
COVID through a statistical lens

Mark S. Handcock
Professor of Statistics

What is Statistics?

The **statistical sciences** focus on the question:

How do we learn from data?

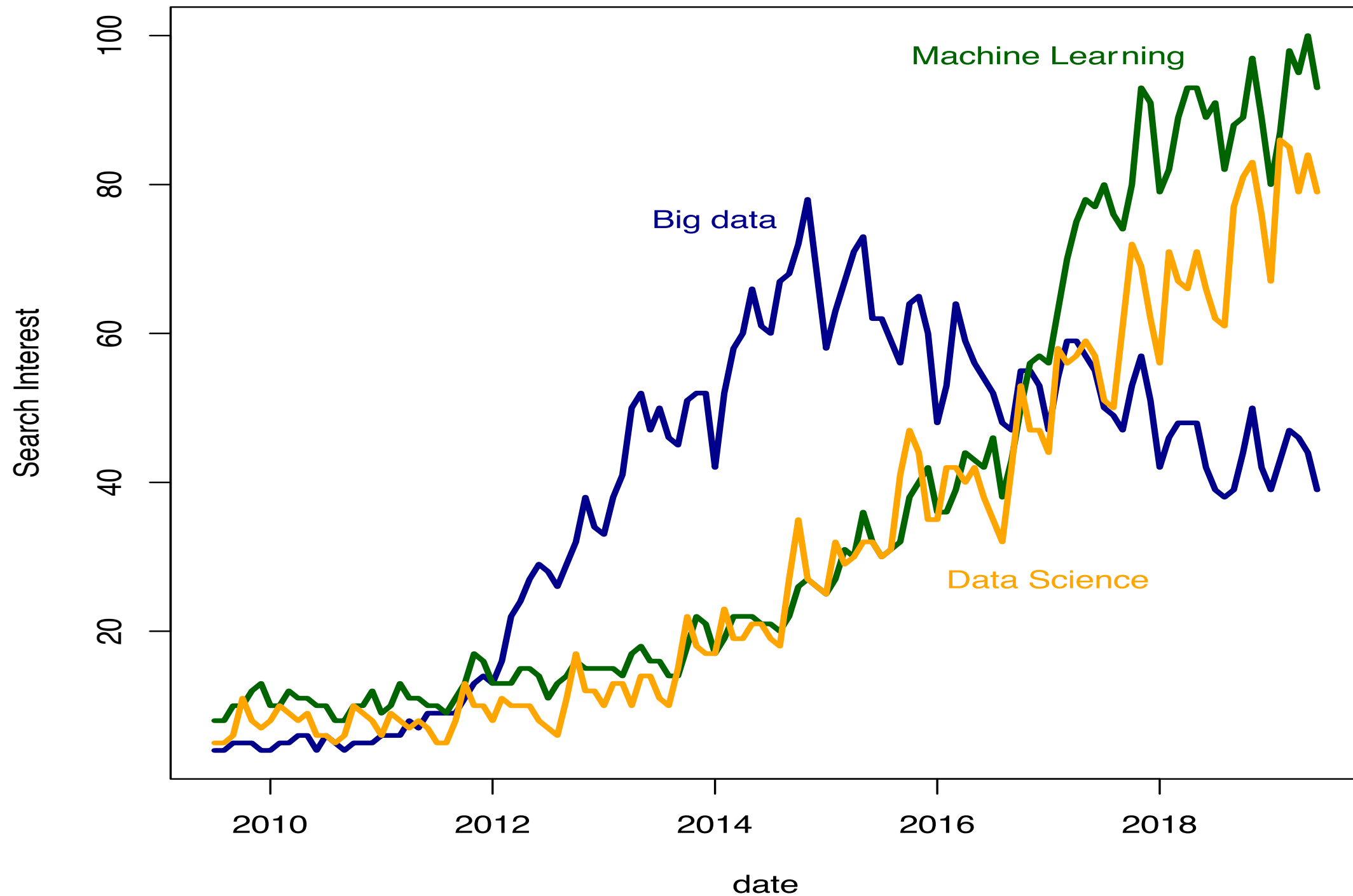
What is Statistics?

The purpose of the discipline is the development of tools
for

designing studies
and
analyzing data

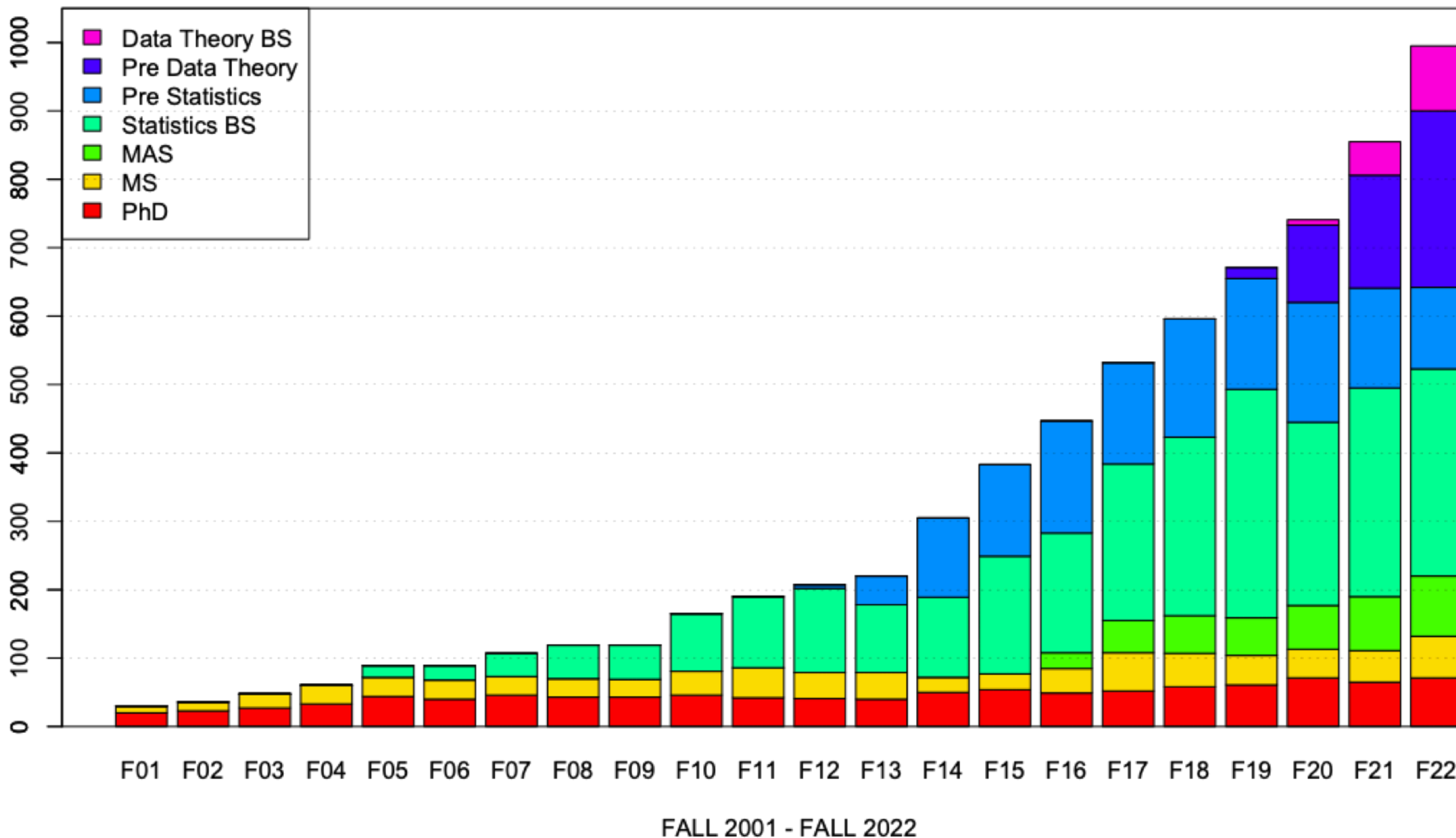
Rapidly changing in response to the remarkable
increase in demand for statistical thinking
in research, industry, business and government

The rise of Statistics, by other names...



Change in the popularity of search terms over the last decade based on Google Search. Numbers represent search interest relative to the highest point on the graph. Created using Google Trend: <https://www.google.com/trends>

The rise of Statistics, by other numbers...



Statistic is the #1 growing department at UCLA over the last 10 years

Some statistical aspects of the COVID-19 pandemic

- Estimating epidemic characteristics using contact-tracing data
- Looking at the spatial patterns of COVID within Los Angeles communities
- Estimating the number of deaths due to COVID in countries with limited data

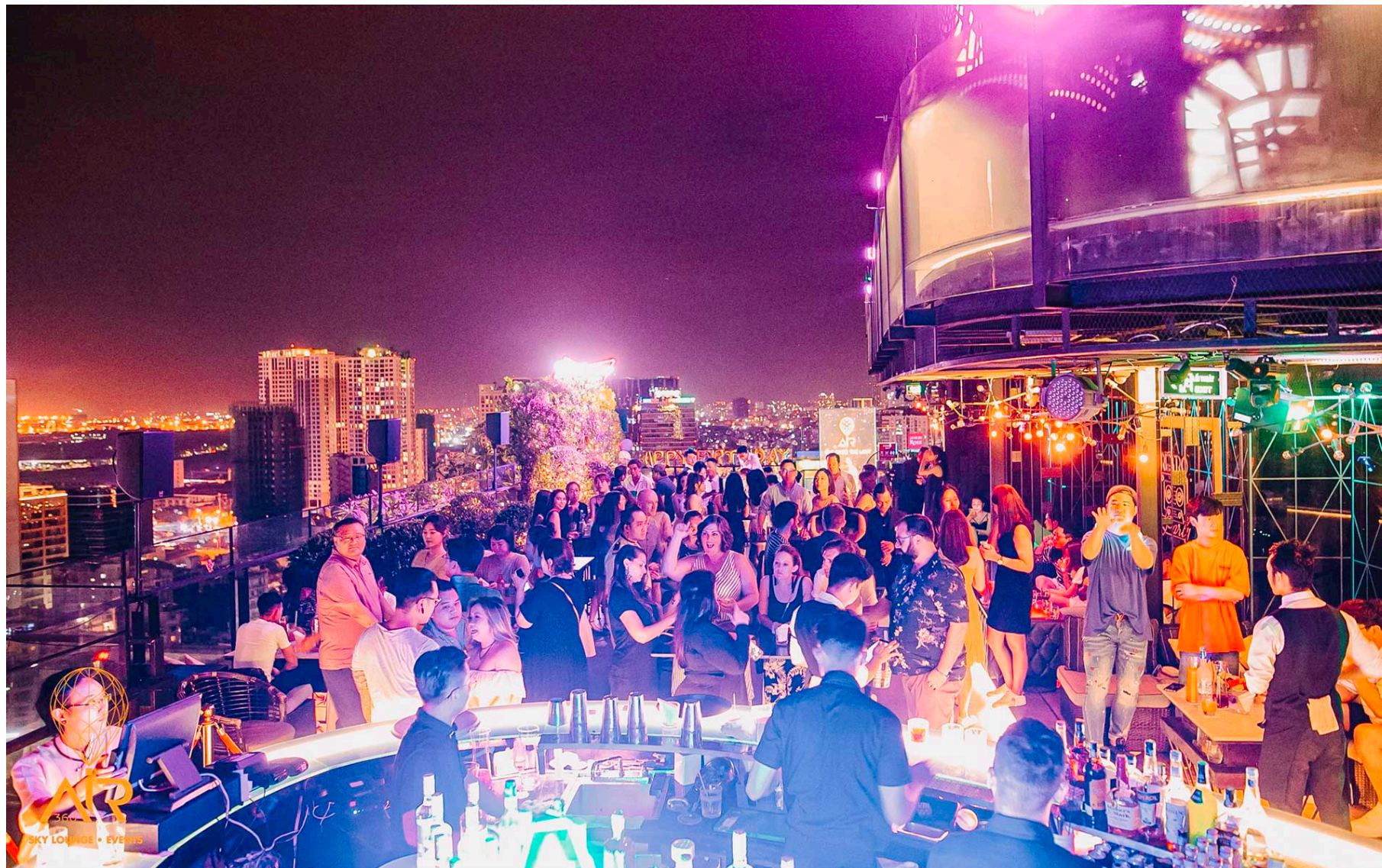
The Big Picture: The Past and Future of Epidemics

- At the beginning of an epidemic (e.g., March 2020) separate **outbreaks** occur among susceptible populations
- The public health system responds by isolating cases and quarantining contacts to **stop the outbreak spreading** far
- **Questions:** Is this necessary? Are stronger efforts needed? What is the risk of a generalized epidemic?



Example: Contact tracing an outbreak in Ho Chi Minh City, Viet Nam

Epidemiological investigation was conducted on a cluster of COVID-19 cases linked to a bar gathering on March 14, 2020

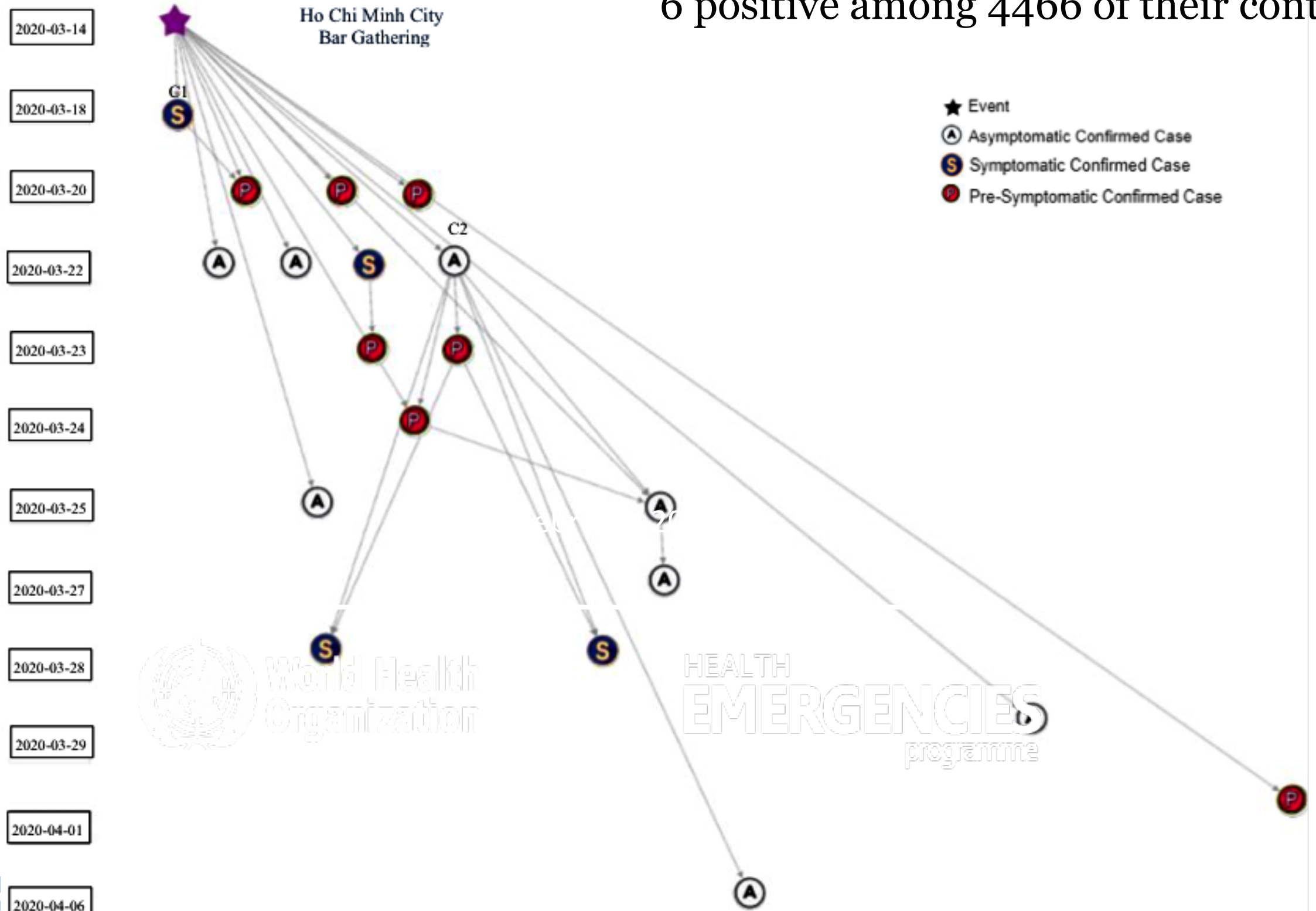


Timeline of outbreak after the bar event

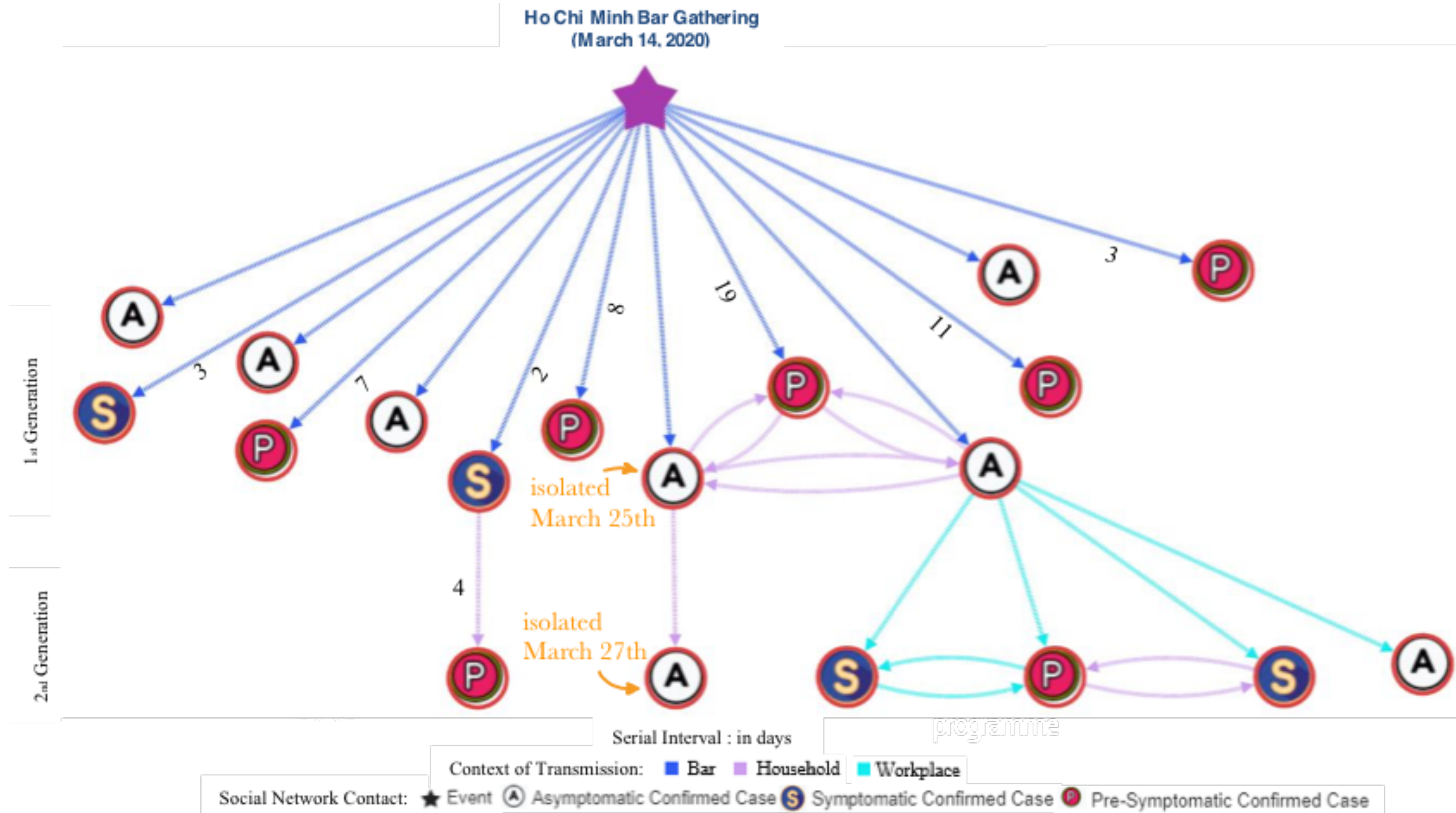
19 total cases

13 positive among 298 bar attendees

6 positive among 4466 of their contacts



Routes of transmission out from people at the bar



Contact tracing: as a data source

While contact tracing is meant to protect people, it is also a source of information:

- “snowball” contact tracing after the outbreak is detected
- anonymized data on who spent time with who (cases and non-cases)



Contact tracing: as a data source

- Information collected as part of a contact tracing system can be used to:
 - determine how isolation and quarantine can control transmission in local epidemics
 - modify contact tracing to make it more efficient or less of a burden on people
- We apply statistical modeling to local contact tracing data to do this.
- Our goal is to determine what the epidemic would have looked like had contact tracing not been done!

Measures of how transmissible an unaltered epidemic is

- The **reproductive number** is a measure of the contagiousness or transmissibility of the infection
- The basic reproductive number (R_0) is **the number of other people, on average, an infectious person will infect.**
- In an outbreak, we start at ground zero - a completely susceptible population.
- It depends on the **given social setting, how susceptible the people are, social mixing, etc.**

Disease	
Influenza(s)	0.9 to 2.1
Measles	12 to 18
SAR-CoV-2 original	2.4 to 3.9

Statistical data

- Focus on the epidemic in a limited spatial-temporal region (an **outbreak**).
- Presume we observe all cases related to outbreak, even if censored.
- We focus on the situation where the source event is identified and we have information on the **exposure network** of all people.
- 19 total cases, 13 positive among 298 bar attendees
- additional 6 positive among 4466 of their contacts



Statistical modeling: Bayesian framing

We take a Bayesian perspective on estimation.



We chose to specify knowledge about unknown quantities via probability distributions about them

- There is much **prior knowledge** about the epidemic outside of the outbreak
- We need a method to **combine the available knowledge** with the outbreak data coherently
- In the Bayesian framework we can **quantify the uncertainty** of the estimates in a natural way



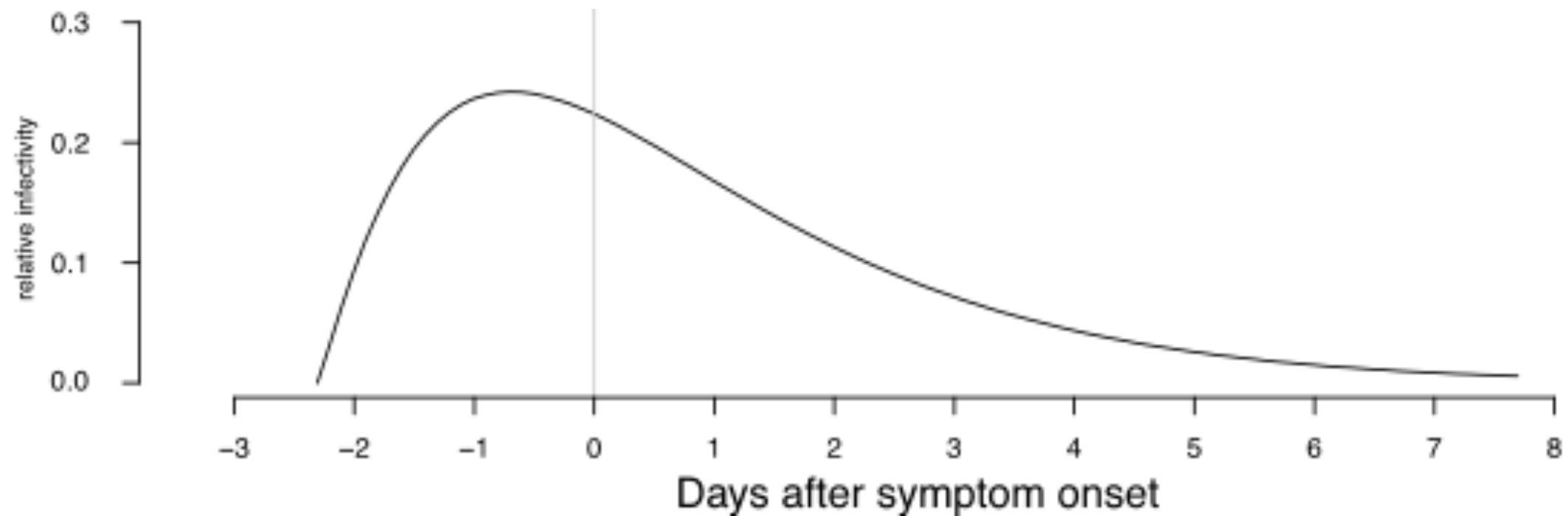
Estimation of the basic reproduction number, R_0

- We assume that people have an **infectivity profile** given by a probability distribution $p_{inf}^A(s)$, dependent on time since infection, s , of an asymptomatic person and the time since onset of symptoms for a symptomatic person.
- We **modeled individual transmission with a Poisson process in time**, so that the instantaneous rate at which a person infects at time $t - s$ is

$$R_0^A p_{inf}^A(t - s)$$


Specification of relative infectivity profiles

Infectivity of symptomatic and pre-symptomatic person, relative to onset time



Estimation of the basic reproduction number, R_0

Hence Y_i , the number of people infected by case i , is Poisson with mean

$$R_0^{A_i} \int_{C_i}^{I_i} p_{inf}^{A_i}(t - S_i) dt$$

where C_i , S_i , I_i and A_i are the times of infection, onset, isolation and the symptomatic status of case i , respectively.

Estimation of the basic reproduction number, R_0

- Hence treating the unknown infection times, C_i , as latent variables and using the exposure network, we can compute

$$P\left(R_0^{pre}, R_0^{sym}, R_0^{asy} \mid \{Y_i\}_{i=1}^n\right)$$

- Implemented via Hamiltonian Markov Chain Monte Carlo (HMC)



Estimation of the basic reproduction number, R_0

Reproduction numbers		90% credibility interval
Overall	2.64	1.41 – 3.68
asymptomatic cases	2.36	0.84–3.56
pre-symptomatic cases	1.97	0.85–5.38
symptomatic cases	0.71	0.28–5.50
secondary cases	1.58	0.83–2.89

Asymptomatic, pre-symptomatic and symptomatic epidemiology

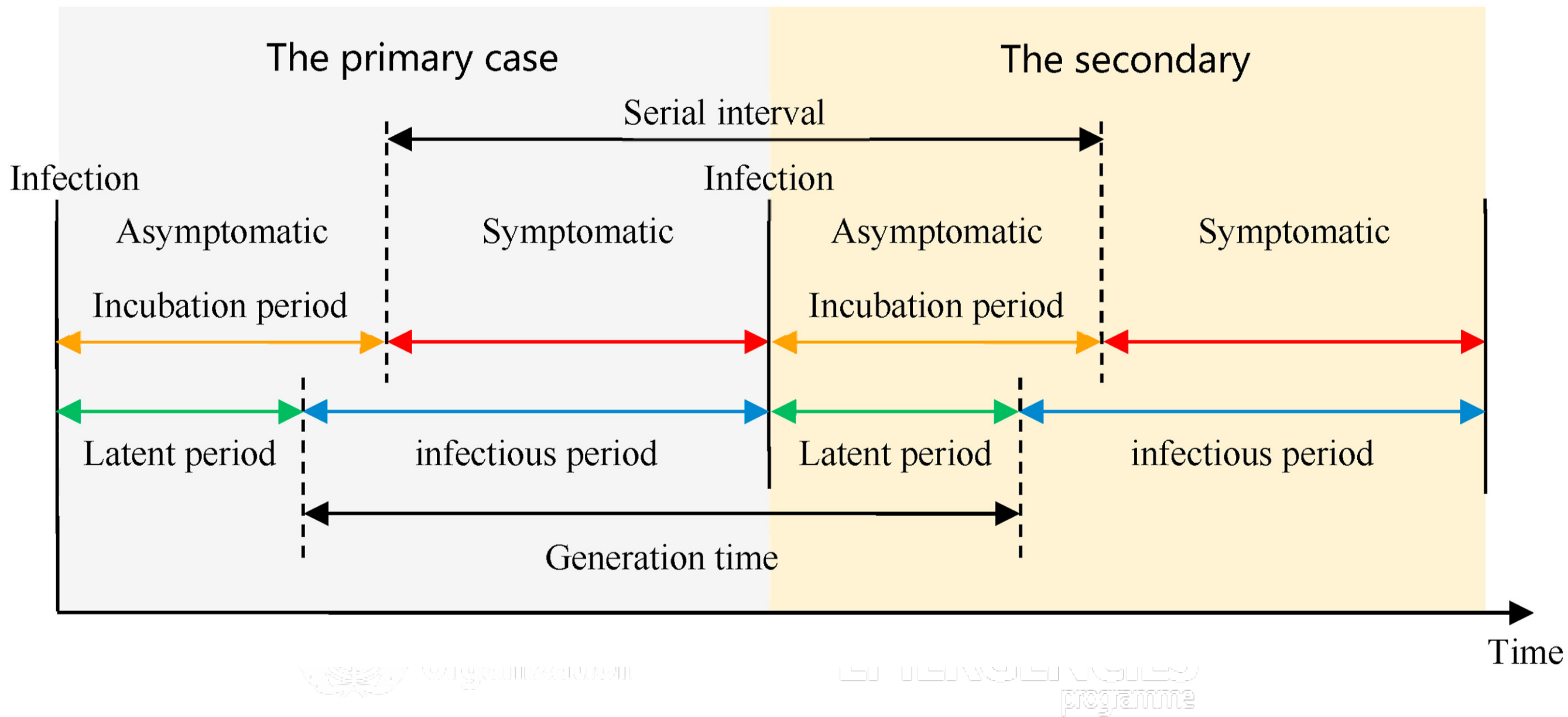
Population proportions of	proportion	90% credibility interval
asymptomatic cases	43%	26% - 60%
pre-symptomatic cases	35%	20% - 52%
symptomatic cases	22%	9% - 37%
asymptomatic transmission	45%	13% - 74%
pre-symptomatic transmission	24%	11% - 38%
symptomatic transmission	31%	15% - 49%

Looking at the spatial patterns of COVID within Los Angeles communities

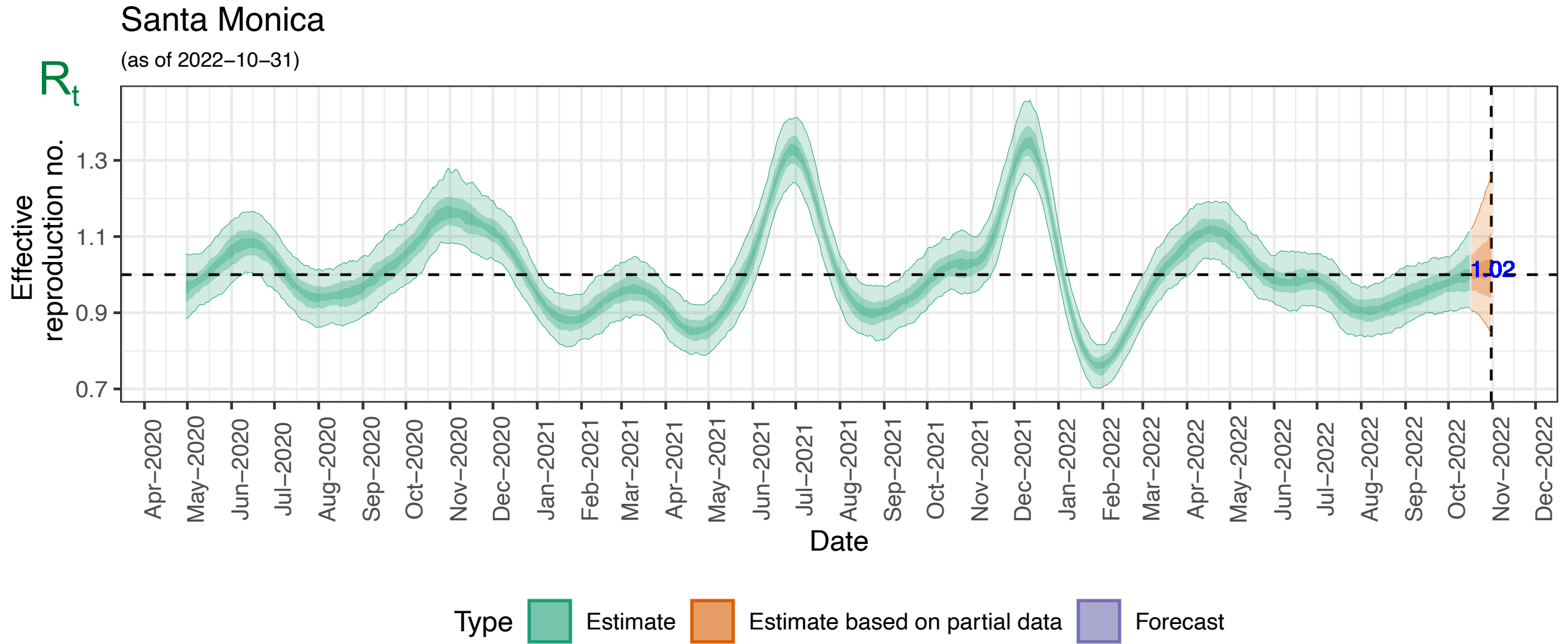
Daily regional reporting of COVID cases and deaths

- LA County Public Health reports cases and deaths for 341 cities/communities each day (or so)
- We can use a simple non-stationary Gaussian temporal process to estimate the **time-varying reproduction numbers**, R_t , and then infer infections separately for each community
- It uses Bayesian inference implemented via R/Stan

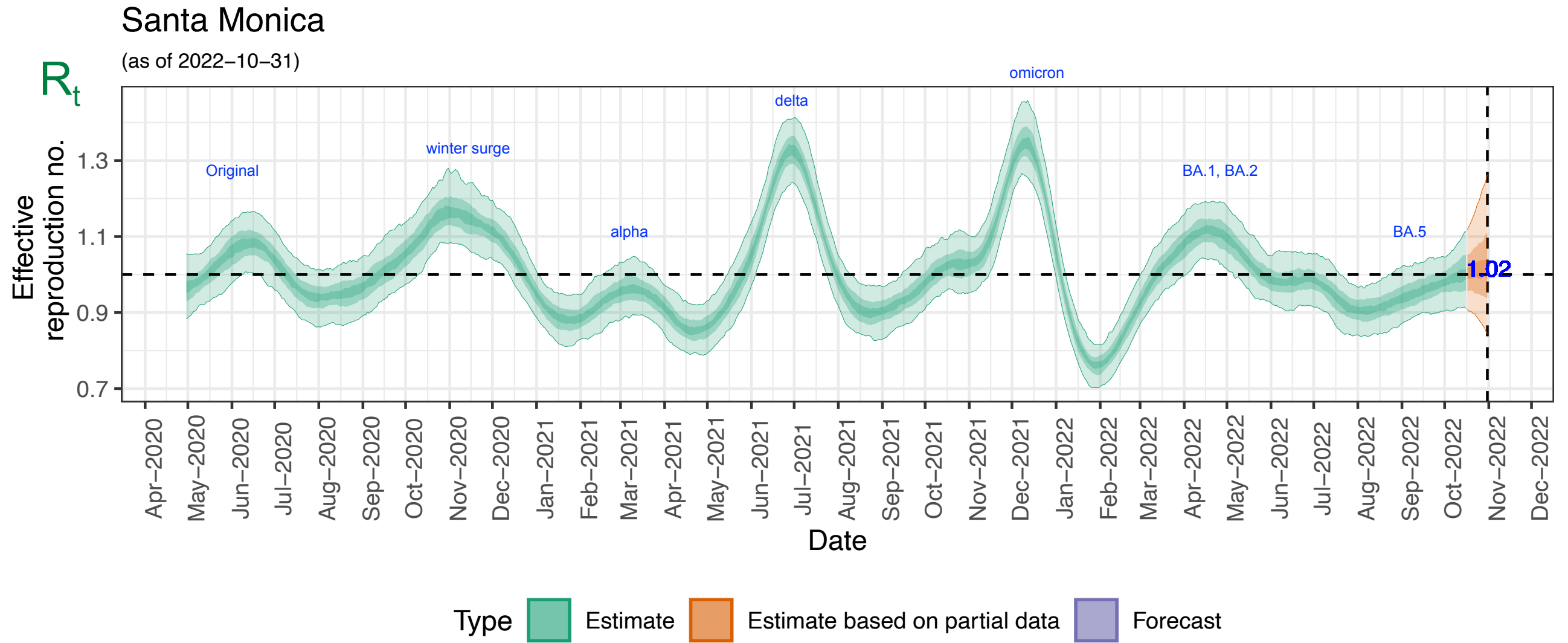
Cartoon of infection from a primary case to another person



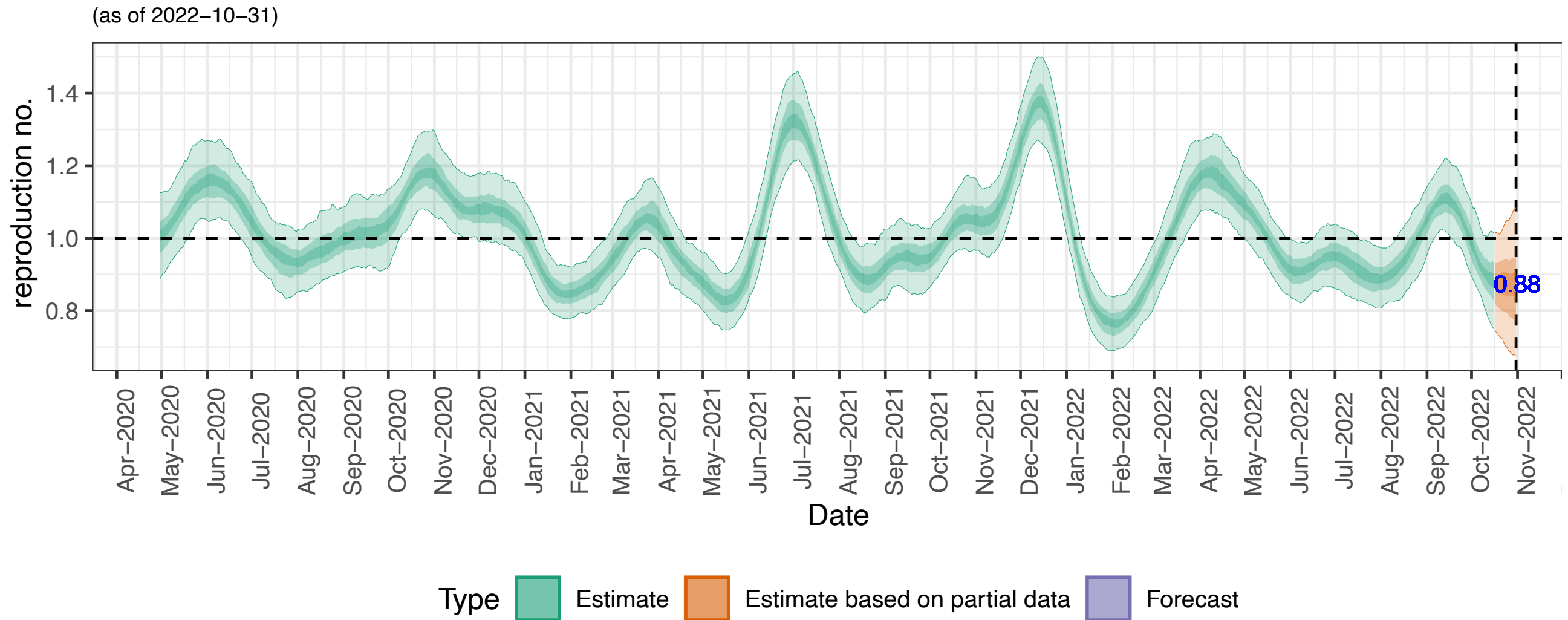
Example: Santa Monica (population 92446)



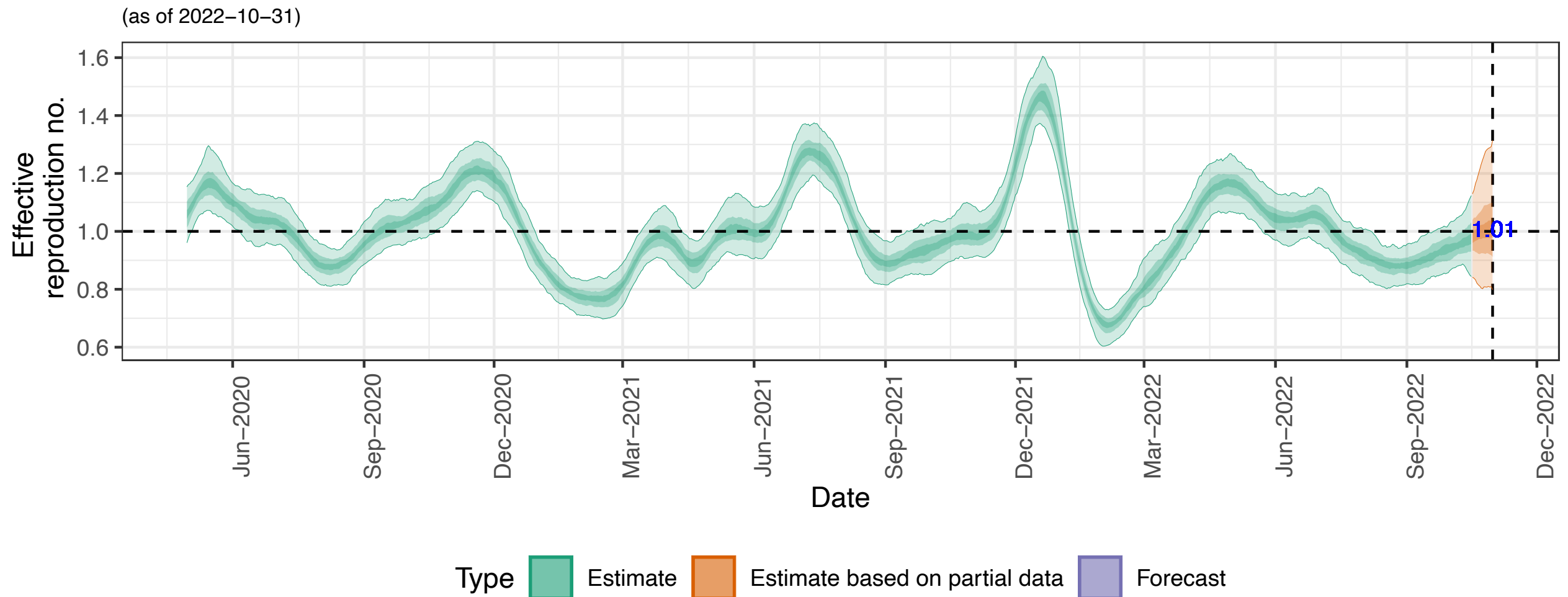
Example: Santa Monica (population 92446)



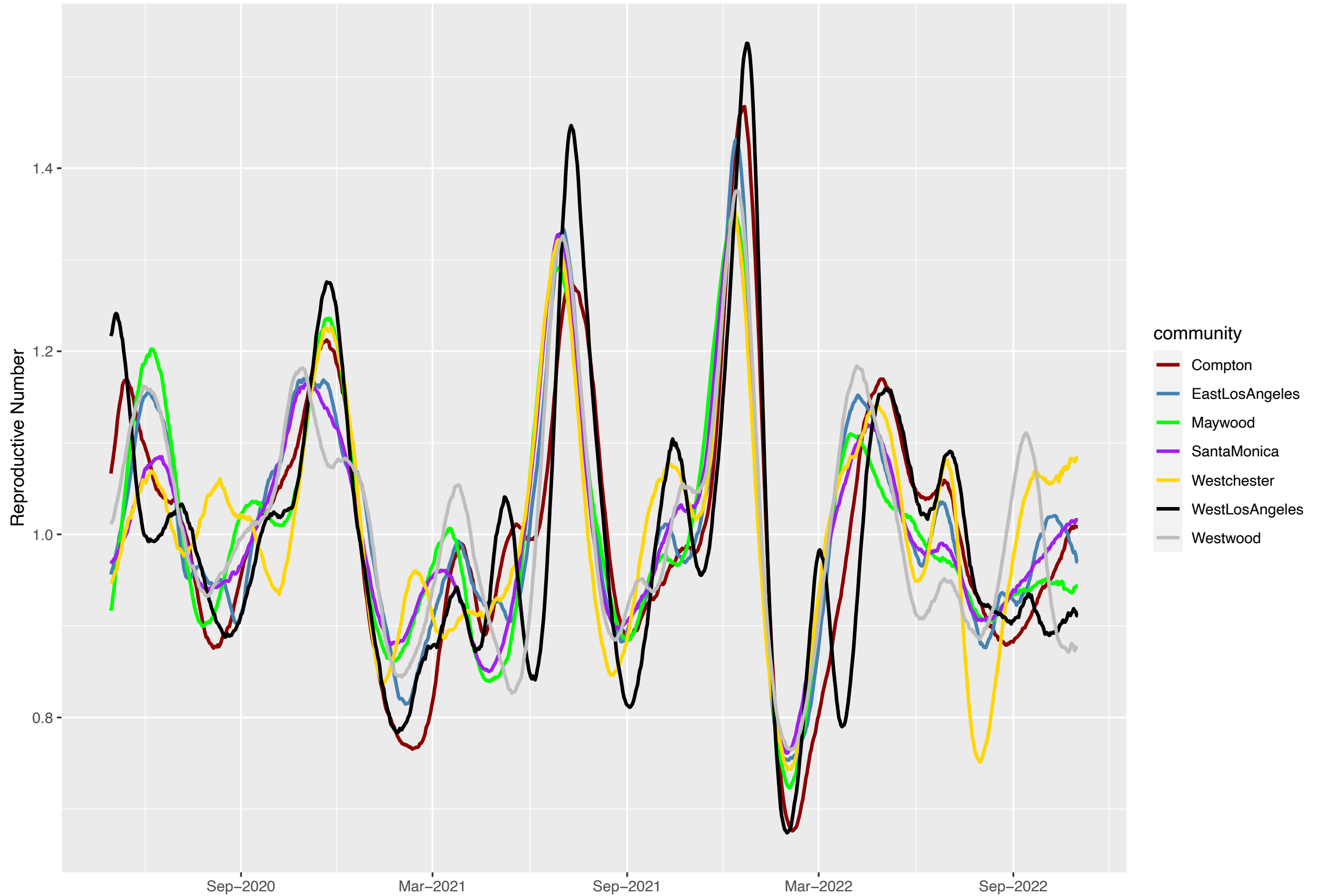
Example: Westwood (population 54109)



Example: East Los Angeles (population 125269)



Comparing communities



Comparing communities

- While there are strong similarities between communities within the county, there are also differences
- There is evidence of lags between communities
 - East Los Angeles lags West Los Angeles by 8-10 days
 - Westwood lags West Los Angeles by 8-10 days
 - Westwood is ahead of Santa Monica and Westchester by about 2 days

**Tracking all-cause-of-death and estimating excess mortality
during the COVID-19 pandemic:
shiny, accurate and secure**

What problem are we trying to solve?

- Quantifying mortality from COVID-19 is difficult, especially in countries with limited resources.
- The challenges may be in part due to:
 - underreporting (spatially, social, economic)
 - limited testing to confirm cases
 - inaccurate diagnosis or reporting
 - delays in determining or reporting cause of death
 - interactions with other causes of death (e.g., flu, heart disease).
- Comparing mortality data between countries is also challenging³² due to differences in methods for reporting COVID mortality

How are we solving it?

- Track over time **all-cause mortality** (ACM) and compare it with the expected ACM from pre-pandemic data
- This can provide an estimate of the overall burden of mortality directly or indirectly related to the COVID-19 pandemic
- Very useful for public health decision-making

Comparing actual to expected deaths in Philippines

All Cause Mortality for Male ≥ 80 in PYF data.xlsx during the Pandemic
deaths in 2020 compared to negative binomial regression on 2015–19

