

Question 15

Yuhan Zhang(A13829264)

7/16/2022

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

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

```
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 csv

```
c19.variant = read.csv("covid19_variants.csv")
```

Represent date using lubridate

```
c19.variant.date = ymd(c19.variant$date)
c19.variant$date = c19.variant.date
```

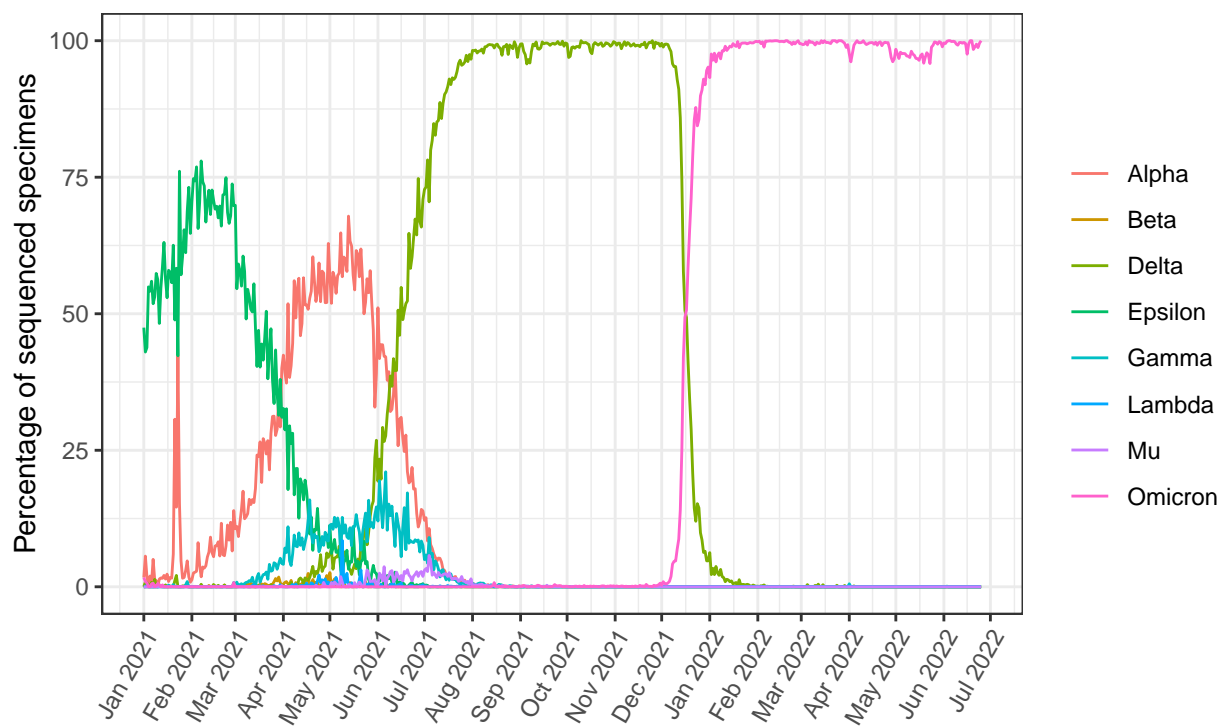
Eliminate rows that are “Other” or “Total”

```
c19.variant.greekOnly = c19.variant %>%
  filter(!variant_name %in% c("Total", "Other"))
```

Plot using ggplot2

```
ggplot(c19.variant.greekOnly, aes(x = date, y = percentage, col = variant_name)) +
  geom_line() +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month") +
  theme_bw() +
  theme(axis.text.x=element_text(angle=60, hjust=1)) +
  ggtitle("Covid-19 Variants in California") +
  xlab("") +
  ylab("Percentage of sequenced specimens") +
  labs(col = "", caption = "Data Source: <https://www.cdph.ca.gov/>")
```

Covid-19 Variants in California



Data Source: <<https://www.cdph.ca.gov/>>