

Visualization - Exercise 01

Silas Hoffmann
22234721

Katharina Walter
21674744

Leonhard Braun
21679325

May 4, 2023

1 Exercise 1.1: applying Tufte's design principles

1.1 Discussion - *baseball.png*

1.1.1 Tufte - Elements that are not data ink

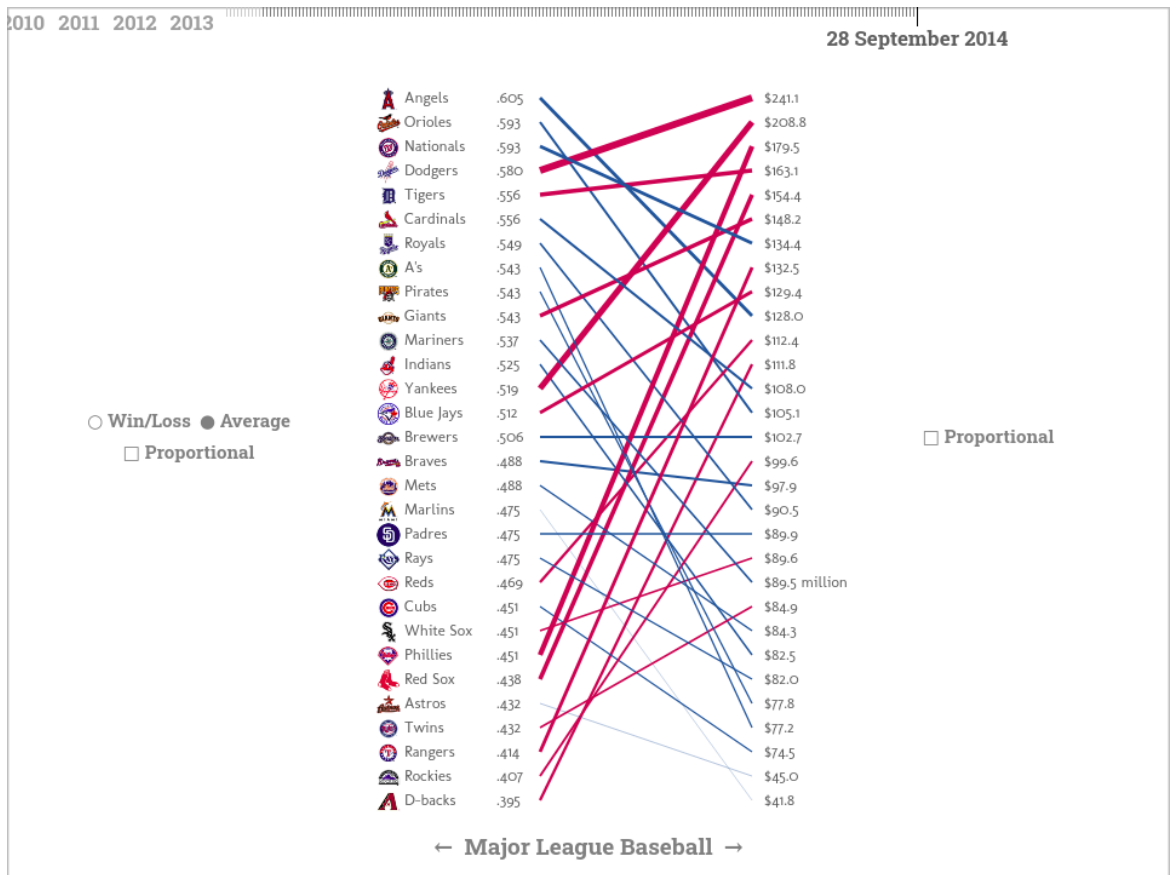
- Arrows in the label
- Time-axis
- Team Logos
- the colors in general since they do not add any meaning

1.1.2 Dos and Dont's

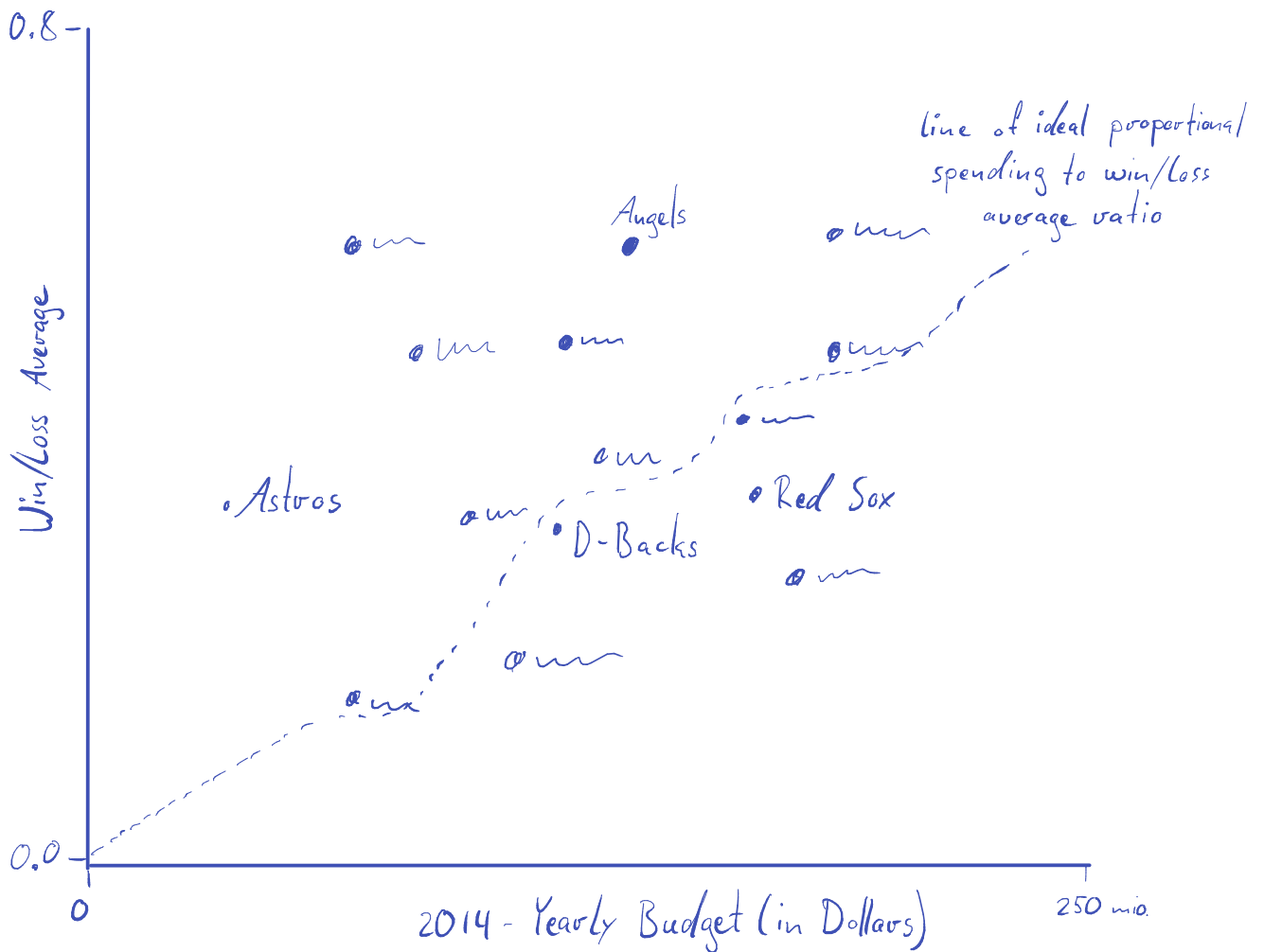
- **show full y-axis:** not present
- **consistent x-axis intervals:** not present
- **Edward Tufte in a nutshell - remove clutter:** is full of clutter, not very neatly organized
- **highlight what's important:** in essence the main points are highlighted, but it can be conveyed much simpler
- **sorting:** Data is sorted
- **do not use 3d or other visual effects:** no 3d effects
- **direct labeling where possible:** done correctly
- **avoid pie charts:** done correctly - isn't a pie chart
- **avoid stacked charts:** done correctly - isn't a stacked chart
- **do not use maps for everything with spatial dimension:** done correctly
- **avoid animations, use small multiples:** done correctly - isn't an animation
- **show level of confidence:** accurate, although not explicitly given
- **tell the 'why' and 'how':** not implemented, hard to understand
- **how to treat missing data:** not indicated
- **do not confuse causation and correlation:** not present
- **do not compare apples with oranges:** at first glance it is unclear what the data should display, but if the underlying article examines the displayed coherence it can be correctly used
- **adjust for inflation:** all data points come from the same year - no adjustment necessary
- **do not forget color deficiency:** no information is bound to a color

The following figure displays how we would implement the principles from the lecture.

baseball.pug



Major League Baseball



source:

1.2 Discussion - *energy.png*

1.2.1 Tufte - Elements that are not data ink

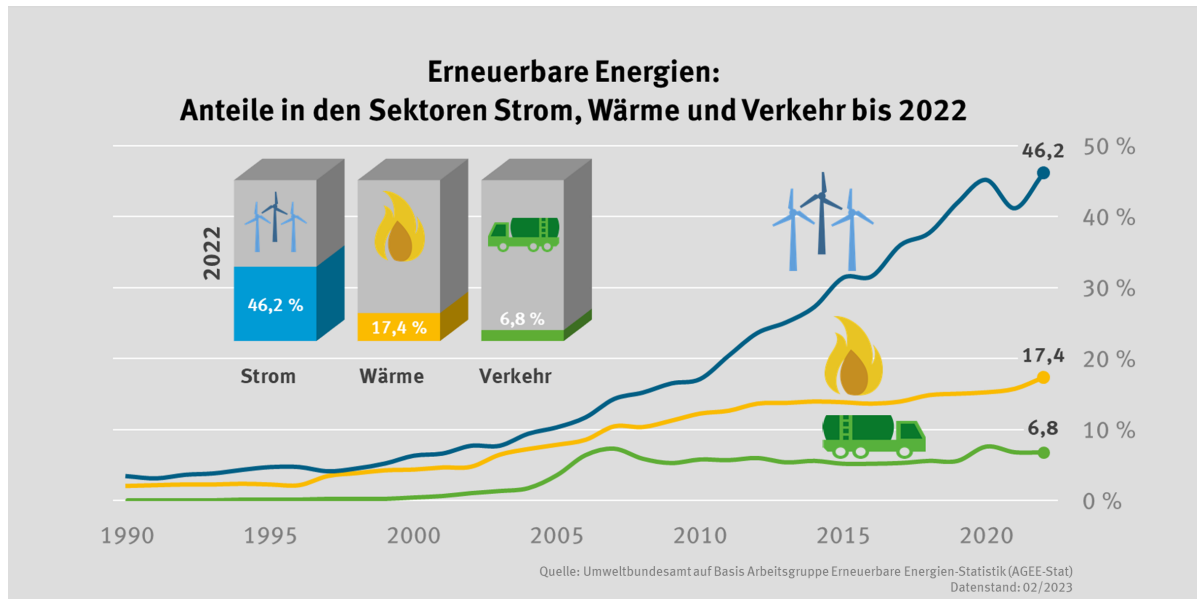
- The logos displaying the different sectors
- 3d Animation

1.2.2 Dos and Dont's

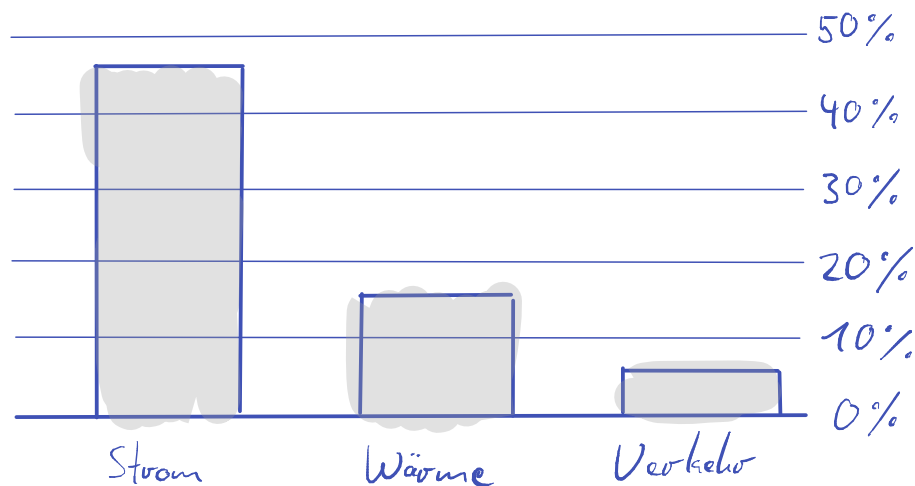
- **show full y-axis:** implemented correctly, although maybe it would be better to ramp up the y-axis to 100 %
- **consistent x-axis intervals:** present and consistent
- **Edward Tufte in a nutshell - remove clutter:** the chart actually displays two separate graphs in one
- **highlight what's important:** done correctly
- **sorting:** Data is sorted
- **do not use 3d or other visual effects:** 3d effects present - could be improved
- **direct labeling where possible:** done partially through the images - could be done better (with text)
- **avoid pie charts:** done correctly - isn't a pie chart
- **avoid stacked charts:** done correctly - isn't a stacked chart
- **do not use maps for everything with spatial dimension:** done correctly
- **avoid animations, use small multiples:** done correctly - isn't an animation
- **show level of confidence:** accurate, although not explicitly given
- **tell the 'why' and 'how':** since the sources is cited, this could be interpreted as the 'how' but the 'why' is missing
- **how to treat missing data:** not indicated
- **do not confuse causation and correlation:** not present
- **do not compare apples with oranges:** all data-points seem compatible
- **adjust for inflation:** not necessary since the data does not display monetary value
- **do not forget color deficiency:** no information is bound to a color. So, although color is used in the chart, a colorblind person still can convey the meaning of the graph without any problems.

The following figure displays how we would implement the principles from the lecture.

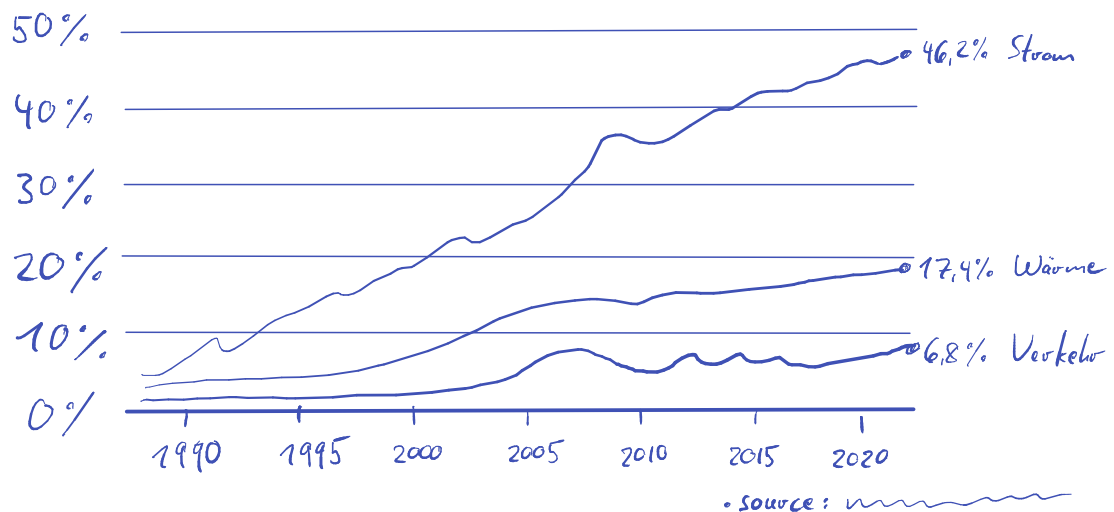
energy. pug



Erneuerbare Energien anteilig im Jahr 2022



Erneuerbare Energien anteilig bis zum Jahr 2022



1.3 Discussion - *machine-learning.png*

1.3.1 Tufte - Elements that are not data ink

- pattern in the posts seem unnecessary distracting

1.3.2 Dos and Dont's

- **show full y-axis:** is present, but no measurement unit provided
- **consistent x-axis intervals:** present and consistent
- **Edward Tufte in a nutshell - remove clutter:** seems ok, but the pattern in the post make the whole chart confusing
- **highlight what's important:** done correctly
- **sorting:** Data is sorted
- **do not use 3d or other visual effects:** 3d effects present - could be improved
- **direct labeling where possible:** implemented through a legend - could be improved
- **avoid pie charts:** done correctly - isn't a pie chart
- **avoid stacked charts:** done correctly - isn't a stacked chart
- **do not use maps for everything with spatial dimension:** done correctly
- **avoid animations, use small multiples:** done correctly - isn't an animation
- **show level of confidence:** not present
- **tell the 'why' and 'how':** not present - no source, no further information, no heading, legend without context highly confusing
- **how to treat missing data:** not indicated
- **do not confuse causation and correlation:** not present
- **do not compare apples with oranges:** due to lack of context not able to draw conclusion
- **adjust for inflation:** not necessary since the data does not display monetary value
- **do not forget color deficiency:** done correctly

The following figure displays how we would implement the principles from the lecture.

machine-learning.pyq

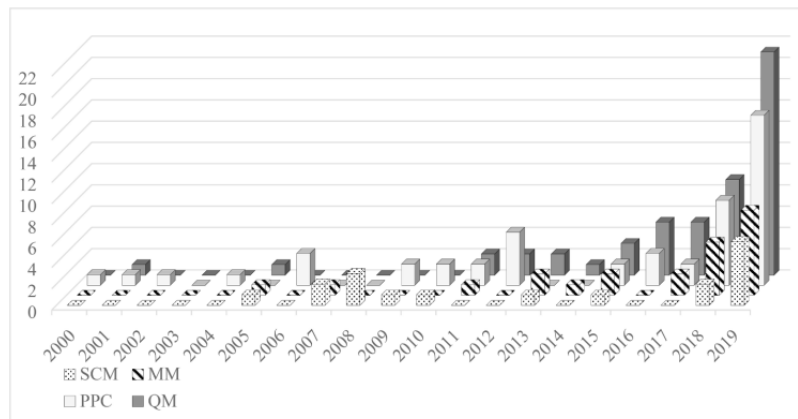
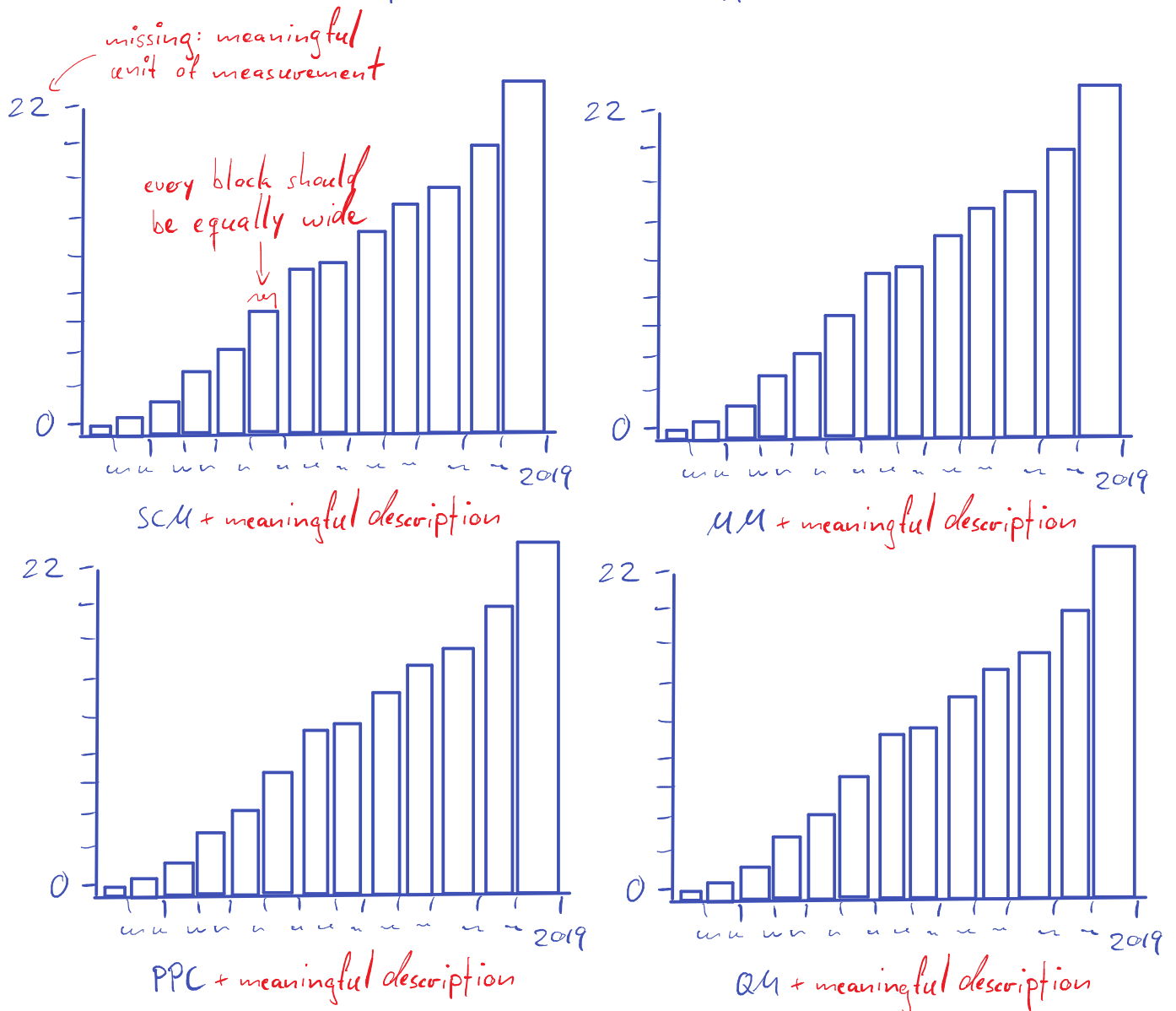


Fig. 3. The trend of publications, for each Application Domain (AD).

The trend of publications, for each Application Domain



+ Source for the data

2 Exercise 1.2: a first visualization task.

```
In [24]: #importing necessary libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from google.colab import files
import io
from sklearn.linear_model import LinearRegression

In [25]: uploaded = files.upload()

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving mpg-data[1].csv to mpg-data[1] (1).csv

In [26]: data = pd.read_csv(io.BytesIO(uploaded['mpg-data[1].csv']))

In [27]: data_hwy = data.applymap(str).groupby('class')['hwy'].apply(list).to_dict()
data_displ = data.applymap(str).groupby('class')['displ'].apply(list).to_dict()

In [29]: for key in data_displ:
X = np.array(data_hwy[key]).reshape(-1,1).astype(float)
Y = np.array(data_displ[key]).reshape(-1,1).astype(float)
linear_reg = LinearRegression()
linear_reg.fit(X,Y)
y_pred = linear_reg.predict(X)
plt.scatter(X,Y,label = key)
plt.plot(X,y_pred)
plt.xlabel("fuel efficiency on highways mi/l")
plt.legend()
plt.ylabel("displacement in l")
plt.title("Engine displacement and fuel efficiency in different car types")
```

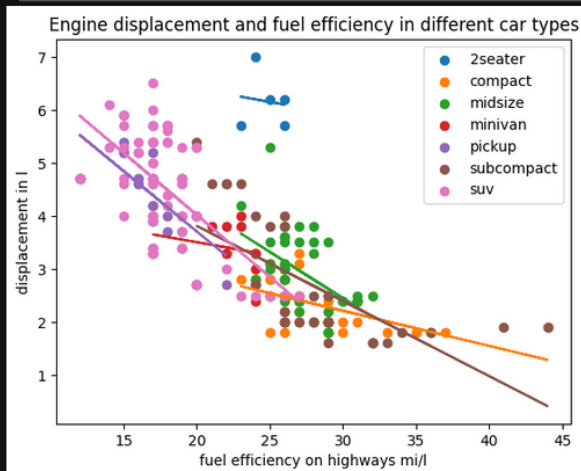


Figure 1: Screenshot from Jupyter Notebook