



Global COVID-19 Visualization

PSDS Capstone Project
Kathleen Smith
14 Jan 2021



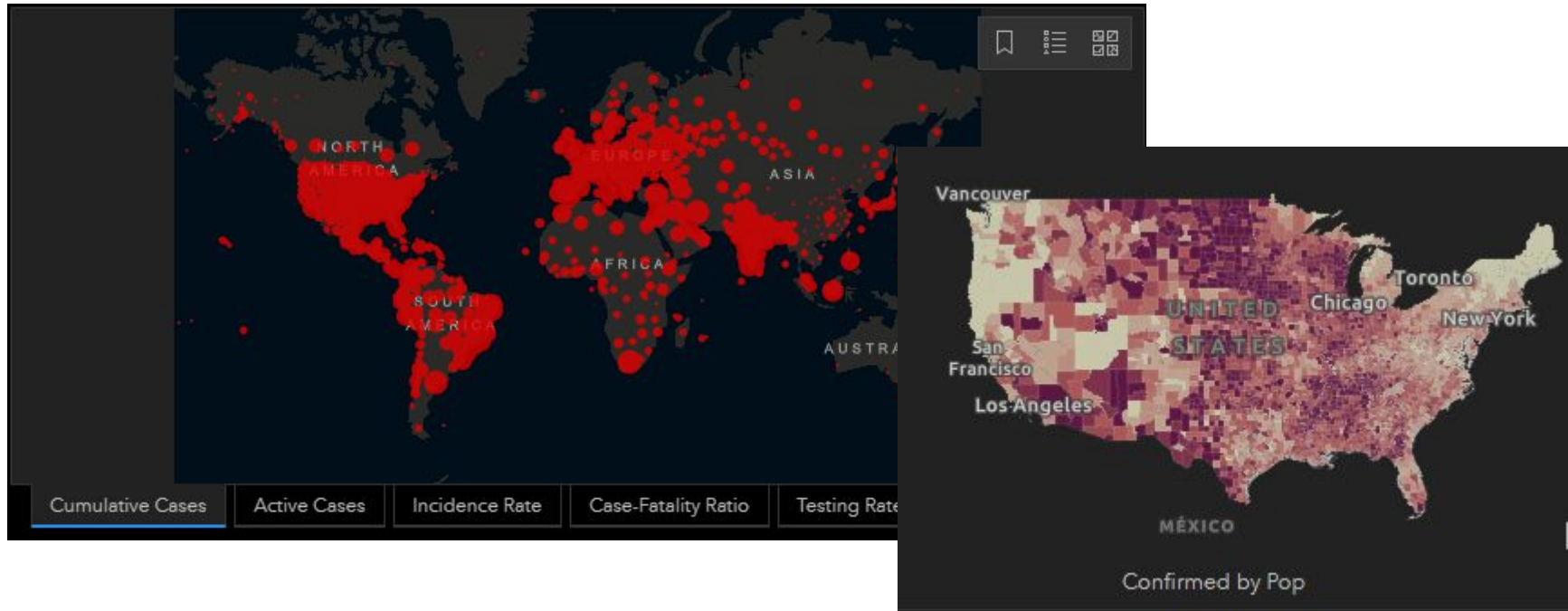
Overview

- Gallery of existing visualisations
- Data sources
- R shiny app deployment
- Further analysis



Existing Visualizations

- Johns Hopkins Coronavirus Resource Center
- Our World in Data: Coronavirus Pandemic Data Explorer
- 3D Map of COVID Cases from r/dataisbeautiful



COVID-19 Map - Johns Hopkins Coronavirus Resource Center

<https://coronavirus.jhu.edu/map.html>

Data Explorer

Download the complete Our World in Data COVID-19 dataset.

Q Type to add a country...

Sort by **Total cases**

<input checked="" type="checkbox"/>	United States	22.61M
<input checked="" type="checkbox"/>	India	10.48M
<input checked="" type="checkbox"/>	United Kingdom	3.13M
<input checked="" type="checkbox"/>	France	2.84M
<input checked="" type="checkbox"/>	Germany	1.94M
<input checked="" type="checkbox"/>	Canada	672,931
<input type="checkbox"/>	World	90.89M
<input type="checkbox"/>	Europe	26.31M
<input type="checkbox"/>	North America	25.87M
<input type="checkbox"/>	Asia	20.76M
<input type="checkbox"/>	European Union	16.14M
<input type="checkbox"/>	South America	14.02M

☒ Confirmed cases

☐ Tests

☐ Confirmed deaths

☐ Share of positive tests

☐ Case fatality rate

☐ Vaccinations

7-day rolling average

☒ Per million people

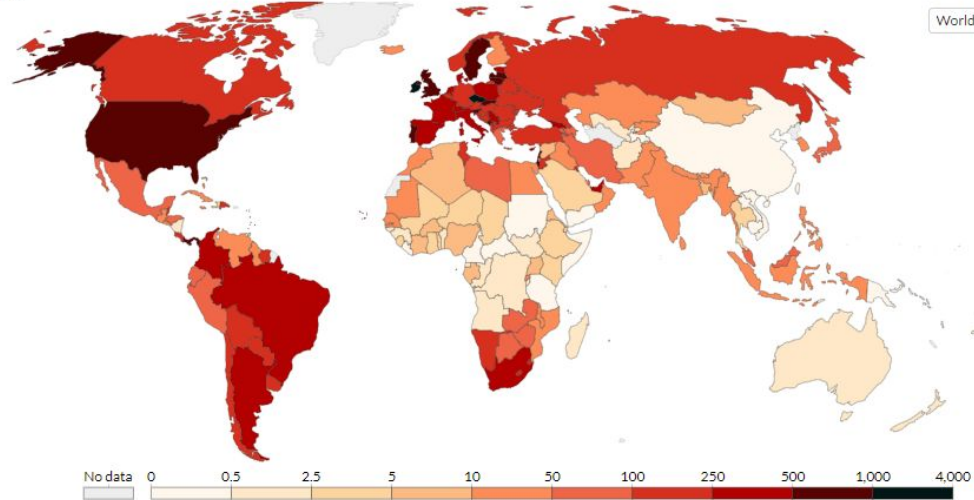
☐ Align outbreaks

Days since the total confirmed cases per million people reached 1

Daily new confirmed COVID-19 cases per million people, Jan 11, 2021

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World in Data



Source: Johns Hopkins University CSSE COVID-19 Data - Last updated 12 January, 06:02 (London time)

CC BY

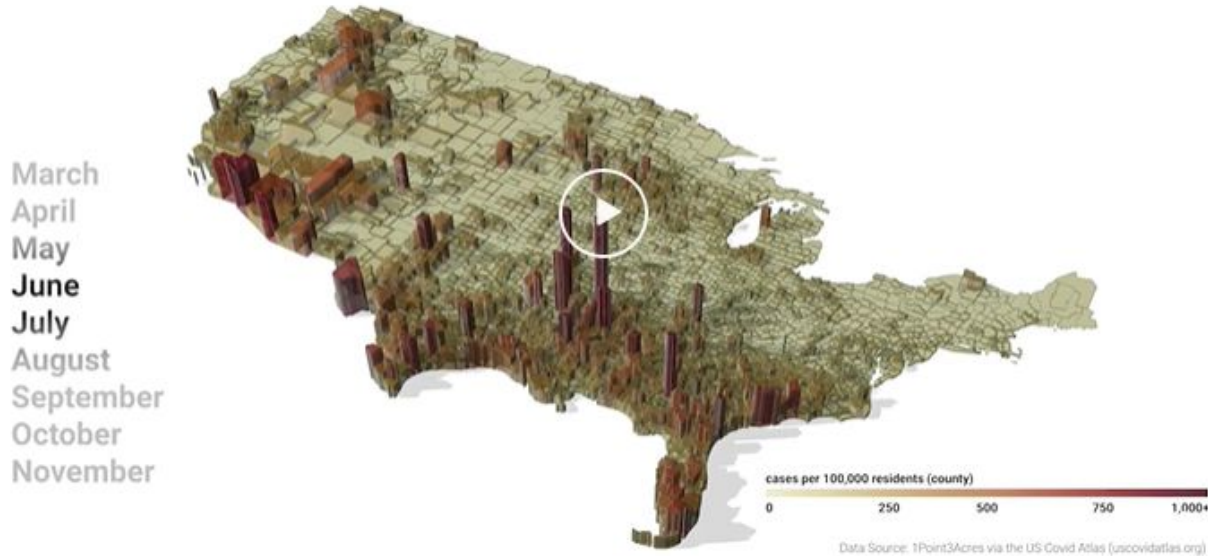
Jan 23, 2020

Jan 11, 2021

Coronavirus Pandemic Data Explorer

<https://ourworldindata.org/coronavirus-data-explorer>

COVID Cases by Population, Week of 2020-07-05



[3D Map of COVID Cases by Population, March through Today \[OC\] : dataisbeautiful](https://www.reddit.com/r/dataisbeautiful/comments/jrkoze/3d_map_of_covid_cases_by_population_march_through_today/)

https://www.reddit.com/r/dataisbeautiful/comments/jrkoze/3d_map_of_covid_cases_by_population_march_through/



Data Sources

- Our World in Data
 - Freely available and easy to download as a csv, enabling live updates
 - Rich data set, collated from various sources
 - For cases and deaths, includes totals, new counts, smoothed counts, and counts per capita
 - Includes testing and vaccination data (limited availability)
 - Includes data on health systems and government response
- CIA World Factbook
 - Added supplementary data on each country from the CIA World Factbook (parsed into JSON format)



R Shiny Web App

https://kes256.shinyapps.io/capstone_basic/

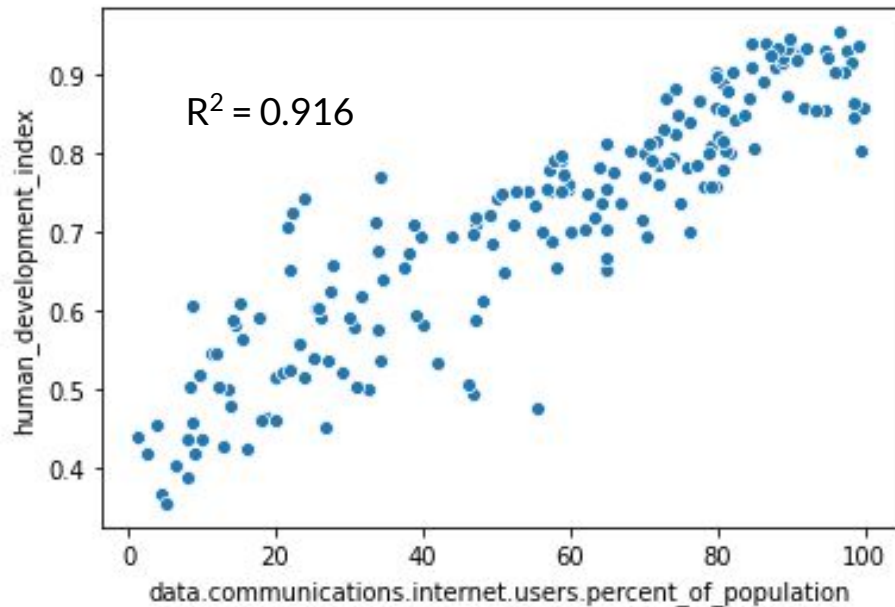


Further Analysis

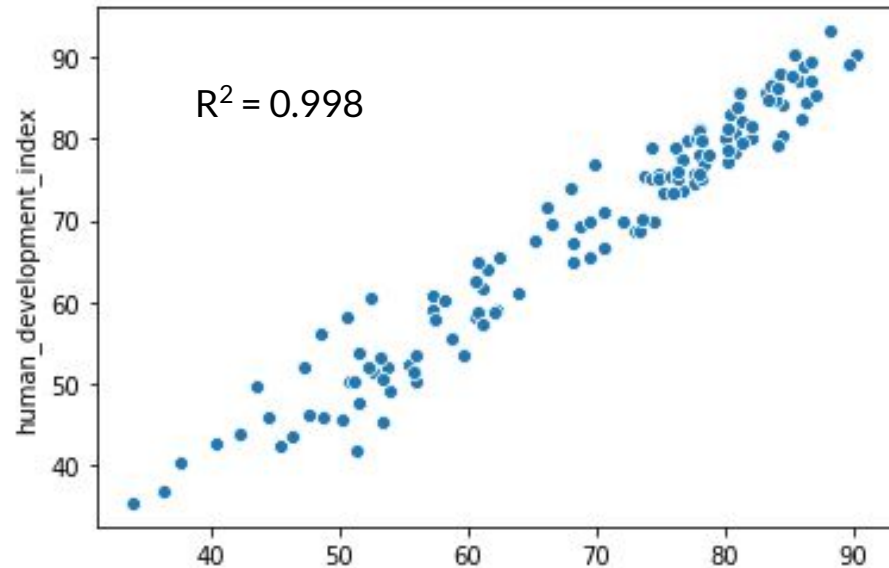
CIA Factbook Analysis

- Initial loading and cleaning of data
- Comparison to Human Development Index
 - Identified strongest correlations
 - Ran multiple linear regression

CIA Factbook Feature	r^2 vs Human Development Index
internet.users.percent_of_population	0.829
school_life_expectancy	0.803
life_expectancy_at_birth	0.794
age_structure.0_to_14.percent	0.775
infant_mortality_rate	0.774
median_age	0.768
electricity.access.total_electrification	0.716
Literacy.total_population.value	0.709
sanitation_facility_access.improved	0.703



Internet Use vs. Human Development Index
Pretty good predictor!



Human Development Index prediction
based on:

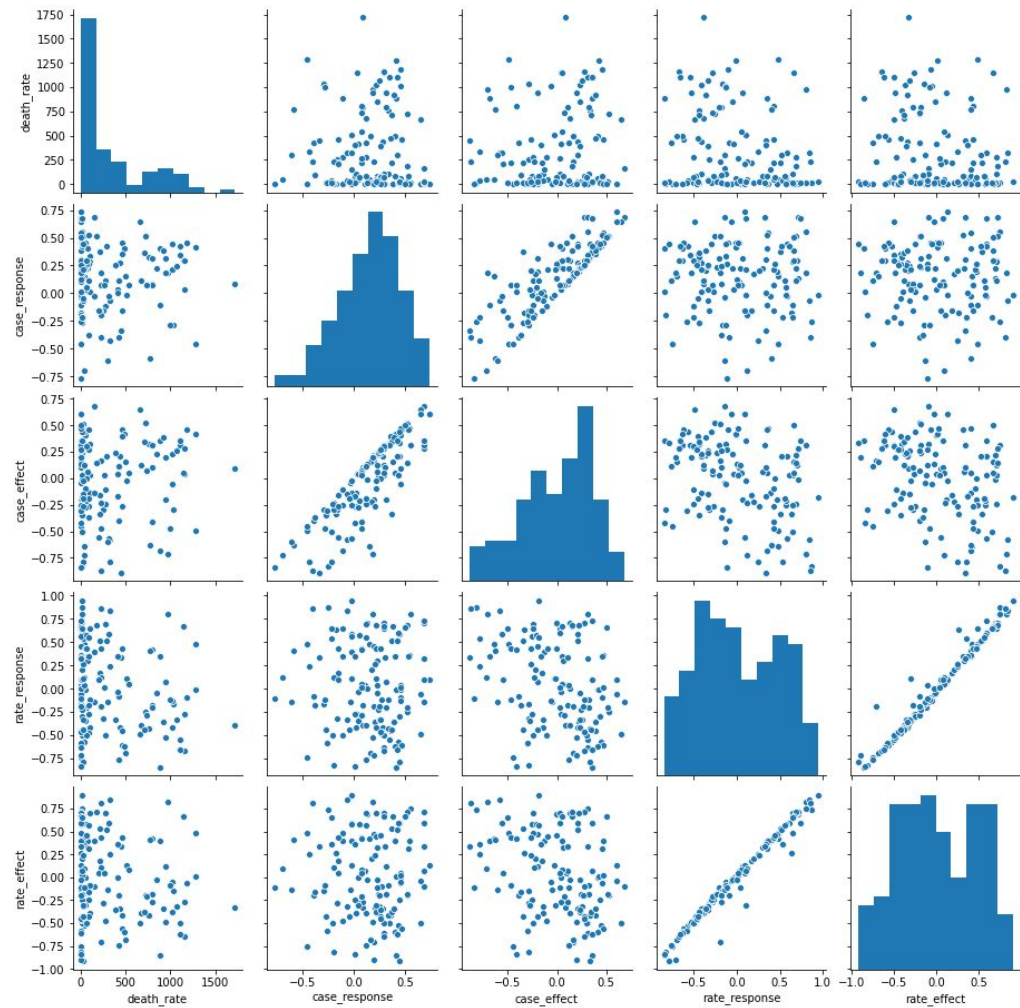
- Internet Use
- School Life Expectancy
- Life Expectancy at Birth
- Literacy



Further Analysis

Time Series Analysis

- Lagged correlations between Government Policy Stringency Index and both New Cases per Million, and Reproduction Rate
- Policy effectiveness: looked for negative correlation between stringency index and future case data
- Policy responsiveness: looked for positive correlation between stringency index and past case data
- No compelling patterns emerged





Further Analysis

Detailed policy data from

[Coronavirus Government Response Tracker | Blavatnik School of Government](#)

Includes stringency broken into more specific features (Mask policy, Economic responses...)

Includes state level data for USA, Canada, and Brazil

Includes risk of openness index



Further Analysis

Detailed policy data from

[Coronavirus Government Response Tracker | Blavatnik School of Government](#)

Challenges:

Haven't merged with OWID dataset, limiting options to measure effective management of pandemic

Haven't accounted for dependencies on date



Questions?



Links and Resources

Code repository:

https://github.com/kes256/PSDS_Capstone

Shiny app:

https://kes256.shinyapps.io/capstone_basic

Data sources:

https://github.com/iancoleman/cia_world_factbook_api

<https://covid.ourworldindata.org/data/owid-covid-data.csv>



Python and R packages attempted

- Python geopandas - poor support for dependencies in python 3
- R ggplot - lacks animation and 3d options
- R plotly - limited options for 3d animations
- Python plotly - animated 3d scatterplot with error bars to connect points to x-y plane
- R shiny - streamlined options to set up a web app; just needed to embed python plot



Python and R packages used

Python

- pandas
- plotly/plotly_express
- seaborn
- statsmodels

R

- tidyverse - dplyr, lubridate
- countrycode
- shiny
- reticulate
- viridis