Epi Final

Margaret Patterson

12/13/2021

#Packages and data import

library(tidyverse)  
library(stringr)  
library(dplyr)  
library(lme4)  
library(broom)  
library(scales)  
library(readxl)  
library(broom.mixed)  
library(ggplot2)  
  
mask <- read\_csv("nytimes\_covid\_masks.csv")   
colleges <- read\_csv("nytimes\_covid\_college\_cases.csv")  
recent <- read\_csv("us-counties-recent.csv")   
election <- read\_csv("countypres\_2000-2020.csv")  
calicoll <- read\_excel("colleges.xlsx")

#Initial clean

set.seed(416)  
  
#county\_state (for joining), matching fips, cleaning imported data  
mask <- mask %>%  
 mutate(fips = str\_pad(COUNTYFP, 5, side = c("left"), pad = "0")) %>%  
 select(fips, NEVER, ALWAYS) %>%  
 rename\_all(str\_to\_lower)  
  
colleges <- colleges %>%  
 mutate(county\_state = paste(county, state, sep = ", ")) %>%  
 select(state, county, city, ipeds\_id, college, cases, cases\_2021, county\_state)  
  
recent <- recent %>%  
 mutate(county\_state = paste(county, state, sep = ", "))  
  
election <- election %>%  
 mutate(fips = str\_pad(county\_fips, 5, side = c("left"), pad = "0")) %>%  
 select(fips, candidate, party, candidatevotes, totalvotes, fips)  
  
calicoll <- calicoll %>%  
 mutate(fips = str\_pad(fips, 5, side = c("left"), pad = "0")) %>%  
 mutate(fips = as.character(fips))

#Joins

recentcoll <- full\_join(recent, colleges, by = "county\_state") #recent, colleges  
collrecoll <- full\_join(recentcoll, election, by = "fips") #recent, colleges, election  
addcali <- full\_join(collrecoll, calicoll, by = "fips") #recent, colleges, election, Cali - uses excel  
addmask <- full\_join(addcali, mask, by = "fips")

#Clean new dataset

cdf <- addmask %>%  
 na.omit(college.x) %>%  
 na.omit(`total enrollment`) %>%  
 select(date, college.x, city.x, state.x, county.x, `county pop`, fips, cases.x, deaths, ipeds\_id.x, `total enrollment`, `student vax`,  
 cases.y, cases\_2021, candidate, party, candidatevotes, totalvotes, always, never) %>%  
 rename(college = college.x) %>%  
 rename(ipeds\_id = ipeds\_id.x) %>%  
 rename(city = city.x) %>%  
 rename(county = county.x) %>%  
 rename(state = state.x) %>%  
 rename(cas\_cty = cases.x) %>%  
 rename(dth\_cty = deaths) %>%  
 rename(cas\_coll = cases.y) %>%  
 rename(cas\_coll\_21 = cases\_2021) %>%  
 mutate(partyratio = candidatevotes/totalvotes) %>%  
 mutate(cs\_cty\_ratio = (cas\_cty/`county pop`)) %>%  
 mutate(dth\_cty\_ratio = (dth\_cty/`county pop`)) %>%  
 mutate(cs\_collratio = (cas\_coll/`total enrollment`)) %>%  
 mutate(cs\_coll\_21\_ratio = (cas\_coll\_21/`total enrollment`)) %>%  
 group\_by(college)

#Finding CA colleges

california <- cdf %>% #75  
 filter(state == "California") %>%  
 nest()

#Filtering out large duplicates, filter to 13 with no vax info

ncdf <- cdf %>%  
 filter(ipeds\_id == c("110617", "110583", "110608", "121150", "120254", "122597", "122755",   
 "110635", "110644", "110662", "123961", "120883", "122612")) %>%  
 group\_by(ipeds\_id) %>%  
 mutate(n = n()) %>%  
 select(date, college, ipeds\_id, cas\_cty, dth\_cty, `county pop`, cas\_coll, cas\_coll\_21,   
 `total enrollment`, `student vax`, party, candidate, candidatevotes, totalvotes, always, never, partyratio, cs\_cty\_ratio,   
 dth\_cty\_ratio, cs\_collratio, cs\_coll\_21\_ratio, n) %>%  
 nest() %>%  
 unnest()  
  
ncdf

## # A tibble: 2,168 x 22  
## # Groups: ipeds\_id [13]  
## ipeds\_id date college cas\_cty dth\_cty `county pop` cas\_coll cas\_coll\_21  
## <chr> <date> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 110635 2021-10-31 Univer~ 121563 1419 1671329 1095 679  
## 2 110635 2021-10-31 Univer~ 121563 1419 1671329 1095 679  
## 3 110635 2021-11-01 Univer~ 122002 1432 1671329 1095 679  
## 4 110635 2021-11-02 Univer~ 122090 1432 1671329 1095 679  
## 5 110635 2021-11-02 Univer~ 122090 1432 1671329 1095 679  
## 6 110635 2021-11-03 Univer~ 122172 1441 1671329 1095 679  
## 7 110635 2021-11-04 Univer~ 122349 1441 1671329 1095 679  
## 8 110635 2021-11-04 Univer~ 122349 1441 1671329 1095 679  
## 9 110635 2021-11-05 Univer~ 122492 1440 1671329 1095 679  
## 10 110635 2021-11-05 Univer~ 122492 1440 1671329 1095 679  
## # ... with 2,158 more rows, and 14 more variables: total enrollment <dbl>,  
## # student vax <dbl>, party <chr>, candidate <chr>, candidatevotes <dbl>,  
## # totalvotes <dbl>, always <dbl>, never <dbl>, partyratio <dbl>,  
## # cs\_cty\_ratio <dbl>, dth\_cty\_ratio <dbl>, cs\_collratio <dbl>,  
## # cs\_coll\_21\_ratio <dbl>, n <int>

#Models, used mod3 for analysis

#Mixed effects  
mod <- ncdf %>%  
 lmer(dth\_cty\_ratio ~ 1 + cs\_collratio + `student vax` + never + cs\_cty\_ratio + cs\_coll\_21\_ratio + (1 | ipeds\_id), data = .)  
summary(mod)

## Linear mixed model fit by REML ['lmerMod']  
## Formula:   
## dth\_cty\_ratio ~ 1 + cs\_collratio + `student vax` + never + cs\_cty\_ratio +   
## cs\_coll\_21\_ratio + (1 | ipeds\_id)  
## Data: .  
##   
## REML criterion at convergence: -47735.5  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.5944 -0.4514 0.0032 0.4627 5.9149   
##   
## Random effects:  
## Groups Name Variance Std.Dev.   
## ipeds\_id (Intercept) 8.971e-08 2.995e-04  
## Residual 1.404e-11 3.747e-06  
## Number of obs: 2168, groups: ipeds\_id, 13  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) -1.179e-04 1.933e-04 -0.610  
## cs\_collratio 3.851e-07 3.378e-06 0.114  
## `student vax` 2.697e-07 8.992e-07 0.300  
## never 1.130e-02 8.095e-03 1.396  
## cs\_cty\_ratio 1.502e-02 8.163e-05 184.039  
## cs\_coll\_21\_ratio -8.453e-07 6.726e-06 -0.126  
##   
## Correlation of Fixed Effects:  
## (Intr) cs\_cll `stvx` never cs\_ct\_  
## cs\_collrati 0.000   
## `studentvx` -0.004 0.021   
## never -0.902 0.000 0.000   
## cs\_cty\_rati -0.038 0.003 0.001 -0.013   
## cs\_cll\_21\_r 0.000 -0.994 -0.042 0.000 -0.003

mod2 <- ncdf %>%  
 lmer(dth\_cty\_ratio ~ 1 + `student vax` + (1 | ipeds\_id), data = .)  
summary(mod2)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: dth\_cty\_ratio ~ 1 + `student vax` + (1 | ipeds\_id)  
## Data: .  
##   
## REML criterion at convergence: -41717.8  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.6296 -0.8619 -0.0456 0.8167 2.2843   
##   
## Random effects:  
## Groups Name Variance Std.Dev.   
## ipeds\_id (Intercept) 7.519e-07 8.671e-04  
## Residual 2.344e-10 1.531e-05  
## Number of obs: 2168, groups: ipeds\_id, 13  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 1.886e-03 2.405e-04 7.842  
## `student vax` 1.224e-07 3.601e-06 0.034  
##   
## Correlation of Fixed Effects:  
## (Intr)  
## `studentvx` -0.014

mod3 <- ncdf %>%  
 lmer(dth\_cty\_ratio ~ 1 + `student vax` + never + (1 | ipeds\_id), data = .)  
summary(mod3)

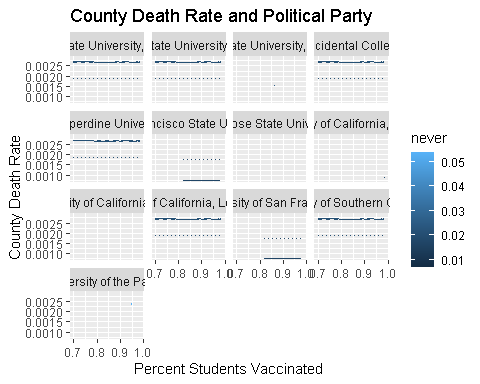
## Linear mixed model fit by REML ['lmerMod']  
## Formula: dth\_cty\_ratio ~ 1 + `student vax` + never + (1 | ipeds\_id)  
## Data: .  
##   
## REML criterion at convergence: -41713.8  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.6301 -0.8619 -0.0456 0.8167 2.2837   
##   
## Random effects:  
## Groups Name Variance Std.Dev.   
## ipeds\_id (Intercept) 7.065e-07 8.405e-04  
## Residual 2.344e-10 1.531e-05  
## Number of obs: 2168, groups: ipeds\_id, 13  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 1.235e-03 5.420e-04 2.279  
## `student vax` 1.224e-07 3.601e-06 0.034  
## never 3.022e-02 2.272e-02 1.331  
##   
## Correlation of Fixed Effects:  
## (Intr) `stvx`  
## `studentvx` -0.006   
## never -0.903 0.000  
## optimizer (nloptwrap) convergence code: 0 (OK)  
## unable to evaluate scaled gradient  
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues

mod3 %>%  
 tidy()

## # A tibble: 5 x 6  
## effect group term estimate std.error statistic  
## <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 fixed <NA> (Intercept) 0.00124 0.000542 2.28   
## 2 fixed <NA> `student vax` 0.000000122 0.00000360 0.0340  
## 3 fixed <NA> never 0.0302 0.0227 1.33   
## 4 ran\_pars ipeds\_id sd\_\_(Intercept) 0.000841 NA NA   
## 5 ran\_pars Residual sd\_\_Observation 0.0000153 NA NA

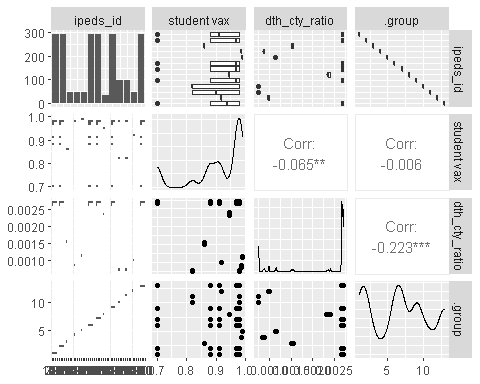
#Facet plot

library(broom.mixed)  
mod3 %>%  
 augment(data = ncdf) %>%  
 ggplot(aes(x = `student vax`, color = never)) +  
 geom\_line(aes(y = dth\_cty\_ratio), linetype = 1) +  
 geom\_line(aes(y = .fitted), linetype = 2) +  
 geom\_line(aes(y = .fixed), linetype = 3) +  
 labs(x = "Percent Students Vaccinated", y = "County Death Rate", title = "County Death Rate and Political Party") +  
 facet\_wrap(~ college, ncol = 4)

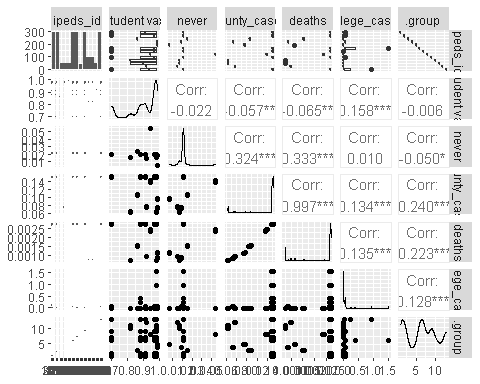


#Pairs plots

library(GGally)  
exposurepairs <- ncdf %>%  
 select(`student vax`, dth\_cty\_ratio) %>%  
 ggpairs()  
exposurepairs



bigpairs <- ncdf %>%  
 select(`student vax`, never, cs\_cty\_ratio, dth\_cty\_ratio, cs\_collratio) %>%  
 rename(county\_cases = cs\_cty\_ratio, deaths = dth\_cty\_ratio, college\_cases = cs\_collratio ) %>%  
 ggpairs()  
bigpairs



#Corr plot

library(ggcorrplot)  
cormod2 <- ncdf %>%  
 select(ipeds\_id, dth\_cty\_ratio, cs\_collratio, cs\_coll\_21\_ratio, cs\_cty\_ratio, `student vax`, never) %>%  
 mutate(dth\_cty\_ratio = as.numeric(dth\_cty\_ratio)) %>%  
 mutate(ipeds\_id = as.numeric(ipeds\_id)) %>%  
 rename(`county deaths` = dth\_cty\_ratio) %>%  
 rename(`college cases` = cs\_collratio) %>%  
 rename(`college cases 2021` = cs\_coll\_21\_ratio) %>%  
 rename(`county cases` = cs\_cty\_ratio) %>%  
 rename(`student vaccination rate` = `student vax`) %>%  
 rename(`never use mask` = never)   
  
cplot2 <- cormod2 %>%  
 cor(x = cormod2) %>%  
 ggcorrplot(method = "circle", type = "lower", title = "Student Vaccination and COVID-19 Cases and Deaths")  
cplot2

