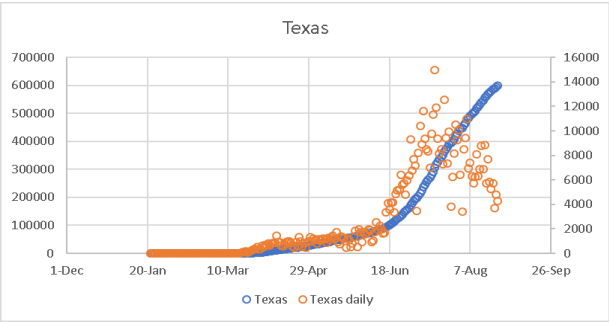
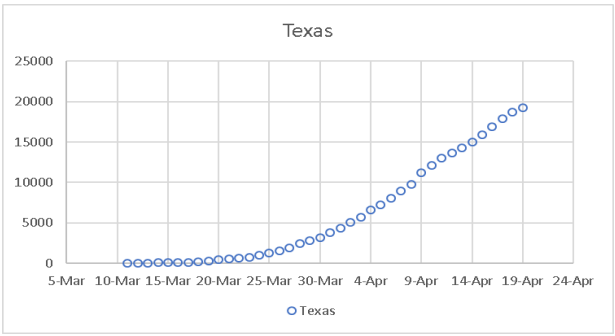
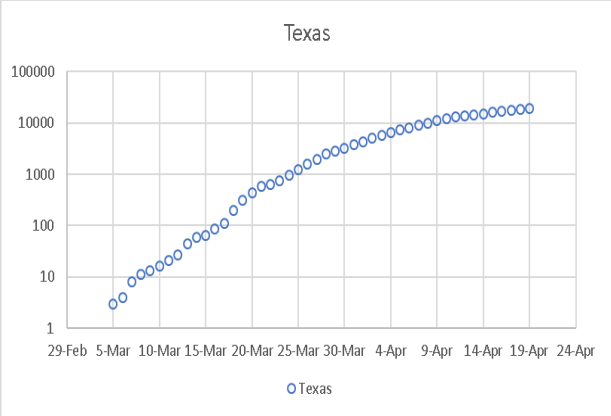
Explanation on R_o: If one can find the exponential part of the cumulative growth curve (a short section near the beginning of the curve that loses validity as N approaches zero), one finds the growth rate constant for an exponential, which is $K = d \ln(N) / dt$. [1]

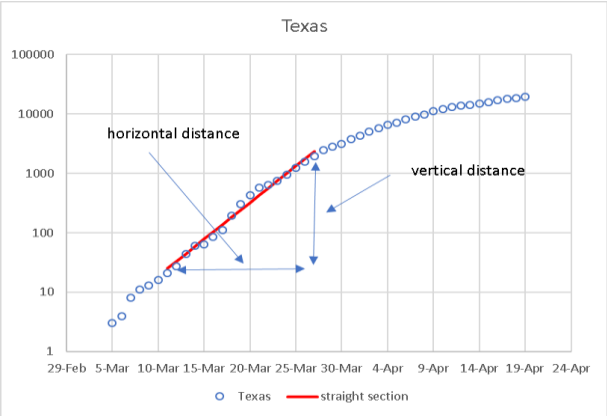
For example, here is some data on Texas' CV19 growth this year (the blue is cumulative), where the second graph zeros in on the very first days of this misery:





If we plot this semi-log (logarithmic vertical axis vs. arithmetic horizontal axis), this is what it looks like. The second graph highlights a straight line section, after which it begins to curve:





The straight section has a slope (ratio of rise over run, vertical over horizontal distance) that is 0.284 (done with a least squares regression). Afterwards, it is no longer exponential growth, since it curves off. This is called K, $K = 0.284$ for Texas. Notice this is a bit arbitrary, and avoids the values less than 40, at the extreme early stage.

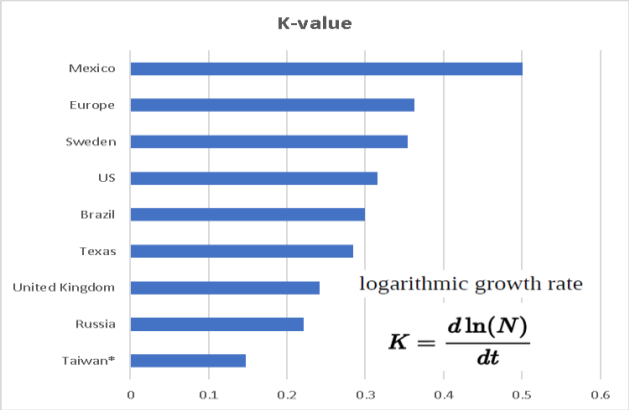
You can use this to find out infectious rate and contact rate, since can be R_o defined as the ratio of the contact rate to the infectious rate. β is often used for contact rate; τ is often used for infectious period, while γ (or κ) is used for infectious rate, its inverse. So,

$$R_o = \frac{\beta}{\gamma} = \beta\tau$$

R_o is also found graphically as $R_o = \exp(K\tau)$, where K is the slope of the exponential part of the cumulative curve just described.

[1] Basic Reproduction Rate. R_o is a measure of transmissibility: $R_o < 1$, disease disappears; $R_o = 1$, it's endemic; $R_o > 1$, epidemic. R_o is mentioned a lot during this epidemic, along with flattening of curves, with not a lot of understanding or relevance. The real trick is figuring out the *effective* reproduction rate, R_e.

$R_o = e^{K\tau}$ where τ is the average infectious period for each person. Doubling time is also often mentioned. Reduce K by reducing contacts, reduce τ by isolation of infected individuals, for example. Here are some K values taken from Johns Hopkins data:



Example ($R_0 = 2.6$, $K_{TX} = 0.284$ / day)

$$R_0 = e^{K\cdot\tau}$$

$$\tau = \frac{ln(R_0)}{K} \qquad \frac{ln(2.6)}{0.284} = 3.4 \text{ days}$$

Example, doubling time (2/1) :

$$2 = 1\cdot e^{K\cdot\tau}$$

$$\tau = \frac{ln(2)}{K} \qquad \frac{ln(2)}{0.284} = 2.4 \text{ days}$$

Interesting comment in open letter from Belgian doctors (Sep 2020):

https://docs4opendebate.be/en/open-letter/

Reminder from introductory statistics: A low incidence infection is very sensitive to false positives. For example, say 6% incidence, and 95% accurate test. Out of 1,000 people that means 60 people are infected. Therefore, out of 1,000 – 60 = 940, 5% are false positives (or 48 people). On the other hand, 95% of the 6% (or, 5.7%) that are infected are true, and so 5.7% of 1000 = 57. The number of false positives is almost equal to the number of true positives in this example! This is a pitfall of widespread testing with a test that is less than 99% accurate. In this example, it would almost double the estimated number of positive cases which could be used to set policy (lockdowns, masks, etc.). However, current practice is to test people with symptoms or prior exposure, so the sample incidence is no doubt much higher than the population incidence, which reduces this effect.[2]

Modeling

Interestingly, one way these epidemics are visualized and modeled is based on susceptibles contacting infected people at some average rate. These susceptibles are quantified as a percentage or fraction of the entire population[3] (which implies that some fraction of the population *isn't* susceptible; think vaccine or herd immunity). So, over time, these susceptible people are catching the disease at some rate. You could put it in pseudo-algebraic terms like this:

rate of change, susceptibles = contact rate, times fraction of total that are susceptible, times number of infectious [4]

That needs to be refined, since the fraction of the population that are susceptibles will decrease, as they are converted to infectious. That means the rate is negative. So, add a negative sign, and abbreviated a bit:

susceptibles change = -contact rate X fraction susceptible X infectious

---

[2] Notice difference between case fatality rate (CFR) and infection fatality rate (IFR). CFR is the ratio of the number of deaths divided by the number of confirmed (preferably through testing) cases of disease. IFR is the ratio of deaths divided by the number of actual infections with SARS-CoV-2. Much uncertainty on false positives, and therefore on the real IFR.

[3] Another way to look at this is the ratio of susceptibles to the whole population is the basis for assigning a *probability* of infection of anyone in the whole population, not just susceptibles: contact rate X probability X infectious. The program of compartmentalizing by susceptibles, infectious, and recovered is known as the SIR model.

[4] Contact rate is average number of people a person has contact with, per unit time. Hence contact tracing, that not only tells how many people a certain infected person has had contact with, but hopefully who they were.

That's a bit cumbersome, so if you follow Leibniz' suggestion, and write a rate of change as the change divided by the corresponding change in time, where the symbol for change (difference) is a *d*:

$$\frac{dS}{dt} = -contact\ rate\ X\ \ fraction\ X\ infectious$$

People have come up with less cumbersome algebraic symbols, like β for the contact rate, and a real fraction for the fraction, *S*/*N*, where *S* is susceptibles and *N* is the number of the total population. Likewise, the average number of infectious would be well served with the letter I, so *I*:

$$\frac{dS}{dt} = -\beta \frac{S}{N} I$$

That's easier on the eye. It's important to remember all these quantities change over time, so they can be shown as functions of time (e.g. *S* written as *S*(*t*)), which clutters things up a bit. For example, deaths from an epidemic (not shown here) would reduce the value of *N*(*t*) over time. *S*/*N* is almost always less than 1.0, except perhaps at the very beginning where just about everyone was susceptible and where R_o was found, keep in mind.

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

Furthermore, the number of infectious changes, too. So, another, new equation:

$$\frac{d\,I(t)}{dt} = \beta \frac{S(t)}{N(t)} I(t) - \gamma I(t)$$

Notice the first term on the right side of the equation is just the negative of the change in susceptibles, since we're *adding* the same amount to infectious, that was *subtracted* from susceptibles before. So the change in infectious relates to the change in susceptibles, since members from one group are going over to the other. There is also the infectious period (not to be confused with contact rate). After a certain time, infectious either recover or die (this is kind of grim). So that has to be subtracted from the right side of the equation (the second term). γ is a common symbol for the frequency associated with the period of infection. Therefore γ^{-1} is the average period of infection. γ is the third letter in the Greek alphabet, gamma. $\gamma I(t)$ should be recognized as the rate of infectious by death or recovery, taken away from the overall rate.

The SIR model includes one more item in addition to the above, the recovered cases. The three should add up to a constant, throughout their evolution. Notice the following drops the *N* denominator in the ratio *S*/*N*, but *S*/*N* is easily recoverable, so:

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

where $S(t) + I(t) + R(t)$ = constant. That's true, because they are all feeding off each other. An approximation to these expressions can be made by adding increments, using Taylor's expansion (sometimes called Euler's method):

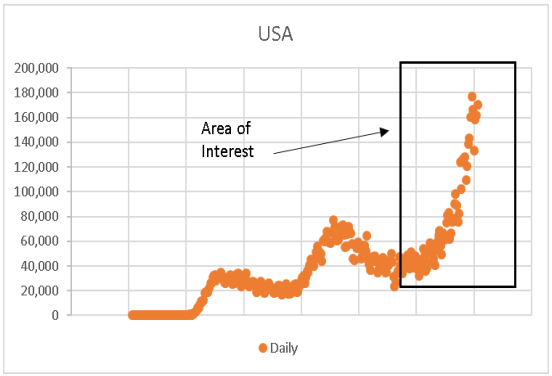
$$f(t + \Delta t) = f(t) + \Delta t\ [\Delta df(t)\ /\Delta t] \text{ and } \frac{df}{dt} \approx \frac{\Delta f}{\Delta t}.$$

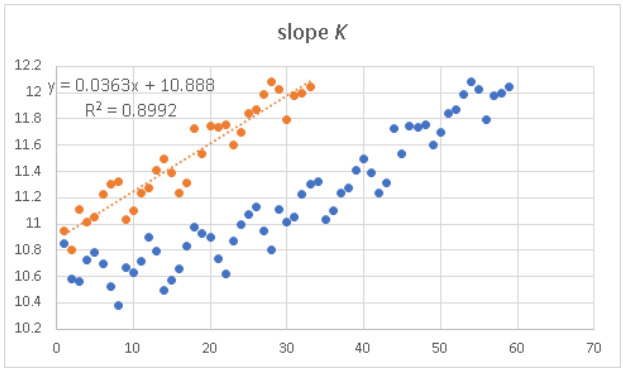
That way a table can be built from *I*(*t*), for example.

$$\frac{dI}{dt} \approx \frac{\Delta I}{\Delta t} = \beta\, S(t) I(t) - \gamma\, I(t), \text{ so } I(t+\Delta t) \approx I(t)\left[1 + \Delta t\left(\beta\, S(t) - \gamma\right)\right].$$

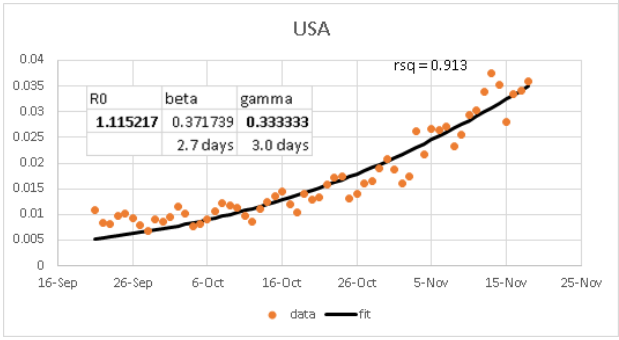
(Also, $S(t+\Delta t) \approx S(t)\left[1 - \Delta t\ \beta\ I(t)\right]$ and $R(t+\Delta t) \approx R(t) + \Delta t\ \gamma\ I(t)$.)

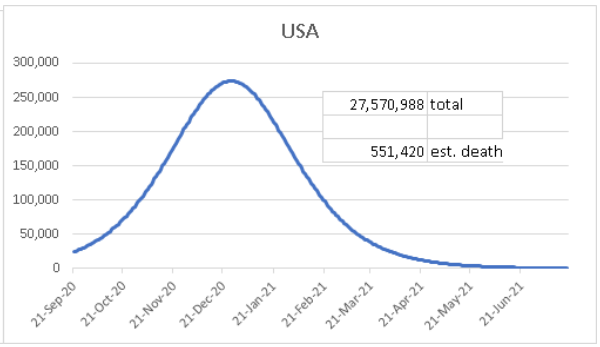
Taking the last few weeks of US numbers (just keep in mind, the US is comprised of thousands of similar sub-waves, at different magnitudes and starting points), and assuming that γ is 1/3, taking daily increments (Δt = 1, in other words), a curve fit can be made, using R_o (use the definition from page 1) and a starting point as two parameters to fit. The area of interest is shown below, with a logarithmic representation to the right, showing a slope of 0.0363. Since R_o can be approximated as $e^{K/\gamma}$, R_o is estimated at 1.115.





One other thing interesting about the table created (it's called numerical integration, by the way) is it is of arbitrary scale. That means the data to be fitted and the table of values have to be scaled to each other.





It's not a great fit, but gives some idea of what's happening. The R_o estimated is almost not an epidemic, but keep in mind this is a composite value, and some locales are much higher (denser populations, for example). Using the scaling factors, the original scale can be recovered, and the table filled out towards the end, as in the chart on the right.

This probably won't happen this way, because of the uncertainty in any projection and also the lockdowns and extreme methods that will likely occur (already occuring), which will change parameters like R_o. The estimated death figure is just assuming a case fatality rate of 2%. It's about the same as died in the Civil War (but that was a much higher proportion of the population, it should be pointed out). Also, this only is a projection of confirmed cases—the actual number would be 5 to 10 times higher.