

Tarefa da Semana 1

Objetivo: Instalar e utilizar o R, Rstudio e o Rmarkdown. Explorar base de dados.

Vamos usar a base de dados disponibilizada no Kaggle chamada “Mental health in tech survey”

<https://www.kaggle.com/osmi/mental-health-in-tech-survey>

Para essa base de dados, utilizando as funções descritas no R de exemplo e outras funções R que você julgar importantes (se você utilizar novos pacotes não utilizados no exemplo, inclua a instalação e uso desse novo pacote no seu código R, pois ao executá-lo eu gostaria que o mesmo funcionasse)

Para essa tarefa descreva:

- Quantos atributos e quantas instâncias
- Quais os tipos de dados dos atributos, intervalo de valores
 - para os atributos numéricos, mostre mínimo, máximo, média, desvio padrão, mediana.
 - Há atributos categóricos? Quais? Quantas instâncias há para cada valor desses atributos categóricos?
- Faça um histograma do atributo Age e mostre com diferentes números de faixas de valores (bins=5, bins=10, bins=20)
- Faça também um boxplot deste atributo. A partir desses 2 gráficos você consegue identificar outliers? Se sim, quais valores dos mesmos?
- Se você encontrou outliers, exclua essas instâncias (ou substitua os valores pela média de idade) e refaça o histograma. Ficou melhor de analisar as faixas de idade dos pacientes analisados por essa base de dados?
- Faça gráficos de matrizes de dados relacionando os atributos. Busque relações entre os atributos que julgar interessantes para serem visualizadas. Por exemplo: work interfere e mental_health_consequence. Você pode interpretar algo dessa relação. Faça outras análises desse tipo (mínimo mais 1, com interpretação).
- Faça pelo menos mais um gráfico que quiser que não tenha sido discutido no documento de exemplo. Dica: olhe a documentação do ggplot, há muitos e muitos gráficos e visualizações de dados interessantes. Pode ser um gráfico simples, mas que a interpretação seja interessante.
- **Comece a perceber que, sem aplicar qualquer algoritmo de mineração de dados, com bons gráficos (e boas técnicas de visualização de dados) você já pode começar a extrair conhecimento dos dados.**
- Ao finalizar sua tarefa, envie no AVA o arquivo RMD e o PDF resultante do Rmarkdown. Esse PDF deve exibir os comandos R executados assim como a saída de todos eles.