Exploring Exponential Models

Students investigate exponential relationships in data about Covid spread, using an inquiry-based model, involving hypothesizing, experimental and computational modeling, and sense-making. They are introduced to table transformations and inverse functions, which are used to fit exponential models onto nonlinear data.

Lesson Goals	Students will be able to… • Read and interpret real-world data, presented in a scatter plot • Recognize exponential growth in tables and graphs • Model exponential relationships using functions
Student-facing Lesson Goals	• Let's use Pyret to model exponential relationships in data • Let's use Pyret to filter a dataset into subsets
Prerequisites	• Simple Data Types • Contracts: Making Tables and Displays • Exploring Linear Models • Row and Column Lookups • Exploring Quadratic Models
Materials	• PDF of all Handouts and Page • Covid Spread Starter File • Modeling Covid Spread (Desmos) • Lesson Slides • Printable Lesson Plan (a PDF of this web page)
Supplemental Materials	• Additional Printable Pages for Scaffolding and Practice

Glossary

asymptote :: A straight line to which a curve gets closer and closer - but never touches - as one of the variables approaches infinity.

exponential decay :: A sequence in which each number is multiplied by a constant amount - less than one - to produce the next, causing the sequence to decay rapidly at first and then slow to smaller and smaller reductions

exponential growth :: A sequence in which each number is multiplied by a constant amount - greater than one - to produce the next, causing the sequence to grow slowly at first and then switch to rapidly-accelerating increases

exponential relationship :: A mathematical relation between two variables x and y, in which the dominant term is raised to the power of x, and the y-values grow by a constant factor over equal intervals in the x-values. When graphed, an exponential relationship appears as a 'hockey stick' curve (sloping up or down). Exponential functions occur widely in the natural and social sciences, as in a self-reproducing population or a fund accruing compound interest.

growth factor :: the amount each term in an exponential sequence is multiplied by to get the next term (either 1 plus the growth rate or 1 minus the decay rate)

When and Why do we Filter Data?   *45 minutes*

Overview

Students discuss an example of Simpson's Paradox, which motivates splitting a dataset into grouped samples using filters. They then explore the Covid Spread Starter File, and discover another motivation for filtering: some scatter plots show *multiple* correlations, instead of just one. Finally, they learn how to filter a dataset and apply that knowledge to filtering the covid dataset into samples grouped by state.

Launch

A college is looking at enrollment and housing data for a sample of students who've decided what their major will be, vs. those who are undecided:

	# On Campus	# Off Campus	% On Campus
Undecided	120	80	120/200 = 60%
Decided	80	100	80/180 = 44%

> - According to the table, how many Undecided Majors live *off*-campus?
>   - *80*
> - How many Decided Majors live *on*-campus?
>   - *80*
> - Who is more likely to live on campus: Decided or Undecided Majors? Let students discuss the last question!

Optional: If your students need printed copies of this table and the accompanying questions, you can use Campus Housing Data.

It *looks* like the two variables are significantly related: undecided majors are more likely to live on campus than decided ones!

But there's a third variable hiding in the background: freshmen college students are far more likely to live on campus, *and* they generally have not picked a major. **We need to take the *year* into account.**

Freshmen	# On Campus	# Off Campus	% On Campus	Non-Freshmen	# On Campus	# Off Campus	% On Campus
Undecided	100	20	100/120 = 83%	Undecided	20	60	20/80 = 25%
Decided	50	10	50/60 = 83%	Decided	30	90	30/120 = 25%

Because **year at school** is an important 3rd variable, we should first filter by that variable. And it turns out that for both Freshmen and Non-Freshmen, there is *no correlation* between between deciding on a major and living on- or off-campus.

How is this possible? When a third variable is lurking in the data, it can play tricks by obscuring relationships between two other variables - or by creating the appearance of a relationship where none exists!

Simpson's Paradox: visible trends in sub-groups *disappear* or even *reverse* when the groups are combined.

Normally we think that "more data means more power!", and that the more data we include in our sample the more clearly we'll see any potential relationships. But in certain circumstances - like in our Covid dataset - the correlations in our sub-groups cancel each other out when we put the groups together. This is called Simpson's Paradox. (Learn more about Simpson's Paradox at Wikipedia.)
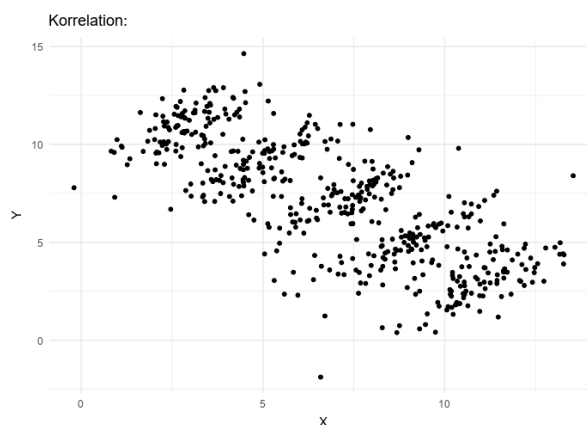
"More data" isn't always better. Simpson's Paradox shows us that being able to filter the data into subsets is sometimes the only way to see what's really going on.

```
In this lesson, we'll explore some real data that's better analyzed
__in pieces...__
```

Investigate

Starting in 2019, Covid spread across the globe. Most of us heard terms like "flatten the curve" and "infection rate" in videos and on the news.

Even in 2020, very few people could have predicted the impact Covid would have on the world. But Data Scientists who were looking at the data knew differently. Let's take a look at some of that data!

- Open the Covid Spread Starter File. This data looks at the Covid infection rates for New England states from summer 2020 until the end of the year.
- From the File menu, select "Save a Copy", and click "Run."
- Working in pairs or small groups, complete Exploring the Covid Dataset.

Note: This dataset is available for all 50 states (and Washington, D.C!), but for pedagogical purposes we've written the starter file to pull only data from New England.
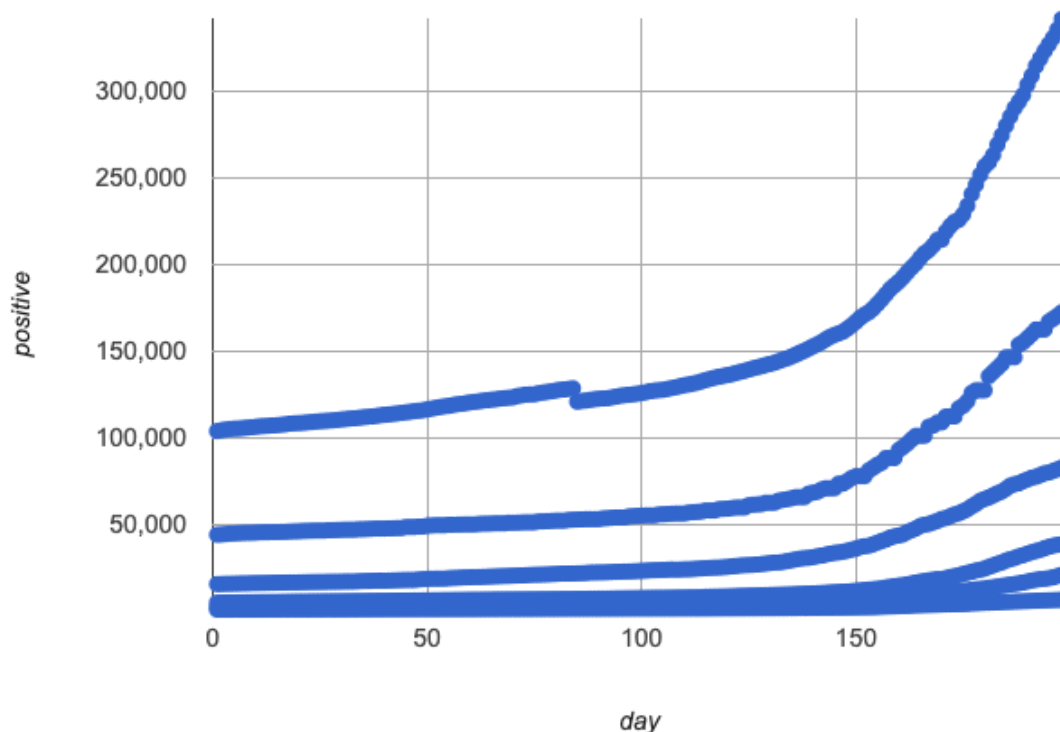
Discuss in groups or pairs, and prepare to share out to the class:

- Based on the look of the scatter plot you just made, do you think there's a strong relationship here?

The scatterplot you made doesn't look much like a scatterplot all! It looks like someone took a marker and drew in five different curvy lines. While it's clear that there are strong patterns in the data, these patterns are so distinct from one another that there isn't really one, single model that fits them all. Each relationship appears very strong, almost as if there is *more than one model* here.

Review student answers to confirm that students have made a number of observations:

- There is *more than one* relationship in this dataset
- Every relationship seems extremely strong
- Most/all relationships appear nonlinear

With all these clear, tight curves, we might think this would be a dataset with a very strong relationship. Unfortunately, that's not what we see when we group all the data together!

Datasets like these are very difficult to model all at once, because there will always be lots of points that are far from any single function. But it's not that there's *no relationship* between the x- and y-variables. Instead, we have several sub-groups each with their own *very strong relationships.*

- We need to break the Covid data up into *grouped samples*, so that all of the data for Rhode Island is in one table, all of the data for Maine is in another, etc.
- How is a grouped sample different from a random sample?
  - *A grouped sample is a non-random subset chosen from a larger set. Grouped samples are non-random by design!*

Working in pairs or small groups, complete [Filtering by State](#).

The `filter` function consumes a Table and **a helper function!** The helper function is used on every Row of the Table, producing true or false. The `filter` function takes all the Rows for which the helper produced true, and combines them all into a new table.

Optional: While filtering is introduced in this lesson, the primary goal is for students to explore exponential functions. If your students need more practice with filtering - or wish to filter their own datasets - we recommend checking out the [Filtering and Building](#) lesson.

Common Misconceptions

It's extremely common for students to think that filtering a table *changes the original table*. This is NOT how it works in Pyret! Instead, the `filter` function always produces a *new* table, containing only the Rows for which the supplied function evaluates to `true`.

Synthesize

- In what other situations would it be useful to filter a dataset?

- Can you think of other examples where Simpson's Paradox might arise?

  - *When comparing one country's schools to another's, a researcher finds that students living in poverty in country A outperform students living in poverty in country B. They also find that the wealthy students in A outperform their wealthy peers in B. In fact, for every income level, country A outperforms country B! But if country B has less child poverty overall, it will still outperform A.*

  - *Another, thoroughly-explained example involving soft drinks can be found [on this web page](#).*

Looking for Patterns   *45 minutes*

Overview

Students explore their newly-filtered `MA-table` dataset, trying to fit different kinds of models to it. This section makes heavy use of interactive slider activities we've built in Desmos to support open-ended experimentation.

Launch

- Open the [Covid Spread Starter File](#).
- Make a scatter-plot showing the Covid infection rate for Massachusetts.
- What kind of model do you think would fit this best?

Why just New England, starting from June 9th?!?

We have *artificially constrained this dataset*, showing only the data from June 9th to December 26th, 2020. We've made this choice in order to showcase the most purely-exponential behavior of the infection curve, for the sake of this lessons' math learning goals.

For students who are farther along, we recommend showing them *all* the data through 2020, starting in January rather than June. The first portion of the infection curve shows a gradual, linear growth pattern before exploding in the Fall of 2020. This is *polynomial* behavior, where a linear term dominates when the exponential term is small.

Based on the strength of your students, we encourage you to choose the data that best fits your learning goals. You may also wish to return to full dataset later on, once students are comfortable with polynomial functions.

To use all available data, open the [Covid Spread Starter File](#) and change the source sheet on line 7 from `"New England"` to `"All"`

Investigate

Linear models capture *straight-line relationships*, where one quantity varies proportionally based on another. In linear models, we expect the response variable to grow by equal amounts over equal intervals in the explanatory variable.

Are linear models a good fit for this data? Why or why not?

Have students share their resulting models. Which one fits best?

If we make the line go from the start to the peak of the curve, almost all of the points bulge out below our line of best fit. If we make the line hit the bottom of the curve, all the points fall above it. Splitting the difference (orange line) is better than both of those options, and we might even get a pretty good R^2! But ultimately, straight-line, linear models just don't behave like this curve, and we'll never get the *best-possible fit* with them. **It's growing too fast to be fit with a linear model that grows at a constant rate!**
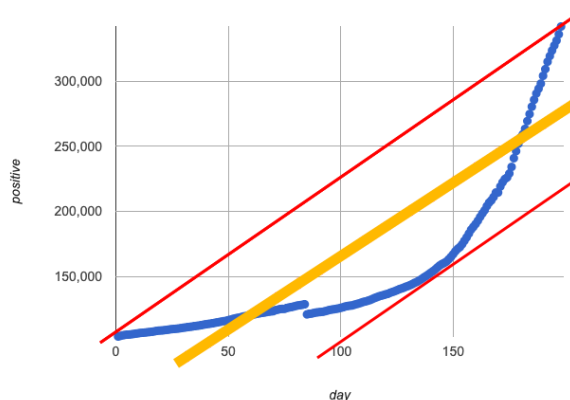
- Are quadratic models a good fit for this data? Why or why not?

Have students share their resulting models. Which one fits best?

Quadratic models capture *parabolic relationships*, where one quantity varies based on the square of another. In quadratic models, we expect the response variable to grow by differing amounts over equal intervals in the explanatory variable.

Quadratic models change their rate of growth over time, which definitely makes them a better fit for this data than linear ones. It's very likely we could find a quadratic model with a pretty good R^2 value! But this data starts out almost flat and then suddenly takes off like a rocket - quadratic models just don't have that kind of explosive growth, so our model will never be as good as it *could* be.
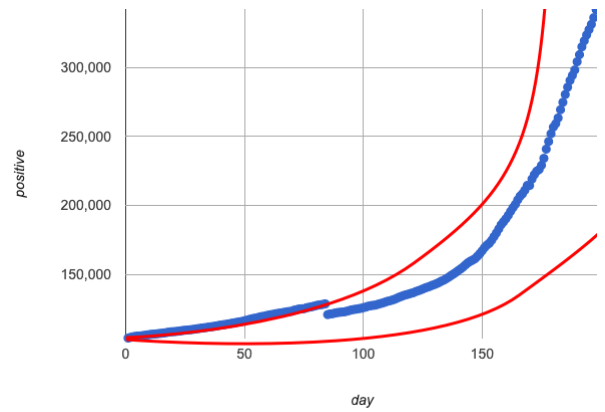
Synthesize

- Do you think the data for MA shows a linear relationship? Why or why not?
- Do you think this data shows a quadratic relationship? Why or why not?
- Do you think this data shows some other kind of relationship? Why or why not?

Exponential Functions   *55 minutes*

Overview

Having identified that the Covid scatter plot is neither linear nor quadratic, students learn about characteristics of exponential functions in tabular, graphical, and function notation form.

Launch

Let's review what we know about the behavior of the models we've seen so far:

Remember that linear functions grow by *fixed intervals,* so the rate of change is *constant*. In the table shown here, each time the x-value increases by 1, we see that the y-value increases by 2. This is true for any set of equal-sized intervals: a line needs to slope up or down at a constant rate in order to be a straight line!

x_1	y_1
1	5
2	7
3	9
4	11

+2
+2
+2

If the "growth" is constant, the relationship is linear.
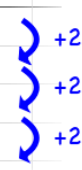
Quadratic functions grow by intervals that *increase by fixed amounts!* In the table to the right, the blue arrows show a differently-sized jump between identical intervals, meaning *the function is definitely not linear!* However, if we take a look at the *difference between those differences* (shown in red), we're back to constant growth!

x_1	y_1
1	5
2	8
3	13
4	20

+3   +2
+5   +2
+7

If the "growth of the growth" is constant, the relationship is quadratic.
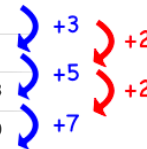
There is, however, a class of functions that grows even faster than quadratics: *exponential functions*.

If we try to calculate the growth between the y-values, we can immediately tell it's not linear. But then if we try to calculate the "growth of the growth", we see that it's not quadratic either. Even if we calculate the "growth of the *growth of the growth*" (shown in green)... we still haven't found a constant. In fact, each of these "growths" just repeats the original pattern of y-values! Something is making this function grow so fast that our attempt to calculate the rate of change fails to simplify anything.
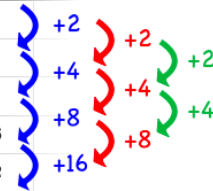
x_1	y_1
1	2
2	4
3	8
4	16
5	32

+2   +2   +2
+4   +4   +4
+8   +8
+16

Exponential functions grow so rapidly that looking for "what is *added* to y?" isn't helpful at all. The only way to talk about their growth is to start noticing "what is y being *multiplied* by?"

In this case, we can see that the y-values are doubling each time!

x_1	y_1	
1	2	×2
2	4	×2
3	8	×2
4	16	×2
5	32	×2

- Complete What Kind of Model? (Tables)
- Be ready to discuss your answers with the class!

Investigate

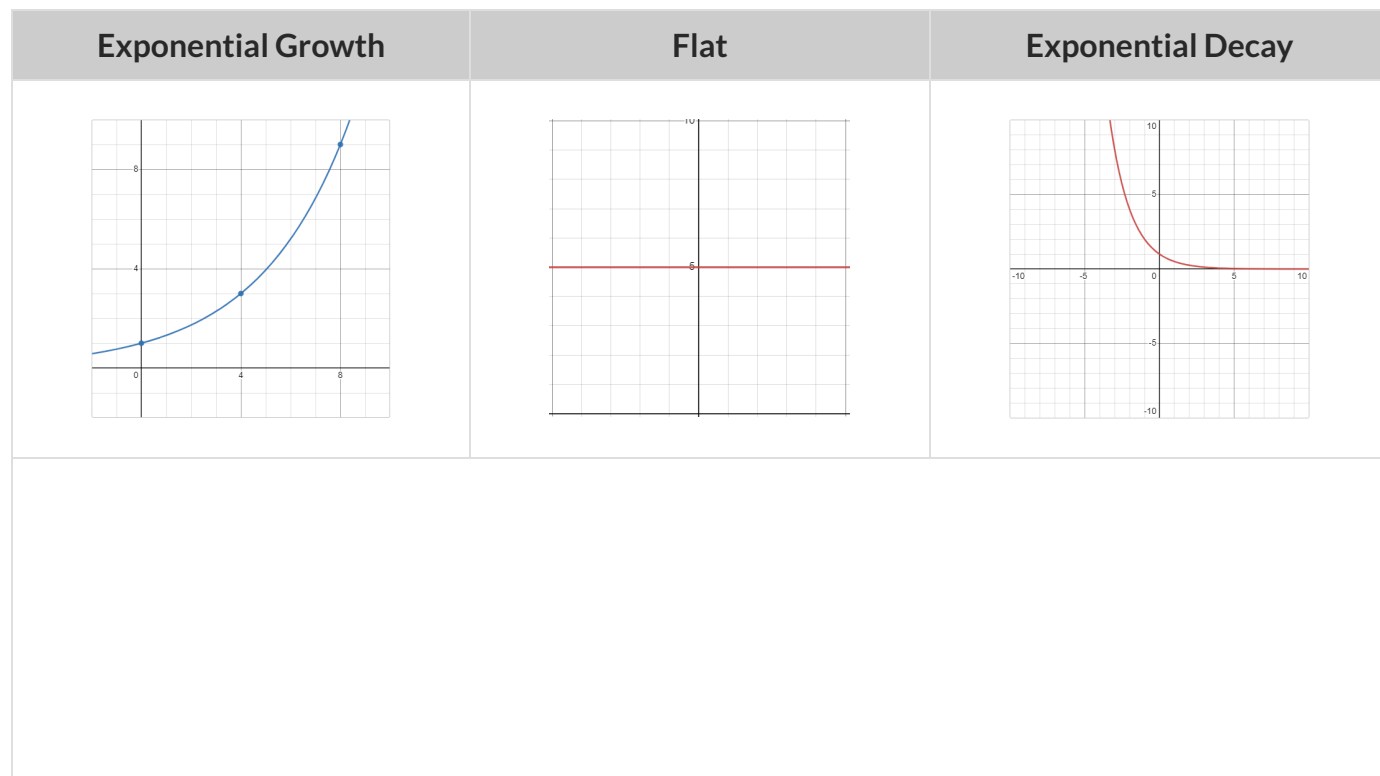We generally write exponential functions like this: $f(x) = ab^x + k$.

Let's explore what each coefficient means!

Use the third slide of Modeling Covid Spread (Desmos) to complete the first section ("base") of Graphing Exponential Models.

Review students answers, and then debrief via class discussion. Invite students to consider what new information they have gained by looking at graphical representations rather than tables.

The base of an exponential function (b) must **always be positive**, because exponential functions grow and decay uniformly. A negative b would bounce from one side of the y-axis to another. When raised to a fractional power, negative values of b might also lead to things like $\sqrt{-2}$!

Exponential Growth	Flat	Exponential Decay

Exponential Growth	Flat	Exponential Decay
$b > 1$	$b = 1$	$0 < b < 1$
When the base is **larger** than 1, the function starts out flat and then grows by the "percentage greater than 1". A base of 1.25 - or $(1 + 0.25)$ - will grow by 25% each time x grows by 1. In this instance, the base is also called the *growth factor*, since it determines how quickly the function grows.	When the base is equal to 1, the function stays flat without any growth of all (raising 1 to *any* power will always produce 1!).	When the base is **smaller** than 1, the function shrinks by the "amount less than 1". A base of 0.25 - or $(1 - 0.75)$ - will shrink by 75% each time x grows by 1. In this instance, the base is also called the *decay factor*, since it determines how quickly the function shrinks.

Use the third slide of Modeling Covid Spread (Desmos) to complete the second section ("vertical shift") of Graphing Exponential Models.

An exponential function with a *growth factor* will always start close to a horizontal line, then gradually shoot up to ever-increasing values. An exponential function with a *decay factor* will drop quickly, then level out close to a horizontal line. This horizontal line is called an *asymptote*, and the equation of the line will always be $y = k$.

Adjusting k shifts the asymptote up and down, along with the rest of the exponential curve that approaches it.

Use the third slide of Modeling Covid Spread (Desmos) to complete the last section ("initial value") of Graphing Exponential Models.

The y-intercept appears *differently* in exponential function definitions than in linear and quadratic definitions:

- In both linear and quadratic functions, we could cross out the linear or quadratic term when $x = 0$ (because anything multiplied by zero is zero) and the constant term being added or subtracted in the equation was our y-intercept.
- But, because any value raised to the power of zero is 1, when $x = 0$ in exponential equations, part of the exponential term *remains*, for example: $4(2^0) = 4(1) = 4.$

- As a result, **the y-intercept of an exponential function is $a + k$.**

- When there is no k-term being added or subtracted, the coefficient a is the initial value where $x = 0$.

- And, if a is "missing", the value of the coefficient is 1.   *After all,* $2^x = 1(2^x)$

  That means that if we don't see a or k in an exponential equation, the y-intercept of the function is 1.

Exponential growth and *exponential decay* show up all the time!

- Most cells (e.g. bacteria, the cells in a growing fetus, etc) divide every few hours, doubling the number of cells each time. A single cell will split into 2, those 2 cells will split to become 4, which will become 8, then 16, and so on.

- Unstable particles degrade into stable particles over time, emitting radiation as a byproduct. We use the term *half-life* to refer to the length of time it takes for 50% of the particles in a sample to become stable, leaving behind the other half as radiation-emitting material.

- Money in a savings account grows by a certain percentage each year. 3% growth on $100 would turn into $103. The next year that would become $106.09. And the next year $109.27. Every year there's a little more money to grow. If you start saving early, the account will grow into quite a lot more money down the road.

In the following two activities, students will decide whether various scenarios and definitions represent quadratic, linear, or exponential functions. They will also have opportunities to think about and apply their knowledge of growth, decay, initial value, and growth factor.

- Complete [What Kind of Model? (Descriptions)](#).
- What strategies did you use to decide if a function was linear, quadratic, or exponential?
- What new insights did you gain about exponential functions by thinking about them in real-world scenarios?

Have students share their answers, asking them to notice and wonder about the sequences for the exponential examples. How are these sequences growing or decaying? How is that growth or decay different from what they've seen before?

- Complete [What Kind of Model? (Definitions)](#).
- What strategies did you use to decide if a function was linear, quadratic, or exponential?
- What new insights did you gain about exponential functions by thinking about their definitions?

As students discuss their answers, pay special attention to their use of vocabulary when describing the initial value and the growth factor.

Synthesize

- You looked at several different representations of exponential functions: tables, graphs, descriptions, and equations.

- Which representation was the most useful for you? Why?

- Which representation was the least useful for you? Why?

Fitting Exponential Models   *30 minutes*

Overview

Students extend their sampling techniques to exponential relationships. Students continue experimenting in Desmos, but eventually switch back to Pyret to formalize their understanding.

Launch

Now that you're familiar with exponential functions, let's use them to model this Covid data!

Direct students to create a scatter plot showing the change in positive Covid cases for `MA-Table`. Then, support them in making educated guesses about the values of a, b, and k. Have students respond to the discussion questions below in pairs or small groups.

- Does your scatter plot show exponential growth or exponential decay?
  - *The scatter plot shows growth. The "hockey stick" is pointing up, meaning that positive cases are increasing.*
- Can we make any conclusions about the value of b? Explain.
  - *Because we see exponential growth, we know that b must be greater than one.*
- Can we make any conclusions about the value of k?
- Can we make any conclusions about the value of a? Explain.
  - *a must be positive, because the curve is consistently above k.*

Investigate

In the next activity, students use Desmos to find promising exponential models, and then fit the model programmatically in Pyret!

- Open to the fourth slide of [Modeling Covid Spread (Desmos)](#), and complete [Exponential Models - MA Table](#).
- Is an exponential model a good fit for this data? Why or why not?

★ *Optional:* Build models for *other states*. How do the coefficients differ from state to state? What differences between states could explain the different values of the coefficients?

Precision v. Efficiency in Computation

On [Exponential Models - MA Table](#) you'll see a note about the use of $*\sim*1$ to tell Pyret to prioritize speed over precision. Unlike most calculators that students will engage with, Pyret usually prioritizes precision.

In a math classroom, this is the difference between $^2/_3$ rendering as $0.\overline{666}$ or being rounded to 0.666666667.

In data processing, opting to round for speed over preserving precision can have ethical or technical consequences. For example:

1) When calculating a path over an extremely long distance, missing decimal places could result in the Mars Rover missing its destination.

2) For an extremely large population like China, rounding to 10 decimal places might result in discounting an entire subpopulation.

Synthesize

- What makes exponential models different from the linear and quadratic models you've seen before?

- How would you describe the shape of the three models you've seen so far (Linear, Quadratic, and Exponential)?

- Is it always okay for Data Scientists to round off their numbers to speed up computation? Why or why not?