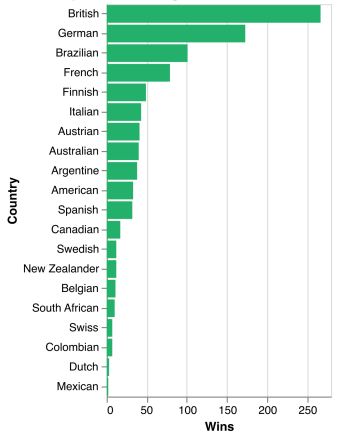
COMP47970 —-Information Visualisation

Anastasiia Deviataieva

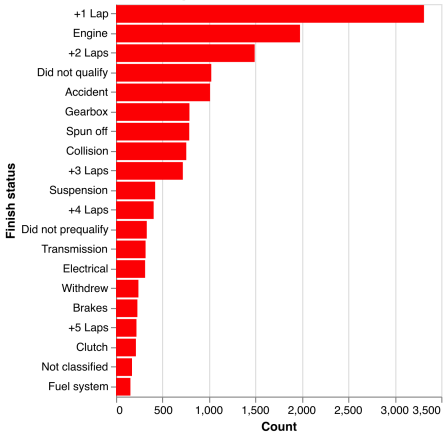
Student Number: 24100519

May 2025

Top 20 Leading Nationalities in F1 Wins

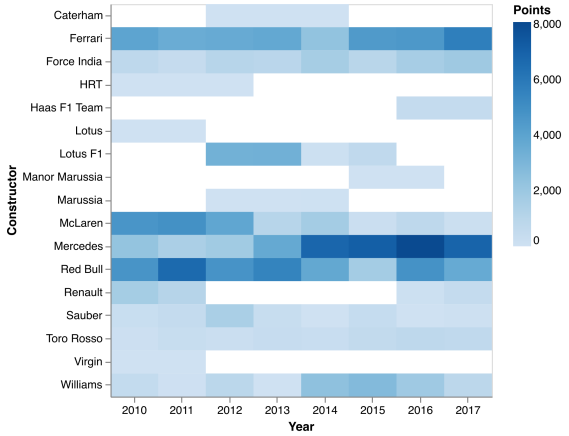


Top 20 Non-Finish Statuses

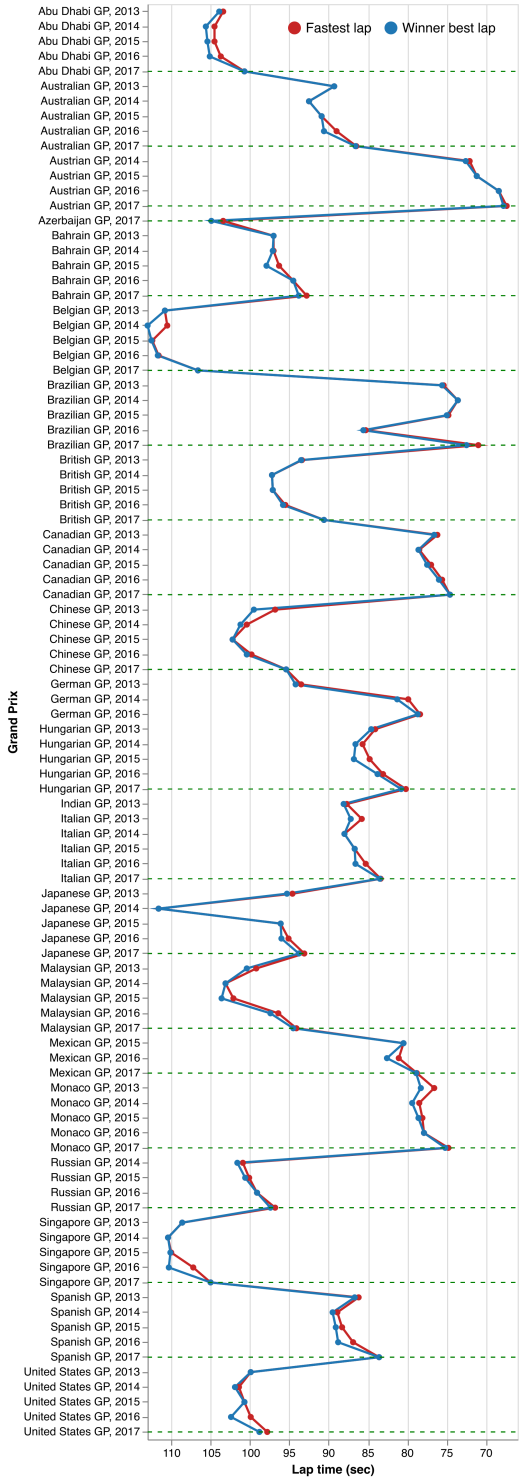


Constructor Points by Season (2010–2017)

Total championship points per team per year

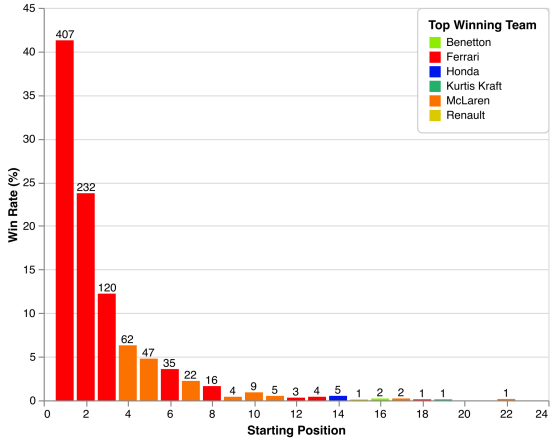


Fastest Lap vs Winner's Best Lap (2013–2017)



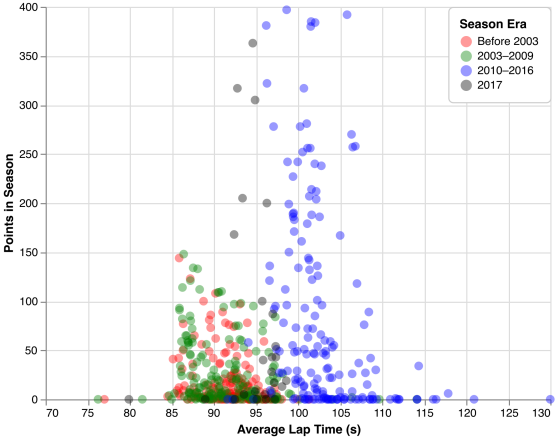
Win Rate by Starting Position

Numbers above bars indicate total wins from each starting position

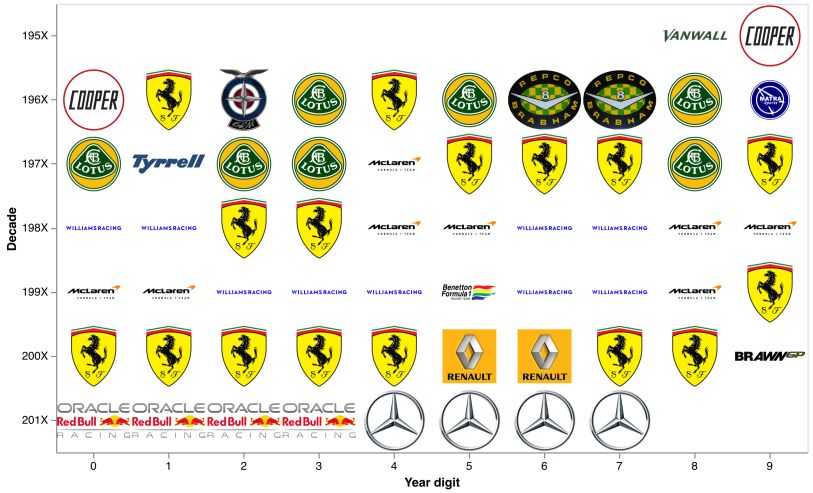


Avg Lap Time vs Season Points per Driver

Each point represents a driver's average pace and total points in one season



Constructor Champions by Season (1958–2017)



For this project, I prepared a dataset of Formula 1 race data spanning from 1950 to 2017. The data was sourced from an open Kaggle repository, which compiles official historical records into structured CSV tables. While standard standings and championship summaries are widely available through the FIA and media outlets, my goal was to uncover fresh perspectives —novel statistics, trends, and comparisonsthat go beyond conventional podium results.While the Fédération Internationale de lAutomobile (FIA) and media outlets frequently publish standard standings and championship tables, my goal was to surface fresh perspectives —novel statistics and crosscomparisonsthat go beyond the usual podium finishes. I processed and combined multiple tables (races, results, lap times, standings, constructors, drivers, and status codes) in Python, then crafted a series of focused Vega-Lite charts to reveal patterns.

Constructor Champions by Season (1958 – 2017)

Chart Type & Encodings I use an image grid: each cell places a constructor logo at the intersection of its championship decade (Y-axis) and the last digit of the year (X-axis). The url channel maps each champions name to a publicly hosted PNG or SVG emblem, while the decade and digit fields drive ordinal band scales.

Why an Image Grid? Logos are instantly recognizable and celebrate the trophy tradition: winners physically receive a constructors trophy each year. A grid of emblems communicates who and when in a way that raw text or colored bars cannot. Alternative designssuch as a colourcoded treemap or stacked bar —would obscure the iconic branding that fans instantly associate with each era.

Insight This chart makes constructor dominance obvious at a glance: Ferraris strong showings in the 1960s, McLarens 1980s peaks, Red Bulls early 2010s streak, and Mercedess recent supremacy. Clusters of logos reveal multiyear dynasties, while scattered emblems (e.g. Vanwall, Cooper) highlight oneoff champions.

Top 20 Nationalities by Career Wins

Chart Type & Encodings A horizontal bar chart shows each country (Y) versus its total driver victories (X). Bars are sorted descending by win count; a single uniform colour emphasizes total tallies without unnecessary distractions.

Why a Bar Chart? Bar length is an effective quantitative channel for comparing discrete categories, especially when the category labels (country names) are textual and ordinal ranking is important. I considered a pie chart or bubble chart but rejected them because precise comparison and clear ranking are critical.

Insight British and German drivers lead by a wide marginreflecting both historical team depth and larger driver poolswhile standout nations like Brazil (largely thanks to legends such as Ayrton Senna and Emerson Fittipaldi) also feature prominently. The chart also hints at regional overrepresentation (Central Europe) influencing totals.

Top 20 Non-Finish Statuses

Chart Type & Encodings Another horizontal bar chart enumerates all non-Finished statuses (e.g. +1 Lap, Engine, Accident) by frequency. I colour all bars red to signal caution or failure.

Why a Simple Histogram? These are nominal categories with a single count metric; a bar chart clearly conveys relative frequency without overengineering. A stacked bar would add complexity without benefit, and a treemap would hide precise counts.

Insight The most common reason for nonfinish is being lapped at least once (+1 Lap), which often stems from underpowered cars or safety cars bunching up the field. Mechanical failures (engine, gearbox) and accidents follow. This chart underscores reliability and equipment gaps as perennial challenges for smaller teams.

Constructor Points by Season (2010 – 2017)

Chart Type & Encodings I chose a heatmap: years (X) versus constructors (Y), with cell colour encoding total points. A sequential blue scale emphasizes intensity.

Why a Heatmap? The two dimensions (time and team) form a natural grid, and colour facilitates quick comparison of annual performance trends. Alternative line charts or slopegraphs would require dozens of overlapping lines and become illegible. A heatmap neatly summarizes eight years at a glance.

Insight The map shows Red Bulls darker cells from 2010 – 2013 and Mercedess from 2014 – 2017, confirming shifts in competitive balance. Mercedess deeper colour saturation suggests more dominant margins (fewer challengers able to score close points) than Red Bulls era. Consistent mid-to-dark blocks for Ferrari and McLaren indicate sustained competitiveness despite not always winning the title.

Fastest Lap vs Winners Best Lap (2013 – 2017)

Chart Type & Encodings I use a dualline chart with overlaid points, plotting each Grand Prix (Y or X axis) with two series: outright fastest lap (red) and winners fastest lap (blue). Point marks highlight each data sample.

Why a DualLine? I wanted to compare two time series across many events in chronological order. A line preserves temporal ordering and reveals event-to-event variation, while points mark exact laptime values. A scatterplot of the two metrics would lose the racebyrace context and chronological narrative.

Insight Certain circuits (Abu Dhabi, Spain) show winners not setting the fastest lap in 4 of 5 seasonsperhaps due to strategic tire management late in the race —whereas in Britain, Brazil, and Australia winners more often secure the fastest lap. This highlights how circuit type and race strategy influence the correlation between lap speed and victory.

Win Rate by Starting Position

Chart Type & Encodings A stacked bar chart with normalized (percentage) heights shows win rate by grid slot (X) and overlays absolute win counts as data labels. Each coloured segment corresponds to the constructor most successful from that position.

Why Stacked Bar with Labels? I combined a percentage axis (to compare relative probabilities) with numeric annotations (to anchor percentages in real counts), and coloured bars to show which teams dominate certain grid slots. An alternative heatmap would hide granular counts; a line chart would not emphasize the discrete nature of starting positions.

Insight About 77 % of wins come from the top three grid positions, underscoring qualifying importance. Ferraris red segments dominate P1 –P3, reflecting both poletowin conversion and the teams overall success. Wins from lower grid slots (10+) are exceedingly rare.

Average Lap Time vs Season Points per Driver (1996 – 2017)

Chart Type & Encodings A scatterplot where each mark represents one drivers performance in a single season:

- X: the drivers seasonlong average lap time (avgLapTimeYear), in seconds.
- Y: the total championship points they scored that year (seasonPoints).
- Color: a computed era field divides seasons into four buckets —Before 2003 (red), 2003−2009 (green), 2010 − 2016 (blue), and 2017 (black) —reflecting changes in scoring rules and technical regulations.
- Opacity: set low (0.3) to expose dense clusters without overwhelming overlap.
- Tooltip: reveals driver name, season, avg lap time, and total points.

Why a Scatterplot? We want to explore how points accumulate relates to season pace across hundreds of driverseasons. A scatterplot shows each pair individually, revealing both the overall trend and exceptional cases. Alternatives like bar charts or heatmaps aggregate away that nuance, while a line chart would misleadingly imply continuity within a drivers career rather than independent season snapshots.

Insight

Points driving pace: Highscoring seasons (points) nearly always correspond to drivers with low (faster) average lap times —but be cautious: a driver with a handful of strong finishes can post a deceptively low average if they retired early in other races. Scoring eras: Red (Before 2003): only top 6 scored, so red points cap out lower on Y; Green (2003 − 2009): expanded to top 8, green dots extend higher; Blue (2010 − 2016): top 10 scored and technical restrictions slowed cars –blue cluster shifts right (slower laps) but reaches much higher on Y; Black (2017): same 10-scorer system but faster regulations push black dots left compared to 2010 − 16 blues; Outliers & samplesize effects: Some drivers appear with very fast averages yet low points –often because they only completed a few races. The scatterplot lets us spot and investigate these anomalies rather than hiding them in aggregate metrics.

By combining diverse chart types a multifaceted picture of Formula 1s competitive landscape was present. Together they reveal not only who won and when, but why and how factors like starting position, vehicle reliability, and circuit characteristics influenced outcomes.