Class 17: Mini Project

Jessica Diaz-Vigil

2023-05-31

Getting Started

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

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction      county
## 1 2021-01-05                   94579                   Alameda     Alameda
## 2 2021-01-05                   93726                    Fresno      Fresno
## 3 2021-01-05                   94305               Santa Clara Santa Clara
## 4 2021-01-05                   93704                    Fresno      Fresno
## 5 2021-01-05                   94403                 San Mateo   San Mateo
## 6 2021-01-05                   93668                    Fresno      Fresno
##   vaccine_equity_metric_quartile                vem_source
## 1                              3 Healthy Places Index Score
## 2                              1 Healthy Places Index Score
## 3                              4 Healthy Places Index Score
## 4                              1 Healthy Places Index Score
## 5                              4 Healthy Places Index Score
## 6                              1    CDPH-Derived ZCTA Score
##   age12_plus_population age5_plus_population tot_population
## 1               19192.7                20872          21883
## 2               33707.7                39067          42824
## 3               15716.9                16015          16397
## 4               24803.5                27701          29740
## 5               37967.5                41530          44408
## 6                1013.4                 1199           1219
##   persons_fully_vaccinated persons_partially_vaccinated
## 1                       NA                           NA
## 2                       NA                           NA
## 3                       NA                           NA
## 4                       NA                           NA
## 5                       NA                           NA
## 6                       NA                           NA
##   percent_of_population_fully_vaccinated
## 1                                     NA
## 2                                     NA
## 3                                     NA
## 4                                     NA
## 5                                     NA
## 6                                     NA
##   percent_of_population_partially_vaccinated
## 1                                         NA
```

```
## 2                                        NA
## 3                                        NA
## 4                                        NA
## 5                                        NA
## 6                                        NA
##   percent_of_population_with_1_plus_dose booster_recip_count
## 1                                     NA                  NA
## 2                                     NA                  NA
## 3                                     NA                  NA
## 4                                     NA                  NA
## 5                                     NA                  NA
## 6                                     NA                  NA
##   bivalent_dose_recip_count eligible_recipient_count
## 1                        NA                        4
## 2                        NA                        2
## 3                        NA                        8
## 4                        NA                        5
## 5                        NA                        7
## 6                        NA                        0
##   eligible_bivalent_recipient_count
## 1                                 4
## 2                                 2
## 3                                 8
## 4                                 5
## 5                                 7
## 6                                 0
##                                                                  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. What column details the total number of people fully vaccinated?

```
colnames(vax)
```

```
##  [1] "as_of_date"
##  [2] "zip_code_tabulation_area"
##  [3] "local_health_jurisdiction"
##  [4] "county"
##  [5] "vaccine_equity_metric_quartile"
##  [6] "vem_source"
##  [7] "age12_plus_population"
##  [8] "age5_plus_population"
##  [9] "tot_population"
## [10] "persons_fully_vaccinated"
## [11] "persons_partially_vaccinated"
## [12] "percent_of_population_fully_vaccinated"
## [13] "percent_of_population_partially_vaccinated"
## [14] "percent_of_population_with_1_plus_dose"
## [15] "booster_recip_count"
## [16] "bivalent_dose_recip_count"
## [17] "eligible_recipient_count"
```

```
## [18] "eligible_bivalent_recipient_count"
## [19] "redacted"
```

The column that represents the total number of people fully vaccinated it "persons_fully_vaccinated"

Q2. What column details the Zip code tabulation area?

```
colnames(vax)
```

```
##  [1] "as_of_date"
##  [2] "zip_code_tabulation_area"
##  [3] "local_health_jurisdiction"
##  [4] "county"
##  [5] "vaccine_equity_metric_quartile"
##  [6] "vem_source"
##  [7] "age12_plus_population"
##  [8] "age5_plus_population"
##  [9] "tot_population"
## [10] "persons_fully_vaccinated"
## [11] "persons_partially_vaccinated"
## [12] "percent_of_population_fully_vaccinated"
## [13] "percent_of_population_partially_vaccinated"
## [14] "percent_of_population_with_1_plus_dose"
## [15] "booster_recip_count"
## [16] "bivalent_dose_recip_count"
## [17] "eligible_recipient_count"
## [18] "eligible_bivalent_recipient_count"
## [19] "redacted"
```

The column that represents the total number of people fully vaccinated it "zip_code_tabulation_area"

Q3. What is the earliest date in this dataset?

```
table(vax$as_of_date)
```

```
##
## 2021-01-05 2021-01-12 2021-01-19 2021-01-26 2021-02-02 2021-02-09 2021-02-16
##       1764       1764       1764       1764       1764       1764       1764
## 2021-02-23 2021-03-02 2021-03-09 2021-03-16 2021-03-23 2021-03-30 2021-04-06
##       1764       1764       1764       1764       1764       1764       1764
## 2021-04-13 2021-04-20 2021-04-27 2021-05-04 2021-05-11 2021-05-18 2021-05-25
##       1764       1764       1764       1764       1764       1764       1764
## 2021-06-01 2021-06-08 2021-06-15 2021-06-22 2021-06-29 2021-07-06 2021-07-13
##       1764       1764       1764       1764       1764       1764       1764
## 2021-07-20 2021-07-27 2021-08-03 2021-08-10 2021-08-17 2021-08-24 2021-08-31
##       1764       1764       1764       1764       1764       1764       1764
## 2021-09-07 2021-09-14 2021-09-21 2021-09-28 2021-10-05 2021-10-12 2021-10-19
##       1764       1764       1764       1764       1764       1764       1764
## 2021-10-26 2021-11-02 2021-11-09 2021-11-16 2021-11-23 2021-11-30 2021-12-07
##       1764       1764       1764       1764       1764       1764       1764
## 2021-12-14 2021-12-21 2021-12-28 2022-01-04 2022-01-11 2022-01-18 2022-01-25
##       1764       1764       1764       1764       1764       1764       1764
## 2022-02-01 2022-02-08 2022-02-15 2022-02-22 2022-03-01 2022-03-08 2022-03-15
##       1764       1764       1764       1764       1764       1764       1764
## 2022-03-22 2022-03-29 2022-04-05 2022-04-12 2022-04-19 2022-04-26 2022-05-03
##       1764       1764       1764       1764       1764       1764       1764
## 2022-05-10 2022-05-17 2022-05-24 2022-05-31 2022-06-07 2022-06-14 2022-06-21
##       1764       1764       1764       1764       1764       1764       1764
```

```
## 2022-06-28 2022-07-05 2022-07-12 2022-07-19 2022-07-26 2022-08-02 2022-08-09
##       1764       1764       1764       1764       1764       1764       1764
## 2022-08-16 2022-08-23 2022-08-30 2022-09-06 2022-09-13 2022-09-20 2022-09-27
##       1764       1764       1764       1764       1764       1764       1764
## 2022-10-04 2022-10-11 2022-10-18 2022-10-25 2022-11-01 2022-11-08 2022-11-15
##       1764       1764       1764       1764       1764       1764       1764
## 2022-11-22 2022-11-29 2022-12-06 2022-12-13 2022-12-20 2022-12-27 2023-01-03
##       1764       1764       1764       1764       1764       1764       1764
## 2023-01-10 2023-01-17 2023-01-24 2023-01-31 2023-02-07 2023-02-14 2023-02-21
##       1764       1764       1764       1764       1764       1764       1764
## 2023-02-28 2023-03-07 2023-03-14 2023-03-21 2023-03-28 2023-04-04 2023-04-11
##       1764       1764       1764       1764       1764       1764       1764
## 2023-04-18 2023-04-25 2023-05-02 2023-05-09 2023-05-16 2023-05-23
##       1764       1764       1764       1764       1764       1764
```

The earliest data in this dataset is 2021-05-05

Q4. What is the latest date in this dataset?

```
table(vax$as_of_date)
```

```
##
## 2021-01-05 2021-01-12 2021-01-19 2021-01-26 2021-02-02 2021-02-09 2021-02-16
##       1764       1764       1764       1764       1764       1764       1764
## 2021-02-23 2021-03-02 2021-03-09 2021-03-16 2021-03-23 2021-03-30 2021-04-06
##       1764       1764       1764       1764       1764       1764       1764
## 2021-04-13 2021-04-20 2021-04-27 2021-05-04 2021-05-11 2021-05-18 2021-05-25
##       1764       1764       1764       1764       1764       1764       1764
## 2021-06-01 2021-06-08 2021-06-15 2021-06-22 2021-06-29 2021-07-06 2021-07-13
##       1764       1764       1764       1764       1764       1764       1764
## 2021-07-20 2021-07-27 2021-08-03 2021-08-10 2021-08-17 2021-08-24 2021-08-31
##       1764       1764       1764       1764       1764       1764       1764
## 2021-09-07 2021-09-14 2021-09-21 2021-09-28 2021-10-05 2021-10-12 2021-10-19
##       1764       1764       1764       1764       1764       1764       1764
## 2021-10-26 2021-11-02 2021-11-09 2021-11-16 2021-11-23 2021-11-30 2021-12-07
##       1764       1764       1764       1764       1764       1764       1764
## 2021-12-14 2021-12-21 2021-12-28 2022-01-04 2022-01-11 2022-01-18 2022-01-25
##       1764       1764       1764       1764       1764       1764       1764
## 2022-02-01 2022-02-08 2022-02-15 2022-02-22 2022-03-01 2022-03-08 2022-03-15
##       1764       1764       1764       1764       1764       1764       1764
## 2022-03-22 2022-03-29 2022-04-05 2022-04-12 2022-04-19 2022-04-26 2022-05-03
##       1764       1764       1764       1764       1764       1764       1764
## 2022-05-10 2022-05-17 2022-05-24 2022-05-31 2022-06-07 2022-06-14 2022-06-21
##       1764       1764       1764       1764       1764       1764       1764
## 2022-06-28 2022-07-05 2022-07-12 2022-07-19 2022-07-26 2022-08-02 2022-08-09
##       1764       1764       1764       1764       1764       1764       1764
## 2022-08-16 2022-08-23 2022-08-30 2022-09-06 2022-09-13 2022-09-20 2022-09-27
##       1764       1764       1764       1764       1764       1764       1764
## 2022-10-04 2022-10-11 2022-10-18 2022-10-25 2022-11-01 2022-11-08 2022-11-15
##       1764       1764       1764       1764       1764       1764       1764
## 2022-11-22 2022-11-29 2022-12-06 2022-12-13 2022-12-20 2022-12-27 2023-01-03
##       1764       1764       1764       1764       1764       1764       1764
## 2023-01-10 2023-01-17 2023-01-24 2023-01-31 2023-02-07 2023-02-14 2023-02-21
##       1764       1764       1764       1764       1764       1764       1764
## 2023-02-28 2023-03-07 2023-03-14 2023-03-21 2023-03-28 2023-04-04 2023-04-11
##       1764       1764       1764       1764       1764       1764       1764
```

```
## 2023-04-18 2023-04-25 2023-05-02 2023-05-09 2023-05-16 2023-05-23
##       1764       1764       1764       1764       1764       1764
```

The latest date in this data set is 2023-05-30

```
skimr::skim_without_charts(vax)
```

Table 1: Data summary

Name	vax
Number of rows	220500
Number of columns	19
Column type frequency:	
character	5
numeric	14
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	125	0
local_health_jurisdiction	0	1	0	15	625	62	0
county	0	1	0	15	625	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
zip_code_tabulation_area	0	1.00	93665.11	1817.38	90001	92257.75	93658.50	95380.50	97635.0
vaccine_equity_metric_quartile	10875	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0
age12_plus_population	0	1.00	18895.04	18993.87	0	1346.95	13685.10	31756.12	88556.7
age5_plus_population	0	1.00	20875.24	21105.97	0	1460.50	15364.00	34877.00	101902.0
tot_population	10750	0.95	23372.77	22628.50	12	2126.00	18714.00	38168.00	111165.0
persons_fully_vaccinated	17711	0.92	14272.72	15264.17	11	954.00	8990.00	23782.00	87724.0
persons_partially_vaccinated	17711	0.92	1711.05	2071.56	11	164.00	1203.00	2550.00	42259.0
percent_of_population_fully_vaccinated	22570	0.90	0.58	0.25	0	0.44	0.62	0.75	1.0
percent_of_population_partially_vaccinated	22570	0.90	0.08	0.09	0	0.05	0.06	0.08	1.0
percent_of_population_with_1_plus_dose	23732	0.89	0.64	0.24	0	0.50	0.68	0.82	1.0
booster_recip_count	74388	0.66	6373.43	7751.70	11	328.00	3097.00	10274.00	60022.0
bivalent_dose_recip_count	159956	0.27	3407.91	4010.38	11	222.00	1832.00	5482.00	29484.0
eligible_recipient_count	0	1.00	13120.40	15126.17	0	534.00	6663.00	22517.25	87437.0
eligible_bivalent_recipient_count	0	1.00	13016.51	15199.08	0	266.00	6562.00	22513.00	87437.0

Q5. How many numeric columns are in this dataset?

By looking at the data summary using skim, there are 14 numeric columns in the dataset

Q6. Note that there are "missing values" in the dataset. How many `NA` values there in the `persons_fully_vaccinated`column?

```
total_na_full_vac <- sum(is.na(vax$persons_fully_vaccinated))
total_na_full_vac
```

```
## [1] 17711
```

There are 17711 NA values in the "persons_fully_vaccinated" column

Q7. What percent of `persons_fully_vaccinated` values are missing (to 2 significant figures)?

```
total_full_vac <- nrow(vax)
total_full_vac
```

```
## [1] 220500
```

```
total_na_full_vac / total_full_vac * 100
```

```
## [1] 8.0322
```

The percent of missing values is 8.03%

Working With Dates

```
library(lubridate)
```

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

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

```
today()
```

```
## [1] "2023-05-31"
```

```
vax$as_of_date <- ymd(vax$as_of_date)
```

```
days_since_first <- today() - vax$as_of_date[1]
days_since_first
```

```
## Time difference of 876 days
```

```
days_in_data <- vax$as_of_date[nrow(vax)] - vax$as_of_date[1]
days_in_data
```

```
## Time difference of 868 days
```

Q9. How many days have passed since the last update of the dataset?

```
days_since_first - days_in_data
```

```
## Time difference of 8 days
```

8 days have passed since the last update

Q10. How many unique dates are in the dataset (i.e. how many different dates are detailed)?

```
unique_dates <- unique(vax$as_of_date)
unique_dates
```

```
##   [1] "2021-01-05" "2021-01-12" "2021-01-19" "2021-01-26" "2021-02-02"
##   [6] "2021-02-09" "2021-02-16" "2021-02-23" "2021-03-02" "2021-03-09"
```

```
##  [11] "2021-03-16" "2021-03-23" "2021-03-30" "2021-04-06" "2021-04-13"
##  [16] "2021-04-20" "2021-04-27" "2021-05-04" "2021-05-11" "2021-05-18"
##  [21] "2021-05-25" "2021-06-01" "2021-06-08" "2021-06-15" "2021-06-22"
##  [26] "2021-06-29" "2021-07-06" "2021-07-13" "2021-07-20" "2021-07-27"
##  [31] "2021-08-03" "2021-08-10" "2021-08-17" "2021-08-24" "2021-08-31"
##  [36] "2021-09-07" "2021-09-14" "2021-09-21" "2021-09-28" "2021-10-05"
##  [41] "2021-10-12" "2021-10-19" "2021-10-26" "2021-11-02" "2021-11-09"
##  [46] "2021-11-16" "2021-11-23" "2021-11-30" "2021-12-07" "2021-12-14"
##  [51] "2021-12-21" "2021-12-28" "2022-01-04" "2022-01-11" "2022-01-18"
##  [56] "2022-01-25" "2022-02-01" "2022-02-08" "2022-02-15" "2022-02-22"
##  [61] "2022-03-01" "2022-03-08" "2022-03-15" "2022-03-22" "2022-03-29"
##  [66] "2022-04-05" "2022-04-12" "2022-04-19" "2022-04-26" "2022-05-03"
##  [71] "2022-05-10" "2022-05-17" "2022-05-24" "2022-05-31" "2022-06-07"
##  [76] "2022-06-14" "2022-06-21" "2022-06-28" "2022-07-05" "2022-07-12"
##  [81] "2022-07-19" "2022-07-26" "2022-08-02" "2022-08-09" "2022-08-16"
##  [86] "2022-08-23" "2022-08-30" "2022-09-06" "2022-09-13" "2022-09-20"
##  [91] "2022-09-27" "2022-10-04" "2022-10-11" "2022-10-18" "2022-10-25"
##  [96] "2022-11-01" "2022-11-08" "2022-11-15" "2022-11-22" "2022-11-29"
## [101] "2022-12-06" "2022-12-13" "2022-12-20" "2022-12-27" "2023-01-03"
## [106] "2023-01-10" "2023-01-17" "2023-01-24" "2023-01-31" "2023-02-07"
## [111] "2023-02-14" "2023-02-21" "2023-02-28" "2023-03-07" "2023-03-14"
## [116] "2023-03-21" "2023-03-28" "2023-04-04" "2023-04-11" "2023-04-18"
## [121] "2023-04-25" "2023-05-02" "2023-05-09" "2023-05-16" "2023-05-23"
```

```
length(unique_dates)
```

```
## [1] 125
```

There are 125 unique dates in the dataset

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
## # i 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>
```

Focus On The San Diego Area

```
sd <- vax[ vax$county == "San Diego" , ]
head(sd)
```

```
##     as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 134 2021-01-05                   91977                San Diego San Diego
## 171 2021-01-05                   92110                San Diego San Diego
## 173 2021-01-05                   92101                San Diego San Diego
## 175 2021-01-05                   92071                San Diego San Diego
## 176 2021-01-05                   92070                San Diego San Diego
## 178 2021-01-05                   92028                San Diego San Diego
##     vaccine_equity_metric_quartile                vem_source
## 134                              2 Healthy Places Index Score
## 171                              3 Healthy Places Index Score
## 173                              2 Healthy Places Index Score
## 175                              3 Healthy Places Index Score
## 176                              2    CDPH-Derived ZCTA Score
## 178                              2 Healthy Places Index Score
##     age12_plus_population age5_plus_population tot_population
## 134               53851.0                59911          64750
## 171               27003.5                28597          30108
## 173               39588.5                40077          41159
## 175               49137.8                53795          57710
## 176                 682.4                  743            786
## 178               41252.1                44782          48173
##     persons_fully_vaccinated persons_partially_vaccinated
## 134                       12                         950
## 171                       19                         550
## 173                       45                         991
## 175                       22                        1417
## 176                       NA                          NA
## 178                       14                         509
##     percent_of_population_fully_vaccinated
## 134                               0.000185
## 171                               0.000631
## 173                               0.001093
## 175                               0.000381
## 176                                     NA
## 178                               0.000291
##     percent_of_population_partially_vaccinated
## 134                                   0.014672
## 171                                   0.018268
## 173                                   0.024077
## 175                                   0.024554
## 176                                         NA
## 178                                   0.010566
```

```
##     percent_of_population_with_1_plus_dose booster_recip_count
## 134                                0.014857                  NA
## 171                                0.018899                  NA
## 173                                0.025170                  NA
## 175                                0.024935                  NA
## 176                                      NA                  NA
## 178                                0.010857                  NA
##     bivalent_dose_recip_count eligible_recipient_count
## 134                        NA                       12
## 171                        NA                       19
## 173                        NA                       45
## 175                        NA                       22
## 176                        NA                        0
## 178                        NA                       14
##     eligible_bivalent_recipient_count
## 134                                12
## 171                                19
## 173                                45
## 175                                22
## 176                                 0
## 178                                14
##                                                                    redacted
## 134 Information redacted in accordance with CA state privacy requirements
## 171 Information redacted in accordance with CA state privacy requirements
## 173 Information redacted in accordance with CA state privacy requirements
## 175 Information redacted in accordance with CA state privacy requirements
## 176 Information redacted in accordance with CA state privacy requirements
## 178 Information redacted in accordance with CA state privacy requirements
```

```
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] 13375
```

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

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

```
unique_zip_sd <- unique(sd$zip_code_tabulation_area)
length(unique_zip_sd)
```

```
## [1] 107
```

There are 107 unique zip codes in San Diego

Q12. What San Diego County Zip code area has the largest population in this dataset?

```
which.max(sd$tot_population)
```

```
## [1] 87
```

```
sd$zip_code_tabulation_area[87]
```

```
## [1] 92154
```

The zip code 92154 has the largest population

Q13. What is the overall average (with 2 decimal numbers) "Percent of Population Fully Vaccinated" value for all San Diego "County" as of "2023-05-23"?

```
sd.1 <- filter(sd.10, county == "San Diego" &
                 as_of_date == "2023-05-23")
head(sd.1)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction    county
## 1 2023-05-23                   91932                San Diego San Diego
## 2 2023-05-23                   92124                San Diego San Diego
## 3 2023-05-23                   92014                San Diego San Diego
## 4 2023-05-23                   92009                San Diego San Diego
## 5 2023-05-23                   92057                San Diego San Diego
## 6 2023-05-23                   92102                San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 1                              2 Healthy Places Index Score
## 2                              3 Healthy Places Index Score
## 3                              4 Healthy Places Index Score
## 4                              4 Healthy Places Index Score
## 5                              2 Healthy Places Index Score
## 6                              1 Healthy Places Index Score
##   age12_plus_population age5_plus_population tot_population
## 1               21968.2                24874          26492
## 2               25422.4                29040          32600
## 3               11942.5                13149          13568
## 4               39183.5                43710          46612
## 5               51927.0                56906          60414
## 6               37042.3                41033          44010
##   persons_fully_vaccinated persons_partially_vaccinated
## 1                    18553                         2530
## 2                    18902                         2508
## 3                    11564                         1132
## 4                    34280                         2815
## 5                    38168                         4091
## 6                    34082                         3761
##   percent_of_population_fully_vaccinated
## 1                               0.700325
## 2                               0.579816
## 3                               0.852300
## 4                               0.735433
## 5                               0.631774
## 6                               0.774415
##   percent_of_population_partially_vaccinated
## 1                                   0.095501
## 2                                   0.076933
```

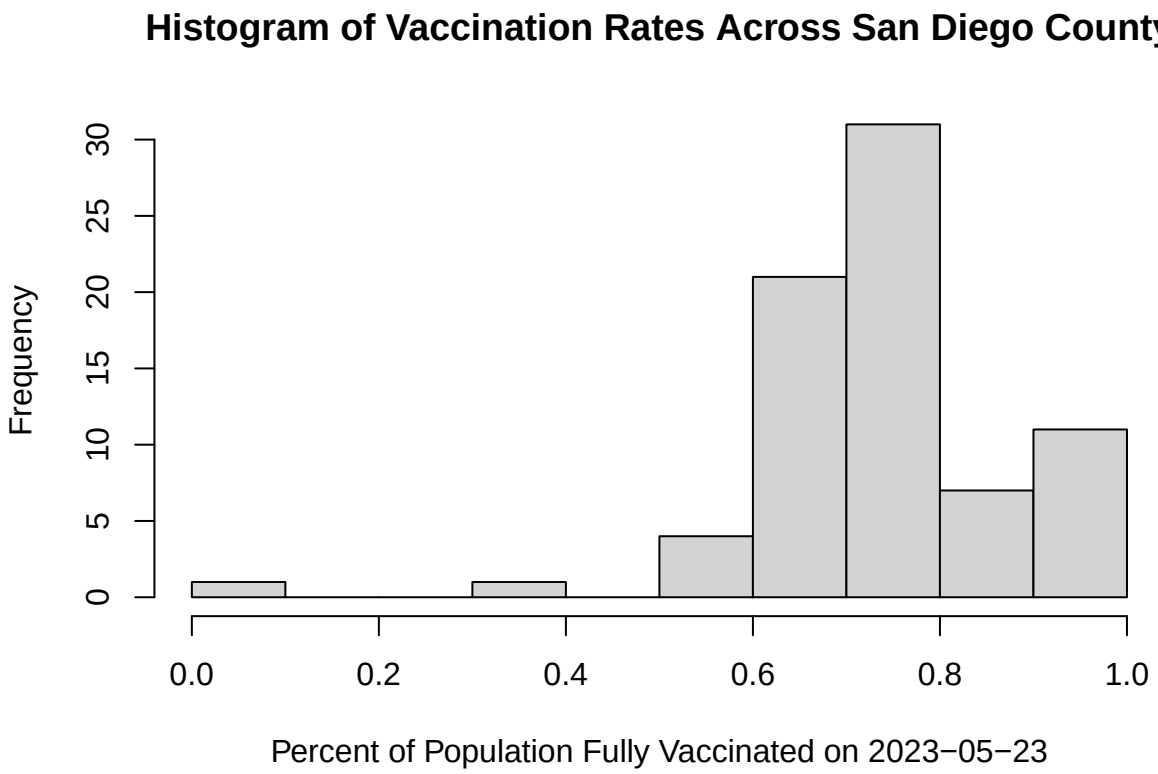
```
## 3                                  0.083432
## 4                                  0.060392
## 5                                  0.067716
## 6                                  0.085458
##   percent_of_population_with_1_plus_dose booster_recip_count
## 1                               0.795826                9928
## 2                               0.656749               12196
## 3                               0.935732                8595
## 4                               0.795825               23395
## 5                               0.699490               22983
## 6                               0.859873               19859
##   bivalent_dose_recip_count eligible_recipient_count
## 1                      3225                    18507
## 2                      5812                    18737
## 3                      4840                    11544
## 4                     11264                    34163
## 5                      9241                    38055
## 6                      6881                    33981
##   eligible_bivalent_recipient_count redacted
## 1                             18507       No
## 2                             18737       No
## 3                             11544       No
## 4                             34163       No
## 5                             38055       No
## 6                             33981       No
```

```
mean(na.omit(sd.1$percent_of_population_fully_vaccinated))
```

```
## [1] 0.7383022
```

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 "2023-05-23"?

```
hist(sd.1$percent_of_population_fully_vaccinated, main = "Histogram of Vaccination Rates Across San Dieg
```



Focus On UCSD/La Jolla

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

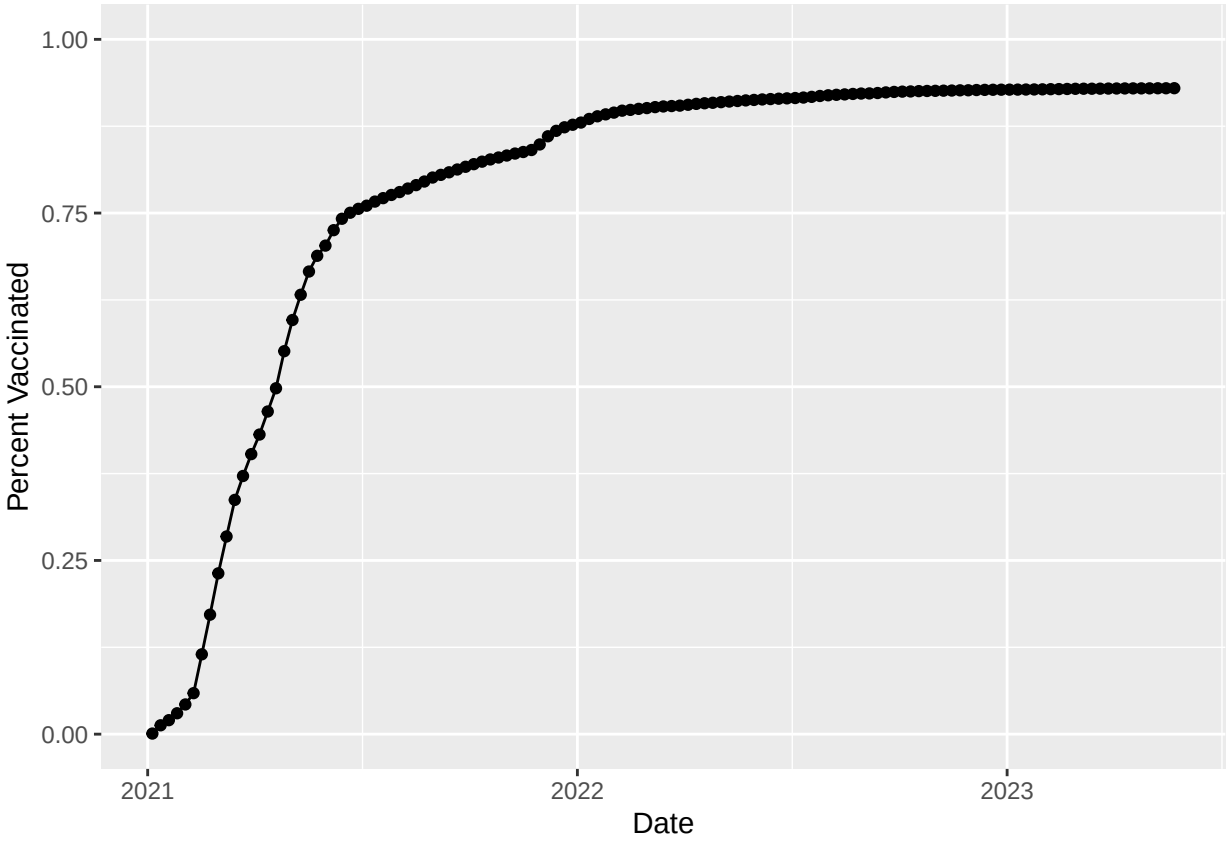
```
## [1] 36144
```

```
dim(ucsd)
```

```
## [1] 125  19
```

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

```
library(ggplot2)
p <- ggplot(ucsd) +
  aes(x = as_of_date,
      y = percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated")
p
```



Comparing To Similar Sized Areas

```
vax.36 <- filter(vax, age5_plus_population > 36144 &
                 as_of_date == "2023-05-23")
head(vax.36)
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction       county
## 1 2023-05-23                   93720                     Fresno       Fresno
## 2 2023-05-23                   95670                 Sacramento   Sacramento
## 3 2023-05-23                   91405                Los Angeles  Los Angeles
## 4 2023-05-23                   94582               Contra Costa Contra Costa
## 5 2023-05-23                   95687                     Solano       Solano
## 6 2023-05-23                   92627                     Orange       Orange
##   vaccine_equity_metric_quartile                vem_source
## 1                              3 Healthy Places Index Score
## 2                              2 Healthy Places Index Score
## 3                              1 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 tot_population
## 1               40357.3                44412          47081
## 2               46783.6                52133          55558
## 3               46561.6                51961          55506
## 4               34809.5                40433          42576
```

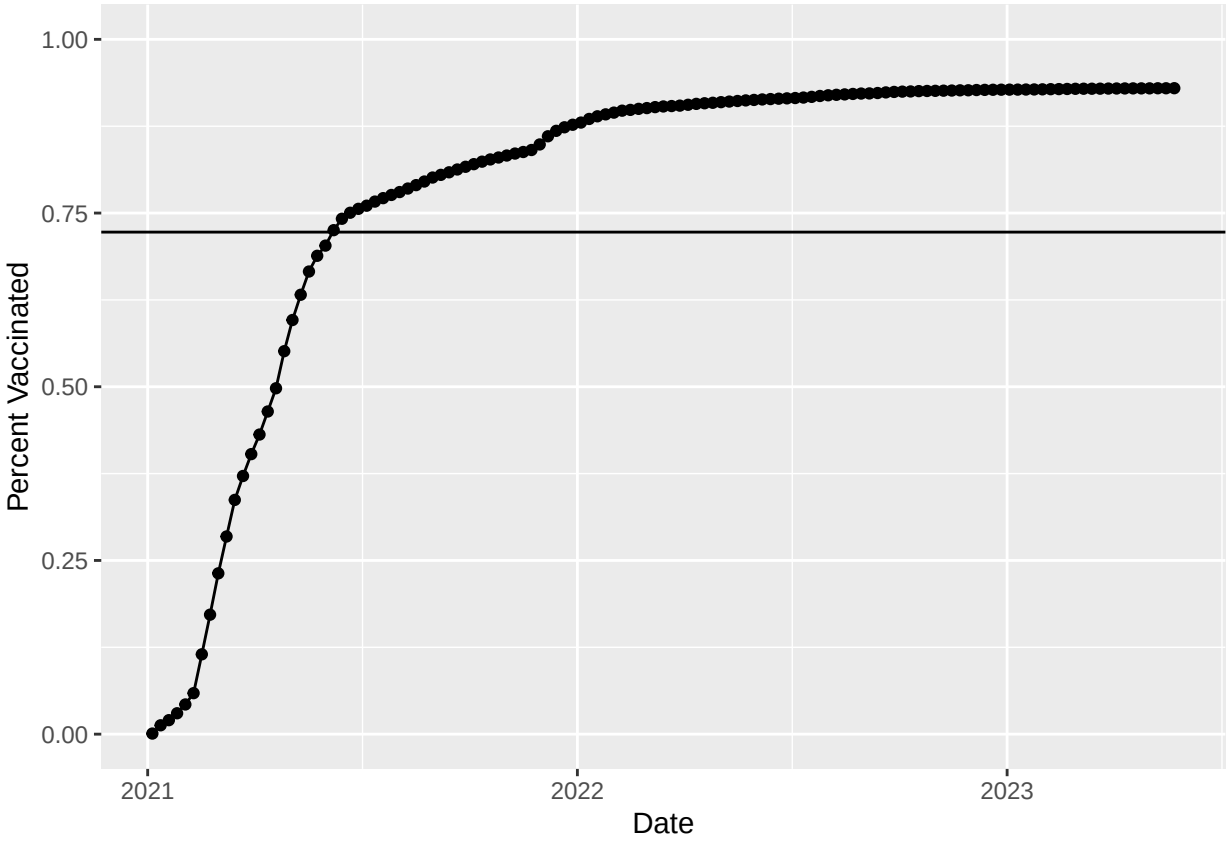
```
## 5               59036.1                65398          69060
## 6               54060.2                59229          63161
##   persons_fully_vaccinated persons_partially_vaccinated
## 1                    33810                         3122
## 2                    35674                         3418
## 3                    37040                         4832
## 4                    44338                         3214
## 5                    40549                         4178
## 6                    40189                         3798
##   percent_of_population_fully_vaccinated
## 1                               0.718124
## 2                               0.642104
## 3                               0.667315
## 4                               1.000000
## 5                               0.587156
## 6                               0.636295
##   percent_of_population_partially_vaccinated
## 1                                   0.066311
## 2                                   0.061521
## 3                                   0.087054
## 4                                   0.075489
## 5                                   0.060498
## 6                                   0.060132
##   percent_of_population_with_1_plus_dose booster_recip_count
## 1                               0.784435               21186
## 2                               0.703625               21712
## 3                               0.754369               18988
## 4                               1.000000               33971
## 5                               0.647654               24494
## 6                               0.696427               21494
##   bivalent_dose_recip_count eligible_recipient_count
## 1                      8056                    33740
## 2                     10016                    35587
## 3                      6688                    36977
## 4                     16642                    44050
## 5                     10308                    40460
## 6                      7819                    40104
##   eligible_bivalent_recipient_count redacted
## 1                             33740       No
## 2                             35587       No
## 3                             36977       No
## 4                             44050       No
## 5                             40460       No
## 6                             40104       No
```

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* "2023-05-23". Add this as a straight horizontal line to your plot from above with the `geom_hline()`function?

```
mean.vax.36 <- mean(vax.36$percent_of_population_fully_vaccinated)
mean.vax.36
```

```
## [1] 0.7225892
```

```
p + geom_hline(yintercept = mean.vax.36)
```



Q17. What is the 6 number summary (Min, 1st Qu., Median, Mean, 3rd Qu., and Max) of the *"Percent of Population Fully Vaccinated"*values for ZIP code areas with a population as large as 92037 (La Jolla) *as_of_date* "2023-05-23"?

```
summary(vax.36$percent_of_population_fully_vaccinated)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3816  0.6469  0.7207  0.7226  0.7924  1.0000
```

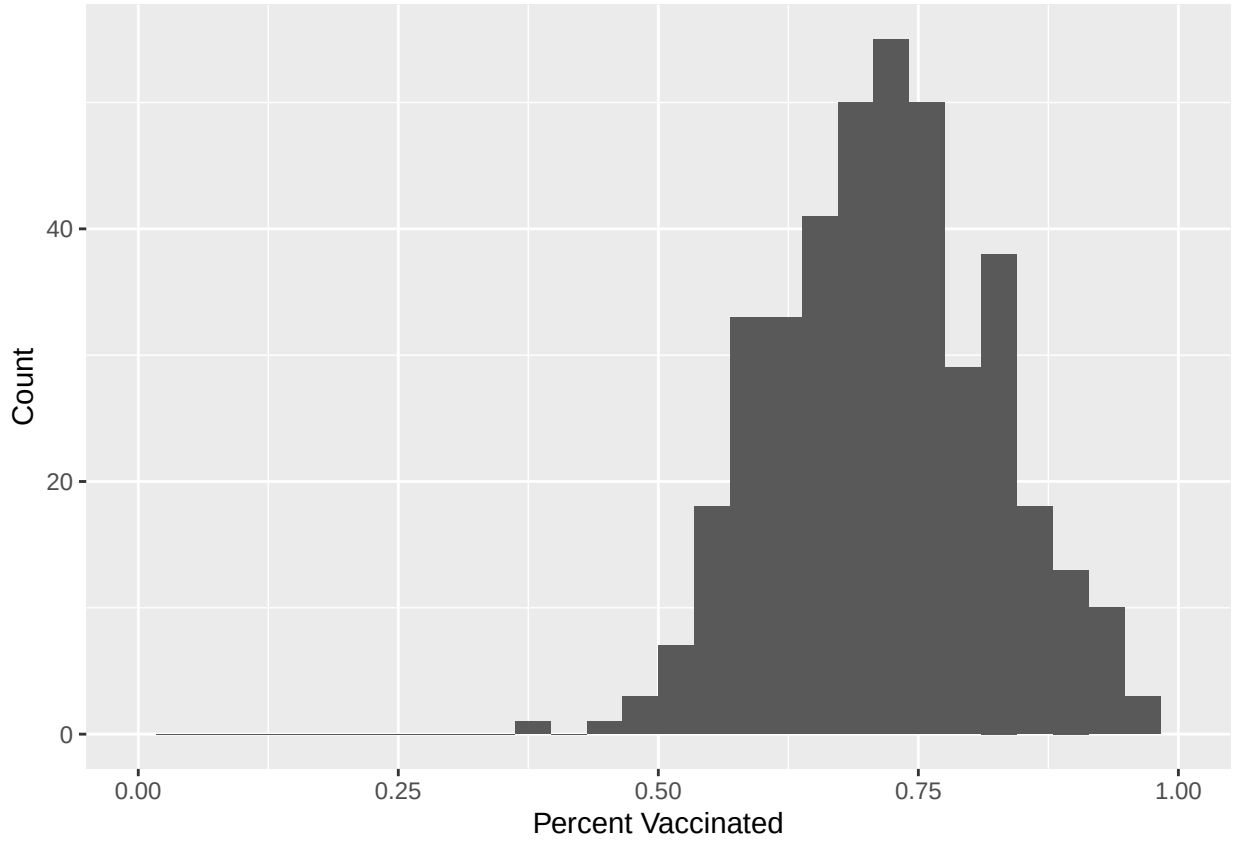
The 6 number summary is above

Q18. Using ggplot generate a histogram of this data

```
ggplot(vax.36) +
  aes(x = percent_of_population_fully_vaccinated) +
  geom_histogram() +
  xlab("Percent Vaccinated") +
  ylab("Count") +
  xlim(c(0,1))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```



Q19. Is the 92109 and 92040 ZIP code areas above or below the average value you calculated for all these above?

```
vax %>% filter(as_of_date == "2023-05-23") %>%
  filter(zip_code_tabulation_area=="92040") %>%
  select(percent_of_population_fully_vaccinated)
```

```
##   percent_of_population_fully_vaccinated
## 1                               0.552434
```

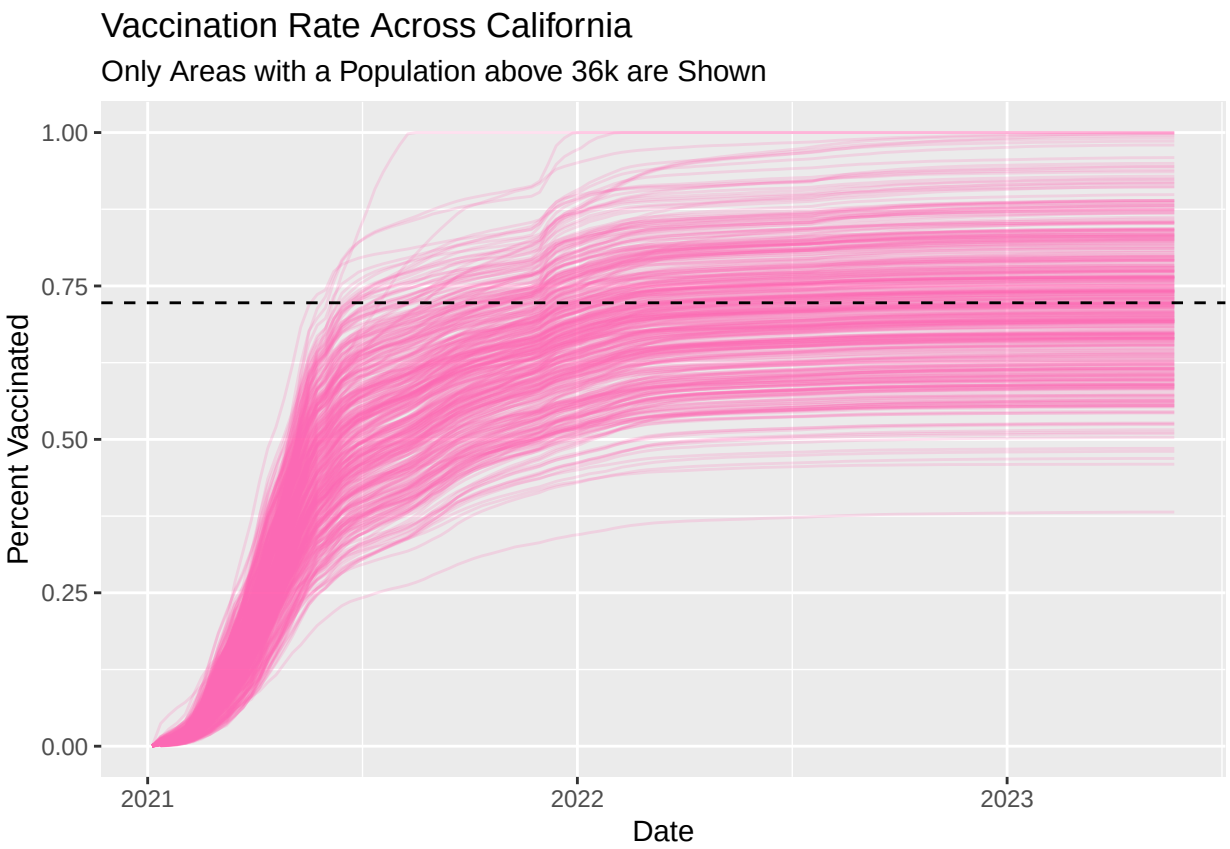
This value is below the other average we found (0.55 compared to 0.72)

Q20. Finally make a time course plot of vaccination progress for all areas in the full dataset with aage5_plus_population > 36144

```
vax.36.all <- filter(vax, age5_plus_population > 36144)

ggplot(vax.36.all) +
  aes(as_of_date,
      percent_of_population_fully_vaccinated,
      group=zip_code_tabulation_area) +
  geom_line(alpha=0.2, color="hotpink") +
  ylim(c(0,1)) +
  labs(x="Date", y="Percent Vaccinated",
       title="Vaccination Rate Across California",
       subtitle="Only Areas with a Population above 36k are Shown") +
  geom_hline(yintercept = mean.vax.36, linetype="dashed")
```

```
## Warning: Removed 185 rows containing missing values (`geom_line()`).
```


Vaccination Rate Across California
Only Areas with a Population above 36k are Shown
Percent Vaccinated
1.00
0.75
0.50
0.25
0.00
2021
2022
2023
Date