

Comcast Telecom Consumer Complaints

Andrei Enescu

10/14/2020

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

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

1. Import data into R environment.
2. Provide the trend chart for the number of complaints at monthly and daily granularity levels.
3. Provide a table with the frequency of complaint types.
4. Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
 - 4.1. 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.
 - 4.2. Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3.
5. Which state has the maximum complaints
6. Which state has the highest percentage of unresolved complaints
 - 6.1 Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

The analysis results to be provided with insights wherever applicable

These are the libraries used

```
library(rio)
library(lubridate)
library(dplyr)
library(ggplot2)
library(scales)
library(tidyselect)
library(wordcloud)
library(plotrix)
```

1. Import data into R environment.

The data will be loaded using the RIO library

```
ComcastDF <- import("Comcast Telecom Complaints data.csv")
head(ComcastDF)
```

```
##      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
```

```
str(ComcastDF)
```

```
## 'data.frame':    2224 obs. of  10 variables:
## $ Ticket #      : chr  "250635" "223441" "242732" "277946" ...
## $ Customer Complaint : chr  "Comcast Cable Internet Speeds" "Payment disappear - service got
## $ 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 30101 49221 94502 94501 ...
## $ Status : chr "Closed" "Closed" "Closed" "Open" ...
## $ Filing on Behalf of Someone: chr "No" "No" "Yes" "Yes" ...
```

Checking for NA / missing values

```
sum(is.na(ComcastDF))
```

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

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

First, I am formatting the Date column and separate the month and day into new columns

```
head(ComcastDF$Date)
```

```
## [1] "22-04-2015" "4/8/2015" "18-04-2015" "5/7/2015" "26-05-2015"
## [6] "6/12/2015"
```

```
ComcastDF$Date <- dmy(ComcastDF$Date)
head(ComcastDF$Date)
```

```
## [1] "2015-04-22" "2015-08-04" "2015-04-18" "2015-07-05" "2015-05-26"
## [6] "2015-12-06"
```

```
class(ComcastDF$Date)
```

```
## [1] "Date"
```

```
ComcastDF$Month <- as.Date(cut(ComcastDF$Date, breaks = "month"))
ComcastDF$Day <- as.Date(cut(ComcastDF$Date, breaks = "day"))
ComcastDF["Count"] <- 1
```

```
ComcastDF1 <-
  aggregate(ComcastDF[c("Count")],
    by = list(Month = ComcastDF$Month),
    FUN = sum)
```

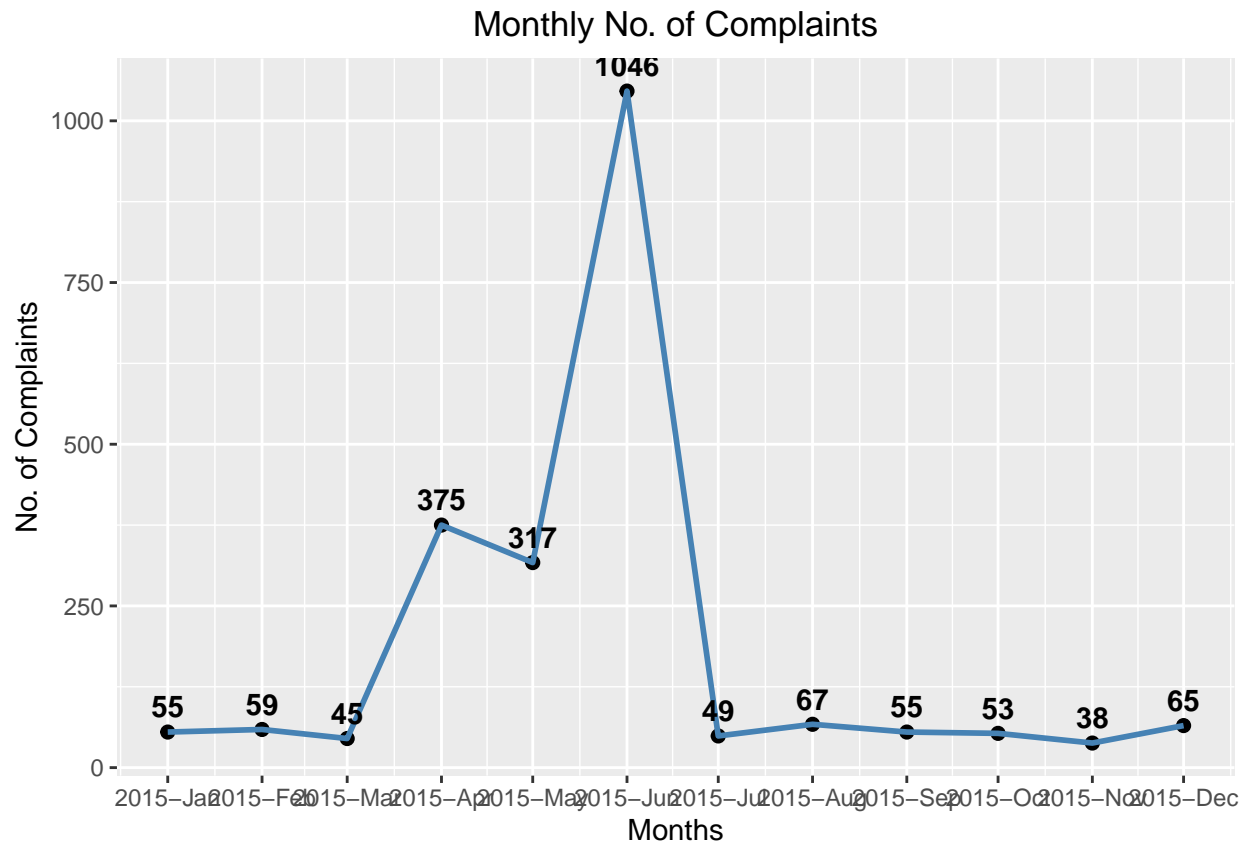
```
ComcastDF2 <-
  aggregate(ComcastDF[c("Count")],
    by = list(Day = ComcastDF$Day),
    FUN = sum)
```

```
ggplot(ComcastDF1,
```

```

aes(x = Month, y = Count, label = Count)) +
geom_point(size = 2) +
geom_line(size = 1, color = "steelblue") +
geom_text(vjust = -0.7, fontface = 2) +
scale_x_continuous(breaks = ComcastDF1$Month,
                    labels = date_format("%Y-%b")) +
labs(title = "Monthly No. of Complaints",
     x = "Months",
     y = "No. of Complaints") +
theme(plot.title = element_text(hjust = 0.5))

```



Observation

From the above trend chart we can see that there is an increase in complaints in the months of April, May and with a peak of 1046 complaints in June.

```

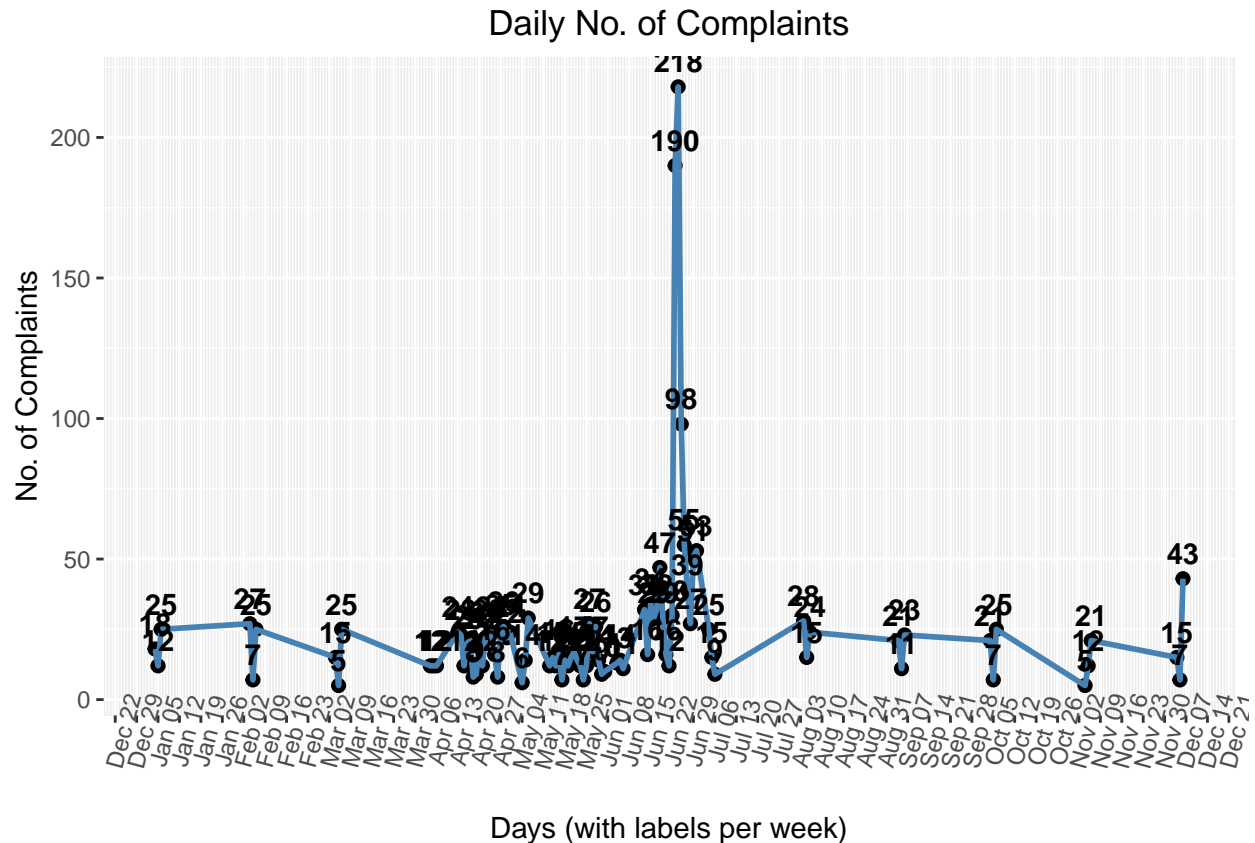
ggplot(ComcastDF2,
       aes(x = Day, y = Count, label = Count)) +
geom_point(size = 2) +
geom_line(size = 1, color = "steelblue") +
geom_text(vjust = -0.7, fontface = 2) +
labs(title = "Daily No. of Complaints",
     x = "Days (with labels per week)",
     y = "No. of Complaints") +
theme(axis.text.x = element_text(angle = 75, size = 9),
      plot.title = element_text(hjust = 0.5)) +
scale_x_date(

```

```

date_breaks = "1 week",
date_labels = "%b %d",
date_minor_breaks = "1 day"
)

```



Observation

Analyzing the above daily chart we can observe that in the second half of June there were more complaints in comparison to the rest of the month / year.

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

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

In order to have the frequency of complaint types I am using wordcloud and unlist.

```

wordcloud(ComcastDF$'Customer Complaint', min.freq = 10, colors = T)

```



```

contains(ComcastDF$`Customer Complaint`,
         match = "pric",
         ignore.case = T)
data.cap_complaints <-
  contains(ComcastDF$`Customer Complaint`,
          match = "data cap",
          ignore.case = T)

ComcastDF$`Complaint Type`[internet_complaints] <- "Internet"
ComcastDF$`Complaint Type`[network_complaints] <- "Network"
ComcastDF$`Complaint Type`[service_complaints] <- "Service"
ComcastDF$`Complaint Type`[billing_complaints] <- "Billing"
ComcastDF$`Complaint Type`[speed_complaints] <- "Speed"
ComcastDF$`Complaint Type`[price_complaints] <- "Price"
ComcastDF$`Complaint Type`[data.cap_complaints] <- "Data cap"

ComcastDF$`Complaint Type`[-c(
  internet_complaints,
  network_complaints,
  service_complaints,
  billing_complaints,
  speed_complaints,
  price_complaints,
  data.cap_complaints
)] <- "Others"

sort.int(table(ComcastDF$`Complaint Type`), decreasing = T)

```

```

##
##   Others  Service  Billing Internet   Speed Data cap   Price  Network
##     865     390     286     251     181     150     100         1

```

Observation

As we can see from the above table, there still are a great number of complaints categorized as “Others”, but we can observe that most of the complaints are referring to the service (with 390 tickets)

4.1. 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.

```

ComcastDF <-
  transform(ComcastDF, `Complaint Status` =
    ifelse((Status == 'Open' |

```

```
      Status == 'Pending'), "Open", "Closed"))
head(ComcastDF$Status, 10)
```

```
## [1] "Closed" "Closed" "Closed" "Open" "Solved" "Solved" "Pending"
## [8] "Solved" "Closed" "Open"
```

```
head(ComcastDF$Complaint.Status, 10)
```

```
## [1] "Closed" "Closed" "Closed" "Open" "Closed" "Closed" "Open" "Closed"
## [9] "Closed" "Open"
```

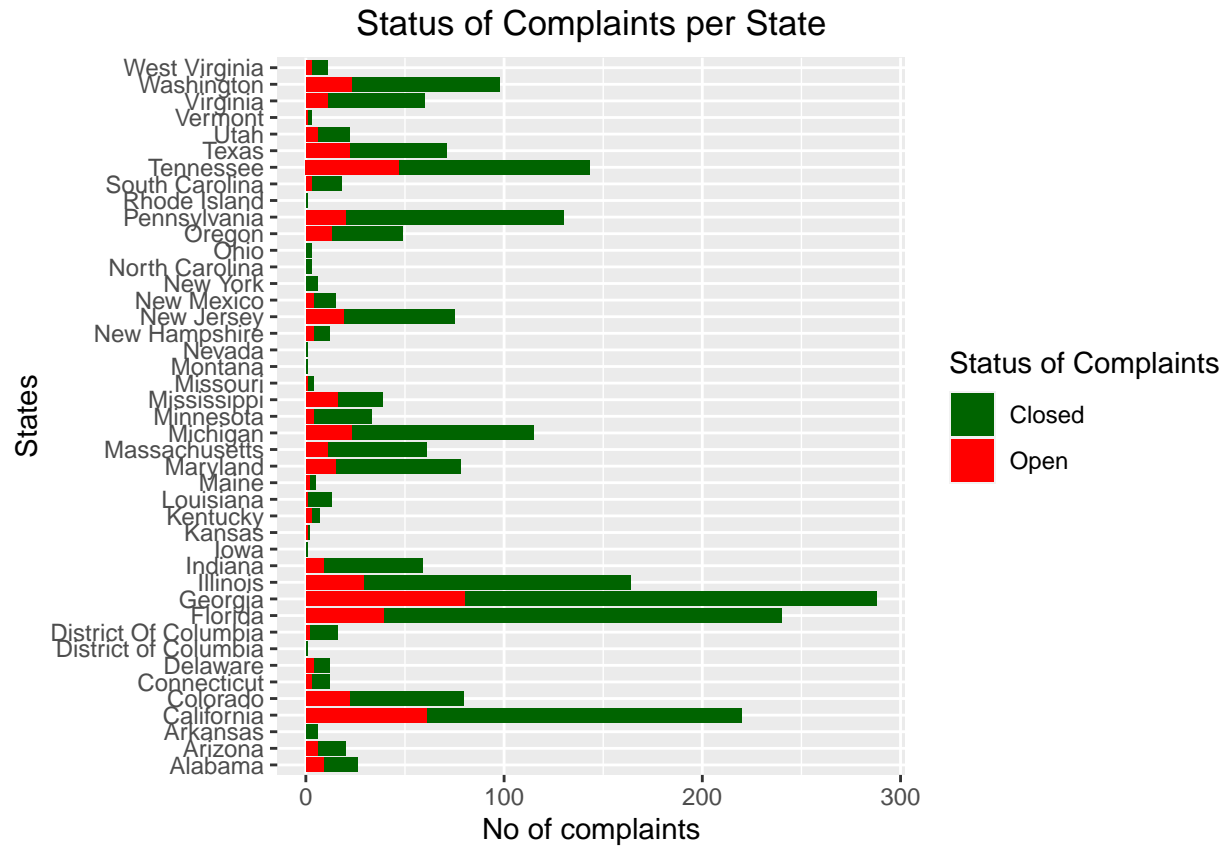
4.2. Provide state wise status of complaints in a stacked bar chart.

Use the categorized variable from Q3. Provide insights on:

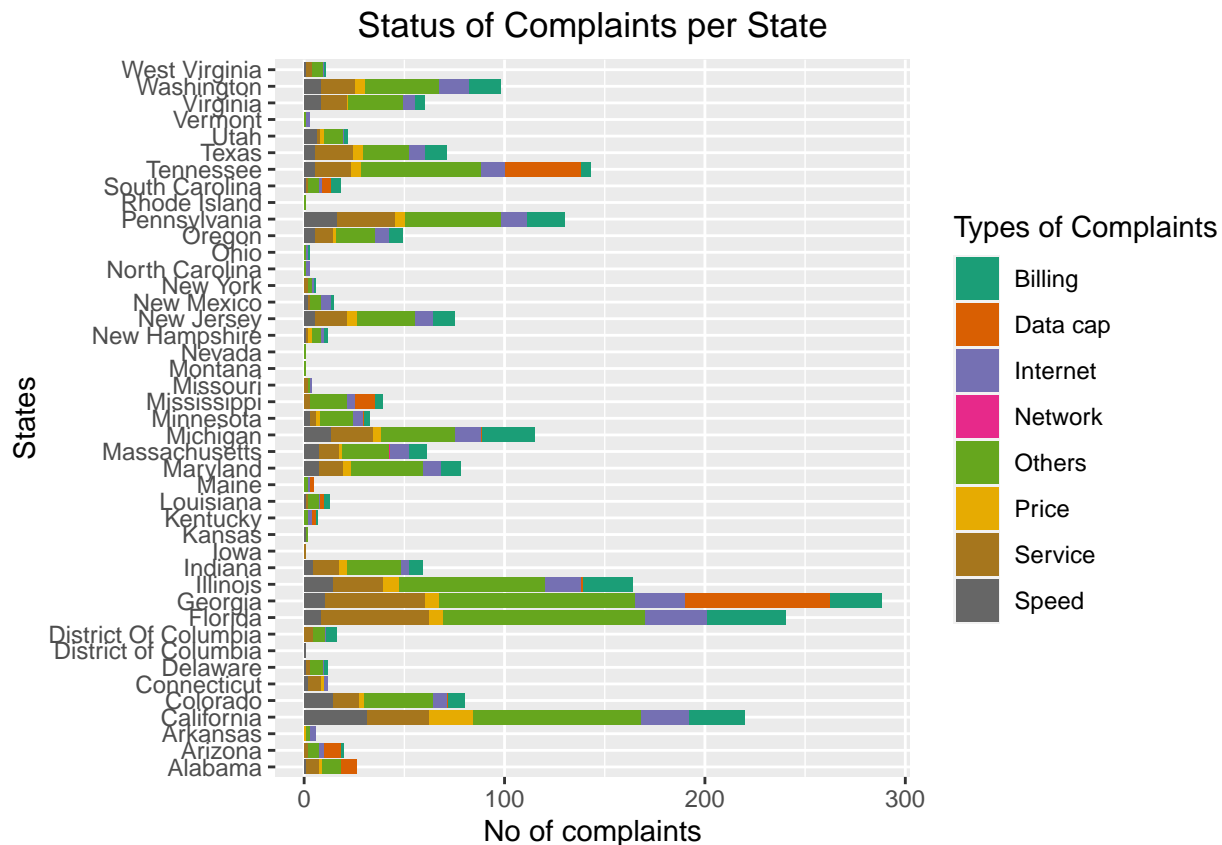
5. Which state has the maximum complaints

```
state_complaints <-
  data.frame(ComcastDF$State,
             ComcastDF$Complaint.Status,
             ComcastDF$Complaint.Type)

ggplot(state_complaints,
       aes(y = ComcastDF.State)) +
  geom_bar(aes(fill = ComcastDF.Complaint.Status)) +
  labs(title = "Status of Complaints per State",
       x = "No of complaints",
       y = "States",
       fill = "Status of Complaints") +
  scale_fill_manual(values=c("dark green", "red"))+
  theme(plot.title = element_text(hjust = 0.5))
```

```
ggplot(state_complaints,
  aes(y = ComcastDF.State)) +
  geom_bar(aes(fill = ComcastDF.Complaint.Type)) +
  labs(title = "Status of Complaints per State",
    x = "No of complaints",
    y = "States",
    fill = "Types of Complaints") +
  scale_fill_brewer(palette = "Dark2") +
  theme(plot.title = element_text(hjust = 0.5))
```



```
ComcastDF <- group_by(ComcastDF, State, Complaint.Status)

all_complaints <- summarise(ComcastDF, Count = n())

## 'summarise()' regrouping output by 'State' (override with '.groups' argument)

all_complaints[all_complaints$Count == max(all_complaints$Count), c(1, 3)]

## # A tibble: 1 x 2
## # Groups:   State [1]
##   State   Count
##   <chr>   <int>
## 1 Georgia    208

open_status <-
  summarise(ComcastDF, Count = n()) %>% filter(Complaint.Status == "Open")

## 'summarise()' regrouping output by 'State' (override with '.groups' argument)

open_status[open_status$Count == max(open_status$Count), c(1, 3)]

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

Observation

The state of Georgia has the highest amount of complaints - 208; and 80 of them still have the status “Open”

6. Which state has the highest percentage of unresolved complaints

```
open_status$Percent <- round(prop.table(open_status$Count)*100,2)
open_status[open_status$Percent == max(open_status$Percent), c(1, 4)]
```

```
## # A tibble: 1 x 2
## # Groups:   State [1]
##   State      Percent
##   <chr>      <dbl>
## 1 Georgia    15.5
```

Observation

The state of Georgia has the most unresolved tickets / complaints - with a percentage of 15.5%

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

```
ComcastDF$Received.Via <- as.factor(ComcastDF$Received.Via)
levels(ComcastDF$Received.Via)
```

```
## [1] "Customer Care Call" "Internet"
```

```
tab <- table(ComcastDF$Received.Via, ComcastDF$Complaint.Status)
round(prop.table(tab, 1)*100,2)
```

```
##
##               Closed  Open
## Customer Care Call  77.21 22.79
## Internet           76.29 23.71
```

Analysis

As per the above analysis we can conclude the following:

1. Comcast had a peak of received complaints in the 2nd part of June
2. Most complaints are regarding Comcast's service
3. The state of Georgia has the most number of complaints
4. Based on the channel of complaint we can conclude that Comcast has resolved 78% of the tickets received via *Customer Care Call* and 76% of the tickets received via *Internet*