FYE_COVID19variants

Gabrielle Meza (A13747395)

7/13/2022

```
#Loading packages needed:
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
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(RColorBrewer)
```

Import COVID 19 variant data

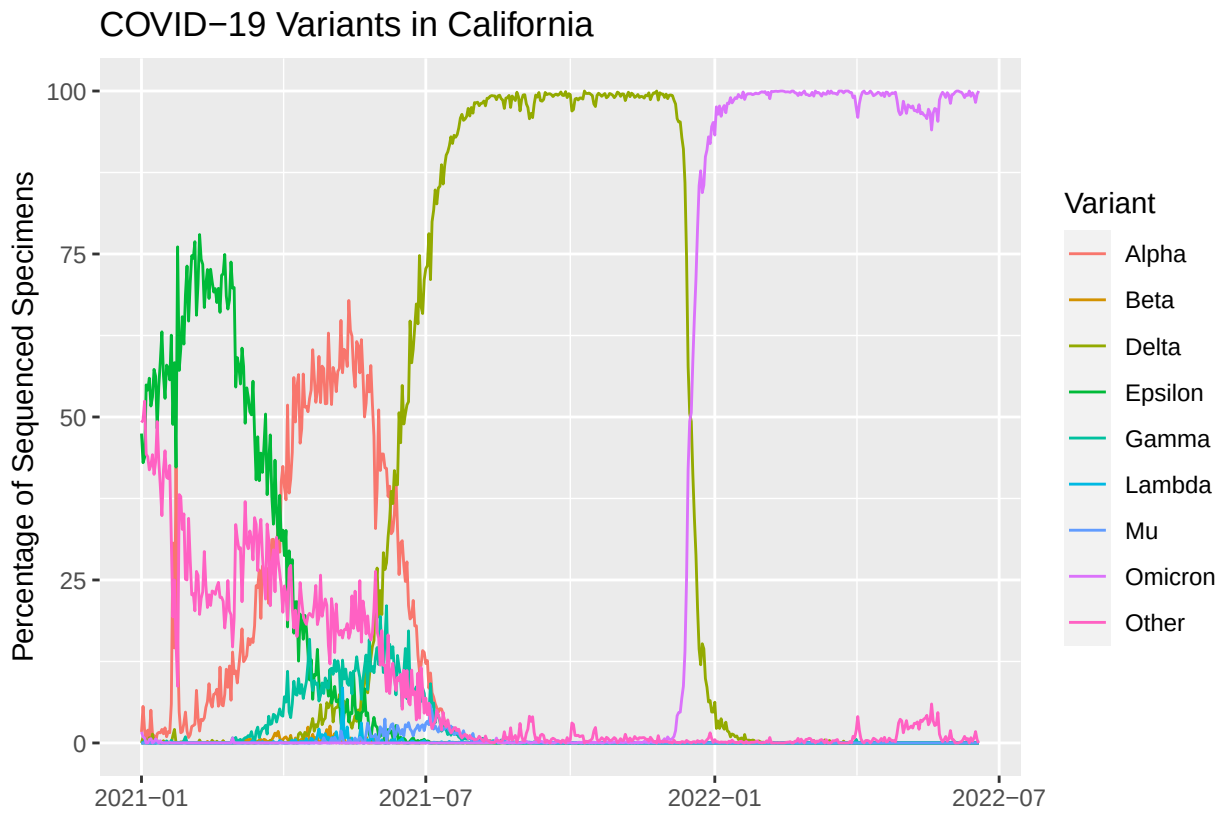
```
# Import COVID 19 variant data
data <- read.csv("covid19_variants.csv")
head(data)
```

```
##         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
```

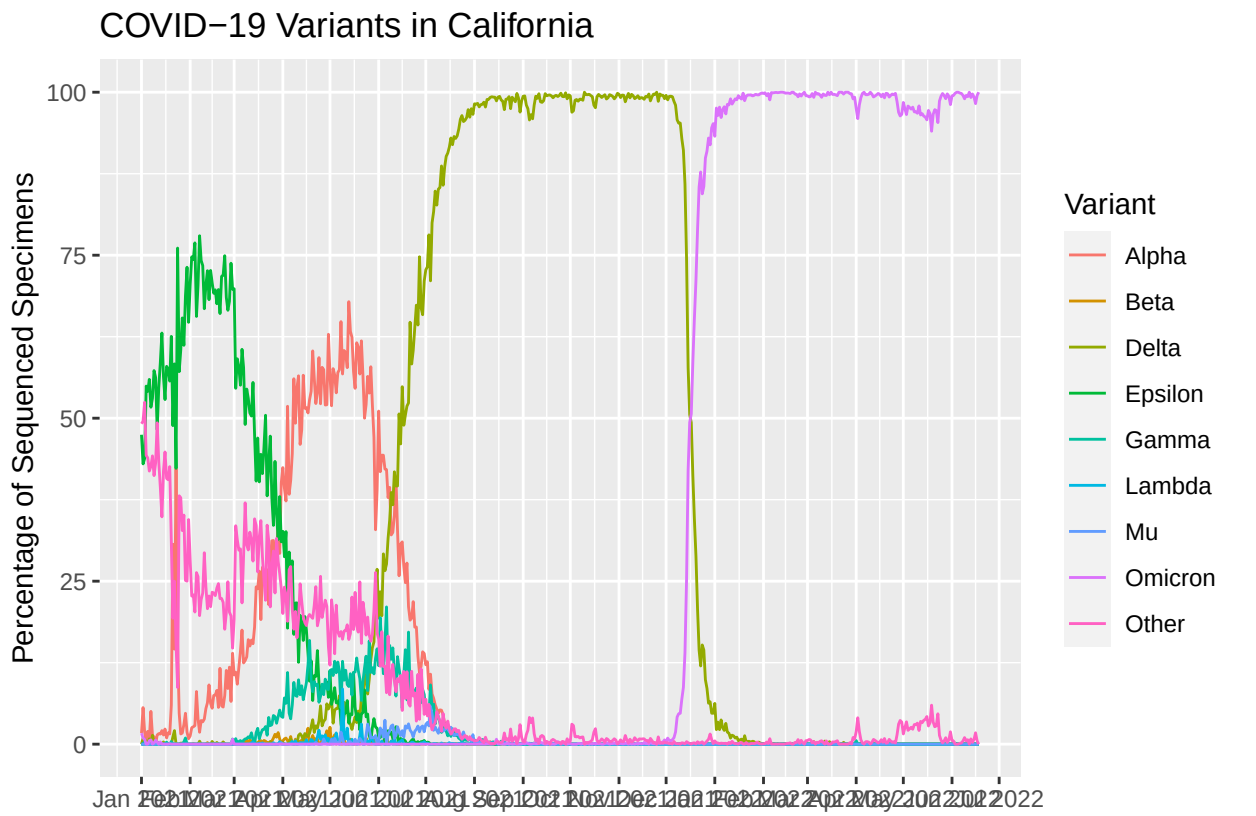
```
## 6 2021-01-01 California     State        Other        29      49.15
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

```
#Using lubridate to specficy that the date column in dataset will use year-month-day format

data$date <- ymd(data$date)

#Remove total values from data
clean.data <- data[data$variant_name != "Total", ]
```

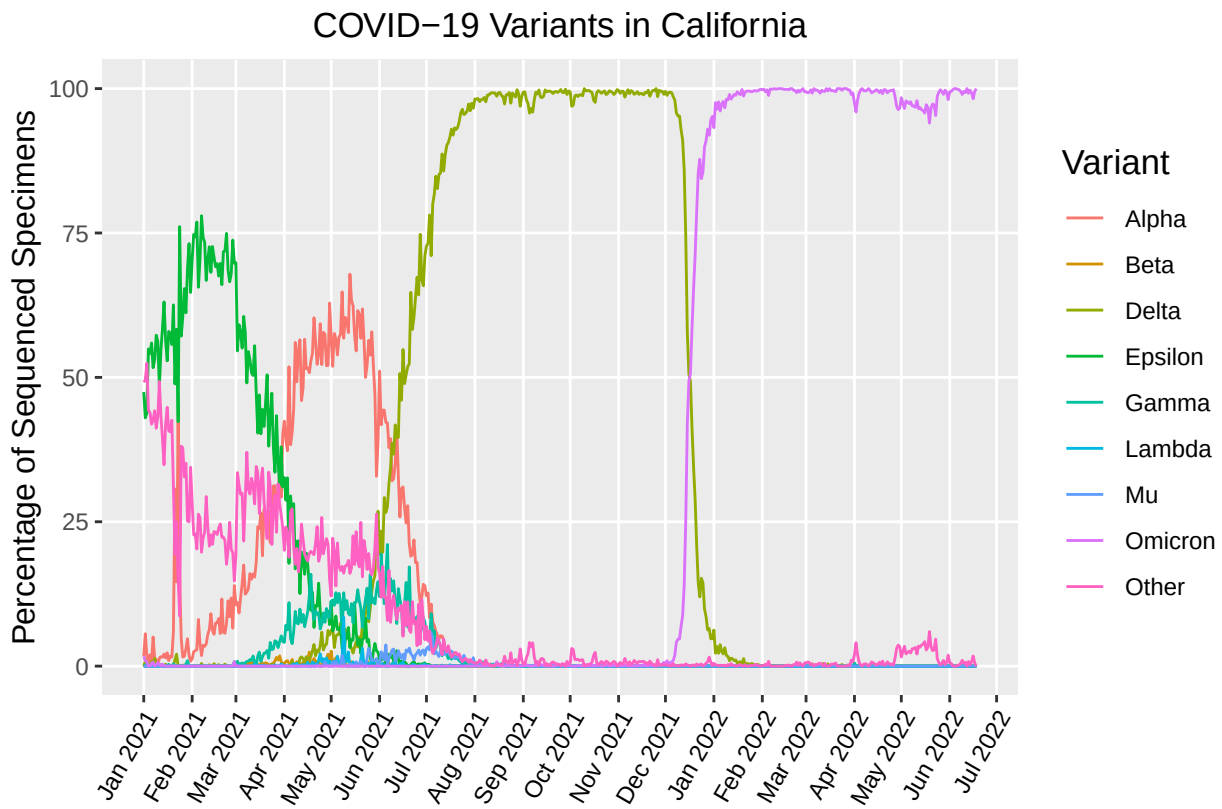
```
#Use ggplot to make initial line graph of % of each variant over time, and modified labels. Assign to v
plot <-ggplot(clean.data, aes(x=date, y=percentage, colour = variant_name)) +
  geom_line() +
  labs (x= "", y="Percentage of Sequenced Specimens", colour="Variant", title = "COVID-19 Variants in Ca

plot
```



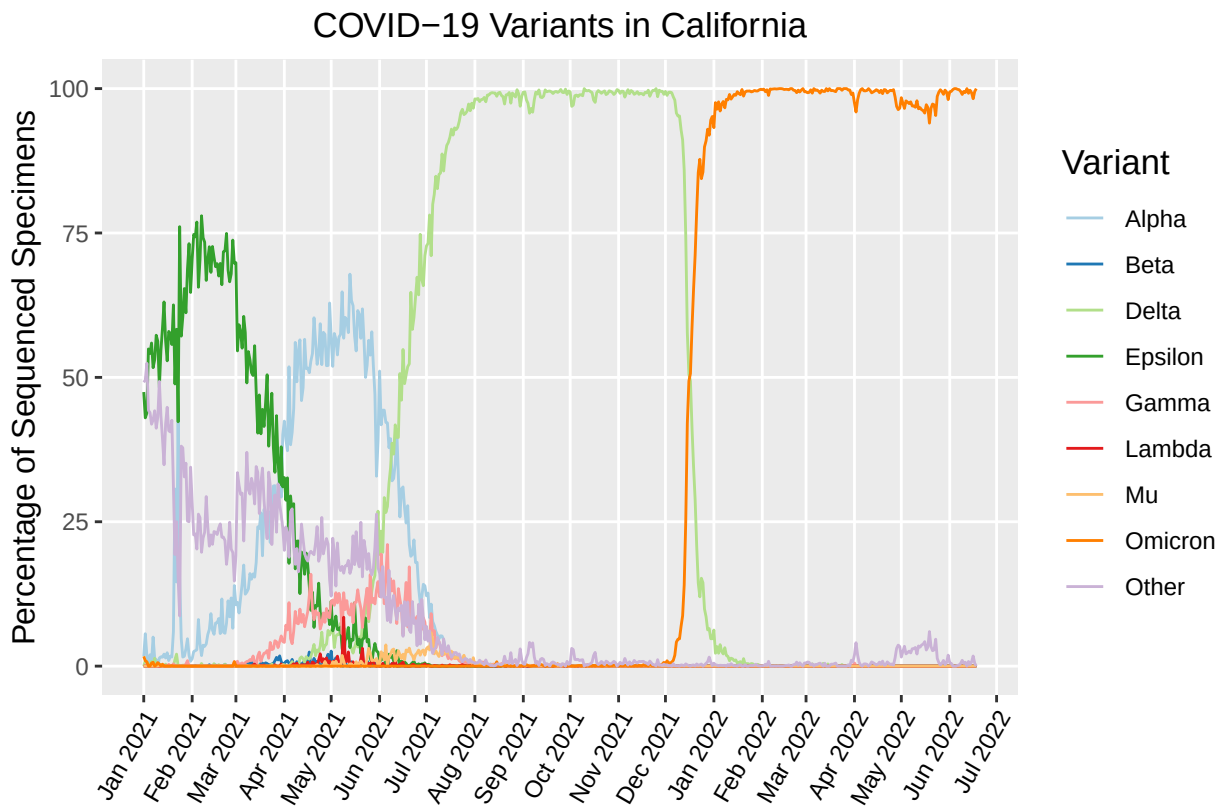
```
#Format dates using dyplr
p <- plot +
  scale_x_date(date_breaks = "1 month", date_labels =  "%b %Y")
p
```




```
#Start changing plot to make it look nice, edits in these to adress the following in order:
#angle/ dates , remove minor axis lines, remove ledgend background, center title, resize axis, resize l

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
```




```
#Finally, changing colors to be prettier using 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