# SDS3386 - Project

## IKEA Team

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```r
library(gbmt)
```

```
## Warning: package 'gbmt' was built under R version 4.2.3

## Loading required package: Matrix

## Warning: package 'Matrix' was built under R version 4.2.3
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3

##
## 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
```

```r
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3

##
## Attaching package: 'tidyr'

## The following objects are masked from 'package:Matrix':
##
##     expand, pack, unpack
```
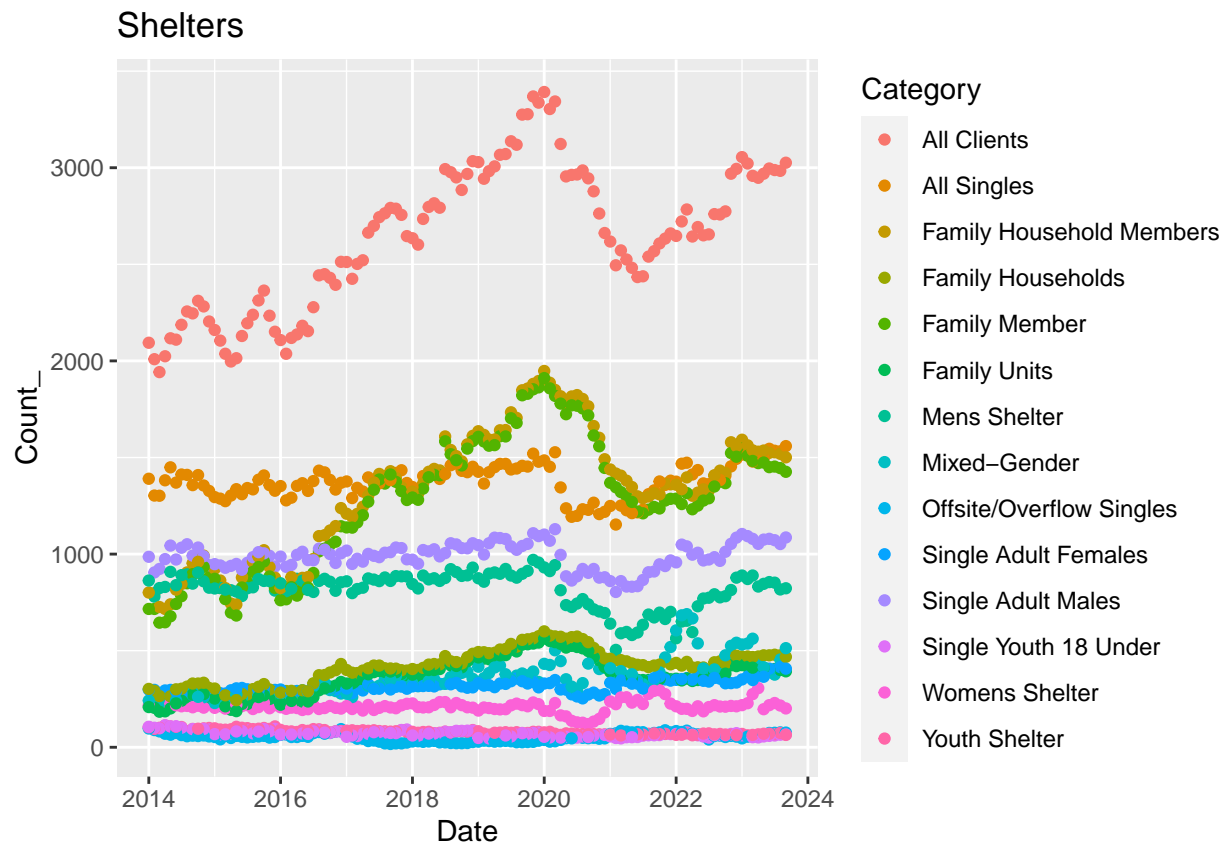
## Shelter Groups

```
dataset_name <- "Shelters"
```

**load data**

```
shelter <- read.csv("../clean_datasets/shelters.csv")
shelter$Date <- as.Date(shelter$Date)

ggplot(data = shelter, aes(x= Date, y= Count_, col = Category)) + geom_point() +
                labs(title = dataset_name)
```



```
ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, ".png"))
```

```
## Saving 6.5 x 4.5 in image
```

**Find Groups**

```
group_traj <- gbmt(x.names = "Count_", unit = "Category", time = "Date",
                data = shelter,scaling=2, d=2, ng=3)
```

```
##  EM iteration 0. Log likelihood: -1856.2178  EM iteration 1. Log likelihood: -1827.6443  EM iteration
```

2

```r
group_traj$assign.list
```

```
## $`1`
## [1] "All Clients"            "Family Household Members"
## [3] "Family Households"      "Family Member"
## [5] "Family Units"
##
## $`2`
## [1] "Mens Shelter"        "Single Youth 18 Under" "Youth Shelter"
##
## $`3`
## [1] "All Singles"            "Mixed-Gender"
## [3] "Offsite/Overflow Singles" "Single Adult Females"
## [5] "Single Adult Males"      "Womens Shelter"
```

```r
png(paste0("..\\Trajectories\\plots\\", dataset_name, "_trajectory.png"),
    width = 800, height = 600)
plot(group_traj, title = dataset_name)
dev.off()
```
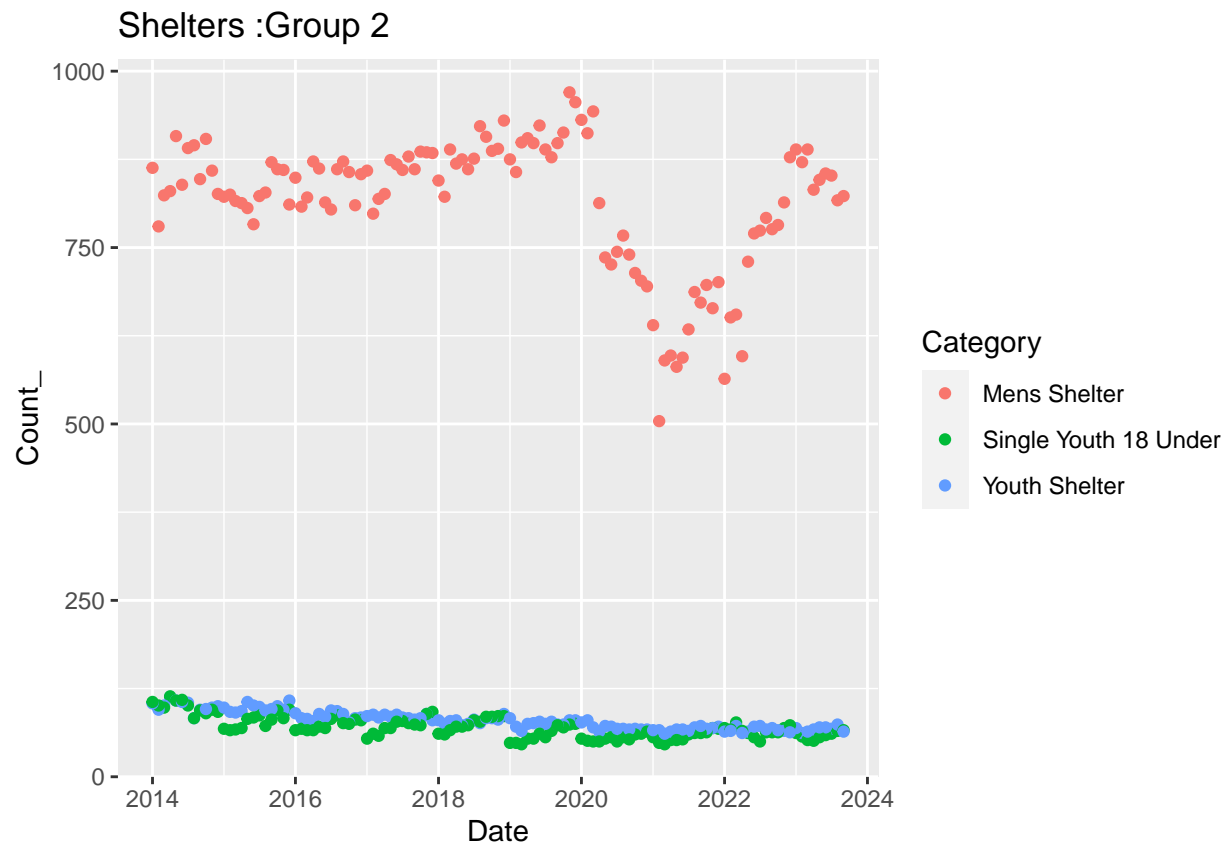
```
## pdf
##   2
```

```r
for (i in 1:length(group_traj$assign.list)){
 print(ggplot(data = shelter[shelter$Category %in%
                                group_traj$assign.list[[i]],],
          aes(x= Date, y= Count_, col = Category)) +   geom_point() +
            labs(title = paste(dataset_name,":Group", i)) )
 ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, "_Group_", i, ".png"))
}
```
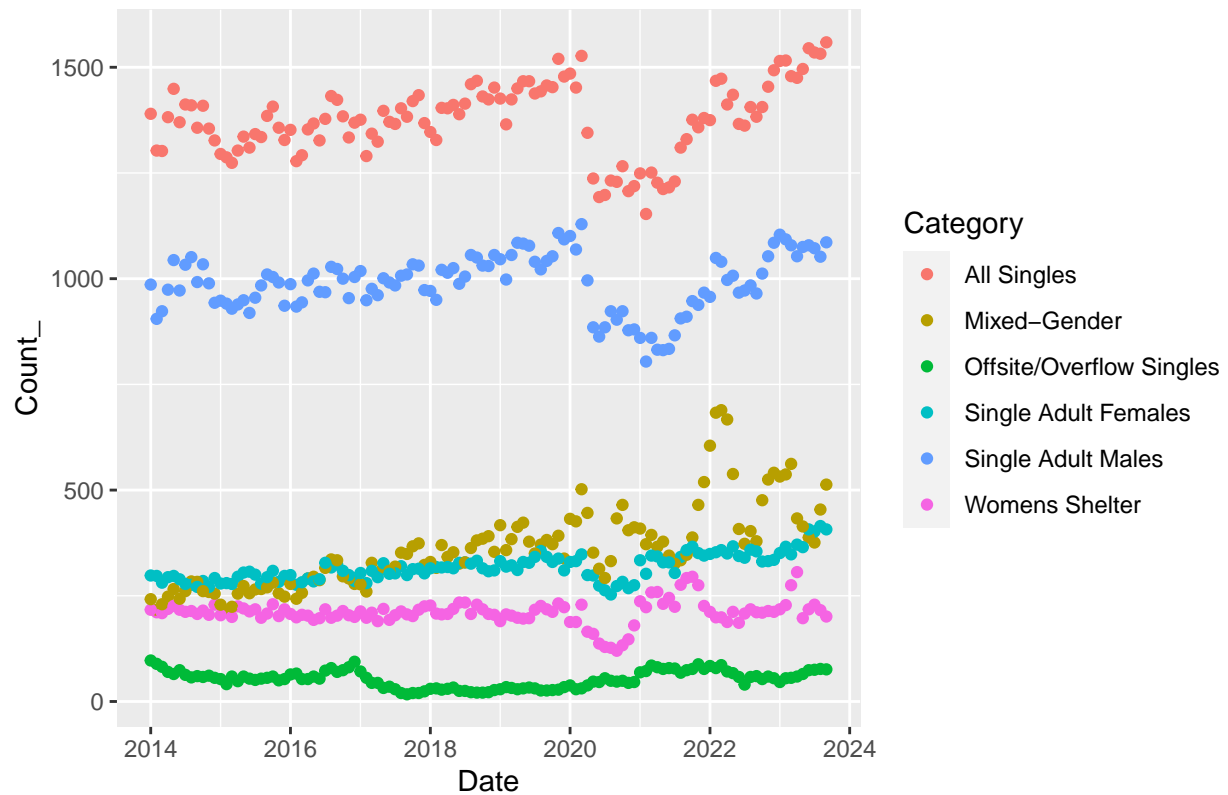
```
## Saving 6.5 x 4.5 in image
```

Shelters :Group 1

Category
- All Clients
- Family Household Members
- Family Households
- Family Member
- Family Units

```
## Saving 6.5 x 4.5 in image
```

Shelters :Group 2

```
## Saving 6.5 x 4.5 in image
```

## Shelter Individuals

```r
dataset_name <- "Shelters Individuals"
```

**load data**

```r
shelter <- read.csv("../clean_datasets/shelters_indvd.csv")
shelter$Date <- as.Date(shelter$Date)

ggplot(data = shelter, aes(x= Date, y= Count_, col = Category)) + geom_point()+
            labs(title = dataset_name)
```

## Shelters Individuals



```r
ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, ".png"))
```

```
## Saving 6.5 x 4.5 in image
```

**Find Groups**

```r
group_traj <- gbmt(x.names = "Count_", unit = "Category", time = "Date",
                   data = shelter,scaling=2, d=2, ng=2)
```

```
##  EM iteration 0. Log likelihood: -563.3023  EM iteration 1. Log likelihood: -563.3023
```

```r
group_traj$assign.list
```

```
## $`1`
## [1] "Family Household Members" "Single Adult Females"
## [3] "Single Adult Males"
##
## $`2`
## [1] "Single Youth 18 Under"
```

```
png(paste0("..\\Trajectories\\plots\\", dataset_name, "_trajectory.png"),
    width = 800, height = 600)
plot(group_traj, title = dataset_name)
dev.off()
```

```
## pdf
##   2
```

```
for (i in 1:length(group_traj$assign.list)){
 print(ggplot(data = shelter[shelter$Category %in%
                                group_traj$assign.list[[i]],],
            aes(x= Date, y= Count_, col = Category)) +  geom_point() +
             labs(title = paste(dataset_name, ":Group", i)) + xlab("Year") + ylab("Number of people in
 ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, "_Group_", i, ".png"))
}
```
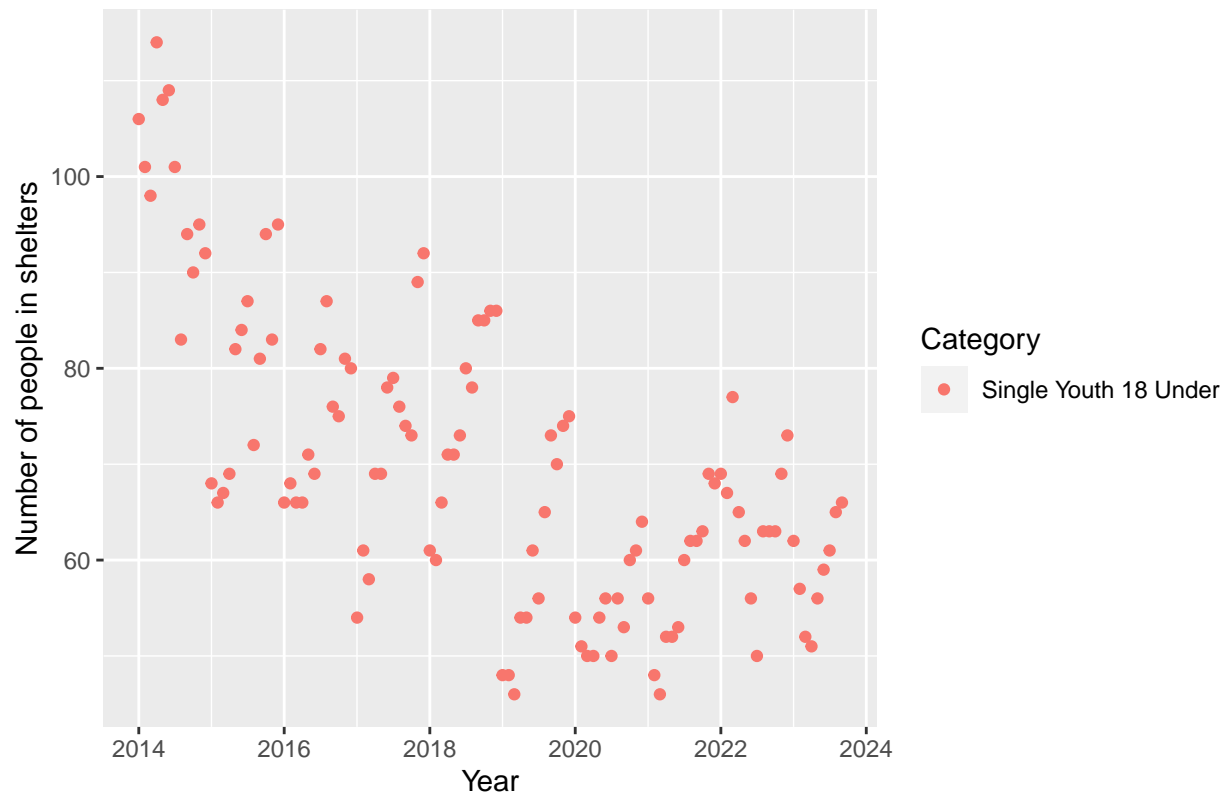
```
## Saving 6.5 x 4.5 in image
```
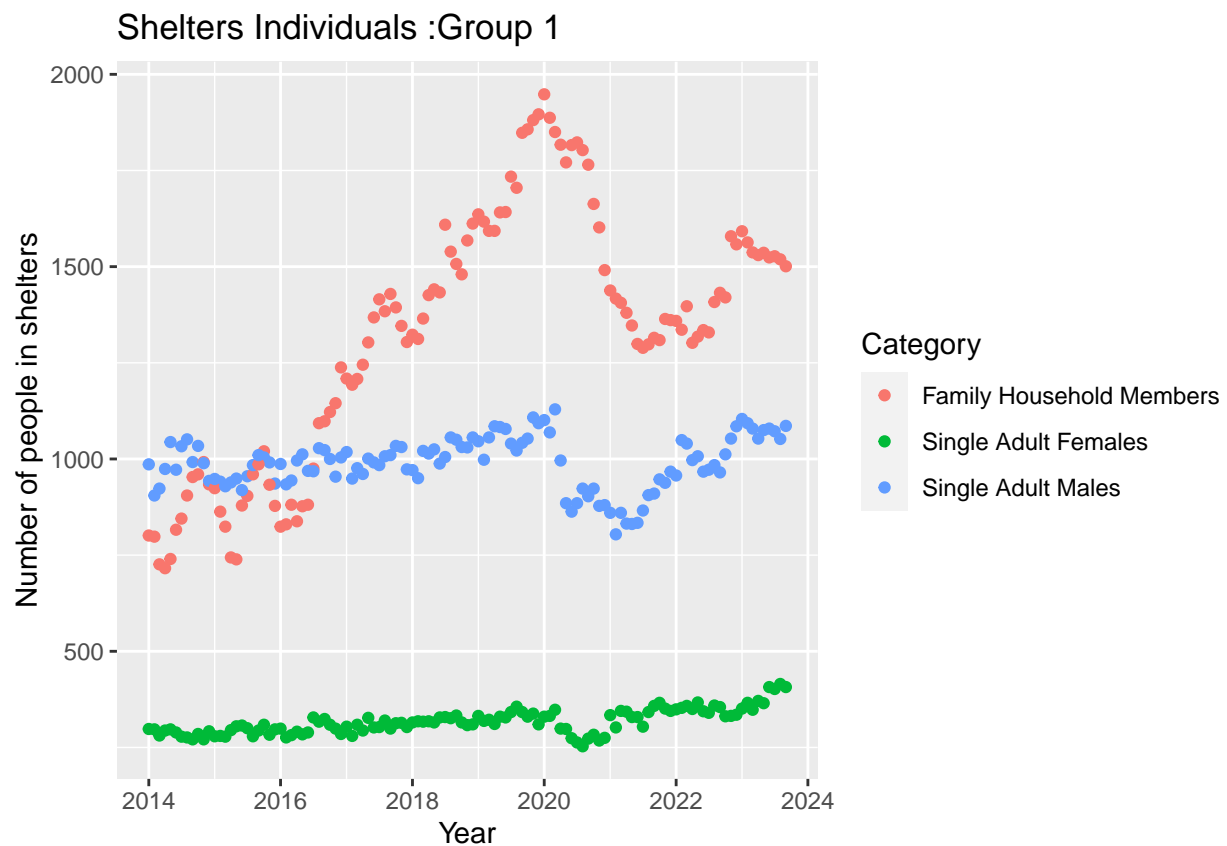


```
## Saving 6.5 x 4.5 in image
```
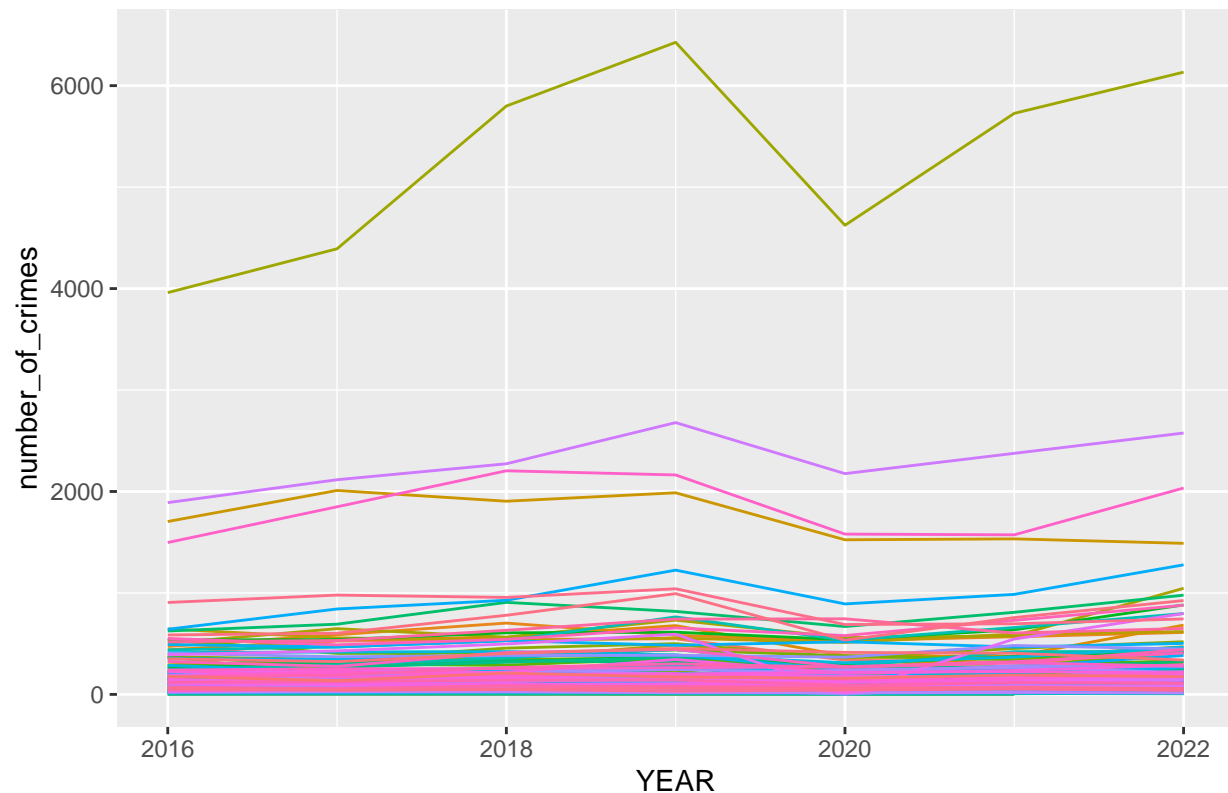
## Shelters Individuals :Group 2



## Crimes By neighbourhood

```r
dataset_name <- "Crimes By neighbourhood"
```

**load data**

```r
crimes <- read.csv("../clean_datasets/crimes_for_traj_by_neighbourhood.csv")
crimes_no_noise <- crimes[crimes$YEAR >= 2016,]
ggplot(data = crimes_no_noise, aes(x= YEAR, y= number_of_crimes, col = NB_NAME_EN)) +
        geom_line()+ theme(legend.position = "none")+
                labs(title = dataset_name)
```

## Crimes By neighbourhood



```r
ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, ".png"))
```

```
## Saving 6.5 x 4.5 in image
```

## fill missing values

```r
## https://stackoverflow.com/questions/43799109/add-rows-with-missing-years-by-group
crimes_filled_na<-crimes_no_noise %>%
  group_by(NB_NAME_EN ) %>%
  complete(YEAR = full_seq(2016:2022, 1)) %>%
  fill(number_of_crimes) %>% ungroup()
crimes_filled_na <- data.frame(crimes_filled_na)
crimes_filled_na[is.na(crimes_filled_na$X),]$number_of_crimes <- 0.1
```

## Find Groups

```r
group_traj <- gbmt(x.names = "number_of_crimes", unit = "NB_NAME_EN",
                   time = "YEAR", data = data.frame(crimes_filled_na) ,
                   scaling=4, d=3, ng=2)
```

```
##  EM iteration 0. Log likelihood: -156.7995  EM iteration 1. Log likelihood: -42.9839  EM iteration 2
```

```
group_traj$assign.list
```

```
## $`1`
##  [1] "Bayshore - Belltown"
##  [2] "Beacon Hill South - Cardinal Heights"
##  [3] "Beaverbrook"
##  [4] "Bells Corners East"
##  [5] "Bells Corners West"
##  [6] "Billings Bridge - Alta Vista"
##  [7] "Blackburn Hamlet"
##  [8] "Borden Farm - Fisher Glen"
##  [9] "Braemar Park - Bel Air Heights - Copeland Park"
## [10] "Briar Green - Leslie Park"
## [11] "Bridlewood - Emerald Meadows"
## [12] "Britannia Village"
## [13] "Brookside - Briarbrook - Morgan's Grant"
## [14] "Byward Market"
## [15] "Carleton Heights - Rideauview"
## [16] "Carlington"
## [17] "Carlingwood West - Glabar Park - McKellar Heights"
## [18] "Carp"
## [19] "Carson Grove - Carson Meadows"
## [20] "Centrepointe"
## [21] "Centretown"
## [22] "Chapel Hill North"
## [23] "Chapman Mills"
## [24] "Cityview - Crestview - Meadowlands"
## [25] "Civic Hospital-Central Park"
## [26] "Constance Bay"
## [27] "Convent Glen - Orléans Woods"
## [28] "Corkery"
## [29] "Crystal Bay - Lakeview Park"
## [30] "Dunrobin"
## [31] "East Industrial"
## [32] "Edwards - Carlsbad Springs"
## [33] "Elmvale - Canterbury"
## [34] "Emerald Woods - Sawmill Creek"
## [35] "Findlay Creek"
## [36] "Fitzroy"
## [37] "Glebe - Dows Lake"
## [38] "Glen Cairn - Kanata South Business Park"
## [39] "Greely"
## [40] "Greenbelt-Findlay"
## [41] "Greenbelt-Merivale"
## [42] "Greenboro West"
## [43] "Hawthorne Meadows - Sheffield Glen"
## [44] "Hintonburg - Mechanicsville"
## [45] "Hunt Club - Ottawa Airport"
## [46] "Hunt Club East - Western Community"
## [47] "Hunt Club South Industrial"
## [48] "Hunt Club Upper -Blossom Park - Timbermill"
## [49] "Hunt Club Woods - Quintarra - Revelstoke"
## [50] "Huntclub Park"
```

```
## [51] "Iris - Queensway Terrance South"
## [52] "Island Park - Wellington Village"
## [53] "Kanata Lakes - Arcardia"
## [54] "Katimavik - Hazeldean"
## [55] "Kinburn"
## [56] "Laurentian"
## [57] "Lebreton Development"
## [58] "Ledbury - Heron Gate - Ridgemont"
## [59] "Lowertown"
## [60] "Manor Park"
## [61] "Manotick East"
## [62] "Manotick West"
## [63] "Marlborough"
## [64] "Merivale Gardens - Grenfell Glen - Pineglen - Country Place"
## [65] "Munster - Ashton"
## [66] "North Gower - Kars"
## [67] "Old Barrhaven East"
## [68] "Old Barrhaven West"
## [69] "Old Ottawa East"
## [70] "Old Ottawa South"
## [71] "Orléans Industrial"
## [72] "Osgoode - Vernon"
## [73] "Overbrook - McArthur"
## [74] "Parkwood Hills - Stewart Farm"
## [75] "Pineview"
## [76] "Playfair Park - Lynda Park - Guildwood Estates"
## [77] "Qualicum - Redwood Park"
## [78] "Richmond"
## [79] "Rideau Crest - Davidson Heights"
## [80] "Riverside Park"
## [81] "Riverside South - Leitrim"
## [82] "Riverview"
## [83] "Rockcliffe Park"
## [84] "Rothwell Heights - Beacon Hill North"
## [85] "Sandy Hill"
## [86] "Skyline - Fisher Heights"
## [87] "South Keys - Greenboro West"
## [88] "Stittsville"
## [89] "Stonebridge - Halfmoon Bay - Heart's Desire"
## [90] "Tanglewood"
## [91] "Trend-Arlington"
## [92] "Vanier North"
## [93] "Wateridge Village"
## [94] "West Centertown"
## [95] "Westboro"
## [96] "Whitehaven - Queensway Terrace North"
## [97] "Woodvale - Craig Henry - Manordale - Estates of Arlington Woods"
##
## $`2`
##  [1] ""                       "Beechwood Cemetery"
##  [3] "Cardinal Creek"         "Carleton University"
##  [5] "Chapel Hill South"      "Chatelaine Village"
##  [7] "Cumberland"             "Fallingbrook"
##  [9] "Greenbelt-Dunrobin"     "Greenbelt-Edwards"
```

```
## [11] "Greenbelt - Shirleys Bay"        "Greenbelt - SouthEast"
## [13] "Lindenlea - New Edinburgh"        "Metcalfe"
## [15] "Navan - Sarsfield"               "Orléans Village - Chateauneuf"
## [17] "Portobello South"                "Queenswood Heights"
## [19] "Vars"
```

```r
## https://stackoverflow.com/questions/55816280/unlisting-a-list-while-keeping-the-indices
traj_nei_results <- data.frame(stack(setNames(group_traj$assign.list, seq_along(group_traj$assign.list))
colnames(traj_nei_results) <- c("Group_ID", "NB_NAME_EN")
write.csv(traj_nei_results, "../Trajectories/traj_nei_results.csv")
```

```r
png(paste0("..\\Trajectories\\plots\\", dataset_name, "_trajectory.png"),
    width = 800, height = 600)
plot(group_traj, title = dataset_name)
dev.off()
```

```
## pdf
##   2
```

```r
for (i in 1:length(group_traj$assign.list)){
 print(ggplot(data = crimes_no_noise[crimes_no_noise$NB_NAME_EN %in%
                                        group_traj$assign.list[[i]],],
            aes(x= YEAR, y= number_of_crimes, col = NB_NAME_EN)) +
              geom_line() + labs(title = paste(dataset_name, ":Group", i))+
              theme(legend.position = "none") )
 ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, "_Group_", i, ".png"))
}
```
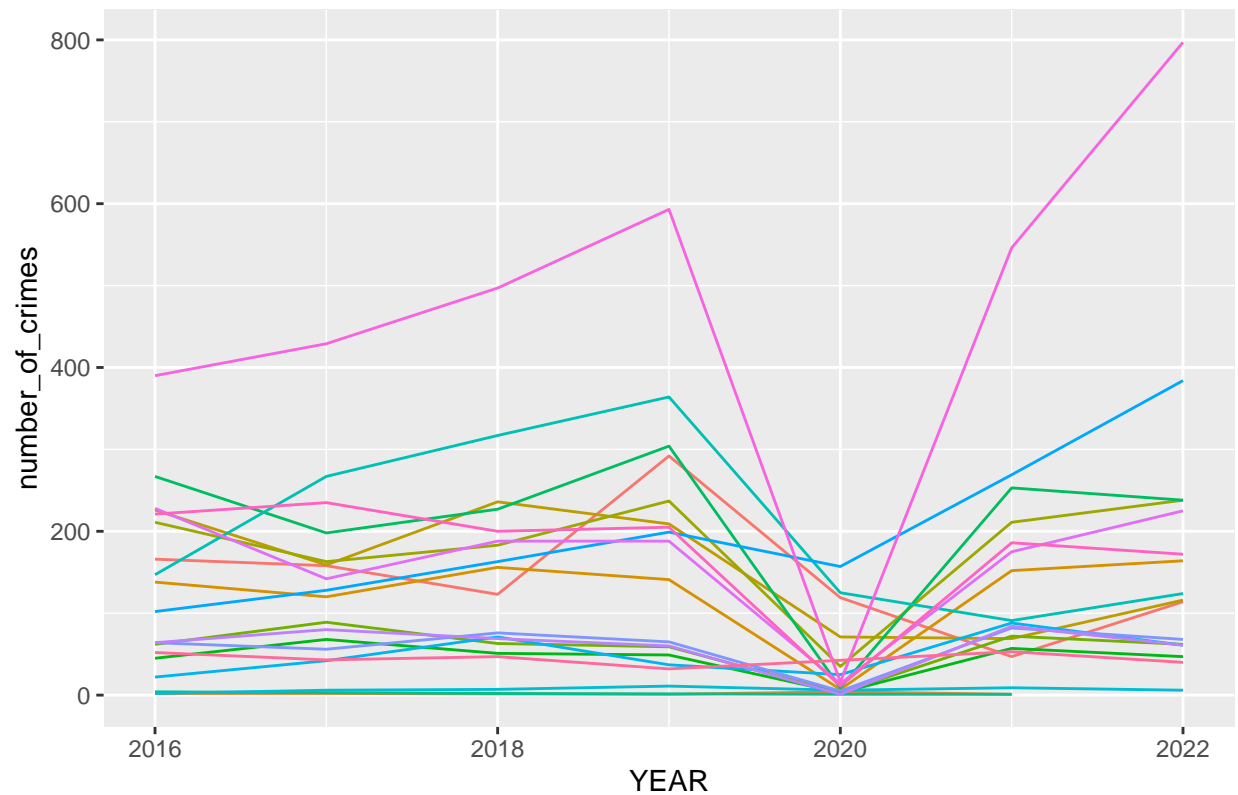
```
## Saving 6.5 x 4.5 in image
```

Crimes By neighbourhood :Group 1

```
## Saving 6.5 x 4.5 in image
```
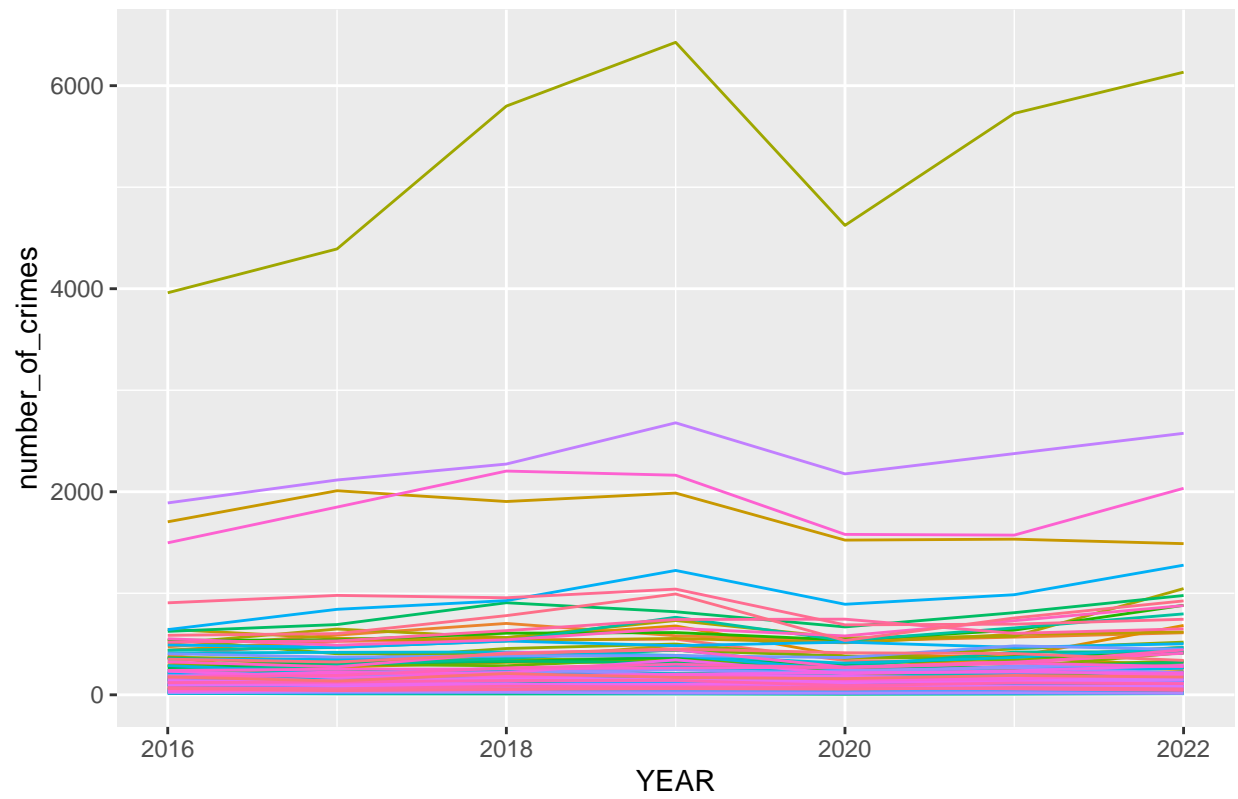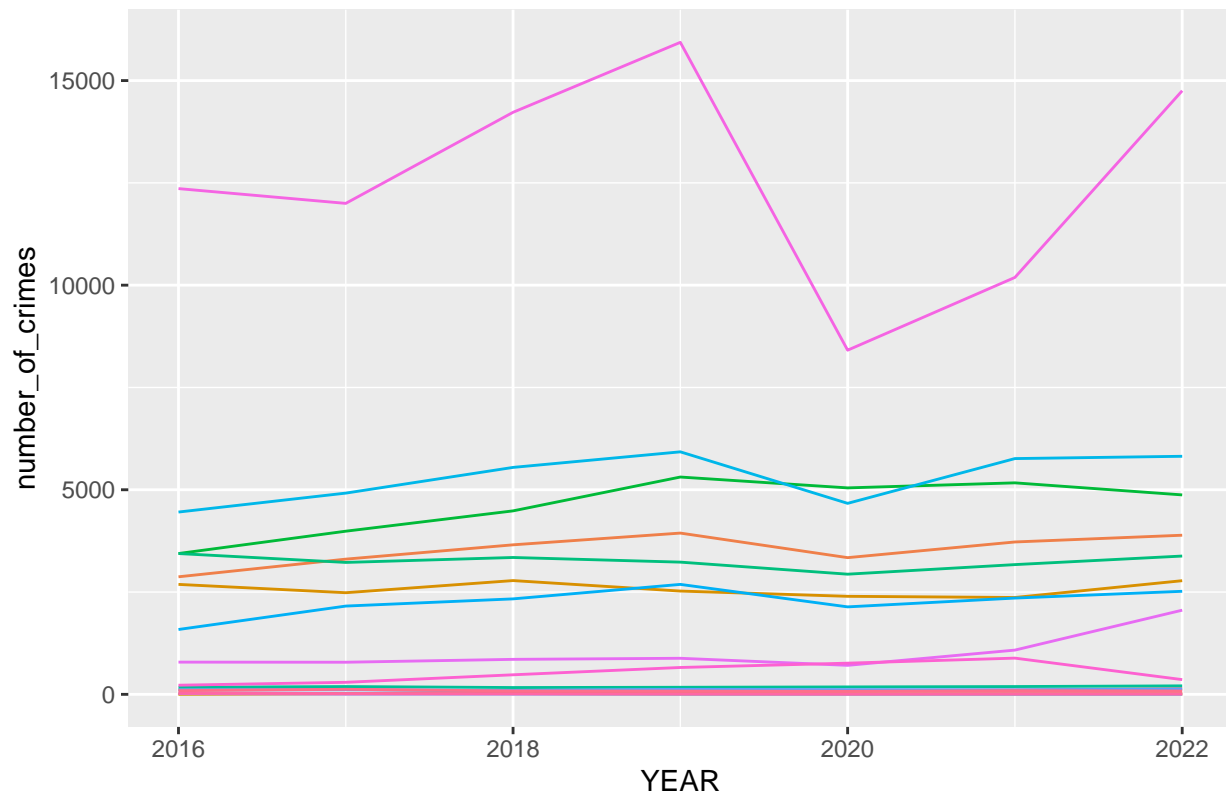
## Crimes By neighbourhood :Group 2



## Crimes By Crime Type

```
dataset_name <- "Crimes By Crime Type"
```

**load data**

```
crimes <- read.csv("../clean_datasets/crimes_for_traj_by_crime.csv")
crimes_no_noise <- crimes[crimes$YEAR >= 2016,]
ggplot(data = crimes_no_noise, aes(x= YEAR, y= number_of_crimes, col = OFF_CATEG)) +
        geom_line()+ theme(legend.position = "none")+
                labs(title = dataset_name)
```

## Crimes By Crime Type



```r
ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, ".png"))
```

```
## Saving 6.5 x 4.5 in image
```

**fill missing values**

```r
## https://stackoverflow.com/questions/43799109/add-rows-with-missing-years-by-group
crimes_filled_na<-crimes_no_noise %>%
  group_by(OFF_CATEG ) %>%
  complete(YEAR = full_seq(2016:2022, 1)) %>%
  fill(number_of_crimes) %>% ungroup()
crimes_filled_na <- data.frame(crimes_filled_na)
crimes_filled_na[is.na(crimes_filled_na$X),]$number_of_crimes <- 0.1
```

**Find Groups**

```r
group_traj <- gbmt(x.names = "number_of_crimes", unit = "OFF_CATEG",
                   time = "YEAR", data = crimes_filled_na,scaling=4, d=3, ng=4)
```

```
##  EM iteration 0. Log likelihood: -237.1282  EM iteration 1. Log likelihood: -224.6046  EM iteration
```
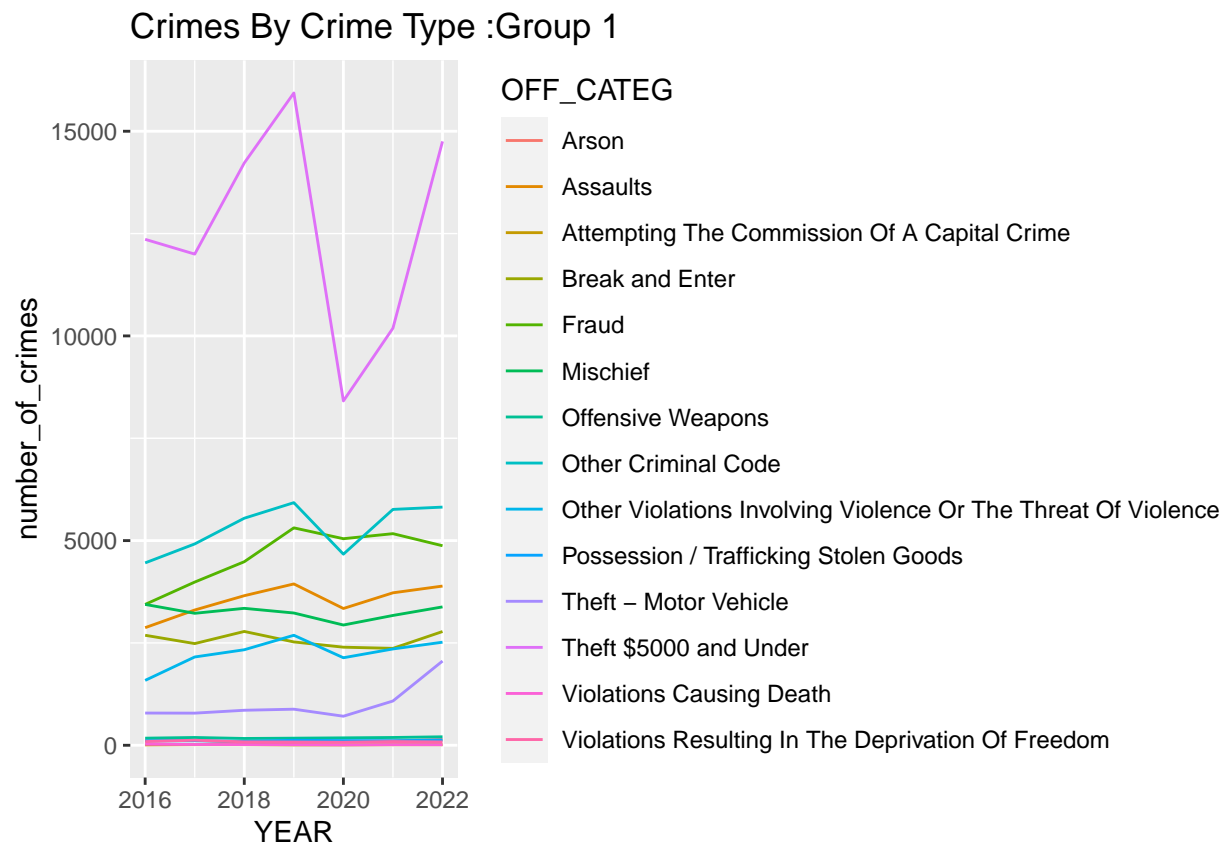
```
group_traj$assign.list
```

```
## $`1`
##  [1] "Arson"
##  [2] "Assaults"
##  [3] "Attempting The Commission Of A Capital Crime"
##  [4] "Break and Enter"
##  [5] "Fraud"
##  [6] "Mischief"
##  [7] "Offensive Weapons"
##  [8] "Other Criminal Code"
##  [9] "Other Violations Involving Violence Or The Threat Of Violence"
## [10] "Possession / Trafficking Stolen Goods"
## [11] "Theft - Motor Vehicle"
## [12] "Theft $5000 and Under"
## [13] "Violations Causing Death"
## [14] "Violations Resulting In The Deprivation Of Freedom"
##
## $`2`
## [1] "Dangerous Operation"
## [2] "Distribution"
## [3] "Failure or Refusal to Comply with Demand"
## [4] "Failure to Stop after Accident"
## [5] "Flight From Peace Officer"
## [6] "Operation while Impaired/Low Blood Drug Concentration Violations"
## [7] "Operation while Prohibited"
## [8] "Possession"
## [9] "Trafficking"
##
## $`3`
## [1] "Commodification Of Sexual Activity" "Gaming and Betting"
## [3] "Other Cannabis Violations"          "Production"
## [5] "Sale"                               "Theft Over $5000"
##
## $`4`
## [1] "Prostitution"
```

```
png(paste0("..\\Trajectories\\plots\\", dataset_name, "_trajectory.png"),
    width = 800, height = 600)
plot(group_traj, title = dataset_name)
dev.off()
```

```
## pdf
##   2
```
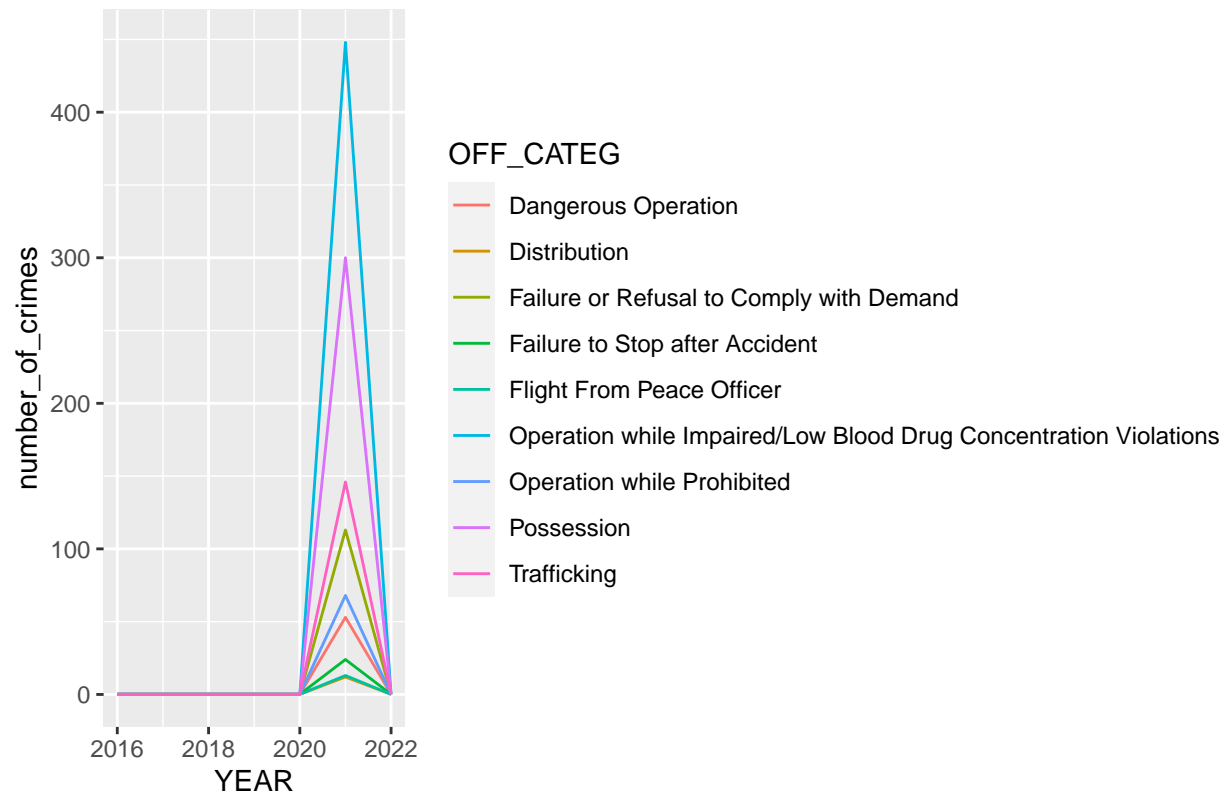
```
for (i in 1:length(group_traj$assign.list)){
 print(ggplot(data = crimes_filled_na[crimes_filled_na$OFF_CATEG %in%
                                        group_traj$assign.list[[i]],],
         aes(x= YEAR, y= number_of_crimes, col = OFF_CATEG) +
           geom_line() + labs(title = paste(dataset_name, ":Group", i)))
 ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, "_Group_", i, ".png"))
}
```
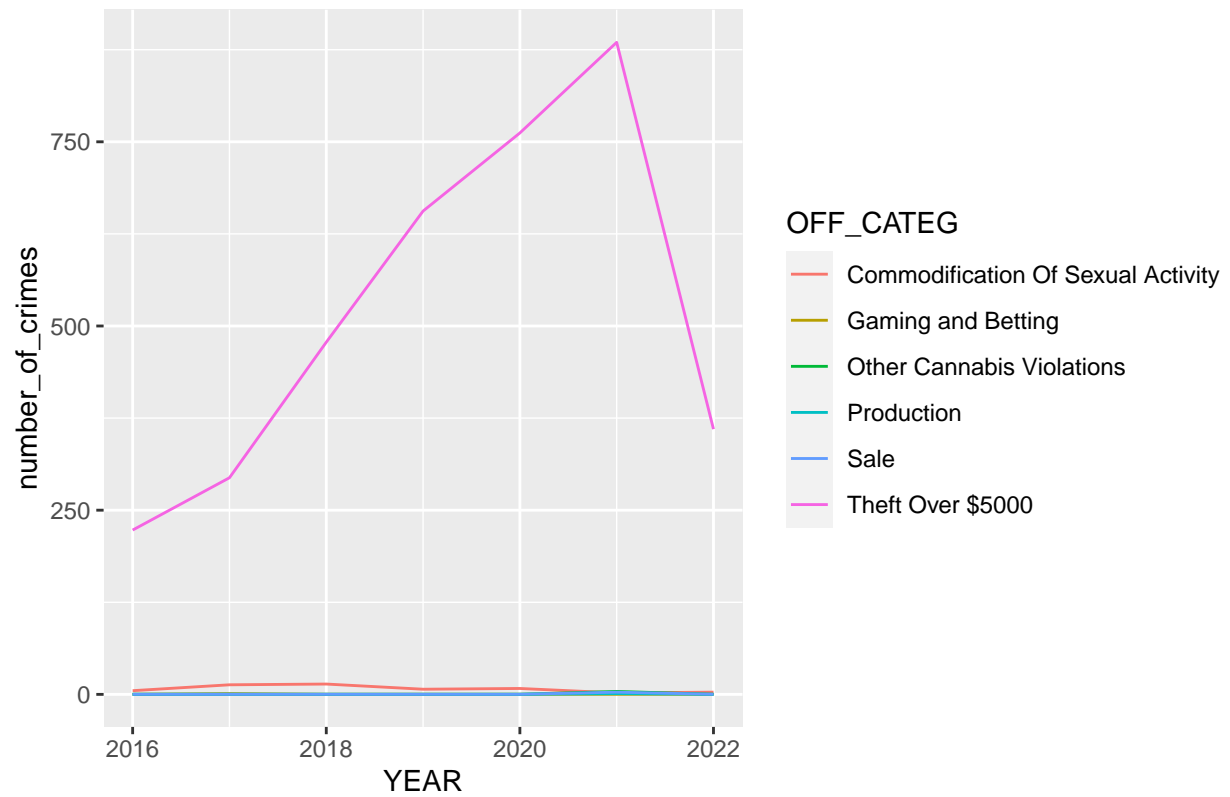
```
## Saving 6.5 x 4.5 in image
```

## Crimes By Crime Type :Group 1



```
## Saving 6.5 x 4.5 in image
```

# Crimes By Crime Type :Group 2



OFF_CATEG

- Dangerous Operation
- Distribution
- Failure or Refusal to Comply with Demand
- Failure to Stop after Accident
- Flight From Peace Officer
- Operation while Impaired/Low Blood Drug Concentration Violations
- Operation while Prohibited
- Possession
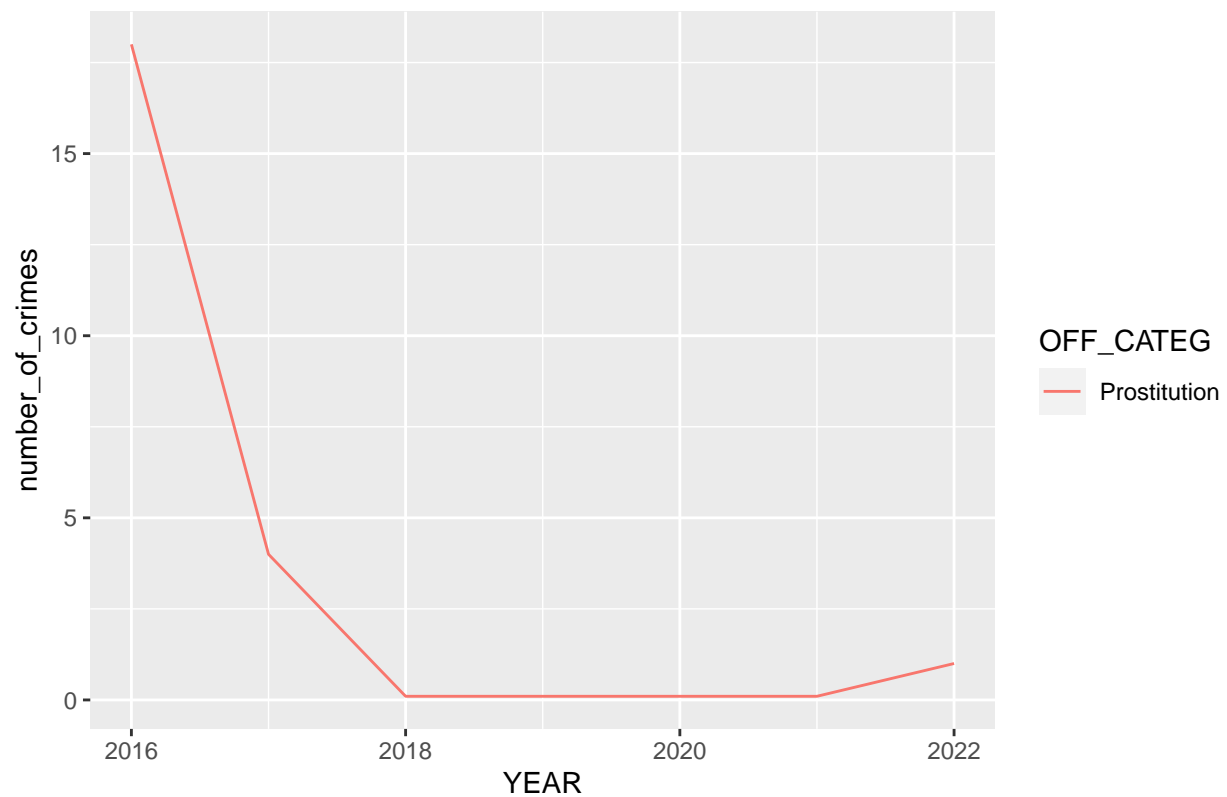- Trafficking

```
## Saving 6.5 x 4.5 in image
```

## Crimes By Crime Type :Group 3



```
## Saving 6.5 x 4.5 in image
```
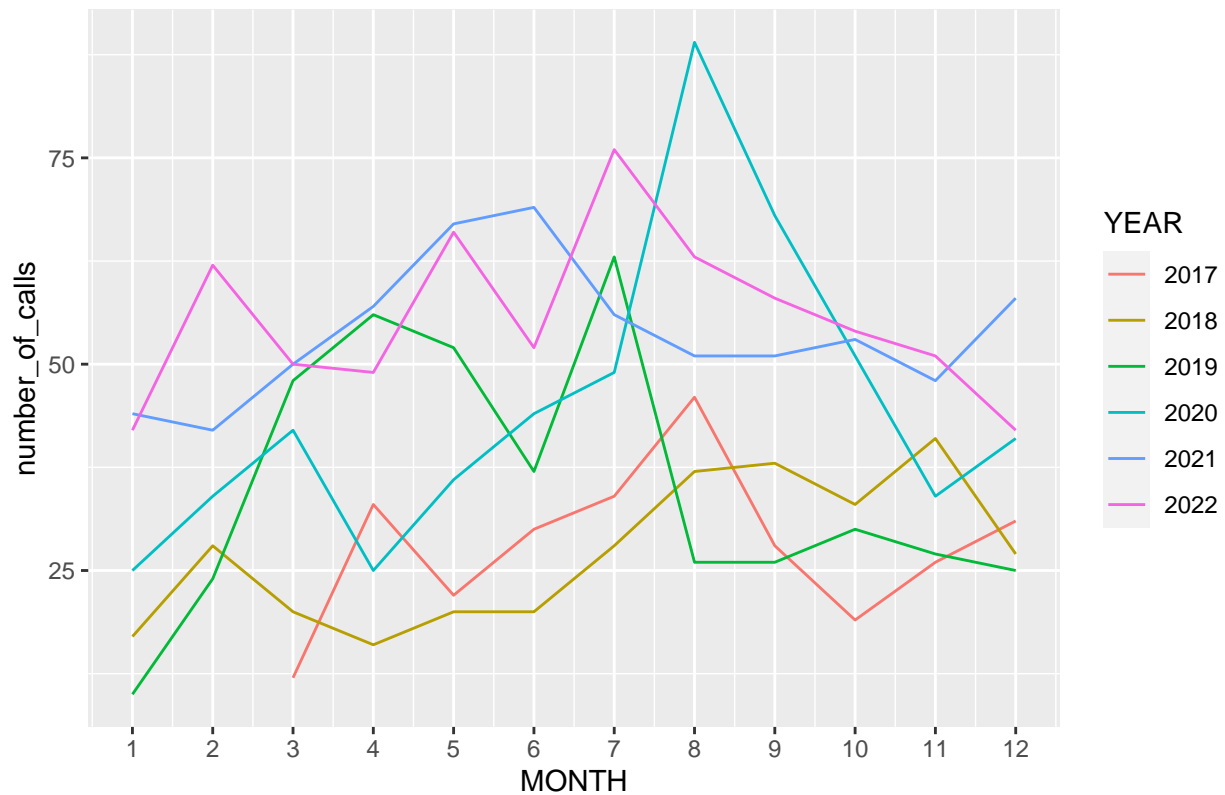
## Crimes By Crime Type :Group 4



## Overdoses

```
dataset_name <- "Overdoses"
```

**load data**

```
overdose <- read.csv("../clean_datasets/Overdose_calls_by_year_and_month.csv")
overdose$YEAR <- as.factor(overdose$YEAR)
ggplot(data = overdose, aes(x= MONTH, y= number_of_calls, col = YEAR)) +
      geom_line()+
            labs(title = dataset_name)  +
            scale_x_continuous(breaks = seq(1, 12, 1))
```

## Overdoses



```r
ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, ".png"))
```

```
## Saving 6.5 x 4.5 in image
```

**fill missing values**

```r
## https://stackoverflow.com/questions/43799109/add-rows-with-missing-years-by-group
overdose<-overdose %>%
  group_by(YEAR ) %>%
  complete(MONTH = full_seq(1:12, 1)) %>%
  fill(number_of_calls) %>% ungroup()
overdose <- data.frame(overdose)
overdose[is.na(overdose$number_of_calls),]$number_of_calls <- 0.1
```

**Find Groups**

```r
group_traj <- gbmt(x.names = "number_of_calls", unit = "YEAR",
                   time = "MONTH", data = overdose,scaling=2, d=3, ng=3)
```

```
##  EM iteration 0. Log likelihood: -75.9994  EM iteration 1. Log likelihood: -75.9938  EM iteration 2.
```

```
group_traj$assign.list
```

```
## $`1`
## [1] "2017" "2018" "2020"
##
## $`2`
## [1] "2019" "2021"
##
## $`3`
## [1] "2022"
```

```
png(paste0("..\\Trajectories\\plots\\", dataset_name, "_trajectory.png"),
    width = 800, height = 600)
plot(group_traj, title = dataset_name)
dev.off()
```

```
## pdf
##   2
```

```
for (i in 1:length(group_traj$assign.list)){
 print(ggplot(data = overdose[overdose$YEAR %in%
                                         group_traj$assign.list[[i]],],
            aes(x= MONTH, y= number_of_calls, col = YEAR)) +
              geom_line() + labs(title = paste(dataset_name, ":Group", i)) +
              scale_x_continuous(breaks = seq(1, 12, 1)))
 ggsave(paste0("..\\Trajectories\\plots\\", dataset_name, "_Group_", i, ".png"))
}
```
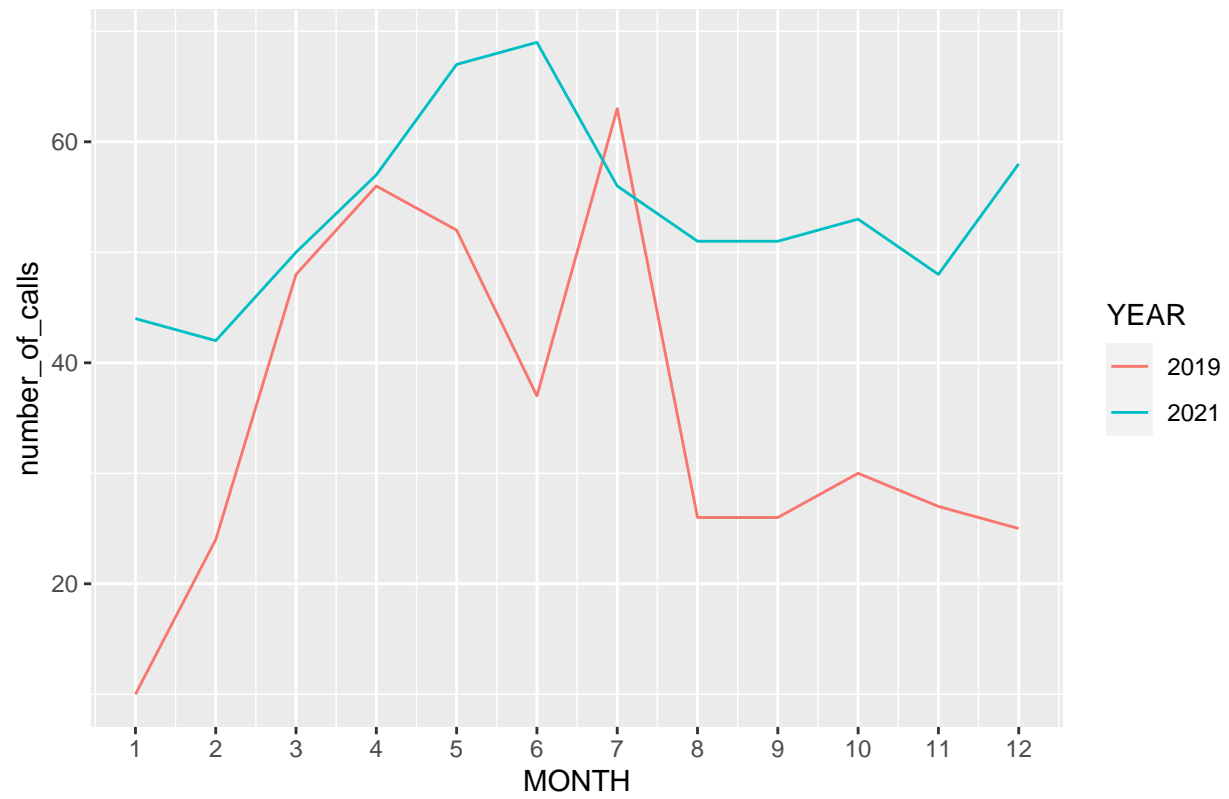
```
## Saving 6.5 x 4.5 in image
```

Overdoses :Group 1
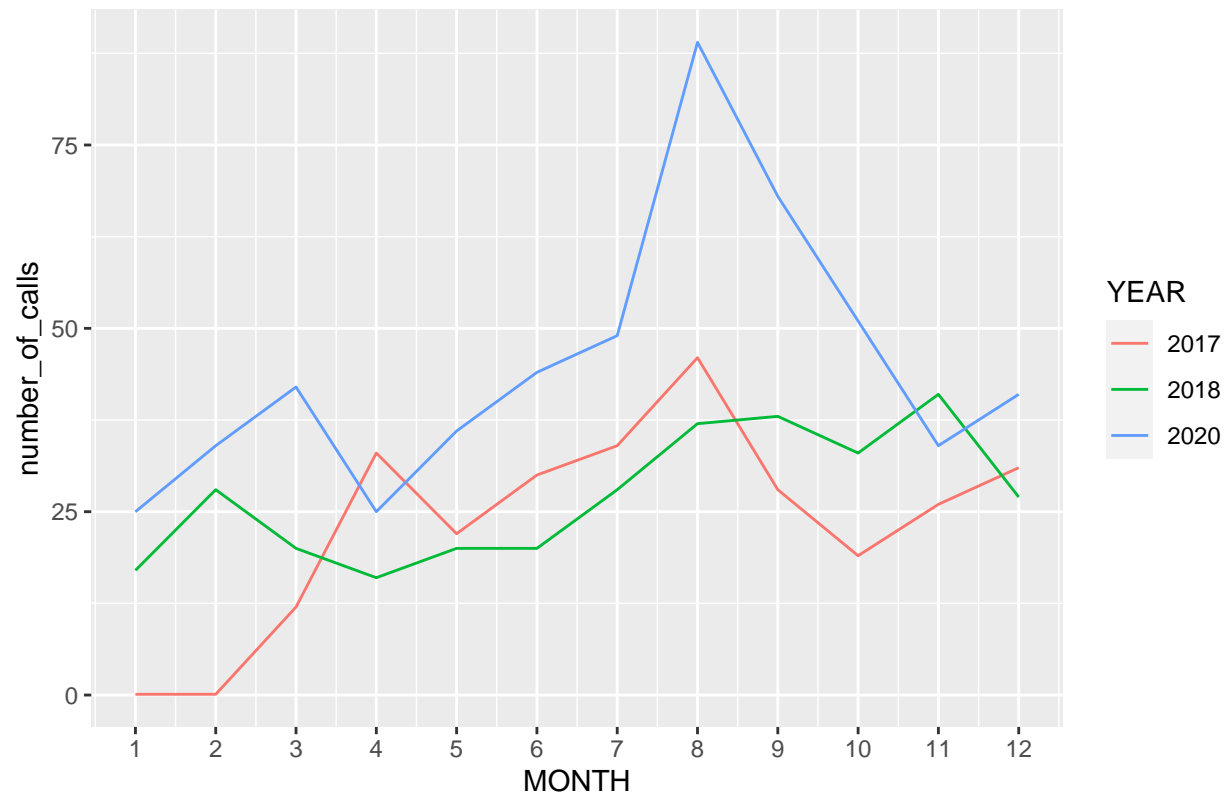
```
## Saving 6.5 x 4.5 in image
```

Overdoses :Group 2

```
## Saving 6.5 x 4.5 in image
```

Overdoses :Group 3