# 1 Epidemic Modelling Intro

McMaster University

Mathematics and Statistics

$$\int_M d\omega = \int_{\partial M} \omega$$

# Mathematics 4MB3/6MB3 Mathematical Biology

Instructor: David Earn

Lecture 1
Epidemic Modelling Intro
Tuesday 3 September 2024

# Course information

- Everyone should have received an e-mail.
  - If not, please e-mail earn@math.mcmaster.ca now from the e-mail address that you use.
- Course web page: `http://ms.mcmaster.ca/earn/4MB3`
  - Click on "Course information".
  - Let's have a look now...

# Mathematical Biology Research Seminar (MBRS)

Most weeks, there is a Mathematical Biology Research Seminar, which you are encouraged to attend if you are available.

- Where: **HH-410**
- When: **Thursdays, 2:30–3:20pm**
- Starts: **Thursday 12 September 2024**

If you would like to be on the e-mail distribution list for these seminars, please send me an e-mail (with MBRS in the subject line).

# In-class polls: a form of participation

- Please log in (right now) to the childsmath web site: https://www.childsmath.ca/childsa/forms/main_login.php
- Click on Math 4MB3.
- Click on Take Class Poll.
- After selecting the person you think is your instructor, click the Submit button.
- Everybody done?
- Let's Deactivate the poll and View Results

# Group formation

**Most work in this course will be done in groups.**

- Form a group of 2 or 3 students **during the break TODAY.**
- **Exactly one** member of your group must `e-mail the instructor` **TODAY during the break:**
  - Include "Math 4MB3" and your proposed group name in the subject line.
  - **Copy your message to all members of your proposed group so I have everyone's e-mail in the thread.**
- *Note: Instructor may change groups based on survey results.*

# Online Surveys

*You will be required to fill in online surveys during this course. Doing so in a timely manner contributes to your participation mark.*

The first online survey has been posted:

- Go to the Surveys page on the course web site.
- Follow the link for Background and Group formation Survey.
  - `https://surveys.mcmaster.ca/limesurvey/index.php/746511`
- Complete the survey ***TODAY***.
- It should take only $\sim 5$ minutes.
- Note that *surveys sometimes fail to save*.
  - Type long answers into a file first and paste them into the survey. Then you won't get as frustrated if it fails to save.

# Software

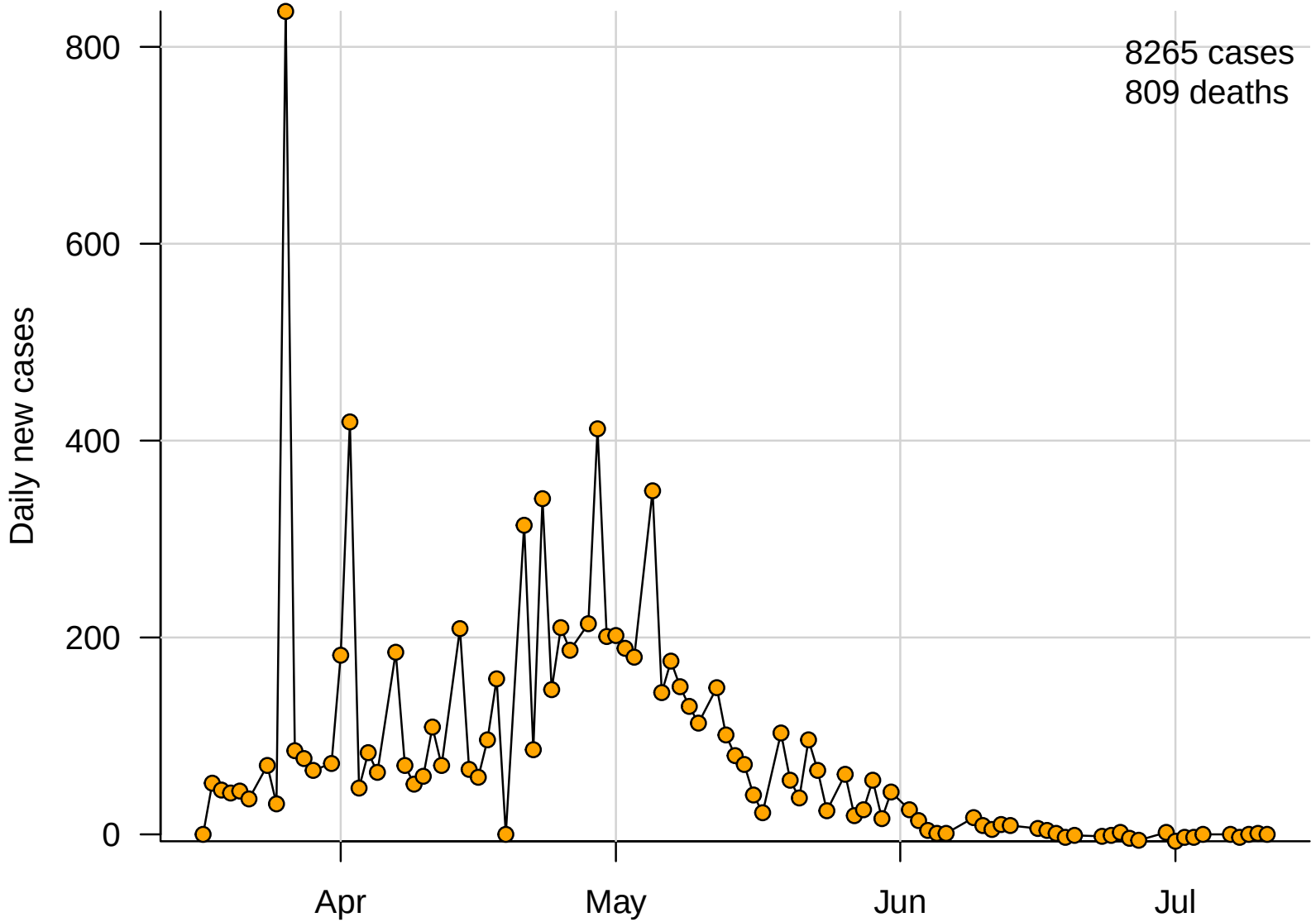
- **ASAP**, install the software discussed on the Software page on the course web site:
  - LaTeX

  - R
  - RStudio
  - XPPAUT
  - Emacs

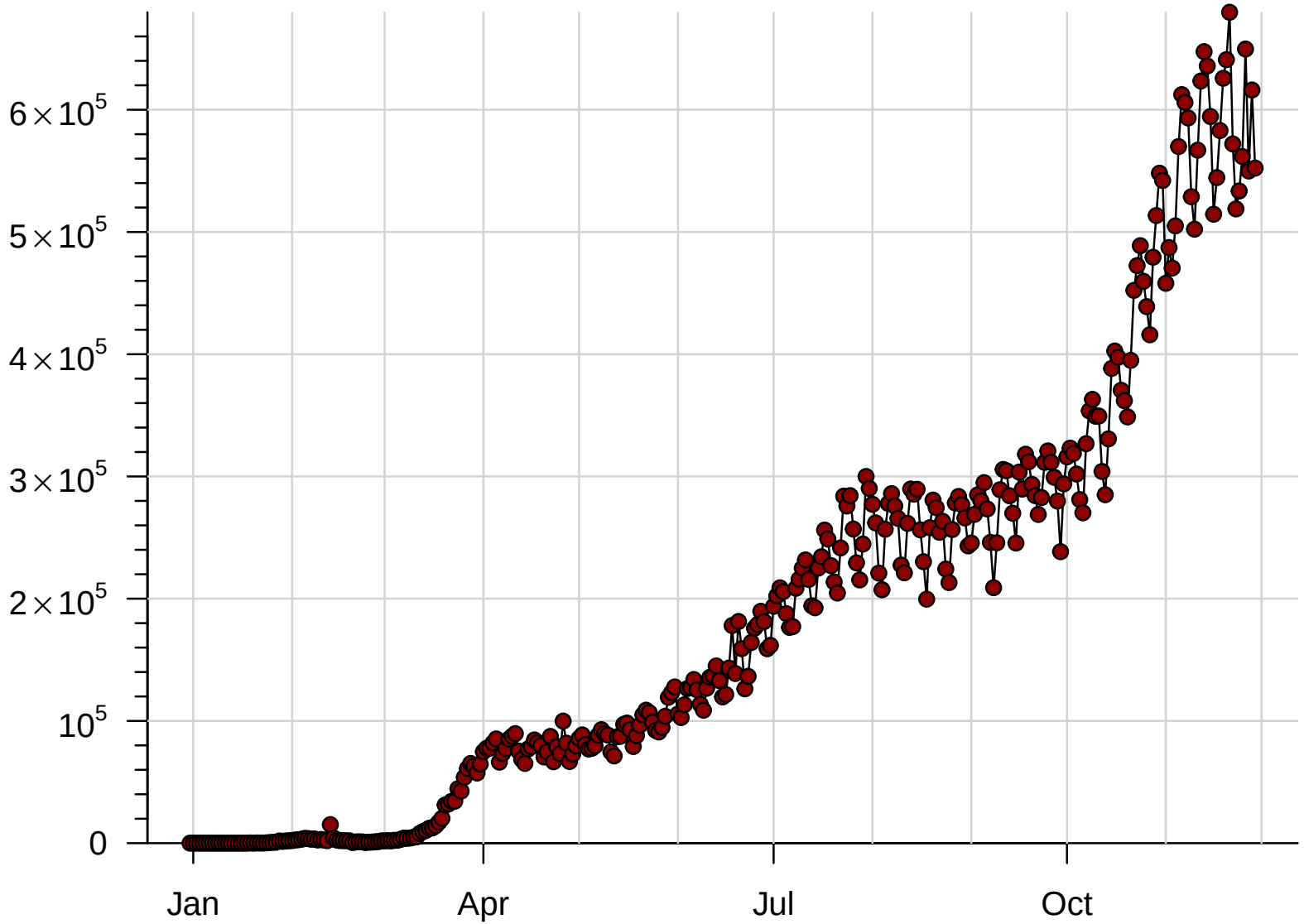

- *Note:* the Software page also contains some info about spell-checking and counting words in LaTeX documents.

# Epidemic Modelling

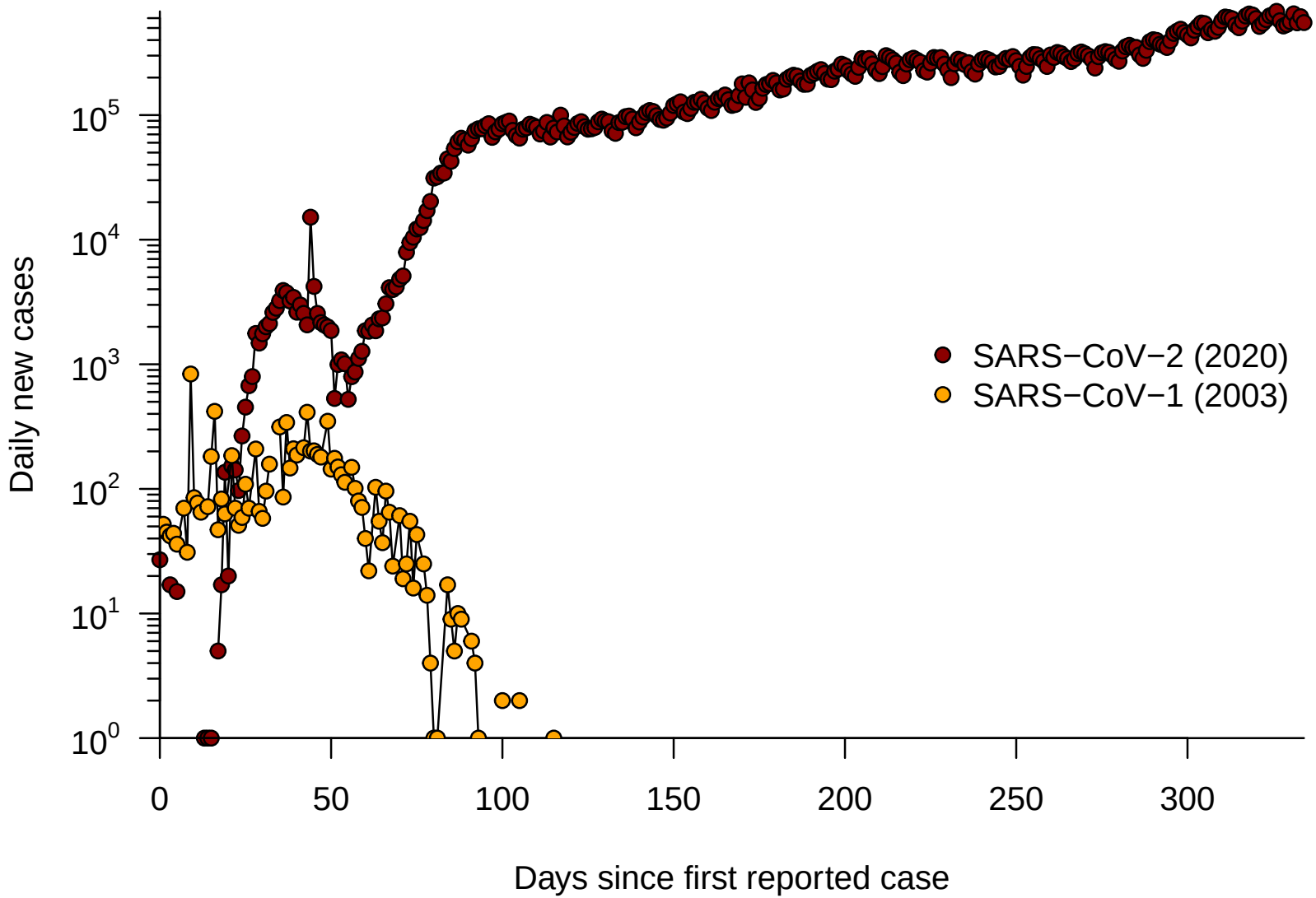
# Daily SARS-CoV-1 in 2003 (Worldwide)



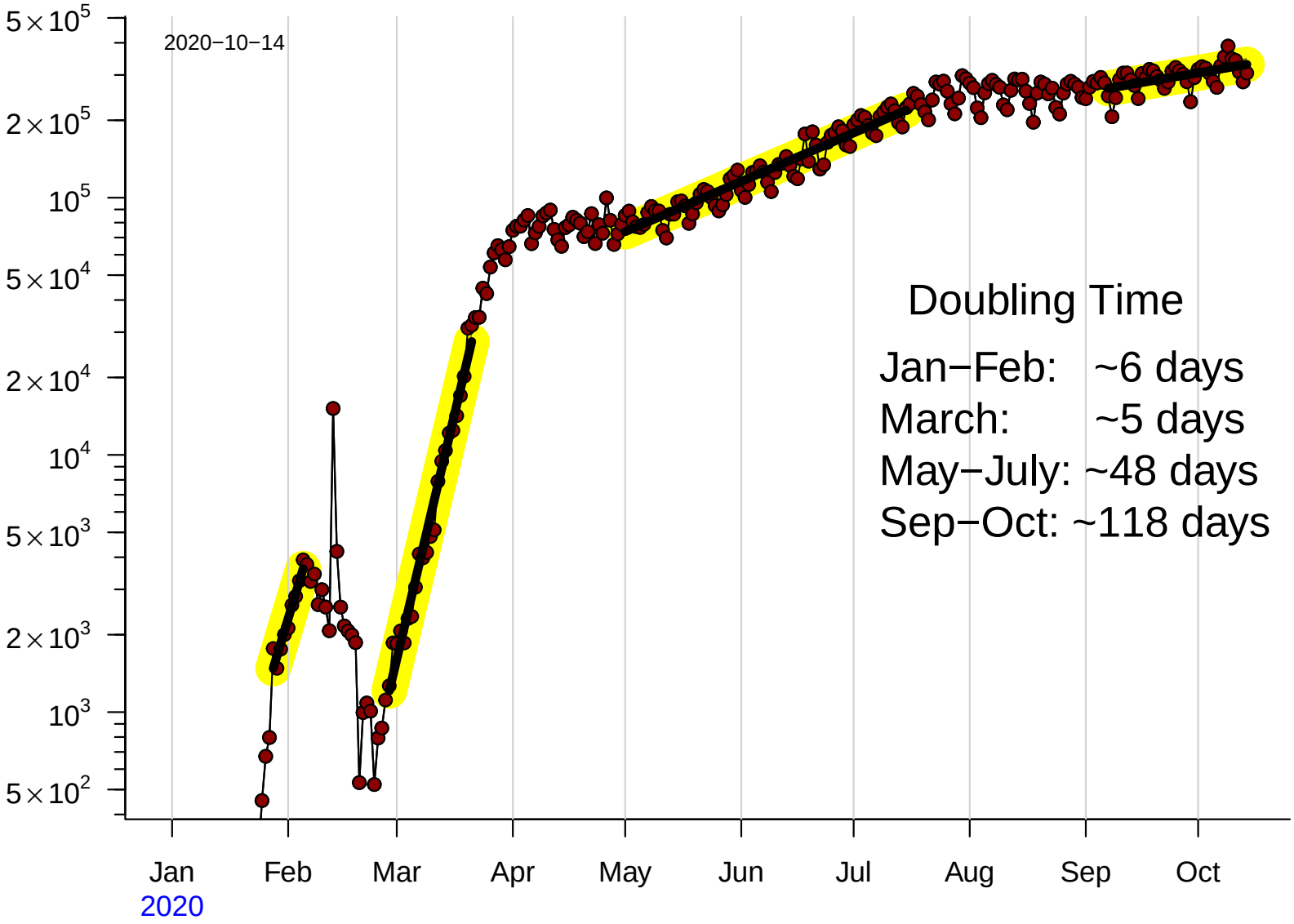
# Daily SARS-CoV-2 in 2020 (Worldwide)



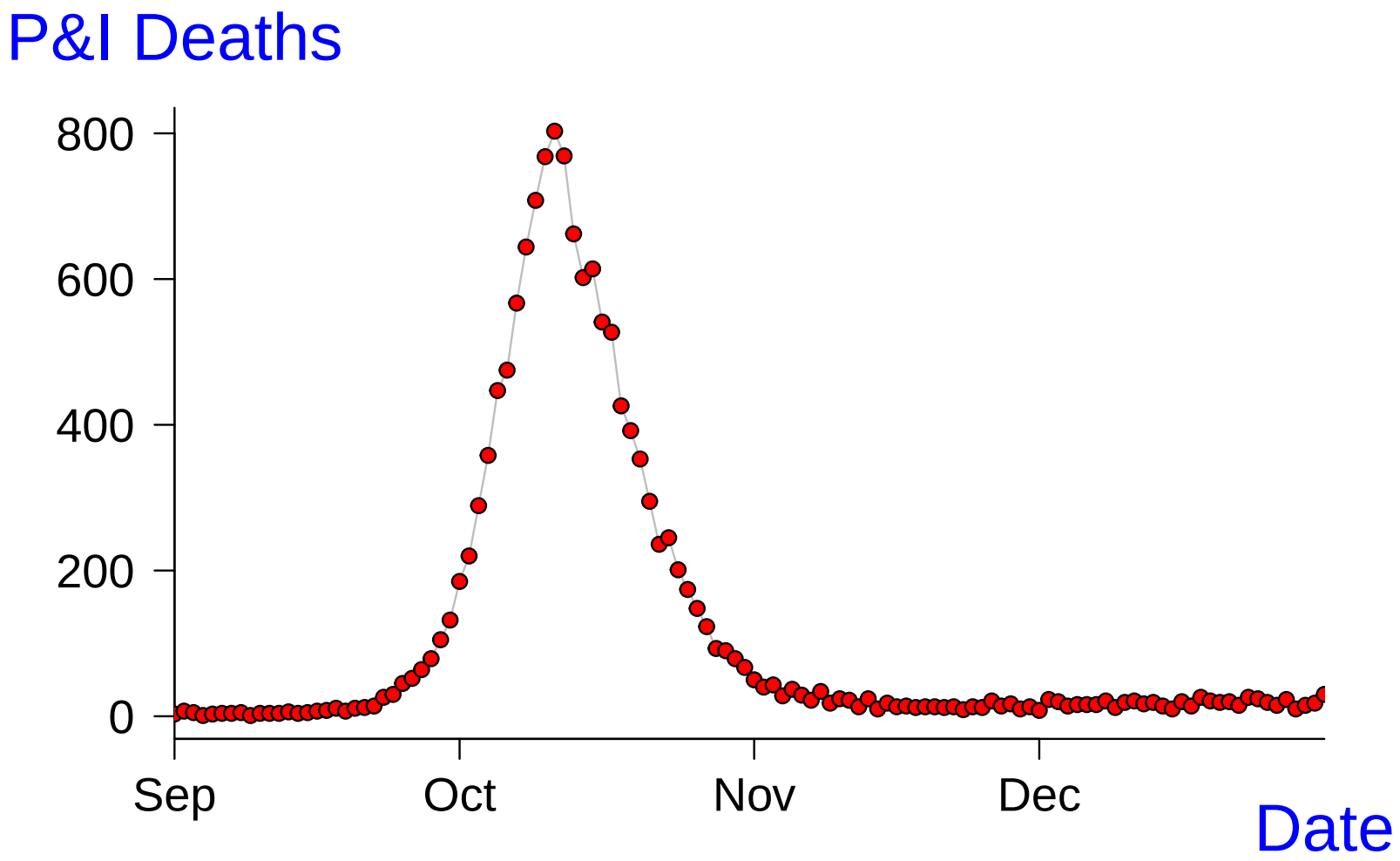
# Daily SARS-CoV-1 vs SARS-CoV-2 (Worldwide)



# Daily SARS-CoV-2 (Worldwide) exponential growth fits



# Pneumonia & Influenza Mortality, Philadelphia, 1918



# Modelling challenge

Develop a model that helps us understand the graph on the previous slide, based on mechanisms of disease spread.

- Only one variable is observed (P&I deaths per day) so construct a model containing only one variable.
- Think about how disease spreads and express your thoughts with mathematical notation.
- Derive a differential equation that models the process of disease transmission.
- Analyze the model and determine its strengths and weaknesses/limitations.

# Make (Biological) Assumptions Clear

1. *Assume* the disease is transmitted by contact between an infected individual and a susceptible individual.
2. *Assume* the latent period (delay between being infected and becoming infectious) is so short that it can be ignored (technically assume it is zero).
3. *Assume* all members of the population are identical and respond identically to the disease. In particular, all susceptible individuals are equally susceptible and all infected individuals are equally infectious.
4. *Assume* the population size is fixed during the epidemic, *i.e.*, ignore births, migration, and deaths from causes other than the disease, and count individuals who have died from the disease as part of the population.

# About Assumptions. . .

- Note that the first assumption on the previous slide is actually correct.
- The other assumptions are wrong, but are reasonable approximations.
- It is best to *start as simple as possible and add complexity later*, in order to:
  - obtain a model that actually succeeds in explaining the data with as few ingredients as possible;
  - check that inferences we draw from our model(s) are robust to the inclusion of more biological details/realism.

# What variables should we include in our model?

- Independent variable: time ($t$)
- Dependent variable: Many options, *e.g.*,
  - Incidence (number of new infections per unit time)
  - Prevalence (total current number of infected individuals)
  - Death rate (number of deaths per unit time)
  - Death toll (number of deaths so far)
- So, what would be best?
- Not deaths, because whether or not you die may be unrelated to how much you transmit.
- But deaths are what we observe! What to do?!?
- Make another assumption...

# Additional assumption(s)

- We actually want to know incidence or prevalence, but we observe deaths.
- Under what circumstances would daily deaths be a good estimate of incidence? (*i.e.*, What must we assume in addition to the assumptions we have already made.)
  5. *Assume* that the time from infection to death is exactly the same (a certain number of days) for every individual who dies.
  6. *Assume* that the probability of dying from the disease is the same for every individual who is infected.
- Then daily death counts are proportional to daily incidence a certain number of days in the past, *i.e.*, the "mortality curve" that we observe is a translated and scaled version of the "epidemic curve" (new cases per day).

# So. . . what variables should we include in our model?

- Independent variable: time ($t$)
- Dependent variable: one of:
  - Incidence (number of new infections per unit time)
  - Prevalence (total current number of infected individuals)
- Which one?
- Choose prevalence ($I$) because anybody who is currently infectious can infect others, so it will probably be easier to formulate a transmission model based on prevalence. (Try not to lose sight of underlying biological mechanisms.)
- But our mortality curve is related to incidence, not prevalence!?! Argh. What to do?!?
- Let's work with prevalence and see how it works out. Maybe we'll be able to derive the incidence curve from a model based on prevalence.

# Notational note

- We use $I$ for prevalence because prevalence is the number of infected individuals.

- So, let's try to write down a model...

# A first (naïve) attempt at an epidemic model

- Variables: time $t$, prevalence $I(t)$
- How does $I$ increase?
- Start with $I_0$ infected individuals at time $t = 0$. What happens for $t > 0$.
- Let $B =$ average number of contacts with susceptible individuals that lead to a new infective per unit time per infective in the population (and suppose $B$ is constant). Then

$$I(t + \Delta t) \simeq I(t) + B\, I(t)\Delta t$$

- In the limit $\Delta t \to 0$, we have

$$\frac{dI}{dt} = BI \quad \Longrightarrow \quad I(t) = I_0 e^{Bt}$$

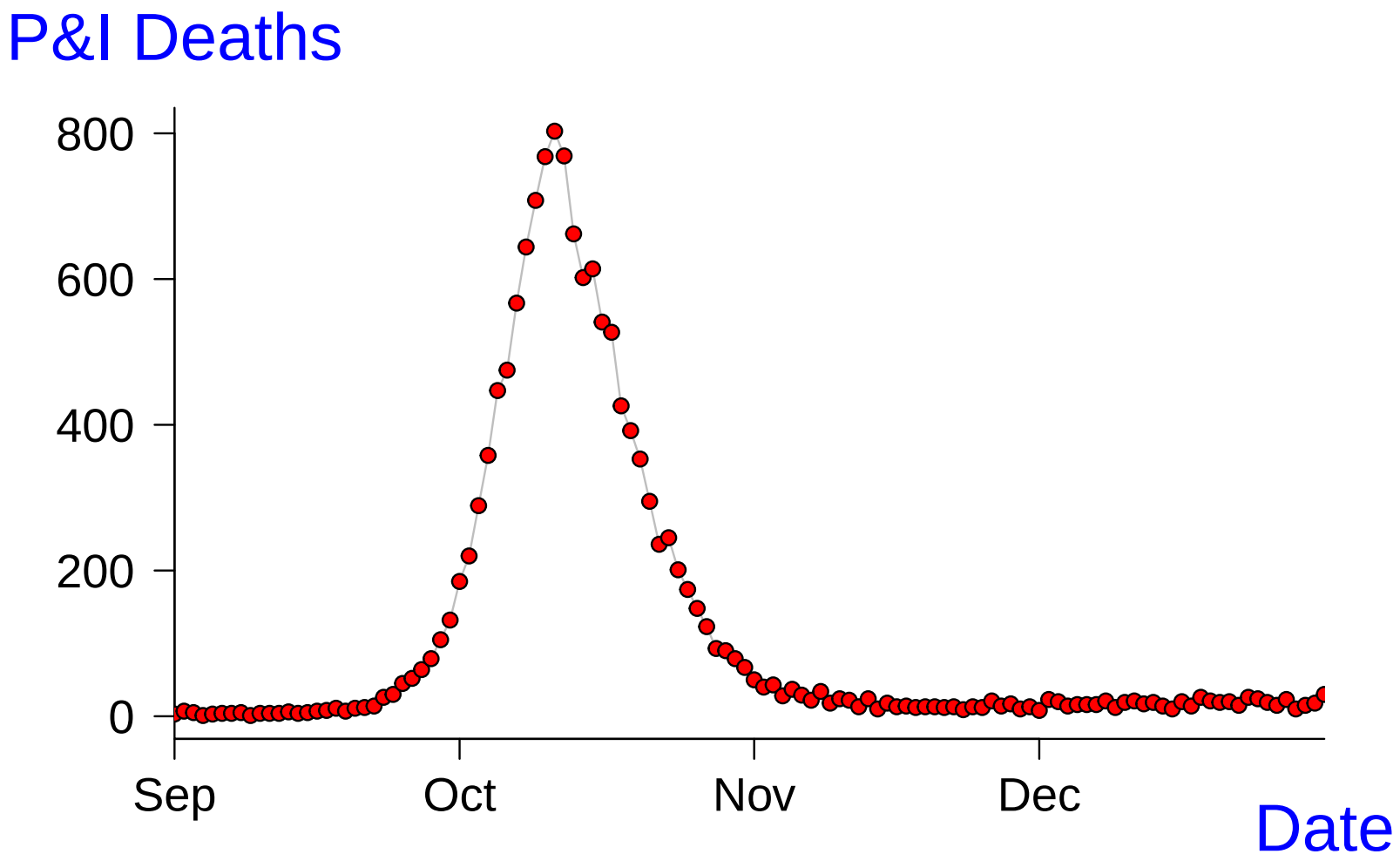
# Beware: implicit assumptions that should be explicit

- Ignored discrete nature of individuals when taking limit.
- *Ignored finite infectious periods!*
  - Sometimes it isn't obvious that we've made some assumptions until after we see what the model predicts.

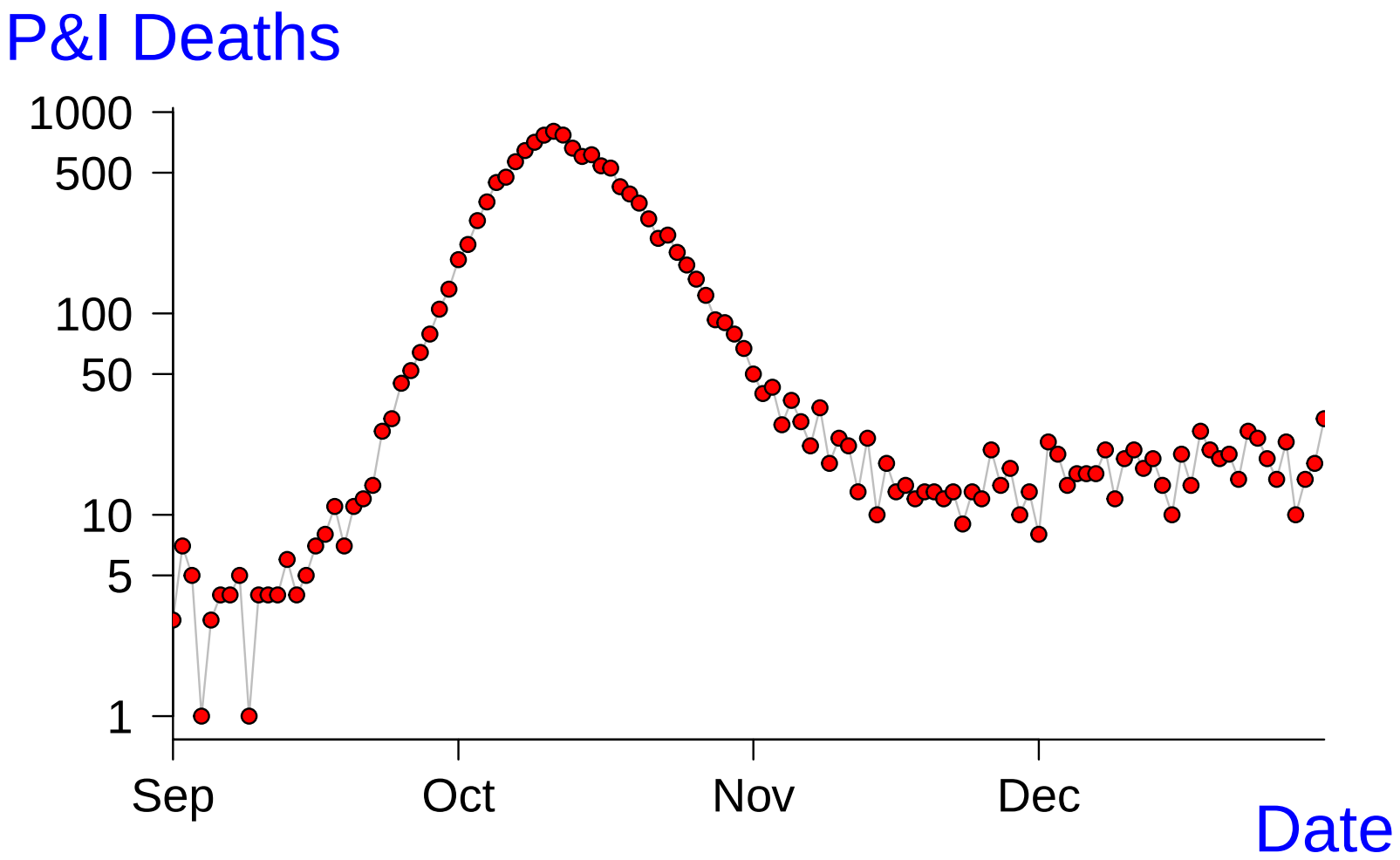
# How can we tell if our model is good?

- Compare model predictions with data.
- What is the best way to do that?
- Depends on what predictions we're trying to test.
- Model predicts exponential growth.
  *How should we test that prediction?*
- Transforming the data might help.

# Original data: P&I Mortality, Philadelphia, 1918



# Logarithmic scale: P&I Mortality, Philadelphia, 1918



# Parameter estimation

*How can we estimate the model parameters, $I_0$ and $B$, from the P&I data?*

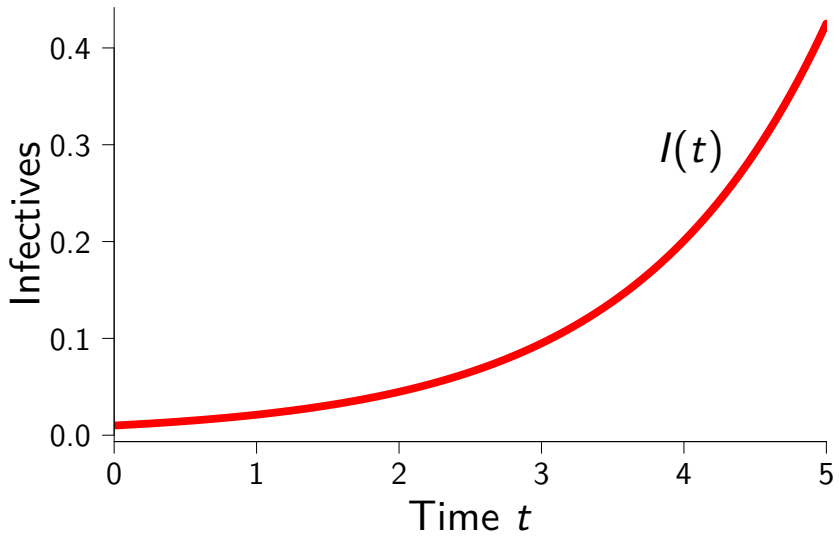
- Fit a straight line through the part of the logarithmic mortality curve that looks straight.
- The slope of the line is $B$.
- The "intercept" is $\log I_0$.
  - "Intercept" in quotes because we need to define $t = 0$ as the time when exponential growth begins.
- *Note:* Parameter estimation is, in general, a very tricky business and deserves a great deal of attention.

Earn, Park, Bolker (2024) "Fitting Epidemic Models to Data..." *Bull. Math. Biol.* **86**, 109

# Naïve epidemic model

- Variables: time $t$, prevalence $I(t)$
- Parameter $B$ = average number of contacts with susceptible individuals that lead to a new infective per unit time per infective in the population

$$\frac{dI}{dt} = BI \quad \Longrightarrow \quad I(t) = I_0 e^{Bt}$$



# Naïve model: the good and the bad

- Good:
  - Makes clear predictions
  - Predictions can be tested
  - Estimation of parameter ($B$) is easy
    - $B$ is the slope of the straight portion of the epidemic curve on the log scale. (Why?)
    - Remember we are imagining that the mortality curve is equivalent to the epidemic curve after translation and scaling.
    - Why do translation and scaling not affect the estimate of $B$? Assignment 2...
- Bad:
  - Model is consistent only with exponential growth phase.
  - Absurd long-term prediction: unbounded growth in $I(t)$
    - Implicitly assumed that population size $N = \infty$.

# How can we improve our model?

- Insist that population size is finite ($N < \infty$).
- Keep track of both infectives $I(t)$ and susceptibles $S(t)$.
- Assume individuals who are *not infected* are susceptible:

$$I(t) + S(t) = N = \text{constant}.$$

# New model parameter(s)?

- $B =$ average number of contacts with susceptible individuals that lead to a new infective *per* unit time *per* infective
- In the naïve model, we assumed $B =$ constant.
  Is $B$ really constant?
- $B$ depends on how many susceptibles there are.
- $B = \beta S(t)$
- $\beta =$ average number of contacts between susceptibles and infectives that lead to a new infective
  *per* unit time
  *per* infective
  *per* susceptible
- $\beta$ is called the ***transmission rate***.

# Revised epidemic model: "SI model"

$$\frac{dI}{dt} = \beta S(t) I(t)$$

- Two state variables. One equation. Problem? No:

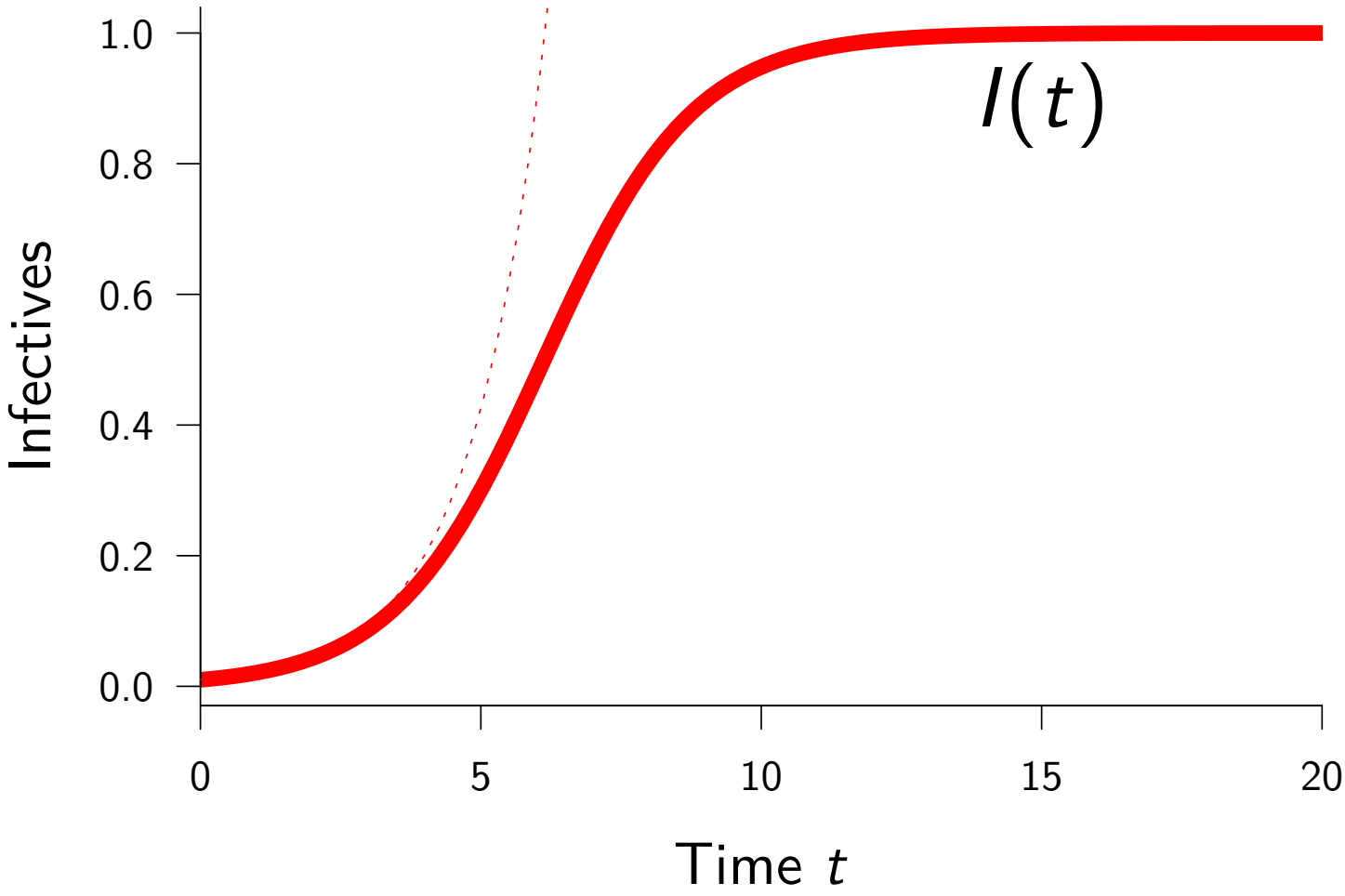
$$\frac{dS}{dt} = -\beta S(t) I(t)$$

- But $S(t) = N - I(t)$ $\quad \Longrightarrow \quad$ $I(t)$ is still the only variable:

$$\frac{dI}{dt} = \beta I(N - I)$$

- Is this a better model?
- What does it predict?

# SI model: Example solution



# SI model: Analysis

$$\frac{dI}{dt} = \beta I(N - I)$$

- We can find the exact solution. How? (Assignment 1)

$$I(t) = \frac{I_0 e^{N\beta t}}{1 + (I_0/N)(e^{N\beta t} - 1)}$$

- But exact solution is not particularly enlightening.
- Qualitative analysis:
  - Initially $I \ll N$. What does the model predict then? Exponential growth. Great!
  - As $I$ grows, growth rate slows. Why? Fewer and fewer susceptibles to infect.
  - Asymptotic behaviour? Equilibria? Periodic orbits? (periodic orbit = recurrent epidemics)
  - (Non-trivial) periodic orbits impossible in one dimension (existence-uniqueness theorem).
  - Consider equilibria...

# SI model: Equilibrium Analysis

$$\frac{dI}{dt} = \beta I(N - I), \qquad I \in [0, N]$$

- Two equilibria:
  - $I = 0$ Disease Free Equilibrium (**DFE**)
  - $I = N$ Endemic Equilibrium (**EE**)
- Stability:
  - DFE is unstable ($0 < I < N \implies dI/dt > 0$)
  - EE is locally asymptotically stable (**LAS**)
  - EE is globally asymptotically stable (**GAS**)
    (stability of EE follows from $0 < I < N \implies dI/dt > 0$)
    (GAS requires a little more analysis... Assignment 1)
  - *Note:* In one dimension, global analysis always easy.
    In higher dimensions, often try to find Lyapunov function.
    (Lyapunov function for EE of SI model?... Assignment 1)
- Conclusions identical for any $\beta > 0$.

# SI model: Biological Inferences

- For *any* transmission rate $\beta$:
  - Initially, exponential growth of cases.
  - Eventually, convergence to equilibrium (EE) at which *everyone* in the population is infective. hmmm...
- Is this model better than our first naïve model?
  YES.
  - Still correctly predict initial exponential growth.
  - Can match epidemic curve for longer (up to the peak).
  - Does not predict absurd unbounded growth in infective population.
  - But this model cannot explain the decline of the epidemic.
- What should we do? Two obvious options:
  1. Get depressed, drop the course.
  2. Try to improve the model.

# Recall motivating data: 1918 flu in Philadelphia

- Mortality curve (linear scale)
- Mortality curve (log scale)

# How can we improve on the SI model?

- Include a key biological fact:
  Individuals do not stay infectious with flu forever
- Either they *recover* and are immune thereafter, or they *die* (doesn't matter which) (well, maybe to them it does)
- Why don't we care if someone recovers or dies?
  (*i.e.*, Why doesn't it affect our dynamical inferences?)
- Because in either case the individual is *removed* from the transmission process, hence cannot affect the future pattern of the epidemic.

# The SIR model

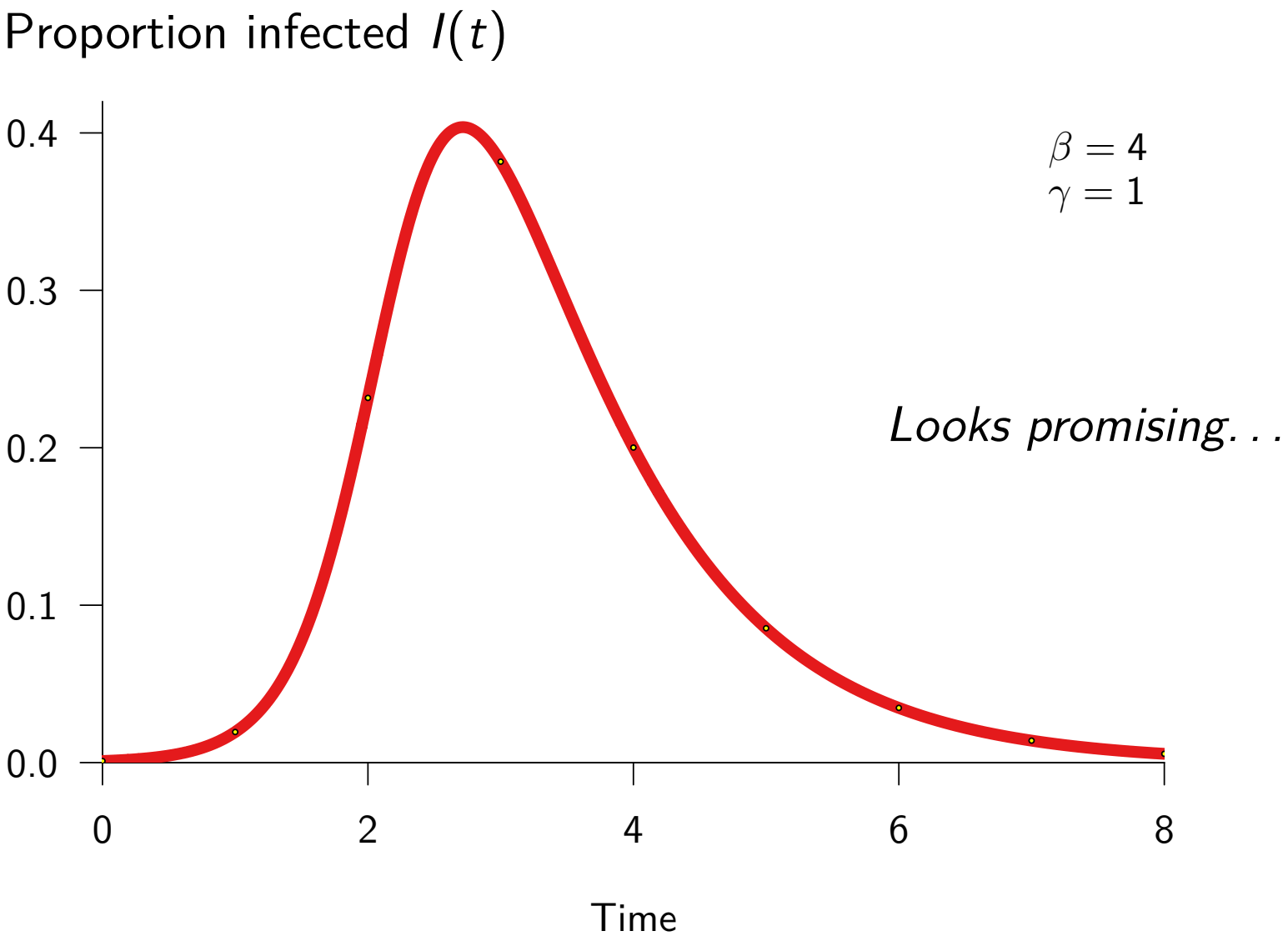
Introduce new class of removed individuals:

- $R(t)$ = number of individuals who have either recovered and are now immune or have died
- Let $\gamma$ = rate of removal from the infective class (via recovery or death)

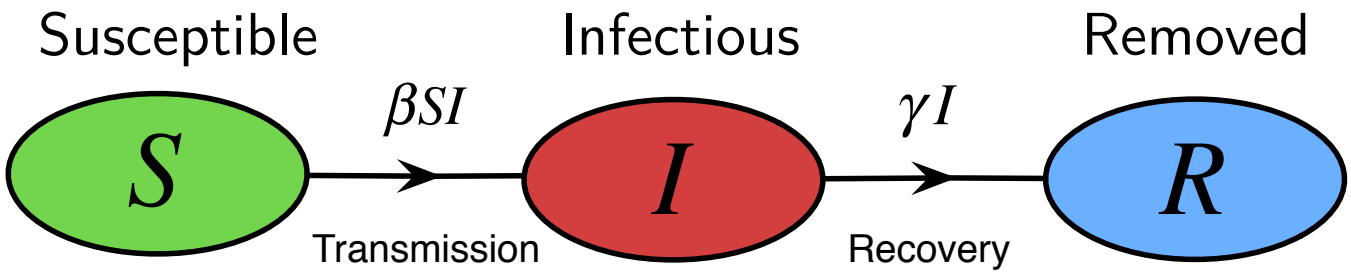
$$\frac{dS}{dt} = -\beta SI$$
$$\frac{dI}{dt} = \beta SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

- Note: $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \implies S + I + R = N = \text{constant}$
- Convenient to rescale variables by $N$ and interpret $S, I, R$ as *proportions* of the population in each disease state.

# The SIR model: Example numerical solution



# The SIR model: Flow Chart and Parameters



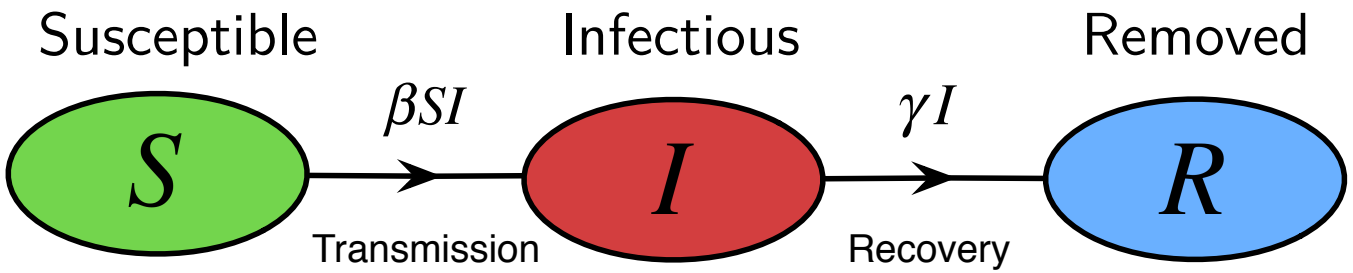

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Parameters:**
  - Transmission rate $\beta$
  - Recovery rate $\gamma$ (or Removal rate)

# The SIR model: Derived Parameters




$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- **Derived Parameters:**
  - Initial growth rate $\beta - \gamma$
  - Mean infectious period $\frac{1}{\gamma}$
  - Basic Reproduction Number

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

# The SIR model: Derived parameters

## The initial growth rate

- How do we calculate the initial growth rate from our model?
- Consider change in prevalence initially (when $I \ll 1$):

$$\begin{aligned}\frac{dI}{dt} &= \beta SI - \gamma I \\ &= \Big(\beta S - \gamma\Big) I \\ &\approx \Big(\beta - \gamma\Big) I \qquad \text{if } S \sim 1 \text{ initially.}\end{aligned}$$

- $\therefore$ Initially $\quad I(t) \approx I_0 e^{(\beta-\gamma)t}$.
- $\therefore$ Initial slope of logged prevalence curve is $\quad \beta - \gamma$.

# The SIR model: Derived parameters

## The mean infectious period

- How do we calculate the mean infectious period from our model?
- Suppose at time 0 there are $I_0$ infectious individuals, and suppose we can *prevent contact with susceptibles*.
- The equation for $I$ then simplifies to

$$\frac{dI}{dt} = -\gamma I\,, \qquad I(0) = I_0$$

- We can solve this immediately to find

$$I(t) = I_0 e^{-\gamma t}$$

# The SIR model: Derived parameters

## The mean infectious period, continued...

- Thus, after time $t$, the number of people still infectious is reduced by a factor $e^{-\gamma t}$
- *i.e.*, the proportion of individuals who have an infectious period shorter than $t$ is $1 - e^{-\gamma t}$
- *i.e.*, the cumulative distribution of the infectious period is $C(t) = 1 - e^{-\gamma t}$.
- Therefore, the probability density of the infectious period is $p(t) = C'(t) = \frac{d}{dt}\left(1 - e^{-\gamma t}\right) = \gamma e^{-\gamma t}$
- The mean of this exponential probability distribution is $\int_0^\infty t\, p(t)\, dt = \int_0^\infty t\, \gamma e^{-\gamma t}\, dt = \frac{1}{\gamma}$

# The SIR model: Derived parameters

## The basic reproduction number $\mathcal{R}_0$

$$\mathcal{R}_0 = \beta \cdot \frac{1}{\gamma}$$
$$= (\text{transmission rate})$$
$$\times (\text{mean infectious period})$$

- $\mathcal{R}_0$ is dimensionless
- $\mathcal{R}_0$ is the average number of secondary cases caused by a primary case (in a wholly susceptible population).
- We must have $\mathcal{R}_0 > 1$ to have an epidemic. Why?
  - $\frac{dI}{dt} = \beta SI - \gamma I = \big(\mathcal{R}_0 S - 1\big)\gamma I$
  - $\therefore \mathcal{R}_0 \leq 1 \implies \frac{dI}{dt} \leq 0$ for all $(S, I) \in [0, 1]^2 \implies$ no growth