**Departamento de Ciência da Computação - UFMG**

**Visualização de Dados – 2018/2**

**Aluna:** Amanda Fernandes Pereira

Relatório – Trabalho Prático 2

**Introdução**

O presente relatório visa descrever as atividades desenvolvidas no Trabalho Prático 2 (TP2) da disciplina de Visualização de Dados, que teve como objetivo colocar em prática os conhecimentos adquiridos na disciplina para propor e desenvolver visualizações interativas. Para isso, foram criadas visualizações que poderão ser aproveitadas para o projeto final, que abordará os conflitos entre países ocorridos no mundo desde 1501.

Dentre as perguntas que o grupo pretende responder com o desenvolvimento das visualizações, foram escolhidas as seguintes:

* *ranking* dos X países que possuem mais relacionamentos (alianças e inimizades) com outros países nos conflitos registrados no período;
* *ranking* com o X conflitos de maior duração.

Além disso, foram desenvolvidas outras visualizações que possibilitem responder as seguintes questões:

* relacionamentos do Brasil ao longo dos anos;
* envolvimento do Brasil em conflitos ao longo dos anos.

As visualizações estão disponíveis em: <https://dataviz-tp2.herokuapp.com> e o código está disponível em <https://github.com/amandap11/tp2-visualizacao/tree/tp2-amanda>. Elas foram desenvolvidas com a biblioteca C3.js (que é baseada na D3.js).

**Visualizações desenvolvidas**

* Visualização 1: Distribuição dos votos dos candidatos à presidência

(<https://amandap11.github.io/tp1-visualizacao/distribuicaoVotosPresidente.html>)

Foi utilizada a técnica de pequenos múltiplos e são exibidos onze pequenos mapas do Brasil (um para cada candidato à presidência, conforme pode ser visto na Figura 1). Os estados foram coloridos em escala de cinza, sendo que os tons mais escuros indicam maior concentração de votos desse candidato no estado.

A decisão de utilizar pequenos múltiplos em formato de mapa do Brasil foi devido aos fatos dos brasileiros reconhecerem com certa facilidade os estados e serem muitos candidatos e estados. Dessa forma, é possível analisar os 26 estados e os 11 estados de maneira generalizada, sem impedir que a pessoa observe mais detalhadamente os detalhes de um candidato específico.



Figura 1 – Distribuição dos votos dos candidatos à presidência

* Visualização 2: Diferença de votos dos candidatos entre os dois turnos

(<https://amandap11.github.io/tp1-visualizacao/diferencaVotosPresidente.html>)

Nessa visualização foram combinadas as visualizações com barras e com pontos (conforme pode ser visto na Figura 2). As barras indicam a diferença de votos de cada candidato entre os turnos e, os pontos indicam a quantidade de votos de cada candidato em cada turno. A opção de mesclar os tipos foi devido à vontade da autora desse trabalho em mostrar, além da diferença, o quanto essa diferença significa no total de votos dos candidatos.

Não foram usados *grids* para não provocar distração com conteúdo que não agrega informação. Então, para saber valores exatos, como a visualização é interativa, o usuário deve passar o mouse sobre as barras ou sobre os pontos para que sejam exibidas a quantidade de votos e/ou a diferença de votos.

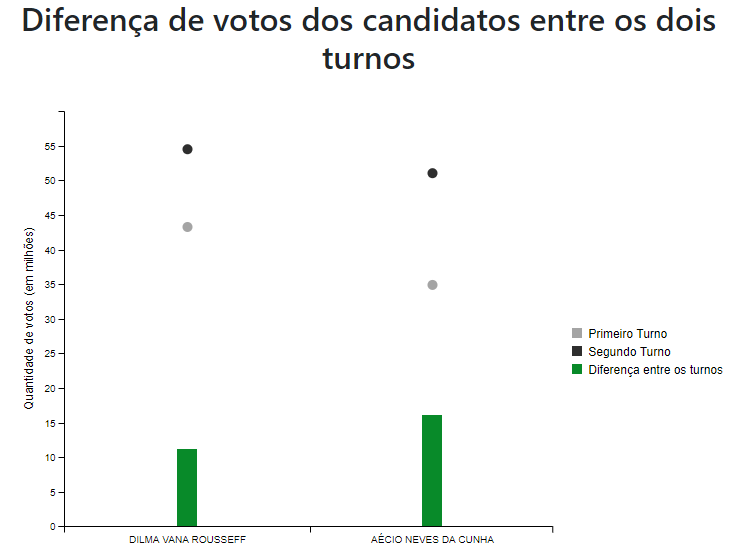


Figura 2 – Diferença de votos dos candidatos entre os dois turnos

**Análises de dados**

* Visualização 1: Distribuição dos votos dos candidatos à presidência

A partir da análise é possível identificar, por exemplo, que todos os candidatos tiveram maior concentração de votos no estado de São Paulo. Porém, alguns estados têm pouca participação, o que é o caso do Acre, Roraima, Rondônia, Amapá, Tocantins, Mato Grosso e Mato Grosso do Sul. Apesar de São Paulo ter maior número de habitantes, seria interessante analisar mais profundamente, em outras bases de dados, os motivos pelos quais esses estados têm pouca representatividade no total de votos de todos os candidatos à presidência, pois mesmo que sejam populações menos numerosas em relação a São Paulo, é no mínimo curioso pensar que esses estados estão com cores bem próximas ao branco para todos os candidatos.

* Visualização 2: Diferença de votos dos candidatos entre os dois turnos

A análise dessa visualização permite ver que, apesar de ter menor quantidade de votos tanto no primeiro turno quanto no segundo, o candidato que conseguiu maior crescimento entre os turnos não foi o que ganhou as eleições. Os dados utilizados para o desenvolvimento dessa visualização não indicam os motivos, mas seria interessante pesquisá-los mais profundamente em outras bases, pois no cenário político brasileiro da época, poderia indicar, por exemplo, que o candidato com maior crescimento era o que tinha menor índice de rejeição dentre os que não votaram em nenhum deles no primeiro turno (ou seja, o crescimento no segundo turno não estaria relacionado à preferência da população, mas sim à rejeição).

**Ferramentas**

Devido à dificuldade em utilizar a biblioteca D3.js (que era a sugestão no enunciado do trabalho), as visualizações foram desenvolvidas com a C3.js. Essa biblioteca foi escolhida por ser baseada na D3.js e por permitir que as visualizações sejam alteradas utilizando os recursos oferecidos pela D3.js. Apesar do uso ser pontual, em alguns itens das *tooltips* a D3.js foi utilizada para alcançar o formato desejado..

A segunda visualização foi feita utilizando a C3.js, cujo aprendizado é muito mais fácil. Por ser baseada na D3.js, caso seja necessário, as visualizações padrão podem ser alteradas, recurso que foi pouco utilizado nesse trabalho. Devido À facilidade de aprendizado e à possibilidade de customização utilizando a D3.js, achei melhor trabalhar com a C3.js.

**Conclusão**

O objetivo deste trabalho, que foi introduzir os alunos a ferramentas/bibliotecas para o desenvolvimento de visualizações de dados, bem como para a exploração e análise visual, foi alcançado. Foram encontradas muitas dificuldades na compreensão da biblioteca sugerida (D3.js) e, por isso, foi feita a opção por utilizar outra biblioteca (C3.js) para criar a segunda visualização. Acredito que essa dificuldade poderia ter sido amenizada se fossem dadas uma ou duas aulas de introdução à biblioteca, pois mesmo tendo algum conhecimento de *javascript*, foi difícil criar a visualização.

Em relação às análises, criei algumas opções de visualização (que não foram apresentadas nesse trabalho) para apresentação dos dados até encontrar uma que considerei melhor para cada situação e, com isso, foi possível ver que algumas visualizações favorecem e outras dificultam a análise dos dados devido à forma como eles são apresentados. Portanto, é imprescindível que o desenvolvedor tenha clareza do que quer transmitir e quais as melhores formas de apresentar os dados de forma a facilitar a identificação de padrões, o que favorece a análise.