Capstone: Predictive modeling for Epidemics

By: Khadija Conteh

Capstone Motivation

- Covid-19
- Identify the most vulnerable countries at risk to an epidemic
    - Prioritizing international and national funding
    - Setup/strengthen country emergency preparedness and response plans
    - Set up/strengthen international and national coordination of resources
- Personal/Professional experience

Ebola

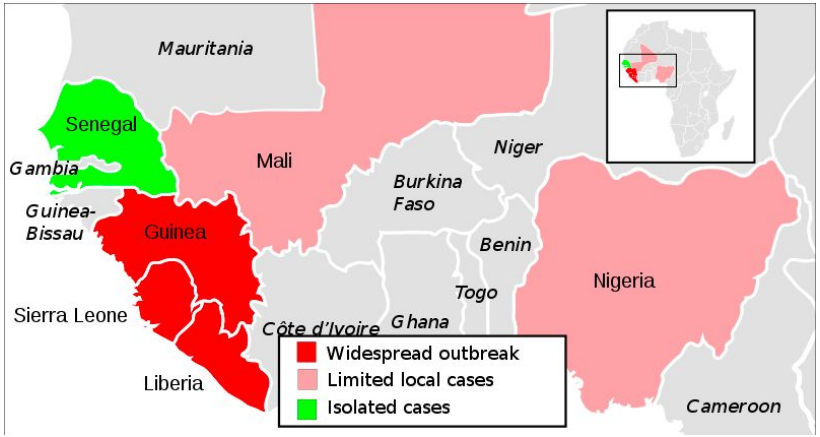
Emergence: Guinea

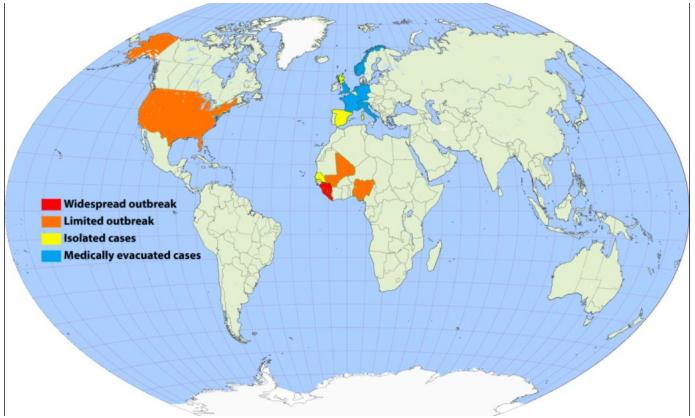
Date: December 2013 - June 2016

Declared Epidemic: 8 August 2014

Cases: 28,646

Deaths: 11,323





Lessons Learned

Emergency Preparedness Planning (i.e. institutional coordination and financing, risk assessments, surveillance)

Health infrastructure (i.e. health workers per capita; #of hospitals; availability of vital medical equipment)

Key non-health factors that severely impacted the crisis:

- Demographics (Population Age, Density, Urban/Rural)
- Economic stability (i.e. GDP, GNI, Employment Sectors)
- Political stability (Governance, Corruptness)
- Compounded crises (natural disaster, conflict)

Ways Forward - Using Data Science

Using regression based modeling, I will explore the correlation between varied health, social and economic indicators and INFORM's country risk score to predict a country's risk to an epidemic.

Data Collection

World Bank

World Health Organization

Index for Risk Management (INFORM)

- Game Changer!
- open source risk assessment s to support decisions about prevention, preparedness and response to humanitarian crises and disasters.

Total indicators collected: 130

INFORM Risk Score and Class

Risk	INFORM EPIDEMIC P2P RISK INDEX				
Dimension	RISK FORMULA				
	Hazard & Exposure		Vulnerability		Lack of Coping Capacity
Category	P2P		GEOMETRIC AVERAGE		GEOMETRIC AVERAGE
	ARITHMETIC AVERAGE		INFORM Vulnerability	Epidemic Vulnerability	INFORM Lack of Coping Capacity	Epidemic Lack of Coping Capacity
	WaSH	Population				

Very High	High	Medium	Low	Very Low
>= 6.5	>= 5	>= 3.5	>= 3.5	< 2

Data Challenges

Completeness

- Nulls

Recentness

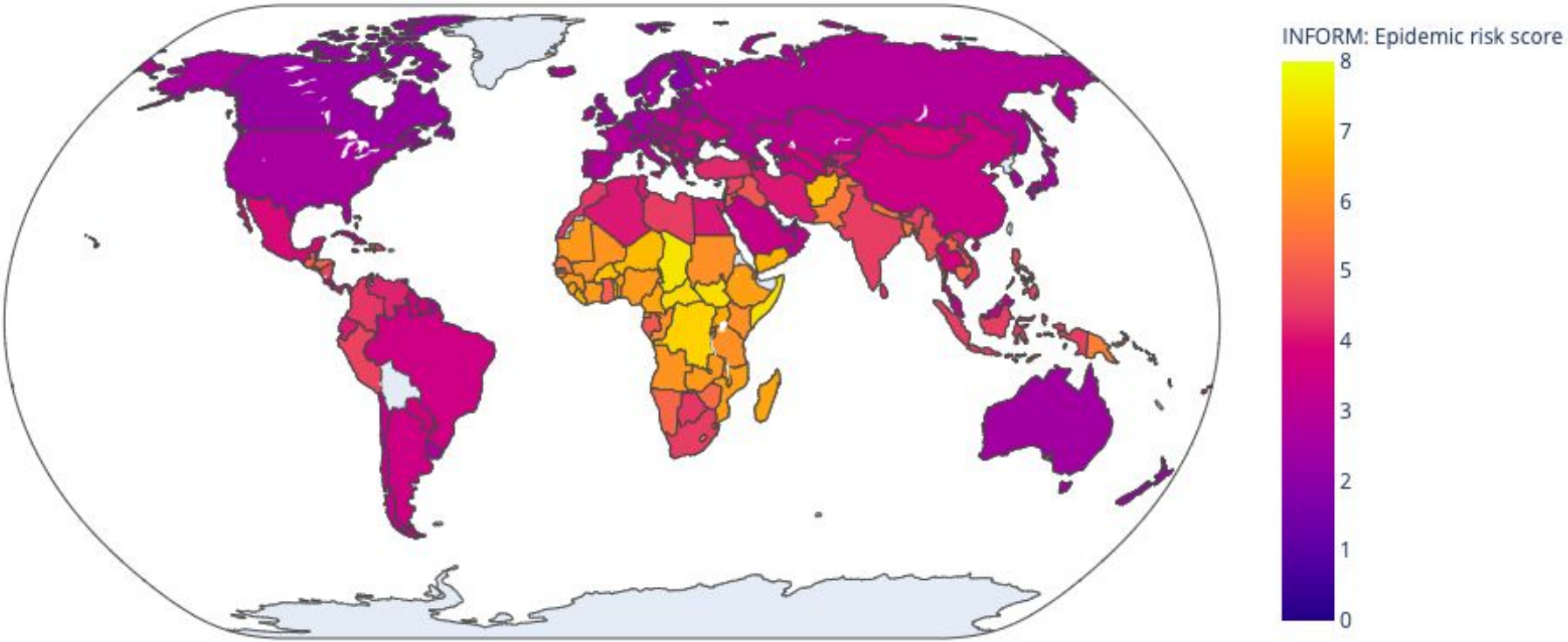
- *How old is too old?*

Standardization

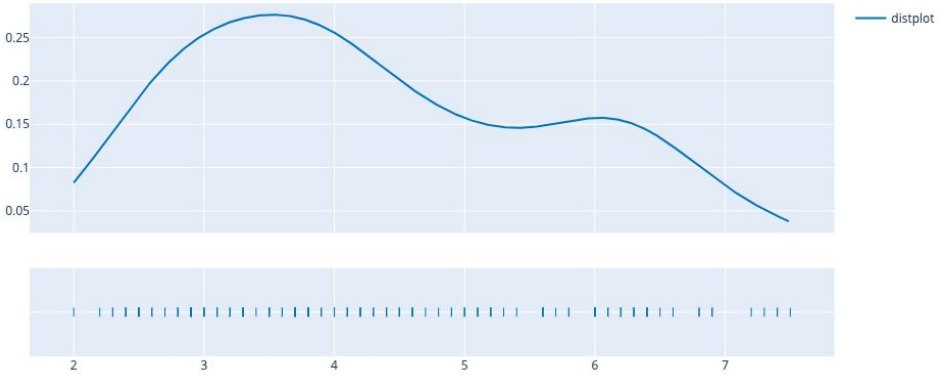
- Naming convention
- Countries/territories used for data collection)

Exploratory Data Analysis

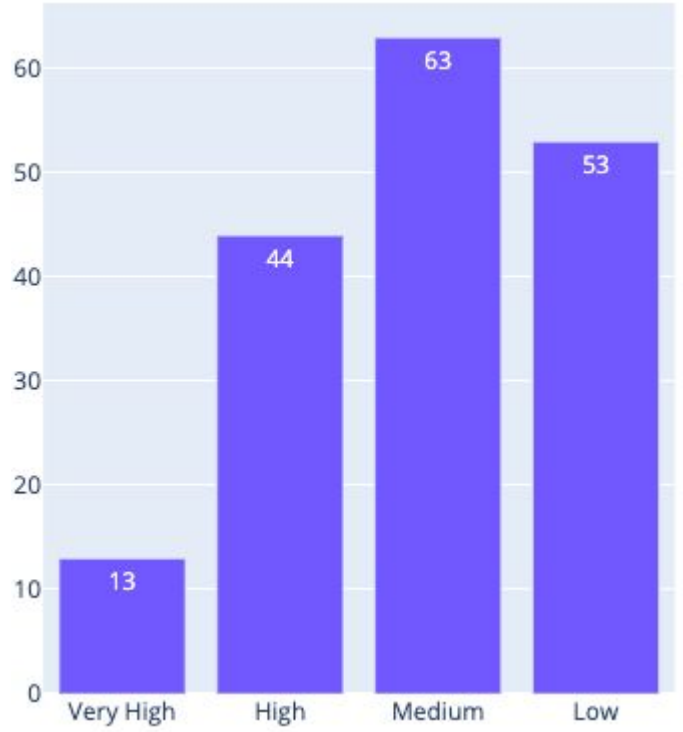
Global Risk of Epidemic



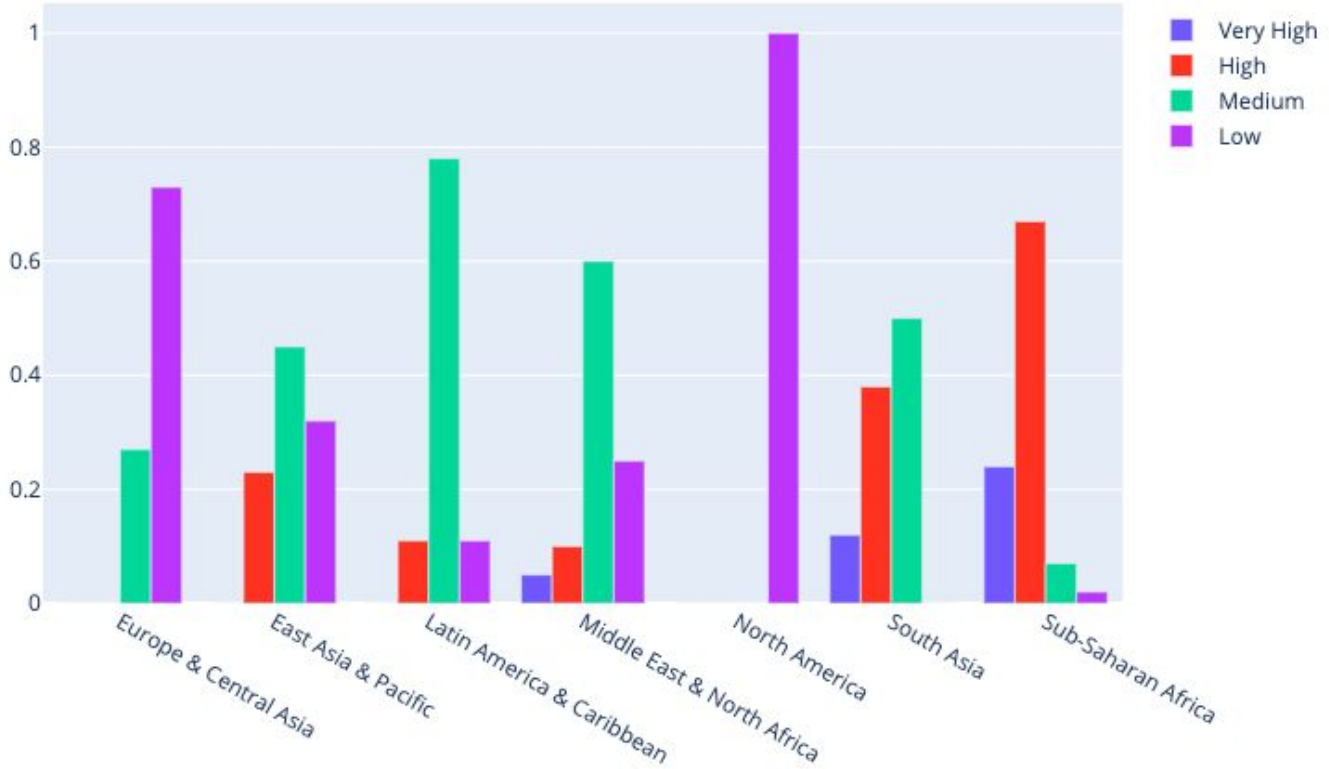
Inform Epidemic Risk Score and Class



Very High	High	Medium	Low	Very Low
>= 6.5	>= 5	>= 3.5	>= 3.5	< 2

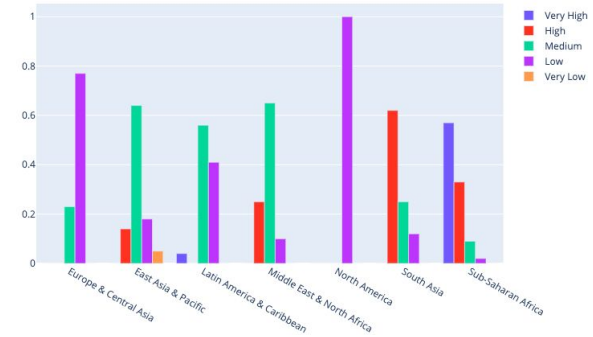


Risk of Epidemic: by Region

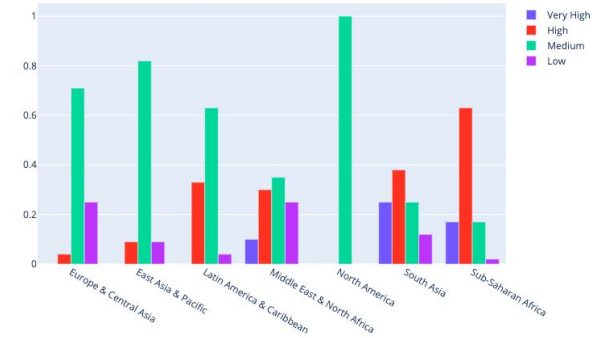


Epidemic Risk Sub-Categories: By Region

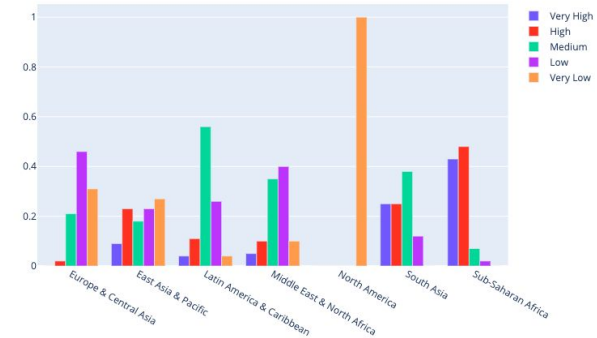
Hazard & Exposure Risk: By Region



Vulnerability Risk: by Region



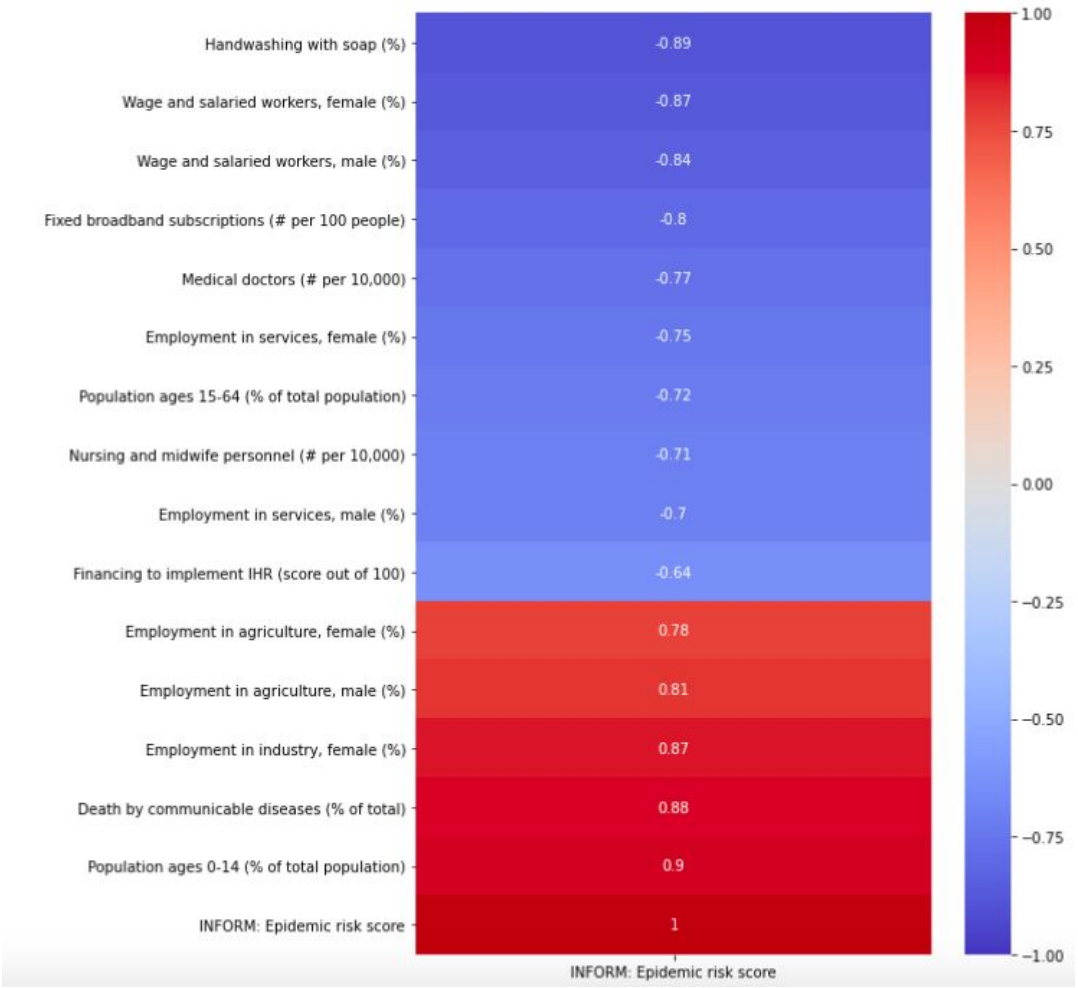
Lack of Coping Capacity Risk: By Region



INFORM Epidemic Risk: Top & Least Countries

#	Country Name	Risk Score
1	Chad	7.5
2	Somalia	7.4
3	South Sudan	7.4
4	Central African Republic	7.3
5	Democratic Republic of the Congo	7.2
6	Burkina Faso	6.9
7	Burundi	6.9
8	Afghanistan	6.8
9	Niger	6.8
10	Libera	6.6
11	Togo	6.6
12	Yemen	6.6
13	Uganda	6.5

#	Country Name	Risk Score
173	Singapore	2.2
172	Norway	2.3
171	United Kingdom	2.4
170	Netherlands	2.4
169	New Zealand	2.4
168	Switzerland	2.4
167	United Arab Emirates	2.5
166	Slovenia	2.5
165	Sweden	2.6
164	United States	2.6
163	Republic of Korea	2.7
162	Samoa	2.8
161	Uruguay	2.9
160	Qatar	2.9



Feature Selection

Vulnerability	9
Hazard & Exposure	3
Lack of Coping Capacity	3
Total	15

Health Specific	4

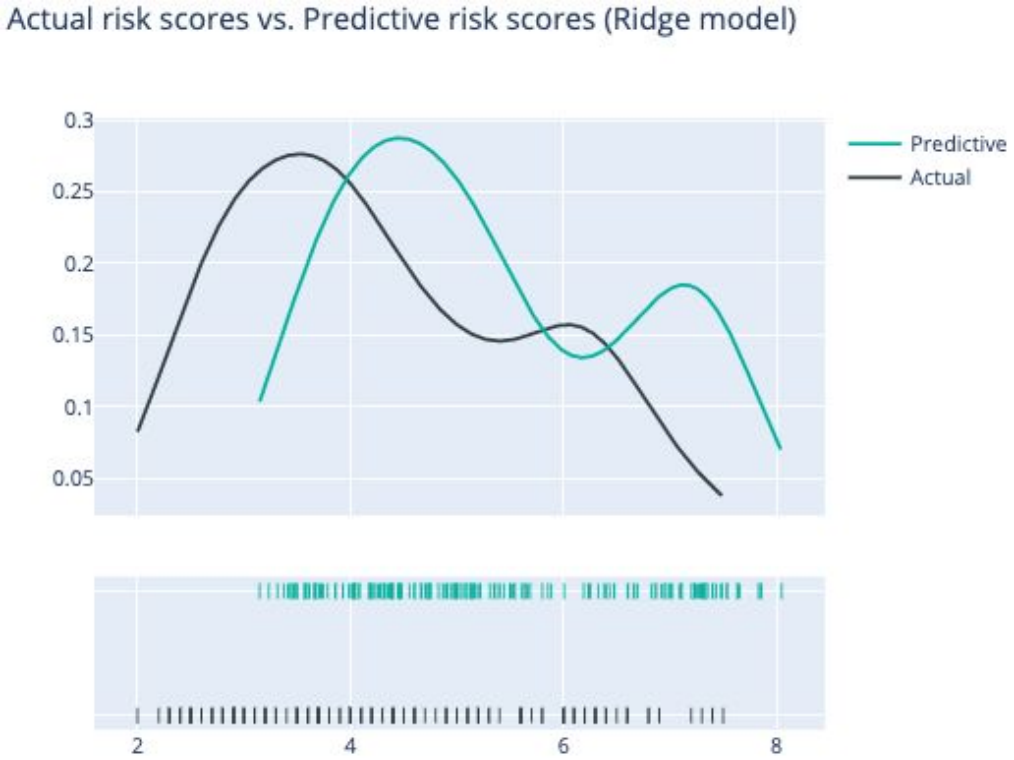
Predictive Models

- Linear Regression
- Lasso
- Ridge (2)
- Random Forest (2)

Predictive Model Results

Model	R2 Score	Train Score	Test Score	Mean Squared Error	Mean Absolute Error	Cross Validation Score
Linear Regression	0.918748	0.917695	0.91726	0.158544	0.309946	0.869057
Lasso	0.90855	0.902004	0.929252	0.178442	0.321385	0.864926
Ridge	0.918754	0.917695	0.917291	0.158532	0.309923	0.861916
RidgeCV	0.91888	0.917685	0.917994	0.158286	0.309414	0.869313
Random Forest: Model #1	0.967889	0.983459	00.900575	0.062658	0.176532	0.866756
Random Forest: Model #2	0.926994	0.930783	0.905891	0.142455	0.282497	0.871353

Actual vs. Pedictive risk scores



Recommendations

- Explore predictive modeling with classification method, using INFORM's epidemic risk class of countries
- Build an interactive web app for interested users to further explore the data!