

# Class 17: Vaccine Mini Project

Aparajita Pranjal

5/31/23

## Importing data

```
# 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	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	4
2	2
3	8
4	5
5	7
6	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? **Column 10 - “persons 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"
```

- **Q2.** What column details the Zip code tabulation area?

**Column 2 - “zip\_code\_tabulation\_area”**

- **Q3.** What is the earliest date in this dataset?

**2021-01-05**

- **Q4.** What is the latest date in this dataset?

**2023-05-23**

Summary of data:

```
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	93665.11817	3890001	192257.79	3658.50	5380.50	7635.0	
vaccine_equity_metric_q1	10875	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0
age12_plus_population	0	1.00	18895.04	8993.87	0	1346.95	13685.10	1756.12	8556.7
age5_plus_population	0	1.00	20875.22	1105.97	0	1460.50	15364.00	4877.00	1902.0
tot_population	10750	0.95	23372.72	2628.50	12	2126.00	18714.00	8168.00	11165.0
persons_fully_vaccinated	17711	0.92	14272.72	25264.17	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	22579	0.90	0.58	0.25	0	0.44	0.62	0.75	1.0
percent_of_population_partially_vaccinated	22579	0.90	0.08	0.09	0	0.05	0.06	0.08	1.0
percent_of_population_waiting_for_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	5126.17	0	534.00	6663.00	22517.25	87437.0
eligible_bivalent_recipient_count	0	1.00	13016.51	5199.08	0	266.00	6562.00	22513.00	87437.0

- **Q5.** How many numeric columns are in this dataset?

#### 14 numeric columns

- **Q6.** Note that there are "missing values" in the dataset. How many NA values there in the `persons_fully_vaccinated` column? **17711**

```
sum( is.na(vax$persons_fully_vaccinated) )
```

```
[1] 17711
```

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

$(17711/220500) * 100 = 8.03\%$

- **Q8.** [Optional]: Why might this data be missing?

**If they were vaccinated in another state their medical records may not have been updated.**

Using lubridate:

```
library(lubridate)
```

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

`date, intersect, setdiff, union`

```
today()
```

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

Incorrect use of package:

```
# This will give an Error!  
#today() - vax$as_of_date[1]
```

Changing the date into the correct format:

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

Difference from today and first date of reported vaccine:

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

Time difference of 876 days

To find the difference between the first and last reported date:

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

Time difference of 868 days

- **Q9.** How many days have passed since the last update of the dataset? **8 days**

```
today() - ymd(vax$as_of_date[nrow(vax)])
```

Time difference of 8 days

- **Q10.** How many unique dates are in the dataset (i.e. how many different dates are detailed)? **125 unique dates**

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

```
[1] 125
```

## Working with zipcodes:

```
library(zipcodeR)
```

The legacy packages `maptools`, `rgdal`, and `rgeos`, underpinning this package will retire shortly. Please refer to R-spatial evolution reports on <https://r-spatial.org/r/2023/05/15/evolution4.html> for details.

This package is now running under evolution status 0

Coordinates of La Jolla zipcode:

```
geocode_zip('92037')
```

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

Calculating the distance between zipcode:

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

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

For census data:

```
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>
```

Census data for vaccine file zipcodes:

```
# Pull data for all ZIP codes in the dataset
zipdata <- reverse_zipcode( vax$zip_code_tabulation_area )
```

Focusing on San Diego using base R:

```
# Subset to San Diego county only areas
sd <- vax[vax$county == "San Diego", ]
dim(sd)
```

```
[1] 13375    19
```

Using 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
```

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

```
[1] 13375
```

Filtering data further:

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

- **Q11.** How many distinct zip codes are listed for San Diego County? **107 unique zip codes**

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

```
[1] 107
```



- **Q12.** What San Diego County Zip code area has the largest population in this dataset?  
**92154**

```
which.max(sd.10$tot_population)
```

```
[1] 64
```

```
sd.10$zip_code_tabulation_area[64]
```

```
[1] 92154
```

- **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"? **0.74**

```
sd.23 <- filter(sd, as_of_date == '2023-05-23')
sd.23
```

	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
7	2023-05-23	92082	San Diego	San Diego
8	2023-05-23	92029	San Diego	San Diego
9	2023-05-23	92131	San Diego	San Diego
10	2023-05-23	92061	San Diego	San Diego
11	2023-05-23	91916	San Diego	San Diego
12	2023-05-23	92026	San Diego	San Diego
13	2023-05-23	91950	San Diego	San Diego
14	2023-05-23	91935	San Diego	San Diego
15	2023-05-23	91934	San Diego	San Diego
16	2023-05-23	91914	San Diego	San Diego
17	2023-05-23	91962	San Diego	San Diego
18	2023-05-23	91978	San Diego	San Diego
19	2023-05-23	91906	San Diego	San Diego
20	2023-05-23	91913	San Diego	San Diego
21	2023-05-23	91942	San Diego	San Diego
22	2023-05-23	91948	San Diego	San Diego
23	2023-05-23	91905	San Diego	San Diego
24	2023-05-23	91910	San Diego	San Diego

25	2023-05-23	92104	San Diego	San Diego
26	2023-05-23	92075	San Diego	San Diego
27	2023-05-23	92083	San Diego	San Diego
28	2023-05-23	92084	San Diego	San Diego
29	2023-05-23	92114	San Diego	San Diego
30	2023-05-23	92115	San Diego	San Diego
31	2023-05-23	92120	San Diego	San Diego
32	2023-05-23	92122	San Diego	San Diego
33	2023-05-23	92154	San Diego	San Diego
34	2023-05-23	92066	San Diego	San Diego
35	2023-05-23	92091	San Diego	San Diego
36	2023-05-23	92103	San Diego	San Diego
37	2023-05-23	92113	San Diego	San Diego
38	2023-05-23	92116	San Diego	San Diego
39	2023-05-23	92118	San Diego	San Diego
40	2023-05-23	92060	San Diego	San Diego
41	2023-05-23	92008	San Diego	San Diego
42	2023-05-23	92036	San Diego	San Diego
43	2023-05-23	92106	San Diego	San Diego
44	2023-05-23	92108	San Diego	San Diego
45	2023-05-23	92004	San Diego	San Diego
46	2023-05-23	92007	San Diego	San Diego
47	2023-05-23	92037	San Diego	San Diego
48	2023-05-23	92020	San Diego	San Diego
49	2023-05-23	92025	San Diego	San Diego
50	2023-05-23	92027	San Diego	San Diego
51	2023-05-23	92065	San Diego	San Diego
52	2023-05-23	92105	San Diego	San Diego
53	2023-05-23	92107	San Diego	San Diego
54	2023-05-23	92109	San Diego	San Diego
55	2023-05-23	92121	San Diego	San Diego
56	2023-05-23	92128	San Diego	San Diego
57	2023-05-23	91915	San Diego	San Diego
58	2023-05-23	91980	San Diego	San Diego
59	2023-05-23	91963	San Diego	San Diego
60	2023-05-23	91941	San Diego	San Diego
61	2023-05-23	91911	San Diego	San Diego
62	2023-05-23	91977	San Diego	San Diego
63	2023-05-23	92021	San Diego	San Diego
64	2023-05-23	92003	San Diego	San Diego
65	2023-05-23	92059	San Diego	San Diego
66	2023-05-23	92024	San Diego	San Diego
67	2023-05-23	92028	San Diego	San Diego

68	2023-05-23	92070	San Diego	San Diego
69	2023-05-23	92071	San Diego	San Diego
70	2023-05-23	92101	San Diego	San Diego
71	2023-05-23	92110	San Diego	San Diego
72	2023-05-23	92111	San Diego	San Diego
73	2023-05-23	92119	San Diego	San Diego
74	2023-05-23	92127	San Diego	San Diego
75	2023-05-23	92129	San Diego	San Diego
76	2023-05-23	92067	San Diego	San Diego
77	2023-05-23	92132	San Diego	San Diego
78	2023-05-23	92081	San Diego	San Diego
79	2023-05-23	92058	San Diego	San Diego
80	2023-05-23	92055	San Diego	San Diego
81	2023-05-23	92011	San Diego	San Diego
82	2023-05-23	92010	San Diego	San Diego
83	2023-05-23	91945	San Diego	San Diego
84	2023-05-23	92019	San Diego	San Diego
85	2023-05-23	92155	San Diego	San Diego
86	2023-05-23	92064	San Diego	San Diego
87	2023-05-23	92069	San Diego	San Diego
88	2023-05-23	92086	San Diego	San Diego
89	2023-05-23	92147	San Diego	San Diego
90	2023-05-23	92126	San Diego	San Diego
91	2023-05-23	92145	San Diego	San Diego
92	2023-05-23	92140	San Diego	San Diego
93	2023-05-23	92130	San Diego	San Diego
94	2023-05-23	92135	San Diego	San Diego
95	2023-05-23	92134	San Diego	San Diego
96	2023-05-23	92056	San Diego	San Diego
97	2023-05-23	92054	San Diego	San Diego
98	2023-05-23	92040	San Diego	San Diego
99	2023-05-23	91902	San Diego	San Diego
100	2023-05-23	91901	San Diego	San Diego
101	2023-05-23	92139	San Diego	San Diego
102	2023-05-23	92123	San Diego	San Diego
103	2023-05-23	92117	San Diego	San Diego
104	2023-05-23	92078	San Diego	San Diego
105	2023-05-23	92173	San Diego	San Diego
106	2023-05-23	91917	San Diego	San Diego
107	2023-05-23	91931	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
7	3 Healthy Places Index Score
8	3 Healthy Places Index Score
9	4 Healthy Places Index Score
10	2 Healthy Places Index Score
11	2 Healthy Places Index Score
12	2 Healthy Places Index Score
13	1 Healthy Places Index Score
14	3 Healthy Places Index Score
15	1 CDPH-Derived ZCTA Score
16	4 Healthy Places Index Score
17	3 Healthy Places Index Score
18	2 Healthy Places Index Score
19	1 Healthy Places Index Score
20	3 Healthy Places Index Score
21	3 Healthy Places Index Score
22	4 CDPH-Derived ZCTA Score
23	1 Healthy Places Index Score
24	2 Healthy Places Index Score
25	3 Healthy Places Index Score
26	4 Healthy Places Index Score
27	2 Healthy Places Index Score
28	2 Healthy Places Index Score
29	2 Healthy Places Index Score
30	2 Healthy Places Index Score
31	4 Healthy Places Index Score
32	4 Healthy Places Index Score
33	2 Healthy Places Index Score
34	1 CDPH-Derived ZCTA Score
35	4 CDPH-Derived ZCTA Score
36	4 Healthy Places Index Score
37	1 Healthy Places Index Score
38	3 Healthy Places Index Score
39	3 Healthy Places Index Score
40	3 CDPH-Derived ZCTA Score
41	3 Healthy Places Index Score
42	2 Healthy Places Index Score
43	4 Healthy Places Index Score
44	3 Healthy Places Index Score
45	2 Healthy Places Index Score

46	4 Healthy Places Index Score
47	4 Healthy Places Index Score
48	2 Healthy Places Index Score
49	2 Healthy Places Index Score
50	2 Healthy Places Index Score
51	3 Healthy Places Index Score
52	1 Healthy Places Index Score
53	3 Healthy Places Index Score
54	3 Healthy Places Index Score
55	4 Healthy Places Index Score
56	4 Healthy Places Index Score
57	4 Healthy Places Index Score
58	NA No VEM Assigned
59	2 CDPH-Derived ZCTA Score
60	3 Healthy Places Index Score
61	2 Healthy Places Index Score
62	2 Healthy Places Index Score
63	2 Healthy Places Index Score
64	3 Healthy Places Index Score
65	2 Healthy Places Index Score
66	4 Healthy Places Index Score
67	2 Healthy Places Index Score
68	2 CDPH-Derived ZCTA Score
69	3 Healthy Places Index Score
70	2 Healthy Places Index Score
71	3 Healthy Places Index Score
72	3 Healthy Places Index Score
73	4 Healthy Places Index Score
74	4 Healthy Places Index Score
75	4 Healthy Places Index Score
76	4 Healthy Places Index Score
77	NA No VEM Assigned
78	2 Healthy Places Index Score
79	1 Healthy Places Index Score
80	3 CDPH-Derived ZCTA Score
81	4 Healthy Places Index Score
82	4 Healthy Places Index Score
83	2 Healthy Places Index Score
84	3 Healthy Places Index Score
85	NA No VEM Assigned
86	4 Healthy Places Index Score
87	2 Healthy Places Index Score
88	1 Healthy Places Index Score

89	NA	No VEM Assigned	
90	4	Healthy Places Index Score	
91	NA	No VEM Assigned	
92	NA	No VEM Assigned	
93	4	Healthy Places Index Score	
94	NA	No VEM Assigned	
95	NA	No VEM Assigned	
96	3	Healthy Places Index Score	
97	2	Healthy Places Index Score	
98	3	Healthy Places Index Score	
99	4	Healthy Places Index Score	
100	3	Healthy Places Index Score	
101	2	Healthy Places Index Score	
102	3	Healthy Places Index Score	
103	3	Healthy Places Index Score	
104	3	Healthy Places Index Score	
105	1	Healthy Places Index Score	
106	1	CDPH-Derived ZCTA Score	
107	3	CDPH-Derived ZCTA 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
7	16113.9	17551	18705
8	16904.2	18441	19382
9	28789.5	32291	34727
10	1981.1	2336	2660
11	1621.4	1812	1939
12	42613.9	46283	50321
13	54341.2	59361	62859
14	7390.0	8101	8550
15	330.7	323	323
16	14156.9	16302	17379
17	1758.7	2020	2106
18	8644.9	9663	10506
19	3594.7	3982	4428
20	43514.7	50461	54114
21	34685.9	37483	40151
22	130.0	130	130
23	1395.4	1451	1472

24	64013.6	70086	74855
25	40343.9	42839	45435
26	11136.3	12177	12752
27	32246.5	36283	39509
28	42677.7	47784	51619
29	59050.7	64945	68851
30	56152.4	60409	64343
31	26372.9	28414	30550
32	44091.1	45951	48071
33	76365.2	82971	88979
34	589.5	685	693
35	1238.3	1303	1313
36	32146.4	33213	34700
37	47799.7	53883	58408
38	30255.7	31673	33408
39	19835.0	21470	22548
40	166.0	166	166
41	24104.7	25958	27151
42	2333.9	2496	2593
43	17253.3	19025	20155
44	20384.0	21182	22280
45	2151.8	2186	2299
46	10061.5	10704	11417
47	33675.6	36144	38168
48	49284.5	54991	59743
49	43598.3	49162	52985
50	47422.5	52758	56788
51	32025.6	35208	37505
52	61097.9	68711	73623
53	28321.0	29863	31223
54	43222.5	44953	47111
55	4134.7	4387	4729
56	44465.5	48329	51357
57	26688.6	30884	33485
58	0.0	0	NA
59	1010.3	1089	1182
60	27354.6	29757	31918
61	71642.8	79225	84026
62	53851.0	59911	64750
63	59134.5	65415	70148
64	4475.4	4803	5160
65	1200.0	1371	1464
66	44405.4	48477	51381

67	41252.1	44782	48173
68	682.4	743	786
69	49137.8	53795	57710
70	39588.5	40077	41159
71	27003.5	28597	30108
72	44075.0	48160	50693
73	21444.8	23472	24831
74	38942.3	46080	49935
75	46449.1	51493	54762
76	6973.9	7480	7724
77	0.0	0	NA
78	25558.0	27632	29596
79	34956.0	39695	46029
80	11548.0	11654	11765
81	20503.6	23247	24144
82	13762.3	14939	16346
83	22820.5	25486	27236
84	37439.4	40464	43272
85	456.0	456	NA
86	42177.1	46855	49805
87	41447.3	46850	50376
88	1460.5	1492	1543
89	518.0	518	NA
90	71820.2	77775	82658
91	1603.5	1821	NA
92	3747.7	3737	NA
93	46300.3	53102	56134
94	635.0	635	NA
95	285.0	285	NA
96	45552.2	49110	52337
97	35176.1	39270	41807
98	39405.0	42833	46306
99	16620.7	18026	18896
100	15549.8	16905	18162
101	30679.9	33923	36105
102	28353.3	30426	32473
103	50041.6	53839	56983
104	41789.5	47476	50510
105	25332.5	28487	31000
106	826.1	939	986
107	475.7	586	613
persons_fully_vaccinated persons_partially_vaccinated			
1	18553	2530	



2	18902	2508
3	11564	1132
4	34280	2815
5	38168	4091
6	34082	3761
7	13370	1125
8	15490	1272
9	30863	2769
10	2142	169
11	1024	73
12	35955	3051
13	47979	6984
14	5856	458
15	423	60
16	16182	2518
17	1028	63
18	7034	719
19	2190	234
20	45329	6529
21	29082	2993
22	NA	NA
23	837	91
24	74184	16454
25	34868	4001
26	10306	1065
27	26272	2570
28	32963	3040
29	50721	5403
30	41108	5862
31	23255	2402
32	37858	5617
33	87724	23997
34	180	20
35	1276	144
36	47993	4840
37	39259	5057
38	24632	2621
39	16574	4307
40	177	17
41	20115	2198
42	1911	200
43	14421	2032
44	20181	23456

45	2584	274
46	8413	768
47	35484	5531
48	41220	5015
49	35490	3777
50	38881	3439
51	21116	1950
52	49846	5700
53	20197	1982
54	32736	4278
55	5346	876
56	40035	3439
57	29120	6477
58	1976	681
59	1205	176
60	23706	2126
61	83845	18395
62	42309	4470
63	43964	4547
64	3696	375
65	983	68
66	39247	3722
67	32318	3288
68	699	60
69	37280	3572
70	37500	13865
71	20965	8770
72	35089	3791
73	18605	1439
74	41418	3648
75	43675	3623
76	8901	935
77	NA	NA
78	20040	1770
79	17565	3026
80	105	91
81	18372	1944
82	13451	1221
83	19623	2117
84	29164	2471
85	34	20
86	37157	2897
87	35096	2985

88	769	82
89	37	11
90	60928	5713
91	140	74
92	26	21
93	52722	6288
94	24	11
95	34	15
96	37647	3567
97	25209	2647
98	25581	2190
99	14968	1808
100	9818	778
101	26603	2788
102	29794	3786
103	39999	3585
104	36785	3024
105	56523	42259
106	1040	169
107	221	20

percent\_of\_population\_fully\_vaccinated

1	0.700325
2	0.579816
3	0.852300
4	0.735433
5	0.631774
6	0.774415
7	0.714782
8	0.799195
9	0.888732
10	0.805263
11	0.528107
12	0.714513
13	0.763280
14	0.684912
15	1.000000
16	0.931124
17	0.488129
18	0.669522
19	0.494580
20	0.837658
21	0.724316
22	NA

23	0.568614
24	0.991036
25	0.767426
26	0.808187
27	0.664962
28	0.638583
29	0.736678
30	0.638888
31	0.761211
32	0.787543
33	0.985896
34	0.259740
35	0.971820
36	1.000000
37	0.672151
38	0.737308
39	0.735054
40	1.000000
41	0.740857
42	0.736984
43	0.715505
44	0.905790
45	1.000000
46	0.736884
47	0.929679
48	0.689955
49	0.669812
50	0.684669
51	0.563018
52	0.677044
53	0.646863
54	0.694870
55	1.000000
56	0.779543
57	0.869643
58	1.000000
59	1.000000
60	0.742716
61	0.997846
62	0.653421
63	0.626732
64	0.716279
65	0.671448

66	0.763843
67	0.670874
68	0.889313
69	0.645989
70	0.911101
71	0.696327
72	0.692186
73	0.749265
74	0.829438
75	0.797542
76	1.000000
77	NA
78	0.677119
79	0.381607
80	0.008925
81	0.760934
82	0.822892
83	0.720480
84	0.673969
85	NA
86	0.746050
87	0.696681
88	0.498380
89	NA
90	0.737110
91	NA
92	NA
93	0.939217
94	NA
95	NA
96	0.719319
97	0.602985
98	0.552434
99	0.792125
100	0.540579
101	0.736823
102	0.917501
103	0.701946
104	0.728272
105	1.000000
106	1.000000
107	0.360522

percent\_of\_population\_partially\_vaccinated

1	0.095501
2	0.076933
3	0.083432
4	0.060392
5	0.067716
6	0.085458
7	0.060144
8	0.065628
9	0.079736
10	0.063534
11	0.037648
12	0.060631
13	0.111106
14	0.053567
15	0.185759
16	0.144888
17	0.029915
18	0.068437
19	0.052846
20	0.120653
21	0.074544
22	NA
23	0.061821
24	0.219812
25	0.088060
26	0.083516
27	0.065048
28	0.058893
29	0.078474
30	0.091105
31	0.078625
32	0.116848
33	0.269693
34	0.028860
35	0.109673
36	0.139481
37	0.086581
38	0.078454
39	0.191015
40	0.102410
41	0.080955
42	0.077131
43	0.100819

44	1.000000
45	0.119182
46	0.067268
47	0.144912
48	0.083943
49	0.071284
50	0.060559
51	0.051993
52	0.077421
53	0.063479
54	0.090807
55	0.185240
56	0.066963
57	0.193430
58	1.000000
59	0.148900
60	0.066608
61	0.218920
62	0.069035
63	0.064820
64	0.072674
65	0.046448
66	0.072439
67	0.068254
68	0.076336
69	0.061896
70	0.336864
71	0.291285
72	0.074784
73	0.057952
74	0.073055
75	0.066159
76	0.121051
77	NA
78	0.059805
79	0.065741
80	0.007735
81	0.080517
82	0.074697
83	0.077728
84	0.057104
85	NA
86	0.058167

87	0.059254	
88	0.053143	
89	NA	
90	0.069116	
91	NA	
92	NA	
93	0.112018	
94	NA	
95	NA	
96	0.068154	
97	0.063315	
98	0.047294	
99	0.095682	
100	0.042837	
101	0.077219	
102	0.116589	
103	0.062914	
104	0.059869	
105	1.000000	
106	0.171400	
107	0.032626	
	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
7	0.774926	7695
8	0.864823	10007
9	0.968468	22538
10	0.868797	1224
11	0.565755	599
12	0.775144	21493
13	0.874386	27475
14	0.738479	3374
15	1.000000	253
16	1.000000	10554
17	0.518044	596
18	0.737959	4010
19	0.547426	1133
20	0.958311	28745
21	0.798860	18553



22	NA	NA
23	0.630435	456
24	1.000000	43433
25	0.855486	23171
26	0.891703	7065
27	0.730010	13579
28	0.697476	18126
29	0.815152	30439
30	0.729993	25450
31	0.839836	15892
32	0.904391	27110
33	1.000000	48492
34	0.288600	95
35	1.000000	905
36	1.000000	22739
37	0.758732	19845
38	0.815762	17255
39	0.926069	10015
40	1.000000	128
41	0.821812	12674
42	0.814115	1176
43	0.816324	9815
44	1.000000	12201
45	1.000000	1663
46	0.804152	5733
47	1.000000	25011
48	0.773898	22000
49	0.741096	19224
50	0.745228	21533
51	0.615011	11756
52	0.754465	28033
53	0.710342	13077
54	0.785677	19740
55	1.000000	3908
56	0.846506	28994
57	1.000000	17960
58	NA	872
59	1.000000	593
60	0.809324	15505
61	1.000000	47021
62	0.722456	23970
63	0.691552	22706
64	0.788953	2297

65	0.717896	495
66	0.836282	26374
67	0.739128	17902
68	0.965649	424
69	0.707885	22312
70	1.000000	22302
71	0.987612	12674
72	0.766970	22324
73	0.807217	12753
74	0.902493	28710
75	0.863701	30810
76	1.000000	6211
77	NA	NA
78	0.736924	12372
79	0.447348	9144
80	0.016660	NA
81	0.841451	12589
82	0.897589	8801
83	0.798208	11142
84	0.731073	16136
85	NA	NA
86	0.804217	24192
87	0.755935	20529
88	0.551523	465
89	NA	NA
90	0.806226	41692
91	NA	30
92	NA	NA
93	1.000000	36938
94	NA	NA
95	NA	NA
96	0.787473	23717
97	0.666300	14588
98	0.599728	14270
99	0.887807	9673
100	0.583416	5563
101	0.814042	16672
102	1.000000	19685
103	0.764860	25989
104	0.788141	23723
105	1.000000	27396
106	1.000000	522
107	0.393148	130

	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
7	2943	13343
8	4229	15426
9	10898	30701
10	471	2137
11	226	1024
12	8118	35871
13	8785	47862
14	1281	5851
15	87	423
16	3969	16138
17	234	1027
18	1348	7020
19	369	2188
20	9967	45108
21	7778	28968
22	NA	49
23	177	836
24	14778	73947
25	10693	34754
26	3577	10276
27	4499	26195
28	6816	32871
29	10281	50586
30	10029	40957
31	7442	23144
32	12846	37658
33	14269	87437
34	43	180
35	459	1273
36	11313	47936
37	5990	39165
38	8496	24539
39	4413	16540
40	54	177
41	5499	20059
42	549	1907

43	4581	14366
44	4956	20090
45	700	2583
46	2738	8387
47	12594	35331
48	7517	41105
49	6737	35412
50	7440	38796
51	4731	21092
52	9510	49691
53	5846	20124
54	8307	32632
55	2072	5321
56	13538	39835
57	6072	28966
58	216	1968
59	152	1205
60	6753	23631
61	14578	83599
62	8181	42195
63	7262	43874
64	1073	3685
65	165	982
66	12808	39145
67	6882	32246
68	198	698
69	8803	37166
70	9589	37432
71	5615	20881
72	9289	34933
73	6044	18515
74	12696	41154
75	14053	43437
76	2895	8882
77	NA	28
78	5108	19955
79	3353	17492
80	NA	105
81	5990	18310
82	4131	13399
83	3995	19583
84	5669	29112
85	NA	34

86	10731	36988
87	7303	34978
88	197	769
89	NA	37
90	16820	60618
91	14	140
92	NA	26
93	17449	52480
94	NA	24
95	NA	34
96	10347	37520
97	5734	25159
98	4926	25541
99	3922	14919
100	2067	9800
101	5690	26535
102	7865	29633
103	11878	39801
104	9948	36630
105	7428	56367
106	140	1035
107	51	221

eligible\_bivalent\_recipient\_count

1	18507
2	18737
3	11544
4	34163
5	38055
6	33981
7	13343
8	15426
9	30701
10	0
11	0
12	35871
13	47862
14	0
15	0
16	0
17	0
18	0
19	2188
20	45108

21	28968
22	0
23	0
24	73947
25	34754
26	10276
27	26195
28	32871
29	50586
30	40957
31	23144
32	37658
33	87437
34	0
35	1273
36	47936
37	39165
38	24539
39	16540
40	0
41	20059
42	1907
43	14366
44	20090
45	2583
46	8387
47	35331
48	41105
49	35412
50	38796
51	21092
52	49691
53	20124
54	32632
55	5321
56	39835
57	28966
58	0
59	0
60	23631
61	83599
62	42195
63	43874

64	0
65	0
66	39145
67	32246
68	0
69	37166
70	37432
71	20881
72	34933
73	18515
74	41154
75	43437
76	8882
77	0
78	19955
79	17492
80	0
81	18310
82	13399
83	19583
84	29112
85	0
86	36988
87	34978
88	769
89	0
90	60618
91	0
92	0
93	52480
94	0
95	0
96	37520
97	25159
98	25541
99	14919
100	9800
101	26535
102	29633
103	39801
104	36630
105	56367
106	0

107

0

redacted

1	No
2	No
3	No
4	No
5	No
6	No
7	No
8	No
9	No
10	No
11	No
12	No
13	No
14	No
15	No
16	No
17	No
18	No
19	No
20	No
21	No
22	Information redacted in accordance with CA state privacy requirements
23	No
24	No
25	No
26	No
27	No
28	No
29	No
30	No
31	No
32	No
33	No
34	No
35	No
36	No
37	No
38	No
39	No
40	No
41	No



42	No
43	No
44	No
45	No
46	No
47	No
48	No
49	No
50	No
51	No
52	No
53	No
54	No
55	No
56	No
57	No
58	No
59	No
60	No
61	No
62	No
63	No
64	No
65	No
66	No
67	No
68	No
69	No
70	No
71	No
72	No
73	No
74	No
75	No
76	No
77	Information redacted in accordance with CA state privacy requirements
78	No
79	No
80	Information redacted in accordance with CA state privacy requirements
81	No
82	No
83	No
84	No

```

85 Information redacted in accordance with CA state privacy requirements
86                                                                 No
87                                                                 No
88                                                                 No
89 Information redacted in accordance with CA state privacy requirements
90                                                                 No
91                                                                 No
92 Information redacted in accordance with CA state privacy requirements
93                                                                 No
94 Information redacted in accordance with CA state privacy requirements
95 Information redacted in accordance with CA state privacy requirements
96                                                                 No
97                                                                 No
98                                                                 No
99                                                                 No
100                                                                No
101                                                                No
102                                                                No
103                                                                No
104                                                                No
105                                                                No
106                                                                No
107                                                                No

```

```
mean(sd.23$percent_of_population_fully_vaccinated, na.rm = T)
```

```
[1] 0.7419654
```

- **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"?

```

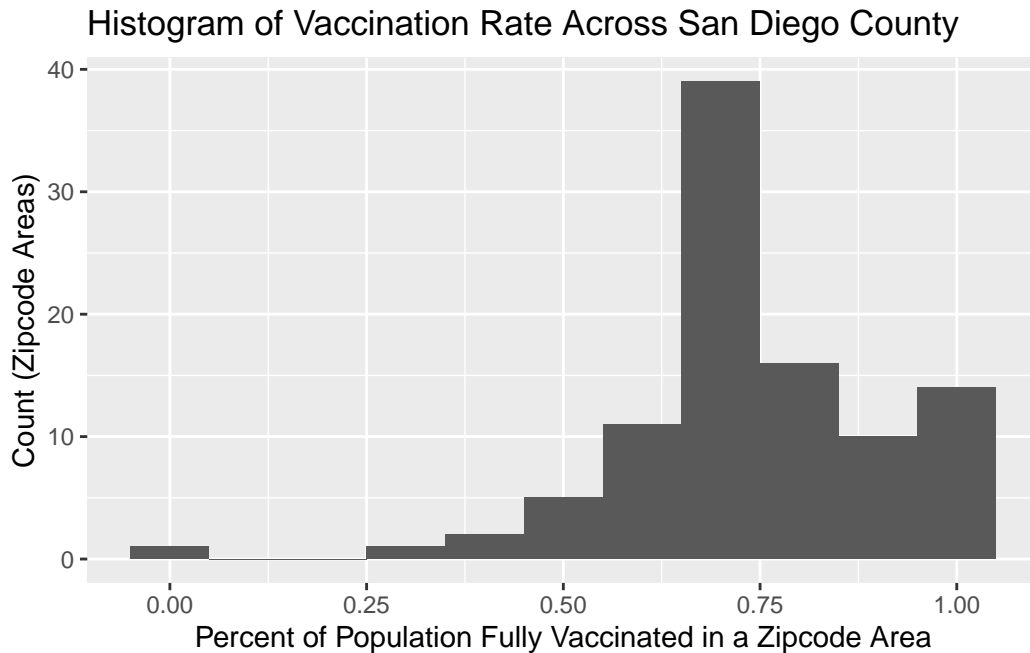
library(ggplot2)

p <- ggplot(sd.23) +
  aes(percent_of_population_fully_vaccinated) +
  geom_histogram(binwidth = 0.1)

p + ggtitle("Histogram of Vaccination Rate Across San Diego County") +
  xlab("Percent of Population Fully Vaccinated in a Zipcode Area") + ylab("Count (Zipcode

```

Warning: Removed 8 rows containing non-finite values (`stat\_bin()`).



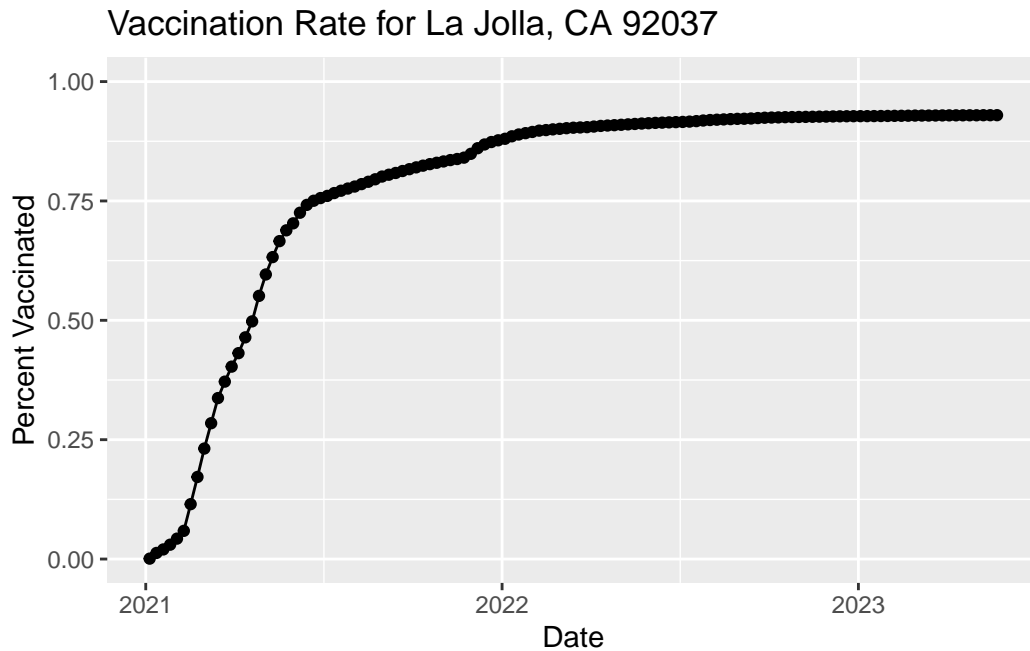
Focusing on UCSD:

```
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:

```
p <- ggplot(ucsd) +
  aes(as_of_date,
      percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x = "Date", y="Percent Vaccinated")
p + ggtitle("Vaccination Rate for La Jolla, CA 92037")
```



Comparing similar sized areas:

```
# Subset to all CA areas with a population as large as 92037
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		
2		
3		
4		
5		
6		

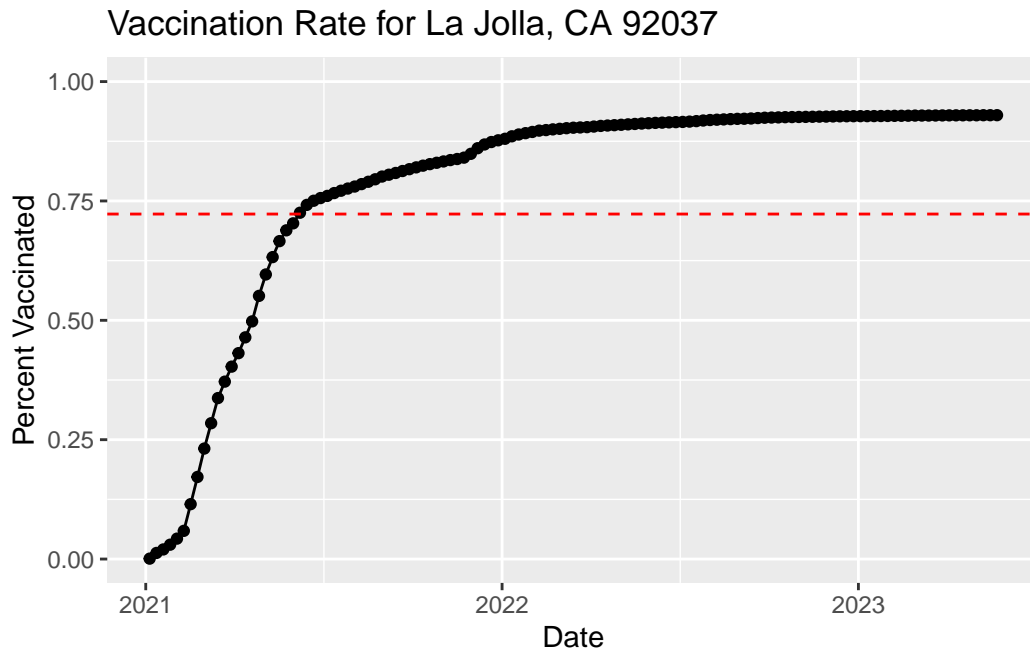
1	33740	No
2	35587	No
3	36977	No
4	44050	No
5	40460	No
6	40104	No

- **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?

```
vax.36_mean <- mean(vax.36$percent_of_population_fully_vaccinated, na.rm = T)
vax.36_mean
```

```
[1] 0.7225892
```

```
p <- ggplot(ucsd) +
  aes(as_of_date,
      percent_of_population_fully_vaccinated) +
  geom_point() +
  geom_line(group=1) +
  ylim(c(0,1)) +
  labs(x = "Date", y="Percent Vaccinated", title = "Vaccination Rate for La Jolla, CA 92037")
  geom_hline(yintercept = vax.36_mean, linetype="dashed", col = "red")
p
```



- **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$persons_fully_vaccinated)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
17565	32272	37303	40126	45420	87724

- **Q18.** Using ggplot generate a histogram of this data.

```
h <- ggplot(vax.36) +
  aes(percent_of_population_fully_vaccinated) +
  geom_histogram() +
  xlim(c(0,1))
labs(x = "Percent Vaccinated", y="Count")
```

```
$x
[1] "Percent Vaccinated"
```

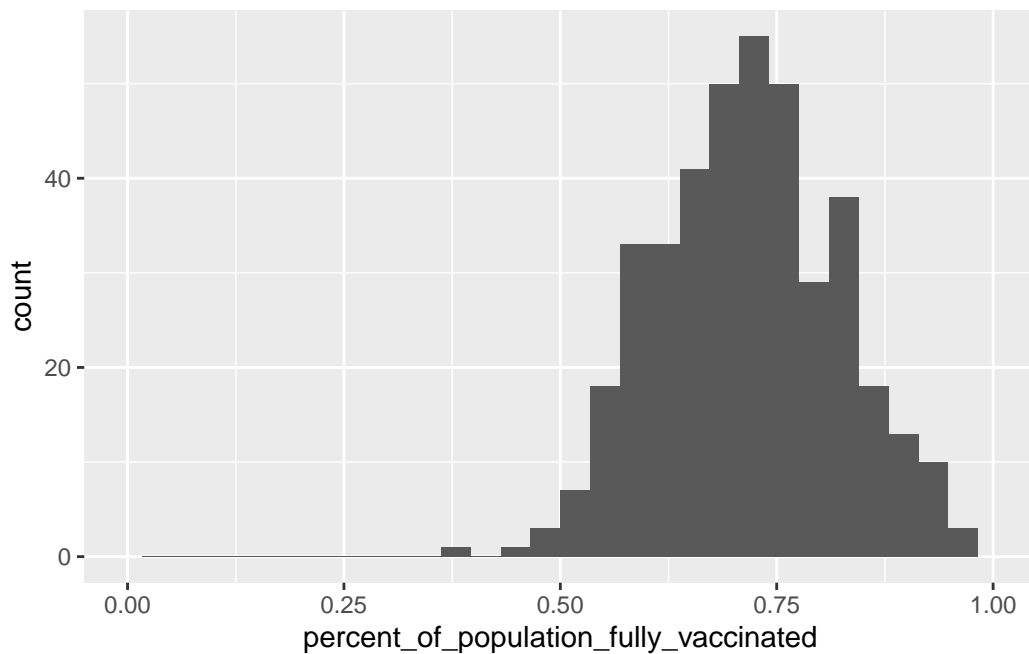
```
$y
[1] "Count"
```

```
attr("class")
[1] "labels"
```

```
h
```

```
`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
```

- **Q20.** Finally make a time course plot of vaccination progress for all areas in the full dataset with a `age5_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="blue") +
  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 = vax.36_mean, linetype="dashed")

```

Warning: Removed 185 rows containing missing values (`geom\_line()`).

