

R ile Veri Görselleştirme

Zeynep Sueda Tuna
Bilgisayar Mühendisliği, 3. Sınıf
21360859005
04.04.2024

İçindekiler

01 - 'Graphics reveal data'

02 - Anscombe's quartet

03 - Birkaç Basit Kural

04 - R

05 - R ve Python: hangisi?

06 - Tidyverse ve ggplot

07 - Uygulama

08 - Kaynaklar

O1 - 'Graphics reveal data'

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Meşhur istatistikçi Edward Tufte, 'The Visual Display of Quantitative Information' kitabına Anscombe Quartet'ini göstererek başlıyor ve görselleştirmenin neden gerekli olduğunu gösteriyor.

01 - 'Graphics reveal data'

Bir öneki sayfadaki veri setlerinde istatistik hesaplamaların sonuçları:

X değerlerinin ortalaması= 9

Y değerlerinin ortalaması = 7.50

X değerlerinin varyansı= 11

Y değerlerinin varyansı = 4.12

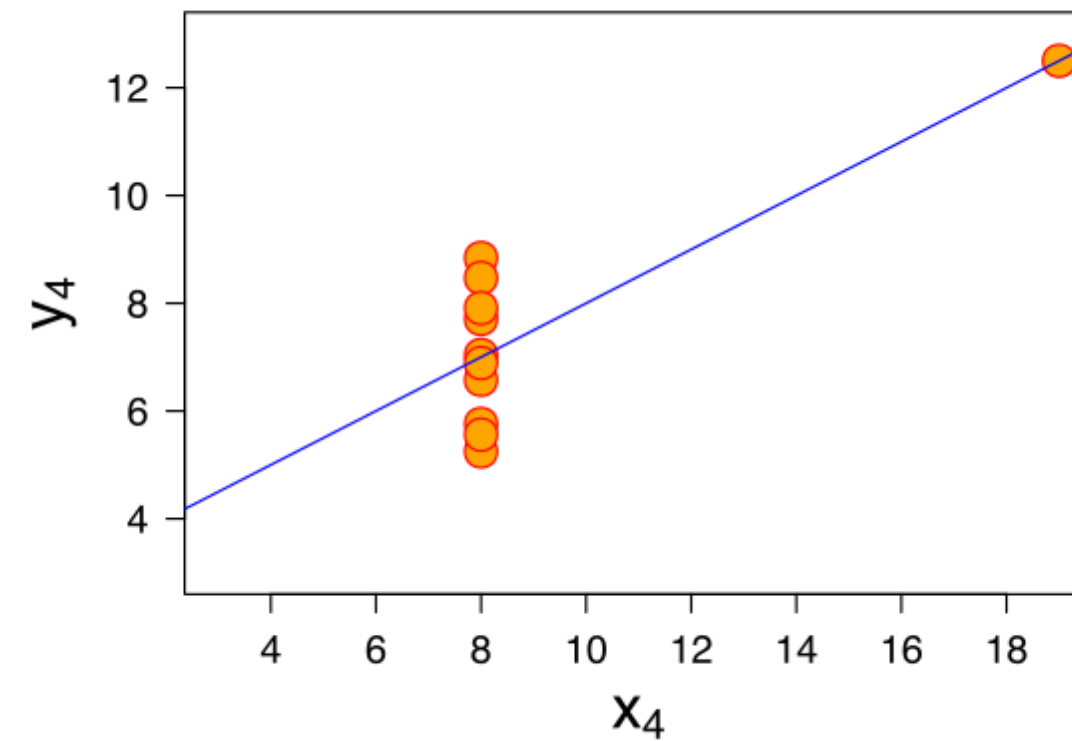
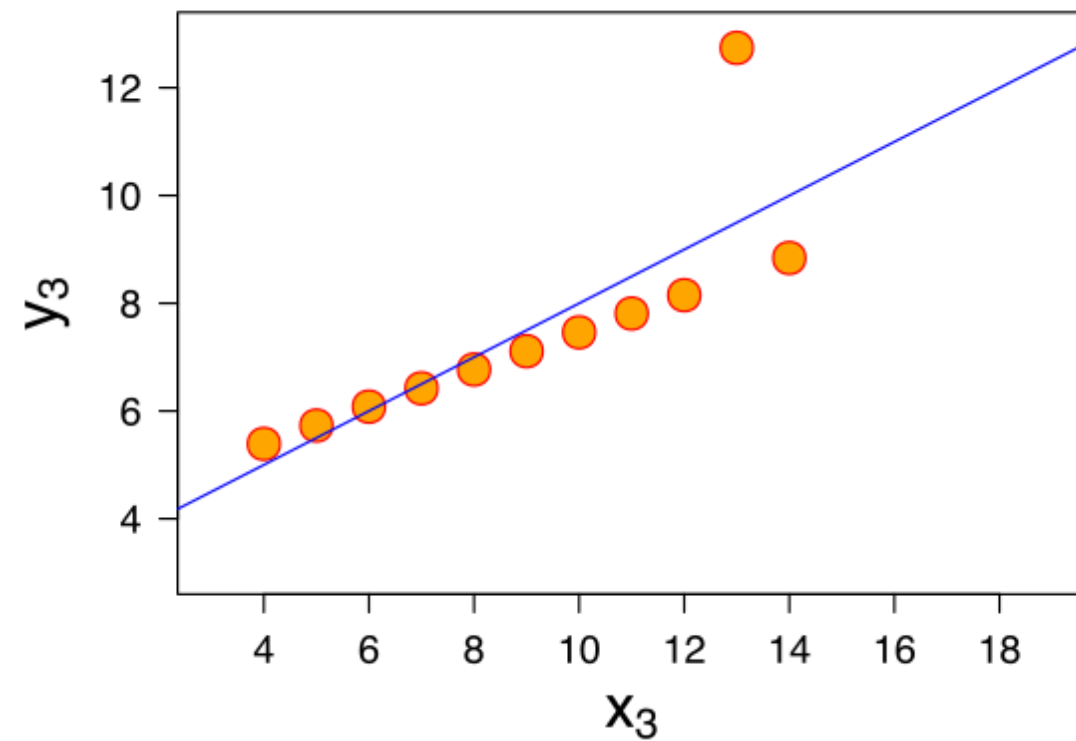
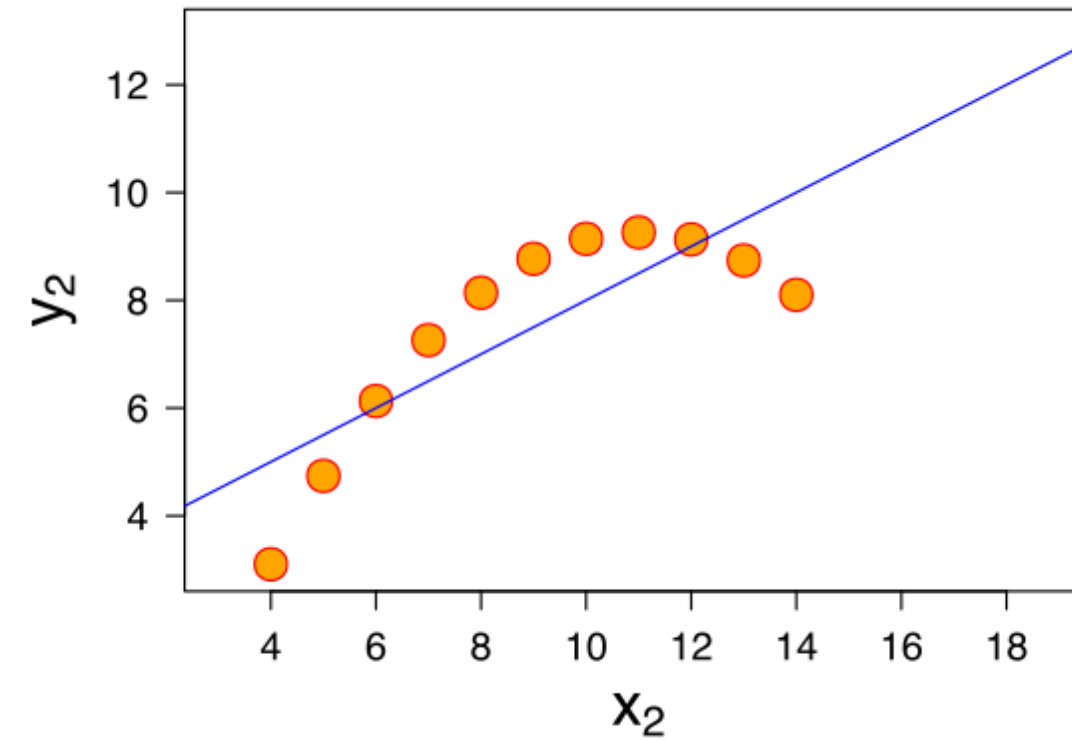
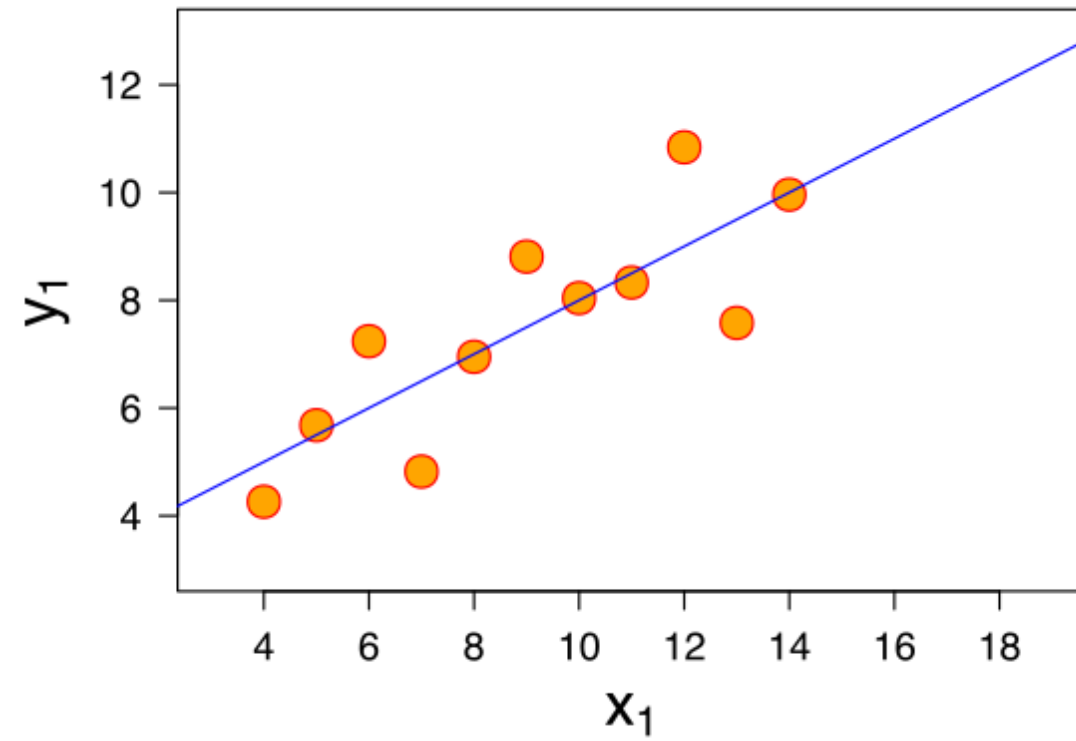
Korelasyon Katsayısı = 0.816

Doğrusal Regresyon Denklemi: $y = 0.5x + 3$

$R^2 = .67$

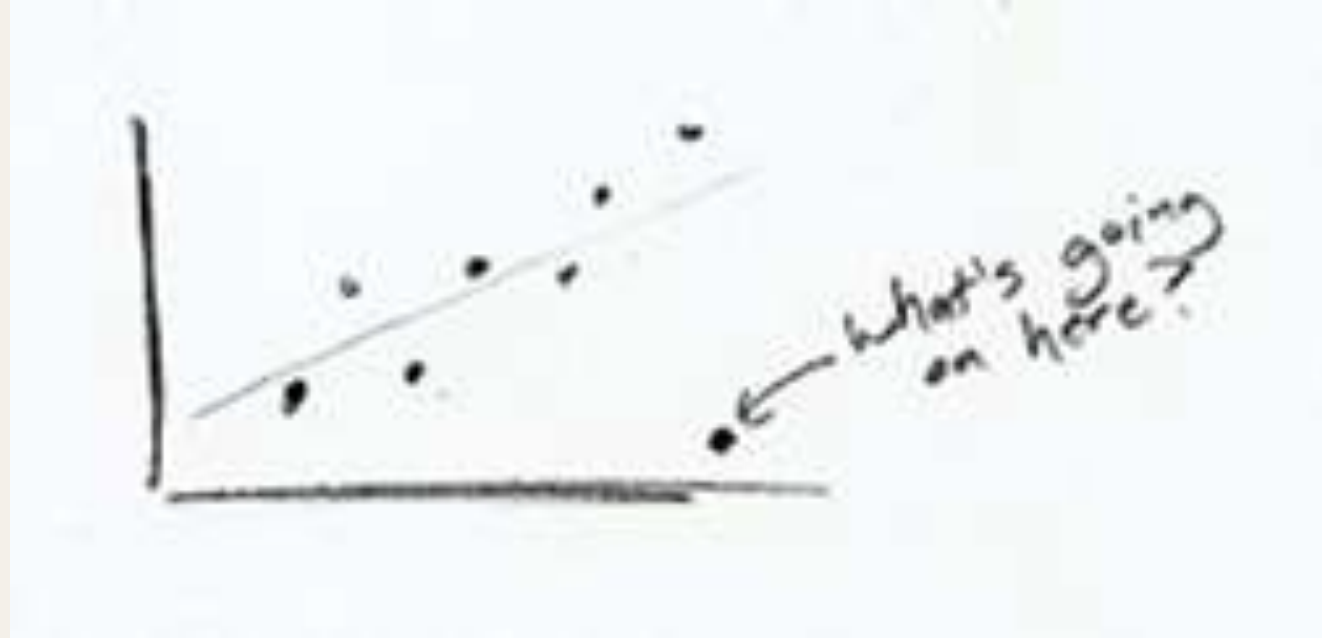
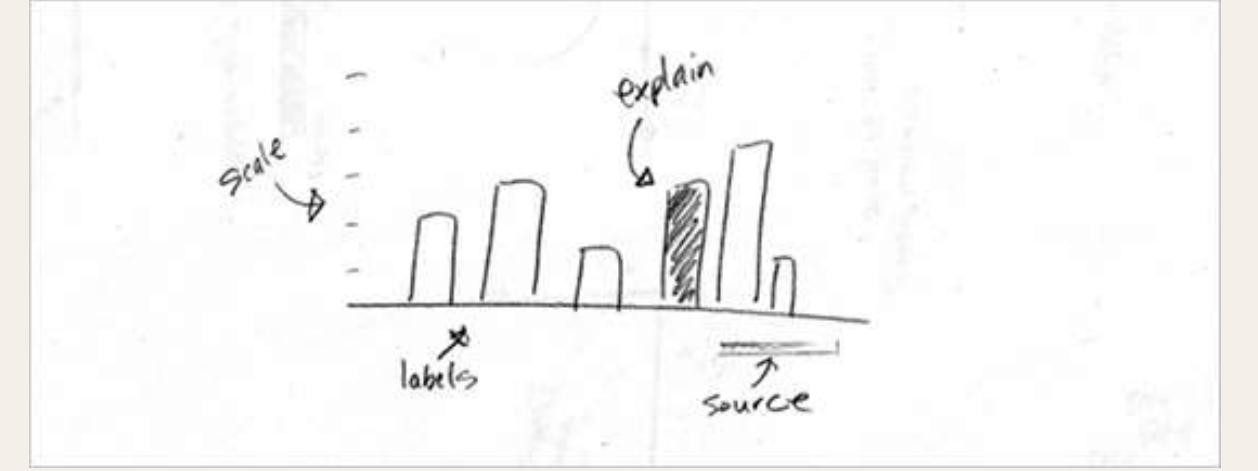
Tüm istatistik hesaplamaları aynı sonuçları veriyor, bu veri setleri üzerinde uygulanacak lineer modellemenin aynı sonucu vereceğini tahmin edebiliriz.

02 - Anscombe's Quartet



03 - Birkaç Basit Kural

Grafik oluştururken birtakım noktalara dikkat etmemiz gerekiyor.

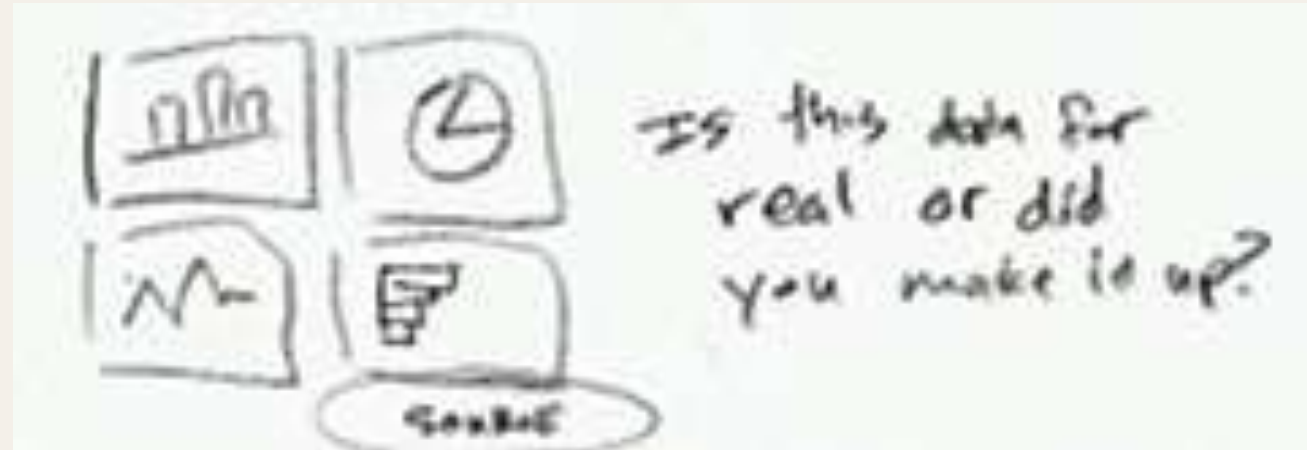


Veriyi kontrol etmek faydanıza.

03 – Birkaç Basit Kural

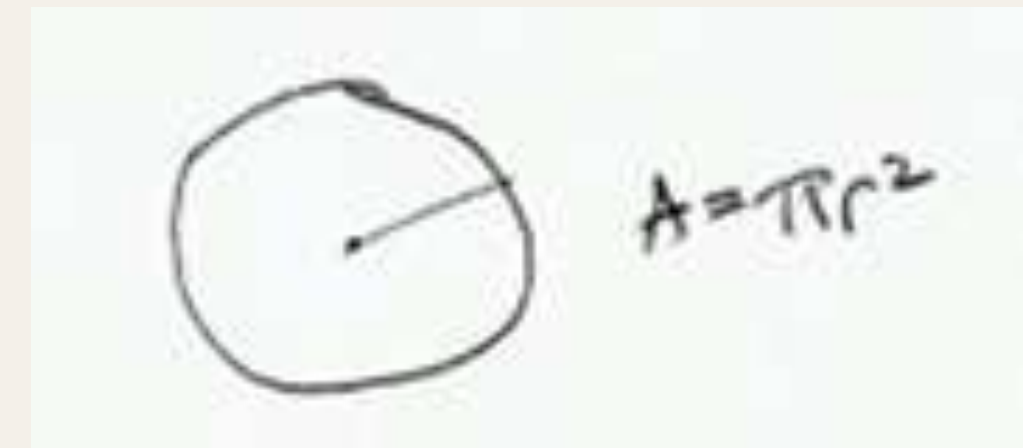
Gerektiğinde lejand kullanmanız veriyi inceleyenler için faydalı olur.

Use a legend when necessary.



Verinin kaynağını belirtmeniz önemli.

Verilerin büyüklüğü ile grafikteki geometrinin uyuşması tutarlılık açısından beklenir.



04 - R



R, istatistiksel hesaplama ve grafikler için bir dil ve ortamdır.

"Ortam" terimi, diğer veri analiz yazılımlarında sıklıkla olduğu gibi, çok özel ve esnek olmayan araçların artımlı bir birikimi yerine, tamamen planlanmış ve tutarlı bir sistem olarak nitelendirmeyi amaçlamaktadır.

05 – R ve Python: hangisi?

R

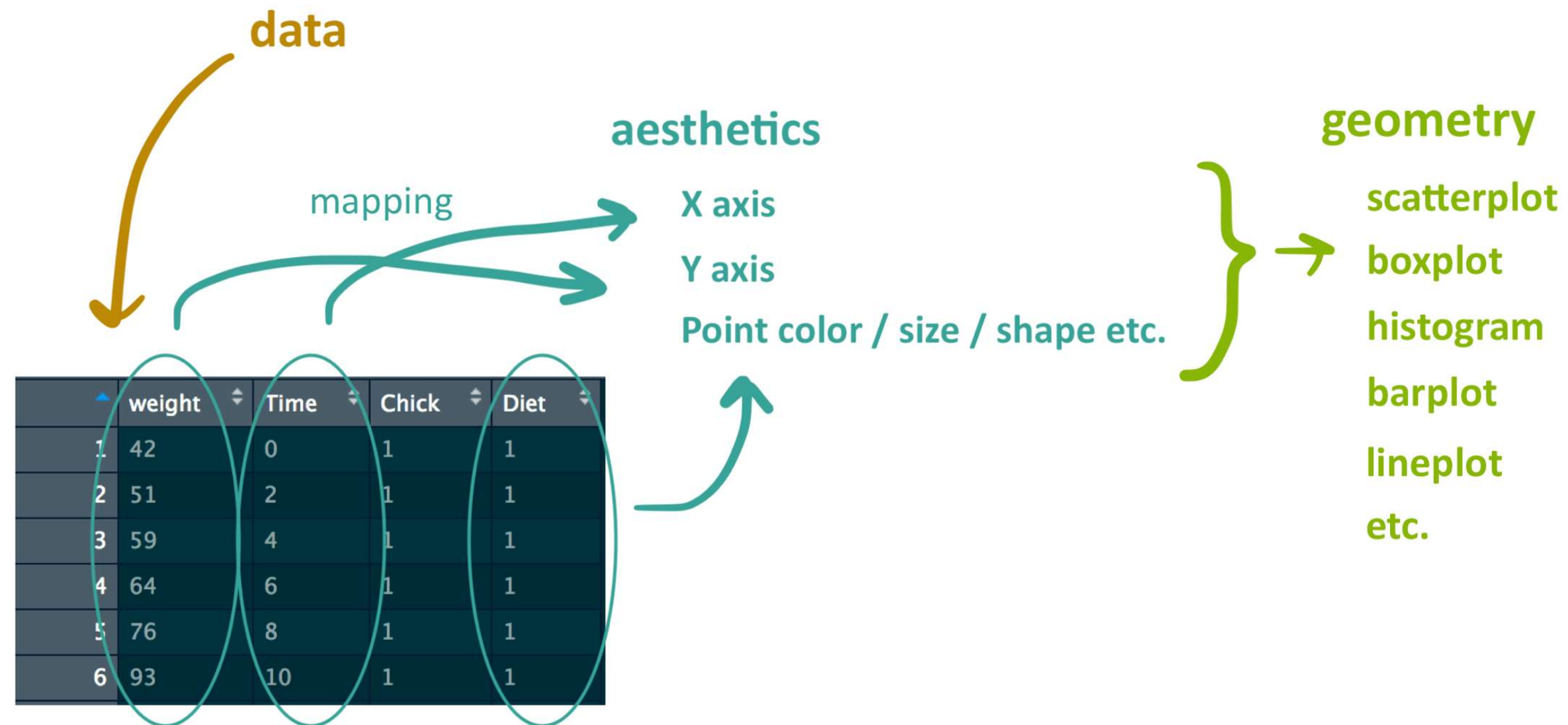
Özel amaçlı
Tidyverse
Ggplot
Vektör veriler
İstatistiki analiz
Veri görselleştirme

Python

Genel amaçlı yazılım dili
Numpy
Pandas
Matplotlib
Entegre yazılımlar

06 - Tidyverse ve ggplot

Wickham'ın düzenli veri yaklaşımı



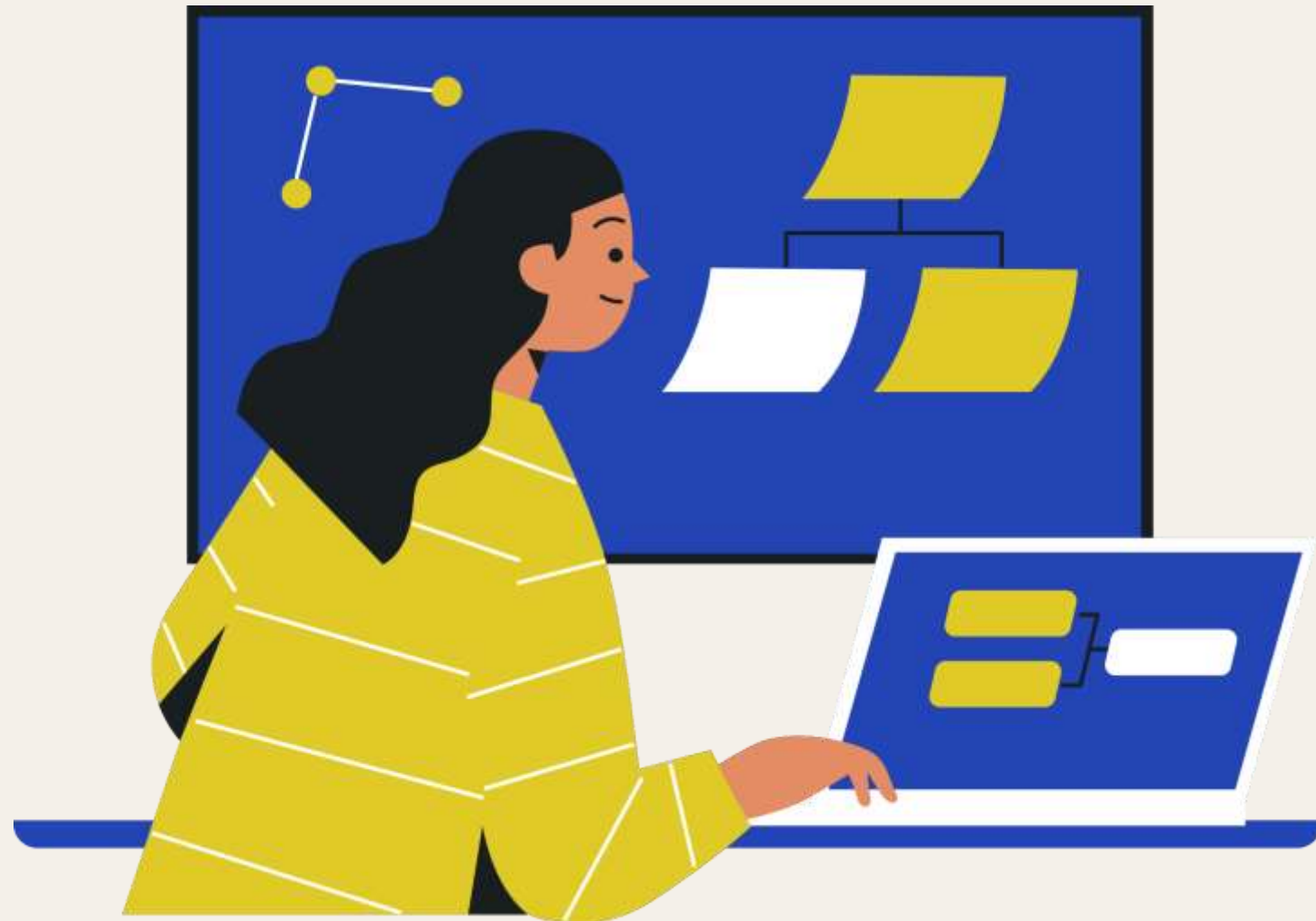
07 – Uygulama

Heat map

Slider : interactive maps

Leaflets

wordcloud



08 – Kaynaklar

The Visual Display of Quantitative Information – Edward Tufte

<https://flowingdata.com/2010/07/22/7-basic-rules-for-making-charts-and-graphs/>

<https://flowingdata.com/2015/08/11/real-chart-rules-to-follow/>

<https://www.r-project.org/about.html>

<https://www.datacamp.com/blog/python-vs-r-for-data-science-whats-the-difference>

<https://www.ibm.com/blog/python-vs-r/>

<https://www.datacamp.com/blog/python-vs-r-for-data-science-whats-the-difference>

https://en.wikipedia.org/wiki/Hadley_Wickham

<https://www.rforecology.com/post/a-simple-introduction-to-ggplot2/>



Sorularınız var mı?

Teşekkürler