

Ciência de Dados com R

Aula 4 | Visualização de dados

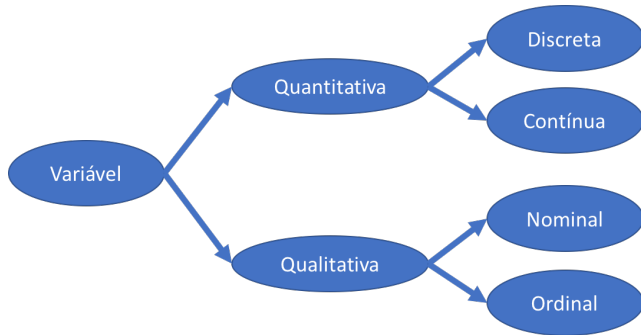
Izabel Nolau

nolau@dme.ufrj.br

Visualização de dados

Tipos de Variáveis

- As variáveis representam características dentro de uma pesquisa
- Esses valores variam de elemento para elemento
- Essas variáveis podem ter valores numéricos ou não numéricos



- São representadas por meio de números resultantes de uma contagem ou mensuração
- Essas variáveis podem ser de dois tipos:
 - *Variáveis discretas*: os valores representam uma contagem. Exemplos: Número de filhos (0,1,2,...), número de bactérias por amostra, número de copos de cerveja tomados por dia,...
 - *Variáveis contínuas*: os valores pertencem a um intervalo de números reais e representam uma mensuração. Exemplo: altura ou peso de uma pessoa, tempo (relógio) e pressão arterial,...

- Representam uma qualidade (ou atributo) de um indivíduo pesquisado, são definidas por várias categorias
- Essas variáveis podem ser de dois tipos:
 - *Variável nominal*: quando não existe nenhuma ordenação nas possíveis representações. Exemplos: sexo, cor dos olhos, fumante/não fumante,...
 - *Variável ordinal*: quando apresentam uma ordem nos seus resultados. Exemplos: escolaridade (1, 2, 3 graus), mês de observação,...

O Pacote ggplot2

- O **ggplot2** é um pacote da linguagem R construído para visualização de dados
- O **gg** é uma abreviatura de *The Grammar of Graphics*
- A gramática de gráficos descreve o método de quebrar gráficos em elementos, construindo cada um deles em uma série de camadas para controlar a representação visual

“Um gráfico estatístico é um mapeamento de dados para atributos estéticos (cor, forma e tamanho) de objetos geométricos (pontos, linhas, barras). É a combinação de todas essas camadas que forma seu gráfico estatístico.”

O Pacote ggplot2

- + é a chave para a construção de gráficos sofisticados do `ggplot2`
- Os componentes para construção de gráficos são:
 - `data`: base de dados a ser plotada
 - `aesthetics`: atributos estéticos utilizados na função `aes()`
 - `geom_`: O nome do objeto geométrico a ser utilizado
 - `stat_`: transformação estatística a ser utilizada
 - `position adjustments`: O método utilizado para ajuste de objetos sobrepostos ao gráfico
- A primeira coisa a se fazer quando se utiliza o `ggplot` é mapear os dados da seguinte maneira:

```
ggplot(data = nome.da.base.de.dados)
```

- Os dados devem ser um DataFrame adequadamente formatado (colunas são variáveis e linhas as observações)
- `ggplot2` usa os dados disponíveis no DataFrame para compor o gráfico
- Cada camada do gráfico deve ter dados associados que não precisam ser iguais!

Conjunto de Dados: mpg

- A base de dados **mpg** fala sobre a economia de combustível dos modelos de carros populares em 1999 e 2008 nos EUA
 - **manufacturer**: fabricante
 - **model**: nome do modelo
 - **displ**: cilindradas, em litros
 - **year**: ano de fabricação
 - **cyl**: número de cilindros
 - **trans**: tipo de transmissão
 - **drv**: f = tração dianteira, r = tração traseira, 4 = 4 rodas
 - **cty**: milhas por galão na cidade
 - **hwy**: milhas por galão na rodovia
 - **fl**: tipo de combustível
 - **class**: "tipo" de carro

Aesthetic Mapping

- O aesthetic mappings, definido com `aes()`, descreve como as variáveis estarão dispostas no gráfico
- Por exemplo, tem-se o mapeamento: x-position como `motor`, y-position como `mpgr`, and `color` como `tipo`, então fazemos:

```
aes(x = motor, y = mpg, color = tipo)
```

- Aesthetic mappings pode ser especificado no `ggplot()` inicial, em layers individuais ou combinando os dois

Controla o tipo de gráfico que está sendo criado, ou seja, definem qual forma geométrica será utilizada para a visualização dos dados no gráfico

- `geom_line`: para retas definidas por pares (x,y)
- `geom_abline`: para retas definidas por um intercepto e uma inclinação
- `geom_hline`: para retas horizontais
- `geom_boxplot`: para boxplots
- `geom_histogram`: para histogramas
- `geom_density`: para densidades
- `geom_area`: para áreas
- `geom_bar`: para barras

Bons estudos! 😊