**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, foi desenvolvida outra visualização que possibilita responder as seguintes questões que comparam o Brasil e o mundo:

* 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, o que possibilitou o uso desta em alguns pontos.

**Visualizações desenvolvidas**

* Visualização 1: Porcentagem de aliados e inimigos

(https://dataviz-tp2.herokuapp.com/visualizacao1)

A Visualização 1 apresenta um *ranking* com os países com maior número de relacionamentos desde 1501. Esses relacionamentos podem ser de alianças ou inimizades, sendo que cada tipo é exibido de uma cor no gráfico. Por ser um *ranking*, foi feito um gráfico de barras na horizontal. Esta visualização apresenta a quantidade (absoluta e percentual) de alianças e inimizades, de acordo com a seleção feita pelo usuário.

Como dito anteriormente, o usuário pode selecionar como prefere visualizar os valores: valores absolutos ou valores percentuais. Dessa forma, caso o usuário marque a opção “*Normalize data*”, a visualização como um todo é alterada para que os valores sejam exibidos em valores percentuais. Inicialmente a visualização é exibida utilizando barras empilhadas e em valores percentuais, porém, além da normalização, o usuário pode alterar a exibição para barras separadas, o que facilita a comparação dos valores números de alianças e inimizades de cada país.

Outros elementos de interação incluem possibilitar que pesquise um país pelo nome, filtro da informação a ser exibida (somente aliados, somente inimigos ou ambos os relacionamentos) e selecione a quantidade de países que devem ser exibidos no *ranking*. Além disso, são exibidas *tooltips* quando o usuário passa o cursor sobre as barras.

Em relação às decisões tomadas durante a implementação, foram utilizadas barras empilhadas como configuração inicial para indicar o todo, já que a visualização começa com a exibição dos valores percentuais.

Em relação às cores utilizadas, apesar de uma cor ser mais forte que a outra, elas foram escolhidas utilizando o Color Brewer combinado com o Color Oracle, a fim de possibilitar que pessoas com diferentes tipos de daltonismo (deuteranopia, protanopia e tritanopia) consigam diferenciá-las, além de possibilitar a correta leitura da informação também em escala de cinza.

A Figura 1 mostra um print da visualização desenvolvida. Ela apresenta a configuração inicial da visualização. Nessa configuração padrão não é possível identificar que se trata de um *ranking* e, portanto, o título da visualização foi escolhido de forma a representar melhor o que é exibido nesse primeiro momento.

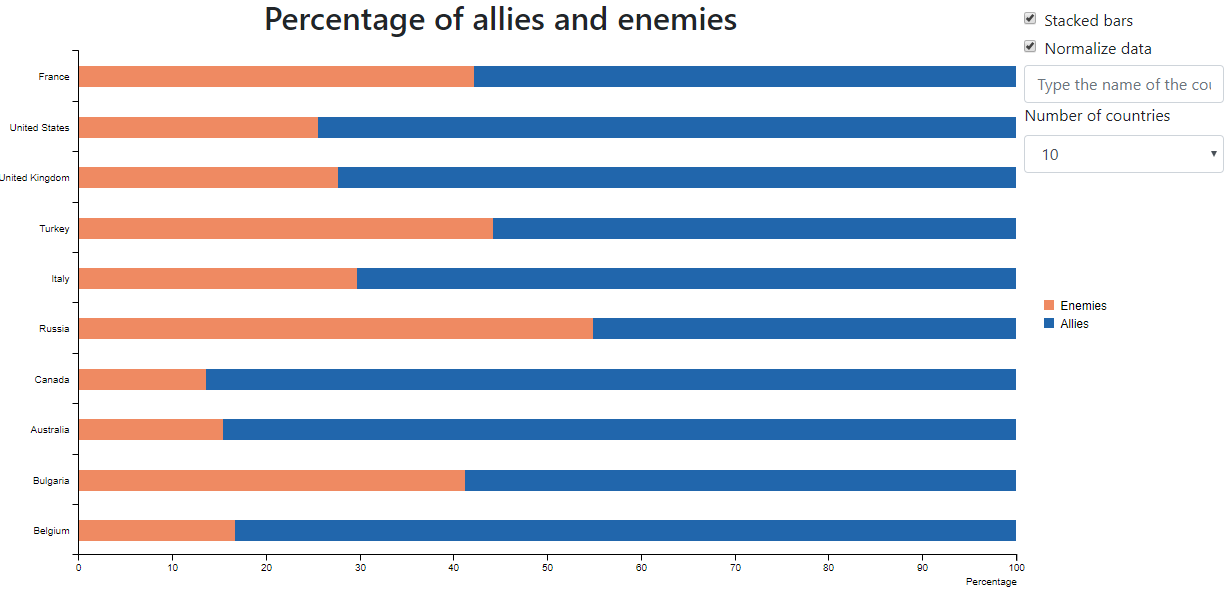


Figura 1 – Pocentagem de aliados e inimigos

* Visualização 2: *Ranking* dos conflitos mais longos

(https://dataviz-tp2.herokuapp.com/visualizacao2)

A Visualização 2 apresenta um ranking dos conflitos mais duradouros desde 1501. Por se tratar de um *ranking*, foi escolhido um gráfico de barras horizontais. Os *grids* não estão presentes porque não são estritamente necessários para uma análise qualitativa e, caso o usuário queira saber mais precisamente quais os valores, basta passar o cursor sobre as barras que são exibidas *tooltips* com os valores absolutos das barras.

Outra possível interação nessa visualização é a alteração da quantidade de conflitos

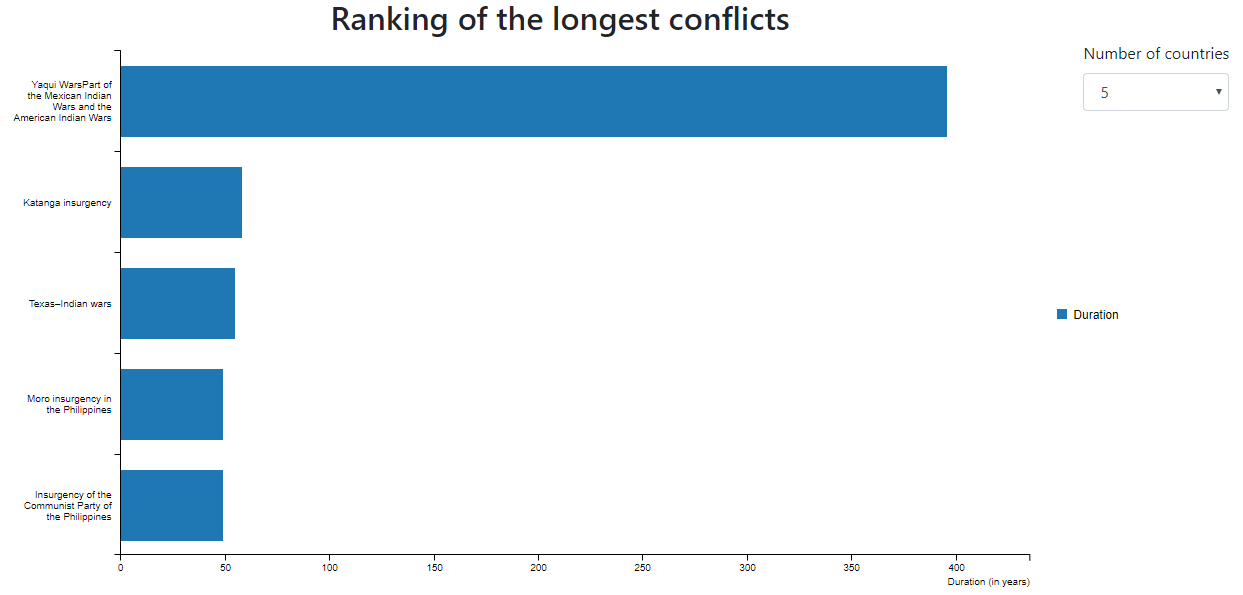


Figura 2 – Diferença de votos dos candidatos entre os dois turnos

* Visualização 2: *Ranking* dos conflitos mais longos

(https://dataviz-tp2.herokuapp.com/visualizacao2)

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.

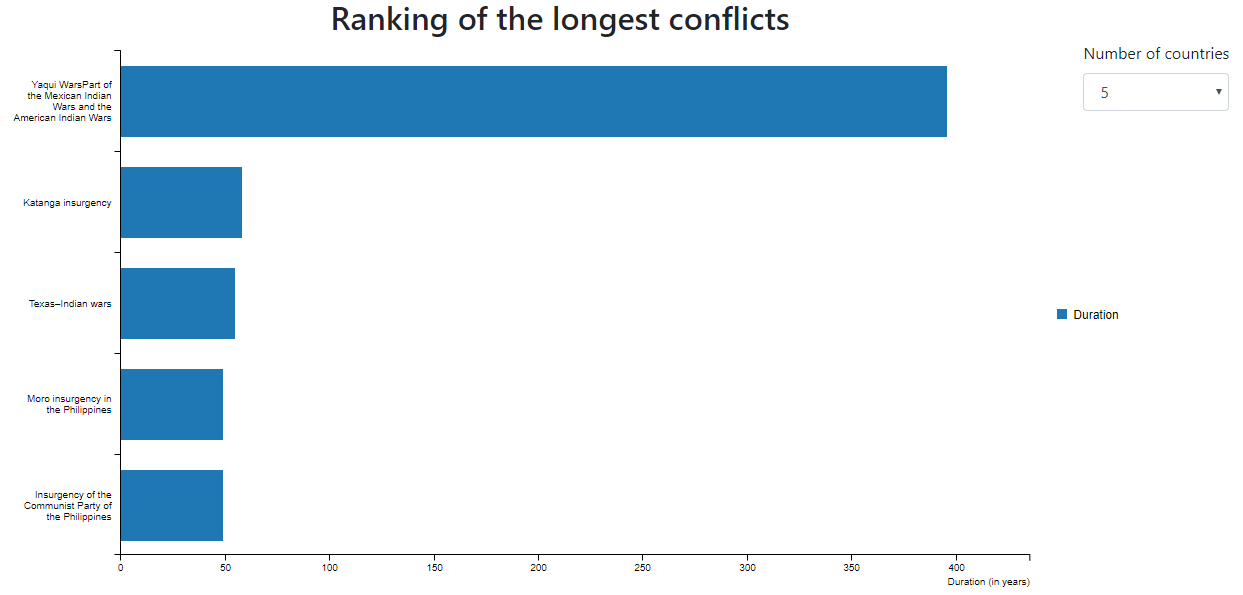


Figura 2 – Diferença de votos dos candidatos entre os dois turnos

**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.