Clustering Analysis: 3865

Nisha Iyer

July 11, 2016

```
setwd("~/Desktop/practicum/venga_practicum/")
library(dplyr)
library(ggplot2)
library(data.table)
library(tidyr)
library(reshape2)
library(stats)
library(caret)
library(corrplot)
library(fpc)
library(cluster)
options(scipen=999)
options( java.parameters = "-Xmx4g" )
load("~/Desktop/practicum/venga_practicum/final_analysis.cluster.RData")
```

Final analysis based on three groups; first,repeat, and all users. I am basing the columns off of segmentation categories from insights and analysis PDF. Will cover the following areas for the three user groups:

- Total Revenue
- Behavior by Hour
- Behavior by Month (Seasonality)
- Holidays (Christmas/New Years/ Valentine's Day)
- Discounts
- Weekend vs. Weekday Diners

This analysis is using the clusters built from the 15 variable reduced set.

```
#These are the columns that built the actual clusters, after dimension reduction:
reduced.names
```

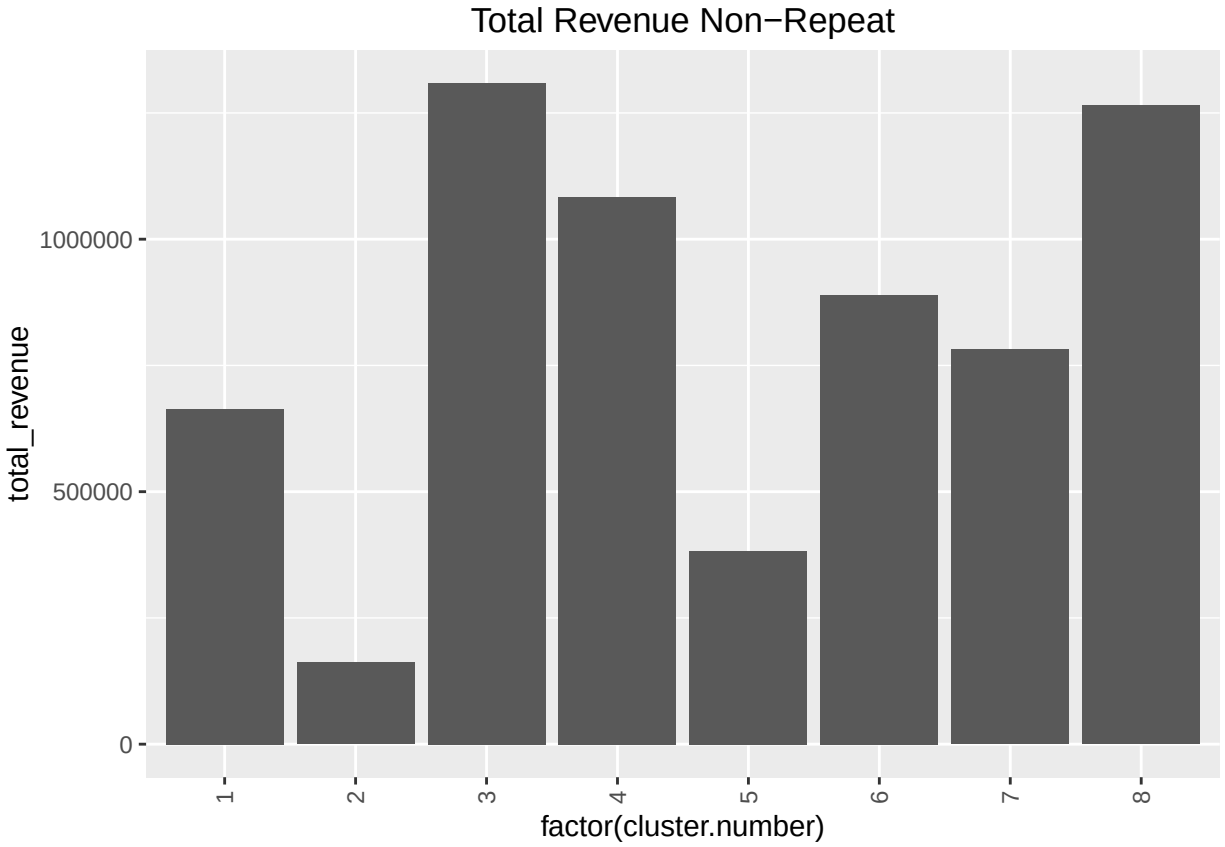
```
##  [1] "num_of_repeats"           "max_party_size"
##  [3] "avg_cover_count"          "max_spend"
##  [5] "avg_diffdays_madeon_open" "avg_diffmins_rsvp_open"
##  [7] "avg_brunch_visits"        "first_weekend_visit"
##  [9] "first_dinner_visit"       "avg_Beverage_spend"
## [11] "avg_Food_qty"             "avg_Liquor.Beer_qty"
## [13] "avg_Wine_qty"             "Cancelled_visits"
## [15] "avg_days_bt_visits"
```

Functions I use throughout:

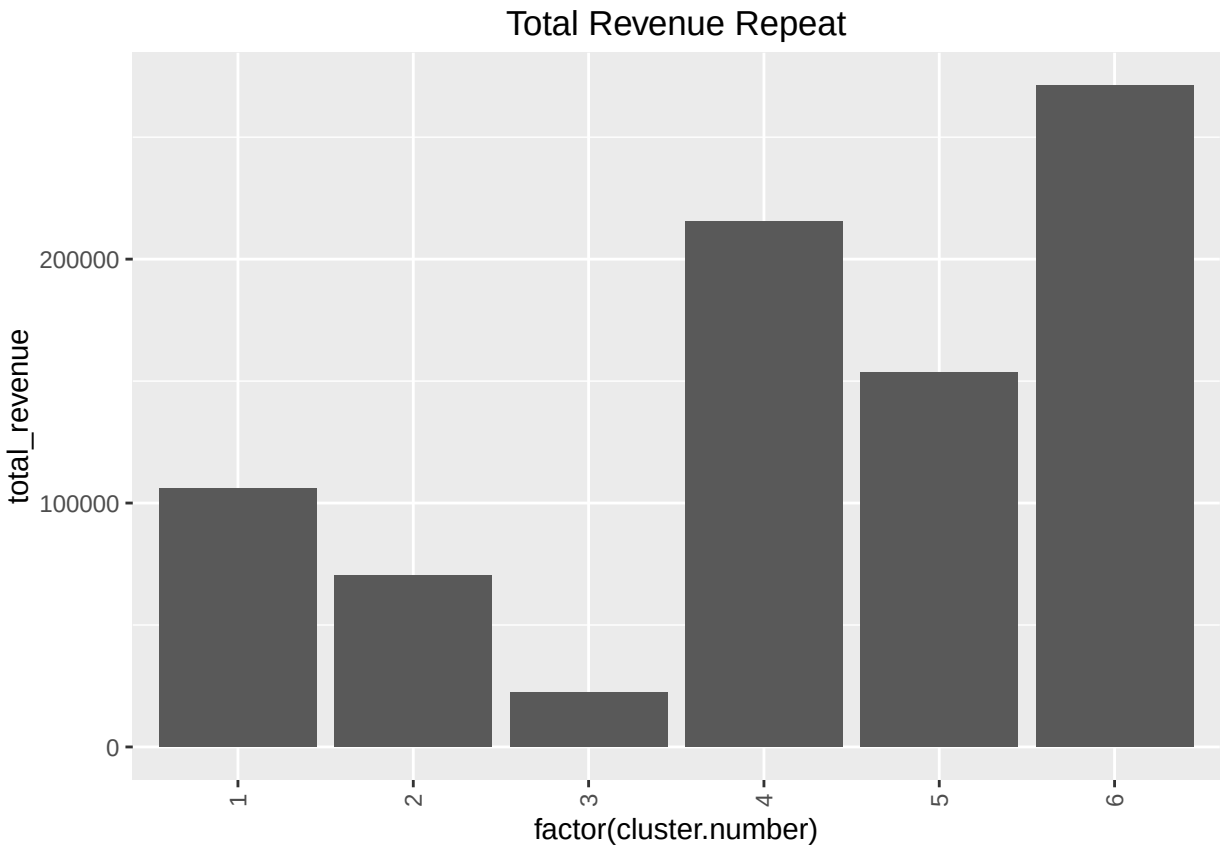
```r
# Adding hour and month to all three data sources:
firstuser.final <- hour_month(firstuser.final)
repeat.final <- hour_month(repeat.final)
user.final <- hour_month(user.final)
```

Total Revenue:

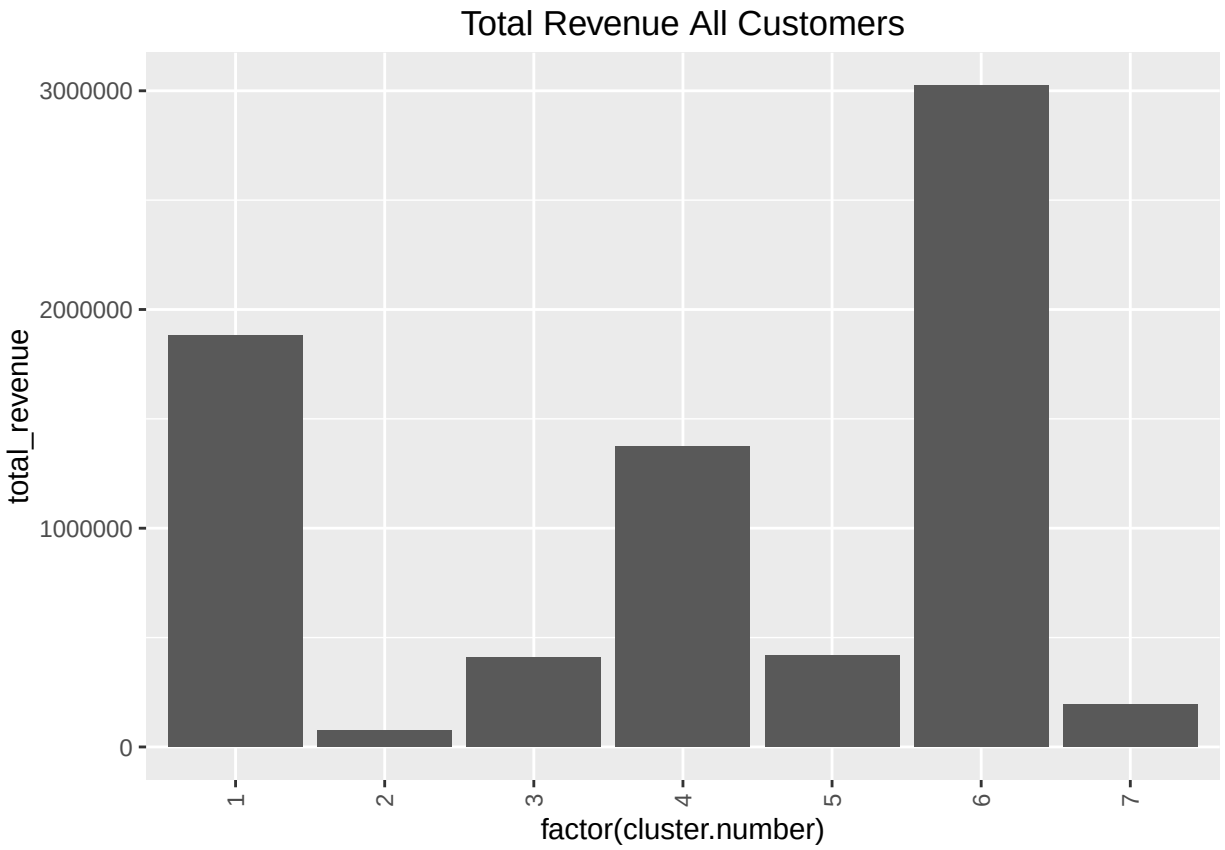
```r
#Total revenue by cluster
first.users.revenue <- total_rev_cluster(firstuser.final)
return.users.revenue <- total_rev_cluster(repeat.final)
users.revenue <- total_rev_cluster(user.final)


#Revenue plot results
ggplot(first.users.revenue, aes(x = factor(cluster.number), y = total_revenue)) +
  geom_bar(stat = "identity") + theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) + ggtitle("
```




```r
ggplot(return.users.revenue, aes(x = factor(cluster.number), y = total_revenue)) +
  geom_bar(stat = "identity") + theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) + ggtitle("
```




```
ggplot(users.revenue, aes(x = factor(cluster.number), y = total_revenue)) +
  geom_bar(stat = "identity") + theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) + ggtitle("
```



Look further in to the highest and lowest revenue clusters for each user type

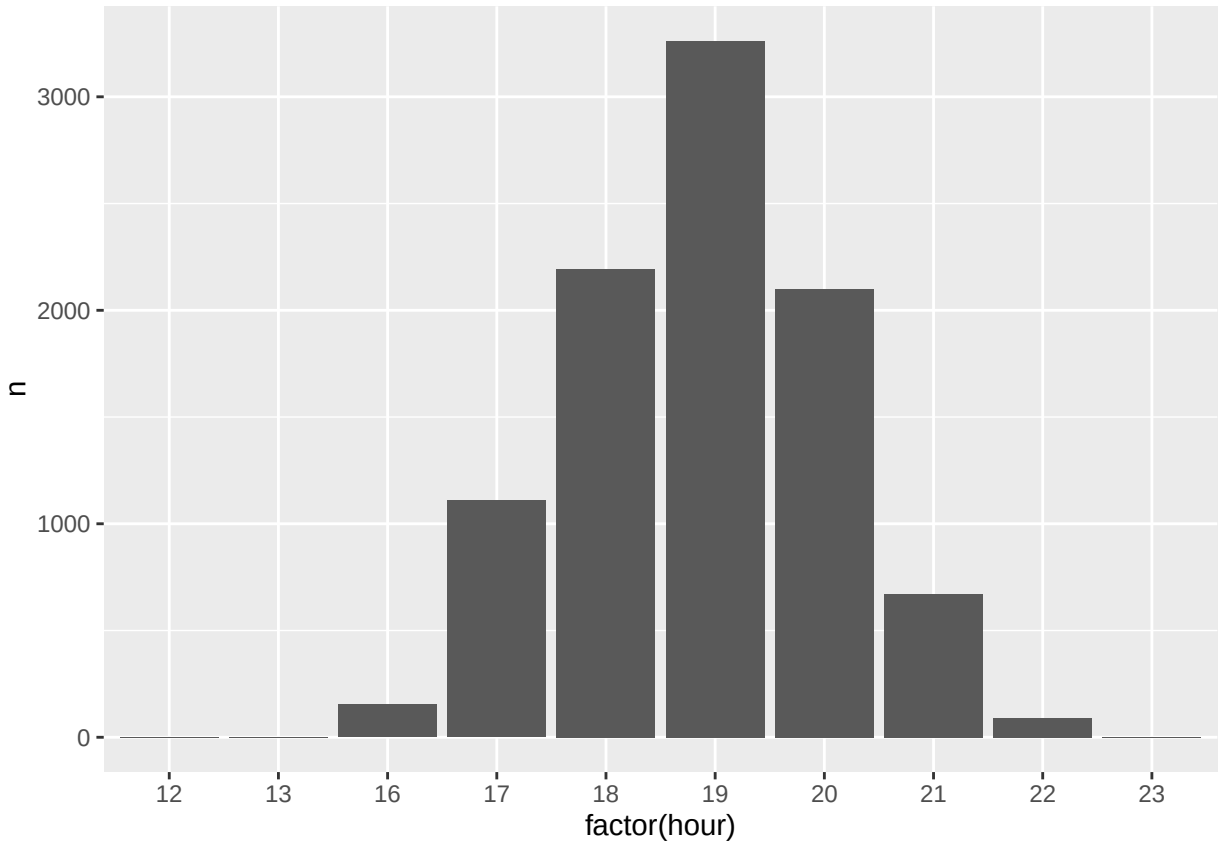
```
#First time users: Highest Revenue - 3, Lowest Revenue - 2

first.list <- highest_lowest(3,2,firstuser.final,"first")
list2env(first.list,environment())
```

```
## <environment: R_GlobalEnv>
```

```
#high
foodvswine.highfirst
```

```
##        food     wine   liquor
## 1: 938680.5 233212.2 108664.5
```

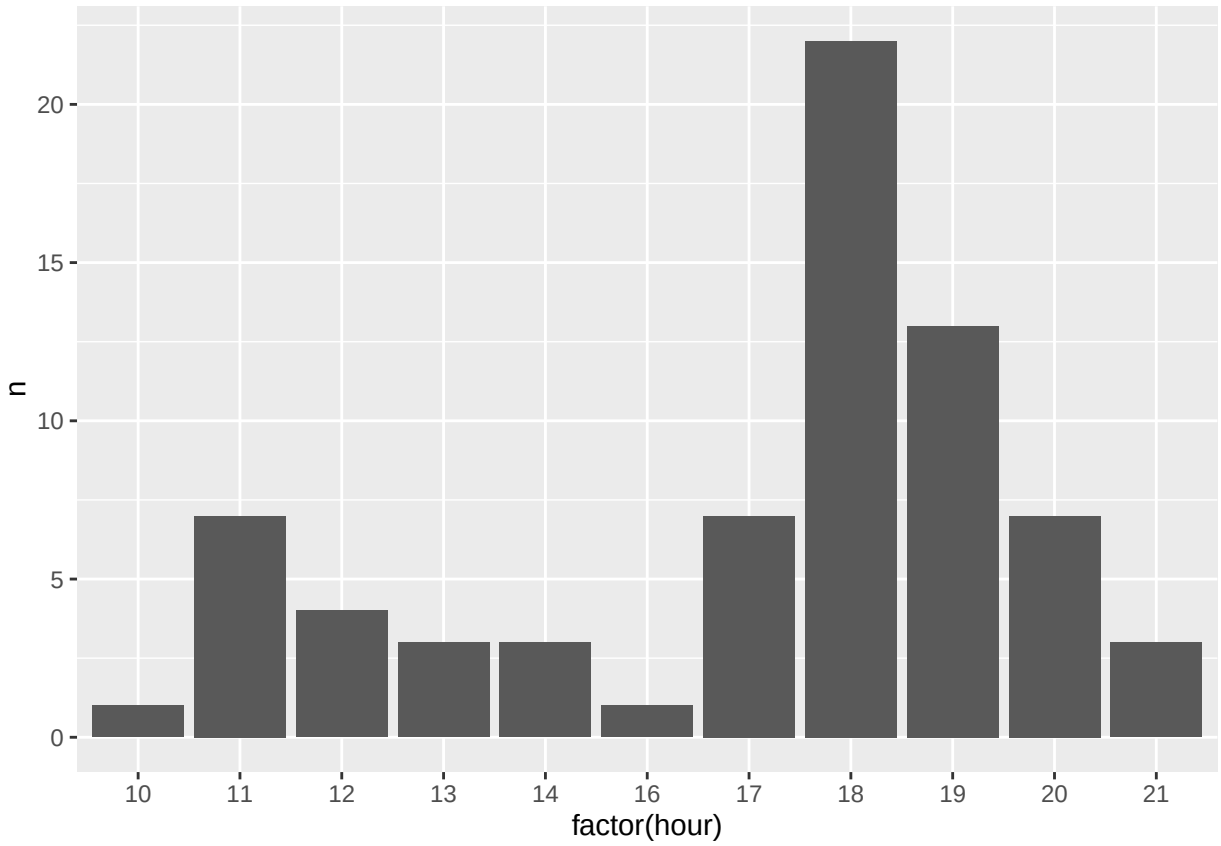
```
ggplot(data=time.highfirst,aes(x=factor(hour), y=n)) + geom_bar(stat="identity")
```




```
#low
foodvswine.lowfirst
```

```
##    food    wine liquor
## 1: 3337 21433.2 8048.4
```

```
ggplot(data=time.lowfirst,aes(x=factor(hour), y=n)) + geom_bar(stat="identity")
```



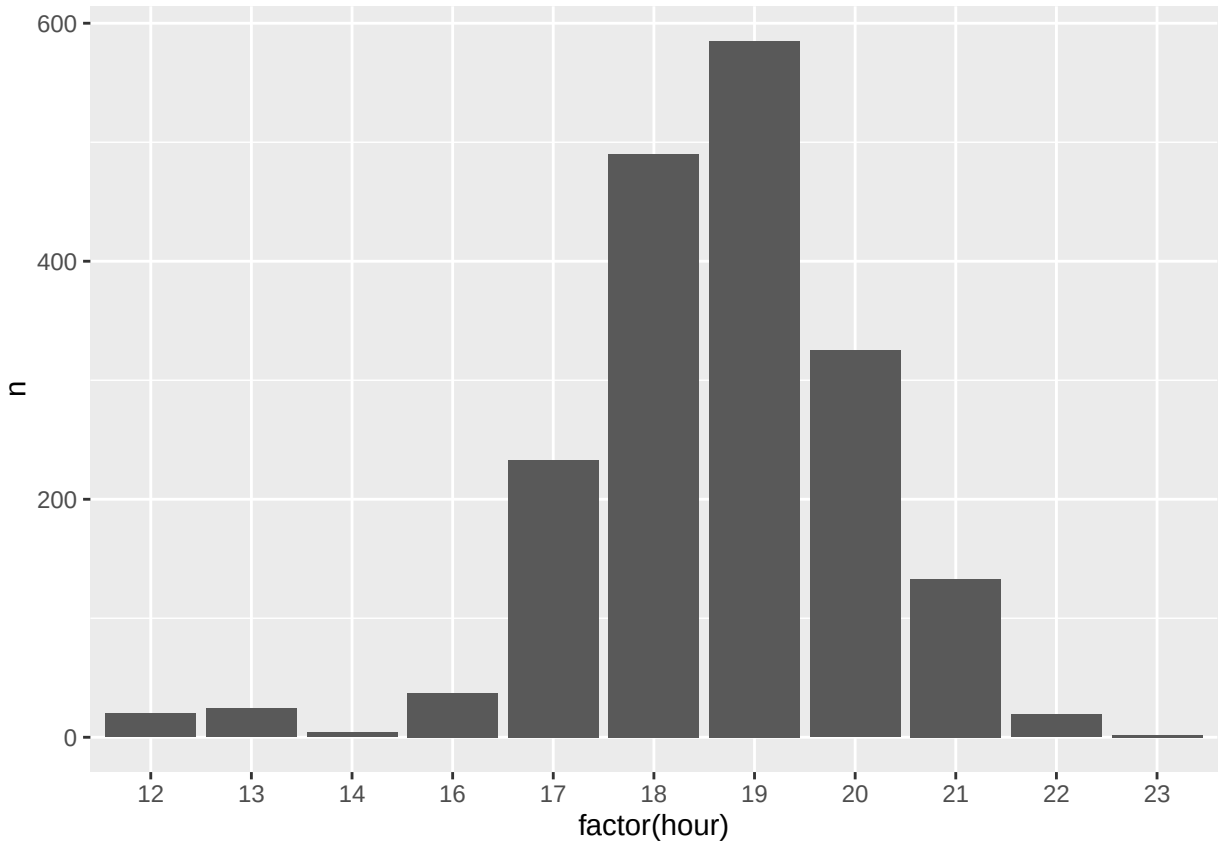

```
#Repeat users
repeat.list <- highest_lowest(6,3,repeat.final,"repeat")
list2env(repeat.list,environment())
```

```
## <environment: R_GlobalEnv>
```

```
#high
foodvswine.highrepeat
```

```
##        food     wine   liquor
## 1: 194819.7 53495.62 19781.02
```

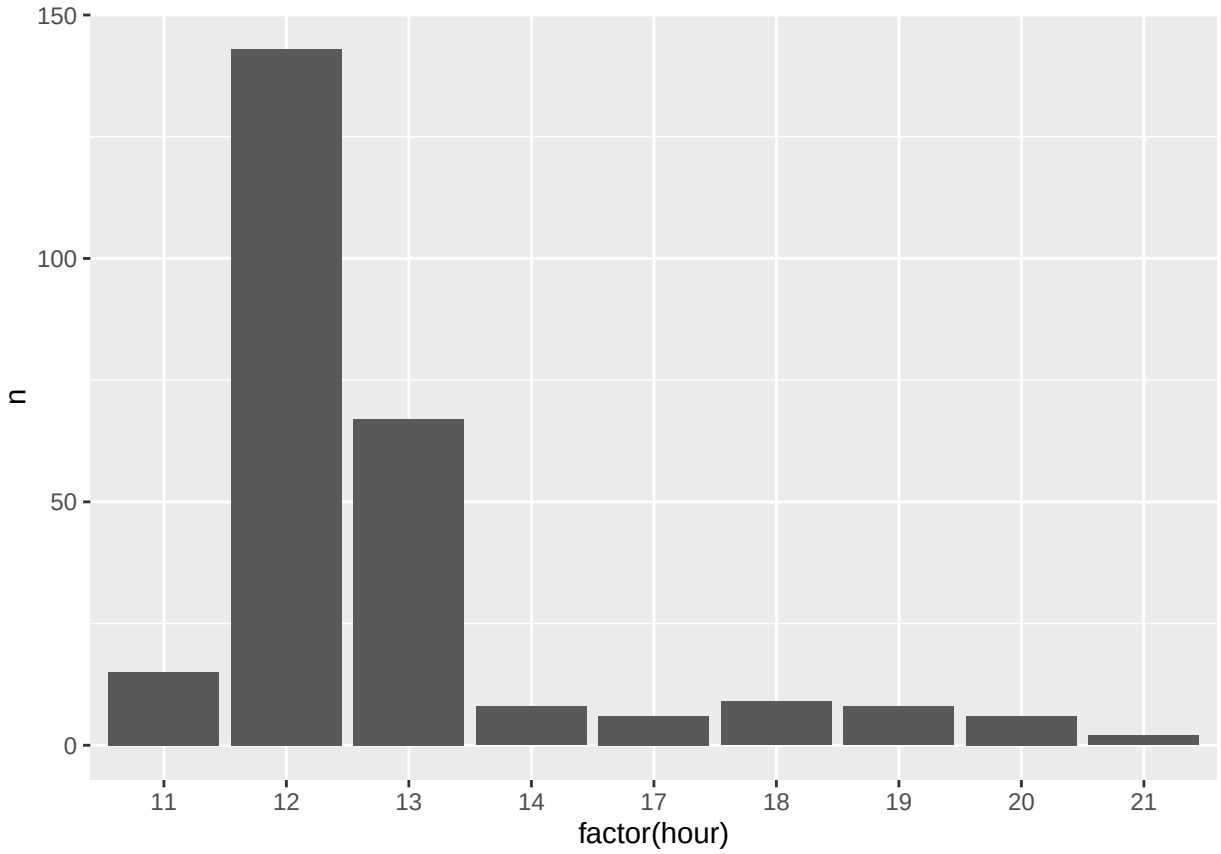
```
ggplot(data=time.highrepeat,aes(x=factor(hour), y=n)) + geom_bar(stat="identity")
```




```
#low
foodvswine.lowrepeat
```

```
##        food     wine   liquor
## 1: 19459.59 1525.289 555.3816
```

```
ggplot(data=time.lowrepeat,aes(x=factor(hour), y=n)) + geom_bar(stat="identity")
```



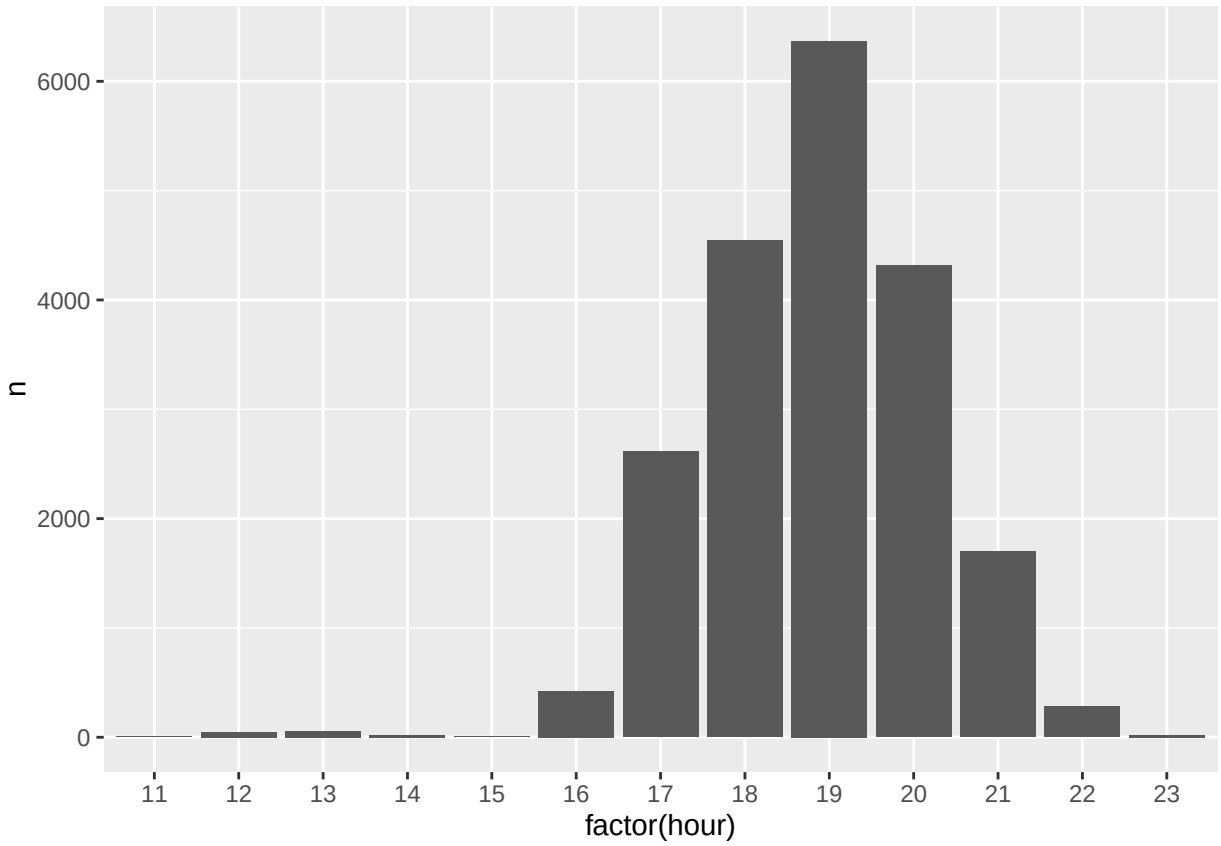

```
#All users
users.list <- highest_lowest(6,2,user.final,"all")
list2env(users.list,environment())
```

```
## <environment: R_GlobalEnv>
```

```
#high
foodvswine.highall
```

```
##       food     wine liquor
## 1: 2154261 571080.9 242364
```

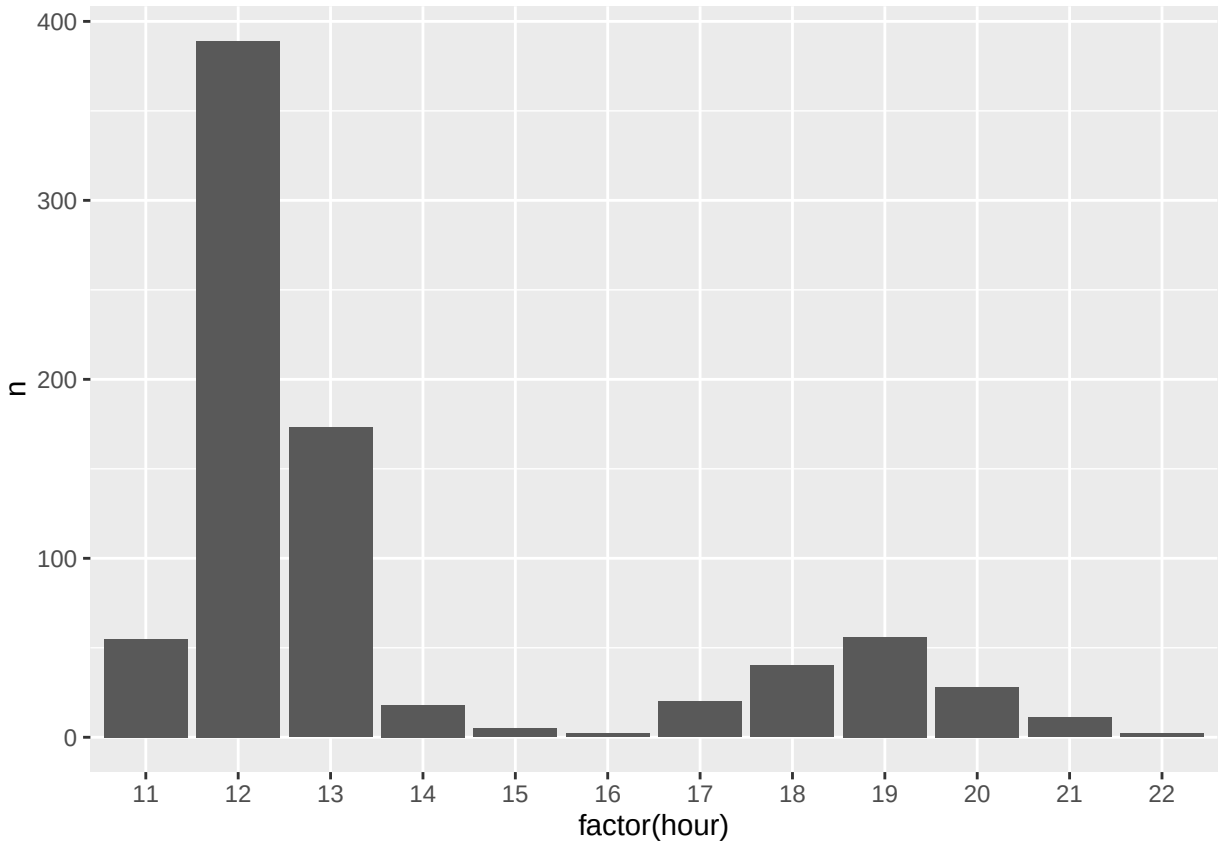
```
ggplot(data=time.highall,aes(x=factor(hour), y=n)) + geom_bar(stat="identity")
```




```
#low
foodvswine.lowall
```

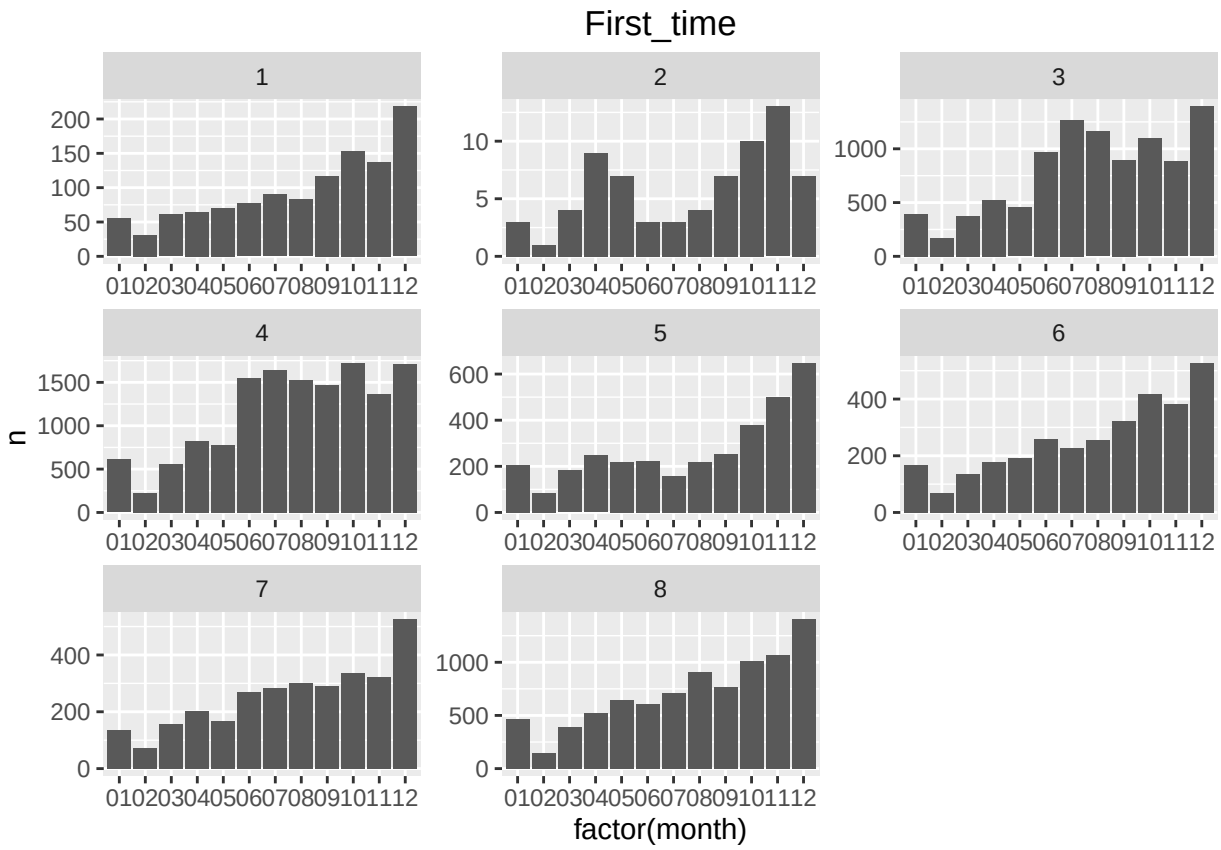
```
##     food     wine   liquor
## 1: 63624 7096.773 2923.327
```

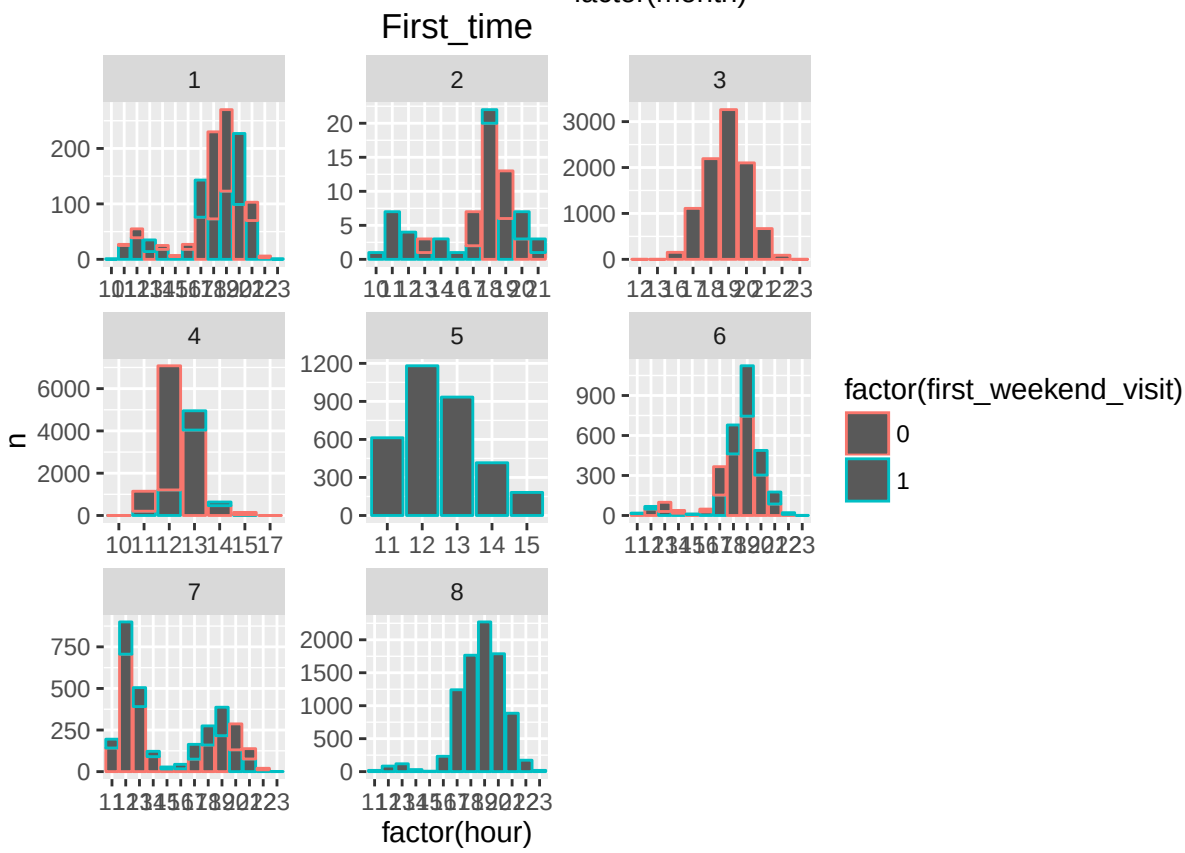
```
ggplot(data=time.lowall,aes(x=factor(hour), y=n)) + geom_bar(stat="identity")
```



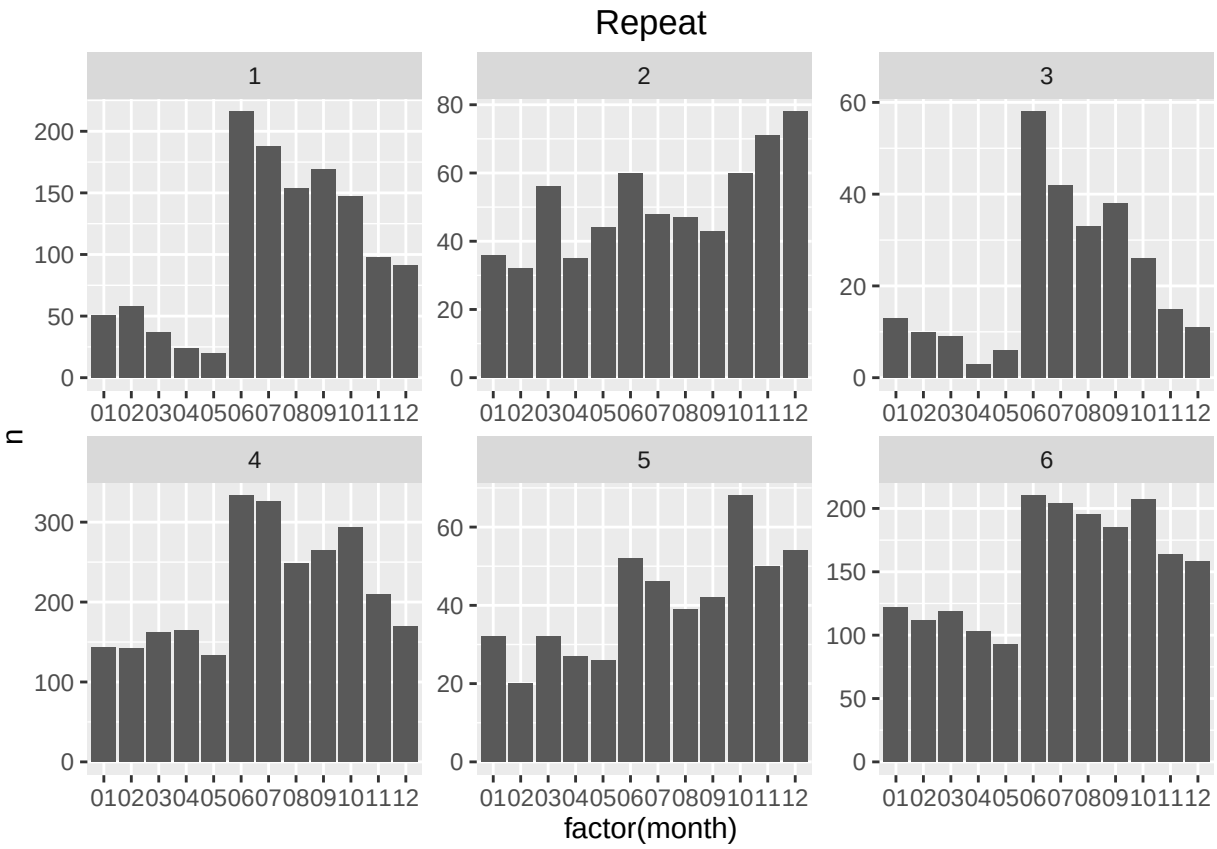
Look for seasonality and each cluster by hour, each graph represents a cluster

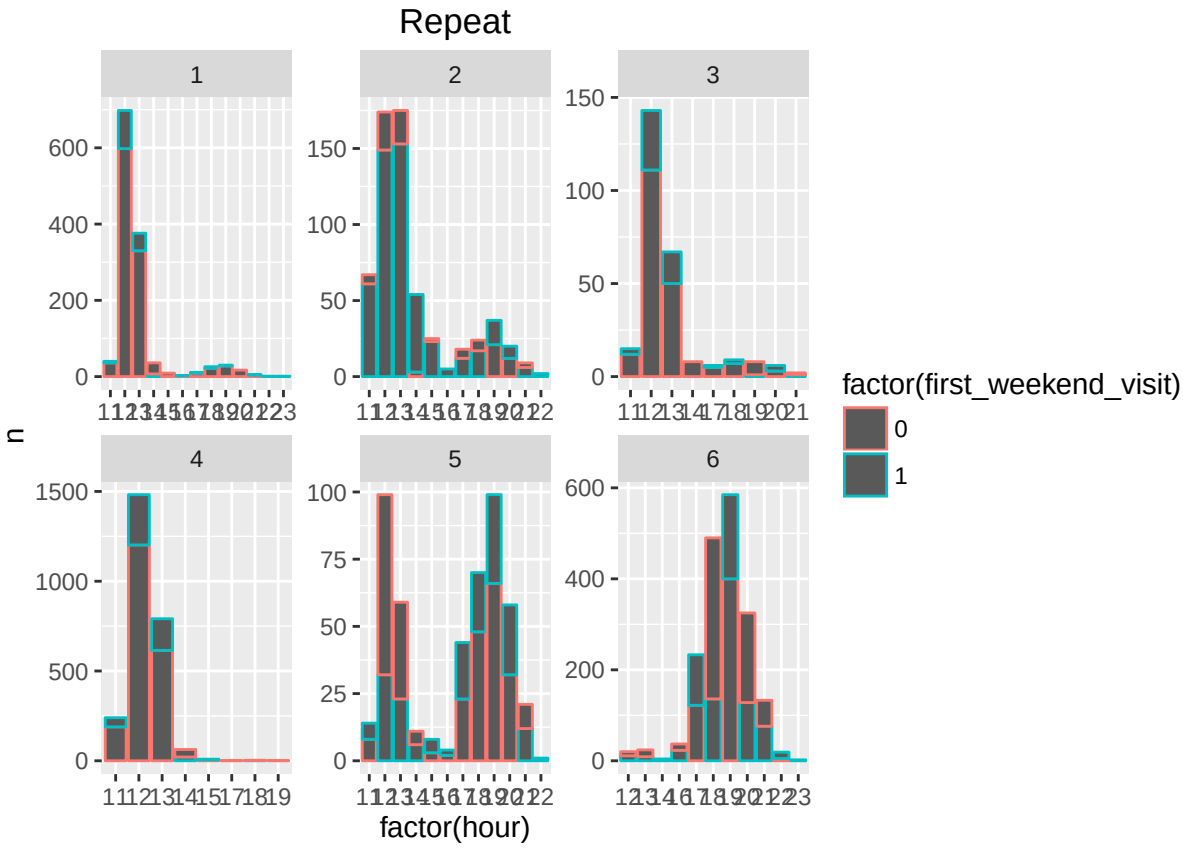
```
#First time
month_hour(firstuser.final,"First_time")
```


First_time
1
2
3
4
5
6
7
8
n
factor(month)

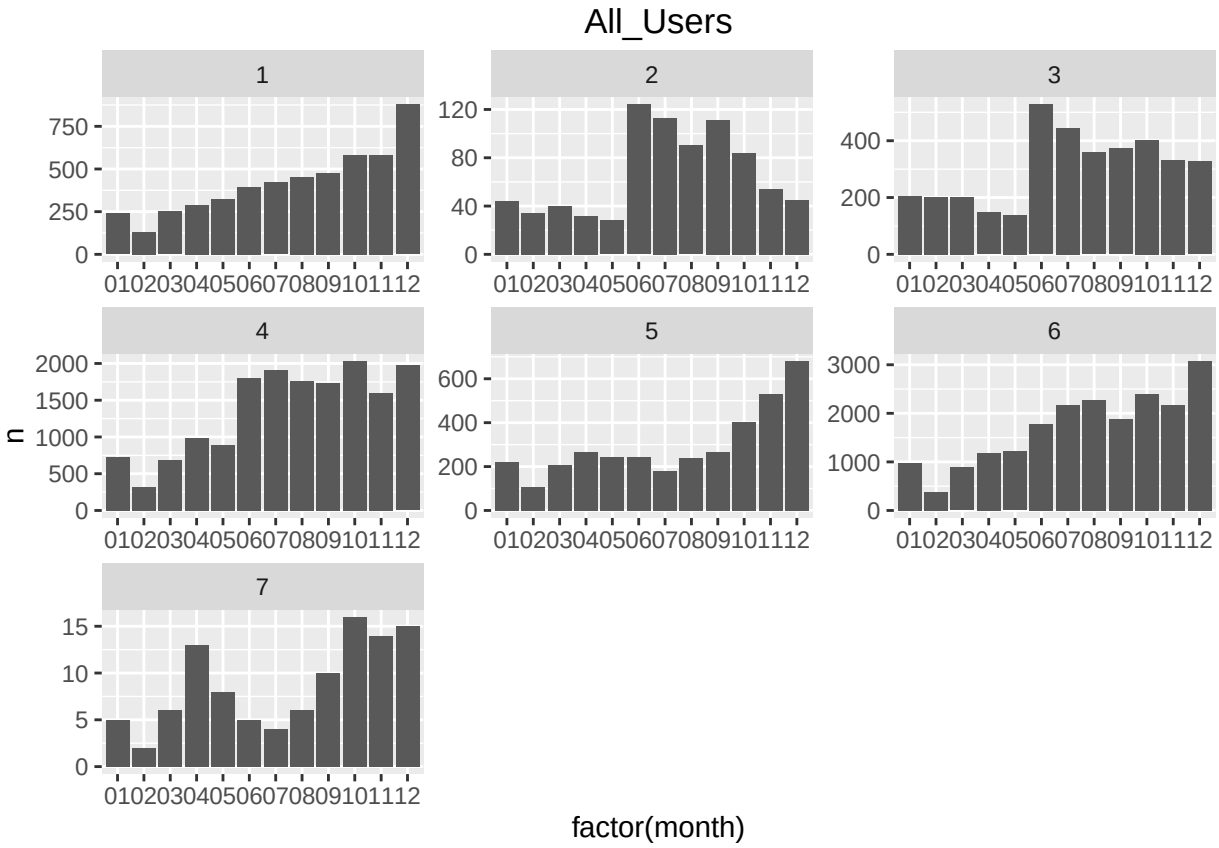

First_time
1
2
3
4
5
6
7
8
n
factor(hour)
factor(first_weekend_visit)
0
1

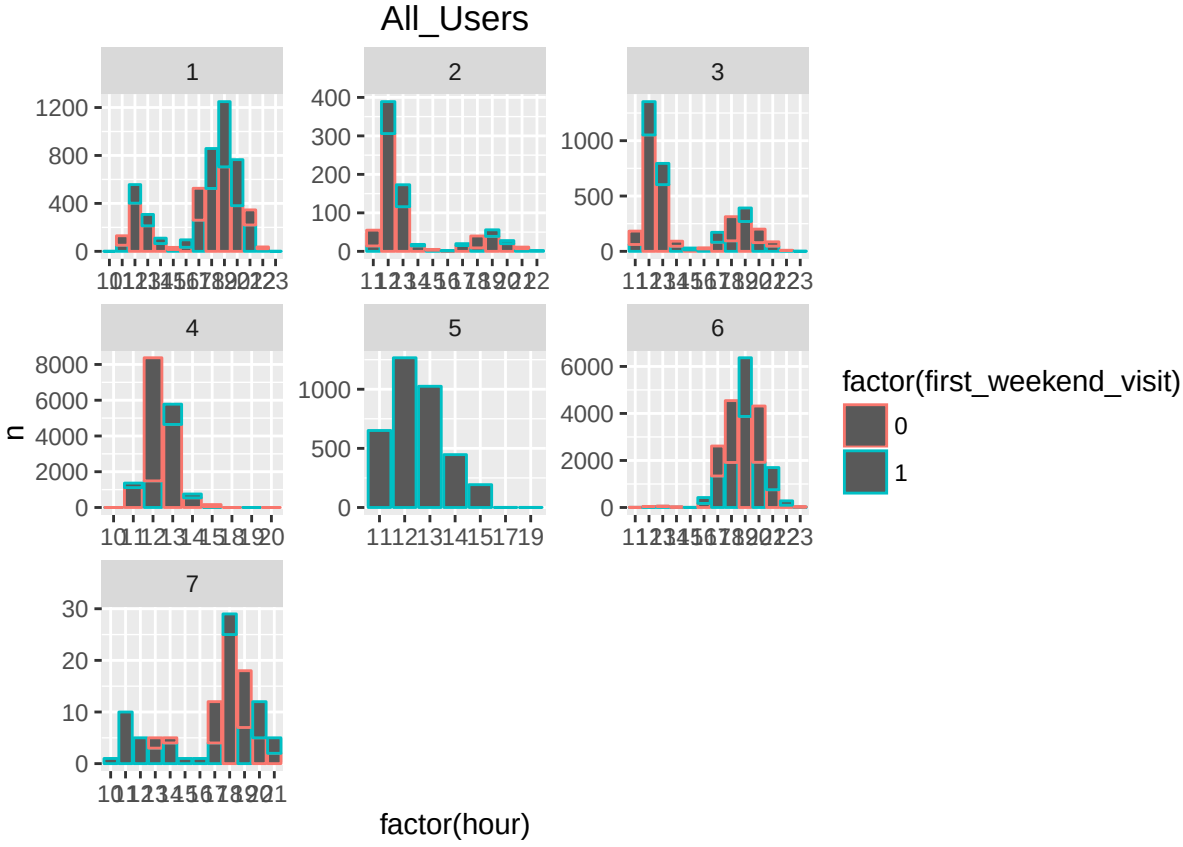
```
#Repeat
month_hour(repeat.final, "Repeat")
```






```
#All
month_hour(user.final, "All_Users")
```

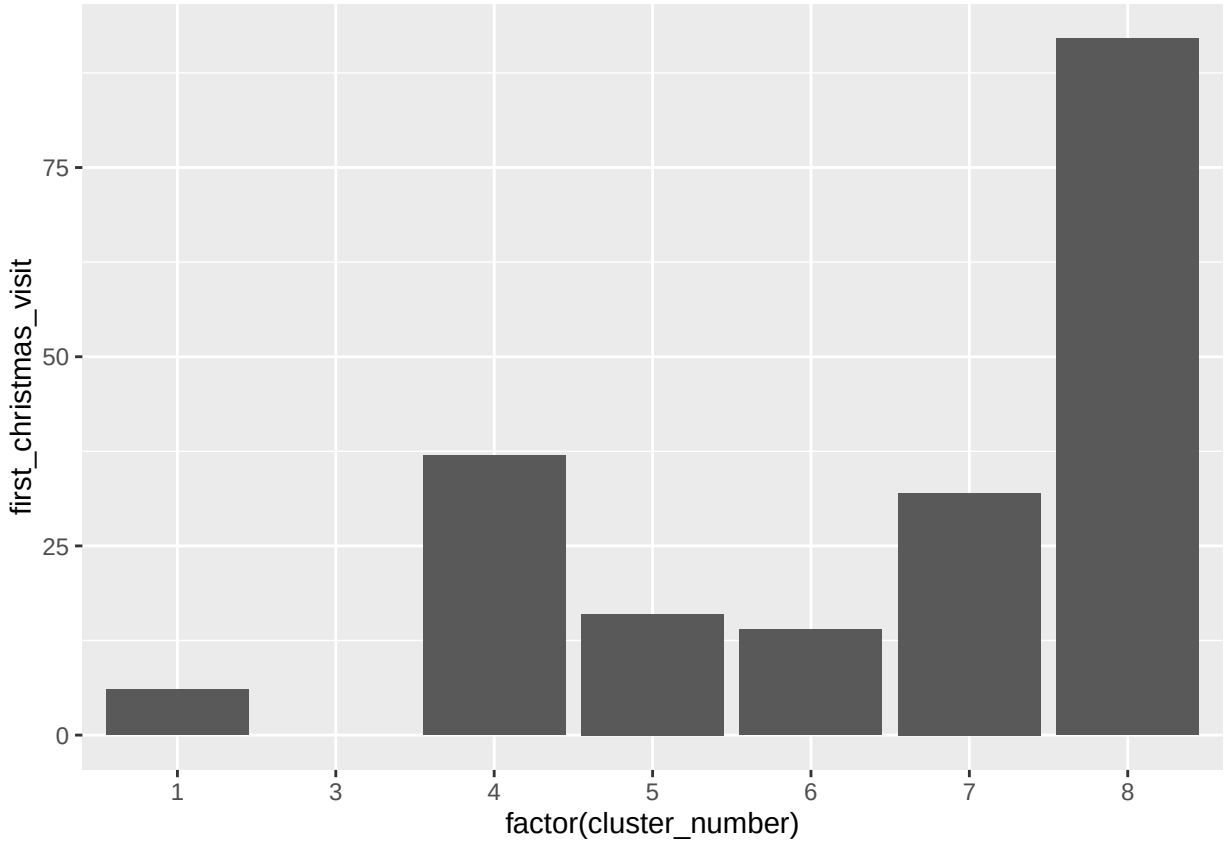

All_Users
1
2
3
4
5
6
7
n
factor(month)


All_Users
1
2
3
4
5
6
7
n
factor(hour)
factor(first_weekend_visit)
0
1

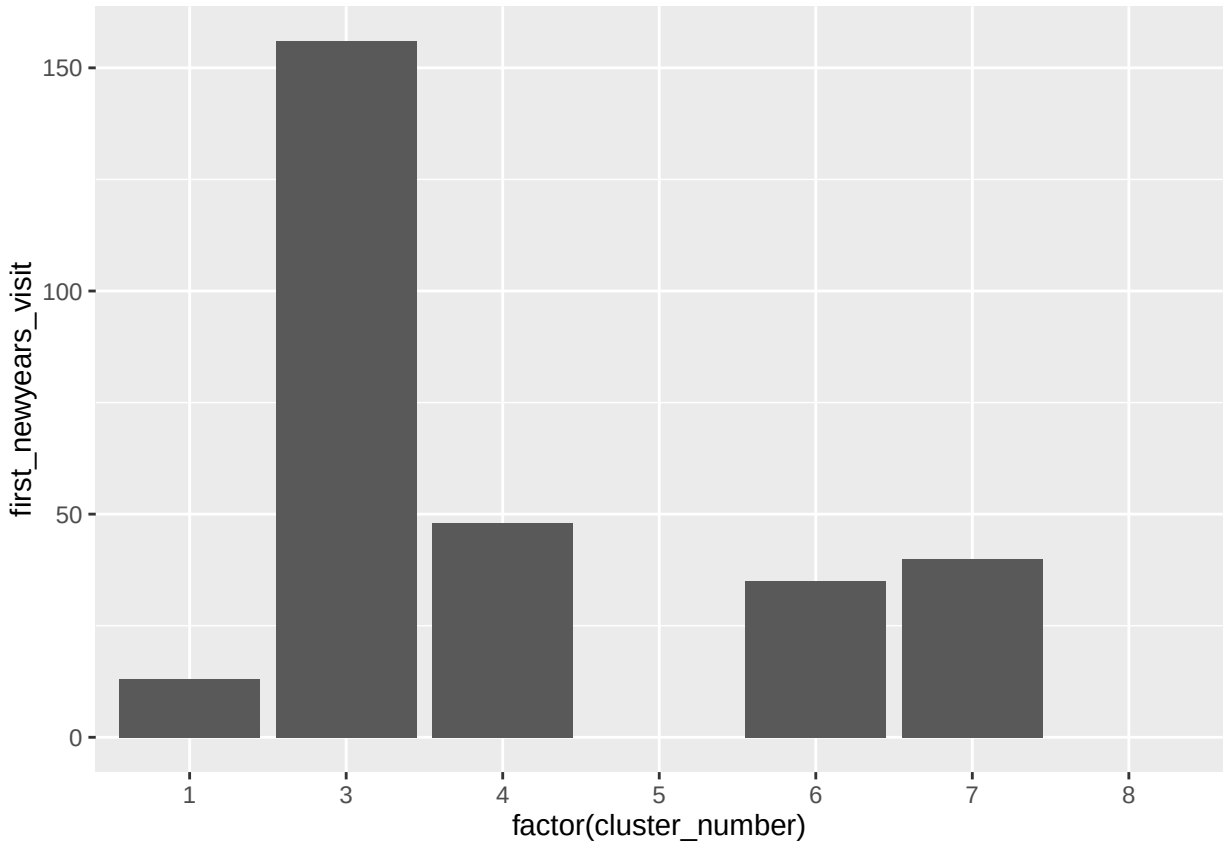
Holiday Behavior

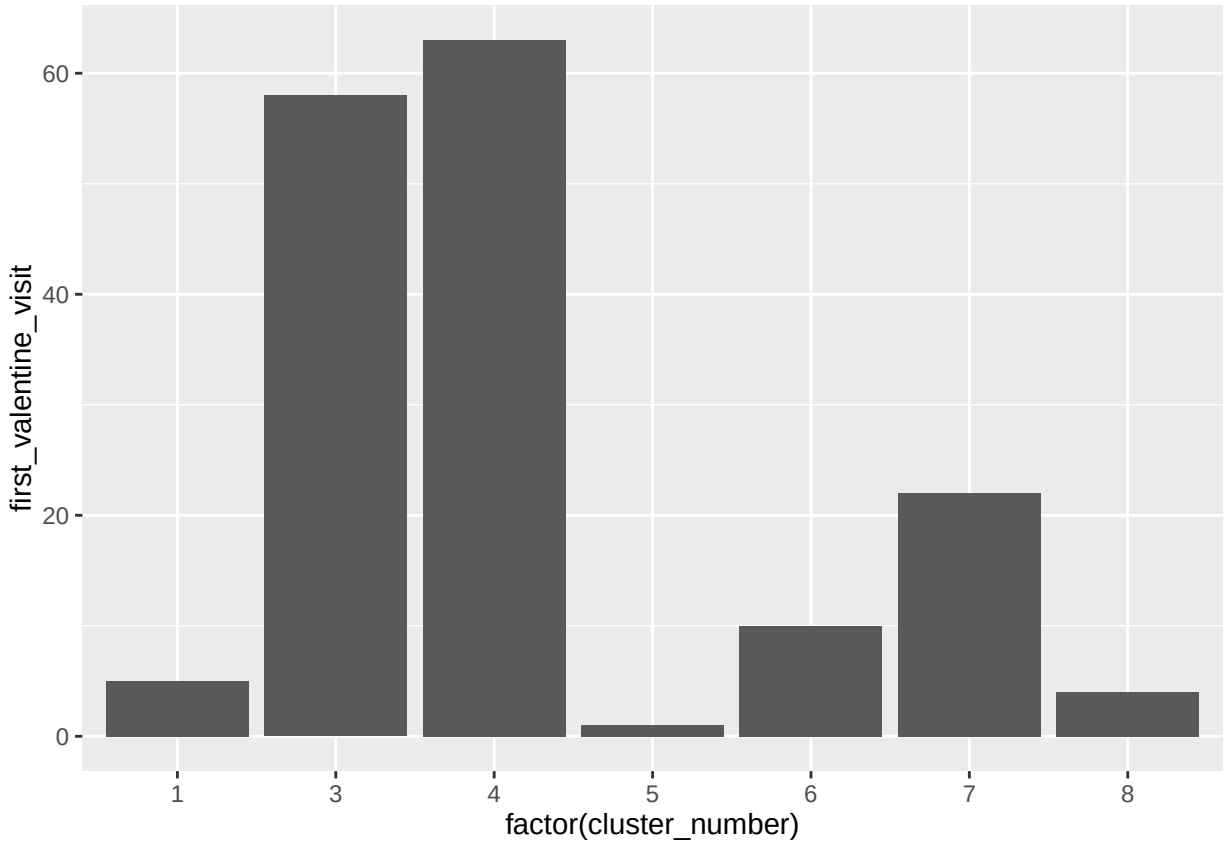
```
#First time
holidays(firstuser.final)
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```

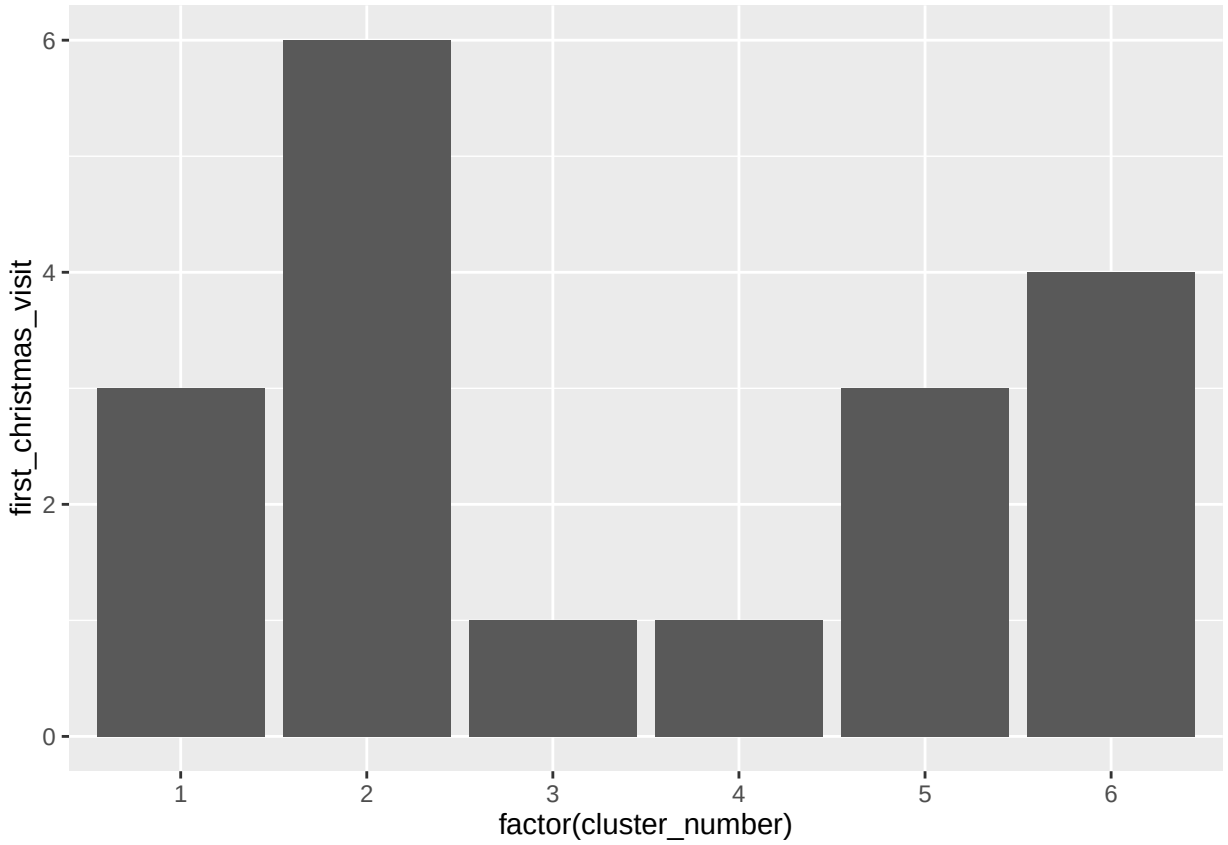



```
## Warning: Removed 2 rows containing missing values (position_stack).
```

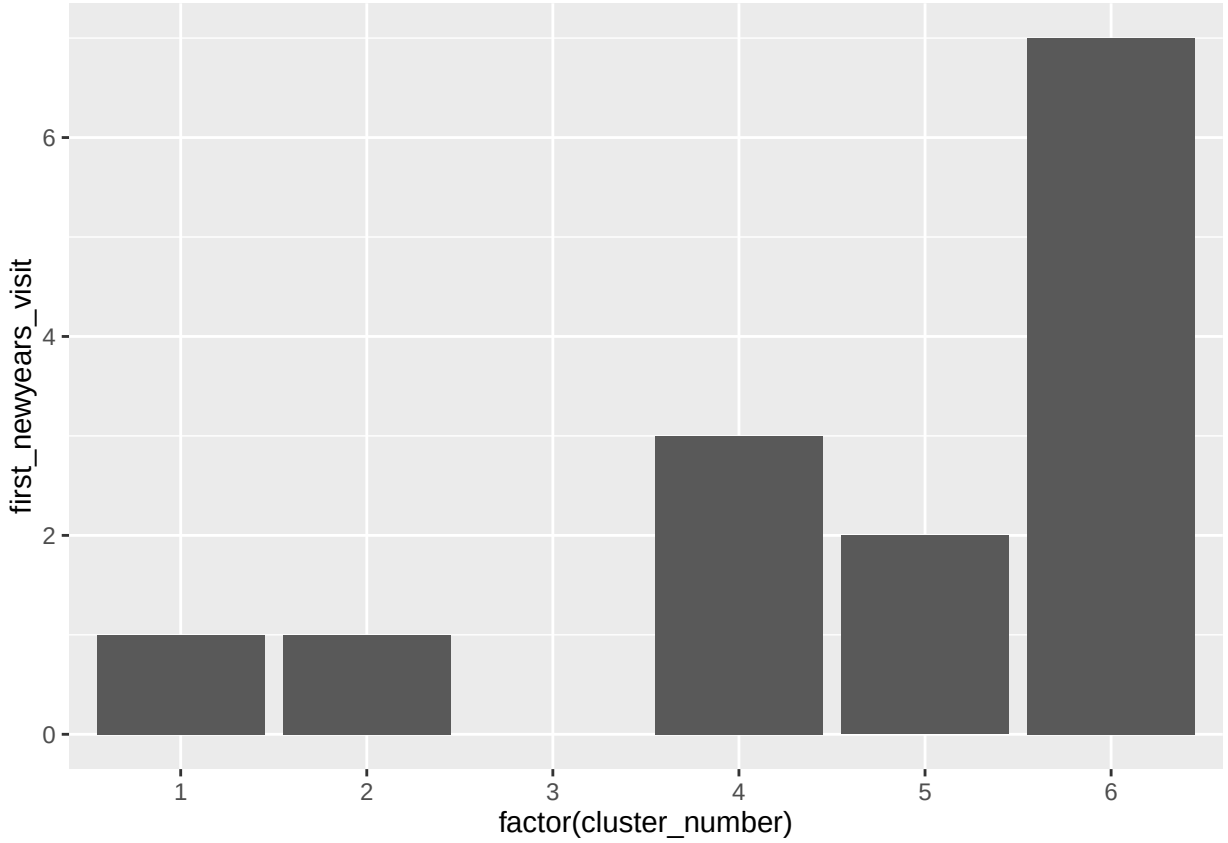

150
100
50
0
first_newyears_visit
1
3
4
5
6
7
8
factor(cluster_number)

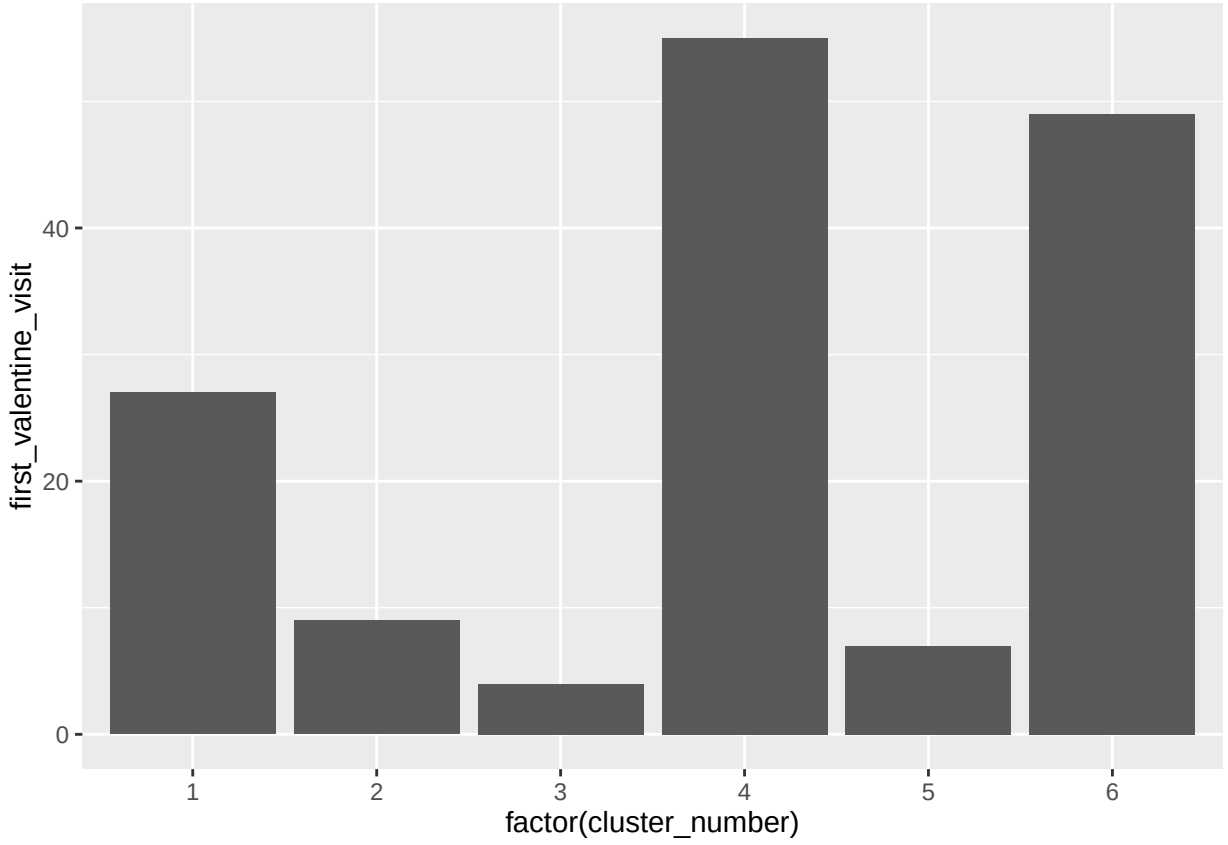

60
40
20
0
first_valentine_visit
1
3
4
5
6
7
8
factor(cluster_number)

```
#Repeat
holidays(repeat.final)
```



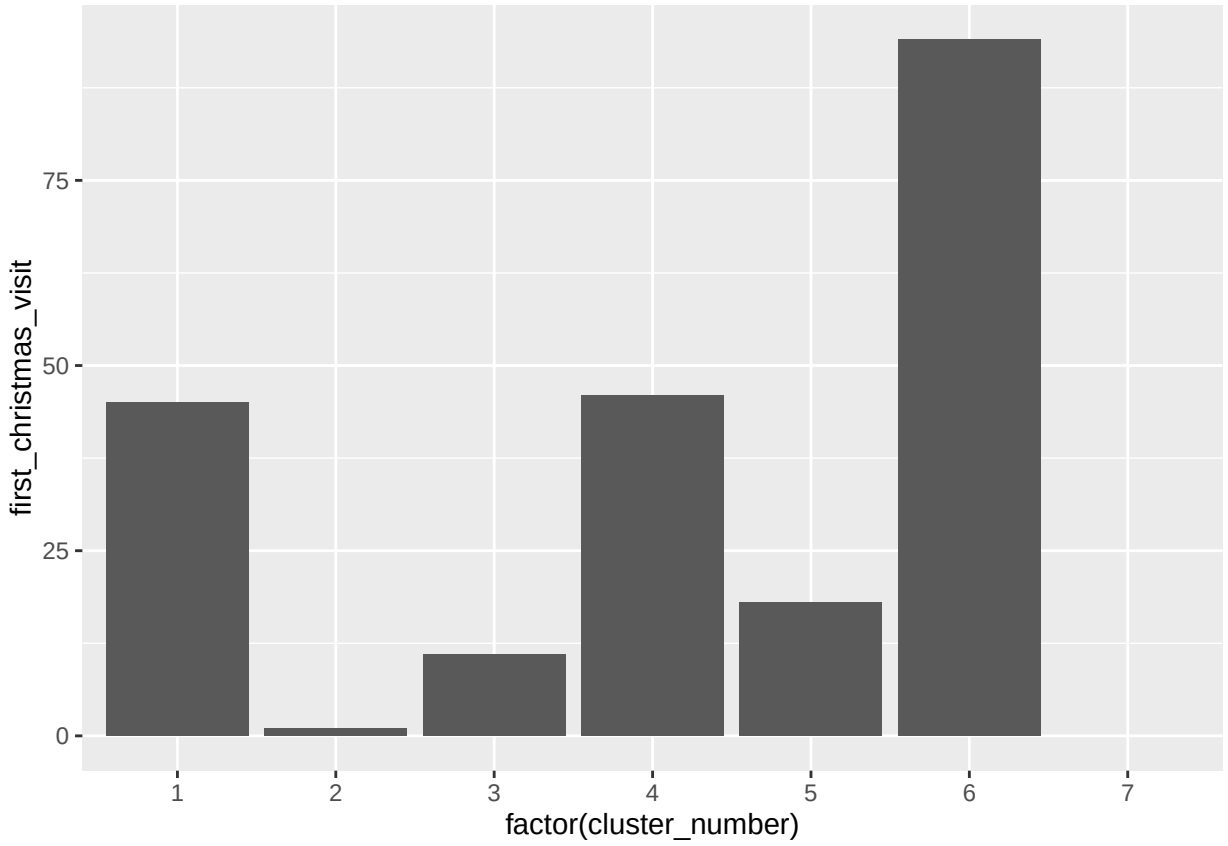

```
## Warning: Removed 1 rows containing missing values (position_stack).
```


first_newyears_visit
0
2
4
6
1
2
3
4
5
6
factor(cluster_number)

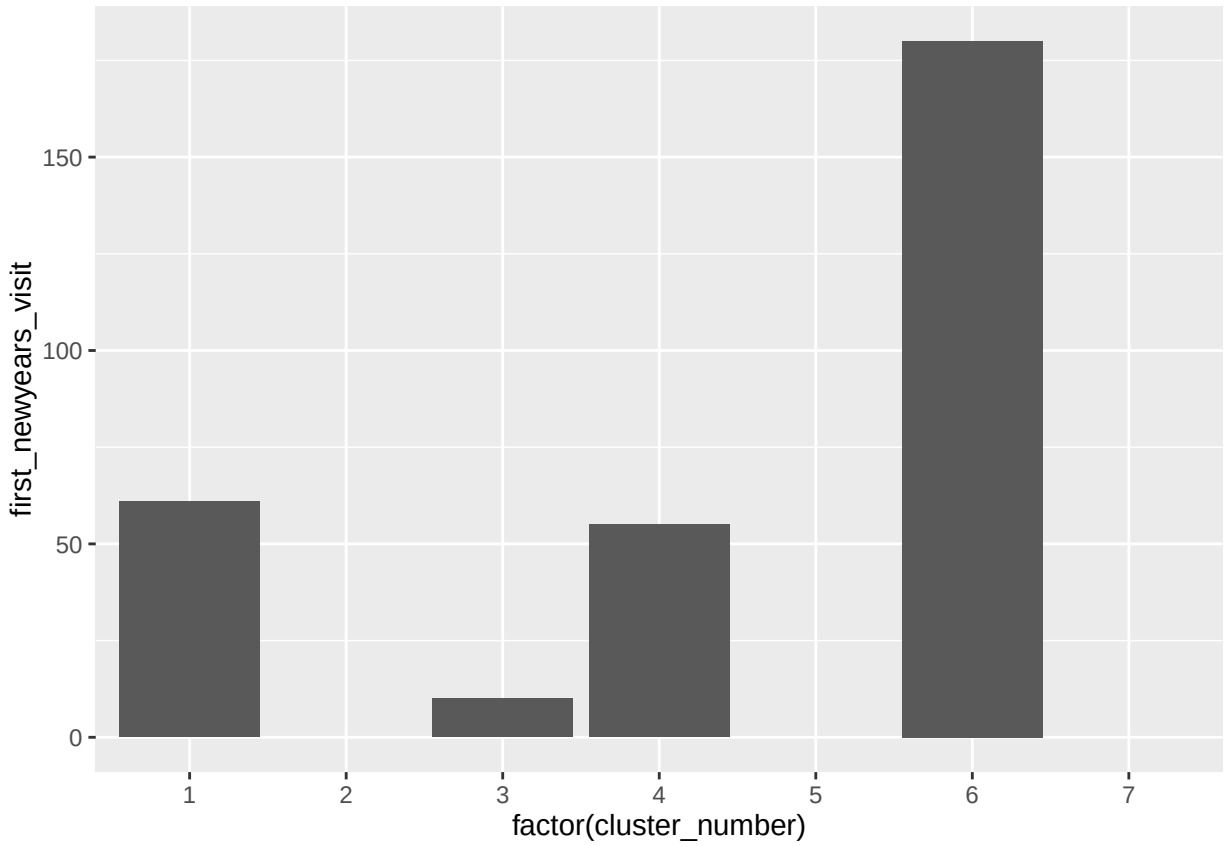

first_valentine_visit
0
20
40
1
2
3
4
5
6
factor(cluster_number)

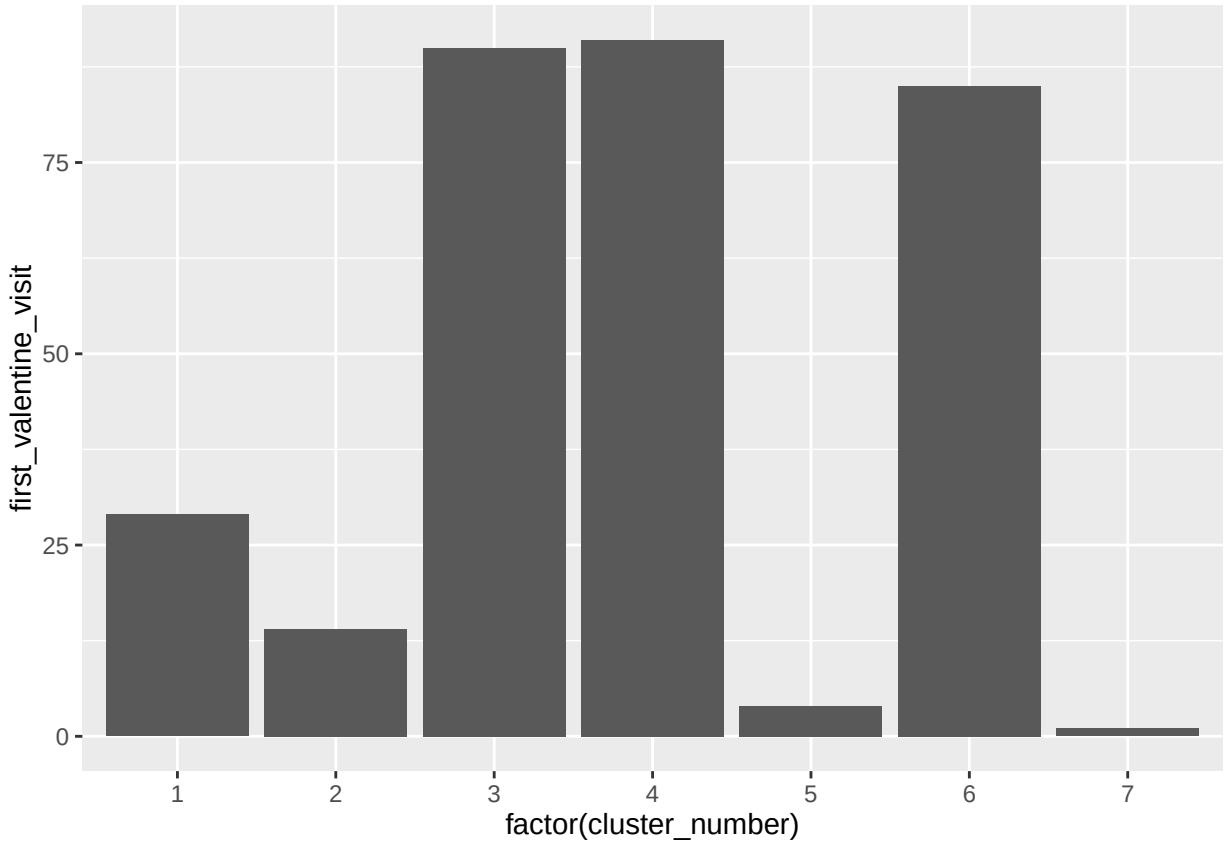
```
#All
holidays(user.final)
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



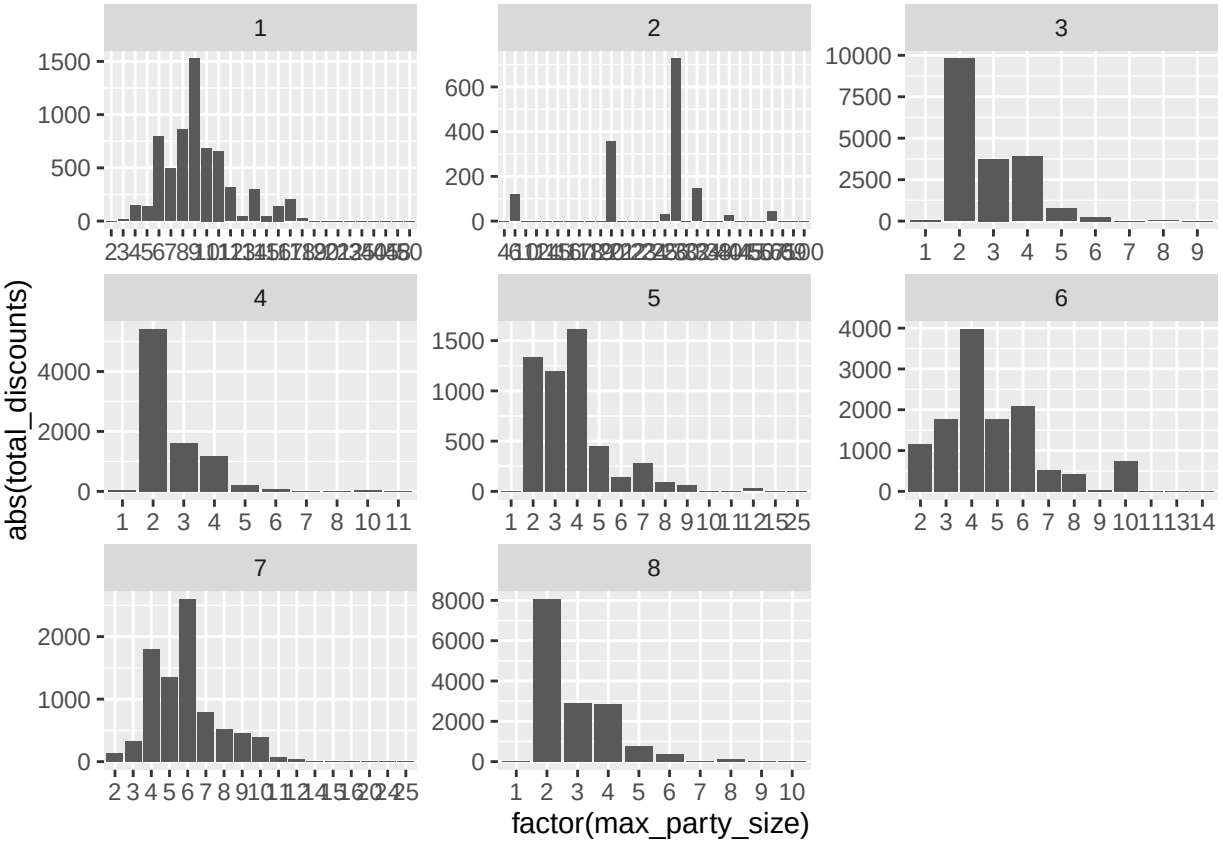

```
## Warning: Removed 3 rows containing missing values (position_stack).
```

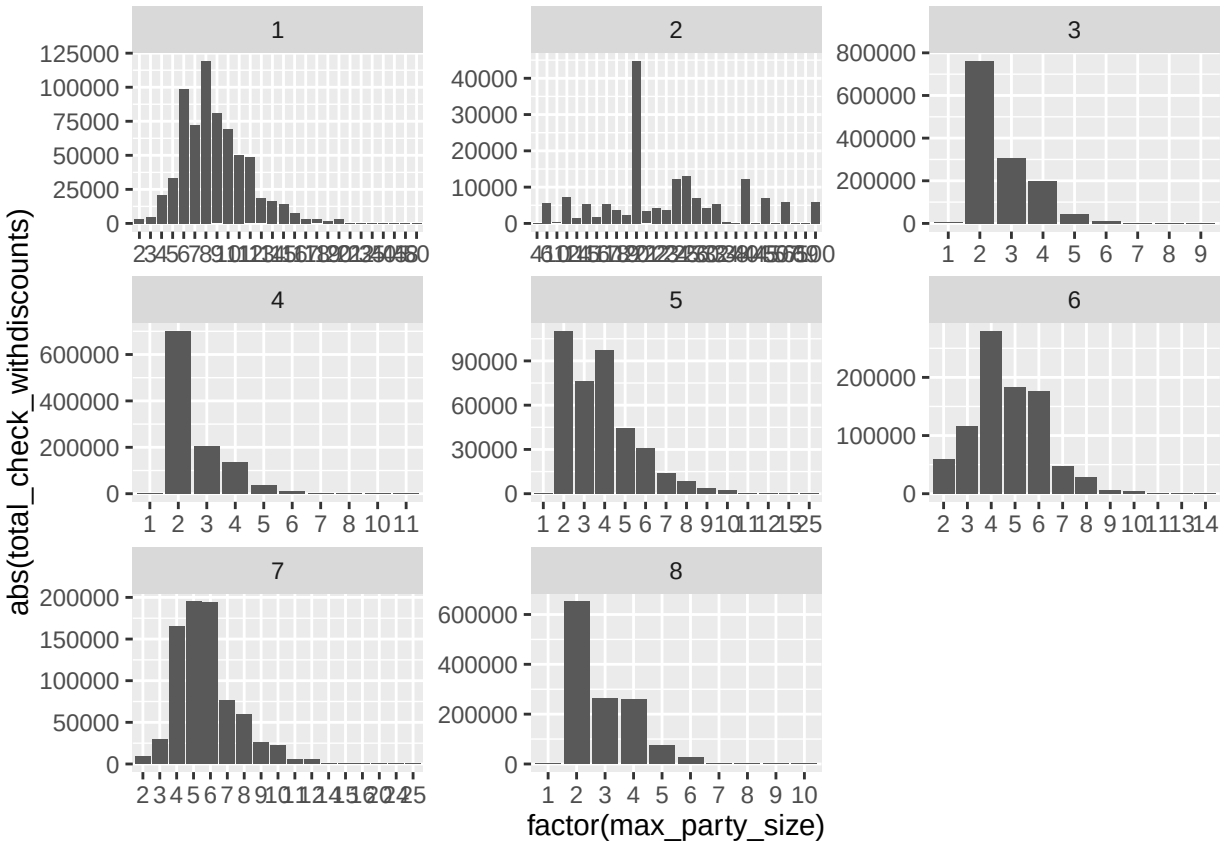

first_newyears_visit
150
100
50
0
1
2
3
4
5
6
7
factor(cluster_number)


first_valentine_visit
75
50
25
0
1
2
3
4
5
6
7
factor(cluster_number)

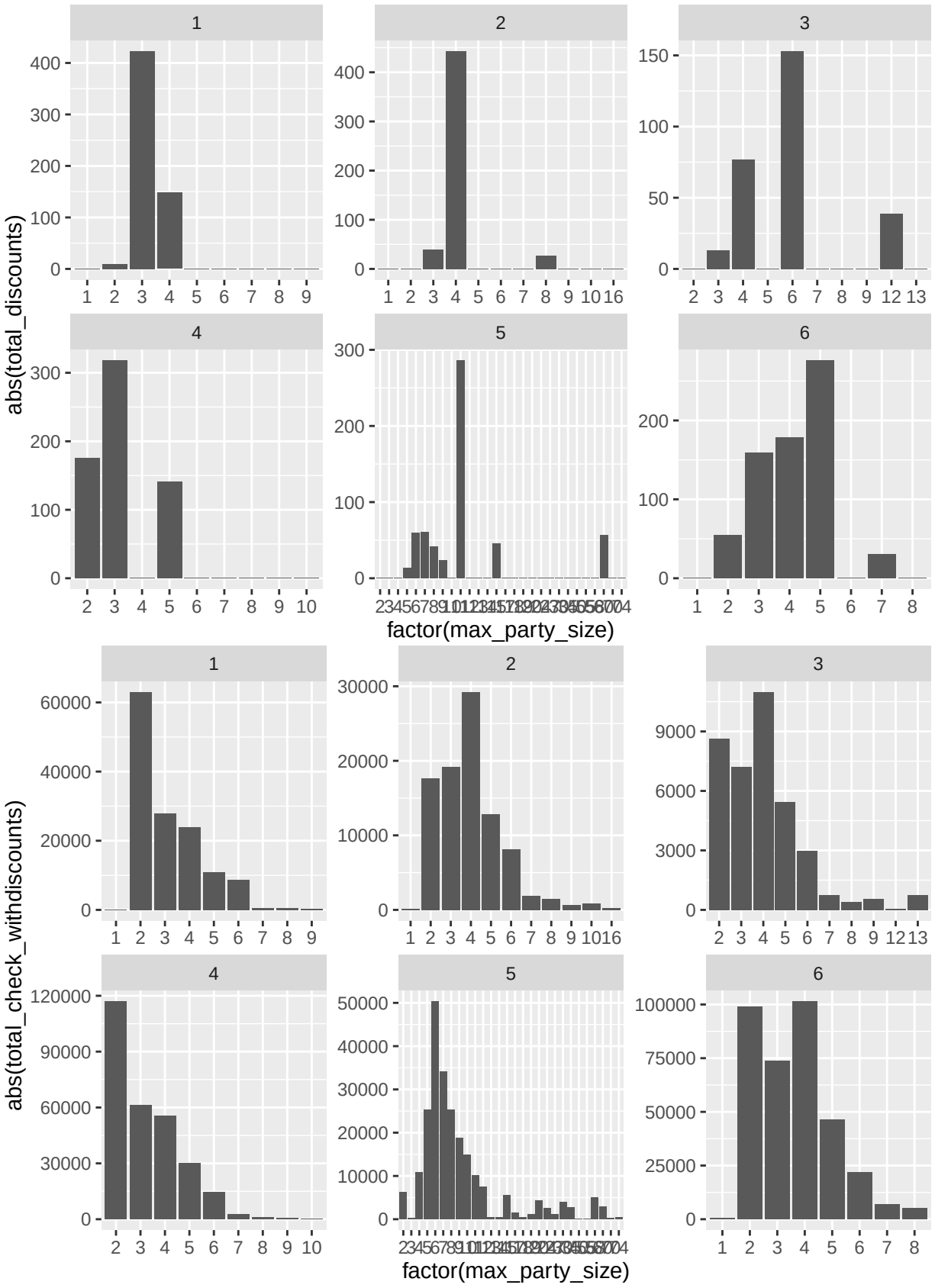
Discounts

```
#First time
discounts(firstuser.final)
```

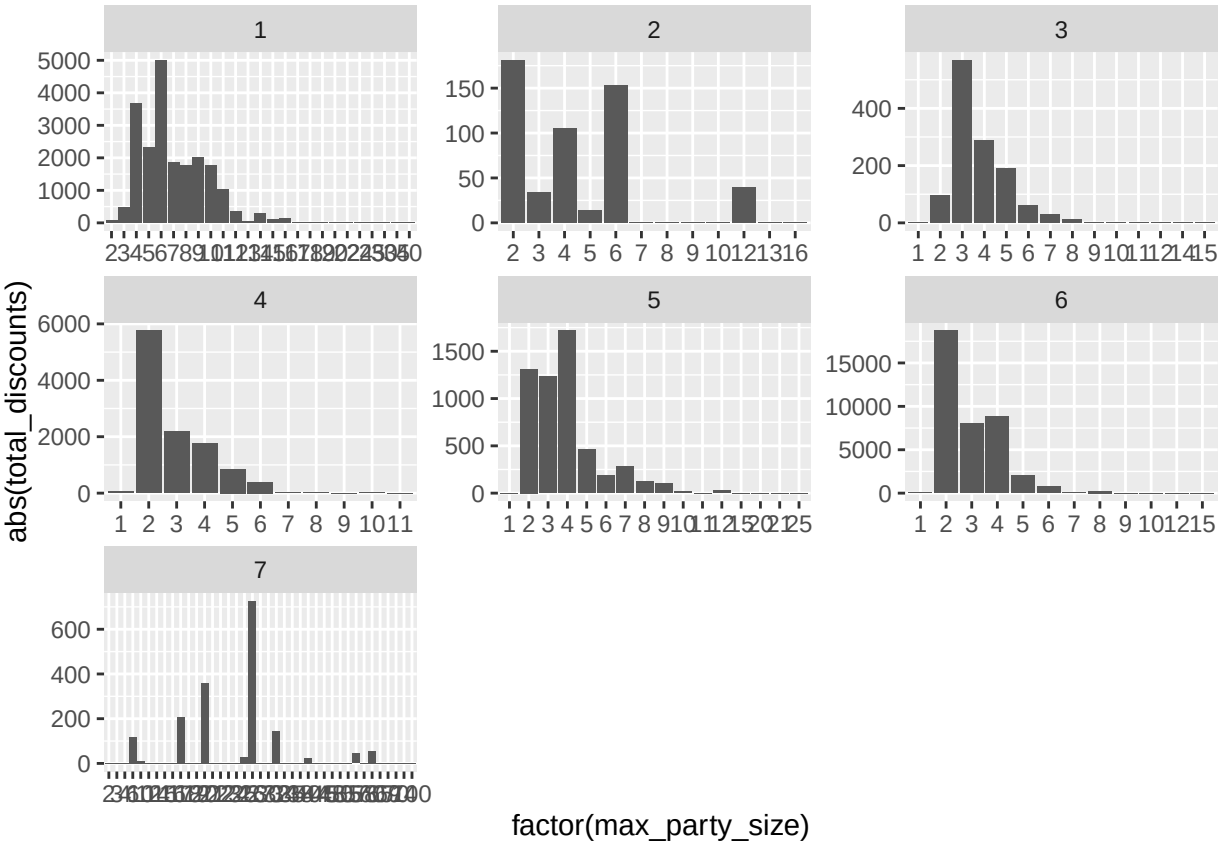


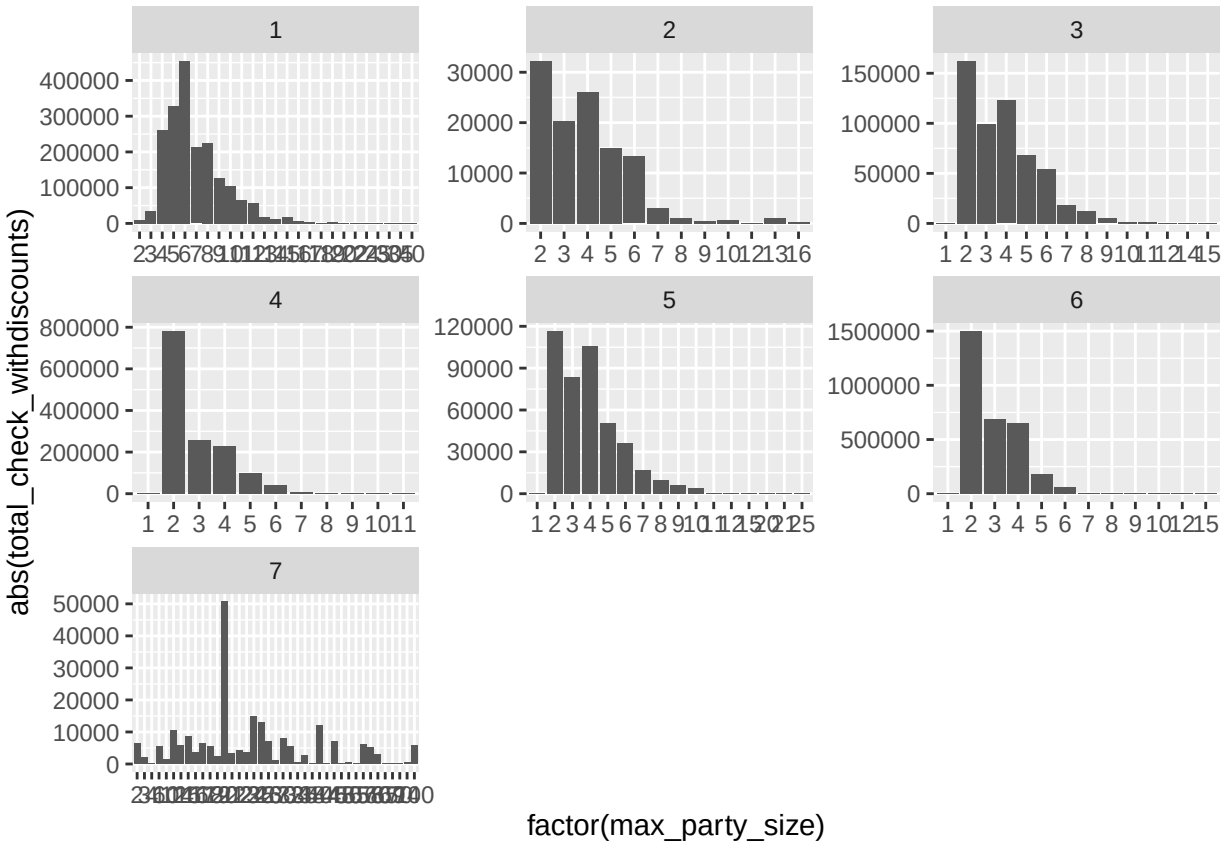



```
#Repeat
discounts(repeat.final)
```


1
2
3
4
5
6
abs(total_discounts)
factor(max_party_size)
abs(total_check_withdiscounts)
factor(max_party_size)

```
#All
discounts(user.final)
```




1
2
3
4
5
6
7
abs(total_check_withdiscounts)
factor(max_party_size)