Sticks Kabob

Yuwen Hsieh

October 24, 2018

Initialization:

- Data import
- Data Cleaning

```
if(!("tidyverse" %in% (.packages()))){library(tidyverse)}
if(!("readxl" %in% (.packages()))){library(readxl)}
df <- read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\km5_main.xlsx")
{
trans_df=function(df,clm,new_name,ex=NA){
  tdf=df[df[,clm]=='1',]
  tdf[,clm]=colnames(tdf)[clm]
  colnames(tdf)[clm]=new_name
  if (is.na(ex)){return(tdf[,c('ID',new_name,'customer_status','group')])}
  else {return(tdf[,c('ID',ex,new_name,'customer_status','group')])}
}
clean_up=function(df,clms,nn,extra=NA){
  tdf=trans_df(df,clms[1],nn,extra)
  for (i in clms[-1]){
  tdf=rbind(tdf,trans_df(df,i,nn,extra))
  }
  return(map_df(tdf,factor))
}
}
theme_update(plot.background = element_rect(fill = "gold"),panel.background = element_rect(fill = "azur
df$Household_Income=factor(df$Household_Income,levels = c('$50k-','$50k~$100k','$100k+'))
df$ID=as.character(df$ID)
df[,c(-1,-8)]=map_df(df[c(-1,-8)],factor)
Activity=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\Activity.xlsx")
Activity=clean_up(Activity,2:18,'Activity')
Profession=read_csv("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\profession.csv")
Profession$ID=as.character(Profession$ID)
Profession$work=factor(Profession$work)
Profession$group=factor(Profession$group)
Profession$customer_status=factor(Profession$customer_status)
visit_reason=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\visit_reason.xlsx")
visit_reason=map_df(visit_reason,factor)
lunch=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\lunch.xlsx")
lunch=map_df(lunch,factor)
important=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\important.xlsx")
important=map_df(important,factor)
house=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\house.xlsx")
clean_house=function(x){return(ifelse(x %in% as.character(0:4),x,NA))}
house[,3:5]=map_df(house[,3:5],clean_house)
house=map_df(house,factor)
coupons=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\coupons.xlsx")
coupons$other=as.numeric(!(is.na(coupons$other)))
```

```
coupons=clean_up(coupons,3:8,nn = 'source',extra = 'coupons_used')
Compare=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\compare.xlsx")
colnames(Compare)[2]='Convenient_place_to_eat'
Compare$other=ifelse(is.na(Compare$other),NA,'other')
Compare$ID=as.character(Compare$ID)
Compare[,-1]=map_df(Compare[,-1],factor)
child_act=read_excel("C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\child_act.xlsx")
child_act=clean_up(child_act,2:11,'Children_Activity')
```

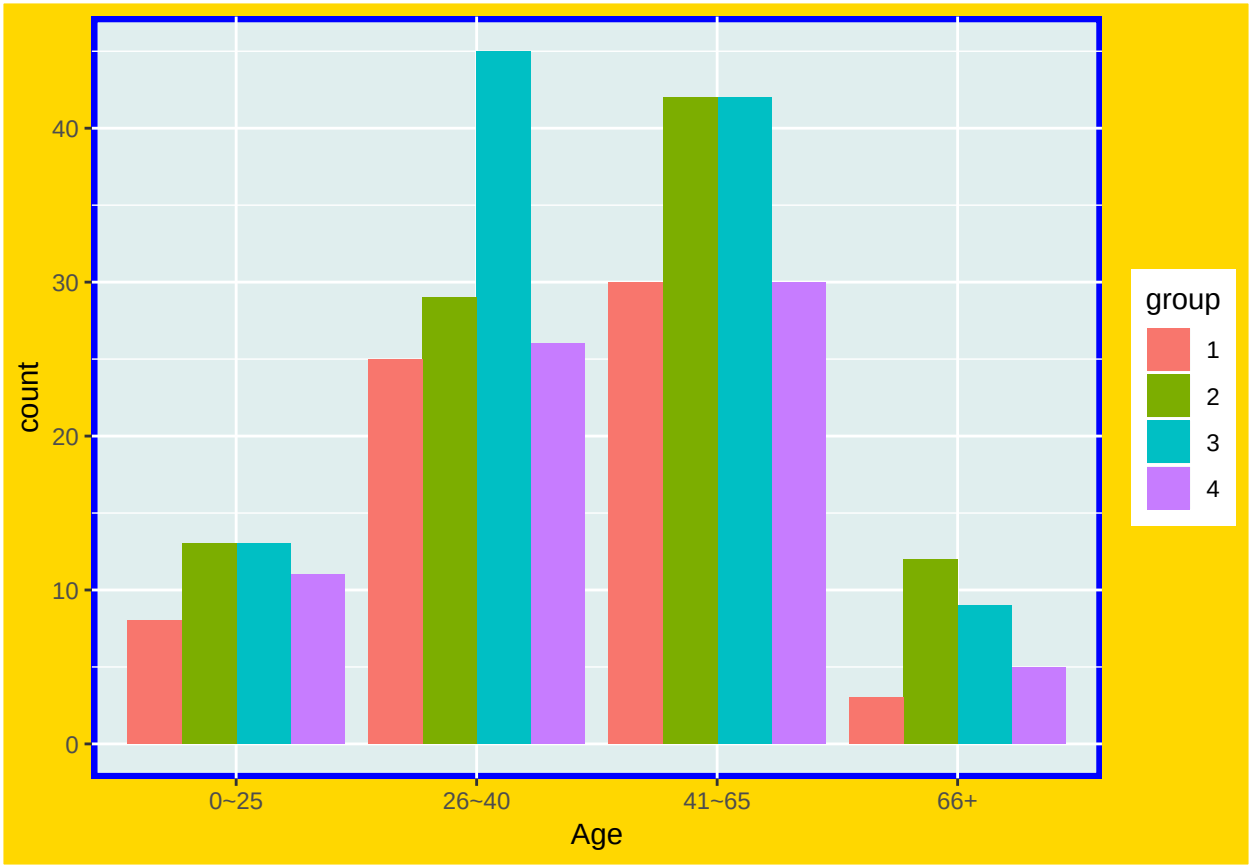
Centroids:

```
(ct=read_excel('C:\\Users\\Arnold\\OneDrive\\R_Python_working_directory\\Centroids.xlsx'))
```

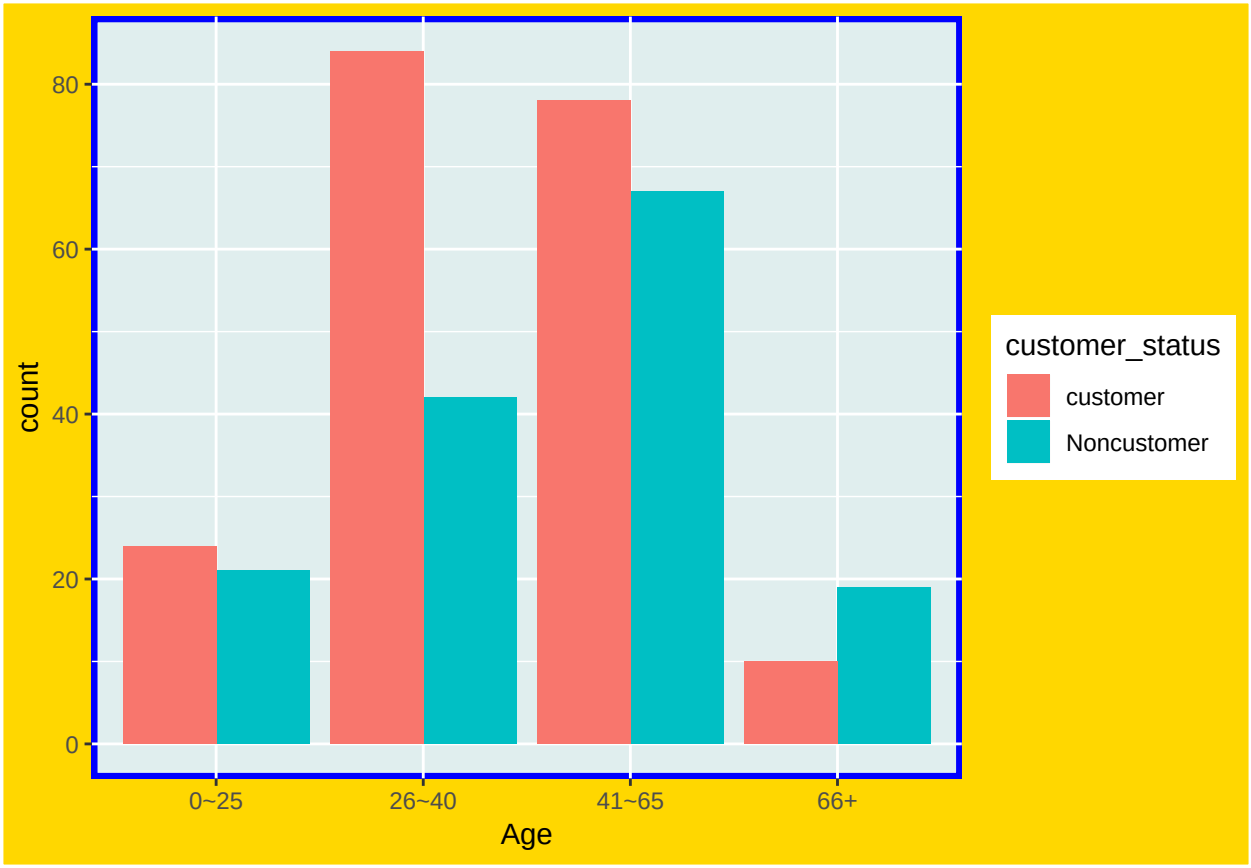
```
## # A tibble: 4 x 5
##   group planning spend_control local_preference healthy
##   <dbl>    <dbl>         <dbl>            <dbl>   <dbl>
## 1     1     1.76          1.92             1.24    1.32
## 2     2     1.25          3.63             1.87    1.34
## 3     3     2.27          3.45             2.02    2.22
## 4     4     1.94          1.84             2.28    2.22
```

Graphs for Age

```
ggdf=df %>% drop_na(Age)
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Age,fill=group),position = 'dodge')
```

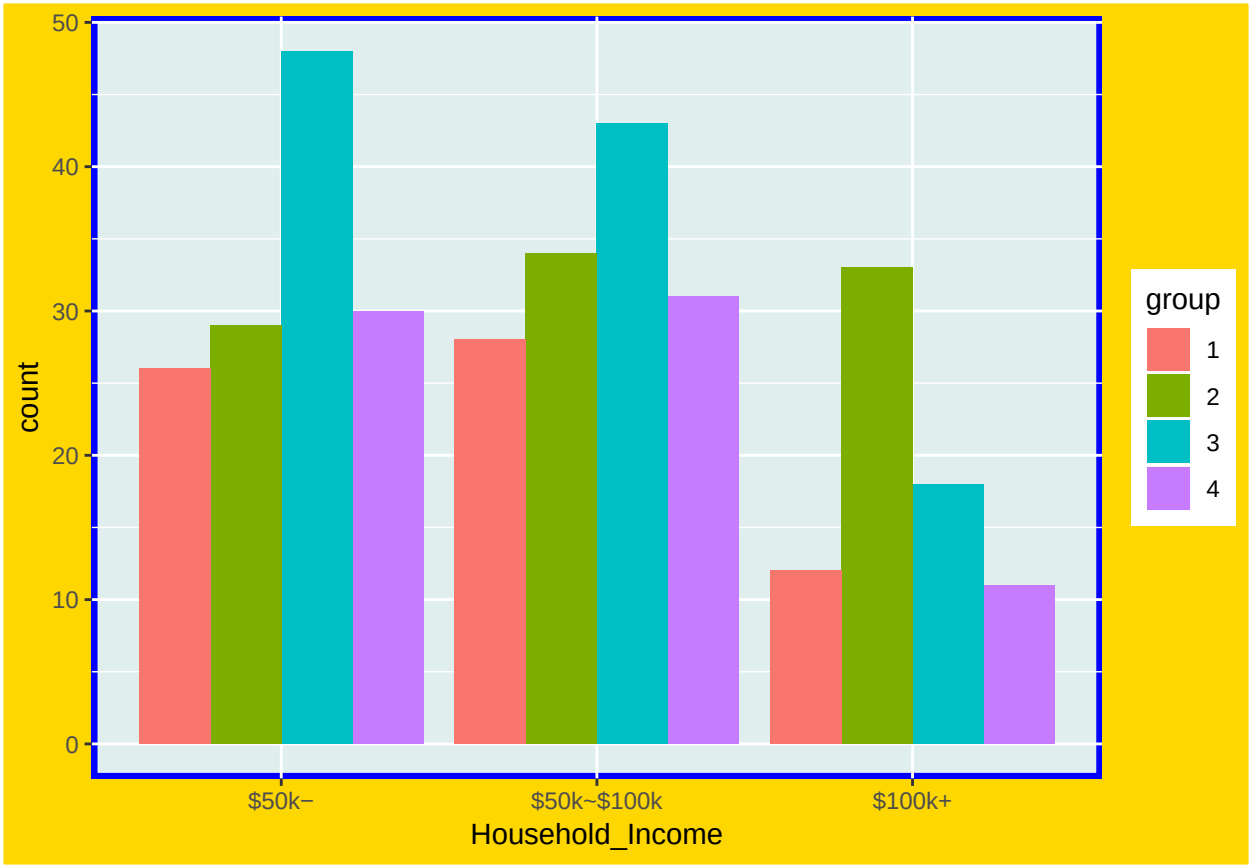



```
base_customer+geom_bar(aes(Age,fill=customer_status),position = 'dodge')
```

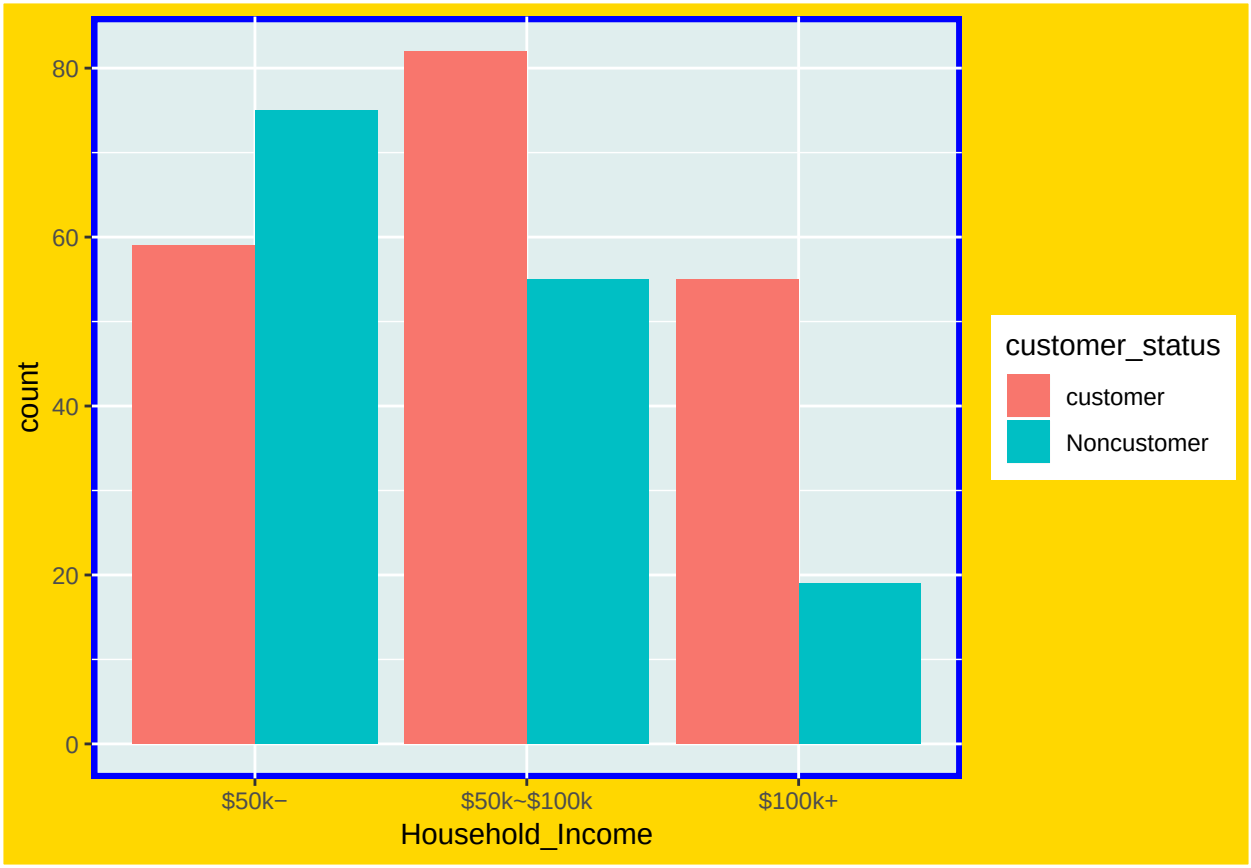


Graphs for Household_Income

```
ggdf=drop_na(df,Household_Income)
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Household_Income,fill=group),position = 'dodge')
```

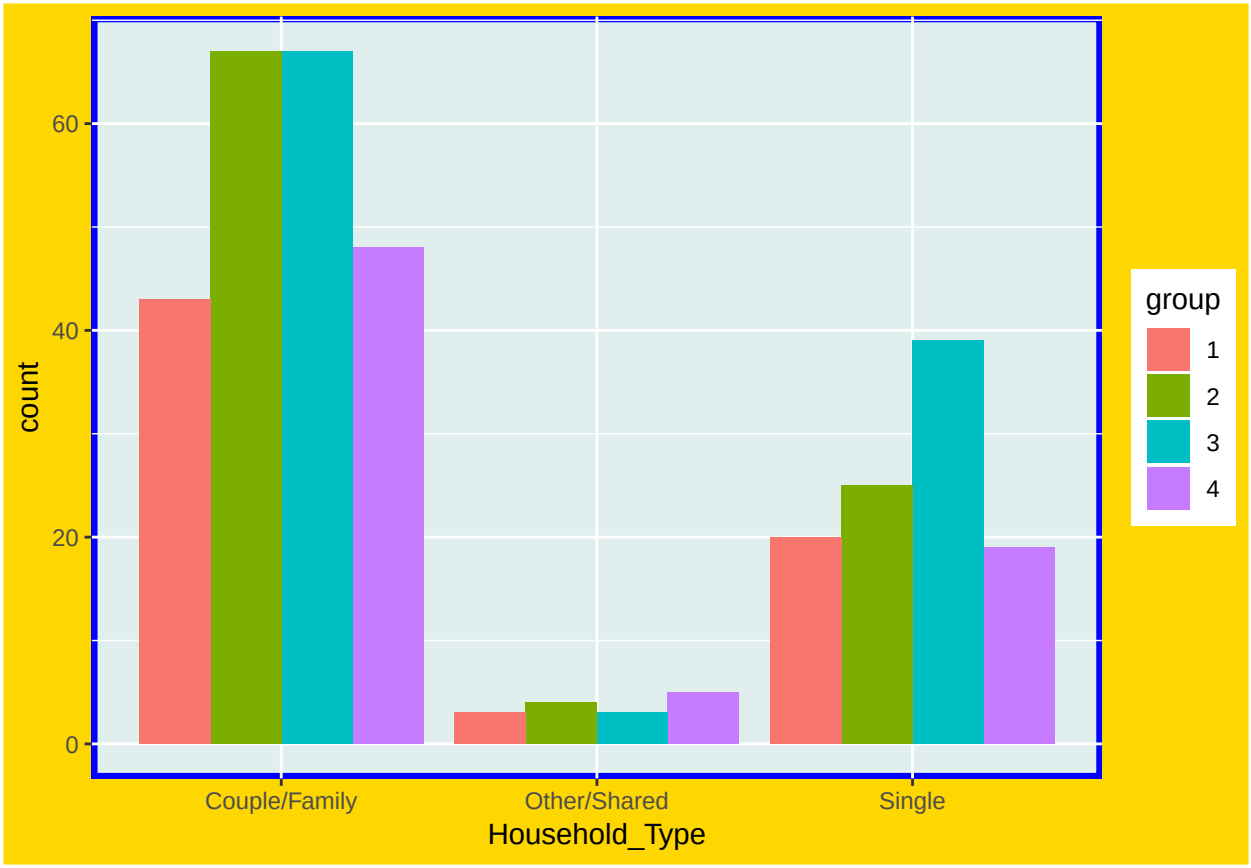



```
base_customer+geom_bar(aes(Household_Income,fill=customer_status),position = 'dodge')
```

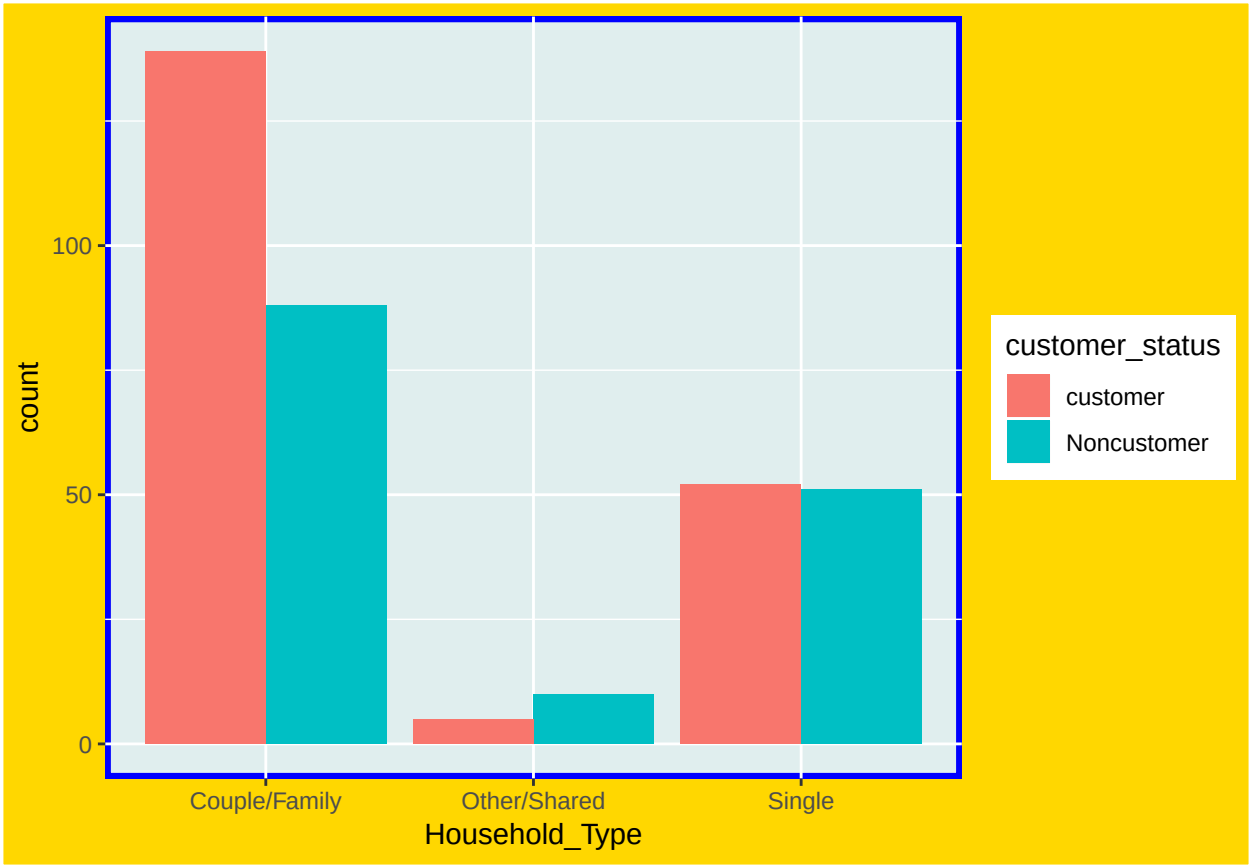


Graphs for Household_Type

```
ggdf=drop_na(df,Household_Type)
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Household_Type,fill=group),position = 'dodge')
```

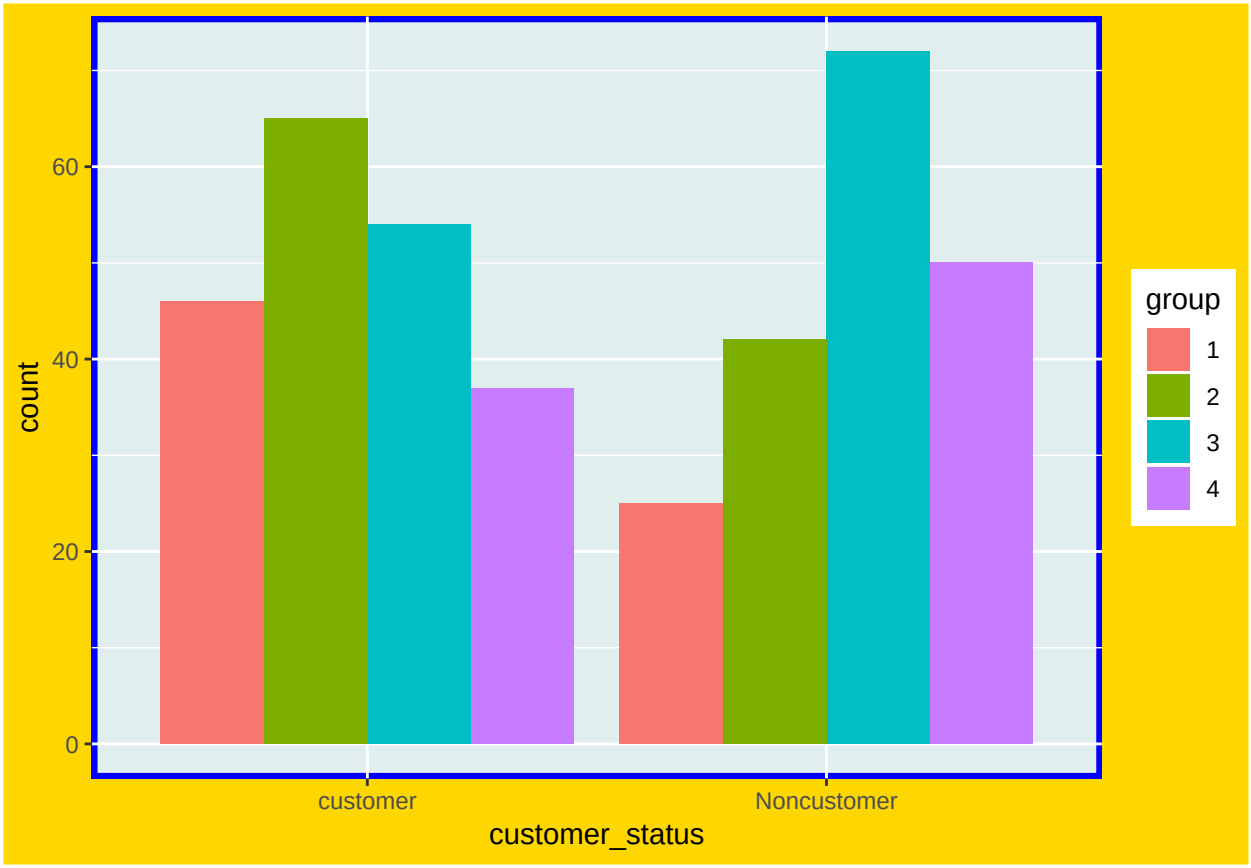



```
base_customer+geom_bar(aes(Household_Type,fill=customer_status),position = 'dodge')
```



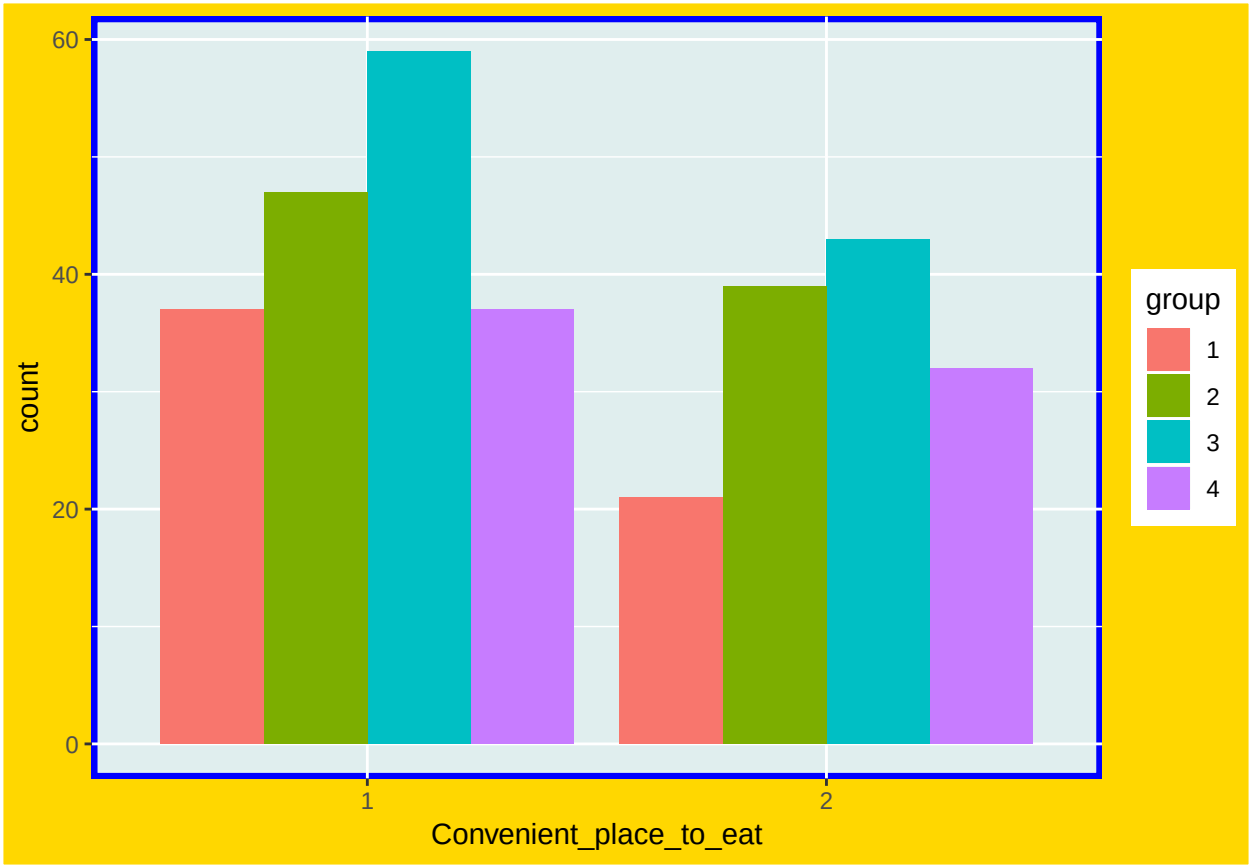
Graphs for customer_status

```
ggdf=drop_na(df,customer_status)
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_group+geom_bar(aes(customer_status,fill=group),position = 'dodge')
```

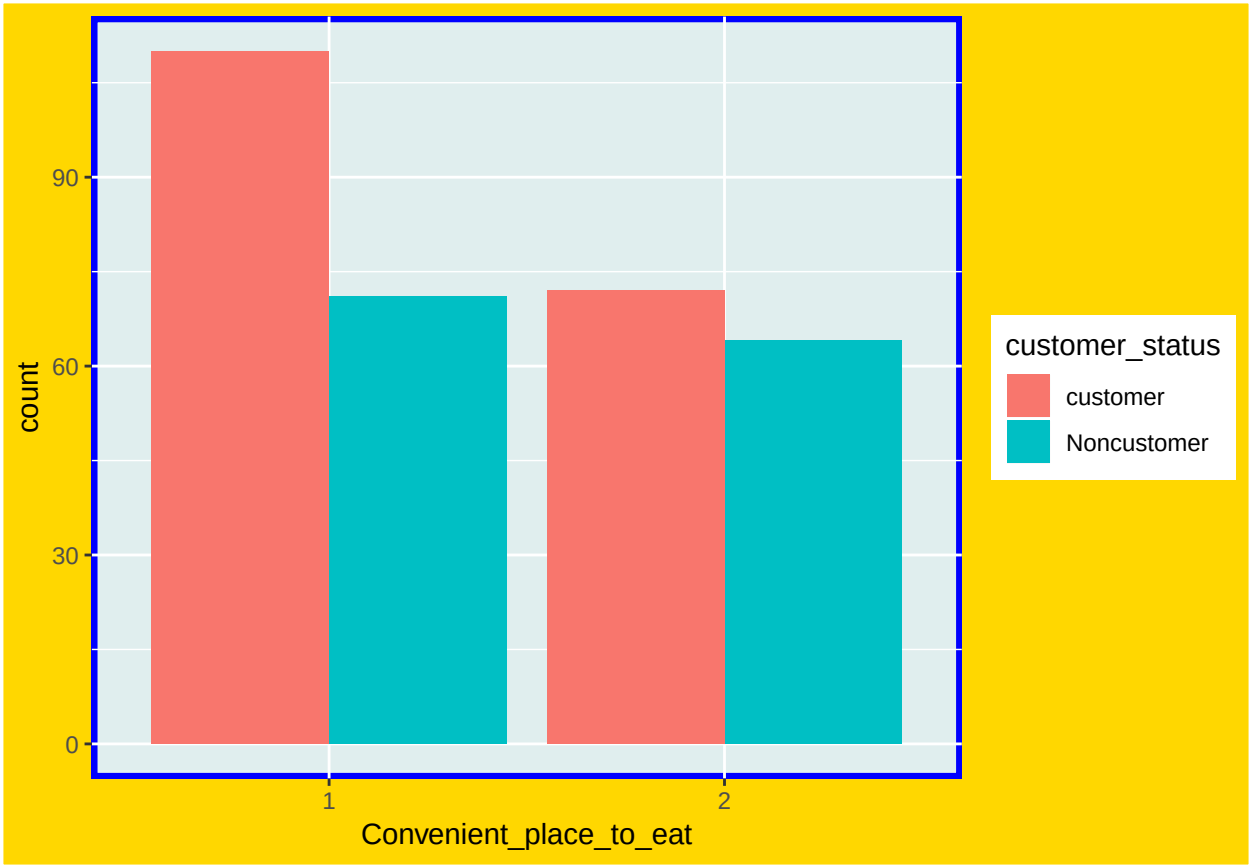


Graphs for "How important is Convenient_place_to_eat"

```
ggdf=drop_na(df,Convenient_place_to_eat) %>% filter(Convenient_place_to_eat %in% c('1','2'))
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Convenient_place_to_eat,fill=group),position = 'dodge')
```

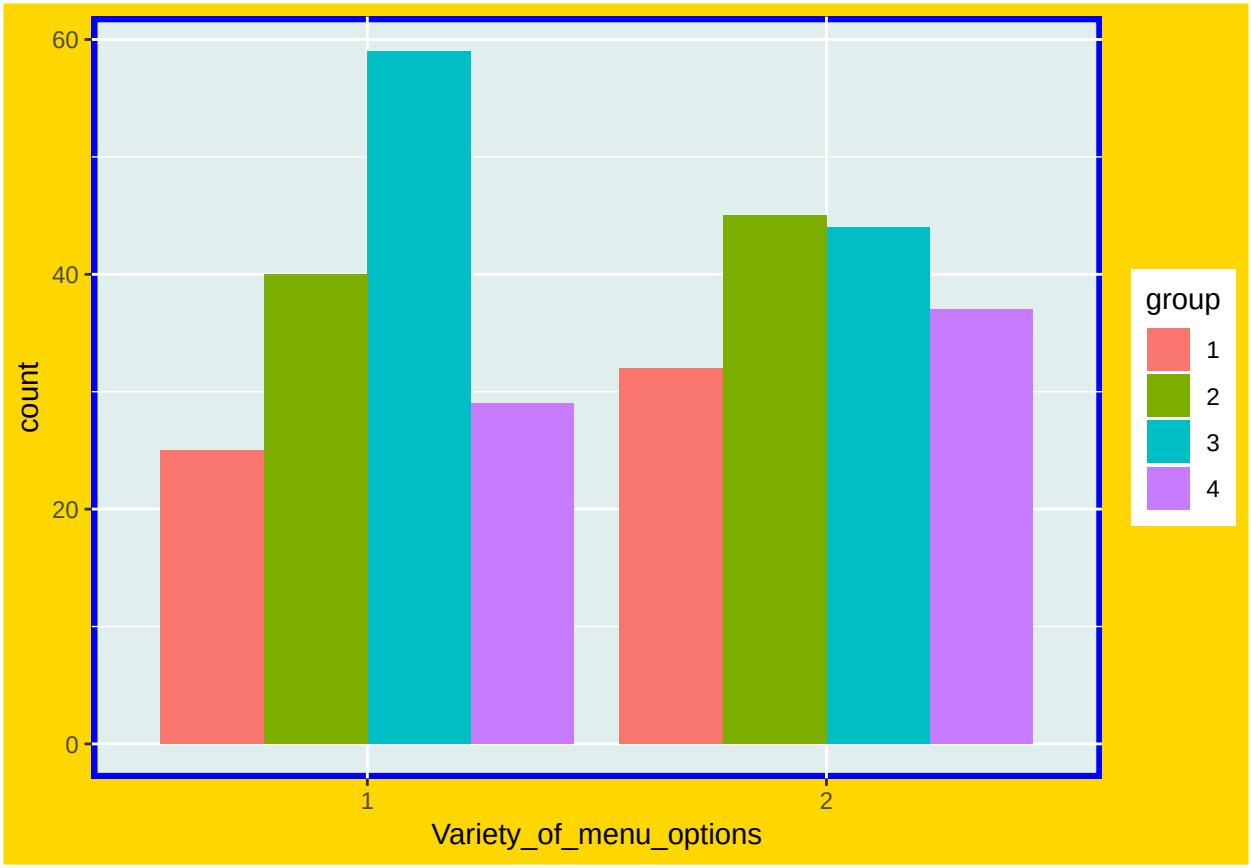



```
base_customer+geom_bar(aes(Convenient_place_to_eat,fill=customer_status),position = 'dodge')
```

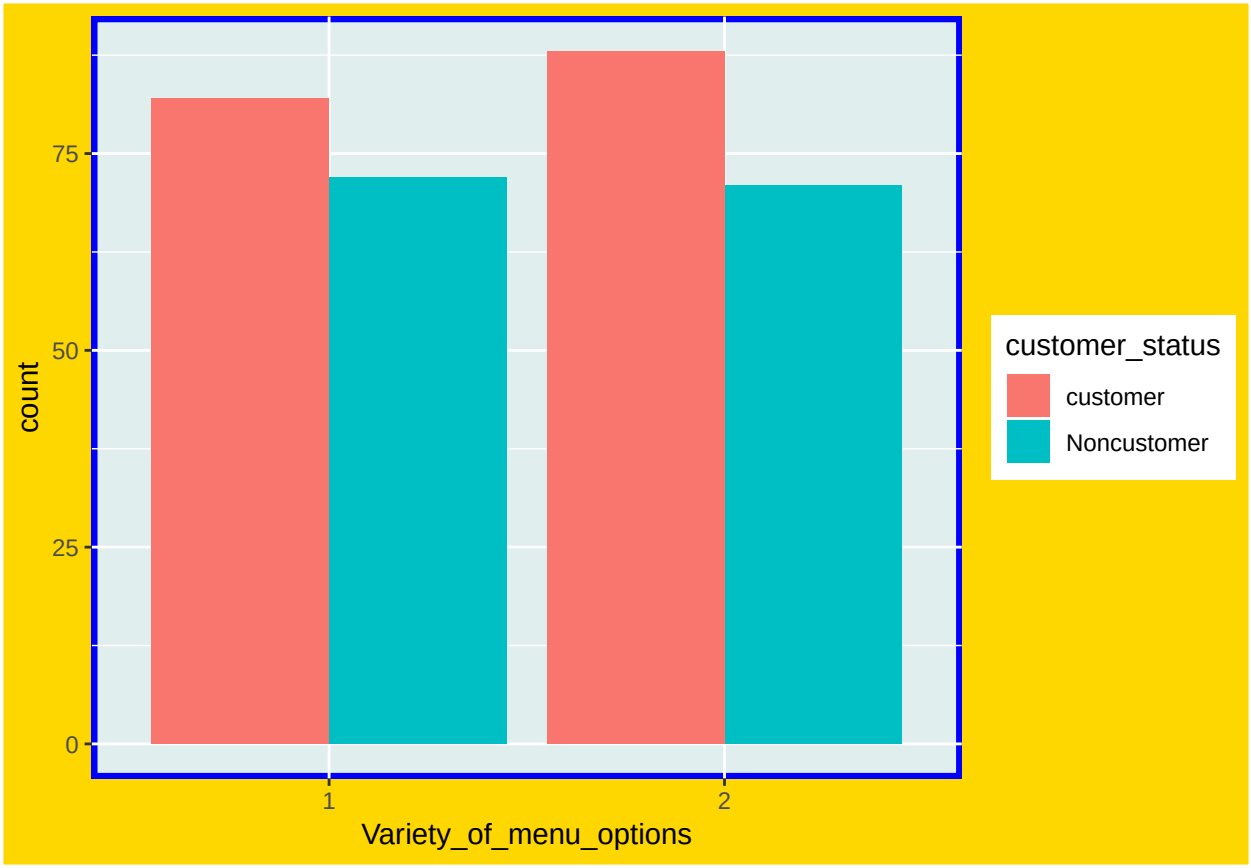


Graphs for "How important is Variety_of_menu_options"

```
ggdf=drop_na(df,Variety_of_menu_options) %>% filter(Variety_of_menu_options %in% c('1','2'))
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Variety_of_menu_options,fill=group),position = 'dodge')
```

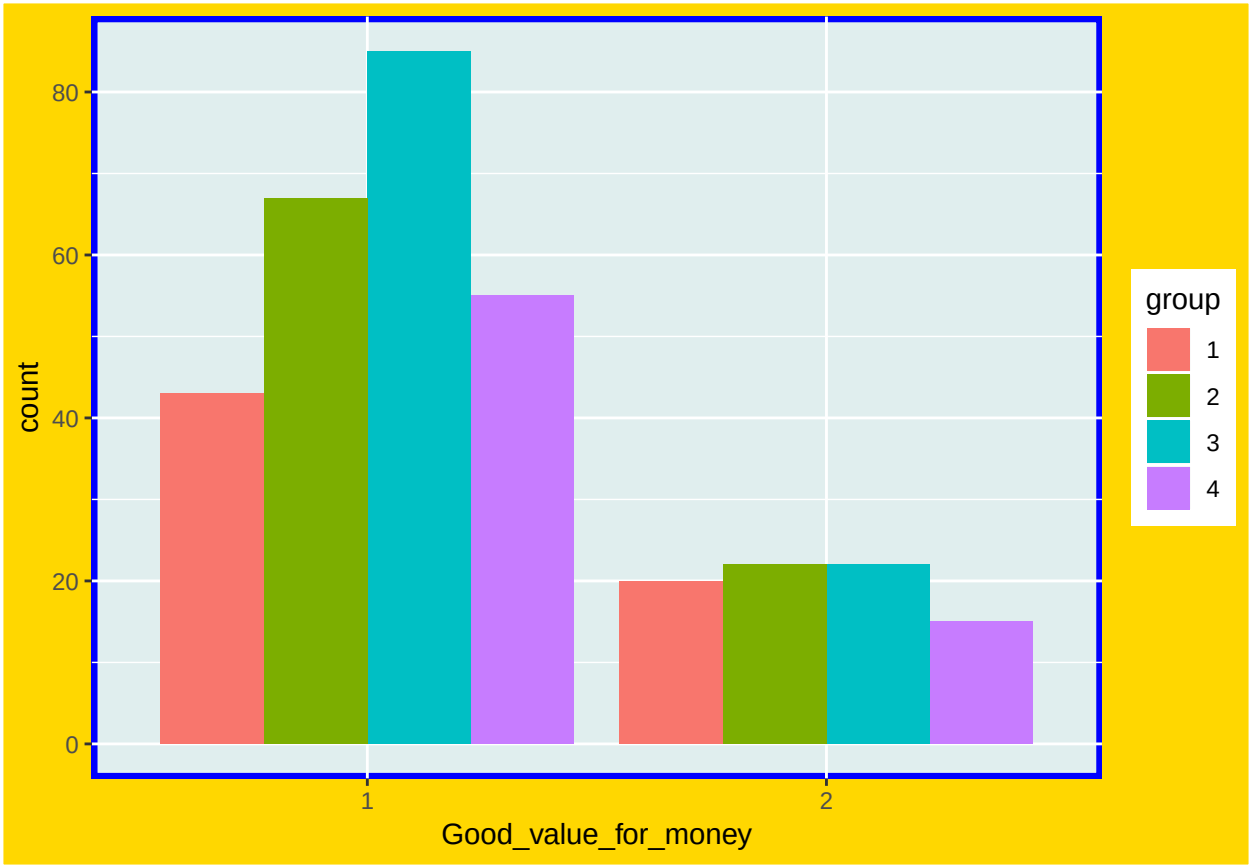



```
base_customer+geom_bar(aes(Variety_of_menu_options,fill=customer_status),position = 'dodge')
```

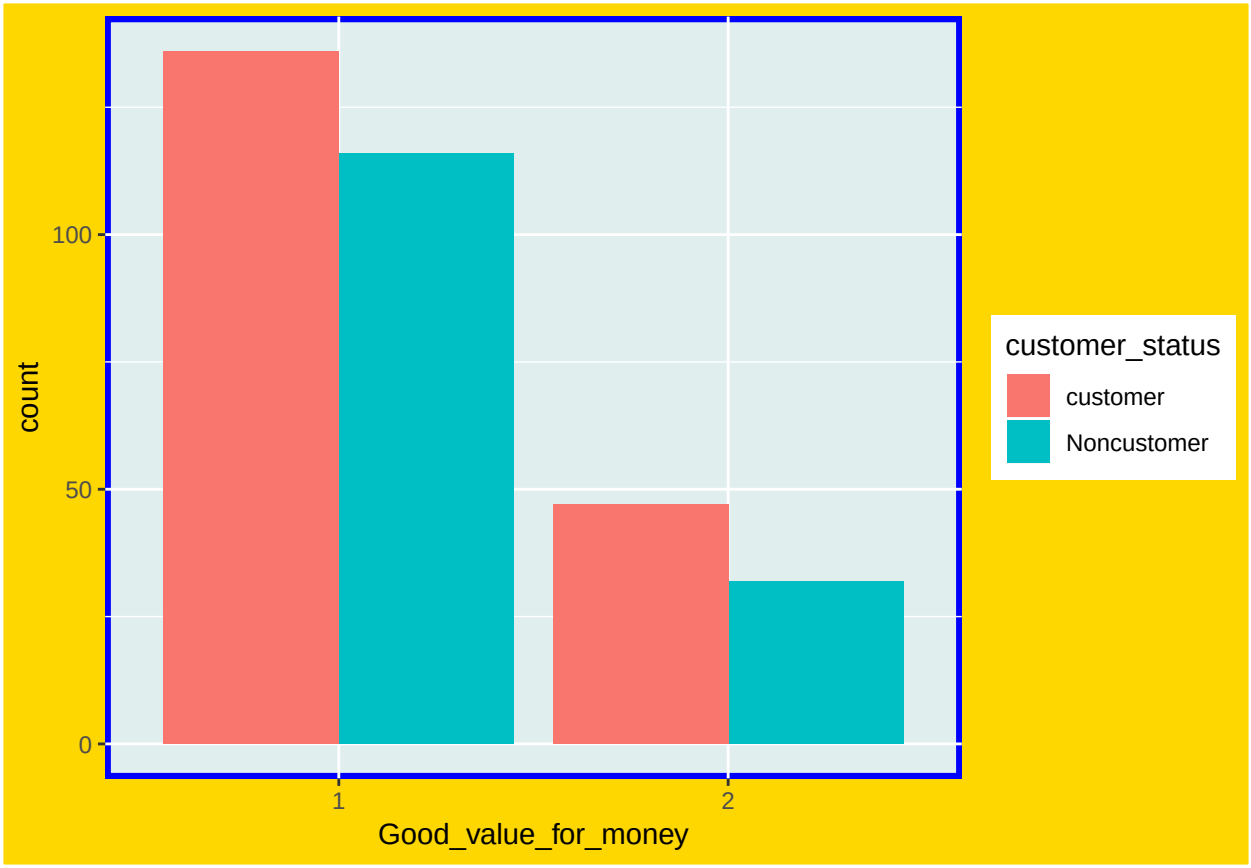


Graphs for "How important is Good_value_for_money"

```
ggdf=drop_na(df,Good_value_for_money) %>% filter(Good_value_for_money %in% c('1','2'))
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Good_value_for_money,fill=group),position = 'dodge')
```

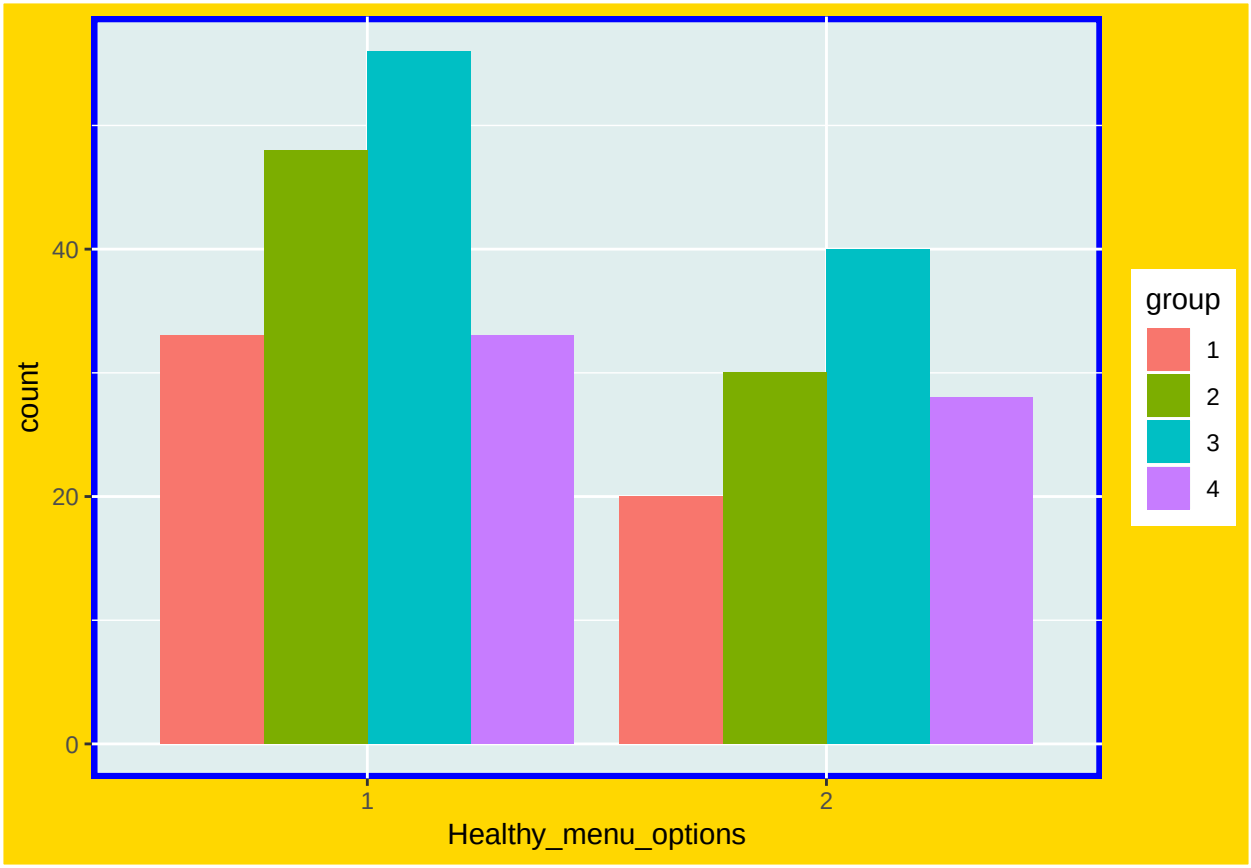



```
base_customer+geom_bar(aes(Good_value_for_money,fill=customer_status),position = 'dodge')
```

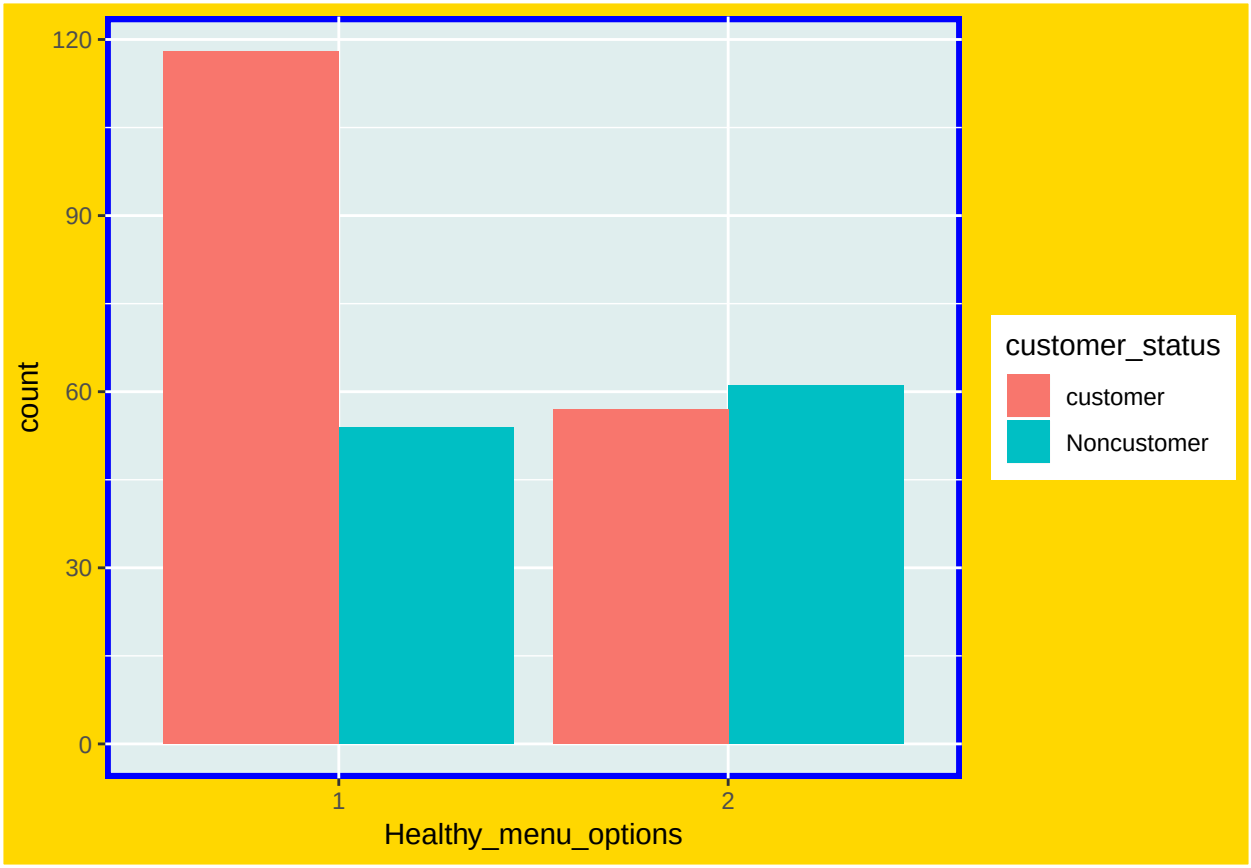


Graphs for "How important is Healthy_menu_options"

```
ggdf=drop_na(df,Healthy_menu_options) %>% filter(Healthy_menu_options %in% c('1','2'))
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Healthy_menu_options,fill=group),position = 'dodge')
```

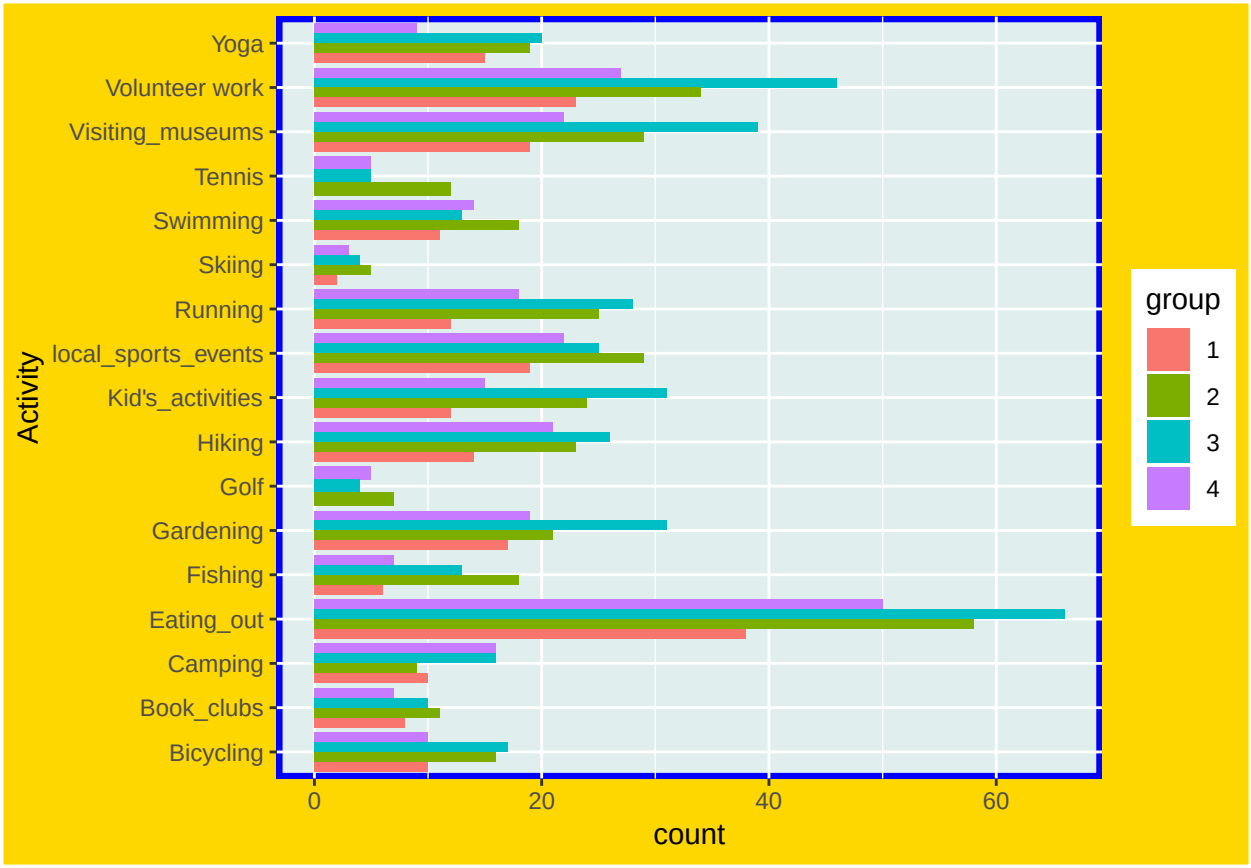



```
base_customer+geom_bar(aes(Healthy_menu_options,fill=customer_status),position = 'dodge')
```

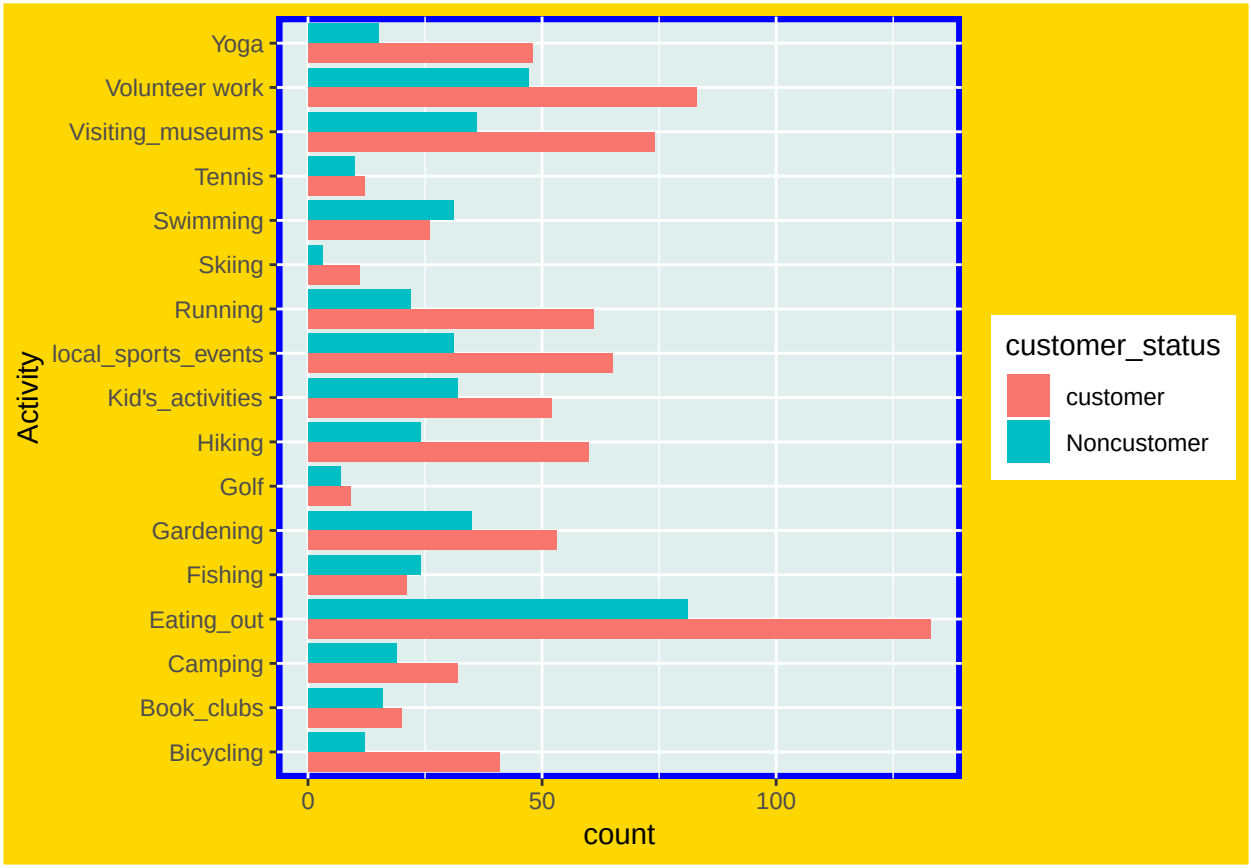


Graphs for activity

```
ggdf=Activity
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Activity,fill=group),position = 'dodge')+coord_flip()
```

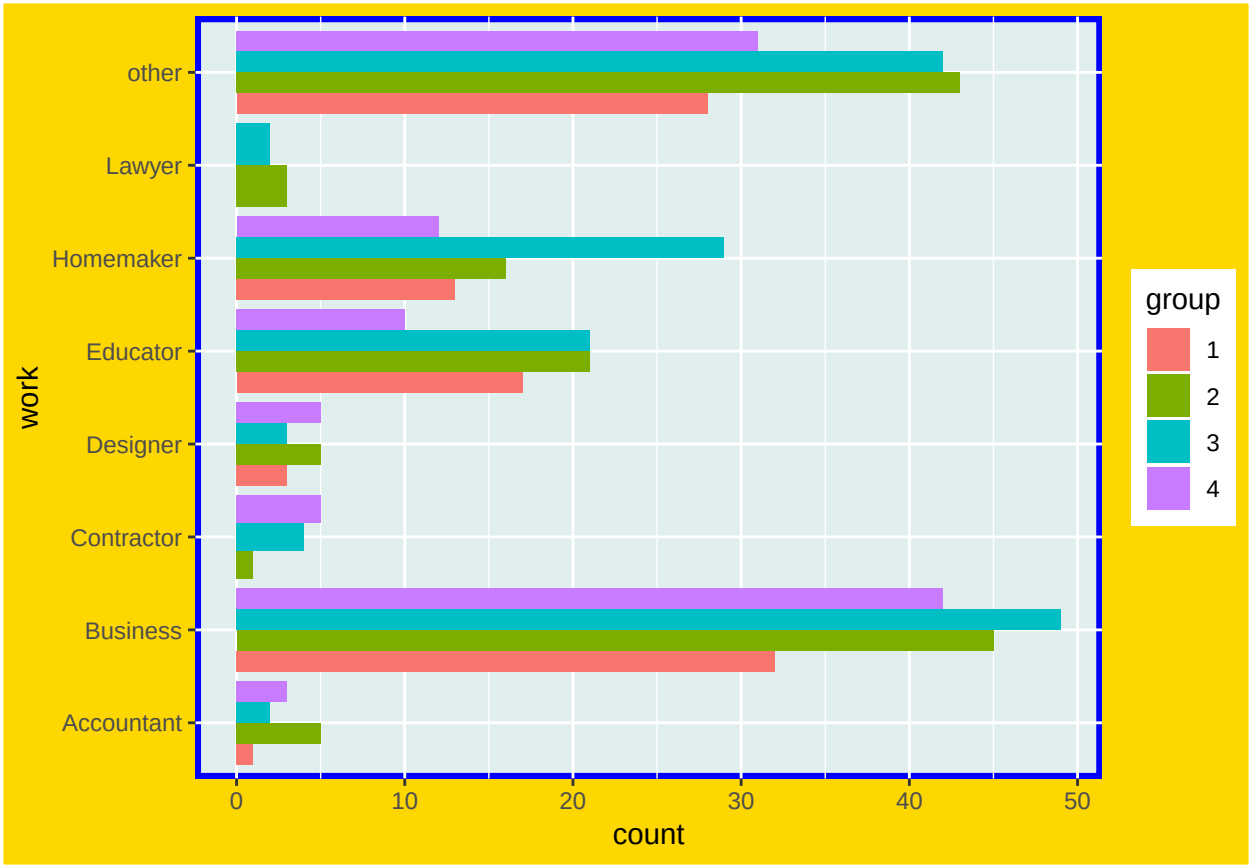



```
base_customer+geom_bar(aes(Activity,fill=customer_status),position = 'dodge')+coord_flip()
```

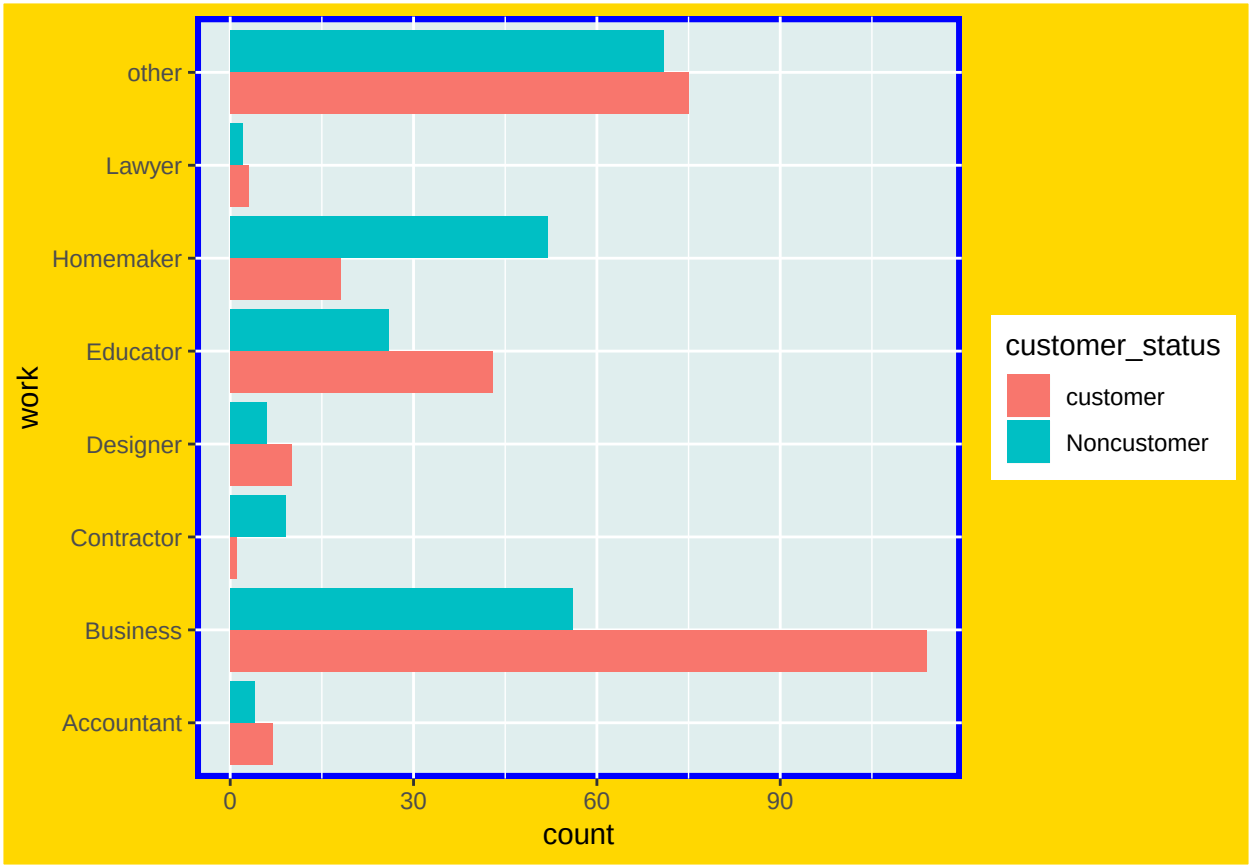


Graphs for profession

```
ggdf=Profession
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(work,fill=group),position = 'dodge')+coord_flip()
```

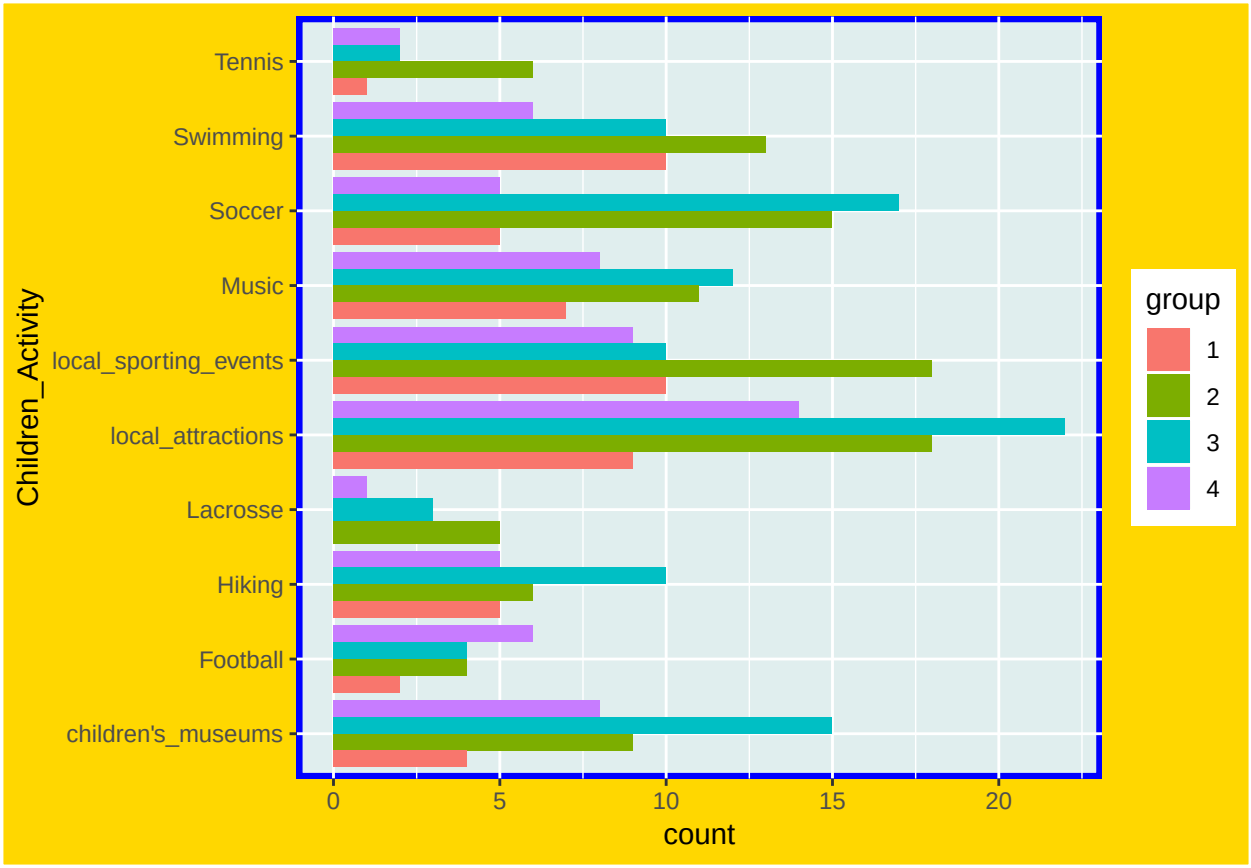



```
base_customer+geom_bar(aes(work,fill=customer_status),position = 'dodge')+coord_flip()
```

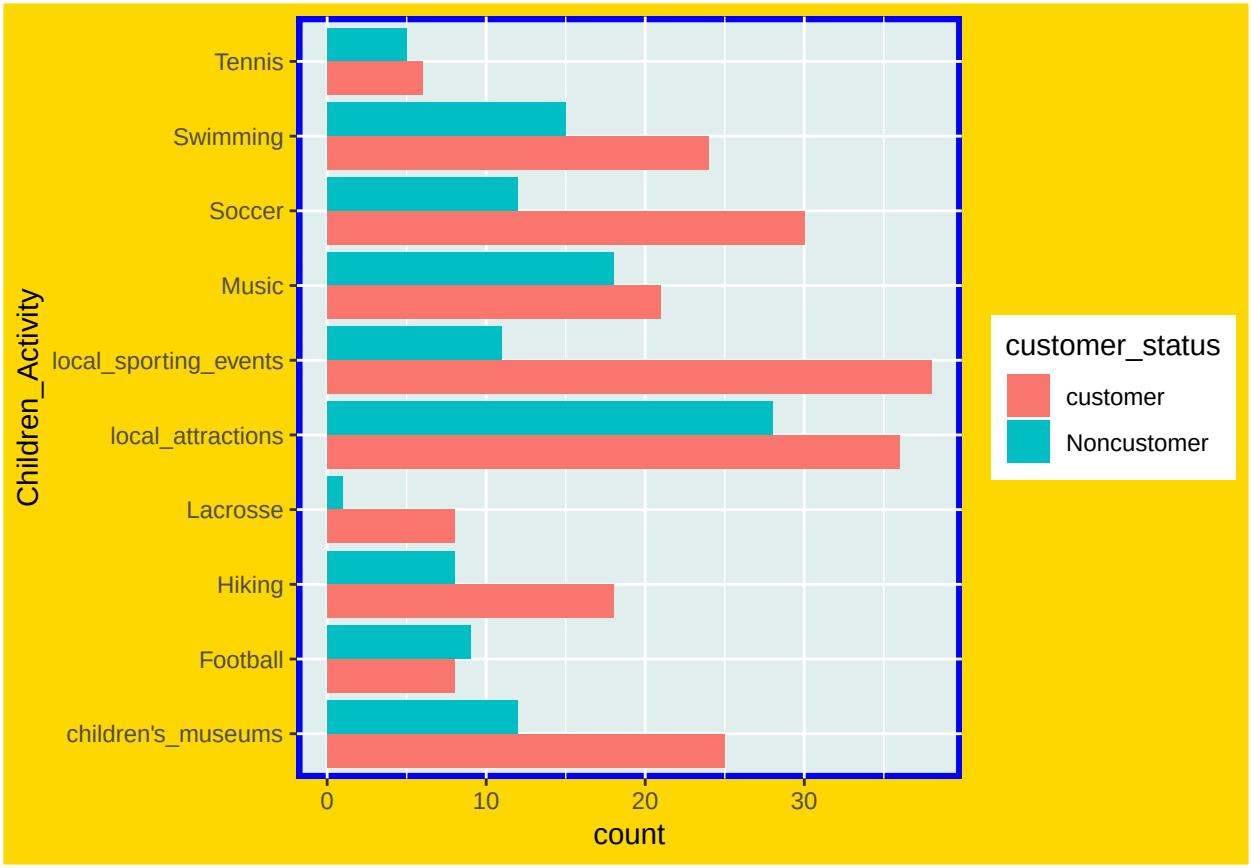


Graphs for children activity

```
ggdf=child_act
base_group=ggplot(drop_na(ggdf,group),aes(group=group))
base_customer=ggplot(drop_na(ggdf,customer_status),aes(group=customer_status))
base_group+geom_bar(aes(Children_Activity,fill=group),position = 'dodge')+coord_flip()
```

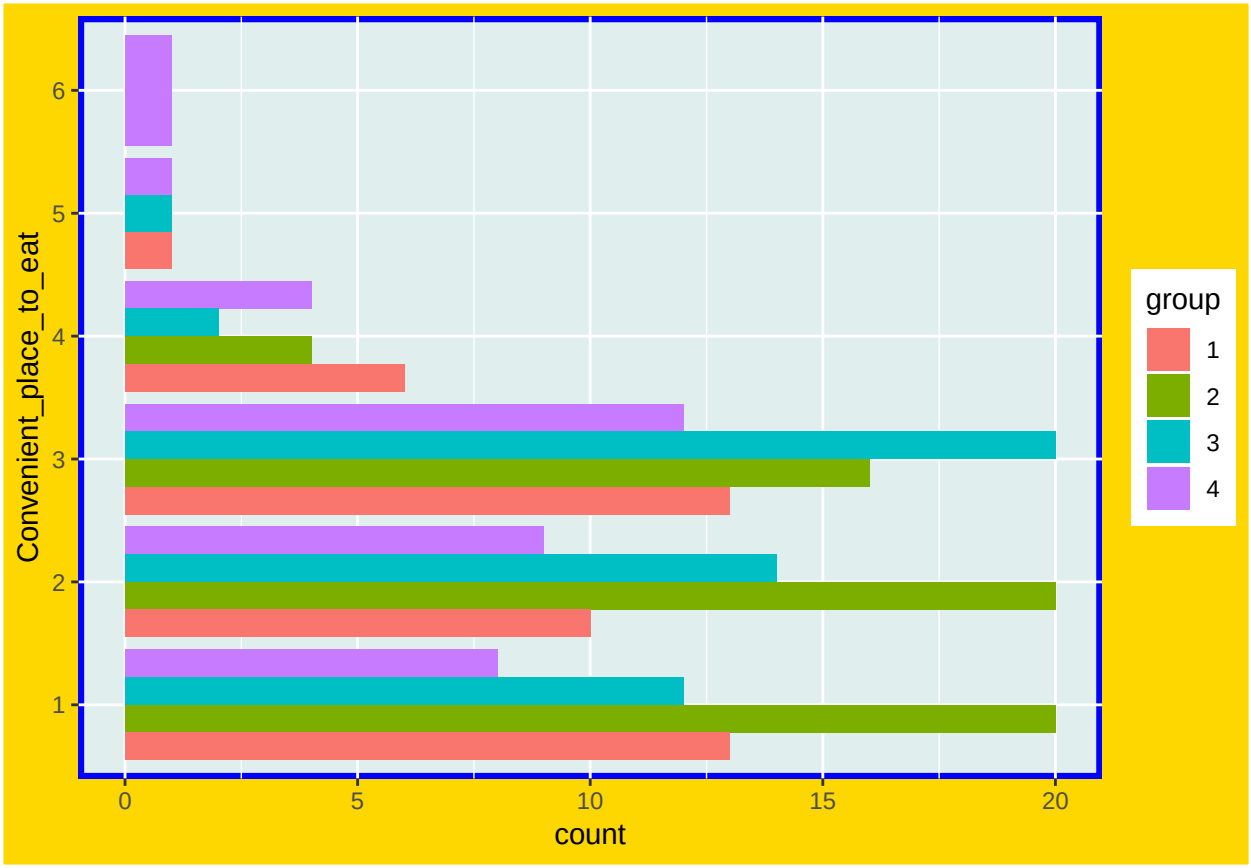



```
base_customer+geom_bar(aes(Children_Activity,fill=customer_status),position = 'dodge')+coord_flip()
```

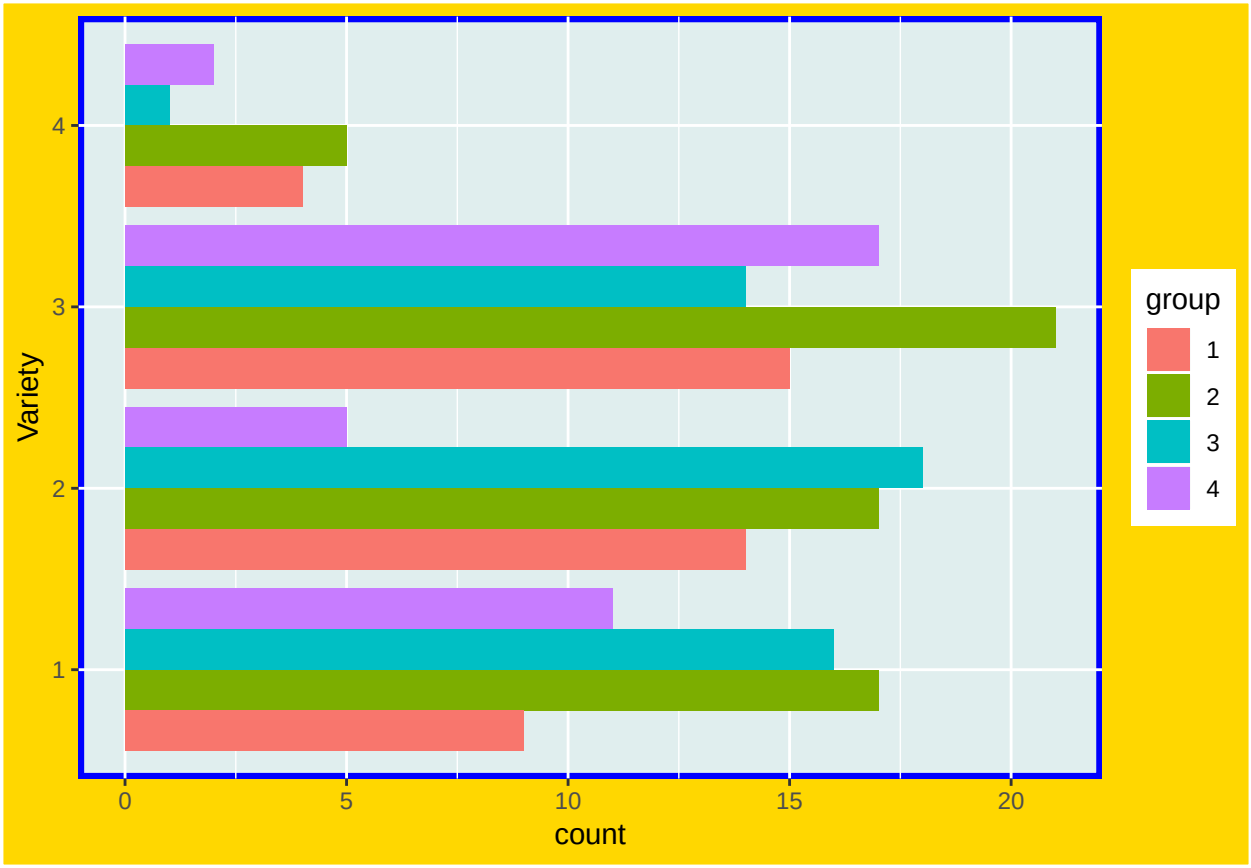


Graphs for the question "Please indicate how you rate Sticks in comparison to similar restaurants that you visit regularly"

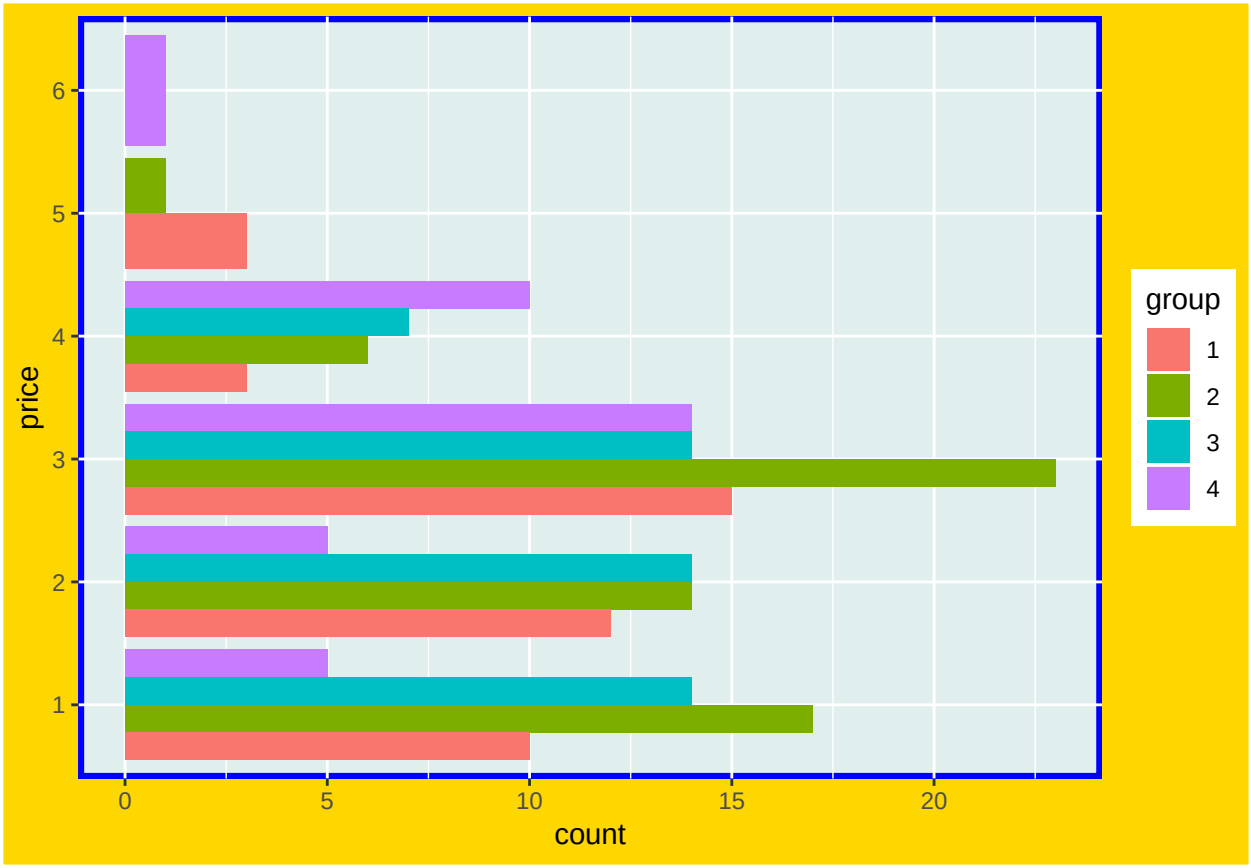
```
ggplot(drop_na(Compare,group,Convenient_place_to_eat),aes(group=group))+geom_bar(aes(Convenient_place_t
```



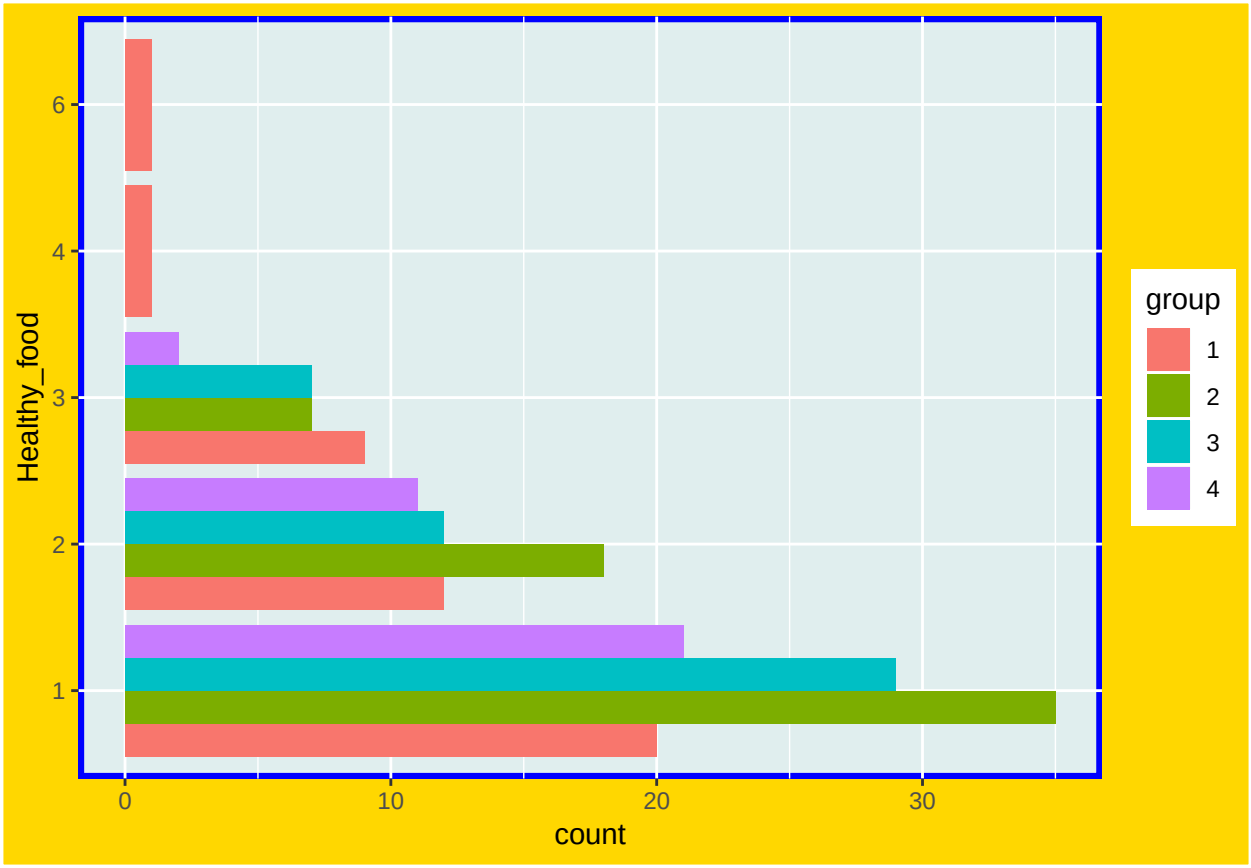

```
ggplot(drop_na(Compare,group,Variety),aes(group=group))+geom_bar(aes(Variety,fill=group),position = 'do
```



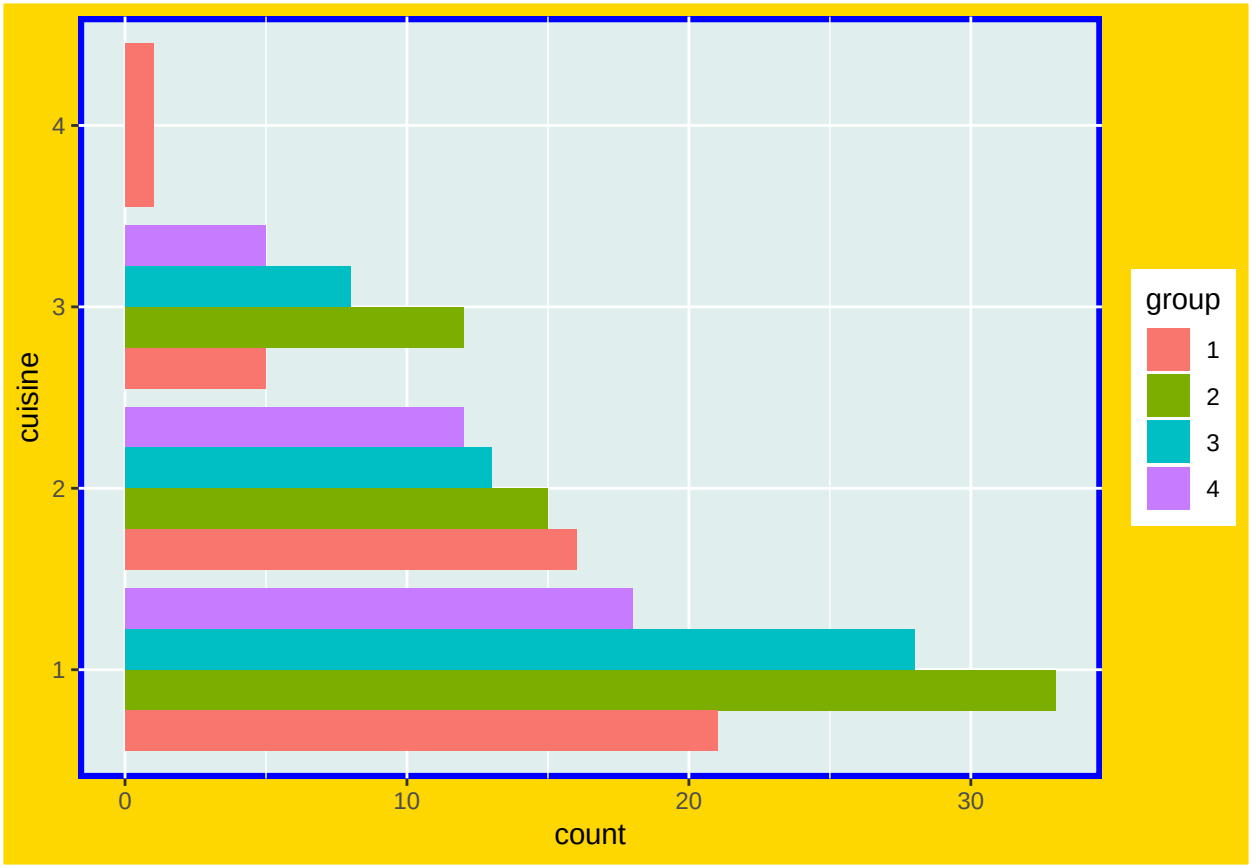

```
ggplot(drop_na(Compare,group,price),aes(group=group))+geom_bar(aes(price,fill=group),position = 'dodge')
```



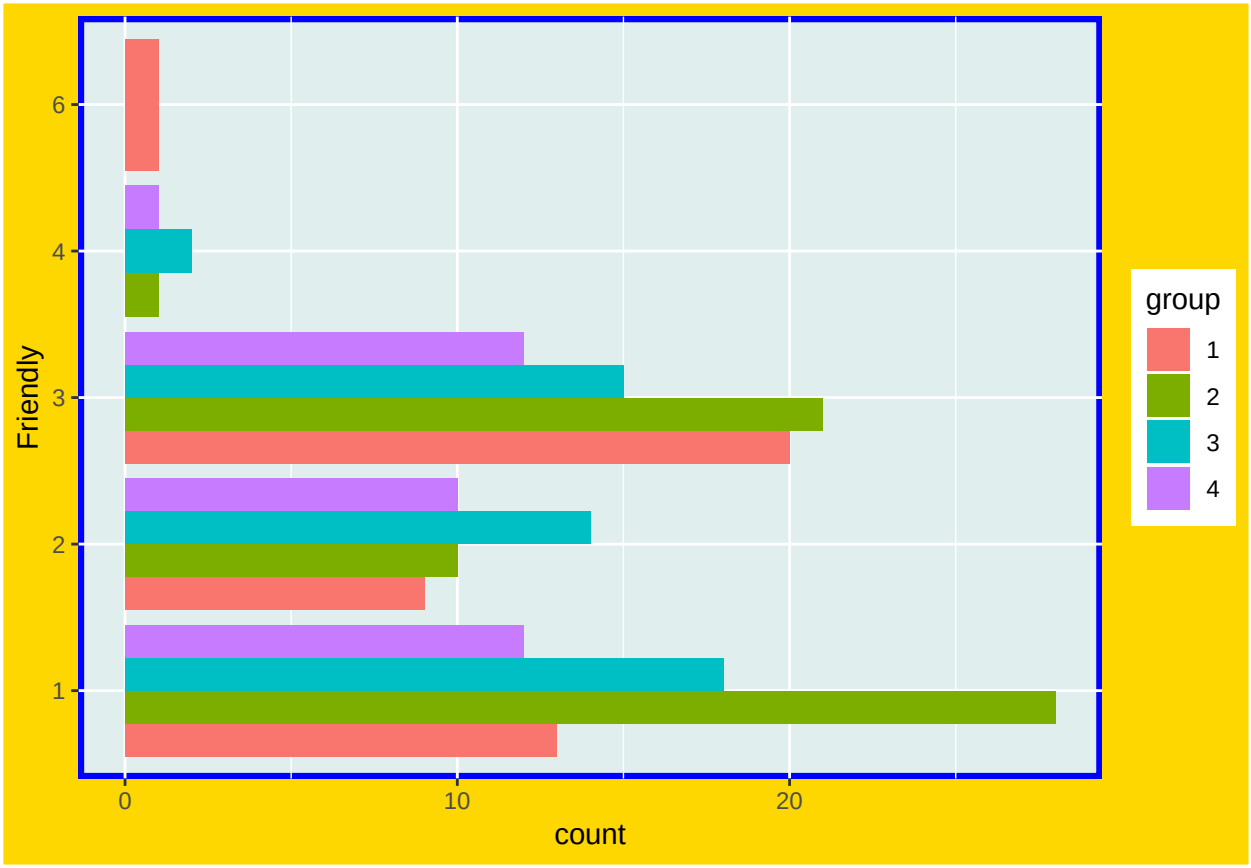

```
ggplot(drop_na(Compare,group,Healthy_food),aes(group=group))+geom_bar(aes(Healthy_food,fill=group),posi
```



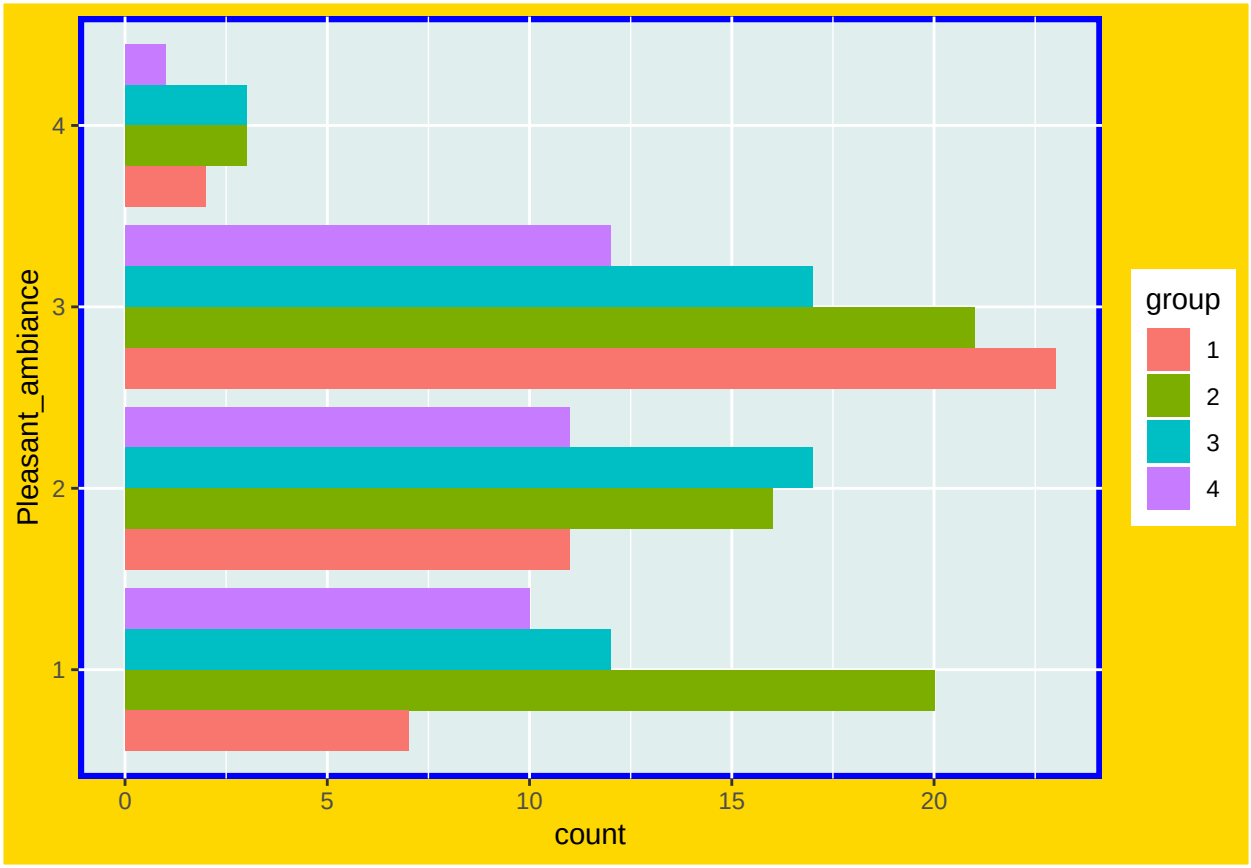

```
ggplot(drop_na(Compare,group,cuisine),aes(group=group))+geom_bar(aes(cuisine,fill=group),position = 'do
```



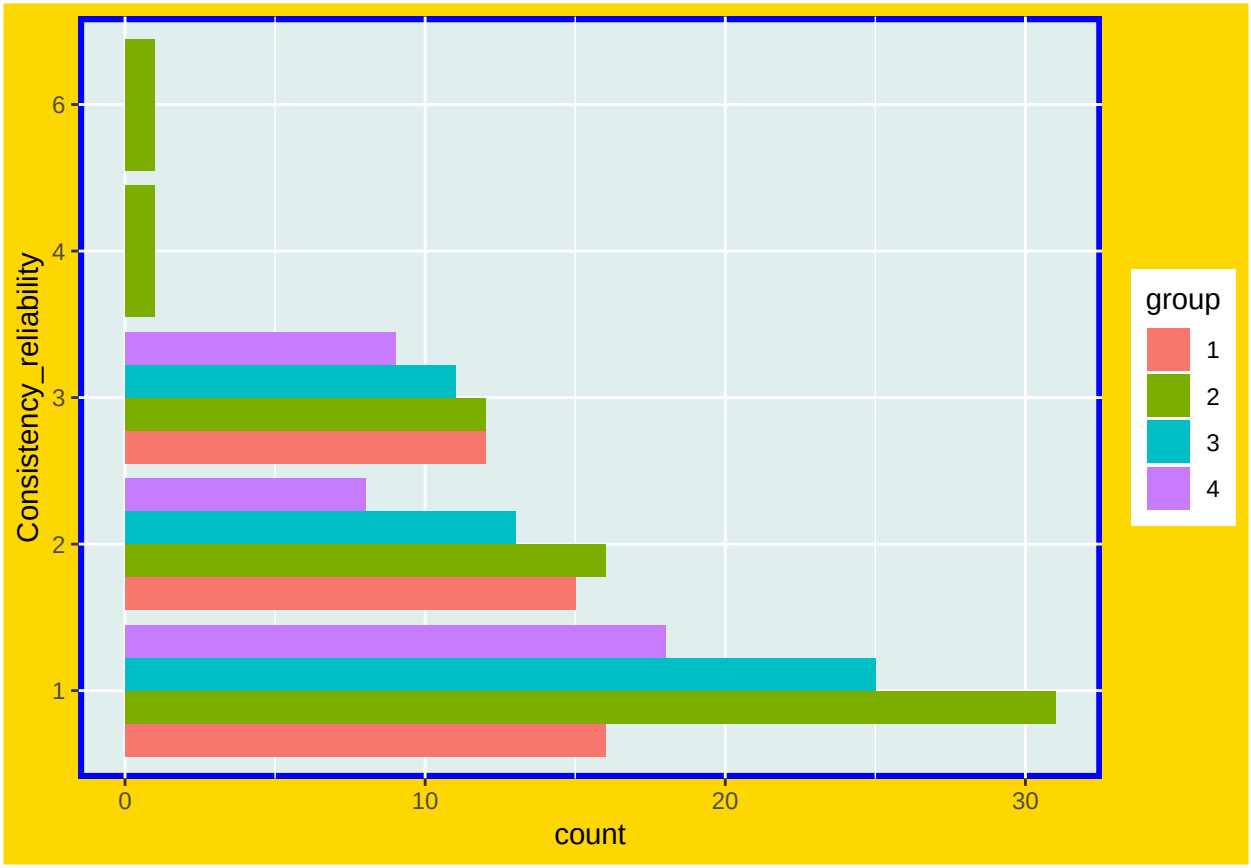

```
ggplot(drop_na(Compare,group,Friendly),aes(group=group))+geom_bar(aes(Friendly,fill=group),position = '
```



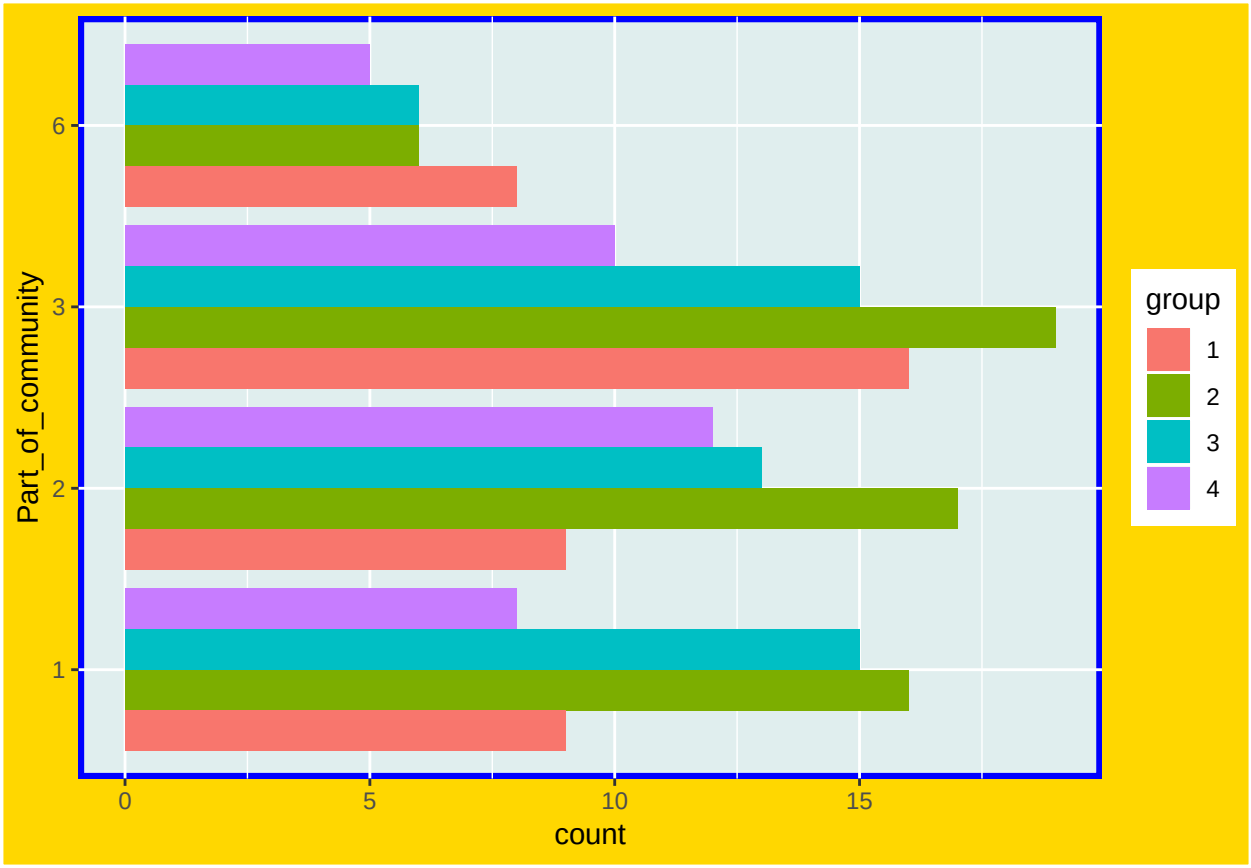

```
ggplot(drop_na(Compare,group,Pleasant_ambiance),aes(group=group))+geom_bar(aes(Pleasant_ambiance,fill=g
```




```
ggplot(drop_na(Compare,group,Consistency_reliability),aes(group=group))+geom_bar(aes(Consistency_reliabi
```

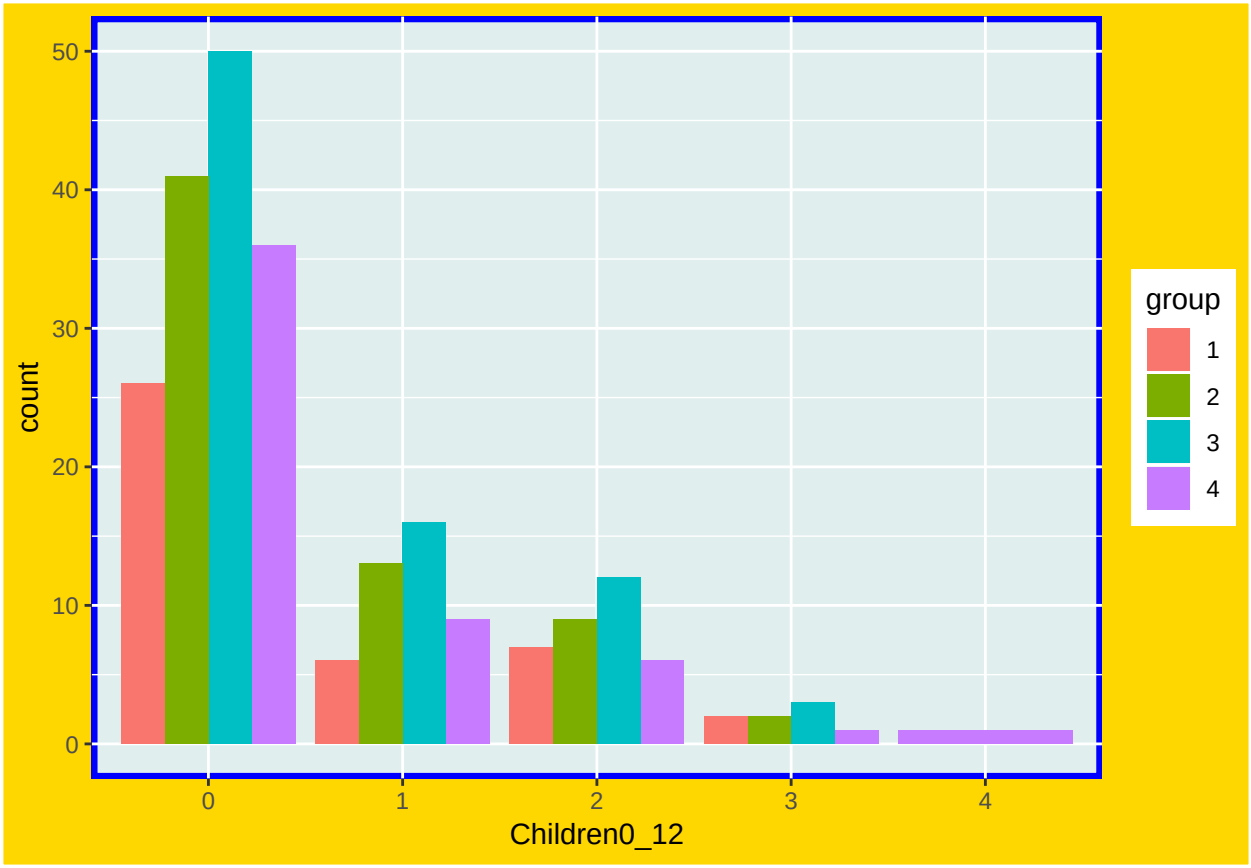



```
ggplot(drop_na(Compare,group,Part_of_community),aes(group=group))+geom_bar(aes(Part_of_community,fill=g
```

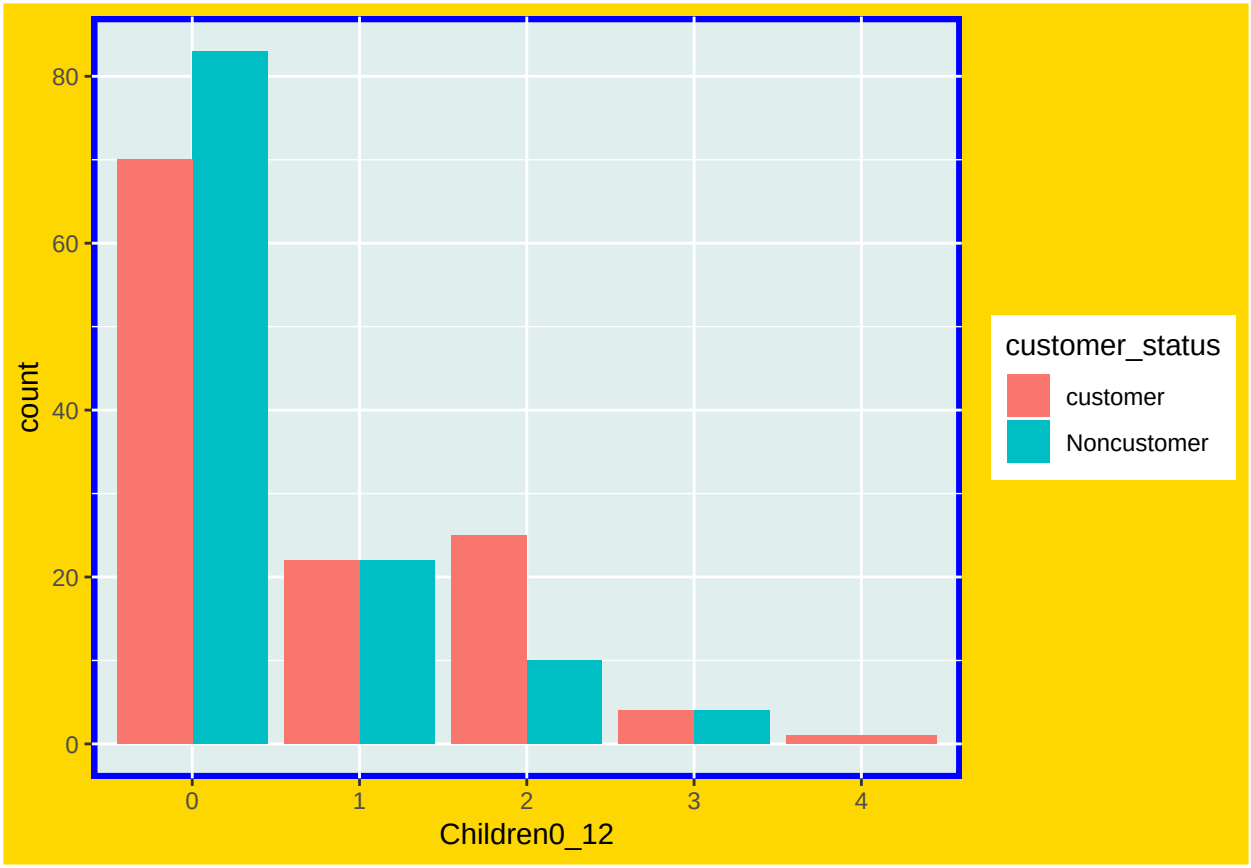


Graphs for house hold imformation

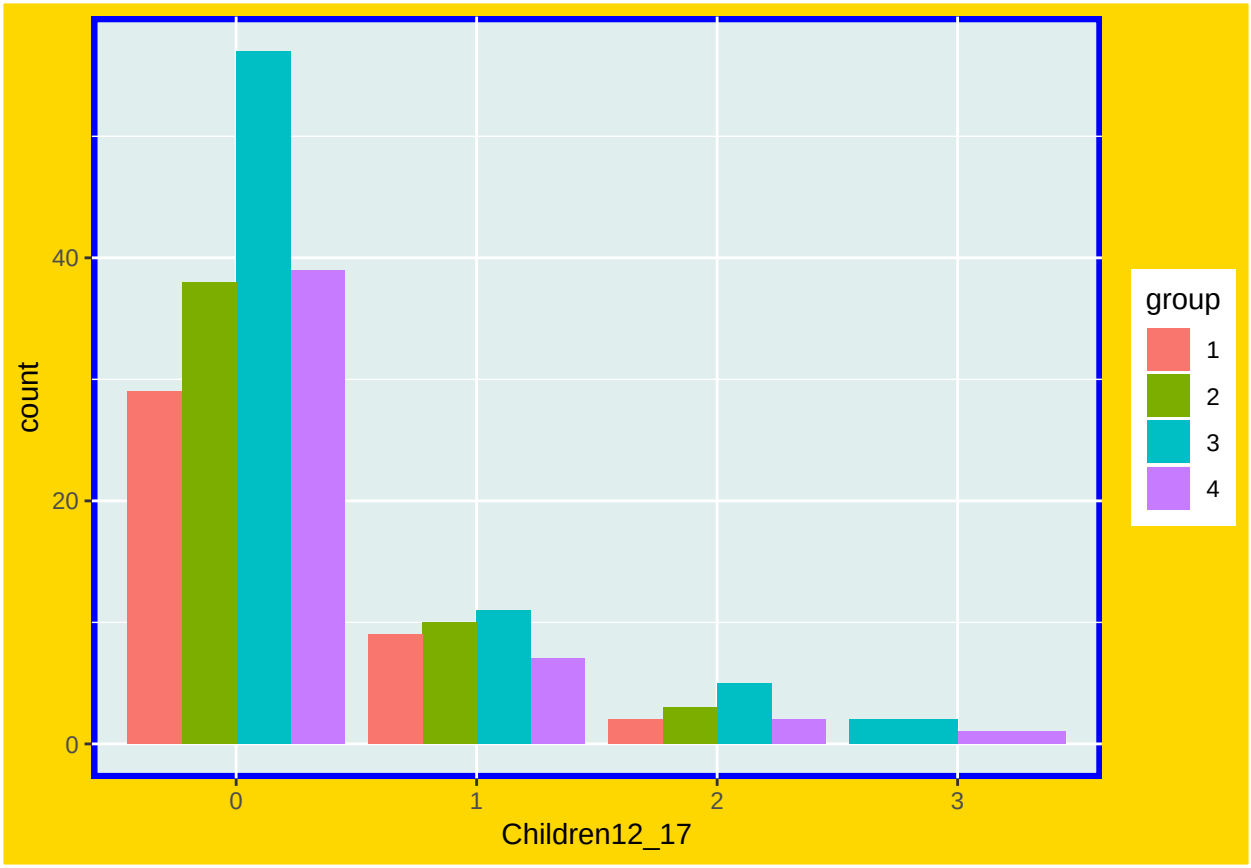
```
gdf=drop_na(house,group)
cdf=drop_na(house,customer_status)
ggplot(drop_na(gdf,Children0_12),aes(group=group))+geom_bar(aes(Children0_12,fill=group),position = 'do
```



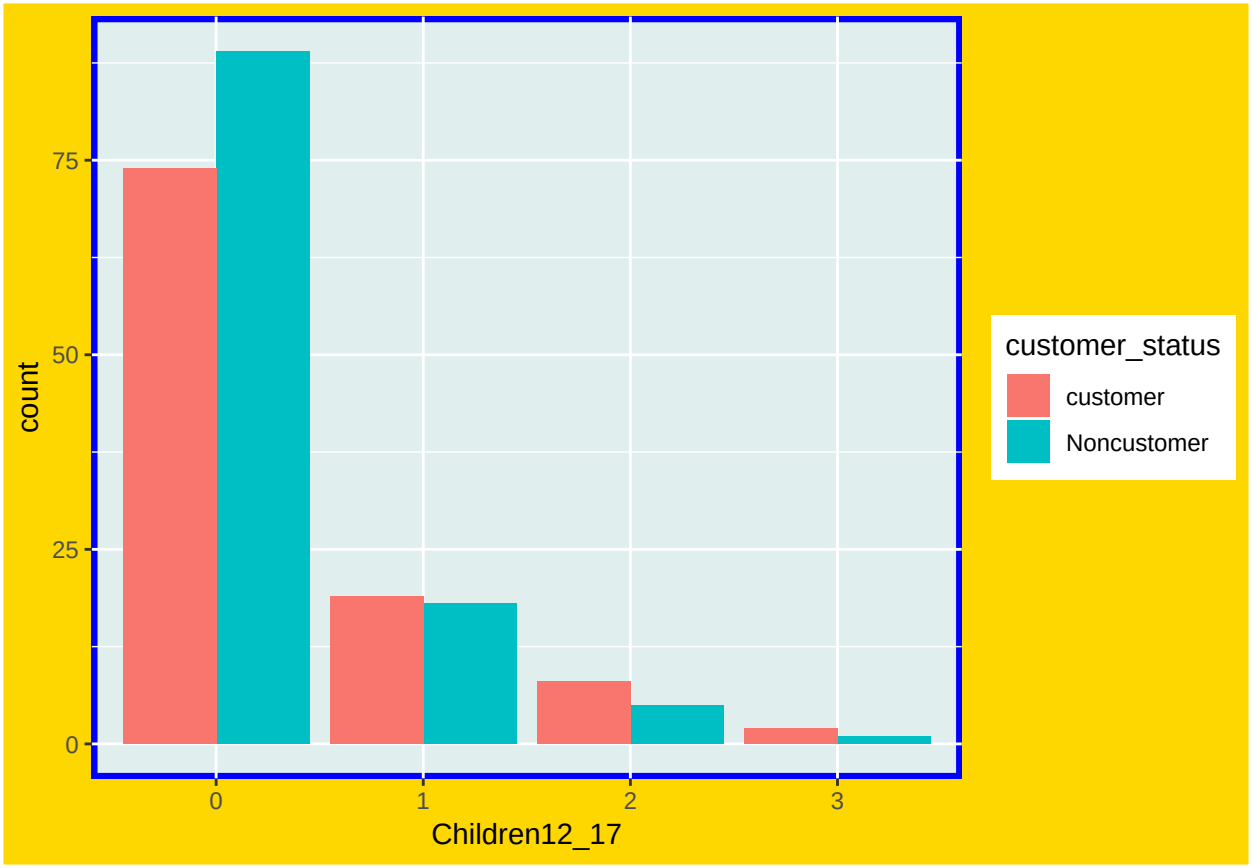

```
ggplot(drop_na(cdf,Children0_12),aes(group=customer_status))+geom_bar(aes(Children0_12,fill=customer_sta
```



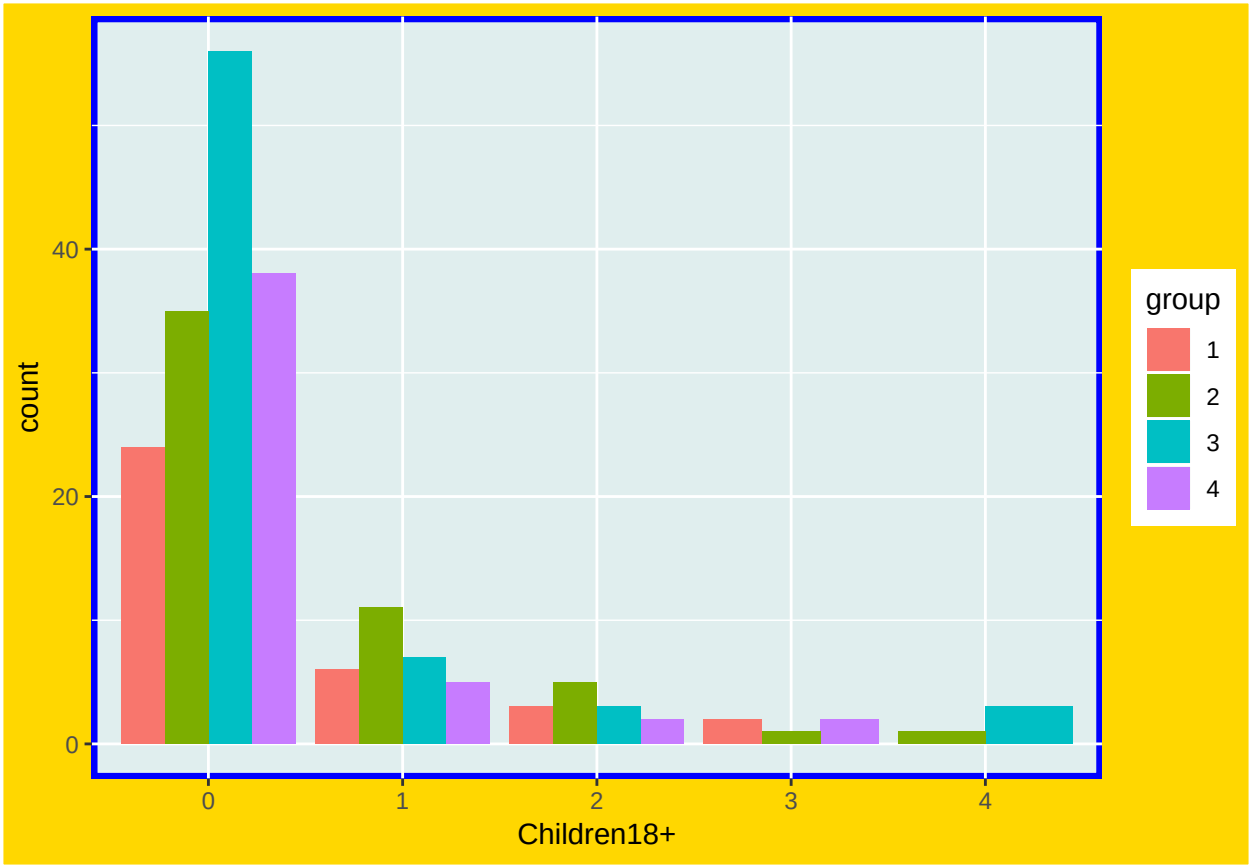

```
ggplot(drop_na(gdf,Children12_17),aes(group=group))+geom_bar(aes(Children12_17,fill=group),position = '
```



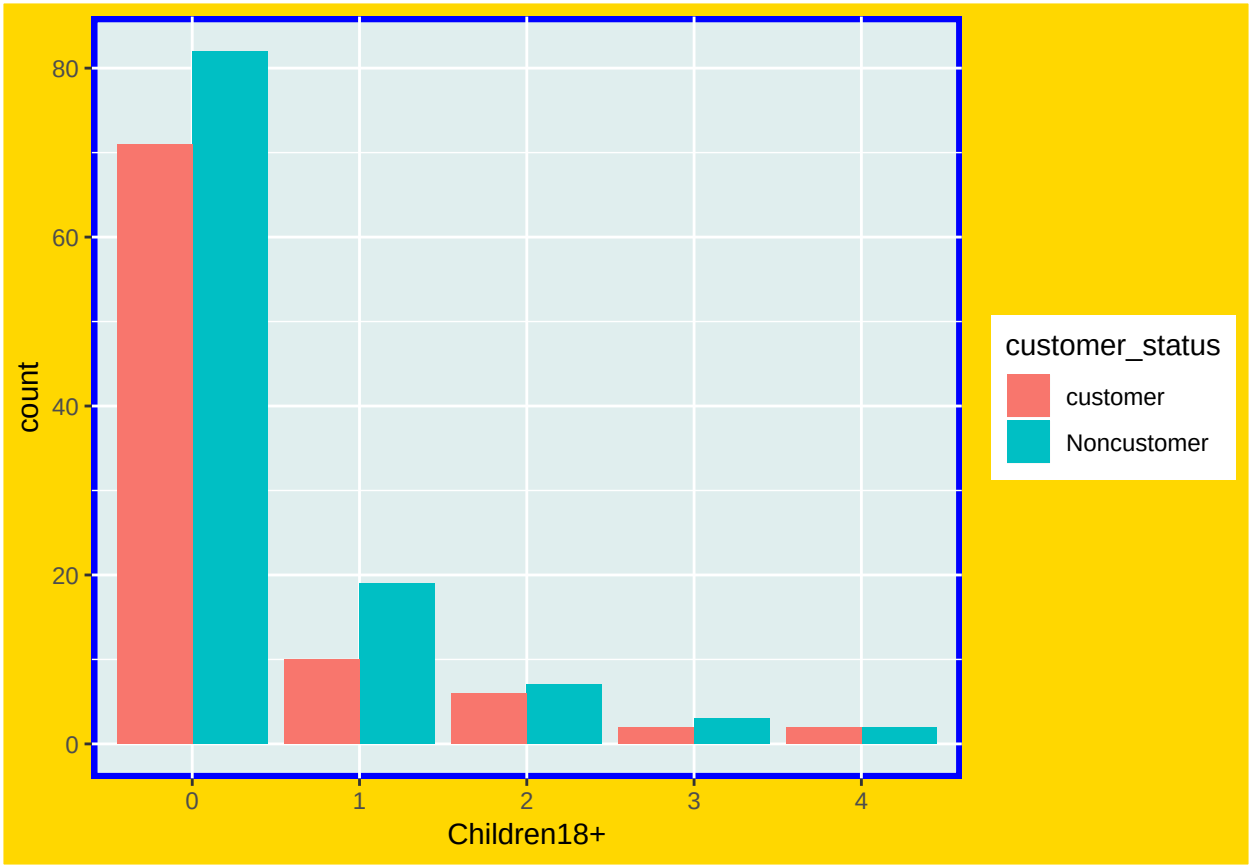

```
ggplot(drop_na(cdf,Children12_17),aes(group=customer_status))+geom_bar(aes(Children12_17,fill=customer_
```




```
ggplot(drop_na(gdf,`Children18+`),aes(group=group))+geom_bar(aes(`Children18+`,fill=group),position = '
```

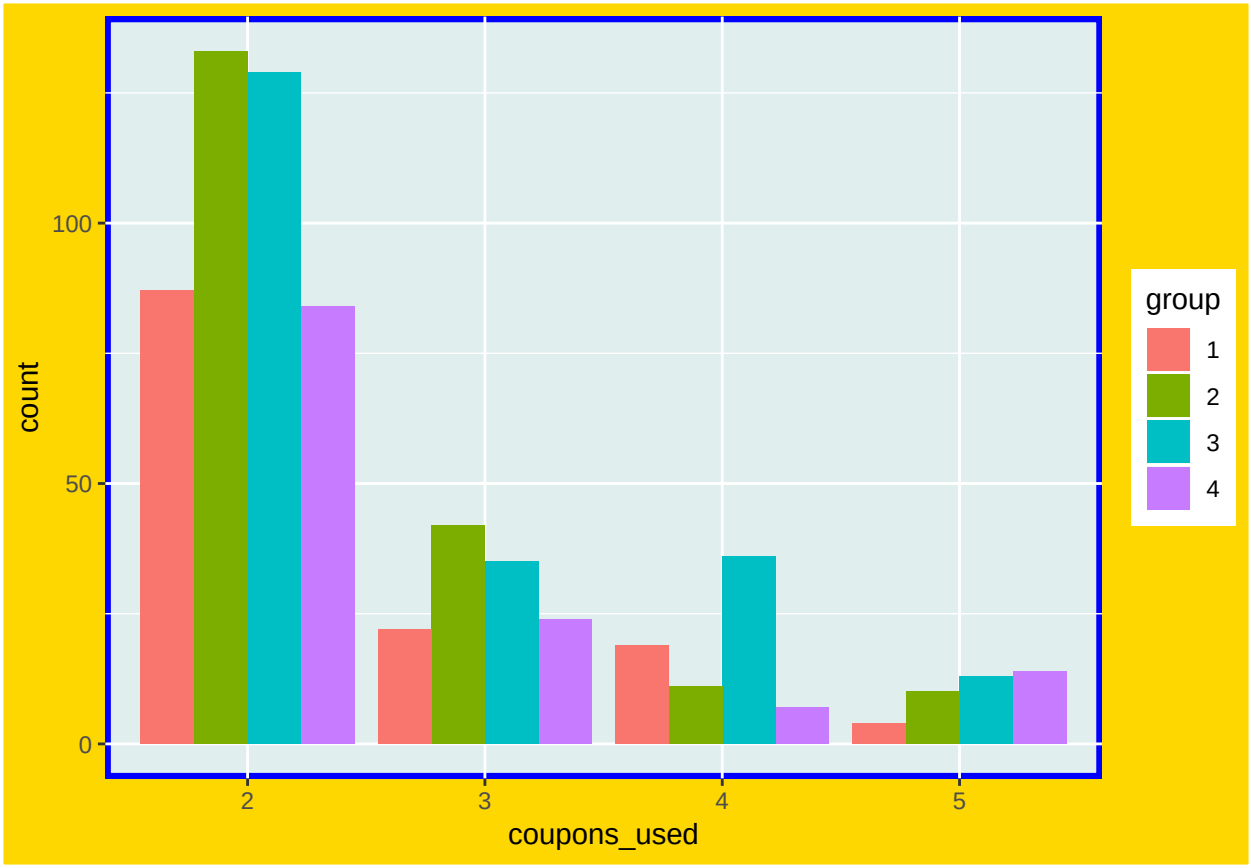



```
ggplot(drop_na(cdf,`Children18+`),aes(group=customer_status))+geom_bar(aes(`Children18+`,fill=customer_
```

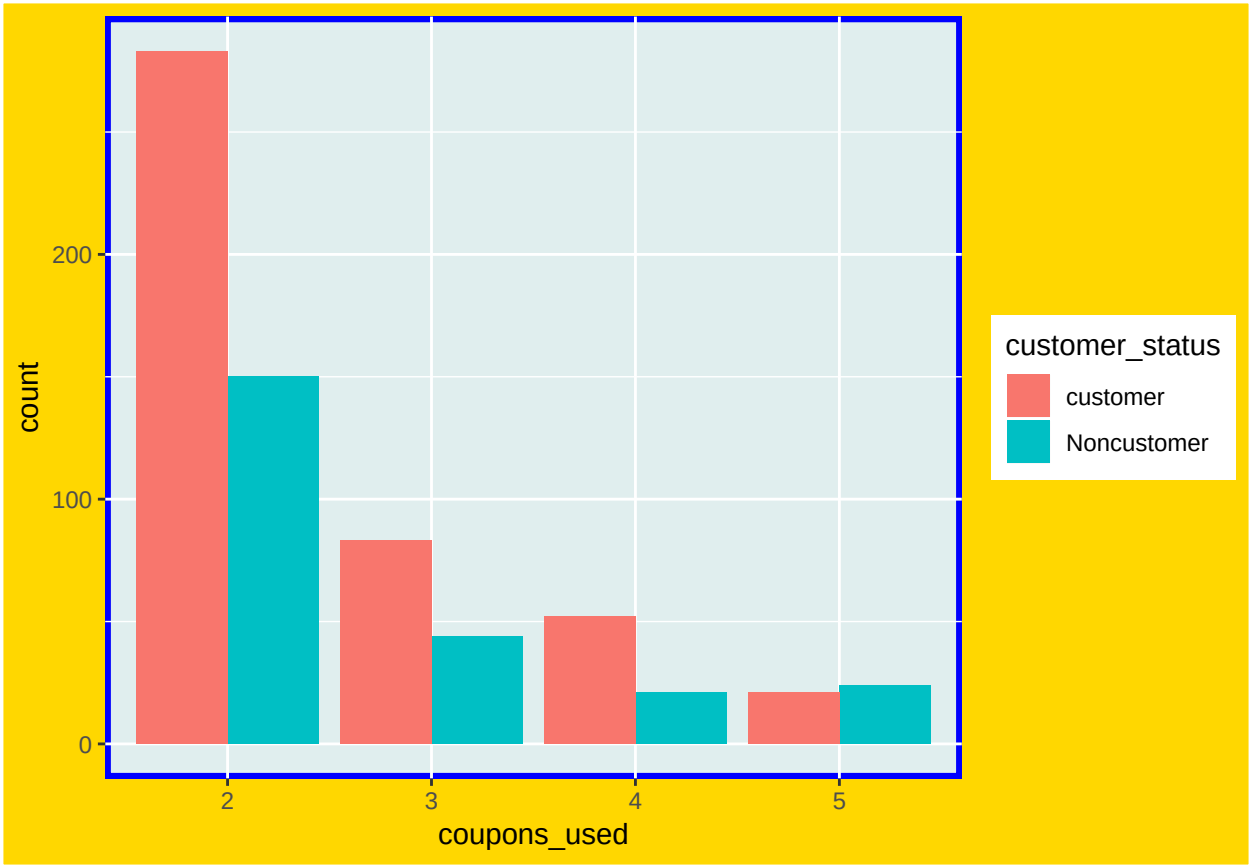


Graphs for coupons imformation

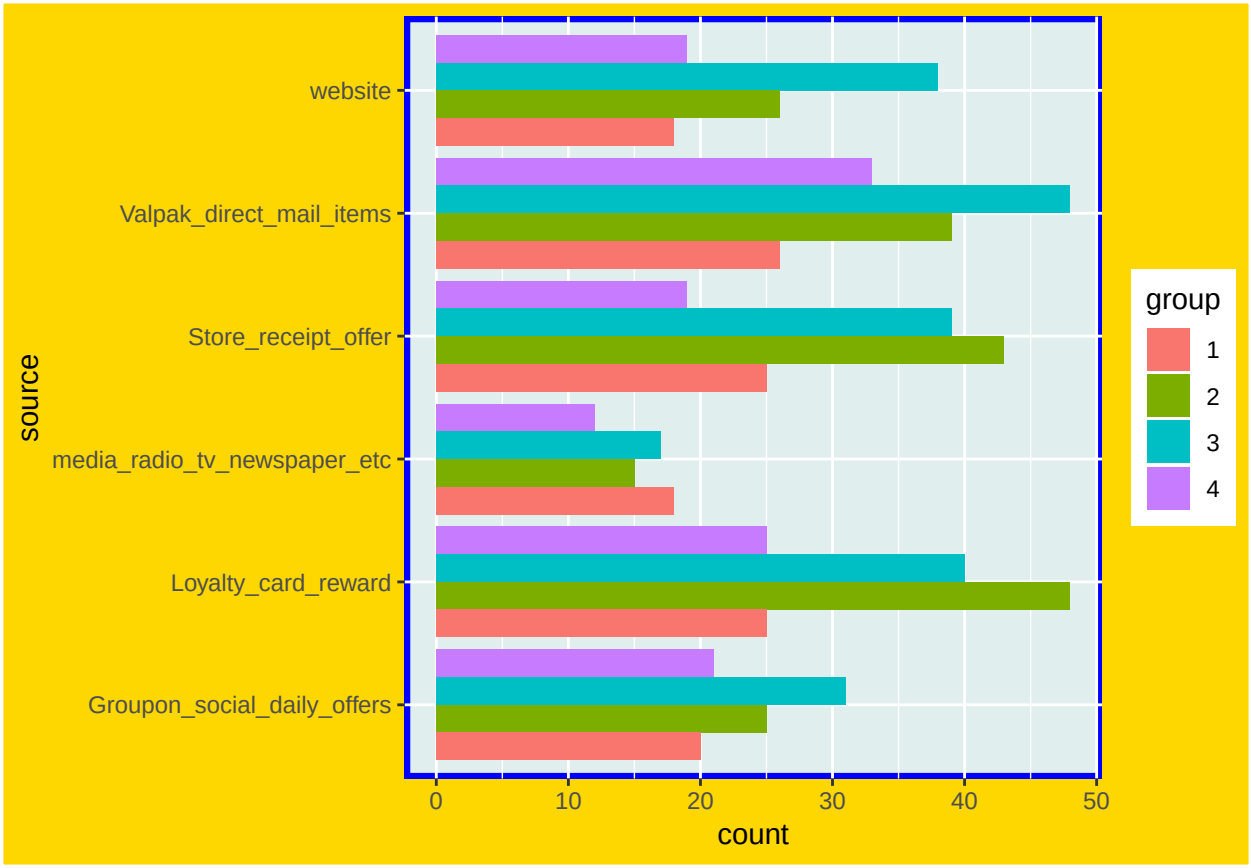
```
gdf=drop_na(coupons,group)
cdf=drop_na(coupons,customer_status)
ggplot(drop_na(gdf,coupons_used),aes(group=group))+geom_bar(aes(coupons_used,fill=group),position = 'do
```



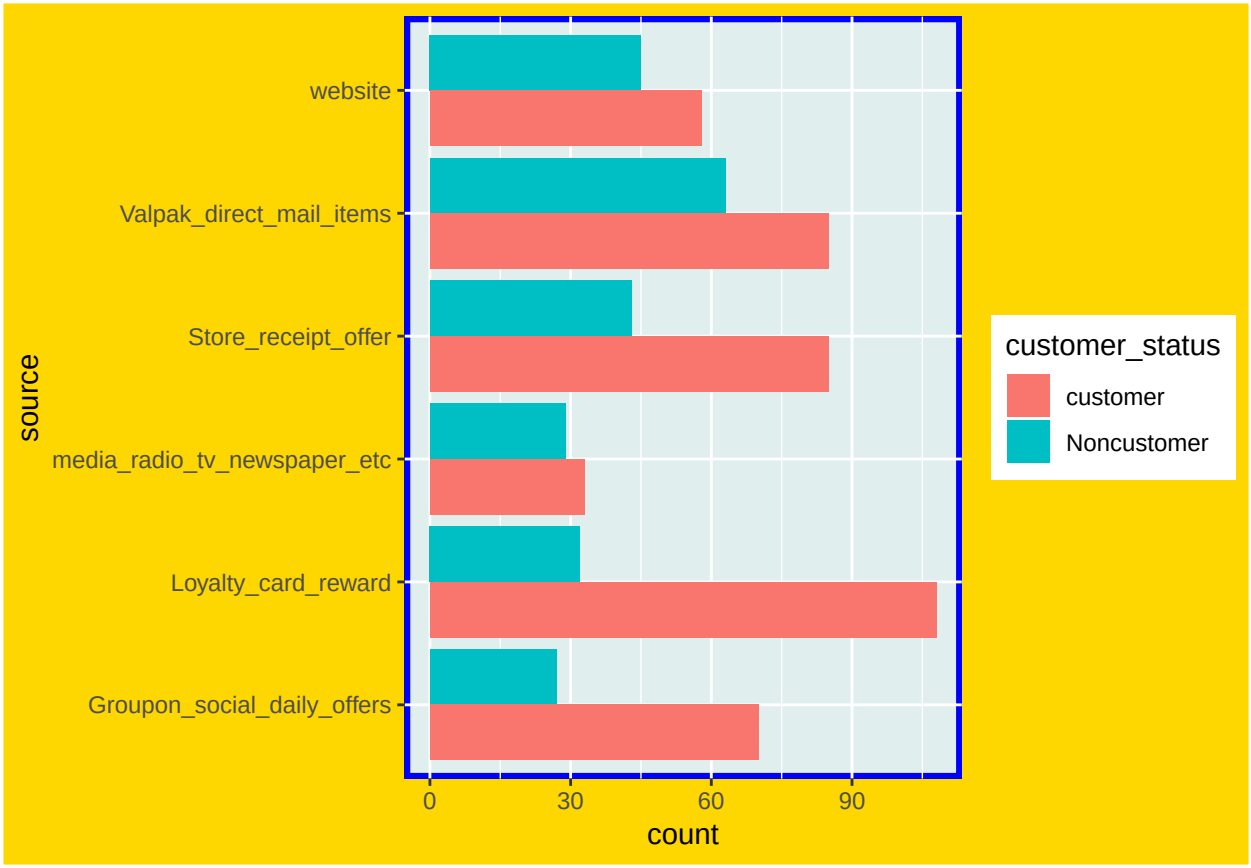

```
ggplot(drop_na(cdf,coupons_used),aes(group=customer_status))+geom_bar(aes(coupons_used,fill=customer_sta
```




```
ggplot(drop_na(gdf,source),aes(group=group))+geom_bar(aes(source,fill=group),position = 'dodge')+coord_
```

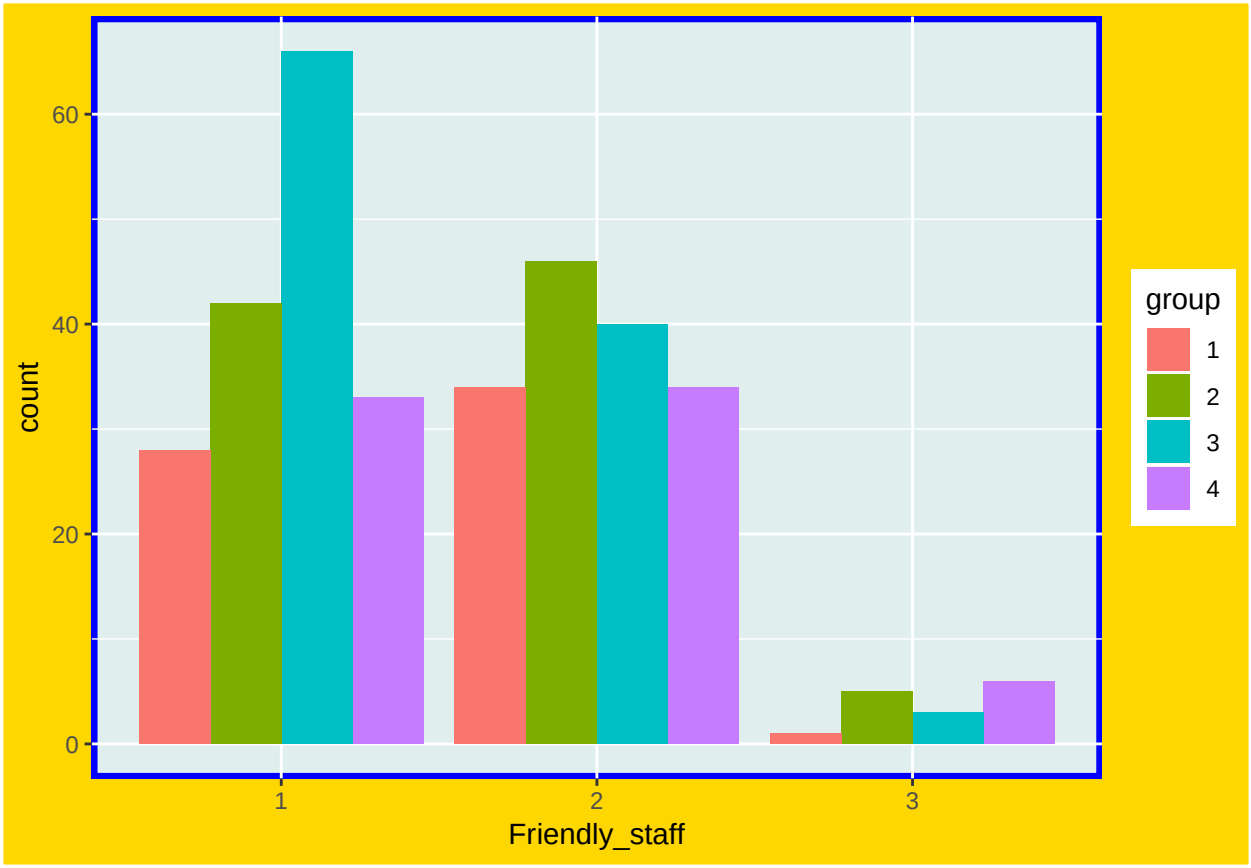



```
ggplot(drop_na(cdf,source),aes(group=customer_status))+geom_bar(aes(source,fill=customer_status),positi
```

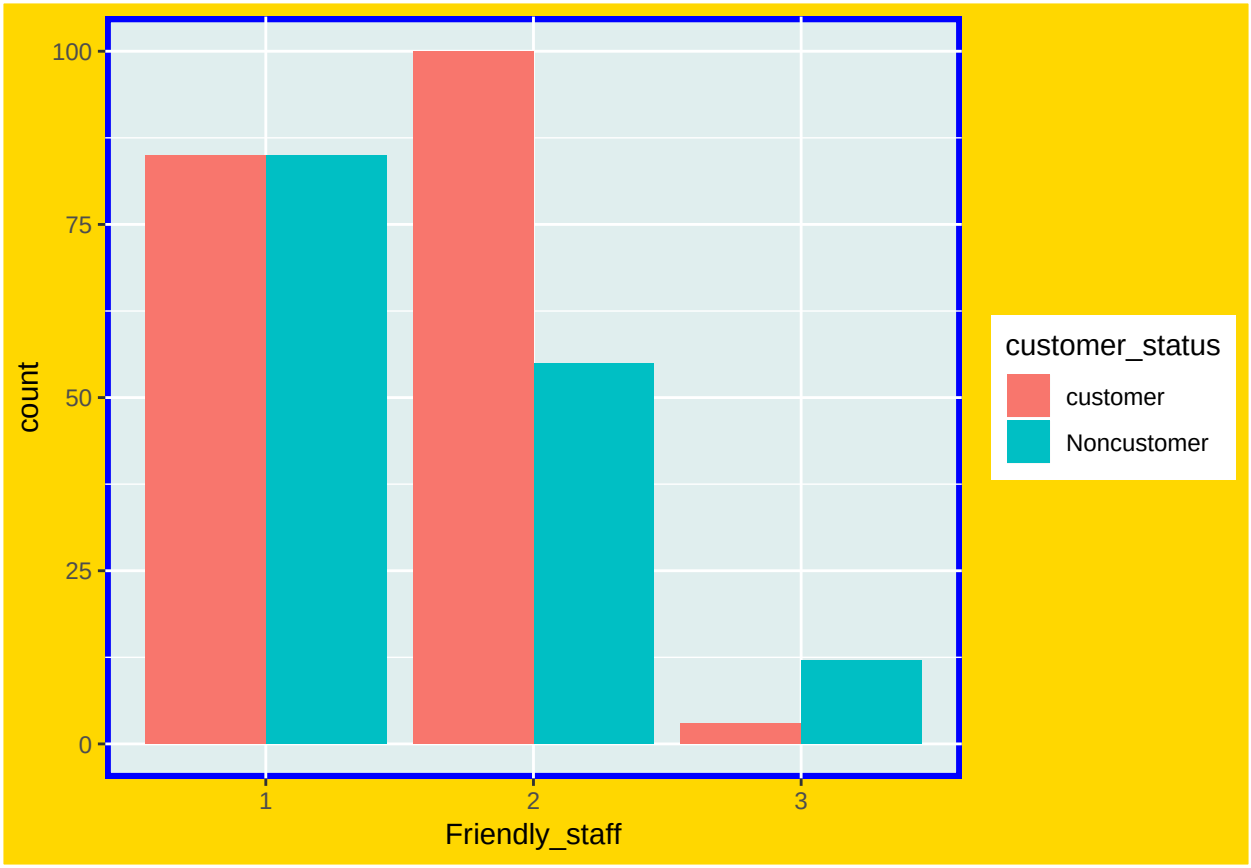


Graphs for the survey "Please indicate how important the following factors are when you visit a restaurant"

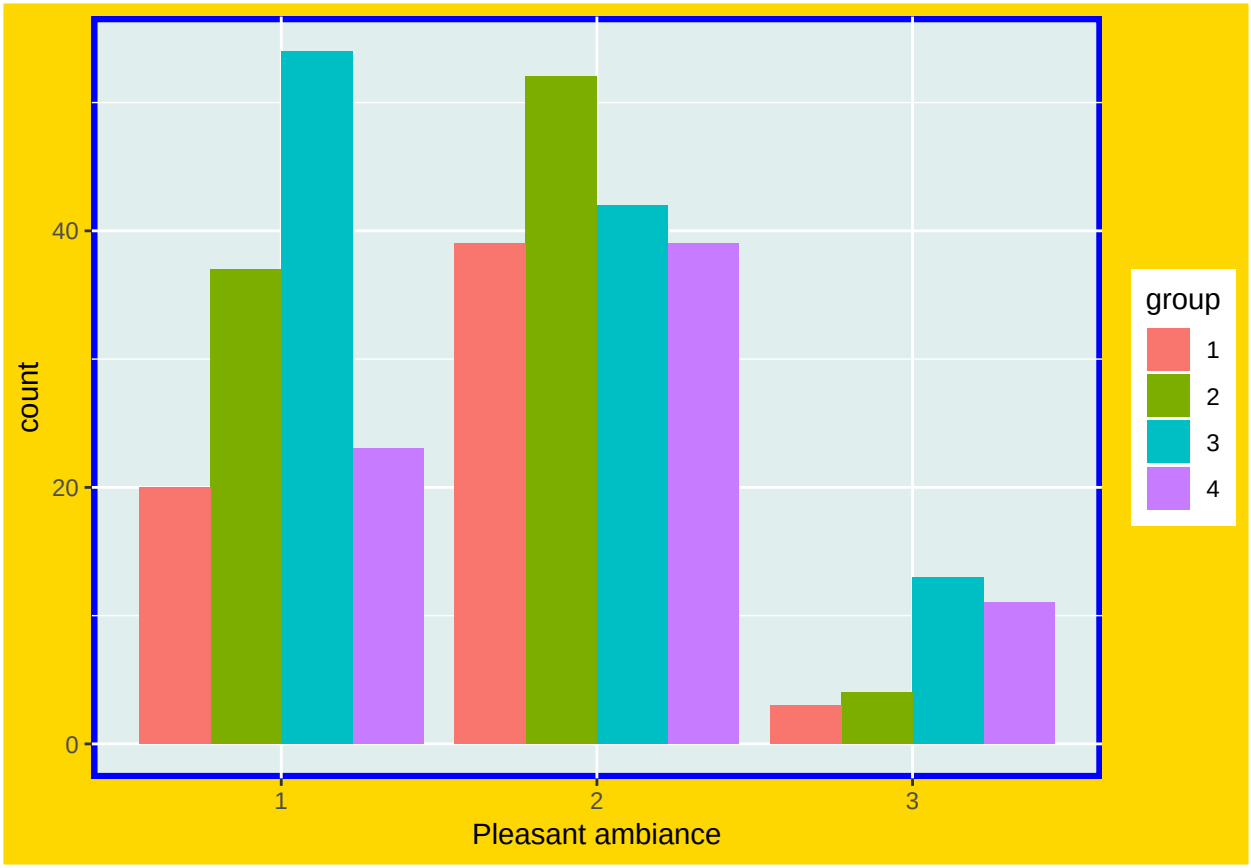
```
fdf=important %>% filter(Friendly_staff %in% as.character(1:3))
pdf=important %>% filter(`Pleasant ambiance` %in% as.character(1:3))
cdf=important %>% filter(Consistency_reliability %in% as.character(1:3))
pcdf=important %>% filter(Part_of_community %in% as.character(1:3))
ggplot(drop_na(fdf,group),aes(group=group))+geom_bar(aes(Friendly_staff,fill=group),position = 'dodge')
```



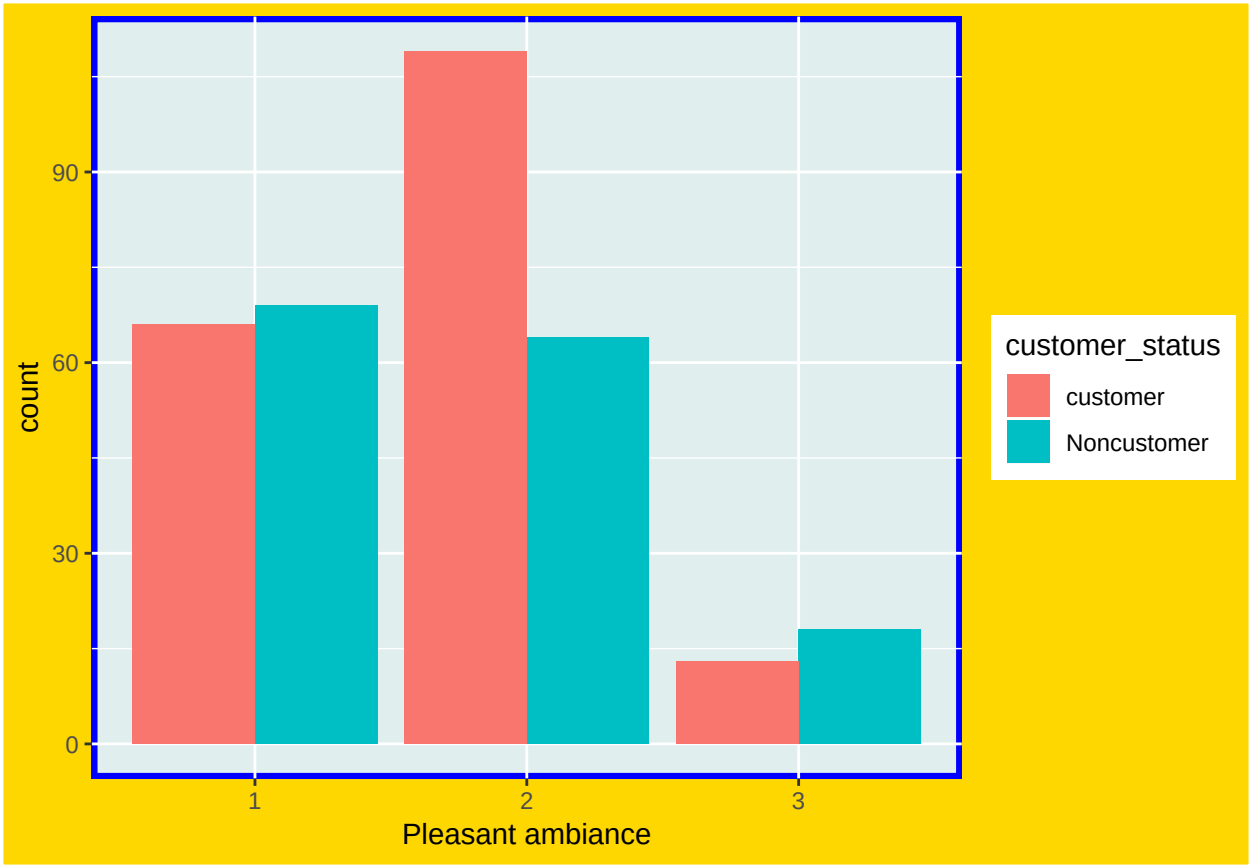

```
ggplot(drop_na(fdf,customer_status),aes(group=customer_status))+geom_bar(aes(Friendly_staff,fill=custome
```



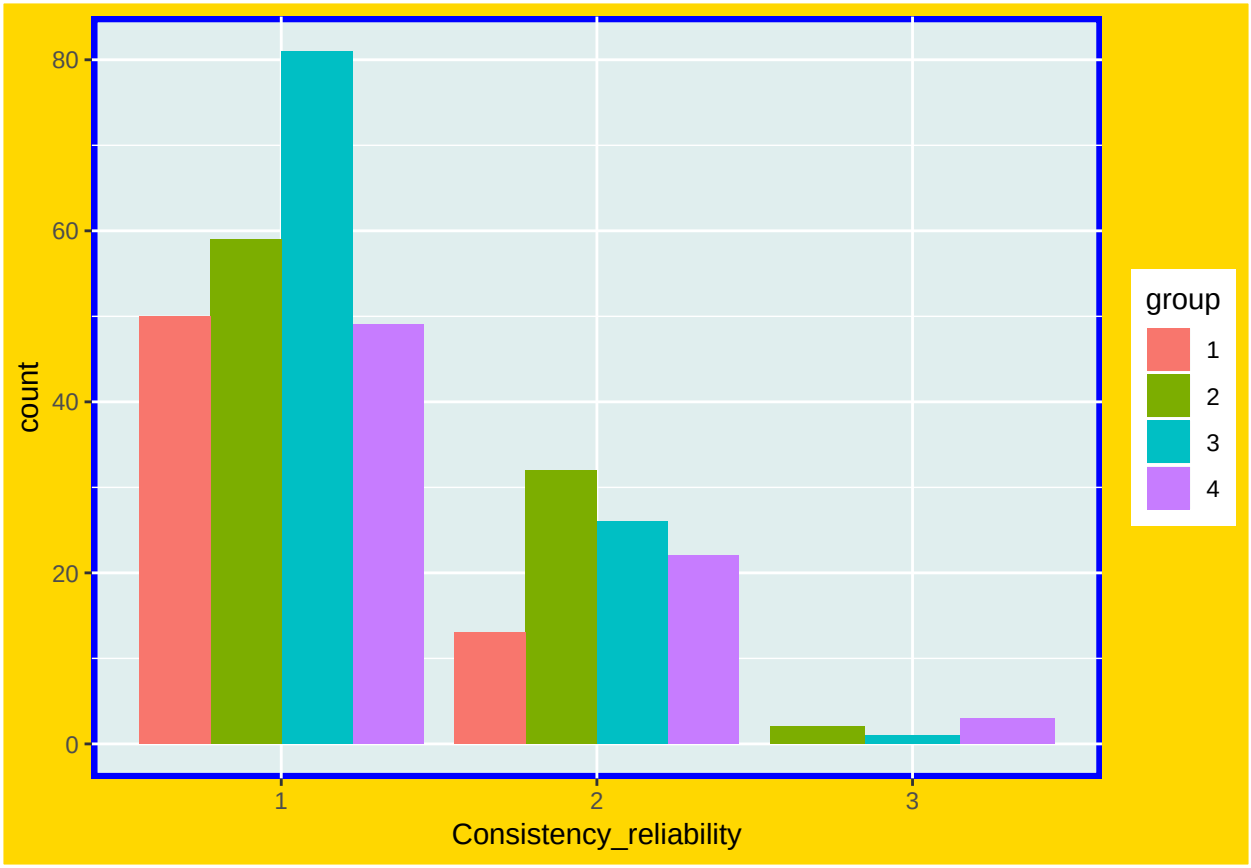

```
ggplot(drop_na(pdf,group),aes(group=group))+geom_bar(aes(`Pleasant ambiance`,fill=group),position = 'do
```



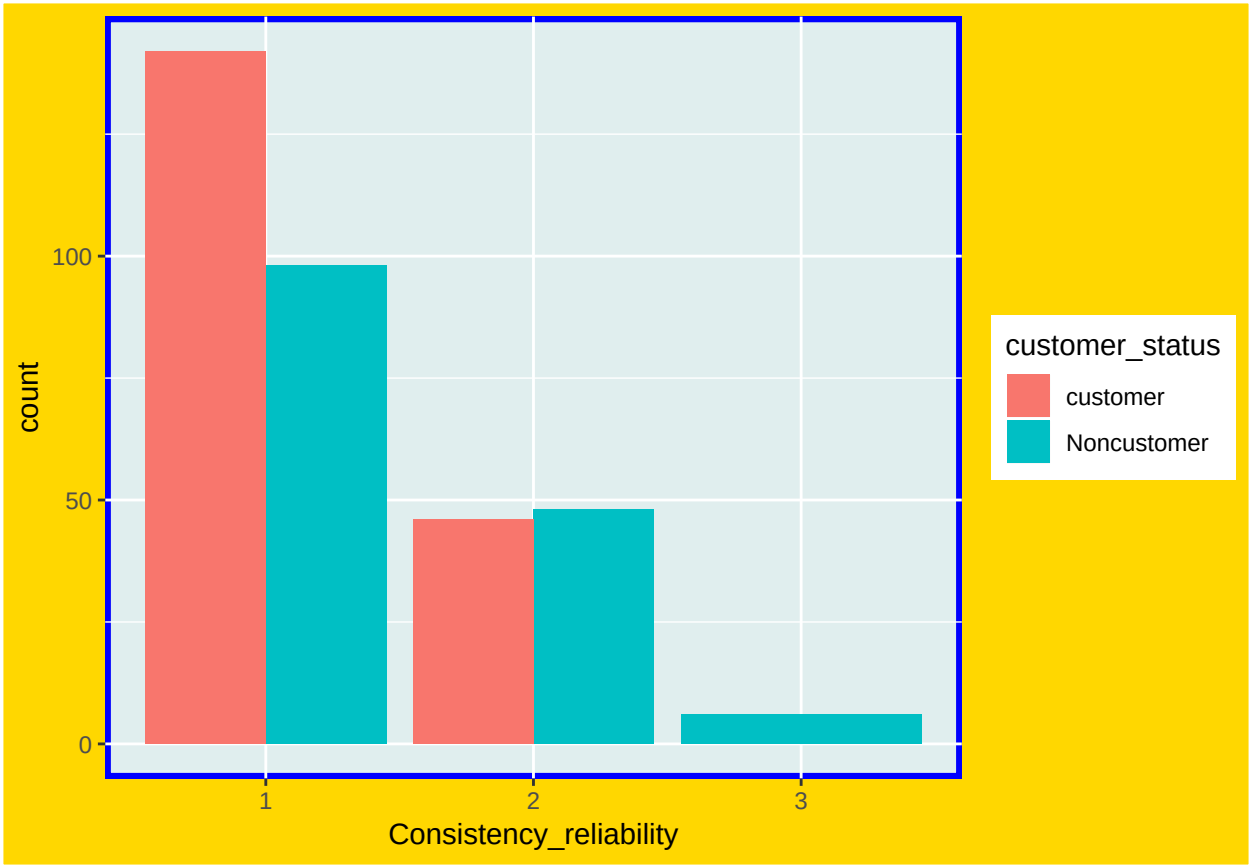

```
ggplot(drop_na(pdf,customer_status),aes(group=customer_status))+geom_bar(aes(`Pleasant ambiance`,fill=cu
```



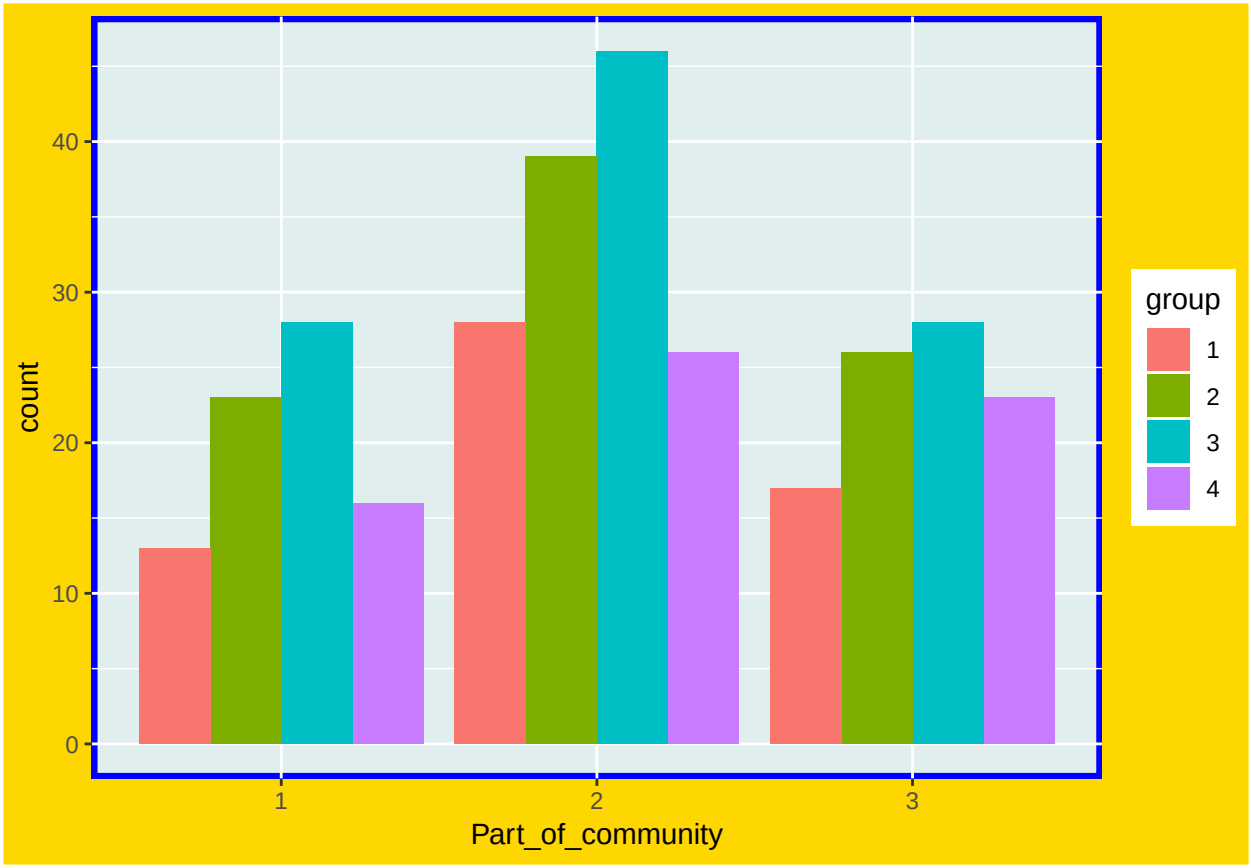

```
ggplot(drop_na(cdf,group),aes(group=group))+geom_bar(aes(Consistency_reliability,fill=group),position =
```



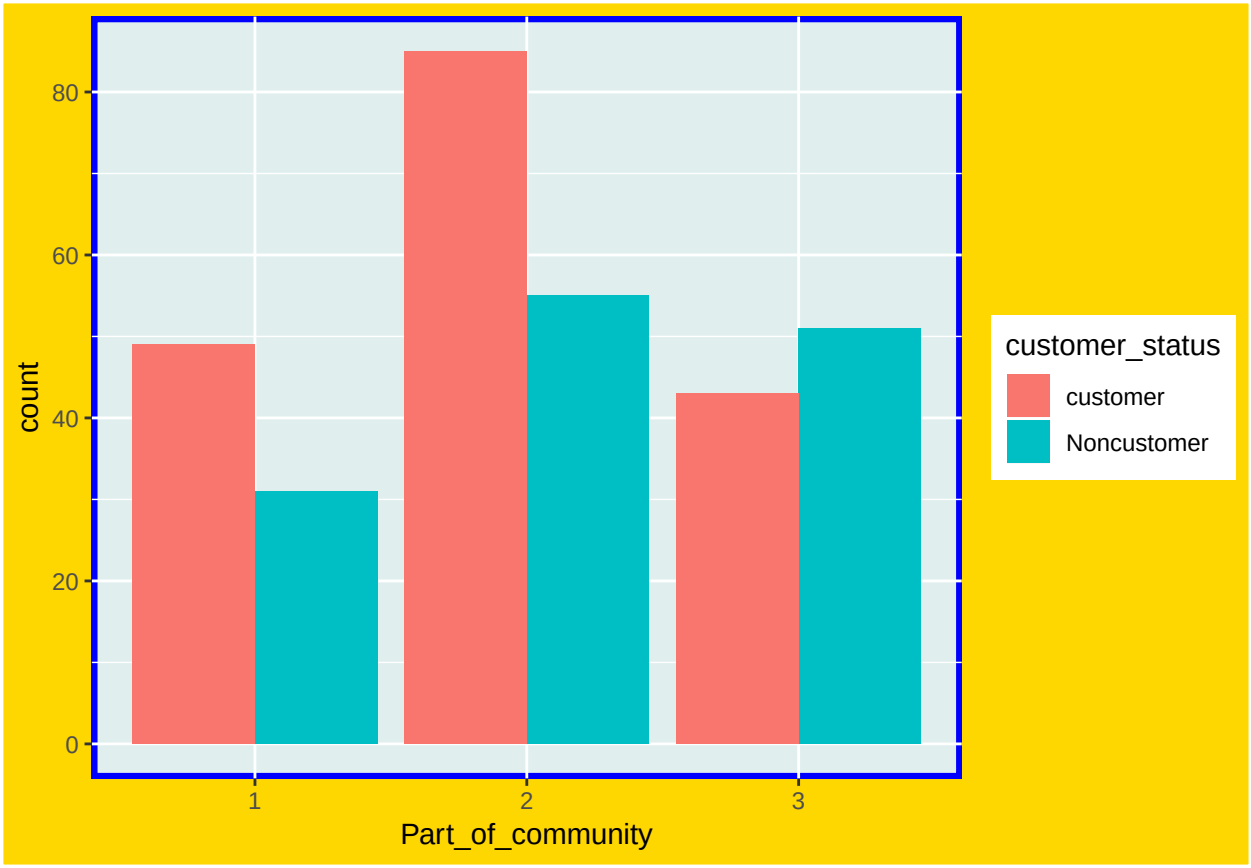

```
ggplot(drop_na(cdf,customer_status),aes(group=customer_status))+geom_bar(aes(Consistency_reliability,fi
```




```
ggplot(drop_na(pcdf,group),aes(group=group))+geom_bar(aes(Part_of_community,fill=group),position = 'dod
```

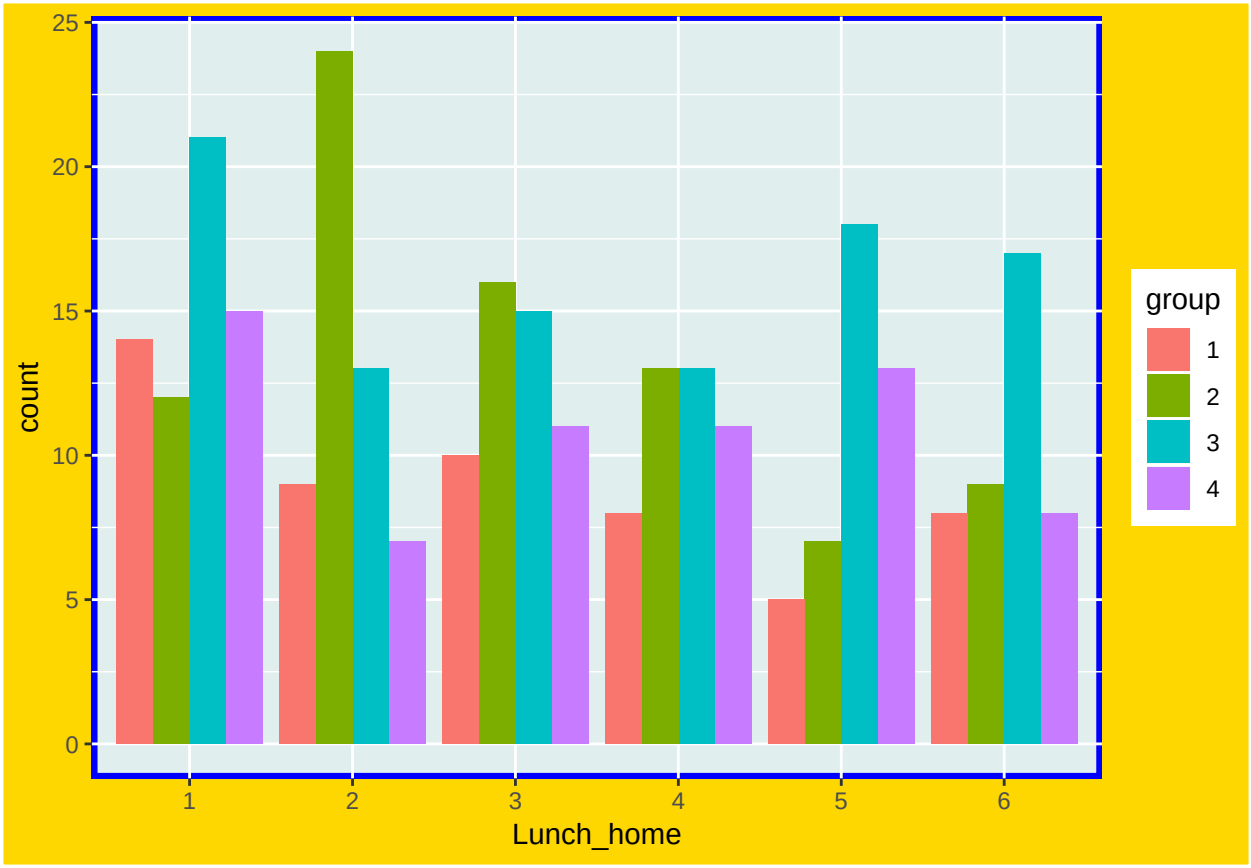



```
ggplot(drop_na(pcdf,customer_status),aes(group=customer_status))+geom_bar(aes(Part_of_community,fill=cus
```

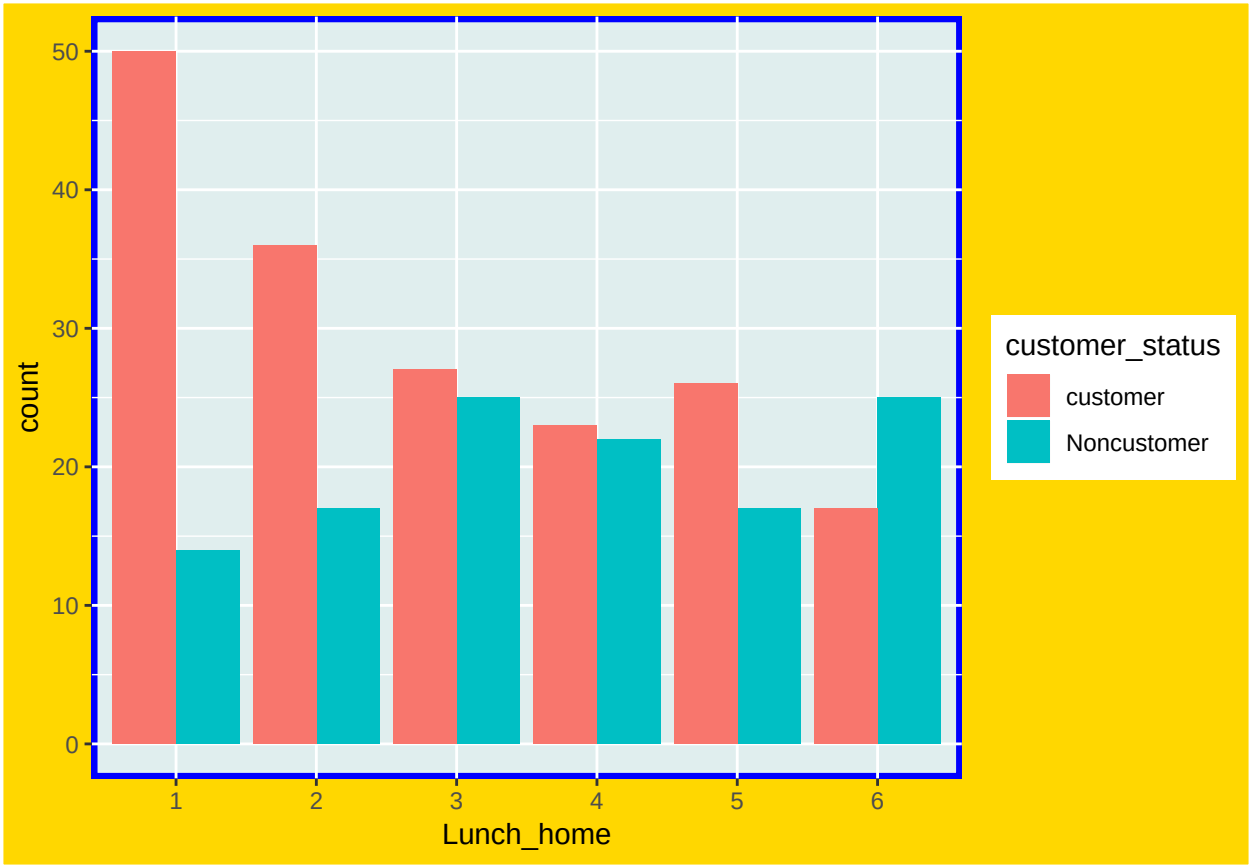


Graphs for the survey "How many times in the last week did you do the following?"

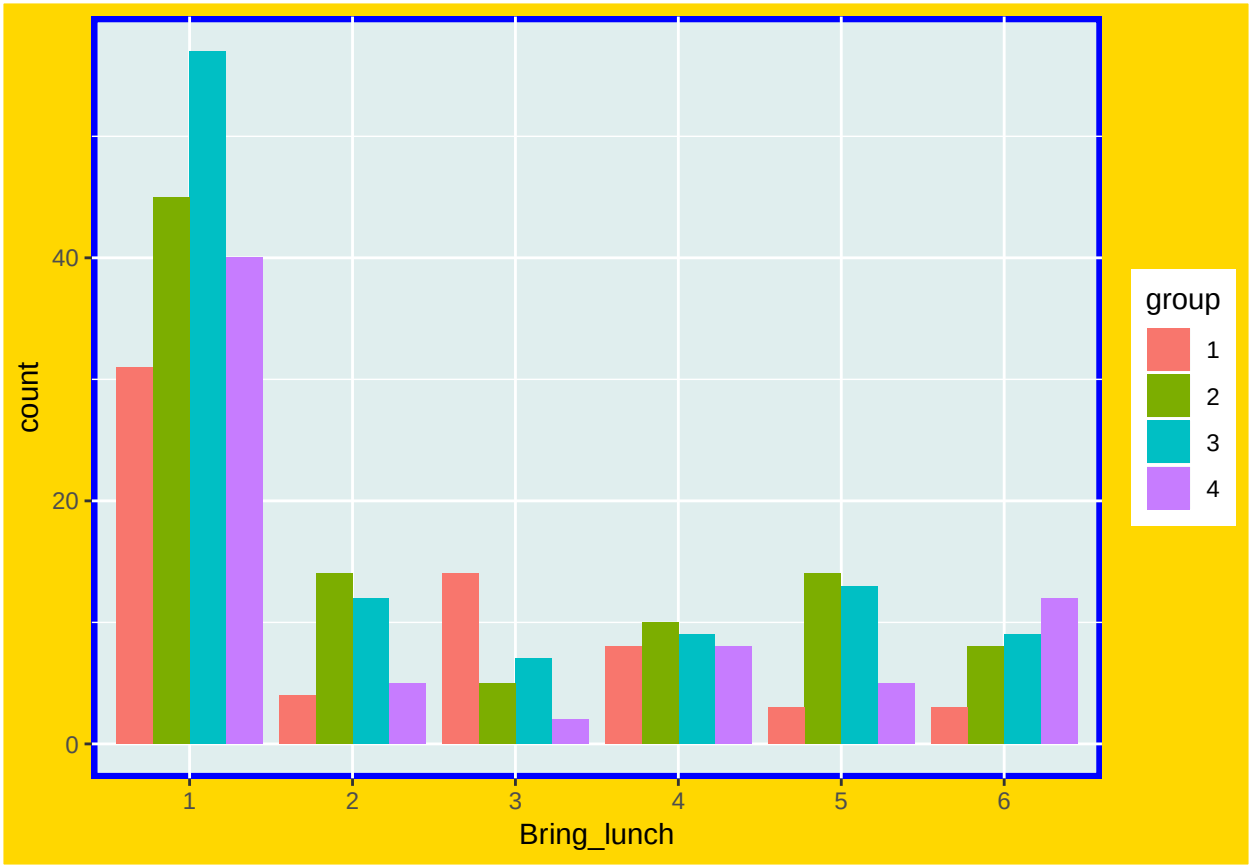
```
home=lunch %>% filter(Lunch_home %in% as.character(1:6))
bring=lunch %>% filter(Bring_lunch %in% as.character(1:6))
wk=lunch %>% filter(Buy_lunch_workplace %in% as.character(1:6))
out=lunch %>% filter(Buy_lunch_out %in% as.character(1:6))
skip=lunch %>% filter(Skip_lunch %in% as.character(1:6))
ggplot(drop_na(home,group),aes(group=group))+geom_bar(aes(Lunch_home,fill=group),position = 'dodge')
```



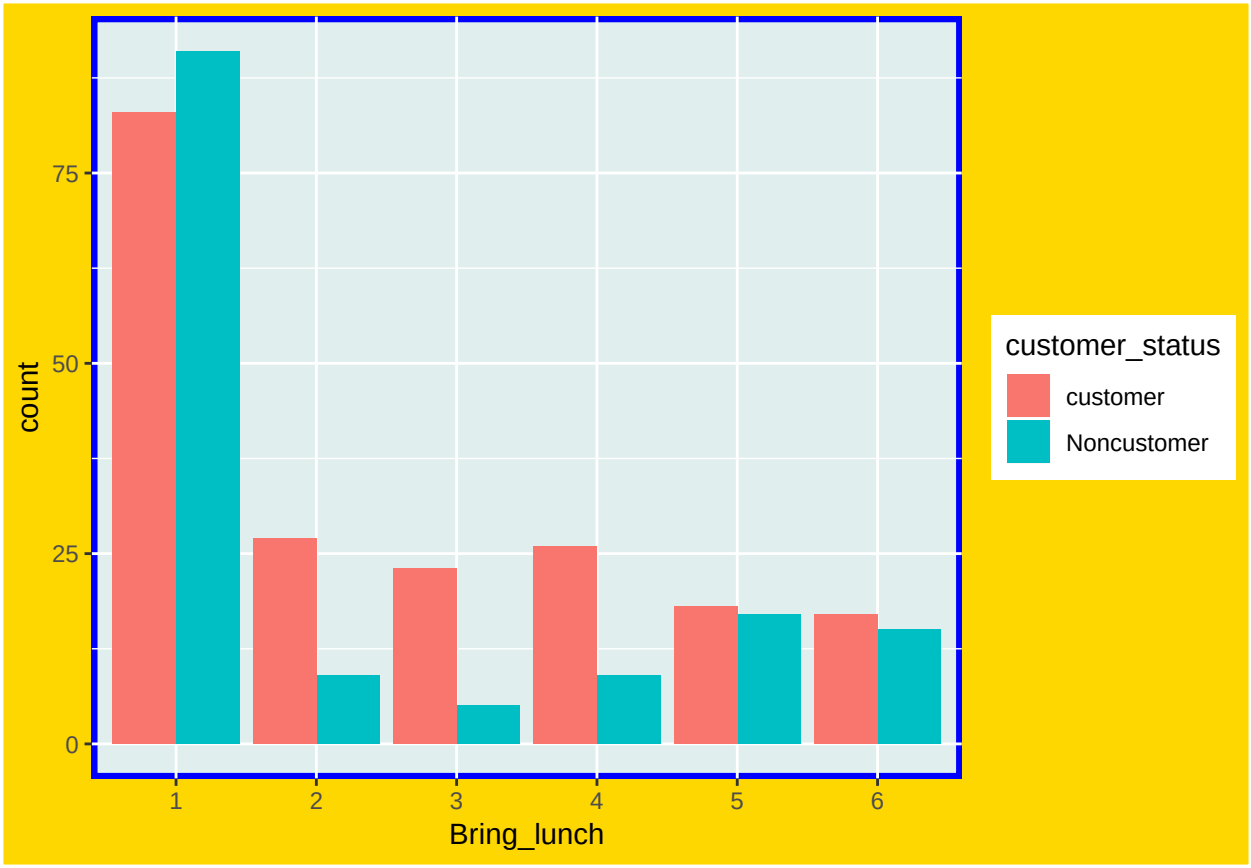

```
ggplot(drop_na(home,customer_status),aes(group=customer_status))+geom_bar(aes(Lunch_home,fill=customer_s
```



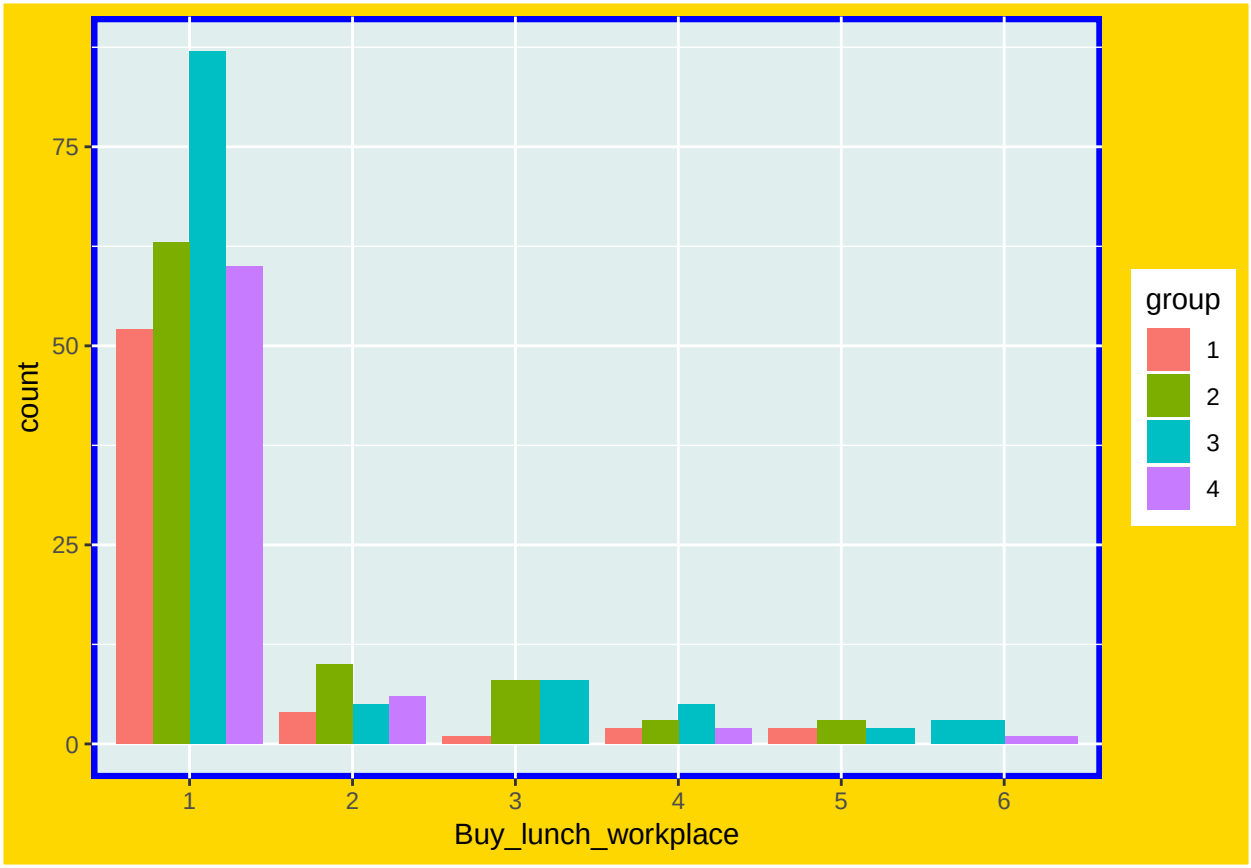

```
ggplot(drop_na(bring,group),aes(group=group))+geom_bar(aes(Bring_lunch,fill=group),position = 'dodge')
```



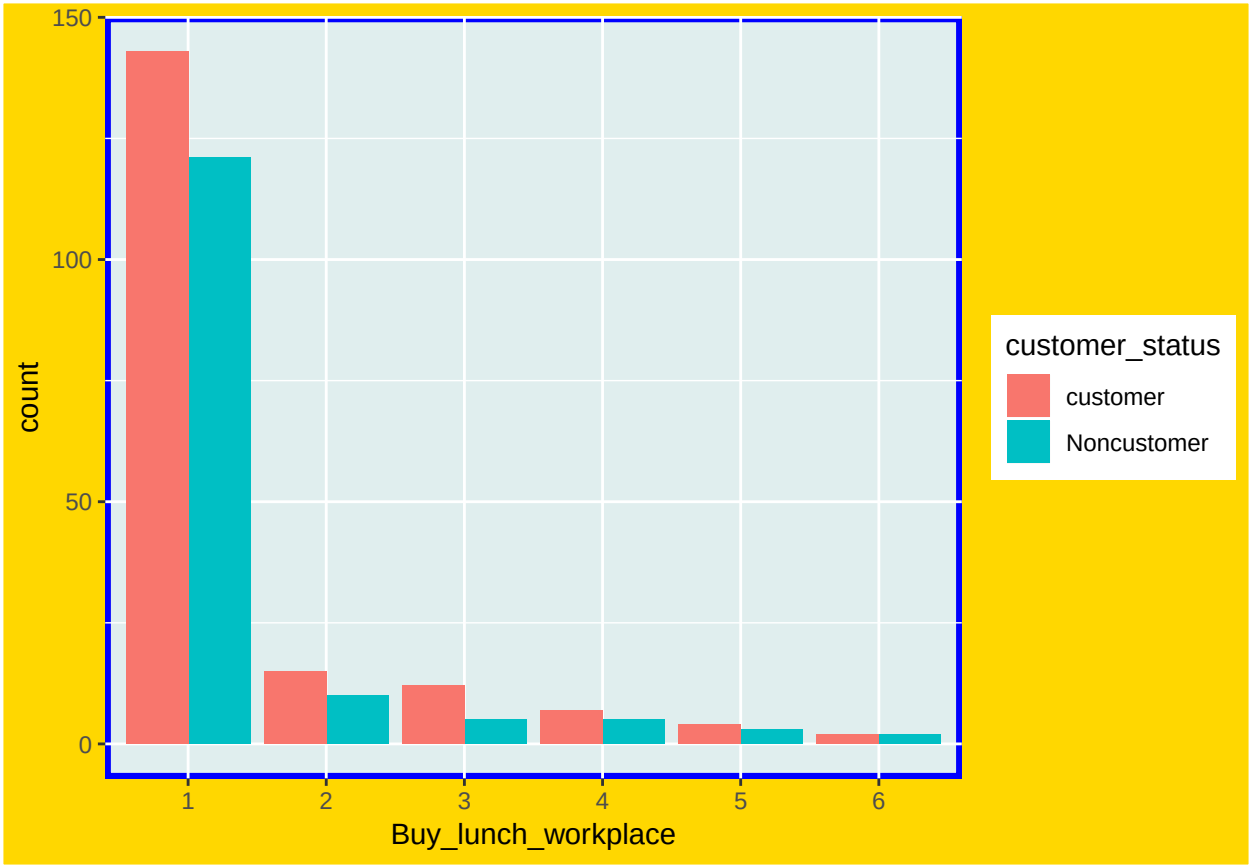

```
ggplot(drop_na(bring,customer_status),aes(group=customer_status))+geom_bar(aes(Bring_lunch,fill=customer
```



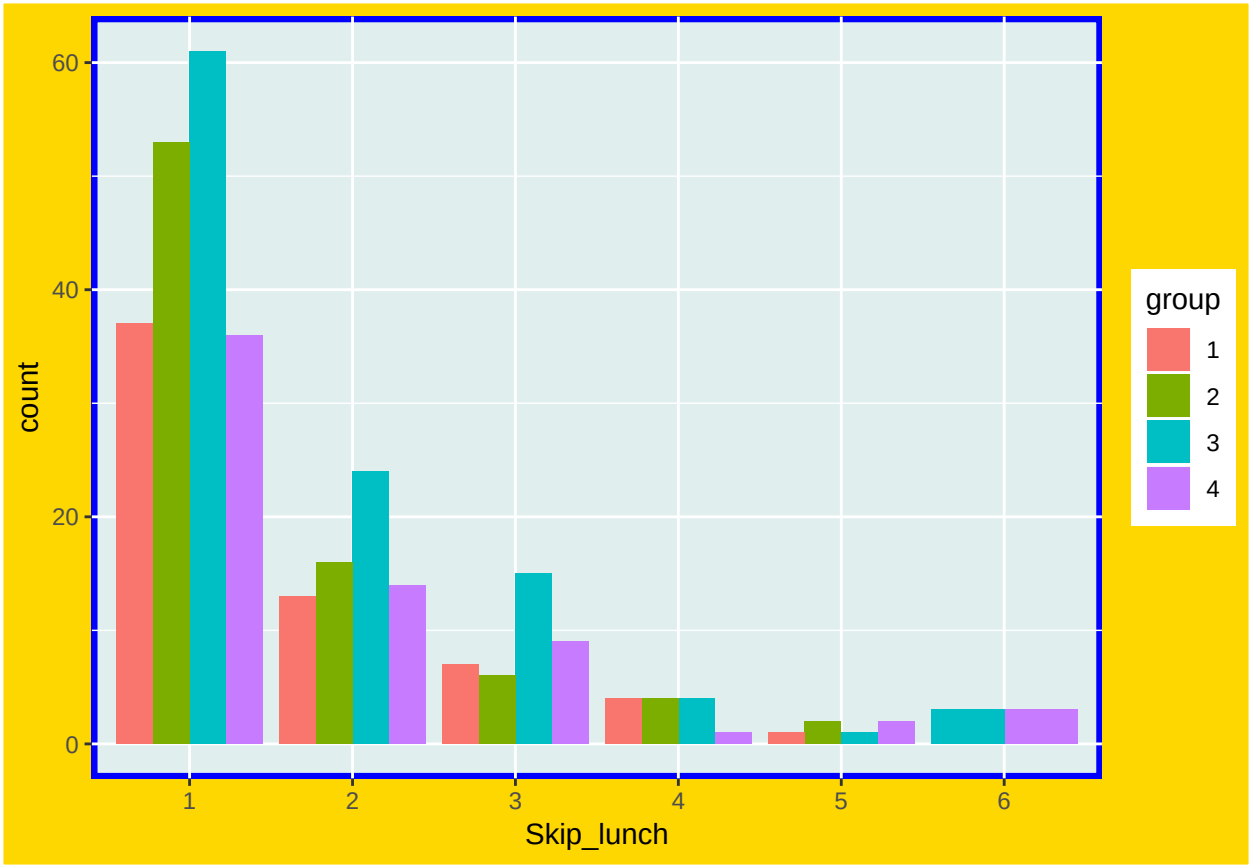

```
ggplot(drop_na(wk,group),aes(group=group))+geom_bar(aes(Buy_lunch_workplace,fill=group),position = 'dod
```



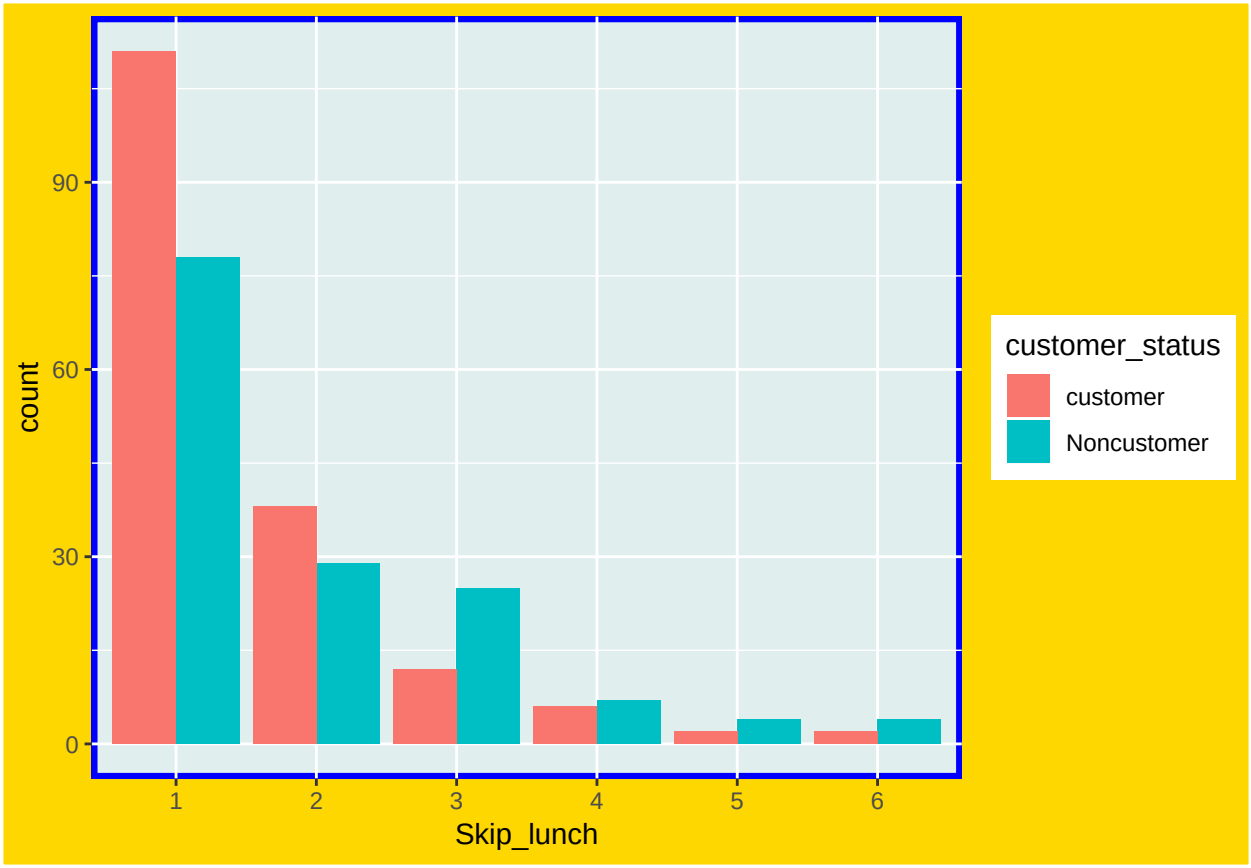

```
ggplot(drop_na(wk,customer_status),aes(group=customer_status))+geom_bar(aes(Buy_lunch_workplace,fill=cus
```




```
ggplot(drop_na(skip,group),aes(group=group))+geom_bar(aes(Skip_lunch,fill=group),position = 'dodge')
```

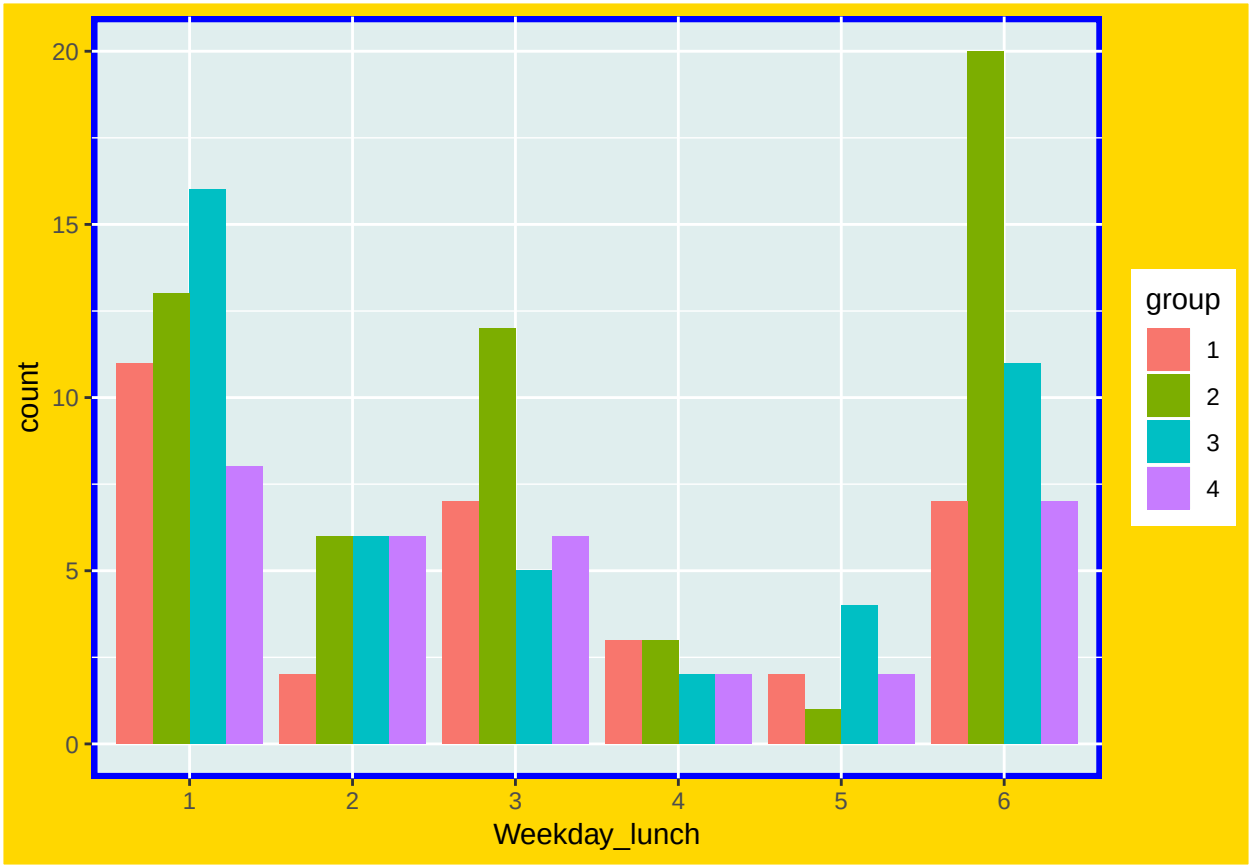



```
ggplot(drop_na(skip,customer_status),aes(group=customer_status))+geom_bar(aes(Skip_lunch,fill=customer_s
```

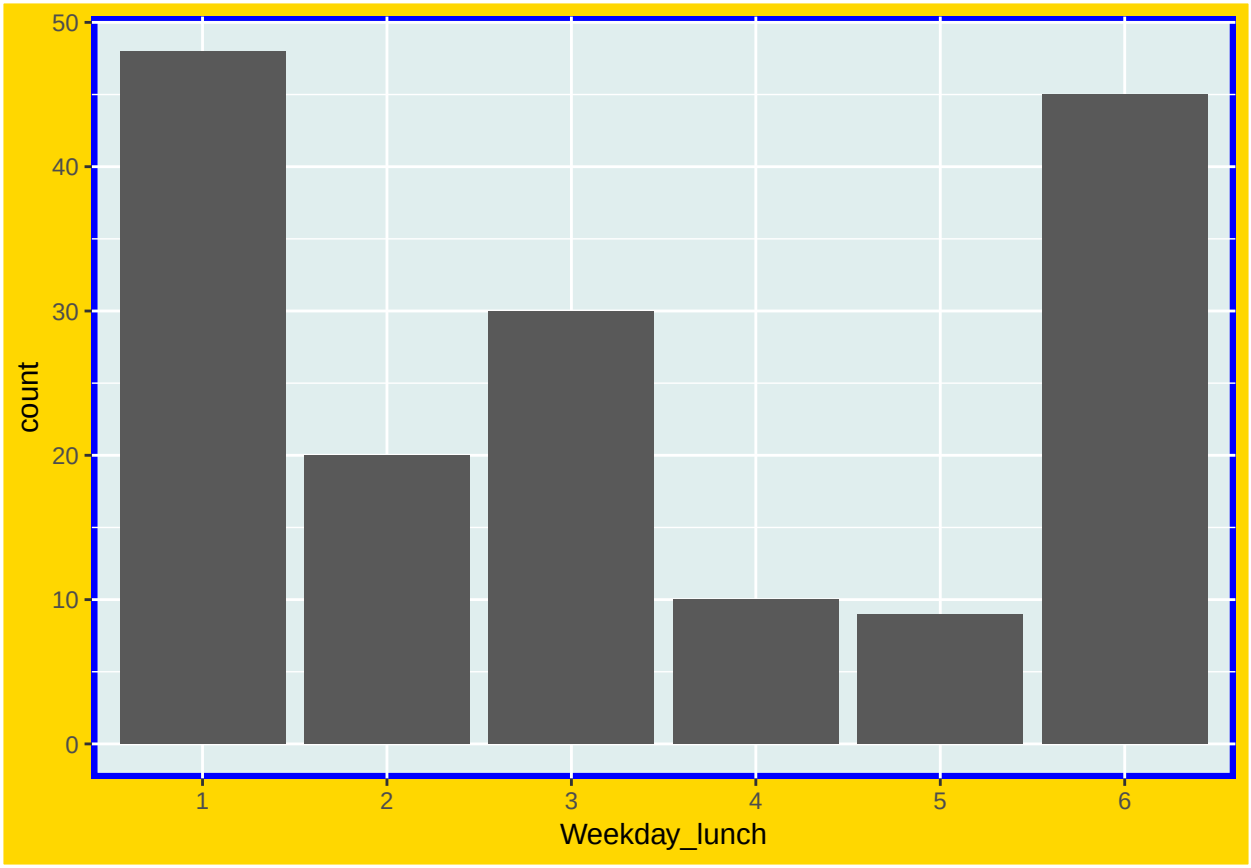


Graphs for the survey "In the last month, how often have you visited Sticks for the following occasions?"

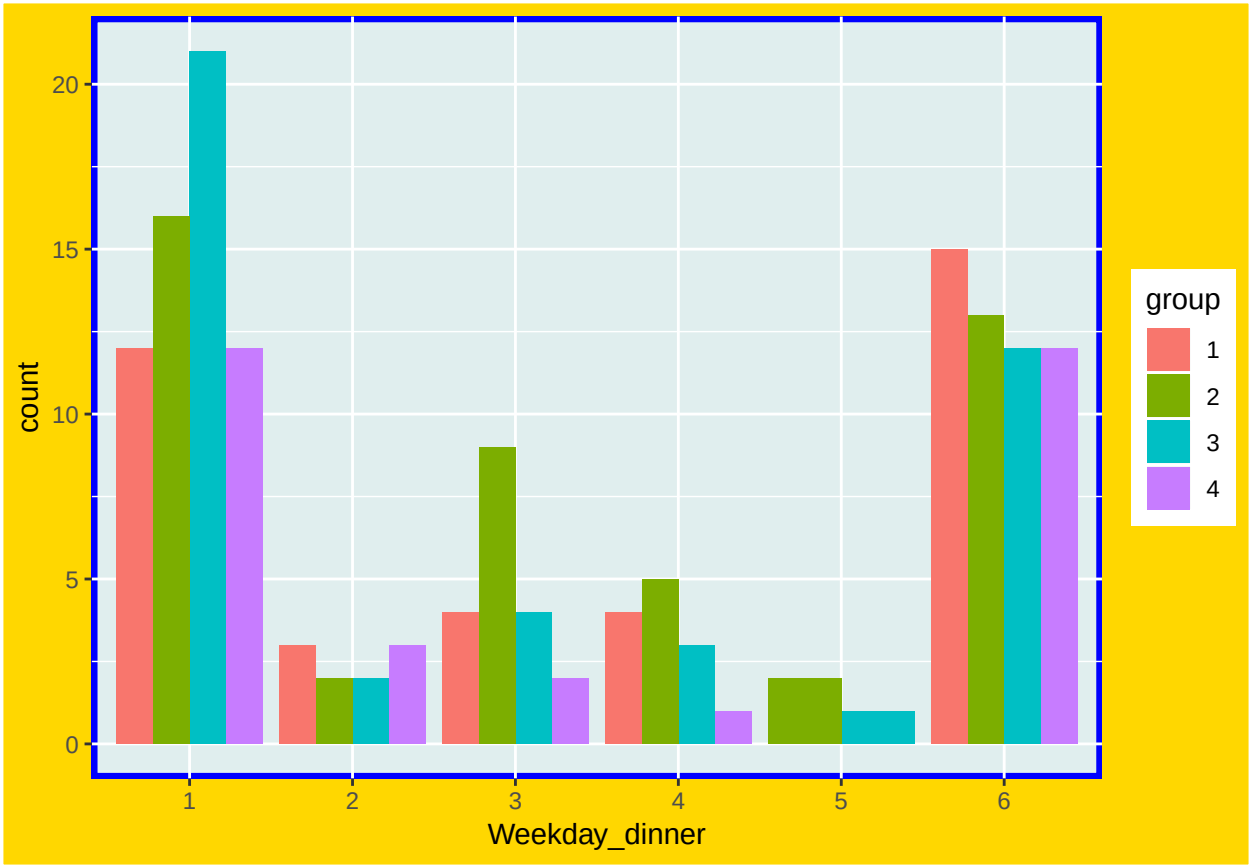
```
wdl=visit_reason %>% filter(Weekday_lunch %in% as.character(1:6))
wdd=visit_reason %>% filter(Weekday_dinner %in% as.character(1:6))
wel=visit_reason %>% filter(Weekend_lunch %in% as.character(1:6))
wed=visit_reason %>% filter(Weekend_dinner %in% as.character(1:6))
event=visit_reason %>% filter(Event %in% as.character(1:6))
as=visit_reason %>% filter(After_school_snack_sports_practice_event %in% as.character(1:6))
ggplot(drop_na(wdl,group),aes(group=group))+geom_bar(aes(Weekday_lunch,fill=group),position = 'dodge')
```



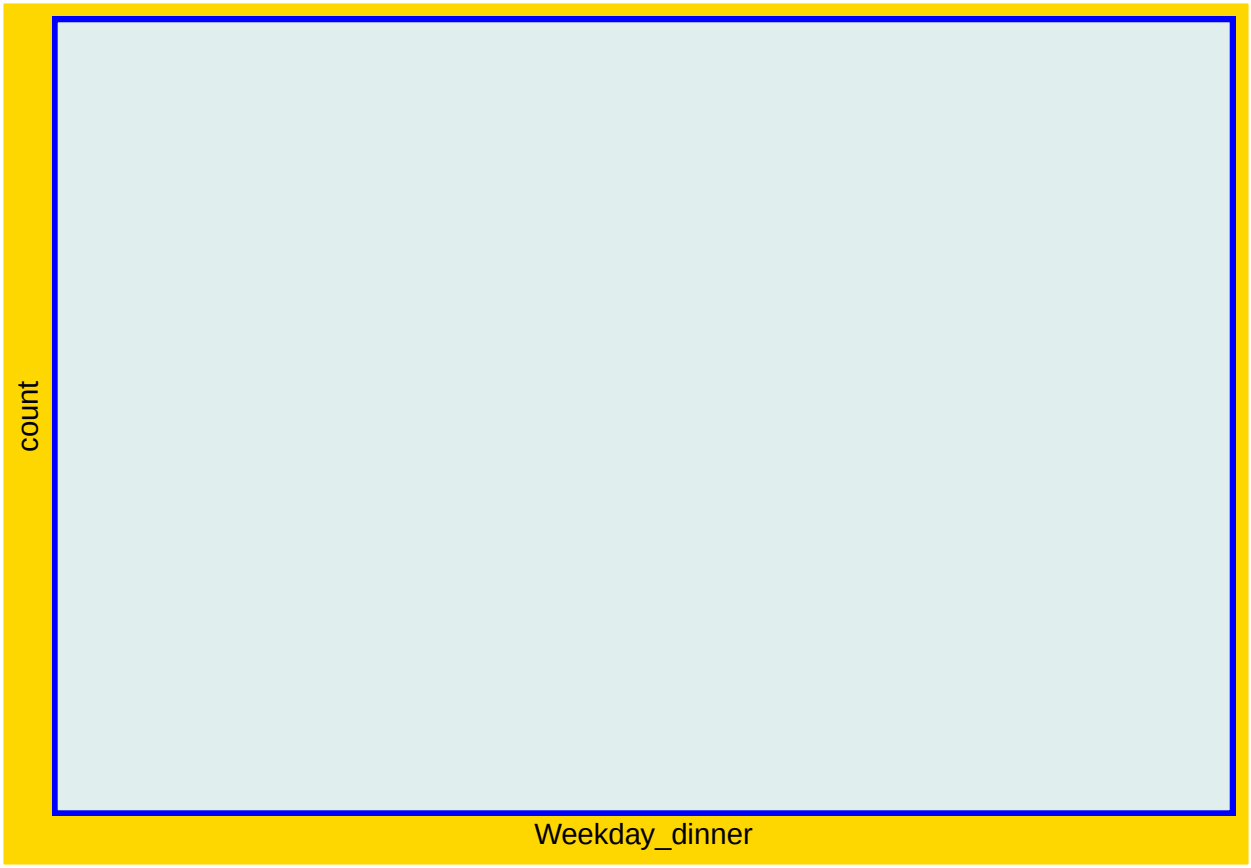

```
ggplot(drop_na(wdl,group))+geom_bar(aes(Weekday_lunch),position = 'dodge')
```



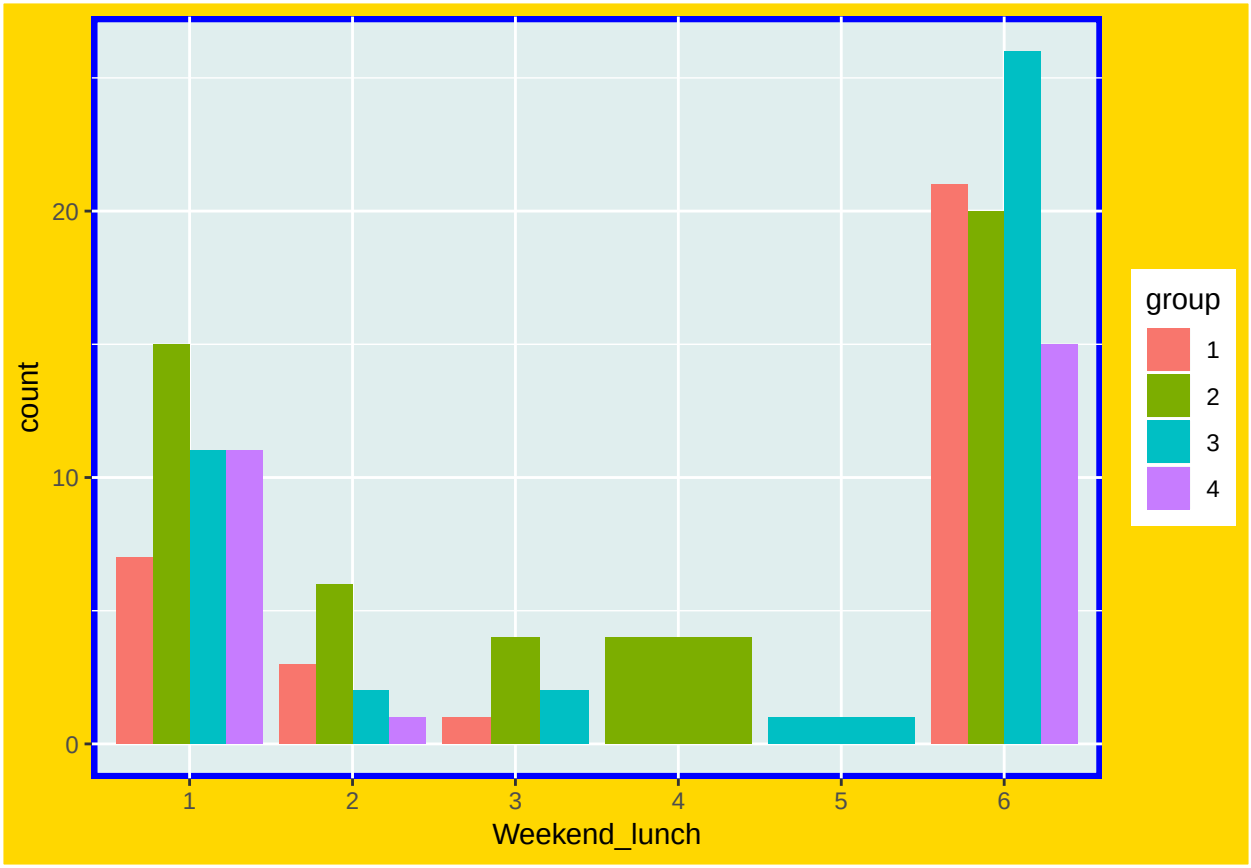

```
ggplot(drop_na(wdd,group),aes(group=group))+geom_bar(aes(Weekday_dinner,fill=group),position = 'dodge')
```



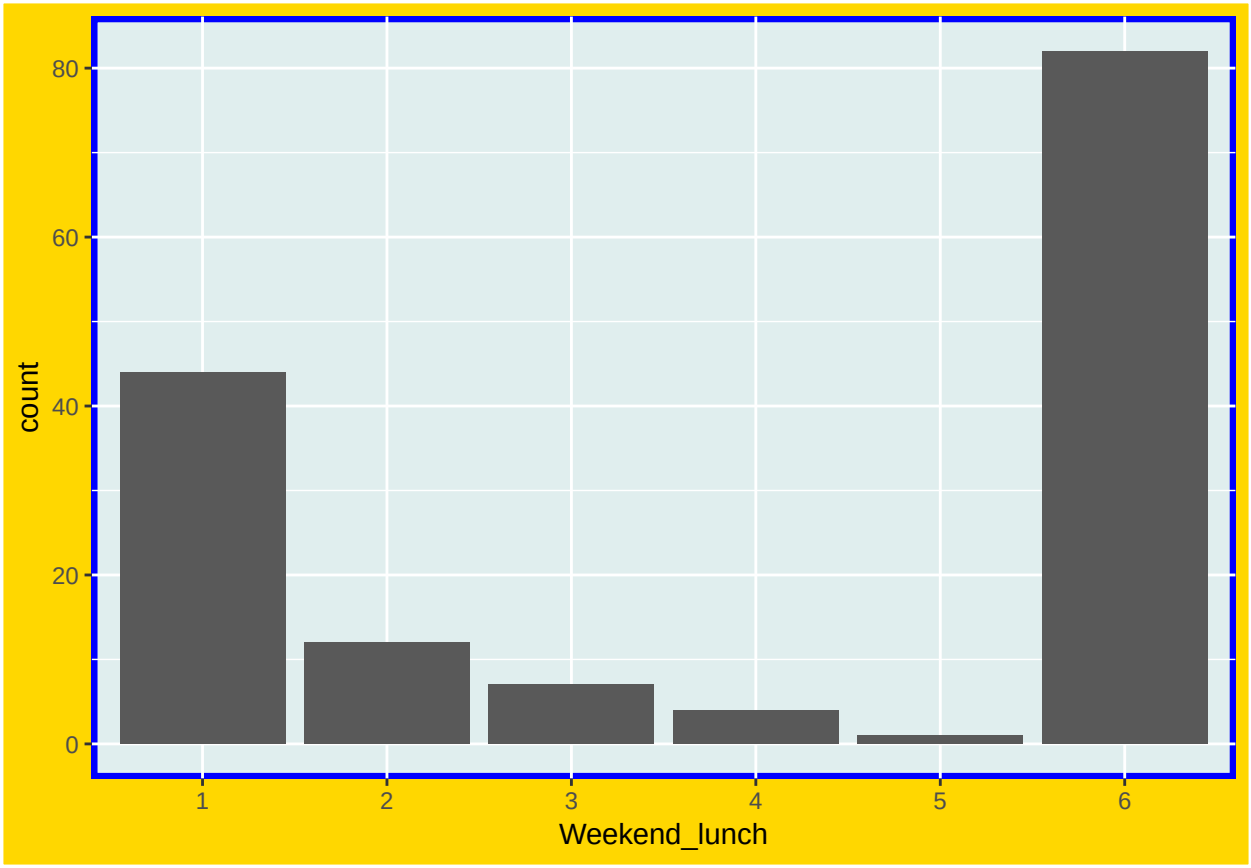

```
ggplot(drop_na(wdd))+geom_bar(aes(Weekday_dinner),position = 'dodge')
```



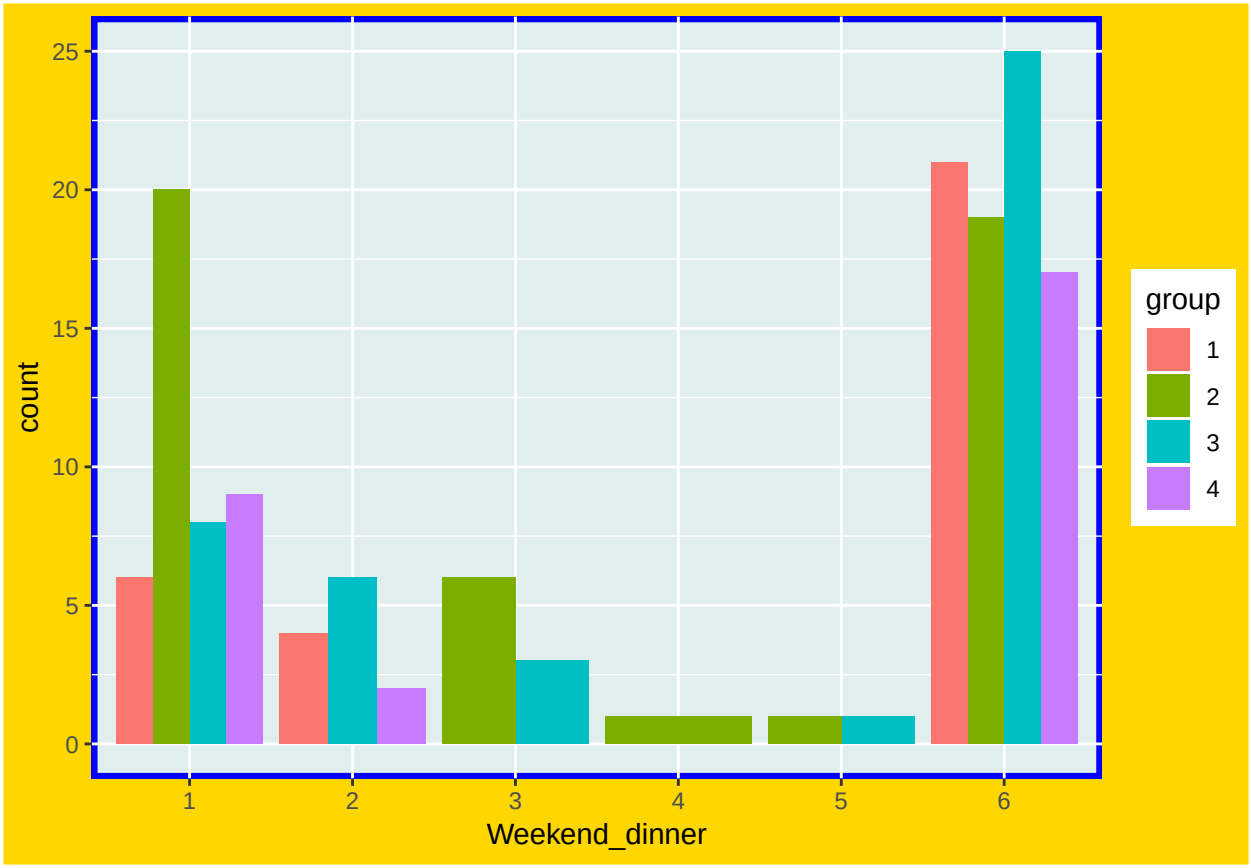

```
ggplot(drop_na(wel,group),aes(group=group))+geom_bar(aes(Weekend_lunch,fill=group),position = 'dodge')
```



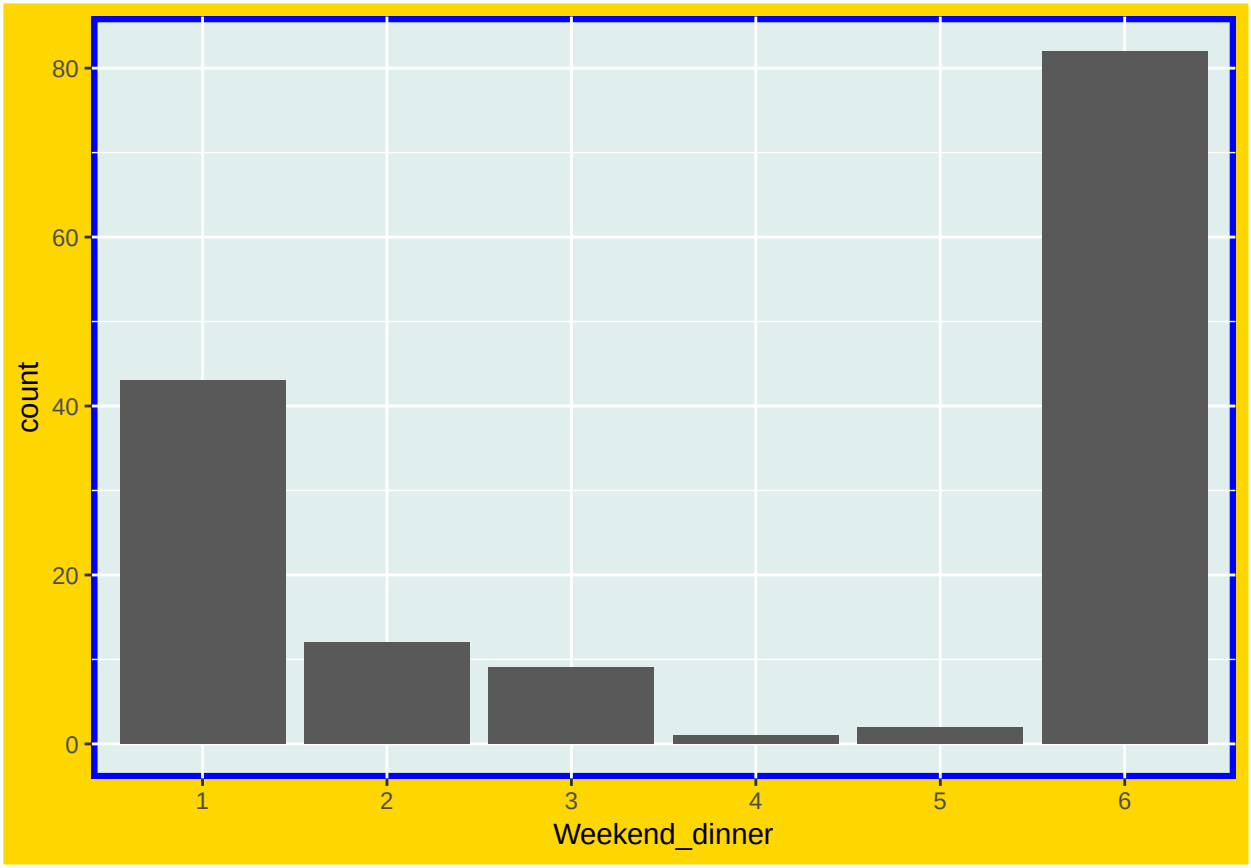

```
ggplot(drop_na(wel,group))+geom_bar(aes(Weekend_lunch),position = 'dodge')
```



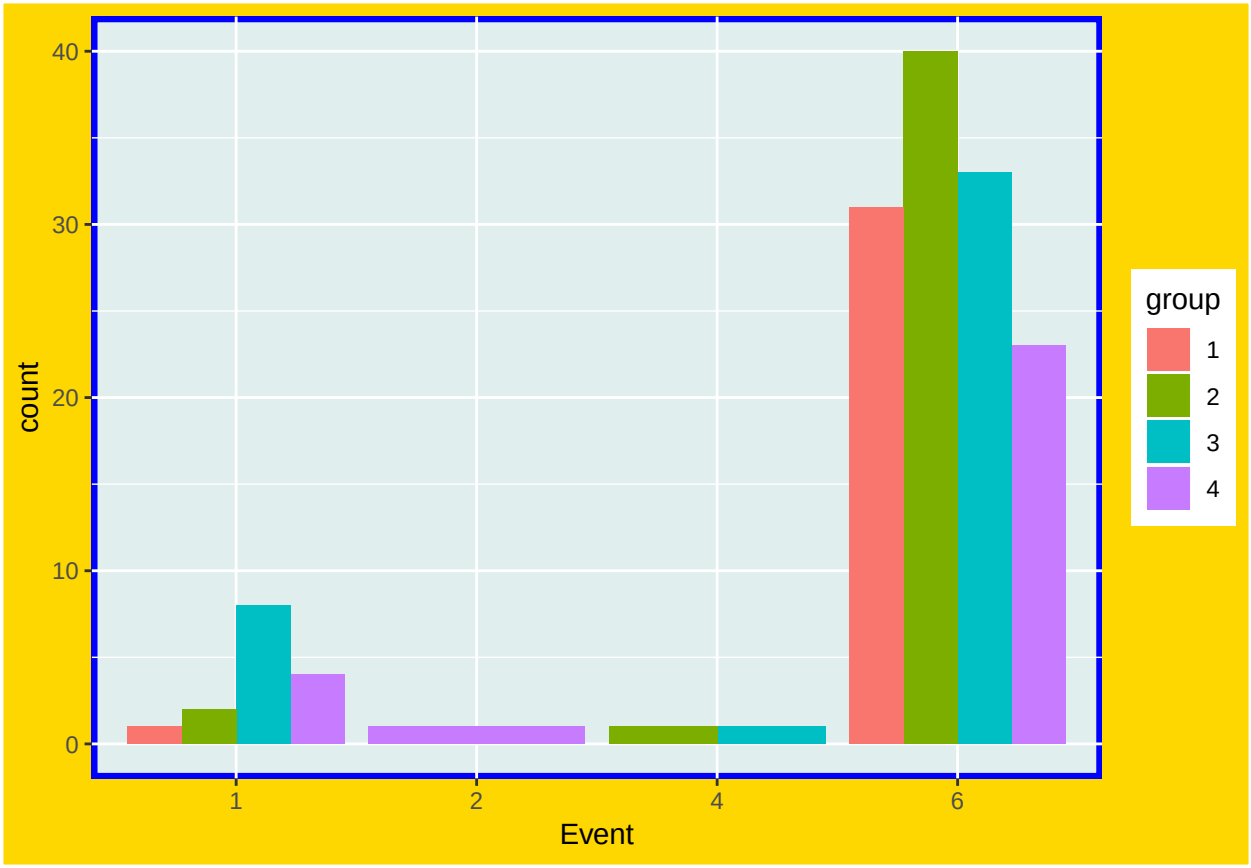

```
ggplot(drop_na(wed,group),aes(group=group))+geom_bar(aes(Weekend_dinner,fill=group),position = 'dodge')
```



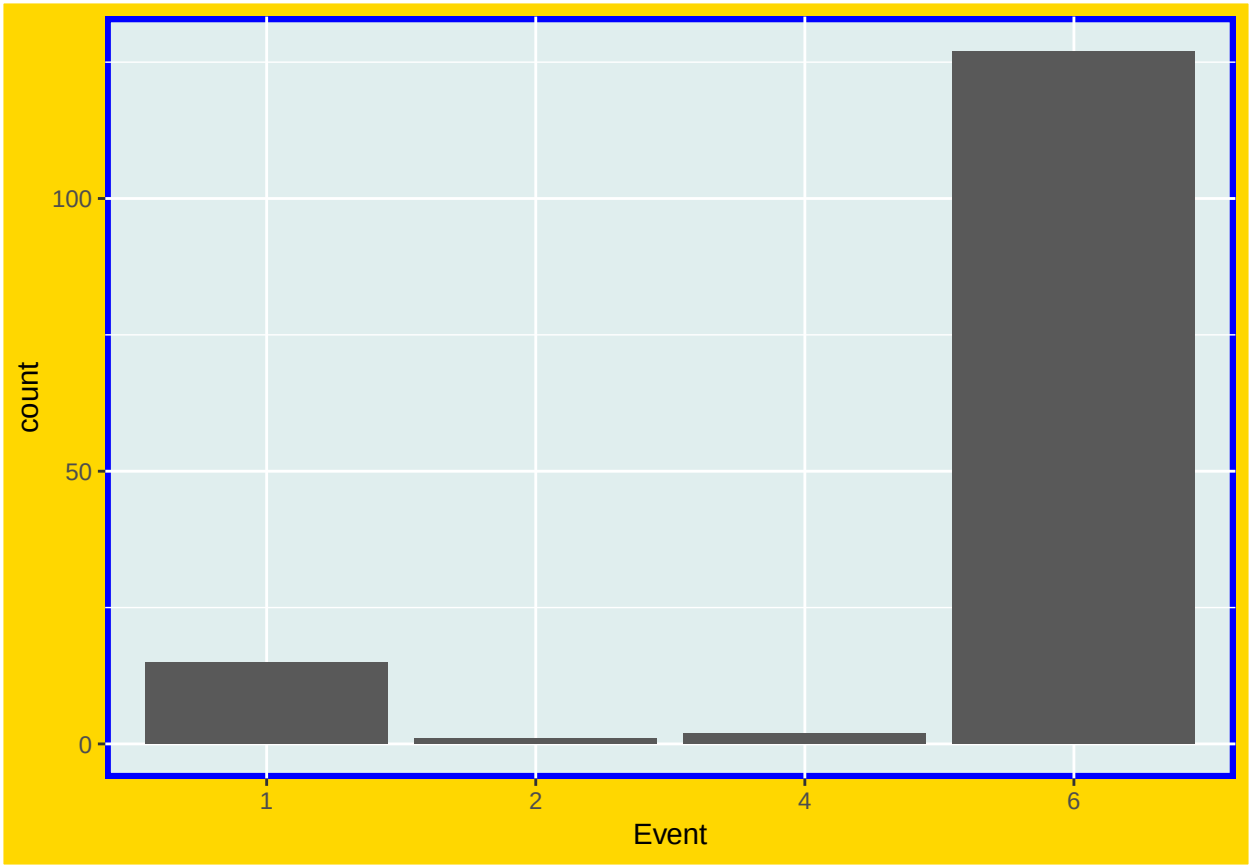

```
ggplot(drop_na(wed,group))+geom_bar(aes(Weekend_dinner),position = 'dodge')
```



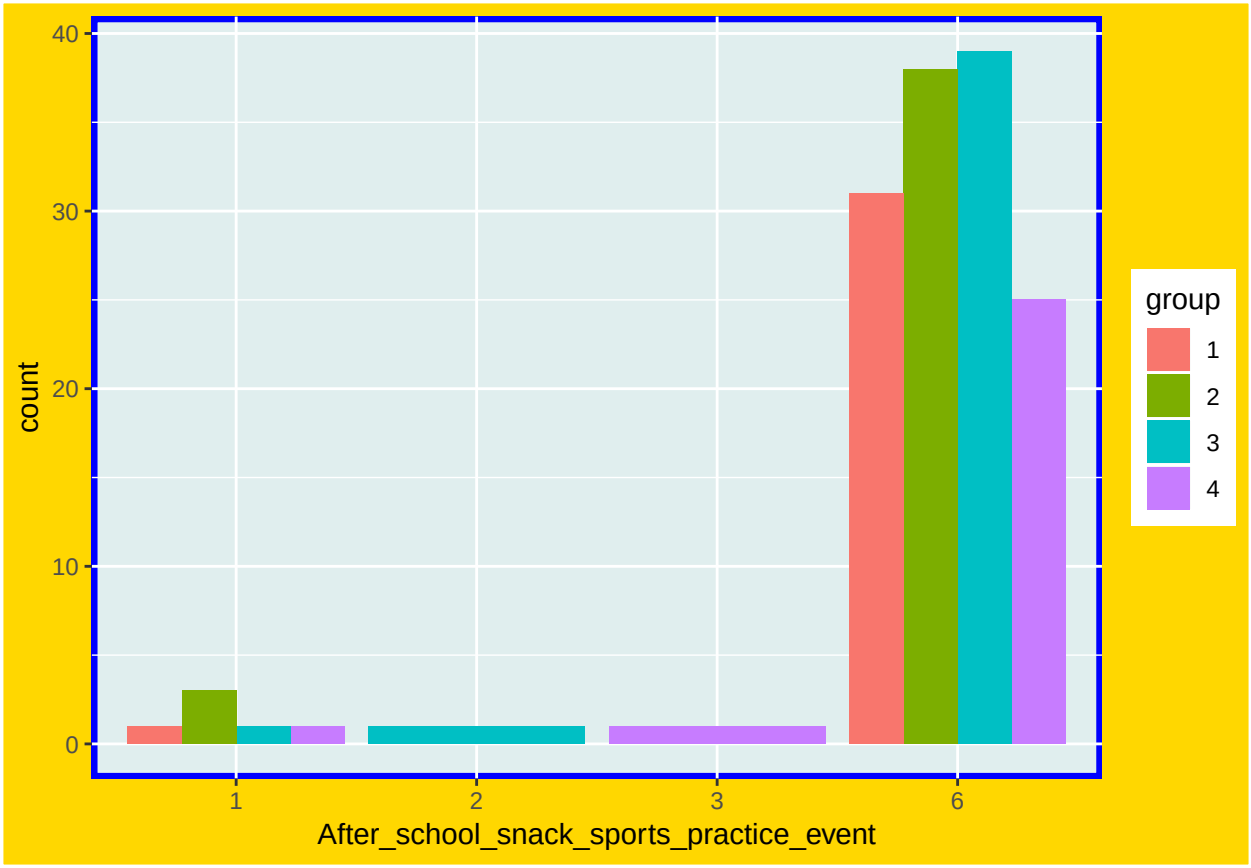

```
ggplot(drop_na(event,group),aes(group=group))+geom_bar(aes(Event,fill=group),position = 'dodge')
```



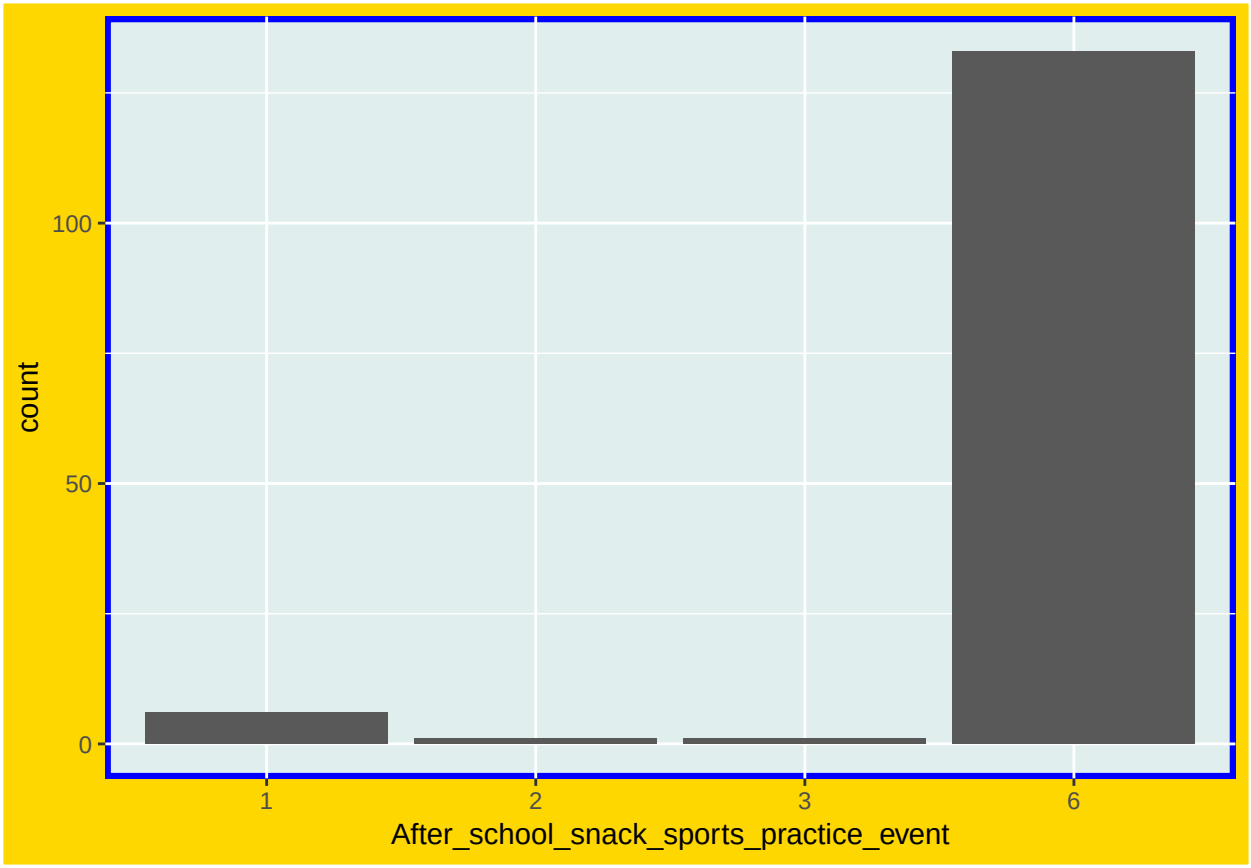

```
ggplot(drop_na(event,group))+geom_bar(aes(Event),position = 'dodge')
```




```
ggplot(drop_na(as,group),aes(group=group))+geom_bar(aes(After_school_snack_sports_practice_event,fill=g
```




```
ggplot(drop_na(as,group))+geom_bar(aes(After_school_snack_sports_practice_event),position = 'dodge')
```



The potential location Sticks can pick for new restaurant.

```
locations=data.frame(Loc.=c('A','B','C','D'),Population=c(29321,34183,42913,57509),Median_Age=c(39.1,32
print(locations)
```

```
##   Loc. Population Median_Age Median_Income Consumer_Spend
## 1    A      29321       39.1         92700          $722M
## 2    B      34183       32.5         31900          $482M
## 3    C      42913       32.5         55700          $754M
## 4    D      57509       34.8         75000         $1184M
##   Consumer_Spend.Household
## 1                    62404
## 2                    36720
## 3                    46828
## 4                    57880
##                                                                  Major_Customer_Profiles
## 1 Blue Blood Estates,Brite Lites,Li'l City,Executive Suites,Upward Bound,Winner's Circle
## 2         City Startups,Family Thrifts,Hometown Retired,New Beginnings,Sunset City Blues
## 3   Brite Lites, Li'l City,Family Thrifts,Up-and-Comers,Upward Bound,White Picket Fences
## 4    Brite Lites,Li'l City,Country Quires,Up-and-Comers,Upward Bound,White Picket Fences
```

Customer's behaviors & characteristics:

- Age mostly 26~65
- Income mostly $50k~$100k
- Mostly in relationship or married.
- Convience is really important.
- Variety is somewhat important.
- Price is really important.
- Healthy menue is very important, but this might be due to the fact that people who took the survey are mostly female.
- Eat outside a lot
- Do volunteer work often
- Often visit museums
- Often running
- Often attend local sports events
- Often do kids activity
- Often gardening
- Majority doing business
- Frequently attend local sports events for kids activity.
- Frequently go to local attractions for kids activity.
- Often go to soccer games for kids activity.
- Often go swimming for kids activity.
- Often go to children's museums for kids activity.
- Use coupons a little bit.
- Get coupons mostly from loyalty card rewards.
- Get coupons often from mails,store receipts, & groupons social daily offers.
- Friendly staff is important.
- Pleasant ambiance is somewhat important.
- Consistency reliability is very important.
- Part of community is somewhat important.
- Mostly don't skip lunch
- Sometimes visit Sticks for weekday lunch
- Visit Sticks a lot for weekend lunch
- Visit Sticks often for weekend dinner
- Visit Sticks a lot for kids' after school event, & some other events.

Which profiles from the potential locations match with customers: (Claritas website)

- Brite Lites, Li'l City
- Executive Suites
- Upward Bound
- Up-and-Comers
- White Picket Fences

Which location to pick?

- C & D looks good, but D might be better due to higher population, Income, & consumer spend.

Which group of people should Sticks target

- Looks like group 2 has most of the customers, so Sticks should target group 2.

What are the characteristics of group 2:

- 41~65 years old
- Income mostly $50k~$100k
- Mostly in relationship or married.
- Convience is really important.
- Variety is somewhat important.
- Price is really important.
- Healthy menue is very important, but this might be due to the fact that people who took the survey are mostly female.
- Eat outside a lot
- Do volunteer work often
- Often visit museums
- Often attend loycal sports events
- Majority doing business
- Frequently attend local sports events for kids activity.
- Frequently go to local attractions for kids activity.
- Often go to soccer games for kids activity.
- Often go swimming for kids activity.
- Think Sticks is more convimience than other restaurants.
- Think Sticks is healthier than other restaurants.
- Think Sticks has better foods than other restaurants.
- Think Sticks is friendlier than other restaurants.
- Think Sticks is more cosistent than other restaurants.
- Mostly without chidren
- Use coupons a little bit.
- Get coupons mostly from loyalty card rewards.
- Get coupons often from mails,store receipts, & groupons social daily offers.
- Friendly staff is important.
- Pleasant ambiance is important.
- Consistency reliability is very important.
- Part of community is somewhat important.
- Mostly lunch outside
- Mostly don't skip lunch
- Often visit Sticks for weekday lunch
- Sometimes visit Sticks for weekday lunch
- Visit Sticks a lot for weekend lunch
- Visit Sticks sometimes for weekend dinner
- Visit Sticks a lot for kids' after school event, & some other events.