CE491/591 Transportation Data Science

Python III: Data process and visualization

March 2, 2021

Content

- Common libraries you will use
  - Numpy, Scipy – computing
  - Pandas – Data processing
  - Matplotlib, seaborn - Visualization
  - Geopandas, networkx - Deal with shapefile and network data

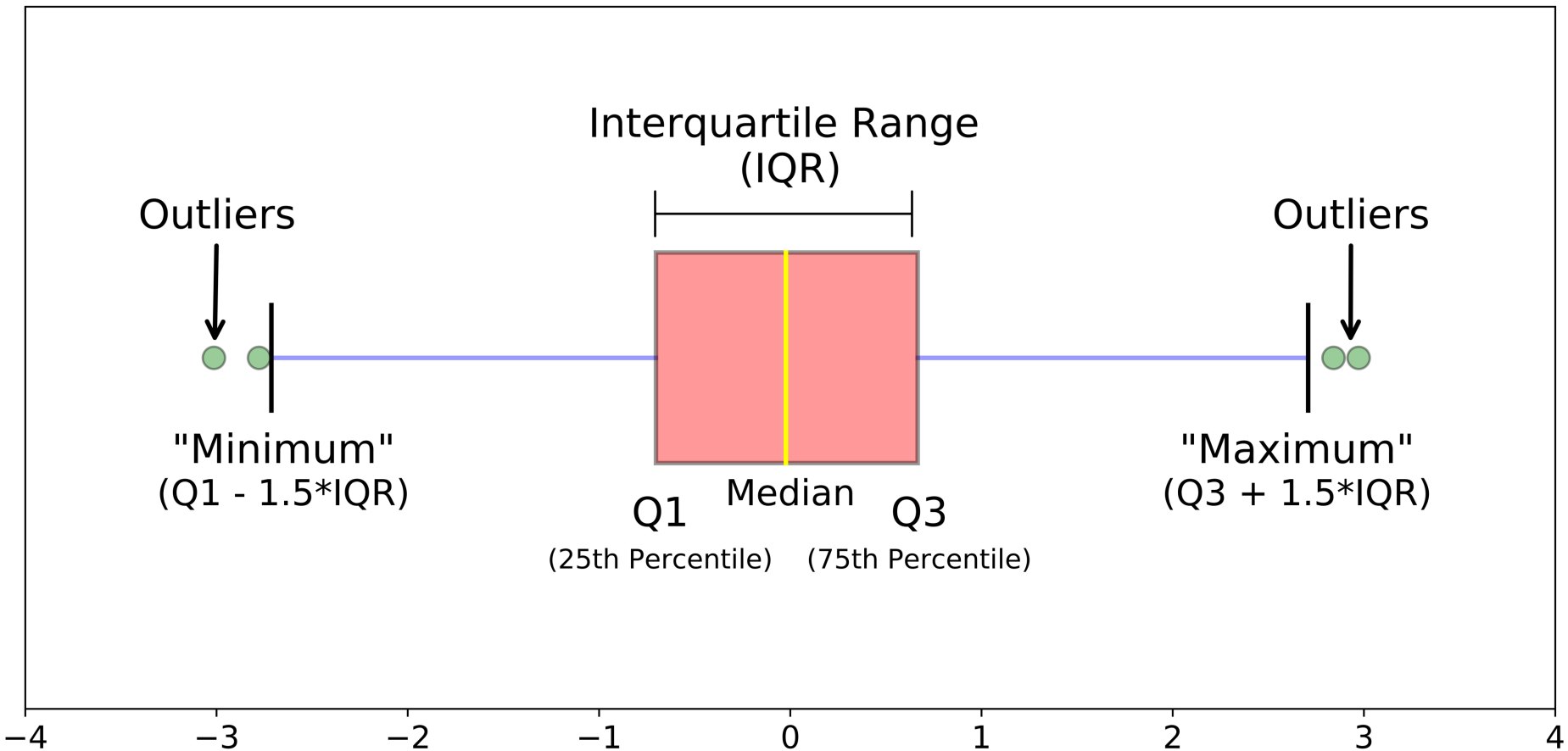
- Different type of visualizations
- Spatial processing
- Network processing

Visualization

- Single type of data
  - Histogram or probability density function (distribution)
- Pairwise comparison of data
  - Histogram, Scatter plot, 2D histogram, kernel-density estimation plot, or joint plot
- Compare data across categories

- **How to best display your information?**
- **How to choose the proper colormap?**

Box plot



Source: towardsdatascience.com

Violin plot

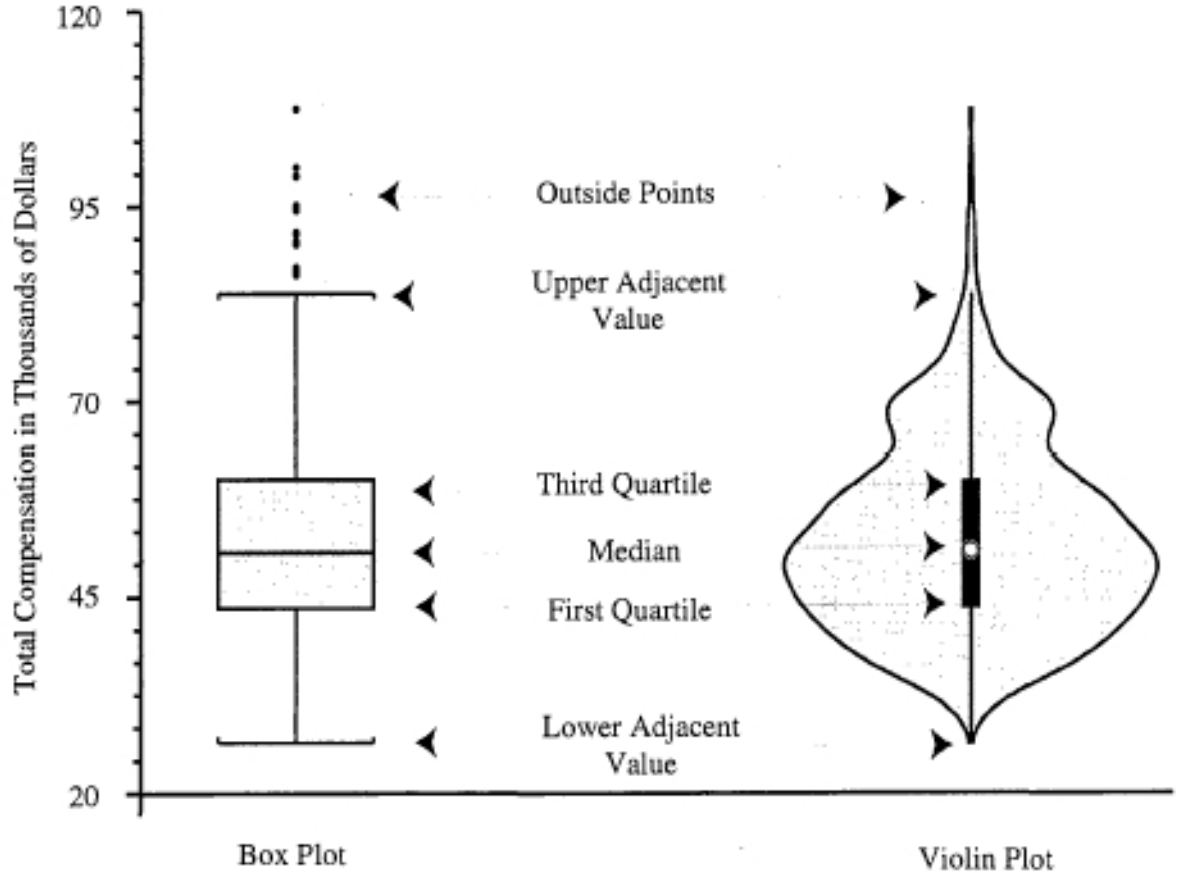
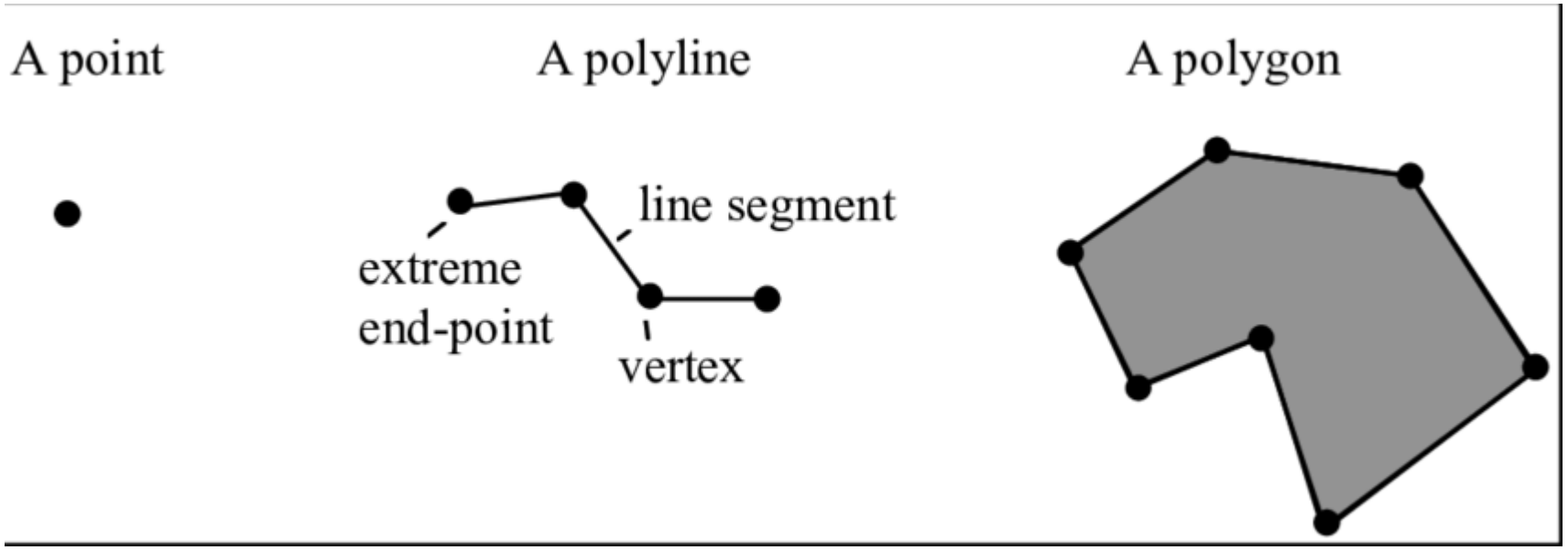


Figure 1. Common Components of Box Plot and Violin Plot. Total compensation for all academic ranks.

Source: towardsdatascience.com

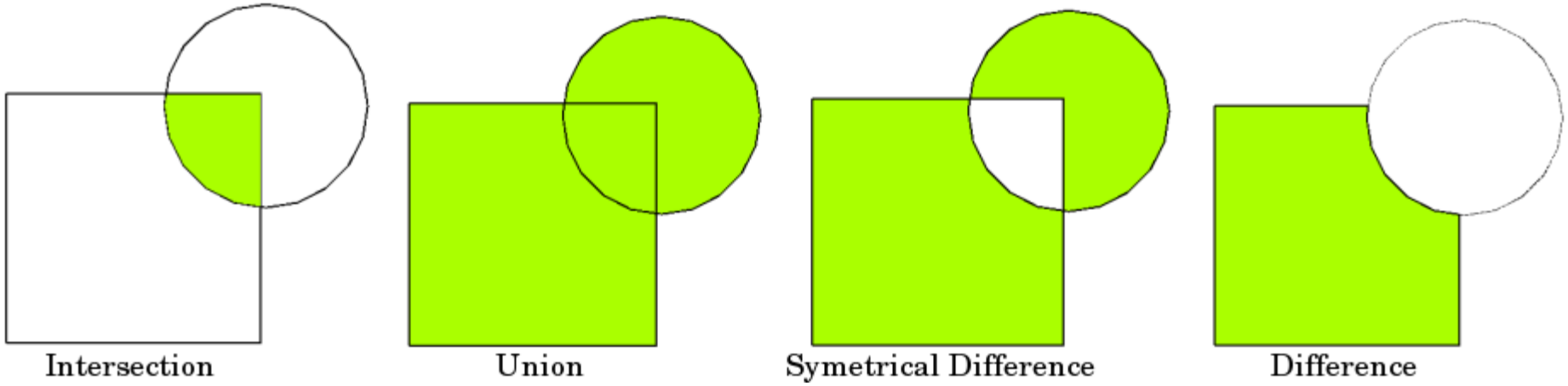
Shapefile

- Data format for spatial objects. You can use ArcGIS, QGIS…
- Type of spatial objects



Spatial processing

- Basic set operations for spatial data



Example – Tuscaloosa Area

- Road shapefile:
- https://catalog.data.gov/dataset/tiger-line-shapefile-2019-county-tuscaloosa-county-al-topological-faces-polygons-with-all-geoco

- Zip code area:
- https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

Example – Tuscaloosa Area

- Common projection system you would remember:
  - EPS: 4326 – used for lat, lon coordinates on sphere surface
  - Example: Google Earth

  - EPS: 3857 – used for Web Mercator projection on flat surface
  - Example: Google Maps, OpenStreetMap
  - Can be used to measure distance in meters!

Example – Tuscaloosa Area

- Visualize the shapefile
- Spatial processing (intersection)
- Obtain key geometry information (e.g., length, area, x,y of points)
- Calculate overlay statistics (road density in this case)
- Annotate the visualization