Project 2Comcast Telecom ConsumerComplaints.

By: Abdullah Alwabel

DESCRIPTION

Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a $2.3 million, after receiving over 1000 consumer complaints.

The existing database will serve as a repository of public customer complaints filed against Comcast.
It will help to pin down what is wrong with Comcast's customer service.

Data Dictionary

Ticket #: Ticket number assigned to each complaint
Customer Complaint: Description of complaint
Date: Date of complaint
Time: Time of complaint
Received Via: Mode of communication of the complaint
City: Customer city
State: Customer state
Zipcode: Customer zip
Status: Status of complaint
Filing on behalf of someone
Analysis Task

- Import data into R environment.
- Provide the trend chart for the number of complaints at monthly and daily granularity levels.
- Provide a table with the frequency of complaint types.

Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

- Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.
- Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:

Which state has the maximum complaints
Which state has the highest percentage of unresolved complaints
- Provide the percentage of complaints resolved till date, which were received through theInternet and customer care calls.

The analysis results to be provided with insights wherever applicable.

Analysis to be done:

1. - Import data into R environment.

CODE:

```
# Import data into R environment.
comcast_d=read.csv("C:/Users/User/Downloads/sda/Comcast Telecom Complaints data.csv")
View(comcast_d)
head(comcast_d)

summary(comcast_d)
str(comcast_d)

library(stringi)
library(lubridate)
library(dplyr)
library(ggplot2)

head(comcast_d)
names(comcast_d)<- stri_replace_all(regex = "\\.",replacement = "",str =names(comcast_d))
names
head(comcast_d)

na_vector <- is.na(comcast_d)
na_vector
length(na_vector[na_vector==T])
```

```
> # Import data into R environment.
> comcast_d=read.csv("C:/Users/User/Downloads/sda/Comcast Telecom Complaints data.csv")
> View(comcast_d)
> head(comcast_d)
  Ticket..
1   250635
2   223441
3   242732
4   277946
5   307175
6   338519
                                                Customer.Complaint
1                                    Comcast Cable Internet Speeds
2                      Payment disappear - service got disconnected
3                                                 Speed and Service
4 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.
5                           Comcast not working and no service to boot
6           ISP Charging for arbitrary data limits with overage fees
        Date        Time       Received.Via     City
1 22-04-2015  3:53:50 PM Customer Care Call Abingdon
2   4/8/2015 10:22:56 AM           Internet  Acworth
3 18-04-2015  9:55:47 AM           Internet  Acworth
4   5/7/2015 11:59:35 AM           Internet  Acworth
5 26-05-2015  1:25:26 PM           Internet  Acworth
6  6/12/2015  9:59:40 PM           Internet  Acworth
     State Zip.code Status
1 Maryland    21009 Closed
2  Georgia    30102 Closed
3  Georgia    30101 Closed
4  Georgia    30101   Open
5  Georgia    30101 Solved
6  Georgia    30101 Solved
  Filing.on.Behalf.of.Someone
1                           No
2                           No
3                          Yes
4                          Yes
5                           No
6                           No
>
> summary(comcast_d)
   Ticket..         Customer.Complaint
 Length:2224        Length:2224
 Class :character   Class :character
 Mode  :character   Mode  :character


     Date               Time
 Length:2224        Length:2224
 Class :character   Class :character
 Mode  :character   Mode  :character
```

```
 Received.Via          City
 Length:2224        Length:2224
 Class :character   Class :character
 Mode  :character   Mode  :character



    State             Zip.code
 Length:2224        Min.   : 1075
 Class :character   1st Qu.:30057
 Mode  :character   Median :37211
                    Mean   :47994
                    3rd Qu.:77059
                    Max.   :99223
    Status          Filing.on.Behalf.of.Someone
 Length:2224        Length:2224
 Class :character   Class :character
 Mode  :character   Mode  :character



> str(comcast_d)
'data.frame':   2224 obs. of  10 variables:
 $ Ticket..                  : chr  "250635" "223441" "242732" "277946" ...
 $ Customer.Complaint        : chr  "Comcast Cable Internet Speeds" "Payment disappear - service got disconnected" "Speed and Service" "Comcast Imposed a New Usage Cap of 300GB that punishes
 streaming." ...
 $ Date                      : chr  "22-04-2015" "4/8/2015" "18-04-2015" "5/7/2015" ...
 $ Time                      : chr  "3:53:50 PM" "10:22:56 AM" "9:55:47 AM" "11:59:35 AM" ...
 $ Received.Via              : chr  "Customer Care Call" "Internet" "Internet" "Internet" ...
 $ City                      : chr  "Abingdon" "Acworth" "Acworth" "Acworth" ...
 $ State                     : chr  "Maryland" "Georgia" "Georgia" "Georgia" ...
 $ Zip.code                  : int  21009 30102 30101 30101 30101 30101 30101 49221 94502 94501 ...
 $ Status                    : chr  "Closed" "Closed" "Closed" "Open" ...
 $ Filing.on.Behalf.of.Someone: chr  "No" "No" "Yes" "Yes" ...
>
> library(stringi)
> library(lubridate)
> library(dplyr)
> library(ggplot2)
>
> head(comcast_d)
  Ticket..
1   250635
2   223441
3   242732
4   277946
5   307175
6   338519
```

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

```
3  Georgia   30101 Closed                    Yes
4  Georgia   30101   Open                    Yes
5  Georgia   30101 Solved                     No
6  Georgia   30101 Solved                     No
>
> na_vector <- is.na(comcast_d)
> na_vector
       Ticket CustomerComplaint  Date  Time
  [1,]  FALSE             FALSE FALSE FALSE
  [2,]  FALSE             FALSE FALSE FALSE
  [3,]  FALSE             FALSE FALSE FALSE
  [4,]  FALSE             FALSE FALSE FALSE
  [5,]  FALSE             FALSE FALSE FALSE
  [6,]  FALSE             FALSE FALSE FALSE
  [7,]  FALSE             FALSE FALSE FALSE
  [8,]  FALSE             FALSE FALSE FALSE
  [9,]  FALSE             FALSE FALSE FALSE
 [10,]  FALSE             FALSE FALSE FALSE
 [11,]  FALSE             FALSE FALSE FALSE
 [12,]  FALSE             FALSE FALSE FALSE
 [13,]  FALSE             FALSE FALSE FALSE
 [14,]  FALSE             FALSE FALSE FALSE
 [15,]  FALSE             FALSE FALSE FALSE
 [16,]  FALSE             FALSE FALSE FALSE
 [17,]  FALSE             FALSE FALSE FALSE
 [18,]  FALSE             FALSE FALSE FALSE
 [19,]  FALSE             FALSE FALSE FALSE
 [20,]  FALSE             FALSE FALSE FALSE
 [21,]  FALSE             FALSE FALSE FALSE
 [22,]  FALSE             FALSE FALSE FALSE
 [23,]  FALSE             FALSE FALSE FALSE
 [24,]  FALSE             FALSE FALSE FALSE
 [25,]  FALSE             FALSE FALSE FALSE
 [26,]  FALSE             FALSE FALSE FALSE
 [27,]  FALSE             FALSE FALSE FALSE
 [28,]  FALSE             FALSE FALSE FALSE
 [29,]  FALSE             FALSE FALSE FALSE
 [30,]  FALSE             FALSE FALSE FALSE
 [31,]  FALSE             FALSE FALSE FALSE
 [32,]  FALSE             FALSE FALSE FALSE
 [33,]  FALSE             FALSE FALSE FALSE
 [34,]  FALSE             FALSE FALSE FALSE
 [35,]  FALSE             FALSE FALSE FALSE
 [36,]  FALSE             FALSE FALSE FALSE
 [37,]  FALSE             FALSE FALSE FALSE
 [38,]  FALSE             FALSE FALSE FALSE
 [39,]  FALSE             FALSE FALSE FALSE
 [40,]  FALSE             FALSE FALSE FALSE
 [41,]  FALSE             FALSE FALSE FALSE
 [42,]  FALSE             FALSE FALSE FALSE
 [43,]  FALSE             FALSE FALSE FALSE
 [44,]  FALSE             FALSE FALSE FALSE
 [45,]  FALSE             FALSE FALSE FALSE
 [46,]  FALSE             FALSE FALSE FALSE
 [47,]  FALSE             FALSE FALSE FALSE
 [48,]  FALSE             FALSE FALSE FALSE
```

```
  [48,]  FALSE                FALSE FALSE FALSE
  [49,]  FALSE                FALSE FALSE FALSE
  [50,]  FALSE                FALSE FALSE FALSE
  [51,]  FALSE                FALSE FALSE FALSE
  [52,]  FALSE                FALSE FALSE FALSE
  [53,]  FALSE                FALSE FALSE FALSE
  [54,]  FALSE                FALSE FALSE FALSE
  [55,]  FALSE                FALSE FALSE FALSE
  [56,]  FALSE                FALSE FALSE FALSE
  [57,]  FALSE                FALSE FALSE FALSE
  [58,]  FALSE                FALSE FALSE FALSE
  [59,]  FALSE                FALSE FALSE FALSE
  [60,]  FALSE                FALSE FALSE FALSE
  [61,]  FALSE                FALSE FALSE FALSE
  [62,]  FALSE                FALSE FALSE FALSE
  [63,]  FALSE                FALSE FALSE FALSE
  [64,]  FALSE                FALSE FALSE FALSE
  [65,]  FALSE                FALSE FALSE FALSE
  [66,]  FALSE                FALSE FALSE FALSE
  [67,]  FALSE                FALSE FALSE FALSE
  [68,]  FALSE                FALSE FALSE FALSE
  [69,]  FALSE                FALSE FALSE FALSE
  [70,]  FALSE                FALSE FALSE FALSE
  [71,]  FALSE                FALSE FALSE FALSE
  [72,]  FALSE                FALSE FALSE FALSE
  [73,]  FALSE                FALSE FALSE FALSE
  [74,]  FALSE                FALSE FALSE FALSE
  [75,]  FALSE                FALSE FALSE FALSE
  [76,]  FALSE                FALSE FALSE FALSE
  [77,]  FALSE                FALSE FALSE FALSE
  [78,]  FALSE                FALSE FALSE FALSE
  [79,]  FALSE                FALSE FALSE FALSE
  [80,]  FALSE                FALSE FALSE FALSE
  [81,]  FALSE                FALSE FALSE FALSE
  [82,]  FALSE                FALSE FALSE FALSE
  [83,]  FALSE                FALSE FALSE FALSE
  [84,]  FALSE                FALSE FALSE FALSE
  [85,]  FALSE                FALSE FALSE FALSE
  [86,]  FALSE                FALSE FALSE FALSE
  [87,]  FALSE                FALSE FALSE FALSE
  [88,]  FALSE                FALSE FALSE FALSE
  [89,]  FALSE                FALSE FALSE FALSE
  [90,]  FALSE                FALSE FALSE FALSE
  [91,]  FALSE                FALSE FALSE FALSE
  [92,]  FALSE                FALSE FALSE FALSE
  [93,]  FALSE                FALSE FALSE FALSE
  [94,]  FALSE                FALSE FALSE FALSE
  [95,]  FALSE                FALSE FALSE FALSE
  [96,]  FALSE                FALSE FALSE FALSE
  [97,]  FALSE                FALSE FALSE FALSE
  [98,]  FALSE                FALSE FALSE FALSE
  [99,]  FALSE                FALSE FALSE FALSE
 [100,]  FALSE                FALSE FALSE FALSE
       ReceivedVia  City State Zipcode Status
  [1,]       FALSE FALSE FALSE   FALSE  FALSE
  [2,]       FALSE FALSE FALSE   FALSE  FALSE
```

[3,]	FALSE	FALSE	FALSE	FALSE	FALSE
[4,]	FALSE	FALSE	FALSE	FALSE	FALSE
[5,]	FALSE	FALSE	FALSE	FALSE	FALSE
[6,]	FALSE	FALSE	FALSE	FALSE	FALSE
[7,]	FALSE	FALSE	FALSE	FALSE	FALSE
[8,]	FALSE	FALSE	FALSE	FALSE	FALSE
[9,]	FALSE	FALSE	FALSE	FALSE	FALSE
[10,]	FALSE	FALSE	FALSE	FALSE	FALSE
[11,]	FALSE	FALSE	FALSE	FALSE	FALSE
[12,]	FALSE	FALSE	FALSE	FALSE	FALSE
[13,]	FALSE	FALSE	FALSE	FALSE	FALSE
[14,]	FALSE	FALSE	FALSE	FALSE	FALSE
[15,]	FALSE	FALSE	FALSE	FALSE	FALSE
[16,]	FALSE	FALSE	FALSE	FALSE	FALSE
[17,]	FALSE	FALSE	FALSE	FALSE	FALSE
[18,]	FALSE	FALSE	FALSE	FALSE	FALSE
[19,]	FALSE	FALSE	FALSE	FALSE	FALSE
[20,]	FALSE	FALSE	FALSE	FALSE	FALSE
[21,]	FALSE	FALSE	FALSE	FALSE	FALSE
[22,]	FALSE	FALSE	FALSE	FALSE	FALSE
[23,]	FALSE	FALSE	FALSE	FALSE	FALSE
[24,]	FALSE	FALSE	FALSE	FALSE	FALSE
[25,]	FALSE	FALSE	FALSE	FALSE	FALSE
[26,]	FALSE	FALSE	FALSE	FALSE	FALSE
[27,]	FALSE	FALSE	FALSE	FALSE	FALSE
[28,]	FALSE	FALSE	FALSE	FALSE	FALSE
[29,]	FALSE	FALSE	FALSE	FALSE	FALSE
[30,]	FALSE	FALSE	FALSE	FALSE	FALSE
[31,]	FALSE	FALSE	FALSE	FALSE	FALSE
[32,]	FALSE	FALSE	FALSE	FALSE	FALSE
[33,]	FALSE	FALSE	FALSE	FALSE	FALSE
[34,]	FALSE	FALSE	FALSE	FALSE	FALSE
[35,]	FALSE	FALSE	FALSE	FALSE	FALSE
[36,]	FALSE	FALSE	FALSE	FALSE	FALSE
[37,]	FALSE	FALSE	FALSE	FALSE	FALSE
[38,]	FALSE	FALSE	FALSE	FALSE	FALSE
[39,]	FALSE	FALSE	FALSE	FALSE	FALSE
[40,]	FALSE	FALSE	FALSE	FALSE	FALSE
[41,]	FALSE	FALSE	FALSE	FALSE	FALSE
[42,]	FALSE	FALSE	FALSE	FALSE	FALSE
[43,]	FALSE	FALSE	FALSE	FALSE	FALSE
[44,]	FALSE	FALSE	FALSE	FALSE	FALSE
[45,]	FALSE	FALSE	FALSE	FALSE	FALSE
[46,]	FALSE	FALSE	FALSE	FALSE	FALSE
[47,]	FALSE	FALSE	FALSE	FALSE	FALSE
[48,]	FALSE	FALSE	FALSE	FALSE	FALSE
[49,]	FALSE	FALSE	FALSE	FALSE	FALSE
[50,]	FALSE	FALSE	FALSE	FALSE	FALSE
[51,]	FALSE	FALSE	FALSE	FALSE	FALSE
[52,]	FALSE	FALSE	FALSE	FALSE	FALSE
[53,]	FALSE	FALSE	FALSE	FALSE	FALSE
[54,]	FALSE	FALSE	FALSE	FALSE	FALSE
[55,]	FALSE	FALSE	FALSE	FALSE	FALSE
[56,]	FALSE	FALSE	FALSE	FALSE	FALSE
[57,]	FALSE	FALSE	FALSE	FALSE	FALSE
[58,]	FALSE	FALSE	FALSE	FALSE	FALSE

```
  [58,]       FALSE FALSE FALSE   FALSE  FALSE
  [59,]       FALSE FALSE FALSE   FALSE  FALSE
  [60,]       FALSE FALSE FALSE   FALSE  FALSE
  [61,]       FALSE FALSE FALSE   FALSE  FALSE
  [62,]       FALSE FALSE FALSE   FALSE  FALSE
  [63,]       FALSE FALSE FALSE   FALSE  FALSE
  [64,]       FALSE FALSE FALSE   FALSE  FALSE
  [65,]       FALSE FALSE FALSE   FALSE  FALSE
  [66,]       FALSE FALSE FALSE   FALSE  FALSE
  [67,]       FALSE FALSE FALSE   FALSE  FALSE
  [68,]       FALSE FALSE FALSE   FALSE  FALSE
  [69,]       FALSE FALSE FALSE   FALSE  FALSE
  [70,]       FALSE FALSE FALSE   FALSE  FALSE
  [71,]       FALSE FALSE FALSE   FALSE  FALSE
  [72,]       FALSE FALSE FALSE   FALSE  FALSE
  [73,]       FALSE FALSE FALSE   FALSE  FALSE
  [74,]       FALSE FALSE FALSE   FALSE  FALSE
  [75,]       FALSE FALSE FALSE   FALSE  FALSE
  [76,]       FALSE FALSE FALSE   FALSE  FALSE
  [77,]       FALSE FALSE FALSE   FALSE  FALSE
  [78,]       FALSE FALSE FALSE   FALSE  FALSE
  [79,]       FALSE FALSE FALSE   FALSE  FALSE
  [80,]       FALSE FALSE FALSE   FALSE  FALSE
  [81,]       FALSE FALSE FALSE   FALSE  FALSE
  [82,]       FALSE FALSE FALSE   FALSE  FALSE
  [83,]       FALSE FALSE FALSE   FALSE  FALSE
  [84,]       FALSE FALSE FALSE   FALSE  FALSE
  [85,]       FALSE FALSE FALSE   FALSE  FALSE
  [86,]       FALSE FALSE FALSE   FALSE  FALSE
  [87,]       FALSE FALSE FALSE   FALSE  FALSE
  [88,]       FALSE FALSE FALSE   FALSE  FALSE
  [89,]       FALSE FALSE FALSE   FALSE  FALSE
  [90,]       FALSE FALSE FALSE   FALSE  FALSE
  [91,]       FALSE FALSE FALSE   FALSE  FALSE
  [92,]       FALSE FALSE FALSE   FALSE  FALSE
  [93,]       FALSE FALSE FALSE   FALSE  FALSE
  [94,]       FALSE FALSE FALSE   FALSE  FALSE
  [95,]       FALSE FALSE FALSE   FALSE  FALSE
  [96,]       FALSE FALSE FALSE   FALSE  FALSE
  [97,]       FALSE FALSE FALSE   FALSE  FALSE
  [98,]       FALSE FALSE FALSE   FALSE  FALSE
  [99,]       FALSE FALSE FALSE   FALSE  FALSE
 [100,]       FALSE FALSE FALSE   FALSE  FALSE
      FilingonBehalfofSomeone
  [1,]                  FALSE
  [2,]                  FALSE
  [3,]                  FALSE
  [4,]                  FALSE
  [5,]                  FALSE
  [6,]                  FALSE
  [7,]                  FALSE
  [8,]                  FALSE
  [9,]                  FALSE
 [10,]                  FALSE
 [11,]                  FALSE
 [12,]                  FALSE
```

```
[13,] FALSE
[14,] FALSE
[15,] FALSE
[16,] FALSE
[17,] FALSE
[18,] FALSE
[19,] FALSE
[20,] FALSE
[21,] FALSE
[22,] FALSE
[23,] FALSE
[24,] FALSE
[25,] FALSE
[26,] FALSE
[27,] FALSE
[28,] FALSE
[29,] FALSE
[30,] FALSE
[31,] FALSE
[32,] FALSE
[33,] FALSE
[34,] FALSE
[35,] FALSE
[36,] FALSE
[37,] FALSE
[38,] FALSE
[39,] FALSE
[40,] FALSE
[41,] FALSE
[42,] FALSE
[43,] FALSE
[44,] FALSE
[45,] FALSE
[46,] FALSE
[47,] FALSE
[48,] FALSE
[49,] FALSE
[50,] FALSE
[51,] FALSE
[52,] FALSE
[53,] FALSE
[54,] FALSE
[55,] FALSE
[56,] FALSE
[57,] FALSE
[58,] FALSE
[59,] FALSE
[60,] FALSE
[61,] FALSE
[62,] FALSE
[63,] FALSE
[64,] FALSE
[65,] FALSE
[66,] FALSE
[67,] FALSE
[68,] FALSE
```

```
[13,] FALSE
[14,] FALSE
[15,] FALSE
[16,] FALSE
[17,] FALSE
[18,] FALSE
[19,] FALSE
[20,] FALSE
[21,] FALSE
[22,] FALSE
[23,] FALSE
[24,] FALSE
[25,] FALSE
[26,] FALSE
[27,] FALSE
[28,] FALSE
[29,] FALSE
[30,] FALSE
[31,] FALSE
[32,] FALSE
[33,] FALSE
[34,] FALSE
[35,] FALSE
[36,] FALSE
[37,] FALSE
[38,] FALSE
[39,] FALSE
[40,] FALSE
[41,] FALSE
[42,] FALSE
[43,] FALSE
[44,] FALSE
[45,] FALSE
[46,] FALSE
[47,] FALSE
[48,] FALSE
[49,] FALSE
[50,] FALSE
[51,] FALSE
[52,] FALSE
[53,] FALSE
[54,] FALSE
[55,] FALSE
[56,] FALSE
[57,] FALSE
[58,] FALSE
[59,] FALSE
[60,] FALSE
[61,] FALSE
[62,] FALSE
[63,] FALSE
[64,] FALSE
[65,] FALSE
[66,] FALSE
[67,] FALSE
[68,] FALSE
```

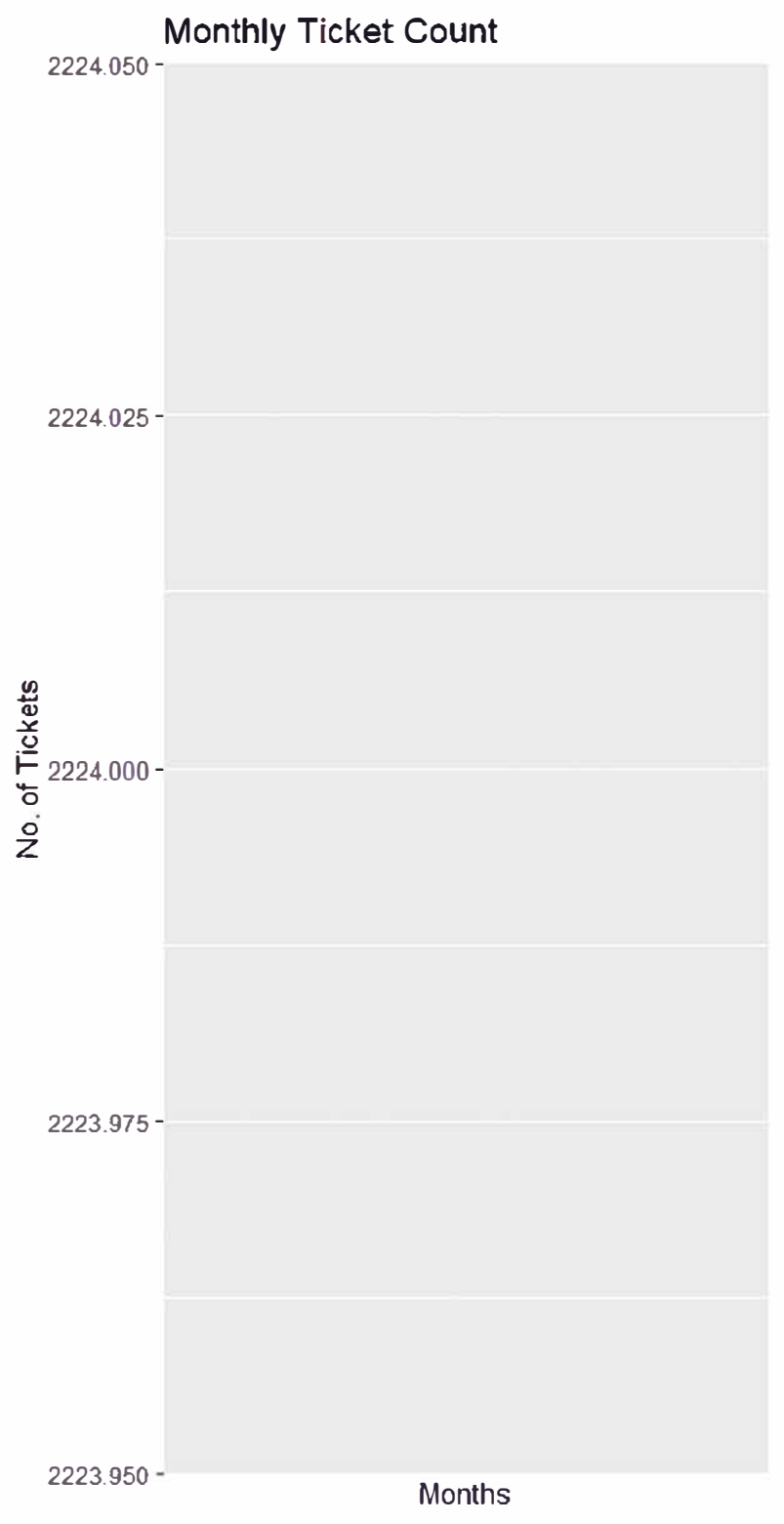
```
  [69,]                    FALSE
  [70,]                    FALSE
  [71,]                    FALSE
  [72,]                    FALSE
  [73,]                    FALSE
  [74,]                    FALSE
  [75,]                    FALSE
  [76,]                    FALSE
  [77,]                    FALSE
  [78,]                    FALSE
  [79,]                    FALSE
  [80,]                    FALSE
  [81,]                    FALSE
  [82,]                    FALSE
  [83,]                    FALSE
  [84,]                    FALSE
  [85,]                    FALSE
  [86,]                    FALSE
  [87,]                    FALSE
  [88,]                    FALSE
  [89,]                    FALSE
  [90,]                    FALSE
  [91,]                    FALSE
  [92,]                    FALSE
  [93,]                    FALSE
  [94,]                    FALSE
  [95,]                    FALSE
  [96,]                    FALSE
  [97,]                    FALSE
  [98,]                    FALSE
  [99,]                    FALSE
 [100,]                    FALSE
 [ reached getOption("max.print") -- omitted 2124 rows ]
> length(na_vector[na_vector==T])
[1] 0
>
```

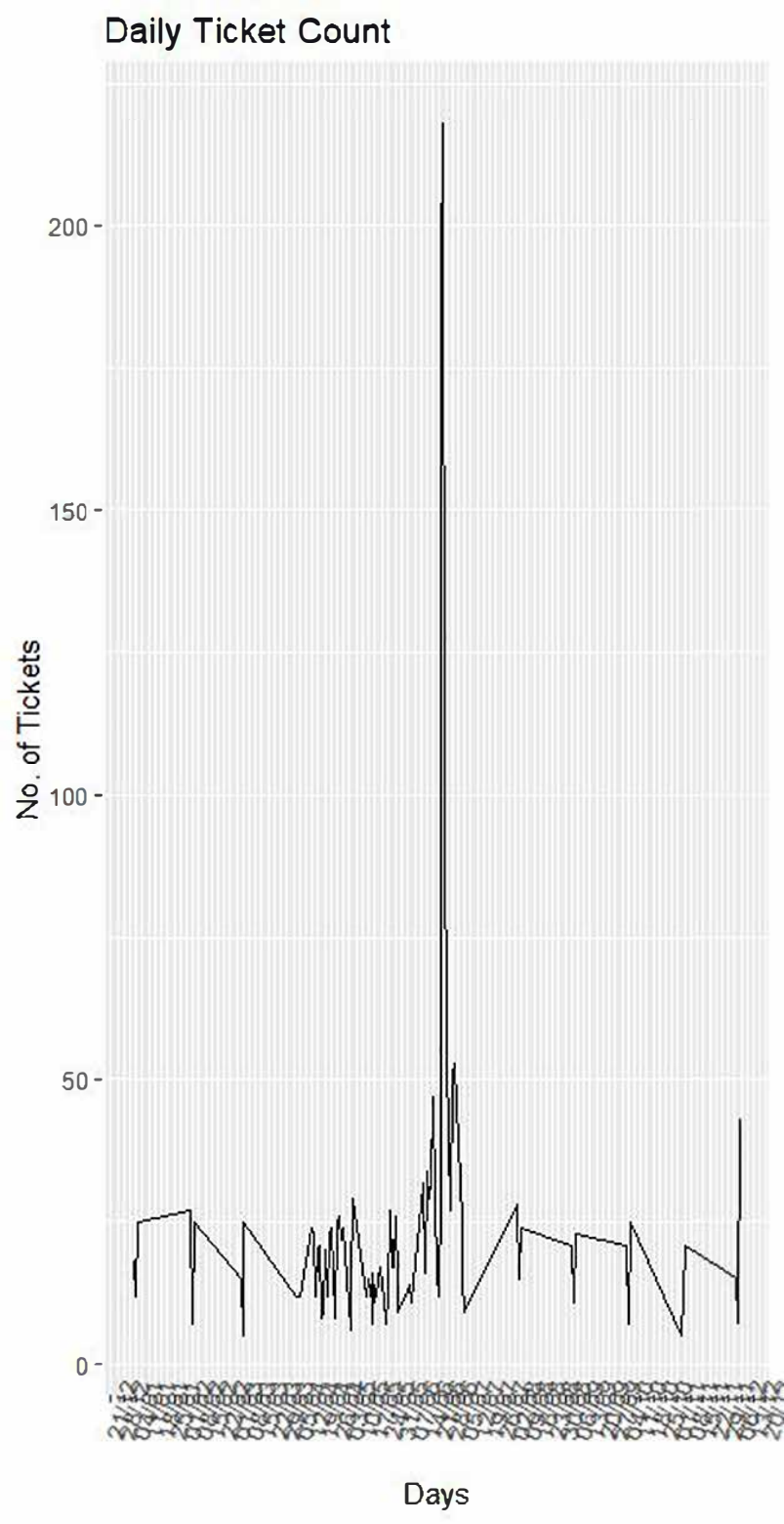
2. Provide the trend chart for the number of complaints at monthly and daily granularity levels..

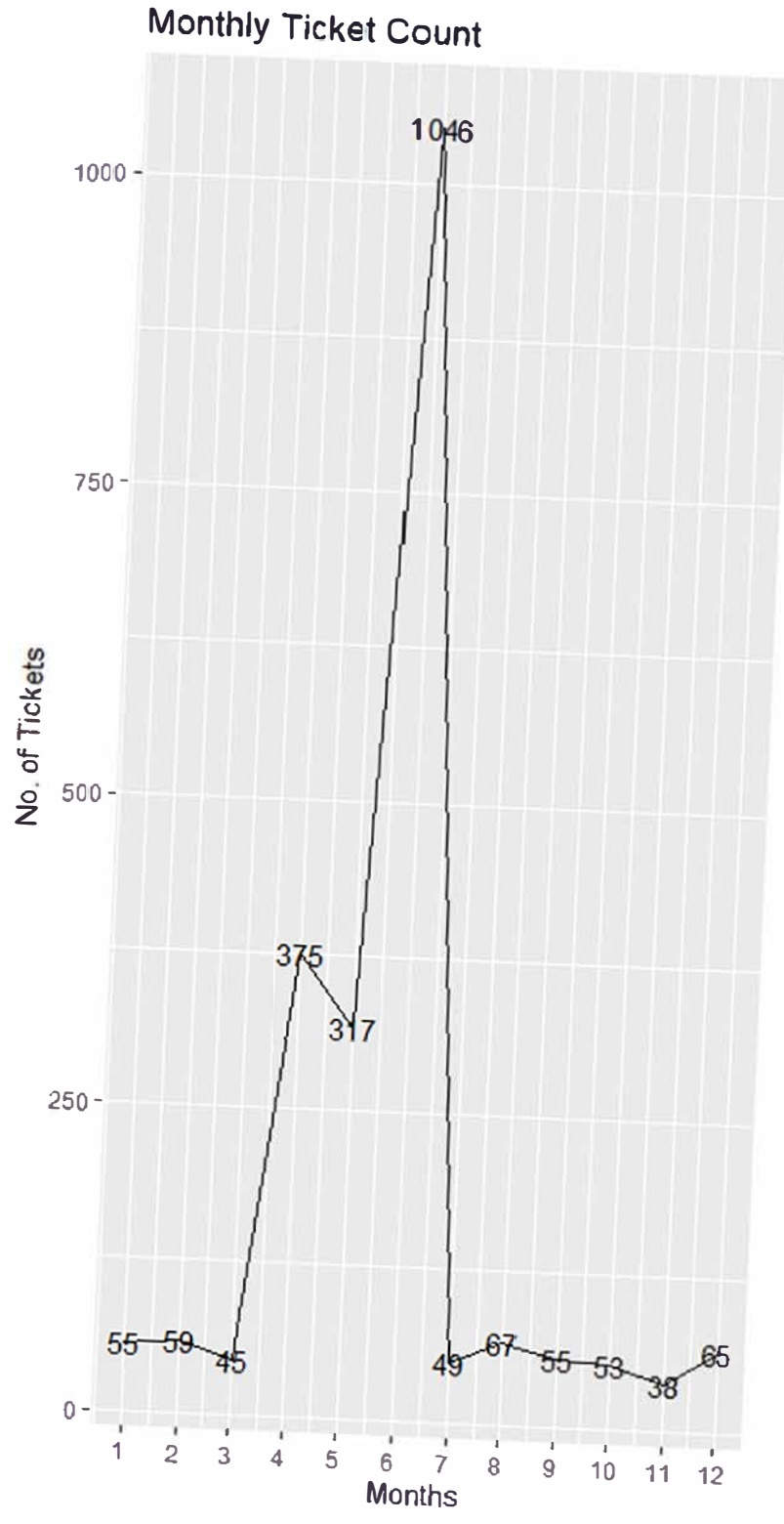
CODE:

```r
#Provide the trend chart for the number of complaints at monthly and daily granularity levels.
comcast_d$Date<- dmy(comcast_d$Date)
head(comcast_d)
monthly_count=arrange(summarise(group_by(comcast_d,month=as.integer(month(Date))),Count=n()),month)
daily_count=summarise(group_by(comcast_d,Date),Count=n())
monthly_count
daily_count
ggplot(data = monthly_count,aes(month,Count,label = Count))+geom_line()+geom_text()+scale_x_continuous(breaks = monthly_count$month)+labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")
ggplot(data = daily_count,aes(as.POSIXct(Date),Count))+geom_line()+theme(axis.text.x = element_text(angle = 75))+scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")+labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")
```

```
> #Provide the trend chart for the number of complaints at monthly and daily granularity levels.
> comcast_d$Date<- dmy(comcast_d$Date)
> head(comcast_d)
  Ticket
1 250635
2 223441
3 242732
4 277946
5 307175
6 338519
                                                CustomerComplaint
1                                  Comcast Cable Internet Speeds
2                       Payment disappear - service got disconnected
3                                               Speed and Service
4 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.
5                            Comcast not working and no service to boot
6          ISP Charging for arbitrary data limits with overage fees
        Date        Time        ReceivedVia     City
1 2015-04-22  3:53:50 PM Customer Care Call Abingdon
2 2015-08-04 10:22:56 AM           Internet  Acworth
3 2015-04-18  9:55:47 AM           Internet  Acworth
4 2015-07-05 11:59:35 AM           Internet  Acworth
5 2015-05-26  1:25:26 PM           Internet  Acworth
6 2015-12-06  9:59:40 PM           Internet  Acworth
     State Zipcode Status FilingonBehalfofSomeone
1 Maryland   21009 Closed                      No
2  Georgia   30102 Closed                      No
3  Georgia   30101 Closed                     Yes
4  Georgia   30101   Open                     Yes
5  Georgia   30101 Solved                      No
6  Georgia   30101 Solved                      No
> monthly_count=arrange(summarise(group_by(comcast_d,month=as.integer(month(Date))),Count=n()),month)
> daily_count=summarise(group_by(comcast_d,Date),Count=n())
> monthly_count
# A tibble: 12 x 2
   month Count
   <int> <int>
 1     1    55
 2     2    59
 3     3    45
 4     4   375
 5     5   317
 6     6  1046
 7     7    49
 8     8    67
 9     9    55
10    10    53
11    11    38
12    12    65
> daily_count
# A tibble: 91 x 2
   Date       Count
   <date>     <int>
 1 2015-01-04    18
 2 2015-01-05    12
 3 2015-01-06    25
 4 2015-02-04    27
 5 2015-02-05     7
 6 2015-02-06    25
 7 2015-03-04    15
 8 2015-03-05     5
 9 2015-03-06    25
10 2015-04-04    12
# ... with 81 more rows
> ggplot(data = monthly_count,aes(month,Count,label = Count))+geom_line()+geom_text()+scale_x_continuous(breaks = monthly_count$month)+labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")
> ggplot(data = daily_count,aes(as.POSIXct(Date),Count))+geom_line()+theme(axis.text.x = element_text(angle = 75))+scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")+labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")
>
```


Monthly Ticket Count
2224.050
2224.025
2224.000
2223.975
2223.950
No. of Tickets
Months


Daily Ticket Count
200
150
100
50
0
No. of Tickets
Days


Monthly Ticket Count
No. of Tickets
Months
0
250
500
750
1000
1 2 3 4 5 6 7 8 9 10 11 12
55
59
45
375
317
1046
49
67
55
53
38
65

3. -Provide a table with the frequency of complaint types.

CODE:

```r
#Provide a table with the frequency of complaint types.
#Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
network_tickets=contains(comcast_d$CustomerComplaint,match='network',ignore.case = T)
internet_tickets=contains(comcast_d$CustomerComplaint,match ='internet',ignore.case = T)
bill_tickets=contains(comcast_d$CustomerComplaint,match='bill',ignore.case = T)
email_tickets=contains(comcast_d$CustomerComplaint,match="email",ignore.case = T)
charge_tickets=contains(comcast_d$CustomerComplaint,match='charge',ignore.case = T)
comcast_d$ComplaintType[internet_tickets]='Internet'
comcast_d$ComplaintType[bill_tickets]='Billing'
comcast_d$ComplaintType[email_tickets]='Email'
comcast_d$ComplaintType[charge_tickets]='Charges'
comcast_d$ComplaintType[network_tickets]='Network'
comcast_d$ComplaintType[- c(network_tickets,internet_tickets,bill_tickets,email_tickets,charge_tickets)]="Others"
View(comcast_d)
table(comcast_d$ComplaintType)

```

```
> #Provide the trend chart for the number of complaints at monthly and daily granularity levels.
> comcast_d$Date<- dmy(comcast_d$Date)
Warning message:
All formats failed to parse. No formats found.
> head(comcast_d)
  Ticket                                           CustomerComplaint Date        Time        ReceivedVia     City    State Zipcode Status FilingonBehalfofSomeone
1 250635                                 Comcast Cable Internet Speeds <NA>  3:53:50 PM Customer Care Call Abingdon Maryland   21009 Closed                     No
2 223441                     Payment disappear - service got disconnected <NA> 10:22:56 AM           Internet  Acworth  Georgia   30102 Closed                     No
3 242732                                  Speed and Service <NA>  9:55:47 AM           Internet  Acworth  Georgia   30101 Closed                    Yes
4 277946 Comcast Imposed a New Usage Cap of 300GB that punishes streaming. <NA> 11:59:35 AM           Internet  Acworth  Georgia   30101   Open                    Yes
5 307175                    Comcast not working and no service to boot <NA>  1:25:26 PM           Internet  Acworth  Georgia   30101 Solved                     No
6 338519          ISP Charging for arbitrary data limits with overage fees <NA>  9:59:40 PM           Internet  Acworth  Georgia   30101 Solved                     No
> monthly_count=arrange(summarise(group_by(comcast_d,month=as.integer(month(Date))),Count=n()),month)
> daily_count=summarise(group_by(comcast_d,Date),Count=n())
> monthly_count
# A tibble: 1 x 2
  month Count
  <int> <int>
1    NA  2224
> daily_count
# A tibble: 1 x 2
  Date       Count
  <date>     <int>
1 NA          2224
> ggplot(data = monthly_count,aes(month,Count,label = Count))+geom_line()+geom_text()+scale_x_continuous(breaks = monthly_count$month)+labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")
Warning messages:
1: Removed 1 row(s) containing missing values (geom_path).
2: Removed 1 rows containing missing values (geom_text).
> ggplot(data = daily_count,aes(as.POSIXct(Date),Count))+geom_line()+theme(axis.text.x = element_text(angle = 75))+scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")+labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")
Error in seq.int(0, to0 - from, by) : 'to' must be a finite number
>
```

4. Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.

CODE:

```r
#Create a new categorical variable with value as Open and Closed.
#Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.
open_complaints=(comcast_d$Status=='Open'|comcast_d$Status=='Pending')
closed_complaints=(comcast_d$Status=='Closed'|comcast_d$Status=='Solved')
comcast_d$ComplaintStatus[open_complaints]="Open"
comcast_d$ComplaintStatus[closed_complaints]='Closed'

```

```
> #Create a new categorical variable with value as Open and Closed.
> #Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.
> open_complaints=(comcast_d$Status=='Open'|comcast_d$Status=='Pending')
> closed_complaints=(comcast_d$Status=='Closed'|comcast_d$Status=='Solved')
> comcast_d$ComplaintStatus[open_complaints]="Open"
> comcast_d$ComplaintStatus[closed_complaints]='Closed'
>
```

5. Provide state wise status of complaints in a stacked bar chart.Use the categorized variable from Q3. Provide insights on:

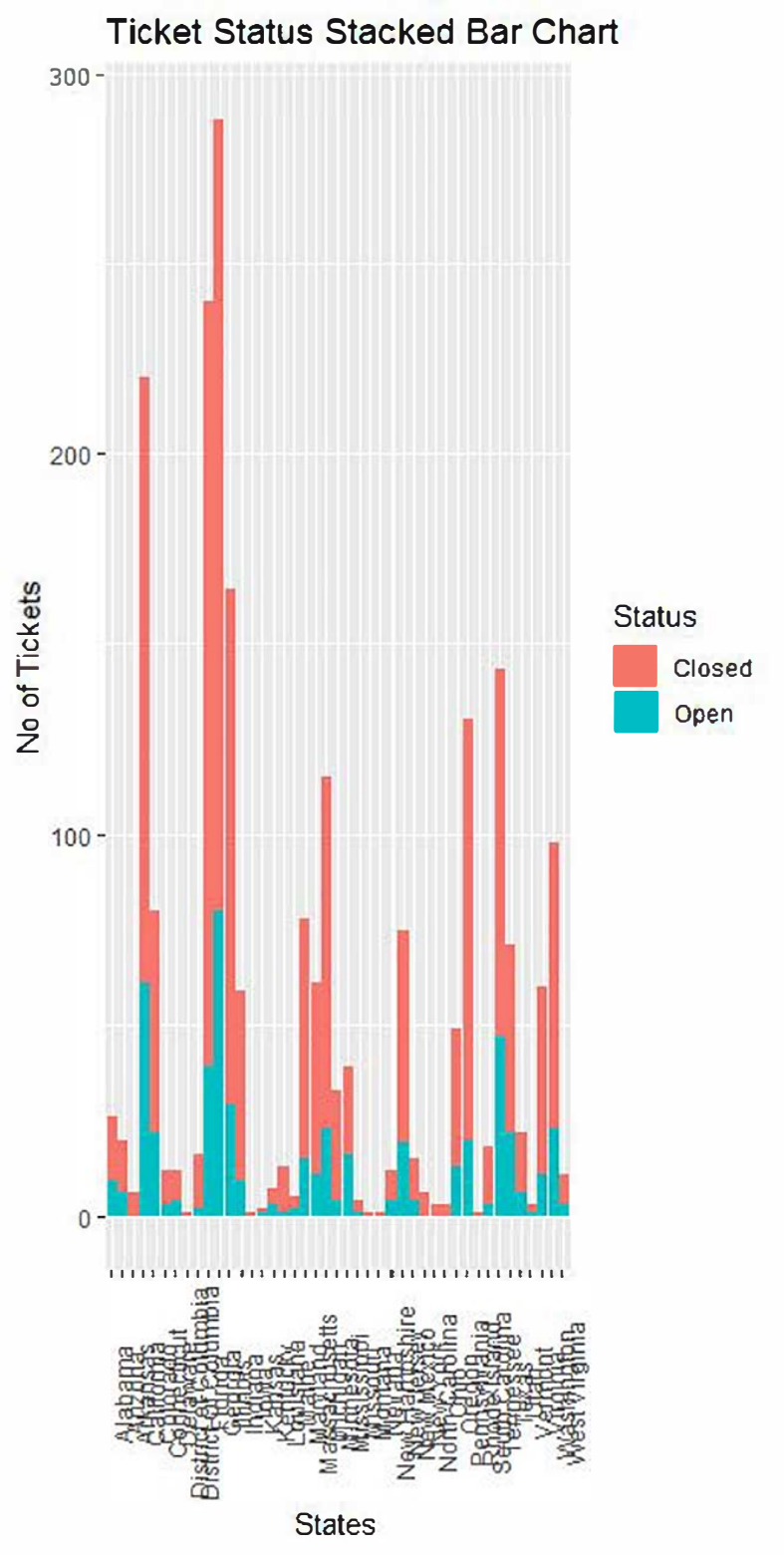
Which state has the maximum complaints
Which state has the highest percentage of unresolved complaints

CODE:

```r
#Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3.
chart_data=summarize(group_by(comcast_d,State,ComplaintStatus),Count=n())
chart_data
chart_data=as.data.frame(chart_data)
chart_data
ggplot(chart_data ,mapping = aes(State,Count))+geom_col(aes(fill = ComplaintStatus),width = 0.95)+theme(axis.text.x = element_text(angle = 90))+labs(title = "Ticket Status Stacked Bar Chart ",x = "States",y = "No of Tickets",fill= "Status")
```

```
> #Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3.
> chart_data=summarize(group_by(comcast_d,State,ComplaintStatus),Count=n())
`summarise()` has grouped output by 'State'. You can override using the `.groups` argument.
> chart_data
# A tibble: 77 x 3
# Groups:   State [43]
   State       ComplaintStatus Count
   <chr>       <chr>           <int>
 1 Alabama     Closed             17
 2 Alabama     Open                9
 3 Arizona     Closed             14
 4 Arizona     Open                6
 5 Arkansas    Closed              6
 6 California  Closed            159
 7 California  Open               61
 8 Colorado    Closed             58
 9 Colorado    Open               22
10 Connecticut Closed              9
# ... with 67 more rows
> chart_data=as.data.frame(chart_data)
> chart_data
                State ComplaintStatus Count
1             Alabama          Closed    17
2             Alabama            Open     9
3             Arizona          Closed    14
4             Arizona            Open     6
5            Arkansas          Closed     6
6          California          Closed   159
7          California            Open    61
8            Colorado          Closed    58
9            Colorado            Open    22
10         Connecticut          Closed     9
11         Connecticut            Open     3
12            Delaware          Closed     8
13            Delaware            Open     4
14 District of Columbia          Closed     1
15 District Of Columbia          Closed    14
16 District Of Columbia            Open     2
17             Florida          Closed   201
18             Florida            Open    39
19             Georgia          Closed   208
20             Georgia            Open    80
21            Illinois          Closed   135
22            Illinois            Open    29
23             Indiana          Closed    50
24             Indiana            Open     9
25                Iowa          Closed     1
26              Kansas          Closed     1
27              Kansas            Open     1
28            Kentucky          Closed     4
29            Kentucky            Open     3
30           Louisiana          Closed    12
31           Louisiana            Open     1
32               Maine          Closed     3
33               Maine            Open     2
34            Maryland          Closed    63
35            Maryland            Open    15
36       Massachusetts          Closed    50
37       Massachusetts            Open    11
38            Michigan          Closed    92
39            Michigan            Open    23
40           Minnesota          Closed    29
41           Minnesota            Open     4
42         Mississippi          Closed    23
43         Mississippi            Open    16
44            Missouri          Closed     3
45            Missouri            Open     1
46             Montana          Closed     1
47              Nevada          Closed     1
48       New Hampshire          Closed     8
49       New Hampshire            Open     4
50          New Jersey          Closed    56
51          New Jersey            Open    19
52          New Mexico          Closed    11
53          New Mexico            Open     4
54            New York          Closed     6
55      North Carolina          Closed     3
56                Ohio          Closed     3
57              Oregon          Closed    36
58              Oregon            Open    13
59        Pennsylvania          Closed   110
60        Pennsylvania            Open    20
```

```
61          Rhode Island          Closed     1
62        South Carolina          Closed    15
63        South Carolina            Open     3
64             Tennessee          Closed    96
65             Tennessee            Open    47
66                 Texas          Closed    49
67                 Texas            Open    22
68                  Utah          Closed    16
69                  Utah            Open     6
70               Vermont          Closed     2
71               Vermont            Open     1
72              Virginia          Closed    49
73              Virginia            Open    11
74            Washington          Closed    75
75            Washington            Open    23
76         West Virginia          Closed     8
77         West Virginia            Open     3
> ggplot(chart_data ,mapping = aes(State,Count))+geom_col(aes(fill = ComplaintStatus),width = 0.95)+theme(axis.text.x = element_text(angle = 90))+labs(title = "Ticket Status Stacked Bar Chart ",x = "States",y = "No of Tickets",fill= "Status")
>
```


Ticket Status Stacked Bar Chart
300
200
100
0
No of Tickets
Status
Closed
Open
States

```
63 #which state has the maximum complaints
64 max(chart_data$Count)
65 arrange(select(chart_data,State,Count),desc(Count))
66
```

```
> #which state has the maximum complaints
> max(chart_data$Count)
[1] 208
> arrange(select(chart_data,State,Count),desc(Count))
                  State Count
1               Georgia   208
2               Florida   201
3            California   159
4              Illinois   135
5          Pennsylvania   110
6             Tennessee    96
7              Michigan    92
8               Georgia    80
9            Washington    75
10             Maryland    63
11           California    61
12             Colorado    58
13           New Jersey    56
14              Indiana    50
15        Massachusetts    50
16                Texas    49
17             Virginia    49
18            Tennessee    47
19              Florida    39
20               Oregon    36
21             Illinois    29
22            Minnesota    29
23             Michigan    23
24          Mississippi    23
25           Washington    23
26             Colorado    22
27                Texas    22
28         Pennsylvania    20
29           New Jersey    19
30              Alabama    17
31          Mississippi    16
32                 Utah    16
33             Maryland    15
34       South Carolina    15
35              Arizona    14
36 District Of Columbia    14
37               Oregon    13
38            Louisiana    12
39        Massachusetts    11
40           New Mexico    11
41             Virginia    11
42              Alabama     9
43          Connecticut     9
44              Indiana     9
45             Delaware     8
46        New Hampshire     8
47        West Virginia     8
48              Arizona     6
49             Arkansas     6
50             New York     6
51                 Utah     6
52             Delaware     4
53             Kentucky     4
54            Minnesota     4
55        New Hampshire     4
56           New Mexico     4
57          Connecticut     3
58             Kentucky     3
59                Maine     3
60             Missouri     3
61       North Carolina     3
62                 Ohio     3
63       South Carolina     3
64        West Virginia     3
65 District Of Columbia     2
66                Maine     2
67              Vermont     2
68 District of Columbia     1
69                 Iowa     1
70               Kansas     1
71               Kansas     1
72            Louisiana     1
73             Missouri     1
74              Montana     1
75               Nevada     1
76         Rhode Island     1
77              Vermont     1
```

```
67  #which state has the highest percentage of unresolved complaints
68  arrange(filter(chart_data,ComplaintStatus=="Open"),desc(Count))
69
```

```
> #which state has the highest percentage of unresolved complaints
> arrange(filter(chart_data,ComplaintStatus=="Open"),desc(Count))
```

	State	ComplaintStatus	Count
1	Georgia	Open	80
2	California	Open	61
3	Tennessee	Open	47
4	Florida	Open	39
5	Illinois	Open	29
6	Michigan	Open	23
7	Washington	Open	23
8	Colorado	Open	22
9	Texas	Open	22
10	Pennsylvania	Open	20
11	New Jersey	Open	19
12	Mississippi	Open	16
13	Maryland	Open	15
14	Oregon	Open	13
15	Massachusetts	Open	11
16	Virginia	Open	11
17	Alabama	Open	9
18	Indiana	Open	9
19	Arizona	Open	6
20	Utah	Open	6
21	Delaware	Open	4
22	Minnesota	Open	4
23	New Hampshire	Open	4
24	New Mexico	Open	4
25	Connecticut	Open	3
26	Kentucky	Open	3
27	South Carolina	Open	3
28	West Virginia	Open	3
29	District Of Columbia	Open	2
30	Maine	Open	2
31	Kansas	Open	1
32	Louisiana	Open	1
33	Missouri	Open	1
34	Vermont	Open	1

```
>
```

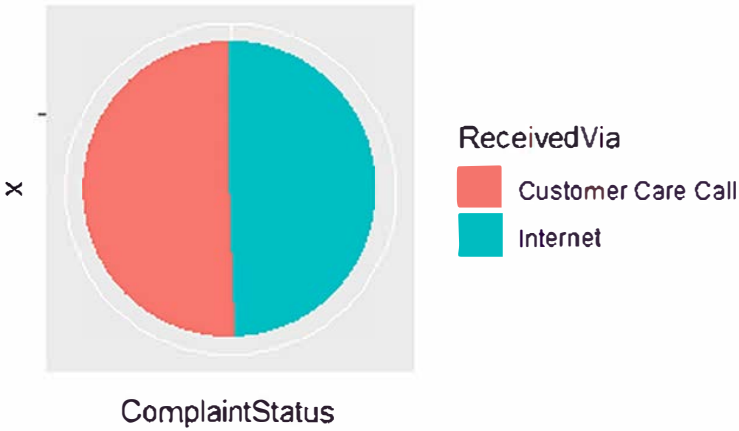
6. Provide the percentage of complaints resolved till date, which were received through theInternet and customer care calls

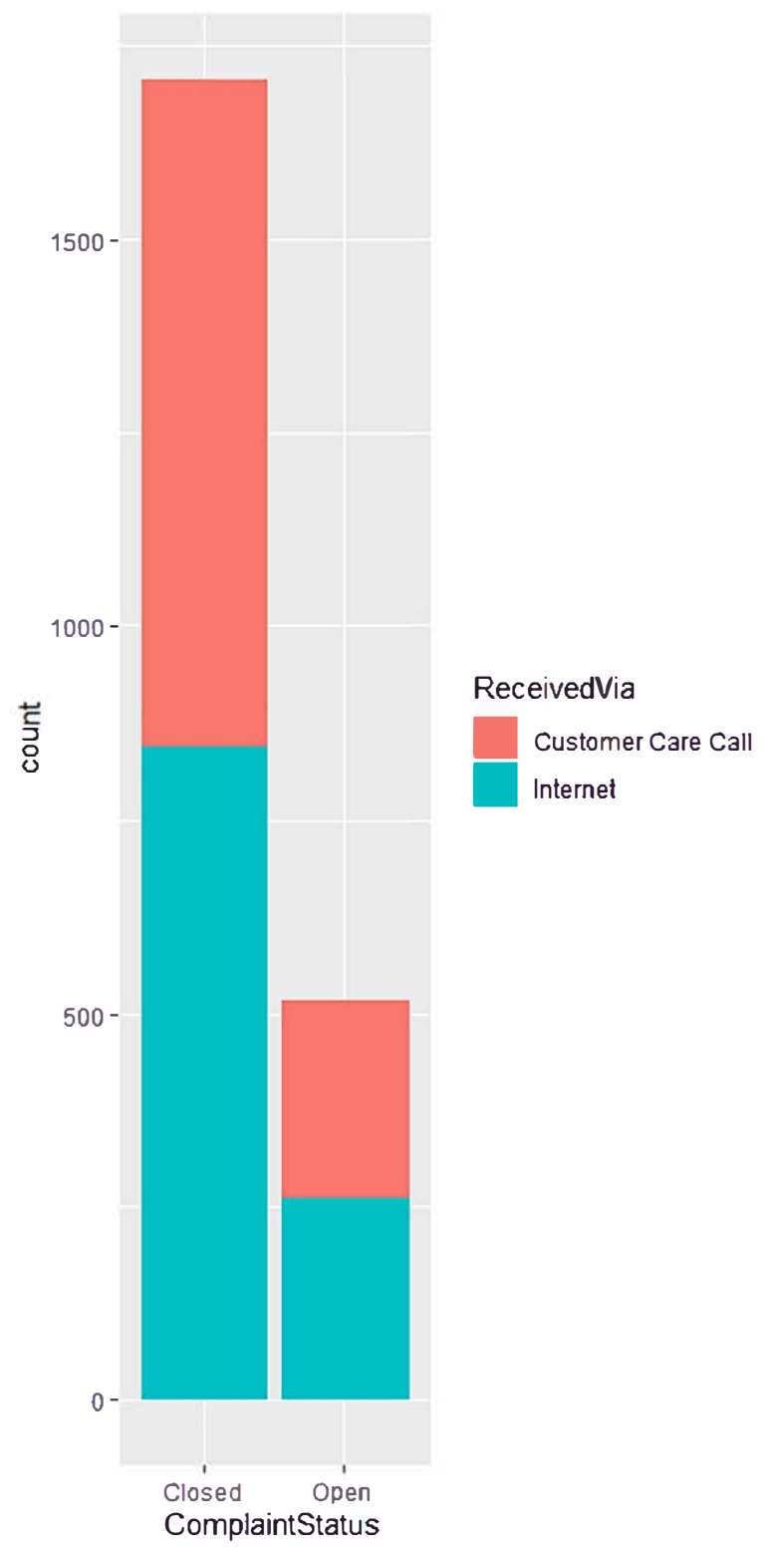
CODE:

```r
#Provide the percentage of complaints resolved till date,
#which were received through the Internet and customer care calls.
resolved=summarise(filter(comcast_d,ComplaintStatus=='Closed'),count=n())
resolved
resolved_internet=summarise(filter(comcast_d,ComplaintStatus=='Closed',ReceivedVia=='Internet'),count=n())
resolved_internet
resolved_CustomerCare=summarise(filter(comcast_d,ComplaintStatus=='Closed',ReceivedVia=='Customer Care Call'),count=n())
resolved_CustomerCare
percentage_internet=(resolved_internet/resolved)*100
percentage_internet
percentage_CustomerCare=(resolved_CustomerCare/resolved)*100
percentage_CustomerCare
table_df=table(comcast_d$ReceivedVia,comcast_d$ComplaintStatus)
table_df
bar=ggplot(comcast_d,aes(ComplaintStatus,fill=ReceivedVia))+geom_bar()
bar
resolved_df=select(filter(comcast_d,ComplaintStatus=='Closed'),ComplaintStatus, ReceivedVia)
pie<- ggplot(resolved_df, aes(x="", y= ComplaintStatus,fill=ReceivedVia)) +geom_bar(width = 1, stat = "identity") +coord_polar("y")
pie
```

```
> #Provide the percentage of complaints resolved till date,
> #which were received through the Internet and customer care calls.
> resolved=summarise(filter(comcast_d,ComplaintStatus=='closed'),count=n())
> resolved
  count
1  1707
> resolved_internet=summarise(filter(comcast_d,ComplaintStatus=='Closed',Receivedvia=='Internet'),count=n())
> resolved_internet
  count
1   843
> resolved_CustomerCare=summarise(filter(comcast_d,ComplaintStatus=='Closed',Receivedvia=='Customer Care Call'),count=n())
> resolved_CustomerCare
  count
1   864
> percentage_internet=(resolved_internet/resolved)*100
> percentage_internet
     count
1 49.38489
> percentage_CustomerCare=(resolved_CustomerCare/resolved)*100
> percentage_CustomerCare
     count
1 50.61511
> table_df=table(comcast_d$Receivedvia,comcast_d$ComplaintStatus)
> table_df

                   Closed Open
  Customer Care Call    864  255
  Internet              843  262
> bar=ggplot(comcast_d,aes(ComplaintStatus,fill=Receivedvia))+geom_bar()
> bar
> resolved_df=select(filter(comcast_d,ComplaintStatus=='Closed'),ComplaintStatus, Receivedvia)
> pie<- ggplot(resolved_df, aes(x="", y= ComplaintStatus,fill=Receivedvia)) +geom_bar(width = 1, stat = "identity") +coord_polar("y")
> pie
>
```


x
ReceivedVia
Customer Care Call
Internet
ComplaintStatus


1500
1000
500
0
count
ReceivedVia
Customer Care Call
Internet
Closed
Open
ComplaintStatus