HW2

Question 1

Question 1(a)

```r
load("~/Documents/2023Fall/P8157/P8157/MACS-VL.RData")
data = macsVL
# number of clusters
length(unique(data$id))
```

```
## [1] 225
```

```r
# number of measurements within each cluster
obs = data |> group_by(id) |> summarize(n_obs = n())
summary(obs$n_obs)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.000   7.000   8.000   7.484   9.000  10.000
```

```r
# follow-up period
fl = data |> group_by(id) |> mutate(max_mon = max(month)) |>
  filter(month == max_mon)
summary(fl$max_mon)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   42.00   45.00   42.22   47.00   48.00
```

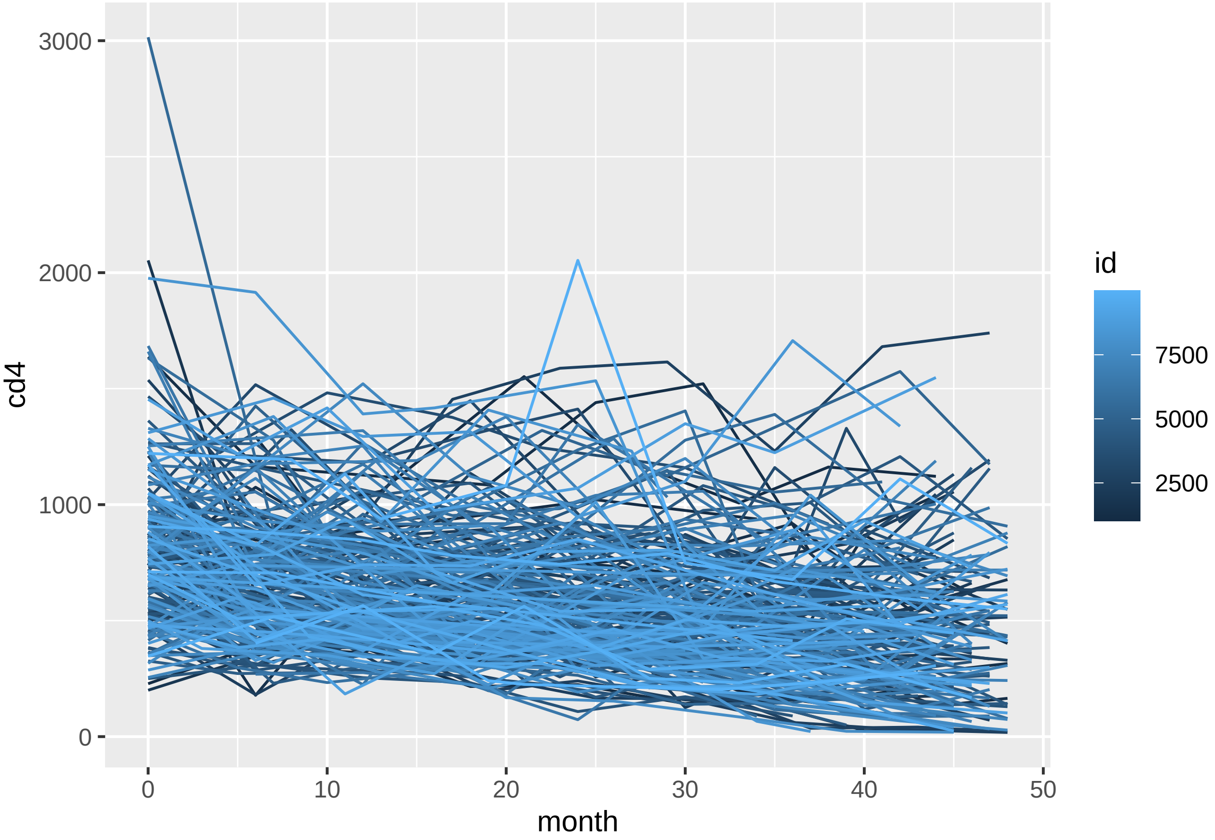
```r
# time interval between measurements within each cluster
int = data |>
  group_by(id) |>
  mutate(delta_mon = month - lag(month))
mean_int = mean(int$delta_mon, na.rm = TRUE)
median_int = median(int$delta_mon, na.rm = TRUE)
# baseline vload
vl = data |> group_by(id) |> summarize(vload = first(vload))
summary(vl$vload)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     300    7928   24573   78348   91195 1026656
```

```r
# cd4+ count
c4 = data |> group_by(id) |> summarize(base_cd4 = first(cd4), last_cd4 = last(cd4)) |>
  mutate(loss_cd4 = base_cd4 - last_cd4)
summary(c4$loss_cd4)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -452.0   115.0   283.0   316.4   467.0  1917.0
```

```
# spaghetti plot
ggplot(data, aes(x = month, y = cd4, group = id, color = id)) +
  geom_line()
```


cd4
month
id