

# Project

*Darya Nemirich*

```
library(dplyr)

##
## 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(tidyr)
library(DataExplorer)
library(ggpubr)

## Loading required package: ggplot2
## Loading required package: magrittr
##
## Attaching package: 'magrittr'
## The following object is masked from 'package:tidyr':
##
##   extract

neurons <- read.csv2("D:/Bioinformatics and System Biology/2nd term/R/R_classwork/Project/data.csv")
neurons2 <- read.csv2("D:/Bioinformatics and System Biology/2nd term/R/R_classwork/Project/data2.csv")
View(neurons)
View(neurons2)

summary(neurons)

##      Treatment  Neurite_length  Soma_diameter
## control   :130   Min.    : 5.522   Min.    :11.30
## GNF       :107   1st Qu.: 26.580   1st Qu.:20.58
## GNF_nutlin:104   Median : 41.141   Median :24.61
## nutlin    : 97   Mean    : 47.885   Mean    :25.73
##           3rd Qu.: 59.943   3rd Qu.:29.74
##           Max.    :252.927   Max.    :55.12
##           NA's    :169

str(neurons)

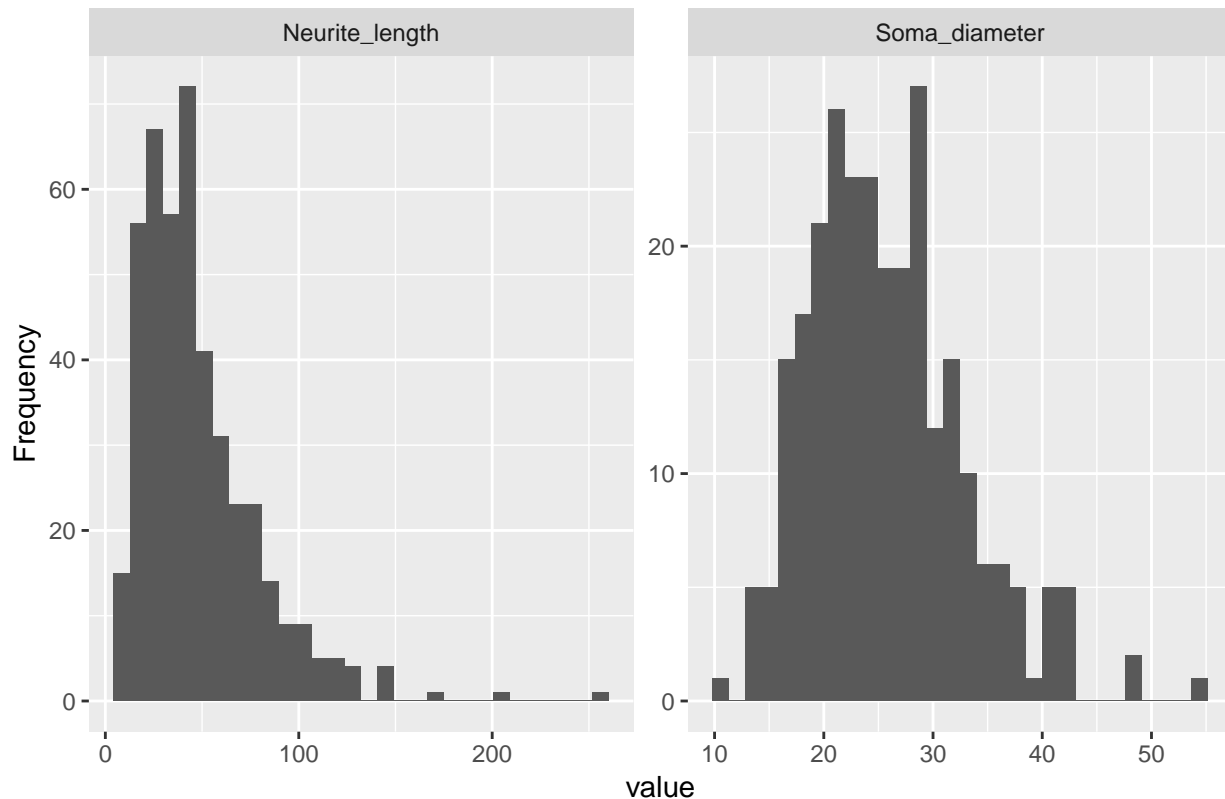
## 'data.frame':   438 obs. of  3 variables:
## $ Treatment      : Factor w/ 4 levels "control","GNF",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Neurite_length: num  21.2 128 17.1 24.4 45.7 ...
## $ Soma_diameter : num  21.6 34.4 17.9 15 20.8 ...

levels(neurons$Treatment)

## [1] "control" "GNF" "GNF_nutlin" "nutlin"
```

```
neurons$Treatment <- ordered(neurons$Treatment,
                             levels = c('control', 'nutlin', 'GNF', 'GNF_nutlin'))

plot_histogram(neurons)
```



```
neurites_summary <- neurons %>%
  group_by(Treatment) %>%
  summarise(
    count = n(),
    neurite_mean = mean(Neurite_length),
    neurite_sd = sd(Neurite_length),
    neurite_median = median(Neurite_length),
    IQR = IQR(Neurite_length)
  )
```

```
neurites_summary
```

```
## # A tibble: 4 x 6
##   Treatment count neurite_mean neurite_sd neurite_median IQR
##   <ord>      <int>      <dbl>      <dbl>      <dbl> <dbl>
## 1 control    130        43.5        31.9        35.4  36.7
## 2 nutlin     97        56.7        34.2        46.8  44.3
## 3 GNF       107        40.8        24.3        34.0  30.4
## 4 GNF_nutlin 104        52.3        29.1        44.4  27.1
```

```
soma_summary <- neurons %>%
  group_by(Treatment) %>%
```

```

summarise(
  count = n(),
  soma_mean = mean(Soma_diameter, na.rm = TRUE),
  soma_sd = sd(Soma_diameter, na.rm = TRUE),
  soma_median = median(Soma_diameter, na.rm = TRUE),
  IQR = IQR(Soma_diameter, na.rm = TRUE)
)

soma_summary

## # A tibble: 4 x 6
##   Treatment count soma_mean soma_sd soma_median IQR
##   <ord>      <int>    <dbl>  <dbl>    <dbl> <dbl>
## 1 control      130     23.4   6.65     22.0  7.88
## 2 nutlin       97     26.0   6.81     25.6  8.2
## 3 GNF         107     26.5   7.60     25.9  9.52
## 4 GNF_nutlin  104     27.5   6.68     27.9  9.41

control <- subset(neurons, Treatment == 'control')
shapiro.test(control$Neurite_length)

##
## Shapiro-Wilk normality test
##
## data: control$Neurite_length
## W = 0.84205, p-value = 1.743e-10

nutlin <- subset(neurons, Treatment == 'nutlin')
shapiro.test(nutlin$Neurite_length)

##
## Shapiro-Wilk normality test
##
## data: nutlin$Neurite_length
## W = 0.82541, p-value = 2.414e-09

GNF <- subset(neurons, Treatment == 'GNF')
shapiro.test(GNF$Neurite_length)

##
## Shapiro-Wilk normality test
##
## data: GNF$Neurite_length
## W = 0.8722, p-value = 3.884e-08

GNF_nutlin <- subset(neurons, Treatment == 'GNF_nutlin')
shapiro.test(GNF_nutlin$Neurite_length)

##
## Shapiro-Wilk normality test
##
## data: GNF_nutlin$Neurite_length
## W = 0.87653, p-value = 8.27e-08

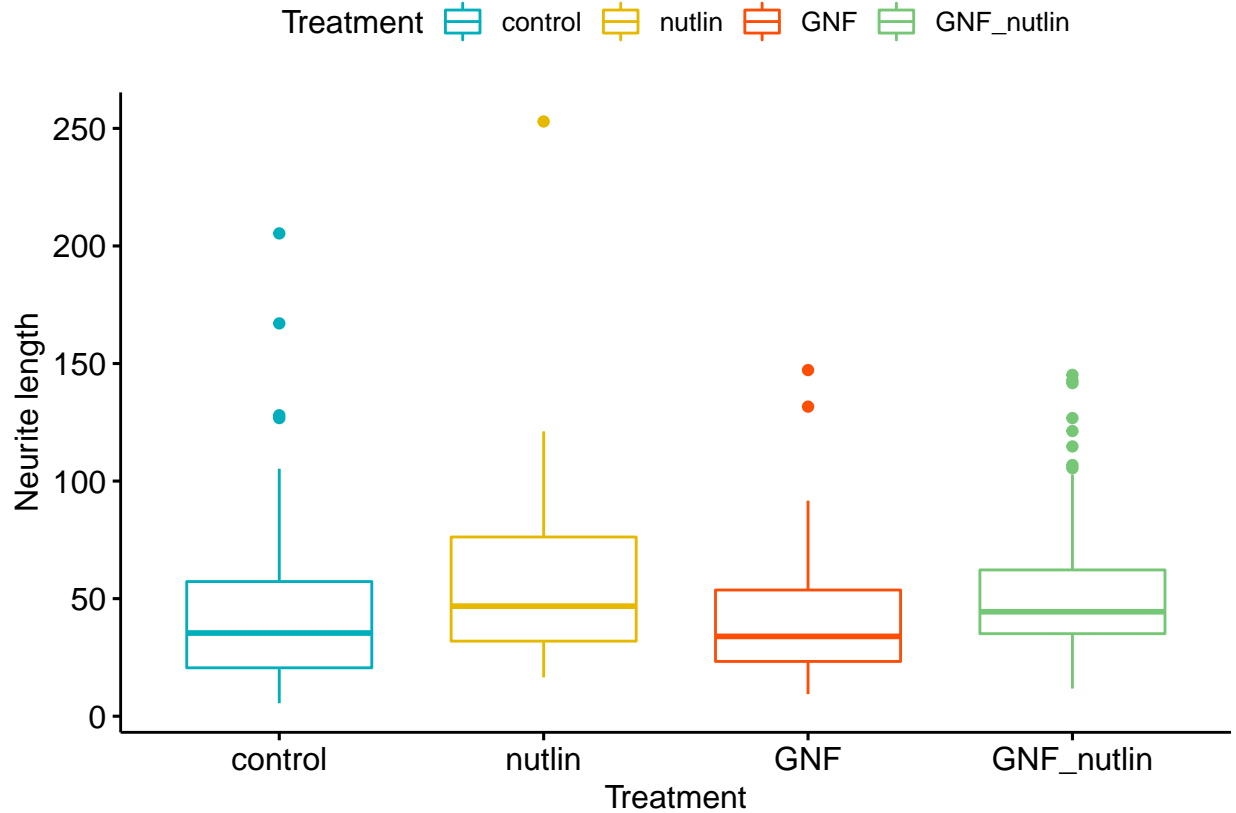
ggboxplot(neurons,
  x = "Treatment",
  y = "Neurite_length",

```

```

color = "Treatment",
palette = c("#00AFBB", "#E7B800", "#FC4E07", "#75c675"),
order = c("control", "nutlin", "GNF", "GNF_nutlin"),
ylab = "Neurite length", xlab = "Treatment")

```

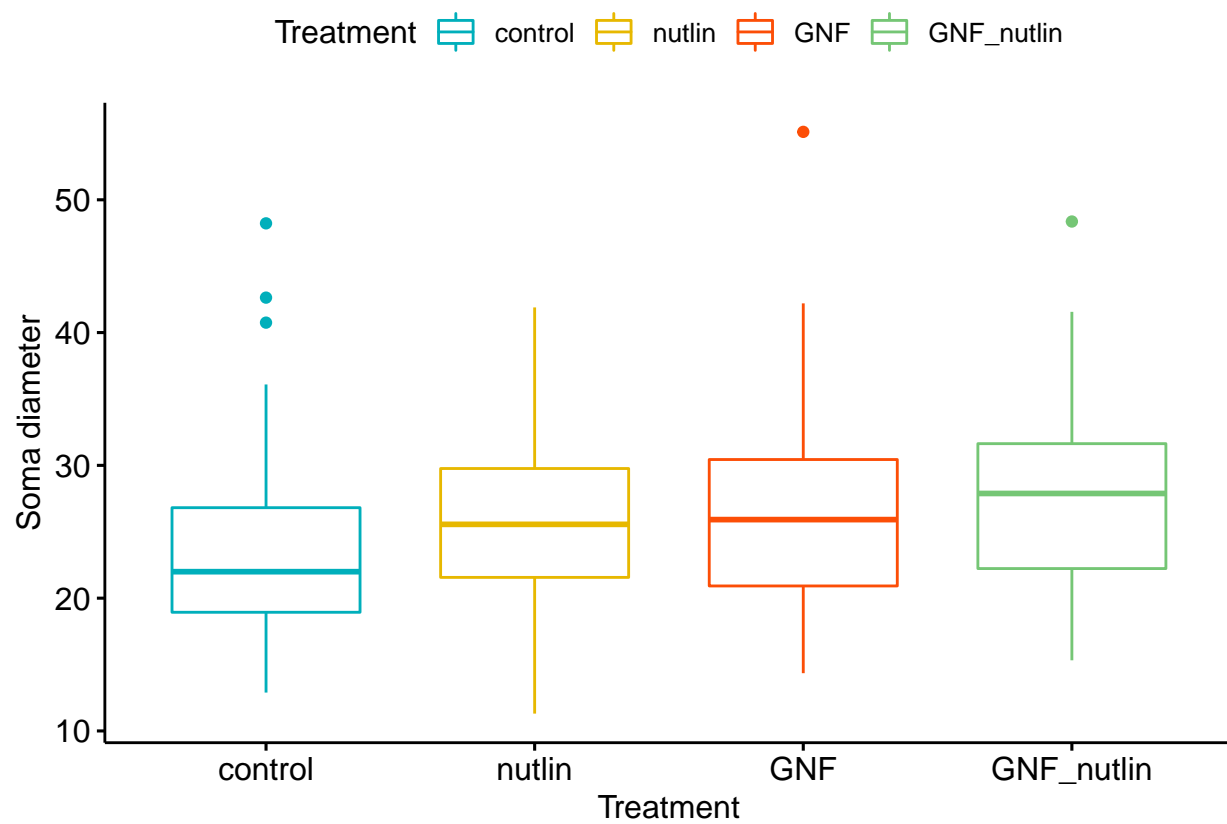


```

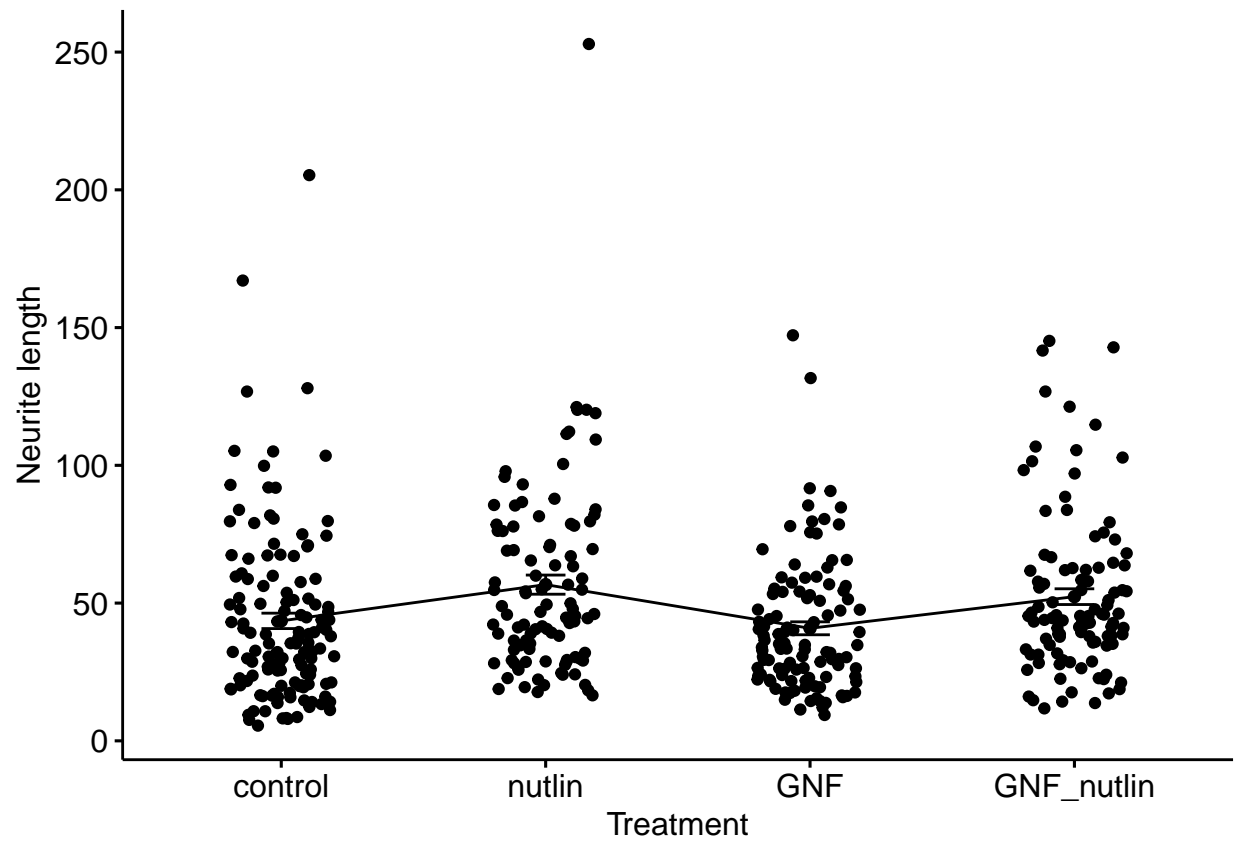
ggboxplot(neurons,
  x = "Treatment",
  y = "Soma_diameter",
  color = "Treatment",
  palette = c("#00AFBB", "#E7B800", "#FC4E07", "#75c675"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Soma diameter", xlab = "Treatment")

```

## Warning: Removed 169 rows containing non-finite values (stat\_boxplot).



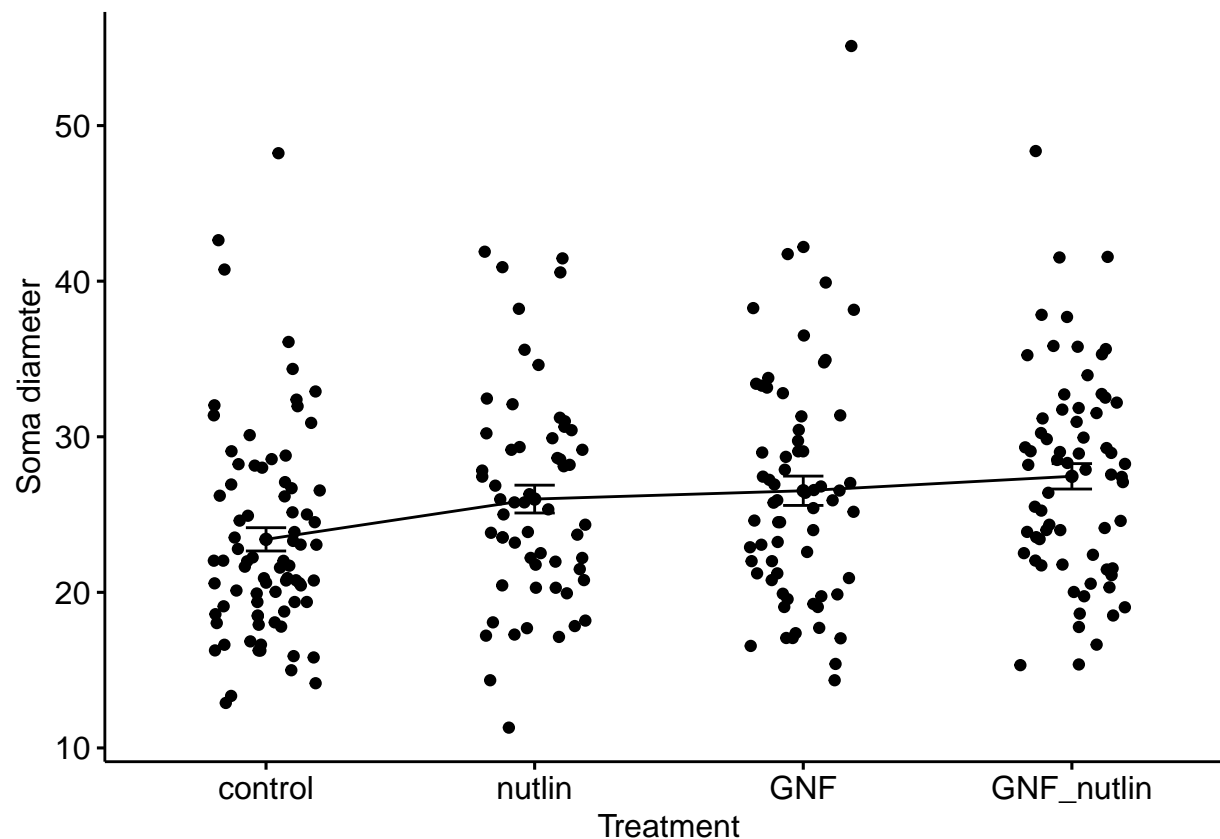
```
ggline(neurons, x = "Treatment", y = "Neurite_length",
  add = c("mean_se", "jitter"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Neurite length", xlab = "Treatment")
```



```
ggline(neurons, x = "Treatment", y = "Soma_diameter",
  add = c("mean_se", "jitter"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Soma diameter", xlab = "Treatment")
```

```
## Warning: Removed 169 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 169 rows containing missing values (geom_point).
```



```
kruskal.test(Neurite_length ~ Treatment, data = neurons)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: Neurite_length by Treatment
## Kruskal-Wallis chi-squared = 25.351, df = 3, p-value = 1.304e-05
```

```
kruskal.test(Soma_diameter ~ Treatment, data = neurons)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: Soma_diameter by Treatment
## Kruskal-Wallis chi-squared = 16.982, df = 3, p-value = 0.0007127
```

```
pairwise.wilcox.test(neurons$Neurite_length, neurons$Treatment, p.adjust.method = "BH")
```

```
##
## Pairwise comparisons using Wilcoxon rank sum test
##
## data: neurons$Neurite_length and neurons$Treatment
##
##          control nutlin  GNF
## nutlin    0.00062 -      -
## GNF       0.98861 0.00041 -
## GNF_nutlin 0.00294 0.49950 0.00164
##
```

```
## P value adjustment method: BH
pairwise.wilcox.test(neurons$Soma_diameter, neurons$Treatment, p.adjust.method = "BH")
```

```
##
## Pairwise comparisons using Wilcoxon rank sum test
##
## data: neurons$Soma_diameter and neurons$Treatment
##
##          control nutlin  GNF
## nutlin    0.02177 -      -
## GNF        0.02094 0.90723 -
## GNF_nutlin 0.00053 0.28684 0.33683
##
## P value adjustment method: BH
```

The second dataset

```
summary(neurons2)
```

```
##      Treatment  Neurite_length  Soma_diameter
## control   :219   Min.   : 8.195   Min.   :10.24
## GNF       :131   1st Qu.: 25.415   1st Qu.:21.44
## GNF_nutlin:122   Median : 38.674   Median :25.91
## nutlin    :169   Mean    : 48.450   Mean    :26.38
##           3rd Qu.: 62.346   3rd Qu.:30.71
##           Max.    :208.204   Max.    :67.50
##           NA's    :248
```

```
str(neurons2)
```

```
## 'data.frame':   641 obs. of  3 variables:
## $ Treatment      : Factor w/ 4 levels "control","GNF",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Neurite_length: num  81.4 67.9 24.8 72.2 14.2 ...
## $ Soma_diameter : num  35.1 32.2 40.7 19.2 21.1 ...
```

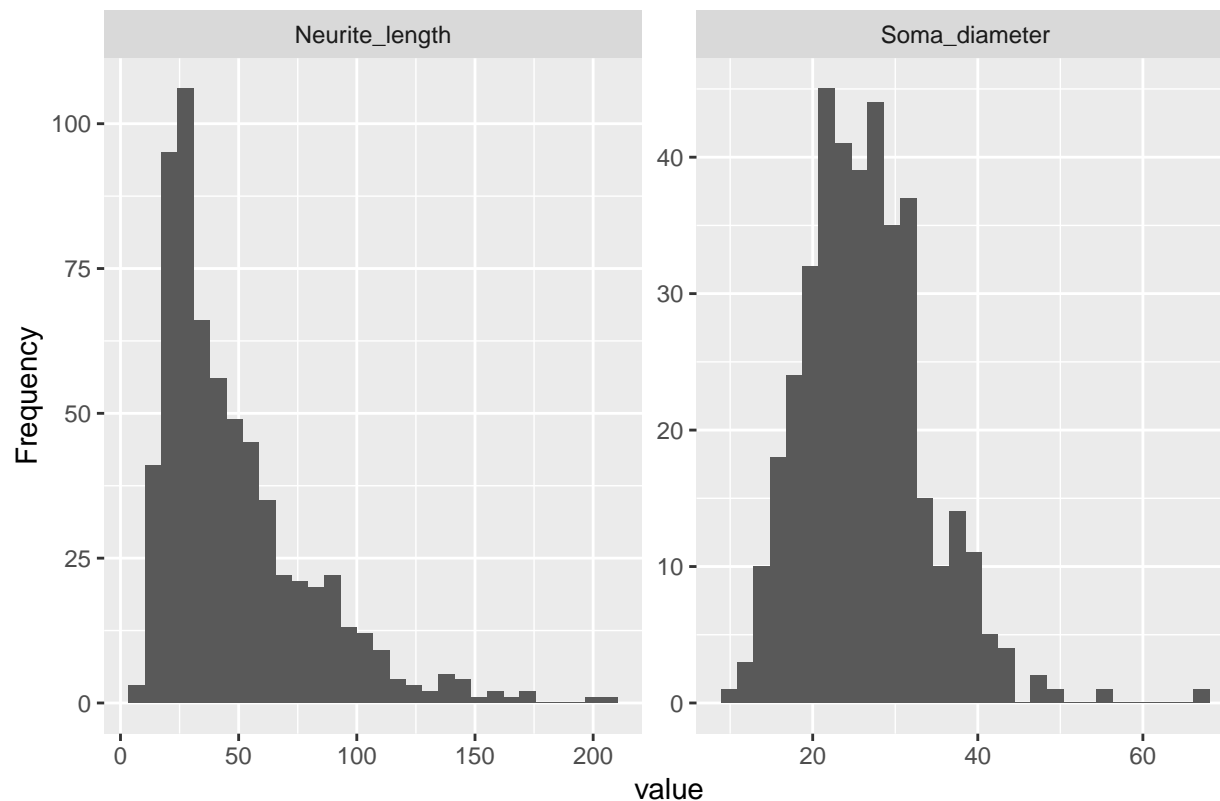
```
levels(neurons2$Treatment)
```

```
## [1] "control"      "GNF"          "GNF_nutlin"  "nutlin"
```

```
neurons2$Treatment <- ordered(neurons2$Treatment,
                              levels = c('control', 'nutlin', 'GNF', 'GNF_nutlin'))
```

```
plot_histogram(neurons2)
```





```
neurites_summary_2 <- neurons2 %>%
  group_by(Treatment) %>%
  summarise(
    count = n(),
    neurite_mean = mean(Neurite_length),
    neurite_sd = sd(Neurite_length),
    neurite_median = median(Neurite_length),
    IQR = IQR(Neurite_length)
  )
```

```
neurites_summary_2
```

```
## # A tibble: 4 x 6
##   Treatment    count neurite_mean neurite_sd neurite_median    IQR
##   <ord>      <int>      <dbl>     <dbl>      <dbl>  <dbl>
## 1 control     219        45.7      26.8        38.3   36.7
## 2 nutlin      169        59.6      34.2        52.1   50.7
## 3 GNF         131        36.2      24.2        28.8   24.0
## 4 GNF_nutlin  122        51.1      36.9        39.3   35.9
```

```
soma_summary_2 <- neurons2 %>%
  group_by(Treatment) %>%
  summarise(
    count = n(),
    soma_mean = mean(Soma_diameter, na.rm = TRUE),
    soma_sd = sd(Soma_diameter, na.rm = TRUE),
    soma_median = median(Soma_diameter, na.rm = TRUE),
```

```

    IQR = IQR(Soma_diameter, na.rm = TRUE)
  )

soma_summary_2

## # A tibble: 4 x 6
##   Treatment count soma_mean soma_sd soma_median IQR
##   <ord>      <int>    <dbl>  <dbl>    <dbl> <dbl>
## 1 control    219     24.2   6.90     23.4  8.64
## 2 nutlin     169     28.6   7.33     28.2  7.88
## 3 GNF        131     25.1   6.37     24.7  9.74
## 4 GNF_nutlin 122     29.3   8.05     27.7  8.28

control2 <- subset(neurons2, Treatment == 'control')
shapiro.test(control2$Neurite_length)

##
##  Shapiro-Wilk normality test
##
## data:  control2$Neurite_length
## W = 0.90421, p-value = 1.224e-10

nutlin2 <- subset(neurons2, Treatment == 'nutlin')
shapiro.test(nutlin2$Neurite_length)

##
##  Shapiro-Wilk normality test
##
## data:  nutlin2$Neurite_length
## W = 0.9178, p-value = 3.627e-08

GNF2 <- subset(neurons2, Treatment == 'GNF')
shapiro.test(GNF2$Neurite_length)

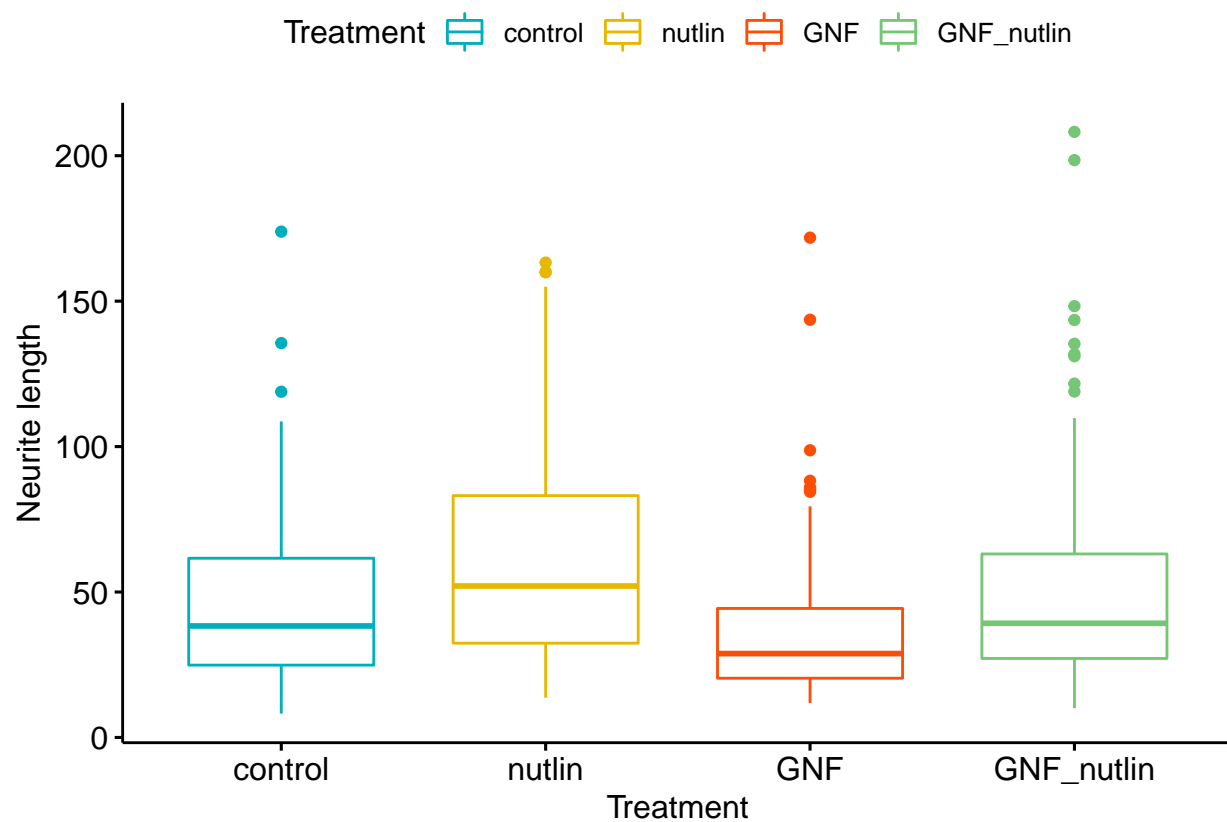
##
##  Shapiro-Wilk normality test
##
## data:  GNF2$Neurite_length
## W = 0.76685, p-value = 3.772e-13

GNF_nutlin2 <- subset(neurons2, Treatment == 'GNF_nutlin')
shapiro.test(GNF_nutlin2$Neurite_length)

##
##  Shapiro-Wilk normality test
##
## data:  GNF_nutlin2$Neurite_length
## W = 0.80974, p-value = 2.862e-11

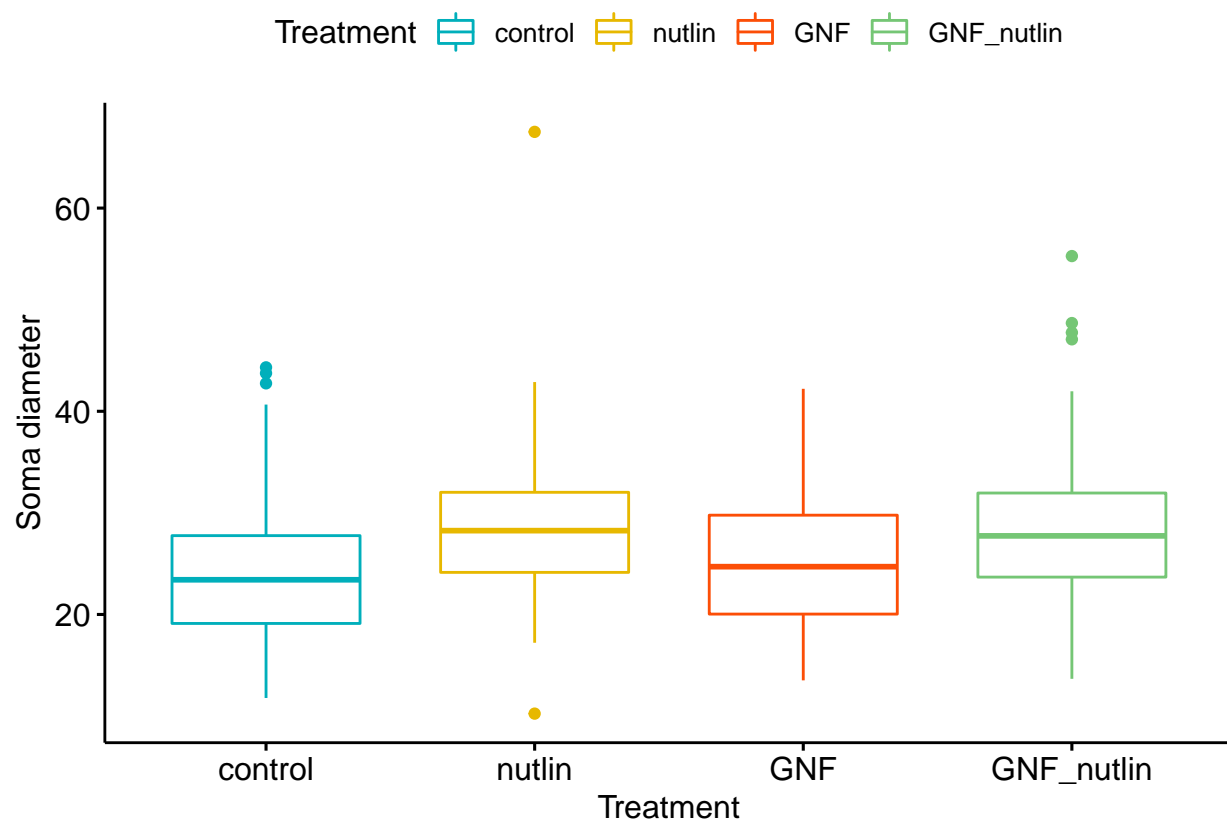
ggboxplot(neurons2,
  x = "Treatment",
  y = "Neurite_length",
  color = "Treatment",
  palette = c("#00AFBB", "#E7B800", "#FC4E07", "#75c675"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Neurite length", xlab = "Treatment")

```

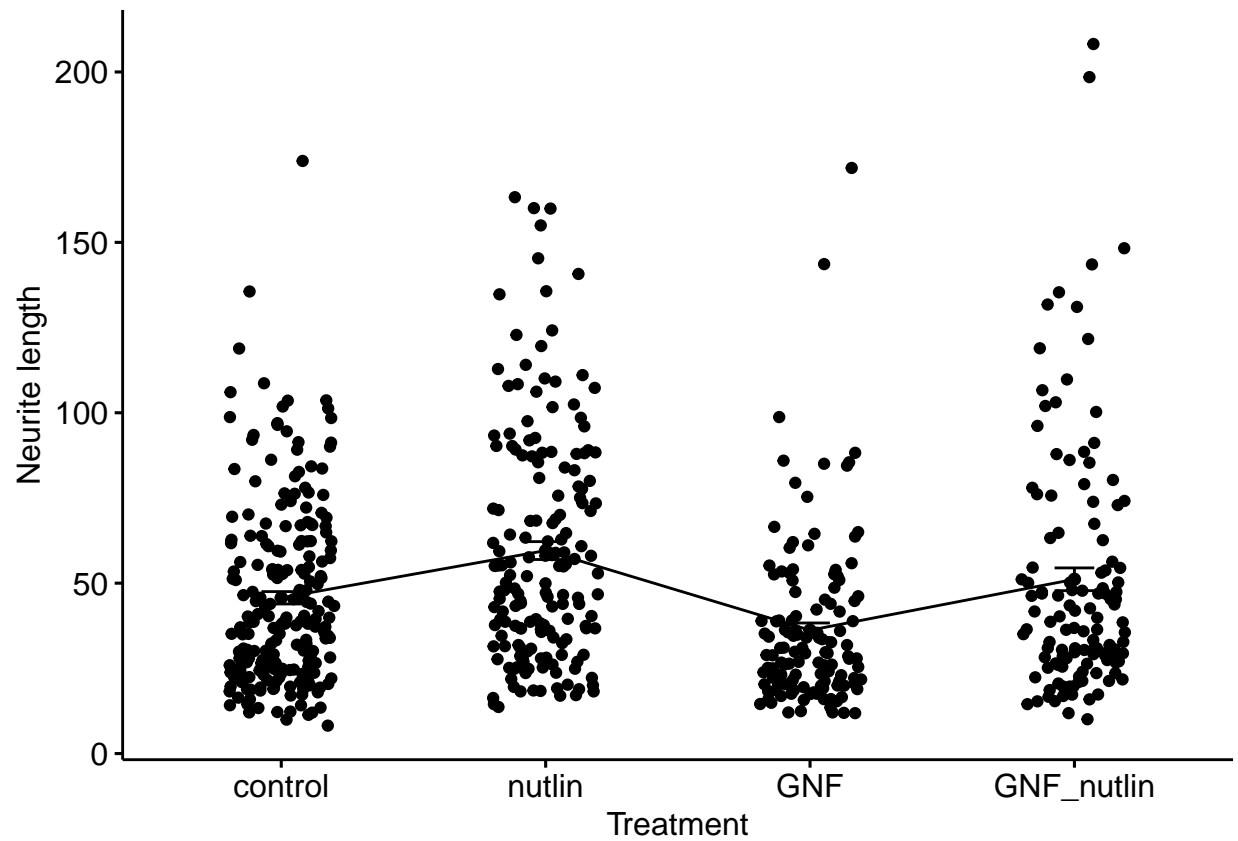


```
ggboxplot(neurons2,
  x = "Treatment",
  y = "Soma_diameter",
  color = "Treatment",
  palette = c("#00AFBB", "#E7B800", "#FC4E07", "#75c675"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Soma diameter", xlab = "Treatment")
```

## Warning: Removed 248 rows containing non-finite values (stat\_boxplot).



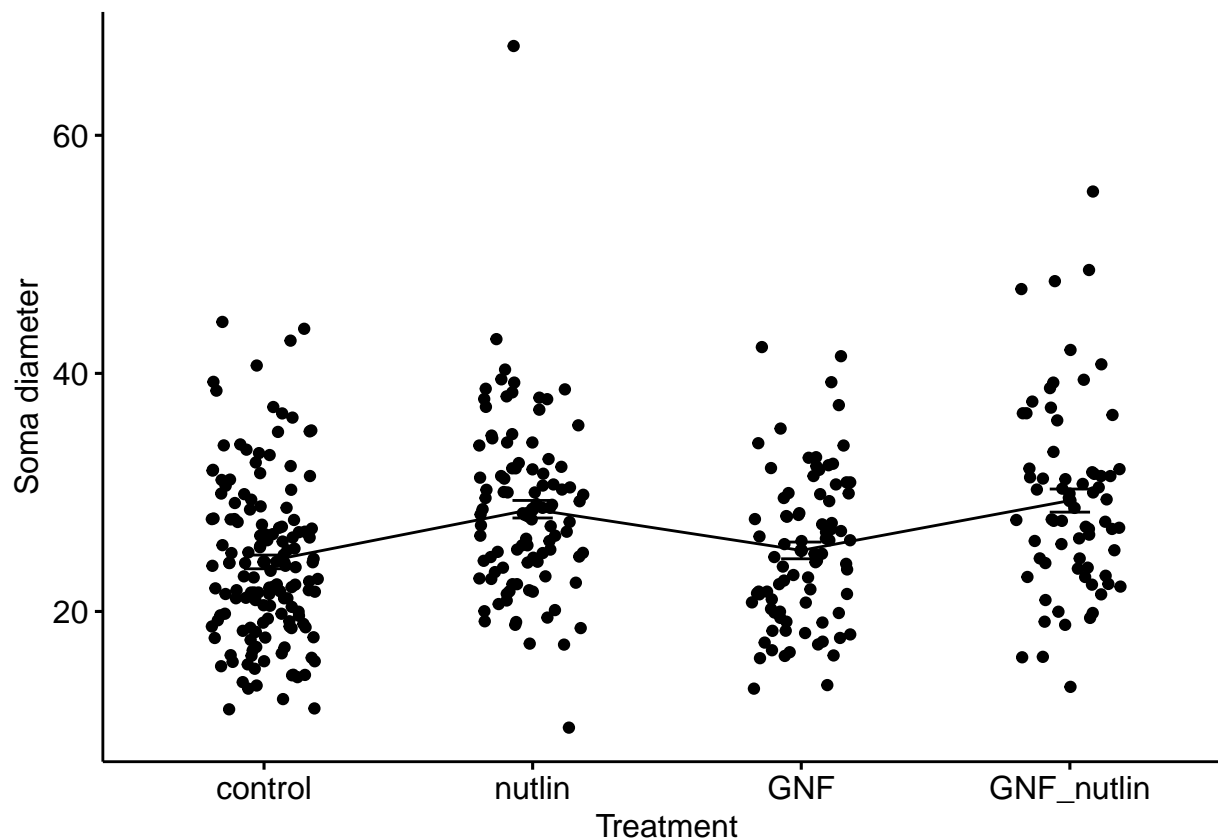
```
ggline(neurons2, x = "Treatment", y = "Neurite_length",
  add = c("mean_se", "jitter"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Neurite length", xlab = "Treatment")
```



```
ggline(neurons2, x = "Treatment", y = "Soma_diameter",
  add = c("mean_se", "jitter"),
  order = c("control", "nutlin", "GNF", "GNF_nutlin"),
  ylab = "Soma diameter", xlab = "Treatment")
```

```
## Warning: Removed 248 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 248 rows containing missing values (geom_point).
```



```
kruskal.test(Neurite_length ~ Treatment, data = neurons2)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: Neurite_length by Treatment
## Kruskal-Wallis chi-squared = 51.152, df = 3, p-value = 4.54e-11
```

```
kruskal.test(Soma_diameter ~ Treatment, data = neurons2)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: Soma_diameter by Treatment
## Kruskal-Wallis chi-squared = 36.231, df = 3, p-value = 6.693e-08
```

```
pairwise.wilcox.test(neurons2$Neurite_length, neurons2$Treatment, p.adjust.method = "BH")
```

```
##
## Pairwise comparisons using Wilcoxon rank sum test
##
## data: neurons2$Neurite_length and neurons2$Treatment
##
##      control nutlin  GNF
## nutlin  0.00011 -      -
## GNF      0.00019 1.6e-11 -
## GNF_nutlin 0.48955 0.00534 0.00019
##
```

```
## P value adjustment method: BH
pairwise.wilcox.test(neurons2$Soma_diameter, neurons2$Treatment, p.adjust.method = "BH")

##
## Pairwise comparisons using Wilcoxon rank sum test
##
## data: neurons2$Soma_diameter and neurons2$Treatment
##
##          control nutlin GNF
## nutlin    4.8e-06 -      -
## GNF       0.2673  0.0016 -
## GNF_nutlin 1.6e-05 0.7084 0.0022
##
## P value adjustment method: BH
```