

PROGRAMAÇÃO EM R

AULA 03 – A CONSTRUÇÃO DE GRÁFICOS COM A LINGUAGEM R

29/01/2024

AGENDA

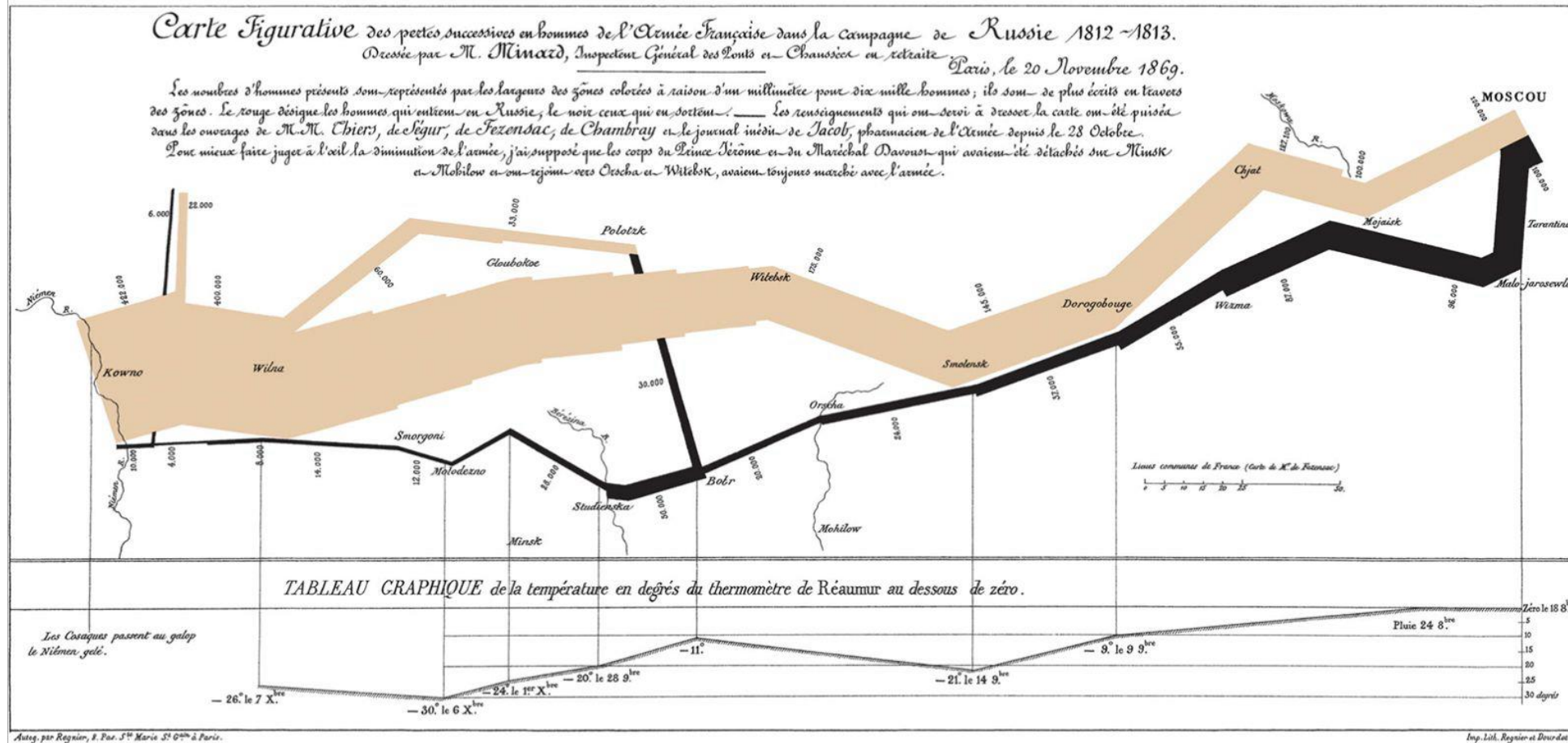
- O que é um gráfico?
- Importância do dataviz;
- Gramática dos gráficos: a construção dos gráficos em camadas;
- Pacotes de gráficos nativos do R: graphics e grdevices;
- Principais tipos de gráficos;
- Ggplot2;
- Vantagens do pacote ggplot2.

O QUE É UM GRÁFICO?

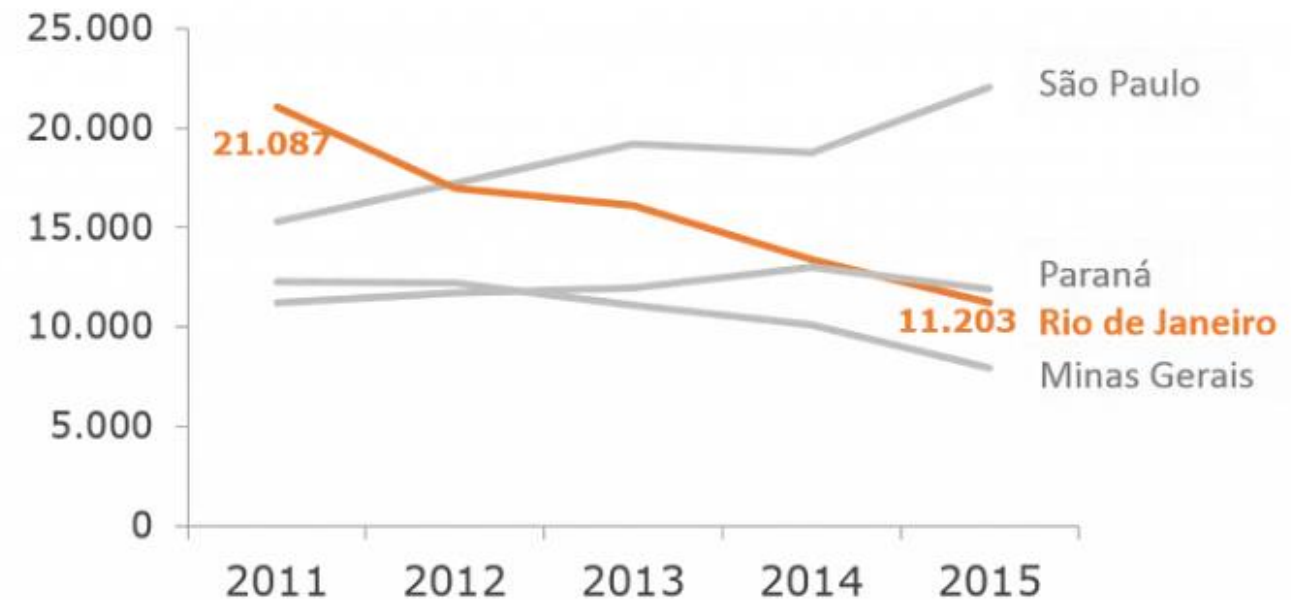
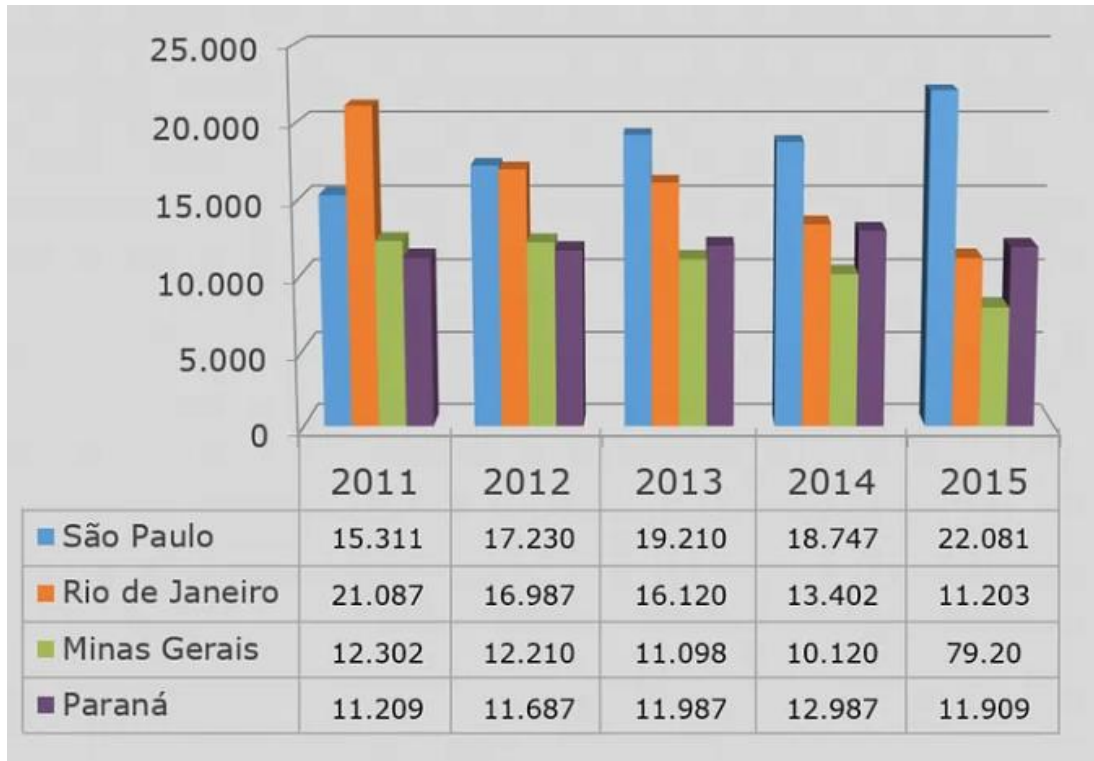
- Os gráficos são recursos visuais muito utilizados para facilitar a leitura e compreensão das informações e divulgação de pesquisas em jornais, revistas, panfletos, livros e televisão. (IBGE);
- O gráfico é uma representação com forma geométrica (desenho) de maneira exata e precisa a partir de informações numéricas obtidas através de pesquisas e organizadas em uma tabela.
- Todo gráfico tem um objetivo, um propósito;
- Os gráficos devem fazer parte de todo o processo de data science, desde a análise exploratório até a apresentação do resultado final.

O QUE É UM GRÁFICO?

■ Os primórdios: Minard



O QUE É UM GRÁFICO?



<https://blog.dp6.com.br/por-que-dataviz-%C3%A9-essencial-e-como-provar-pro-seu-chefe-dataviz-intermediate-1-de-4-b3ce7e8b9917>

IMPORTÂNCIA E BENEFÍCIOS DO DATAVIZ

- A mudança de cultura das empresas para uma cultura orientada por dados, ou “data-driven” exige que as decisões estratégicas sejam tomadas baseadas em dados e de forma mais ágil e assertiva.
- Para interpretar dados de forma rápida e precisa é preciso transformar números em informações visuais.
- Com a crescente quantidade de dados gerados diariamente, a visualização de dados é uma habilidade essencial para empresas, governos e indivíduos que desejam tomar decisões informadas com base em fatos e evidências.
- A visualização de dados é uma habilidade fundamental, pois pode transformar conjuntos complexos de informações em insights claros e acionáveis;
- Dados só têm valor quando são entendidos, não basta estarem disponíveis.

IMPORTÂNCIA E BENEFÍCIOS DO DATAVIZ

- Data Visualization ou Visualização de Dados significa estudar e criar representações visuais para dados. Seu grande objetivo é construir uma ponte de comunicação entre operacional e executivo, para que as análises e relatórios de resultados se tornem decisões e ações efetivas.
- A visualização de dados é uma ferramenta poderosa para comunicar insights e informações de maneira fácil de entender. É útil para simplificar dados complexos, comunicar insights efetivamente, explorar dados em tempo real e criar painéis interativos. Sendo assim, os principais benefícios resumidamente são:
 - Simplificar dados complexos;
 - Comunicar insights efetivamente;
 - Explorar dados em tempos real;
 - Criar painéis interativos.

GRAMÁTICA DOS GRÁFICOS

- A gramática de gráficos é uma estrutura que segue uma abordagem em camadas para descrever e construir visualizações ou gráficos de maneira estruturada.
- Os gráficos são construídos em camadas, e cada uma delas adiciona um elemento visual. Quando os elementos se juntam, acontece o que chamamos de visualização dos dados. Existem 7 camadas que compõem a gramática dos dados:
 - **Dados – camada obrigatória;**
 - **Estética – camada obrigatória;**
 - **Geometria – camada obrigatória;**
 - Facets;
 - Estatística;
 - Coordenadas;
 - Temas.



<https://www.science-craft.com/2014/07/08/introducing-the-grammar-of-graphics-plotting-concept/>

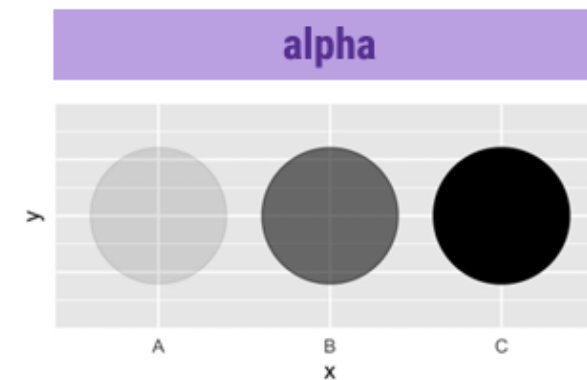
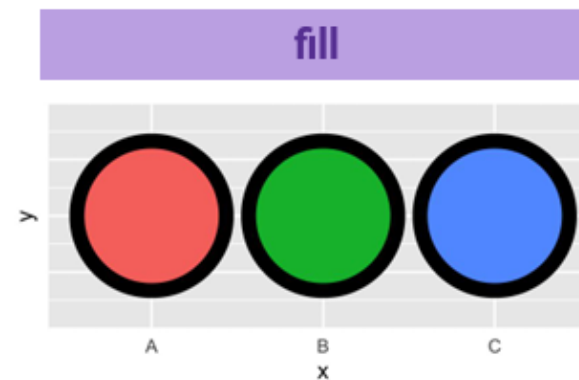
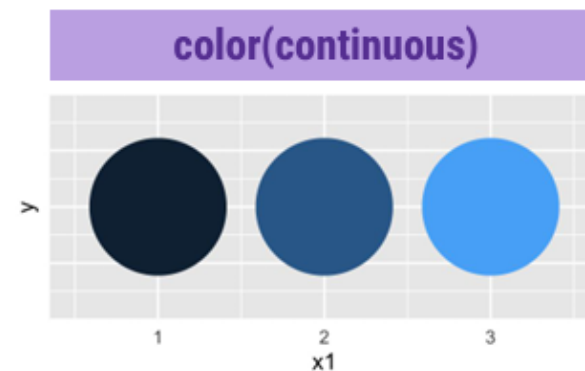
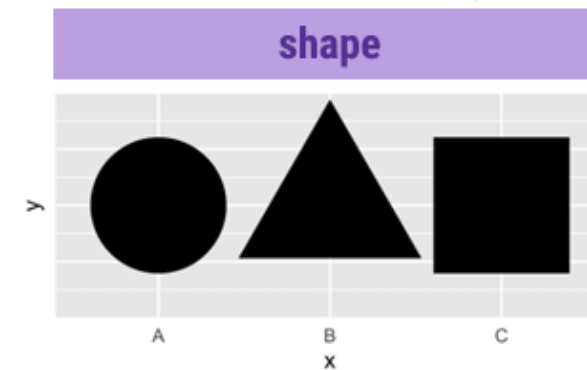
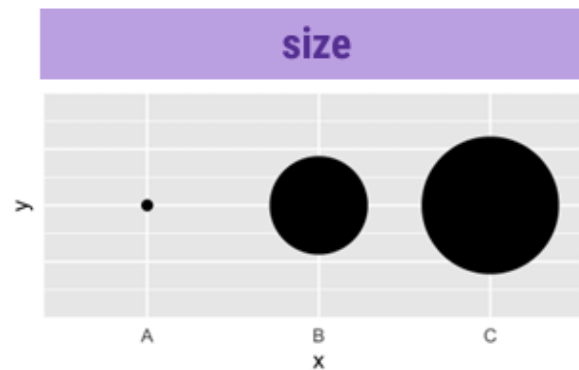
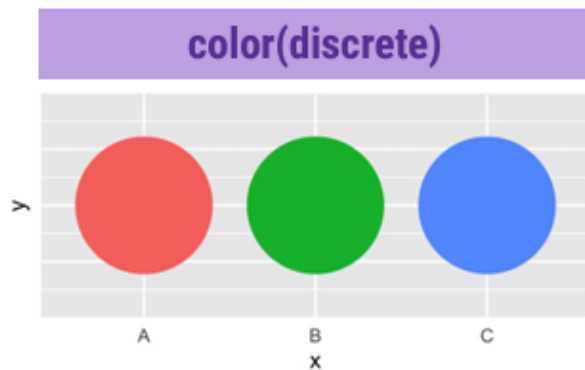
GRAMÁTICA DOS GRÁFICOS

■ Detalhando cada uma das camadas, temos:

- **Dados:** é a primeira camada e base para que as outras camadas sejam adicionadas. Nesta etapa são decididas as variáveis e dimensão que serão analisadas.
- **Estética:** nessa camada é definida as escalas dos dados mapeados na primeira etapa.
- **Geometria:** nessa camada são definidas forma, o tamanho e a posição relativa do gráfico. A representação pode ser feitas com pontos, barras, linhas, etc. Basicamente, todos os elementos visuais serão utilizados para representar os dados, devem ser inseridos na camada atual, o que inclui inclusive textos, como legendas.
- **Facets:** essa camada é usada quando precisamos criar subparcelas com base em dimensões de dados específicas. Assim, o gráfico é particionado, sendo possível evidenciar mais de uma visualização lado a lado.
- **Estatística:** caso seja necessário mostrar algumas medidas estatísticas na visualização, como medidas de tendência central, spreads, intervalos de confiança e outros, incluimos tais elementos nesta camada.
- **Coordenadas:** nessa etapa é definido em que tipo de sistema de coordenadas a visualização deverá ser feita: cartesiano ou polar.
- **Temas:** a última camada é o gráfico pronto. Toda a parte visual do gráfico e que não faz parte dos dados, como fontes, linhas, entre outros é especificada nessa camada. Portanto, evidencia a soma da aplicação de todas as camadas na construção de um gráfico, as quais devem estar em harmonia.






ASPECTOS ESTÉTICOS

- Definindo a forma como os dados serão apresentados:



ASPECTOS GEOMÉTRICOS

- Definindo o tipo de gráfico:

	Example geom	What it makes
	<code>geom_col()</code>	Bar charts
	<code>geom_text()</code>	Text
	<code>geom_point()</code>	Points
	<code>geom_boxplot()</code>	Boxplots
	<code>geom_sf()</code>	Maps

PACOTES DE GRÁFICOS NATIVOS DO R: GRAPHICS E GRDEVICES

- Os pacotes básicos de plotagem existentes no R são:
 - Graphics: contém as funções gráficas básicas – plot, hist. E boxplot
 - grDevices – contém as implementações de dispositivos gráficos como X11, png, PostScript, etc.
- Os plots são objetos construídos por meio de funções e possuem atributos configuráveis;
- Gráficos disponíveis nos pacotes básicos do R:
 - Colunas; barras; linhas; dispersão; área; bolhas; pizza; cone e superfície
- Função demo(“graphics”);
- Função par(): definir parâmetros dos gráficos;
- Possibilidade de salvar os gráficos como png e pdf.

PACOTES DE GRÁFICOS NATIVOS DO R: GRAPHICS E GRDEVICES

■ Principais parâmetros do Plot:

- col - cor do plot;
- lty - tipo de linha;
- lwd - largura de linha;
- pch - símbolo no plot;
- xlab - label do eixo x;
- ylab - label do eixo y;
- Main – título do gráfico;
- las - como os labels dos eixos são orientados;
- bg - background color;
- mfrow - número de plots por linha;
- mfcol - número de plots por coluna .

■ Funções básicas do Plot:

- plot() – scatterplots;
- lines() - adiciona linhas ao gráfico;
- points() - adiciona pontos ao gráfico;
- text() - adiciona label ao gráfico;
- title() - adiciona título ao gráfico.

PRINCIPAIS TIPOS DE GRÁFICOS

- Dispersão – plot()
 - Relação entre variáveis;
- Boxplot – boxplot()
 - Evidencia medidas descritivas;
- Histograma – hist()
 - Distribuição de frequência de uma variável;
 - Parâmetro breaks;
- Barras – barplot()
 - Tipo de gráfico mais utilizado;
 - As barras podem estar empilhadas ou lado a lado (beside);
- Pizza – pie()

GGPLOT2

- É uma alternativa ao sistema básico de gráficos do R, um pacote mais completo para construção de gráficos;
- É possível claramente ver a aplicação da construção dos gráficos camada a camada;
- É totalmente integrado com estatística. Sendo possível inclusive adicionar alguns modelos na representação gráfica, como o modelo de regressão;
- Tipos de gráficos:
 - Dispersão: `geom_point()`;
 - Boxplot: `geom_boxplot()`;
 - Histograma: `geom_histogram()`;
 - Barras: `geom_bar()`.
- Cheatsheet ggplot2: <https://rstudio.github.io/cheatsheets/html/data-visualization.html>

VANTAGENS DO GGLOT2

- É uma alternativa ao sistema básico de gráficos do R;
- É baseado na gramática dos gráficos, o que torna o processo de criação dos gráficos mais intuitivo e flexível;
- Permite adicionar camadas aos gráficos, como pontos, linhas e barras e também manipular as propriedades estéticas dos gráficos de forma independente;
- Oferece maior controle sobre a estética do gráfico, através de cores, temas e tamanhos de fontes, permitindo a criação de visualizações mais personalizadas e profissionais;
- Oferece legendas prontas, não é necessário definir explicitamente;
- Possui um sistema de formatação mais sólido;
- Possui uma ampla gama de extensões, que adicionam funcionalidades, como a criação de gráficos interativos, mapas, gráficos complexos, etc;
- Há uma grande comunidade de usuários e desenvolvedores em torno do ggplot2, proporcionando uma rica variedade de recursos de aprendizado, exemplos e suporte.