

Data Processing

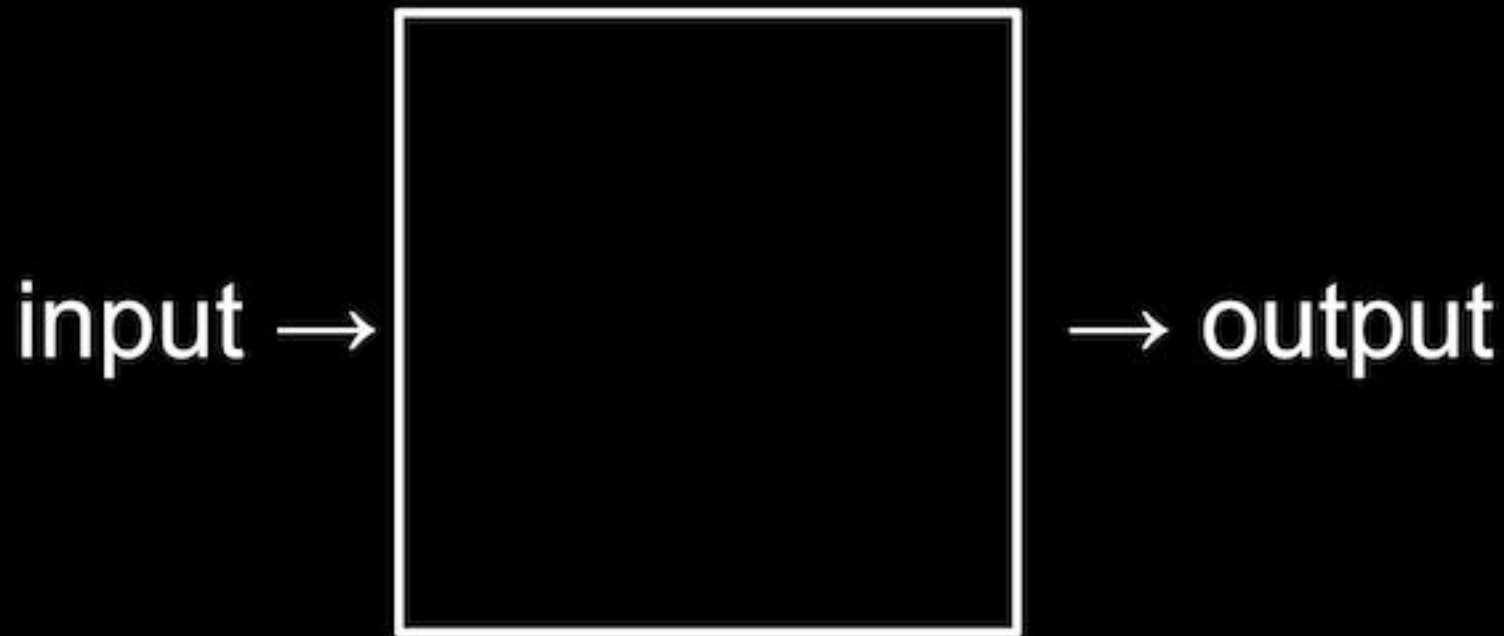
Python

Agenda

- Python
- Scraper
- Code review
- EDA
- JSON

Python

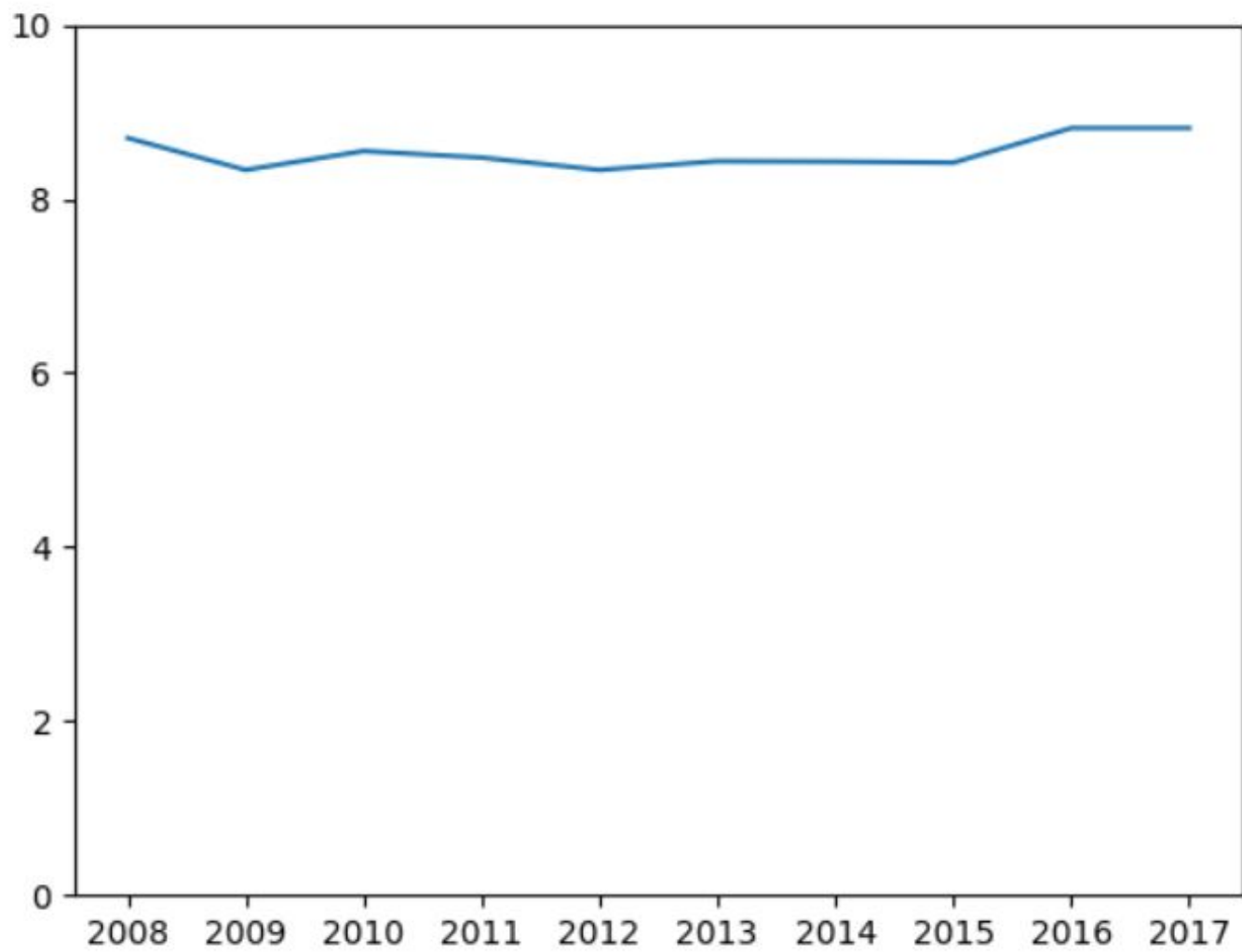
High-level language



Scraper

Webpage => .html => list/dictionary => .csv

Csv => datastructuur => visualisatie



Code review

- Groepjes van 2/3
- Vergelijk moviescraper.py en visualizer.py

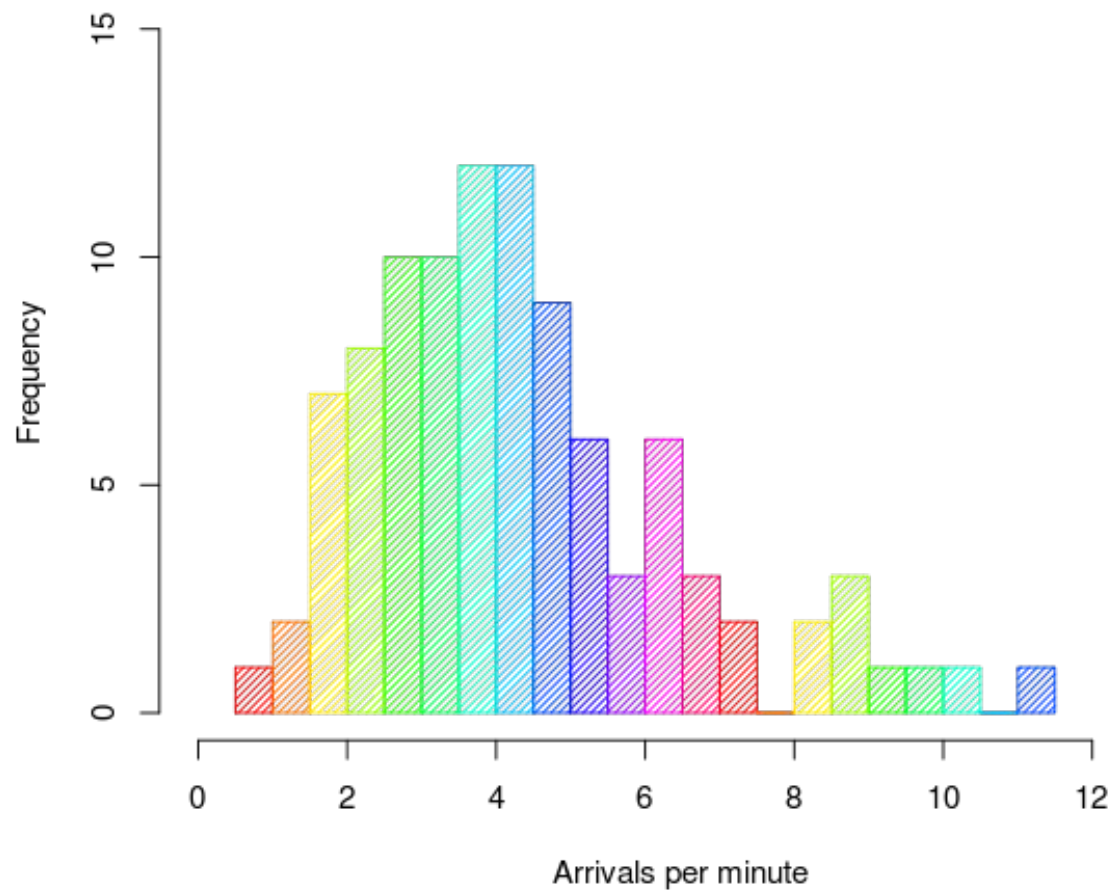
EDA (Exploratory Data Analysis)

- Homework/Week_2
- Download input.csv
- eda.py
- Parsing and preprocessing
- Central Tendency
- Five Number Summary
- JSON

Central Tendency

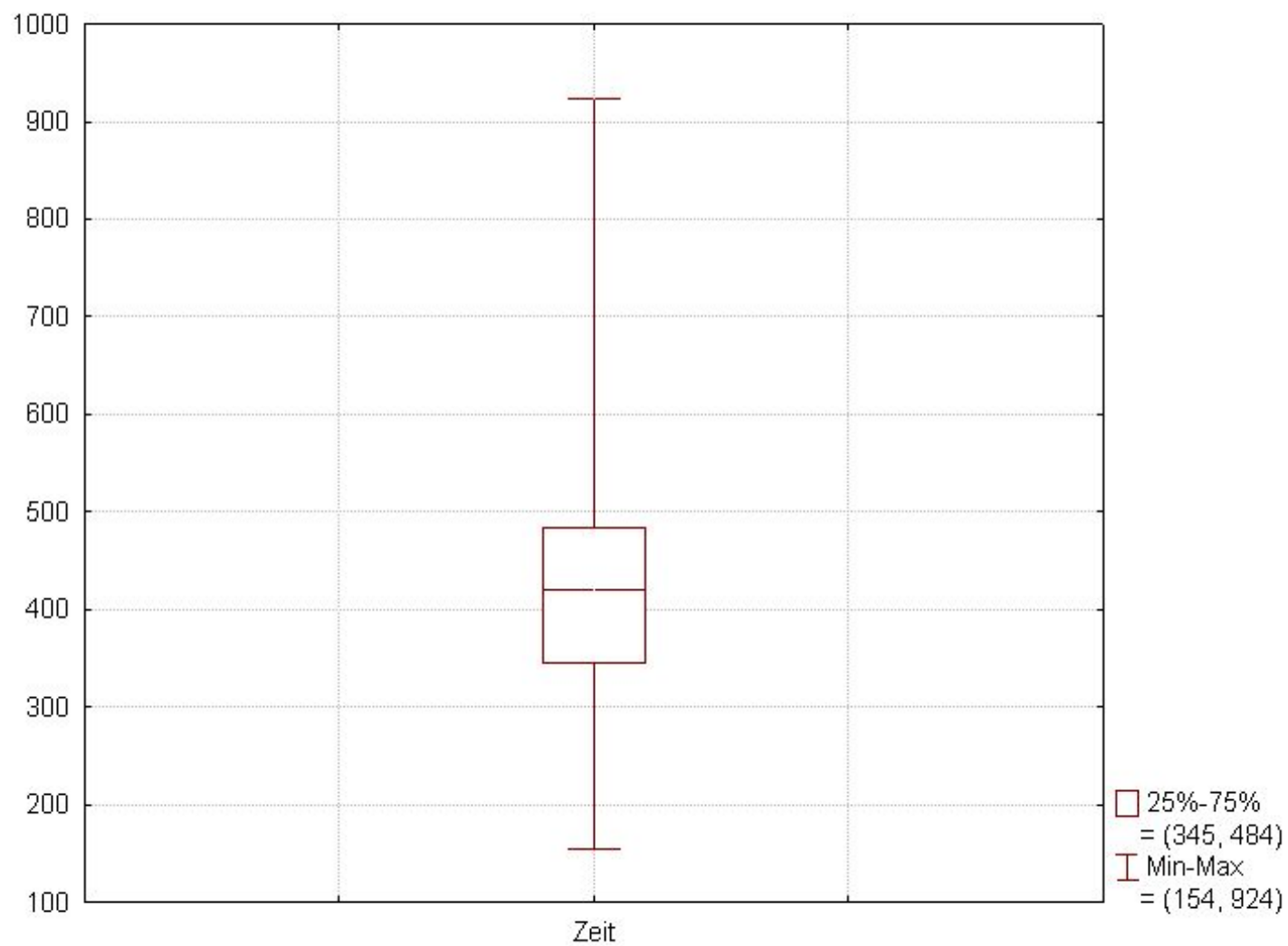
- Mean
 - Median
 - Mode
-
- Standard deviation

Histogram of arrivals



Five Number Summary

- Minimum
- First quartile
- Median
- Third quartile
- Maximum



JSON

- JavaScript Object Notation
- Object in JavaScript \approx Dictionary in Python
- Keys and values
- Example