Effect of School Teaching Posture on COVID-19 Transmission

By Cheyenne Ehman, Yixuan Luo, Zi Yang, and Ziyan Zhu

Department of Statistics and Data Science, Carnegie Mellon University
Master's of Statistical Practice
cehman@andrew.cmu.edu; yixuanlu@andrew.cmu.edu;
ziyang@andrew.cmu.edu; ziyanzhu@andrew.cmu.edu

May 17, 2021

Abstract

We are interested in the research question of whether the teaching posture has an impact on the transmission of COVID-19. To answer this question, we first use the visual comparison of the death numbers to capture the differences in the outcome of COVID by majority teaching posture and identify some potential confounders. We then develop an exponential growth model to measure the state of the disease and derive two summary metrics to reveal the transmission of COVID. One of them captures the severity of COVID during the entire fall semester and the other one captures the direct schooling effect at the beginning of the semester. Moving forward, we incorporate several confounders to analyze how they are impacting the relationship between the transmission of COVID and teaching posture. Meanwhile, we find that by blocking on micropolitan counties, we eliminate the effect of some confounders not limited to the urban setting, population, and geography. We thus continue our study with these micropolitan counties and we find out that counties with Online Only learning are having faster transmission of COVID than counties with On Premises learning. This suggests that there may be a schooling effect. However, we need more evidence from other States to come to a solid conclusion.

1 Introduction

The outbreak of SARS-CoV-2 led to the ongoing global pandemic of COVID-19, which was first identified in December of 2019. The virus quickly spread to the United States, with the first known incidence occurring on January 21, 2020. On March 13th, U.S President Donald Trump declared a national emergency. During the month of March, various counties and states started implementing public health interventions in an attempt to mitigate the spread of the virus. These interventions included stay-at-home orders, limited capacities at bars and restaurants, school closures, and eventually face mask mandates. However, cases continued to increase throughout the summer and into the 2020-2021 academic school year. Schools and school districts were left with the decision on how to proceed with the upcoming school year, implementing policies of their own regarding teaching posture.

When compared to cases of COVID-19 in adults, children represent fewer cases both in the United States and globally. Notably, hospitalization rates in children are significantly lower than the rates for adults. Despite evidence of children being less susceptible to COVID-19, the question still remains on whether or not children are potential vectors of transmission in the community, passing the virus to the school staff, faculty members, and their own families. This question is posed by Seema Lakdawala and the Lakdawala Lab at the University of Pittsburgh, who is working on the research project, Public Health Interventions aGainst Human-to-Human Transmission of COVID-19 (PHIGHT COVID).

The Master of Statistical Practice team at Carnegie Mellon University has partnered with the Lakdawala Lab to answer this question. Our analyses aim to discover whether or not children are acting as a vector of transmission by unpacking the relationship between K-12 school teaching posture and the spread of COVID-19. School districts across the country varied in their approaches of how to deliver instructions with some choosing to continue providing in-person classes, some choosing to carry out all instructions online in a virtual classroom setting, and others opted for more of a Hybrid approach of learning. These Hybrid approaches also varied greatly from rotating students to and from virtual classrooms to limiting class size requirements for in-person instructions. We believe that the mobility of children between 5 and 18 years old is highly subject to the teaching posture in K-12 schools during the shutdown. Consequently, the childrens' parents' mobility would also be limited by whether their children are staying at home. Thus, by looking at whether K-12 children are taking in-person classes, we would be able to investigate if the transmission of COVID-19 in a county is impacted by children's interactions. In doing so, we are also interested in the best way to measure and assess the difference in COVID transmissibility by teaching posture. With so many possible influential factors, we question whether or not our variable of interest, teaching posture is actually responsible for the effect on transmission.

We believe it is important and urgent to answer this question, since the relevant information would influence schools' decision on how to proceed with the fall 2021 semester and also aid the future decision-making on school operations during the pandemic, particularly in terms of teaching posture carried out in the K-12 schools.

1.1 A motivating example

With the vast variation in policies regarding the everyday lives of children, we question whether or not it is reasonable to look further into their roles in the spread of COVID-19. Below we see a motivating example for further analysis into the issue.

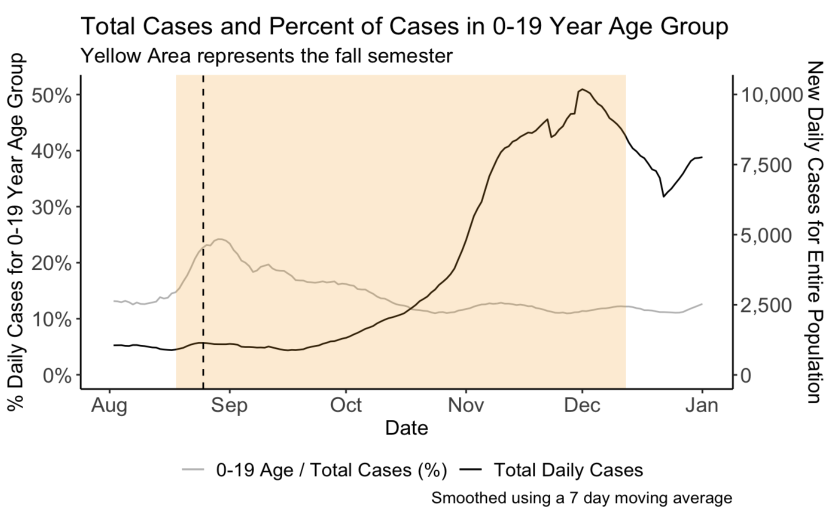


Figure 1: Proportion of Cases by Age

In Figure 1, we see the plot of two metrics in the duration of the fall semester (represented by the orange area) where teaching posture might be influencing transmission. The first metric denoted in gray represents the proportion of COVID-19 cases of people between the ages 0 to 19 in Ohio state. We see that approximately a week after the beginning of the semester there is a peak in the gray line, meaning between late August to late September (when school reopened), younger and school-aged people are representing more of the total cases in Ohio. The second metric denoted by the black line is the total number of cases in the entire Ohio state. Following the first peak, it shows that a larger peak in cases in the entire population occurred around the holiday season in the United States. This graph may suggest that school-aged children are transmitting the virus during the fall semester to the general population. While this evidence is striking, the trend may be due to many other unobservable confounding factors. However, it is enough to probe further analysis into the issue of whether or not children are spreading COVID-19.

2 Data

2.1 Motivation behind Using Ohio Data

Due to the potentially serious long-term effects of the COVID-19 disease and the high risk of dying in the senior population, we believe it would be unethical and unfeasible to even conduct a designed experiment for a causal inference study on children subjects. The available data we will be using in our analyses are purely observational, making it difficult to draw inferential conclusions on any of the relationships we find. To mitigate this, we have narrowed the scope of the analysis, focusing only on the counties in the state of Ohio. The goal was to simulate an experimental design, in which we would have control for confounding variables as much as possible. Ohio was chosen because the public health interventions implemented in the state mostly occurred at the state level and fewer at the county level. This means that the counties in Ohio are similar in their governmental policies yet differ in their school-regulated policies, such as teaching posture, making them potentially comparable to one another for the purposes of this analysis.

2.2 Deaths over Cases

In this analysis, we use deaths instead of case data to extract information on COVID-19 transmission. The intention is to avoid the systematic bias in the case records: case data may be reported only once in a while; thus, peaks could be due to aggregations; an increase in cases could also be explained by the test volume and the test availability in certain counties. We should be aware that deaths could also be biased due to reporting issues, age distribution, county's hospital admission capacity, etc. Overall, it is relatively more feasible to adjust for the bias in death numbers given our data availability because the reporting of deaths is much more accurate than cases. However, we can derive a way mathematically to use death incidence as an approximation of infections and the state of the disease at a given time. More details will be given in the Method section.

2.3 Data Description

Our analysis is based on four datasets, which describe the school's COVID-19 intervention by school districts in Ohio state (Ohio K12 data) [1], county-level daily deaths and cases from COVID-19 in Ohio State (John Hopkins Open-Source Tracking Data)[2], county-level population mobility (SafeGraph data from COVID-CAST API created by CMI DELPHI group)[3], and county-level demographic information of Ohio State (Ohio profile data from CDC)[4]. The datasets are linked by the county names and date of the records in the manners shown in Figure 2.

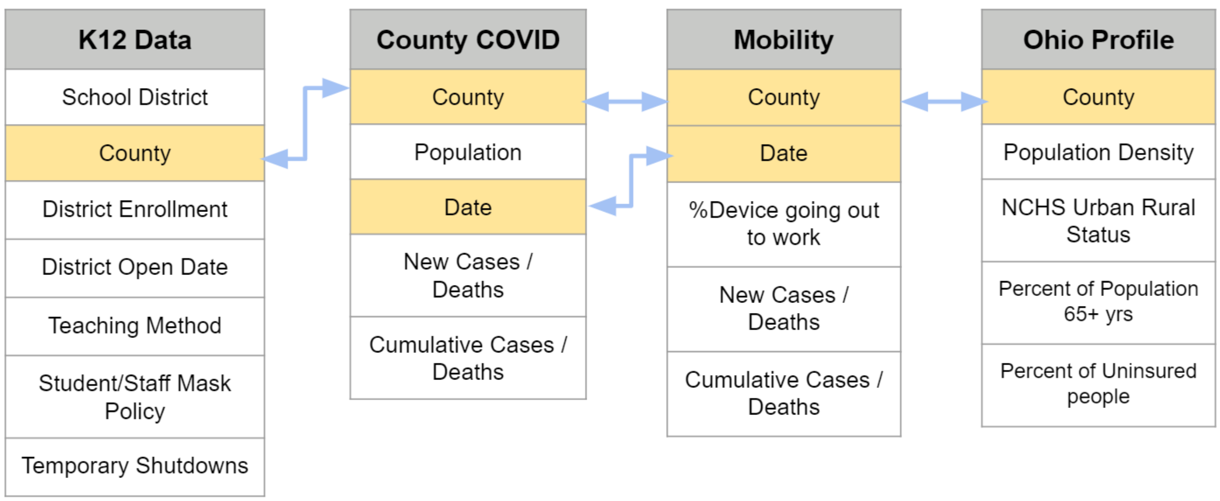


Figure 2: Data Linkage

The Ohio K12 Data consists of 35 variables and 2786 observations. The County COVID data and Mobility data have 14 variables and 9 variables respectively, and all have 35024 observations. The Ohio Profile data gives 21 variables and one observation for each county in Ohio. The time-series data of COVID starts from January 22nd, 2020, to February 22nd, 2021. We drop Harrison and Vinton County, whose school status information is missing; we then screen the variable of interest relevant to teaching posture. The data used for our analysis are described in Table 1, 2, 3 and 4. It is also important to note that the data for school policies is at the individual school level. However, we know that there can be multiple schools in a school district and multiple school districts in a county. Because of this, we eventually aggregate information about the schools to the county level. The one variable we are most interested in is "Majority teaching posture." This variable characterizes each of the counties in Ohio by the teaching posture that the majority of students in the county are following.

---

[1] MCHdata.com
[2] https://coronavirus.jhu.edu/covid-19-daily-video
[3] COVIDcast | DELPHI - CMU Delphi
[4] https://data.cdc.gov

Table 1: Ohio K12 (Dataset 1)

Variables	Values	Descriptions
School Name, School District Code, Physical City, Physical County	e.g. North High School, 3904348,AKRON, SUMMIT	Name of the school and its school district, also the city/county it physically located in
Student enrollment by school district	e.g. 914	The number of students enrolled in the certain school district
School opening date by district	e.g. 9/9/2020	Fall semester start date for the school districts in 2020 Fall. Notes that some districts reopen in late October
Teaching posture	Online Only, Hybrid, On Premises, Pending, Other	School teaching posture conducted in the 2020 Fall semester
Temporary School Shutdowns	Close 1-5 days; close 6-14 days; never closed ; unknown	How many days the school had temporarily shut down during the 2020 Fall

Table 2: Ohio Cases & Deaths and Demographic information (Dataset 2)

Variables	Values	Description
County	e.g. ADAMS	Physical county where the cases or deaths occured
Date	Range from 2020-01-22 to 2021-02-22	The date when data was reported
NewDeaths	e.g. 4	New deaths in the county on the reporting date
CumDeaths	e.g. 482	Cumulative deaths in the county on the reporting date
Population	e.g. 1256007	The total population of the county

Table 3: Ohio Cell Phone Mobility Data (Dataset 3)

Variables	Values	Description
County	e.g. ADAMS	County name for the data reported
Date	Range from 2020-01-22 to 2021-02-22	The date when data was reported
full_work_prop_7d	e.g. 0.0643	The fraction of mobile devices that spent more than 6 hours at a location other than their home during the daytime; ought to suggest mobility of full-time workers/students

Table 4: Ohio Profile (Dataset 4)

Variables	Values	Description
County	e.g. ADAMS	County name for the data reported
Population.density	e.g. 47.44	Number of people per square mile
NCHS.Urban.Rural.Status	Noncore, small metro, Micropolitan, Large fringe metro, Medium metro, and Large central metro	A six-level urban-rural classification scheme developed by NCHS
Percent of Uninsured people	e.g. 8.7%	The percentage of people do not have insurance in the county
Percent of 65+ years old population	e.g. 18.46%	The percentage of population older than 65 years old in the county

2.4 Data Cleaning and Wrangling

Most of the missing values occurred in the Ohio K12 dataset; for example, 4 schools did not report their COVID-19 public health intervention policies by the time when data was collected; some schools do not have county information; some schools have an empty value, or an incorrect number of students enrolled. Since only 4 out of 2876 schools do not have public intervention information that we are interested in recorded, we simply drop those 4 rows of data. In addition, we were able to remove duplicated columns and impute missing counties with the city information, remove COVID cases observations with missing values in cases & deaths, and drop missing values case by case during EDA.

As part of the data cleaning process, we also notice that there were occasionally negative new deaths reported. We believe this is due to the adjustment for over-reporting of deaths in previous days. Since we will take log transformation of the new deaths data, we manually correct the negative entries by distributing the negative values equally to the positive new deaths' records in the previous week.

Using the cleaned data, we aggregate the original data to extract county-level information. We scale the cumulative deaths by the population in the counties, and then calculate the proportion of students either taking On Premises classes, Online Only classes or Hybrid classes. Because we only have county-level death incidence, we are interested in the schooling effect in a county. We then assign the majority teaching posture to each county by its most popular teaching posture. The detailed aggregation rules are shown in Table 5.

Table 5: Aggregation Rules

Generated Variables	Aggregation Rules
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 posture	Teaching posture in county with highest proportion

2.5 Summary Information of Ohio

In this section, we will show the demographics information of the counties in Ohio state.

Population

Figure 3 shows the distribution of population across all counties in Ohio adjusted by 1000. We identified that the most populated counties are Franklin and Cuyahoga marked with the darkest colors.

Proportion of Teaching Posture

Figure 4 displays the proportion of schools implementing different teaching posture for each county, namely On Premises, Hybrid and Online Only. We can see that there is a wide range of school teaching posture across counties which validates our objective in studying their effect on COVID-19 transmission. If we compare the three graphs, we see that Hybrid counties in Figure 4 (a) generally have darker colors meaning the proportion of schools conducting Hybrid learning is the highest.

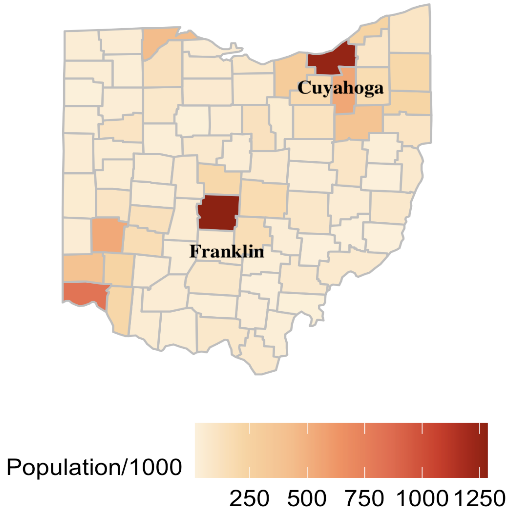
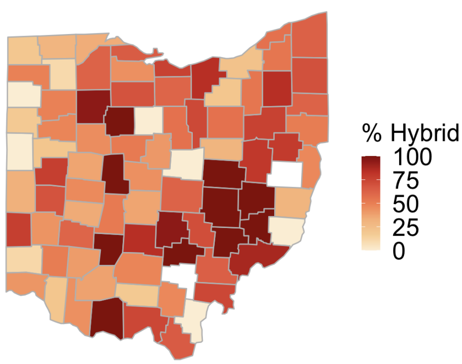
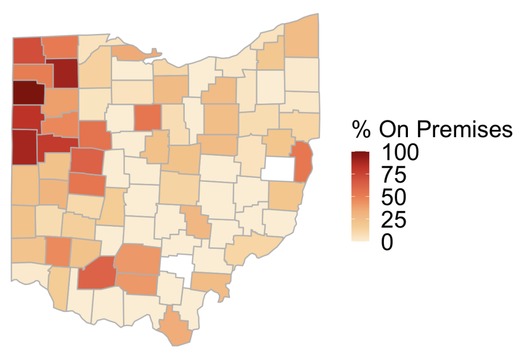


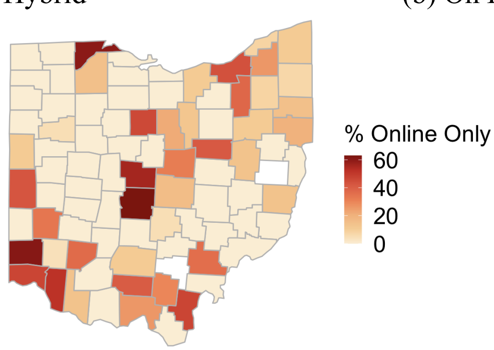
Figure 3. Number of Population by 1000



(a) Hybrid



(b) On Premises



(c) Online Only

Figure 4: Proportion of student enrollments by teaching posture

3 Methods

In investigating whether school teaching posture has an effect on the COVID-19 transmission, we can visually compare the death numbers and transmission metrics depicting the state of the pandemic among groups to capture any differences between counties that adopted Online Only and those adopted On Premises teaching posture. More specifically, we first visualize geographical information using maps; we then use the time series plots and boxplots to summarize the aggregated information at the county level grouped by the three types of teaching posture. We plot the death incidence per 1000 people as a function of time for each of the three types of teaching posture. The trend will allow us to see the trajectory of the disease in each of the types of counties. While this aggregated trend line will show us the trajectory for the types of counties, we also want to look at the spread of death incidence amongst the different types of counties specifically during the fall semester. This can be done with the use of boxplots for data in the fall semester. We initially ran the ANOVA test on the null hypothesis that the death incidence per 1000 people is equal for all three types of teaching posture. After getting a significant result, indicating that the death incidences are different in the three groups, we perform a Duncan multiple range test to compare the means of the death incidences by teaching posture.

Finally, we investigate the individual counties with scatter plots to visualize the relationship between the state of pandemic and school teaching posture while blocking for possible confounding variables that interfere with our interpretation of the school posture effect. In the rest of section 3.1, we want to compute the exponential growth coefficient with the infection model and implied deaths model to summarize the state of the disease at a given time. We dive into two characteristics of the state of pandemic to measure the impact of school teaching posture.

The first transmission metric we will look into is the maximum exponential growth coefficient between 08/18/2020 to 12/15/2020, which corresponds to the COVID-19 transmission severity in the 2020 Fall semester. We also want to understand if limiting in-person components in schools would affect the change in growth coefficients at the early stage of school reopening, thus the second metric we examine is the difference in exponential growth coefficients at the first and second 3 weeks of school reopening, which correspond to how fast the disease expand in a small time window. By calculating the change in growth over a short time window, we would also be able to control for unmeasured confounding variables by assuming these factors remain constant. For example, we can assume people's compliance with mask-wearing and social distancing did not change over 3 weeks; or the mobility and public interventions in the act stay the same over short periods.

We are interested in if the school teaching posture is significantly correlated with the severity of the pandemic and how fast the disease expands over the Fall semester. However, because this study is retrospective and the observations are not from a designed experiment, we will block the data by several possible confounding or lurking variables that could also be attributed to the differences we see, for example, population density and urban-rural status of the county. By looking at these other variables, we will be able to block some confounding variables to find the most comparable set of counties to work with and determine the relationship between teaching posture and transmission.

3.1 Measurement of State of Pandemic

One major challenge we encountered in this analysis is to extract information of the state of COVID-19 transmission from the daily death numbers. To capture the state of the pandemic, we use an exponential growth model to locally measure the infection size at time t given preceding infections; then we use probability of deaths from COVID-19 to link the deaths number with infections; eventually, we describe the distribution of time from infections to deaths of COVID-19. By looking at the instantaneous growth rate, we would be able to capture the changes in the disease dynamics better than observed cumulative death incidence. We will show how we arrive at the estimation of the growth coefficients in the below section.[5] [6]

3.1.1 Exponential Growth Model of Infections

We assume that the infection number follows an exponential growth:

$$I_t = I_{t-1}e^{B_t} + \varepsilon_t \qquad (1)$$
$$E[I_t] = I_1 e^{\sum_{r=1}^{t} B_t}$$

where I_t is the new infections in day t and ε_t is the random error with zero mean, and B is the exponential growth rate. Note that B should vary by county and by time due to mask-wearing willingness, public interventions, etc. B summarizes the state of pandemic: a positive B means that infections at time t is larger than infections at time t-1, thus the disease is expanding; similarly, a negative B means the disease is shrinking. I_t is not observed in our data, so we carry out a recursive procedure of first equation in (1) to find an approximation of I_t using the infections at the beginning of the pandemic, as shown in the second equation in (1).

3.1.2 Implied Death Model

As deaths is a consequence of infections, we bridge the time lag between death in day t and infection in day s by following model:

$$D_t = \sum_{s=1}^{t} f(s,t) I_s + \xi_t \qquad (2)$$

where D_t is the new deaths in day t, ξ_t is the random noise with zero mean, $f(s,t)$ denotes the probability that someone infected with COVID-19 on day s will die on day t.

Let $d(s)$ denotes the probability that someone infected on day s will eventually die from the disease and $f_0(s,t)$ denotes the density function of the distribution of time from infection to deaths. We write

$$f(s,t) = d(s)\, f_0(s,t) \qquad (3)$$

[5] Ventura, Valerie. (2021). PHIGHT notes.

[6] Bonvini, M., Kennedy, E.H., Ventura, V., Wasserman, L.. (2021) Causal inference in the time of COVID-19. [*Preprint*]. Mar 7, 2021. Available from: https://arxiv.org/abs/2103.04472

From established study by Unwin, H.J.T., Mishra, S., Bradley, V.C. et al. in 2020[7], we know that the distribution of time from infection to deaths can be approximated by a Gamma distribution with mean 23.9 days and coefficient of variation 0.40.

By plugging in the recursion of I_t in Equation (1), we arrive at

$$\mathbb{E}[D_t] = \sum_{s=1}^{t} f(s,t) I_1\, e^{\sum_{r=1}^{s} B_r} \tag{4}$$

Since we are mostly interested in deriving the exponential growth coefficient B from the death model, we take $f_0(s,t)$ in Equation (3) to be a point mass probability at $t - \delta\ =\ 24$days, which means we assume all patients infected will die exactly after 24 days (average time from infections to deaths). Then we get

$$\mathbb{E}[D_t] \approx I_1 d(t-\delta) e^{\sum_{r=1}^{t-\delta} B_r} \tag{5}$$

If we approximate $log(\mathbb{E}[D_t])$ with $\mathbb{E}[log(D_t)]$ we further obtain

$$\mathbb{E}[L_t] =\ log(d(t-\delta)) + log I_1 + \sum_{r=1}^{t-\delta} B_r = v(t) \tag{6}$$

Where $L_t = log(D_t + 1)$, note that we add 1 to the new deaths to avoid taking log on zero new death; or equivalently,

$$\mathbb{E}[L_{t+\delta}] =\ log(d(t-\delta)) + log I_1 + \sum_{r=1}^{t-\delta} B_r = v(t+\delta) \tag{7}$$

Note that if the exponential growth rate is constant where $B_r = B$, then $v(t+\delta)$ with derivative

$$\partial\mathbb{E}[L_{t+\delta}]/\partial t =\ d'(t)/d(t) + B \tag{8}$$

Hence if B is not constant,

$$\partial\mathbb{E}[L_{t+\delta}]/\partial t =\ d'(t)/d(t) + B_t \tag{9}$$

where $B_t = B(t)$ is the exponential growth of the epidemic at time t.

Note that the probability of dying $d(t)$ is allowed to change smoothly over time, which could be subject to the hospital capacity and treatment development. But as we will look at a short window of time, we can assume the probability of dying from COVID-19 is constant, then

$$\partial\mathbb{E}[L_{t+\delta}]/\partial t = B(t) \tag{10}$$

Following Equation (10), we should be able to summarize the state of pandemic on day t by taking the derivative of log new deaths on day t+δ to compute the exponential growth coefficient B at time t.

The cumulative death count at time t+$\delta$is

$$\mathbb{E}[C_{\,t+\delta}] = \sum_{r=1}^{t+\delta} \mathbb{E}[D_t]\ = I_1 \sum_{r=1}^{t+\delta} d(s-\delta)\, e^{\sum_{r=1}^{t-\delta} B(r)} \tag{11}$$

[7] Unwin, H. J. T., Mishra, S., Bradley, V. C., Gandy, A., Mellan, T. A., Coupland, H., Ish-Horowicz, J., Vollmer, M. A., Whittaker, C., Filippi, S. L. et al. (2020). State-level tracking of COVID-19 in the United States. *Nature communications* **11** 1–9.

and we assume the probability of dying from COVID-19 on day $s - \delta$ to be constant, then

$$\mathbb{E}[C_{t+\delta}] = I_1 \sum_{r=1}^{t+\delta} d\, e^{\sum_{r=1}^{t-\delta} B(r)} \tag{12}$$

4 Results

By looking at the majority teaching posture for each county in Ohio, we can then compare the different counties by their transmission rates of COVID-19. We will examine transmission by looking at cumulative death incidence and the exponential growth coefficient described in previous sections; both metrics will help us discover if teaching posture is associated with the differences in transmission between different counties.

4.1 Majority Teaching Posture Vary across the Ohio State

As mentioned above, we have aggregated the teaching posture from the school level to the county level. Figure 5 below therefore shows the majority teaching posture we assigned for each county also in the geographical map format in Ohio. Note that Harrison and Vinton counties did not report their K-12 school teaching policies and thus they are not colored on this map. We can see that the On Premises counties are mainly located in the upper left corner of Ohio, while the Online Only counties scatter across the state; this mirrors our observations on the teaching posture proportions in Figure 4.

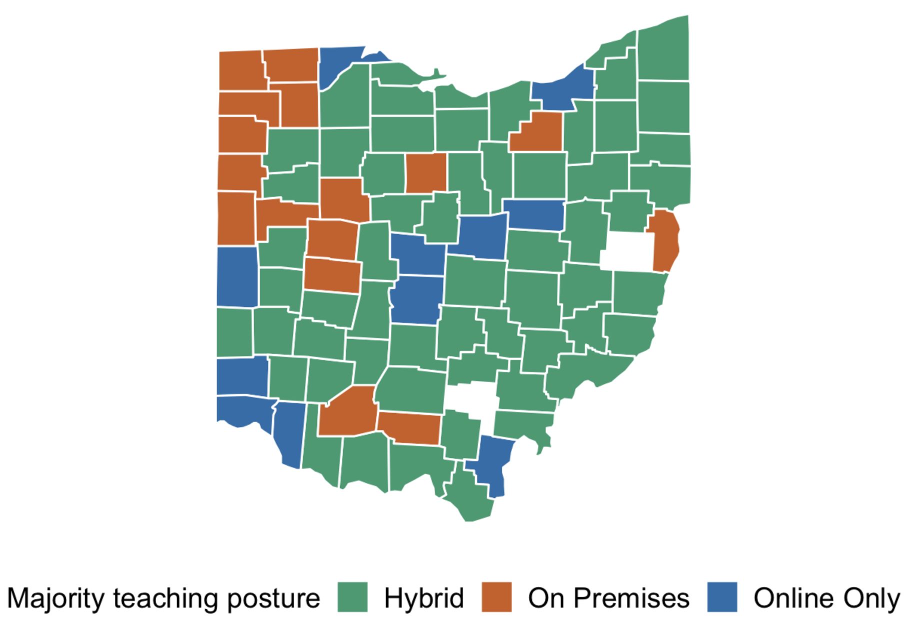
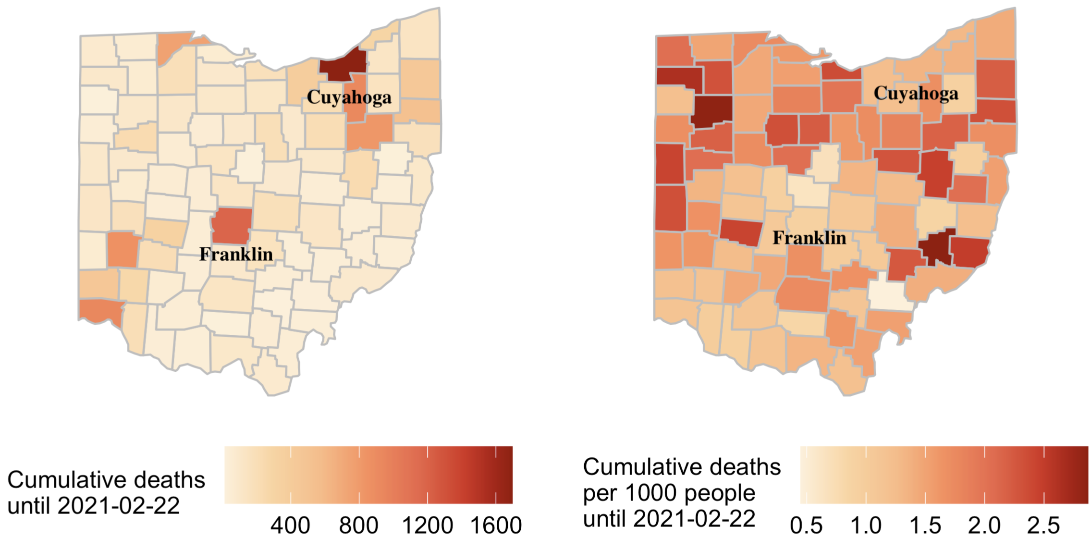


Figure 5: Majority teaching posture

We start by using the majority teaching posture as the main explanatory variable that represents the school policy for each county to analyze its effect on COVID-19 transmission. However, we are aware that there will be potential issues regarding the loss of information while aggregating the data and we will elaborate on this in the Discussion section.

4.2 Higher death proportions in Northwest and Southeast Ohio



(a) Cumulative deaths (b) Death incidence

Figure 6: Covid-19 cumulative deaths until 02/22/2021

Figure 6 shows the death condition in different counties. Compared to the population map in Figure 2, we find that counties with large populations often have larger numbers of cumulative deaths as well. However, large numbers of cumulative deaths do not necessarily mean high death incidences. For instance, Franklin County, which has the most population and the second most cumulative deaths, has a much lower death incidence. According to the map, counties in the northwestern region of Ohio have relatively higher death incidence.

4.3 Death Incidence Increases Faster in On Premises Counties

We wished to study the schooling effect on COVID-19 transmission. Therefore, below we have plotted the cumulative death incidence as a function of time by teaching posture in Figure 7. In the plot, the yellow area represents the duration of time in which the fall semester would take place. This is approximately between 08/18/2020 to 12/15/2020. In this way, we can observe the speed of COVID-19 transmission reflected by death numbers. By comparing the different colored lines, we wanted to know if there are differences among the three types of teaching posture.

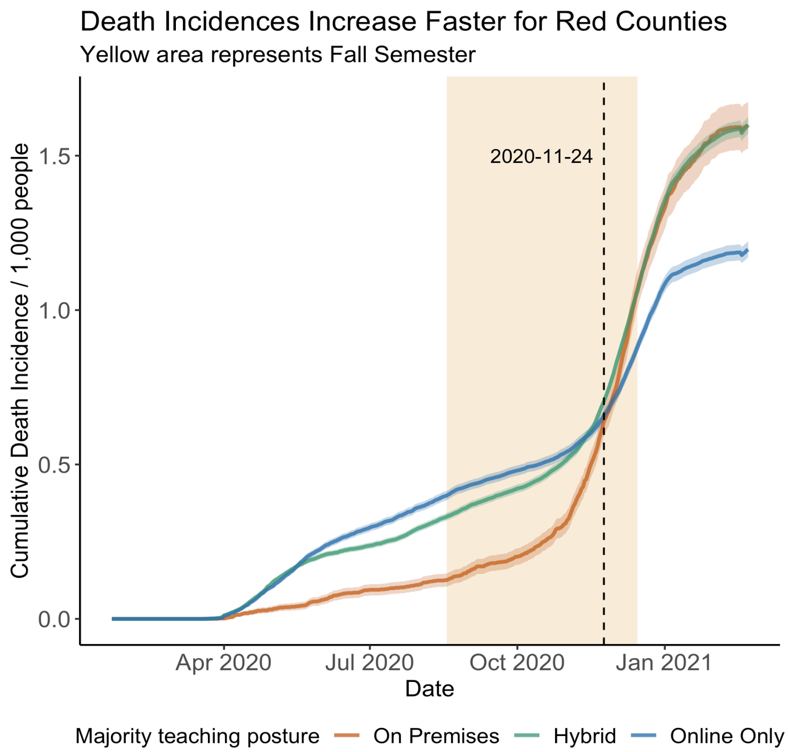


Figure 7: Death Incidence by Majority Teaching Posture

We see that the number of deaths remained 0 or very little at the beginning of the pandemic. Starting from April, the death incidence started to grow for all three types of teaching posture. However, the green and blue lines are increasing at a faster speed than the red line. This difference remains until around October, or the middle of the fall semester. This indicates that Hybrid and Online Only counties have a consistently larger death incidence in the beginning stages of the pandemic. Starting around October however, we notice that the red line denoting the On Premises counties, begins to rise steeply and actually surpasses the blue line around 11/24/2020 (represented by the dashed line) and then proceeds to closely follow the green line. This is very striking evidence that teaching posture may be causing a large increase in the death incidence in these counties.

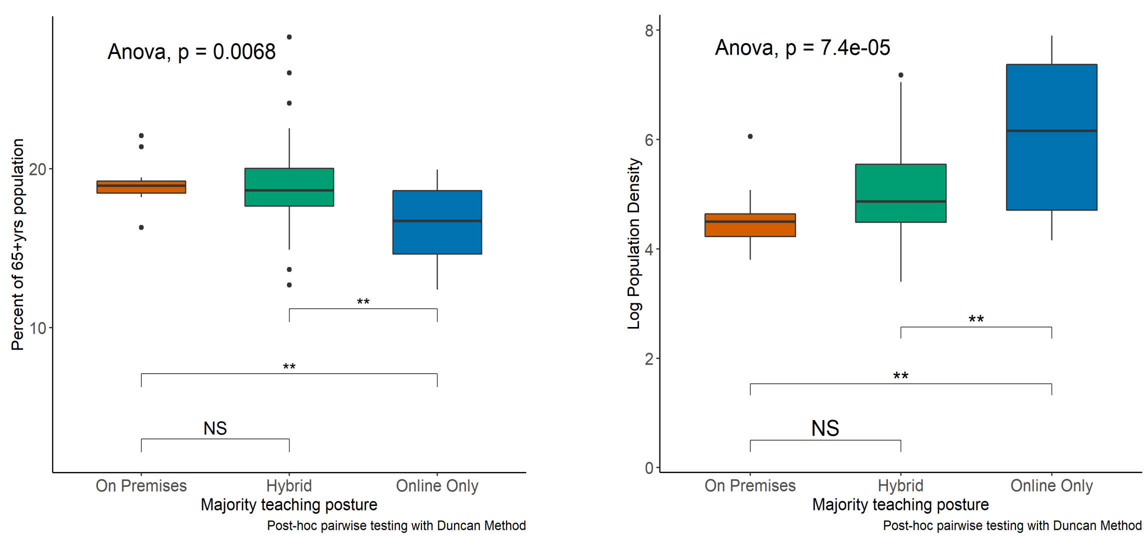
We narrow our analysis to the fall semester represented by the yellow area. We see that the red line is increasing faster and sooner than the other two lines. Therefore, we suspect that the differences in teaching posture could potentially have an association with the transmission of COVID-19. However, we recognize that many covariates could contribute to the differences in death incidence we observed. We performed a one-way ANOVA test to determine if the death incidence during the fall semester was statistically significant amongst the three types of different posture (supplemental Figure 18). The results of that test affirmed that the incidence was statistically different among the three groups. After performing post hoc pairwise comparison tests (Duncan, Dunnet, Scheffé), we see that the death incidence in On Premises counties was consistently significantly different from that in Online Only counties. The combination of Figure 6 and our statistical analysis may suggest that teaching posture does influence death incidence.

However, it is important to note that using cumulative death incidence as a measure of transmission may overinflate the severity of COVID-19 in an area for a given time because it is always increasing. Death incidence is also unable to take into account possible confounders that we could not control. One such confounder for the deaths during the fall semester may be the death incidence before the fall semester. From Figure 6 above, we see that the death incidence before the semester is very similar to the death incidences after the start of the semester. All local governments were monitoring the state of COVID-19 in their respective counties. It is not unreasonable to assume that perhaps some school districts decided whether or not they would be learning virtually by analyzing the severity of the disease right before school started. Note that the On Premises counties had fairly low death incidences before the semester started, indicating that it was probably safe to continue learning on school premises. Likewise, the blue line is the highest amongst the three lines a few months before the semester, so maybe the school districts decided those high levels were not safe enough to resume in-person learning.

There are so many possible non-school-related factors that could be attributed to the difference we see, so we cannot yet conclude that teaching posture is associated with transmission. Below we will look further into some of the possible confounding factors that could be influencing the differences in transmission in the counties of Ohio.

4.4 Confounding Variables

There are many non-school related factors that could explain any differences we see in the counties. Below we see two variables that could potentially influence the results of our analysis.



(a) Percent of 65+ years old population (b) Log population density

Figure 8: Distribution of Confounding Variables by Majority Teaching Posture

The first confounding variable we could look at is the proportion of seniors in each of the different types of counties. Here we define a senior as a person who is 65 years old or older. Counties with higher proportions of senior citizens would intuitively be more susceptible to larger death incidences. The boxplots in Figure 8 (a) shows that On Premises counties have significantly higher average percent of senior citizens than Online Only counties. It is possible that the higher deaths incidence we observe in On Premises counties could be attributed simply by a larger group of subjects susceptible to dying from COVID-19.

Likewise, we can also look at the population density in Figure 8 (b); we rescaled the population density with log transformation for better visualization. We see that On Premises and Hybrid counties have significantly lower population densities than Online Only counties. This is consistent with Figure 3 and Figure 5 previously, where we see that the counties we have assigned to the On Premises category are usually smaller in terms of population. This variable could also lead to the differences in transmission in the different counties. The more dense a county is, it is reasonable to assume that there are more opportunities to spread the disease from person to person because the people are physically closer to each other.

These are important variables to consider in continuing our analysis, but it is impossible to take into account every possible confounding variable that explains any differences in transmission. Other variables could include the percentage of people that are uninsured (supplemental Figure 19), rural-urban status of each county, as well as mobility, or the amount of time spent away from home on a given day. Because of these variables, it should be noted that we cannot come to a solid conclusion as to whether or not school teaching posture affects transmission. However, we can use a better metric for transmission like the exponential growth coefficient described above to better isolate the effect of school learning. With this metric, we can block the data such that we can take into account at least some of these confounding variables.

4.5 Exponential Growth Coefficients

Next, we find the exponential growth coefficient (referred to as B(t) in section 3.1.2) as a function of time for each of the three types of majority teaching posture. Figure 8 below shows the exponential coefficient we estimated using Equation (10), and the orange area represents the fall semester. We see that the growth coefficient for Hybrid and Online Only counties experience large growth together around April while the On Premises counties had consistently smaller growths.

If we look at the summer before school started, the slopes of the state of pandemic are all negative and they all fall under zero line before the first day of school, meaning that the disease is decelerating and starts to shrink before the school posture directly affects the transmission. Afterwards we observe an upward slope right before the school starts where the disease in Online Only (blue) counties picked up first, followed by Hybrid (green) counties, and then On Premises (red) counties. About 3 weeks into the fall semester (9/8/2020), the growth coefficients become positive for all counties. Each of the three lines during the fall semester shows approximately the same profile and shape. It's important to note that the red line, which represents the On Premises counties, appears to be shifted upwards during this time period. This however may be due to the fact that when the curves start increasing, the growth coefficients for On Premises counties were closer to zero, making it easier for growth to increase more quickly. Overall, Figure 9 appears to disprove the idea that the teaching posture in the counties has an effect on transmission.

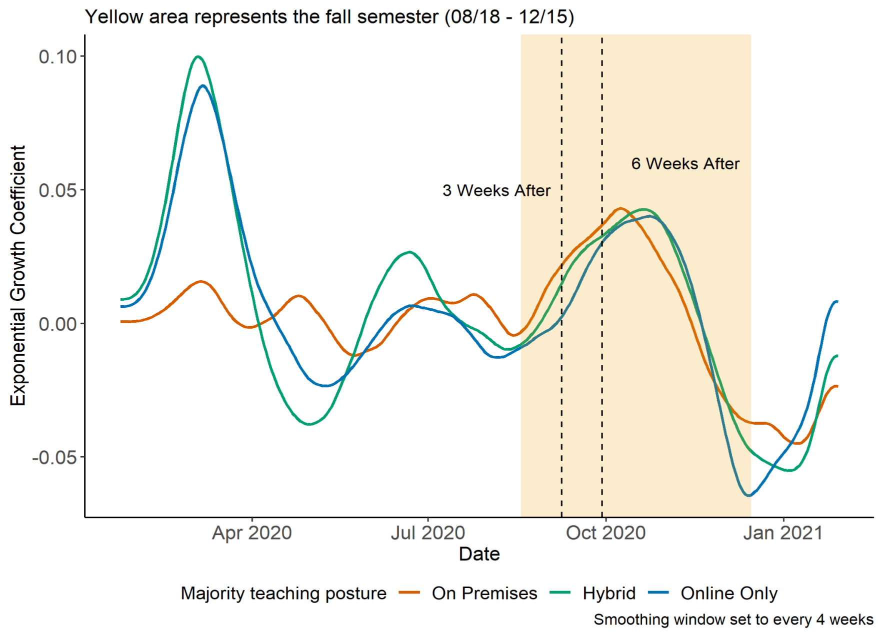


Figure 9: Exponential Growth Coefficient B(t) by Majority Teaching Posture

4.6 Maximum Growth Coefficient and Blocking

From this section onwards, we will derive metrics from the exponential growth model to explore if the school has an effect on the transmission of COVID. The first metric we will look into is the maximum exponential growth coefficient during the fall semester. This metric captures the severity of the pandemic during the fall semester. The larger the maximum growth coefficient shown, the higher the transmission and therefore, the worse the pandemic.

Figure 10 shows us that On Premises counties experienced the lowest transmission rate while the Online Only counties experienced the highest. This is different from what we see previously in Figure 9 where the peaks for three types of majority teaching posture are pretty even during the fall. The differences are due to the fact that we were looking at the schooling effect on an aggregate level while we now look at them on county levels. Additionally, by intuition, we hoped to see that the counties conducting On Premises teaching posture should have a higher transmission rate because of more gatherings in general, given that there indeed existed a schooling effect. However, Figure 10 does not show what we would expect to see. We suspect that we are seeing this result as a consequence of the confounding variables previously mentioned. Therefore, we will incorporate some of them to visually compare how they are related to this maximum severity metric.

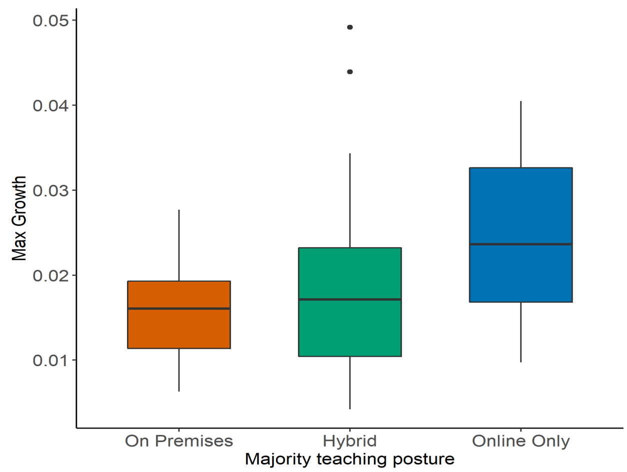


Figure 10: Distribution of Maximum Growth Coefficients by Majority Teaching Posture for All Counties

The first confounder we investigate is average mobility level, measured by the percentage of people away from home for over 6 hours daily. We can see from Figure 11 that overall, there is a negative relationship between mobility and the maximum growth coefficient for each county. This trend is mostly driven by Online Only counties in blue. This is quite surprising because intuitively, higher mobility should have higher severity of COVID. This phenomenon makes us wonder what other factors are influencing transmission. One potential confounding factor could be population density as mentioned earlier. As shown in the graph, the size of the dots represents the size of the population. Counties with larger populations are somewhat clustered together in the top left corner, whereas counties with smaller populations are in the bottom right. One possibility is that people realized how severe COVID was at the time as reflected by the maximum growth coefficient and decided to take more precautions by limiting their mobility or the time spent away from their homes. Another possibility is that perhaps counties with larger population densities tended to adhere more to the masking and social distancing policies set forth by the state, limiting overall mobility.

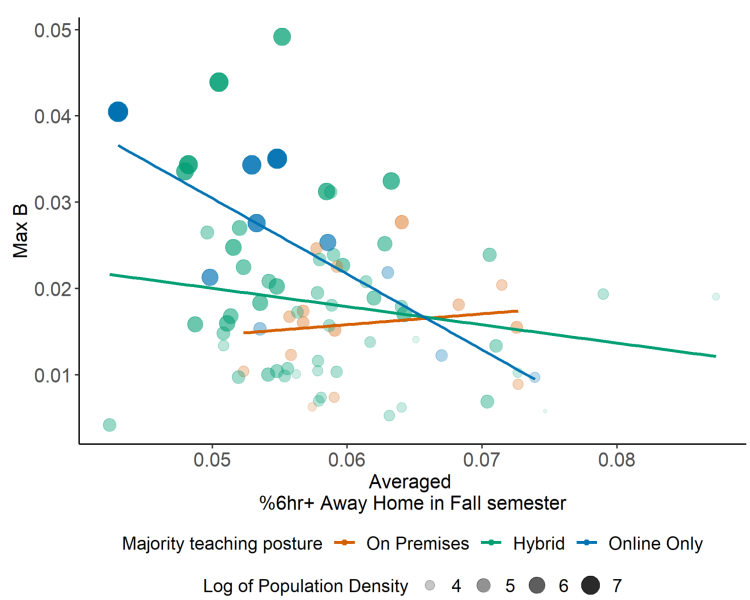


Figure 11: Maximum Growth Coefficient versus Average Mobility for All Counties

Therefore, we move on and examine log population density and how it affects the maximum severity. We can see from Figure 12 below that there is a positive relationship between the maximum growth coefficient and log population density, regardless of teaching posture. This gives the information that the counties with larger population density are likely to have more severe COVID-19 outbreaks. This is not surprising to see. However, by looking at the differences in slope of these different-colored lines, we see that the red line that consists of On Premises counties is steeper than the other two lines in green and blue. We conclude that the effect of population density is larger for On Premises counties than for Hybrid and Online Only counties. Another way to interpret this is, if we block on the log population density, let's say around 5, and we compare the counties. The On Premises counties had a more severe outbreak than the other two types of teaching posture.

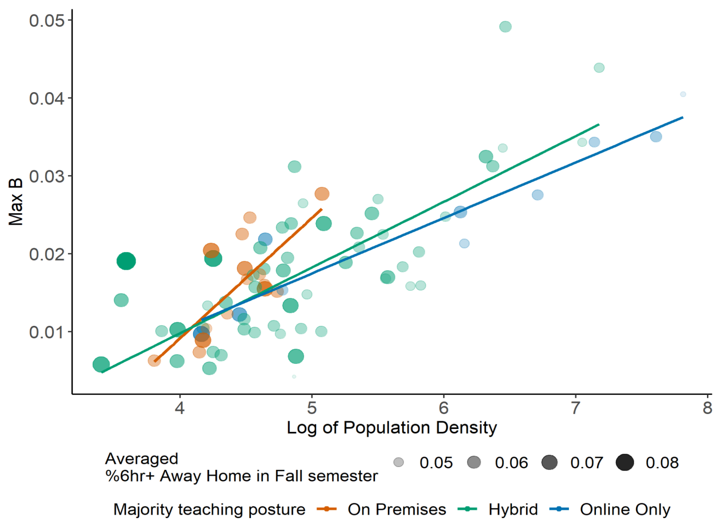


Figure 12: Maximum Growth Coefficient versus Log Population Density for All Counties

We now take a step further to think about if there is an efficient way to block on various confounders at a time. We have found a way by looking at the rural-urban status of these counties. In Figure 13, we notice that the only category of rural-urban status that contains all types of teaching posture is "Micropolitan". On top of that, we see that these counties have similar population densities between 4 and 5. By narrowing our analysis to these micropolitan counties, we can not only block on urban-rural status, but also block on population density which is previously shown as an important factor affecting the severity of COVID-19. The Appendix 5.4 also indicates that the relationship between maximum growth coefficient and mobility is much weaker now since micropolitan counties have moderate population density, which means that our previous guess of mobility probably makes sense, and this blocking is reasonable.

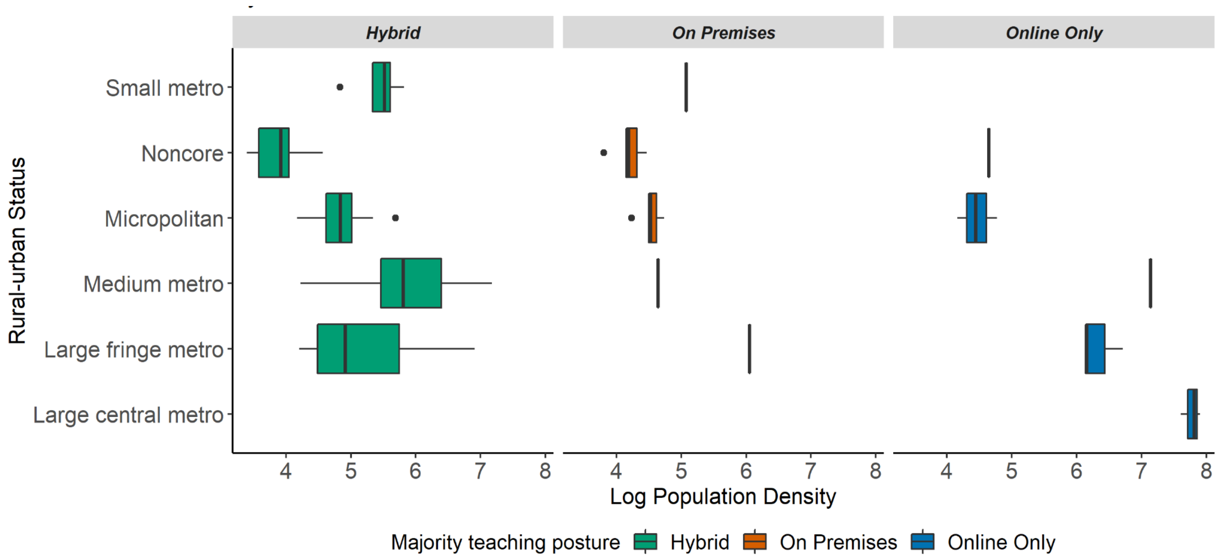


Figure 13: Distribution of Log Population Density by Rural-urban Status and Majority Teaching Posture

A map of where these micropolitan counties can be found in the 7 Supplemental Figures and Analysis section, Figure 20. We see that these counties with the same sort of urban type, same types of density are having a wide range of distribution across Ohio which can be helpful in eliminating some geographical factors. We therefore conclude that these micropolitan counties are more comparable than if we look at the entire sample of all the counties across Ohio.

Finally, with the micropolitan counties, we look at their maximum growth coefficients by teaching posture in Figure 14. We see those counties with the majority of On Premises learning tended to have more severe outbreaks of COVID-19. It is therefore striking if we compare this result with what we saw previously with all the counties included, as shown in Figure 10. These two figures are seemingly contradicting each other. However, blocking on micropolitan counties gives a more accurate assessment of how teaching posture affects maximum severity, because these counties are the most comparable in terms of population density. Despite this promising result, it is important to note that the sample sizes are too small to perform any statistical tests. It would be necessary for future analysis to incorporate the counties from other states so fully validate the findings seen here.

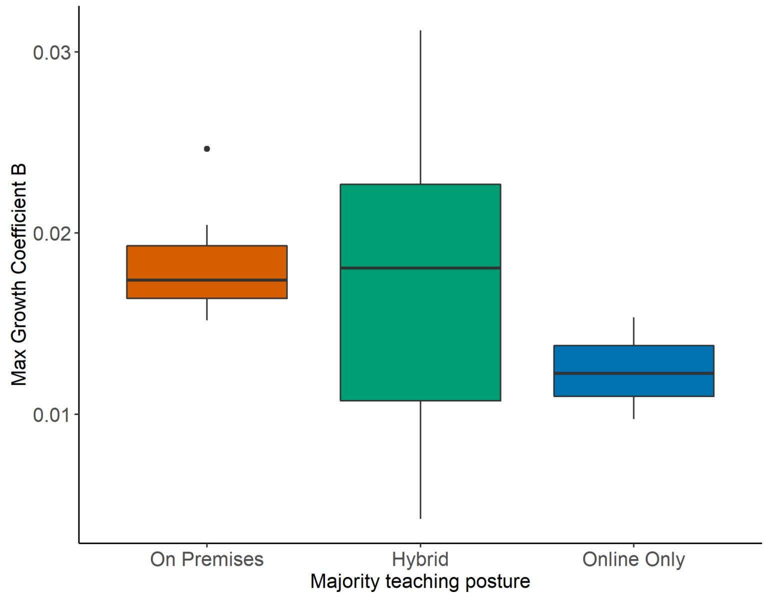


Figure 14: Distribution of Maximum Growth Coefficient by Majority Teaching Posture in Micropolitan Counties

However, we are not yet confident about our conclusion on the schooling effect because we have realized that the worst of the pandemic happened during the second half of the semester. Any interventions that happen during the first half of the semester may contribute to the differences that we see. Therefore, we will then examine another metric, the change of growth, in the next section.

4.7 Change in growth

As we mentioned in the Methods section, we also want to understand if limiting in-person components in schools affects the change in growth coefficients at the early stage of school reopening. By calculating the change in growth over a short time window, we can also account for unmeasured confounding variables by assuming these factors remain constant. For example, we can assume people's compliance with mask-wearing and social distancing did not change over 3 weeks; or the mobility and public interventions in the act stay the same over short periods.

We also assume that school posture takes approximately 3 weeks to reflect on the change in the growth coefficient, so we look at the difference in this time period. By comparing these two changes in growth in such a short time window, we can reduce the effect brought by other confounding factors during the rest of fall semester. We define the change in growth at the start of Fall semester as the difference between exponential growth at 3 weeks and 0 weeks (B(3) - B(0) in reference to equation 10). Likewise, we define the change in growth three weeks after school reopens as the difference between exponential growth at 6 weeks and 3 weeks. (B(6) - B(3)) as shown in Figure 15.

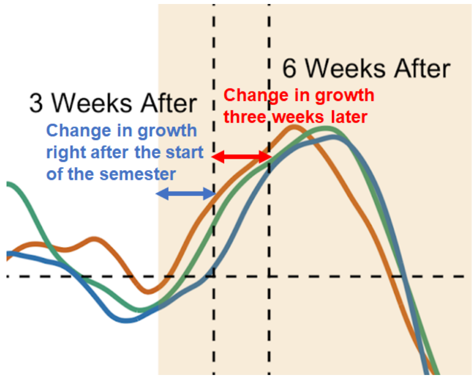
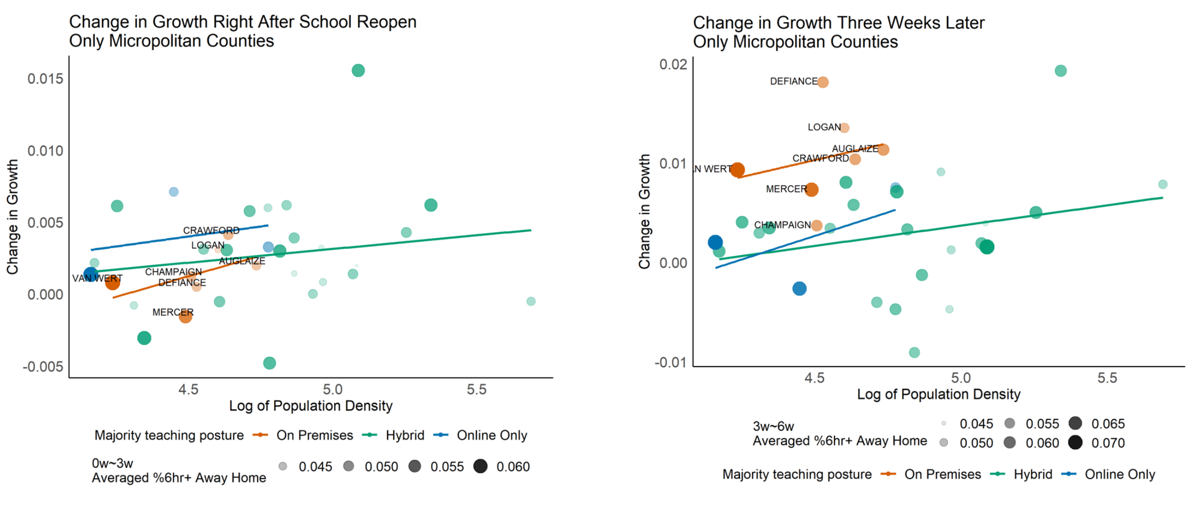


Figure 15: Definition of Change in Growth

Similar to Figure 12, we have plotted the change in growth versus log population density for better visualization and comparison (Figure 16). The graph on the left in Figure 16 approximates the state of COVID-19 at the beginning of the semester, when the teaching posture has yet to take effect on transmission. During this time we find that all of the red points denoting the On Premises counties are centered around zero. In other words the growth of the pandemic isn't growing much at that time. In contrast, the blue, or Online Only counties are all above 0, meaning that these counties are growing faster in terms of the pandemic. However, when we look at the plot of three weeks later on the right of Figure 16, we see that the change in growth of On Premises counties are all above 0 now. Contextually, these counties with On Premises learning are experiencing faster growth in the COVID-19 pandemic after school had the chance to take effect. The shape for Online Only and Hybrid counties does not change greatly, but the red points have shifted above all of the other points after the first 3 weeks of school. This is strong evidence to suggest that On Premises learning may lead to fast growth of COVID-19.



(a) (b)

Figure 16: Shifting of Change in Growth vs Log Population Density for On Premises Counties

Similarly, we plot the change in growth against average mobility during the given time frames (Figure 17). We find that despite the time window, the correlation between change in growth and mobility is quite weak here, meaning our blocking is effective at controlling for mobility as well as population density, making micropolitan counties even more comparable. As before in Figure 16, we see that the red points have been shifted above all others.

(a) (b)

Figure 17: Shifting of Change in Growth vs Mobility for On Premises Counties

This result is quite interesting because we have already blocked our data to micropolitan counties and shortened the time window, which means that the effect of other confounders is quite small now. So, this shifting is probably brought by the reopen of school. We also performed a sensitivity analysis to observe how things changed before and after the start of the school and if our results are consistent when varying our defined time windows. This analysis can be found in the Supplemental Figures and Analysis section in Figure 22 and Figure 23.

5 Discussion

This study demonstrates the effect of school teaching posture on COVID transmission. In the initial stages of our analysis, the cumulative death incidence plot (Figure 7) suggested a striking difference by school posture. We saw that the red line grew much faster than either the blue or green line, indicating that On Premises learning was associated with higher death incidence. However, we identified several confounding variables that could potentially explain the result we saw. We then developed the exponential growth model to directly measure the state of COVID-19 pandemic at a given time. Therefore, we continued our analysis with two summary measurements, namely the maximum growth coefficient and change in growth. With these measures of transmission, we were able to identify some confounding variables, such as population density and urban-rural status that could potentially be skewing our initial results. By incorporating these confounders and applying blocking, the results indicate that On Premises counties had faster transmission of COVID-19 than Online Only counties. Our current analysis might suggest that school teaching posture does indeed have an effect on COVID-19 transmission.

5.1 Discussion for Maximum Growth Coefficients

We first examined the maximum growth coefficient to measure the severity of the pandemic during the fall semester. We have found that when blocking on micropolitan counties, On Premises counties have higher maximum exponential growth than Online Only counties, indicating faster transmission and higher severity for these counties. In addition, we also realized that blocking is an important and useful tool that helps remove the effect of some important confounding variables like population density.

5.2 Discussion for Change in Growth

From Figure 16 and Figure 17, we find that the change in growth of On Premises counties three weeks later after the schools reopen is higher than that of Hybrid and Online Only counties. That is to say, the red line shifted from the bottom to top when teaching posture would have the chance to take effect. This phenomenon is most likely brought about by the majority teaching posture taken by counties. However, a notable limitation to our current analysis is that we only have 7 micropolitan counties whose majority of students are learning on the school premises and 3 micropolitan counties adopting the Online Only teaching posture in Ohio. Because of this, we cannot draw an exact conclusion that school posture led to such faster growth in these types of counties.

5.3 Limitations

We recognized some limitations of this study. The reliability of the data may be impacted by the sourcing and how schooling and COVID numbers are reported. Also due to the lack of schooling data, we cannot confirm if there are changes for these school districts on teaching posture during the fall semester. Although we have information regarding whether or not a school temporarily shut down, we do not know when the shutdown occurred, or if multiple shutdowns occurred. Besides, the methodological choices are limited to visualizations as a part of the exploratory data analysis. We haven't applied more sophisticated testing methods, due to the complex nature of purely observational data. Concerning the exponential growth modeling we used for this study, the method is built on existing evidence that the average time lag between infection to death is 24 days. We therefore cannot confirm if the results will remain solid if this number is to change. Finally, the generalizability of the results can be constrained by the fact that our conclusion is derived from only the data of Ohio, specifically in micropolitan counties.

5.4 Next steps

We now have some evidence that suggests school teaching posture may have some effects on COVID transmission. However, since we only have 7 samples of On Premises counties and 3 samples of Online Only counties in Ohio, further research study is needed to establish if these effects do truly exist. Moving forward, it would be beneficial to look at some other states, such as Texas, meaning more samples or counties to be studied. If we do observe consistent outcomes, we will be more confident that the probability of teaching posture affecting COVID transmission is high.

6 References

Bonvini, M., Kennedy, E.H., Ventura, V., Wasserman, L.. (2021) Causal inference in the time of COVID-19. [*Preprint*]. Mar 7, 2021. Available from: *https://arxiv.org/abs/2103.04472*

Unwin, H. J. T., Mishra, S., Bradley, V. C., Gandy, A., Mellan, T. A., Coupland, H., Ish-Horowicz, J., Vollmer, M. A., Whittaker, C., Filippi, S. L. et al. (2020). State-level tracking of COVID-19 in the United States. *Nature communications* **11** 1–9.

Ventura, Valerie. (2021). PHIGHT notes.

Williams, N., Radia, T., Harman, K. *et al.* COVID-19 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection in children and adolescents: a systematic review of critically unwell children and the association with underlying comorbidities. *Eur J Pediatr* 180, 689–697 (2021). *https://doi.org/10.1007/s00431-020-03801-6*

7 Supplementary Figures and Analysis

7.1 Death Incidence

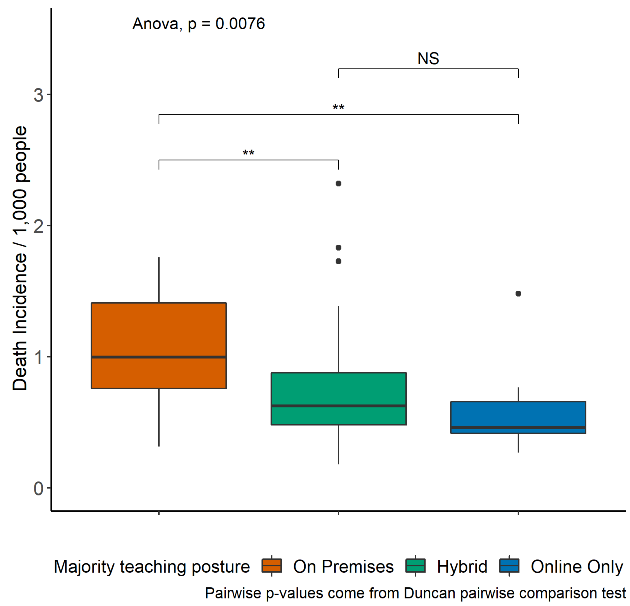


Figure 18: Death Incidence in the Fall Semester

Above we see the distribution of death incidence during the fall semester broken down by teaching posture. An ANOVA test produced a value of .0076, indicating that there is a significant difference in the incidence for the three types of counties. The post hoc analysis using the duncan multiple range comparison test showed that the On Premises counties' death incidence was higher than the Hybrid and Online Only counties.

7.2 Confounders

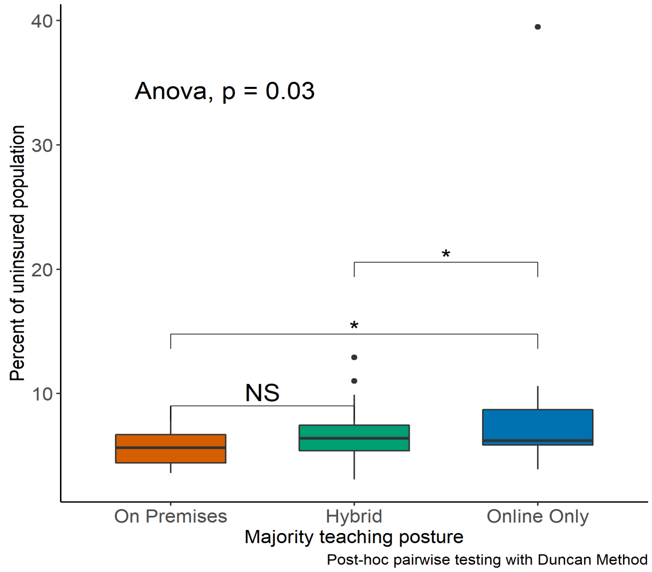


Figure 19: Percentage of Uninsured People by County

The boxplots above in Figure 19 show the distribution for the proportion of uninsured people in each of the three types of counties. An ANOVA test proved that the proportion was significantly different for the teaching posture, and the Duncan test showed that On Premises counties have lower uninsured populations than Hybrid and Online Only counties.

7.3 Micropolitan Counties

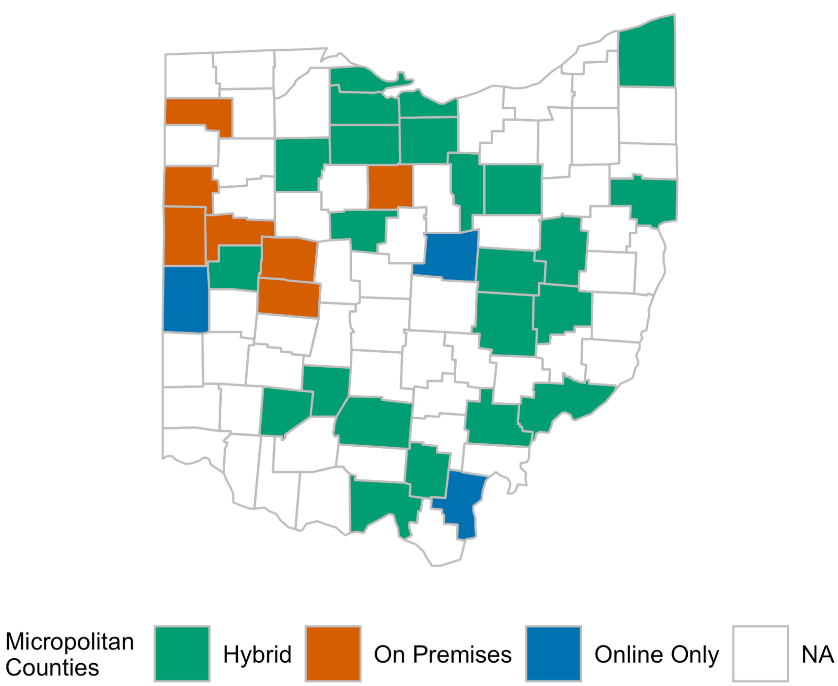


Figure 20: Map of Micropolitan Counties by Teaching Posture

In our analysis, we block on micropolitan counties because it was found that these counties were the most comparable in terms of population density, a known impactful confounder when looking at the maximum severity of the pandemic. Figure 20 shows which counties are considered micropolitan as well as their teaching posture characterization.

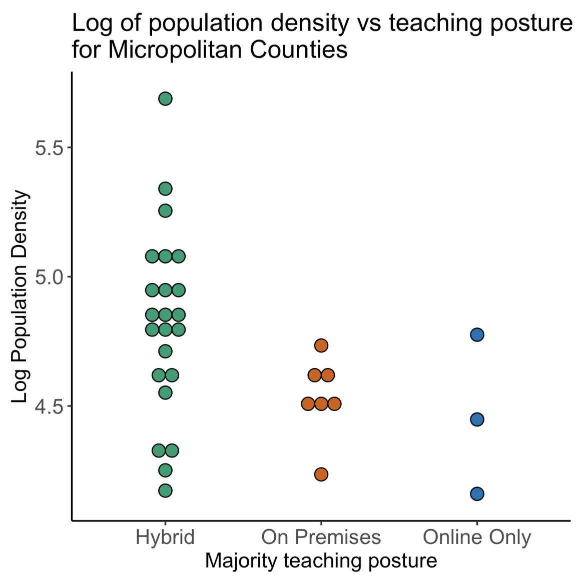


Figure 21: Distribution of log population density by teaching posture for micropolitan counties

Figure 21 shows the distribution of the log population density in the micropolitan counties. Here we see that the range of log population density is larger for Hybrid counties, but is somewhat comparable to the other two types of counties with the majority of the values between 4 and 5. Because of this, it is justified to block on these types of counties.

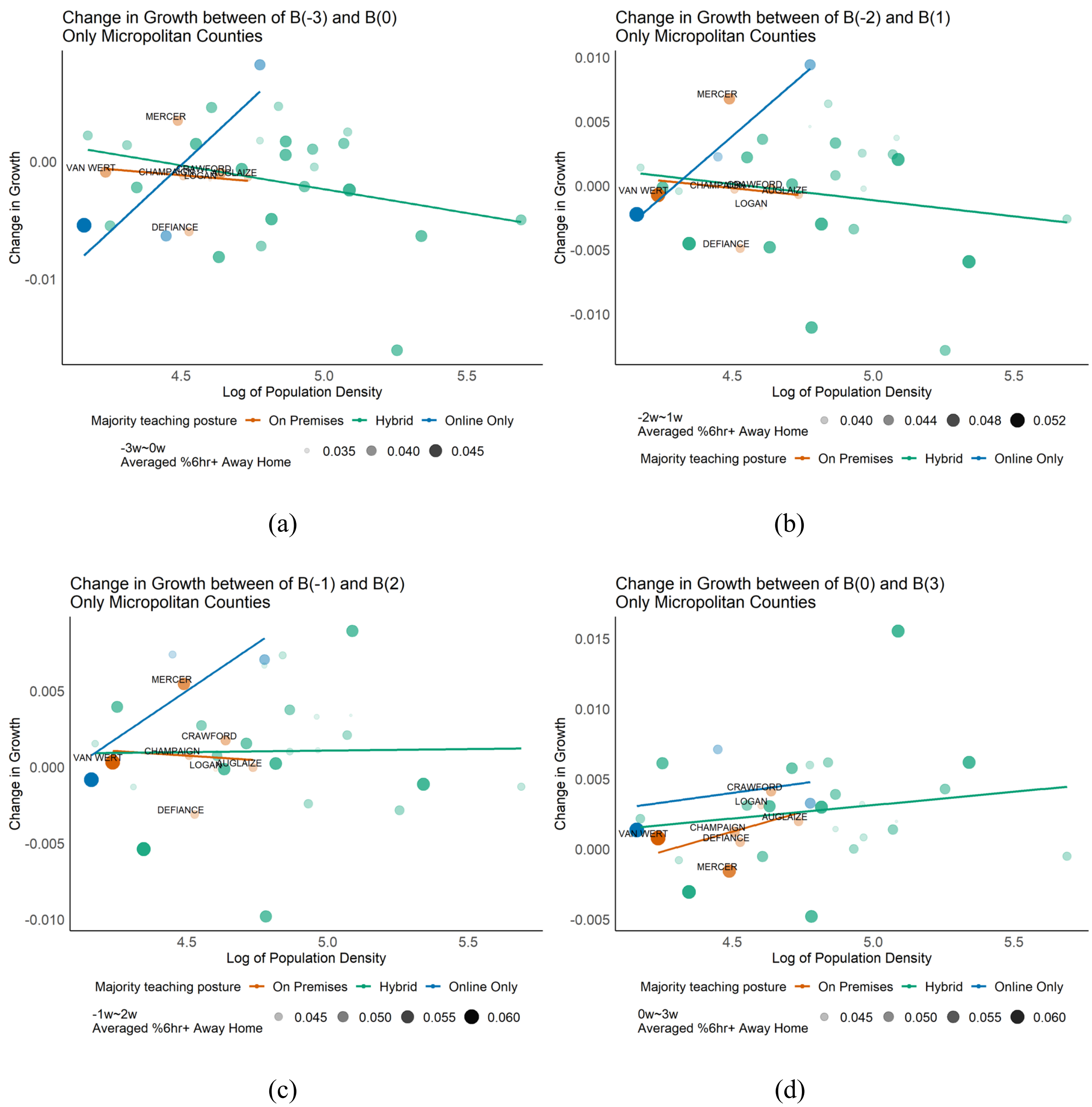
7.4 Sensitivity Analysis

In order to make sure that the shifting of the red line (On Premises counties) is not brought by chance, we will conduct a sensitive analysis to detect how the change in growth varies throughout time. The whole sensitive analysis is based on ‘Change in growth versus Log Population Density.

Since we assume that the school posture takes three weeks to reflect on the growth coefficient, the growth coefficients before 3 weeks after the start of school are all regarded as not taking effect. So, we use B(3) as a turning point. The changes in growth we want to test are as below:

(1) Before school posture taking effect: B(0)-B(-3), B(1)-B(-2), B(2)-B(-1), B(3)-B(0) [also known as change in growth right after the start of school reopen]

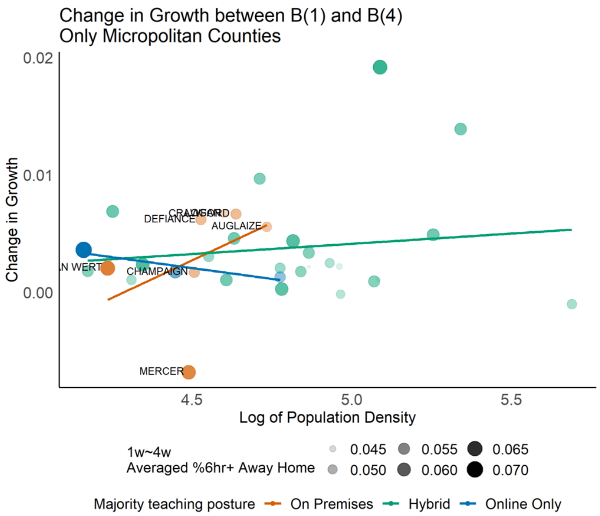
(2) After school posture taking effect: B(4)-B(1), B(5)-B(2), B(6)-B(3) [also known as change in growth three weeks later], B(7)-B(4)



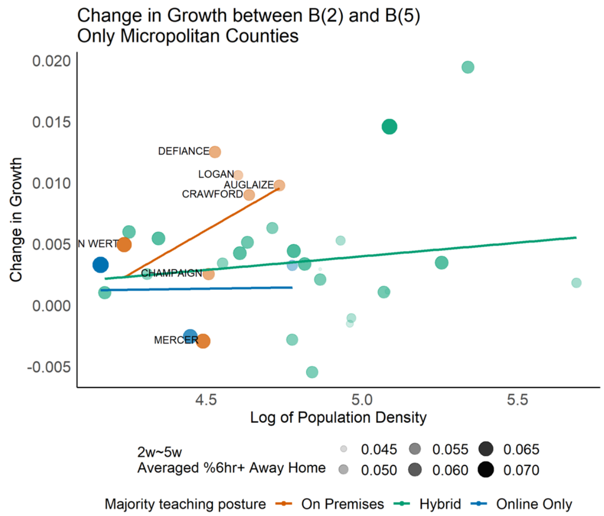
(a) (b) (c) (d)

Figure 22: Change in Growth Before School Reopen Takes Effect

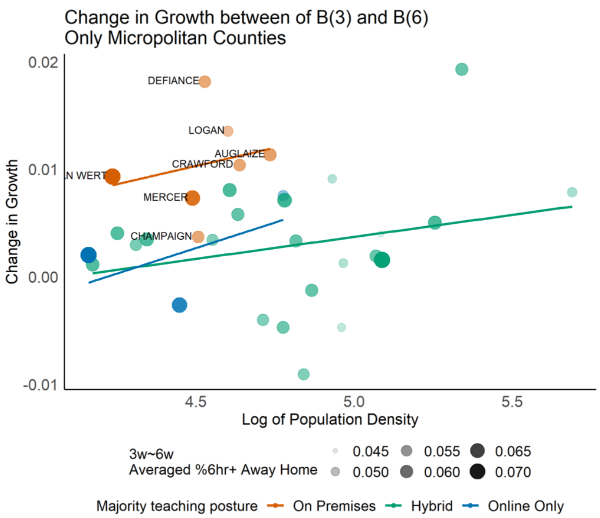
Before schooling posture taking effect, the changes in growth for On Premises Counties (seen in Figure 22) are almost below the other two lines (Hybrid and Online Only). In addition, we do not observe that the red line increased with time. The changes in growth before schooling posture taking effect almost remain constant.



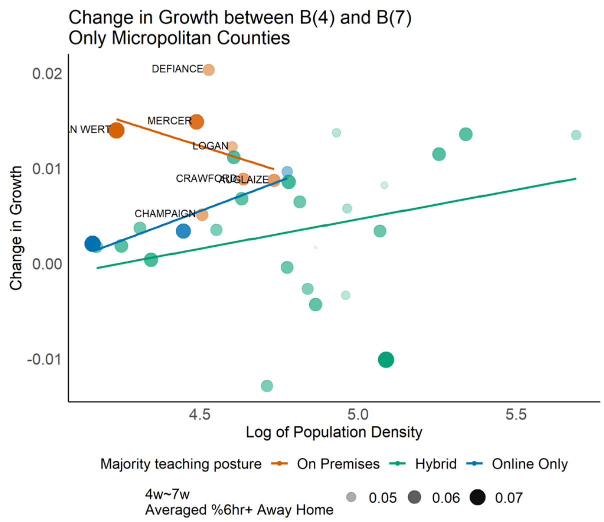
(a)



(b)



(c)



(d)

Figure 23: Change in Growth After School Reopen Takes Effect

When we believe the school posture starts to take effect, things change a lot. The red line started to shift from bottom to top(seen in Figure 23). This shifting is probably due to the schooling effect, which is aligned with our analysis in the 4.7 section.

Technical Appendix

Ziyan Zhu, Cheyenne Ehman, Yixuan Luo, Zi Yang

Appendix 1: Motivations

```
## Set up aesthetic theme for all graphs generated in the report
Sys.setlocale("LC_TIME", "English")
```

```
## [1] "English_United States.1252"
```

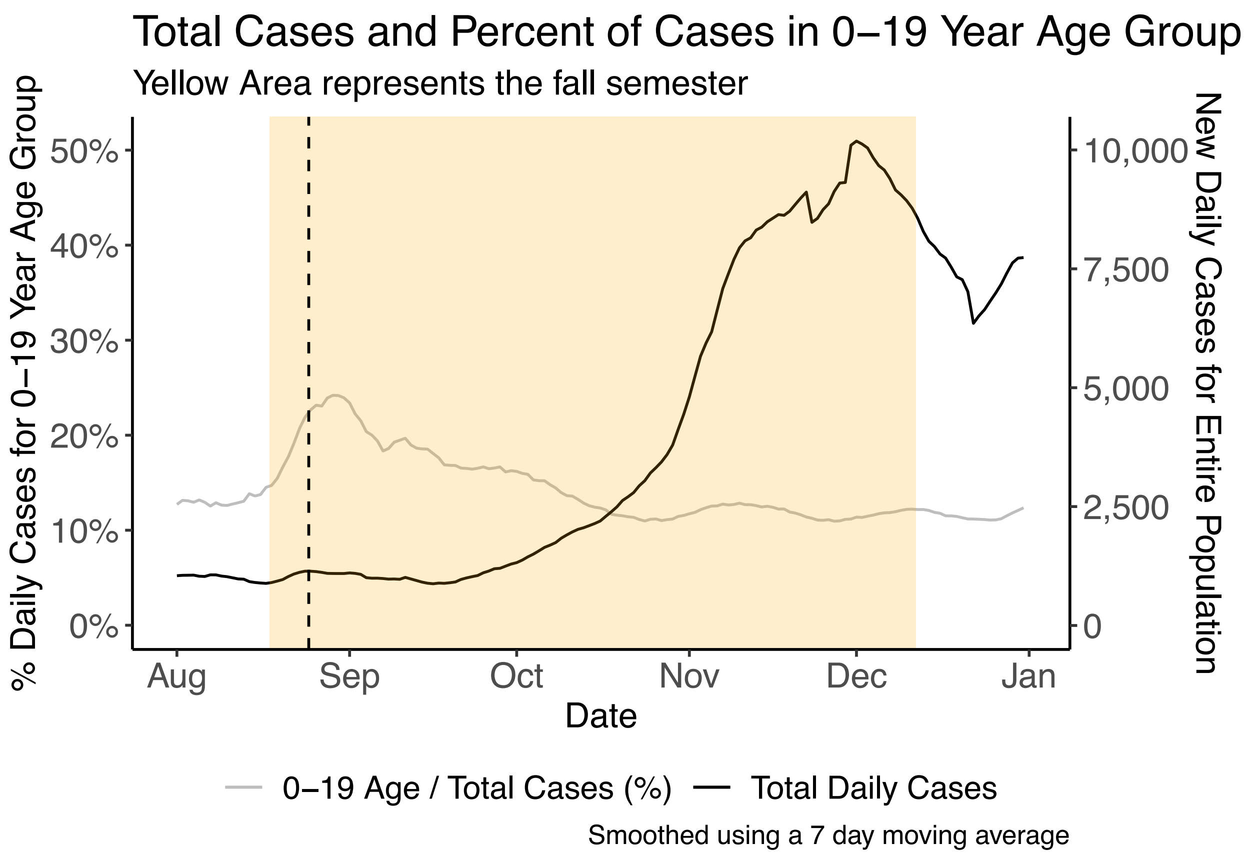
```
library(ggrepel)
library(tidyverse)
library(lubridate)
require(scales)
library(readxl)
library(ggpubr)
library(PMCMRplus)
require(DescTools)
library(cowplot)
library(sp)
source("step2_new.R")
# color blind friendly Palette
library(ggthemes)
col_theme <- c("Hybrid"="#009E73","On Premises"="#D55E00","Online Only"="#0072B2")
## plot theme
grid_theme <- theme(axis.line = element_line(colour = "black"),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.border = element_blank(),
    legend.key = element_blank(),
    panel.background = element_blank(),
    legend.box="vertical", legend.margin=margin())
team_theme <- grid_theme+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        axis.text = element_text(size=13),
        title=element_text(size=13),
        strip.text.x = element_text(size = 10, face = "bold.italic"))
```

1.1 Time series plot of daily cases of children under 19 years old

This plot serves as an motivation example to investigate on the transmissibility of children under 18 years old; then we are interested in whether children act as an vector of transmission after school reopening in the 2020 Fall semester.

```
## process Age data
cases_by_age <- read_excel("OhiobyAge.xlsx")
```

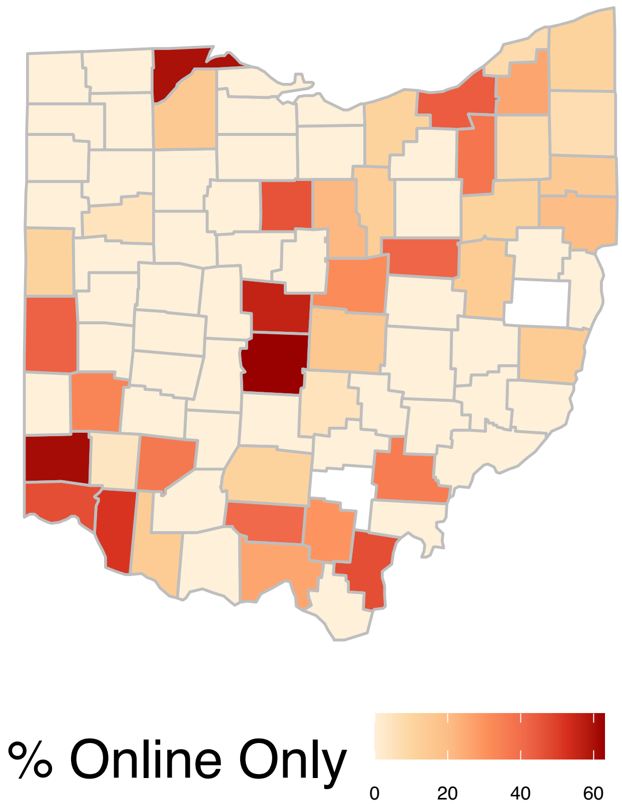
```r
rolling_age_cases <- cases_by_age %>%
  mutate(youth_prop_roll = zoo::rollmean(`00_19/total(%)`, k = 7, fill = NA),
         all_roll = zoo::rollmean(`00_80+`, k = 7, fill = NA))
colors <- c("Total Daily Cases" = "black",
            "0-19 Age / Total Cases (%)" = "gray")
coeff <- 200
cases_by_age_long <- cases_by_age %>%
  gather(age_group, percent_cases,
         `00_19/total(%)`:`80+/total(%)`,
         factor_key=TRUE) %>%
  group_by(age_group) %>%
  mutate(roll_percent_cases= zoo::rollmean(percent_cases, k = 7, fill = NA))

colors <- c("Total Daily Cases" = "black",
            "0-19 Age / Total Cases (%)" = "gray")
coeff <- 200
ggplot(rolling_age_cases, aes(x=Date)) +
  geom_line( aes(y=youth_prop_roll,
                 color = "0-19 Age / Total Cases (%)"),
             na.rm = T)+
  geom_line( aes(y=all_roll/coeff,
                 color = "Total Daily Cases"),
             na.rm = T) +
  scale_y_continuous(
    # Features of the first axis
    name = "% Daily Cases for 0-19 Year Age Group",
    labels = function(x){paste0(x, "%")},
    # Add a second axis and specify its features
    sec.axis = sec_axis(~.*coeff, name="New Daily Cases for Entire Population",
                        label=comma)
  ) +
  geom_rect(data=rolling_age_cases[1,],
            aes(xmin=as.POSIXct ("2020/08/18"), xmax=as.POSIXct ("2020/12/12"),
                ymin=-Inf,ymax=Inf),
            color = NA,alpha=0.2, show.legend = F, fill = "orange") +
  geom_vline(xintercept = as.POSIXct ("2020/08/18") + days(7),lty = 2)+
  xlim(c(as.POSIXct ("2020/08/01"),as.POSIXct ("2021/01/01"))) +
  labs(title = "Total Cases and Percent of Cases in 0-19 Year Age Group",
       subtitle = "Yellow Area represents the fall semester",
       caption = "Smoothed using a 7 day moving average",
       color = "")+
  scale_color_manual(values = colors)+
  team_theme +
  theme(legend.position='bottom')
```



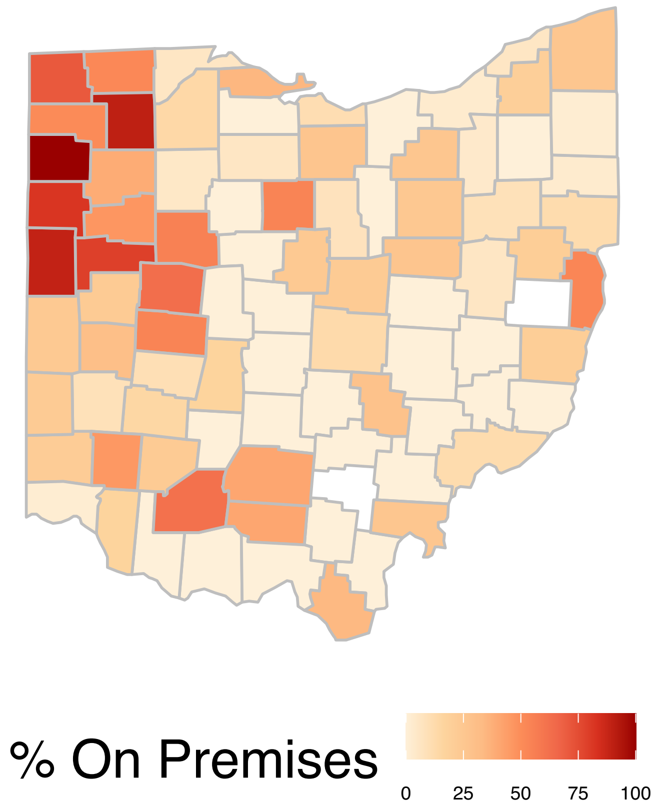
Appendix 2: Maps of Ohio

2.1 Geographical distribution of the teaching posture proportions, population and student enrollmen tat county-level

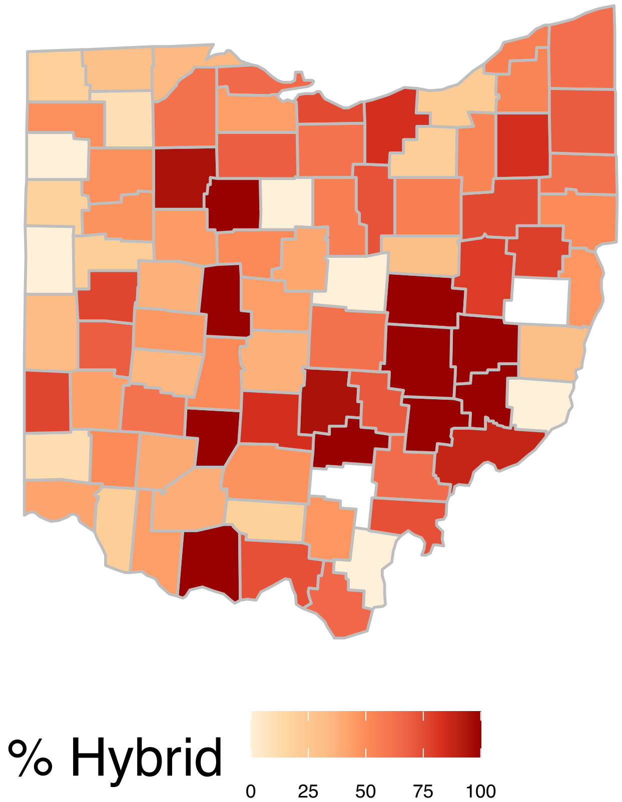
```
ohio_map <- map_data("county") %>%subset(region=="ohio")%>%
  mutate(county=toupper(subregion))%>%dplyr::select(long,lat,county,group)

# Map of proportion of students taking online-only classes
wide_teaching_enroll%>%
  left_join(ohio_map,by='county')%>%
  mutate(Online_Only= Online_Only*100)%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = Online_Only), color = "gray") +
  coord_fixed(1.3) + theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='% Online Only')+
  theme(legend.position = "bottom",
        legend.text = element_text(size=),
        legend.title = element_text(size=20))
```



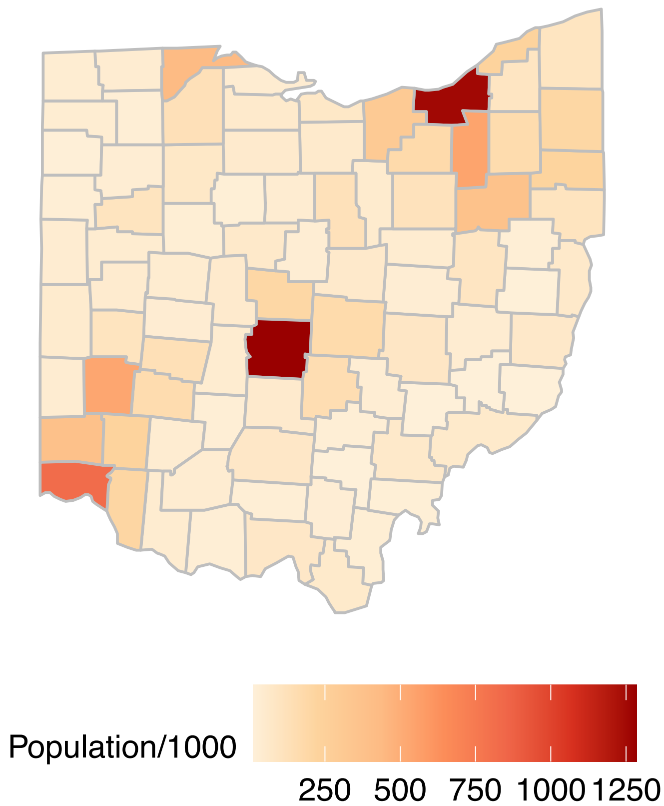

```r
# Map of proportion of students taking on-premises classes
wide_teaching_enroll%>%
  left_join(ohio_map,by='county')%>%
  mutate(On_Premises= On_Premises*100)%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = On_Premises), color = "gray") +
  coord_fixed(1.3) + theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='% On Premises')+
  theme(legend.position = "bottom",
        legend.text = element_text(size=),
        legend.title = element_text(size=20))
```



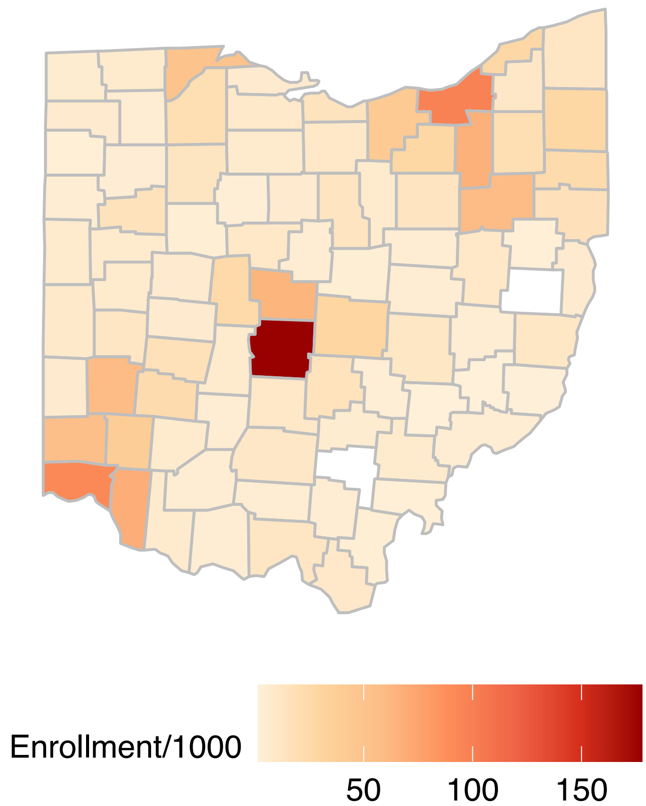

```r
# Map of proportion of students taking hybrid classes
wide_teaching_enroll%>%
  left_join(ohio_map,by='county')%>%
  mutate(Hybrid= Hybrid*100)%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = Hybrid), color = "gray") +
  coord_fixed(1.3) +
  theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='% Hybrid')+
  theme(legend.position = "bottom",
        legend.text = element_text(size=),
        legend.title = element_text(size=20))
```




```r
# Map of population size
cases%>%
  distinct(COUNTY,POPULATION)%>%
  left_join(ohio_map,by=c('COUNTY'='county'))%>%
  mutate(population = POPULATION/1000)%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = population), color = "gray") +
  coord_fixed(1.3) + theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='Population/1000')+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        legend.position = "bottom",
        legend.key.size = unit(2,"lines"))
```

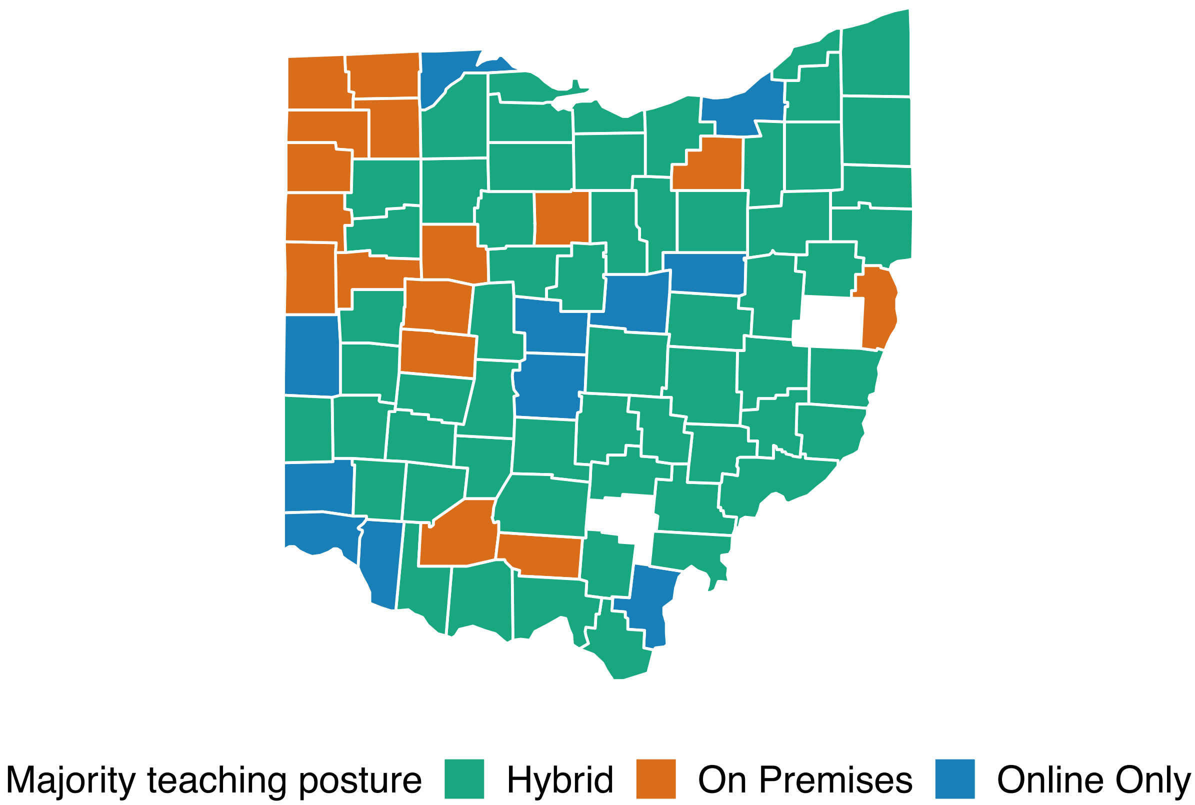



```r
# Map of student enrollments
teachingmethod_enroll%>%
  distinct(county,county_enroll)%>%
  left_join(ohio_map,by=c('county'))%>%
  mutate(county_enroll = county_enroll/1000)%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = county_enroll), color = "gray") +
  coord_fixed(1.3) + theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='Enrollment/1000')+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        legend.position = "bottom",
        legend.key.size = unit(2,"lines"))
```



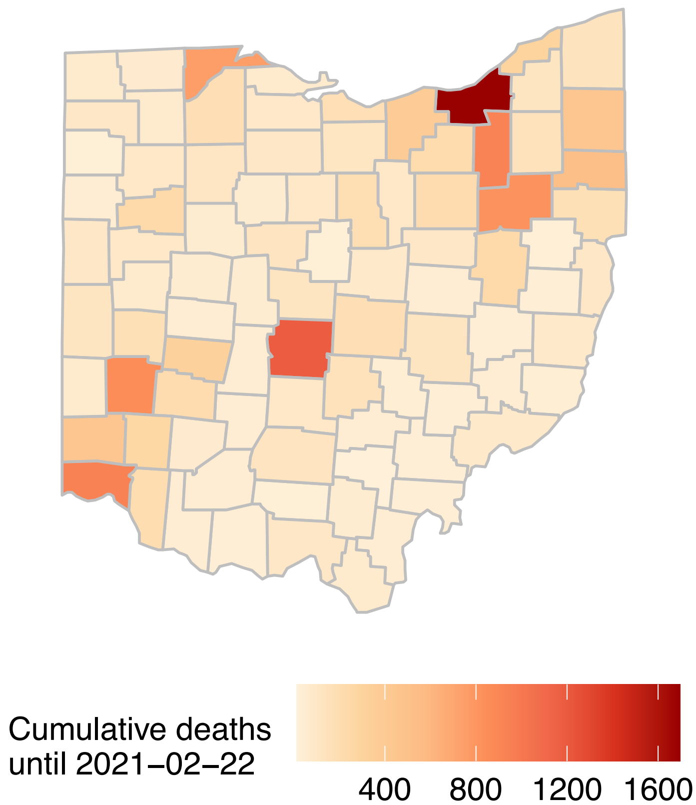
2.2 Geographical distribution of the majority teaching posture at county-level

```
wide_teaching_enroll%>%
  left_join(ohio_map,by='county')%>%
  mutate(On_Premises= On_Premises*100)%>%
  ggplot() + geom_polygon(aes(x = long, y = lat, group = group,
                              fill = as.factor(major_teaching)),
                          color = "white",alpha=0.9) +
  coord_fixed(1.3) + theme_map() +
  scale_fill_manual(values=col_theme)+
  labs(fill='Majority teaching posture')+
  theme(legend.position = "bottom",
        legend.text = element_text(size=14),
        legend.title = element_text(size=14))
```

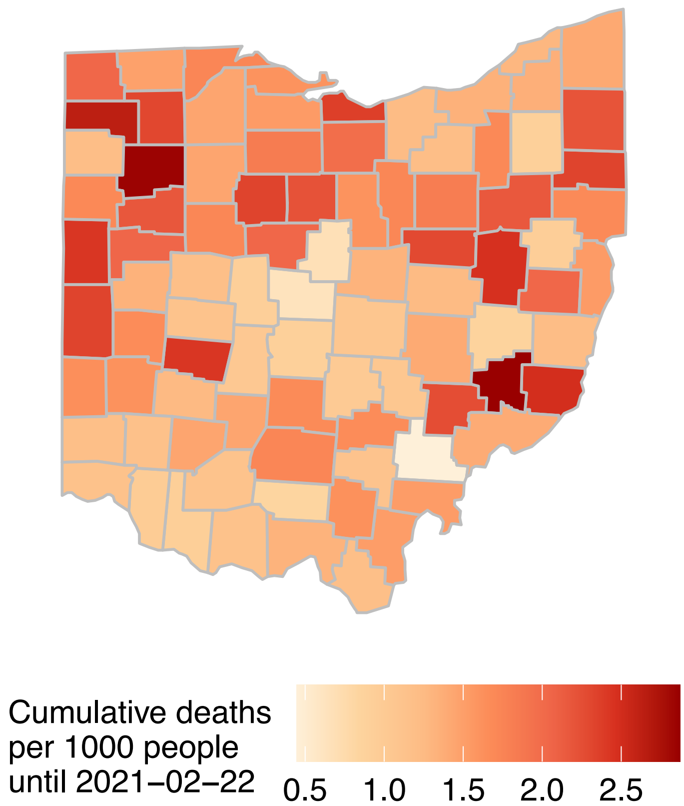


2.3 Geographical distribution of cumulative COVID-19 deaths and deaths incidence until 02/22/2021.

```
death_prop%>%
  left_join(ohio_map,by=c("COUNTY"='county'))%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group=group,fill = CUMDEATHS), color = "gray")+
  coord_fixed(1.3) + theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='Cumulative deaths \nuntil 2021-02-22')+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        legend.position = "bottom",
        legend.key.size = unit(2,"lines"))
```




```
death_prop%>%
  left_join(ohio_map,by=c("COUNTY"='county'))%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group=group,fill = death_per_1000),
               color = "gray") +
  coord_fixed(1.3) + theme_map() +
  scale_fill_distiller(palette = "OrRd",direction = 1)+
  labs(fill='Cumulative deaths \nper 1000 people \nuntil 2021-02-22')+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        legend.position = "bottom",
        legend.key.size = unit(2,"lines"))
```



Appendix 3: Difference in death incidence during the Fall semester

3.1 One-way anova test on death incidence by teaching posture

```
county_open_teaching_enroll <- OH_K12%>%
  distinct(county,leaid,teachingmethod,county_enroll,district_enroll,date)%>%
  group_by(county,date,teachingmethod)%>%
  summarise(open_county_enroll = sum(district_enroll),
            opendate_prop = sum(district_enroll)/county_enroll)%>%
  rename(opendate = date)

major_reopening <- county_open_teaching_enroll%>%
  group_by(county)%>%
  slice(which.max(opendate_prop))%>%
  rename(COUNTY=county,major_opendate=opendate)%>%
  distinct(COUNTY,major_opendate,opendate_prop)


# see when the intesection happens
date.intercept <- as.Date("2020-11-24")
# add 95% confidence bans
confidence_level <- .95
z_cl <- qnorm(confidence_level)
# case_policy_wide
case_policy_wide <- cases %>%
```

```r
  left_join(county_policy_wide[,c("county","major_teaching",
                                "Online_Only","Hybrid","On_Premises")],
          by = c("COUNTY" = "county")) %>%
  mutate(death_prop = CUMDEATHS/POPULATION)
opendate_cases <- case_policy_wide%>%
  inner_join(major_reopening%>%dplyr::select(COUNTY,major_opendate),by=c('COUNTY'))

# filter and summarize deaths incidence in the Fall
fall_cases <- opendate_cases %>%
  filter(DATE >= major_opendate & DATE <= as.Date("2020/12/15")) %>%
  group_by(COUNTY) %>%
  arrange(DATE) %>%
  filter(row_number()==1 | row_number()==n()) %>%
  mutate(death_incidence = diff(CUMDEATHS),
         death_incidence_per_1000 = death_incidence*1000/POPULATION) %>%
  distinct(COUNTY,POPULATION,major_teaching,
           death_incidence,death_incidence_per_1000)

# one-way ANOVA
fall_major_teaching.aov <- aov(death_incidence_per_1000 ~ major_teaching,
                               data = fall_cases)

summary(fall_major_teaching.aov) # p-value of .012
```

```
##                Df Sum Sq Mean Sq F value  Pr(>F)
## major_teaching  2  1.653  0.8264   5.205 0.00761 **
## Residuals      76 12.067  0.1588
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

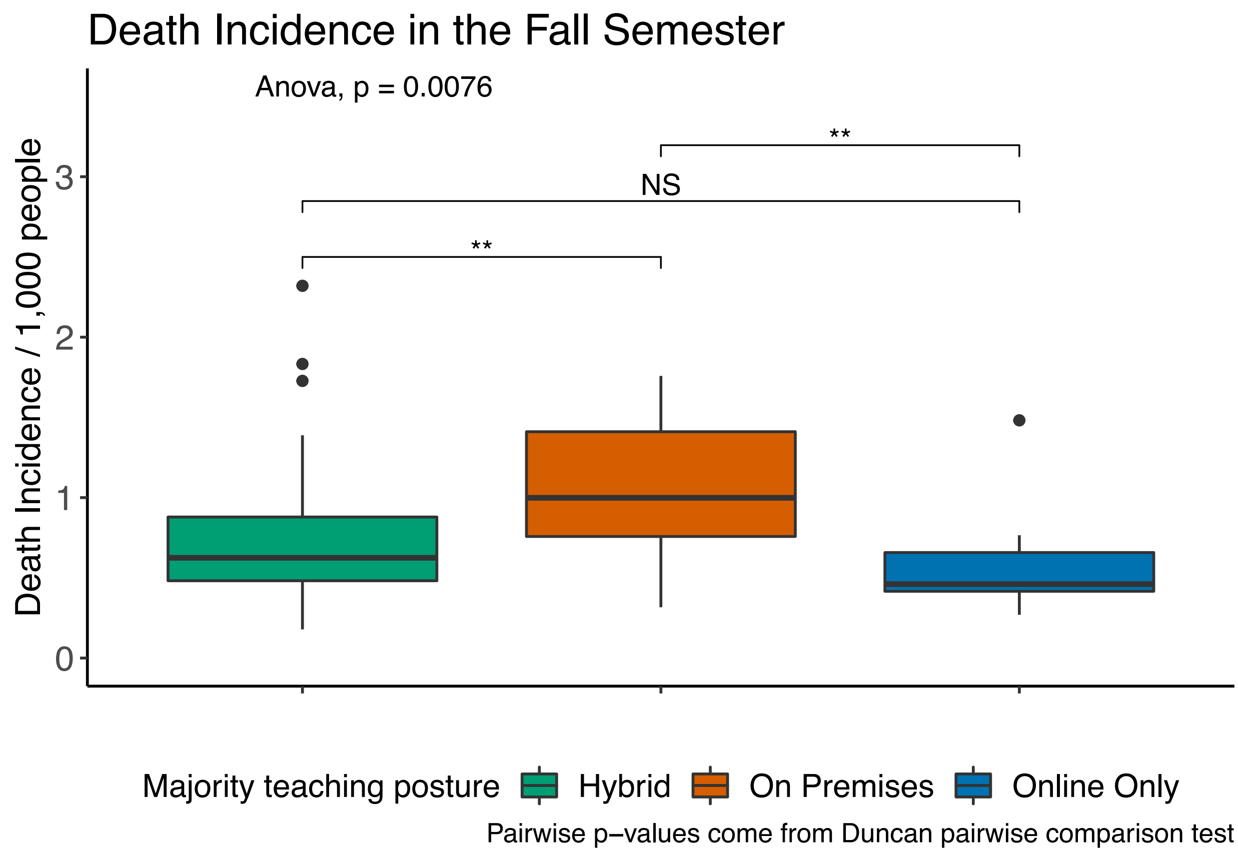
The one-way ANOVA test shows that there is at least one pair of counties grouped by majority teaching posture that have significantly different death incidences during Fall.

3.2 Pairwise test of death incidences by the majority teaching posture

Following the significant result in the one-way ANOVA test, we conduct a Duncan posthoc test to figure out which pair of teaching postures have significantly different death incidences during Fall.

```r
## Duncan test after significant ANOVA test
stat.test <- PostHocTest(fall_major_teaching.aov, method = "duncan")$major_teaching%>%
  as.data.frame()%>%
  rownames_to_column("group") %>%
  separate(group,"-", into = c("group1","group2")) %>%
  mutate(pval = round(pval,3),
         p = case_when(pval <= .01~ "**",
                       pval <= .05 ~ "*",
                       TRUE ~ "NS"))%>%
  dplyr::select(group1, group2, pval, p)

# Box Plots with test statistics
ggplot(fall_cases,aes(y = death_incidence_per_1000, x = major_teaching)) +
  geom_boxplot(aes(fill = major_teaching))+
  stat_compare_means(method = "anova")+
  stat_pvalue_manual(stat.test, label = "p",y.position = 2.5, step.increase = 0.15)+
  ylim(c(0,3.5))+
```

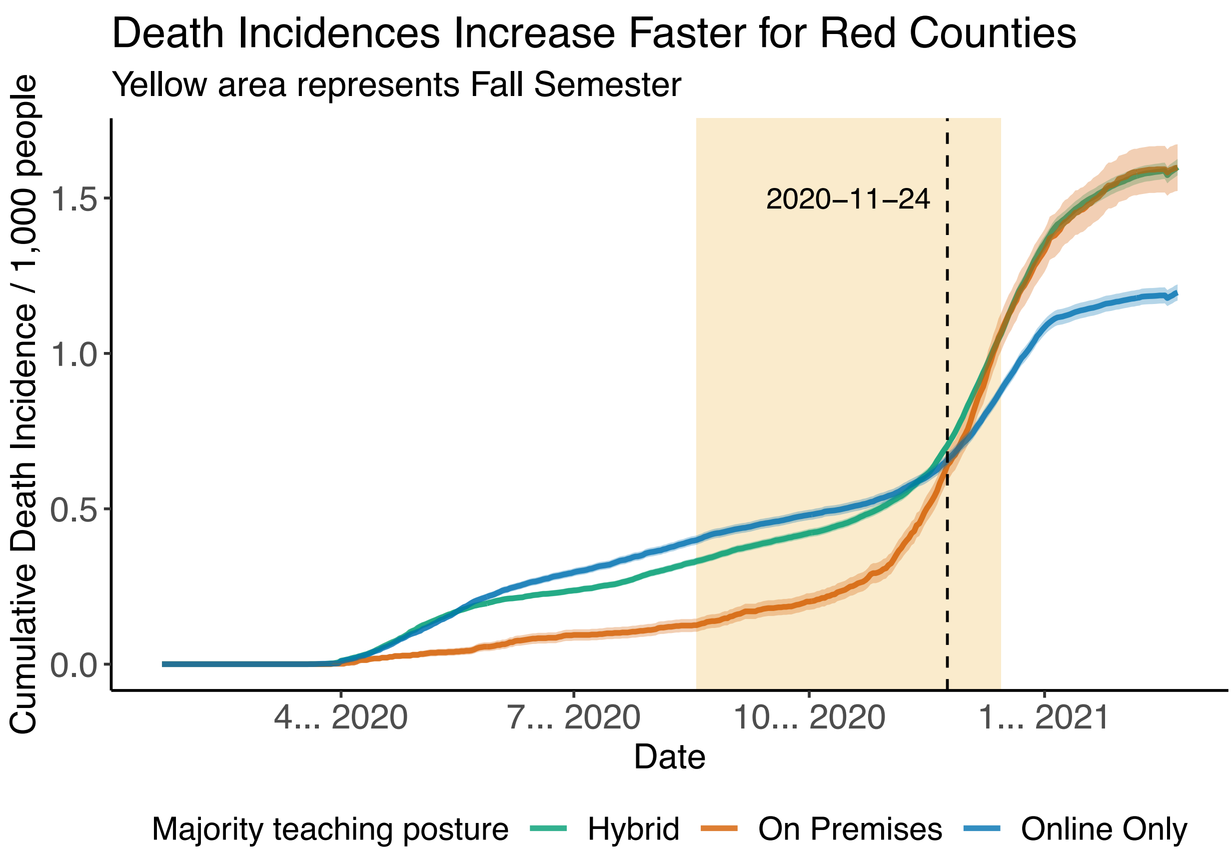
```r
  theme_bw()+
  labs(y = "Death Incidence / 1,000 people",x = "",
       fill = "Majority teaching posture",
       title = "Death Incidence in the Fall Semester",
       caption = "Pairwise p-values come from Duncan pairwise comparison test") +
  theme(legend.position = "bottom",
        axis.text.x=element_blank())+team_theme+
  scale_colour_manual(values=col_theme)+scale_fill_manual(values=col_theme)
```



3.3 Time series of death incidences by the majority teaching posture

```r
case_policy_wide%>%
  group_by(DATE, major_teaching) %>%
  drop_na(major_teaching)%>%
  summarise(total_deaths = sum(CUMDEATHS),
            total_pop = sum(POPULATION),
            death_prop = total_deaths/total_pop,
            death_prop_upper = death_prop + z_cl*sqrt(death_prop*(1 - death_prop)/total_pop),
            death_prop_lower = death_prop - z_cl*sqrt(death_prop*(1 - death_prop)/total_pop),
            .groups = "drop") %>%
  ggplot(aes(x = DATE, y = death_prop*1000, group = major_teaching))+
  geom_rect(data=case_policy_wide[1,],
            aes(xmin=as.Date("2020/08/18"), xmax=as.Date("2020/12/15"),
                ymin=-Inf,ymax=Inf),
```

```
            color = NA,alpha=0.2, show.legend = F, fill = "#E69F00") +
  geom_line(aes(color = major_teaching),size = 1, alpha = .8) +
  geom_ribbon(aes(ymin = 1000*death_prop_lower, ymax = 1000*death_prop_upper,
                  fill= major_teaching),
              alpha = .3, show.legend = F)+
  geom_vline(xintercept = date.intercept, linetype = "dashed") +
  annotate("text",x = date.intercept,y = 1.5,
           label = date.intercept,
           hjust = 1.1) +
  team_theme + theme(legend.position = "bottom")+
  ggtitle("Death Incidences Increase Faster for Red Counties ")+
  labs(x = "Date", y = "Cumulative Death Incidence / 1,000 people",
       subtitle = "Yellow area represents Fall Semester",
       color = "Majority teaching posture") +team_theme+
  scale_colour_manual(values=col_theme)+scale_fill_manual(values=col_theme)
```



Appendix 4: Confounding Variables

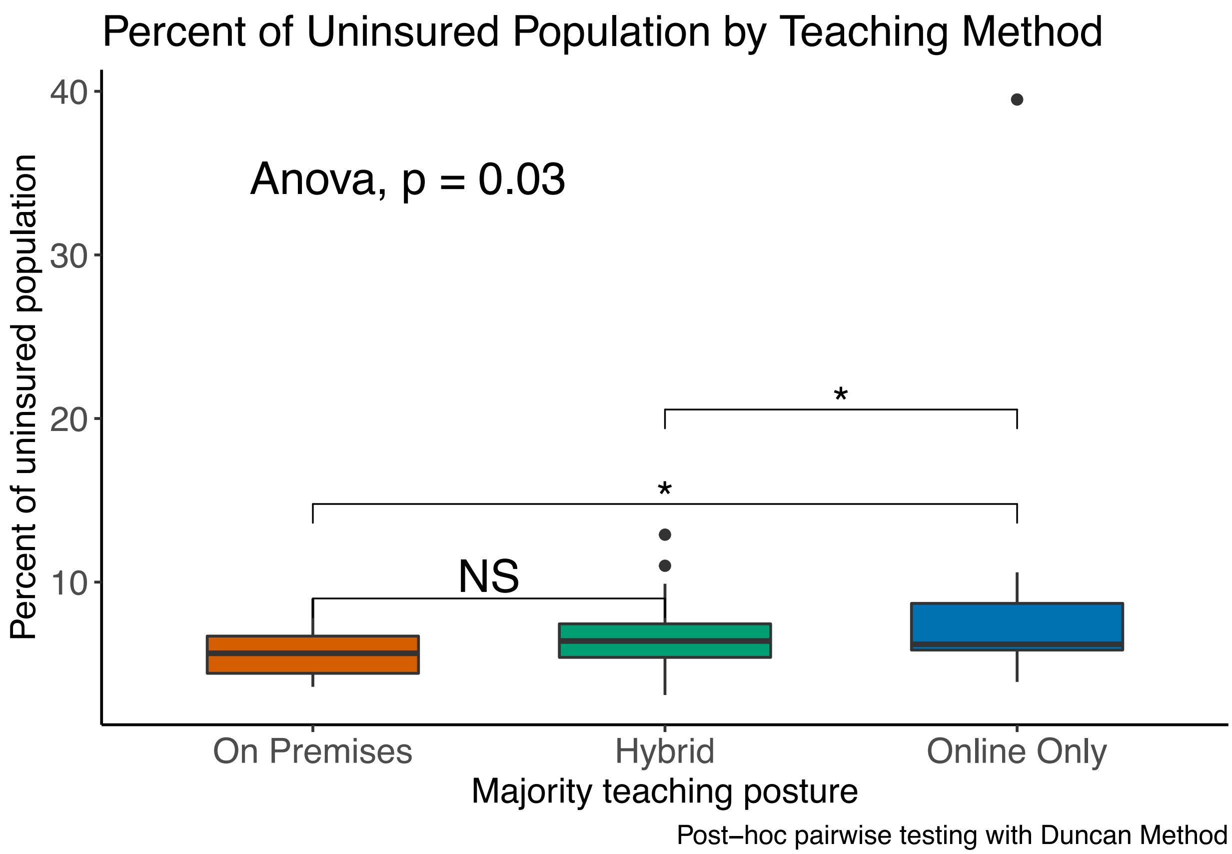
Since what we show above are results aggregated by majority teaching posture. We should notice that there could be other essential factors that are significantly different between counties taking different act in school operations.

4.1 Distribution of percent of uninsured population by majority teaching posture

```
library(ggpubr)
library(PMCMRplus)
require(DescTools)

ohio_profile <- read.csv("county_level_latest_data_for_ohio.csv")
ohio_profile <- ohio_profile[,c(1,14:20,38:50)]
names(ohio_profile)[1]<-"County"
ohio_profile$County <- toupper(ohio_profile$County)

# set up data
teaching_profile <- ohio_profile%>%
  inner_join(wide_teaching_enroll,by=c("County"="county"))
teaching_profile$major_teaching <- factor(teaching_profile$major_teaching,
                                          levels = c("On Premises","Hybrid","Online Only"))

# one-way ANOVA test
profile_major_teaching.aov <- aov(Percent.uninsured ~ major_teaching,data = teaching_profile)
summary(profile_major_teaching.aov)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## major_teaching  2  108.8   54.38   3.645 0.0304 *
## Residuals      83 1238.2   14.92
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Duncan test, p-value of .012
stat.test <- PostHocTest(profile_major_teaching.aov, method = "duncan")$major_teaching %>%
  as.data.frame()%>%
  rownames_to_column("group") %>%
  separate(group,"-", into = c("group1","group2")) %>%
  mutate(pval = round(pval,3),
         p = case_when(pval <= .01~ "**",
                       pval <= .05 ~ "*",
                       TRUE ~ "NS"))%>%
  dplyr::select(group1, group2, pval, p)

teaching_profile%>%
  ggplot(aes(x=major_teaching,y=Percent.uninsured))+
  geom_boxplot(aes(fill=major_teaching),width=0.6)+
  stat_compare_means(method = "anova",size=6,label.y.npc=0.85)+
  stat_pvalue_manual(stat.test, label = "p",y.position = 1,
                     step.increase = 0.15,size = 6,bracket.nudge.y = 8)+
  labs(title="Percent of Uninsured Population by Teaching Method",
       x="Majority teaching posture",
       y="Percent of uninsured population",
       caption = "Post-hoc pairwise testing with Duncan Method")+
  team_theme+theme(legend.position = "")+scale_fill_manual(values=col_theme)
```



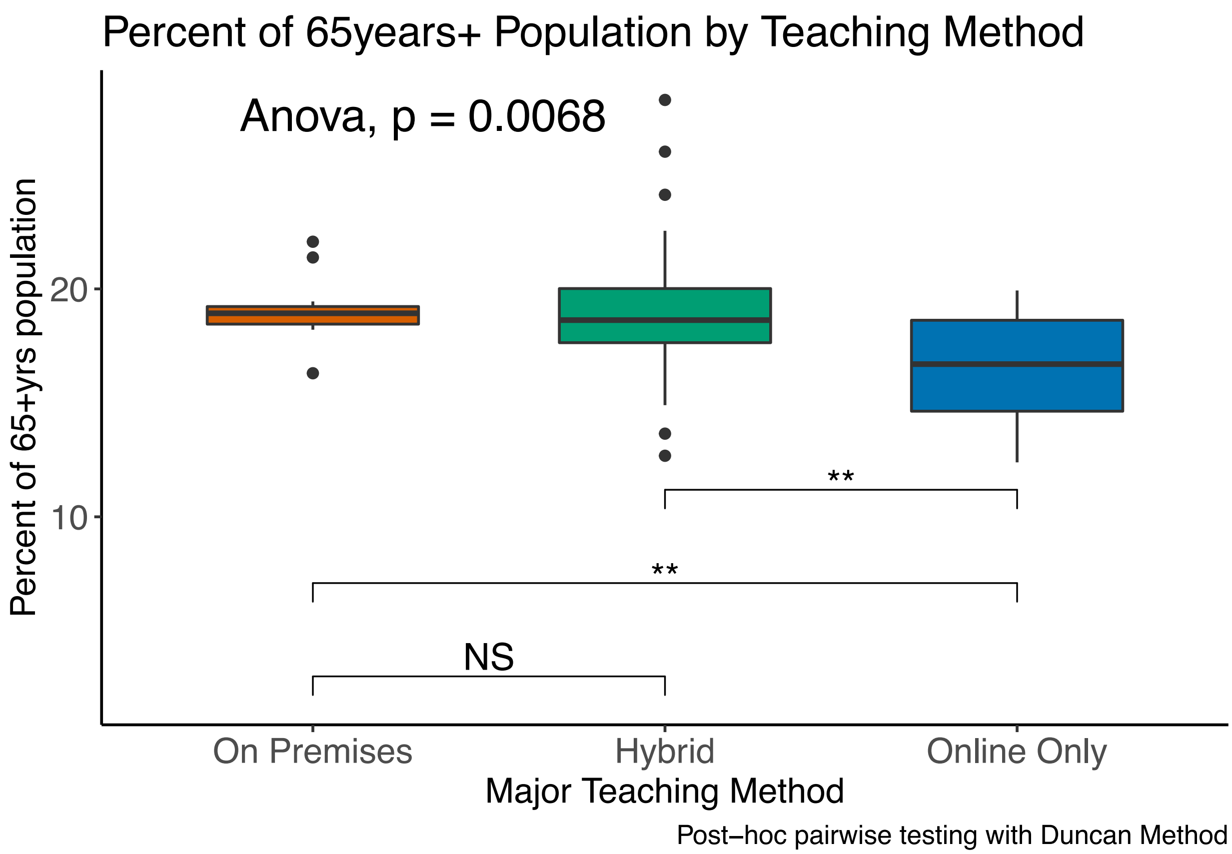
4.2 Distribution of percent of senior population by majority teaching posture

```
# one-way ANOVA
senior_major_teaching.aov <- aov(Percent.Population.65..yrs~ major_teaching,data = teaching_profile)

summary(senior_major_teaching.aov)
```

```
##                Df Sum Sq Mean Sq F value  Pr(>F)
## major_teaching  2   61.8   30.88   5.297 0.00684 **
## Residuals      83  483.9    5.83
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Duncan test: p-value of .012
stat.test <- PostHocTest(senior_major_teaching.aov, method = "duncan")$major_teaching %>%
  as.data.frame()%>%
  rownames_to_column("group") %>%
  separate(group,"-", into = c("group1","group2")) %>%
  mutate(pval = round(pval,3),
         p = case_when(pval <= .01~ "**",
                       pval <= .05 ~ "*",
                       TRUE ~ "NS"))%>%
  dplyr::select(group1, group2, pval, p)
# boxplot
teaching_profile%>%
  ggplot(aes(x=major_teaching,y=Percent.Population.65..yrs))+
```

```r
geom_boxplot(aes(fill=major_teaching),width=0.6)+
stat_compare_means(method = "anova",size=6,label.y.npc=0.95)+
stat_pvalue_manual(stat.test, label = "p",y.position = 1,
                   step.increase = 0.15,size = 5,bracket.nudge.y = 2)+
labs(title="Percent of 65years+ Population by Teaching Method",
     x="Major Teaching Method",y="Percent of 65+yrs population",
     fill="Majority teaching posture",
     caption = "Post-hoc pairwise testing with Duncan Method")+team_theme+
theme(legend.position = "")+
scale_fill_manual(values=col_theme)
```



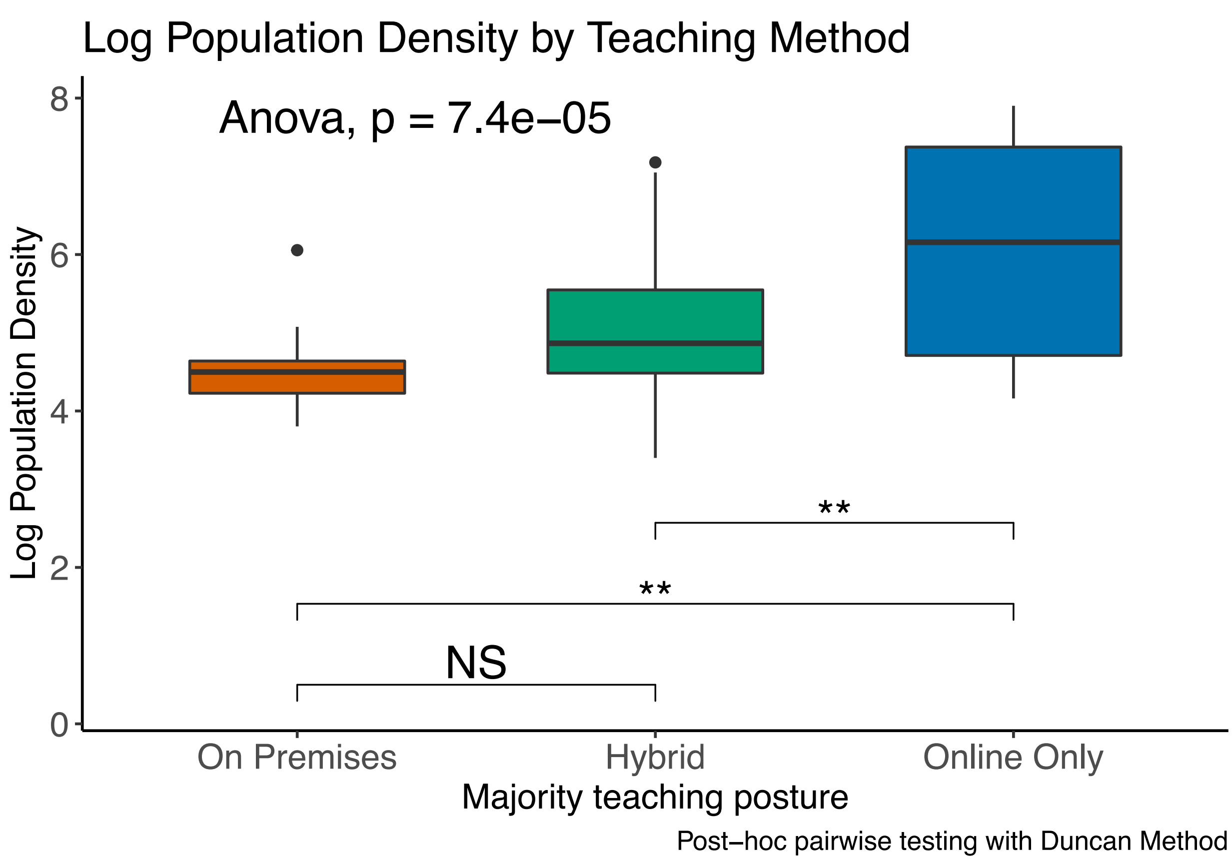
4.3 Distribution of log population density by majority teaching posture

We rescale the population density by log transformation because we want to better show the ylab.

```r
# one-way ANOVA
pop_den_major_teaching.aov <- aov(log(Population.density)~ major_teaching,data = teaching_profile)

summary(pop_den_major_teaching.aov)
```

```
##                Df Sum Sq Mean Sq F value   Pr(>F)    
## major_teaching  2  17.04   8.520   10.69 7.41e-05 ***
## Residuals      83  66.16   0.797                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Duncan test: p-value of .012
stat.test <- PostHocTest(pop_den_major_teaching.aov, method = "duncan")$major_teaching %>%
  as.data.frame()%>%
  rownames_to_column("group") %>%
  separate(group,"-", into = c("group1","group2")) %>%
  mutate(pval = round(pval,3),
         p = case_when(pval <= .01~ "**",
                       pval <= .05 ~ "*",
                       TRUE ~ "NS"))%>%
  dplyr::select(group1, group2, pval, p)

# boxplot
teaching_profile%>%
  ggplot(aes(x=major_teaching,y=log(Population.density)))+
  geom_boxplot(aes(fill=major_teaching),width=0.6)+
  stat_compare_means(method = "anova",size=6,label.y.npc=0.95)+
  stat_pvalue_manual(stat.test, label = "p",y.position = 1, step.increase = 0.15,
                     size = 6,bracket.nudge.y = -0.5)+
  labs(title="Log Population Density by Teaching Method",
       x="Majority teaching posture",
       y="Log Population Density",
       caption = "Post-hoc pairwise testing with Duncan Method")+
  team_theme+theme(legend.position = "")+scale_fill_manual(values=col_theme)
```



Appendix 5: Exponential growth model

We construct the exponential growth model to measure the state of pandemic. Please refer to the details in our Methods section.

5.1 Process growth coefficent

```
# read in the estimated coefficients
cases_slope <- read.csv("county_splines.csv", header = T)%>%
  dplyr::select(COUNTY,DATE,POPULATION,CUMDEATHS,
                log_tot_deaths,tot.slope,NEWDEATHS,rev_NEWDEATHS,
                log_new_deaths,new.slope)

# SHIFT THE DATE 24 days forward
cases_slope$DATE <- as.Date(cases_slope$DATE)-24

# get Majority teaching posture wide_teaching_enroll
cases_slope_teach <-death_teaching%>%
  dplyr::select(-DATE,-POPULATION,-CUMDEATHS,-NEWDEATHS)%>%
  distinct()%>%
  right_join(cases_slope,by=c("COUNTY"))%>%
  filter(DATE>as.Date("2020-01-23"))
write.csv(cases_slope_teach,"cases_slope_teach.csv",row.names = F)


## ordering the teaching method factor to ensure the color order
cases_slope_teach$major_teaching <- factor(cases_slope_teach$major_teaching,
                                           levels = c("On Premises","Hybrid","Online Only"))
cases_slope_teach$DATE <- as.Date(cases_slope_teach$DATE)
```

5.2 Compute the maximum B values for each county during the Fall semester

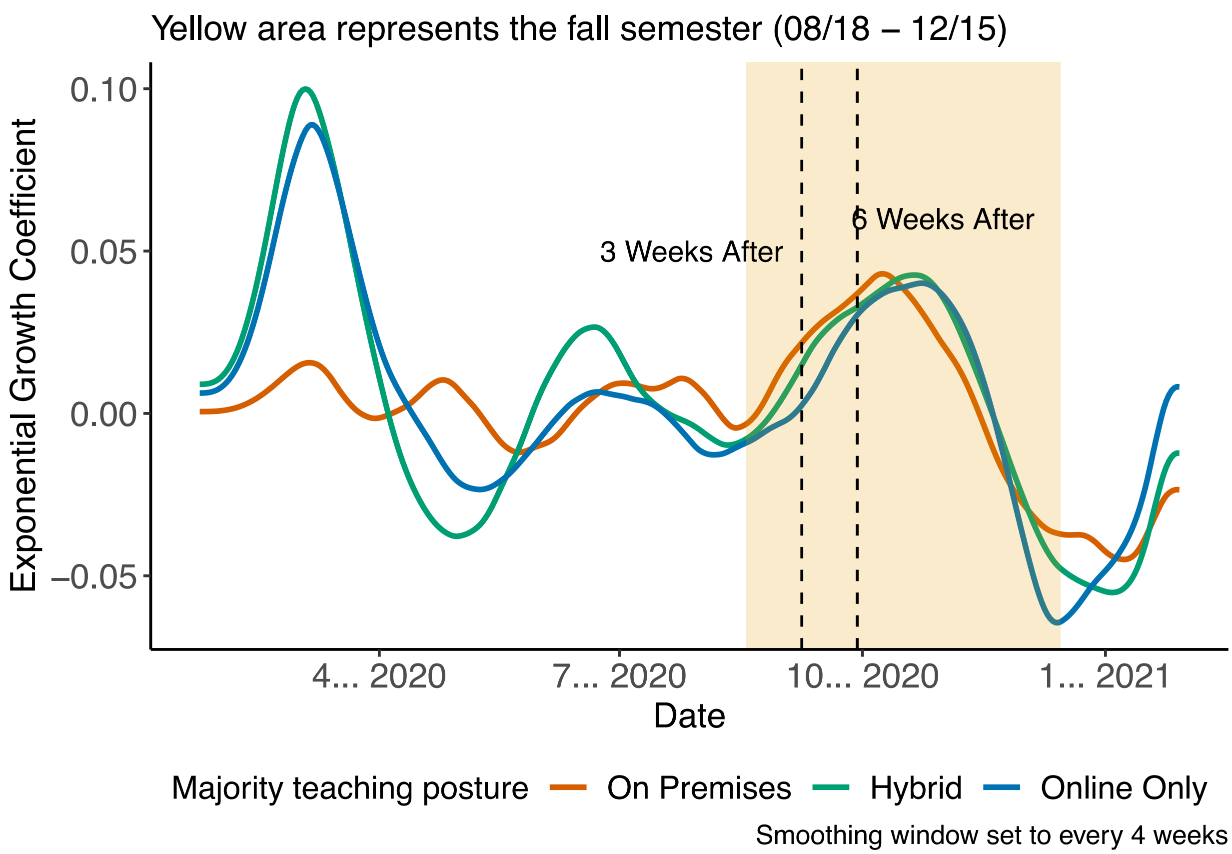
Maximum B values represent the severity of the disease for the county during the Fall semester.

```
maxB1 <- cases_slope_teach%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18") & DATE<=as.Date("2020-12-15"))%>%
  summarise(max_B1 = max(new.slope), .groups = 'drop')
avgB1 <- cases_slope_teach%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18") & DATE<=as.Date("2020-12-15"))%>%
  summarise(avg_B1 = mean(new.slope), .groups = 'drop')
## avg3w_B0
## average B0 of the first 3 weeks of school reopening
## avg1w_2w_B0
## OR average B0s between  2020-08-18 -7days and +14days
##[before the rate bounce back around the dashed line]
## avg3w_bf_B0 ## OR average B0s between  2020-08-18 -21days and 2020-08-18
##[before the rate bounce back around the dashed line]
avgB0 <- cases_slope_teach%>%
  group_by(COUNTY)%>%
  filter(DATE > as.Date("2020-08-18") & DATE<as.Date(major_opendate)+21)%>%
  summarise(avg3w_B0 = mean(new.slope), .groups = 'drop')%>%
  left_join(cases_slope_teach%>%
```

```
  group_by(COUNTY)%>%
  filter(DATE > as.Date("2020-08-18")-7 & DATE<as.Date("2020-08-18")+14)%>%
  summarise(avg1w_2w_B0 = mean(new.slope)),by="COUNTY", .groups = 'drop')%>%
  left_join(cases_slope_teach%>%
  group_by(COUNTY)%>%
  filter(DATE < as.Date("2020-08-18") & DATE>=as.Date("2020-08-18")-21)%>%
  summarise(avg3w_bf_B0 = mean(new.slope)),by="COUNTY", .groups = 'drop')
#  B0 and B1
B0B1 <- death_teaching%>%
  distinct(COUNTY,POPULATION,NCHS.Urban.Rural.Status,Population.density)%>%
  left_join(maxB1,by="COUNTY")%>%
  left_join(wide_teaching_enroll, by = c("COUNTY" = "county"))%>%
  left_join(avgB1,by="COUNTY")%>%
  left_join(avgB0,by="COUNTY") %>%
  left_join(avg_mobility,by="COUNTY")
## ordering the teaching method factor to ensure the color order
B0B1$major_teaching <- factor(B0B1$major_teaching,levels = c("On Premises","Hybrid","Online Only"))
```

5.3 Time series of growth coefficient aggregation by counties majority teaching posture

```
cases_slope_teach_agg <- cases_slope_teach %>%
  drop_na(major_teaching)%>%
  group_by(DATE, major_teaching) %>%
  summarise(total_new_deaths = sum(rev_NEWDEATHS), .groups = "drop") %>%
  mutate(log_new_deaths = log(total_new_deaths + 1)) %>%
  group_by(major_teaching) %>%
  mutate(smooth.spline = smooth.spline(DATE,log_new_deaths,df = 398/28)$y,
         B = predict(smooth.spline(DATE,log_new_deaths,df = 398/28),deriv = 1)$y,
         B2 = predict(smooth.spline(DATE,log_new_deaths,df = 398/28),deriv = 2)$y)
week3_after_start <- as.Date("2020/08/18") + 21

####
ggplot(cases_slope_teach_agg, aes(x = DATE, color = major_teaching)) +
  geom_line(aes(y = B), size = 1) +
  geom_rect(data = cases_slope_teach_agg[1,],
            aes(xmin=as.Date("2020/08/18"), xmax=as.Date("2020/12/15"),
                ymin=-Inf,ymax=Inf),
            color = NA,alpha=0.2, show.legend = F, fill = "#E69F00") +
  geom_vline(xintercept = week3_after_start, lty = 2) +
  annotate("text",label = "3 Weeks After",
           x = week3_after_start, y = .05, hjust = 1.1)+
  geom_vline(xintercept = as.Date("2020/08/18")+42, lty = 2) +
  annotate("text",label = "6 Weeks After",
           x = as.Date("2020/08/18")+130, y = .06, hjust = 1.3)+
  labs(x = "Date", y = "Exponential Growth Coefficient",
       color = "Majority teaching posture",
       caption = "Smoothing window set to every 4 weeks",
       subtitle = "Yellow area represents the fall semester (08/18 - 12/15)") +
  theme(legend.position = "bottom")+
  team_theme+scale_color_manual(values=col_theme)
```



5.4 Compute the change in growth B values and corresponding average mobility for each county during the Fall semester

```
B0w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18"))%>%
  drop_na(major_teaching)%>%
  rename(new.slope0w=new.slope)
B1w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+7)%>%
  drop_na(major_teaching)%>%
  rename(new.slope1w=new.slope)
B2w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+14)%>%
  drop_na(major_teaching)%>%
  rename(new.slope2w=new.slope)
B3w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+21)%>%
  drop_na(major_teaching)%>%
  rename(new.slope3w=new.slope)
B4w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+28)%>%
  drop_na(major_teaching)%>%
  rename(new.slope4w=new.slope)
B5w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+35)%>%
```

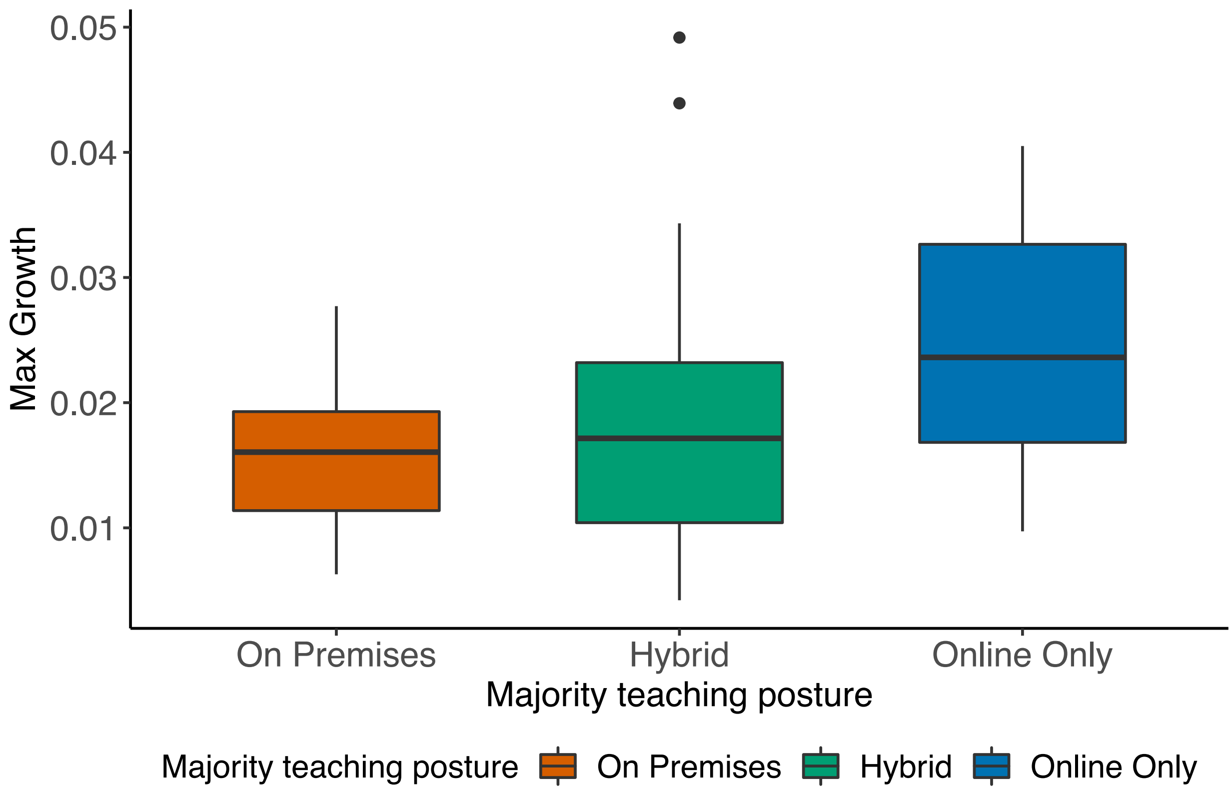
```
  drop_na(major_teaching)%>%
  rename(new.slope5w=new.slope)
B6w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+42)%>%
  drop_na(major_teaching)%>%
  rename(new.slope6w=new.slope)
B7w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")+49)%>%
  drop_na(major_teaching)%>%
  rename(new.slope7w=new.slope)

Bm1w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")-7)%>%
  drop_na(major_teaching)%>%
  rename(new.slopem1w=new.slope)
Bm2w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")-14)%>%
  drop_na(major_teaching)%>%
  rename(new.slopem2w=new.slope)
Bm3w <- cases_slope_teach%>%
  filter(DATE==as.Date("2020-08-18")-21)%>%
  drop_na(major_teaching)%>%
  rename(new.slopem3w=new.slope)

avg_mobi_0w3w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")& DATE <as.Date("2020-08-18") + 21)%>%
  summarise(avg_full_work_prob = mean(full_work_prop_7d))

avg_mobi_3w6w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")+ 21 & DATE <=as.Date("2020-08-18") + 42)%>%
  summarise(avg2_full_work_prob = mean(full_work_prop_7d))

# Before slope mobility
avg_mobi_m1w2w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")-7 & DATE <=as.Date("2020-08-18") + 14)%>%
  summarise(avg_full_work_prob_m1w2w = mean(full_work_prop_7d))

avg_mobi_m2w1w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")-14 & DATE <=as.Date("2020-08-18") + 7)%>%
  summarise(avg_full_work_prob_m2w1w = mean(full_work_prop_7d))

avg_mobi_m3w0w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")-21 & DATE <=as.Date("2020-08-18"))%>%
```

```r
  summarise(avg_full_work_prob_m3w0w = mean(full_work_prop_7d))

# After slope mobility
avg_mobi_1w4w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")+7 & DATE <=as.Date("2020-08-18")+28)%>%
  summarise(avg_full_work_prob_1w4w = mean(full_work_prop_7d))

avg_mobi_2w5w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")+14 & DATE <=as.Date("2020-08-18")+35)%>%
  summarise(avg_full_work_prob_2w5w = mean(full_work_prop_7d))

avg_mobi_4w7w <- case_mobility%>%
  left_join(major_reopening,by=c("COUNTY"))%>%
  group_by(COUNTY)%>%
  filter(DATE >= as.Date("2020-08-18")+28 & DATE <=as.Date("2020-08-18")+49)%>%
  summarise(avg_full_work_prob_4w7w = mean(full_work_prop_7d))

# Construct B_diff
B_diff <- B6w[,c(1:9,13,20)]%>%
  left_join(B3w%>%dplyr::select(COUNTY,new.slope3w),by="COUNTY")%>%
  left_join(B0w%>%dplyr::select(COUNTY,new.slope0w),by="COUNTY")%>%
  left_join(B1w%>%dplyr::select(COUNTY,new.slope1w),by="COUNTY")%>%
  left_join(B2w%>%dplyr::select(COUNTY,new.slope2w),by="COUNTY")%>%
  left_join(B4w%>%dplyr::select(COUNTY,new.slope4w),by="COUNTY")%>%
  left_join(B5w%>%dplyr::select(COUNTY,new.slope5w),by="COUNTY")%>%
  left_join(B7w%>%dplyr::select(COUNTY,new.slope7w),by="COUNTY")%>%
  left_join(Bm1w%>%dplyr::select(COUNTY,new.slopem1w),by="COUNTY")%>%
  left_join(Bm2w%>%dplyr::select(COUNTY,new.slopem2w),by="COUNTY")%>%
  left_join(Bm3w%>%dplyr::select(COUNTY,new.slopem3w),by="COUNTY")%>%
  mutate(new.slope.diff = new.slope3w-new.slope0w,
         new.slope.diff2 = new.slope6w-new.slope3w,
         new.slope.diff2m1 = new.slope2w-new.slopem1w,
         new.slope.diff1m2 = new.slope1w-new.slopem2w,
         new.slope.diff0m3 = new.slope0w-new.slopem3w,
         new.slope.diff52 = new.slope5w-new.slope2w,
         new.slope.diff41 = new.slope4w-new.slope1w,
         new.slope.diff74 = new.slope7w-new.slope4w)%>%
  left_join(avg_mobi_0w3w,by="COUNTY")%>%
  left_join(avg_mobi_3w6w,by="COUNTY")%>%
  left_join(avg_mobi_m1w2w,by="COUNTY")%>%
  left_join(avg_mobi_m2w1w,by="COUNTY")%>%
  left_join(avg_mobi_m3w0w,by="COUNTY")%>%
  left_join(avg_mobi_1w4w,by="COUNTY")%>%
  left_join(avg_mobi_2w5w,by="COUNTY")%>%
  left_join(avg_mobi_4w7w,by="COUNTY")
B_diff$major_teaching <- factor(B_diff$major_teaching,
                                levels = c("On Premises","Hybrid","Online Only"))
```

5.5 Distribution of maximum growth coefficient for all counties by majority teaching posture

Online-only counties have highest max B value, which is not what we expected.

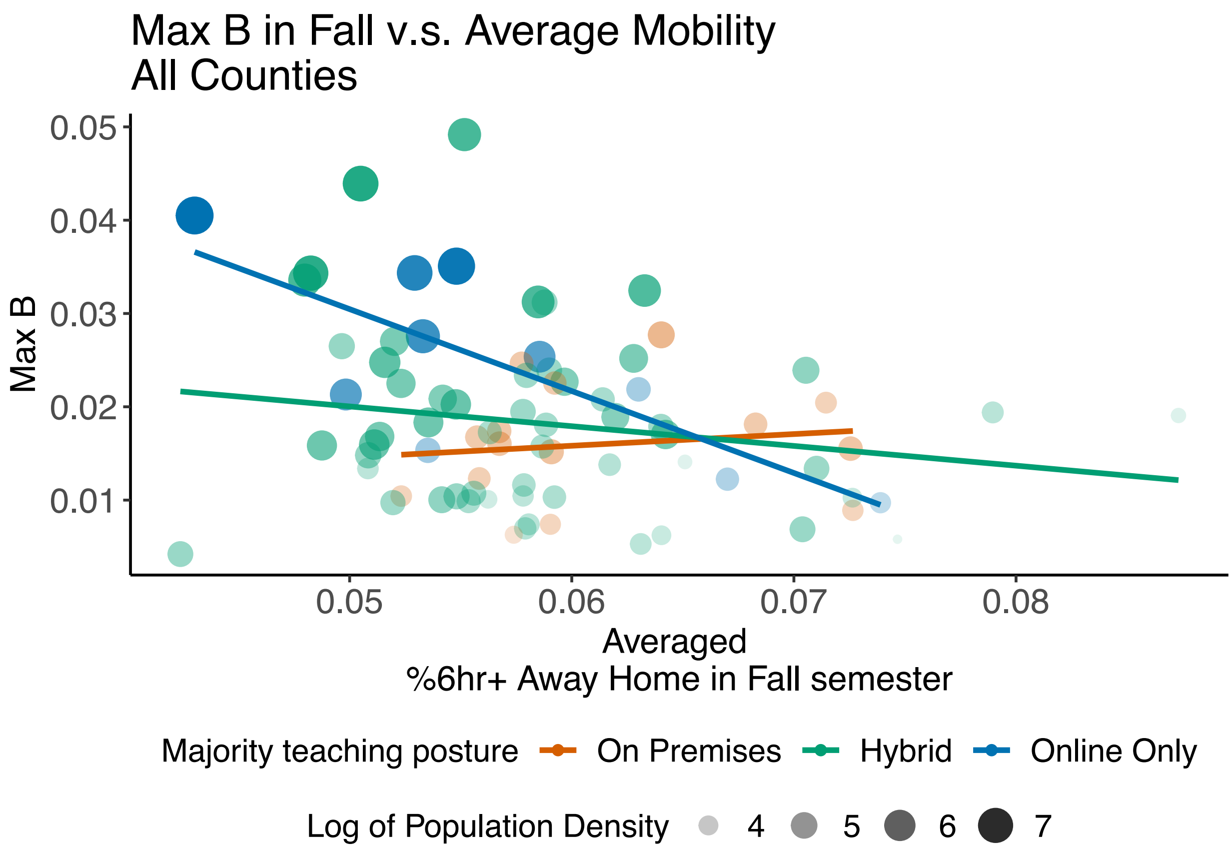
```
na.omit(B0B1)%>%
  ggplot(aes(x=major_teaching,y=max_B1))+geom_boxplot(aes(fill=major_teaching),width=0.6)+
  labs(title="",x="Majority teaching posture",y="Max Growth",fill="Majority teaching posture")+
  team_theme+theme(legend.position = "bottom")+
  scale_fill_manual(values=col_theme)
```



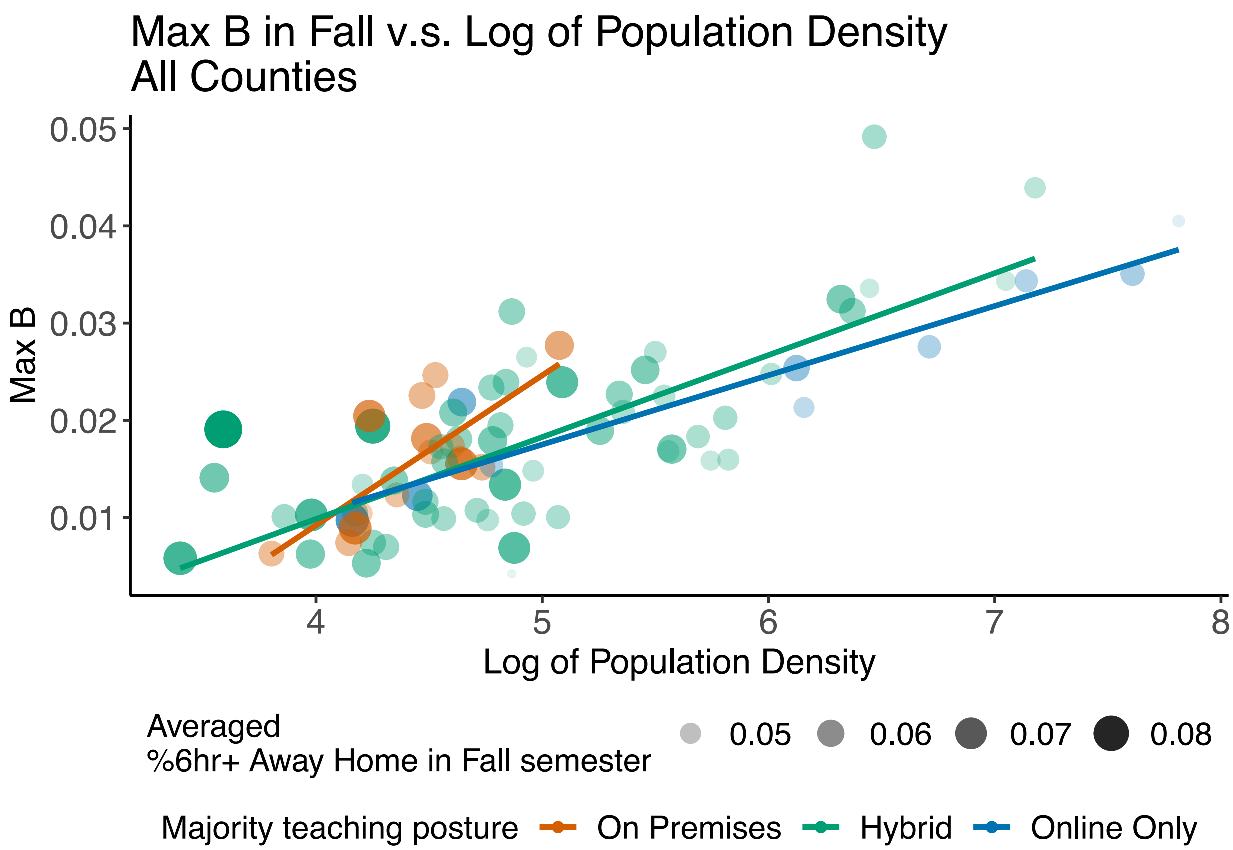
5.6 Maximum growth coefficient vs. mobility and population density for all counties

```
na.omit(B0B1)%>%
  drop_na(major_teaching)%>%
  ggplot(aes(x=avg_full_work_prob,y=max_B1,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=log(Population.density),alpha=log(Population.density)))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  labs(y="Max B",
       x="Averaged \n%6hr+ Away Home in Fall semester",
       title="Max B in Fall v.s. Average Mobility \nAll Counties",
       color="Majority teaching posture",
       size = "Log of Population Density",
       alpha= "Log of Population Density" )+
  team_theme+theme(legend.position = "bottom")+
```

```r
  scale_color_manual(values=col_theme)
```

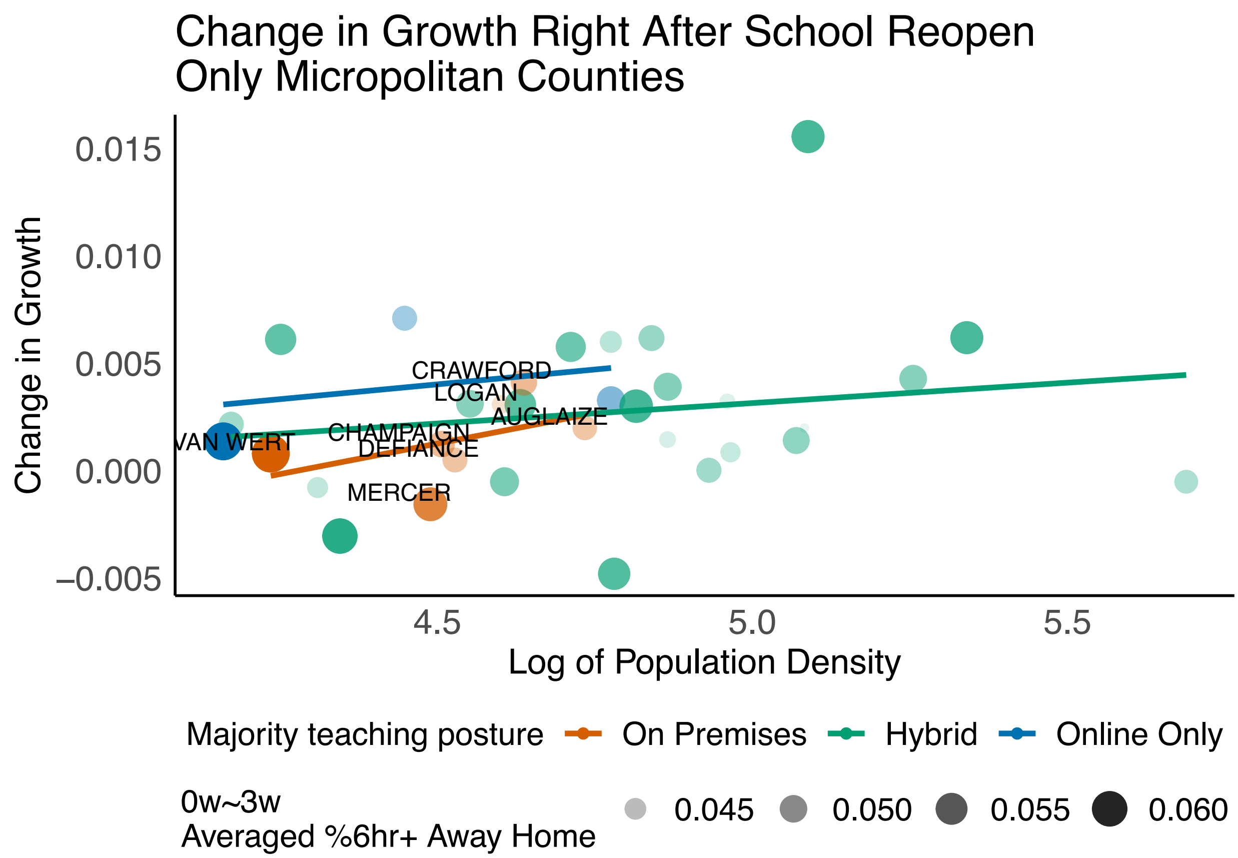



```r
na.omit(BOB1)%>%
  drop_na(major_teaching)%>%
  ggplot(aes(x=log(Population.density),y=max_B1,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob,alpha=avg_full_work_prob))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  labs(y="Max B",x="Log of Population Density",
       title="Max B in Fall v.s. Log of Population Density \nAll Counties",
       color="Majority teaching posture",
       size = "Averaged \n%6hr+ Away Home in Fall semester",
       alpha= "Averaged \n%6hr+ Away Home in Fall semester" )+
  team_theme+theme(legend.position = "bottom")+
  scale_color_manual(values=col_theme)
```

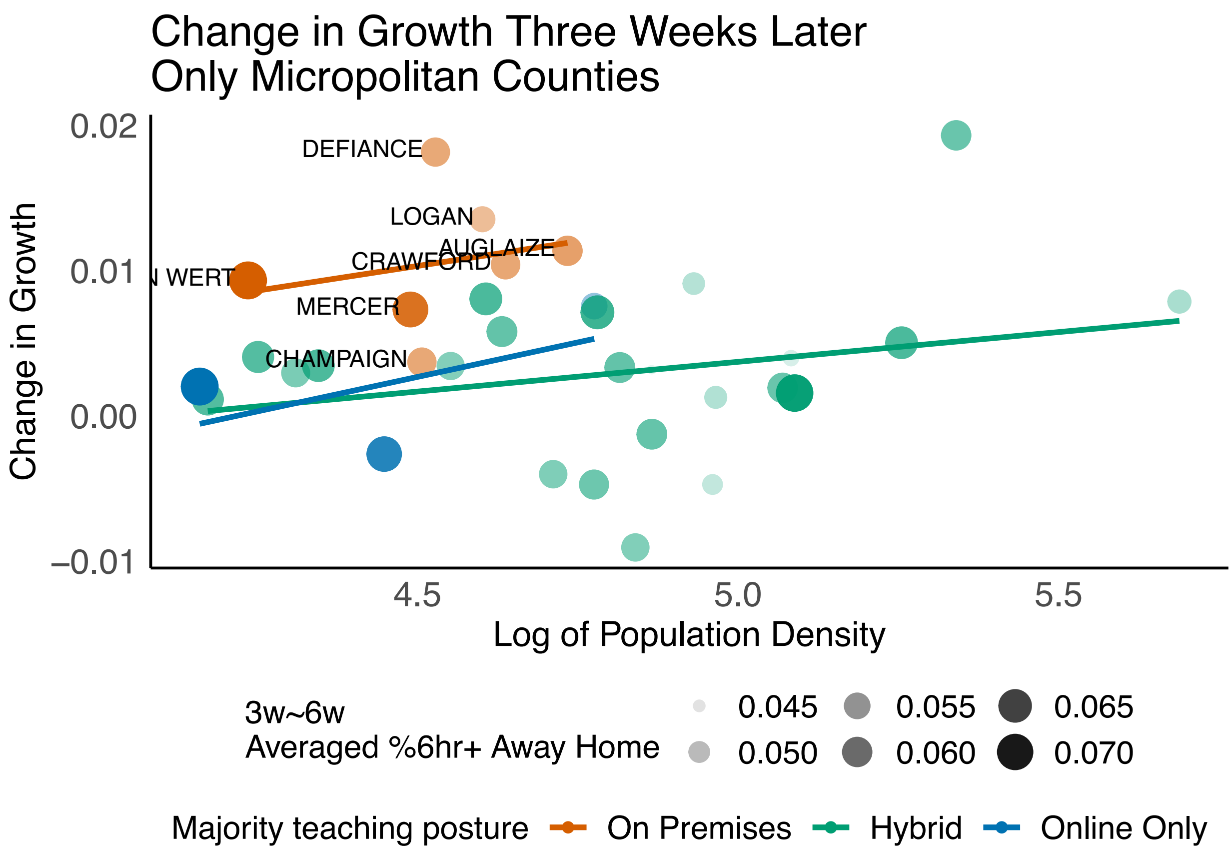


5.7 Change in growth vs. mobility and population density for all counties

```
B_diff_micro <- B_diff%>%
  drop_na(major_teaching)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan") %>%
  mutate(acc = new.slope.diff2 - new.slope.diff)
#At start of reopen 3w-0w
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff,group=major_teaching,
             color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob,alpha=avg_full_work_prob))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth Right After School Reopen\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "0w~3w\nAveraged %6hr+ Away Home",
       alpha= "0w~3w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+
  theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
              filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',size=3,hjust=0.8, vjust=-0.2)
```




```
#After reopen for 6w-3w
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff2,
             group=major_teaching,
             color=major_teaching))+
  geom_point(aes(size=avg2_full_work_prob,alpha=avg2_full_work_prob))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth Three Weeks Later\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "3w~6w\nAveraged %6hr+ Away Home",
       alpha= "3w~6w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
             filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',size=3,hjust=1.1, vjust=0.3)
```

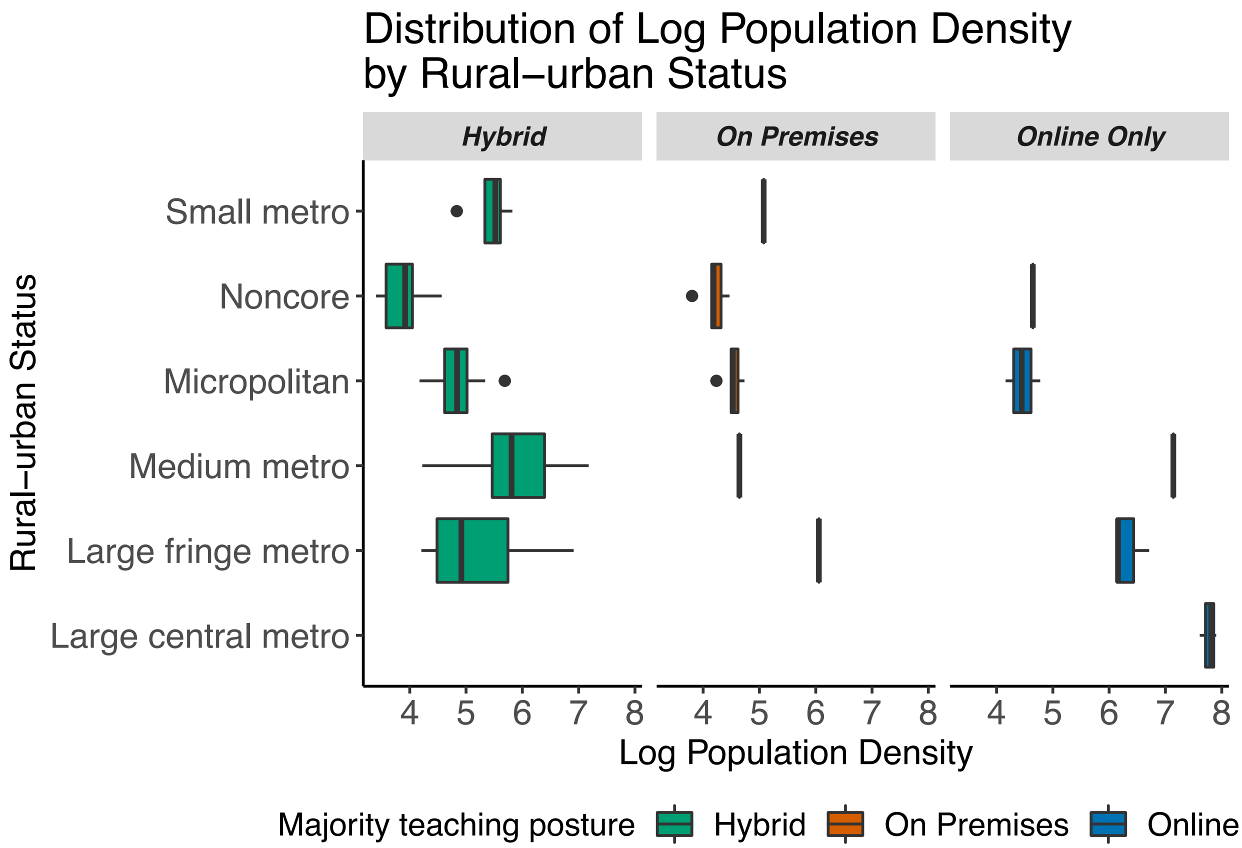


Appendix 6: Micropolitan Counties

Here we found that the rural-urban status is differentitating three groups of counties. Thus we block on the Micropolitan counties where the three groups are most comparable to reduce the confounding effect of urban-rural status.

6.1 Distribution of log population density by rural-Urban status and majority teaching posture

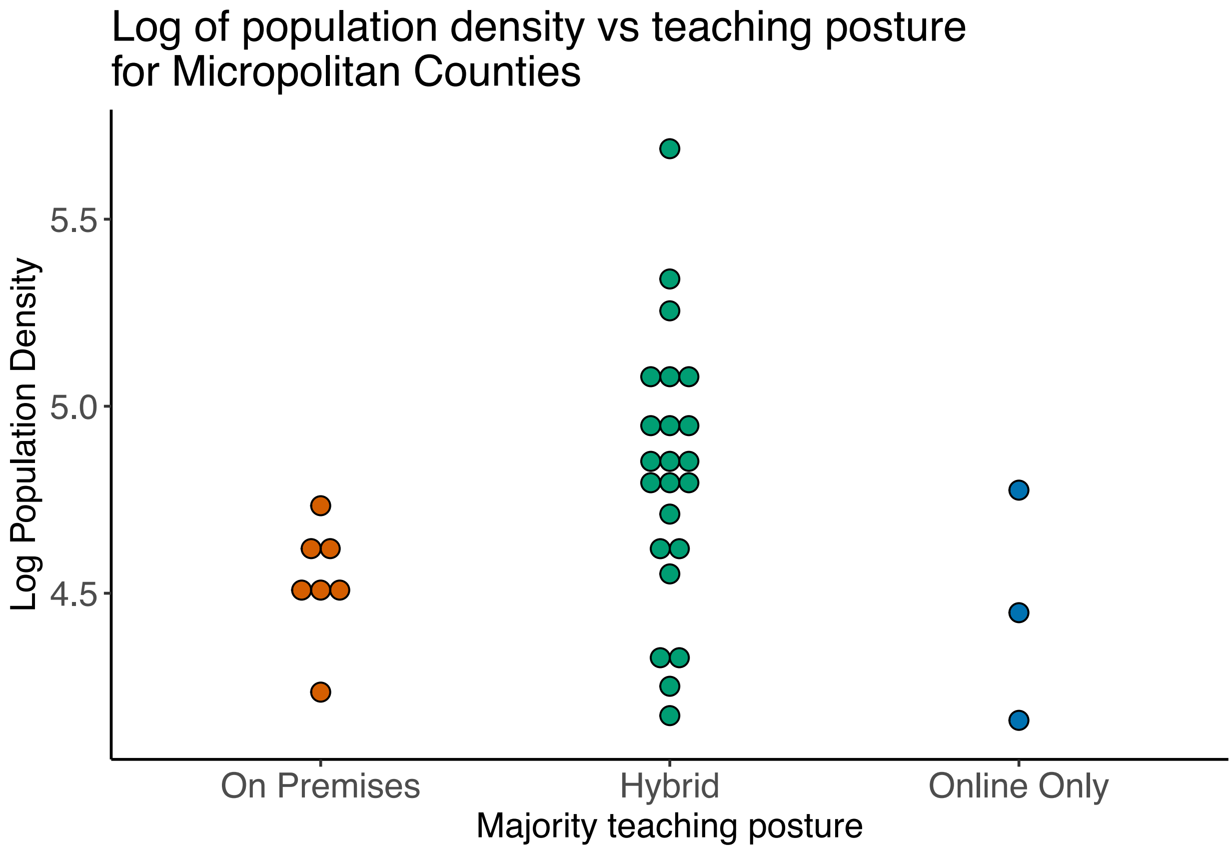
```
# Pop density vs RURAL-Urban status
ohio_profile%>%
  left_join(wide_teaching_enroll[,c("county","major_teaching")],
            by = c("County" = "county"))%>%
  drop_na(major_teaching)%>%
  ggplot(aes(y=NCHS.Urban.Rural.Status,x=log(Population.density),
             fill=major_teaching))+
  facet_grid(~major_teaching)+
  geom_boxplot()+
  labs(fill="Majority teaching posture",
       size="Averaged \n%6hr+ Away Home",
       title="Distribution of Log Population Density \nby Rural-urban Status",
       x="Log Population Density", y= "Rural-urban Status")+
  team_theme+
  scale_fill_manual(values=col_theme)+
  theme(legend.position = "bottom")
```



6.2 Distribution of log population density majority teaching postue in Micropolitan counties

Here we can see that in the micropolitan counties, online-only and on-premises ones have similar population density.

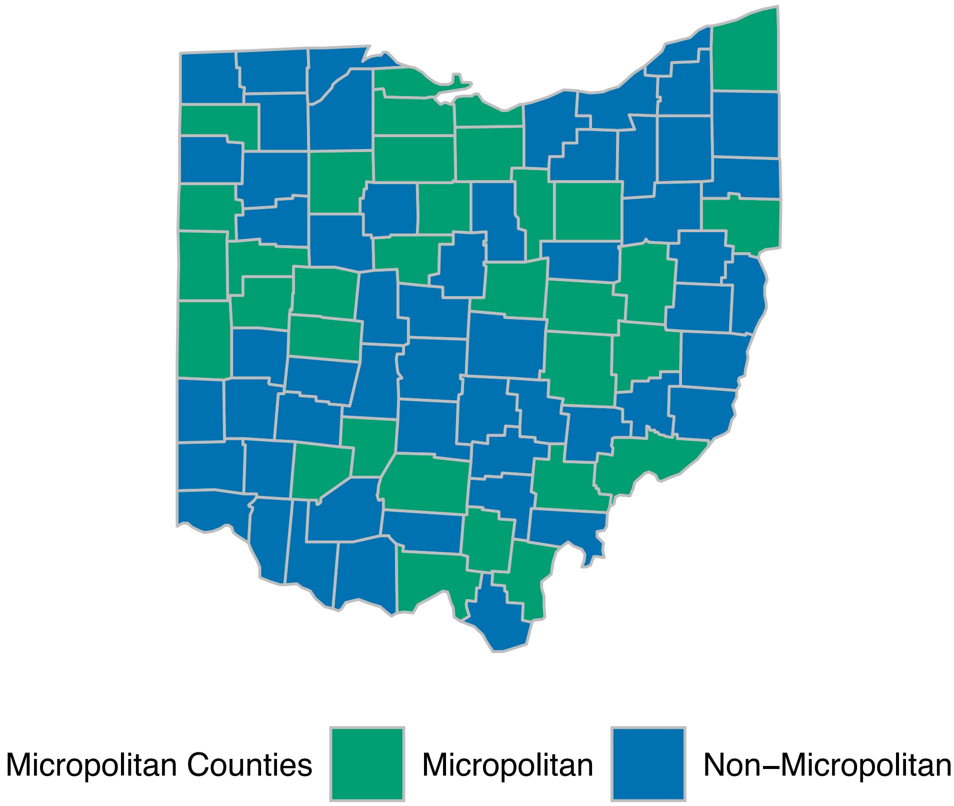
```
ohio_profile%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan")%>%
 left_join(BOB1%>%dplyr::select(COUNTY,major_teaching),by=c("County"="COUNTY"))%>%
  ggplot(aes(x=major_teaching))+
  geom_dotplot(aes(y=log(Population.density),fill=major_teaching),
             binaxis='y', stackdir='center')+
  team_theme+guides(fill=FALSE)+
  labs(y="Log Population Density",x="Majority teaching posture",
       title="Log of population density vs teaching posture\nfor Micropolitan Counties")+
  scale_fill_manual(values=col_theme)
```



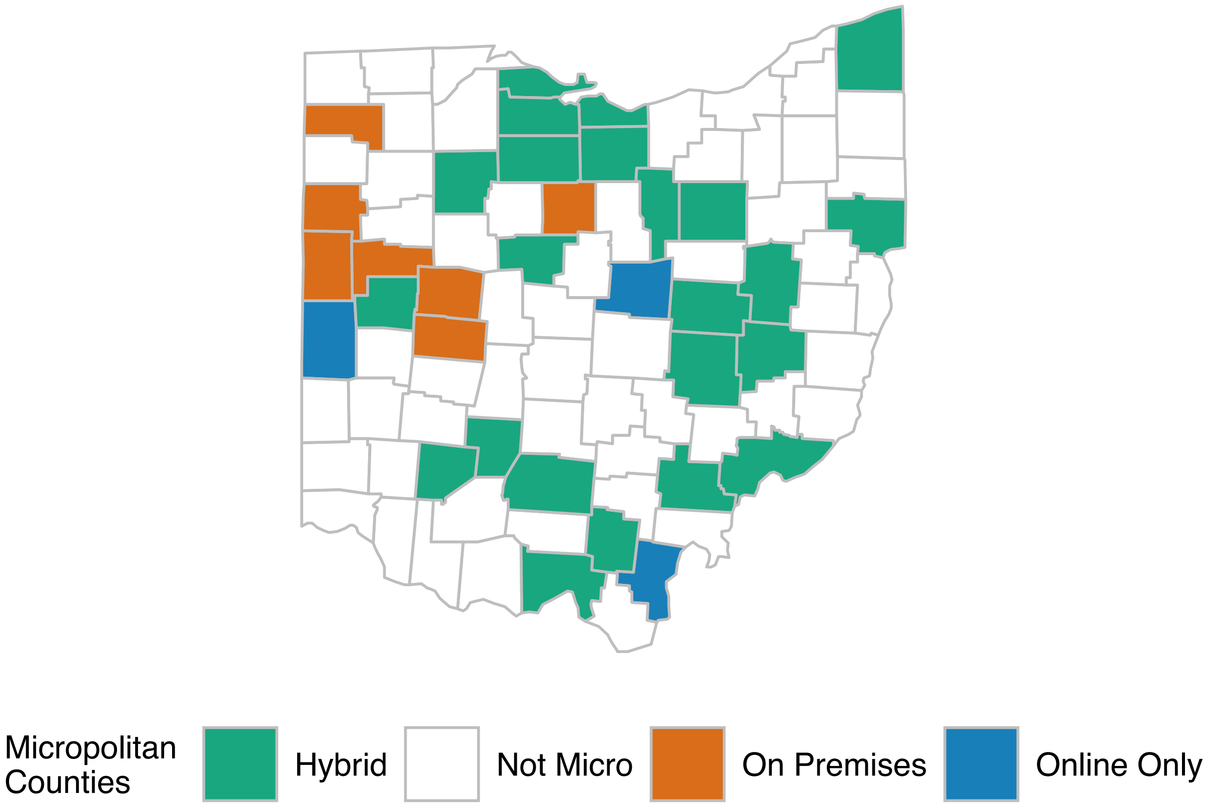
6.3 Geographical distribution of counties' Micropolitan status

We can see that micropolitan counties are well-spread across the state.

```
# Map of Micropolitan status
ohio_profile%>%
  distinct(County,NCHS.Urban.Rural.Status) %>%
  mutate(is_micro = factor(ifelse(NCHS.Urban.Rural.Status == "Micropolitan",
                                  "Micropolitan","Non-Micropolitan")))%>%
  left_join(ohio_map,by=c('County'='county'))%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = is_micro),
               color = "gray") +
  coord_fixed(1.3) + theme_map() +
  scale_fill_manual(values = c("Non-Micropolitan" ="#0072B2",
                    "Micropolitan" ="#009E73"))+
  labs(fill='Micropolitan Counties')+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        legend.position = "bottom",
        legend.key.size = unit(2,"lines"))
```




```r
# Map of majority teaching posture in Micropolitan counties
ohio_profile%>%
  distinct(County,NCHS.Urban.Rural.Status) %>%
  left_join(wide_teaching_enroll[,c("county","major_teaching")],
          by = c("County" = "county"))%>%
  mutate(is_micro = factor(ifelse(NCHS.Urban.Rural.Status == "Micropolitan",1,0)),
         micro_teach = factor(ifelse(is_micro == 1, major_teaching, "Not Micro")))%>%
  left_join(ohio_map,by=c('County'='county'))%>%
  ggplot() +
  geom_polygon(aes(x = long, y = lat, group = group, fill = micro_teach),
               color = "gray",alpha=0.9) +
  coord_fixed(1.3) + theme_map() +
  scale_fill_manual(values = c(col_theme, "Not Micro" = "white"))+
  labs(fill='Micropolitan \nCounties')+
  theme(legend.text = element_text(size=12),
        legend.title = element_text(size=12),
        legend.position = "bottom",
        legend.key.size = unit(2,"lines"))
```



6.4 Distribution of maximum growth ccoefficient in Micropolitan counties

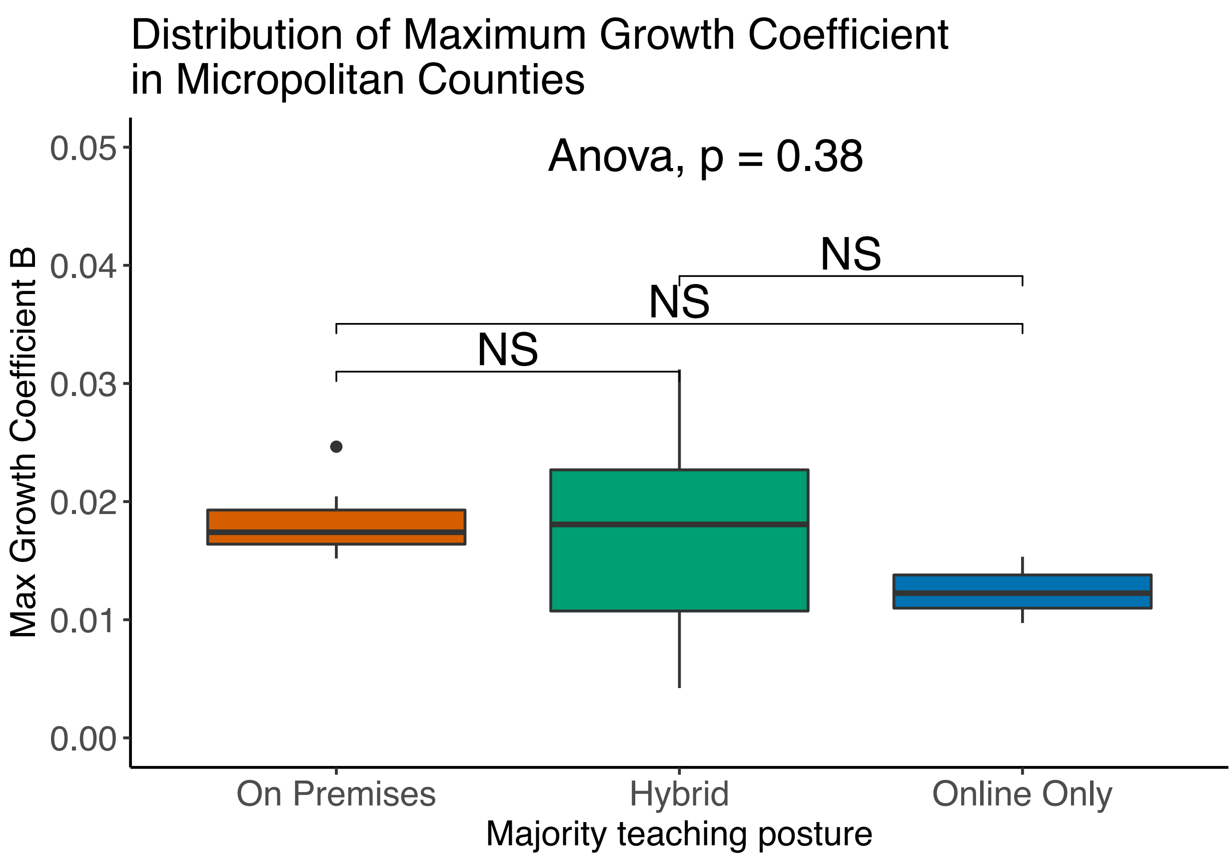
On Premises counties have significant higher maximum growth coefficient than online only counties.

```
# one-way ANOVA
maxB_major_teaching.aov <- aov(max_B1 ~ major_teaching,data = na.omit(BOB1)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan"))
summary(maxB_major_teaching.aov)
```

```
##                Df    Sum Sq   Mean Sq F value Pr(>F)
## major_teaching  2 0.0000764 3.821e-05   1.002   0.38
## Residuals      28 0.0010679 3.814e-05
```

```
# Duncan test
stat.test <- PostHocTest(maxB_major_teaching.aov, method = "duncan")$major_teaching %>%
  as.data.frame()%>%
  rownames_to_column("group") %>%
  separate(group,"-", into = c("group1","group2")) %>%
  mutate(pval = round(pval,3),
         p = case_when(pval <= .01~ "**",
                       pval <= .05 ~ "*",
                       TRUE ~ "NS"))%>%
  dplyr::select(group1, group2, pval, p)

# boxplot
na.omit(BOB1)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan")%>%
```

```
  drop_na(major_teaching)%>%
  ggplot(aes(x=major_teaching,y=max_B1))+
  geom_boxplot(aes(fill=major_teaching))+
  ylim(c(0,0.05))+
  stat_compare_means(method = "anova",size=6,label.y.npc=0.96,label.x.npc = 0.4)+
  stat_pvalue_manual(stat.test, label = "p",y.position = 0.03,
                     step.increase = 0.15,size = 6,bracket.nudge.y = 0.001)+
  team_theme+
  theme(legend.position = " ")+
  labs(y="Max Growth Coefficient B",
       x="Majority teaching posture",
       title="Distribution of Maximum Growth Coefficient \nin Micropolitan Counties",
       fill="Majority teaching posture")+
  scale_fill_manual(values=col_theme)
```

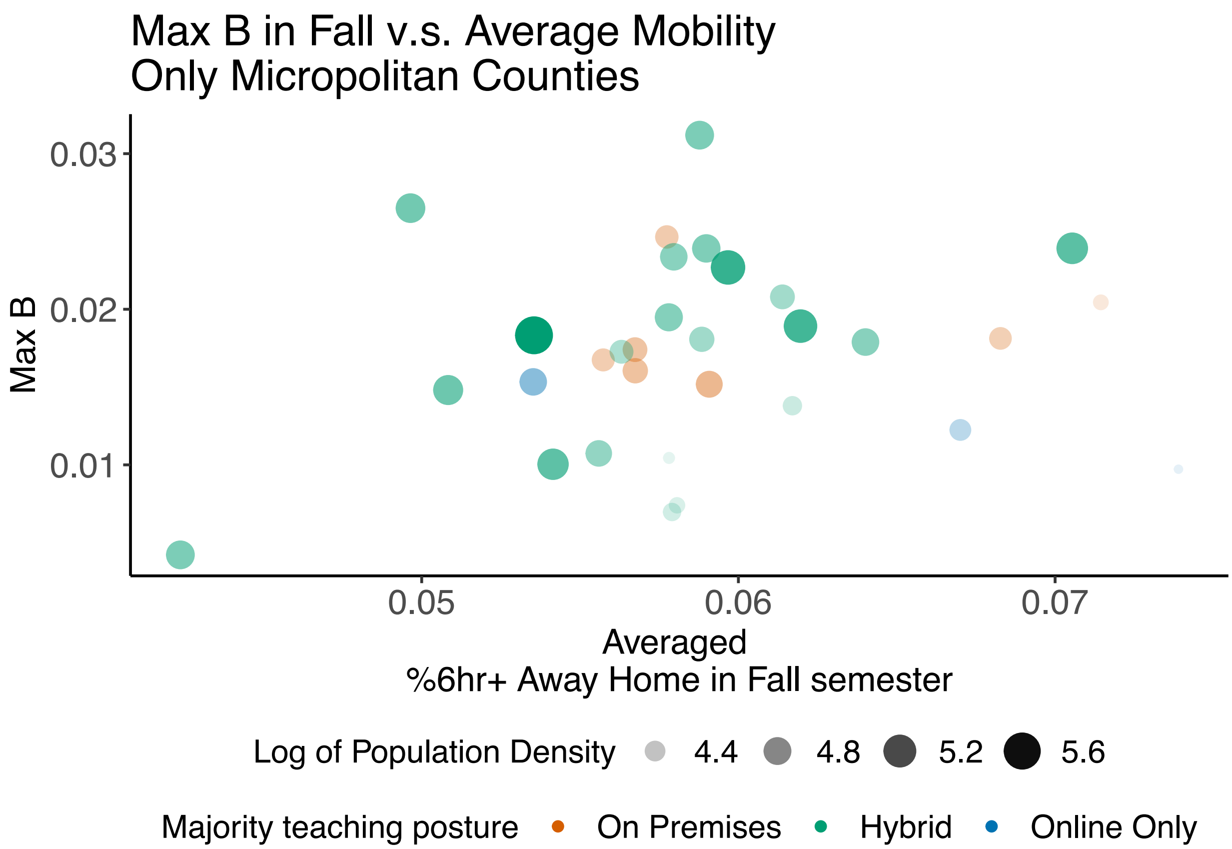


6.5 Maximum B vs. mobility and population density in Micropolitan counties

We can see the relationship between severity of the disease and mobility or population density become unclear after we block the data on its Micropolitan status and the sample size is limited.

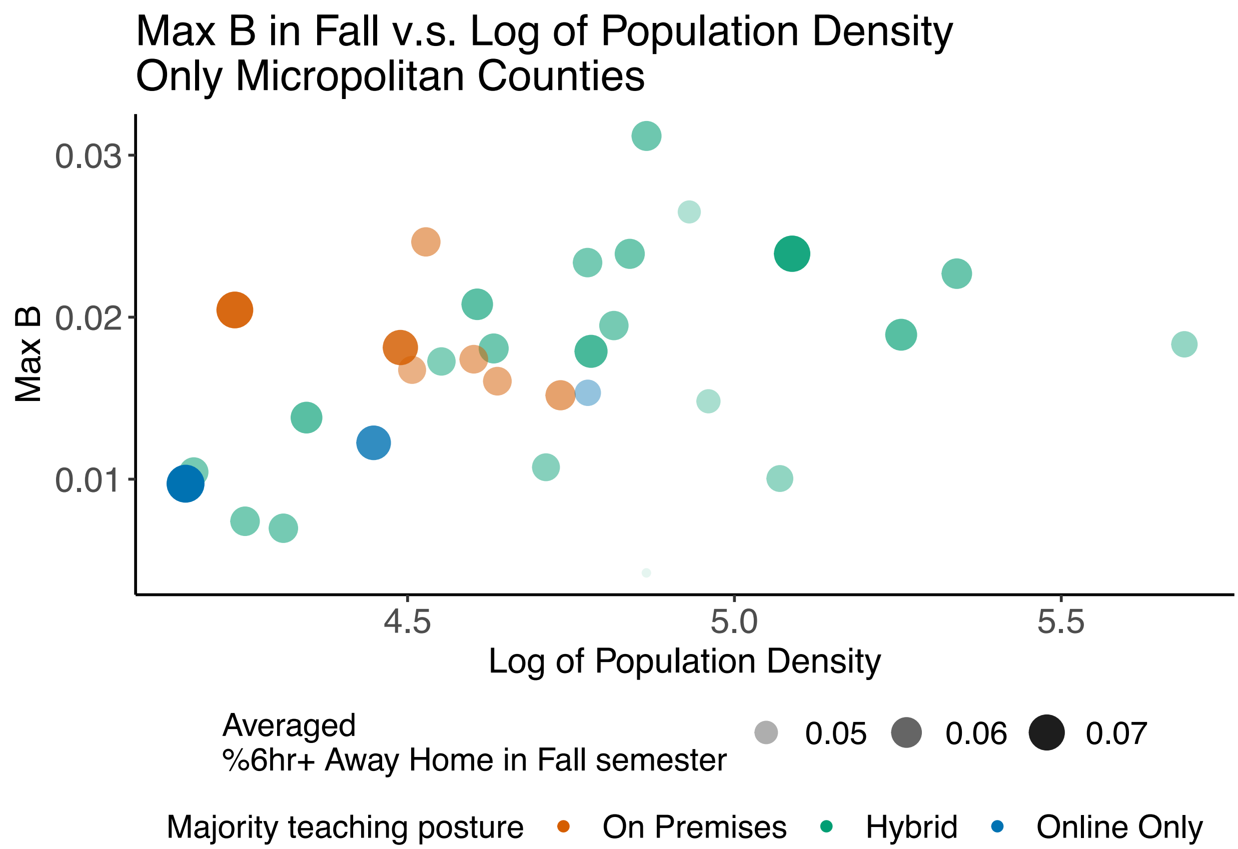
```
na.omit(BOB1)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan")%>%
  drop_na(major_teaching)%>%
  ggplot(aes(x=avg_full_work_prob,y=max_B1,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=log(Population.density),alpha=log(Population.density)))+
  #geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
```

```r
  labs(y="Max B",x="Averaged \n%6hr+ Away Home in Fall semester",
       title="Max B in Fall v.s. Average Mobility \nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "Log of Population Density",
       alpha= "Log of Population Density" )+
  team_theme+theme(legend.position = "bottom")+
  scale_color_manual(values=col_theme)+
  guides(
    size = guide_legend(order = 1),
    alpha = guide_legend(order = 1),
    fill = guide_legend(order = 0)
  )
```




```r
na.omit(BOB1)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan")%>%
  drop_na(major_teaching)%>%
  ggplot(aes(x=log(Population.density),
             y=max_B1,group=major_teaching,
             color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob,alpha=avg_full_work_prob))+
  #geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  labs(y="Max B",x="Log of Population Density",
       title="Max B in Fall v.s. Log of Population Density \nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "Averaged \n%6hr+ Away Home in Fall semester",
       alpha= "Averaged \n%6hr+ Away Home in Fall semester" )+
```

```
  team_theme+theme(legend.position = "bottom")+
  scale_color_manual(values=col_theme)
```



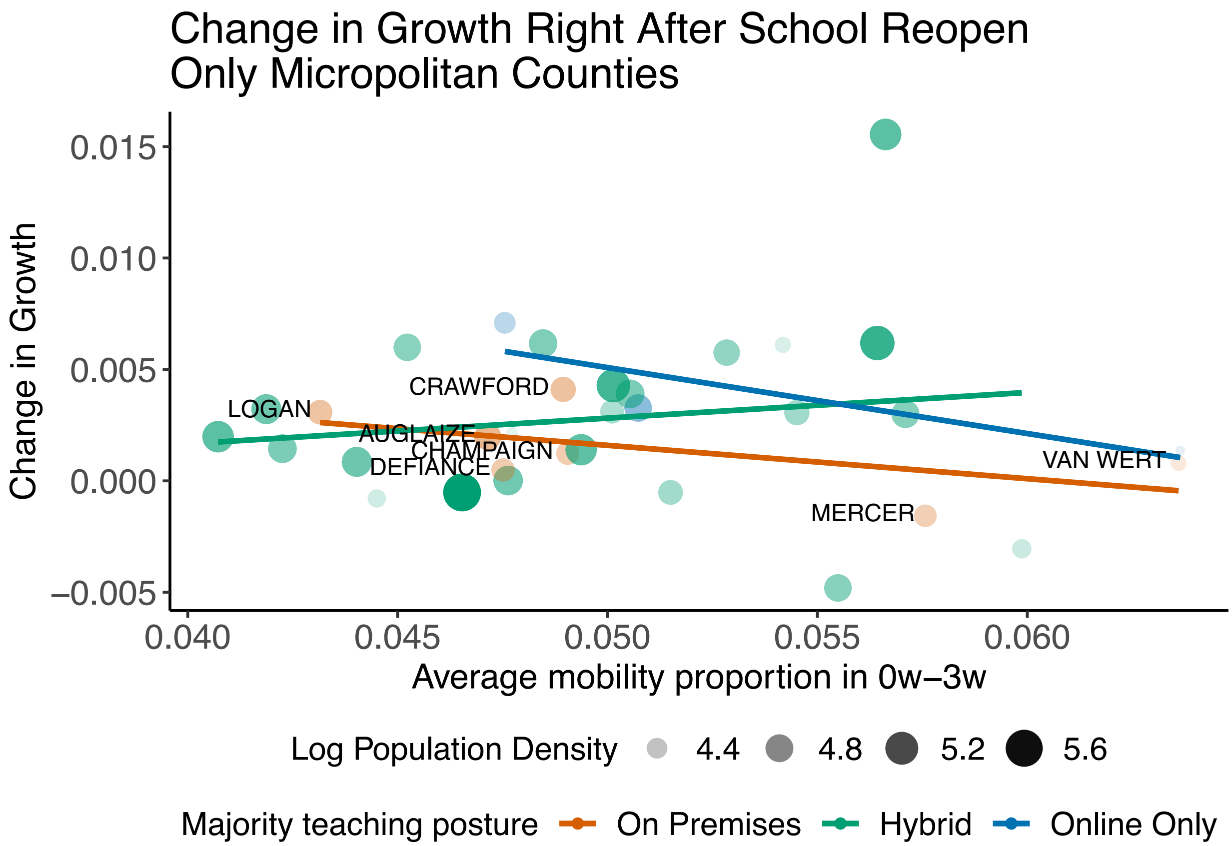
6.6 Change in Growth vMaximum growth coefficient vs. mobility and population density for Micropolitan counties

```
## filter only micro
B_diff_micro <- B_diff%>%
  drop_na(major_teaching)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan") %>%
  mutate(acc = new.slope.diff2 - new.slope.diff)
```

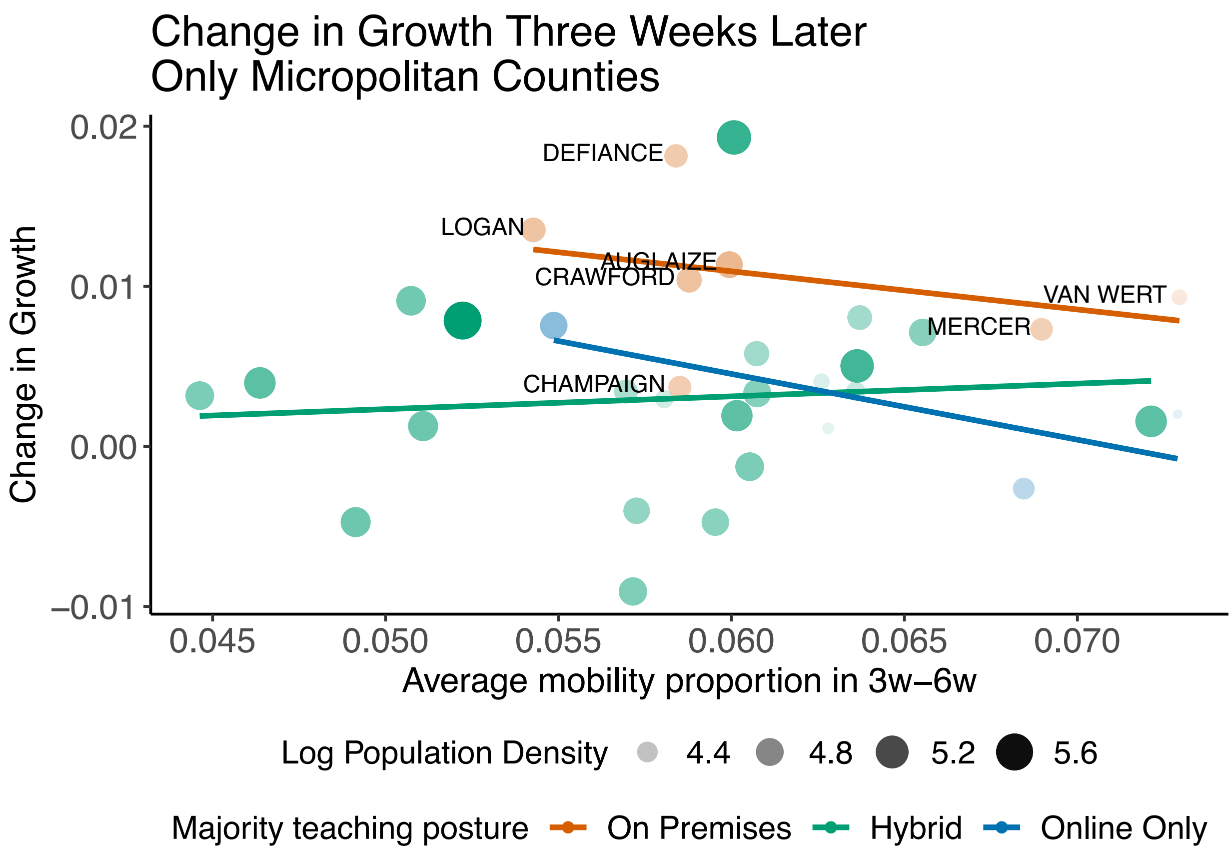
Mobility

```
B_diff_micro%>%
  ggplot(aes(x=avg_full_work_prob,
             y=new.slope.diff,
             group=major_teaching,color=major_teaching))+
  geom_point(aes(size=log(Population.density),alpha=log(Population.density)))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  labs(y="Change in Growth",x="Average mobility proportion in 0w-3w",
       title="Change in Growth Right After School Reopen\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "Log Population Density",
       alpha= "Log Population Density")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+team_theme+
```

```
  geom_text(data =B_diff_micro%>%
            filter(major_teaching=="On Premises"),
          aes(label=COUNTY),color='black',
          size=3,hjust=1.1, vjust=0.3,size=6)
```

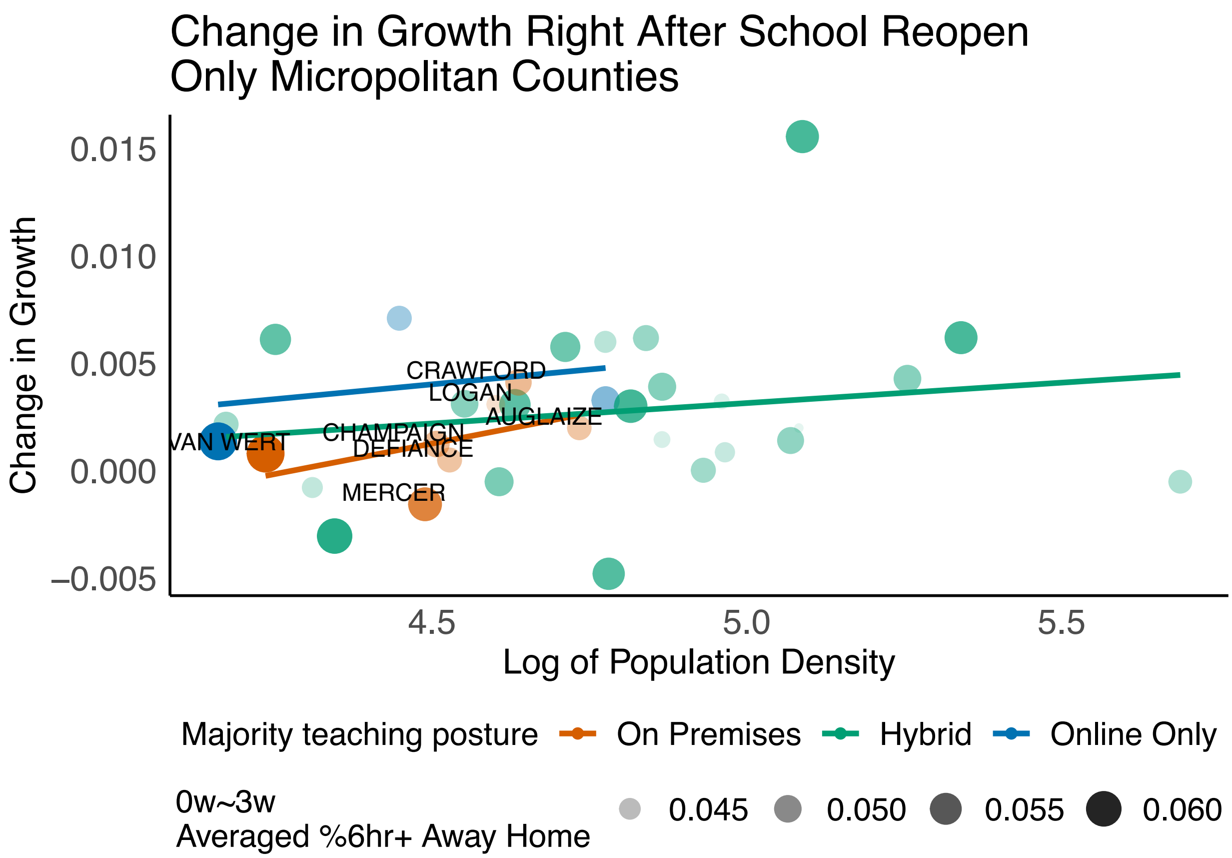



```
B_diff_micro%>%
  ggplot(aes(x=avg2_full_work_prob,
             y=new.slope.diff2,
             group=major_teaching,
             color=major_teaching))+
  geom_point(aes(size=log(Population.density),alpha=log(Population.density)))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  labs(y="Change in Growth",x="Average mobility proportion in 3w-6w",
       title="Change in Growth Three Weeks Later\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size ="Log Population Density",
       alpha= "Log Population Density")+
  scale_color_manual(values=col_theme)+
  theme(legend.position = "bottom")+
  team_theme+geom_text(data =B_diff_micro%>%
                        filter(major_teaching=="On Premises"),
                      aes(label=COUNTY),color='black',
                      size=3,hjust=1.1, vjust=0.3,size=6)
```

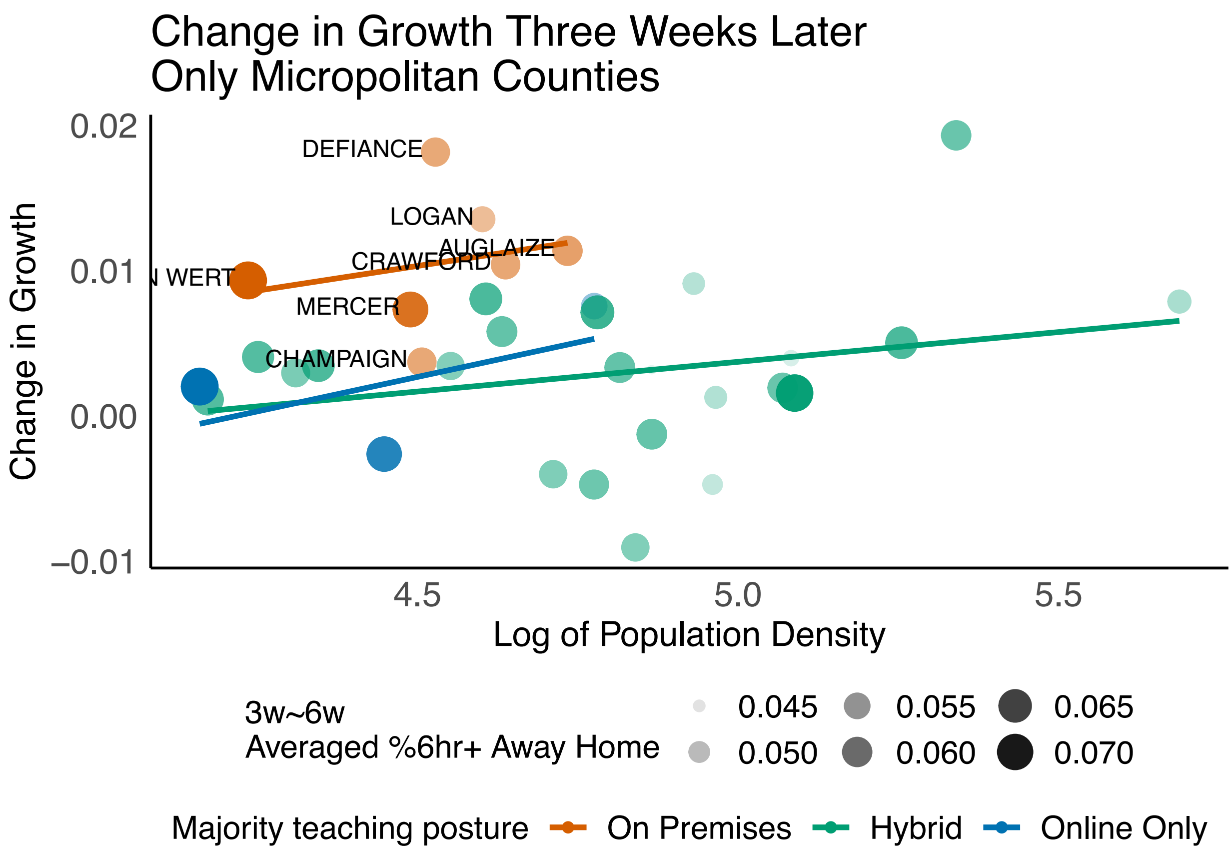


Log Population Density

```
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff,
             group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob,alpha=avg_full_work_prob))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth Right After School Reopen\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "0w~3w\nAveraged %6hr+ Away Home",
       alpha= "0w~3w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+
  theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
              filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',
            size=3,hjust=0.8, vjust=-0.2)
```




```r
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff2,
             group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg2_full_work_prob,
                 alpha=avg2_full_work_prob))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth Three Weeks Later\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "3w~6w\nAveraged %6hr+ Away Home",
       alpha= "3w~6w\nAveraged %6hr+ Away Home" ,
       fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
              filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',size=3,hjust=1.1, vjust=0.3)
```



Appendix 7: Sensitive Analysis

In order to make sure that the shifting of the red line (On Premises counties) is not brought by chance, we will conduct a sensitive analysis to detect how the change in growth varies throughout time. The whole sensitive analysis is based on 'Change in growth versus Log Population Density.

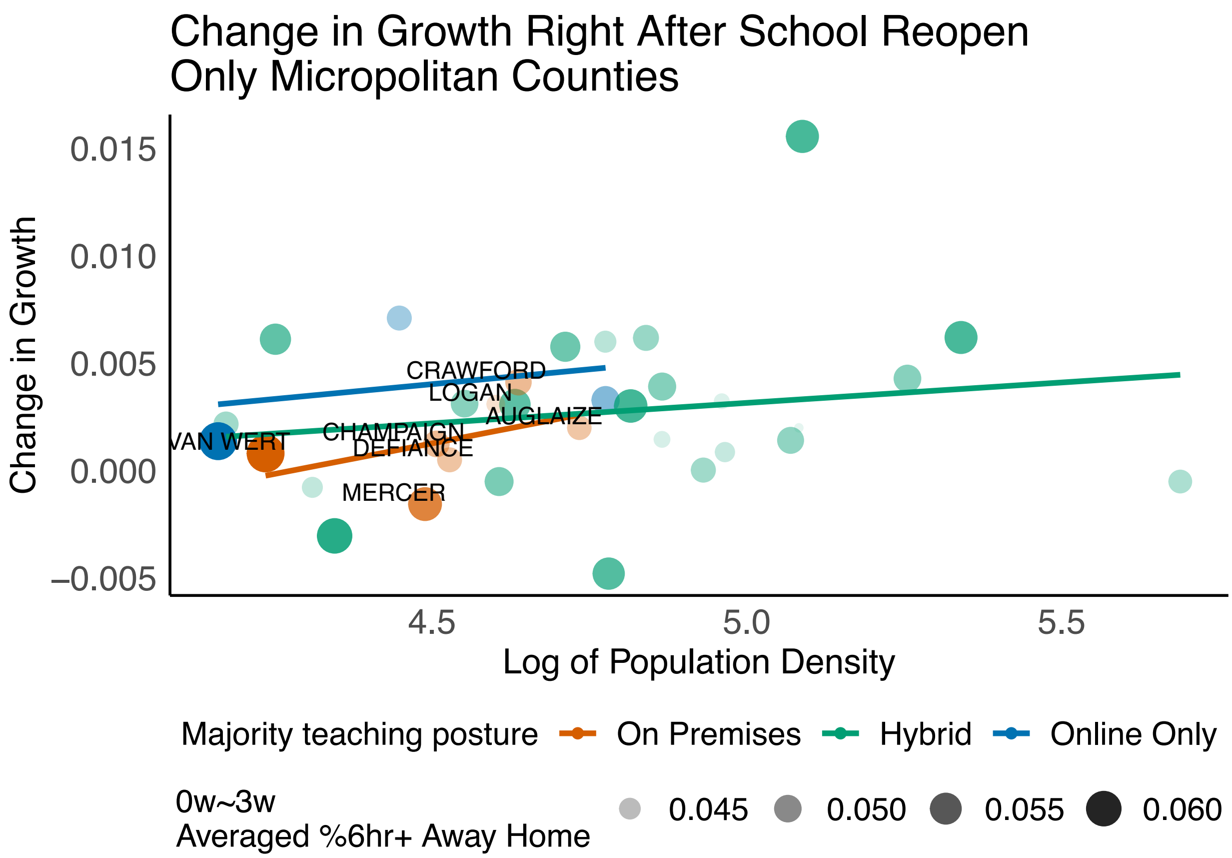
Since we assume that the school posture takes three weeks to reflect on the growth coefficient, the growth coefficients before 3 weeks after the start of school are all regarded as not taking effect. So, we use B(3) as a turning point. The changes in growth we want to test are as below:

Before school posture taking effect: B(0)-B(-3), B(1)-B(-2), B(2)-B(-1), B(3)-B(0) (also known as change in growth right after the start of school reopen).

After school posture taking effect: B(4)-B(1), B(5)-B(2), B(6)-B(3) (also known as change in growth three weeks later), B(7)-B(4)

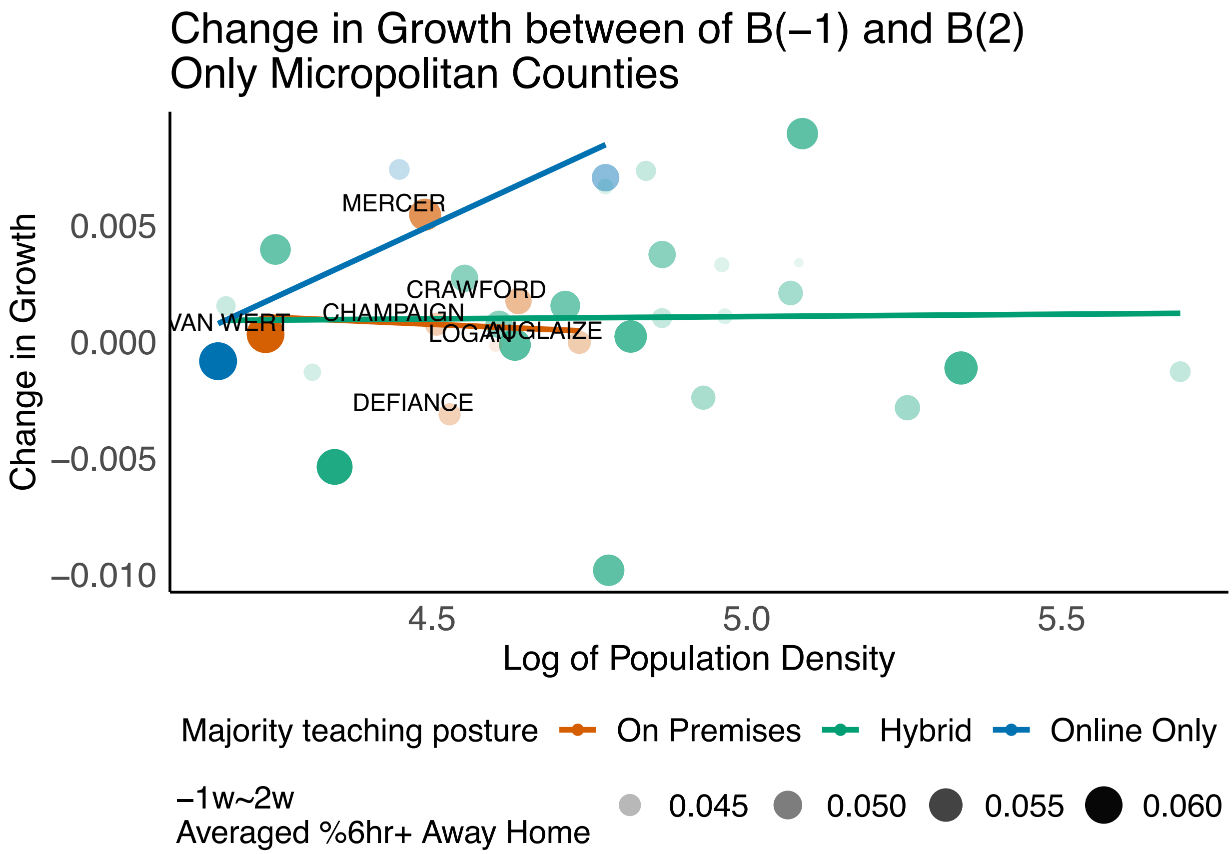
```
B_diff_micro <- B_diff%>%
  drop_na(major_teaching)%>%
  filter(NCHS.Urban.Rural.Status=="Micropolitan") %>%
  mutate(acc = new.slope.diff2 - new.slope.diff)
#Before reopen
##3w-0w
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff,
             group=major_teaching,
             color=major_teaching))+
```

```
  geom_point(aes(size=avg_full_work_prob,alpha=avg_full_work_prob))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth Right After School Reopen\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "0w~3w\nAveraged %6hr+ Away Home",
       alpha= "0w~3w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',size=3,hjust=0.8, vjust=-0.2)
```

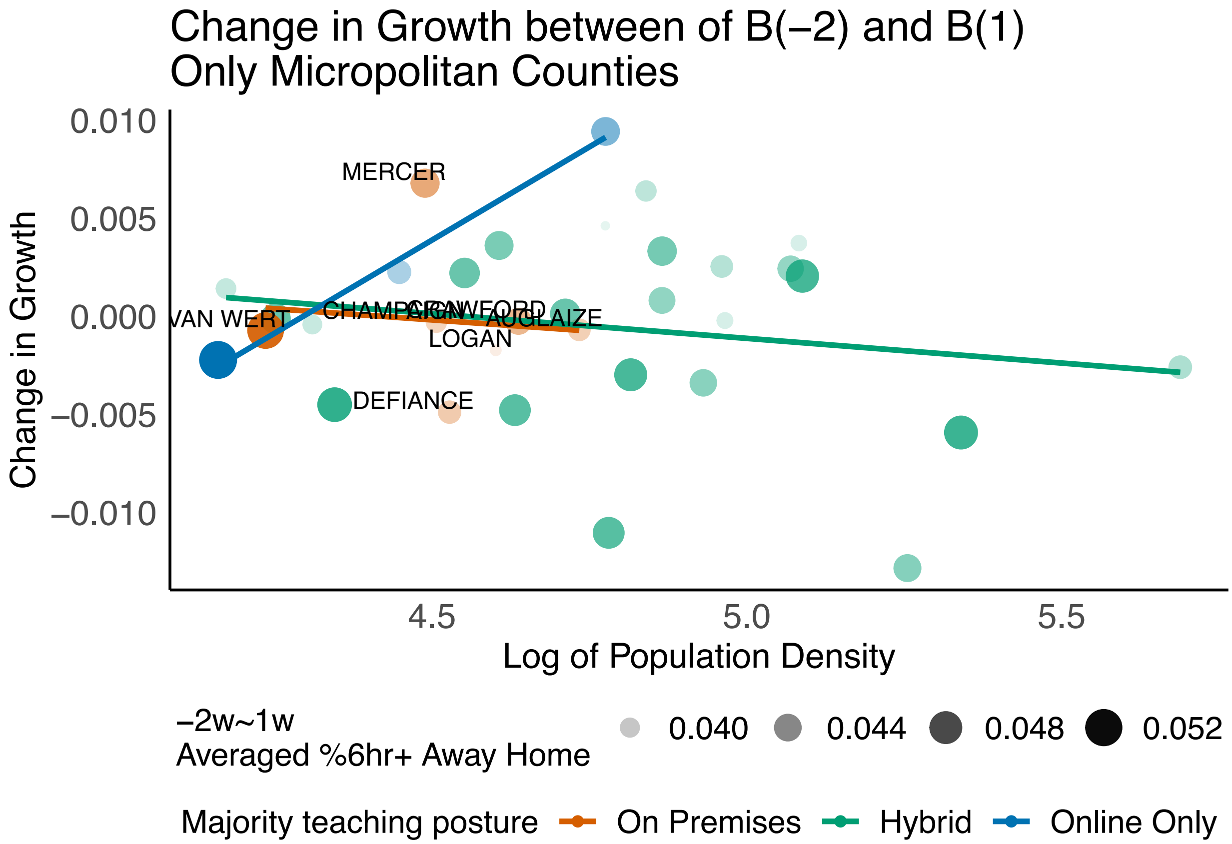



```
##2w-(-1w)
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff2m1,group=major_teaching,
             color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob_m1w2w,alpha=avg_full_work_prob_m1w2w))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth between of B(-1) and B(2)\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "-1w~2w\nAveraged %6hr+ Away Home",
       alpha= "-1w~2w\nAveraged %6hr+ Away Home" ,
       fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+
```

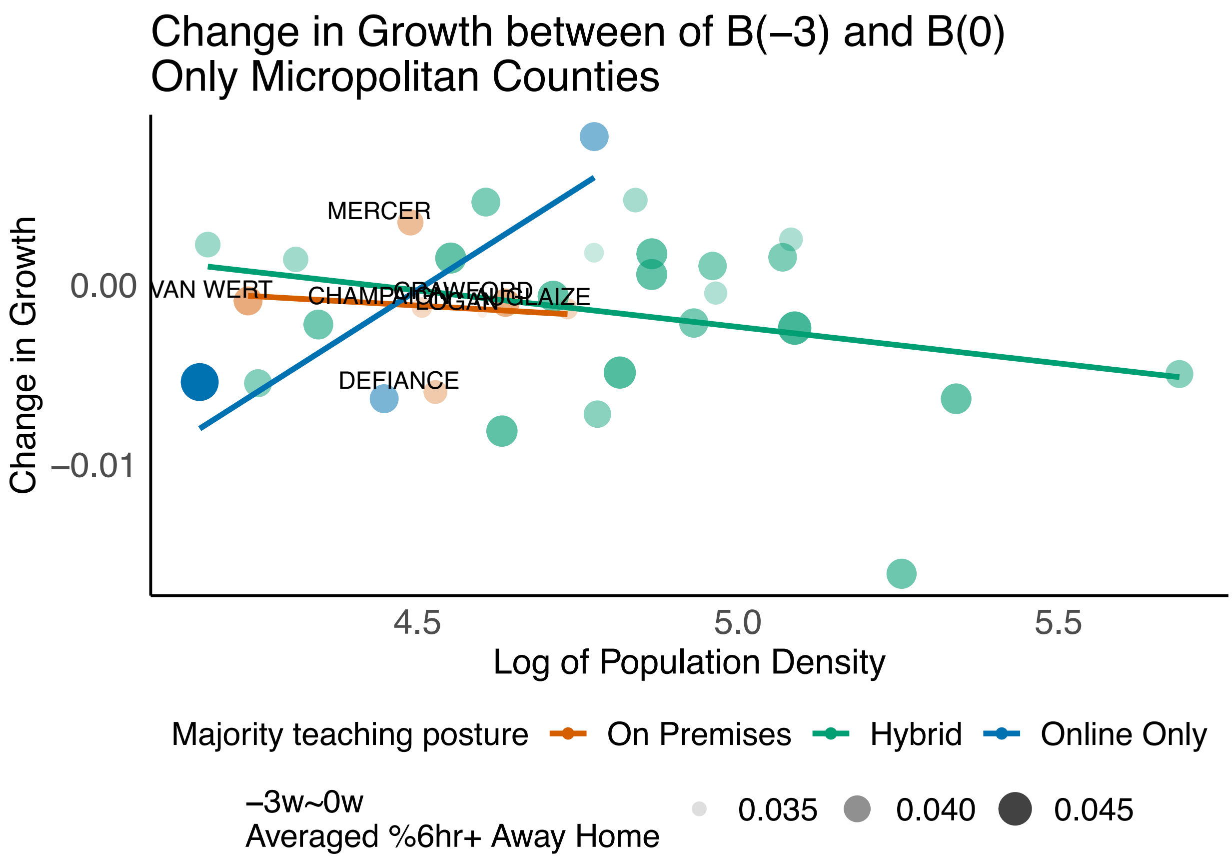
```
  theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
              filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',size=3,hjust=0.8, vjust=-0.2)
```



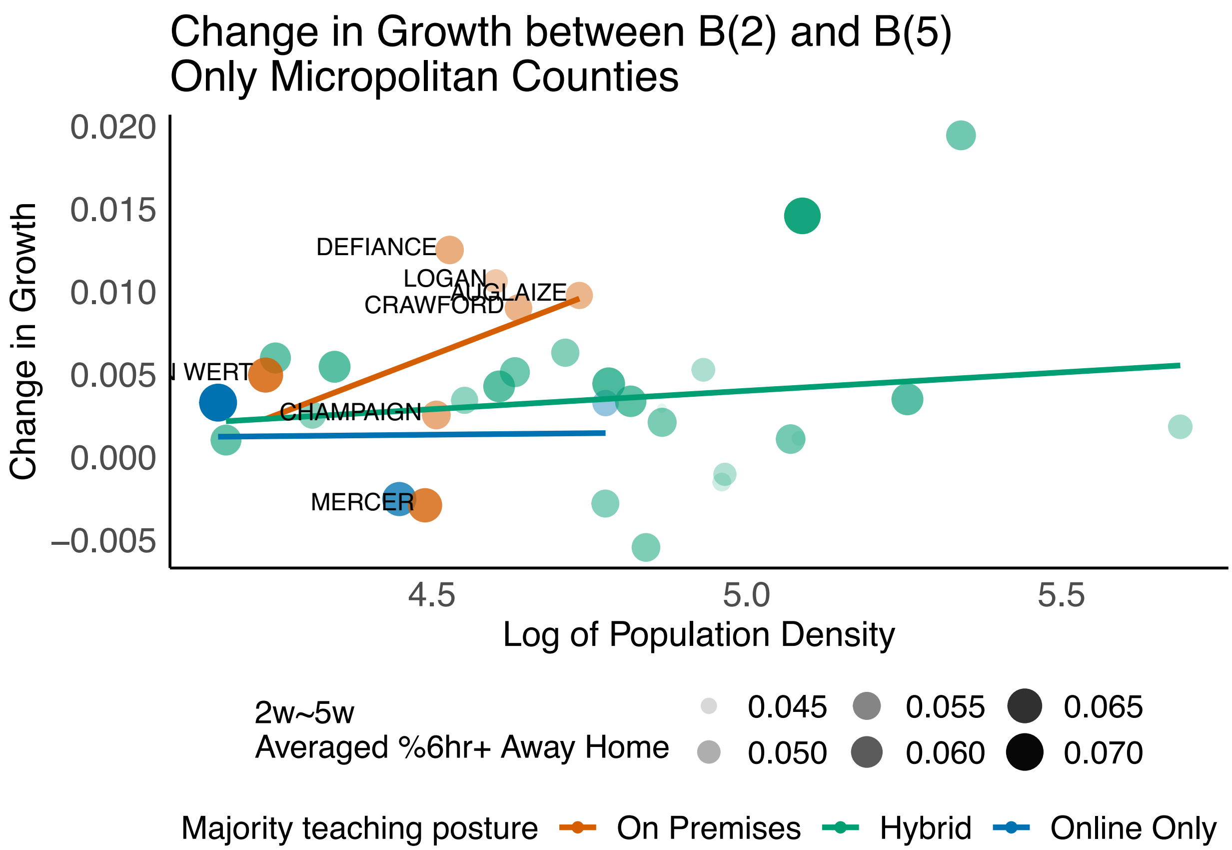

```
##1w-(-2w)
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff1m2,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob_m2w1w,alpha=avg_full_work_prob_m2w1w))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth between of B(-2) and B(1)\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "-2w~1w\nAveraged %6hr+ Away Home",
       alpha= "-2w~1w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
              filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',size=3,hjust=0.8, vjust=-0.2)
```



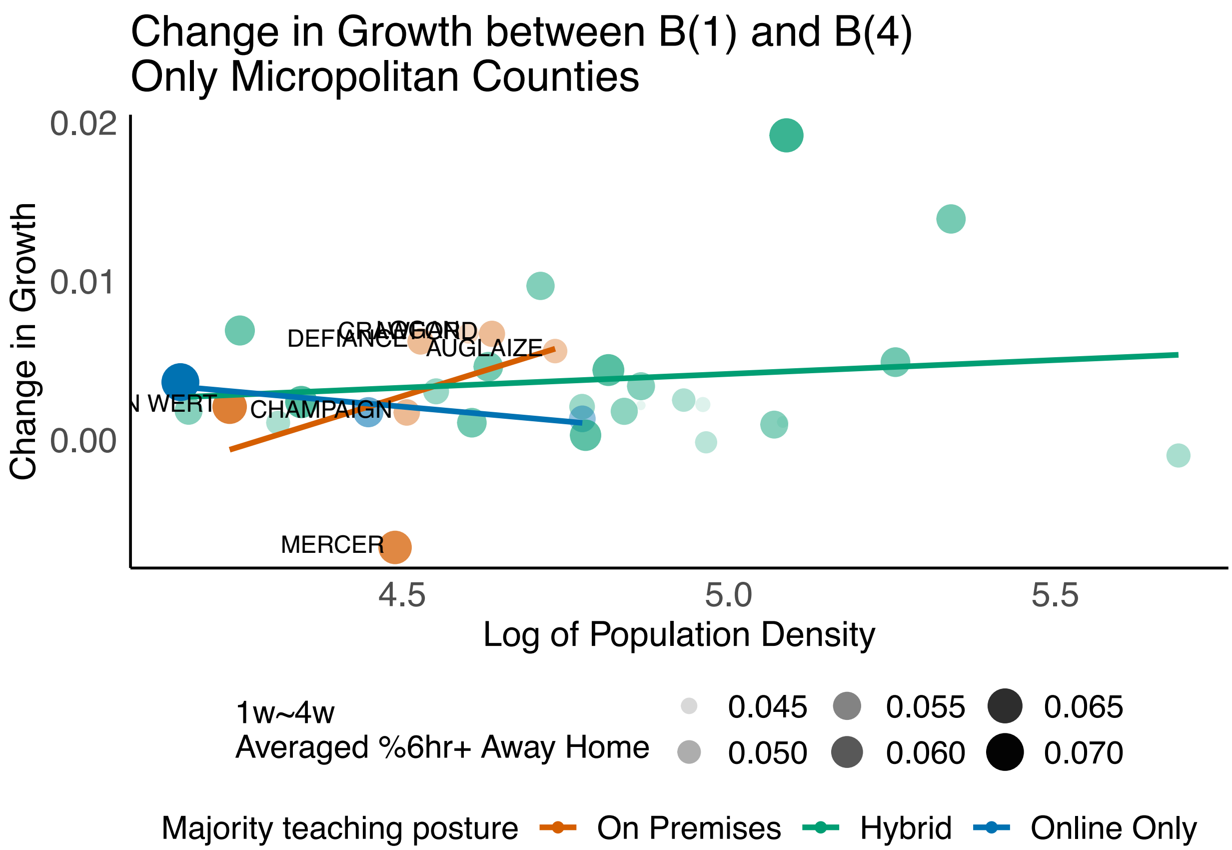

```r
##0w-(-3w)
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff0m3,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob_m3w0w,alpha=avg_full_work_prob_m3w0w))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth between of B(-3) and B(0)\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "-3w~0w\nAveraged %6hr+ Away Home",
       alpha= "-3w~0w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',
            size=3,hjust=0.8, vjust=-0.2)
```



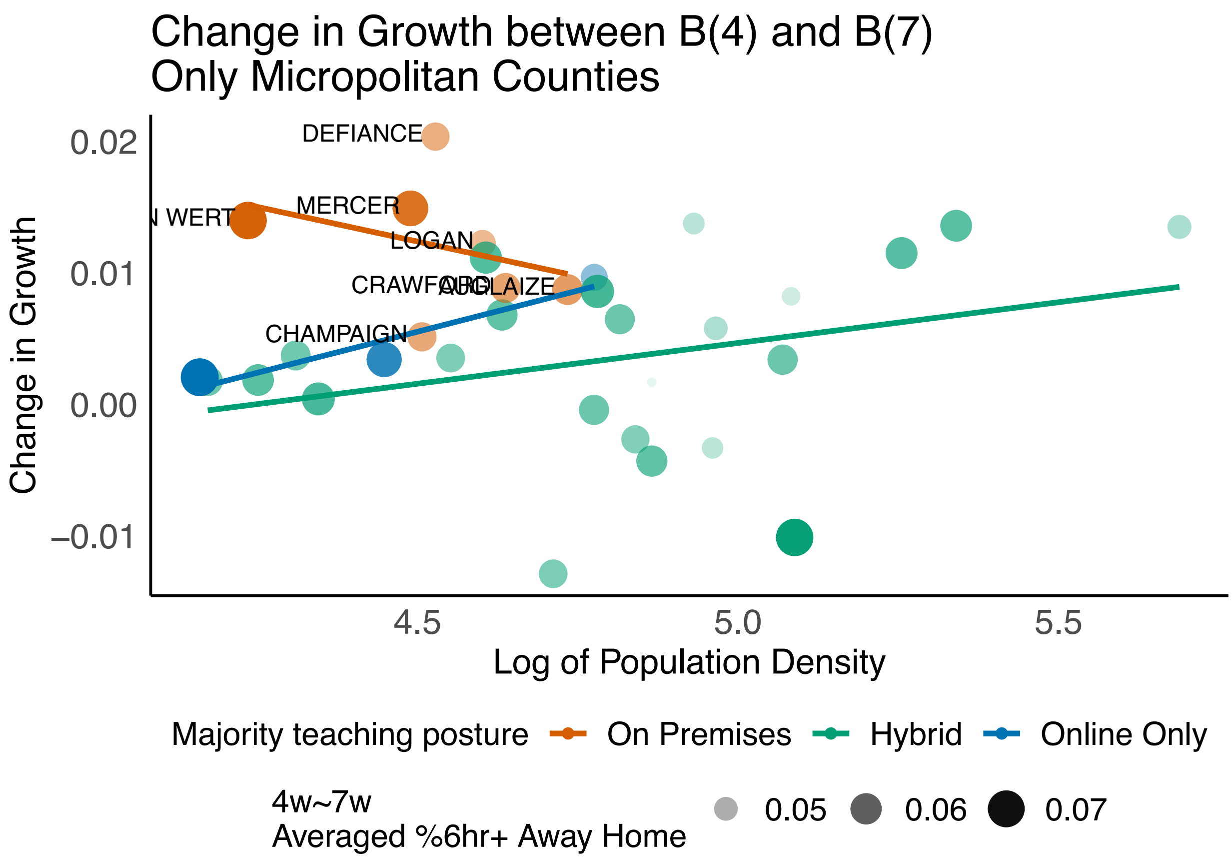

```
##5w-2w
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),y=new.slope.diff52,
             group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob_2w5w,alpha=avg_full_work_prob_2w5w))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+
  theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth between B(2) and B(5)\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "2w~5w\nAveraged %6hr+ Away Home",
       alpha= "2w~5w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',
            size=3,hjust=1.1, vjust=0.3)
```




```
##4w-1w
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff41,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob_1w4w,alpha=avg_full_work_prob_1w4w))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth between B(1) and B(4)\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "1w~4w\nAveraged %6hr+ Away Home",
       alpha= "1w~4w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%
              filter(major_teaching=="On Premises"),
            aes(label=COUNTY),color='black',
            size=3,hjust=1.1, vjust=0.3)
```




```r
##7w-4w
B_diff_micro%>%
  ggplot(aes(x=log(Population.density),
             y=new.slope.diff74,group=major_teaching,color=major_teaching))+
  geom_point(aes(size=avg_full_work_prob_4w7w,alpha=avg_full_work_prob_4w7w))+
  geom_smooth(method = "lm", se=F, formula = y ~ x,alpha=0.1)+theme_minimal()+team_theme+
  labs(y="Change in Growth",x="Log of Population Density",
       title="Change in Growth between B(4) and B(7)\nOnly Micropolitan Counties",
       color="Majority teaching posture",
       size = "4w~7w\nAveraged %6hr+ Away Home",
       alpha= "4w~7w\nAveraged %6hr+ Away Home" ,fill="Majority teaching posture")+
  scale_color_manual(values=col_theme)+
  theme(legend.position = "bottom")+
  geom_text(data =B_diff_micro%>%filter(major_teaching=="On Premises"),
            aes(label=COUNTY),
            color='black',size=3,hjust=1.1, vjust=0.3)
```

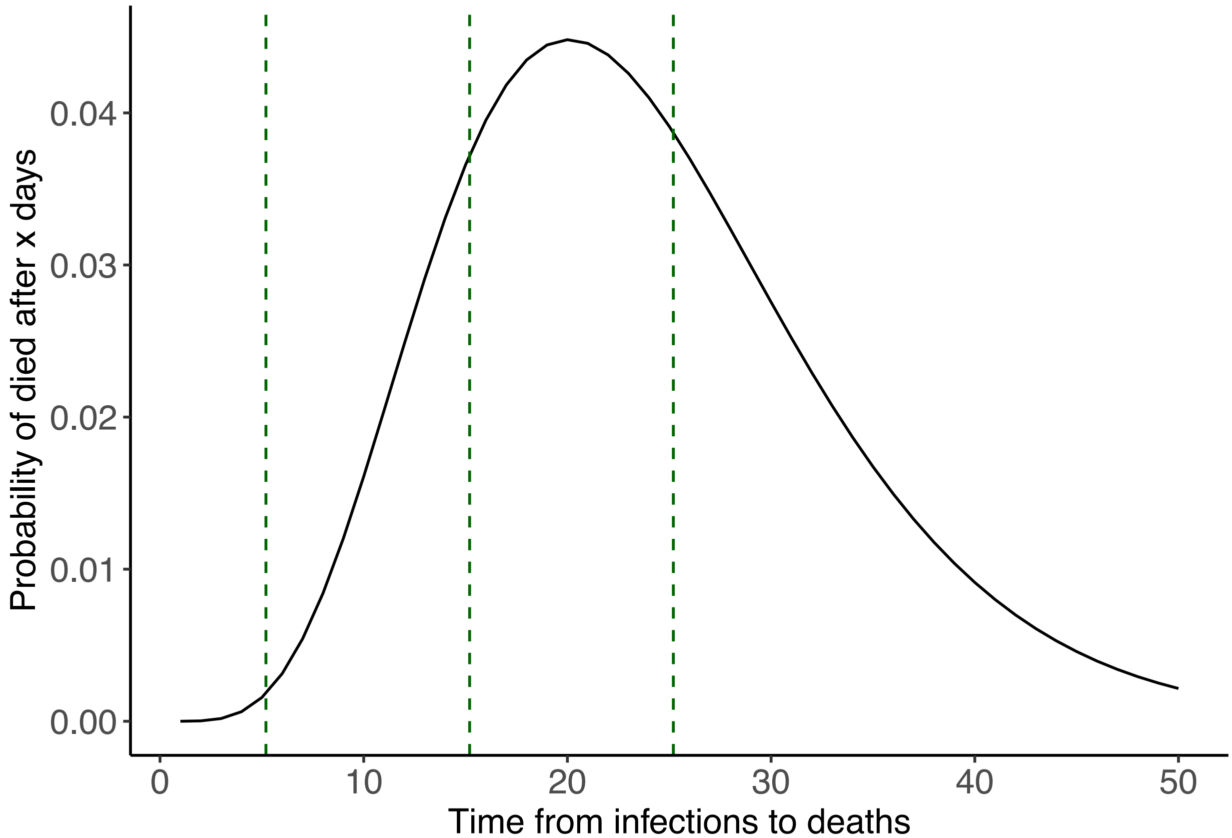


Appendix 8: Math Plots

Gamma distribution for the time lengths from infections to deaths

We know from previous study that the mean for this Gamma distribution is 23.9, with a coefficient of variation being 0.4.

```
# package for The Gamma Distribution (Alternative Parameterization)
# install.packages("EnvStats")
library(EnvStats)
time_to_deaths <- 1:50
prob_time_to_deaths <- dgammaAlt(x = time_to_deaths,mean = 23.9, cv = 0.4)
## shift x
gamma_plot <- data.frame(prob_time_to_deaths,time_to_deaths,
                         time_to_deaths+5,time_to_deaths+10,time_to_deaths+15)
colnames(gamma_plot) <- c("prob","time1","time2","time3","time4")
ggplot(gamma_plot)+
  geom_line(aes(x=time1,y=prob),colour = "black")+
  geom_vline(xintercept = 5.2,lty=2,colour="darkgreen")+
  geom_vline(xintercept = 15.2,lty=2,colour="darkgreen")+
  geom_vline(xintercept = 25.2,lty=2,colour="darkgreen")+
  labs(x="Time from infections to deaths",
       y="Probability of died after x days")+team_theme+
  theme(legend.position = "bottom")
```


Probability of died after x days
0.04
0.03
0.02
0.01
0.00
0
10
20
30
40
50
Time from infections to deaths