

Geospatial Basics

Michele Coscia

First Year Project #1

February 8th, 2022

Lecture Plan

- 1) (February 8th) Intro
- 2) (Today) Geospatial Basics**
- 3) (February 15th) Estimating Associations
- 4) (February 17th) Multivariate Regression
- 5) (February 22nd) Interventions
- 6) (February 24th) Project Run Through
- 7) (March 1st) Q&A – Open Supervision
- 8) (March 3rd) Q&A – Open Supervision

Outline

- Resources
- Projections
- Choropleths

Resources

With some variable of interest


#1: Data about borders!

Border Resources

- Lots of great free ones
- For the course, we use <https://www.naturalearthdata.com/>
- Not just borders!

Admin 1 – States, Provinces

Internal administrative boundaries.




- [Download states and provinces](#) (13.92 MB) version 4.1.0
- [Download as scale ranks](#) (12.83 MB) version 4.1.0
- [Download without large lakes](#) (14.07 MB) version 4.1.0
- [Download boundary lines](#) (5.58 MB) version 4.1.0

[About](#) | [Issues](#) | [Version History](#) »

Populated Places

City and town points, from Tokyo to Wasilla, Cairo to Kandahar

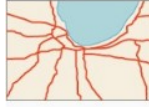


- [Download populated places](#) (2.68 MB) version 4.1.0
- [Download simple \(less columns\)](#) (751.96 KB) version 4.1.0

[About](#) | [Issues](#) | [Version History](#) »

Roads

Transportation.

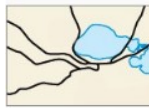


- [Download roads](#) (8.65 MB) version 4.0.0
- [Download North America supplement](#) (45.57 MB) version 4.0.0

[About](#) | [Issues](#) | [Version History](#) »

Railroads

Transportation.




- [Download railroads](#) (14.42 MB) version 4.0.0
- [Download North America supplement](#) (405.5 KB) version 4.0.0

[About](#) | [Issues](#) | [Version History](#) »

Airports

Transportation.



- [Download airports](#) (227.43 KB) version 4.1.0

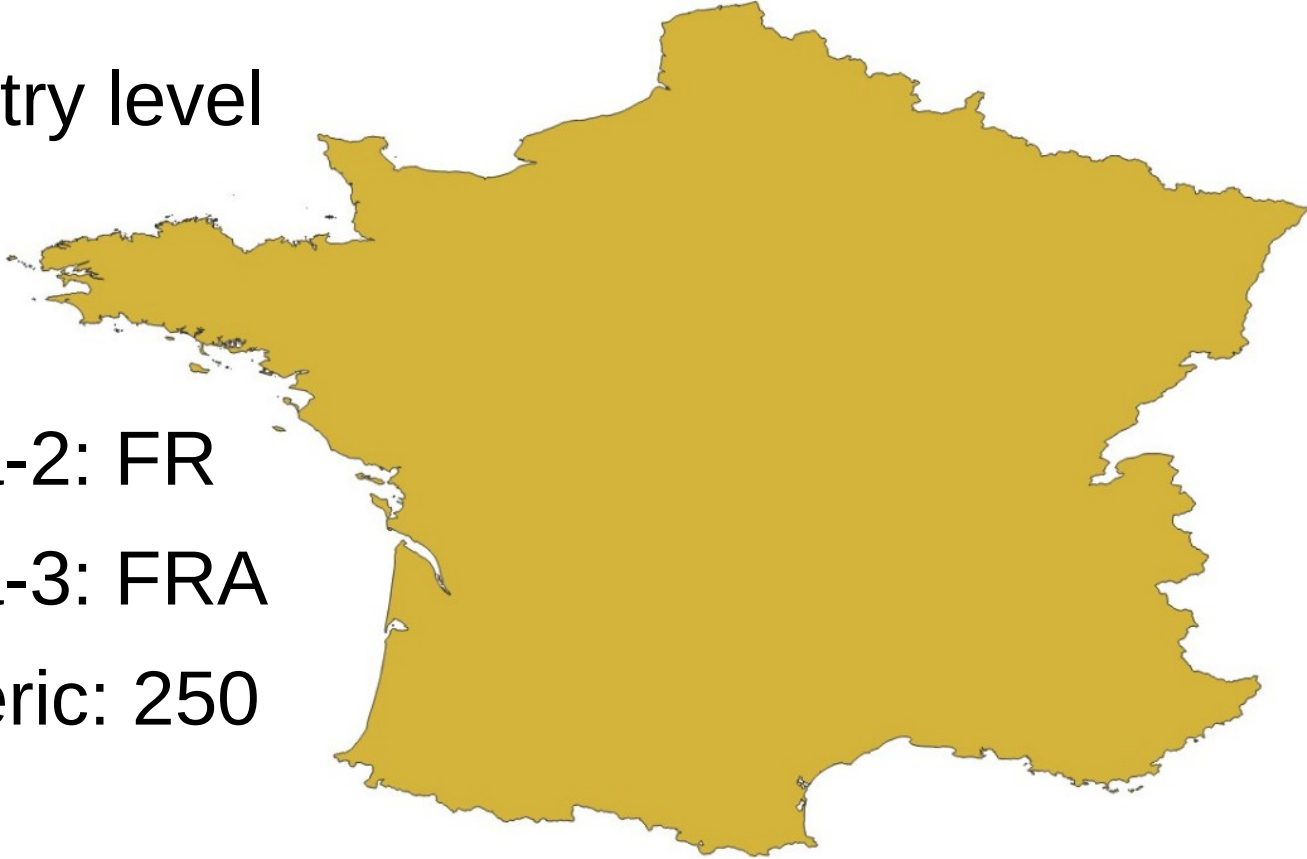
[About](#) | [Issues](#) | [Version History](#) »

GADM.org if
you really
want to go
deep into
borders...



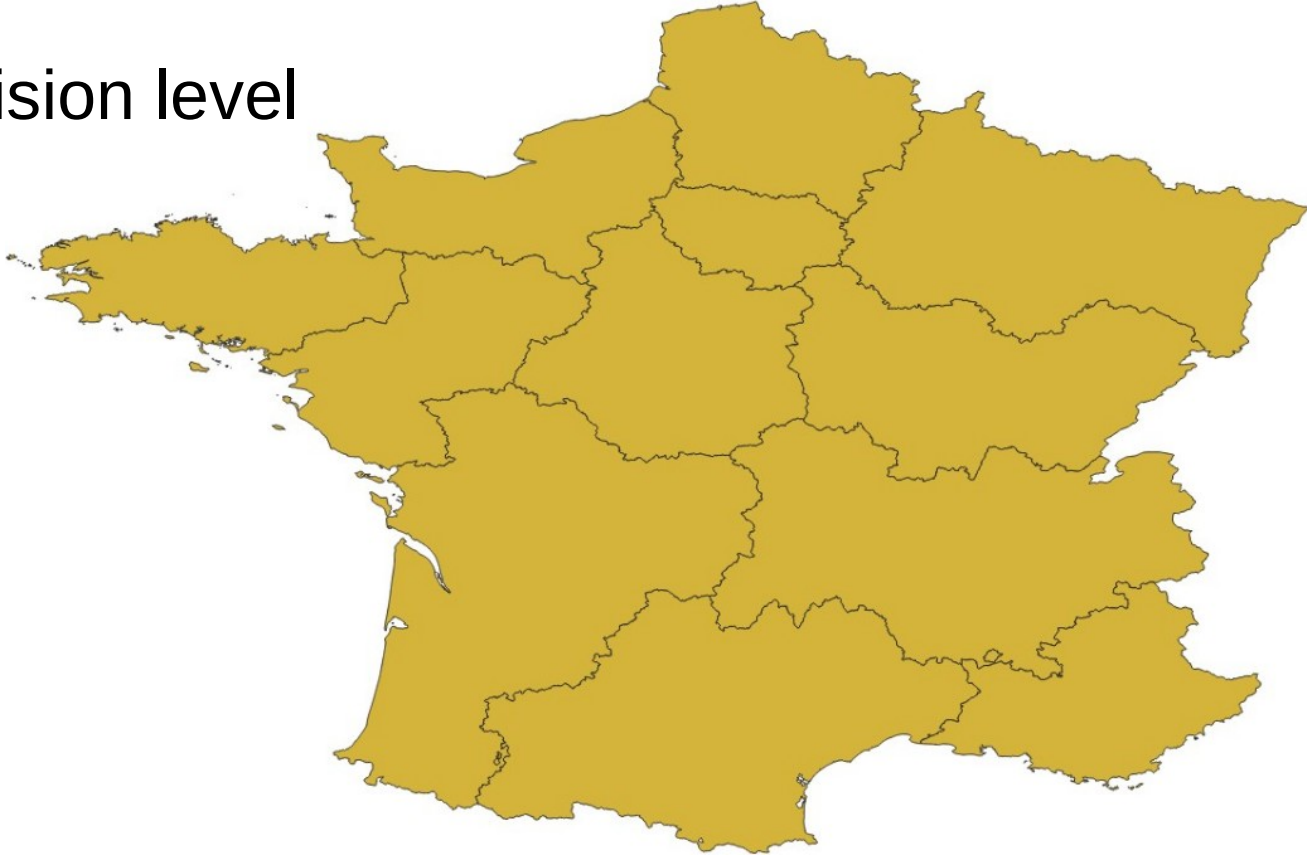
The ISO 3166 System

- ISO-3166-1: country level
- ISO-3166-1-alpha-2: FR
- ISO-3166-1-alpha-3: FRA
- ISO-3166-1-numeric: 250

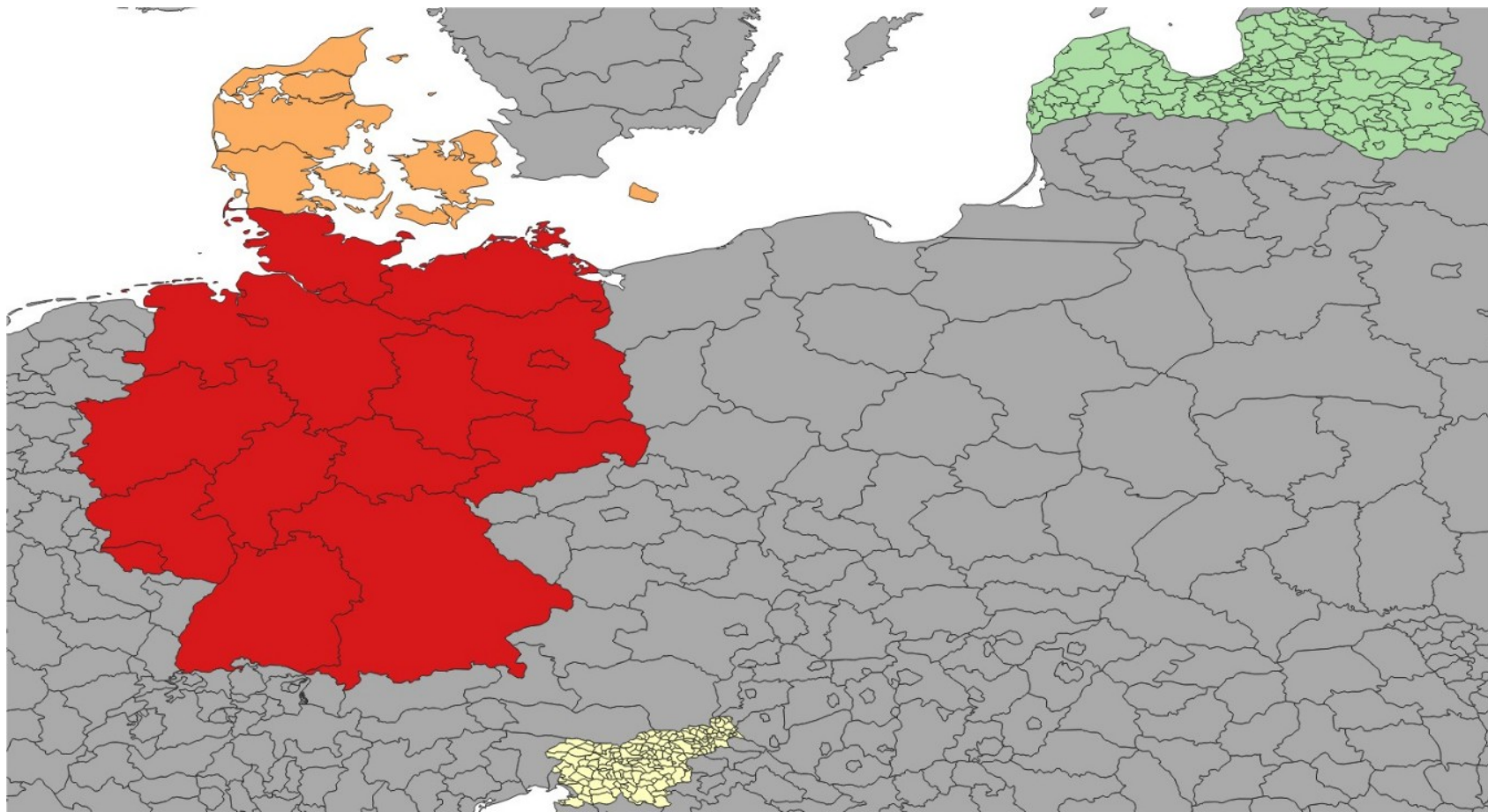


The ISO 3166 System

- ISO-3166-2: subdivision level
- ISO-3166-2-codes:
 - FR-ARA
 - FR-BFC
 - ...



Not so ISO across countries...



Projections

Projection Matters!



The World is One, but the Map is not (<https://xkcd.com/977/>)

WHAT YOUR FAVORITE
MAP PROJECTION
SAYS ABOUT YOU

MERCATOR



YOU'RE NOT REALLY INTO MAPS.

ROBINSON



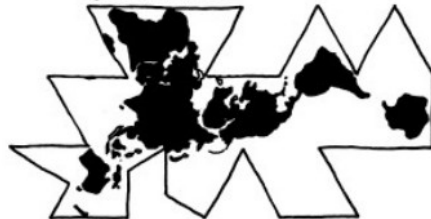
YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.

VAN DER GRINTEN



YOU'RE NOT A COMPLICATED PERSON. YOU LOVE THE MERCATOR PROJECTION; YOU JUST WISH IT WEREN'T SQUARE. THE EARTH'S NOT A SQUARE, IT'S A CIRCLE. YOU LIKE CIRCLES. TODAY IS GONNA BE A GOOD DAY!

DYMAXION



YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TOES. YOU THINK THE SEGWAY GOT A BAD RAP. YOU OWN 3D GOGGLES, WHICH YOU USE TO VIEW ROTATING MODELS OF BETTER 3D GOGGLES. YOU TYPE IN DVORAK.

WINKEL-TRIEPEL



NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIEPEL IN 1998, BUT YOU'VE BEEN A WAT FAN SINCE LONG BEFORE "NAT GEO" SHOWED UP. YOU'RE WORRIED IT'S GETTING PLAYED OUT, AND ARE THINKING OF SWITCHING TO THE KAVRAYSKY. YOU ONCE LEFT A PARTY IN DISGUST WHEN A GUEST SHOWED UP WEARING SHOES WITH TOES. YOUR FAVORITE MUSICAL GENRE IS "POST-".

HOB0-DYER



YOU WANT TO AVOID CULTURAL IMPERIALISM, BUT YOU'VE HEARD BAD THINGS ABOUT GAIL-PETERS. YOU'RE CONFLICT-AVERSE AND BUY ORGANIC. YOU USE A RECENTLY-INVENTED SET OF GENDER-NEUTRAL PRONOUNS AND THINK THAT WHAT THE WORLD NEEDS IS A REVOLUTION IN CONSCIOUSNESS.

GOODE HOMOLoSINE



THEY SAY MAPPING THE EARTH ON A 2D SURFACE IS LIKE FLATTENING AN ORANGE PEEL, WHICH SEEMS EASY ENOUGH TO YOU. YOU LIKE EASY SOLUTIONS. YOU THINK WE WOULDN'T HAVE SO MANY PROBLEMS IF WE'D JUST ELECT *NORMAL* PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY FOOD FROM THE RESTAURANTS NEAR THE GATES AND SERVE *THAT* ON BOARD. YOU CHANGE YOUR CAR'S OIL, BUT SECRETLY WONDER IF YOU REALLY *NEED* TO.

PLATE CARRÉE
(EQUIRECTANGULAR)



YOU THINK THIS ONE IS FINE. YOU LIKE HOW X AND Y MAP TO LATITUDE AND LONGITUDE. THE OTHER PROJECTIONS OVERCOMPLICATE THINGS. YOU WANT ME TO STOP ASKING ABOUT MAPS SO YOU CAN ENJOY DINNER.

Let's Make a Quick Dive

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
pr01_e01.l...	5 hours ago
pr01_e02.l...	4 minutes ago
pr02_e98.l...	a year ago
pr02_e99.l...	a year ago

pr01_e02.ipynb

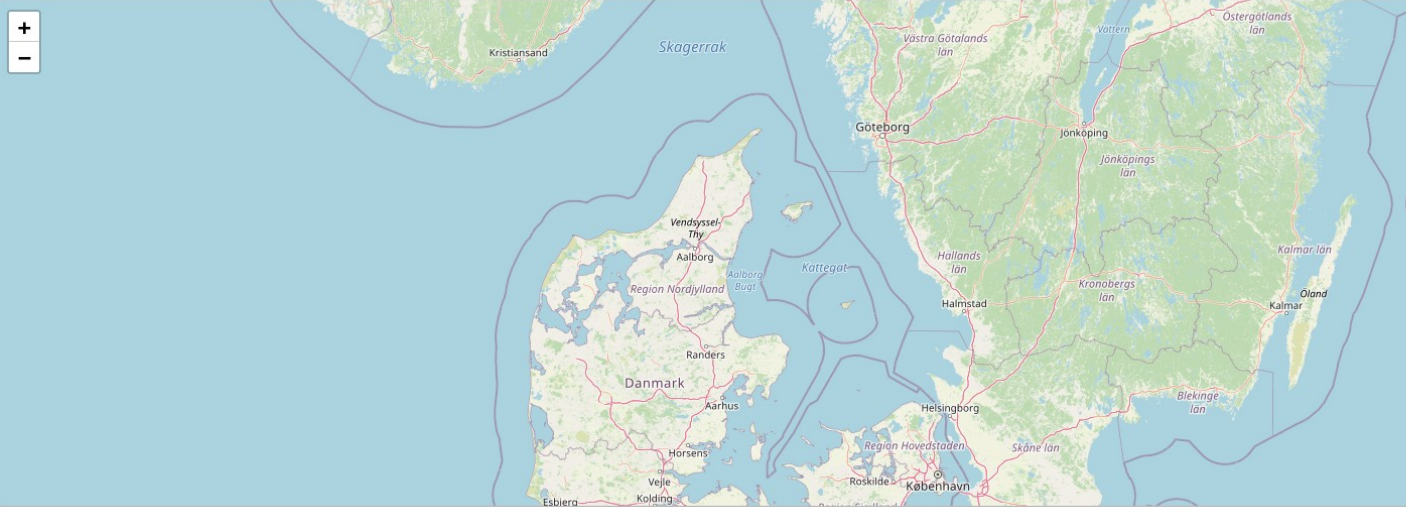
```
[1]: # We start by loading the libraries we're going to use
import folium, json
import pandas as pd

[25]: # Here we store the paths to the data files
# Remember to change this in case your folder structure is different
geo_json_path = "../data/raw/shapefiles/dk.geojson"
corona_df = pd.read_csv("../data/raw/corona/dk_corona.csv", sep = "\t")

# Loading the country metadata directly, since it is a JSON file
with open("../data/raw/metadata/dk_metadata.json", 'r') as f:
    country_metadata = json.load(f)

[49]: # Time to look at some maps
# First, let's initialize folium by finding the geographic center of the country and an appropriate zoom level
m = folium.Map(location = [56, 10], zoom_start = 7, crs = "EPSG3857")
m

[49]:
```



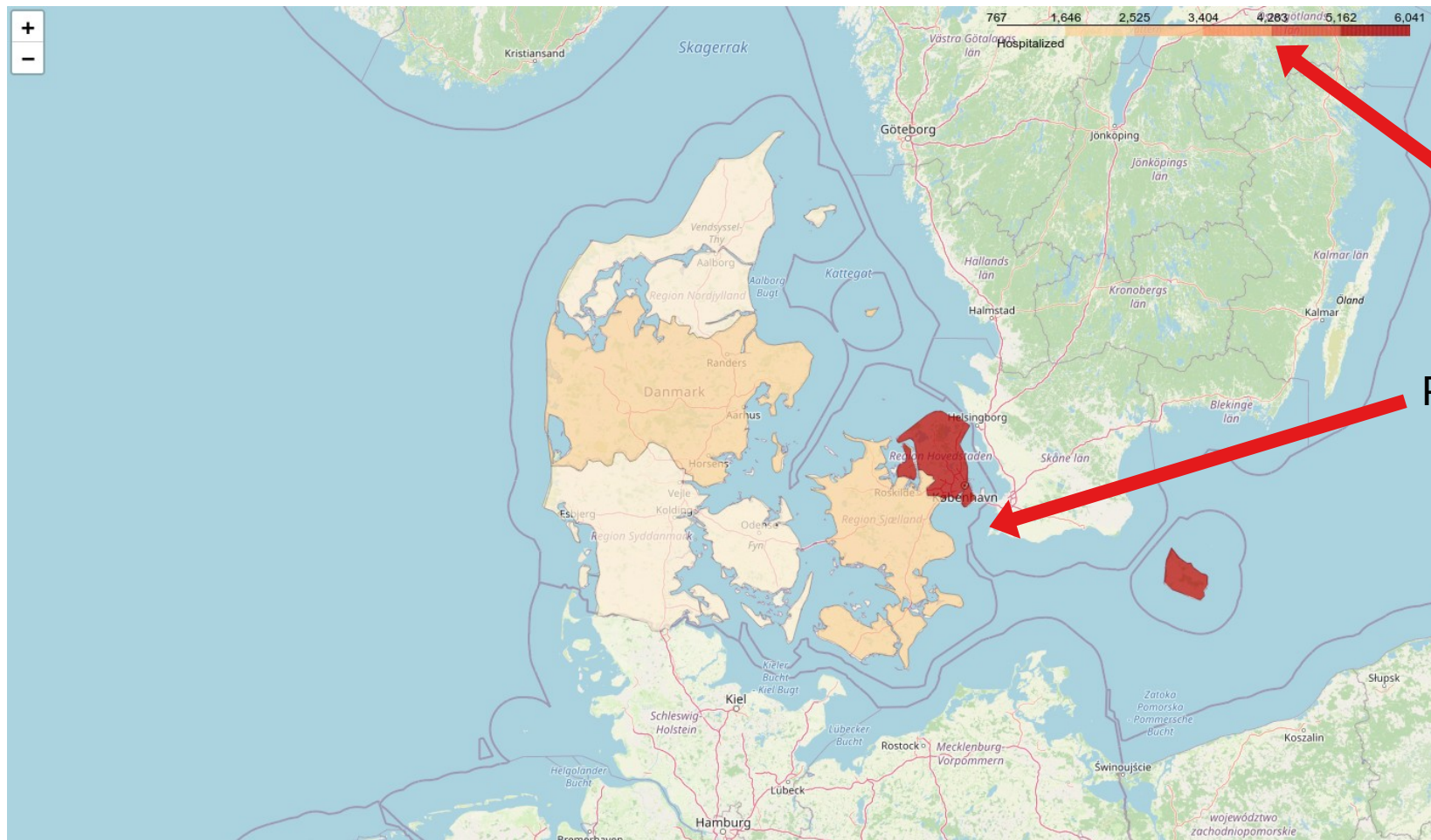
Simple 0 2 Python 3 (ipykernel) | Idle Mode: Command Ln 3, Col 17 pr01_e02.ipynb

Choropleths

Definition

A type of thematic map in which a set of pre-defined areas is colored or patterned in proportion to a statistical variable that represents an aggregate summary of a geographic characteristic within each area.

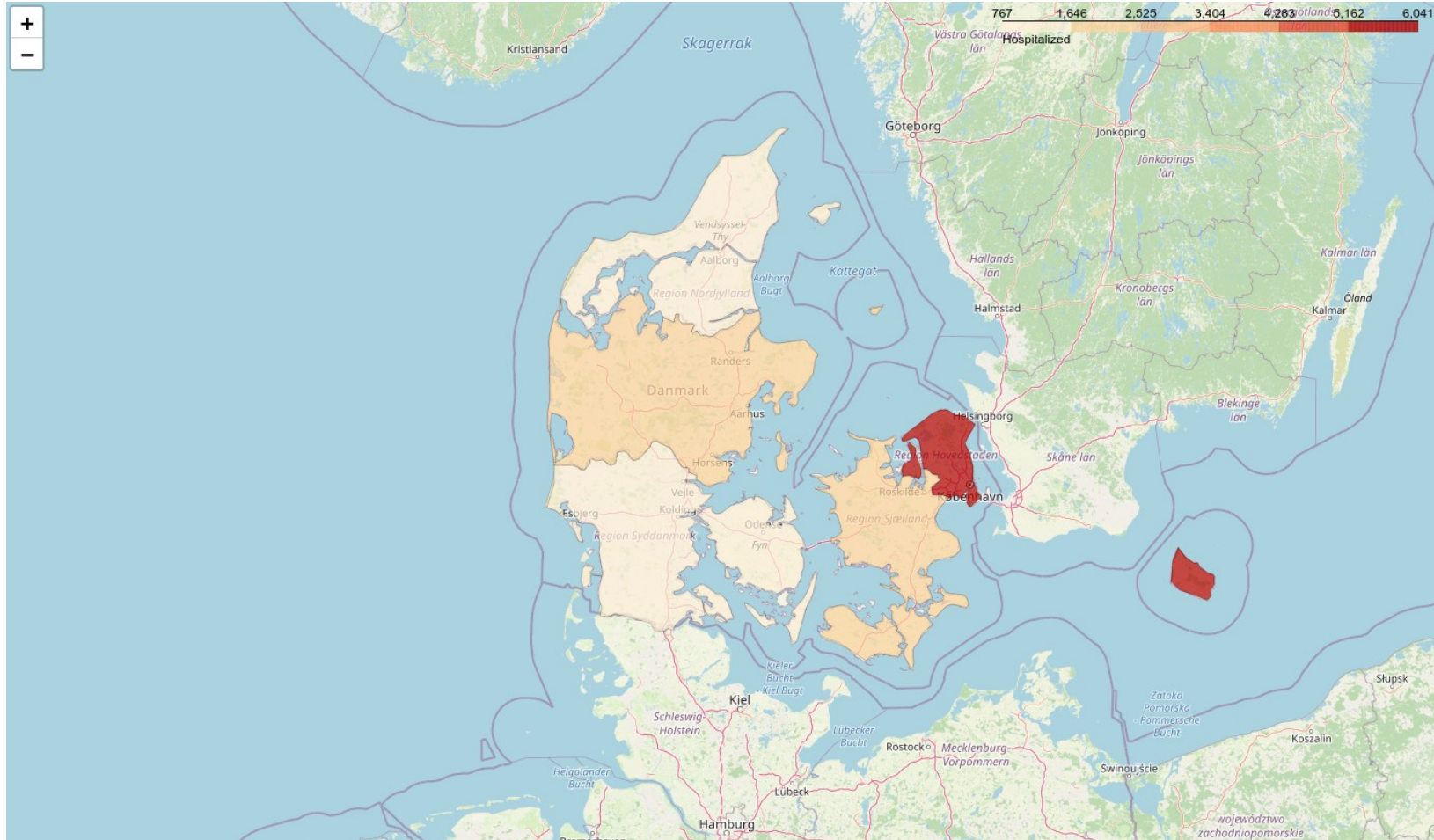
Definition



Statistic

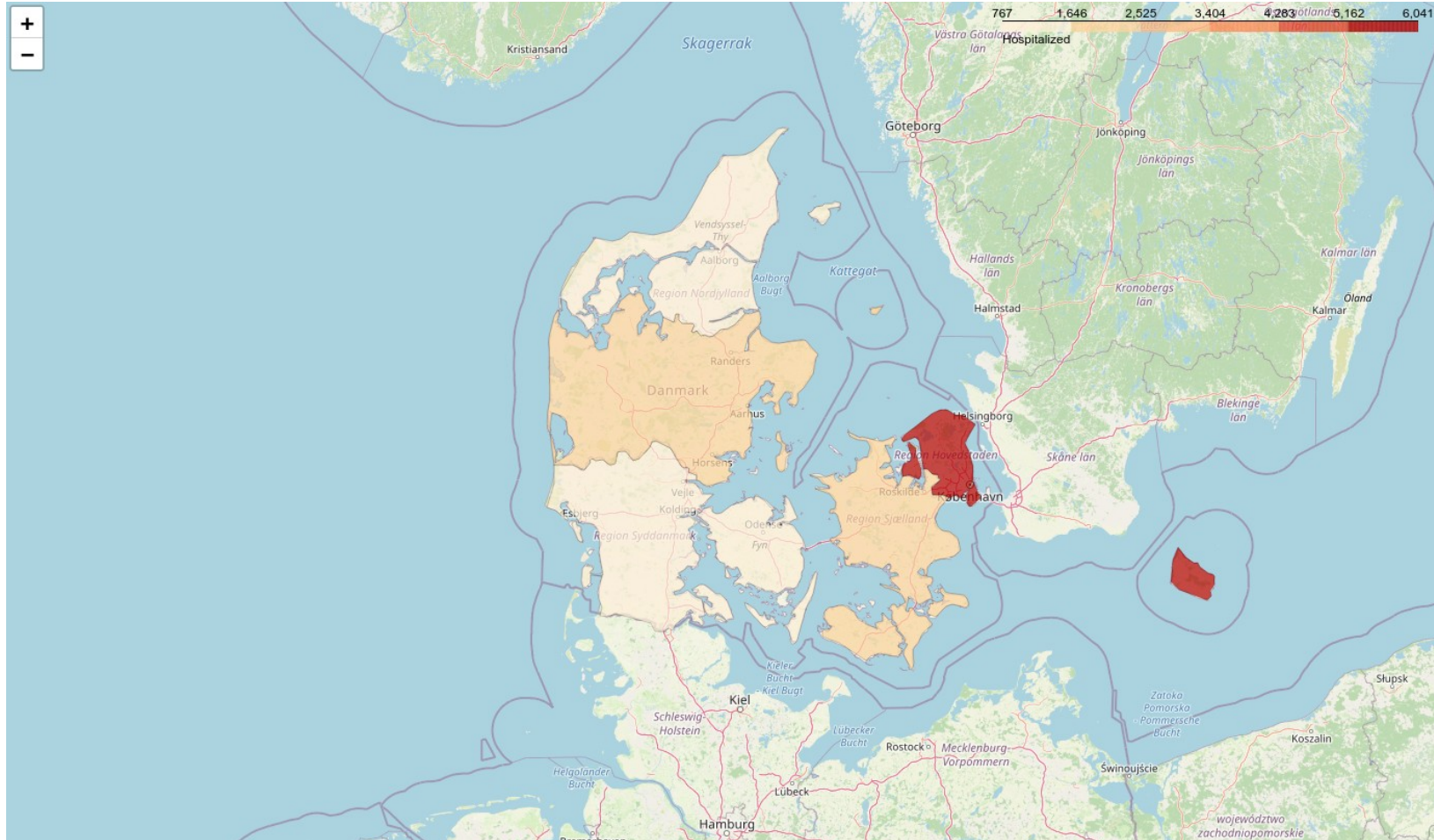
Pre-defined areas

Why choropleths are great



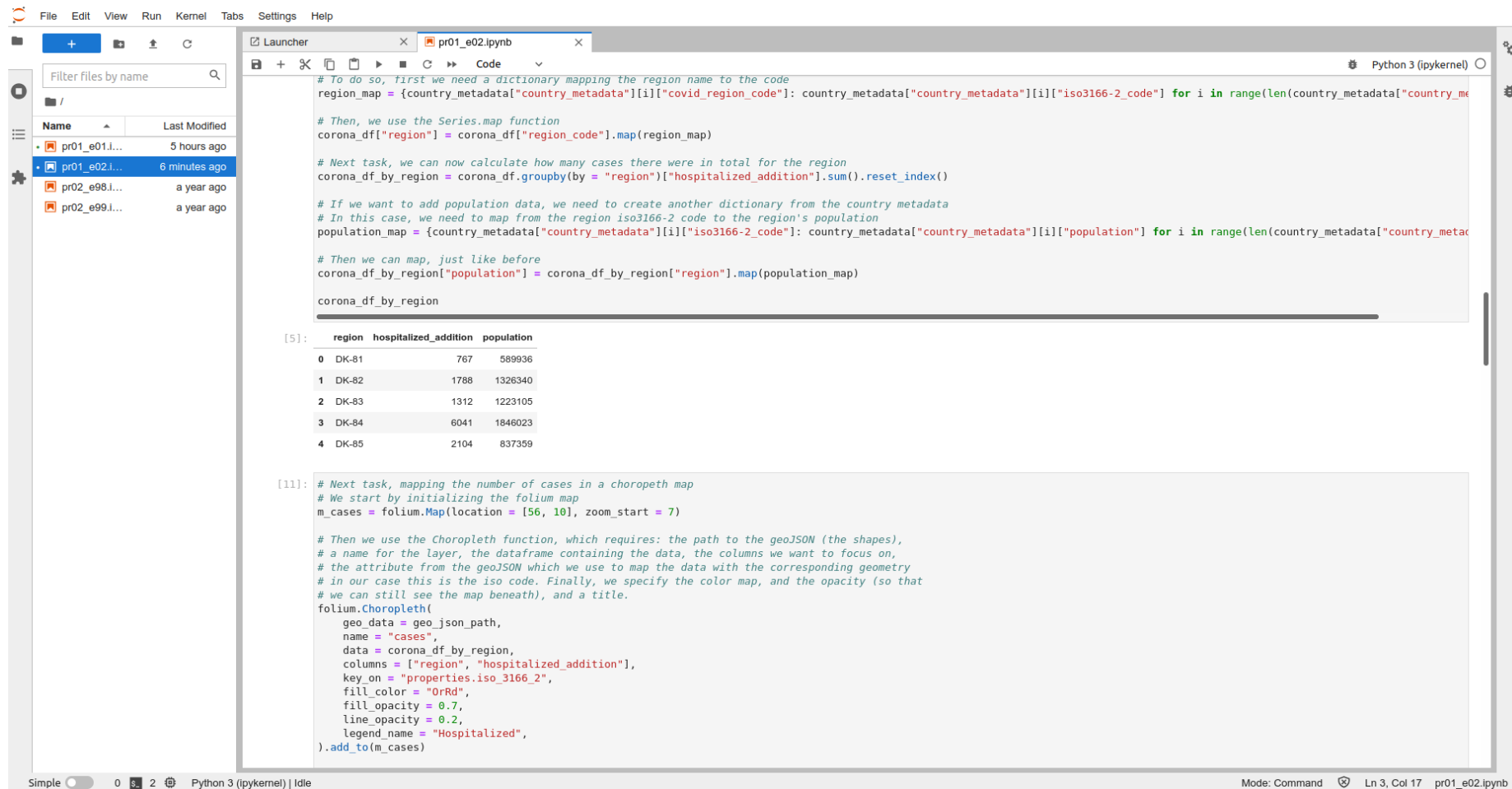
- Familiarity
- Simplicity
- Easy to read

Why choropleths are dangerous



- Size = importance?
- Forced location
- Easy to misread

Let's Make Some Maps



The screenshot shows a Jupyter Notebook interface with a file explorer on the left and a code editor on the right. The file explorer shows a list of files with names like 'pr01_e01...', 'pr01_e02...', 'pr02_e98...', and 'pr02_e99...'. The code editor shows a Python script for mapping COVID-19 data. The script includes comments and code for creating a region map, calculating hospitalized cases by region, and initializing a folium map. A table of results for hospitalized cases is displayed below the code.

```
# To do so, first we need a dictionary mapping the region name to the code
region_map = {country_metadata["country_metadata"][i]["covid_region_code"]: country_metadata["country_metadata"][i]["iso3166-2_code"] for i in range(len(country_metadata["country_metadata"]))}

# Then, we use the Series.map function
corona_df["region"] = corona_df["region_code"].map(region_map)

# Next task, we can now calculate how many cases there were in total for the region
corona_df_by_region = corona_df.groupby(by = "region")["hospitalized_addition"].sum().reset_index()

# If we want to add population data, we need to create another dictionary from the country metadata
# In this case, we need to map from the region iso3166-2 code to the region's population
population_map = {country_metadata["country_metadata"][i]["iso3166-2_code"]: country_metadata["country_metadata"][i]["population"] for i in range(len(country_metadata["country_metadata"]))}

# Then we can map, just like before
corona_df_by_region["population"] = corona_df_by_region["region"].map(population_map)

corona_df_by_region
```

	region	hospitalized_addition	population
0	DK-81	767	589936
1	DK-82	1788	1326340
2	DK-83	1312	1223105
3	DK-84	6041	1846023
4	DK-85	2104	837359

```
[5]:
```

```
# Next task, mapping the number of cases in a choropleth map
# We start by initializing the folium map
m_cases = folium.Map(location = [56, 10], zoom_start = 7)

# Then we use the Choropleth function, which requires: the path to the geoJSON (the shapes),
# a name for the layer, the dataframe containing the data, the columns we want to focus on,
# the attribute from the geoJSON which we use to map the data with the corresponding geometry
# in our case this is the iso code. Finally, we specify the color map, and the opacity (so that
# we can still see the map beneath), and a title.
folium.Choropleth(
    geo_data = geo_json_path,
    name = "cases",
    data = corona_df_by_region,
    columns = ["region", "hospitalized_addition"],
    key_on = "properties.iso_3166_2",
    fill_color = "OrRd",
    fill_opacity = 0.7,
    line_opacity = 0.2,
    legend_name = "Hospitalized",
).add_to(m_cases)
```

```
[11]:
```

Simple 0 2 Python 3 (ipykernel) | Idle Mode: Command Ln 3, Col 17 pr01_e02.ipynb

Q&A