A/B Test in R

Mumitul Al-Anik

December 17, 2020

```
options(warn = -1)
```

```
#Installing necessary packages
```

```
library(dplyr)
library(readr)
library(tidyverse)
library(lubridate)
library(ggplot2)
```

```
#Loading data and making it machine readable
```

```
ab_data <- read_csv("C:/Users/Mumitul/Desktop/ab_data.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   user_id = col_double(),
##   timestamp = col_datetime(format = ""),
##   group = col_character(),
##   landing_page = col_character(),
##   converted = col_double()
## )
```

```
View(ab_data)
df<- read.csv("C:/Users/Mumitul/Desktop/ab_data.csv")
df <- data.frame(df)
colnames(df)
```

```
## [1] "user_id"      "timestamp"    "group"        "landing_page" "converted"
```

```
nrow(df)
```

```
## [1] 294478
```

```
#find out not aligned info between 'group' and 'landing_page'

notaligned_user=df %>% filter((df$group=='treatment'& landing_page == "old_page")|(df$group=='control'& landing_page == "new_page"))

#aligned info between 'group' and 'landing_page'
df1=df[(df$group=='control'& df$landing_page == "old_page")|(df$group=='treatment'& df$landing_page == "new_page"),]

# I am only keeping the unique ids for the anlysis
unique_id <- unique(df1$user_id)
length(unique_id)
```

```
## [1] 290584
```

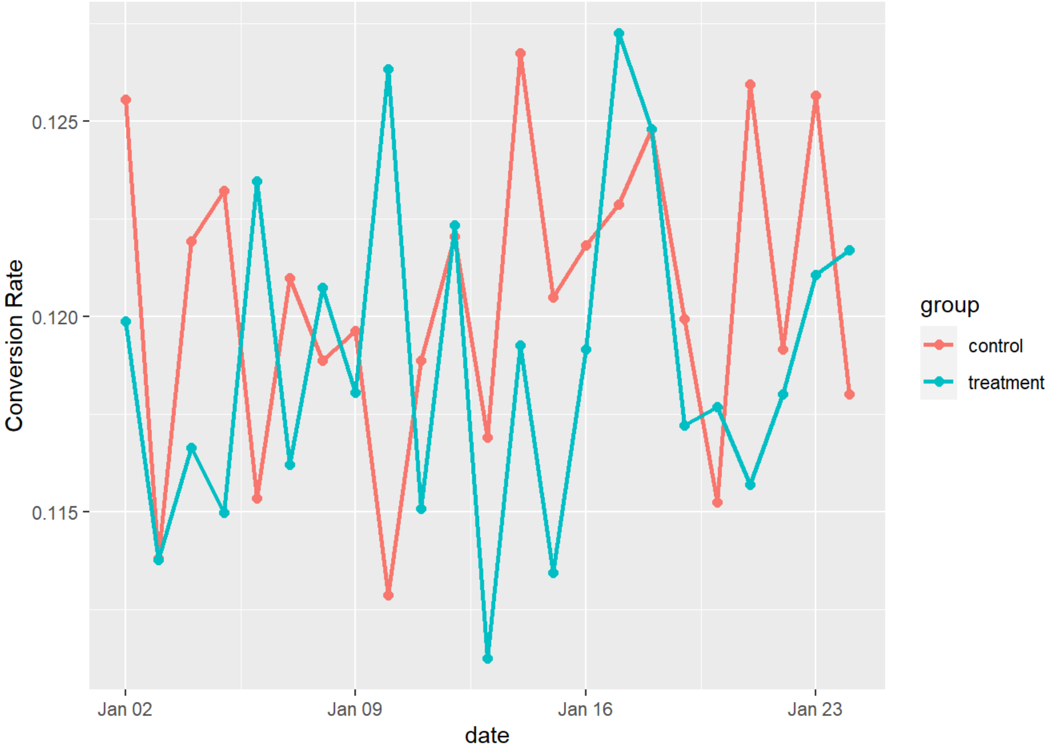
```
#if a user clicked several times, only keep the first result for analysis

df1 <- df1 %>% group_by(user_id) %>% arrange(timestamp)
df2=df1[!duplicated(df1$user_id),]
nrow(df2)
```

```
## [1] 290584
```

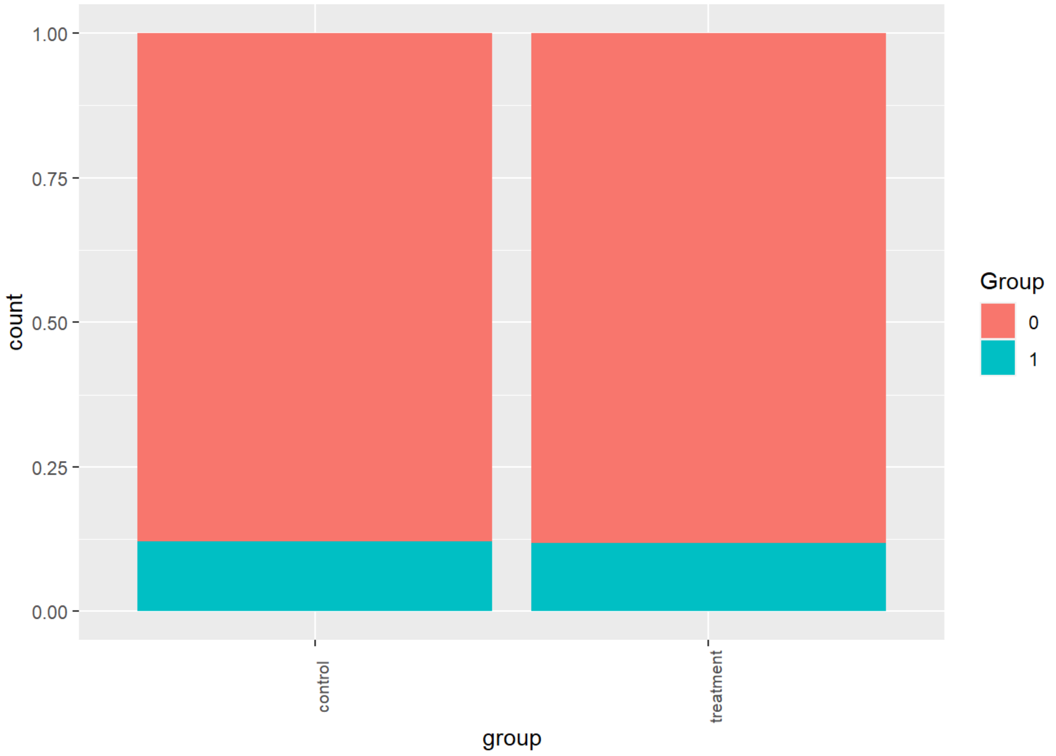
```
#Converting the time stamp to yyyy-mm-dd format
#Calculating the conversion rate by the mean

df2$date<-as.Date(df2$timestamp)
df_new<-df2 %>%
group_by(date,group) %>%
summarize(conversion_rate = mean(converted))
```

```
## `summarise()` regrouping output by 'date' (override with `.groups` argument)
```

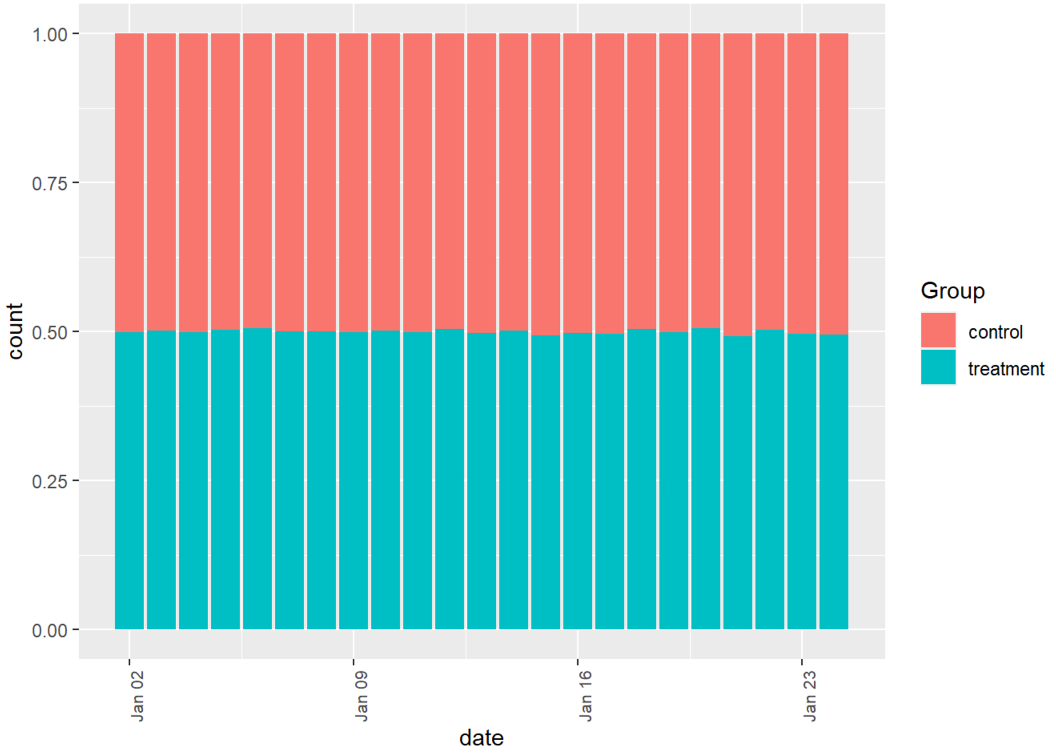
```
# Some desciptive data anlysis using ggplot

ggplot(df_new,aes(x=date,y=conversion_rate,
  color = group,group = group)) +
  geom_point(size = 2) +
  geom_line(lwd = 1) +
  labs(x = "date",  y = "Conversion Rate")
```



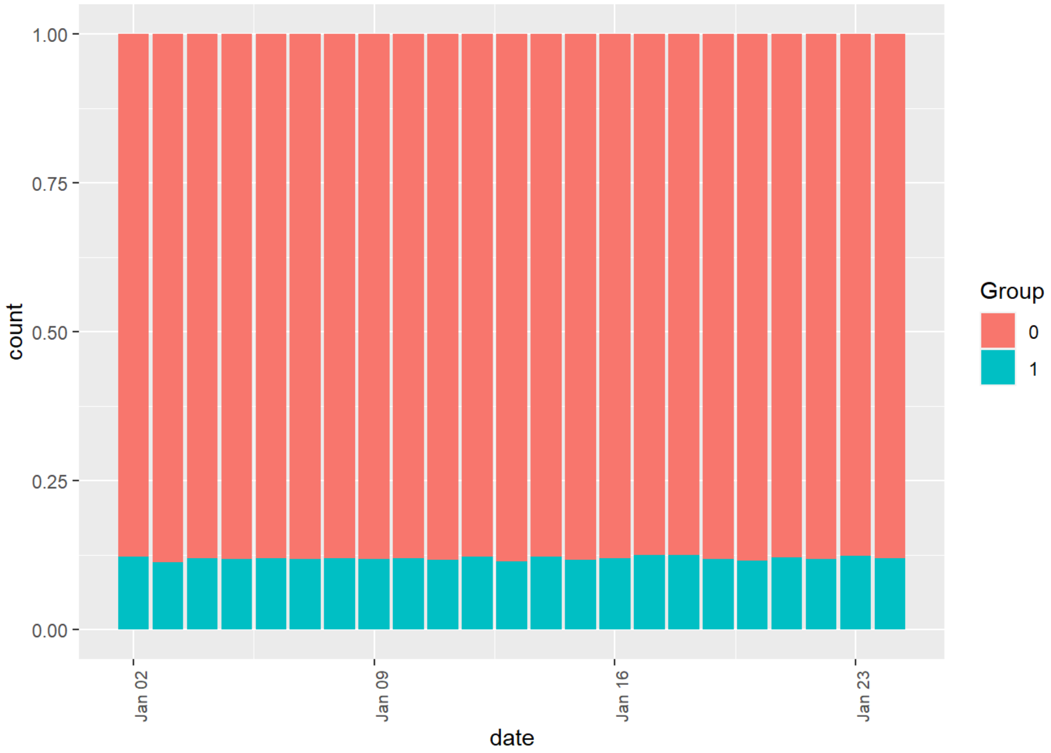
```
ggplot(df2, aes(x = group, fill = factor(converted)) +geom_bar(position = "fill") + theme(axis.text.x = element_text( size=8, angle=90)) + scale_fill_discrete(name = "Group")
```



```
ggplot(df2, aes(x = date, fill = factor(group))) +geom_bar(position = "fill") + theme(axis.text.x = element_
text( size=8, angle=90)) + scale_fill_discrete(name = "Group")
```



```
ggplot(df2, aes(x = date, fill = factor(converted))) +geom_bar(position = "fill") + theme(axis.text.x = elem
ent_text( size=8, angle=90)) + scale_fill_discrete(name = "Group")
```



```
#AB Testing: Organize variables and run logistic regression
#H0:Pnew<=Pold  H1:Pnew>Pold

df2$group = factor(df2$group, levels = c("control", "treatment"))
fit<-glm(converted ~ group, family = "binomial", data =df2)
summary(fit)
```

```
##
## Call:
## glm(formula = converted ~ group, family = "binomial", data = df2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.5065  -0.5065  -0.5030  -0.5030   2.0641
##
## Coefficients:
##                 Estimate Std. Error  z value Pr(>|z|)
## (Intercept)    -1.988777   0.008062 -246.671   <2e-16 ***
## grouptreatment -0.014989   0.011434   -1.311     0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212778  on 290583  degrees of freedom
## Residual deviance: 212776  on 290582  degrees of freedom
## AIC: 212780
##
## Number of Fisher Scoring iterations: 4
```

```
# group treatment p-value 0.19>0.05
#Finding is not statistically significant, cannot reject H0
#Visitors do not like the new interface of the webpage
```