

# COVID-19 DEEP DIVE

Team 125: Han Li, Rui Nah Koh, Wenjun Wu, Ying Zhang, Jiani Qian

## Summary (Motivation & Introduction)

In this study, we aim to understand what factors contribute to the spread and containment of COVID-19, by exploring the two topics: **1) the correlation between countries’ development indicators and their pandemic situations**, and **2) the differentiating success factors in controlling the spread of the virus**. Through data analysis and visualization on COVID-19 cases and deaths, countries’ development indicators, and control and containment measures, we hope to shed some light on factors correlating with the transmission, mortality, and control of COVID-19, and also provide some lessons learnt for governments, world organizations, and even the entire human race.

## Data

Our raw data are **downloaded directly** and processed in **Jupyter notebook**. Apart from Country development indicators, COVID-19 cases, death, vaccine, and NPI datasets need to be regularly updated.

Data source	Usage	Size on disk	Number of records	Duration
World Bank	Country development indicator	201.6 MB	378576	Only 2018 used
Kaggle	COVID-19 cases and death	4.1 MB	85653	2020/02/15-2021/03/11
Our World in Data	Vaccine administered	651 KB	10782	2020/12/13-2021/04/05
CDC COVID data tracker	Vaccine administered in US	25 KB	63	Updated on 2021/04/07
ECDC COVID vaccine tracker	Vaccine administered in EU	1.1 MB	20605	2020/12-2021/04
ACAPS	NPI implemented by governments	4.2 MB	23923	2020/01/01-2021/01/15

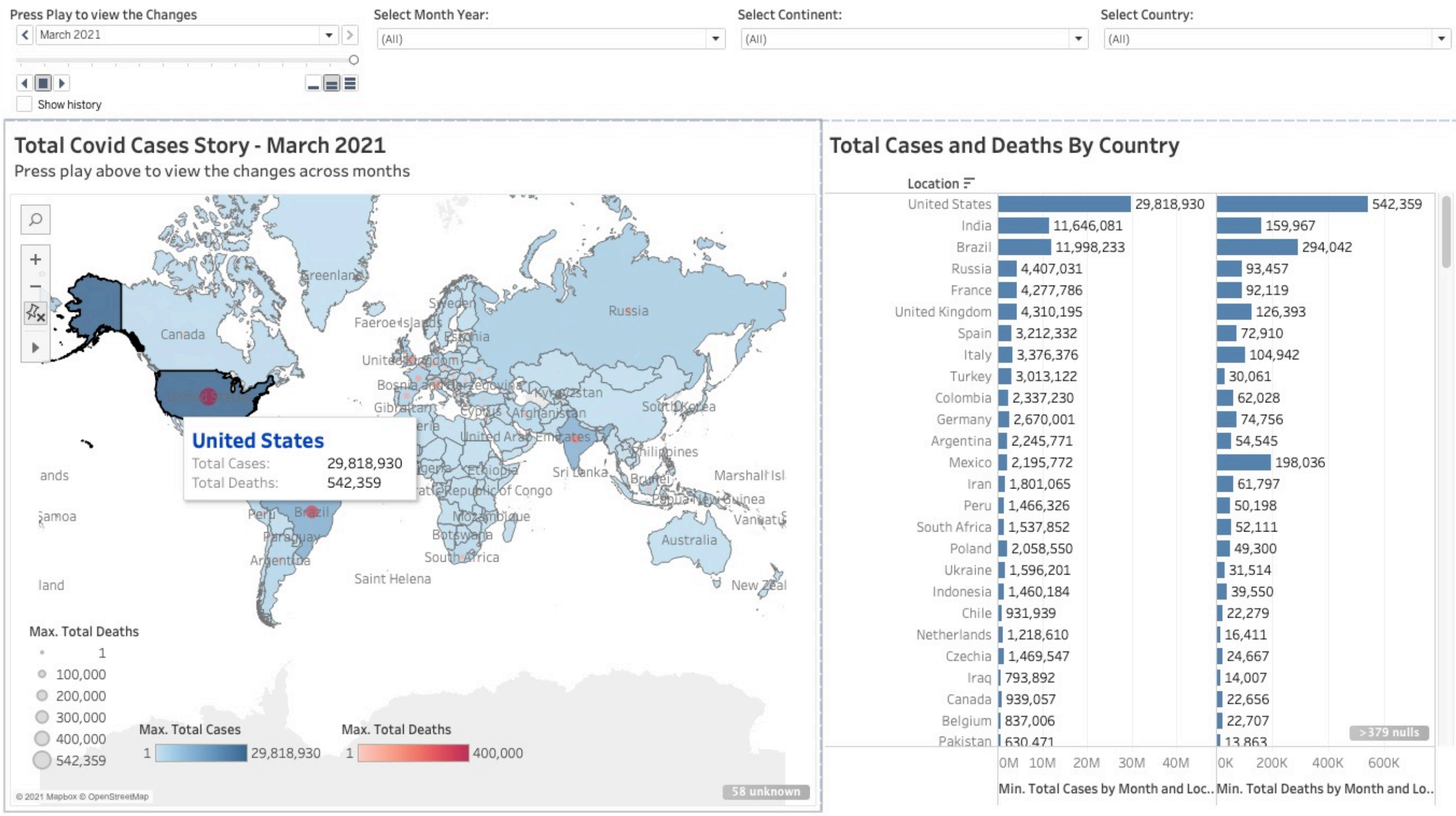
## Our Approach

We chose **Tableau** as our user interface to provide a holistic view of COVID-19 evolvement, including case spread, vaccine distribution and NPI tooltips across location and over time. It is a one-stop dashboard for all COVID-19 related data visualization. All graphs and charts are provided with interactive toolbars for any deep dive needed by users. Innovative **machine learning models** (clarified below) are explored to investigate the impact and the magnitude of impact of development factors, control measures and vaccinations on COVID-19 spread and containment.

➤ Animation

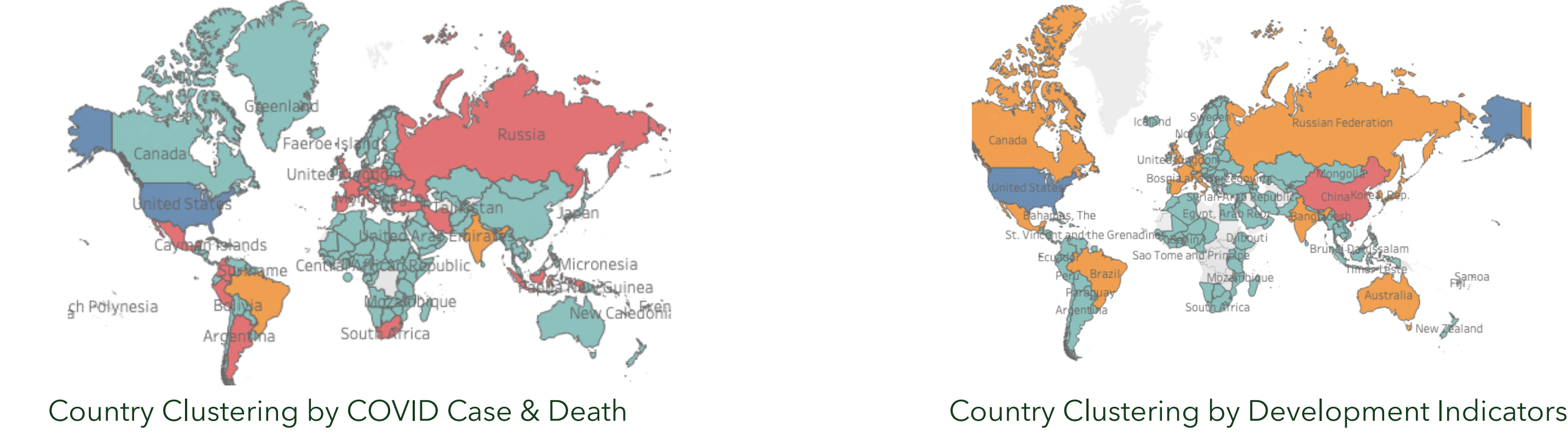
➤ Selection

➤ Tooltip



## What are the factors affecting COVID Development (Experiments + Results)

### Development Indicators



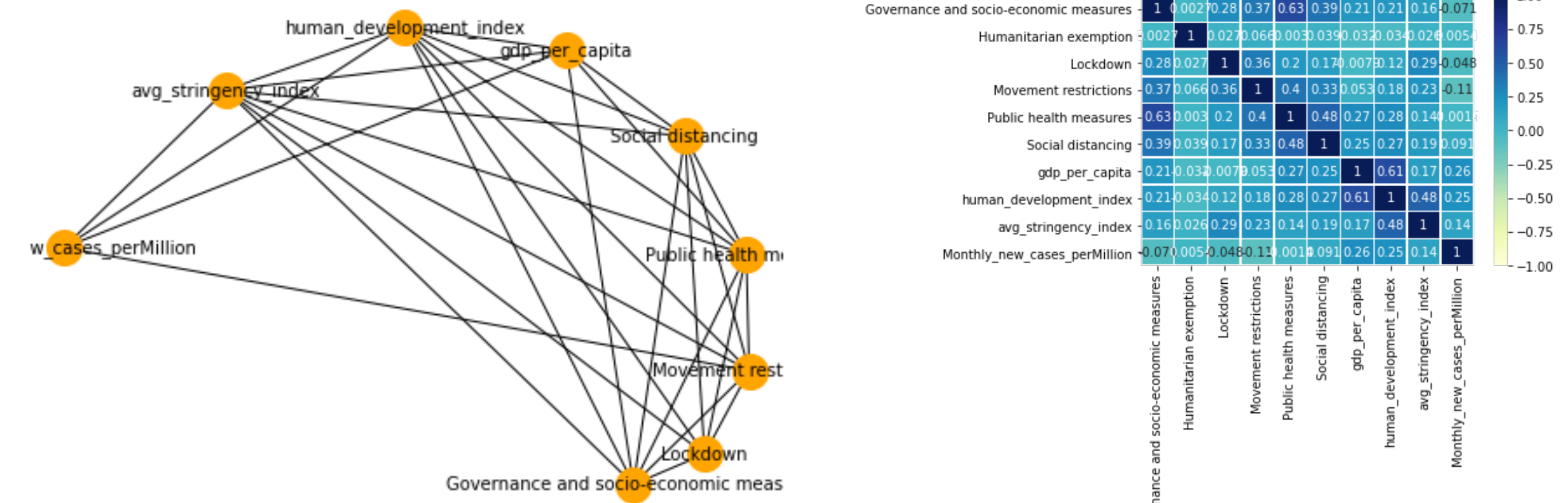
**No strong co-trend found** between COVID spread and development indicators given very different clustering results.

Data Processing	Independent Variables	Dependent Variable	Model	R squared
PCA	Absolute value of development indicators in 2018	Total Case		0.229
		Total Case		0.611
	Absolute value of development indicators in 2018	Infection Rate	Linear Regression	0.331
Collinearity Check		Fatality Rate		0.104
	Moving Average of development indicator values in 2018	Total Case		0.626
	YOY % change of development indicators in 2018	Total Case		-0.113
	<b>Absolute value of development indicators in 2018</b>	<b>Total Case</b>	<b>Step-wise Linear Regression</b>	<b>0.839</b>
	Absolute value of development indicators in 2018	Total Case	Decision Tree Regressor	-0.077
	Absolute value of development indicators in 2018	Total Case	Random Forest Regressor	0.28

- Combination of *Absolute value of development indicators*, *Total Case* and *Step-wise Regression* achieved the highest R-squared.
- More developed countries with more advanced economic development, wider education coverage and better infrastructure support tend to have more infected cases.

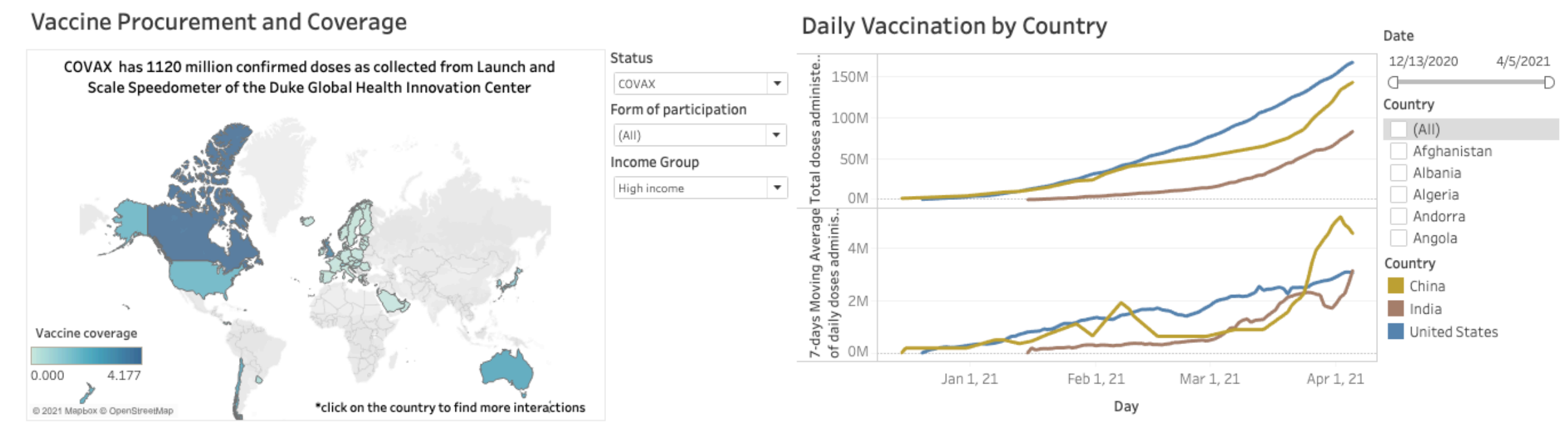
### Containment measures

#### ➤ NPI analysis



- Graph network: **stringency index** and **movement restrictions** are **more correlated** to the number of **cases** per million capita
- Correlation matrix: **negative correlation** between **cases** and **movement restrictions**

#### ➤ Vaccination progress analysis



#### ➤ The top 3 most important factors in predicting new cases

- Decision tree: stringency index, new vaccination smoothed per million and GDP per capita

## Discussion

Compared to current COVID-19 dashboards available, our analysis and visualization is **global and comprehensive**.

- Possible reasons for **developed countries** having more cases could be they have **more resources** to test and report cases, a **higher population density**, and also **better mobility** with more extensive transportation networks.
- Development indicators** alone do not have strong explanatory power for COVID development while **adding NPI and vaccinations into the picture** increases the overall model performance to predict COVID new infection.
- Issues of inequality and justice around vaccine delivery** still exists especially for low-income countries and non-participants of COVAX. With our vaccine coverage analysis, we hope more targeted support can be provided to some countries or regions.
- The huge gap in full vaccination rate between countries** draws attention to educate the community about safety and effectiveness of vaccination.