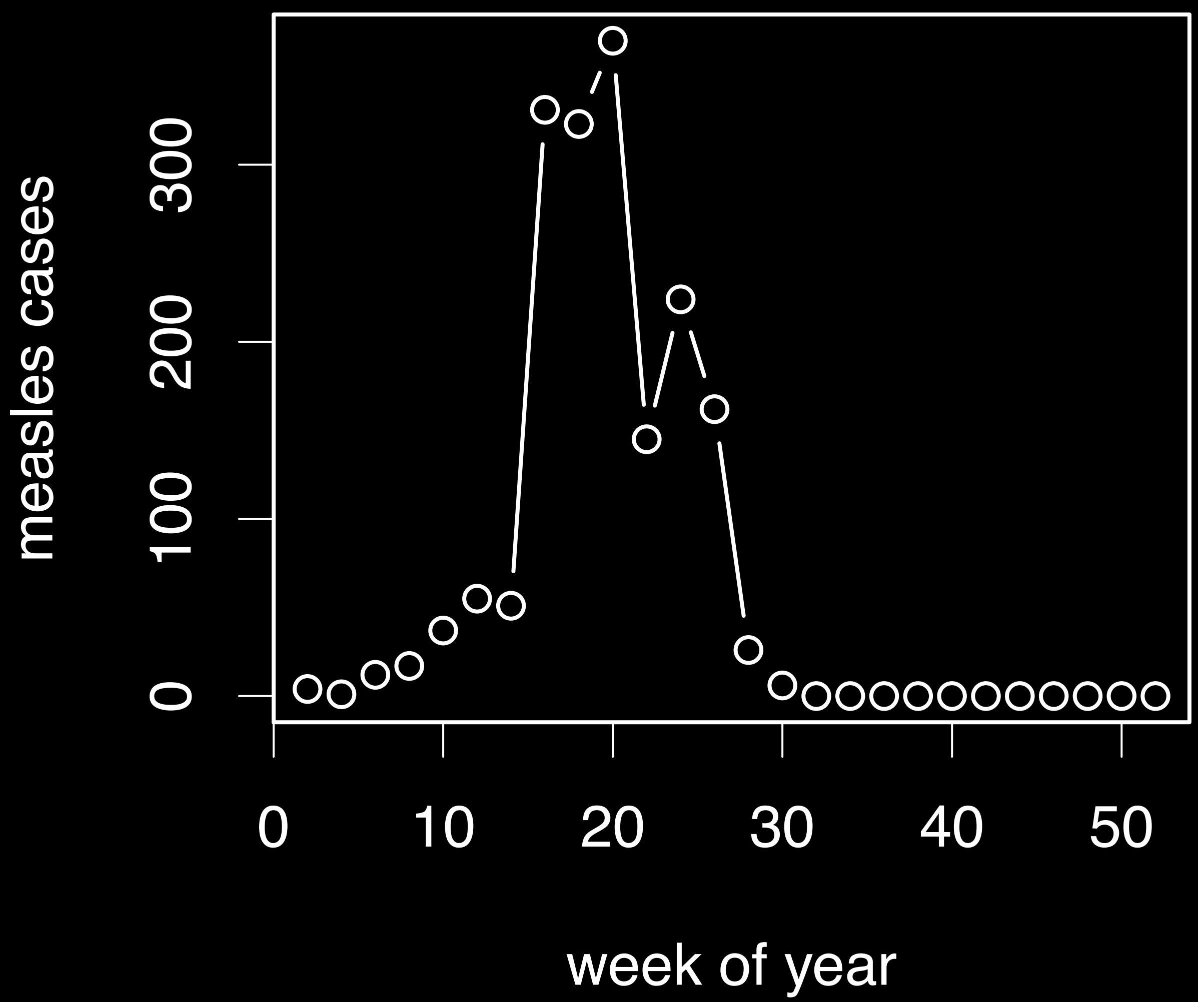
Model Fitting: The Basic Concept

Cara Brook

E^2M^2, Centre ValBio

Ranomafana National Park, Madagascar

What happened in Niamey?



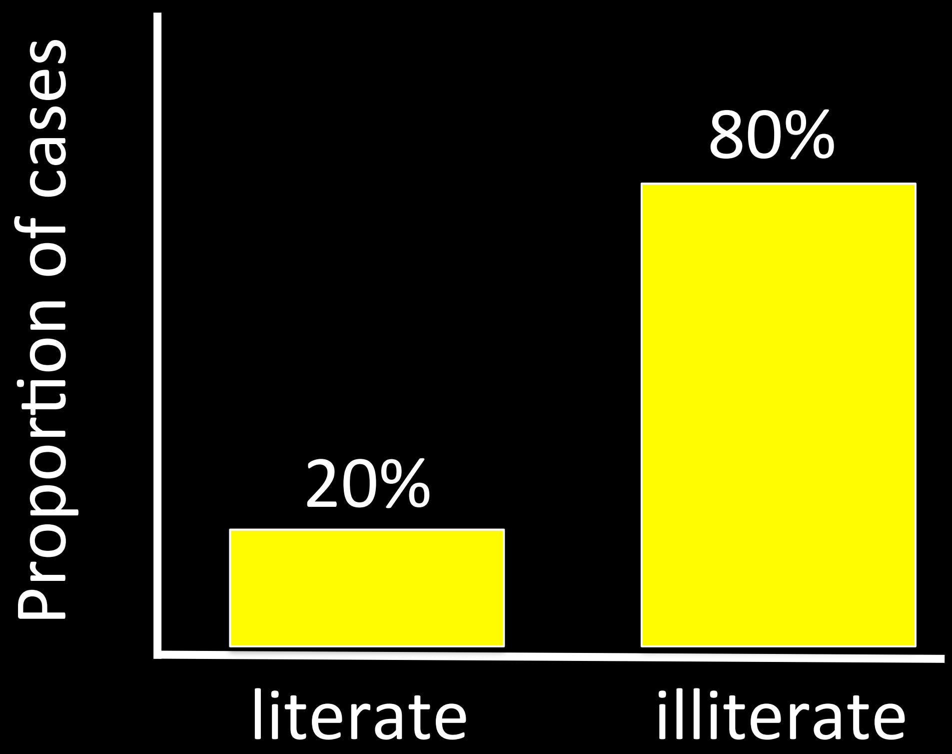
Traditional statistics is data-driven…

- We might ask questions about these data:
  - What proportion of cases occurred in males vs. females?
  - In winter vs. spring?
  - In Maradi vs. Dosso arrondisement?

- Goal: find correlations that imply causality

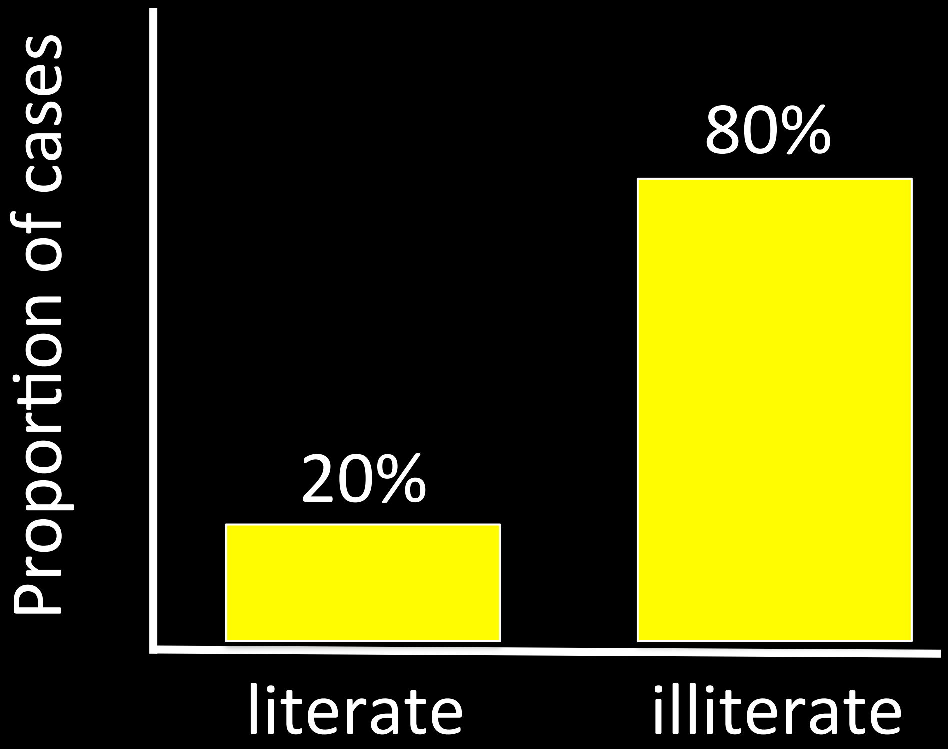
Traditional statistics is data-driven…

- Goal: find correlations that imply causality
- Imagine you discover:



Traditional statistics is data-driven…

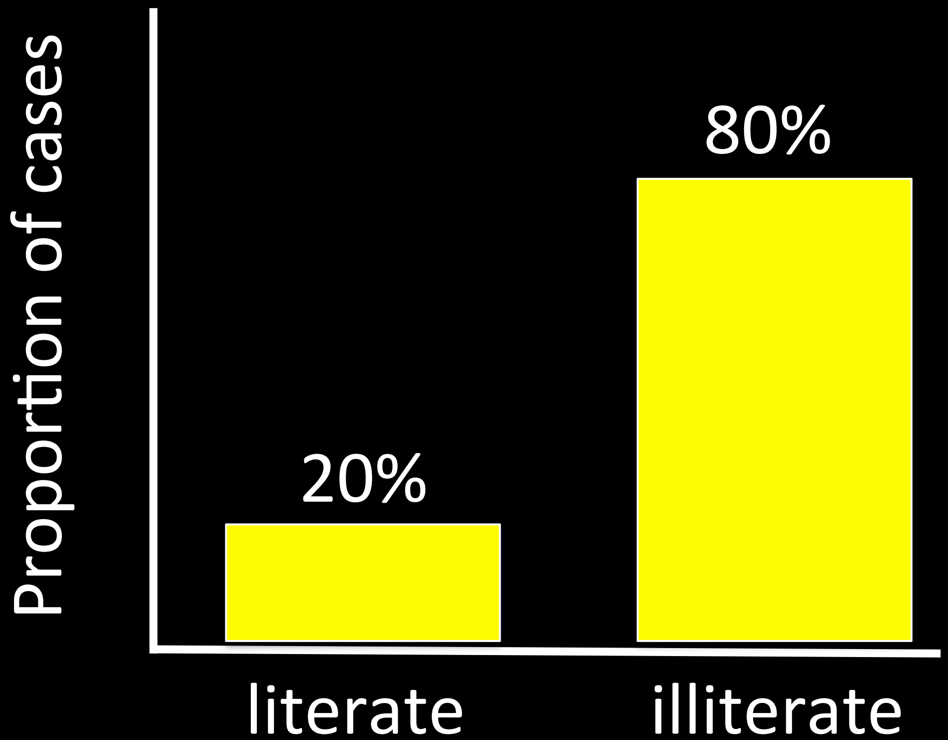
- Goal: find correlations that imply causality
- Imagine you discover:



Does literacy protect against measles?

Traditional statistics is data-driven…

- Goal: find correlations that imply causality
- Imagine you discover:



Must account for confounding, bias, and random error

- *Literacy is confounded with age!*

Mechanistic modeling is process-driven…

- We want to understand *what* happened, *when* it happened, and *why* it happened

Mechanistic modeling is process-driven…

- We want to understand *what* happened, *when* it happened, and *why* it happened

- Allows us to scale up from individual-level processes to population-level patterns

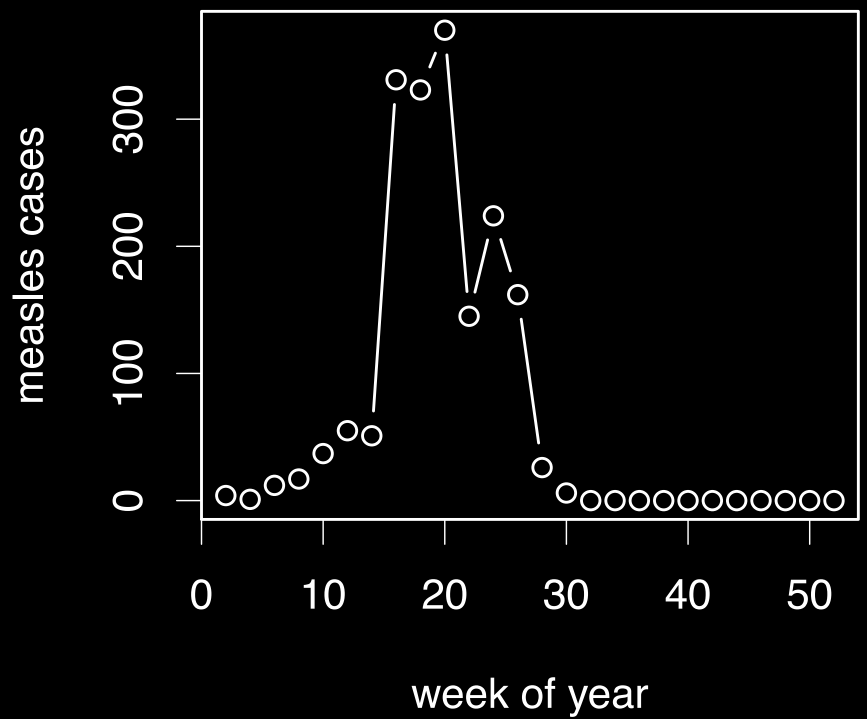
Mechanistic modeling is process-driven…

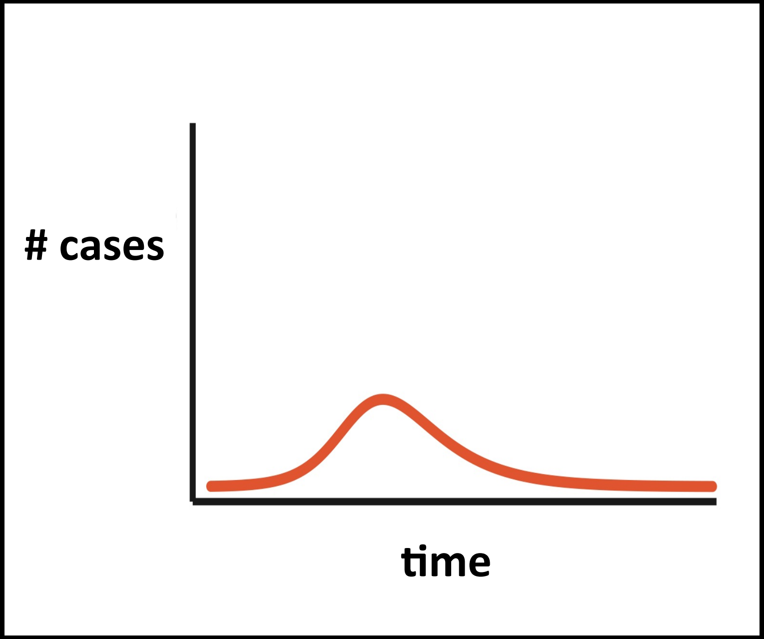
- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns
- We start by building a model that uses explicit processes to recover the same outcomes ("states") as our data

Mechanistic modeling is process-driven…

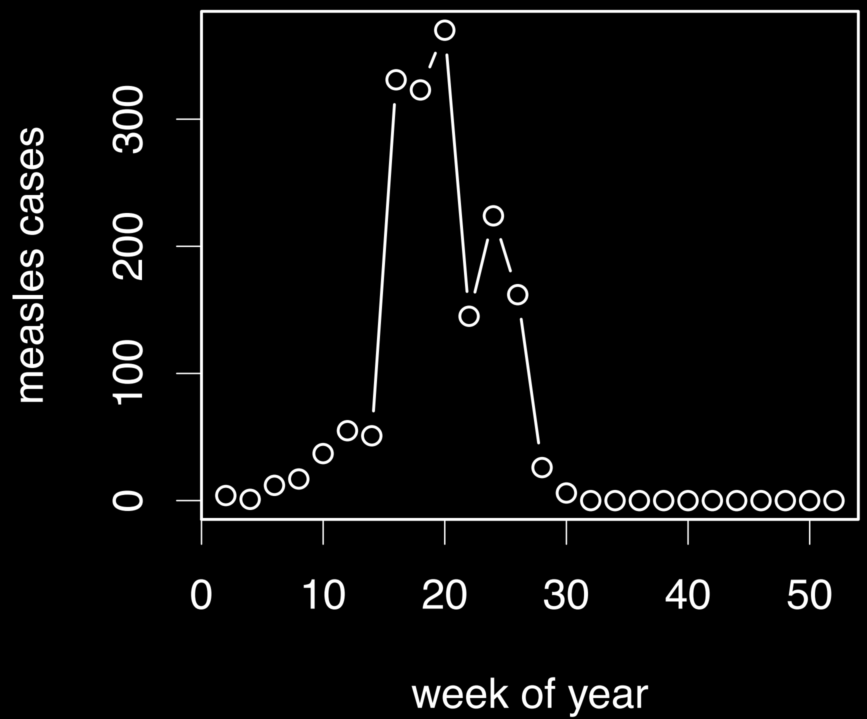
- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns
- We start by building a model that uses explicit processes to recover the same outcomes ("states") as our data
- *What state variables are captured in our data?*

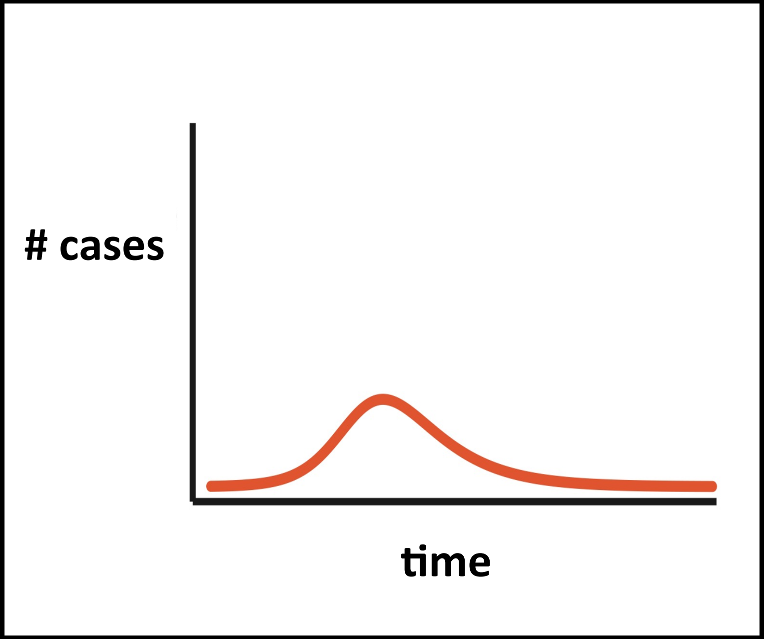
These data give us infecteds over time…

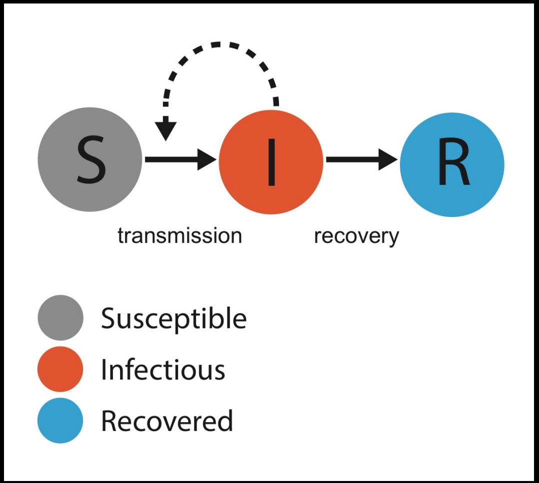




These data give us infecteds over time…







What *processes* contribute to the "infected" state in our system?

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.
2. Use any statistical tool (i.e. maximum likelihood, least squares) to ask, assuming our model is true, how likely are we to recover the observed data?
3. Optimize the parameters behind the processes to make the model most likely to recover the data.
4. If need be, restructure your model to better match your data.

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

Discrete time models are simple:

$$S_{t+1} = S_t - beta*I_t*S_t/N$$

$$I_{t+1} = I_t + beta*I_t*S_t/N - gamma*I_t$$

where beta = transmission coefficient and gamma = recovery rate

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

If we set the timestep = 1/gamma, we can reduce the system to:

$$S_{t+1} = S_t - beta*I_t*S_t/N_t$$

$$I_{t+1} = beta*I_t*S_t/N_t$$

This means we make the assumption that there are no overlapping infectious generations.

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

$I_2 = I_1 + beta*I_1*S_1/N - gamma*I_1$

$I_2 = 1 + (.2)(1)(9/10) - (1)(1)$	$I_1 = 1$ person $S_1 = 9$ persons $N = 10$ beta = .2 hour^{-1} gamma = 1 hour^{-1} $\Delta t = 1$ hour

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

$I_2 = I_1 + beta*I_1*S_1/N - gamma*I_1$

$I_2 = 1 + (.2)(1)(9/10) – (1)(1)$

$I_2 = 1 + (.2)(1)(9/10) – 1$

$I_2 = .18$ at 2 hours

cancel!

$I_1 = 1$ person
$S_1 = 9$ persons
$N = 10$
beta = .2 $hour^{-1}$
gamma = 1 $hour^{-1}$
$\Delta t = 1$ hour

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

$I_2 = I_1 + beta*I_1*S_1/N - gamma*I_1$

$I_2 = 1 + (.2)(1)(9/10) – (1)(1)$

$I_2 = 1 + (.2)(1)(9/10) – 1$

$I_2 = .18$ at 2 hours

cancel!

I_1 = 1 person
S_1 = 9 persons
N = 10
beta = .2 $hour^{-1}$
gamma = 1 $hour^{-1}$
Δt = 1 hour

$I_2 = 1 + (.2)(2)(1)(9/10) – (1)(1)(2)$

I_1 = 1 person
S_1 = 9 persons
N = 10
beta = .2 $hour^{-1}$
gamma = 1 $hour^{-1}$
Δt = 2 hours

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

$I_2 = I_1 + beta*I_1*S_1/N - gamma*I_1$

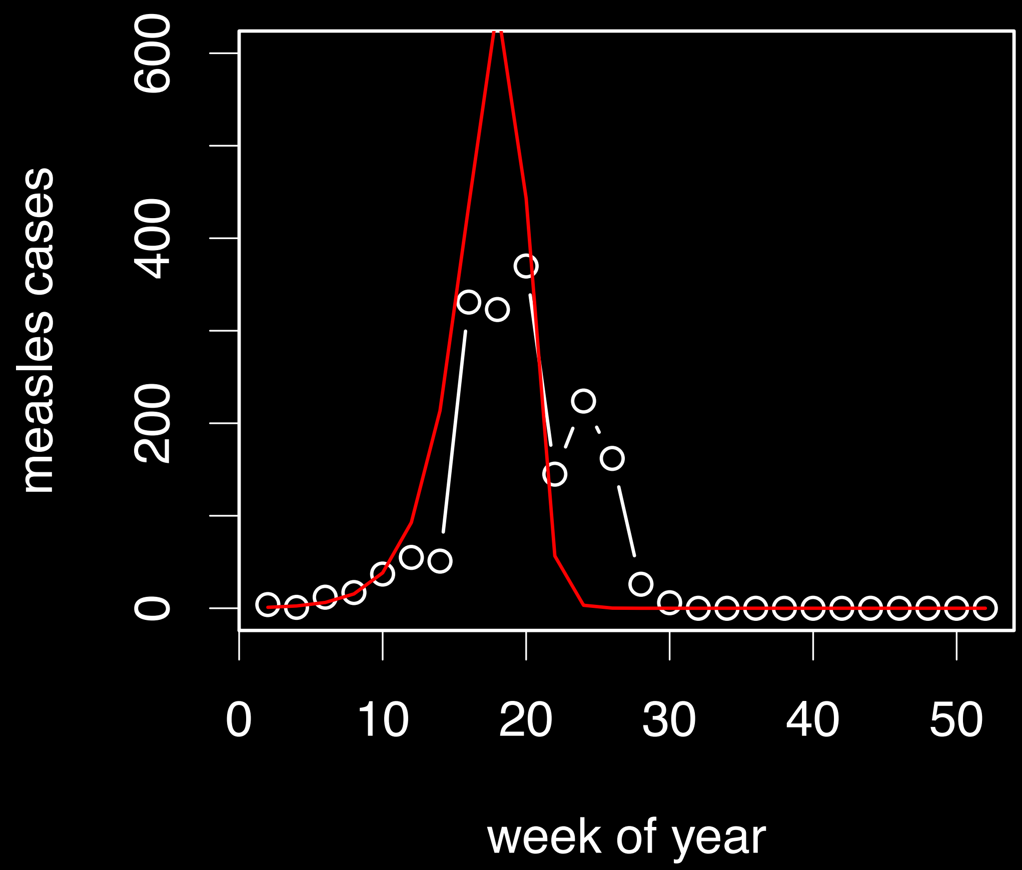
$I_2 = 1 + (.2)(1)(9/10) – (1)(1)$ $I_2 = 1 + (.2)(1)(9/10) – 1$ $I_2 = .18$ at 2 hours cancel!	I_1 = 1 person S_1 = 9 persons N = 10 beta = .2 $hour^{-1}$ gamma = 1 $hour^{-1}$ Δt = 1 hour

$I_2 = 1 + (.2)(2)(1)(9/10) – (1)(1)(2)$ $I_2 = 1 + (.2)(2)(1)(9/10) – 2$ $I_2 = -.64$ at 4 hours Don't cancel!	I_1 = 1 person S_1 = 9 persons N = 10 beta = .2 $hour^{-1}$ gamma = 1 $hour^{-1}$ Δt = 2 hours

How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

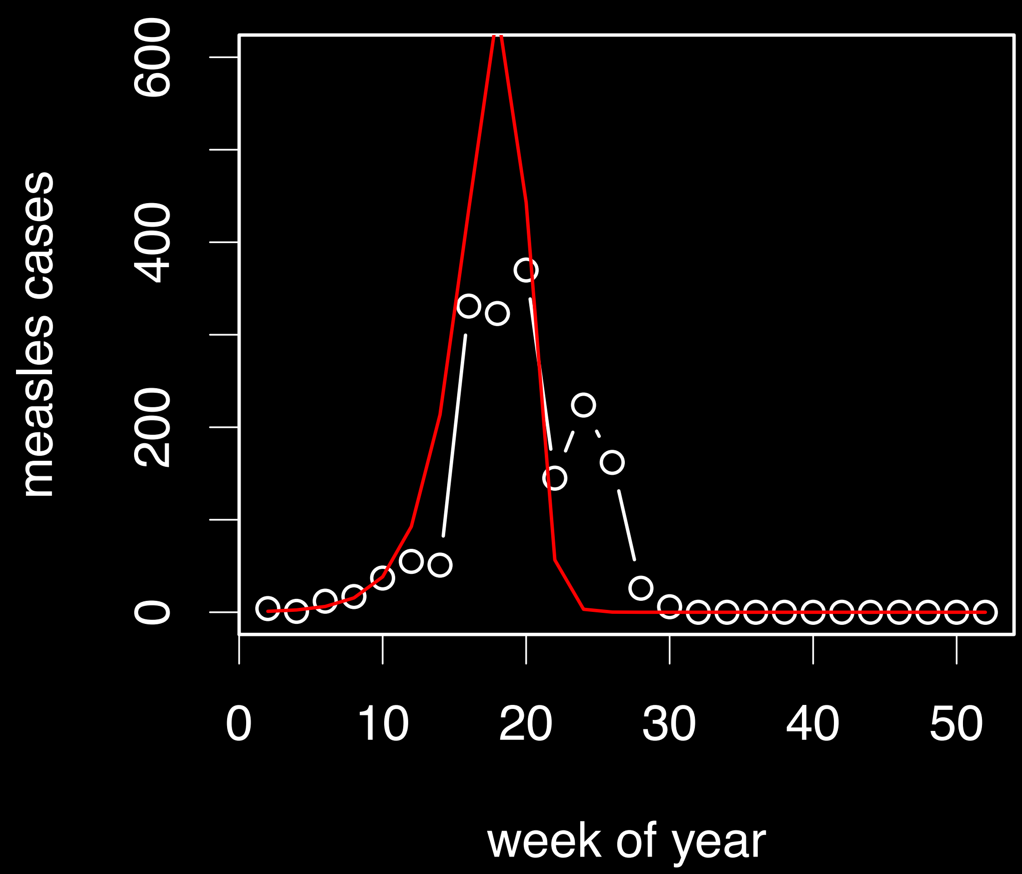
simA = det.SIR(par = c(2000, 2.5))



How to fit a model to data

1. Build a model that uses explicit processes to recover the same states as the data.

simA = det.SIR(par = c(2000, 2.5))



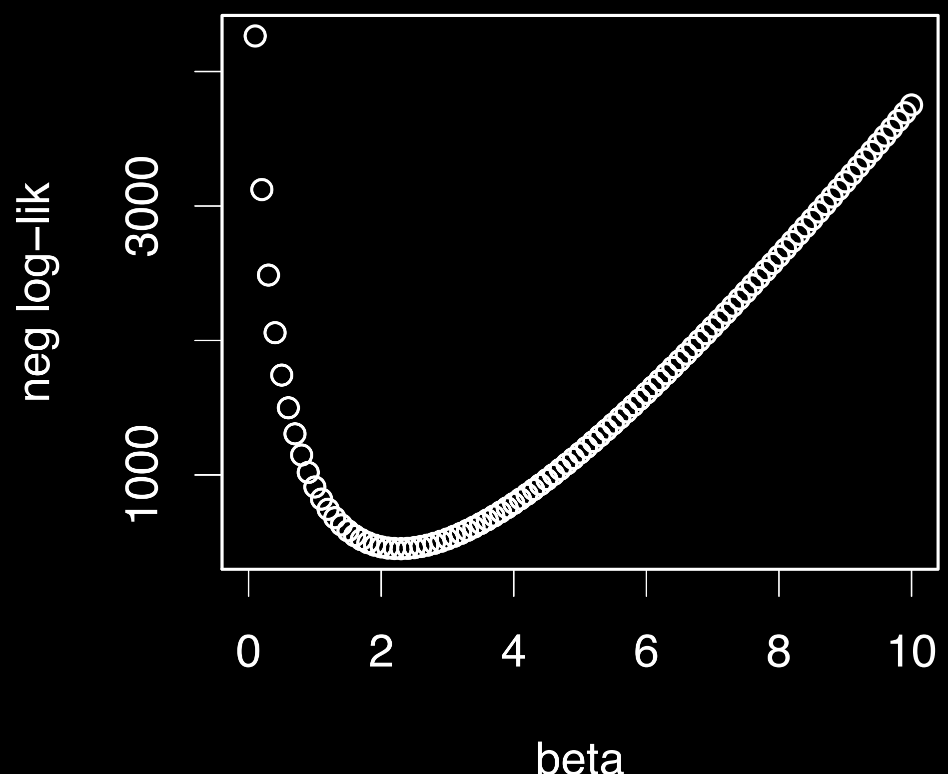
Model does a pretty good job, but but it overshoots our data by quite a bit.

What does this suggest about our guess for beta?

How to fit a model to data

2. Use any maximum likelihood to ask, assuming our model is true, how likely are we to recover the observed data?

beta_guess[which.min(ll)] = 2.3

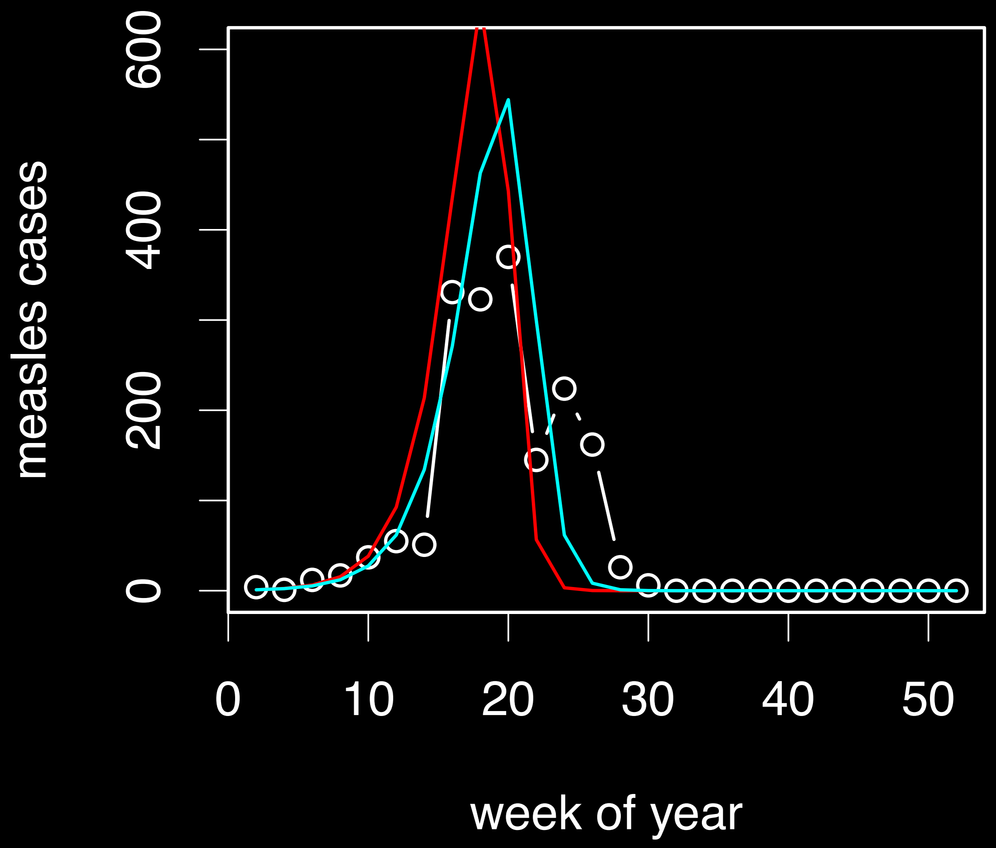


How to fit a model to data

2. Use any maximum likelihood to ask, assuming our model is true, how likely are we to recover the observed data?

simB = det.SIR(par = c(2000, 2.3))

New beta fits even better!



How to fit a model to data

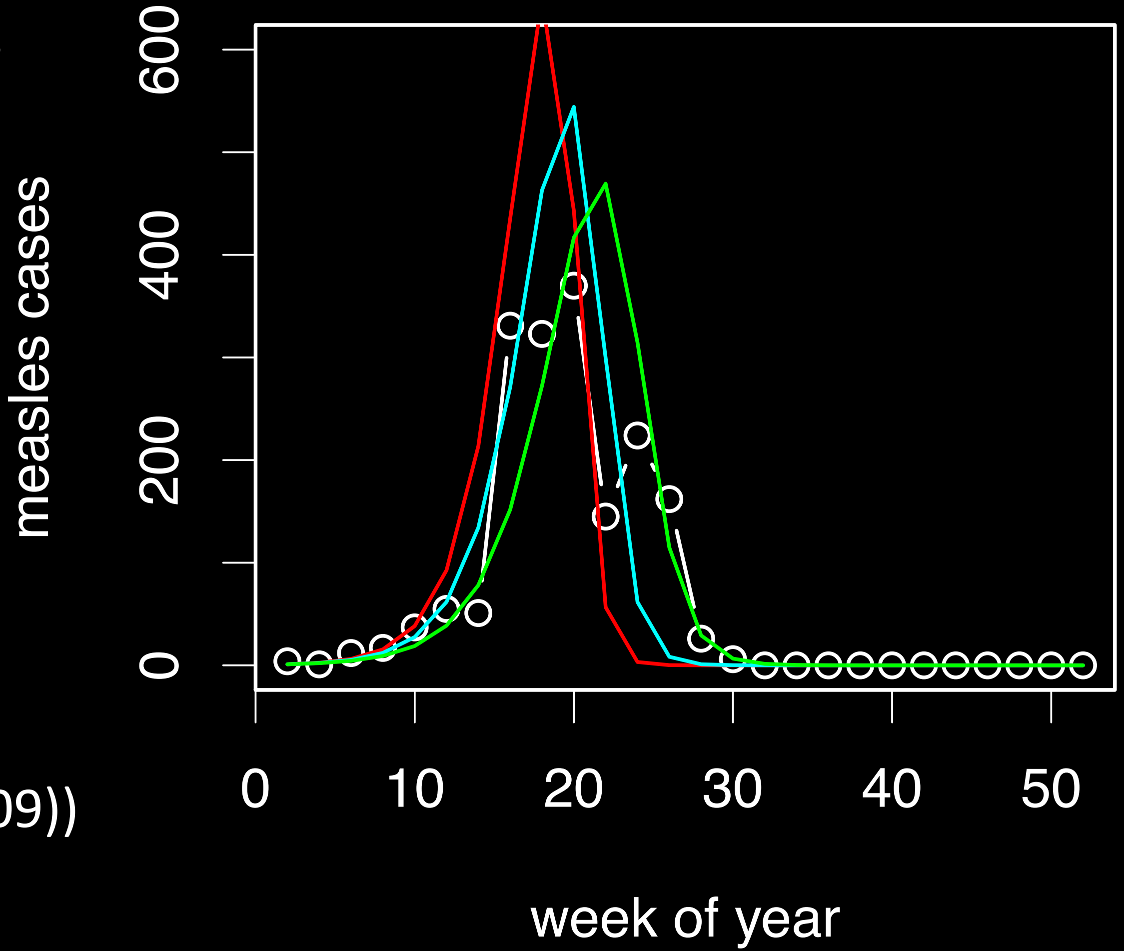
3. Optimize the parameters behind the processes to make the model most likely to recover the data.

```
out <- optim(par = c(2000, 2.5),
      likelihood,
      method = "Nelder-Mead",
      I = dat$cases)

out$par = c(2152, 2.09)
```

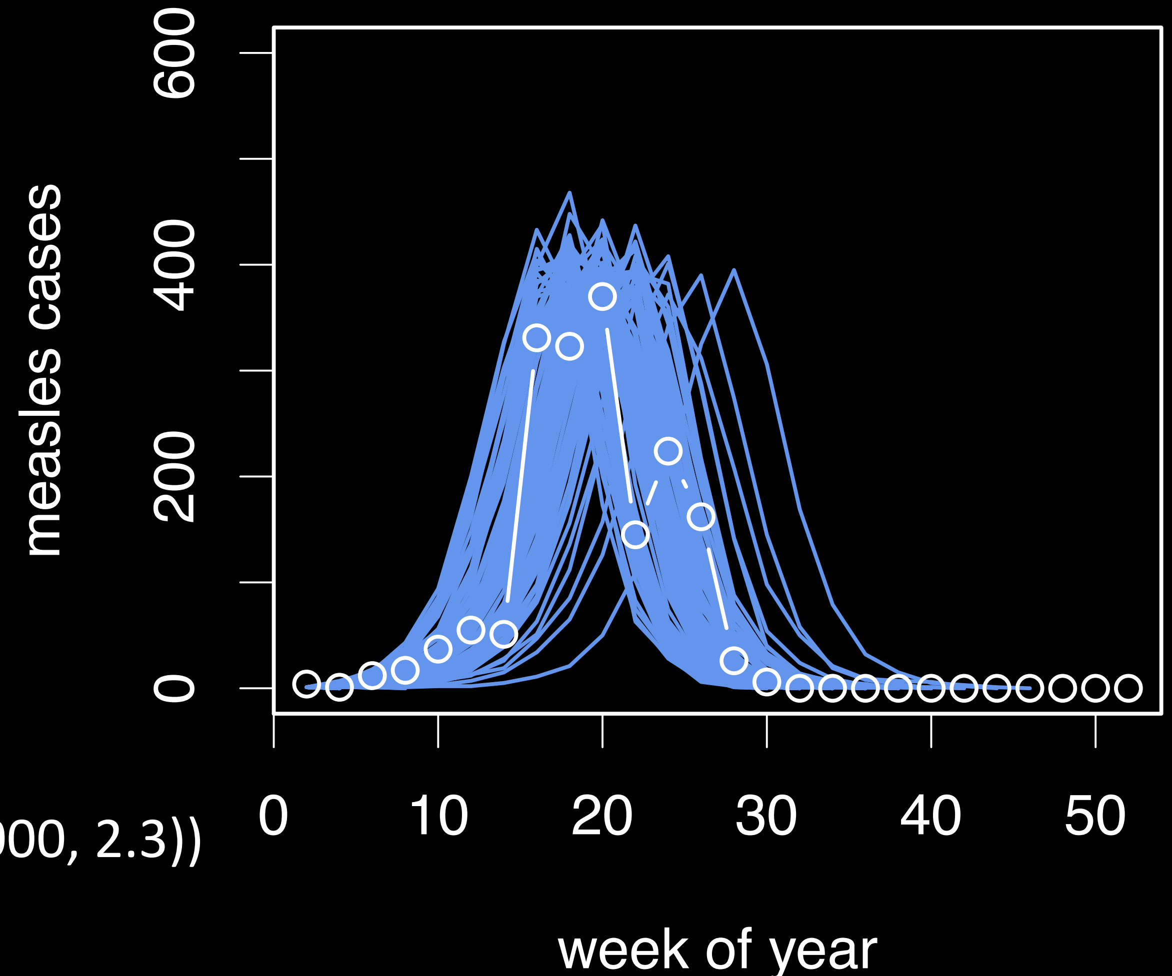
Looks great!
But why not perfect?

```
simC = det.SIR(par = c(2152, 2.09))
```



How to fit a model to data

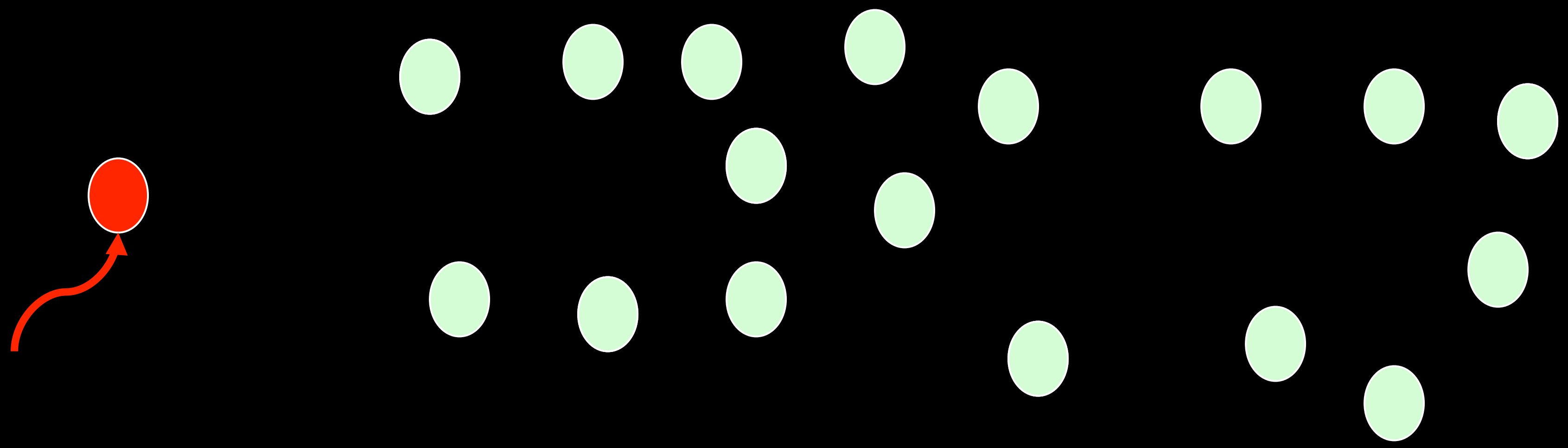
4. If need be, restructure your model to better match your data.




```
for (i in 1:100) {
   stoch = stoch.SIR(par=c(2000, 2.3))
}
```

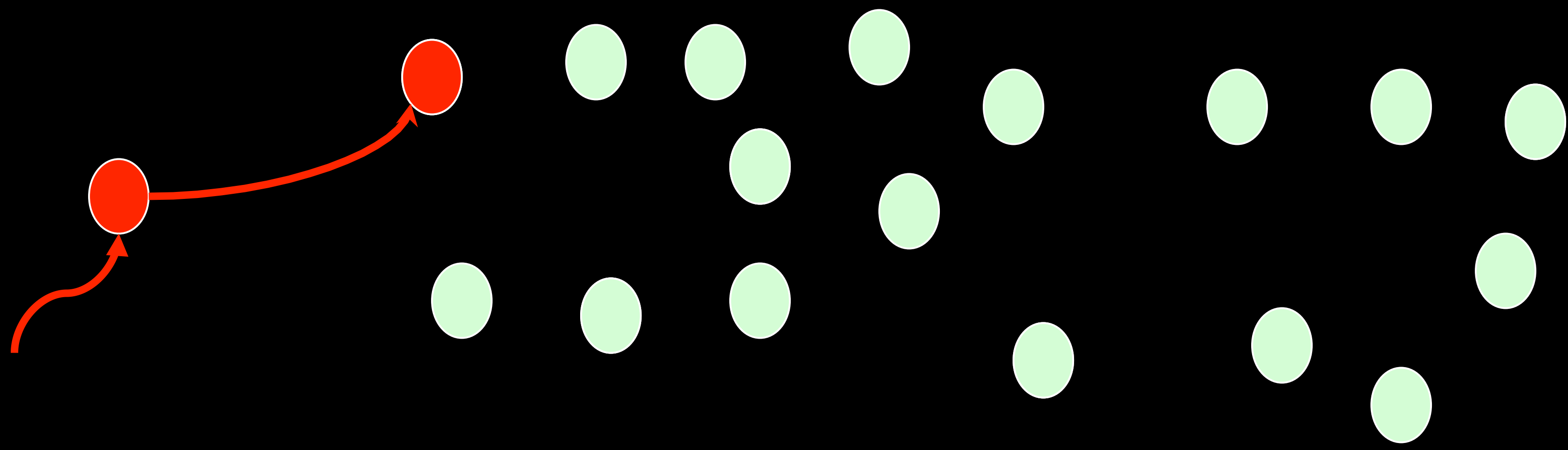
Insights from fitting dynamic models: R0

- The basic reproduction number for a pathogen
- Defined as the number of new infections generated by one existing infection in a completely susceptible host population
- In these discrete time models, R0 = beta



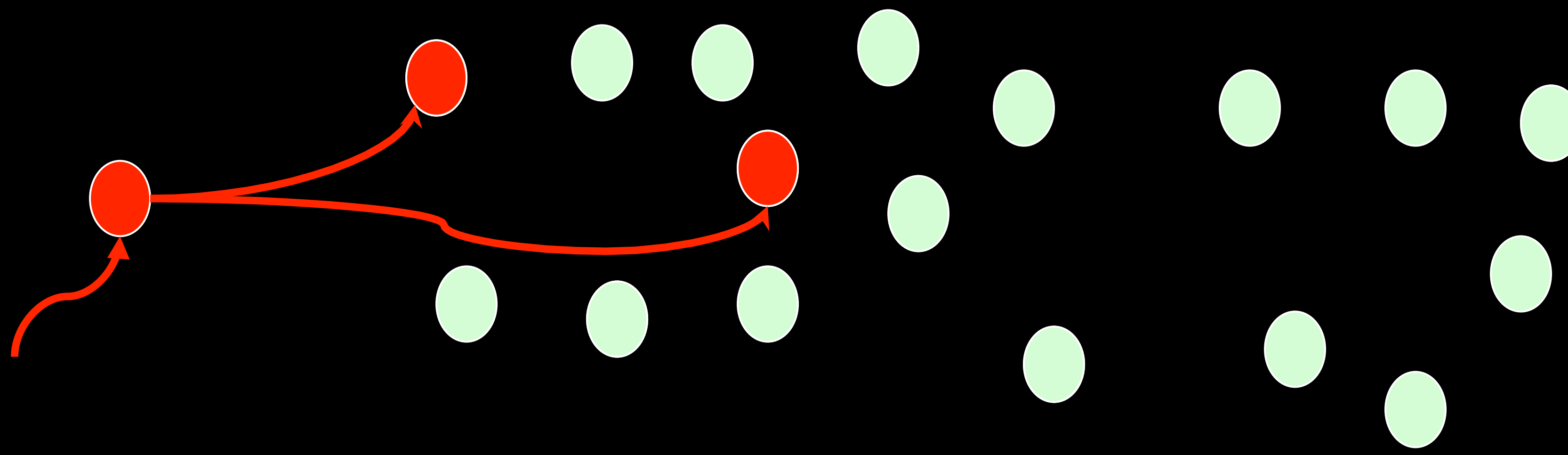
Insights from fitting dynamic models: R0

- The basic reproduction number for a pathogen
- Defined as the number of new infections generated by one existing infection in a completely susceptible host population
- In these discrete time models, R0 = beta



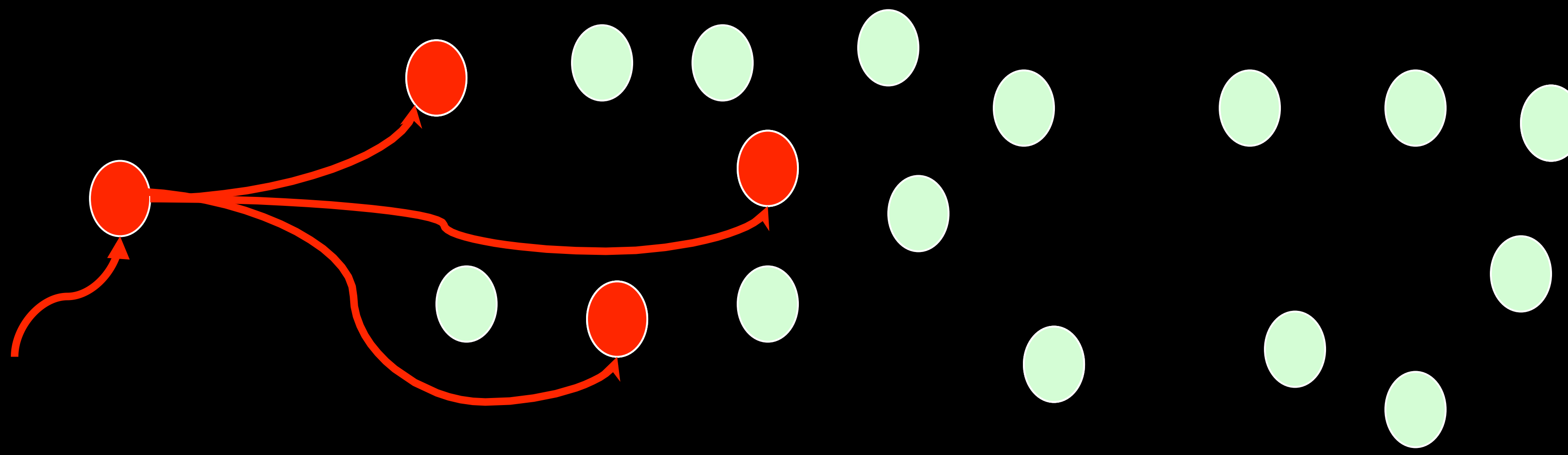
Insights from fitting dynamic models: R0

- The basic reproduction number for a pathogen
- Defined as the number of new infections generated by one existing infection in a completely susceptible host population
- In these discrete time models, R0 = beta



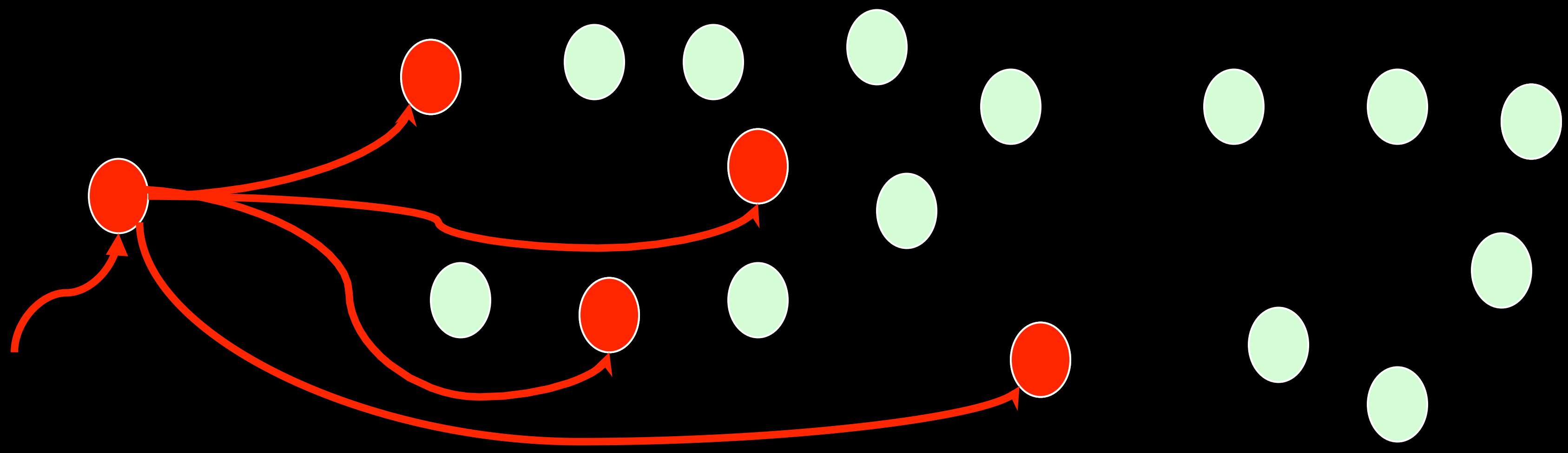
Insights from fitting dynamic models: R0

- The basic reproduction number for a pathogen
- Defined as the number of new infections generated by one existing infection in a completely susceptible host population
- In these discrete time models, R0 = beta



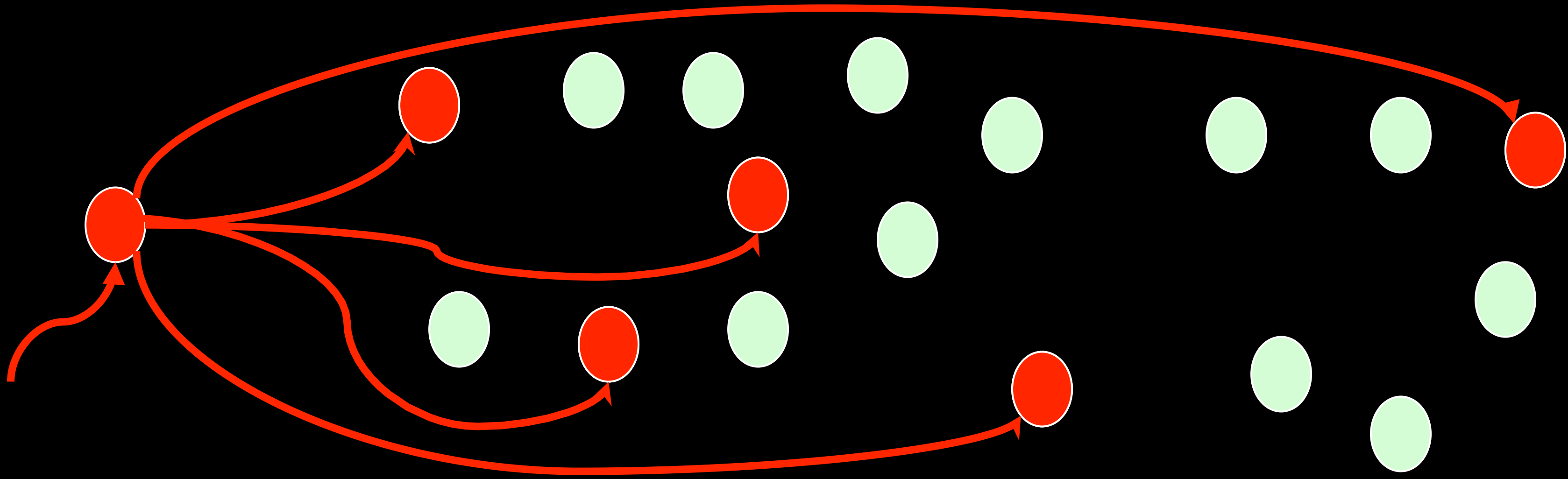
Insights from fitting dynamic models: R0

- The basic reproduction number for a pathogen
- Defined as the number of new infections generated by one existing infection in a completely susceptible host population
- In these discrete time models, R0 = beta



Insights from fitting dynamic models: R0

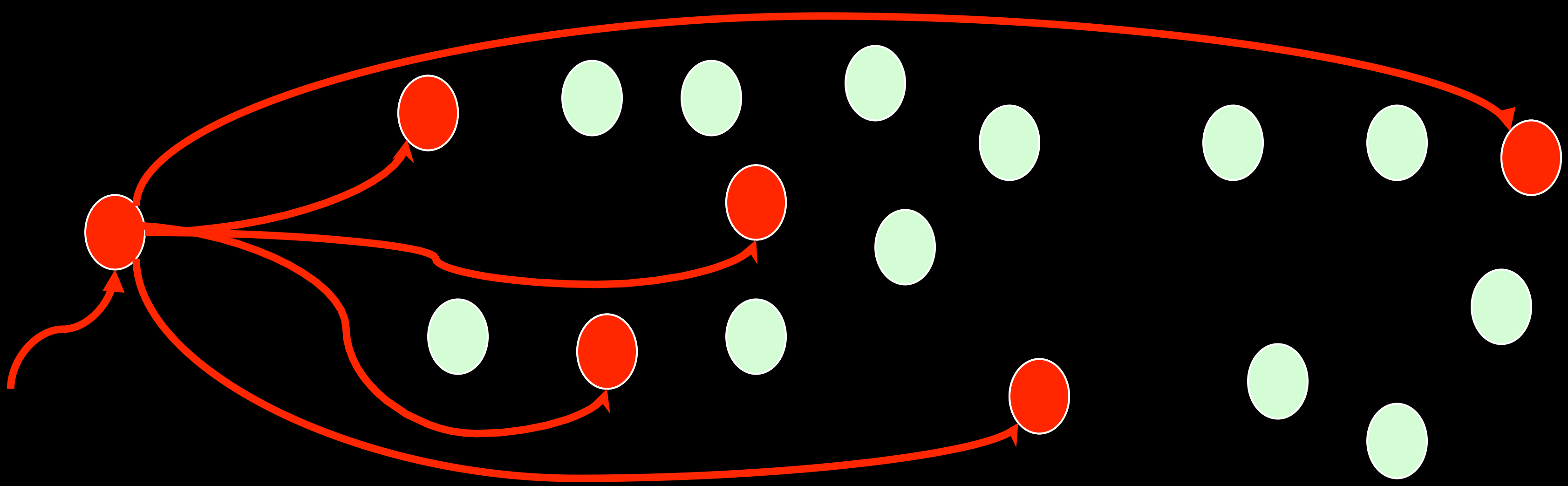
- The basic reproduction number for a pathogen
- Defined as the number of new infections generated by one existing infection in a completely susceptible host population
- In these discrete time models, R0 = beta



$R_0 = 5$

Insights from fitting dynamic models: R-effective

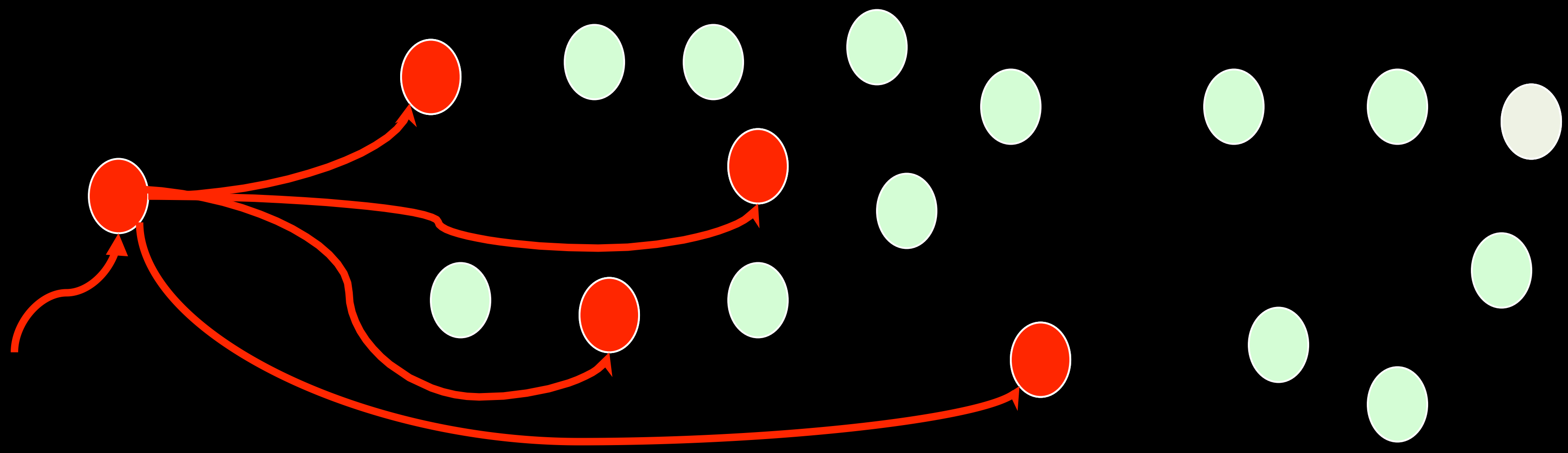
- The number of new infections generated by one existing infection in a partially susceptible host population
- Reff = R0*(S/N)
- Changes over time!



$R_{eff} = (5)*(12/17) = 3.53$

Insights from fitting dynamic models: R-effective

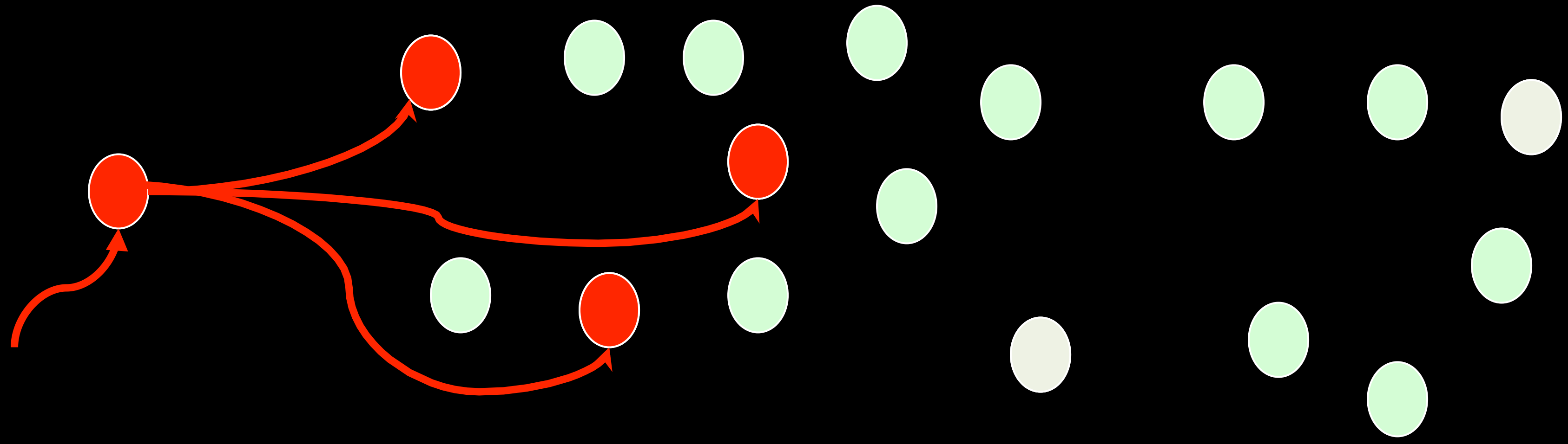
- The number of new infections generated by one existing infection in a partially susceptible host population
- Reff = R0*(S/N)
- Changes over time!



$R_{eff} = (5)*(13/17) = 3.82$

Insights from fitting dynamic models: R-effective

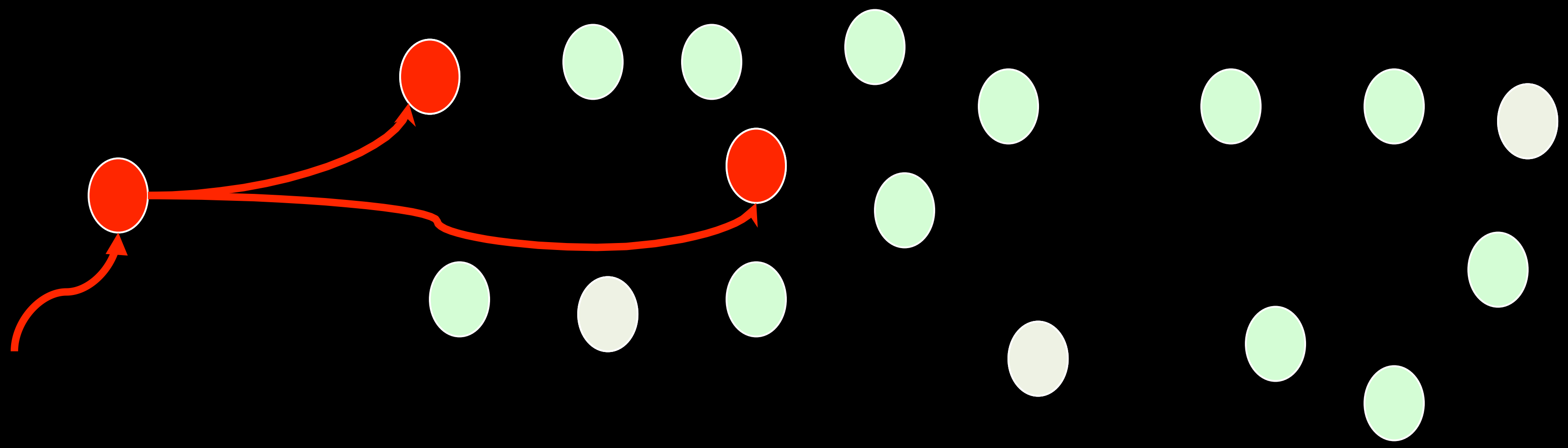
- The number of new infections generated by one existing infection in a partially susceptible host population
- Reff = R0*(S/N)
- Changes over time!



$R_{eff} = (5)*(14/17) = 4.12$

Insights from fitting dynamic models: R-effective

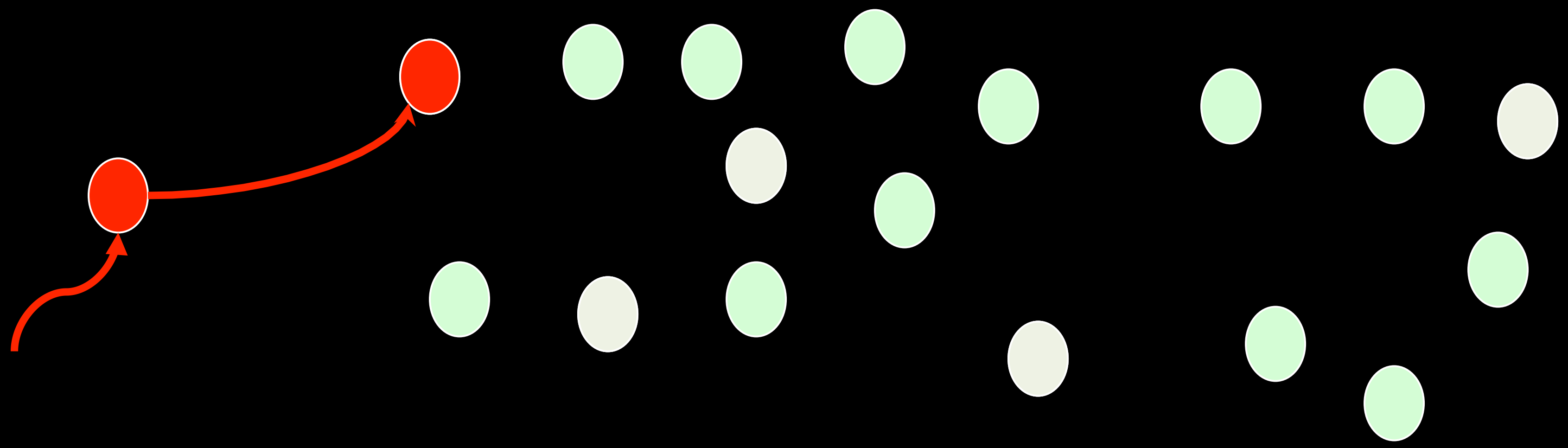
- The number of new infections generated by one existing infection in a partially susceptible host population
- Reff = R0*(S/N)
- Changes over time!



R_{eff} = (5)*(15/17) = 4.41

Insights from fitting dynamic models: R-effective

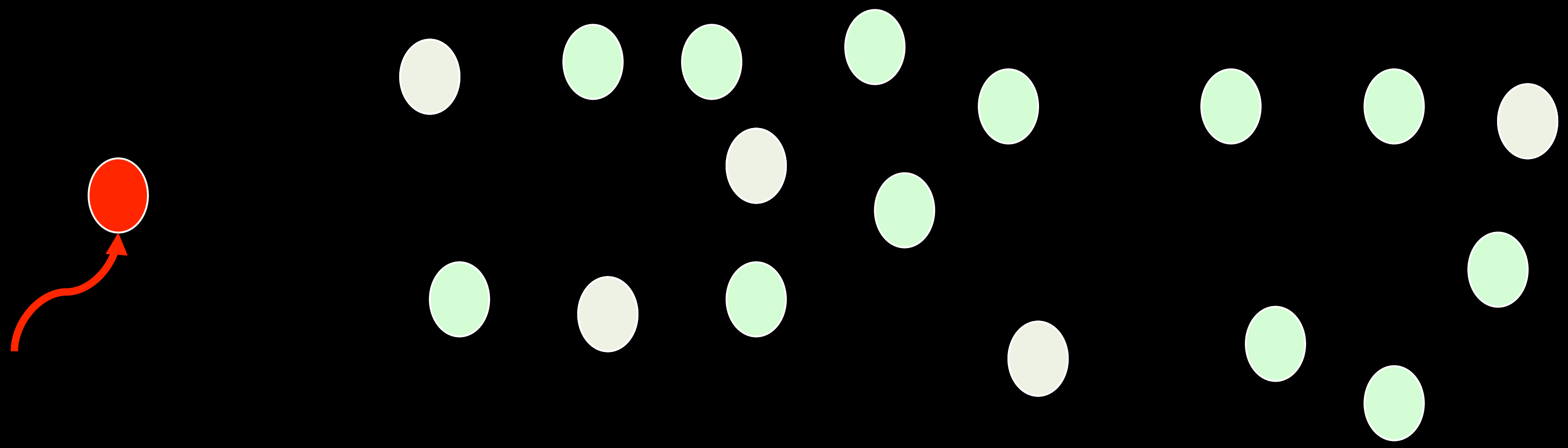
- The number of new infections generated by one existing infection in a partially susceptible host population
- Reff = R0*(S/N)
- Changes over time!



$R_{eff} = (5)*(16/17) = 4.71$

Insights from fitting dynamic models: R-effective

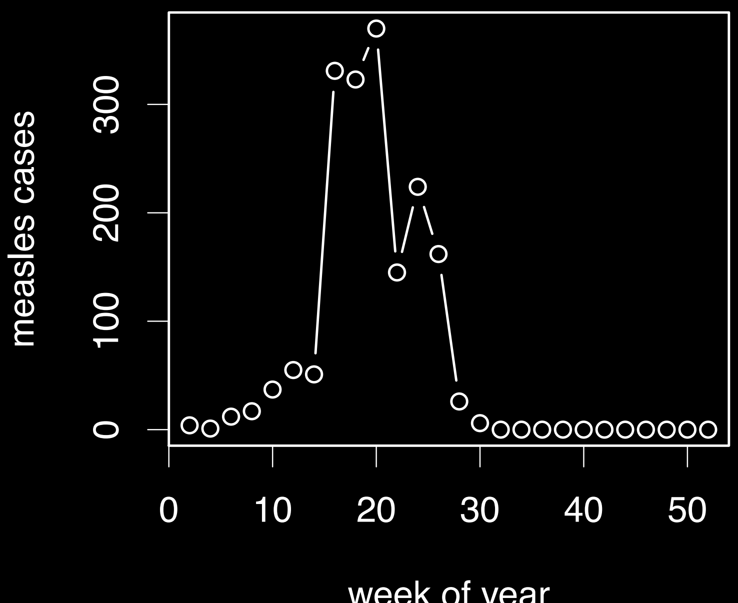
- The number of new infections generated by one existing infection in a partially susceptible host population
- Reff = R0*(S/N)
- Changes over time!

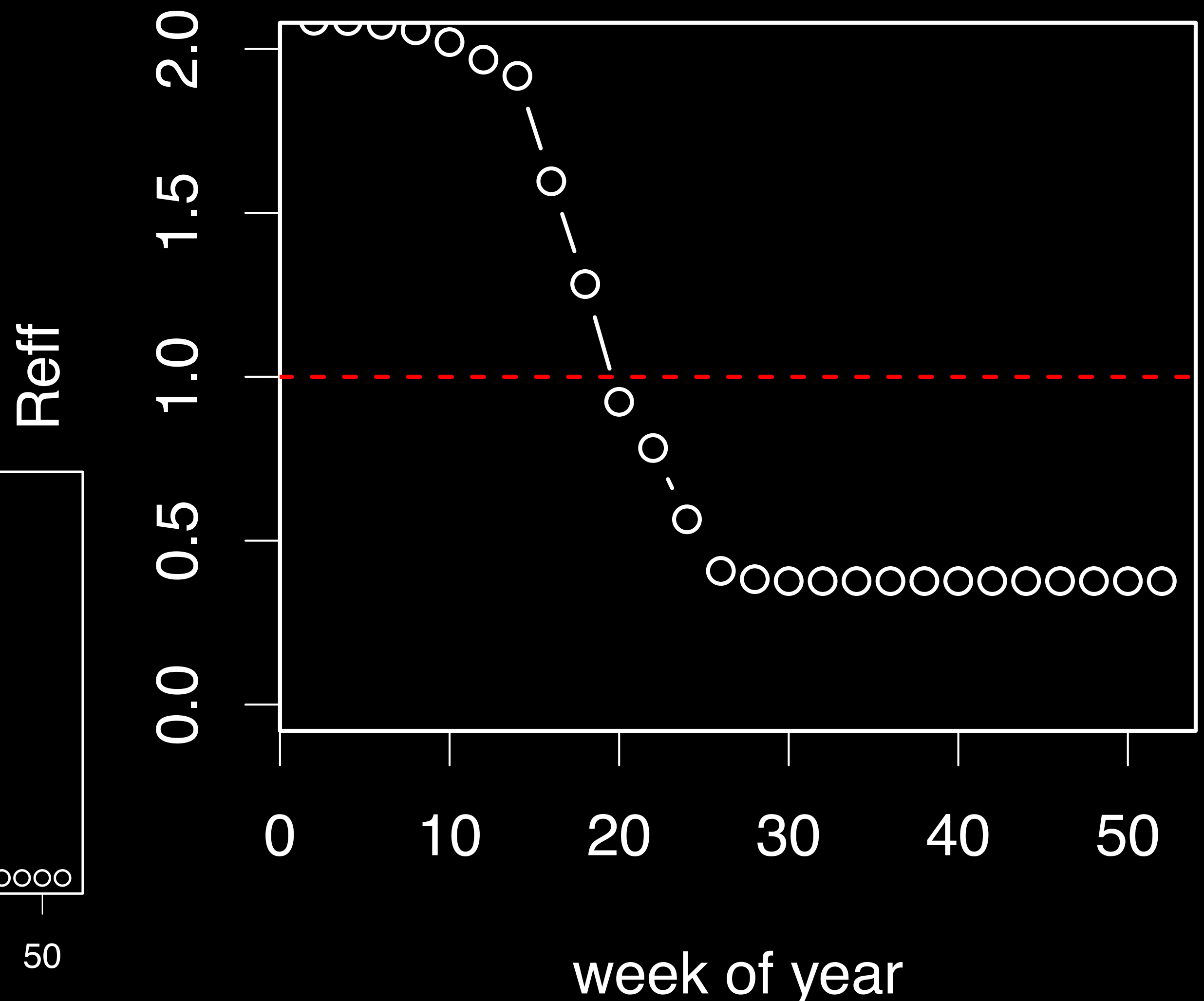


$R_{eff} = (5)*(17/17) = 5$

Insights from fitting dynamic models: R-effective

An epidemic spreads when Reff > 1 and declines when Reff < 1



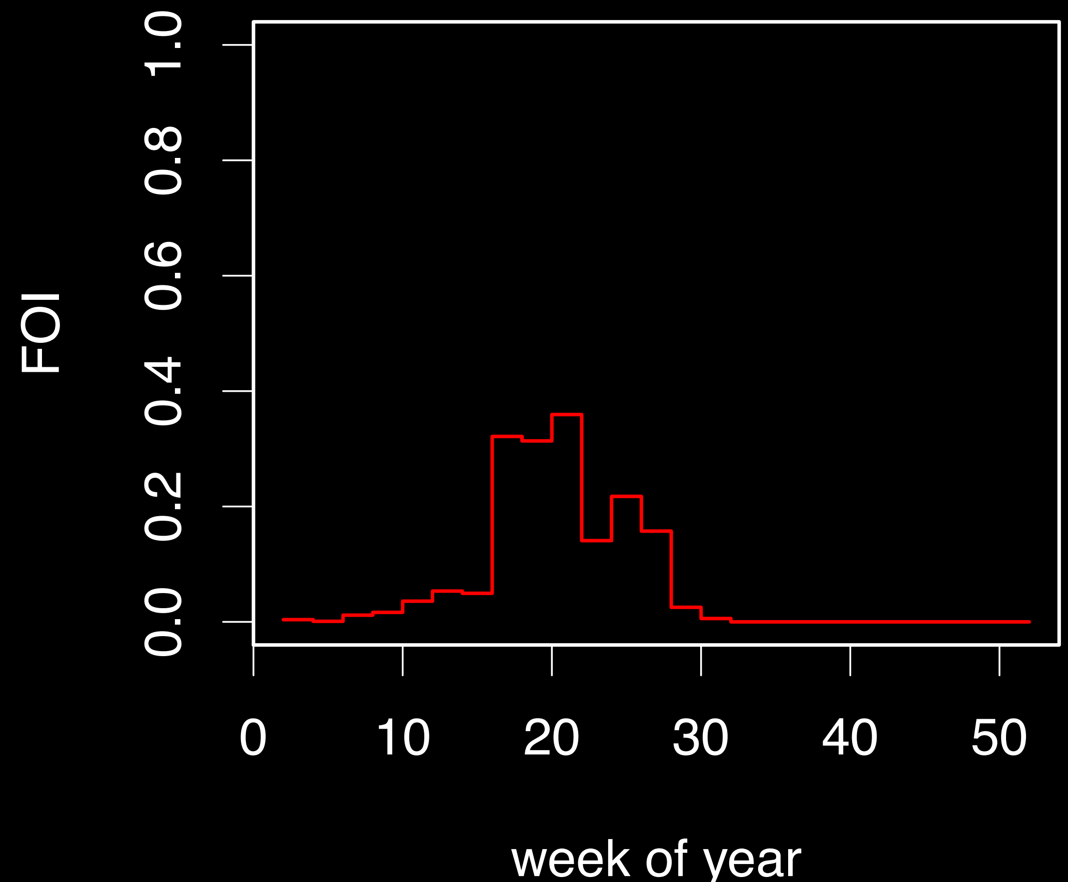


Insights from fitting dynamic models: Force of Infection

- Defined as the rate at which susceptibles become infected

- FOI = R0*(I/N)

- Changes over time, but constant across each timestep



Mechanistic modeling is process-driven…

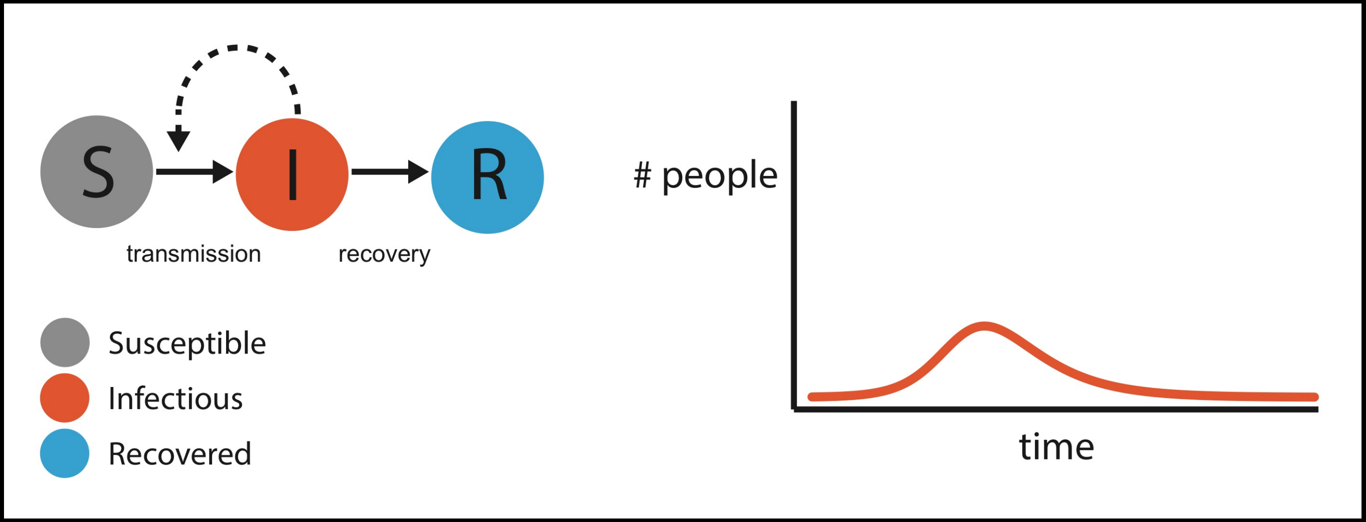
- We want to understand *what* happened, *when* it happened, and *why* it happened
- Allows us to scale up from individual-level processes to population-level patterns
- We start by building a model that uses explicit processes to recover the same outcomes ("states") as our data

Mechanistic modeling is process-driven…

- Test “what if” scenarios not amenable to experimentation

Mechanistic modeling is process-driven…

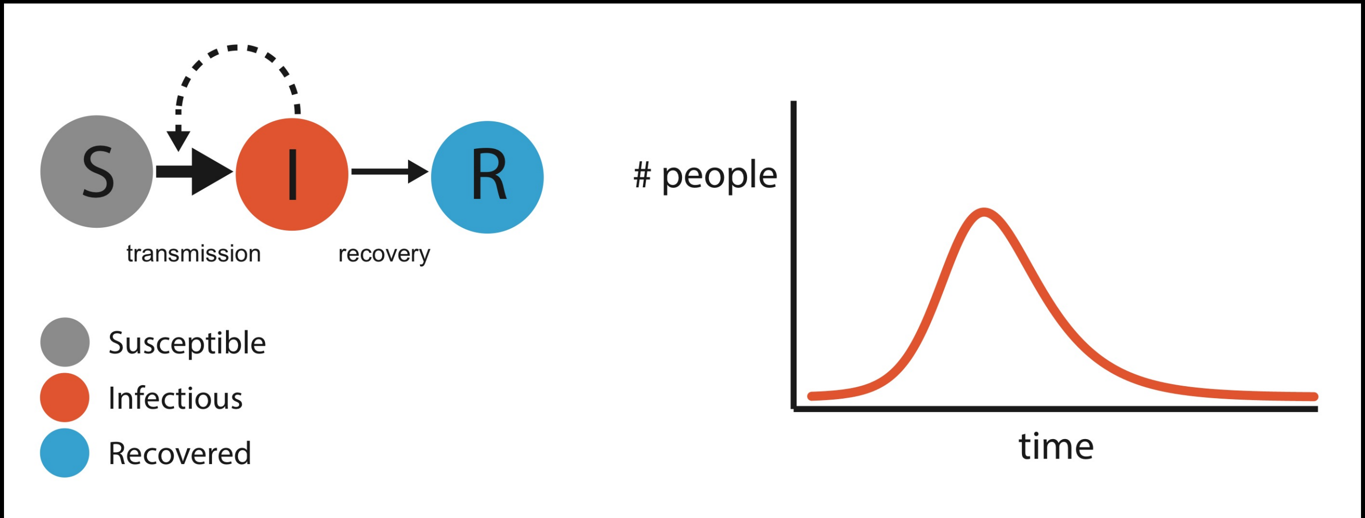
- Test "what if" scenarios not amenable to experimentation



Mechanistic modeling is process-driven…

- Test "what if" scenarios not amenable to experimentation

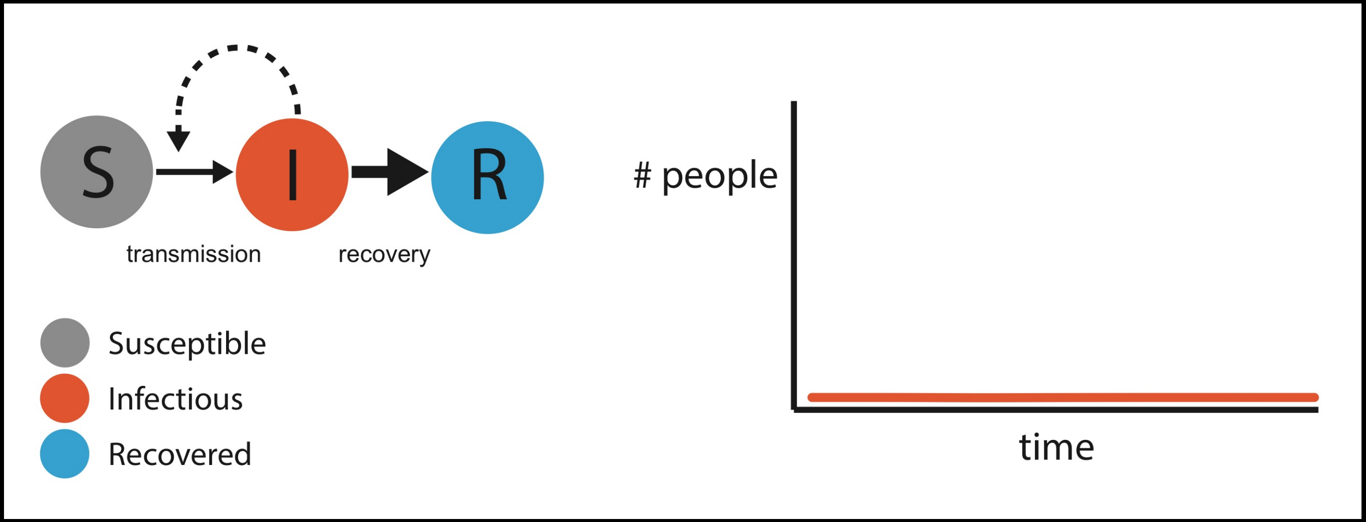
What if each person exposed 50% more people?



Mechanistic modeling is process-driven…

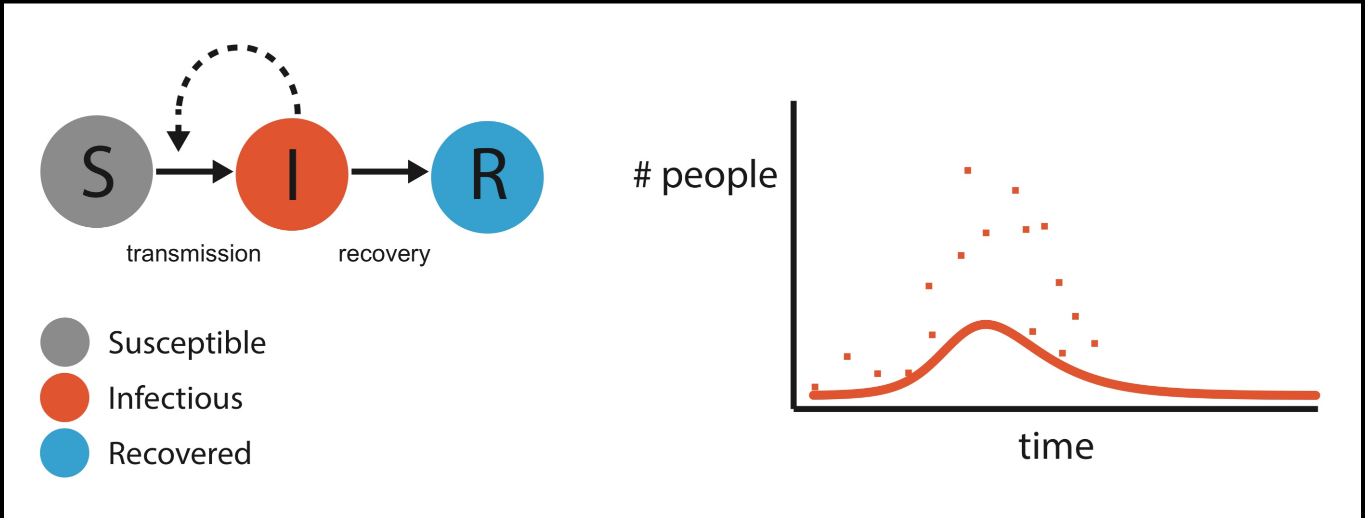
- Test "what if" scenarios not amenable to experimentation

What if we treated people and doubled the rate of recovery?



Mechanistic modeling is process-driven…

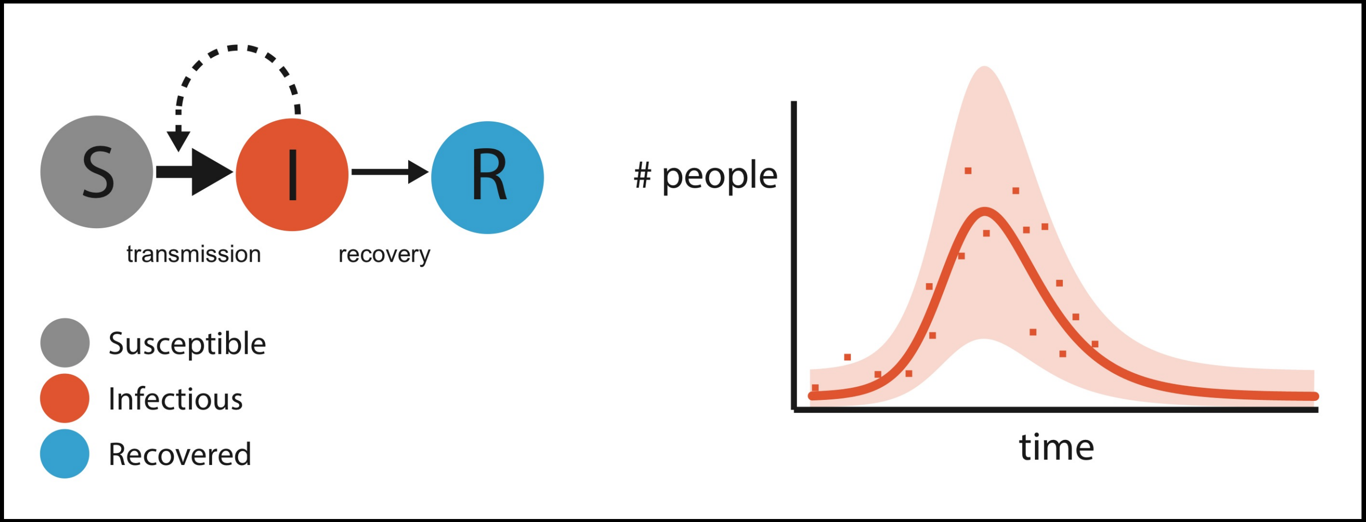
- Test “what if” scenarios not amenable to experimentation
- Estimate parameters that are difficult to measure by fitting models to available data



Mechanistic modeling is process-driven…

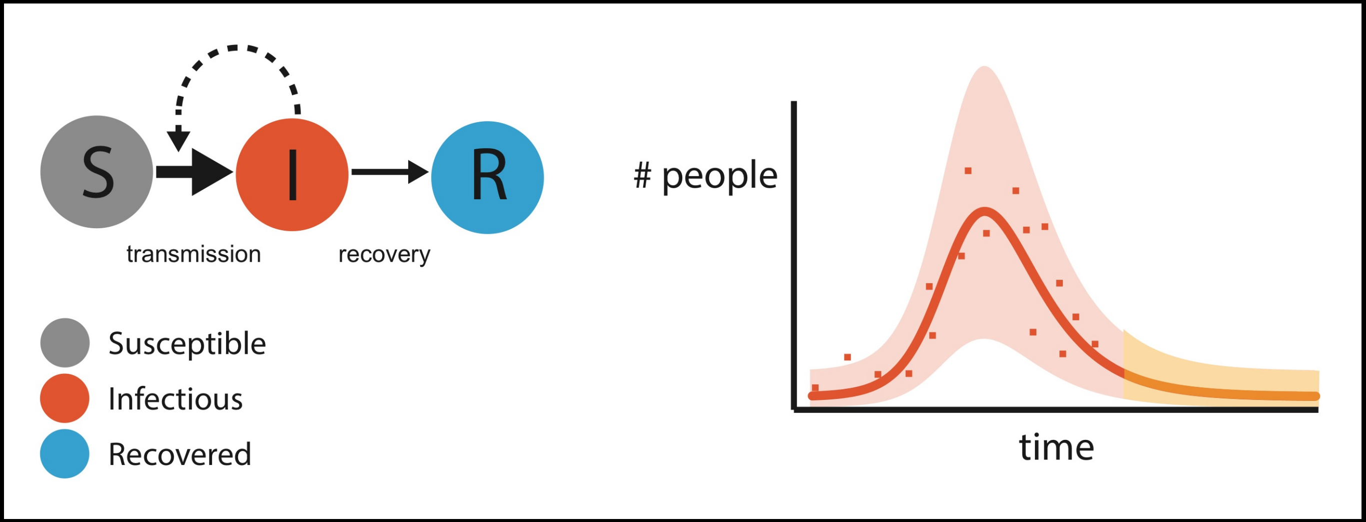
- Test "what if" scenarios not amenable to experimentation
- Estimate parameters that are difficult to measure by fitting models to available data

Estimate transmission rate or other model parameters
(with confidence intervals)



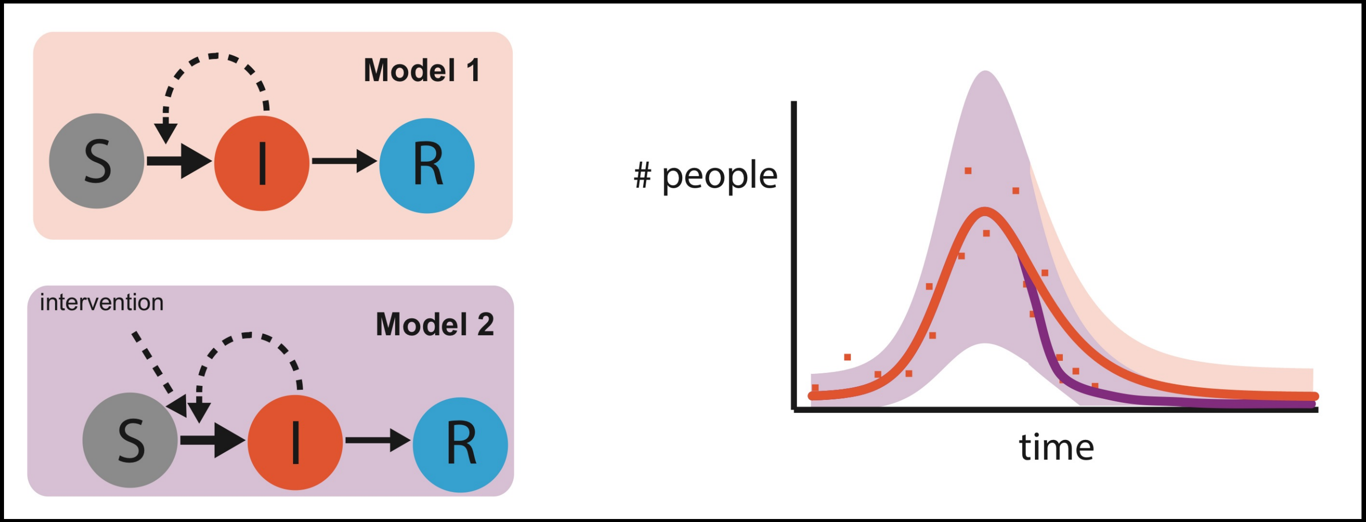
Mechanistic modeling is process-driven…

- Test "what if" scenarios not amenable to experimentation
- Estimate parameters that are difficult to measure by fitting models to available data
- Forecast forward in time



Mechanistic modeling is process-driven…

- Test "what if" scenarios not amenable to experimentation
- Estimate parameters that are difficult to measure by fitting models to available data
- Forecast forward in time
- Select between models of differing hypotheses



Mechanistic modeling is process-driven…

- Estimate: time series of state variables/ parameters of interest → DATA
- Inference: Build model to recapture data. Fit to optimize parameters and "infer" the process underlying the data
- Model assessment: Assess plausibility or model comparison
- End goal: explain observed patterns or predict