Assignment 11

Code

Yiman Liu

1. Follow this tutorial on MongoDB (Links to an external site.) and build an R Notebook that implements all of the steps.

Hide

```
library(ggplot2)
library(dplyr)
library(maps)
library(ggmap)
library(mongolite)
library(lubridate)
library(gridExtra)

crimes=data.table::fread("Crimes_-_2001_to_Present.csv")
```

```
|----------------------------------------------------|
|====================================================|
```

Hide

```
names(crimes)
```

```
 [1] "ID"                    "Case Number"           "Date"                  "Block"
 [5] "IUCR"                  "Primary Type"          "Description"           "Location Desc

ription"
 [9] "Arrest"                "Domestic"              "Beat"                  "District"
[13] "Ward"                  "Community Area"        "FBI Code"              "X Coordinate"
[17] "Y Coordinate"          "Year"                  "Updated On"            "Latitude"
[21] "Longitude"             "Location"
```

Hide

```
names(crimes) = gsub(" ","",names(crimes))
names(crimes)
```

```
 [1] "ID"                    "CaseNumber"            "Date"                  "Block"
 [5] "IUCR"                  "PrimaryType"           "Description"           "LocationDescript
ion"
 [9] "Arrest"                "Domestic"              "Beat"                  "District"
[13] "Ward"                  "CommunityArea"         "FBICode"               "XCoordinate"
[17] "YCoordinate"           "Year"                  "UpdatedOn"             "Latitude"
[21] "Longitude"             "Location"
```

Hide

```
my_collection = mongo(collection = "crimes", db = "Chicago") # create connection, databa
se and collection
my_collection$insert(crimes)
```

```
List of 5
 $ nInserted  : num 7242630
 $ nMatched   : num 0
 $ nRemoved   : num 0
 $ nUpserted  : num 0
 $ writeErrors: list()
```

Hide

```
my_collection$count()
```

```
[1] 7242630
```

Hide

```
my_collection$iterate()$one()
```

```
$ID
[1] 11034701

$CaseNumber
[1] "JA366925"

$Date
[1] "01/01/2001 11:00:00 AM"

$Block
[1] "016XX E 86TH PL"

$IUCR
[1] "1153"

$PrimaryType
[1] "DECEPTIVE PRACTICE"

$Description
[1] "FINANCIAL IDENTITY THEFT OVER $ 300"

$LocationDescription
[1] "RESIDENCE"

$Arrest
[1] FALSE

$Domestic
[1] FALSE

$Beat
[1] 412

$District
[1] 4

$Ward
[1] 8

$CommunityArea
[1] 45

$FBICode
[1] "11"

$Year
[1] 2001

$UpdatedOn
[1] "08/05/2017 03:50:08 PM"

$Location
[1] ""
```

Hide

```
length(my_collection$distinct("PrimaryType"))
```

```
[1] 36
```

Hide

```
my_collection$count('{"PrimaryType" : "ASSAULT", "Domestic" : true }')
```

```
[1] 102162
```

Hide

```
query1= my_collection$find('{"PrimaryType" : "ASSAULT", "Domestic" : true }')
query2= my_collection$find('{"PrimaryType" : "ASSAULT", "Domestic" : true }',
                           fields = '{"_id":0, "PrimaryType":1, "Domestic":1}')
ncol(query1) # with all the columns
```

```
[1] 22
```

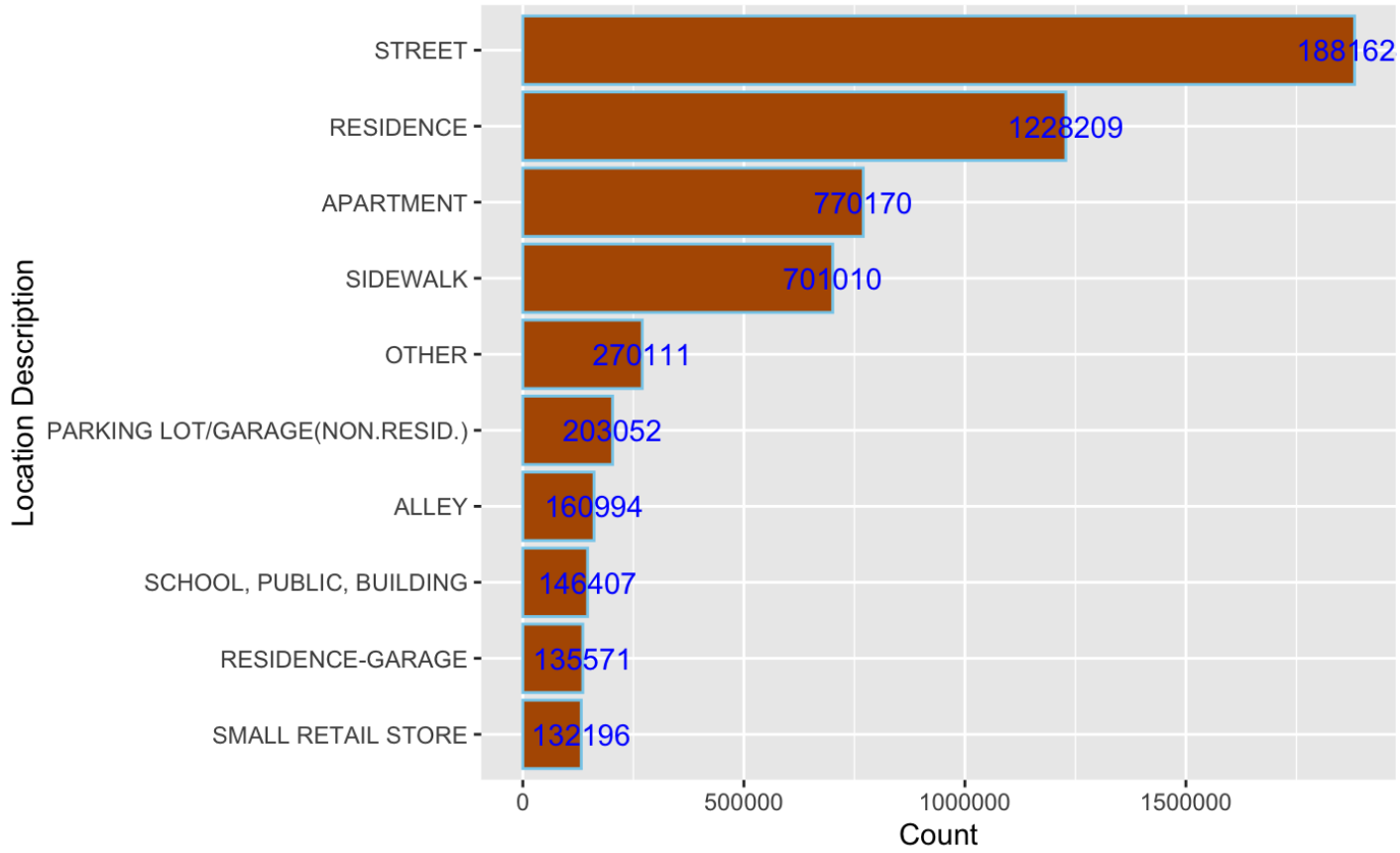
Hide

```
ncol(query2) # only the selected columns
```

```
[1] 2
```

Hide

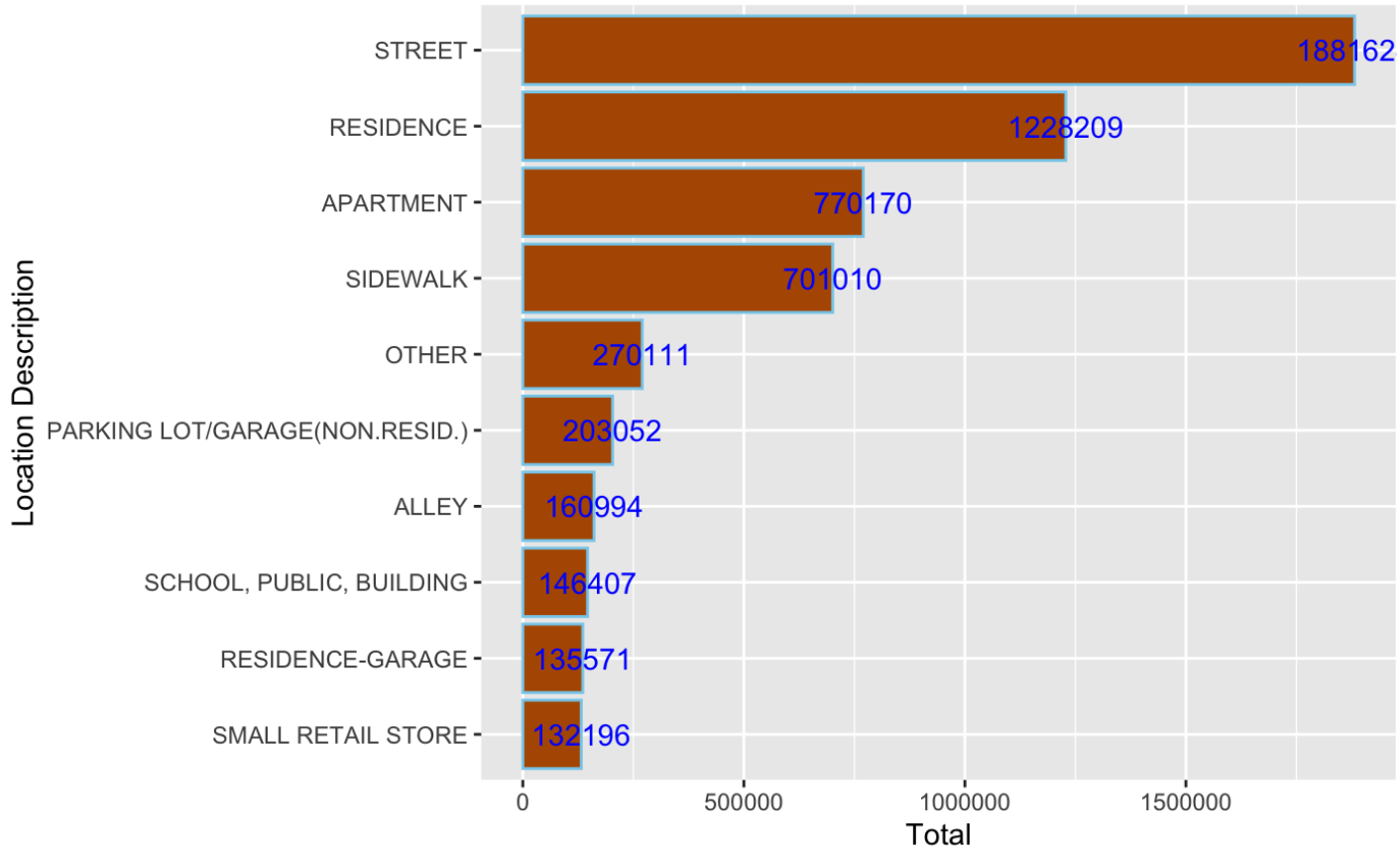
```
my_collection$aggregate('[{"$group":{"_id":"$LocationDescription", "Count": {"$sum":
1}}}]')%>%na.omit()%>%
arrange(desc(Count))%>%head(10)%>%
ggplot(aes(x=reorder(`_id`,Count),y=Count))+
geom_bar(stat="identity",color='skyblue',fill='#b35900')+geom_text(aes(label = Count), c
olor = "blue") +coord_flip()+xlab("Location Description")
```



Hide

```
crimes%>%group_by(`LocationDescription`)%>%summarise(Total=n())%>% arrange(desc(Total))%
>%head(10)%>%
ggplot(aes(x=reorder(`LocationDescription`,Total),y=Total))+
geom_bar(stat="identity",color='skyblue',fill='#b35900')+geom_text(aes(label = Total), c
olor = "blue") +coord_flip()+xlab("Location Description")
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

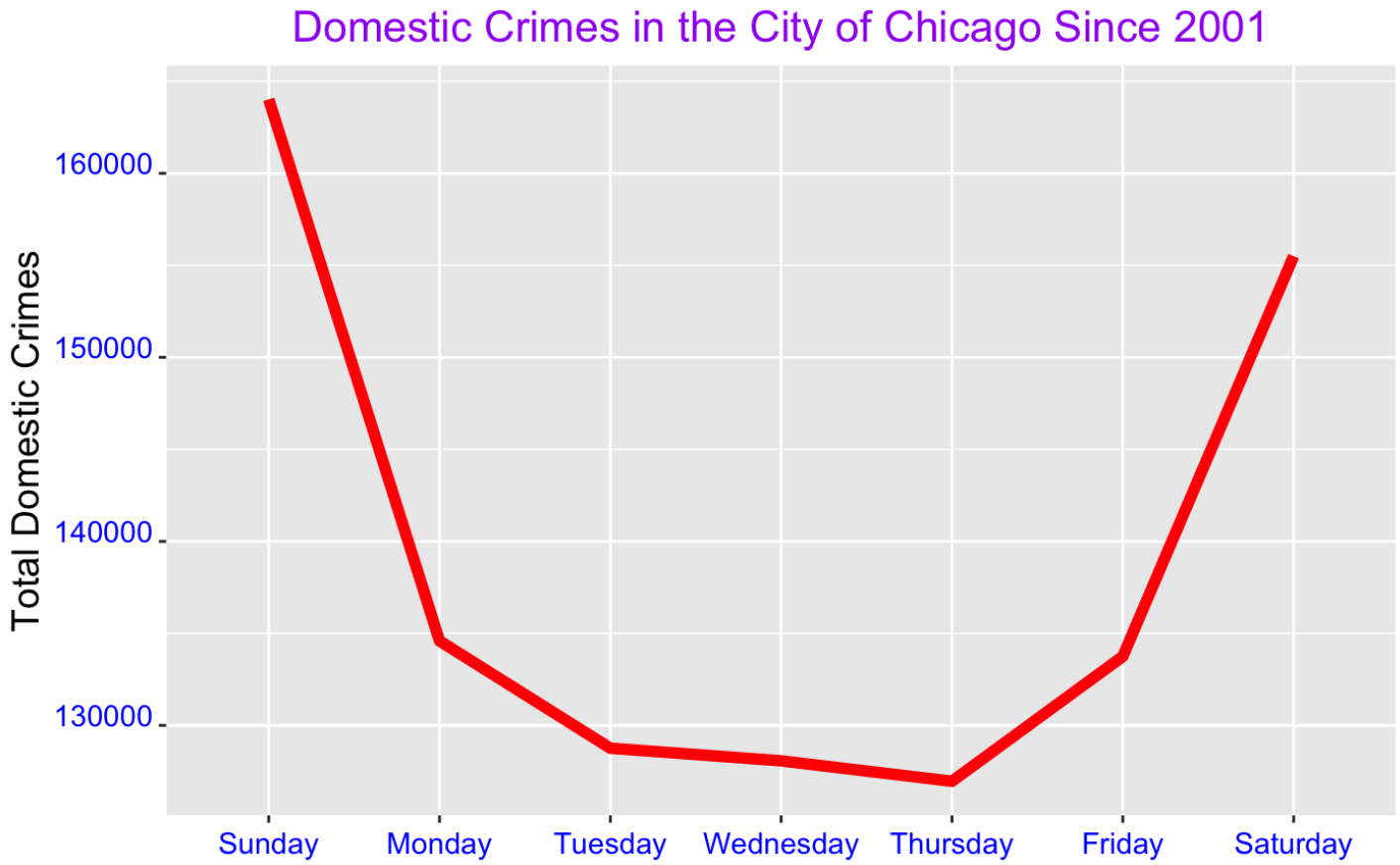


Hide

```
query3= my_collection$find('{}', fields = '{"_id":0, "Latitude":1, "Longitude":1,"Year":
1}')
```

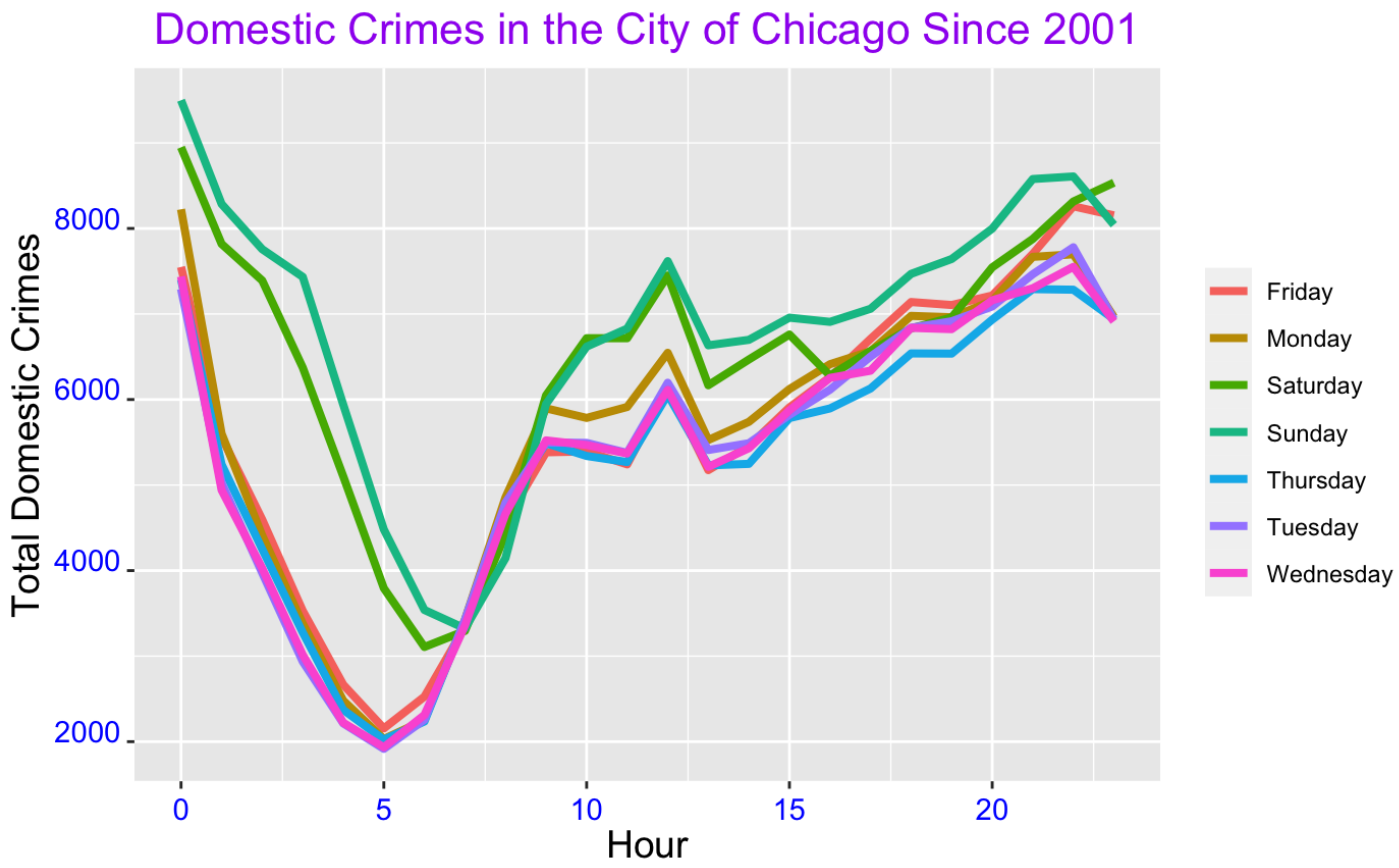
Hide

```
domestic=my_collection$find('{"Domestic":true}', fields = '{"_id":0, "Domestic":1,"Dat
e":1}')
domestic$Date= mdy_hms(domestic$Date)
domestic$Weekday = weekdays(domestic$Date)
domestic$Hour = hour(domestic$Date)
domestic$month = month(domestic$Date)

WeekdayCounts = as.data.frame(table(domestic$Weekday))
WeekdayCounts$Var1 = factor(WeekdayCounts$Var1, ordered=TRUE, levels=c("Sunday", "Monda
y", "Tuesday", "Wednesday", "Thursday", "Friday","Saturday"))
ggplot(WeekdayCounts, aes(x=Var1, y=Freq)) + geom_line(aes(group=1),size=2,color="red")
 + xlab("Day of the Week") + ylab("Total Domestic Crimes")+
ggtitle("Domestic Crimes in the City of Chicago Since 2001")+
theme(axis.title.x=element_blank(),axis.text.y = element_text(color="blue",size=11,angle
=0,hjust=1,vjust=0),
      axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
     axis.title.y = element_text(size=14),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
```



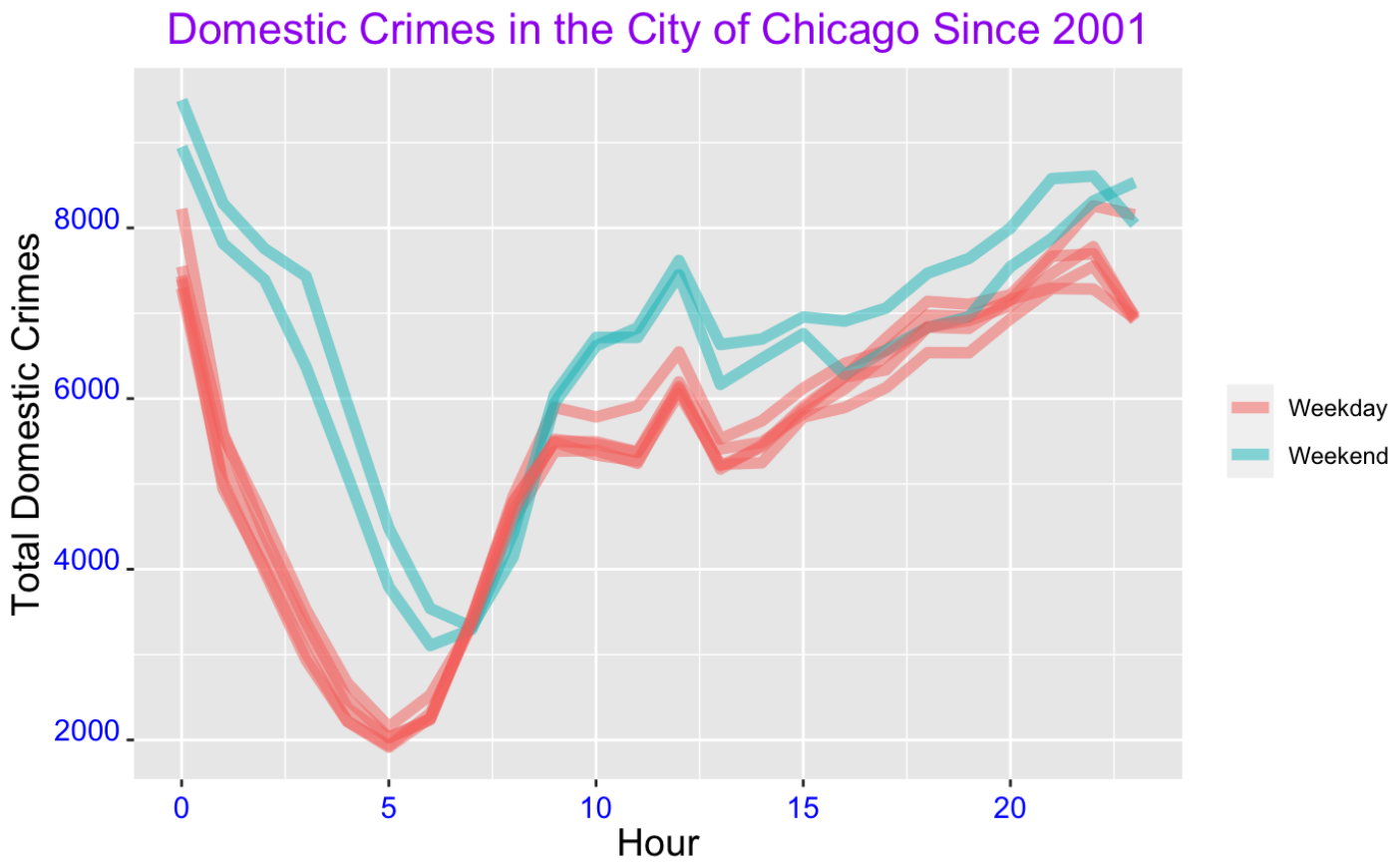
Hide

```
DayHourCounts = as.data.frame(table(domestic$Weekday, domestic$Hour))
DayHourCounts$Hour = as.numeric(as.character(DayHourCounts$Var2))
ggplot(DayHourCounts, aes(x=Hour, y=Freq)) + geom_line(aes(group=Var1, color=Var1), size
=1.4)+ylab("Count")+
ylab("Total Domestic Crimes")+ggtitle("Domestic Crimes in the City of Chicago Since 200
1")+
theme(axis.title.x=element_text(size=14),axis.text.y = element_text(color="blue",size=11
,angle=0,hjust=1,vjust=0),
      axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
     axis.title.y = element_text(size=14),
     legend.title=element_blank(),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
```



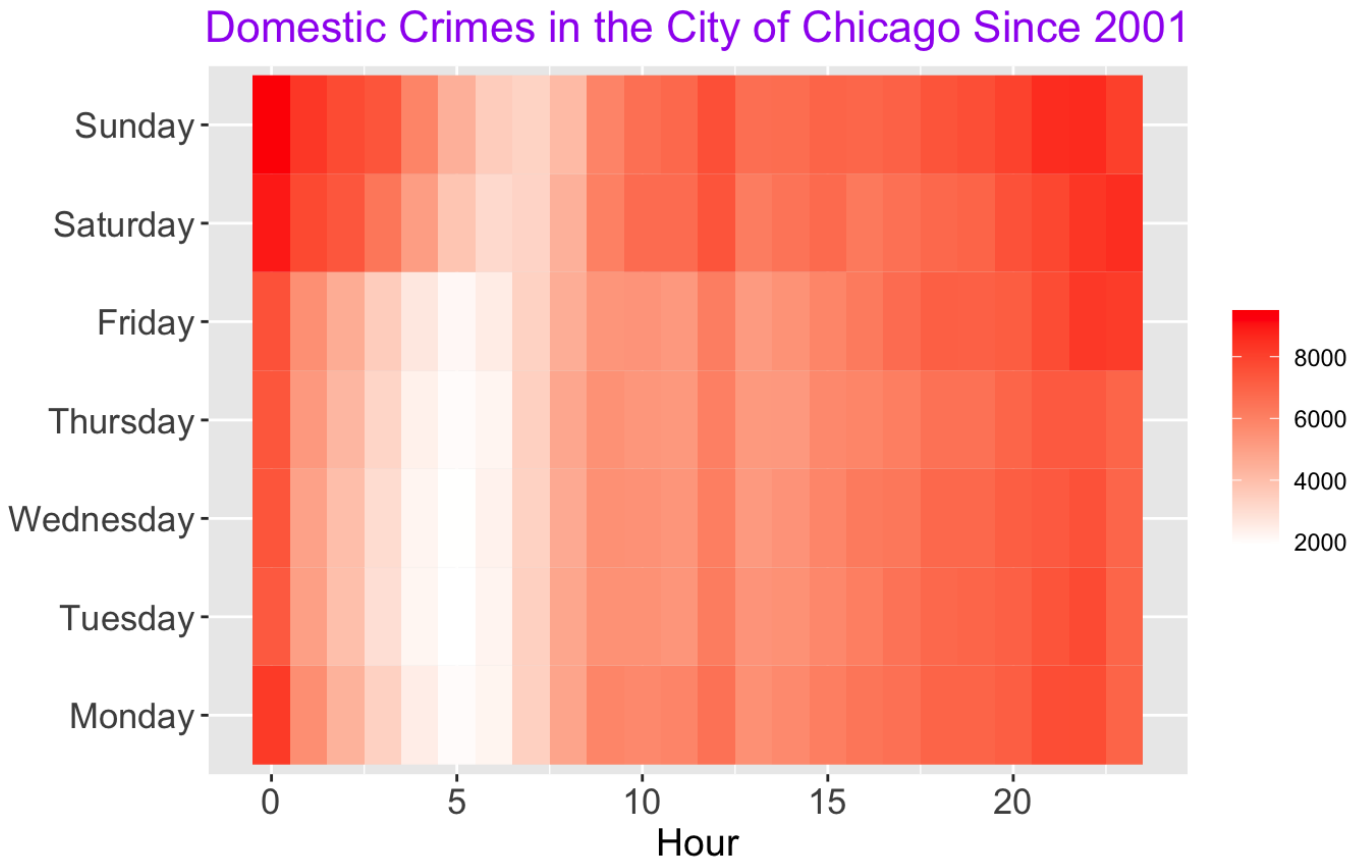
Hide

```
DayHourCounts$Type = ifelse((DayHourCounts$Var1 == "Sunday") | (DayHourCounts$Var1 == "S
aturday"), "Weekend", "Weekday")

ggplot(DayHourCounts, aes(x=Hour, y=Freq)) + geom_line(aes(group=Var1, color=Type), size
=2, alpha=0.5) +
ylab("Total Domestic Crimes")+ggtitle("Domestic Crimes in the City of Chicago Since 200
1")+
theme(axis.title.x=element_text(size=14),axis.text.y = element_text(color="blue",size=11
,angle=0,hjust=1,vjust=0),
      axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
     axis.title.y = element_text(size=14),
     legend.title=element_blank(),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
```



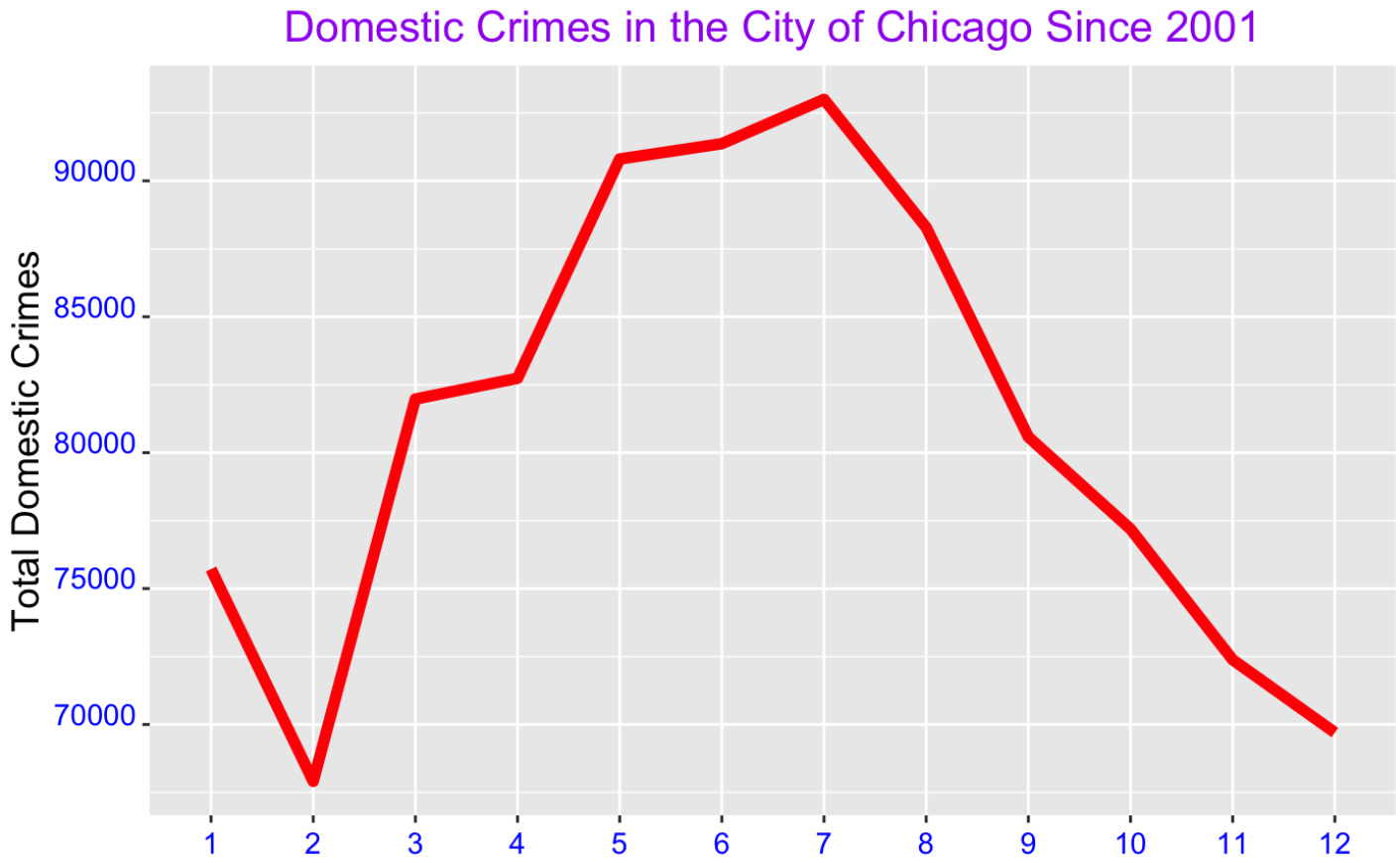
Hide

```
DayHourCounts$Var1 = factor(DayHourCounts$Var1, ordered=TRUE, levels=c("Monday", "Tuesda
y", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))
ggplot(DayHourCounts, aes(x = Hour, y = Var1)) + geom_tile(aes(fill = Freq)) + scale_fil
l_gradient(name="Total MV Thefts", low="white", high="red") +
ggtitle("Domestic Crimes in the City of Chicago Since 2001")+theme(axis.title.y = elemen
t_blank())+ylab("")+
theme(axis.title.x=element_text(size=14),axis.text.y = element_text(size=13),
      axis.text.x = element_text(size=13),
     axis.title.y = element_text(size=14),
     legend.title=element_blank(),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
```



Hide

```
monthCounts = as.data.frame(table(domestic$month))
ggplot(monthCounts, aes(x=Var1, y=Freq)) + geom_line(aes(group=1),size=2,color="red") +
 xlab("Day of the Week") + ylab("Total Domestic Crimes")+
ggtitle("Domestic Crimes in the City of Chicago Since 2001")+
theme(axis.title.x=element_blank(),axis.text.y = element_text(color="blue",size=11,angle
=0,hjust=1,vjust=0),
      axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
     axis.title.y = element_text(size=14),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
```



Hide

```
crimes=my_collection$find('{}', fields = '{"_id":0, "PrimaryType":1,"Year":1}')
crimes%>%group_by(PrimaryType)%>%summarize(Count=n())%>%arrange(desc(Count))%>%head(4)
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

PrimaryType <chr>	Count <int>
THEFT	1528017
BATTERY	1326755
CRIMINAL DAMAGE	824874
NARCOTICS	735228

4 rows

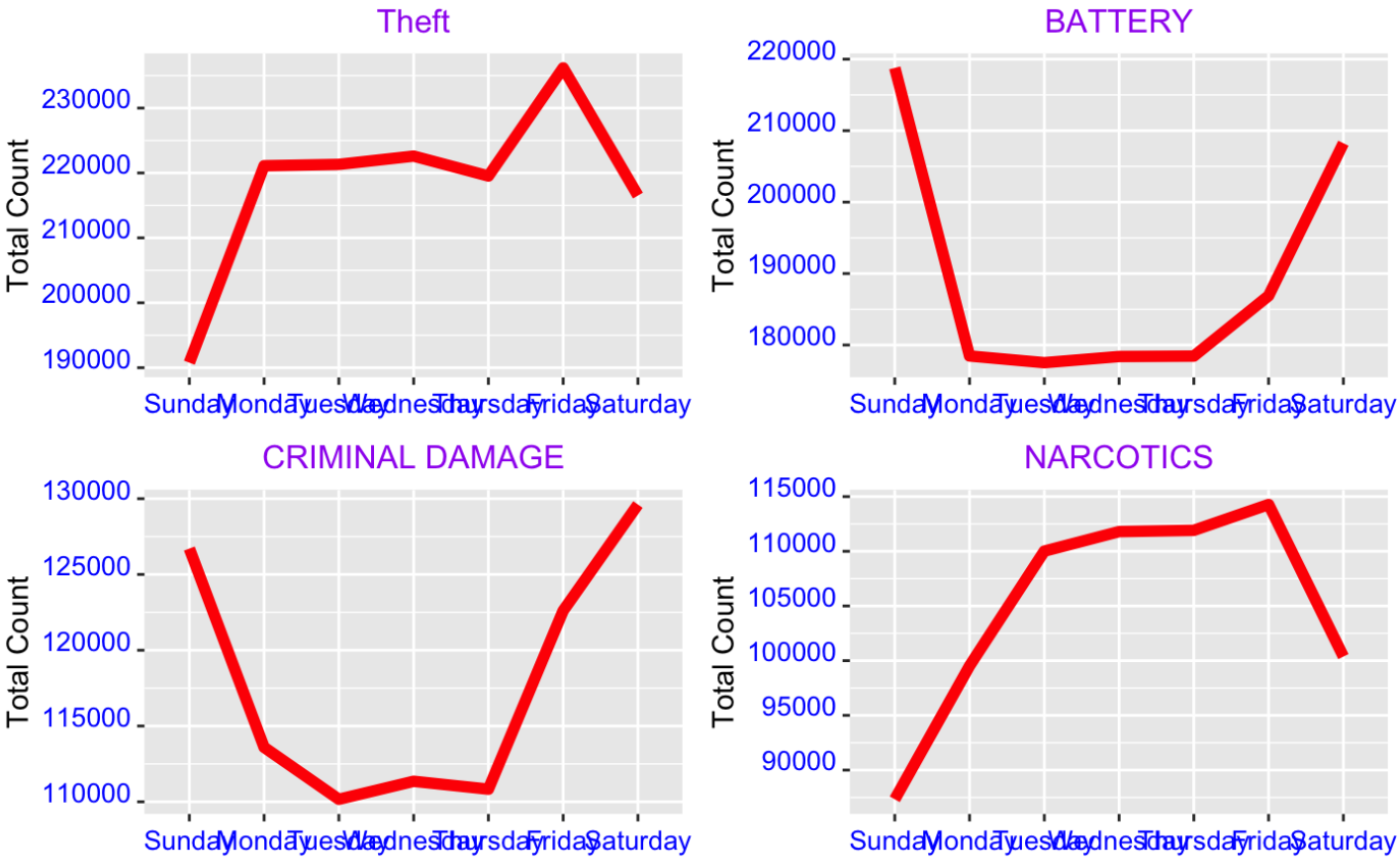
Hide

```
four_most_common=crimes%>%group_by(PrimaryType)%>%summarize(Count=n())%>%arrange(desc(Co
unt))%>%head(4)
```

```
`summarise()` ungrouping output (override with `.groups` argument)
```

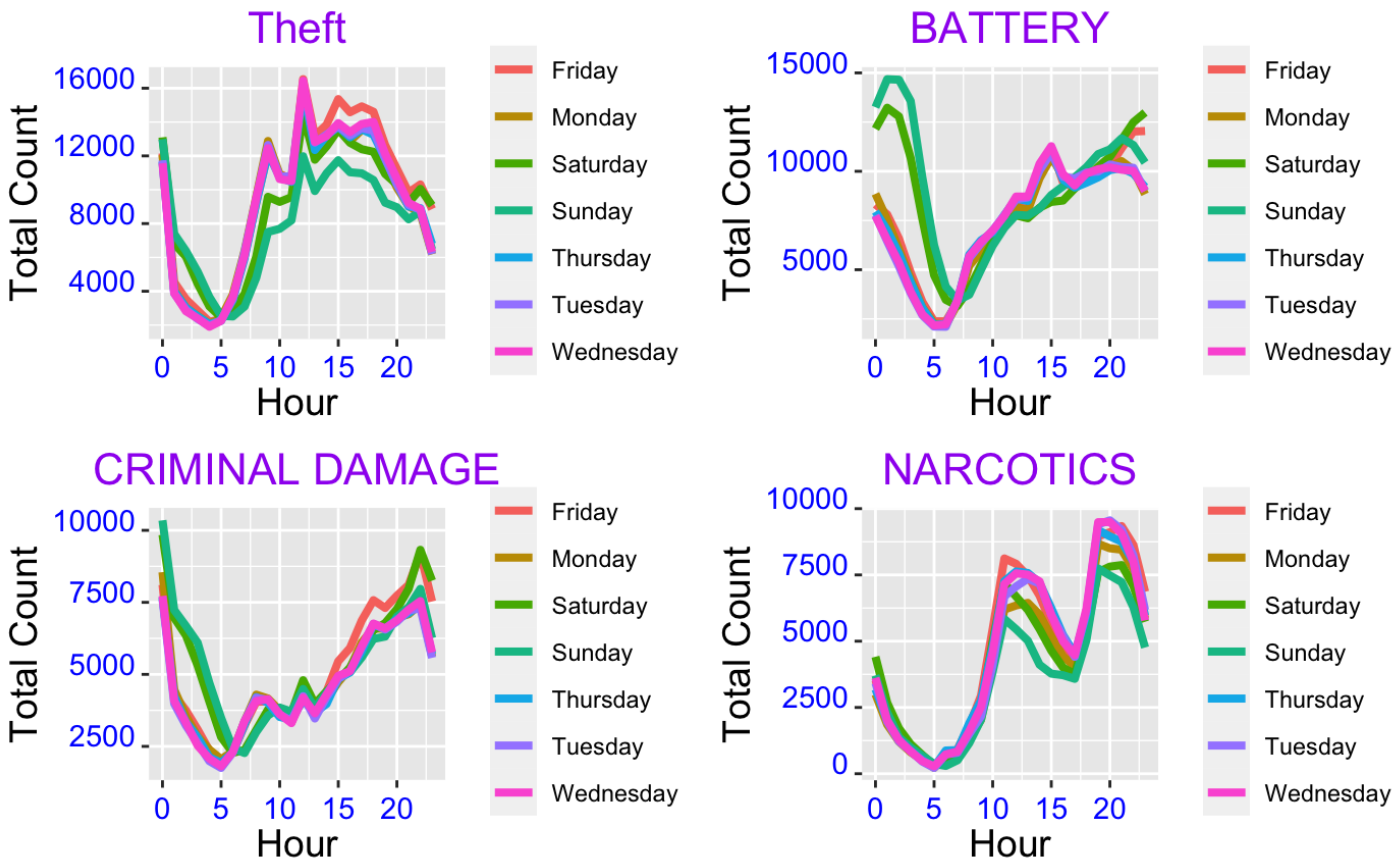
Hide

```
four_most_common=four_most_common$PrimaryType

crimes=my_collection$find('{}', fields = '{"_id":0, "PrimaryType":1,"Date":1}')
crimes=filter(crimes,PrimaryType %in%four_most_common)
crimes$Date= mdy_hms(crimes$Date)
crimes$Weekday = weekdays(crimes$Date)
crimes$Hour = hour(crimes$Date)
crimes$month=month(crimes$Date)

g = function(data){
    WeekdayCounts = as.data.frame(table(data$Weekday))
    WeekdayCounts$Var1 = factor(WeekdayCounts$Var1, ordered=TRUE, levels=c("Sunday", "Mo
nday", "Tuesday", "Wednesday", "Thursday", "Friday","Saturday"))
    ggplot(WeekdayCounts, aes(x=Var1, y=Freq)) + geom_line(aes(group=1),size=2,color="re
d") + xlab("Day of the Week") +
    theme(axis.title.x=element_blank(),axis.text.y = element_text(color="blue",size=10,a
ngle=0,hjust=1,vjust=0),
          axis.text.x = element_text(color="blue",size=10,angle=0,hjust=.5,vjust=.5),
         axis.title.y = element_text(size=11),
         plot.title=element_text(size=12,color="purple",hjust=0.5))
    }

g1=g(filter(crimes,PrimaryType=="THEFT"))+ggtitle("Theft")+ylab("Total Count")
g2=g(filter(crimes,PrimaryType=="BATTERY"))+ggtitle("BATTERY")+ylab("Total Count")
g3=g(filter(crimes,PrimaryType=="CRIMINAL DAMAGE"))+ggtitle("CRIMINAL DAMAGE")+ylab("Tot
al Count")
g4=g(filter(crimes,PrimaryType=="NARCOTICS"))+ggtitle("NARCOTICS")+ylab("Total Count")
grid.arrange(g1,g2,g3,g4,ncol=2)
```



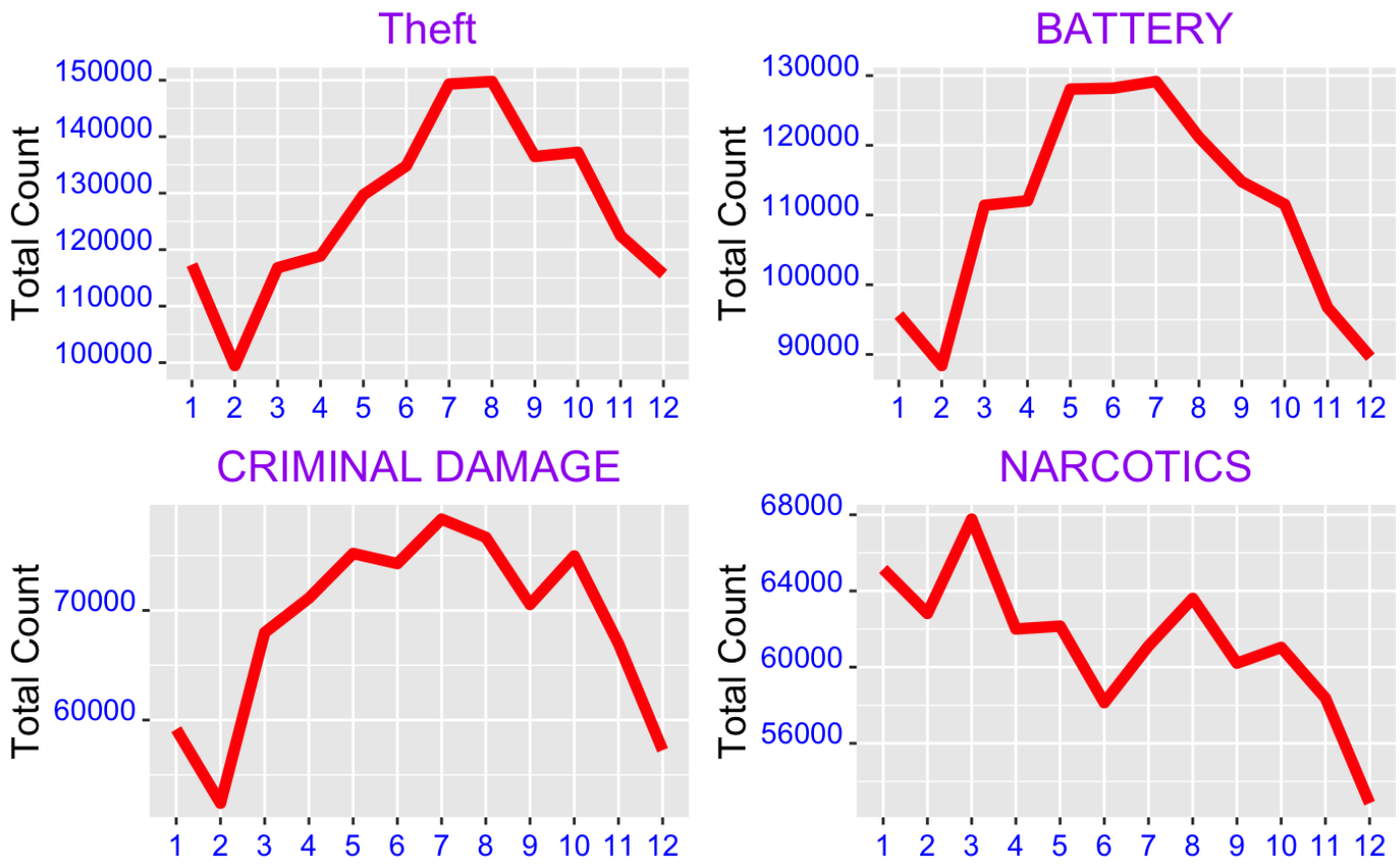
Hide

```
g=function(data){
    DayHourCounts = as.data.frame(table(data$Weekday, data$Hour))
    DayHourCounts$Hour = as.numeric(as.character(DayHourCounts$Var2))
    ggplot(DayHourCounts, aes(x=Hour, y=Freq)) + geom_line(aes(group=Var1, color=Var1),
 size=1.4)+ylab("Count")+
    theme(axis.title.x=element_text(size=14),axis.text.y = element_text(color="blue",siz
e=11,angle=0,hjust=1,vjust=0),
      axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
     axis.title.y = element_text(size=14),
     legend.title=element_blank(),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
    }

g1=g(filter(crimes,PrimaryType=="THEFT"))+ggtitle("Theft")+ylab("Total Count")
g2=g(filter(crimes,PrimaryType=="BATTERY"))+ggtitle("BATTERY")+ylab("Total Count")
g3=g(filter(crimes,PrimaryType=="CRIMINAL DAMAGE"))+ggtitle("CRIMINAL DAMAGE")+ylab("Tot
al Count")
g4=g(filter(crimes,PrimaryType=="NARCOTICS"))+ggtitle("NARCOTICS")+ylab("Total Count")
grid.arrange(g1,g2,g3,g4,ncol=2)
```



Hide

```
g=function(data){
    monthCounts = as.data.frame(table(data$month))
    ggplot(monthCounts, aes(x=Var1, y=Freq)) + geom_line(aes(group=1),size=2,color="red"
) + xlab("Day of the Week") +
    theme(axis.title.x=element_blank(),axis.text.y = element_text(color="blue",size=11,a
ngle=0,hjust=1,vjust=0),
          axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
         axis.title.y = element_text(size=14),
         plot.title=element_text(size=16,color="purple",hjust=0.5))
    }

g1=g(filter(crimes,PrimaryType=="THEFT"))+ggtitle("Theft")+ylab("Total Count")
g2=g(filter(crimes,PrimaryType=="BATTERY"))+ggtitle("BATTERY")+ylab("Total Count")
g3=g(filter(crimes,PrimaryType=="CRIMINAL DAMAGE"))+ggtitle("CRIMINAL DAMAGE")+ylab("Tot
al Count")
g4=g(filter(crimes,PrimaryType=="NARCOTICS"))+ggtitle("NARCOTICS")+ylab("Total Count")
grid.arrange(g1,g2,g3,g4,ncol=2)
```



Hide

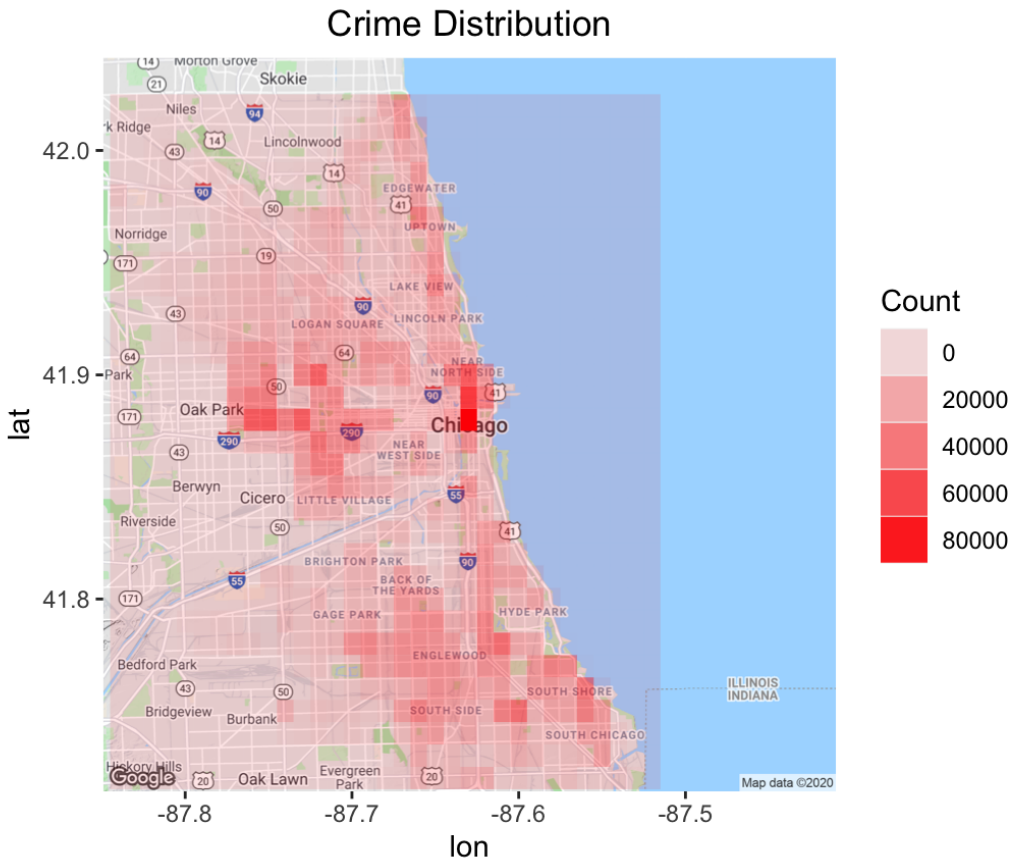
```
chicago = get_map(location = "chicago", zoom = 11) # Load a map of Chicago into R:
```

Hide

```
query3= my_collection$find('{}', fields = '{"_id":0, "Latitude":1, "Longitude":1,"Year":
1}')
LatLonCounts=as.data.frame(table(round(query3$Longitude,2),round(query3$Latitude,2)))
```

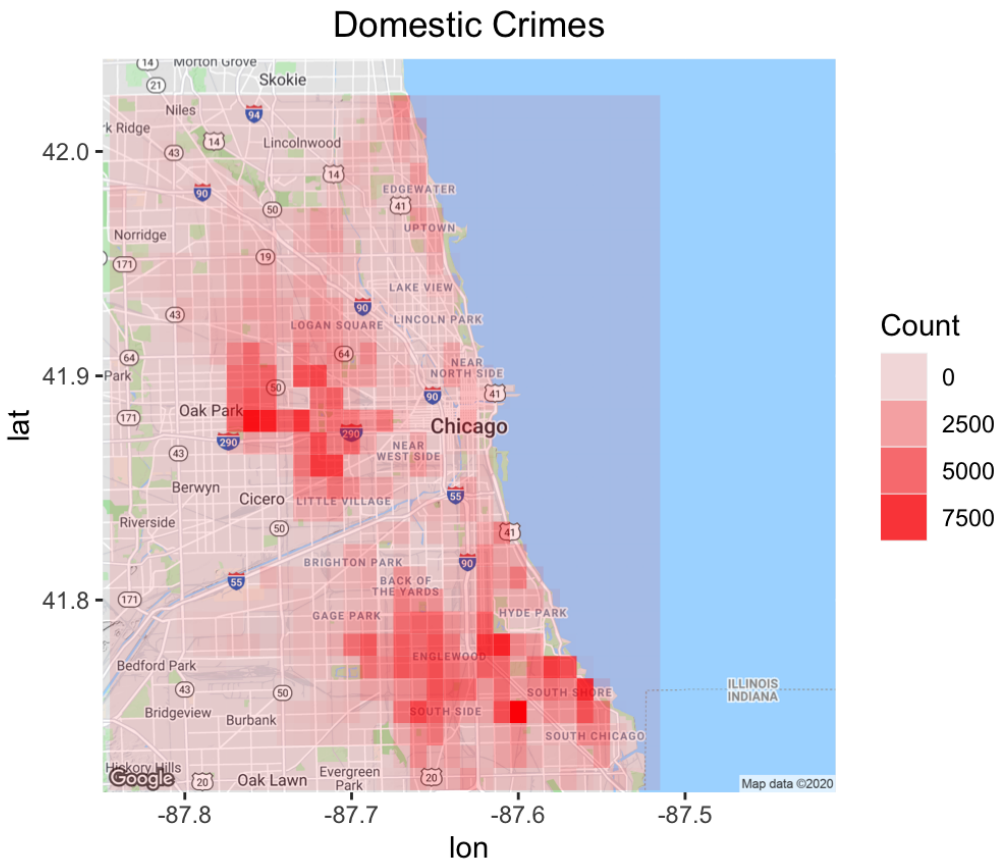
Hide

```
LatLonCounts$Long = as.numeric(as.character(LatLonCounts$Var1))
LatLonCounts$Lat = as.numeric(as.character(LatLonCounts$Var2))
ggmap(chicago) + geom_tile(data = LatLonCounts, aes(x = Long, y = Lat, alpha = Freq), fi
ll="red")+
ggtitle("Crime Distribution")+labs(alpha="Count")+theme(plot.title = element_text(hjust=
0.5))
```



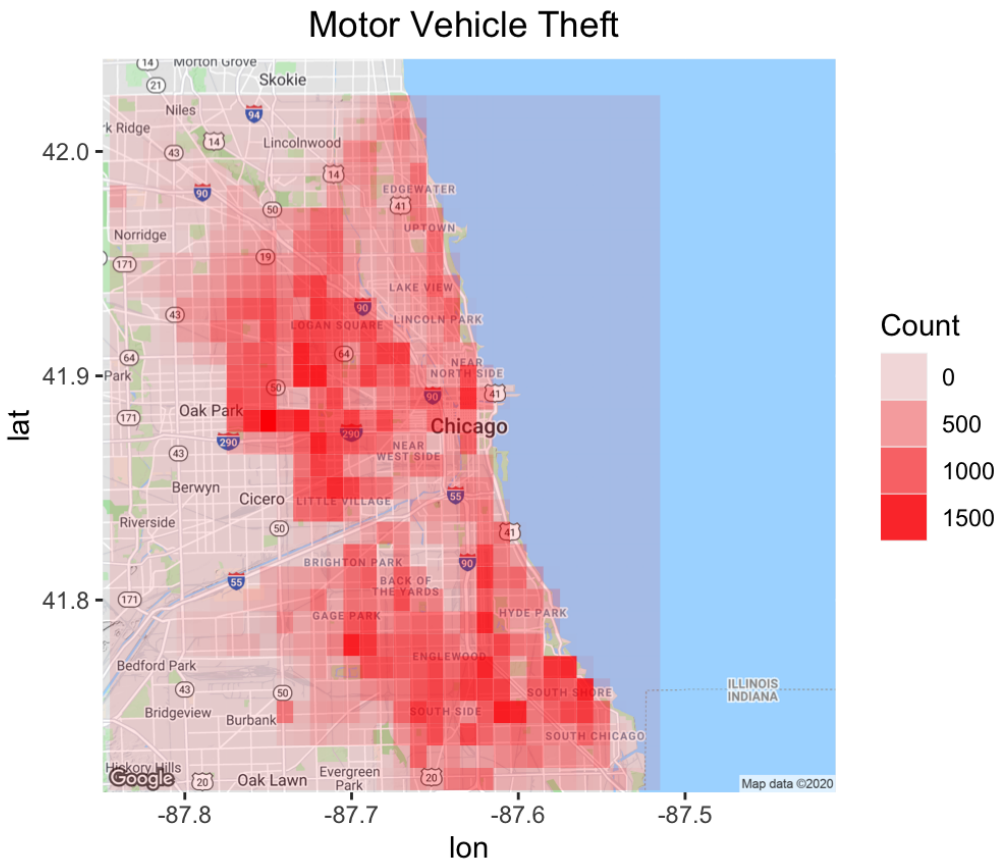
Hide

```
domestic=my_collection$find('{"Domestic":true}', fields = '{"_id":0, "Latitude":1, "Long
itude":1,"Year":1}')
LatLonCounts=as.data.frame(table(round(domestic$Longitude,2),round(domestic$Latitude,2
)))
LatLonCounts$Long = as.numeric(as.character(LatLonCounts$Var1))
LatLonCounts$Lat = as.numeric(as.character(LatLonCounts$Var2))
ggmap(chicago) + geom_tile(data = LatLonCounts, aes(x = Long, y = Lat, alpha = Freq), fi
ll="red")+
ggtitle("Domestic Crimes")+labs(alpha="Count")+theme(plot.title = element_text(hjust=0.5
))
```



Hide

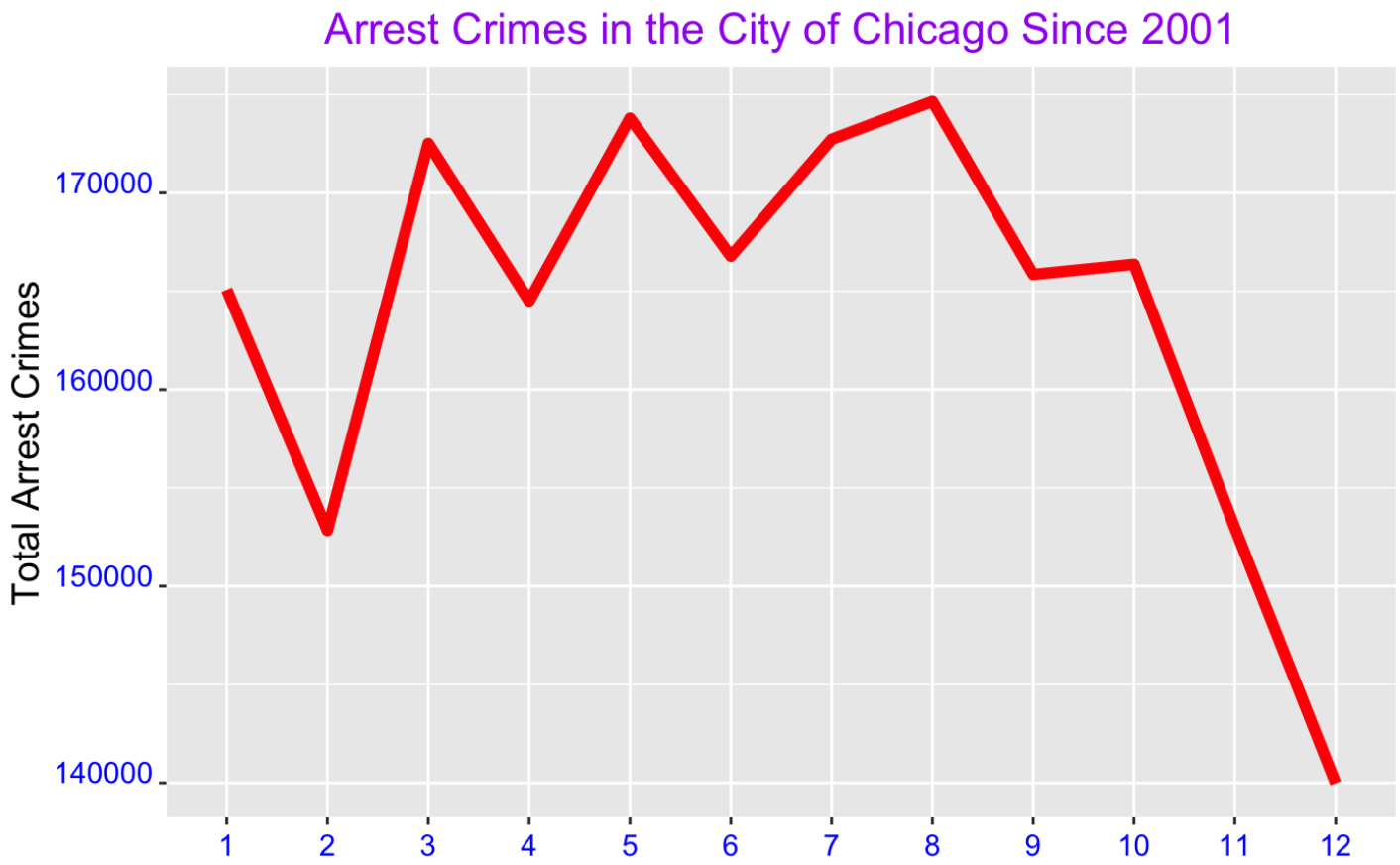
```
mtheft=my_collection$find('{"PrimaryType":"MOTOR VEHICLE THEFT"}', fields = '{"_id":0,
 "Latitude":1, "Longitude":1,"Year":1}')
LatLonCounts=as.data.frame(table(round(mtheft$Longitude,2),round(mtheft$Latitude,2)))
LatLonCounts$Long = as.numeric(as.character(LatLonCounts$Var1))
LatLonCounts$Lat = as.numeric(as.character(LatLonCounts$Var2))
ggmap(chicago) + geom_tile(data = LatLonCounts, aes(x = Long, y = Lat, alpha = Freq), fi
ll="red")+
ggtitle("Motor Vehicle Theft ")+labs(alpha="Count")+theme(plot.title = element_text(hjus
t=0.5))
```



2. Build one additional query (of your choice) to retrieve data from MongoDB into a dataframe and show that the retrieval was successful by using the result in some way.

Hide

```
arrest=my_collection$find('{"Arrest":true}', fields = '{"_id":0, "Arrest":1,"Date":1}')
arrest$Date= mdy_hms(arrest$Date)
arrest$Weekday = weekdays(arrest$Date)
arrest$Hour = hour(arrest$Date)
arrest$month = month(arrest$Date)

MonthCounts = as.data.frame(table(arrest$month))
MonthCounts$Var1 = factor(MonthCounts$Var1, ordered=TRUE, levels=c("1", "2", "3", "4",
"5", "6","7", "8", "9", "10", "11", "12"))
ggplot(MonthCounts, aes(x=Var1, y=Freq)) + geom_line(aes(group=1),size=2,color="red") +
 xlab("Month") + ylab("Total Arrest Crimes")+
ggtitle("Arrest Crimes in the City of Chicago Since 2001")+
theme(axis.title.x=element_blank(),axis.text.y = element_text(color="blue",size=11,angle
=0,hjust=1,vjust=0),
      axis.text.x = element_text(color="blue",size=11,angle=0,hjust=.5,vjust=.5),
     axis.title.y = element_text(size=14),
     plot.title=element_text(size=16,color="purple",hjust=0.5))
```


Arrest Crimes in the City of Chicago Since 2001
Total Arrest Crimes
170000
160000
150000
140000
1
2
3
4
5
6
7
8
9
10
11
12