

# Atividade 01

José Felipe Custódia de Azevedo

25/09/2020

## Análise de dados

Em estatística, a **análise exploratória de dados** (AED) é uma abordagem à análise de conjuntos de dados de modo a resumir suas características principais, frequentemente com métodos visuais. Um modelo estatístico pode ou não ser usado, mas primariamente a **AED** tem como objetivo observar o que os dados podem nos dizer além da modelagem formal ou do processo de teste de hipóteses. A análise exploratória de dados foi promovida pelo estatístico norte-americano John Tukey.



John Tukey



John Tukey

A **análise exploratória de dados** emprega grande variedade de técnicas gráficas e quantitativas, visando maximizar a obtenção de informações ocultas na sua estrutura, descobrir variáveis importantes em suas tendências, detectar comportamentos anômalos do fenômeno, testar se são válidas as hipóteses assumidas, escolher modelos e determinar o número ótimo de variáveis.

## Técnicas

Há várias ferramentas úteis para a análise exploratória de dados, mas a **AED** é mais caracterizada pela atitude tomada do que por técnicas particulares. Técnicas gráficas típicas usadas na análise exploratória de dados são: (Mais informações acesse as respectivas páginas)

Diagrama de caixa;

Histograma;

Diagrama de Pareto;

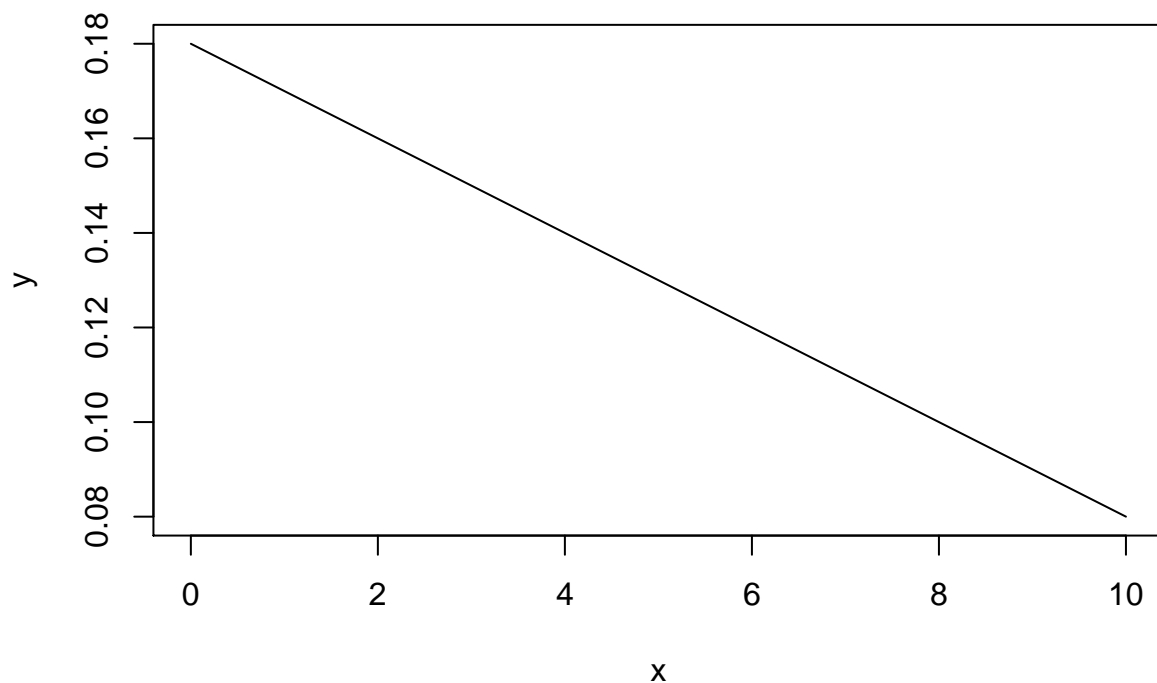
Existem diversas técnicas gráficas que podemos encontrar na **Wikipedia** ou no **Medium**.

**Exemplo:**

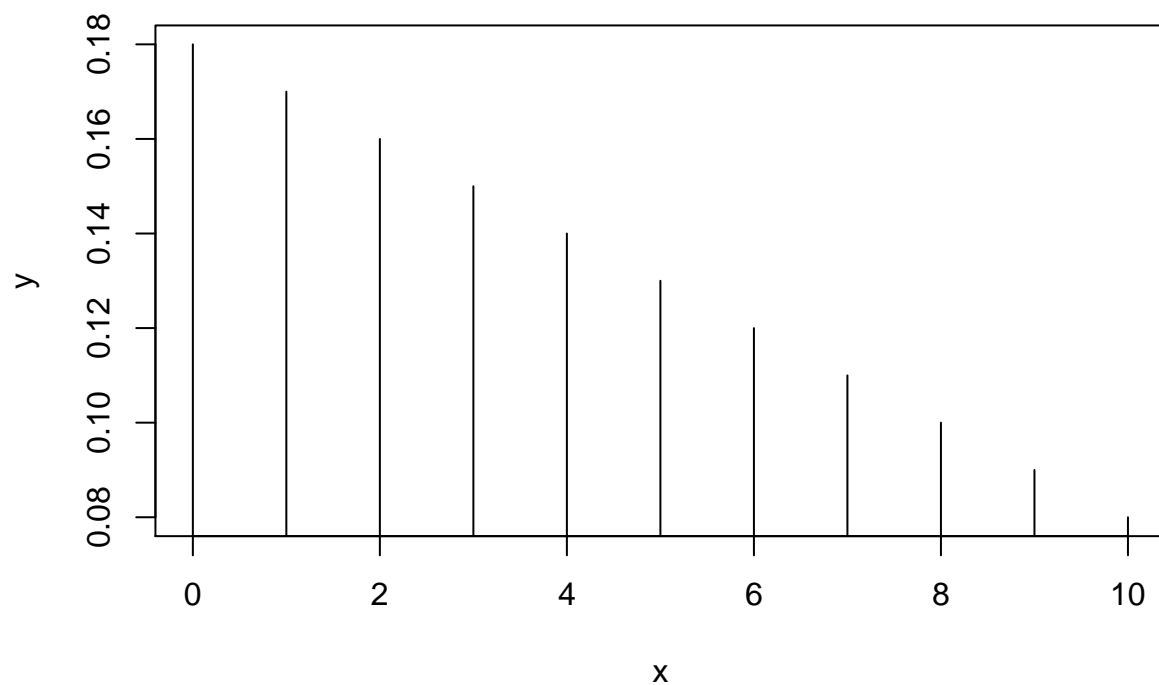
Para ilustrar, considere um exemplo em que a tarefa da análise é encontrar as variáveis que melhor preveem a gorjeta que os clientes de uma mesa em um jantar darão ao garçom.[21] As variáveis disponíveis nos dados coletados para esta tarefa são: o montante da gorjeta, o montante da conta, o gênero do pagante, seção de fumantes/não fumantes, a hora do dia, o dia da semana e o tamanho da mesa. A tarefa da análise primária é aproximada ao ajustar um modelo de regressão em que a razão da gorjeta é a variável de resposta. O modelo ajustado é:

$$Y = 0.18 - 0.01m$$

*Portanto, o gráfico da função é:*



```
x=10:0  
y=(0.18 - 0.01*x)  
plot(x,y, 'h')
```



Tal que  $x$  é o tamanho da mesa e  $y$  a gorjeta.

```
m <-2  
y = 0.18 - 0.01*m  
print("m com valor de 2:")
```

```
## [1] "m com valor de 2:"
```

```
print(y)
```

```
## [1] 0.16
```