Chapter 1 : Data Preprocessing & EDA & Visualization

문제1.

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────
──────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.4.0      ✓ purrr   1.0.0
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.1      ✓ stringr 1.5.0
## ✓ readr   2.1.3      ✓ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────
──────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(data.table)
```

```
##
## 다음의 패키지를 부착합니다: 'data.table'
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
##
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(magrittr)
```

```
##
## 다음의 패키지를 부착합니다: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##     set_names
##
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
train<-fread('train.csv')
```

```
head(train)
```

```
##    V1     id Gender     Customer Type Age  Type of Travel    Class
## 1:  0  70172   Male    Loyal Customer  13 Personal Travel Eco Plus
## 2:  1   5047   Male disloyal Customer  25 Business travel Business
## 3:  2 110028 Female    Loyal Customer  26 Business travel Business
## 4:  3  24026 Female    Loyal Customer  25 Business travel Business
## 5:  4 119299   Male    Loyal Customer  61 Business travel Business
## 6:  5 111157 Female    Loyal Customer  26 Personal Travel      Eco
##    Flight Distance Inflight wifi service Departure/Arrival time convenient
## 1:             460                     3                                 4
## 2:             235                     3                                 2
## 3:            1142                     2                                 2
## 4:             562                     2                                 5
## 5:             214                     3                                 3
## 6:            1180                     3                                 4
##    Ease of Online booking Gate location Food and drink Online boarding
## 1:                      3             1              5                3
## 2:                      3             3              1                3
## 3:                      2             2              5                5
## 4:                      5             5              2                2
## 5:                      3             3              4                5
## 6:                      2             1              1                2
##    Seat comfort Inflight entertainment On-board service Leg room service
## 1:            5                      5                4                3
## 2:            1                      1                1                5
## 3:            5                      5                4                3
## 4:            2                      2                2                5
## 5:            5                      3                3                4
## 6:            1                      1                3                4
##    Baggage handling Checkin service Inflight service Cleanliness
## 1:                4                4                5           5
## 2:                3                1                4           1
## 3:                4                4                4           5
## 4:                3                1                4           2
## 5:                4                3                3           3
## 6:                4                4                4           1
##    Departure Delay in Minutes Arrival Delay in Minutes           satisfaction
## 1:                         25                       18 neutral or dissatisfied
## 2:                          1                        6 neutral or dissatisfied
## 3:                          0                        0               satisfied
## 4:                         11                        9 neutral or dissatisfied
## 5:                          0                        0               satisfied
## 6:                          0                        0 neutral or dissatisfied
```

```
tail(train)
```

```
##        V1    id Gender     Customer Type Age  Type of Travel    Class
## 1: 103898 60666   Male    Loyal Customer  50 Personal Travel      Eco
## 2: 103899 94171 Female disloyal Customer  23 Business travel      Eco
## 3: 103900 73097   Male    Loyal Customer  49 Business travel Business
## 4: 103901 68825   Male disloyal Customer  30 Business travel Business
## 5: 103902 54173 Female disloyal Customer  22 Business travel      Eco
## 6: 103903 62567   Male    Loyal Customer  27 Business travel Business
##    Flight Distance Inflight wifi service Departure/Arrival time convenient
## 1:            1620                     3                                1
## 2:             192                     2                                1
## 3:            2347                     4                                4
## 4:            1995                     1                                1
## 5:            1000                     1                                1
## 6:            1723                     1                                3
##    Ease of Online booking Gate location Food and drink Online boarding
## 1:                      3             4              2               3
## 2:                      2             3              2               2
## 3:                      4             4              2               4
## 4:                      1             3              4               1
## 5:                      1             5              1               1
## 6:                      3             3              1               1
##    Seat comfort Inflight entertainment On-board service Leg room service
## 1:            2                      2                4                3
## 2:            2                      2                3                1
## 3:            5                      5                5                5
## 4:            5                      4                3                2
## 5:            1                      1                4                5
## 6:            1                      1                1                1
##    Baggage handling Checkin service Inflight service Cleanliness
## 1:                4                2                4           2
## 2:                4                2                3           2
## 3:                5                5                5           4
## 4:                4                5                5           4
## 5:                1                5                4           1
## 6:                4                4                3           1
##    Departure Delay in Minutes Arrival Delay in Minutes           satisfaction
## 1:                          0                       0 neutral or dissatisfied
## 2:                          3                       0 neutral or dissatisfied
## 3:                          0                       0              satisfied
## 4:                          7                      14 neutral or dissatisfied
## 5:                          0                       0 neutral or dissatisfied
## 6:                          0                       0 neutral or dissatisfied
```

```
summary(train)
```

```
##        V1               id            Gender          Customer Type
##  Min.   :     0   Min.   :     1   Length:103904      Length:103904
##  1st Qu.: 25976   1st Qu.: 32534   Class :character   Class :character
##  Median : 51952   Median : 64857   Mode  :character   Mode  :character
##  Mean   : 51952   Mean   : 64924
##  3rd Qu.: 77927   3rd Qu.: 97368
##  Max.   :103903   Max.   :129880
##
##       Age        Type of Travel        Class           Flight Distance
##  Min.   : 7.00   Length:103904      Length:103904      Min.   :  31
##  1st Qu.:27.00   Class :character   Class :character   1st Qu.: 414
##  Median :40.00   Mode  :character   Mode  :character   Median : 843
##  Mean   :39.38                                         Mean   :1189
##  3rd Qu.:51.00                                         3rd Qu.:1743
##  Max.   :85.00                                         Max.   :4983
##
##  Inflight wifi service Departure/Arrival time convenient Ease of Online booking
##  Min.   :0.00          Min.   :0.00                      Min.   :0.000
##  1st Qu.:2.00          1st Qu.:2.00                      1st Qu.:2.000
##  Median :3.00          Median :3.00                      Median :3.000
##  Mean   :2.73          Mean   :3.06                      Mean   :2.757
##  3rd Qu.:4.00          3rd Qu.:4.00                      3rd Qu.:4.000
##  Max.   :5.00          Max.   :5.00                      Max.   :5.000
##
##  Gate location   Food and drink  Online boarding  Seat comfort
##  Min.   :0.000   Min.   :0.000   Min.   :0.00    Min.   :0.000
##  1st Qu.:2.000   1st Qu.:2.000   1st Qu.:2.00    1st Qu.:2.000
##  Median :3.000   Median :3.000   Median :3.00    Median :4.000
##  Mean   :2.977   Mean   :3.202   Mean   :3.25    Mean   :3.439
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:4.00    3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.00    Max.   :5.000
##
##  Inflight entertainment On-board service Leg room service Baggage handling
##  Min.   :0.000          Min.   :0.000    Min.   :0.000    Min.   :1.000
##  1st Qu.:2.000          1st Qu.:2.000    1st Qu.:2.000    1st Qu.:3.000
##  Median :4.000          Median :4.000    Median :4.000    Median :4.000
##  Mean   :3.358          Mean   :3.382    Mean   :3.351    Mean   :3.632
##  3rd Qu.:4.000          3rd Qu.:4.000    3rd Qu.:4.000    3rd Qu.:5.000
##  Max.   :5.000          Max.   :5.000    Max.   :5.000    Max.   :5.000
##
##  Checkin service Inflight service  Cleanliness    Departure Delay in Minutes
##  Min.   :0.000   Min.   :0.00     Min.   :0.000   Min.   :   0.00
##  1st Qu.:3.000   1st Qu.:3.00     1st Qu.:2.000   1st Qu.:   0.00
##  Median :3.000   Median :4.00     Median :3.000   Median :   0.00
##  Mean   :3.304   Mean   :3.64     Mean   :3.286   Mean   :  14.82
##  3rd Qu.:4.000   3rd Qu.:5.00     3rd Qu.:4.000   3rd Qu.:  12.00
##  Max.   :5.000   Max.   :5.00     Max.   :5.000   Max.   :1592.00
##
##  Arrival Delay in Minutes satisfaction
##  Min.   :   0.00          Length:103904
##  1st Qu.:   0.00          Class :character
##  Median :   0.00          Mode  :character
##  Mean   :  15.18
##  3rd Qu.:  13.00
```

```
##  Max.   :1584.00
##  NA's   :310
```

```
str(train)
```

```
## Classes 'data.table' and 'data.frame':   103904 obs. of  25 variables:
##  $ V1                              : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ id                              : int  70172 5047 110028 24026 119299 111157 82113 96462
79485 65725 ...
##  $ Gender                          : chr  "Male" "Male" "Female" "Female" ...
##  $ Customer Type                   : chr  "Loyal Customer" "disloyal Customer" "Loyal Custo
mer" "Loyal Customer" ...
##  $ Age                             : int  13 25 26 25 61 26 47 52 41 20 ...
##  $ Type of Travel                  : chr  "Personal Travel" "Business travel" "Business tra
vel" "Business travel" ...
##  $ Class                           : chr  "Eco Plus" "Business" "Business" "Business" ...
##  $ Flight Distance                 : int  460 235 1142 562 214 1180 1276 2035 853 1061 ...
##  $ Inflight wifi service           : int  3 3 2 2 3 3 2 4 1 3 ...
##  $ Departure/Arrival time convenient: int  4 2 2 5 3 4 4 3 2 3 ...
##  $ Ease of Online booking          : int  3 3 2 5 3 2 2 4 2 3 ...
##  $ Gate location                   : int  1 3 2 5 3 1 3 4 2 4 ...
##  $ Food and drink                  : int  5 1 5 2 4 1 2 5 4 2 ...
##  $ Online boarding                 : int  3 3 5 2 5 2 2 5 3 3 ...
##  $ Seat comfort                    : int  5 1 5 2 5 1 2 5 3 3 ...
##  $ Inflight entertainment          : int  5 1 5 2 3 1 2 5 1 2 ...
##  $ On-board service                : int  4 1 4 2 3 3 3 5 1 2 ...
##  $ Leg room service                : int  3 5 3 5 4 4 3 5 2 3 ...
##  $ Baggage handling                : int  4 3 4 3 4 4 4 5 1 4 ...
##  $ Checkin service                 : int  4 1 4 1 3 4 3 4 4 4 ...
##  $ Inflight service                : int  5 4 4 4 3 4 5 5 1 3 ...
##  $ Cleanliness                     : int  5 1 5 2 3 1 2 4 2 2 ...
##  $ Departure Delay in Minutes      : int  25 1 0 11 0 0 9 4 0 0 ...
##  $ Arrival Delay in Minutes        : num  18 6 0 9 0 0 23 0 0 0 ...
##  $ satisfaction                    : chr  "neutral or dissatisfied" "neutral or dissatisfie
d" "satisfied" "neutral or dissatisfied" ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
glimpse(train)
```

```
## Rows: 103,904
## Columns: 25
## $ V1                                <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, …
## $ id                                <int> 70172, 5047, 110028, 24026, 119299…
## $ Gender                            <chr> "Male", "Male", "Female", "Female"…
## $ `Customer Type`                   <chr> "Loyal Customer", "disloyal Custom…
## $ Age                               <int> 13, 25, 26, 25, 61, 26, 47, 52, 41…
## $ `Type of Travel`                  <chr> "Personal Travel", "Business trave…
## $ Class                             <chr> "Eco Plus", "Business", "Business"…
## $ `Flight Distance`                 <int> 460, 235, 1142, 562, 214, 1180, 12…
## $ `Inflight wifi service`           <int> 3, 3, 2, 2, 3, 3, 2, 4, 1, 3, 4, 2…
## $ `Departure/Arrival time convenient` <int> 4, 2, 2, 5, 3, 4, 4, 3, 2, 3, 5, 4…
## $ `Ease of Online booking`          <int> 3, 3, 2, 5, 3, 2, 2, 4, 2, 3, 5, 2…
## $ `Gate location`                   <int> 1, 3, 2, 5, 3, 1, 3, 4, 2, 4, 4, 2…
## $ `Food and drink`                  <int> 5, 1, 5, 2, 4, 1, 2, 5, 4, 2, 2, 1…
## $ `Online boarding`                 <int> 3, 3, 5, 2, 5, 2, 2, 5, 3, 3, 5, 2…
## $ `Seat comfort`                    <int> 5, 1, 5, 2, 5, 1, 2, 5, 3, 3, 2, 1…
## $ `Inflight entertainment`          <int> 5, 1, 5, 2, 3, 1, 2, 5, 1, 2, 2, 1…
## $ `On-board service`                <int> 4, 1, 4, 2, 3, 3, 3, 5, 1, 2, 3, 1…
## $ `Leg room service`                <int> 3, 5, 3, 5, 4, 4, 3, 5, 2, 3, 3, 2…
## $ `Baggage handling`                <int> 4, 3, 4, 3, 4, 4, 4, 5, 1, 4, 5, 5…
## $ `Checkin service`                 <int> 4, 1, 4, 1, 3, 4, 3, 4, 4, 4, 3, 5…
## $ `Inflight service`                <int> 5, 4, 4, 4, 3, 4, 5, 5, 1, 3, 5, 5…
## $ Cleanliness                       <int> 5, 1, 5, 2, 3, 1, 2, 4, 2, 2, 2, 1…
## $ `Departure Delay in Minutes`      <int> 25, 1, 0, 11, 0, 0, 9, 4, 0, 0, 0,…
## $ `Arrival Delay in Minutes`        <dbl> 18, 6, 0, 9, 0, 0, 23, 0, 0, 0, 0,…
## $ satisfaction                      <chr> "neutral or dissatisfied", "neutra…
```

문제2.

```
train %<>% select(-c(V1,id))
```

문제3.

```
train %>% lapply(n_distinct)
```

```
## $Gender
## [1] 2
##
## $`Customer Type`
## [1] 2
##
## $Age
## [1] 75
##
## $`Type of Travel`
## [1] 2
##
## $Class
## [1] 3
##
## $`Flight Distance`
## [1] 3802
##
## $`Inflight wifi service`
## [1] 6
##
## $`Departure/Arrival time convenient`
## [1] 6
##
## $`Ease of Online booking`
## [1] 6
##
## $`Gate location`
## [1] 6
##
## $`Food and drink`
## [1] 6
##
## $`Online boarding`
## [1] 6
##
## $`Seat comfort`
## [1] 6
##
## $`Inflight entertainment`
## [1] 6
##
## $`On-board service`
## [1] 6
##
## $`Leg room service`
## [1] 6
##
## $`Baggage handling`
## [1] 5
##
## $`Checkin service`
## [1] 6
##
## $`Inflight service`
```

```
## [1] 6
##
## $Cleanliness
## [1] 6
##
## $`Departure Delay in Minutes`
## [1] 446
##
## $`Arrival Delay in Minutes`
## [1] 456
##
## $satisfaction
## [1] 2
```

문제4.

```
train %<>%
  mutate(`Customer Type`=
           recode(`Customer Type`,"Loyal Customer"="Loyal","disloyal Customer"="Disloyal"),
         `Type of Travel`=
           recode(`Type of Travel`,"Personal Travel"="Personal","Business travel"="Business"))
%>%
  rename('Time Convenient'='Departure/Arrival time convenient',
         'Departure Delay'='Departure Delay in Minutes',
         'Arrival Delay'='Arrival Delay in Minutes')
```

```
train %>% select(c(`Customer Type`,`Type of Travel`))%>%lapply(unique)
```

```
## $`Customer Type`
## [1] "Loyal"    "Disloyal"
##
## $`Type of Travel`
## [1] "Personal" "Business"
```

```
train %>% colnames
```

```
##  [1] "Gender"                "Customer Type"          "Age"
##  [4] "Type of Travel"        "Class"                  "Flight Distance"
##  [7] "Inflight wifi service" "Time Convenient"        "Ease of Online booking"
## [10] "Gate location"         "Food and drink"         "Online boarding"
## [13] "Seat comfort"          "Inflight entertainment" "On-board service"
## [16] "Leg room service"      "Baggage handling"       "Checkin service"
## [19] "Inflight service"      "Cleanliness"            "Departure Delay"
## [22] "Arrival Delay"         "satisfaction"
```

##문제5.

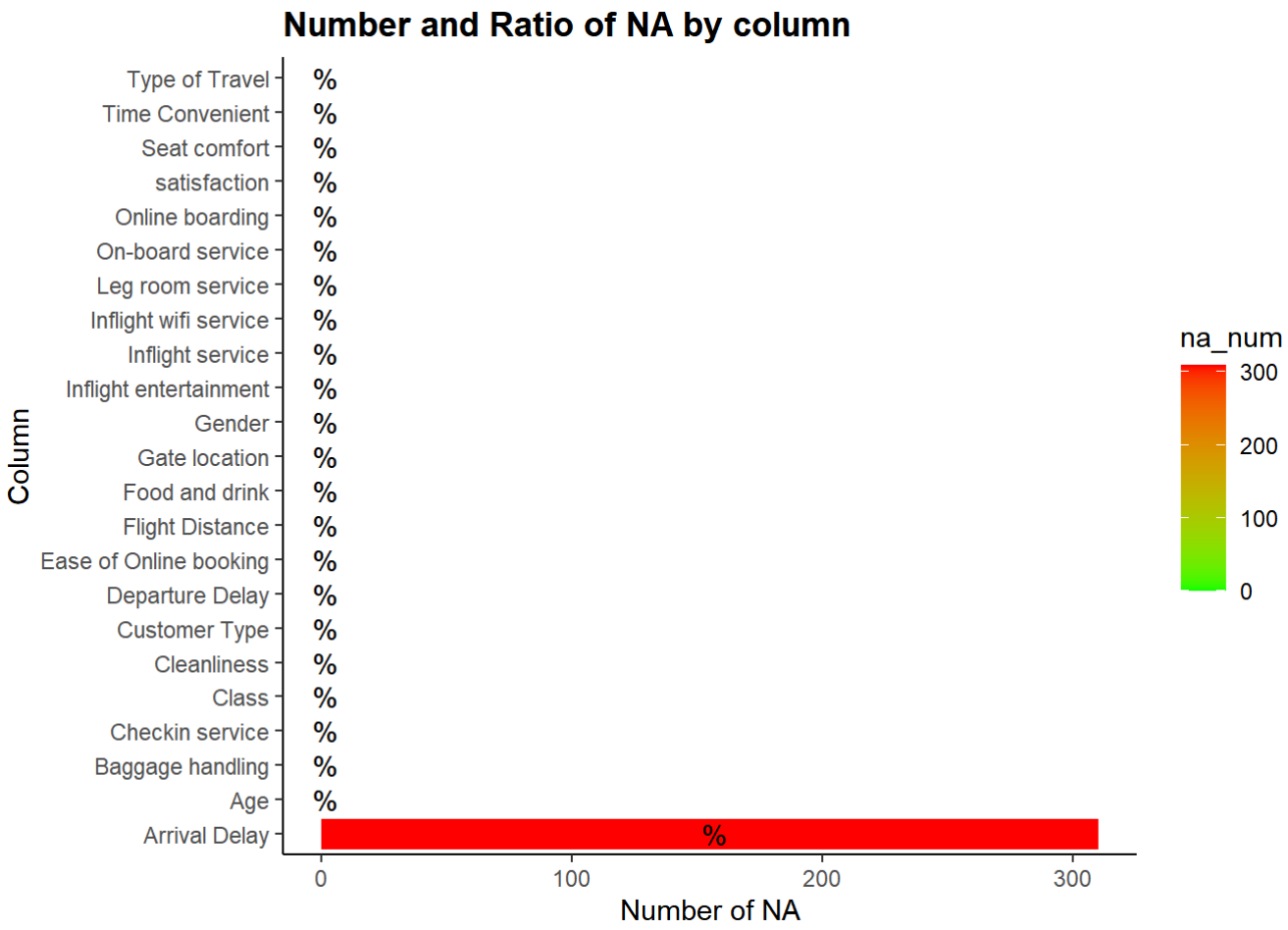
```
num<-c('Age','Flight Distance','Departure Delay','Arrival Delay')
cate<-setdiff(colnames(train),num)

train %<>%
  mutate_at(num,as.numeric) %>%
  mutate_at(cate,as.factor)
```

문제6.

```
train %>% is.na %>% colSums
```

```
##                 Gender          Customer Type                    Age
##                      0                      0                      0
##         Type of Travel                  Class         Flight Distance
##                      0                      0                      0
##  Inflight wifi service        Time Convenient Ease of Online booking
##                      0                      0                      0
##          Gate location         Food and drink        Online boarding
##                      0                      0                      0
##           Seat comfort Inflight entertainment       On-board service
##                      0                      0                      0
##       Leg room service       Baggage handling        Checkin service
##                      0                      0                      0
##       Inflight service            Cleanliness        Departure Delay
##                      0                      0                      0
##          Arrival Delay           satisfaction
##                    310                      0
```

```
col<-colnames(train)
na<-colSums(is.na(train)) %>% as.vector
percent<-((na/nrow(data))*100) %>% round(2)


na_data<-data.frame(col=col,
           na=na,
           percentage=paste(percent %>% as.character,'%'))

na_data %>%
  ggplot(aes(reorder(col,-na),na,fill=na))+
  geom_bar(stat='identity')+
  scale_fill_gradient(low='green',high='red')+
  labs(title='Number and Ratio of NA by column',
       x='Column',
       y='Number of NA',
       fill='na_num')+
  geom_text(aes(x=col,label=percentage),
            position=position_stack(vjust=0.5))+
  theme_classic()+
  coord_flip()+
  theme(plot.title=element_text(face="bold"))
```



문제7.

```
train$`Arrival Delay` <-ifelse(train$`Arrival Delay` %>% is.na,
                               median(train$`Arrival Delay`,na.rm=T),
                               train$`Arrival Delay`)
```

문제8.

```
train %<>%
  mutate_if(is.numeric,as.numeric) %>%
  mutate_if(is.factor,as.factor)
```

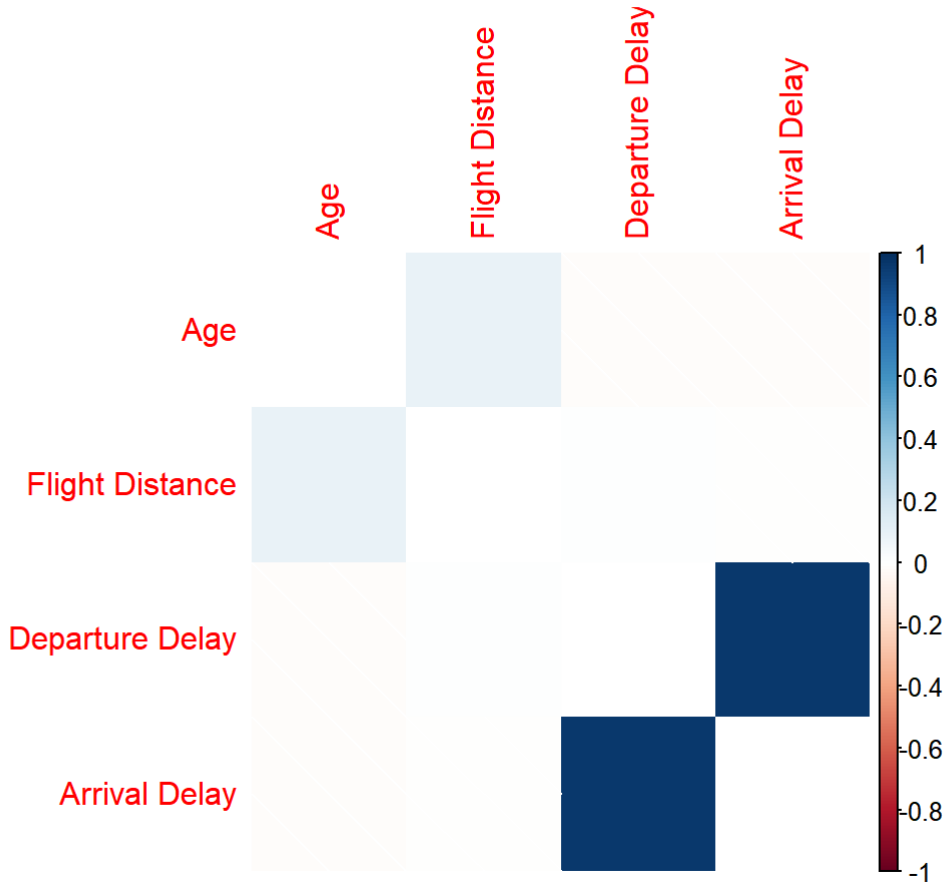
문제9.

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
num_cor<-cor((train %>% select_if(is.numeric)),method='pearson')
```

```
corrplot(num_cor,method='shade',order='AOE',diag=FALSE,
         title="Correlation of Numeric Variables",
         mar=c(0,0,2,0))
```

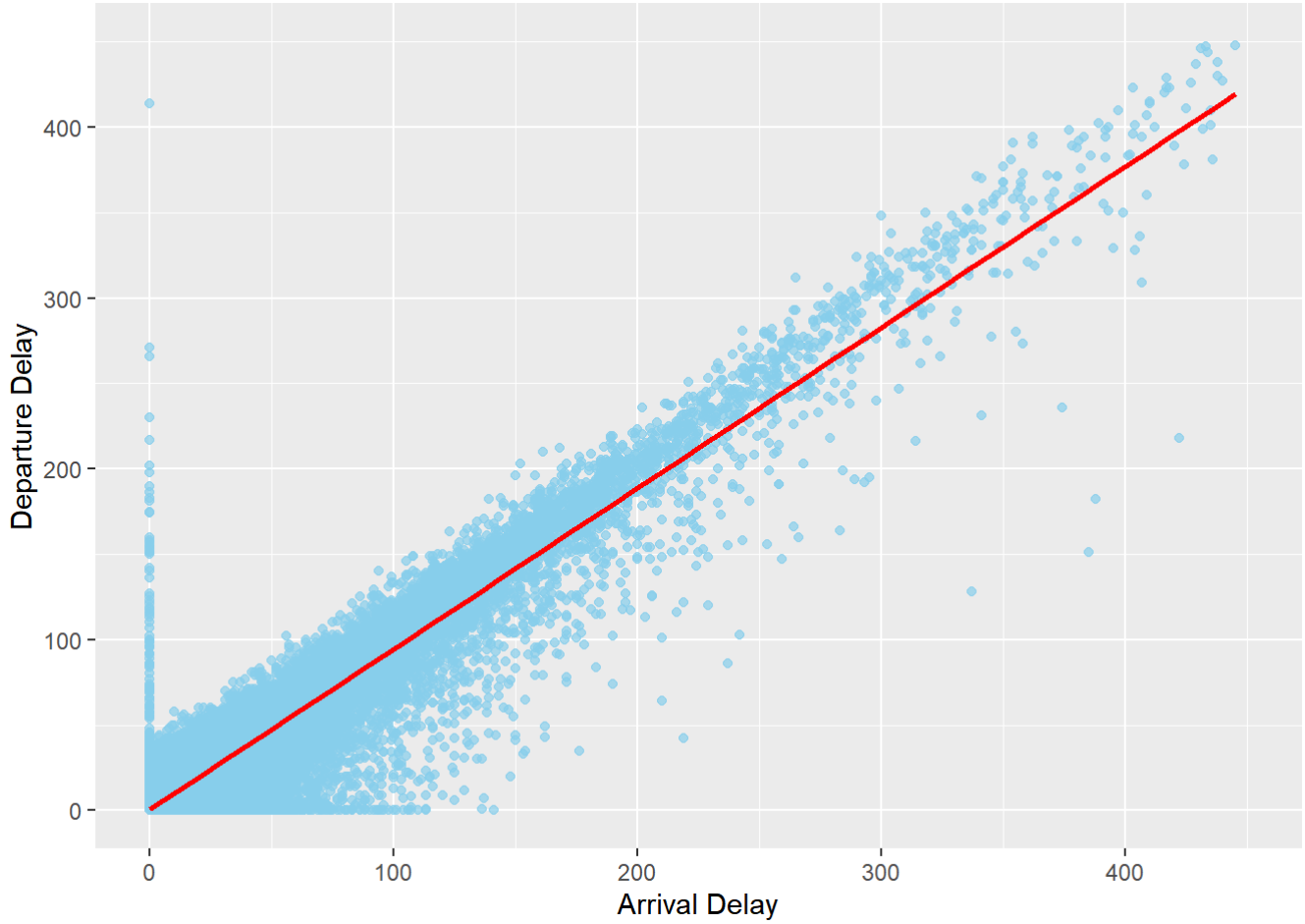



```
train %>%
  ggplot(aes(`Arrival Delay`,`Departure Delay`))+
  geom_point(color='skyblue',alpha=0.7)+
  scale_x_continuous(limits=c(0,450))+
  scale_y_continuous(limits=c(0,450))+
  geom_smooth(method='lm',colour='red')
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 55 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 55 rows containing missing values (`geom_point()`).
```



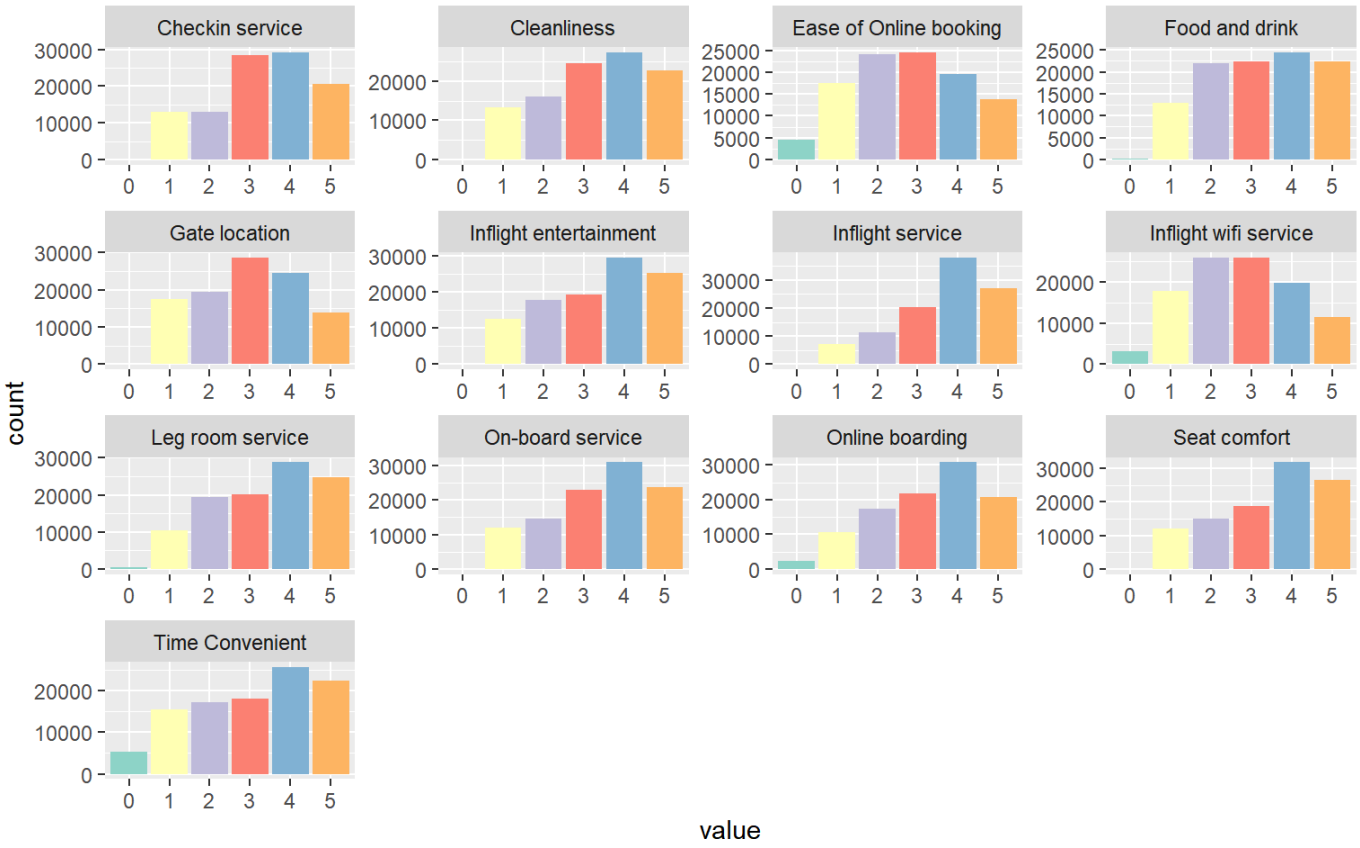

```
unique<-train %>% summarise_all(n_distinct) %>% as.vector
levels_6<-colnames(train)[which(unique==6)]
under_6<-colnames(train)[which(unique<=3)];under_6<-under_6[-length(under_6)]
```

문제10.

```
train %>% select(levels_6) %>%
  gather %>%
  group_by(key,value) %>%
  summarise(count=n()) %>%
  ggplot(aes(value,count))+
  geom_bar(aes(fill=value),stat='identity')+
  scale_fill_brewer(palette='Set3')+
  facet_wrap(vars(key),ncol=4,scales='free')+
  theme(legend.position='none')
```

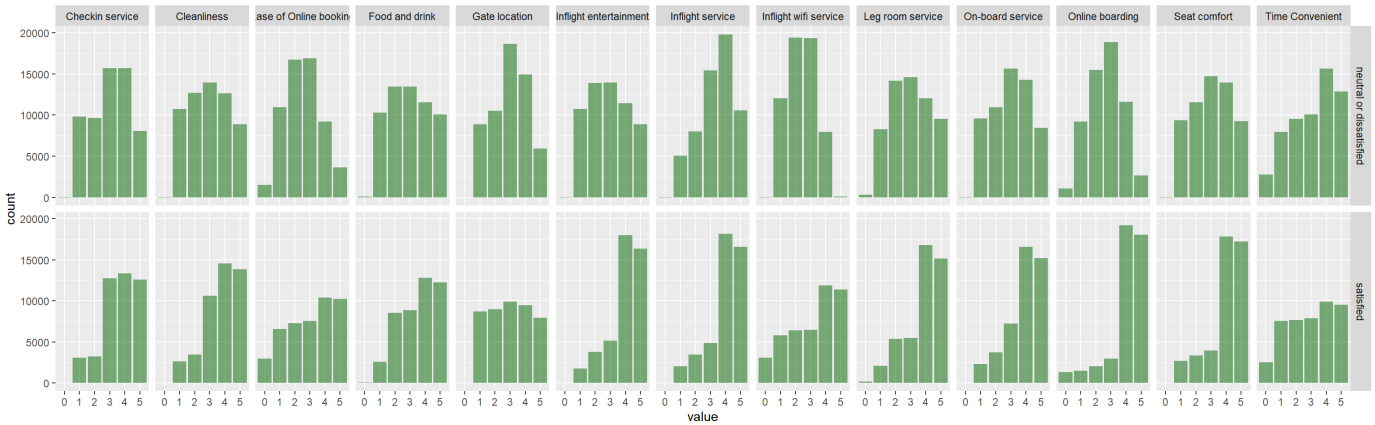
```
## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
## ℹ Please use `all_of()` or `any_of()` instead.
##   # Was:
##   data %>% select(levels_6)
##
##   # Now:
##   data %>% select(all_of(levels_6))
##
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```

```
## `summarise()` has grouped output by 'key'. You can override using the `.groups`
## argument.
```




```
train %>% select(c(levels_6,'satisfaction')) %>%
  gather(levels_6,key='key',value='value') %>%
  group_by(satisfaction,key,value) %>%
  summarise(count=n()) %>%
  ggplot(aes(value,count))+
  geom_bar(aes(fill=value),stat='identity',
           fill='darkgreen',alpha=0.5)+
  facet_grid(satisfaction~key)+
  theme(legend.position='none')
```

```
## `summarise()` has grouped output by 'satisfaction', 'key'. You can override
## using the `.groups` argument.
```



```
nd<-train %>% filter(satisfaction=='neutral or dissatisfied') %>%
  select(all_of(under_6)) %>%
  gather %>%
  group_by(key,value) %>%
  summarise(count=n()) %>%
  ggplot(aes(value,count))+
  geom_bar(aes(fill=value),stat='identity')+
  scale_fill_brewer(palette='Set3')+
  facet_wrap(vars(key),ncol=4,scales='free')+
  theme(legend.position='none')+
  labs(x='',title='Neutral or Dissatisfied')
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
## `summarise()` has grouped output by 'key'. You can override using the `.groups`
## argument.
```

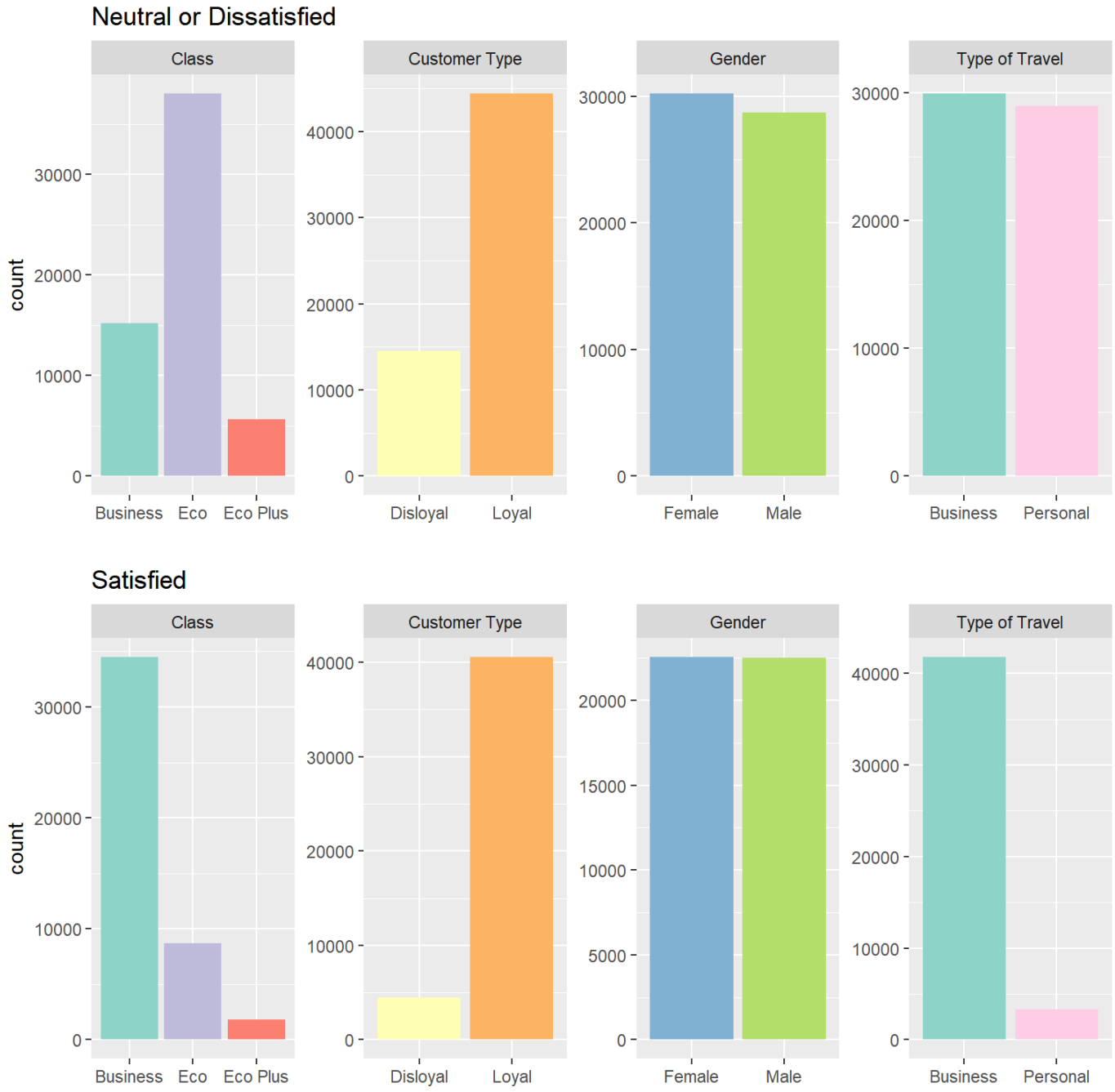
```
sa<-train %>% filter(satisfaction=='satisfied') %>%
  select(all_of(under_6))%>%
  gather %>%
  group_by(key,value) %>%
  summarise(count=n()) %>%
  ggplot(aes(value,count))+
  geom_bar(aes(fill=value),stat='identity')+
  scale_fill_brewer(palette='Set3')+
  facet_wrap(vars(key),ncol=4,scales='free')+
  theme(legend.position='none')+
  labs(x='',title='Satisfied')
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
## `summarise()` has grouped output by 'key'. You can override using the `.groups`
## argument.
```

```
library(ggpubr)
```

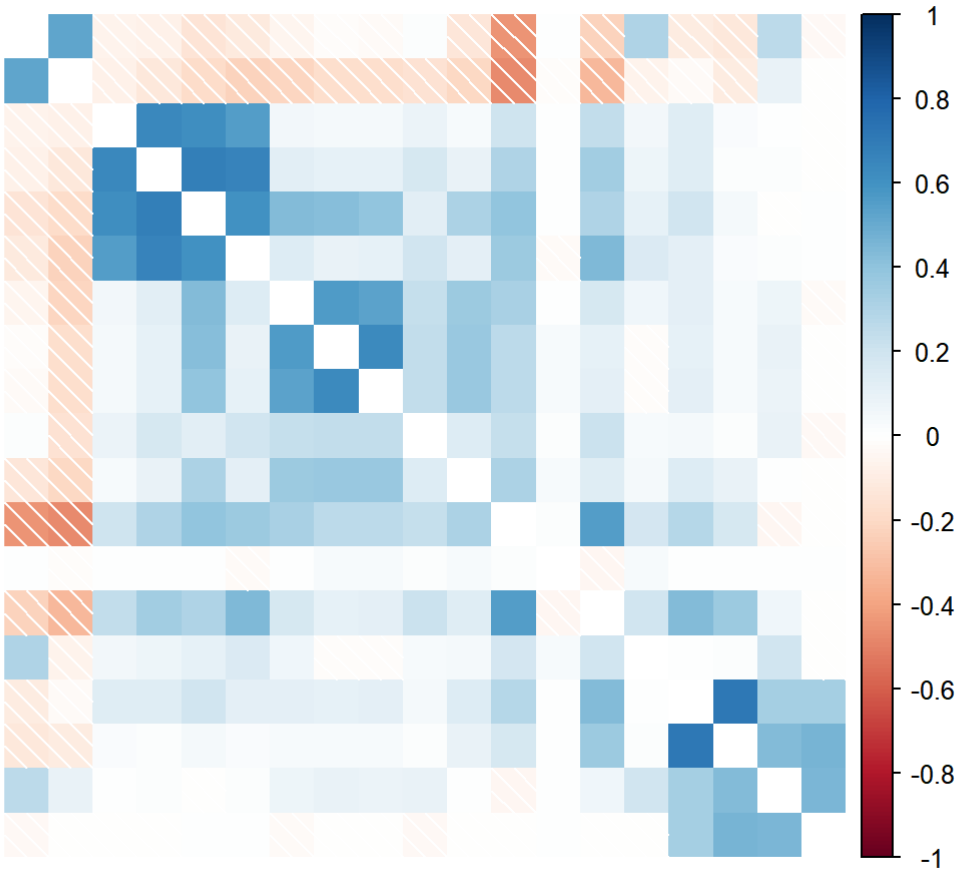
```
ggarrange(nd,sa,ncol=1,legend='none')
```



##문제 11.

```
category<-train %>%
  select(all_of(cate)) %>%
  mutate_at(cate,as.numeric)

cat_cor<-cor(category,method='spearman')
```

```
corrplot(cat_cor,method='shade',order='AOE',diag=FALSE,
         title="Correlation of Categorical Variables",
         tl.pos='n',
         mar=c(1,1,1,1))
```



##문제13.

```
for_pi<-train %>%
  group_by(satisfaction) %>%
  summarise(count=n()) %>%
  mutate(percent=count/sum(count),
         ymax=cumsum(percent),
         ymin=ymax-percent,
         label=paste((round(percent,3)*100) %>% as.character,'%'),
         labelpos=ymax-percent/2)
```

```
theme_clean=function(base_size=12){
  theme_grey(base_size) %+replace%
    theme(
      axis.title=element_blank(),
      axis.text=element_blank(),
      panel.background=element_blank(),
      panel.grid=element_blank(),
      axis.ticks.length=unit(0,"cm"),
      axis.ticks.margin=unit(0,"cm"),
      panel.margin=unit(0,"lines"),
      plot.margin=unit(c(0,0,0,0),"lines"),
      complete=TRUE
    )
}
```

```
donut<-for_pi %>% ggplot+
  geom_rect(aes(xmin=2,xmax=4,ymin=ymin,ymax=ymax,fill=satisfaction))+
  coord_polar(theta='y')+
  xlim(0,4)+
  geom_text(aes(x=3.2,y=labelpos,label=label))+
  theme_clean()+
  scale_fill_brewer(palette='Pastel1')
```

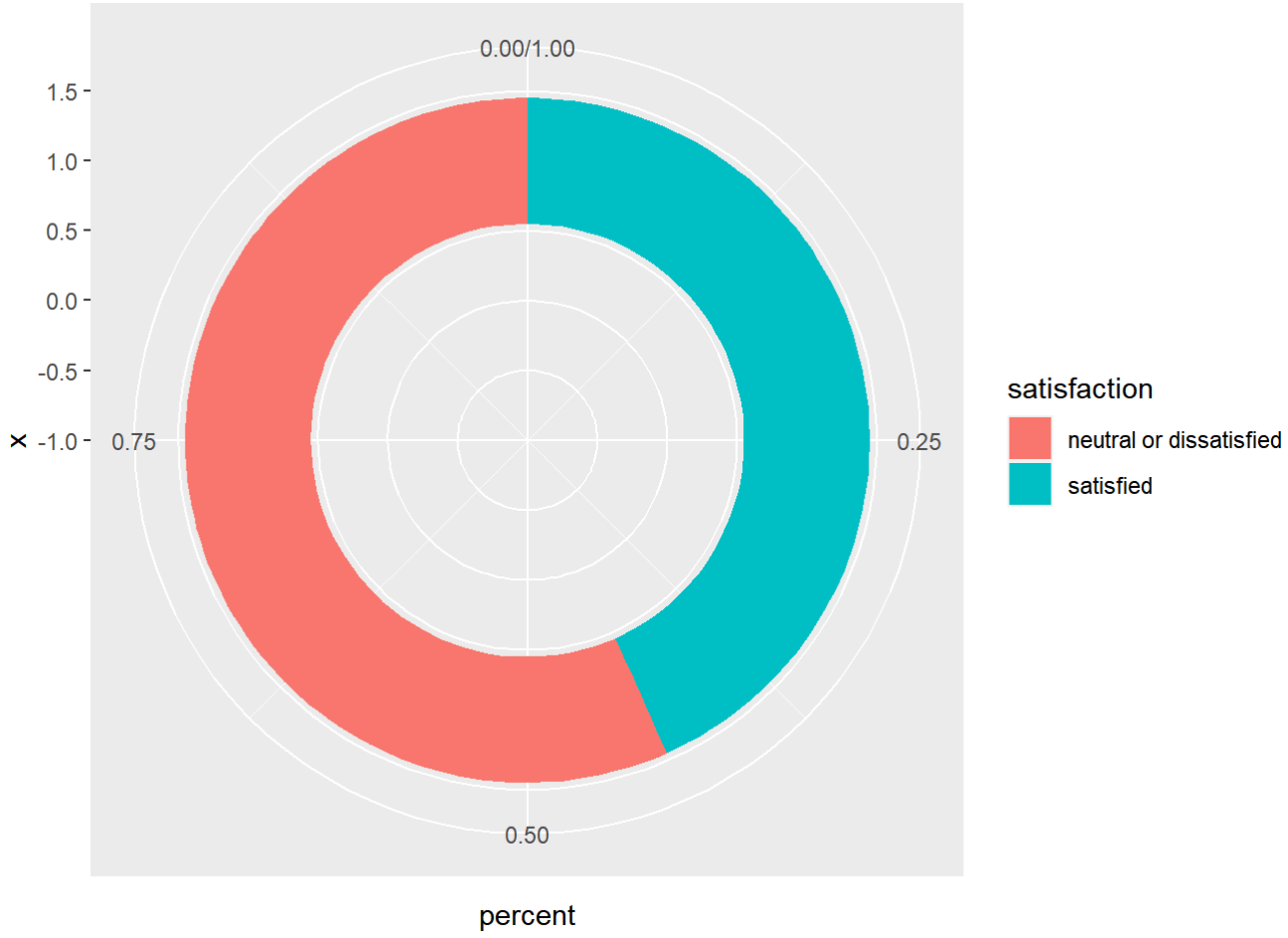
```
## Warning: The `axis.ticks.margin` argument of `theme()` is deprecated as of ggplot2
## 2.0.0.
## ℹ Please set `margin` property of `axis.text` instead
```

```
## Warning: The `panel.margin` argument of `theme()` is deprecated as of ggplot2 2.2.0.
## ℹ Please use the `panel.spacing` argument instead.
```

```
bar<-for_pi %>% ggplot(aes(x=satisfaction,y=percent))+
  geom_bar(aes(fill=satisfaction),stat='identity',width=0.6)+
  scale_fill_brewer(palette='Pastel1')+
  geom_text(aes(x=satisfaction,y=percent,label=label),
            position=position_stack(vjust=0.5))+
  theme_classic()
```

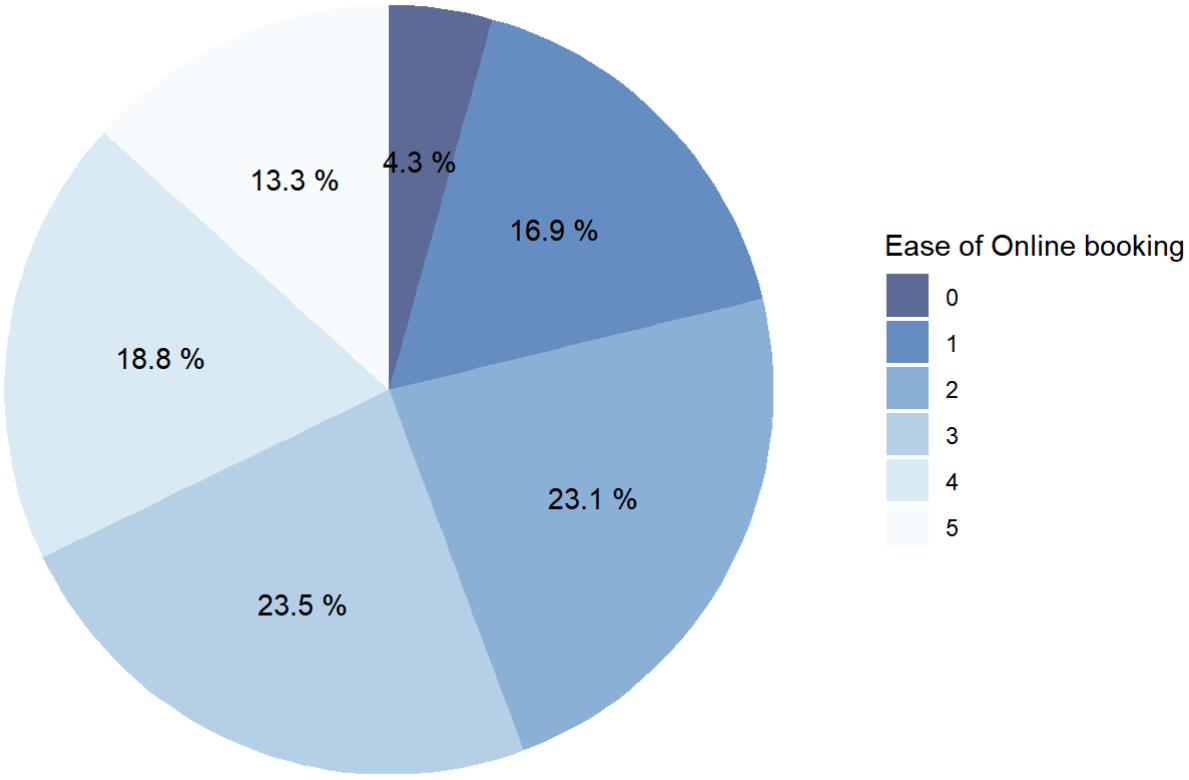
```
satisfaction<-train %>%
  group_by(satisfaction) %>%
  summarise(count=n()) %>%
  mutate(percent=count/sum(count),
         label=paste((round(percent,3)*100) %>% as.character,'%'))
```

```
ggplot(satisfaction,aes(x=1,y=percent,fill=satisfaction))+
  geom_bar(stat='identity')+
#  theme_void()+
  coord_polar('y', start=0)+
  xlim(c(-1, 1.5))
```

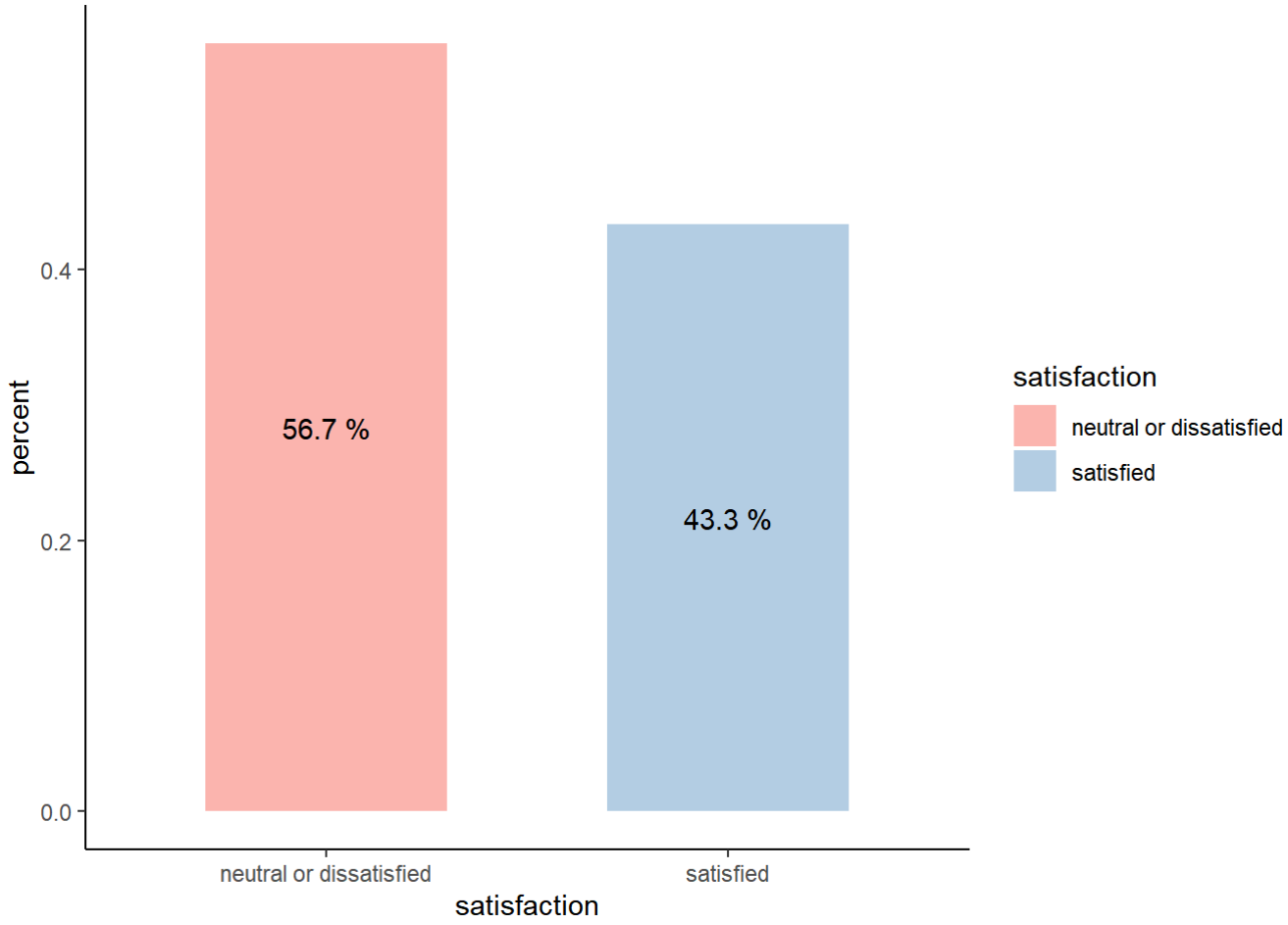



```
c_chart<-function(full=T,data,var,text=T,alpha=1){

  graph<-data %>%
    group_by({{var}}) %>%
    summarise(count=n()) %>%
    mutate(percent=count/sum(count),
           ymax=cumsum(percent),
           ymin=ymax-percent,
           label=paste((round(percent,3)*100) %>% as.character,'%'),
           labelpos=ymax-percent/2) %>%
    ggplot()+
    geom_rect(aes(xmin=2,xmax=4,ymin=ymin,ymax=ymax,fill={{var}}),
              alpha=alpha)+
    coord_polar(theta='y')+
    theme_void()

  if(full==T){graph<-graph}else{
    graph<-graph+xlim(0,4)}

  if(text==T){
    graph<-graph+geom_text(aes(x=3.2,y=labelpos,label=label))}else{
      graph<-graph
    }

  return(graph)
  }
```

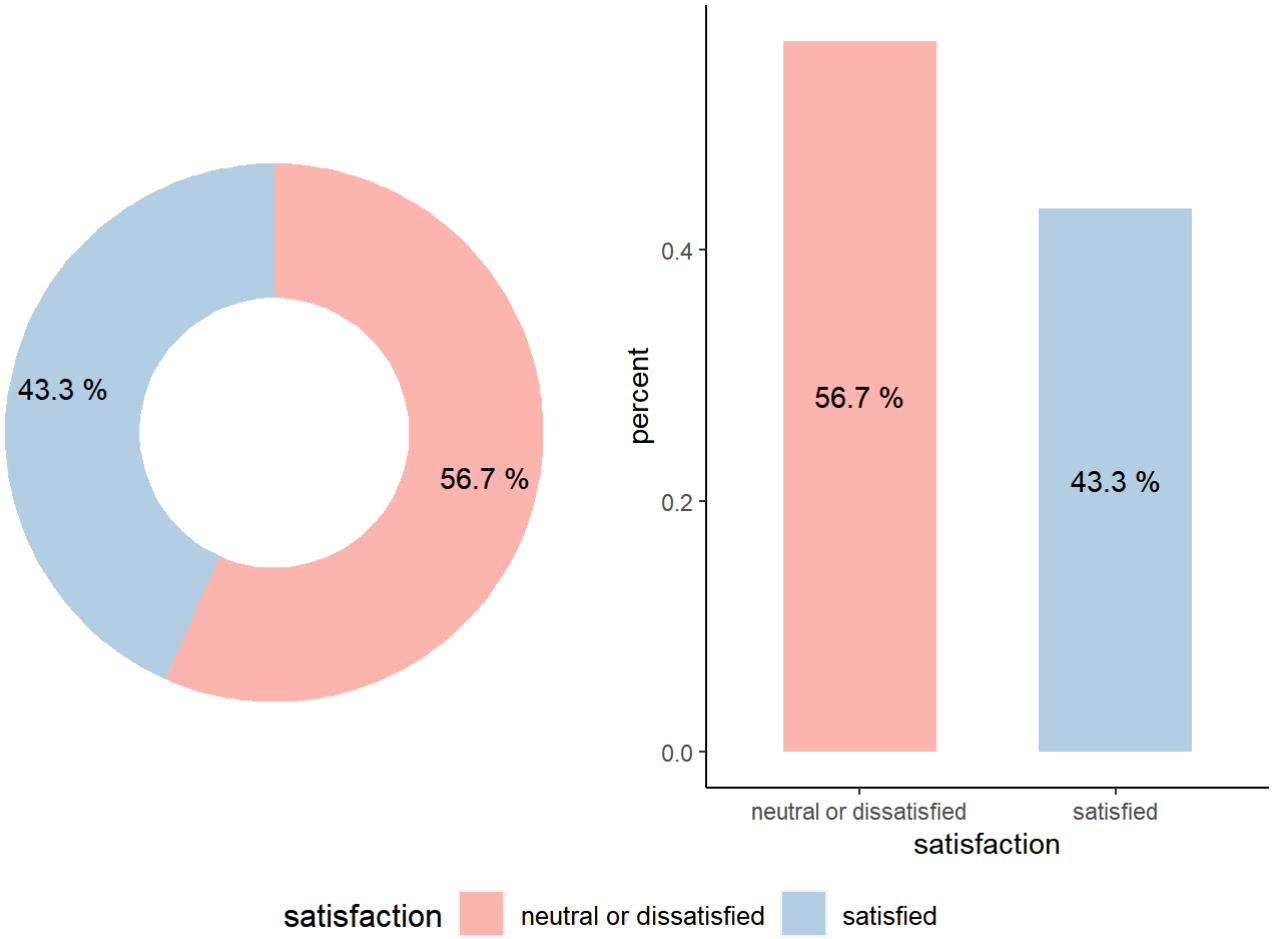
```
col <- hcl.colors(6, palette ="Blues")

c_chart(data=train,var=`Ease of Online booking`,
        text=T,full=T,alpha=0.75)+
  scale_fill_manual(values=col)
```




```
bar
```

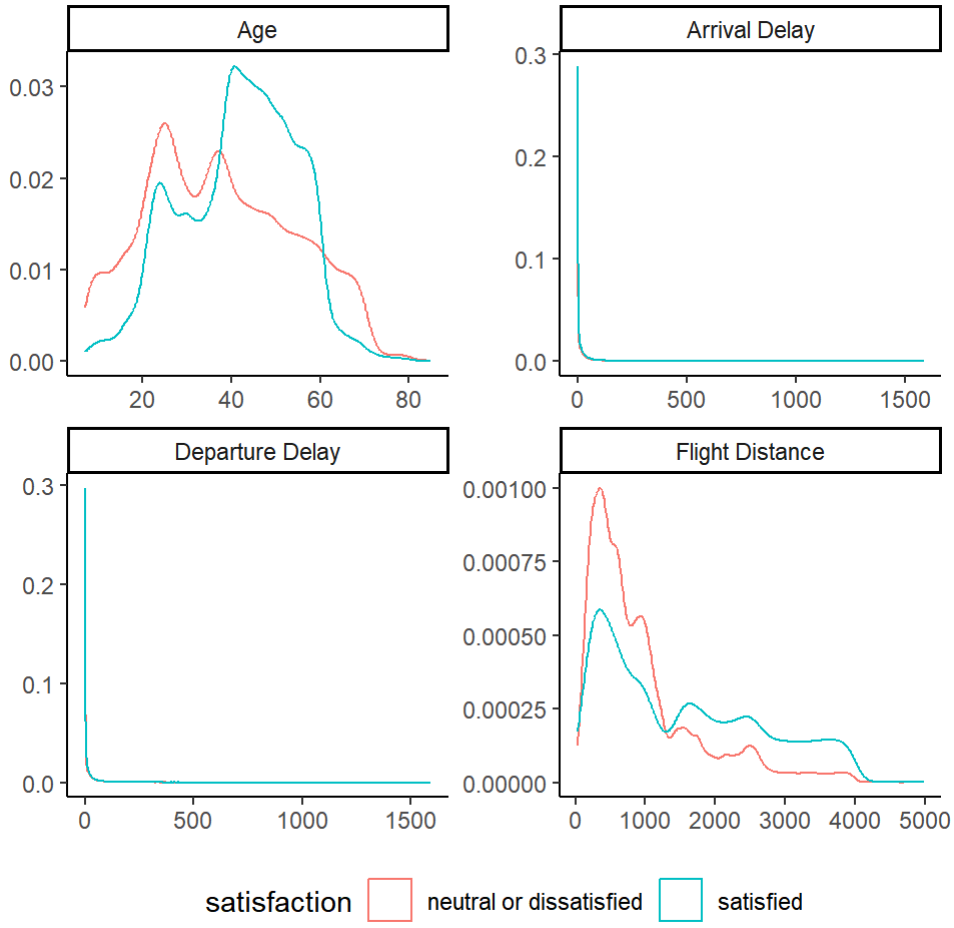



```
ggarrange(donut,bar,ncol=2,common.legend=TRUE,legend='bottom')+
  theme(plot.title=element_text(hjust=0.5,vjust=-1,
                                size=20,face='bold'))
```




```
train %>% select(c(num,'satisfaction')) %>%
  gather(num,key='key',value='value') %>%
  group_by(satisfaction,key,value) %>%
  ggplot()+
  geom_density(mapping=aes(x=value,color=satisfaction,group=satisfaction))+
  facet_wrap(vars(key),ncol=length(num)/2,scales="free")+
  theme_classic()+
  theme(legend.position = "bottom")+
  labs(x=NULL,y=NULL)
```

```
## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
## ℹ Please use `all_of()` or `any_of()` instead.
##   # Was:
##   data %>% select(num)
##
##   # Now:
##   data %>% select(all_of(num))
##
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```


Age
0.03
0.02
0.01
0.00
20
40
60
80
Arrival Delay
0.3
0.2
0.1
0.0
0
500
1000
1500
Departure Delay
0.3
0.2
0.1
0.0
0
500
1000
1500
Flight Distance
0.00100
0.00075
0.00050
0.00025
0.00000
0
1000
2000
3000
4000
5000
satisfaction
neutral or dissatisfied
satisfied