FYE_COVID19variants

Gabrielle Meza (A13747395)

7/13/2022

Initial Set up

Load Packages

Removed messages for loading to make Rmarkdown cleaner (message=FALSE)

Loaded RcolorBrewer for last step to changing colors

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

Import data

Import COVID 19 variant data. Use head to get sense of data rows/columns

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

```
##         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
```

Modify dataset

Lubridate and filter data

Use lubridate to specify that the date column in data set will use year-month-day format.

Using data$colun name != (does not equal total) to get new dataset that does not include total values for each date. Can also use dyplr and » % filtering if data set is more complex/ want to filter out specfic parts in columns or rows

```
data$date <- ymd(data$date)

clean.data <- data[data$variant_name != "Total", ]
```

Using GGplot

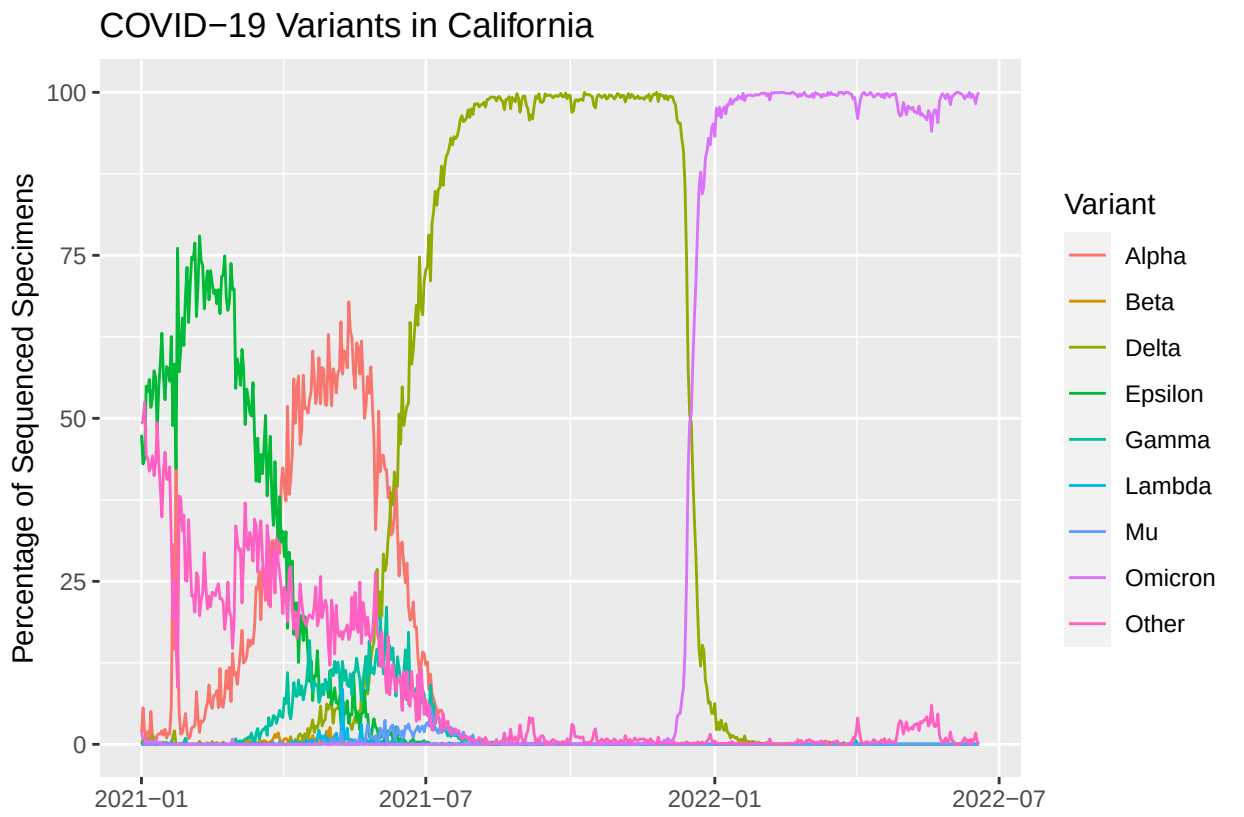
Getting started

Use ggplot to make initial line graph of % of each variant over time, and modified labels.

Assign initial plot to variable so later modifications are easier

```
plot <-ggplot(clean.data, aes(x=date, y=percentage, colour = variant_name)) +
  geom_line() +
  labs (x= "",
        y="Percentage of Sequenced Specimens",
        colour="Variant",
        title = "COVID-19 Variants in California")

plot
```

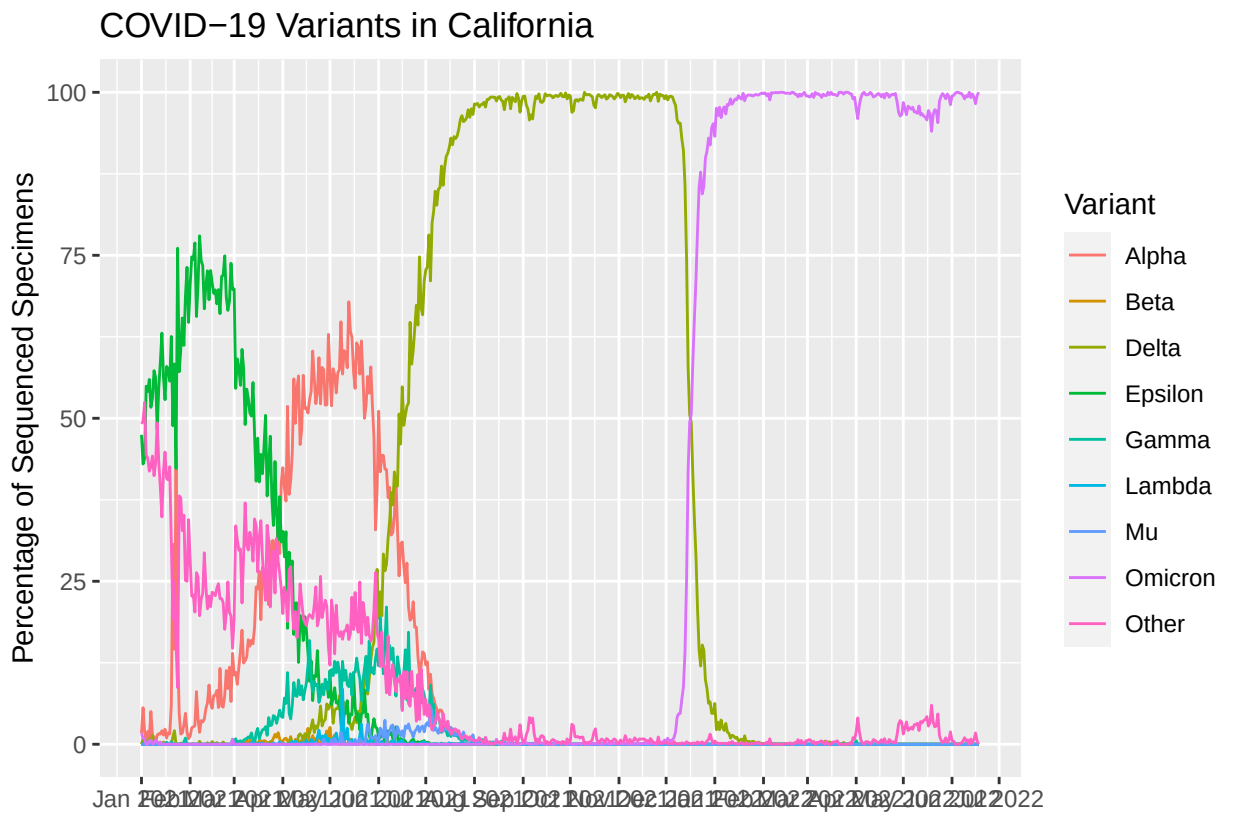


Plot Modifications

Format dates

Assigned 1 month interval for x-axis, and used dyplr to put in format Month, year. then view graph to ensure it worked. Assign to new variable to make next modifications easier.

```
p <- plot +
  scale_x_date(date_breaks = "1 month", date_labels =  "%b %Y")
p
```



Visual edits

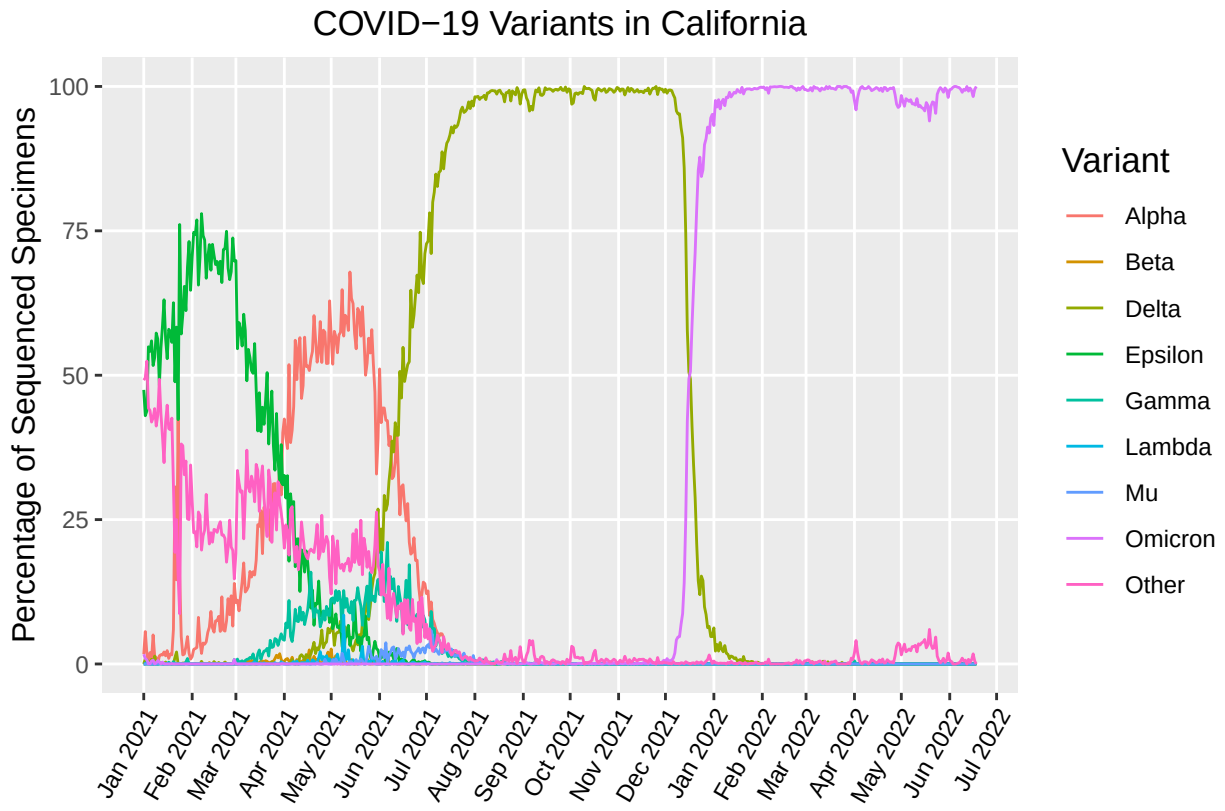
Changing plot to make it look nice.

Edits made to address the following (in order):

- angle/ dates
- Remove minor axis lines (looks less cluttered)
- Removed legend background
- Center graph title
- Resize text of axis (Bigger)

- Resize Legend title (Bigger)

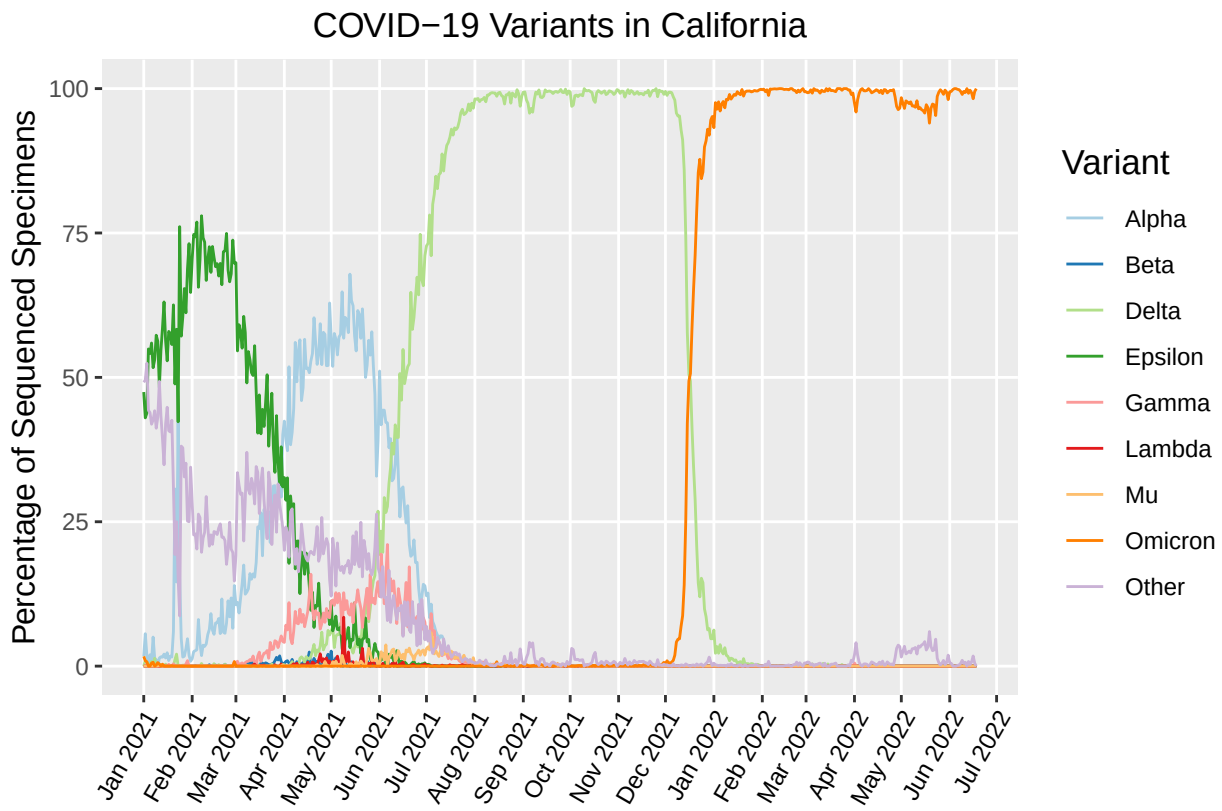
```
p1 <-p +
  theme(axis.text.x=element_text(angle=60, hjust=1, colour = "black"),
        panel.grid.minor = element_blank(),
        legend.key= element_blank(),
        plot.title = element_text(hjust = 0.5),
        axis.title = element_text(size = "12"),
        legend.title = element_text(size = "13"))
p1
```



Custom colors

Finally, changing colors to be prettier using 'Paired' pallet from RcolorBrewer and assign final graph:)

```
p.final <- p1 + scale_color_brewer(palette = "Paired")
p.final
```


COVID−19 Variants in California
Percentage of Sequenced Specimens
100
75
50
25
0
Jan 2021
Feb 2021
Mar 2021
Apr 2021
May 2021
Jun 2021
Jul 2021
Aug 2021
Sep 2021
Oct 2021
Nov 2021
Dec 2021
Jan 2022
Feb 2022
Mar 2022
Apr 2022
May 2022
Jun 2022
Jul 2022
Variant
Alpha
Beta
Delta
Epsilon
Gamma
Lambda
Mu
Omicron
Other