

Data wrangling

AVOCADO

Read and format data

Block and randomization experimental design. Three blocks and 12 subjects each

```
da <- fread('avocado_blackness.csv')
dim(da)
```

```
## [1] 50 37
```

```
head(da)
```

```
##   hue_index  Bc1  Bc6  Bc8  Bc9  Bc11  Bc12  Bt2  Bt3  Bt4  Bt5  Bt7
## 1:         0 23.71 45.27 43.63 41.31 37.85 52.14 36.47 40.1 49.59 35.40 44.60
## 2:         1 23.71 45.27 43.63 41.31 37.87 52.43 36.56 40.1 49.71 35.52 44.87
## 3:         2 23.71 45.27 43.63 41.31 37.87 52.43 36.56 40.1 49.71 35.52 44.87
## 4:         3 23.71 45.27 43.63 41.31 37.87 52.43 36.56 40.1 49.71 35.52 44.87
## 5:         4 23.71 45.27 43.63 41.31 37.87 52.43 36.56 40.1 49.71 35.52 44.87
## 6:         5 23.71 45.27 43.63 41.31 37.87 52.43 36.56 40.1 49.71 35.52 44.87
##   Bt10  Nt1  Nc2  Nt3  Nc4  Nc5  Nt6  Nc7  Nt8  Nc9  Nt10  Nc11  Nt12
## 1: 45.68 28.10 30 39.39 28.56 28.83 30.35 29.79 29.93 32.58 23.50 30.05 25.81
## 2: 45.72 28.11 30 39.39 28.56 28.83 30.36 29.79 29.94 32.58 23.51 30.09 25.81
## 3: 45.72 28.11 30 39.39 28.56 28.83 30.36 29.79 29.94 32.58 23.51 30.09 25.81
## 4: 45.72 28.11 30 39.39 28.56 28.83 30.36 29.79 29.94 32.58 23.51 30.09 25.81
## 5: 45.72 28.11 30 39.39 28.56 28.83 30.36 29.79 29.94 32.58 23.51 30.09 25.81
## 6: 45.72 28.11 30 39.39 28.56 28.83 30.36 29.79 29.94 32.58 23.51 30.09 25.81
##   Jt1  Jt4  Jt7  Jt8  Jt9  Jt11  Jc2  Jc3  Jc5  Jc6  Jc10  Jc12
## 1: 23.37 25.69 24.22 26.37 26.62 21.58 29.4 27.77 28.5 25.84 28.69 29.58
## 2: 23.38 25.69 24.22 26.38 26.62 21.59 29.4 27.77 28.5 25.84 28.69 29.58
## 3: 23.38 25.69 24.22 26.38 26.62 21.59 29.4 27.77 28.5 25.84 28.69 29.58
## 4: 23.38 25.69 24.22 26.38 26.62 21.59 29.4 27.77 28.5 25.84 28.69 29.58
## 5: 23.38 25.69 24.22 26.38 26.62 21.59 29.4 27.77 28.5 25.84 28.70 29.58
## 6: 23.38 25.70 24.22 26.38 26.62 21.59 29.4 27.77 28.5 25.84 28.70 29.58
```

```
tail(da)
```

```
##   hue_index  Bc1  Bc6  Bc8  Bc9  Bc11  Bc12  Bt2  Bt3  Bt4  Bt5  Bt7  Bt10
## 1:         44 100 100 100 100 99.97 99.61 99.76 100 99.65 99.81 99.78 99.91
## 2:         45 100 100 100 100 99.97 99.61 99.76 100 99.65 99.81 99.78 99.91
## 3:         46 100 100 100 100 99.97 99.61 99.76 100 99.65 99.81 99.78 99.91
## 4:         47 100 100 100 100 99.97 99.61 99.76 100 99.65 99.81 99.78 99.91
## 5:         48 100 100 100 100 99.97 99.61 99.76 100 99.65 99.81 99.78 99.91
```

```
## 6:      49 100 100 100 100 99.97 99.61 99.76 100 99.65 99.81 99.78 99.91
##      Nt1 Nc2      Nt3      Nc4 Nc5      Nt6 Nc7      Nt8 Nc9 Nt10 Nc11 Nt12 Jt1 Jt4
## 1: 99.96 100 99.99 99.99 100 99.98 100 99.92 100 100 99.99 99.99 99.97 99.99
## 2: 99.96 100 100.00 99.99 100 99.98 100 99.93 100 100 99.99 99.99 99.97 99.99
## 3: 99.97 100 100.00 99.99 100 99.99 100 99.94 100 100 99.99 99.99 99.97 99.99
## 4: 99.97 100 100.00 99.99 100 99.99 100 99.95 100 100 99.99 99.99 99.97 99.99
## 5: 99.97 100 100.00 99.99 100 99.99 100 99.95 100 100 99.99 99.99 99.97 99.99
## 6: 99.97 100 100.00 99.99 100 99.99 100 99.95 100 100 99.99 99.99 99.97 99.99
##      Jt7      Jt8      Jt9 Jt11      Jc2      Jc3 Jc5 Jc6      Jc10 Jc12
## 1: 99.99 99.99 99.98 99.98 99.99 99.99 100 100 99.99 100
## 2: 99.99 99.99 99.99 99.98 100.00 99.99 100 100 99.99 100
## 3: 99.99 100.00 99.99 99.99 100.00 100.00 100 100 99.99 100
## 4: 100.00 100.00 99.99 99.99 100.00 100.00 100 100 100.00 100
## 5: 100.00 100.00 99.99 99.99 100.00 100.00 100 100 100.00 100
## 6: 100.00 100.00 99.99 99.99 100.00 100.00 100 100 100.00 100
```

```
colnames(da)
```

```
## [1] "hue_index" "Bc1"      "Bc6"      "Bc8"      "Bc9"      "Bc11"
## [7] "Bc12"      "Bt2"      "Bt3"      "Bt4"      "Bt5"      "Bt7"
## [13] "Bt10"      "Nt1"      "Nc2"      "Nt3"      "Nc4"      "Nc5"
## [19] "Nt6"      "Nc7"      "Nt8"      "Nc9"      "Nt10"     "Nc11"
## [25] "Nt12"      "Jt1"      "Jt4"      "Jt7"      "Jt8"      "Jt9"
## [31] "Jt11"      "Jc2"      "Jc3"      "Jc5"      "Jc6"      "Jc10"
## [37] "Jc12"
```

```
#convert wide to long extract data attribute
#and create new columns
df <- (melt(da, id.vars=c("hue_index")))
df$block <- substring(df$variable,0,1)
df$treatment <- substring(df$variable,2,2)
df$id <- substring(df$variable,3,4)
head(df)
```

```
## hue_index variable value block treatment id
## 1      0      Bc1 23.71      B      c 1
## 2      1      Bc1 23.71      B      c 1
## 3      2      Bc1 23.71      B      c 1
## 4      3      Bc1 23.71      B      c 1
## 5      4      Bc1 23.71      B      c 1
## 6      5      Bc1 23.71      B      c 1
```

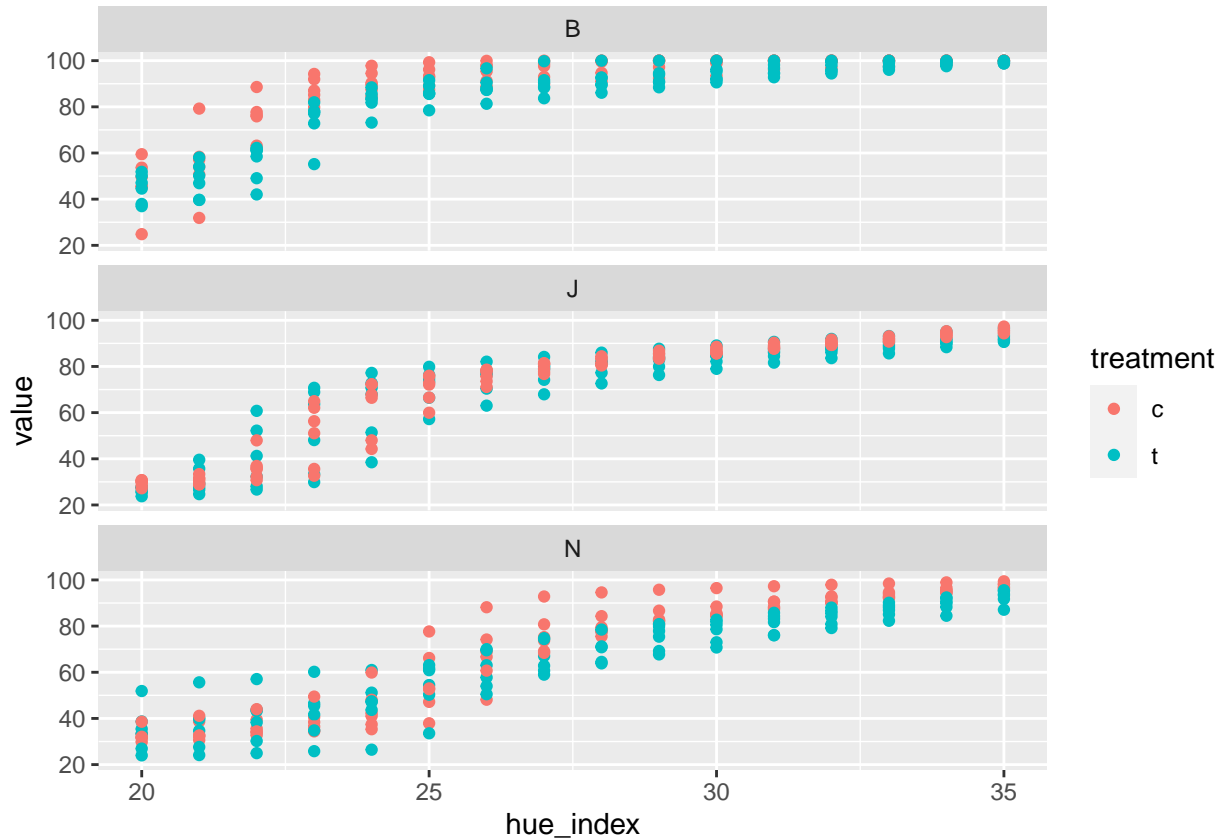
```
d_avo_raw <- df[,c("block", "treatment", "id", "hue_index", "value")]
dim(d_avo_raw)
```

```
## [1] 1800      5
```

The following figure shows both fixed effect for block and individual level Y axis shows percent
[TODO]

```
d_avo_raw %>% ggplot(aes(x = hue_index, y = value, color=treatment)) +
  geom_point(aes(color = treatment)) + facet_wrap(~block,ncol = 1) + xlim(20,35)
```

```
## Warning: Removed 1224 rows containing missing values (geom_point).
```



Get individual hue count data

TODO The higher the value of hue, the stronger the filter effect. For an example, For Bill, the background was black For Nobu and Justin, it was white. So, we also need to evaluate the incremental effect

```
## Get the length of data, hue_range
len <- dim(da)[2]

## empty matrix that will get the frequency data
avo_frequency <- as.matrix(0:49)

for (i in colnames(da)){
  if (i == "hue_index"){
  }
  else{
    ##Do something
    temp <- getIncre(as.matrix(da[[i]]))
  }
}
```

```

temp <- as.matrix(temp)
##get the increment
avo_frequency <- cbind(avo_frequency ,temp)
##start adding them
}
}

#now add column names
d_avo <- data.frame(avo_frequency)
colnames(d_avo) <- colnames(da)
head(d_avo)

```

```

##   hue_index Bc1 Bc6 Bc8 Bc9 Bc11 Bc12 Bt2 Bt3 Bt4 Bt5 Bt7 Bt10 Nt1 Nc2 Nt3
## 1         0  0  0  0  0 0.00 0.00 0.00  0 0.00 0.00 0.00 0.00 0.00  0  0
## 2         1  0  0  0  0 0.02 0.29 0.09  0 0.12 0.12 0.27 0.04 0.01  0  0
## 3         2  0  0  0  0 0.00 0.00 0.00  0 0.00 0.00 0.00 0.00 0.00  0  0
## 4         3  0  0  0  0 0.00 0.00 0.00  0 0.00 0.00 0.00 0.00 0.00  0  0
## 5         4  0  0  0  0 0.00 0.00 0.00  0 0.00 0.00 0.00 0.00 0.00  0  0
## 6         5  0  0  0  0 0.00 0.00 0.00  0 0.00 0.00 0.00 0.00 0.00  0  0
##   Nc4 Nc5 Nt6 Nc7 Nt8 Nc9 Nt10 Nc11 Nt12 Jt1 Jt4 Jt7 Jt8 Jt9 Jt11 Jc2 Jc3
## 1  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00  0 0.00  0  0
## 2  0  0 0.01  0 0.01  0 0.01 0.04  0 0.01 0.00  0 0.01  0 0.01  0  0
## 3  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00  0 0.00  0  0
## 4  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00  0 0.00  0  0
## 5  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00  0 0.00  0  0
## 6  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.01  0 0.00  0 0.00  0  0
##   Jc5 Jc6 Jc10 Jc12
## 1  0  0 0.00  0
## 2  0  0 0.00  0
## 3  0  0 0.00  0
## 4  0  0 0.00  0
## 5  0  0 0.01  0
## 6  0  0 0.00  0

```

- Now, d_avo_frequency contains frequency of pixels whose color changed when hue was incremented by 1

```

#convert wide to long extract data attribute
#and create new columns
dff <- (melt(d_avo, id.vars=c("hue_index")))
dff$block<- substring(dff$variable,0,1)
dff$treatment <- substring(dff$variable,2,2)
dff$id <- substring(dff$variable,3,4)
d_avo_frequency <- dff[,c("block","treatment","id","hue_index","value")]

```

TODO: NEED TO CONVERT TO DENSITY PLOT

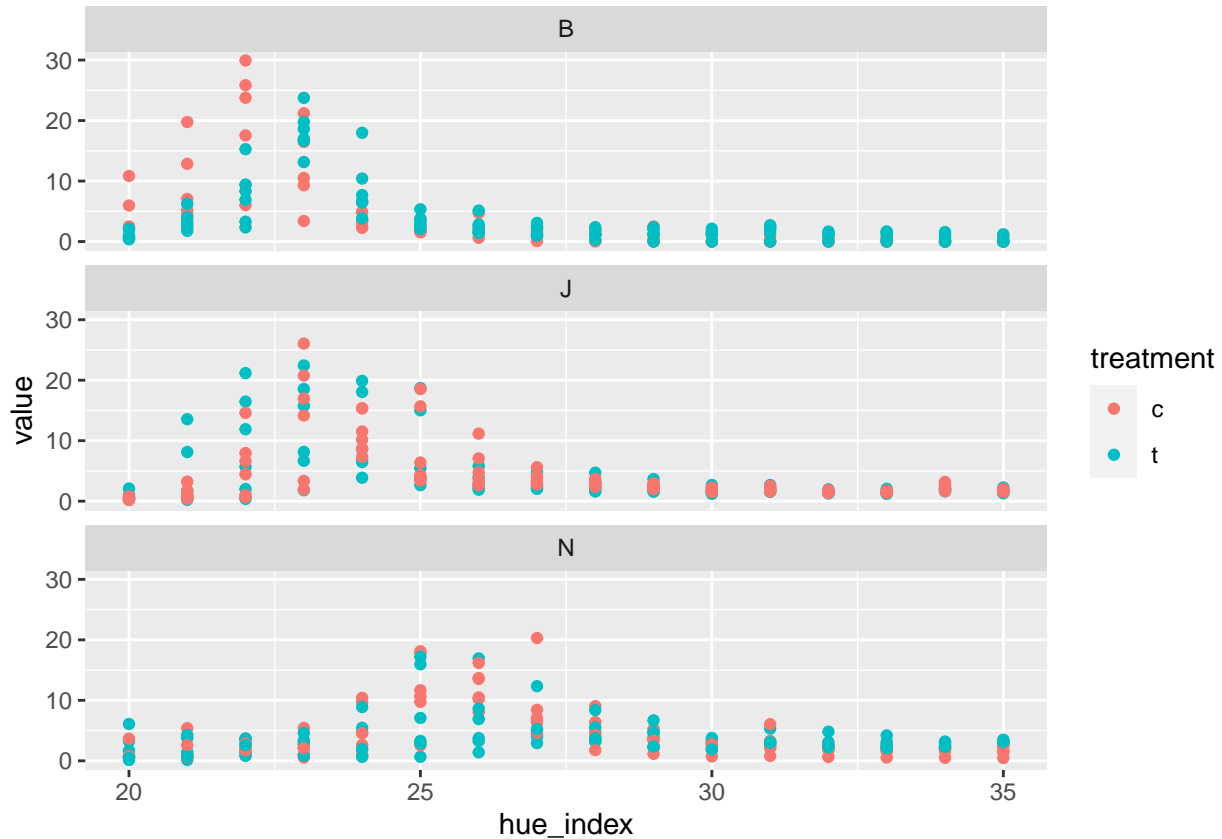
Color wise, N > J and B. [TODO] 20 HUE indicate percent of avocado whose that became black when hue was changed from 19 to 20. (really bad)

25 HUE indicate percent of avocado that turn black when hue was increased from 24 to 25. (still good)

30 HUE indicate percent of avocado that turn black when hue changed from 29 to 30.

```
d_avo_frequency %>% ggplot(aes(x = hue_index, y = value, color=treatment)) +  
  geom_point(aes(color = treatment)) + facet_wrap(~block,ncol = 1) + xlim(20,35)
```

```
## Warning: Removed 1224 rows containing missing values (geom_point).
```



Is color good indicator?

Create avocado pdf

```
#now df_avo cotains percent of pixles whose color changed when  
head(d_avo)
```

```
##   hue_index Bc1 Bc6 Bc8 Bc9 Bc11 Bc12 Bt2 Bt3 Bt4 Bt5 Bt7 Bt10 Nt1 Nc2 Nt3  
## 1         0   0   0   0   0 0.00 0.00 0.00   0 0.00 0.00 0.00 0.00 0.00   0   0  
## 2         1   0   0   0   0 0.02 0.29 0.09   0 0.12 0.12 0.27 0.04 0.01   0   0  
## 3         2   0   0   0   0 0.00 0.00 0.00   0 0.00 0.00 0.00 0.00 0.00   0   0  
## 4         3   0   0   0   0 0.00 0.00 0.00   0 0.00 0.00 0.00 0.00 0.00   0   0  
## 5         4   0   0   0   0 0.00 0.00 0.00   0 0.00 0.00 0.00 0.00 0.00   0   0  
## 6         5   0   0   0   0 0.00 0.00 0.00   0 0.00 0.00 0.00 0.00 0.00   0   0  
##   Nc4 Nc5  Nt6 Nc7  Nt8 Nc9 Nt10 Nc11 Nt12  Jt1  Jt4 Jt7  Jt8 Jt9 Jt11 Jc2 Jc3
```

```
## 1  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00 0  0
## 2  0  0 0.01  0 0.01  0 0.01 0.04  0 0.01 0.00  0 0.01 0  0
## 3  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00 0  0
## 4  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00 0  0
## 5  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.00  0 0.00 0  0
## 6  0  0 0.00  0 0.00  0 0.00 0.00  0 0.00 0.01  0 0.00 0  0
##   Jc5 Jc6 Jc10 Jc12
## 1  0  0 0.00  0
## 2  0  0 0.00  0
## 3  0  0 0.00  0
## 4  0  0 0.00  0
## 5  0  0 0.01  0
## 6  0  0 0.00  0
```

empty matrix that will get the frequency data

```
avo_pdf <- as.matrix(0:49)
```

```
for (i in colnames(d_avo)){
  if (i == "hue_index"){
  }
  else{
    ##Do something
    temp <- sum(d_avo[[i]])
    temp <- d_avo[[i]]/temp
    ##start adding them
    avo_pdf <- cbind(avo_pdf ,temp)
  }
}
```

##now add column names

```
d_avo_pdf <- data.frame(avo_pdf)
colnames(d_avo_pdf) <- colnames(da)
head(d_avo_pdf)
```

```
##   hue_index Bc1 Bc6 Bc8 Bc9          Bc11          Bc12          Bt2 Bt3
## 1         0  0  0  0  0 0.0000000000 0.0000000000 0.0000000000  0
## 2         1  0  0  0  0 0.0003219575 0.006109122 0.001422026  0
## 3         2  0  0  0  0 0.0000000000 0.0000000000 0.0000000000  0
## 4         3  0  0  0  0 0.0000000000 0.0000000000 0.0000000000  0
## 5         4  0  0  0  0 0.0000000000 0.0000000000 0.0000000000  0
## 6         5  0  0  0  0 0.0000000000 0.0000000000 0.0000000000  0
##          Bt4          Bt5          Bt7          Bt10          Nt1 Nc2 Nt3 Nc4 Nc5
## 1 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  0  0  0  0
## 2 0.002397123 0.001863065 0.004893077 0.0007375991 0.0001391401  0  0  0  0
## 3 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  0  0  0  0
## 4 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  0  0  0  0
## 5 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  0  0  0  0
## 6 0.0000000000 0.0000000000 0.0000000000 0.0000000000 0.0000000000  0  0  0  0
##          Nt6 Nc7          Nt8 Nc9          Nt10          Nc11 Nt12          Jt1
## 1 0.0000000000  0 0.0000000000  0 0.0000000000 0.0000000000  0 0.0000000000
## 2 0.0001435956  0 0.0001428163  0 0.000130719 0.0005719188  0 0.0001305483
## 3 0.0000000000  0 0.0000000000  0 0.0000000000 0.0000000000  0 0.0000000000
## 4 0.0000000000  0 0.0000000000  0 0.0000000000 0.0000000000  0 0.0000000000
## 5 0.0000000000  0 0.0000000000  0 0.0000000000 0.0000000000  0 0.0000000000
```

```
## 6 0.0000000000 0 0.0000000000 0 0.0000000000 0.0000000000 0 0.0000000000
##      Jt4 Jt7      Jt8 Jt9      Jt11 Jc2 Jc3 Jc5 Jc6      Jc10
## 1 0.0000000000 0 0.0000000000 0 0.0000000000 0 0 0 0 0.0000000000
## 2 0.0000000000 0 0.0001358142 0 0.0001275348 0 0 0 0 0.0000000000
## 3 0.0000000000 0 0.0000000000 0 0.0000000000 0 0 0 0 0.0000000000
## 4 0.0000000000 0 0.0000000000 0 0.0000000000 0 0 0 0 0.0000000000
## 5 0.0000000000 0 0.0000000000 0 0.0000000000 0 0 0 0 0.0001402328
## 6 0.0001345895 0 0.0000000000 0 0.0000000000 0 0 0 0 0.0000000000
##      Jc12
## 1      0
## 2      0
## 3      0
## 4      0
## 5      0
## 6      0
```

```
#convert wide to long extract data attribute
#and create new columns
dff <- (melt(d_avo_pdf, id.vars=c("hue_index")))
dff$block<- substring(dff$variable,0,1)
dff$treatment <- substring(dff$variable,2,2)
dff$id <- substring(dff$variable,3,4)
d_avo_pdf_long <- dff[,c("block", "treatment", "id", "hue_index", "value")]
head(d_avo_pdf_long)
```

```
##   block treatment id hue_index value
## 1     B          c 1          0      0
## 2     B          c 1          1      0
## 3     B          c 1          2      0
## 4     B          c 1          3      0
## 5     B          c 1          4      0
## 6     B          c 1          5      0
```

Create sample data based on the pdf

TODO Need to create function

```
BT <- d_avo_pdf_long %>% filter(block=="B" & treatment == "t") %>%
  group_by(hue_index) %>% dplyr::summarize(Mean = mean(value, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
BC <- d_avo_pdf_long %>% filter(block=="B" & treatment == "c") %>%
  group_by(hue_index) %>% dplyr::summarize(Mean = mean(value, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
NT <- d_avo_pdf_long %>% filter(block=="N" & treatment == "t") %>%
  group_by(hue_index) %>% dplyr::summarize(Mean = mean(value, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
NC <- d_avo_pdf_long %>% filter(block=="N" & treatment=="c")%>%  
  group_by(hue_index) %>% dplyr::summarize(Mean = mean(value, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
JT <- d_avo_pdf_long %>% filter(block=="J" & treatment=="t")%>%  
  group_by(hue_index) %>% dplyr::summarize(Mean = mean(value, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
JC <- d_avo_pdf_long %>% filter(block=="J" & treatment=="c")%>%  
  group_by(hue_index) %>% dplyr::summarize(Mean = mean(value, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
#treatment  
s1 <- get_ind_data(BT)  
t1 <- data.frame(block="B", control = "treatment", value = s1)  
  
s2 <- get_ind_data(JT)  
t2 <- data.frame(block="J", control = "treatment", value = s2)  
  
s3 <- get_ind_data(NT)  
t3 <- data.frame(block="N", control = "treatment", value = s3)  
  
three_treats <- rbind(t1,t2,t3)  
  
#control  
s4 <- get_ind_data(BC)  
t4 <- data.frame(block="B", control = "control", value = s4)  
  
s5 <- get_ind_data(JC)  
t5 <- data.frame(block="J", control = "control", value = s5)  
  
s6 <- get_ind_data(NC)  
t6 <- data.frame(block="N", control = "control", value = s6)  
three_control <- rbind(t4,t5,t6)  
  
data <- rbind(three_treats,three_control)  
  
p1 <- data%>% ggplot(.,aes(x=value)) +  
  geom_density(aes(fill=control),adjust=1.5,alpha=0.3) +  
  facet_wrap(~block, ncol = 1) +  
  xlim(20, 45) +  
  theme(  
    legend.position="top",  
    panel.spacing = unit(0.1, "lines"),  
    axis.ticks.x=element_blank(),  
    plot.title = element_text(hjust = 0.5)
```



```

    ) +
    ggtitle("emprical pdf") +
    xlab("Hue") + ylab("Probability")

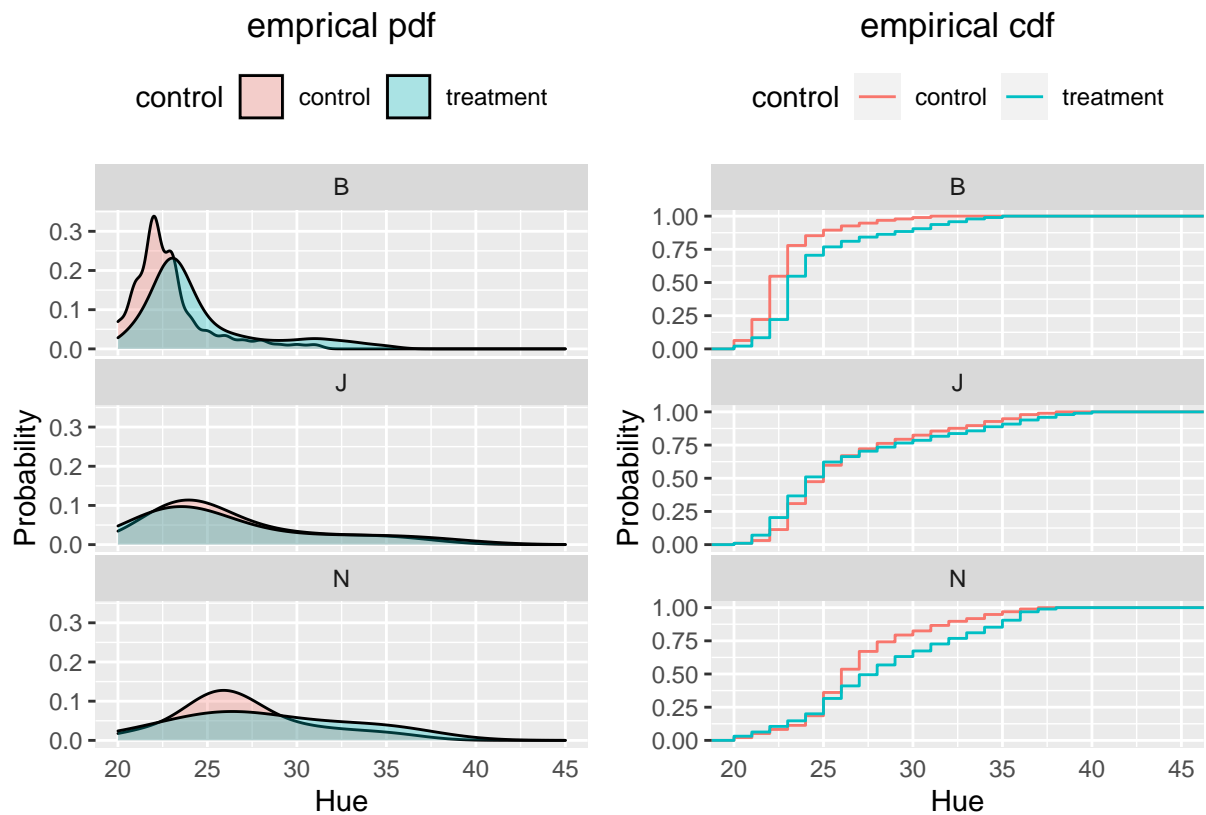
p2 <- data %>% ggplot(.,aes(x=value, colour = control)) + stat_ecdf() +
  facet_wrap(~block, ncol = 1) +
  xlim(20, 45) +
  theme(
    legend.position="top",
    panel.spacing = unit(0.1, "lines"),
    axis.ticks.x=element_blank(),
    plot.title = element_text(hjust = 0.5)
  ) +
  ggtitle("empirical cdf") +
  xlab("Hue") + ylab("Probability")

p1 | p2

```

Warning: Removed 10 rows containing non-finite values (stat_density).

Warning: Removed 10 rows containing non-finite values (stat_ecdf).



```
#control <- getIncr(df1)
#treatment <- getIncr(df2)
control <- BT$Mean
treatment <- BC$Mean

#sharp null distribution
par(mfrow=c(3,1))
get_ks_permutation(BT$Mean,BC$Mean,5000)
```

10

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## [1] 0.5118
```

```
get_ks_permutation(JT$Mean, JC$Mean, 5000)
```

```
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## [1] 0.9624
```

```
get_ks_permutation(NT$Mean,NC$Mean,5000)
```

```
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
## Warning in ks.test(A, B): cannot compute exact p-value with ties
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

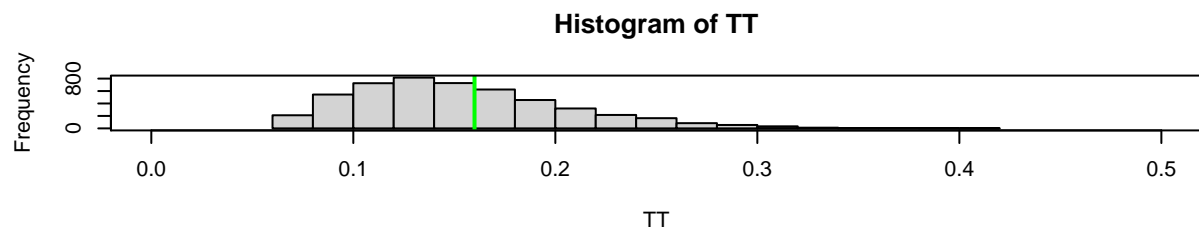
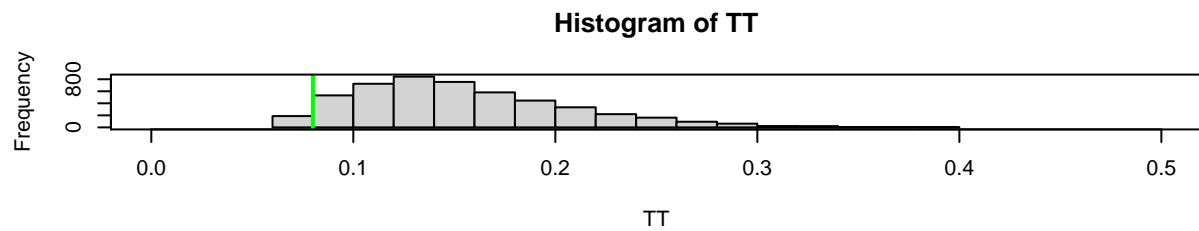
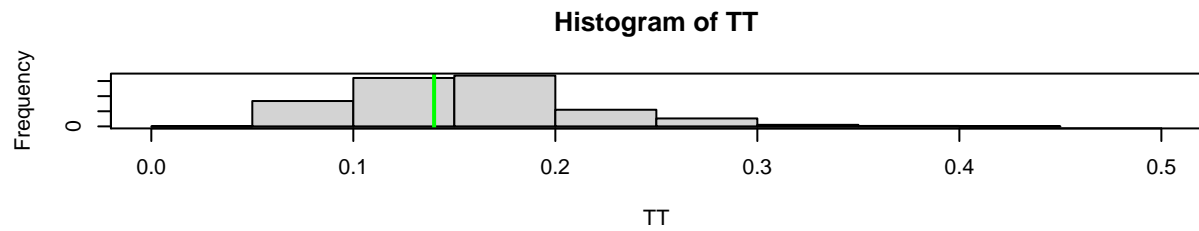
[illegible]

[illegible]

[illegible]

```
## Warning in ks.test(A, B): cannot compute exact p-value with ties
```

```
## Warning in ks.test(A, B): cannot compute exact p-value with ties
```



```
## [1] 0.395
```

```
par(mfrow=c(1,1))
Z <- c(control,treatment)
n <- length(control )
m <- length(treatment)
N <- length(Z)
```

Measure 2

```
#raw data converted to long format
head(d_avo_raw)
```

```
##   block treatment id hue_index value
## 1     B         c 1          0 23.71
## 2     B         c 1          1 23.71
## 3     B         c 1          2 23.71
## 4     B         c 1          3 23.71
## 5     B         c 1          4 23.71
## 6     B         c 1          5 23.71
```

```
#arrange the data by block, treatment, id and hue_index
d <- as.data.table(d_avo_raw)
d <- d %>% filter(hue_index > 18 & hue_index < 44)
db <- d[order(rank(block), treatment,id,hue_index)]
head(db)
```

```
##      block treatment id hue_index value
## 1:      B          c  1         19 24.21
## 2:      B          c  1         20 24.84
## 3:      B          c  1         21 31.88
## 4:      B          c  1         22 61.81
## 5:      B          c  1         23 83.02
## 6:      B          c  1         24 89.56
```

```
dim(db)/36
```

```
## [1] 25.0000000 0.1388889
```

Code for mannual confirmation

```
##      block treatment id hue_index value
## 1:      N          t  8         38 98.67
## 2:      N          t  8         39 99.35
## 3:      N          t  8         40 99.63
## 4:      N          t  8         41 99.77
## 5:      N          t  8         42 99.84
## 6:      N          t  8         43 99.89
```

```
## [1] 4
## [1] 23
## [1] 3
## [1] 22
## [1] 4
## [1] 23
## [1] 3
## [1] 22
## [1] 3
## [1] 22
## [1] 2
## [1] 21
## [1] 4
## [1] 23
## [1] 5
## [1] 24
## [1] 4
## [1] 23
## [1] 4
## [1] 23
## [1] 4
## [1] 23
## [1] 4
## [1] 23
```

```

## [1] 6
## [1] 25
## [1] 7
## [1] 26
## [1] 5
## [1] 24
## [1] 5
## [1] 24
## [1] 4
## [1] 23
## [1] 5
## [1] 24
## [1] 4
## [1] 23
## [1] 6
## [1] 25
## [1] 4
## [1] 23
## [1] 5
## [1] 24
## [1] 4
## [1] 23
## [1] 7
## [1] 26
## [1] 7
## [1] 26
## [1] 8
## [1] 27
## [1] 8
## [1] 27
## [1] 6
## [1] 25
## [1] 8
## [1] 27
## [1] 6
## [1] 25
## [1] 12
## [1] 31
## [1] 8
## [1] 27
## [1] 7
## [1] 26
## [1] 11
## [1] 30
## [1] 7
## [1] 26
## [1] 10
## [1] 29

##      block avocado_number treatment hue_turn
## 1      B              1         c      23
## 2      B             11         c      22
## 3      B             12         c      23
## 4      B              6         c      22

```


## 5	B	8	c	22
## 6	B	9	c	21
## 7	B	10	t	23
## 8	B	2	t	24
## 9	B	3	t	23
## 10	B	4	t	23
## 11	B	5	t	23
## 12	B	7	t	23
## 13	J	10	c	25
## 14	J	12	c	26
## 15	J	2	c	24
## 16	J	3	c	24
## 17	J	5	c	23
## 18	J	6	c	24
## 19	J	1	t	23
## 20	J	11	t	25
## 21	J	4	t	23
## 22	J	7	t	24
## 23	J	8	t	23
## 24	J	9	t	26
## 25	N	11	c	26
## 26	N	2	c	27
## 27	N	4	c	27
## 28	N	5	c	25
## 29	N	7	c	27
## 30	N	9	c	25
## 31	N	1	t	31
## 32	N	10	t	27
## 33	N	12	t	26
## 34	N	3	t	30
## 35	N	6	t	26
## 36	N	8	t	29

Abrupt change detection in HUE

TODO there is an error in the code here. Need to fix it

```
# #raw data converted to long format
# head(d_avo_raw)
#
# #arrange the data by block, treatement, id and hue_index
# d <- as.data.table(d_avo_raw)
# db <- d[order(rank(block), treatment, id,hue_index)]
# head(db)
#
# d <- as.data.table(d_avo_raw)
# d <- d %>% filter(hue_index > 18 & hue_index < 44)
# db <- d[order(rank(block), treatment,id,hue_index)]
#
# b_result <- data.frame(block = numeric(0),avocado_number= numeric(0),
#                         treatment = numeric(0),
#                         hue_turn = numeric(0) )
# #has 160 rows
```

```

# for (val in 1:36)
# {
#   #selected only upto min hue + 25
#   start <- 1 + (val-1)*24
#   end <- (25*val)
#   #get the subject information
#   val_block <- db[start,block]
#   val_subject <- db[start,id]
#   val_treat <- db[start,treatment]
#
#   # # print(paste(start,":",end))
#   d_temp <- db[start:end,]
#   #abrupt change detection point
#   dg <- d_temp$value
#   dg.amoc=cpt.mean(dg)
#   #19 is the minimum number of Hue when the black ratio started to change
#   hue_turn <- cpts(dg.amoc)
#   b_result[val,] <- c(val_block,val_subject,val_treat,hue_turn)
# }
#
# b_result

```

Regression Estimator for Block Randomization

We have 3 blocks, B, N, and J with 12 samples in each block.

b_result

##	block	avocado_number	treatment	hue_turn
## 1	B	1	c	23
## 2	B	11	c	22
## 3	B	12	c	23
## 4	B	6	c	22
## 5	B	8	c	22
## 6	B	9	c	21
## 7	B	10	t	23
## 8	B	2	t	24
## 9	B	3	t	23
## 10	B	4	t	23
## 11	B	5	t	23
## 12	B	7	t	23
## 13	J	10	c	25
## 14	J	12	c	26
## 15	J	2	c	24
## 16	J	3	c	24
## 17	J	5	c	23
## 18	J	6	c	24
## 19	J	1	t	23
## 20	J	11	t	25
## 21	J	4	t	23
## 22	J	7	t	24
## 23	J	8	t	23

```
## 24      J           9      t      26
## 25      N          11      c      26
## 26      N           2      c      27
## 27      N           4      c      27
## 28      N           5      c      25
## 29      N           7      c      27
## 30      N           9      c      25
## 31      N           1      t      31
## 32      N          10      t      27
## 33      N          12      t      26
## 34      N           3      t      30
## 35      N           6      t      26
## 36      N           8      t      29
```

```
mod_b1 <- lm(hue_turn ~ as.factor(treatment) + as.factor(block), data = b_result)
mod_b2 <- lm(hue_turn ~ as.factor(treatment)*as.factor(block), data = b_result)

mod_b2
```

```
##
## Call:
## lm(formula = hue_turn ~ as.factor(treatment) * as.factor(block),
##     data = b_result)
##
## Coefficients:
##                (Intercept)
##                   22.167
##      as.factor(treatment)t
##                   1.000
##      as.factor(block)J
##                   2.167
##      as.factor(block)N
##                   4.000
## as.factor(treatment)t:as.factor(block)J
##                   -1.333
## as.factor(treatment)t:as.factor(block)N
##                   1.000
```

```
coefficients(mod_b1)
```

```
##                (Intercept) as.factor(treatment)t      as.factor(block)J
##                22.2222222      0.8888889      1.5000000
##      as.factor(block)N
##                4.5000000
```

```
coeftest(mod_b1, vcov = vcovHC(mod_b1, type = "HC1"))
```

```
##
## t test of coefficients:
##
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      22.22222      0.29200 76.1042 < 2.2e-16 ***
```

```
## as.factor(treatment)t 0.88889 0.42853 2.0743 0.0461824 *
## as.factor(block)J 1.50000 0.41002 3.6584 0.0009047 ***
## as.factor(block)N 4.50000 0.52347 8.5966 7.994e-10 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(mod_b2, vcov = vcovHC(mod_b2, type = "HC1"))
```

```
##
## t test of coefficients:
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.16667 0.30732 72.1294 < 2.2e-16
## as.factor(treatment)t 1.00000 0.34960 2.8604 0.0076326
## as.factor(block)J 2.16667 0.52175 4.1527 0.0002506
## as.factor(block)N 4.00000 0.50553 7.9126 7.862e-09
## as.factor(treatment)t:as.factor(block)J -1.33333 0.75277 -1.7712 0.0866826
## as.factor(treatment)t:as.factor(block)N 1.00000 1.02198 0.9785 0.3356562
##
## (Intercept) ***
## as.factor(treatment)t **
## as.factor(block)J ***
## as.factor(block)N ***
## as.factor(treatment)t:as.factor(block)J .
## as.factor(treatment)t:as.factor(block)N
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
stargazer(mod_b1, mod_b2, type = "text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               hue_turn
##                               (1)           (2)
## -----
## as.factor(treatment)t 0.889**           1.000
##                       (0.429)           (0.704)
##
## as.factor(block)J 1.500***           2.167***
##                   (0.525)           (0.704)
##
## as.factor(block)N 4.500***           4.000***
##                   (0.525)           (0.704)
##
## as.factor(treatment)t:as.factor(block)J -1.333
##                                           (0.996)
##
## as.factor(treatment)t:as.factor(block)N 1.000
##                                           (0.996)
##
## Constant 22.222***           22.167***
```

```
##                                (0.429)                                (0.498)
##
## -----
## Observations                    36                                36
## R2                             0.716                            0.760
## Adjusted R2                    0.689                            0.720
## Residual Std. Error            1.286 (df = 32)                1.220 (df = 30)
## F Statistic                    26.846*** (df = 3; 32)          18.985*** (df = 5; 30)
## =====
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

(see page 77 of Analysis of Categorical data) `anova()` from the `stat` package performs type I test (i.e., sequentially adding the additional terms) while `Anova()` from `car` package performs type II test (i.e.,)

```
#anova(long_mod, short_mod, test = 'F')
anova(mod_b2, mod_b1, test = 'F')
```

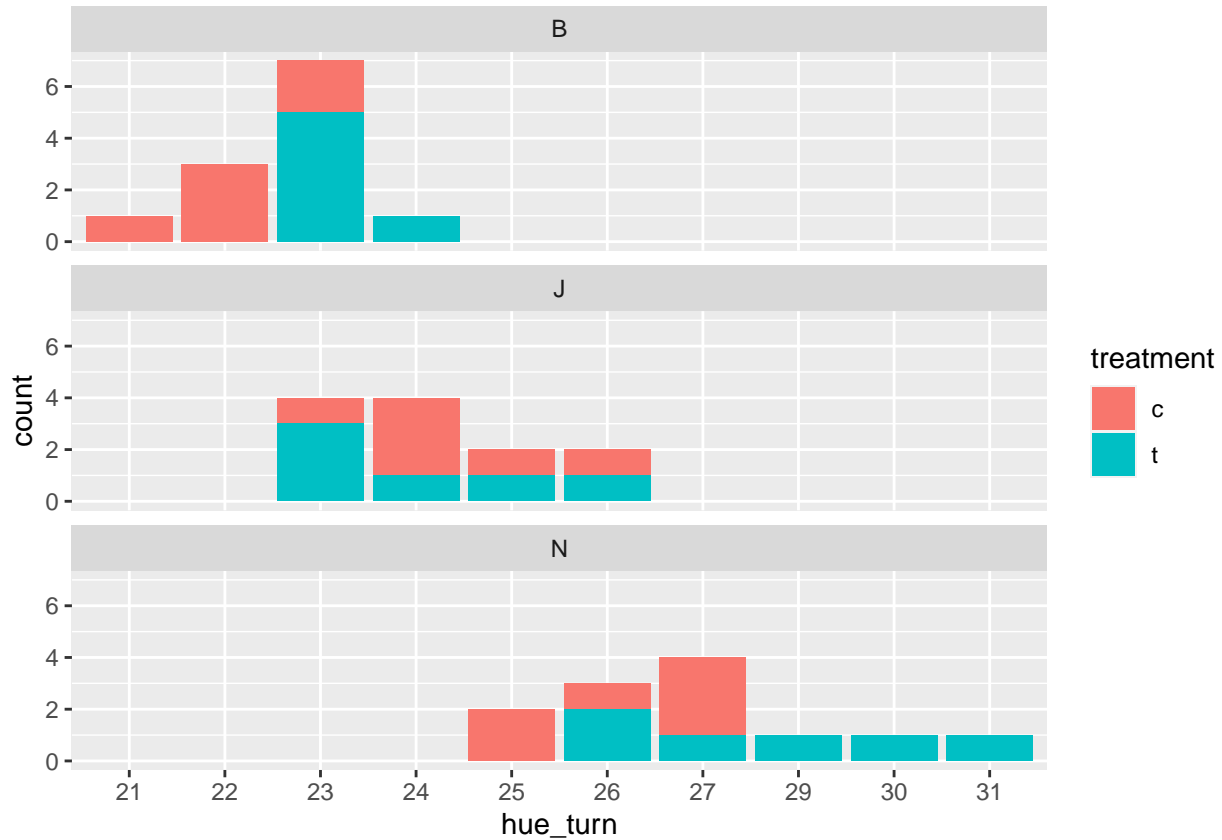
```
## Analysis of Variance Table
##
## Model 1: hue_turn ~ as.factor(treatment) * as.factor(block)
## Model 2: hue_turn ~ as.factor(treatment) + as.factor(block)
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      30 44.667
## 2      32 52.889 -2   -8.2222 2.7612 0.0793 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
b_result
```

```
##   block avocado_number treatment hue_turn
## 1     B             1         c      23
## 2     B            11         c      22
## 3     B            12         c      23
## 4     B             6         c      22
## 5     B             8         c      22
## 6     B             9         c      21
## 7     B            10         t      23
## 8     B             2         t      24
## 9     B             3         t      23
## 10    B             4         t      23
## 11    B             5         t      23
## 12    B             7         t      23
## 13    J            10         c      25
## 14    J            12         c      26
## 15    J             2         c      24
## 16    J             3         c      24
## 17    J             5         c      23
## 18    J             6         c      24
## 19    J             1         t      23
## 20    J            11         t      25
## 21    J             4         t      23
```

```
## 22      J           7           t           24
## 23      J           8           t           23
## 24      J           9           t           26
## 25      N          11           c           26
## 26      N           2           c           27
## 27      N           4           c           27
## 28      N           5           c           25
## 29      N           7           c           27
## 30      N           9           c           25
## 31      N           1           t           31
## 32      N          10           t           27
## 33      N          12           t           26
## 34      N           3           t           30
## 35      N           6           t           26
## 36      N           8           t           29
```

```
b_result %>% ggplot(aes(x = hue_turn, fill=treatment)) + geom_bar() + facet_wrap(~block,ncol = 1)
```



```
db <- as.data.table(b_result)
b_result2 <- db[, .(round(mean(as.integer(hue_turn))),0), by = .(block,treatment)][,1:3]
print(b_result2)
```

```
##      block treatment V1
## 1:      B          c 22
## 2:      B          t 23
```

## 3:	J	c 24
## 4:	J	t 24
## 5:	N	c 26
## 6:	N	t 28