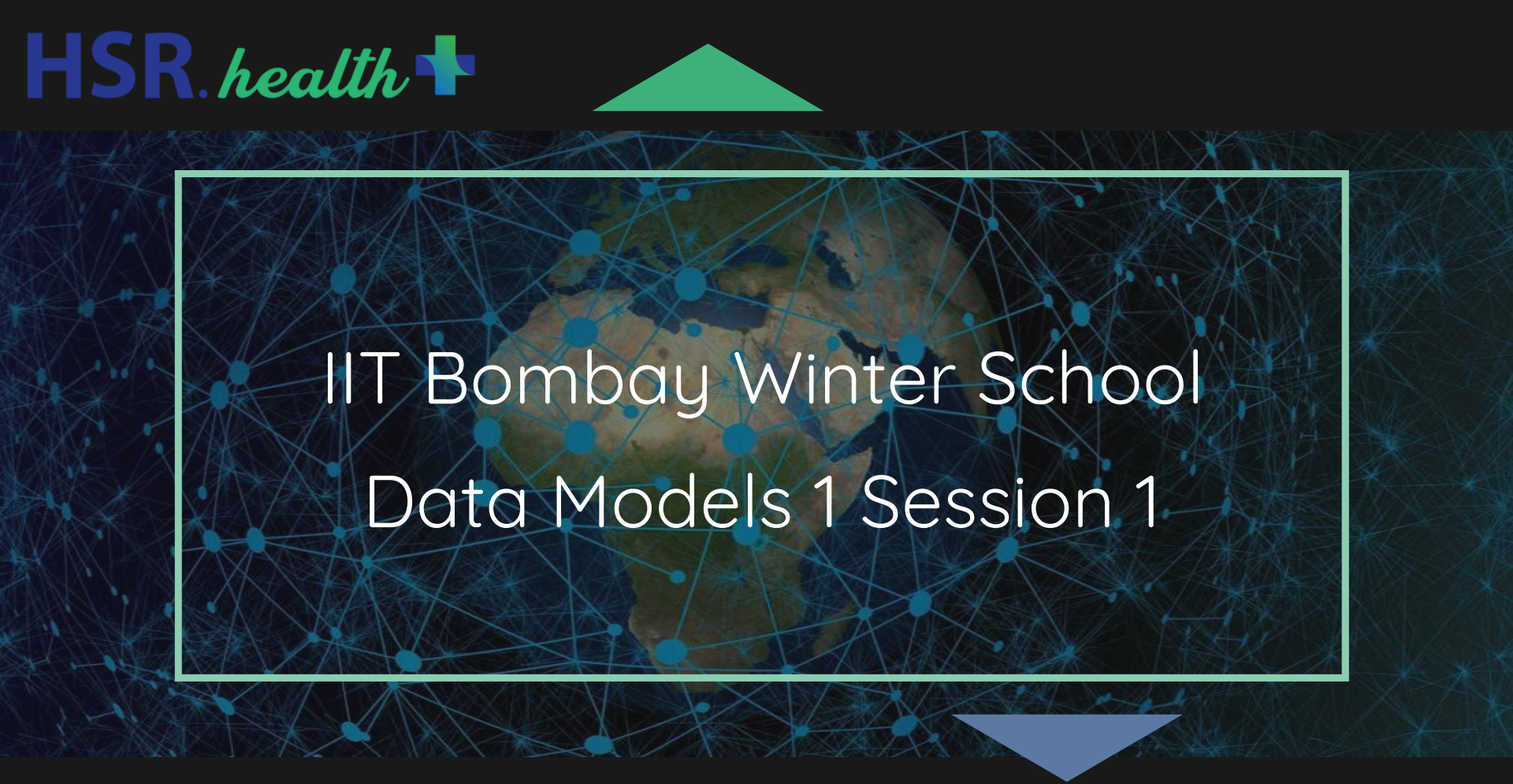

HSR.health
IIT Bombay Winter School
Data Models 1 Session 1

Introduction

- Paul A Churchyard
- Chief Geospatial Engineer @ HSR.*health*
- BS Earth Science, B.S. Geography,
- MPS Geographic Information Systems

- Ajay K Gupta, CISSP, MBA
- CEO @ HSR.*health*
- BS Electrical Engineering
- MS Electrical Engineering
- MBA

Introduction

- Structure of course
  - Day 1: Intro to data models
  - Day 2: Geospatial data models
  - Day 3: GIS and health data models
  - Day 4: EO data models

- Lecture and labs
  - Each day lecture material will be explored in labs
  - There will be a final Lab and presentation during the afternoon session of Day 4

HSR.health

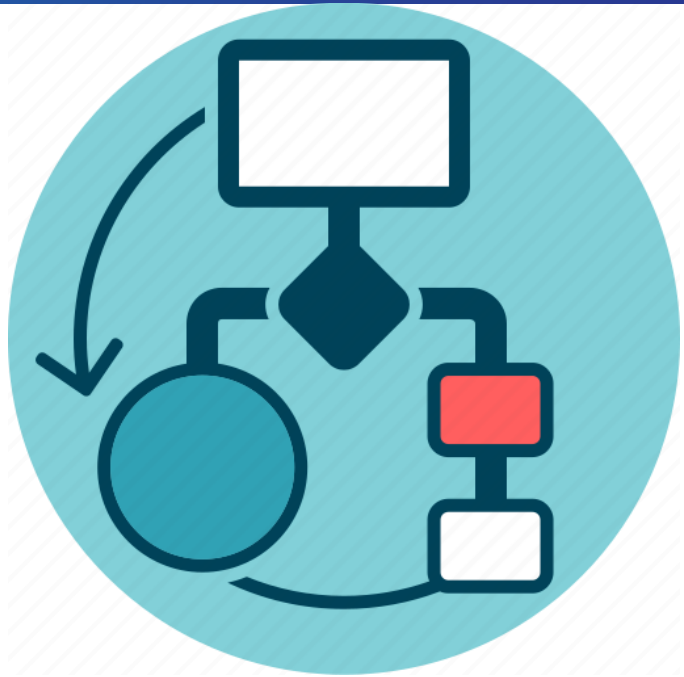
Open Geospatial Consortium

Agenda

- What are data models?
  - Components
    - Source
      - Source and data types
    - Transformations
      - Aggregations and mathematical operations
    - Output
      - Data and Inferences
      - Presentation
- Lab
- Discussion

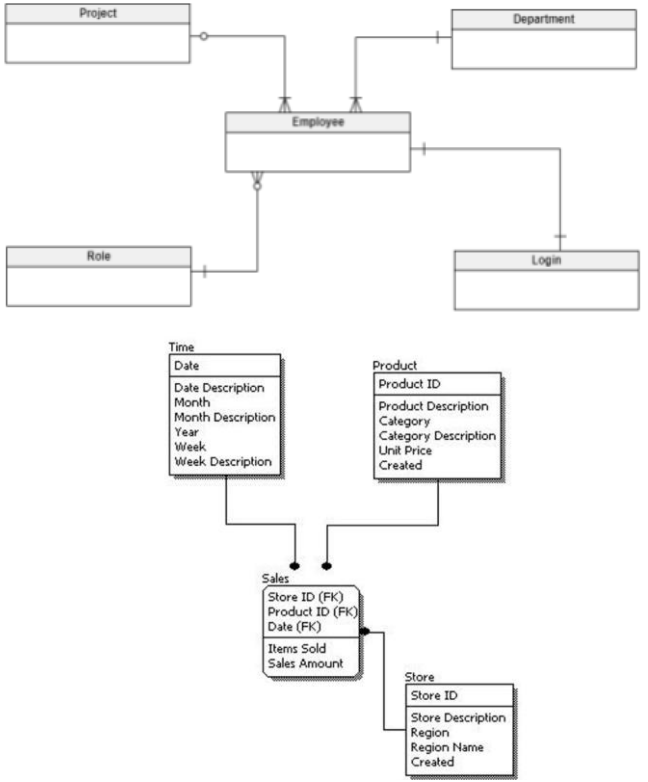
What are Data Models?

An abstract model that organizes elements of data and standardizes how they relate to one another and to the properties of real-world entities.



Types of Data Models

- Conceptual
  - High level model of how information is organized and relates to each other.
- Logical
  - Strategy for representing conceptual model in existing software, hardware, or standards.
- Physical
  - Detailed description of how data will be organized within software, hardware, and standards constraints.



Components of Data Models

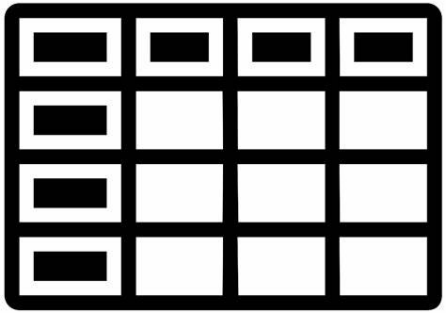
- **Source**
  - **What does the data look like and where is it coming from?**
- **Transformations/Relationships**
  - **What's happening to the data?**
- **Output**
  - **What is the data telling you?**

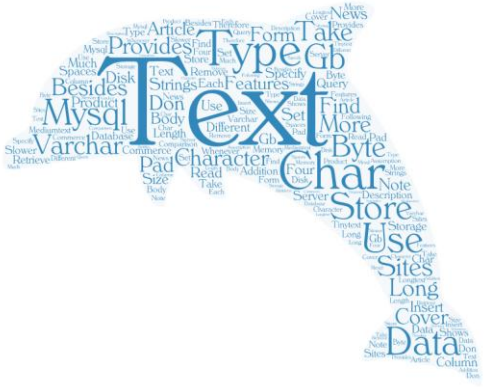
Components of Data Models: Source

- **Data types**
  - **Qualitative VS Quantitative data**
  - **Quantitative**
    - **Integer**
    - **Float**
    - **Boolean**
    - **Character**
    - **Date**

Components of Data Models: Source

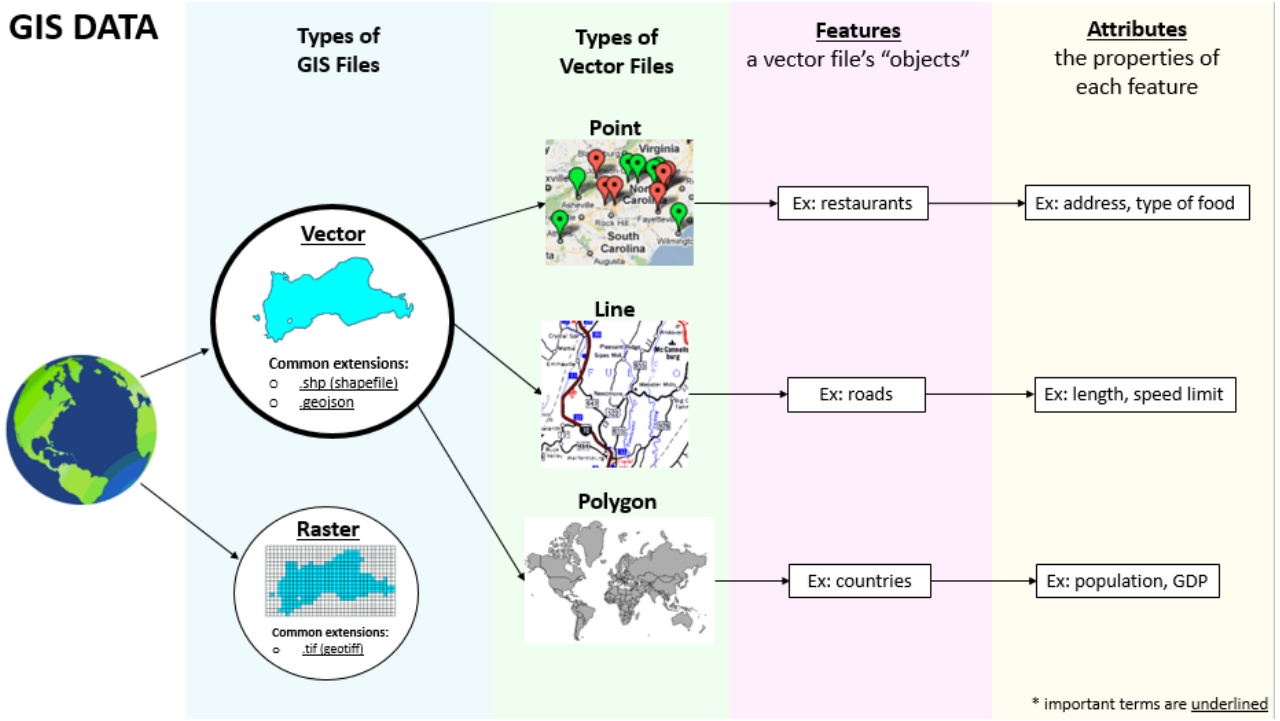
- **Data formats**
  - **Text (TXT)**
  - **Image (PNG, JPEG)**
  - **Tabular (CSV, Parquet)**
  - **Dictionary (JSON)**





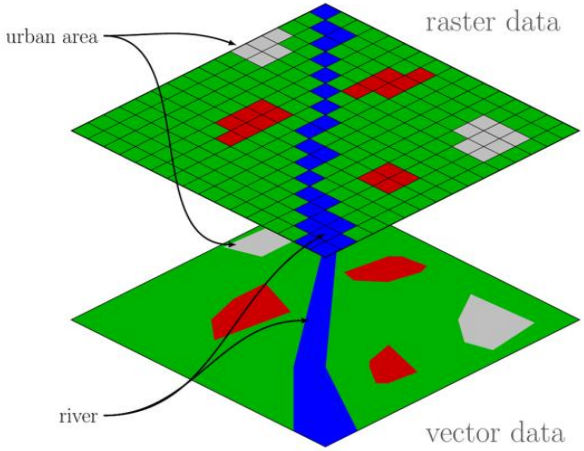
Components of Data Models: Source

- **Geospatial data types**



Components of Data Models: Source

- **GIS specific formats**
  - **Vector**
    - **Shapefiles**
    - **GeoJSON**
  - **Raster**
    - **Geotiff**

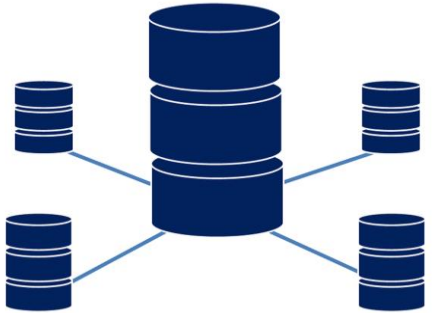


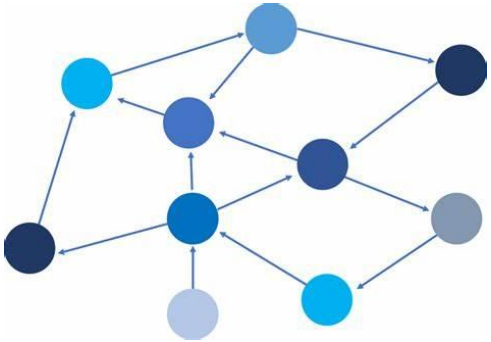
Components of Data Models: Source

- **Scales**
  - **Temporal**
  - **Spatial**
  - **Population vs Individual**

Components of Data Models: Source

- **Databases**
  - **Relational**
  - **Non-Relational**
  - **Graph**





Components of Data Models: Source

- **Relational Databases**
  - **Pros**
    - **Powerful Query Languages (SQL)**
  - **Cons**
    - **Low velocity**
    - **Only can handle structured data**
  - **Use cases**
    - **Private Patient Information**
    - **Spatial and Non-Spatial data**

Components of Data Models: Source

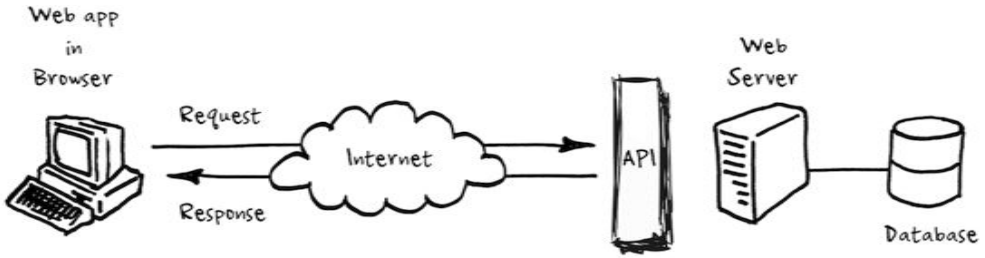
- **Non-relational (NoSQL) Databases**
  - **Pros**
    - **High Velocity**
    - **Can handle all kinds of data**
  - **Cons**
    - **Weak query language**
  - **Use cases**
    - **Key-value databases**
    - **Document databases**

Components of Data Models: Source

- **Graph databases**
  - **Pros**
    - **Allows semantic queries**
    - **High Velocity**
  - **Cons**
    - **No standard query language**
  - **Use cases**
    - **Transaction data**
    - **Supply chain**

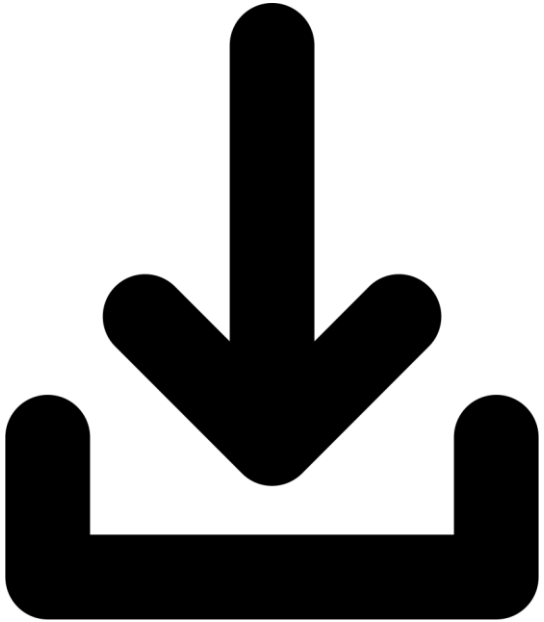
Components of Data Models: Source

- **Application Programming interface (API)**
  - **REST**
  - **SOAP**
  - **RPC**



Components of Data Models: Source

- One time file download and other sources



Components of Data Models: Transformations

- **What is happening to the data?**
  - **Joins/Relationships**
  - **Aggregations**
  - **Scaling and Normalization**

Components of Data Models: Transformations

- **Joins**
  - **For joining data with a hierarchical relationship**
    - **Primarily SQL and NoSQL**
  - **Join types**


Left outer join

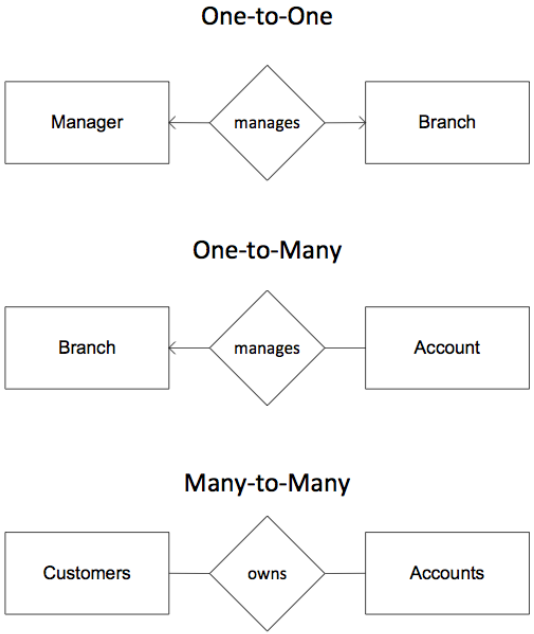

Inner join


Right outer join


Full outer join

Components of Data Models: Transformations

- **Relationships**

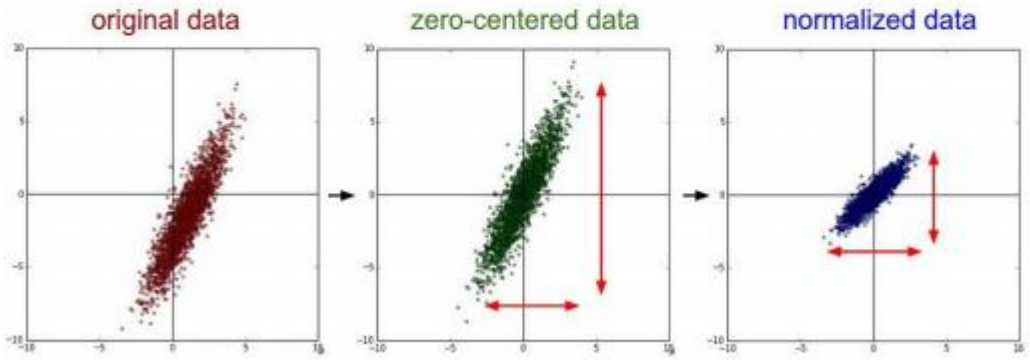


Components of Data Models: Transformations

- **Aggregations**
  - **Sum**
  - **Min**
  - **Max**
  - **Binning**
  - **Count**
  - **Central Tendency**
    - **Mean, Median, Mode**

Components of Data Models: Transformations

- **Scaling and Normalization**



Components of Data Models: Output

- **What does the data mean once transformed**
- **Data**
  - **Types**
  - **Formats**
  - **Scales**
- **Inferences**
  - **5 W's**
- **Presentation**
  - **Display and visualization**

Components of Data Models: Output

- **Data**
  - **Types**
  - **Formats**
  - **Scales**
  - **All of these variables influence how meaningful output can be**

Components of Data Models: Output

- **Inferences**
  - **Who**
  - **What**
  - **Where**
  - **When**
  - **Why**

Components of Data Models: Output

- **Presentation**
  - **Display and visualization**
  - **ArcGIS, ESRI, Qgis**

Summary

- What are data models?
  - Components
    - Source
      - Source and data types
    - Transformations
      - Aggregations and mathematical operations
    - Output
      - Data and Inferences
      - Presentation

Activity

- Background and introduction
  - Land Suitability Assessment and Land Suitability Index (LSI)
- Activity
  - Dissect the LSI using the Source, Transformations, Output (STO) framework
  - http://currikicdn.s3.amazonaws.com/resourcedocs/54d33dece9c78.pdf
- Discussion

Time to complete: 1 hour
1 hour discussion afterwards

Activity Review

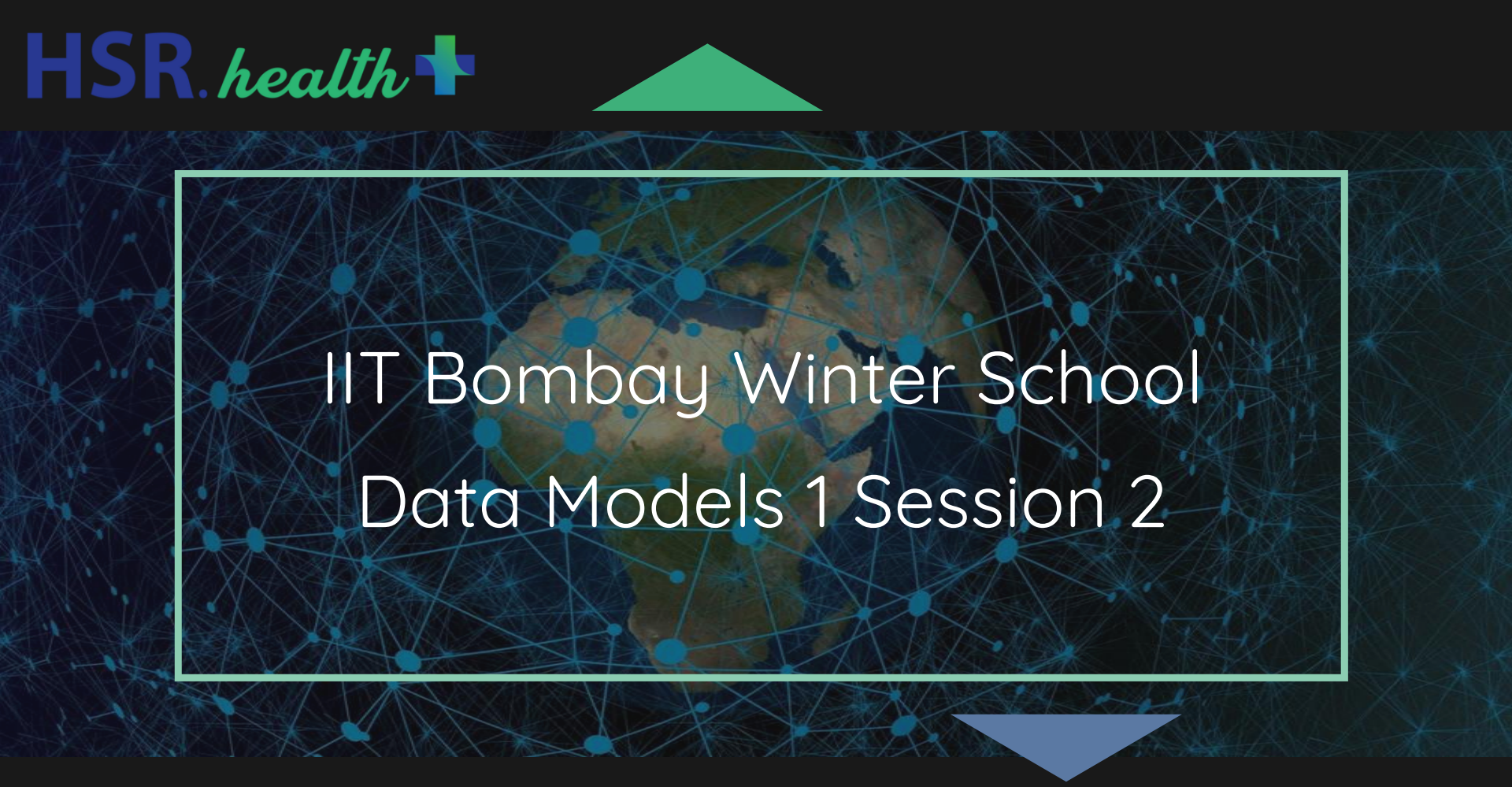
- Going over land suitability assessment activity
  - Source
    - Data types
    - Formats
    - Scales
  - If you had to store this data in one place, what storage option would you choose?
    - Relational
    - Non-Relational
    - Graph
  - Why?

Activity Review

- Going over land suitability assessment activity
  - Transformations
    - Joins/Relationships
    - Aggregations
    - Scaling/Normalizing
  - Would you recommend any changes to the index calculation?

Activity Review

- Going over land suitability assessment activity
  - Output data
    - Format
    - Type
    - Scale
  - Inferences
    - 5 W’s
  - Presentation
    - If you had to present the index to a stakeholder, how would you?



IIT Bombay Winter School

Data Models 1 Session 2

Agenda

- Going over land suitability assessment activity (recap)
  - Source
  - Transformations
  - Output
- Case study
  - Social vulnerability index (SVI)
- Activity
- Final discussion

Activity Review

- Going over land suitability assessment activity
  - Source
    - Data types
    - Formats
    - Scales
  - If you had to store this data in one place, what storage option would you choose?
    - Relational
    - Non-Relational
    - Graph
  - Why?

Activity Review

- Going over land suitability assessment activity
  - Transformations
    - Joins/Relationships
    - Aggregations
    - Scaling/Normalizing
  - Would you recommend any changes to the index calculation?

Activity Review

- Going over land suitability assessment activity
  - Output data
    - Format
    - Type
    - Scale
  - Inferences
    - 5 W’s
  - Presentation
    - If you had to present the index to a stakeholder, how would you?

Natural Disasters, Health, and GIS

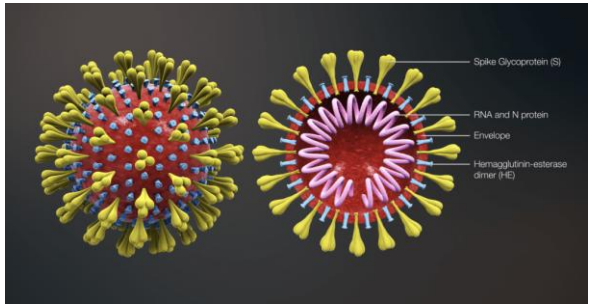
- Health and disaster response domains
  - GIS applications

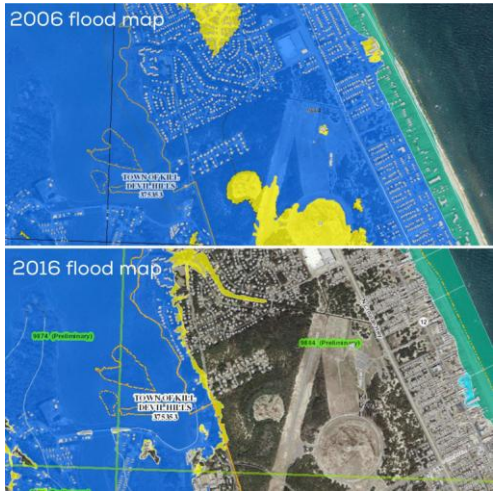
Natural Disasters, Health, and GIS

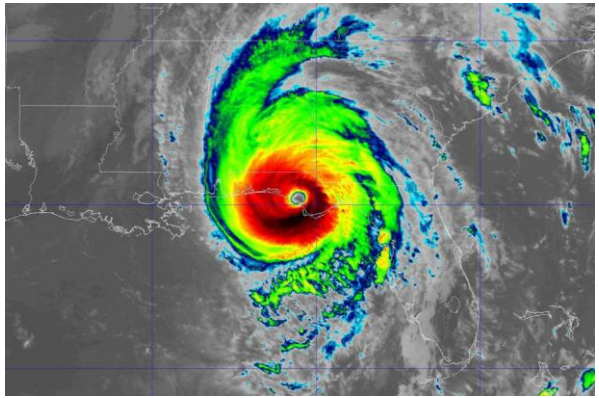
- Social Vulnerability Index (SVI)
  - “a database that helps emergency response planners and public health officials identify, map, and plan support for communities that will most likely need support before, during, and after a public health emergency. The tool is commonly used across CDC/ATSDR, in addition to many emergency preparedness and response organizations.”

Natural Disasters, Health, and GIS

- Use cases







Natural Disasters, Health, and GIS

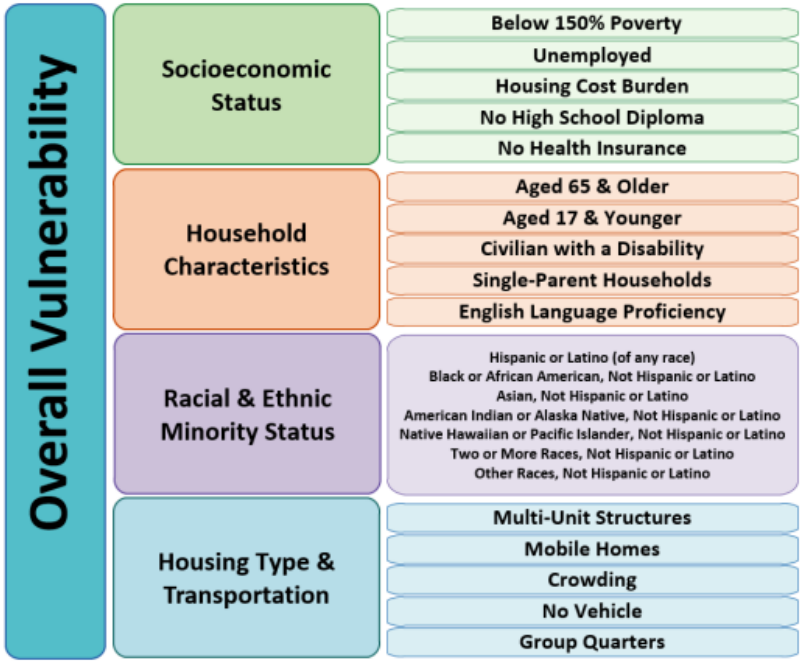
- SVI Source
  - 16 social factors
    - Unemployment
    - Racial and Ethnic Minority
    - Disability
    - Etc
  - Collected from American Community Survey (ACS)
    - Household level survey
      - One time file download or API access

Natural Disasters, Health, and GIS

- SVI Source
  - Format
    - Tabular
  - Types
    - Text
      - Place descriptions (location name)
    - Int
      - Population statistics
  - Scales
    - Geospatial
      - State, County, Tract
    - Temporal
      - 1 year span
    - Population level

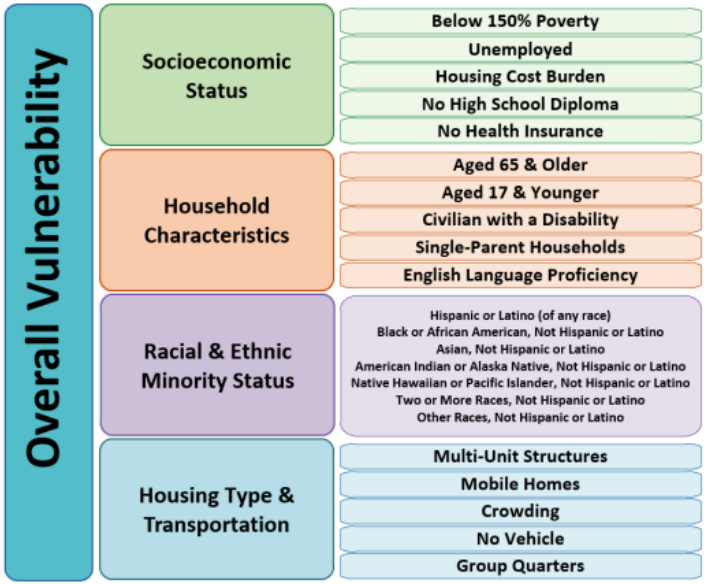
Natural Disasters, Health, and GIS

- SVI Source



Natural Disasters, Health, and GIS

- SVI transformations goal:
  - Relate data from 4 themes in a comprehensive way that allows for easy comparison across geographic areas.



Natural Disasters, Health, and GIS

- SVI transformations
  - Joins
    - Based on geographic unit (tract-level FIPS) and year
  - Aggregations
    - Sum
      - Ex: Summing fields based on those with income less than threshold
  - Scaling/Normalizing
    - Scaling
      - Square root transformation to calculate margin of error (MOE)
    - Normalization
      - (Persons below 150% poverty estimate / Population for whom poverty status is determined estimate) * 100

Natural Disasters, Health, and GIS

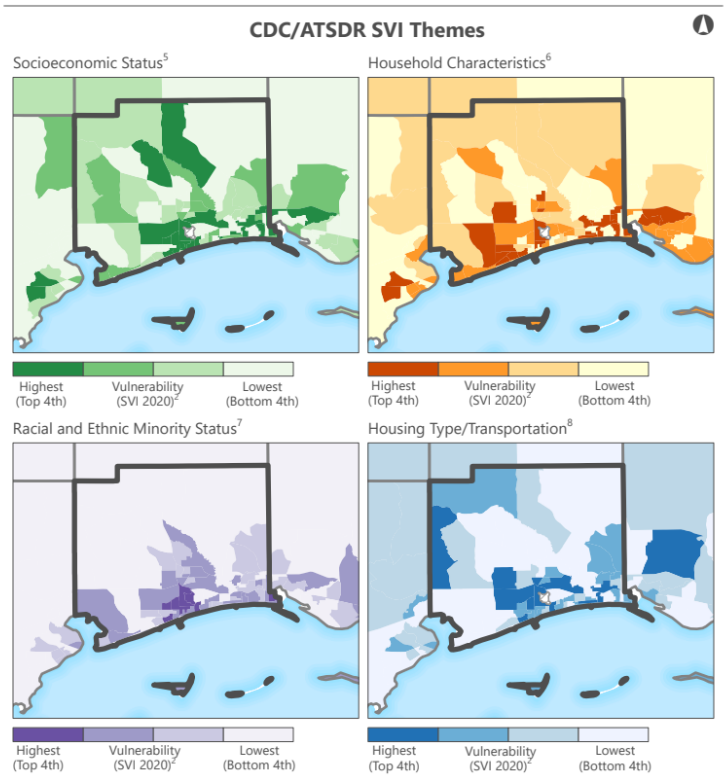
2020 VARIABLE NAME	2020 DESCRIPTION	THEME	CENSUS or SVI TABLE(S)	FIELD NAME CHANGED SINCE 2018?	2020 TABLE FIELD CALCULATION	CALCULATION DESCRIPTION	NOTES	2018 TABLE FIELD CALCULATION if changed
	than rooms estimate, 2016-2020 ACS					Occupants per room, occupied housing units, 1.51 or more		
M_CROWD	At household level (occupied housing units), more people than rooms estimate MOE, 2016-2020 ACS	4	DP04	No	SQRT (DP04_0078M ^2 + DP04_0079M ^2)	SQRT (MOE Occupants per room, occupied housing units, 1.01 to 1.50^2 + MOE Occupants per room, occupied housing units, 1.51 or more^2)		
E_NOVEH	Households with no vehicle available estimate, 2016-2020 ACS	4	DP04	No	DP04_0058E			
M_NOVEH	Households with no vehicle available estimate MOE, 2016-2020 ACS	4	DP04	No	DP04_0058M			
E_GROUPQ	Persons in group quarters estimate, 2016-2020 ACS	4	B26001	No	B26001_001E			
M_GROUPQ	Persons in group quarters estimate MOE, 2016-2020 ACS	4	B26001	No	B26001_001M			
EP_POV150	Percentage of persons below 150% poverty estimate	1	SVI and S1701	Yes	(E_POV150 / S1701_C01_001E) * 100	(Persons below 150% poverty estimate / Population for whom poverty status is determined estimate) * 100		S0601_C01_049E
MP_POV150	Percentage of persons below 150% poverty estimate MOE	1	SVI and S1701	Yes	((SQRT (M_POV150^2-((EP_POV150 / 100)^2 * S1701_C01_001M ^2))) / S1701_C01_001E) * 100	((SQRT(MOE Persons below 150% poverty^2 - ((Estimated proportion of persons below 150% poverty)^2 * MOE Population for whom poverty status is determined^2))) / Population for whom poverty status is determined estimate) * 100		S0601_C01_049M

Natural Disasters, Health, and GIS

- SVI Output
  - Data
    - Type: index value from 0-100
    - Format: Tabular
    - Scales: State, County, Tract

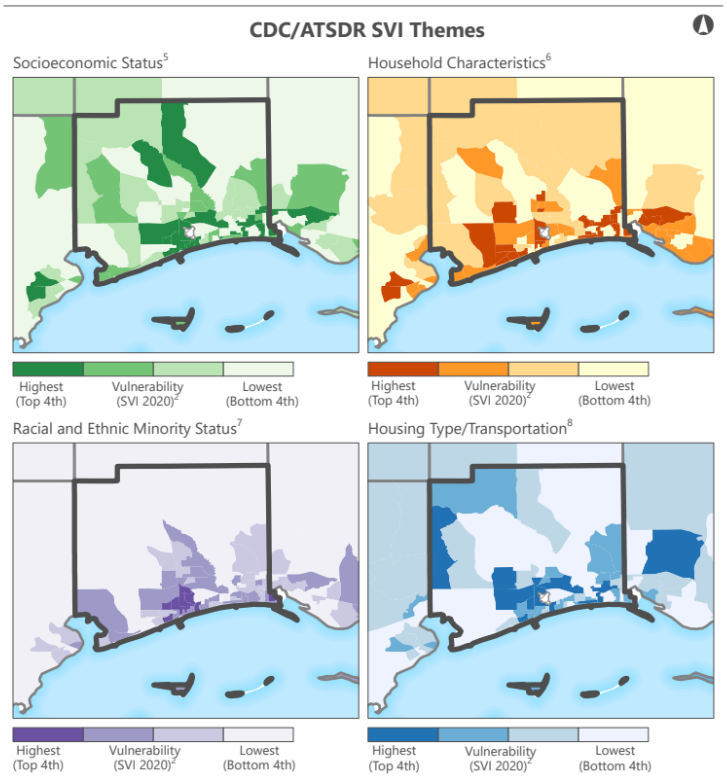
Natural Disasters, Health, and GIS

- What does the output tell you?



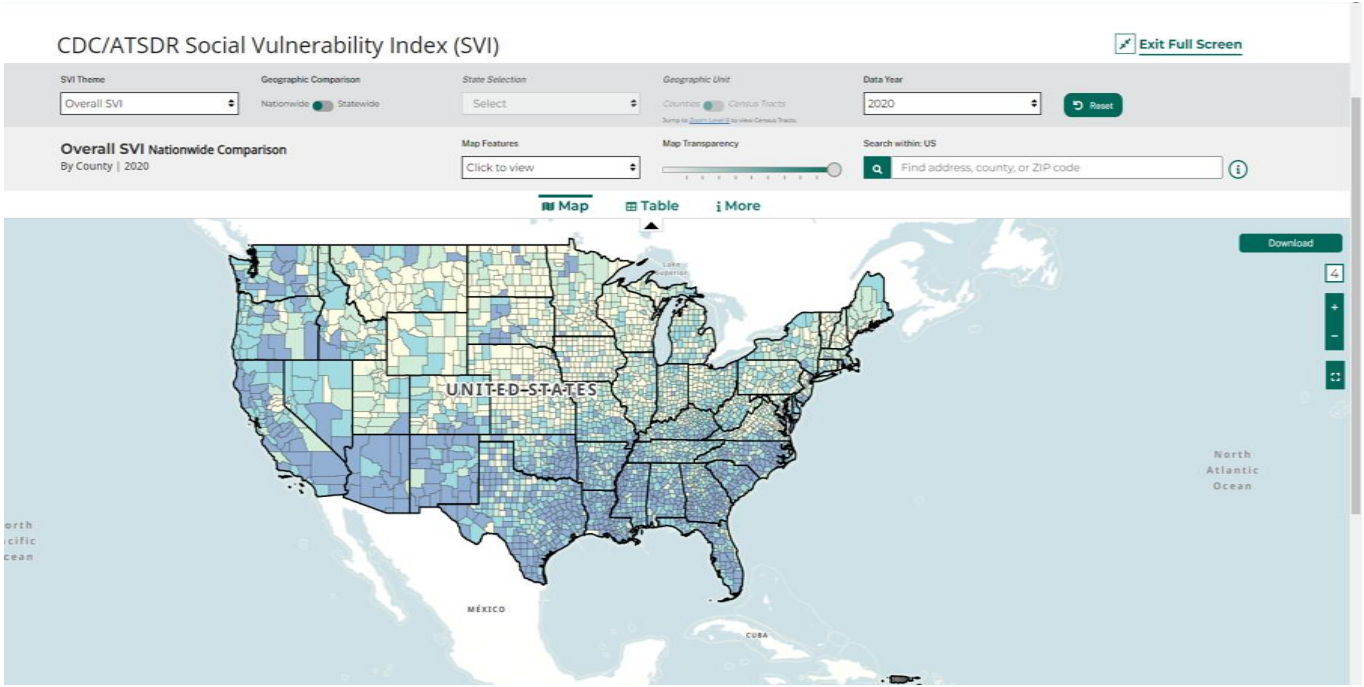
Natural Disasters, Health, and GIS

- Whos is benefiting from the output?
  - First responders
  - Public health departments
  - NGOs/special interest groups



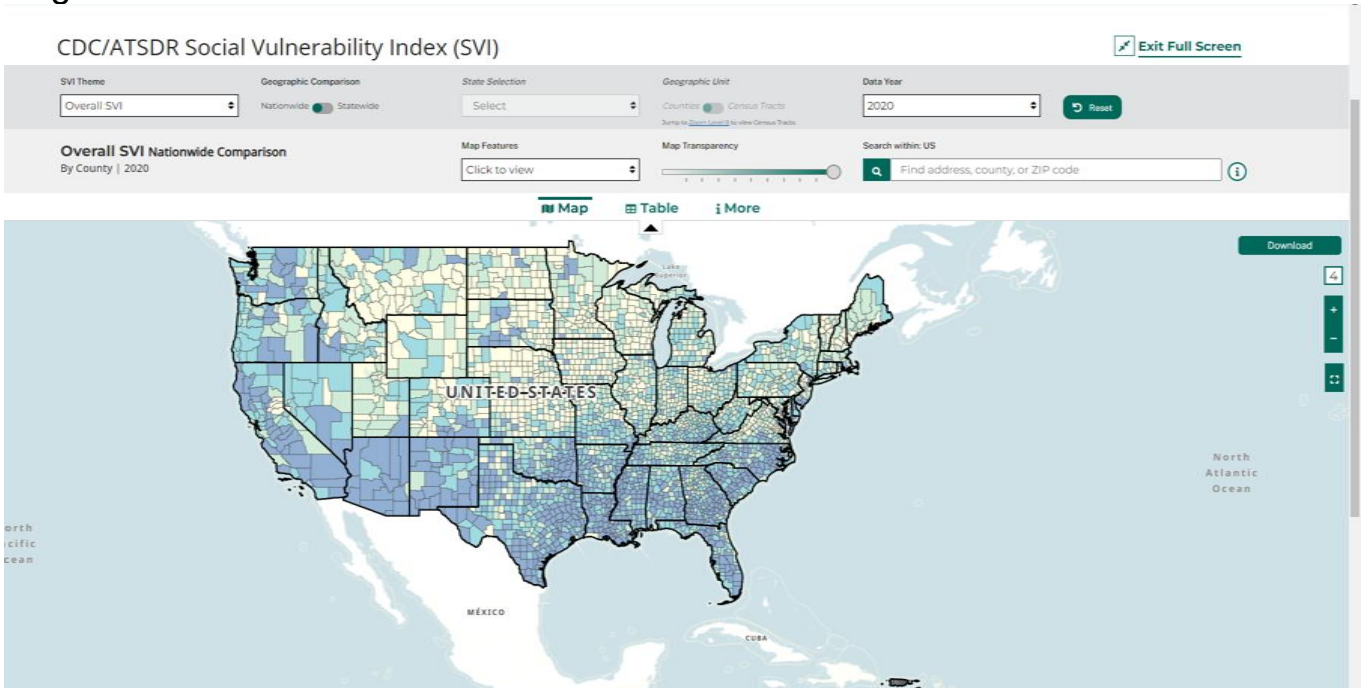
Natural Disasters, Health, and GIS

- Presentation



Natural Disasters, Health, and GIS

- Accessibility?
- Matching stakeholder with access



Natural Disasters, Health, and GIS

- Summary
  - Social Vulnerability Index (SVI)
    - Source
    - Transformations
    - Output
- Activity
  - Given a spatially relevant problem of your choosing, how would you use a data model to solve it?

Key Points to Remember

Source:

- What are the relevant factors?
- What are the data sources for these factors?

Transformation:

- How can you combine to relevant factors?

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