Comcast_Data.R | comcast_data

Source on Save | Run | Source

```r
setwd("C:/Users/prave/Downloads/")
getwd()

#importing dataset
comcast_data <- read.csv("Comcast Telecom Complaints data.csv")

#importing packagaes
library(dplyr)
library(ggplot2)
library(lubridate)

#Manipulating field names
names(comcast_data)
names(comcast_data) <- gsub(pattern='//.',replacement="",x=names(comcast_data))
names(comcast_data)
View(comcast_data)

#Processing Date
comcast_data$Date <- dmy(comcast_data$Date)
View(comcast_data)

#The trend chart for the number of complaints at monthly and daily granularity levels
#Complaints on daily level basis
ans <- comcast_data %>% group_by(Date)%>% summarise(Numberofcomplaints=n())

#Trend chart for daily granularity level
ggplot(data=ans,aes(as.POSIXct(Date),Numberofcomplaints))+
  geom_line(color="green")+
  geom_point(size=1,color="red")+
  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))

#For monthly
ans1 <- comcast_data %>% group_by(Month=as.integer(month(Date)))%>% summarise(Numberofcomplaints=n())

ans1 <- na.omit(ans1)
#Trend Chart for monrhly granularity level
```

Comcast_Data.R | comcast_data

Source on Save | Run | Source

```r
#Trend Chart for monrhly granularity level
ggplot(data=ans1,aes(Month,Numberofcomplaints,label=Numberofcomplaints))+
         geom_line(color="green")+
         geom_point(size=1,color="red")+
         scale_x_continuous(breaks=ans1$Month)+
         labs(title="MONTHLY TICKET COUNT",x="MONTH",y="NUM OF TICKETS")+
         theme( plot.title = element_text(hjust=0.5))

#Table with frequency of complaint type
network_tickets <- contains(comcast_data$Customer.Complaint,match='network',ignore.case=T)
internet_tickets <- contains(comcast_data$Customer.Complaint,match='internet',ignore.case=T)
billing_tickets <- contains(comcast_data$Customer.Complaint,match='bill',ignore.case=T)
email_tickets <- contains(comcast_data$Customer.Complaint,match='email',ignore.case=T)
charges_tickets <- contains(comcast_data$Customer.Complaint,match='charges',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_tickets] <- "charges"

comcast_data$ComplaintType[-c(internet_tickets,network_tickets,billing_tickets,email_tickets,charges_tickets)] <- "

table(comcast_data$ComplaintType)

#from above table we can see that internet type complaints are maximum
#making a new categorical variable for complaint status

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'

#plotting a state wise status ofcomplaints in a stacked bar chart.
table(comcast_data$ComplaintStatus,comcast_data$State)

comcast_data <- group_by(comcast_data,State,ComplaintStatus)
chart_data <- summarise(comcast_data,count=n())
```

Comcast_Data.R | comcast_data

Source on Save | Run | Source

```r

#Bar chart
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")

#from above chart we can clearly see that Georgia has maximum complaints
#The highest percentage of resolved complaint
comcast_data %>% filter(ComplaintStatus=='Open') %>% group_by(State) %>% summarise(Numberofcomplaints=n())

tot <- comcast_data %>% group_by(ComplaintStatus) %>% summarise(Numberofcomplaints=n())
tot

slices <- tot$Numberofcomplaints
pct <- round(slices/sum(slices)*100,2)
lbls <- paste(tot$ComplaintStatus," ",pct,"%",sep="")

#plotting pie chart
pie(slices,labels=lbls)

#from the pie chart we can clearly see that 76.75% complaints are being resolved
int <- comcast_data %>% filter(Received.Via=="Internet",ComplaintStatus=="Closed") %>% group_by(Received.Via,Compla
ccc <- comcast_data %>% filter(Received.Via=="Customer Care Call",ComplaintStatus=="Closed") %>% group_by(Received.

#percentage of resolved internet complaint
intpct <- round(int$Numberofcomplaints/sum(tot$Numberofcomplaints)*100,2)
intpct

#percentage of resolved internet complaint
cccpct <- round(ccc$Numberofcomplaints/sum(tot$Numberofcomplaints)*100,2)
cccpct



```

Console | Terminal | Render | Background Jobs

R 4.3.2 · C:/Users/prave/Downloads/

```
> setwd("C:/Users/prave/Downloads/")
> getwd()
[1] "C:/Users/prave/Downloads"
>
> #importing dataset
> comcast_data <- read.csv("Comcast Telecom Complaints data.csv")
>
> #importing packagaes
> 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(lubridate)

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union

>
> #Manipulating field names
> names(comcast_data)
 [1] "Ticket.."                  "Customer.Complaint"        "Date"
 [4] "Time"                      "Received.Via"              "City"
 [7] "State"                     "Zip.code"                  "Status"
[10] "Filing.on.Behalf.of.Someone"
> names(comcast_data) <- gsub(pattern='//.',replacement="",x=names(comcast_data))
> names(comcast_data)
 [1] "Ticket.."                  "Customer.Complaint"        "Date"
 [4] "Time"                      "Received.Via"              "City"
```

Console | Terminal | Render | Background Jobs

R 4.3.2 · C:/Users/prave/Downloads/

```
 [7] "State"                         "Zip.code"                      "Status"
[10] "Filing.on.Behalf.of.Someone"
> View(comcast_data)
>
> #Processing Date
> comcast_data$Date <- dmy(comcast_data$Date)
> View(comcast_data)
>
> #The trend chart for the number of complaints at monthly and daily granularity levels
> #Complaints on daily level basis
> ans <- comcast_data %>% group_by(Date)%>% summarise(Numberofcomplaints=n())
>
> #Trend chart for daily granularity level
> ggplot(data=ans,aes(as.POSIXct(Date),Numberofcomplaints))+
+   geom_line(color="green")+
+   geom_point(size=1,color="red")+
+   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))
>
> #For monthly
> ans1 <- comcast_data %>% group_by(Month=as.integer(month(Date)))%>% summarise(Numberofcomplaints=n())
>
> ans1 <- na.omit(ans1)
> #Trend Chart for monrhly granularity level
> ggplot(data=ans1,aes(Month,Numberofcomplaints,label=Numberofcomplaints))+
+          geom_line(color="green")+
+          geom_point(size=1,color="red")+
+          scale_x_continuous(breaks=ans1$Month)+
+          labs(title="MONTHLY TICKET COUNT",x="MONTH",y="NUM OF TICKETS")+
+          theme( plot.title = element_text(hjust=0.5))
>
> #Table with frequency of complaint type
> network_tickets <- contains(comcast_data$Customer.Complaint,match='network',ignore.case=T)
> internet_tickets <- contains(comcast_data$Customer.Complaint,match='internet',ignore.case=T)
> billing_tickets <- contains(comcast_data$Customer.Complaint,match='bill',ignore.case=T)
> email_tickets <- contains(comcast_data$Customer.Complaint,match='email',ignore.case=T)
> charges_tickets <- contains(comcast_data$Customer.Complaint,match='charges',ignore.case=T)
>
> comcast_data$ComplaintType[internet_tickets] <- "internet"
```

Console | Terminal | Render | Background Jobs

R 4.3.2 · C:/Users/prave/Downloads/

```
> comcast_data$ComplaintType[-c(internet_tickets,network_tickets,billing_tickets,email_tickets,charges_tickets)] <- "other
s"
>
> table(comcast_data$ComplaintType)

 billing  charges    email internet  network   others
     370       77       16      479        1     1281
>
> #from above table we can see that internet type complaints are maximum
> #making a new categorical variable for complaint status
>
> 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'
>
> #plotting a state wise status ofcomplaints in a stacked bar chart.
> table(comcast_data$ComplaintStatus,comcast_data$State)

       Alabama Arizona Arkansas California Colorado Connecticut Delaware District of Columbia District Of Columbia
  Closed      17      14        6        159       58           9        8                    1                   14
  Open         9       6        0         61       22           3        4                    0                    2

       Florida Georgia Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Maryland Massachusetts Michigan
  Closed     201     208      135      50    1      1        4        12     3       63            50       92
  Open        39      80       29       9    0      1        3         1     2       15            11       23

       Minnesota Mississippi Missouri Montana Nevada New Hampshire New Jersey New Mexico New York North Carolina
  Closed        29          23        3       1      1             8         56         11        6              3
  Open           4          16        1       0      0             4         19          4        0              0

       Ohio Oregon Pennsylvania Rhode Island South Carolina Tennessee Texas Utah Vermont Virginia Washington
  Closed    3     36          110            1             15        96    49   16       2       49         75
  Open      0     13           20            0              3        47    22    6       1       11         23

       West Virginia
  Closed             8
  Open               3
>
> comcast_data <- group_by(comcast_data,State,ComplaintStatus)
> chart_data <- summarise(comcast_data,count=n())
```

Console | Terminal | Render | Background Jobs

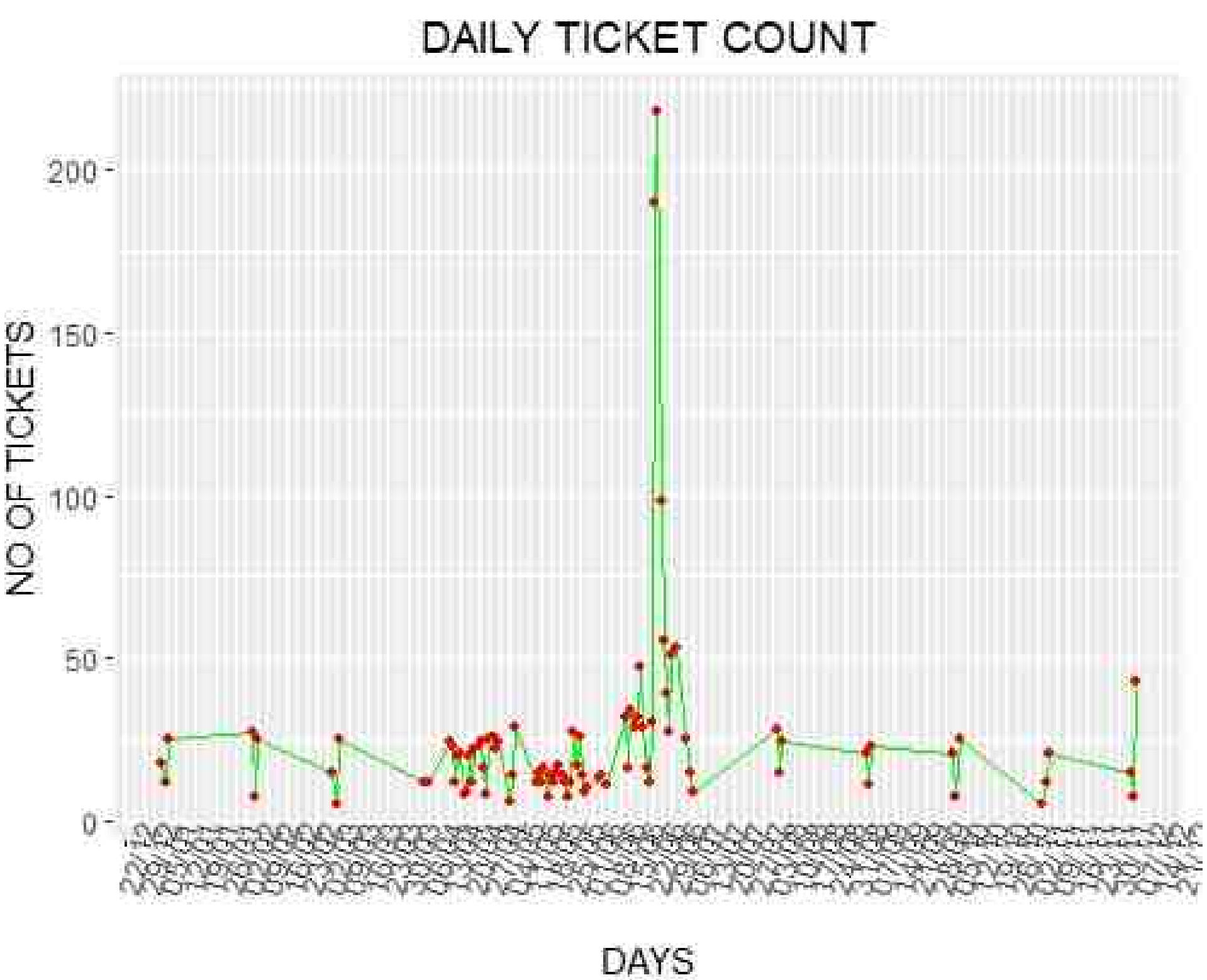
R 4.3.2 · C:/Users/prave/Downloads/

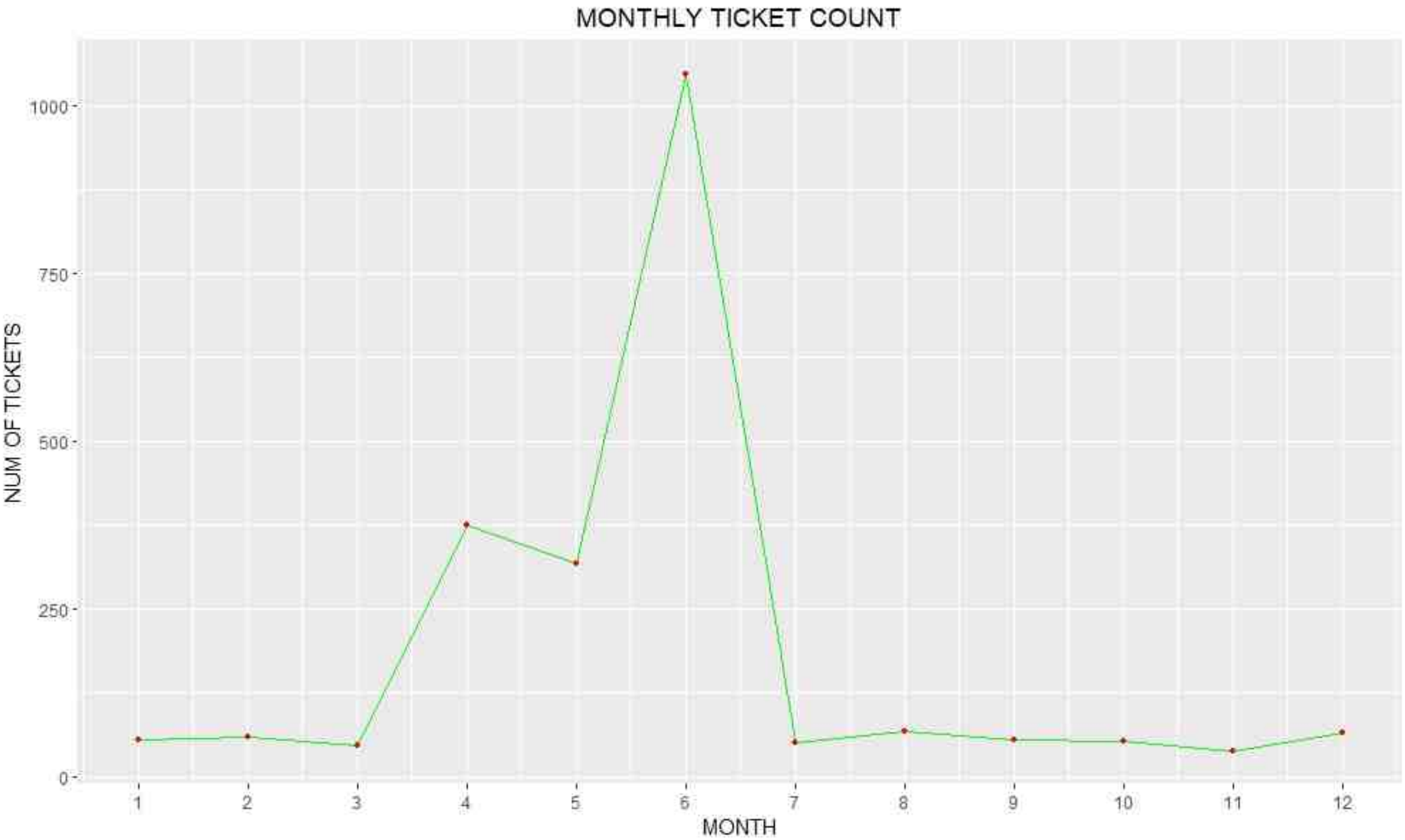
```
`summarise()` has grouped output by 'State'. You can override using the `.groups` argument.
>
> #Bar chart
> ggplot(as.data.frame(chart_data),mapping=aes(State,count))+
+   geom_col(aes(fill=ComplaintStatus),width=0.95)+
+   theme(axis.text.x=element_text(angle=90),
+         axis.title.y=element_text(size=15),
+         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")
>
> #from above chart we can clearly see that Georgia has maximum complaints
> #The highest percentage of resolved complaint
> comcast_data %>% filter(ComplaintStatus=='Open') %>% group_by(State) %>% summarise(Numberofcomplaints=n())
# A tibble: 34 × 2
   State                Numberofcomplaints
   <chr>                             <int>
 1 Alabama                               9
 2 Arizona                               6
 3 California                           61
 4 Colorado                             22
 5 Connecticut                           3
 6 Delaware                              4
 7 District Of Columbia                  2
 8 Florida                              39
 9 Georgia                              80
10 Illinois                             29
# i 24 more rows
# i Use `print(n = ...)` to see more rows
>
> tot <- comcast_data %>% group_by(ComplaintStatus) %>% summarise(Numberofcomplaints=n())
> tot
# A tibble: 2 × 2
  ComplaintStatus Numberofcomplaints
  <chr>                        <int>
1 Closed                        1707
2 Open                           517
>
> slices <- tot$Numberofcomplaints
> pct <- round(slices/sum(slices)*100,2)
```

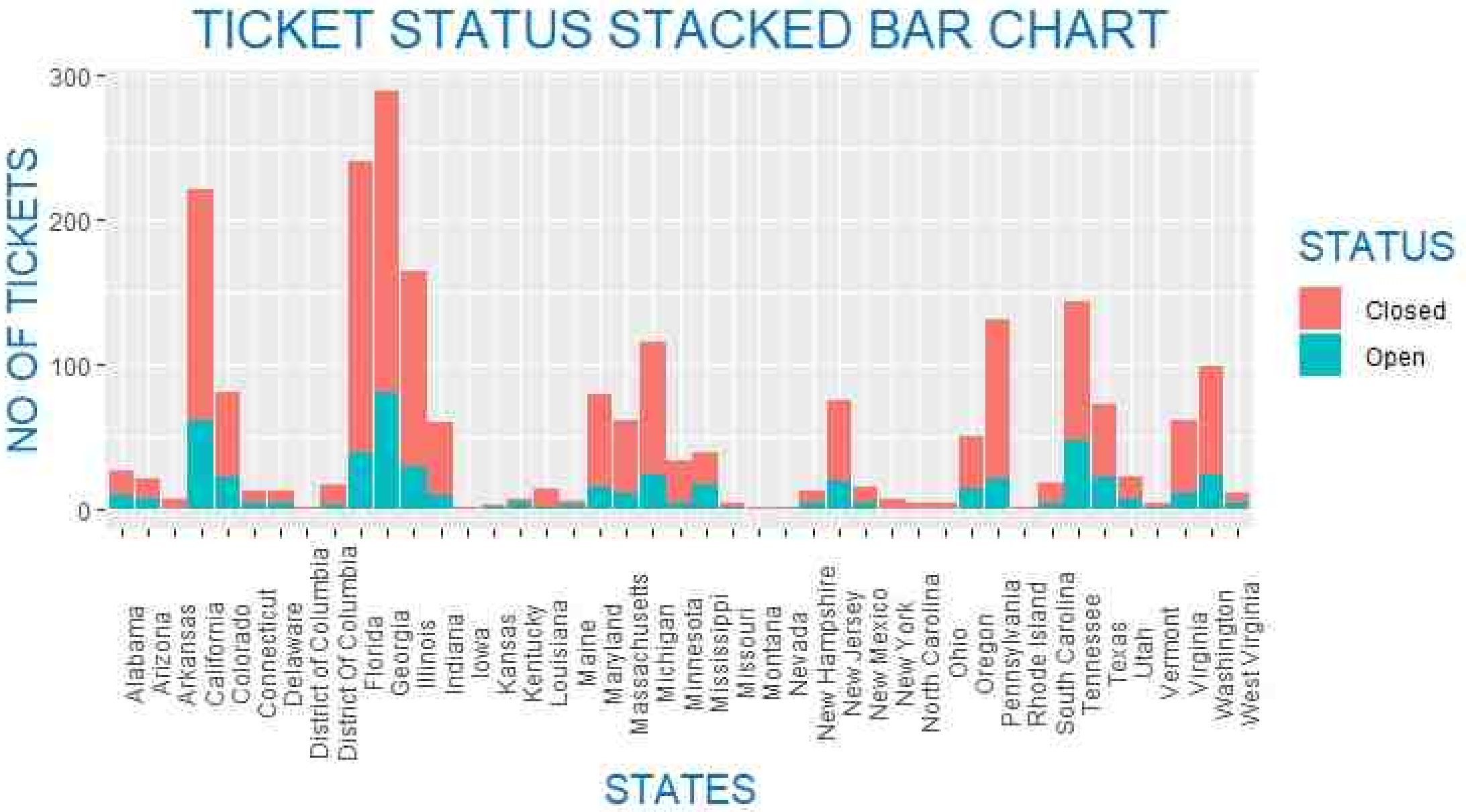
Console | Terminal | Render | Background Jobs

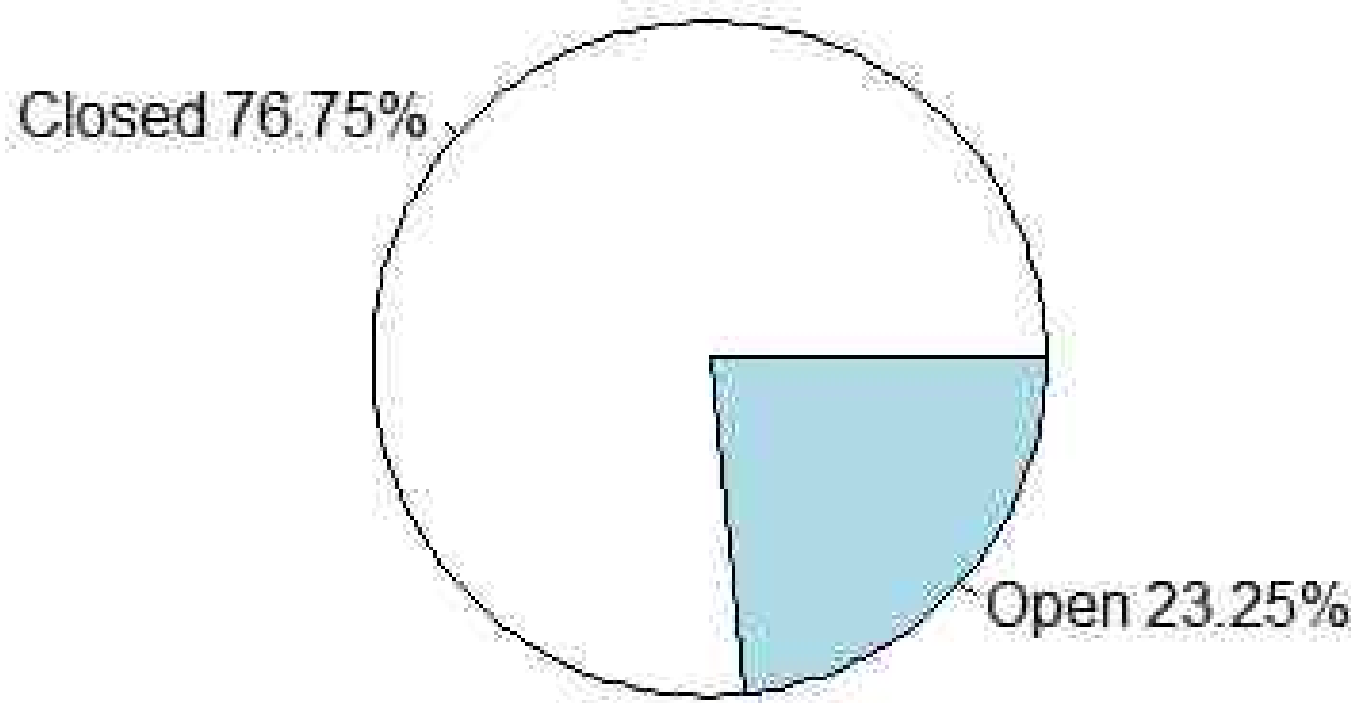
R 4.3.2 · C:/Users/prave/Downloads/

```
 7 District Of Columbia                2
 8 Florida                            39
 9 Georgia                            80
10 Illinois                           29
# i 24 more rows
# i Use `print(n = ...)` to see more rows
>
> tot <- comcast_data %>% group_by(ComplaintStatus) %>% summarise(Numberofcomplaints=n())
> tot
# A tibble: 2 x 2
  ComplaintStatus Numberofcomplaints
  <chr>                        <int>
1 Closed                        1707
2 Open                           517
>
> slices <- tot$Numberofcomplaints
> pct <- round(slices/sum(slices)*100,2)
> lbls <- paste(tot$ComplaintStatus," ",pct,"%",sep="")
>
> #plotting pie chart
> pie(slices,labels=lbls)
>
> #from the pie chart we can clearly see that 76.75% complaints are being resolved
> int <- comcast_data %>% filter(Received.Via=="Internet",ComplaintStatus=="Closed") %>% group_by(Received.Via,ComplaintSt
atus) %>% summarise(Numberofcomplaints=n())
`summarise()` has grouped output by 'Received.Via'. You can override using the `.groups` argument.
> ccc <- comcast_data %>% filter(Received.Via=="Customer Care Call",ComplaintStatus=="Closed") %>% group_by(Received.Via,C
omplaintStatus) %>% summarise(Numberofcomplaints=n())
`summarise()` has grouped output by 'Received.Via'. You can override using the `.groups` argument.
>
> #percentage of resolved internet complaint
> intpct <- round(int$Numberofcomplaints/sum(tot$Numberofcomplaints)*100,2)
> intpct
[1] 37.9
>
> #percentage of resolved internet complaint
> cccpct <- round(ccc$Numberofcomplaints/sum(tot$Numberofcomplaints)*100,2)
> cccpct
[1] 38.85
>
```


DAILY TICKET COUNT
NO OF TICKETS
200
150
100
50
0
DAYS


MONTHLY TICKET COUNT
NUM OF TICKETS
1000
750
500
250
0
1
2
3
4
5
6
7
8
9
10
11
12
MONTH


TICKET STATUS STACKED BAR CHART
NO OF TICKETS
300
200
100
0
STATUS
Closed
Open
Alabama
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
District Of Columbia
Florida
Georgia
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Ohio
Oregon
Pennsylvania
Rhode Island
South Carolina
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
STATES


Closed 76.75%
Open 23.25%