HSR.health

IIT Bombay Winter School

Public Health Data Models

Session 1

Schedule

Time	Session	Speakers
0930-1000	Recap	Sumit Sen
1000-1100	HealthCare Background and Ebola Usecase	Paul Churchyard, Ajay Gupta
1100-1145	Activity 1: Create a Data Model for and Ebola Outbi	Paul Churchyard, Ajay Gupta
1145-1200	Tea	
1200-1300	First Discussion Session	Paul Churchyard, Ajay Gupta, Sumit Sen
1300-1400	Lunch	
1400-1500	Exploring a variety of panedmic related data model	Paul Churchyard, Ajay Gupta
1500-1545	Activity 2: Create a Data Model for the Health Impa	Paul Churchyard, Ajay Gupta
1545-1600	Tea	
1600-1700	Second Disucssion Session	Paul Churchyard, Ajay Gupta, Sumit Sen

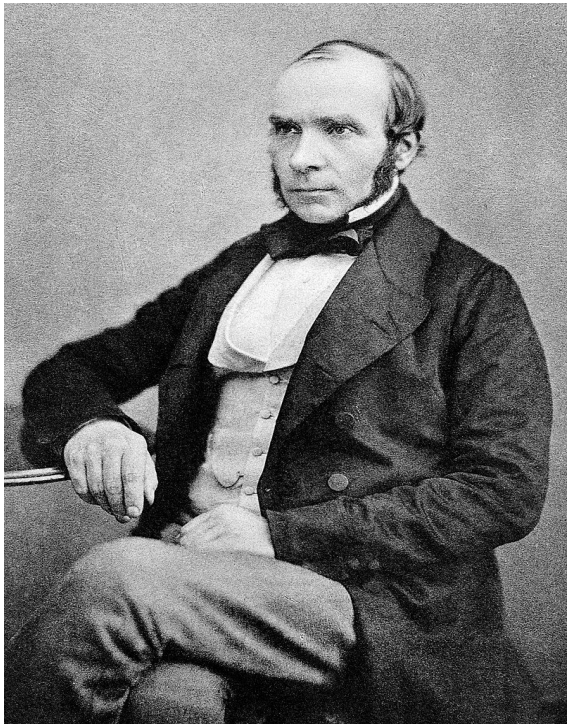
What is Public Health?

- Public Health is:
  - *"...the science of protecting and improving the health of people and their communities"*. CDC Foundation
  - *"...the art and science of preventing disease, prolonging life and promoting health…"*, World Health Organization
  - *"...identifies the causes of disease and disability and implement large-scale solutions…"*, Johns Hopkins School of Public Health

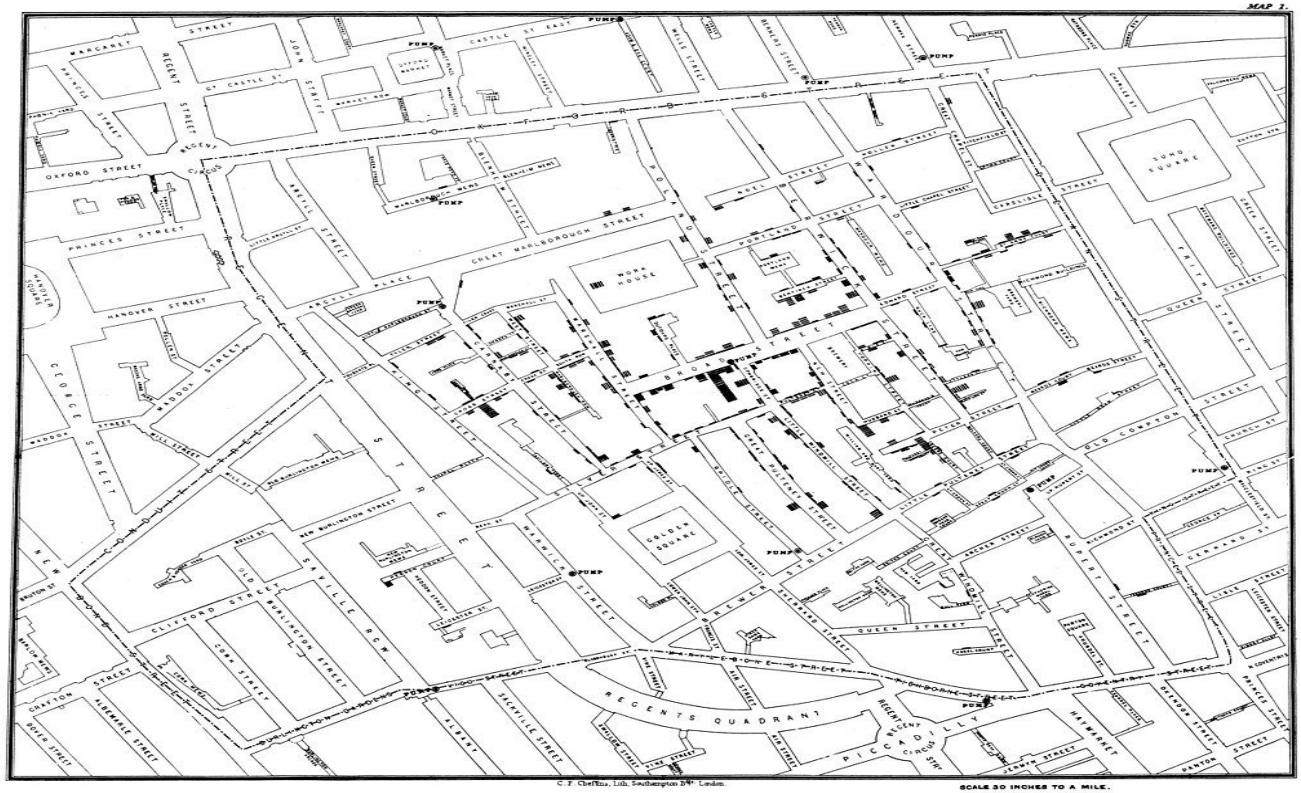
John Snow and the Origins of Epidemiology

- John Snow (1813-1858)
  - English physician. Father of Epidemiology
  - Stopped the cholera outbreak in Soho, London in 1864
  - Traced the infection to an infected water source

- Pattern recognition in the data
  - Recognized a pattern of incidence from a map of cases (the “Ghost Map”)
  - Showed the importance of data-driven intervention
  - Early form of GIS (Geographic Information Systems)



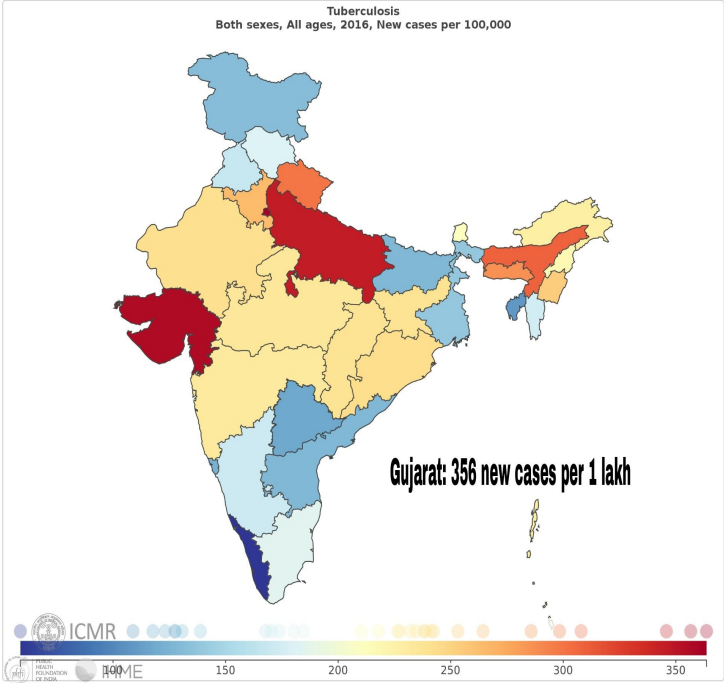
John Snow and the Ghost Map



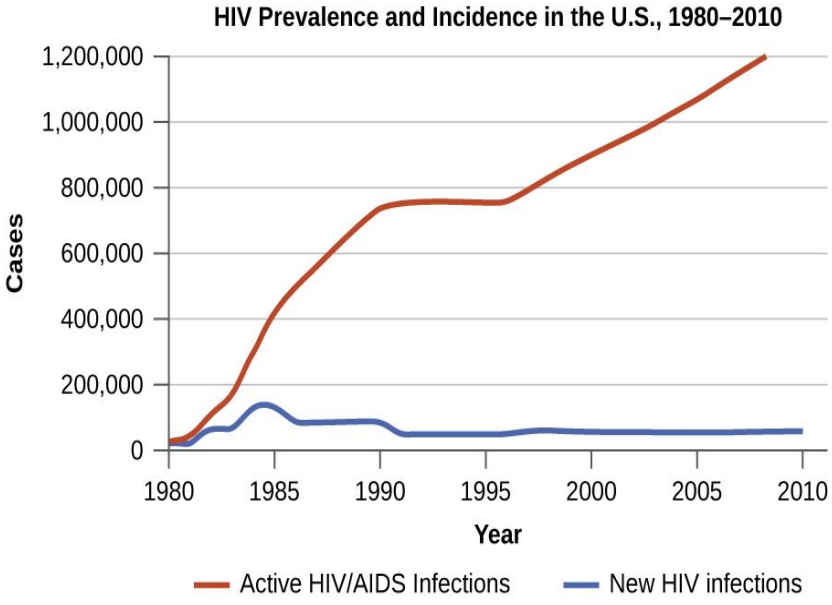
What is Epidemiology

- Epidemiology
  - “Branch of medical science that investigates all the factors that determine the presence or absence of diseases and disorders”. National Institutes of Health.
  - Seek to understand, investigate, control and prevent the spread of disease.
  - Built for GIS

- Subdisciplines of Epidemiology:
  - Infectious Disease Epidemiology
  - Chronic Disease Epidemiology
  - Environmental / Occupational Epidemiology



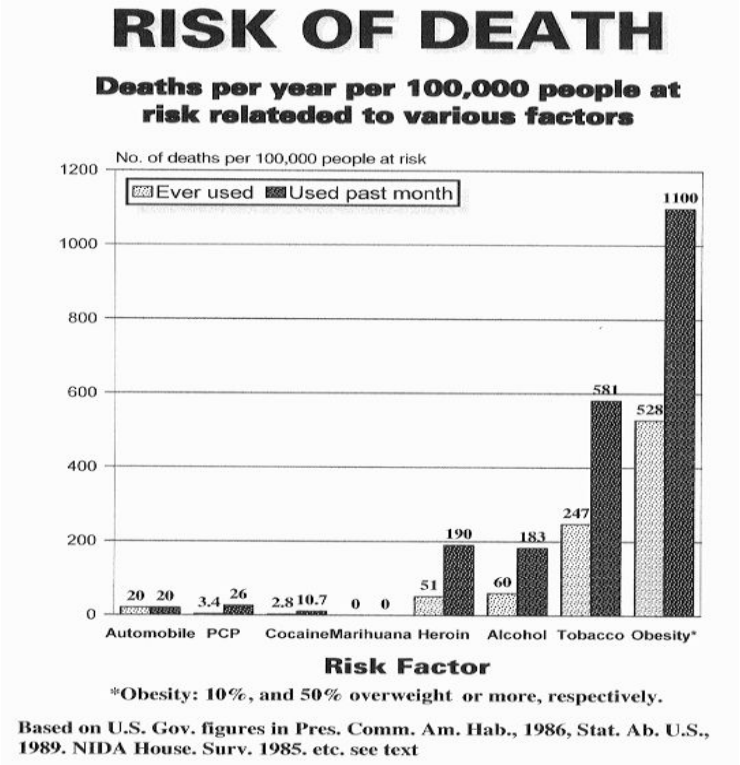
What is Biostatistics



- What is Biostatistics
  - "...the branch of statistics that deals with data relating to living organisms", UW School of Public Health
  - Turning data into knowledge
- Used together with Epidemiology to describe and analyze population health
- Useful terms:
  - Incidence
  - Prevalence
  - R0 / reproductive number

Quantifying and Illustrating Risk

- What do we mean by risk?
  - The probability of an outcome, be it death, infection or disability, in the near future.
- Rates / proportion
  - Morbidity
  - Mortality
- Odds ratios / risk ratios
  - Definition
- Why is this necessary?
  - Quantification of outcomes
  - Better understanding of magnitude of disease
  - Strategic intervention



Sources of Data in Public Health

- The importance of data sources
  - Where's your data coming from?
  - How and from where is that data being obtained?
  - Accuracy / timeliness

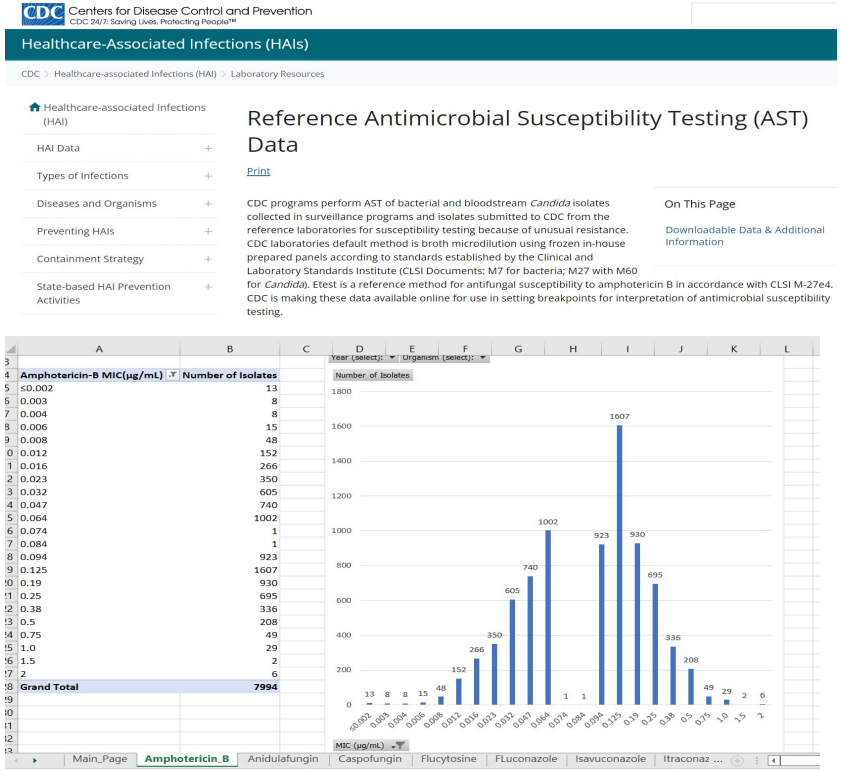
- Some common sources of public health data:
  - The World Health Organization
  - Ministries of Health
  - Local health departments
  - Hospitals
  - NGOs / think tanks

- Common formats:
  - Mostly files and APIs

Source of Data in Public Health (continued…)

- How is the data being collected?
  - Census
  - Research efforts
  - Health authorities / different levels
    - Local
    - State / province
    - Government
    - International

- Surveys
  - Sampling science and methodology



CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

Healthcare-Associated Infections (HAIs)

CDC > Healthcare-associated Infections (HAI) > Laboratory Resources

Healthcare-associated Infections (HAI)
HAI Data
Types of Infections
Diseases and Organisms
Preventing HAIs
Containment Strategy
State-based HAI Prevention Activities

Reference Antimicrobial Susceptibility Testing (AST) Data

Print

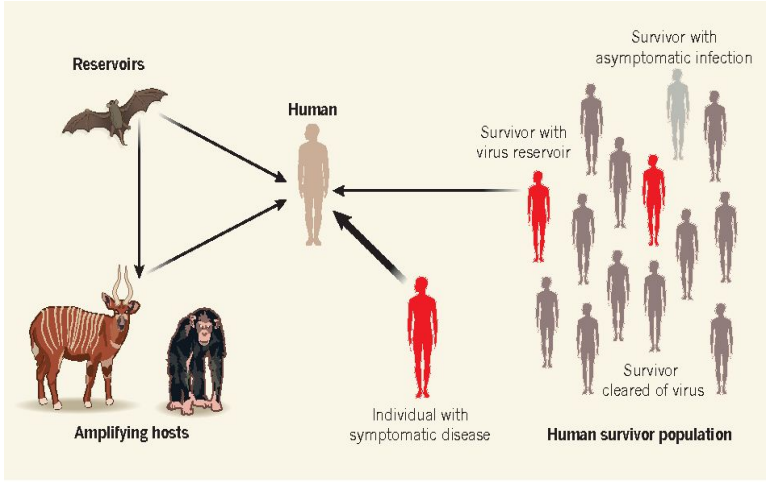
CDC programs perform AST of bacterial and bloodstream *Candida* isolates collected in surveillance programs and isolates submitted to CDC from the reference laboratories for susceptibility testing because of unusual resistance. CDC laboratories default method is broth microdilution using frozen in-house prepared panels according to standards established by the Clinical and Laboratory Standards Institute (CLSI Documents: M7 for bacteria; M27 with M60 for *Candida*). Etest is a reference method for antifungal susceptibility to amphotericin B in accordance with CLSI M-27e4. CDC is making these data available online for use in setting breakpoints for interpretation of antimicrobial susceptibility testing.

On This Page
Downloadable Data & Additional Information

Amphotericin-B MIC(µg/mL)	Number of Isolates
≤0.002	13
0.003	8
0.004	8
0.006	15
0.008	48
0.012	152
0.016	266
0.023	350
0.032	605
0.047	740
0.064	1002
0.074	1
0.084	1
0.094	923
0.125	1607
0.19	930
0.25	695
0.38	336
0.5	208
0.75	49
1.0	29
1.5	2
2	6
Grand Total	7994

Year (select): Organism (select):
Number of Isolates
MIC (µg/mL)
Main_Page | Amphotericin_B | Anidulafungin | Caspofungin | Flucytosine | Fluconazole | Isavuconazole | Itraconaz …

Public Health in the Real World: Zoonotic Disease Spillover

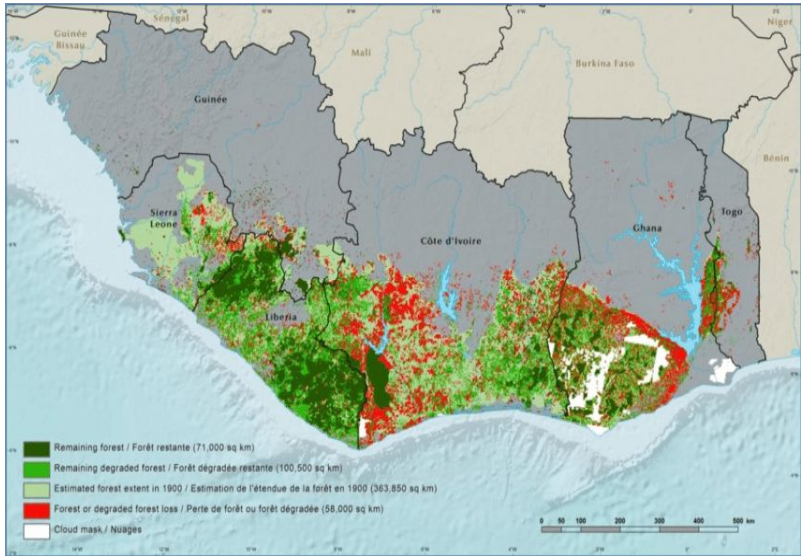


- Zoonotic disease
  - Any disease that spreads from animals to humans
  - Ex: HIV/AIDS, Ebola, Influenza
- Zoonotic spillover
  - A transmission event, where an animal pathogen infects a human host (patient zero)
  - Outsized representation in global pandemics
- Different pathogens
  - Viral
  - Bacterial
  - Fungal

The Mechanisms of Zoonotic Spillovers

- Zoonotic Spillover risk
  - Animal reservoirs
  - Social and economic factors
  - Geographic factors

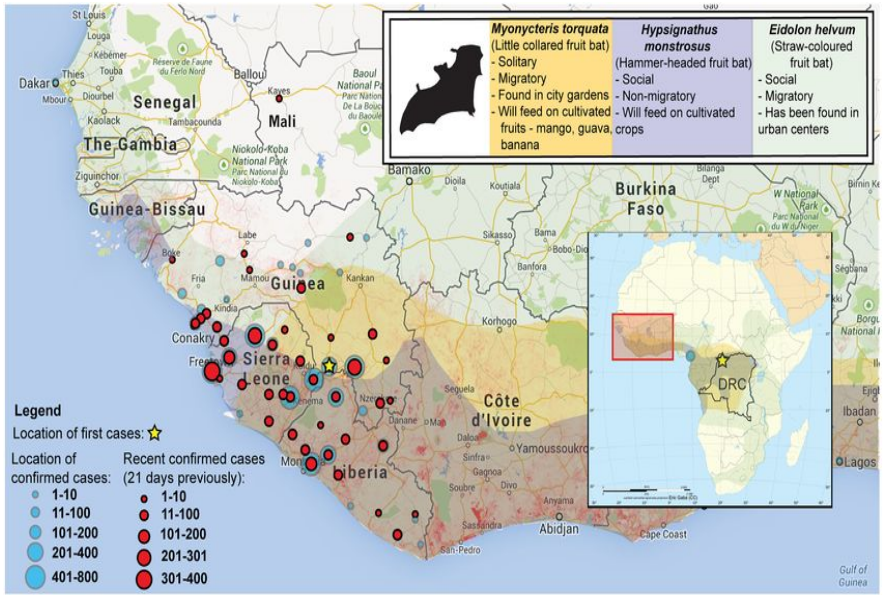
- Geospatial analytics
  - Focus on analyzing the intersection between human effect on geography and geographical effects on humans
  - Positioned as a tool in order to identify high risk areas
  - Quantifying risk to the human population.



Zoonotic Spillover Case Study: Ebola

- Ebola
  - Viral hemorrhagic fever, zoonotic, bats likely animal reservoir
  - History
    - First seen in Africa, 1970s
    - Multiple outbreaks
  - Transmission is through contact with bodily fluids of the infected or recently deceased.
  - High mortality
    - 53-68%
    - In comparison, smallpox has a 30% mortality rate
- Ebola as a zoonotic disease
  - Spillover events in Western and Central Africa
  - Wet markets and bush meat
  - Habitat destruction, fragmentation and new human habitation

Zoonotic Spillover Case Study: Ebola

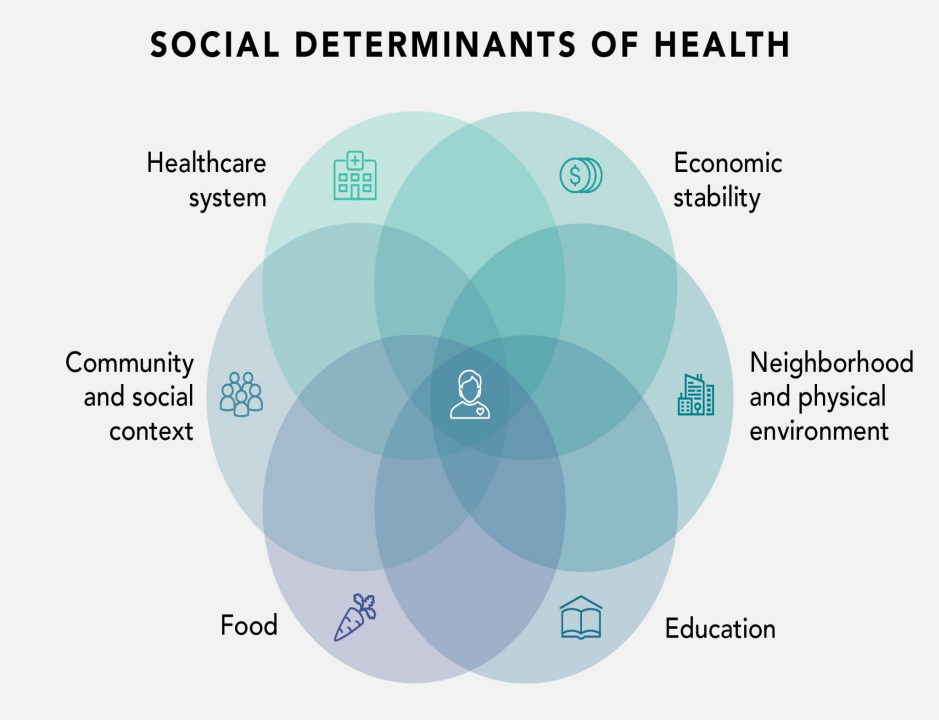


Ebola and Pandemic Diseases: Future Predictive Needs

- The potential effects of an Ebola global pandemic
- The challenge of prevention
  - Global travel and commerce
  - Non-specific prodromal symptoms
- Data modeling and predictive analytics as a solution
  - Patterns of spread
  - High risk areas
  - Intervention points
- Proactive vs reactive disease surveillance and detection

Social Determinants of Health

- SDOH for short, these are non-medical factors that influence health outcomes.
- Present where humans live, work and play
- Common SDOH factors:
  - Income
  - Housing type
  - Behavior (Smoking, Drinking, drug use)
  - Transportation
  - Education
  - Employment
  - Environmental / Air quality



SDOH and Health Outcomes

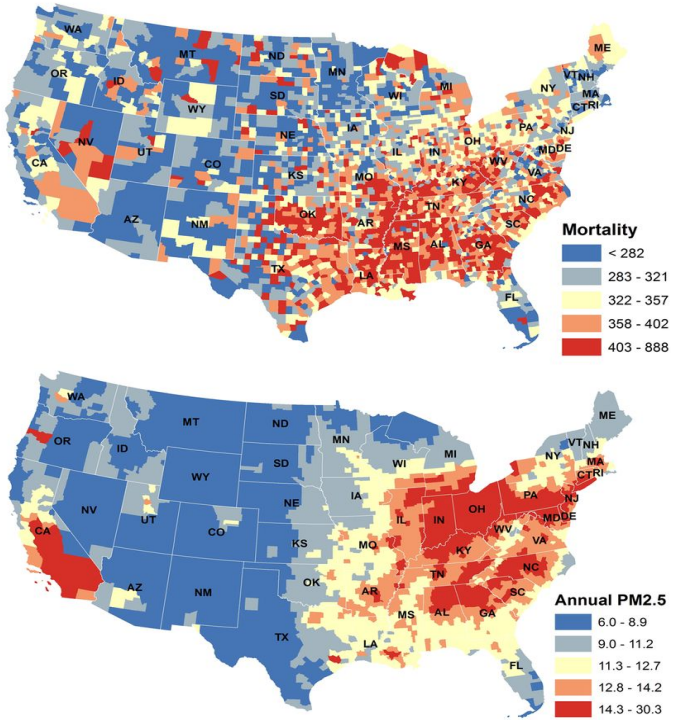
- Your education affects your health:
  - less educated adults reporting worse general health, a greater number of chronic health conditions, and more functional limitations and disabilities.

- Your neighborhood affects your health:
  a. Lower income neighborhoods have positive correlations with low birth weight, childhood injury and abuse, and teenage pregnancy risk.

- Your income affects your health:
  a. in the United States, between 2001 and 2014, difference in life expectancy between the top 1% and the bottom 1% of the economic spectrum was 14.6 years.

SDOH and Geospatial Analytics

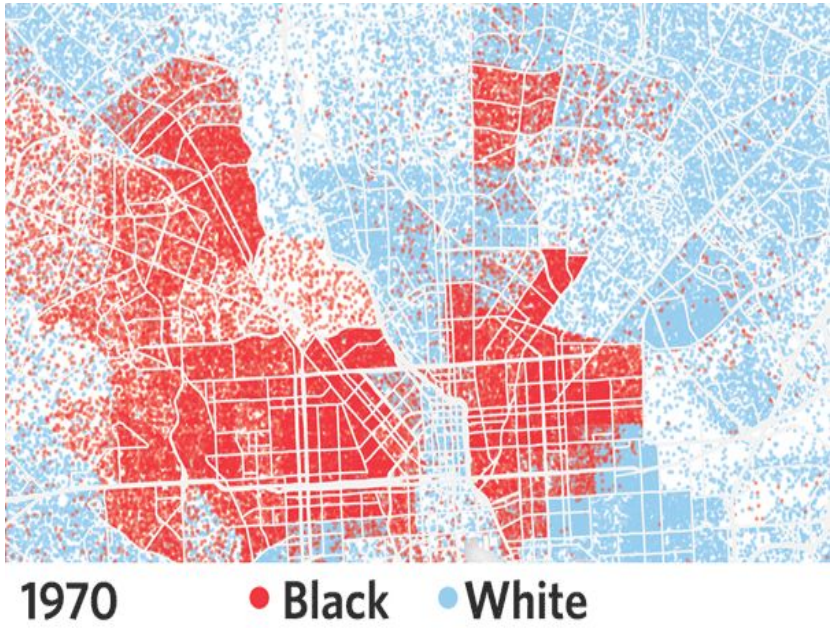
- Geospatial analytics and data models can uncover:
  - Social and economic patterns
  - Health outcomes patterns
  - Vulnerable populations and at risk localities
- Air quality and neighborhoods

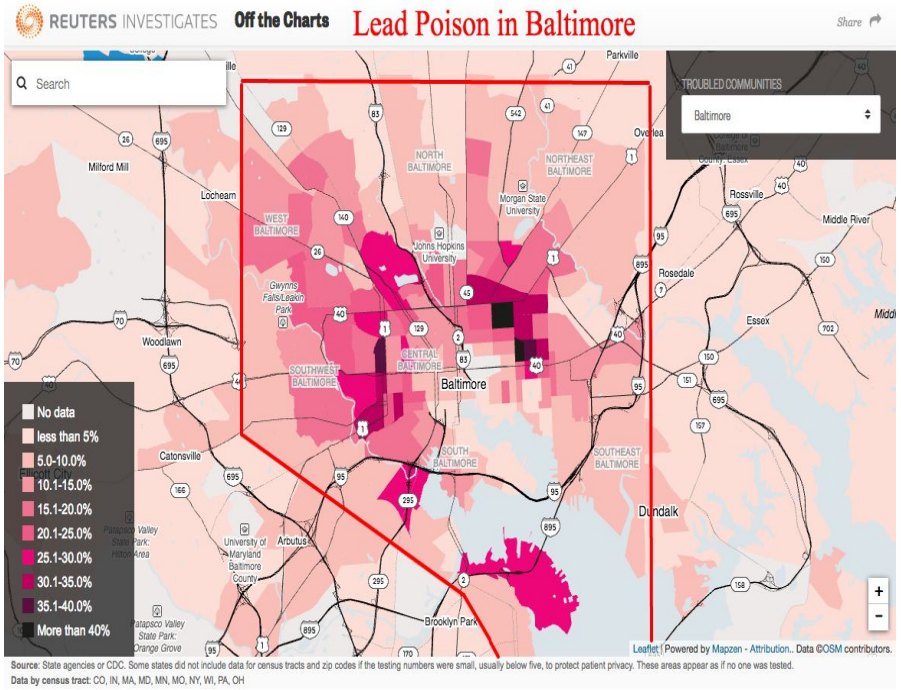


Lead Contamination in the USA: SDOH Case Study

- Lead is known to have multiple effects on health specially in growing brains and intelligence
- Lead was a component of paint for decades
- Research found correlation between lead toxicity and paint exposure
- Laws and regulations prohibited the use of lead in paint manufacturing starting in 1978
- Lead toxicity has decreased
- Still a problem in many localities in the U.S.
    - Flint, Michigan
- The fight against lead shows:
    - Data at work
    - Recognition of patterns
    - Data-driven interventions

Lead Contamination in the USA: SDOH Case Study





Activity 1: Using Data to Prevent an Ebola Pandemic

Activity 1: Uganda Ebola Outbreak, September 2022 | Ebola (Ebola Virus Disease) | CDC

Task: Come up with a transmission risk index that quantifies the risk of transmission in an area.

Source:

- What are the relevant factors?
- Where are they coming from?

Transformation:

- What combination of factors quantifies risk?
- How do the factors relate?
- Weights?

Output:

- What does the output mean?
- Who and how could use the output?
- What is the best way to display the output?

Time to complete: 1 hour
1 hour discussion afterwards

To get you started:

2014-2016 Ebola Outbreak in West Africa | History | Ebola (Ebola Virus Disease) | CDC

2014 Ebola Outbreak in Democratic Republic of the Congo | Democratic Republic of Congo | Outbreaks | Ebola (Ebola Virus Disease) | CDC

Activity Review

Source:

- What are the relevant factors?
- Where are they coming from?

Transformation:

- What combination of factors quantifies risk?
- How do the factors relate?
- Weights?

Output:

- What does the output mean?
- Who and how could use the output?
- What is the best way to display the output?

HSR.health

IIT Bombay Winter School

Public Health Data Models

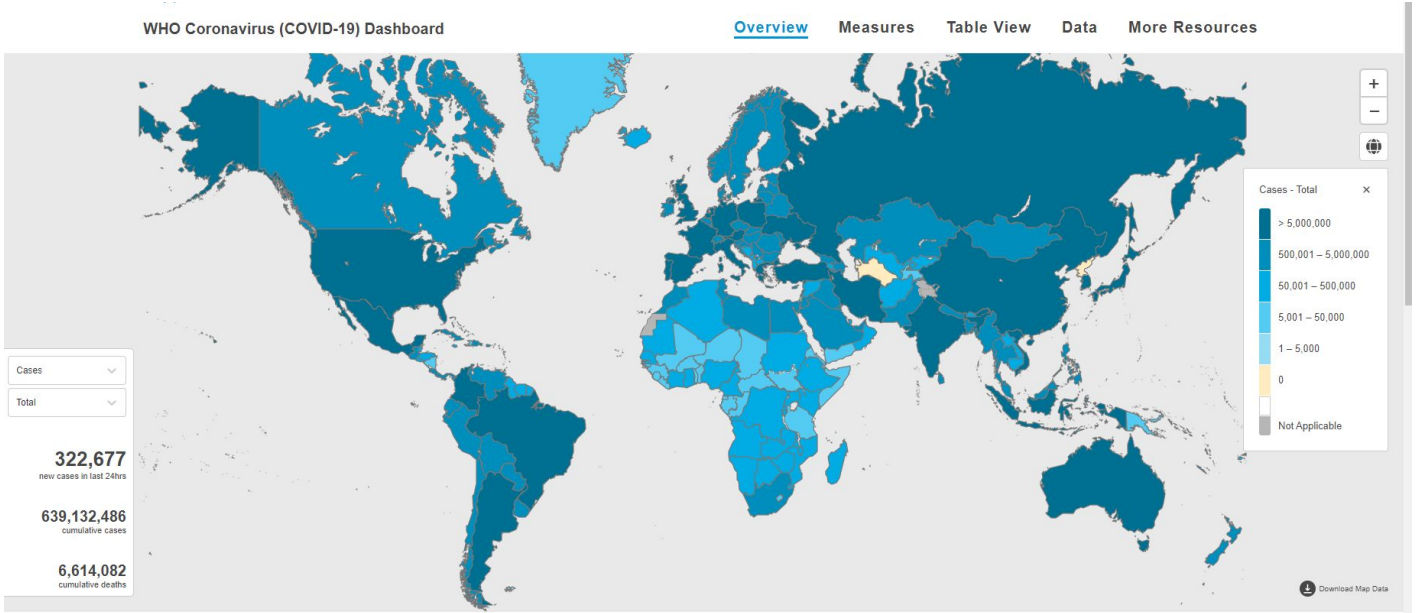
Session 2

Agenda

- COVID-19
- Health SDI
  - TRI/MRI
  - Medical Supplies need index
  - BRI
- Activity
- Final discussion

COVID-19

- Overview



COVID-19

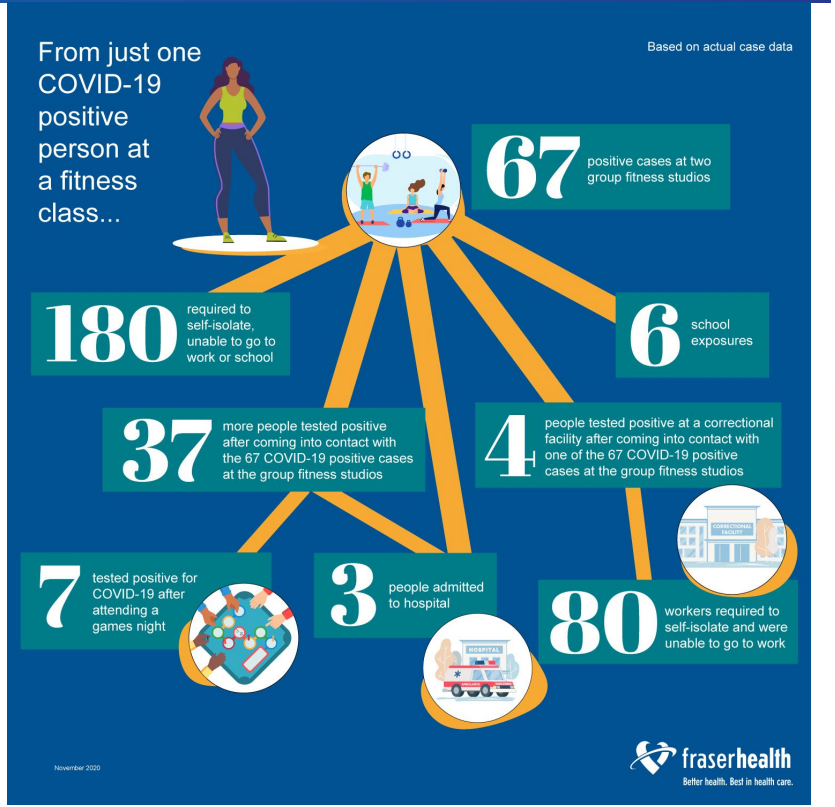
- Overview

Situation by Region, Country, Territory & Area

Name	Cases - cumulative total	Cases - newly reported in last 7 days	Deaths - cumulative total	Deaths - newly reported in last 7 days	Total vaccine doses administered per 100 population	Persons fully vaccinated with last dose of primary series per 100 population	Persons Boosted per 100 population
Global	**639,572,819**	**2,681,201**	**6,615,258**	**7,639**	**167.32**	**64.34**	**29.9**
+ By WHO Region							
+ By World Bank Income Group							
United States of America	97,329,491	296,882	1,069,757	2,611	195.04	68.16	33.98
India	44,672,638	2,155	530,622	21	159.36	68.9	16.04
France	36,768,890	350,774	155,331	438	227.4	77.48	60.47
Germany	36,499,600	168,434	157,943	70	226.9	78.04	62.46
Brazil	35,232,625	180,473	689,665	510	231.79	79.18	49.88
Republic of Korea	27,155,813	371,671	30,568	345	257.01	87.17	65.63
Japan	24,793,166	664,252	49,644	872	271.49	81.42	66.47
Italy	24,260,660	36,907	181,098	89	248.9	82.95	75.04

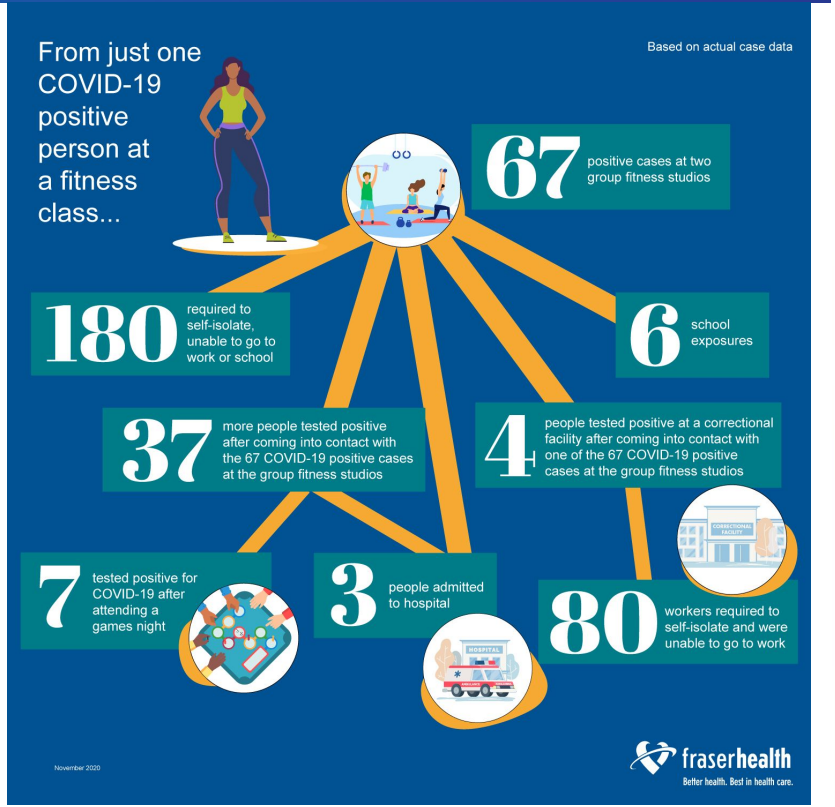
COVID-19

- Epidemiological Characteristics
  - Mode of transmission
    - Droplet
  - R0
    - 3.3 - 18.6 (depending on variant)
  - Mortality rate
    - 0.1 - 5.0% varying by
      - Age
      - Comorbidities
      - Vaccination status
      - Variant
  - Possibility of asymptomatic spread



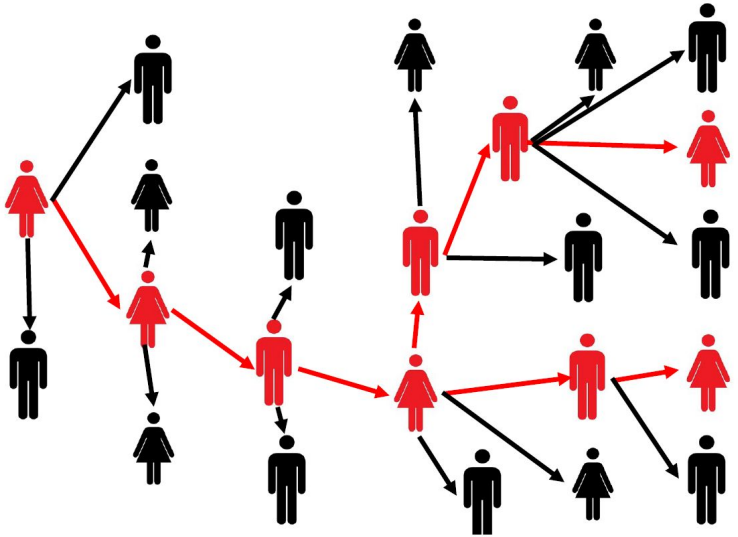
COVID-19

- Two major problems
  - Hospital capacity and load balancing
  - Containing community transmission



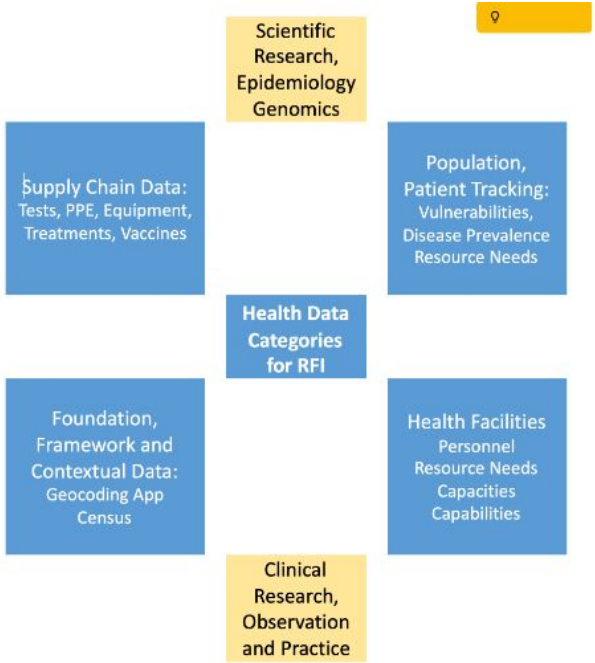
COVID-19

- Interventions
  - Pharmaceutical
    - Vaccinations
      - Targeted vs Mass
  - Non-Pharmaceutical
    - Contact tracing
    - Isolation
    - Lockdowns
  - Hospital
    - Medical supplies routing
    - Load balancing
    - Triage



COVID-19

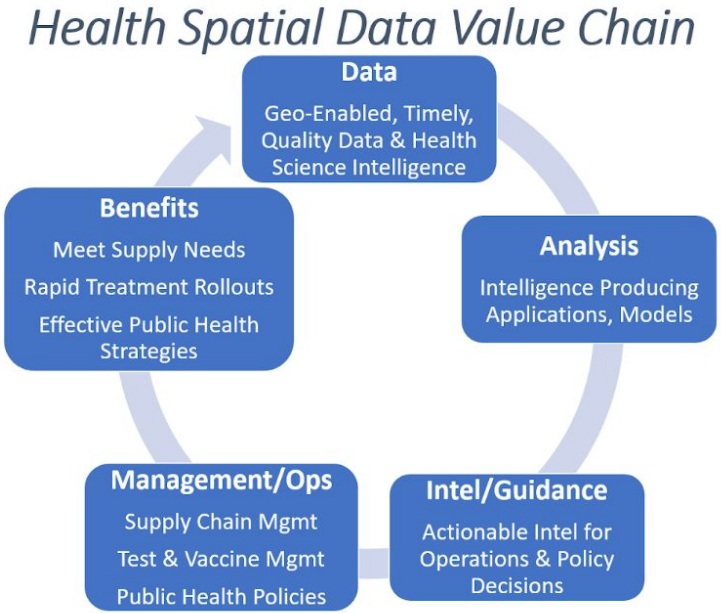
- Main challenge
  - Gathering and making available all of the required data to inform interventions.



Health Spatial Data Infrastructure

Health Spatial Data Infrastructure

- Allows for the creation and support of:
  - Transmission Risk Index
  - Mortality Risk Index
  - Medical Supplies needs Index
  - Business Risk Index



[Open Geospatial Consortium Health Spatial Data Infrastructure Concept Development Study]

Transmission Risk Index (TRI)

Sources

COVID-19 data (Johns Hopkins)

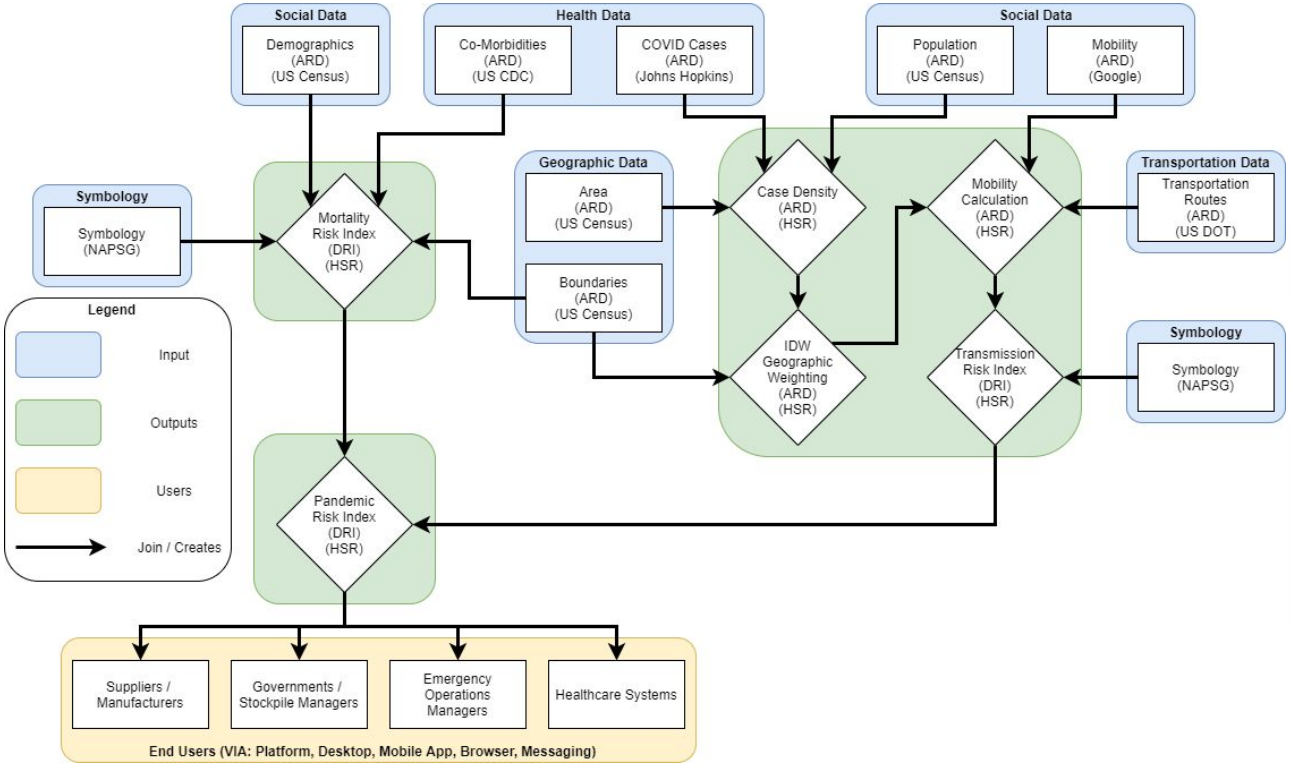
- Case counts
- Deaths
- Vaccinations

Mobility data

- Google mobility report

Demographics

- U.S. Census Bureau



Transmission Risk Index (TRI)

- Transformations

$$\frac{\{Cases/\frac{Population}{100{,}000}\}}{\frac{Area}{10km^2}} = Case\ Density$$

$$\frac{\sum_{0\to j}\{\frac{Case\ Density_j}{Distance_j\ from\ Current\ Admin}\}}{\sum_{0\to j}\{\frac{1}{Distance_j\ from\ Current\ Admin}\}} = IDW\ for\ Current\ Admin\ Region$$

Where the j subscript denotes the neighboring Admin Regions

$$Case\ Density\ *\ (1\ +\ Average\ Change\ in\ Mobility\ as\ a\ percent) = Mobility\ Factor$$

$$\frac{(Case\ Density + IDW + Mobility\ Factor)}{3} = Final\ Raw\ Transmission\ Risk\ Index\ Value$$

Transmission Risk Index (TRI)

- Transformations

Normalizing the Transmission Risk Index

Log Transformed Raw TRI (LTR_TRI) = $\log_{10}$(Final Raw Transmission Risk Index Value)

CMax_LTR_TRI = Maximum LTR _TRI of all states since the beginning of the pandemic

DMax_LTR_TRI = Maximum LTR _TRI of all states for a specific day

CMin_LTR_TRI = Minimum LTR _TRI of all states since the beginning of the pandemic

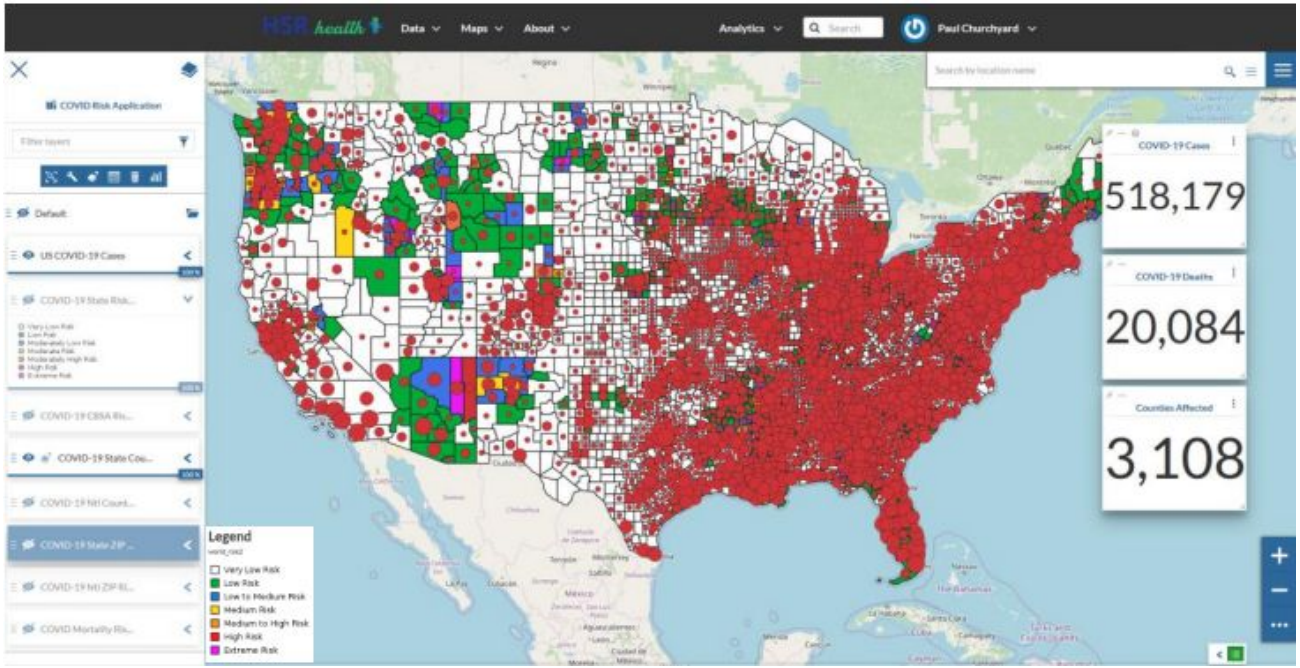
DMin_LTR_TRI = Minimum LTR _TRI of all states for a specific day

Cumulatively Normalized TRI = (LTR_TRI – CMin_LTR_TRI) * (100 / (CMax_LTR_TRI-CMin_LTR_TRI))

Daily Normalized TRI = (LTR_TRI – DMin_LTR_TRI) * (100 / (DMax_LTR_TRI-DMin_LTR_TRI))

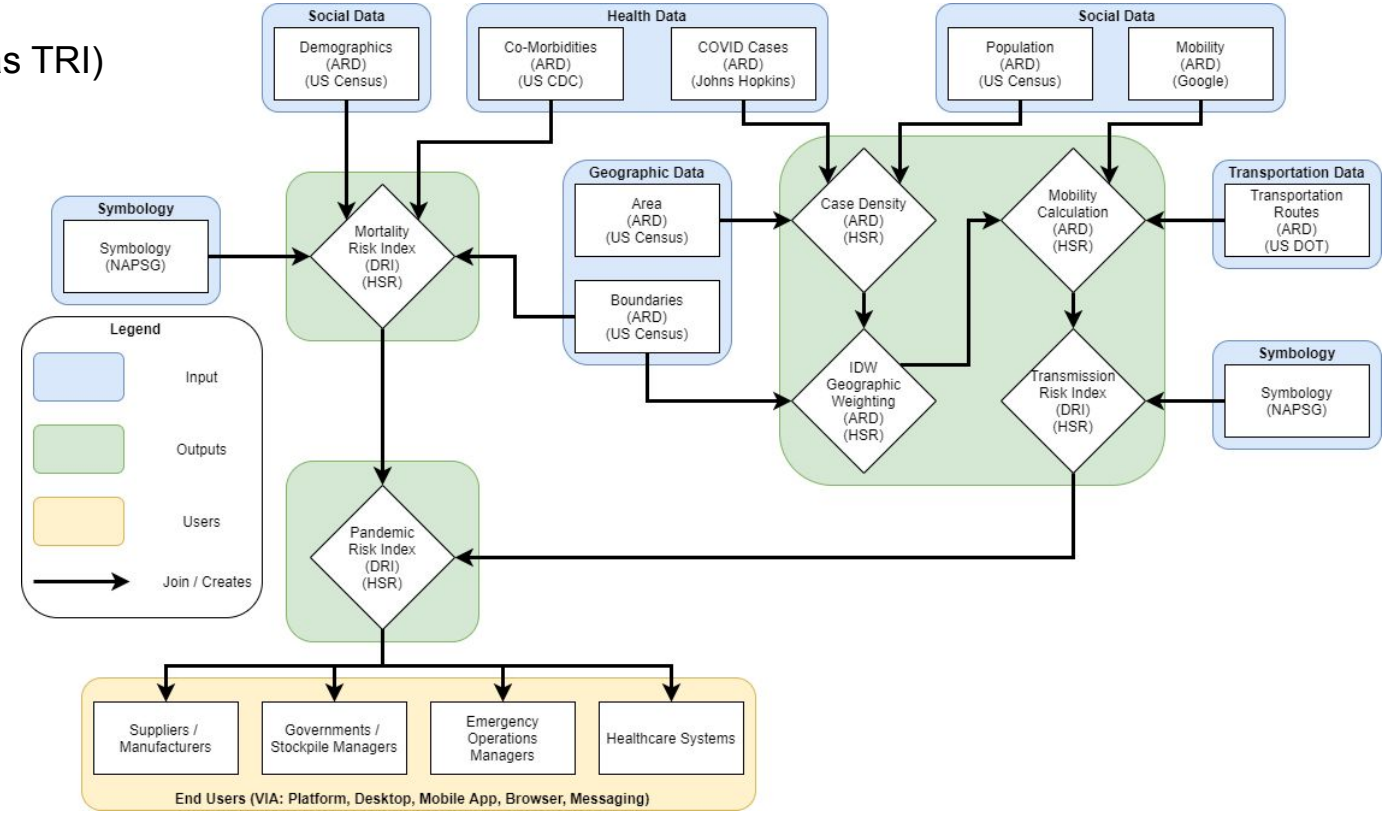
Transmission Risk Index (TRI)

- Output



Mortality Risk Index (MRI)

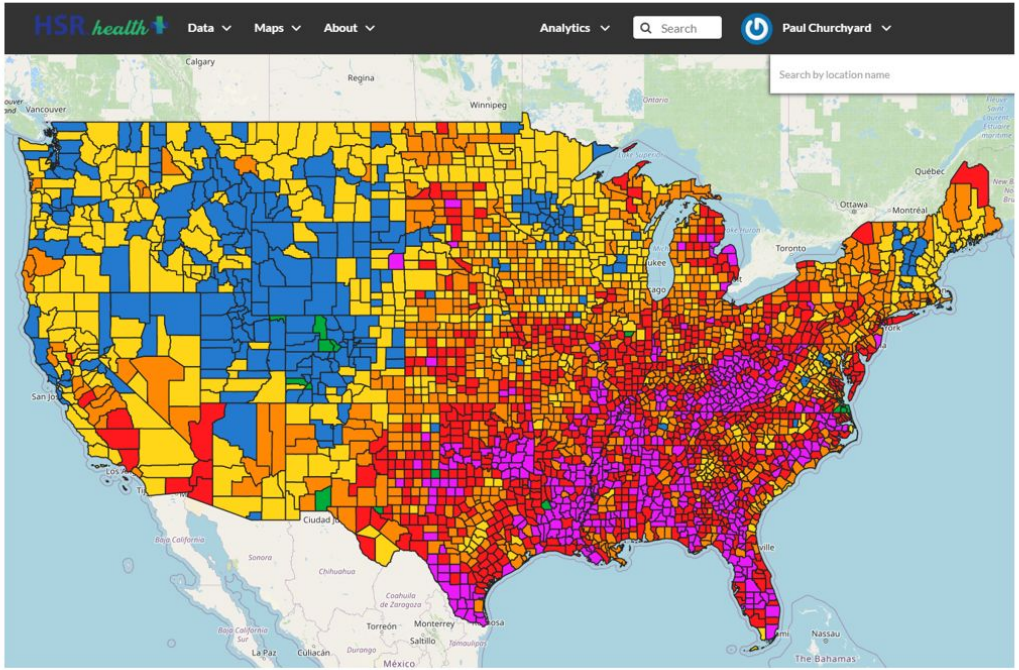
- Source (same as TRI)
- transformations



Mortality Risk Index (MRI)

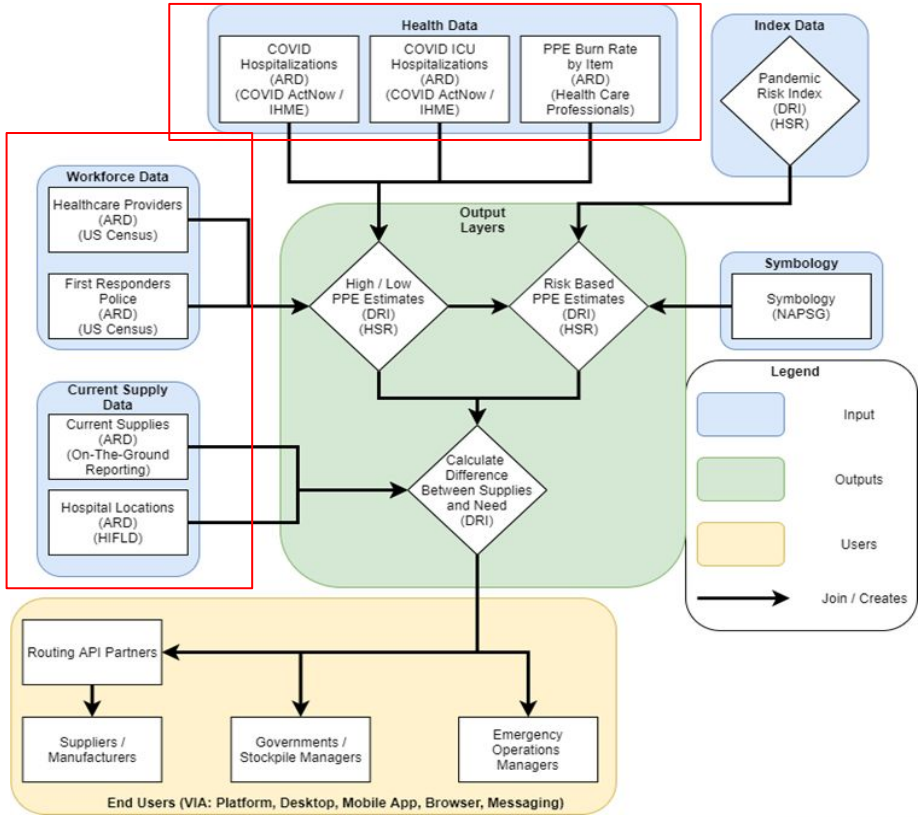
- Output

National Mortality Risk Index



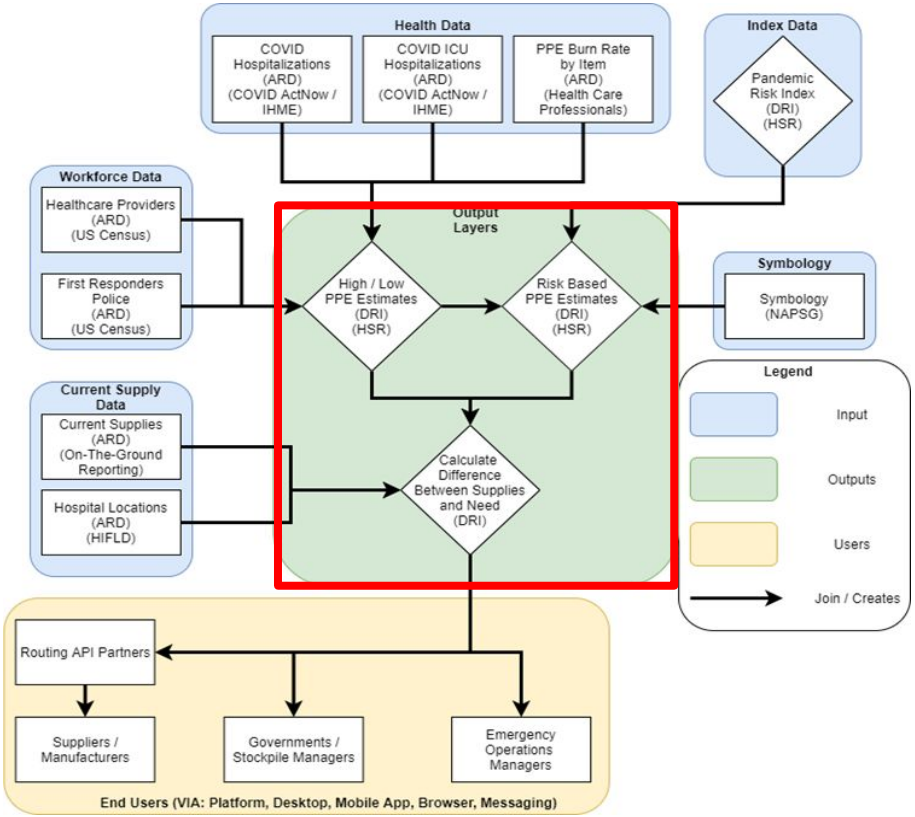
Medical Supplies Needs Index

Sources

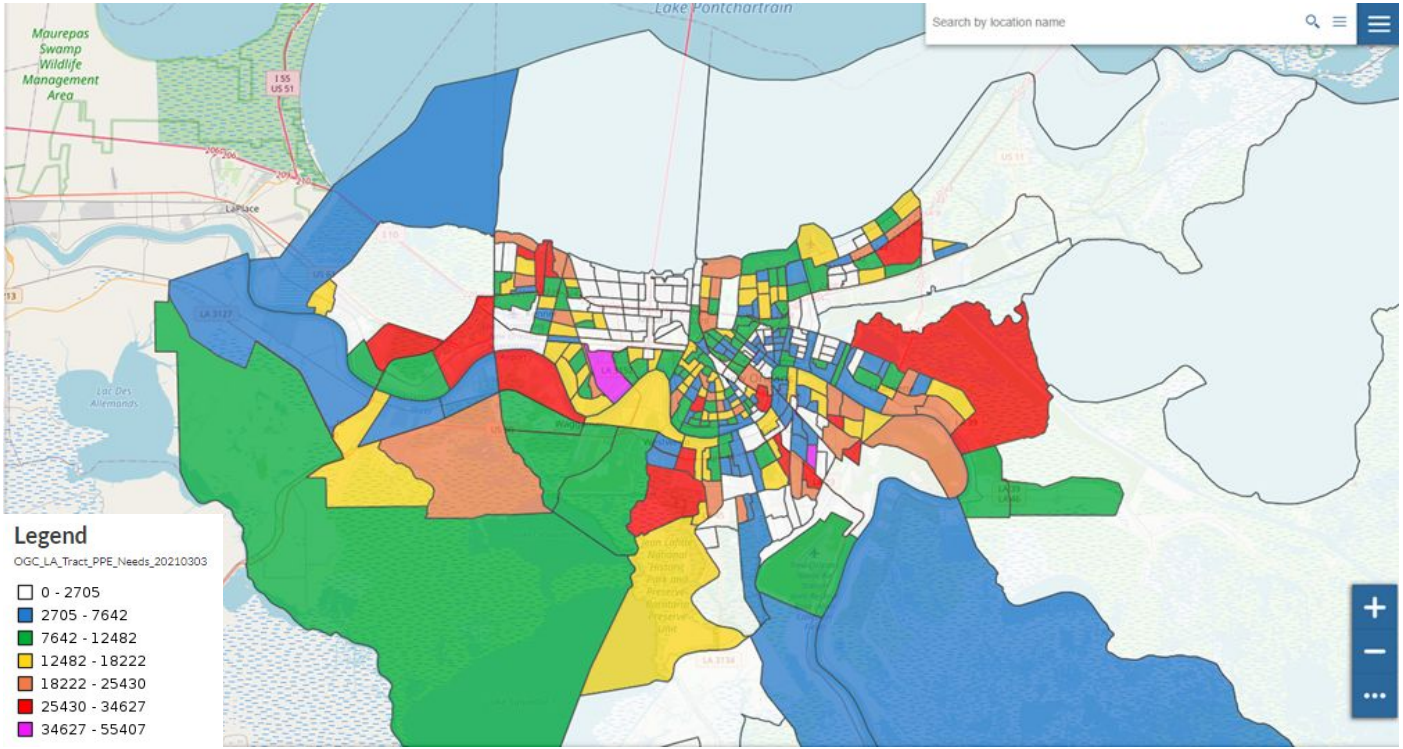


Medical Supplies Needs Index

Transformations

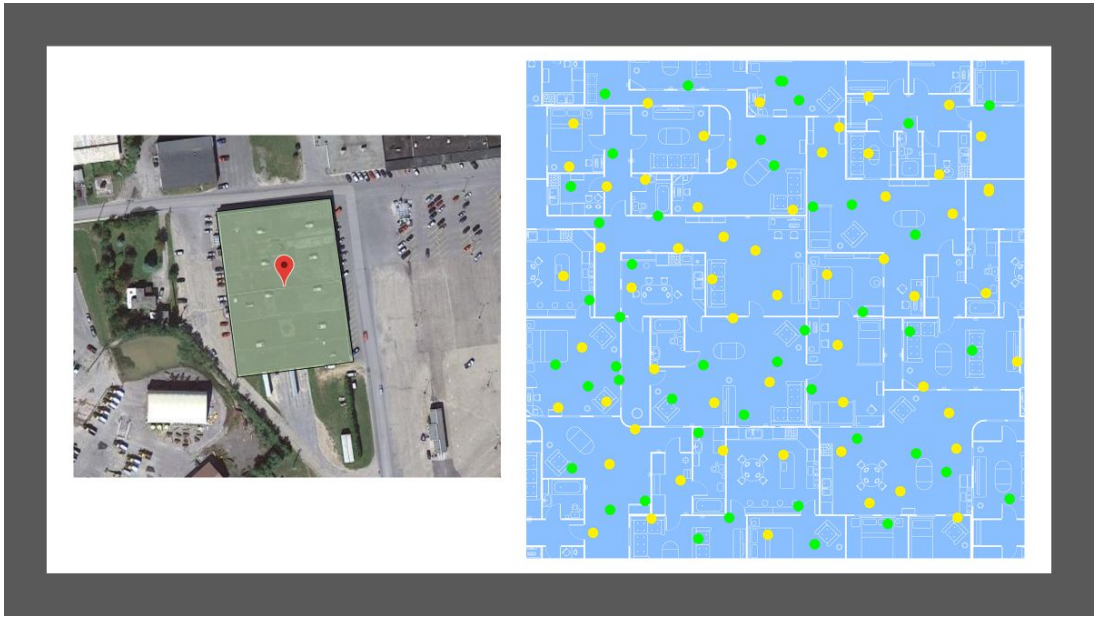


Medical Supplies Needs Index

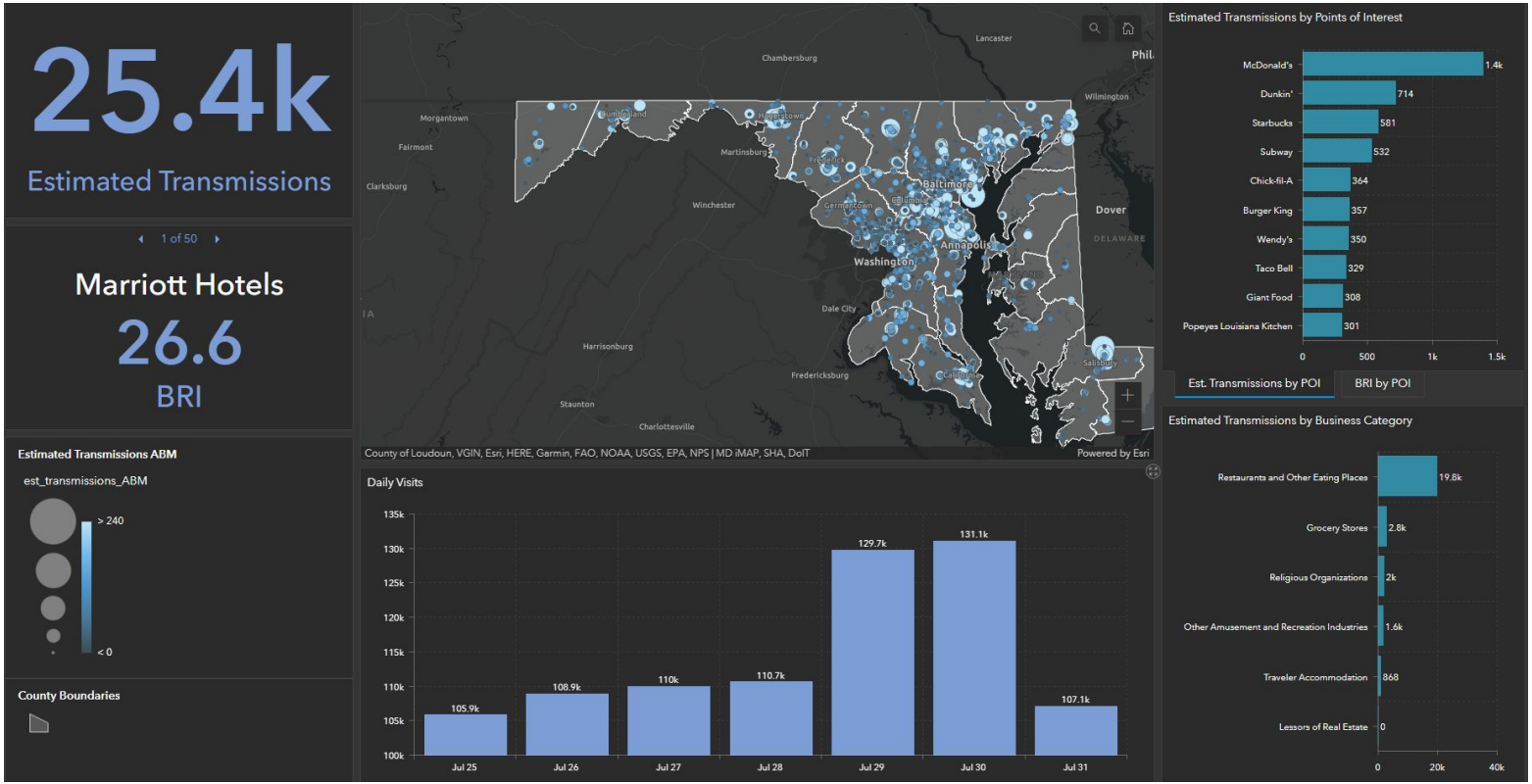


Business Risk Index (BRI)

- Transformations
  - Agent Based model



Business Risk Index (BRI)



Business Risk Index (BRI)

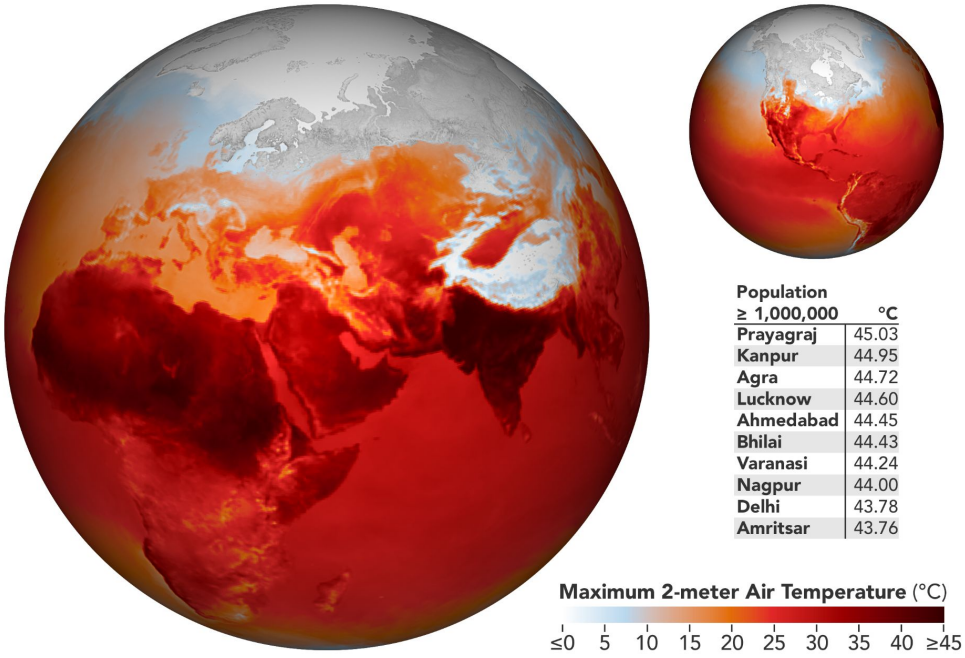
- Source
  - Point of Interest (POI) data (Safegraph/Advan)
    - Building size
    - Location (lat/long)
    - Visitors
      - Arrival
      - Dwell
      - Home CBG
  - COVID-19
    - Case counts
    - Vaccinations

Summary

- COVID-19
- Health SDI
- TRI/MRI
- Medical Supplies need index
- BRI

Activity

- Heat models and health



Activity 2: Using Data to Assist with Extreme Heat

Activity 2: :Early Season Heat Waves Strike India (nasa.gov)

Time to complete: 1 hour
1 hour discussion afterwards

Task: Come up with an index that quantifies the risk of mortality from heat waves.

Source:

- What are the relevant factors?
  - Regions previously affected by heat waves
  - Population characteristics
    - Age
    - Comorbidities
  - Housing characteristics
    - AC
    - Urban characteristics
    - vulnerability
- Where are they coming from?

Transformation:

- What combination of factors quantifies risk?
- How do the factors relate?
- Weights?

Output:

- What does the output mean?
- Who and how could use the output?
- What is the best way to display the output?

To get you started:

Heat-related mortality in India: excess all-cause mortality associated with the 2010 Ahmedabad heat wave - PubMed (nih.gov)

IJERPH | Free Full-Text | Mortality during a Large-Scale Heat Wave by Place, Demographic Group, Internal and External Causes of Death, and Building Climate Zone (mdpi.com)