

## PI IV - Extração e Visualização de Dados – 2º Semestre de 2019













### Introdução

Até o presente momento trabalhamos com dados numéricos. A vantagem de trabalhar com este tipo de informação é que ela pode ser facilmente apresentada através de eixos que representam cada uma das grandezas utilizadas.

Porém, existem diversos tipos de dados não numéricos e muitas vezes precisamos mostrar como estes dados estão relacionados. Por exemplo considere um sistema de informação que indique a distância entre duas cidades que foram sedes da copa do mundo de 2014 e como esta informação pode ser apresentada:

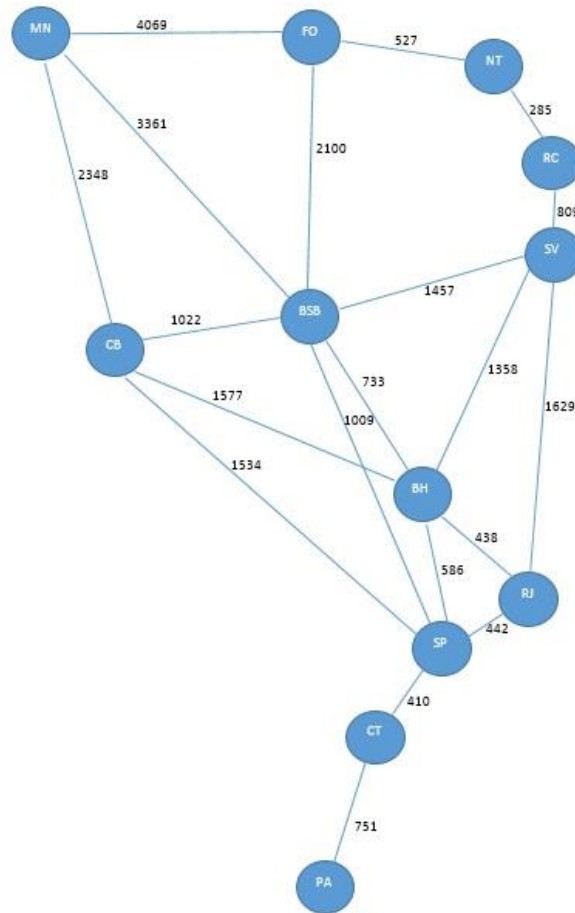
#### Distâncias rodoviárias (em Km) entre as cidades-sede

Road distances (in km) between host cities  
Distancias por carretera (en Km) entre las ciudades sede

|                | Belo Horizonte   | Brasília  | Cuiabá  | Curitiba  | Fortaleza   | Manaus  | Natal   | Porto Alegre   | Recife  | Rio de Janeiro  | Salvador  | São Paulo   |
|----------------|--|---|---|---|---|---|---|--|---|---|---|---|
| Belo Horizonte |  | 733   | 1577  | 995   | 2349  | 3923  | 2292  | 1731   | 2028  | 438   | 1358  | 586   |
| Brasília       | 733  |  | 1022  | 1390  | 2100  | 3361  | 2420  | 2123   | 2130  | 1162  | 1457  | 1009  |
| Cuiabá         | 1577   | 1022  |  | 1723  | 3205  | 2348  | 3576  | 2345   | 3148  | 1930  | 2478  | 1534  |
| Curitiba       | 995  | 1390  | 1723  |  | 3379  | 4072  | 3322  | 751  | 3058  | 840   | 2388  | 410   |
| Fortaleza      | 2349   | 2100  | 3205  | 3379  |  | 4069  | 527   | 4106   | 794   | 2614  | 1228  | 2961  |
| Manaus         | 3923   | 3361  | 2348  | 4072  | 4069  |  | 4546  | 4691   | 4508  | 4277  | 4529  | 3880  |
| Natal          | 2292   | 2420  | 3576  | 3322  | 527   | 4546  |  | 4049   | 285   | 2557  | 1093  | 2904  |
| Porto Alegre   | 1731   | 2123  | 2345  | 751   | 4106  | 4691  | 4049  |  | 3793  | 1575  | 3123  | 1145  |
| Recife         | 2028   | 2130  | 3148  | 3058  | 794   | 4508  | 285   | 3793   |  | 2297  | 809   | 2644  |
| Rio de Janeiro | 438  | 1162  | 1930  | 840   | 2614  | 4277  | 2557  | 1575   | 2297  |  | 1629  | 442   |
| Salvador       | 1358   | 1457  | 2478  | 2388  | 1228  | 4529  | 1093  | 3123   | 809   | 1629  |  | 1971  |
| São Paulo      | 586  | 1009  | 1534  | 410   | 2961  | 3880  | 2904  | 1145   | 2644  | 442   | 1971  |  |

Fonte: DNIT - <http://www.rotasdascidades.com.br/>

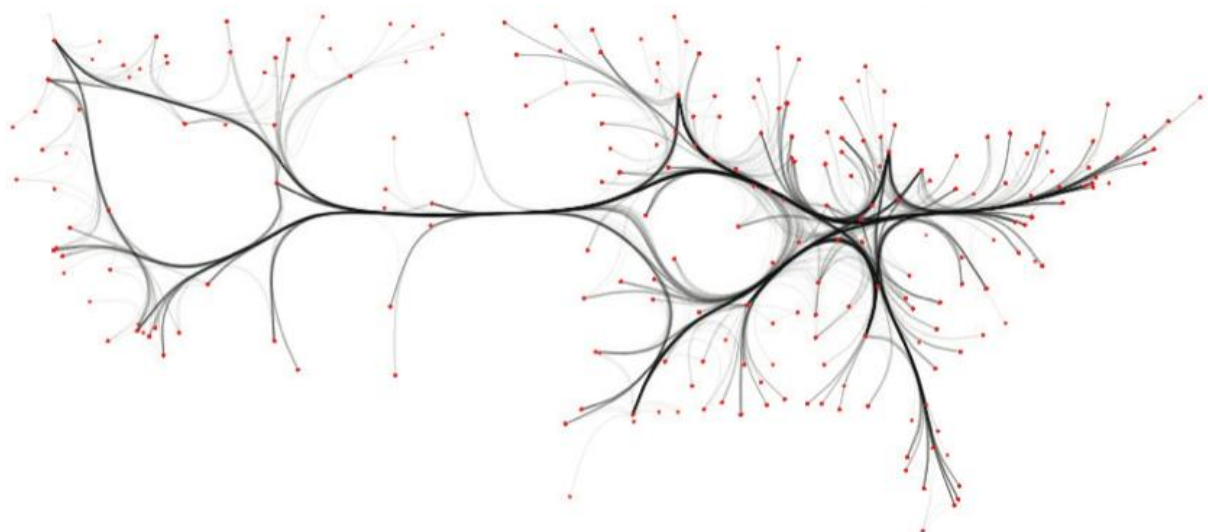
Uma outra forma de apresentar a mesma informação está apresentada na figura seguinte. Comente as vantagens e desvantagens de cada uma destas formas de representar informações.



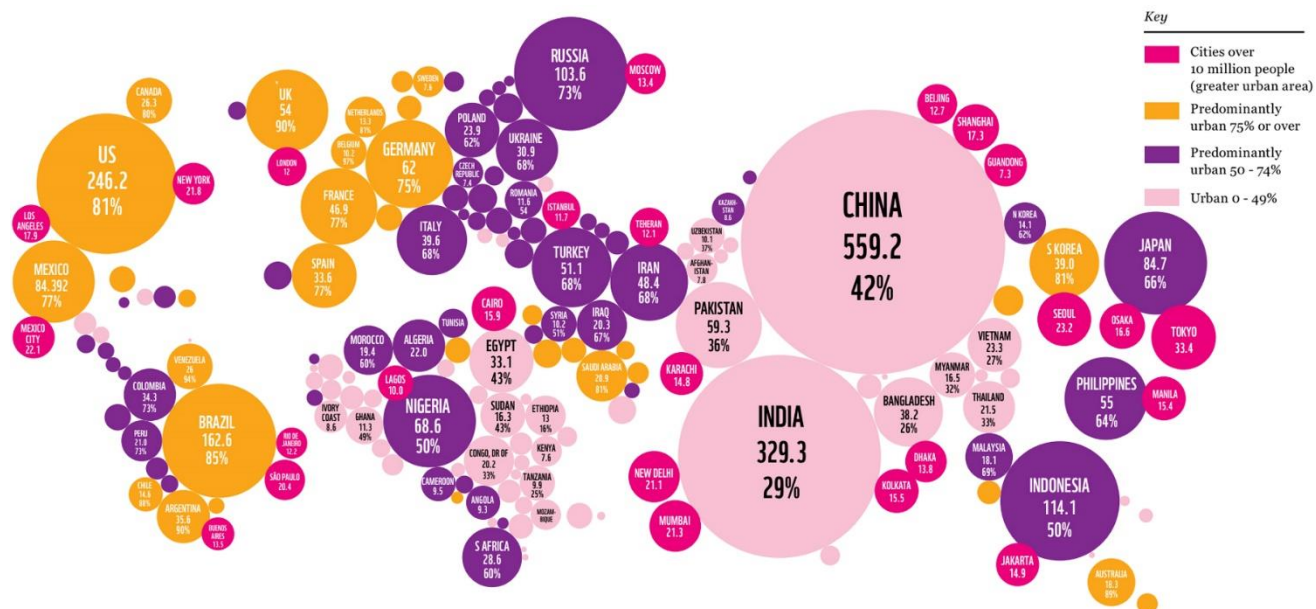
### Visualização de dados não numéricos

Além dos exemplos apresentados acima existem ainda outros tipos de dados não numéricos e formas distintas de representá-los. Considere por exemplo o caso de linhas aéreas nos Estados Unidos das principais empresas aéreas americanas e as cidades atendidas por elas, a figura abaixo representa este caso com 235 nós e 2099 arestas. As linhas mais espessas representam as rotas mais frequentes.

Existe alguma outra forma, que seja mais intuitiva, de se representar este tipo de informação?



Um outro exemplo de representação visual de dados é apresentado abaixo para a população do planeta num determinado ano. Que tipo de conceito está sendo utilizado neste tipo de representação?



No link a seguir você encontrará outras formas de se representar esta mesma informação e ainda adicionar informação extra:

[https://archian.files.wordpress.com/2013/09/urban\\_world.jpg](https://archian.files.wordpress.com/2013/09/urban_world.jpg)

[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_and\\_dependencies\\_by\\_population](https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population)

<http://www.ctbuh.org/Home/FactsData/TallBuildingsandPopulation/tabid/2775/language/en-US/Default.aspx>

## Atividade

Nesta atividade você deve trabalhar com dados do IBGE (<https://www.ibge.gov.br/estatisticas-novoportal/sociais/educacao/9662-censo-demografico-2010.html?edicao=9753>) e selecione 5 estados diferentes. Considere entre os dados disponíveis as tabelas 3 e 4 que apresentam o número de alunos matriculados por idade e nível de escolaridade da rede pública e da rede particular. Considere 3 cidades ou micro-regiões de cada estado e apresente os dados de forma que seja possível fazer uma análise da situação. Mostre os dados de duas formas diferentes e comente das vantagens e desvantagens de cada uma delas.

Fique a vontade para apresentar algum outro tipo de dados não numérico diferente do apresentado acima mas que tenha pelo menos 5 tipos de informação diferente.