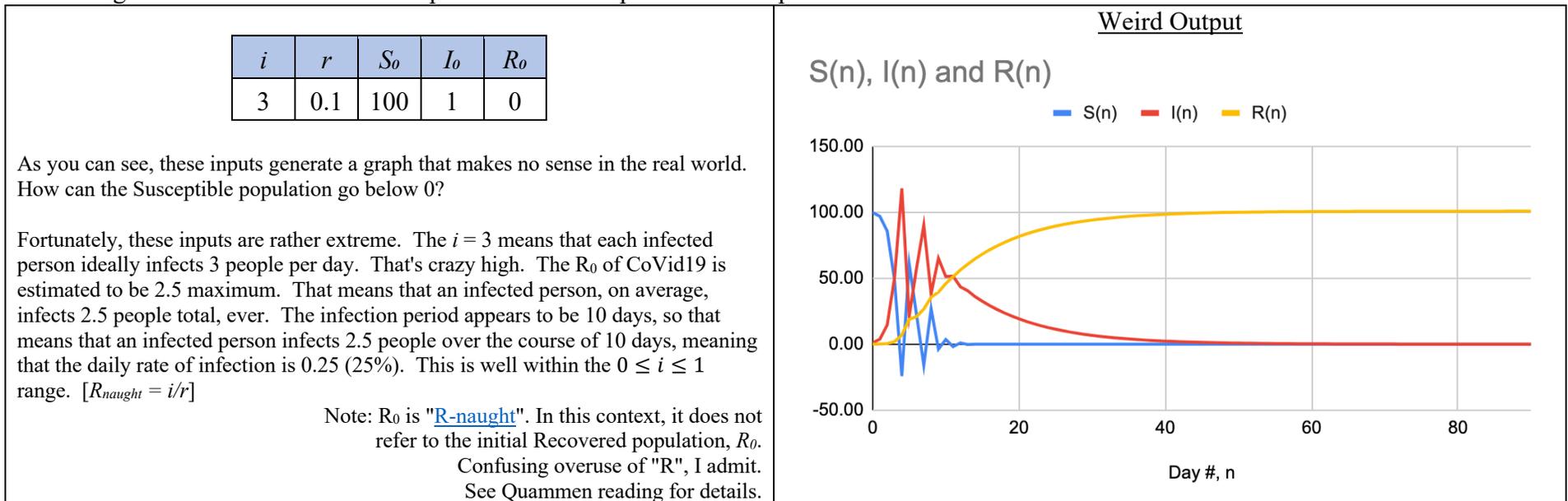# Making an SIR Model in GoogleSheets (or Excel or Apple Numbers)

The image below is a screen shot of the formula view of an SIR Model. I'd like you to make your own version of this. Let me walk you through it.

	A	B	C	D	E	F	G	H
17		Infection rate	recovery rate	Initial Sus. Pop	Initial Inf. Pop	Initial Rec. Pop	Total Pop.	
18	Input in	i	r	S_0	I_0	R_0	T	
19	blue cells:	0.5	0.25	10	1	0	=D19+E19+F19	
20								
21			Day #, n	S(n)	I(n)	R(n)	T(n)	S/T
22			0	=D19	=E19	=F19	=D22+E22+F22	=D22/G22
23			1	=D22-$B$19*E22*D22/G22	=E22+$B$19*E22*D22/G22-E22*$C$19	=F22+$C$19*E22	=D23+E23+F23	=D23/G23
24			2	=D23-$B$19*E23*D23/G23	=E23+$B$19*E23*D23/G23-E23*$C$19	=F23+$C$19*E23	=D24+E24+F24	=D24/G24
25			3	=D24-$B$19*E24*D24/G24	=E24+$B$19*E24*D24/G24-E24*$C$19	=F24+$C$19*E24	=D25+E25+F25	=D25/G25
26			4	=D25-$B$19*E25*D25/G25	=E25+$B$19*E25*D25/G25-E25*$C$19	=F25+$C$19*E25	=D26+E26+F26	=D26/G26

The input cells are in blue and are specifically B19-F19. These are the initial conditions that prime the pump of this SIR Model. The reason I made this starting on row 17 is so that I'd have room above it for a graph of the data. I'll discuss graphing after I discuss the mathematical details.

B19: This is where you input the infection rate. In this specific model the infection rate, i, should stay somewhere between 0 and 1 (0% and 100%). Values greater than 1 will sometimes produce weird output. Weird output looks like this:

i	r	S_0	I_0	R_0
3	0.1	100	1	0

As you can see, these inputs generate a graph that makes no sense in the real world. How can the Susceptible population go below 0?

Fortunately, these inputs are rather extreme. The $i = 3$ means that each infected person ideally infects 3 people per day. That's crazy high. The R_0 of CoVid19 is estimated to be 2.5 maximum. That means that an infected person, on average, infects 2.5 people total, ever. The infection period appears to be 10 days, so that means that an infected person infects 2.5 people over the course of 10 days, meaning that the daily rate of infection is 0.25 (25%). This is well within the $0 \le i \le 1$ range. [$R_{naught} = i/r$]

Note: R_0 is "R-naught". In this context, it does not refer to the initial Recovered population, R_0. Confusing overuse of "R", I admit. See Quammen reading for details.

Weird Output

S(n), I(n) and R(n)

— S(n) — I(n) — R(n)

Day #, n

There are ways to avoid this and allow for $i > 1$, but it complicates the formulas and for now this limitation is perfectly adequate.

Note: Measles, for example, has a particularly high reproduction rate, R_0 (R-naught), of 18 max. An infected person is typically infectious for about 8 days, meaning that for our SIR Model the infection rate, i, would be about 18/8, which is 2.25 per day. [Where $i = \frac{R_0}{P}$, and $P = duration\ of\ infectiousness$.] So, a person with Measles infects 2.25 new people every day for the duration of the infectious period, which is about 8 days. $8\ days \times \frac{2.25\ infections}{day} = 18\ new\ infections$.

Sorry for that digression.

C19: This cell has r, the recovery rate. Generally speaking, it is the rate at which an infected person "recovers," or at least becomes *not infectious*. CoVid19 seems to be infectious for 10 days (on average). Thus a recovery rate for CoVid19 is 0.1/day, or 10%/day. Meaning that after 10 days an Infected population individual will no longer be infectious and will be moved into the Recovered population, R. The interpretation of this number is somewhat problematic in this model because if 10 people are infected and have r = 10%, then one of them recovers in a day.

$$10\ Infected \times 10\% = 1\ Recovered.$$

You would expect that if 1 person started the infection on Day-0, that person would be the first to recover, and it would take 10 days, but this isn't what happens in this SIR Model. As the numbers of Infected people rise their collective recovery also rises, so that in this model you see people recovering in less than 10 days from the start of the outbreak... because on Day-6, for example, there might be 10 Infected people with r = 10%, leading to a recovered person.

D19: This one is pretty straight forward. It is the initial population of people Susceptible to the infection. For CoVid19 this appears to be everybody, however it may turn out that there are immune people out there. So for cell D19 you simply input whatever size population you want to look at. Bard College would be about 2000 people. New York City would be 8 million.

E19: This is also straight forward. It is the initial population of Infected people. I usually set it to 1, a Patient-Zero. But you could set it to whatever you like.

F19: Again, straight forward. The initial Recovered population. I usually set it to 0. However, you can use it to model a population of immune people or people who are vaccinated or people who are quarantined. If you are looking to find the necessary number of people to vaccinate in order to acheive "herd immunity," this might be the population to work with.

That's it for the inputs. Now the formulas.
-D22, E22, and F22 simply copy the inputs from D19, E19, and F19.
-All the formulas found in column G just add up the S, I, and R populations of each row. I used this to see if everything was working.
-Formulas in column H just keep tabs on the ratio between Susceptible and the Total populations. Interesting to see this change.

And here are the actual SIR formulas:

D23:	S_1	=D22-$B$19*E22*D22/G22	Susceptible population on Day-1	$S_1 = S_0 - iI_0 S_0 / T$
E23:	I_1	=E22+$B$19*E22*D22/G22-$C$19*E22	Infected population on Day-1	$I_1 = I_0 + iI_0 \left(\dfrac{S_0}{T}\right) - rI_0$
F23:	R_1	=F22+$C$19*E22	Recovered population on Day-1	$R_1 = R_0 + rI_0$

Most of the stuff in these formulas is pretty straight forward, but the dollar signs might be confusing. Let me explain. If I simply typed "=F22+C19*E22" into cell F23, without dollar signs, it would work fine in cell F23. But if I selected cell 23 and dragged it down to copy the forumla into the cells below, I'd get this. See picture to the left. You'll note that all of the cell references change. They all count upwards as they go down in the row. F22 becomes F23, then F24, F25, etc. Similarly, C19 turns to C20, C21, etc. And E22 goes to E23, E24, E25, etc. But if you look at where these cells refer to, the formula no longer makes sense with the C-cell references. The references to the F and E cells are fine, but the C-references are no good. That's why the $. The $ tells the spreadsheet not to incrementally change the cell reference when you drag-copy it. This explains the =F22+ **$C$19***E22. See picture to the right. I typed in the top formula, drag-copied it, and it kept the reference to C19 intact. That's why the $.

=F22+C19*E22
=F23+C20*E23
=F24+C21*E24
=F25+C22*E25
=F26+C23*E26
=F27+C24*E27

=F22+$C$19*E22
=F23+$C$19*E23
=F24+$C$19*E24
=F25+$C$19*E25
=F26+$C$19*E26
=F27+$C$19*E27

All the other cells in the spreadsheet are labels, identifying what a particular cell is. I then color coded various parts, added borders to various cells, etc. Decoration with a functional purpose. I won't explain all the reasons behind my choice in formatting, that's up to you.

Also be aware that the cells can have different number formats. E.g. percent, with commas, rounding off limits, etc. You'll have to figure out how that stuff works. I may make a video showing off some of these features.

Now it is time for you to make this SIR Model spreadsheet.

Make one for at least 30 days (that data found in column C).

Also figure out how to get a nice line graph of the output. [Also called "charts".]
This takes some poking around the spreadsheet program and is sometimes pretty annoying getting it to be the way you want it.

I don't need to see this spreadsheet , but I want you to do it. Learning how to model stuff in a spreadsheet is very, very useful.