LETTERS IN BIOMATHEMATICS
An International Journal

Intercollegiate
Biomathematics
Alliance

EDUCATION ARTICLE

OPEN ACCESS

Resources for Supporting Mathematics and Data Science Instructors During COVID-19

E. Cabral Balreira[a], Casey Hawthorne[b], Grace Stadnyk[b], Zeynep Teymuroglu[c], Marcella Torres[d], Joanna R. Wares[d]

[a]Department of Mathematics, Trinity University, San Antonio, TX; [b]Department of Mathematics, Furman University; [c]Department of Mathematics and Computer Science, Rollins College; [d]Department of Mathematics and Computer Science, University of Richmond

ABSTRACT
In late May of 2020, a few months after the raging COVID-19 pandemic forced university faculty to quickly switch to online teaching, the Associated Colleges of the South (ACS) released a call for grant applications to support working groups "to help faculty within our consortium who will be teaching during the pandemic (e.g., from hybrid courses with some remote/online components to fully remote/online courses; socially distanced face-to-face courses)." We replied to this call and the ACS awarded the six of us (from four ACS schools) a Summer Rapid Response Grant in early June. The grant funded our efforts to create and provide to other faculty Mathematics and Data Science Resources to Support Socially Relevant Teaching in the Time of COVID-19. This paper summarizes our efforts and includes the resources that we developed.

ARTICLE HISTORY
Received October 7, 2020
Accepted March 12, 2021

KEYWORDS
COVID-19, SIR Model, R_0, Data Visualization, Calculus, Differential Equations

Introduction

From June to August 2020, we participated in weekly virtual meetings and worked asynchronously to collaboratively develop modules for mathematics and data science courses for use by ourselves and other faculty for the fall of 2020 and beyond. The goals of our efforts were twofold: to create inclusive foundational and upper level activities about modeling COVID-19 that could be utilized in a variety of settings, while at the same time providing advice and support to faculty who want to discuss social justice issues surrounding COVID-19 in their data and modeling oriented STEM classes. In our modules, we focused on creating activities, lessons, and projects that address complex issues related to COVID-19 such as inequality in testing, wealth distributions, or race/ethnicity issues. In addition, we tried to incorporate different teaching pedagogies that fit well with hybrid, face-to-face, and virtual teaching modalities. Each module comes with supplementary resources: YouTube videos, peer-reviewed journal articles, textbooks, and simulation tools that are easy to access.

The nine modules work well alone or in sequence. The module names are:

1. Let's Talk About COVID-19
2. Persuasion and Data Interpretation
3. Exploring COVID-19 Through Charts and Graphs
4. Estimating R_0 from Real Data
5. Race and Ethnicity in COVID-19 Numbers: FL case study
6. Hypothesis Testing: A quantitative approach to the scientific method
7. Mathematical Modeling of Epidemics: From influenza to COVID-19
8. Flatten the Curve
9. Differential Equations Semester-Long Project

On our website (Wares et al., 2020) where you can access these modules along with student handouts, we suggested courses where these modules would be useful and suggested particular orders to present them. These are of course just suggestions and you can leave out any particular modules and the other should still stand alone. For all courses, preparing the class with the "Let's Talk

CONTACT Joanna R. Wares `jwares@richmond.edu`

About COVID-19" module can make students feel more comfortable while providing the faculty member knowledge about what particular students are experiencing in their lives that might help them be sensitive to these issues in the classroom.

As an example of a sequence for a particular class, we thought an Intro to Modeling course might benefit from the presentation of **1) Hypothesis Testing: A quantitative approach to the scientific method** and then **2) Flatten the Curve**, followed by **3) Estimating R_0 from Real Data** and finishing with **4) Exploring COVID-19 Through Charts and Graphs.**

Below, we discuss each of the activities in brief, suggesting how they might be useful in your own class. At the end of the paper, we provide an instructional guide and links to supplementary material for each module. Our hope is that the pertinency of these activities will outlast the virus. We encourage instructors to download or record Youtube videos for future use. A URL for a source might not stay active. Instructors should save or print out a copy of the resources for these modules. In addition, libraries might choose to participate in perma.cc's (`https://lil.law.harvard.edu/projects/perma-cc/`) network to preserve online resources.

1 Let's Talk About COVID-19

We are all adapting to changes that are resulted from the COVID-19 pandemic. As instructors, we want to use this opportunity to create culturally and socially relevant assignments and projects for our upcoming mathematics courses. What a better time to show that an exponential growth function, logarithmic scales, SIR models, etc. are in fact applicable and useful in solving real-world problems? As we update our course material and make it directly related to the COVID-19 pandemic mathematical models, it is important to remember that some of our students are personally affected by COVID-19. We all know that providing opportunities for students to gain hands-on experience with real data is a good thing. However, the upcoming semester might be too soon for our students (and for some of the instructors) to work on homework assignments or projects dealing with the COVID-19 pandemic. Students might have had the virus themselves or lost a family member or a loved one to the pandemic. We encourage instructors to create a safe environment where students can freely and comfortably share their concerns and questions about assignments and projects that directly deal with the pandemic. This discussion could take place in the classroom, in one-to-one meetings with the students, or in the form of a writing assignment. The module provides a template for a writing assignment to help students discuss their experience with the COVID-19 pandemic. In addition, the writing assignment prompts them to discuss COVID-19 related death rates through a social justice lens connected to racial inequalities, health disparities, and socio-economic status.

Instructors should make it clear that these assignments and projects are targeted to help students analyze, understand, and critically evaluate a public health problem and its proposed solutions. It is not intended to make students feel sad or fearful. Students should be given an option of opting-out of completing COVID-19 related assignments and projects. For students that choose to opt-out, these projects can be substituted for similar projects with the same learning goals. For example, the prompt for another reflection paper could be as follows: How have your race/ethnicity, socioeconomic status, and gender impacted your mathematics education?

2 Persuasion and Data Interpretation

Mathematics holds a particular position of authority in Western society. Whenever we are presented a numerical argument, we tend to believe it is true. While data can elevate aspects of the world that otherwise go unnoticed, it can also obscure other facets. We must remember that by the time we as consumers receive information, the data has potentially been filtered through many different people or groups. Each person analyzing the data has conscious and unconscious biases that are intentionally or unintentionally reflected in the delivery of the data. This module serves to highlight that how we define and measure different qualities is subjective. Moreover, such choices are made by those in power, shaped by their implicit and explicit biases, inherently to serve their own benefit.

To support students in developing the skills to interpret and question data driven claims presented to them, we designed four different activities all centered around President Trump's repeated claim that the United States should reduce testing for COVID-19. Students begin by exploring how to define and measure a quality such as being good at testing. They then analyze different ways countries are attempting to answer this question and what type of data they are using. This work involves discussing the affordances and limitations of different approaches, but also different visual representations used to communicate these data. The module culminates in a final activity in which students use data to write three press releases reporting on the US's COVID-19 testing program, each from a different point of view: one in which a student is asked to support Trump's claim, one in which they are asked to refute Trump's claim, and one in which they attempt to remain unbiased in presenting the data.

3 Exploring COVID-19 Through Charts and Graphs

We are inundated with graphical representations of data in our daily lives: in our classrooms, in our jobs, on the news, on social media, in advertising, and more. We often process these visuals quickly and subconsciously, so without careful consideration, it is easy to miss important pieces of the story, or misinterpret the data presented entirely. Furthermore, as creators of different graphs try to generate more detailed and eye-catching visual representations, graphs have become more and more complex. Whether intentionally or unintentionally, different choices made about the design of graphs and charts can obscure some facets of the data while emphasizing others. While graphs can be a powerful way to communicate data relatively quickly, informed citizens must possess the ability to critically interpret them.The objective of this module is to support students in developing this skill.

This module consists of four activities, but as with most of the modules included in this collection, these activities can generally be completed independently of the others. The first activity introduces students to some of the ways in which graphs can be misleading. The second activity invites students to dive into a wide variety of charts, especially observing both benefits and limitations that underlie different design elements from the different graphs. This activity would work well as a regular (daily or weekly) activity through which students can develop a range of analytical skills and an orientation to reflect and ask thoughtful questions when digesting visual representations of data in their everyday lives. Students explore the disproportionate effect of COVID-19 on minorities in the third activity, analyzing a collection of graphs displaying race/ethnicity-related COVID-19 data. The culminating activity asks students to reflect on and summarize what they've discovered throughout the module. Instructors can choose to supplement the graphs and charts in this module with additional location-specific charts or other charts they deem especially relevant and thus potentially more engaging to their students.

4 Estimating $\mathbf{R_0}$ from Real Data

In today's media coverage of COVID-19, students are becoming familiar with epidemiological terms. One of the main terms that has kept the media's attention from day 1 is the basic reproduction number, R_0. Most of our students might look at the value of R_0 and quantify whether there will be a disease outbreak. If $R_0 > 1$, on average an infected person can infect more than one person during the period of infectiousness in a completely susceptible population. Therefore, it is likely to expect a disease outbreak. If $R_0 < 1$, we expect that the number of new cases is not high enough to cause an outbreak. R_0 is also a key metric that policy makers closely followed in their preparation to allow states to open back up after stay-at-home orders were put in place.

This module presents a mathematical framework for estimating R_0 from real data. It uses an early data set of daily positive COVID-19 cases in Franklin City and South Milwaukee for the time period, April 4 – 10, 2020. We should note that the number of infected cases appear to grow exponentially in early April as the epidemic is still at an on-going stage. As students make an observation about the exponential growth model fit to the early stages of the epidemic, we encourage instructors to address why the exponential growth model is not sustainable to capture the long-term behavior of the disease spread by introducing the Flatten the Curve module in an earlier class. By fitting an exponential growth model to the new cases of COVID-19 in the early stages, we can calculate the growth factor, which then can be used to estimate R_0. In summary, this module helps students understand that key factors in mathematical models are in fact estimated from real data. Therefore, students will learn that the R_0 estimate is as good as the data itself and it is not the be-all and end-all result to fight the COVID-19 pandemic.

5 Race and Ethnicity in COVID-19 Numbers: FL case study

In this module we are interested in quantifying if there is any relationship between COVID-19 case numbers and demographics. We provide students with an interactive map of Florida where case numbers per 100,000 inhabitants are provided per county. The data for the activity is frozen with data from Aug. 1, 2020. In addition, we provide maps with Race, namely percentages of African Americans per county and Ethnicity, namely the percentage of Hispanic or Latino residents per county. We expect students to conjecture and visualize if there is any relationship between the number of cases and Race and Ethnicity. Although there is visual evidence, the goal is for the students to provide quantitative arguments to support that observation. In the second activity, we provide the data behind the maps so students can plot the data and use trendlines to compute correlations. We start with the computation of cases per 100,000 inhabitants so the data is comparable, then we look at Race, then Ethnicity, and finally combining the minority populations students confirm that there is a high correlation between Race and Ethnicity in COVID-19 numbers in FL. Finally, in the last activity, students are asked to analyze and explore if there are reasons for these correlations and propose other explanations.

6 Hypothesis Testing: A quantitative approach to the scientific method

This module was offered as part of a two-part mathematics workshop for underrepresented and first generation freshman undergraduate students in a summer research training program focused on biology and chemistry. The first goal was to introduce the idea of mathematics and statistics as quantitative approaches to experimentation and hypothesis generation and testing to show the connection between mathematics and other STEM disciplines. The second goal was to motivate further study in this area by applying hypothesis testing to social justice and public health issues. Since no statistics experience is assumed, we scaffold the activity with preparatory videos and use an intuitive random sampling approach to hypothesis testing.

We begin the activities by watching a video presentation on YouTube from Dr. Omayra Ortega on The Statistics of Driving While Black. After discussing her approach as a class, we apply the same methods to COVID-19 data from Washington, D.C. to test the hypothesis racial disparity exists in health outcomes. Group and class discussion is woven throughout, and the scientific process is emphasized: students are asked how they would use our results to generate new hypotheses or to determine what data would be necessary to strengthen our existing hypothesis.

Because the topics covered are uncomfortable, instructors should let students know they will be covered in advance and ideally frame the module as part of a larger class discussion about social justice and how mathematics and data can help or harm in these cases. Overall, student comments in the "1-minute paper" follow-up activity to the module reflected a theme of empowerment and the necessity of addressing these topics.

7 Mathematical Modeling of Epidemics: From influenza to COVID-19

This module introduces calculus students to SIR (Susceptible-Infectious-Removed) infectious disease models in the mathematical context of bridging discrete to continuous models, connecting limits to derivatives, comparing transmission and recovery rate parameters between two disease models, and simple data fitting. An ideal placement in the calculus curriculum for Activities 1 and 2 is early in the transition from limits to derivatives. Activities 3 and 4 stand alone as introductions to coding of SIR models using Google Collab and parameter estimation using a spreadsheet. All of the activities completed together have made a nice introduction to the modeling process for a differential equations class.

In the preview work for the modules, students are introduced to SIR models by watching a short video and reading an article. Activity 1 walks students through modeling the first 10 days of an influenza epidemic in a boarding school using a discrete SIR model in a spreadsheet. This discrete exercise, in which the time step is 1 day, allows students to get a sense of how transitions between subpopulations occur at each time step. In Activity 2, students are shown how limits are used to shrink the time step until we have a continuous set of derivative equations. Students use the intuition they developed in the first exercise to analyze these equations.

Activity 3 provides a gentle introduction to reading and altering code for SIR models. A linked Google Collab worksheet provides some Python code for the influenza SIR model, which can be accessed and run without downloading software. Students are asked to read and interpret the code. Next, students are asked to copy the code and alter the parameters to create a model for COVID-19. This second part of the activity allows students to speculate on how parameters might change between the influenza and COVID-19. Finally, students are asked to reflect on such questions as the reasonableness of their simulation results, where they might look up information on model parameters, whether they would expect parameters to vary by location, and what limitations might exist in the current model -do they think all populations are equally susceptible?

The final activity allows students to attempt to fit real influenza outbreak data to the discrete model that was developed in Activity 1. The purpose is to both introduce students to dependence of model predictions on parameters and the need for parameter estimation, and to introduce them to the difficulty in manipulating more than one parameter to fit data. This is done simply by changing parameters in the second tab of the discrete model spreadsheet.

8 Flatten the Curve

In this foundational exercise, students are introduced to mathematical modeling as a predictive tool and also as a method of examining the underlying dynamics of a system. Preparatory work asking the students to watch a video and answer questions is included. After a brief introduction into the components of modeling and an example of a good predictive model that gets the dynamics wrong (this example can be easily removed), students begin to work with real data about COVID-19 taken from the CDC (data in Google Sheet included).

The first exercise uses Total Number of Cases data that the students create from the New Daily Cases data from the CDC. The students separate this data into the early days and the latter days. Early data is almost exactly exponential and the latter is much closer to linear. Students are walked through creating these fits in Google sheets and are asked questions about using these fits as models. The main revelation is that neither model is a good predictive model.

The second part of the module utilizes a premade website that leads students through many different scenarios using the SIR differential equations model. At the bottom of the website, students are able to adjust parameters on their own. They are then asked to answer questions about different parameter choices, how different interventions enter the model, and how one could use the model to investigate the inequities seen in the COVID-19 data

9 Differential Equations Semester-Long Project

This module provides structure for team projects where students develop and analyze their own differential equations models. This module could easily be adapted for other courses where a semester-long project would be appropriate, by changing the differential equations specific language.

In this module, students are split into teams. At the beginning of the semester, students work to choose a question about COVID-19 that they would like to attempt to answer using a differential equations model. The focus of the project is on creating the model (which can be an extension of known models) and on being able to explain the assumptions of the model and also the variables and parameter choices. Analysis is done in the end and students try to answer their original question, but the focus is not on the results as much as the modeling.

Throughout the semester classes are reserved for groups to work together. Mid-way through the semester, teams present a "pitch" where they explain their question and assumptions. It is a good idea to have the other students offer feedback and questions in whatever way you find best. Students take the feedback given in the pitch presentations and continue to work on the rest of their projects for the rest of the semester. In the last week of classes, teams present their final work. You could also do this as a paper write-up if you prefer.

The instructor's guide provides topic ideas in case the students are having difficulty choosing a project. This module would nicely follow the Flatten the Curve Module and/or the Mathematical Modeling of Epidemics: From Influenza to COVID-19 Module.

10 Modules

10.1 Let's Talk About COVID-19

The novel coronavirus has affected lives all around the world, including all of ours. Before we discuss it in class, I want to know how it has been affecting yours. If you feel comfortable sharing your personal journey with COVID-19, please complete the following assignment for me. If not, please use news articles to complete the same assignment but from a less personal point of view, or talk to me about how we can accommodate you to do other work.

This exercise asks you to write a short paper reflecting on your personal experiences (or those you have read about) with COVID-19. Please only share what you feel comfortable sharing. Please do not share your health history. The prompt for this reflection paper is as follows: How is the COVID-19 pandemic affecting your life? Do you know people who have been affected by COVID-19? Do you believe that race/ethnicity, socioeconomic status, gender, lack of health insurance, and other connected factors impacted your/their experience with COVID-19? Do you believe that it is important to address those issues in the class? Why? Why not?

Please keep your paper short and succinct. If there is anything that you want to tell me in private in a non-written manner, I would love to speak with you about this also, at any time during the semester. The precise format for the paper is open.

Modified from Math for Social Justice Writing Assignment by Dr. Nathan Alexander (Morehouse College) Instructors might design an alternative assignment for students to discuss prior experiences of learning mathematics if some students do not feel comfortable with the assignment above.

10.2 Persuasion and Data Interpretation

Optional tech resources used: Zoom for remote learning, access to Google Sheets for collective use of spreadsheets or documents, access to internet to explore other data.

How to use this guide: Notes to instructors appear in italics.

Discussion activities are intended to engage students in the question and to help them understand the questions that researchers must ask and answer. We suggest you include at least one discussion activity from this module, though it is reasonable that you might not want to include all of them due to time constraints or constraints resulting from the mode of course delivery.

Possible Courses: Introduction to Statistics, Finite Mathematics, Intro to Political Science/Analysis

Objectives/Purpose:

1. Understand and investigate what limitations and challenges arise in attempting to find data and answer questions using statistics.
2. Understand that data does not speak for itself. As such, data analysis is inherently subjective and can be used to support or refute politically-motivated claims or other biases in general.
3. Read and analyze different graphical representations of the same or similar data. Learn to identify various affordances and limitations of different graphs.
4. Engage in the critical reading of statistical analyses.

Introduction: Math tends to be viewed as universally correct. If data supports a particular claim, then we believe it must be true. However, as we will see, there are many choices that scientists make that can affect the outcome and results of a study or an analysis of data. By the time the end consumers (such as those people who read news articles or watch the news on TV) receive the information, the data has potentially been filtered through many different people or groups of people. Each person analyzing the data has conscious and unconscious biases that can be intentionally or unintentionally reflected in the delivery of the data. This module serves to highlight that how we define and measure different qualities is subjective. In terms of equity, we must remind ourselves that such choices are made by those in power. They inherently bring in their biases, implicitly, and at times explicitly, serving their own benefit. Helping students to use mathematics to critique and understand the world empowers them to question our present social order and to act to hold those in power accountable.

When asked about the rate of Coronavirus testing performed in the United States in a TV interview on June 21st, 2020, Trump claimed "If you want to know the truth, we've done too good a job." Vazquez (2020) Are we doing a "good job" at testing for COVID-19 in the United States? What does it mean to do "too good" a job? What point was Trump trying to make when he said this? According to Johns Hopkins' Coronavirus Resource Center, the United States has performed more Coronavirus tests than any other country in the world. But is this information enough to determine whether the testing protocol in the United States is sufficient?

You will investigate what information you need in order to answer these questions (Activity 1, 3) and how to think critically about how this data is presented to you (Activity 2). Then you will attempt to present the data in three different points of view: one in which you support Trump's claim, one in which you refute Trump's claim, and one where you attempt to remain unbiased in presenting the data (Activity 4).

Suggestion to Structure Group Discussions: With all breakout sessions, start by allowing each person at least 1 minute to individually think about the task so everyone has different ideas and perspectives to contribute. Then assign roles to distribute participation.

Scribe This person takes notes of the different ideas that emerge. The scribe could use a different color pen/font for each person's contribution. This way the students and teachers can all see how the ideas are being distributed.

Facilitator You should start each breakout session having each group member share his/her idea. This person calls on the people and ensures no one talks too much or too little.

Presenter This person will present the group's ideas to the whole class discussion. With each activity, the group should rotate through these roles. If a group has more than 3 people, then some might not have a role with every iteration.

10.2.1 Activity 1: Choosing Data

Part I – Brainstorm Ideas:

Think/pair/share; Zoom breakout discussion rooms

Students will brainstorm different ways to effectively measure testing by answering the following questions. Groups should take notes on the different ways they come up with. If done electronically, students can write up their ideas on a collective Google Doc. Afterwards, different groups will share their ideas to the class with the instructor taking notes.

- How do you think we should measure testing effectiveness?
- What data would we need in order to measure this?
- Why is this an effective way to measure testing? (This prompt can be omitted initially and added as a separate breakout, see Part II)
- What problems might arise in the collection or interpretation of this data? (Similarly, this question could be included in the initial breakout session, although we suggest making this a separate breakout, see Part III)

In general, the gross number of tests might not be the best measure to compare countries. Instead, the rate of testing must be a relative measurement (a ratio), taking into consideration such things as the population of the country, the current spread of the disease in the country, or the wealth of the country. One might also look at how testing is changing, meaning the number of tests now relative to the number of tests in the past.

Possible student answers or talking points to extend discussion:

- Measure number of tests as a percentage of the total population of a country
    - Need: number of tests administered, total population per country.
    - Countries with more people will obviously be testing more.
- Measure number of positive results (i.e. have COVID-19) as a percentage of the total number of tests administered (*this is what we explore in the next activities and discussions).
    - Need: number of tests administered, number of positive results.
    - If countries have a larger spread of the disease, they need to be testing more.
- Number of tests per confirmed case (This is simply the inverse of the above measurement which a lot of sources use.)
    - Need: number of tests administered, number of positive cases in the country at the time
    - Whereas the above statistic seems to put the focus more on the spread, the inverse seems to put the focus more on testing.
- Measure number of tests as a percentage of hospitalizations or deaths.
    - Need: number of tests administered, number of hospitalizations or deaths.
    - Similar to the above issue. How serious a health concern the disease is should dictate the amount of testing.
- Number of tests relative to countries' wealth
    - Need: number of tests administered, GNI (or similar statistics) per capita
    - The effectiveness of testing should be measured by the resources a country has. It is unfair to compare a wealthy country to poor country when the poorer country will have less money to administer tests.
- Tests per day
    - Need: number of tests administered on the most recent day (This might also be measured relative to the population size.)
    - The quality of testing differs over time. While a country might have conducted many tests overall, we should compare how they are performing currently.
- Percent increase/decrease of tests
    - Need: number of tests administered now, number of tests administered at a date in the past
    - Similar to above. If a country has improved over time, it is the trajectory of the testing program that distinguishes countries.

If some of these ideas do not emerge in discussion, teachers might consider introducing them and having students discuss why or why not they might be productive ways to measure effective testing. We suggest introducing ideas in a neutral way. One way to do this is to attribute an idea to someone else, rather than one that you as the authority endorses: "In the last class, one group said…" or "Last night, I brought up this idea to my wife and kids, and they said…"

Part II – Justify Choices: Once all the groups have shared, the teacher can list all the different ideas that emerge (or pick the 4–6 most viable) and then ask students to choose which one they believe is the "best" way. This can be done physically by having students move to different parts of the room or electronically on a Google doc where students type their name or drag an icon in different places on the document (see Activity I-Vote).

Students are then asked to defend their chosen method. With this option, students should only answer Questions 1 and 2 in Part I, as this activity explores Question 3.

Alternatively, once students have made their choices, teachers can place students in groups or breakout rooms based on their particular choice. Students are then asked to collectively make an argument for why their choice is optimal.

Part III – Anticipate Problems:

Think/pair/share; return to Zoom breakout discussion rooms

In this part, students investigate potential issues with obtaining their data or measuring the above quantities. The main takeaway is for students to realize that while data might be labeled the same way, the methods for obtaining this data or measuring might be quite distinct, resulting in qualitatively different data.

- Some countries may choose to mask their numbers for political (or other) reasons (e.g. China, Brazil),
- Some countries may have differing methods of counting these things (for example, deaths may not accurately be recorded as due to coronavirus, one country might consider each test conducted while another only measures each person tested)
- Some countries may be using different kinds of tests that may have different sensitivity or specificity rates.
- Another consideration is how to even calculate the number of tests. Should we consider each test conducted or each person who is tested? Many of the same people are tested quite frequently.
- Depending on the health care system established in certain countries, if the cost of the test is not covered by the government, this might prohibit poorer people from accessing testing resources. Furthermore, poorer people might tend to be those who end up working in closer proximity to others and would thus be more susceptible to contracting the virus. Rather than looking simply at total numbers, we might look at policies that ensure access to testing or other issues of equity around testing such as the distribution of testing among the population.
- Other considerations might be: How might we handle concerns about reliability of the data? Are numbers updated in real time? Are they delayed? Backfilled? How might this affect any analyses?

10.2.2 Activity 2: Analyzing Graphs

Part I – Discussion:

Think/pair/share; Zoom breakout discussion rooms

As an effective measurement, many of you may have suggested looking at:

1. the number of tests conducted as a percentage of the total population in a country (also known as the tests per capita) or
2. the positive testing rate.

Both of these statistics are presented in three different representations provided by the University of Oxford, in the two sets of Google slides below. The first three graphs display the number of tests relative to 1000 people. The second three show the number of tests per confirmed case. All graphs were retrieved on June 26. See the links on the slides attached below for updated versions.

Divide the class up with half of the groups analyzing the testing per capita graphs and the other half analyzing the tests per confirmed case graphs. Each group should compare their three graphs and then answer the following questions:

- What can you see in one graph that you cannot see in the others and vice versa? In other words, do you come to different conclusions depending on what graph you are looking at?
- What questions do you have when looking at these graphs? What other data would you like to see and why?

Activity II-Analyzing Graphs (a)
Activity II-Analyzing Graphs (b)

Tests per Capita

Time series:

- *It is clear that Iceland initially had a huge testing program and now is similar to other OECD countries.*
- *Because of Iceland's large initial numbers, it is harder to compare the other countries.*
- *Most wealthier countries have had a similar, gradual increase in testing with the exception of South Korea which had more initially and has tapered off.*
- *You can see that Mexico and India, countries that have low GDP per capita have not tested much.*

Bar graph:

- *Since the countries are listed greatest to least, it is very easy to see which countries are currently testing the most per capita.*
- *You can easily read the actual numbers.*
- *You can only see a portion of the countries.*

Map:

- *Geographically you can see that wealthier countries (US, Canada, EU, Australia, New Zealand) have more testing.*
- *The countries that have no data are prominent. You can see that there is no data from about half of the world.*

Tests per Confirmed Case

Time Series:

- *Australia is testing the most according to this metric, as of June 26th. It is conducting 2000 tests per confirmed case.*
- *Among the set of countries shown, the US ranks 8th in the number of tests per confirmed case, as of June 26th.*
- *Since Australia's number of tests per confirmed case is so much larger, it is nearly impossible to see how many tests are being conducted per confirmed case for the other countries shown. It is also nearly impossible to see the trend, except for the three largest testers (Australia, South Korea, and Iceland).*

Bar Graph:

- *It is much easier to see the numbers on this graph than it was on the time series.*
- *According to this graph, it is easy to calculate that Australia is conducting over 116 times more tests per confirmed case than the US is!*
- *The US is testing at roughly the same rate as India, according to this statistic.*
- *We cannot see the trend in testing over time.*

Map:

- *We can see data for many more countries at once using this graphic.*
- *The data for the different countries is not precise in this graph. Instead, we are given ranges of tests conducted per confirmed test*
- *The scale is inconsistent for the different colors. For example, the orange color indicates that there are between 10 and 30 tests conducted per confirmed case, while the dark blue indicates there are over 1000 tests conducted per confirmed case in that country.*
- *It is easier to see which countries conduct a similar number of tests per confirmed case.*

Part II – Comparing Graphs: In the US we often report the Positive Testing Rate as a way to measure the spread of the virus. This measures the percent of all people who tested positive out of all the people tested. For example, on July 1, 2020, this rate was just over 25% in South Carolina. That means for every 100 people who took a test, 25 of them tested positive. This data point is explored in more detail in Activity 3. For now, we will compare this data point with the "Tests per Confirmed Case" data point.

Now that students have discussed affordances of different types of graphs, they are going to compare different data points across various pairs of graphs. The first data point is the Positive Testing Rate and the second data point is "Tests per Confirmed Case," which was represented in the diagrams Group (b) just analyzed.

Compare the two sets of graphs in the slides here: Part II-Comparing Graphs. What do you notice between each corresponding set?

Time Series:

- *I'm not sure I see a relationship. In the first one, the lines are more spread out throughout and in the second they are initially close, but then spread out. The countries on the top in one graph are on the bottom of the other.*

Bar Graph:

- *The countries in one graph are in reverse order (except Taiwan and Australia) than their order in the other graph.*

Map:

- *These maps look mostly the same except for a few countries in Europe.*
  - Follow up: *Look at a single country or two countries. How do they change throughout the different representations?*
  - *What relationship, if any, do you see between the two statistics?*

10.2.3 Activity 3: The % Positive Statistic

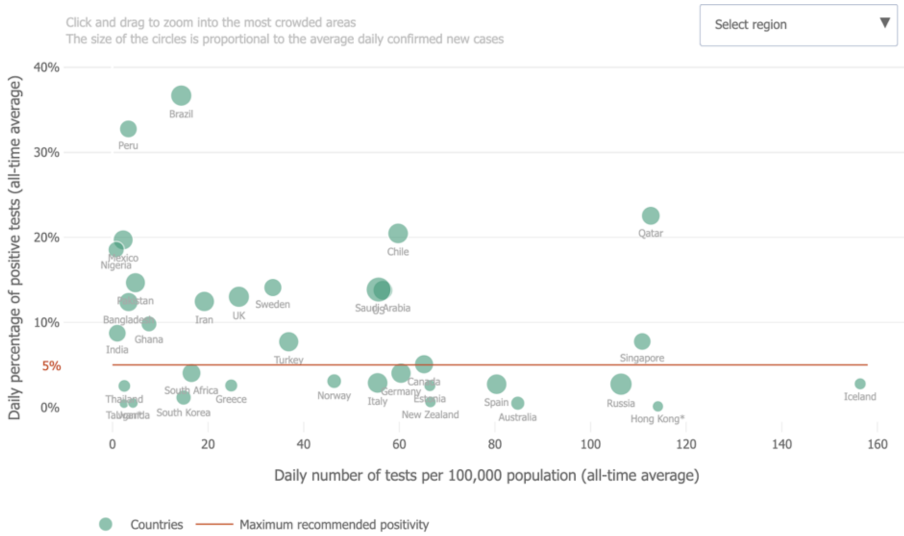
Part I – Discussion:

Think/pair/share; Zoom breakout discussion rooms

According to the World Health Organization, the % positive statistic is the best indicator of testing effectiveness. The following scatter plot provided by John Hopkins displays this statistic along with the tests per capita statistic.

In your groups, analyze this new representation and answer the following questions:

- What observations can you make and what conclusions might you draw from this graph?
- What is visible in this graph that is not visible in the other graphs?



(See Johns Hopkins University and Medicine, 2021 for the most current data.)

Possible student answers or talking points to extend discussion:

- *The x-axis measures the average daily number of tests of all time. This means the data might not reflect a country's changing or delayed approach to managing the virus.*
- *There is very little correlation between these two variables. The r-value would be quite low.*
  - Follow up: *This data is clearly not linear. Is that what you would expect? Why?*
- *Poorer countries tend to perform far fewer tests per day on average. Some wealthier countries seem to be testing far more, but this does not necessarily coincide with a lower % of tests positive for those countries (e.g. Qatar, which by IMF estimates, has the largest GDP per capita in the world. Although their Gini Coefficient puts them in the bottom third in terms of wealth distribution).*
- *Italy seems to have a relatively low % positive statistic (2.67%), despite being one of the hardest hit countries, at least initially.*
  - Follow up: *Donald Trump has argued that the reason the number of cases in the US has risen is because we are testing "too well". Does this graph support this claim or not? Explain.*
    - *The fact that the relationship between the rate of positive tests and the number of overall tests per capita is not an increasing linear relationship, this claim is false. For example, Iceland is an outlier in terms of the amount of tests they are conducting per capita, but they have a very small number of positive tests.*

Part II – Discussion:

Think/pair/share; return to Zoom breakout discussion rooms

1. Look at the data and compare Germany and Chile. Contextualize these statistics. What do they mean in terms of the pandemic for these two countries? How is what is happening in these two countries similar? different?
2. Do you believe the World Health Organization's claim that % positive tests is a more accurate way to measure testing effectiveness than number of tests or number of tests per capita? Why or why not?

3. What additional data might you want or need to make a more convincing argument about whether % of tests positive is a good measurement of a country's testing program?
4. What problems do you think might arise by using percent positive tests as a measure of testing-effectiveness for all countries?

Possible student answers or talking points to extend discussion:

- *In this graph, Chile and Germany have roughly equal testing rates per capita (60 tests per 100k people). However, Chile has 20.46% of its tests coming back positive, while Germany has only 4.02% of tests coming back positive. This would suggest that Chile is being hit harder by the pandemic than Germany (at least at the time that this data was collected).*
- *Looking at just the number of tests performed in a country relative to the total population (the statistic measured on the x-axis of the graph), assumes that COVID-19 is affecting every country equally. Since Chile is being affected more severely by the pandemic, it would make sense that Chile should perform MORE tests per capita than Germany. For example, we wouldn't test for a virus if we were not currently in a pandemic.*
- *Average daily new cases as a percent of the population could be a useful data point. Testing trends over time could be useful: are countries tending to increase or decrease testing? When testing increases and % of tests positive decreases, does this correlate with a reduction in daily # of new cases? How to handle false positive/false negative tests? Are all countries reliably reporting this data? Are all countries using the same test? If not, is it safe to compare these numbers or is this not an "apple to apples" comparison?*

Part III – Comparing Graphs: *In this part, students are led through a series of questions asking them to look at two different graphs, both of which include data related to the % Positive statistic. Students are asked to explore strengths and weaknesses of each graph, differences between the graphs, and which graph is more effective or impactful. Students are also asked to determine whether and how the choices made by the creators of the graphs affects their interpretation of the data. See Activity III, Part III - Comparing Graphs Worksheet. For an updated version of this graph, see Ritchie et al. (2020).*

Part IV – Determining the Number of Tests:

Think/pair/share; return to Zoom breakout discussion rooms

The World Health Organization has recommended that a country should ensure their % of tests positive is below 5% for 14 days before relaxing social distancing recommendations and requirements.

1. How many tests should a country be performing? In particular, what does this mean for Chile and Germany?
2. Assume that all those not being tested do not have the Coronavirus. Use the data from the graph in Part I and determine what percentage of the population each country needs to test each day to achieve 5%.
3. In reality, we know there are many people who have the virus, but are asymptomatic and therefore do not get tested. In fact, researchers now estimate that for every person who has the virus and is symptomatic, there are 2 people that have the virus but are asymptomatic.

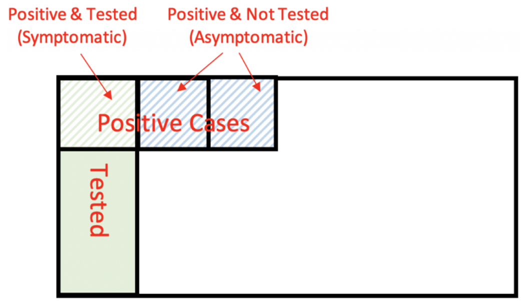
Let's assume that all people with symptoms get tested. Take a hypothetical country A in which 10% of the population gets tested with a positive testing rate of 20%. Using a rectangle to represent the entire population, highlight on the rectangle the part of the population that gets tested each day and how many of these people have the virus. Assuming that for each person who has tested positive, there are two people who have the virus, but are asymptomatic and consequently have not been tested. Include these people on the diagram. Using this representation, determine the following:

1. Percent of the population of country A with COVID-19.
2. Percent of untested people with COVID-19.
3. Explain why Country A, no matter what percentage of their population they test, will be unable to get their % positive testing rate below 5%.

Possible student answers:

1. Since Chile's 20% positive rate is too high, it needs to increase testing. Assuming the population that is not yet tested has a much lower % positive rate, this will lower the rate of those tested. Conversely, since Germany's rate is below 5% they can continue as they are now or even begin to perform fewer tests.
2. Chile's positive testing rate is approximately 4 times what it needs to be to relax social distancing $\left(\frac{20.5\%}{5\%} \approx 4\right)$ They are currently performing 60 tests per 100k, which needs to be 240 tests per 100k, which is 0.24% of the population. This is a minimum amount. Similarly, on the other hand, Germany's 4% positive rate is lower than 5%. They could continue with their current testing or, since they could reduce it by a factor of 1.25. This would mean they only need to conduct 48 tests per 100k or 0.045% of the population.

3.



- The percent of the population who has been tested and has the virus is 20% of 10%, or 2% of the population. The non-symptomatic people are double this or 4% of the total population. This makes a total of 6%.
- This would be the ratio of 4% to 90% or 4.44%.
- This is because more than 5% of the whole population has the virus. They could test everyone and the lowest positive rate would be 5%.

10.2.4 Activity 4: Applying the Data

Part I – Crafting a Persuasive Argument from Data:

Suggested asynchronous

Consider the table provided in the "Testing-International" file, which provides a variety of data points for 12 selected countries. You are tasked with writing three different news briefs (between 100–200 words each), all of which claim to answer the question: "Is the United States performing enough tests to accurately reflect the prevalence of COVID-19 in the country?" Your three news briefs differ in that:

- One is written by a member of the Trump administration
- One is written by someone working for an anti-Trump organization
- One is written by a journalist for a moderate news organization, who is attempting to present an unbiased analysis of the U.S. testing program

Modification:

- *Students split into groups of 3. Each student selects one side to argue and writes a "full" article (suggested 250-500 words). Students within groups then exchange articles and critique the claims made in their peers' articles.*
- *Students can choose to argue these three sides for another COVID-19 related question of their choosing, for which they may find data themselves. See Activity 6.*

Activity 4 - Sample Student Responses

Part II – Reflection:

Suggested asynchronous

Write a short reflection (~500 words) on the process of answering this question from three perspectives. Some questions you might want to address in your reflection are:

1. What additional data would you need to present a more convincing article in each of these cases?
2. What choices did you have to make when writing the different perspectives?
3. What choices were made by someone else before you even started to write your articles?
4. What was the most challenging point of view to present and why?
5. What other (potentially "better") questions might you be able to answer with this set of data ?

Rubric

10.2.5 Activity 5: Reading Critically

Select another article related to COVID-19 and read it critically. Come up with a list of 5 questions that you would ask the author if you were "grilling" them on their analysis.

10.3 Exploring COVID-19 Through Charts and Graphs

Minimum tech requirements: Pdf file

Optional tech resources used: Zoom for remote learning, access to Google Slides to view graphs and have class make comments, access to YouTube, access to the internet to find more context and interact with graphs as well as final assignment.

Possible Courses: Introduction to Statistics, Data Analysis, Finite Mathematics, Intro to Political Science

Objectives/Purpose:

1. Support students in becoming graphically literate
   (a) Understand the quantities and relationships embedded in the diagrams. Be able to interpret the "story" graphs present.
   (b) Identify ways graphs present information in a misleading fashion (ie. manipulate axis, leave out data, etc),
   (c) Learn to question the message(s) communicated, validity, etc.
   (d) Provide a low stakes environment for students to discuss mathematics and share mathematical thinking.

Student Handout: Supplementary Material for Students

Introduction: Turn on the news, go to your favorite website, or click on social media, and you will be immediately inundated with a wide range of charts. While it is common to believe that such charts and numbers are objective and precise–and consequently convincing–many of these charts present data in ambiguous, incomprehensible, or even deceiving ways. Even well-designed graphs have limitations. Decisions about how to display the data inherently emphasize certain characteristics while hiding others. Moreover, as we develop new and creative ways to encode data, many charts contain so much information, they are challenging to interpret. While graphs can be a powerful way to communicate a range of data relatively quickly, informed citizens must possess the ability to critically interpret them.The point of this module is to support students in developing this understanding. While we have included a specific introductory activity and culminating assignment, in general, we suggest using this module as a daily or weekly activity through which students can develop a range of analytical skills and an orientation to reflect and ask thoughtful questions. It is not designed to be content-heavy, though modifications may be made by the instructor to emphasize the content more if the instructor feels this decision best supports the objectives of their course.

Background: These graphs have been collected from the internet since the outbreak of the Coronavirus in February, 2020. Most come from reputable websites, although a few are from social media.

10.3.1 Activity 1: Introducing Graphs

The goal of this activity is to introduce students to some of the characteristics they might look for when analyzing graphs. In part I, students reflect on features of graphs that have been poorly designed and that intentionally alter what one sees. In parts II and III, they look at a well designed graph, analyzing how decisions about what quantities to include and how they are represented impact what is emphasized and what is hidden. All three of these parts involve the following document (Activity I - Parts 1-3)) with the corresponding slides noted throughout.

Part I – Distorting Perception by Poorly Scaled Axis:

Full class discussion

- Present the MSNBC graph (Slide 2 of Activity I - Parts 1-3) displaying the cases of Coronavirus on various days between January 21 and March 13. Ask students to spend 1-2 minutes looking at the graph and reflecting on what they notice and what they wonder. Since this is the first activity, this prompt might be supplemented with a leading question such as, "What might be a way the design of this graph distorts how you understand the graph?"
- Have students share their conclusions in a whole class discussion. Ideally, students will observe that the x-scale is not consistent. If students identify the inconsistencies along the x-axis, follow up by asking students what effect these inconsistencies have on the way someone might interpret this graph. Illustrate how the poor scaling makes the increase look less dramatic by showing the same data with the x-scale properly drawn (Slide 3 of Activity I - Parts 1-3).
- Now, provide another example of how scales can be misrepresented, by showing the Argentine graph (Slide 4 of Activity I - Parts 1-3) of tests per million and ask students if they see a similar error. Again, show the graph of the same data (Slide 5 of Activity I - Parts 1-3) with the y-axis drawn proportionally to highlight the difference.

Part II – Graphical Choices: What is Emphasized? What is Hidden?

Think/pair/share; Zoom breakout discussion rooms

Preface this part of the activity by noting that the previous graphs contained design choices that intentionally altered how the data is perceived. Other graphs are well designed, but still include design choices that emphasize certain characteristics and hide others.

- Show the graph of Cumulative Coronavirus Cases in different countries (Slide 6 of Activity I - Parts 1-3) and ask students the general questions, "What do you notice? What do you wonder?" Again, since this is the first such activity a more direct question might be, "What choices were made about what data is included and how it is displayed affect what you see?" Have students discuss the graph in groups, either in class or in Zoom breakout rooms.
- Afterwards, different groups will share their ideas to the class with the professor taking notes.

Part III – Video: In the following video (Vox, 2020)(Slide 7 of Activity I - Parts 1-3), , the narrator identifies four design features on the graph. Have students watch it in groups and record these four features and what each feature might emphasize and hide. The video can also be shown to the whole class, but students may find it more useful to watch in groups, so they can rewind and support each other in processing the information.

Summary of 4 points from video:

1. Does not take into consideration testing. A country that tests more will have more identified cases. Japan had fewer reported cases than South Korea early on, but also did not test as much. Now that they have increased testing, they have identified more cases.
    - This hides how many cases there really are as well as the rate of growth.
    - No significant advantage. This does not really emphasize any particular feature.
2. Logarithmic scale. Scale is the same distance apart for multiplicative comparisons, not additive.
    - Emphasizes the multiplicative rate of growth. While points get closer as you move up the graph, they are the same distance for multiplicative factors. For example, if we look at doubling, 100, 200, and 400 are the same distance apart on the graph as 1000, 2000, and 4000 even though the additive difference is 10 times much.
    - Hides differences in cases between countries. For example, the US has hundreds of thousands of cases more than the UK.
3. Does not account for population size
    - Hides cases relative to the country's size. Smaller countries look like they are doing better.
    - Emphasizes growth rate. The rate of increase is still comparable, regardless of country size.
4. Number of days since country reported the first 100 cases
    - Hides when the pandemic hit each country, which countries had more time to prepare than others.
    - Allows reader to compare growth rate between countries relative to how the pandemic has progressed in each country.

10.3.2 Activity 2: Brainstorm

Think/pair/share; Zoom breakout discussion rooms

1. Start class by displaying a graph and ask students two questions:
    (a) "What do you notice?"
    (b) "What do you wonder?"

    These questions provide a format where students display their curiosity, not misunderstandings. As such, the hope is that more students will be inclined to share their thinking and perspective, increasing participation.
2. Charts to display in class are provided in Google Slides Background information about the charts as well as commentary to guide your discussion for each graph is provided on a separate Google Slides. We suggest presenting the graphs in Slides 7-10 together as these were published in the same article and highlight how choices made about the method used to present the data can skew the reader's perception. We also suggest presenting the graphs in Slides 12-14 together as these were published in the same article and highlight different ways of displaying similar data.

3. Provide students a couple of minutes to make sense of the graph on their own and record their thoughts individually. If you would like to increase accountability, you can give each student a recording sheet. If you are doing this activity remotely, you can provide each student a blank slide in a Google Slide where they write a few ideas.
4. Have students share their ideas with the class, while you record them at the front of the room. If you are doing this remotely over Zoom, you can put students into breakout rooms and have them share with each other. You can then have each group present one idea to the whole class.
5. We encourage the instructor to maintain a list of graph characteristics that students identify as unusual, misleading, or confusing throughout the course of the activity (potentially all semester). Ideally, students will begin to see characteristics recurring and grow more adept at identifying aspects of graphs that may alter a consumer's perception of the data. Students will use this list in Activity 4.

10.3.3 Activity 3: Assignment: Race/Ethnicity vs COVID-19

Full class discussion (Slides, Activity 4)

1. Start the class with a discussion of this article (Blow, 2020). According to this article, what is the major problem in identifying the effects of COVID-19 on African-American and/or Hispanic/Latino communities?
2. Do we do a good job of gathering and sharing information on the racial breakdown of positive COVID-19 cases? (Activity 4, slide 1) Why? Why not?
3. Refer to Activity 4, slide 4, what do you notice about this graph? Check out the APM Research Lab website DataViz tab. Do you think that Activity 4 Slide 4, compared to the graphs in DataViz tab, is simple/complicated/eye-catching?
4. Show the graph of death rates by Age and Race (Activity 4, slide 2) and ask students the general questions, "What do you notice? What do you wonder?" "What choices were made about what data is included?" and "How do choices about the way the data is displayed affect what you see?" Have students discuss the graph in groups, either in class or in Zoom breakout rooms.
5. Repeat the previous question (Activity 4, slide 3) for Slide 8.
6. These graphs accompanied an article on a blog by the Brookings Institute. Now ask students to read the article and discuss the following statement "In every age category, Black people are dying from COVID-19 at roughly the same rate as white people more than a decade older."

10.3.4 Activity 4: Assignment

Throughout the semester, students will have identified various characteristics that make graphs challenging to interpret. While these can be organized in many different ways, this assignment considers two broad categories of graphs. The instructor may want to note that these categories are neither exhaustive nor mutually exclusive–some graphs will fall into one category, some will fall into both, and some will fall into neither. For this assignment, students will find one example from each category related to COVID-19. These can be found in articles from various media outlets or social media. Regardless, the source should be cited.

Although this is designed to be a culminating assignment due at the end of the course, it can and should be worked on throughout the semester. The point is to spend time looking at different graphical representations and reflecting on the affordances and limitations of their different features.

Category I – Biased Graphs: These are graphs that include a feature that manipulates how you interpret the data. Examples include graphs with inconsistent scaling, uneven bin sizes, or trendlines fit to data that has no correlation. Another example could be a graph that excludes data that is critical to understanding the full picture. For this part of the assignment, find an example of a graph with such a characteristic and analyze the graph. In the analysis,

1. identify the exact design feature that distorts how one interprets the data and explain how it alters what you see
2. any questions you might have

Category II – Complicated Graphs: These are graphs that possess a characteristic that make it time consuming or challenging to interpret. This might be because there are multiple variables displayed or an unfamiliar design feature such as the use of a logarithmic scale. Again, choose a current graph that you feel falls in this category and analyze it. Include in your analysis the following:

1. why this graph is challenging to interpret.
2. your interpretation of the graph
3. a reflection on the design choices (what do these choices emphasize and what do they obscure?)
4. any questions you might have

For examples of graphs that would work well for this assignment, the instructor may want to consider the following two sites:

- The New York Times posts a weekly chart they post for collective analysis in a section called "What's Going on in this Graph?" (The New York Times, 2021)
- The Economist posts a daily chart (The Economist, 2021).

For a general resource on common attributes among misleading graphs, see this page from Calling Bullshit. (Bergstrom and West, 2019)

10.4 Estimating R_0 from Real Data

Minimum tech requirements: Pdf file

Optional tech resources used: Zoom for remote learning, Excel spreadsheet, access to YouTube, access to the internet to find more context and download data.
(Barton, 2016)
Watch (BBCNews, 2020) and (Fefferman, 2020)
Read (Eisenberg, 2020)

Possible Courses: Introduction to Mathematical Modeling and Introduction to Data Science (virtual or in-person)

Student Handouts: Understanding R_0.

Review: How can we quantify R_0?

Objectives/Purpose: This module shows students how to estimate the basic reproduction number, R_0, from real data. It is best when it follows the Flatten the Curve, Exponential Growth Simulation and the Mathematical Modeling of Epidemics teaching modules.

- Help students understand R_0 : the basic reproduction number and estimating its value from real COVID-19 data
- Collect data about positive COVID-19 cases from early March to July
- Use scatter plots to predict the early behaviour of the infected cases

Provide a low stakes environment for students to discuss mathematics and share mathematical thinking. Students should start this module after mastering exponential growth models and SIR models.

Introduction: In this module, our main goal is to estimate a basic reproduction number, R_0, from real data. We will use a discrete SIR model framework. In this framework, all people are either susceptible to being infected, currently infected, or removed which can mean immune or dead. The model is discrete because changes in population size occur at fixed time steps.

The basic reproduction number, R_0 , represents the average number of secondary cases caused by a single infected case in a susceptible population. So if $R_0 = 3$, then the average infected person causes 3 more infections over time.

The main intuition behind understanding R_0 is that if $R_0 > 1$, then each person causes more than one new infection, and an epidemic will occur. If $R_0 < 1$, then the disease will die out before an epidemic occurs because each infection causes less than one other new infection. They are not replacing themselves before they recover.

10.4.1 Activity 1

Class Discussion: Use the Figure in Activity 1 in Understanding R_0.

- What does $R_0 > 1$ say about the spread of the disease?
- What does $R_0 < 1$ say about the spread of the disease?

In the SIR model, $R_0 = \beta * \delta$, where is the average number of effective contacts (contacts with other people that cause them to become infected) that an infected person makes each day, and is the average length of time, in days, that an infected person remains infectious. So if someone makes 2 effective contacts for 3 days, then they will infect 6 people over the full time of their infection.

R_0 is used in epidemiology to determine whether an emerging infectious disease, one that is not already prevalent in a population, can spread in the population and also to determine how fast it is spreading.

10.4.2 Activity 2

The goal of this exercise is to help students practice scatter plots and to better understand the exponential growth model.

Use the Table in Activity 2 Understanding R_0.

1. Use Excel or Google Sheets to graph a scatter plot with y-axis (Number of Infectives) and x-axis (Dates)
2. Write down your observations about the graph. Do the early data in the pandemic look like they follow an exponential growth curve? Why or Why not? What defines an exponential growth curve?
3. Remembering that an exponential function assumes there is a constant multiplicative growth factor, r, in the number of infected people per day. For example, if there are 100 infected people today, and $r = 1.2$, then there will be 120 infected people tomorrow and 144 infected people the next day. Note that r can be any real number, it does not have to be an integer. Find the growth factor in the number of infected cases between each subsequent day for Franklin City.
4. Alternatively, we can use the first and final day in a sequence to estimate the constant growth rate of infected people. Let's do this next.
   (a) In Franklin City, there were 15 positive COVID-19 cases on April 4. On April 10, the number of cases had increased to 27. From these two values, estimate the exponential growth factor, r (Hint: $I(t) = ((1+r)^t I(0), I(t)$: number of infectives at time t)
   (b) In S. Milwaukee City, there were 14 positive COVID-19 cases on April 4. In a week, the case number increased to 28 on April 10. From these two values, estimate the exponential growth rate, r (Hint: $I(t) = (1+r)^t I(0), I(t)$: number of infectives, COVID-19 patients)
5. Assuming that the number of infected people is estimated by an exponential growth function with the factor r you found in question $4 - b$, what is the (estimated) net increase in the number of infected people in Franklin City on April 5 given that there are 15 infected people on April 4? How does this compare to the actual increase value on April 5?
6. Assuming that the number of infected people is estimated by an exponential growth function with the factor r you found in question $4 - b$, what is the (estimated) net increase in the number of infected people in South Milwaukee on April 5 given that there are 14 infected people on April 4? How does this compare to the actual increase value on April 5?

10.4.3 Activity 3

The goal of this exercise is to help students understand the difficulty in calculating, interpreting, and estimating R_0

In the following exercise, we will discuss how one would go about calculating, interpreting, and estimating R_0. Remember that $R_0 = \beta * \delta$, where is the number of effective contacts that infected person makes per day and is the length of time that an infected person remains infectious.

1. How easily can we measure, the number of effective contacts that an infected person makes with others each day? Remember that an effective contact is a contact that an infected person makes with another person who is not infected which results in them becoming infected.
2. What factors will increase/reduce the value of β?
3. The following link provides some examples of for different diseases: Infection Exposure Questions. If we did not know for a particular disease, what data might we want to collect in order to predict δ?

10.4.4 Activity 4

The goal of this exercise is to show students how to use exponential growth factor to estimate R_0. Pair/Share exercise

1. How can you formulate the number of new infected people by using R0 and I(0) during the first days?
2. After days, you know that people who are initially infected, $I(0)$ are not infected anymore. How can you formulate the net number of infected people at the end of days by using R_0 and $I(0)$?
3. Assume that the number of infected people is estimated by exponential growth during the first days, what's the net increase in the number of infected people on day 1? How do you formulate that using exponential growth factor r and initial number of infected people, $I(0)$?
4. True or False $rI(0) = R_0 I(0) - I(0)$. Explain your answer.
5. Based on your answer 4, estimate a basic reproduction number for Franklin City. Use the exponential growth factor in Activity 2, Question 5. Based on your R_0 estimate, write a statement about whether the disease is expected to spread in Franklin City, WI.
6. Based on your answer 4, estimate a basic reproduction number in S. Milwaukee. Use the exponential growth factor in Activity 2, Question 6. Based on your R_0 estimate, write a statement about whether the disease is expected to spread in S. Milwaukee, WI.

7. Now that you have an estimate for R_0 for Franklin City, WI, please revisit Activity 3, Question1, how would you answer that question? Similarly, you have an estimate for R_0 for S. Milwaukee, WI, how would you answer Activity 3, Question 1 in this case?
8. Based on your answers, can you say that one city is doing better than the other one in terms of reducing effective contacts among its residents?
9. What other measures are used to calculate a potential for epidemic spread? (Hint: You can try searching for "effective reproduction number.") Please include a short description in your answer.

10.5 Race and Ethnicity in COVID-19 Numbers: FL case study

Minimum tech requirements: Google Sheets, Web access of Public Tableau

Optional tech resources used: Zoom for remote learning, Excel spreadsheet, access to YouTube, access to the internet to find more context and download data.
Notes to instructors appear in *italics*.

Resources used: United States Census Data: Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2010 to July 1, 2019, Census (2020) (link to dataset)

Possible Courses: Intro to Modeling, data science, or statistics (virtual or in-person)

Student Handouts: Link to student activities

Review: Is there a relation between Race and COVID-19 cases? Case demographics can provide a glimpse of the relationship between the population distribution by county and the observed cases by county. We shall discuss how to explore maps, manipulate data charts, and analyze visualizations to detect if there is a pattern.

Objectives/Purpose:

1. Understand how to compare cases of COVID-19 infections between counties using per capita measures.
2. Learn how to make conjectures from data visualizations and provide data-driven decisions to support those conclusions.
3. Interact with various publicly available technologies to help draw conclusions about the relationship between counties that have experienced larger rates of COVID-19 infection and counties that have a higher percentage of African Americans or Hispanics.

Note to Instructors: The following three activities are available as individual student handouts at `https://sites.google.com/view/covid-19teachingmodules/covid-19-teaching-modules`

10.5.1 Activity 1 - Exploring Maps

Let us explore the map of the State of Florida with information on COVID-19 infection cases, the percentage of Hispanics and African Americans per county, and population density.

1. Access the FL map in the interactive tableau map.
2. Consider map A, where are the higher cases? What explains those numbers?
3. Map A does not take population into account, but map B looks at cases per 100,000 people. Can you provide an explanation of why some counties have a higher rate? Discuss why these are different than the counties in step 2.
4. During the COVID-19 pandemic, one important recommendation is to maintain social distance. Does population density account for higher cases? Explore map C and address this issue.
5. Looking at the map D, is there any evidence that COVID-19 cases per 100,000 people and Hispanic population distribution are related?
6. Looking at map E, is there evidence that COVID-19 cases per 100,000 people and African American population distribution are related?
7. In map F, we compare the two maps above at once with the COVID-19 cases per 100,000 people. Is there any pattern between those maps?
8. Considering the aggregate of Hispanics and African Americans in map G, can you discuss, using all the maps if necessary, if there is any relation between Race and Ethnicity and Covid-19 numbers in FL?

10.5.2 Activity 2 - Manipulate Data

Let us determine if there is any relation between COVID-19 cases and race or ethnicity in Florida. We will look at the data from August 1, 2020, though this data is publicly available and so this activity can be updated to reflect more current data.

1. Open the data set in Google sheets - FL Covid-19 and Race Data - Aug 01 2020
2. Save yourself a copy (*You could make a few groups and put the group sheets online*)
3. What is this data? Familiarize yourself with the data. You may want to address the following questions:
    a. Is any data obviously missing?
    b. What does each row tell you? What does each column tell you?
    c. Can you draw any conclusions immediately from looking at the spreadsheet?
    d. What kind of questions can you answer using these data points?
4. Your first goal is to compare the number of cases and the percent of the population that is Hispanic by county.
    a. Select Columns F and J by clicking on the column letter. You will need to hold command (on a Mac) or ctrl (on a PC) to select multiple columns at a time.
    b. Click on Insert > Chart.
    c. Analyze your scatterplot:
        i. Does there seem to be a correlation between these two data sets? If there is a correlation, is it positive or negative? Is it weak or strong?
        ii. Are there any outliers? How would the correlation change if you removed any outliers?
        iii. If you think there is a correlation, add a trendline:
            A. Double click on your graph.
            B. In the Chart Editor panel on the right, click on the Customize tab.
            C. Scroll down in this panel and click on the Series section.
            D. Check the box for Trendline.
            E. Check the box for Show R^2.
        iv. What does the R^2 value suggest about the correlation between these two data sets? Does it support or refute your initial interpretation?
    d. Based on a-c, can you conclude that counties with larger Hispanic populations can be expected to have a larger number of COVID-19 cases?
    e. Does this chart and the data used to create it give you a fair picture on how COVID-19 cases relates to Hispanic populations? Hint: does it omit any potentially important factors?

    In a virtual class, you can use `https://jamboard.google.com/` *to create a canvas where all students can share and annotate the graphs they are creating.*
5. On second thought, your chart from step 4 does not take into account the population in each county. You realize that you should instead be looking at cases per capita. Your second goal is thus to compare the number of cases per capita and the percent of the population that is Hispanic by county.
    a. Label column M CasesPer100k.
    b. Each row in column M should contain the number of Cases in the county divided by the Total Population multiplied by 100,000. For example, in cell M2, type **=100000*F2/H2**.
    c. Copy this formula down to row 68 by either clicking on the bottom right square in cell M2 and dragging directly down to row 68, or highlighting all cells in column M from row 2 to row 68 and then typing command+d (for Mac) or Ctrl+d (for PC). select Columns J and M. Click Insert > Chart.
    d. Analyze your scatterplot:
        i. Does there seem to be a correlation between these two data sets? If there is a correlation, is it positive or negative? Is it weak or strong?
        ii. Are there any outliers? How would the correlation change if you removed any outliers? [Try this, if you'd like, by copying the spreadsheet to a new tab, removing the row(s) corresponding to any outlier(s) and creating a new chart.]
        iii. If you think there is a correlation, add a trendline:
            A. Double click on your graph.
            B. In the Chart Editor panel on the right, click on the Customize tab.
            C. Scroll down in this panel and click on the Series section.

        D. Check the box for Trendline.
        E. Check the box for Show R^2.
    iv. What does the R^2 value suggest about the correlation between these two data sets? Does it support or refute your initial interpretation?
  e. Based on a-c, can you conclude that counties with larger Hispanic populations can be expected to have a larger number of COVID-19 cases?
  f. To what extent does your answer to this question based on Cases per 100k differ from your answer to this question when you were conducting your analysis based on total number of cases?

6. Repeat step 5 for CasesPer100k and African American, that is columns K and M.
7. Your fourth goal is to see if there is a correlation between number of cases of COVID-19 and the percentage of the population that is African American OR Hispanic, by county.
   a. Label column N "African American or Hispanic"
   b. Each row in column N should contain the sum of the corresponding numbers in columns J and K. Thus, in cell N2, type =**J2+K2**. Fill down this formula to row 68, as you did in step 5.
   c. Create a new chart by highlighting columns M and N and clicking Insert > Chart. . Note: Make sure the x-axis is African American or Hispanic and Series is CasesPer100k (you can check this by double-clicking your graph and looking in the Chart Editor panel to the right).

10.5.3 Activity 3 - Analysis.

After completing the two previous activities, students should use map visualizations and their data analysis to address the following questions.

1. Is it accurate to use Total Cases to compare counties if you are interested in determining the extent to which the counties are being affected by COVID-19? Why or why not?
2. If the number of cases were independent of population racial or ethnic makeup, what should each of your scatterplots (from Activity 2, steps 5–7) look like? Sketch an example.
3. Using the charts you created in Activity 2, steps 5–7, is there evidence of a relationship between COVID-19 cases per county and racial or ethnic makeup of that county?
4. What do you think is the relationship between population density and COVID-19 rates? cases? What might you look at (in the provided data set) to help you answer this question? Does the data support your hypothesis?

One can introduce multiple linear regression analysis by looking at Cases per capita as the dependent variable and racial and ethnic makeup as the independent variables. For example, one can lead with a question such as: Are there other ways to look at the relationship between race/ethnicity and COVID-19 rates?

10.6 Hypothesis Testing: A quantitative approach to the scientific method

Minimum tech requirements: Pdf file, access to Youtube

Optional tech resources used: Zoom for remote, Edpuzzle or Youtube, Google Sheets, Slide presentation software

Additional Resources:

- Youtube video: "The statistics behind driving while Black" by Dr. Omayra Ortega `https://youtu.be/Tkph3i3snnI` Ortega (2020)
- Examples of surprising math applications, suggestions are cited that can be replaced with others:
  - The physics of tossing fried rice (Ko and Hu, 2020)
  - The spread of Bieber fever (Tweedle and Smith, 2012)
- Youtube videos that give a quick summary of p-values, null hypothesis, and alternative hypothesis
- Web-based statistics app Statkey (StatKey, 2020)
- The Color of Coronavirus dataset: COVID-19 deaths, analyzed by race and ethnicity, from APM Research Lab (APM, 2020)

Possible Courses: introductory statistics, introductory modeling, social sciences, biology, workshops or presentations for high school students

Student Handout: Can be created by removing the comments to instructors, which appear in italcs, and adding desired links.

Objectives/Purpose:

- Introduce students to how math and statistics are used to carry out experiments, beyond just the data analysis stage of the scientific method.
- Introduce students to modeling, which can be further addressed in a follow-up module.
- Introduce students to social justice issues in the context of both criminal justice and public health.

10.6.1 Preparatory work

(Optional) for students with minimal or no statistics background, assign both a short Youtube video or text summary of p-values and short Youtube video or text summary of null and alternative hypotheses to watch before class.

10.6.2 Introduction

The introductory material could be given as a handout or as a slide presentation.

Close your eyes and think “the scientific method”. What springs to mind?

...Probably not math, right? Even I think of white lab coats, test tubes, and, oddly, Sesame Street.

Review: The scientific method can be thought of as a problem-solving approach that includes five basic steps, plus one feedback step:

1. Make an observation.
2. Ask a question.
3. Form a hypothesis, or testable explanation.
4. Make a prediction based on the hypothesis.
5. Test the prediction.
6. Iterate: use the results to make new hypotheses or predictions → back to step 1.

Mathematics, statistics, data science, and computer science happen to be excellent tools for experimentation. It is natural to think of applying these fields in step 5, testing predictions by analyzing data, but mathematics can be used at every step. Some unexpected ways math has been applied to the scientific method include making the best fried rice (Ko, 2020), the spread of Bieber fever (Tweedle 2012), and yes, sending large populations in and out of lockdown in a global pandemic.

There are (arguably) two main ways we use mathematics as applied to any STEM or business field:

Analyzing Data We collect data and try to understand how the world works - can also be used to create statistical models for predictions about what might happen

Modeling We take our assumptions about how the world works, translate these assumptions into math, and use this computer version of reality to predict what might happen. If we combine our model with real data, we can test our predictions and use the results to generate new hypotheses or predictions.

Usually each of these areas informs the other. For example, we use data to inform our assumptions for modeling, and we can use modeling to determine what data is necessary to make good predictions.

10.6.3 Activity 1

Mathematics, statistics, computer science, and other STEM fields have a reputation for being unbiased and objective. Given my definition of the two main applications of mathematics, can you think of any ways that social injustice or implicit bias could interfere with good data analysis or modeling?

Suggested format: Zoom breakout session or class poll with multi-select options prefilled. Suggestions for collecting responses: Mentimeter voting, Zoom polls, Zoom whiteboard annotation. Possible student answers or talking points to extend discussion:

- *Data collection - underrepresented groups in samples used for analysis. A problem especially in drug trials and other clinical studies*
- *Bias in assumptions - if we assume all populations are equally susceptible to infection of COVID-19, for example, we may not appropriately distribute resources or vaccines*

10.6.4 Activity 2

This activity requires first watching "The statistics of driving while Black" by Dr. Omayra Ortega, from time 0:00 to 31:05, on Youtube. This can be done asynchronously in advance, or as a class. Remind students to take notes, since you will be using the examples discussed in the video to complete the activities. The following questions can be asked at the given time points in the video, and the video paused, if watched as a class, or can be given to the students to answer individually while watching.

1. (16:40): In the State vs. Pedro Soto case, 43,000 cars were observed and 2,000 of them were speeding. What proportion of drivers were speeding?
2. (20:39): Remember, our claim was that Black/African American drivers get pulled over more often than other drivers. In other words, that being Black or African-American is a risk factor for getting pulled over by the police. By testing our hypothesis, we have (choose one):
   - Strongly contradicted (effectively DISPROVED) the null hypothesis that being Black or African-American is NOT as risk factor for being pulled over.
   - PROVEN beyond all doubt the alternative hypothesis, that being Black or African-American IS a risk factor for being pulled over
3. (31:05): Use the results discussed in the video to generate either a new hypothesis or a new prediction. This is the hard part of the scientific method!

10.6.5 Activity 3

Suggested format: whole class discussion.

Let's walk through how a quantitative approach to the scientific method relates to the State vs. Pedro Soto case as presented by Dr. Omayra Ortega.

1. Make an observation: 43,000 cars were observed, with 13.5% Black or African-American drivers. 2,000 drivers were speeding. 15% of drivers speeding were Black or African-American. 35% of drivers stopped by police were Black or African-American.
2. Ask a question: How likely is it that this disparity is due to chance?
3. Form a hypothesis, or testable explanation by constructing a hypothesis test.

   Claim: Black or African-American drivers get pulled over more often than other drivers. In other words, being Black or African-American is a risk factor for getting pulled over.

   Hypothesis Test: The null hypothesis is $H_0 : p = 0.15$. In words: Black or African-American drivers get pulled over at a rate you would expect given their proportion of the driving population (is NOT a risk factor).

   The alternative hypothesis is $H_a : p > 0.15$. In words: Black or African-American drivers get pulled over at higher rates than you would expect given their proportion of the driving population (IS a risk factor).
4. Make a prediction based on the hypothesis: This is the alternative hypothesis.
5. Test the prediction: They estimated the proportion of Black and African-American drivers getting pulled over by sampling the population of drivers at one location on the New Jersey Turnpike over set time periods, getting a point estimate of $\hat{p} = .35$ (pronounced "p-hat"). To test how likely it is that this disparity is due to chance, a statistical hypothesis test was performed.

   In the video, Dr. Ortega used a cool method: she essentially created a computer model of reality using the online tool StatKey. This generated a bunch of random samples from an underlying population that has the characteristics of our population of interest (15% proportion Black or African-American), to see how often such a selection would result in a sample with 35% Black/African American. The random sampling mimics the "getting pulled over" process. And we can see, with a p-value ≈ 0, the chance of getting a sample of drivers like this is vanishingly small. So we reject the null hypothesis.

   Another option is to calculate a z-statistic for a proportion.
6. Iterate: use the results to make new hypotheses or predictions: Is there anything we could investigate using THIS data? What could we do to strengthen our hypothesis? Does this "experiment" spark any ideas for new work to be done?

10.7 Mathematical Modeling of Epidemics: From influenza to COVID-19

Minimum tech requirements: Youtube, Google spreadsheet, Google Colaboratory (Part 2 only)

Optional tech resources used: Any type of statistic or programming language could be used for solving the SIR model instead of Colaboratory notebooks.

Additional Resources:

- Data: Influenza in a British boarding school
- Google Colaboratory notebook, available for free in Google drive

Possible Courses: Calculus I (after limits or during derivatives), Differential Equations

This module is suitable for Calculus I, to be introduced between limits and derivatives, or as a bridge from calculus topics to the start of a differential equations course. In a calculus class, I would use the activity from Joanna Wares, "Flattening the curve" at the start of the semester when no knowledge of limits or derivatives is assumed. I would tell the students that we will revisit the ideas when we know more. We would then complete the following activities at an appropriate point mid-semester. Following both activities, I would give students an assignment that calls on them to reflect on the growth in their knowledge of mathematical modeling and tools they have learned to use.

In a differential equations class, I would start the semester with this activity, probably as a precursor to Joanna Ware's "DE Semester Modeling Project" that we would work on throughout the remainder of class.

Student Handout: Can be created by removing the comments to instructors, which appear in italcs, and adding desired links.

Objectives/Purpose:

- Students learn the basics of compartmental modeling in an infectious disease context.
- Students see the connection between discrete and continuous models.
- Students learn about the relationship between parameters and model behavior; in particular, students learn how the same basic model can be applied to different diseases or populations by varying parameters.
- Students are introduced to reading and making small changes to code for simulations in one of the activities.

10.7.1 Preparatory work

The following can be assigned as homework prior to completing the activities in class. A quick recap of the preparatory work during class can begin the activities.

- Read: Any recent news article on infectious disease modeling with math.
- Watch: A short, basic Youtube video on introduction to SIR models. Calcvids has an excellent short video at `https://youtu.be/IwZ9DVxf82s`.
- The introduction can optionally be assigned as homework.

10.7.2 Introduction

In an SIR model, We divide the population into 3 groups:

- Susceptible individuals who can catch the disease
- Infectious individuals who can spread the disease
- Removed individuals who have recovered or otherwise developed immunity to the disease

We denote $S(t)$ as the number of individuals who are susceptible at time t, $I(t)$ as the number of infectious individuals at time t, and $R(t)$ as the number of removed individuals at time t. We make the following assumptions:

- The number of individuals N in the population is large and constant, so for all time t, Total population = Susceptible + Infectious + Removed or $N = S(t) + I(t) + R(t)$. There is no birth, death, emigration, or immigration. Individuals move between the subpopulations only. It is helpful to think of this as a closed community, like a college dorm.
- The population is well mixed. This means we are not considering spatial distribution of people and we can picture them as moving around within the space.
- There is no latent period: once you contract the disease, you are immediately infectious. Is this always true?
- Individuals transition from susceptible to infected at transmission rate β (Greek letter beta). The rate $\beta = c \cdot x$, where c is the likelihood of contacts between individuals, and x is the likelihood the contact results in infection.
- Individuals transition from infectious to removed at rate γ (Greek letter gamma). This rate is inversely proportional to the usual duration of an infection D, so diseases with longer duration of infection have smaller recovery rate. This means that people would remain in the infectious population longer.

Equations of change If $S(t), I(t), R(t)$ are the number of susceptibles, infectious, and removed individuals at time t, then to find the number of individuals in each group at the next time step $t+1$ we have:

- $S(t+1) - S(t) = -S(t)I(t)$, where $S(t+1) - S(t)$ is the change in susceptible in 1 day
- $I(t+1) - I(t) = S(t)I(t) - I(t)$, where $I(t+1) - I(t)$ is the change in infected in 1 day
- $R(t+1) - R(t) = I(t)$, where $R(t+1) - R(t)$ is the change in recovered in 1 day

Reflection questions *These questions can be turned in as part of a preparatory homework assignment, or they can kick off a review and class discussion before beginning the activities.*

1. What are we leaving out in our assumptions? Is it reasonable to assume that all people are equally susceptible to a disease? Explain.
2. What about our assumption regarding a latency period? Name a disease that DOES have a latency period where there is a lag between contact and infectiousness.
3. Notice the signs on the terms in the equations of change. For example, in the equation that describes the change in the susceptible population, the term $\beta \cdot S(t) \cdot I(t)$ is negative. In the equation that describes the change in the infectious population, the term $\beta \cdot S(t) \cdot I(t)$ is positive. Why?

Influenza is the disease modeled in the first Activity as a warm-up to COVID-19 in Activity 2, but these activities can be collapsed into one shorter activity by changing the parameters here to be COVID-19 specific or by only investigating influenza. I think it is worthwhile to both show students different diseases and allow them to "discover" a model of COVID-19 but it is lengthy unless some work is completed out of class.

10.7.3 Activity 1: A discrete math model of influenza

1. Notice that we are making some simplifying assumptions with our SIR model. For example, we are assuming that when you recover from an infection, you are immune to the disease...like chicken pox, for example. Is this true for any disease? Give an example of a disease for which infection does NOT give immunity:
2. We can rearrange our equations of daily change to find the number of people in each group S, I, and R at the next time step given our knowledge of the current time step:

$$\begin{aligned} S(t+1) &= S(t) - \beta S(t)I(t) \\ I(t+1) &= I(t) + \beta S(t)I(t) - \gamma I(t) \\ R(t+1) &= R(t) + \gamma I(t) \end{aligned}$$

   We will use these equations to compute 10 days of a influenza epidemic, with initial populations $S(0) = 760$, $I(0) = 3$, $R(0) = 0$. This is case data from a British boarding school outbreak. For influenza, assume the usual duration of infection is $D = 2$ days and in this setting the transmission rate was $\beta = 0.00218$.

   (a) Find the total number of individuals (N) in the population.
   (b) Compute the recovery rate $\gamma = \frac{1}{D}$.
   (c) Use the equations of change to complete a table showing the change in each subpopulation (S, I, R) over the first 10 days of a influenza epidemic. For example, the number of susceptible people on day one would be calculated as $S(1) = S(0) - \beta S(0)I(0)$ by substituting in known quantities $S(0)$, $I(0)$, and β. *In a remote format, or where students have access to laptops, I walk them through the creation of a spreadsheet and linked graph, or provide them a link to a pre-filled spreadsheet where they can extend existing formulas.*

Time (t)	0	1	2	3	4	5	6	7	8	9	10
S	760										
I	3										
R	0										

   (d) Verify that $S + I + R = N$ on each day – this will help you know if your model is working correctly!
   (e) Use your completed table to sketch a very rough graph of S, I, and R versus t (all on the same axis), or generate a graph generated in your spreadsheet to copy and paste.

10.7.4 Activity 2: Calculus takes us from discrete to continuous

In Activity 1, we used a discrete model in which we are looking at the rate of change in these populations over time steps of one day. However, using calculus, we can take the limit of these discrete rates of change as the time step goes gets infinitely small:

$$\begin{aligned} S'(t) &= \lim_{\Delta t \to 0} \frac{\Delta S}{\Delta t} = \lim_{\Delta t \to 0} \frac{S(t+\Delta t) - S(t)}{\Delta t} \\ I'(t) &= \lim_{\Delta t \to 0} \frac{\Delta I}{\Delta t} = \lim_{\Delta t \to 0} \frac{I(t+\Delta t) - I(t)}{\Delta t} \\ R'(t) &= \lim_{\Delta t \to 0} \frac{\Delta R}{\Delta t} = \lim_{\Delta t \to 0} \frac{R(t+\Delta t) - R(t)}{\Delta t} \end{aligned}$$

These are *derivatives*, and thus our model becomes a system of differential equations below! (Input variable t is not shown on the right-hand side for readability).

$$\begin{aligned} S'(t) &= -\beta SI \\ I'(t) &= \beta SI - \gamma I \\ R'(t) &= \gamma I \end{aligned}$$

1. Based on the formula for $S'(t)$ above (the derivative of $S(t)$), do you expect the population of susceptible individuals to increase or decrease?
2. Based on the formula for $R'(t)$ above, which is the derivative of $R(t)$, do you expect the population of removed individuals to increase or decrease?
3. Notice the connection between the equations: Individuals leave the susceptible group ($-SI$) and join the infectious group ($+SI$). Infectious individuals leave the infectious group ($-I$) and join the removed group as they recover ($+I$).
4. Replace parameters β and γ in the general differential equations SIR model above to write the specific SIR model of influenza in the British boarding school.

10.7.5 Activity 3: Simulating continuous SIR models of influenza and COVID-19

This could be done offline as a group project or in class in a group. This requires that at least 1 student in each group have a Google account to use Colab notebooks to run and edit the Python code for the SIR model. The code necessary to create the Colab file shared with the class is included below the activity.

The solutions to the SIR differential equation system are the population functions $S(t)$, $I(t)$, and $R(t)$ that we would like to see: we want to see how these populations will change over time. We can't, however, solve these equations analytically by hand - we must solve them numerically with a computer. We can use some Python code to solve the differential equations for us.

1. Each group should have at least one student with a Google account who can share a screen, since it is required to use Colab notebooks for Python coding. Go to the Colab notebook and save a copy of the file to your drive, or designate one or more group members for this.
2. Complete the section titled "A continuous model of an influenza epidemic" – this just involves reading through and gaining some familiarity with the code.
3. This could be discussed in groups, with someone designated to report for the group. We will generate a new model specific to COVID-19, first assuming that the population remains the same as in the influenza case: 763 susceptible and 1 infected initially. How do you think the transmission rate will change for COVID-19? Will it be smaller or larger? How do you think =1D, where D is the duration of illness, will change? Will it be smaller or larger?
4. Return to the Colab notebook and follow instructions for the next section titled "A continuous model of COVID-19". Choose parameters that reflect your thinking in part (3).
5. Sketch or copy and paste your resulting plot below.
6. *This could be discussed in groups, with someone designated to report for the group or assigned as a post-class writing assignment.* Reflection questions:
    (a) Do you think your results are reasonable?
    (b) Where might you look to identify parameters based on data? Did you try to look these up?
    (c) Do you think both parameters might be different in different locations?
    (d) What are some limitations in the SIR model? For example, do you think all members of the population are equally susceptible? *This could be an opportunity to talk about structured models in a DEs course.*

The activity could wrap up here or could continue offline as a research project to identify parameters through guided research of articles or even data fitting. Alternatively, you can share actual parameters and results that you have found for a particular location with your students for closure.

To create the Colab notebook for the activity, first create a blank Colab notebook. In Google drive go to New > More > Google Colaboratory. Once created, the Colab notebook code and text blocks can be copied and pasted.

Paste the following into a text block:
A continuous model of an influenza epidemic

Reading the SIR model code:

The "#" symbol indicates explanatory text in the code. For example, "# Total population, N" tells you that the line of code below (‘N=1000‘) gives the total population.
We will only need to change code between the lines indicated: values for ‘N‘, ‘I0‘, ‘R0‘, ‘beta‘, ‘gamma‘, and time interval ‘t‘ must be changed to model different situations, but the system of differential equations remains the same. First, read through the code. Next, we will alter it to model a COVID-19 epidemic.
1. Make sure you can identify:
* ‘N‘
* ‘I0‘
* ‘R0‘
* ‘beta‘
* ‘gamma‘
* ‘t‘: how many days is this epidemic?

2. Can you find where the differential equations appear in the code?
3. Notice the code that generates the plot: what is the significance of the line ‘ax.set_ylim(0,800)‘?

Paste the following into a code block:

```
import numpy as np
from scipy.integrate import odeint
import matplotlib.pyplot as plt
#-------------CHANGES CAN BE MADE BETWEEN THESE LINES------------#
# Total population, N.
N = 763
# Initial number of infected and recovered individuals, I0 and R0.
I0, R0 = 1, 0
# Everyone else, S0, is susceptible to infection initially.
S0 = N - I0 - R0
# Contact rate, beta, and mean recovery rate, gamma, (in 1/days).
beta, gamma = 0.00218, 1./2
# A grid of time points (in days)
t = np.linspace(0, 10, 10)
#-------------CHANGES CAN BE MADE BETWEEN THESE LINES------------#

# The SIR model differential equations.
def deriv(y, t, beta, gamma):
    S, I, R = y
    dSdt = -beta * S * I
    dIdt = beta * S * I  - gamma * I
    dRdt = gamma * I
    return dSdt, dIdt, dRdt

# Initial conditions vector
y0 = S0, I0, R0
# Integrate the SIR equations over the time grid, t.
ret = odeint(deriv, y0, t, args=(beta, gamma))
```

```
S, I, R = ret.T

# Plot the data on three separate curves for S(t), I(t) and R(t)
fig = plt.figure(facecolor='w')
ax = fig.add_subplot(111, facecolor='#dddddd', axisbelow=True)
ax.plot(t, S, 'b', alpha=0.5, lw=2, label='Susceptible')
ax.plot(t, I, 'r', alpha=0.5, lw=2, label='Infected')
ax.plot(t, R, 'g', alpha=0.5, lw=2, label='Recovered')
ax.set_xlabel('Time /days')
ax.set_ylabel('Number')
ax.set_ylim(0,800)
ax.yaxis.set_tick_params(length=0)
ax.xaxis.set_tick_params(length=0)
ax.grid(b=True, which='major', c='w', lw=2, ls='-')
legend = ax.legend()
legend.get_frame().set_alpha(0.5)
for spine in ('top', 'right', 'bottom', 'left'):
    ax.spines[spine].set_visible(False)
plt.show()
```

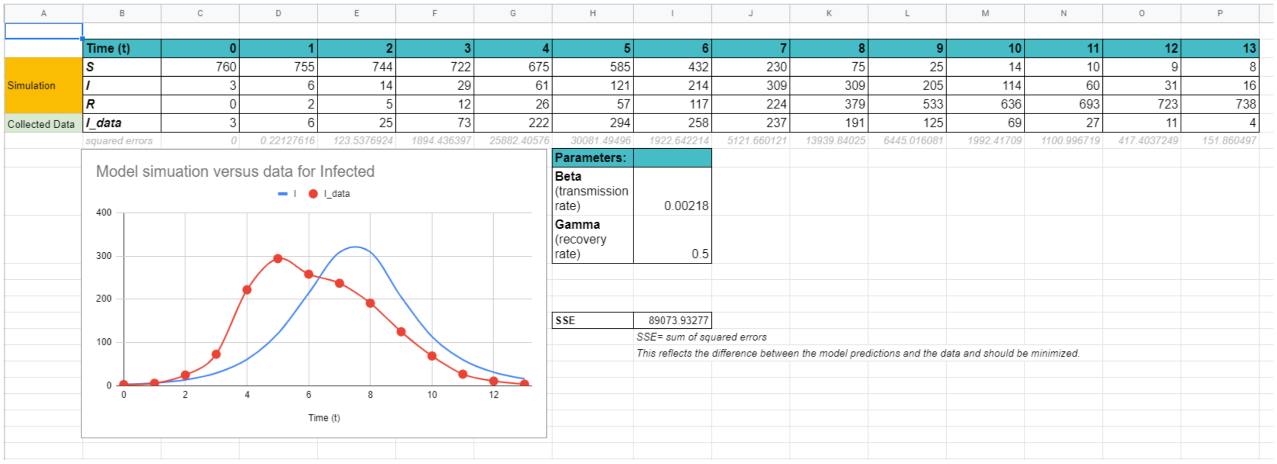
Paste the following into a text block:
A continuous model of COVID-19

Values for ‘N‘, ‘I0‘, ‘R0‘, ‘beta‘, ‘gamma‘, and time interval ‘t‘ must be changed to model a new disease in a new setting. For now, let’s imagine that the population values are all the same and only the transmission rate beta and recovery rate gamma will change.

1. Copy and paste the code for an influenza model below. You can do this by clicking the button above that says "+ Code" to add a code cell. Then right-click the existing code for the influenza model, select "Copy cell", and paste the code into your new code cell below.
2. Change parameters beta and gamma in the code.
3. Press the "Play" button to generate a plot with the new parameters.

10.7.6 Activity 4: Data fitting

For a differential equations course, since this module is introductory, you may want to give your students a first sense of parameter estimation that is very low stakes. This activity uses a Google spreadsheet that contains data for number of infected over the time course of the British boarding school flu outbreak pictured in Figure 1 .



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	Time (t)	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Simulation	***S***	760	755	744	722	675	585	432	230	75	25	14	10	9	8
	I	3	6	14	29	61	121	214	309	309	205	114	60	31	16
	R	0	2	5	12	26	57	117	224	379	533	636	693	723	738
Collected Data	***I_data***	3	6	25	73	222	294	258	237	191	125	69	27	11	4
	squared errors	*0*	*0.22127616*	*123.5376924*	*1894.436397*	*25882.40576*	*30081.49496*	*1922.642214*	*5121.660121*	*13939.84025*	*6445.016081*	*1992.41709*	*1100.996719*	*417.4037249*	*151.860497*

Parameters:	
Beta (transmission rate)	0.00218
Gamma (recovery rate)	0.5

SSE	89073.93277

SSE= sum of squared errors

This reflects the difference between the model predictions and the data and should be minimized.

Figure 1: Screenshot of a simple spreadsheet used to manipulate parameters β and γ to fit data.

An important part of the mathematical modeling process is fitting the model to real data. If the model can generate predictions that are close to real data collected over time, this gives us some comfort that it is a good predictive model. The connection between the data and the model is in the parameters; ideally the parameters will be able to be tweaked until the model output matches the data reasonably closely, and this process is called data fitting or parameter estimation.

Here we will get a quick sense of what can be a complex process by attempting to match the model values for infected individuals I to real data collected on the number of infected individuals at 14 time points over the course of the outbreak, I_{data}.

1. Go to the second tab in the Google "Discrete SIR" worksheet, called "Data Fitting". This can be found at `https://tinyurl.com/y9rc4duy`. Notice that the graph shows only the infected population graph, which is in blue for the model output $I(t)$ and red for the data I_{data}.
2. The goal is to change parameters Beta and Gamma until you get a reasonably good match between the model predictions and the data. This can be evaluated visually and by minimizing the Sum of Squared Errors (SSE) metric on the sheet. The error at each of the 14 time points when data was collected is the difference between the model values and the data values, squared to account for differences in sign. After varying both parameters, write down your best estimates:

Following this activity, groups can share their graphs and parameters, and it may be interesting to compare how different parameter combinations can produce similar results. Discussion could touch on methods of parameter estimation and on the importance of quantifying uncertainty in model predictions and knowing the model limitations given relationships between parameters and output.

10.8 Flatten the Curve

Minimum tech requirements: Google Sheets or spreadsheet software

Optional tech resources used: YouTube

Objectives/Purpose:

- Students learn about modeling in general and think about assumptions of models
- Students learn about exponential and linear models
- Students learn about the S-E-I-R model
- Students explore how changing parameters of the SIR model affect the solutions.
- Students think about how they would change the model or parameters to simulate different interventions or model heterogeneity.

Possible Courses: Calculus 1 or 2, Intro to modeling, data science, or statistics, Advanced modeling or data science in the first week.

Student Handout:
`https://docs.google.com/document/d/12h-RP0uvftnLGTk84DR8_HHOHtKooQk69QuQs6jvM8U/edit?usp=sharing`

Preparatory Work: *I had students complete preview work in advance of this module, and shared a Google doc where they could post questions or reflections from what they had read, which I responded to and we briefly discussed as part of the module.*

Read (short): `https://theconversation.com/how-maths-can-help-us-fight-infectious-diseases-44848` (Laffery and Knight, 2015)

Watch: Introduction to an infectious disease model `https://youtu.be/XWXqXzAYe4E` (Nykamp, 2013)

10.8.1 Pre-activity 1: Disease modeling

Read (long): *`https://science.sciencemag.org/content/347/6227/aaa4339.full`* (Heesterbeek, Anderson, Andreasen, Shwetae, De Angelis, Dye, and et al., 2015)

Read the full article. This is a very thorough overview and I just want you to especially notice a few things as you read:

1. At the top of the article, in the "Structured Abstract", look at the figure called Modeling for Public Health. What do you notice about this process?
2. What are some of the reasons they give for using computational tools to study infectious diseases?
3. In Box 1, what do they mean by "stochastic"?

4. In Box 2, what are 3 major issues in infectious disease that a quantitative approach has been used for?
5. In Box 3, notice how modeling was used to develop treatments for HIV - there is a section on this in your Infinite Powers book and you will see it again in Calculus II when we study differential equations.
6. In Box 4, note some key terms. You've probably been hearing some of these in the news, and we will be looking at SIR models in our lab.
7. My favorite part is Figure 2 - the caption describes all of the ways that a quantitative approach can help us identify unexpected patterns and relationships in infectious disease that are completely counter-intuitive. You can click on the image to enlarge it, and mouse over to read the caption.

10.8.2 Pre-activity 2: Learning about Mathematical Models

Question: Can you name some successful mathematical models? (Here's an example: F = ma.) What makes the model successful?

Two different goals of modeling:

1. Modeling to make accurate predictions
2. Modeling to reveal the true mathematical properties underlying a phenomenon

You can delete from here to pre-activity 2 if you want the activity to be purely COVID-19 Focused. Just make sure to delete it in the student handout also.

Example of a model that achieves goal 1 but fails badly on goal 2

Scientific Question - Can we predict future positions of the planets and the stars?

Ptolemy (Claudius Ptolemaeus AD 83 - c.168) was a Greek astronomer who believed the earth was the center of the universe. He created a mathematical model that could predict future positions of planets and the sun very accurately for his time - success.

Back in his time, people believed that the earth was the center of the universe and that all the planets and the sun and stars revolved around the earth. They also believed that godly motion was perfectly circular motion and that all movement of the planets and stars must be on perfectly circular trajectories.

Ptolemy created a model of the universe that used perfect circular motion but he had to have circles going around other circles to make it work. His predictions about the locations of the planets and the sun, as they traveled in time, were pretty accurate, but the underlying model did not uncover the real dynamics of planetary motion.

Watch the first 4 minutes: `https://www.youtube.com/watch?v=ZGr1nHdzLyk` (Farina, 2018)

Watch Ptolemy's model in action: `https://www.youtube.com/watch?v=2HS3-UgqMmI` (Vanh, 2019)

So, this model was successful at making predictions and used for a long time until Copernicus came around and developed a model that centered the sun in the "solar system":

- Was much simpler
- Made better predictions about planetary motion
- Gave scientists a new platform for testing new hypotheses to make predictions
- Revealed the true nature of solar system

Let's do some activities to see how researchers use mathematical and statistical models to make predictions and understand the underlying phenomena that we perceive. The best models achieve both goals 1 and 2 above.

10.8.3 Activity 1

Question: How do we predict how many people will be infected with SARS-CoV-2 in the USA at any time? The CDC is reporting data for daily counts of people with positive test results: `https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html` (CDC, 2020)

I have downloaded this and put it in a google sheet for you, but you could download the latest data yourself and use that: `https://docs.google.com/spreadsheets/d/139tQ_0ou7Oc8wGiR7esAE8sIKpW8bWD9EmM2FPjq2-M/edit?usp=sharing`

Let's create a first model to say how many people are infected at any time, $I(t)$. We are going to find a function $I(t)$, where t is time in days (independent variable) and $I(t)$ is the total number of people that have been infected time t (I is the dependent variable).

1. Open data set in google sheets - COVID Data from CDC - (Data Link)
2. Save yourself a copy *(You could make a few groups and put the group sheets online)*
3. What is this data? Make a chart of New Cases.
4. Make a new row of data that represents Total Cases (=sum(B2:B2) let's you dynamically add when you drag the box to the right. Figure out how it works. This goes in B3. Label A3 as Total Cases).
5. Make a second separate chart of that row (Total Cases). Make sure to label the vertical axis.
6. Limit the Total Cases chart to just A3 through Z3 data. (Insert → Chart) right menu (Setup) Data range A3:Z3
7. Make a copy of the chart and limit that new chart to data Z3 to DD3.
8. How do the 2 charts differ in shape? What do you think is happening in this data?

Now let's find the function that looks like it would intersect the most points in our partial data sets, or that would come closest to intersecting a lot of them. In statistics, this is called finding the least squares best fit or a regression model. We are going to ASSUME that the model of the phenomenon is a particular type of function and find the parameters for that function that are closest to our data. In doing so, we are trying to achieve goal 1 above. We are going to find a function that our data seems to come from, and then use the function to make predictions at times in the future that we don't have data about yet.

Using only the Total Cases data, not the daily counts (so have this on a separate plot):

1. Let's start with the data from Z to DD. This would relate to starting 26 days into the pandemic. What type of function does this look like if we restrict our data to later data?
2. Now let's find the best fit function to use as our model.
3. Click the 3 dots in the upper right corner and select "Edit Chart".
4. Go to Customize > Series > Trendline
5. Under "Type" choose a type and see if it looks like a close fit or choose the one you think will be good.
6. Change Label to "Use Equation"
7. Repeat this for the chart where you use data A to Z.

(The A to Z data is exponential and the Z to DD data is close to linear).

Let's investigate the second model (for data A to Z).

Q1: What function are you going to use to predict the total number of infected people at any time t, $I(t)$?

Q2: Using your model, how many people would be infected after 10 days? Is that close to the real data point?

Q3: Again using your model how many people would be infected after 100 days? Does this make sense?

Q4: Edit the chart of New Cases to include all of the data again. Is the model still a good one if we consider all of that data?

Q5: What constant factor r is the total number of cases increasing by each day? Explain this in terms of the model. If you do this again or first with the linear model, ask how many new cases are being added each day. Compare this with exponential growth where the new cases are a multiple of the current cases.

You could do the same thing for the linear model. If so, ask how the growth is different from the exponential model. Could also ask here if either of these models is a good fit for the whole data set. Come back to the point that these models are good short term models for goal 1 but not for goal 2.

Take away: Infections grow exponentially early on, but slow down after some time. We can use an exponential model to predict infections for a short period of time but need a more interesting and dynamic model for predictions about the longer term future.

10.8.4 Activity 2

Now we are going to try to better understand the Flatten the Curve models that were in the media early on in the pandemic. The idea behind "Flatten the Curve" was that we could reduce infection and death rates, by slowing down the rate of new infections (Figure 2). Unlike the models above which aim at achieving goal 1, the following model tries to get at the underlying dynamics of the pandemic (goal 2).

To do so, let's watch the first 9 minutes of `https://www.youtube.com/watch?v=Qrp40ck3WpI` (Bazett, 2020) to get a foundational understanding of the standard SIR model. *You could skip this.*

So we see, all people are considered either to be in one of three states: 1) $S(t)$ – susceptible to being infected at time t, 2) $I(t)$ – infectious at time t, 3) Recovered (presumably immune) at time t (see Figure 3).

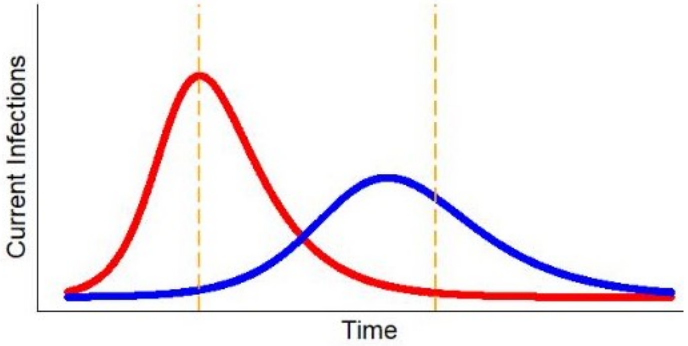


Figure 2: Flatten the Curve. Different trajectories of $I(t)$ in the SIR model that have different parameter choices.

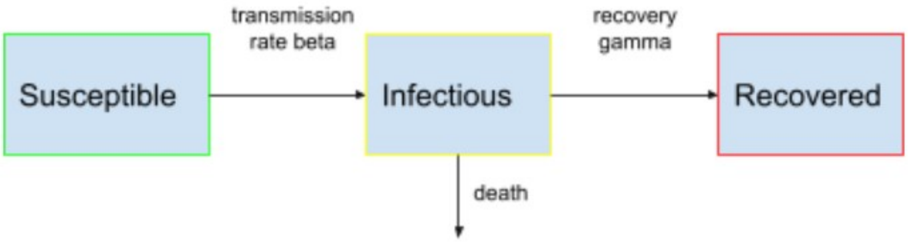


Figure 3: SIR Model.

There are really 3 basic parameters here too, 1) the transmission rate (which has to do with the number of contacts an infected person makes with susceptible people and the likelihood of any contact causing the disease to be transmitted), 2) the recovery rate (which has to do with how long they are infectious), and 3) the amount of people that die and leave the system.

Q1: The SIR model is being used to predict how many people will get infected with SARS-CoV-2. Can we discuss what assumptions people are making when they use this model? What is being left out?

Q2: When we talk about "flatten THE curve" which curve are we talking about? The S, I, or R curve (Figure 4)?

Q3: Let's try to better understand why we would want to flatten the curve and what the results of flattening the curve are.

Spend some time going through the website tutorial linked below and try to figure out why one would want to flatten the curve.

Scroll down to this image first (Figure 5) and read about the reproductive number R_0 (the basic reproductive number is $R_0 = \frac{\beta}{\gamma}$ so it depends on both transmission and recovery rate).

Note that the big takeaway is $R_0 < 1$ is good and $R_0 > 1$ is bad.

Then scroll through the section "The Next Few Months" and try running some of the simulations. (see Salathé and Case, 2021)

Q4: Let's think about how interventions for COVID-19 could be added to the SIR model.

1. Remember, we assume that the population is homogeneous (everyone has about the same susceptibility, etc.). Is this true? What would change if not?
2. What would change if we implemented social distancing?
3. What would change if we found a drug that helped patients recover more quickly?
4. What would change if we vaccinated?
5. How would the parameters change if you work masks? What if some people wore masks and others didn't?
6. How could you model different parameters for different groups?

Q5: Preliminary data suggests that COVID-19 disproportionately impacts minorities and low-income households. What are some ways that you could use the SIR model to think about this and ask and answer questions about disproportionate effects on certain groups?

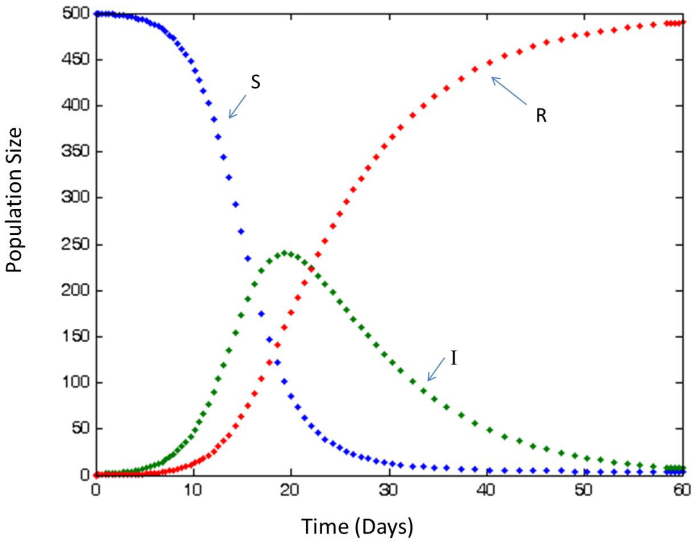


Figure 4: SIR Model Trajectories.

Figure 5: Symbol to find.

10.8.5 Future Reading

- Harrod, 2020: A nice video going over what we just talked about from a data point of view.
- Wares and Krehbiel, 2020: A link to an article Dr. Joanna Wares of University of Richmond wrote in 2020 about herd immunity that uses the SIR model and talks about flattening the curve.

10.9 Differential Equations Semester-Long Project

Minimum tech requirements: Could do this purely theoretically with no tech but is better done with any kind of computational tool, like R Studio or MATLAB

Optional tech resources used: Any type of statistic or programming language could be used for simulating models.

Objectives/Purpose: Students learn all that goes into developing a differential equations model about something real Students learn how to ask and answer questions using the model Understand how DEs are a powerful tool for understanding real world phenomena

Possible Courses: This is written for Differential Equations in particular but could be easily modified for other courses.

Student Handout: Differential Equations Project.

10.9.1 Assignment

So much about our world is constantly changing (in time and in terms of other variables). Differential equations are the language of change. Creating your own differential equations model will help you learn how to characterize relationships in the world, and, in particular, how to capture the essence of those relationships. You can use your model to answer questions or make predictions about the future. During this semester, you will also learn how to present your ideas clearly by creating presentations about your model and results.

Preparatory Work: Watch the video `https://www.youtube.com/watch?v=NKMHhm2Zbkw` (Crawford, 2020). Notice how he describes the assumptions, model, and the questions he is trying to answer.

In groups of two *(or more)*, students will create a Differential Equations Model. You must choose your idea and start formulating your idea in the first month of classes. Each group will present their ideas to the class at two different times in the semester. The first presentation will occur when you are part way through the project and is called the "pitch presentation." In the pitch presentation, you want to toss your idea around with the class and get input to adjust your model. The second presentation is the final presentation at the end of the semester.

Directions: Find a topic that interests you about COVID-19. It would be good if you found a data source, or an article of interest that inspires you to ask a new question. The idea is to build a *Differential Equations (DE) Model* that will help you answer this question. Extending a previously created model counts as creating a model, but you need to add something new. The something new might be the data set and finding parameters that match the data set well.

Remember, there are many types of mathematical and statistical modeling techniques that are useful, but this idea should be able to be addressed with a DE model, so you want to have some quantity changing with respect to some other quantity (likely time).

Create a presentation for the class describing how modeling could improve our understanding of your chosen phenomenon. The presentations can be video or slides. The pitch presentation should include:

10.9.2 Pitch Presentation

1. A description of the your question and why it is important or interesting
2. A list of the variables included in the model (distinguishing between dependent/independent variables)
3. A list of parameters. (Variables and parameters are different.)
4. A description of the assumptions you are thinking about including in the model (why you included or left out pieces)
5. A compartmental diagram when appropriate
6. Questions you will try to answer using the model
7. What you would expect the model to show (including a hypothetical graph of the results where applicable)
8. Limitations of the model based on the assumptions you made and how they could be improved for future work
9. Citations

Note that the pitch presentation does not include equation formulation. That is in the final presentation. You should take feedback from the class to help you develop your model for the final project.

10.9.3 Final Presentation

Along with all of the items from the pitch presentation add the following:

1. The main point of this is to tell a story about your project. One way to do this is to think about how you would explain it to someone on a job interview. The presentation should be centered around your story.
2. Create your model with equations and describe them. You can extend a previous model if you like.
3. Explain your initial conditions.
4. Draw a diagram of your model, probably compartmental
5. Simulate your model and answer your question(s).
6. Extra credit will be given for real attempts at using real data and parameterizing the model well.

Suggested topics if your students need help finding them:

1. *Model Social Distancing and see how infection numbers change*
2. *Create more compartments and separate groups by income level (could just use two income levels or maybe 3) and make different assumptions about the parameters for the different groups. Model how different interventions affect different groups.*
3. *Investigate the racial disparities seen in COVID-19 infection rates.*
4. *Find R0 after you include various interventions and talk about how the different parameters are tied to the different interventions and outcomes.*
5. *Include a quarantine state.*
6. *Investigate herd immunity and overshoot.*
7. *Use either SIR or SEIR on a particular data set from a particular location and ask questions about interventions and spread of COVID-19.*

8. *How do you model the starting and stopping of interventions? Ask questions about how starting and stopping interventions change the infection rates.*
9. *Model the introduction of a vaccine.*

Other Tips for Instructors:

1. *It is helpful to have a few classes throughout the semester where the class gets together in their teams and you interact with each group and help them hash out their ideas, especially at the beginning of the semester. Frequently, they have ideas that are good research questions but do not lend themselves to a good DE formulation and it is important to find this out early.*
2. *Sometimes students find a complicated DE model that they are interested in, like something that has stochastic DEs or Delay DEs. In this case, you could allow them to do a thorough investigation of the model and explanation. They could possibly even try to reproduce results that have already been determined. This usually involves them having to understand material outside the scope of the class and is a project in itself.*

References

APM (2020). Color of Coronavirus: COVID-19 Deaths, Analyzed by Race and Ethnicity. https://www.apmresearchlab.org/covid/deaths-by-race. 68

Barton, J. T. (2016). *Models for life: An introduction to discrete mathematical modeling with Microsoft Office Excel*. John Wiley & Sons. 64

Bazett, T. (2020). The mathematics of epidemics. https://www.youtube.com/watch?v=Qrp40ck3WpI. 78

BBCNews (2020). What's the r number and why does it matter? https://www.youtube.com/watch?v=iaOm-eF9p-8. Accessed 2020-06-26. 64

Bergstrom, C. T. and J. D. West (2017-2019). Calling bullshit: Tools and tricks. https://www.callingbullshit.org/tools.html. Accessed 2020-06-26. 64

Blow, C. M. (2020). Focus the covid-19 fight in black cities. *The New York Times*. 63

CDC (2020). Covid data tracker. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html. 77

Census, U. (2020). Annual county resident population estimates by age, sex, race, and hispanic origin: april 1, 2010 to july 1, 2019 (cc-est2019-alldata). https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/counties/asrh/cc-est2019-alldata.csv. 66

Crawford, T. (2020). Oxford mathematician explains sir disease model for covid-19. https://www.youtube.com/watch?v=NKMHhm2Zbkw&t=593s. 81

Eisenberg, J. (2020). How scientists quantify the intensity of an outbreak. https://labblog.uofmhealth.org/rounds/how-scientists-quantify-intensity-of-an-outbreak-like-covid-19. Accessed 2020-06-26. 64

Farina, D. (2018). History of astronomy part 3: Copernicus and heliocentrism. https://www.youtube.com/watch?v=ZGr1nHdzLyk. 77

Fefferman (2020). The role of applied math in real-time pandemic response: How basic disease models work. https://www.youtube.com/watch?v=Ewuo_2pzNNw. Accessed 2020-06-26. 64

Harrod, J. (2020). How does outbreak prediction work?. https://www.youtube.com/watch?v=jj0FMoFX4TM. 80

Heesterbeek, H., R. Anderson, V. Andreasen, B. Shwetae, D. De Angelis, C. Dye, and et al. (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science 347*(6227). 76

Johns Hopkins University and Medicine (2020-2021). How does testing in the u.s. compare to other countries? https://coronavirus.jhu.edu/testing/international-comparison. Accessed 2020-06-26. 58

Ko, H. and D. L. Hu (2020). The physics of tossing fried rice. *Journal of the Royal Society Interface 17*(163), 20190622. 68

Laffery, E. and G. Knight (2015). How maths can help us fight infectious diseases. *The Conversation*. 76

Nykamp, D. (2013). Introduction to an infectious disease mode. https://youtu.be/XWXqXzAYe4E. 76

Ortega, O. (2020). Inspiring talks-ortega. 68

Ritchie, H., E. Ortiz-Ospina, D. Beltekian, E. Mathieu, J. Hasell, B. Macdonald, C. Giattino, C. Appel, L. Rodés-Guirao, and M. Roser (2020). Coronavirus (covid-19) testing. https://ourworldindata.org/coronavirus-testing. Accessed 2020-06-26. 59

Salathé, M. and N. Case (2021). What happens next? https://ncase.me/covid-19/?fbclid=IwAR2_atnrmhsWIqb5odjmN8zEAA4I5wCm-jg6vOe7xPHXel11NiufOWOYptI. 79

StatKey (2020). Randomization Test for a Proportion. https://www.lock5stat.com/StatKey/randomization_1_cat/randomization_1_cat.html. 68

The Economist (2021). Graphic detail. https://www.economist.com/graphic-detail. 64

The New York Times (2021). What's going on in this graph. https://www.nytimes.com/column/whats-going-on-in-this-graph. 64

Tweedle, V. and R. J. Smith (2012). A mathematical model of bieber fever: The most infectious disease of our time. *Understanding the dynamics of emerging and re-emerging infectious diseases using mathematical models. Transworld Research Network*. 68

Vanh, D. (2019). Ptolemy's model of the universe. https://www.youtube.com/watch?v=2HS3-UgqMmI. 77

Vazquez, M. (2020). Trump now says he wasn't kidding when he told officials to slow down coronavirus testing, contradicting staff. *CNN*. 54

Vox (2020). How coronavirus charts can mislead us. https://www.youtube.com/watch?v=O-3Mlj3MQ_Q. Youtube. 62

Wares, J. and S. Krehbiel (2020). Herd immunity won't solve america's covid-19 problem. *The Conversation*. 80

Wares, J., M. Torres, Z. Teymuroglu, G. Stadnyk, C. Hawthorne, and E. C. Balreira (2020). COVID-19 teaching modules. https://sites.google.com/view/covid-19teachingmodules/covid-19-teaching-modules. 49