christineRogers\_explore\_linelist\_20210114.R

gabrielodom

2021-01-14

# Query State Linelist  
# Gabriel Odom  
# 2021-01-14  
  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.2 ✓ purrr 0.3.4  
## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.0

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

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(readxl)  
  
linelist\_df <- read\_csv(  
 file = "../../data/deaths/Case\_Data\_arcGIS\_20210110.csv"  
)

##   
## ── Column specification ────────────────────────────────────────────────────────  
## cols(  
## County = col\_character(),  
## Age = col\_double(),  
## Age\_group = col\_character(),  
## Gender = col\_character(),  
## Jurisdiction = col\_character(),  
## Travel\_related = col\_character(),  
## Origin = col\_character(),  
## EDvisit = col\_character(),  
## Hospitalized = col\_character(),  
## Died = col\_character(),  
## Case = col\_character(),  
## Contact = col\_character(),  
## Case\_ = col\_character(),  
## EventDate = col\_datetime(format = ""),  
## ChartDate = col\_character()  
## )

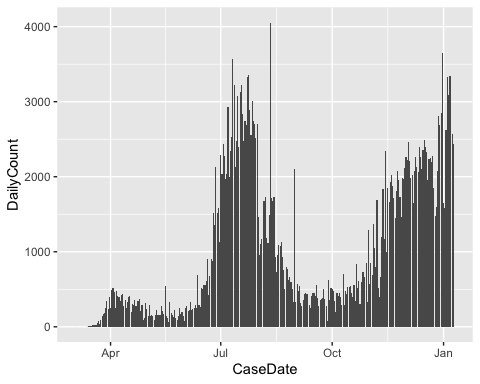
### Questions for MDC ###  
mdcLinelist\_df <-  
 linelist\_df %>%  
 filter(County == "Dade") %>%   
 rename(CaseDate = Case\_) %>%   
 # Regular expression to remove " 5:00" from the end of the date  
 mutate(CaseDate = str\_remove(CaseDate, pattern = " .\*")) %>%  
 mutate(CaseDate = as.Date(CaseDate, format = "%m/%d/%Y")) %>%   
 mutate(CaseDate = as\_date(CaseDate))  
  
# Christine Rogers hs asked 3 questions concerning the state linelist:  
# 1. Number of FL resident cases per day (the linelist data will not enable us  
# to indicate if the case was the first, or if the person tested positive  
# multiple days in a row). FLDoH measures this as "percent positivity for  
# new cases in FL residents".  
# 2. Number of non-FL resident cases per day (similar comments to above)  
# 3. Number of FL resident deaths daily (we are aware that this number is   
# delayed, up to weeks or months at a time)  
  
  
  
###### Question 1: FL Resident Cases / Day ##################################  
mdcLinelist\_df %>%   
 pull(Jurisdiction) %>%   
 table(useNA = "always")

## .  
## FL resident Non-FL resident   
## 322768 3838   
## Not diagnosed/isolated in FL <NA>   
## 1 0

dailyMDCflResCount\_df <-   
 mdcLinelist\_df %>%   
 filter(Jurisdiction != "Non-FL resident") %>%   
 group\_by(CaseDate) %>%   
 summarise(  
 DailyCount = n()  
 )

## `summarise()` ungrouping output (override with `.groups` argument)

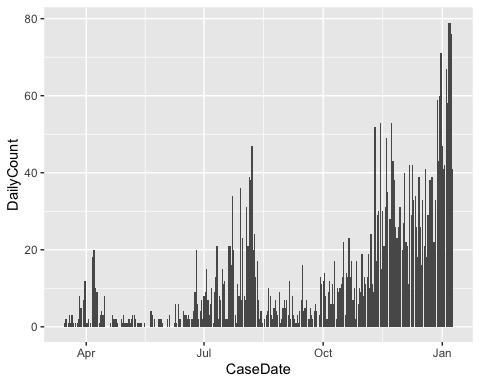
# write\_csv(dailyMDCflResCount\_df, "mdcDailyResCases\_20210114.csv")  
  
ggplot(data = dailyMDCflResCount\_df) +  
 aes(x = CaseDate, y = DailyCount) +  
 geom\_col()



###### Question 2: Non-FL Resident Cases / Day ##############################  
  
dailyMDCflNonResCount\_df <-   
 mdcLinelist\_df %>%   
 filter(Jurisdiction == "Non-FL resident") %>%   
 group\_by(CaseDate) %>%   
 summarise(  
 DailyCount = n()  
 )

## `summarise()` ungrouping output (override with `.groups` argument)

# write\_csv(dailyMDCflNonResCount\_df, "mdcDailyNonResCases\_20210114.csv")  
  
ggplot(data = dailyMDCflNonResCount\_df) +  
 aes(x = CaseDate, y = DailyCount) +  
 geom\_col()



###### Question 3: FL Resident Deaths / Day #################################  
  
dailyMDCflDeathsCount\_df <-   
 mdcLinelist\_df %>%   
 filter(Died %in% c("Recent", "Yes")) %>%   
 group\_by(CaseDate) %>%   
 summarise(  
 DailyCount = n()  
 )

## `summarise()` ungrouping output (override with `.groups` argument)

# write\_csv(dailyMDCflDeathsCount\_df, "mdcDailyDeaths\_20210114.csv")  
  
ggplot(data = dailyMDCflDeathsCount\_df) +  
 aes(x = CaseDate, y = DailyCount) +  
 geom\_col() +  
 # Add line chart  
 stat\_smooth(method = "gam")

## `geom\_smooth()` using formula 'y ~ s(x, bs = "cs")'

