**INTRODUÇÃO**

O projeto abaixo consiste na análise do dataset “Iris.csv”. É uma base de dados que contém 150 observações com as dimensões das pétalas e sépalas de 3 espécies diferentes de flores.

O intuito desse projeto foi de desenvolver habilidades e treinar as técnicas mostradas no decorrer do curso de Data Science. Serão utilizadas técnicas como analise exploratória, Data Wrangling, Análise fatorial e Predição a partir de uma Logística Multinominal.

**DESENVOLVIMENTO**

Abaixo vemos o dataset que foi utilizado no desenvolvimento desse projeto.



*Dataset “Iris.csv”. Inicialmente estruturado.*

*Descrição das Variáveis*

**Id** = Identificação da Observação

**SepalLenghtcm** = Comprimento da sépala em cm

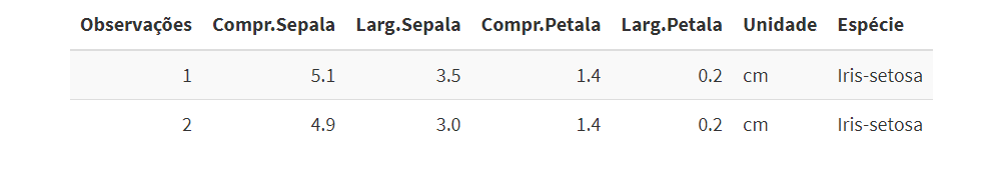
**SepalWidthcm** = Largura da sépala em cm

**PetalLenghtcm** = Comprimento da pétala em cm

**PetalWidthcm** = Largura da pétala em cm

**Species** = Espécie

O dataset vem com uma estrutura padrão, que torna mais fácil de ser manipulado, as primeiras alterações para facilitar o desenvolvimento é mudar os nomes das variáveis. Tambem foi adicionada a variável “Unidade” que indica a unidade dos valores numéricos das observações.



*Dataset “Iris.csv”. Após alteração dos nomes das variáveis.*

Ao iniciar a analise exploratória é notado que o dataset contém 150 observações de 3 espécies diferentes. O gráfico abaixo ilustra o histograma de cada espécie.



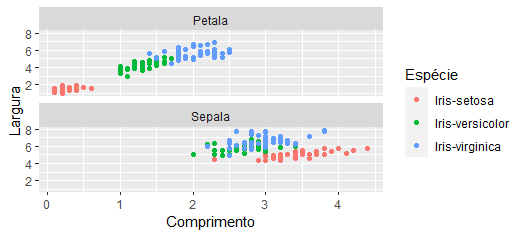
*Histograma Quantidade de Observações x Espécies*

Após algumas transformações no dataset, é construída uma estrutura que facilita a manipulação e visualização de informações que antes não estavam tão visíveis. Com um agrupamento e transformação das colunas dos valores dimensionais das pétalas e sépalas em linhas, é notado que há mais facilidade em comparar dados. Veja que temos duas observações ‘1’, uma para os valores da pétala e o outro para sépala.

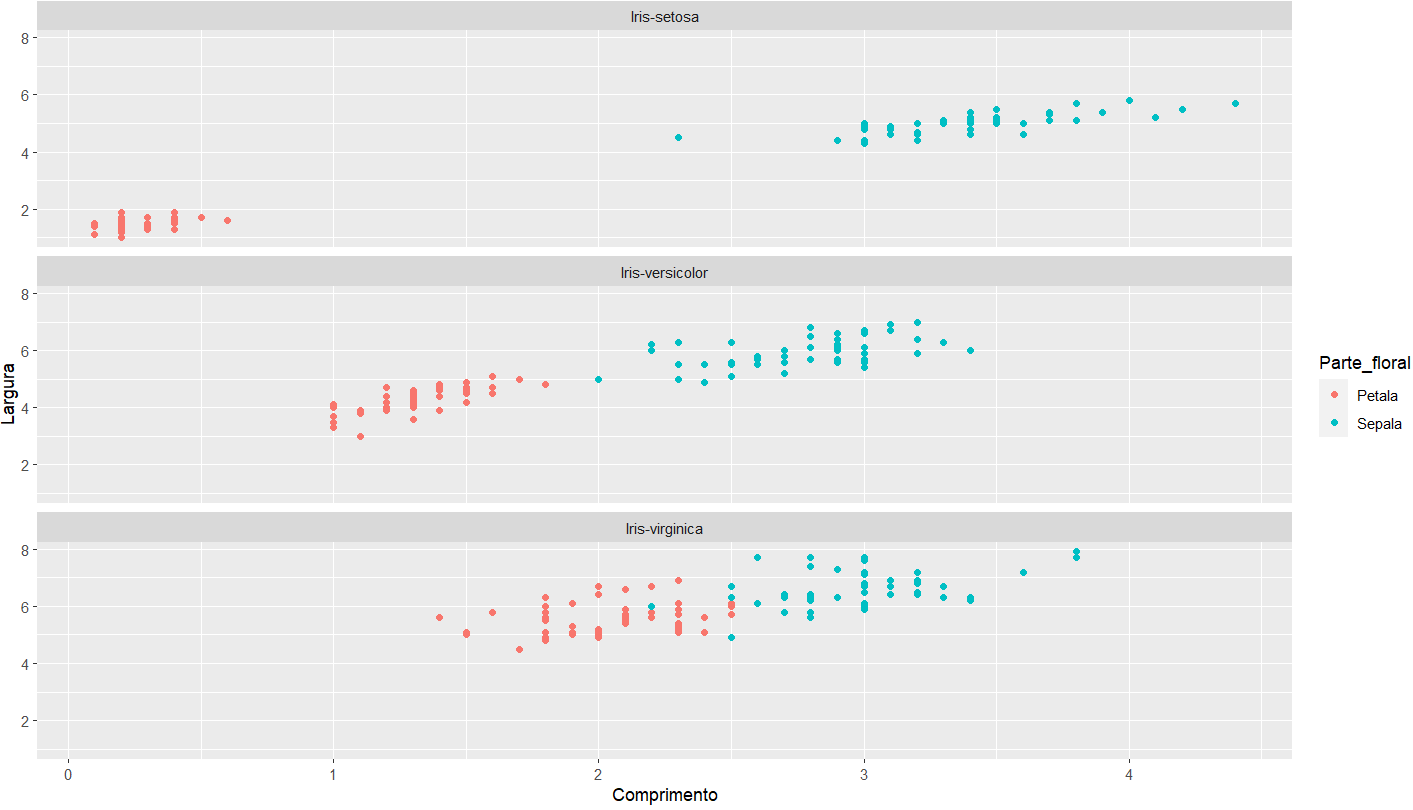


*Dataset após Wrangle.*

A partir daqui será feita as plotagens comparando as informações extraídas do dataset.

Abaixo vemos a relação comprimento x largura das partes florais de cada espécie.

*Gráficos de relação entre pétala e sépala.*

E a partir desses gráficos é possível extrair ainda mais informações e visualizar as diferenças das observações.

Abaixo é visto numericamente os valores médios para cada parte floral de cada espécie.



*Valores médios das partes florais de cada espécie.*

Analisando os gráficos e as médias, pode se afirmar que a as partes florais da Iris-Virginica são as maiores dentre as três.