Introduction to Data Science

Metro Health Department

NASHVILLE SOFTWARE SCHOOL

Goals for today

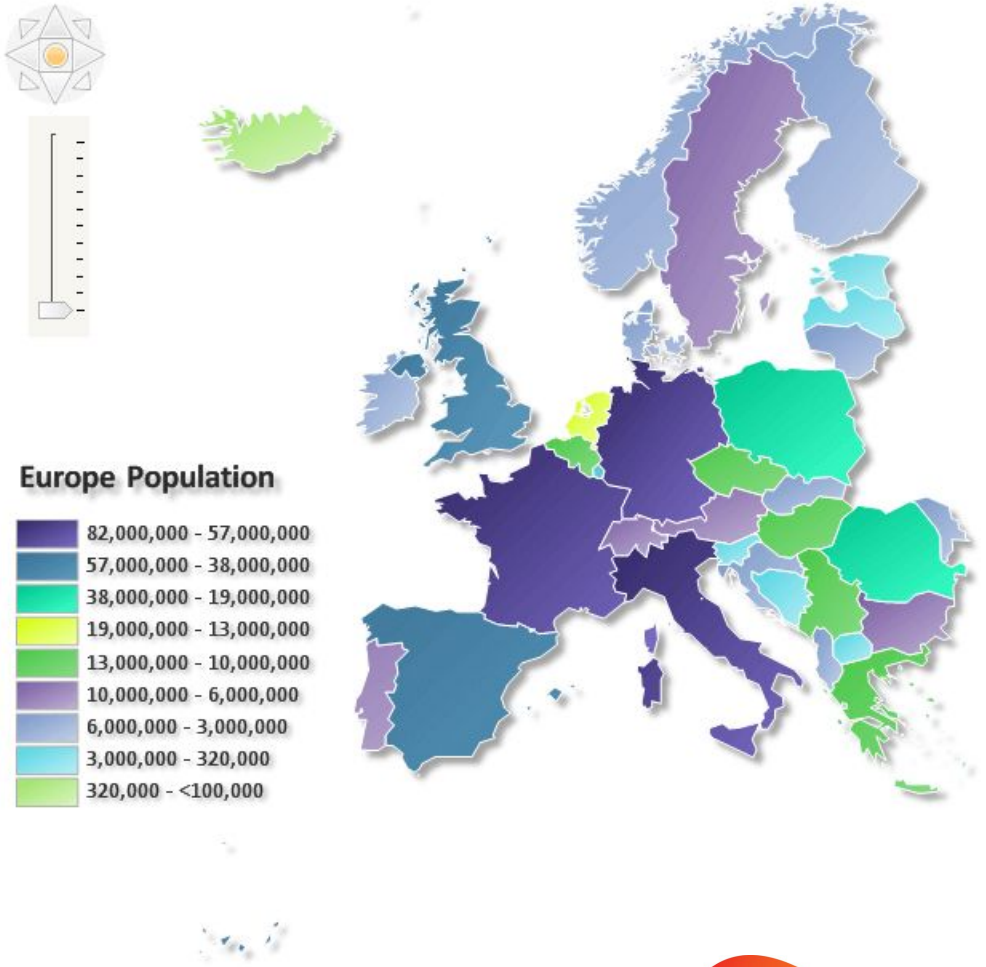
- **Finish additional practice in EDA notebook**
- **Learn about merging in *pandas***
- **Use folium to build a choropleth**
- **Start exploring the data with your group**

NASHVILLE SOFTWARE SCHOOL

Choropleths

A choropleth is a map where areas are colored or shaded according to the value of some aggregate statistic for that area (eg. average income, population density, unemployment rate, etc.)

We will create choropleths in Python by using the *folium* library, which you will most likely need to install.



NASHVILLE SOFTWARE SCHOOL

Group Project

Practice_NB_04.ipynb gave you a chance to explore possible associations between demographic characteristics and the number of COVID infections by zip code.

Your task after finishing Practice_NB_04 is to, with your group, continue to explore the data provided (or bring in other sources of data), come up with questions to ask, and put together some visualizations to show what you are able to find.

Reminder of the datasets you have:

- *COVID_CountByZipPerDate 03292021.csv* - number of COVID cases by zip code
- *vaccine_with_protocol.csv* - number of vaccinations by zip code
- *census_data.sqlite* - a SQLite database containing 4 tables of data obtained from the US Census Bureau

On our last class (August 16), each group will have the opportunity to show off your findings.

Questions?