Homework 4

Mengkun Chen (mengkun21@vt.edu)

Problem 3

```
library(data.table)
library(tidyverse)
library(ggplot2)
```

Part A

```
thick<-fread("https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/ThicknessGauge.dat",
             header = FALSE, data.table = FALSE)
```

- Rename the column: column 2 and 3 are measurements under operator 1, column 4 and 5 are measurements under operator 2, column 6 and 7 are measurements under operator 3.

- Compute mean thickness for 10 different parts using different operators.

- Keep the mean thickness as the measure of wall thickness

```
thick<- thick %>%
  rename(Part = V1, G11 = V2, G12 = V3, G21 = V4, G22 = V5, G31 = V6, G32 = V7) %>%
  mutate(G1 = (G11+G12)/2, G2 = (G21+G22)/2, G3 = (G31+G32)/2) %>%
  select(Part, G1, G2, G3)

knitr::kable(thick, digits = 4, caption = "Wall Thickness")
```

Table 1: Wall Thickness

| Part	G1	G2	G3
1	0.9525	0.9540	0.9550
2	0.9560	0.9565	0.9575
3	0.9555	0.9555	0.9565
4	0.9570	0.9575	0.9575
5	0.9570	0.9570	0.9580
6	0.9580	0.9570	0.9580
7	0.9565	0.9575	0.9575
8	0.9560	0.9565	0.9570
9	0.9540	0.9540	0.9550
10	0.9545	0.9550	0.9545

Summary of dataset:

```
knitr::kable(summary(thick[,-1]), digits = 4, caption = "Summary of Wall Thickness")
```

Table 2: Summary of Wall Thickness

G1	G2	G3
Min. :0.9525	Min. :0.9540	Min. :0.9545
1st Qu.:0.9547	1st Qu.:0.9551	1st Qu.:0.9554
Median :0.9560	Median :0.9565	Median :0.9573
Mean :0.9557	Mean :0.9560	Mean :0.9566
3rd Qu.:0.9569	3rd Qu.:0.9570	3rd Qu.:0.9575
Max. :0.9580	Max. :0.9575	Max. :0.9580

Draw bar plot for wall thickness of 10 parts of the wall.

```
thick %>% gather(key = Guage, value = measure, G1:G3) %>%
  ggplot(aes(x = Guage, y = measure)) + geom_boxplot() +
  theme_light()
```
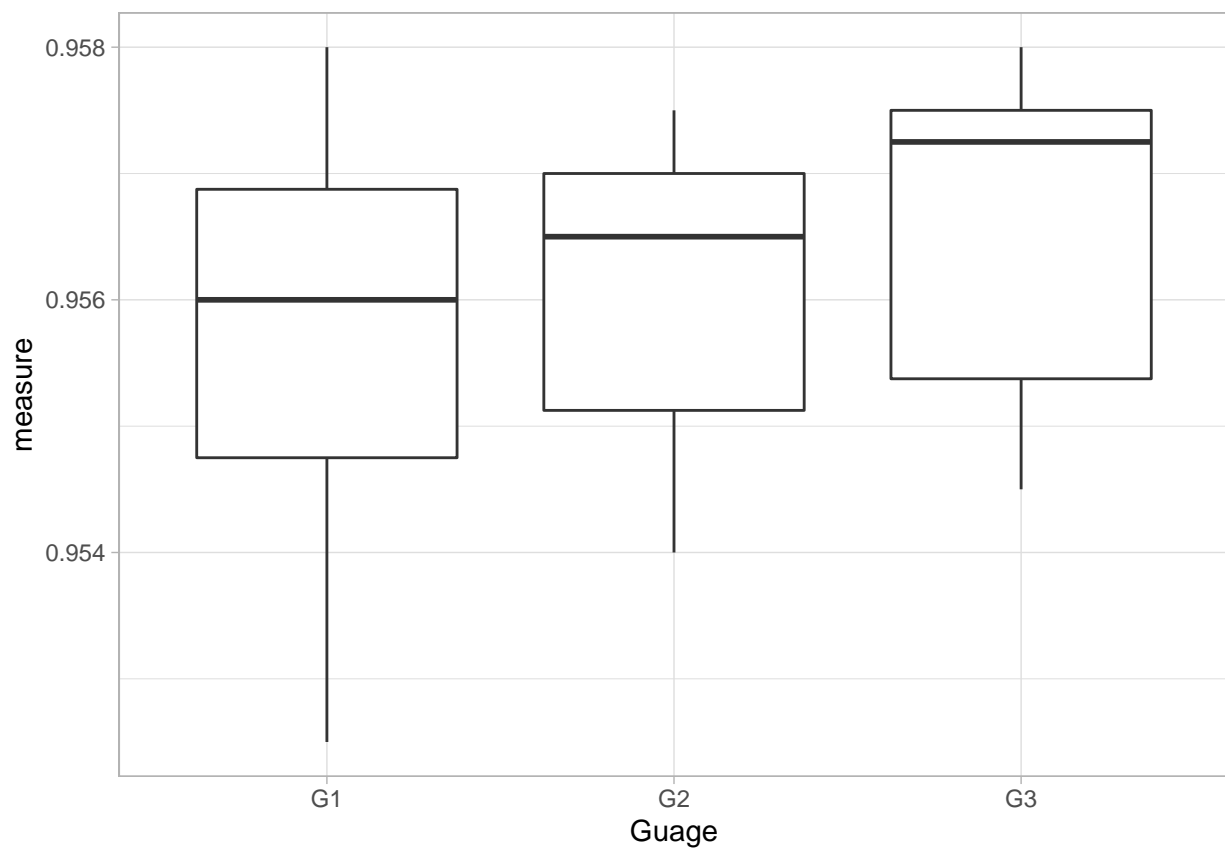

Operator 3 tend to get higher thickness measurements.

Part B

```
bbweight<-fread("http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat", header =
colnames(bbweight)<-rep(c("Body Wt", "Brain Wt"), 3)
```

- Combine all the body weight and brain weight into two separate columns.

```
body<-c(bbweight[,1],bbweight[,3],bbweight[,5])
brain<-c(bbweight[,2],bbweight[,4],bbweight[,6])
bb<-data.frame("Body Wt" = body, "Brain Wt" = brain)
knitr::kable(head(bb), digits = 4, caption = "Body-Brain Weight")
```

Table 3: Body-Brain Weight

Body.Wt	Brain.Wt
3.385	44.5
0.480	15.5
1.350	8.1
465.000	423.0
36.330	119.5
27.660	115.0

Summary of dataset:

```
knitr::kable(summary(bb), digits = 4, caption = "Body-Brain Weight Summary")
```

Table 4: Body-Brain Weight Summary

Body.Wt	Brain.Wt
Min. : 0.005	Min. : 0.100
1st Qu.: 0.532	1st Qu.: 3.975
Median : 3.150	Median : 16.250
Mean : 198.131	Mean : 277.422
3rd Qu.: 35.333	3rd Qu.: 128.875
Max. :6654.000	Max. :5712.000

From the summary table, the dataset may contain several mis-recorded observation, which are far away from the majority observations.

Draw dot plot to show the relationship between body weight and brain weight ignorning the observations whose body weight and brain weight are greater than the third quartile respectively.

```
plot(bb$Body.Wt, bb$Brain.Wt, xlab = "Body Weight", ylab = "Brain Weight",
     main = "Brain Weight vs. Body Weight", xlim = c(0,40), ylim = c(0,130))
```

Brain Weight vs. Body Weight
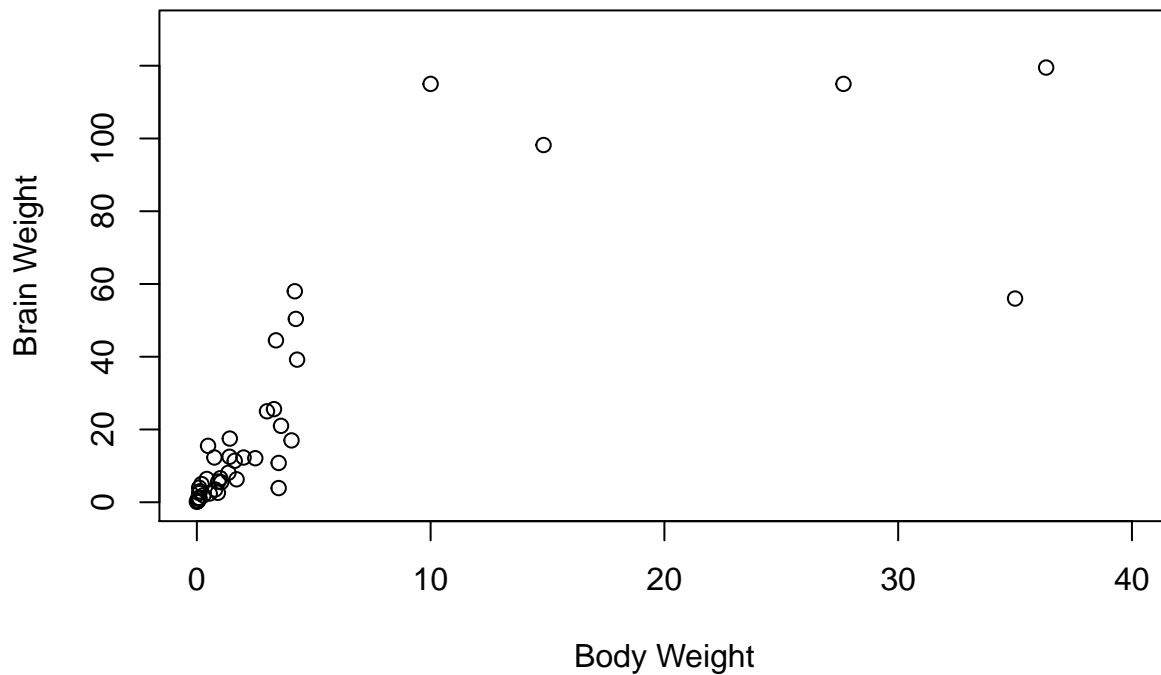

```r
lm(bb$Brain.Wt~bb$Body.Wt)
```

```
##
## Call:
## lm(formula = bb$Brain.Wt ~ bb$Body.Wt)
##
## Coefficients:
## (Intercept)    bb$Body.Wt
##     85.9552        0.9664
```

Brain weight and body weight are positive related ignorning some obvious outliers.

Part C

```r
jump<-fread("http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat", data.table = FALSE
colnames(jump)<-rep(c("Year", "Long Jump"),4)
```

- Combine all year and long jump columns into two seperate columns.

```r
year<-c(jump[,1], jump[,3], jump[,5], jump[,7])
long.jump<-c(jump[,2], jump[,4], jump[,6], jump[,8])
lj<-data.frame("Year" = year, "Long Jump" = long.jump)
knitr::kable(head(lj), caption = "Gold Medal Performance for Long Jump")
```

Table 5: Gold Medal Performance for Long Jump

Year	Long.Jump
-4	249.75
0	282.88
4	289.00
8	294.50
24	293.13
28	304.75

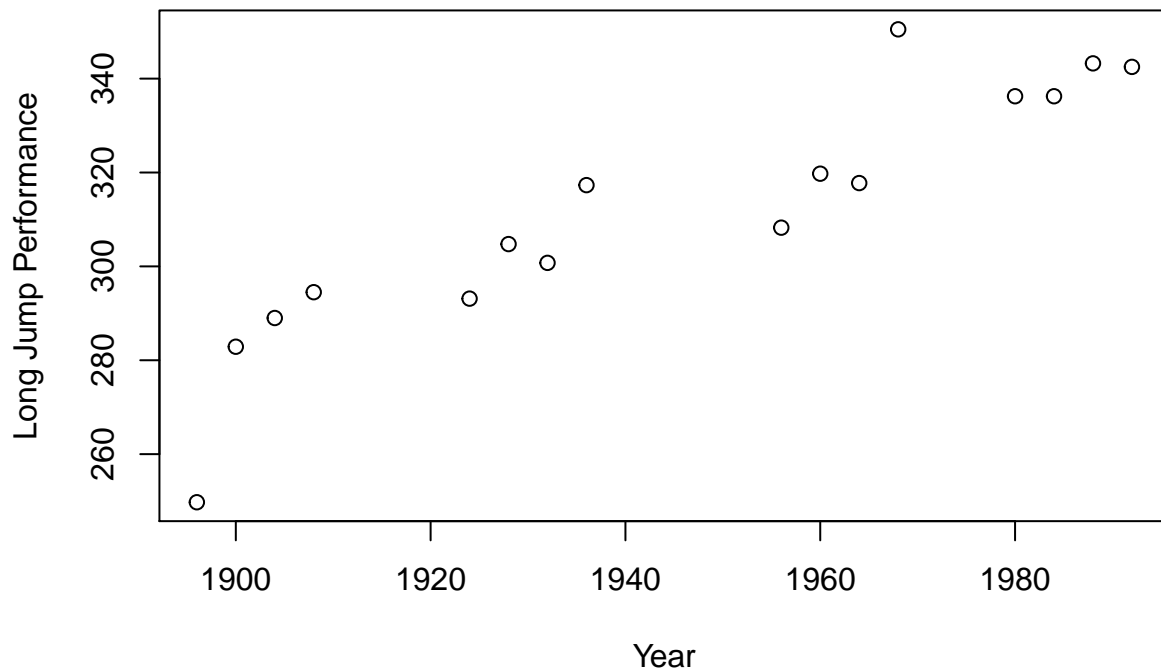Summary of Gold Medal performance for Olympic Men's Long Jump.

```r
knitr::kable(summary(lj), caption = "Summary of Gold Medal Performance for Long Jump")
```

Table 6: Summary of Gold Medal Performance for Long Jump

Year	Long.Jump
Min. :-4	Min. :249.8
1st Qu.:20	1st Qu.:294.2
Median :46	Median :312.8
Mean :45	Mean :311.7
3rd Qu.:71	3rd Qu.:336.2
Max. :92	Max. :350.5

```r
plot(lj$Year+1900, lj$Long.Jump, xlab = "Year", ylab = "Long Jump Performance",
     main = "Gold Medal Performance for Long Jump")
```

Gold Medal Performance for Long Jump

Gold Medal performance for Olympic Men's Long Jump increase as year increase.

Part D

```
yield<-fread("http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat",
             header = FALSE)
colnames(yield)<-c("Variety", "D1", "D2", "D3")
```

- Separate tomato yields under different densities.

```
d1<-yield %>%
  select(D1) %>%
  separate(col = D1, into = c("y1", "y2", "y3"), sep = ",") %>%
  mutate(Density = c("D1","D1"))
d2<-yield %>%
  select(D2) %>%
  separate(col = D2, into = c("y1", "y2", "y3"), sep = ",") %>%
  mutate(Density = c("D2","D2"))
d3<-yield %>%
  select(D3) %>%
  separate(col = D3, into = c("y1", "y2", "y3"), sep = ",") %>%
  mutate(Density = c("D3","D3"))
tomato<-cbind("Variety" = rep(c("v1", "v2"), 2), rbind(d1, d2, d3)) %>%
  gather(key = y, value = yield, y1:y3) %>%
```

```
  select(Variety, Density, yield)
knitr::kable(head(tomato), caption = "Tomator Yield")
```

Table 7: Tomator Yield

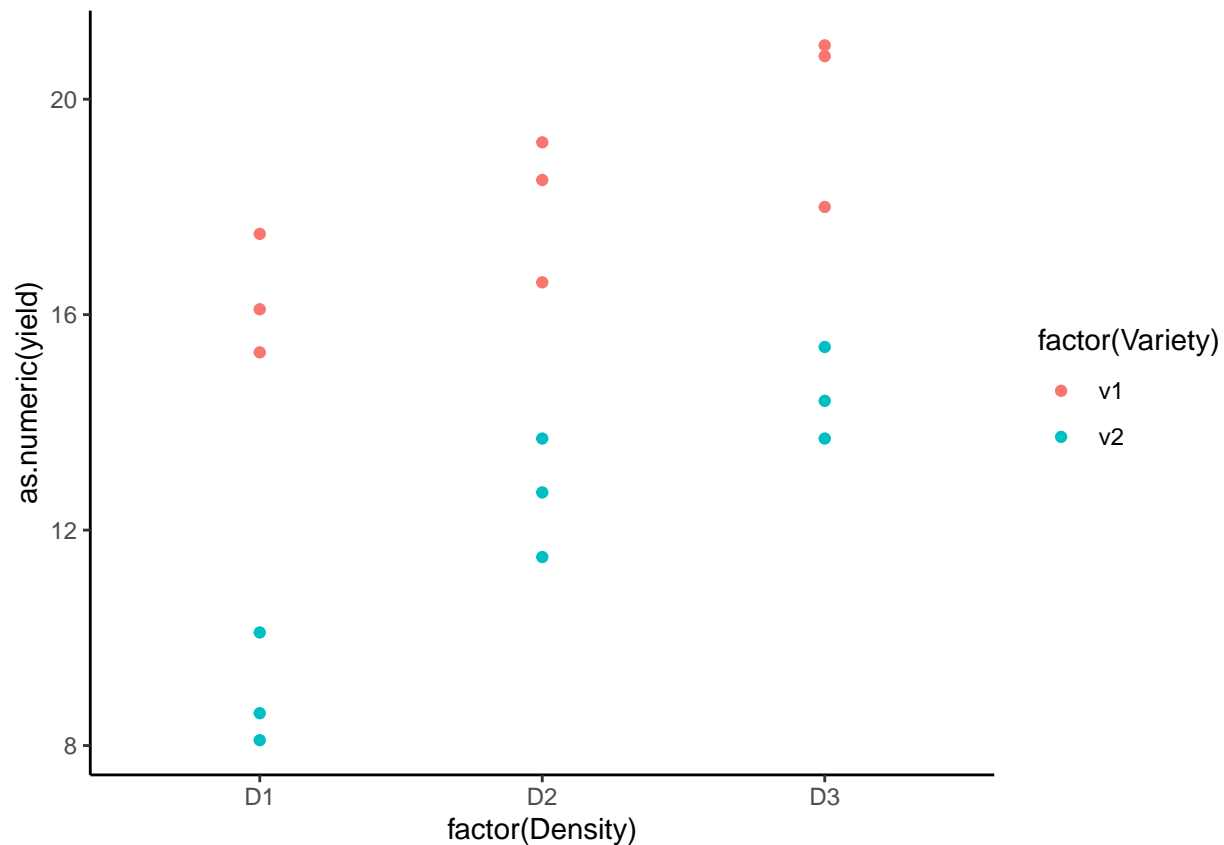| Variety	Density	yield
v1	D1	16.1
v2	D1	8.1
v1	D2	16.6
v2	D2	12.7
v1	D3	20.8
v2	D3	14.4

Summary of tomato yield:

```
knitr::kable(summary(tomato), caption = "Summary of Tomato Yield")
```

Table 8: Summary of Tomato Yield

Variety	Density	yield
Length:18	Length:18	Length:18
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

```
tomato %>% ggplot(aes(x = factor(Density), y = as.numeric(yield),
                col = factor(Variety))) +
  geom_point() +
  theme_classic()
```

Variety 2 has lower yield than variety 1. Tomato yields increase when planting density increase.

Part E

```
larvae<-fread("https://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LarvaeControl.dat", header = TRUE
```

- Separate data according to age.

- Rearrange the dataset into four columns: block, treat, count and age.

```
A1<-larvae[,1:6] %>%
  rename("T1" = "1", "T2" = "2", "T3" = "3", "T4" = "4", "T5" = "5") %>%
  gather(key = Treat, value = count, T1:T5) %>%
  mutate(Age = rep(1, 40))
A2<-larvae[,c(1,7:11)] %>%
  rename("T1" = "\t1", "T2" = "2", "T3" = "3", "T4" = "4", "T5" = "5") %>%
  gather(key = Treat, value = count, T1:T5) %>%
  mutate(Age = rep(2, 40))
larvae_count<-rbind(A1, A2)
knitr::kable(head(larvae_count), caption = "Larvae Counts")
```

Table 9: Larvae Counts

| Block	Treat	count	Age
1	T1	13	1
2	T1	29	1
3	T1	5	1
4	T1	5	1
5	T1	0	1
6	T1	1	1

Summary of dataset:

```r
larvae_count$Block<-as.factor(larvae_count$Block)
larvae_count$Treat<-as.factor(larvae_count$Treat)
larvae_count$Age<-as.factor(larvae_count$Age)
knitr::kable(summary(larvae_count), caption = "Summary of Larvae Counts")
```

Table 10: Summary of Larvae Counts

Block	Treat	count	Age
1 :10	T1:16	Min. : 0.00	1:40
2 :10	T2:16	1st Qu.: 2.75	2:40
3 :10	T3:16	Median : 5.50	NA
4 :10	T4:16	Mean :10.50	NA
5 :10	T5:16	3rd Qu.:13.00	NA
6 :10	NA	Max. :61.00	NA
(Other):20	NA	NA	NA

```r
larvae_count %>% ggplot(aes(x = Treat, y = count, col = Age, alpha = Block)) +
  geom_point() +
  theme_classic()
```
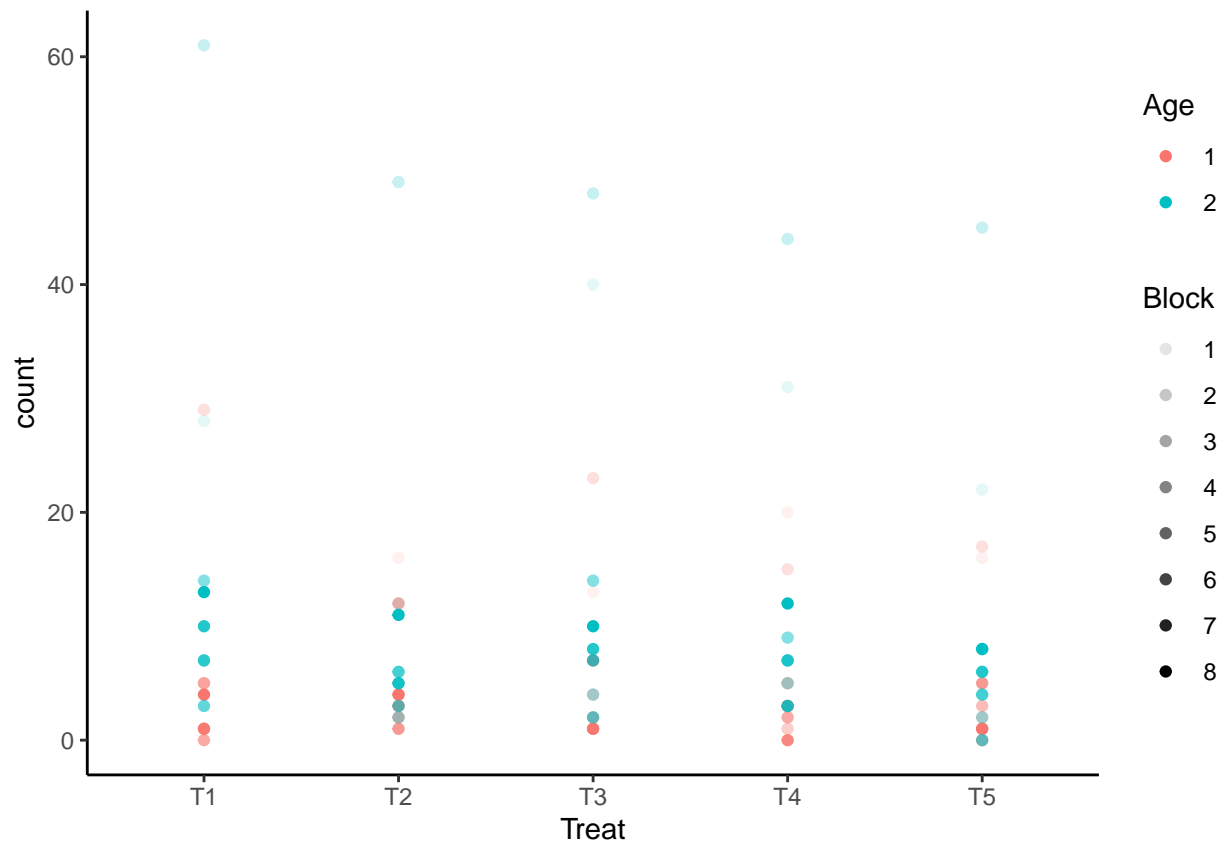
Larvae counts are larger in age 2 than those in age 1. Blocks have significant effect on larvae counts and the effect decrease as block number increase. No obvious treatment effect is shown in the plot.