

Visualização de dados com plotnine e ggplot no Python

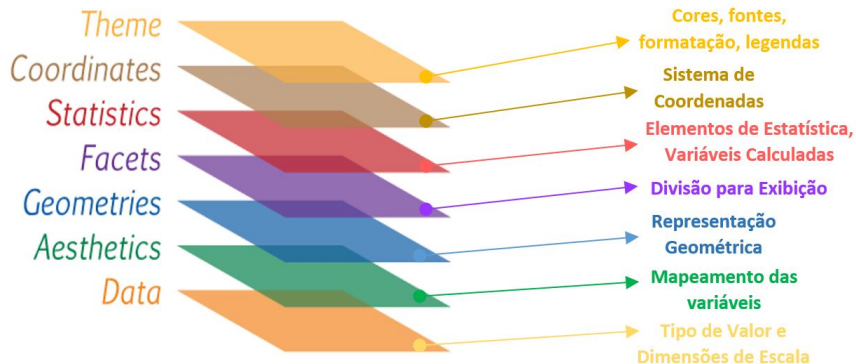
João Matheus S. K. T. Hneda

ggplot2

- Documentação: <https://ggplot2.tidyverse.org/>
- Pacote desenvolvido para a linguagem de programação estatística R.
- Pacote utilizado para a geração de gráficos.
- Desenvolvido por **Hadley Wickham**
- Implementação baseada no **The Grammar of Graphics**.
- Os pacotes **ggplot** e **plotnine** foram desenvolvidos para o Python tendo o ggplot2 como base.

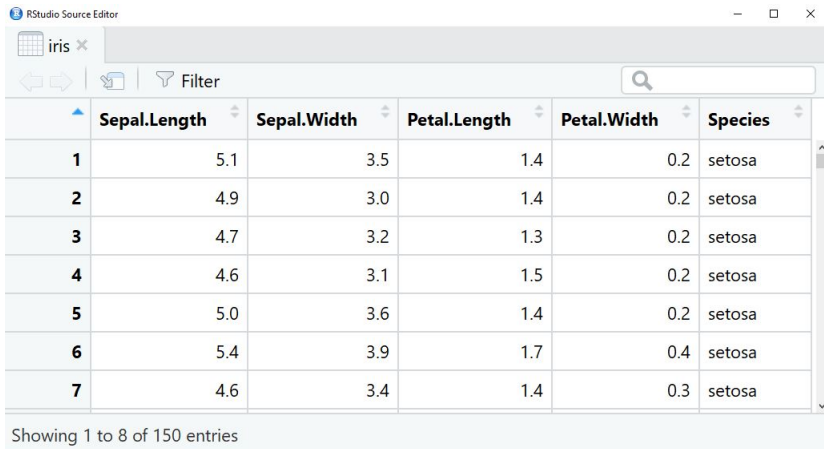


Gramática dos Gráficos



Camada 1: Dados (Data)

- Essa é a 1ª camada na construção de um gráfico. É nela que decidimos quais são os dados que iremos analisar.



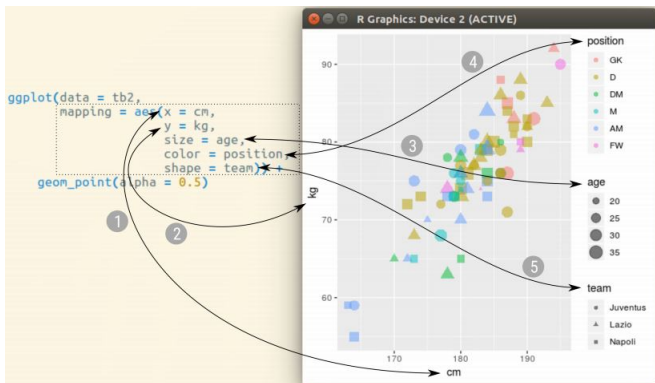
The screenshot shows the RStudio Source Editor window with the 'iris' dataset loaded. The table has 6 columns: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. The first 7 rows are visible, all belonging to the 'setosa' species. The status bar at the bottom indicates 'Showing 1 to 8 of 150 entries'.

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa

Showing 1 to 8 of 150 entries

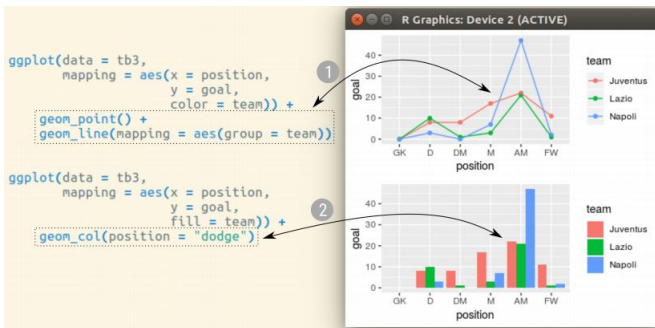
Camada 2: Mapeamento em Variáveis Visuais (Aesthetics)

- Essa é a 2ª camada na construção de um gráfico. É nela que começamos a mexer na estrutura do gráfico.
- É durante essa camada que fazemos o mapeamento das variáveis visuais.



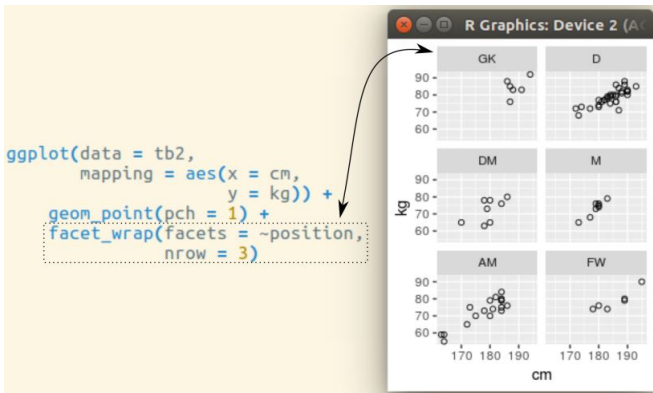
Camada 3: Geometria (Geometries)

- Essa é a 3ª camada na construção de um gráfico. É nela que definimos a forma, o tamanho e a posição relativa de um gráfico.
- É nesse momento que o gráfico começa a tomar forma, e elementos como, pontos, linhas, intervalos, textos e legendas são inseridos dentro do gráfico.



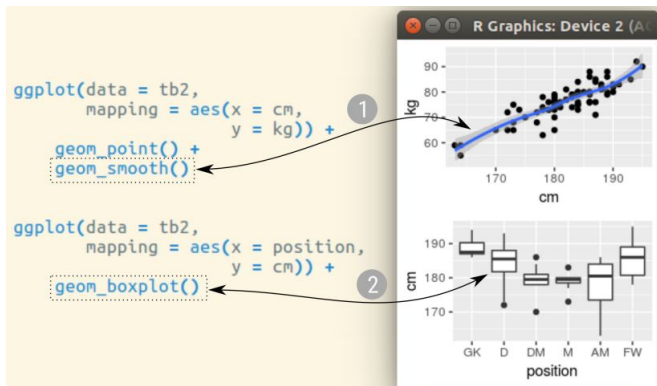
Camada 4: Divisão em Facetas (Facets)

- Essa é a 4ª camada na construção de um gráfico. Ela é utilizada quando você precisa visualizar muitos gráficos para fazer comparações ou tem muita informação no gráfico.



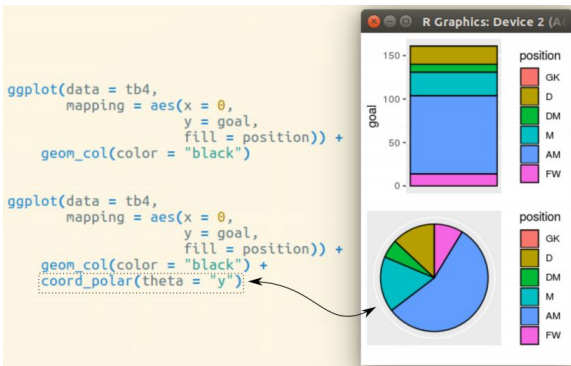
Camada 5: Estatística (Statistics)

- Essa é a 5ª camada na construção de um gráfico. Ela é utilizada quando você precisa adicionar cálculos estatísticos nos gráficos ou ajustes de modelo como a regressão.



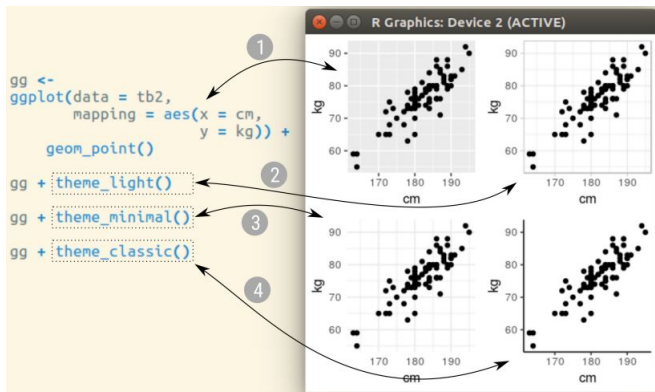
Camada 6: Coordenadas (Coordinates)

- Essa é a 6ª camada na construção de um gráfico. Ela é utilizada quando precisamos passar as informações da área na qual o gráfico será construído, como exemplo é possível utilizar essa camada quando queremos transformar as coordenadas em coordenadas polares.



Camada 7: Temas (Theme)

- Essa é a 7ª camada na construção de um gráfico. Ela é utilizada quando já fizemos todas as adaptações e modificações das camadas anteriores e queremos mostrar o gráfico para alguém.



Referências

- Livro The Grammar of Graphics
- Leland Wilkinson
- Hadley Wickham
- A Layered Grammar of Graphics
- ggplot2 (r)
- plotnine (python)
- ggplot (python)
- Visualização de dados com ggplot2 - Prof. Walmes Zeviani

Tá na hora de praticar!

Vamos ao Python!