OptPark EDA

2023-10-19

EDA

```
library(tidyverse)
library(lubridate)
library(readxl)
library(stringr)
library(ggplot2)
```

```
df_10 <- read.csv("../data/df_10.csv")
df_5 <- read.csv("../data/df_5.csv")
df_6b <- read.csv("../data/df_6b.csv")
df_4 <- read.csv("../data/df_4.csv")
df_3 <- read.csv("../data/df_3.csv")
df_5b <- read.csv("../data/df_5b.csv")
df_3a <- read.csv("../data/df_3a.csv")
```

```
df_3 %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##            53410             1007             4559              775
```

```
df_3a %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##            49190             5227             9630             7620
```

```
df_4 %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##            71654             2295             5866            16392
```

```
df_5 %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##            48473              518             3259             7746
```

```r
df_5b %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##            20822             1072             7793                2
```

```r
df_6b %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##            37128             1163            32207                1
```

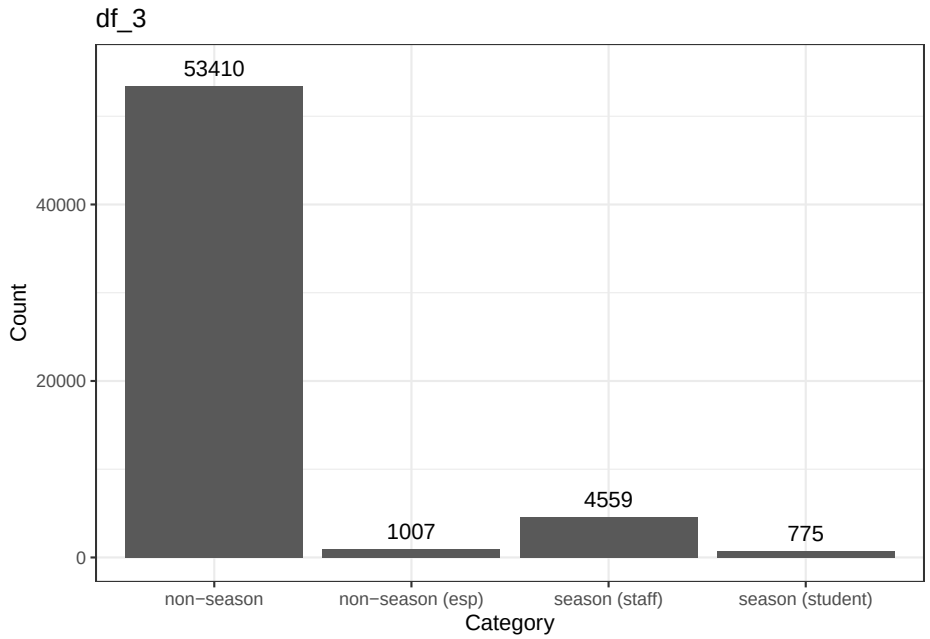
```r
df_10 %>% select("type") %>% table()
```

```
## type
##       non-season non-season (esp)   season (staff) season (student)
##              367                1               11               15
```

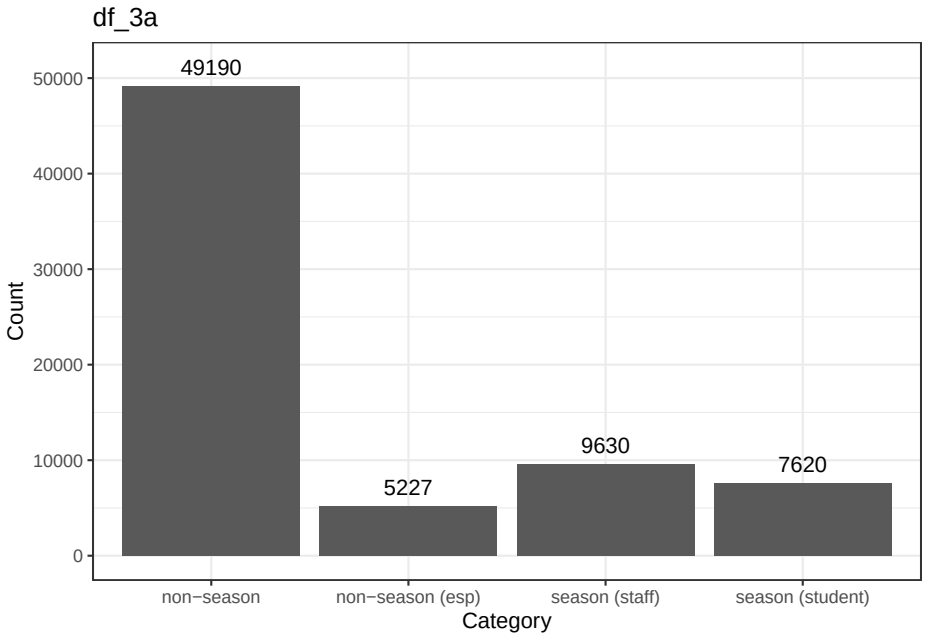
```r
# Create bar plots for each data frame to observe the "type" of parkers in each carpark
barplot_df_3 <- ggplot(df_3, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=2000)+
  labs(x = "Category", y = "Count") +
  ggtitle("df_3") +
  theme_bw()

barplot_df_3a <- ggplot(df_3a, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=2000)+
  labs(x = "Category", y = "Count") +
  ggtitle("df_3a") +
  theme_bw()

barplot_df_4 <- ggplot(df_4, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=2000)+
  labs(x = "Category", y = "Count") +
  ggtitle("df_4") +
  theme_bw()

barplot_df_5 <- ggplot(df_5, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=2000)+
  labs(x = "Category", y = "Count") +
  ggtitle("df_5") +
  theme_bw()

barplot_df_5b <- ggplot(df_5b, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=1000)+
  labs(x = "Category", y = "Count") +
```

```
  ggtitle("df_5b") +
  theme_bw()

barplot_df_6b <- ggplot(df_6b, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=1500)+
  labs(x = "Category", y = "Count") +
  ggtitle("df_6b") +
  theme_bw()

barplot_df_10 <- ggplot(df_10, aes(x = type)) +
  geom_bar() +
  geom_text(aes(label=..count..),stat="count",nudge_y=10)+
  labs(x = "Category", y = "Count") +
  ggtitle("df_10") +
  theme_bw()

# Print or display the bar plots
print(barplot_df_3)
```

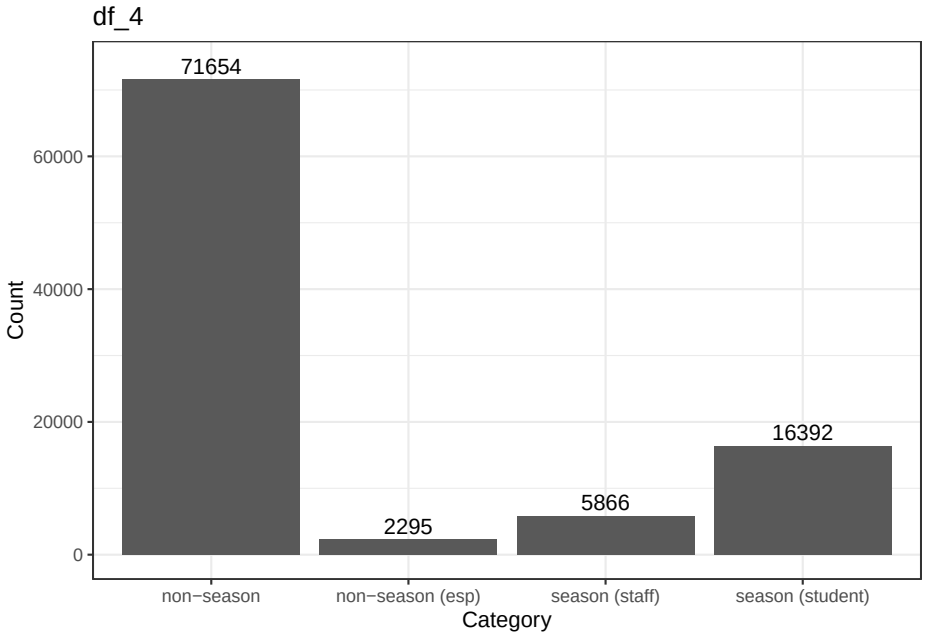
```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.
## i Please use ‘after_stat(count)‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



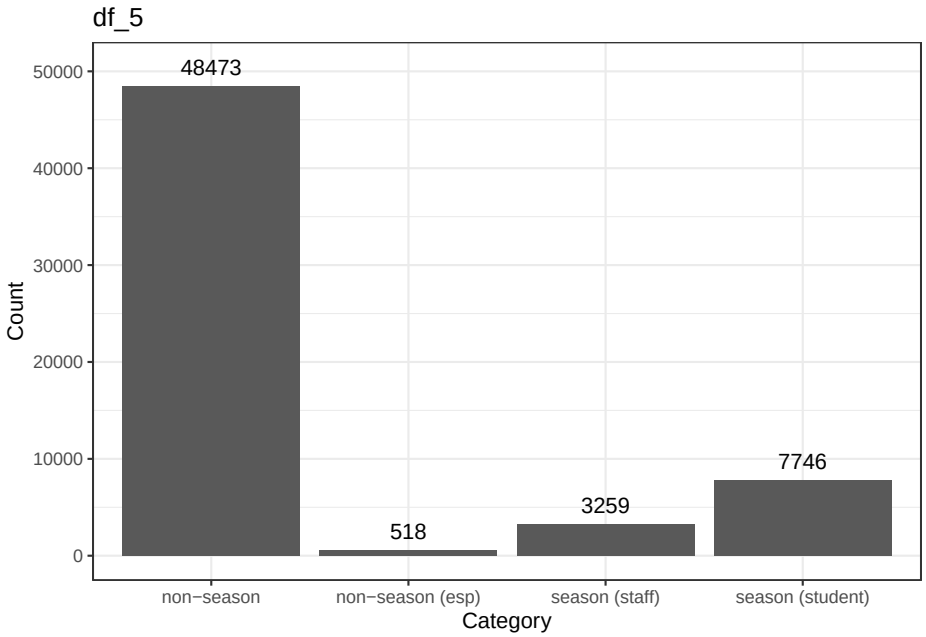

```
print(barplot_df_3a)
```



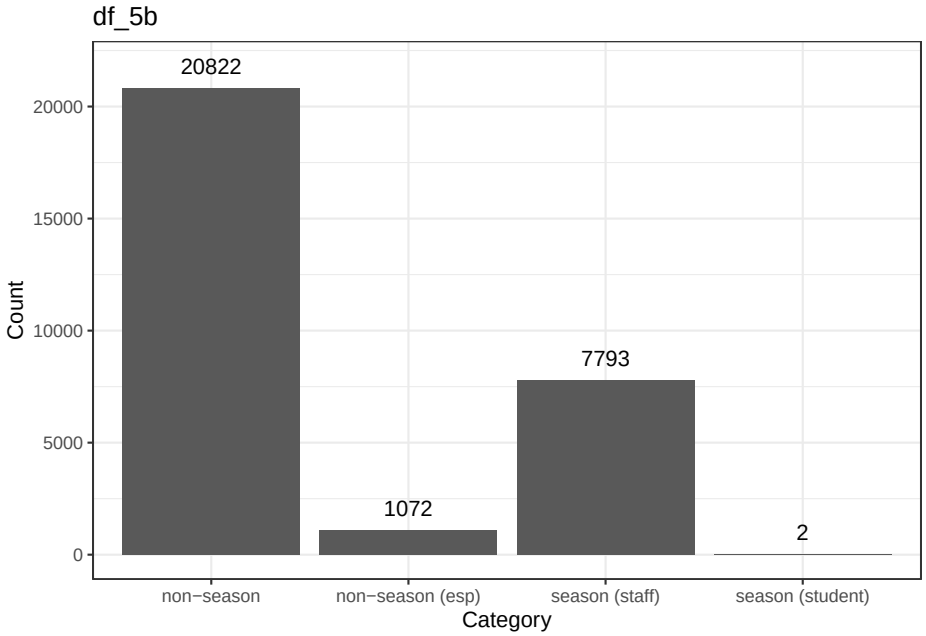

```
print(barplot_df_4)
```



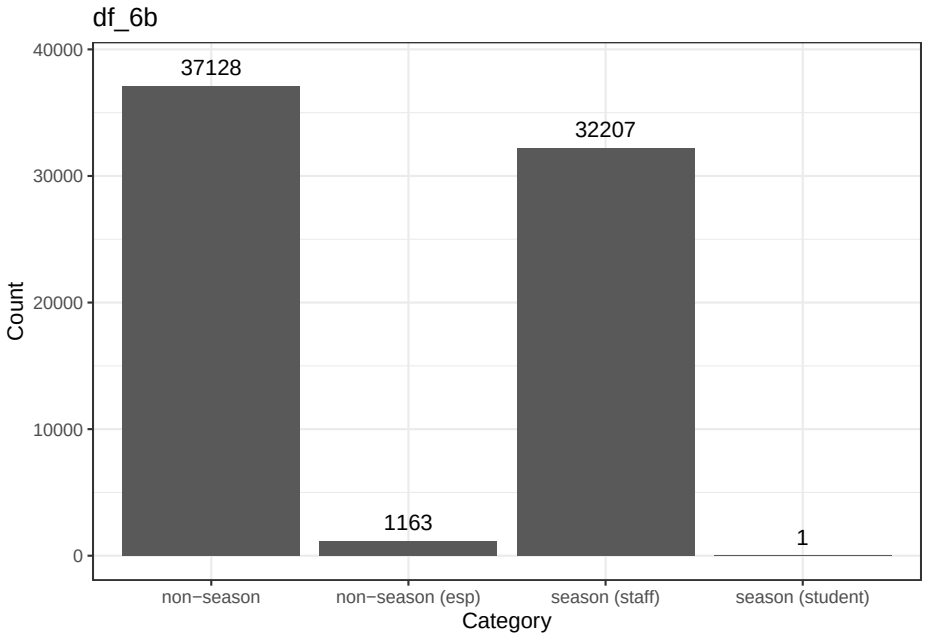

```
print(barplot_df_5)
```



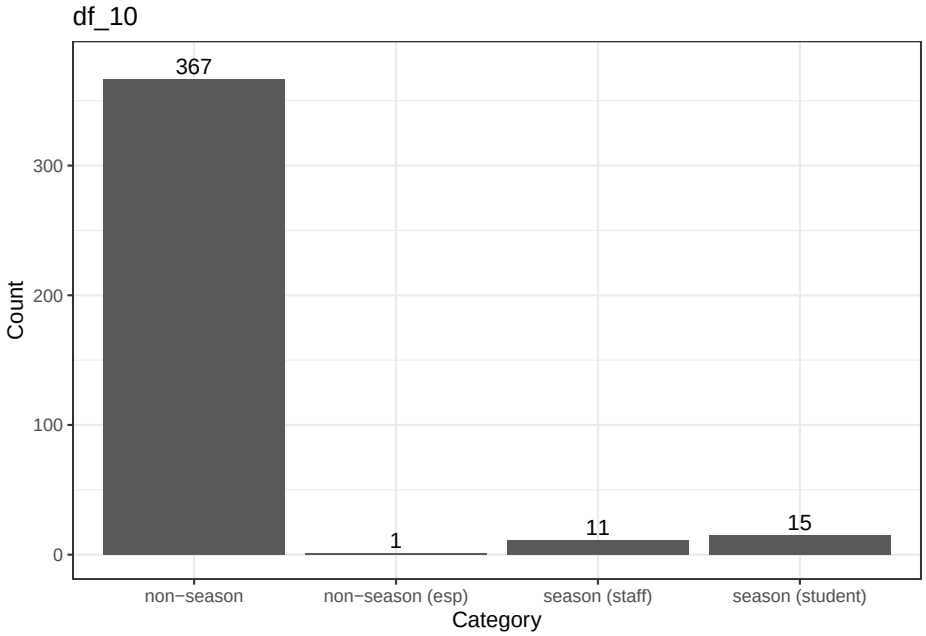

```
print(barplot_df_5b)
```




```
print(barplot_df_6b)
```

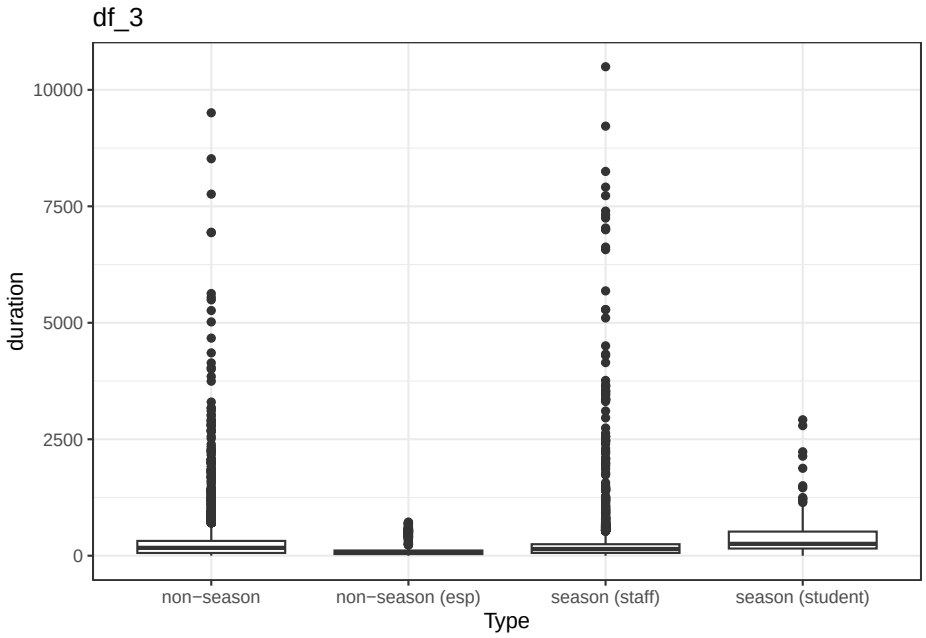



```
print(barplot_df_10)
```

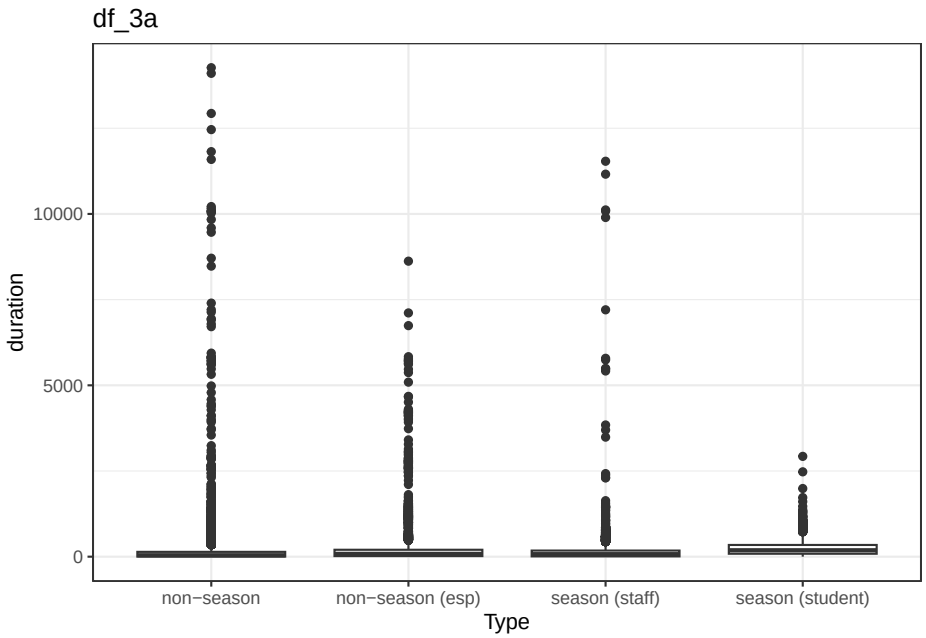



```
# Create boxplots for each data frame
boxplot_df_3 <- ggplot(df_3, aes(x = type, y = du_val)) +
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_3") +
  theme_bw()

boxplot_df_3a <- ggplot(df_3a, aes(x = type, y = du_val)) +
```

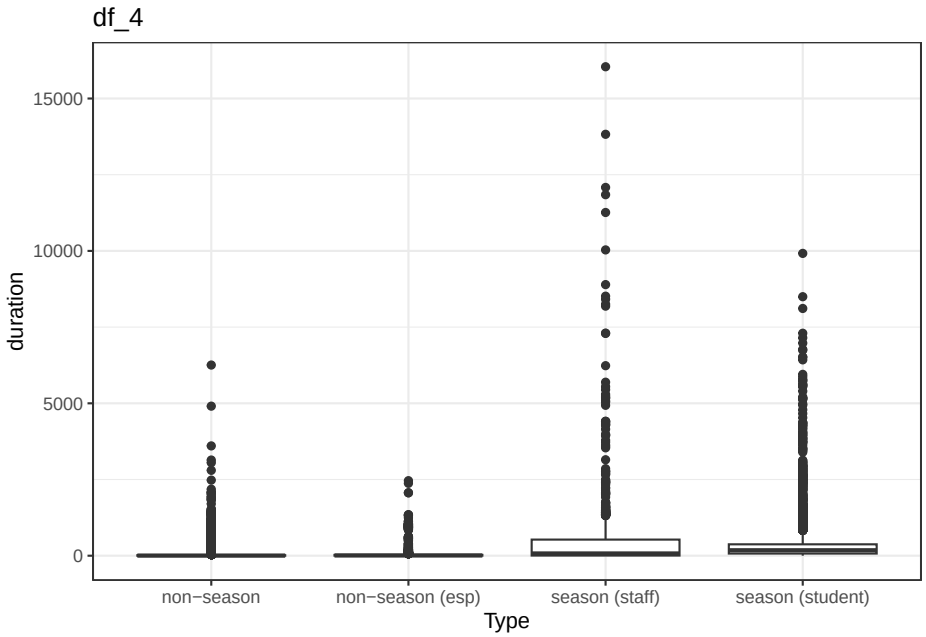
```
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_3a") +
  theme_bw()

boxplot_df_4 <- ggplot(df_4, aes(x = type, y = du_val)) +
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_4") +
  theme_bw()

boxplot_df_5 <- ggplot(df_5, aes(x = type, y = du_val)) +
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_5") +
  theme_bw()

boxplot_df_5b <- ggplot(df_5b, aes(x = type, y = du_val)) +
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_5b") +
  theme_bw()

boxplot_df_6b <- ggplot(df_6b, aes(x = type, y = du_val)) +
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_6b") +
  theme_bw()

boxplot_df_10 <- ggplot(df_10, aes(x = type, y = du_val)) +
  geom_boxplot() +
  labs(x = "Type", y = "duration") +
  ggtitle("df_10") +
  theme_bw()

# Print or display the boxplots
print(boxplot_df_3)
```



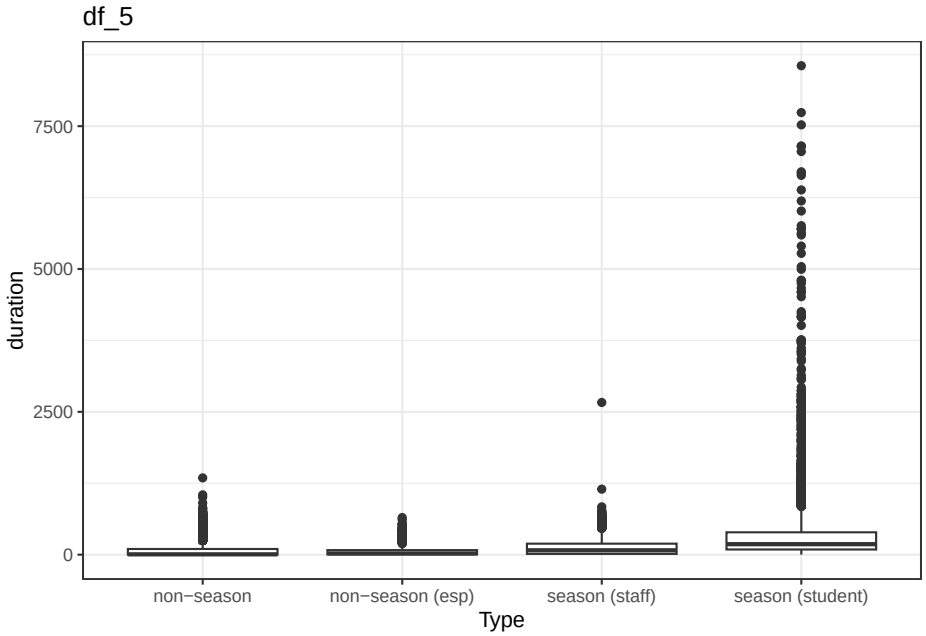

```
print(boxplot_df_3a)
```



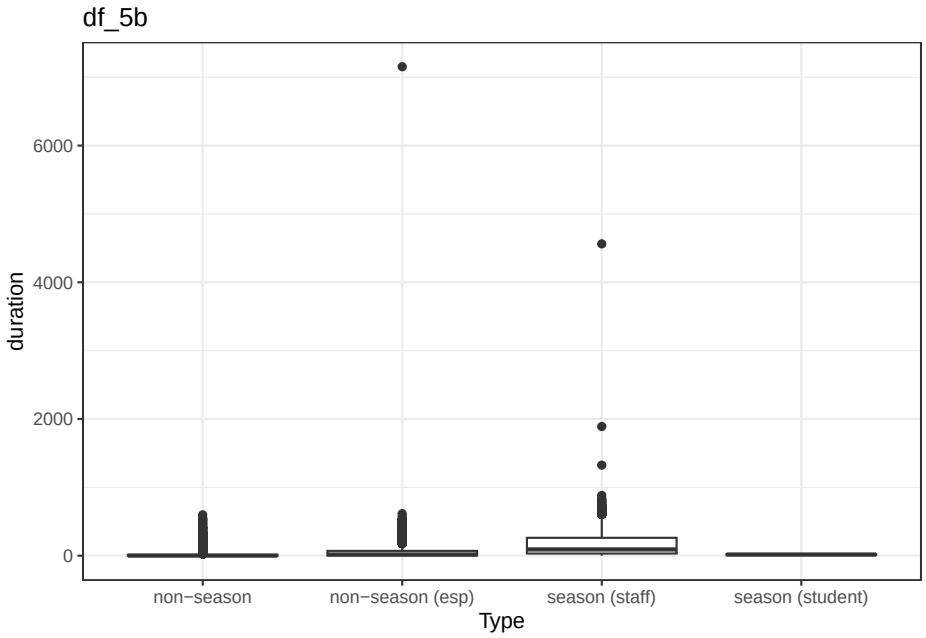

```
print(boxplot_df_4)
```



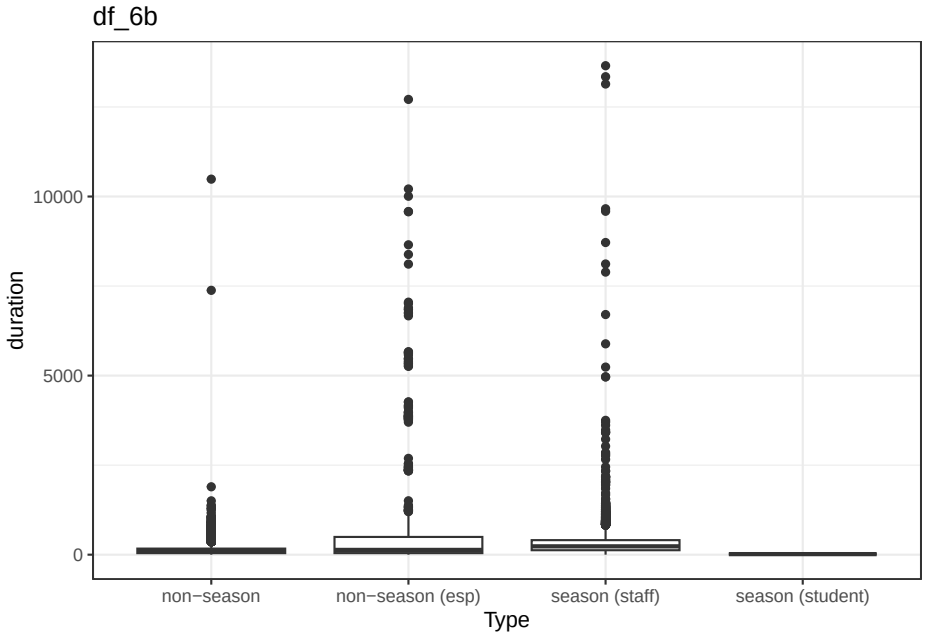

```
print(boxplot_df_5)
```



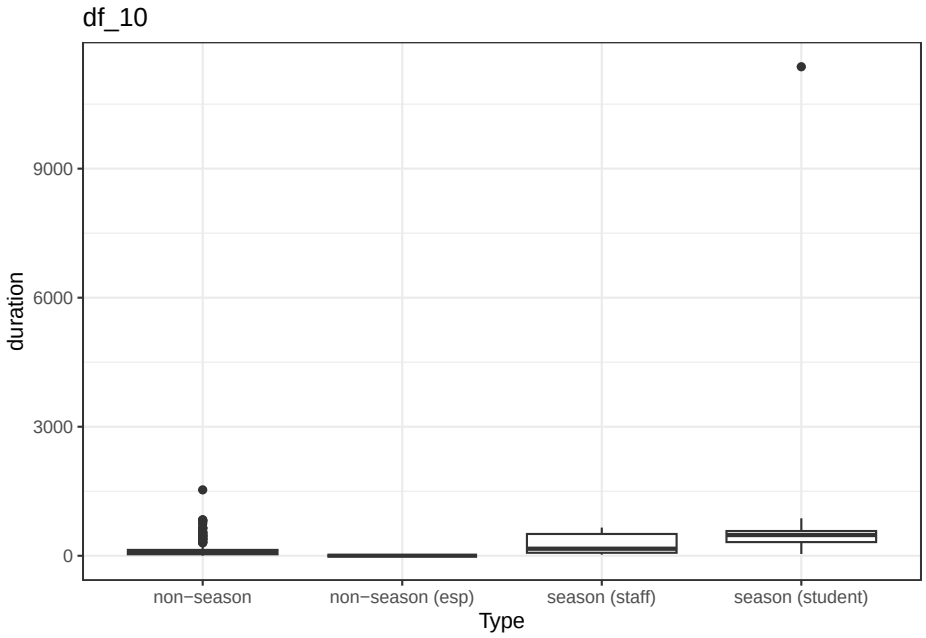

```
print(boxplot_df_5b)
```




```
print(boxplot_df_6b)
```

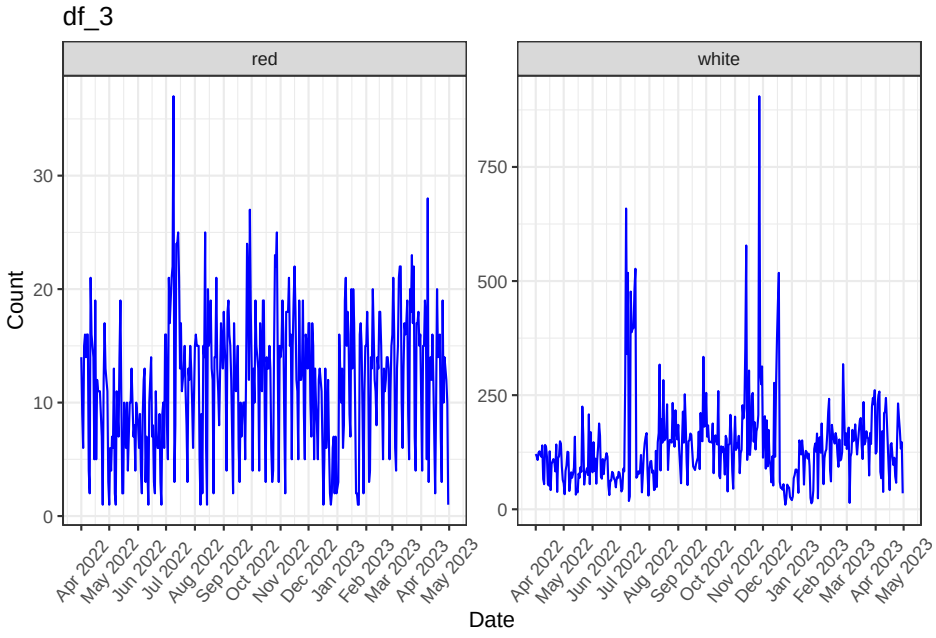



```
print(boxplot_df_10)
```

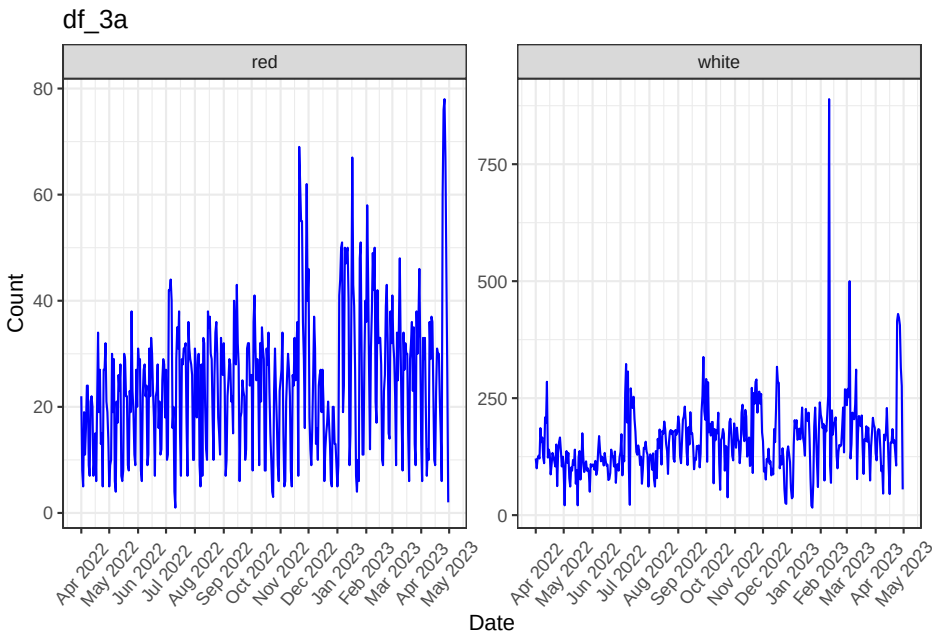



```
# Create entry timeline for each data frame, faceted by slot type
timeline_df_3 <- ggplot(df_3) +
  geom_line(aes(x = as.Date(enter_time)),stat = "count",color="blue") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_3") +
  facet_wrap(~slot, scales = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

timeline_df_3a <- ggplot(df_3a) +
  geom_line(aes(x = as.Date(enter_time)),stat = "count",color="blue") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_3a") +
  facet_wrap(~slot, scales = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

timeline_df_4 <- ggplot(df_4) +
  geom_line(aes(x = as.Date(enter_time)),stat = "count",color="blue") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_4") +
  facet_wrap(~slot, scales = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

timeline_df_5 <- ggplot(df_5) +
  geom_line(aes(x = as.Date(enter_time)),stat = "count",color="blue") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
```

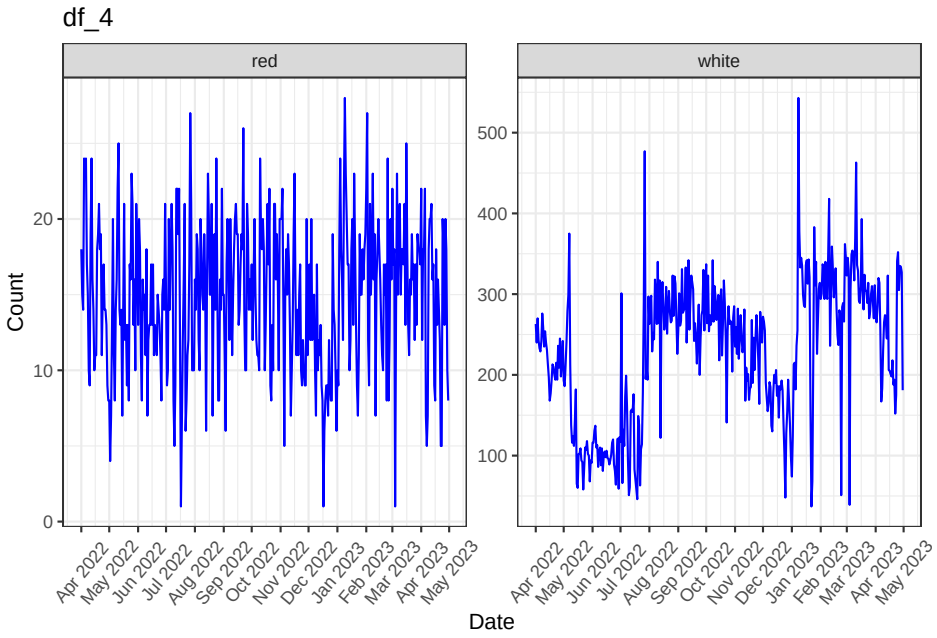
```
  ggtitle("df_5") +
  facet_wrap(~slot, scales = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

timeline_df_5b <- ggplot(df_5b) +
  geom_line(aes(x = as.Date(enter_time)),stat = "count",color="blue") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_5b") +
  facet_wrap(~slot, scales = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

timeline_df_6b <- ggplot(df_6b) +
  geom_line(aes(x = as.Date(enter_time)),stat = "count",color="blue") +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_6b") +
  facet_wrap(~slot, scales = "free_y") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

# Print or display the timeline
print(timeline_df_3)
```



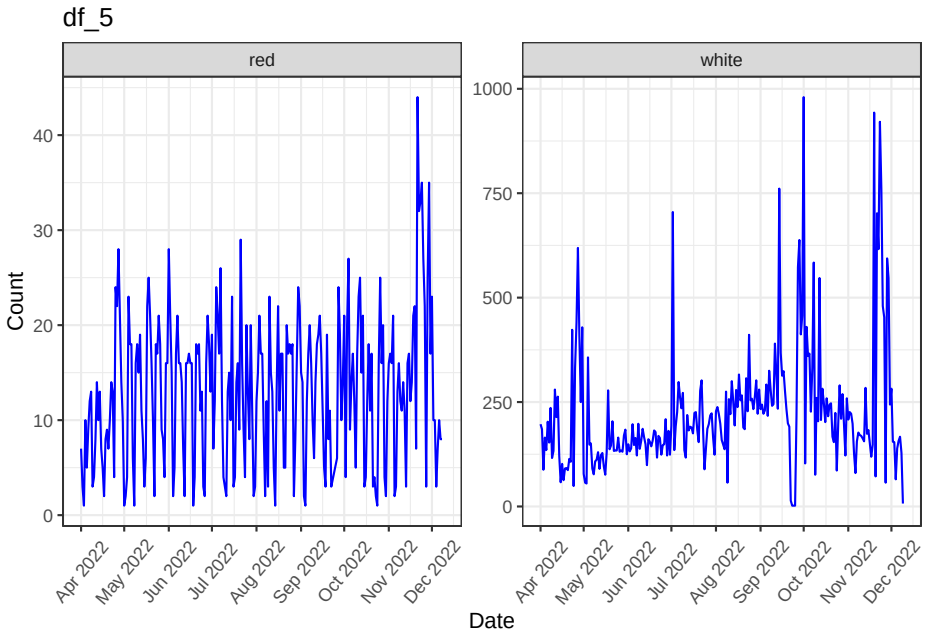

```
print(timeline_df_3a)
```



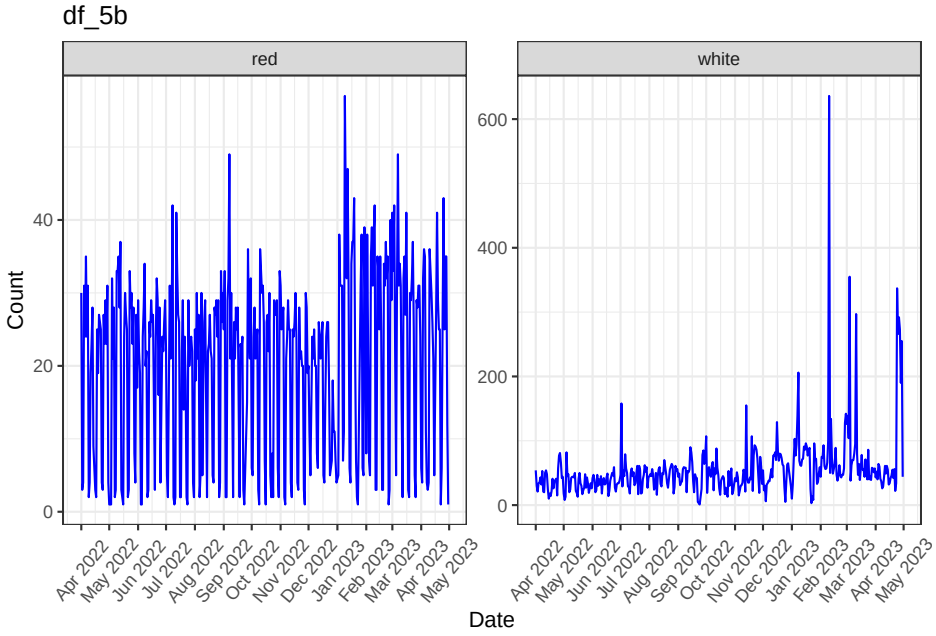

```
print(timeline_df_4)
```



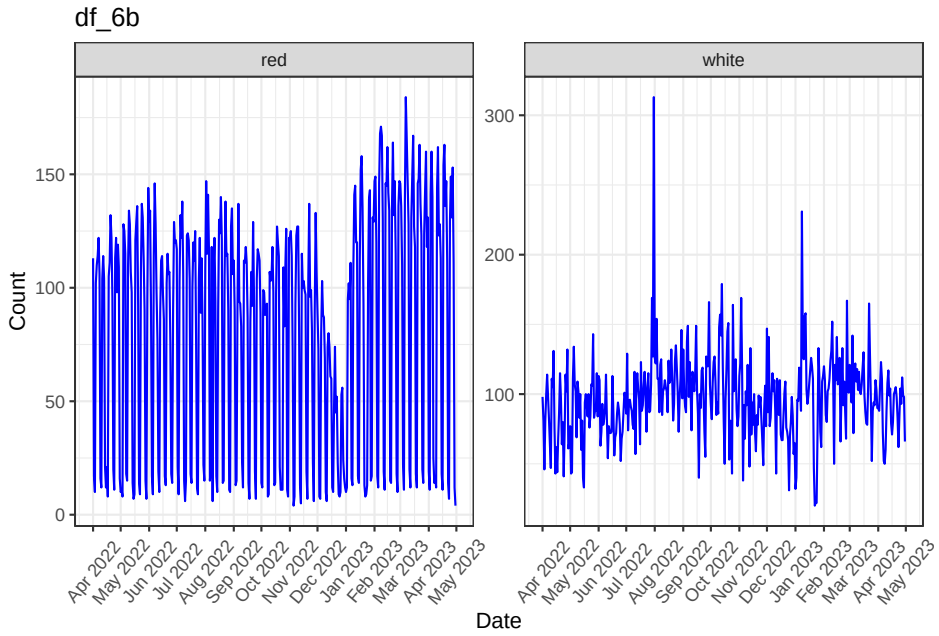

```
print(timeline_df_5)
```




```
print(timeline_df_5b)
```

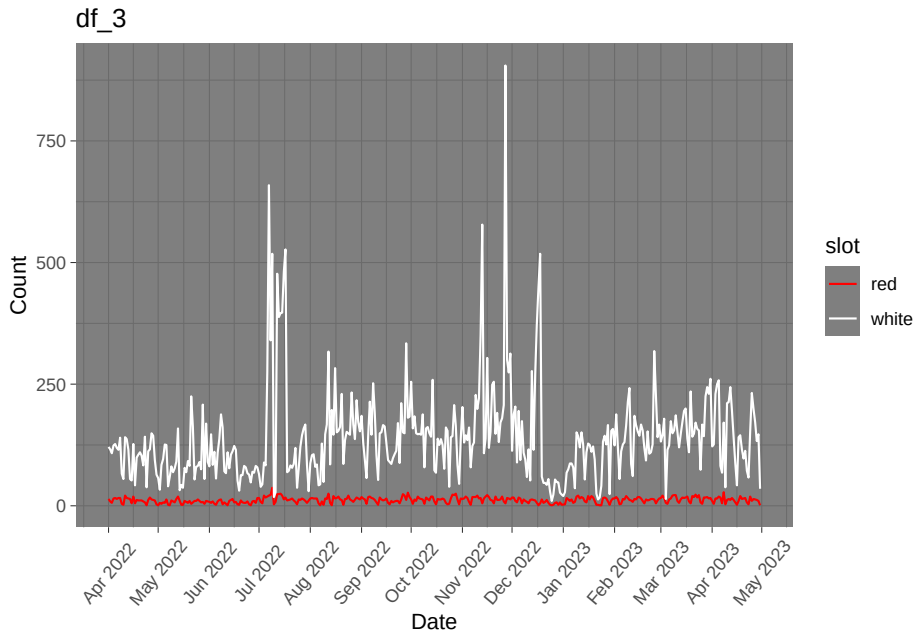



```
print(timeline_df_6b)
```

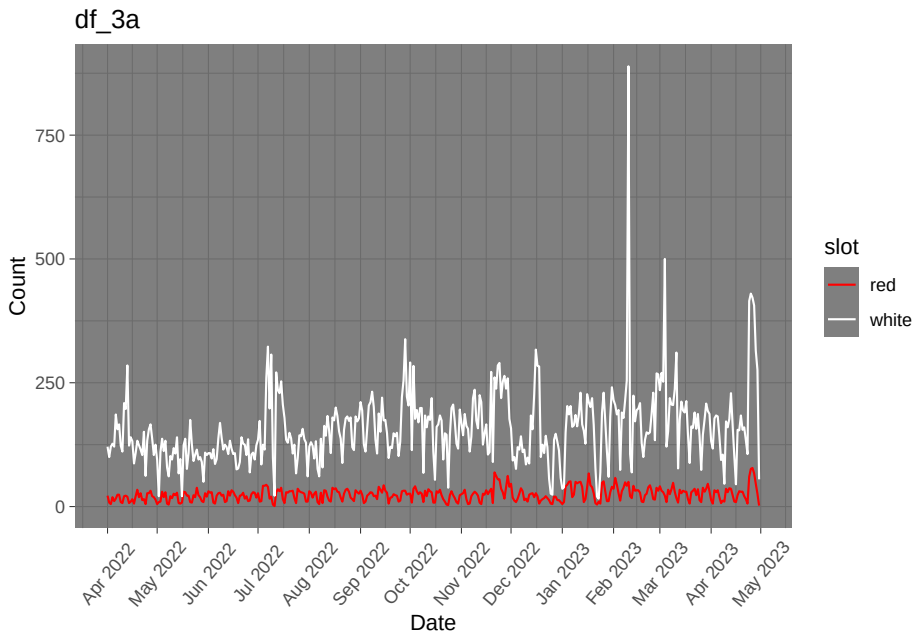



```
# Create line plot for each data frame, lines colored according by the slot type
line_split_df3 <- ggplot(df_3) +
  geom_line(aes(x = as.Date(enter_time),color = slot),stat = "count") +
  scale_color_manual(values = c("red", "white")) +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_3") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

line_split_df3a <- ggplot(df_3a) +
  geom_line(aes(x = as.Date(enter_time),color = slot),stat = "count") +
  scale_color_manual(values = c("red", "white")) +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_3a") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

line_split_df4 <- ggplot(df_4) +
  geom_line(aes(x = as.Date(enter_time),color = slot),stat = "count") +
  scale_color_manual(values = c("red", "white")) +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_4") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

line_split_df5 <- ggplot(df_5) +
  geom_line(aes(x = as.Date(enter_time),color = slot),stat = "count") +
  scale_color_manual(values = c("red", "white"))+
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
```

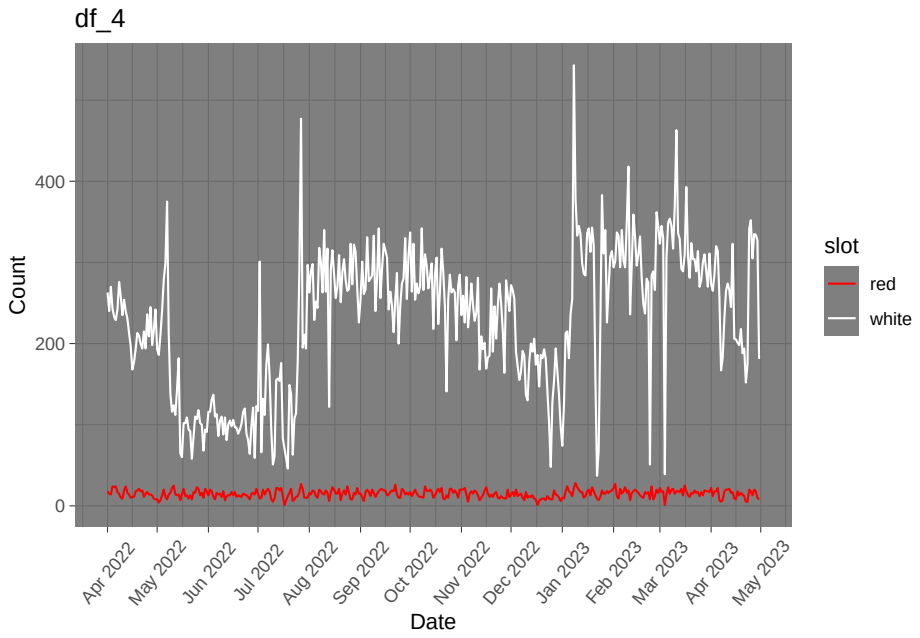
```
  labs(x = "Date", y = "Count") +
  ggtitle("df_5") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

line_split_df5b <- ggplot(df_5b) +
  geom_line(aes(x = as.Date(enter_time),color = slot),stat = "count") +
  scale_color_manual(values = c("red", "white")) +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_5b") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))

line_split_df6b <- ggplot(df_6b) +
  geom_line(aes(x = as.Date(enter_time),color = slot),stat = "count") +
  scale_color_manual(values = c("red", "white")) +
  scale_x_date(date_labels = "%b %Y", date_breaks = "1 month")+
  labs(x = "Date", y = "Count") +
  ggtitle("df_6b") +
  theme_dark() +
  theme(axis.text.x = element_text(angle = 50, vjust = 0.5))


print(line_split_df3)
```



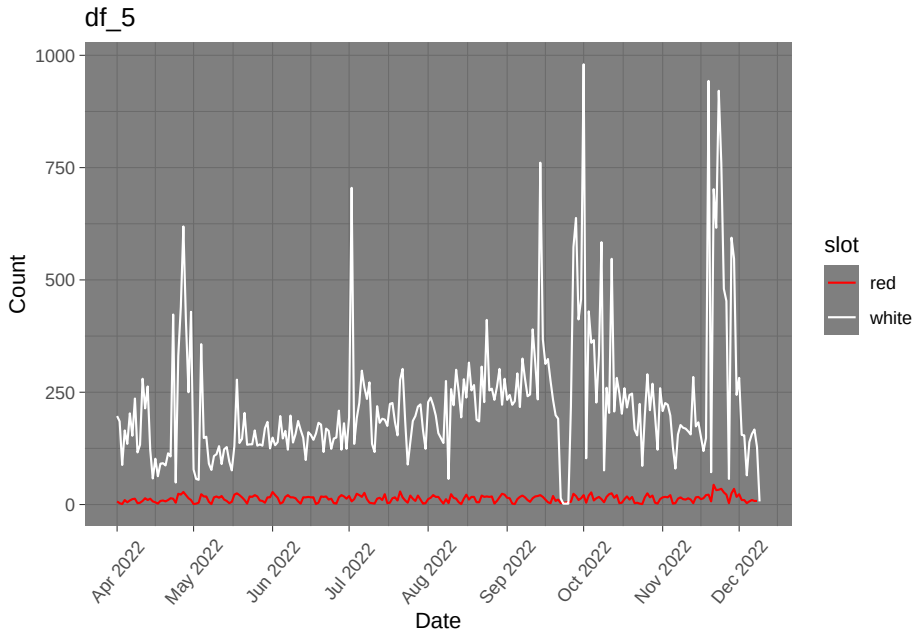

```
print(line_split_df3a)
```



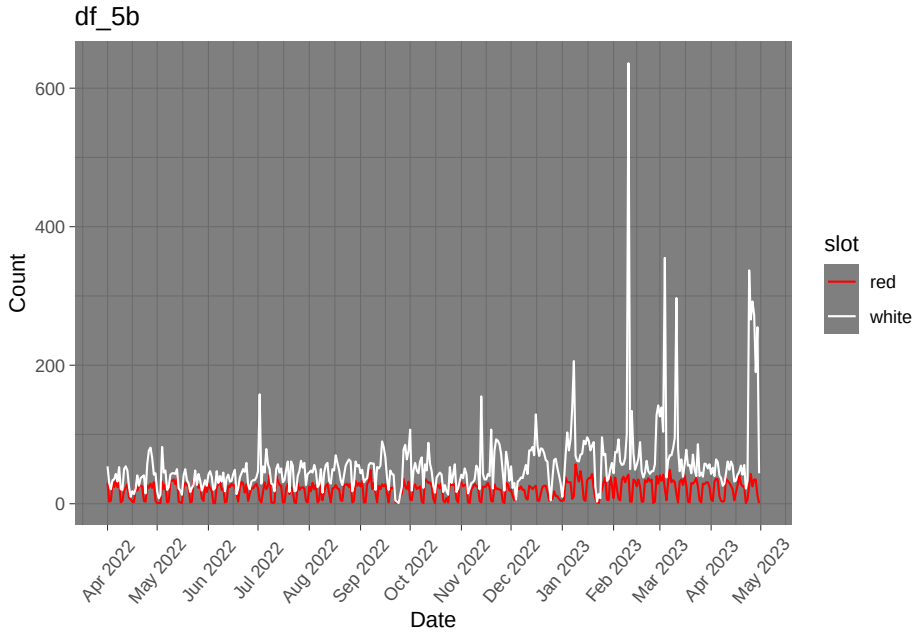

```
print(line_split_df4)
```



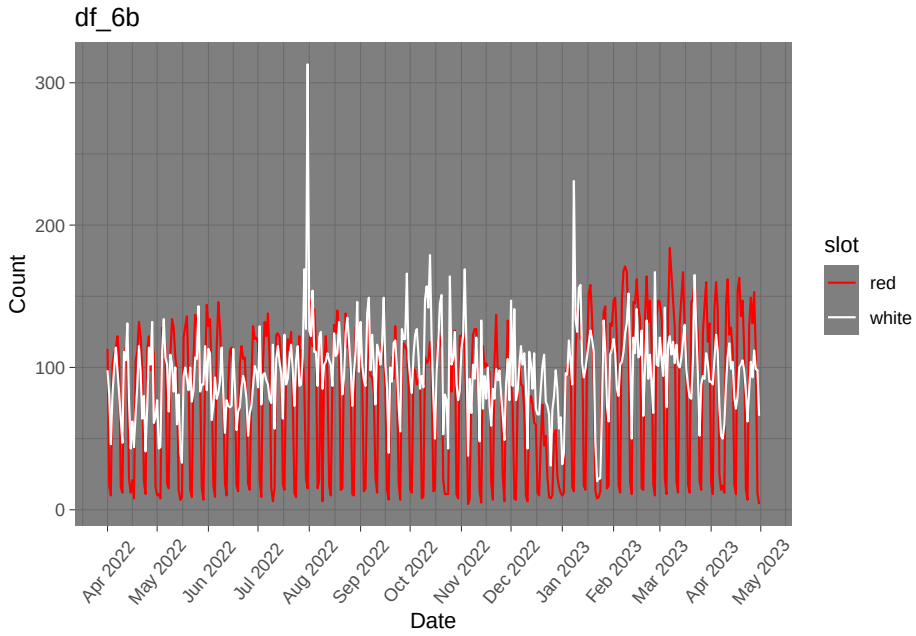

```
print(line_split_df5)
```




```
print(line_split_df5b)
```




```
print(line_split_df6b)
```


df_6b
300
200
100
0
Count
Apr 2022
May 2022
Jun 2022
Jul 2022
Aug 2022
Sep 2022
Oct 2022
Nov 2022
Dec 2022
Jan 2023
Feb 2023
Mar 2023
Apr 2023
May 2023
Date
slot
red
white