



# Visualização de dados com {ggplot2}

A gramática para construção de gráficos

Prof. Dr. **Walmes Zeviani** · UFPR

2021-11-02

# **Recursos gráficos do R**

# Visão geral

O landscape de recursos para visualização de dados no R pode ser representado por uma divisão em 4 territórios.

1. O pacote `graphics` e derivados.
2. O pacote `lattice` e derivados.
3. O pacote `ggplot2` e derivados.
4. Pacotes para gráficos interativos.

## O pacote `graphics`

- Contém os recursos mais primitivos: com funções de alto e baixo nível.
- Pouco suporte para mapeamento em variáveis visuais retinais.
- Vários pacotes que complementam suas funcionalidades: `plotrix` e `gplots`.
- Usado em métodos gráficos de saídas de análises: dendrogramas, biplots, etc.

## O pacote `lattice`

- Desenvolvido por Deepayan Sarkar.
- Plotagem multi-painel e mapeamento em variáveis visuais retinais.
- Já vem com a instalação básica do R.
- Também é utilizado na implementação de métodos gráficos.

## O pacote `ggplot2`

- Desenvolvido por Hadley Wickham.
- Baseado na *Grammar of Graphics*.
- Plotagem multi-painel e mapeamento em variáveis visuais retinais.
- A importância da `ggplot2` está no modelo mental mais claro.

# Pacotes para recursos interativos

A visualização interativa é voltada para exibição na WEB. Alguns dos pacotes para isso são estes:

- ▶ `plotly`.
- ▶ `highcharter`.
- ▶ `googleVis`.
- ▶ `echarts4r`.
- ▶ `leaflet`.
- ▶ `iplots`.
- ▶ `rgl`.
- ▶ `animation`.

Para mais detalhes sobre os recursos gráficos, siga esse link:

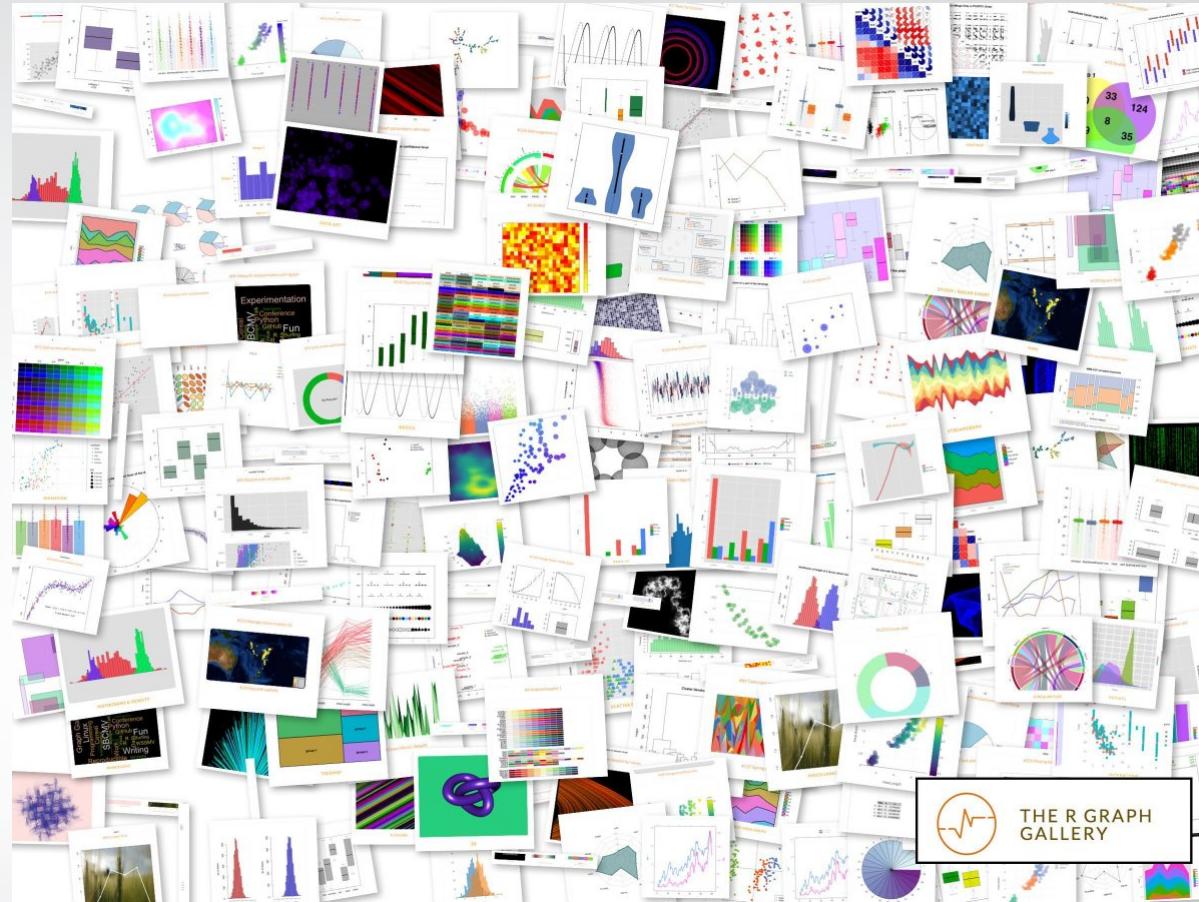
[https://www.stat.ubc.ca/~jenny/STAT545A/block90\\_baseLatticeGgplot2.html](https://www.stat.ubc.ca/~jenny/STAT545A/block90_baseLatticeGgplot2.html).

Para uma descrição completa da comparação entre `lattice` e `ggplot2`, siga esse link:

<https://learnr.wordpress.com/2009/08/26/ggplot2-version-of-figures-in-lattice-multivariate-data-visualization-with-r-final-part/>.

Confira no **R Graph Gallery** a variedade de gráficos confeccionados com o R.

# The R Graph Gallery



Gráficos da galeria de gráficos do R. Fonte: <https://pbs.twimg.com/media/DeUD4fKW0AAcyqO.jpg>.

# Galerias

## Gráficos

- ▶ [https://www.r-graph-gallery.com/portfolio/ggplot2-package/.](https://www.r-graph-gallery.com/portfolio/ggplot2-package/)
- ▶ [http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html.](http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html)
- ▶ [https://r4stats.com/examples/graphics-ggplot2/.](https://r4stats.com/examples/graphics-ggplot2/)
- ▶ [http://girke.bioinformatics.ucr.edu/GEN242/pages/mydoc/Rgraphics.html.](http://girke.bioinformatics.ucr.edu/GEN242/pages/mydoc/Rgraphics.html)

## Mapas

- ▶ [https://rstudio-pubs-static.s3.amazonaws.com/176768\\_ec7fb4801e3a4772886d61e65885fbdd.html.](https://rstudio-pubs-static.s3.amazonaws.com/176768_ec7fb4801e3a4772886d61e65885fbdd.html)
- ▶ [http://girke.bioinformatics.ucr.edu/GEN242/pages/mydoc/Rgraphics.html.](http://girke.bioinformatics.ucr.edu/GEN242/pages/mydoc/Rgraphics.html)
- ▶ [https://cran.r-project.org/web/packages/geobr/vignettes/intro\\_to\\_geobr.html.](https://cran.r-project.org/web/packages/geobr/vignettes/intro_to_geobr.html)

# Alguns tutoriais em português

- ▶ <https://rpubs.com/mnunes/ggplot2>.
- ▶ <https://analisereal.com/2015/09/19/introducao-ao-ggplot2/>.
- ▶ <https://timogrossenbacher.ch/2016/12/beautiful-thematic-maps-with-ggplot2-only/>.
- ▶ <http://recologia.com.br/tag/graficos/>.
- ▶ [http://rstudio-pubs-static.s3.amazonaws.com/24563\\_3b7b0a6414824e3b91769a95309380f1.html](http://rstudio-pubs-static.s3.amazonaws.com/24563_3b7b0a6414824e3b91769a95309380f1.html).
- ▶ <http://eduardogutierrezes.com/inteligencia-geografica-gerando-mapas-em-r/>.
- ▶ <https://pt.stackoverflow.com/questions/332053/r-mapa-de-cidades-brasileiras>.

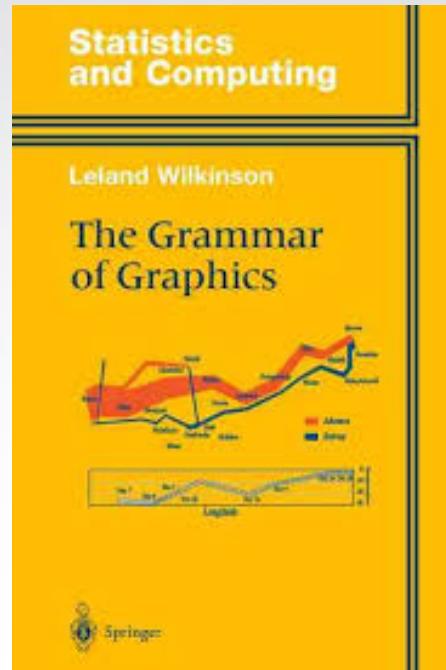
# Um overview do {ggplot2}

A implementação da gramática dos gráficos

# **Cartão de referência de visualização de dados com o {ggplot2}**

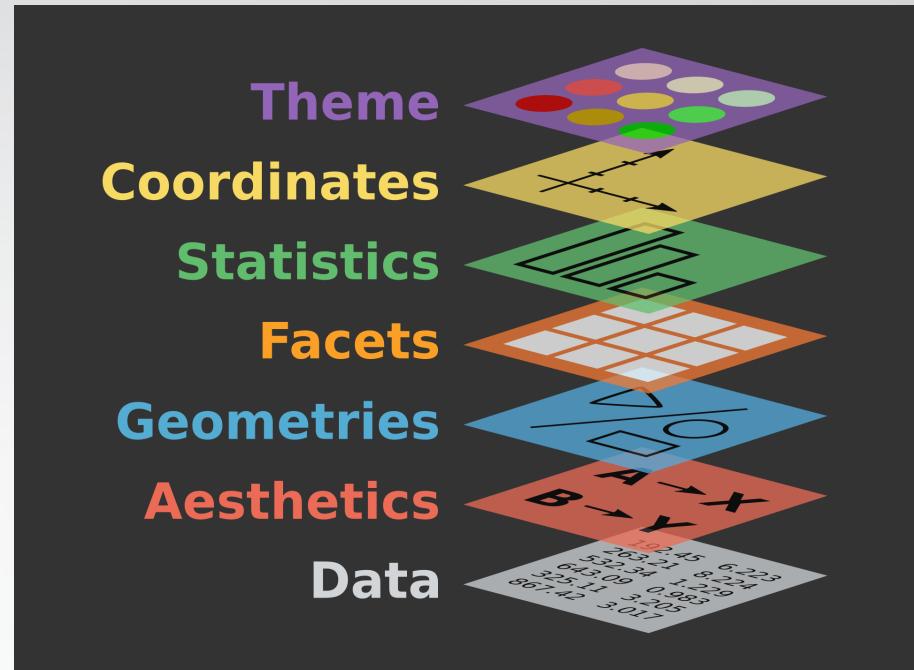
Clique no texto para abrir o arquivo

# Gramática dos gráficos



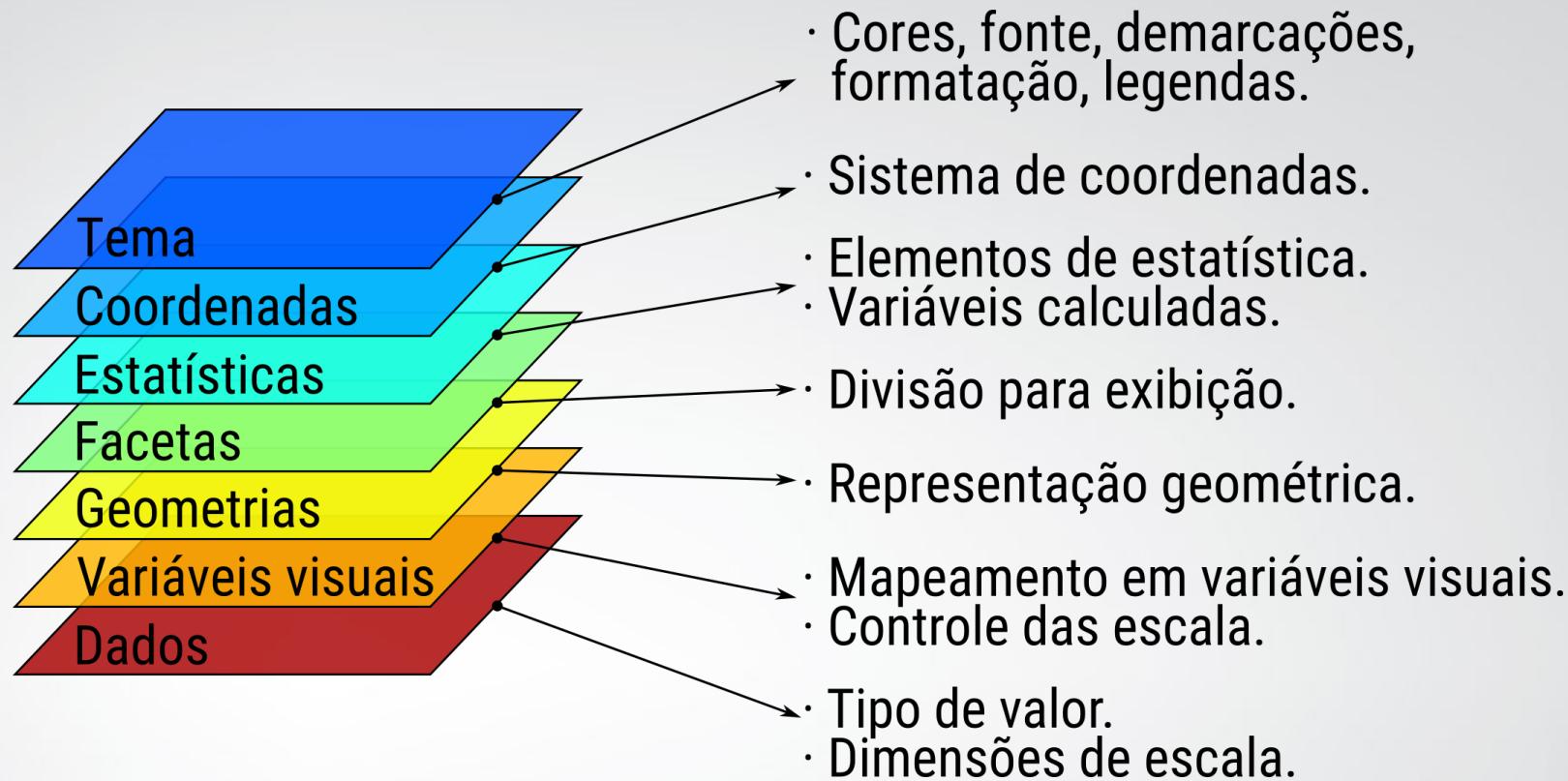
Leland Wilkinson.

A capa do livro.



As camadas da gramática de gráficos que são nos gráficos.

# Gramática dos gráficos

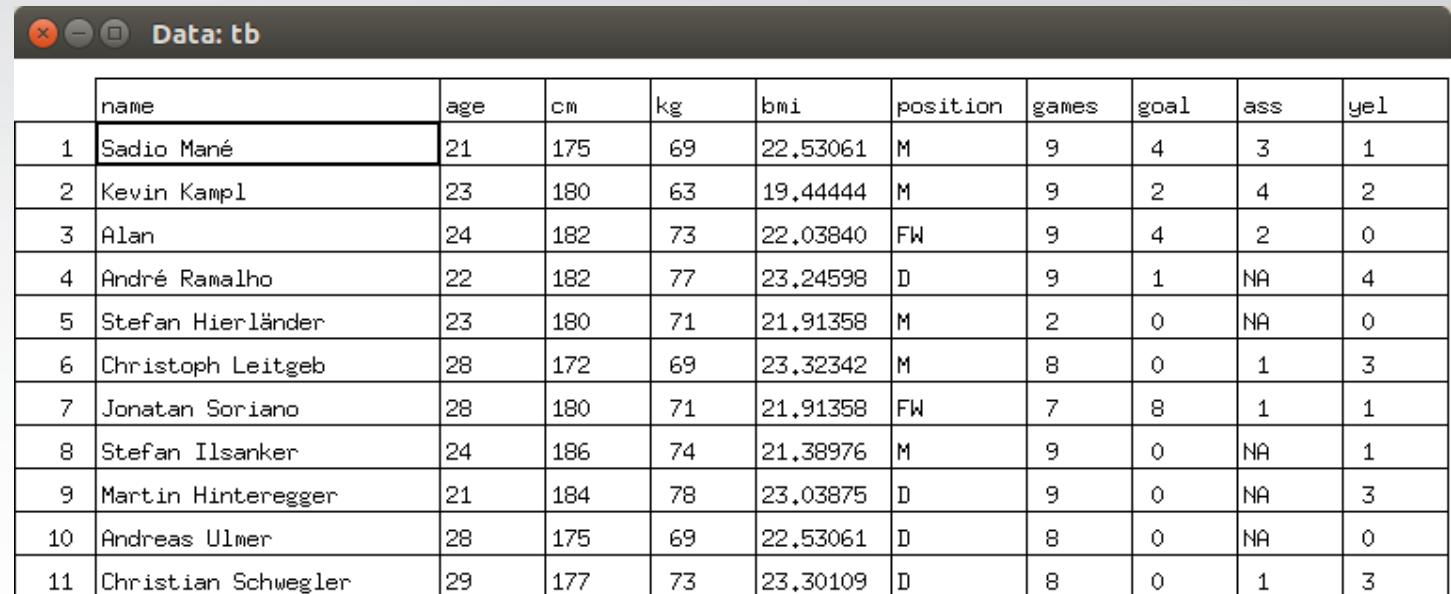


As camadas na gramática dos gráficos.

# Camada 1: dados

Deve se estar atento ao tipo de valor/objeto.

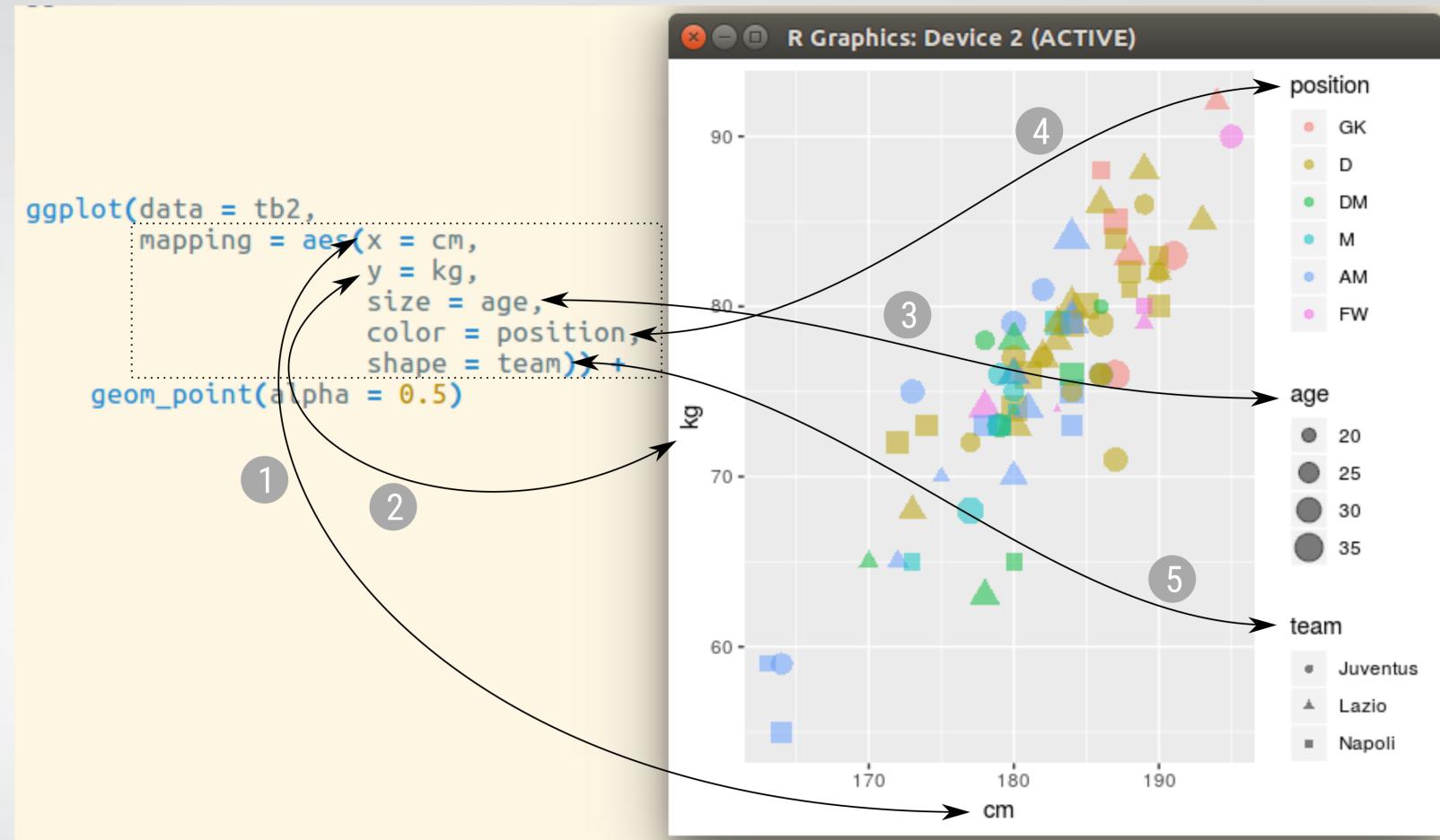
- ▶ **Quantitativa:** variável numérica discreta ou contínua.
- ▶ **Qualitativa:** variável nominal ou ordinal.
- ▶ **Cronológica:** variável de data ou data-tempo.
- ▶ **Geográfica:** objeto como polígonos, por exemplo.



|    | name                | age | cm  | kg | bmi      | position | games | goal | ass | yel |
|----|---------------------|-----|-----|----|----------|----------|-------|------|-----|-----|
| 1  | Sadio Mané          | 21  | 175 | 69 | 22.53061 | M        | 9     | 4    | 3   | 1   |
| 2  | Kevin Kampl         | 23  | 180 | 63 | 19.44444 | M        | 9     | 2    | 4   | 2   |
| 3  | Alan                | 24  | 182 | 73 | 22.03840 | FW       | 9     | 4    | 2   | 0   |
| 4  | André Ramalho       | 22  | 182 | 77 | 23.24598 | D        | 9     | 1    | NA  | 4   |
| 5  | Stefan Hierländer   | 23  | 180 | 71 | 21.91358 | M        | 2     | 0    | NA  | 0   |
| 6  | Christoph Leitgeb   | 28  | 172 | 69 | 23.32342 | M        | 8     | 0    | 1   | 3   |
| 7  | Jonatan Soriano     | 28  | 180 | 71 | 21.91358 | FW       | 7     | 8    | 1   | 1   |
| 8  | Stefan Ilsanker     | 24  | 186 | 74 | 21.38976 | M        | 9     | 0    | NA  | 1   |
| 9  | Martin Hinteregger  | 21  | 184 | 78 | 23.03875 | D        | 9     | 0    | NA  | 3   |
| 10 | Andreas Ulmer       | 28  | 175 | 69 | 22.53061 | D        | 8     | 0    | NA  | 0   |
| 11 | Christian Schwegler | 29  | 177 | 73 | 23.30109 | D        | 8     | 0    | 1   | 3   |

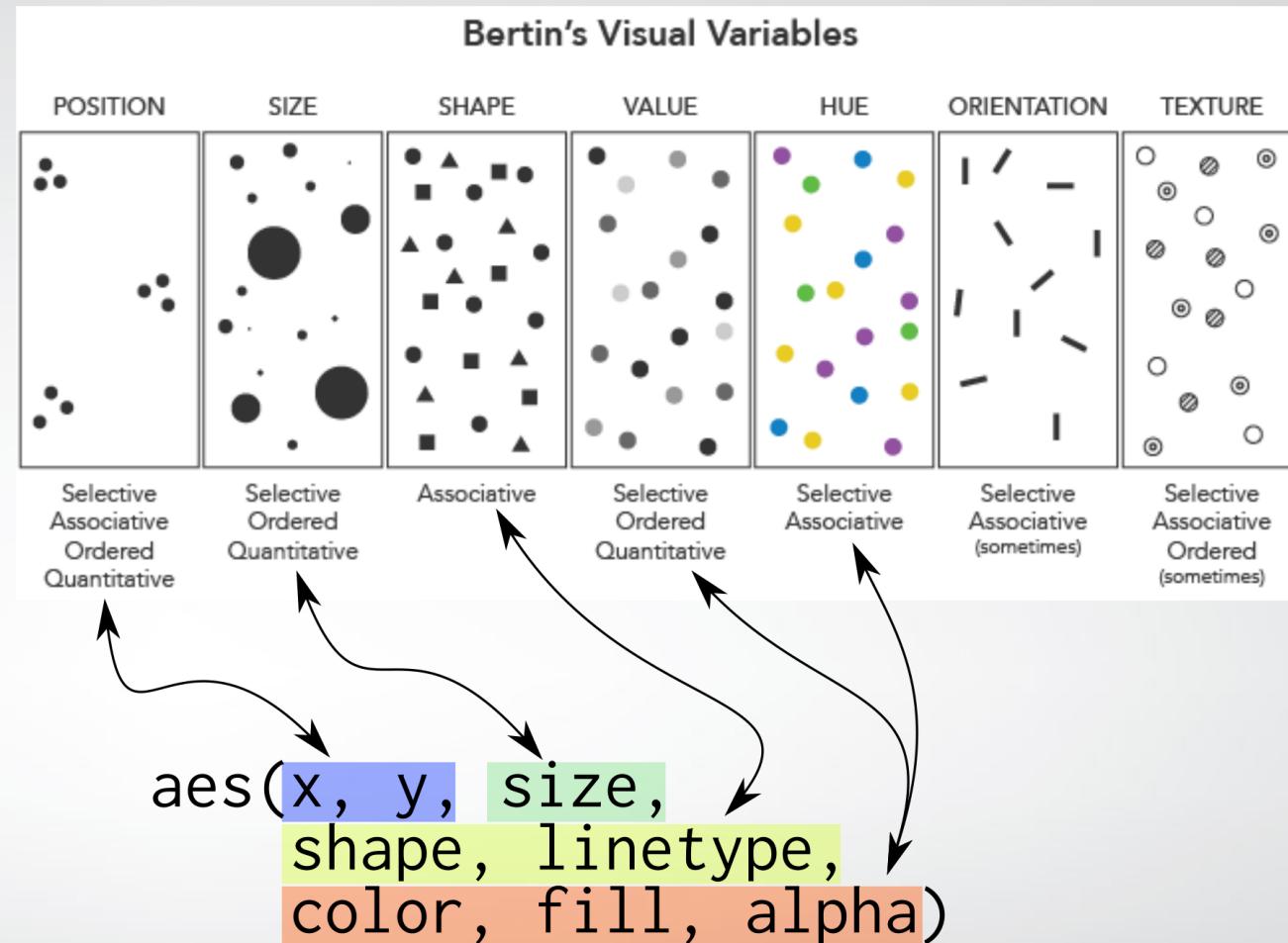
A camada dos dados.

## Camada 2: mapeamento em variáveis visuais



A camada de mapeamento dos valores em variáveis visuais.

## Camada 2: mapeamento em variáveis visuais



Variáveis visuais disponíveis no {ggplot2}.

## Camada 2: mapeamento em variáveis visuais

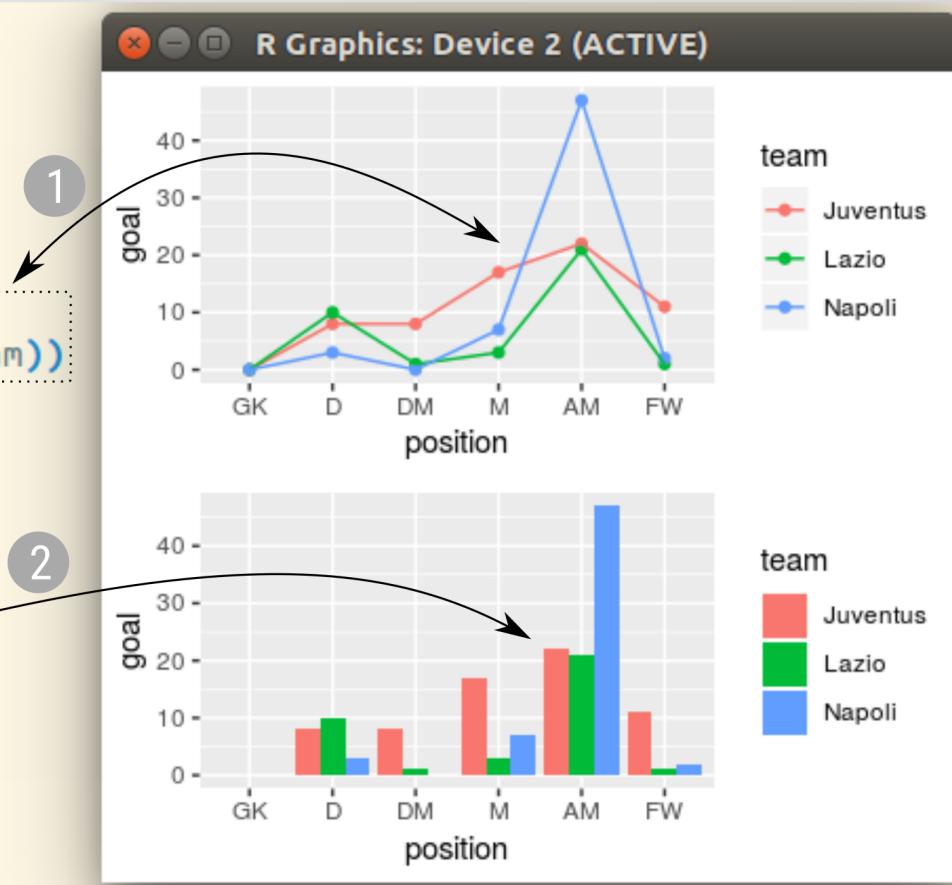
Funções para controle de escala:

```
scale_alpha_:
  binned, continuous, date, datetime, discrete, identity, manual, ordinal
scale_colour_:
  binned, brewer, continuous, date, datetime, discrete, distiller, fermenter, gradient,
  gradient2, gradientn, grey, hue, identity, manual, ordinal, steps, steps2, stepsn
scale_continuous_:
  identity
scale_discrete_:
  identity, manual
scale_fill_:
  binned, brewer, continuous, date, datetime, discrete, distiller, fermenter, gradient,
  gradient2, gradientn, grey, hue, identity, manual, ordinal, steps, steps2, stepsn
scale_linetype_:
  binned, continuous, discrete, identity, manual
scale_shape_:
  binned, continuous, discrete, identity, manual, ordinal
scale_size_:
  area, binned, continuous, date, datetime, discrete, identity, manual, ordinal
scale_x_:
  binned, continuous, date, datetime, discrete, log10, reverse, sqrt, time
scale_y_:
  binned, continuous, date, datetime, discrete, log10, reverse, sqrt, time
```

# Camada 3: geometrias

```
ggplot(data = tb3,  
       mapping = aes(x = position,  
                      y = goal,  
                      color = team)) +  
  geom_point() +  
  geom_line(mapping = aes(group = team))
```

```
ggplot(data = tb3,  
       mapping = aes(x = position,  
                      y = goal,  
                      fill = team)) +  
  geom_col(position = "dodge")
```



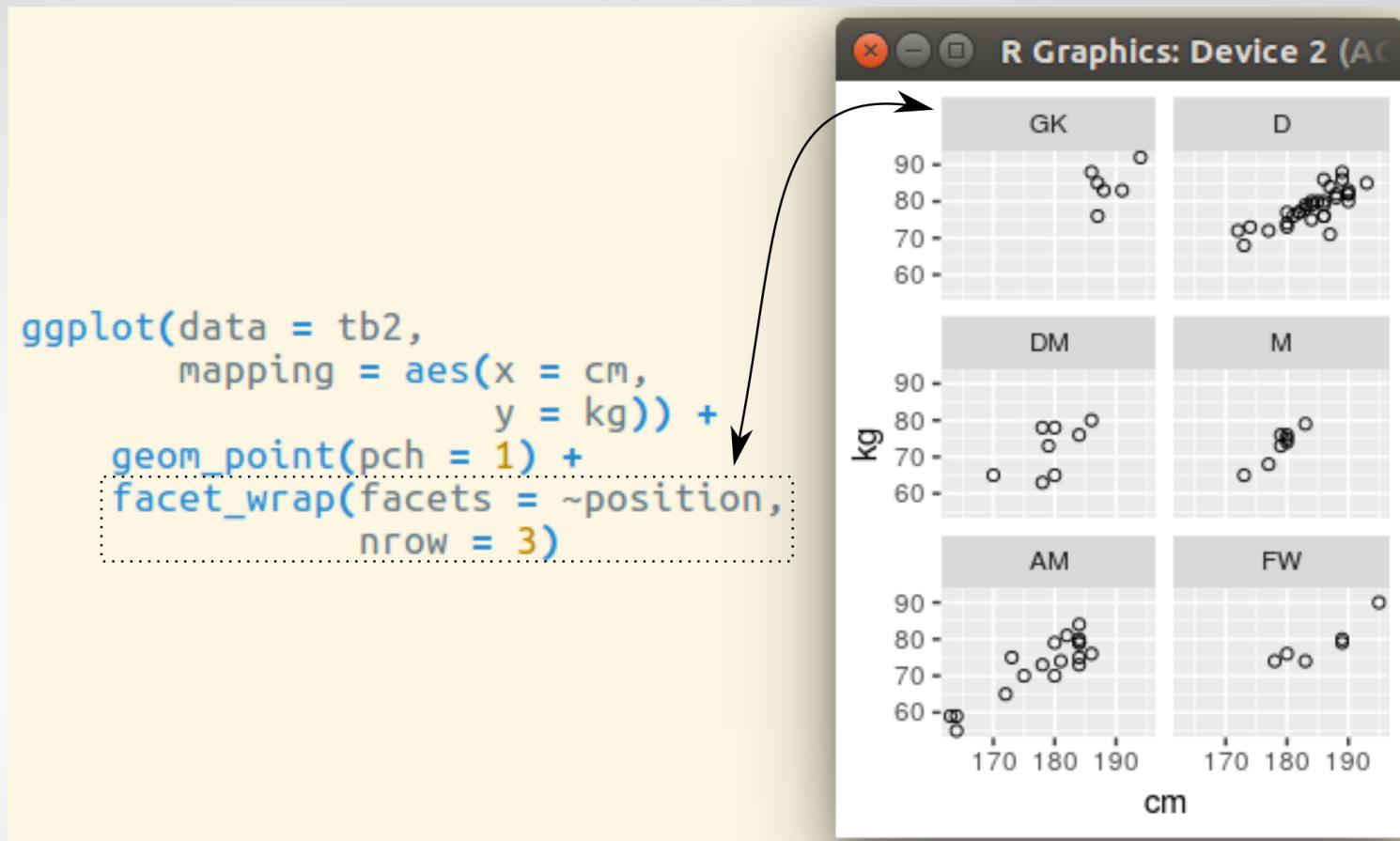
A camada de elementos geométricos.

# Camada 3: geometrias

Funções disponíveis para a camada de geometria:

|                        |                 |               |
|------------------------|-----------------|---------------|
| geom_abline            | geom_dotplot    | geom_qq_line  |
| geom_area              | geom_errorbar   | geom_quantile |
| geom_bar               | geom_errorbarh  | geom_raster   |
| geom_bin_2d            | geom_freqpoly   | geom_rect     |
| geom_bin2d             | geom_function   | geom_ribbon   |
| geom_blank             | geom_hex        | geom_rug      |
| geom_boxplot           | geom_histogram  | geom_segment  |
| geom_col               | geom_hline      | geom_sf       |
| geom_contour           | geom_jitter     | geom_sf_label |
| geom_contour_filled    | geom_label      | geom_sf_text  |
| geom_count             | geom_line       | geom_smooth   |
| geom_crossbar          | geom_linerange  | geom_spoke    |
| geom_curve             | geom_map        | geom_step     |
| geom_density           | geom_path       | geom_text     |
| geom_density_2d        | geom_point      | geom_tile     |
| geom_density_2d_filled | geom_pointrange | geom_violin   |
| geom_density2d         | geom_polygon    | geom_vline    |
| geom_density2d_filled  | geom_qq         |               |

## Camada 4: divisão em facetas



A camada da divisão em facetas.

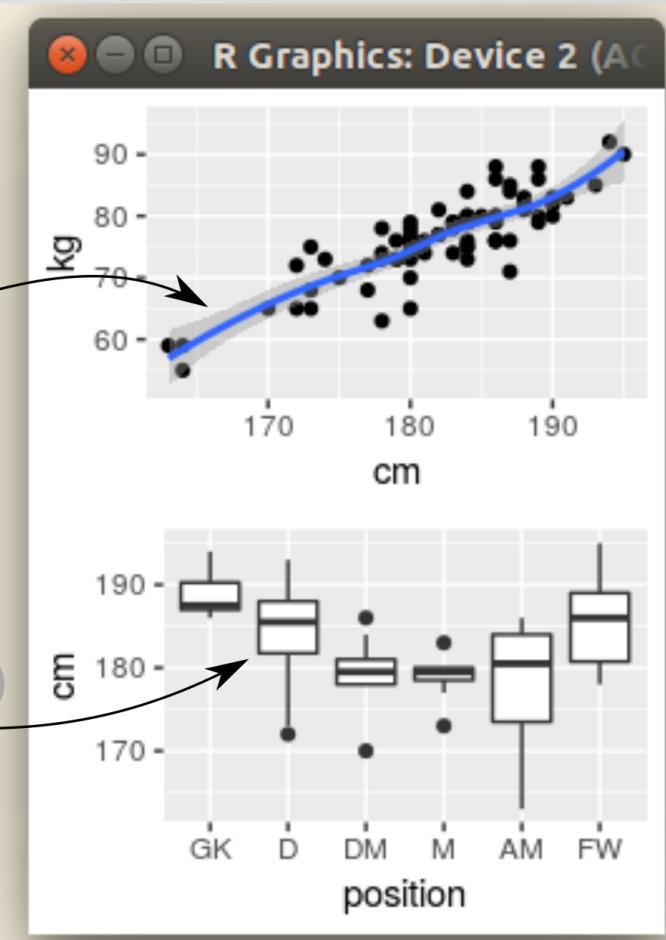
## Camada 4: divisão em facetas

Funções disponíveis para divisão em facetas:

```
facet_grid  
facet_null  
facet_wrap
```

# Camada 5: estatística

```
ggplot(data = tb2,  
       mapping = aes(x = cm,  
                      y = kg)) +  
  geom_point() +  
  geom_smooth()  
  
ggplot(data = tb2,  
       mapping = aes(x = position,  
                      y = cm)) +  
  geom_boxplot()
```



A camada de elementos de estatística.

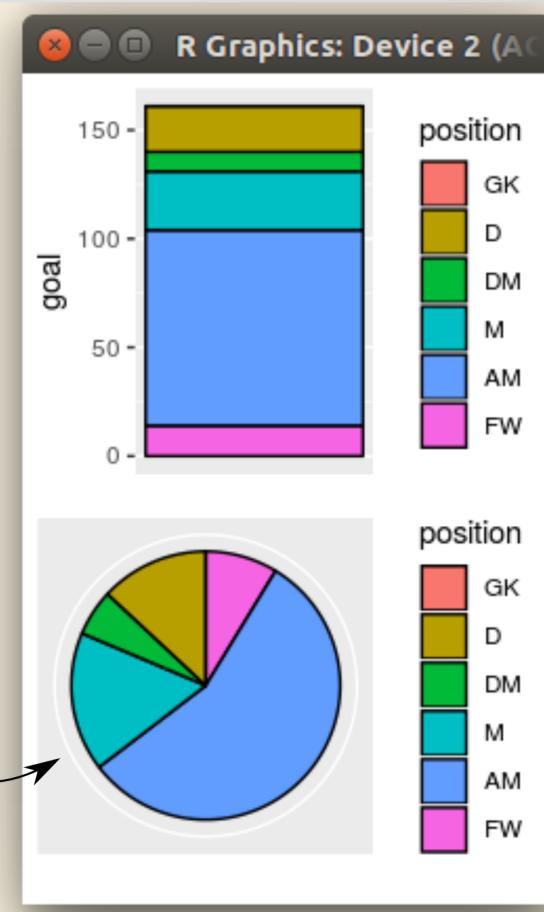
# Camada 5: estatística

Funções disponíveis para a camada estatística:

|                     |                        |                     |
|---------------------|------------------------|---------------------|
| stat_bin            | stat_density_2d_filled | stat_sf_coordinates |
| stat_bin_2d         | stat_density2d         | stat_smooth         |
| stat_bin_hex        | stat_density2d_filled  | stat_spoke          |
| stat_bin2d          | stat_ecdf              | stat_sum            |
| stat_binhex         | stat_ellipse           | stat_summary        |
| stat_boxplot        | stat_function          | stat_summary_2d     |
| stat_contour        | stat_identity          | stat_summary_bin    |
| stat_contour_filled | stat_qq                | stat_summary_hex    |
| stat_count          | stat_qq_line           | stat_summary2d      |
| stat_density        | stat_quantile          | stat_unique         |
| stat_density_2d     | stat_sf                | stat_ydensity       |

# Camada 6: coordenadas

```
ggplot(data = tb4,  
       mapping = aes(x = 0,  
                      y = goal,  
                      fill = position)) +  
  geom_col(color = "black")  
  
ggplot(data = tb4,  
       mapping = aes(x = 0,  
                      y = goal,  
                      fill = position)) +  
  geom_col(color = "black") +  
  coord_polar(theta = "y")
```



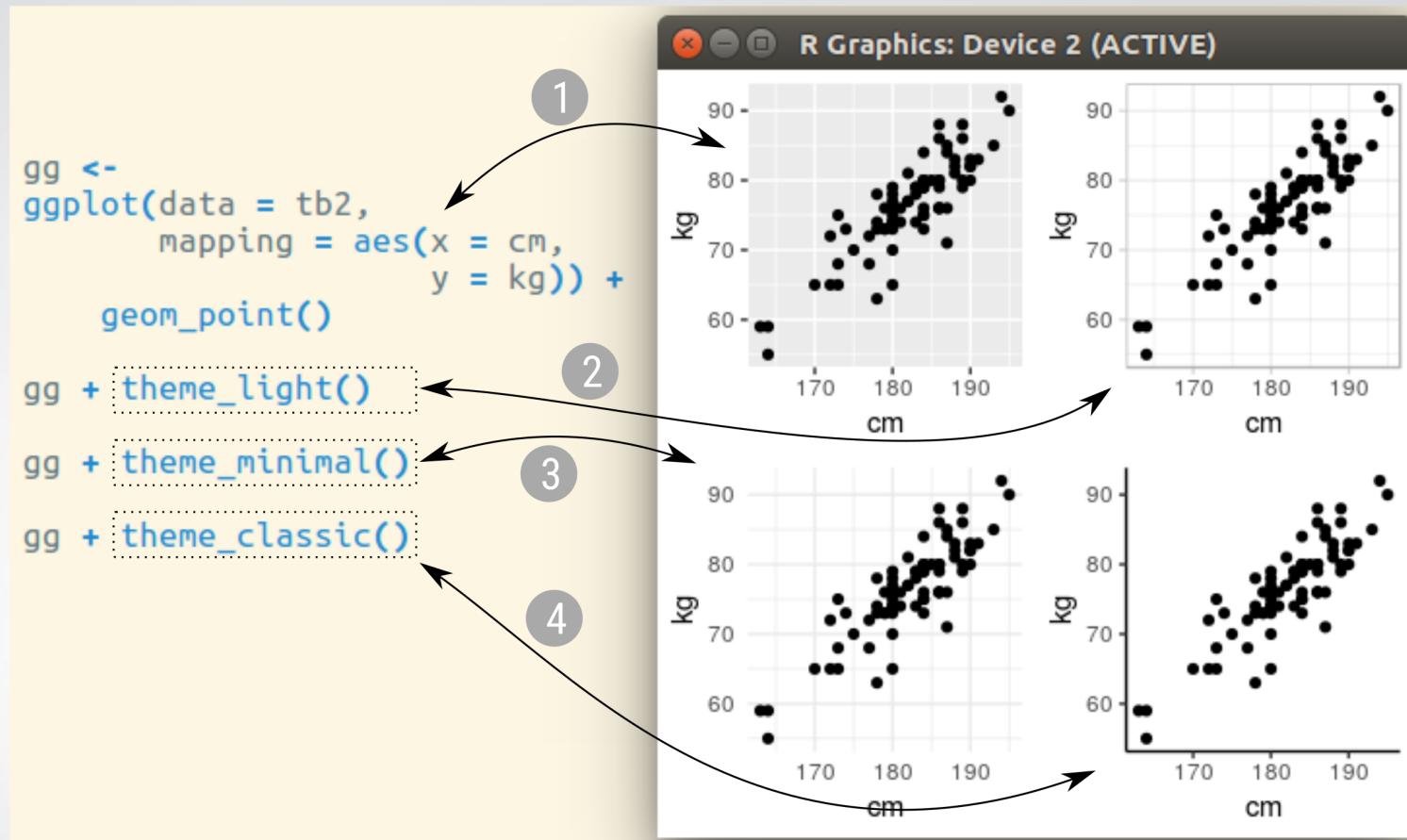
A camada do sistema de coordenadas.

## Camada 6: coordenadas

Funções disponíveis para a camada de coordenadas:

```
coord_cartesian  
coord_equal  
coord_fixed  
coord_flip  
coord_map  
coord_munch  
coord_polar  
coord_quickmap  
coord_sf  
coord_trans
```

# Camada 7: tema



A camada de tema.

# Camada 7: tema

Funções disponíveis para a camada de tema:

```
element_blank  
element_grob  
element_line  
element_rect  
element_render  
element_text  
theme_bw  
theme_classic  
theme_dark  
theme_get  
theme_gray  
theme_grey  
theme_light  
theme_linedraw  
theme_minimal  
theme_replace  
theme_set  
theme_test  
theme_update  
theme_void
```

# Considerações finais

- ▶ A análise exploratória de dados está presente em todos os projetos de ciência de dados.
- ▶ Ela serve para tirar impressões iniciais dos dados como a distribuição das variáveis, a relação entre elas, problemas com valores atípicos, etc.
- ▶ Uma boa análise exploratória subsidia o levantamento de hipóteses, as decisões de negócio, a engenharia de características e a construção e validação de modelos.
- ▶ Dominar ferramentas para uma análise exploratória ágil é importante para celeridade dos projetos e ganho de conhecimento.
- ▶ O R é extraordinário em termos de visualização com vários pacotes e frameworks para visualização de dados.
- ▶ Dispõe ainda de inúmeros recursos para a visualização interativa.