Practical Session 8

Foundations Spatial Data Science

UCL

Today Goals & Aims

Practical goals

- Get familiar with transforming data, selecting subsets suitable for plotting
- Map and plot elements of the Airbnb dataset
- Get familiar with Matplotlib and Seaborn libraries

Why are we doing this

- Visualisation of data is essential to explore and make sense of data.
- Visualisation is a way of communication. Communication skills is critical to successfully share your research and engage with diverse audiences
- Communicating your results is an important part of assessment 3. It is good to start thinking how you might do so now.

Today Goals & Aims

Term Calendar

	Weekly Topic		Lead	WORKSHOP Date (Monday)	PRACTICAL Date Groups 1,2,3 (Tuesday)	Groups 4,5,6 (Wednesday)
1	Getting Oriented	initiate	David, Nicolas	4 Oct	4 Oct	5 Oct
2	Foundations (Part 1)	initiate	Nicolas	11 Oct	11 Oct	12 Oct
3	Foundations (Part 2)	initiate	Nicolas	18 Oct	18 Oct	19 Oct
4	Objects & Classes	İnitiate	David	25 Oct	25 Oct	26 Oct
5	Numeric Data	engage	David	1 Nov	1 Nov	2 Nov
	Reading Week					
6	Spatial Data	engage	Nicolas	15 Nov	15 Nov	16 Nov
7	Textual Data	engage	Nicolas	22 Nov	22 Nov	23 Nov
8	Visualising Data	solve	David	29 Nov	29 Nov	30 Nov
9	Classifying Data	solve	David	6 Dec	6 Dec	7 Dec
10	Clustering Data	solve	Nicolas	13 Dec	13 Dec	14 Dec

Visualisation and Linked Data

Image source: https://realpython.com/python-statistics/
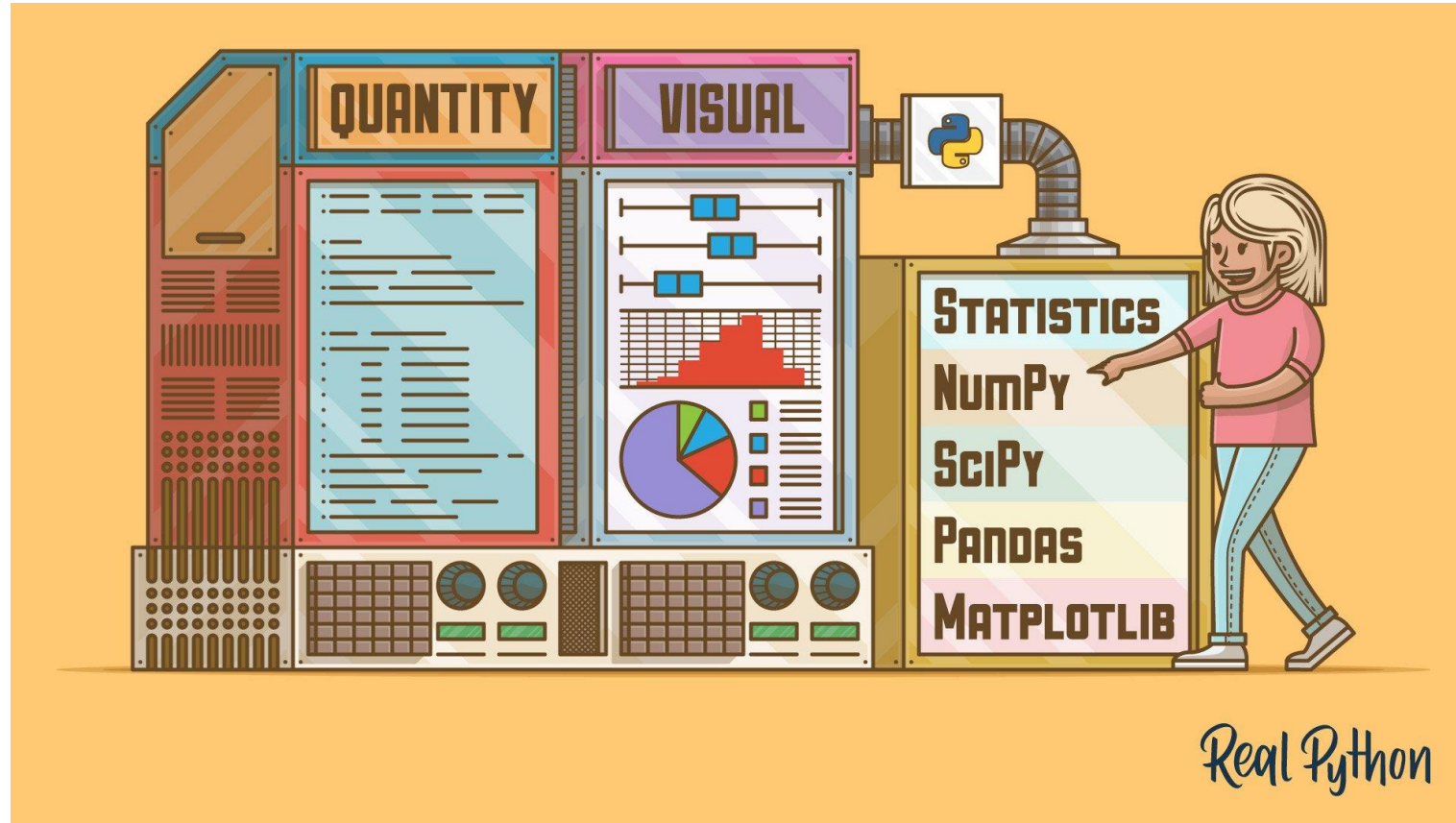
Today's Agenda

1. Where to find Spatial Data?
2. Merging Data - Spatial joins and Attribute Joins
3.
4.

Where to find Spatial Data!

1. Office for National Statistics UK (ONS) - Geoportal.
   View on: https://geoportal.statistics.gov.uk/

2. OpenStreetMap (OSM) - Crowdsourced spatial data. GIS data created by the public. View on: https://www.openstreetmap.org/

3. OpenTopography - High-resolution, Earth Science-oriented topography data. View on: https://opentopography.org/

Others:

1. Have a look at the blogpost "*10 free GIS Data Sources*", View on: https://gisgeography.com/best-free-gis-data-sources-raster-vector/

Merging Data

"Data merging is the process of combining two or more data sets into a single data set. Necessary process when raw data are stored in multiple data tables."
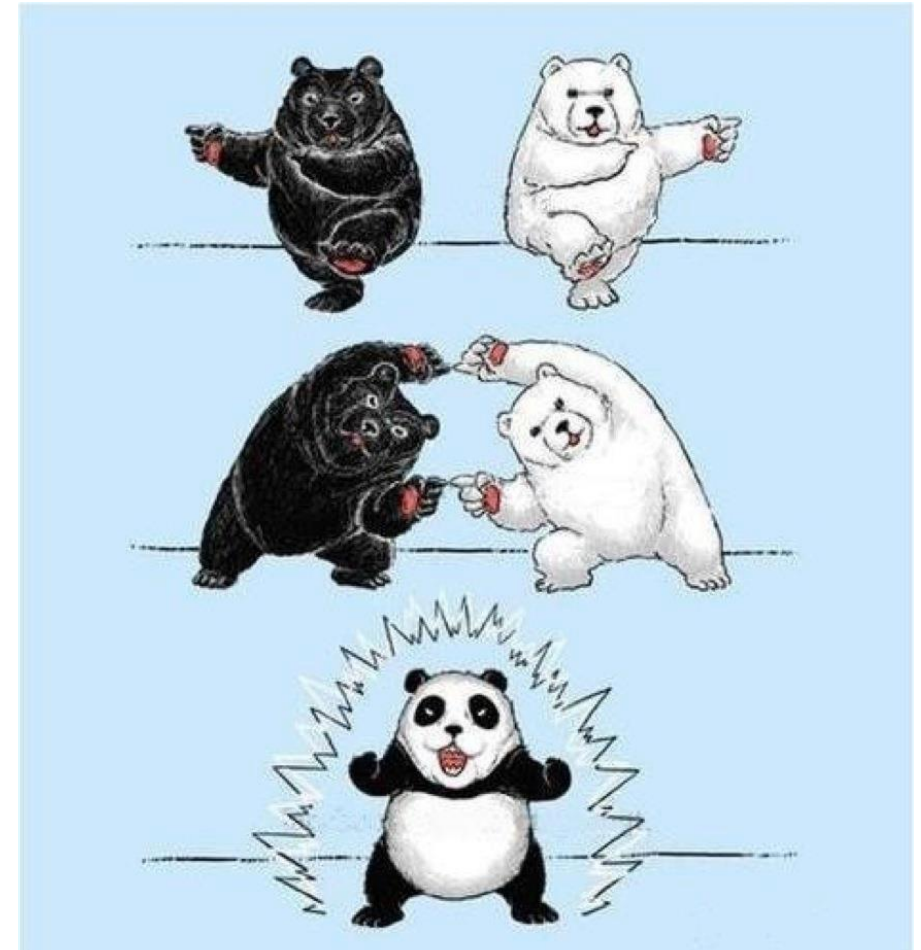
Spatial Joins vs. Attribute Joins

Spatial Joins

Observation from two
GeoSeries or **Geo**DataFrame.

Objects are combined based
on their spatial relationships.

Attribute Joins

Observation from a **Geo**Series
or **Geo**DataFrame is combine
with a Series or DataFrame.

Objects are combined based
on common variables.

Analogue to pandas merge.

Spatial Joins

Find the spatial relationship between geometries such as point, line and polygon.

How: **left, right, inner**.

Binary predicates. Valid value are determined ba the spatial index used: **intersects, contains, within**
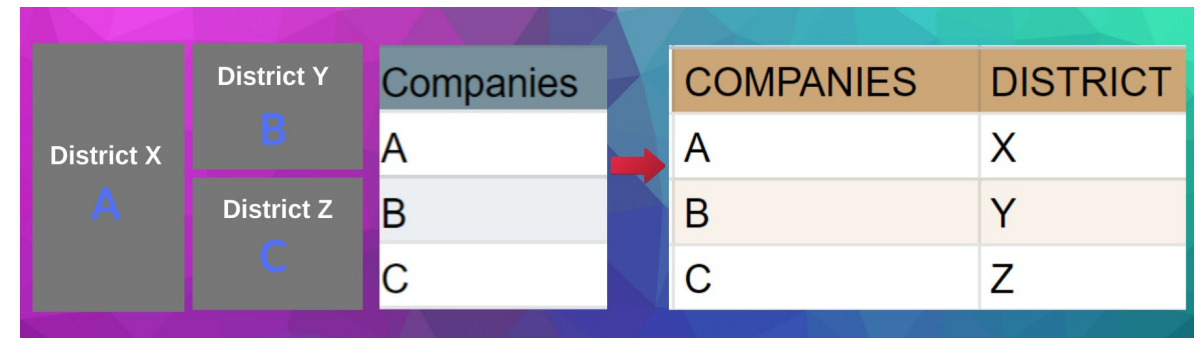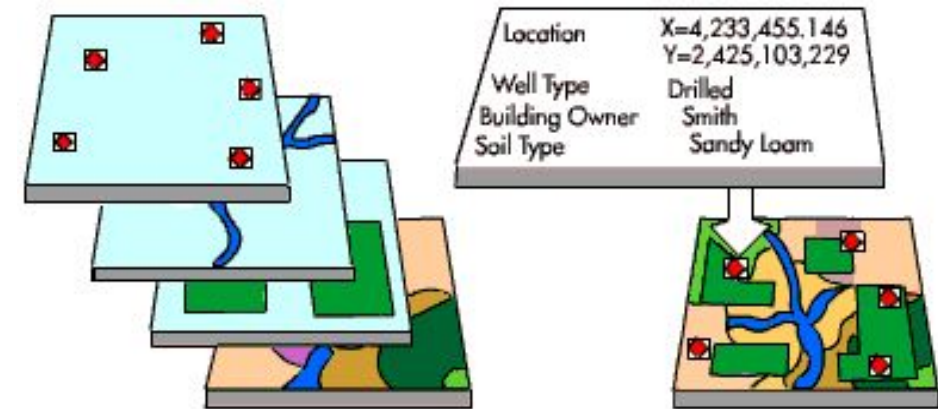
Spatial Joins

How do I know in which London borough are the Airbnb listings?

How many Airbnb listings are there for each London Boroughs?

Dataframe : airbnb_listings

| listings_name	city	geometry
modern apartment, 4-room with garden	London	Point(-0.0317 51.3214)
1-room in the attic of an old victorian house	London	Point(0.0926 51.3229)

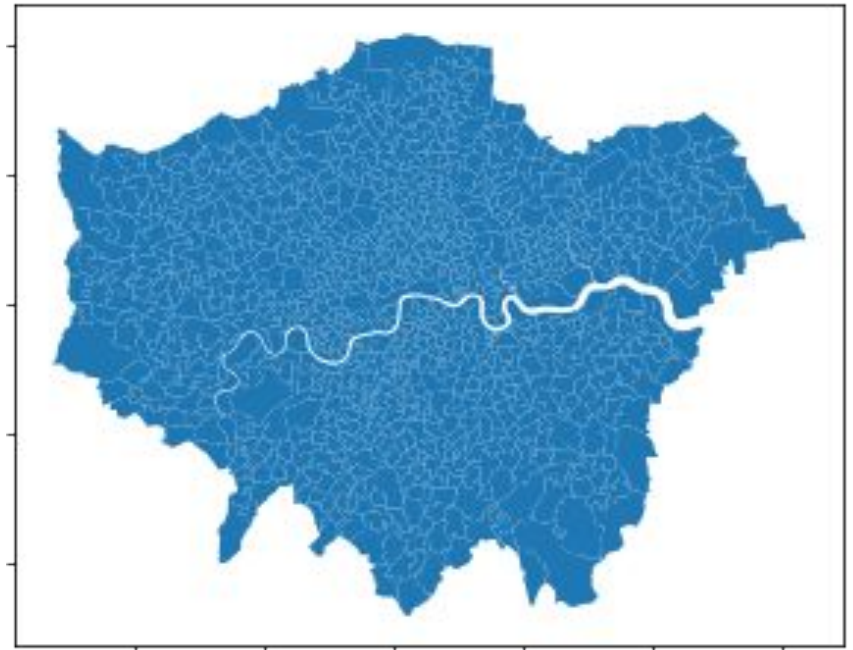Dataframe : ldn_borough

borough_name	city	geometry
Camden	London	Polygon ((-0.03870 51.500 -0.04290 51.4830))
Hackney	London	Polygon ((0.1130 51.2490 0.1070 51.4250))

Spatial Joins

ldn_sjoined_listings = gpd.sjoin(airbnb_listings, lnd_borough, op= "within")

Dataframe : airbnb_listings

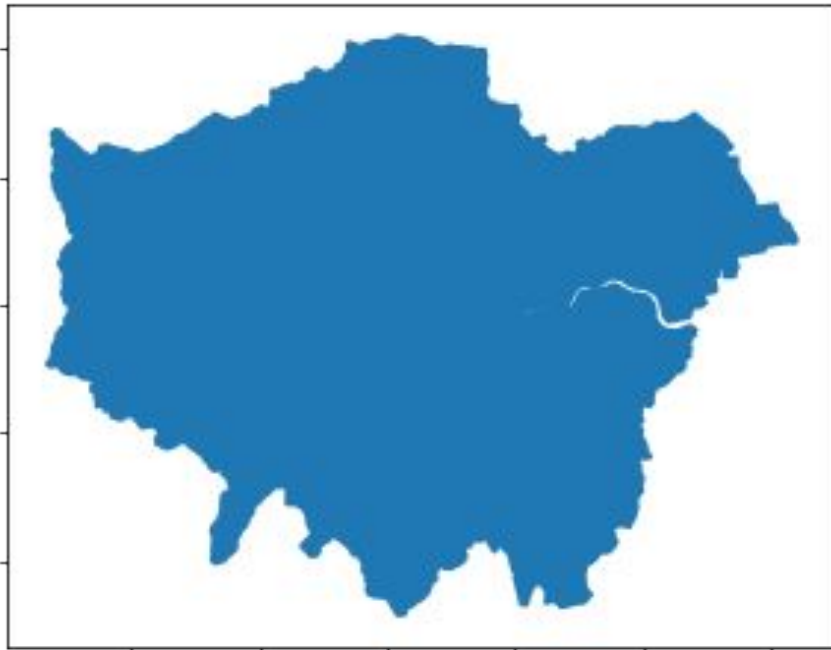| listings_name	city	geometry
modern apartment, 4-room with garden	London	Point(-0.0317 51.3214)
1-room in the attic of an old victorian house	London	Point(0.0926 51.3229)

Dataframe : ldn_borough

borough_name	city	geometry
Camden	London	Polygon ((-0.03870 51.500 -0.04290 51.4830))
Hackney	London	Polygon ((0.1130 51.2490 0.1070 51.4250))

Dataframe : ldn_sjoined_listings

listings_name	city	borough_name	geometry
modern apartment, 4-room with garden	London	Hackney	Point(-0.0317 51.3214)
1-room in the attic of an old victorian house	London	Camden	Point(0.0926 51.3229)

Spatial Aggregation

What is the granularity of the spatial data? What is the spatial level we are studying patterns? LSOA, Borough, County?
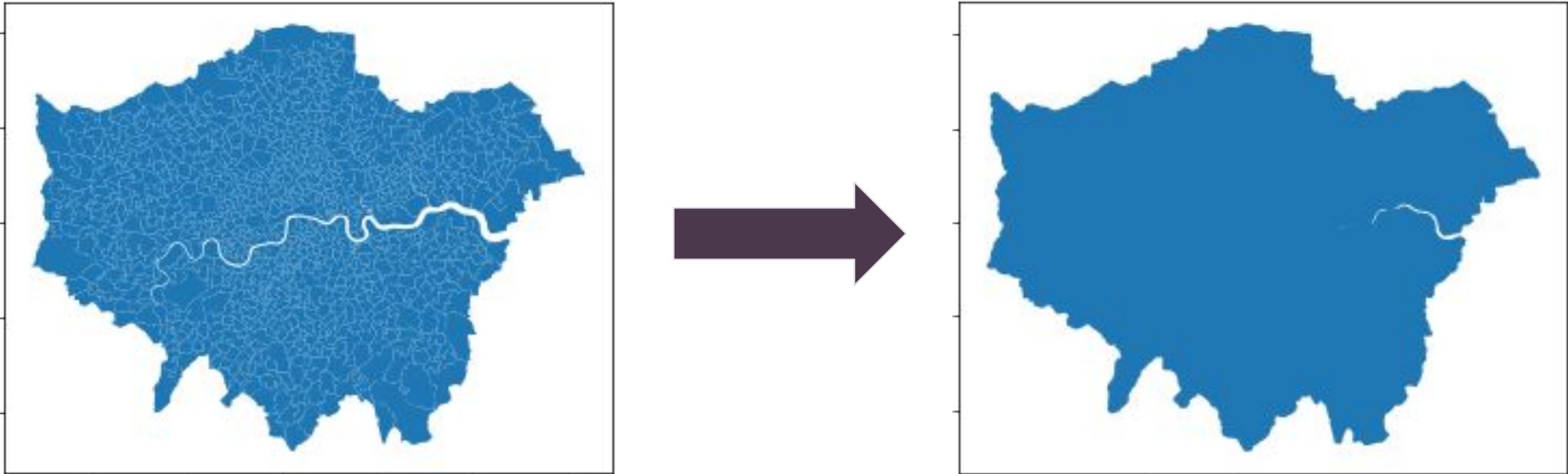
dissolve()

Dissolve all the geometries within a given group together in a single geometric feature.

What are the different .plot() options?

1. "area" is for area plots.
2. "bar" is for vertical bar charts.
3. "barh" is for horizontal bar charts.
4. "box" is for box plots.
5. "hexbin" is for hexbin plots.
6. "hist" is for histograms.
7. "kde" is for kernel density estimate.
8. "density" is an alias for "kde".
9. "line" is for line graphs.
10. "pie" is for pie charts.
11. "scatter" is for scatter plots

Image Link: https://towardsdatascience.com/visualizations-with-matplotlib-part-1-c9651008b6b8

Introduction to Seaborn

1. Plot background. How to style the plot grid?
   a. sns.set_style() -> 'whitegrid', 'darkgrid', 'dark', 'white'
   b. sns.set_theme(style= '', palette= '')
2. How to choose the plot?
   a. sns.CHOOSE_PLOT -> boxplot, lmplot, displot, catplot
3. How to visualise different data with different colors?
   a. sns.lmplot(x= 'median', y= 'lnd_borough', **hue**= 'aibnb_room_type', data=lnd_airbnb_data)
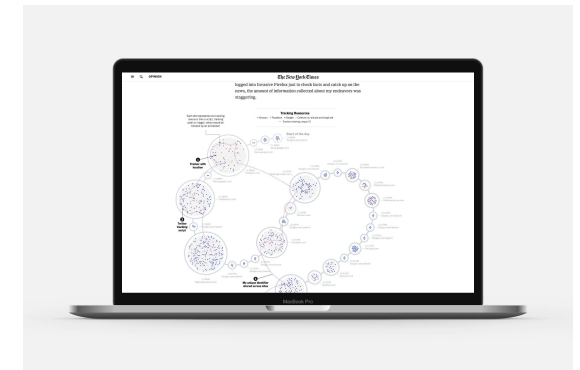
Data Visualisation - Information Designer

Visualisingdata by Andy Kirk

Work by Giorgia Lupi

Work by Nadieh Bremer

Next Week

Continue to explore the last part of the module. Review methods for problem solving, exploring, classifying data. Discuss the data space: transformation, dimensionality!

Week 9 READINGS - Dimension in Data and the Statistician

- B.D Bunday, *A Final Tale or You Can Prove Anything with Figures,* URL
- Harris (2019), *The Certain Uncertainty of University Rankings,* URL
- Cima (2014), *The Most and Least Diverse Cities in America, Priceonomics,* URL (PDF with Figures)
- Lu (2013), *Are statisticians cold-blooded bosses? a new perspective on the 'old' concept of statistical population* DOI

Time to practice !

Notes

1. Taks 1.1 - URL may give you an error. Change url to "https://raw.githubusercontent.com/jreades/fsds/master/data/2020-08-24-sample-listings-simple.csv"