First Year Exam: Question 15

Rachel Weinstein (A11836450)

7/8/2022

```
# Load necessary packages

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(ggplot2)
```

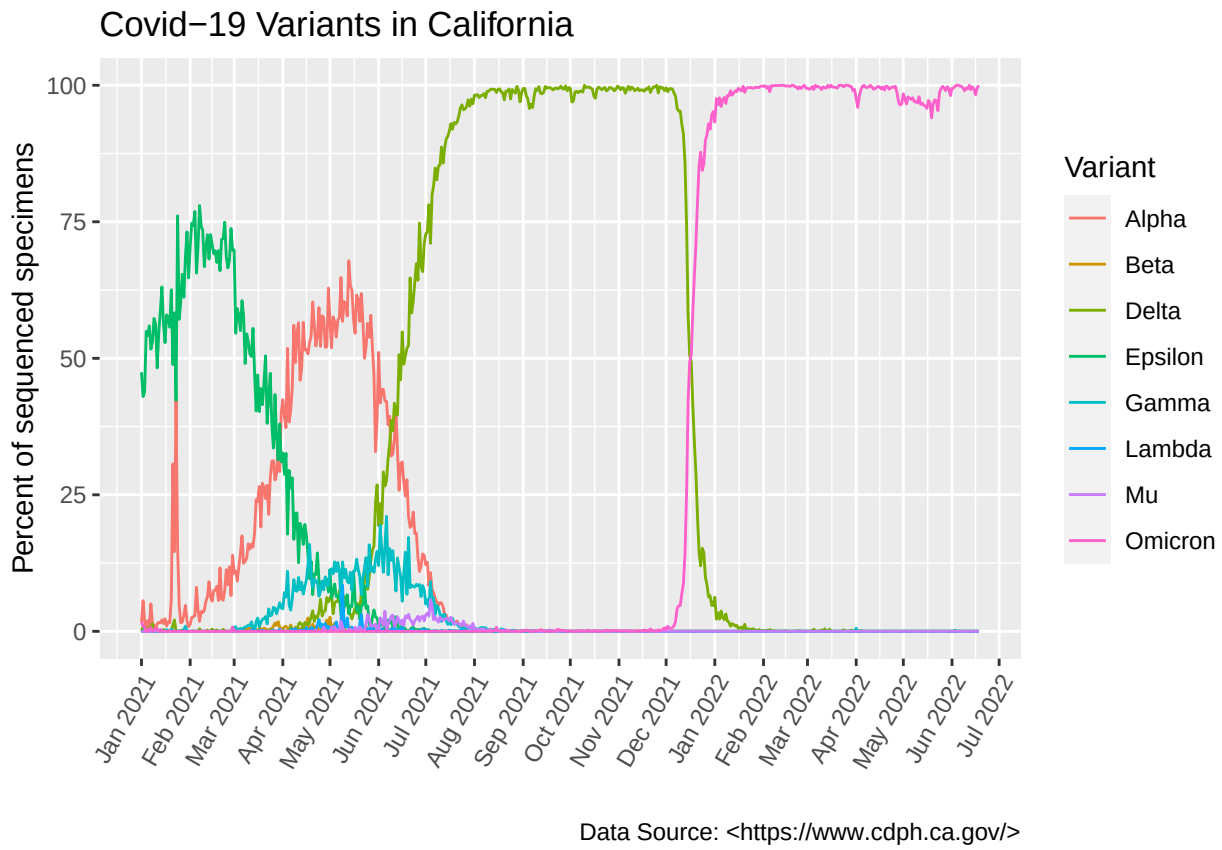
```
# Import the data

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

```
# Filter out 'Other' and 'Total' from variants

clean_data <- filter(data, variant_name != "Other", variant_name != "Total")
```

```
# Change name of column "variant_name" to "Variant" (an aesthetic change for the final figure)

clean_data <- rename(clean_data, Variant = variant_name)
```

```r
# Make data lubridate compatible

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

```r
# Plot the data

ggplot(clean_data, aes(date, percentage, color = Variant))+
  geom_line()+
  scale_x_date(date_breaks = "1 month", date_labels =  "%b %Y")+
  theme(axis.text.x=element_text(angle=60, hjust=1))+
  labs(x = "", y = "Percent of sequenced specimens", title = "Covid-19 Variants in California", caption
```



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