

Covid-19 Vaccination Rates Mini-Project

Katie Chau

3/5/2022

```
# 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           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. What column details the total number of people fully vaccinated?

persons_fully_vaccinated

Q2. What column details the Zip code tabulation area?

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?

```
nrow(vax)
```

```
## [1] 107604
```

The number of rows in the mini-project site was 105840, so to get the right answer on the website, I used that row's date and not the date that I got when I downloaded the csv file with more information.

```
vax$as_of_date[105840]
```

```
## [1] "2022-02-22"
```

```
vax$as_of_date[107604]
```

```
## [1] "2022-03-01"
```

2022-03-01

```
#Overview of dataset
#install.packages("skimr")
skimr::skim(vax)
```

Table 1: 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	93665.111817.39	90001	92257.7593658.5095380.5097635.0					
vaccine_equity_metric_quarter	1807	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	
age12_plus_population	0	1.00	18895.0418993.91	0	1346.95	13685.1031756.1288556.7				
age5_plus_population	0	1.00	20875.2421106.02	0	1460.50	15364.0034877.00101902.0				
persons_fully_vaccinated	18338	0.83	12155.6113063.88	11	1066.25	7374.50	20005.0077744.0			
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_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	

Q5. How many numeric columns are in this dataset?

10 (on the website it has 9 as the correct answer but for my data as well as the data retrieved from skim(vax) on the website it says 10)

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

18338 (18174 on the website)

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

```
## [1] 18338
```

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

```
18174/77213*100
```

```
## [1] 23.53749
```

I tried this calculation for the data on the website but I couldn't get the correct number to show green.

```
#The same equation but with my data set.  
18338/77744*100
```

```
## [1] 23.58767
```

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

This data may be missing because this many people in the population aren't vaccinated. This could be due to being too young to receive a vaccination since a person must be 5 years or older.

#Working with dates

```
#install.packages("lubridate")
```

```
library(lubridate)
```

```
##
```

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

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

```
##
```

```
## date, intersect, setdiff, union
```

```
today()
```

```
## [1] "2022-03-08"
```

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

```
#specify using year-month-day format  
vax$as_of_date <- ymd(vax$as_of_date)
```

```
#how many days since first vaccination reported in dataset  
today() - vax$as_of_date[1]
```

```
## Time difference of 427 days
```

```
#days in dataset span  
vax$as_of_date[nrow(vax)] - vax$as_of_date[1]
```

```
## Time difference of 420 days
```

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

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

```
## Time difference of -7 days
```

6 days have passed since the last update of the dataset.

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

```
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
```

```
n_distinct(vax$as_of_date)
```

```
## [1] 61
```

61 unique dates are detailed in my data set. For the website, the answer is 60 for that dataset.

Working with ZIP codes

```
#library(zipcodeR)
```

```
#install.packages("sf")
```

```
#install.packages("gdal-config")
```

Cannot load zipcodeR package.Kept getting error "there is no package called 'sf'. Couldn't install package 'sf' due to gdal-config not found. gdal-config would not install for this version of R.

```
#distance between centroids of any two zip codes in miles
```

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

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

```
#census data on zip code areas
```

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

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

Focus on the San Diego Area

```
library(dplyr)

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

nrow(sd)
```

```
## [1] 6527
```

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

```
distinctZip <- filter(vax, county == "San Diego")
uniqueZip <- unique((distinctZip))
nrow(uniqueZip)
```

```
## [1] 6527
```

107 distinct zip codes in San Diego. Not sure why my output does not show the correct info with the unique function.

```
which.max(distinctZip$age12_plus_population)
```

```
## [1] 91
```

```
distinctZip[91,]
```

```
##   as_of_date zip_code_tabulation_area local_health_jurisdiction   county
## 91 2021-01-05           92154                San Diego San Diego
##   vaccine_equity_metric_quartile                vem_source
## 91                2 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 91                76365.2                82971                18
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 91                22                0.000217
##   percent_of_population_partially_vaccinated
## 91                0.000265
##   percent_of_population_with_1_plus_dose booster_recip_count
## 91                0.000482                NA
##
## 91 Information redacted in accordance with CA state privacy requirements
```

92154 is the zip code with highest age12 plus population.

Q13. What is the overall average “Percent of Population Fully Vaccinated” value for all San Diego “County” as of “2022-02-22”?

```
pfv <- filter(vax, county == "San Diego", as_of_date == "2022-02-22")
pfv
```

##	as_of_date	zip_code_tabulation_area	local_health_jurisdiction	county
## 1	2022-02-22	92064	San Diego	San Diego
## 2	2022-02-22	92103	San Diego	San Diego
## 3	2022-02-22	92118	San Diego	San Diego
## 4	2022-02-22	92083	San Diego	San Diego
## 5	2022-02-22	92056	San Diego	San Diego
## 6	2022-02-22	92069	San Diego	San Diego
## 7	2022-02-22	92066	San Diego	San Diego
## 8	2022-02-22	92060	San Diego	San Diego
## 9	2022-02-22	91917	San Diego	San Diego
## 10	2022-02-22	91977	San Diego	San Diego
## 11	2022-02-22	92130	San Diego	San Diego
## 12	2022-02-22	92086	San Diego	San Diego
## 13	2022-02-22	92113	San Diego	San Diego
## 14	2022-02-22	92104	San Diego	San Diego
## 15	2022-02-22	91931	San Diego	San Diego
## 16	2022-02-22	91945	San Diego	San Diego
## 17	2022-02-22	92116	San Diego	San Diego
## 18	2022-02-22	92054	San Diego	San Diego
## 19	2022-02-22	92040	San Diego	San Diego
## 20	2022-02-22	92091	San Diego	San Diego
## 21	2022-02-22	92126	San Diego	San Diego
## 22	2022-02-22	92084	San Diego	San Diego
## 23	2022-02-22	92025	San Diego	San Diego
## 24	2022-02-22	92004	San Diego	San Diego
## 25	2022-02-22	91915	San Diego	San Diego
## 26	2022-02-22	92036	San Diego	San Diego
## 27	2022-02-22	92106	San Diego	San Diego
## 28	2022-02-22	92037	San Diego	San Diego
## 29	2022-02-22	92075	San Diego	San Diego
## 30	2022-02-22	92020	San Diego	San Diego
## 31	2022-02-22	92007	San Diego	San Diego
## 32	2022-02-22	92065	San Diego	San Diego
## 33	2022-02-22	92070	San Diego	San Diego
## 34	2022-02-22	92108	San Diego	San Diego
## 35	2022-02-22	92008	San Diego	San Diego
## 36	2022-02-22	92027	San Diego	San Diego
## 37	2022-02-22	92101	San Diego	San Diego
## 38	2022-02-22	92107	San Diego	San Diego
## 39	2022-02-22	92128	San Diego	San Diego
## 40	2022-02-22	92109	San Diego	San Diego
## 41	2022-02-22	92028	San Diego	San Diego
## 42	2022-02-22	92024	San Diego	San Diego
## 43	2022-02-22	92003	San Diego	San Diego
## 44	2022-02-22	92129	San Diego	San Diego
## 45	2022-02-22	92110	San Diego	San Diego
## 46	2022-02-22	92071	San Diego	San Diego
## 47	2022-02-22	92021	San Diego	San Diego

## 48	2022-02-22	92119	San Diego	San Diego
## 49	2022-02-22	92059	San Diego	San Diego
## 50	2022-02-22	92105	San Diego	San Diego
## 51	2022-02-22	92111	San Diego	San Diego
## 52	2022-02-22	92121	San Diego	San Diego
## 53	2022-02-22	92127	San Diego	San Diego
## 54	2022-02-22	91914	San Diego	San Diego
## 55	2022-02-22	91910	San Diego	San Diego
## 56	2022-02-22	92055	San Diego	San Diego
## 57	2022-02-22	92081	San Diego	San Diego
## 58	2022-02-22	91978	San Diego	San Diego
## 59	2022-02-22	92078	San Diego	San Diego
## 60	2022-02-22	92019	San Diego	San Diego
## 61	2022-02-22	92134	San Diego	San Diego
## 62	2022-02-22	92139	San Diego	San Diego
## 63	2022-02-22	92155	San Diego	San Diego
## 64	2022-02-22	92135	San Diego	San Diego
## 65	2022-02-22	92145	San Diego	San Diego
## 66	2022-02-22	91948	San Diego	San Diego
## 67	2022-02-22	91941	San Diego	San Diego
## 68	2022-02-22	91963	San Diego	San Diego
## 69	2022-02-22	92173	San Diego	San Diego
## 70	2022-02-22	91962	San Diego	San Diego
## 71	2022-02-22	91916	San Diego	San Diego
## 72	2022-02-22	91913	San Diego	San Diego
## 73	2022-02-22	92011	San Diego	San Diego
## 74	2022-02-22	92140	San Diego	San Diego
## 75	2022-02-22	92117	San Diego	San Diego
## 76	2022-02-22	92114	San Diego	San Diego
## 77	2022-02-22	92122	San Diego	San Diego
## 78	2022-02-22	92147	San Diego	San Diego
## 79	2022-02-22	92010	San Diego	San Diego
## 80	2022-02-22	92132	San Diego	San Diego
## 81	2022-02-22	92154	San Diego	San Diego
## 82	2022-02-22	91902	San Diego	San Diego
## 83	2022-02-22	91901	San Diego	San Diego
## 84	2022-02-22	92058	San Diego	San Diego
## 85	2022-02-22	92123	San Diego	San Diego
## 86	2022-02-22	92115	San Diego	San Diego
## 87	2022-02-22	92014	San Diego	San Diego
## 88	2022-02-22	92057	San Diego	San Diego
## 89	2022-02-22	91911	San Diego	San Diego
## 90	2022-02-22	92026	San Diego	San Diego
## 91	2022-02-22	91935	San Diego	San Diego
## 92	2022-02-22	92009	San Diego	San Diego
## 93	2022-02-22	92061	San Diego	San Diego
## 94	2022-02-22	92131	San Diego	San Diego
## 95	2022-02-22	92029	San Diego	San Diego
## 96	2022-02-22	92124	San Diego	San Diego
## 97	2022-02-22	92120	San Diego	San Diego
## 98	2022-02-22	91950	San Diego	San Diego
## 99	2022-02-22	91905	San Diego	San Diego
## 100	2022-02-22	91934	San Diego	San Diego
## 101	2022-02-22	92102	San Diego	San Diego

##	102	2022-02-22	92082	San Diego	San Diego
##	103	2022-02-22	92067	San Diego	San Diego
##	104	2022-02-22	91906	San Diego	San Diego
##	105	2022-02-22	91932	San Diego	San Diego
##	106	2022-02-22	91980	San Diego	San Diego
##	107	2022-02-22	91942	San Diego	San Diego
##		vaccine_equity_metric_quartile		vem_source	
##	1		4	Healthy Places Index Score	
##	2		4	Healthy Places Index Score	
##	3		3	Healthy Places Index Score	
##	4		2	Healthy Places Index Score	
##	5		3	Healthy Places Index Score	
##	6		2	Healthy Places Index Score	
##	7		1	CDPH-Derived ZCTA Score	
##	8		3	CDPH-Derived ZCTA Score	
##	9		1	CDPH-Derived ZCTA Score	
##	10		2	Healthy Places Index Score	
##	11		4	Healthy Places Index Score	
##	12		1	Healthy Places Index Score	
##	13		1	Healthy Places Index Score	
##	14		3	Healthy Places Index Score	
##	15		3	CDPH-Derived ZCTA Score	
##	16		2	Healthy Places Index Score	
##	17		3	Healthy Places Index Score	
##	18		2	Healthy Places Index Score	
##	19		3	Healthy Places Index Score	
##	20		4	CDPH-Derived ZCTA Score	
##	21		4	Healthy Places Index Score	
##	22		2	Healthy Places Index Score	
##	23		2	Healthy Places Index Score	
##	24		2	Healthy Places Index Score	
##	25		4	Healthy Places Index Score	
##	26		2	Healthy Places Index Score	
##	27		4	Healthy Places Index Score	
##	28		4	Healthy Places Index Score	
##	29		4	Healthy Places Index Score	
##	30		2	Healthy Places Index Score	
##	31		4	Healthy Places Index Score	
##	32		3	Healthy Places Index Score	
##	33		2	CDPH-Derived ZCTA Score	
##	34		3	Healthy Places Index Score	
##	35		3	Healthy Places Index Score	
##	36		2	Healthy Places Index Score	
##	37		2	Healthy Places Index Score	
##	38		3	Healthy Places Index Score	
##	39		4	Healthy Places Index Score	
##	40		3	Healthy Places Index Score	
##	41		2	Healthy Places Index Score	
##	42		4	Healthy Places Index Score	
##	43		3	Healthy Places Index Score	
##	44		4	Healthy Places Index Score	
##	45		3	Healthy Places Index Score	
##	46		3	Healthy Places Index Score	
##	47		2	Healthy Places Index Score	

## 48	4 Healthy Places Index Score
## 49	2 Healthy Places Index Score
## 50	1 Healthy Places Index Score
## 51	3 Healthy Places Index Score
## 52	4 Healthy Places Index Score
## 53	4 Healthy Places Index Score
## 54	4 Healthy Places Index Score
## 55	2 Healthy Places Index Score
## 56	3 CDPH-Derived ZCTA Score
## 57	2 Healthy Places Index Score
## 58	2 Healthy Places Index Score
## 59	3 Healthy Places Index Score
## 60	3 Healthy Places Index Score
## 61	NA No VEM Assigned
## 62	2 Healthy Places Index Score
## 63	NA No VEM Assigned
## 64	NA No VEM Assigned
## 65	NA No VEM Assigned
## 66	4 CDPH-Derived ZCTA Score
## 67	3 Healthy Places Index Score
## 68	2 CDPH-Derived ZCTA Score
## 69	1 Healthy Places Index Score
## 70	3 Healthy Places Index Score
## 71	2 Healthy Places Index Score
## 72	3 Healthy Places Index Score
## 73	4 Healthy Places Index Score
## 74	NA No VEM Assigned
## 75	3 Healthy Places Index Score
## 76	2 Healthy Places Index Score
## 77	4 Healthy Places Index Score
## 78	NA No VEM Assigned
## 79	4 Healthy Places Index Score
## 80	NA No VEM Assigned
## 81	2 Healthy Places Index Score
## 82	4 Healthy Places Index Score
## 83	3 Healthy Places Index Score
## 84	1 Healthy Places Index Score
## 85	3 Healthy Places Index Score
## 86	2 Healthy Places Index Score
## 87	4 Healthy Places Index Score
## 88	2 Healthy Places Index Score
## 89	2 Healthy Places Index Score
## 90	2 Healthy Places Index Score
## 91	3 Healthy Places Index Score
## 92	4 Healthy Places Index Score
## 93	2 Healthy Places Index Score
## 94	4 Healthy Places Index Score
## 95	3 Healthy Places Index Score
## 96	3 Healthy Places Index Score
## 97	4 Healthy Places Index Score
## 98	1 Healthy Places Index Score
## 99	1 Healthy Places Index Score
## 100	1 CDPH-Derived ZCTA Score
## 101	1 Healthy Places Index Score

## 102		3 Healthy Places Index Score	
## 103		4 Healthy Places Index Score	
## 104		1 Healthy Places Index Score	
## 105		2 Healthy Places Index Score	
## 106	NA	No VEM Assigned	
## 107		3 Healthy Places Index Score	
##	age12_plus_population	age5_plus_population	persons_fully_vaccinated
## 1	42177.1	46855	34266
## 2	32146.4	33213	46456
## 3	19835.0	21470	14954
## 4	32246.5	36283	24146
## 5	45552.2	49110	34782
## 6	41447.3	46850	32505
## 7	589.5	685	181
## 8	166.0	166	163
## 9	826.1	939	968
## 10	53851.0	59911	38823
## 11	46300.3	53102	50300
## 12	1460.5	1492	731
## 13	47799.7	53883	36354
## 14	40343.9	42839	33667
## 15	475.7	586	216
## 16	22820.5	25486	18125
## 17	30255.7	31673	24093
## 18	35176.1	39270	23944
## 19	39405.0	42833	23614
## 20	1238.3	1303	1213
## 21	71820.2	77775	58503
## 22	42677.7	47784	29940
## 23	43598.3	49162	31104
## 24	2151.8	2186	2623
## 25	26688.6	30884	25915
## 26	2333.9	2496	1851
## 27	17253.3	19025	13842
## 28	33675.6	36144	34859
## 29	11136.3	12177	10056
## 30	49284.5	54991	37783
## 31	10061.5	10704	8044
## 32	32025.6	35208	19451
## 33	682.4	743	637
## 34	20384.0	21182	18118
## 35	24104.7	25958	19152
## 36	47422.5	52758	34494
## 37	39588.5	40077	34339
## 38	28321.0	29863	19756
## 39	44465.5	48329	37924
## 40	43222.5	44953	32503
## 41	41252.1	44782	29341
## 42	44405.4	48477	37429
## 43	4475.4	4803	3118
## 44	46449.1	51493	41622
## 45	27003.5	28597	19998
## 46	49137.8	53795	34481
## 47	59134.5	65415	39893

## 48	21444.8	23472	17840
## 49	1200.0	1371	906
## 50	61097.9	68711	46545
## 51	44075.0	48160	33715
## 52	4134.7	4387	4611
## 53	38942.3	46080	39294
## 54	14156.9	16302	15112
## 55	64013.6	70086	65376
## 56	11548.0	11654	136
## 57	25558.0	27632	18615
## 58	8644.9	9663	6516
## 59	41789.5	47476	33706
## 60	37439.4	40464	27272
## 61	285.0	285	52
## 62	30679.9	33923	25420
## 63	456.0	456	66
## 64	635.0	635	36
## 65	1603.5	1821	156
## 66	130.0	130	49
## 67	27354.6	29757	22437
## 68	1010.3	1089	1070
## 69	25332.5	28487	48629
## 70	1758.7	2020	968
## 71	1621.4	1812	986
## 72	43514.7	50461	41413
## 73	20503.6	23247	16741
## 74	3747.7	3737	38
## 75	50041.6	53839	38537
## 76	59050.7	64945	47012
## 77	44091.1	45951	37493
## 78	518.0	518	NA
## 79	13762.3	14939	12393
## 80	0.0	0	88
## 81	76365.2	82971	77457
## 82	16620.7	18026	13987
## 83	15549.8	16905	9170
## 84	34956.0	39695	15406
## 85	28353.3	30426	23525
## 86	56152.4	60409	40630
## 87	11942.5	13149	11169
## 88	51927.0	56906	35165
## 89	71642.8	79225	73649
## 90	42613.9	46283	32057
## 91	7390.0	8101	5526
## 92	39183.5	43710	32863
## 93	1981.1	2336	1896
## 94	28789.5	32291	29537
## 95	16904.2	18441	14175
## 96	25422.4	29040	17712
## 97	26372.9	28414	22720
## 98	54341.2	59361	44024
## 99	1395.4	1451	764
## 100	330.7	323	390
## 101	37042.3	41033	31663

## 102	16113.9	17551	11738
## 103	6973.9	7480	8193
## 104	3594.7	3982	1935
## 105	21968.2	24874	16797
## 106	0.0	0	1629
## 107	34685.9	37483	27145
##	persons_partially_vaccinated	percent_of_population_fully_vaccinated	
## 1	6861		0.731320
## 2	8434		1.000000
## 3	7405		0.696507
## 4	5924		0.665491
## 5	7362		0.708247
## 6	7043		0.693810
## 7	35		0.264234
## 8	52		0.981928
## 9	355		1.000000
## 10	8749		0.648011
## 11	12642		0.947234
## 12	173		0.489946
## 13	10564		0.674684
## 14	7621		0.785896
## 15	61		0.368601
## 16	3964		0.711175
## 17	5291		0.760679
## 18	5880		0.609728
## 19	4529		0.551304
## 20	341		0.930929
## 21	12369		0.752208
## 22	6929		0.626570
## 23	9759		0.632684
## 24	917		1.000000
## 25	9092		0.839108
## 26	596		0.741587
## 27	3545		0.727569
## 28	10997		0.964448
## 29	2720		0.825819
## 30	9452		0.687076
## 31	1702		0.751495
## 32	3803		0.552460
## 33	147		0.857335
## 34	6986		0.855349
## 35	4467		0.737807
## 36	9105		0.653816
## 37	16002		0.856826
## 38	4080		0.661554
## 39	8448		0.784705
## 40	8066		0.723044
## 41	6744		0.655196
## 42	8743		0.772098
## 43	680		0.649178
## 44	8077		0.808304
## 45	6553		0.699304
## 46	6591		0.640970
## 47	9109		0.609845

## 48	3065	0.760055
## 49	165	0.660832
## 50	11359	0.677402
## 51	7587	0.700062
## 52	1599	1.000000
## 53	8830	0.852734
## 54	4550	0.927003
## 55	24868	0.932797
## 56	81	0.011670
## 57	4035	0.673675
## 58	1413	0.674325
## 59	7107	0.709959
## 60	5454	0.673982
## 61	22	0.182456
## 62	5881	0.749344
## 63	31	0.144737
## 64	18	0.056693
## 65	77	0.085667
## 66	18	0.376923
## 67	4869	0.754007
## 68	419	0.982553
## 69	20743	1.000000
## 70	190	0.479208
## 71	189	0.544150
## 72	11624	0.820693
## 73	3910	0.720136
## 74	26	0.010169
## 75	7448	0.715782
## 76	11102	0.723874
## 77	10131	0.815934
## 78	NA	NA
## 79	2546	0.829574
## 80	33	1.000000
## 81	29331	0.933543
## 82	3761	0.775935
## 83	1684	0.542443
## 84	4363	0.388109
## 85	7498	0.773187
## 86	9991	0.672582
## 87	2960	0.849418
## 88	8334	0.617949
## 89	27682	0.929618
## 90	7855	0.692630
## 91	1041	0.682138
## 92	7025	0.751842
## 93	467	0.811644
## 94	5982	0.914713
## 95	3103	0.768668
## 96	3750	0.609917
## 97	4466	0.799606
## 98	13435	0.741632
## 99	172	0.526533
## 100	111	1.000000
## 101	8237	0.771647

## 102	2501	0.668794
## 103	3056	1.000000
## 104	452	0.485937
## 105	4898	0.675283
## 106	963	1.000000
## 107	6102	0.724195
##	percent_of_population_partially_vaccinated	
## 1	0.146430	
## 2	0.253937	
## 3	0.344900	
## 4	0.163272	
## 5	0.149908	
## 6	0.150331	
## 7	0.051095	
## 8	0.313253	
## 9	0.378062	
## 10	0.146033	
## 11	0.238070	
## 12	0.115952	
## 13	0.196054	
## 14	0.177899	
## 15	0.104096	
## 16	0.155536	
## 17	0.167051	
## 18	0.149733	
## 19	0.105736	
## 20	0.261704	
## 21	0.159036	
## 22	0.145007	
## 23	0.198507	
## 24	0.419488	
## 25	0.294392	
## 26	0.238782	
## 27	0.186334	
## 28	0.304255	
## 29	0.223372	
## 30	0.171883	
## 31	0.159006	
## 32	0.108015	
## 33	0.197847	
## 34	0.329808	
## 35	0.172086	
## 36	0.172580	
## 37	0.399281	
## 38	0.136624	
## 39	0.174802	
## 40	0.179432	
## 41	0.150596	
## 42	0.180354	
## 43	0.141578	
## 44	0.156856	
## 45	0.229150	
## 46	0.122521	
## 47	0.139249	

## 48	0.130581
## 49	0.120350
## 50	0.165316
## 51	0.157537
## 52	0.364486
## 53	0.191623
## 54	0.279107
## 55	0.354821
## 56	0.006950
## 57	0.146026
## 58	0.146228
## 59	0.149697
## 60	0.134786
## 61	0.077193
## 62	0.173363
## 63	0.067982
## 64	0.028346
## 65	0.042284
## 66	0.138462
## 67	0.163625
## 68	0.384757
## 69	0.728157
## 70	0.094059
## 71	0.104305
## 72	0.230356
## 73	0.168194
## 74	0.006957
## 75	0.138338
## 76	0.170945
## 77	0.220474
## 78	NA
## 79	0.170426
## 80	1.000000
## 81	0.353509
## 82	0.208643
## 83	0.099615
## 84	0.109913
## 85	0.246434
## 86	0.165389
## 87	0.225112
## 88	0.146452
## 89	0.349410
## 90	0.169717
## 91	0.128503
## 92	0.160718
## 93	0.199914
## 94	0.185253
## 95	0.168266
## 96	0.129132
## 97	0.157176
## 98	0.226327
## 99	0.118539
## 100	0.343653
## 101	0.200741

## 102	0.142499	
## 103	0.408556	
## 104	0.113511	
## 105	0.196912	
## 106	1.000000	
## 107	0.162794	
##	percent_of_population_with_1_plus_dose	booster_recip_count
## 1	0.877750	15499
## 2	1.000000	14627
## 3	1.000000	5721
## 4	0.828763	7322
## 5	0.858155	15441
## 6	0.844141	12168
## 7	0.315329	58
## 8	1.000000	73
## 9	1.000000	263
## 10	0.794044	13723
## 11	1.000000	23272
## 12	0.605898	293
## 13	0.870738	9821
## 14	0.963795	14950
## 15	0.472697	73
## 16	0.866711	6477
## 17	0.927730	11413
## 18	0.759461	9014
## 19	0.657040	9070
## 20	1.000000	598
## 21	0.911244	26579
## 22	0.771577	10450
## 23	0.831191	9644
## 24	1.000000	1098
## 25	1.000000	9099
## 26	0.980369	675
## 27	0.913903	6596
## 28	1.000000	16380
## 29	1.000000	4635
## 30	0.858959	13055
## 31	0.910501	3770
## 32	0.660475	7084
## 33	1.000000	246
## 34	1.000000	7124
## 35	0.909893	8460
## 36	0.826396	11614
## 37	1.000000	12257
## 38	0.798178	8881
## 39	0.959507	18793
## 40	0.902476	13251
## 41	0.805792	10837
## 42	0.952452	17482
## 43	0.790756	1218
## 44	0.965160	20425
## 45	0.928454	7896
## 46	0.763491	14195
## 47	0.749094	13171

## 48	0.890636	8791
## 49	0.781182	274
## 50	0.842718	15198
## 51	0.857599	14122
## 52	1.000000	1921
## 53	1.000000	18024
## 54	1.000000	5632
## 55	1.000000	22477
## 56	0.018620	NA
## 57	0.819701	7990
## 58	0.820553	2480
## 59	0.859656	14847
## 60	0.808768	10211
## 61	0.259649	14
## 62	0.922707	9694
## 63	0.212719	NA
## 64	0.085039	NA
## 65	0.127951	18
## 66	0.515385	21
## 67	0.917632	10201
## 68	1.000000	265
## 69	1.000000	10879
## 70	0.573267	372
## 71	0.648455	385
## 72	1.000000	15485
## 73	0.888330	7920
## 74	0.017126	NA
## 75	0.854120	17263
## 76	0.894819	17437
## 77	1.000000	17150
## 78	NA	NA
## 79	1.000000	5791
## 80	1.000000	16
## 81	1.000000	23701
## 82	0.984578	5770
## 83	0.642058	3757
## 84	0.498022	5011
## 85	1.000000	11919
## 86	0.837971	15875
## 87	1.000000	5670
## 88	0.764401	13757
## 89	1.000000	23267
## 90	0.862347	12346
## 91	0.810641	2168
## 92	0.912560	15837
## 93	1.000000	655
## 94	1.000000	15211
## 95	0.936934	6424
## 96	0.739049	8053
## 97	0.956782	10816
## 98	0.967959	13713
## 99	0.645072	279
## 100	1.000000	119
## 101	0.972388	10942

## 102	0.811293	4396	
## 103	1.000000	3362	
## 104	0.599448	628	
## 105	0.872195	5310	
## 106	1.000000	300	
## 107	0.886989	11514	
##			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			No
## 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	Information redacted in accordance with CA state privacy requirements
## 57	No
## 58	No
## 59	No
## 60	No
## 61	No
## 62	No
## 63	Information redacted in accordance with CA state privacy requirements
## 64	Information redacted in accordance with CA state privacy requirements
## 65	No
## 66	No
## 67	No
## 68	No
## 69	No
## 70	No
## 71	No
## 72	No
## 73	No
## 74	Information redacted in accordance with CA state privacy requirements
## 75	No
## 76	No
## 77	No
## 78	Information redacted in accordance with CA state privacy requirements
## 79	No
## 80	No
## 81	No
## 82	No
## 83	No
## 84	No
## 85	No
## 86	No
## 87	No
## 88	No
## 89	No
## 90	No
## 91	No
## 92	No
## 93	No
## 94	No
## 95	No
## 96	No
## 97	No
## 98	No
## 99	No
## 100	No
## 101	No

```
## 102 No
## 103 No
## 104 No
## 105 No
## 106 No
## 107 No
```

```
noNA <- pfv$percent_of_population_fully_vaccinated[!is.na(pfv$percent_of_population_fully_vaccinated)]
mean(noNA)
```

```
## [1] 0.7041551
```

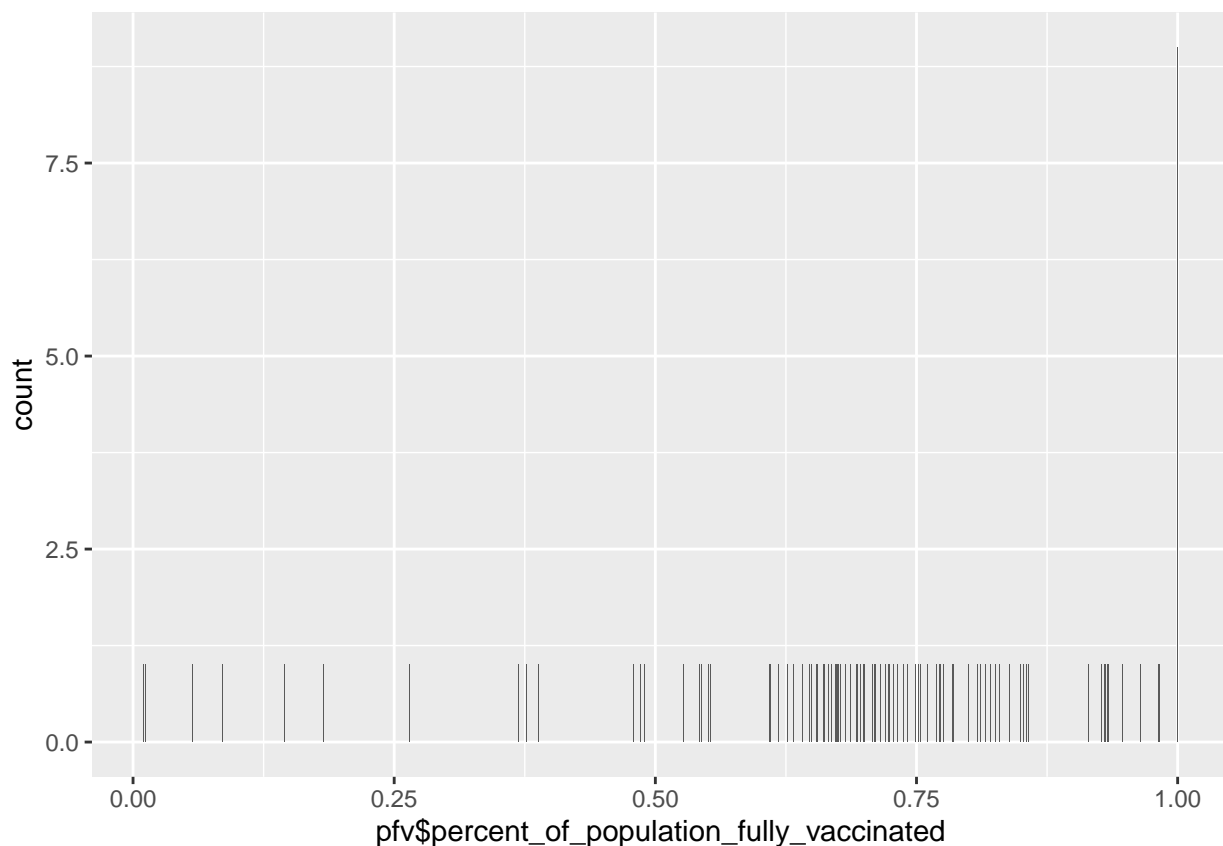
0.7041551 is the average “Percent of Population Fully Vaccinated” in San Diego county as of 2022-02-22.

```
library(ggplot2)
```

```
ggplot(pfv, aes(pfv$percent_of_population_fully_vaccinated,)) + geom_bar()
```

```
## Warning: Use of `pfv$percent_of_population_fully_vaccinated` is discouraged. Use
## `percent_of_population_fully_vaccinated` instead.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_count).
```

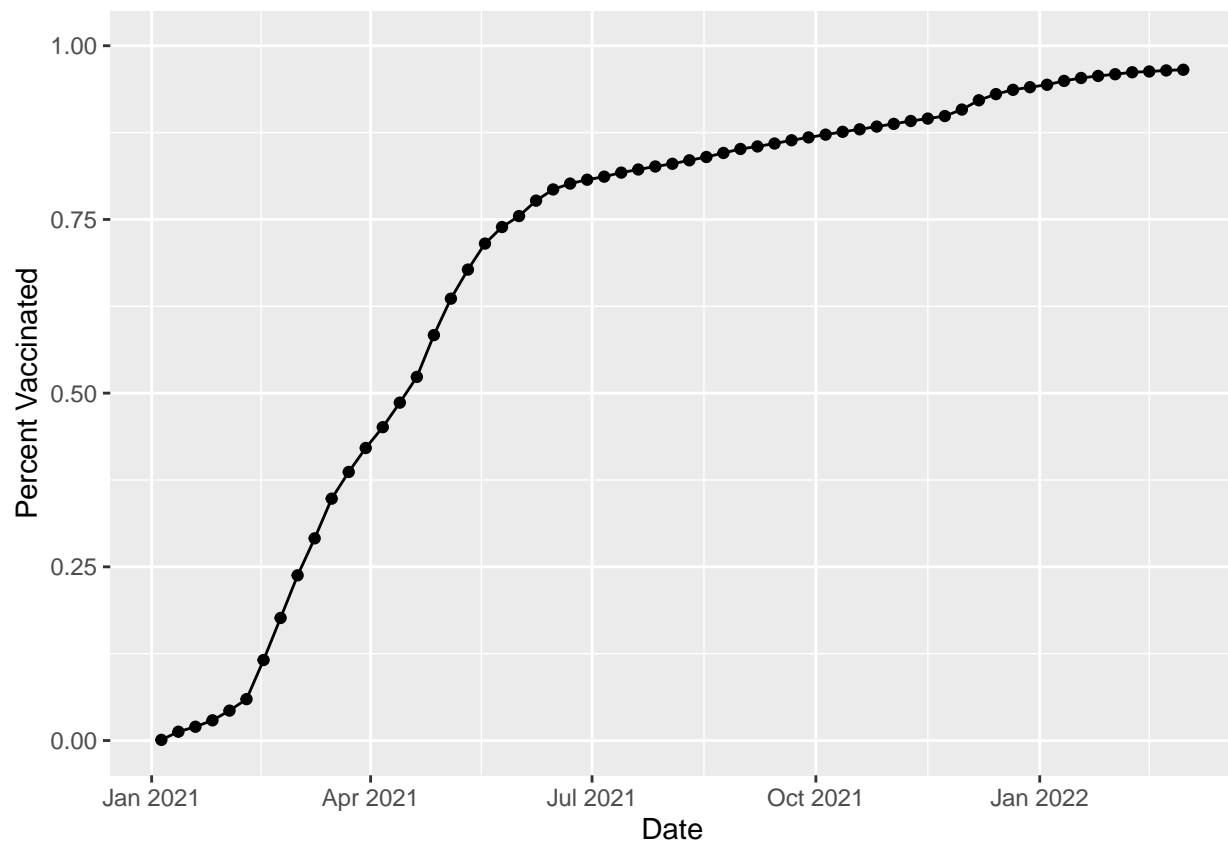


Focus on UCSD/La Jolla

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

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

```
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")
```



```
## Comparing to similar sized areas
```

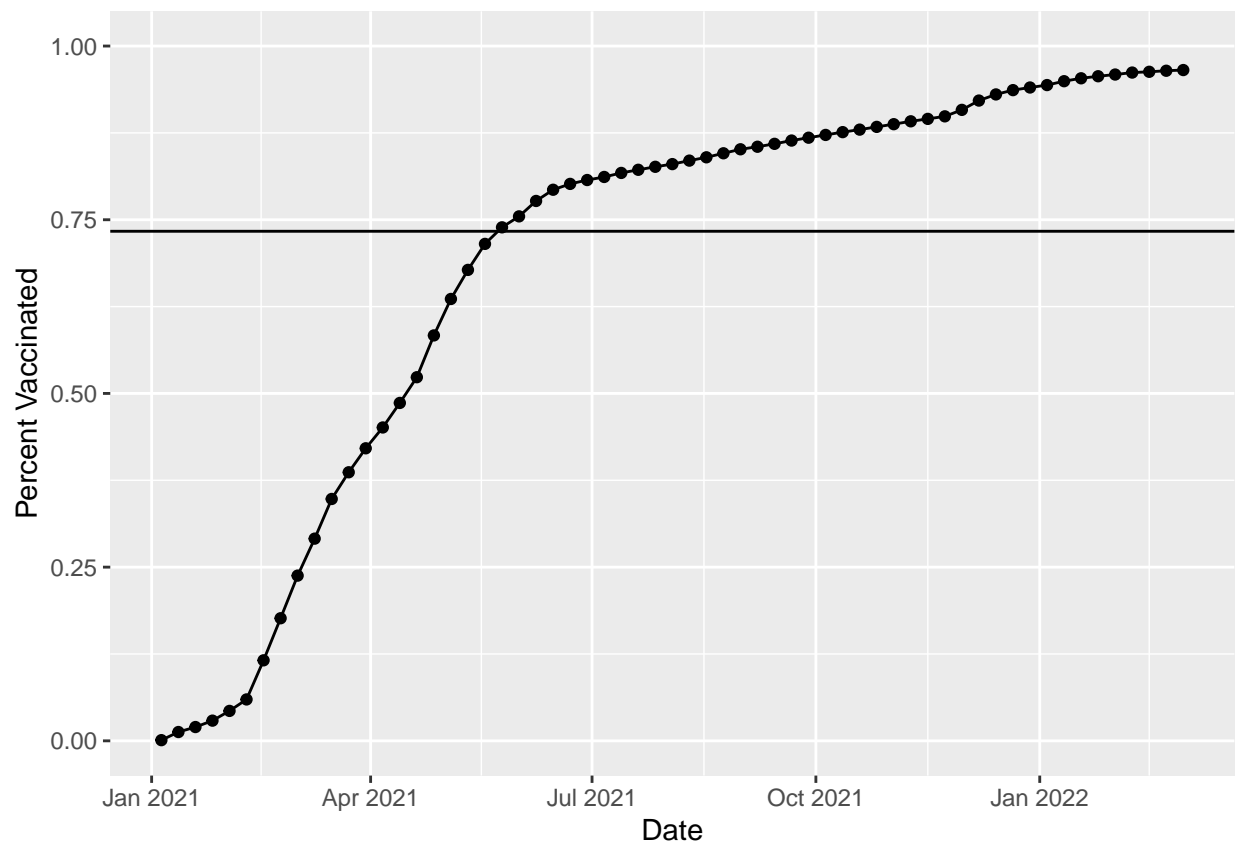
```
# Subset to all CA areas with a population as large as 92037
vax.36 <- filter(vax, age5_plus_population > 36144 &
  as_of_date == "2022-02-22")
#head(vax.36)
```

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 “2022-02-22”. Add this as a straight horizontal line to your plot from above with the `geom_hline()` function?

```
vaxNew <- vax.36$percent_of_population_fully_vaccinated[!is.na(vax.36$percent_of_population_fully_vaccinated)]
hline <- mean(vaxNew)
hline
```

```
## [1] 0.733385
```

```
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")+
  geom_hline(yintercept = hline)
```



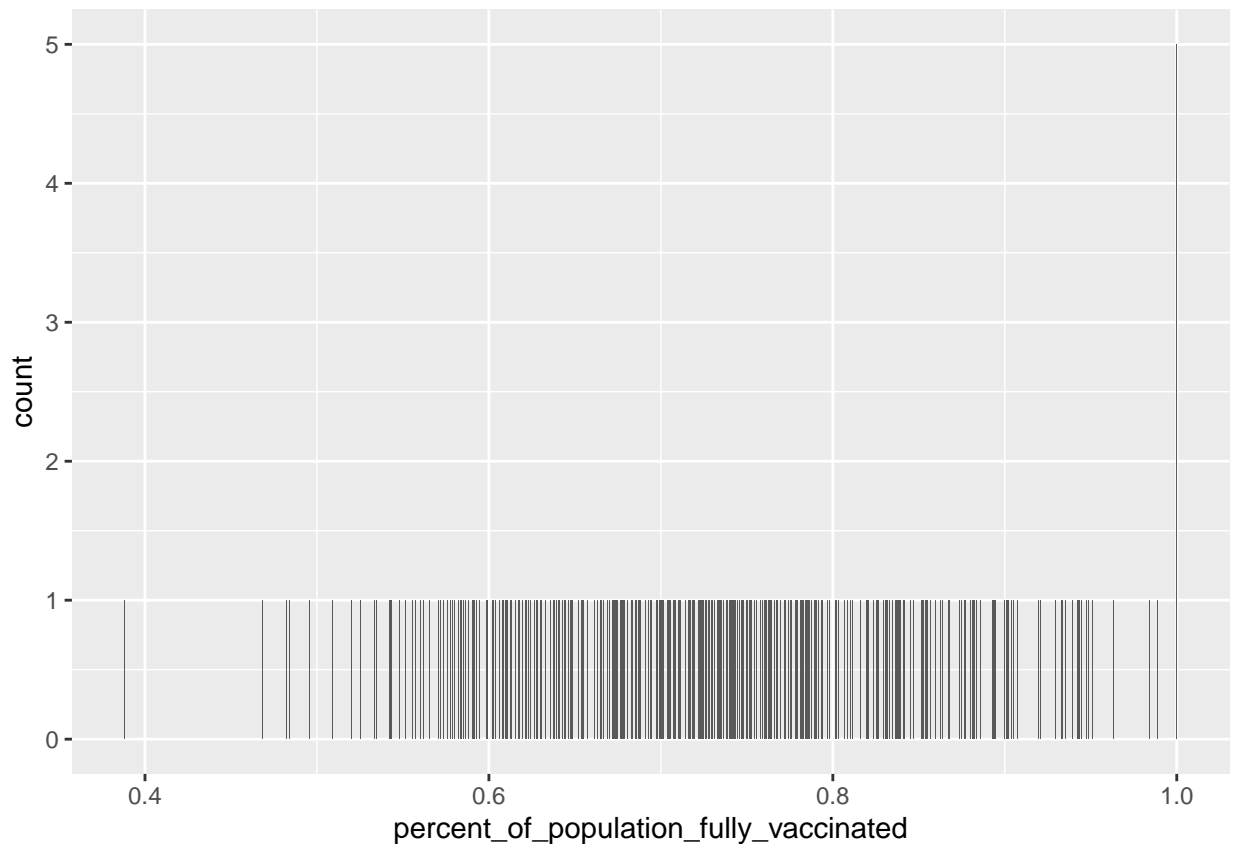
> Q17. What is the 6 number summary (Min, 1st Qu., Median, Mean, 3rd Qu., and Max) of the “Percent of Population Fully Vaccinated” values for ZIP code areas with a population as large as 92037 (La Jolla) as_of_date “2022-02-22”?

```
summary(vaxNew)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3881  0.6539  0.7333  0.7334  0.8027  1.0000
```

Q18. Using ggplot generate a histogram of this data.

```
ggplot(vax.36)+
  aes(x=percent_of_population_fully_vaccinated)+
  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 == "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
```

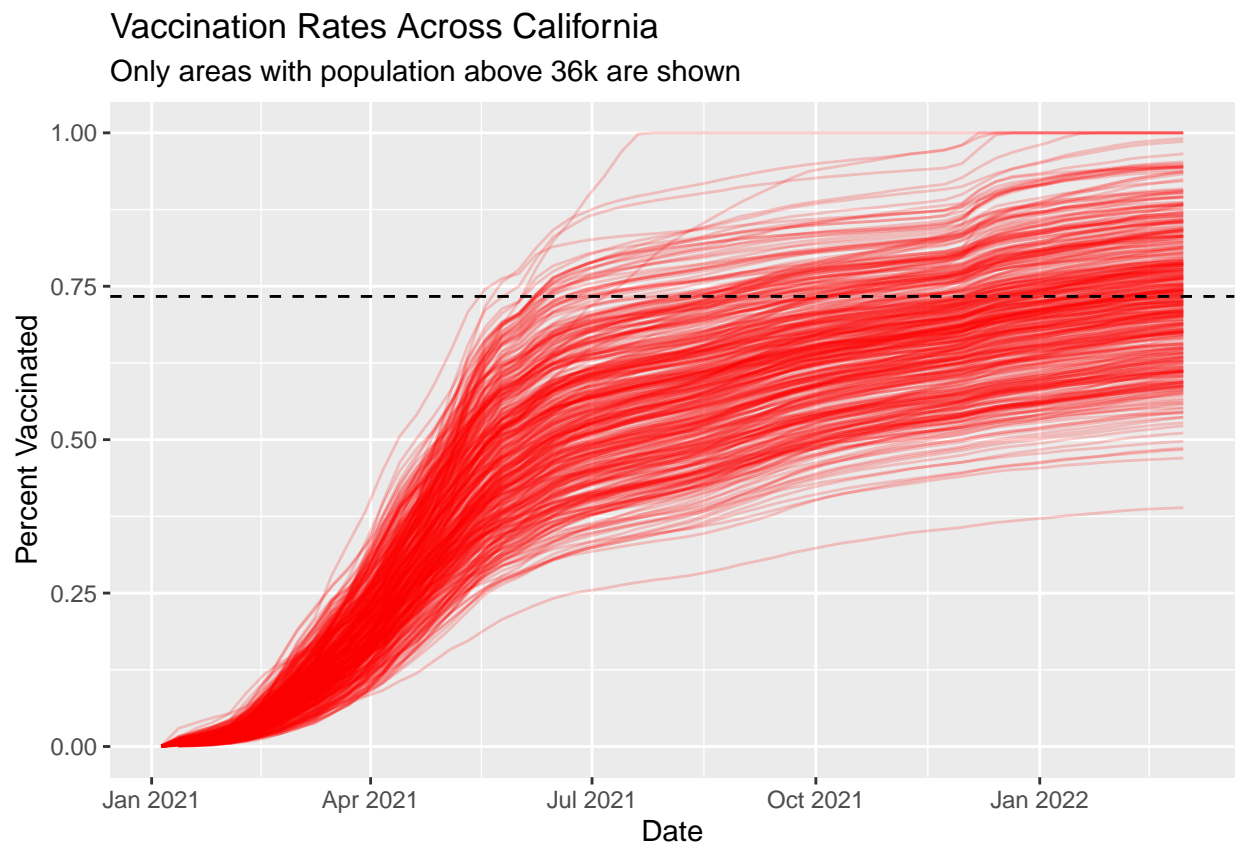
Both are below the average value for the calculated mean of the 92037 zip code that we calculated above.

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= "red") +
  ylim(0,1) +
  labs(x="Date", y="Percent Vaccinated",
       title="Vaccination Rates Across California",
       subtitle="Only areas with population above 36k are shown") +
  geom_hline(yintercept = hline, linetype="dashed")
```

Warning: Removed 311 row(s) containing missing values (geom_path).



Q21. How do you feel about traveling for Spring Break and meeting for in-person class afterwards?

I would like things to go back to normal but I am scared with going back in person.