

First Year Exam

Andrew Kapinos | PID: A12708564

July 3, 2022

Load dataset and assign to object:

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

Load necessary packages:

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

Convert dates to proper format using lubridate and append to dataset as new column:

```
data$lubridate <- ymd(data$date)
```

Generate plot:

```
# Filter data to exclude rows containing "Total" in variant_name column using dplyr
data %>%
  filter(variant_name != "Total") %>%
  # Generate ggplot for data
  # Set x to formatted dates and y to percentage of total sequenced specimens
  ggplot(aes(lubridate,percentage)) +
  # Add geom_line and group by variant
  geom_line(aes(color = variant_name)) +
  # Set theme to minimal
  theme_minimal() +
  # Format x axis using Month Year labels and 1 month intervals
  # Set x axis to start at minimum date value
  scale_x_date(date_labels = "%b %Y",
               date_breaks = "1 month",
               expand = c(0,0)) +
  # Format y axis to begin at 0 and end at 100
  scale_y_continuous(limits = c(0,100),
                    expand = c(0,0)) +
  # Add y axis title
  ylab("Percentage of sequenced specimens") +
  # Edit theme and add axis lines
  theme(axis.line = element_line(),
        # Remove x axis title
        axis.title.x = element_blank(),
```

```

# Rotate and position x axis labels
axis.text.x = element_text(angle = 60,
                             hjust = 1),

# Change axis text color to black
axis.text = element_text(color = "black"),
# Add axis ticks
axis.ticks = element_line(),
# Remove legend title
legend.title = element_blank() +
# Add plot title
ggtitle("Covid-19 Variants in California") +
# Add caption with data source
labs(caption = "
      Data Source: <https://data.chhs.ca.gov/dataset/covid-19-variant-data>")

```

