

Government response to COVID-19 pandemic:

KPI and correlation analysis

Daria Siniatulova

Abstract

In this project I create the tool to conduct automated monitoring of both COVID statistics and government measures to understand situation and government reaction in different countries on the threats of 2020.

I explore the correlation between COVID spread statistics and government response to it.

I used 3 datasets in order to gather pandemic statistics (JHU CSSE COVID-19 Dataset), structured this data by country, region, income level (WDI Dataset) and date since January 1st 2020 till today, and then combined resulting dataset with government response index (Oxford Covid-19 Government Response Tracker) to calculate correlations and get insights on different countries reaction to this threat.

I mostly use pandas built-in methods to work with tabular data, along with several visualization techniques to present the research results.

Datasets

Dataset #1:

Name: JHU CSSE COVID-19 Dataset.

Source: <https://github.com/CSSEGISandData/>

Dataset #2:

Name: Oxford Covid-19 Government Response Tracker

Source: <https://github.com/OxCGRT/>

Dataset #3 (supplementary):

Name: World Development Indicators Dataset (WDI)

Source: <http://api.worldbank.org/v2/country/all/indicator/>

Motivation

I decided to explore what new indicators could reveal summary on governments respond to COVID-19 pandemic. Being a municipal servant myself I would like to have the tool to monitor this activity to understand reaction in different countries on the threats of 2020. My goal was to write the code that would let me update the information on the fly daily to prepare summary of situation monitoring automatically.

I consider government policies indicators and pandemic statistics to be one of the most interesting datasets to research in mid 2020. I specifically use visualization technics that are different from JHU approach to their famous [dashboard](#) and suggest supplementary visual solutions.

I believe the ideas of this project could be valuable for many people like me who is concerned with pandemic issues or try to compare different governments approaches to this problem.

Data preparation and cleaning

I described my approach to cleaning and preparation of the data in special paragraphs of each chapter of this project. Particular steps are always related to the structure and completeness degree of the data, as well as my goals in terms of using data: for calculating, for creating animated map, for plotting interactive line graphs, bubble graphs or sunbursts.

The problems I have encountered with the dataset were the following: lack of values for particular countries or dates, excessive values that can not be plotted on map or can not be visible there even if plotted (for example, COVID cases on cruise ships "Diamond Princess and "MS Zaandam", or in Vatican), wrong ISO-3 codes for the countries, outdated choropleth maps (for example, Kosovo could not be plotted, even though the shape is there, and I cleaned the dataset, and I fixed the ISO-3 codes).

Research Questions

I have formulated 3 research questions to answer with this project:

- is there a significant correlation between cumulative cases of COVID-19 and cumulative deaths related to it?
- is there a significant correlation between maximum government response index and government healthcare expenditure in previous years?
- is there strong positive correlation between confirmed COVID-19 cases and government response index over time?

Hypothesis to check

- yes, there is significant positive correlation between cases of COVID-19 and deaths;
- no, strict response to COVID-19 in 2020 and better funding healthcare systems in 2019 are not significantly correlated;
- yes, correlation between confirmed cases and government response over time is strong.

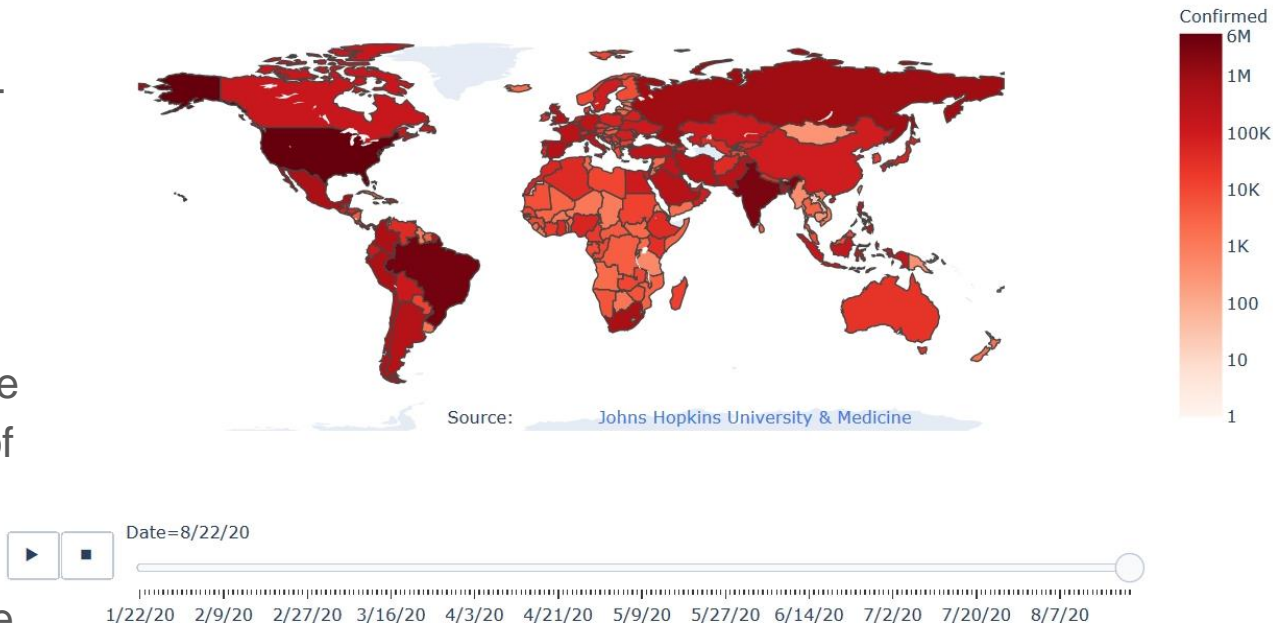
Note: I explore indicators using Russia, Canada, USA, China and Brazil to describe and compare trends. I also plotted world's average data to provide additional context to comparisons.

Methods

- I mostly used pandas to work with tabular data. I used plotly express, plotly graphic objects, choropleth and matplotlib to visualize the results, and then used the visualizations as tools to guide my data analysis.
- I read scientific research by the largest universities on the topic of interest.
- I used pandas built-in methods to access the datasets, connecting my code directly to the repositories (Datasets 1 and 2) or API (Dataset 3).
- I implemented 3 different visualization techniques to present the research results.
- I used several comparison methods: correlation coefficients, scatter plots, merging datasets, and others using examples of Russia, Canada, USA, China and Brazil.
- To describe and compare trends I also plotted world and world's average to provide additional context.
- I use interactive plots and animated world maps where everyone can find their countries and explore the depths of the data.

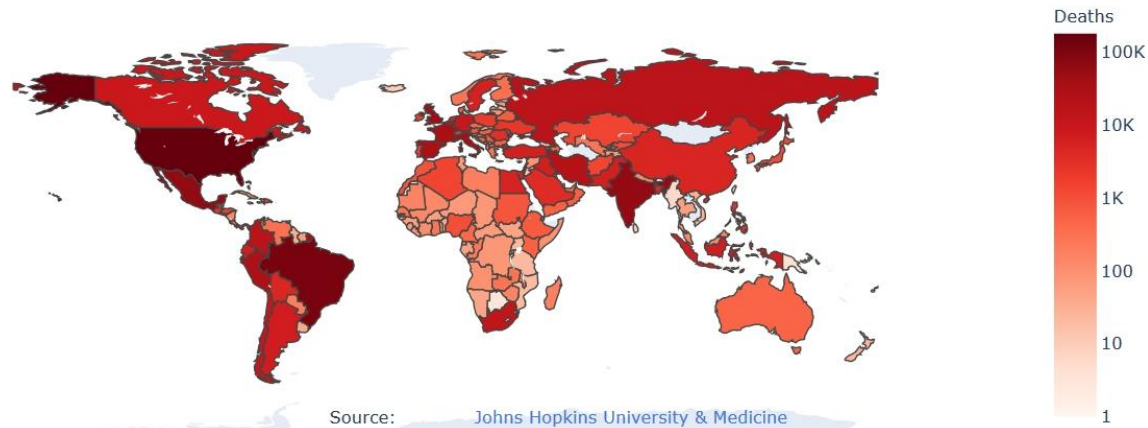
Confirmed cases

As of August 20, 2020
maximum number of cases -
5.6 M - were confirmed in
USA. There were performed
about 69.6 M tests there, so
testing has covered every
210 out of 1000 people in the
country. This could be one of
the reasons of the record
high number of confirmed
cases in comparison with the
rest of the world.



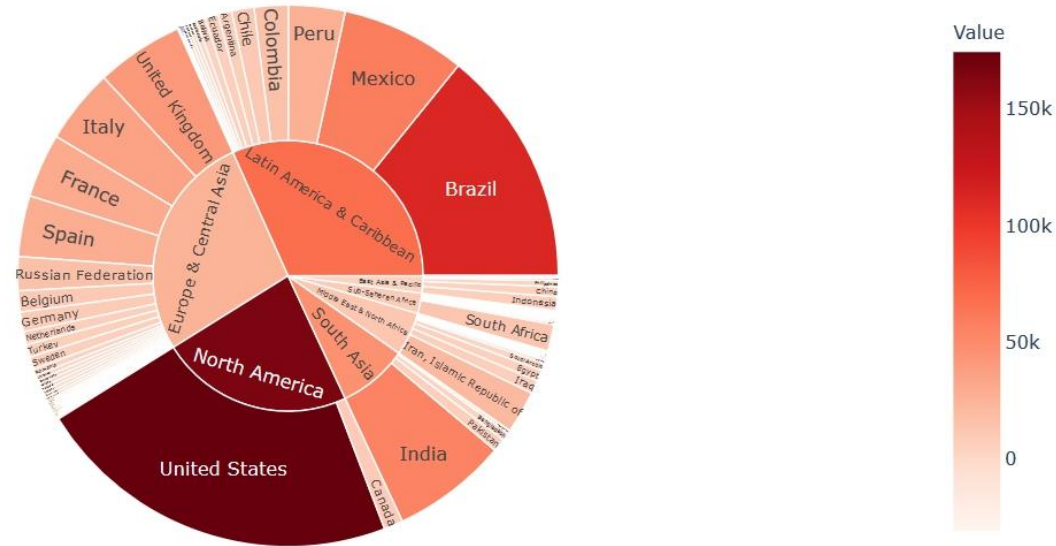
Deaths

As of August 2020 maximum number of lethal end cases were registered in USA. Mongolia and Turkmenistan have not reported COVID-related deaths.



Deaths: Structure by regions

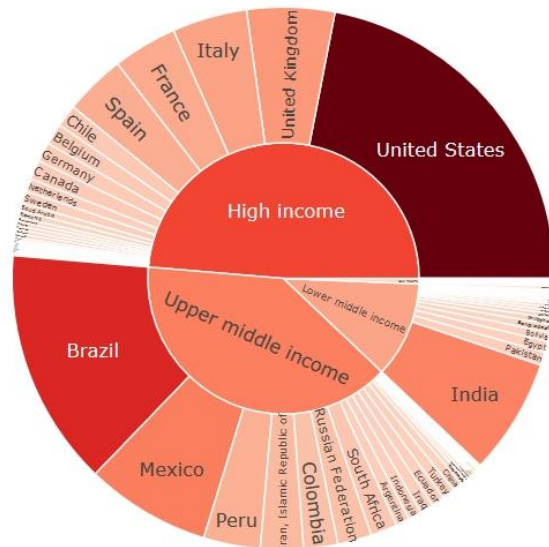
United States in North American Region, Brazil and Mexico in Latin America & Caribbean region, India in South Asia and UK, Italy, France and Spain in Europe region are the countries most affected by COVID-related deaths.



Deaths: Structure by income level

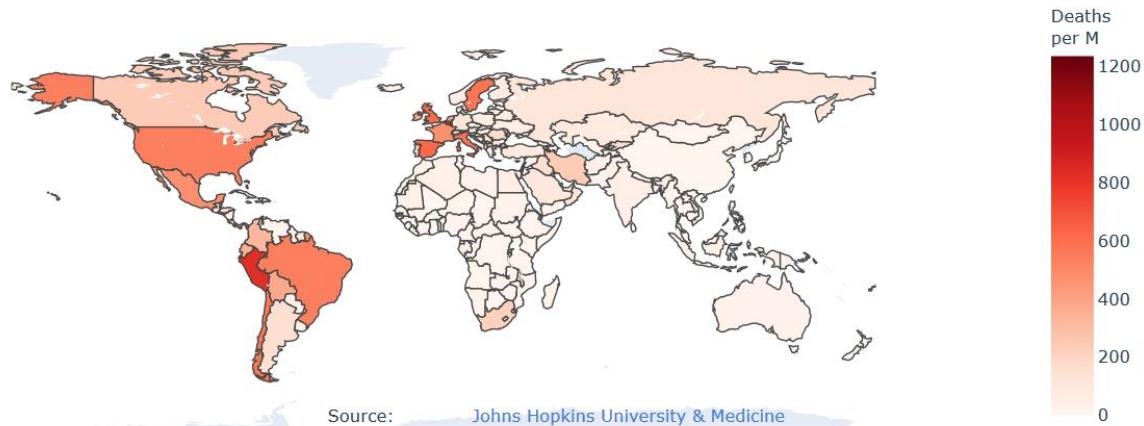
“High income” countries and “Upper medium income” countries cover 41% and 39% of total COVID-19 deaths respectively.

“Low” and “Lower medium income” countries cover other 20% cases.



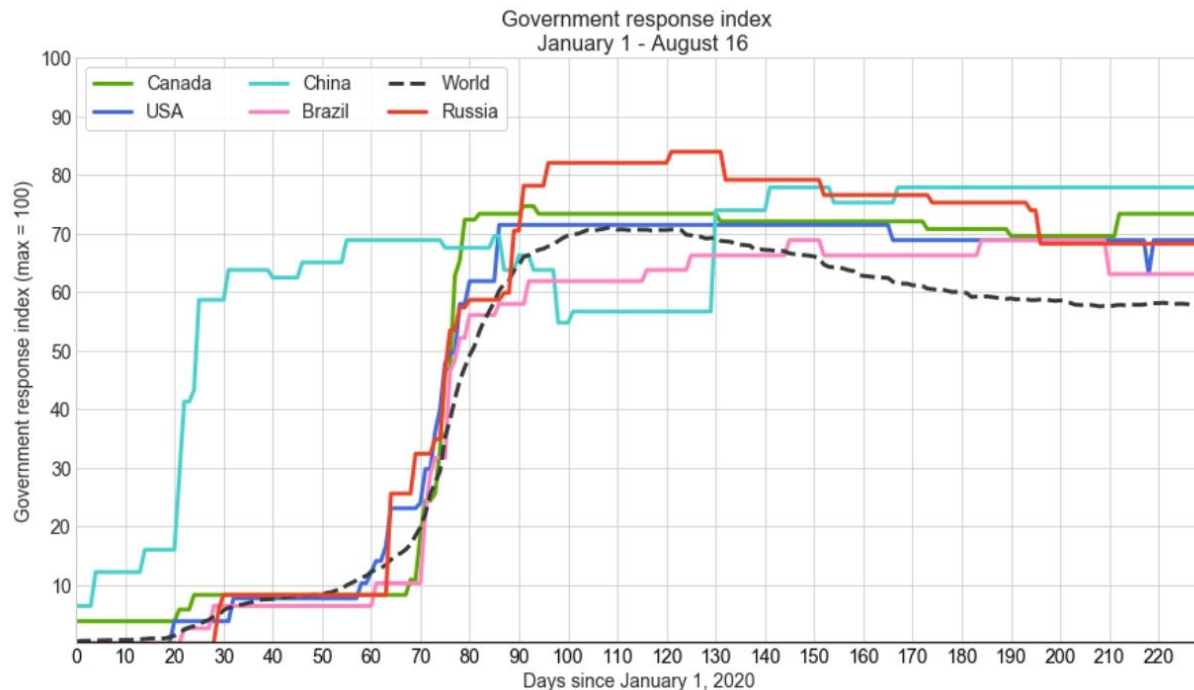
Deaths per million population

As of August 21, 2020
maximum number of lethal
end cases per million
population was registered in
Belgium (873.4) and Peru
(838.8); losses are high in
UK (622.7), Spain (616.7)
and Italy (586.1).



Government response index

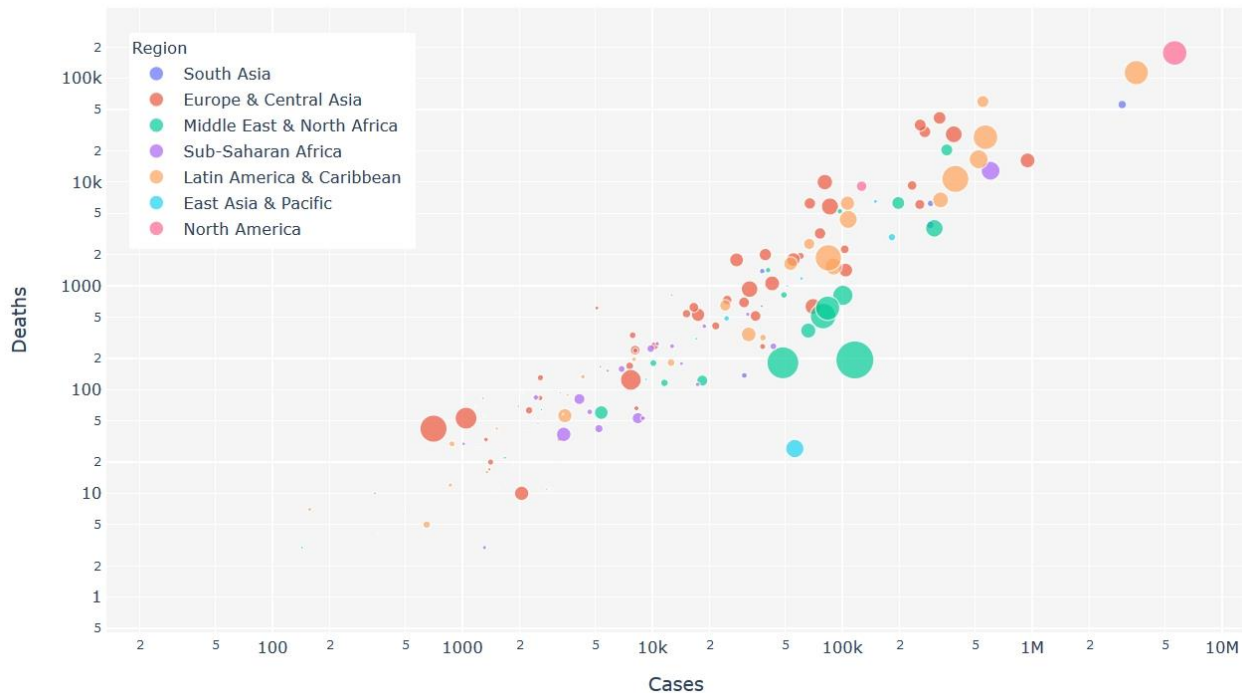
Government of China responded first, dealing with Wuhan outbreak in December 2019. It took actions 50-70 days ahead of the rest of the world. Other countries preferred not to make costly decisions and were not enforcing strict measures till mid-March.



Confirmed cases and deaths

Number of confirmed cases and COVID-related deaths over time have strong positive correlation.

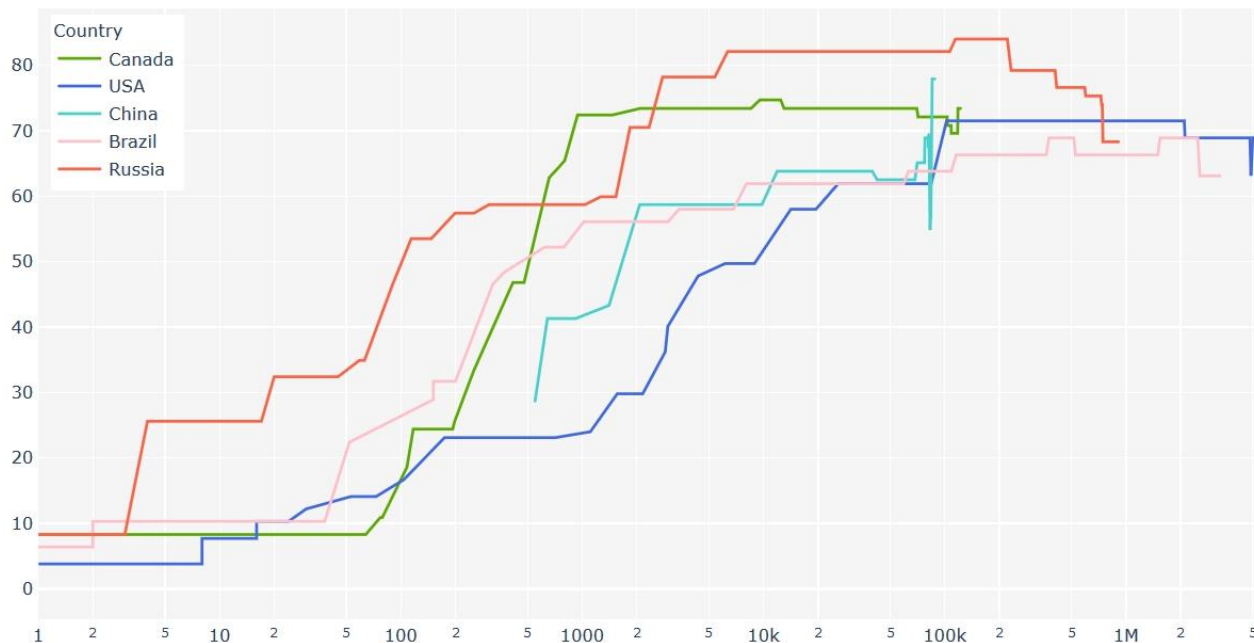
That is why the governments need to react timely and must be proactive to prevent number of new cases from rising. This issue is crucial and time-sensitive.



Confirmed cases and government response index

Number of confirmed cases (x axis) and government response index (y axis, max = 100) over time have strong positive correlation: correlation coefficient is equal 0.94.

The plot allows to compare different countries approach to response measures and compare the effects they may have on a daily basis.



Findings 1 – Cases and Deaths

- Yes, there is significant positive correlation between cumulative number of cases and deaths from COVID-19. That is why it was (and still is) so important for the governments to act strict and timely to prevent further virus spread and it's damage to society and economy.

Findings 2 – Response and Health expenditures

- No, there is no correlation between maximum government response index and government healthcare expenditure in previous years. The governments that demonstrate higher response index to COVID-19 in 2020 were not necessarily better in funding their healthcare systems in previous years.

Findings 3 – Cases and Response

Yes, correlation between confirmed cases over time and government response index over time is very strong (coefficient is equal to 0.94). At the same time, correlation between maximum cumulative number of cases and maximum government response is very weak (0.26), and correlation between deaths and maximum government response is even weaker (0.19).

Time factor is crucial to government response in order for measures to have desired effect and prevent virus from rapid spreading and it's damage to society and economy.

Limitations

The Dataset 2 has inherent limitations. As the indicators are work in progress, and research team has to monitor 17 sub-indicators to calculate another 4 aggregated ones for 200+ countries to publish daily, their data has the lag for 1-2 weeks back. This factor should be taken into account at all times when comparing with the daily updated pandemic statistics.

In order to check the results of this research the indicators should be monitored longer, that current available period (January-July or January-August).

Some of the described trends and findings are not fully confirmed for the countries of the lower middle income and low income due to limitations of the dataset and COVID-19 stage in these countries.

Conclusions

I have confirmed all 3 hypothesis: I found out that positive correlation between cases of COVID-19 and deaths from it is very strong, as well as between confirmed cases and government response. Government response index and government healthcare expenditure in previous years are not correlated. These findings are relevant for different regions, but their relevancy for the lower middle income countries and low income countries is not fully confirmed yet due to limitations of the dataset.

It is too early to make final conclusions about COVID mortality for different regions and income levels as disease statistics database is only growing and information available now is incomplete. The indicators that we could use to describe the situation unfolding now are work in progress themselves. I would do my best to add my share to the knowledge and ideas database for this important work.

Acknowledgements

1. Johns Hopkins University & Medicine provided the dataset “JHU CSSE COVID-19”
2. Blavatnik School of Government, University of Oxford provided the dataset “Oxford Covid-19 Government Response Tracker”.
3. World Bank provided the dataset “World Development Indicators”

References

I have used a research paper by Blavatnik School of Government on their project “Oxford Covid-19 Government Response Tracker”:

“Variation in government responses to COVID-19”, May 2020 - Blavatnik School of Government, University of Oxford, by team of authors:

Thomas Hale, Noam Angrist, Beatriz Kira,
Anna Petherick, Toby Phillips, Samuel Webster

Source: www.bsg.ox.ac.uk

Annex:

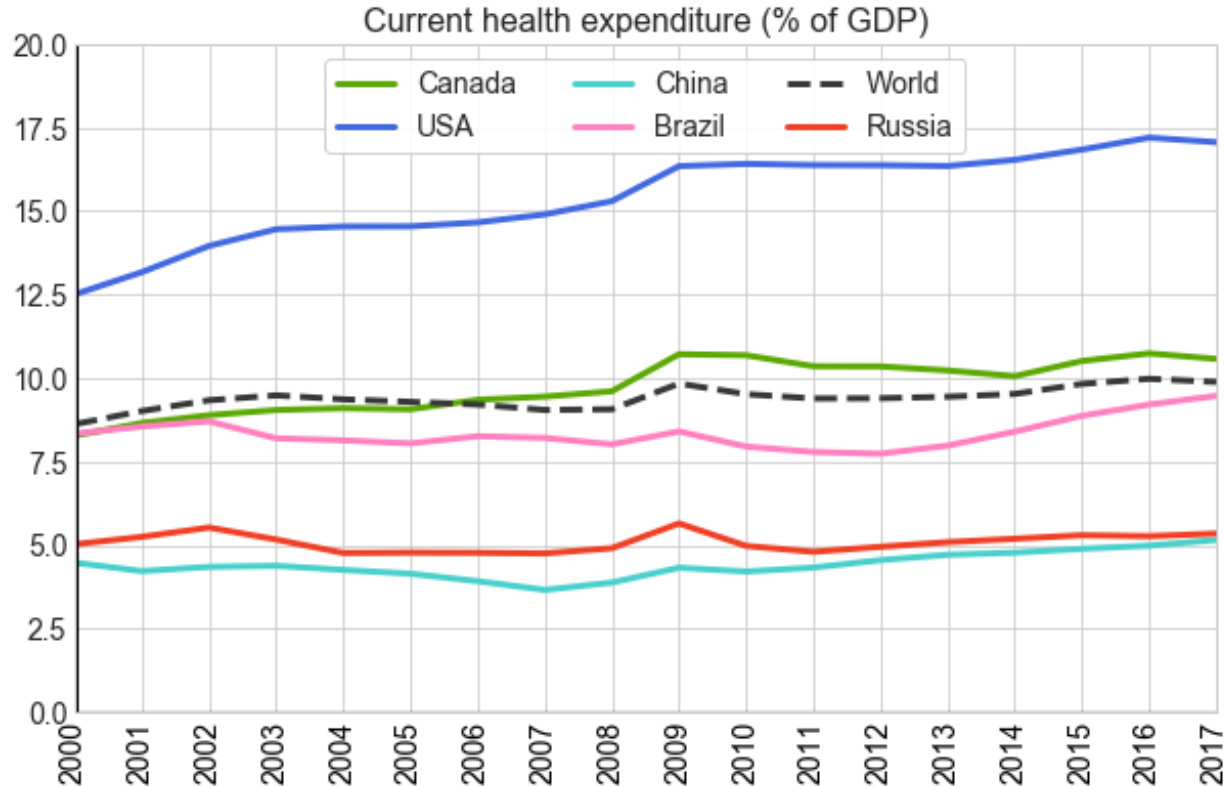
Government health expenditure trends and maps

Additional slides from Week 6: Mini-Project with
World Development Indicators Dataset

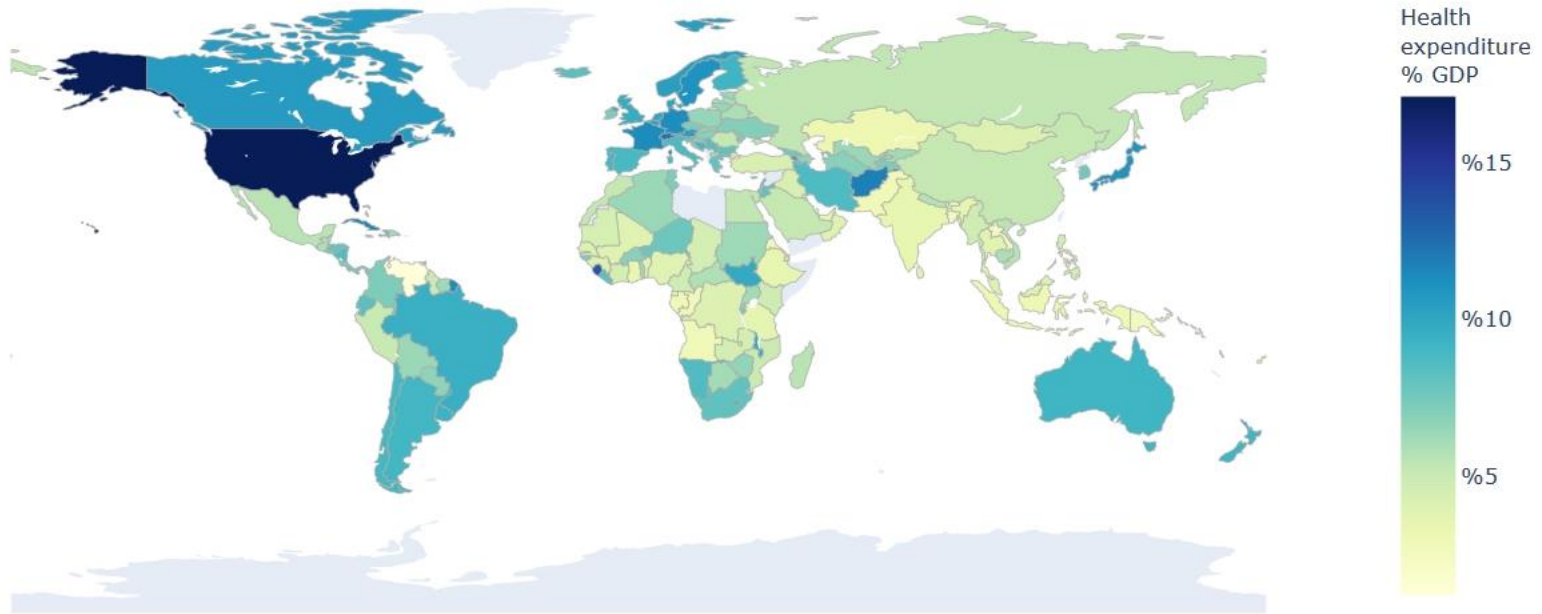
Health expenditure trends (% GDP)

Canada increased health expenditure from 8.3 to 10.7% of GDP, while Russia - from 4.7 to 5.6%.

Russia's expenditure on healthcare system are twice lower than world's average, although its GDP per capita is greater than average since 2003.



Health expenditure map (% GDP), 2017

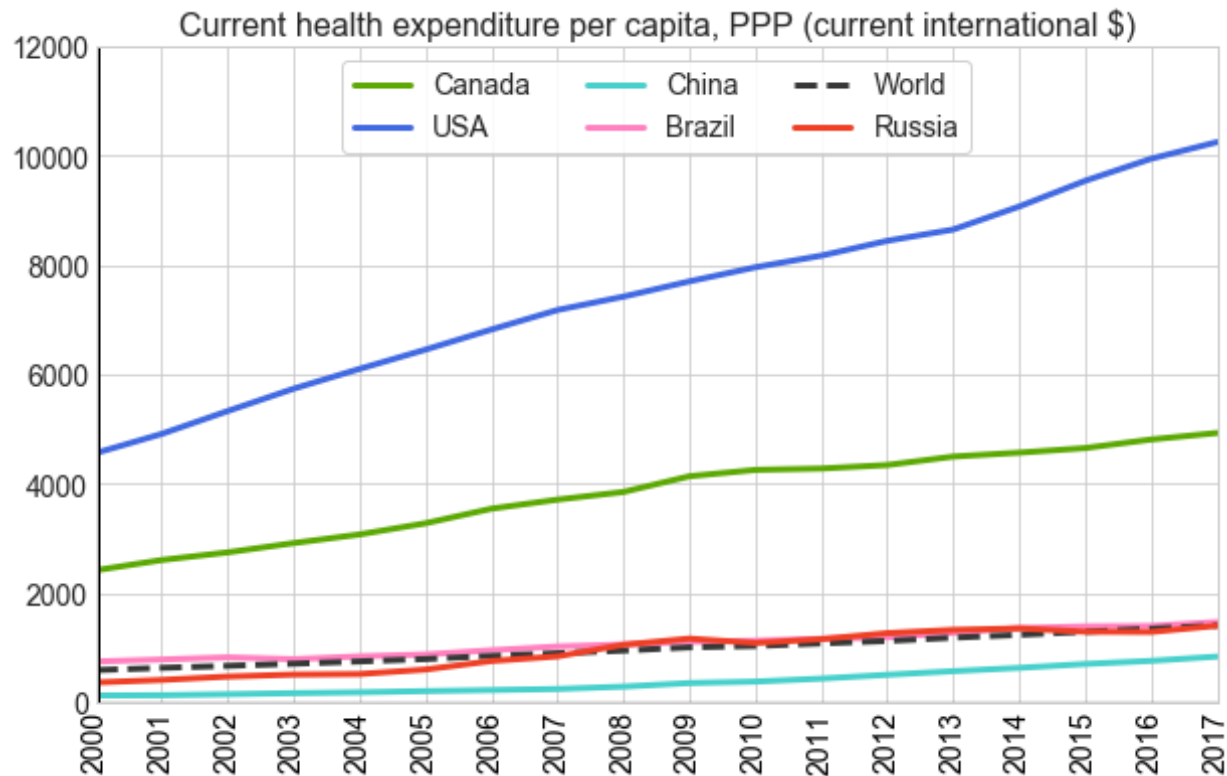


Health expenditure trends (int. \$ by PPP)

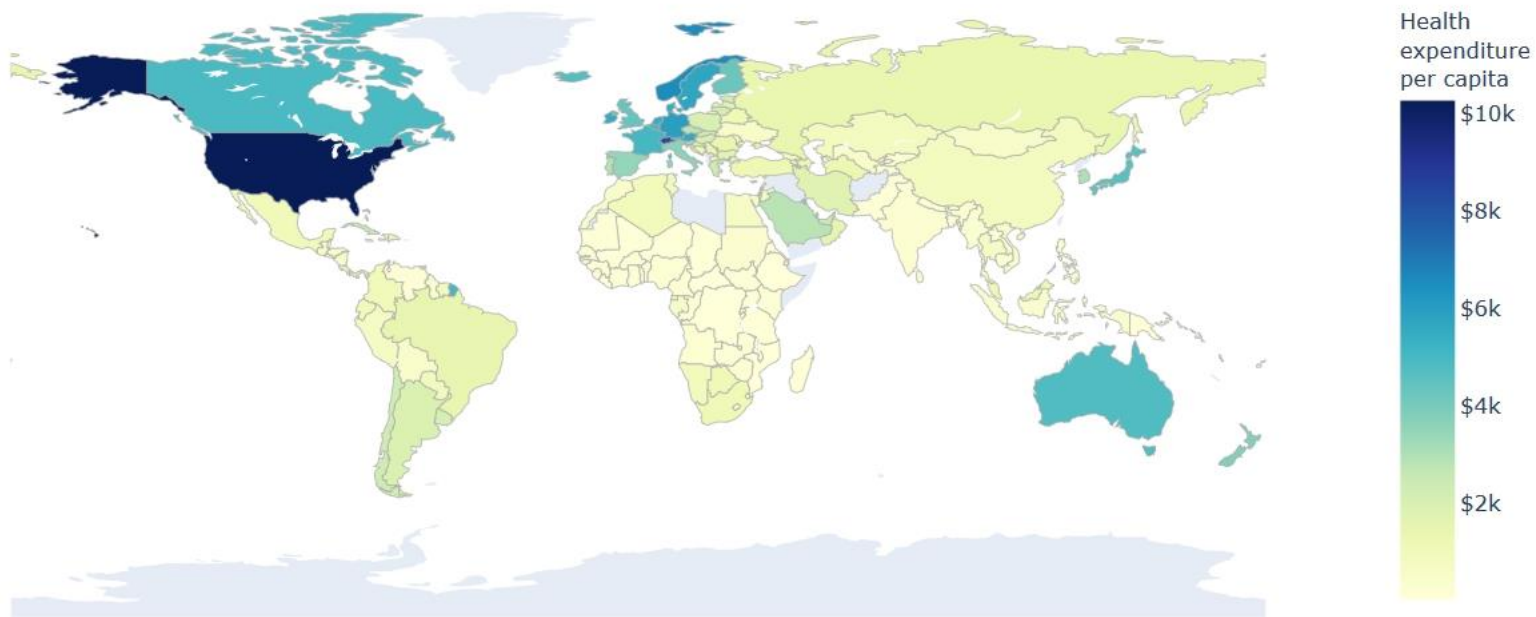
Canada increased health expenditure from 2418 to 4929 int. \$, and Russia - from 368 to 1404 int. \$.

In Canada health expenditure were 3.5 times greater than world's average in 2017.

In Russia they were lower than world's average.



Health expenditure map (int. \$), 2017

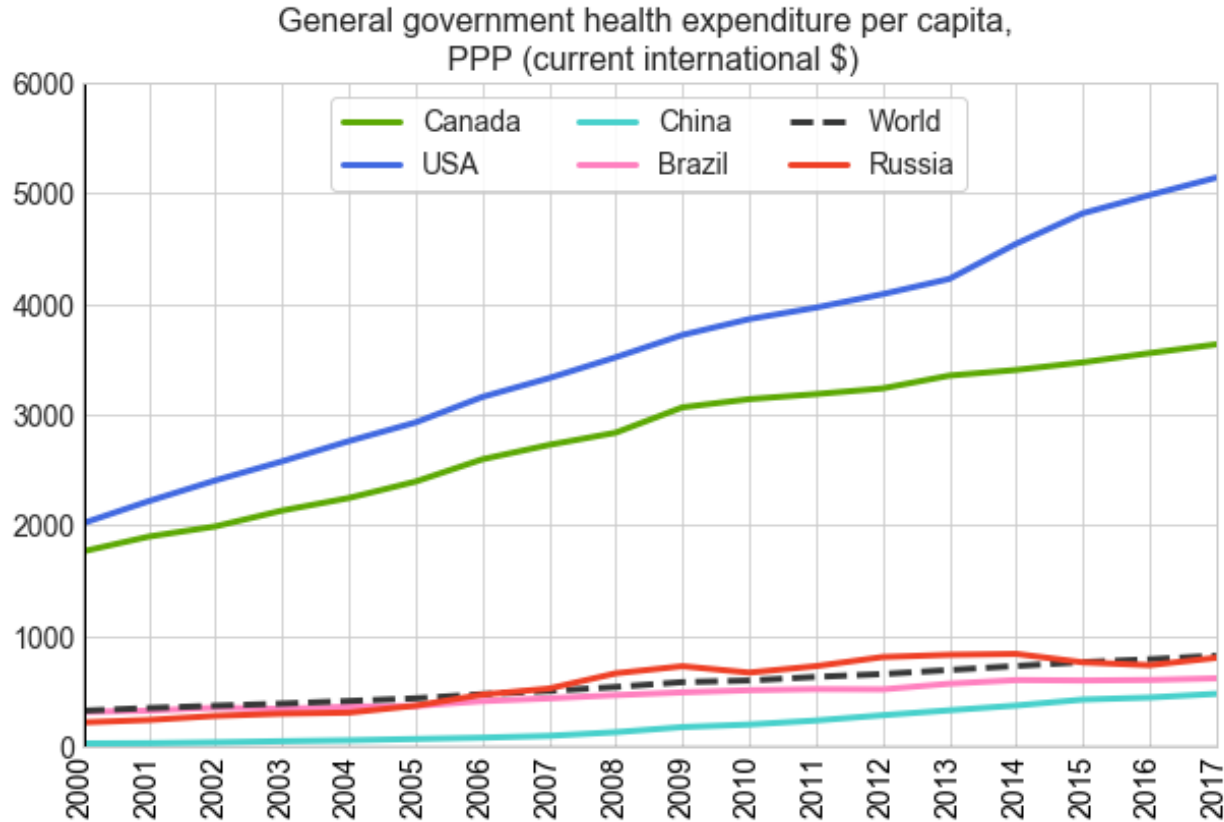


Government health expenditure trends

In Canada government health expenditure were 3633 int. \$, and Russia - from 802 int. \$ in 2017.

In Canada government health expenditure were 4.4 times greater than average in 2017.

In Russia they were lower than world's average.



Government health expenditure map (int. \$), 2017

