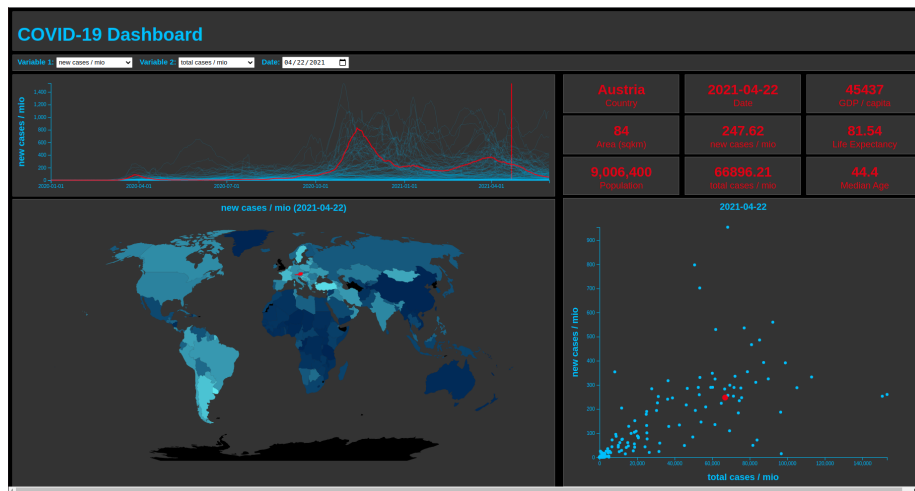


Vis A5

Michael Raffelsberger

June 2021



1 Motivation

The dashboard is supposed to present different aspects of the Covid pandemic for countries all around the world from the very beginning of the pandemic until today¹. It shall support different users with respect to different tasks.

In particular, a public health expert will use it to find out how well vaccination works in different countries. Based on that he or she will try to spot some trends about what impact the vaccination has with respect to infection and particularly death rates. He or she will also have to be able to find out how his or her own country compares to other countries and what hope other countries with more progress provide when performing as expert in news shows, press conferences, etc.

On the other hand, it will support a government official who also wants to compare his or her own country to other countries. In particular, it is generally about finding out what factors might have had a very bad influence (find patterns) to approach those issues and also whether comparable countries (similar gdp, etc.) were similarly affected by the pandemic and why (not).

¹the time I downloaded the dataset

There are of course many other tasks of these two and also other users the dashboard may be useful for. I decided to use a mix of my two A4 prototypes. I am basically using the first design with a time series linechart, a choropleth map and fact fields (i.e. text), but replace the parallel-coordinate-density plot by a scatterplot because the implementation is easier, users are more used to it and it is also better-suited for detecting correlations which is a very important task for both users.²

2 Prototyping / Design Process

2.1 Prototype

As indicated before, the result is a dashboard containing four views in total: A linechart on the top left, a choropleth map on the bottom left, text fields to have same balance between visualization and hard facts on the top right and a scatterplot on the bottom right.

The linechart shows the “time axis” of the pandemic. You can see a thin line for every country for the currently chosen main variable. This helps the user to figure out how a country has behaved over time and also how it compares to all the other countries. Does it fit the general trend? Note here that all variables in use are in some way relative to population size so that comparisons also make sense. The linechart remains unchanged when the selected date changes since it always covers all dates in the data set. However, a vertical line indicates where the currently selected date is located in the linechart to have more context and make more sense out of the other plots that only convey information for one specific date!

The choropleth map is mainly addressing the spatial aspect of the topic. It shows the main variable for one specific chosen date (date selector!). Unlike the other plots it contains information on the location/distance of countries and is the place to choose and highlight a desired country (no country dropdown!). The coloring always adapts to the current variable-date combination to show which countries have low/high values for the current variable 1 and the chosen date.

The text fields provide information on the currently highlighted country (name, population and area which I derived from the variables population and population density). It also shows the currently selected date and the values of the two selected variables for the selected country and date. Finally, one can also see the median age, life expectancy and gdp/capita of the country which are date independent but provide some impression on how well developed the country is. The scatterplot is the place to look for correlations. Here one can see the values of the two chosen variables for the chosen date. Observing how correlation develops over time is a little cumbersome unfortunately.

²vaccination vs. infections/deaths (public health expert) and risk factors vs. infections/deaths (government official)

2.2 What can be found

One can find trends in 6³ different Covid variables over time with the help of the line chart and also detect lines (i.e. countries) that do not share the overall pattern (outliers). One can dive deeper with respect to (extreme) observations in the linechart by choosing the corresponding date in the date selector and then find the brightest country in the map or the outermost points in the scatterplot to select it. (selecting lines directly is cumbersome since they are very thin)

One always gets a rough overview over a selected country with the help of the fact fields. This might be interesting for the government official to briefly compare whether some country has reasonably similar stats to his or her own country.

The map is the best place to highlight a country (hover or click) which is then highlighted in all(!) elements of the dashboard.

The scatterplot though not so prominent is extremely important for correlations. Since the constant variables (smokers, diabetes, etc.) are also included like normal time series variables it is really possible here to for example study the correlation of a risk factor and total deaths for some late date (new deaths probably not so interesting but rather the whole “bill” a country had to pay). Another interesting correlation analysis (for the public health expert) could be to look at total vaccinations against new deaths. Hopefully, countries with many vaccinations do not have too many new deaths.

2.3 Issues

I think that most ideas that I did not implement were simply hard to implement within a short amount of time. I lost my nerve a bit when trying to create a color legend for the map and ended up thinking it is ok to lack it since the values are in the text fields anyway and the range of values also gets clear by looking at the scatterplot which always displays the same variable-date combination as the map in the y-direction. The date indicator line is also not horizontally shiftable to choose a date as initially planned. The parallel-coordinate-density plot would have been more work but less useful than a simple scatterplot in my opinion.

One thing that would have been great for task where comparing specific countries is required is to allow for highlighting multiple countries which technically would not be a problem but hard without losing the meaningfulness of the fact fields. Maybe an extension to two highlighted countries with different colors (also including both in the fact fields) would be worth a try.

3 Implementation Details

My implementation consists of an `index.html` file, three `CSS` files and 7 `JS` files using 5 Javascript libraries. There is a Javascript file for every plot, one config

³The other 6 variables are constant over time but still included for correlation analyses in the scatterplot. They also give some indication of missing values.

file creating the SVGs and some globally relevant variables, one file for helper functions and especially one file that handles all the interaction between the elements (selectors, mousovers, mouseouts, click events, ...).

On the top of the page there are two dropdowns (variable 1 and variable 2) and a date selector. Whenever they change, all elements will adjust to the new variable/date. However, a highlighted country will remain highlighted which enables the user to do extensive analyses for a country of interest without having to reselect it all the time.

The linechart, map and scatter plot allow for highlighting a country (mouseover and click) which will affect all four views and highlight everything related to the country with a dominant red color. Unfortunately, Firefox and in rare cases also Chrome were not 100 per-cent reliable when it came to mouse-events for the scatterplot.

A highlighted country will be unhighlighted only if the user clicks on some grey space surrounding the linechart, the map or the scatterplot. By default the text fields show data for the whole world by the way.

For details, please have a look at my code! I tried as well as I could to structure and comment it so that others and also I myself after some time would be able to understand how it works and what it does. I hope together with the earlier sections some possible connections to some central tasks I had in mind are clear without elaborating on it again.

3.1 Note on Preprocessing

For preprocessing, I used R. All the relevant code is in the R folder. The `main.R` script reads in the Covid data and the world countries json file (for harmonizing country names) from the data folder and also writes output (prepared datasets) for the dashboard elements to that folder. The preprocessing decisions executed in `main.R` are based on a thorough exploratory data analysis where I extended my analysis from A4. You can find a markdown file with this EDA (`eda.html`) in the R folder to better understand my approach.

4 Discussion

Actually, apart from the small issues already mentioned above, I would do most things as I did them again: After thinking about the prototype from A4, I created the basic functionality just with two (uncleaned) variables (no dropdown with 12 variables as in the final version) before I did the preprocessing and added further variables to see if it had still sufficient performance. In the end, I just added some minor improvements like the date indicator line, comma separation for the population or axis labels and subplot titles. I think the biggest issue still remaining is that of the sizes. It seems that unless very carefully designed a dashboard will not look equally good on different screens. I also learned that preprocessing plays an important role and in practice more thought would have to go into the question of how to deal with missing data, especially how to

communicate it to the user. Also improving the dashboard with a frontend framework is something that I would certainly do if I had more time. One very specific thing that I learned was that paths are dependent on the ordering of the data: Due to many missing values in Jan/Feb/Mar 2020 these dates were located at the very end of my linechart data which led to straight lines in the linechart paths from end of May 2021 back to January 2020. I solved this by correctly arranging the rows of the dataframe in R.

5 Conclusion

All in all, it worked quite well to create a dashboard with different views allowing for various “axes of analysis”. The parts are all interconnected and at least on my computer I think it looks really decent and also performs very well. The preprocessing is done using R on the downloaded dataset. The `main.R` file could be easily run again on the current version of the dataset as long as all file names are identical (otherwise just the `config.R` file has to be adapted). As already indicated multiple improvements are imaginable but in total I came surprisingly close to my A4 plans.

I put quite some effort and thought into the dashboard even though I initially wanted to finish a little earlier and feel like I cannot really transport it with just a couple of report pages. I would encourage doing “Abgabegespräche” instead or on top of report/readme in the future. I think that might be a good possibility to clarify some things and also get some further tips on how to improve some aspects of the implementation.

6 References

As in A4, one particular dashboard that somewhat inspired me is the Corona Dashboard of the State of Tyrol: <https://experience.arcgis.com/experience/295e492032ff446ca13d40d47193cfd7> Thanks to it, I implemented the idea with the fact fields, decided for a black background for the body and dark grey background with thin lightgrey borders for all the elements.