# Homework 9

STAT 601

Please do the following problems from the text book R Handbook and discuss your analysis with the rest of the class. Remember to have an annotated RMD file that will allow the rest of us to see what you are doing!

1. Following up with the Beat the Blues data from the video (package HSAUR3) do the following

   a. Construct boxplots to compare the factor variable in an analogous way to how we constructed boxplots in the video for the treatment variable. Discuss the results.

	drug <fct>	length <fct>	treatment <fct>	bdi.pre <dbl>	bdi.2m <dbl>	bdi.3m <dbl>	bdi.5m <dbl>	bdi.8m <dbl>
1	No	>6m	TAU	29	2	2	NA	NA
2	Yes	>6m	BtheB	32	16	24	17	20
3	Yes	<6m	TAU	25	20	NA	NA	NA
4	No	>6m	BtheB	21	17	16	10	9
5	Yes	>6m	BtheB	26	23	NA	NA	NA
6	Yes	<6m	BtheB	7	0	0	0	0

6 rows

```
## [1] "data.frame"
```

Treated as Usual
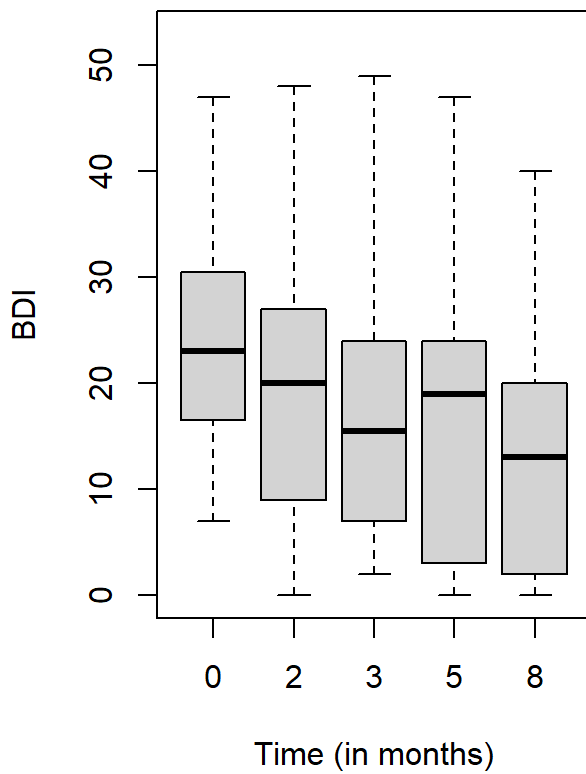
Beat the Blues
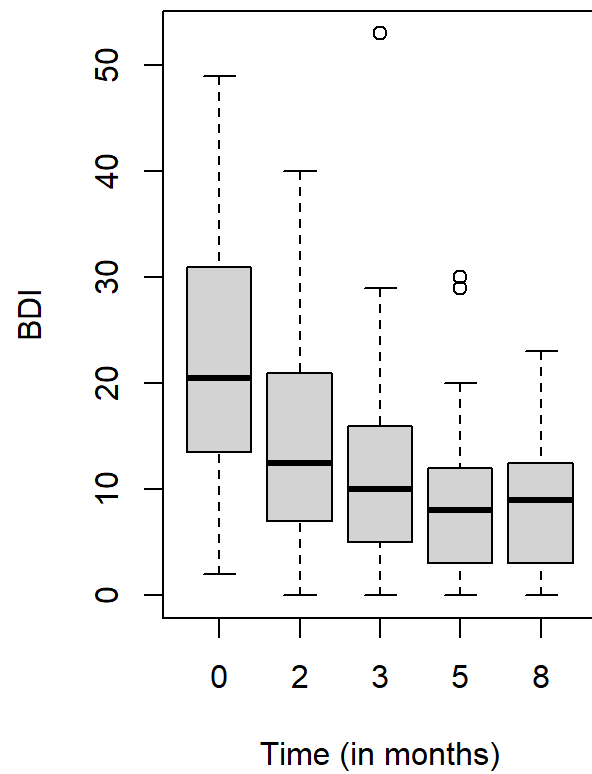
Time Depressed >6m

Time Depressed <6m

Antidepressants = Yes
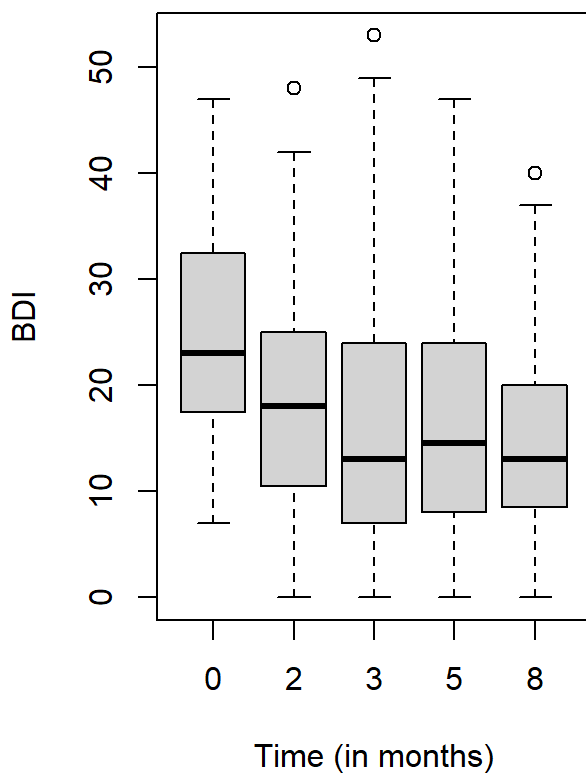
Antidepressants = No
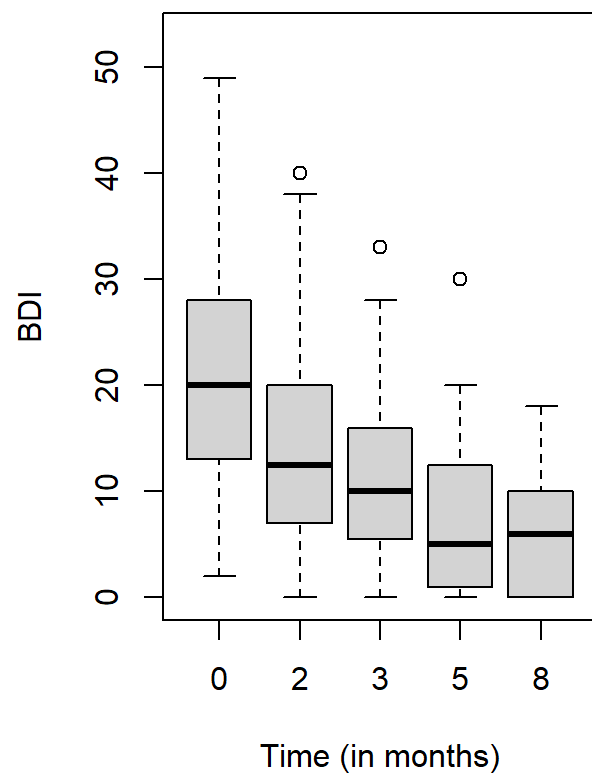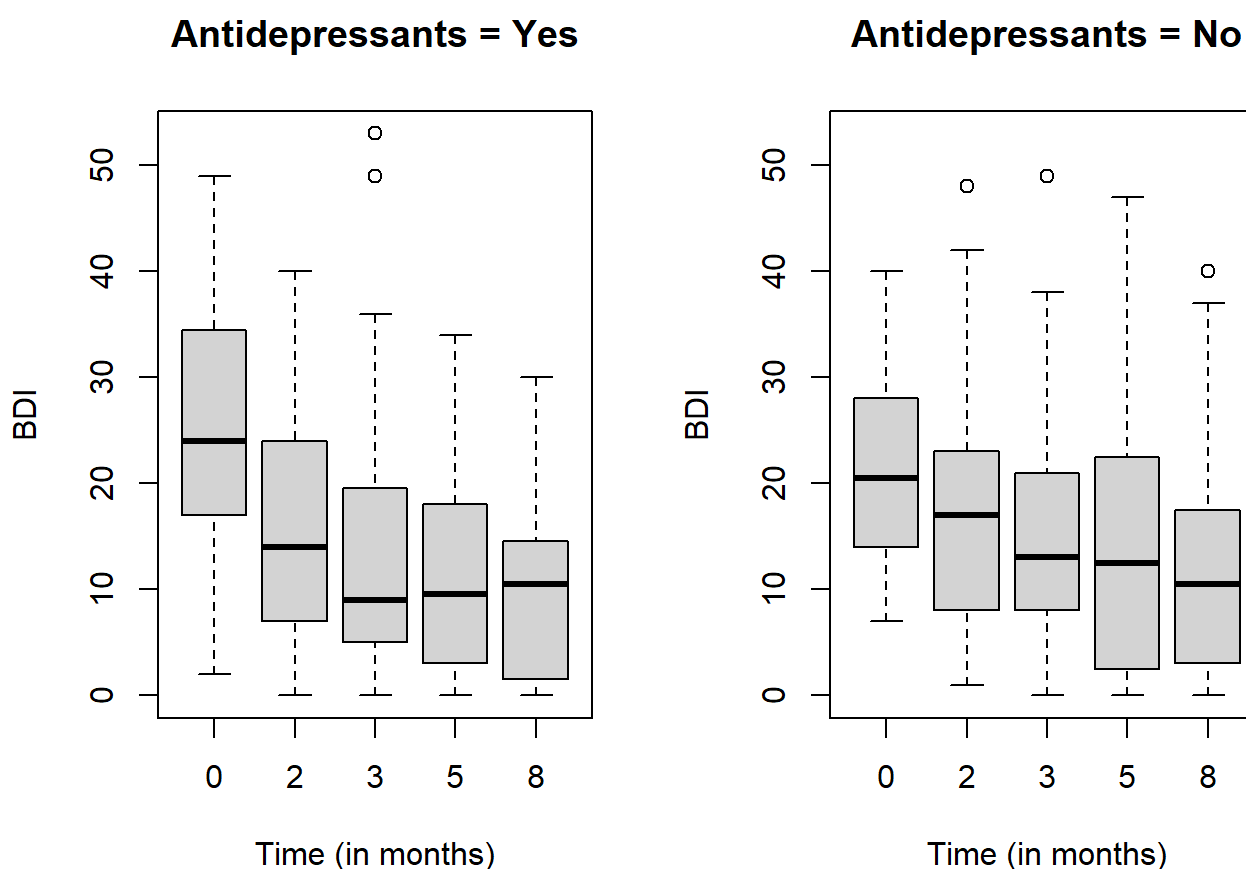

```
## [1] No  Yes
## Levels: No Yes
```

I created new box plots showing the other factors of the analysis which include whether the patient took anti-depressant drugs, and whether the patient currently had an episode of depression more or less than 6 months. Results are similar to the original boxplot comparing the treatments of the patients although there are more outliers. Also, I'm noticing more of a decrease over time in the Beck Depression Inventory II measure when applying treatments on patients who are in a depression episode less than 6 months.

```
   b. Repeat (a) for the \textbf{length} variable. Discuss the results.
```

see part a

  c. Use the \textit{lm} function to fit a model to the Beat the Blues data that assumes that
the repeated measurements are independent. Compare the results to those from fitting the rand
om intercept model \textit{BtheB_lmer1} from the video.

```
<div data-pagedtable="false">
  <script data-pagedtable-source type="application/json">
```

{"columns":[{"label":[""],"name":["_rn_"],"type":[""],"align":["left"]},{"label":["npar"],"na
me":[1],"type":["dbl"],"align":["right"]},{"label":["AIC"],"name":[2],"type":["dbl"],"alig
n":["right"]},{"label":["BIC"],"name":[3],"type":["dbl"],"align":["right"]},{"label":["logLi
k"],"name":[4],"type":["dbl"],"align":["right"]},{"label":["deviance"],"name":[5],"type":["db
l"],"align":["right"]},{"label":["Chisq"],"name":[6],"type":["dbl"],"align":["right"]},{"labe
l":["Df"],"name":[7],"type":["dbl"],"align":["right"]},{"label":["Pr(>Chisq)"],"name":[8],"ty
pe":["dbl"],"align":["right"]}],"data":[{"1":"8","2":"1887.492","3":"1916.570","4":"-935.746
0","5":"1871.492","6":"NA","7":"NA","8":"NA","_rn_":"BtheB_lmer1"},{"1":"10","2":"1891.038","
3":"1927.386","4":"-935.5189","5":"1871.038","6":"0.4541645","7":"2","8":"0.7968552","_rn_":"
BtheB_lmer2"}],"options":{"columns":{"min":{},"max":[10]},"rows":{"min":[10],"max":[10]},"pag
es":{}}}

```
  </script>
</div>
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = bdi ~ bdi.pre + time + treatment + drug + length +
##     (1 | subject), data = BtheB_long, REML = FALSE, na.action = na.omit)
##
## Linear Hypotheses:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0    5.59239    2.24244   2.494   0.0126 *
## bdi.pre == 0        0.63968    0.07789   8.212 2.22e-16 ***
## time == 0          -0.70476    0.14639  -4.814 1.48e-06 ***
## treatmentBtheB == 0 -2.32908   1.67036  -1.394   0.1632
## drugYes == 0       -2.82495    1.72684  -1.636   0.1019
## length>6m == 0      0.19708    1.63832   0.120   0.9043
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = bdi ~ bdi.pre + time + treatment + drug + length +
##     (time | subject), data = BtheB_long, REML = FALSE, na.action = na.omit)
##
## Linear Hypotheses:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0    5.61391    2.24541   2.500   0.0124 *
## bdi.pre == 0        0.64185    0.07784   8.245 2.22e-16 ***
```

```
## time == 0            -0.70214     0.15407   -4.557  5.18e-06 ***
## treatmentBtheB == 0 -2.37598     1.67129   -1.422   0.1551
## drugYes == 0        -2.86942     1.72766   -1.661   0.0967 .
## length>6m == 0       0.14041     1.63854    0.086   0.9317
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```


```
##
## Call:
## lm(formula = bdi ~ bdi.pre + time + treatment + drug + length,
##     data = BtheB_long, na.action = na.omit)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.6779  -5.4177   0.0151   5.3268  27.2688
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.32290    1.72765   4.239 3.08e-05 ***
## bdi.pre          0.57396    0.05497  10.440  < 2e-16 ***
## time            -0.93784    0.23650  -3.965 9.36e-05 ***
## treatmentBtheB  -3.32254    1.10069  -3.019  0.00278 **
## drugYes         -3.56866    1.14717  -3.111  0.00206 **
## length>6m        1.71067    1.11056   1.540  0.12463
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.674 on 274 degrees of freedom
##   (120 observations deleted due to missingness)
## Multiple R-squared:  0.395,  Adjusted R-squared:  0.384
## F-statistic: 35.78 on 5 and 274 DF,  p-value: < 2.2e-16
```

When comparing the random intercept models with a linear model of the data, the linear model, along with bdi.pre, and time, find that whether the Beat the Blues Treatment was used and whether the patient was already medicated with anti-depressants are significant factors for determining patient bdi.

d. Investigate and discuss whether there is any evidence of an interaction between treatment and time for the Beat the Blues data.

```
##
## Call:
## lm(formula = time ~ treatment, data = BtheB_long, na.action = na.omit)
##
## Residuals:
##    Min    1Q Median    3Q    Max
##  -2.50  -1.75  -0.50   1.25   3.50
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.500e+00  1.658e-01   27.14   <2e-16 ***
## treatmentBtheB 8.889e-16  2.299e-01    0.00        1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.297 on 398 degrees of freedom
## Multiple R-squared:  1.118e-29,  Adjusted R-squared:  -0.002513
## F-statistic: 4.449e-27 on 1 and 398 DF,  p-value: 1
```

```
##
## Call:
## lm(formula = bdi ~ bdi.pre + time * treatment + drug + length,
##     data = BtheB_long, na.action = na.omit)
##
## Residuals:
##      Min      1Q   Median      3Q     Max
## -23.7000  -5.4490   0.0309   5.3395  27.2771
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.36399    1.98536   3.709 0.000252 ***
## bdi.pre             0.57399    0.05508  10.421  < 2e-16 ***
## time               -0.94830    0.34266  -2.767 0.006035 **
## treatmentBtheB     -3.40264    2.19366  -1.551 0.122029
## drugYes            -3.56797    1.14938  -3.104 0.002108 **
## length>6m           1.71097    1.11261   1.538 0.125259
## time:treatmentBtheB 0.02000    0.47345   0.042 0.966336
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.689 on 273 degrees of freedom
##   (120 observations deleted due to missingness)
## Multiple R-squared:  0.395,  Adjusted R-squared:  0.3817
## F-statistic: 29.71 on 6 and 273 DF,  p-value: < 2.2e-16
```

When fitted to a linear model, it seems like there is do interaction between treatment and time.

e. Construct a plot of the mean profiles of both treatment groups in the Beat the Blues data, showing also standard deviation bars at each time point.
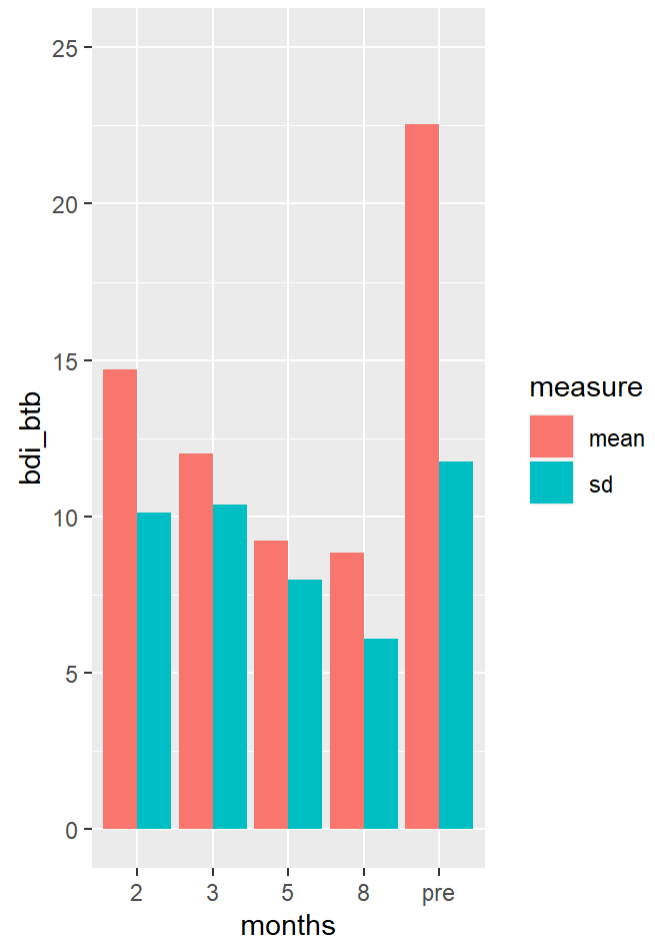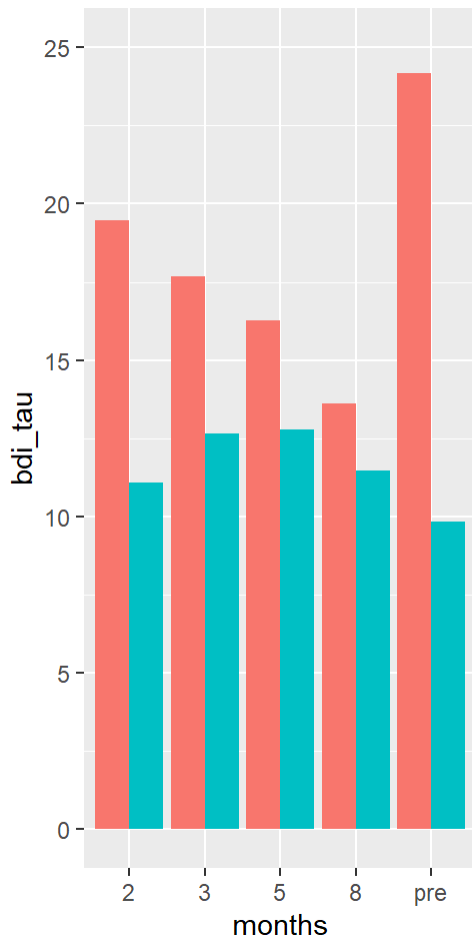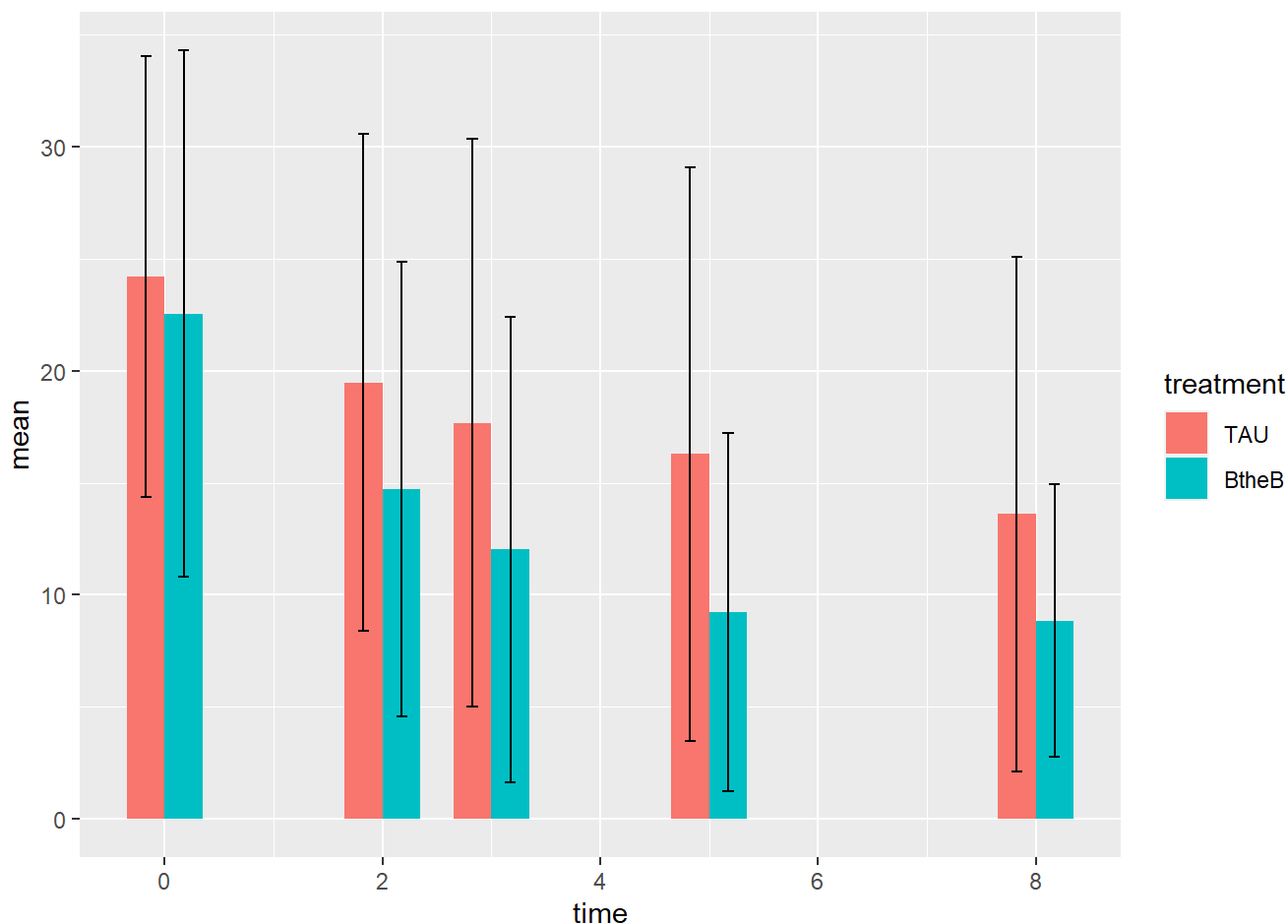
These bar charts show that the means of the two treatments tend to be noticeably lower for patients treated with the beat the blues therapy compared to the normal treatment and the standard deviations tend to be noticeably lower for months 5 and 8.

2. Consider the data from the package HSAUR3. This data shows the plasma inorganic phosphate levels for 33 subjects, 20 of whom are controls and 13 of whom have been classified as obese (Davis, 2002). Perform the following on this dataset

    a. Construct boxplots by group and discuss.

```
## [1] control obese
## Levels: control obese
```

	group	t0	t0.5	t1	t1.5	t2	t3	t4	t5
	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	control	4.3	3.3	3.0	2.6	2.2	2.5	3.4	4.4
2	control	3.7	2.6	2.6	1.9	2.9	3.2	3.1	3.9
3	control	4.0	4.1	3.1	2.3	2.9	3.1	3.9	4.0
4	control	3.6	3.0	2.2	2.8	2.9	3.9	3.8	4.0
5	control	4.1	3.8	2.1	3.0	3.6	3.4	3.6	3.7
6	control	3.8	2.2	2.0	2.6	3.8	3.6	3.0	3.5

6 rows

```
## [1] 33
```
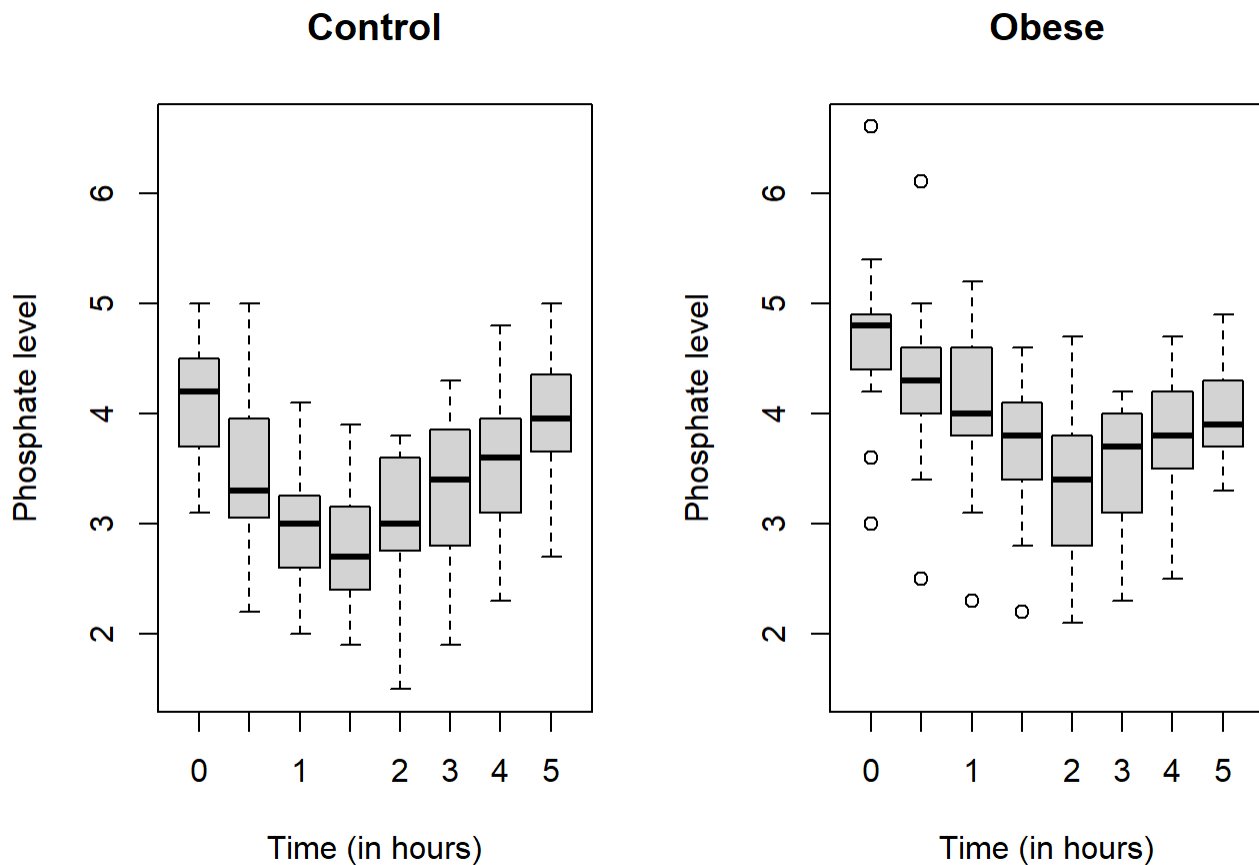
Control

Obese

Results of the boxplots seem to differ between the control and obese groups, with the obese group having more outliers and a trend with a trough at 2 hours in compared to the control group which has a trough at 1.5 hours in.

```
    b. Produce separate plots of the profiles of the individuals in each group.
```

Control Subjects

Obese Subjects

c. Guided by how these plots fit, which linear mixed effects models do you think might be sensible? (Hint: Discuss intercept and slope, intercept and interaction).

After comparing the two types of linear mixed effects models, I'd say that the random intercept model would be more sensible to fit to this dataset due to the measures not really having a particularly noticeable pattern in variance other than some subjects having a high variance while others having a low variance. Time between when the measures were take does not seem to be having a noticeable impact on variance.

d. Convert the data to long version and fit the model of your choice and discuss the results.

	group <fct>	t0 <dbl>	subject <fct>	hour <dbl>	Phos_Level <dbl>
1.control.4.3.0.5	control	4.3	1	0.5	3.3
2.control.3.7.0.5	control	3.7	2	0.5	2.6
3.control.4.0.5	control	4.0	3	0.5	4.1
4.control.3.6.0.5	control	3.6	4	0.5	3.0
5.control.4.1.0.5	control	4.1	5	0.5	3.8
6.control.3.8.0.5	control	3.8	6	0.5	2.2

6 rows

	n…	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
phosphate_lmer1	6	411.7727	432.4272	-199.8864	399.7727	NA	NA	NA
phosphate_lmer2	8	404.2254	431.7647	-194.1127	388.2254	11.54737	2	0.003108287

2 rows

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Phos_Level ~ group + hour + t0 + (1 | subject)
##    Data: phosphate_long_2
##
##      AIC      BIC   logLik deviance df.resid
##    411.8    432.4   -199.9    399.8      225
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3400 -0.6734  0.0299  0.7167  3.1689
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 0.05969  0.2443
##  Residual             0.29099  0.5394
## Number of obs: 231, groups:  subject, 33
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.45577    0.36579   1.246
## groupobese   0.15611    0.12197   1.280
## hour         0.09110    0.02332   3.906
## t0           0.63084    0.08537   7.389
##
## Correlation of Fixed Effects:
##            (Intr) gropbs hour
## groupobese  0.244
## hour       -0.155  0.000
## t0         -0.969 -0.369  0.000
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Phos_Level ~ group + hour + t0 + (hour | subject)
##    Data: phosphate_long_2
##
##      AIC      BIC   logLik deviance df.resid
##    404.2    431.8   -194.1    388.2      223
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4685 -0.5788  0.0747  0.5836  3.0899
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  subject  (Intercept) 4.385e-05 0.006622
##           hour        1.261e-02 0.112297 0.96
##  Residual             2.588e-01 0.508769
## Number of obs: 231, groups:  subject, 33
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.04788    0.32455   0.148
## groupobese   0.34732    0.10874   3.194
## hour         0.09110    0.02943   3.096
## t0           0.70716    0.07611   9.291
##
## Correlation of Fixed Effects:
##            (Intr) gropbs hour
## groupobese  0.245
## hour       -0.121  0.000
## t0         -0.973 -0.369  0.000
## optimizer (nloptwrap) convergence code: 0 (OK)
## unable to evaluate scaled gradient
## Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = Phos_Level ~ group + hour + t0 + (1 | subject),
##     data = phosphate_long_2, REML = FALSE, na.action = na.omit)
##
## Linear Hypotheses:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0  0.45577    0.36579   1.246    0.213
## groupobese == 0   0.15611    0.12197   1.280    0.201
## hour == 0         0.09110    0.02332   3.906 9.37e-05 ***
## t0 == 0           0.63084    0.08537   7.389 1.48e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

```
##
##    Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = Phos_Level ~ group + hour + t0 + (hour | subject),
##     data = phosphate_long_2, REML = FALSE, na.action = na.omit)
##
## Linear Hypotheses:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept) == 0  0.04788    0.32455   0.148  0.88271
## groupobese == 0   0.34732    0.10874   3.194  0.00140 **
## hour == 0         0.09110    0.02943   3.096  0.00196 **
## t0 == 0           0.70716    0.07611   9.291  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Univariate p values reported)
```

The random intercept model seems to be a better fit based on AIC and BIC than the fitted random intercept model. Furthermore, when determining phosphate levels, it seems like whether the subject is in the control or obese group has a significant effect.