```r
library(dplyr)
```

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

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

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

```r
library(readr)
library(ggplot2)
library(ggthemes)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(gridExtra)
```

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

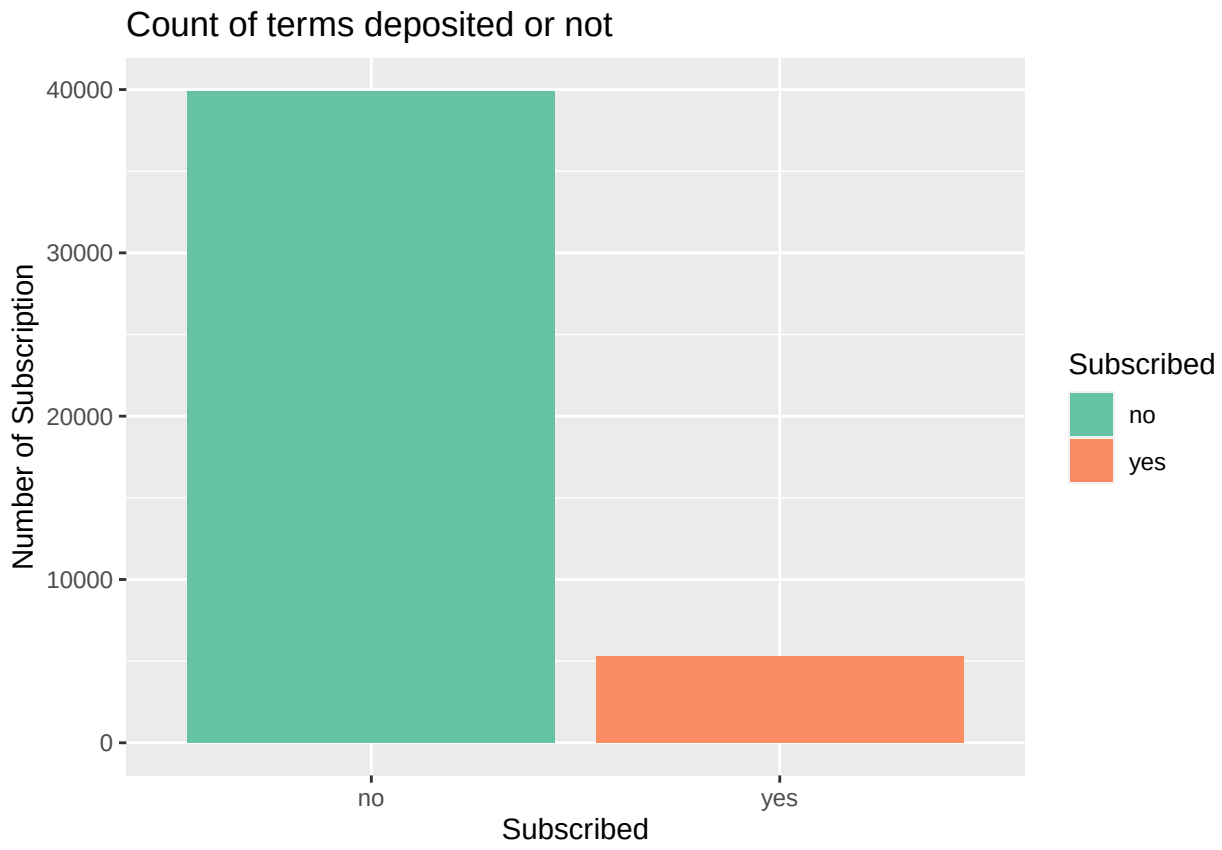
```r
library(readxl)
library(ggcorrplot)
# Change according to your file location, use backslash
bank <- 'C:\\Users\\smeet\\Desktop\\bankmarketing\\data\\bank-full.csv'
bank <-  read_csv2(bank)
```

```
## i Using "',' as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.
## Rows: 45211 Columns: 17-- Column specification --------------------------------------------------------
## Delimiter: ";"
## chr (10): job, marital, education, default, housing, loan, contact, month, p...
## dbl  (7): age, balance, day, duration, campaign, pdays, previous
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

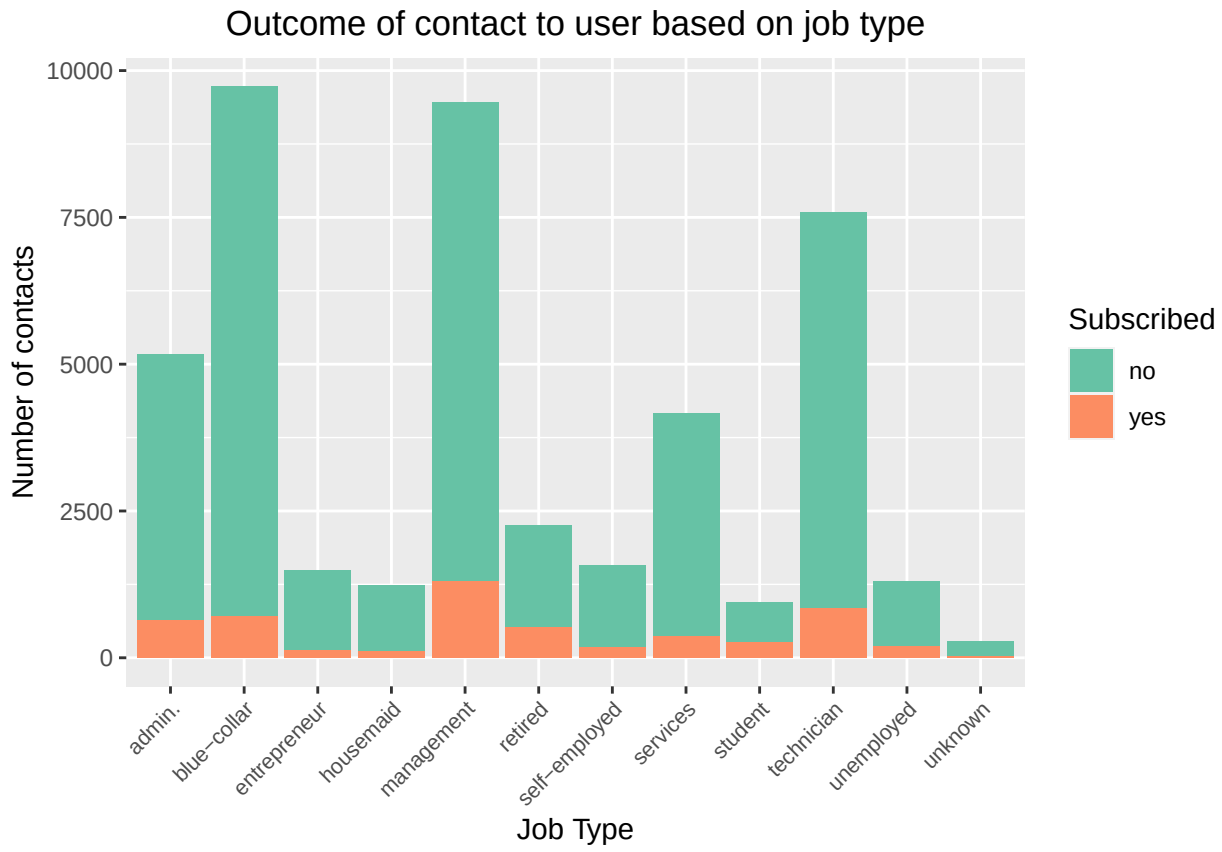
```r
bank <- na.omit(bank)
```

```r
bank <- bank %>%
  mutate(month = factor(month, levels = c("jan","feb","mar","apr","may",'jun','jul','aug','sep','oct','
```

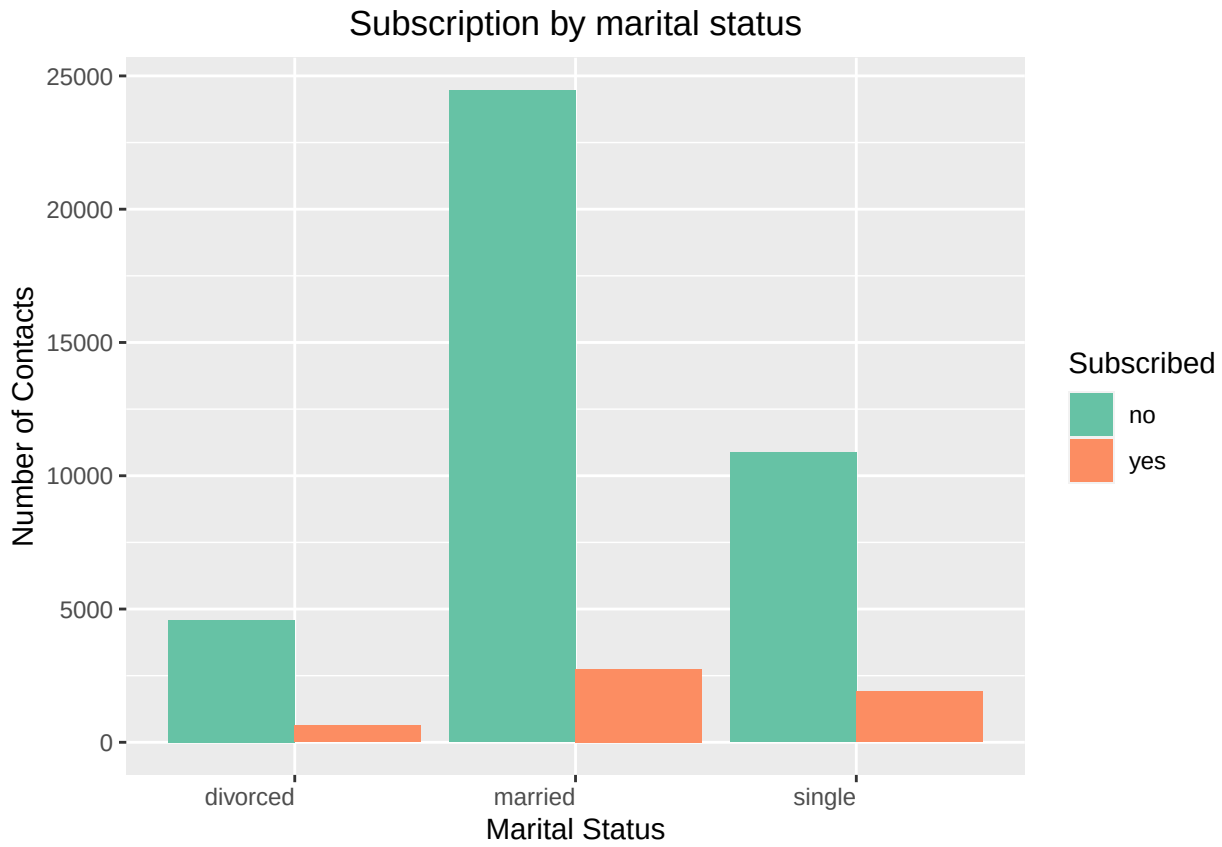
```r
# Far more people did not subscribe to term deposit than they did.
# Class of Outcomes
ggplot(data=bank,mapping= aes(x=y,fill=y)) +
  geom_bar()+ labs(x='Subscribed',fill='Subscribed',y='Number of Subscription',
  title='Count of terms deposited or not')+ scale_fill_brewer(palette = 'Set2')
```




```r
# Technicians and management tend to subscribe more as compared to entrepreneurs and
# house-maids.
ggplot(data=bank,mapping= aes(x = job,fill=y)) +
  geom_bar()+
  labs(title="Outcome of contact to user based on job type",fill='Subscribed',y='Number of contacts',x=
  theme(axis.text.x = element_text(size=8,angle = 45,hjust=1.0))+
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```

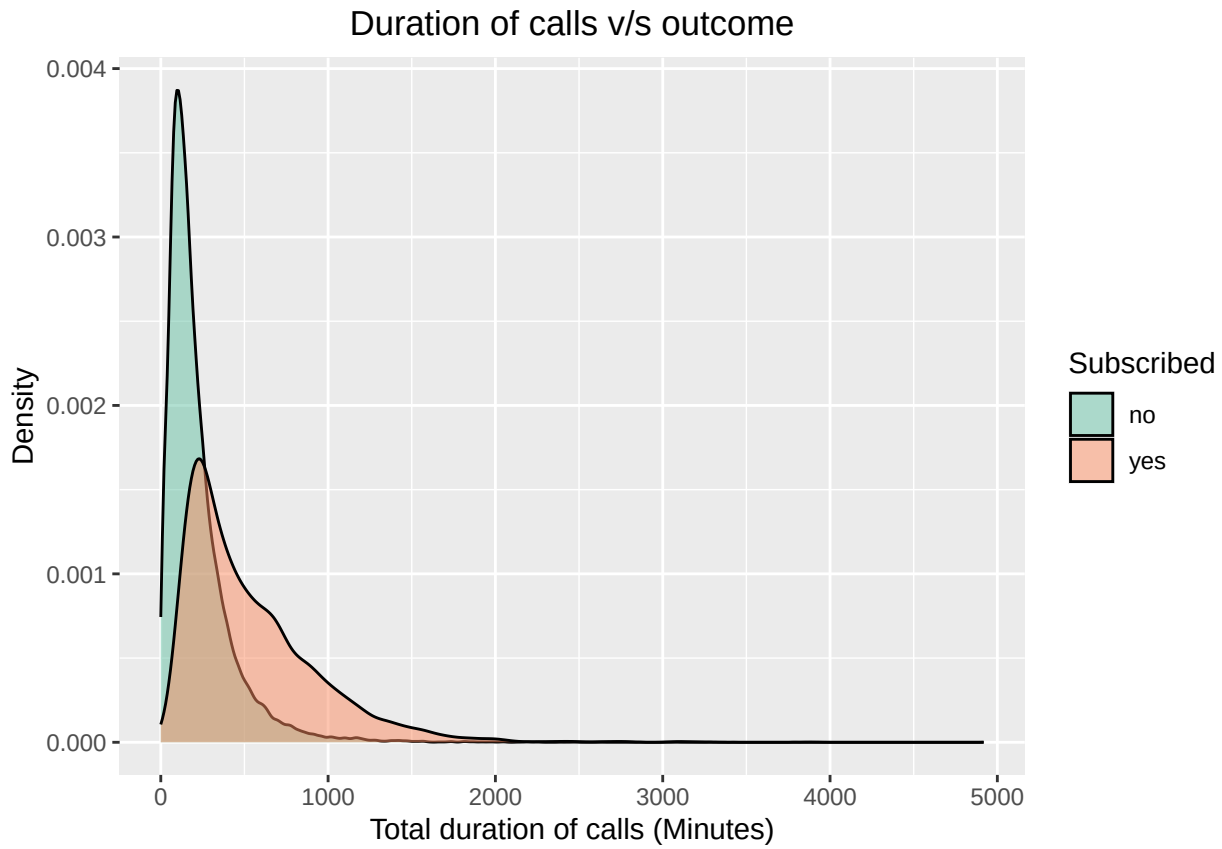



```
# Singles tend to subscribe more to term deposit compared to married and divorced.


# Deposit by Marital-type
ggplot(data=bank,mapping= aes(x=marital,fill=y)) +
  geom_bar(position='dodge')+
  labs(x ='Marital Status',y='Number of Contacts',fill='Subscribed',title='Subscription by marital statu
  theme(axis.text.x = element_text(hjust=1.0))+
    scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```

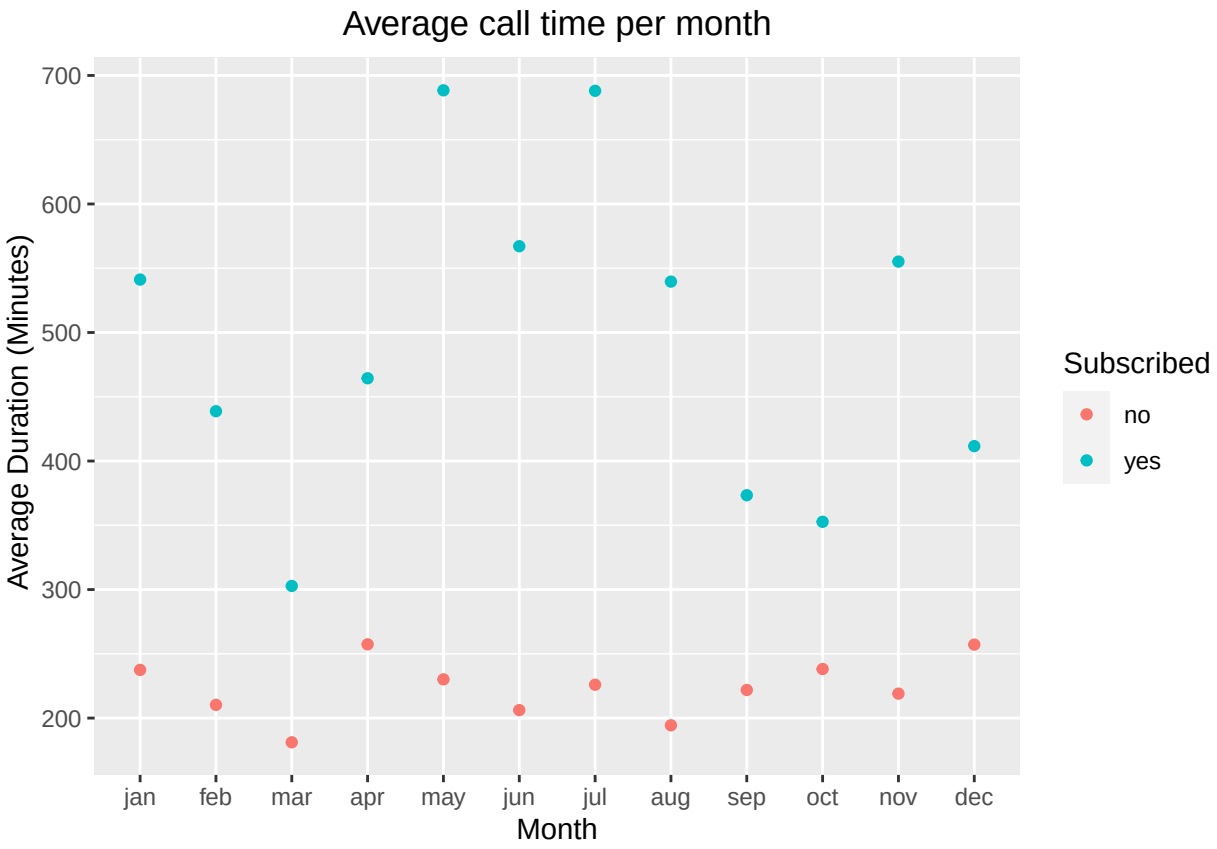



```
# March,April and September,October got more term deposit success compared to rejection

# Deposit based on month
ggplot(bank,aes(x=y,fill=month)) +
  geom_bar(position='fill') +
  labs(title= 'Subscription to term deposit by months',y='Proportion',x='Subscribed',fill='Months')+
   theme(plot.title = element_text(hjust = 0.5))
```

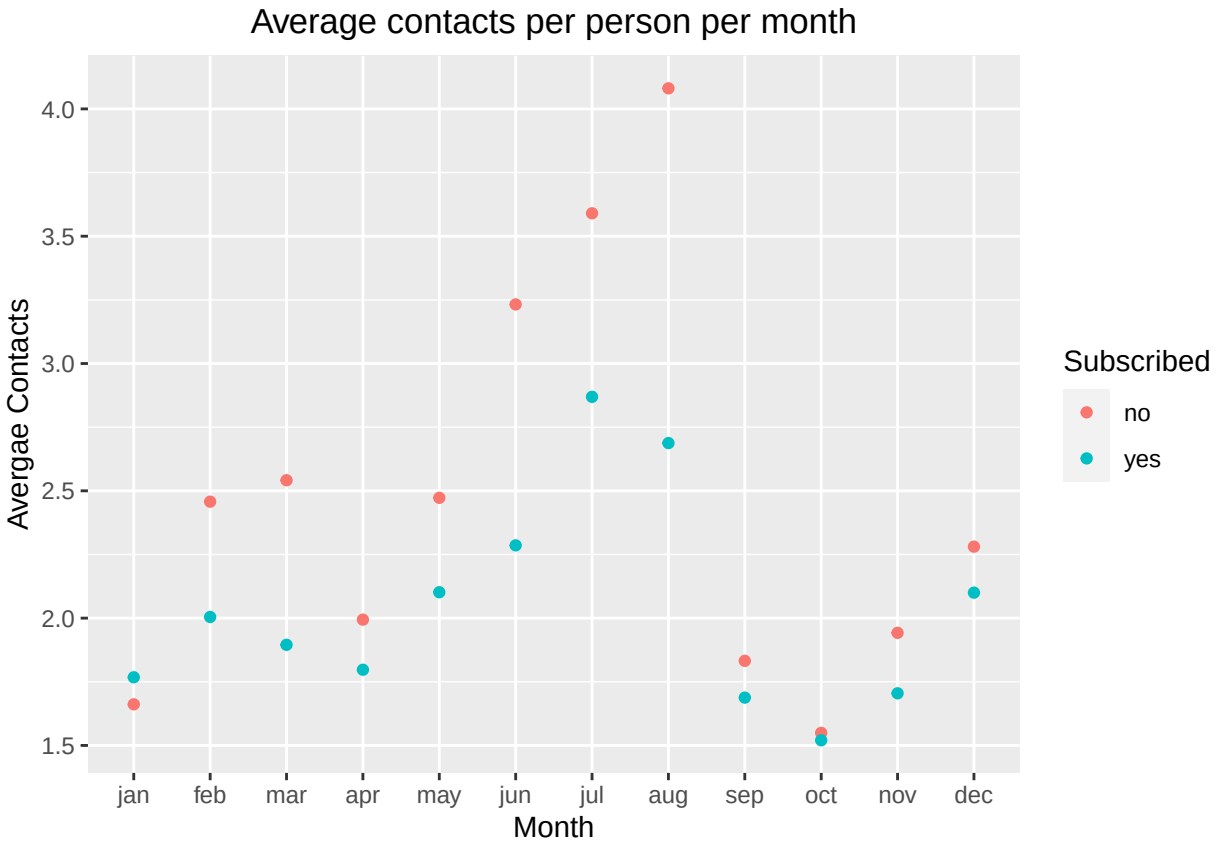

```
# Longer Call duration can be associated with Subscription
# Deposit based on duration
ggplot(bank, aes(duration, fill = y)) +
  geom_density(alpha = 0.5)+ labs(x='Total duration of calls (Minutes)',y='Density',fill='Subscribed',
                                  title= 'Duration of calls v/s outcome')+
    scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```



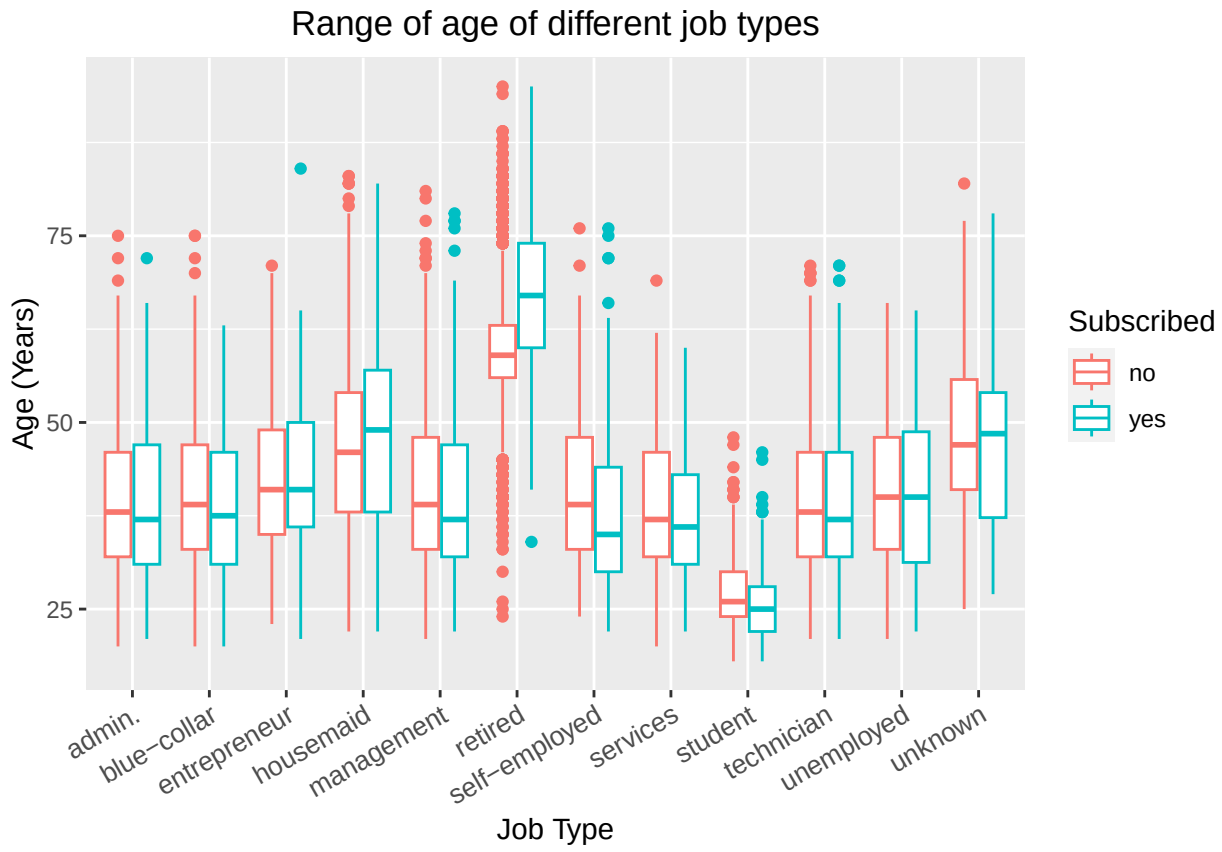

```
# Average duration of calls by month
ggplot(data =bank,
       mapping = aes(x= month,
                     y= duration,color=y)) +
  geom_point(stat='summary',fun=mean) +
  labs(color='Subscribed',x= 'Month',y=' Average Duration (Minutes)',
       title='Average call time per month')+
     scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```



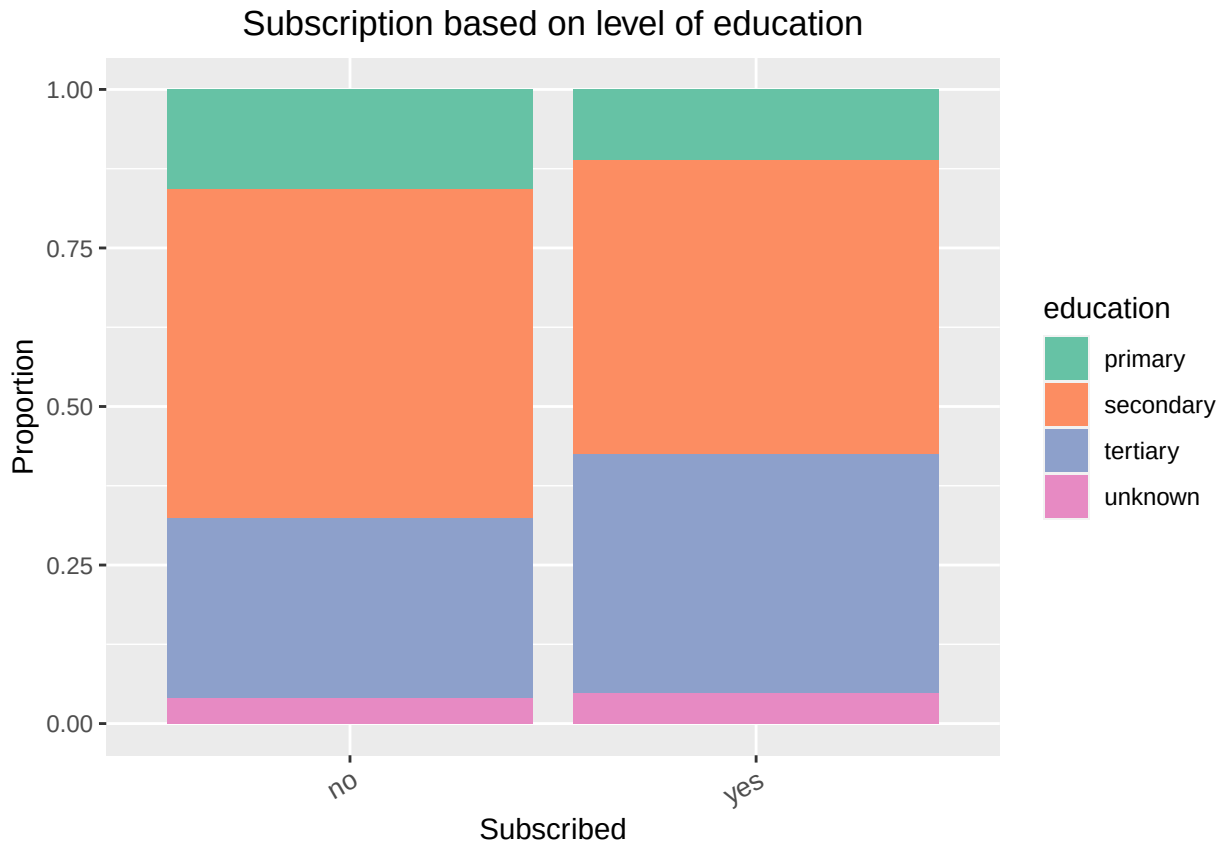

```
ggplot(data =bank,
       mapping = aes(x= month,
                     y= campaign,color=y)) +
  geom_point(stat='summary',fun=mean) +
  labs(color='Subscribed',x= 'Month',y='Avergae Contacts',
       title='Average contacts per person per month')+
    scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```




```
# Age of different job types
ggplot(bank,aes(x= job,y= age,color=y))+ geom_boxplot()+
  labs(title='Range of age of different job types',x='Job Type',y='Age (Years)',color='Subscribed')+
  theme(axis.text.x = element_text(size=10,angle =30,hjust=1.0))+
    scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```



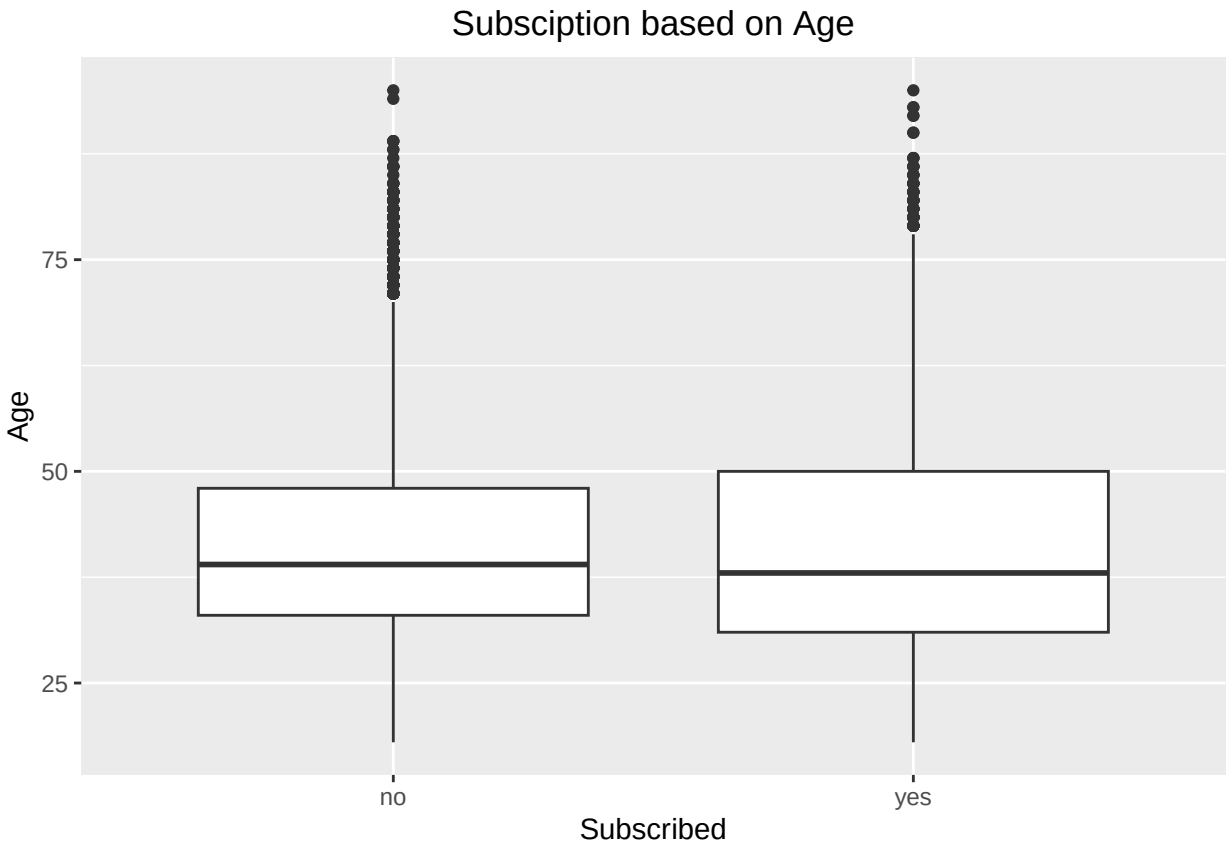

```
# deposit based on month
ggplot(bank ,  aes(x=y,fill=education))+ geom_bar(position='fill')+
  theme(axis.text.x = element_text(size=10,angle =30,hjust=1.0))+ labs(y='Proportion',
                                                                       x = 'Subscribed',
                                                                       title='Subscription based on leve
    scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```



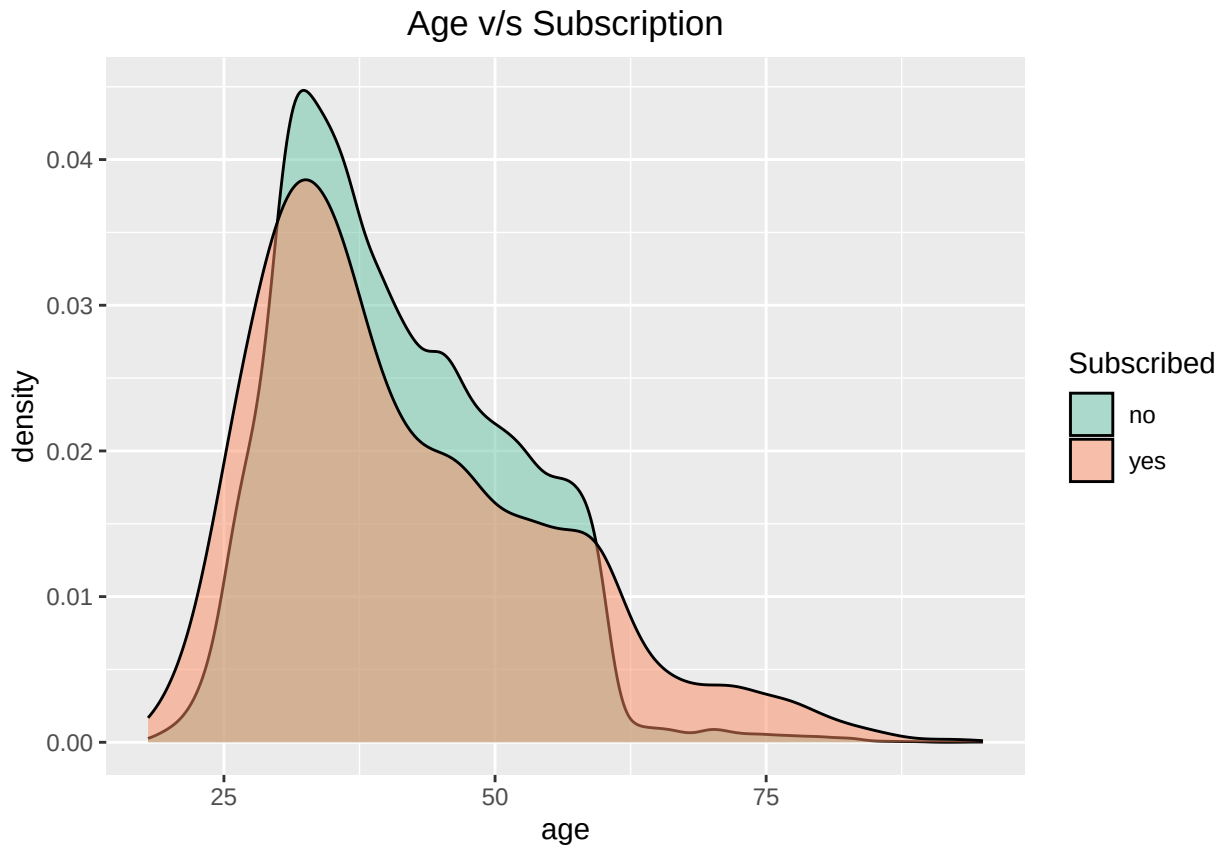

```
bank
```

```
## # A tibble: 45,211 x 17
##      age job       marital education default balance housing loan  contact   day
##    <dbl> <chr>     <chr>   <chr>     <chr>     <dbl> <chr>   <chr> <chr>   <dbl>
##  1    58 manageme~ married tertiary  no         2143 yes     no    unknown     5
##  2    44 technici~ single  secondary no           29 yes     no    unknown     5
##  3    33 entrepre~ married secondary no            2 yes     yes   unknown     5
##  4    47 blue-col~ married unknown   no         1506 yes     no    unknown     5
##  5    33 unknown   single  unknown   no            1 no      no    unknown     5
##  6    35 manageme~ married tertiary  no          231 yes     no    unknown     5
##  7    28 manageme~ single  tertiary  no          447 yes     yes   unknown     5
##  8    42 entrepre~ divorc~ tertiary  yes           2 yes     no    unknown     5
##  9    58 retired   married primary   no          121 yes     no    unknown     5
## 10    43 technici~ single  secondary no          593 yes     no    unknown     5
## # i 45,201 more rows
## # i 7 more variables: month <fct>, duration <dbl>, campaign <dbl>, pdays <dbl>,
## #   previous <dbl>, poutcome <chr>, y <chr>
```

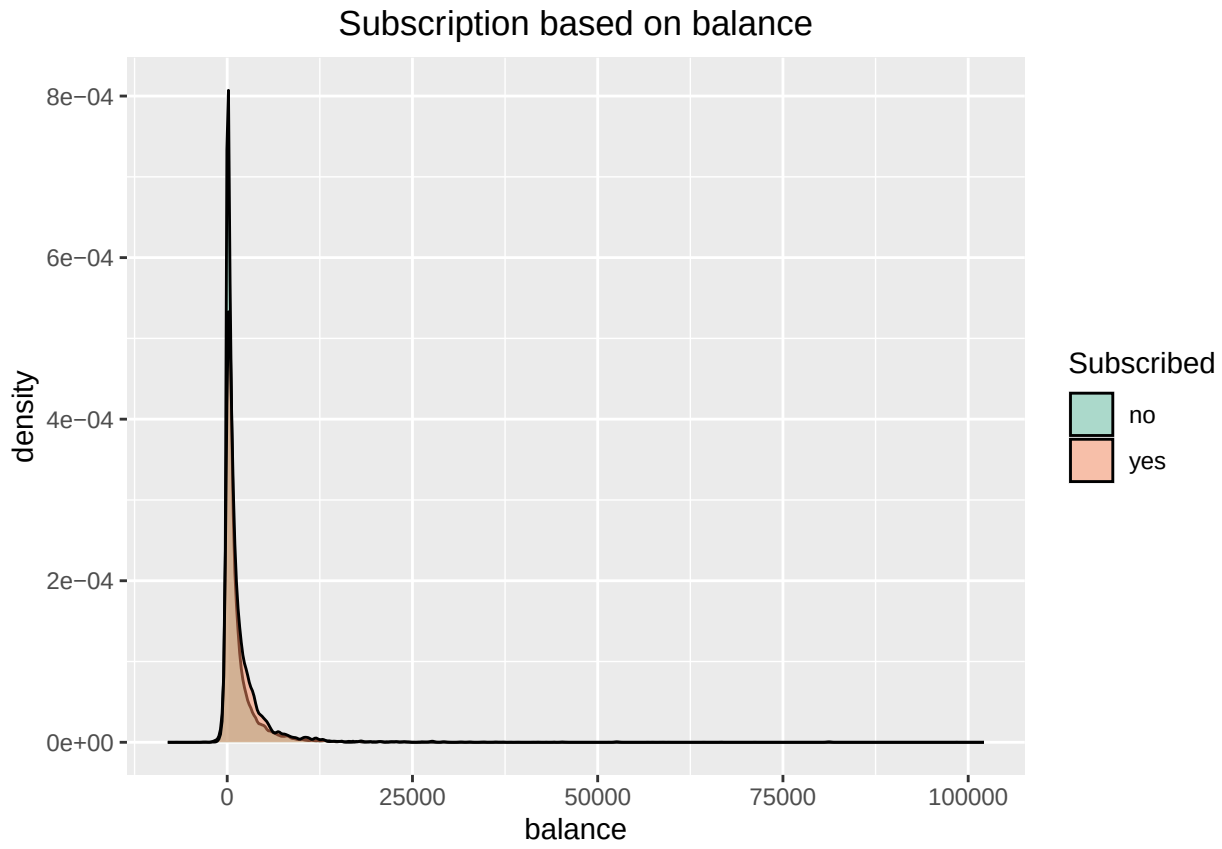
```
# deposit based on age
ggplot(bank,aes(x=y,y=age)) + geom_boxplot() + labs(x='Subscribed',y='Age',
                                                    title='Subsciption based on Age')+
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```



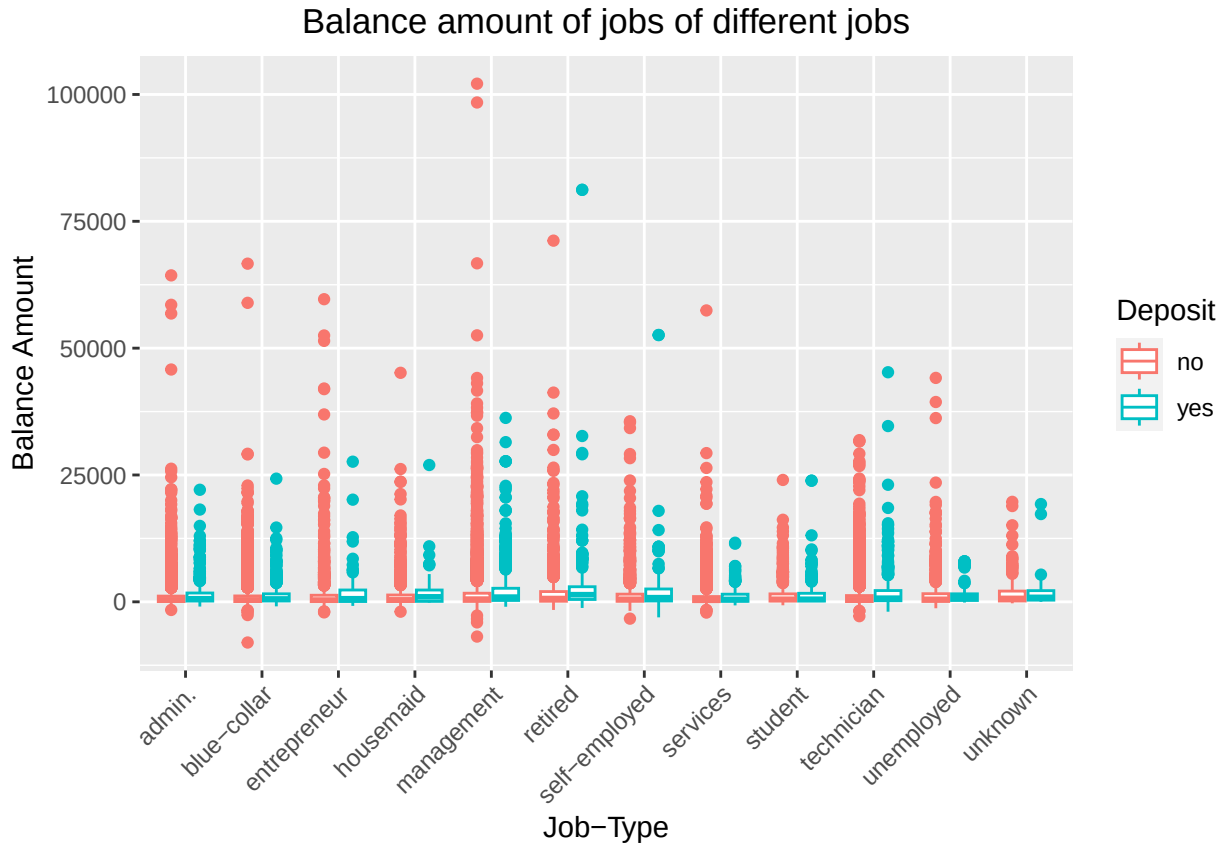

```
# Deposit based on age
ggplot(bank,aes(age,fill=y))+ labs(fill='Subscribed',title='Age v/s Subscription')+ geom_density(alpha=
```



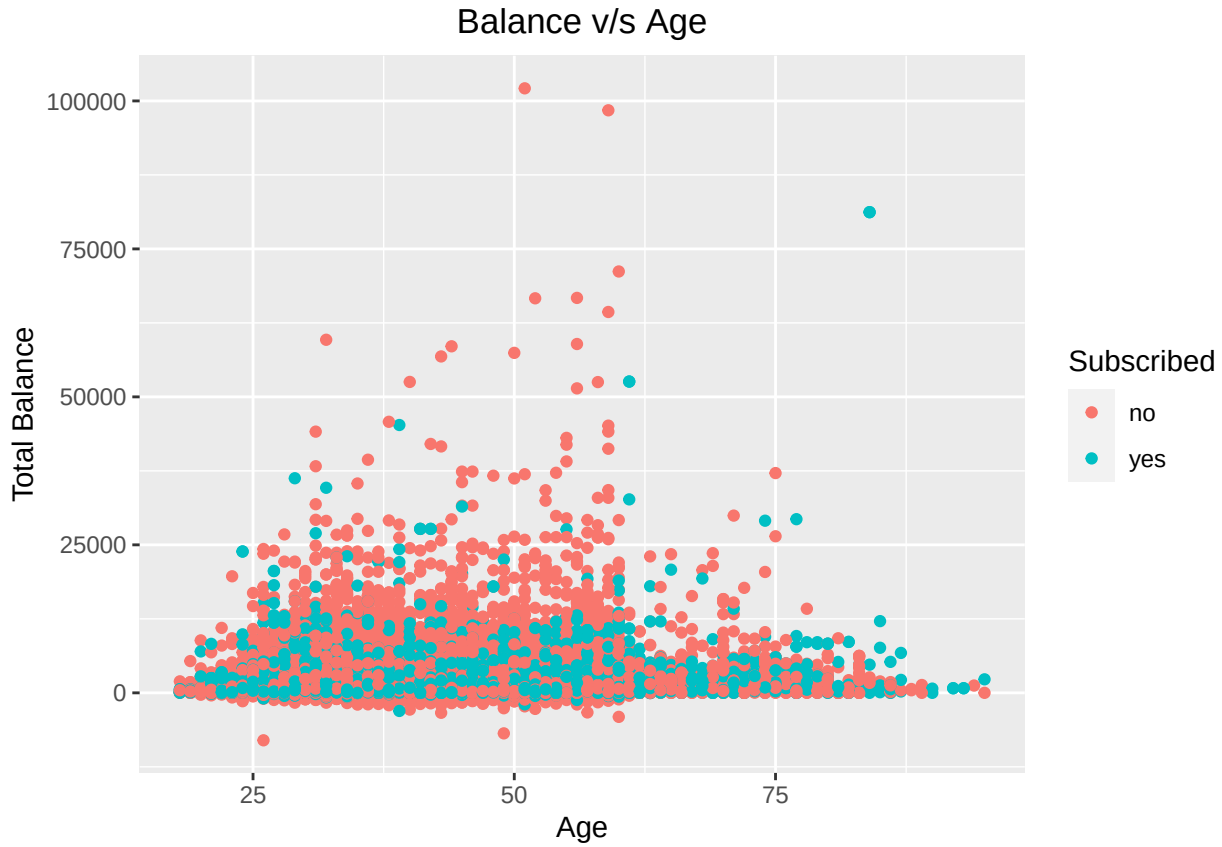

```
# Subscription and age
ggplot(bank, aes(balance,fill=y)) + labs(fill='Subscribed',title='Subscription based on balance')+ geom_
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```




```
# Balance of different types of jobs
ggplot(bank,aes(x=job,y=balance,color=y))+ geom_boxplot()+
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5),
                                             axis.text.x =
                                               element_text(angle=45,hjust=1.0))+ labs(title='Balance ar
      ,color='Deposit')
```

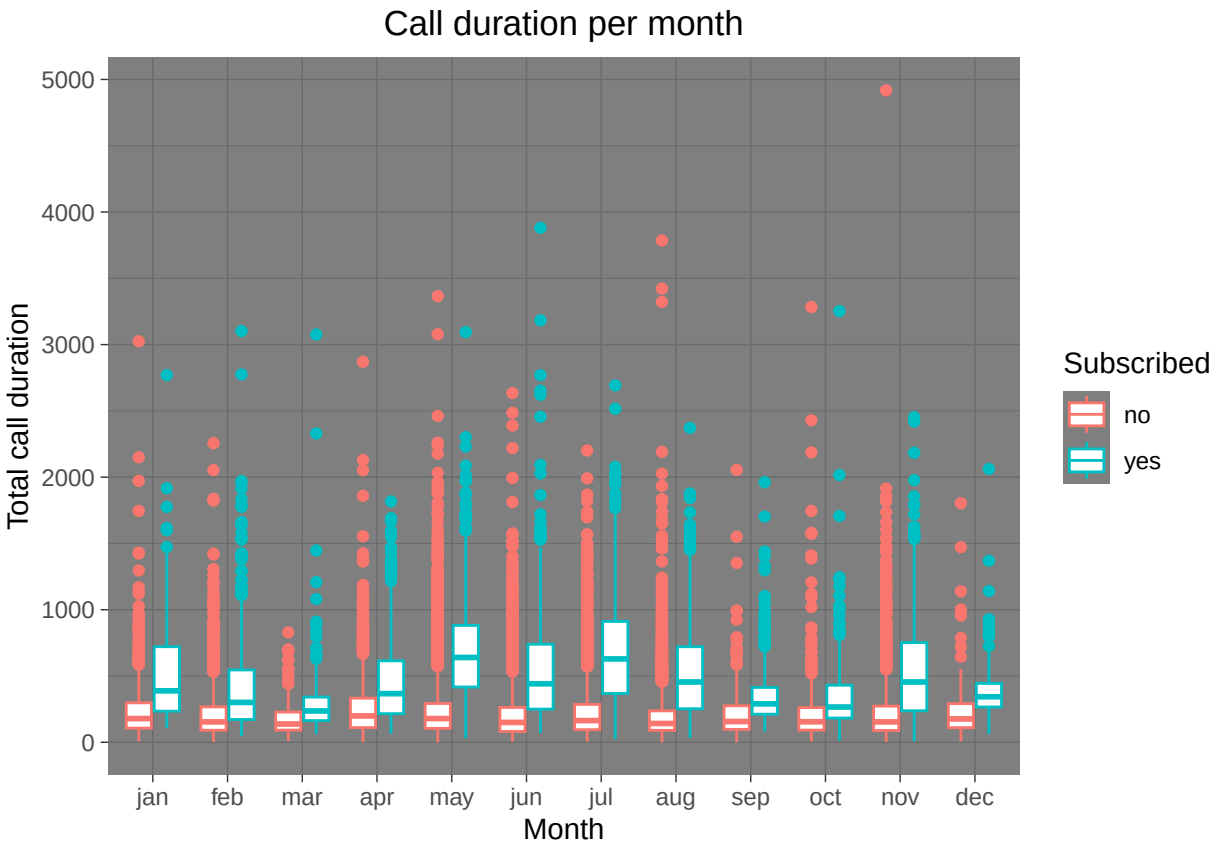



```
# how age changes with balance
ggplot(data=bank,
       mapping=aes(x=age,
                   y=balance,
                   color=y))+
  geom_point()+
  labs(x='Age',y='Total Balance',title= 'Balance v/s Age',color= 'Subscribed')+
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```



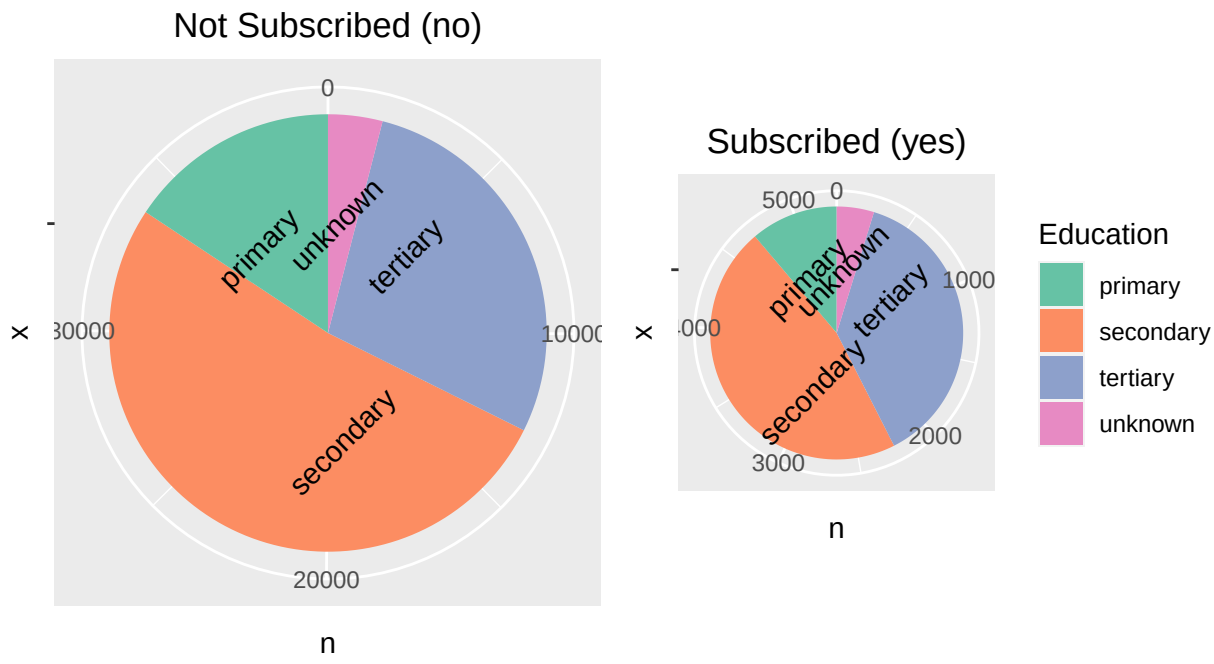

```
# Count of contacts per month during this campaign by month
ggplot(data=bank,
       mapping=aes(x=month,
                   y=campaign,
                   color=y))+
  geom_boxplot()+labs(title='Contacts per month',color='Subscribed')+
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```


```
# call duration per month boxplot
ggplot(data=bank,
       mapping=aes(x=month,
                   y=duration,color=y))+
  geom_boxplot() + theme_dark() + labs(title='Call duration per month',y='Total call duration',x='Month
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))
```




```
bank_yes <-  bank[bank$y=='yes',]

bank_no <- bank[bank$y=='no',]


education_yes <- bank_yes %>%
  count(education, sort = TRUE) %>%
  mutate(edcuation = reorder(education, n)) %>%
  ggplot(aes(x="", y= n,fill=education)) +
  geom_bar(stat="identity") +
  geom_text(aes(label = education,angle=45),
            position = position_stack(vjust = 0.3),
            show.legend = FALSE)+
  coord_polar("y",start=0)+ labs(title='Subscribed (yes)',fill='Education')+
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))

education_no <- bank_no %>%
  count(education, sort = TRUE) %>%
  mutate(edcuation = reorder(education, n)) %>%
  ggplot(aes(x="", y= n,fill=education)) +
  geom_bar(stat="identity") +
  geom_text(aes(label = education,angle=45),
            position = position_stack(vjust = 0.3),
            show.legend = FALSE)+
  coord_polar("y",start=0)+ labs(title='Not Subscribed (no)')+
  theme(legend.position = 'none')+
```

```
  scale_fill_brewer(palette = 'Set2')+ theme(plot.title = element_text(hjust = 0.5))


grid.arrange(education_no,education_yes, ncol= 2)
```




```
month_yes <- bank_yes %>%
  count(month, sort = TRUE) %>%
  mutate(month = reorder(month, n)) %>%
  ggplot(aes(x="", y= n,fill=month)) +
  geom_bar(stat="identity") +
  geom_text(aes(label = month,angle=90),
            position = position_stack(vjust = 0.3),
            show.legend = FALSE)+
  coord_polar("y",start=0)+ labs(title='Subscribed("yes")',fill='Education',x='',y='')+ theme(plot.titl


month_no <- bank_no %>%
  count(month, sort = TRUE) %>%
  mutate(month = reorder(month, n)) %>%
  ggplot(aes(x="", y= n,fill=month)) +
  geom_bar(stat="identity") +
  geom_text(aes(label = month,angle=90),
            position = position_stack(vjust = 0.3),
            show.legend = FALSE)+
  coord_polar("y",start=0)+ labs(title='Not Subscribed("no")',x='',y='')+
```

```
  theme(legend.position = 'none')+ theme(plot.title = element_text(hjust = 0.5))


grid.arrange(month_no,month_yes, ncol= 2)
```