

Iniciando a brincadeira com R

Objetivos

- Ficar íntimo dos comandos básicos do R;
 - Praticar os comandos que servem para ler e escrever uma tabela de dados no R;
 - Criar o hábito de conhecer os dados que tem em mãos para análise e de entender os tipos de dados;
 - Criar o hábito de pensar em perguntas interessantes que podem ser investigadas com os dados que tem em mãos, ie. ter a prática de planejar a análise dos dados para extrair valor deles;
 - Despertar para diferenças entre média/mediana e o conceito de outliers;
 - Integrar na sua rotina de análise comandos básicos como mean, median, summary, sort, max;
 - Dominar comandos mais avançados que permitam trabalhar com dados em formato de tabelas (data frames), agrupar dados por um atributo específico e sumarizar resultados com base em um atributo específico.
-

O resultado

O que esperamos como resultado deste problema é um relatório respondendo e comentando os pontos levantados abaixo. Lembre que o importante é comunicar suas conclusões e *como você chegou nelas* a partir dos dados. Procure ser preciso nas suas afirmações, e use os dados para embasar as afirmações.

Questões práticas:

- Seu resultado deverá estar em um repositório no github
 - Use o [R Markdown](#) para fazer o relatório. Simplificará a sua vida.
 - Você pode publicar o seu relatório usando o [Rpubs](#).
-

Antes de começar

1. Crie um repositório no github para esta atividade e [coloque o link aqui](#)
 2. Instale o R em sua máquina
 3. Instale o R Studio
 4. Leia dicas básicas [aqui](#).
-

O problema: entender os salários de profissionais de TI no Brasil

Estamos tentando entender um pouco a distribuição dos salários pagos a profissionais de TI (tecnologia da informação). Algumas informações foram coletadas sobre esses profissionais e estão [aqui](#). Abra o arquivo fora do R, e entenda o formato deles. Em seguida, os carregue no R.

Dica para passar esses dados para o R em formato de tabela: use `read.csv`.

Use essas informações para gerar ~~em um google doc~~ um relatório de análise que responde as seguintes questões:

1. Assista [esse vídeo](#) 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.
 - a. Trace um esboço do caminho que ele percorreu analisando os dados para responder essas perguntas.

Dica: Anote os termos ou ideias que você não entende ainda. Saber o que você ainda não sabe é bem importante.

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. Dica 1: Quando temos um conjunto de dados para analisar é muito importante ter ciência de todos os campos de informação que temos disponíveis. Com o comando `head`, por exemplo, você consegue ter uma visão geral do cabeçalho do seu conjunto de dados e dos tipos de dados que estão envolvidos. Ao ter conhecimento do seu conjunto de dados e de que tipo de informação cada campo traz, fica mais fácil ter boas ideias para analisar.
 - b. Faça o exercício de fazer perguntas aos dados!
 - c. O comando `table` pode ajudar nessa tarefa.
3. 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*)?
 - a. Dicas: `summary`, `boxplot`, `hist`, `density`.
4. 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?
5. 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`.

Dica 2: um dos pacotes mais úteis para seu dia-a-dia com dados é o `plyr`: <http://plyr.had.co.nz/>. Se sua solução já funciona e você quer deixá-la ótima, ou aprender algo pra melhorar seu futuro, dá uma olhada.
6. Se um profissional de TI deseja ter um salario melhor, qual a(s) dica(s) para ele? *Dica: Use sua criatividade para brincar com os dados e descobrir informações relevantes/interessantes.*
7. *[Opcional] Use o pacote [ggplot2](#) do R para ilustrar com gráficos suas conclusões.*

8. *[Opcional 2] Use [rCharts](#) para ilustrar com um gráfico interativo alguma de suas conclusões.*