

First Year Exam (Bioinfo)

Jibin Zhang (PID: A53300326)

2022/07/07

Import data

```
var <- read.csv( "covid19_variants.csv" )  
head(var)
```

```
##      date      area area_type variant_name specimens percentage  
## 1 2021-01-01 California      State      Gamma          0         0.00  
## 2 2021-01-01 California      State      Beta           0         0.00  
## 3 2021-01-01 California      State     Lambda          0         0.00  
## 4 2021-01-01 California      State     Alpha          1         1.69  
## 5 2021-01-01 California      State    Epsilon         28        47.46  
## 6 2021-01-01 California      State     Other         29        49.15  
## specimens_7d_avg percentage_7d_avg  
## 1              NA              NA  
## 2              NA              NA  
## 3              NA              NA  
## 4              NA              NA  
## 5              NA              NA  
## 6              NA              NA
```

Notice that one of these column is a date column. Working with time and dates get's annoying quickly. We can use the **lubridate** package to make this easy

```
library("lubridate")
```

I will make the as of date column Data format

```
var$as_of_date <- ymd(var$date)
```

Use the **filter()** function to do Subsetting

```
library("dplyr")
```

```
var_in_ca <- filter(var, variant_name != "Total"& variant_name != "Other")
```

Use the ggplot() function to plot the data

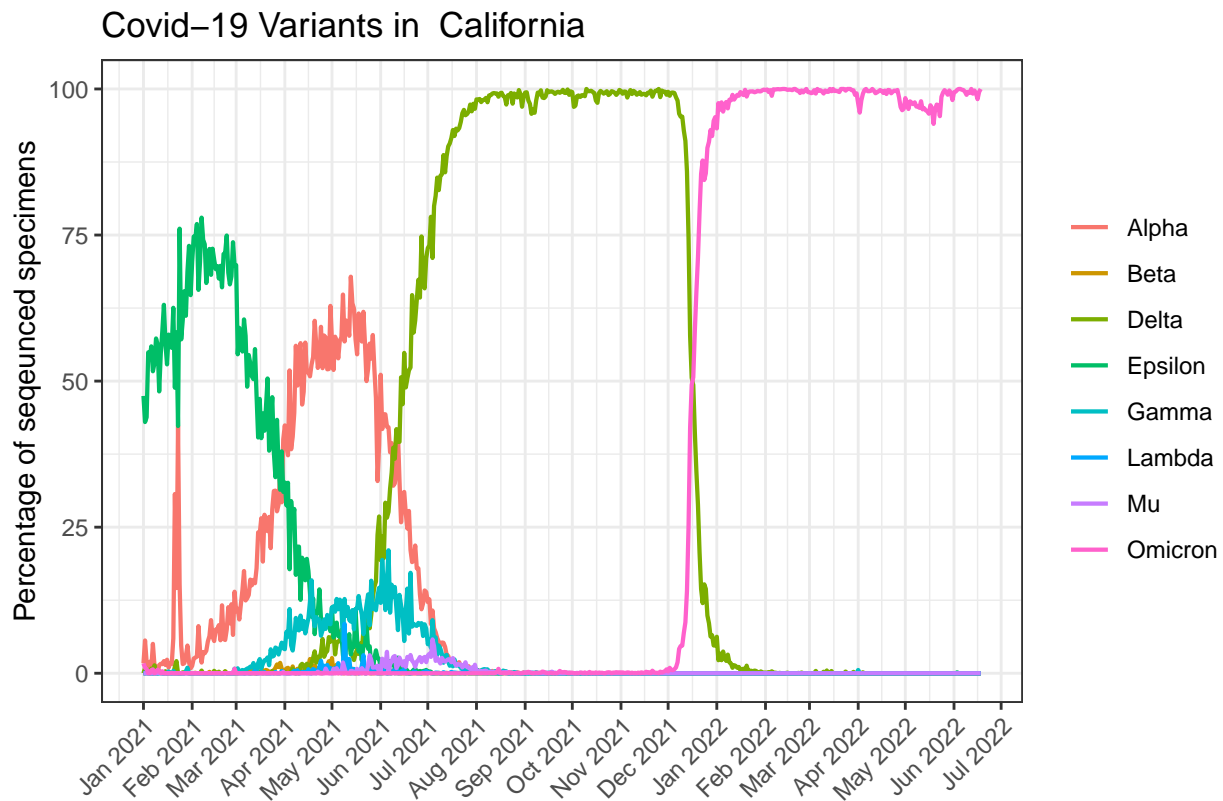
```
library("ggplot2")
```

Set the data to English format

```
Sys.setlocale(locale = "English")
```

```
## [1] "LC_COLLATE=English_United States.1252;LC_CTYPE=English_United States.1252;LC_MONETARY=English_U
```

```
ggplot(var_in_ca) +
  aes(as_of_date,
      percentage, colour = variant_name) +
  geom_line(size=0.75) +
  labs(x=NULL, y="Percentage of sequeunced specimens",
       title="Covid-19 Variants in California",
       caption="Data Source: <https://data.chhs.ca.gov/>")+
  scale_x_date(date_breaks = "1 month", date_labels = " %b %Y")+
  theme_bw()+
  theme(axis.text.x= element_text( angle = 45, hjust = 1),
        legend.title = element_blank())
```



Data Source: <https://data.chhs.ca.gov/>