

# DADOS, PYTHON E ESTATÍSTICA: CAMINHOS PARA STEM

Ana Júlia Amaro Pereira Rocha

**SEMANA 1**  
**DIA 4**

# ESTATÍSTICA

**Definição:** área que estuda como coletar, organizar e analisar dados. Ajuda a transformar números em informação e a entendermos melhor o mundo ao nosso redor.

A estatística é muito usada para tomar decisões, aparece em pesquisas, notícias, ciência e no dia a dia.

**Ex:** pesquisas eleitorais, notas escolares, estudos de saúde...

**Sempre que vemos números explicando algo, há estatística envolvida.**

# DADOS E ESTATÍSTICA

Dados são informações coletadas sobre pessoas, objetos ou situações e a estatística trabalha a partir desses dados.

Os dados podem ser classificados de diferentes formas, sendo a distinção entre dados **quantitativos** e **qualitativos** uma das mais importantes, já que é a partir dela que escolhemos como analisar os dados.

**qualitativos:** descrevem características. Ex: tipo de escola, cor, cidade...

**quantitativos:** representam quantidades numéricas. Ex: idade, altura...

→ São subclassificados ainda em **discretos** e **contínuos**.

# POPULAÇÃO E AMOSTRA

**População** é o conjunto completo que queremos estudar. Pode ser um grupo de pessoas, objetos ou eventos. **Estudar toda a população muitas vezes não é possível.**

**Amostra** é uma parte da população, usada para representar o todo. Uma boa amostra deve ser colhida com cuidado uma vez que se a amostra não representa a população o resultado pode ser enganoso.

O uso de amostras economiza tempo e recursos. Com a estatística, podemos tirar conclusões da população inteira a partir desse “pedaço”.

# MEDIDAS DE TENDÊNCIA CENTRAL

**Medidas de tendência central** ajudam a resumir um conjunto de dados ao indicar um valor “típico” ou central. São usadas para entender o comportamento geral dos dados.

**Média** é a soma de todos os valores dividida pela quantidade de dados. É muito usada, mas pode ser influenciada por valores extremos. Assim, nem sempre a média representa bem todos os dados.

# MEDIDAS DE TENDÊNCIA CENTRAL

**Mediana** é o valor que fica no meio quando os dados são ordenados. Metade fica abaixo e metade acima da mediana, sendo menos sensível a extremos.

**Moda** é o valor que aparece com mais frequência. Um conjunto de dados pode ter uma ou mais modas. Às vezes pode não haver nenhuma também.

Cada medida de tendência central tem um uso diferente a depender do tipo de dado e da distribuição. Comparar elas ajuda a entender os dados.

# QUANTIS E QUARTIS

**Quantis** dividem os dados ordenados em partes iguais, ajudando a entender como os dados estão distribuídos. São úteis para identificar padrões e desigualdades.

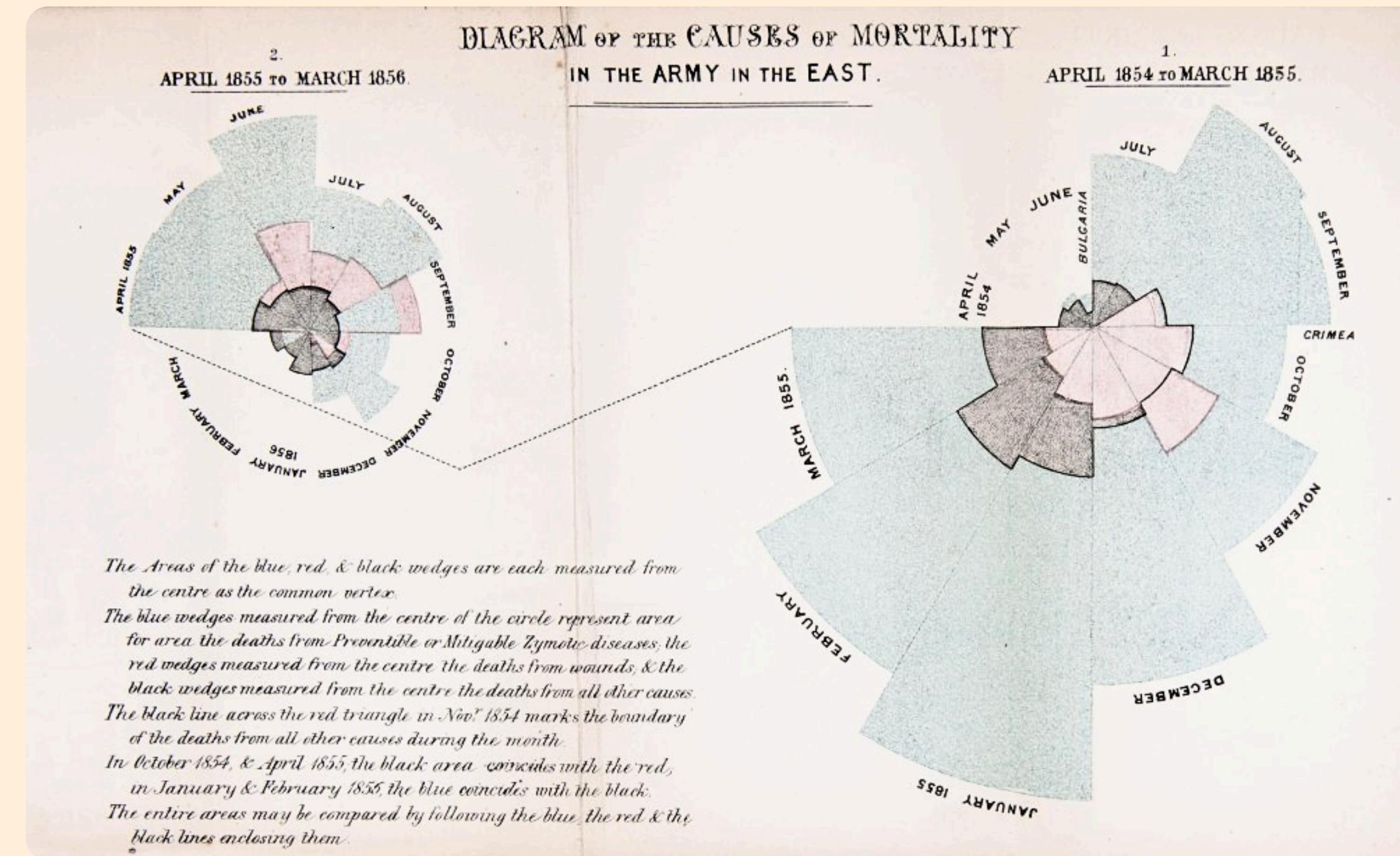
**Quartis** dividem os dados em quatro partes, ajudando a visualizar a dispersão dos dados. O segundo quartil corresponde à mediana.

# POR QUE VISUALIZAR DADOS?

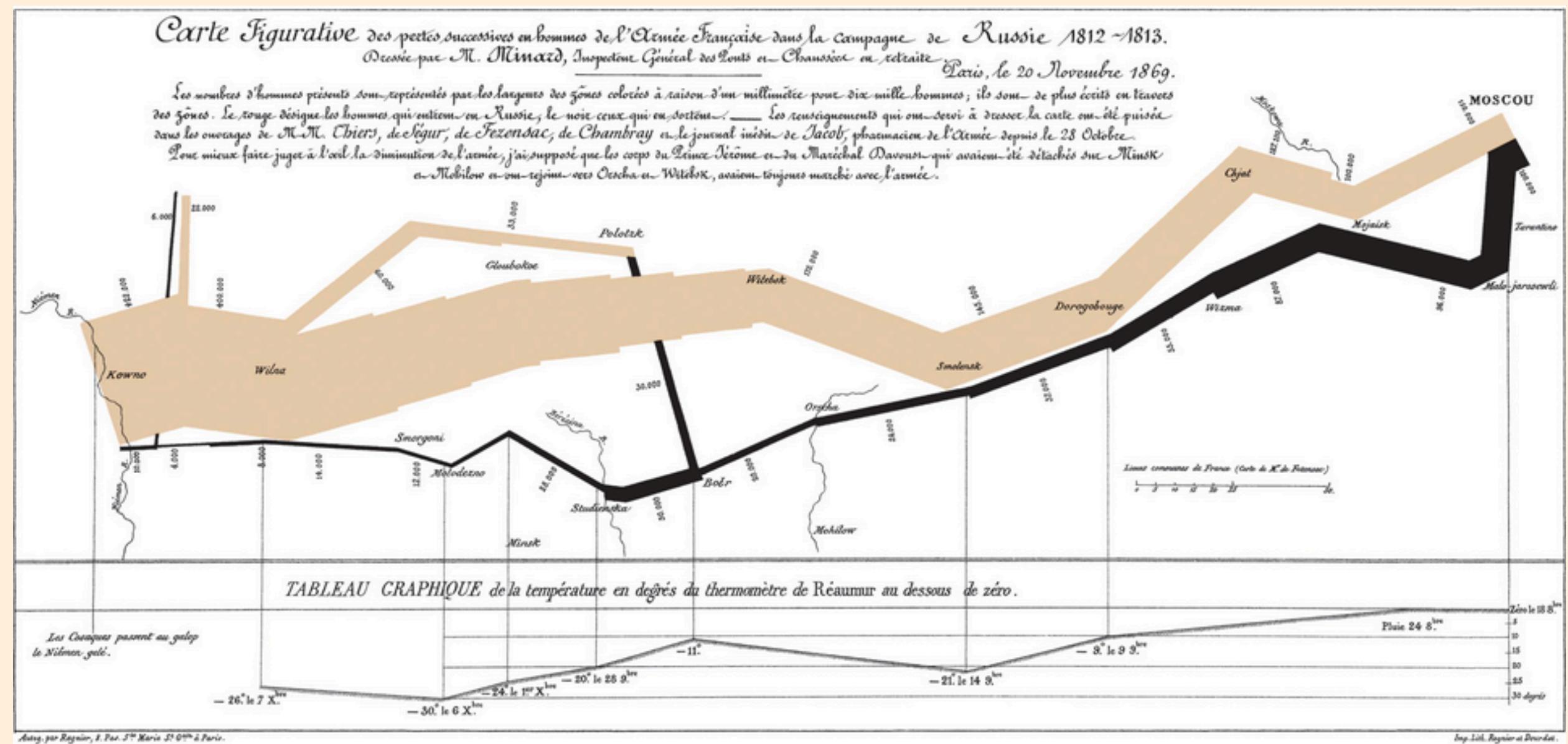
Gráficos facilitam a compreensão dos dados. Eles permitem identificar padrões rapidamente e, muitas vezes, explicam mais do que uma tabela.

Diferentes gráficos são usados para diferentes tipos de dados. Gráficos de barras, linhas e histogramas são comuns. Devemos sempre ter em mente que a **escolha do gráfico influencia a interpretação**. Então, antes de qualquer coisa, pense no que exatamente quer mostrar ao leitor.

# EXEMPLOS FAMOSOS



# EXEMPLOS FAMOSOS



# VISUALIZAÇÃO NO PAPEL

Mesmo sem computador, conseguimos visualizar dados. Desenhar gráficos no papel ajuda a entender os conceitos e por isso vamos fazer isso hoje.

Gráficos mostram tendências, comparações e diferenças, ajudando a contar a história por trás dos dados. Uma boa visualização torna a análise mais clara. Assim, a partir de um dos datasets a seguir, gostaria que fizessem uma visualização simples e calculassem as respectivas medidas de tendência central.