Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019)

Examples of Chapter 8: Analysis of Longitudinal Data I: Graphical Displays and Summary Measure Approach

Kimmo Vehkalahti and Brian S. Everitt

1 October 2018

Contents

Examples

Table 8.1: BPRS Measurements from 40 Subjects

```
library(tidyr); library(dplyr); library(ggplot2)
```

```
##
## 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
```

```
BPRS <- read.table("data/BPRS.txt", sep = ' ', header = TRUE)
BPRS <- within(BPRS, {
    treatment <- factor(treatment)
      subject <- factor(subject)
})

glimpse(BPRS)
```

```
## Observations: 40
## Variables: 11
## $ treatment <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ week0     <int> 42, 58, 54, 55, 72, 48, 71, 30, 41, 57, 30, 55, 36, ...
## $ week1     <int> 36, 68, 55, 77, 75, 43, 61, 36, 43, 51, 34, 52, 32, ...
## $ week2     <int> 36, 61, 41, 49, 72, 41, 47, 38, 39, 51, 34, 49, 36, ...
## $ week3     <int> 43, 55, 38, 54, 65, 38, 30, 38, 35, 55, 41, 54, 31, ...
## $ week4     <int> 41, 43, 43, 56, 50, 36, 27, 31, 28, 53, 36, 48, 25, ...
## $ week5     <int> 40, 34, 28, 50, 39, 29, 40, 26, 22, 43, 36, 43, 25, ...
## $ week6     <int> 38, 28, 29, 47, 32, 33, 30, 26, 20, 43, 38, 37, 21, ...
## $ week7     <int> 47, 28, 25, 42, 38, 27, 31, 25, 23, 39, 36, 36, 19, ...
## $ week8     <int> 51, 28, 24, 46, 32, 25, 31, 24, 21, 32, 36, 31, 22, ...
```

```
head(BPRS)
```

```
##   treatment subject week0 week1 week2 week3 week4 week5 week6 week7 week8
## 1         1       1    42    36    36    43    41    40    38    47    51
## 2         1       2    58    68    61    55    43    34    28    28    28
## 3         1       3    54    55    41    38    43    28    29    25    24
## 4         1       4    55    77    49    54    56    50    47    42    46
## 5         1       5    72    75    72    65    50    39    32    38    32
## 6         1       6    48    43    41    38    36    29    33    27    25
```

```
tail(BPRS)
```

```
##    treatment subject week0 week1 week2 week3 week4 week5 week6 week7 week8
## 35         2      15    40    36    55    55    42    30    26    30    37
## 36         2      16    54    45    35    27    25    22    22    22    22
## 37         2      17    33    41    30    32    46    43    43    43    43
## 38         2      18    28    30    29    33    30    26    36    33    30
```

```
## 39         2      19     52     43     26     27     24     32     21     21     21
## 40         2      20     47     36     32     29     25     23     23     23     23
```

```
# Convert data to long form:

BPRSL <- gather(BPRS, key = weeks, value = bprs, -treatment, -subject) %>%
  mutate(week = as.integer(substr(weeks, 5, 5)))

glimpse(BPRSL)
```

```
## Observations: 360
## Variables: 5
## $ treatment <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ weeks     <chr> "week0", "week0", "week0", "week0", "week0", "week0"...
## $ bprs      <int> 42, 58, 54, 55, 72, 48, 71, 30, 41, 57, 30, 55, 36, ...
## $ week      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
```

Figure 8.1

```
p1 <- ggplot(BPRSL, aes(x = week, y = bprs, linetype = subject))
p2 <- p1 + geom_line() + scale_linetype_manual(values = rep(1:10, times=4))
p3 <- p2 + facet_grid(. ~ treatment, labeller = label_both)
p4 <- p3 + theme_bw() + theme(legend.position = "none")
p5 <- p4 + theme(panel.grid.minor.y = element_blank())
p6 <- p5 + scale_y_continuous(limits = c(min(BPRSL$bprs), max(BPRSL$bprs)))
p6
```

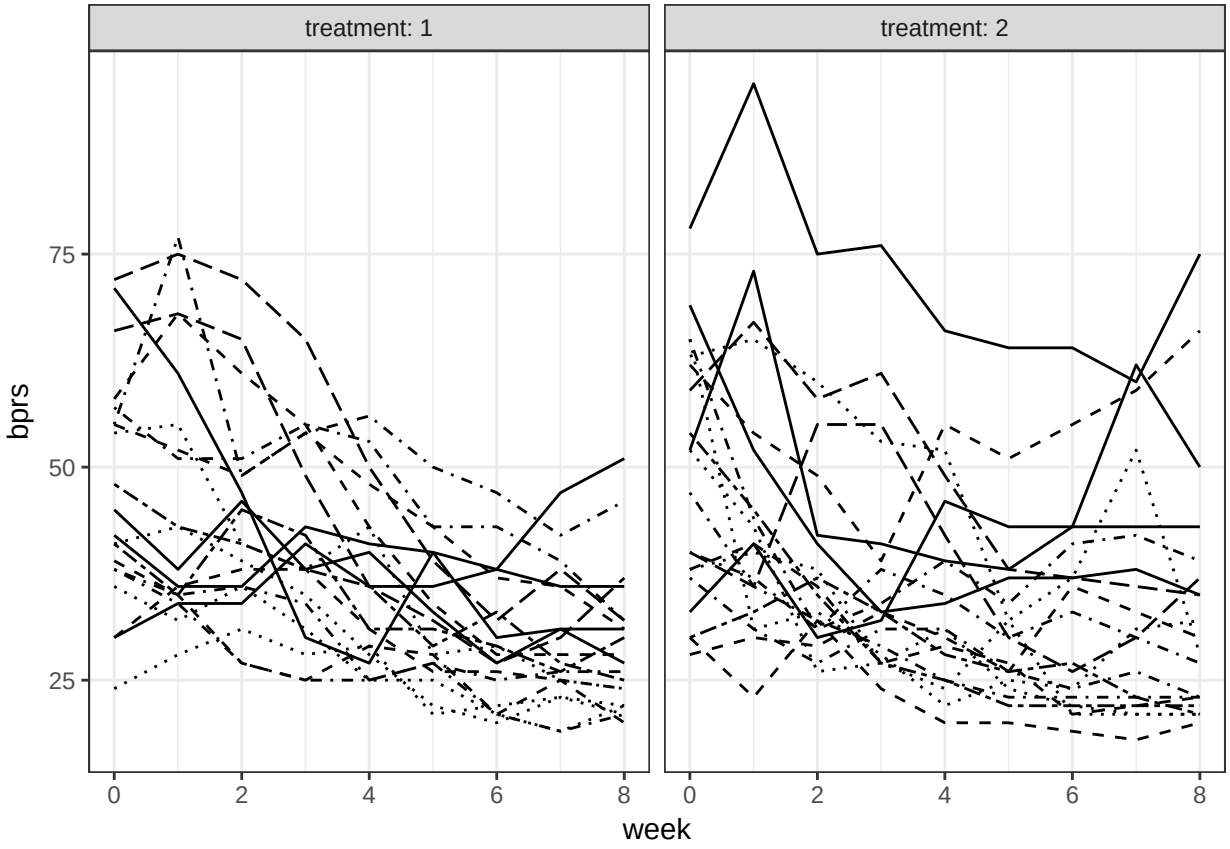

Figure 8.2

```
# Standardise the scores:
BPRSL <- BPRSL %>%
  group_by(week) %>%
  mutate( stdbprs = (bprs - mean(bprs))/sd(bprs) ) %>%
  ungroup()
glimpse(BPRSL)
```

```
## Observations: 360
## Variables: 6
## $ treatment <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ weeks     <chr> "week0", "week0", "week0", "week0", "week0", "week0"...
## $ bprs      <int> 42, 58, 54, 55, 72, 48, 71, 30, 41, 57, 30, 55, 36, ...
## $ week      <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ stdbprs   <dbl> -0.4245908, 0.7076513, 0.4245908, 0.4953559, 1.69836...
```

```
p1 <- ggplot(BPRSL, aes(x = week, y = stdbprs, linetype = subject))
p2 <- p1 + geom_line() + scale_linetype_manual(values = rep(1:10, times=4))
p3 <- p2 + facet_grid(. ~ treatment, labeller = label_both)
p4 <- p3 + theme_bw() + theme(legend.position = "none")
p5 <- p4 + theme(panel.grid.minor.y = element_blank())
p6 <- p5 + scale_y_continuous(name = "standardized bprs")
p6
```

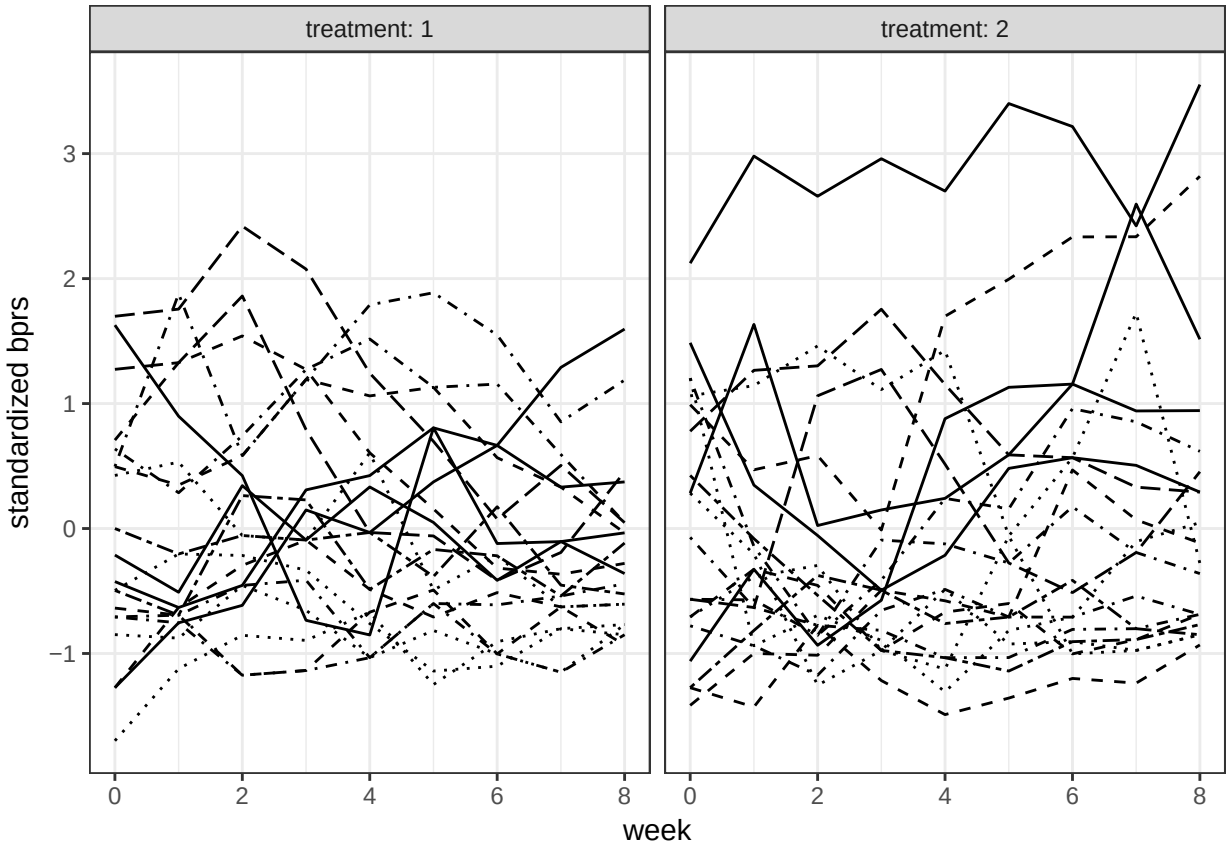

Figure 8.3

```
# Make a summary data:
n <- length(0:8) # weeks, incl. baseline (0)
BPRSS <- BPRSL %>%
  group_by(treatment, week) %>%
  summarise( mean=mean(bprs), se=sd(bprs)/sqrt(n) ) %>%
  ungroup()
glimpse(BPRSS)
```

```
## Observations: 18
## Variables: 4
## $ treatment <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2
## $ week      <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 0, 1, 2, 3, 4, 5, 6, 7, 8
## $ mean      <dbl> 47.00, 46.80, 43.55, 40.90, 36.60, 32.70, 29.70, 29....
## $ se        <dbl> 4.534468, 5.173708, 4.003617, 3.744626, 3.259534, 2....
```

```
p1 <- ggplot(BPRSS, aes(x = week, y = mean, linetype = treatment, shape = treatment))
p2 <- p1 + geom_line() + scale_linetype_manual(values = c(1,2))
p3 <- p2 + geom_point(size=3) + scale_shape_manual(values = c(1,2))
p4 <- p3 + geom_errorbar(aes(ymin=mean-se, ymax=mean+se, linetype="1"), width=0.3)
p5 <- p4 + theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
p6 <- p5 + theme(legend.position = c(0.8,0.8))
p7 <- p6 + scale_y_continuous(name = "mean(bprs) +/- se(bprs)")
p7
```

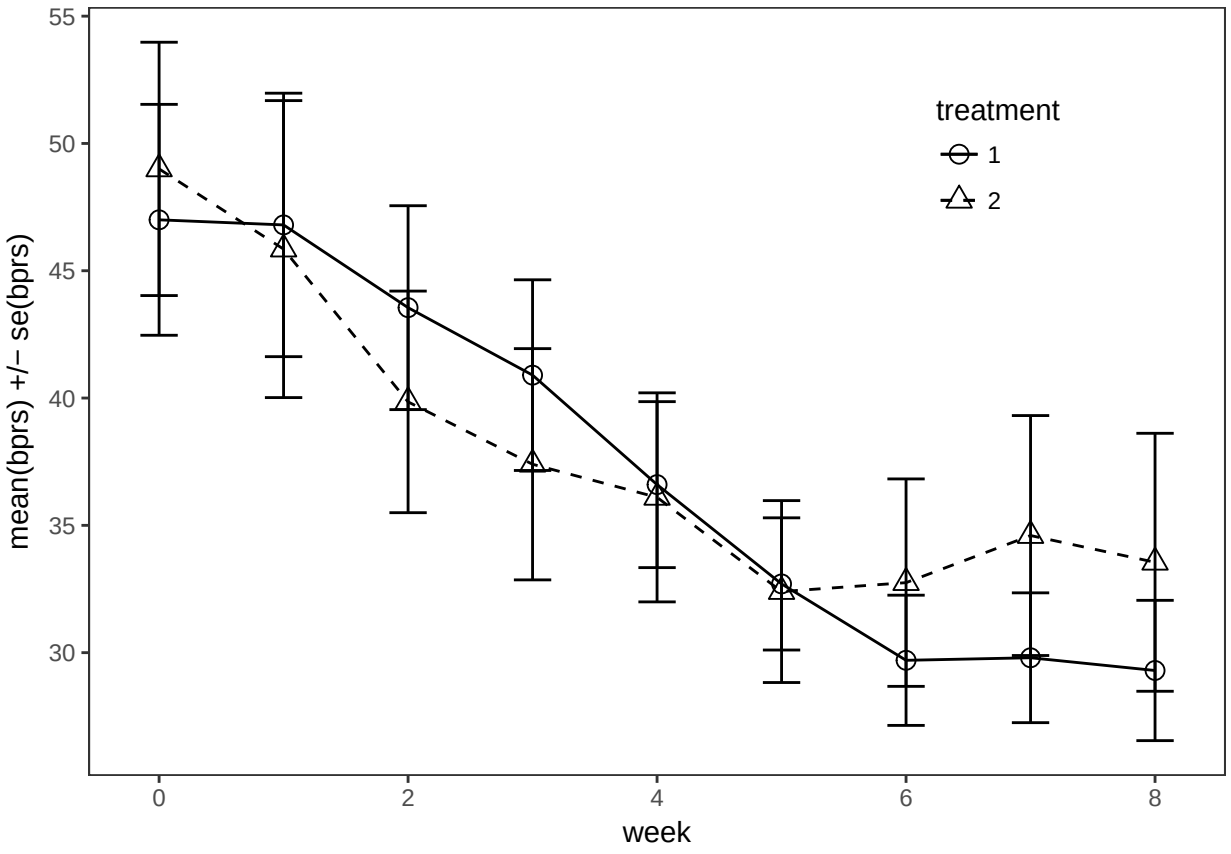

Figure 8.4

```
p1 <- ggplot(BPRSL, aes(x = factor(week), y = bprs, fill = treatment))
p2 <- p1 + geom_boxplot(position = position_dodge(width = 0.9))
p3 <- p2 + theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
p4 <- p3 + theme(legend.position = c(0.8,0.8))
p5 <- p4 + scale_x_discrete(name = "week")
# Black & White version:
#p6 <- p5 + scale_fill_grey(start = 0.5, end = 1)
p5
```

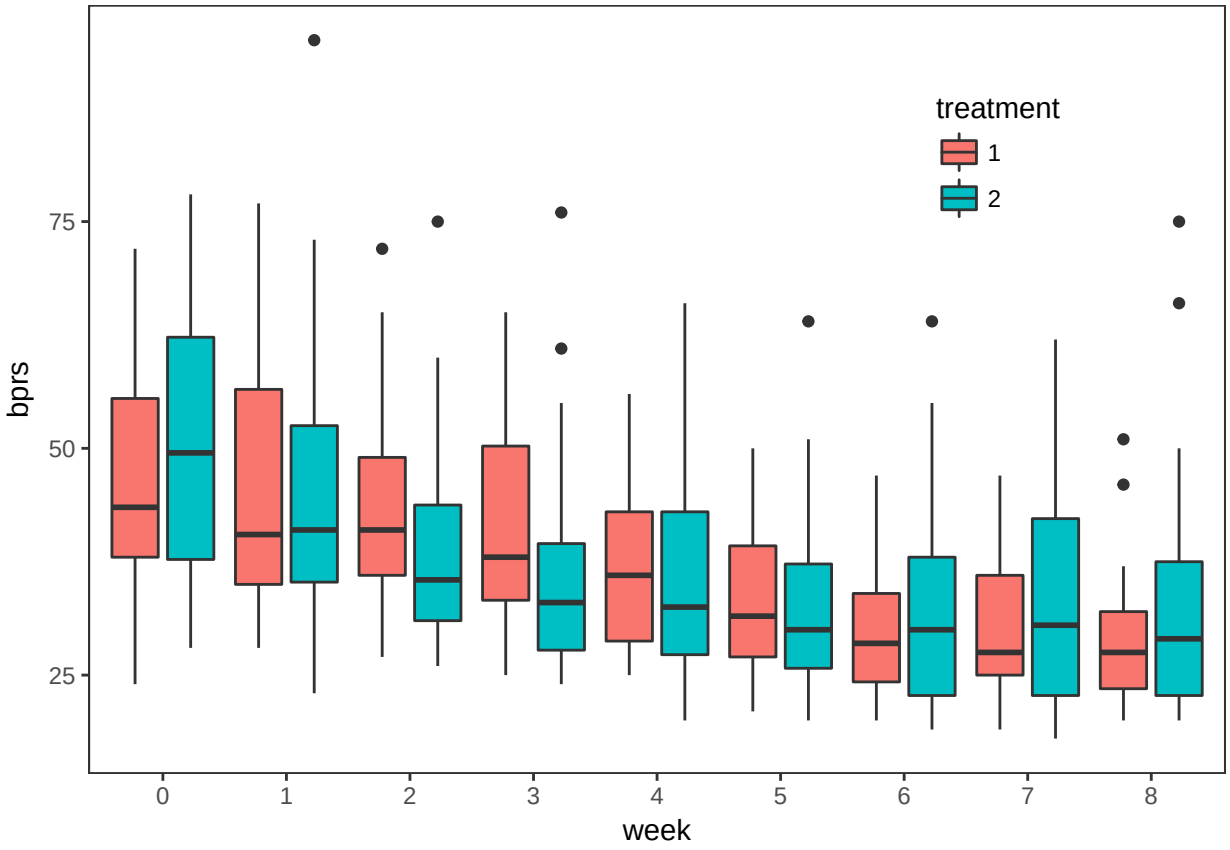

Figure 8.5

```
# Make a summary data of the post treatment weeks (1-8)
BPRSL8S <- BPRSL %>%
  filter(week > 0) %>%
  group_by(treatment, subject) %>%
  summarise( mean=mean(bprs) ) %>%
  ungroup()

glimpse(BPRSL8S)
```

```
## Observations: 40
## Variables: 3
## $ treatment <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ mean      <dbl> 41.500, 43.125, 35.375, 52.625, 50.375, 34.000, 37.1...
```

```
p1 <- ggplot(BPRSL8S, aes(x = treatment, y = mean))
p2 <- p1 + geom_boxplot()
p3 <- p2 + theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
p4 <- p3 + stat_summary(fun.y = "mean", geom = "point", shape=23, size=4, fill = "white")
p5 <- p4 + scale_y_continuous(name = "mean(bprs), weeks 1-8")
p5
```

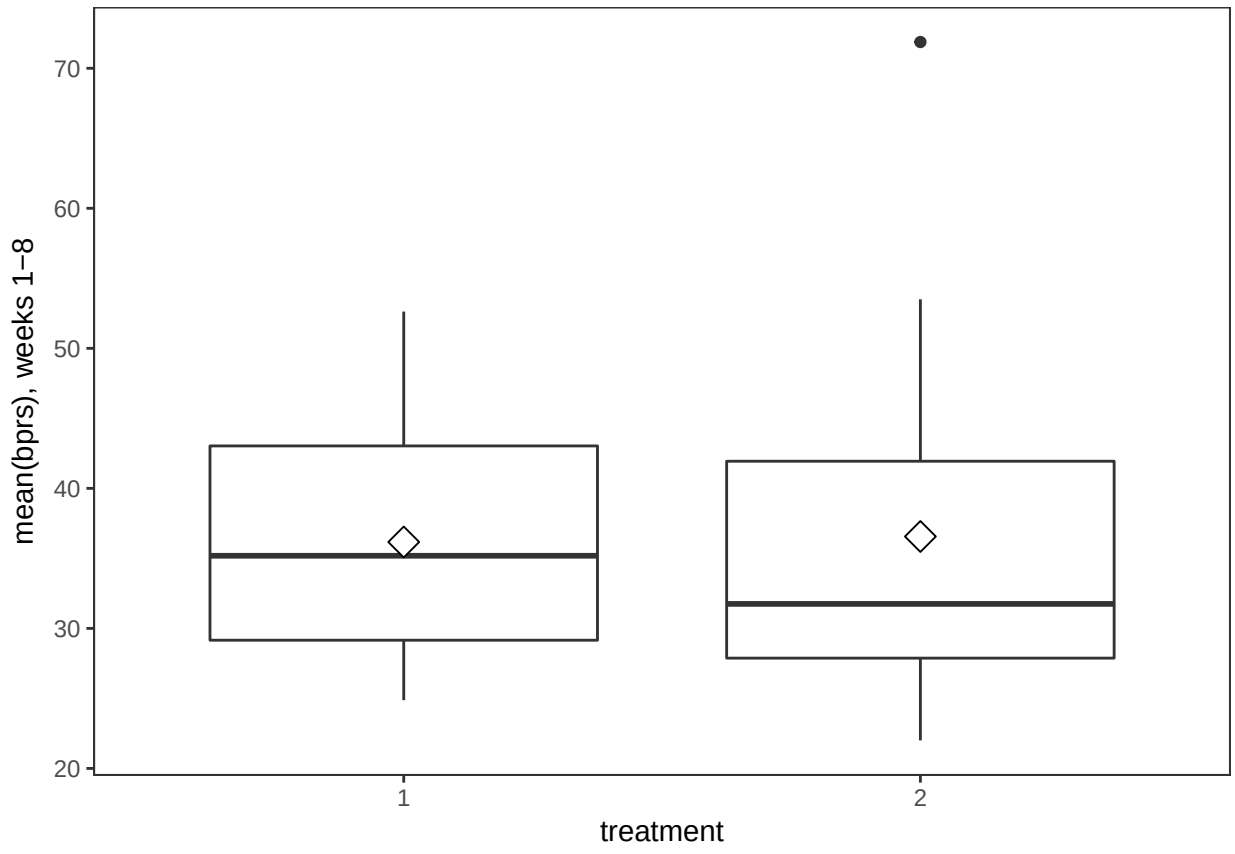

Figure 8.6

```
# Remove the outlier:
BPRSL8S1 <- BPRSL8S %>%
  filter(mean < 60)

glimpse(BPRSL8S1)
```

```
## Observations: 39
## Variables: 3
## $ treatment <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ subject   <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1...
## $ mean      <dbl> 41.500, 43.125, 35.375, 52.625, 50.375, 34.000, 37.1...
```

```
p1 <- ggplot(BPRSL8S1, aes(x = treatment, y = mean))
p2 <- p1 + geom_boxplot()
p3 <- p2 + theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
p4 <- p3 + stat_summary(fun.y = "mean", geom = "point", shape=23, size=4, fill = "white")
p5 <- p4 + scale_y_continuous(name = "mean(bprs), weeks 1-8")
p5
```

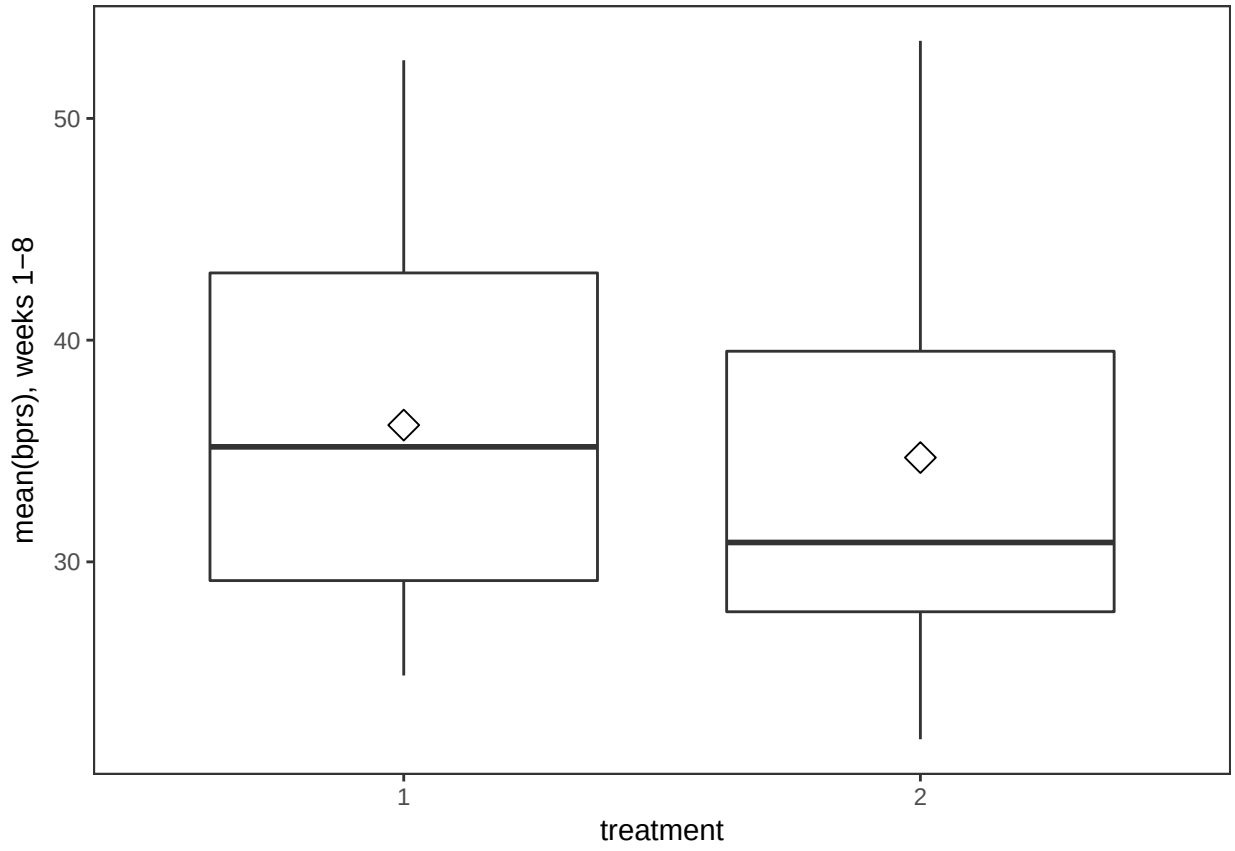

Table 8.3

```r
# Without the outlier, apply Student's t-test, two-sided:

t.test(mean ~ treatment, data = BPRSL8S1, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  mean by treatment
## t = 0.52095, df = 37, p-value = 0.6055
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.232480  7.162085
## sample estimates:
## mean in group 1 mean in group 2
##        36.16875        34.70395
```

Table 8.4

```
# Add the baseline from the original data as a new variable to the summary data:

baseline <- BPRS$week0
BPRSL8S2 <- BPRSL8S %>%
  mutate(baseline)

# Fit the ANCOVA model and see the results:

fit <- lm(mean ~ baseline + treatment, data = BPRSL8S2)

summary(fit)
```

```
##
## Call:
## lm(formula = mean ~ baseline + treatment, data = BPRSL8S2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.1729  -4.5994   0.1088   4.6703  21.0656
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.07897    4.55710   2.870  0.00675 **
## baseline     0.49127    0.08943   5.493 3.05e-06 ***
## treatment2  -0.58879    2.49584  -0.236  0.81480
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.872 on 37 degrees of freedom
## Multiple R-squared:  0.4494, Adjusted R-squared:  0.4196
## F-statistic:  15.1 on 2 and 37 DF,  p-value: 1.605e-05
```

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: mean
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## baseline   1 1868.07 1868.07 30.1437 3.077e-06 ***
## treatment  1    3.45    3.45  0.0557    0.8148
## Residuals 37 2292.97   61.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 8.5: Pain Scores from 83 Women in Labor: 43 Subjects in Group 1 (Medication) and 40 Subjects in Group 2 (Placebo)

```
PAIN <- read.table("data/pain.txt", header = TRUE, sep = "\t")

PAIN <- within(PAIN, {
    group <- factor(group)
       id <- factor(id)
})

glimpse(PAIN)
```

```
## Observations: 83
## Variables: 9
## $ group <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ id    <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1...
## $ m0    <dbl> 0.0, 0.0, 38.0, 6.0, 19.0, 7.0, 44.0, 1.0, 24.5, 1.0, 35...
## $ m30   <dbl> 0.0, 0.0, 5.0, 48.0, 5.0, 0.0, 42.0, 0.0, 35.0, 30.5, 44...
## $ m60   <dbl> 0.0, 0.0, 1.0, 85.0, NA, 0.0, 42.0, 0.0, 13.0, 81.5, 55....
## $ m90   <dbl> 0.0, 0.0, 1.0, 0.0, NA, 0.0, 45.0, 0.0, NA, 67.5, 69.0, ...
## $ m120  <dbl> NA, 2.5, 0.0, 0.0, NA, NA, NA, 0.0, NA, 98.5, 72.5, 0.0,...
## $ m150  <dbl> NA, 2.3, 5.0, NA, NA, NA, NA, 6.0, NA, 97.0, 39.5, 0.0, ...
## $ m180  <dbl> NA, 14.0, NA, NA, NA, NA, NA, 24.0, NA, NA, 26.0, 0.0, 9...
```

Table 8.6

```
# Convert data to long form:
PAINL <- gather(PAIN, key = mins, value = pain, -group, -id) %>%
  mutate(mins = as.integer(substr(mins,2,4)))

glimpse(PAINL)
```

```
## Observations: 581
## Variables: 4
## $ group <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ id    <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1...
## $ mins  <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ pain  <dbl> 0.0, 0.0, 38.0, 6.0, 19.0, 7.0, 44.0, 1.0, 24.5, 1.0, 35...
```

```
# (1) Make a summary data of group means,
# removing subjects with any missing values:

PAINLS1 <- PAINL %>%
  group_by(group, id) %>%
  summarise( mean=mean(pain) ) %>%
  ungroup()

glimpse(PAINLS1)
```

```
## Observations: 83
## Variables: 3
## $ group <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ id    <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1...
## $ mean  <dbl> NA, 2.6857143, NA, NA, NA, NA, NA, 4.4285714, NA, NA, 48...
```

```
# (2) Make a summary data of group means,
# now using the mean of available responses for each subject:

PAINLS2 <- PAINL %>%
  group_by(group, id) %>%
  summarise( mean=mean(pain, na.rm = TRUE) ) %>%
  ungroup()

glimpse(PAINLS2)
```

```
## Observations: 83
## Variables: 3
## $ group <fct> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ id    <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1...
## $ mean  <dbl> 0.0000000, 2.6857143, 8.3333333, 27.8000000, 12.0000000,...
```

```
# Compare the t-test results:

t.test(mean ~ group, data = PAINLS1, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  mean by group
## t = -3.8465, df = 32, p-value = 0.000538
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -45.97397 -14.14032
## sample estimates:
## mean in group 1 mean in group 2
##        16.74286        46.80000
```

```
t.test(mean ~ group, data = PAINLS2, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  mean by group
## t = -4.1455, df = 81, p-value = 8.293e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -33.58920 -11.80261
## sample estimates:
## mean in group 1 mean in group 2
##        18.34181        41.03771
```