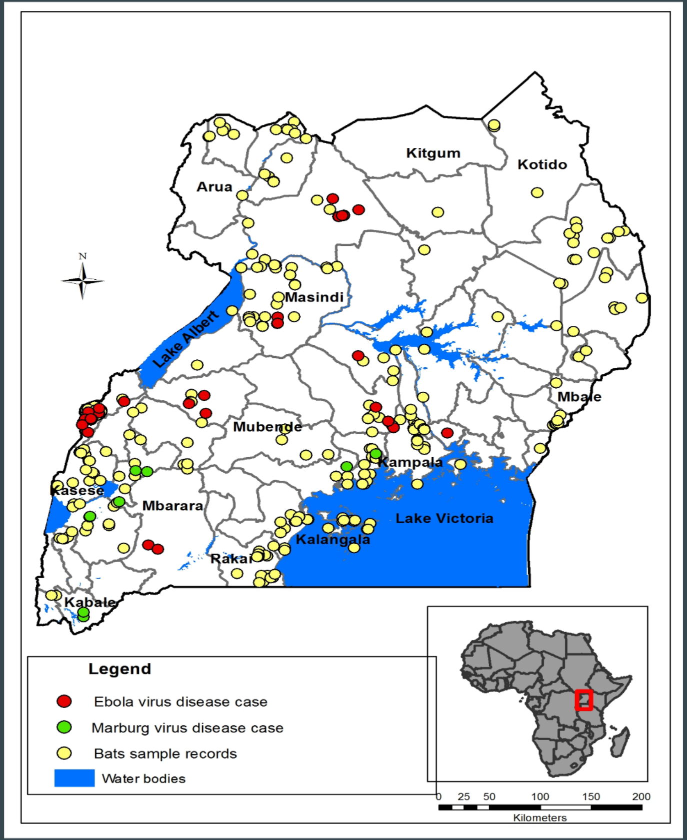
Disease Mapping

Mentee: Wuwei Zhang
Mentor: Taylor Okonek

Main Topics

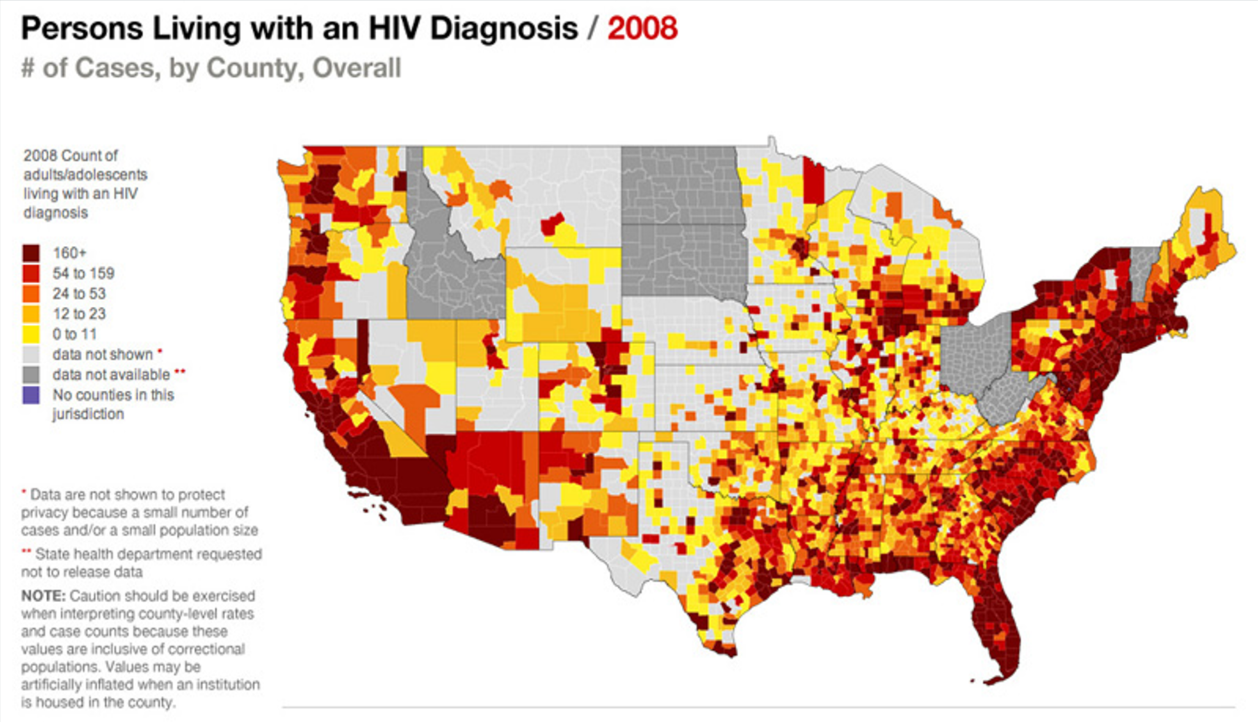
- Data types and maps: Unit-level vs. Area-level vs. Pixel-level, Proportion vs. Count
- Bayesian data analysis
- Uncertainty in maps

Data Types and Maps: Unit-level



Data Types and Maps: Area-level

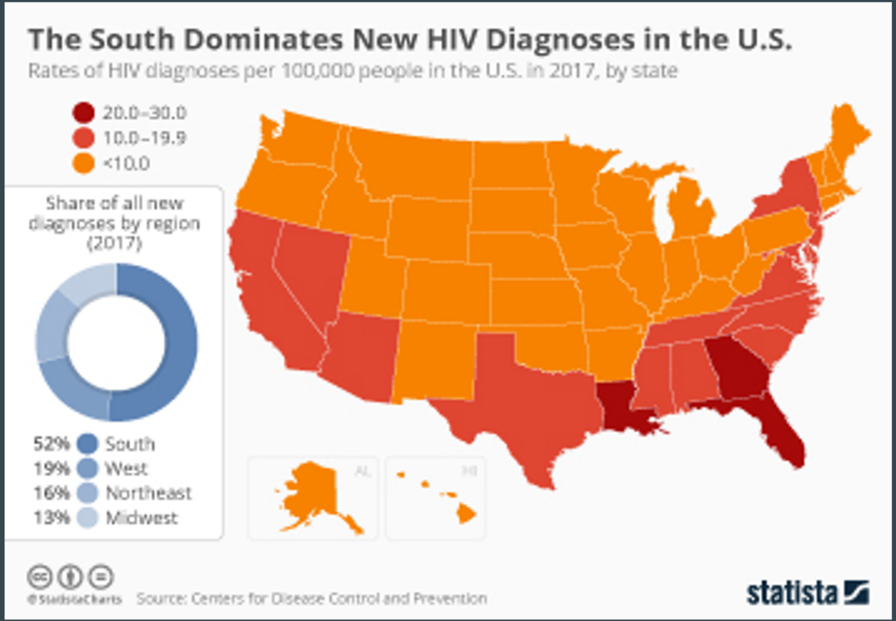
- Aggregation level: by county

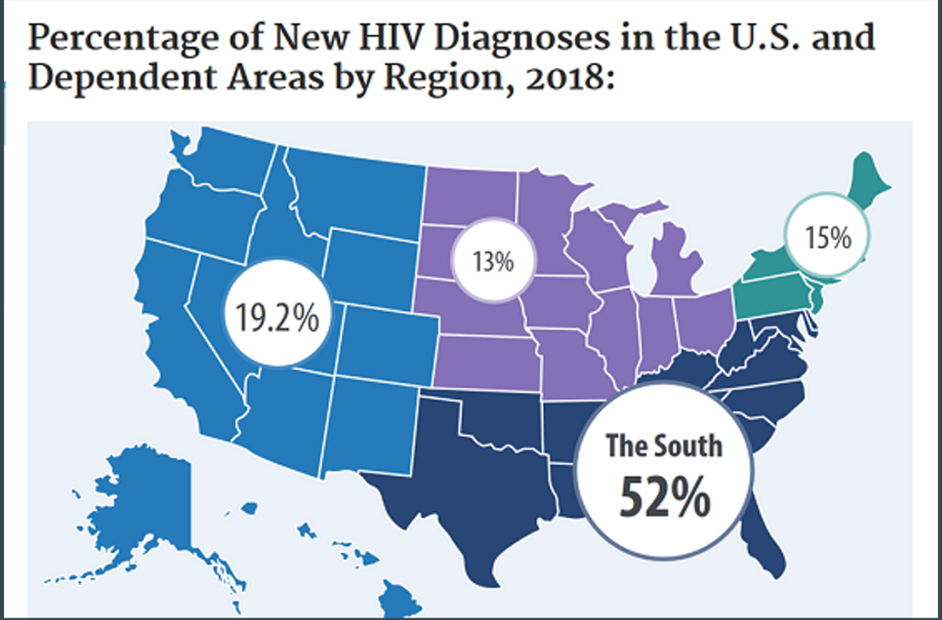


Data Types and Maps: Area-level

- Aggregation level: by state

Aggregation level: by region





Data Types and Maps

Unit-level vs. Area-level

Unit-level	Area-level
Benefit: Describes a set of points. We can know exactly where each case is located. Drawback: 1) Seems very hard to record every single case. 2) It is difficult to tell from the map what an area-level estimate would be. (simply looking at the map)	Benefit: User-friendly. Shows an overall prevalence of certain disease, i.e. which areas are at relatively higher risks. Drawback: 1) Cannot estimate disease at smaller levels of aggregation. 2) Also, need to consider the possible effect of the different population size.

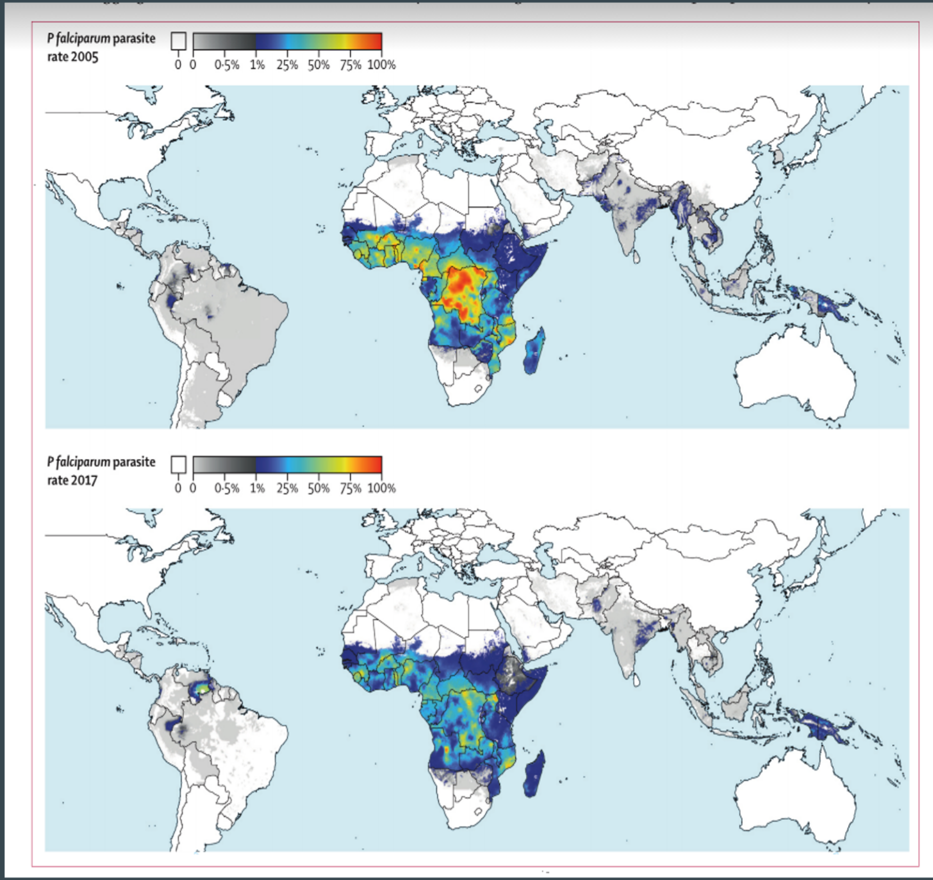
Data Types and Maps: Pixel-level map

- Need models if we want to fill in every pixel

Unit-level data/Cluster-level data → statistical models → pixel-level data

- Drawbacks:

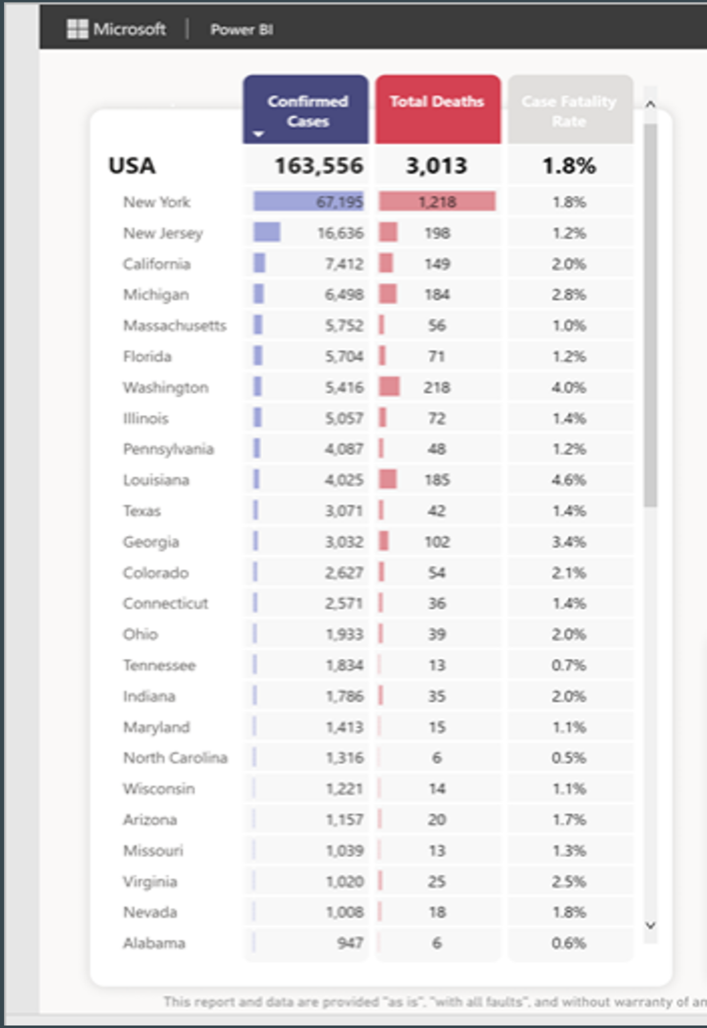
1) Has high uncertainty (because it is possible that many pixels don't have data).
2) Hard to visualize uncertainty at each pixel.



Data Types and Maps

Proportion & Count

Proportion	Count
Benefit: Takes the effect of population size into account. Drawback: Lose the information of exact number (may need to make public health policies based on exact number, e.g. health resource.	Benefit: Has access to population totals, which are sometimes of interest for policy Drawback: doesn't show the overall picture of the prevalence of disease, because different areas may have different land size and population size.



Microsoft | Power BI

	Confirmed Cases	Total Deaths	Case Fatality Rate
USA	163,556	3,013	1.8%
New York	67,195	1,218	1.8%
New Jersey	16,636	198	1.2%
California	7,412	149	2.0%
Michigan	6,498	184	2.8%
Massachusetts	5,752	56	1.0%
Florida	5,704	71	1.2%
Washington	5,416	218	4.0%
Illinois	5,057	72	1.4%
Pennsylvania	4,087	48	1.2%
Louisiana	4,025	185	4.6%
Texas	3,071	42	1.4%
Georgia	3,032	102	3.4%
Colorado	2,627	54	2.1%
Connecticut	2,571	36	1.4%
Ohio	1,933	39	2.0%
Tennessee	1,834	13	0.7%
Indiana	1,786	35	2.0%
Maryland	1,413	15	1.1%
North Carolina	1,316	6	0.5%
Wisconsin	1,221	14	1.1%
Arizona	1,157	20	1.7%
Missouri	1,039	13	1.3%
Virginia	1,020	25	2.5%
Nevada	1,008	18	1.8%
Alabama	947	6	0.6%

This report and data are provided "as is", "with all faults", and without warranty of an

Main Topics

- ✓Data types and maps: Unit-level vs. Area-level vs. Pixel-level, Proportion vs. Count
- Bayesian data analysis
- Uncertainty in maps

Bayesian data analysis

Frequentist vs. bayesian inference

Context: Most research in disease mapping uses Bayesian data analysis

Bayesian inference:

Models uncertainty by a probability distribution over prior belief (hypothesis), and then uses new data (likelihood) to update the distribution (posterior).

1. Use probabilities for both hypotheses and data.
2. Need to construct a "subjective prior".

$$\text{Posterior} = \frac{\text{Probability of the data} \times \text{Prior}}{\text{Average probability of the data}}$$

Main Topics

- ✓Data types and maps: Unit-level vs. Area-level vs. Pixel-level, Proportion vs. Count
- ✓Bayesian data analysis
- Uncertainty in maps

Uncertainty in maps

Modeling and presentation of vaccination coverage estimates using data from household surveys (by Tracy Qi Dong, Jon Wakefield)

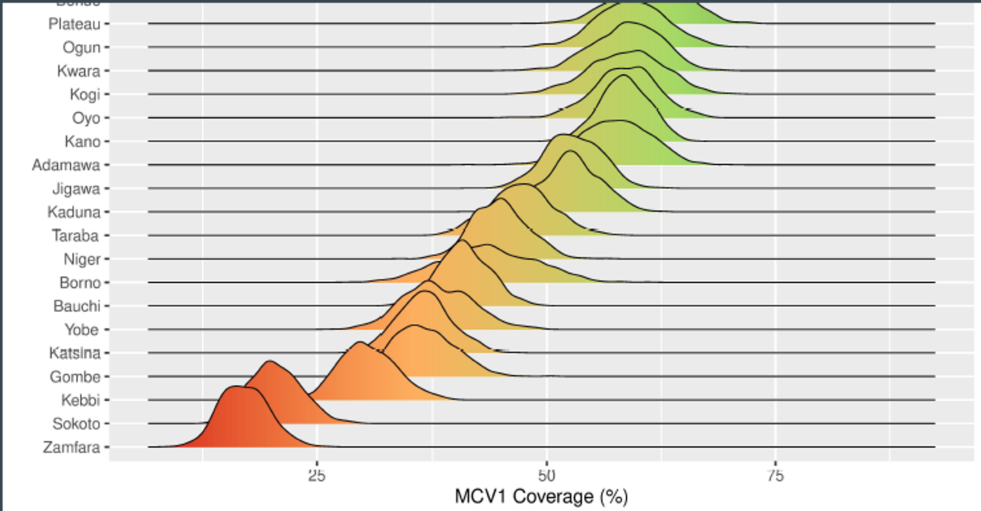
- To identify inequalities in vaccination coverage, visualize the uncertainty associated with the ranking

1) Uncertainty within state: spread of posterior distribution
2) Uncertainty between state: overlaps

<u>How</u>: Use ridgeline plots to show the posterior distribution of coverage estimate for each area.

<u>Conclusion</u>:

The last two states are separated from the rest.
Cannot really tell difference for most states.



663 Figure 5: Ridgeline plots of the posterior distributions of the MCV1 coverage estimates for
664 Nigeria's 37 states, ordered by posterior median, based on 1000 posterior samples from the
665 *Binomial TS* model that includes the urban/rural strata variable.

Uncertainty in maps

Modeling and presentation of vaccination coverage estimates using data from household surveys (by Tracy Qi Dong, Jon Wakefield)

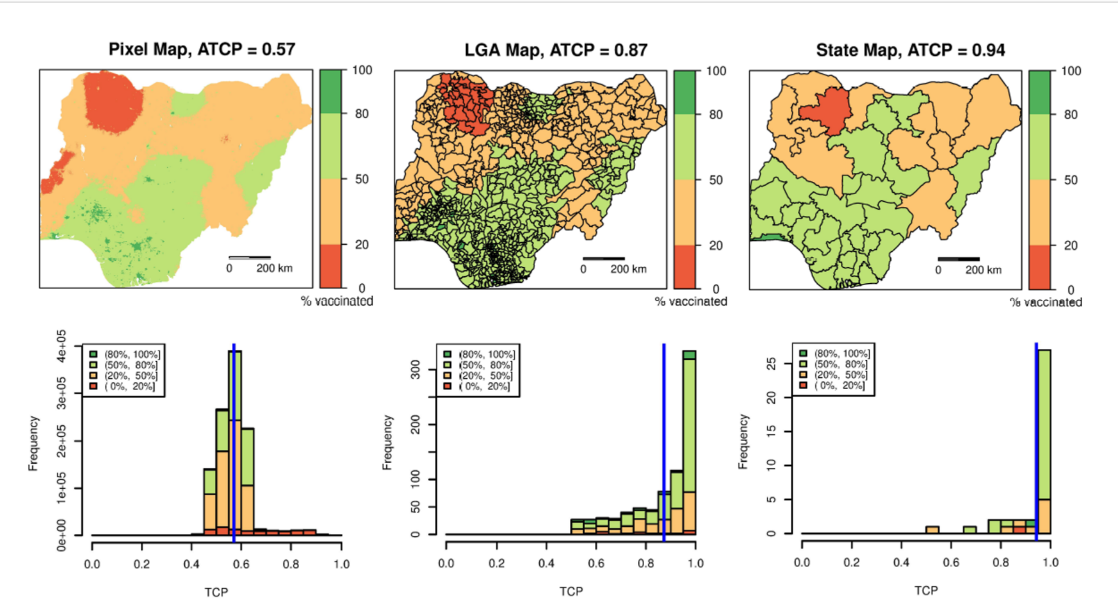
- Propose a new approach to control overall map uncertainty, which allows comparing uncertainty between maps.

1) Use a discrete set of colors to represent a partition of [0%, 100%], the range of vaccination coverage.
2) Examine the posterior distribution of the coverage estimate for each area and assign each area to the interval that contains the greatest posterior probability. And call this maximum the true classification probability (TCP).
3) Calculate the average of the TCPs across all areas and call it the average true classification probability (ATCP) of the map.

Uncertainty in maps

Modeling and presentation of vaccination coverage estimates using data from household surveys (by Tracy Qi Dong, Jon Wakefield)

- Conclusion:
  state map has the highest ATCP (0.94), followed by the LGA map (ATCP = 0.87) and the pixel map (ATCP = 0.57). This reflects that the precision is associated with the spatial resolution.
  Coverage estimates at a finer spatial resolution tend to have larger associated uncertainty, and hence poorer precision.



Main Topics

- ✓Data types and maps: Unit-level vs. Area-level vs. Pixel-level, Proportion vs. Count
- ✓Uncertainty in maps
- ✓Bayesian data analysis