Carnegie Mellon University

Progress Report II

PHIGHT COVID RESEARCH PROJECT

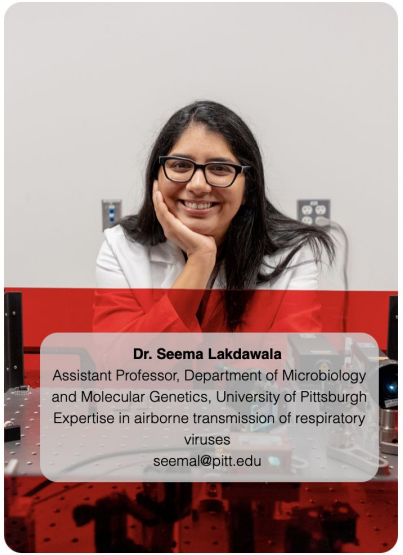
Lakdawala Lab: Seema Lakdawala
MSP Team: Cheyenne Ehman, Yixuan Luo, Zi Yang, Ziyan Zhu
Faculty Advisor: Valerie Ventura

03/22/2021

Introduction

PHIGHT COVID (Public Health Intervention aGainst Human-to-Human Transmission)

Lakdawala Lab - University of Pittsburgh



Carnegie Mellon University

Introduction

We seek to analyze the effectiveness of public health interventions (PHI) on the transmission of COVID-19, particularly how school policies affect transmission on the county level.

- Some stakeholders believe children act as a “factor” for COVID-19
- We are looking at Ohio
    - Ohio is good as a control as it has many statewide PHI’s, but fewer individual county interventions

Questions:

- How can we measure transmission?
- Does teaching method affect case/death rates in a county? If so, how?
- What covariates lead to lower transmission?

Datasets

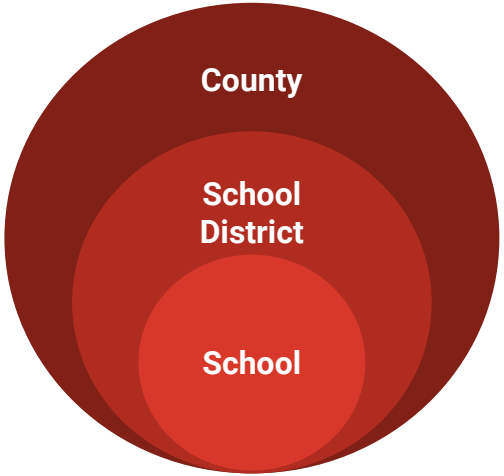
All data was collected and shared to us by Annika Avery

Ohio County Level Cases

- 35,820 observations, 17 Variables
- Date range: 01/22/2020-02/22/2021
- 88 different counties
- **Variables of Interest:**

County	Date	Population	New Confirmed	Cumulative Confirmed	New deaths	Cumulative deaths
Franklin	12/15/2020	1283688	918	70879	4	1004

Datasets



Ohio K12

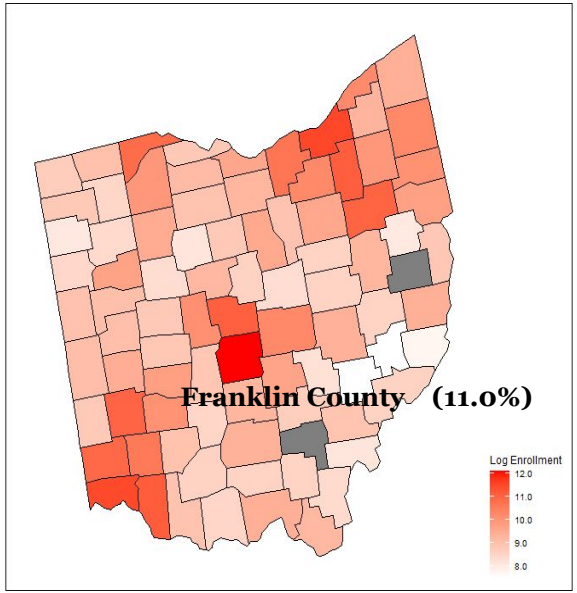
- Obtained from MCHdata.com
- 2786 observations (schools), 35 Variables
- **Variables of Interest:**

School Name, School District, City, County	North High School, Akron Public Schools, Akron, Summit
Enrollment by school, district	914, 21579
School opening date by district	9/9/2020
Teaching method	Online Only, Hybrid, On Premises, Pending, Other
Student & Staff Mask Policy	Required all * to wear mask; not required; pending; unknown
Temporary School Shutdowns	Close 1-5 days; close 6-14 days; never closed ; unknown

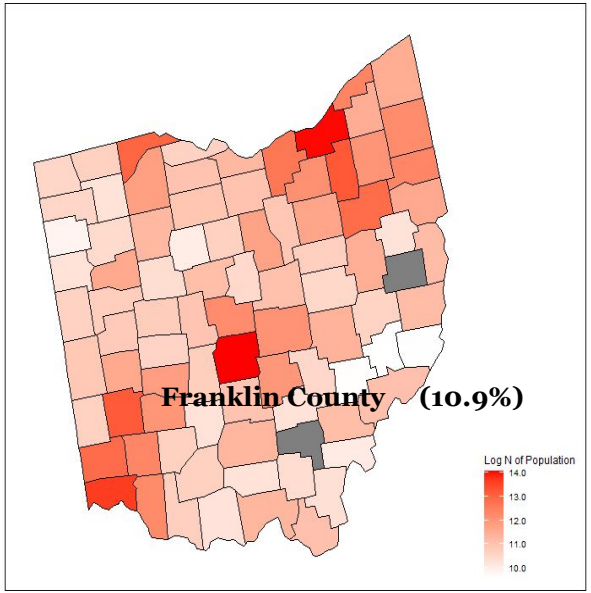
Carnegie Mellon University

Summary Statistics

- **Ohio States**
  - 88 Counties (86 counties enclosed in data)
  - Population: 11,755,535
  - Student enrollment: 1,615,134 (13.7%)
  - Number of schools: 2,871



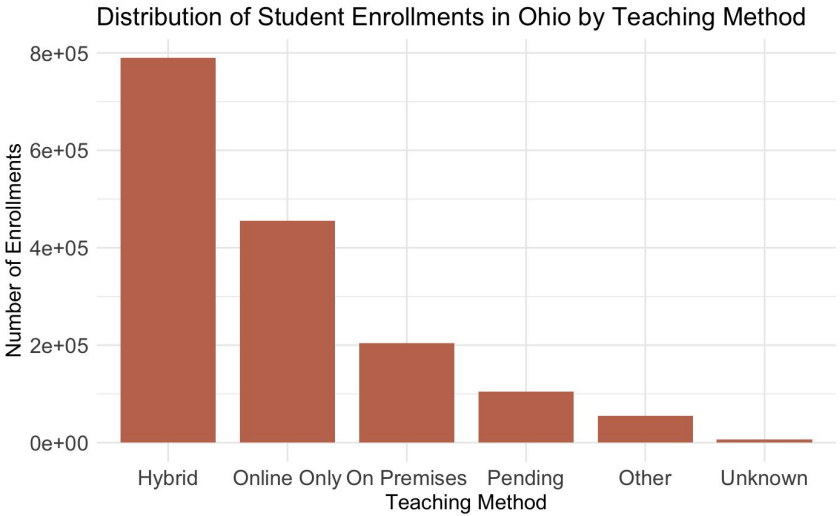
Log of Number of Enrollment



Log of Number of Population

Carnegie Mellon University

Teaching Method

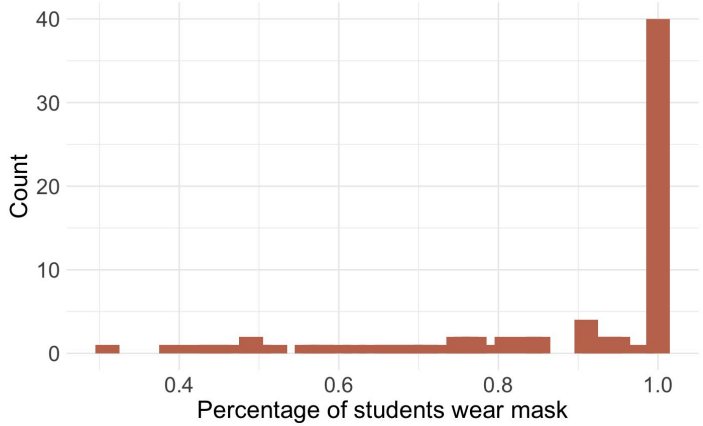


- **Majority Teaching Method (county level):**
  - Hybrid: 59 counties
  - On Premises: 16 counties
  - Online Only: 11 counties

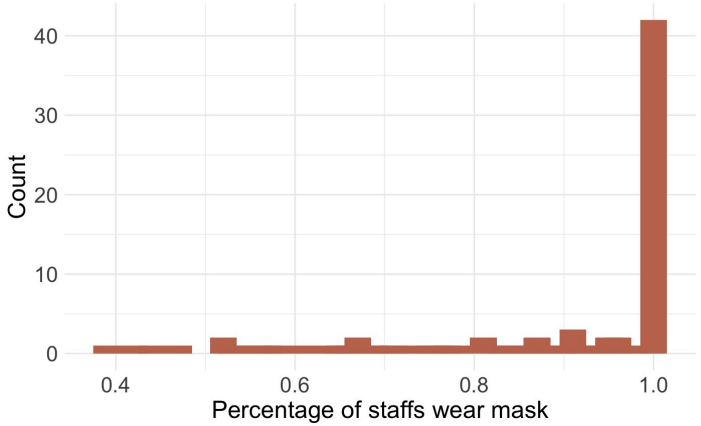
Carnegie Mellon University

Mask Wearing Policy

- **Percentage of Students Required to Wear Mask:**
  - Range: 31% ~ 100%
  - Large than 80%: 57 out of 86 counties



- **Percentage of Staff Required to Wear Mask:**
  - Range: 39% ~ 100%
  - Large than 80%: 61 out of 86 counties



Carnegie Mellon University

Data Cleaning & Wrangling

- Manually drop redundant columns
- Manually correct wrong entries and NA values
- Missing values:
  - Only impute missing county with the city information
  - Remove COVID cases observations with missing values in cases & deaths
  - Drop missing values case by case during EDA

- Bracket COVID cases data between school reopen dates and 10 days before Christmas break
- Death Increase Proportion
  - Cumulative cases/deaths on 2020/12/15 minus reopen date cumulative cases/deaths
  - Deaths Increase divided by population

Data Wrangling

County Level Aggregation: Weighted statistics by #students in the district / #students in the county

- Teaching method (*hybrid, online, on-premises = in person*)
  - Example, if county Adam has:

    50% students doing online only, 20% hybrid, 30% on premises

    The most common teaching method for Adam is *"Online Only"*.
- Online only (*online only, not online only*)
  - Example, if county Adam has:

    40% students doing online only, 20% hybrid, 30% on premises

    Adam is *"Not Online Only"* since (20%+30%) > 40%
- Required all students to wear mask
  - If proportion of students wearing mask > 0.8
- Required all staff to wear mask
  - If proportion of staff wearing mask > 0.8
- Most common school reopen date

Carnegie Mellon University

Methods

To find covariates that affect transmission with data on 02/22/2021

- Use deaths proportions to measure transmission
- Exploratory Data Analysis on Death Proportions by Different Policies
- Hypothesis Testing on Difference in Average Death Proportions
    - ANOVA on 3-level Teaching Method
    - Pairwise Comparison by Duncan Multiple Testing after significant ANOVA

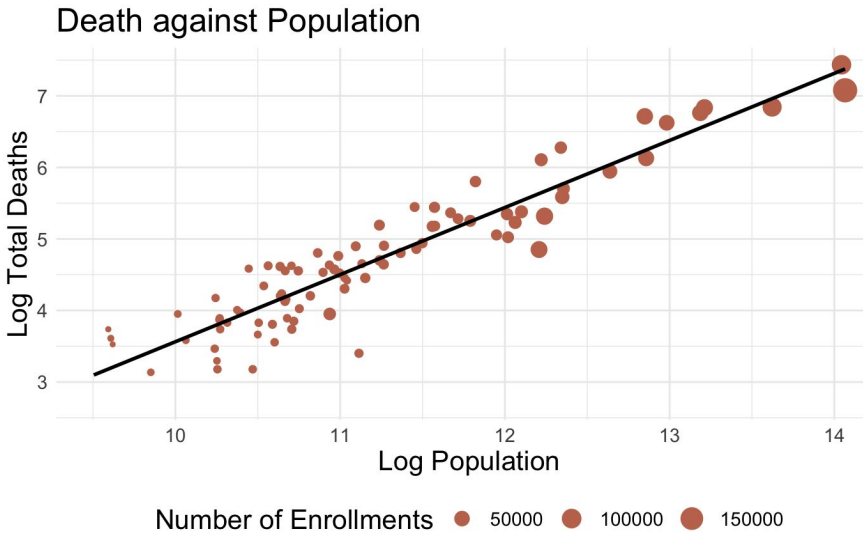
To see how interventions affect transmission over time (next step)

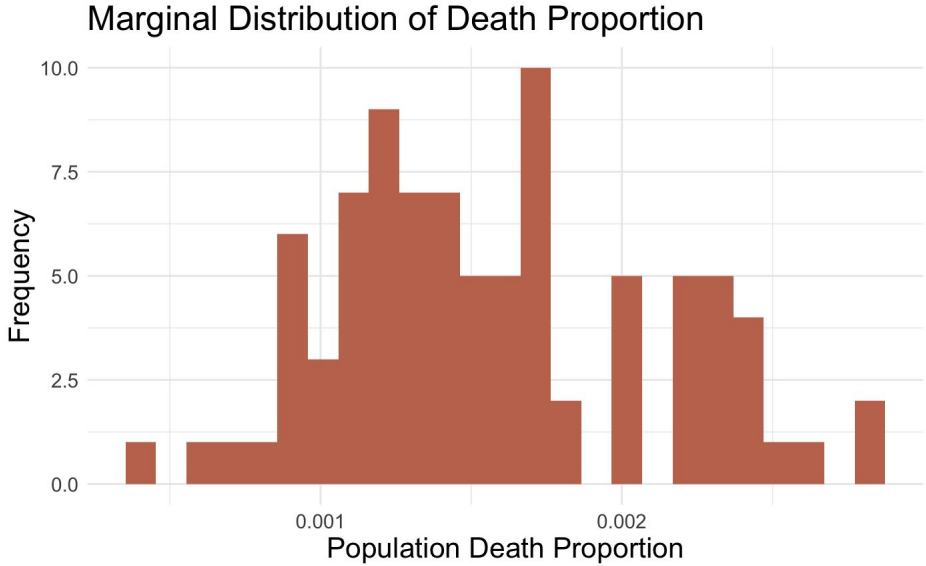
- Time series modeling

Carnegie Mellon University

Death is proportionate to the population

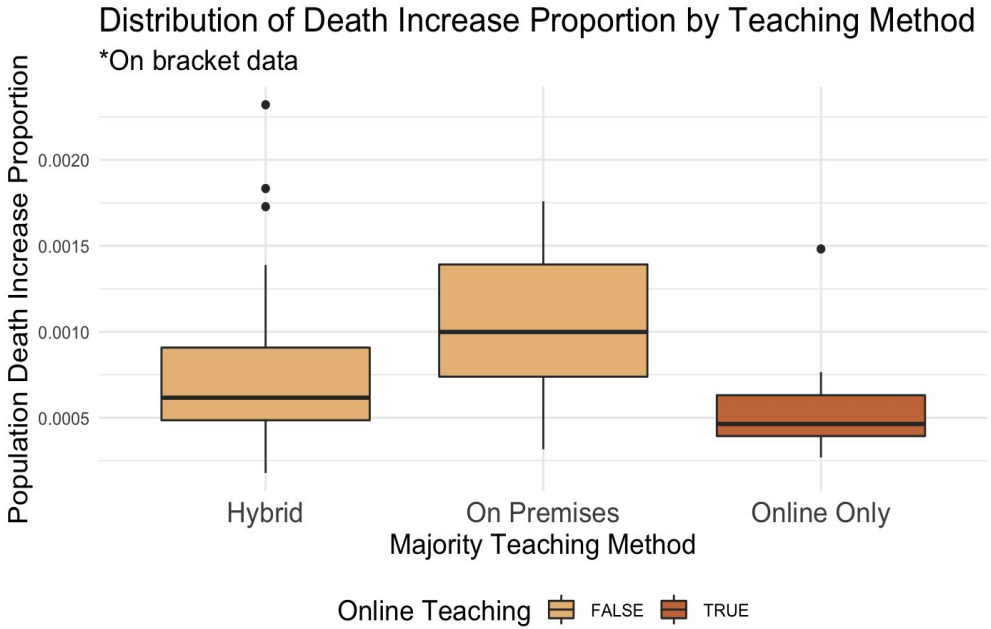
Death Proportion = # Deaths in Time Period/ Population





Carnegie Mellon University

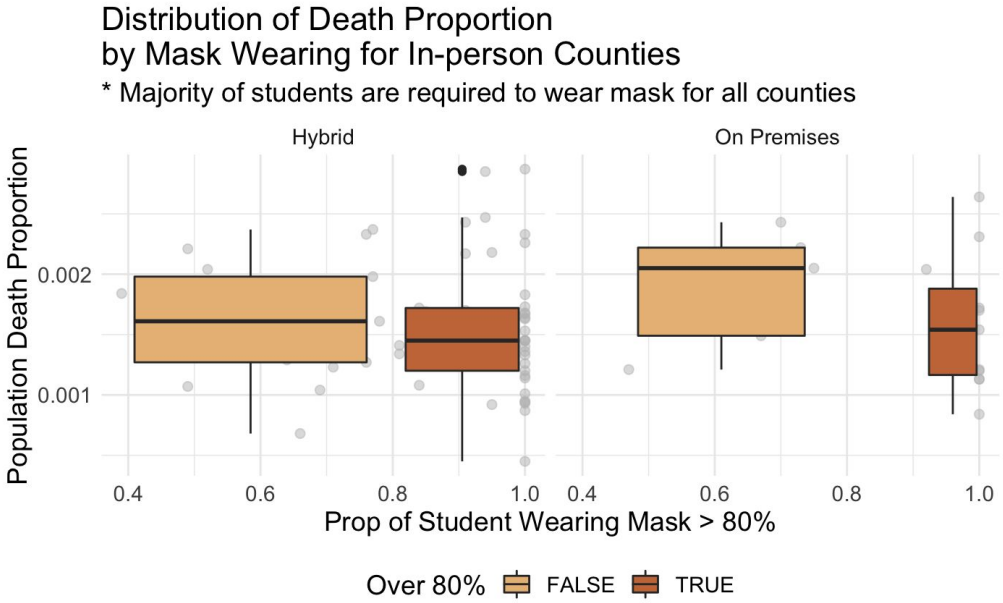
Death increase proportion significantly different by teaching method



- Possible stratification in death proportion by teaching method
- Counties with Online Only instruction have lower death proportions
- 1 way ANOVA significant on bracket data but insignificant on all time data
- Duncan's Multiple Range Test are significant for :
  Hybrid -- On Premises
  Online Only -- On Premises

Carnegie Mellon University

Death proportion not significantly different by 80%+ of students wearing masks or not



58 Counties

16 Counties

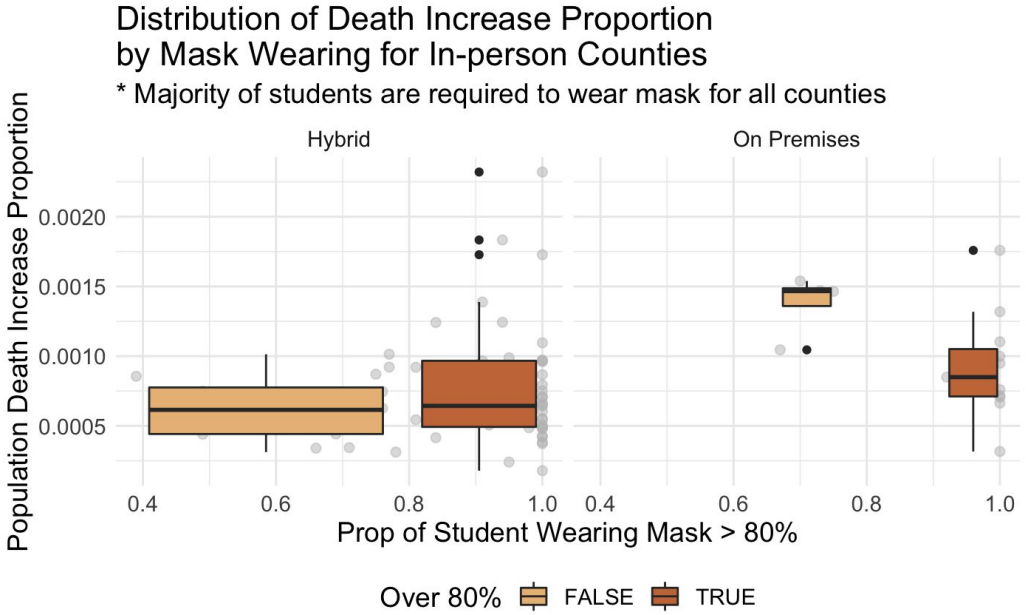
At a significance level of 0.05, for either Hybrid counties or On-Premises counties

1-way ANOVA insignificant on 80% cutoff

(same for cutoffs from 40% to 99%)

Carnegie Mellon University

Death increase proportion significantly different by 80%+ of students wearing masks for full in-person classes



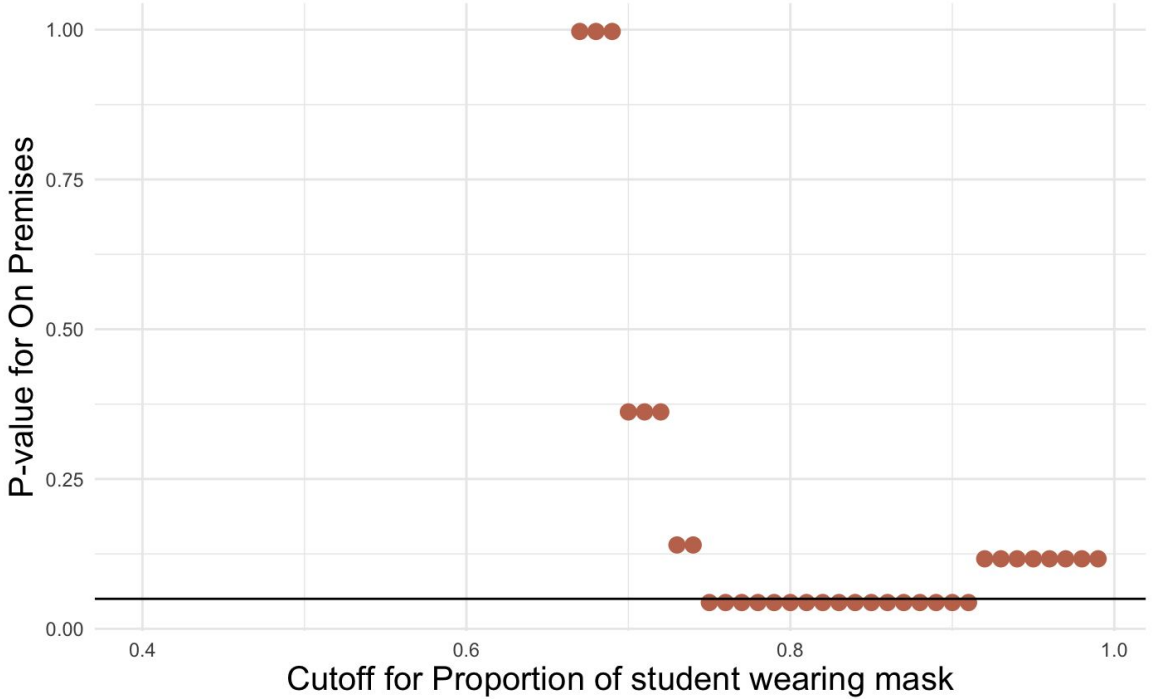
At a significance level of 0.05,

1-way ANOVA significant for wearing mask in On-premises counties

But insignificant in Hybrid counties

Carnegie Mellon University

Significant differences exist for cut-off between 75% to 91% when students taking full in-person classes



At a significance level of 0.05, for On-Premises counties

1-way ANOVA significant if we split the counties by a point between 75% ~ 91% students wearing mask.

Carnegie Mellon University

Road Blockers

- No data about the start and end date of the active school district public health policy
  - Status on ‘Last Verified Date’
  - What if there is a change?
- How to connect students’ behavior with the transmission in communities?
  - Mobility
- How to explain the differences of death proportions between Mask v.s. No Mask for Online Only? What are the potential covariates?

Next Steps

- Improve Data Bracketing Method
  - Find when log total deaths changes in trend after fall semester to set a better bracket window
  - Plot log total deaths as a time series
    - Assuming cumulative deaths is exponential, so taking the log makes it linear
    - Fit a linear regression only on fall semester data to see when the rest of the data deviates
    - Look only at counties where the majority of students are in person
  - Learn how long deaths/cases are influenced by the effect of school policy

Next Steps

- Connect with newly introduced stakeholders
  - University of Nabraska Medical Center:
    - James Lawler (Associate Professor)
    - Sarah Donavan (PHD Student)
  - Department of Veterans Affairs, Public Health:
    - Carter Mecher (Senior Medical Advisor)
- Summarize current EDA results on teaching methods (requested by client)
  - To be delivered to Nebraska public officials - Several Slides
- Explore the relationship between teaching method & deaths, start with simple linear regression

Carnegie Mellon University

Carnegie Mellon University

Thank you!

Q&A