By. Data "Bike_Sharing_Demand

김슬기 이찬희

By. Data Bike Sharing Demand

CONTENTS 1
기본 library 설정

CONTENTS 3
자전거 수요 상관관계 (선택 요소)

CONTENTS 2
자전거 수요 상관관계 (모든 요소)

CONTENTS 4
최종결론

By. Data Bike Sharing Demand

CONTENTS 1

```
install.packages("dplyr") #for mutate()
library(dplyr)

install.packages("stringr")   #for substr()
library(stringr)

df<-df%>%mutate(work=ifelse(workingday==0,0,
                            ifelse(holiday==1,0,1 )))

df<-df%>%mutate(rest=ifelse(workingday==0,1,
                            ifelse(holiday==1,1,0)))

df<-df%>%mutate(Date=substr(df$datetime,0,10))
df$date<-weekdays(as.Date(df$Date))
```

패캐지 “dplr”을 설치

함수 mutate()를 사용

패캐지 “stringr”을 설치

함수 substr()을 사용

datetime에서 날짜를 나타내는 Date 생성

Date를 통해 weekdays()함수로 요일을 나타내는 date를 생성

By. Data Bike Sharing Demand

CONTENTS 1

```
Install.packages('ggplot2')
library(ggplot2)

Install.packages('gridExtra')
library(gridExtra)

grid.arrange(p11,p12, p21,p22,p23,p24,p25,p26,p27,p28, p3, p4,p51,p52,p6,p71,p72,p73,p74,p75,p76,p77,p78,
             ncol = 5, nrow = 5, top = "Bike_Sharing_Demand")

grid.arrange(p12, p21,p22,p23,p24, p3,p51,p6,p77,
             ncol = 3, nrow = 3, top = "Bike_Sharing_Demand")
```

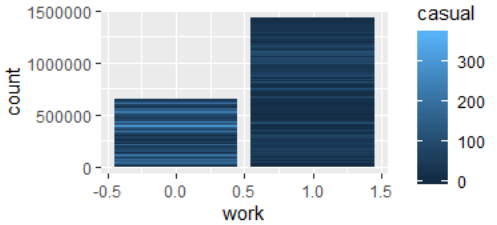
패키지를 사용해서 데이터를 효과적으로 시각화

복잡성 높은 시각화 산출물을 단계별 접근법을 통해 생성

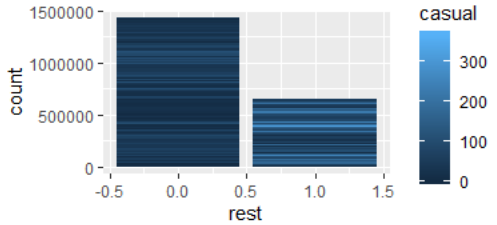
산점도, 상자그림, 시계열 그래프를 생성할 수 있게 된다.

Bike_Sharing_Demand

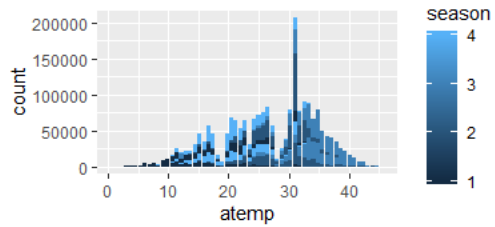
노동과 자전거 수요의 상관관계



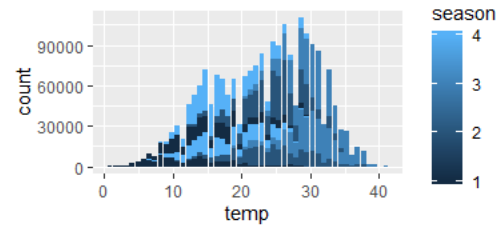
휴일과 자전거 수요의 상관관계



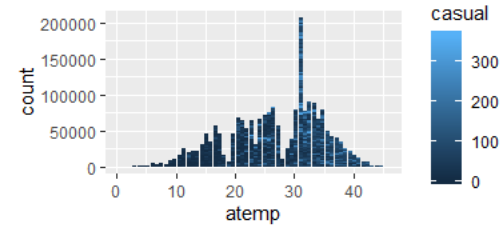
체감기온과 자전거 수요의 상관관계



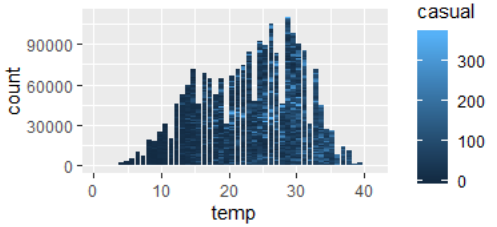
기온과 자전거 수요의 상관관계



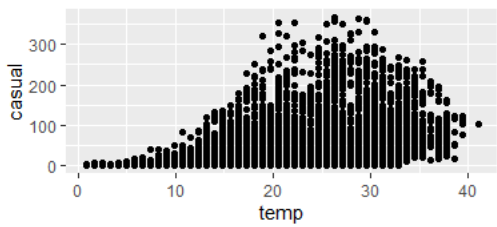
체감기온과 자전거 수요의 상관관계



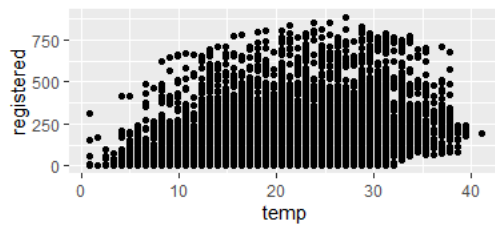
기온과 자전거 수요의 상관관계



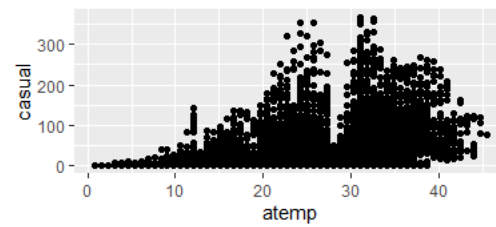
기온과 비등록자 자전거 수요의 상관관계



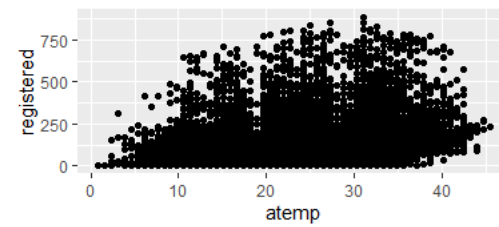
기온과 등록자 자전거 수요의 상관관계



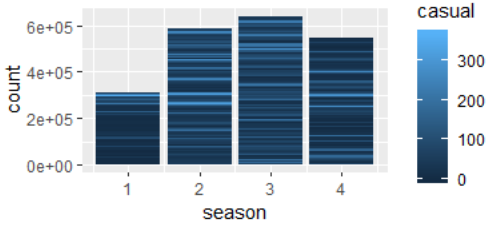
체감기온과 비등록자 자전거 수요의 상관관계



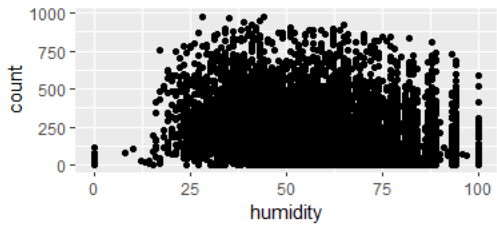
체감기온과 등록자 자전거 수요의 상관관계



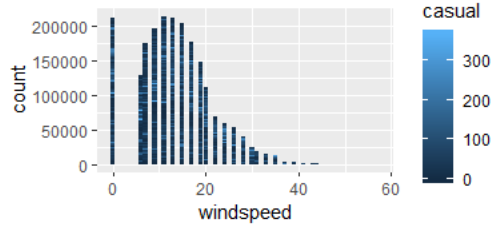
계절과 자전거 수요의 상관관계



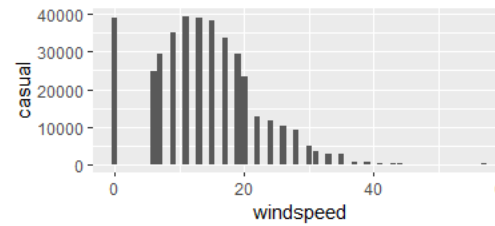
습도와 자전거 수요의 상관관계



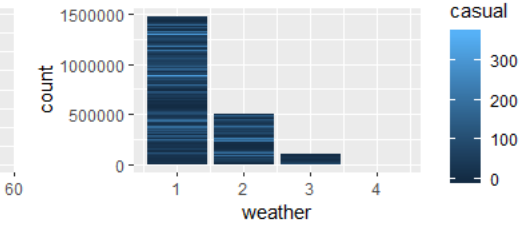
풍속과 자전거 수요의 상관관계



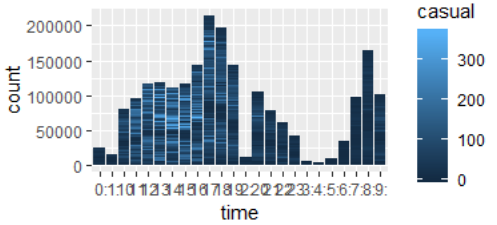
풍속과 비등록자자전거 수요의 상관관계



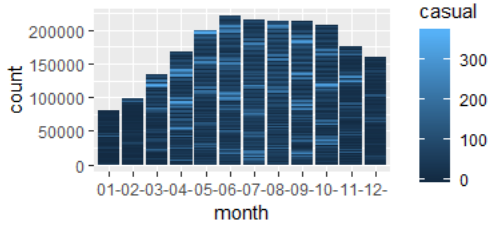
날씨와 자전거 수요의 상관관계



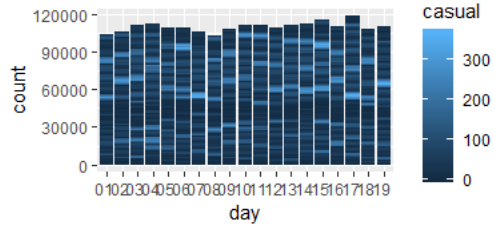
시간대와 자전거 수요의 상관관계



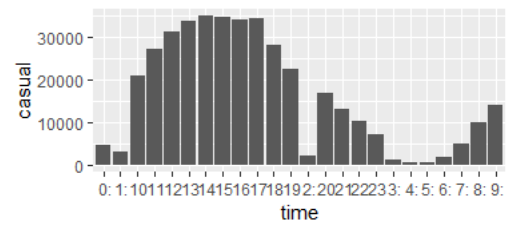
월별 자전거 수요



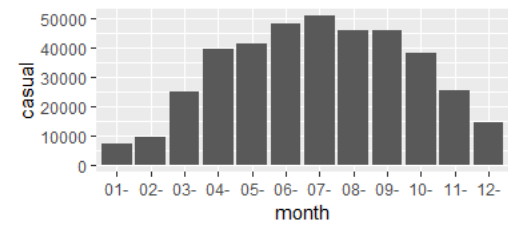
날짜와 자전거 수요의 상관관계



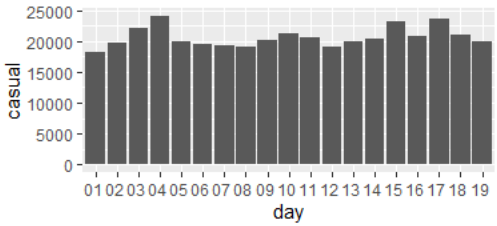
시간대와 비등록자 자전거 수요의 상관관계



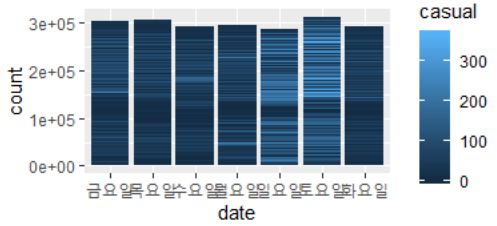
월별 비등록자 자전거 수요



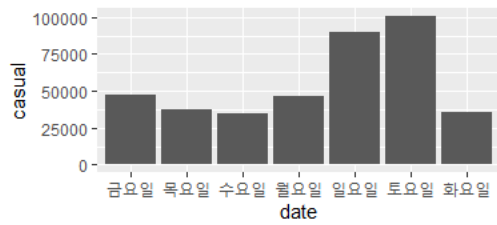
날짜와 비등록자 자전거 수요의 상관관계



요일과 자전거 수요의 상관관계



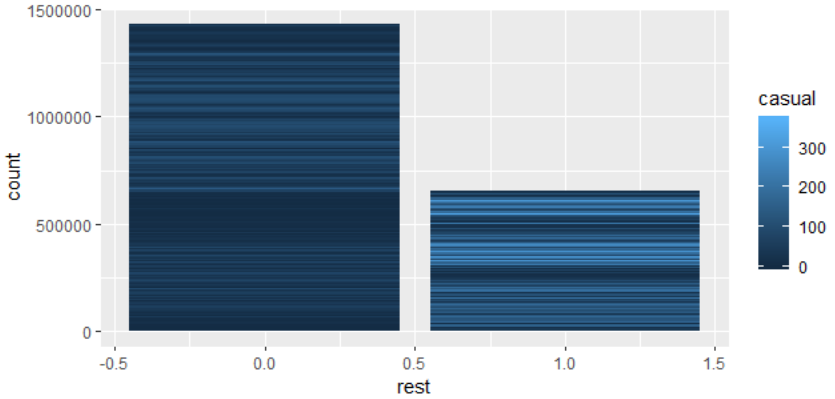
요일과 비등록자 자전거 수요의 상관관계



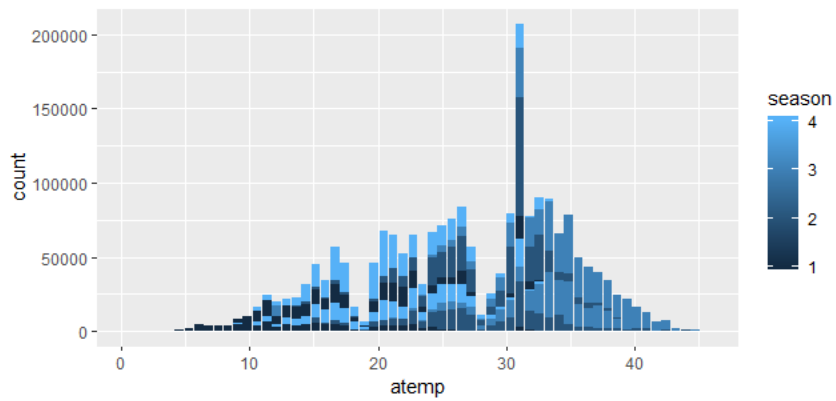
By. Data Bike Sharing Demand

Bike_Sharing_Demand

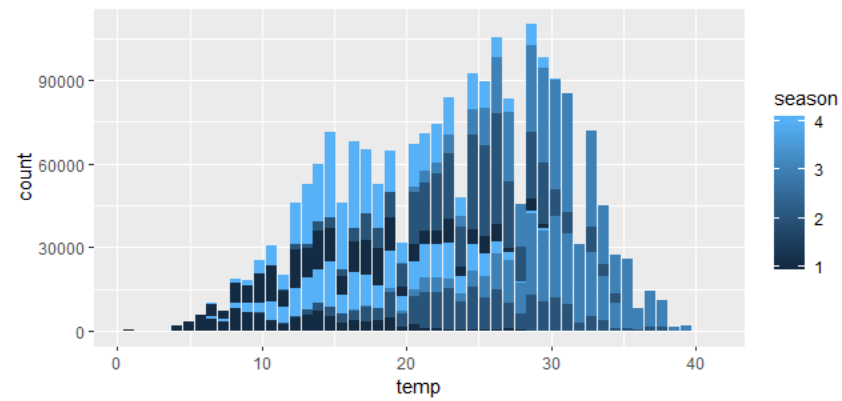
휴일과 자전거 수요의 상관관계



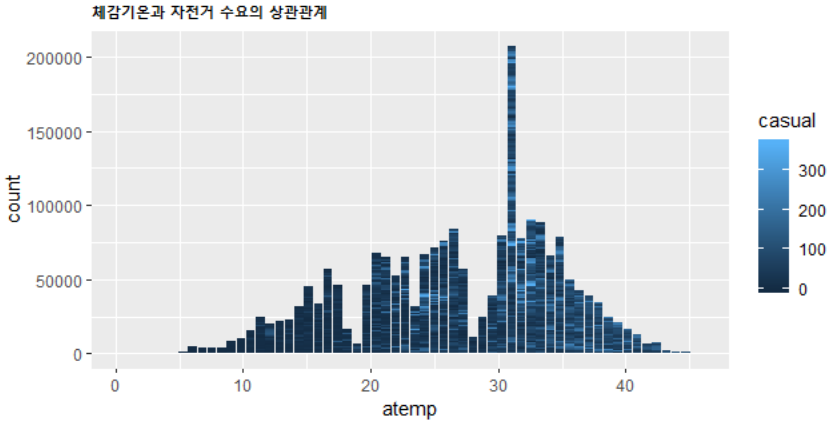
체감기온과 자전거 수요의 상관관계



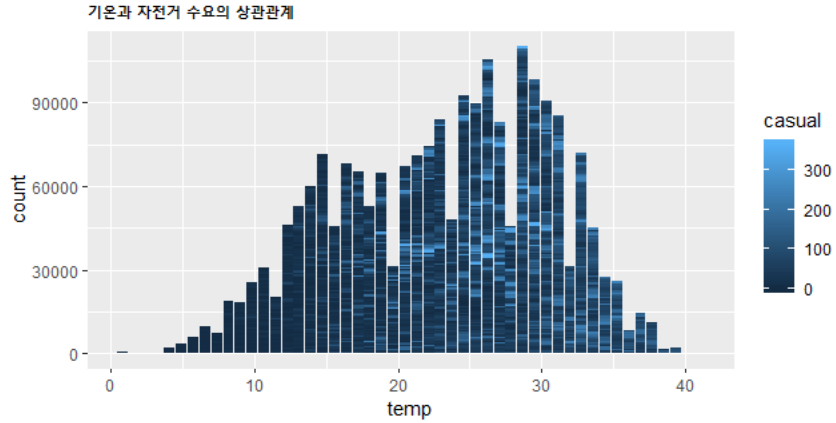
기온과 자전거 수요의 상관관계



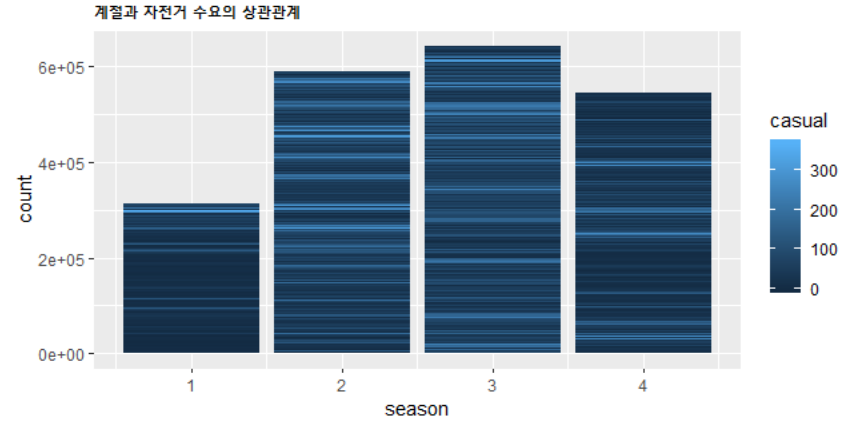
체감기온과 자전거 수요의 상관관계



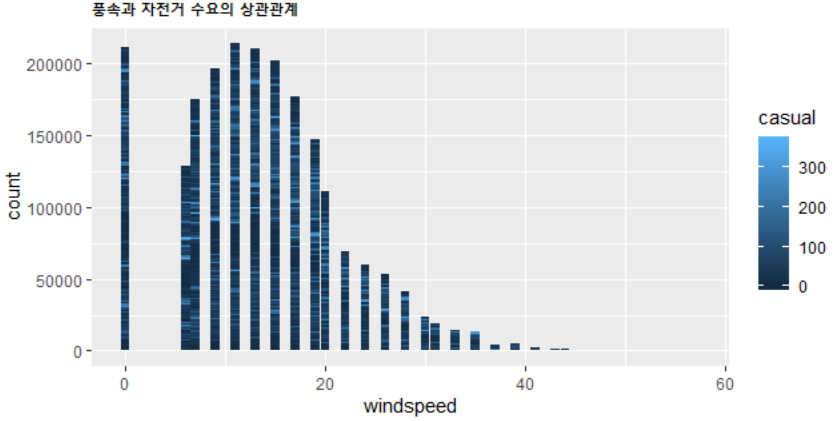
기온과 자전거 수요의 상관관계



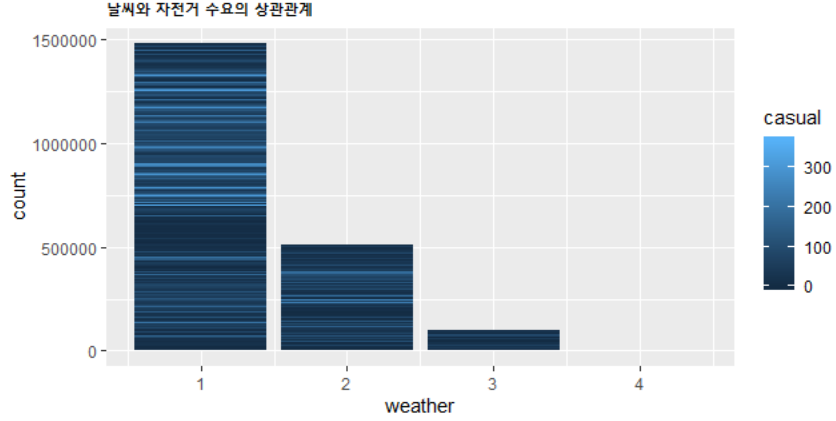
계절과 자전거 수요의 상관관계



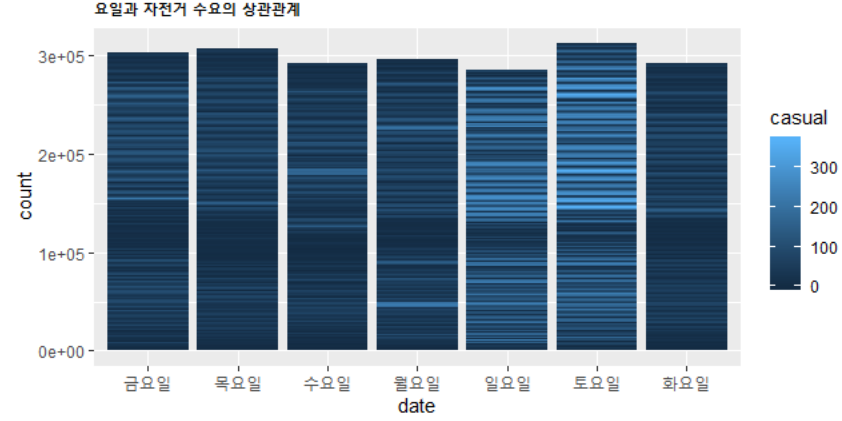
풍속과 자전거 수요의 상관관계



날씨와 자전거 수요의 상관관계



요일과 자전거 수요의 상관관계



By. Data Bike Sharing Demand

CONTENTS 2

```
> summary(df$count)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0    42.0   145.0   191.6   284.0   977.0
> df<-df%>%mutate(users=ifelse(count>191.6,1,0))
```

```
df<-df%>%mutate(tem=ifelse(temp<20,0,
                           ifelse(temp<25,1,
                                  ifelse(temp<34,2,3))))
df<-df%>%mutate(atem=ifelse(atemp<20,0,
                            ifelse(atemp<25,1,
                                   ifelse(atemp<30,2,
                                          ifelse(atemp<35,3,4)))))
df<-df%>%mutate(wind=ifelse(windspeed<10,0,
                            ifelse(windspeed<20,1,
                                   ifelse(windspeed<30,2,3))))
```

수치 데이터를
일정한 개수로
단위를 나누어
새로운 컬럼 형성

```
> D<-df%>%filter(users==1)
> dim(D)
[1] 4356   23
> d<-D%>%select(date)
> table(d)
d
금요일 목요일 수요일 월요일 일요일 토요일 화요일
   687    639    584    614    589    645    598
> D<-D%>%filter(date=='금요일')
> dim(D)
[1] 687  23
> d<-D%>%select(rest)
> table(d)
d
  0   1
670  17
> D<-D%>%filter(rest==0)
> dim(D)
[1] 670  23
> d<-D%>%select(season)
> table(d)
d
  1   2   3   4
 90 184 222 174
> D<-D%>%filter(season==3)
> dim(D)
[1] 222  23
```

```
> d<-D%>%select(tem)
> table(d)
d
  0   1   2   3
  3  12 185  22
> D<-D%>%filter(tem==2)
> dim(D)
[1] 185  23
> d<-D%>%select(atem)
> table(d)
d
  0   2   3   4
 12  13 113  47
> D<-D%>%filter(atem==3)
> dim(D)
[1] 113  23
> d<-D%>%select(weather)
> table(d)
d
 1  2  3
82 30  1
> D<-D%>%filter(weather==1)
> dim(D)
[1] 82 23
> d<-D%>%select(Wind)
> table(d)
d
 0  1
32 50
```

By. Data Bike Sharing Demand

CONTENTS 2

```
#가을을 기준으로
D<-df%>%mutate(situation=ifelse(date!='금요일',0,
                                ifelse(rest!=0,0,
                                       ifelse(season!=3,0,
                                              ifelse(tem!=2,0,
                                                     ifelse(atem!=3,0,
                                                            ifelse(weather!=1,0,
                                                                   ifelse(wind!=1,0,1)))))),

               people=ifelse(count<42,"1.more less",
                             ifelse(count<192,"2.less",
                                    ifelse(count<284,"3.more","4.too more"))))



#봄을 기준으로
d<-df%>%mutate(situation=ifelse(date!='금요일',0,
                                ifelse(rest!=0,0,
                                       ifelse(season==1,0,
                                              ifelse(tem!=2,0,
                                                     ifelse(atem!=3,0,
                                                            ifelse(weather!=1,0,
                                                                   ifelse(wind!=1,0,1)))))),
               unregister=ifelse(casual<4,"1.more less",
                                 ifelse(casual<36,"2.less",
                                        ifelse(casual<49,"3.more","4.too more"))),
               register=ifelse(registered<36,"1.more less",
                               ifelse(registered<156,"2.less",
                                      ifelse(registered<222,"3.more","4.too more"))),
               people=ifelse(count<42,"1.more less",
                             ifelse(count<192,"2.less",
                                    ifelse(count<284,"3.more","4.too more"))))
```

어떤 조건에

자전거 수요가 많은지

알아보기 위한 데이터를

시각화하기 위해

새로운 컬럼 형성

Bike_Sharing_Demand of number of total rentals