ADAPTIVE SAMPLING DESIGN FOR ESTIMATING SPATIOTEMPORAL PATHOGEN PREVALENCE IN CITIES

Katherine McLaughlin

Associate Professor
Department of Statistics
Oregon State University

katherine.mclaughlin@oregonstate.edu

Jeffrey Bethel
Nicole Breuner
Benjamin Dalziel
Kathryn Higley
Allison Myers
Justin Preece
Tyler Radniecki

JSM Session on Big Data Initiatives in Survey Statistics

Portland, Oregon || August 6, 2024

OUTLINE

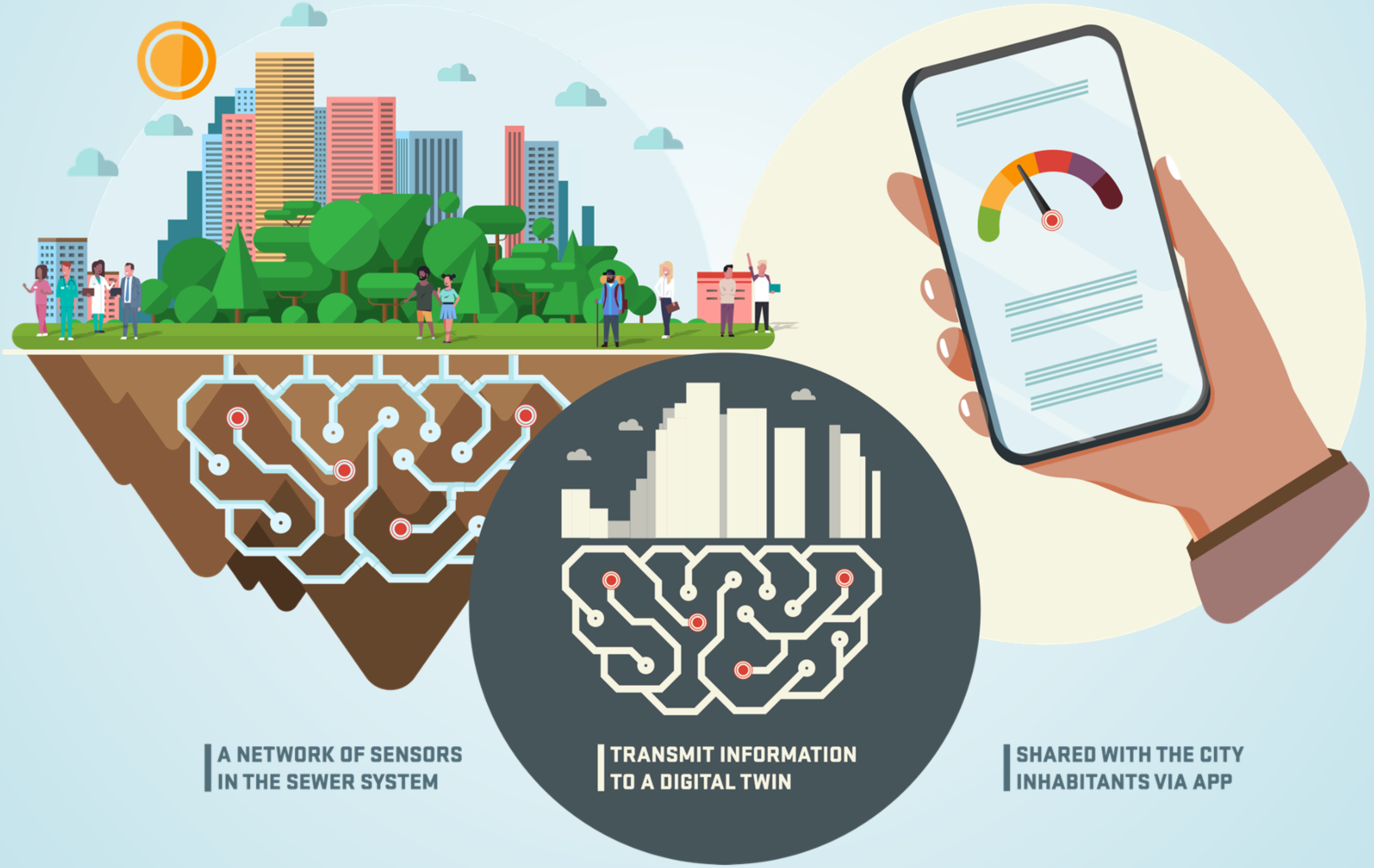
- Pandemic Resilient Cities Project
- Motivating examples
- Conceptualizing the adaptive sampling design
  - Creating sewersheds
  - Aligning with Census data
  - Unequal probability sampling
- Results
  - Viral concentration estimation
  - Preliminary results in City 1

PANDEMIC RESILENT CITIES PROJECT

- *Goal*: Partner with communities to create systems that produce city-level infectious disease forecasts
  - Build interdisciplinary partnerships between science & academia, public health professionals, and community leaders
  - Build mathematical models that demonstrate stepping back across epidemic tipping points
  - Identify feedback loops that predict and alter transmission rates
- Key components
  - Wastewater-based epidemiology
  - Surveillance by sequencing
  - Adaptive sampling

LONG-RANGE GOAL

PANDEMIC-RESILIENT CITY



A NETWORK OF SENSORS IN THE SEWER SYSTEM

TRANSMIT INFORMATION TO A DIGITAL TWIN

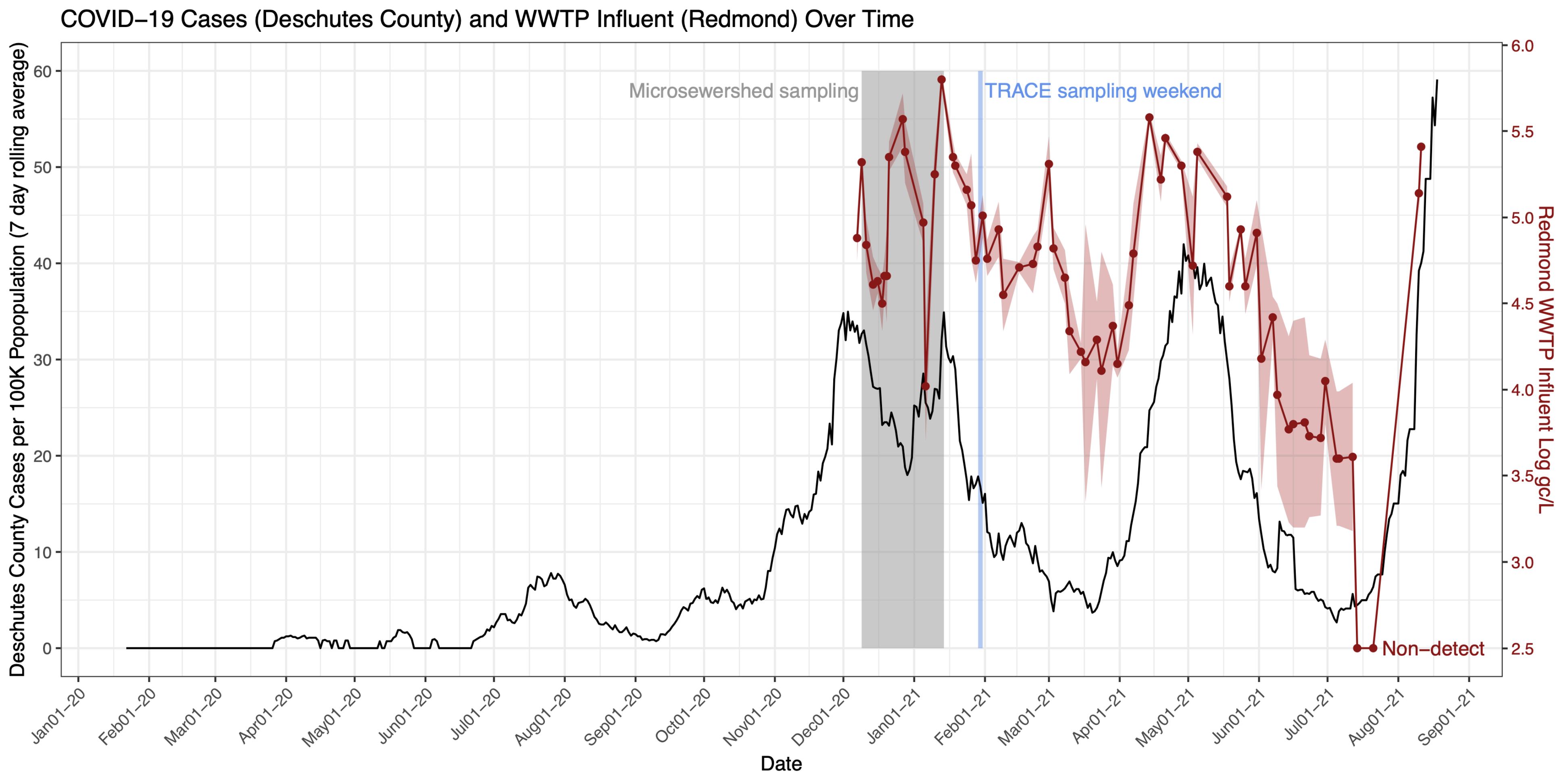
SHARED WITH THE CITY INHABITANTS VIA APP

WORK WITH CITIES

- Workshop last spring/summer with community stakeholders (public health, utilities, etc.)
- Prototype development starting this flu season, continuing through next
- Weekly wastewater monitoring at different point in city and at influent of wastewater treatment plant for 4 pathogens:
  - Influenza A
  - Influenza B
  - SARS-CoV-2 (COVID)
  - RSV
- Have time series of data from 3 cities

MOTIVATING EXAMPLES

WHY WASTEWATER?



Oregon State University

WHY UNEQUAL PROBABILITY SAMPLING?

Simulation example:

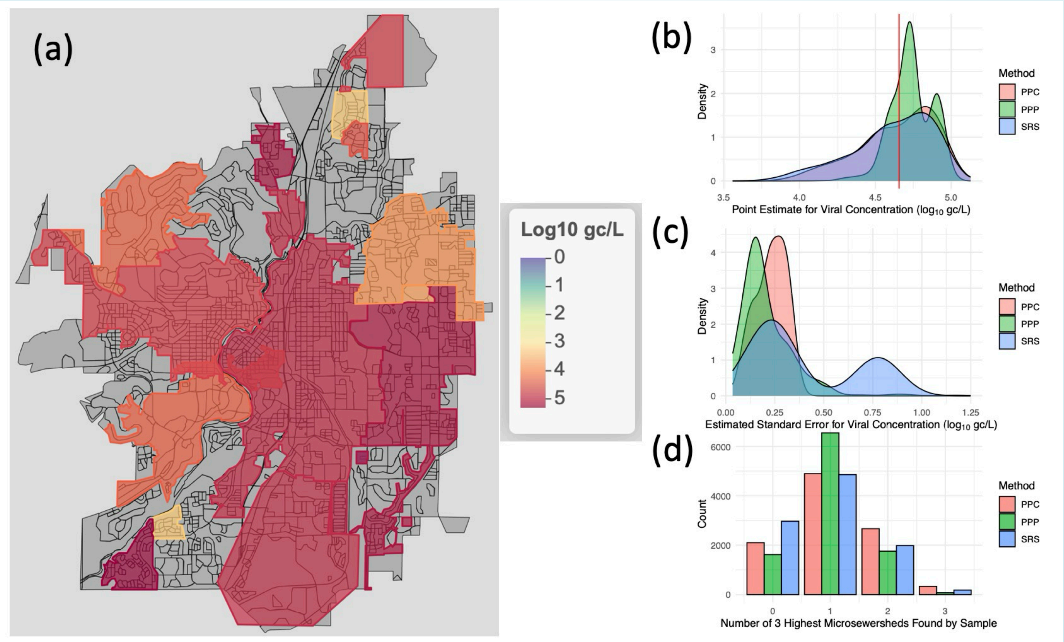
- A city has been split into 16 sampling locations
- Take 10,000 samples of size 5 using:
  - simple random sample (SRS)
  - probability proportional to population (PPP)
  - probability proportional to viral concentration (PPC)

WHY UNEQUAL PROBABILITY SAMPLING?

(a) Viral concentration data

(b) Point estimates and (c) SE of estimates by method for 10,000 simulated samples

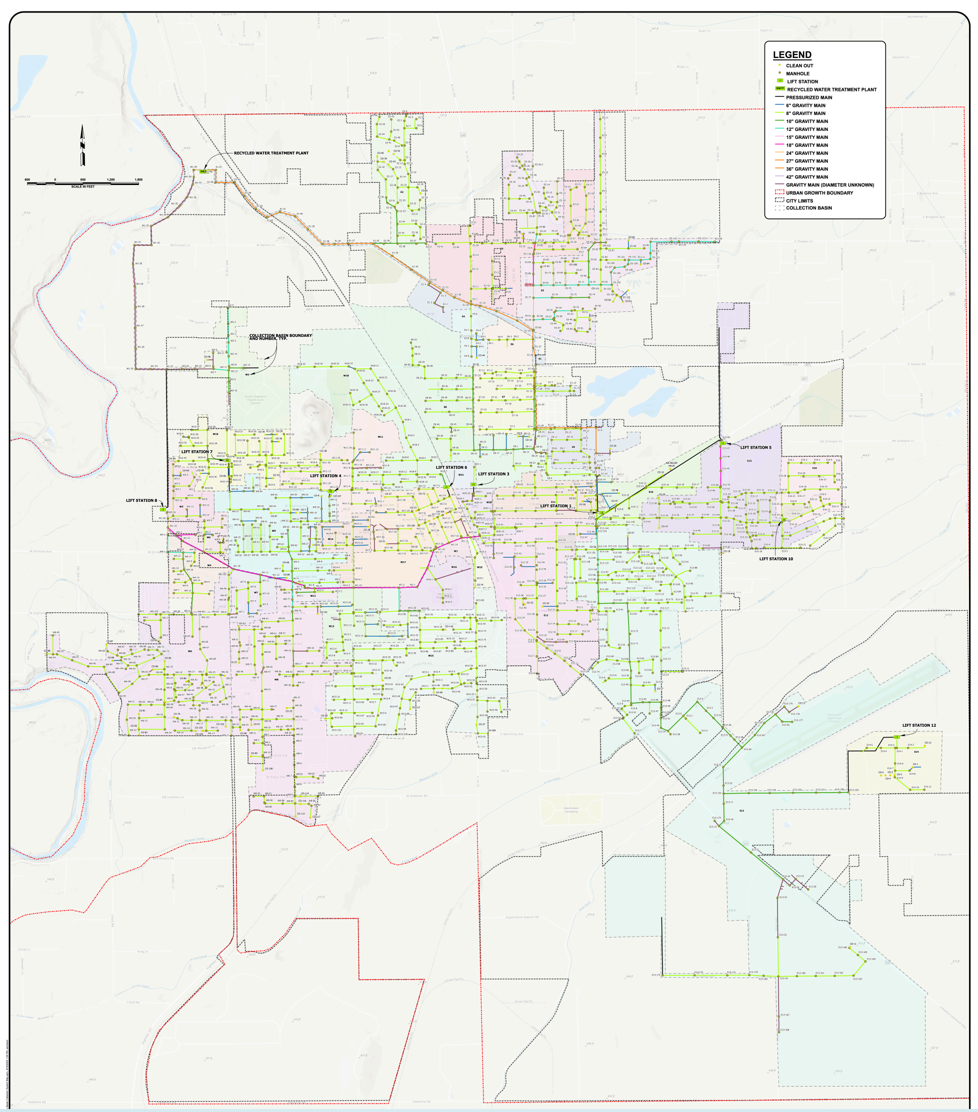
(d) PPC is the best method at sampling higher concentration MH, resulting in samples that contain at least two of the three highest concentration MH 29.8% of the time, compared with 18.1% of the time for PPP and 21.2% for SRS.



Oregon State University

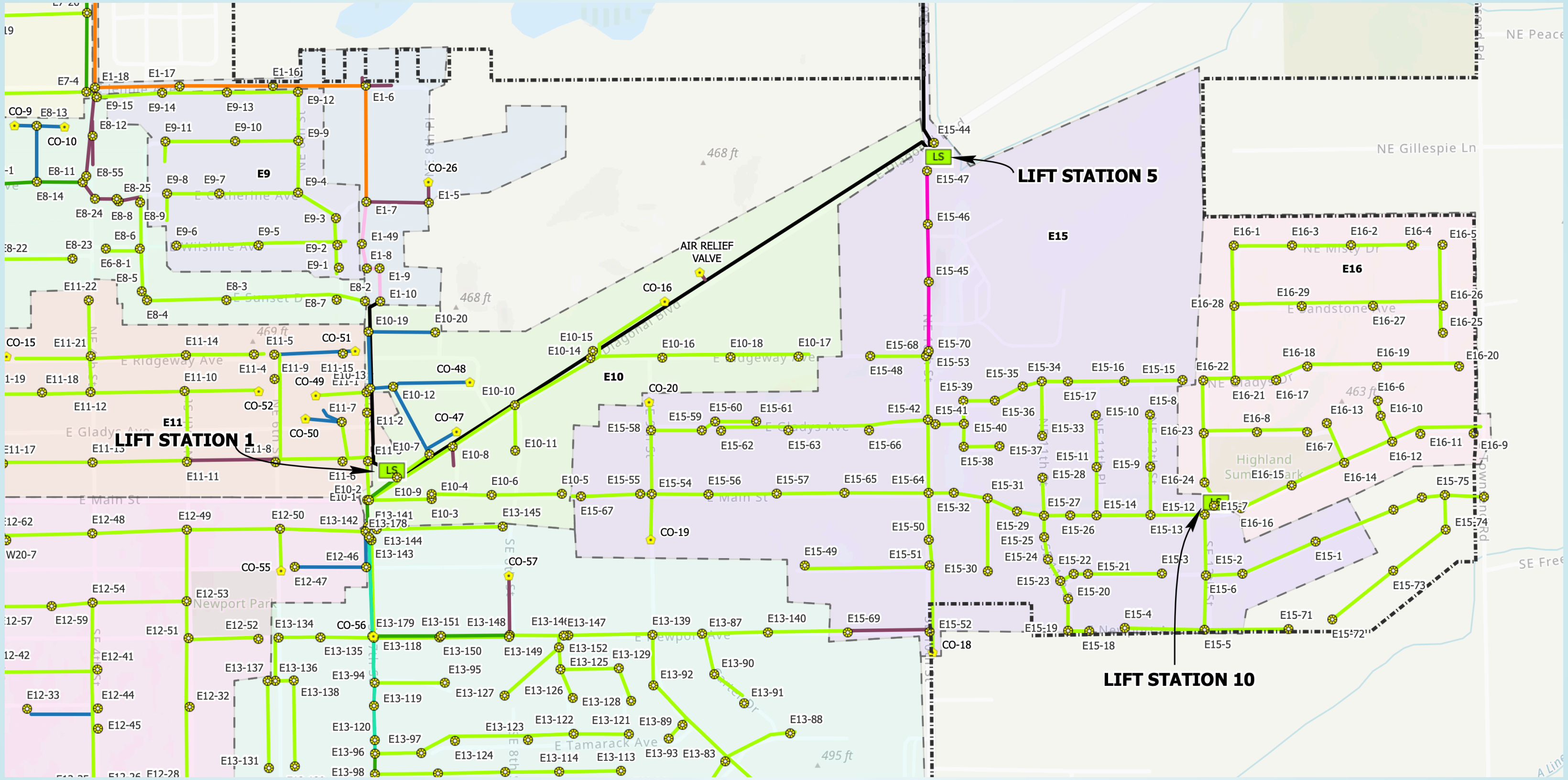
CREATING THE SAMPLING FRAME

CREATING SEWERSHEDS



Oregon State University

CREATING SEWERSHEDS

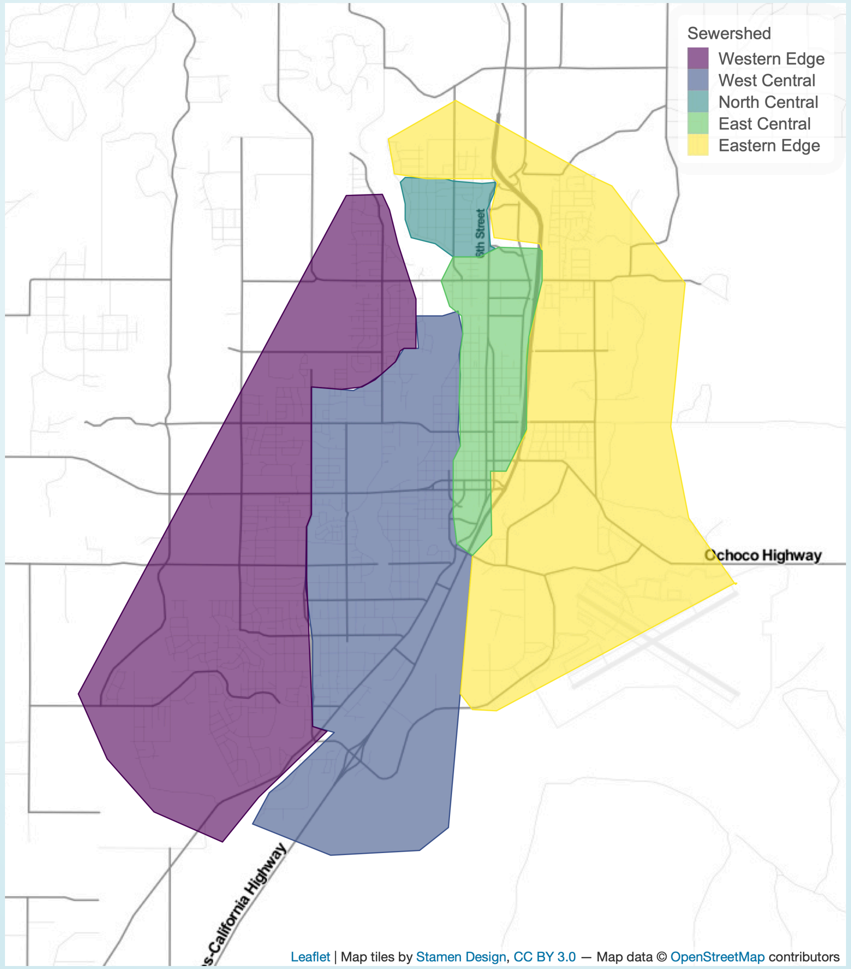


- How many?
- Identify Maintenance Hole (MH) that captures all flow from sewershed
- Attempt to make sewersheds disjoint

Oregon State University

SEWERSHEDS

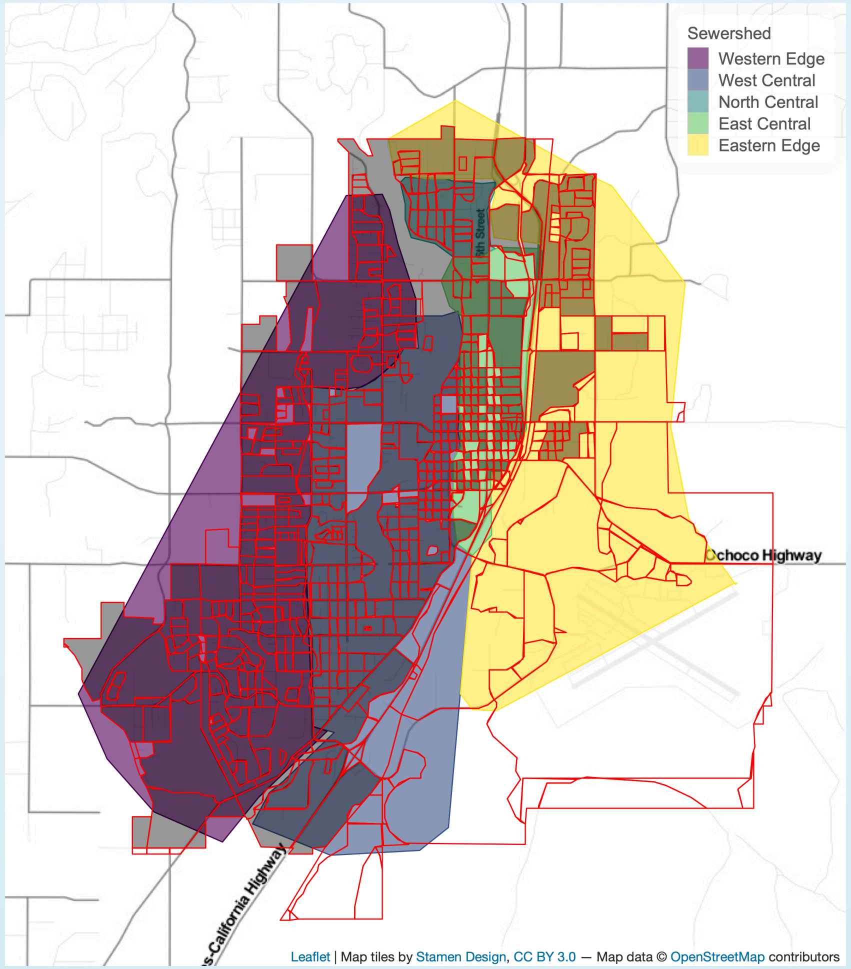
- Each sewershed will be a sampling unit



Oregon State University

CENSUS BLOCKS

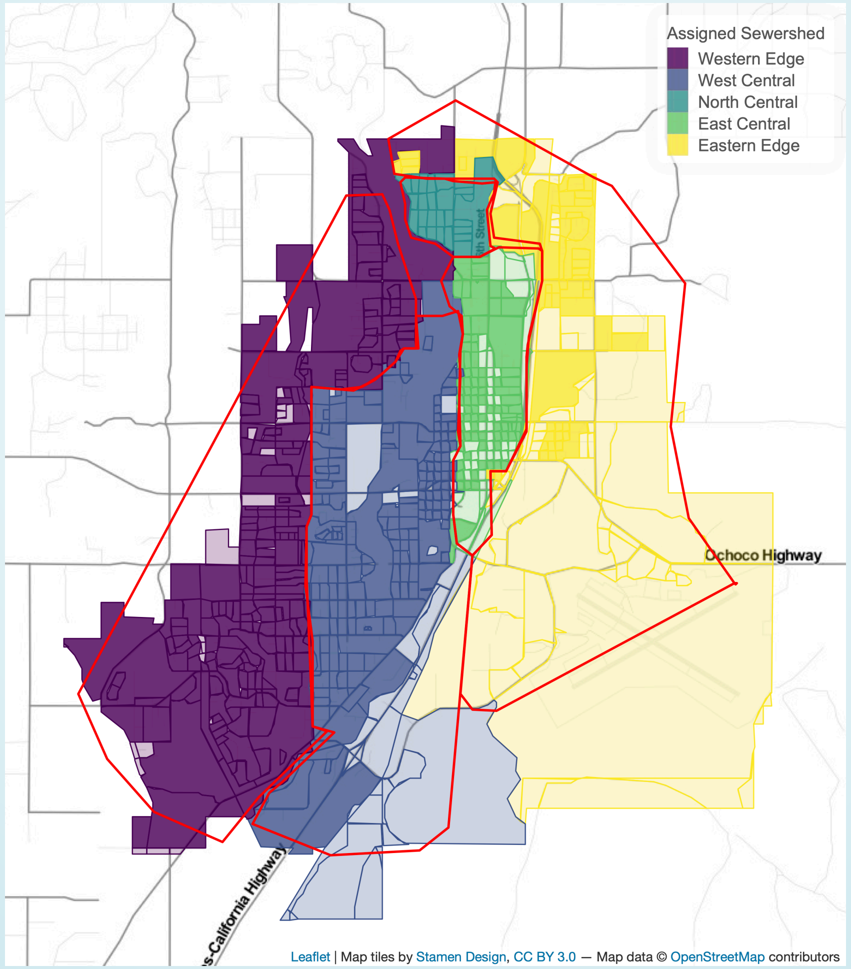
- Use Census data to add auxiliary information to inform selection probabilities



Oregon State University

DEVELOPMENT OF SAMPLING FRAME

- Intersect Census blocks and sewersheds to assign demographic variables



Oregon State University

UNEQUAL PROBABILITY SAMPLING DESIGN

- Selection probabilities for each sewershed are determined by a weighted average of population size and viral concentration of pathogen targets
- Updated each week based on previous weeks' data and logistical constraints

The probability of selecting sewershed i during week k is

$$\pi_i^{(k)} = \frac{\sum_{t \in \tau} \left(\frac{w_t^{(k)} x_{it}^{(k-1)}}{\sum_{i \in M} x_{it}^{(k-1)}} \right)}{\sum_{t \in \tau} w_t^{(k)}}$$

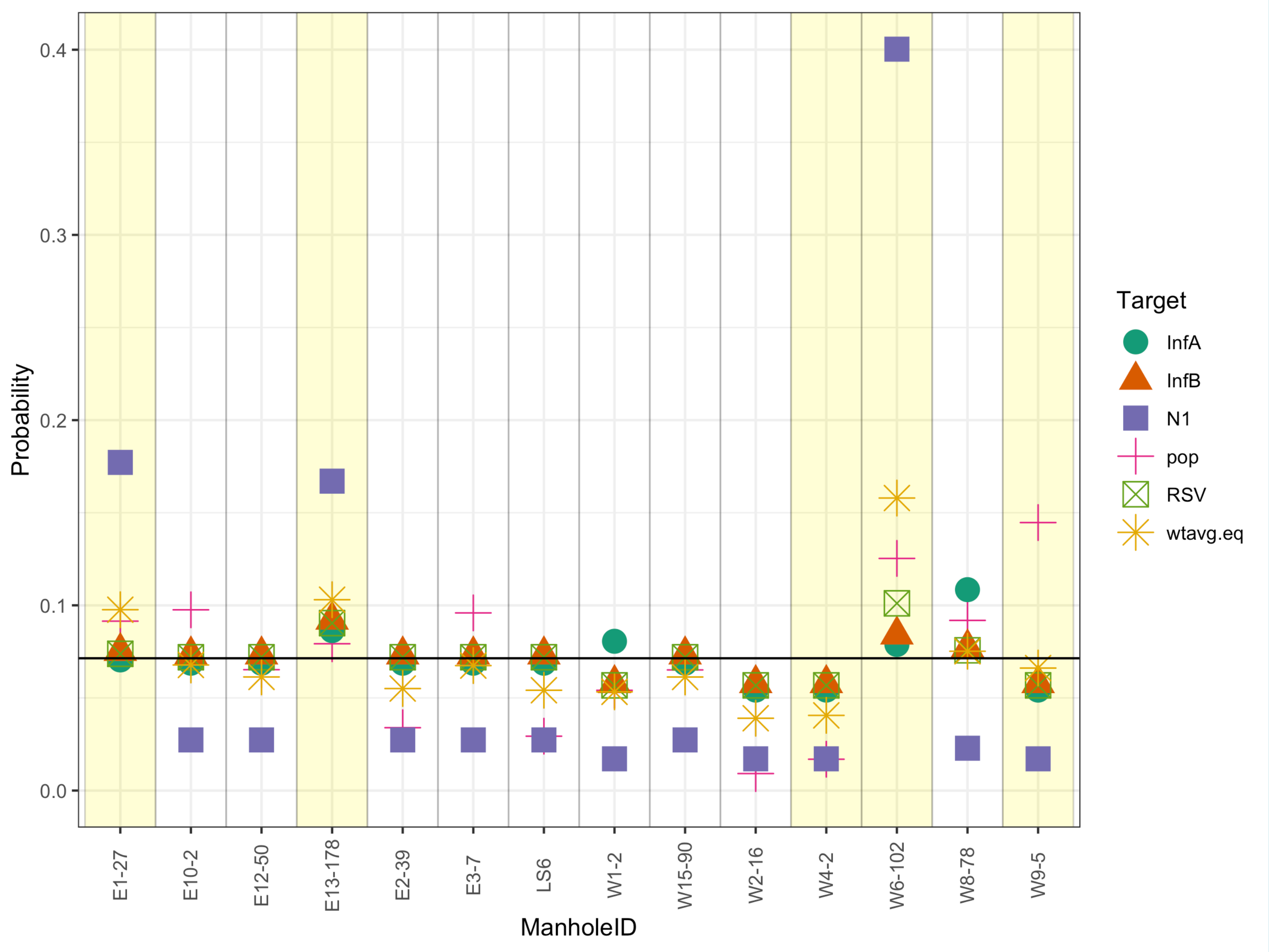
where

- τ is the set of targets used to determine probabilities (including population size and pathogen targets)
- $w_t^{(k)}$ is the weight given to item t of τ for week k
- $x_{it}^{(k-1)}$ is the value of item t for week $k-1$ (for example, the measured viral concentration of SARS-CoV-2 during the previous week)

ADVANTAGES OF THIS APPRAOCH

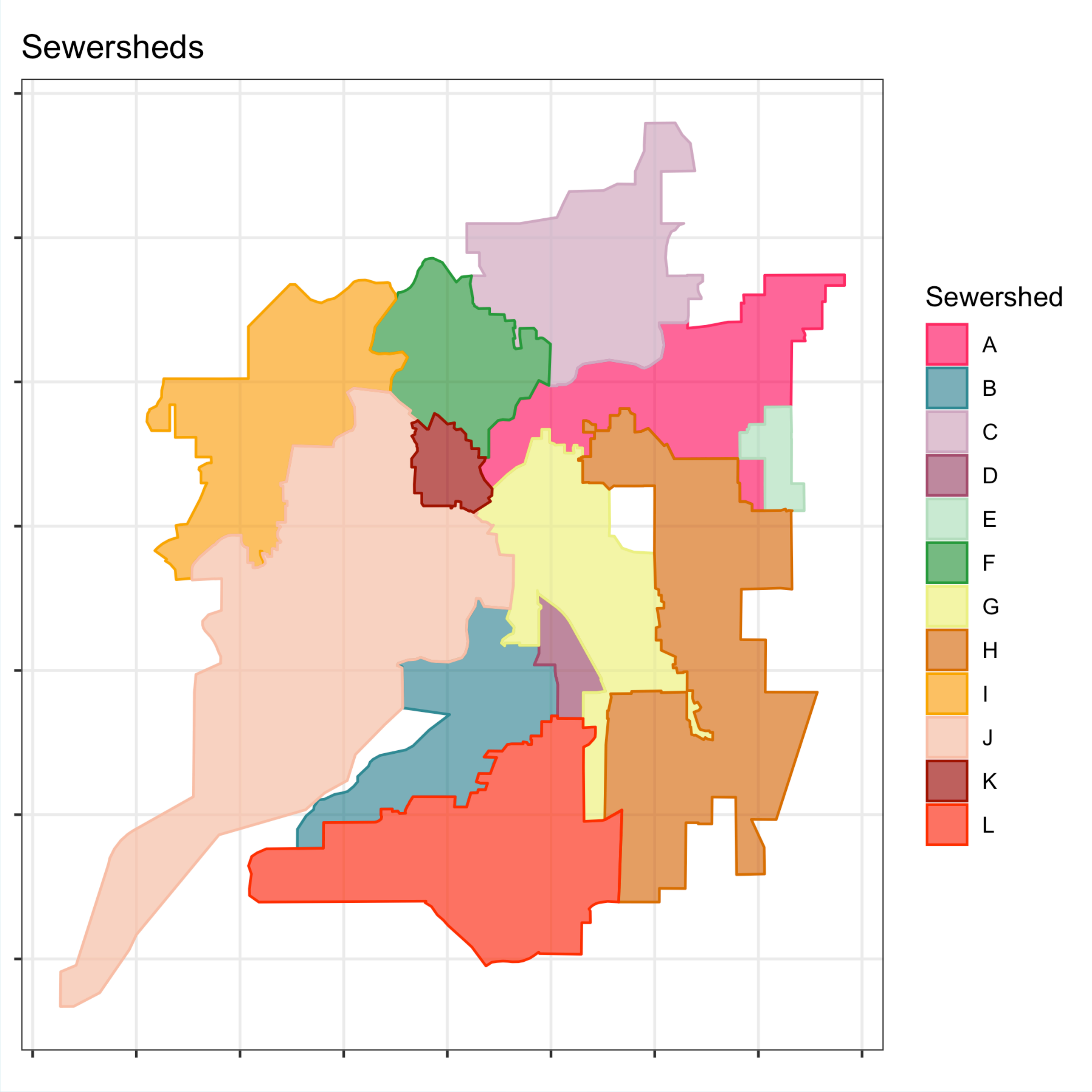
- Combines information from multiple targets and demographic variables (or other auxiliary sources)
- Allows information from past week(s) to be incorporated
- Can be tuned based on local needs, expert opinion (e.g. give influenza a lower weight when it is not flu season)
- Flexible pipeline that can handle delayed/missing data
- Sample with higher probability from 'hotspots' and other areas of interest
- Still a random, probability-based sample and allows generalizable inference

EXAMPLE



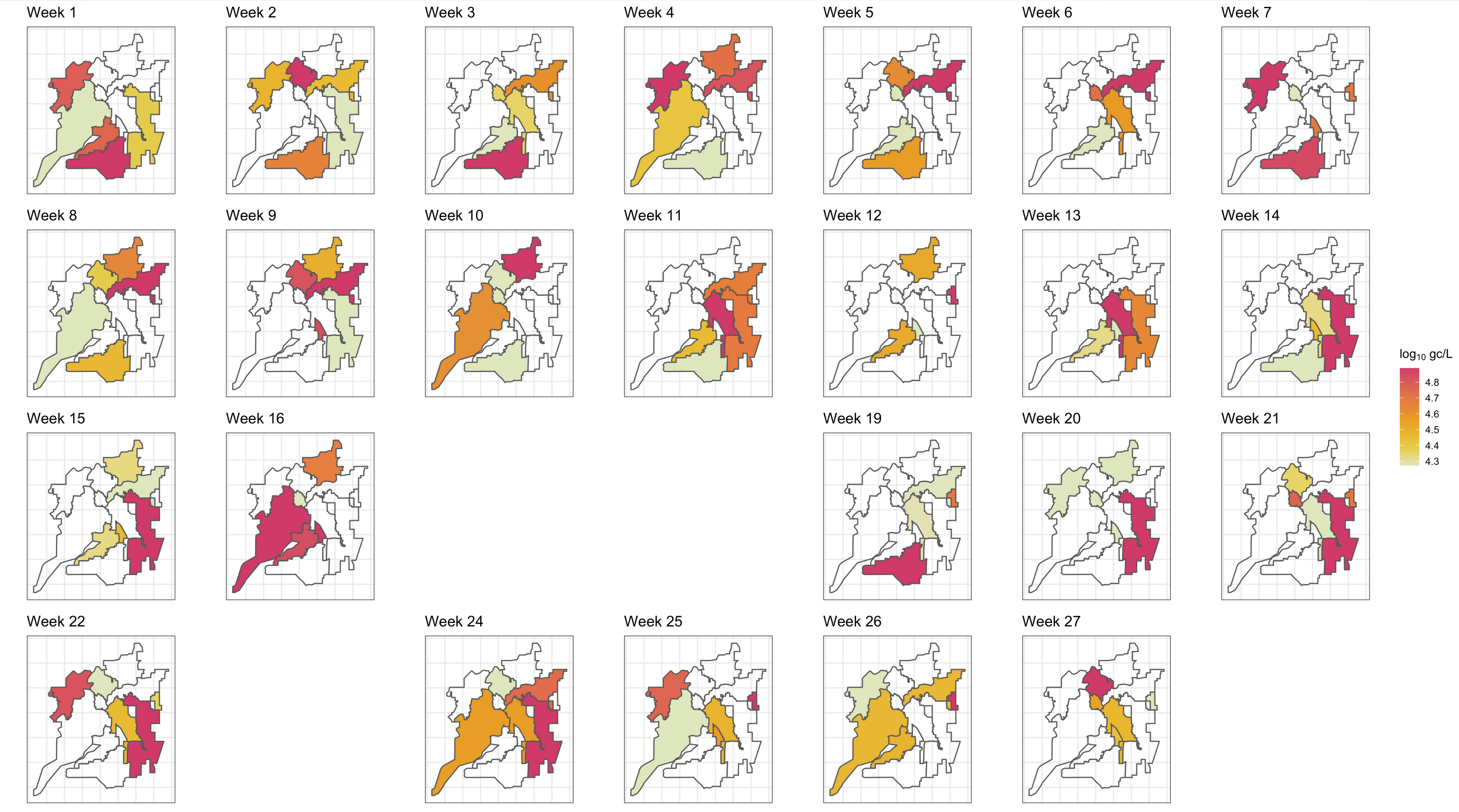
Oregon State University

RESULTS



Oregon State University

VIRAL CONCENTRATION OF SARS-COV-2



Oregon State University

VIRAL CONCENTRATION ESTIMATION

The estimated viral concentration based on the unequal probability sampling design can be found using a Horvitz-Thompson estimator:

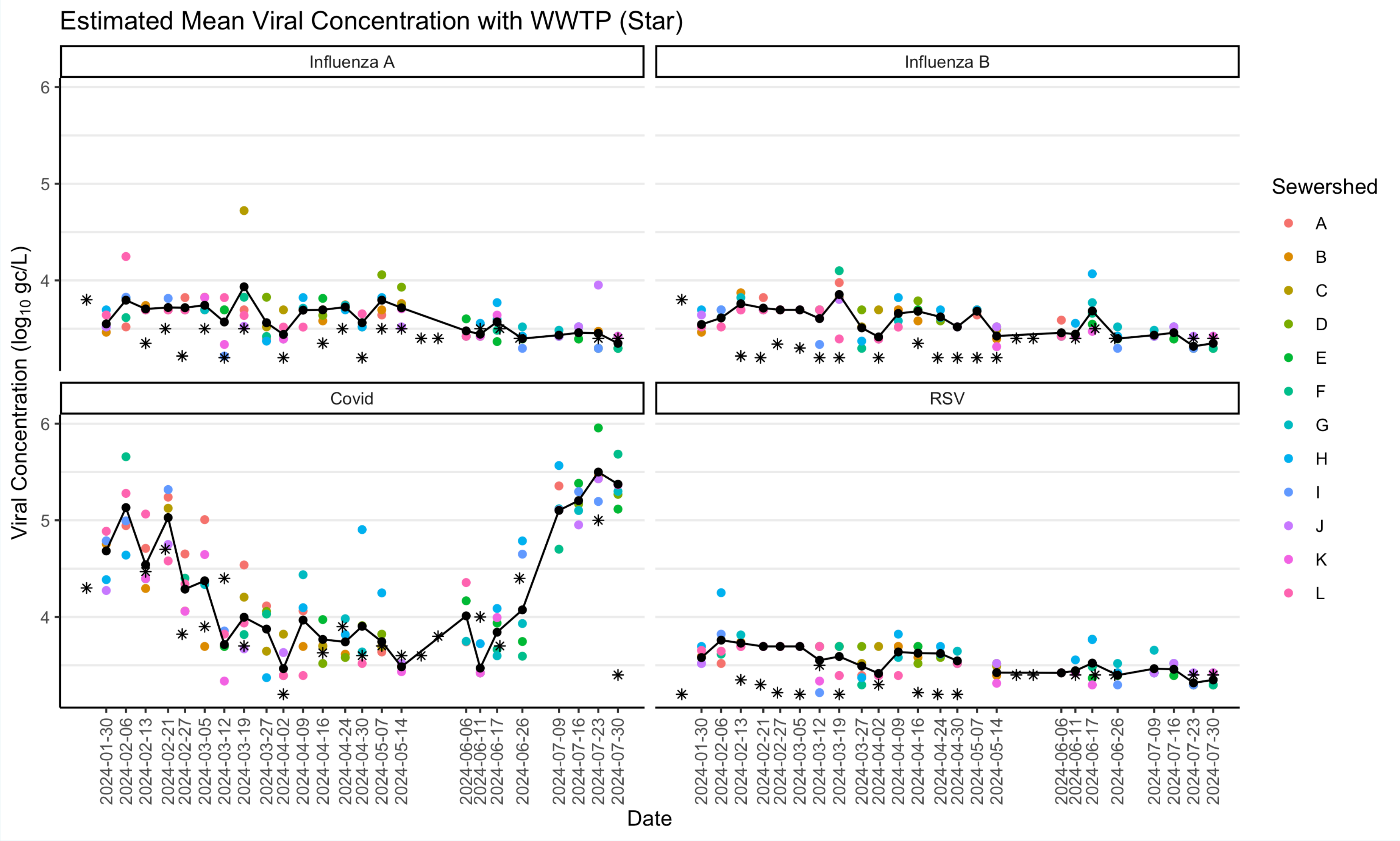
$$\hat{\bar{x}}^{(k)} = \frac{\sum\limits_{i \in \mathcal{S}^{(k)}} x_i^{(k)} / \pi_i^{(k)}}{\sum\limits_{i \in \mathcal{S}^{(k)}} 1 / \pi_i^{(k)}}$$

where

- $\pi_i^{(k)}$ is the selection probability for sewershed i during week k
- $x_i^{(k)}$ is the observed viral concentration (of the particular target being estimated) in sewershed i during week k

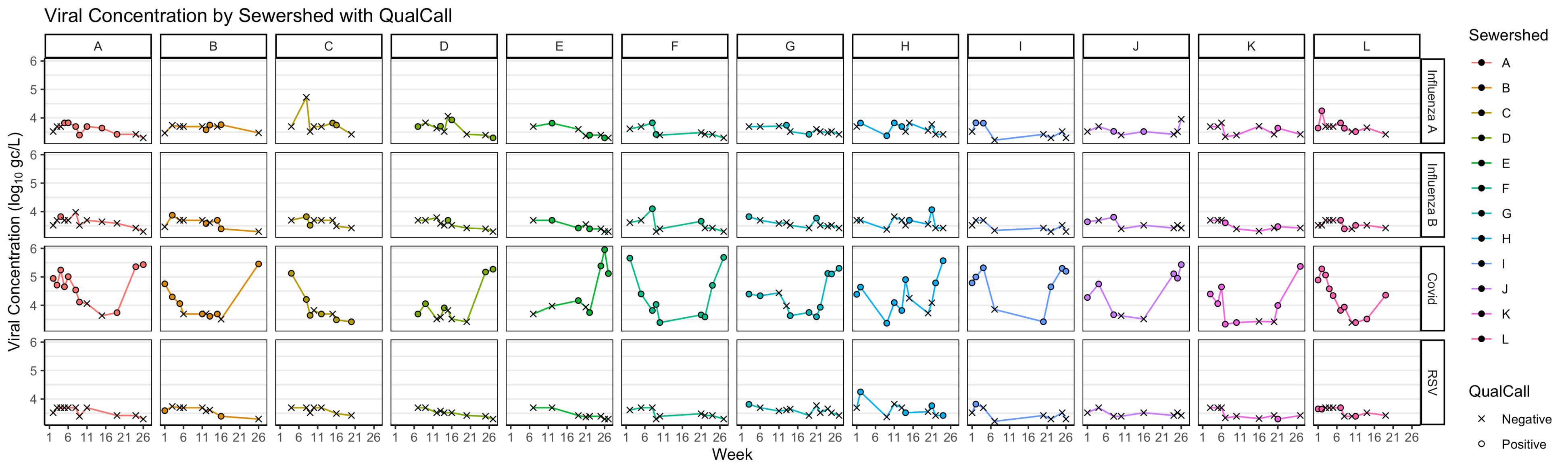
Implementation via the survey package in R.

RESULTS IN CITY 1



Oregon State University

CITY 1 BY SEWERSHED



Oregon State University

NEXT STEPS

- Use 'Rosetta Stone' to translate between log_{10} gc/L to cases
- Incorporate sampled data into mathematical models to nowcast and forecast case counts
- Incorporate other data streams (e.g. hospitalizations, demographic data)
- Infer concentrations for non-sampled locations at a point in time to create a heatmap
- Work with community partners on how to make the data products most actionable

CONCLUSIONS

- Flexible design-based sampling framework that incorporates previous weeks of data
- Cooperative feedback loop with wastewater & mathematical modeling teams, as well as community partners
- Potential for incorporation of a variety of data streams
- Can be tailored for unique characteristics, needs of cities
- Still a work in progress!!

THANK YOU!

(katherine.mclaughlin@oregonstate.edu)

Ben Dalziel (PI)
population biology, epidemic dynamics, collective behavior
benjamin.dalziel@oregonstate.edu

Kathryn Higley (co-PI)
radioecology, emergency response, radiation protection, risk communication
kathryn.higley@oregonstate.edu

Justin Preece (Project Coordinator)
computational infrastructure, data management, communication
preecej@oregonstate.edu

Joseph Agor (co-PI)
data-driven optimization with applications in health care
joe.agor@oregonstate.edu

Allison Meyers (co-PI)
health systems innovation, building workforce capacity in public health
allison.myers@oregonstate.edu

Tyler Radniecki (co-PI)
sustainable biological wastewater/stormwater treatment, wastewater-based epidemiology
tyler.radniecki@oregonstate.edu

Jeff Bethel (co-PI)
infectious disease epidemiology, community-based public health, disaster research methods
jeff.bethel@oregonstate.edu

Katherine McLaughlin (co-PI)
statistical sampling methods, social network analysis
katherine.mclaughlin@oregonstate.edu

Justin Sanders (co-PI)
microbiology, ecology and evolution of infectious diseases
justin.sanders@oregonstate.edu