

COVID-19 Variant Analysis

AUTHOR

Shane Desfor - PID: A59018963

Step-By-Step

#Read in and examine csv file.

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

	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

#Going to need ggplot for this.

```
library(ggplot2)
```

#...Realized waaay down below that this needed to change. And it needed to be here because of the way I had layered in date.

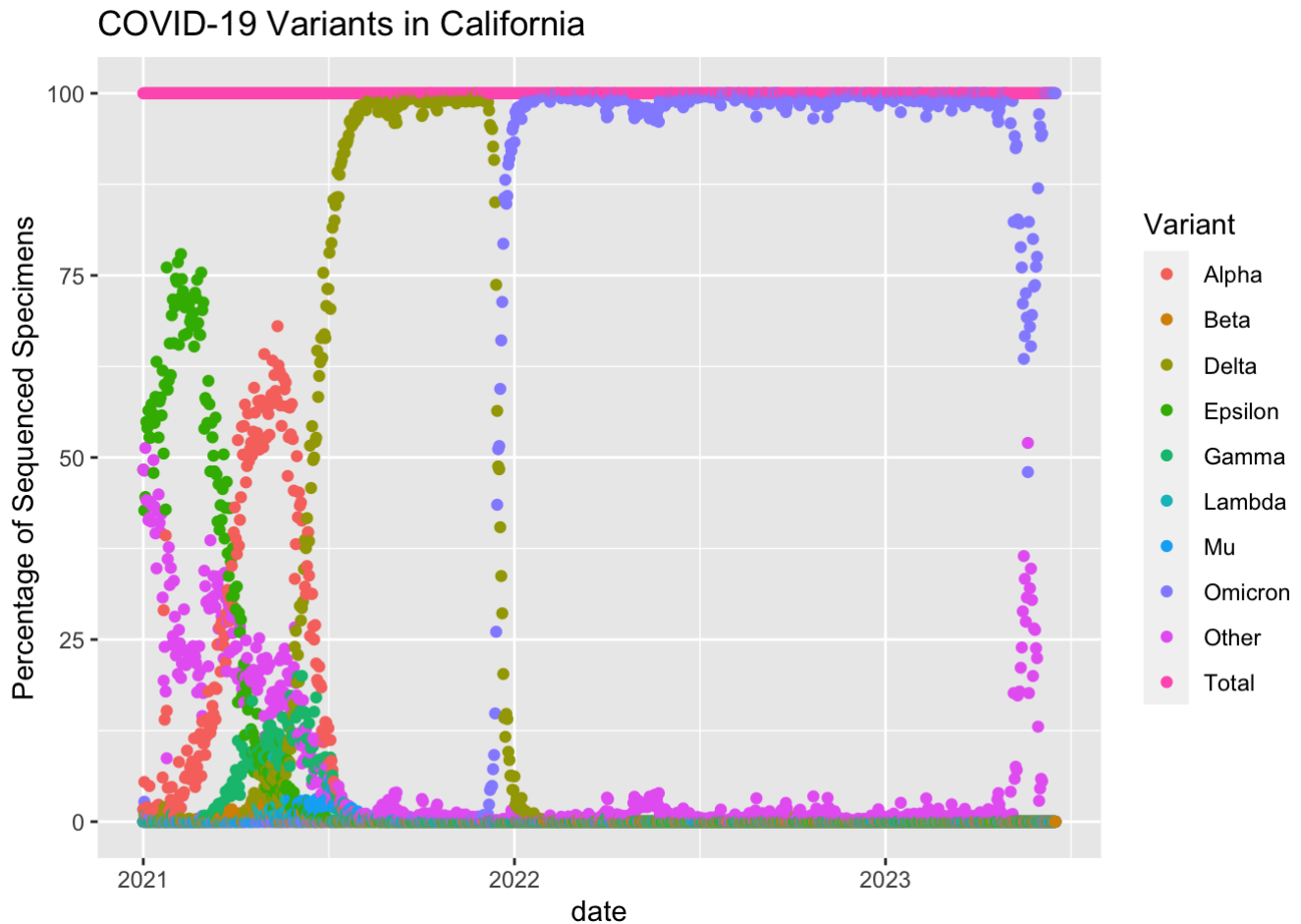
```
covid$date <- as.Date(covid$date, format = "%Y-%m-%d")
```

#After much tweeking, I got a workerable plot to play with, but it needs some tuning in.

```
covid_plot <- ggplot(covid,
                     aes(x=date,
                         y=percentage,
                         color=variant_name))

covid_plot +
  geom_point() +
  labs(title = "COVID-19 Variants in California",
```

```
y= "Percentage of Sequenced Specimens",
color="Variant")
```



#First, remove the 'Total' from being plotted

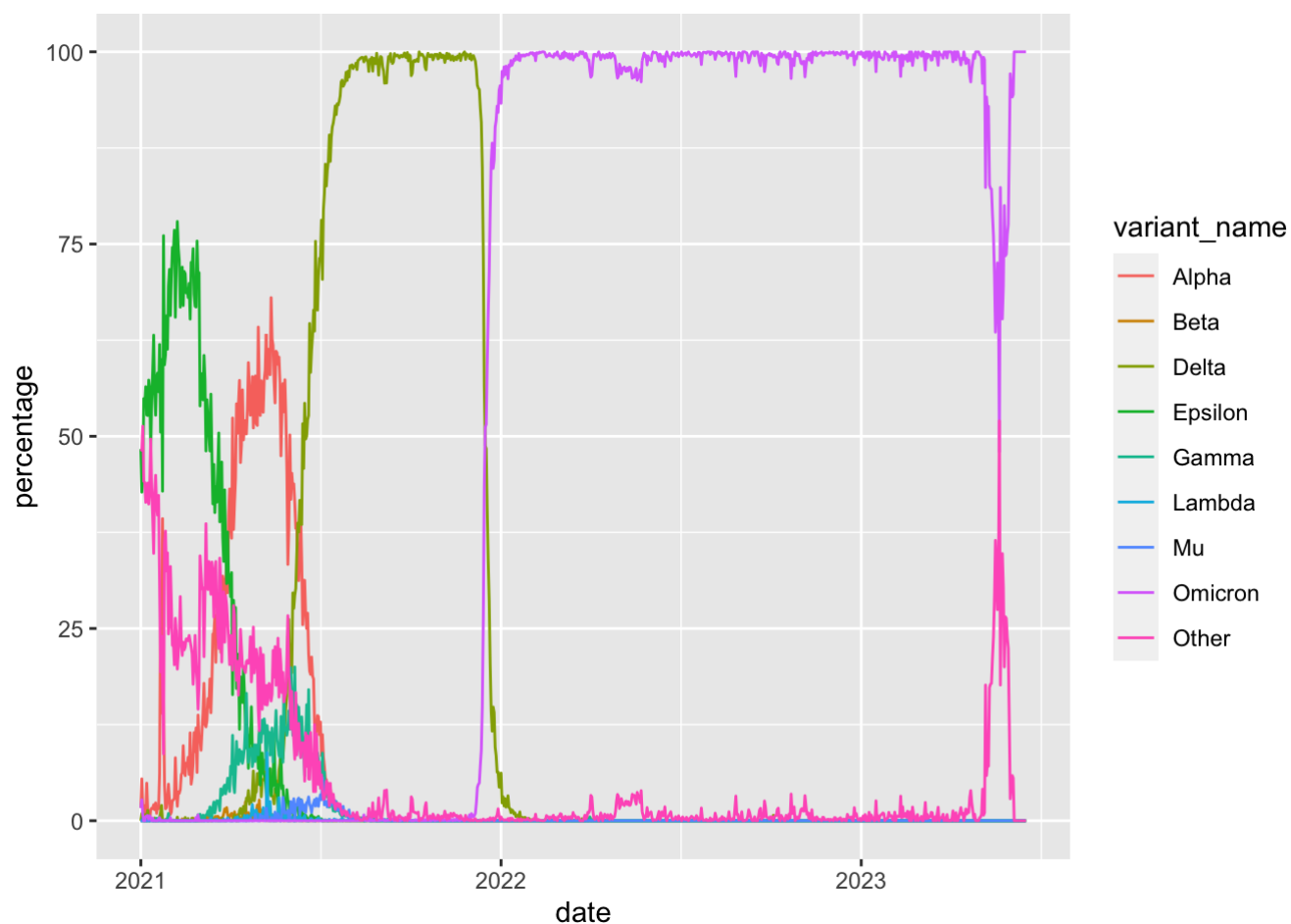
```
covid_filtered <- covid[covid$variant_name != "Total", ]
```

#Could not get geom_line to work, would not display any variants, going to try and group by variant, as a factor.

```
covid_filtered$variant_name <- factor(covid_filtered$variant_name)
```

#Adjusted line plot

```
covid_plot2 <- ggplot(covid_filtered,
  aes(x = date,
      y = percentage,
      color = variant_name,
      group = variant_name)) +
  geom_line()
covid_plot2
```

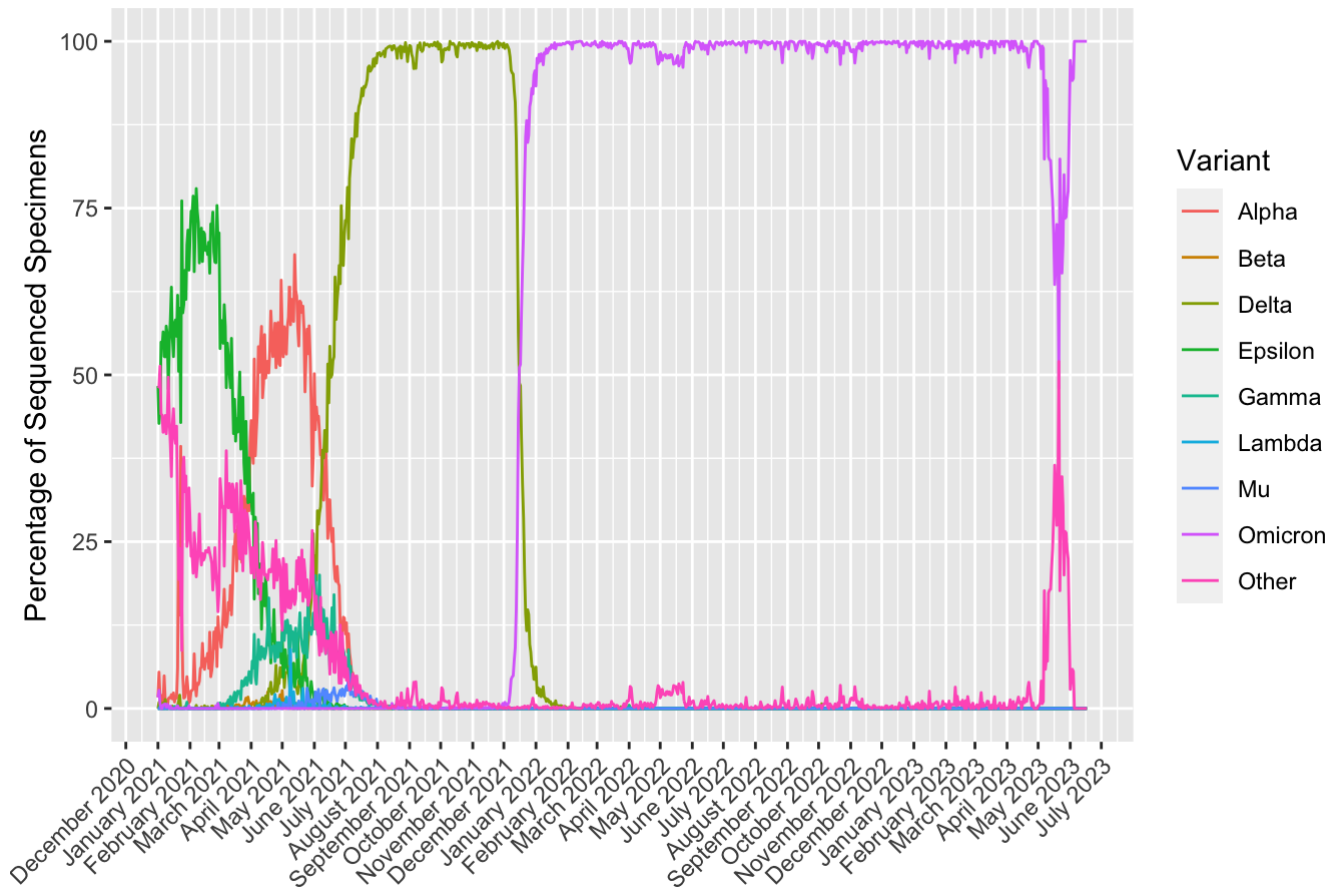


#Worked! But it's a little sloppy.

#Added labels and adjusted the x-axis...still needs some more tweeking though.

```
covid_plot2 +  
  labs(title = "COVID-19 Variants in California",  
        x= NULL,  
        y= "Percentage of Sequenced Specimens",  
        color="Variant") +  
  scale_x_date(date_labels = "%B %Y", date_breaks = "1 month") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

COVID-19 Variants in California



#Get better x-axis starting and ending points, also discovered I needed a 'breaks vector'

```
start_date <- min(covid_filtered$date)
end_date <- max(covid_filtered$date)
breaks.vec <- seq(min(covid_filtered$date),
                  max(covid_filtered$date), by= "1 month")
```

##Final Product #This is the one!

```
covid_plot2 +
  labs(title = "COVID-19 Variants in California",
       x= NULL, y= "Percentage of Sequenced Specimens",
       color="Variant") +
  scale_x_date(breaks = breaks.vec, date_labels = "%B %y") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

COVID-19 Variants in California

