fist-year-exam

May Wu PID:A59010588

6/20/2022

COVID-19 Variant Data

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## 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
```

```
covid = read.csv('d7f9acfa-b113-4cbc-9abc-91e707efc08a.csv')
head(covid)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State      Epsilon        28      47.46
## 2 2021-01-01 California     State        Other        29      49.15
## 3 2021-01-01 California     State        Alpha         1       1.69
## 4 2021-01-01 California     State         Beta         0       0.00
## 5 2021-01-01 California     State        Total        59     100.00
## 6 2021-01-01 California     State        Gamma         0       0.00
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

```r
data = filter(covid, variant_name!='Other' & variant_name!='Total')
head(data)
```

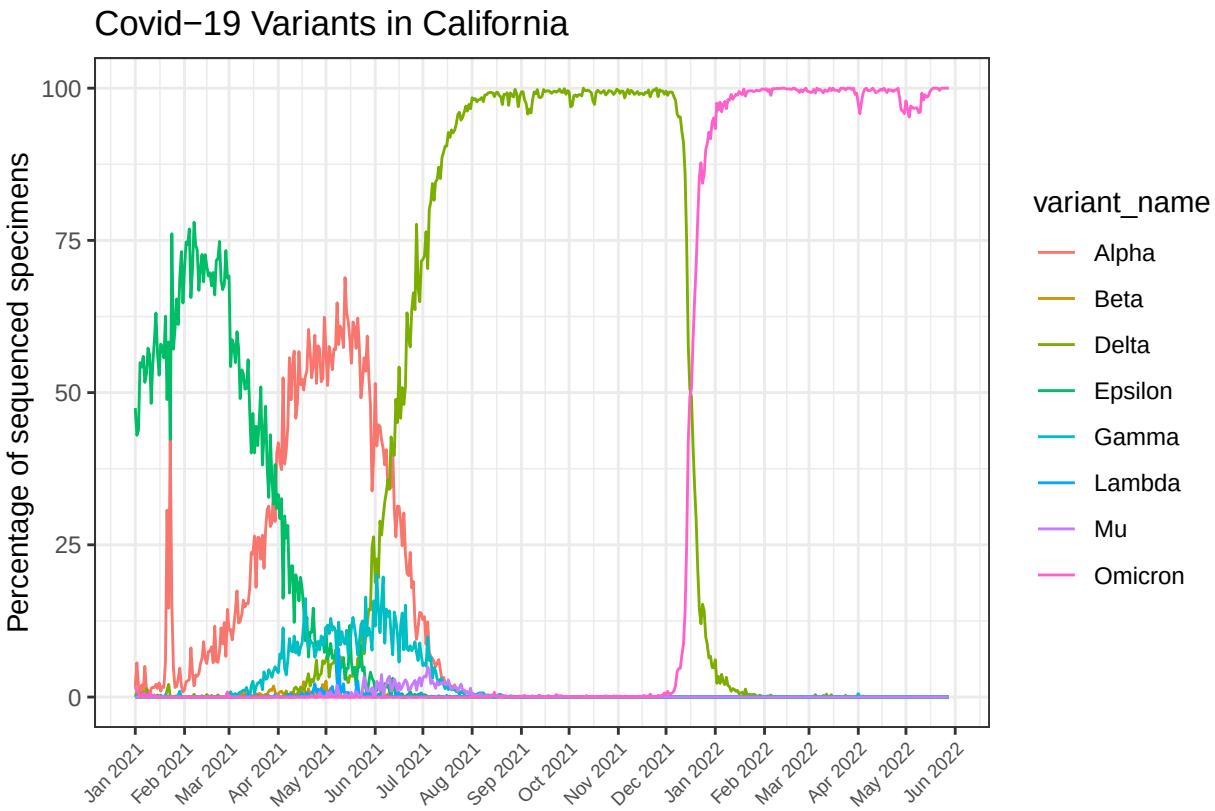
```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State      Epsilon        28      47.46
## 2 2021-01-01 California     State        Alpha         1       1.69
## 3 2021-01-01 California     State         Beta         0       0.00
## 4 2021-01-01 California     State        Gamma         0       0.00
## 5 2021-01-01 California     State           Mu         0       0.00
## 6 2021-01-01 California     State        Delta         0       0.00
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

```r
data['Date'] = as.Date(data$date, format = "%Y-%m-%d")
```

```r
table(data$area)
```

```
##
## California
##       4104
```

```r
ggplot(data=data, aes(x=Date,y=percentage,group=variant_name)) + geom_line(aes( color=variant_name)) +tl
scale_x_date(date_breaks = "1 month", date_labels =  "%b %Y") + labs(x = "", y = "Percentage of sequenc
```


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
variant_name
Alpha
Beta
Delta
Epsilon
Gamma
Lambda
Mu
Omicron