



Quem se prepara, não para.

# Business Intelligence

4º período

Professora: Michelle Hanne

A **Estatística Descritiva** Resume ou descrição das características importantes de um conjunto conhecido de dados populacionais... está relacionada com a organização e descrição de dados associada a *cálculos de médias, variâncias, estudo de gráficos, tabelas, etc.*

# Principais Conceitos em Estatística

- **População:**
  - E todo o conjunto de elementos que possuam ao menos uma característica comum observável.
- **Amostra:**
  - E uma parte da população que será avaliada por um critério comum.
- **Dados estatísticos:**
  - São os valores associados as variáveis de pesquisas.
- **Frequências:**
  - O numero de vezes em que a variável ocorre e chamado **frequência absoluta** e é indicado por  $n_i$ .
  - Definimos **frequência relativa** ( $f_i$ ) como a razão entre a frequência absoluta ( $n_i$ ) e o numero total de observações ( $n$ ) , ou seja:

$$f_i = n_i/n$$

# Principais Conceitos em Estatística

- **Parâmetros:**

- São valores singulares que existem na população e que servem para caracterizá-la. Para definirmos um parâmetro devemos examinar toda a população. *Ex: Os alunos da Newton Paiva têm em média 1,70 metros de estatura.*

- **Estimativa:**

- É um valor aproximado do parâmetro e é calculado com o uso da amostra.

- **Atributo:**

- Quando os dados estatísticos apresentam um caráter qualitativo, o levantamento e os estudos necessários ao tratamento desses dados são designados genericamente de estatística de atributo.

- **Variável:**

- É o conjunto de resultados possíveis de um fenômeno.

# Principais Conceitos em Estatística

- **Variável Qualitativa:**
  - Quando seu valores são expressos por **atributos**: sexo, cor da pele, etc.
    - **Nominais**
    - **Ordinais**
- **Variável Quantitativa:**
  - Quando os **dados são de caráter nitidamente quantitativo**, e o conjunto dos resultados **possui uma estrutura numérica**, trata-se portanto da estatística de variável e se dividem em :
    - **Variável discreta ou descontínua**
    - **Variável contínua**

# Variável Qualitativa

- **Variável nominal:** valores que expressam atributos, sem nenhum tipo de ordem. Ex: cor dos olhos, sexo, estado civil, etc.
- **Variável ordinal:** quando existe uma ordem nos seus valores. Por exemplo, a variável “Grau de instrução” pode ter seus valores ordenados (fundamental, médio, superior, etc). O mesmo não ocorre com a variável “cor da pele”.

# Variável Quantitativa

- **Variável Discreta ou Descontínua:** Seus valores são expressos geralmente através de números inteiros não negativos. Resulta normalmente de contagens.
  - *Ex:* N<sup>o</sup> de alunos presentes às aulas de introdução à estatística econômica no 1<sup>o</sup> semestre de 1997: mar = 18 , abr = 30 , mai = 35 , jun = 36.
- **Variável Contínua:** Resulta normalmente de uma **mensuração**, e a **escala numérica de seus possíveis valores corresponde ao conjunto  $\mathbb{R}$**  dos números Reais, ou seja, podem assumir, teoricamente, qualquer valor entre dois limites.
  - *Ex.:* Quando você vai medir a temperatura de seu corpo com um termômetro de mercúrio o que ocorre é o seguinte: O filete de mercúrio, ao dilatar-se, passará por todas as temperaturas intermediárias até chegar na temperatura atual do seu corpo.



# Tipos de Variáveis

- Cor dos olhos das alunas: **qualitativa**
- Índice de liquidez nas indústrias: **quantitativa contínua**
- Produção de café no Brasil: **quantitativa contínua**
- Número de defeitos em aparelhos de TV: **quantitativa discreta**
- Comprimento dos pregos produzidos por uma empresa: **quantitativa contínua**
- O ponto obtido em cada jogada de um dado: **quantitativa discreta**

# Análise Descritiva Bivariada

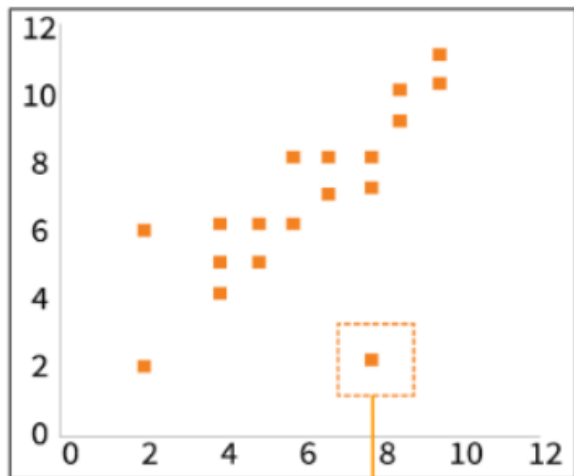
- Variável Quantitativa X Variável Quantitativa  
Gráfico de dispersão, outliers, regressão linear.
- Variável Qualitativa X Variável Quantitativa  
Box Plot+ média + desvio padrão
- Variável Qualitativa X Qualitativa

# Análise Descritiva Bivariada

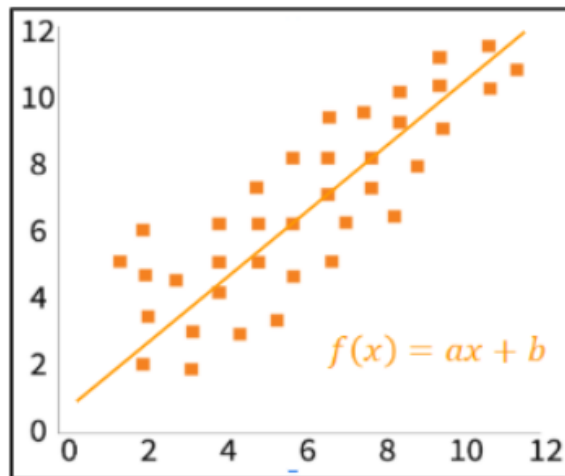
Envolve a análise de duas variáveis, com o objetivo de determinar a relação empírica entre elas. A análise bivariada pode ser útil para testar hipóteses simples de associação:

- **Variável Quantitativa X Variável Quantitativa**
  - Gráfico de dispersão, outliers, regressão linear.
- **Variável Qualitativa X Variável Quantitativa**
  - Box Plot+ média + desvio padrão
- **Variável Qualitativa X Qualitativa**

# Variável Quantitativa X Variável Quantitativa

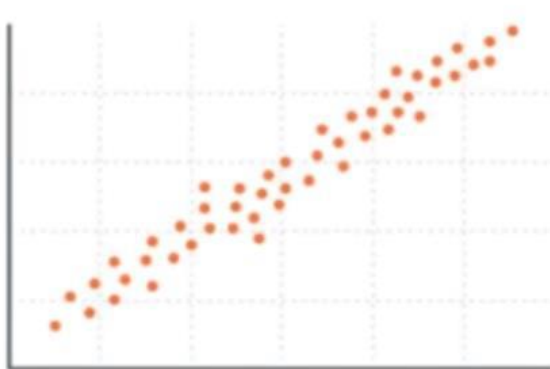


Ponto muito discrepante (*outlier*) com relação ao conjunto de dados. É necessário o isolamento e tratamento do mesmo para eliminação corretiva de suas causas.



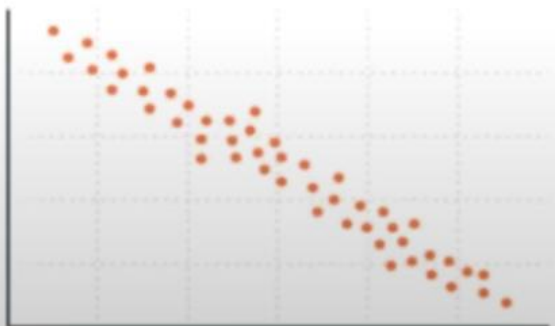
O Diagrama de Dispersão pode ser utilizado para auxiliar na modelagem de Regressão Linear, o qual busca, a partir de uma equação matemática, resumir a distribuição de dados em torno de uma reta.

# Variável Quantitativa X Variável Quantitativa



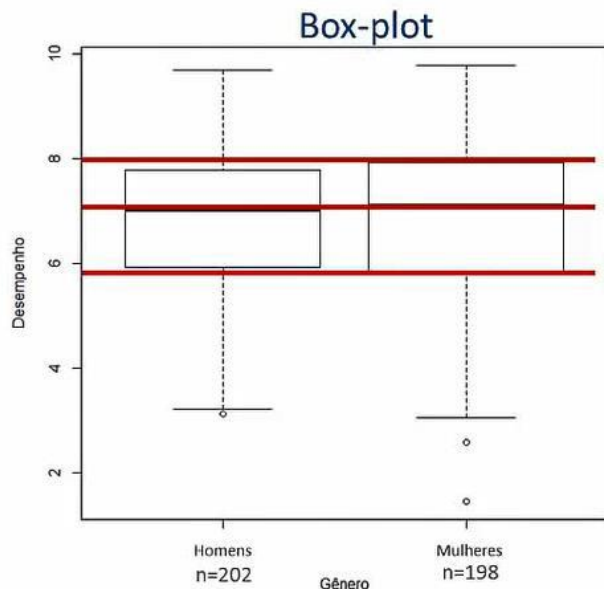
## Relação Linear:

- Relação Fraca
- Relação Forte
- Nenhuma Relação



# Variável Qualitativa X Variável Quantitativa

## Análise do desempenho por gênero



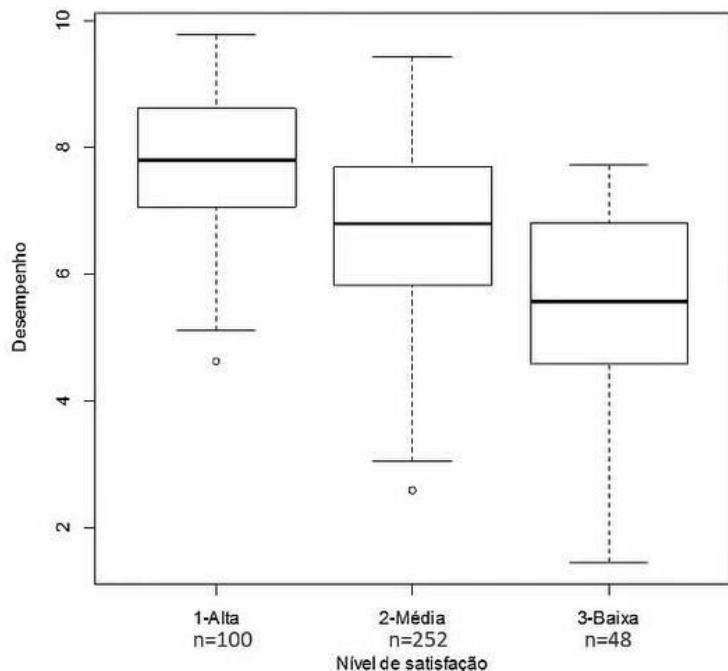
Rótulos de Linha	Média de Desempenho acadêmico	DesvPad de Desempenho acadêmico
Homem	6,83	1,45
Mulher	6,85	1,45
Total Geral	6,84	1,45

Não há relação entre gênero e desempenho, pois:

- ✓ Box-plots do desempenho para homens e mulheres são muito parecidos (mediana e quartis são praticamente iguais para homens e mulheres),
- ✓ Desempenho médio de homens e mulheres são similares;
- ✓ Desvios-padrão do desempenho, para homens e mulheres, não são menores que o desvio-padrão do desempenho em geral.

# Variável Qualitativa X Variável Quantitativa

## Análise do desempenho por nível de satisfação



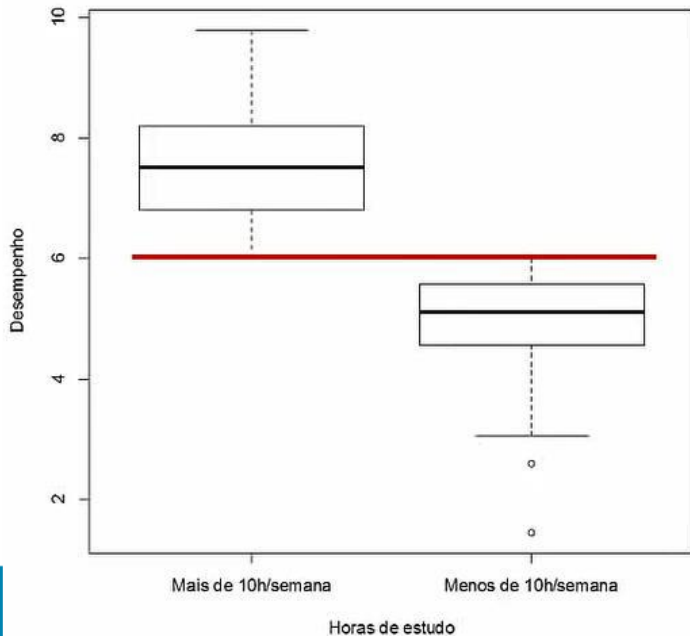
Rótulos de Linha	Média de Desempenho acadêmico	DesvPad de Desempenho acadêmico
Alta	7,78	1,17
Média	6,70	1,34
Baixa	5,64	1,40
Total Geral	6,84	1,45

Há relação entre nível de satisfação e desempenho, pois:

- ✓ Box-plots do desempenho são diferentes para cada nível (mediana e quartis maiores quanto maior o nível de satisfação);
- ✓ Desempenho médio aumenta conforme aumenta o nível de satisfação;
- ✓ Desvio-padrão do desempenho dentro de cada nível de satisfação é menor que o desvio-padrão do desempenho geral.

# Variável Qualitativa X Variável Quantitativa

## Análise do desempenho por horas de estudo



Rótulos de Linha <input type="checkbox"/>	Média de Desempenho acadêmico	DesvPad de Desempenho acadêmico2
Mais de 10h/semana	7,54	0,91
Menos de 10h/semana	4,96	0,82
Total Geral	6,84	1,45

Há **forte** relação entre horas de estudo e desempenho, pois:

- ✓ Box-plots do desempenho são diferentes para cada grupo (mediana e quartis maiores quando horas=Mais de 10/semana);
- ✓ Desempenho médio, conforme horas de estudo, difere muito;
- ✓ Desvio-padrão do desempenho, dentro de cada grupo é menor que o desvio-padrão do desempenho geral.



- **Associação entre uma variável qualitativa e uma quantitativa:**  
*Há associação quando:*
  - os box-plots são diferentes;
  - as médias dos grupos são diferentes e
  - os desvios-padrão dentro dos grupos são menores que o desvio-padrão geral.
- **Quanto maior a diferença entre as médias e menores forem os desvios-padrão, mais forte a associação!**

# Análise Multivariada

Refere-se a um conjunto de métodos estatísticos que torna possível a análise simultânea de medidas múltiplas para cada indivíduo, objeto ou fenômeno observado.

Todas as variáveis envolvidas devem ser aleatórias e relacionadas, de modo que seus efeitos não sejam expressivamente interpretados de maneira separada.

# Análise Multivariada

Essa classe de técnicas é voltada a ajudar-nos a compreender os relacionamentos entre duas ou mais variáveis.

**Medidas de associação:** permitem quantificar a força e a direção do relacionamento entre variáveis.

Tais medidas podem ajudar a investigar duas questões de grande importância teórica e prática:

- **Causalidade** (relações de causa e efeito);
- **Previsão:** quando duas variáveis têm uma alta medida de associação, é possível prever (com certo grau de convicção) o valor de uma a partir da outra.

Importante: forte associação entre dois fatores não implica imediatamente em relação causal entre os fatores.

- Mas ajuda o investigador a pesquisar possíveis causas para essa associação.

Suponha que estejamos analisando os dados de um grupo de estudantes, e o interesse seja analisar o desempenho de duas variáveis:

- Quantidade de tempo de estudo e nota final obtida.
- Suponha que tenhamos encontrado uma associação positiva forte entre essas duas variáveis.
- Isso indicaria que "tempo de estudo" e "nota" teriam relação próxima (força do relacionamento) e à medida em que uma variável aumentasse de valor, a outra também aumentaria (direção do relacionamento).
- Você poderia fazer previsões de uma variável a partir de outra:
- "Quanto maior o tempo de estudo, maior a nota final"

Questão: Todavia, seria possível estabelecer uma relação causal?

- "Um maior tempo de estudo leva a uma maior nota final"
- Resposta: não, pois outras variáveis podem interferir, e precisariam ser analisadas em conjunto com "tempo de estudo" e "nota final".

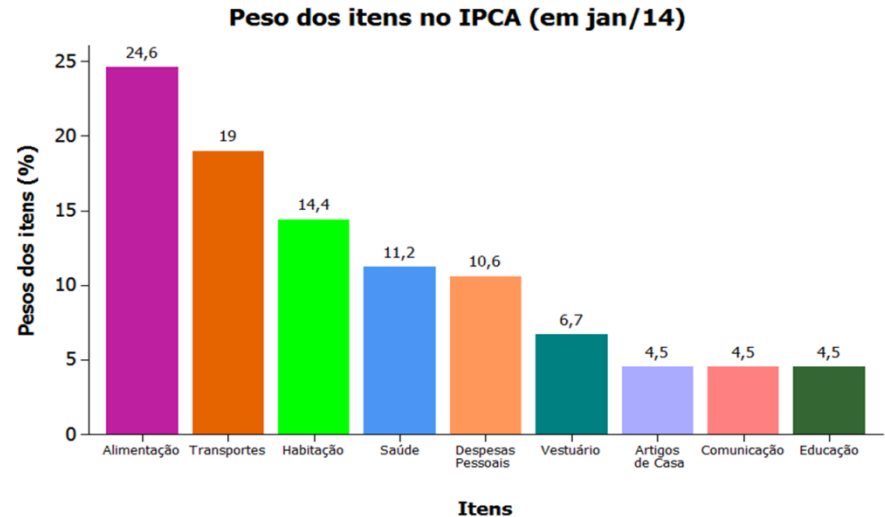
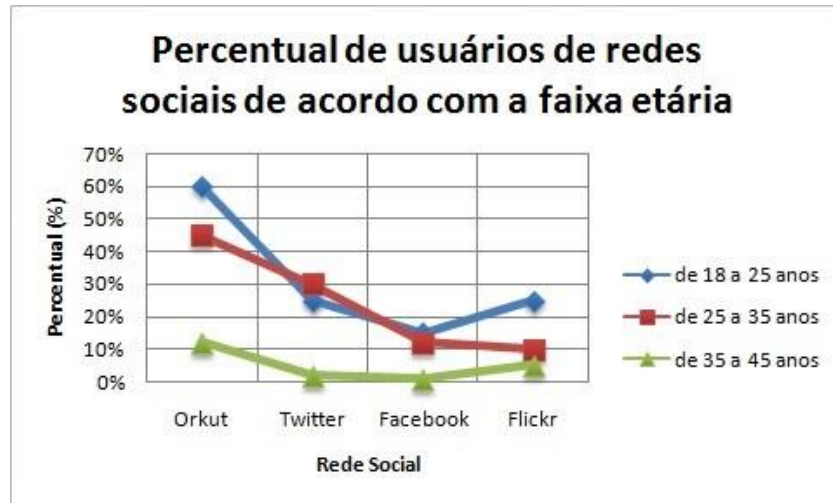
# ***Representação Gráfica de uma Distribuição***

# Gráficos

- São representações visuais dos dados estatísticos que devem corresponder, mas nunca substituir as tabelas estatísticas.
- Características: Uso de escalas, sistema de coordenadas, simplicidade, clareza e veracidade.
- **Classificação dos gráficos**: Diagramas, Estereogramas, Pictogramas e Cartogramas.

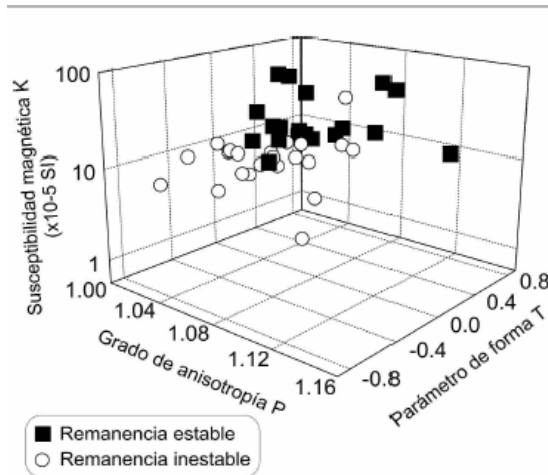
# Gráficos

- **Diagramas:** São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas. Eles podem ser: Barras, Linhas, Colunas e Setores

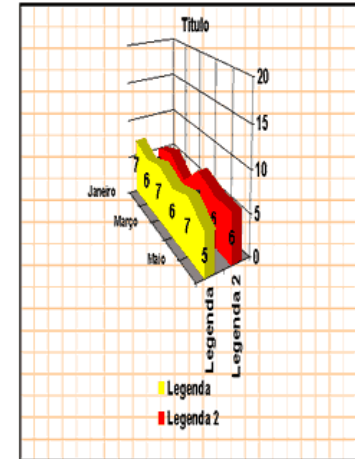


# Gráficos

- **Estereogramas**: São gráficos geométricos dispostos em três dimensões, pois representam volume. São usados nas representações gráficas das tabelas de dupla entrada. Em alguns casos este tipo de gráfico fica difícil de ser interpretado dada a pequena precisão que oferecem.



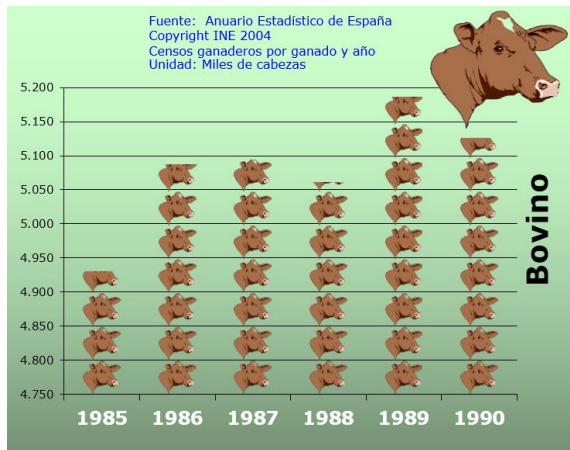
	Legenda	Legenda 2
Janeiro	7	5
Fevereiro	6	6
Março	7	4
Abril	6	7
Maió	7	6
Junho	5	6
Julho	6	5





# Gráficos

- **Pictogramas**: São construídos a partir de figuras representativas da intensidade do fenômeno. Este tipo de gráfico tem a vantagem de despertar a atenção do público leigo, pois sua forma é atraente e sugestiva. Os símbolos devem ser autoexplicativos. A desvantagem dos pictogramas é que apenas mostram uma visão geral do fenômeno, e não de detalhes minuciosos. Veja o exemplo abaixo:



# Gráficos

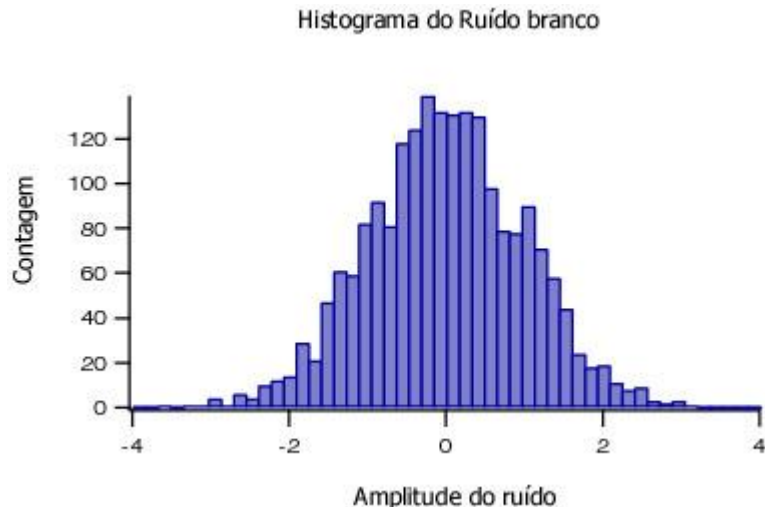
- **Cartogramas**: São ilustrações relativas a cartas geográficas (mapas). O objetivo desse gráfico é o de figurar os dados estatísticos diretamente relacionados com áreas geográficas ou políticas.



# Gráficos

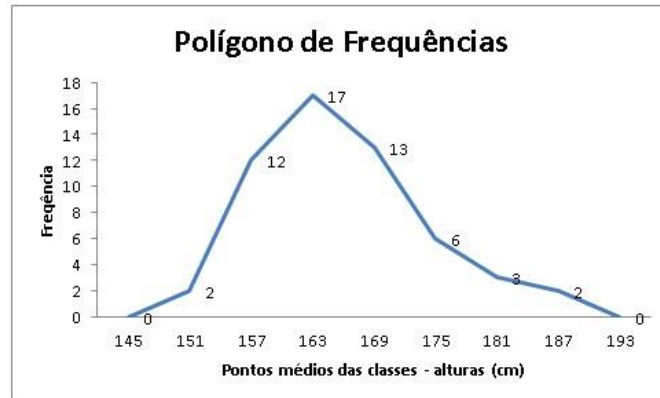
- **Histograma:**

- formado por um conjunto de retângulos justapostos, cujas bases se localizam sobre o eixo horizontal, de tal modo que seus pontos médios coincidam com os pontos médios dos intervalos de classe. A área de um histograma é proporcional à soma das frequências simples ou absolutas.



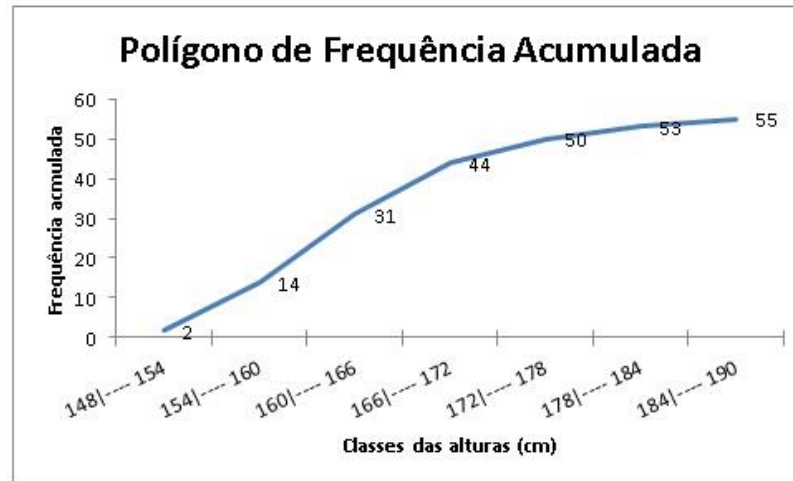
# Gráficos

- **Polígono de frequência:** é um gráfico em linha, sendo as frequências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe. Para realmente obtermos um polígono (linha fechada), devemos completar a figura, ligando os extremos da linha obtida aos pontos médios da classe anterior à primeira e da posterior à última, da distribuição.



# Gráficos

- **Polígono de frequência acumulada:** é traçado marcando-se as frequências acumuladas sobre perpendiculares ao eixo horizontal, levantadas nos pontos correspondentes aos limites superiores dos intervalos de classe.



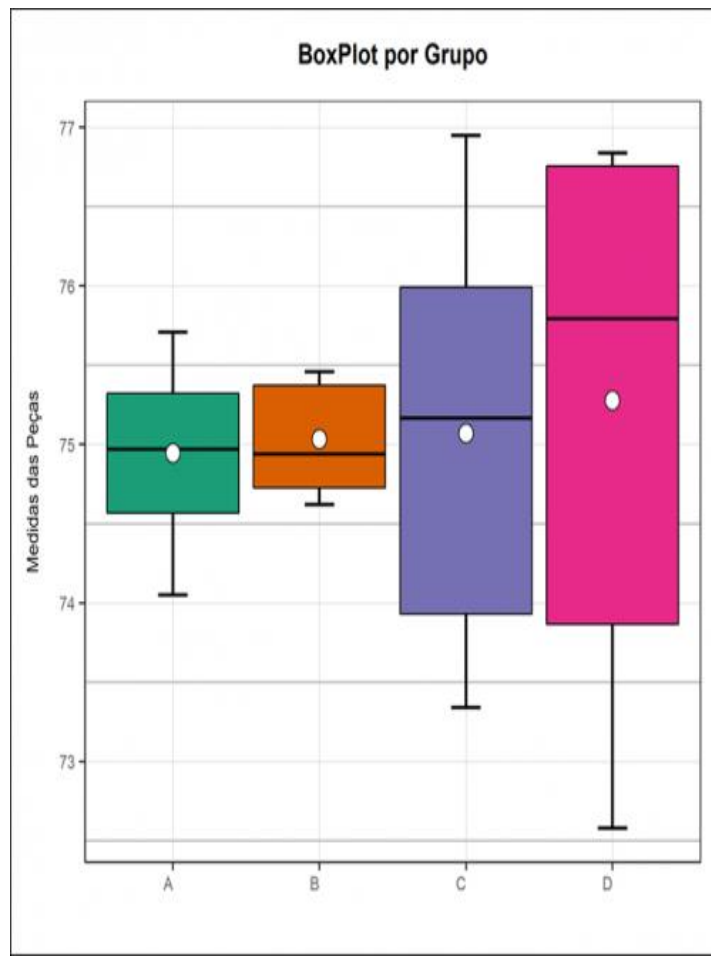
# Exemplo

Para verificar a eficiência do treinamento, foram selecionadas 10 peças produzidas pelas equipes A e B e 10 peças produzidas pelas equipes C e D que não participaram do treinamento

A		B		C		D	
75,27	74,93	74,94	74,75	75,93	73,34	75,98	76,75
75,33	74,72	75,25	74,65	76,95	74,04	75,61	76,78
74,58	74,53	75,44	74,94	75,47	75	74,2	74,74
75,01	75,32	74,62	74,92	73,6	76,18	76,44	72,58
75,71	74,05	75,35	75,46	74,85	75,33	76,84	72,86

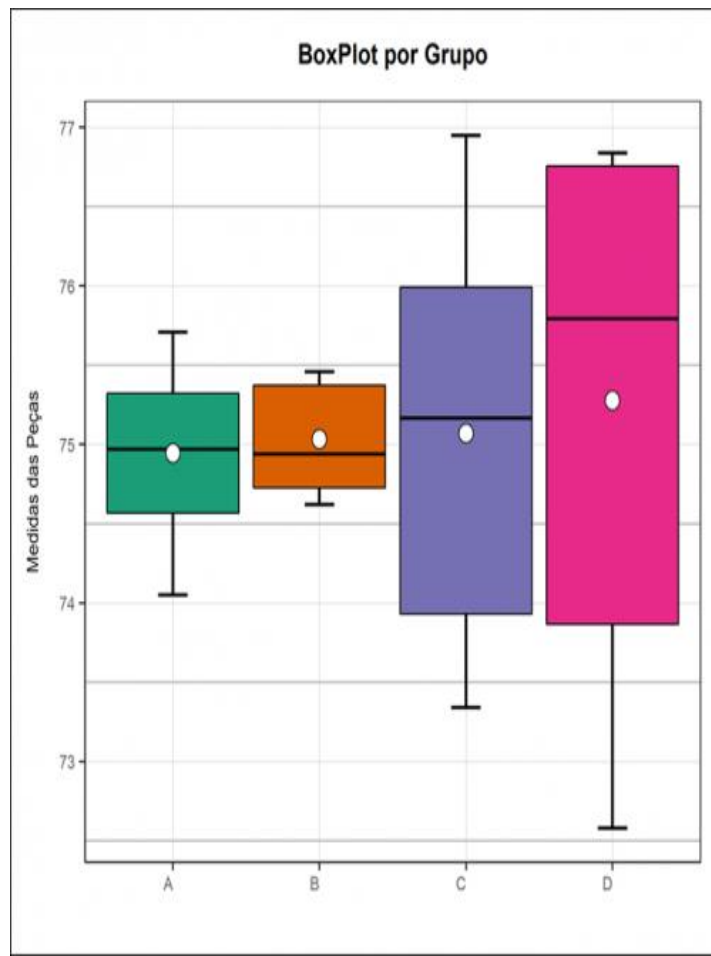
Analisando o gráfico podemos observar que:

- As equipes **A e B produzem peças com menor variabilidade**, indicando que o treinamento teve o efeito desejado;
- A equipe **D é a que produz peças com maior variabilidade**;
- A equipe **B é a que produz peças com menor variabilidade**.



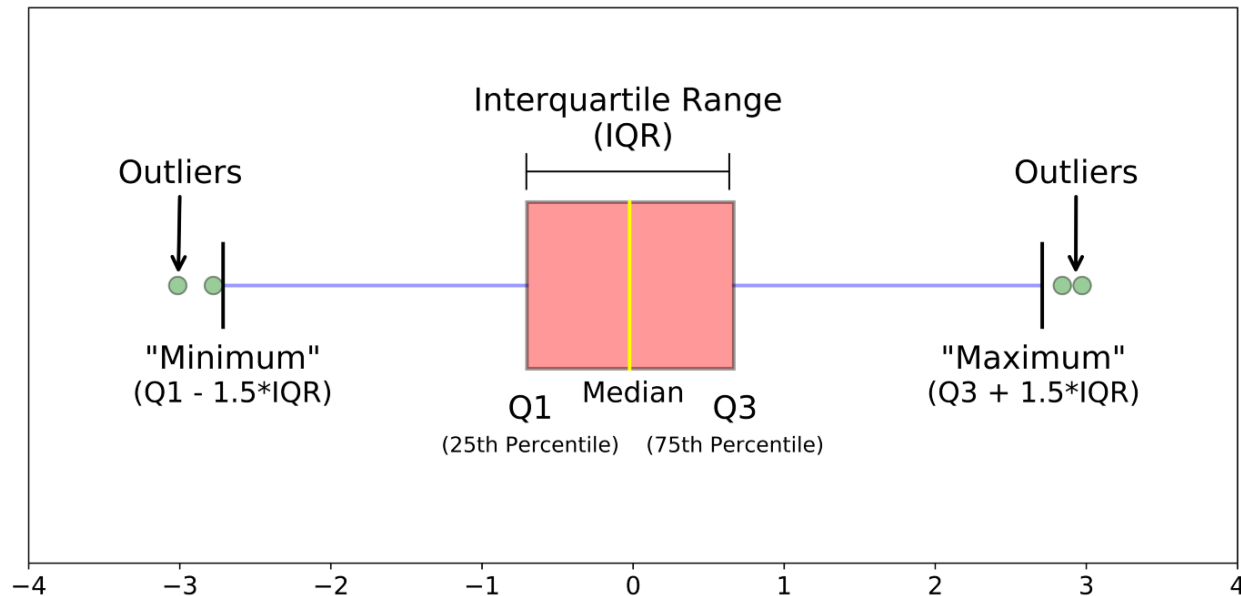
## Considerações:

Como as peças das **equipes A e B** tem **menor variabilidade** e com valor **médio próximo do valor de referência**, vale a pena enviar as demais equipes para o treinamento.





# Cálculo de Outliers



Basicamente, este tipo de visualização é construído através dos resultados das medidas abaixo:

- Mediana (também considerado 2º quartil — 50% da distribuição) — *Median*;
- Quartis (Q1 e Q3) — 25th percentile e 75th percentile;
- Amplitude Interquartil (diferença entre o 3º quartil e o 1º quartil) — *Interquartile Range (IQR)*;
- Limite inferior (Mínimo) — *Minimum*;  
Limite inferior =  $Q1 - 1,5 * (Q3 - Q1)$
- Limite superior (Máximo) — *Maximum*;  
Limite superior =  $Q3 + 1,5 * (Q3 - Q1)$

# Box-plot e Histograma no PowerBI

- Clique no ícone (...) obter mais visuais
- Adicionar uma conta válida da Microsoft.
- Na janela de Visuais da Microsoft procurar por box, escolha a opção “Box and Whisker chart by MAC”

# Box-plot e Histograma no PowerBI

Visuais do Power BI

① Ao clicar em "Adicionar" e/ou "Baixar Amostra" e baixar um visual, você concorda com os Termos e Condições e a Política de Privacidade do provedor na página do visual e concorda que a Microsoft pode compartilhar os detalhes da sua conta com o provedor para fins transacionais. O uso do Microsoft AppSource está sujeito aos [Termos do Marketplace Comercial da Microsoft](#) e [Declaração de Privacidade](#).

Todos os visuais Visuais organizacionais AppSource visuais

Explore todos os visuais disponíveis para ampliar seus insights de negócios [Saiba mais](#)

Filtrar por Todos

Classificar por: Popularidade

Box and Whisker c...  
MAQ LLC  
★★★★★ (58)  
Understand the mean, median, and quartile distribution of your data.

Filter by List  
DEVSCOPE  
★★★★★ (4)

Mapbox Visual  
Mapbox  
★★★★★ (28)

Box and Whisker c...  
DataScenarios  
★★★★★ (84)

Timeline Box  
TheraTraQ  
★★★★★ (2)

Mapbox Custom Vi...  
Starschema Kereskedelmi ...

DataText Box  
北京上北智信科技有限公司

swBoxPlot  
上北智信

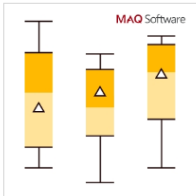
Add Intelligent Nar...  
Arria NLG

# Box-plot e Histograma no PowerBI

AppSource

Aplicativos para Power BI visuals

< Aplicativos



**Box and Whisker chart by MAQ Software**  
MAQ LLC  
★★★★★ 3.8 (58)  
PBI Certified

**Visão geral** Classificações + revisões

**Adicionar**

Baixar Amostra

[Instruções de amostra](#)

**Preços**  
Gratuito

**Produtos**  
[Visuais do Power BI](#)

**Editor**  
MAQ LLC

**Adquirir Usando**  
Conta corporativa ou de estudante

**Versão**  
4.0.0.0

**Atualizado**  
10/03/2022

**Suporte**  
[Suporte](#)  
[Ajuda](#)

Understand the mean, median, and quartile distribution of your data.

Compare large data sets in a compact space. Box and Whisker Chart by MAQ Software enables you to compare the distribution of values across several sets of data. In addition to showing the median, first and third quartiles, and the maximum and minimum values, Box and Whisker Chart by MAQ Software displays the mean, standard deviation, and quartile deviation.

**Business Uses:**

- Sales– Compare quarter over quarter sales across geographic regions or market demographics
- Project Managers – Analyze work output after major process changes
- Product Developers – Compare different processes to develop the same product at an improved speed, accuracy, or overall functionality

**Key Features:**


- Clear display of both broad categories (such as financial quarter or geographic region) and the specific data sets within those categories

Whisker Type:

Whisker Type: Min/Max  
FY20 Sales by Quarter

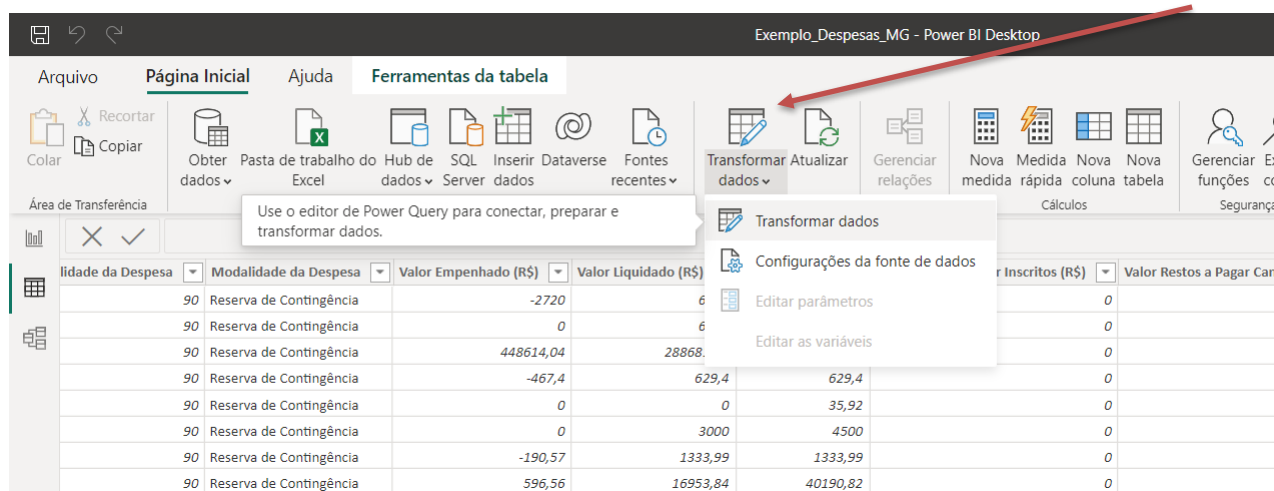
One Standard Deviation  
FY20 Sales by Quarter

Whisker Type: Custom  
FY20 Sales by Quarter



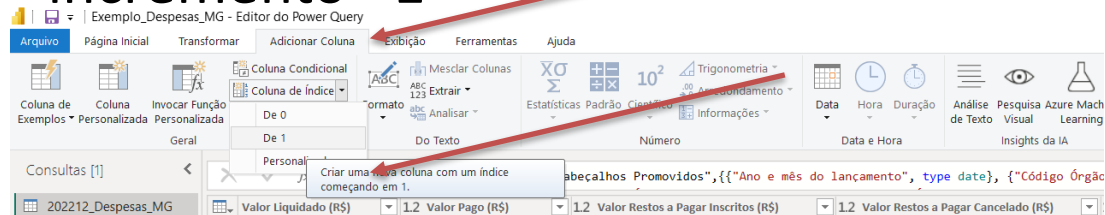
# Box-plot e Histograma no PowerBI

- Para fazer um box-plot precisamos criar uma coluna do tipo índice (auto numeração).
- Acesse o Power Query ou clique Transformar Dados



# Box-plot e Histograma no PowerBI

- No Power Query – selecione Adicionar Coluna > Coluna de índice, selecione a Personalizado. Configure inicial=1 e incremento =1



- Salve e feche o Power Query
- Atualize os dados

## Adicionar Coluna de Índice

Adicione uma coluna de índice com um índice inicial especificado e incremento.

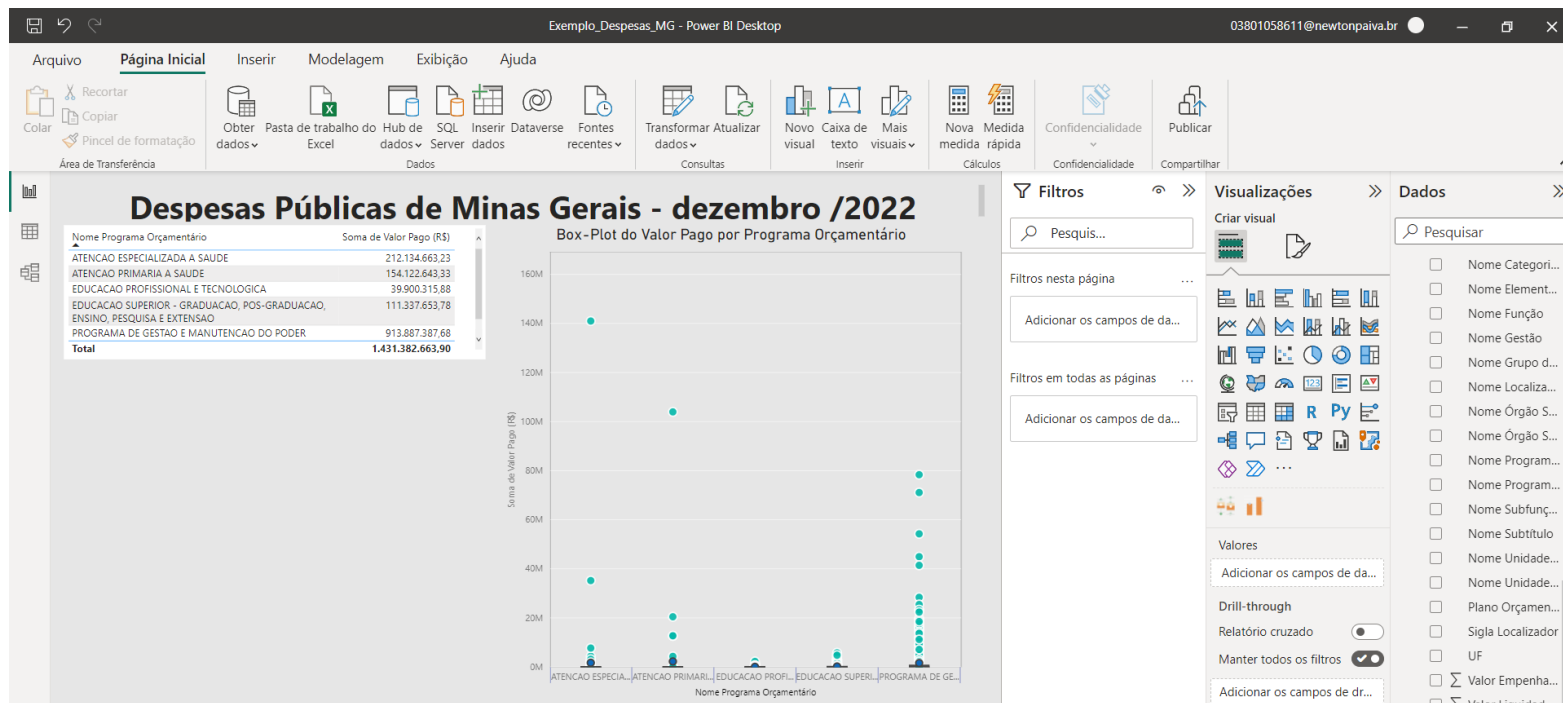
Índice Inicial

Incremento

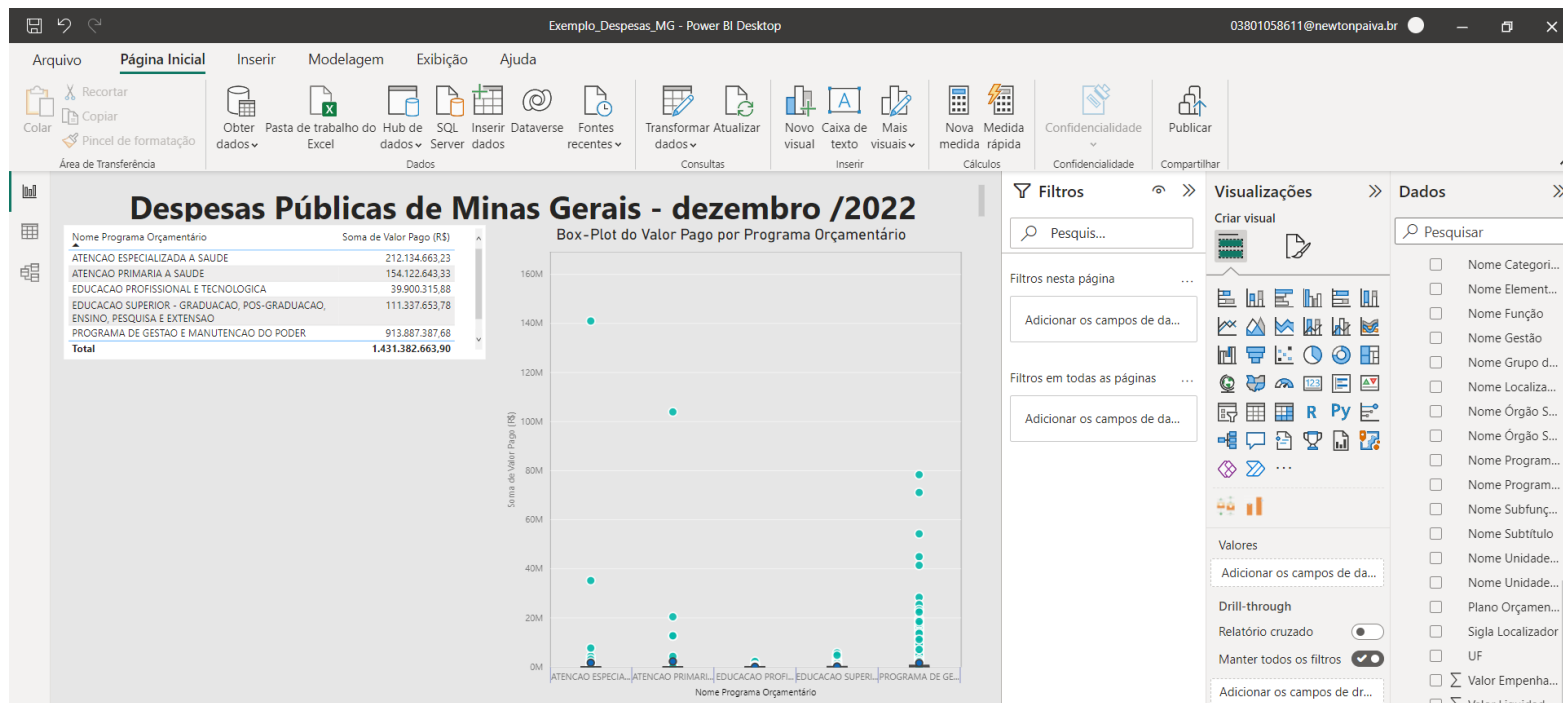
OK

Cancelar

# Box-plot e Histograma no PowerBI



# Box-plot e Histograma no PowerBI





# Análise Bivariada

- **Vídeo 1:** Relação entre duas variáveis quantitativas:  
<https://www.youtube.com/watch?v=nW4yMYf8YDg>
- **Vídeo 2:** Relação entre uma variável quantitativa e uma variável qualitativa:  
<https://www.youtube.com/watch?v=qMVALuutgU0>
- **Vídeo 3:** Relação entre duas variáveis qualitativas  
<https://www.youtube.com/watch?v=7zzQAPuATK8>

# Referências

BRAGHITTONI, Ronaldo. **Business Intelligence**: Implementar do jeito certo e a custo zero. São Paulo: Casa do Código, 2017.

GONÇALVES, Glauber Rogério Barbieri. **Sistemas de informação**. Porto Alegre: SAGAH, 2017.

KIMBALL, Ralph. **Data warehouse toolkit** - Técnicas para Construção de Data Warehouses Dimensionais. São Paulo: Makron Books, 1998.

TURBAN, Efraim et al. **Business Intelligence**: um enfoque gerencial para a inteligência do negócio. Porto Alegre: Bookman, 2009.