First Year Exam Q15

Shivani Khosla (PID: A59010433)

2022-07-15

importing data and viewing the first few rows

```
variants_data <- read.csv('covid19_variants.csv')
head(variants_data)
```

```
##         date       area area_type variant_name specimens percentage
## 1 2021-01-01 California     State        Alpha         1       1.69
## 2 2021-01-01 California     State         Beta         0       0.00
## 3 2021-01-01 California     State           Mu         0       0.00
## 4 2021-01-01 California     State        Gamma         0       0.00
## 5 2021-01-01 California     State        Total        59     100.00
## 6 2021-01-01 California     State      Omicron         1       1.69
##   specimens_7d_avg percentage_7d_avg
## 1               NA                NA
## 2               NA                NA
## 3               NA                NA
## 4               NA                NA
## 5               NA                NA
## 6               NA                NA
```

selecting all columns except "Total" and "Other" since these are not shown in the plot we are instructed to make

```
no_total_and_other <- variants_data[variants_data$variant_name != "Total"
                                    & variants_data$variant_name != "Other",]
```

call ggplot2 from library to use the package for creating the graph

```
library(ggplot2)
```

use the ggplot function and subfunctions to create the graph (details explained in comments below)

```
ggplot(data=no_total_and_other, aes(x=as.Date(date), y=percentage, color=variant_name)) +
    # tell ggplot what data to use; also specify date is on the x-axis
    # (converted from character to date format)
    # and percentage is on y-axis; color/group by variant
  geom_line() + # making a line graph/time series
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month") +
    # change date format to month/Year (e.g. Mar 2022)
  theme(axis.text.x=element_text(angle = 60, hjust = 1, vjust = 1),
        legend.title = element_blank()) +
    # set x-axis tick labels (dates) at an angle and re-position for easier
    # reading; also remove title from variant legend
  xlab(NULL) + # remove x-axis label
  ylab("Percentage of sequenced specimens") + # change y-axis label
  ggtitle("Covid-19 Variants in California") # add graph title
```