Before we start

Schedule:

- Lecture
- Coding Lab

For the coding lab, please install R and R studio on your computer:

- https://posit.co/download/rstudio-desktop/
- R version >= 4.2.0

Code, data and presentation are all available from:

- https://github.com/wangzhan90/2023DSTEM

The Magic of Where:
How Spatial Data Strengthen Insights into the World

Zhan Wang

Department of Agricultural Economics

Purdue University

June 21, 2023

PURDUE UNIVERSITY | College of Agriculture

Data science in my view

The science of how to gain insight from data

Data ≠ Insights

- Number, text, image, …
- Static, temporal, spatial …
- Observation, experiment …

- Understanding
- Relationship / causality
- Implication

Example #1

Suppose you are a data scientist. During the pandemic of an unknown disease, your job is to analyze the data of patients and develop a disease control and prevention plan.

ID	Patient	Age	Profession	Date reported	Address
1	Mr. Davis	24	Student	June 3rd	101 Maple Lane
2	Ms. Smith	32	Nurse	June 5th	123 Main Street
3	Mr. Jones	45	Worker	June 12th	456 Elm Avenue
4	Ms. Lee	28	Accountant	June 19th	789 Pine Road
5	Mr. Miller	52	Salesman	June 26th	202 Oak Street

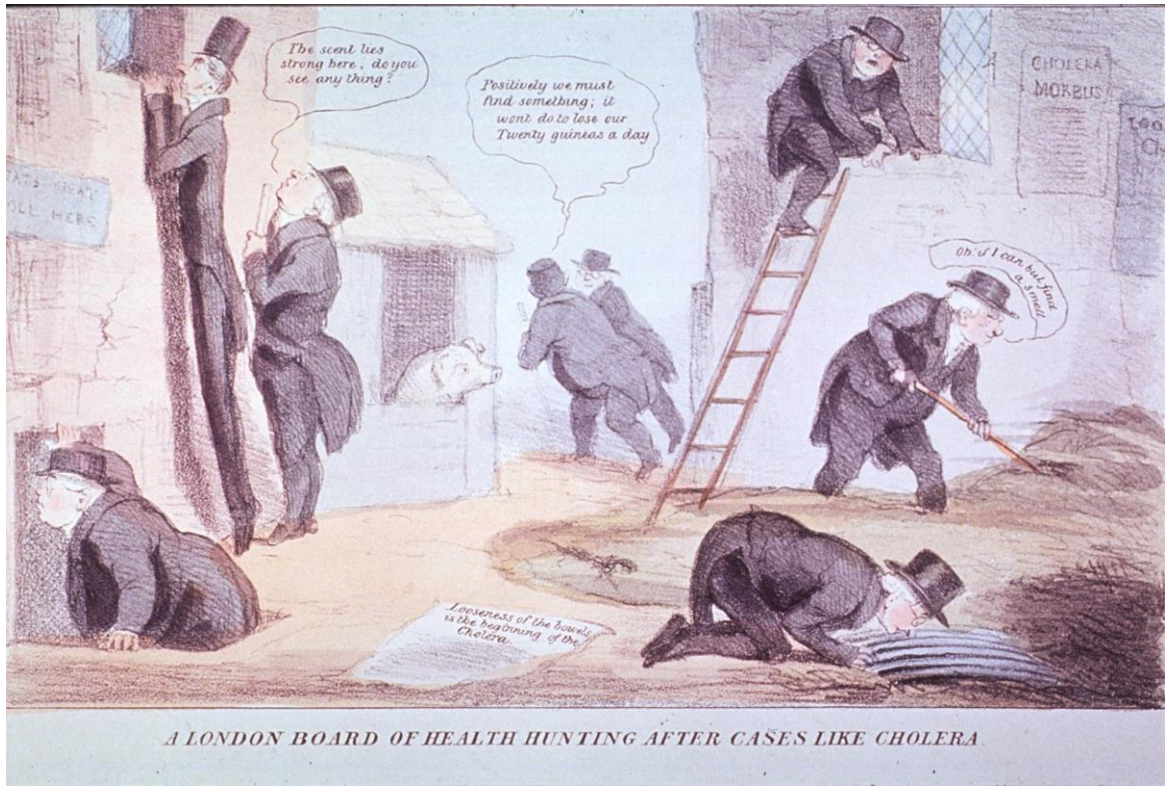
Questions:

- What data do you get? What insight do you need?
- How can you gain insight from data?

Story behind the example

This example is based on a real-world story: the cholera outbreak in London in the nineteenth century.

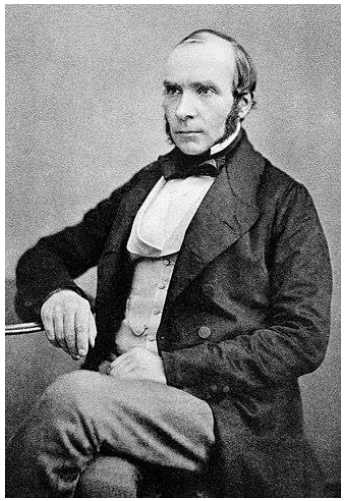
Before the bacteria causing cholera was found, people knew almost nothing about how the cholera is transmitted and how to control its outbreak.



"A London Board of Health Hunting After Cases Like Cholera" by Robert Seymour (1832)

Source: http://resource.nlm.nih.gov/101393365

Story behind the example

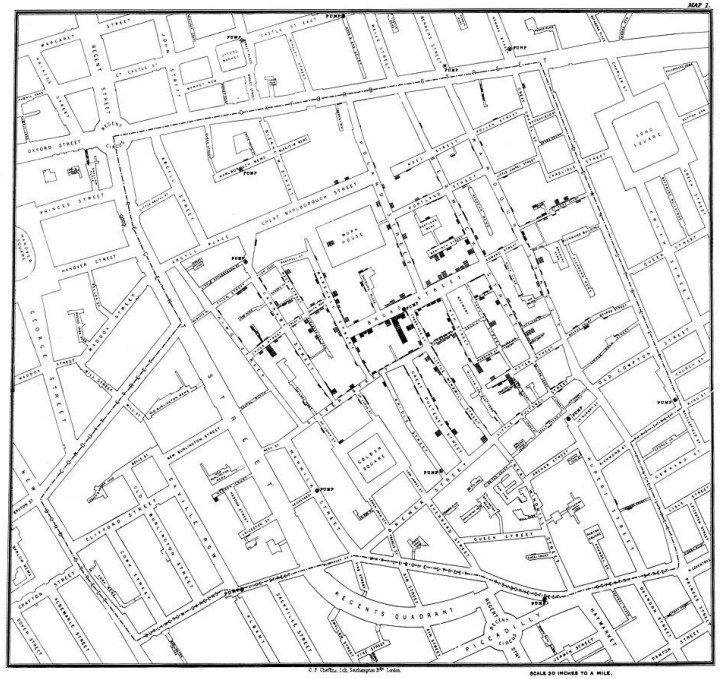


John Snow
(1813 – 1858)

John Snow, an English physician, introduced spatial data analysis into epidemiology

By mapping patients' address, he idented the center of disease cluster is the Broad Street pump, and concluded cholera is transmitted with polluted water.

This insight helps to stope the cholera pandemic and save lives, even before we identify the bacteria with modern medical researches!



Source: https://en.wikipedia.org/wiki/John_Snow

Take-home messages of example 1

Spatial features can help us to gain insight from data, because they provide information on:

1) A better understanding of data on geographic level

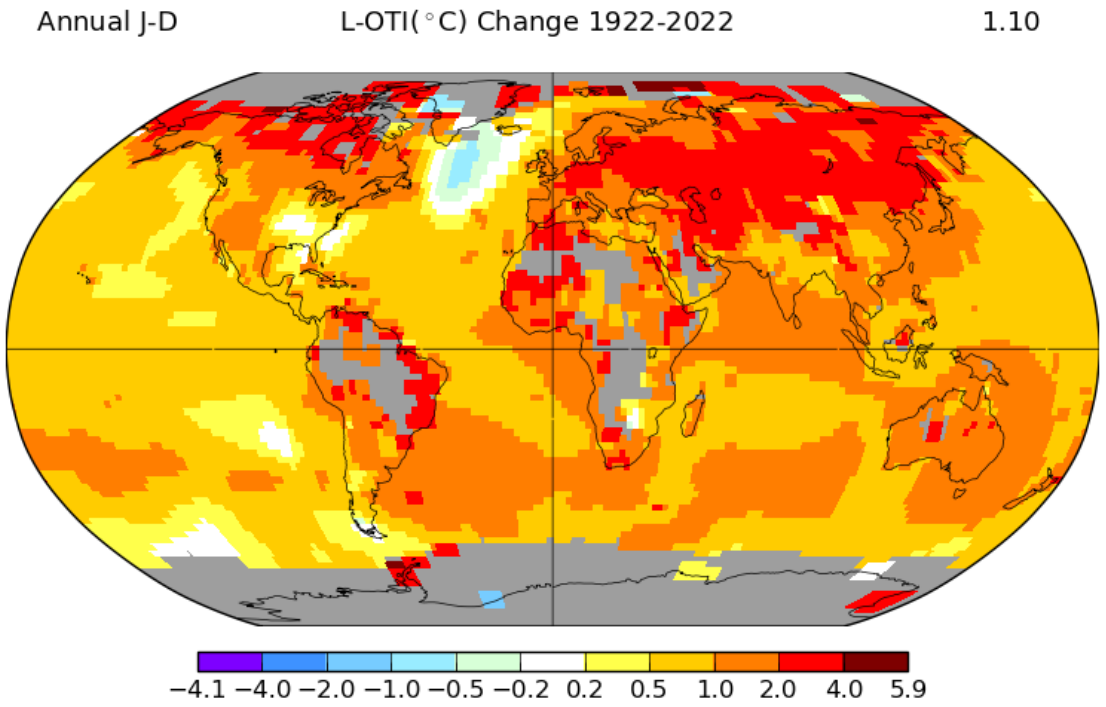
2) Relationships from spatial closeness or distribution, which may further imply causality

3) Implications on real-world actions / interventions

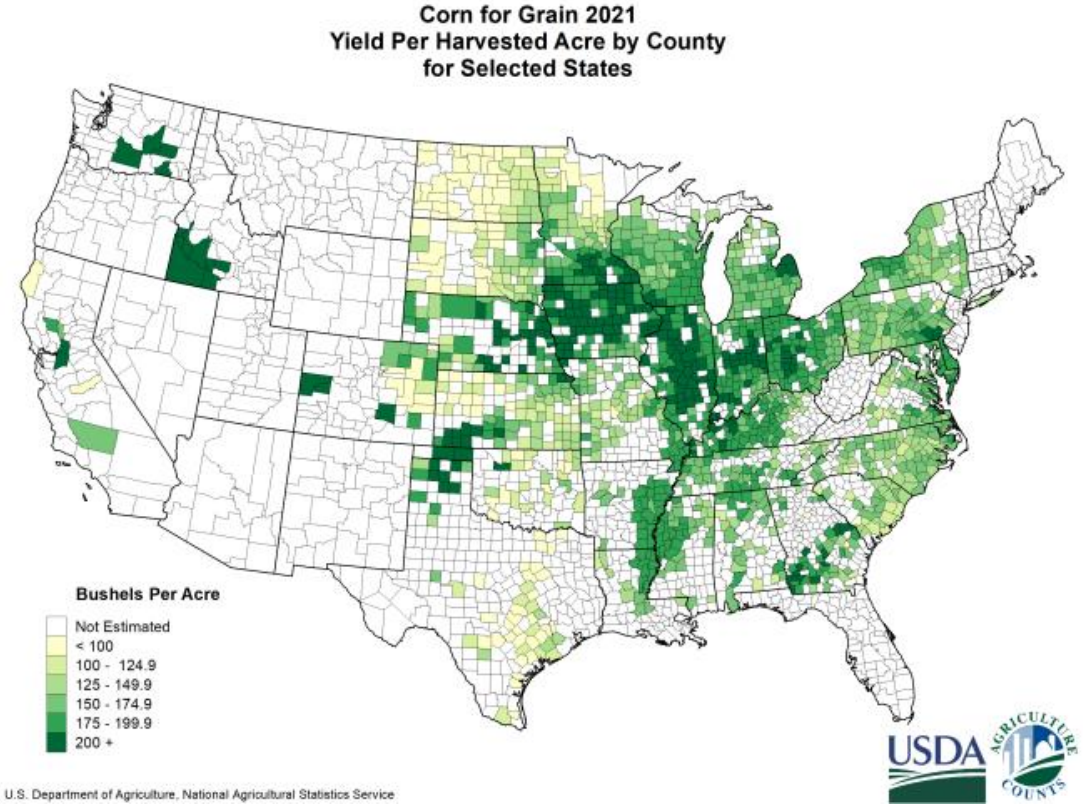
Example #2

Suppose you are a data scientist. USDA ask you to analyze climate change's impact on corn production by 2050.

You need to work with spatial data, because both the impact of climate change and corn yield vary by locations.



Source: https://data.giss.nasa.gov/gistemp/maps/index_v4.html



Source: https://www.nass.usda.gov/Charts_and_Maps/Crops_County/cr-yi.php

Coding lab

- In the coding lab, I will guide you to use R to process spatial data for this task (taking Indiana as an example)

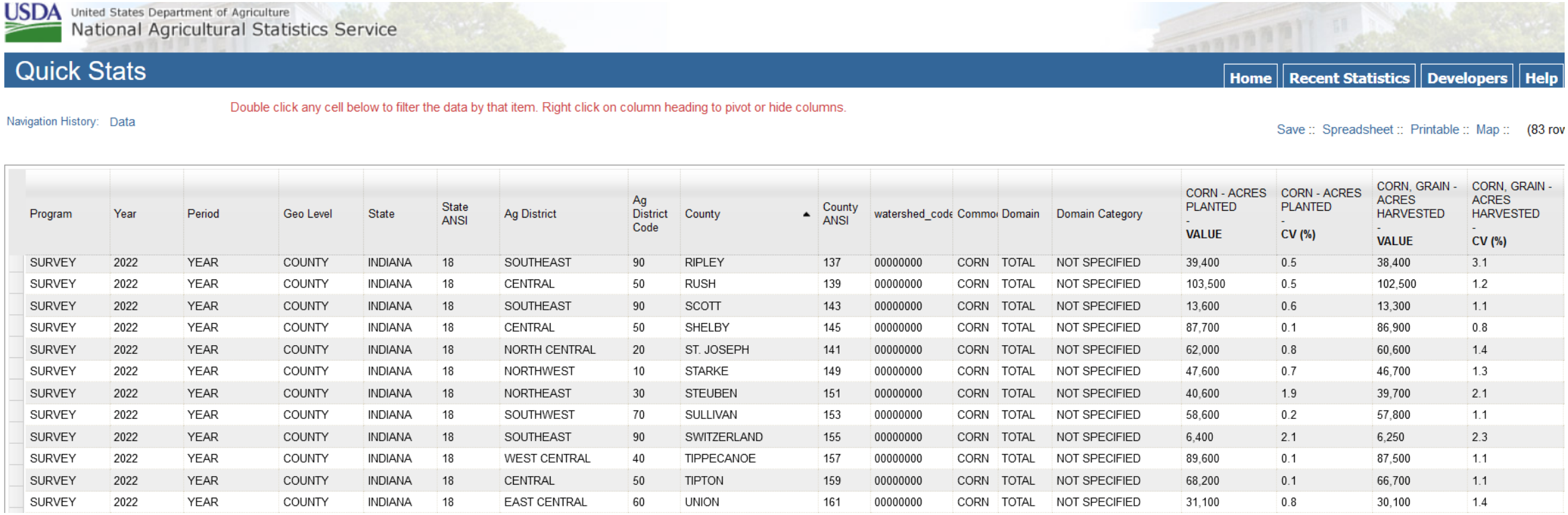
- Why R?
  - Free and open-source
  - A large community
  - Powerful packages for analyzing geodata

Skills we will learn

- Loading and pre-treatment of all three major formats of data
  - Table / Shapefile / Raster
- Merging table and shapefile
- Raster calculation
- Zonal statistics
- Visualization of spatial data

Data sources

- County level corn yield data in 2022 from USDA
  - A Table of data
  - Data source: https://www.nass.usda.gov/Statistics_by_State/



USDA United States Department of Agriculture
National Agricultural Statistics Service

Quick Stats

Home | Recent Statistics | Developers | Help

Double click any cell below to filter the data by that item. Right click on column heading to pivot or hide columns.

Navigation History: Data

Save :: Spreadsheet :: Printable :: Map :: (83 rov

Program	Year	Period	Geo Level	State	State ANSI	Ag District	Ag District Code	County	County ANSI	watershed_code	Commo	Domain	Domain Category	CORN - ACRES PLANTED - VALUE	CORN - ACRES PLANTED - CV (%)	CORN, GRAIN - ACRES HARVESTED - VALUE	CORN, GRAIN - ACRES HARVESTED - CV (%)
SURVEY	2022	YEAR	COUNTY	INDIANA	18	SOUTHEAST	90	RIPLEY	137	00000000	CORN	TOTAL	NOT SPECIFIED	39,400	0.5	38,400	3.1
SURVEY	2022	YEAR	COUNTY	INDIANA	18	CENTRAL	50	RUSH	139	00000000	CORN	TOTAL	NOT SPECIFIED	103,500	0.5	102,500	1.2
SURVEY	2022	YEAR	COUNTY	INDIANA	18	SOUTHEAST	90	SCOTT	143	00000000	CORN	TOTAL	NOT SPECIFIED	13,600	0.6	13,300	1.1
SURVEY	2022	YEAR	COUNTY	INDIANA	18	CENTRAL	50	SHELBY	145	00000000	CORN	TOTAL	NOT SPECIFIED	87,700	0.1	86,900	0.8
SURVEY	2022	YEAR	COUNTY	INDIANA	18	NORTH CENTRAL	20	ST. JOSEPH	141	00000000	CORN	TOTAL	NOT SPECIFIED	62,000	0.8	60,600	1.4
SURVEY	2022	YEAR	COUNTY	INDIANA	18	NORTHWEST	10	STARKE	149	00000000	CORN	TOTAL	NOT SPECIFIED	47,600	0.7	46,700	1.3
SURVEY	2022	YEAR	COUNTY	INDIANA	18	NORTHEAST	30	STEUBEN	151	00000000	CORN	TOTAL	NOT SPECIFIED	40,600	1.9	39,700	2.1
SURVEY	2022	YEAR	COUNTY	INDIANA	18	SOUTHWEST	70	SULLIVAN	153	00000000	CORN	TOTAL	NOT SPECIFIED	58,600	0.2	57,800	1.1
SURVEY	2022	YEAR	COUNTY	INDIANA	18	SOUTHEAST	90	SWITZERLAND	155	00000000	CORN	TOTAL	NOT SPECIFIED	6,400	2.1	6,250	2.3
SURVEY	2022	YEAR	COUNTY	INDIANA	18	WEST CENTRAL	40	TIPPECANOE	157	00000000	CORN	TOTAL	NOT SPECIFIED	89,600	0.1	87,500	1.1
SURVEY	2022	YEAR	COUNTY	INDIANA	18	CENTRAL	50	TIPTON	159	00000000	CORN	TOTAL	NOT SPECIFIED	68,200	0.1	66,700	1.1
SURVEY	2022	YEAR	COUNTY	INDIANA	18	EAST CENTRAL	60	UNION	161	00000000	CORN	TOTAL	NOT SPECIFIED	31,100	0.8	30,100	1.4

Source: https://www.nass.usda.gov/Statistics_by_State/Indiana/Publications/County_Estimates/index.php

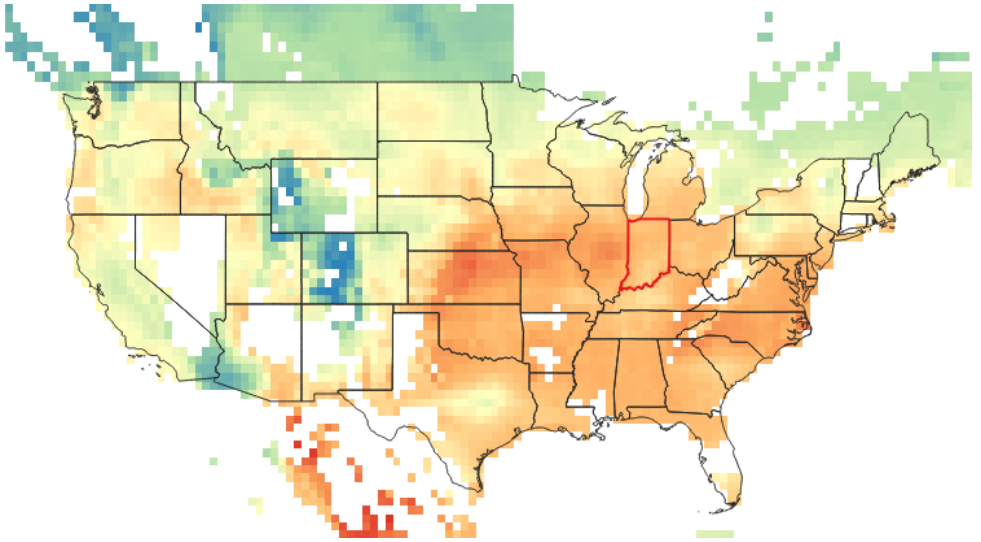
Data sources

- County level corn yield data in 2022 from USDA
- County boundary and location in Indiana
  - A shapefile
  - Shapefile: the vector data format (point, polyline, polygon) for geographic objects
  - Data source: https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

Source: https://www.stats.indiana.edu/maptools/boundary.asp

Data sources

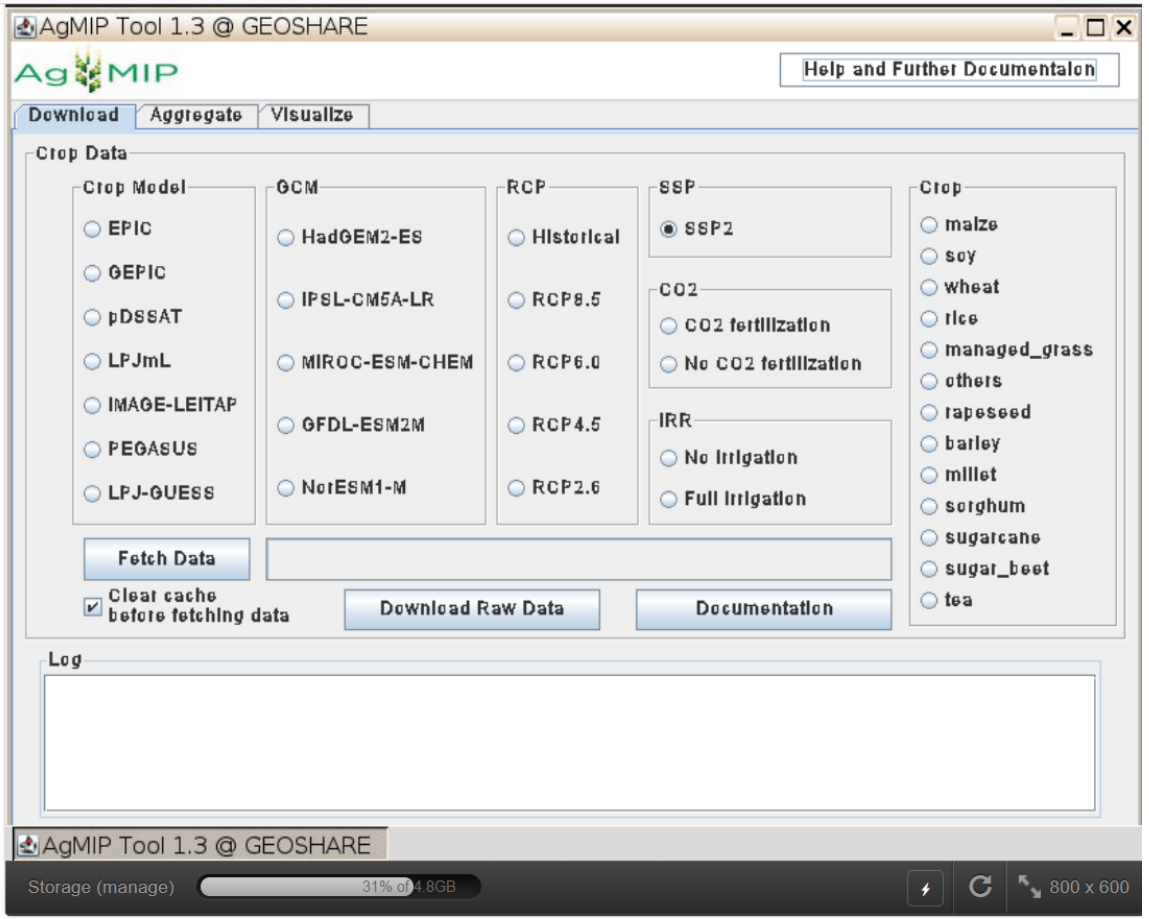
- County level corn yield data in 2022 from USDA
- County boundary and location in Indiana
- Projected yield from climate and crop model
  - A raster data
  - Raster: the gridded data for a given geographic location (longitude, latitude)
  - Data source: AgMIP Tool https://mygeohub.org/tools/agmip



About AgMIP Tool

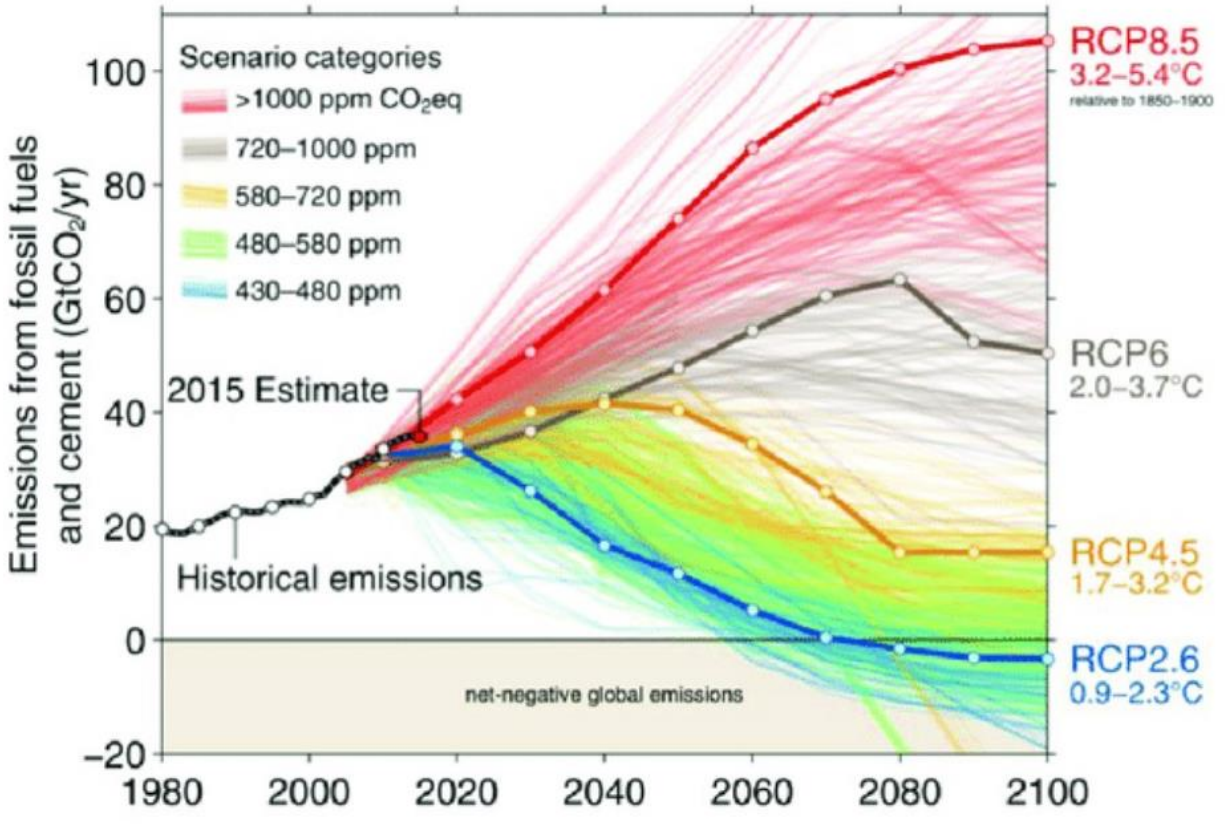
- Agricultural Model Intercomparison Project (AgMIP) Tool is a data accessing tool to obtain simulation results of climate change on crop yield

- Free access with MyGeoHub account
- Multiple Crop model and global climate models
- Here we used the results from:
  - Crop model: Environmental Policy Integrated Climate Model (EPIC)
  - Climate model: Hadley Centre Global Environment Model, version 2-Earth System (HadGEM2-Es)

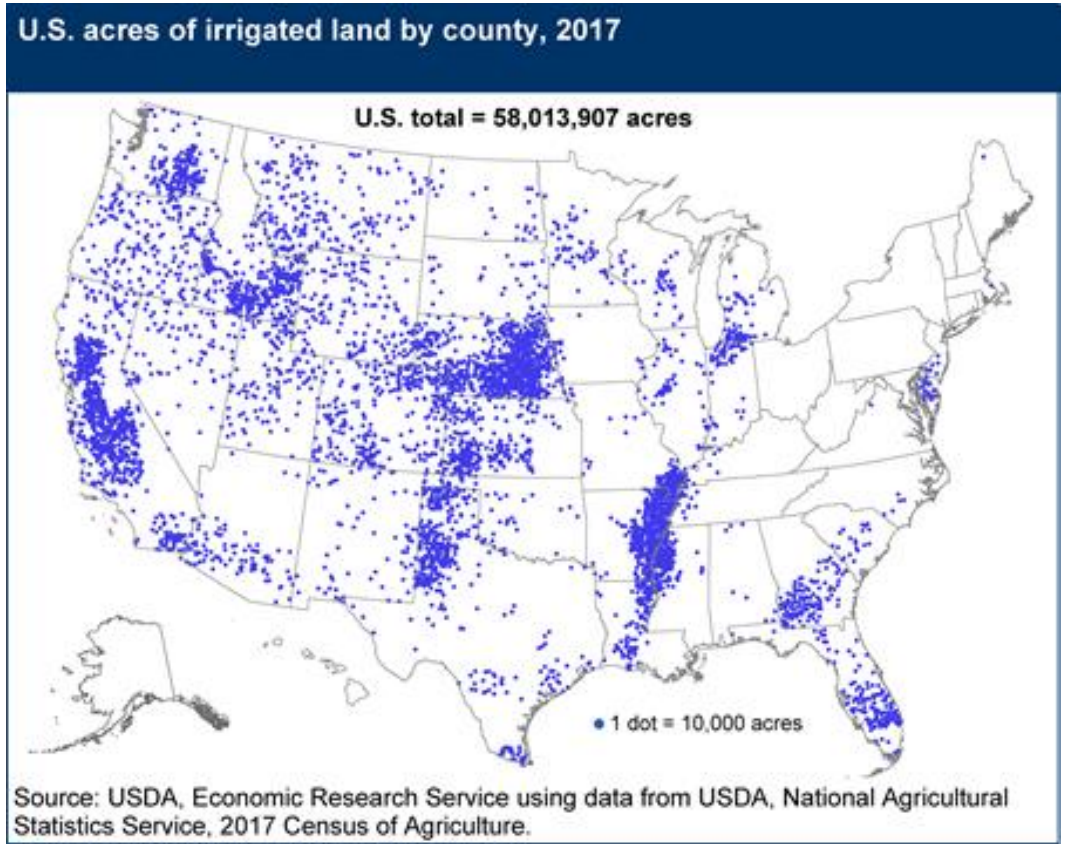


About AgMIP Tool

- The climate change scenario is RCP8.5, and I focus on the change of non-irrigated corn yield



Source: https://climatenexus.org/climate-change-news/rcp-8-5-business-as-usual-or-a-worst-case-scenario/



Source:https://www.ers.usda.gov/topics/farm-practices-management/irrigation-water-use/

Coding lab

- Now, let us switch to the hand-on coding lab!

- If you have not installed R and Rstudio on your computer, please download and install them from:

  https://posit.co/download/rstudio-desktop/

- Code, data and presentation are all available from:

  https://github.com/wangzhan90/2023DSTEM

Implications of example 2

- What have you learnt from this example?

- What implications do you have from this example?
  - Policy suggestions
  - Research questions

  - Hint 1: impacts on stakeholders from different sectors
  - Hint 2: impacts from outside of the US

Implications of example 2

- What have you learnt from this example?

- What implications do you have from this example?

Agriculture and the environment

Domestic food supply and food price

Farmer's response (land expansion, switching to irrigated farming, increase input use …)

Land use conversion and greenhouse gas emission

Non-agricultural sectors and welfare

Production of sectors using corn as input

Impacts on daily life

Global impacts

Corn supply and demand from other countries

Global food trade

THANK YOU!

Zhan Wang

Department of Agricultural Economics

College of Agriculture, Purdue University, USA

zhanwang@purdue.edu

PURDUE UNIVERSITY | College of Agriculture