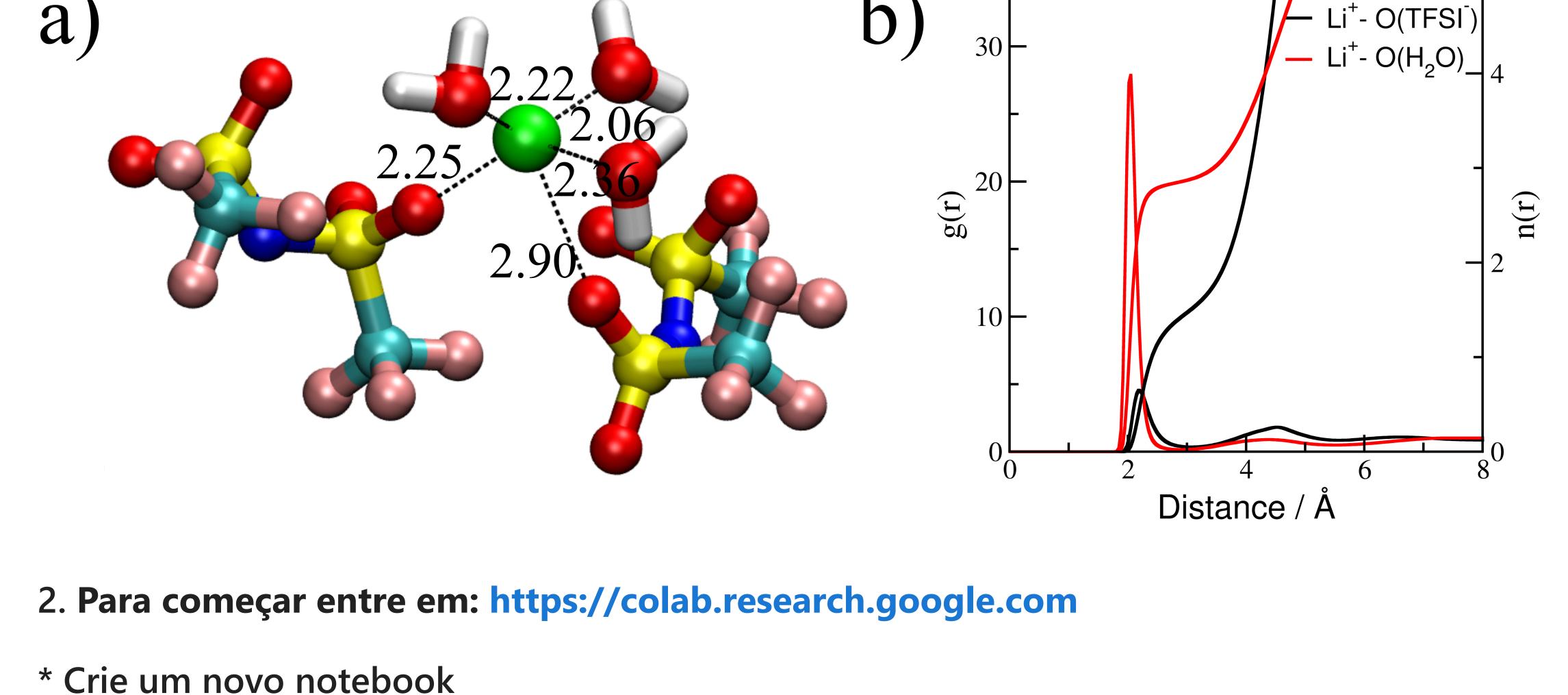
## USANDO PYTHON PARA ESTATÍSTICA DESCRITIVA BÁSICA COM VISUALIZAÇÃO DE **GRAFICOS**

- \* Este script realiza a estatística descritiva, plota um boxplot e um histograma de um conjunto de dados
- \* Criado por: Isabel Amaral da Silva
- \* Linguagem: Python 3
- 1. Para essa introdução, usaremos dados do número de coordenação dos cátions Li+obtidos de uma simulação computacional de uma solução para LiTFSI 20 molal



- 3. Importando os pacotes necessários:
- \* Pandas é um pacote gratuito para tratamento e organização de dados;
- \* Matplotlib é um pacote gratuito para plotar e visualizar gráficos;

In [ ]: #df=pd.read\_excel('/data-CoordNumber-LiTFSi.xlsx')

**Electrolyte** n(r)\_Li-Oanion n(r)\_Li-Owater

1.910891

1.943925

1.942308

1.942308

import pandas as pd

import matplotlib as plt

df = pd.read\_excel(url)

LiTFSI

LiTFSI

LiTFSI

LiTFSI

96

- 4. Importando seu conjunto de dados:
- \* Pode ser do GitHub ou do Google Drive ou por upload no próprio colab In [ ]: url = 'https://github.com/isabelamaral/Python-basic-stats-CQ-1sem2024/raw/main/data-CoordNumber-LiTFSi.xlsx'
  - \* pandas.read\_excel() ----> Função do pandas que lê um arquivo excel e transforma num conjunto de dados legível para o pandas ("Pandas DataFrame")

2.643564

2.598131

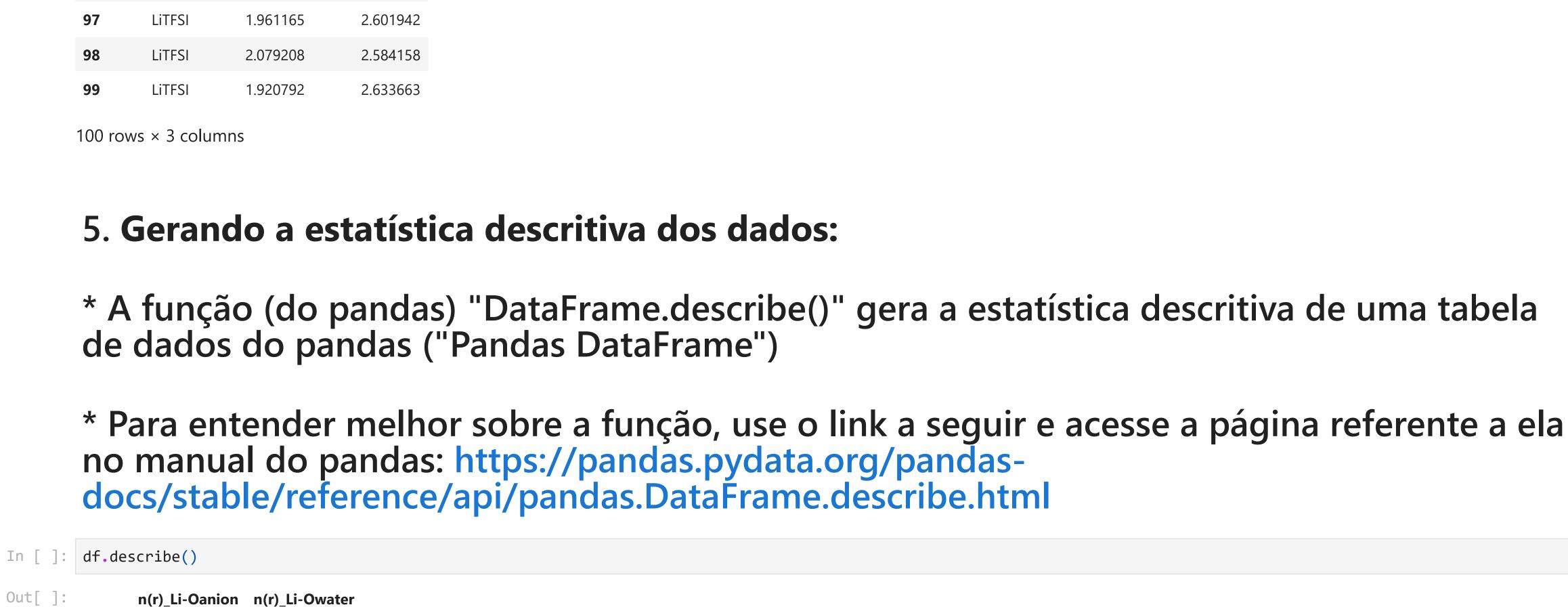
2.701923

2.615385

no manual do pandas: https://pandas.pydata.org/pandasdocs/stable/reference/api/pandas.read\_excel.html

\* Para entender melhor sobre a função, use o link a seguir e acesse a página referente a ela

LiTFSI 1.961165 2.631068 LiTFSI 1.864078 2.621359 1.971698 95 LiTFSI 2.584906



100.000000

1.889775

0.086542

1.898599

1.942712

2.079208

count

mean

**50%** 

**75%** 

max

2.0

1.8

<Axes: >

2.1

2.60

2.55

2.50

20

15

15

10

In [ ]:

df.boxplot(column='n(r)\_Li-Oanion')

100.000000

2.657014

0.065681

2.649281

2.702633

2.820755

um box plot como o abaixo:

Box Plot Simples

## 2.485714 1.843891 2.612406 **25**%

Maior valor Y

Quartil ou

Mediana

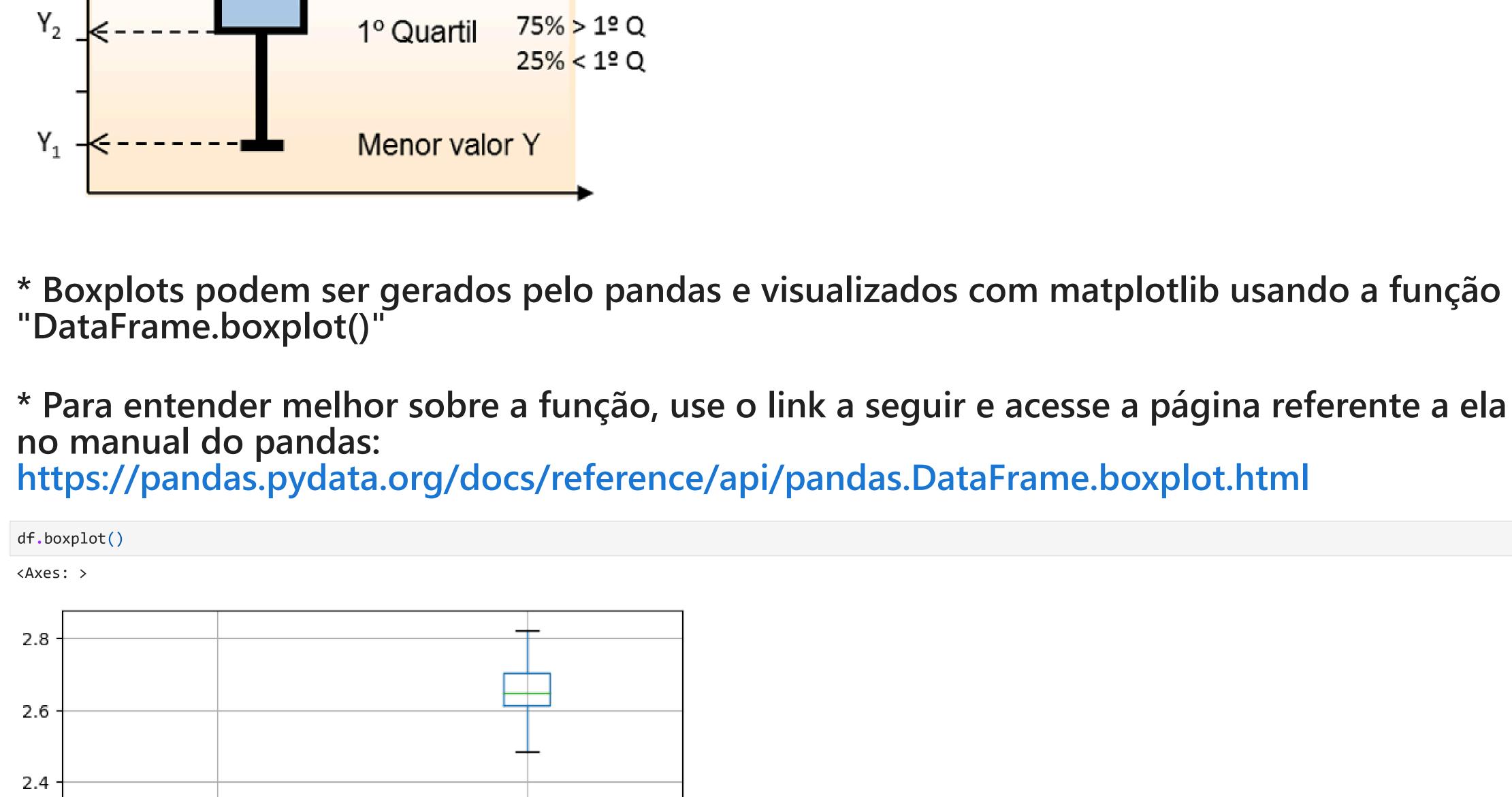
6. Visualizando os dados com gráficos

25% > 3º Q 3° Quartil 75% < 3º Q

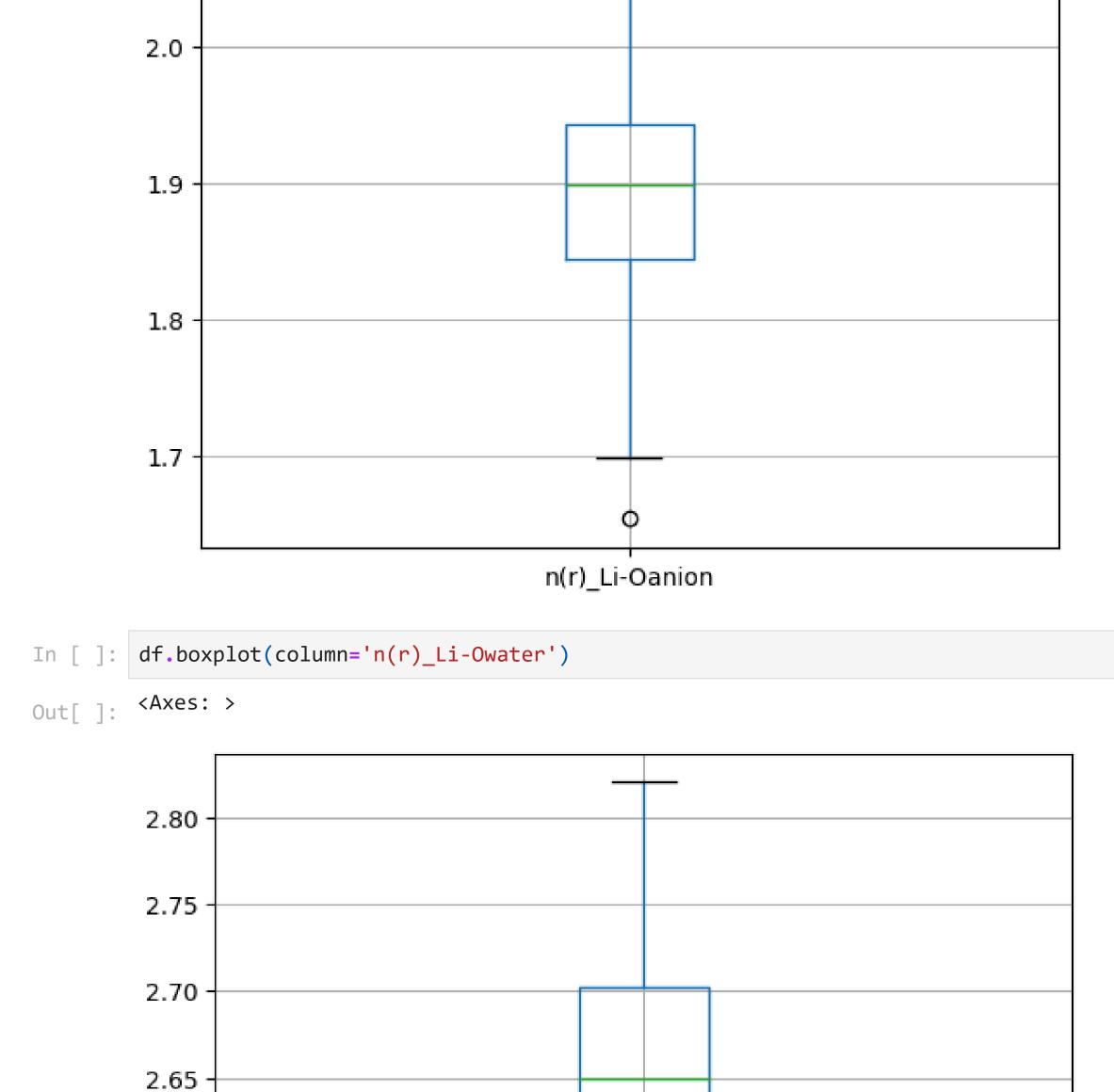
50% > 2º Q

50% < 2º Q

\* Uma maneira de visualizar os dados gerados pela função "DataFrame.describe()" é usar



- 2.2
- 1.6 n(r)\_Li-Oanion n(r)\_Li-Owater



\*Histogramas podem ser gerados pelo pandas e visualizados com matplotlib usando a função "DataFrame.hist()" \*Para entender melhor sobre a função, use o link a seguir e acesse a página referente a ela no manual do pandas: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.hist.html In [ ]: df.hist() n(r)\_Li-Oanion n(r)\_Li-Owater 25 25

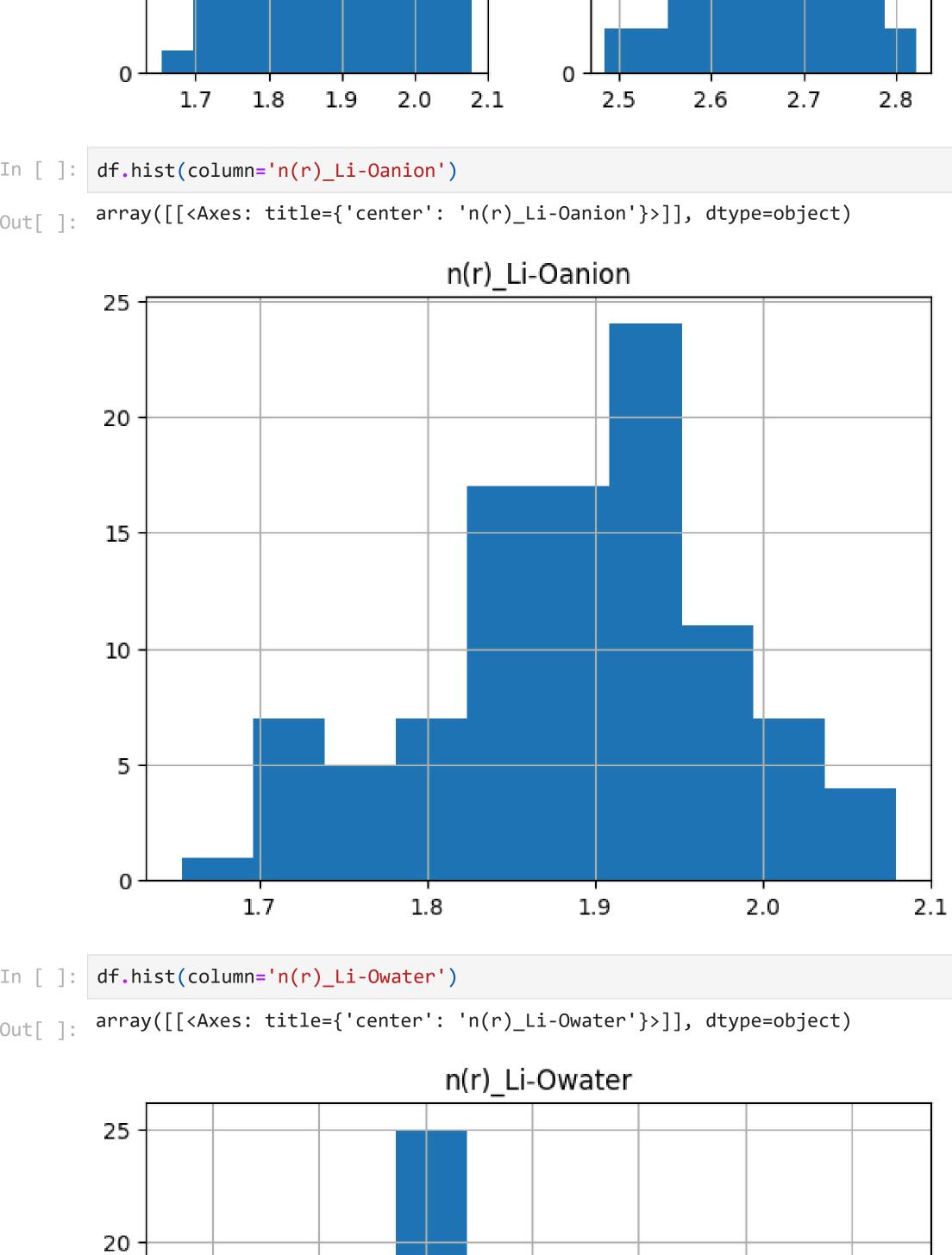
20

15

n(r)\_Li-Owater

\* Uma maneira de vizualisar os dados é usar um histograma:

10 10



5 2.55 2.50 2.60 2.65 2.70 2.75 2.80 O material de aula está disponível em:https://github.com/isabelamaral/Python-basic-stats-**CQ-1sem2024**