Iniciando a brincadeira com R

1. **Assista** [**esse vídeo**](https://www.ted.com/playlists/56/making_sense_of_too_much_data) **com atenção. Perceba que a grande pergunta que Hans Rosling tenta responder é: temos dois mundos separados; o mundo ocidental norte rico e o resto do mundo pobre? Talvez você precise pausar o vídeo para entender, assistir mais de uma vez, etc.** 
   1. **Trace um esboço do caminho que ele percorreu analisando os dados para responder essas perguntas.**
   2. **Anote os termos ou ideias que você não entende ainda.**

// TODO

**2. Qual o panorama geral dos profissionais de TI no Brasil sem ainda se preocupar com seus salários, apenas com sua localização, função, expertise...?**

A maioria dos profissionais de TI do Brasil estão na iniciativa privada e trabalhando na região Sudeste do Brasil, principalmente em Belo Horizonte.

Metade dos profissionais atuantes não possuem uma Pós Graduação ou uma Certificação. Dos que possuem

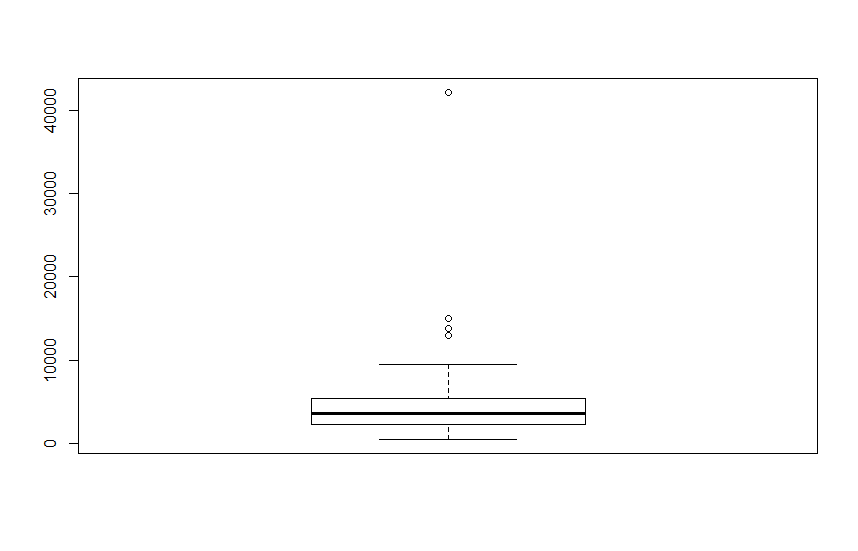
Porém podemos perceber que possuir

1. **Os salários estão bem distribuídos no intervalo que vai do menor salário para o maior salário? Ou será que os salários são muito concentrados em um intervalo pequeno de valores? Será que existem salários atípicos (*outliers*)?**

Como podemos observar no Box Plot dos salários, a maioria está concentrado em um intervalo pequeno entre 2281 e 4233. Também podemos observar que existem um salário atípicos de 42120.

**Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** |
| 500 | 2281 | 3583 | 4233 | 5332 | 42120 |

**Box Plot - Salários**

1. **A distribuição que você encontrou é a que você esperava sem olhar os dados? Se não é, você tem alguma suspeita da razão para a diferença?**

Não, já que eu esperava um intervalo de salário mais alto. O fato de vários profissionais não terem formação acadêmica deve ter influenciado negativamente no valor dos salários.

1. **É interessante ver a média (*mean*) salarial para cada região. Como resultado, apresente uma tabela com o nome da região e a sua média salarial ordenada da região com a maior média salarial para a menor. Se você tivesse usado a mediana (*median*) em vez da média como métrica de interesse, o resultado seria o mesmo? Analise a situação e justifique. Aproveite e salve esta tabela em um arquivo do tipo csv.** Dica: use *tapply, write.csv* e *sort*.

|  |  |
| --- | --- |
| **Região** | **Média Salarial** |
| Centro-Oeste | 5620.474 |
| Nordeste | 4565.706 |
| Sudeste | 3962.227 |
| Sul | 3657.538 |
| Norte | 2552.500 |

Podemos ver abaixo como ficaria a tabela se tivesse usado a mediana.

|  |  |
| --- | --- |
| **Região** | **Mediana Salarial** |
| Centro-Oeste | 5000.00 |
| Sudeste | 3582.95 |
| Sul | 3500.00 |
| Nordeste | 2900.00 |
| Norte | 2552.500 |

Podemos observar o Nordeste “caiu” para quarto lugar. Isso se deve ao fato dos salários dos outros estados estarem mais agrupados e não terem grandes outliers como o caso do profissional que tem um salário de 42mil no Nordeste.

1. **Se um profissional de TI deseja ter um salário melhor, qual a(s) dica(s) para ele?** Dica: Use sua criatividade para brincar com os dados e descobrir informações relevantes/interessantes.

// TODO