

```
# Note to group: Emily loaded HSAUR2, not HSAUR as the Rpubs document state.  
# They look the same  
library(HSAUR2)  
  
## Loading required package: tools  
  
library(ggplot2)  
theme_set(theme_bw())  
attach(BtheB)
```

Background/Context

The data we are using is from a clinical trial called “Beat the Blues”. The Beat the Blues computer program was designed to deliver cognitive behavioural therapy to depressed patients. This dataset is from the R package “HSAUR2”. The data contains 200 observations each relating to a study subject. The study measured 8 variables: drug, length, treatment, bdi.pre, bdi.2m, bdi.3m, bdi.5m and bdi.8m. The variable “drug” refers to if the subject took anti-depressants, “length” is the length of the subjects current episode of depression (either less than 6 months or more than six months), “treatment” is the treatment the subject was placed on (either treatment as usual(TAU), or Beat the Blues (BtheB)). The bdi variables refer to the Beck Depression Inventory at certain points during the study. The variables are titled 3m, 5m, and 8m to refer to 1, 3, and 6 months followup *relative to the 2-month post-treatment visit*.

```
summary(BtheB[, c(1:4, 6, 8)])
```

##	drug	length	treatment	bdi.pre	bdi.3m	bdi.8m
##	No :56	<6m:49	TAU :48	Min. : 2.0	Min. : 0.0	Min. : 0.0
##	Yes:44	>6m:51	BtheB:52	1st Qu.:15.0	1st Qu.: 6.0	1st Qu.: 3.0
##				Median :22.0	Median :13.0	Median :10.5
##				Mean :23.3	Mean :14.8	Mean :11.1
##				3rd Qu.:30.2	3rd Qu.:20.0	3rd Qu.:15.2
##				Max. :49.0	Max. :53.0	Max. :40.0
##					NA's :27	NA's :48

Variable Descriptions

The variables we are interested in are “bdi.pre” and “bdi.8m”, Beck Depression Inventory II before treatment and after six months follow-up.

```
# Subsetting the data
BtheB.tau = BtheB[treatment == "TAU", ]
BtheB.btb = BtheB[treatment == "BtheB", ]
```

```
# This function allows us to put multiple plots on the same page.
multiplot <- function(..., plotlist = NULL, file, cols = 1, layout = NULL) {
  require(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel ncol: Number of columns of plots nrow: Number of rows
    # needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)), ncol = cols,
                     nrow = ceiling(numPlots/cols))
  }

  if (numPlots == 1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

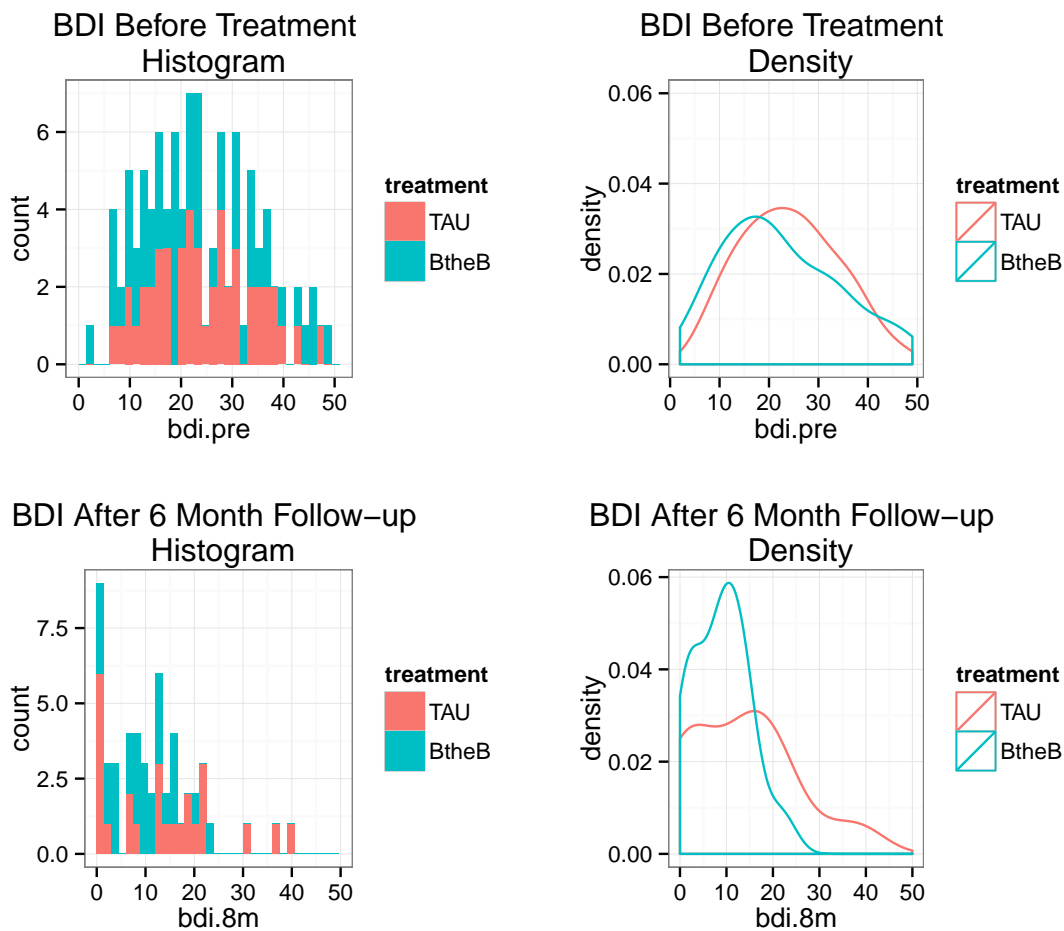
      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row, layout.pos.col =
    }
  }
}
```

```
h1 = qplot(bdi.pre, fill = treatment, main = "BDI Before Treatment \n Histogram",
           binwidth = 1.5)
h2 = qplot(bdi.8m, fill = treatment, main = "BDI After 6 Month Follow-up \n Histogram",
           binwidth = 1.5, xlim = c(0, 50))

p1 = qplot(bdi.pre, geom = "density", color = treatment, main = "BDI Before Treatment \n De
           ylim = c(0, 0.06))
p2 = qplot(bdi.8m, geom = "density", color = treatment, main = "BDI After 6 Month Follow-up
           xlim = c(0, 50))

multiplot(h1, h2, p1, p2, cols = 2)

## Loading required package: grid
```



```
# multiplot(p1,p2, cols=2)

# I am not sure about this part. Why are we looking at BDI after one month?
# Also the second plot is either mislabeled in the title or the wrong
# variable is being called - I'm not sure which so they are just commented
# out for now.

# qplot(bdi.3m, geom = 'density', color = treatment, main = 'BDI After 1
# Month Follow-up') qplot(bdi.3m, fill = treatment, main = 'BDI After 6 Month
# Follow-up')
```

The before treatment groups look similar. However, the after 1 month follow-up treatment groups look severely different. After 1 month follow-up, it seems more BtheB subjects score

lower on the BDI scale than the TAU group (which means the subjects are less depressed). This suggests the BtheB treatment works better than treatment as usual.

Maritza's suggestion for this section:

Before treatment, both groups report BDIs that range between 0 and 50. The distribution is similar for both, as shown above. At this point we have not differentiated between those who are on drug treatment and those who are not.

At the six months followup date (or 8 months since the initial data were taken), we see shifts in the distribution of the BDI for both groups. The BtheB treatment group reported no BDI levels above about 25, and the TAU group reported no BDI levels above 40. The density figures also show a shift to the left, or to lower BDI levels, for both treatment groups. The density graph in particular illustrates a shift in the mean BDI level from about 20-25 pre-treatment to about 10 8-months post-treatment.

Hypothesis

1. Patients receiving either treatment will experience a decline in depression symptoms as measured by the Beck Depression Inventory.
2. Patients receiving BtheB treatment will experience a greater decline in depression as measured by the Beck Depression Inventory compared to those receiving Treatment As Usual.

Missing Data

```
# M: I am not sure what is going on here - I never saw a discussion where we  
# decided to change our variables. Emily sent this to me:
```

```
# ``I choose 3 instead of 8. Once he said we have to incorporate missing  
# variables, i checked the missing counts and 27 sounds better than half of  
# the days... 48'
```

```
# But I don't get this because I don't understand the last sentence about  
# half of the days and how the number of days is related to the number of  
# participants in the study. I ran the numbers and the makeup of the  
# remaining participants after you take out the NAs is basically the same.  
# Is there some rule on how many people have to remain in a study to use its  
# data even if people drop out? Anyways I think the point here as far as the  
# homework goes is for us to discuss it so hopefully other folks also have  
# some thoughts.
```

Here is my code that checked this:

```
BtheB8m = na.omit(BtheB)
sum(BtheB8m$treatment == "BtheB")

## [1] 27

sum(BtheB8m$treatment == "TAU")

## [1] 25

sum(BtheB8m$drug == "Yes")

## [1] 24

sum(BtheB8m$drug == "No")

## [1] 28
```

We chose these two variables to measure the association using a baseline measure and a midpoint measure. Using the 3 month mark (after 1 month follow-up), only 27 subjects having missing data. We originally wished to use the end of the study, however 48 subjects had missing data.

```
sum(is.na(bdi.pre) == TRUE)

## [1] 0

sum(is.na(bdi.3m) == TRUE)

## [1] 27

sum(is.na(bdi.8m) == TRUE)

## [1] 48
```

Analysis and Results

We plotted linear regressions for the BDI levels pre-treatment and three months/six months post-followup.

```
blm <- lm(bdi.3m ~ bdi.pre)
blm

##
## Call:
## lm(formula = bdi.3m ~ bdi.pre)
##
## Coefficients:
## (Intercept)      bdi.pre
##      0.0645      0.6369

blm1 <- lm(bdi.3m ~ bdi.pre, data = BtheB.tau)
blm1

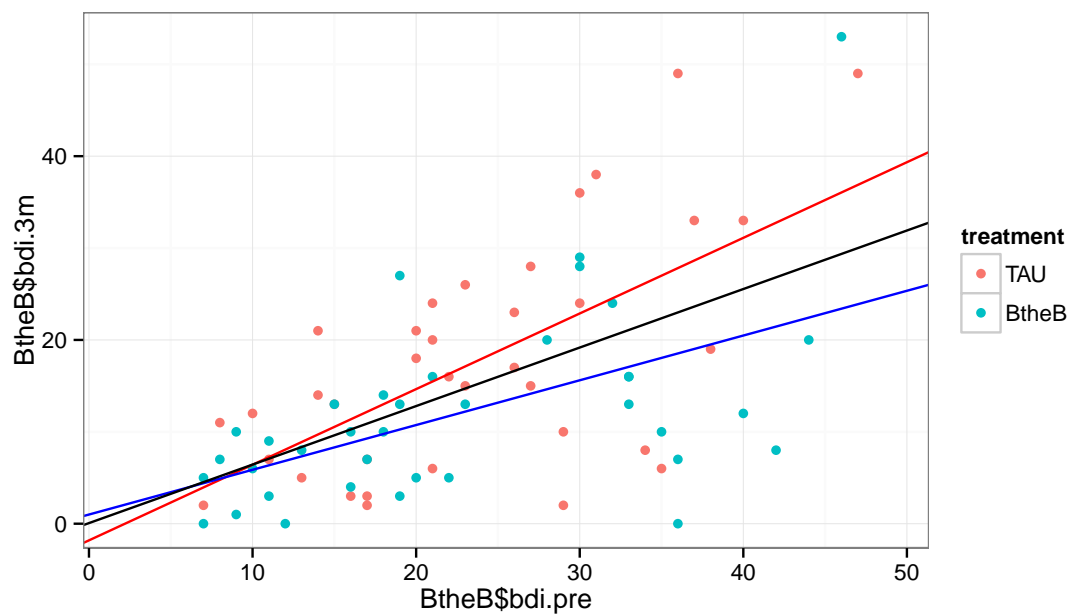
##
## Call:
## lm(formula = bdi.3m ~ bdi.pre, data = BtheB.tau)
##
## Coefficients:
## (Intercept)      bdi.pre
##      -1.814      0.823

blm2 <- lm(bdi.3m ~ bdi.pre, data = BtheB.btb)
blm2

##
## Call:
## lm(formula = bdi.3m ~ bdi.pre, data = BtheB.btb)
##
## Coefficients:
## (Intercept)      bdi.pre
##      0.995      0.487

qplot(BtheB$bdi.pre, BtheB$bdi.3m, data = BtheB, color = treatment) + geom_abline(intercept = 0.06449, slope = 0.8231, colour = "red") + geom_abline(intercept = 0.9949, slope = 0.4871, colour = "blue") + geom_abline(intercept = 0.06449, slope = 0.63686, colour = "black")

## Warning: Removed 27 rows containing missing values (geom_point).
```



```
blm3 <- lm(bdi.8m ~ bdi.pre, data = BtheB.tau)
blm4 <- lm(bdi.8m ~ bdi.pre, data = BtheB.btb)
qplot(BtheB$bdi.pre, BtheB$bdi.8m, data = BtheB, color = treatment) + geom_abline(intercept = 2.5855,
  slope = 0.5779, colour = "red") + geom_abline(intercept = 3.7618, slope = 0.2314,
  colour = "blue") + geom_abline(intercept = 2.5855, slope = 0.3714, colour = "black")

## Warning: Removed 48 rows containing missing values (geom_point).
```