

Time series analysis with updated window definition

Cheyenne Ehman, Ziyang Zhu

3/29/2021

Select variable of interests

```
## If you don't have the covidcast package, run following line

#devtools::install_github("cmu-delphi/covidcast", ref = "main", subdir = "R-packages/covidcast",dependencies = "R")

source("step2_data_wrangle.R")
##### school reopen dates #####
district_policies <- OH_K12 %>%
  distinct(county, county_enroll, leaid, district_enroll, schooltemporaryshutdown, opendategrouped, teachingmethod)

# Calculate the proportion and generate date brackets
major_opendate <- district_policies%>%
  filter(!schooltemporaryshutdown %in% c('Closed indefinitely', 'Pending', 'Unknown'))%>%
  group_by(county, county_enroll, opendategrouped)%>%
  summarise(n_opendate = sum(district_enroll))%>% # number of students under certain date for each county
  mutate(prop_opendate = round(n_opendate/county_enroll,2))%>% # proportion
  group_by(county)%>%
  #filter(prop_opendate>0.6)%>%
  slice(which.max(prop_opendate))%>% # filter large proportions of students with same reopen dates #can
  mutate(reopen_3w_after = opendategrouped + 21)%>%
  select(-n_opendate)

opendate_cases <- case_mobility%>%
  inner_join(major_opendate, by=c('COUNTY'='county'))%>%
  group_by(COUNTY)%>%
  filter(DATE>=reopen_3w_after - 42 & DATE<= as.Date('2020-12-25'))%>%
  ungroup()%>%
  mutate(window_id = case_when(
    DATE <= reopen_3w_after~"2month_before_3wafteropen",
    reopen_3w_after<DATE & DATE<=reopen_3w_after+42~"2month_after_3wafteropen",
    TRUE ~ 'Other'))%>%
  mutate(death_prop_1000 = round(CUMDEATHS/POPULATION,5)*1000,
    window_id = as.factor(window_id))%>%
  left_join(wide_teaching_enroll, by=c('COUNTY'='county', 'county_enroll'))
# select the start date and end date data for each window of time

death_incident_window <- opendate_cases%>%
  group_by(COUNTY, window_id)%>%
  arrange(DATE)%>%
  mutate(avg_work_7d = mean(work_prop_7d, na.rm = T), avg_res_visit = mean(res_visit_prop, na.rm = T), avg_death_prop_1000 = mean(death_prop_1000, na.rm = T))%>%
  filter(row_number()==1 | row_number()==n())%>%
```

```

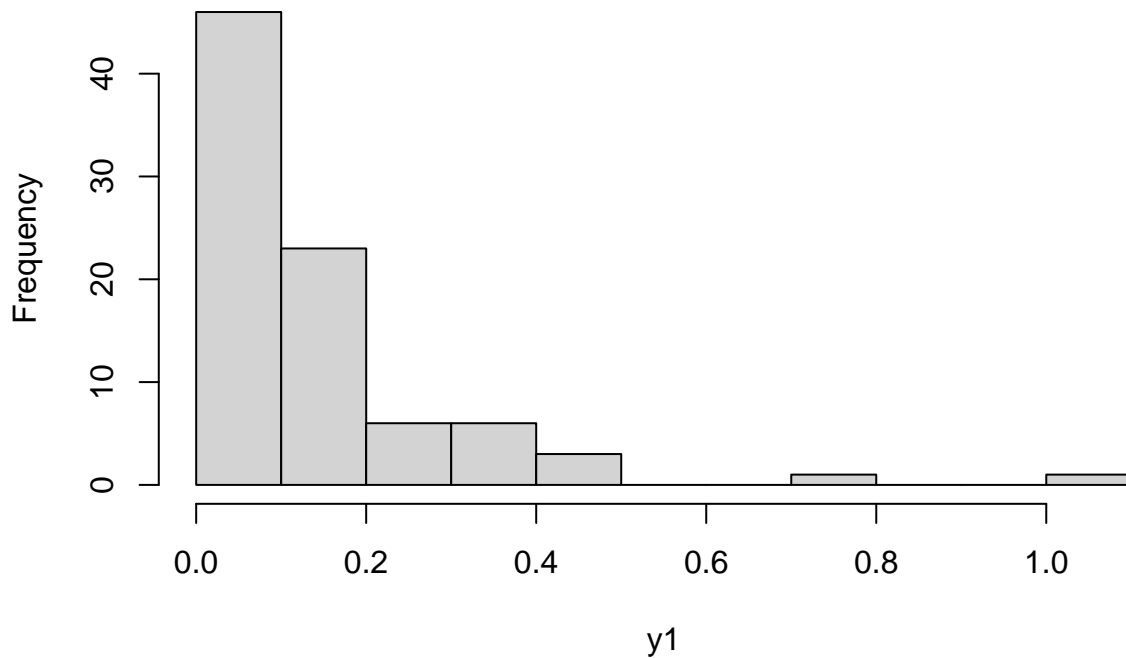
mutate(death_incident = diff(CUMDEATHS),death_incident_per_1000 = diff(CUMDEATHS)*1000/POPULATION)%>%
distinct(COUNTY,POPULATION,avg_work_7d,avg_res_visit,avg_bar_visit,Online_Only,On_Premises,Hybrid,dea

y1y0 <- death_incident_window %>%
  filter(window_id!='Other')%>%
  group_by(COUNTY)%>%
  mutate(y1= death_incident_per_1000, y0 = lag(death_incident_per_1000,n=1))%>%
  drop_na(y0)

hist(y1y0$y1,main = "Histogram of Death Incidents per 1000", xlab = "y1")

```

Histogram of Death Incidents per 1000

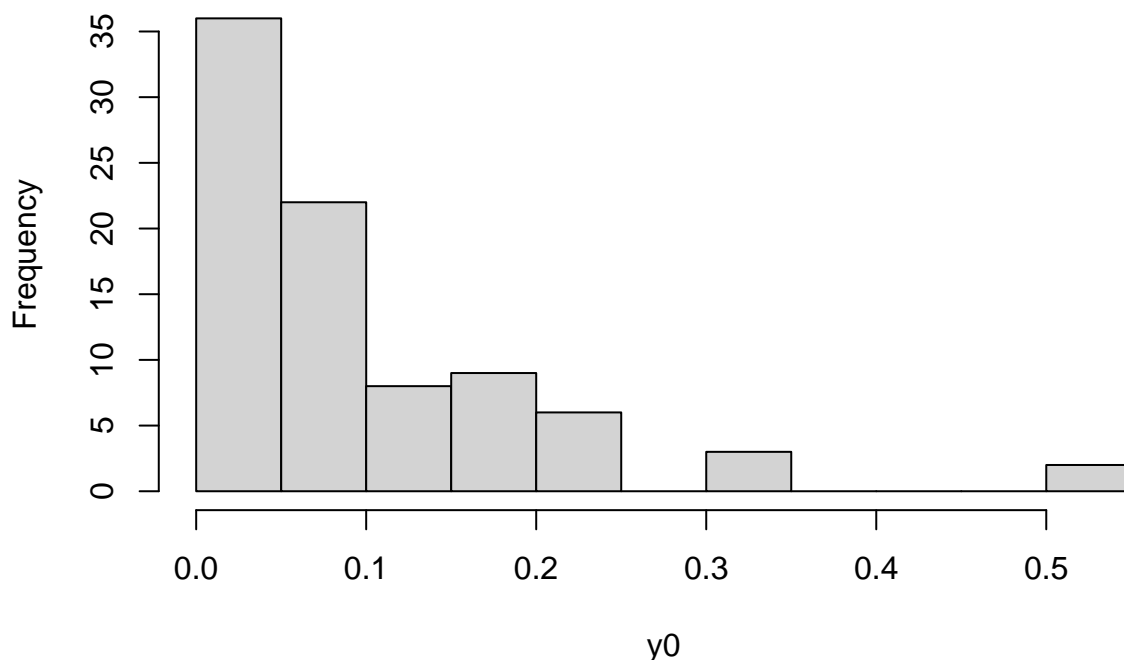


```

hist(y1y0$y0,main = "Histogram of Death Incidents per 1000", xlab = "y0")

```

Histogram of Death Incidents per 1000



```
summary(lm(y1~y0+avg_work_7d,na.action='na.omit',data = y1y0))
```

```
##
## Call:
## lm(formula = y1 ~ y0 + avg_work_7d, data = y1y0, na.action = "na.omit")
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.18251	-0.08679	-0.03874	0.02605	0.90385

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.27452	0.18236	-1.505	0.1360
y0	-0.07253	0.17265	-0.420	0.6755
avg_work_7d	4.70300	2.04689	2.298	0.0241 *

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1615 on 83 degrees of freedom
## Multiple R-squared:  0.05986,    Adjusted R-squared:  0.03721
## F-statistic: 2.642 on 2 and 83 DF,  p-value: 0.07717
```

```
summary(lm(y1~y0+avg_res_visit,na.action='na.omit',data = y1y0))
```

```
##
## Call:
## lm(formula = y1 ~ y0 + avg_res_visit, data = y1y0, na.action = "na.omit")
##
## Residuals:
```

	Min	1Q	Median	3Q	Max

```

## -0.17593 -0.08869 -0.05354 0.02291 0.93278
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.748e-01  3.801e-02  4.598 1.56e-05 ***
## y0          1.978e-02  1.759e-01   0.112   0.911
## avg_res_visit -1.024e-04  9.239e-05 -1.108   0.271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1672 on 81 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.01495,    Adjusted R-squared:  -0.009374
## F-statistic: 0.6146 on 2 and 81 DF,  p-value: 0.5434
summary(lm(y1~y0+avg_bar_visit,na.action='na.omit',data = y1y0))

##
## Call:
## lm(formula = y1 ~ y0 + avg_bar_visit, data = y1y0, na.action = "na.omit")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.050295 -0.030496  0.001266  0.024663  0.063801
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.862e-02  2.039e-02  4.347 0.00067 ***
## y0          9.951e-02  1.492e-01   0.667 0.51568
## avg_bar_visit -2.417e-05  1.269e-04 -0.190 0.85167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03945 on 14 degrees of freedom
## (69 observations deleted due to missingness)
## Multiple R-squared:  0.04531,    Adjusted R-squared:  -0.09107
## F-statistic: 0.3322 on 2 and 14 DF,  p-value: 0.7228
summary(lm(y1~y0+Hybrid,data = y1y0))

##
## Call:
## lm(formula = y1 ~ y0 + Hybrid, data = y1y0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14660 -0.09253 -0.05015  0.01704  0.95739
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.149441  0.043319   3.450 0.000884 ***
## y0          0.009918  0.174278   0.057 0.954752
## Hybrid      -0.014977  0.063674  -0.235 0.814616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
##
## Residual standard error: 0.1665 on 83 degrees of freedom
## Multiple R-squared: 0.0007324, Adjusted R-squared: -0.02335
## F-statistic: 0.03042 on 2 and 83 DF, p-value: 0.9701
```

```
summary(lm(y1~y0+On_Premises,data = y1y0))
```

```
##
## Call:
## lm(formula = y1 ~ y0 + On_Premises, data = y1y0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16950 -0.07433 -0.03554  0.02238  0.97266
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.11919    0.02708   4.402 3.17e-05 ***
## y0          -0.03909    0.17367  -0.225  0.8224
## On_Premises  0.12272    0.07238   1.696  0.0937 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1638 on 83 degrees of freedom
## Multiple R-squared: 0.03354, Adjusted R-squared: 0.01025
## F-statistic: 1.44 on 2 and 83 DF, p-value: 0.2427
```

```
summary(lm(y1~y0+Online_Only,data = y1y0))
```

```
##
## Call:
## lm(formula = y1 ~ y0 + Online_Only, data = y1y0)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15576 -0.08727 -0.04625  0.02067  0.93948
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.152376    0.027490   5.543 3.44e-07 ***
## y0           0.006622    0.173250   0.038  0.97
## Online_Only -0.084704    0.098041  -0.864  0.39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1658 on 83 degrees of freedom
## Multiple R-squared: 0.008979, Adjusted R-squared: -0.0149
## F-statistic: 0.376 on 2 and 83 DF, p-value: 0.6878
```

```
summary(lm(y1~y0+avg_work_7d+On_Premises,na.action='na.omit',data = y1y0))
```

```
##
## Call:
## lm(formula = y1 ~ y0 + avg_work_7d + On_Premises, data = y1y0,
##      na.action = "na.omit")
##
```

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18883 -0.08298 -0.03494  0.03069  0.92393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.21844    0.19304  -1.132   0.2611
## y0          -0.08787    0.17370  -0.506   0.6143
## avg_work_7d  3.92864    2.22461   1.766   0.0811 .
## On_Premises  0.06942    0.07759   0.895   0.3735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1617 on 82 degrees of freedom
## Multiple R-squared:  0.06895,    Adjusted R-squared:  0.03489
## F-statistic: 2.024 on 3 and 82 DF,  p-value: 0.1169

```