

---

# CS-GY 6513 Big Data

## Effectiveness of Government Measurements on Virus Spread

Team Members:

Yixiu Chen <[yc3691@nyu.edu](mailto:yc3691@nyu.edu)>

Hao Huang <[hh1811@nyu.edu](mailto:hh1811@nyu.edu)>

---

# Project Overview

- Government reaction/response towards COVID-19 has become a topic.
- Different level of restrictions could lead to different impacts on:
  - The number of confirmed and death cases
  - Spread trend
  - Other fields, e.g., education and economy, etc.
- In this project, we attempt to analyze the effectiveness of government measures on virus spread.
- By looking at dates when government measures are implemented across different countries, we aim to find relationship between government response and virus spread.



# Preliminary Research Questions

**Research question(s) we are interested in:**

- Is there a correlation between government measurements and virus spread?
- What is the effect of government measurements on virus spread?
  - How effective at reducing virus spread are each of the government measures taken?
  - Whether there exists any pattern between government measurements and virus spread?
  - Can we make predictions based on the patterns?
- Summarize suggestions based on our analysis.

# Data Cleaning and Wrangling

- The schema of cleaned data (an example of the few first lines):

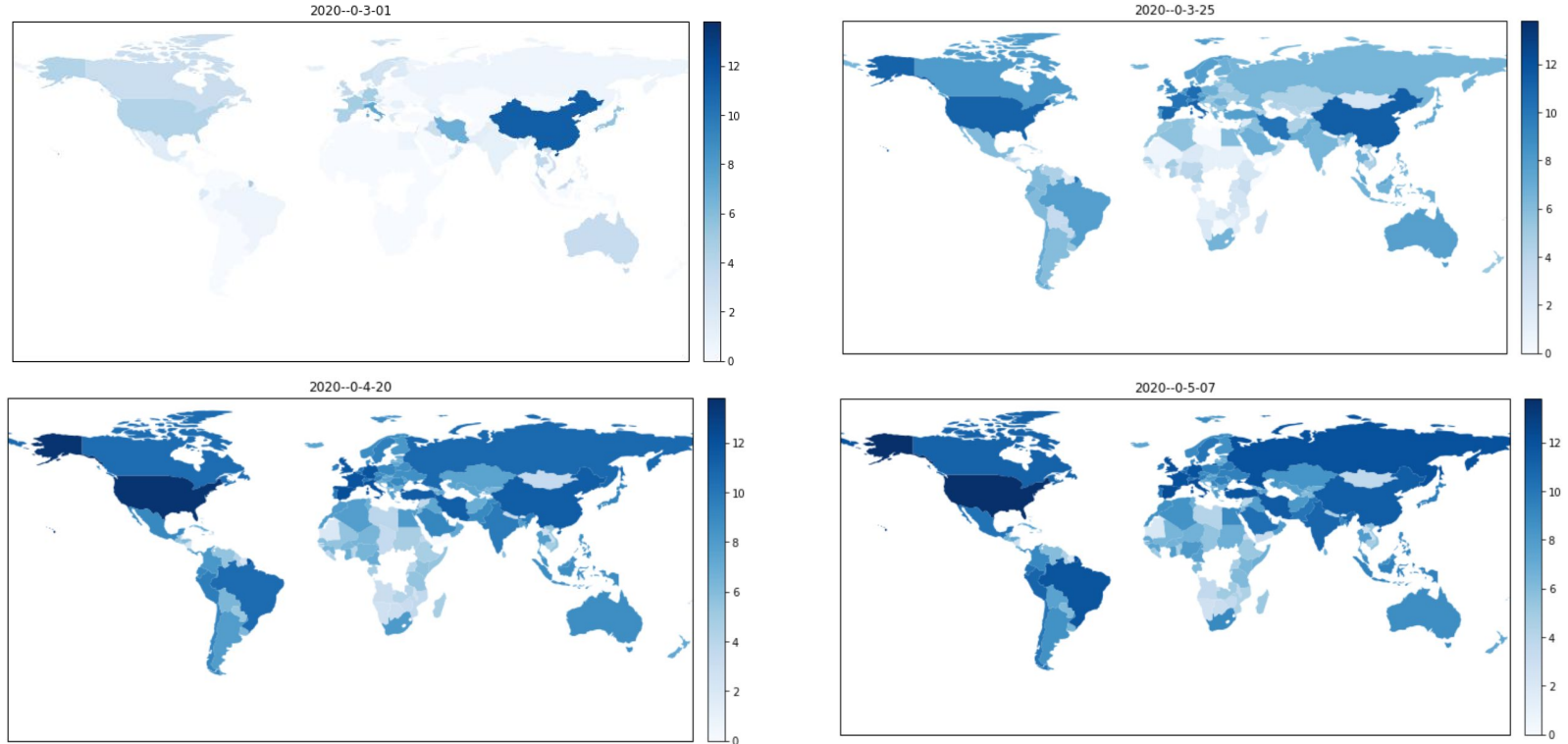
	Date	Country	School closing	Workplace closing	Cancel public events	Restrictions on gatherings	Close public transport	Stay at home requirements	Restrictions on internal movement	International travel controls	Public information campaigns	stringency	Confirmed cumulative	Confirmed daily
0	2020-04-08	Argentina	3.0	3.0	2.0	4.0	2.0	3.0	2.0	4.0	2.0	100.00	1715	87
1	2020-04-09	Argentina	3.0	3.0	2.0	4.0	2.0	3.0	2.0	4.0	2.0	100.00	1795	80
2	2020-04-10	Argentina	3.0	3.0	2.0	4.0	2.0	3.0	2.0	4.0	2.0	100.00	1975	180
3	2020-04-11	Argentina	3.0	3.0	2.0	4.0	2.0	3.0	2.0	4.0	2.0	100.00	1975	0
4	2020-04-13	Argentina	3.0	3.0	2.0	4.0	2.0	3.0	2.0	4.0	2.0	100.00	2208	66
5	2020-04-14	Argentina	3.0	3.0	2.0	4.0	2.0	3.0	2.0	4.0	2.0	100.00	2277	69

- Nine categories of restrictions, each is digitized 3/4 levels represented as a number.
- The stringency index is a sum of all restrictions and scaled between 0~100, which is an overall indicator of government response.

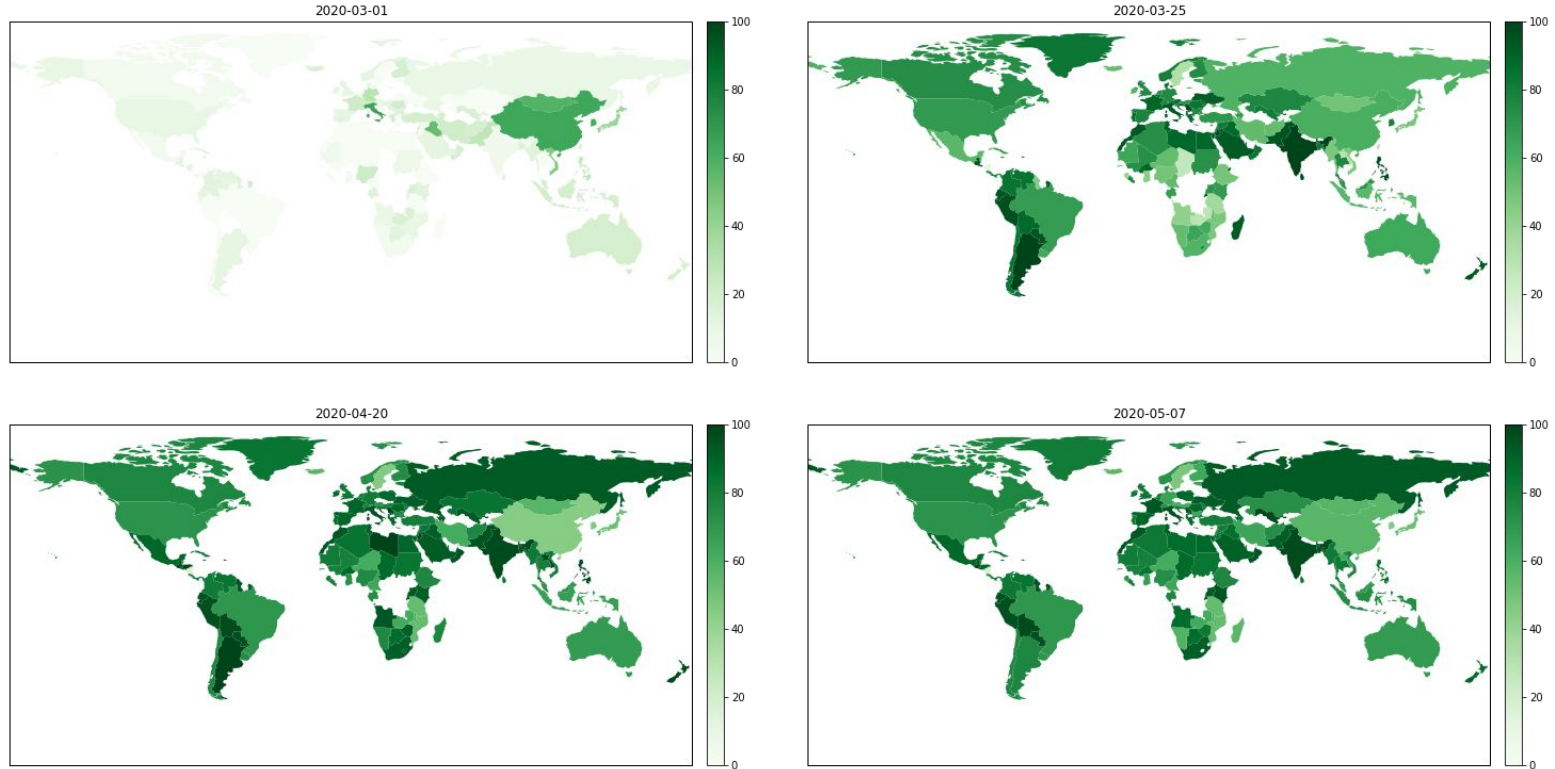
# Research Methods

- Visualization
  - Plotting figures to qualitatively demonstrate the correlation between government measurements and virus spread
- Regression/Classification
  - Machine learning tool, i.e., decision tree, to find the most effective government restrictions.
  - Time series analysis tool i.e., ARMA and VAR model, to
    - quantitatively analyze the pattern between government measurements and virus spread
    - make some prediction about the future

# Visualization: Confirmed Cases

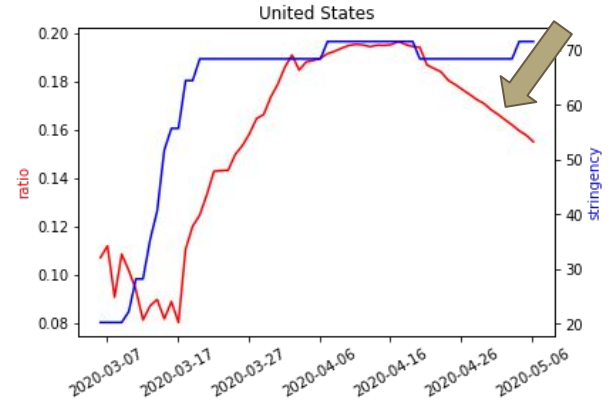
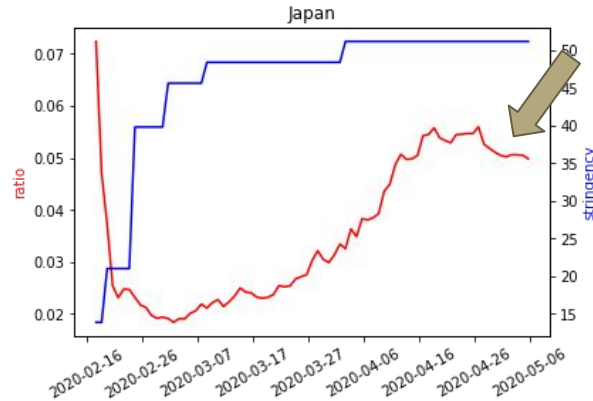
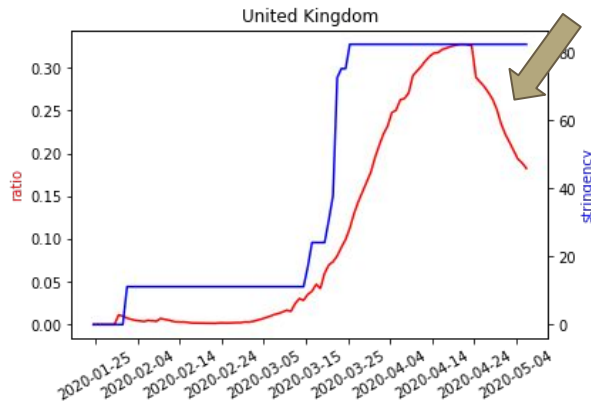


# Visualization: Stringency Index



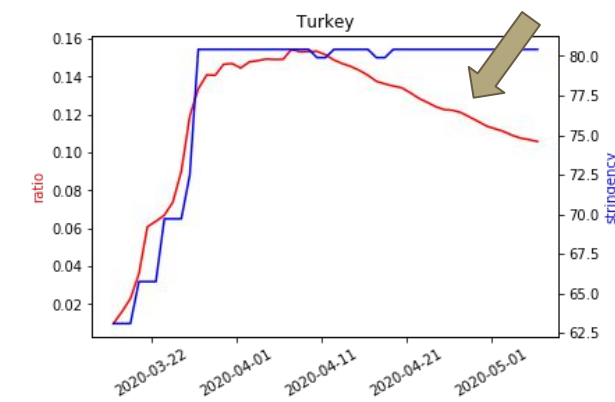
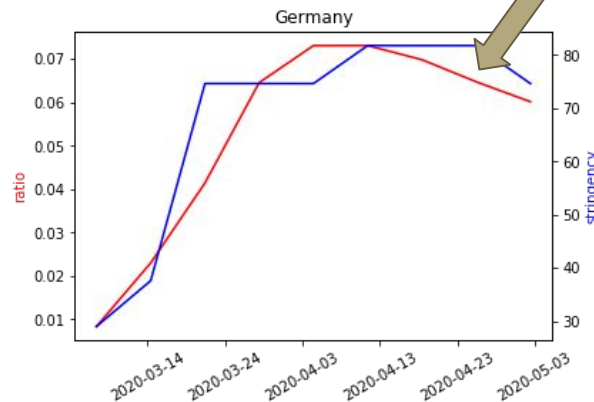
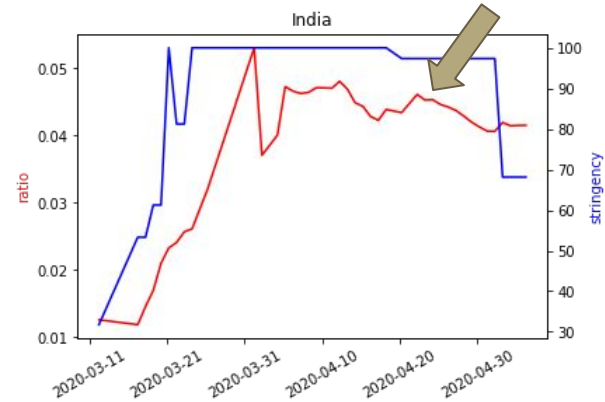
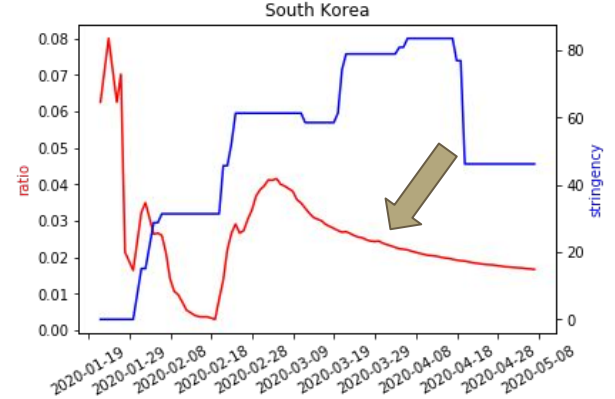
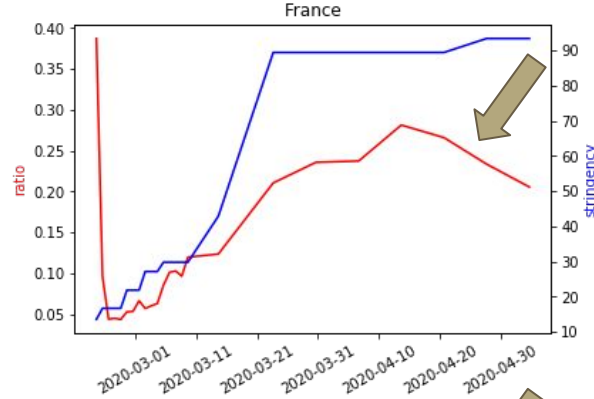
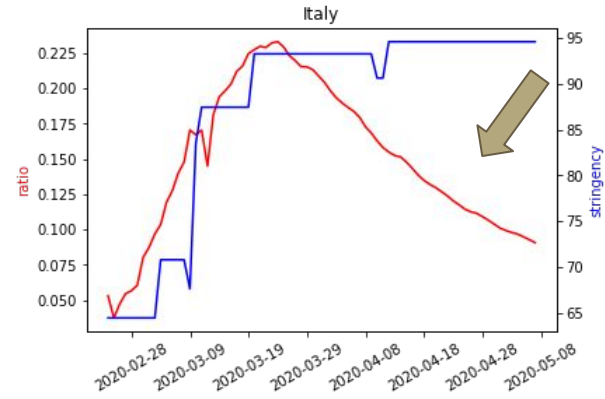
# Qualitative Analysis

- Ratio = cumulative confirmed cases / cumulative testing cases
- Nearly negative correlation between stringency index and the ratio





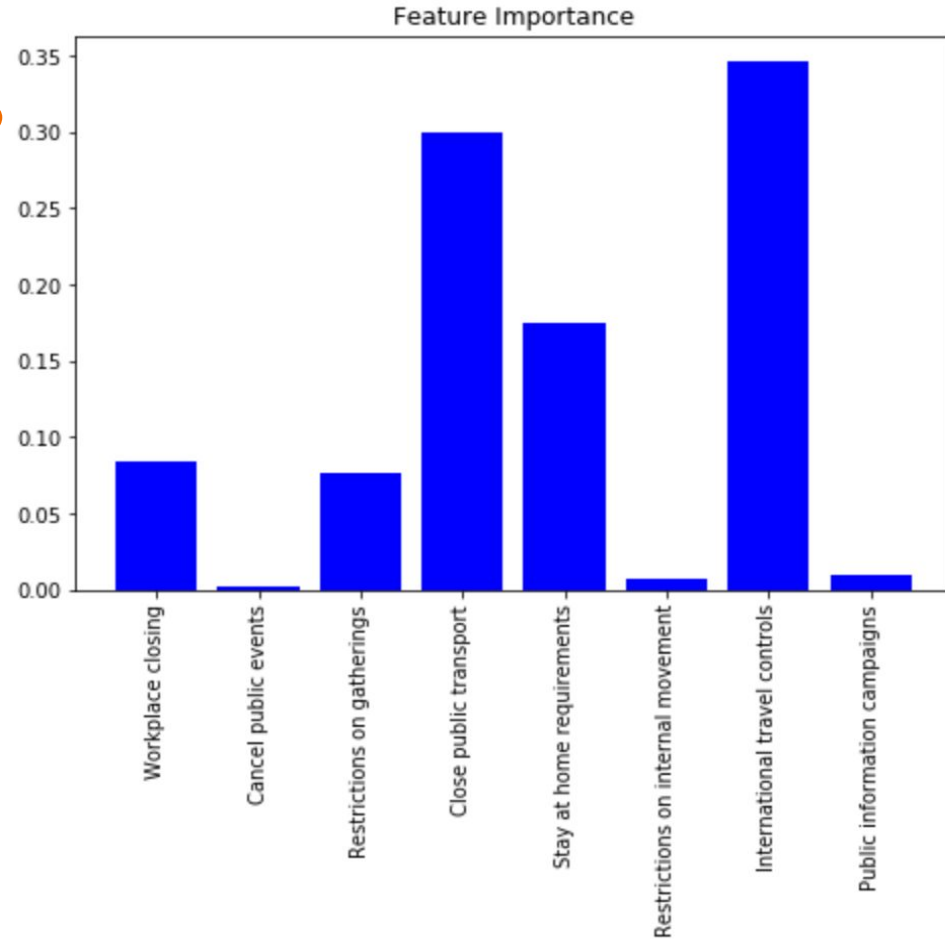
# Qualitative Analysis (cond.)



# Quantitative Analysis

The most effective restrictions (features):

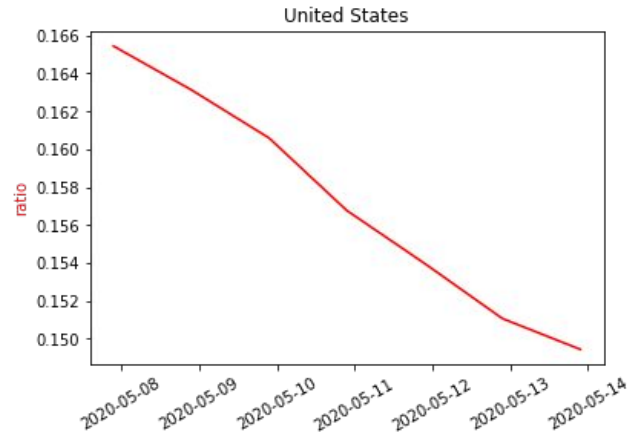
1. International travel controls
2. Close public transport.
3. Stay at home requirements.
4. Workplace closing.
5. Restrictions on gatherings.
6. Public Information Campaigns.
7. ...



# Quantitative Analysis

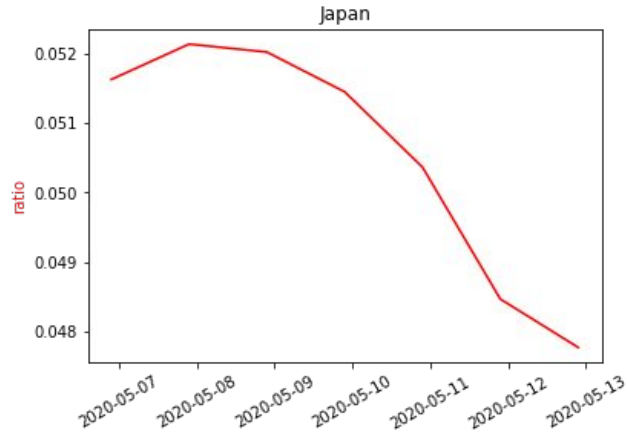
- Time series analysis (single variable - **ratio**)
- Autoregression-moving-average (ARMA) Model

$$r_t = c + \epsilon_t + \sum_{i=1}^p \phi_i r_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

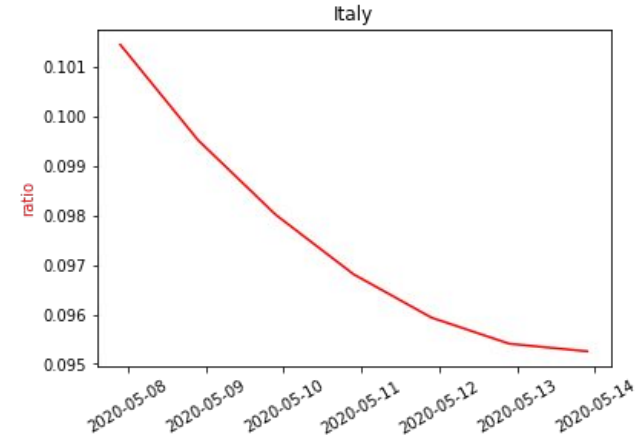


2020-05-08	2020-05-09	2020-05-10	2020-05-11	2020-05-12	2020-05-13	2020-05-14
0.165445	0.163134	0.160590	0.156757	0.153961	0.151051	0.149425

# Quantitative Analysis (cond.)



05-08	05-09	05-10	05-11	05-12	05-13	05-14
0.101435	0.099511	0.098005	0.096807	0.095935	0.095408	0.095256

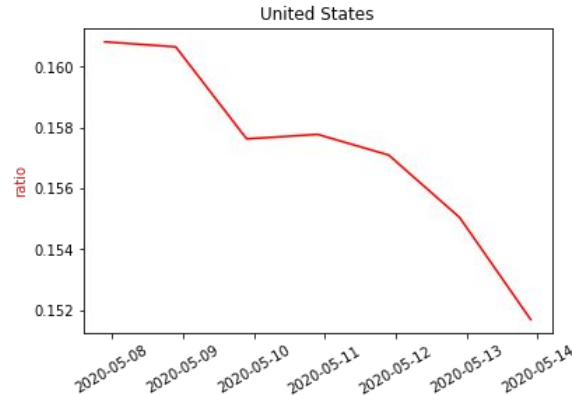


05-07	05-08	05-09	05-10	05-11	05-12	05-13
0.051623	0.052132	0.052018	0.051446	0.050363	0.048468	0.047773

# Quantitative Analysis

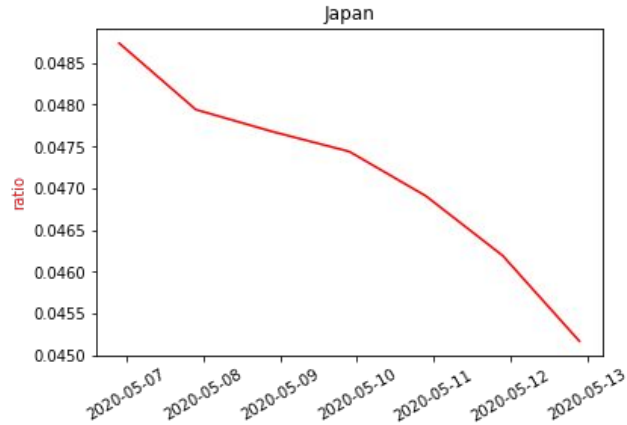
- Time series analysis (multiple variable - **ratio** and **stringency**)
- Vector Autoregression (VAR) Model

$$\begin{bmatrix} r_t \\ s_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} r_{t-1} \\ s_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

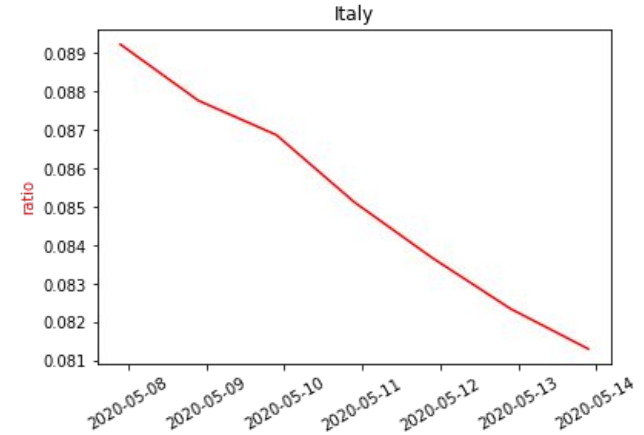


2020-05-08	2020-05-09	2020-05-10	2020-05-11	2020-05-12	2020-05-13	2020-05-14
0.160823	0.160658	0.157629	0.157777	0.157099	0.155044	0.151690

# Quantitative Analysis (cond.)



05-08	05-09	05-10	05-11	05-12	05-13	05-14
0.089224	0.087762	0.086870	0.085123	0.083675	0.082357	0.081309



05-07	05-08	05-09	05-10	05-11	05-12	05-13
0.048736	0.047940	0.047677	0.047439	0.046906	0.046193	0.045170

# Limitations

- Limitations
  - Data is low-volume and biased.
    - Lower granularity data is needed.
    - Missing data due to incomplete government responses.
  - Need to combine data from other perspectives.
    - Different countries have different situations, e.g., socioeconomic, political and geographic environments, etc.
    - Some factors are still hard to be quantized.
  - Linear correlation might be inaccurate.
    - But the dataset is too small to train a non-linear regression model, e.g. neural network.

# Summary

- Governments should take actions as soon as possible and keep the current restrictions for a period of time.
  - In general, international travel controls is relatively effective. Then, it comes stay at home requirements, restrictions on gatherings, workplace closing and close public transport, ..., and so on.
  - Every time a government raises the restriction level of one measurement and lasts for a period of time, we can observe a deduction of confirmed ratio.



# Summary (Cond.)

What we have learned:

- Different methods analyzing data
  - Visualization
  - Machine learning
  - Time series analysis
- Further research is needed
  - More complete and descriptive dataset
  - More reasonable assumption when applying different methods
  - More factors to consideration, e.g., whether people obey governments' restrictions
  - ...

**Thank You!**  
**Questions?**