Olvera_yr1Exam

Jocelyn Olvera/PID:A12459601

7/18/2022

Import packages to be used in plotting COVID-19 data.

```
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
```

Read in csv file and inspect the first few lines.

```
# Read in csv
covid <- read.csv("../Bioinformatics_yr1_Exam/covid19_variants.csv")

# Inspect first 6 lines of the data
head(covid)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Total        59     100.00
## 2 2021-01-01 California     State       Lambda         0       0.00
## 3 2021-01-01 California     State        Delta         0       0.00
## 4 2021-01-01 California     State        Other        29      49.15
## 5 2021-01-01 California     State           Mu         0       0.00
## 6 2021-01-01 California     State         Beta         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
```

Remove columns with NA values.

```
covid <- select(covid, -specimens_7d_avg, -percentage_7d_avg)
head(covid)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Total        59     100.00
## 2 2021-01-01 California     State       Lambda         0       0.00
## 3 2021-01-01 California     State        Delta         0       0.00
## 4 2021-01-01 California     State        Other        29      49.15
## 5 2021-01-01 California     State           Mu         0       0.00
## 6 2021-01-01 California     State         Beta         0       0.00
```

Remove row that are not individual variants ("Other" and "Total")

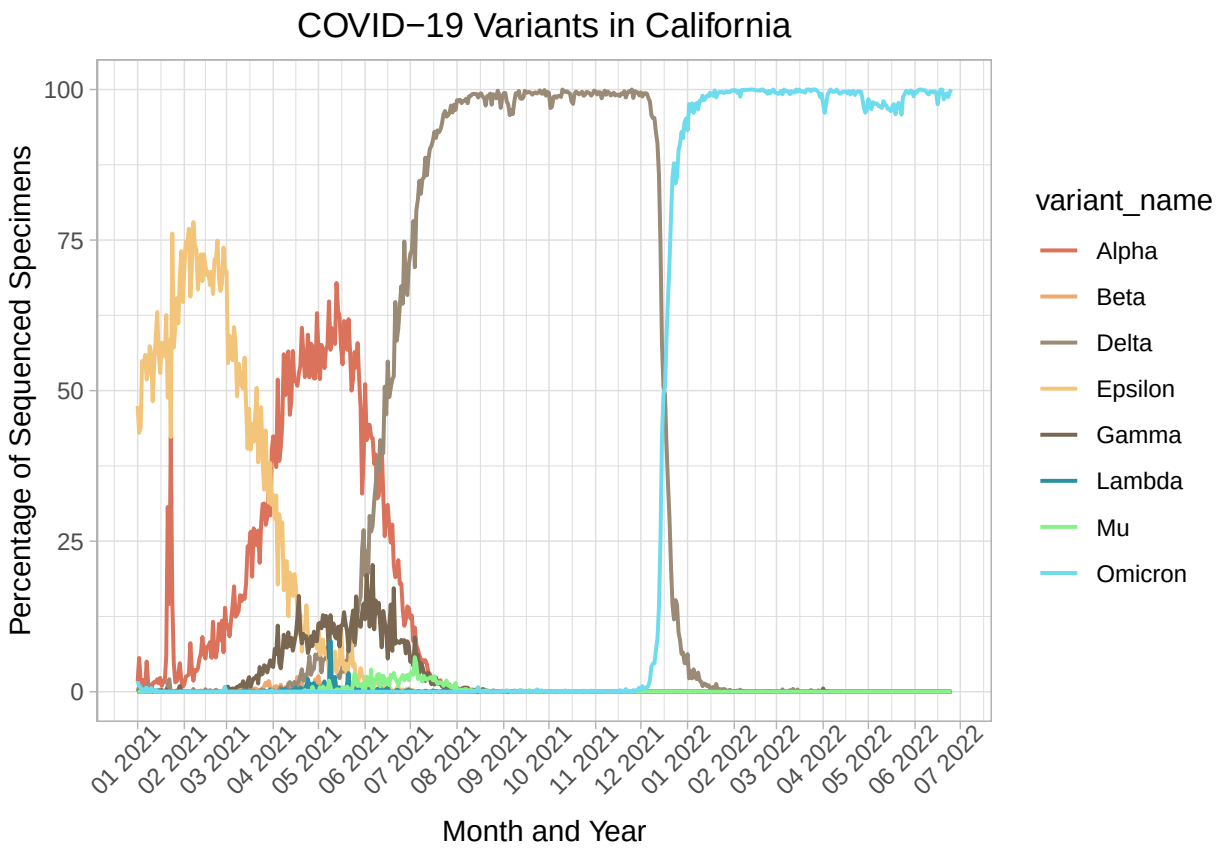
```
covid19 <- covid %>%
  filter(variant_name !="Other") %>% filter(variant_name !="Total")
head(covid19)
```

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

Generating graph using ggplot

```
library(ggthemr)
ggthemr("dust")

ggplot(covid19, aes(x=as.Date(date), y=percentage, group=variant_name, col=variant_name))+
    theme_light() + # light minimal background
    geom_line(size=0.7) + # making lines thicker
    ggtitle("COVID-19 Variants in California") + # setting title
    theme(plot.title = element_text(hjust = 0.5)) +
    ylab("Percentage of Sequenced Specimens") + # Renaming the Y axis
    xlab("Month and Year") + #Renaming the X axis
    scale_x_date(date_labels= "%m %Y", date_breaks = "1 month") + #labeling by month number and year
    theme(axis.text.x = element_text(angle=45, hjust=0.8)) # angle text for and adjust height to avoid
```


COVID−19 Variants in California
Percentage of Sequenced Specimens
100
75
50
25
0
01 2021
02 2021
03 2021
04 2021
05 2021
06 2021
07 2021
08 2021
09 2021
10 2021
11 2021
12 2021
01 2022
02 2022
03 2022
04 2022
05 2022
06 2022
07 2022
Month and Year
variant_name
Alpha
Beta
Delta
Epsilon
Gamma
Lambda
Mu
Omicron