Comcast Telecom Consumer Complaint

R Project

```
install.packages("pscl", repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/HP/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)
```

```
## package 'pscl' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\HP\AppData\Local\Temp\RtmpGYayzu\downloaded_packages
```

```
library(stringi)
library(lubridate)
```

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

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

```
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
```

```
library(ggplot2)
library(ggpubr)

# Loading Dataset:
comcast_data<- read.csv("Comcast_telecom_complaints_data.csv",header = TRUE)

head(comcast_data)
```

```
##   Ticket..                                                Customer.Complaint
## 1   250635                                    Comcast Cable Internet Speeds
## 2   223441                      Payment disappear - service got disconnected
## 3   242732                                                 Speed and Service
## 4   277946 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.
## 5   307175                          Comcast not working and no service to boot
## 6   338519          ISP Charging for arbitrary data limits with overage fees
##         Date        Time       Received.Via     City    State Zip.code Status
## 1 22-04-2015  3:53:50 PM Customer Care Call Abingdon Maryland    21009 Closed
## 2   4/8/2015 10:22:56 AM           Internet  Acworth  Georgia    30102 Closed
## 3 18-04-2015  9:55:47 AM           Internet  Acworth  Georgia    30101 Closed
## 4   5/7/2015 11:59:35 AM           Internet  Acworth  Georgia    30101   Open
## 5 26-05-2015  1:25:26 PM           Internet  Acworth  Georgia    30101 Solved
## 6  6/12/2015  9:59:40 PM           Internet  Acworth  Georgia    30101 Solved
##   Filing.on.Behalf.of.Someone
## 1                          No
## 2                          No
## 3                         Yes
## 4                         Yes
## 5                          No
## 6                          No
```

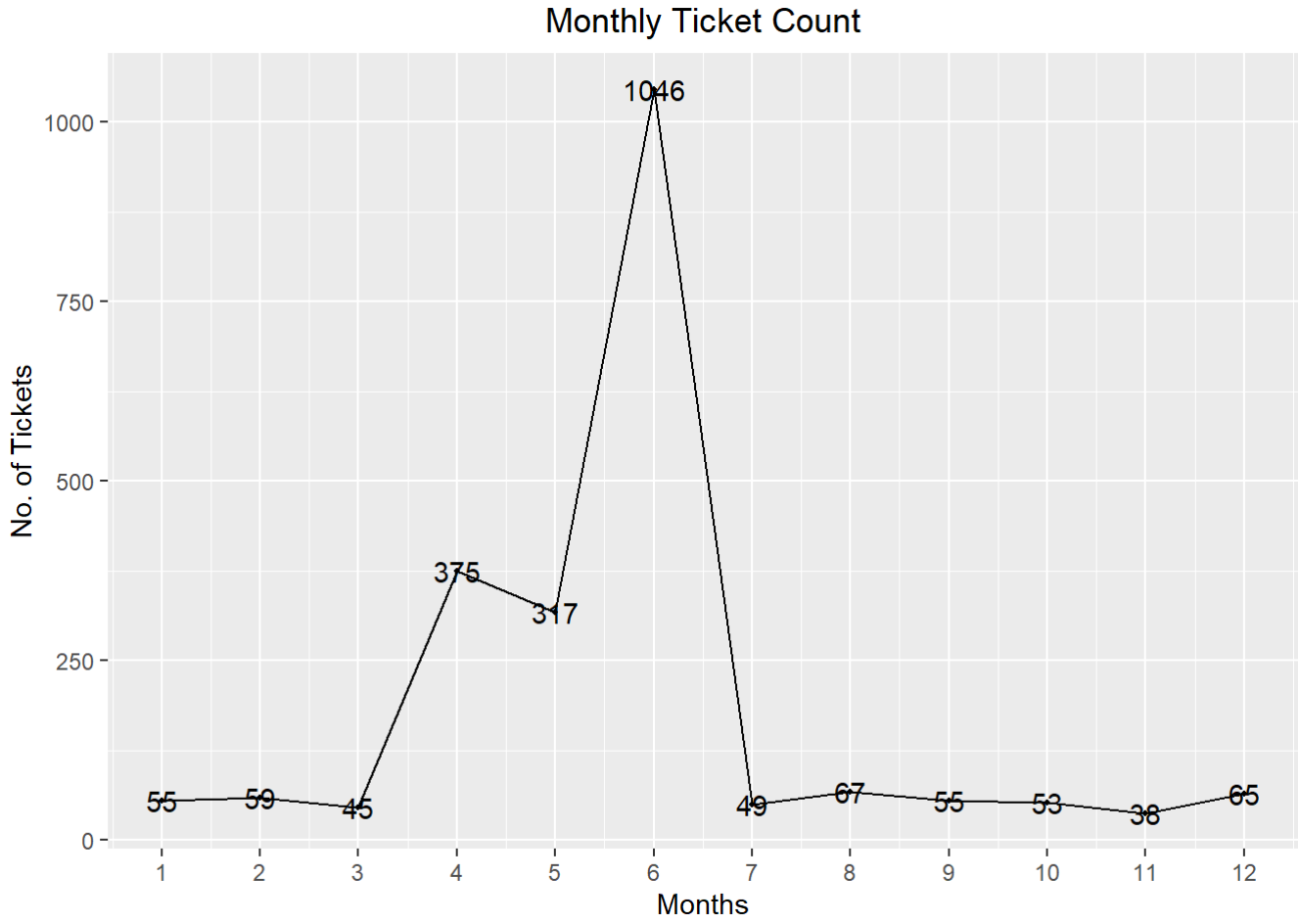
```
#Manipulating column names
names(comcast_data)<- stri_replace_all(regex =  "\\.",replacement = "",str =names(comcast_dat
a))
head(comcast_data)
```

```
##   Ticket                                                 CustomerComplaint
## 1 250635                                      Comcast Cable Internet Speeds
## 2 223441                        Payment disappear - service got disconnected
## 3 242732                                                   Speed and Service
## 4 277946 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.
## 5 307175                            Comcast not working and no service to boot
## 6 338519          ISP Charging for arbitrary data limits with overage fees
##         Date        Time        ReceivedVia     City    State Zipcode Status
## 1 22-04-2015  3:53:50 PM Customer Care Call Abingdon Maryland   21009 Closed
## 2   4/8/2015 10:22:56 AM           Internet  Acworth  Georgia   30102 Closed
## 3 18-04-2015  9:55:47 AM           Internet  Acworth  Georgia   30101 Closed
## 4   5/7/2015 11:59:35 AM           Internet  Acworth  Georgia   30101   Open
## 5 26-05-2015  1:25:26 PM           Internet  Acworth  Georgia   30101 Solved
## 6  6/12/2015  9:59:40 PM           Internet  Acworth  Georgia   30101 Solved
##   FilingonBehalfofSomeone
## 1                      No
## 2                      No
## 3                     Yes
## 4                     Yes
## 5                      No
## 6                      No
```

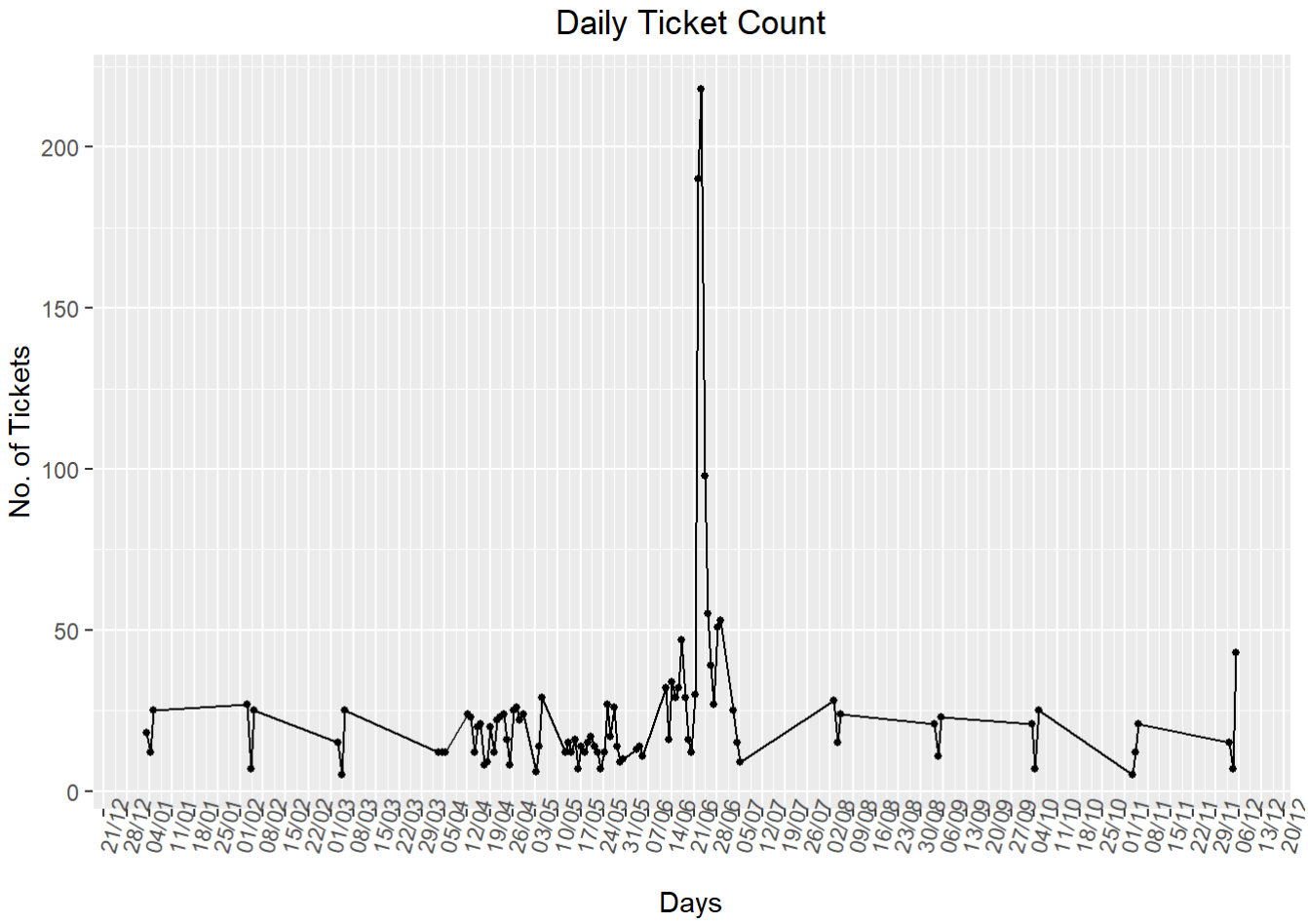
```
# Now data is loaded into R, now its available to process further. .   Finding NAs in Dataset
na_vector <- is.na(comcast_data)
length(na_vector[na_vector==T])
```

```
## [1] 0
```

```
# This shows that there is no missing values in dataset,so now data is tidy and available to
 process further or do EDA based on requirment. . Processing Date.
comcast_data$Date<- dmy(comcast_data$Date)

# Extracting Monthly and Daily Ticket Count.
monthly_count<- summarise(group_by(comcast_data,Month =as.integer(month(Date))),Count = n())
daily_count<- summarise(group_by(comcast_data,Date),Count =n())
monthly_count<-arrange(monthly_count,Month)

# Comparing Monthly and Daily Complaints
ggplot(data = monthly_count,aes(Month,Count,label = Count))+
  geom_line()+
  geom_point(size = 0.8)+
  geom_text()+        scale_x_continuous(breaks = monthly_count$Month)+
  labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")+
  theme(plot.title = element_text(hjust = 0.5))
```




```
# As we can see that in the month of April,May the tickets are increses but in the month  of
 June it increases drastically, so there might be some reason for which they received high am
ount of tickets.

ggplot(data = daily_count,aes(as.POSIXct(Date),Count))+
  geom_line()+
  geom_point(size = 1)+
  scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")+
  labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")+
  theme(axis.text.x = element_text(angle = 75),
        plot.title = element_text(hjust = 0.5))
```



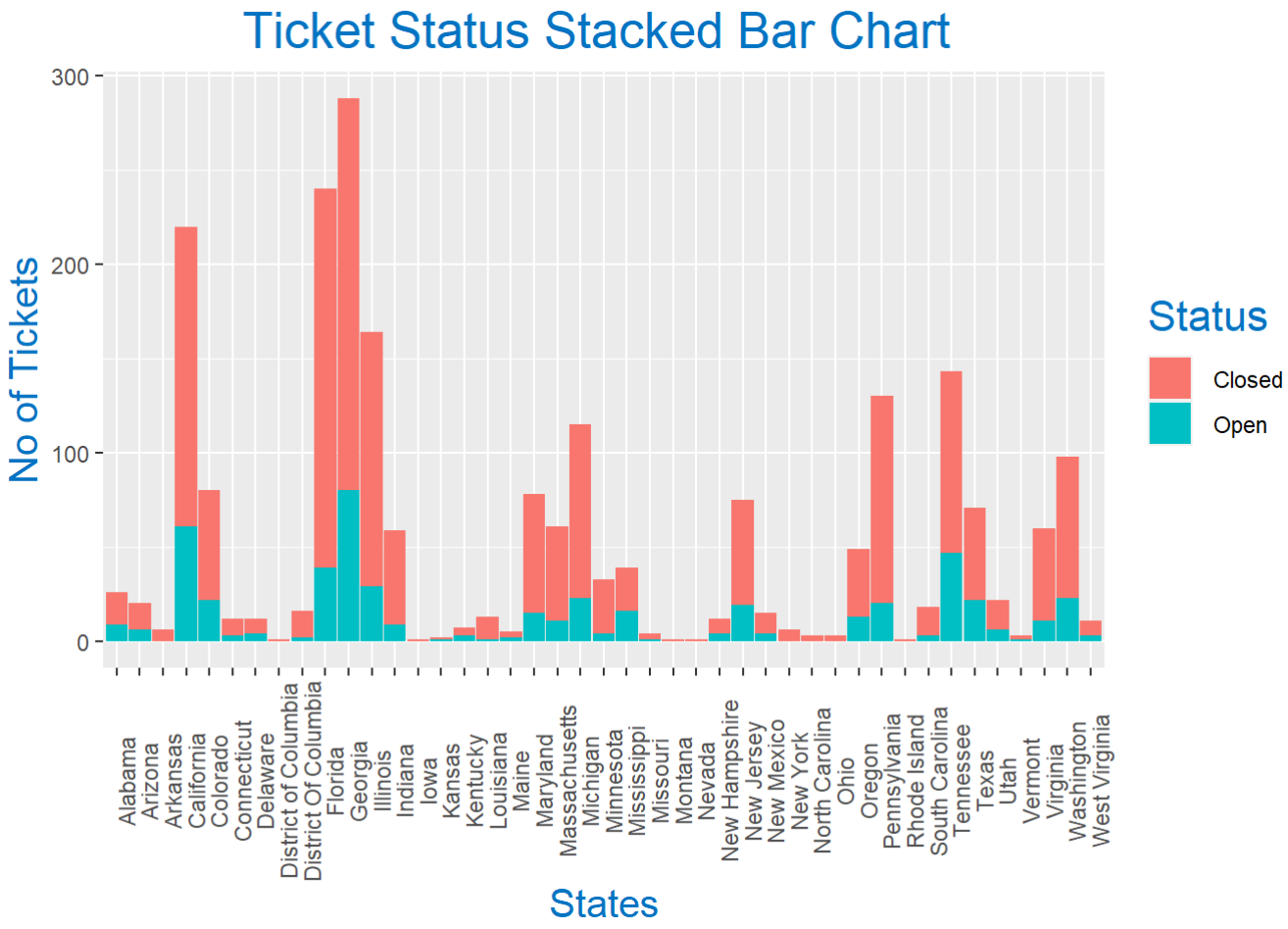

```
# And with the help of above daily chart of tickets we can observe that in second half of
# June month we received more tickets with respect to normal days
# Complaint Type Processing
network_tickets<- contains(comcast_data$CustomerComplaint,match = 'network',ignore.case = T)
internet_tickets<- contains(comcast_data$CustomerComplaint,match = 'internet',ignore.case = T
)
billing_tickets<- contains(comcast_data$CustomerComplaint,match = 'bill',ignore.case = T)
email_tickets<- contains(comcast_data$CustomerComplaint,match = 'email',ignore.case = T)
charges_ticket<- contains(comcast_data$CustomerComplaint,match = 'charge',ignore.case = T)
comcast_data$ComplaintType[internet_tickets]<- "Internet"
comcast_data$ComplaintType[network_tickets]<- "Network"
comcast_data$ComplaintType[billing_tickets]<- "Billing"
comcast_data$ComplaintType[email_tickets]<- "Email"
comcast_data$ComplaintType[charges_ticket]<- "Charges"
comcast_data$ComplaintType[-c(internet_tickets,network_tickets,billing_tickets,charges_ticke
t,email_tickets)]<- "Others"
table(comcast_data$ComplaintType)
```

```
##
##  Billing  Charges    Email Internet  Network   Others
##      363      139       16      472        1     1233
```

```
# As we can observe that there are some complaints from different-different categories and we combined them into one, i.e.- others. So most of the complaints are related to Internet issue. . Creating new Variable ComplaintStatus with values Open and Closed.
open_complaints<- (comcast_data$Status == "Open"| comcast_data$Status =="Pending")
closed_complaints<-(comcast_data$Status == "Closed"| comcast_data$Status =="Solved")
comcast_data$ComplaintStatus[ open_complaints]<-"Open"
comcast_data$ComplaintStatus[closed_complaints]<- "Closed"

# Creating Stacked barchart for complaints based on State and Status.
comcast_data<- group_by(comcast_data,State,ComplaintStatus)
chart_data<- summarise(comcast_data,Count = n())
```

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

```
ggplot(as.data.frame(chart_data), mapping = aes(State,Count))+
  geom_col(aes(fill = ComplaintStatus),width = 0.95)+
  theme(axis.text.x = element_text(angle = 90),
        axis.title.y = element_text(size = 15),
        axis.title.x = element_text(size = 15),
        title = element_text(size = 16,colour = "#0073C2FF"),
        plot.title = element_text(hjust =  0.5))+
  labs(title = "Ticket Status Stacked Bar Chart ",
       x = "States",y = "No of Tickets",
       fill= "Status")
```



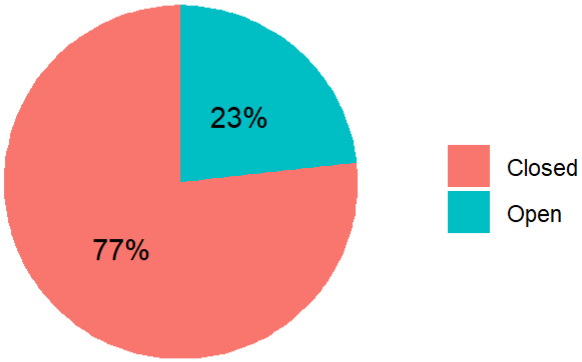
```
# Now it`s clearly shown that the highest number of complaints recorded from the state Georgi
a and the second highest number of complaints recorded from the state Florida. . Finding Stat
e which has Highest number of Unresolved Tickets.
chart_data%>%
  filter(ComplaintStatus == "Open")->
  open_complaints
open_complaints[open_complaints$Count == max(open_complaints$Count),c(1,3)]
```

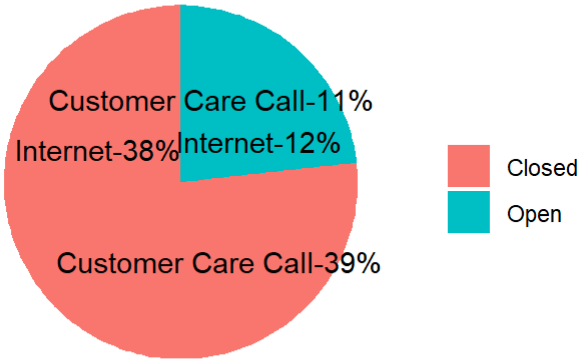
```
## # A tibble: 1 x 2
## # Groups:   State [1]
##   State   Count
##   <chr>   <int>
## 1 Georgia    80
```

```
# As we can observe that State Georgia has maximum number of unresolved tickets and these tic
ket count is 80. . Calculating Resolution Percentage based on Total and Catagory .
resolved_data <- group_by(comcast_data,ComplaintStatus)
total_resloved<- summarise(resolved_data ,percentage =(n()/nrow(resolved_data)))
resolved_data <- group_by(comcast_data,ReceivedVia,ComplaintStatus)
Category_resloved<- summarise(resolved_data ,percentage =(n()/nrow(resolved_data)))
```

```
## `summarise()` has grouped output by 'ReceivedVia'. You can override using the `.groups` ar
gument.
```

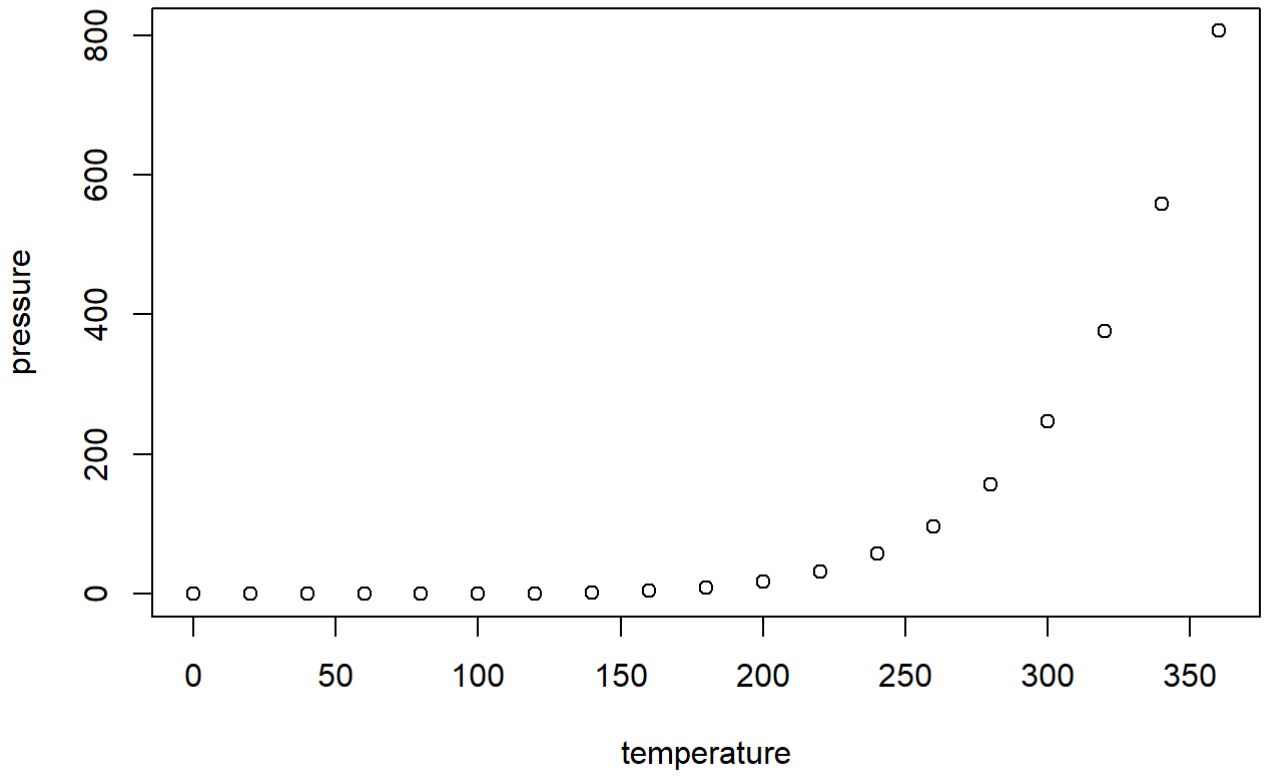
```
# Ploting Pie Chart for Total Resolved Vs Category Resolved
par(mfrow = c(1,2))
total<-ggplot(total_resloved,
              aes(x= "",y =percentage,fill = ComplaintStatus))+
  geom_bar(stat = "identity",width = 1)+
  coord_polar("y",start = 0)+
  geom_text(aes(label = paste0(round(percentage*100),"%")),
            position = position_stack(vjust = 0.5))+
  labs(x = NULL,y = NULL,fill = NULL)+
  theme_classic()+theme(axis.line = element_blank(),
                        axis.text = element_blank(),
                        axis.ticks = element_blank())
# Pie Chart for Category wise Ticket Status
category<-ggplot(Category_resloved,
                 aes(x= "",y =percentage,fill = ComplaintStatus))+
  geom_bar(stat = "identity",width = 1)+
  coord_polar("y",start = 0)+
  geom_text(aes(label = paste0(ReceivedVia,"-",round(percentage*100),"%")),                po
sition = position_stack(vjust = 0.5))+
  labs(x = NULL,y = NULL,fill = NULL)+
  theme_classic()+theme(axis.line = element_blank(),
                        axis.text = element_blank(),
                        axis.ticks = element_blank())
ggarrange(total, category, nrow = 1, ncol = 2)
```





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.