# Stats 112 Homework 1

Owen Lin

Problem 1

1a and 1b:

```
epilepsy = read.table("D:\\Coding\\Stats112\\Data\\epilepsy.txt", header=TRUE)
epilepsy[,4] = epilepsy[,4]/8
epilepsy[,c(5:8)] = epilepsy[,c(5:8)]/2
epi.long = reshape(epilepsy, idvar="ID", varying=list(4:8),v.names="Rate", timevar="Time", times
=c(0,2,4,6,8),direction="long")

means = tapply(epi.long$Rate, list(epi.long$Time,epi.long$trt), mean)

matplot(c(0,2,4,6,8),means,col=c(1,1),lty=c(3,1),type="o",pch=c(1,16),xlab="Time (weeks)",ylab
="Mean rate of seizures (per week)",ylim=c(2.5,5.0),main="Figure 1b: Mean Rate of Seizures by Tr
eatment Group")

legend(3.5,3.0, c("Placebo","Progabide"), lty=c(3,1))
```


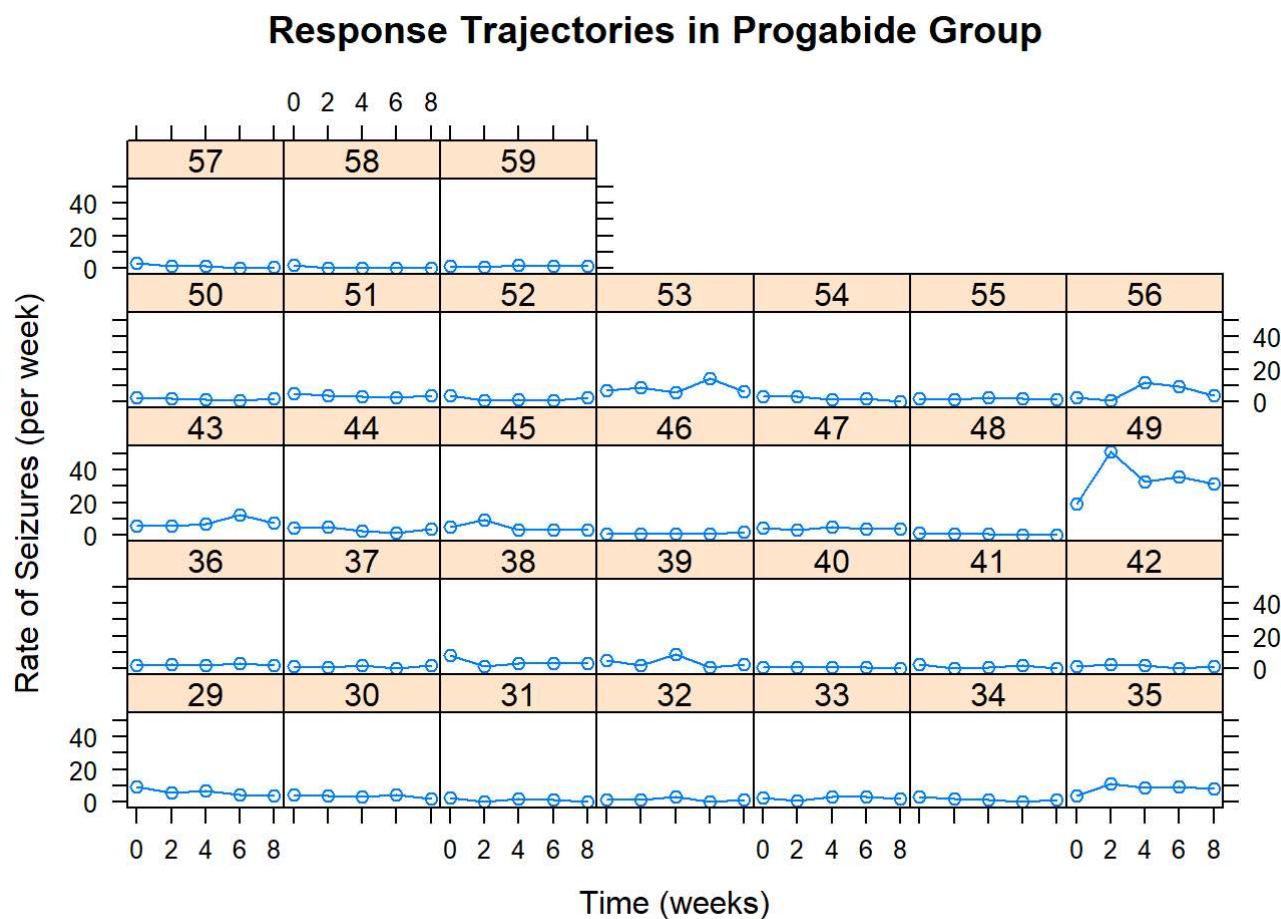Figure 1b: Mean Rate of Seizures by Treatment Group
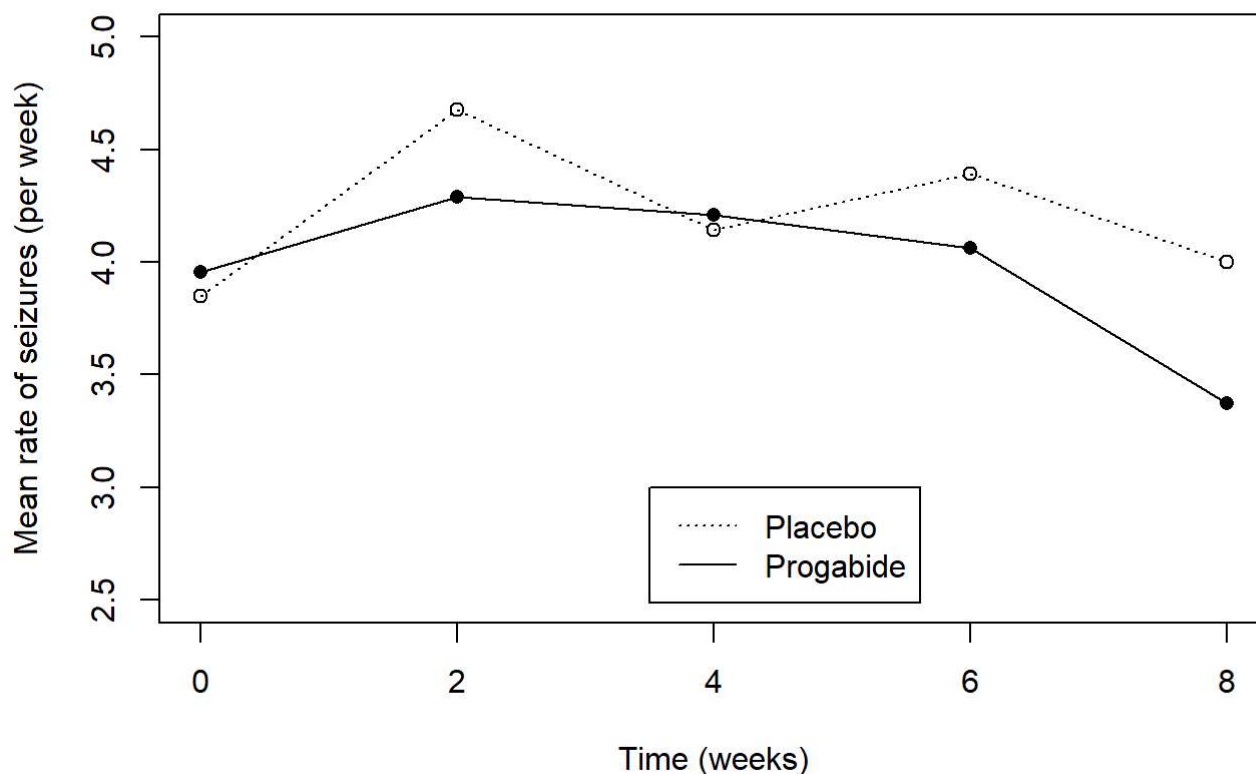
1c:

There is no particular time trends reflected across the individual plots for the Progabide group

```
Prog = epi.long[epi.long$trt=="Progabide",]

# Progabide group
xyplot(Rate ~ Time | factor(ID), data=Prog, type="o",main="Response Trajectories in Progabide Gr
oup",xlab="Time (weeks)", ylab="Rate of Seizures (per week)")
```

## Response Trajectories in Progabide Group

1d: Similarly, there is no particular time trends reflected across the individual plots for the Placebo group

```
Plac = epi.long[epi.long$trt=="Placebo",]

# Placebo group
xyplot(Rate ~ Time | factor(ID), data=Plac, type="o",main="Response Trajectories in Placebo Grou
p",xlab="Time (weeks)", ylab="Rate of Seizures (per week)")
```

Response Trajectories in Placebo Group
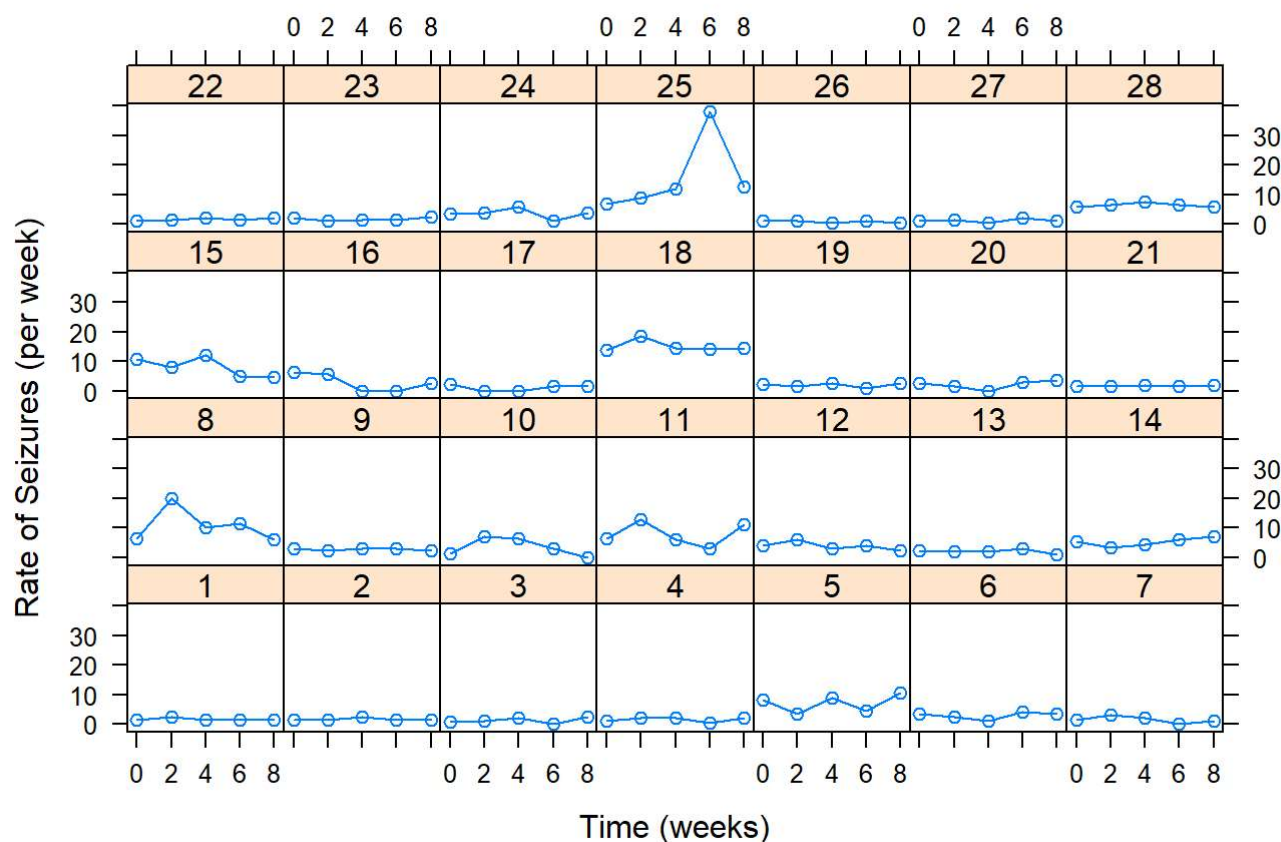
1e:

It seems that individuals in the Progabide group have higher variance across all weeks (except in week 6)

```
cov(epilepsy[epilepsy$trt=="Placebo",4:8])
```

```
##          Week0    Week2    Week4    Week6    Week8
## Week0 10.64740 12.30820 11.07341 11.80489 10.18519
## Week2 12.30820 25.68915 16.18651 18.85317 13.06481
## Week4 11.07341 16.18651 16.66402 19.79365 12.12963
## Week6 11.80489 18.85317 19.79365 53.82143 18.89815
## Week8 10.18519 13.06481 12.12963 18.89815 14.48148
```

```
cov(epilepsy[epilepsy$trt=="Progabide",4:8])
```

```
##          Week0    Week2    Week4    Week6    Week8
## Week0 12.24432 27.24872 17.54919 20.28421 17.18575
## Week2 27.24872 83.17957 49.05376 57.81398 49.83038
## Week4 17.54919 49.05376 35.16290 38.10269 31.55296
## Week6 20.28421 57.81398 38.10269 48.26237 37.20860
## Week8 17.18575 49.83038 31.55296 37.20860 31.66613
```

1f:

H_0 : The mean change in rate of seizures from the baseline to the last week of the study is equal to zero.

H_a : The mean change is not equal to zero.

Test Statistic: 1.054 ~ t(30)

P-value: 0.3003

Conclusion: We fail to reject the null hypothesis and conclude that the mean change is not significantly different from 0 for the Progabide group.

```
Pro.W0 = epilepsy$Week0[epilepsy$trt=="Progabide"]
Pro.W8 = epilepsy$Week8[epilepsy$trt=="Progabide"]
t.test(Pro.W0, Pro.W8, paired=T, alternative="two.sided")
```

```
##
##  Paired t-test
##
## data:  Pro.W0 and Pro.W8
## t = 1.054, df = 30, p-value = 0.3003
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.5482002  1.7175551
## sample estimates:
## mean difference
##       0.5846774
```

1g:

H_0 : The mean change in rate of seizures from the baseline to the last week of the study is the same in both groups.

H_a : The mean change differs between the two groups.

Test Statistic: -1.0656 ~ t(54)

P-value: 0.2913

Conclusion: We fail to reject the null and conclude that the mean change in rate of seizures from the baseline to the last week of the study is not significantly different from each other.

```
Diffs = epilepsy$Week8 - epilepsy$Week0
Pro.D = Diffs[epilepsy$trt=="Placebo"]
Pla.D = Diffs[epilepsy$trt=="Progabide"]
t.test(Pla.D, Pro.D, mu=0, alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  Pla.D and Pro.D
## t = -1.0656, df = 53.99, p-value = 0.2913
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.1220927  0.6491664
## sample estimates:
##  mean of x  mean of y
## -0.5846774  0.1517857
```

Problem 2

2a:

For all weeks, the summary statistics are pretty comparable to each other.

```
exercise = read.csv("D:\\Coding\\Stats112\\Data\\exercise.csv")
summary(exercise)
```

```
##        id          program             week_0          week_1
##  Min.   : 1    Length:37         Min.   :74.00   Min.   :75.00
##  1st Qu.:10    Class :character  1st Qu.:79.00   1st Qu.:79.00
##  Median :19    Mode  :character  Median :80.00   Median :81.00
##  Mean   :19                      Mean   :80.46   Mean   :81.42
##  3rd Qu.:28                      3rd Qu.:83.00   3rd Qu.:84.00
##  Max.   :37                      Max.   :87.00   Max.   :91.00
##                                                  NA's   :1
##      week_2           week_3           week_4
##  Min.   :76.00    Min.   :75.00    Min.   :75.0
##  1st Qu.:80.00    1st Qu.:80.00    1st Qu.:80.0
##  Median :82.00    Median :82.00    Median :82.0
##  Mean   :82.08    Mean   :81.88    Mean   :81.8
##  3rd Qu.:84.00    3rd Qu.:84.00    3rd Qu.:85.0
##  Max.   :90.00    Max.   :91.00    Max.   :87.0
##  NA's   :1        NA's   :3        NA's   :7
```

2b:

It seems that the muscle strength in all the weeks are highly positively correlated with each other. We exclude the id because it doesn't convey any useful information in the analysis.
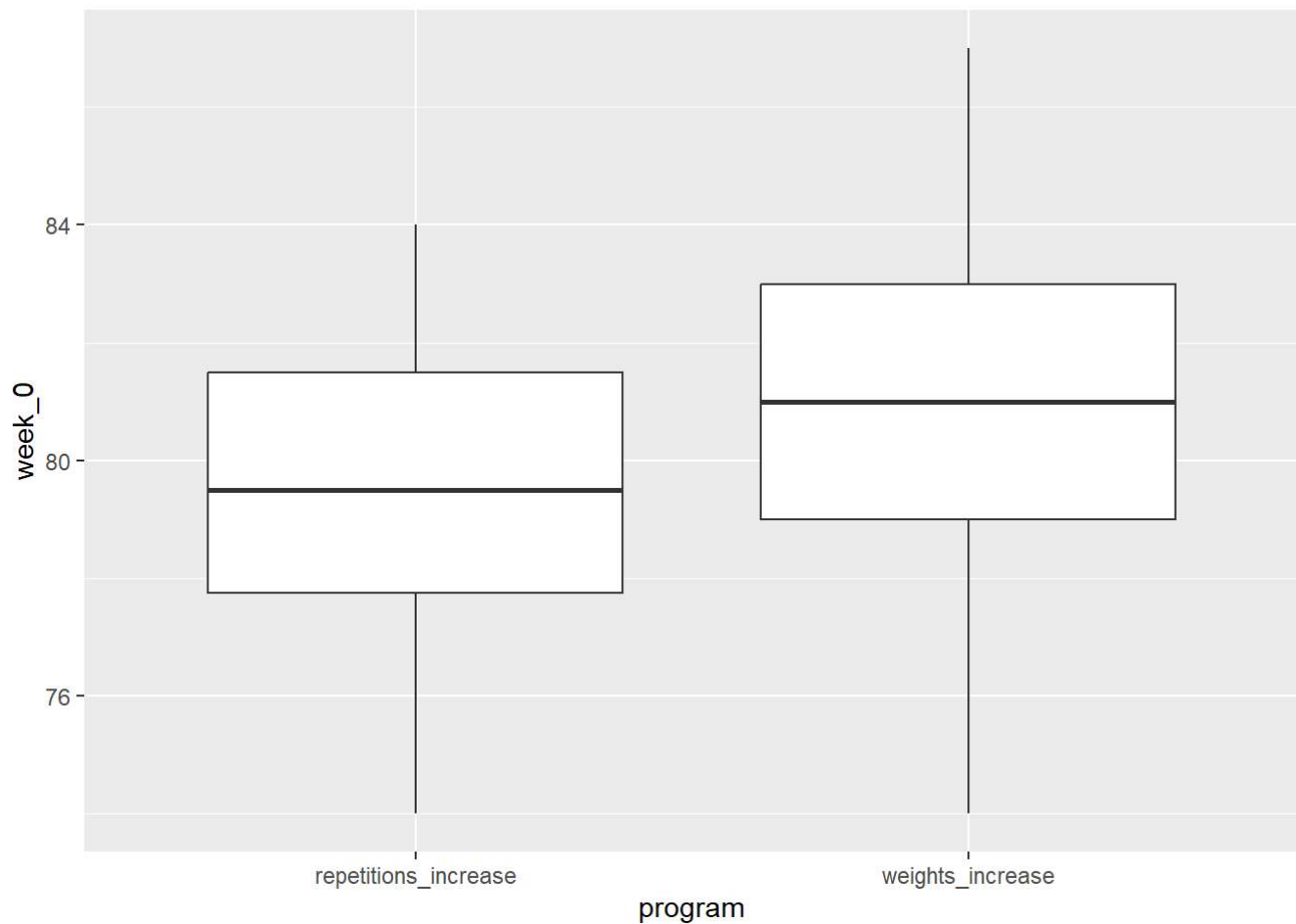
```
ggpairs(select(exercise, -id))
```
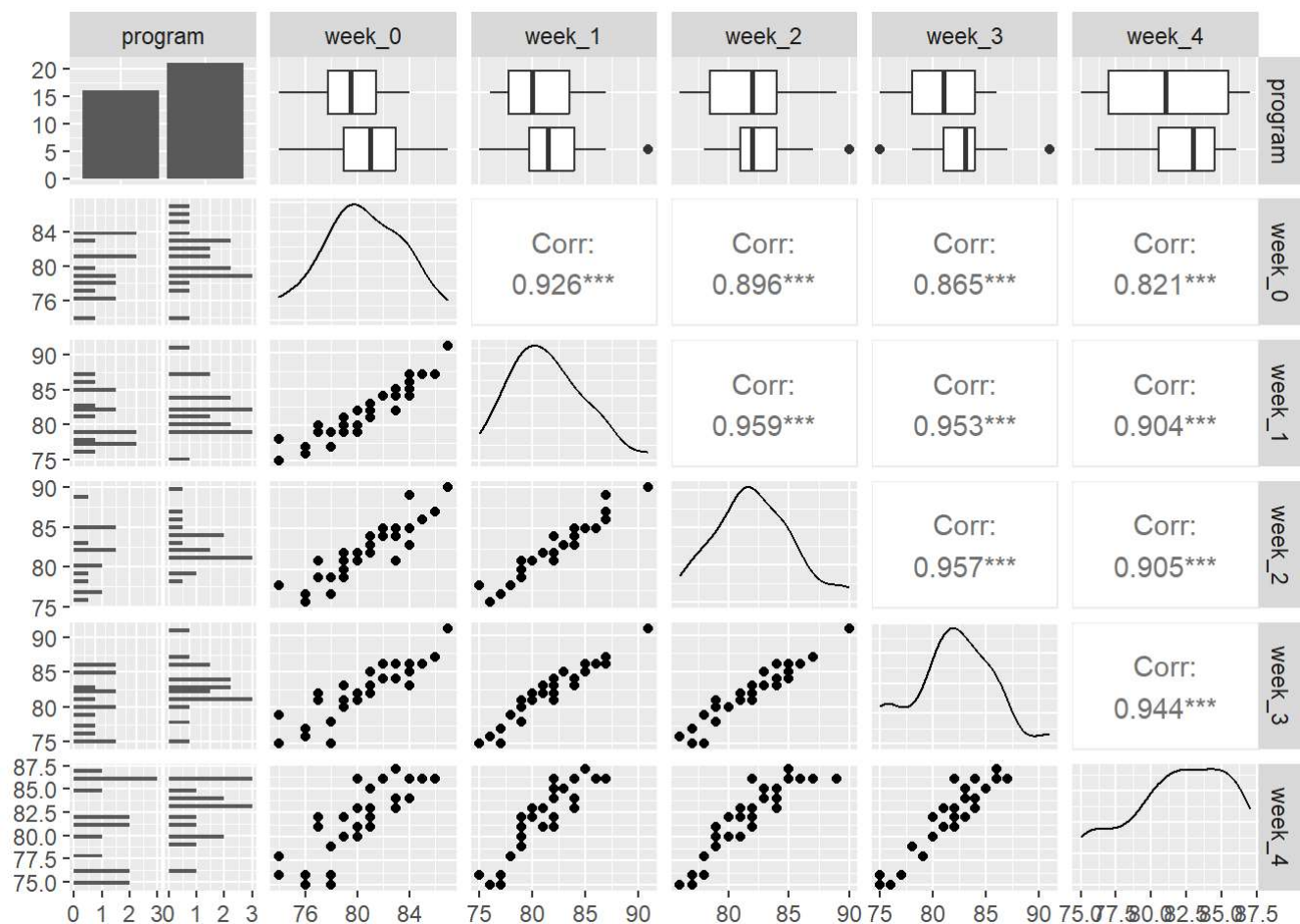
2c:

On week 0, participants in the weight-increase program have a higher muscle strength on average.

```
exercise %>%
ggplot(aes(x = program, y = week_0)) + geom_boxplot()
```

2d:

```
long_exercise <- exercise %>%
pivot_longer(cols = starts_with("week"), names_to = "week", values_to = "muscle_strength")
glimpse(long_exercise)
```

```
## Rows: 185
## Columns: 4
## $ id              <int> 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, …
## $ program         <chr> "repetitions_increase", "repetitions_increase", "repet…
## $ week            <chr> "week_0", "week_1", "week_2", "week_3", "week_4", "wee…
## $ muscle_strength <int> 79, 79, 80, 80, 80, 83, 85, 85, 86, 87, 81, 82, 82, 83…
```

2e:
The change in the magnitude of muscle strength for the repetitions_increase group is not obvious, it fluctuates. However, it seems that the spread of the observation is getting larger.

```
long_exercise %>%
ggplot(aes(x = program, y = muscle_strength)) +
geom_boxplot() +
facet_wrap(~week, nrow = 3)
```
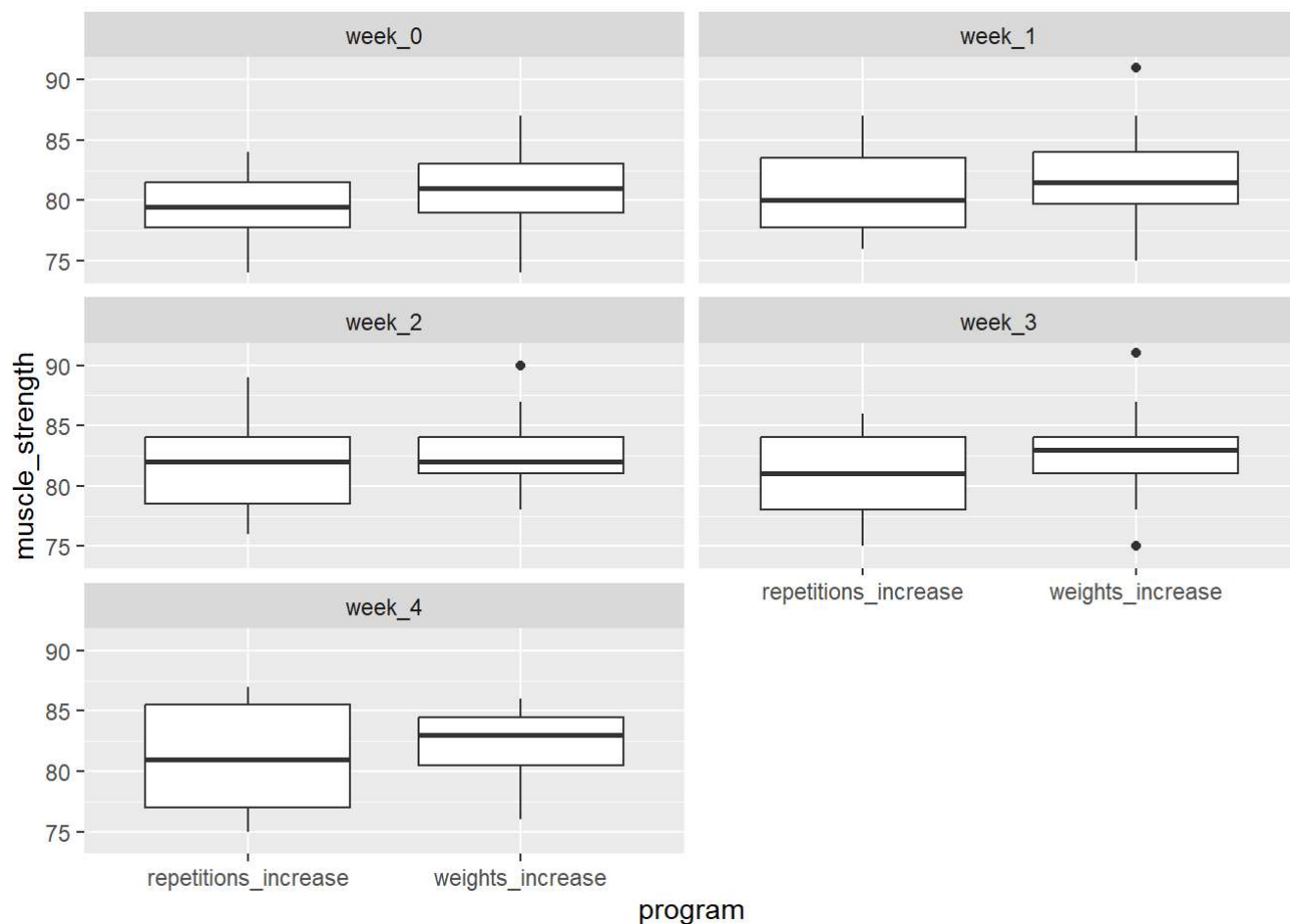
2f:

The result aligns with the observation in part e.

```
long_exercise %>%
group_by(program, week) %>%
summarize(mean_muscle_strength = mean(muscle_strength, na.rm = TRUE),
sd_muscle_strength = sd(muscle_strength, na.rm = TRUE))
```

```
## # A tibble: 10 × 4
## # Groups:    program [2]
##     program               week    mean_muscle_strength sd_muscle_strength
##     <chr>                 <chr>                  <dbl>              <dbl>
##  1 repetitions_increase week_0                   79.7               3.11
##  2 repetitions_increase week_1                   80.8               3.58
##  3 repetitions_increase week_2                   81.3               3.68
##  4 repetitions_increase week_3                   80.8               3.84
##  5 repetitions_increase week_4                   81.1               4.32
##  6 weights_increase      week_0                   81.0               3.11
##  7 weights_increase      week_1                   81.9               3.57
##  8 weights_increase      week_2                   82.6               2.85
##  9 weights_increase      week_3                   82.7               3.46
## 10 weights_increase      week_4                   82.5               2.90
```

2g:

No much trend is presented in this individual plot.
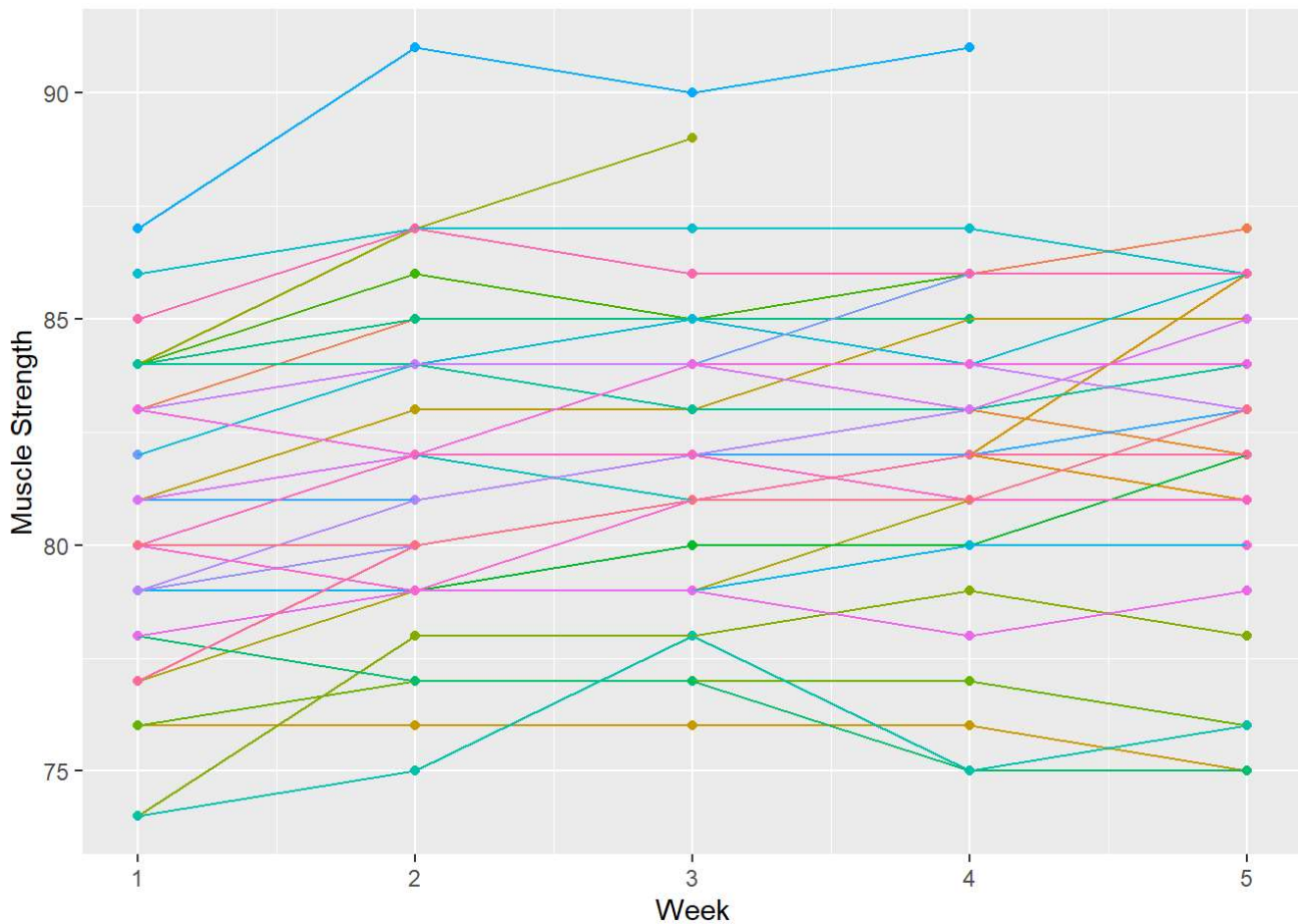
```
long_exercise <- long_exercise %>%
mutate(week_numeric = str_sub(week, 5, 6),
week_numeric = as.numeric(as.factor(week_numeric))) %>%
relocate(week_numeric, .after = week)

long_exercise %>%
ggplot(aes(x = week_numeric, y = muscle_strength,
group = id, color = factor(id))) + geom_point(show.legend = FALSE) +
geom_line(show.legend = FALSE) + labs(x = "Week", y = "Muscle Strength")
```


2h:

For both programs, the mean muscle strength first increases and then decreases.
Interestingly, repetitions_increase group experience a big drop in muscle strength going from week 4 to week 6

```
long_exercise %>%
group_by(week_numeric, program) %>%
summarize(mean_muscle_strength = mean(muscle_strength, na.rm = TRUE)) %>%
ggplot(aes(x = week_numeric, y = mean_muscle_strength, color = program)) +
geom_point() +
geom_line() +
labs(x = "Week", y = "Mean Muscle Strength")
```