Carnegie Mellon University

Effect of School Policies on COVID Transmission

PHIGHT COVID RESEARCH PROJECT

PHIGHT COVID: Seema Lakdawala, Annika Avery, Rebecca Nugent
MSP Team: Cheyenne Ehman, Yixuan Luo, Zi Yang, Ziyan Zhu
Faculty Advisor: Valerie Ventura

04/06/2021

Introduction - Questions

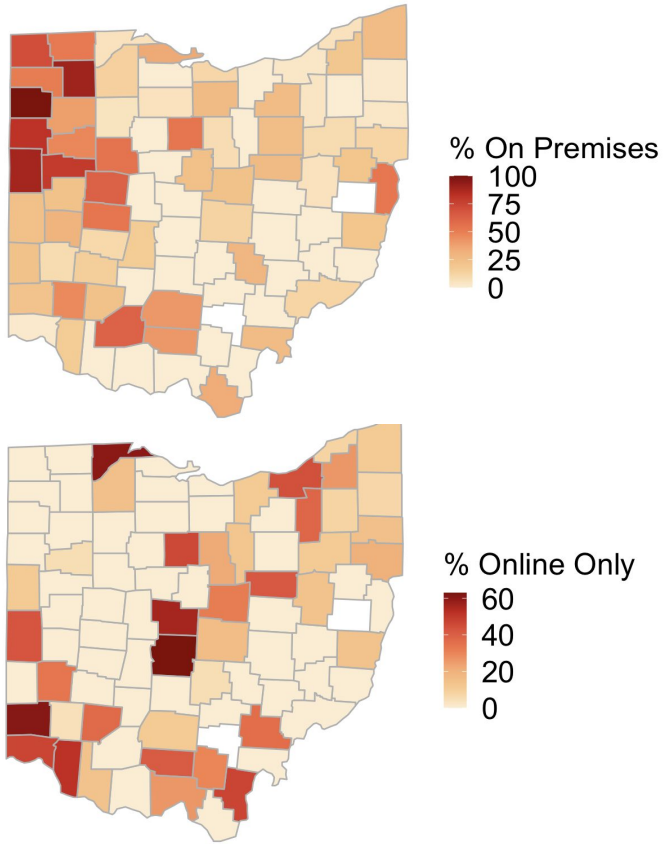
Main Research Questions:

- Does teaching method have effect on the COVID-19 transmission in Ohio?
  - How can we measure transmissibility with available data?
  - What other factors are affecting the transmission?
  - Is the effect still significant after we adjust for confounders and all possible covariates we can get?

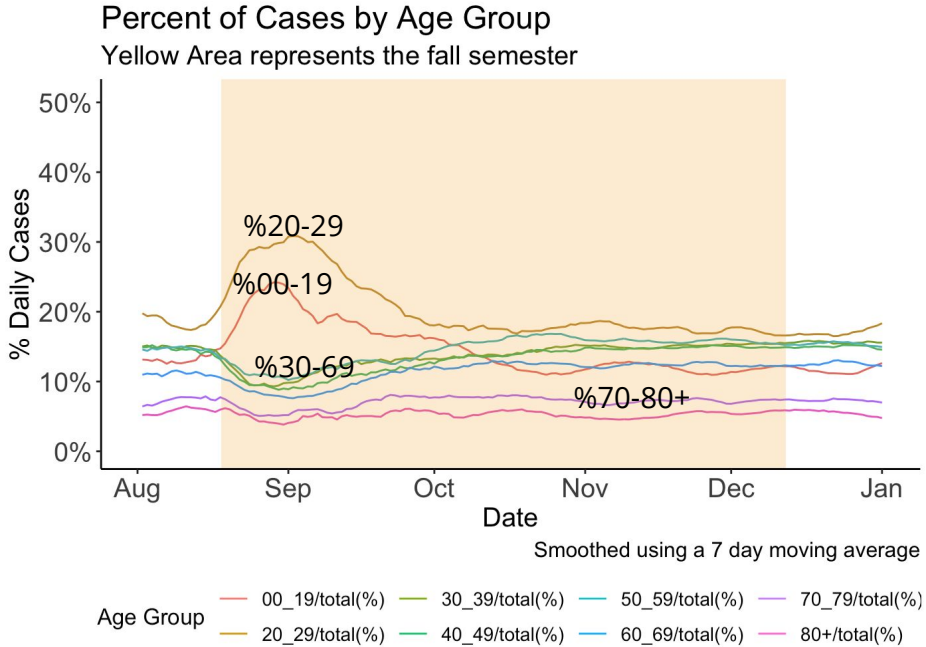
Introduction - Motivation

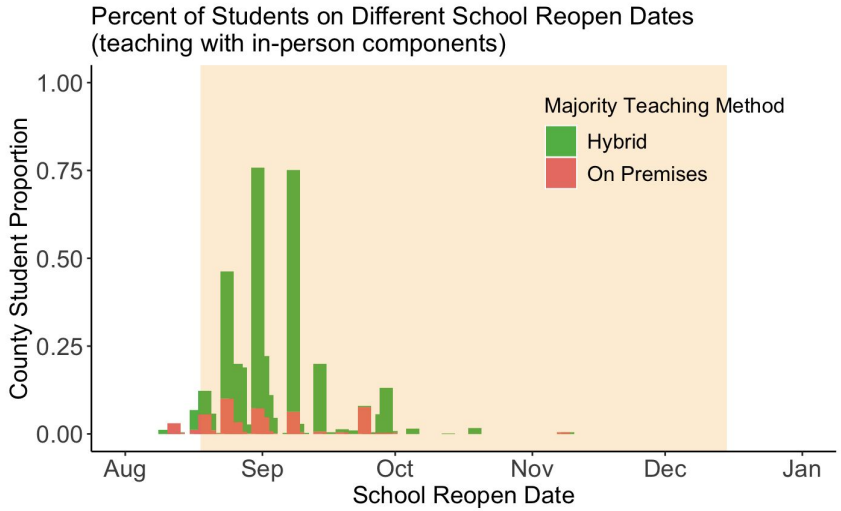
Why Ohio?

- Counties are comparable with respect to public health interventions;
  - Most interventions are statewide,
  - Few are at county-level.
- But there is a wide range of school teaching methods so we can study their effect on covid infections.



Introduction - Motivation





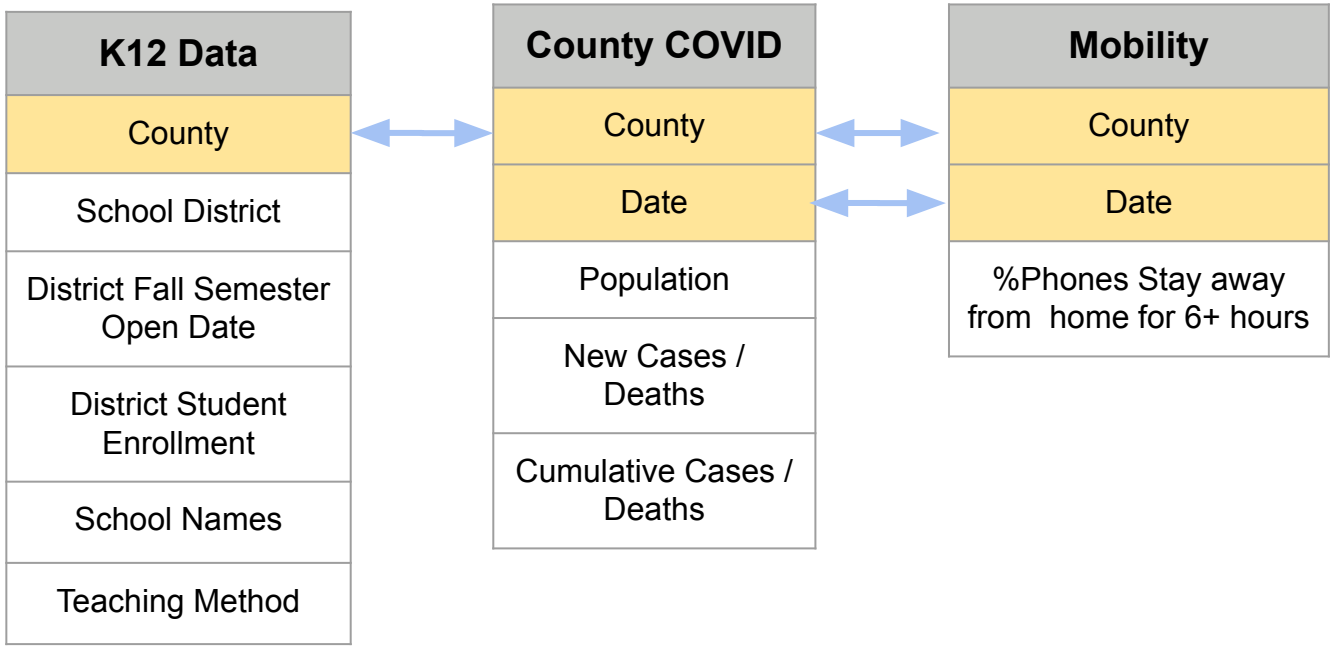
Fall Semester: 08/18 - 12/15

% Daily cases under 29 years old peaks in late August, overlapping with school reopening

Data

- **Data Sources**
  - Cases & Deaths: John Hopkins Open Source Data API
  - K12 school policies: MCH.com
  - Mobile Mobility: SafeGraph.com via CMU DELPHI Group
- **Time Range:** 01/22/2020 - 02/22/2021
- **About Ohio State:**
  - 86 counties (2 dropped due to missing data)
  - 11,755,535 Population
  - 1,615,134 student enrolled in K12 schools (13.7% of population)
  - 2,871 schools

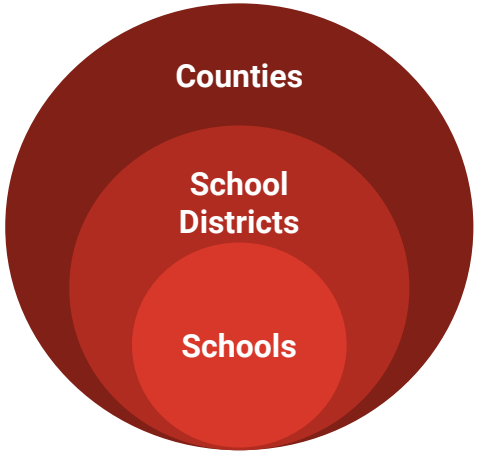
Data - Relation



Data - Aggregation

County-wide variables

Death Incidence per 1000	Cumulative Deaths * 1000 / population
Online Only Proportion	#Student went **Online Only /** County Student Enrollment
Hybrid Proportion	#Student went **Hybrid /** County Student Enrollment
On Premises Proportion	#Student went **On Premises /** County Student Enrollment
Majority Teaching Method	Teaching method in county with highest proportion
Percent Mobiles Stay Away Home for 6 hours +	Ought to suggest mobility of full-time workers/students



* We use death counts as the response variable because it is more reliable than case counts.

Method

- **Is there an effect of school teaching method?**

  What are the other confounders and covariates?
  - Exploratory Data Analysis
    - Times series plot of cumulative death incidence
    - Boxplot of death incidence during fall semester
  - ANOVA and Post-hoc Testing of death incidence
  - Identify confounders based on the availability of data
    - Death incidence before the fall semester
  - Adjust for confounders but realize a potential problem 😭

Method

- **How can we measure the transmission better?**

  Assume the disease follows an exponential growth and the process is stationary, then

$$E(logN_t) = N_0 + B \times t$$

  Nt = number of new infections on day t

  N0 = log number of new infections on day 0

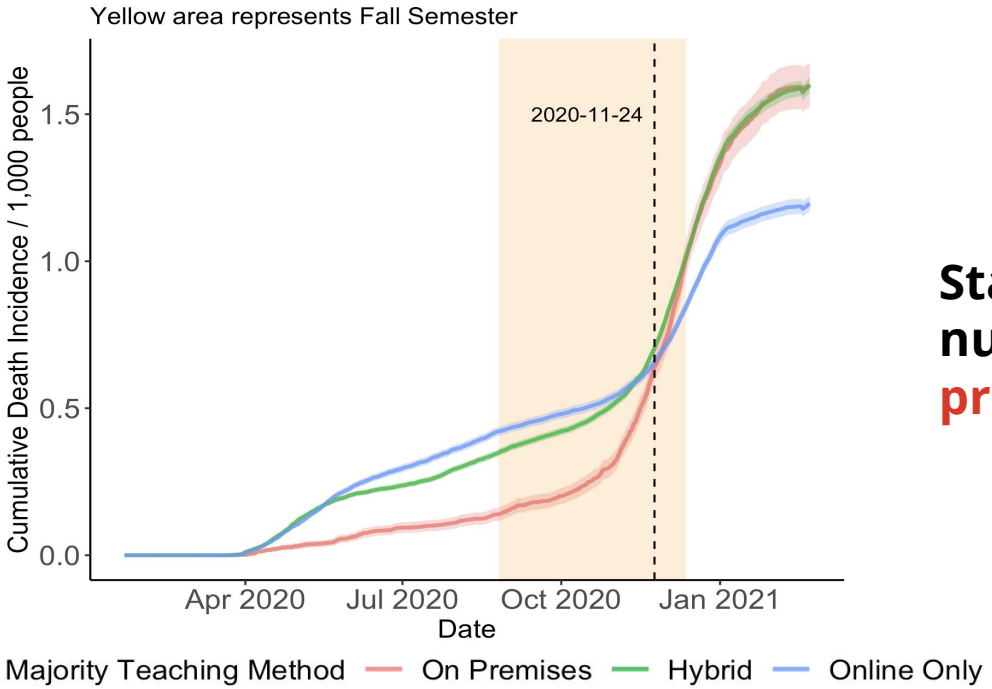
  B = exponential growth coefficient.

  - Use New deaths + 1 as new infections (adjust for 0 new deaths)
  - Time series plot of Log(New deaths + 1) and B for different teaching methods

CAUSAL INFERENCE IN THE TIME OF COVID-19: https://arxiv.org/pdf/2103.04472.pdf

Results - EDA (Time Series)

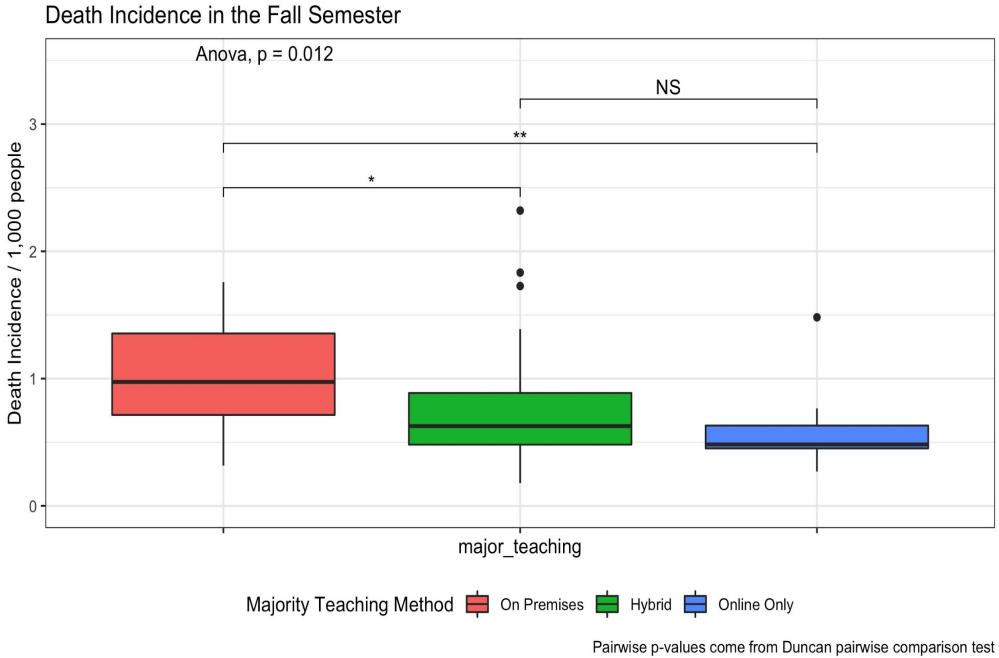
Are the county death trends significantly different by school teaching methods?



Starting mid-semester, death numbers increase faster for on premises counties

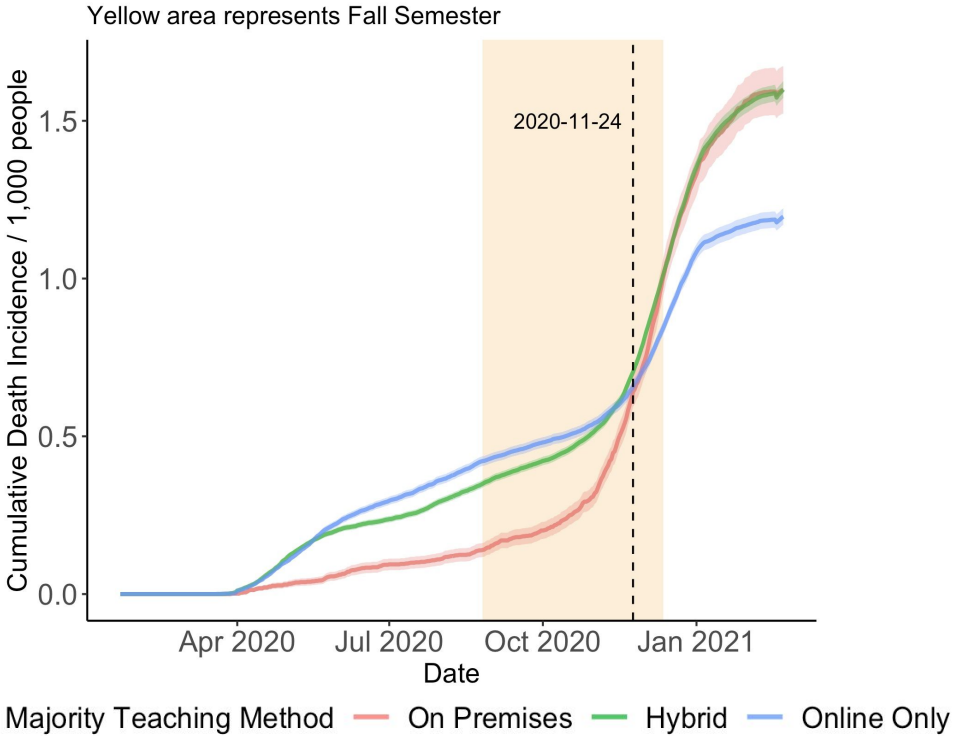
Results - EDA & ANOVA, Post-hoc Testing

How is teaching method related to death proportions?



Death proportions averaged within red, green and blue counties are significantly different (p= .012)

Results - Confounding Effect

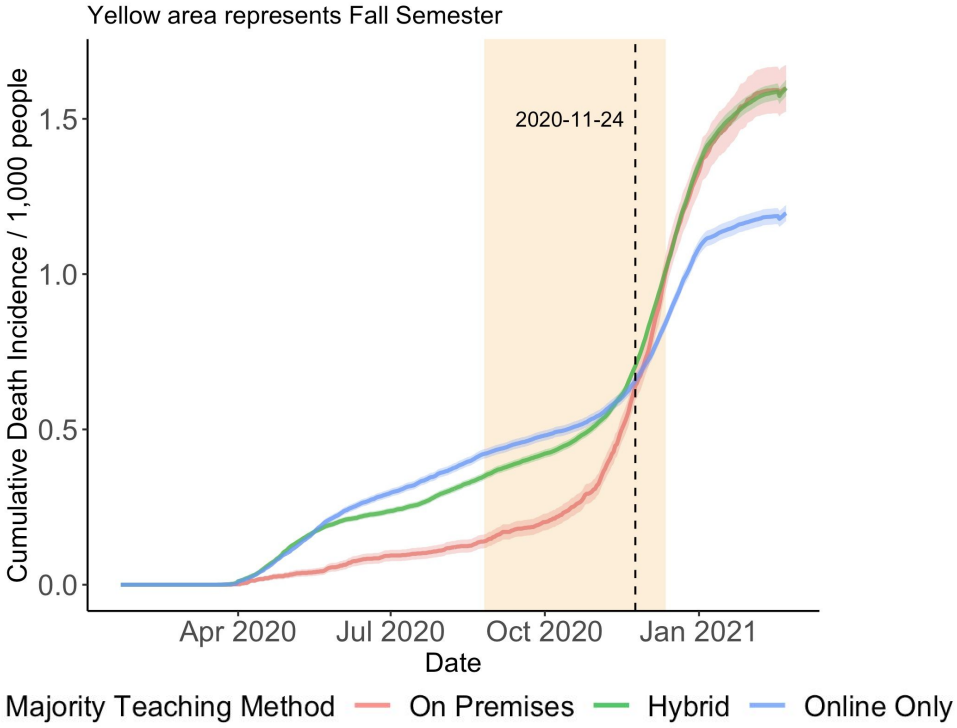


Note:

- Low (high) death rates before the semester implies low (high) death rates during the semester
- Low (high) death rates before the semester implies mostly on premises (online) teaching

→ Death rates *before the semester* is a **confounder**

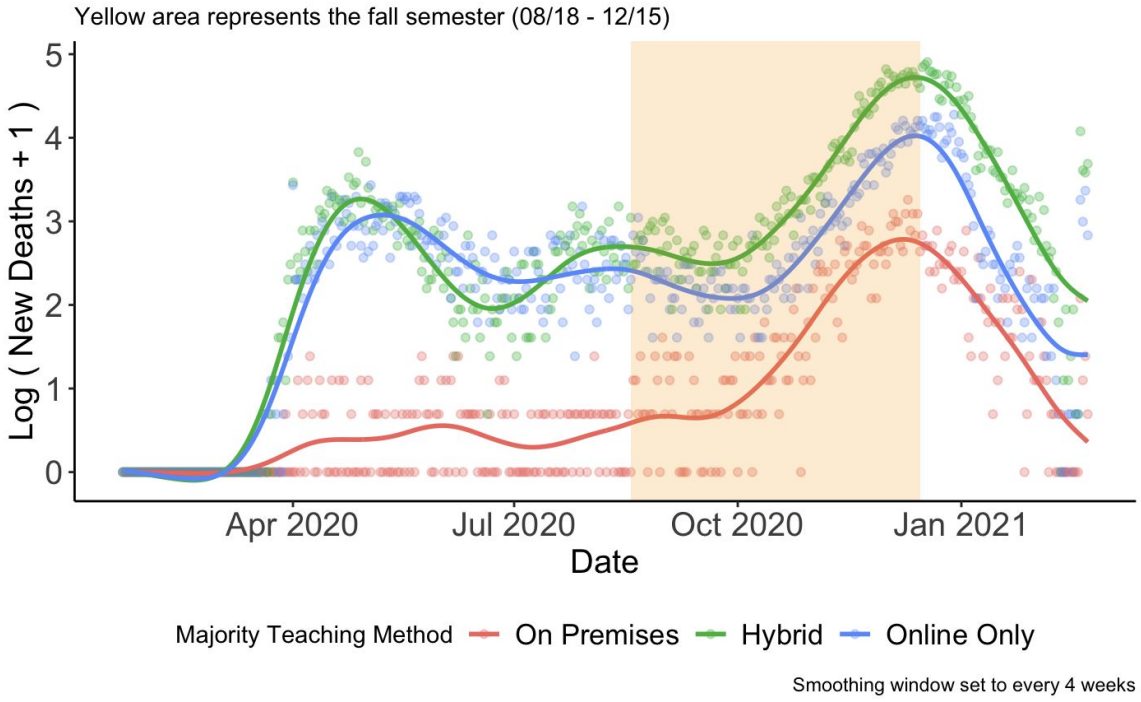
Results - Adjust for Confounder



Death proportions averaged within red, green and blue counties remain significantly different after adjusting for Y0 (p= .011)

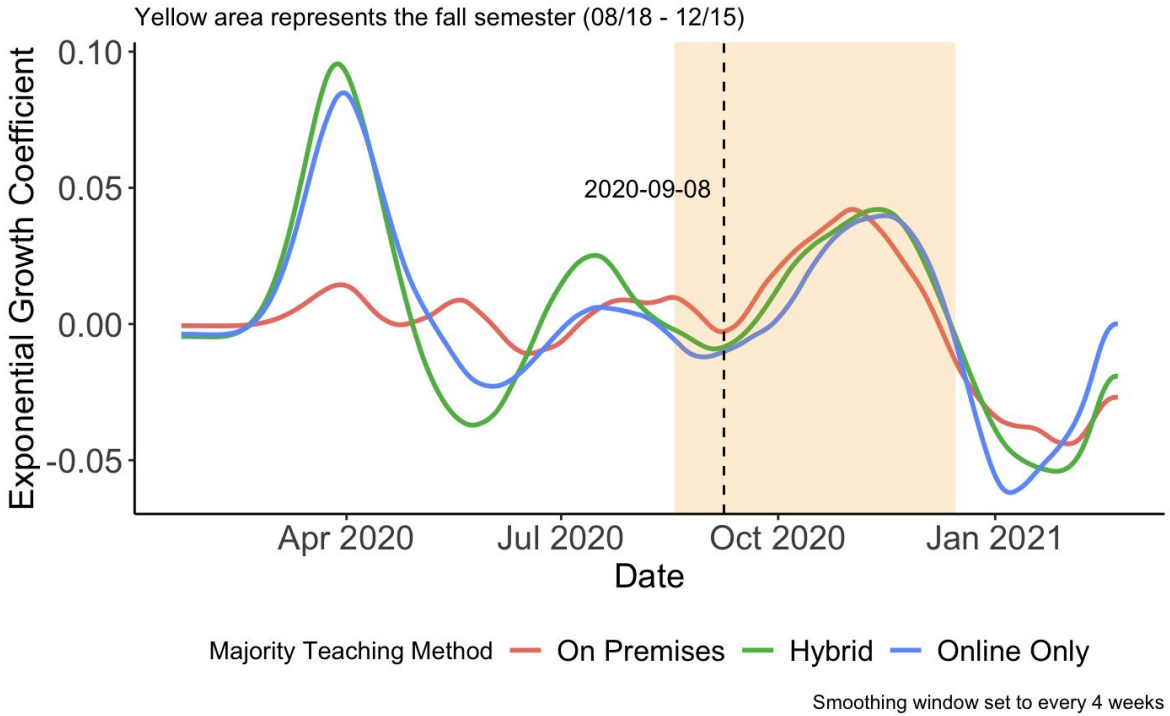
Results - New Model

How can we better measure transmissibility?



$$E(logN_t) = N_0 + B \times t$$

Results - New Model

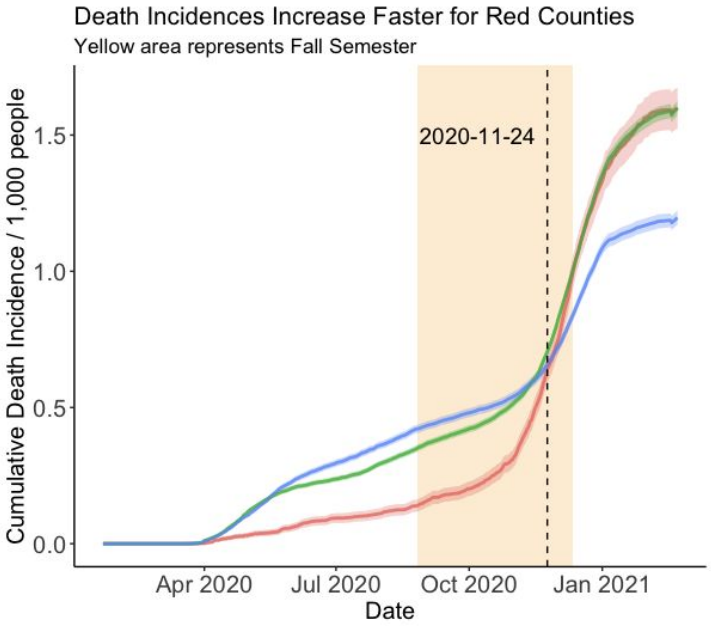


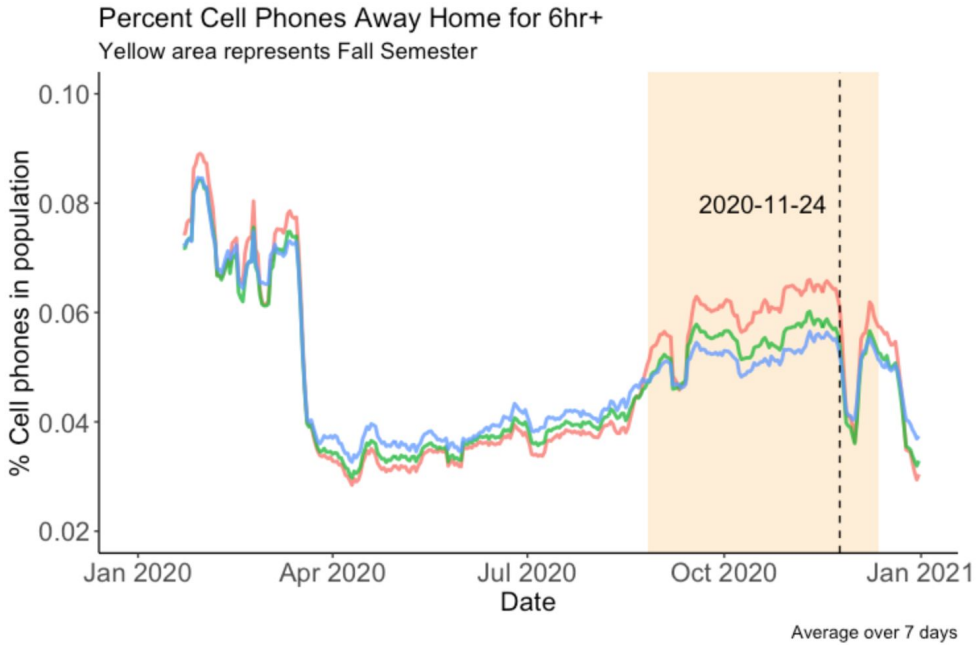
Exponential growth coefficients *B* change with time

$$E(logN_t) = N_0 + B \times t$$

Results - Other Factors

Similar ordering in death numbers and cell phone mobility for on-premises & online-only counties





Majority Teaching Method — On Premises — Hybrid — Online Only

Discussion - Recap & Next Steps

- Does teaching method have effect on the COVID-19 transmission in Ohio?
  - Death proportions averaged within On-Premises, Hybrid and Online Only counties remain significantly different after adjusting for confounder Y0
  - We have not yet come to the conclusion, we have to adjust for other important factors like mobility
- What other factors are expected to affect the transmission?
  - Mobility, Population size, Urban/Rural Status, Testing volume
  - Incorporate serology data (time permitting)

Carnegie Mellon University

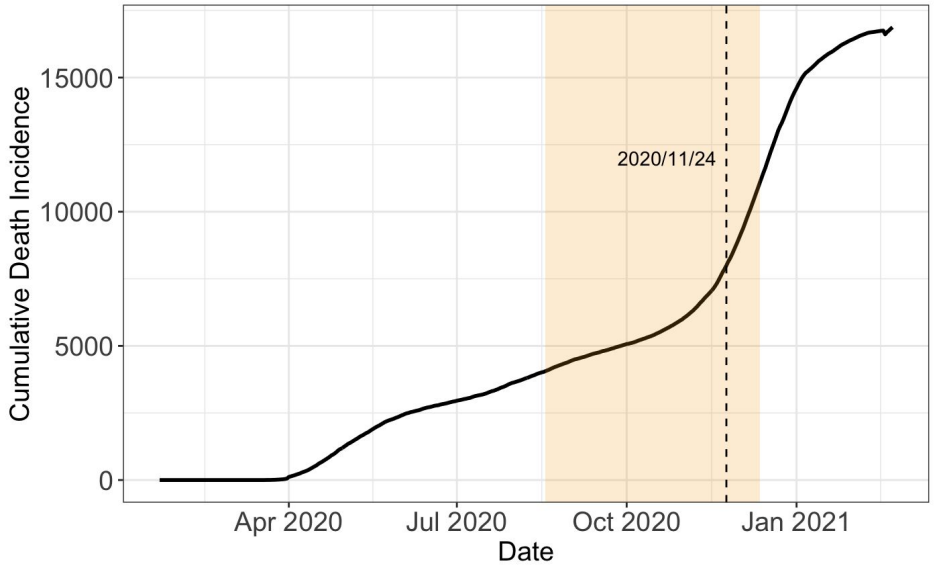
Thank you!

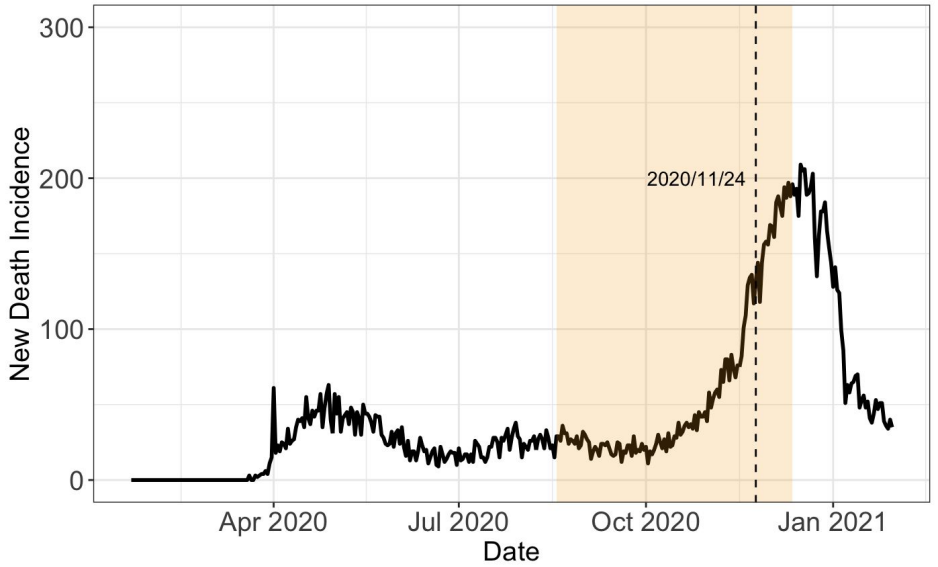
Q&A

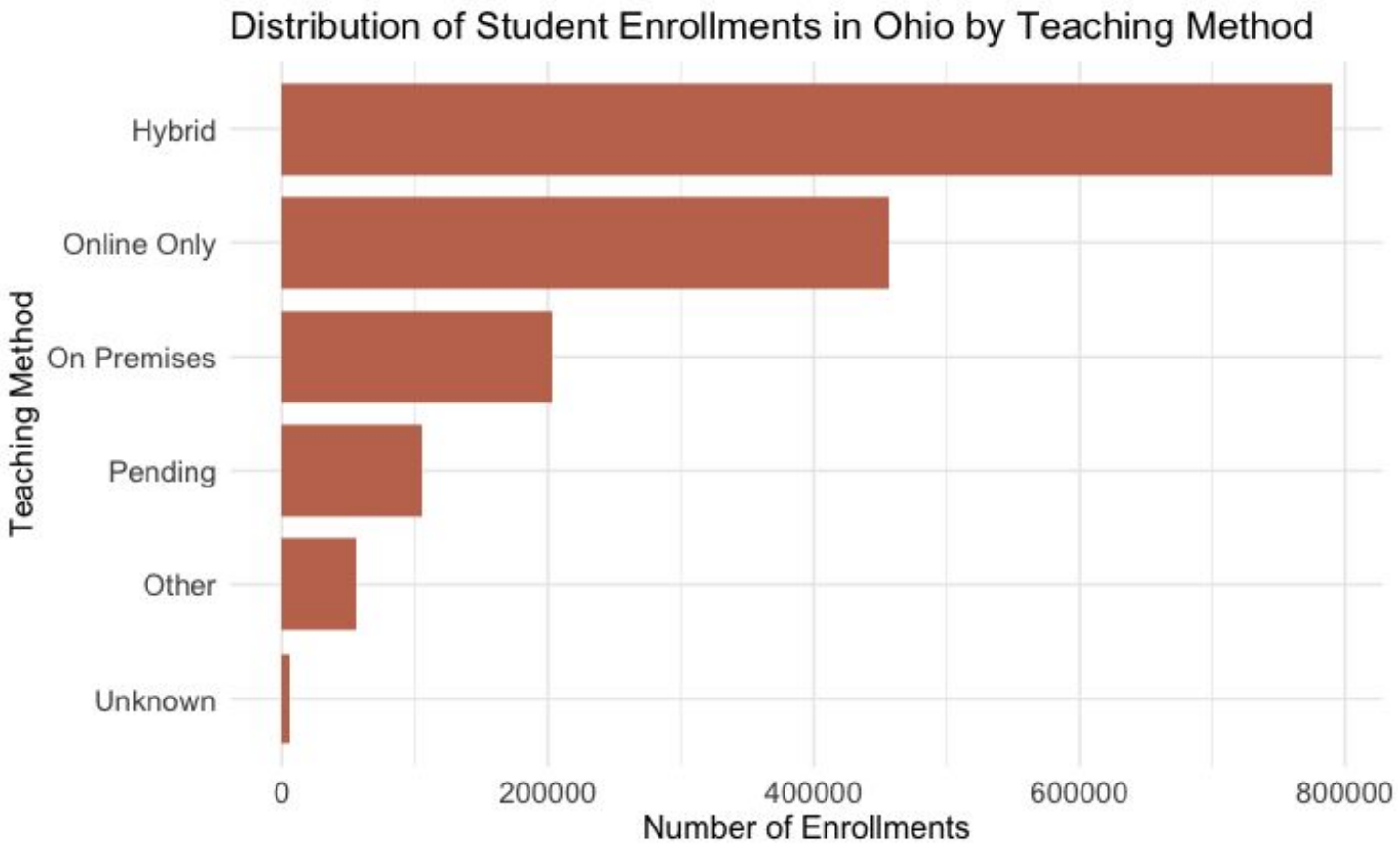
Carnegie Mellon University

Appendix

COVID Death Trend in Ohio State

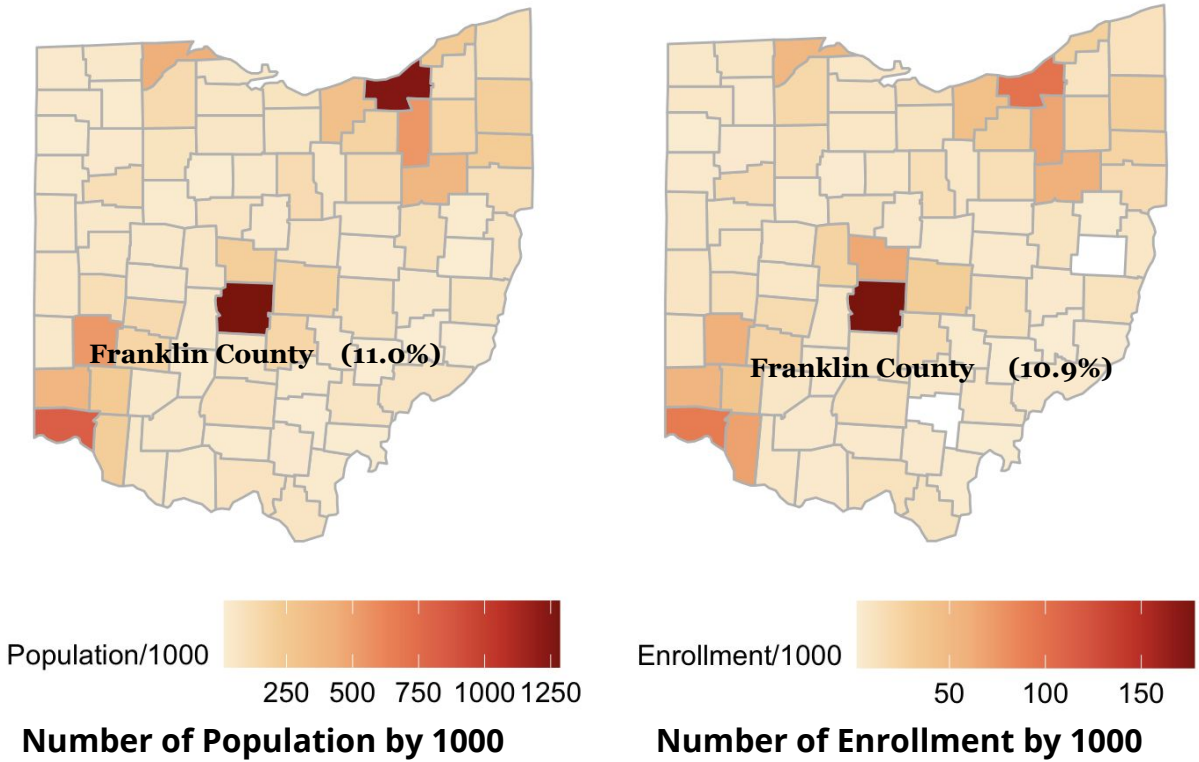





Distribution of Student Enrollments in Ohio by Teaching Method
Hybrid
Online Only
On Premises
Pending
Other
Unknown
Teaching Method
0
200000
400000
600000
800000
Number of Enrollments

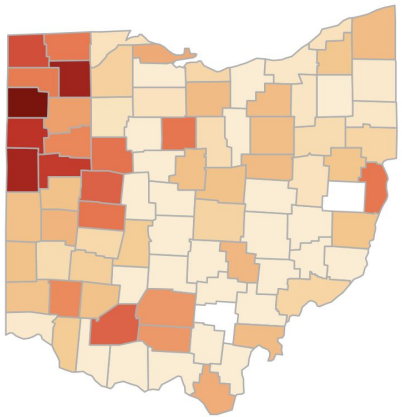
Summary Statistics

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

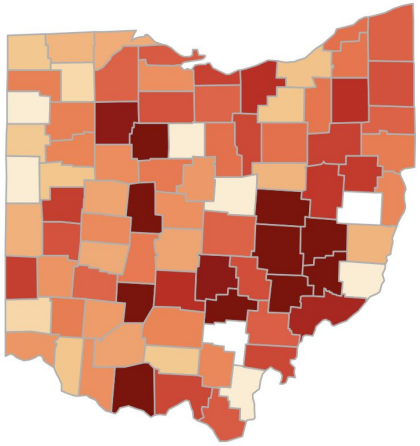


Number of Population by 1000

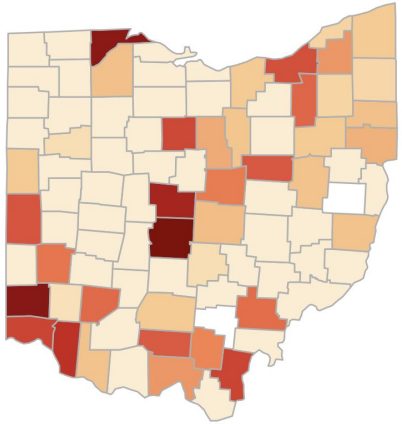
Number of Enrollment by 1000



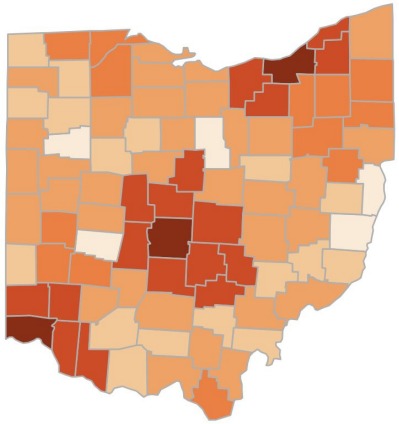
% On Premises
100
75
50
25
0



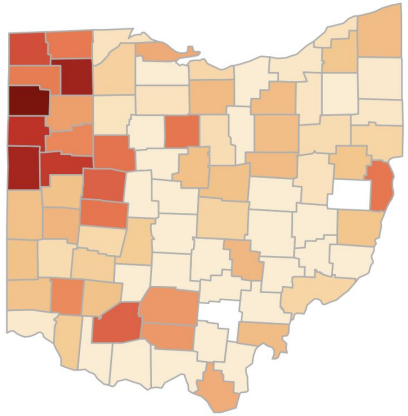
% Hybrid
100
75
50
25
0



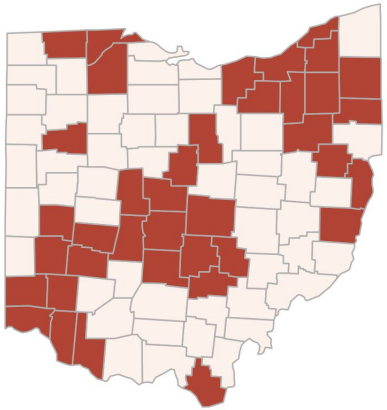
% Online Only
60
40
20
0



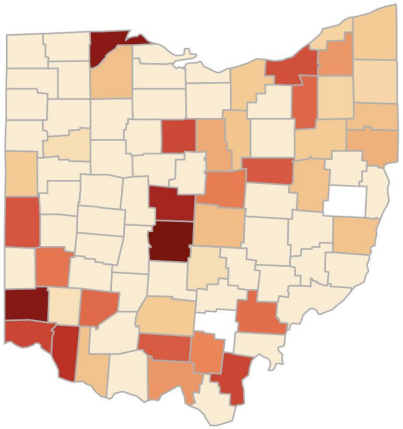
Urban Rural Status
Large central metro
Large fringe metro
Medium metro
Micropolitan
Noncore
Small metro



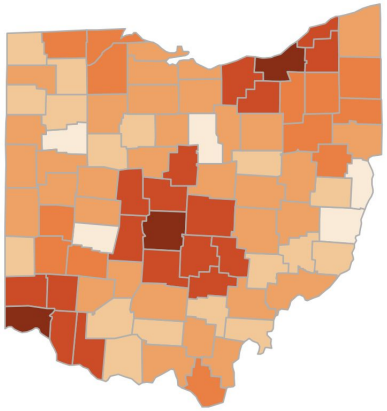
% On Premises
100
75
50
25
0



Metropolitan Status
Metro
Non-metro



% Online Only
60
40
20
0



Urban Rural Status
Large central metro
Large fringe metro
Medium metro
Micropolitan
Noncore
Small metro

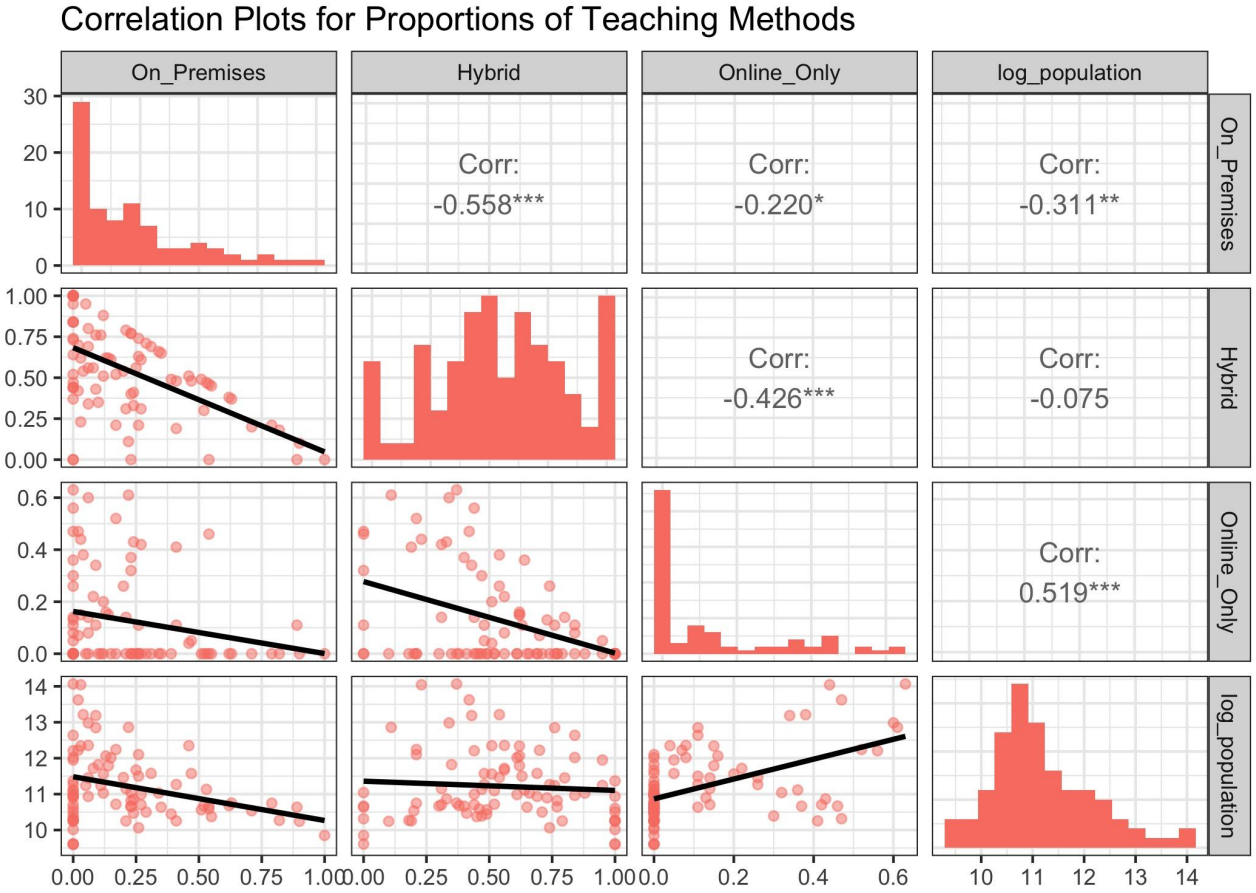
Method (more details on the model)

- **Are the county death trends significantly different by school teaching methods?**
  - Exploratory Data Analysis
  - Pair Test with control for confounder
- **How can we measure the transmission better? [Ongoing]**

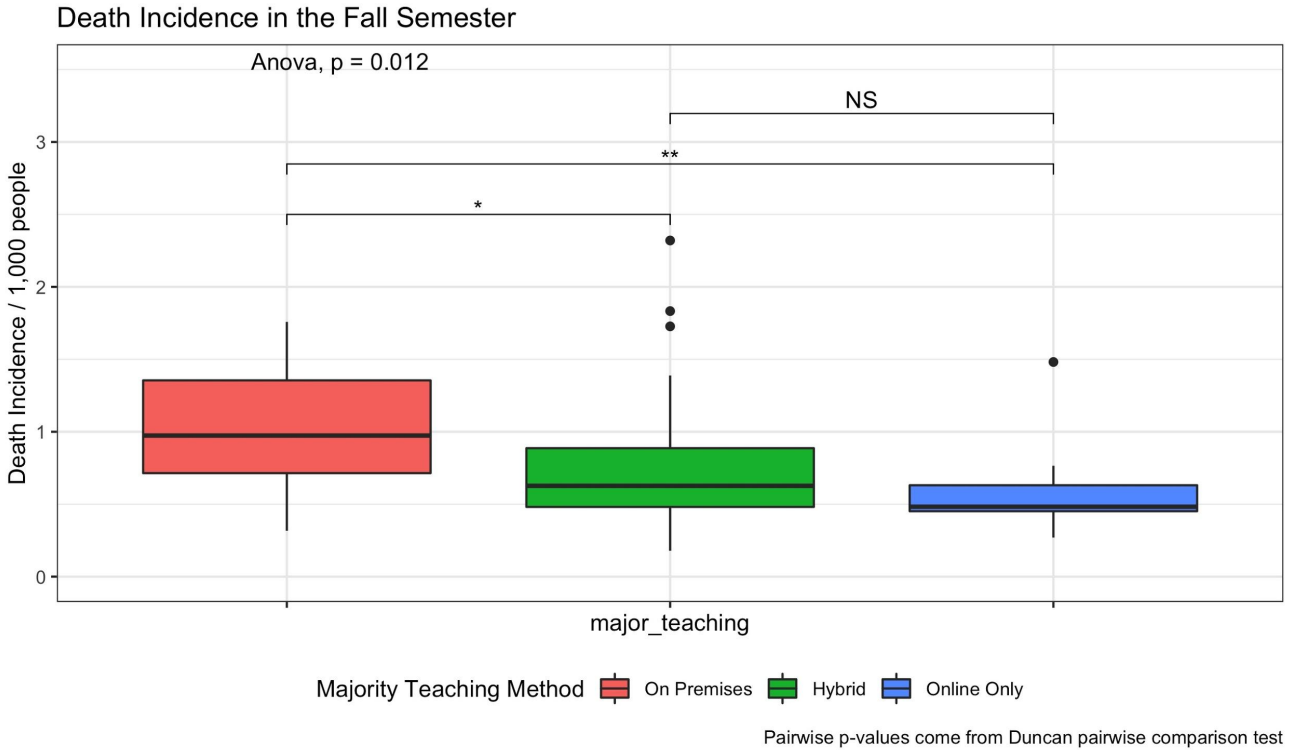
  Assume the disease follows an exponential growth and the process is stationary, then

  $$E(\log N_t) = N_0 + B*t$$

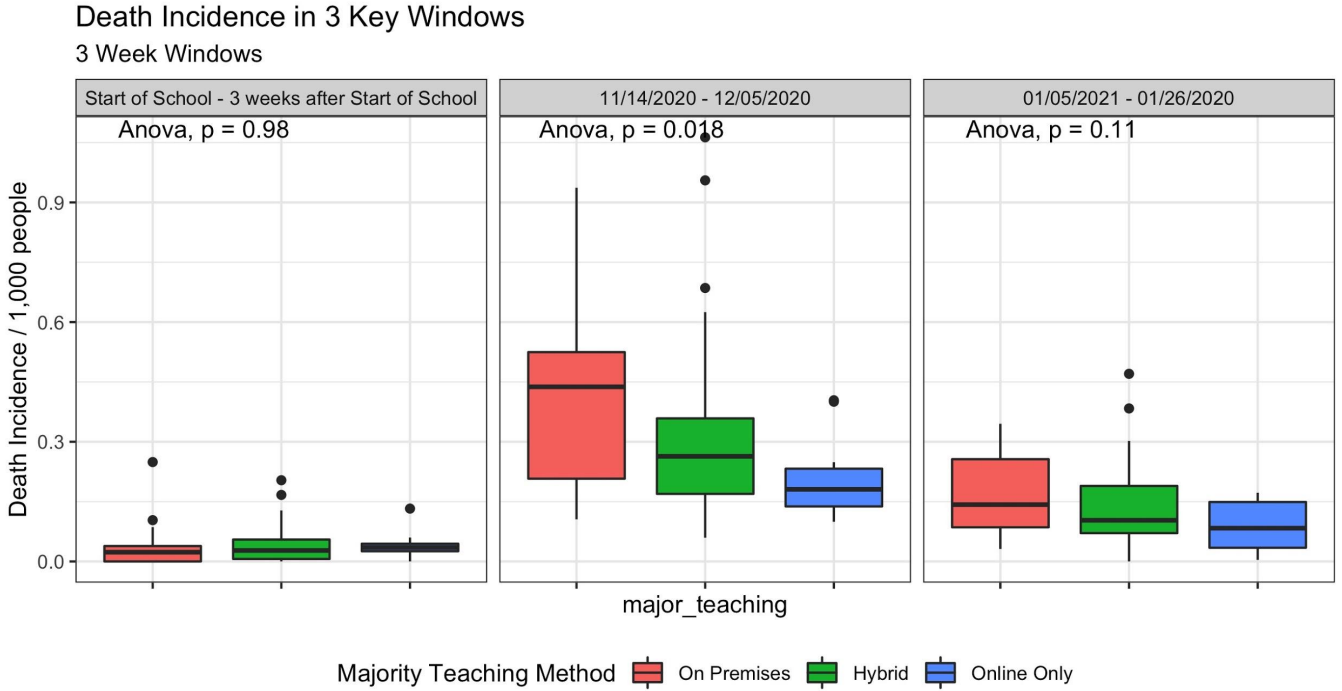
  N_t = number of new infections on day t, N_0 = log number of infections on day 0, B = exponential growth coefficient.
  - Step 1: Transform New Deaths: Log (New Deaths + 1)
  - Step 2: Smoothing Splines on Log (New Deaths + 1) against time t
  - Step 3: Compute the first derivative of Log (New Deaths + 1) to get death rate B
- **What factors might contribute to the change in exponential growth coefficient B?**


Correlation Plots for Proportions of Teaching Methods
On_Premises
Hybrid
Online_Only
log_population
Corr:
-0.558***
Corr:
-0.220*
Corr:
-0.311**
Corr:
-0.426***
Corr:
-0.075
Corr:
0.519***
On_Premises
Hybrid
Online_Only
log_population
30
20
10
0
1.00
0.75
0.50
0.25
0.00
0.6
0.4
0.2
0.0
14
13
12
11
10
0.00
0.25
0.50
0.75
1.00
0.0
0.2
0.4
0.6
10
11
12
13
14

Death numbers are different in In-person Counties

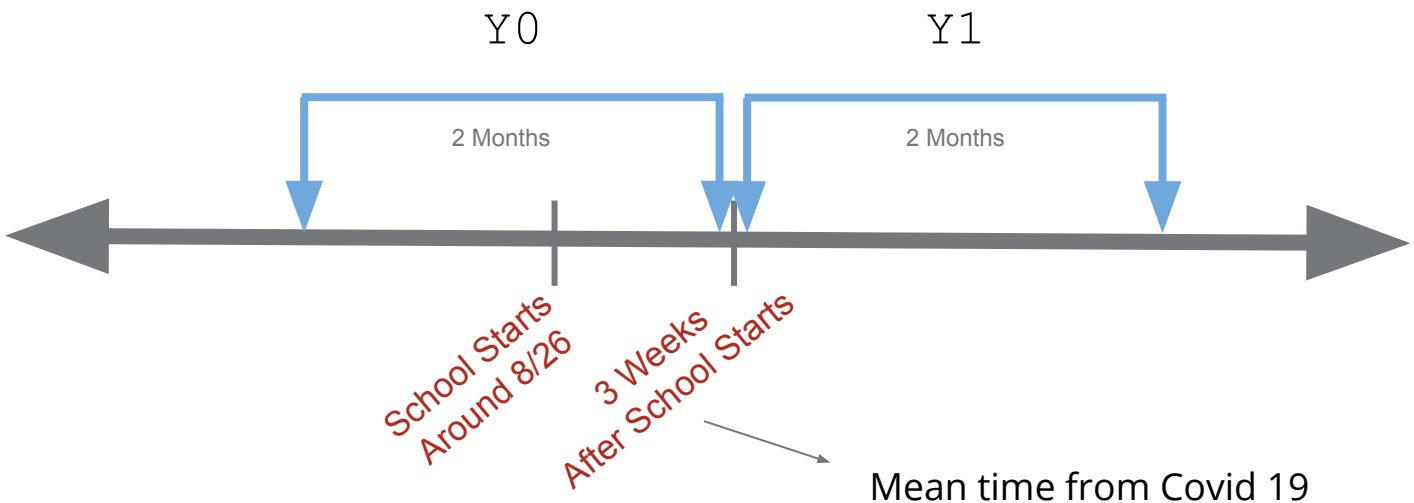


Death numbers increase faster for in-person counties



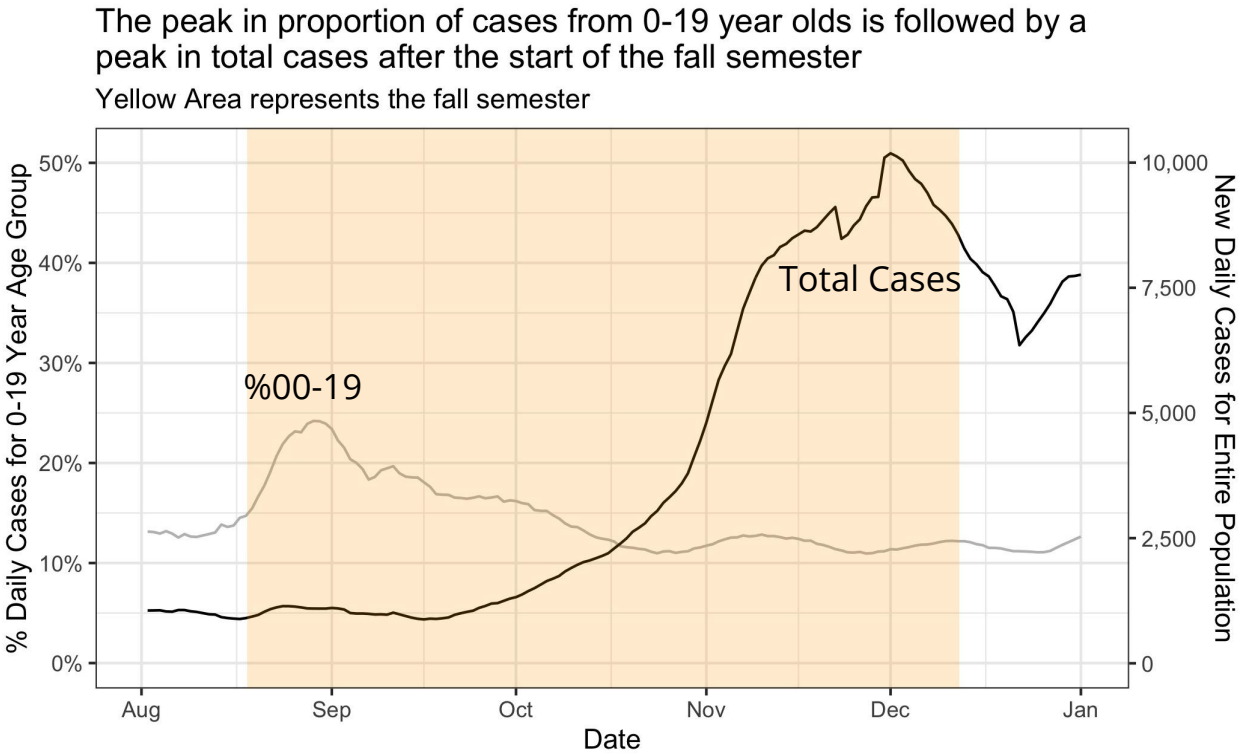
The first window looks at the 3 weeks after the start of school,
the second is the widndow around the point of intersection,
and the last window is 3 weeks after the end of the fall semester.

Time Designation for Death Calculation

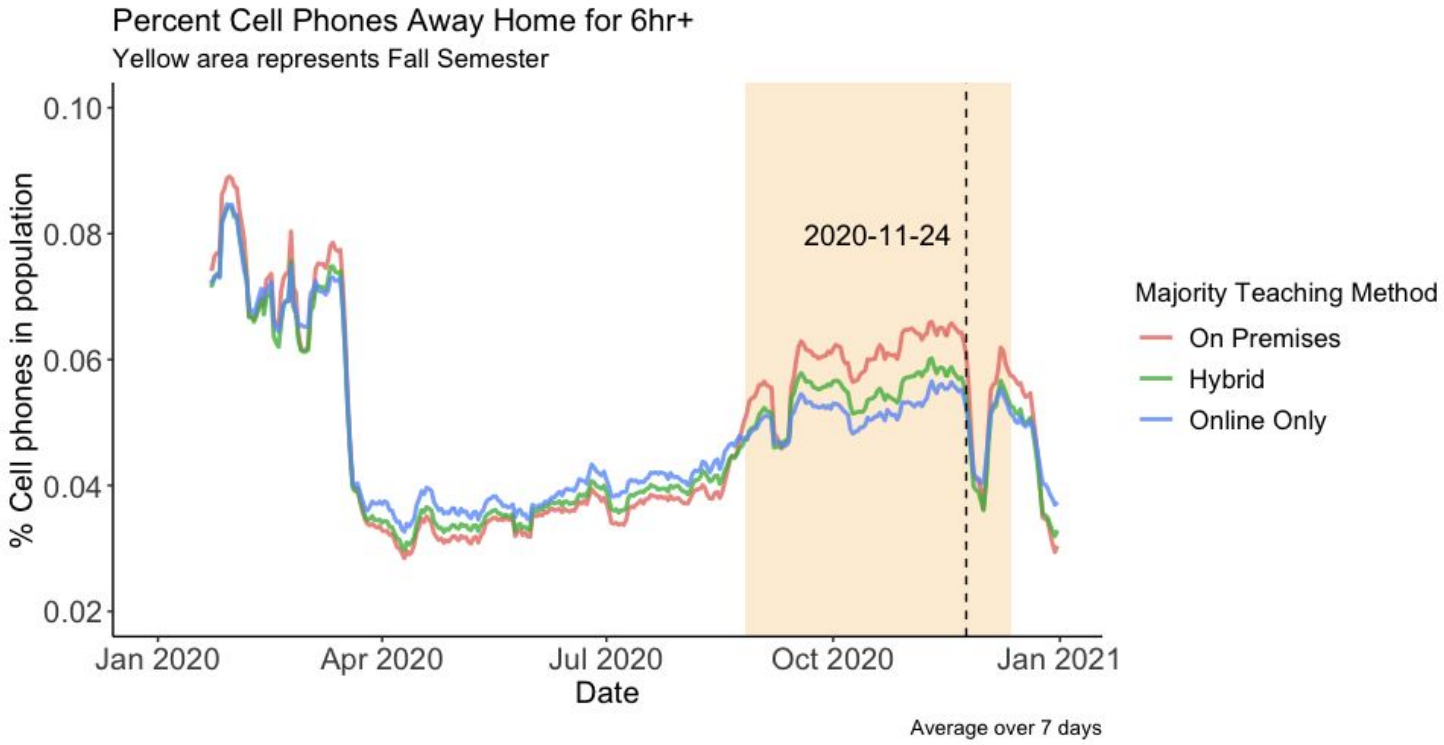


Mean time from Covid 19 infection to death = 3 weeks.

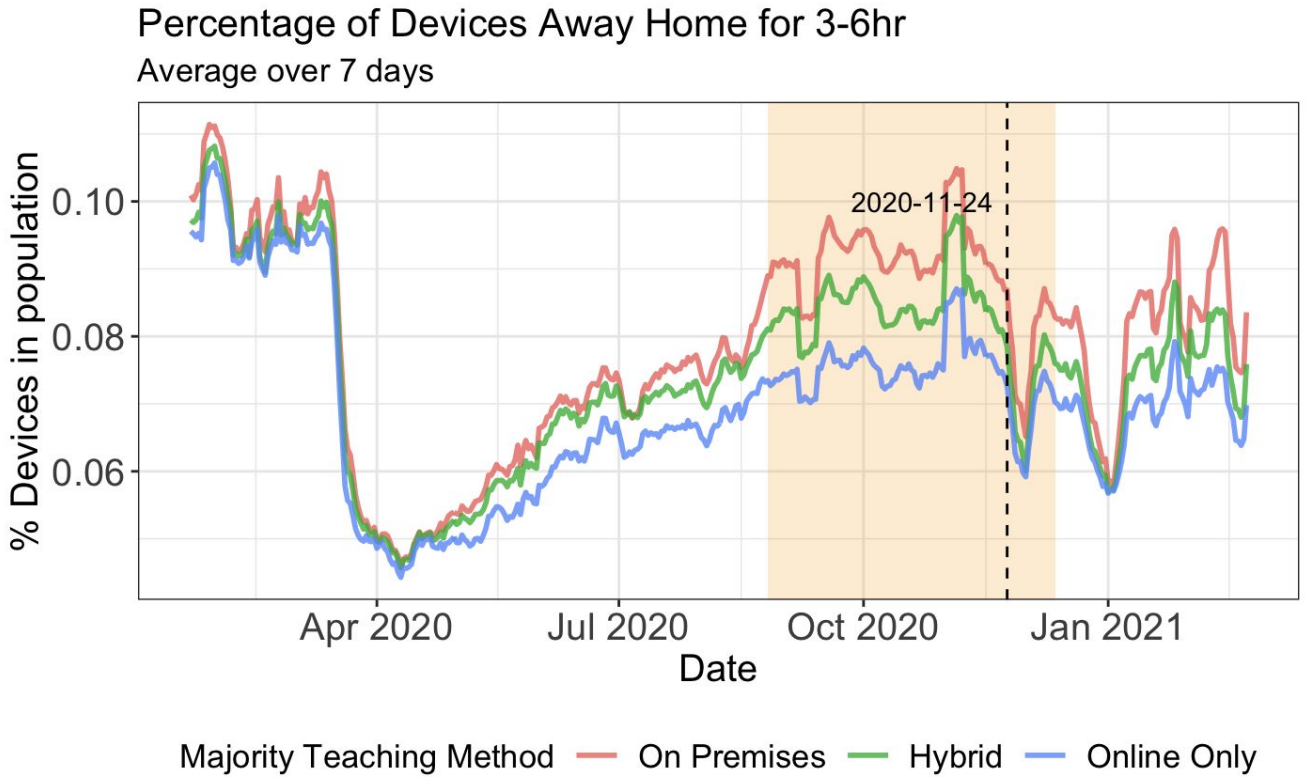
Student enrollments back to school



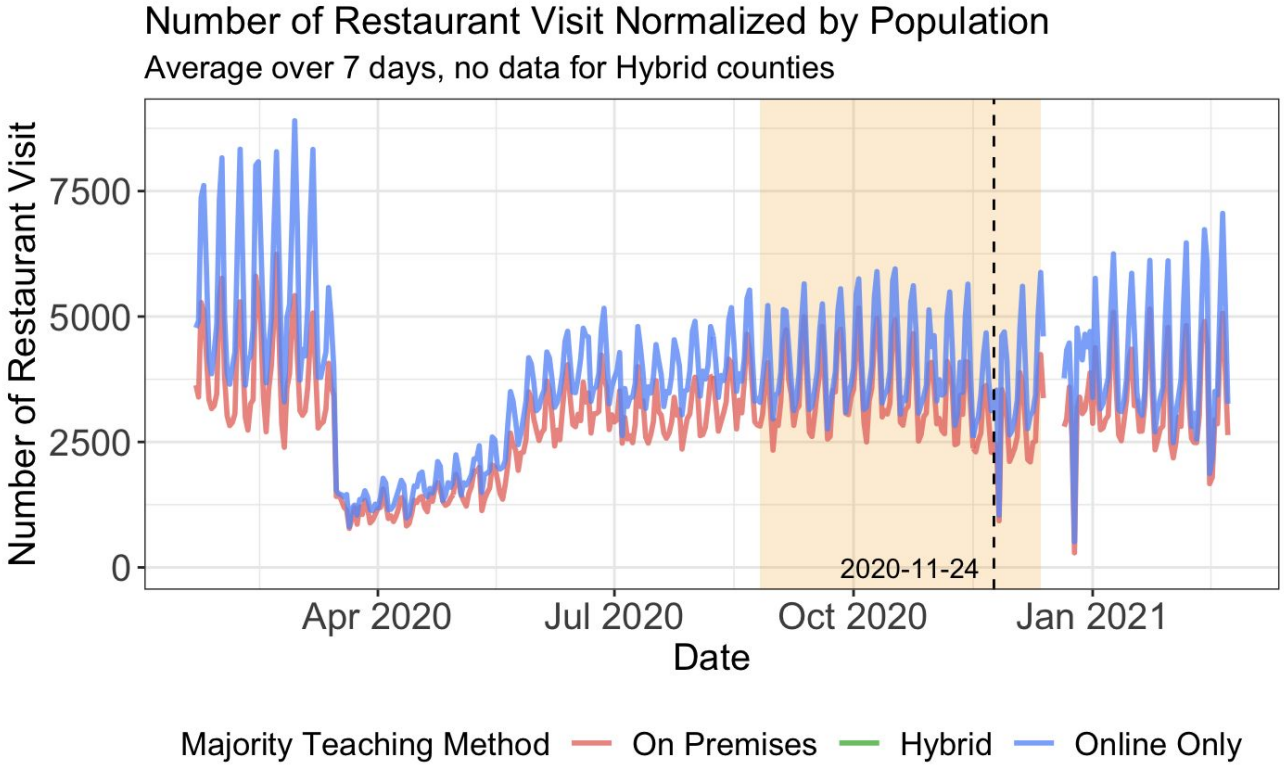
On premises counties have higher percent of cell phones away from home for 6 hours + in Fall



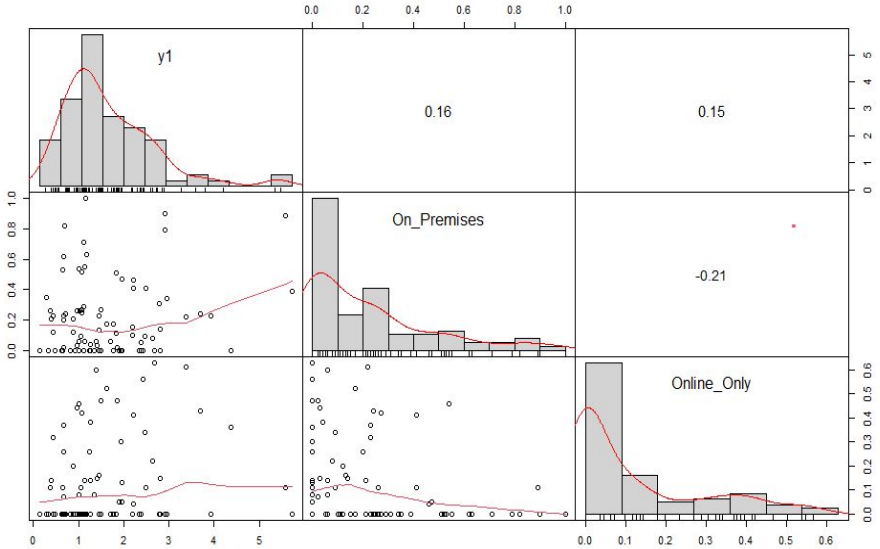
Part-time work -- different peaks?



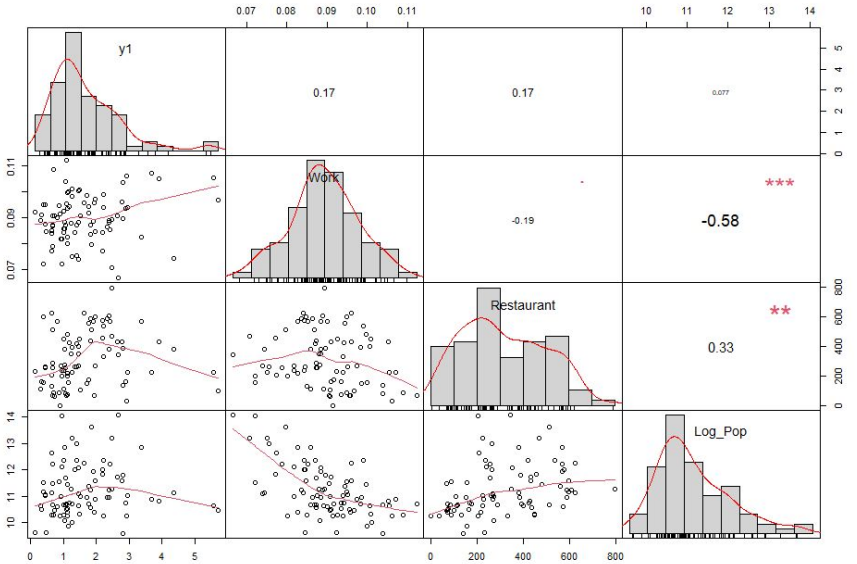
Counties majority going online have more restaurant visits



Pair Plots



Confounder vs Teaching Method



Y1 vs Confounders

Univariate Linear Regression

Variables of Interest

Confounders

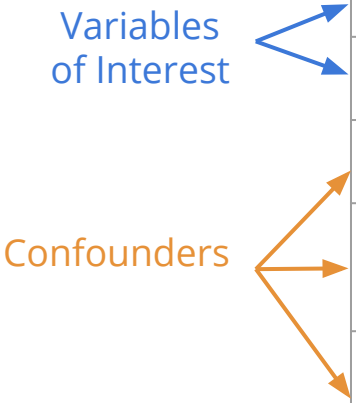
Relation	Significance
Y1 ~ On Premises	.004
Y1 ~ Online Only	.037
Y1 ~ Log Population	.004
Y1 ~ % Away home for 6hrs+	.001
Y1 ~ Average restaurant visits	.069

Insignificant Variables:

- Proportion of students do Hybrid
- Death incidence two months before school reopening
- Average Bar Visits

Schooling method is no longer significant after we adjust for possible confounders.

Y1 ~ On Premises + Online Only + Log Population + Avg. time at Work + Avg. restaurant visits



Relation	Significance
On Premises	.262
Online Only	.425
Log Population	.447
% Away home for 6hrs+	.281
Average restaurant visits	.439

We adjust for confounders by including them as explanatory variables in statistical analyses