

Assignment_03

```
# Cleaning the memory and loading packages.
rm(list=ls())
suppressPackageStartupMessages(library(tidyverse))
suppressPackageStartupMessages(library(jsonlite))
suppressPackageStartupMessages(library(httr))
```

Task 1)

Scrapping data

```
# Create a GET response to call the API.
us_corona <- httr::GET("https://static01.nyt.com/newsgraphics/2021/12/20/us-coronavirus-deaths-")

# View API GET response result.
str(us_corona)
```

List of 10

```
$ url      : chr "https://static01.nyt.com/newsgraphics/2021/12/20/us-coronavirus-deaths-"
$ status_code: int 200
$ headers  :List of 29
..$ connection      : chr "keep-alive"
..$ content-length   : chr "1867"
..$ x-guploader-uploadid : chr "ADPycds20xaeJVcpnN6epfXTcjhExwtkNeLqt1Vn1rcukpH9YI"
..$ cache-control    : chr "max-age=5"
..$ expires          : chr "Thu, 09 Mar 2023 19:12:13 GMT"
..$ last-modified     : chr "Wed, 29 Dec 2021 22:06:11 GMT"
..$ etag             : chr "\"3fc9f17e8bf0ee91cb8666d834ec478b\""
..$ x-goog-generation : chr "1640815571754608"
..$ x-goog-metageneration : chr "1"
```

```

..$ x-goog-stored-content-encoding: chr "identity"
..$ x-goog-stored-content-length : chr "7559"
..$ content-type                  : chr "application/json"
..$ x-goog-hash                   : chr "crc32c=xoibdQ=="
..$ x-goog-hash                   : chr "md5=P8nxfov7pHLhmbYN0xHiw=="
..$ x-goog-storage-class          : chr "MULTI_REGIONAL"
..$ server                        : chr "UploadServer"
..$ content-encoding              : chr "gzip"
..$ via                           : chr "1.1 varnish, 1.1 varnish"
..$ accept-ranges                 : chr "bytes"
..$ date                          : chr "Thu, 09 Mar 2023 20:23:07 GMT"
..$ age                           : chr "0"
..$ x-served-by                   : chr "cache-iad-kiad7000179-IAD, cache-bma1672-BMA"
..$ x-cache                       : chr "MISS, HIT"
..$ x-cache-hits                  : chr "0, 1"
..$ x-timer                       : chr "S1678393387.919465,VS0,VE125"
..$ vary                          : chr "Origin, Accept-Encoding"
..$ access-control-allow-origin   : chr "*"
..$ timing-allow-origin           : chr "*"
..$ strict-transport-security     : chr "max-age=63072000; preload; includeSubdomains"
..- attr(*, "class")= chr [1:2] "insensitive" "list"
$ all_headers:List of 1
..$ :List of 3
.. ..$ status : int 200
.. ..$ version: chr "HTTP/1.1"
.. ..$ headers:List of 29
.. .. ..$ connection          : chr "keep-alive"
.. .. ..$ content-length      : chr "1867"
.. .. ..$ x-guploader-uploadid : chr "ADPycds20xaeJVcpnN6epfXTcjhExwtkNeLqt1Vn1rc"
.. .. ..$ cache-control       : chr "max-age=5"
.. .. ..$ expires             : chr "Thu, 09 Mar 2023 19:12:13 GMT"
.. .. ..$ last-modified       : chr "Wed, 29 Dec 2021 22:06:11 GMT"
.. .. ..$ etag                : chr "\"3fc9f17e8bf0ee91cb8666d834ec478b\""
.. .. ..$ x-goog-generation    : chr "1640815571754608"
.. .. ..$ x-goog-metageneration : chr "1"
.. .. ..$ x-goog-stored-content-encoding: chr "identity"
.. .. ..$ x-goog-stored-content-length : chr "7559"
.. .. ..$ content-type        : chr "application/json"
.. .. ..$ x-goog-hash          : chr "crc32c=xoibdQ=="
.. .. ..$ x-goog-hash          : chr "md5=P8nxfov7pHLhmbYN0xHiw=="
.. .. ..$ x-goog-storage-class : chr "MULTI_REGIONAL"
.. .. ..$ server               : chr "UploadServer"
.. .. ..$ content-encoding     : chr "gzip"

```

```

.. .. ..$ via : chr "1.1 varnish, 1.1 varnish"
.. .. ..$ accept-ranges : chr "bytes"
.. .. ..$ date : chr "Thu, 09 Mar 2023 20:23:07 GMT"
.. .. ..$ age : chr "0"
.. .. ..$ x-served-by : chr "cache-iad-kiad7000179-IAD, cache-bma1672-BM"
.. .. ..$ x-cache : chr "MISS, HIT"
.. .. ..$ x-cache-hits : chr "0, 1"
.. .. ..$ x-timer : chr "S1678393387.919465,VS0,VE125"
.. .. ..$ vary : chr "Origin, Accept-Encoding"
.. .. ..$ access-control-allow-origin : chr "*"
.. .. ..$ timing-allow-origin : chr "*"
.. .. ..$ strict-transport-security : chr "max-age=63072000; preload; includeSubdomain"
.. .. ..- attr(*, "class")= chr [1:2] "insensitive" "list"
$ cookies : 'data.frame': 0 obs. of 7 variables:
..$ domain : logi(0)
..$ flag : logi(0)
..$ path : logi(0)
..$ secure : logi(0)
..$ expiration: 'POSIXct' num(0)
..$ name : logi(0)
..$ value : logi(0)
$ content : raw [1:7559] 5b 7b 22 6e ...
$ date : POSIXct[1:1], format: "2023-03-09 20:23:07"
$ times : Named num [1:6] 0 0.0717 0.1015 0.1679 0.3226 ...
..- attr(*, "names")= chr [1:6] "redirect" "namelookup" "connect" "pretransfer" ...
$ request :List of 7
..$ method : chr "GET"
..$ url : chr "https://static01.nyt.com/newsgraphics/2021/12/20/us-coronavirus-death"
..$ headers : Named chr "application/json, text/xml, application/xml, */*"
.. ..- attr(*, "names")= chr "Accept"
..$ fields : NULL
..$ options :List of 2
.. ..$ useragent: chr "libcurl/7.64.1 r-curl/4.3.2 httr/1.4.4"
.. ..$ httpget : logi TRUE
..$ auth_token: NULL
..$ output : list()
.. ..- attr(*, "class")= chr [1:2] "write_memory" "write_function"
..- attr(*, "class")= chr "request"
$ handle :Class 'curl_handle' <externalptr>
- attr(*, "class")= chr "response"

```

```
# View main source of the data needed from the API - Normally content.  
str(us_corona$content)
```

```
raw [1:7559] 5b 7b 22 6e ...
```

```
# Convert the raw content to text (JSON string).  
us_corona_content <- httr::content(us_corona, as = "text")  
  
# View JSON string result.  
str(us_corona_content)
```

```
chr "[{"name\":\"Alabama\", \"geoid\":\"USA-01\", \"deaths_before\":10790, \"deaths\":5590, \"T
```

```
# Convert the JSON string to a dataframe.  
us_corona_JSON <- jsonlite::fromJSON(us_corona_content)  
  
# View data in a table.  
view(us_corona_JSON)
```

Creating new variable

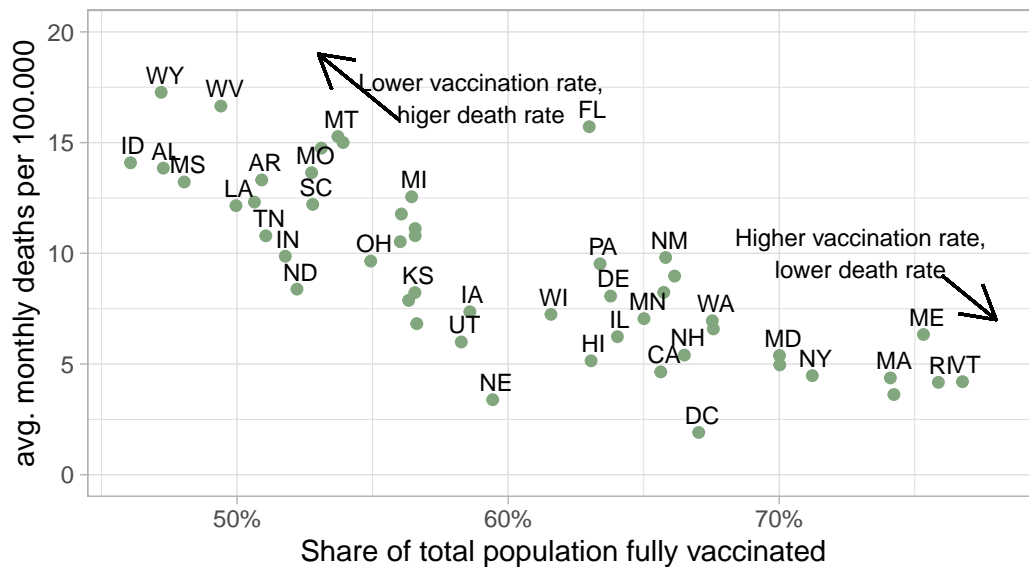
```
# Getting short names for US states.  
states = state.abb  
  
# Adding DC since its not in the dataset.  
states = append(states, "DC", after = 8)  
  
# Removing short names that will create clutter.  
replace_str = c("CT" = " ", "NJ" = " ", "VA" = " ", "CO" = " ", "OR" = " ", "NC" = " ", "T  
  
states <- str_replace_all(states, replace_str)  
  
# Adding the short names.  
us_corona_JSON <- us_corona_JSON %>%  
  mutate(short = states)
```

Creating figure

```
# Creating the figure.
us_corona_figure <- us_corona_JSON %>%
  ggplot(aes(x = fully_vaccinated_pct_of_pop, y = deaths_per_100k,
             label = short)) +
  geom_point(color = "#82A67D", size = 1.6) +
  geom_text(hjust = 0.4, vjust = -0.5, size = 3) +
  ylim(0.00, 20.0) +
  labs(title = "Covid-19 deaths since universal adult vaccine eligibility
compared with vaccination rates",
       y = "avg. monthly deaths per 100.000",
       x = "Share of total population fully vaccinated")+
  scale_x_continuous(label = scales::percent) +
  annotate("text", x = 0.59, y = 17, label = "Lower vaccination rate,
higer death rate", size = 3) +
  annotate("text", x = 0.73, y = 10, label = "Higher vaccination rate,
lower death rate", size = 3) +
  geom_segment(aes(x = 0.76, y = 9, xend = 0.78, yend = 7),
              arrow = arrow(length = unit(.5, "cm")))+
  geom_segment(aes(x = 0.56, y = 16, xend = 0.53, yend = 19),
              arrow = arrow(length = unit(.5, "cm")))+
  theme_light()

# Showing figure.
us_corona_figure
```

Covid-19 deaths since universal adult vaccine eligibility compared with vaccination rates



Task 2)

Running `lm()` and adding to figure

```
# Running lm() function
lm_result <- lm(deaths_per_100k ~ fully_vaccinated_pct_of_pop, us_corona_JSON)

# Showing result
lm_result
```

Call:

```
lm(formula = deaths_per_100k ~ fully_vaccinated_pct_of_pop, data = us_corona_JSON)
```

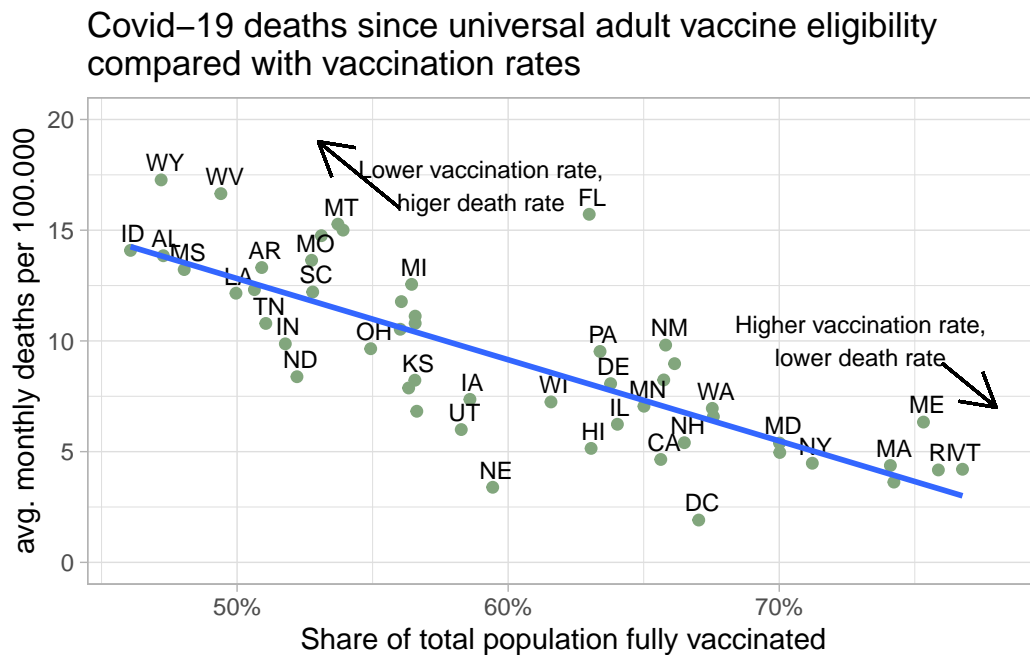
Coefficients:

```
(Intercept)  fully_vaccinated_pct_of_pop
      31.15                -36.66
```

The `lm()` (linear models) function is used to fit linear regression models. The output shows us the equation to that line. So the equation to this linear function would be: $31.15 - 36.66x$

Since the figure has percentage on the x-axis, one “step” to the right would mean 100% (1.00). That suggests the line would go beneath the x-axis before it hits 100%, and avg. monthly deaths would go below 0. In theory the data suggest that not everyone would need to be vaccinated for the US to get 0 deaths from corona. We know this is not true and its just a simplification of reality, but its a good method to easily spot trends and see the effect of the vaccine.

```
# Adding the custom line and "removing standard error bar" to the previous figure.
us_corona_figure + geom_smooth(method = lm, se = FALSE)
```



Source

- Dean Chereden. (15.08.2022). *How to GET data from an API using R in RStudio*. Youtube. <https://www.youtube.com/watch?v=AhZ42vSmDmE&t=1s>