class 14

Ramola Baviskar (PID A12228297)

3/7/2022

Important vaccination data

vax <- read.csv("vaccinedata.csv")  
head(vax)

## as\_of\_date zip\_code\_tabulation\_area local\_health\_jurisdiction county  
## 1 2021-01-05 92549 Riverside Riverside  
## 2 2021-01-05 92130 San Diego San Diego  
## 3 2021-01-05 92397 San Bernardino San Bernardino  
## 4 2021-01-05 94563 Contra Costa Contra Costa  
## 5 2021-01-05 94519 Contra Costa Contra Costa  
## 6 2021-01-05 91042 Los Angeles Los Angeles  
## vaccine\_equity\_metric\_quartile vem\_source  
## 1 3 Healthy Places Index Score  
## 2 4 Healthy Places Index Score  
## 3 3 Healthy Places Index Score  
## 4 4 Healthy Places Index Score  
## 5 3 Healthy Places Index Score  
## 6 2 Healthy Places Index Score  
## age12\_plus\_population age5\_plus\_population persons\_fully\_vaccinated  
## 1 2348.4 2461 NA  
## 2 46300.3 53102 61  
## 3 3695.6 4225 NA  
## 4 17216.1 18896 NA  
## 5 16861.2 18678 NA  
## 6 23962.2 25741 NA  
## persons\_partially\_vaccinated percent\_of\_population\_fully\_vaccinated  
## 1 NA NA  
## 2 27 0.001149  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA  
## percent\_of\_population\_partially\_vaccinated  
## 1 NA  
## 2 0.000508  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA  
## percent\_of\_population\_with\_1\_plus\_dose booster\_recip\_count  
## 1 NA NA  
## 2 0.001657 NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA  
## redacted  
## 1 Information redacted in accordance with CA state privacy requirements  
## 2 Information redacted in accordance with CA state privacy requirements  
## 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. persons\_fully\_vaccinated Q2. zip\_code\_tabulation\_area Q3. 2021-01-05 Q4. 2022-02-22

skimr::skim(vax)

Data summary

|  |  |
| --- | --- |
| Name | vax |
| Number of rows | 107604 |
| Number of columns | 15 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 5 |
| numeric | 10 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| as\_of\_date | 0 | 1 | 10 | 10 | 0 | 61 | 0 |
| local\_health\_jurisdiction | 0 | 1 | 0 | 15 | 305 | 62 | 0 |
| county | 0 | 1 | 0 | 15 | 305 | 59 | 0 |
| vem\_source | 0 | 1 | 15 | 26 | 0 | 3 | 0 |
| redacted | 0 | 1 | 2 | 69 | 0 | 2 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| zip\_code\_tabulation\_area | 0 | 1.00 | 93665.11 | 1817.39 | 90001 | 92257.75 | 93658.50 | 95380.50 | 97635.0 | ▃▅▅▇▁ |
| vaccine\_equity\_metric\_quartile | 5307 | 0.95 | 2.44 | 1.11 | 1 | 1.00 | 2.00 | 3.00 | 4.0 | ▇▇▁▇▇ |
| age12\_plus\_population | 0 | 1.00 | 18895.04 | 18993.91 | 0 | 1346.95 | 13685.10 | 31756.12 | 88556.7 | ▇▃▂▁▁ |
| age5\_plus\_population | 0 | 1.00 | 20875.24 | 21106.02 | 0 | 1460.50 | 15364.00 | 34877.00 | 101902.0 | ▇▃▂▁▁ |
| persons\_fully\_vaccinated | 18338 | 0.83 | 12155.61 | 13063.88 | 11 | 1066.25 | 7374.50 | 20005.00 | 77744.0 | ▇▃▁▁▁ |
| persons\_partially\_vaccinated | 18338 | 0.83 | 831.74 | 1348.68 | 11 | 76.00 | 372.00 | 1076.00 | 34219.0 | ▇▁▁▁▁ |
| percent\_of\_population\_fully\_vaccinated | 18338 | 0.83 | 0.51 | 0.26 | 0 | 0.33 | 0.54 | 0.70 | 1.0 | ▅▅▇▇▃ |
| percent\_of\_population\_partially\_vaccinated | 18338 | 0.83 | 0.05 | 0.09 | 0 | 0.01 | 0.03 | 0.05 | 1.0 | ▇▁▁▁▁ |
| percent\_of\_population\_with\_1\_plus\_dose | 18338 | 0.83 | 0.54 | 0.28 | 0 | 0.36 | 0.58 | 0.75 | 1.0 | ▅▃▆▇▅ |
| booster\_recip\_count | 64317 | 0.40 | 4100.55 | 5900.21 | 11 | 176.00 | 1136.00 | 6154.50 | 50602.0 | ▇▁▁▁▁ |

sum(is.na(vax$persons\_fully\_vaccinated))

## [1] 18338

nrow(vax)

## [1] 107604

sum(is.na(vax$persons\_fully\_vaccinated)) / nrow(vax)

## [1] 0.1704212

Q5. 9 Q6. 18338 Q7. ~17% Q8. Some counties may not have collected this data yet or haven’t yet submitted it.

library(lubridate)

##   
## Attaching package: 'lubridate'

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

vax$as\_of\_date <- ymd(vax$as\_of\_date)  
today <- today() - vax$as\_of\_date  
now <- vax$as\_of\_date[nrow(vax)] - vax$as\_of\_date

date <- vax[, c("as\_of\_date")]  
length(unique(date))

## [1] 61

Q9. 5=425-420 Q10. 61

library(zipcodeR)

geocode\_zip('92037')

## # A tibble: 1 × 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 × 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>

sd <- vax[ "San Diego" , ]  
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

sd <- filter(vax, county == "San Diego")  
nrow(sd)

## [1] 6527

sd.10 <- filter(vax, county == "San Diego" & age5\_plus\_population > 10000)

sd <- filter(vax, county == "San Diego")  
code <- sd[, c("zip\_code\_tabulation\_area")]  
length(unique(code))

## [1] 107

Q11. 107

sd12 <- filter(vax, county == "San Diego" & age12\_plus\_population > 60000)  
sd12[, c("age12\_plus\_population")]

## [1] 71820.2 61097.9 76365.2 64013.6 71642.8 71642.8 64013.6 76365.2 71820.2  
## [10] 61097.9 61097.9 71642.8 76365.2 64013.6 71820.2 61097.9 71820.2 76365.2  
## [19] 71642.8 64013.6 76365.2 64013.6 71642.8 71820.2 61097.9 76365.2 64013.6  
## [28] 71820.2 61097.9 71642.8 71642.8 76365.2 61097.9 64013.6 71820.2 64013.6  
## [37] 76365.2 61097.9 71642.8 71820.2 71642.8 76365.2 64013.6 61097.9 71820.2  
## [46] 61097.9 71820.2 71642.8 64013.6 76365.2 76365.2 71642.8 64013.6 61097.9  
## [55] 71820.2 71820.2 61097.9 64013.6 76365.2 71642.8 71642.8 64013.6 76365.2  
## [64] 61097.9 71820.2 71820.2 61097.9 71642.8 76365.2 64013.6 71642.8 76365.2  
## [73] 64013.6 71820.2 61097.9 71820.2 61097.9 64013.6 76365.2 71642.8 71642.8  
## [82] 64013.6 76365.2 71820.2 61097.9 71820.2 61097.9 76365.2 64013.6 71642.8  
## [91] 71642.8 64013.6 76365.2 71820.2 61097.9 71820.2 61097.9 71642.8 64013.6  
## [100] 76365.2 71642.8 64013.6 61097.9 71820.2 76365.2 61097.9 71642.8 64013.6  
## [109] 76365.2 71820.2 71642.8 76365.2 71820.2 64013.6 61097.9 71820.2 61097.9  
## [118] 64013.6 76365.2 71642.8 71642.8 71820.2 61097.9 76365.2 64013.6 71820.2  
## [127] 61097.9 71642.8 76365.2 64013.6 64013.6 76365.2 71642.8 61097.9 71820.2  
## [136] 76365.2 61097.9 71820.2 71642.8 64013.6 71642.8 64013.6 76365.2 61097.9  
## [145] 71820.2 71820.2 61097.9 71642.8 76365.2 64013.6 64013.6 71642.8 71820.2  
## [154] 76365.2 61097.9 71642.8 71820.2 61097.9 64013.6 76365.2 76365.2 64013.6  
## [163] 71820.2 61097.9 71642.8 71820.2 61097.9 71642.8 76365.2 64013.6 76365.2  
## [172] 64013.6 71642.8 71820.2 61097.9 71820.2 61097.9 71642.8 76365.2 64013.6  
## [181] 64013.6 76365.2 61097.9 71820.2 71642.8 61097.9 71642.8 64013.6 76365.2  
## [190] 71820.2 61097.9 71820.2 71642.8 64013.6 76365.2 64013.6 76365.2 71642.8  
## [199] 71820.2 61097.9 71820.2 61097.9 71642.8 76365.2 64013.6 76365.2 71642.8  
## [208] 64013.6 61097.9 71820.2 71820.2 61097.9 76365.2 64013.6 71642.8 71642.8  
## [217] 64013.6 61097.9 71820.2 76365.2 61097.9 71642.8 76365.2 64013.6 71820.2  
## [226] 71642.8 61097.9 64013.6 71820.2 76365.2 61097.9 71642.8 76365.2 71820.2  
## [235] 64013.6 64013.6 71642.8 76365.2 61097.9 71820.2 71642.8 76365.2 61097.9  
## [244] 64013.6 71820.2 76365.2 61097.9 71820.2 71642.8 64013.6 76365.2 64013.6  
## [253] 71642.8 61097.9 71820.2 61097.9 71820.2 71642.8 76365.2 64013.6 64013.6  
## [262] 71642.8 76365.2 61097.9 71820.2 61097.9 71820.2 76365.2 64013.6 71642.8  
## [271] 71642.8 64013.6 76365.2 61097.9 71820.2 61097.9 71820.2 71642.8 76365.2  
## [280] 64013.6 64013.6 76365.2 71642.8 71820.2 61097.9 71820.2 61097.9 76365.2  
## [289] 64013.6 71642.8 71642.8 76365.2 64013.6 71820.2 61097.9 71820.2 61097.9  
## [298] 64013.6 76365.2 71642.8 71642.8 64013.6 76365.2 71820.2 61097.9

which.max("age12\_plus\_population")

## Warning in which.max("age12\_plus\_population"): NAs introduced by coercion

## integer(0)

Q12. 92154

fullyvaxxed <- filter(vax, county == "San Diego", as\_of\_date == "2022-02-22")  
nrow(fullyvaxxed)

## [1] 107

percentvaxxed <- fullyvaxxed[, c("percent\_of\_population\_fully\_vaccinated")]  
percentvaxxedomit <- na.omit(percentvaxxed)  
sum(percentvaxxedomit)/nrow(fullyvaxxed) \* 100

## [1] 69.75742

Q13. 69.75742%

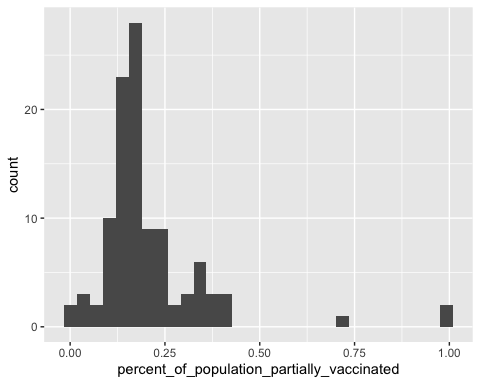
library(ggplot2)

Q14.

ggplot(data=fullyvaxxed, aes(percent\_of\_population\_partially\_vaccinated)) + geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1 rows containing non-finite values (stat\_bin).

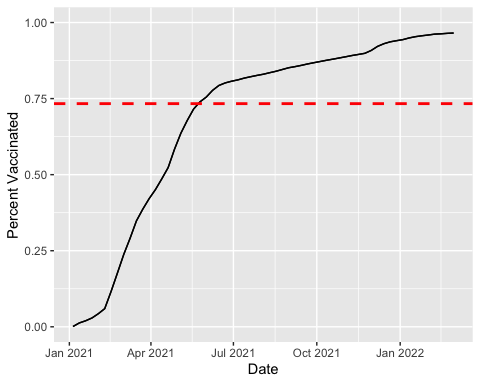


ucsd <- filter(sd, zip\_code\_tabulation\_area=="92037")  
ucsd[1,]$age5\_plus\_population

## [1] 36144

Q15. Q16.

lineplot <- ggplot(ucsd) +  
 aes(x=as\_of\_date, y=percent\_of\_population\_fully\_vaccinated) +  
 geom\_line() +  
 geom\_line(group=1) +  
 ylim(c(0,1)) +  
 labs(x="Date", y="Percent Vaccinated")   
lineplot + geom\_hline(yintercept=0.733385, linetype="dashed",   
 color = "red", size=1)



vax.36 <- filter(vax, age5\_plus\_population > 36144 & as\_of\_date == "2022-02-22")  
mean(vax.36[, c("percent\_of\_population\_fully\_vaccinated")])

## [1] 0.733385

min(vax.36[, c("percent\_of\_population\_fully\_vaccinated")])

## [1] 0.388109

median(vax.36[, c("percent\_of\_population\_fully\_vaccinated")])

## [1] 0.733275

max(vax.36[, c("percent\_of\_population\_fully\_vaccinated")])

## [1] 1

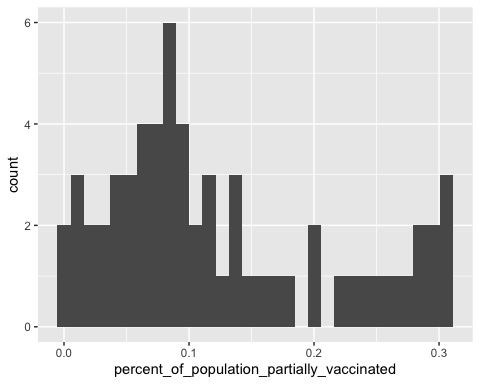
quantile(vax.36[, c("percent\_of\_population\_fully\_vaccinated")])

## 0% 25% 50% 75% 100%   
## 0.3881090 0.6539015 0.7332750 0.8027110 1.0000000

Q17. Min:0.3881090 1st Q:0.6539015 Median:0.7332750 3rd Q:0.8027110 Max:1.0000000 Mean:0.733385 Q18.

ggplot(data=ucsd, aes(percent\_of\_population\_partially\_vaccinated)) + geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



vax %>% filter(as\_of\_date == "2022-02-22") %>%   
 filter(zip\_code\_tabulation\_area=="92040") %>%  
 select(percent\_of\_population\_fully\_vaccinated)

## percent\_of\_population\_fully\_vaccinated  
## 1 0.551304

vax %>% filter(as\_of\_date == "2022-02-22") %>%   
 filter(zip\_code\_tabulation\_area=="92109") %>%  
 select(percent\_of\_population\_fully\_vaccinated)

## percent\_of\_population\_fully\_vaccinated  
## 1 0.723044

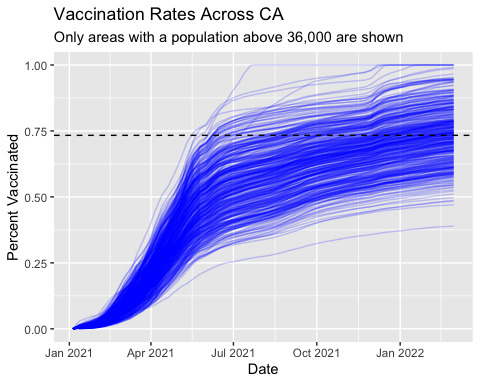
Q19: 92040: 0.551304 92019: 0.723044 Both below average.

vax.36.all <- filter(vax, age5\_plus\_population > 36144)

Q20.

ggplot(vax.36.all) +  
 aes(x=as\_of\_date, y=percent\_of\_population\_fully\_vaccinated,   
 group=zip\_code\_tabulation\_area) +  
 geom\_line(alpha=0.2, color="blue") +  
 ylim(c(0,1)) +  
 labs(x="Date", y="Percent Vaccinated",  
 title="Vaccination Rates Across CA",  
 subtitle="Only areas with a population above 36,000 are shown") +  
 geom\_hline(yintercept = 0.733385, linetype="dashed")

## Warning: Removed 311 row(s) containing missing values (geom\_path).



Q21. I have mixed feelings about it, since there are parts of the state with quite low vaccination rates as of just a few days ago; anyone who travels there has a greater risk of coming into contact with the virus.