

# Visualização Computacional

## Análise de dados utilizando D3

---

Edesio Alcobaça, 8578872

10 de Novembro de 2017

## 1 CONSIDERAÇÕES INICIAIS

Todos os códigos desenvolvidos neste trabalho estão disponíveis no *github*<sup>1</sup> e *online* no *github.io*<sup>2</sup>. Para executar o código *offline* consulte o '*README.md*' e '*make help*'.

## 2 PRÉ-PROCESSAMENTO

Os dados utilizados neste estudo foram produzidos pela "*Times World University Ranking*" ao ranquear as melhores universidades do mundo e podem ser encontrados aqui<sup>3</sup>.

O conjunto de dados é composto por 800 universidades (linhas) e 13 atributos (colunas) e há valores faltantes. Os atributos estão descritos na Tabela 2.1.

Tabela 2.1: Atributos do conjunto de dados.

ID	Descrição
V1	World Rank
V2	University
V3	Country
V4	Teaching Rating
V5	Inter Outlook Rating
V6	Research Rating
V7	Citations Rating
V8	Industry Income Rating
V9	Total Score
V10	Num Students
V11	Student Staff Ratio
V12	% Female Students
V13	% Inter Students

---

<sup>1</sup><https://github.com/EAlcobaca/Visualization>

<sup>2</sup><https://ealcobaca.github.io/Visualization/>

<sup>3</sup>[https://github.com/ttm/dataVisualization/raw/master/data/Times%20World%20University%20Rankings%20\(2016\).xlsx](https://github.com/ttm/dataVisualization/raw/master/data/Times%20World%20University%20Rankings%20(2016).xlsx)

## 2.1 ETAPA 1

Nesta primeira etapa de pré-processamento verificamos que: (i) o atributo V1 possui ranqueamento até o *rank* 200 e depois usa faixas (e.g. 201-250), (ii) o atributo V2 é único para cada amostra, (iii) o atributo V3 é categórico, porém não numérico. Os dados foram tratados como segue: (i) no atributo V1 as faixas foram substituídas pela suas respectivas médias (e.g. a faixa 600-800 por 700), (ii) no atributo V2 os nomes das universidades foram substituídos por números inteiros de 1 a 800, (iii) o atributo V3 foi substituído por categoria numérica.

Por fim, foi feita a imputação dos valores faltantes utilizando a média de suas respectivas colunas.

## 2.2 ETAPA 2

Não contente com os resultados iniciais, optou-se por remover os dois primeiros atributos V2. Selecionamos apenas as 200 melhores universidades, uma vez que a partir deste valor elas são ranqueadas em faixas extensas. Desta maneira obteve-se resultados mais interessantes. Além disso, utilizamos V3 apenas no gráfico de barras.

Os códigos utilizados na etapa 1 e 2 estão disponíveis em "*preprocessing.py*, *preprocessing2.py* e *dataSumm.R*"

## 3 COORDENADAS PARALELAS

O Gráfico da figura 3.1 é de coordenadas paralelas. Cor quente indica ranqueamento alto, cor fria indica ranqueamento baixo. As cores utilizadas levam em consideração pessoas com daltonismo. É possível observar que nos atributos *Teaching\_Rating* e *Research\_Rating* as cores estão bem visíveis, indicando que há boa separação e possivelmente correlação entre estes atributos e o ranque.

Vamos visualizar mais de perto e entender como um sub-conjunto de linhas se comportam. No gráfico da figura 3.2 e 3.3 selecionamos as universidades com ranque entre [1-40] e [160-200] respectivamente. É possível observar que para a figura 3.2, os 40 melhores ranqueados tem alta *Teaching\_Rating* e *Citation\_Rating* e *Research\_Rating* enquanto que para a figura 3.3 o oposto é observado. Isto indica que estes três atributos são muito importantes para obter bom ranqueamento. Em contrapartida *Student\_Staff\_Ratio*, *Female\_Students* e *Inter\_Students* estão muito misturados, indicando que são atributos mais fracos e menos determinantes para ranquear as universidades.

Figura 3.1: Gráfico de coordenadas paralelas. Os eixos são os atributos (colunas) da tabela e as cores das linhas variam de acordo com o ranque. As cores utilizadas levam em conta pessoas com daltonismo.

Select the set of lines that you want to study

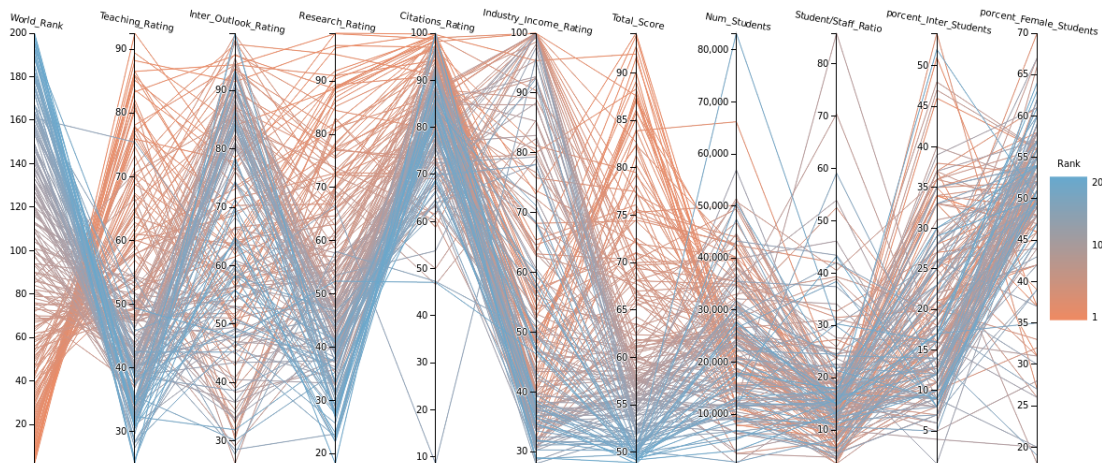


Figura 3.2: Gráfico de coordenadas paralelas. Os eixos são os atributos (colunas) da tabela e as cores das linhas variam de acordo com o ranque. Selecionando os 40 melhores

Select the set of lines that you want to study

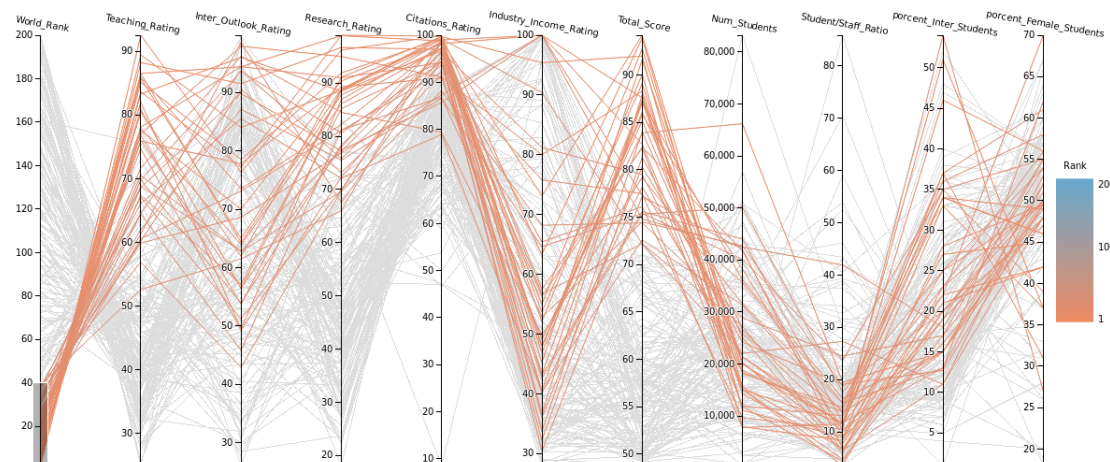
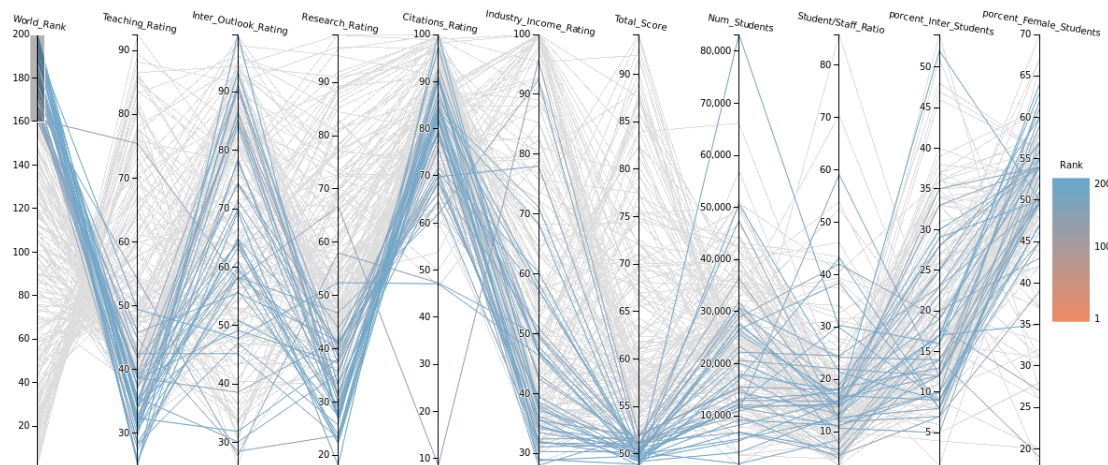


Figura 3.3: Gráfico de coordenadas paralelas. Os eixos são os atributos (colunas) da tabela e as cores das linhas variam de acordo com o ranque. Seleccionando universidades com ranque entre 160-200.

Select the set of lines that you want to study



## 4 SCATTER PLOT

Os scatter plots corroboram com os resultados obtidos anteriormente e mais, podemos visualizar o tipo de relação que as variáveis tem com o ranque. Os gráficos das figuras 4.1 e 4.2 indicam claramente que a medida que chegamos mais perto do ranque 1, maior é a taxa de Teaching e Research. O mesmo não é observado para o Citation\_Rating, pois este se mantém entre [90,100] para os 40 melhores e vai diminuindo lentamente. Ainda, é possível notar que há um *outlier* próximo do ranque 140, com Citation\_Rating muito baixa (aproximadamente 10).

Por fim, no gráfico da figura 4.4 conseguimos notar que não existe nenhuma relação entre Inter\_Outlook\_Rating e o ranqueamento, uma vez que o gráfico aparenta estar uniformemente distribuído.

Figura 4.1: Scatter plot entre os atributos World\_Rank e Teaching\_Rating

Select the attributes that you want to study

x-axis: World\_Rank y-axis: Teaching\_Rating Change

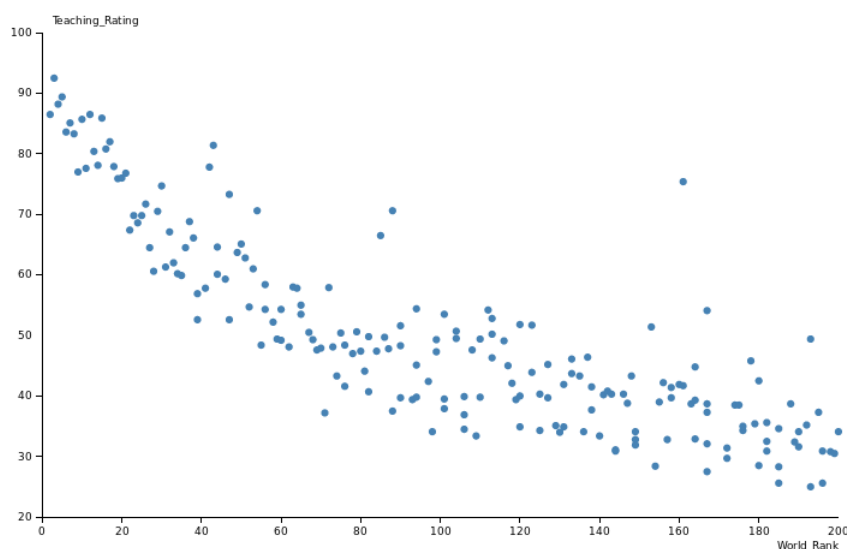


Figura 4.2: Scatter plot entre os atributos World\_Rank e Research\_Rating

Select the attributes that you want to study

x-axis: World\_Rank y-axis: Research\_Rating Change

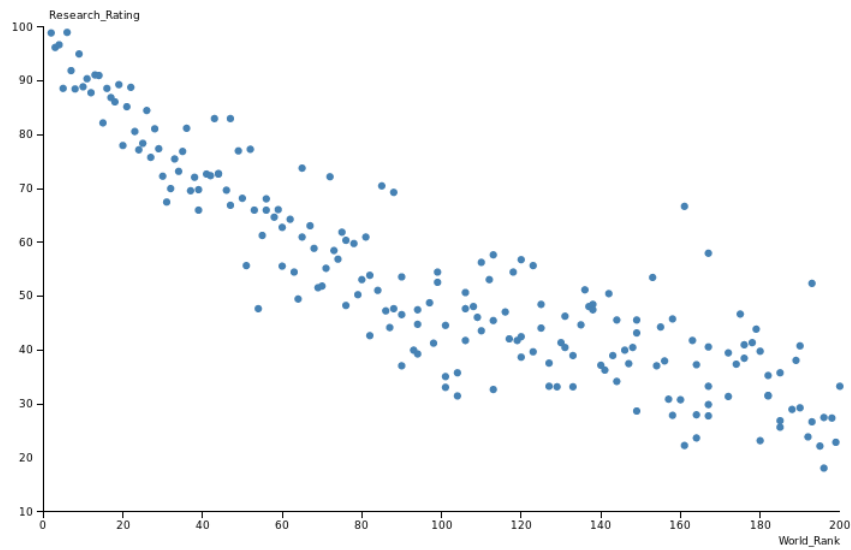


Figura 4.3: Scatter plot entre os atributos World\_Rank e Citations\_Rating

Select the attributes that you want to study

x-axis: World\_Rank y-axis: Citations\_Rating Change

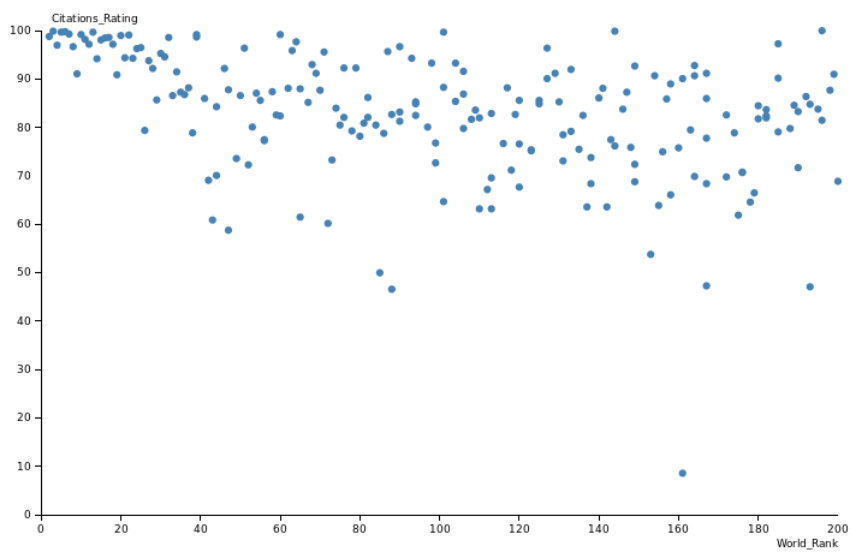
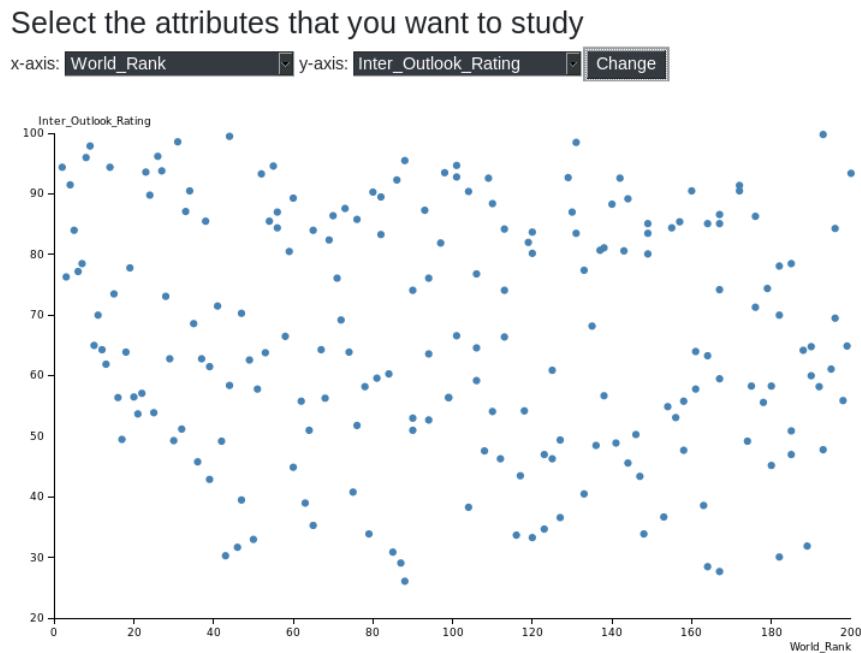


Figura 4.4: Scatter plot entre os atributos World\_Rank e Inter\_Outlook\_Rating



## 5 GRÁFICO DE BARRAS

Analisando o gráfico de barras 5.1 é possível observar que os Estados Unidos tem mais que um quarto das universidades entre as 200 melhores, e quase o dobro do segundo colocado, o Reino Unido. Os 6 primeiros países com mais universidades entre as 200 melhores (Estados Unidos, Reino Unido, Alemanha, Holanda, Austrália e Canadá) são países considerados desenvolvidos (primeiro mundo). A África do Sul, considerado país em desenvolvimento, tal como o Brasil, aparece nesta lista. Fica evidente que os países de primeiro mundo possuem mais universidades entre as 200 melhores.

Figura 5.1: Gráfico de barras representando a quantidade de universidades por país de origem.  
Number of occurrence of the country among the 200 best ranked

