Covid19

```
library(ggplot2)
```

```
Warning: package 'ggplot2' was built under R version 4.3.1
```

```
vax <- read.csv("d4ff0cd9-d6d7-47a7-a5cb-28572c552298.csv")
head(vax)
```

```
  X_id as_of_date zip_code_tabulation_area local_health_jurisdiction
1    1 2021-01-05                    91402               Los Angeles
2    2 2021-01-05                    91780               Los Angeles
3    3 2021-01-05                    91792               Los Angeles
4    4 2021-01-05                    91917                 San Diego
5    5 2021-01-05                    91931                 San Diego
6    6 2021-01-05                    91325               Los Angeles
       county vaccine_equity_metric_quartile                 vem_source
1 Los Angeles                              1 Healthy Places Index Score
2 Los Angeles                              3 Healthy Places Index Score
3 Los Angeles                              2 Healthy Places Index Score
4   San Diego                              1    CDPH-Derived ZCTA Score
5   San Diego                              3    CDPH-Derived ZCTA Score
6 Los Angeles                              3 Healthy Places Index Score
  age12_plus_population age5_plus_population persons_fully_vaccinated
1               59710.6                67118                       15
2               30388.1                33143                       16
3               25684.7                27834                       15
4                 826.1                  939                       NA
5                 475.7                  586                       NA
6               30419.9                33115                       13
  persons_partially_vaccinated percent_of_population_fully_vaccinated
1                         1255                                0.000223
```

```
2                           755                                  0.000483
3                           759                                  0.000539
4                            NA                                        NA
5                            NA                                        NA
6                           707                                  0.000393
  percent_of_population_partially_vaccinated
1                                   0.018698
2                                   0.022780
3                                   0.027269
4                                         NA
5                                         NA
6                                   0.021350
  percent_of_population_with_1_plus_dose booster_recip_count
1                               0.018921                  NA
2                               0.023263                  NA
3                               0.027808                  NA
4                                     NA                  NA
5                                     NA                  NA
6                               0.021743                  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 data

Q3. What is the earliest date in this dataset?

01-2022

Q4. What is the latest date in this dataset?

01-2021

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

Table 1: Data summary

Name	vax
Number of rows	139356
Number of columns	16
_____	
Column type frequency:	
character	5
numeric	11
_____	
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	79	0
local_health_jurisdiction	0	1	0	15	395	62	0
county	0	1	0	15	395	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
X_id	0	1.00	69678.50	40228.76	1	34839.75	69678.50	104517.25	139356.0	
zip_code_tabulation_area	0	1.00	93665.11	1817.39	90001	92257.75	93658.50	95380.50	97635.0	
vaccine_equity_metric_quartile	6873	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	
age12_plus_population	0	1.00	18895.04	18993.89	0	1346.95	13685.50	31756.12	88556.7	
age5_plus_population	0	1.00	20875.22	21105.99	0	1460.50	15364.00	34877.00	101902.0	
persons_fully_vaccinated	12194	0.91	12497.41	14052.32	11	776.00	6861.50	21023.75	85274.0	
persons_partially_vaccinated	12194	0.91	1760.75	2012.95	11	180.00	1217.50	2627.00	38927.0	
percent_of_population_fully_vaccinated	12194	0.91	0.54	0.27	0	0.36	0.58	0.75	1.0	
percent_of_population_partially_vaccinated	12194	0.91	0.09	0.11	0	0.05	0.07	0.10	1.0	
percent_of_population_with_1_plus_dose	12194	0.91	0.62	0.27	0	0.45	0.66	0.82	1.0	
booster_recip_count	66169	0.53	6235.02	7865.10	11	286.00	2356.00	10417.50	55671.0	

Q5. How many numeric columns are in this dataset? 9

Q6. Note that there are "missing values" in the dataset. How many NA values there in the persons_fully_vaccinated column? 4

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

```
library(lubridate)
```

```
Warning: package 'lubridate' was built under R version 4.3.1
```

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

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

    date, intersect, setdiff, union
```

```
today()
```

```
[1] "2024-03-17"
```

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

```
Time difference of 1167 days
```

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

```
Time difference of 546 days
```

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

546 days

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

1160