STATS 112 HW1

Viraj Vijaywargiya

2023-04-13

```
library(lattice)
library(nlme)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
##
##     lmList
```

```
library(survival)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::collapse() masks nlme::collapse()
## x tidyr::expand()   masks Matrix::expand()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x tidyr::pack()     masks Matrix::pack()
## x tidyr::unpack()   masks Matrix::unpack()
```

1.
```
epilepsy = read.table("/Users/virajvijaywargiya/Downloads/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(
```

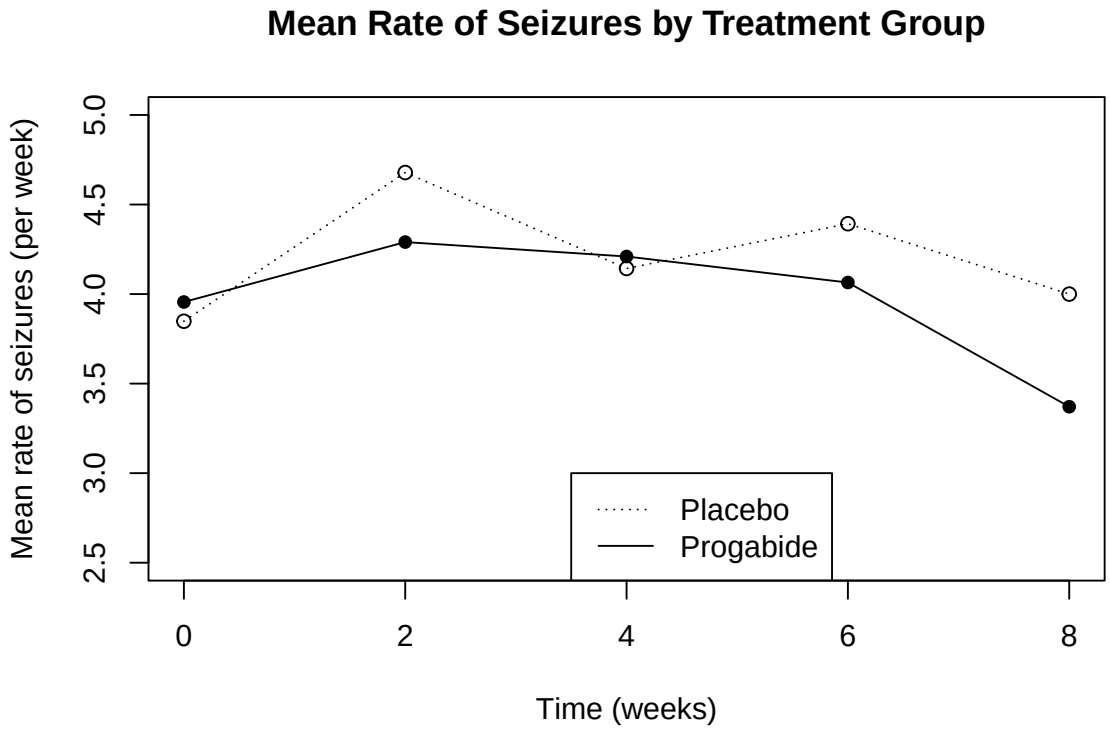
1a)

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

```
##    Placebo Progabide
## 0 3.848214  3.955645
## 2 4.678571  4.290323
## 4 4.142857  4.209677
## 6 4.392857  4.064516
## 8 4.000000  3.370968
```

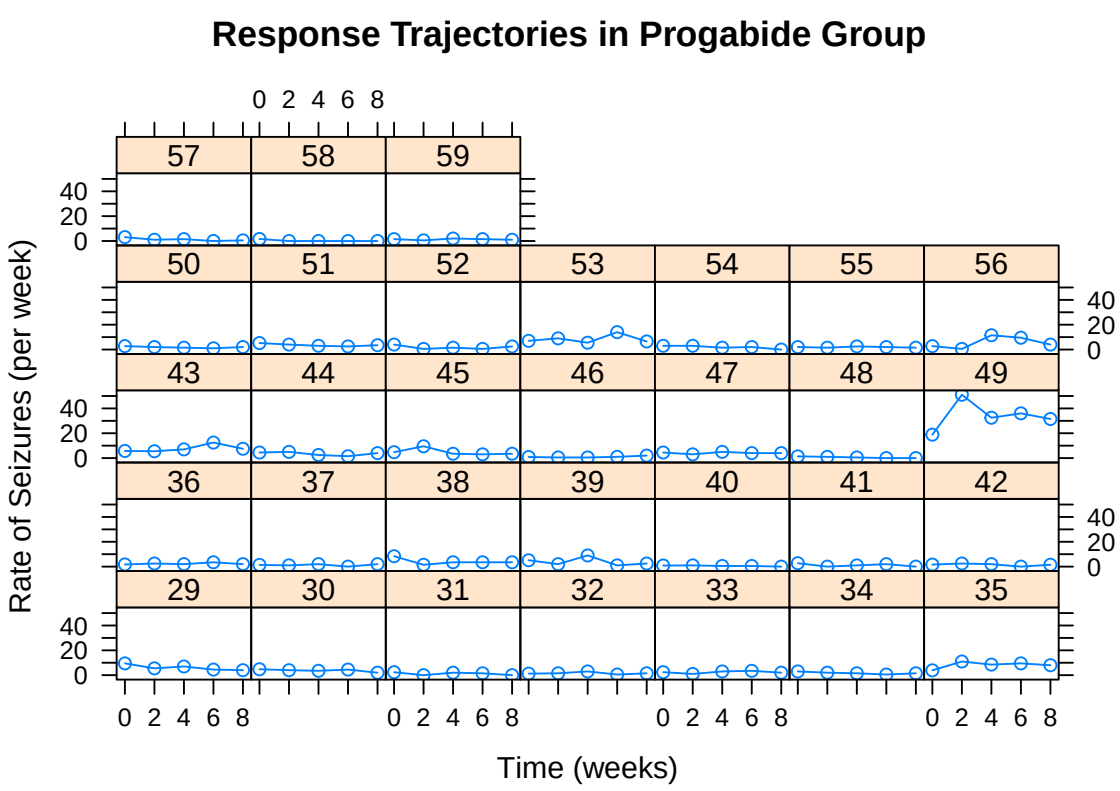
1b)

```
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="Mea

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



1c)

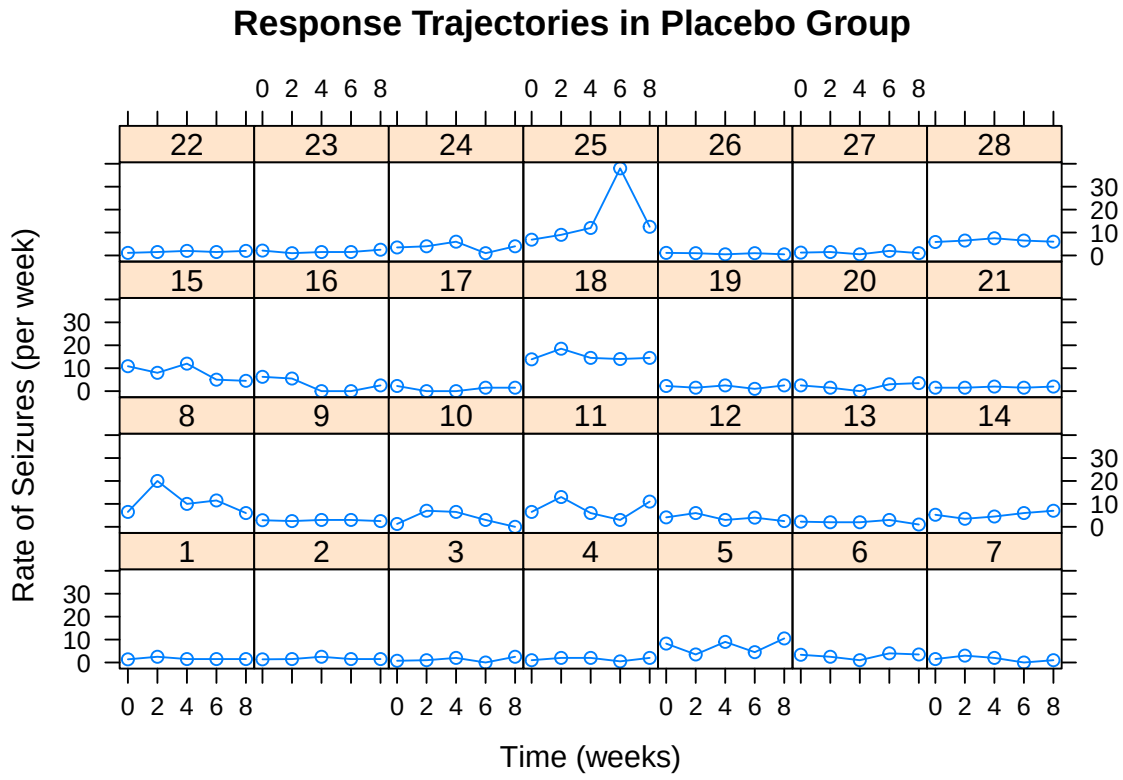
```
Prog = epi.long[epi.long$trt=="Progabide",]
Plac = epi.long[epi.long$trt=="Placebo",]

# Progabide group
xyplot(Rate ~ Time | factor(ID), data=Prog, type="o",main="Response Trajectories in Progabide Group
```



There seems to be a slight downwards trend with time moving on.

1d)

```
# Placebo group
xyplot(Rate ~ Time | factor(ID), data=Plac, type="o",main="Response Trajectories in Placebo Group",
```



This also seems to have a slight downwards trend with time moving on.

1e) Sample Covariance matrix for the Placebo group:

```
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
```

Sample covariance matrix for the Progabide group:

```
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
```

The sample covariance between weeks vary quite a bit across time. They seem somewhat similar across the time lags in the placebo group, however, they differ quite a bit in the Progabide group. For example: In the Progabide group, the sample covariance at the baseline is 12.24 but in week 2 it jumps to 83.18.

1f)

```r
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 difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5482002  1.7175551
## sample estimates:
## mean of the differences
##               0.5846774
```

Null Hypothesis H0: u(d) = 0 vs Alternate Hypothesis HA: u(d) != 0. Test statistic t = 1.054, p-value = 0.3003.
Therefore, 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 equal to zero.

1g)

```r
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
```

Null Hypothesis H0: u1 = u2 vs Alternate Hypothesis HA: u1 != u2 (u1 is the mean change for Placebo and u2 is the mean change for Progabide). Test statistic t = -1.0656. p-value = 0.2913.
Therefore, we fail to reject the null hypothesis and conclude that the mean change in rate of seizures from the baseline to the last week of the study is the same in both groups.

2. **2a)**

```r
exercise = read.csv("/Users/virajvijaywargiya/Downloads/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
```

From the output we can see that the averages for each week are around 80-82, and the inter-quartile ranges are similar. In addition, there are missing observations in week 1, 2, 3, 4.

2b)

```
library(GGally)
```

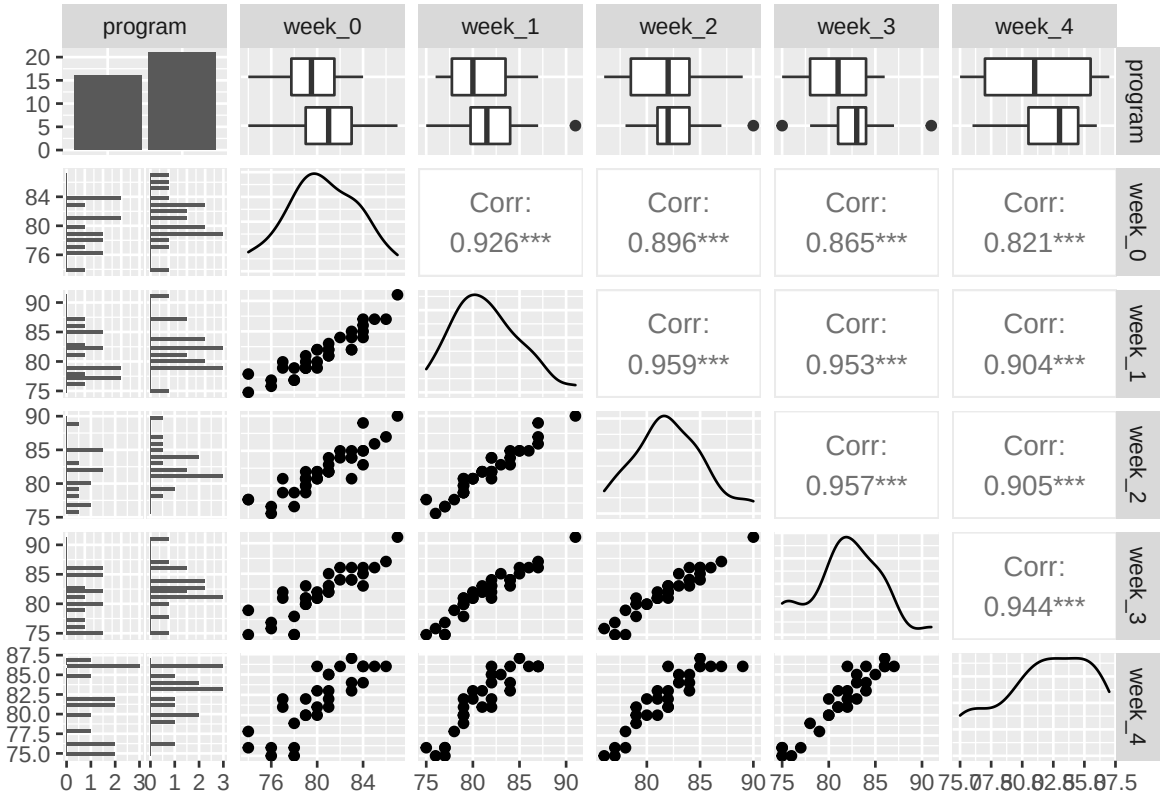
```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
ggpairs(select(exercise, -id))
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
## Removed 1 rows containing non-finite values (stat_boxplot).

## Warning: Removed 3 rows containing non-finite values (stat_boxplot).

## Warning: Removed 7 rows containing non-finite values (stat_boxplot).

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removing 1 row that contained a missing value

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 3 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 7 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1 rows containing non-finite values (stat_bin).

## Warning: Removed 1 rows containing missing values (geom_point).
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 2 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 4 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 7 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1 rows containing non-finite values (stat_bin).

## Warning: Removed 1 rows containing missing values (geom_point).

## Warning: Removed 2 rows containing missing values (geom_point).

## Warning: Removed 1 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 4 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 8 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 3 rows containing non-finite values (stat_bin).

## Warning: Removed 3 rows containing missing values (geom_point).

## Warning: Removed 4 rows containing missing values (geom_point).
## Removed 4 rows containing missing values (geom_point).

## Warning: Removed 3 rows containing non-finite values (stat_density).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 9 rows containing missing values

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 7 rows containing non-finite values (stat_bin).

## Warning: Removed 7 rows containing missing values (geom_point).
## Removed 7 rows containing missing values (geom_point).

## Warning: Removed 8 rows containing missing values (geom_point).

## Warning: Removed 9 rows containing missing values (geom_point).
```

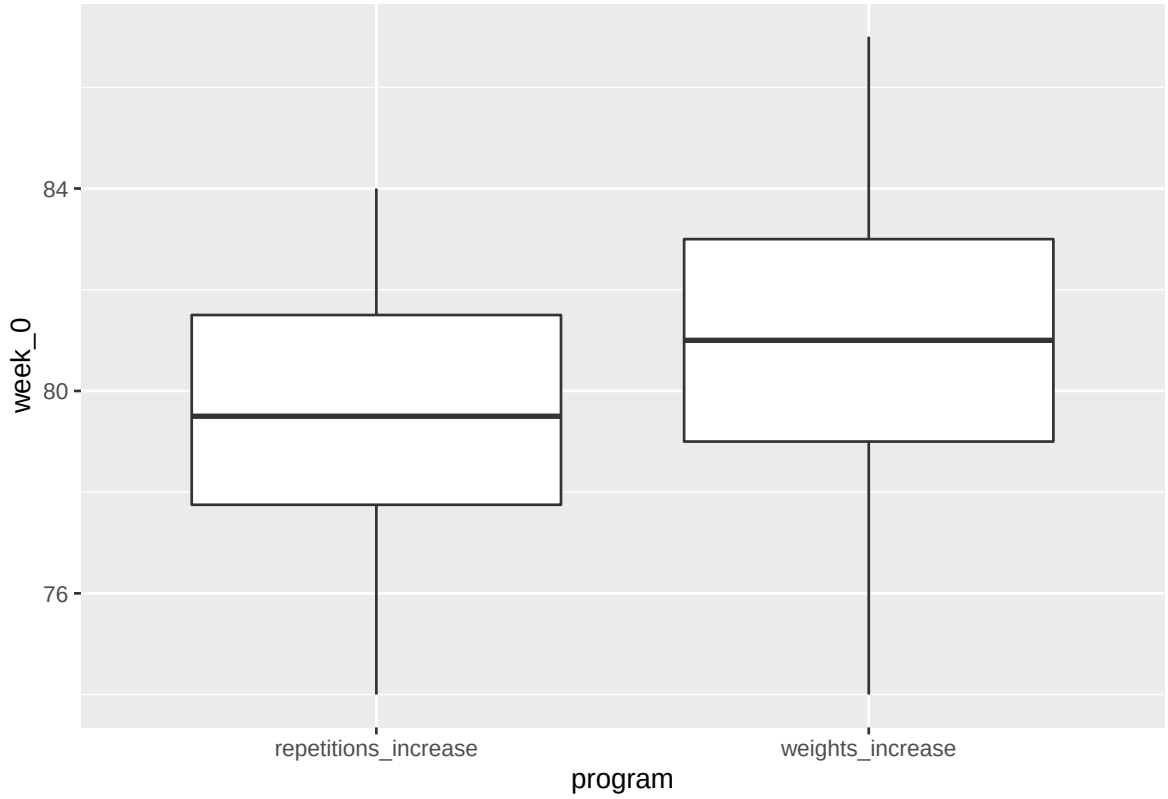
```
## Warning: Removed 7 rows containing non-finite values (stat_density).
```



This plot shows all the possible two-way relationships. It will do a scatterplot if both variables from the dataset are quantitative, for example, week measurements. The top row shows boxplots of the weekly measurements by different groups.
We are excluding the IDs because it plays no role in the dataset as a covariate.

2c)

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



From the boxplots above, we can see that the weights_increase group has a slightly higher muscle strength measurement than the repetitions_increase group, for week 0.

2d)

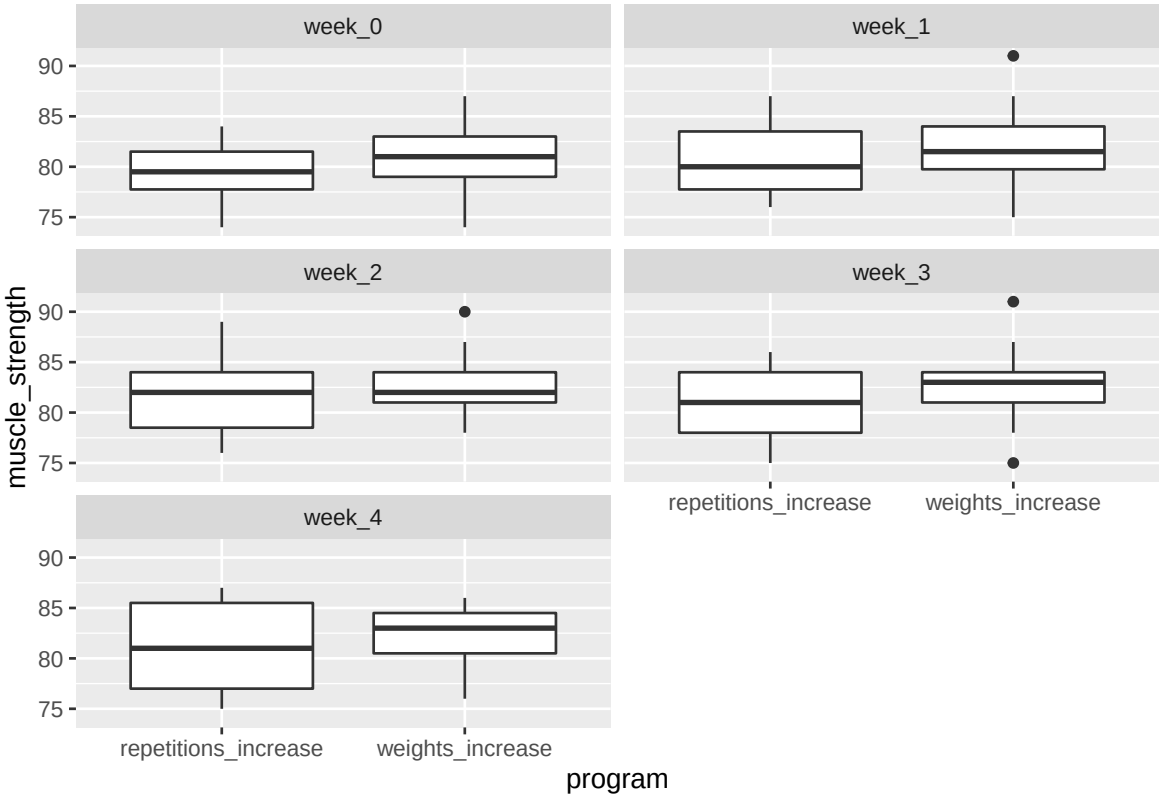
```
long_exercise <- exercise %>%
pivot_longer(cols = starts_with("week"),
names_to = "week",
values_to = "muscle_strength")
glimpse(long_exercise)
```

```
## 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)

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

```
## Warning: Removed 12 rows containing non-finite values (stat_boxplot).
```



For the repetitions_increase group, the average muscle strength across the weeks does not have much difference, it is mainly similar. However, the range for the muscle strength increases across the week, reaching maximum around 85.

2f)

```
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))
```

```
## `summarise()` has grouped output by 'program'. You can override using the
## `.groups` argument.
```

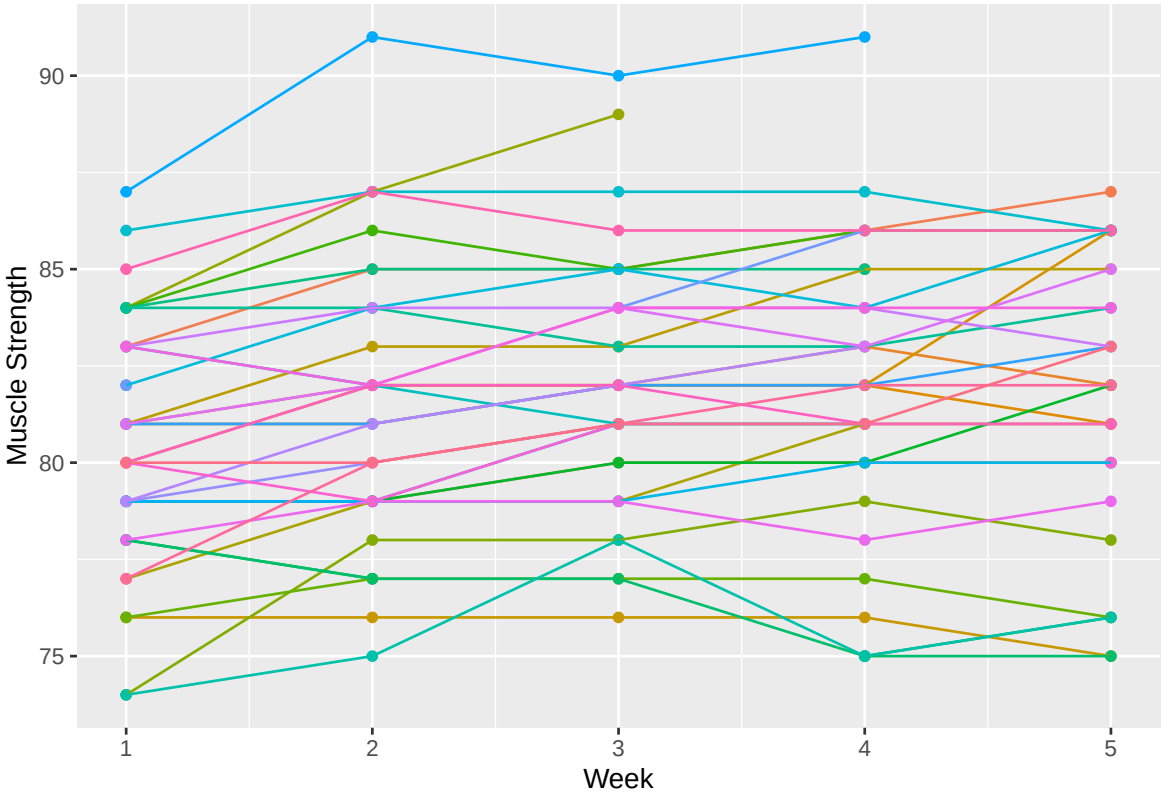
```
## # A tibble: 10 x 4
## # Groups:   program [2]
##    program              week   mean_muscle_strength sd_muscle_strength
##    <chr>                <chr>                 <dbl>              <dbl>
##  1 repetitions_increase week_0                79.7               3.11
##  2 repetitions_increase 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
```

This output relates to what was stated in part e as from the output we can see that the mean_muscle_strength is approximately the same across the weeks, however, the sd_muscle_strength increases across the weeks.

2g)

```
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")
```

```
## Warning: Removed 12 rows containing missing values (geom_point).

## Warning: Removed 8 row(s) containing missing values (geom_path).
```



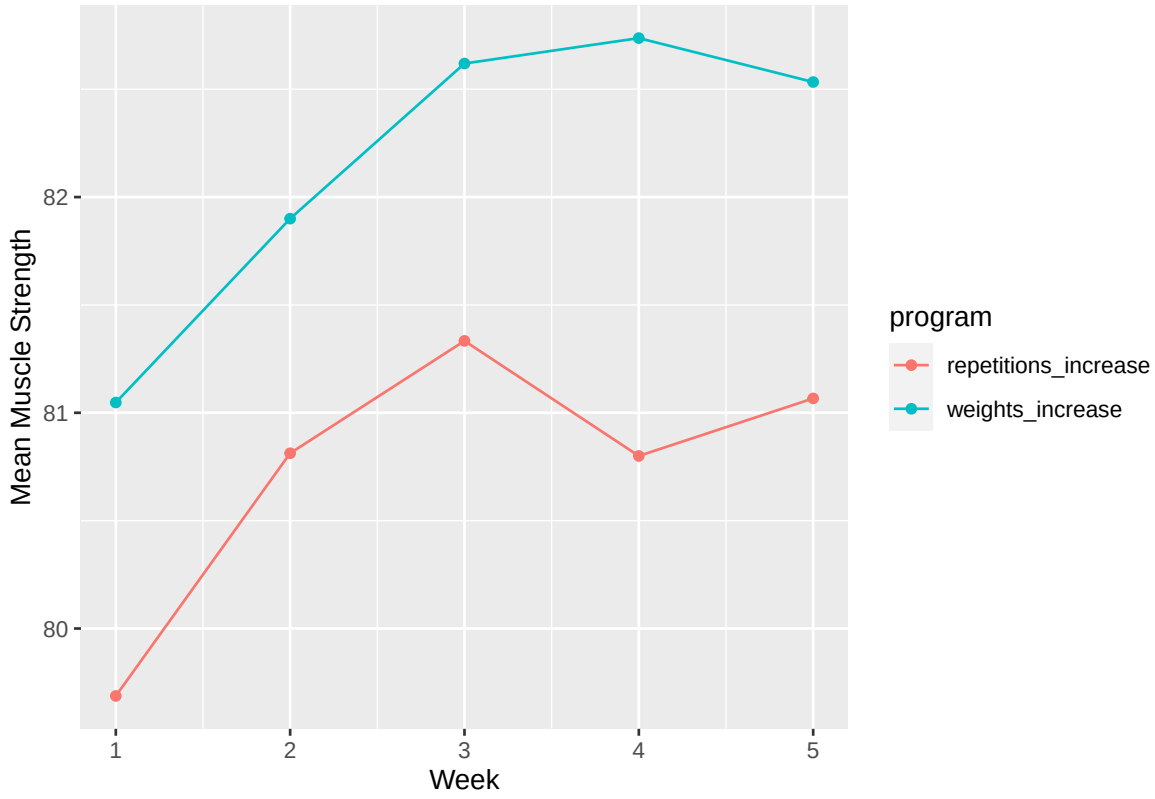
There isn't any fixed trend in the time plot above, some increase across weeks and some decrease. The time trends are all over the place.

2h)

```
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")
```

```
## `summarise()` has grouped output by 'week_numeric'. You can override using the
## `.groups` argument.
```



The mean muscle strength for both groups increase as the weeks increase, therefore, there is an increase in the time trend.