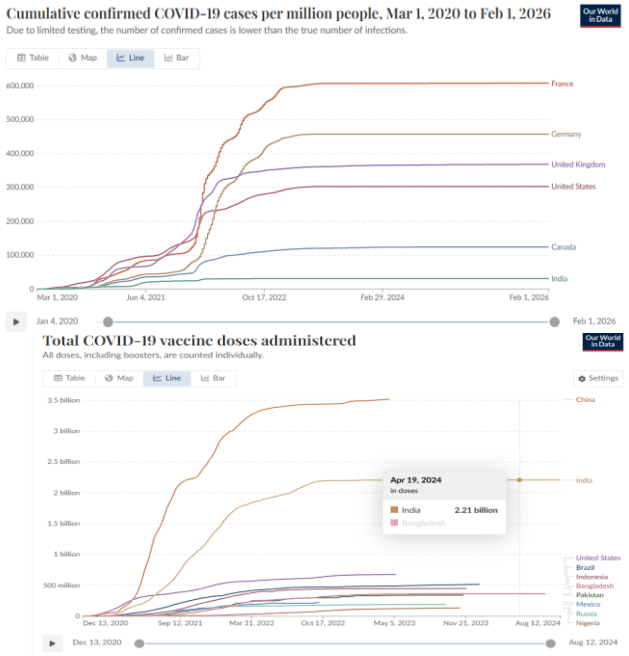
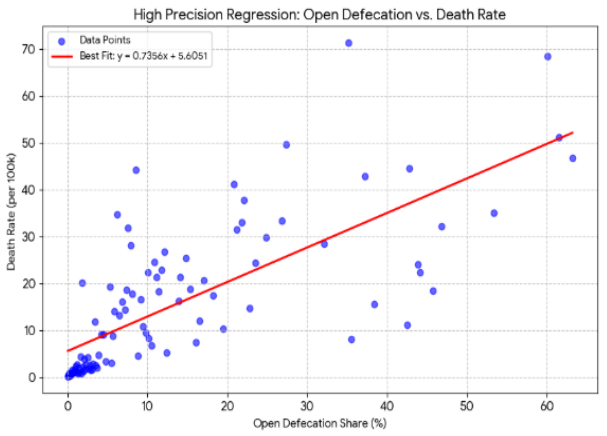
Contagious Disease Model: Visual Impact of Pathogen Spread on Selected Populations

Current disease simulation models do not accept input parameters; only modeling based on density

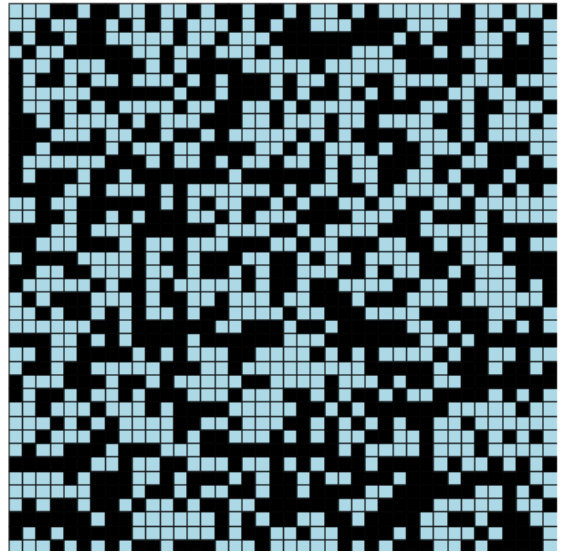
Public epidemiolocal data can be analyzed to find connections between factors of disease and disease spread



That data can be used to make graphs, best fit lines, and regressions of disease factors and disease spread



These correlations can be inputted into the model to show the effect of disease factors on disease spread over time

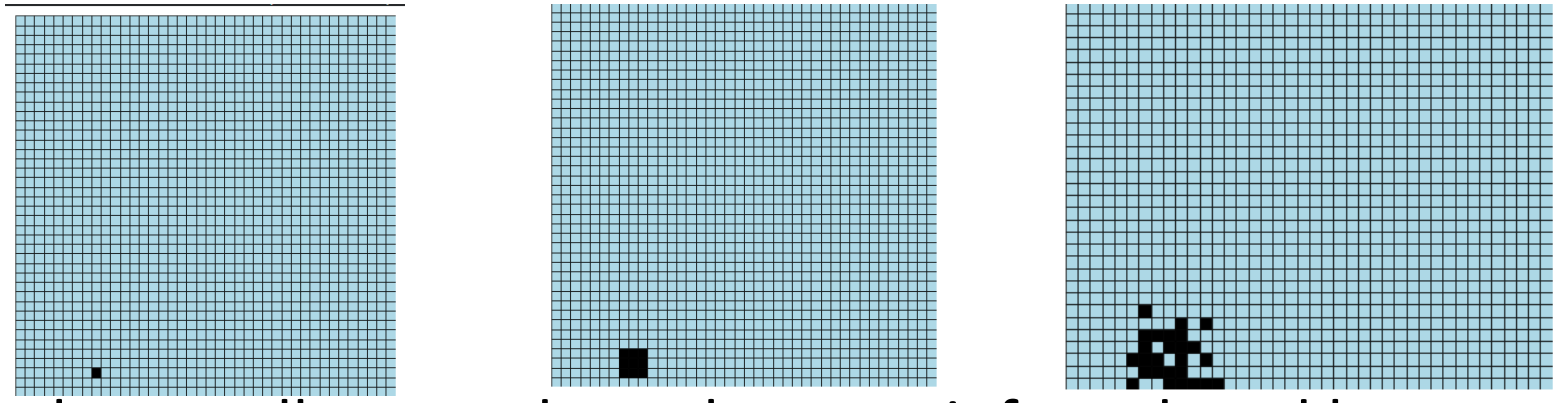


Question

- How effectively can a code-based model predict the spread of disease based on input parameters?
- Changing each parameter should change the model's spread a certain way
  - Connections should be made based on correlations in epidemiological data (how strongly does each factor influence the spread of disease)
  - Eg: How vaccination/sanitation factors influence disease spread

Investigation

- Python grid of 1600 cells (40 X 40), begins with 1 random infected cell



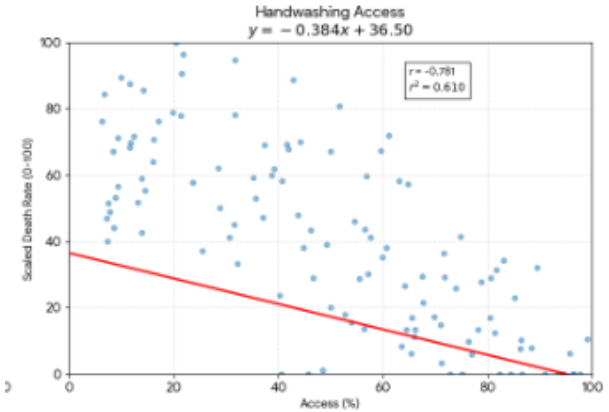
- From there, cells around may become infected, and have a chance to recover or stay infected/die
  - Cells affected in a radius of 1-3 with radius to infected cell $\propto$ (1/likeliness to catch a disease) (radius of 4+ is negligible)
- 3 mathematical determinants: likeliness to catch, likeliness to revive, and vaccination

Factors of Disease: Sanitation

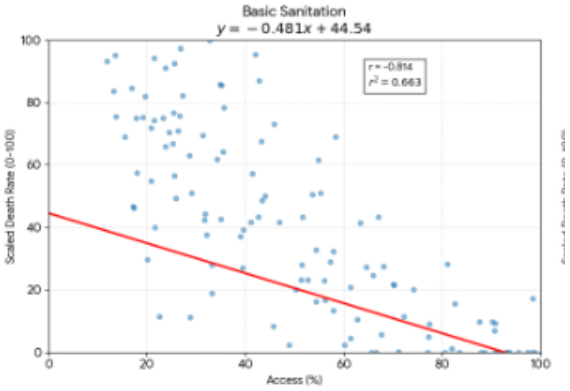
- Primary Method: best fit lines and linear regression analysis
  - 4 Sanitation determinants: determinant index vs deaths from unsafe sanitation
    - Unsafe sanitation deaths implies likeliness to spread pathogen

- Open Defecation: excretion outside toilets (eg: water sources)
  - Positive correlation
- Y = 1.153x + 5.56
- R**2 = 0.709

- Access to handwashing facilities
  - Negative correlation
- Y = -0.384x + 36.5
- R**2 = 0.61



- Access to basic sanitation facilities
  - Negative correlation
- Y = -0.481 + 44.54x
- R**2 = 0.663



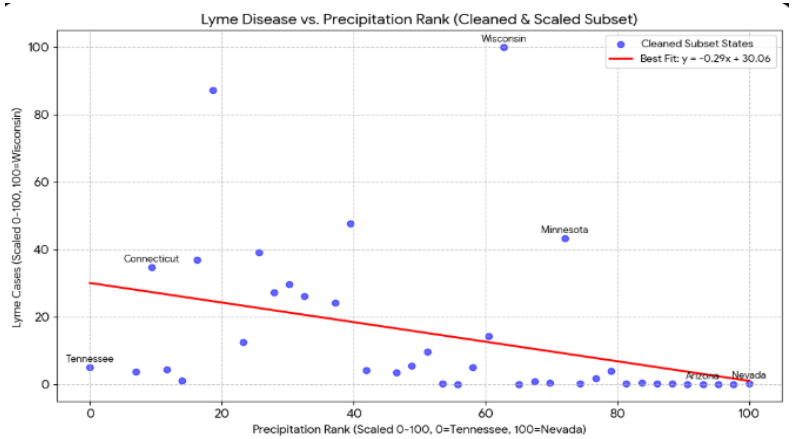
- Basic Water Access
  - Negative correlation
- Y = -0.77x + 73.97
- R**2 = 0.748
  - Strongest predictor

X = % of population, Y = percentile of death rate from unsafe sanitation (higher = more death)
0-100 scaling method
R**2 : data accuracy: how much does regression explain variation in data (0-1)

Factors of Disease: Climate

- Accounting for vector transmitted diseases as well (Eg; mosquitos)
- If vector based: Climate regression (rainfall vs cases)
  - Vectors are primarily based in tropical (mosquitos) or temperate regions (ticks); Climate predicts cases
  - Area's climate can predict abundancy of certain vector-borne diseases
  - Temperate vs Tropical based vectors: regressions: temperate diseases vs climate/tropical diseases vs climate

- <u>Temperate: Lyme Disease vs Rainfall</u>
  - Main temperate vector disease (ticks)
- Lyme cases vs Rain (Positive correlation)
- Y = -0.29x + 30.06
  - X is dryness percentile, Y is case percentile
- R**2 = 0.1262 (weak predictor)

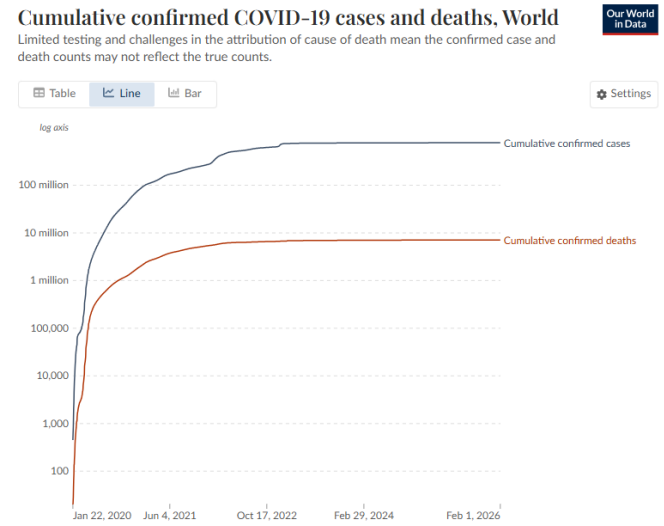


- <u>Tropical: Malaria vs Rainfall</u>
  - Main tropical disease vector (mosquitos)
- Malaria vs Rain (Positive correlation)
- Y = -0.3415x + 53.3
  - X is dryness percentile, Y is case percentile
- R**2 = 0.113 (weak predictor)

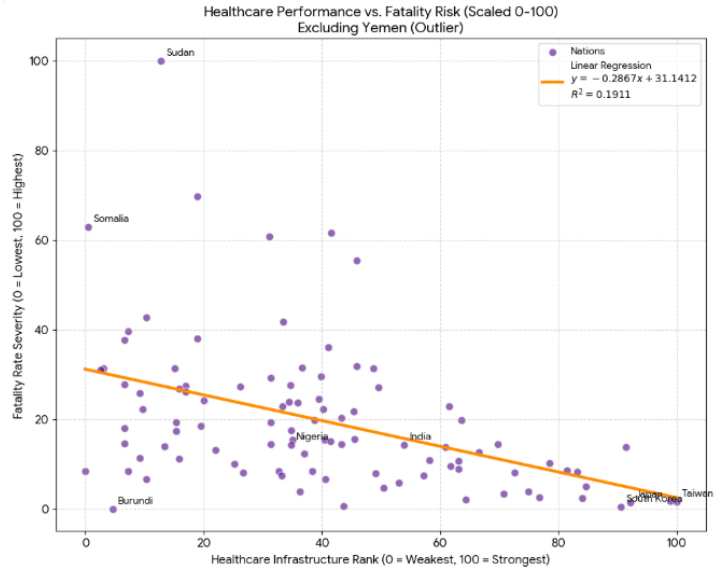
- Conclusion: rainfall weakly predicts abundancy of temperate/tropical vector borne diseases
  - Temperature was a weaker predictor of rainfall
  - Future studies: more in-depth climate analysis

Likeliness to Revive

- Much less dependent on external factors
- Predicted mainly by type of disease/random host factors health, etc.)/vaccination (see p.8)
  - Model inputs all of these



- Last predictor: Healthcare
  - Cannot use disease spread for regressions, use death rate
    - Healthcare changes likeliness to die
  - Synthesize COVID data death rates
    - Cumulative cases and deaths for every nation
  - Hypothesis: For every nation:
    - Healthcare $\propto$ (1/(Covid deaths/Covid cases))
- Healthcare vs. Death Rate (Negative correlation)
  - Y = -0.2867x + 31.1412
  - R**2 = 0.1911
  - X = Healthcare percentile, Y = Death Rate percentile
  - R**2 ~ 0.2: only 1/5 variation explained (low): healthcare cannot do much to decrease affliction with common pathogen until vaccine is implemented (best way to decrease death rate)



Likeliness to Revive

- Randonized factor: Expressed as random integer (1, 100)(~40% determination)
- Also includes vegetable consumption; predicts likelihood to live in ~2% of data
  - Given very low weight, but interesting factor in influencing recovery
- Death rate: variable factor
  - Death rate measures cells that aren't revived
  - In reality, some cells that will revive will persist being infected for multiple cycles before they recover, so death rate must expand to represent this
  - Death rate calculation: death rate X (time to recover / cyclelength)
    - Eg: 10 days to recover and a cycle length of 5 means that each iteration should have double the death rate to represent cells that are infected and will recover, but aren't healthy yet

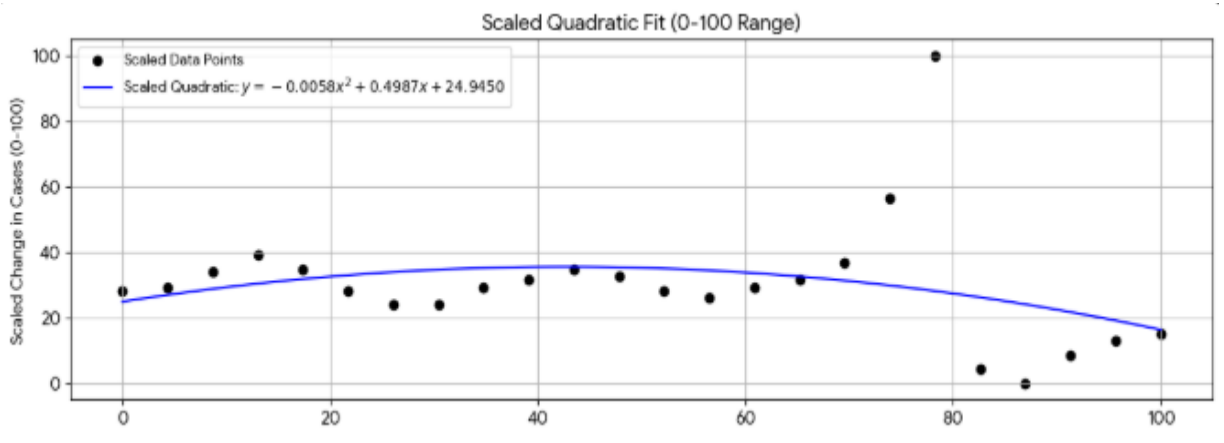
Final Reviving factors:

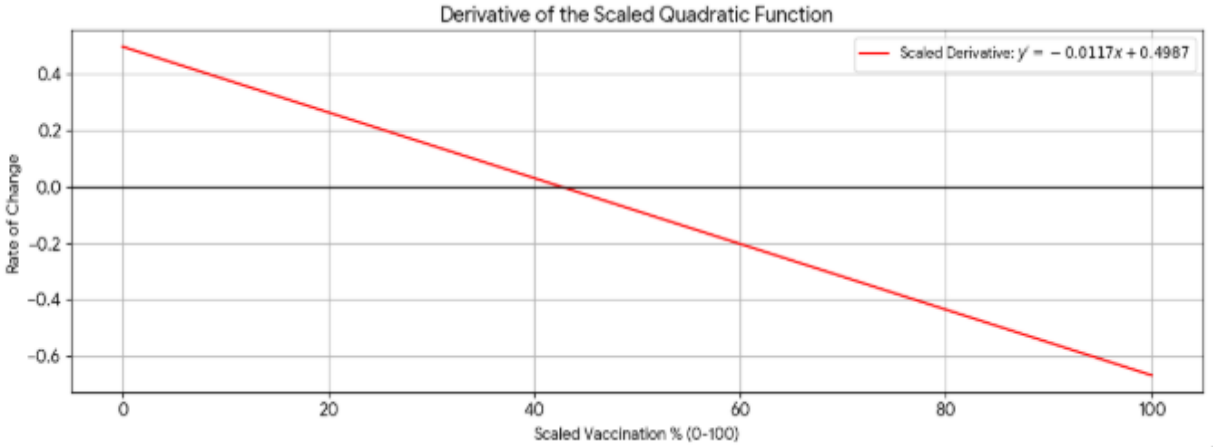
Healthcare =	((-0.2867 * healthcareindex) + 31.1412) * 0.191
Random =	random.randint(0, 100)
Death Rate =	deathrateindex * (recovery / cyclelength)
Vegetable Consumption =	((-0.1575 * vegetableindex) + 36.64) * 02.29

Vaccination

- Does not use regression calculation.
- First attempt: use derivative of change in vaccination vs change in Covid cases to represent the change in cases at each percentage of people vaccinated (tangent line)
  - Problem: data skewed by seasonal variations, new COVID variants, and uniqueness of COVID vaccine (other vaccines may be more or less effective)

Quadratic fit of COVID vaccination percentages and change in cases (R**2 ~ 0.42)accompanied with linear derivative

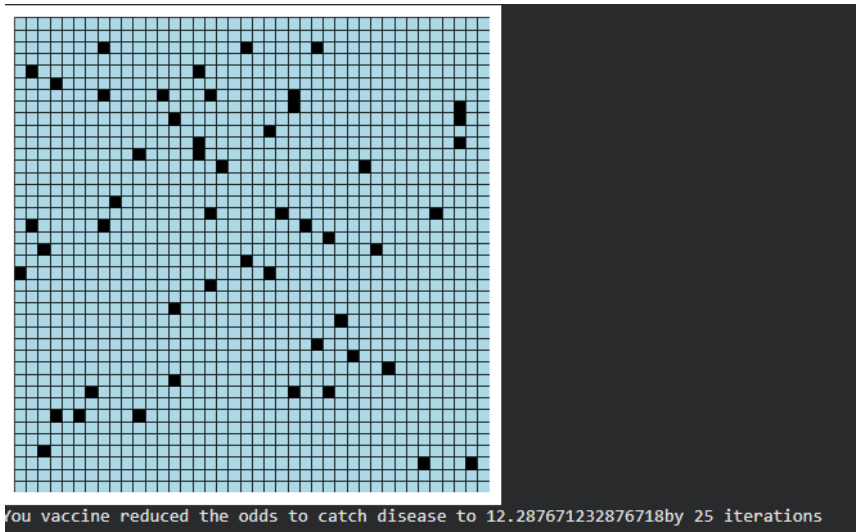




Vaccination

- End Vaccination calculation: user inputs vaccination effectiveness and yearly vaccination rates
- Vaccination updated every iteration with increasing vaccination rates
- Temporal adjustment: each cycle represents time to transmit disease from when disease is caught
  - Shorter cycles: lower vaccination change per cycle
  - Recommended input: Each iteration represents 10-15 cycles to properly show effect of vaccination
    - End of code running: shows percent of population vaccinated after 25 cycles

```
How much does vaccine reduce likliness to get disease (Example: Covid reduced likliness by 95%): 2
Generally, how much more percent of the world's population was vaccinated per year (Example: Covid was 16% at max): 1
```



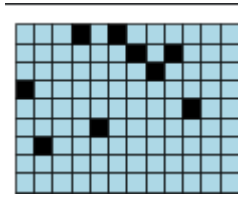
Procedure and Time Management

- Data plugged into code and model created
- Random cell infected, user inputs parameters
  - Likeliness to catch disease and likeliness to revive runs for each specified cycle
- Time Variables
  - Cycle Length (time to transmit disease once infected) adjusts for longer diseases and can specify multiple iterations of the code)
    - Essential for defining vaccination rates
  - Time to Recover estimates time to recover once infected
    - Essential to calculate what cells look infected (still black) when actually recovering, increasing the likelihood to die function

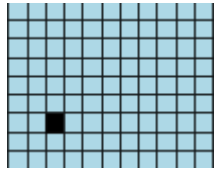
Results and Interpretation

- Likelihood to Catch: all variables are strong predictors of vaccination spread (high R**2 values); combine to make reliable function
- Likelihood to Die: highly dependent on individual disease characteristics/randomness
- Strongest Variables: access to basic water facilities, open defecation rates, disease death rates

Model after 6 iterations: access to basic water index of 15 (right)



Identical model after 6 iterations: access to basic water index of 75 (right)



- Weakest Variables: population density, vegetable consumption
  - Population density is traditional predictor for traditional disease models. However, all population density regression yielded very low R**2 values with almost horizontal slopes (poor predictor)
    - Suggests policy and societal conditions have become more important drivers for disease spread

Model Accuracy Analysis

- High Accuracy: high catch rates (for areas with high catch rates, the model predicts spread well)(Eg: Peru)
- High Accuracy: High density cities with high open defecation and low water access rates (Eg: Urbanized Indian regions)
- High Accuracy: High sanitation (correctly models resulting low spread rates)
- Low Accuracy: Government Policy (masking mandates/low enforcement rates from government influence spread (poor predictor for US/China)(low/high enforcement)
- Low Accuracy: Rural areas (smaller towns tend to have lower disease spread than the model predicts for the nation)
- Low Accuracy: Areas with high interregional travel rates (spread to isolated areas in nation)

High Prediction (left)

Implications/Future Research

- A model can accurately predict disease spread in areas with neutral governments, or more extreme sanitation rates (high/low)
- Future Modeling:
  - Model government intervention/policy
  - Model congestion/contact within rural areas
  - Model interregional contact within nations
  - Instead of raising death rate to model elongated recovery time, keep death rate the same, but specify a recovery period for each cell
  - Model changing vaccination rates every year
  - Model new strains of viruses
  - Model age demographics in a nation

Sources

- COVID-19 Vaccine Doses Administered." *Our World in Data*, ourworldindata.org/grapher/cumulative-covid-vaccinations
- International Citizens Insurance. "Ranking Best Health Care Systems in the World by Country." *International Travel Insurance Group*, 2024, www.internationalinsurance.com/health/systems/.
- Our World In Data. "Average Annual Precipitation." *Our World in Data*, 2019, ourworldindata.org/grapher/average-precipitation-per-year.
- Our World in Data. "Coronavirus (COVID-19) Cases - Statistics and Research." *Our World in Data*, 2020, ourworldindata.org/covid-cases.
- Ritchie, Hannah, and Max Roser. "Clean Water and Sanitation." *Our World in Data*, 1 July 2021, ourworldindata.org/hygiene.
- Roser, Max, and Hannah Ritchie. "Malaria." *Our World in Data*, Nov. 2015, ourworldindata.org/malaria.
- "Share of People Practicing Open Defecation." *Our World in Data*, ourworldindata.org/grapher/people-practicing-open-defecation-of-population.

- "Share of the Population Using at Least Basic Sanitation Services." *Our World in Data*, 2024, ourworldindata.org/grapher/share-of-population-with-improved-sanitation-faciltities
- "Vegetable Consumption by Country 2024." *Worldpopulationreview.com*, worldpopulationreview.com/country-rankings/vegetable-consumption-by-country.