IOD_mini_project1

Ng Jing Kang

IOD_mini_project1

API: https://covid-api.com/api/ (Public api that houses COVID related data)

Endpoints	Description of data
https://covid-api.com/api/regions	List of region names (subset of provinces data)
https://covid-api.com/api/provinces/{iso}	List of provinces by ISO code (subset of reports data under column “regions”)
https://covid-api.com/api/reports	List of reports
https://covid-api.com/api/reports/total	Total data by date (sum of report columns)

IOD_mini_project1

Profiling report: merged_profiling_report.html

Columns	Description
confirmed, deaths, fatality_rate, region	Usable columns (Columns of interest)
date, recovered, confirmed_diff, deaths_diff, recovered_diff, last_update, active, active_diff	Only has one value/Multiple zeros/Derivative column

IOD_mini_project1

Additional datasets:

File	Description of data
Statistic_id1183370_population-density-in-china-2020-by-region.xlsx (Sheet_name = Data)	Average population density in China in 2020, by province or region (https://www.statista.com/statistics/1183370/china-population-density-by-region-province/)
Statistic_id1183370_population-density-in-china-2020-by-region.xlsx (Sheet_name = Data_gdp)	List of Chinese administrative divisions by GDP (https://en.wikipedia.org/wiki/List_of_Chinese_administrative_divisions_by_GDP#cite_note-data2022-1)
province.shp	Shapefile for china provinces

IOD_mini_project1

Assumptions:

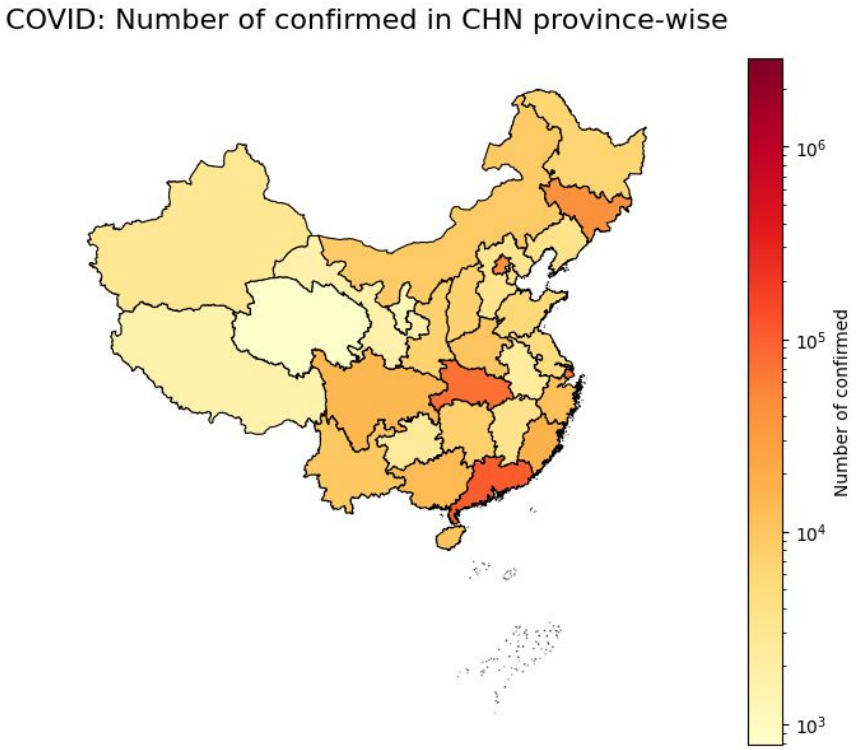
- All datasets values will be assumed to be up to date with actual values having very little variance (Statistics like gdp and population density of a certain country could be non-publicly available)

Initial investigation: To find out if there are any patterns relating geographical location and the modes of transmission of the virus. (e.g. Densely populated areas like major cities/maritime hub)

A short google search summarizes the modes of transmission: close contact/proximity (within 1 m), direct/indirect contact (surfaces/objects within the immediate environment of an infected person) (https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations)

IOD_mini_project1

- Investigation of ISO = “CHN”, excluding Taiwan (ISO = “TWN”)

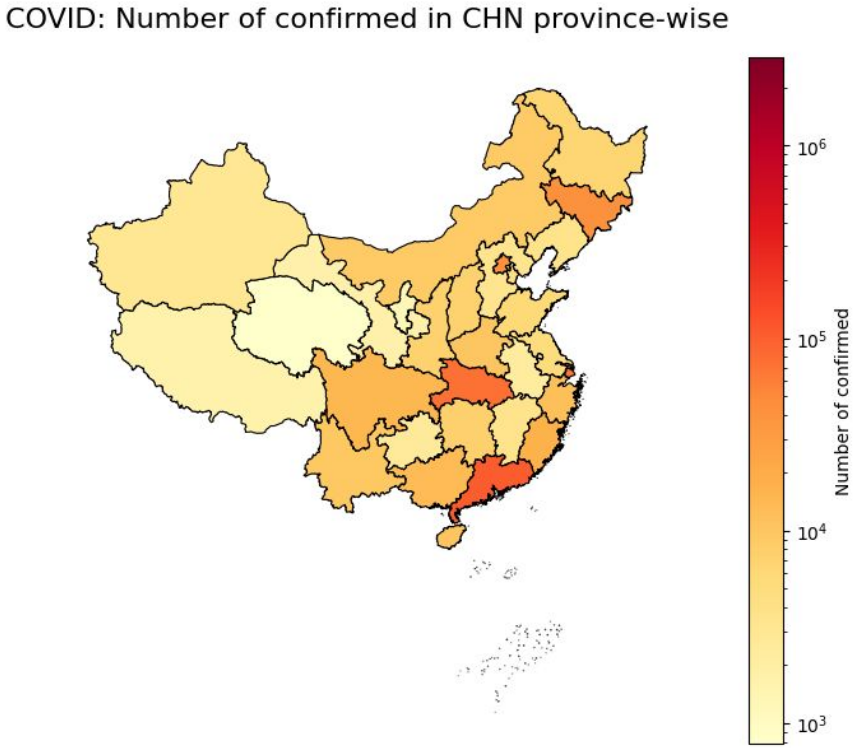


IOD_mini_project1

- Summary statistics:

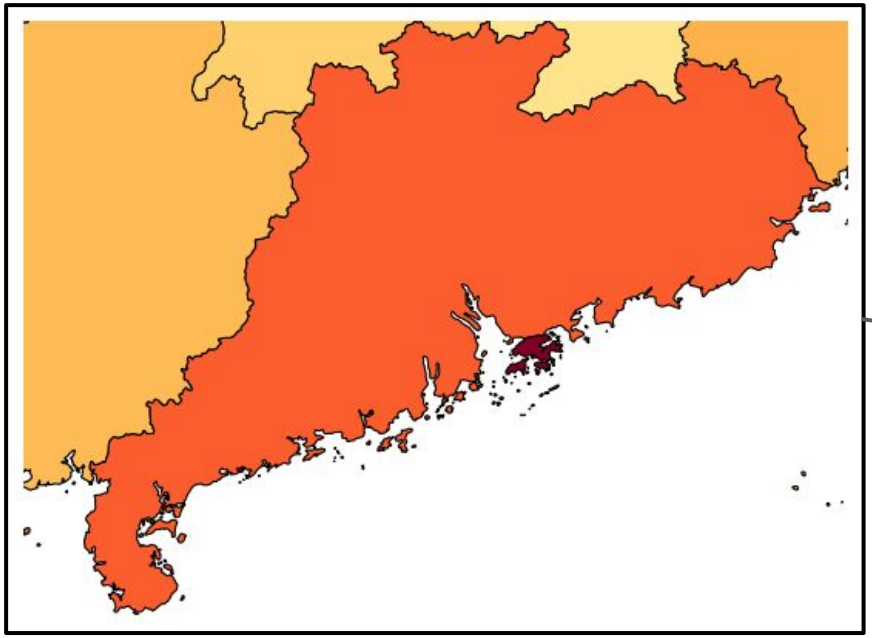
	Confirmed	Deaths	Fatality rate	Population density
Mean	102476	571.54	0.004006	1383.31
Std	498463	2445.25	0.012132	4240.07
Min	782	0	0	23472.00
Max	2876106	13467	0.062600	3.03

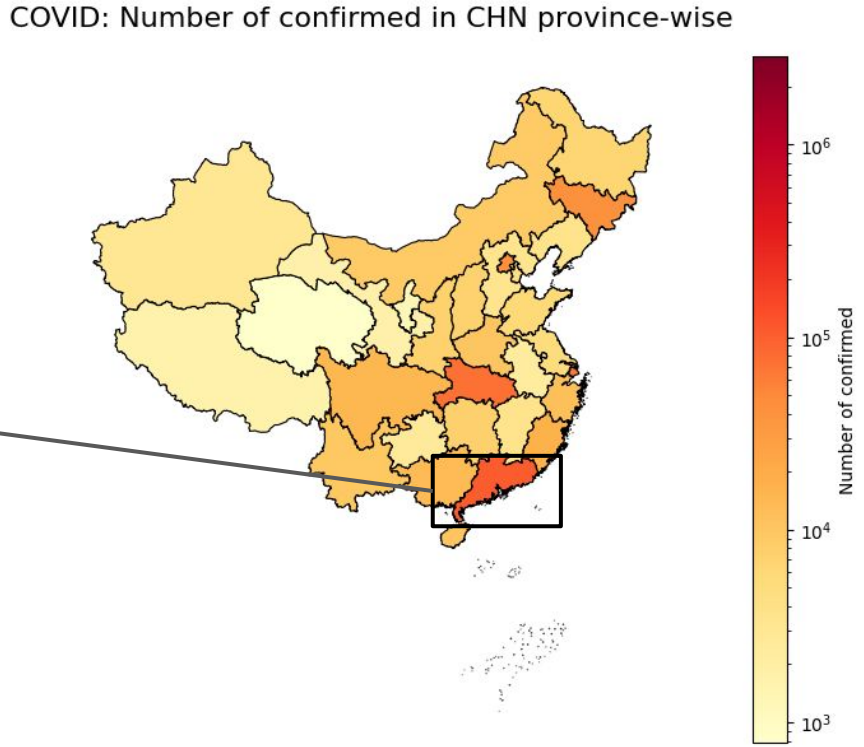
Note: using log scaling due to order of magnitude difference of min and max



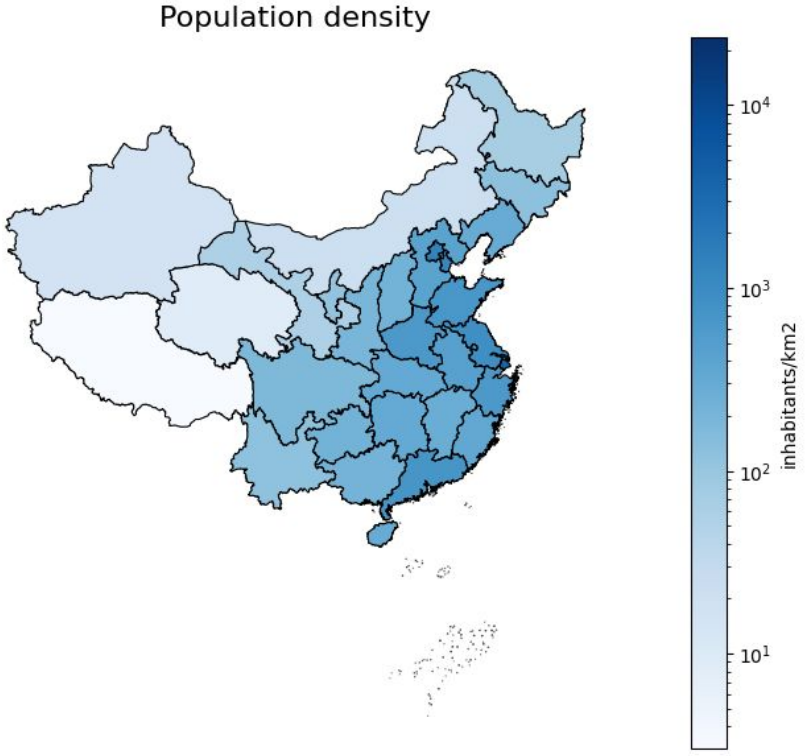
IOD_mini_project1

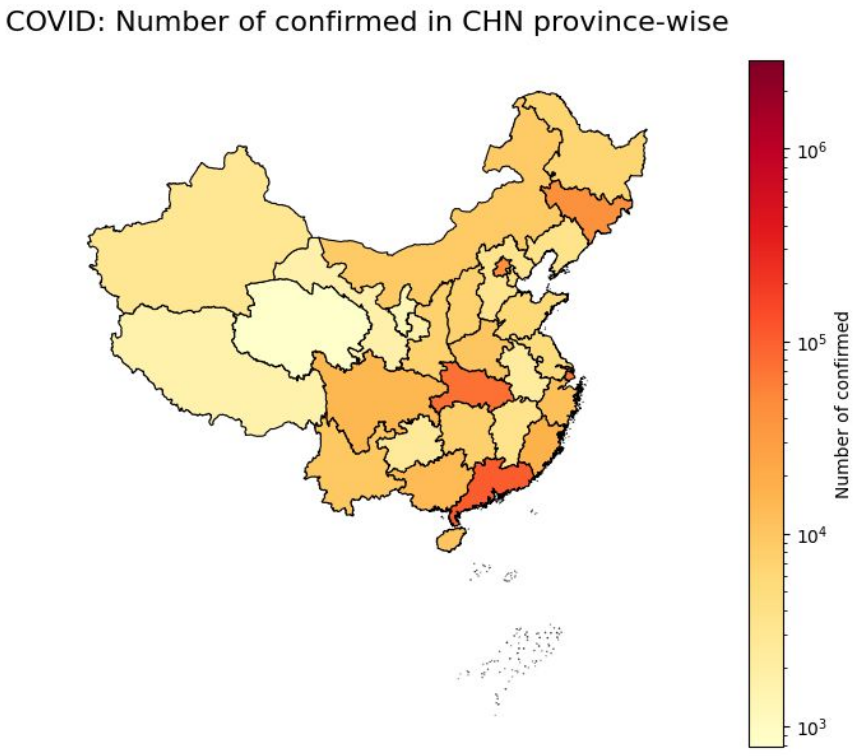
Rank	Province	Confirmed
1	HongKong	2876106
2	Guangdong	103248





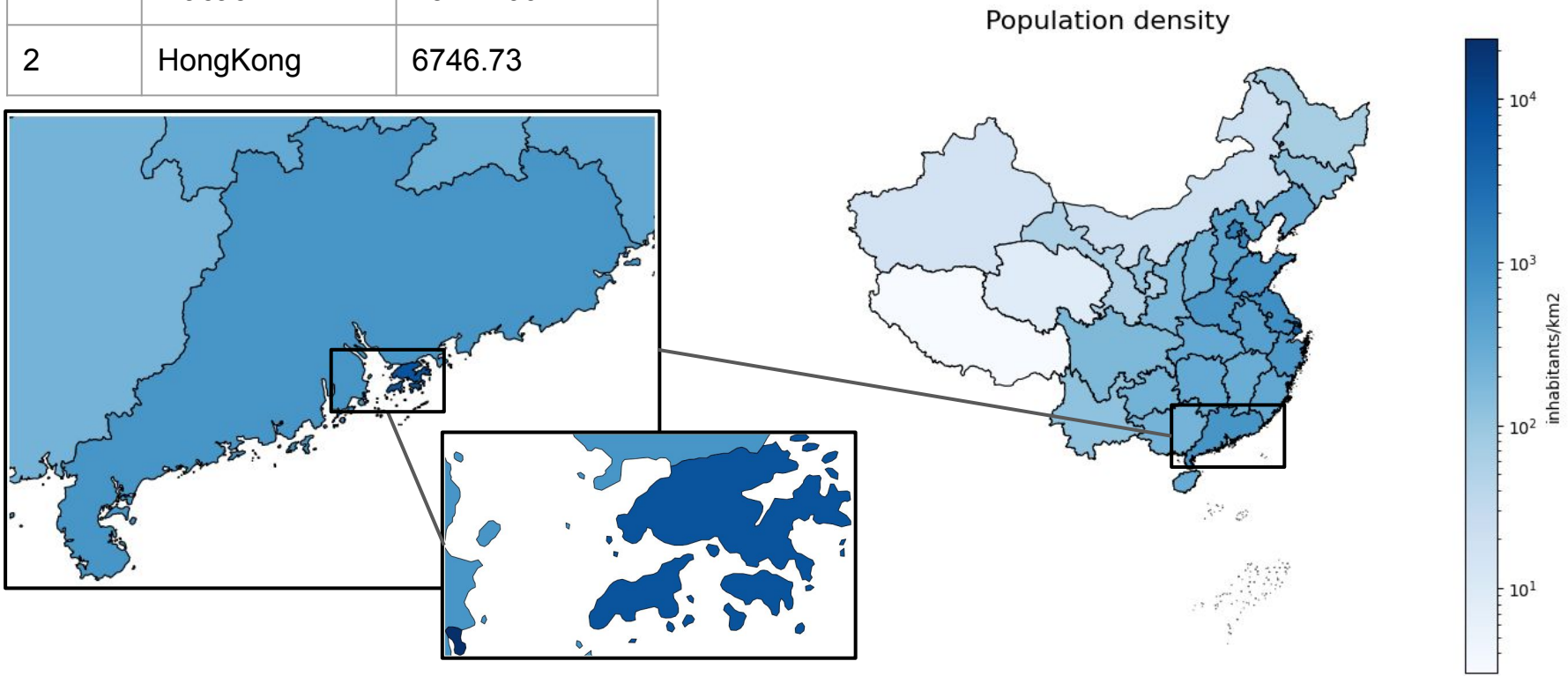
IOD_mini_project1



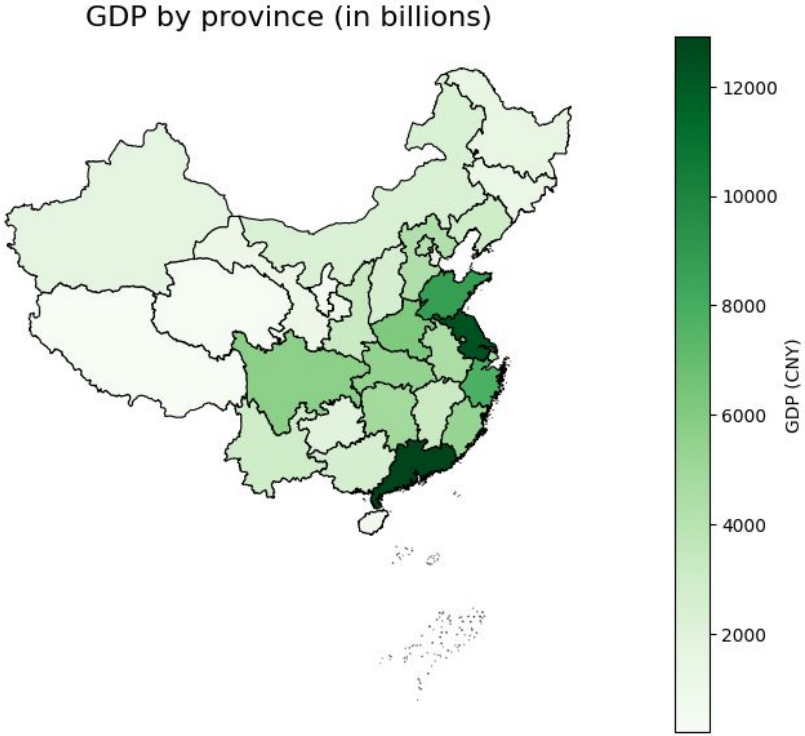


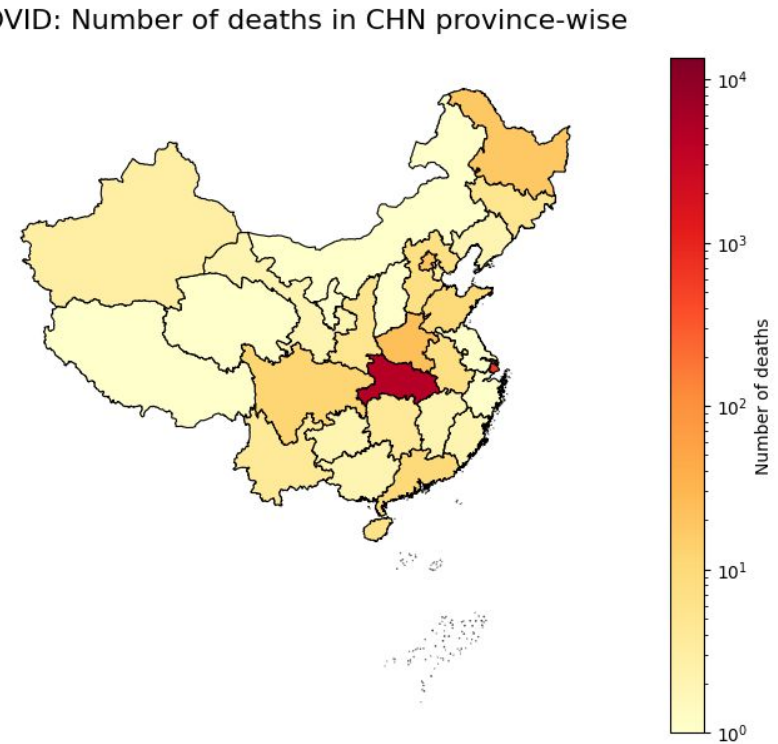
IOD_mini_project1

Rank	Province	Inhabitants/km2
1	Macao	23472.00
2	HongKong	6746.73



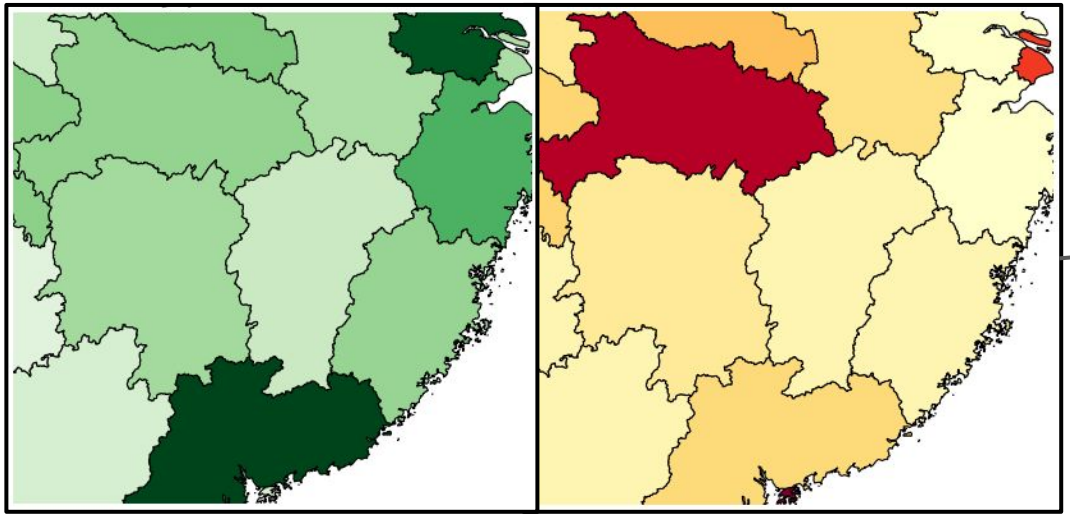
IOD_mini_project1

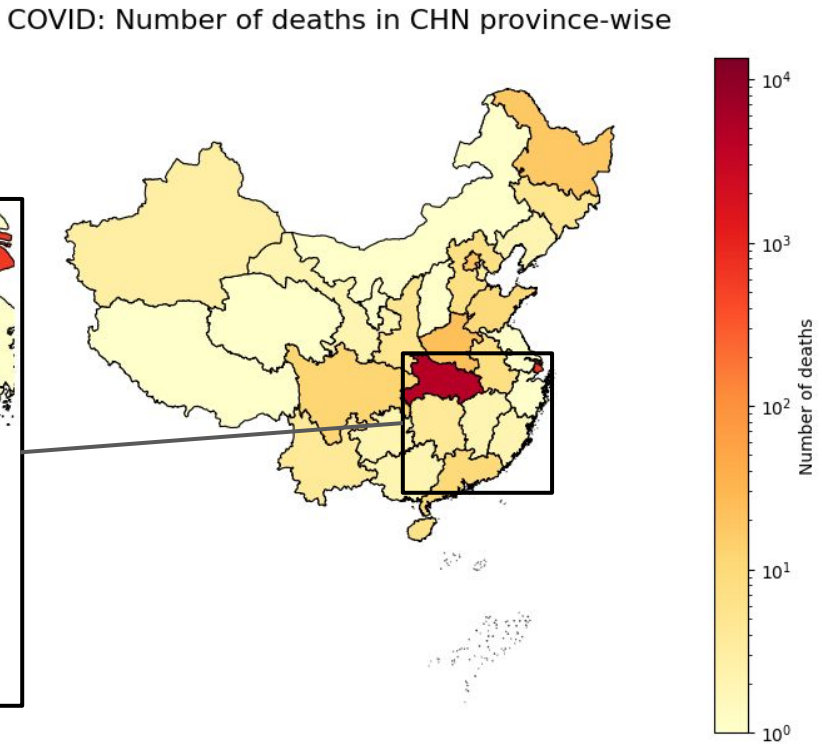




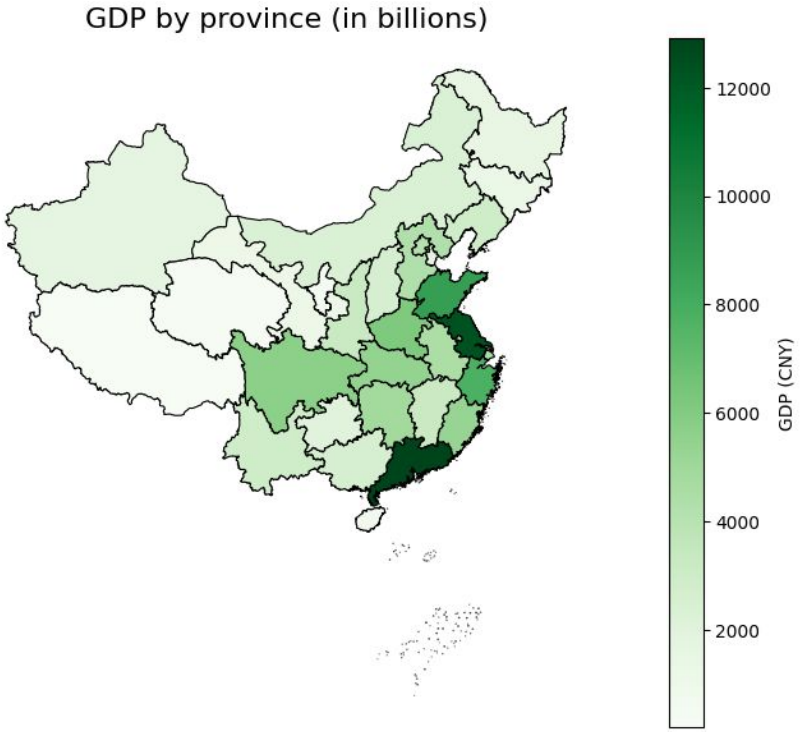
IOD_mini_project1

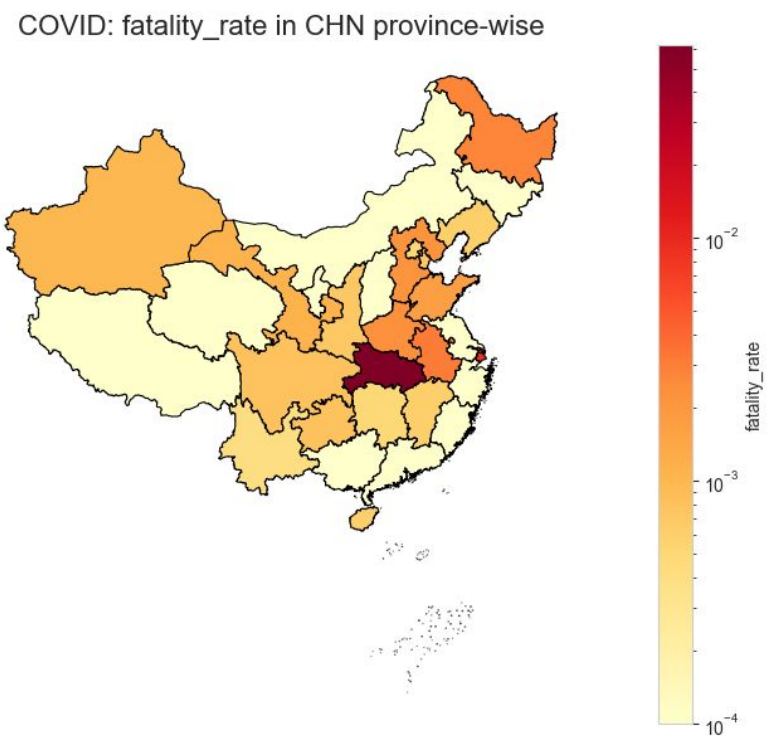
Rank	Province	Deaths	GDP (CNY in billions)
(1,18)	HongKong	13467	2655.65
(2,7)	Hubei	4515	5373.49
(3,11)	Shanghai	595	4465.28





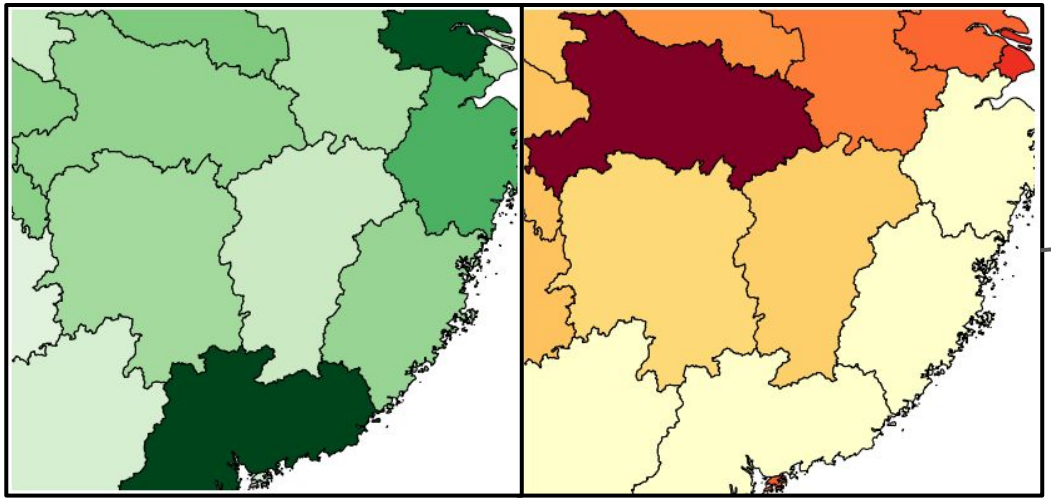
IOD_mini_project1

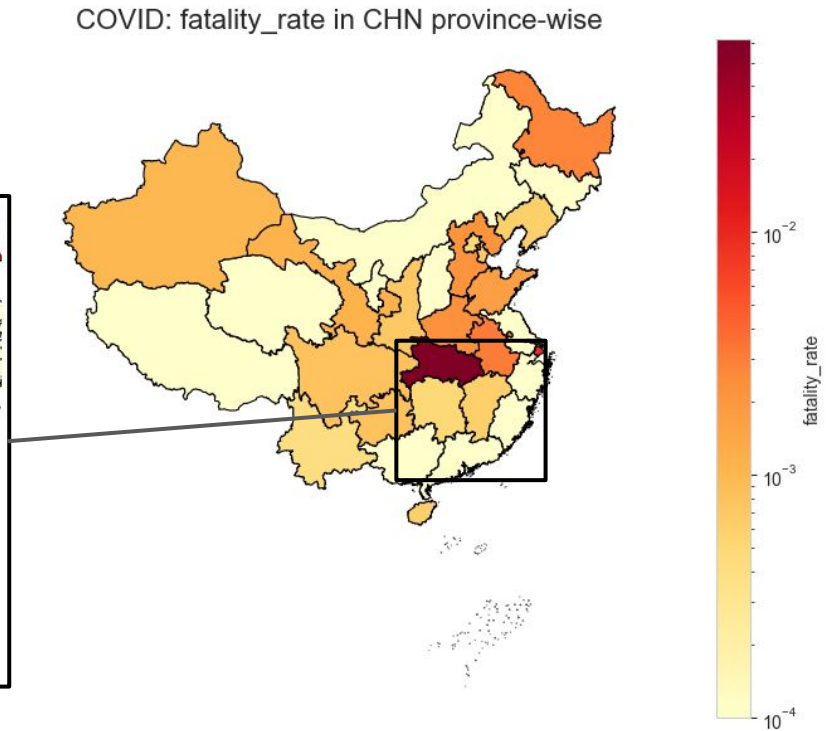




IOD_mini_project1

Rank	Province	Fatality rate	GDP (CNY in billions)
(1,7)	Hubei	6.26%	5373.49
(2,31)	Macao	3.44%	216
(3,11)	Shanghai	0.89%	4465.28





IOD_mini_project1

Findings and enhancements:

1. Population density and confirmed cases
    a. Densely populated regions seems to have increased number of confirmed cases (although a better comparison could be “infection rate” because each region is upper bounded by population size; e.g. macao)
    b. There are also additional factors like the major facilities of a region (Tourism, business/trade, residential) that can affect confirmed cases/infection rate.
2. GDP and fatality rate
    a. GDP is assumed to be the measure for the development level of a region. Some correlation can be drawn where fatality rate is higher in underdeveloped regions. (Rural vs Urban)
    b. Though the data shows that some major cities have higher fatality rate, it can be due to additional factors like lack of medical facilities/manpower.