

Visual F1

Trabalho Final - Visualização de Dados

Ana Carolina Erthal
FGV EMap
Niterói, Brasil
acarolerthal@gmail.com

Felipe Lamarca
FGV EMap, FGV CPDOC
Rio de Janeiro, Brasil
felipe.lamarca@hotmail.com

Guilherme de Melo
FGV EMap
Rio de Janeiro, Brasil
guilhermedemelo@gmail.com

Abstract—Esse artigo aborda diferentes visualizações para representar informações sobre Fórmula 1. O esporte depende diretamente do fluxo constante de informações e dados, já que é a partir das análises de desempenho que ocorre o desenvolvimento das construtoras. Mas, em geral, essas informações não chegam a um público mais amplo — e, nos raros casos, as representações são pouco gráficas e nada atrativas. O objetivo principal do desenvolvimento do projeto é, portanto, abordar técnicas relevantes de representação de dados de Fórmula 1 de forma útil para um público mais específico, e de forma atrativa para o público em geral.

Index Terms—Fórmula 1, Visualização de Dados, *Storytelling*, Séries Temporais, Bar Race, Violin Plot.

I. INTRODUÇÃO

O mundo dos esportes automobilísticos é uma fonte rica de dados. Em uma só corrida, conseguimos extrair milhares de informações, desde as posições dos pilotos a cada milésimo de segundo, observações sobre a performance do carro em si e do desempenho da própria equipe em manutenções — os famosos PitStops. Dezenas de corridas são realizadas todos os anos, e a modalidade se aproxima de completar um século de existência. O fluxo de dados é gigantesco.

Além disso, essas informações são de fundamental importância para o desenvolvimento do esporte. Para as construtoras, não acompanhar seus próprios dados — e também de seus rivais — é sinônimo de fracasso. Cada uma das equipes conta com times especializados em analisar esses dados e extrair qualquer vantagem possível. Naturalmente, é um interesse particular dos competidores e das empresas financiadoras que qualquer tendência no desempenho das equipes não seja divulgada.

Para os fãs, apesar do desenrolar animador que se dá a partir da disputa nas pistas, restam poucas informações visuais sobre o assunto, mas muitas na forma de tabelas e números. O leque de visualizações úteis sobre o assunto é muito restrito, e é objeto deste trabalho abordar técnicas e formatos para desenvolver visualizações relativos aos aspectos mais relevantes da Fórmula 1.

O desenvolvimento da página, produto final desse projeto, conta com uma forte abordagem de *storytelling*, em que buscamos inserir até mesmo o usuário desinteressado no contexto da Fórmula 1¹. Fazemos um passeio pelo passado e presente

do esporte, apresetando dados históricos da disputa, desenvolvendo uma aplicação esteticamente atrativa para diversos públicos.

II. TRABALHOS RELACIONADOS

A. *O storytelling*

O *storytelling* um dos pilares que fundamenta este projeto, já que a empreitada de mobilizar o campo de visualizações de F1 é intrínseca à questão de convencer o público de que há relevância nesse objeto.

A discussão sobre a importância do *storytelling* é bastante presente na literatura de Visualização de forma geral. [1] aborda estratégias de visualização narrativa, e buscamos seguir o caminho de estruturar a aplicação de forma lógica e interessante. O objetivo é que qualquer usuário consiga acompanhar e desenvolva interesse pelo que a aplicação e as visualizações informam.

Em particular, este artigo utiliza a estratégia do **Linear Storytelling**, que consiste em estruturar a exploração do leitor de forma a direcionar sua atenção a uma sequência organizada de informações e explorações que maximizem seu entendimento. Assim, foco da estruturação da página principal é iniciar contextualizando o leitor com o passado do esporte, passando por uma linha do tempo desde a criação até o estado atual do esporte, destacando as principais características da F1.

A organização dos elementos ao longo da página VisualF1 é pensada para que o leitor siga essa construção linear, mas que possa explorar mais a fundo os tópicos de seu interesse, retornando à página inicial quando desejar explorar outros tópicos.

B. Visualizações na Fórmula 1

O levantamento da literatura sobre visualizações gráficas de Fórmula 1 revelou um nicho bastante limitado. Uma das poucas contribuições nesse sentido é [2]. Apesar de trazer uma perspectiva interessante sobre a visualização de dados de Fórmula 1, o artigo é datado e utiliza técnicas que, frente às ferramentas de visualização atuais, já se tornaram obsoletas. O gráfico abaixo, retirado do artigo, é uma das visualizações propostas pelos autores:

Apesar de representar um avanço na época, essa é uma visualização que provavelmente não captaria a atenção do público leigo. Atualmente, é possível elaborar formas mais

¹Frequentemente nos referimos ao nome Fórmula 1 pela sigla F1.

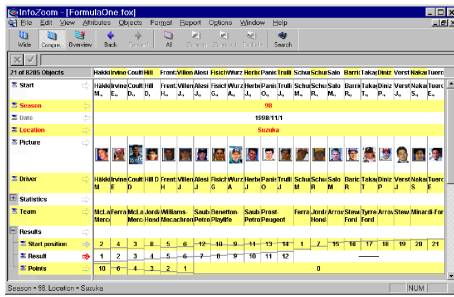


Fig. 1. Representação de dados de Fórmula 1. [2]

eficazes e atrativas de se representar esses dados visualmente. É o caso de [3], que traz algumas contribuições mais recentes que chamam mais atenção por conta da estética e da abordagem utilizada para representar as informações. No entanto, trata-se de um conteúdo que não permite interação do usuário com os dados e não explora a dimensão temporal da F1 ao longo do tempo. Esses são os dois principais objetivos que este trabalho pretende alcançar com a aplicação desenvolvida.

III. MÉTODOS

A metodologia do VisualF1 é abordar cada uma das três principais partes da Fórmula 1: as construtoras, os pilotos e os PitStops (que representam o desempenho da equipe de manutenção). O objetivo é, para cada um desses subtemas, abordar as melhores formas de visualizar cada tipo de dado e implementar essas peças de visualização na aplicação seguindo a estrutura de *storytelling* linear.

O público alvo do projeto é geral, e não apenas o público que se limita aos fãs, patrocinadores e competidores de F1. Por isso, é necessário manter uma visualização de simples entendimento, mas que passe a informação com eficácia e de forma interessante. Para alcançar o nível de interatividade necessário aos objetivos do trabalho, a ferramenta utilizada foi principalmente D3.js e mecanismos de marcação e estilização — HTML e CSS, respectivamente.

A. Construtoras e bar races

Para avaliar o desempenho das construtoras, a variável principal de analisar é, intuitivamente, o número de vitórias. O acúmulo de tantos prêmios da Ferrari é o que a torna alvo de tanta fama, e as construtoras que têm obtido significativo renome nos anos mais recentes do esporte naturalmente devem isso ao número de vitórias que estão acumulando, como é o caso da Mercedes.

Nesse caso, a escolha de um *bar race* é bastante direta. Em uma visualização que trata de vitórias em corridas, uma animação que simula uma corrida de barras representando vitórias acumuladas não só é eficaz para representar os dados de vitórias ao longo do tempo, mas também complementa bem a página.

Implementamos o *bar race* por vitórias em temporadas ou corridas, já que uma temporada é composta de aproximadamente 20 corridas. A forma animada da representação gráfica capta o interesse do usuário na medida em que a simulação

de uma “competição” entre as barras chama a atenção e faz com que o usuário aguarde o resultado final.

Do ponto de vista de estruturação, a visualização foi implementada em um notebook Observable e adicionado ao HTML da página. Ela conta com um botão de *replay* para que o usuário consiga voltar ao início da visualização quando desejar. No que diz respeito à frequência de atualização do ano, optou-se por definir um tempo que parecesse rápido o suficiente para manter a atenção do usuário, mas não tão rápida a ponto de fazer com que ele seja incapaz de acompanhar a atualização das barras.

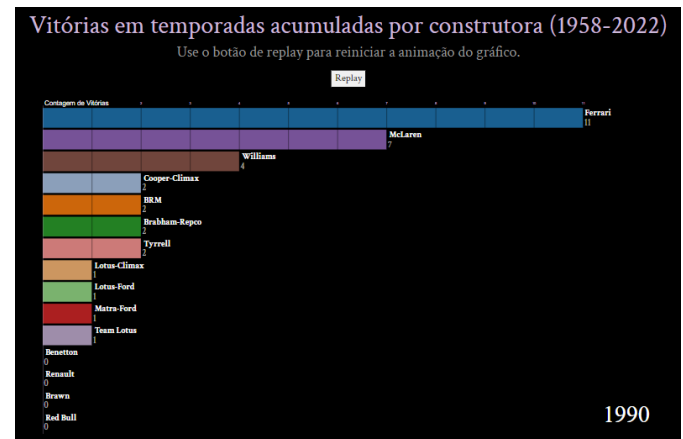


Fig. 2. Gráfico de violino: distribuição de tempos de PitStop

B. Pilotos e séries temporais

Para observar dados sobre cada piloto, a abordagem novamente é motivada por qual pergunta os dados sugerem. Normalmente apresentar o desempenho de um piloto ao longo dos anos não é uma tarefa produtiva, porque o tempo de carreira costuma ser curto. Portanto, para efeito de representação gráfica, faz mais sentido acompanhar o desempenho dos pilotos ao longo da temporada.

Como o objetivo era retratar e informar o usuário da aplicação sobre o desempenho absoluto e comparativo de pilotos com o passar de meses, a abordagem dos gráficos de série temporal é ideal. Os gráficos de linha, em geral, são muito utilizados para representar dados com passagem de tempo, e também fazem um bom trabalho em informar comparações através do tempo, que foi nossa escolha final do que a visualização deveria representar.

A visualização mostra pontos acumulativos por piloto, e após escolher o ano, pode-se escolher também quais pilotos que participaram do Grand Prix naquele ano deseja-se visualizar. Assim, a abordagem de utilizar séries temporais é bem justificada, já que é simples olhar para linhas de pontos acumulados e entender quem performou melhor ou pior ao longo da temporada, principalmente podendo escolher quantos pilotos deseja-se observar, o que evita poluir o gráfico.

O gráfico conta, ainda, com pontos no momento em que ocorreu cada corrida ao longo do ano, para que o usuário

possa explorar mais a fundo informações utilizando um *tooltip*, gerando uma interação maior entre usuário e visualização.

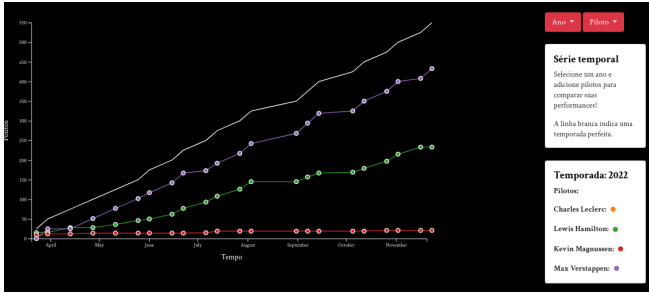


Fig. 3. Gráfico de violino da distribuição de tempos de PitStop

C. PitStops e violin plots

Para a discussão sobre as visualizações feitas para os dados de PitStops, é importante observar, de antemão, que há uma grande quantidade de dados disponíveis. A cada corrida, pilotos realizam em média 3 PitStops, e um GrandPrix conta com aproximadamente 20 corridas e com 12 pilotos. A quantidade de dados quando acumulamos todos os anos da Fórmula 1 é muito grande!

Além disso, analisar dados de PitStop imediatamente desperta uma questão. Queremos saber quais equipes realizam um melhor trabalho, isto é, conseguem realizar as paradas mais rápidas. Também é interessante e possível observar como os PitStops evoluíram ao longo dos anos. Sabemos que se tornaram mais rápidos, mas a partir de quando? Essa mudança ocorreu gradualmente? Infelizmente, algumas dessas questões não puderam ser exploradas pela falta de dados: só é possível obter dados sobre os tempos de PitStop a partir da competição de 2011.

Observar a distribuição desses tempos em cada ano foi, portanto, a abordagem escolhida. A comparação desejada deve se dar entre as equipes, e o objetivo é observar a distribuição de densidade, então a abordagem escolhida foi a utilização de *violin plots*. Essa ferramenta visual tem o objetivo de efetivamente representar distribuições de densidade em variáveis numéricas contínuas de forma suave.

Ter um tooltip sobre cada ponto de dado para que o usuário pudesse obter mais informações sobre cada ponto de dado também era altamente desejável. A observação de alguns *outliers* — isto é, um PitStop que levou bem mais tempo que o comum ou, quem sabe, bem menos tempo que o comum — pode gerar interesse de saber se, por exemplo, o piloto acabou se classificando muito mal ou muito bem na corrida, dependendo do caso. Por isso, decidimos implementar um gráfico de violino integrado com um *scatterplot*, em que podemos ver a distribuição suavizada mas também ponto a ponto, contando também com a seleção por ano e a opção de selecionar por corrida.

IV. RESULTADOS

Além da implementação de uma série de visualizações voltadas para a exploração de dados da Fórmula 1, este

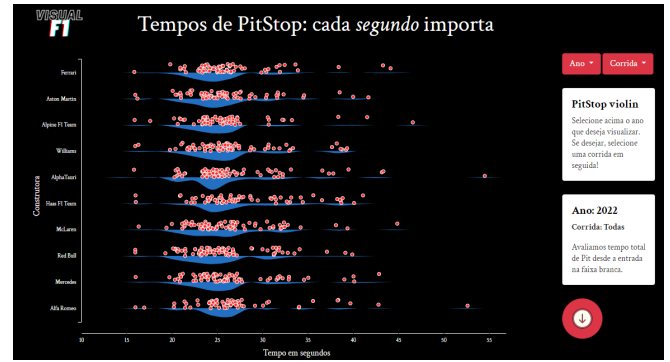


Fig. 4. Gráfico de violino da distribuição de tempos de PitStop

trabalho também faz uma composição dessas visualizações através de uma página atrativa para os usuários. Em particular, as visualizações são apresentadas em contexto usando técnicas de Linear Storytelling.

A aplicação como um todo contribui para ampliar o ainda limitado escopo de visualizações sobre F1, bem como permite a exploração de tendências gerais e específicas das principais features do esporte em questão: os pilotos, as construtoras e os PitStops.

Algumas hipóteses levantadas por este trabalho são efetivamente respondidas por essa visualização. Por exemplo, quão impactante é o tempo de PitStop no desempenho do piloto? Essa é uma questão fundamental, porque um desempenho negativo está associado diretamente a quanto tempo o piloto fica parado. Os violin plots permitem investigar essa informação e observar que outras questões também impactam o desempenho do piloto. Muitas vezes, por mais rápido que seja o PitStop, o piloto ainda não consegue um desempenho satisfatório na corrida.

As bar races apresentam a dinâmica da F1 ao longo do tempo. O usuário consegue visualizar não só a contagem geral de vitórias por ano, mas também por corrida ao longo dos anos. Dessa forma, a dinâmica das disputas da Fórmula 1 é ressaltada.

Por fim, os gráficos de série temporal complementam a composição geral a partir da inclusão de uma forma de o usuário comparar desempenhos entre pilotos ao longo do tempo. Mais do que isso, dada uma temporada, é possível identificar momentos em que o piloto teve um desempenho satisfatório ou não, considerando a maior pontuação acumulada possível.

V. DISCUSSÕES

A etapa de revisão por pares foi fundamental para a implementação de melhoras no projeto antes da entrega final. A maior parte das sugestões foram implementadas, mas algumas não foram levadas a frente. A seguir, discutimos mais detalhadamente sobre o processo de atualização do projeto a partir das sugestões dos colegas.

A. Sugestões implementadas

Um problema apontado foi a (i) não responsividade da *timeline* na página principal, muito importante na construção do *storytelling*, que ocorria devido à necessidade de usar um *image-map*, que é não responsivo. Resolvemos esse problema, e ela passou a se adequar ao tamanho da tela do usuário. Também quanto à *timeline*, já tínhamos o plano de (ii) estilizar o *tooltip* e também recebemos essa sugestão, então implementamos essa estilização não apenas na *timeline*, mas também nos gráficos. Além disso, (iii) tornamos as distâncias entre as bandeiras (anos na *timeline*) relativas, refletindo a distância entre os anos conforme uma crítica.

Foi mencionado que (iv) os *labels* de eixos nos gráficos estavam pequenos, então fizemos a pequena correção de aumentar os textos. No *bar race*, recebemos a sugestão de (v) passar a exibir os textos que descrevem cada barra à direita delas, ao invés de dentro da própria barra, evitando que o texto seja "cortado" quando o tamanho da barra é pequeno demais. Acabamos optando por (vi) usar uma paleta com mais cores (recebemos a sugestão da *tableau20*) para aumentar a diferenciação entre barras.

No gráfico de série temporal, que ainda estava menos desenvolvido no MVP, fizemos as seguintes implementações seguindo sugestões para tornar o gráfico mais informativo:

- Criamos pontos para cada corrida que ocorreu no ano selecionado, clarificando quando os eventos ocorreram;
- Uso de *tooltips* nesses pontos, para que seja possível visualizar informações extras sobre a corrida em questão;
- Tornamos o gráfico acumulativo, para que seja possível acompanhar a evolução do total de pontos por piloto por temporada, evitando um "zigue-zague" quando não acumulamos;
- Comparação entre pilotos: após escolher o ano, pode-se selecionar diversos pilotos para comparar a performance deles.

B. Sugestões não implementadas

No caso da sugestão de (i) trocar as *labels* do *bar race* para a esquerda, preferimos a sugestão mencionada acima, que foi implementada, porque observamos que essa modificação tornaria a visualização confusa quando ocorresse troca de barras e geraria sobreposição frequente de textos se eles acompanhassem a troca de barras.

Outra sugestão foi a de (ii) adicionar o símbolo da equipe ou uma foto do piloto nas barras para facilitar a identificação das equipes/corredores. Não foi possível implementar essa modificação por conta da dificuldade de encontrar símbolos ou fotos padronizadas. Mesmo assim, acreditamos que essa medida não seria efetiva na medida em que o público, de forma geral, não reconhece os corredores ou as equipes por símbolos ou fotos, principalmente a parcela que não é conhecedora de F1 - parte do nosso público alvo -, de modo que as imagens contribuiriam pouco no resultado final da visualização.

No que diz respeito à sugestão de (iii) adicionar marcas de textura para diferenciar barras, optou-se pela não implementação por entendermos que essa marca visual é

pouco efetiva para a apreensão da informação por parte do usuário. Avaliamos a possibilidade de (iv) diminuir o número de barras sendo apresentadas no *bar race*, conforme uma das sugestões, evitando a repetição de cores. No entanto, não acatamos a ideia porque a diminuição do número de barras implicaria um grande número de anos em que a visualização não seria atualizada na parte principal: se as vitórias ocorressem entre novas equipes ou novos corredores, as barras que contavam com mais vitórias acumuladas não seriam atualizadas. Isso transmitiria a sensação de que nada estaria ocorrendo na aplicação. O número de barras apresentado na entrega final minimiza esse efeito.

Uma sugestão para o *violin plot* foi a de (v) permitir quaisquer combinações de ano e corrida. No entanto, essa é uma limitação da própria competição: como as corridas não necessariamente ocorrem nos mesmos locais em todos os anos, seria impossível permitir sempre as mesmas combinações. Houve também uma ideia de (vi) quebrar fluxos na página (retirar, por exemplo, a opção de trocar qual *bar race* se deseja visualizar quando estamos na página de outro *bar race*), mas acreditamos que essa escolha torna o site mais fluido, sem que haja necessidade de retornar a todo tempo para a página inicial.

VI. TRABALHOS FUTUROS

Esse projeto, apesar de ter como objetivo abordar os principais aspectos da Fórmula 1 e estratégias de visualização eficientes e informativas, deixa abertura para a exploração futura de outras possibilidades, em especial quanto a dados de telemetria veicular.

Atualmente, os carros de Fórmula 1 contam com uma tecnologia de telemetria veicular, isto é, captação de todas as métricas dos carros durante a corrida a todo momento. Com esses dados, que podem ser acessados via API, torna-se possível explorar a visualização de dados de um veículo durante o desenvolvimento de uma corrida, tratando das velocidades e posições a cada momento, frenagens etc.

Assim, uma abordagem interessante para dar continuidade ao projeto Visual F1 seria trabalhar com essas informações para adicionar ao site uma visualização relacionada às telemetrias e desenvolvimento de corridas, possivelmente representando o deslocamento de cada carro junto a gráficos de seu desempenho.

REFERENCES

- [1] C. Tong, R. Roberts, R. Borgo, S. Walton, R. Laramée, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo, and X. Ma, "Storytelling and Visualization: An Extended Survey," *Information*, vol. 9, p. 65, 03 2018.
- [2] M. Spence and C. Beilken, "Infozoom analysing formula one racing results with an interactive data mining and visualisation tool," *WIT Transactions on Information and Communication Technologies*, vol. 25, 2000.
- [3] J. Smith, "Formula 1 2020: Sports Viz Sunday Data Challenge," *Medium*, 2020.