# Exploring Modeling Assumptions with Census Data: the United States and Guatemala

Gary Epp and Jean Marie Linhart

Central Washington University

2/12/2022

Given at SIMIODE Expo 2022

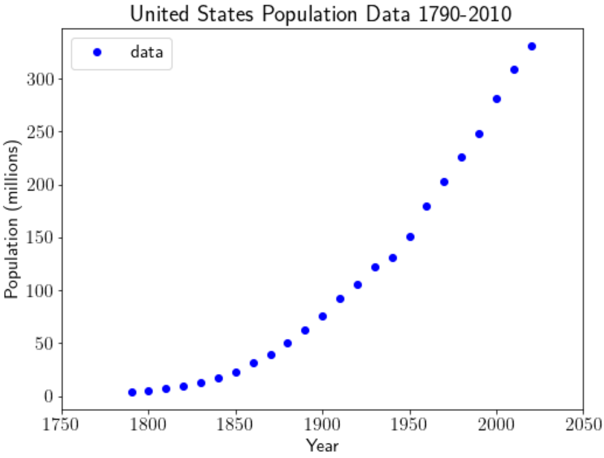
# Native Land Acknowledgement

As we begin, I want to acknowledge that I am on the traditional land of the Yakama people. I want to pay my respects to the native people of this land past, present and future, and to offer my gratitude for the land and water that supports me.

https://native-land.ca

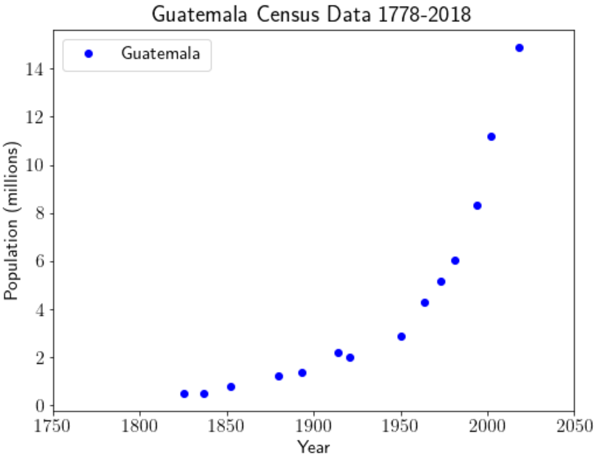
# United States Census Data

| Year | Population |
|---|---|
| 1790 | 3929214 |
| 1800 | 5308483 |
| 1810 | 7239861 |
| 1820 | 9638453 |
| 1830 | 12866020 |
| 1840 | 17069453 |
| 1850 | 23191876 |
| 1860 | 31443321 |
| 1870 | 39818449 |
| ... | ... |
| 1980 | 226545805 |
| 1990 | 248718302 |
| 2000 | 281424603 |
| 2010 | 308745538 |
| 2020 | 331449281 |

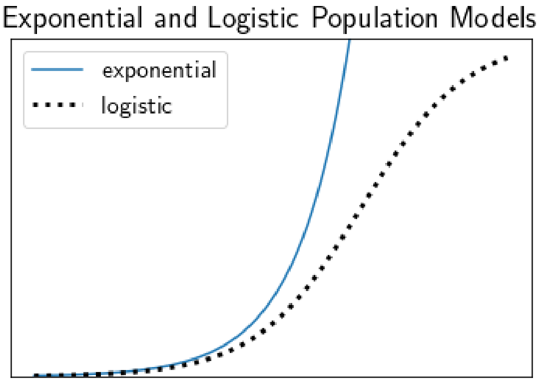


# Guatemala Census Data

| Year | Population |
|---|---|
| 1778 | 430859 |
| 1825 | 507126 |
| 1837 | 700000 |
| 1852 | 787000 |
| 1880 | 1224602 |
| 1893 | 1364678 |
| 1914 | 2183166 |
| 1921 | 2004900 |
| 1950 | 2870272 |
| 1964 | 4287997 |
| 1973 | 5160221 |
| 1981 | 6054227 |
| 1994 | 8321067 |
| 2002 | 11183388 |
| 2018 | 14901286 |



# Exponential and Logistic Models




How do the graphs of these functions look like the graph of the United States Census Data and the Guatemala Census Data and how do they differ?

# Modeling Assumptions 101: Simplify and Capture the Physics

- Exponential Model: $\dfrac{dP}{dt} = rP, \; P(0) = P_0$
- Solution: $P(t) = P_0 e^{rt}$

"The rate of change of the population is directly proportional to the size of the population."

If a population of 100 people has 2 births and 1 death in a year, what do we expect will be the births and deaths for a population of 1000 in a year?

# Modeling Assumptions 101: Capturing the Physics

- Exponential Model: $\dfrac{dP}{dt} = rP$

"The rate of change of the population is directly proportional to the size of the population."

If a population of 100 people has 2 births and 1 death in a year, what do we expect will be the births and deaths for a population of 1000 in a year?

We expect 20 births and 10 deaths for this new population: total change $+10$ people/year. In this model $r = .01$.

# Critical Thinking About the Exponential Model

We want to start with the *simplest model*:

- Exponential Model: $\dfrac{dP}{dt} = rP$

Do you expect the same proportionality constant to work on populations throught every historical era? Do you expect the same proportionality constant to work for every human population across the globe? Why or why not?

# Modeling assumptions 101: Simplify and Capture the Physics 2

The *logistic model* includes a second cause of deaths, deaths from two person interactions.

- $P$ possible choices of the first person
- $P-1$ possible choices for the second person

Logistic model:
$$\begin{aligned} \frac{dP}{dt} &= aP - kP(P-1) \\ &= aP - kP^2 + kP \\ &= (a+k)P - kP^2 \\ &= (a+k)P\left(1 - \frac{k}{a+k}P\right) \\ &= rP\left(1 - \frac{P}{L}\right) \end{aligned}$$

where $r = a + k$ and $L = \frac{k}{a+k}$

# Critical Thinking About the Logistic Model

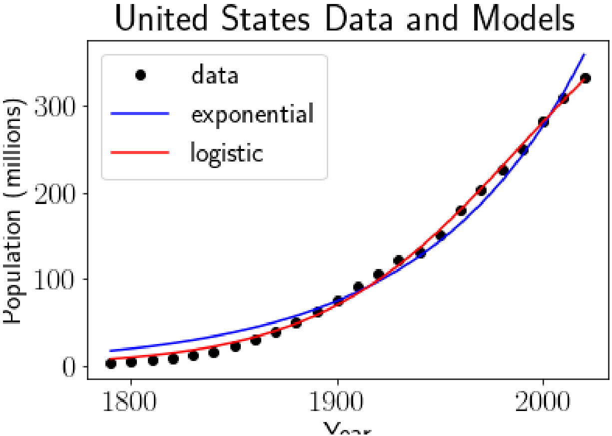
- Logistic Model: $\dfrac{dP}{dt} = rP\left(1 - \dfrac{P}{L}\right), P(0) = P_0$
- Solution: $P(t) = \dfrac{L}{1 + \frac{(L-P_0)}{P_0}e^{-rt}}$

Both the exponential and the logistic models are simplifications of a more complicated reality.

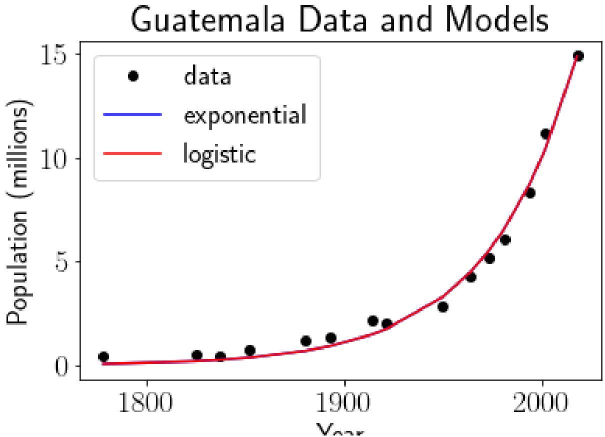
What other factors influence the size of human populations like the population of the United States that we have omitted from both the exponential and the logistic model?

# Modeling results with the United States Data

We use nonlinear least squares parameter estimation with the United States and Guatemala Data.



# Modeling results with the Guatemala Data




The carrying capacity of Guatemala was estimated to be $L = 2.6 \times 10^8 \pm 1 \times 10^{29}$ million people, compared to $L = 501 \pm 29$ million people for the United States.

# Trouble

Something definitely went wrong when we tried to get best fit parameters with the Guatemala model. We estimated an extremely high carrying capacity with an even more extreme error and the logistic model ended up being almost identical to our exponential model. Why is that and what's going on here?

To answer, we have to look more closely at our modeling assumptions, some that are hidden from our original derivation of these models.

# Per Capita Population Growth Rate (PPGR)

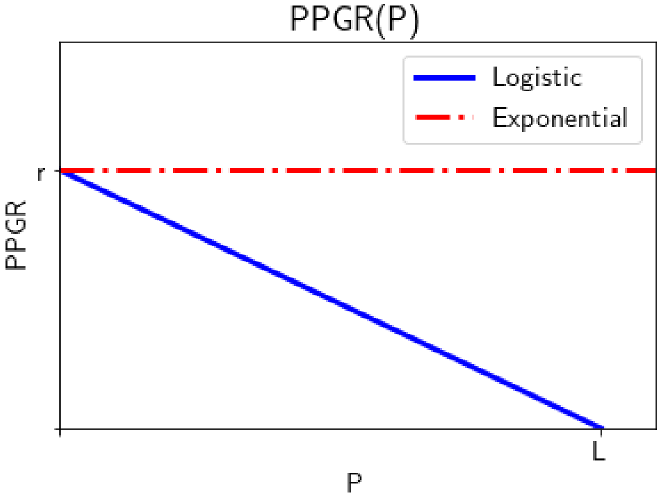
- Growth rate per person (average).
- Formula: $\frac{1}{P}\frac{dP}{dt}$
- Exponential:

$$\frac{dP}{dt} = rP \quad \longrightarrow \quad \frac{1}{P}\frac{dP}{dt} = r$$

- Logistic:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{L}\right) \quad \longrightarrow \quad \frac{1}{P}\frac{dP}{dt} = r\left(1 - \frac{P}{L}\right)$$

# PPGR assumptions of our models



# Calculating PPGR from our data

| $i$ | Year ($t_i$) | Population ($P_i$) |
|---|---|---|
| 1 | 1778 | 430859 |
| 2 | 1825 | 507126 |
| 3 | 1837 | 700000 |
| 4 | 1852 | 787000 |
| 5 | 1880 | 1224602 |
| 6 | 1893 | 1364678 |
| 7 | 1914 | 2183166 |
| 8 | 1921 | 2004900 |
| 9 | 1950 | 2870272 |
| 10 | 1964 | 4287997 |
| 11 | 1973 | 5160221 |
| 12 | 1981 | 6054227 |
| 13 | 1994 | 8321067 |
| 14 | 2002 | 11183388 |
| 15 | 2018 | 14901286 |

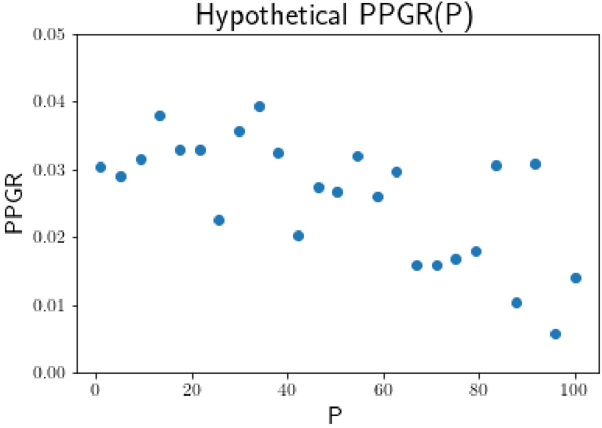
We use the forward difference to approximate the derivative

$$\frac{dP}{dt} \approx \frac{P(t+h) - P(t)}{h}$$

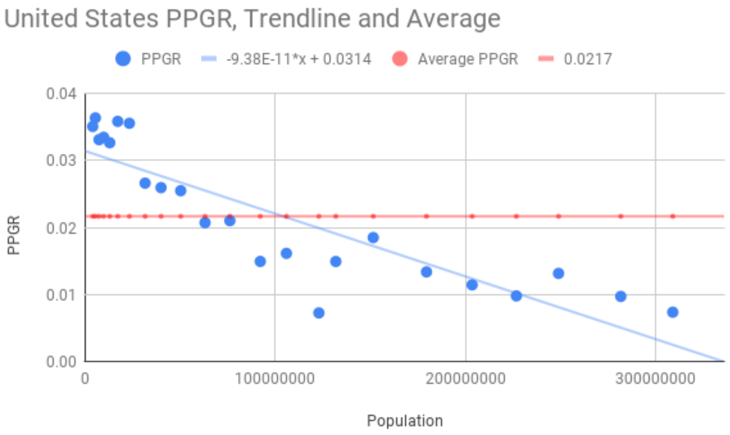
$$\begin{aligned} \text{PPGR} &= \frac{1}{P}\frac{dP}{dt} \\ &\approx \frac{1}{P_i}\frac{P_{i+1} - P_i}{t_{i+1} - t_i} \end{aligned}$$

# PPGR Pseudo Data

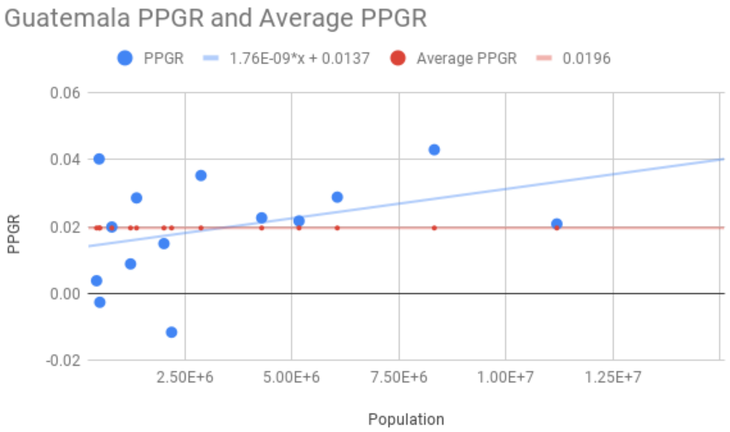
How would you get the horizontal line from the exponential PPGR model from this data? How would you get the line for the logistic PPGR model from this data?



# PPGR and United States Data



# PPGR and Guatemala Data




Note the trendline (best fit line) is increasing not decreasing. That's why we ran into problems trying to fit a logistic model to the Guatemala Data!

# Where can we go from here? Superexponential Growth

Logistic Growth (United States)

$$\frac{1}{P}\frac{dP}{dt} = r\left(1 - \frac{P}{L}\right)$$

$$r\,dt = \frac{1}{P\left(1 - \frac{P}{L}\right)}\,dP$$

$$P(t) = \frac{L}{1 + \frac{L}{A}e^{-rt}}$$

What happens?
$P(t) \to L$ as $t \to \infty$

Superexponential Growth (Guatemala)

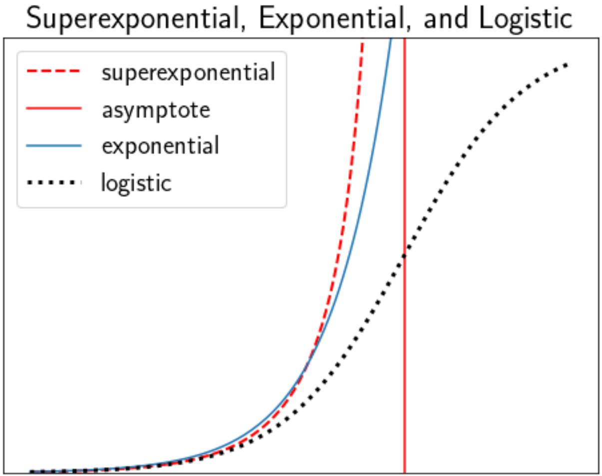
$$\frac{1}{P}\frac{dP}{dt} = r\left(1 + \frac{P}{L}\right)$$

$$r\,dt = \frac{1}{P\left(1 + \frac{P}{L}\right)}\,dP$$

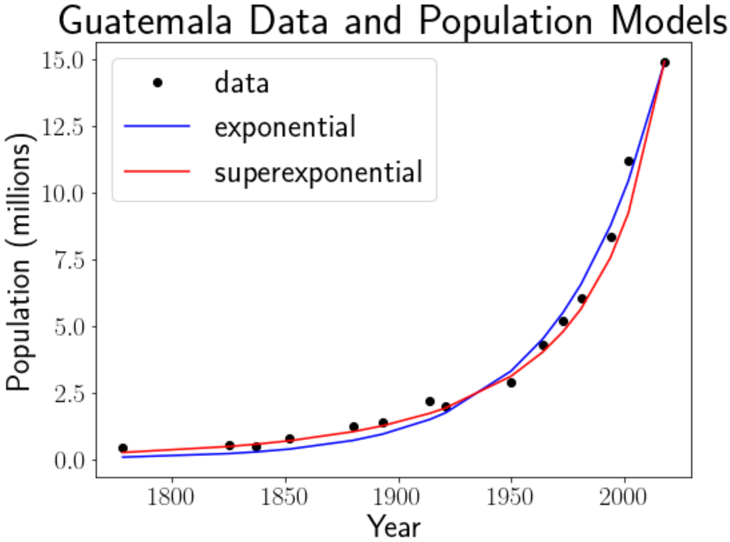
$$P(t) = \frac{L}{\frac{L}{A}e^{-rt} - 1}$$

What happens?
$P(t) \to \infty$ as $t \to \frac{1}{r}\ln\left(\frac{L}{A}\right)$

# Model comparison



# Guatemala




The year of the asymptote is 2054.......

# Thanks for listening!

Helpful resources:

- Linhart, Jean Marie (2021) "1-170-S-CensusModeling-StudentVersion", https://simiode.org/resources/8914.
- Linhart, Jean Marie (2017) "1-066-S-USCensusModeling," https://simiode.org/resources/4210.
- Linhart, Jean Marie (2019) "Using the United States Census Data to Introduce Differential Equations," *PRIMUS*. 29(7):702-711.
- A SIMIODE module on the superexponential model is in progress!