lab17

Ziyuan_Han

12/1/2021

Import vaccination data

```
vax <- read.csv("covid19vaccinesbyzipcode_test.csv")
head(vax)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction       county
## 1 2021-01-05                   92091                San Diego    San Diego
## 2 2021-01-05                   92116                San Diego    San Diego
## 3 2021-01-05                   95360               Stanislaus   Stanislaus
## 4 2021-01-05                   94564             Contra Costa Contra Costa
## 5 2021-01-05                   95501                 Humboldt     Humboldt
## 6 2021-01-05                   95492                   Sonoma       Sonoma
##   vaccine_equity_metric_quartile               vem_source
## 1                              4    CDPH-Derived ZCTA Score
## 2                              3 Healthy Places Index Score
## 3                              1 Healthy Places Index Score
## 4                              4 Healthy Places Index Score
## 5                              2 Healthy Places Index Score
## 6                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1                1238.3                 1303                       NA
## 2               30255.7                31673                       45
## 3               10478.5                12301                       NA
## 4               17033.0                18381                       NA
## 5               20566.6                22061                       NA
## 6               25076.9                28024                       NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                     NA
## 2                          898                                0.001421
## 3                           NA                                     NA
## 4                           NA                                     NA
## 5                           NA                                     NA
## 6                           NA                                     NA
##   percent_of_population_partially_vaccinated
## 1                                         NA
## 2                                   0.028352
## 3                                         NA
## 4                                         NA
## 5                                         NA
```

```
## 6                                         NA
##   percent_of_population_with_1_plus_dose
## 1                                     NA
## 2                                0.029773
## 3                                     NA
## 4                                     NA
## 5                                     NA
## 6                                     NA
##                                                                    redacted
## 1 Information redacted in accordance with CA state privacy requirements
## 2                                                                        No
## 3 Information redacted in accordance with CA state privacy requirements
## 4 Information redacted in accordance with CA state privacy requirements
## 5 Information redacted in accordance with CA state privacy requirements
## 6 Information redacted in accordance with CA state privacy requirements
```

##Q1. What column details the total number of people fully vaccinated? age12_plus_population and age5_plus_population ##Q2. What column details the Zip code tabulation area? column 2 "zip_code_tabulation_area" ##Q3. What is the earliest date in this dataset? 2021.1.5 ##Q4. What is the latest date in this dataset? 2021.11.30

```
head(sort(vax$as_of_date,decreasing=TRUE))
```

```
## [1] "2021-11-30" "2021-11-30" "2021-11-30" "2021-11-30" "2021-11-30"
## [6] "2021-11-30"
```

```
#skimr::skim(vax)
```

##Q5. How many numeric columns are in this dataset? 9 ##Q6. Note that there are "missing values" in the dataset. How many NA values there in the persons_fully_vaccinated column? 8472 ##Q7. What percent of persons_fully_vaccinated values are missing (to 2 significant figures)? 10%

```
8472/84672
```

```
## [1] 0.1000567
```

#Working with dates

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

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

```
today
```

```
## function (tzone = "")
## {
##     as_date(now(tzone))
## }
## <bytecode: 0x7fa1ec474c78>
## <environment: namespace:lubridate>
```

```
# Specify that we are using the year-month-day format
vax$as_of_date <- ymd(vax$as_of_date)
#today() - vax$as_of_date[1]
vax$as_of_date[nrow(vax)] - vax$as_of_date[1]
```

```
## Time difference of 329 days
```

```
length(unique(vax$as_of_date))
```

```
## [1] 48
```

##How many days have passed since the last update of the dataset? 329 days ## How many unique dates are in the dataset (i.e. how many different dates are detailed)? 48 #Working with zip codes

```
#install.packages("zipcodeR")
library(zipcodeR)
geocode_zip('92037')
```

```
## # A tibble: 1 x 3
##   zipcode   lat   lng
##   <chr>   <dbl> <dbl>
## 1 92037    32.8 -117.
```

```
zip_distance('92037','92109')
```

```
##   zipcode_a zipcode_b distance
## 1     92037     92109     2.33
```

```
reverse_zipcode(c('92037', "92109") )
```

```
## # A tibble: 2 x 24
##   zipcode zipcode_type major_city post_office_city common_city_list county state
##   <chr>   <chr>        <chr>      <chr>                     <blob> <chr>  <chr>
## 1 92037   Standard     La Jolla   La Jolla, CA           <raw 20 B> San D~ CA
## 2 92109   Standard     San Diego  San Diego, CA          <raw 21 B> San D~ CA
## # ... with 17 more variables: lat <dbl>, lng <dbl>, timezone <chr>,
## #   radius_in_miles <dbl>, area_code_list <blob>, population <int>,
## #   population_density <dbl>, land_area_in_sqmi <dbl>,
## #   water_area_in_sqmi <dbl>, housing_units <int>,
## #   occupied_housing_units <int>, median_home_value <int>,
## #   median_household_income <int>, bounds_west <dbl>, bounds_east <dbl>,
## #   bounds_north <dbl>, bounds_south <dbl>
```

```
zipdata <- reverse_zipcode(vax$zip_code_tabulation_area )
# Subset to San Diego county only areas
#install.packages("dplyr")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
#Focus on San Deigo
sd <- filter(vax, county == "San Diego")

nrow(sd)
```

```
## [1] 5136
```

```
sd.10 <- filter(vax, county == "San Diego" &
                age5_plus_population > 10000)
```

#Q11. How many distinct zip codes are listed for San Diego County? 76

```
length(unique(sd.10$zip_code_tabulation_area))
```

```
## [1] 76
```

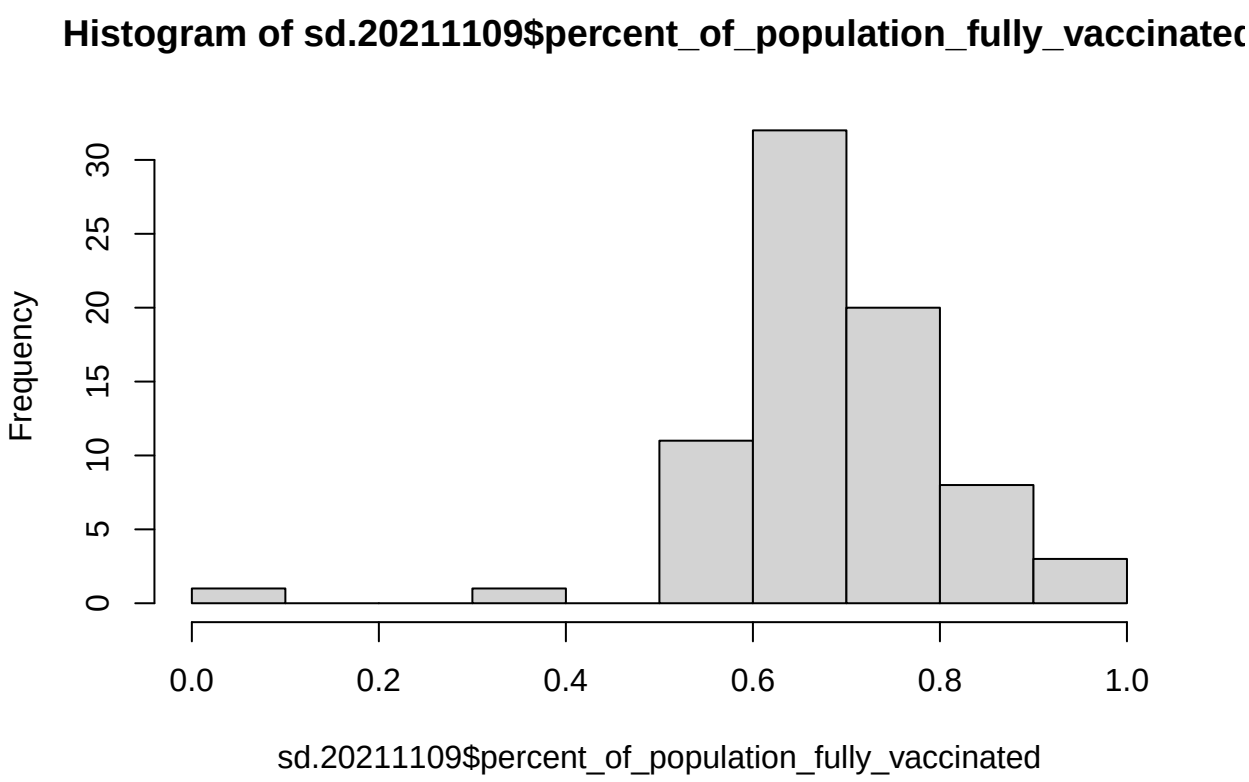
#Q12. What San Diego County Zip code area has the largest 12 + Population in this dataset? 92154
#Q13. What is the overall average "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2021-11-09"? 0.6802456

```
sd.20211109 <- filter(sd.10,as_of_date=="2021-11-09")
mean(sd.20211109$percent_of_population_fully_vaccinated)
```

```
## [1] 0.6802456
```

#Q14. Using either ggplot or base R graphics make a summary figure that shows the distribution of Percent of Population Fully Vaccinated values as of "2021-11-09"?

```
hist(sd.20211109$percent_of_population_fully_vaccinated)
```



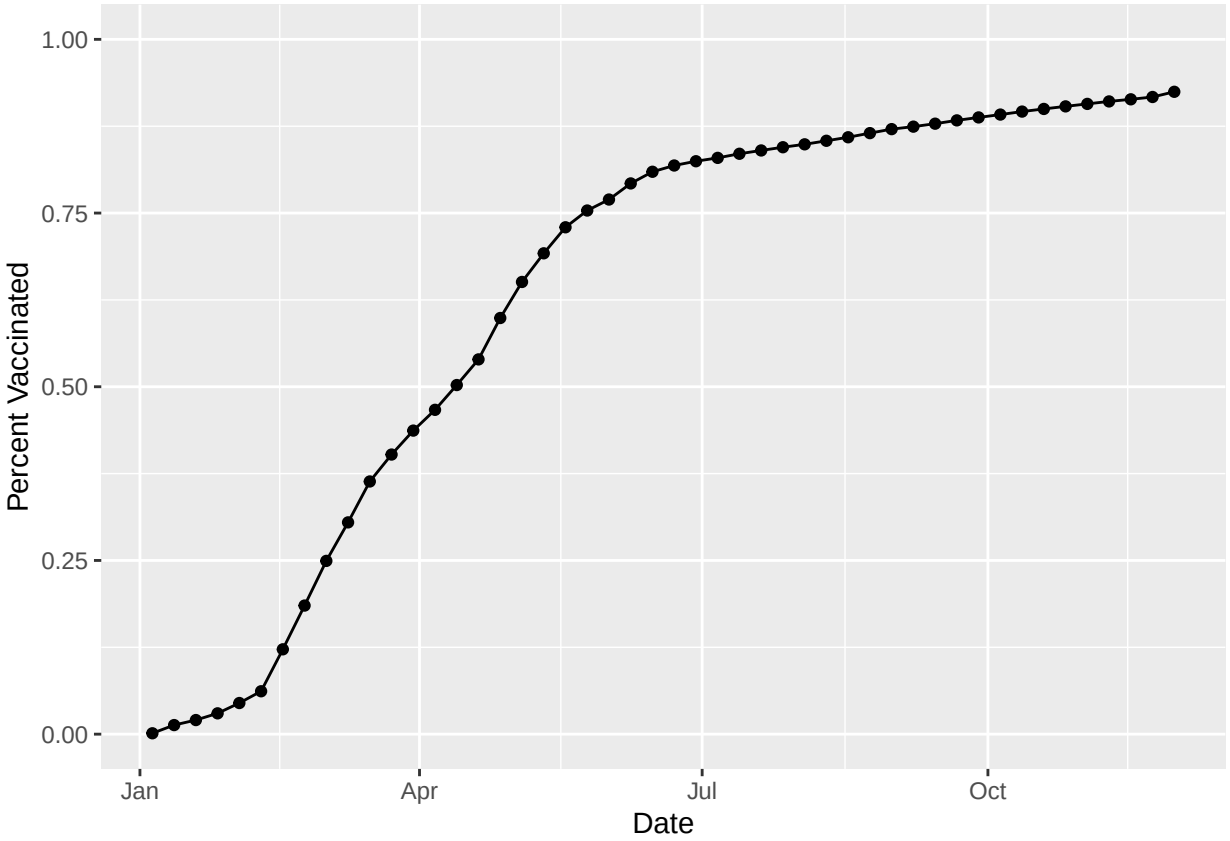
#focus on UCSD/La Jolla

```
ucsd <- filter(sd,zip_code_tabulation_area=="92037")
ucsd[1,]$age5_plus_population
```

```
## [1] 36144
```

#Q15. Using ggplot make a graph of the vaccination rate time course for the 92037 ZIP code area:

```
library(ggplot2)
p<-ggplot() +
  aes(ucsd$as_of_date,
      ucsd$percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated",main="Vaccination rate for La Jolla CA92109")
p
```

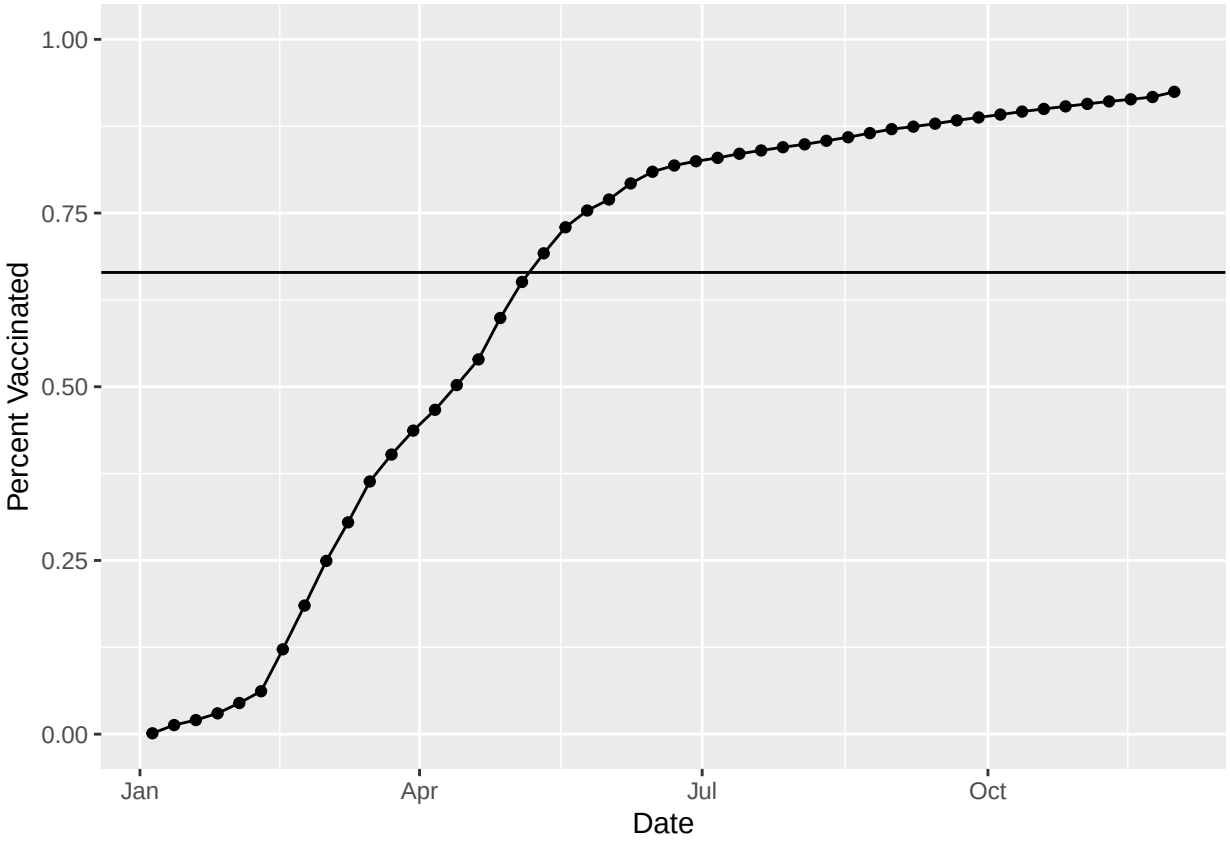


Subset to all CA areas with a population as large as 92037

```
library(dplyr)
vax.36 <- filter(vax, age5_plus_population > 36144 &as_of_date == "2021-11-16")
```

#Q16. Calculate the mean "Percent of Population Fully Vaccinated" for ZIP code areas with a population as large as 92037 (La Jolla) as_of_date "2021-11-16". Add this as a straight horizontal line to your plot from above with the geom_hline() function?

```
mean<-mean(vax.36$percent_of_population_fully_vaccinated)
p<- p +
  geom_hline(yintercept=mean)
p
```




```
sessionInfo()
```