

First Year Exam - Bioinformatics

Eric Jordahl (A59109089)

Process the Data

Read the Data

First we will read the data frame and output the first 6 lines to check that the data looks correct.

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

	date	area	area_type	variant_name	specimens	percentage
1	2021-01-01	California	State	Omicron	1	1.67
2	2021-01-01	California	State	Mu	0	0.00
3	2021-01-01	California	State	Gamma	0	0.00
4	2021-01-01	California	State	Epsilon	29	48.33
5	2021-01-01	California	State	Other	29	48.33
6	2021-01-01	California	State	Total	60	100.00

	specimens_7d_avg	percentage_7d_avg
1	NA	NA
2	NA	NA
3	NA	NA
4	NA	NA
5	NA	NA
6	NA	NA

Filter out Variants

Then we will to work with the data by removing variant categories that we do not want, like Other and Total.

```
library(ggplot2)
library(dplyr)
```

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

```
filtered_c <- c %>% filter(variant_name != "Other" &
                           variant_name != "Total")
```

Change Date Format

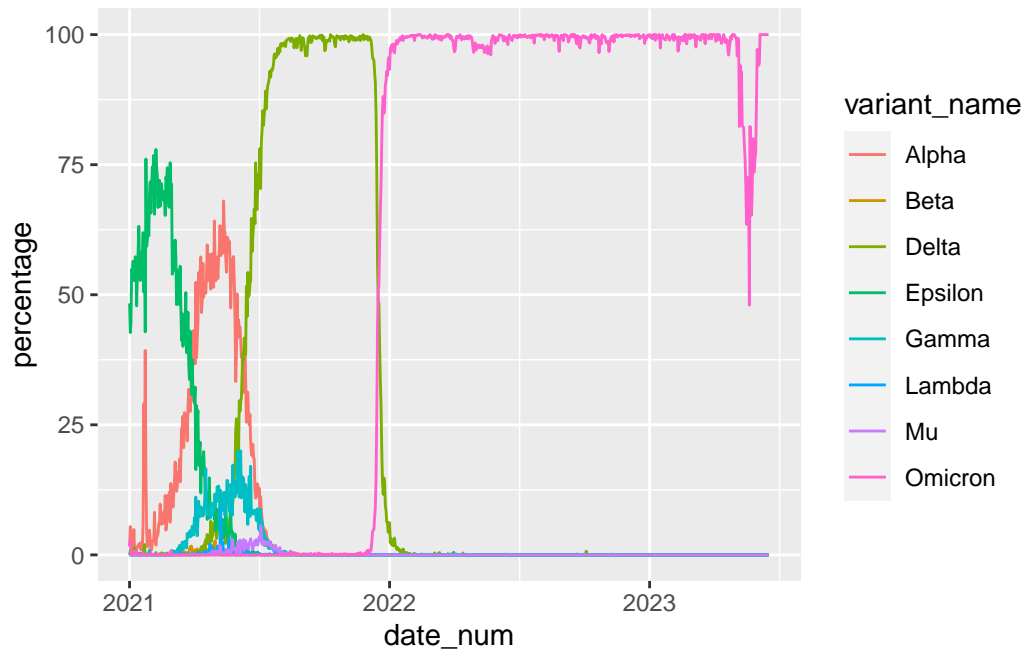
We will then change the date column to be a numerical variable rather than a character one

```
date_num <- as.Date(filtered_c$date)
```

Plot the Data

Then we will plot the line graph for the first time with date on the x and percentage on the y.

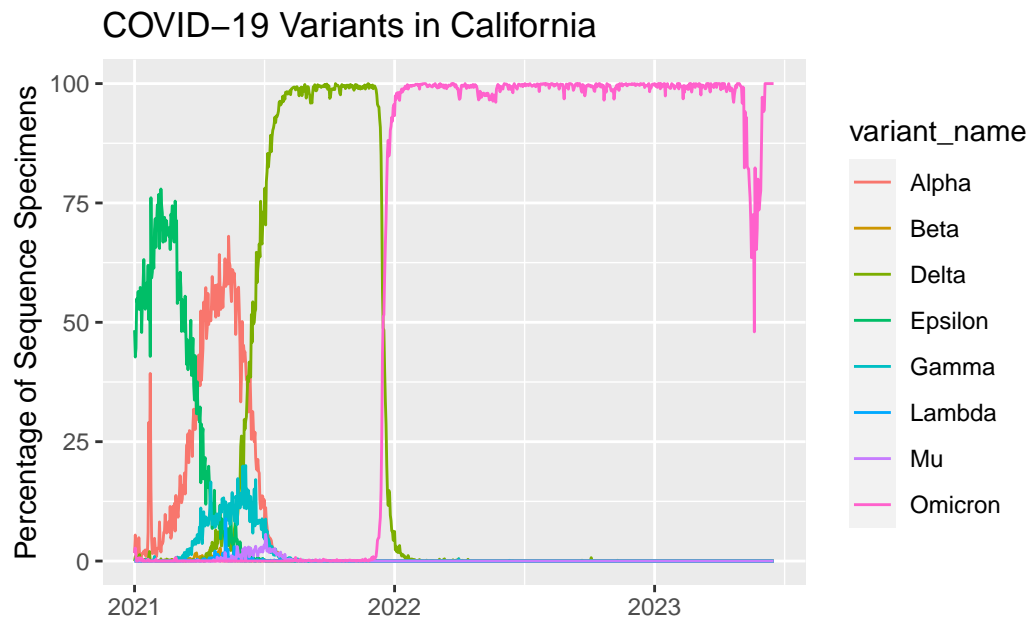
```
plot1 <- ggplot(data=filtered_c,
                aes(x=date_num, y=percentage,
                    col=variant_name)) +
  geom_line()
plot1
```



Add Axis Labels

Now we will add a title and correct the axis labels.

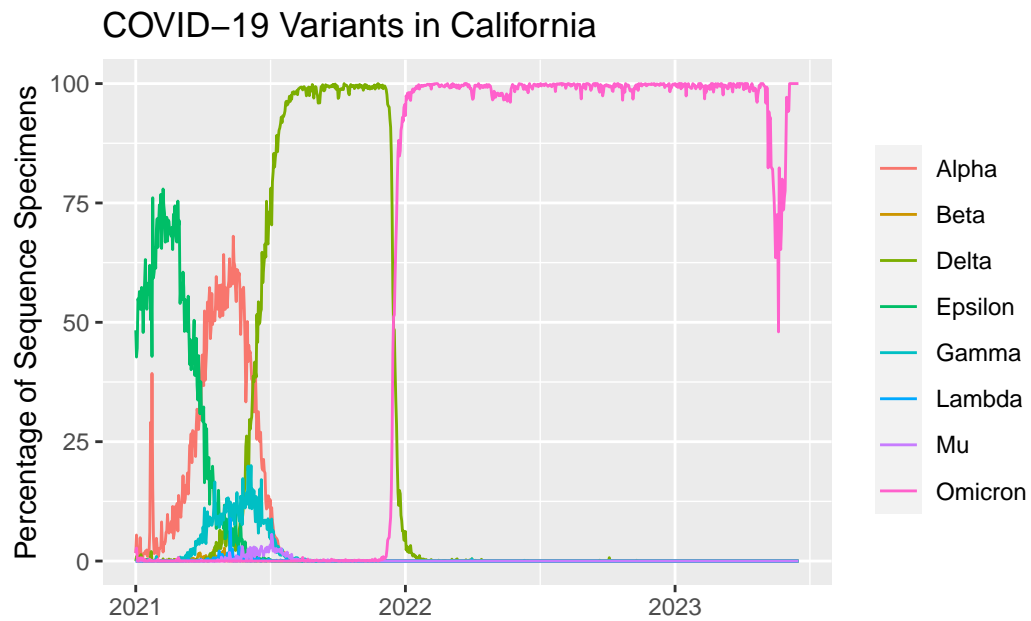
```
plot2 <- plot1 + labs(y="Percentage of Sequence Specimens",
                      x=" ",
                      title = "COVID-19 Variants in California")
plot2
```



Remove Legend Title

We will then remove the legend title

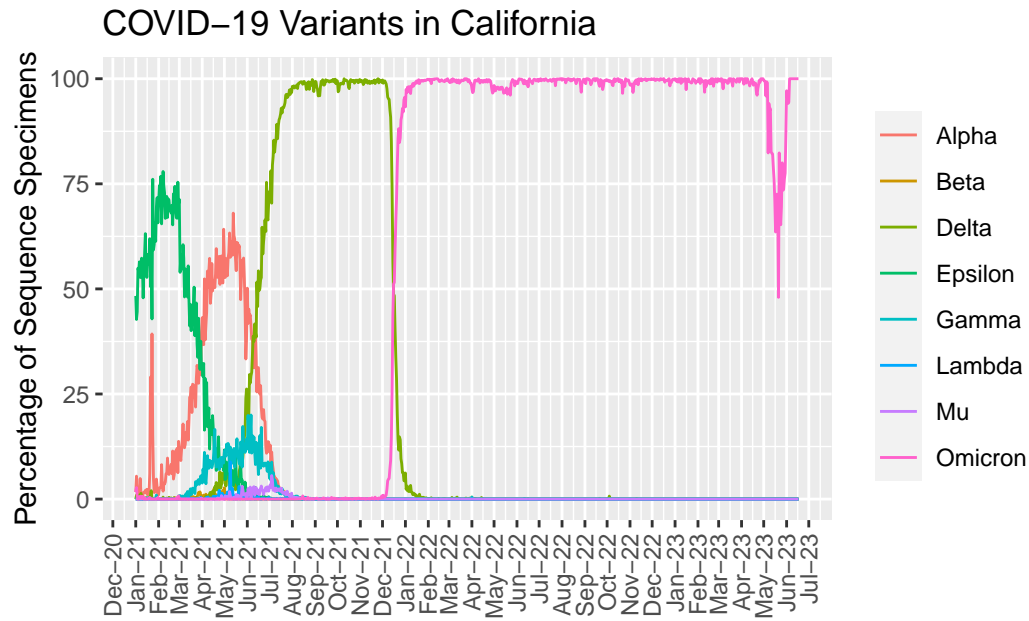
```
plot3 <- plot2 + theme(legend.title=element_blank())
plot3
```



Adjust X Axis Breaks and Labeling

Now we will adjust the x axis to show each month not years

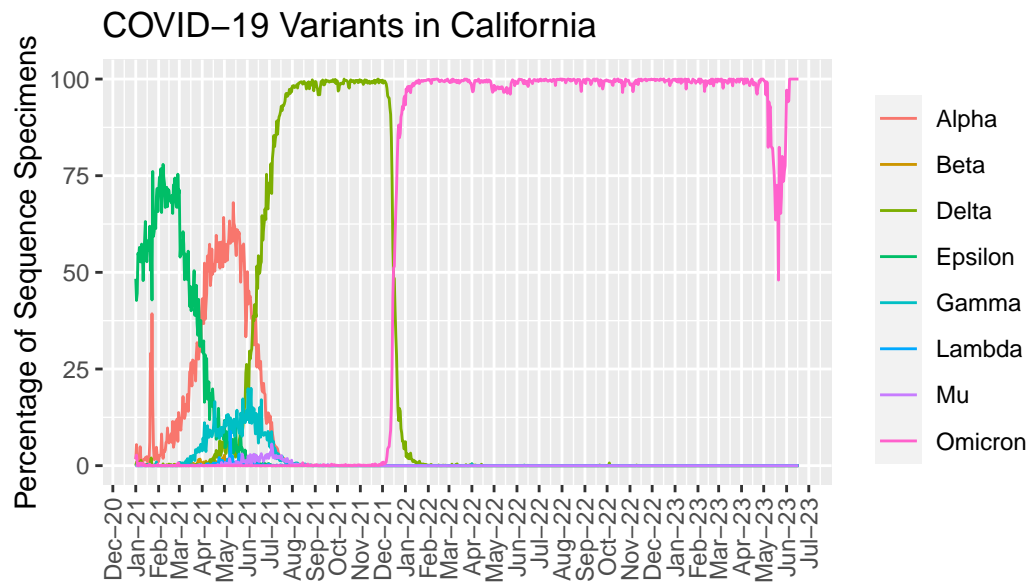
```
plot4 <- plot3 + scale_x_date(date_breaks="1 month", date_labels="%b-%y") +  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))  
plot4
```



Add data source

Now we will add a caption with where the data was obtained from

```
plot5 <- plot4 +
  labs(caption="Data Source:https://data.chhs.ca.gov/") +
  theme(plot.caption=element_text(hjust=0.5))
plot5
```



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