

# First Year Exam Question 10

Julia Napoli

7/9/2022

PID: A59010574

## Question 10

First, we must download the most updated covid-19 data from the website provided by Barry.

Now, let's read it into the workspace.

```
covid_data <- read.csv("covid-19-variant-data.csv")
```

And now we can work on plotting the data. Let's get the ggplot, lubridate and dplyr libraries uploaded.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(lubridate, warn.conflicts = FALSE)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
```

```
## 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
```

Now we can think about plotting the data. First, let's erase the "total" and "other" rows that we don't need.

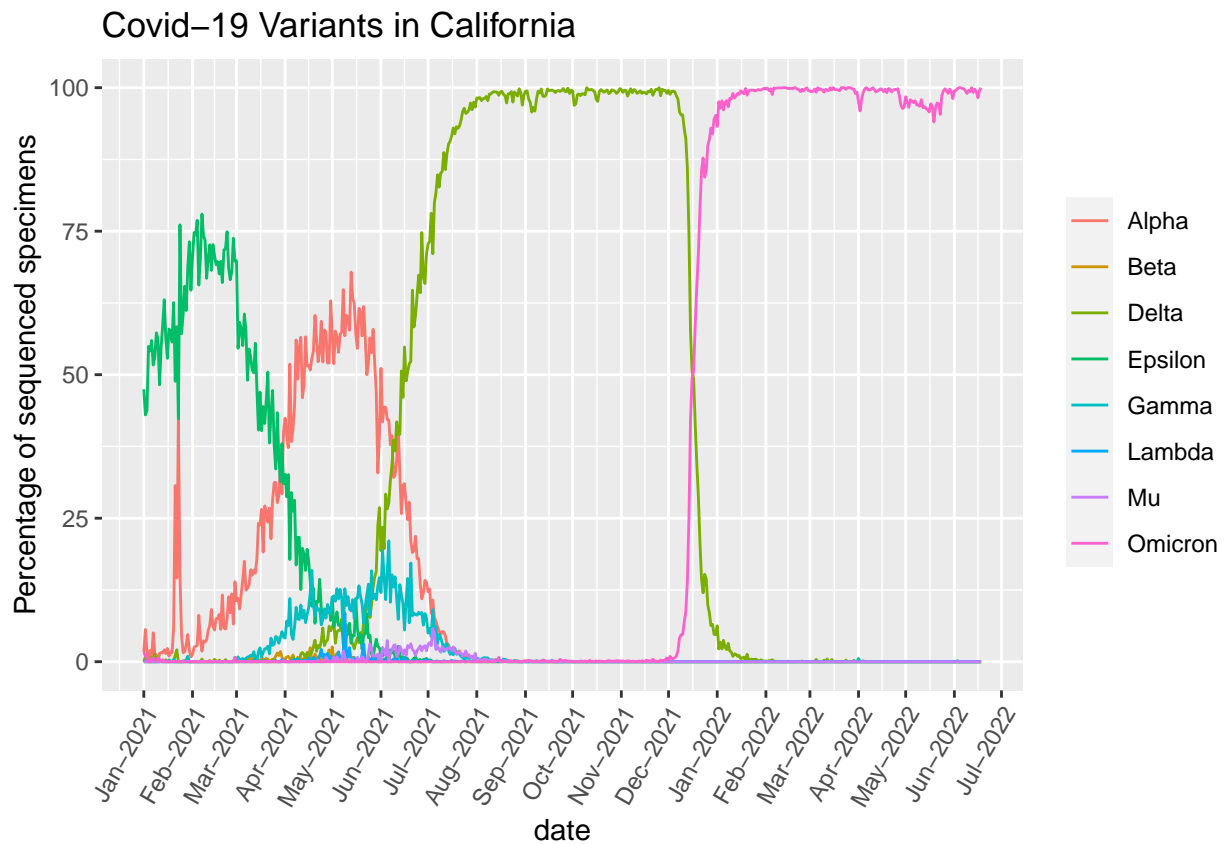
```
# erase the "total" rows
no_total_covid_data <- covid_data %>% filter(!grepl('Total', variant_name))

# erase the "other" rows
new_covid_data <- no_total_covid_data %>% filter(!grepl('Other', variant_name))
```

Now let's convert the date data into date variables in R.

```
new_covid_data$date <- as.POSIXct(new_covid_data$date)
```

```
covid_plot <- ggplot(data=new_covid_data, aes(x=date, y=percentage, group=variant_name, colour = variant_name)) +  
  geom_line() + theme(legend.title=element_blank()) + scale_x_datetime(date_breaks = '1 month', date_labels = '%b-%Y')  
print(covid_plot + ggtitle("Covid-19 Variants in California") + labs(y = "Percentage of sequenced specimens"))
```



All done!