

US State Vaccinations

Jefferson Santos - ANA 515

11/09/2021

This dataset provides information about all the vaccinations by US State. It has been collected from two main websites: <https://ourworldindata.org/us-states-vaccinations> (<https://ourworldindata.org/us-states-vaccinations>) and https://covid.ourworldindata.org/data/vaccinations/us_state_vaccinations.csv (https://covid.ourworldindata.org/data/vaccinations/us_state_vaccinations.csv) It has vaccines data like: total distributed, people vaccinated, daily vaccinations e share doses used. This is saved as .csv file (comma-separated file). We can open it using microsoft excel or any text editor.

```
# using read.csv function and assigning it to an object
# We will use the built in read.csv(...) function call, which reads the data in as a data frame, and assign the data frame to a variable (using <-) so that it is stored in R's memory.
data <- read.csv('us_state_vaccinations.csv')
```

```
# getting the column names
colnames(data)
```

```
## [1] "date" "location"
## [3] "total_vaccinations" "total_distributed"
## [5] "people_vaccinated" "people_fully_vaccinated_per_hundred"
## [7] "total_vaccinations_per_hundred" "people_fully_vaccinated"
## [9] "people_vaccinated_per_hundred" "distributed_per_hundred"
## [11] "daily_vaccinations_raw" "daily_vaccinations"
## [13] "daily_vaccinations_per_million" "share_doses_used"
```

```
# renaming columns
data <- rename(data, state = location)
```

```
# removing few columns from the dataset
data_subset <- select(data, date, state, total_vaccinations, total_distributed, people_fully_vaccinated, daily_vaccinations, share_doses_used)
```

```
# getting the column names again to confirm the changes
colnames(data_subset)
```

```
## [1] "date" "state"
## [3] "total_vaccinations" "total_distributed"
## [5] "people_fully_vaccinated" "daily_vaccinations"
## [7] "share_doses_used"
```

```
# getting the dimensions of the data subset
dimensions <- dim(data_subset)
```

This dataframe has 18784 rows and 7 columns. The names of the columns and a brief description of each are in the table below:

```
library(knitr)

columns_summary <- data.frame(
  Columns = c(colnames(data_subset)),
  Description = c(
    "Date of the information been provided",
    "US State/Location",
    "Cumulative vaccinations at that date per state",
    "Cumulative distribution at that date per state",
    "Cumulative people fully vaccinated at that date per state",
    "Total of people vaccinated on that date per state",
    "The ratio between Cumulative distribution by Cumulative vaccinations"
  )
)

kable(columns_summary, caption = "US State Vaccinations Columns Description")
```

US State Vaccinations Columns Description

Columns	Description
date	Date of the information been provided
state	US State/Location
total_vaccinations	Cumulative vaccinations at that date per state
total_distributed	Cumulative distribution at that date per state
people_fully_vaccinated	Cumulative people fully vaccinated at that date per state
daily_vaccinations	Total of people vaccinated on that date per state
share_doses_used	The ratio between Cumulative distribution by Cumulative vaccinations

```

#filtering the data by location = entire US instead of by State to avoid an incorret summary below
data_us <- filter(data_subset, state == "United States")

# picking three columns to use summary function
data_pick3 <- select(data_us, people_fully_vaccinated, daily_vaccinations, share_doses_used)

# Summary functions (min, max, mean, missing values)
pfv_min <- format(big.mark="," , scientific=FALSE,round(min(data_pick3$people_fully_vaccinated, na.rm=TRUE),2), nsmall = 2)
pfv_max <- format(big.mark="," , scientific=FALSE,round(max(data_pick3$people_fully_vaccinated, na.rm=TRUE),2), nsmall = 2)
pfv_mean <- format(big.mark="," , scientific=FALSE,round(mean(data_pick3$people_fully_vaccinated, na.rm=TRUE),2), nsmall = 2)
pfv_sumna <- sum(is.na(data_pick3$people_fully_vaccinated))

dv_min <- format(big.mark="," , scientific=FALSE,round(min(data_pick3$daily_vaccinations, na.rm=TRUE),2), nsmall = 2)
dv_max <- format(big.mark="," , scientific=FALSE,round(max(data_pick3$daily_vaccinations, na.rm=TRUE),2), nsmall = 2)
dv_mean <- format(big.mark="," , scientific=FALSE,round(mean(data_pick3$daily_vaccinations, na.rm=TRUE),2), nsmall = 2)
dv_sumna <- sum(is.na(data_pick3$daily_vaccinations))

sdu_min <- format(big.mark="," , scientific=FALSE,round(min(data_pick3$share_doses_used, na.rm=TRUE),2), nsmall = 2)
sdu_max <- format(big.mark="," , scientific=FALSE,round(max(data_pick3$share_doses_used, na.rm=TRUE),2), nsmall = 2)
sdu_mean <- format(big.mark="," , scientific=FALSE,round(mean(data_pick3$share_doses_used, na.rm=TRUE),2), nsmall = 2)
sdu_sumna <- sum(is.na(data_pick3$share_doses_used))

```

```

summary_stats <- data.frame(
  Summary = c(colnames(data_pick3)),
  Min = c(pfv_min,
          dv_min,
          sdu_min
        ),
  Max = c(pfv_max,
          dv_max,
          sdu_max
        ),
  Mean = c(pfv_mean,
           dv_mean,
           sdu_mean
        ),
  NAs = c(pfv_sumna,
           dv_sumna,
           sdu_sumna
        )
)

kable(t(summary_stats), caption = "US Vaccinations Summary Stats")

```

US Vaccinations Summary Stats

Summary	people_fully_vaccinated	daily_vaccinations	share_doses_used
---------	-------------------------	--------------------	------------------

Min	782,228.00	57,909.00	0.34
Max	193,227,813.00	3,384,387.00	0.87
Mean	115,895,398.59	1,330,246.28	0.79
NAs	34	1	33