

# Tidying and Transforming Vaccination Data

Nnaemeka Newman Okereafor

9/25/2021

## About the Assignment

The dataset describes August 2021 data for Israeli hospitalization (“Severe Cases”) rates for people under 50 (assume “50 and under”) and over 50, for both un-vaccinated and fully vaccinated populations. The data was analyzed, and miscellaneous questions were also answered.

```
library(tidyverse)
library(stringr)

ulr <- "https://raw.githubusercontent.com/nnaemeka-git/global-datasets/main
/israeli_vaccination_data.csv"
df<-read.csv(ulr,header = FALSE,skip=2,nrow=7)
df
```

```
##      V1        V2        V3   V4   V5   V6   V7   V8   V9   V10  V11  V12  V13
## 1 <50 1,116,834 3,501,118  43   11  NA  NA  NA  NA  NA  NA  NA  NA  NA
## 2          23.3%    73.0%  NA  NA
## 3 >50 186,078 2,133,516 171  290  NA  NA  NA  NA  NA  NA  NA  NA  NA
## 4          7.9%    90.4%  NA  NA
## 5          NA  NA
## 6          NA  NA
## 7          NA  NA
```

```
glimpse(df)
```

```
## #> #> Rows: 7
## #> #> Columns: 13
## #> #> $ V1 <chr> "<50", " ", ">50", "", "", "", ""
## #> #> $ V2 <chr> "1,116,834", "23.3%", "186,078", "7.9%", "", "", ""
## #> #> $ V3 <chr> "3,501,118", "73.0%", "2,133,516", "90.4%", "", "", ""
## #> #> $ V4 <int> 43, NA, 171, NA, NA, NA, NA
## #> #> $ V5 <int> 11, NA, 290, NA, NA, NA, NA
## #> #> $ V6 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V7 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V8 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V9 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V10 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V11 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V12 <lgl> NA, NA, NA, NA, NA, NA, NA
## #> #> $ V13 <lgl> NA, NA, NA, NA, NA, NA, NA
```

```
df<-df%>%select(c('V1','V2','V3','V4','V5','V6','V7'))
df
```

```
##      V1          V2          V3    V4    V5    V6    V7
## 1 <50 1,116,834 3,501,118   43    11   NA   NA
## 2           23.3%     73.0%   NA   NA   NA   NA
## 3 >50 186,078 2,133,516  171   290   NA   NA
## 4           7.9%     90.4%   NA   NA   NA   NA
## 5                   NA   NA   NA   NA
## 6                   NA   NA   NA   NA
## 7                   NA   NA   NA   NA
```

## Age Range

```
#Extract population Age
Age <- unlist(str_extract_all(df$V1, "\\<\\d+.|\\>\\d+"))
Age<-as.data.frame(Age)
colnames(Age)<-"Age"
Age

##      Age
## 1 <50
## 2 >50
```

## Percentage figures

```
#Extract population percent not vaccinated
perc_pop_not_vax <- unlist(str_extract_all(df$V2, "\\d+..\\%"))
perc_pop_not_vax <-str_remove_all(perc_pop_not_vax,"\\%")
perc_pop_not_vax <-as.numeric(perc_pop_not_vax)
perc_pop_not_vax<-as.data.frame(perc_pop_not_vax)
colnames(perc_pop_not_vax)<-"Percent Population not vax"
perc_pop_not_vax

##      Percent Population not vax
## 1                      23.3
## 2                      7.9

#Extract population percent Fully vaccinated
perc_pop_fully_vax <- unlist(str_extract_all(df$V3, "\\d+..\\%"))
perc_pop_fully_vax <-str_remove_all(perc_pop_fully_vax,"\\%")
perc_pop_fully_vax <-as.numeric(perc_pop_fully_vax)
perc_pop_fully_vax

## [1] 73.0 90.4
```

```
#Convert vector to dataframe
perc_pop_fully_vax<-as.data.frame(perc_pop_fully_vax)
colnames(perc_pop_fully_vax)<-"Percent Population fully vax"
perc_pop_fully_vax
```

```
## Percent Population fully vax
## 1 73.0
## 2 90.4
```

## Population figures

```
#Extract population not vaccinated
pop_not_vax <- unlist(str_extract_all(df$V2, "[[:digit:]]+.[[:digit:]]+\.\d+"))
pop_not_vax <-str_remove_all(pop_not_vax, "\\",")
pop_not_vax <-as.numeric(pop_not_vax)
pop_not_vax
```

```
## [1] 1116834 186078
```

```
#Convert vector to dataframe
pop_not_vax<-as.data.frame(pop_not_vax)
colnames(pop_not_vax)<-"Population not vax"
pop_not_vax
```

```
## Population not vax
## 1 1116834
## 2 186078
```

```
#Extract population fully vaccinated
pop_fully_vax <- unlist(str_extract_all(df$V3, "[[:digit:]]+.[[:digit:]]+\.\d+"))
pop_fully_vax <-str_remove_all(pop_fully_vax, "\\",")
pop_fully_vax <-as.numeric(pop_fully_vax)
pop_fully_vax
```

```
## [1] 3501118 2133516
```

```
#Convert vector to dataframe
pop_fully_vax<-as.data.frame(pop_fully_vax)
colnames(pop_fully_vax)<-"Population fully vax"
pop_fully_vax
```

```
## Population fully vax
## 1 3501118
## 2 2133516
```

## Severe case figures

```

#Extract severe cases not vaccinated
sev_cases_not_vax <- unlist(str_replace_na(df$V4, "0"))
sev_cases_not_vax <- unlist(str_extract_all(sev_cases_not_vax, "[[:digit:]]{2,}"))
sev_cases_not_vax <- as.numeric(sev_cases_not_vax)
sev_cases_not_vax

## [1] 43 171

#Convert vector to dataframe
sev_cases_not_vax<-as.data.frame(sev_cases_not_vax)
colnames(sev_cases_not_vax)<-"Severe cases not vax"
sev_cases_not_vax

## Severe cases not vax
## 1 43
## 2 171

#Extract severe cases fully vaccinated
sev_fully_vax <- unlist(str_replace_na(df$V5, "0"))
sev_fully_vax <-unlist(str_extract_all(sev_fully_vax, "[[:digit:]]{2,}"))
sev_fully_vax <-as.numeric(sev_fully_vax)
sev_fully_vax

## [1] 11 290

#Convert vector to dataframe
sev_fully_vax<-as.data.frame(sev_fully_vax)
colnames(sev_fully_vax)<-"Severe cases fully vax"
sev_fully_vax

## Severe cases fully vax
## 1 11
## 2 290

```

## New table

```

#Combine dataframe
table<-data.frame(Age,pop_not_vax,perc_pop_not_vax,pop_fully_vax,perc_pop_fully_vax,sev_cases_not_vax,sev_fully_vax)
table

## Age Population.not.vax Percent.Population.not.vax Population.fully.vax
## 1 <50 1116834 23.3 3501118
## 2 >50 186078 7.9 2133516
## Percent.Population.fully.vax Severe.cases.not.vax Severe.cases.fully.vax
## 1 73.0 43 11
## 2 90.4 171 290

```

## Derive other columns

```

table<-table%>%
  mutate(severe.fully.vax.rate = round((table$Severe.cases.fully.vax/table$Population.fully.vax)*100000,
    severe.cases.not.vax.rate = round((table$Severe.cases.not.vax/table$Population.not.vax)*100000,
    Efficacy =round(((1-(severe.fully.vax.rate/severe.cases.not.vax.rate))*100),1),
    Severe.case.rate=round(severe.cases.not.vax.rate/severe.fully.vax.rate),1)
  
```

### Select Columns in order

```

table<-table %>% select(Age,
  Population.not.vax,
  Percent.Population.not.vax,
  Population.fully.vax,
  Percent.Population.fully.vax,
  Severe.cases.not.vax,severe.cases.not.vax.rate,
  Severe.cases.fully.vax,
  severe.fully.vax.rate,
  Efficacy,
  Severe.case.rate)
table

##   Age Population.not.vax Percent.Population.not.vax Population.fully.vax
## 1 <50          1116834                  23.3        3501118
## 2 >50          186078                   7.9        2133516
##   Percent.Population.fully.vax Severe.cases.not.vax severe.cases.not.vax.rate
## 1                 73.0                  43            3.9
## 2                 90.4                 171           91.9
##   Severe.cases.fully.vax severe.fully.vax.rate Efficacy Severe.case.rate
## 1                  11                  0.3       92.3         13
## 2                 290                 13.6      85.2          7

names(table)

##  [1] "Age"                      "Population.not.vax"
##  [3] "Percent.Population.not.vax" "Population.fully.vax"
##  [5] "Percent.Population.fully.vax" "Severe.cases.not.vax"
##  [7] "severe.cases.not.vax.rate"   "Severe.cases.fully.vax"
##  [9] "severe.fully.vax.rate"      "Efficacy"
## [11] "Severe.case.rate"
  
```

### Miscellaneous Questions

- (1) Do you have enough information to calculate the total population? What does this total population represent?

No. The total population comprises of people that are not vaccinated and people fully vaccinated and are above 16 years of age.

**(2) Calculate the Efficacy vs. Disease; Explain your results.**

The efficacy of the vaccine was calculated based on the normalized values of severe cases per 100,000 people. The efficacy of approximately 92% and 85% for people <50 (more than 16years) years category and >50 years category respectively indicates that the vaccines are effective in preventing severe disease in people of those age range. However, one could speculate that the vaccines are more effective in people in the age less than 50 years.

**(3) From your calculation of efficacy vs. disease, are you able to compare the rate of severe cases in unvaccinated individuals to that in vaccinated individuals?**

From the table above, the rate of severe cases is approximately 13x higher in severe not vaccinated people than severe fully vaccinated people for individuals in <50 years age category. While it is approximately 7x higher in severe not vaccinated people than severe fully vaccinated people for individual in >50 years age category.