Epi Final

Margaret Patterson

12/13/2021

#Packages and data import

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.3 v dplyr 1.0.7  
## v tidyr 1.1.4 v stringr 1.4.0  
## v readr 2.0.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(stringr)  
library(dplyr)  
library(lme4)

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

library(broom)  
library(scales)

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

library(readxl)  
library(broom.mixed)  
library(ggplot2)  
  
mask <- read\_csv("nytimes\_covid\_masks.csv")

## Rows: 3142 Columns: 6

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## dbl (6): COUNTYFP, NEVER, RARELY, SOMETIMES, FREQUENTLY, ALWAYS

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

colleges <- read\_csv("nytimes\_covid\_college\_cases.csv")

## Rows: 1948 Columns: 9

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (7): date, state, county, city, ipeds\_id, college, notes  
## dbl (2): cases, cases\_2021

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

recent <- read\_csv("us-counties-recent.csv")

## Rows: 97502 Columns: 6

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (3): county, state, fips  
## dbl (2): cases, deaths  
## date (1): date

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

election <- read\_csv("countypres\_2000-2020.csv")

## Rows: 72617 Columns: 12

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (8): state, state\_po, county\_name, county\_fips, office, candidate, party...  
## dbl (4): year, candidatevotes, totalvotes, version

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

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")