Brain Volumes Analysis

Isadora Sinha

26/02/2021

R Analysis of TP1, TP2 and TP3 Brain Volumes

Set working directory to where Matlab brain volume output CSV is located and load Tidyverse to work with a tibble.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
setwd("~/Documents/Rotation_1/final_data_analysis")
```

Load in CSVs as tibbles with column names set to true then check class of each column to ensure the brain volumes are numeric. *Note: brain volumes are in mm^3!*

```
TP3_brain_vols <- read_csv("TP3_brain_vols.csv",col_names = TRUE)
```

```
## Parsed with column specification:
## cols(
##   SampleID = col_character(),
##   TP3 = col_double()
## )
```

```
TP3_brain_vols
```

```
## # A tibble: 18 x 2
##    SampleID   TP3
##    <chr>    <dbl>
##  1 A48P      535.
##  2 A49P      543.
##  3 A51P      526.
##  4 A52P      561.
##  5 A53P      519.
##  6 A54P      527.
##  7 A60P      568.
##  8 A61P      504.
##  9 A62P      543.
## 10 A63P      559.
## 11 A65P      499.
```

```
## 12 A66P      444.
## 13 A67P      480.
## 14 A68P      469.
## 15 A69P      602.
## 16 A70P      523.
## 17 A71P      549.
## 18 A74P      554.
```

```
class(TP3_brain_vols$SampleID)
```

```
## [1] "character"
```

```
class(TP3_brain_vols$TP3)
```

```
## [1] "numeric"
```

```
TP2_brain_vols <- read_csv("TP2_brain_vols.csv",col_names = TRUE)
```

```
## Parsed with column specification:
## cols(
##   SampleID = col_character(),
##   TP2 = col_double()
## )
```

```
TP2_brain_vols
```

```
## # A tibble: 18 x 2
##    SampleID   TP2
##    <chr>    <dbl>
##  1 A48P      536.
##  2 A49P      539.
##  3 A51P      523.
##  4 A52P      558.
##  5 A53P      526.
##  6 A54P      516.
##  7 A60P      562.
##  8 A61P      497.
##  9 A62P      547.
## 10 A63P      561.
## 11 A65P      494.
## 12 A66P      453.
## 13 A67P      491.
## 14 A68P      473.
## 15 A69P      601.
## 16 A70P      537.
## 17 A71P      541.
## 18 A74P      543.
```

```
TP1_brain_vols <- read_csv("TP1_brain_vols.csv",col_names = TRUE)
```

```
## Parsed with column specification:
## cols(
##   SampleID = col_character(),
##   TP1 = col_double()
## )
```

```
TP1_brain_vols
```

```
## # A tibble: 18 x 2
```

```
##    SampleID   TP1
##    <chr>    <dbl>
##  1 A48P      507.
##  2 A49P      534.
##  3 A51P      513.
##  4 A52P      538.
##  5 A53P      523.
##  6 A54P      501.
##  7 A60P      548.
##  8 A61P      478.
##  9 A62P      526.
## 10 A63P      532.
## 11 A65P      496.
## 12 A66P      420.
## 13 A67P      482.
## 14 A68P      475.
## 15 A69P      574.
## 16 A70P      512
## 17 A71P      531.
## 18 A74P      525.
```

Load data key then ensure column headings are correct and not placed as row 1.

```
Key <- read_csv("R_Data_Key.csv", col_names = T)
```

```
## Parsed with column specification:
## cols(
##   SampleID = col_character(),
##   Genotype = col_character()
## )
```

```
Key
```

```
## # A tibble: 18 x 2
##    SampleID Genotype
##    <chr>    <chr>
##  1 A48P     WT
##  2 A49P     WT
##  3 A51P     KO
##  4 A52P     KO
##  5 A53P     WT
##  6 A54P     WT
##  7 A60P     KO
##  8 A61P     KO
##  9 A62P     WT
## 10 A63P     KO
## 11 A65P     WT
## 12 A66P     WT
## 13 A67P     WT
## 14 A68P     WT
## 15 A69P     KO
## 16 A70P     WT
## 17 A71P     WT
## 18 A74P     KO
```

```
# If the headers of the Key are in Row 1 rather than in the column names then use the code below:
# names(Key) <- Key %>% slice(1) %>% unlist()
# Key <- Key %>% slice(-1)
```

Only keep the *first 4 characters of the SampleID* to enable merging the datasets of TP3 brain volumes and the data key by column (that is also why both contain the column name SampleID) #{r} `TP3_brain_vols$SampleID <- substr(TP3_brain_vols$SampleID, 0, 4) TP3_brain_vols` Merge data by SampleID column.

```
TP3_Genotyped_Brain_Vols <- left_join(TP3_brain_vols, Key, by = "SampleID", copy = FALSE, suffix = c(".
TP3_Genotyped_Brain_Vols
```

```
## # A tibble: 18 x 3
##    SampleID   TP3 Genotype
##    <chr>    <dbl> <chr>
##  1 A48P      535. WT
##  2 A49P      543. WT
##  3 A51P      526. KO
##  4 A52P      561. KO
##  5 A53P      519. WT
##  6 A54P      527. WT
##  7 A60P      568. KO
##  8 A61P      504. KO
##  9 A62P      543. WT
## 10 A63P      559. KO
## 11 A65P      499. WT
## 12 A66P      444. WT
## 13 A67P      480. WT
## 14 A68P      469. WT
## 15 A69P      602. KO
## 16 A70P      523. WT
## 17 A71P      549. WT
## 18 A74P      554. KO
```

```
TP2_Genotyped_Brain_Vols <- left_join(TP2_brain_vols, Key, by = "SampleID", copy = FALSE, suffix = c(".
TP2_Genotyped_Brain_Vols
```

```
## # A tibble: 18 x 3
##    SampleID   TP2 Genotype
##    <chr>    <dbl> <chr>
##  1 A48P      536. WT
##  2 A49P      539. WT
##  3 A51P      523. KO
##  4 A52P      558. KO
##  5 A53P      526. WT
##  6 A54P      516. WT
##  7 A60P      562. KO
##  8 A61P      497. KO
##  9 A62P      547. WT
## 10 A63P      561. KO
## 11 A65P      494. WT
## 12 A66P      453. WT
## 13 A67P      491. WT
## 14 A68P      473. WT
## 15 A69P      601. KO
```

```
## 16 A70P      537. WT
## 17 A71P      541. WT
## 18 A74P      543. KO
```

```
TP1_Genotyped_Brain_Vols <- left_join(TP1_brain_vols, Key, by = "SampleID", copy = FALSE, suffix = c(".x
TP1_Genotyped_Brain_Vols
```

```
## # A tibble: 18 x 3
##    SampleID   TP1 Genotype
##    <chr>    <dbl> <chr>
##  1 A48P      507. WT
##  2 A49P      534. WT
##  3 A51P      513. KO
##  4 A52P      538. KO
##  5 A53P      523. WT
##  6 A54P      501. WT
##  7 A60P      548. KO
##  8 A61P      478. KO
##  9 A62P      526. WT
## 10 A63P      532. KO
## 11 A65P      496. WT
## 12 A66P      420. WT
## 13 A67P      482. WT
## 14 A68P      475. WT
## 15 A69P      574. KO
## 16 A70P      512  WT
## 17 A71P      531. WT
## 18 A74P      525. KO
```

Now the genotyped data is in one dataset - filter the WT and PKO into separate two tibbles at each timepoint.

```
Control_TP3 <- filter(TP3_Genotyped_Brain_Vols, Genotype == "WT")
PKO_TP3 <- filter(TP3_Genotyped_Brain_Vols, Genotype == "KO")

Control_TP2 <- filter(TP2_Genotyped_Brain_Vols, Genotype == "WT")
PKO_TP2 <- filter(TP2_Genotyped_Brain_Vols, Genotype == "KO")

Control_TP1 <- filter(TP1_Genotyped_Brain_Vols, Genotype == "WT")
PKO_TP1 <- filter(TP1_Genotyped_Brain_Vols, Genotype == "KO")
```

Save all tibbles as CSVs for my records. #“‘{r} write_csv(Control_TP3, “Control_TP3_vols.csv”) write_csv(PKO_TP3, “PKO_TP3_vols.csv”)

write_csv(Control_TP2, “Control_TP2_vols.csv”) write_csv(PKO_TP2, “PKO_TP2_vols.csv”)

write_csv(Control_TP1, “Control_TP1_vols.csv”) write_csv(PKO_TP1, “PKO_TP1_vols.csv”)

```
T-tests have two main assumptions:
1. The datasets follow a normal distribution
2. The datasets contain equal variance

Shapiro-Wilk normaility test was done.


```r
shapiro.test(PKO_TP3$TP3)
```
```

```
##
##  Shapiro-Wilk normality test
##
## data:  PKO_TP3$TP3
## W = 0.9545, p-value = 0.7704
```

```
shapiro.test(Control_TP3$TP3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Control_TP3$TP3
## W = 0.89557, p-value = 0.1631
```

```
shapiro.test(PKO_TP2$TP2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  PKO_TP2$TP2
## W = 0.9644, p-value = 0.8555
```

```
shapiro.test(Control_TP2$TP2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Control_TP2$TP2
## W = 0.8823, p-value = 0.1113
```

```
shapiro.test(PKO_TP1$TP1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  PKO_TP1$TP1
## W = 0.97731, p-value = 0.9455
```

```
shapiro.test(Control_TP1$TP1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Control_TP1$TP1
## W = 0.86512, p-value = 0.06708
```

From the outputs, the all the p-value > 0.05 implying that the distribution of the datasets are not significantly different from normal distribution. So we can assume the normality.

Then F-test was used to test for equal variance.

```
var.test(Control_TP3$TP3, PKO_TP3$TP3, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  Control_TP3$TP3 and PKO_TP3$TP3
## F = 1.2219, num df = 10, denom df = 6, p-value = 0.8403
## alternative hypothesis: true ratio of variances is not equal to 1
```

```
## 95 percent confidence interval:
##  0.2237354 4.9757034
## sample estimates:
## ratio of variances
##           1.221892
```

```
var.test(Control_TP2$TP2, PKO_TP2$TP2, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  Control_TP2$TP2 and PKO_TP2$TP2
## F = 0.91501, num df = 10, denom df = 6, p-value = 0.8571
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1675438 3.7260437
## sample estimates:
## ratio of variances
##          0.9150107
```

```
var.test(Control_TP1$TP1, PKO_TP1$TP1, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  Control_TP1$TP1 and PKO_TP1$TP1
## F = 1.2075, num df = 10, denom df = 6, p-value = 0.8523
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2211016 4.9171277
## sample estimates:
## ratio of variances
##           1.207507
```

All the p-value $>$ 0.05 implying there is no significant difference between the two variances.

Thus, a t-test was conducted between the PKO and Control datasets at each timepoint. The test was specified as unpaired as the data came from subjects in two distinct groups. The variance was set to equal due to the result of the F-test (aka var.test).

```
 t.test(Control_TP3$TP3, PKO_TP3$TP3, paired=FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Control_TP3$TP3 and PKO_TP3$TP3
## t = -2.5872, df = 16, p-value = 0.01985
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -75.559924  -7.500439
## sample estimates:
## mean of x mean of y
##  511.9898  553.5200
```

```
 t.test(Control_TP2$TP2, PKO_TP2$TP2, paired=FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
```

```
##
## data:  Control_TP2$TP2 and PKO_TP2$TP2
## t = -2.2711, df = 16, p-value = 0.0373
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -67.906740  -2.338663
## sample estimates:
## mean of x mean of y
##  514.0247  549.1474
```

```
t.test(Control_TP1$TP1, PKO_TP1$TP1, paired=FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  Control_TP1$TP1 and PKO_TP1$TP1
## t = -1.8865, df = 16, p-value = 0.07751
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -61.757902   3.598146
## sample estimates:
## mean of x mean of y
##  500.5407  529.6206
```

Join up datasets

```
All_Genotyped_Brain_Vols <- left_join(TP3_Genotyped_Brain_Vols, TP2_Genotyped_Brain_Vols, by = "SampleI
All_Genotyped_Brain_Vols <- left_join(All_Genotyped_Brain_Vols, TP1_Genotyped_Brain_Vols, by = "SampleI
All_Genotyped_Brain_Vols
```

```
## # A tibble: 18 x 7
##    SampleID   TP3 Genotype.x   TP2 Genotype.y   TP1 Genotype
##    <chr>    <dbl> <chr>      <dbl> <chr>      <dbl> <chr>
##  1 A48P      535. WT          536. WT          507. WT
##  2 A49P      543. WT          539. WT          534. WT
##  3 A51P      526. KO          523. KO          513. KO
##  4 A52P      561. KO          558. KO          538. KO
##  5 A53P      519. WT          526. WT          523. WT
##  6 A54P      527. WT          516. WT          501. WT
##  7 A60P      568. KO          562. KO          548. KO
##  8 A61P      504. KO          497. KO          478. KO
##  9 A62P      543. WT          547. WT          526. WT
## 10 A63P      559. KO          561. KO          532. KO
## 11 A65P      499. WT          494. WT          496. WT
## 12 A66P      444. WT          453. WT          420. WT
## 13 A67P      480. WT          491. WT          482. WT
## 14 A68P      469. WT          473. WT          475. WT
## 15 A69P      602. KO          601. KO          574. KO
## 16 A70P      523. WT          537. WT          512  WT
## 17 A71P      549. WT          541. WT          531. WT
## 18 A74P      554. KO          543. KO          525. KO
```

```
# Remove the duplicated genotype columns
All_Genotyped_Brain_Vols <- All_Genotyped_Brain_Vols[-c(3,5)]
All_Genotyped_Brain_Vols
```

```
## # A tibble: 18 x 5
```

```
##    SampleID   TP3   TP2   TP1 Genotype
##    <chr>    <dbl> <dbl> <dbl> <chr>
##  1 A48P      535.  536.  507. WT
##  2 A49P      543.  539.  534. WT
##  3 A51P      526.  523.  513. KO
##  4 A52P      561.  558.  538. KO
##  5 A53P      519.  526.  523. WT
##  6 A54P      527.  516.  501. WT
##  7 A60P      568.  562.  548. KO
##  8 A61P      504.  497.  478. KO
##  9 A62P      543.  547.  526. WT
## 10 A63P      559.  561.  532. KO
## 11 A65P      499.  494.  496. WT
## 12 A66P      444.  453.  420. WT
## 13 A67P      480.  491.  482. WT
## 14 A68P      469.  473.  475. WT
## 15 A69P      602.  601.  574. KO
## 16 A70P      523.  537.  512  WT
## 17 A71P      549.  541.  531. WT
## 18 A74P      554.  543.  525. KO
```

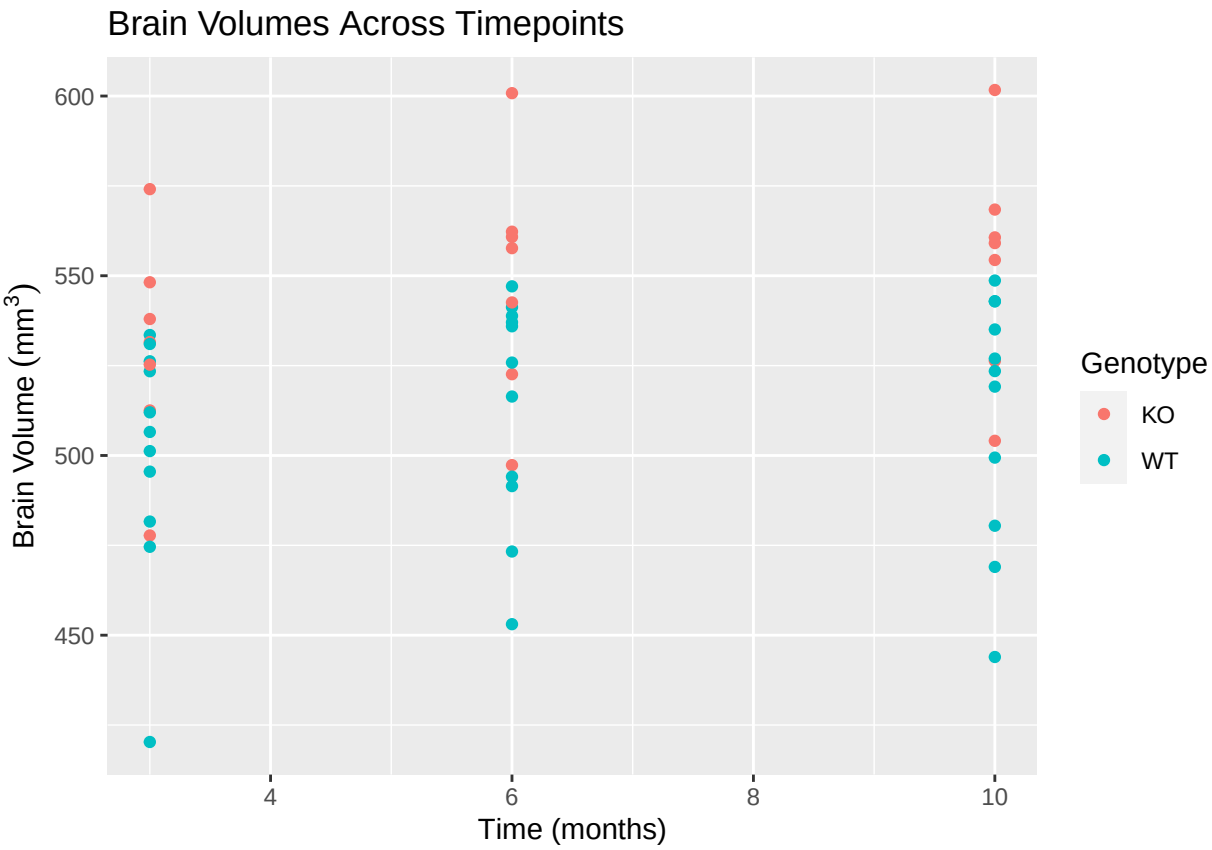
```
# write_csv(All_Genotyped_Brain_Vols , "All_Genotyped_Brain_Vols.csv")
```

Now the dataset is plotted in different ways to visualise the differences between the genotypes, timepoints and also the general trends.

```
library(ggplot2)
require(ggplot2)
plot_brain_vols <- read_csv("Plotting_brain_vols.csv",col_names = TRUE)
```

```
## Parsed with column specification:
## cols(
##   SampleID = col_character(),
##   Time_months = col_double(),
##   Brain_Volume_mm3 = col_double(),
##   Genotype = col_character()
## )
```

```
p <- ggplot(data = plot_brain_vols, aes(x = plot_brain_vols$Time_months, y = plot_brain_vols$Brain_Volu
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  ggtitle("Brain Volumes Across Timepoints") +
  scale_colour_discrete("Genotype")
p + geom_point(aes(colour = factor (plot_brain_vols$Genotype)))
```



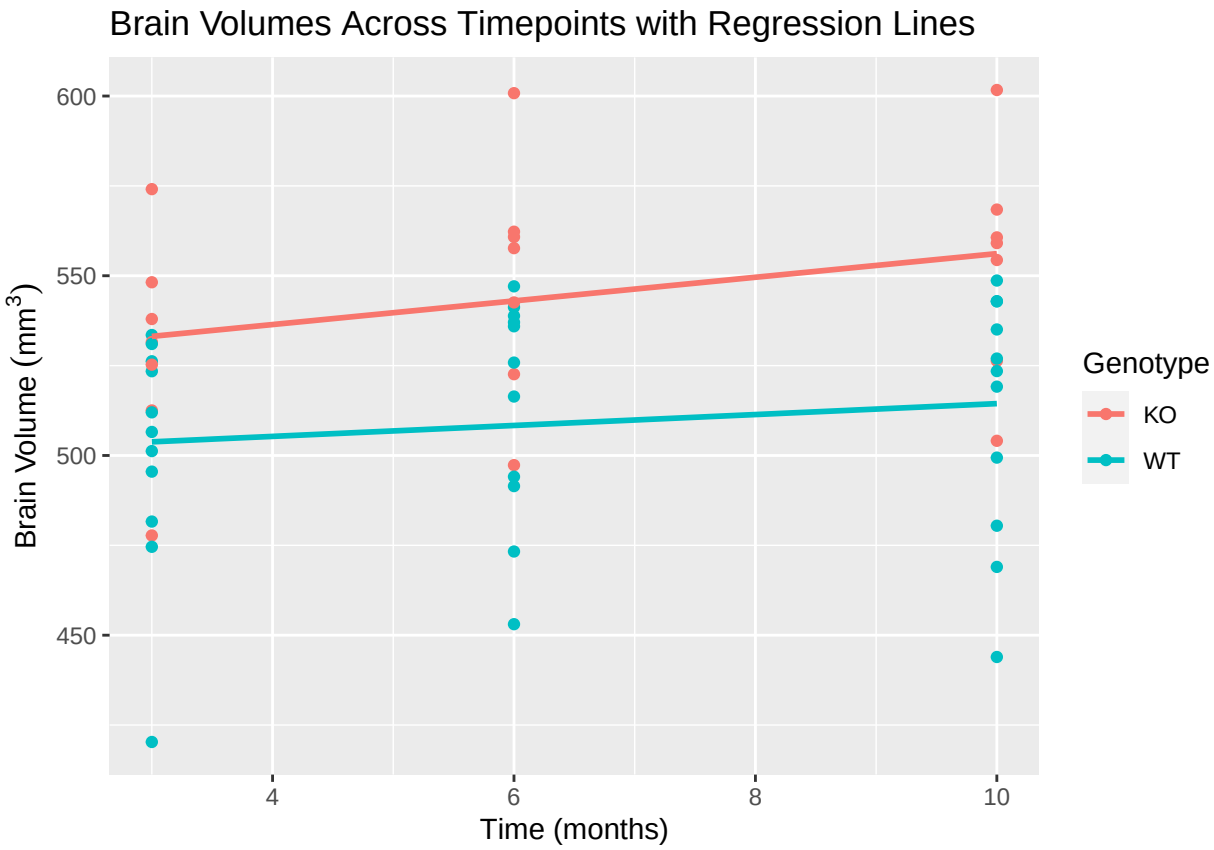

```
# ggsave("P0 Brain Volumes Across Timepoints.png")
```

```
library(ggplot2)
require(ggplot2)
plot_brain_vols <- read_csv("Plotting_brain_vols.csv",col_names = TRUE)
```

```
## Parsed with column specification:
## cols(
##   SampleID = col_character(),
##   Time_months = col_double(),
##   Brain_Volume_mm3 = col_double(),
##   Genotype = col_character()
## )
```

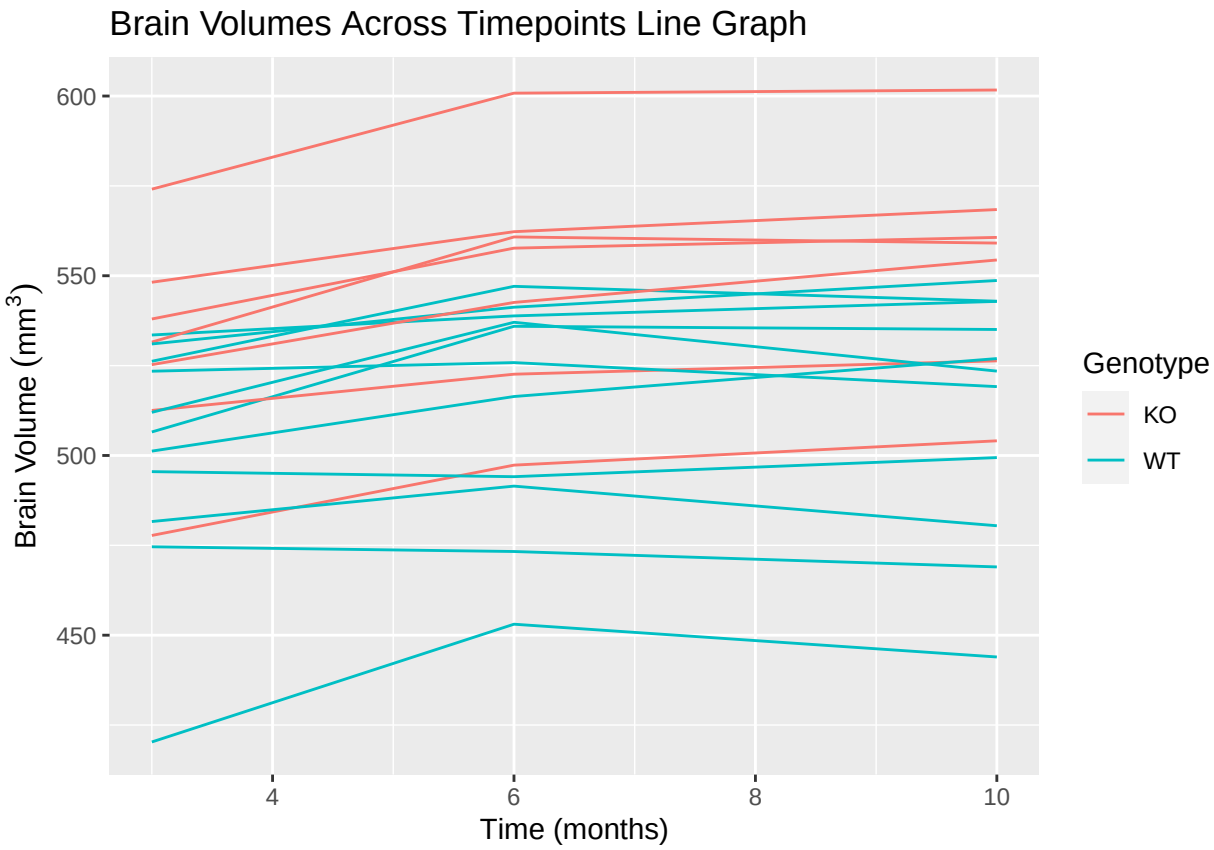
```
p3 <- ggplot(data = plot_brain_vols, aes(x = plot_brain_vols$Time_months, y = plot_brain_vols$Brain_Volu
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  ggtitle("Brain Volumes Across Timepoints with Regression Lines") +
  scale_colour_discrete("Genotype")
p3 + geom_point(aes(colour = factor (plot_brain_vols$Genotype))) +
  geom_smooth(method = "lm",
                    se = FALSE,
                    aes(x = Time_months, y = Brain_Volume_mm3, group = Genotype, color = factor(Genotype
```

```
## `geom_smooth()` using formula 'y ~ x'
```



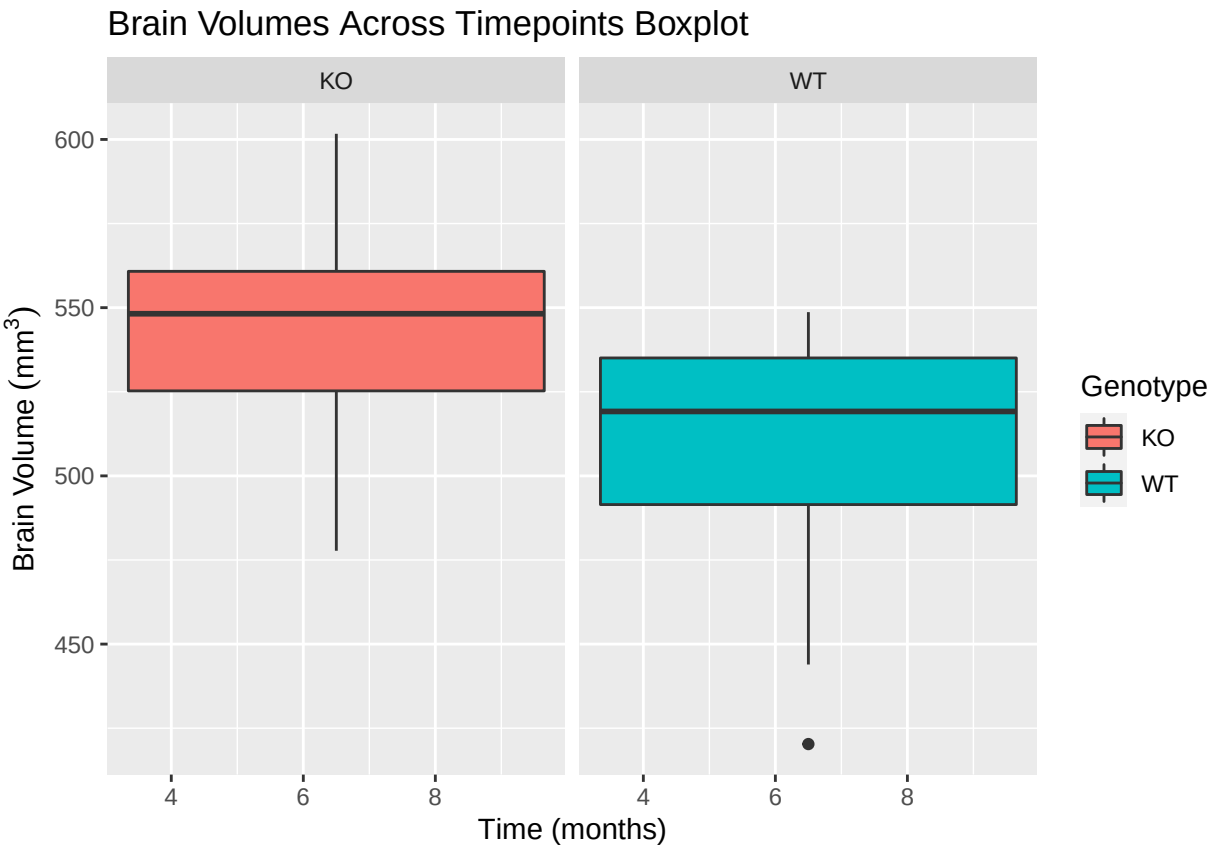

```
# ggsave("P1 Brain Volumes Across Timepoints with Regression Lines.png")
```

```
p1 <- ggplot(data = plot_brain_vols, aes(x = plot_brain_vols$Time_months, y = plot_brain_vols$Brain_Volu
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  ggtitle("Brain Volumes Across Timepoints Line Graph") +
  scale_colour_discrete("Genotype")
p1 + geom_line(aes(colour = factor (plot_brain_vols$Genotype)))
```

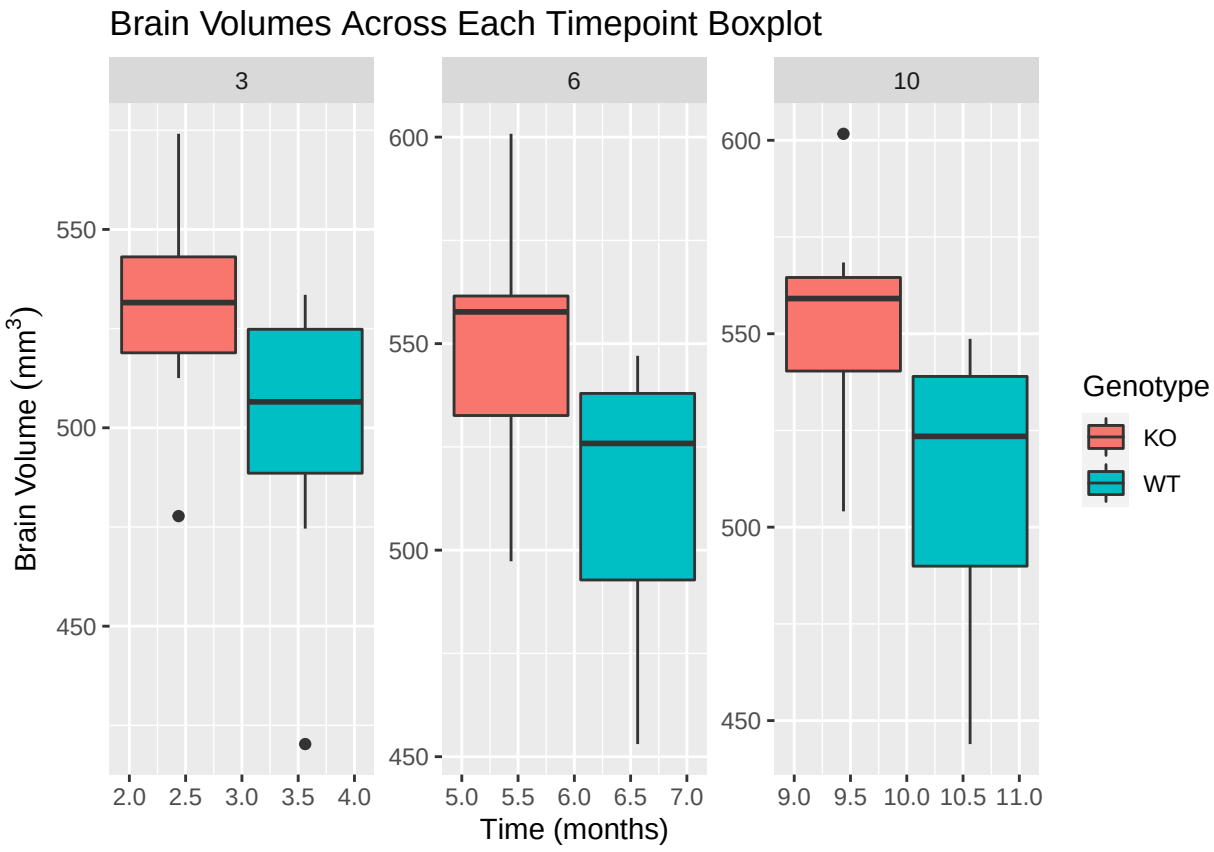



```
# ggsave("P2 Brain Volumes Across Timepoints Line Graph.png")
```

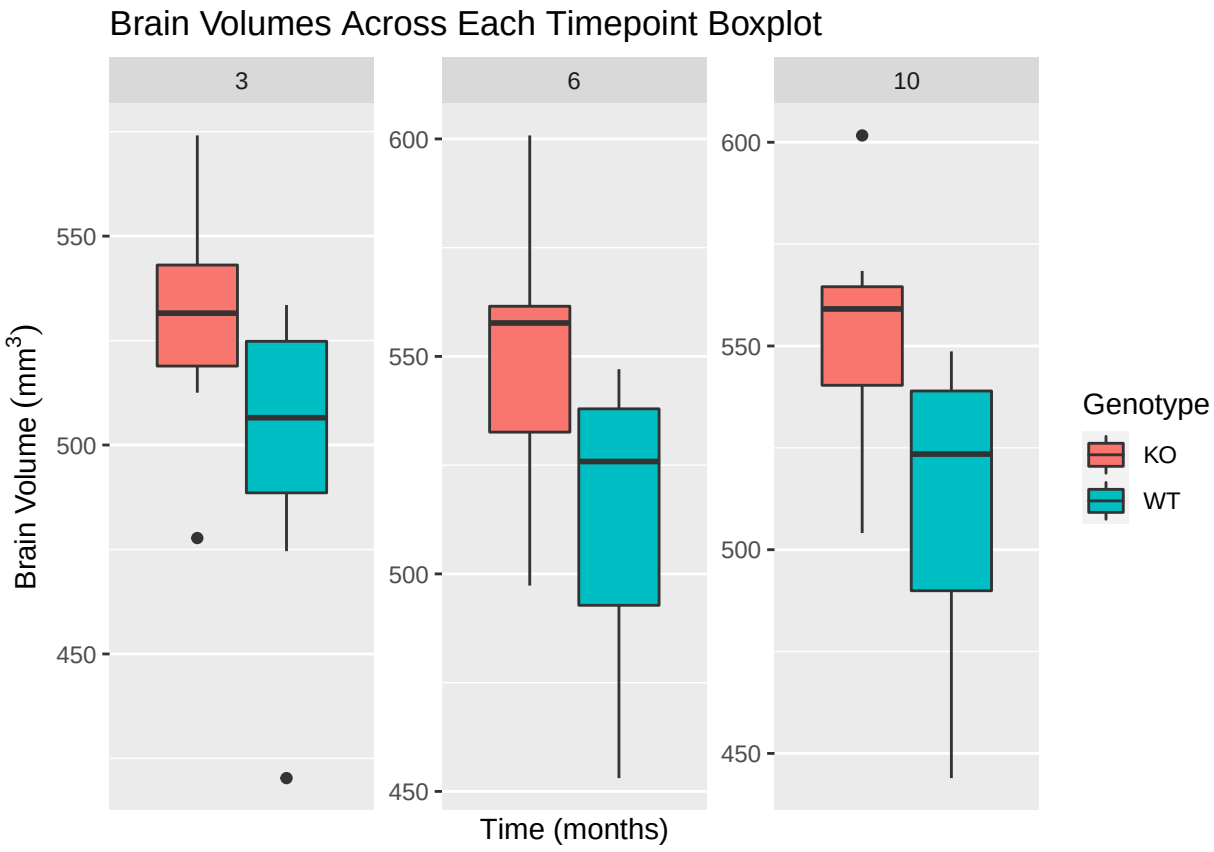
```
p2 <- ggplot(plot_brain_vols, aes(x=Time_months, y=Brain_Volume_mm3, fill=Genotype)) +
    geom_boxplot() +
    facet_wrap(~Genotype) +
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  ggtitle("Brain Volumes Across Timepoints Boxplot")
# ggsave("P3 Brain Volumes Across Timepoints Boxplot.png")
p2
```



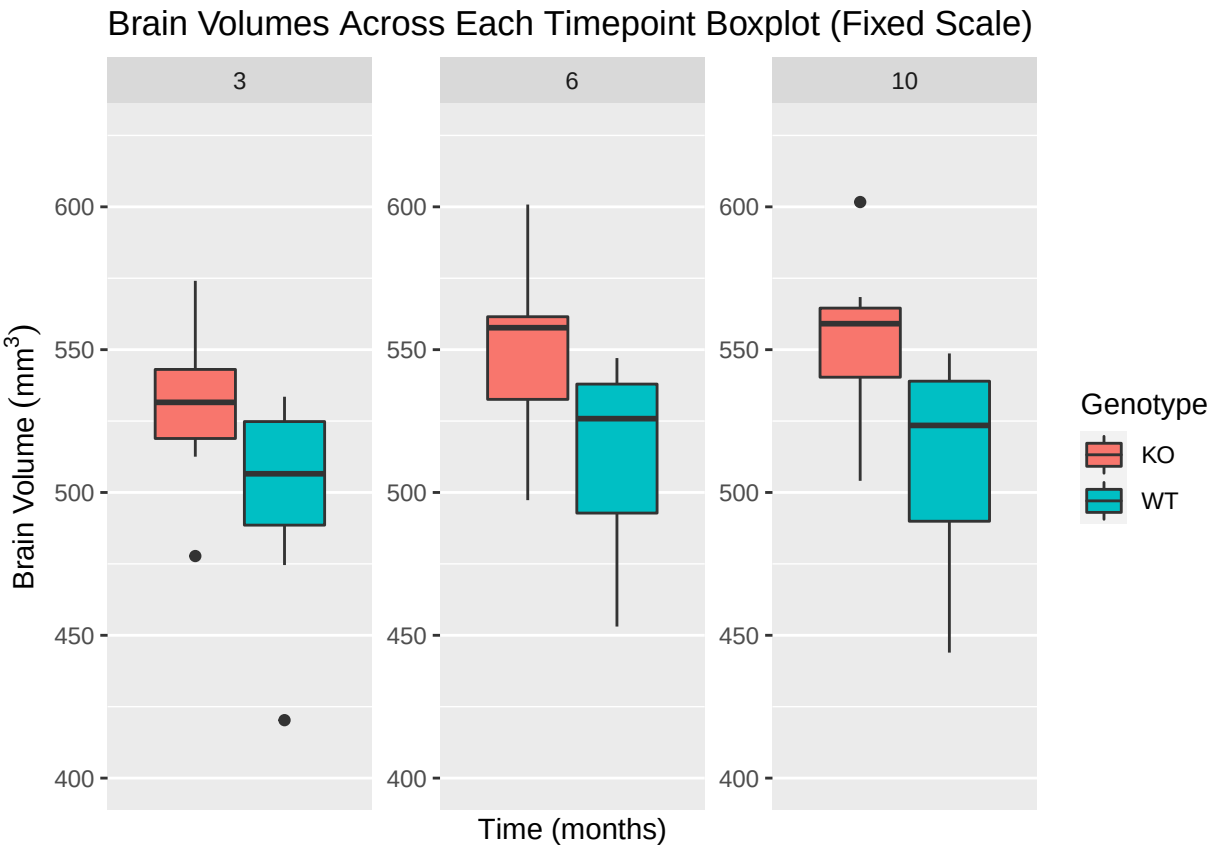

```
p4 <- ggplot(plot_brain_vols, aes(x = plot_brain_vols$Time_months, y = plot_brain_vols$Brain_Volume_mm3
    geom_boxplot() +
    facet_wrap(~Time_months, scale="free") +
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  ggtitle("Brain Volumes Across Each Timepoint Boxplot")
# ggsave("P4 Brain Volumes Across Each Timepoint Boxplot.png")
p4
```




```
p5 <- ggplot(plot_brain_vols, aes(x = plot_brain_vols$Time_months, y = plot_brain_vols$Brain_Volume_mm3
    geom_boxplot() +
    facet_wrap(~Time_months, scale="free") +
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  scale_x_discrete("Time (months)") +
  ggtitle("Brain Volumes Across Each Timepoint Boxplot")
# ggsave("P5 Brain Volumes At Each Timepoint Boxplot.png")
p5
```




```
p6 <- ggplot(plot_brain_vols, aes(x = plot_brain_vols$Time_months, y = plot_brain_vols$Brain_Volume_mm3
    geom_boxplot() +
    facet_wrap(~Time_months, scale="free") +
  ylab(expression("Brain Volume"~(mm^3))) +
  xlab("Time (months)") +
  scale_x_discrete("Time (months)") +
  scale_y_continuous(limits = c(400, 625)) +
  ggtitle("Brain Volumes Across Each Timepoint Boxplot (Fixed Scale)")
# ggsave("P6 Brain Volumes At Each Timepoint Boxplot (fixed scales).png")
p6
```




```
library(tidyverse)
library(ggpubr)
library(rstatix)
```

Summary statistics:

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
plot_brain_vols %>%
  group_by(Time_months, Genotype) %>%
  get_summary_stats(Brain_Volume_mm3, type = "mean_sd")
```

```
## # A tibble: 6 x 6
##   Time_months Genotype variable             n  mean    sd
##         <dbl> <chr>    <chr>            <dbl> <dbl> <dbl>
## 1           3 KO       Brain_Volume_mm3     7  530.  30.0
## 2           3 WT       Brain_Volume_mm3    11  501.  33.0
## 3           6 KO       Brain_Volume_mm3     7  549.  32.9
## 4           6 WT       Brain_Volume_mm3    11  514.  31.4
## 5          10 KO       Brain_Volume_mm3     7  554.  31.1
## 6          10 WT       Brain_Volume_mm3    11  512.  34.4
```

***Check assuptions for ANOVA* (https://www.datanovia.com/en/lessons/mixed-anova-in-r/):** No significant outliers in any cell of the design. This can be checked by visualizing the data using box plot methods and by using the function identify_outliers() [rstatix package]. Normality: the outcome (or dependent) variable should be approximately normally distributed in each cell of the design. This can be checked using the Shapiro-Wilk normality test (shapiro_test() [rstatix]) or by visual inspection using QQ plot (ggqqplot() [ggpubr package]). Homogeneity of variances: the variance of the outcome variable should be equal between the groups of the between-subjects factors. This can be assessed using the Levene's test for equality of variances (levene_test() [rstatix]). Assumption of sphericity: the variance of the differences between within-subjects groups should be equal. This can be checked using the Mauchly's test of sphericity, which is automatically reported when using the anova_test() R function. Homogeneity of covariances tested by Box's M. The covariance matrices should be equal across the cells formed by the between-subjects factors.

```
plot_brain_vols %>%
  group_by(Time_months, Genotype) %>%
  identify_outliers(Brain_Volume_mm3)
```

```
## # A tibble: 3 x 6
##   Time_months Genotype SampleID Brain_Volume_mm3 is.outlier is.extreme
##         <dbl> <chr>    <chr>               <dbl> <lgl>      <lgl>
## 1           3 KO       A61P                 478. TRUE       FALSE
## 2           3 WT       A66P                 420. TRUE       FALSE
## 3          10 KO       A69P                 602. TRUE       FALSE
```

Three outliers identified so an ANOVA should be done on one complete dataset and one without the outliers. Now (as before) double check for normality (code below is the quick way).

```
plot_brain_vols %>%
  group_by(Time_months, Genotype) %>%
  shapiro_test(Brain_Volume_mm3)
```

```
## # A tibble: 6 x 5
##   Time_months Genotype variable         statistic      p
##         <dbl> <chr>    <chr>                <dbl>  <dbl>
## 1           3 KO       Brain_Volume_mm3     0.977 0.945
## 2           3 WT       Brain_Volume_mm3     0.865 0.0671
## 3           6 KO       Brain_Volume_mm3     0.964 0.856
## 4           6 WT       Brain_Volume_mm3     0.882 0.111
## 5          10 KO       Brain_Volume_mm3     0.955 0.770
## 6          10 WT       Brain_Volume_mm3     0.896 0.163
```

All p-values are above 0.05 so normality can be assumed.

Homogeneity of covariances assumption:

```
box_m(plot_brain_vols[, "Brain_Volume_mm3", drop = FALSE], plot_brain_vols$Genotype)
```

```
## # A tibble: 1 x 4
##   statistic p.value parameter method
##       <dbl>   <dbl>     <dbl> <chr>
## 1    0.0183   0.892         1 Box's M-test for Homogeneity of Covariance Matric~
```

There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices ($p > 0.001$).

ANOVA The assumption of sphericity will be automatically checked during the computation of the ANOVA test using the R function anova_test() [rstatix package]. The Mauchly's test is internally used to assess the sphericity assumption. By using the function get_anova_table() [rstatix] to extract the ANOVA table,

the Greenhouse-Geisser sphericity correction is automatically applied to factors violating the sphericity assumption.

```
# Two-way mixed ANOVA test
res.aov <- anova_test(
  data = plot_brain_vols, dv = Brain_Volume_mm3, wid = SampleID,
  between = Genotype, within = Time_months
  )
get_anova_table(res.aov)
```

```
## ANOVA Table (type III tests)
##
##                 Effect DFn DFd      F       p p<.05   ges
## 1             Genotype   1  16  5.217 3.6e-02     * 0.241
## 2          Time_months   2  32 38.782 2.8e-09     * 0.062
## 3 Genotype:Time_months   2  32  3.847 3.2e-02     * 0.007
```

There is a statistically significant two-way interaction between genotype and time (months) on whole brain volume, F(32, 3.847) = 110.18, p = 0.032, p < 0.05.

Interpreting the results: From the ANOVA results, we can conclude the following, based on the p-values and a significance level of 0.05:

> the p-value of Genotype is 0.036 (significant), which indicates that the genoype is associated with significant different brain weight. the p-value of Time (months) is 2.8e-09 (significant), which indicates that the age of the mice are associated with significant different brain weight. the p-value for the interaction between gneotype and time (months) is 0.032 (significant), which indicates that the relationships between time (months) and brain weight depends on the genotype.

http://www.sthda.com/english/wiki/two-way-anova-test-in-r

#####Post-hoc tests for significant two-way interaction

```
# Effect of genotype at each time point
one.way <- plot_brain_vols %>%
  group_by(Time_months) %>%
  anova_test(dv = Brain_Volume_mm3, wid = SampleID, between = Genotype) %>%
  get_anova_table() %>%
  adjust_pvalue(method = "bonferroni")
```

```
## Coefficient covariances computed by hccm()
## Coefficient covariances computed by hccm()
## Coefficient covariances computed by hccm()
```

```
one.way
```

```
## # A tibble: 3 x 9
##   Time_months Effect     DFn   DFd     F     p `p<.05`   ges p.adj
##         <dbl> <chr>    <dbl> <dbl> <dbl> <dbl> <chr>   <dbl> <dbl>
## 1           3 Genotype     1    16  3.56 0.078 ""      0.182 0.234
## 2           6 Genotype     1    16  5.16 0.037 "*"     0.244 0.111
## 3          10 Genotype     1    16  6.69 0.02  "*"     0.295 0.06
```

```
# Pairwise comparisons between genotypes
pwc <- plot_brain_vols %>%
  group_by(Time_months) %>%
  pairwise_t_test(Brain_Volume_mm3 ~ Genotype, p.adjust.method = "bonferroni")
pwc
```

```
## # A tibble: 3 x 10
```

```
##   Time_months .y.   group1 group2    n1    n2      p p.signif  p.adj
## *       <dbl> <chr> <chr>  <chr>  <int> <int>  <dbl> <chr>     <dbl>
## 1           3 Brai~ KO     WT         7    11 0.0775 ns       0.0775
## 2           6 Brai~ KO     WT         7    11 0.0373 *        0.0373
## 3          10 Brai~ KO     WT         7    11 0.0199 *        0.0199
## # ... with 1 more variable: p.adj.signif <chr>
```

```
# Effect of time at each genotype
one.way2 <- plot_brain_vols %>%
  group_by(Genotype) %>%
  anova_test(dv = Brain_Volume_mm3, wid = SampleID, within = Time_months) %>%
  get_anova_table() %>%
  adjust_pvalue(method = "bonferroni")
one.way2
```

```
## # A tibble: 2 x 9
##   Genotype Effect        DFn   DFd     F           p `p<.05`   ges       p.adj
##   <chr>    <chr>       <dbl> <dbl> <dbl>       <dbl> <chr>   <dbl>       <dbl>
## 1 KO       Time_months     2    12 72.5  0.000000199 *       0.114 0.000000398
## 2 WT       Time_months     2    20  9.76 0.001       *       0.034 0.002
```

```
# Pairwise comparisons between time points at each genotype
# Paired t-test is used because we have repeated measures by time
pwc2 <- plot_brain_vols %>%
  group_by(Genotype) %>%
  pairwise_t_test(
    Brain_Volume_mm3 ~ Time_months, paired = TRUE,
    p.adjust.method = "bonferroni"
    ) %>%
  select(-df, -statistic, -p) # Remove details
pwc2
```

```
## # A tibble: 6 x 8
##   Genotype .y.              group1 group2    n1    n2     p.adj p.adj.signif
##   <chr>    <chr>            <chr>  <chr>  <int> <int>     <dbl> <chr>
## 1 KO       Brain_Volume_mm3 3      6          7     7 0.000753  ***
## 2 KO       Brain_Volume_mm3 3      10         7     7 0.0000726 ****
## 3 KO       Brain_Volume_mm3 6      10         7     7 0.117     ns
## 4 WT       Brain_Volume_mm3 3      6         11    11 0.013     *
## 5 WT       Brain_Volume_mm3 3      10        11    11 0.032     *
## 6 WT       Brain_Volume_mm3 6      10        11    11 1         ns
```

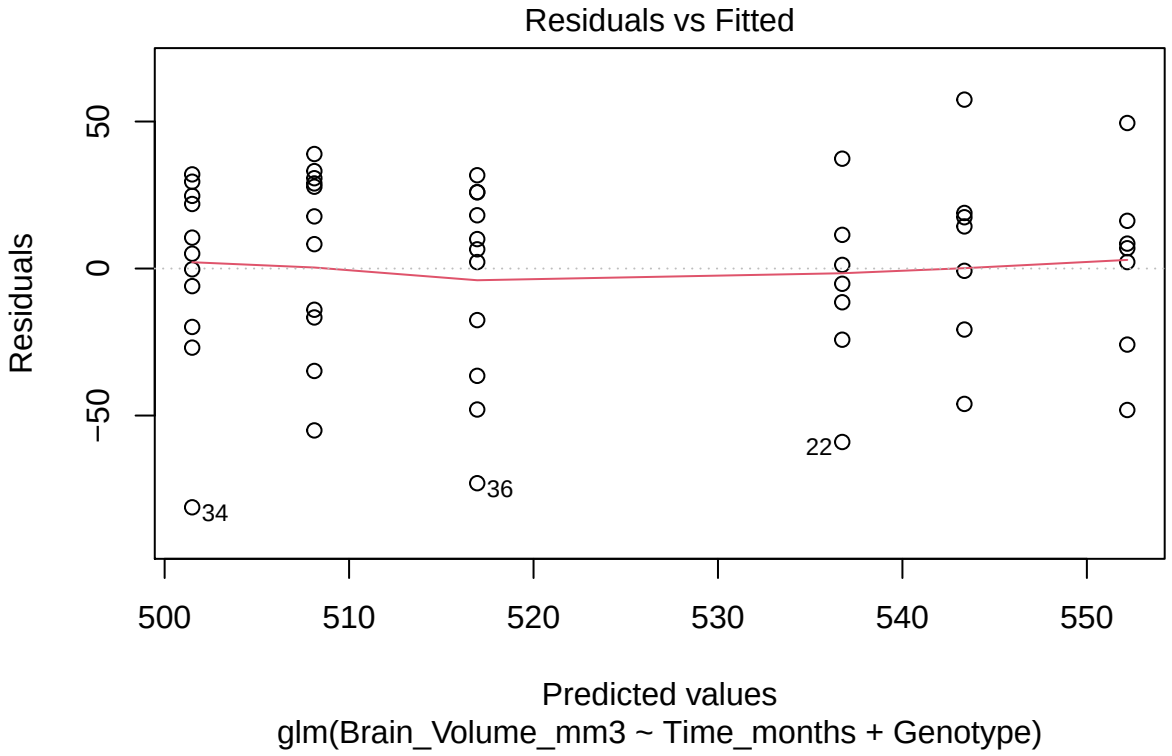
```
library(jtools)
model <- glm(formula= Brain_Volume_mm3 ~ Time_months + Genotype, data=plot_brain_vols, family=gaussian()
summary (model)
```

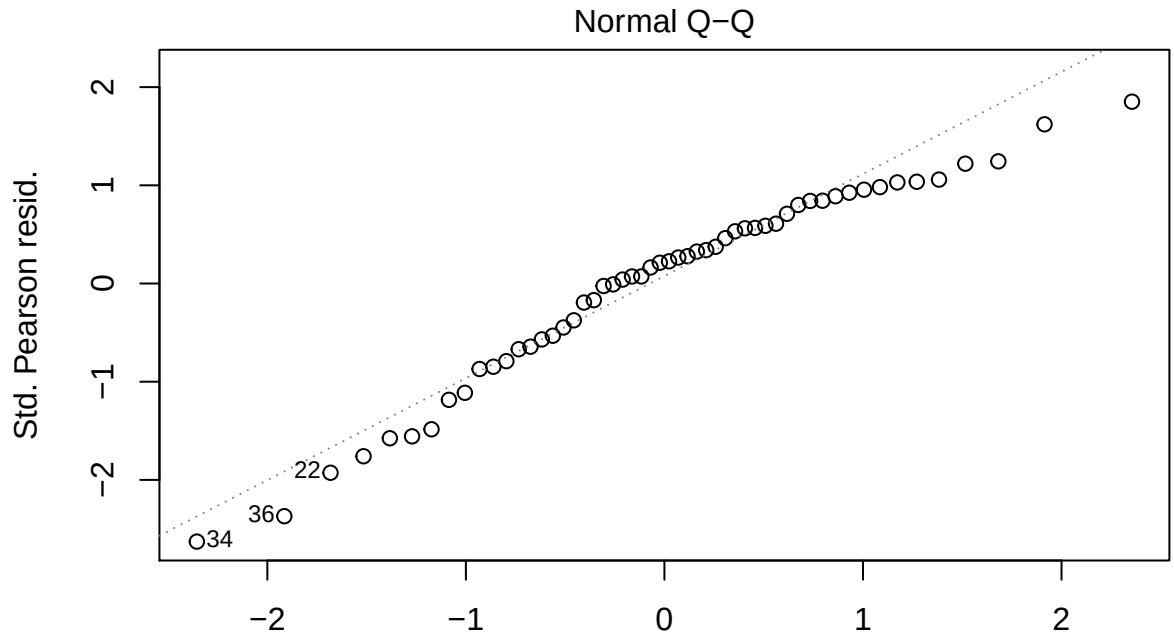
GLM of Whole Dataset

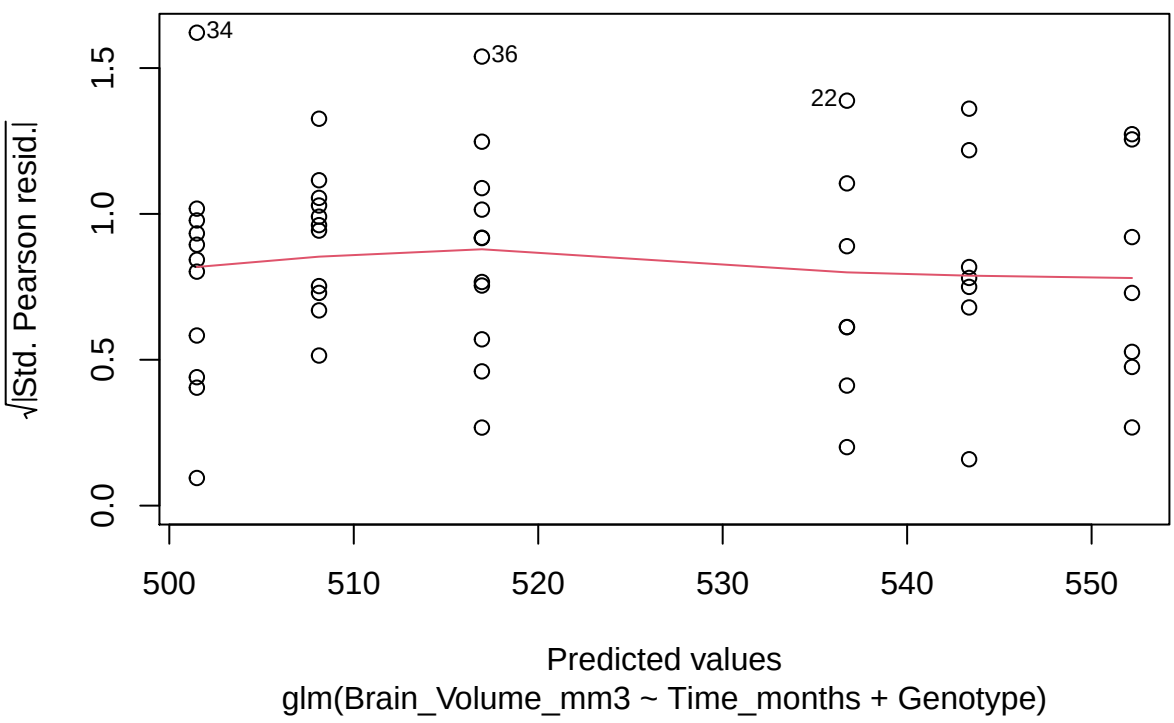
```
##
## Call:
## glm(formula = Brain_Volume_mm3 ~ Time_months + Genotype, family = gaussian(),
##     data = plot_brain_vols)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

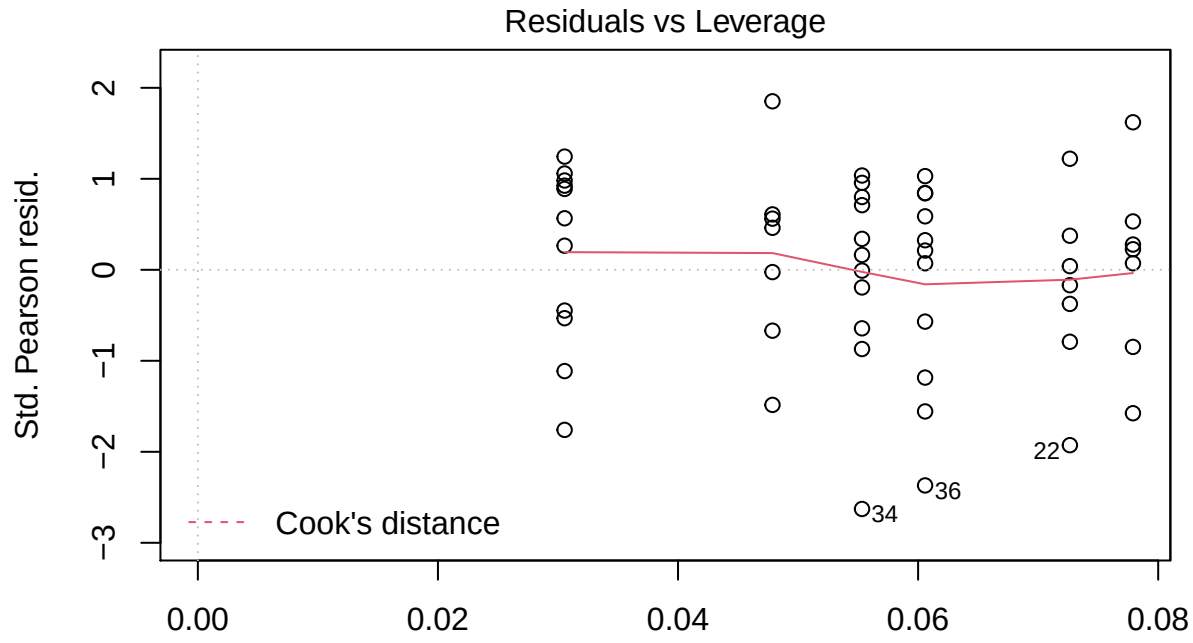
```
## -81.205  -19.292    6.711   24.023   57.440
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  530.114     11.807  44.898  < 2e-16 ***
## Time_months    2.208      1.509   1.463 0.149503
## GenotypeWT   -35.244      8.874  -3.972 0.000225 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1010.499)
##
##     Null deviance: 69640  on 53  degrees of freedom
## Residual deviance: 51535  on 51  degrees of freedom
## AIC: 531.74
##
## Number of Fisher Scoring iterations: 2
```

```
plot(model)
```

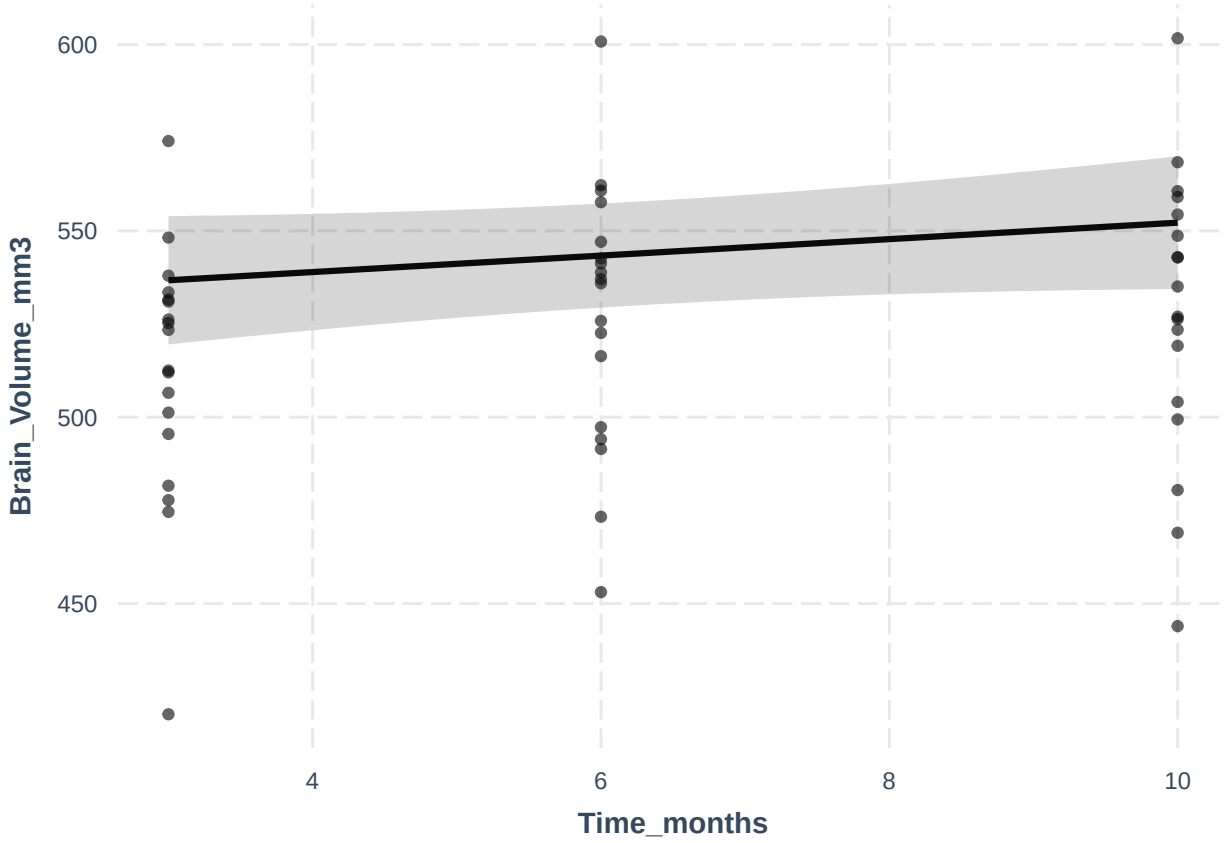



Normal Q−Q
Std. Pearson resid.
Theoretical Quantiles
glm(Brain_Volume_mm3 ~ Time_months + Genotype)
34
36
22


Scale−Location
√|Std. Pearson resid.|
Predicted values
glm(Brain_Volume_mm3 ~ Time_months + Genotype)
34
36
22

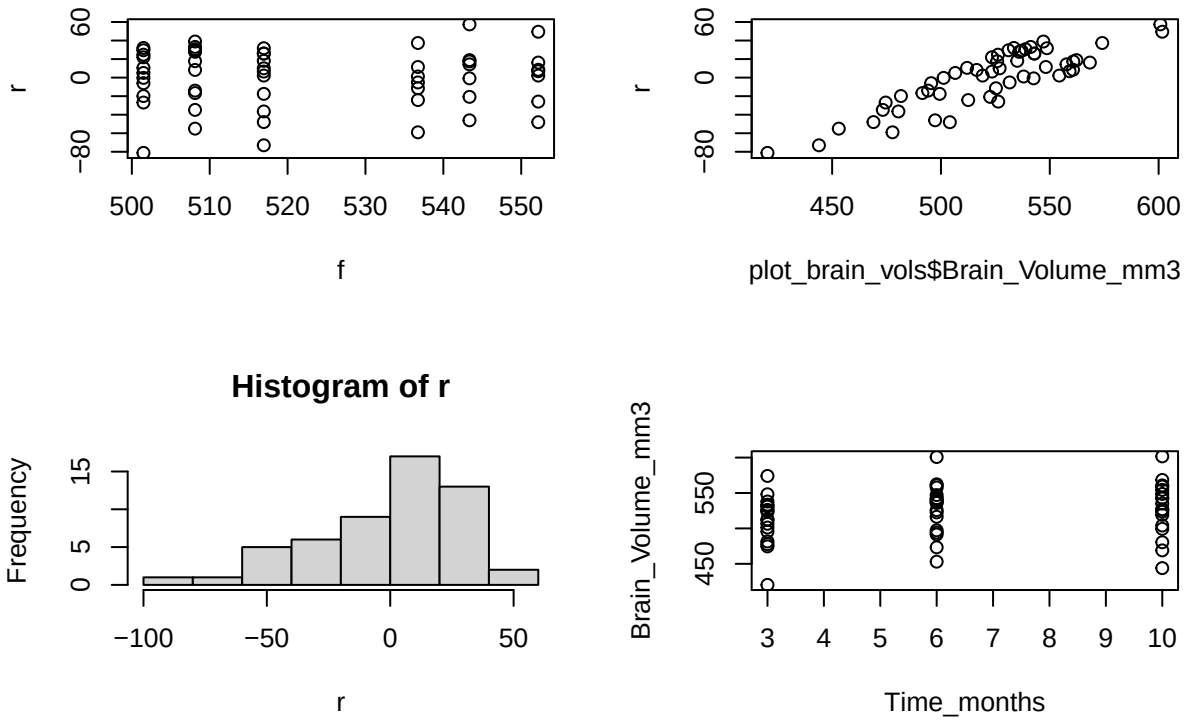



```
effect_plot(model, pred = Time_months, interval = TRUE, plot.points = TRUE)
```



```r
#model validation plot, residuals vs fitted values, residuals vs predictors and histogramm of #residual
 r<-residuals(model)
 f<-fitted.values(model)
 par(mfrow=c(2,2))
 plot(r~f)
 plot(r~plot_brain_vols$Brain_Volume_mm3)
 hist(r)

#add a plot
 plot(Brain_Volume_mm3~Time_months,plot_brain_vols)
```




```r
# Remove rows containing outliers
plot_brain_vols2 <- plot_brain_vols[-c(22,34,45),]
plot_brain_vols2
```

ANOVA Without Outliers

```
## # A tibble: 51 x 4
##    SampleID Time_months Brain_Volume_mm3 Genotype
##    <chr>          <dbl>            <dbl> <chr>
##  1 A48P               3             507. WT
##  2 A48P               6             536. WT
##  3 A48P              10             535. WT
##  4 A49P               3             534. WT
##  5 A49P               6             539. WT
##  6 A49P              10             543. WT
##  7 A51P               3             513. KO
##  8 A51P               6             523. KO
##  9 A51P              10             526. KO
## 10 A52P               3             538. KO
```

```
## # ... with 41 more rows
```

```
plot_brain_vols2 %>%
  group_by(Time_months, Genotype) %>%
  get_summary_stats(Brain_Volume_mm3, type = "mean_sd")
```

```
## # A tibble: 6 x 6
##   Time_months Genotype variable             n  mean    sd
##         <dbl> <chr>    <chr>            <dbl> <dbl> <dbl>
## 1           3 KO       Brain_Volume_mm3     6  538.  21.3
## 2           3 WT       Brain_Volume_mm3    10  509.  20.5
## 3           6 KO       Brain_Volume_mm3     7  549.  32.9
## 4           6 WT       Brain_Volume_mm3    11  514.  31.4
## 5          10 KO       Brain_Volume_mm3     6  545.  24.9
## 6          10 WT       Brain_Volume_mm3    11  512.  34.4
```

```
plot_brain_vols2 %>%
  group_by(Time_months, Genotype) %>%
  identify_outliers(Brain_Volume_mm3)
```

```
## # A tibble: 1 x 6
##   Time_months Genotype SampleID Brain_Volume_mm3 is.outlier is.extreme
##         <dbl> <chr>    <chr>               <dbl> <lgl>      <lgl>
## 1           3 KO       A69P                 574. TRUE       FALSE
```

```
plot_brain_vols2 %>%
  group_by(Time_months, Genotype) %>%
  shapiro_test(Brain_Volume_mm3)
```

```
## # A tibble: 6 x 5
##   Time_months Genotype variable         statistic     p
##         <dbl> <chr>    <chr>                <dbl> <dbl>
## 1           3 KO       Brain_Volume_mm3     0.961 0.826
## 2           3 WT       Brain_Volume_mm3     0.939 0.545
## 3           6 KO       Brain_Volume_mm3     0.964 0.856
## 4           6 WT       Brain_Volume_mm3     0.882 0.111
## 5          10 KO       Brain_Volume_mm3     0.852 0.165
## 6          10 WT       Brain_Volume_mm3     0.896 0.163
```

```
box_m(plot_brain_vols2[, "Brain_Volume_mm3", drop = FALSE], plot_brain_vols2$Genotype)
```

```
## # A tibble: 1 x 4
##   statistic p.value parameter method
##       <dbl>   <dbl>     <dbl> <chr>
## 1     0.212   0.646         1 Box's M-test for Homogeneity of Covariance Matric~
```

```
# Two-way mixed ANOVA test
res.aov2 <- anova_test(
  data = plot_brain_vols2, dv = Brain_Volume_mm3, wid = SampleID,
  between = Genotype, within = Time_months
  )
get_anova_table(res.aov2)
```

```
## ANOVA Table (type III tests)
##
##                 Effect DFn DFd      F        p p<.05   ges
## 1             Genotype   1  13  5.992 2.90e-02     * 0.303
## 2          Time_months   2  26 25.110 8.47e-07     * 0.102
```

```
## 3 Genotype:Time_months   2  26  2.961 6.90e-02       0.013
```

There is NOT a statistically significant two-way interactions between genotype and time (months) on total brain volume when outliers are removed, $p = 0.069$, $p > 0.05$.

```
plot_brain_vols2 %>%
  pairwise_t_test(
    Brain_Volume_mm3 ~ Genotype,
    p.adjust.method = "bonferroni"
  )
```

Post-hoc tests for non-significant two-way interaction

```
## # A tibble: 1 x 9
##   .y.            group1 group2    n1    n2       p p.signif   p.adj p.adj.signif
## * <chr>          <chr>  <chr>  <int> <int>   <dbl> <chr>      <dbl> <chr>
## 1 Brain_Volume_~ KO     WT        19    32 0.00016 ***      0.00016 ***
```

```
plot_brain_vols2 %>%
  pairwise_t_test(
    Brain_Volume_mm3 ~ Time_months,
    p.adjust.method = "bonferroni"
  )
```

```
## # A tibble: 3 x 9
##   .y.              group1 group2    n1    n2     p p.signif p.adj p.adj.signif
## * <chr>            <chr>  <chr>  <int> <int> <dbl> <chr>    <dbl> <chr>
## 1 Brain_Volume_mm3 3      6         16    18 0.477 ns           1 ns
## 2 Brain_Volume_mm3 3      10        16    17 0.717 ns           1 ns
## 3 Brain_Volume_mm3 6      10        18    17 0.725 ns           1 ns
```

```
model1 <- glm(formula= Brain_Volume_mm3 ~ Time_months + Genotype, data=plot_brain_vols2, family=gaussian
summary (model1)
```

GLM Without Outliers

```
##
## Call:
## glm(formula = Brain_Volume_mm3 ~ Time_months + Genotype, family = gaussian(),
##     data = plot_brain_vols2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -69.810  -18.600    5.614   21.383   56.431
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 540.7864    10.8749  49.728  < 2e-16 ***
## Time_months   0.5971     1.3883   0.430 0.669057
## GenotypeWT  -33.0108     8.1231  -4.064 0.000178 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 786.307)
##
```

```
##     Null deviance: 50822  on 50  degrees of freedom
## Residual deviance: 37743  on 48  degrees of freedom
## AIC: 489.67
##
## Number of Fisher Scoring iterations: 2
```