Data visualization in Python

matplotlib

Data Science Hub

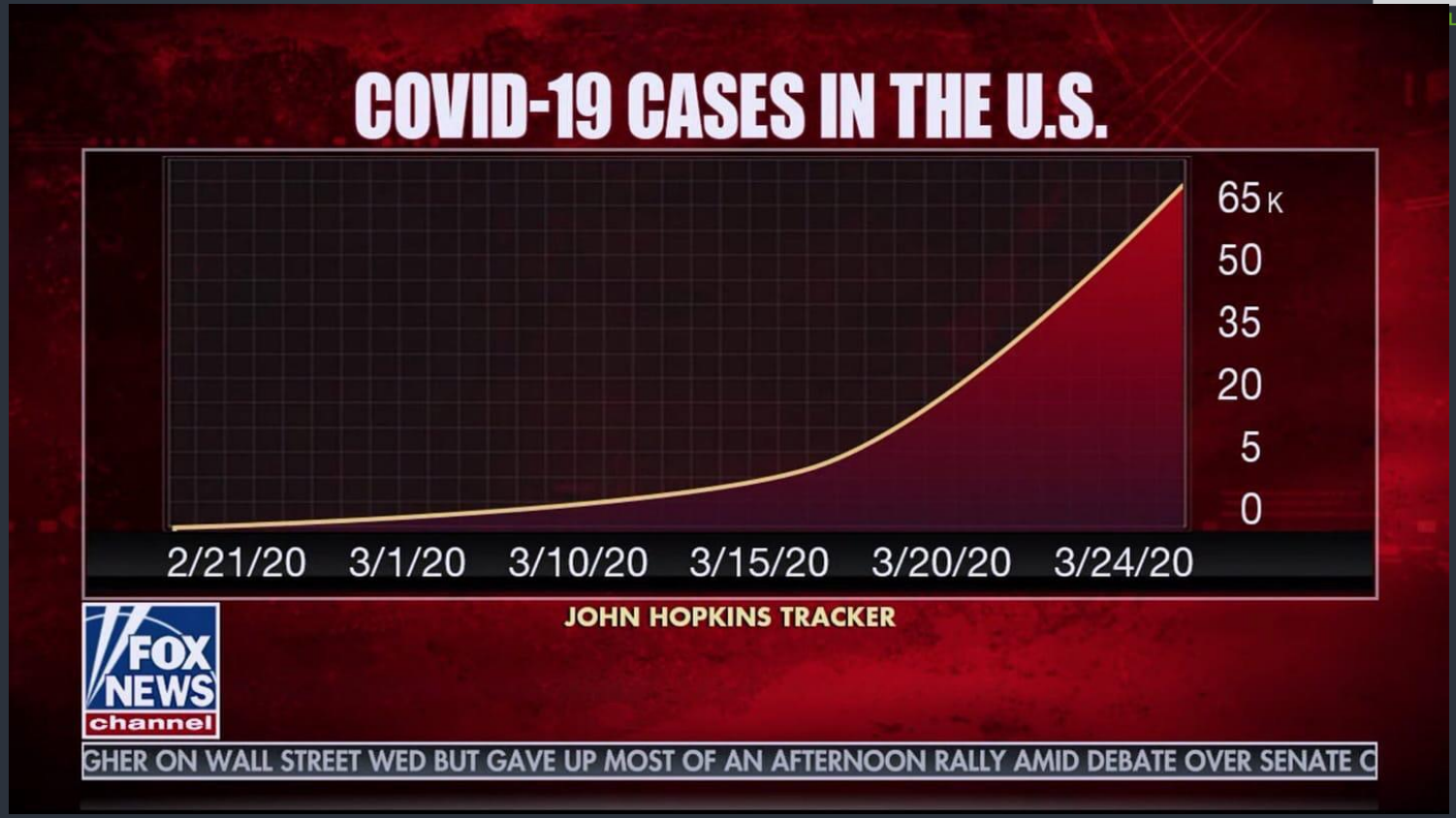
DeNederlandscheBank
EUROSYSTEEM

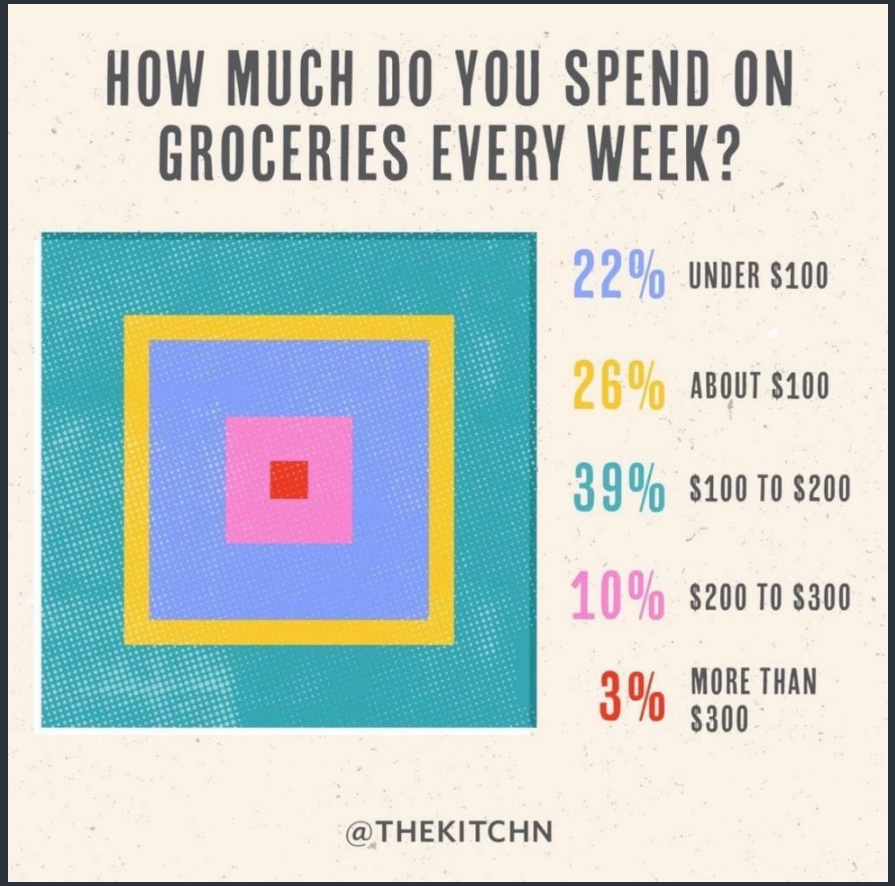
Content

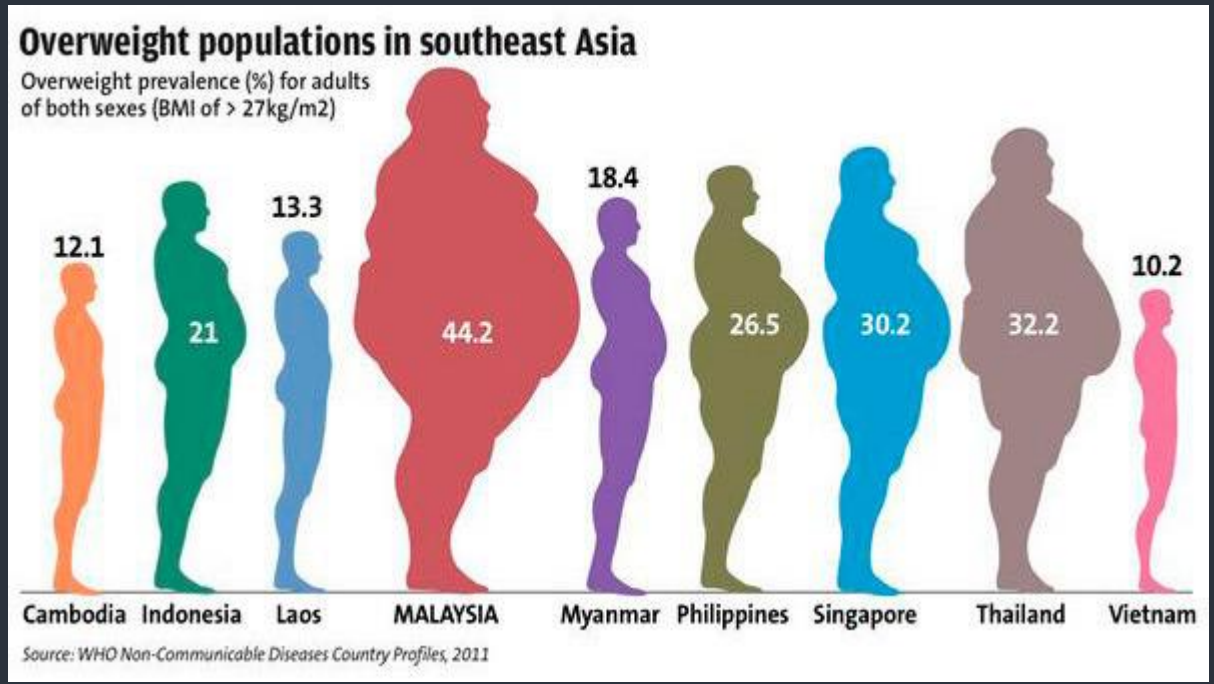
- Some visualizations
- Perceptional cues
- Chart types
- Tufte's design principals
- Matplotlib
- Practical
- Assignment I
- Assignment II

Course material

https://github.com/DeNederlandscheBank/OpenSourceWorkshop/


COVID-19 CASES IN THE U.S.
65K
50
35
20
5
0
2/21/20
3/1/20
3/10/20
3/15/20
3/20/20
3/24/20
JOHN HOPKINS TRACKER
FOX NEWS channel
GHER ON WALL STREET WED BUT GAVE UP MOST OF AN AFTERNOON RALLY AMID DEBATE OVER SENATE C


HOW MUCH DO YOU SPEND ON GROCERIES EVERY WEEK?
22% UNDER $100
26% ABOUT $100
39% $100 TO $200
10% $200 TO $300
3% MORE THAN $300
@THEKITCHN


Overweight populations in southeast Asia
Overweight prevalence (%) for adults of both sexes (BMI of > 27kg/m2)
12.1
21
13.3
44.2
18.4
26.5
30.2
32.2
10.2
Cambodia
Indonesia
Laos
MALAYSIA
Myanmar
Philippines
Singapore
Thailand
Vietnam
Source: WHO Non-Communicable Diseases Country Profiles, 2011

Verdeling positief geteste mensen in Nederland

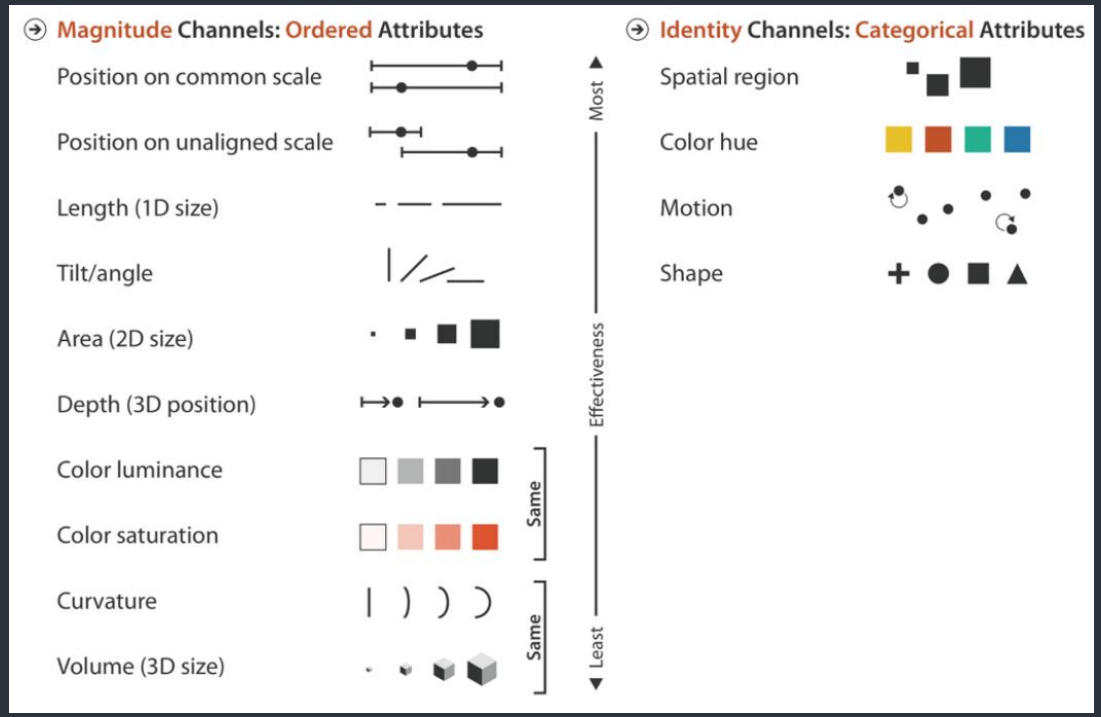
Deze kaarten laten zien van hoeveel mensen gisteren is gemeld dat ze positief getest zijn op COVID-19, per 100.000 inwoners.

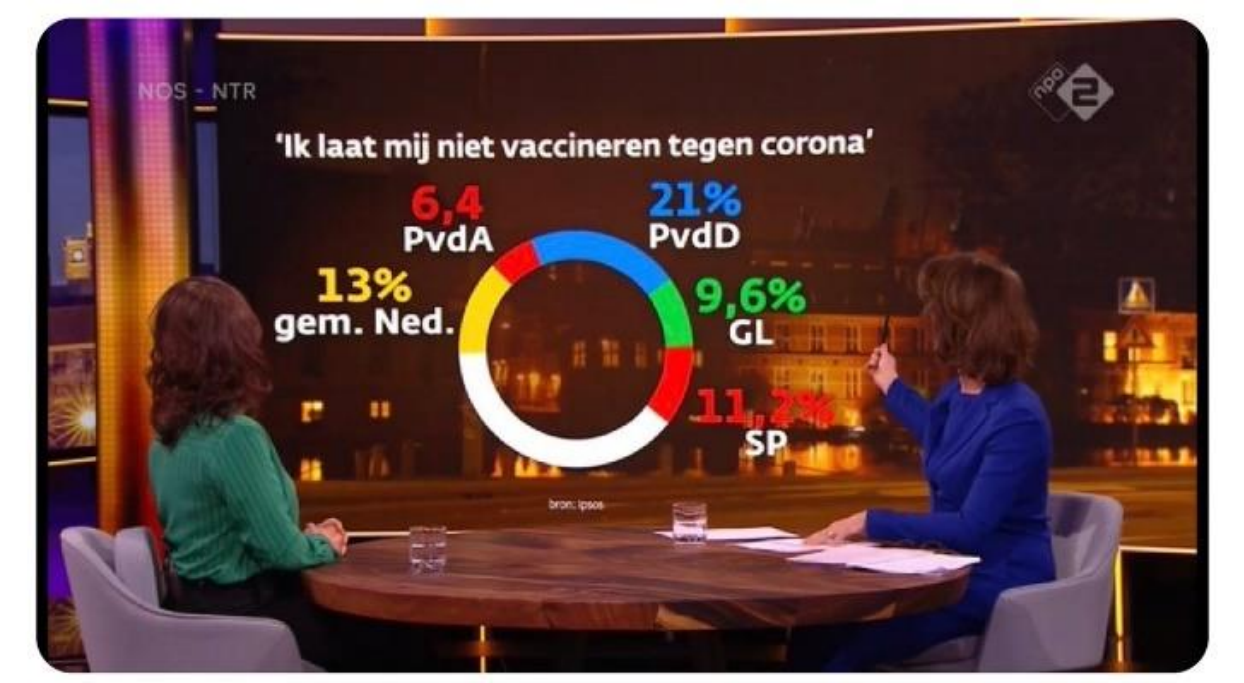
Per gemeente | Per veiligheidsregio

Aantal per 100.000 inwoners

0 4 7 10 20 30

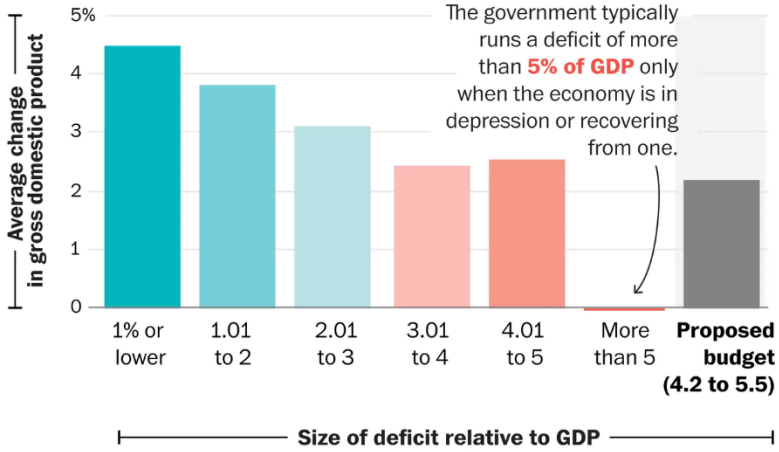
Perceptional cues




NOS - NTR
'Ik laat mij niet vaccineren tegen corona'
6,4
PvdA
21%
PvdD
13%
gem. Ned.
9,6%
GL
11,2%
SP
bron: Ipsos

Strange time for a stimulus

What annual **economic growth** averaged under various deficit-to-GDP ratios, since 1967



Notes: To capture the environment in which the budget was set, deficit-to-GDP ratios are compared with the economic climate of the prior fiscal year. GDP growth is adjusted for inflation and seasonality. Indicators for the current budget are based on the average of available data in fiscal 2017 and 2018 years. Fiscal years end in September.

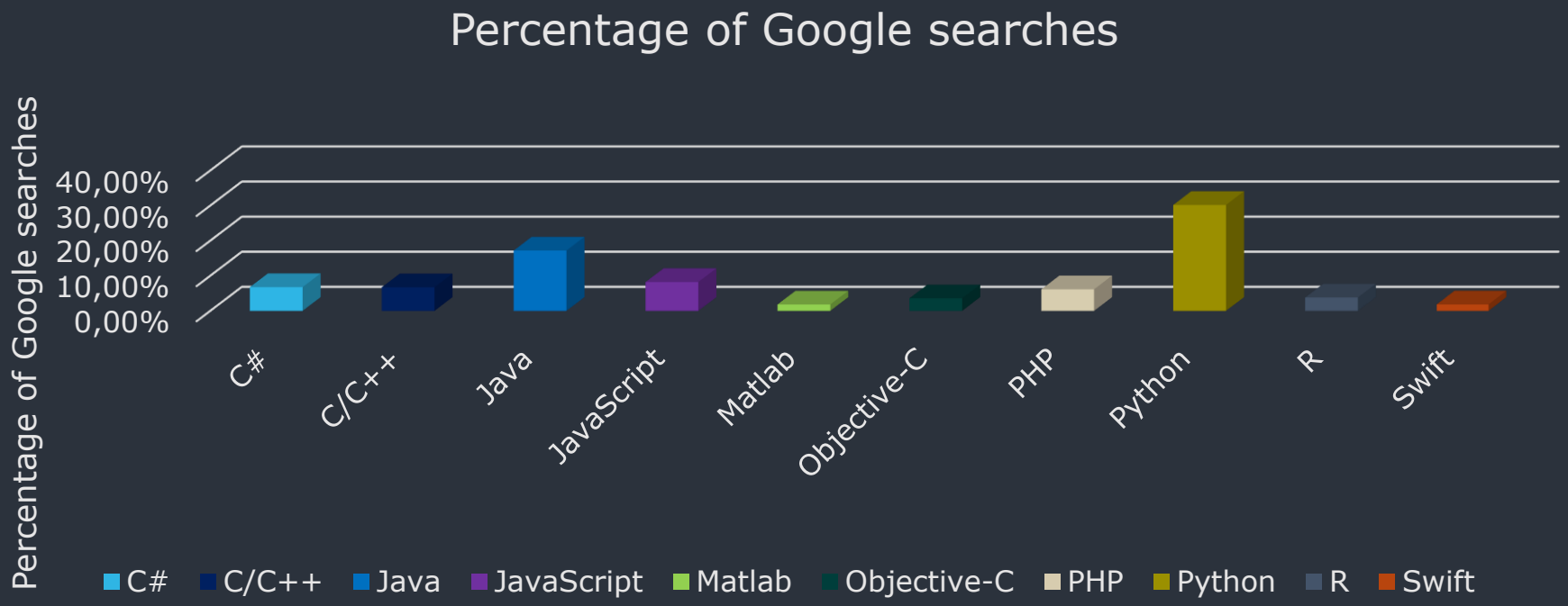
Sources: Commerce Department (GDP); Congressional Budget Office (historical deficit); Committee for a Responsible Federal Budget (deficit forecasts, budget changes)

THE WASHINGTON POST

Tufte’s design principals

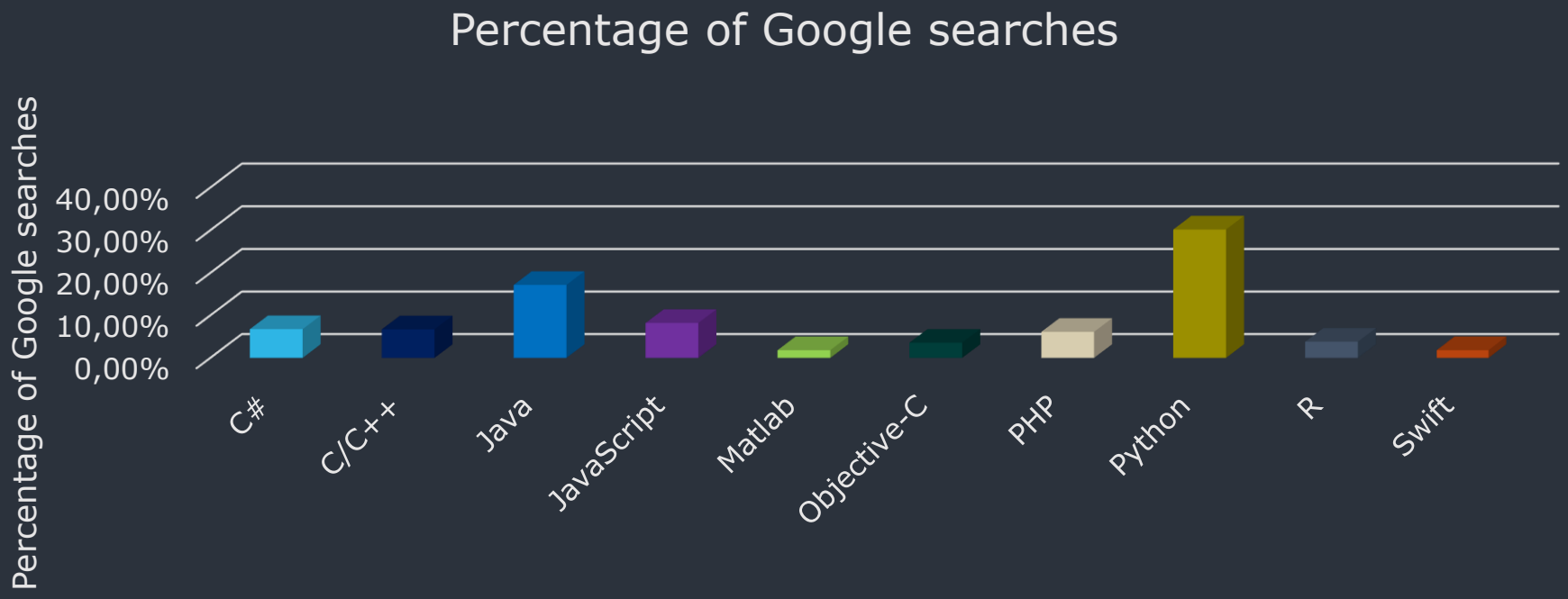
- Clear detailed labeling and appropraite scales
- Maximize data-ink ratio
- Avoid chart junk

Tufte’s design principals



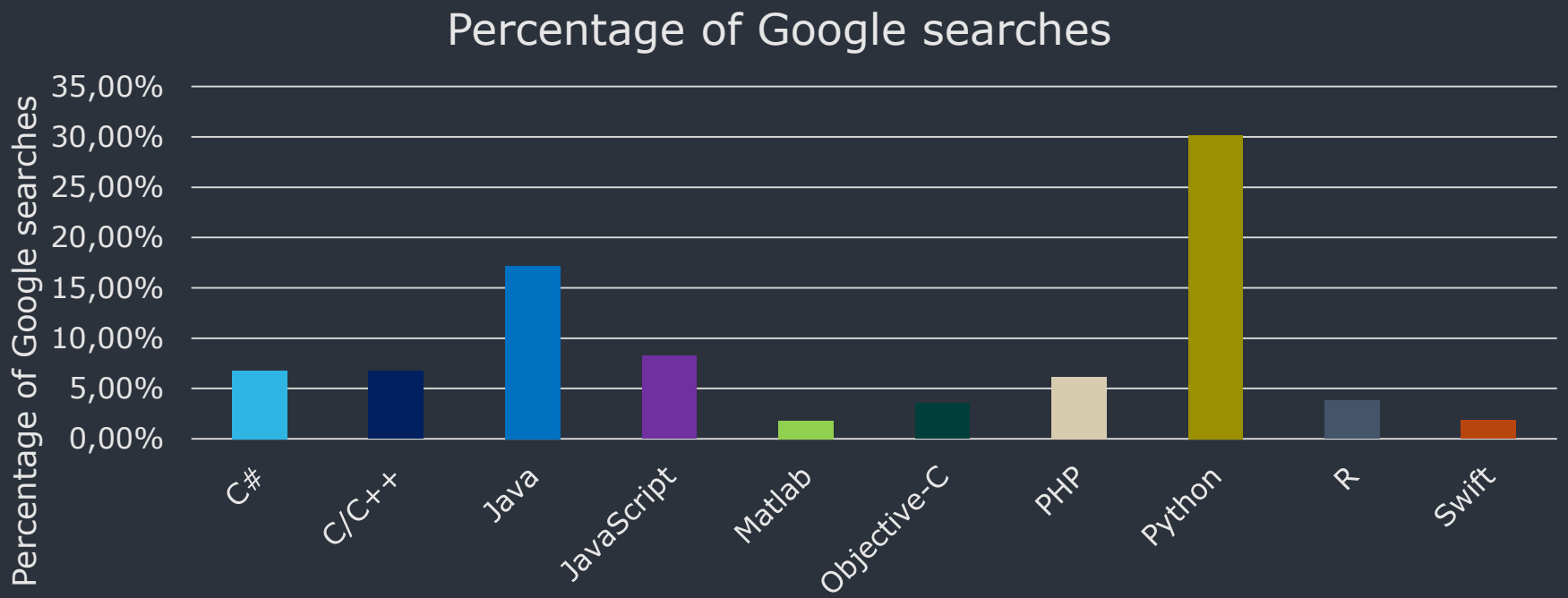
Tufte’s design principals

DNB PUBLIC



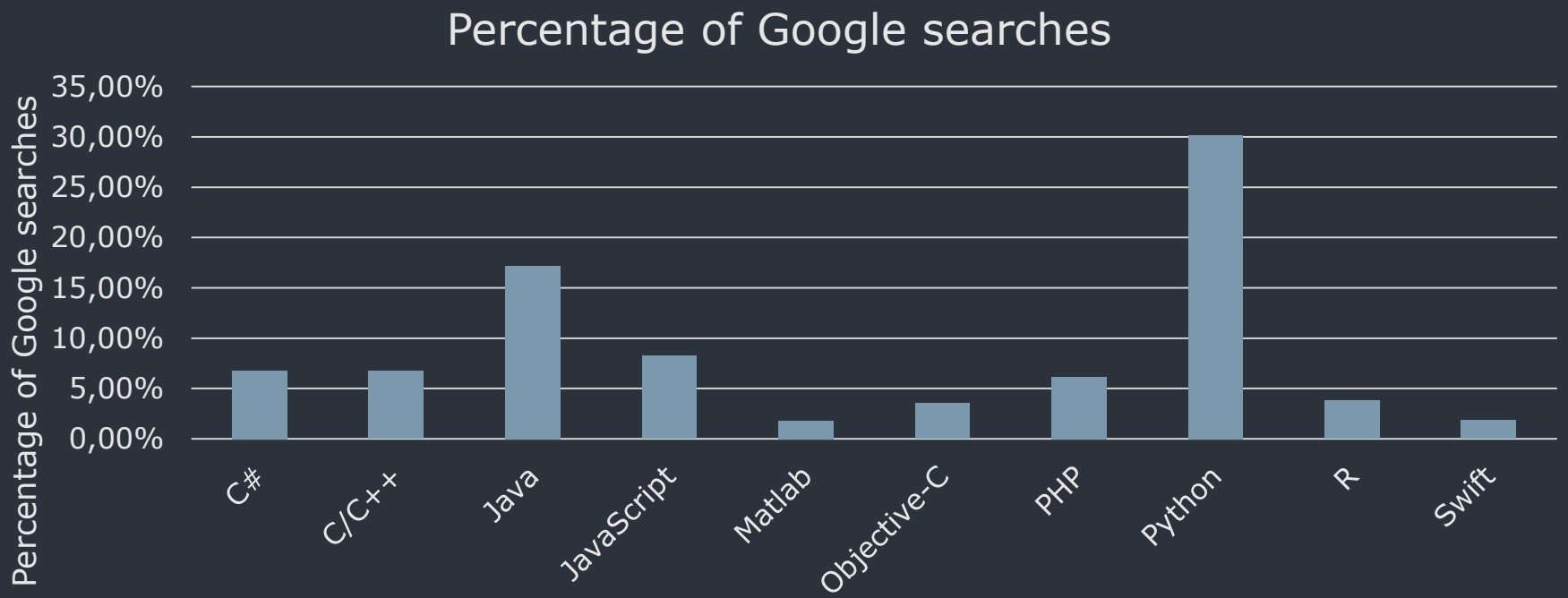
Tufte’s design principals

DNB PUBLIC



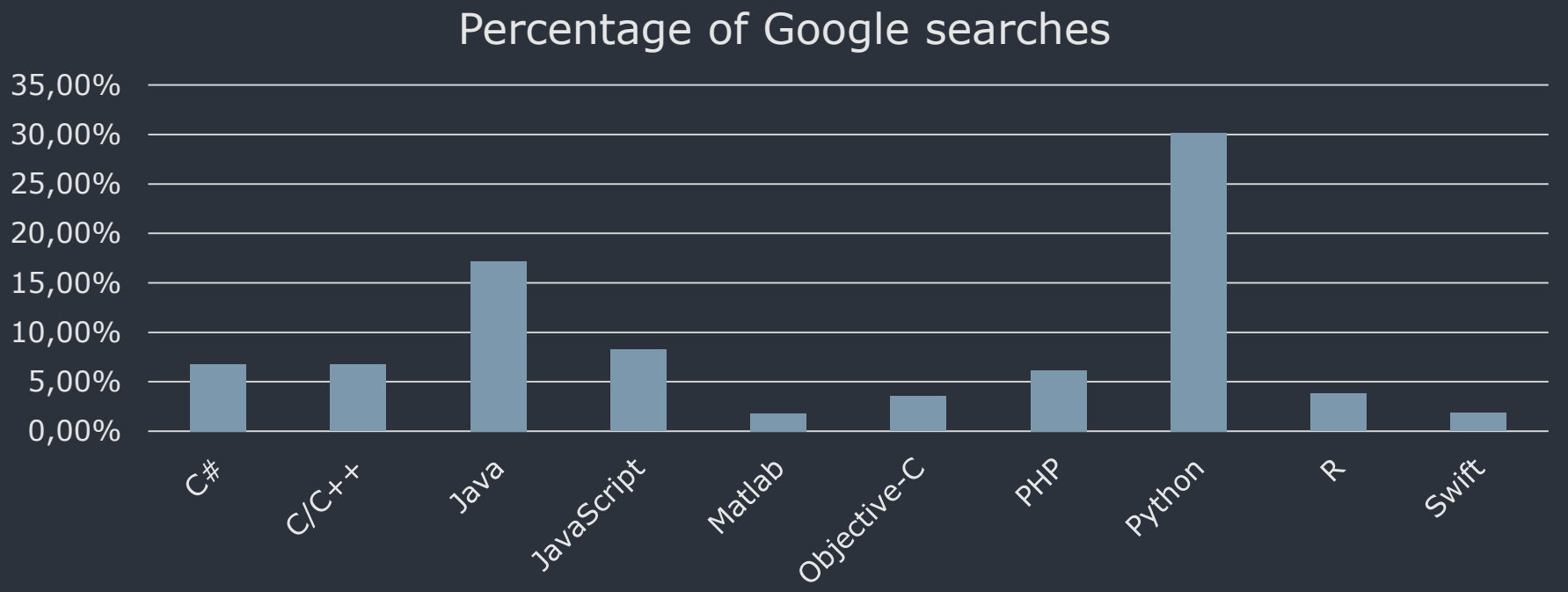
Tufte’s design principals

DNB PUBLIC



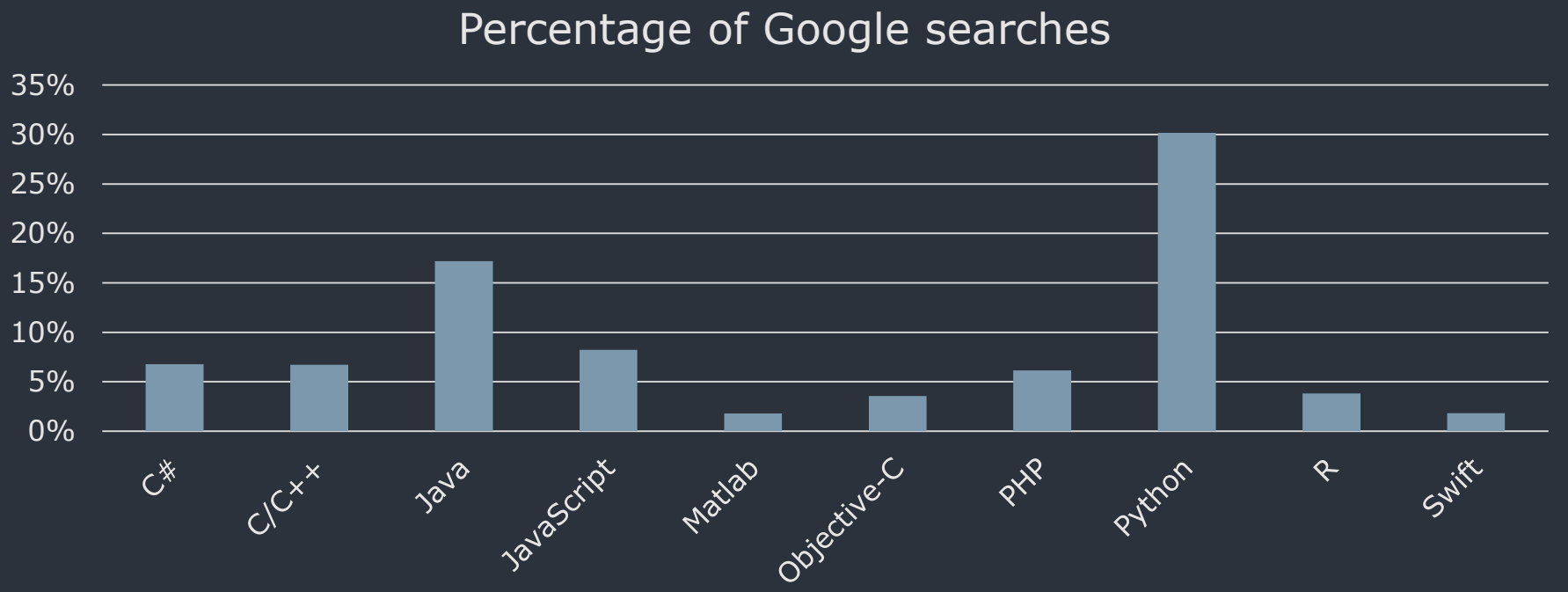
Tufte’s design principals

DNB PUBLIC



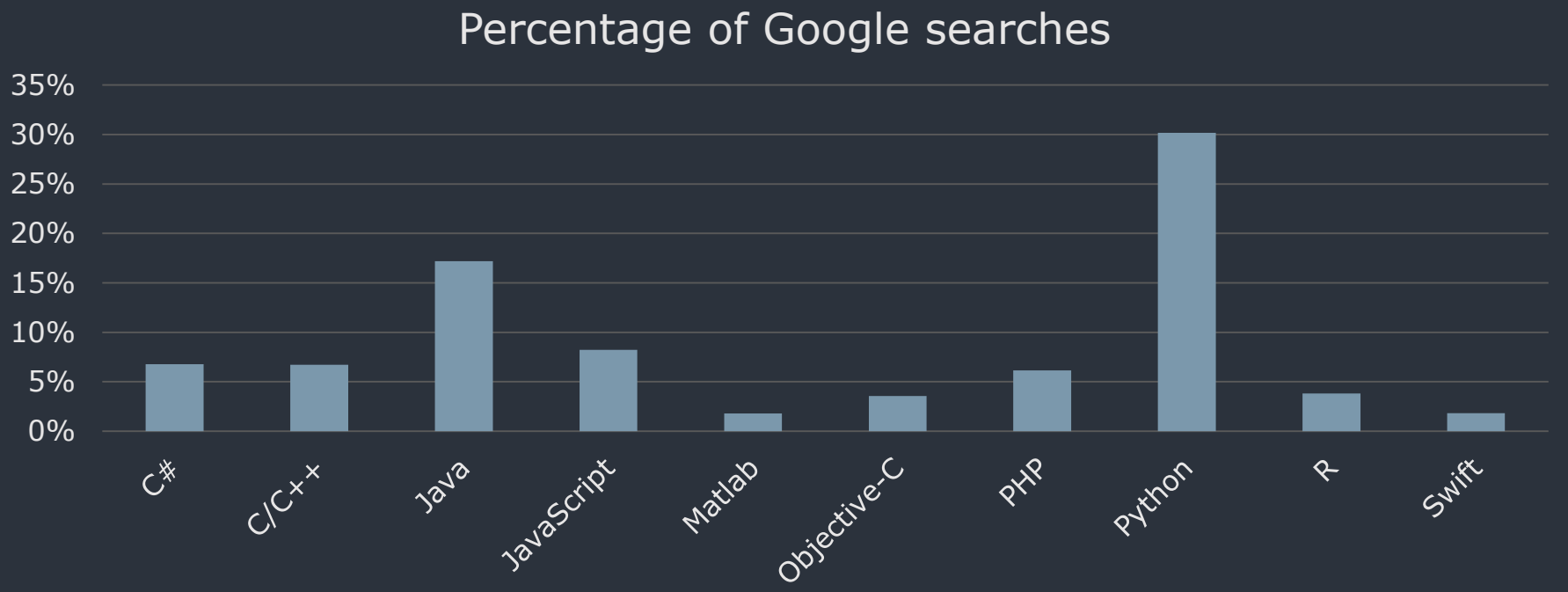
Tufte’s design principals

DNB PUBLIC



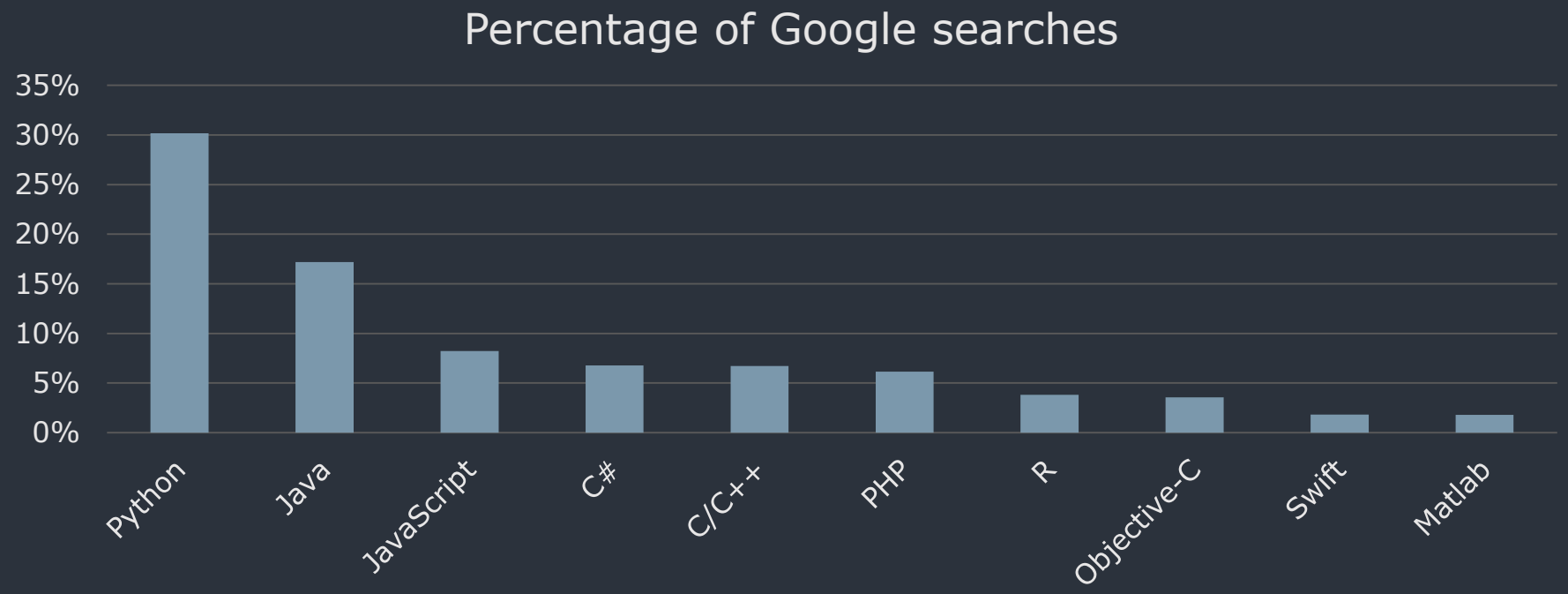
Tufte’s design principals

DNB PUBLIC



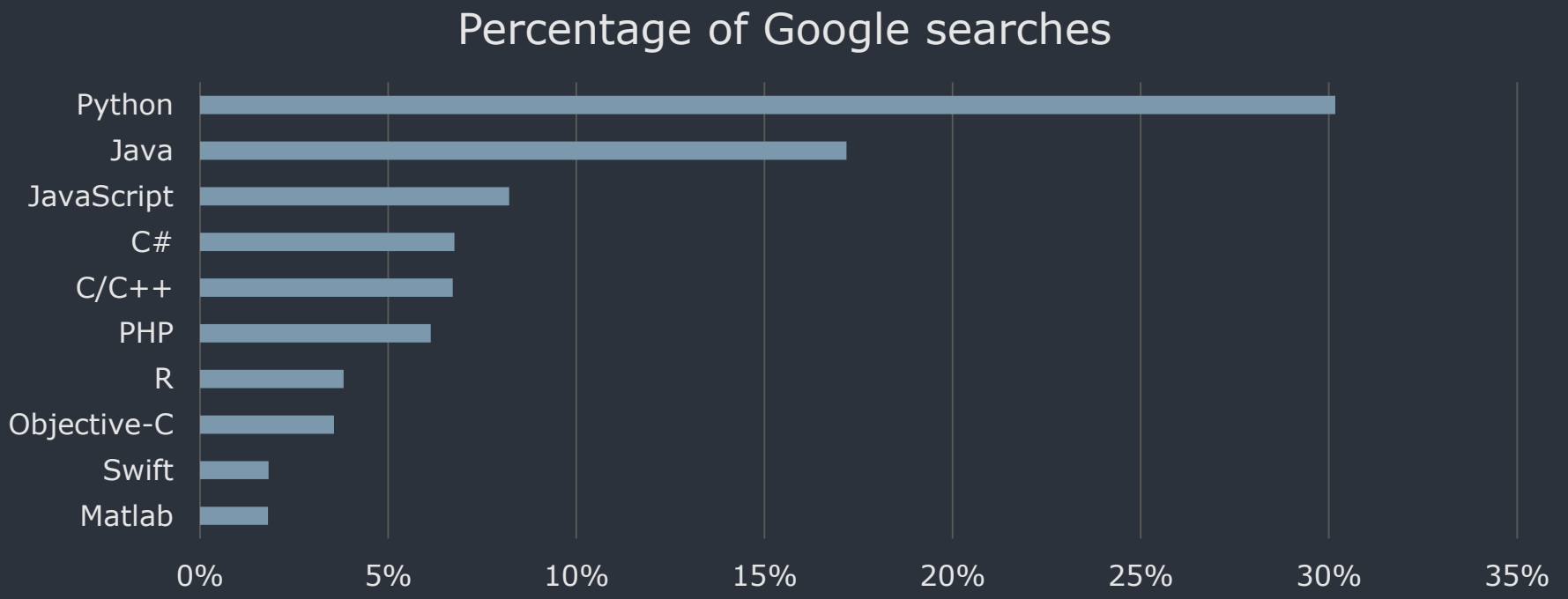
Tufte’s design principals

DNB PUBLIC



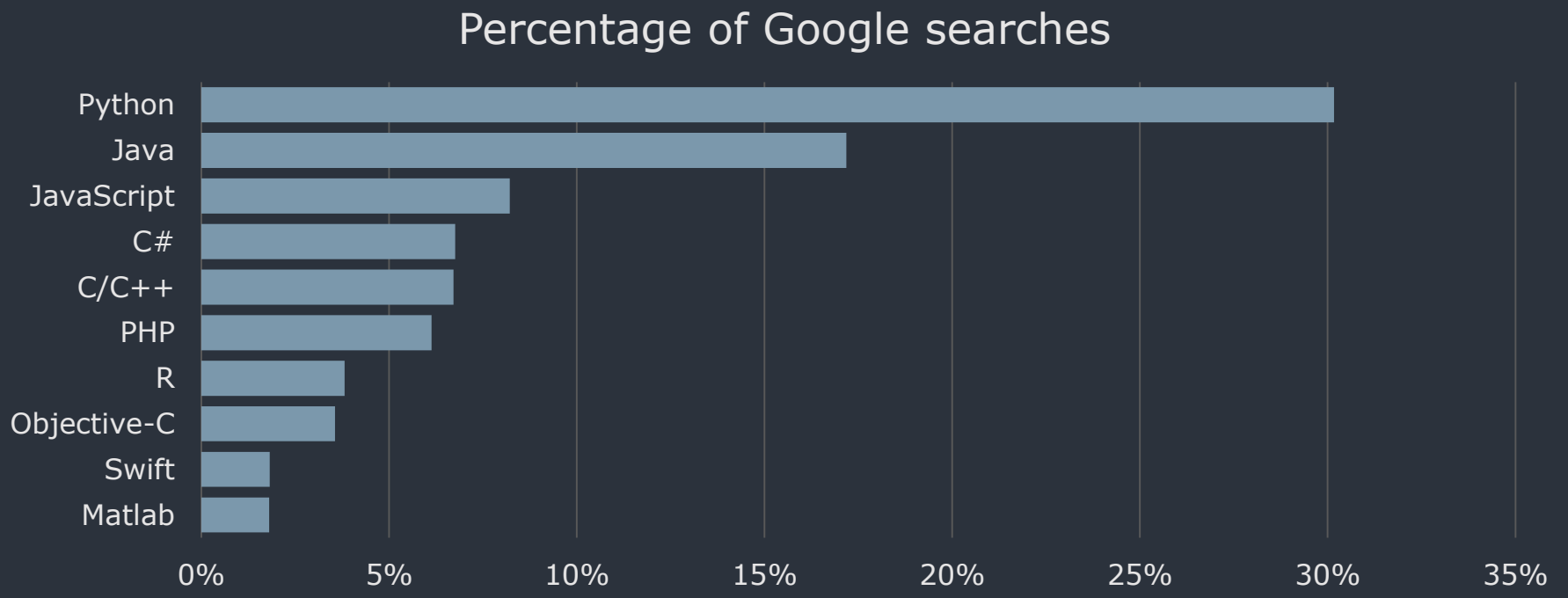
Tufte’s design principals

DNB PUBLIC



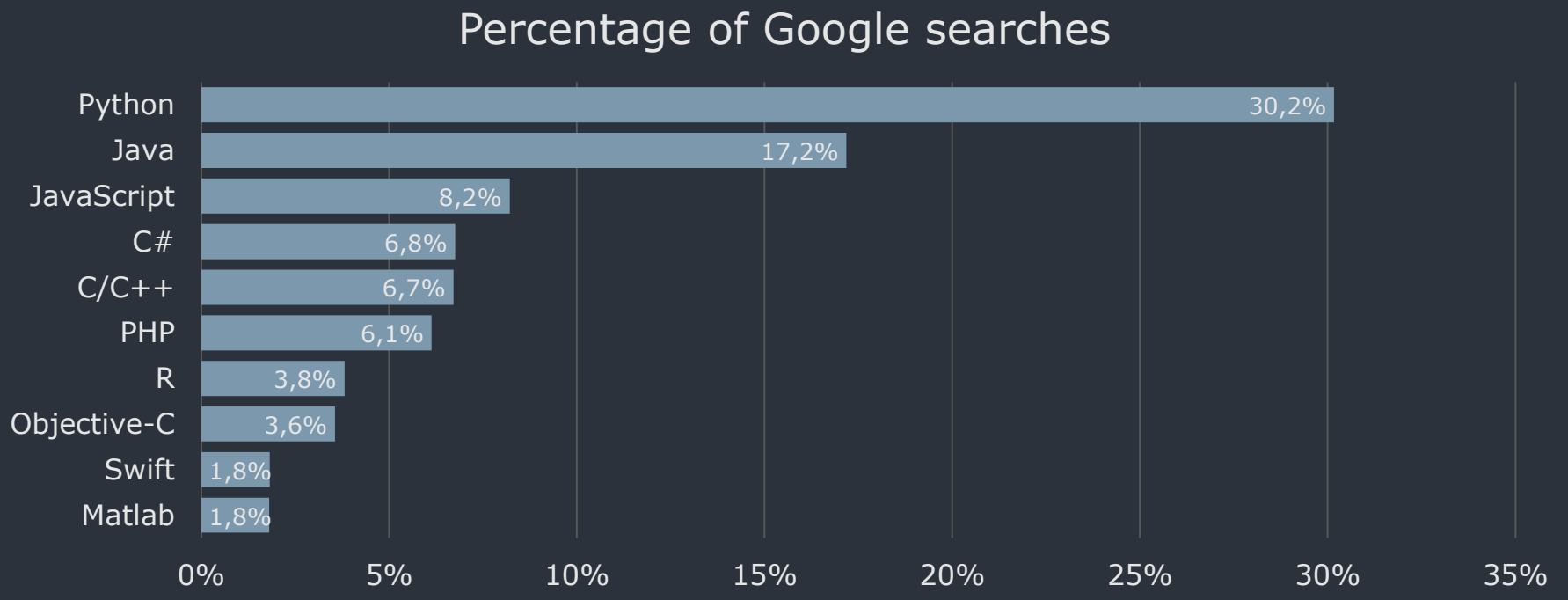
Tufte’s design principals

DNB PUBLIC

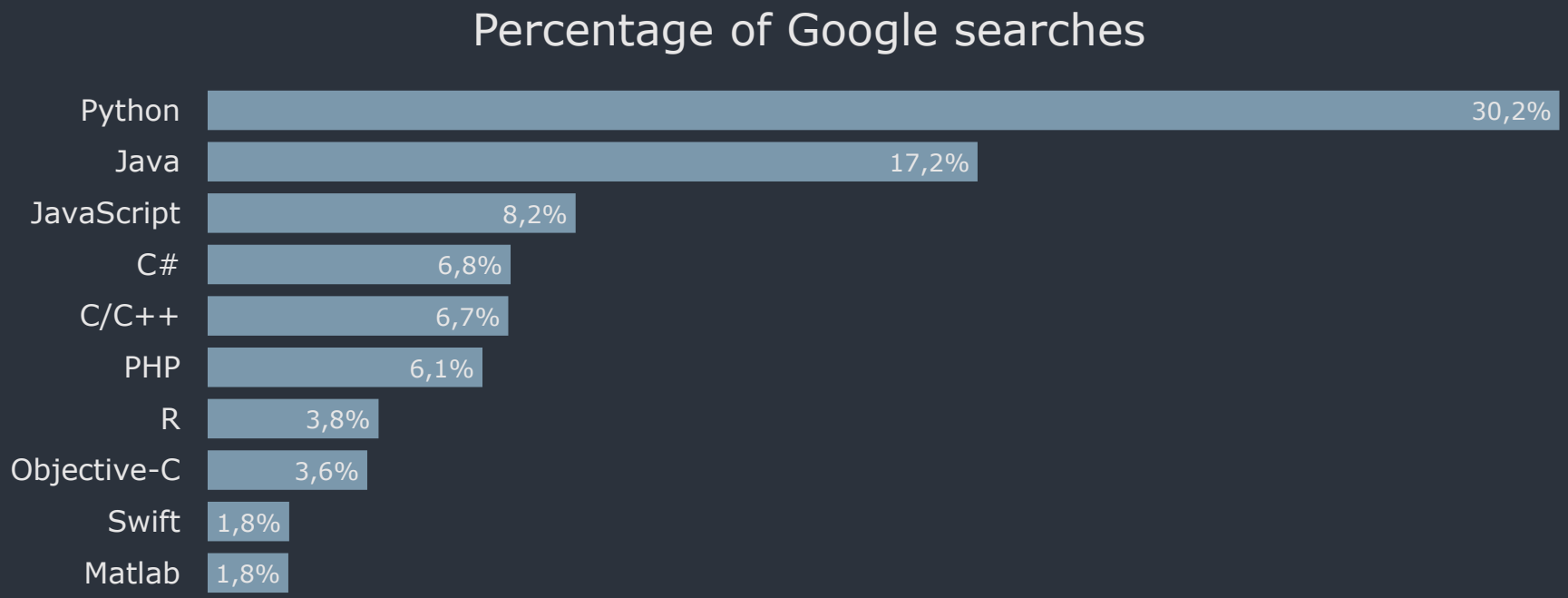


Tufte’s design principals

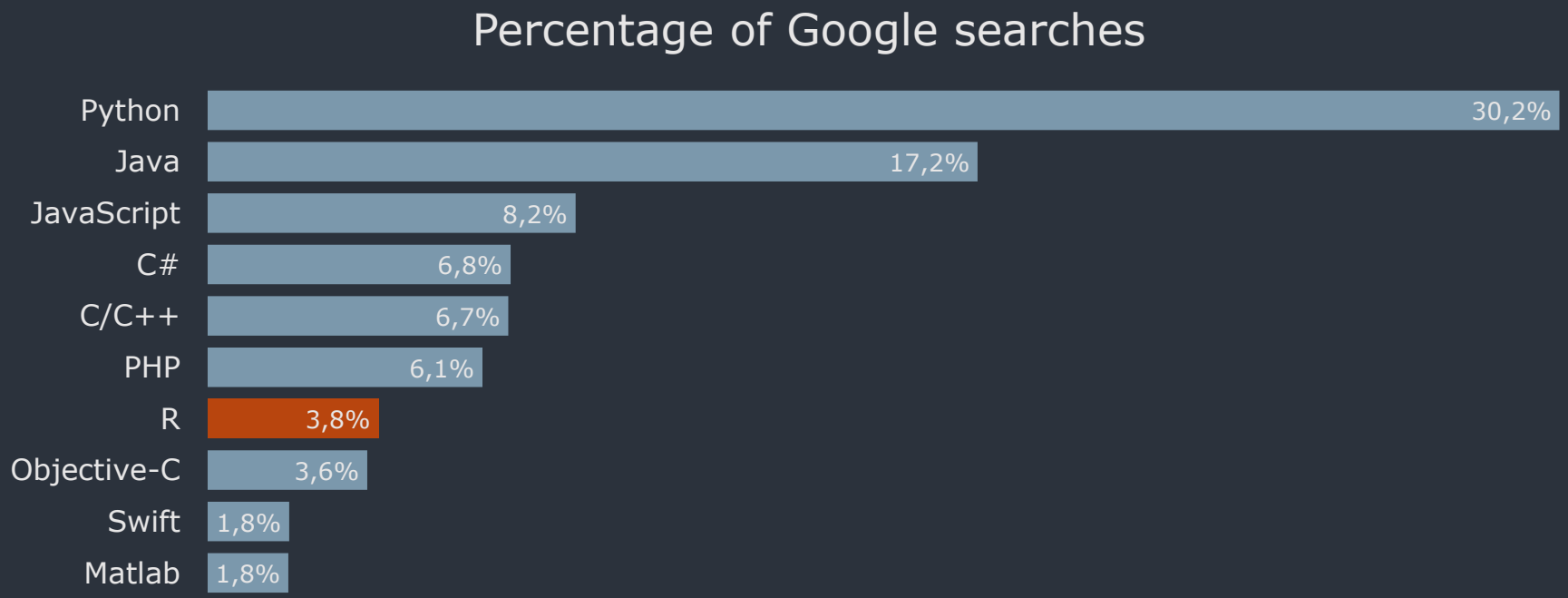
DNB PUBLIC



Tufte’s design principals

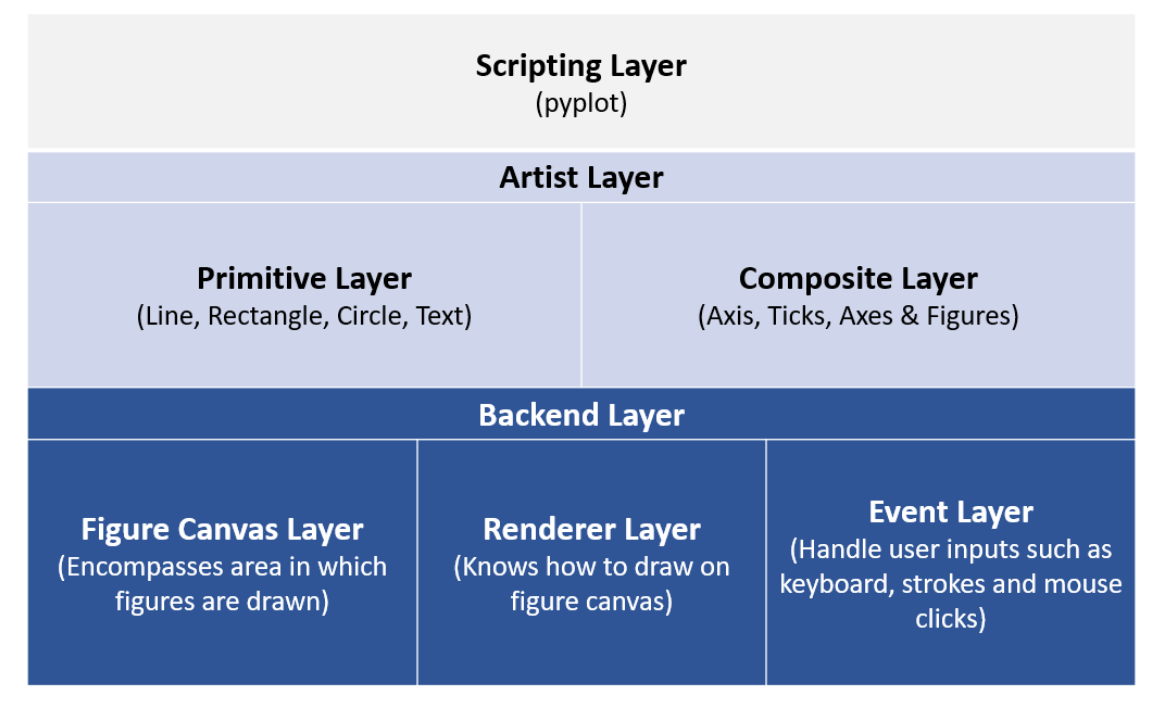


Tufte’s design principals

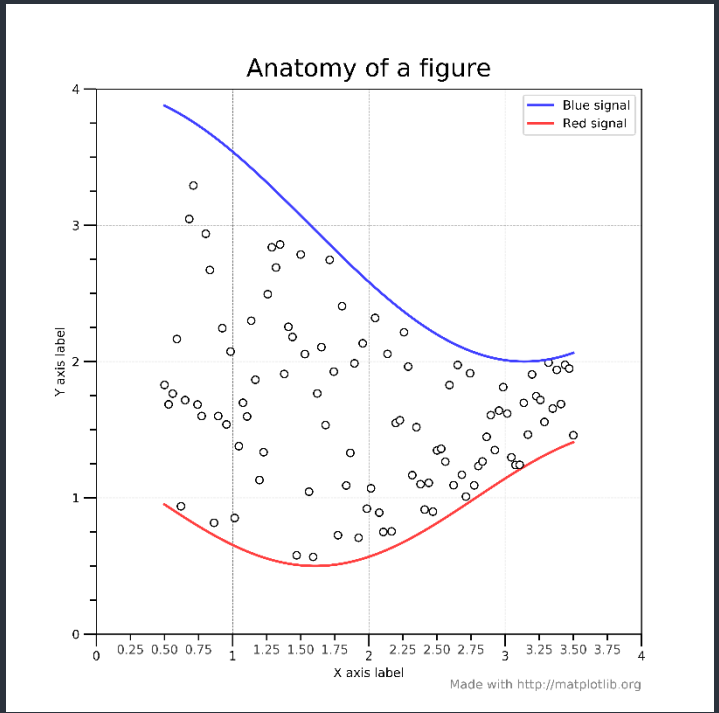


Matplotlib

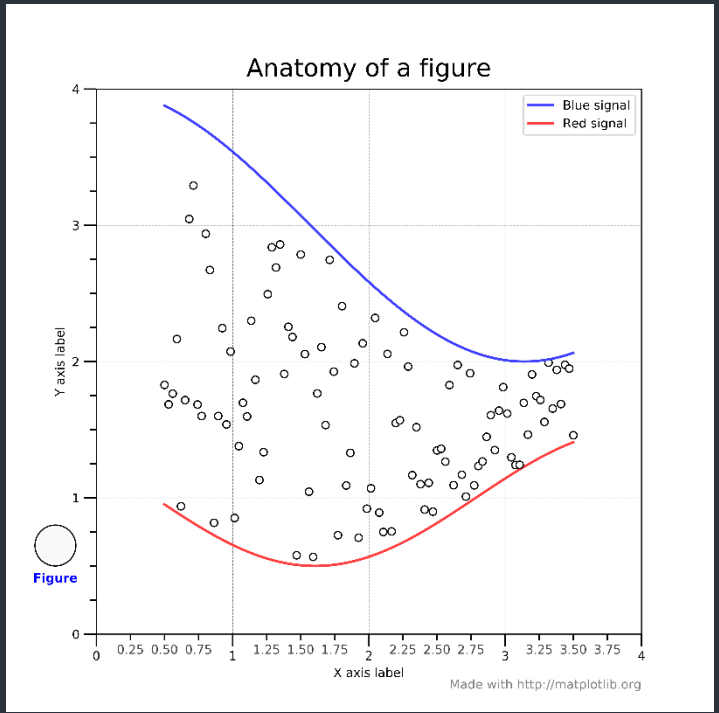
Matplotlib


Scripting Layer
(pyplot)
Artist Layer
Primitive Layer
(Line, Rectangle, Circle, Text)
Composite Layer
(Axis, Ticks, Axes & Figures)
Backend Layer
Figure Canvas Layer
(Encompasses area in which figures are drawn)
Renderer Layer
(Knows how to draw on figure canvas)
Event Layer
(Handle user inputs such as keyboard, strokes and mouse clicks)

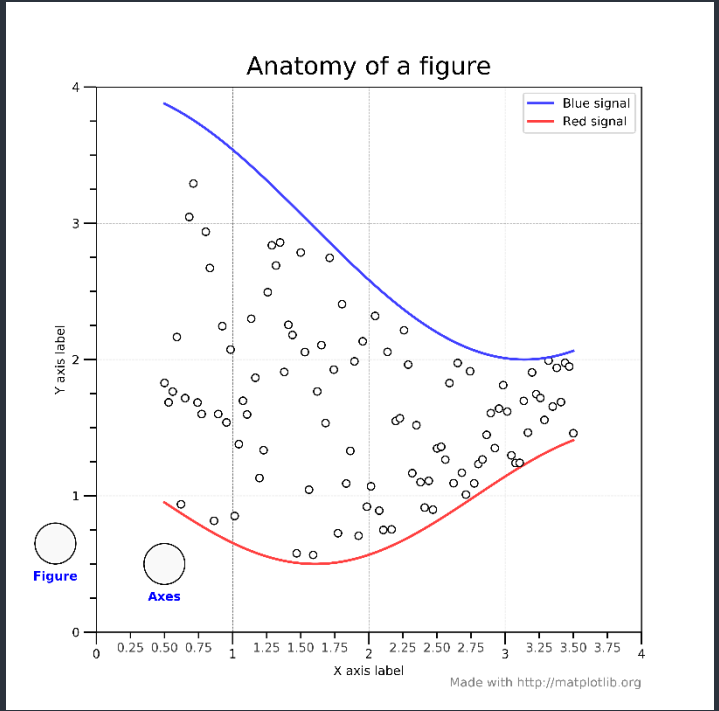
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Y axis label
X axis label
Made with http://matplotlib.org
0
1
2
3
4
0.25 0.50 0.75
1.25 1.50 1.75
2.25 2.50 2.75
3.25 3.50 3.75

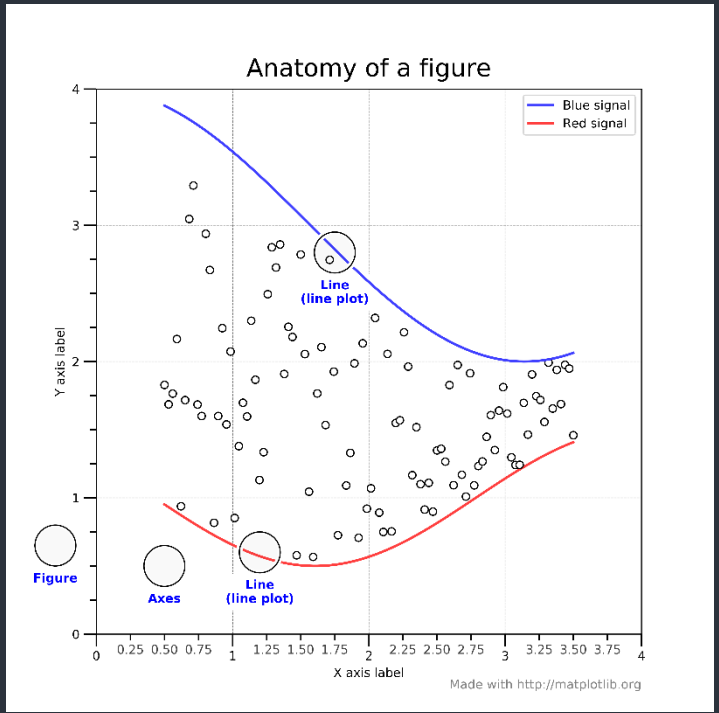
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Y axis label
X axis label
Figure
Made with http://matplotlib.org

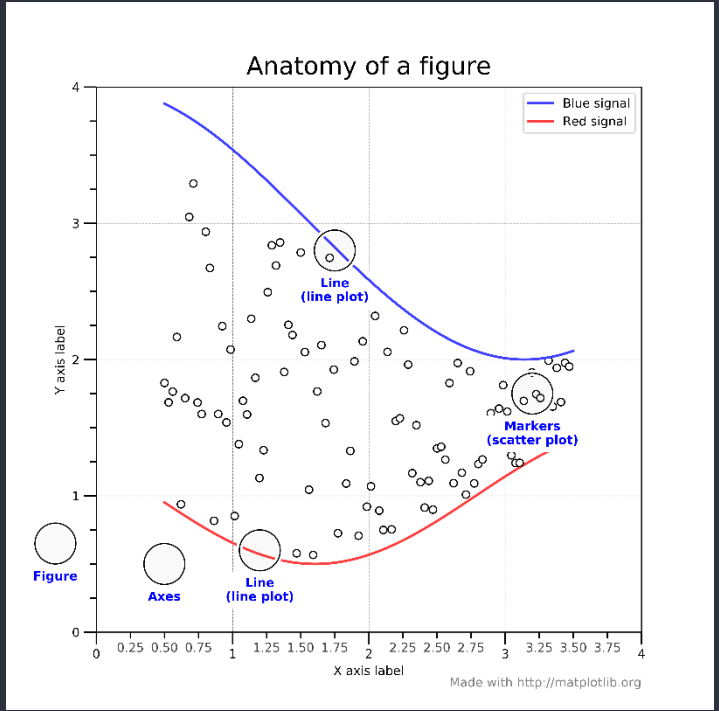
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Y axis label
X axis label
Figure
Axes
Made with http://matplotlib.org

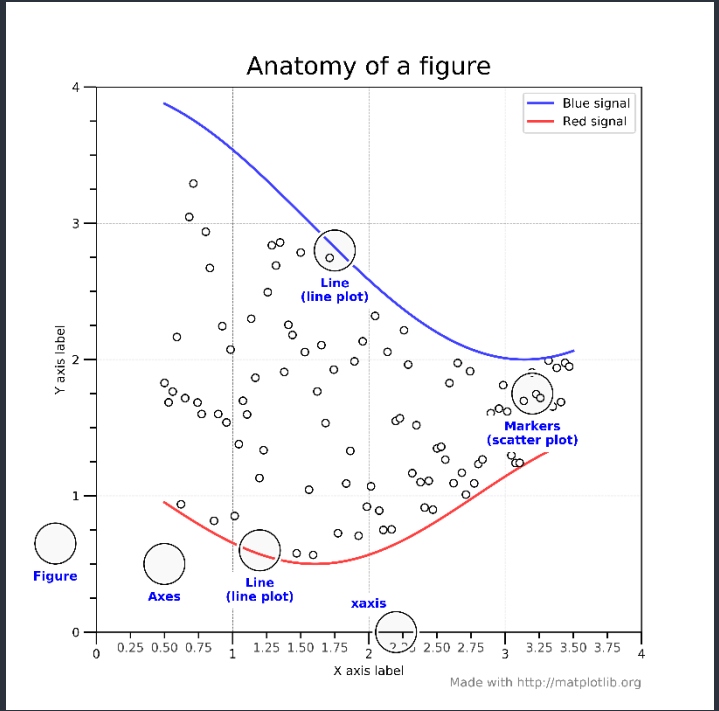
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Line
(line plot)
Line
(line plot)
Figure
Axes
Y axis label
X axis label
Made with http://matplotlib.org

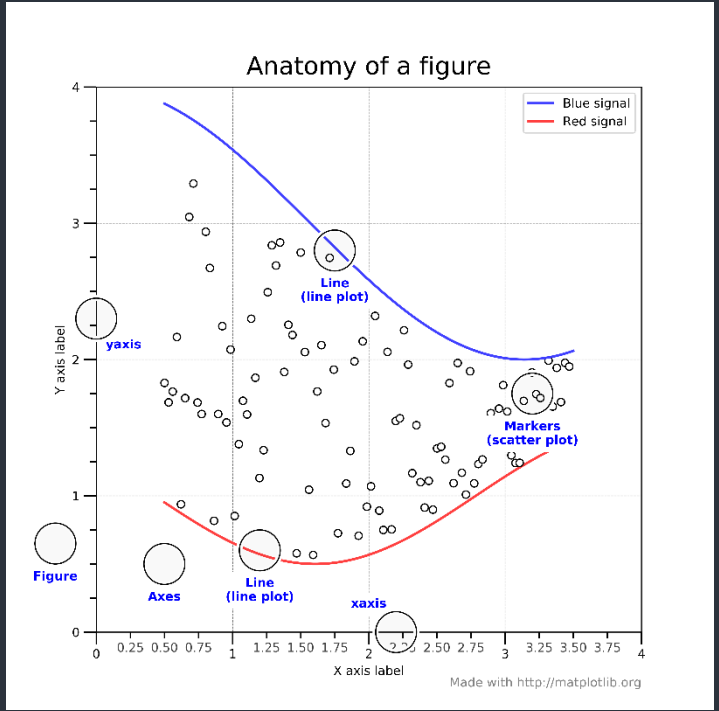
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Line
(line plot)
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
Y axis label
X axis label
Made with http://matplotlib.org

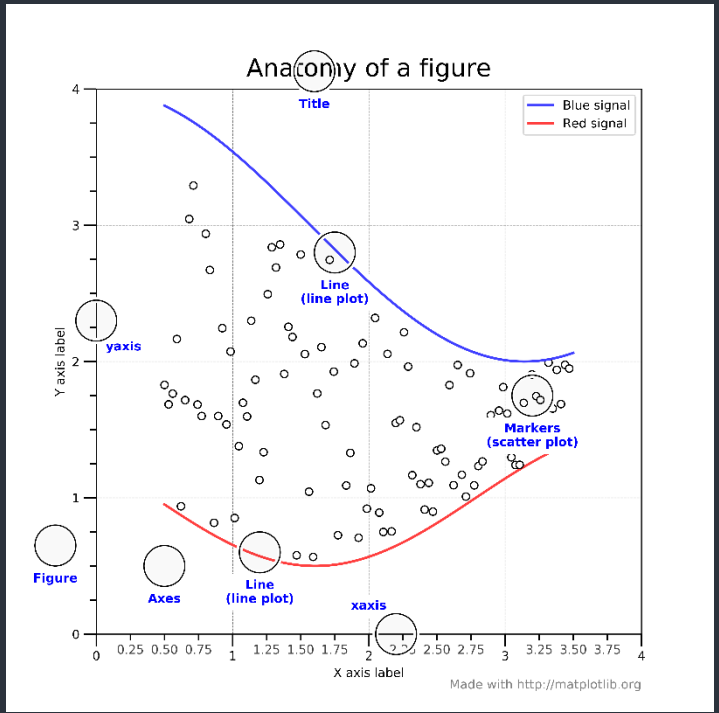
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Line
(line plot)
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
Y axis label
X axis label
Made with http://matplotlib.org

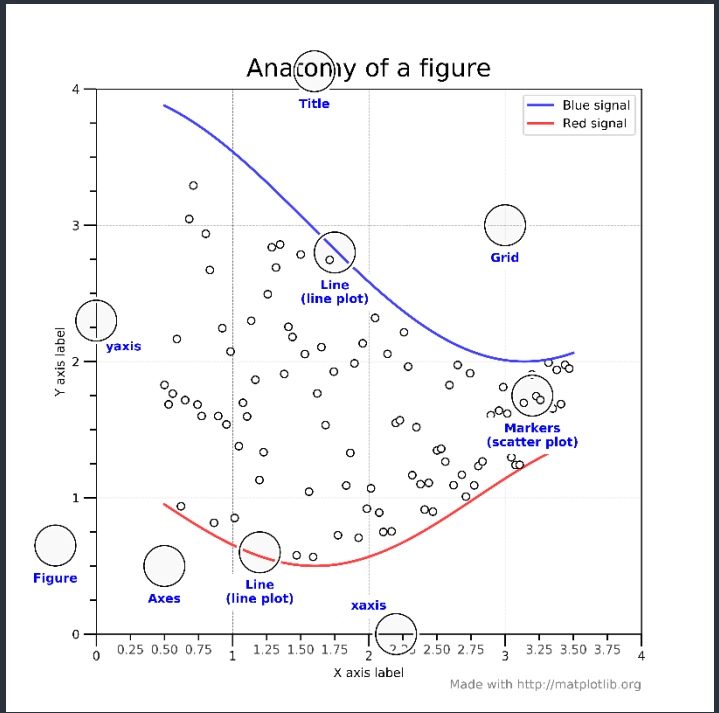
Matplotlib


Anatomy of a figure
Blue signal
Red signal
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
Made with http://matplotlib.org

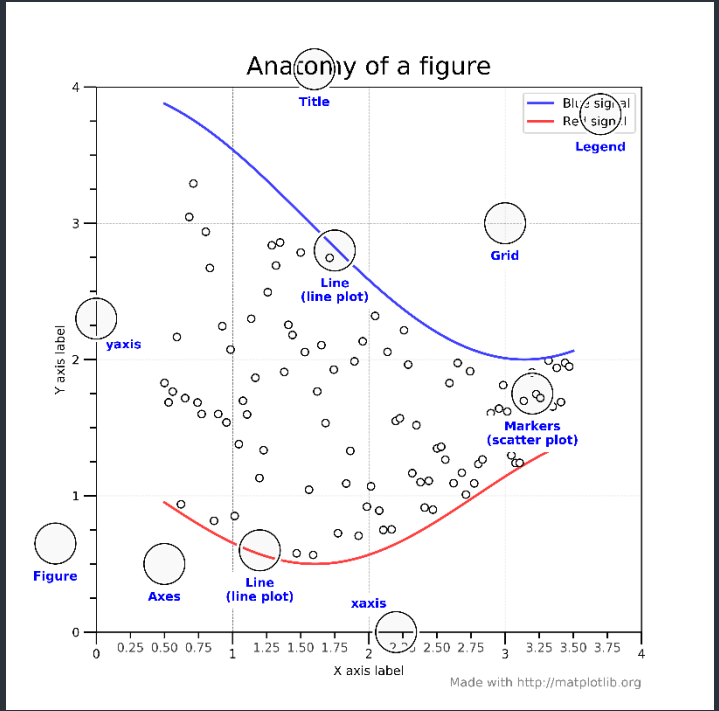
Matplotlib


Anatomy of a figure
Title
Blue signal
Red signal
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
Made with http://matplotlib.org

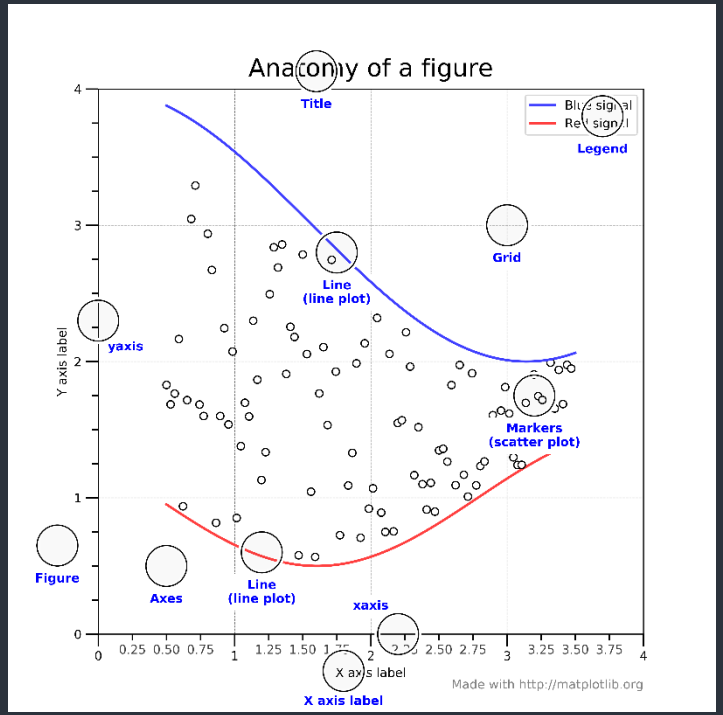
Matplotlib


Anatomy of a figure
Title
Blue signal
Red signal
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
Made with http://matplotlib.org

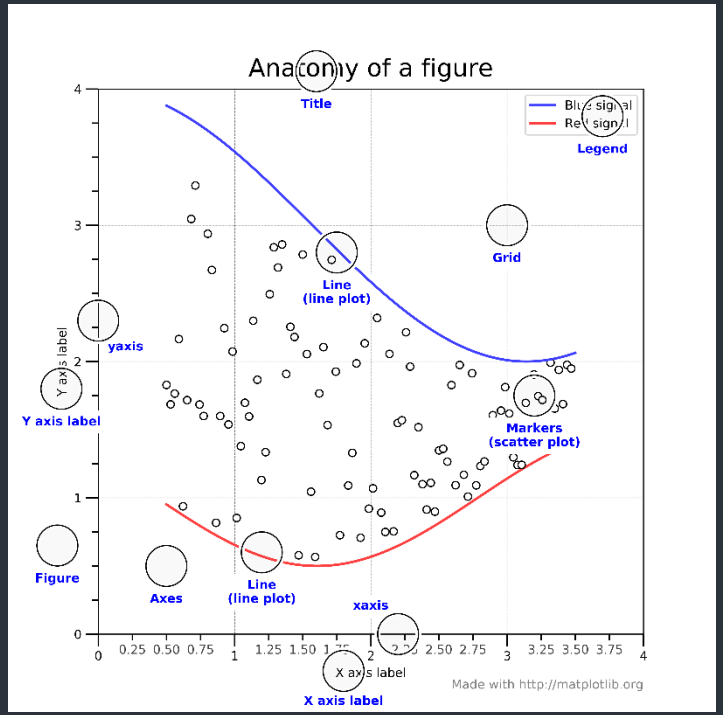
Matplotlib


Anatomy of a figure
Title
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
Made with http://matplotlib.org

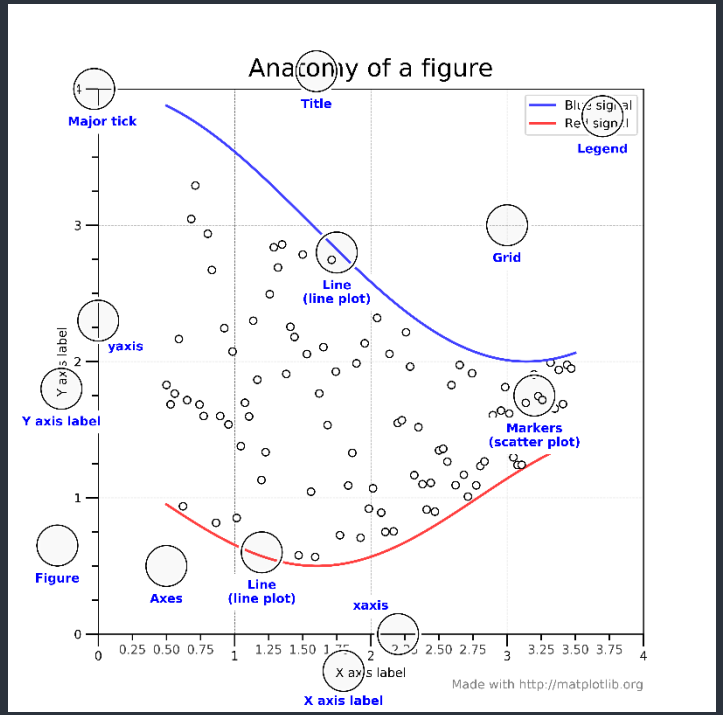
Matplotlib


Anatomy of a figure
Title
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
X axis label
Made with http://matplotlib.org

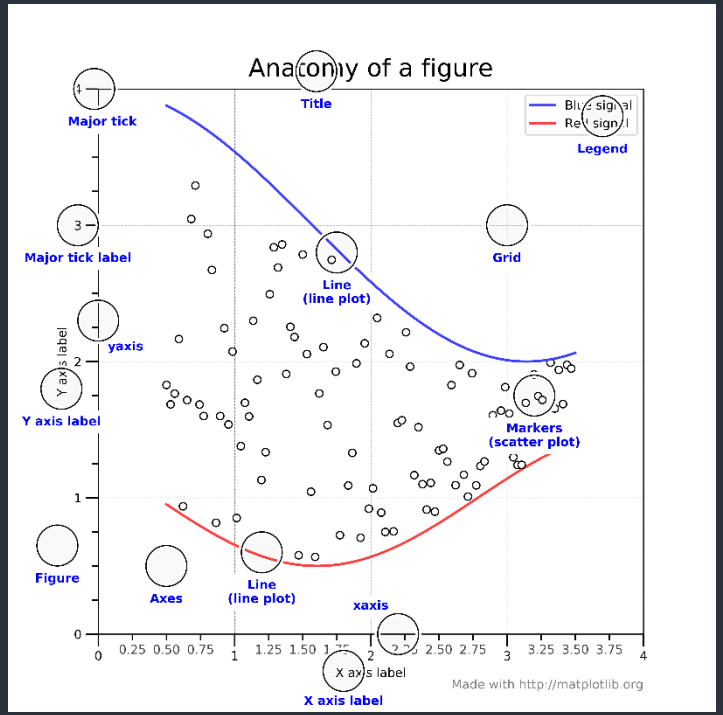
Matplotlib


Anatomy of a figure
Title
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
X axis label
Made with http://matplotlib.org

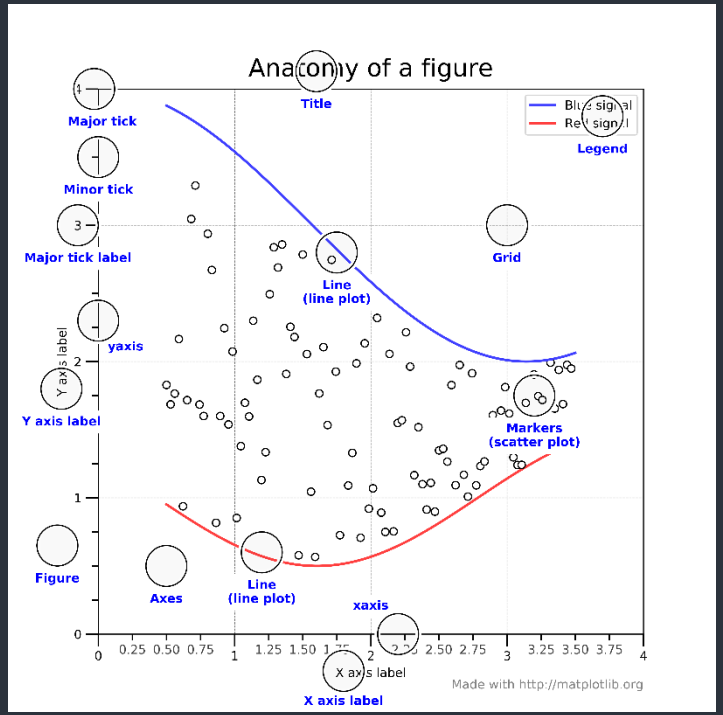
Matplotlib


Anatomy of a figure
Title
Major tick
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
Made with http://matplotlib.org

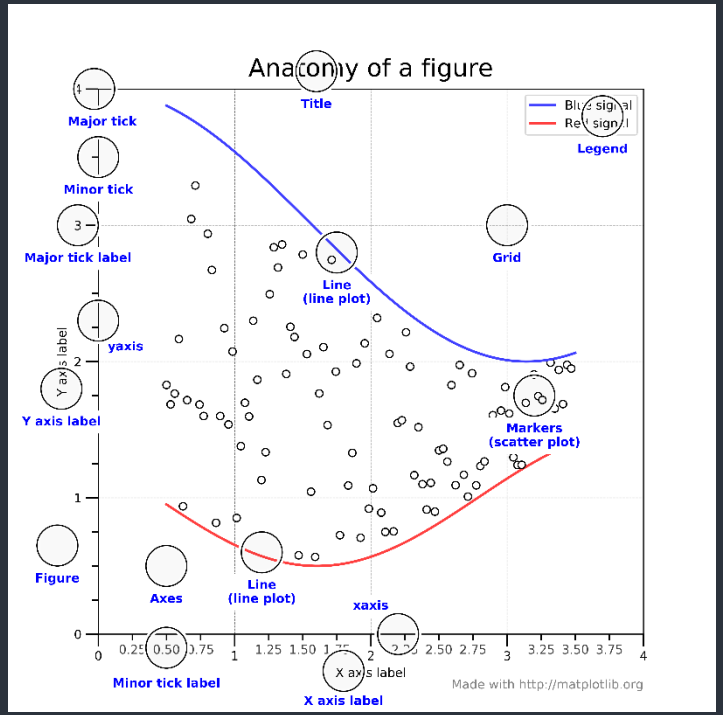
Matplotlib


Anatomy of a figure
Title
Major tick
Legend
Major tick label
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
X axis label
Made with http://matplotlib.org

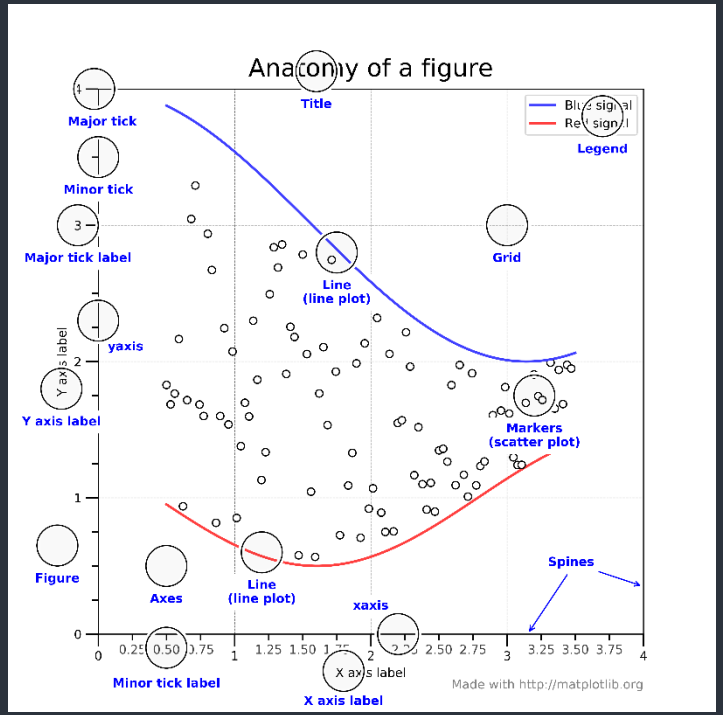
Matplotlib


Anatomy of a figure
Title
Major tick
Minor tick
Major tick label
yaxis
Y axis label
Figure
Axes
Line
(line plot)
Legend
Grid
Markers
(scatter plot)
xaxis
X axis label
Made with http://matplotlib.org

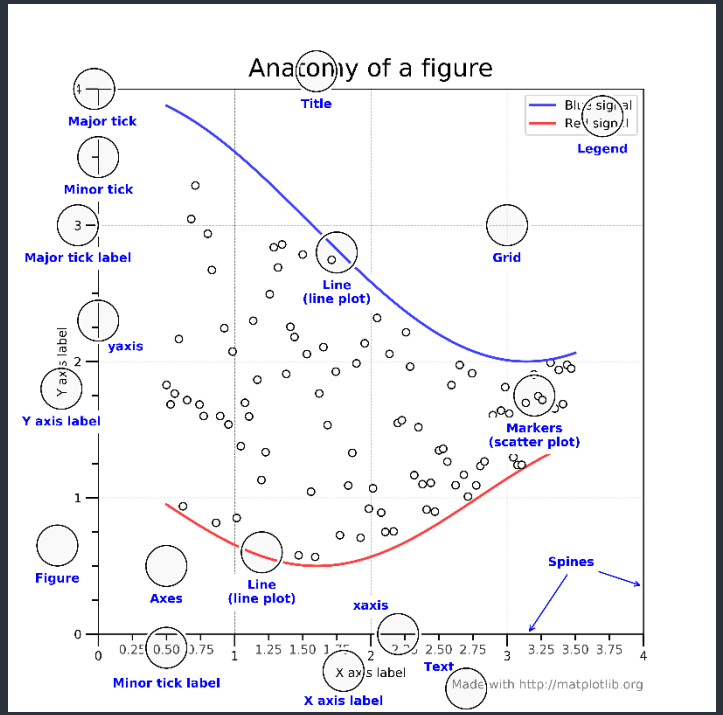
Matplotlib


Anatomy of a figure
Title
Major tick
Minor tick
Major tick label
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
xaxis
Minor tick label
X axis label
X axis label
Made with http://matplotlib.org

Matplotlib


Anatomy of a figure
Title
Major tick
Minor tick
Major tick label
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
Spines
xaxis
Minor tick label
X axis label
X axis label
Made with http://matplotlib.org

Matplotlib


Anatomy of a figure
Title
Major tick
Minor tick
Major tick label
Legend
Grid
Line
(line plot)
yaxis
Y axis label
Markers
(scatter plot)
Figure
Axes
Line
(line plot)
Spines
xaxis
Text
Minor tick label
X axis label
X axis label
Made with http://matplotlib.org

Practical

Notebook

Assignments

In +/- 20 min you will have to produce a plot with your group. One person in the group shares the screen of a notebook via RAN-S, Google Colab, Kaggle within the group. Together you will come to a plot and will present it after the time has finished.

I. Train delays
II. Bechdel Index

Assignment I

Not just in the Netherlands, but also in France, train delays regularly happen. The Lyon-Part-Dieu station has the reputation to be the worst station when it comes to trains departing late. You have been given monthly average data for French train network for the years 2015-2018. The goal is to show whether or not the reputation of Lyon-Part_Dieu is deserved.

https://github.com/DeNederlandscheBank/OpenSourceWorkshop/tree/main/Matplotlib/Assignment

The data: Data/full_trains.csv

A small start: Trains_begin_I.ipynb

Assignment II

To get a feel for female representation in movies, the Bechdel test has been developed. According to the Bechdel test a movie is awarded the following ratings:

- 0 = Less then 2 women in the movie
- 1 = At least two women in the movie
- 2 = The women need to talk to each other
- 3 = They need to talk to each other about something other than a man

Show how the Bechdel test scores vary across the years.
https://github.com/DeNederlandscheBank/OpenSourceWorkshop/tree/main/Matplotlib/Assignment
The data: Data/Bechdel.json
A small start: Bechdel_begin.ipynb

Survey

Link