Kumi Michael Osei

2023-10-22

Part A

Ten parts were selected randomly from a line and duplicate measurements of the part's wall thickness were taken by each of three operators of the measurement apparatus. Click on the link and you will see the data in your browser.

```
operator<-c(1 ,2 ,3)
op_rep<-rep(operator, each =2)
op_rep1<-seq(1,length(op_rep), by = 1)
op_rep2<-c(1,2)
var<-rep("op", length(op_rep))
lab<-c("obs", paste(var,op_rep,op_rep1, sep = "."))
dat1<-read.csv("partA_data.csv", header = F )
colnames(dat1)<-c(lab)
dat<-dat1 %>%
  gather("obs_code","value",op.1.1:op.3.6)%>%
  mutate(obs=as.numeric(obs))%>%
  mutate(value=as.numeric(value))
dat<-dat[order(dat$obs), ]
dat$operator<-rep(op_rep, times = length(dat1$obs))
dat$replicate<-rep(op_rep2, times = 3*length(dat1$obs))
dat$part<-rep(seq(1,10,by =1), each = length(op_rep1))
tidy_data<-(dat[, c(3,4,5,6)])
rownames(tidy_data) <- NULL
tidy_data<-tidy_data %>%
  mutate(operator=as.factor(operator))%>%
  mutate(replicate=as.factor(replicate))%>%
  mutate(part=as.factor(part))
head<-head(tidy_data,15)
kable(head,caption="Tidy data for the duplicate
      measurements of part's wall thickness by
      three operators for 10 parts")
```

Table 1: Tidy data for the duplicate measurements of part's wall thickness by three operators for 10 parts

value	operator	replicate	part
0.953	1	1	1
0.952	1	2	1
0.954	2	1	1
0.954	2	2	1
0.954	3	1	1

value	operator	replicate	part
0.956	3	2	1
0.956	1	1	2
0.956	1	2	2
0.956	2	1	2
0.957	2	2	2
0.958	3	1	2
0.957	3	2	2
0.956	1	1	3
0.955	1	2	3
0.956	2	1	3

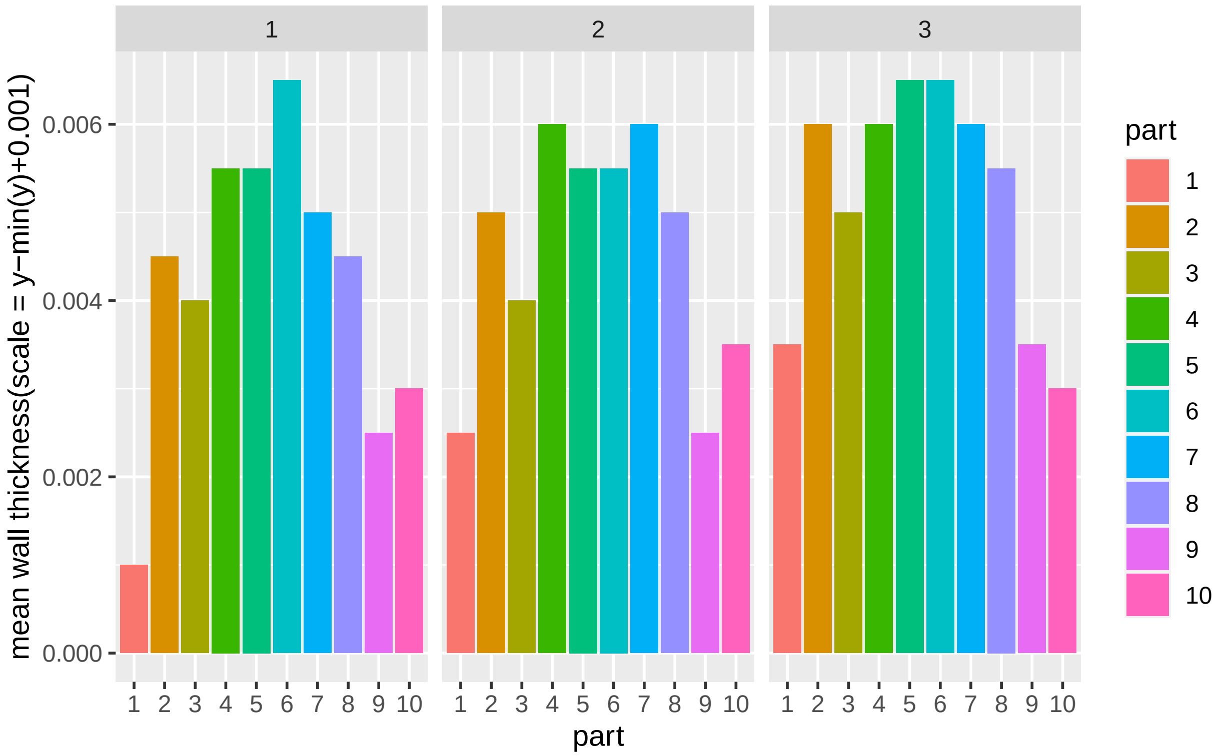
```r
#Averaging the replicates of each operator using the mean
meanrep<-rep(seq(1,length(tidy_data$operator)/2), each = 2)
tidy_data$meanrep<-as.factor(meanrep)
library(dplyr)

tidy2 <- tidy_data %>%
  group_by(meanrep, operator, part) %>%
  summarize(mean_value = mean(value))

head<-head(tidy2,15)
kable(head,caption="Summarized data with mean of replicates")
```

Table 2: Summarized data with mean of replicates

meanrep	operator	part	mean_value
1	1	1	0.9525
2	2	1	0.9540
3	3	1	0.9550
4	1	2	0.9560
5	2	2	0.9565
6	3	2	0.9575
7	1	3	0.9555
8	2	3	0.9555
9	3	3	0.9565
10	1	4	0.9570
11	2	4	0.9575
12	3	4	0.9575
13	1	5	0.9570
14	2	5	0.9570
15	3	5	0.9580

```r
library(ggplot2)
tidy2 %>%
  ggplot(aes(x=part, y=tidy2$mean_value-(min(tidy2$mean_value))+0.001,
             fill =part)) +geom_col()+
  facet_wrap(~ operator)+
  labs(caption = "We have a histogram of the parts
      and their mean wall thickness. Each facet
```

```
      represents the operator used ",
      y = "mean wall thickness(scale = y-min(y)+0.001)")
```



We have a histogram of the parts
and their mean wall thickness. Each facet
represents the operator used

Issues

The data had to scaled for the best visualization because it all the measurements were very close.

Part B

Brain weight (g) and body weight (kg) for 62 species.

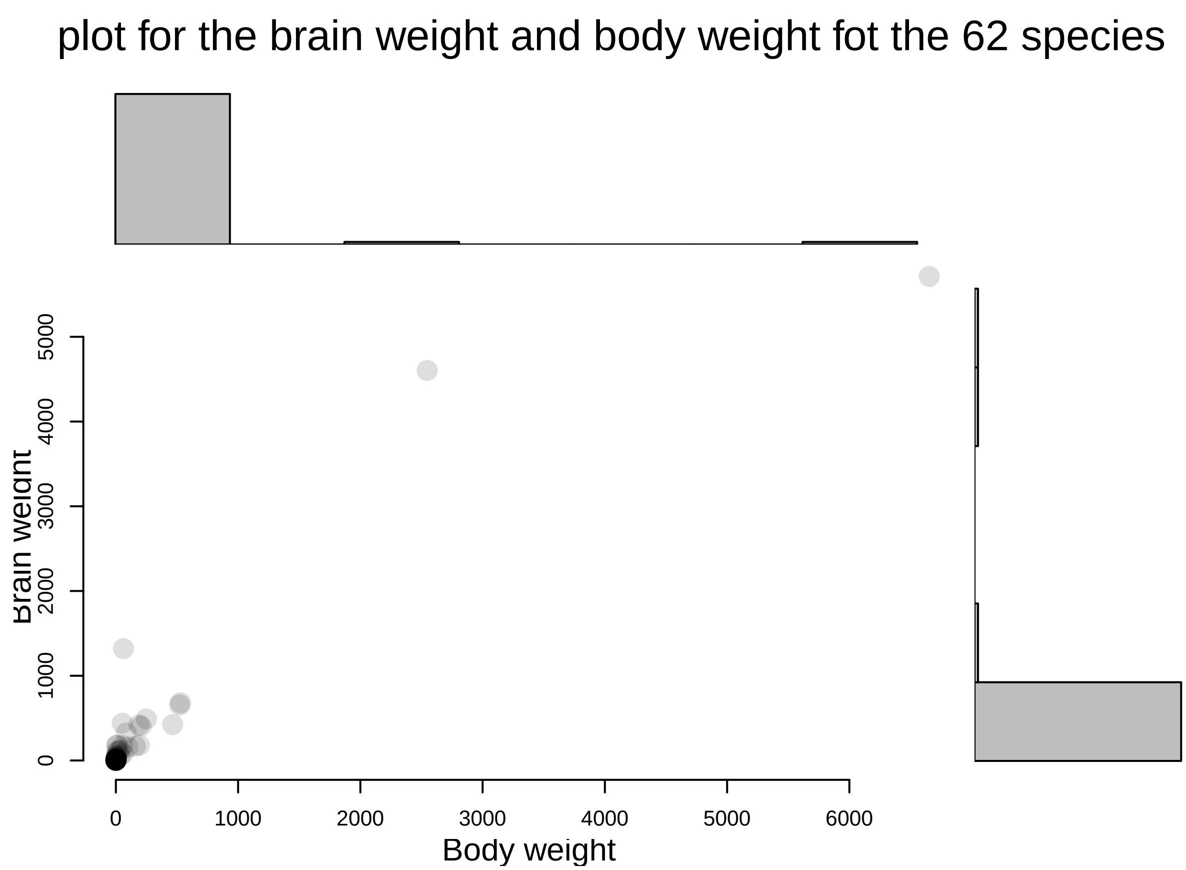
```
mat<-as.matrix(read.csv("partB_data.csv", header = F ))
colnames(mat) <- NULL
vec<-numeric()
for(i in 1:nrow(mat)){
  vec <- c(vec, mat[i, ])
}
vec<-vec[complete.cases(vec)]
type<-c("Body_Wt","Brain_Wt")
M.type<-rep(type, times = length(vec)/2)
rep<-seq(1,length(vec)/2, by =1)
Species<-rep(rep,each =2)
```

```
data<-data.frame(vec,M.type,Species)
tidy_data<-spread(data, M.type, vec)
tidy_data<-tidy_data %>%
  mutate(Species=as.factor(Species))
head<-head(tidy_data,15)
kable(head,caption="Tidy data for Brain weight (g)
      and body weight (kg) for 62 species")
```

Table 3: Tidy data for Brain weight (g) and body weight (kg) for 62 species

Species	Body_Wt	Brain_Wt
1	3.385	44.5
2	521.000	655.0
3	2.500	12.1
4	0.480	15.5
5	0.785	3.5
6	55.500	175.0
7	1.350	8.1
8	10.000	115.0
9	100.000	157.0
10	465.000	423.0
11	3.300	25.6
12	52.160	440.0
13	36.330	119.5
14	0.200	5.0
15	10.550	179.5

```
scatterhist(x, y, xlab="Body weight",
            ylab="Brain weight",
  "plot for the brain weight and body weight fot the 62 species",
  xsize =2)
```



Issues

The outliers in the data set interfered with the clarity of our scatter plot especially for our lower values.

Part C

Gold Medal performance for Olympic Men's Long Jump, year is coded as 1900=0. Goodness, ragged arrays. Check out fread in the data.table package.

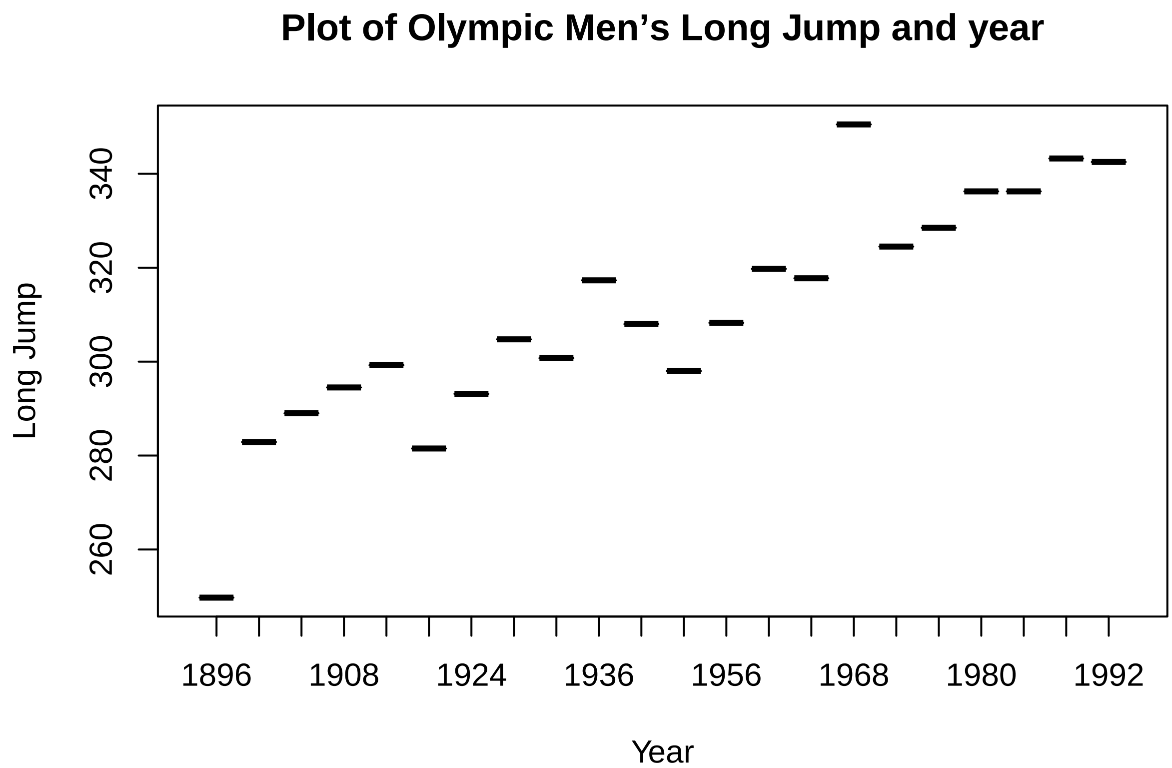
```
mat<-as.matrix(read.csv("partC_data.csv", header = F ))
colnames(mat) <- NULL
vec<-numeric()
for(i in 1:nrow(mat)){
  vec <- c(vec, mat[i, ])
}
vec<-vec[complete.cases(vec)]
dat<-data.frame(matrix(vec, ncol = 2, byrow = T))
dat$X1<-dat$X1+1900
colnames(dat)<-c("year", "Jump")
dat<-dat[order(dat$year), ]
tidy_data<-dat
rownames(tidy_data) <- NULL
tidy_data<-tidy_data %>%
```

```r
  mutate(year=as.factor(year))
head<-head(tidy_data,15)
kable(head,caption="Tidy data  of  performance for
      Olympic Men’s Long Jump and year")
```

Table 4: Tidy data of performance for Olympic Men’s Long Jump and year

year	Jump
1896	249.75
1900	282.88
1904	289.00
1908	294.50
1912	299.25
1920	281.50
1924	293.13
1928	304.75
1932	300.75
1936	317.31
1948	308.00
1952	298.00
1956	308.25
1960	319.75
1964	317.75

```r
plot(x=tidy_data$year, y=tidy_data$Jump,
     main = "Plot of Olympic Men’s Long Jump and year",
     ylab="Long Jump", pch = 19, xlab ="Year")
```



Issues

Data description was not very clear on the unit of measurement for the long jump.

Part D

Triplicate measurements of tomato yield for two varieties of tomatoes at three planting densities.

```
mat<-as.matrix(read.csv("partD_data.csv", header = F ))
colnames(mat) <- NULL
vec<-numeric()
for(i in 1:nrow(mat)){
  vec <- c(vec, mat[i, ])
}
yield<-vec
Variety<-c("Ife","PusaED")
Variety<-rep(Variety, each = length(vec)/2)
densities<-rep(c("10000","20000","30000"),each = 3)
densities<-rep(densities, times = 2)
tidy_data<-data.frame(yield,Variety,densities)
head<-head(tidy_data,15)
kable(head,caption="Tidy data for the triplicate measurements
      of tomato yield for two varieties of tomatoes
      at three planting densities.")
```

Table 5: Tidy data for the triplicate measurements of tomato yield for two varieties of tomatoes at three planting densities.

yield	Variety	densities
16.1	Ife	10000
15.3	Ife	10000
17.5	Ife	10000
16.6	Ife	20000
19.2	Ife	20000
18.5	Ife	20000
20.8	Ife	30000
18.0	Ife	30000
21.0	Ife	30000
8.1	PusaED	10000
8.6	PusaED	10000
10.1	PusaED	10000
12.7	PusaED	20000
13.7	PusaED	20000
11.5	PusaED	20000

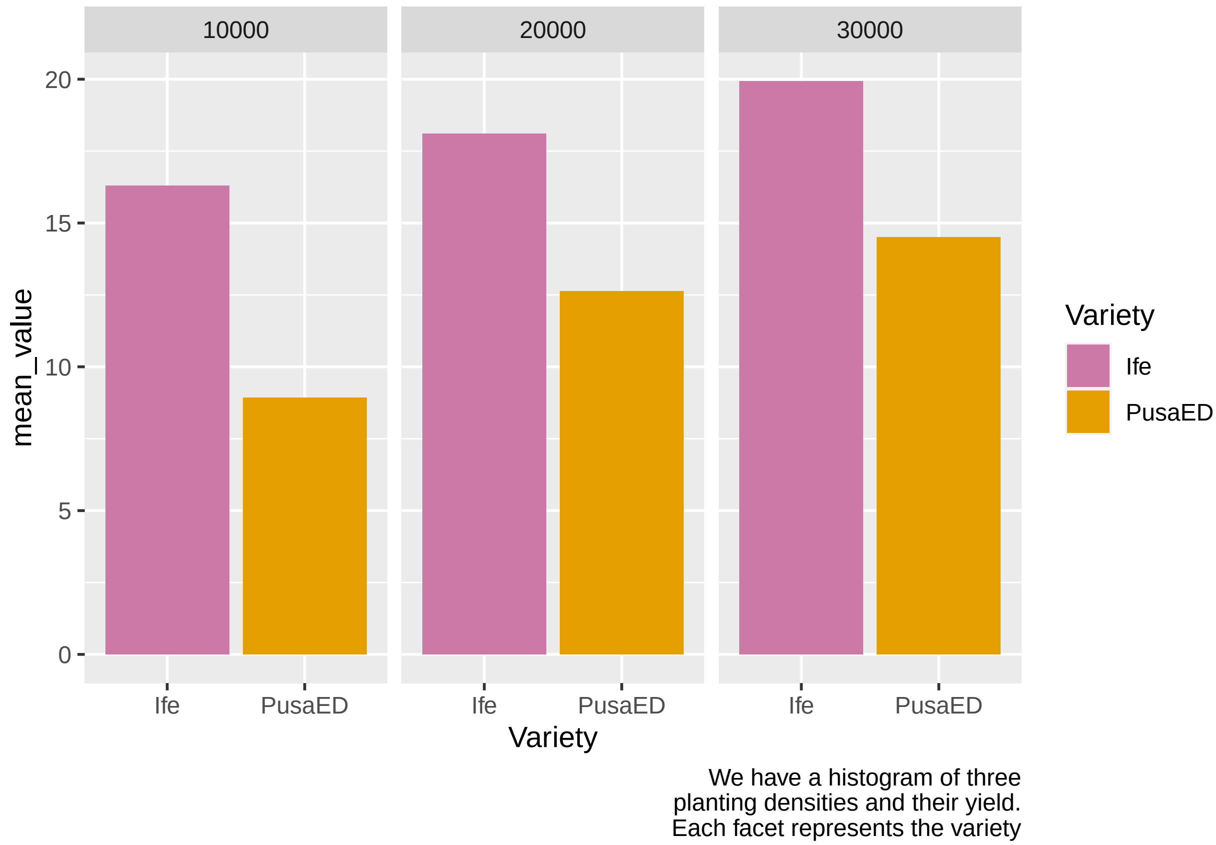
```r
# Mean of triplicate measurements
meanrep<-rep(seq(1,length(tidy_data$densities)/3), each = 3)
tidy_data$meanrep<-as.factor(meanrep)
tidy2 <- tidy_data %>%
  group_by(meanrep, densities, Variety) %>%
  summarize(mean_value = mean(yield))

head<-head(tidy2,15)
kable(head,caption="summarized tidy data with mean of
      triplicates of each density")
```

Table 6: summarized tidy data with mean of triplicates of each density

meanrep	densities	Variety	mean_value
1	10000	Ife	16.300000
2	20000	Ife	18.100000
3	30000	Ife	19.933333
4	10000	PusaED	8.933333
5	20000	PusaED	12.633333
6	30000	PusaED	14.500000

```r
tidy2 %>%
  ggplot(aes(x=Variety, y=mean_value)) +
  geom_col(aes(fill=Variety)) +
  facet_wrap(~ densities)+
  labs(caption = "We have a histogram of three
  planting densities and their yield.
```

```
    Each facet represents the variety")+
  scale_fill_manual(values=cbPalette)
```



Issues

The data did not have any issues.

Part E

Larvae counts at two ages given 5 different treatments in 8 blocks.

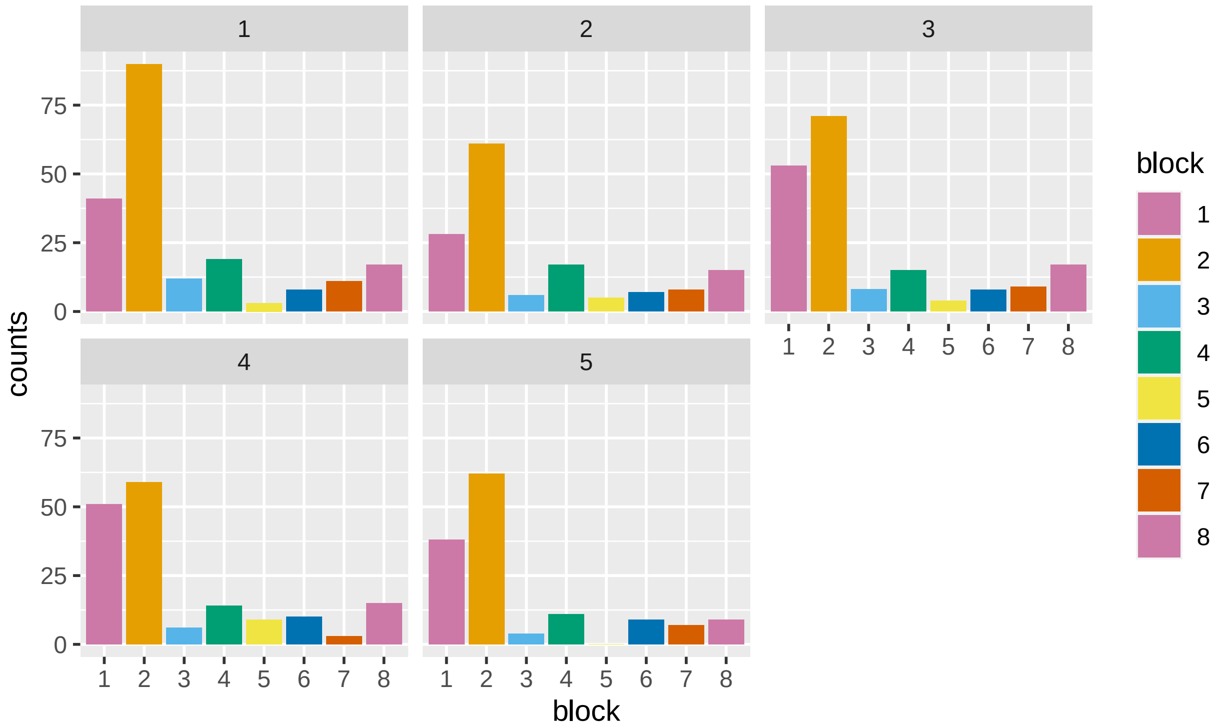
```
mat<-as.matrix(read.csv("partE_data.csv", header = F ))
colnames(mat) <- NULL
mat<-mat[,2:11]
vec<-numeric()
for(i in 1:nrow(mat)){
  vec <- c(vec, mat[i, ])
}
counts<-vec
b<-seq(1,8)
block<-rep(b, each = 10)
A<-rep(c(1,2), each = 5)
```

```r
Age<-rep(A, times = 8)
t<-seq(1,5)
treatment<-rep(t, times = 8)
dat<-data.frame(counts, block, treatment, Age)
tidy_data<-dat %>%
  mutate(block=as.factor(block))%>%
  mutate(treatment=as.factor(treatment))
head<-head(tidy_data,15)
kable(head,caption="Tidy data for our larvae counts
      at two ages given 5 different treatments in 8 blocks")
```

Table 7: Tidy data for our larvae counts at two ages given 5 different treatments in 8 blocks

counts	block	treatment	Age
13	1	1	1
16	1	2	1
13	1	3	1
20	1	4	1
16	1	5	1
28	1	1	2
12	1	2	2
40	1	3	2
31	1	4	2
22	1	5	2
29	2	1	1
12	2	2	1
23	2	3	1
15	2	4	1
17	2	5	1

```r
# To use for fills, add
library(ggplot2)
tidy_data %>%
  ggplot(aes(x=block, y=counts)) +
  geom_col(aes(fill=block)) +
  facet_wrap(~ treatment)+
  scale_fill_manual(values=cbPalette)+
  labs(caption = "We have a histogram of blocks and
       their larvae counts. Each facet represents
       the Treatments. Treatment 1 to 5 for Age 1
       and the rest for age 2")
```



We have a histogram of blocks and their larvae counts. Each facet represents the Treatments. Treatment 1 to 5 for Age 1 and the rest for age 2

Issues

No Issues with data.