Healthcare Appointment No-Show

Samiksha Borade

2023-02-15

```
getwd()
```

```
## [1] "C:/Users/HP/Documents"
```

```
appointments <- read.csv("C:/Users/HP/Desktop/appointments.csv")
View(appointments)
library(arules)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'arules'
```

```
## The following objects are masked from 'package:base':
##
##     abbreviate, write
```

```
library(ggplot2)
library(gridExtra)
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following object is masked from 'package:gridExtra':
##
##     combine
```

```
## The following objects are masked from 'package:arules':
##
##     intersect, recode, setdiff, setequal, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
## Loading required package: timechange
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
## The following objects are masked from 'package:arules':
##
##     intersect, setdiff, union
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(caTools)
library("MASS")
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(magrittr)
library(rpart)
library(rpart.plot)
```

```
head(appointments)
```

```
##   Age Gender AppointmentRegistration      AppointmentDate Diabetes Alcohol
ism
## 1  19      M    2014-12-16T14:46:25Z 2015-01-14T00:00:00Z        0
0
## 2  24      F    2015-08-18T07:01:26Z 2015-08-19T00:00:00Z        0
0
## 3   4      F    2014-02-17T12:53:46Z 2014-02-18T00:00:00Z        0
0
## 4   5      M    2014-07-23T17:02:11Z 2014-08-07T00:00:00Z        0
0
## 5  38      M    2015-10-21T15:20:09Z 2015-10-27T00:00:00Z        0
0
## 6   5      F    2014-06-17T06:47:27Z 2014-07-22T00:00:00Z        0
0
##   HyperTension Handicap Smokes Scholarship Tuberculosis Sms_Reminder  Stat
us
```

```
## 1            0        0      0           0            0            0 Show-
Up
## 2            0        0      0           0            0            0 Show-
Up
## 3            0        0      0           0            0            0 Show-
Up
## 4            0        0      0           0            0            1 Show-
Up
## 5            0        0      0           0            0            1 Show-
Up
## 6            0        0      0           0            0            1 No-Sh
ow
```

```
names(appointments)
```

```
##  [1] "Age"                     "Gender"
##  [3] "AppointmentRegistration" "AppointmentDate"
##  [5] "Diabetes"                "Alcoholism"
##  [7] "HyperTension"            "Handicap"
##  [9] "Smokes"                  "Scholarship"
## [11] "Tuberculosis"            "Sms_Reminder"
## [13] "Status"
```

```
str(appointments)
```

```
## 'data.frame':    300000 obs. of  13 variables:
##  $ Age                   : int  19 24 4 5 38 5 46 4 20 51 ...
##  $ Gender                : chr  "M" "F" "F" "M" ...
##  $ AppointmentRegistration: chr  "2014-12-16T14:46:25Z" "2015-08-18T07:01:
26Z" "2014-02-17T12:53:46Z" "2014-07-23T17:02:11Z" ...
##  $ AppointmentDate       : chr  "2015-01-14T00:00:00Z" "2015-08-19T00:00:
00Z" "2014-02-18T00:00:00Z" "2014-08-07T00:00:00Z" ...
##  $ Diabetes              : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Alcoholism            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ HyperTension          : int  0 0 0 0 0 0 0 0 0 1 ...
##  $ Handicap              : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Smokes                : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Scholarship           : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Tuberculosis          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Sms_Reminder          : int  0 0 0 1 1 1 1 1 0 1 ...
##  $ Status                : chr  "Show-Up" "Show-Up" "Show-Up" "Show-Up" .
..
```

```
appointments$Gender <- as.factor(appointments$Gender)
appointments$Diabetes <- as.factor(appointments$Diabetes)
appointments$Alcoholism <- as.factor(appointments$Alcoholism)
appointments$HyperTension <- as.factor(appointments$HyperTension)
appointments$Handicap <- as.factor(appointments$Handicap)
appointments$Smokes <- as.factor(appointments$Smokes)
appointments$Scholarship <- as.factor(appointments$Scholarship)
appointments$Tuberculosis <- as.factor(appointments$Tuberculosis)
```

```r
appointments$Sms_Reminder <- as.factor(appointments$Sms_Reminder)
appointments$Status <- as.factor(appointments$Status)

sapply(appointments,function(x)sum(is.na(x)))
```

```
##                      Age                   Gender AppointmentRegistration
##                        0                        0                        0
##          AppointmentDate                 Diabetes               Alcoholism
##                        0                        0                        0
##             HyperTension                 Handicap                   Smokes
##                        0                        0                        0
##              Scholarship             Tuberculosis             Sms_Reminder
##                        0                        0                        0
##                   Status
##                        0
```

```r
lapply(appointments, class)
```

```
## $Age
## [1] "integer"
##
## $Gender
## [1] "factor"
##
## $AppointmentRegistration
## [1] "character"
##
## $AppointmentDate
## [1] "character"
##
## $Diabetes
## [1] "factor"
##
## $Alcoholism
## [1] "factor"
##
## $HyperTension
## [1] "factor"
##
## $Handicap
## [1] "factor"
##
## $Smokes
## [1] "factor"
##
## $Scholarship
## [1] "factor"
##
## $Tuberculosis
## [1] "factor"
```

```
##
## $Sms_Reminder
## [1] "factor"
##
## $Status
## [1] "factor"
```

```
summary(appointments)
```

```
##       Age          Gender     AppointmentRegistration AppointmentDate
##  Min.   : -2.00   F:200505   Length:300000           Length:300000
##  1st Qu.: 19.00   M: 99495   Class :character        Class :character
##  Median : 38.00              Mode  :character        Mode  :character
##  Mean   : 37.81
##  3rd Qu.: 56.00
##  Max.   :113.00
##  Diabetes   Alcoholism HyperTension Handicap   Smokes     Scholarship
##  0:276610   0:292497   0:235233     0:294403   0:284289   0:270931
##  1: 23390   1:  7503   1: 64767     1:  5098   1: 15711   1: 29069
##                                     2:   449
##                                     3:    39
##                                     4:    11
##
##  Tuberculosis Sms_Reminder     Status
##  0:299865     0:128547     No-Show: 90731
##  1:   135     1:170654     Show-Up:209269
##               2:   799
##
##
##
```

```
date_2 = "2014-12-16T14:46:25Z"
date_2=as.POSIXct(sub("T"," ",date_2),format="%Y-%m-%d %H:%M:%S",tz=Sys.timez
one())
date_2
```

```
## [1] "2014-12-16 14:46:25 IST"
```

```
strftime(date_2,format="%H")
```

```
## [1] "14"
```

```
strftime(date_2,format = "%M")
```

```
## [1] "46"
```

```
strftime(date_2,format = "%S")
```

```
## [1] "25"
```

```
date_3 = "2015-08-19T00:00:00Z"
date_3=as.POSIXct(sub("T"," ",date_3),format="%Y-%m-%d %H:%M:%S",tz=Sys.timez
```

```
one())
date_3

## [1] "2015-08-19 IST"

strftime(date_3,format="%H")

## [1] "00"

strftime(date_3,format = "%M")

## [1] "00"

strftime(date_3,format = "%S")

## [1] "00"

date_4 =  "2013-12-30T08:56:51Z"
date_4=as.POSIXct(sub("T"," ",date_4),format="%Y-%m-%d %H:%M:%S",tz=Sys.timez
one())
date_4

## [1] "2013-12-30 08:56:51 IST"

strftime(date_4,format="%H")

## [1] "08"

strftime(date_4,format = "%M")

## [1] "56"

strftime(date_4,format = "%S")

## [1] "51"

date_5 =  "2014-07-25T15:02:33Z"
date_5=as.POSIXct(sub("T"," ",date_5),format="%Y-%m-%d %H:%M:%S",tz=Sys.timez
one())
date_5

## [1] "2014-07-25 15:02:33 IST"

strftime(date_5,format="%H")

## [1] "15"

strftime(date_5,format = "%M")

## [1] "02"

strftime(date_5,format = "%S")

## [1] "33"
```

```
appointments[appointments$Age < 0, ]
```

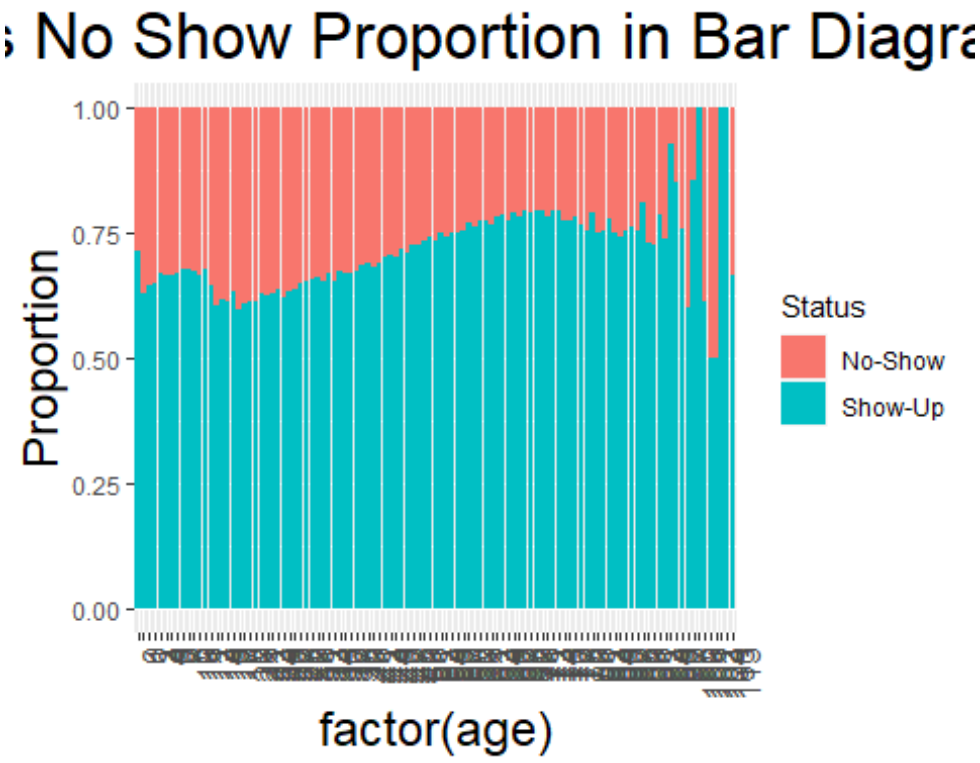
```
##        Age Gender AppointmentRegistration      AppointmentDate Diabetes
## 63391   -1      F    2014-03-14T11:39:20Z 2014-03-21T00:00:00Z        0
## 90974   -1      F    2013-12-26T08:21:55Z 2014-01-30T00:00:00Z        0
## 170052  -1      F    2015-07-30T07:48:10Z 2015-08-20T00:00:00Z        0
## 170599  -1      F    2015-05-11T15:39:33Z 2015-05-18T00:00:00Z        0
## 272821  -2      F    2013-12-18T16:09:18Z 2014-01-07T00:00:00Z        0
## 281910  -1      F    2014-01-23T11:23:22Z 2014-01-24T00:00:00Z        0
##        Alcoholism HyperTension Handicap Smokes Scholarship Tuberculosis
## 63391           0            0        0      0           0            0
## 90974           0            0        0      0           0            0
## 170052          0            0        0      0           0            0
## 170599          0            0        0      0           0            0
## 272821          0            0        0      0           0            0
## 281910          0            0        0      0           0            0
##        Sms_Reminder  Status
## 63391             1 No-Show
## 90974             1 Show-Up
## 170052            1 Show-Up
## 170599            1 Show-Up
## 272821            1 No-Show
## 281910            0 Show-Up
```

```
appointments <-appointments[!(appointments$Age<0),]
summary(appointments$Age)
```

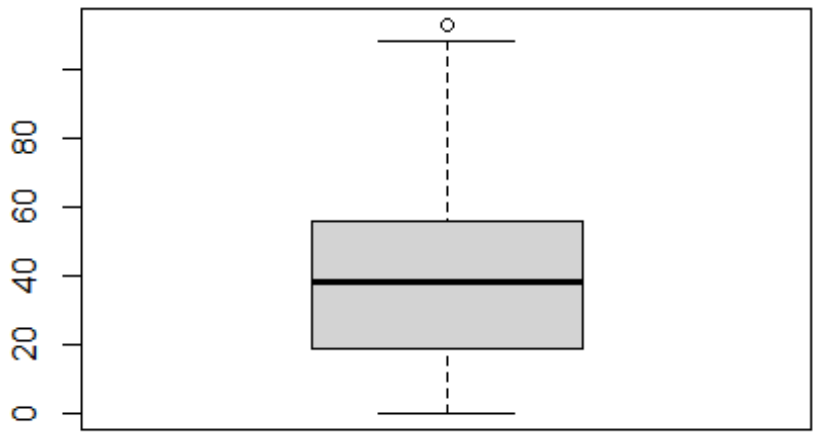
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   19.00   38.00   37.81   56.00  113.00
```

```
names(appointments)<- c('age','gender','appointment_Registration','appointmen
t_Date','diabetes',
                        'alcoholism','hypertension','handicap','smokes',
                        'scholoarship','tuberculosis','sms_reminder','Status'
)
```

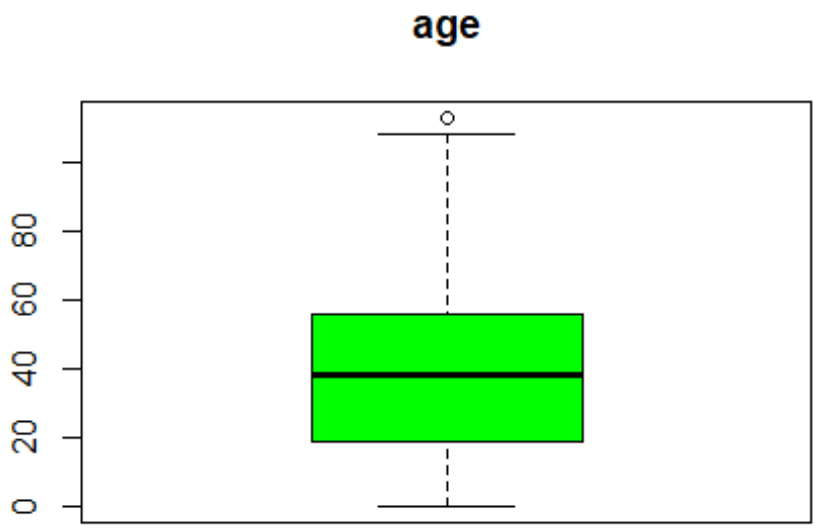
```
ggplot(data = appointments)+
geom_bar(aes(factor(age), fill = Status), position = position_fill())+
ggtitle("Age vs No Show Proportion in Bar Diagram")+
ylab('Proportion')+
theme(plot.title = element_text(hjust = 0.5, size = 24))+
theme(axis.title.y = element_text(size =18))+
theme(axis.title.x = element_text(size =18))+
theme(axis.text.x = element_text(size= 12, angle = 90, hjust = 1))
```



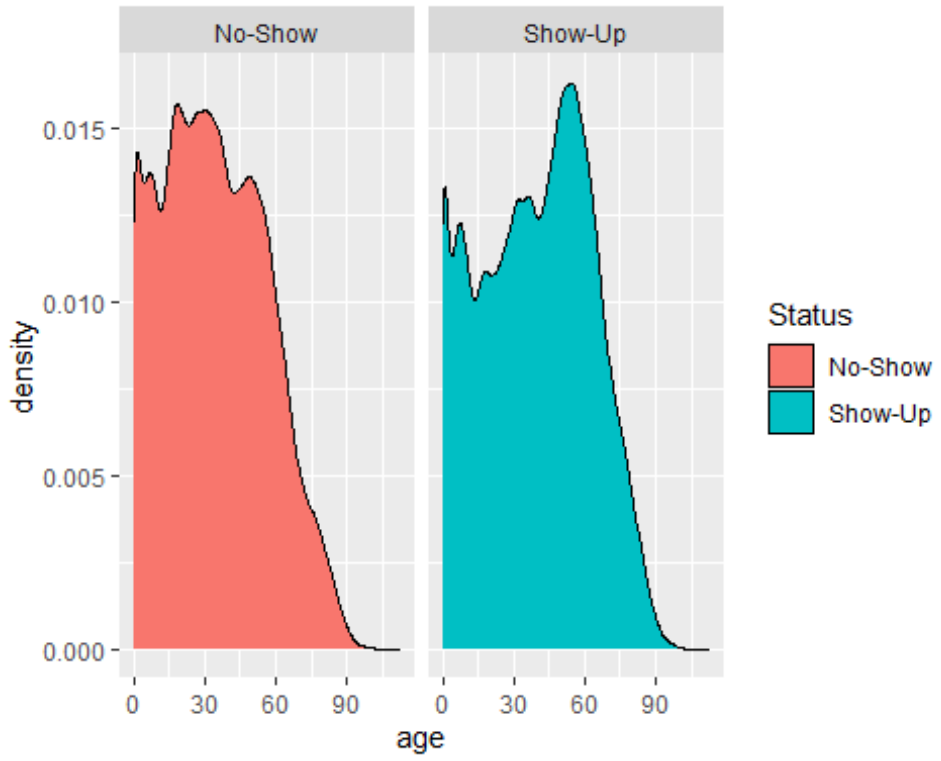

```
boxplot(appointments$age)
```



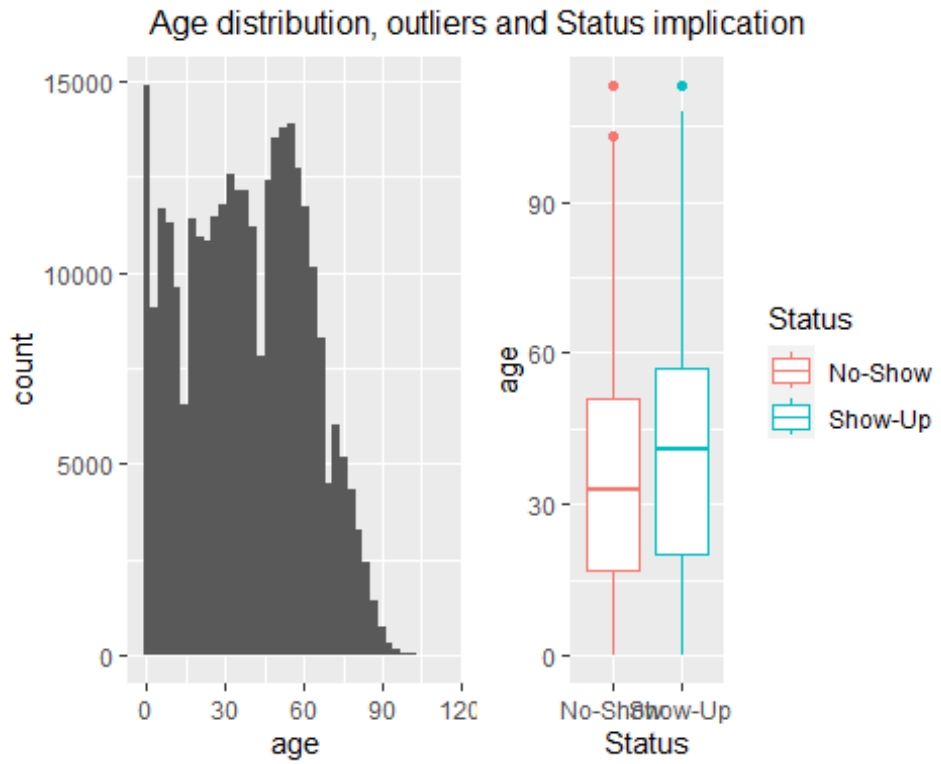
```
boxplot(appointments$age,
        main = "age",
        col = "green")
```



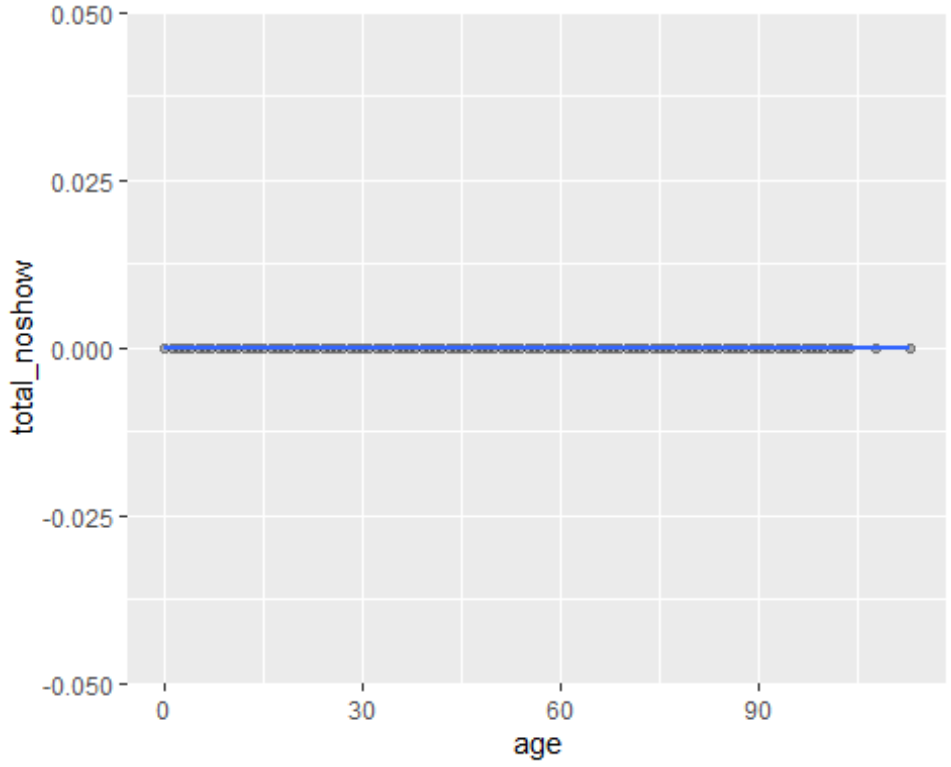

```
ggplot(appointments, aes(x=age, fill=Status)) + geom_density() + facet_grid(.
~Status)
```



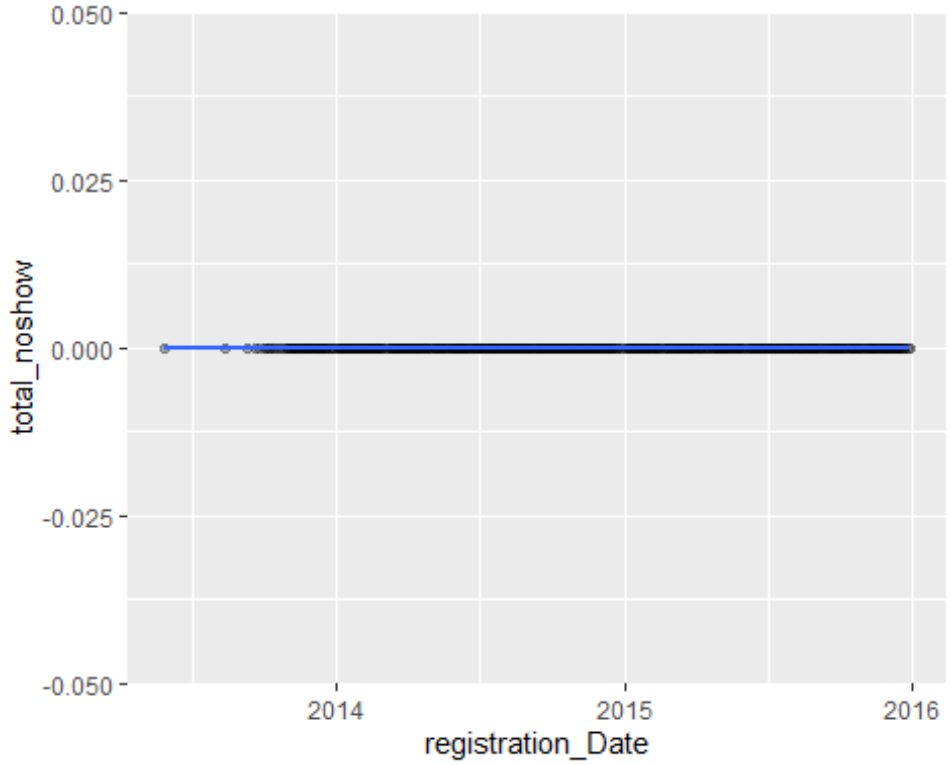

```
g_Age_1 <- ggplot(appointments, aes(x=age)) + geom_histogram(bins=40)
g_Age_2 <- ggplot(appointments, aes(x=Status, y=age, col=Status)) + geom_boxp
lot()
grid.arrange(g_Age_1, g_Age_2,ncol=2, top='Age distribution, outliers and Sta
tus implication')
```




```
appointments %>% group_by(age) %>% summarise(total_noshow=sum(Status=="No.Sho
w")/n()) %>% ggplot(aes(x=age, y=total_noshow)) +
  geom_point(alpha=0.3) + geom_smooth(method = "lm")

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




```
appointments %>% group_by(registration_Date=as.Date(appointment_Registration)
) %>% summarise(total_noshow=sum(Status=="No.Show")/n()) %>% ggplot(aes(x=reg
istration_Date, y=total_noshow)) + geom_point(alpha=0.3) + geom_smooth(method
= "lm")

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



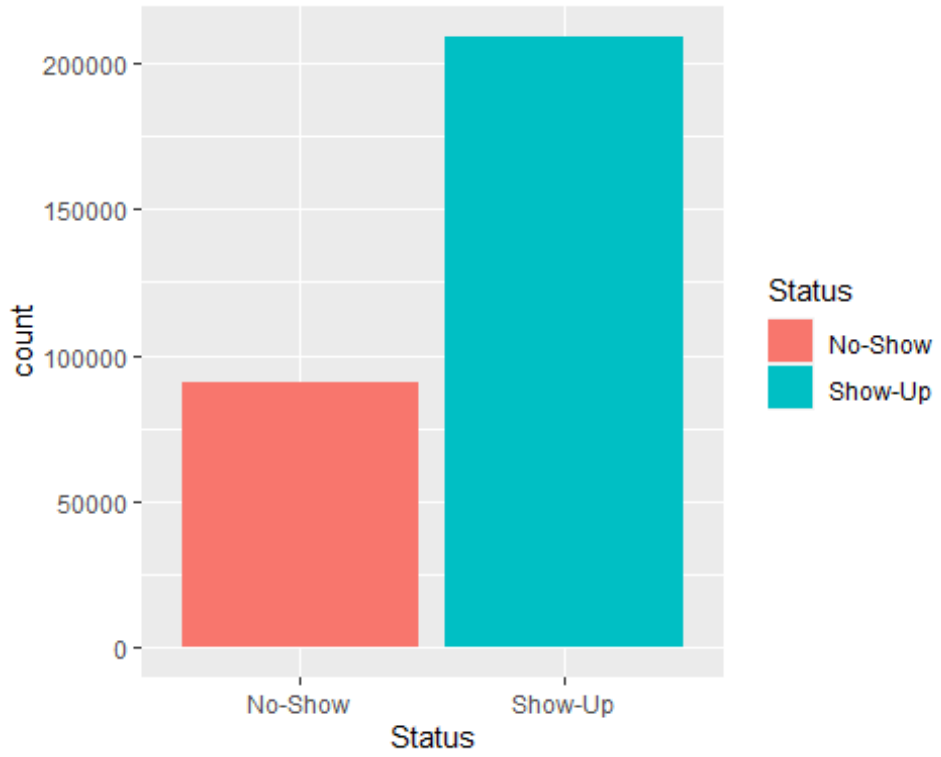

```
table(appointments$gender, appointments$Status)
```

```
##
##     No-Show Show-Up
##   F   59889  140610
##   M   30840   68655
```

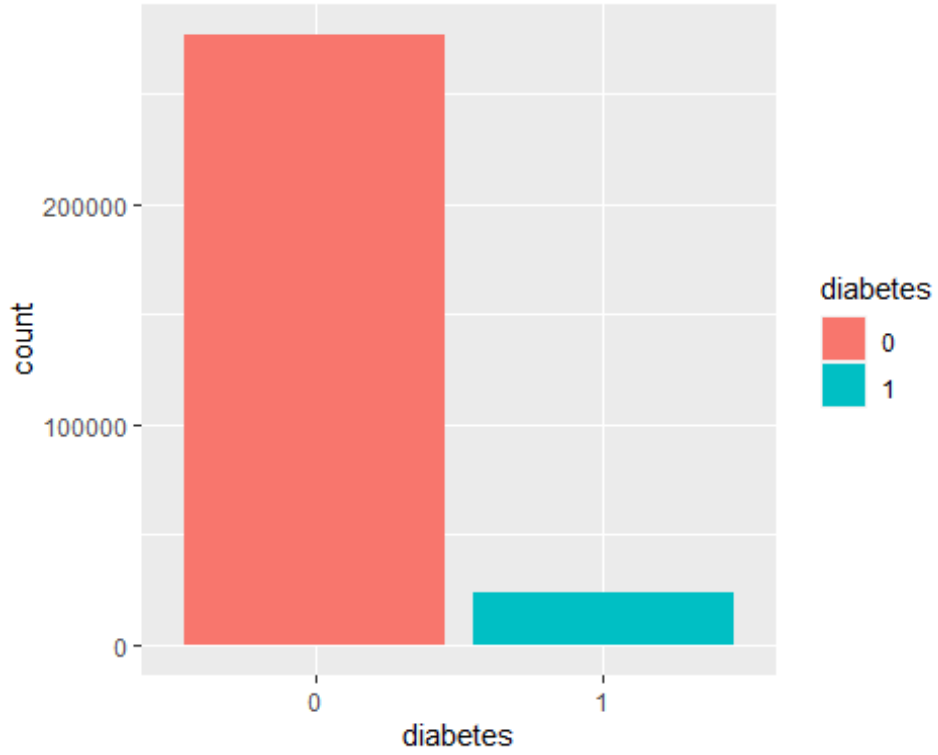
```
table(appointments$Status)
```

```
##
## No-Show Show-Up
##   90729  209265
```

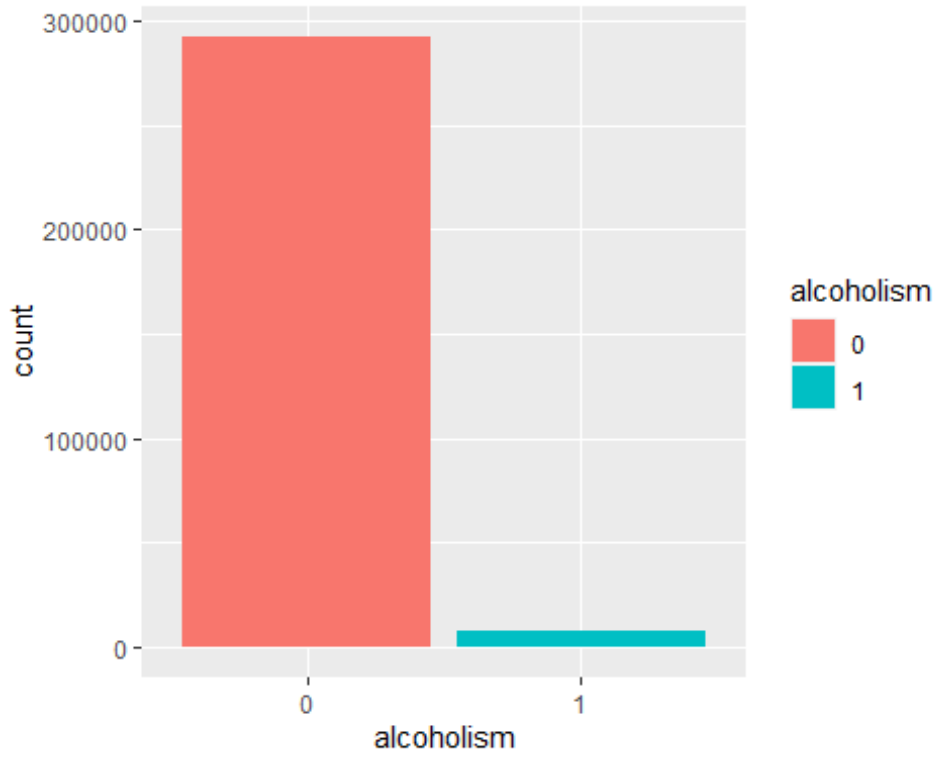
```
ggplot(appointments, aes(x=Status, fill=Status)) + geom_bar()
```




```
Status_table <- table(appointments$diabetes)
Status_table

##
##      0      1
## 276604  23390

options(scipen = 999)
ggplot(appointments, aes(x=diabetes, fill=diabetes)) + geom_bar()
```



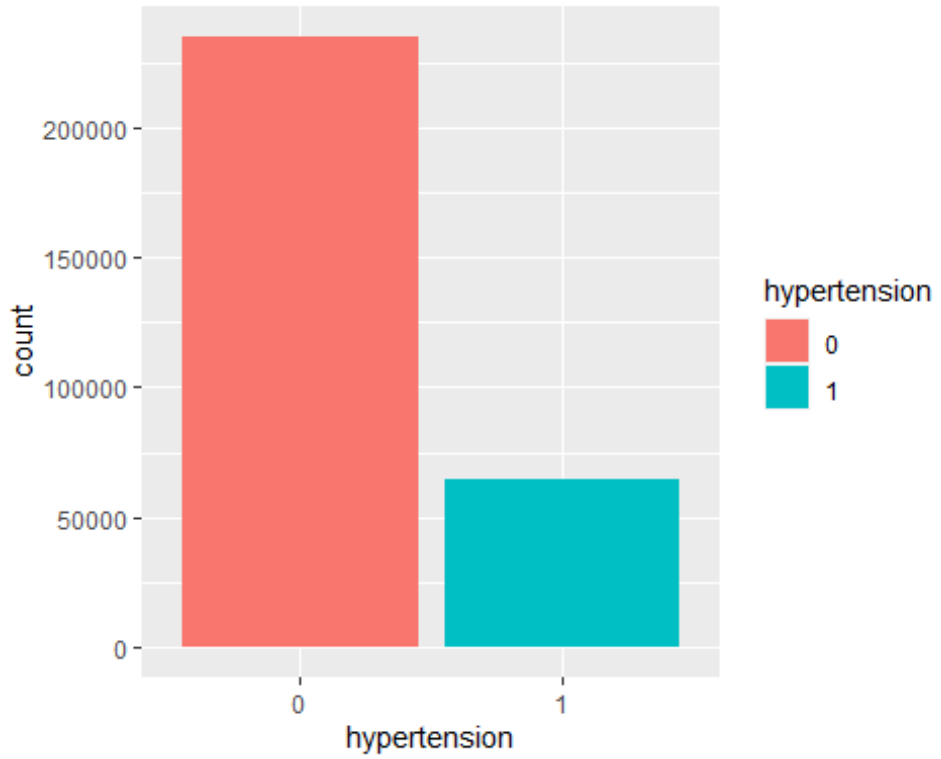

```
Status_table <- table(appointments$alcoholism)
Status_table

##
##      0      1
## 292491   7503

options(scipen = 999)
ggplot(appointments, aes(x=alcoholism, fill=alcoholism)) + geom_bar()
```



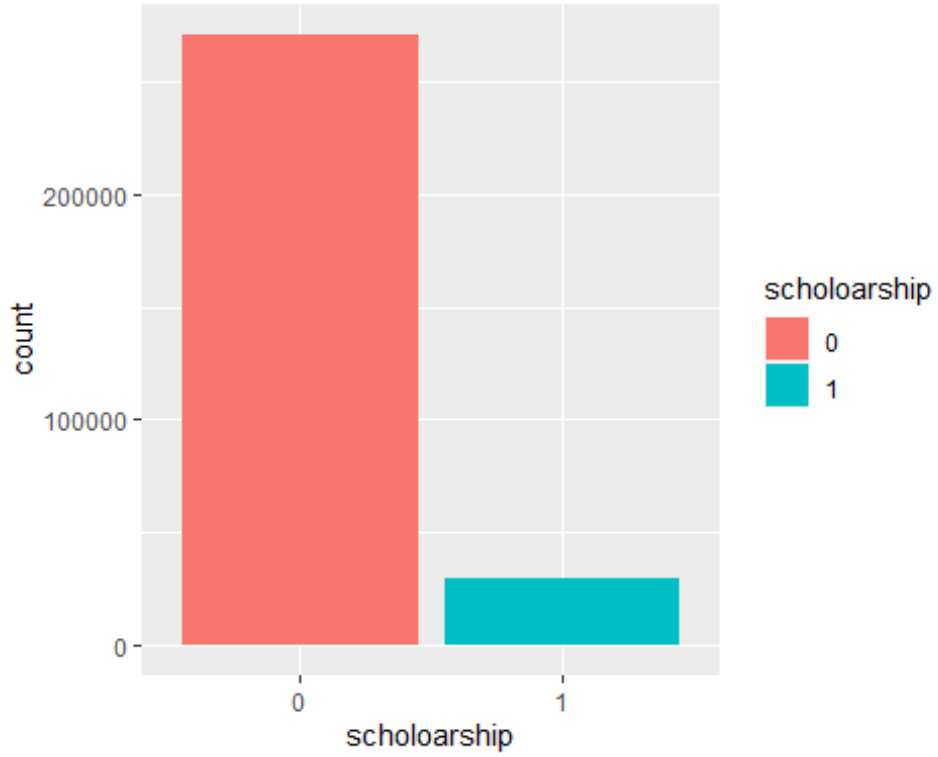

```
Status_table <- table(appointments$hypertension)
Status_table
```

```
##
##      0      1
## 235227  64767
```

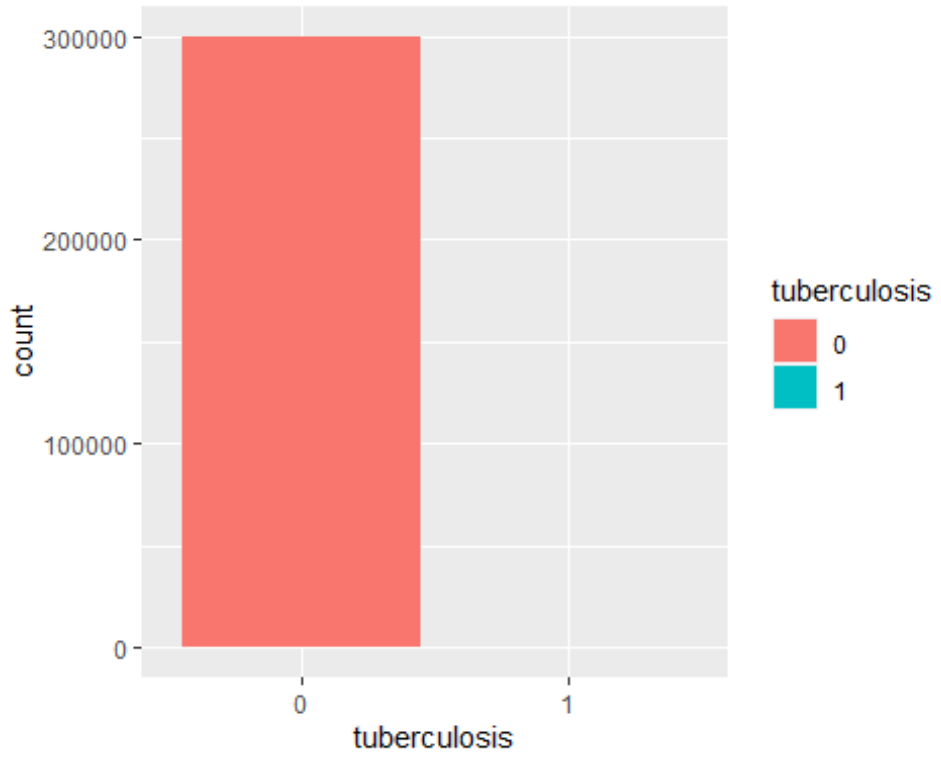
```
options(scipen = 999)
ggplot(appointments, aes(x=hypertension, fill=hypertension)) + geom_bar()
```



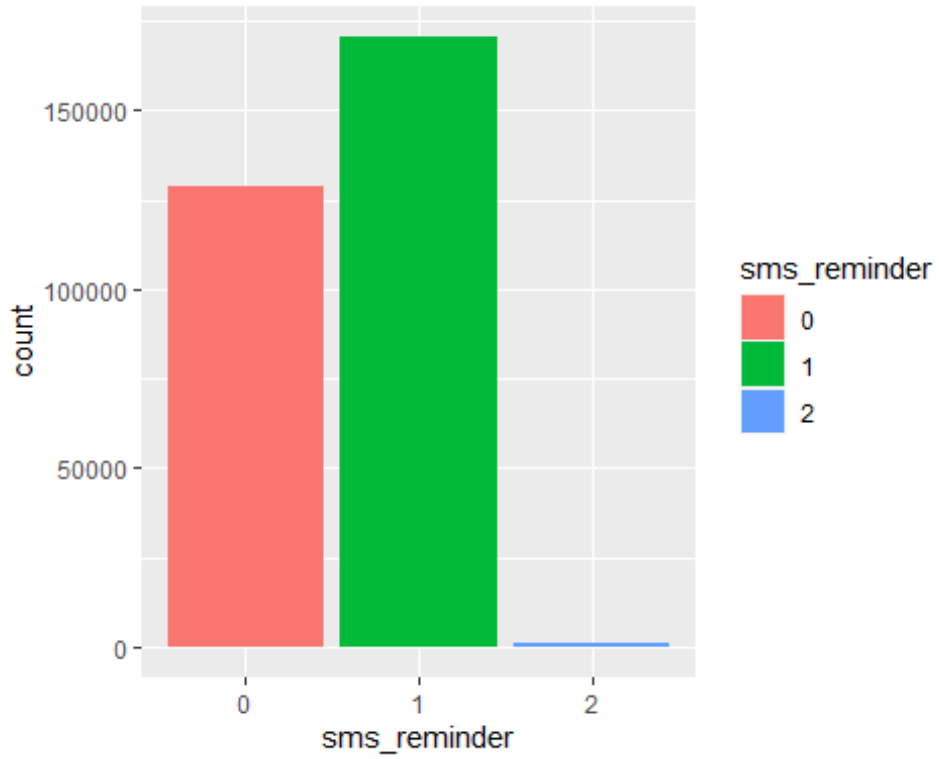

```
Status_table <- table(appointments$scholoarship)
Status_table

##
##      0      1
## 270925  29069

options(scipen = 999)
ggplot(appointments, aes(x=scholoarship, fill=scholoarship)) + geom_bar()
```



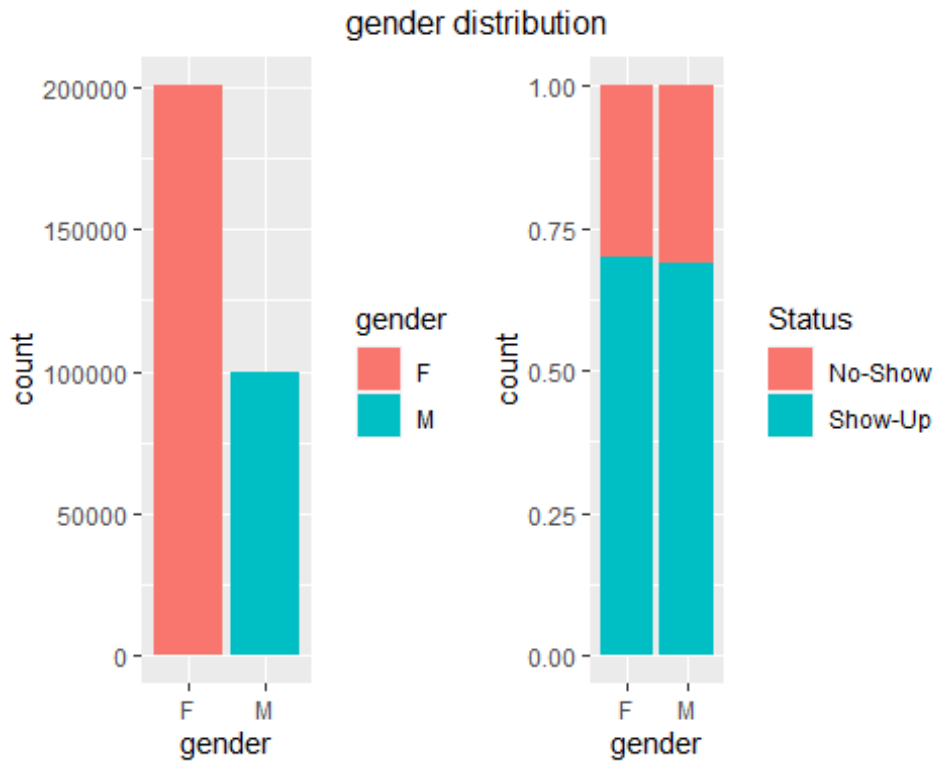

```
Status_table <- table(appointments$tuberculosis)
Status_table

##
##      0      1
## 299859    135

options(scipen = 999)
ggplot(appointments, aes(x=tuberculosis, fill=tuberculosis)) + geom_bar()
```



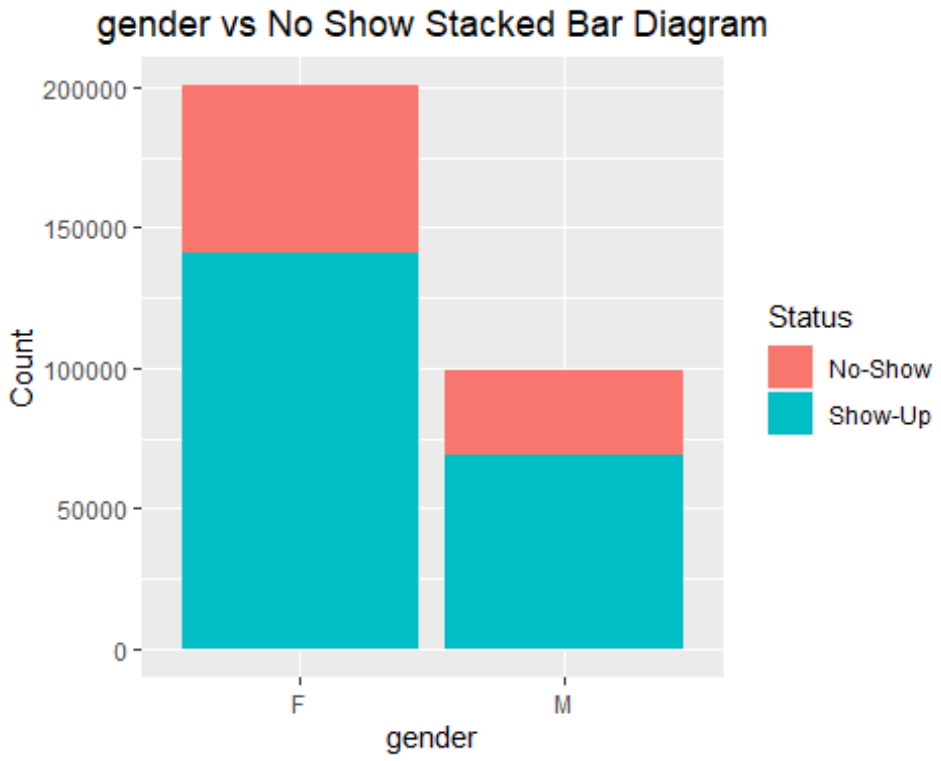

```
Status_table <- table(appointments$sms_reminder)
Status_table

##
##      0      1      2
## 128546 170649    799

options(scipen = 999)
ggplot(appointments, aes(x=sms_reminder, fill=sms_reminder)) + geom_bar()
```




```
g_gender_1 <- ggplot(appointments, aes(x=gender, fill=gender)) + geom_bar(pos
ition="dodge")
g_gender_2 <- ggplot(appointments, aes(x=gender, fill=Status)) + geom_bar(pos
ition="fill")
grid.arrange(g_gender_1, g_gender_2,ncol=2, top='gender distribution')
```




```
ggplot(appointments)+
  geom_bar(aes(x = gender, fill = Status))+
  ggtitle("gender vs No Show Stacked Bar Diagram")+
  theme(plot.title = element_text(hjust = 0.5))+
  ylab("Count")+
  xlab("gender")
```



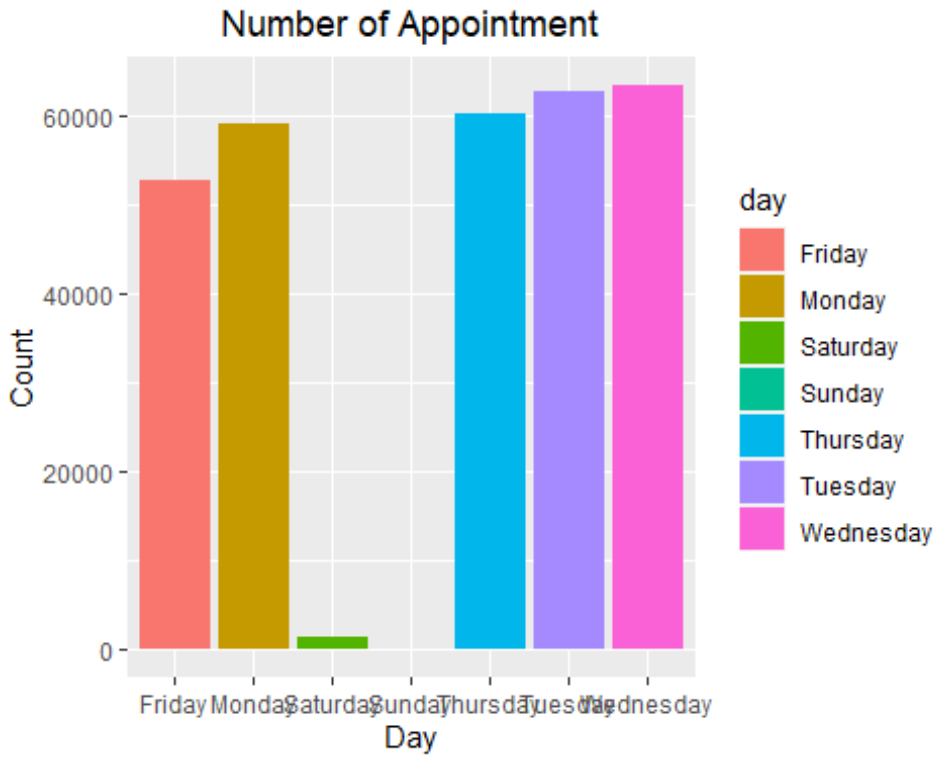

```
tab_gender <- table(appointments$gender, appointments$Status)
addmargins(tab_gender)
```

```
##
##       No-Show Show-Up    Sum
##   F     59889  140610 200499
##   M     30840   68655  99495
##   Sum   90729  209265 299994
```

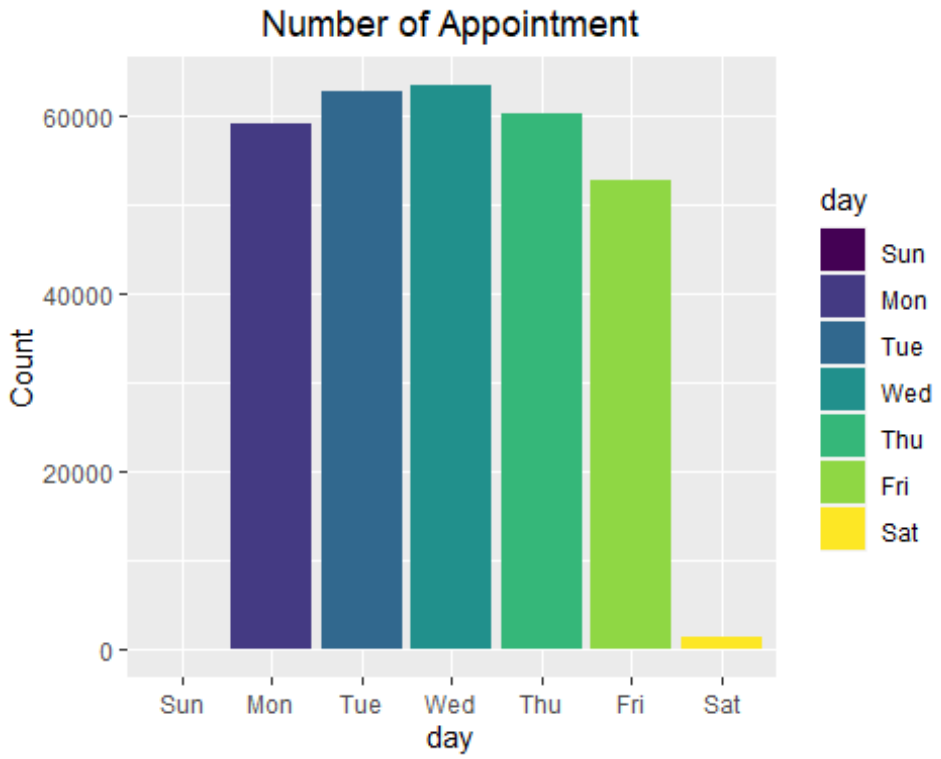
```
prop.table(tab_gender,2)
```

```
##
##       No-Show   Show-Up
##   F 0.6600866 0.6719232
##   M 0.3399134 0.3280768
```

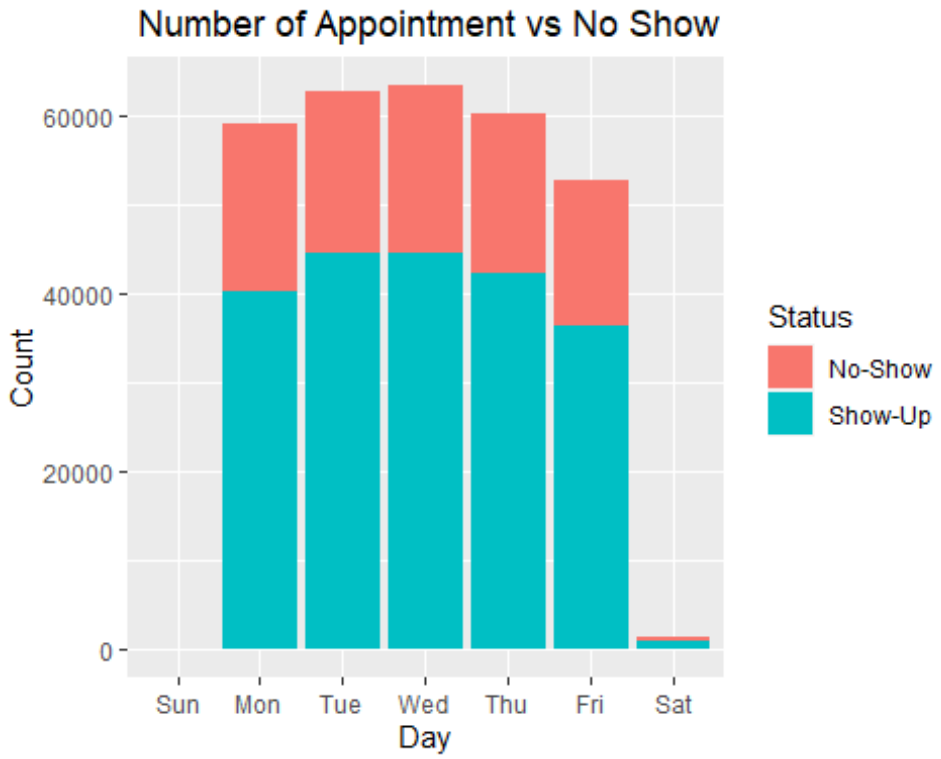
```
appointments$day <- weekdays(as.Date(appointments$appointment_Date))

ggplot(appointments)+geom_bar(aes(day, fill = day))+
  ggtitle("Number of Appointment")+
  ylab('Count')+
  xlab('Day')+
  theme(plot.title = element_text(hjust = 0.5))
```



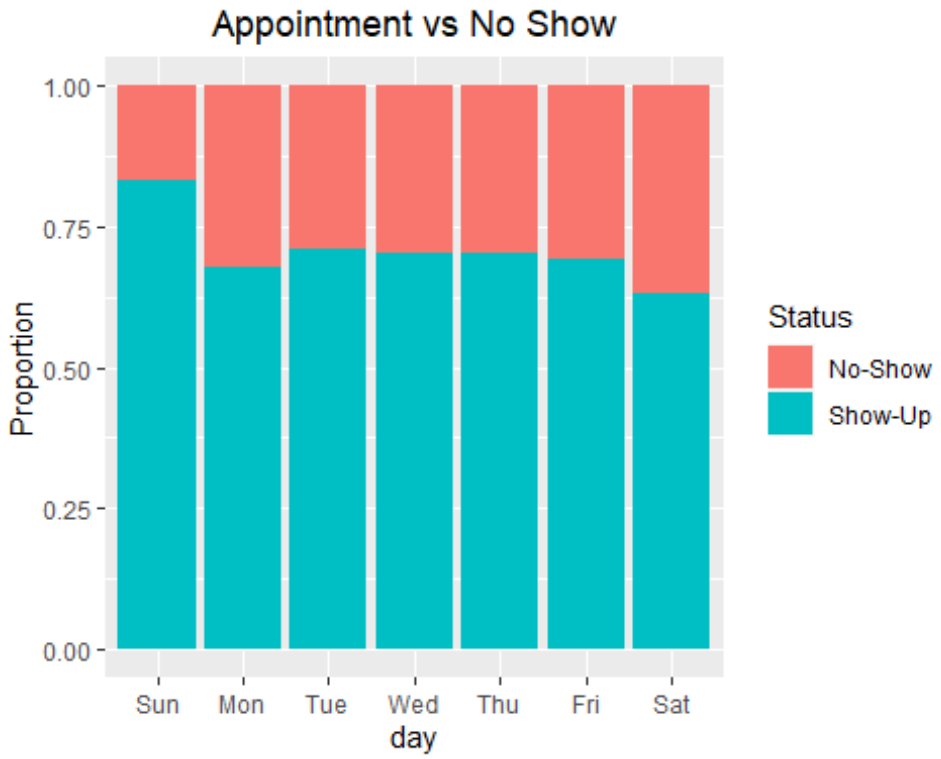

```r
# make days column, with lebel true
appointments$date <- as.Date(appointments$appointment_Date)
appointments$day <- wday(appointments$date, label=TRUE)

ggplot(appointments)+geom_bar(aes(day, fill = day))+
  ggtitle("Number of Appointment")+
  ylab('Count')+
  xlab('day')+
  theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(appointments)+geom_bar(aes(day, fill = Status))+
  ggtitle(" Number of Appointment vs No Show")+
  ylab('Count')+
  xlab('Day')+
  theme(plot.title = element_text(hjust = 0.5))
```

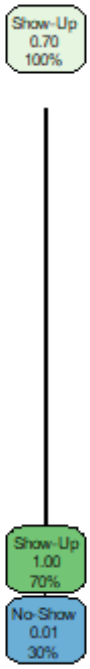



```
ggplot(appointments)+geom_bar(aes(day, fill = Status), position = position_fi
ll())+
  ggtitle("Appointment vs No Show")+
  ylab('Proportion')+
  xlab('day')+
  theme(plot.title = element_text(hjust = 0.5))
```




```
table(appointments$day, appointments$Status)

##
##       No-Show Show-Up
##   Sun       1       5
##   Mon   19053   40244
##   Tue   18148   44626
##   Wed   18919   44576
##   Thu   17850   42410
##   Fri   16245   36524
##   Sat     513     880

day <- table(appointments$Status, appointments$day)
addmargins(day)

##
##               Sun    Mon    Tue    Wed    Thu    Fri    Sat    Sum
##   No-Show       1  19053  18148  18919  17850  16245    513  90729
##   Show-Up       5  40244  44626  44576  42410  36524    880 209265
##   Sum           6  59297  62774  63495  60260  52769   1393 299994

prop.table(day,2)

##
##                 Sun       Mon       Tue       Wed       Thu       Fri
Sat
##   No-Show 0.1666667 0.3213147 0.2891006 0.2979605 0.2962164 0.3078512 0.36
82699
```

```
##   Show-Up 0.8333333 0.6786853 0.7108994 0.7020395 0.7037836 0.6921488 0.63
17301

# Decision Tree
library(caret)

## Loading required package: lattice

set.seed(1234)
ind <- sample(2, nrow(appointments), replace = T, prob = c(0.5, 0.5))
train <- appointments[ind == 1,]
test <- appointments[ind == 2,]
tree <- rpart(Status ~., data = train)
rpart.plot(tree)
```




```
tree <- rpart(Status ~., data = train,cp=0.07444)

pred <- predict(tree, train, type = 'class')
confusionMatrix(pred, train$Status)

## Confusion Matrix and Statistics
##
##           Reference
## Prediction No-Show Show-Up
##    No-Show   45181     333
##    Show-Up      18  104343
##
##                Accuracy : 0.9977
```

```
##                  95% CI : (0.9974, 0.9979)
##     No Information Rate : 0.6984
##     P-Value [Acc > NIR] : < 0.00000000000000022
##
##                   Kappa : 0.9945
##
##  Mcnemar's Test P-Value : < 0.00000000000000022
##
##             Sensitivity : 0.9996
##             Specificity : 0.9968
##          Pos Pred Value : 0.9927
##          Neg Pred Value : 0.9998
##              Prevalence : 0.3016
##          Detection Rate : 0.3015
##    Detection Prevalence : 0.3037
##       Balanced Accuracy : 0.9982
##
##        'Positive' Class : No-Show
##
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.