Covid-19 Variant Data

Sam Altshuler

7/14/2022

Load in correct libraries

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

Read in the data from a csv. Data downloaded from the California Health and Human Services website.

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

```
##          date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Alpha         1       1.69
## 2 2021-01-01 California     State         Beta         0       0.00
## 3 2021-01-01 California     State           Mu         0       0.00
## 4 2021-01-01 California     State        Gamma         0       0.00
## 5 2021-01-01 California     State        Total        59     100.00
## 6 2021-01-01 California     State      Omicron         1       1.69
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

Next we can preliminarily graph the data to see what it looks like.

```
ggplot(total_data, aes(x = date, y = percentage, group = variant_name, color = variant_name))+
  geom_line()
```
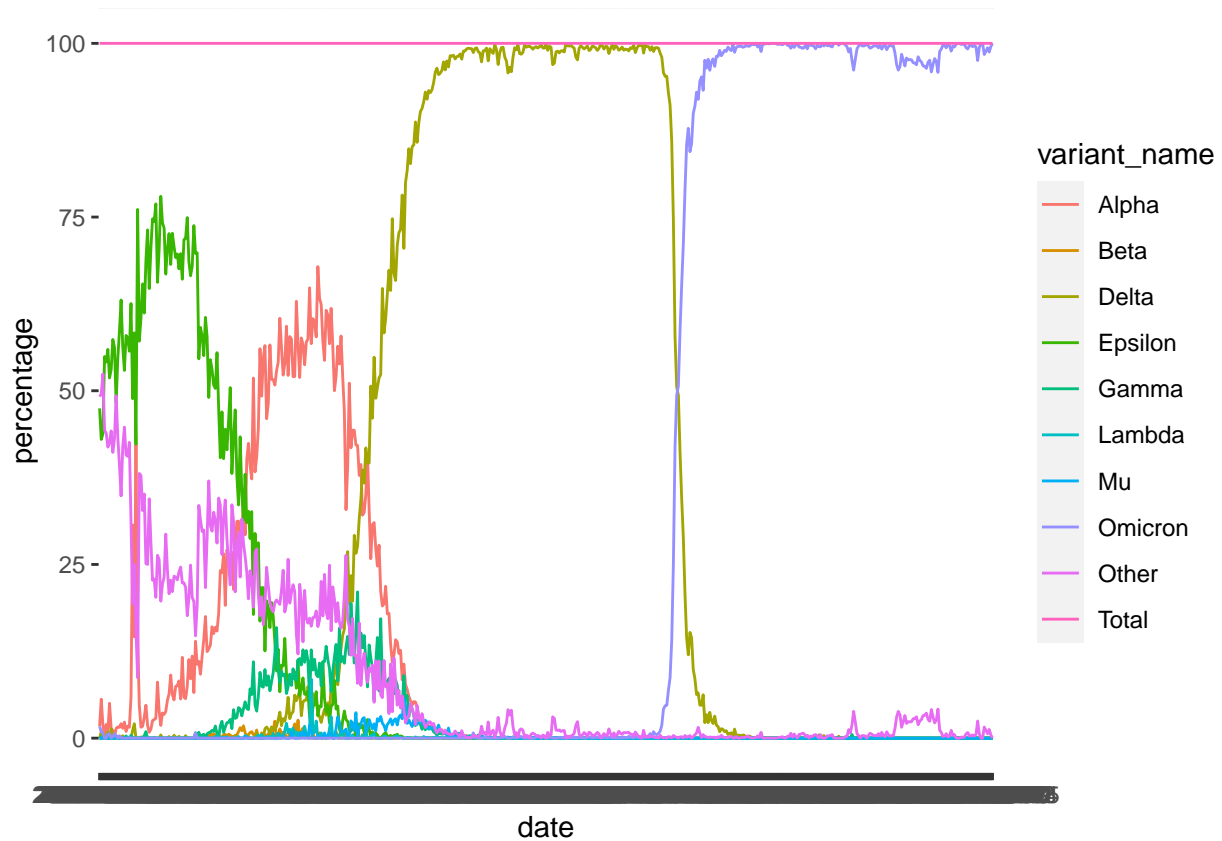
Now this looks similar to example graph, but the next step is to remove all rows that have "Total" variants as this row will always be 100%.

```r
variants <- total_data %>% filter(variant_name != "Total")
ggplot(variants, aes(x = date, y = percentage, group = variant_name, color = variant_name))+
  geom_line()
```
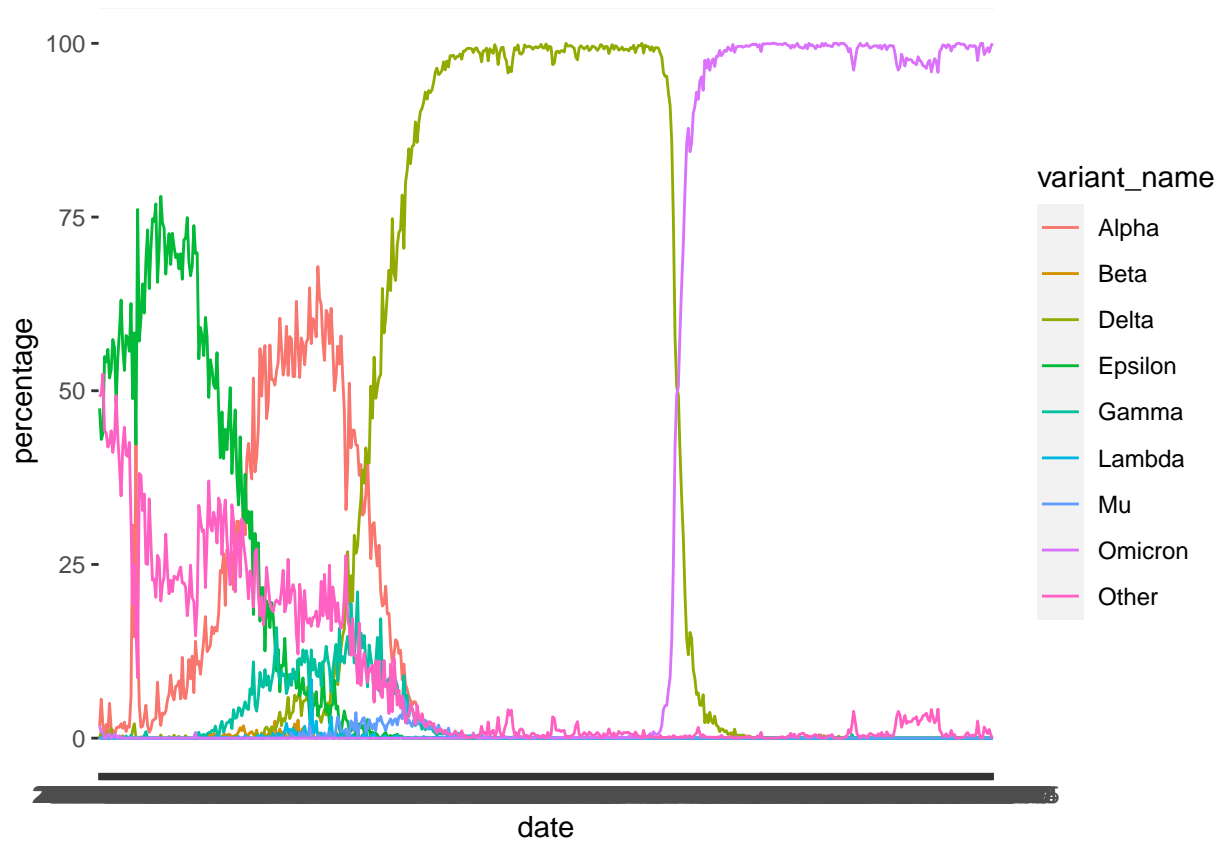
Next we want to change the x axis to show the 3 letters of the month with the year. This is done by using the lubridate package.

```
variants$date <- as_date(variants$date)
ggplot(variants, aes(x = date, y = percentage, group = variant_name, color = variant_name))+
  geom_line()+
  scale_x_date(date_breaks = "1 month", date_labels = "%b-%Y")
```
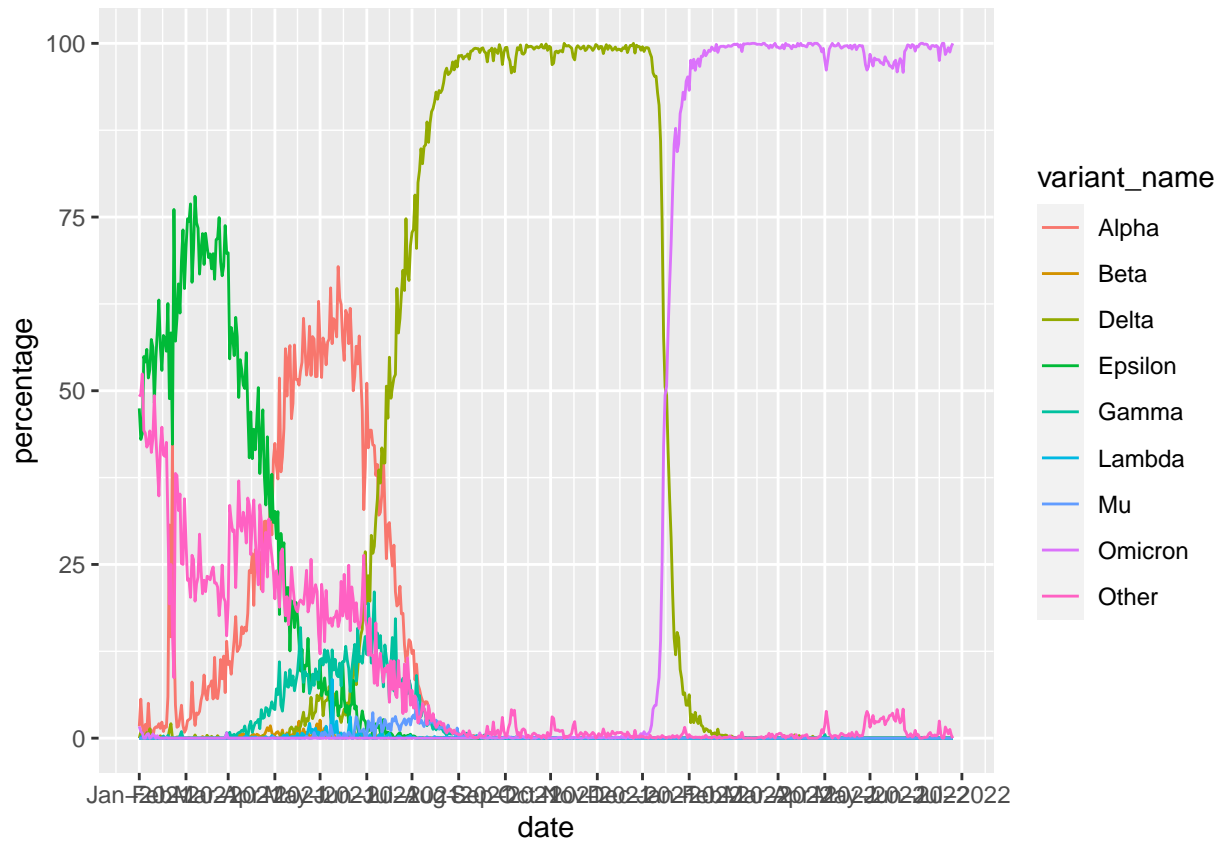
Now that our graph now has the correct axes, let's make it more readable.

```r
ggplot(variants, aes(x = date, y = percentage, group = variant_name, color = variant_name))+
  geom_line()+
  scale_x_date(date_breaks = "1 month", date_labels = "%b-%Y")+
  ylab("Percentage of sequenced specimens")+
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust=1), axis.title.x = element_blank())
```