

Aula 3 - Visualização de Dados

Introdução ao aprendizado de máquina - UEMA 2025

Thiago S. F .Silva

2025-12-10

Parte III - Visualização de dados

Analisar dados sem uma exploração gráfica é como ir a um encontro às escuras.

Análise visual

Analisar dados sem uma exploração gráfica é como ir a um encontro às escuras.



Analisar dados sem uma exploração gráfica é como ir a um encontro às escuras.

Nós somos seres essencialmente visuais, com uma capacidade incrível de processar imagens.

Por que gráficos?

O “Quarteto de Anscombe”

Anscombe, F.J., 1973. Graphs in Statistical Analysis. The American Statistician 27, 17–21

```
head(anscombe)
```

	x1	x2	x3	x4	y1	y2	y3	y4
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.10	8.84	7.04

Quarteto de Anscombe

```
round(lm(y1 ~ x1, data = anscombe)$coefficients,2)
```

(Intercept)	x1
3.0	0.5

```
round(lm(y2 ~ x2, data = anscombe)$coefficients,2)
```

(Intercept)	x2
3.0	0.5

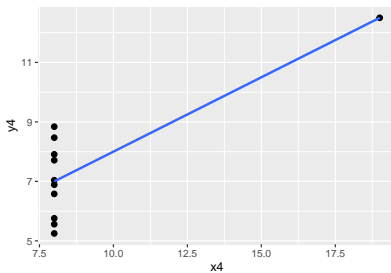
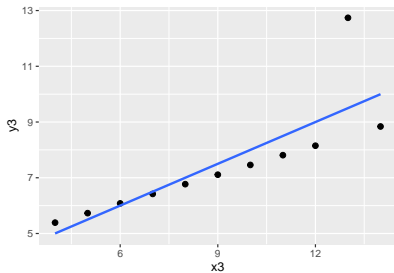
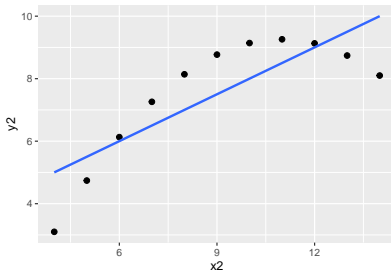
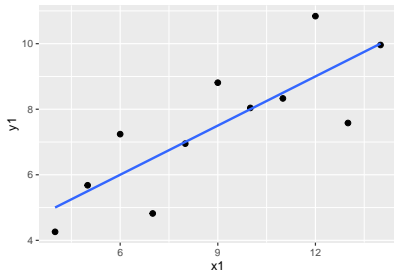
```
round(lm(y3 ~ x3, data = anscombe)$coefficients,2)
```

(Intercept)	x3
3.0	0.5

```
round(lm(y4 ~ x4, data = anscombe)$coefficients,2)
```

(Intercept)	x4
3.0	0.5

Quarteto de Anscombe



Melhor que o quarteto de Anscombe

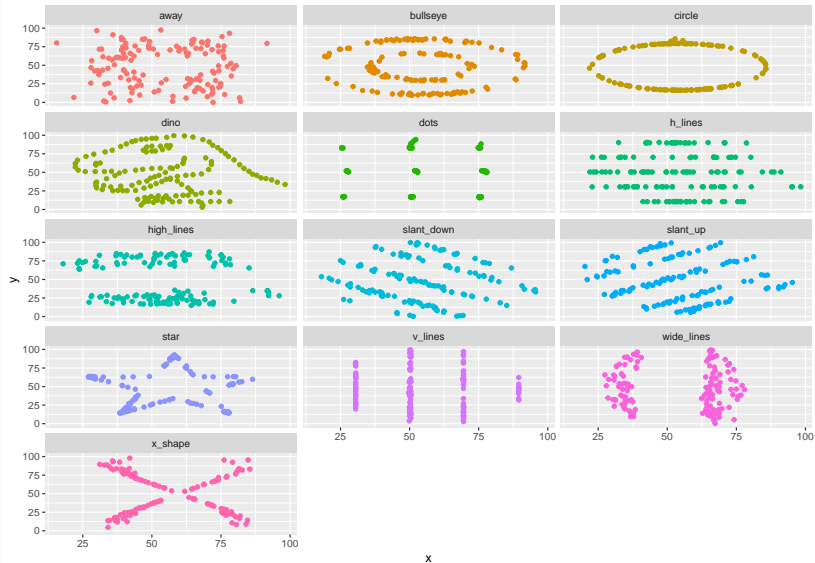
```
library(datasauRus)
library(dplyr)

datasaurus_dozen %>% group_by(dataset) %>%
  summarise(mean_x = mean(x), mean_y = mean(y), sd_x = sd(x),
            sd_y = sd(y))
```

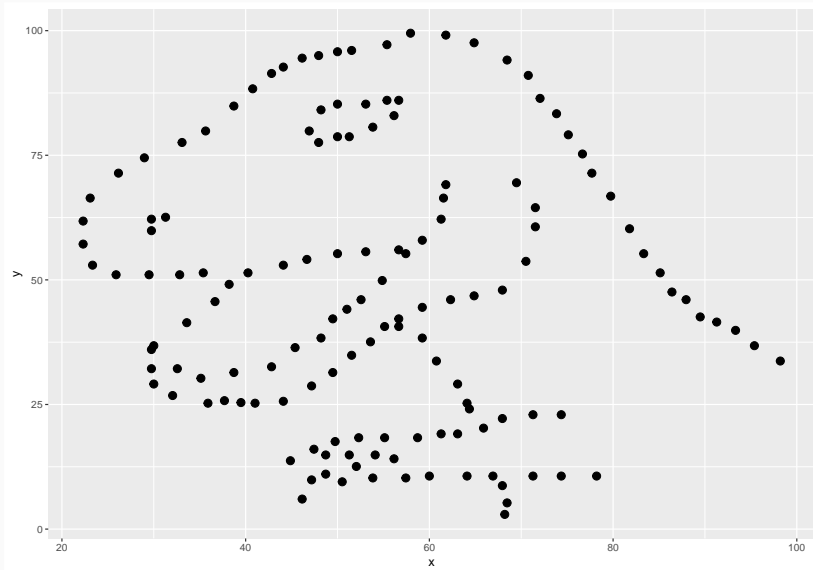
A tibble: 13 x 5

	dataset	mean_x	mean_y	sd_x	sd_y
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	away	54.3	47.8	16.8	26.9
2	bullseye	54.3	47.8	16.8	26.9
3	circle	54.3	47.8	16.8	26.9
4	dino	54.3	47.8	16.8	26.9
5	dots	54.3	47.8	16.8	26.9
6	h_lines	54.3	47.8	16.8	26.9
7	high_lines	54.3	47.8	16.8	26.9
8	slant_down	54.3	47.8	16.8	26.9
9	slant_up	54.3	47.8	16.8	26.9

Melhor que o quarteto de Anscombe

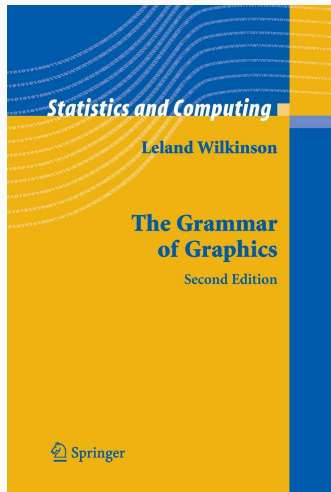


Melhor que o quarteto de Anscombe



A gramática dos gráficos

- Conceito criado por Leland Wilkinson
- Como pensar em gráficos de maneira sistemática



A gramática dos gráficos

Um gráfico é composto pelos seguintes elementos:

- **Dados (*data*):** quais dados se está plotando

A gramática dos gráficos

Um gráfico é composto pelos seguintes elementos:

- **Dados (*data*):** quais dados se está plotando
- **Estética (*aesthetics*):** como estes dados são representados (eixos, forma, - tamanho, cor)

A gramática dos gráficos

Um gráfico é composto pelos seguintes elementos:

- **Dados (*data*):** quais dados se está plotando
- **Estética (*aesthetics*):** como estes dados são representados (eixos, forma, - tamanho, cor)
- **Escala (*Scale*):** qual o intervalo dos dados a ser visualizado?

A gramática dos gráficos

Um gráfico é composto pelos seguintes elementos:

- **Dados (*data*):** quais dados se está plotando
- **Estética (*aesthetics*):** como estes dados são representados (eixos, forma, - tamanho, cor)
- **Escala (*Scale*):** qual o intervalo dos dados a ser visualizado?
- **Geometria (*geometry*):** que tipo de representação gráfica será usada?

A gramática dos gráficos

Um gráfico é composto pelos seguintes elementos:

- **Dados (*data*):** quais dados se está plotando
- **Estética (*aesthetics*):** como estes dados são representados (eixos, forma, - tamanho, cor)
- **Escala (*Scale*):** qual o intervalo dos dados a ser visualizado?
- **Geometria (*geometry*):** que tipo de representação gráfica será usada?
- **Estatísticas (*statistics*):** como os dados originais estão sendo resumidos no - gráfico?

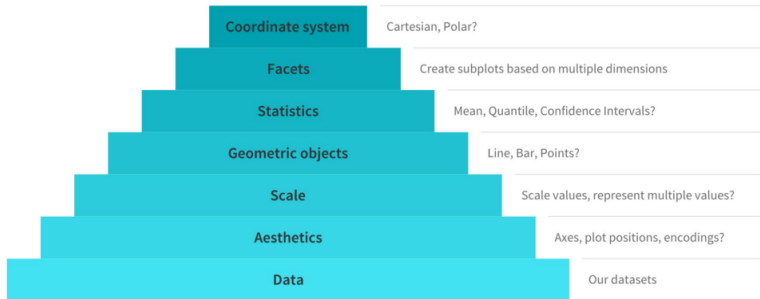
A gramática dos gráficos

Um gráfico é composto pelos seguintes elementos:

- **Dados (*data*):** quais dados se está plotando
- **Estética (*aesthetics*):** como estes dados são representados (eixos, forma, - tamanho, cor)
- **Escala (*Scale*):** qual o intervalo dos dados a ser visualizado?
- **Geometria (*geometry*):** que tipo de representação gráfica será usada?
- **Estatísticas (*statistics*):** como os dados originais estão sendo resumidos no - gráfico?
- **Facetas (*facets*):** como os dados devem ser separados em painéis?

A gramática dos gráficos

Major Components of the Grammar of Graphics



Exemplo

Dados: `iris` , Estética:

`x = Sepal.Length`, `y = Sepal.Width`, `cor = Species`, `valores = círculos` ,

Escala: `x=0:8`, `y=2:4.5` , Geometria = `pontos` , Estatística = `regressão` ,

Facetas = `Species`

