

Visualizing COVID-19 in Italy

(as of April 2020)

Part I

Elizabeth Lentine, Rohan Bansal, Drew Williams, Divya Ramakrishnan

Description of Dataset

For this project, we decided to visualize the data from a dataset containing information regarding the spread of Coronavirus in Italy. We wanted to explore a topic for this project that would provide interesting and important information on subject matter that concerns us all. We felt that there was no better topic to explore than the COVID-19 pandemic currently affecting the world. As a part of this milestone, we brainstormed the areas in the world that the virus had a large reach in, but areas that also implemented rigid social distancing measures. Specifically, Italy, China, Singapore, and Korea came to mind, so we analyzed sources of information on these locations^[A.1]. The Italy dataset we found, sourced from the Italian government, contains 19 features^[A.2] and is updated several times a day. This data includes information on the country, each province, and each region. Since it provided so many useful measures and we can count on current information being added, we felt this was a good dataset to start answering some of our questions and chose to focus on Italy for our visualizations.

Through this project, we will explore how social distancing and testing measures have had an effect on Italy, and how the spread has been contained from the initial outbreak to now. We also would like to know what the growth model of the virus looks like; we are assuming that the virus experienced exponential growth in Italy until the strict social distancing measures were put into place. Furthermore, we hope to see how effective the social distancing measures were on the rate of infection. We would also like to see if there were any surprising spikes shown from the dataset in regards to the spread of the virus. If there were random spikes in the number of infections, despite strict distancing measures in place, it would be interesting to further examine that.

Visualizations of Data + Potential Insights

[Figure 1]



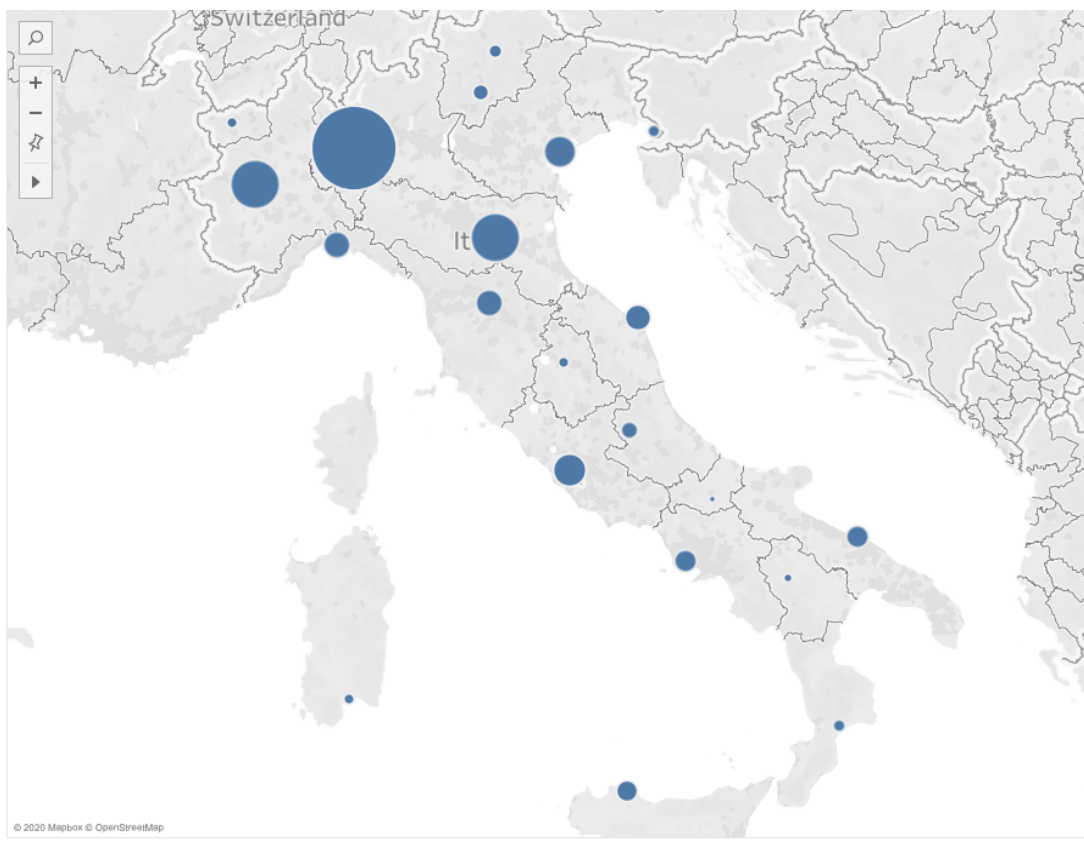
Italy: total confirmed COVID-19 cases and proportion of patients in the ICU by region.

Updated April 18th, 2020

Perhaps the biggest strain on many healthcare systems due to COVID-19 has been the lack of available equipment in intensive care units, specifically ventilators. Physicians in areas hit hard by the virus have been making incredibly difficult triage decisions as to who to treat with the devices they do have^[1]. Understanding the distribution of high-risk patients could be useful in allocating medical equipment and specialized healthcare workers. In this visualization, the size of the mark indicates the total number of confirmed cases in the region. A darker red color indicates a higher proportion of COVID-19 patients in intensive care. Clearly the regions with many more confirmed cases tend to have a lower

rate of patients ending up in the ICU. This suggests that the number of Italian ICU hospitalizations cannot simply be determined by the number of total confirmed cases. In other words, this map would refute the idea that any X% of Italian COVID-19 patients will require intensive care, regardless of region.

[Figure 2]



Italy total confirmed Covid-19 deaths by region.

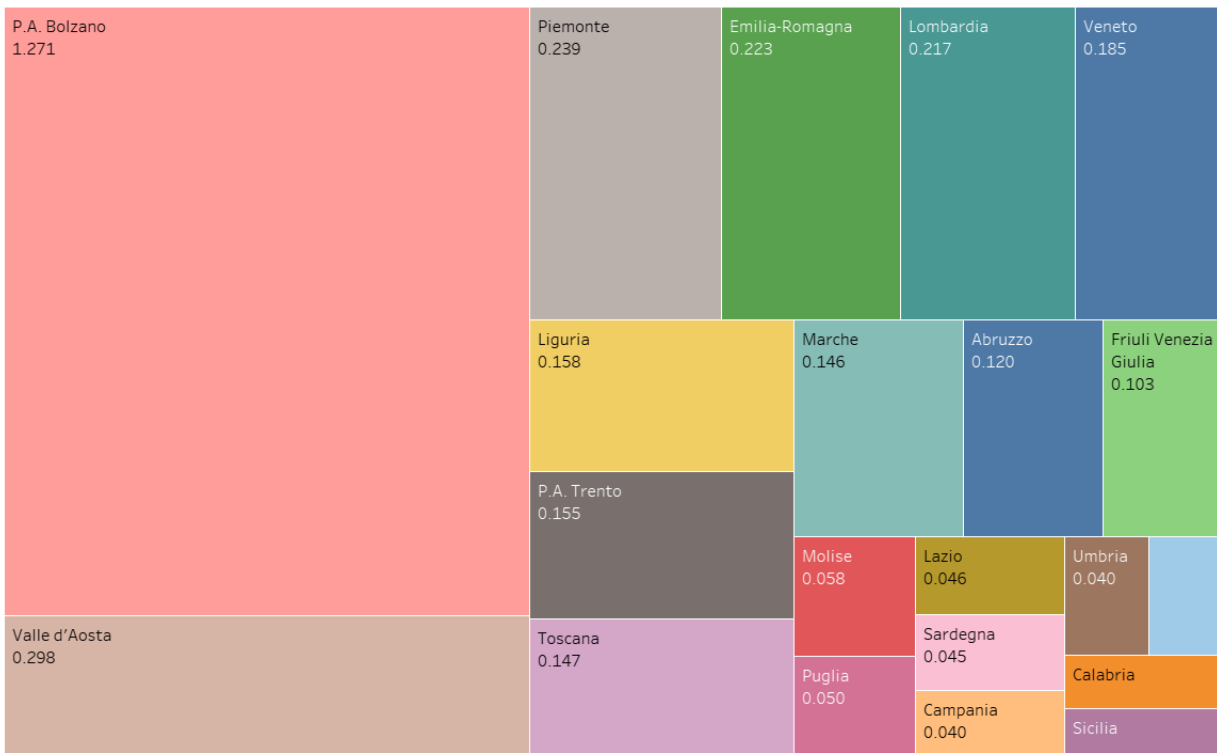
Updated April 18th, 2020

A major concern with Covid-19 is the death rate amongst people over the age of 60, and using this map we are able to see the location of deaths, scaled to the number of deaths in each location. Using this visualization and gathering more data on the population, we will be able to compare the average age of people in the region and how many people in the region who got Covid-19 with the number of people who died to see how strong the correlation of age and death are. Along with this, it is likely that early cases of Covid-19

weren't treated as well as newer cases. This is accurate, as the outbreak in Italy started in Milan. Looking at the visualization, we can see that the most deaths are in the Milan region. Deaths also appear to be clustered in regions with high population density, as when looking at the visualization the clusters of deaths appear to be in urban areas rather than the countryside, where someone is less likely to get Covid-19.

[Figure 3]

Rate of In-Home Isolation



Current percentage of population currently isolated at home by region.

Updated April 17th, 2020

In order to determine how many Italians are currently self-isolating in their residences, we found recent population data on each region^[2] and joined it with the COVID-19 data. The above visualization shows what percentage of the population in each region is currently in isolation due to COVID-19 symptoms, but not hospitalized, in intensive care, or necessarily a confirmed positive case. This is important because tests are still being produced and have not reached full capacity to test every individual with symptoms, so the positive number of cases may not tell the whole story.

Appendix

[A.1] Datasets we explored

Metadata for 52k articles on COVID-19

<https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>

Singapore

<https://towardsdatascience.com/covid-19-the-second-wave-in-singapore-91e886b3d444>

Italy

<https://github.com/pcm-dpc/COVID-19>

[A.2] Italy dataset: features

1. Date
2. Country
3. Region code
4. Region name
5. Latitude
6. Longitude
7. Hospitalized with symptoms
8. Intensive care
9. Total hospitalized
10. In-home isolation
11. Current confirmed cases (isolation + hospital)
12. Change in current confirmed cases from previous day
13. Change in total confirmed cases
14. Recovered
15. Deceased
16. Total positive cases
17. Tests performed
18. Notes in Italian
19. Notes in English

References

- [1] Ferraresi M. A coronavirus cautionary tale from Italy: don't do what we did. Boston Globe. March 13, 2020. [URL](#)

[2] Istat. (November 5, 2019). Resident population of Italy in 2019, by region [Graph]. In Statista. Retrieved April 19, 2020. [URL](#)