1st Year Exam Q15

Shitian Li (PID: A13294481)

7/14/2022

Import the CSV file into R:

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

Inspect the data:

```
head(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
```

Import ggplot2 and dplyr:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## 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(scales)
```

```
## Warning: package 'scales' was built under R version 4.1.2
```

Trim out other and total:

```
trim_dat <- data %>%
  filter(variant_name != "Other") %>%
  filter(variant_name != "Total")
```

```
ggplot(trim_dat, aes(x=as.Date(date), y=percentage, group=variant_name, col=variant_name)) +
  geom_line() +
  theme(legend.title=element_blank()) +
  xlab("") +
  ylab("Percentage of Sequenced Specimens") +
  ggtitle("COVID-19 Variants in California") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 months") +
  theme(axis.text.x = element_text(angle=45, hjust = 1))
```