Class14

Ayse

2022-02-24

```
# 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                    92140                 San Diego     San Diego
## 2 2021-01-05                    94133             San Francisco San Francisco
## 3 2021-01-05                    94523              Contra Costa  Contra Costa
## 4 2021-01-05                    94005                 San Mateo     San Mateo
## 5 2021-01-05                    94104             San Francisco San Francisco
## 6 2021-01-05                    94549              Contra Costa  Contra Costa
##   vaccine_equity_metric_quartile               vem_source
## 1                             NA          No VEM Assigned
## 2                              3 Healthy Places Index Score
## 3                              4 Healthy Places Index Score
## 4                              4 Healthy Places Index Score
## 5                             NA          No VEM Assigned
## 6                              4 Healthy Places Index Score
##   age12_plus_population age5_plus_population persons_fully_vaccinated
## 1               3747.7                3737                       NA
## 2              25070.5               25957                       NA
## 3              30457.9               32828                       NA
## 4               3996.1                4364                       NA
## 5                387.8                 399                       NA
## 6              25393.8               28468                       NA
##   persons_partially_vaccinated percent_of_population_fully_vaccinated
## 1                           NA                                    NA
## 2                           NA                                    NA
## 3                           NA                                    NA
## 4                           NA                                    NA
## 5                           NA                                    NA
## 6                           NA                                    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
##                                                                    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 9

Q2. What column details the Zip code tabulation area?

column 2

Q3. What is the earliest date in this dataset?

January 5 2021

> Q4. What is the latest date in this dataset?

February 22, 2022

```
#overview
#install.packages("skimr")
library(skimr)
skimr::skim(vax)
```

Data summary

Name	vax
Number of rows	105840
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	60	0
local_health_jurisdiction	0	1	0	15	300	62	0
county	0	1	0	15	300	59	0
vem_source	0	1	15	26	0	3	0
redacted	0	1	2	69	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
zip_code_tabulation_area	0	1.00	93665.11	1817.39	90001	92257.75	93658.50	95380.50	97635.0	▁▁▇▇▇
vaccine_equity_metric_quartile	5220	0.95	2.44	1.11	1	1.00	2.00	3.00	4.0	▇▇▁▇▁
age12_plus_population	0	1.00	18895.04	18993.92	0	1346.95	13685.10	31756.12	88556.7	▇▃▂▁▁
age5_plus_population	0	1.00	20875.24	21106.02	0	1460.50	15364.00	34877.00	101902.0	▇▃▂▁▁
persons_fully_vaccinated	18174	0.83	12064.29	12983.91	11	1059.00	7287.50	19859.00	77213.0	▇▃▁▁▁
persons_partially_vaccinated	18174	0.83	820.71	1318.77	11	76.00	370.00	1066.00	31869.0	▇▁▁▁▁
percent_of_population_fully_vaccinated	18174	0.83	0.51	0.26	0	0.33	0.54	0.70	1.0	▂▇▇▇▇
percent_of_population_partially_vaccinated	18174	0.83	0.05	0.09	0	0.01	0.03	0.05	1.0	▇▁▁▁▁
percent_of_population_with_1_plus_dose	18174	0.83	0.54	0.27	0	0.35	0.58	0.75	1.0	▃▃▅▇▇
booster_recip_count	64191	0.39	3923.43	5704.10	11	169.00	1072.00	5803.00	49951.0	▇▁▁▁▁

> 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?

18174

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

17%

```
fv_na <- sum( is.na(vax$persons_fully_vaccinated) )
(fv_na/length(vax$persons_fully_vaccinated))*100
```

```
## [1] 17.1712
```

#working with dates

Q9. How many days have passed since the last update of the dataset? 0 Q10. How many unique dates are in the dataset (i.e. how many different dates are detailed)? 413

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

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

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

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

```
## Date of length 0
```

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

```
## Time difference of 413 days
```

#working with zipcodes

```
#install.packages("zipcodeR")
library(zipcodeR)
geocode_zip('92037')
```

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

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

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

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

```
## # A tibble: 2 × 24
##   zipcode zipcode_type major_city post_office_city common_city_list county state
##   <chr>   <chr>        <chr>      <chr>                      <blob> <chr>  <chr>
## 1 92037   Standard     La Jolla   La Jolla, CA            <raw 20 B> San D… CA
## 2 92109   Standard     San Diego  San Diego, CA           <raw 21 B> San D… CA
## # … with 17 more variables: lat <dbl>, lng <dbl>, timezone <chr>,
## #   radius_in_miles <dbl>, area_code_list <blob>, population <int>,
## #   population_density <dbl>, land_area_in_sqmi <dbl>,
## #   water_area_in_sqmi <dbl>, housing_units <int>,
## #   occupied_housing_units <int>, median_home_value <int>,
## #   median_household_income <int>, bounds_west <dbl>, bounds_east <dbl>,
## #   bounds_north <dbl>, bounds_south <dbl>
```

```
Q11. How many distinct zip codes are listed for San Diego County? 6420
Q12. What San Diego County Zip code area has the largest 12 + Population in this dataset?
```

```r
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
```

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

nrow(sd)
```

```
## [1] 6420
```

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

uni <- length(unique(sd))
```

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

```r
avg <- mean(sd$percent_of_population_fully_vaccinated)
```

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 "2022-02-22"?