

Conteúdo

1. Visão geral	2
2. Quando os crimes ocorrem?	9
3. Onde os crimes ocorrem?	12
4. Explorando dados temporais	15
5. Conclusão	20
6. Código	21
7. Referência	27

1. VISÃO GERAL

Exploração de um dataset público de crimes ocorridos na cidade de Boston - EUA, a fim de se obter conhecimentos e insights como quando e onde normalmente os crimes ocorrem e com qual frequência partir de uma análise de série temporal.

O dataset foi obtido a partir repositório disponível no site Kaggle [1] apresentando 319073 linhas, 17 colunas e 58MB. A menor data de ocorrência de crime registrada foi em 15-06-2015 e a última data registrada foi 03-09-2018.

EXPLORAÇÃO INICIAL DO DATASET

Para uma melhor exploração foi importado o dataset de maneira a preencher qualquer registro vazio com NA (*not available*), assim padronizando a tratativa com os registros sem algum valor.

A partir de uma amostra de 5 linhas foi possível visualizar características das features desse dataset.

INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING
I182070945	619	Larceny	LARCENY ALL OTHERS	D14	808	NA
I182070943	1402	Vandalism	VANDALISM	C11	347	NA
I182070941	3410	Towed	TOWED MOTOR VEHICLE	D4	151	NA
I182070940	3114	Investigate Property	INVESTIGATE PROPERTY	D4	272	NA
I182070938	3114	Investigate Property	INVESTIGATE PROPERTY	B3	421	NA

OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location
2018-09-02 13:00:00	2018	9	Sunday	13	Part One	LINCOLN ST	42.35779	-71.13937	(42.35779134, -71.13937053)
2018-08-21 00:00:00	2018	8	Tuesday	0	Part Two	HECLA ST	42.30682	-71.06030	(42.30682138, -71.06030035)
2018-09-03 19:27:00	2018	9	Monday	19	Part Three	CAZENOVE ST	42.34659	-71.07243	(42.34658879, -71.07242943)
2018-09-03 21:16:00	2018	9	Monday	21	Part Three	NEWCOMB ST	42.33418	-71.07866	(42.33418175, -71.07866441)
2018-09-03 21:05:00	2018	9	Monday	21	Part Three	DELHI ST	42.27537	-71.09036	(42.27536542, -71.09036101)

Desta forma já foi possível identificar características como na coluna *SHOOTING* contém muitos dados perdidos. Sabendo disso, foi feito um cálculo para quantificar a quantidade de dados vazios em cada coluna, o resultado assim como na análise simples visual, o coluna *SHOOTING* apresentou o maior número de

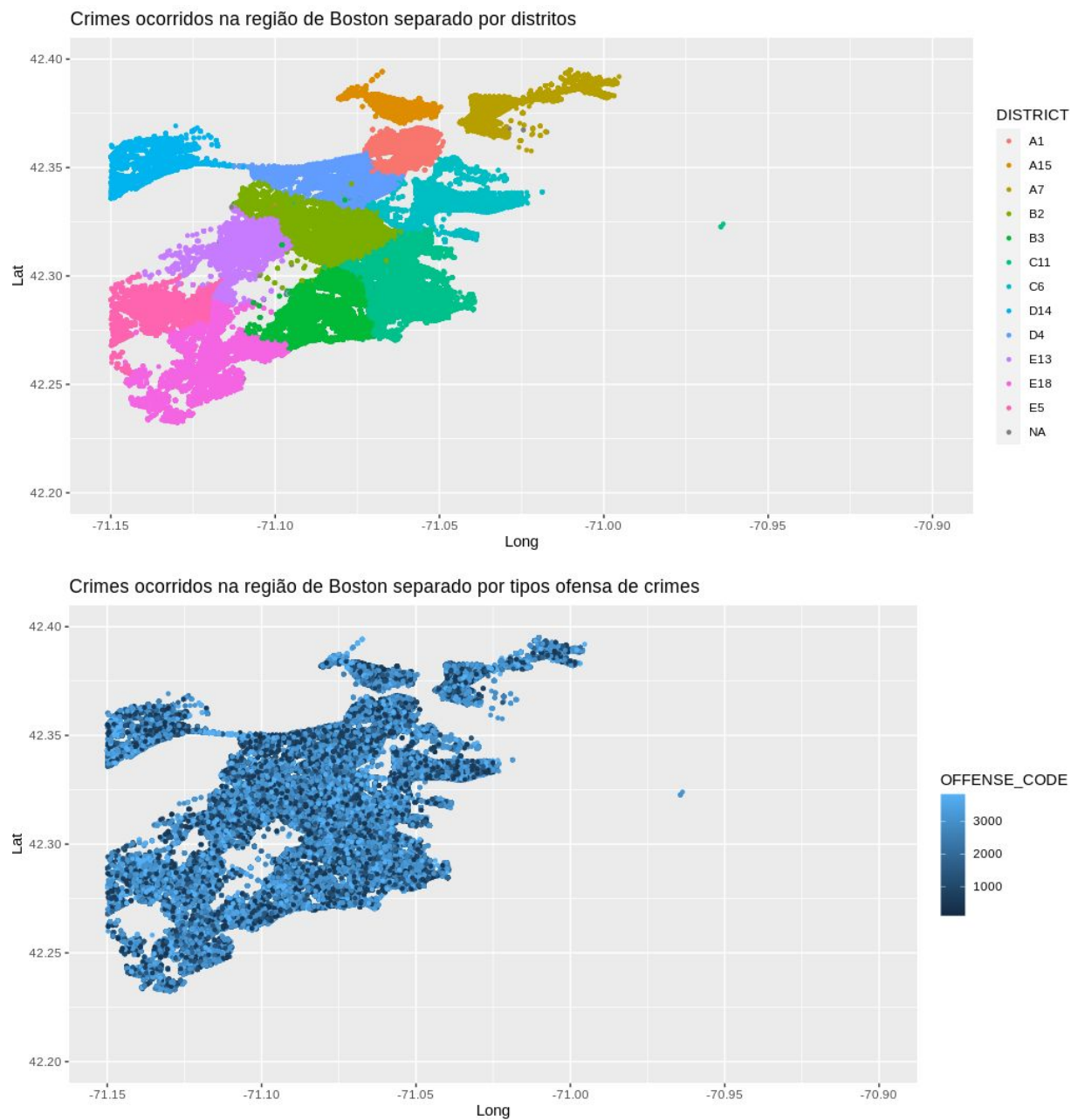
valores não disponíveis do dataset com 318054 dados perdidos, seguidos de *REPORTING ÁREA* com 20250 e *LAT* e *LONG* com 19999 ambos.

INCIDENT_NUMBER	0
OFFENSE_CODE	0
OFFENSE_CODE_GROUP	0
OFFENSE_DESCRIPTION	0
DISTRICT	1765
REPORTING_AREA	20250
SHOOTING	318054
OCCURRED_ON_DATE	0
YEAR	0
MONTH	0
DAY_OF_WEEK	0
HOUR	0
UCR_PART	90
STREET	10871
Lat	19999
Long	19999
Location	0

Na exploração inicial foi feito um mapa de crimes separados por distritos e uma mapa de crimes separados por tipo de ofensa a partir das medidas de longitude e latitude. Foi feito um levantamento do menor e maior índice, encontrado, valores discrepantes relação ao todo como o -1, podendo ser outliers para algum eixo.

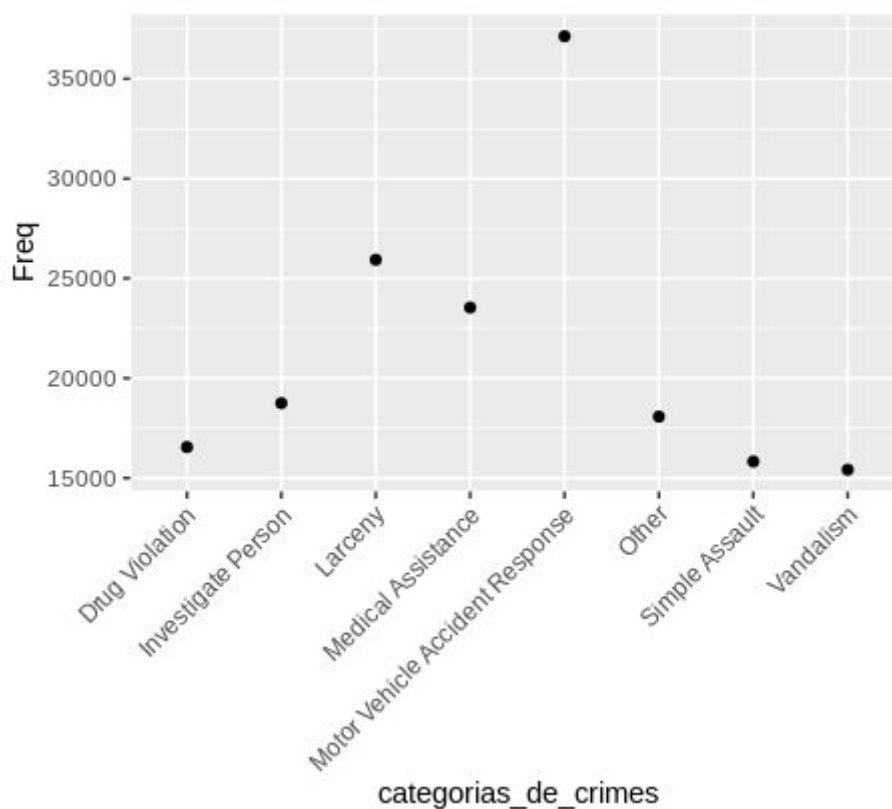
```
maior_lat    42.39504158
maior_long   -1
menor_lat    -1
menor_long   -71.17867378
```

Escala adotada foi longitude -71.15 a -70.90 e latitude 42.2 e 42.4

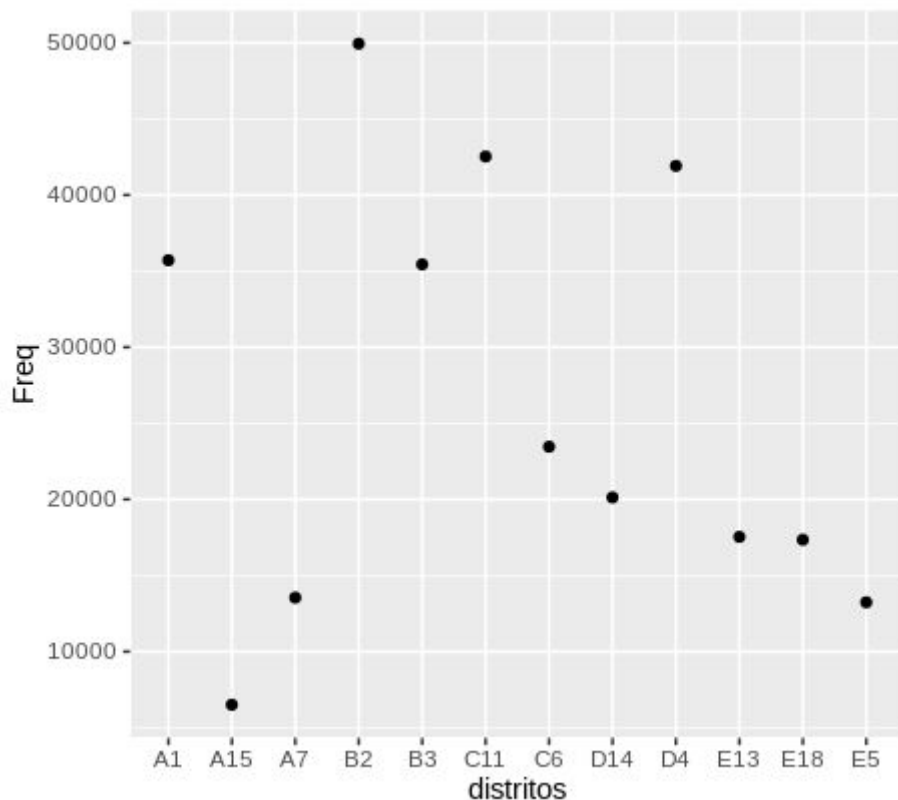


Após esses plots que mapearam a cidade de Boston por pontos foi analisado que existem distritos marcados como NA que pertencem ao distrito A7 pela localidade dos pontos e que variadas tipos de ofensas ocorrem por diversas regiões de boston, ou seja, inicialmente todas regiões pode acontecer todos os tipos de ofensas criminais.

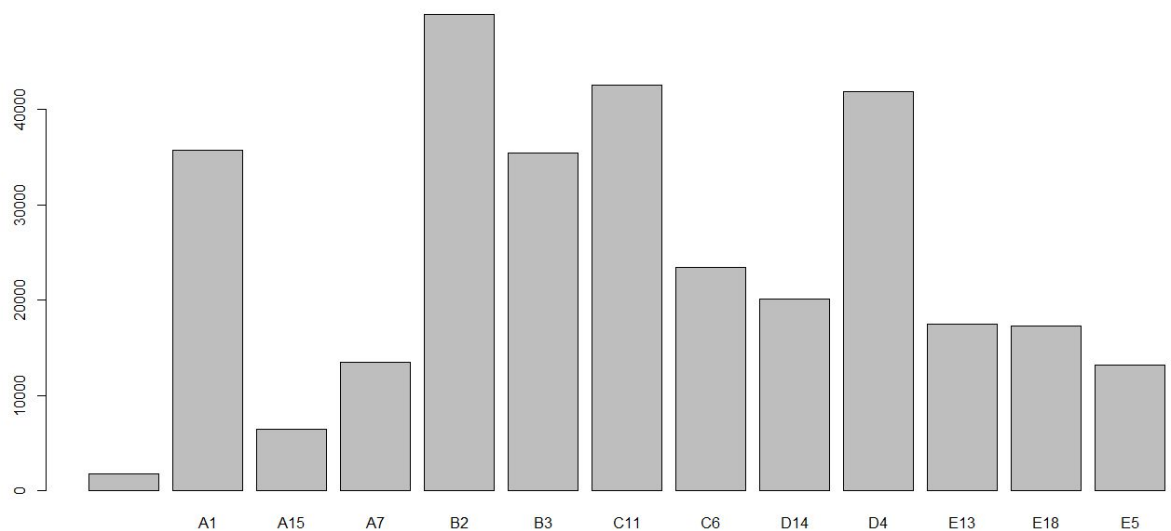
Para ter uma dimensão melhor das ocorrências de crimes foi aplicado medidas estática de exploração de dados relacionando a frequência de ocorrência de crimes pelas categorias de crimes e por distrito, com mais gráficos que permitiram uma melhor análise.



Observações	Código
<p>Ao filtrar buscando as maiores ocorrências, com a tendência a acidentes de veículo motorizado.</p> <p>O crime de roubo também apresentou uma frequência muito alta sendo a segunda posição mais alta ficando acima de 25000 casos.</p>	<pre>categorias_de_crimes <- factor(crimes\$OFFENSE_CODE_GROUP) categorias_de_crimes categoryTable <- table(categorias_de_crimes) converted <- as.data.frame(categoryTable) crimes_mais_cometidos <- filter(converted, Freq > 15000) ggplot(crimes_mais_cometidos ,aes(x = categorias_de_crimes, y = Freq)) + geom_point() + theme(axis.text.x = element_text(angle = 45, hjust = 1))</pre>

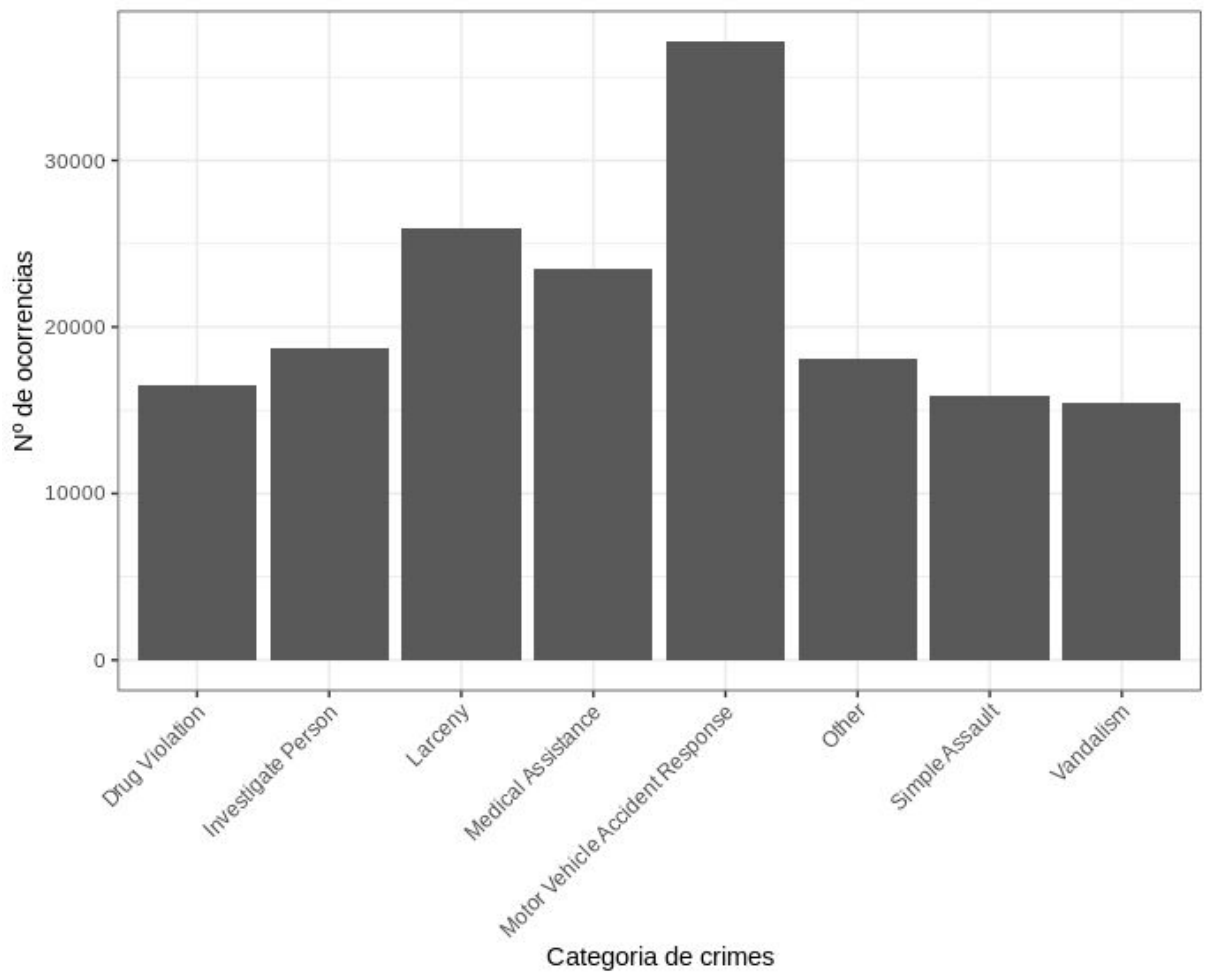


Observações	Código
<p>Observamos a distribuição de frequência dos distritos , indicando o distrito B2 liderando como o de maior ocorrência.</p> <p>Desta forma os distritos mais violentos são B2 , C11 e D4 respectivamente considerando todos os tipos de ocorrências.</p>	<pre>distritos <- factor(crimes\$DISTRICT) distritosTable <- table(distritos) convertedDistrict <- as.data.frame(distritosTable) convertedDistrict</pre>



Observações	Código
<p>Como o gráfico de pontos não foi a melhor escolha visual, mudamos para histograma.</p> <p>Confirmando assim que os os distritos mais violentos são B2 , C11 e D4 respectivamente considerando todos os tipos de ocorrências.</p>	<pre>barplot(convertedDistrict\$Freq, names.arg = convertedDistrict\$distritos)</pre>

Crimes com mais ocorrências



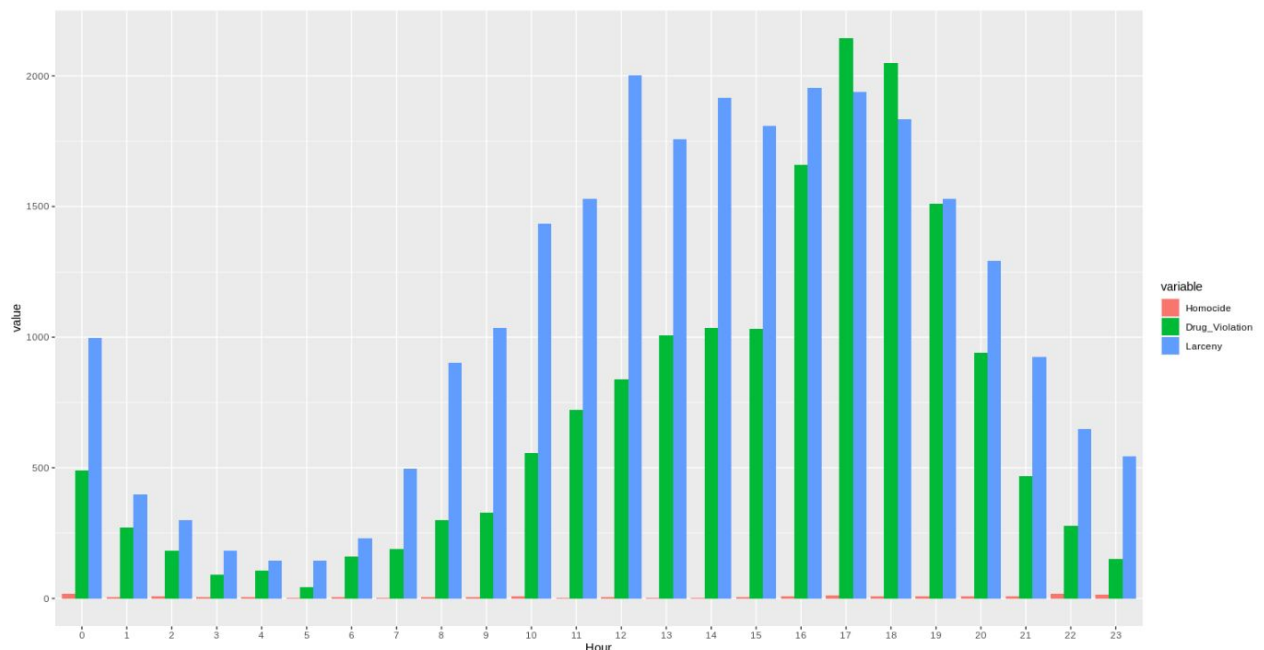
Observações	Código
Como o gráfico de pontos não foi a melhor escolha visual, mudamos para histograma.	<pre>ggplot(crimes_mais_cometidos,aes(x = categorias_de_crimes, y = Freq)) + geom_bar(stat='identity') + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + labs(title="Crimes com mais ocorrências", x="Categoria de crimes", y="Nº de ocorrências")</pre>

2. QUANDO OS CRIMES OCORREM?

Observações	Código
Para iniciar o estudo de gráficos em cima das features indicadas, primeiramente filtramos e remodelamos os dados necessários do dataset com o seguinte código.	<pre># Identificar todos os tipos de crimes presentes unique(crimes\$OFFENSE_CODE_GROUP) # Filtro de todos os crimes relacionado a homicidio, drogas e roubo homicideFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Homicide") drugFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Drug Violation") larcenyFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Larceny") # Tabela de quantidade de crimes de homicidio, drogas e roubo por hora do dia homicide_hour<-as.data.frame(table(unlist(homicideFilter\$HOUR))) drug_hour<-as.data.frame(table(unlist(drugFilter\$HOUR))) larceny_hour<-as.data.frame(table(unlist(larcenyFilter\$HOUR))) # Tabela de quantidade de crimes de homicidio, drogas e roubo por dia da semana homicide_day_of_week<-as.data.frame(table(unlist(homicideFilter\$DAY_OF_WEEK))) drug_day_of_week<-as.data.frame(table(unlist(drugFilter\$DAY_OF_WEEK))) larceny_day_of_week<-as.data.frame(table(unlist(larcenyFilter\$DAY_OF_WEEK))) # Criando dataframes estruturado por horas e dias da semana para os crimes filtrados df_crimes_hour <- data.frame(Hour = homicide_hour\$Var1, Homocide = homicide_hour\$Freq, Drug_Violation = drug_hour\$Freq, Larceny = larceny_hour\$Freq)</pre>

	<pre>df_crimes_day_of_week <- data.frame(Hour = homicide_day_of_week\$Var1, Homocide = homicide_day_of_week\$Freq, Drug_Violation = drug_day_of_week\$Freq, Larceny = larceny_day_of_week\$Freq)</pre>
--	--

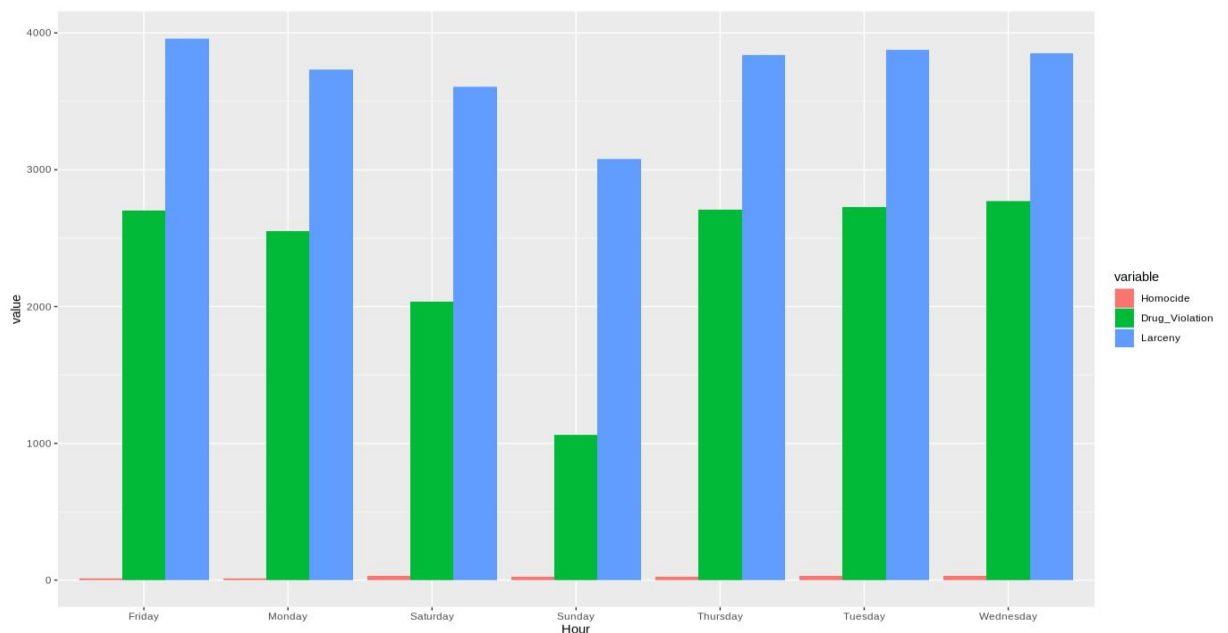
- Para o crime de assassinato, drogas (drug violation) e furto (larceny), plote um gráfico que apresente índices de acordo com os horários do dia (independente do ano).



Observações	Código
<p>A tendência horas no dia tem aumento de frequência numa faixa de 16 - 19 horas do dia. Mostrando um pico maior nessa faixa do que nas demais horas do dia.</p> <p>Em geral esses crimes têm grande baixa entre 1h às 7h, devido o tempo que as pessoas em geral estão nos seus lares dormindo.</p> <p>É possível notar também</p>	<pre>df_hour <- melt(df_crimes_hour, id.vars='Hour') ggplot(df_hour, aes(x=Hour, y=value, fill=variable)) + geom_bar(stat='identity', position='dodge')</pre>

que as 12h há uma crescente no crime de roubo, uma das possíveis causa pode ser o horário de almoço, onde as pessoas saem mais para a rua se tornando vítimas de tal crime.

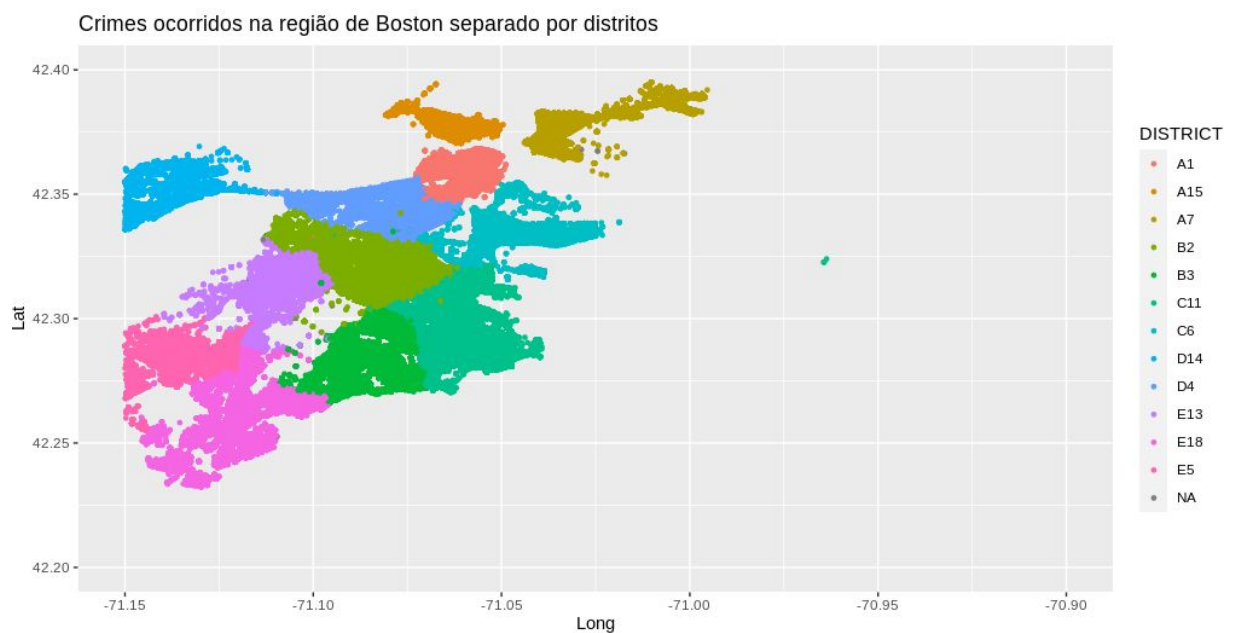
- Para o crime de assassinato, drogas e roubo, plote um gráfico que apresente índices de acordo com os dias da semana (independente do ano).



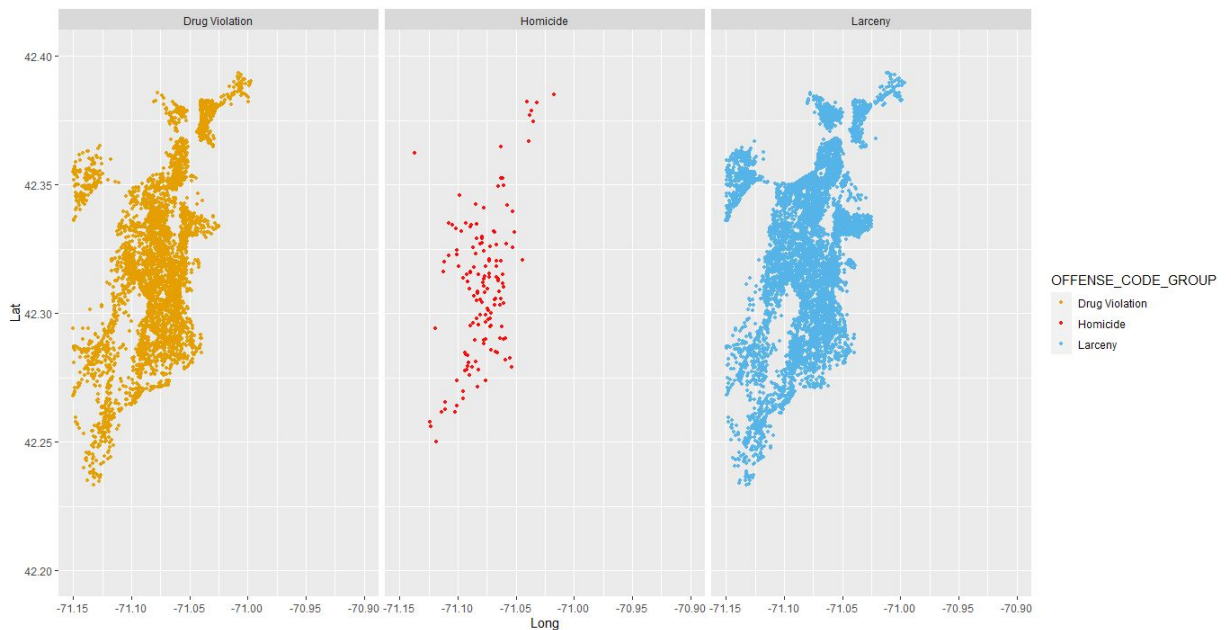
Observações	Código
A tendência em dias de semana se manteve linear para os as 3 ocorrências de crimes. Notamos essa leve diminuição na ocorrências nos fim de semana, mais evidenciado essa queda aos domingos tanto para drogas(drug violation) e furto	<pre>df_day_of_week <- melt(df_crimes_day_of_week, id.vars='Hour' ggplot(df_day_of_week, aes(x=Hour, y=value, fill=variable)) + geom_bar(stat='identity', position='dodge')</pre>

3. ONDE OS CRIMES OCORREM?

- Para os crimes de roubo, assassinato e drogas, plote um gráfico de pontos (Latitude e longitude), apresentando locais onde ocorrem mais crimes. Plote um gráfico para cada tipo de crime.
- Exemplo do mapa colorido por distrito para identificação dos gráficos gerados.



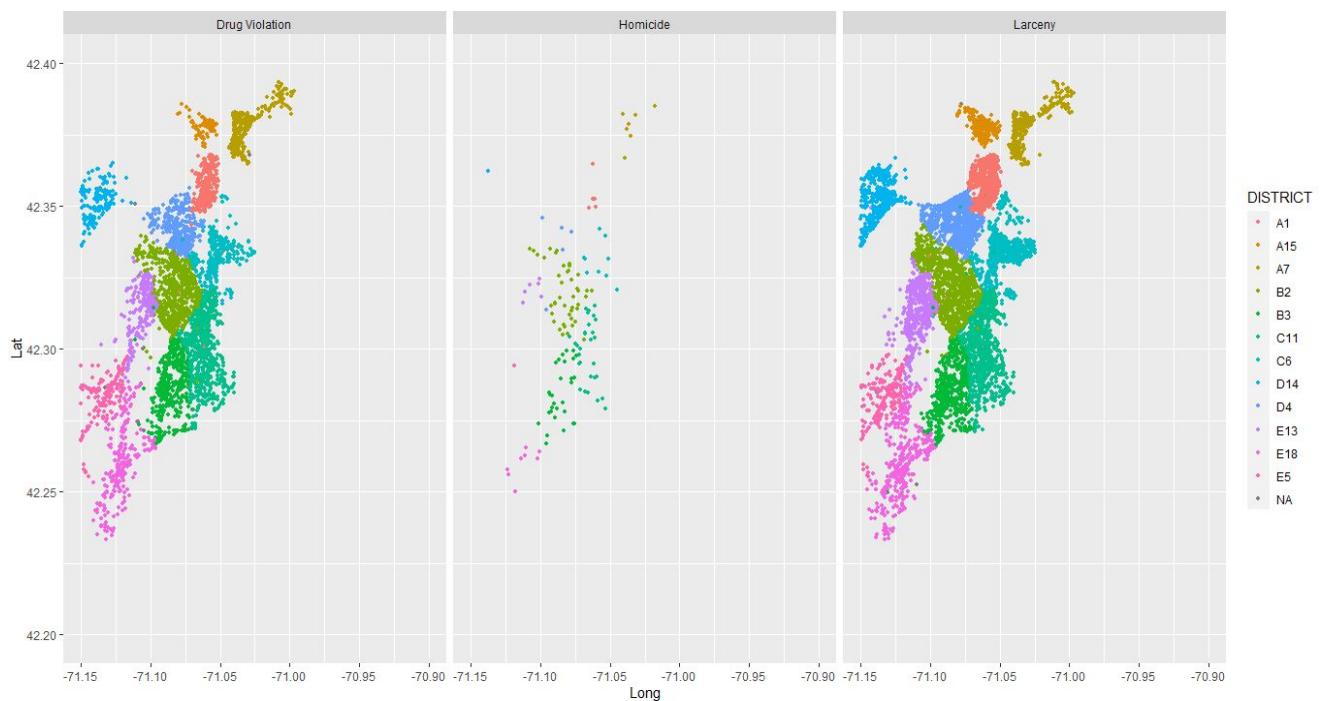
- Para os crimes de roubo, assassinato e drogas, plote um gráfico de pontos (Latitude e longitude), apresentando locais onde ocorrem mais crimes. Plote um gráfico para cada tipo de crime.



- Existe algum tipo de concentração?

Observações	Código
<p>Para os crimes de roubo e drogas há uma concentração bem similar aos distritos do centro de Boston, a distribuição também se espalha de maneira muito similar aos distritos longe do centro mais com um grau menor de concentração. Já para crimes de assassinato também há um espalhamento maior na região central. Pode-se dizer que dada ao centro de Boston possuir um desenvolvimento maior e posteriormente maior população, explicaria a concentração de crimes recorrentes.</p>	<pre> geralFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Homicide" OFFENSE_CODE_GROUP == "Drug Violation" OFFENSE_CODE_GROUP == "Larceny") grafico <- ggplot(geralFilter, aes(x = Long, y = Lat, group=OFFENSE_CODE_GROUP, color = OFFENSE_CODE_GROUP))+ geom_point(size = 1.2, alpha = 0.9) + xlim(-71.15, -70.90) + ylim(42.2, 42.4) + facet_grid(.~OFFENSE_CODE_GROUP) grafico + scale_color_manual(values=c("#E69F00", "red", "#56B4E9")) </pre>

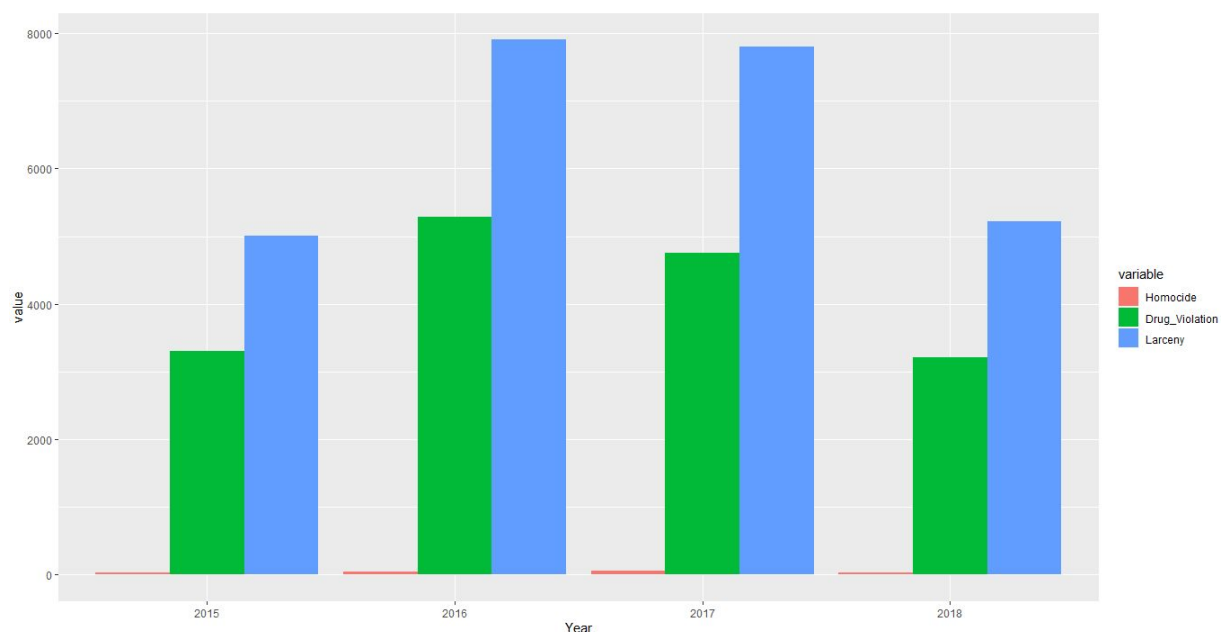
- Apresente este mesmo gráfico com os pontos coloridos pela região.



Observações	Código
Disparadamente o distrito B2 possui uma grande concentração do crimes de roubos e drogas. O Distrito A15 é afastado do centro mais possui uma concentração considerável de roubo. Para o crime de homicídio a concentração é maior também no centro de boston. O motivo principal pode ser o desenvolvimento urbano nesses distritos.	<pre> geralFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Homicide" OFFENSE_CODE_GROUP == "Drug Violation" OFFENSE_CODE_GROUP == "Larceny") ggplot(geralFilter, aes(x = Long, y = Lat, group=DISTRICT, colour = DISTRICT))+ geom_point(size = 1, alpha = 1) + xlim(-71.15, -70.90) + ylim(42.2,42.4) + facet_grid(.~OFFENSE_CODE_GROUP) </pre>

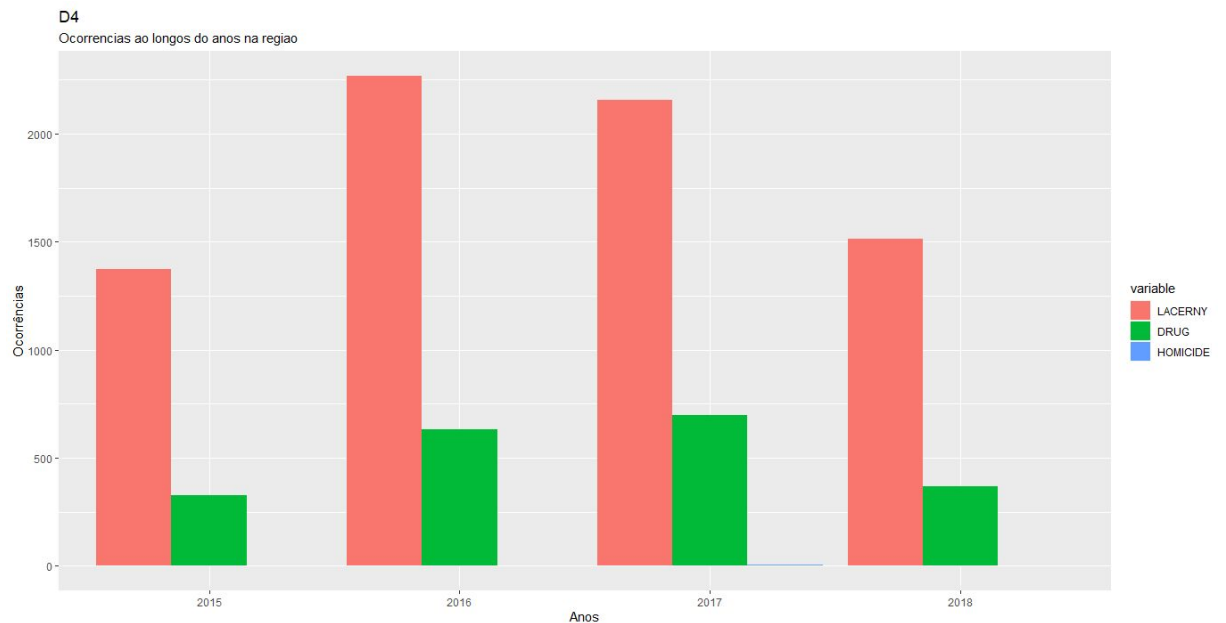
4. EXPLORANDO OS DADOS TEMPORAIS

- Para os crimes tratados anteriormente, plote séries temporais que apresentem como estes crimes tem evoluído ao longo dos anos, no geral.



Observações	Código
<p>A frequência dos crimes teve a seguinte relação, roubo aparece com mais ocorrência, seguido de drogas com uma frequência também alto ao longo do tempo e seguida de uma taxa relativamente baixa de homicídios. Destaque para os anos de 2016 e 2017 onde todos os três tipos de crimes obtiveram uma alta.</p>	<pre> homicide_year<-as.data.frame(table(unlist(homicideFilter\$ YEAR))) drug_year<-as.data.frame(table(unlist(drugFilter\$YEAR))) larceny_year<-as.data.frame(table(unlist(larcenyFilter\$YE AR))) crimesYears <- data.frame(Year = homicide_year\$Var1, Homocide = homicide_year\$Freq, Drug_Violation = drug_year\$Freq, Larceny = larceny_year\$Freq) View(crimesYears) df_years <- melt(crimesYears, id.vars='Year') View(df_years) ggplot(df_years, aes(x=Year, y=value, fill=variable)) + geom_bar(stat='identity', position='dodge') </pre>

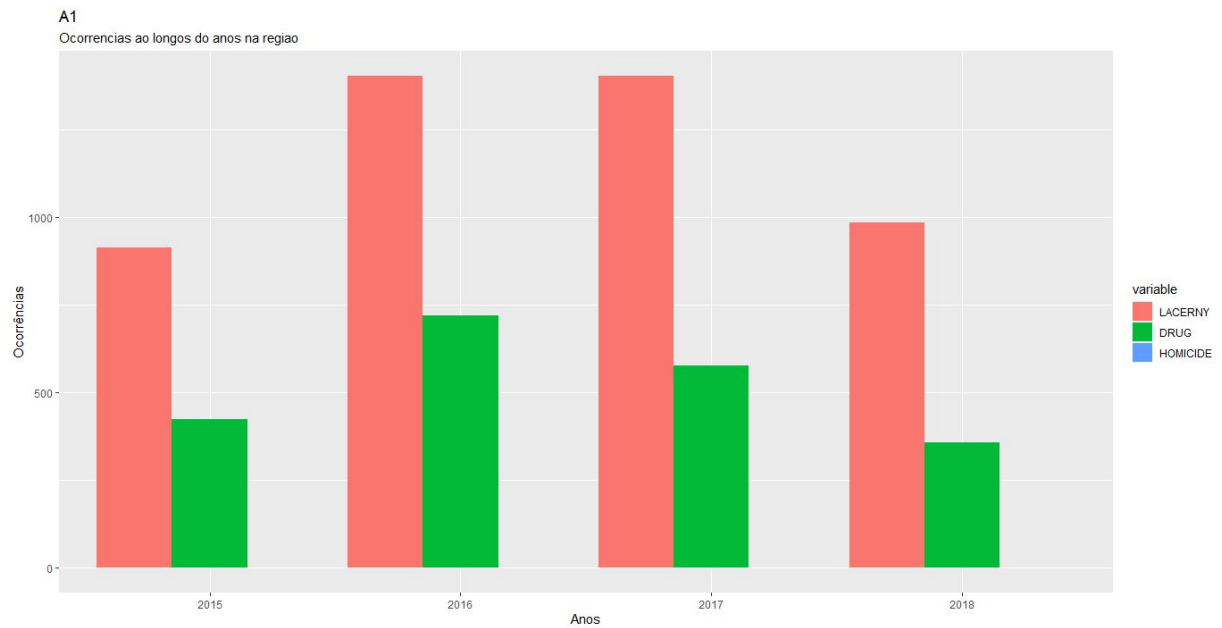
- Para os crimes tratados anteriormente, plote séries temporais, considerando apenas as 3 regiões mais violentas.



Código

```
#Preparação dos dados para todos os crimes
geral_distric<-as.data.frame(table(unlist(geralFilter$DISTRICT)))
distOrder <- geral_distric[order(- geral_distric$Freq), ]

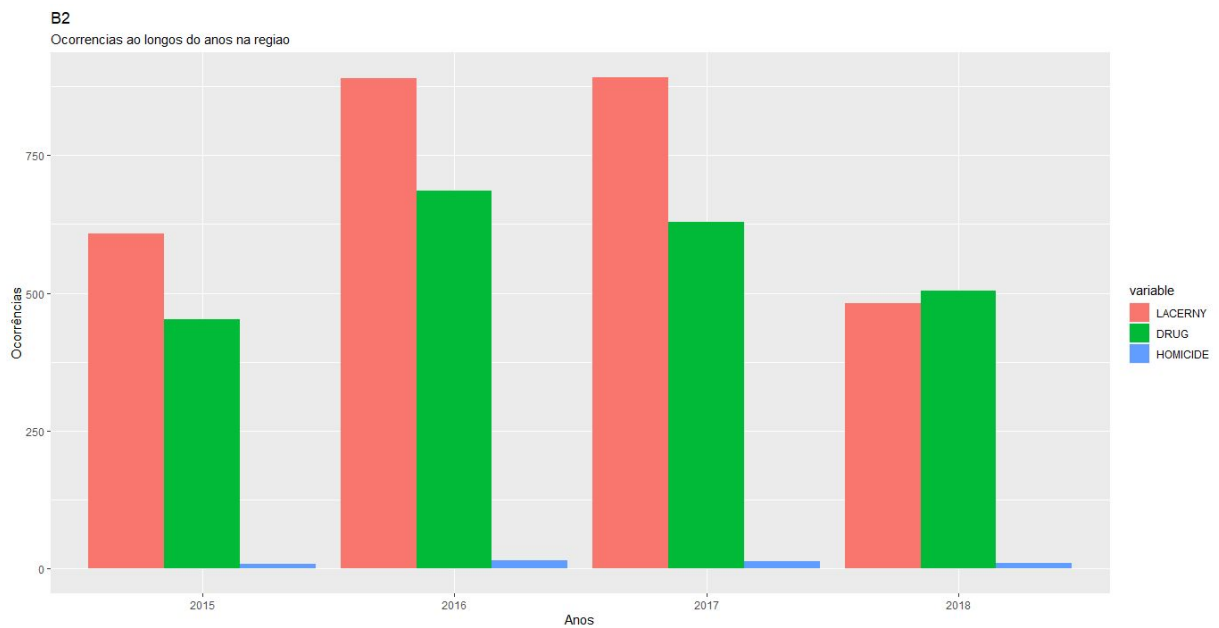
#3 Distritos mais violentos
firstViolent <- distOrder$Var1[1]
firstViolent_graph <- plot_temporal_grap(firstViolent,
geralFilter)
```

Código

```
#Preparação dos dados para todos os crimes
geral_distric<-as.data.frame(table(unlist(geralFilter$DISTRICT)))
distOrder <- geral_distric[order(- geral_distric$Freq), ]

secondViolent <- distOrder[2,1]
secondViolent_graph <- plot_temporal_grap(secondViolent,
geralFilter
```

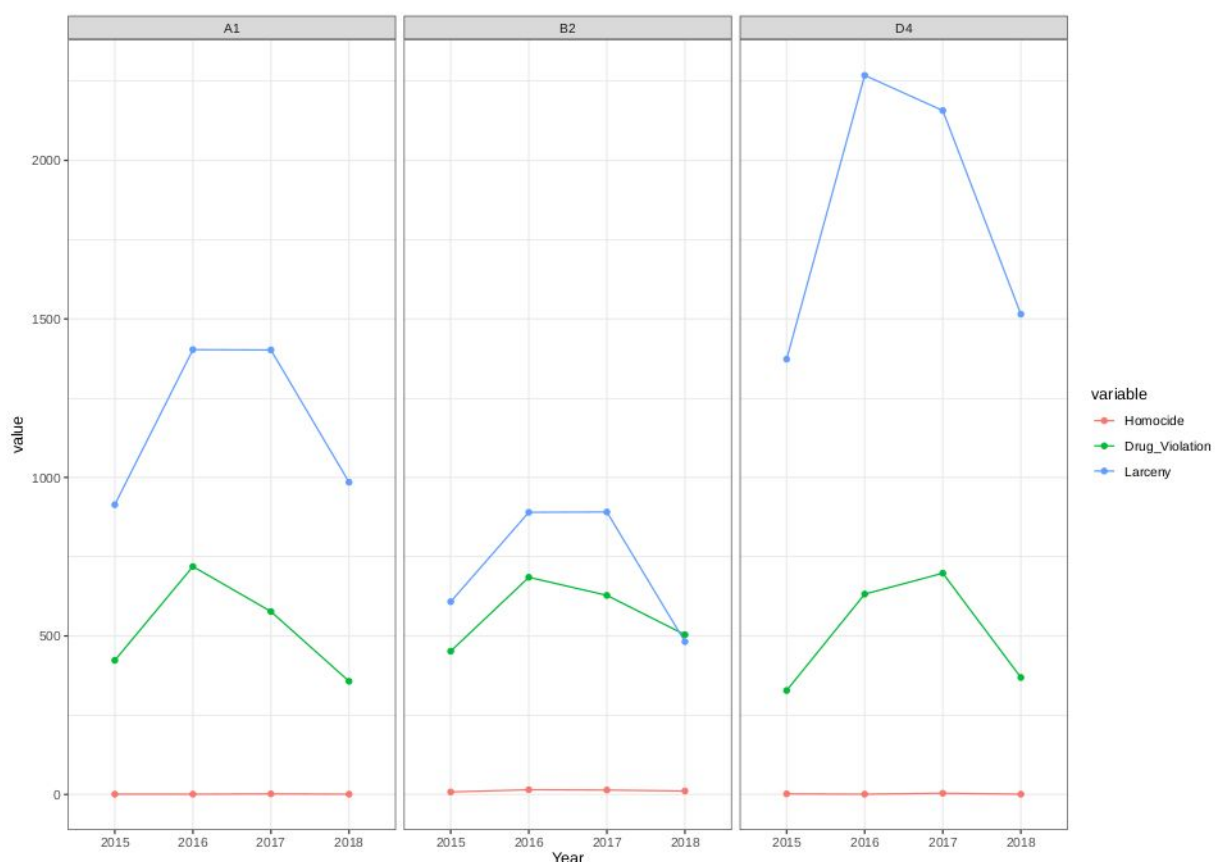


Código

```
#Preparação dos dados para todos os crimes
geral_distric<-as.data.frame(table(unlist(geralFilter$DISTRICT)))
distOrder <- geral_distric[order(- geral_distric$Freq), ]

thirdViolent <- distOrder[3,1]

thirdViolent_graph <- plot_temporal_grap(thirdViolent, geralFilter)
```



Observações	Código (Parcial - apenas o plot do gráfico)
<p>De 2015 a 2018 em todas distrito analisadas no caso os 3 mais violentas apresentam uma estabilidade para o crime de homicídio, sendo que o distrito B2 permanece com leve maior número ocorrências em relação aos outros ao decorrer dos tempo.</p> <p>Em 2016 é bem marcante o elevado número de casos de drogas em ambas regiões, assim como uma queda nos próximos anos indicando que o estado adotou medidas para o combate desse tipo de crime no ano de 2016.</p>	<pre># Gráfico de série temporal do tipo linha separado em grid dos 3 distritos mais violentos ggplot(data=new_df, aes(x=Year, y=value, group=variable, color=variable)) + geom_line() + geom_point() + facet_grid(.~District)</pre>

<p>Podemos também notar o mesmo comportamento para o crime de roubo que tem um número de ocorrência muito elevado, principalmente se considerar o distrito D4. De 2015 para 2016 teve um grande aumento nos casos chegando ao pico de ocorrências, e da mesma forma para o crime de drogas de 2016 a 2018 houve uma queda nesse tipo de criminalidade indicando possíveis medidas de segurança adotada no estado de Boston a partir desse período que contribui para tal queda.</p>	
---	--

- **Comente os resultados obtidos. Há algum aumento ou diminuição dos crimes em alguma região?**

Observações
<p>Os distritos de A1 (<i>segundo distrito com mais crimes</i>) e D4 (<i>distrito com mais crimes</i>), obtiveram um comportamento parecido com o da cidade de Boston como um todo, sendo os anos de 2016 e 2017 os anos com uma alta frequência de ocorrências, com os casos mais recorrentes de roubos e drogas.</p> <p>O distrito B2 (<i>terceiro distrito com mais crimes</i>) registrou uma diferença no ano de 2018, sendo uma grande queda no número de ocorrências relacionadas a roubo, ficando atrás do número de casos relacionados a drogas, o que não se vê nos demais distritos.</p>

5. CONCLUSÃO

Concluimos o trabalho destacando o amadurecimento na linguagem R que obtivemos durante o desenvolvimento do projeto, tentamos ao máximo otimizar o código, pesquisar novos conceitos e implementar ao contexto do trabalho. Trabalhamos o conceito de visualização de dados com diversos gráficos visando sempre extrair o conhecimento correto dos mesmos.

Nosso maior desafio foi trabalhar com dados ordinais e nominais, onde até então não tínhamos contato com um dataset que fosse estruturado dessa forma, conseguimos chegar a uma forma consistente de manipular o dataset para gerar sub datasets e a partir desse ponto seguir normalmente com o estudo.

Sobre o conceito de visualização de dados, vimos a importância de se usar a melhor tratativa possível gerando uma interpretação correta e intuitiva dos dados. Vimos que para uma mesma pergunta, diferentes gráficos podem gerar diferentes interpretações e gerar conflito no ponto de vista.

Sobre o dataset analisado, Boston nos pareceu uma cidade com um índice de criminalidade elevado, mas com baixo índice de homicídios e a última questão nos mostra queda nos distritos com maior número de casos, isso nos leva a interpretar que a cidade tomou medidas mais brandas para combater os casos de crimes da cidade.

6. Código

```
library("tidyr")
library(dplyr)
library(ggplot2)
library(dslabs)
library(reshape2) # melt a data.frame

# setwd("~/Documents/topicos_BD/Projeto1 - Crimes in Boston")

#----- Leitura do CSV -----
#-----
crimes <- read.csv("crime.csv", sep=",", na.strings =
c(' ', 'NA', 'na', 'N/A', 'n/a', 'NaN', 'nan'))

# ----- Exploração do dataset -----
nrow(crimes)
```

```

ncol(crimes)
dim(crimes)

View(head(crimes, n=5))
str(crimes)

# Após a visualização de uma amostra do dataset com 10 linhas
# foi possível perceber valores vazios para coluna SHOOTING,
# então foi realizado um exploração para entender o número de dados
vazios nas features.
# Desta forma SHOTTING apresentou o maior número de dados vazios com
318054 registros NA.

View(sapply(crimes, function(x) sum(is.na(x))))

# Retorna a menor e maior data de ocorrência de crime presente no
dataset
min(crimes$OCCURRED_ON_DATE)
max(crimes$OCCURRED_ON_DATE)

# ----- Exploração inicial com medidas estatísticas
head(crimes$YEAR, 10)
tail(crimes$YEAR, 10)
data_2015_2010<- filter(crimes, YEAR%in%c(2015, 2010))

# Retornar o menor e o maior índice de latitude e longitude para
realizar a escala do mapa de crimes de boston
maior_lat = max(crimes$Lat, na.rm=T)
menor_lat = min(crimes$Lat, na.rm=T)

maior_long = max(crimes$Long, na.rm=T)
menor_long = min(crimes$Long, na.rm=T)

# Mapa dos crimes agrupado por distrito de Boston
ggplot(crimes, aes(x = Long, y = Lat, group=DISTRICT, colour =
DISTRICT))+
  geom_point(size = 1, alpha = 1) +
  xlim(-71.15, -70.90) +
  ylim(42.2,42.4) +
  ggtitle("Crimes ocorridos na região de Boston separado por distritos")

# Mapa dos crimes agrupados por tipo de ofensa

```

```

ggplot(crimes, aes(x = Long, y = Lat, group=OFFENSE_CODE, colour =
OFFENSE_CODE))+
  geom_point(size = 1, alpha = 1) +
  xlim(-71.15, -70.90) +
  ylim(42.2, 42.4) +
  ggtitle("Crimes ocorridos na região de Boston separado por tipos
ofensa de crimes")

# Nesse decidimos aplicar algo para calcular frequência de dados
nominais/ordinais
# Chamados com mais frequência
categorias_de_crimes <- factor(crimes$OFFENSE_CODE_GROUP)
categorias_de_crimes
categoryTable <- table(categorias_de_crimes)
converted <- as.data.frame(categoryTable)

crimes_mais_cometidos <- filter(converted, Freq > 15000)
crimes_mais_cometidos

ggplot(crimes_mais_cometidos ,aes(x = categorias_de_crimes, y = Freq))
+
  geom_point() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Histogramas de tipos de ocorrencia
barplot(crimes_mais_cometidos$Freq,
names.arg =
crimes_mais_cometidos$categorias_de_crimes)

# Histograma de tipos de ocorrencia com rotação na label x para melhor
visualização
ggplot(crimes_mais_cometidos,aes(x = categorias_de_crimes, y = Freq)) +
  geom_bar(stat='identity') +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  labs(title="Crimes com mais ocorrencias", x="Categoria de crimes",
y="N° de ocorrencias")

# Distritos com mais ocorrências
distritos <- factor(crimes$DISTRICT)
distritosTable <- table(distritos)
convertedDistrict <- as.data.frame(distritosTable)
convertedDistrict

ggplot(convertedDistrict ,aes(x = distritos, y = Freq)) +

```

```

geom_point()

ggplot(convertedDistrict, aes(x = distritos, y = Freq)) +
  geom_point()

# ----- Análises -----
# Insight buscado -> Quando os crimes ocorrem?

# Identificar todos os tipos de crimes presentes
unique(crimes$OFFENSE_CODE_GROUP)

# Filtro de todos os crimes relacionado a homicídio, drogas e roubo
homicideFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Homicide")
drugFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Drug Violation")
larcenyFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Larceny")

# Tabela de quantidade de crimes de homicídio, drogas e roubo por hora
do dia
homicide_hour<-as.data.frame(table(unlist(homicideFilter$HOURL)))
drug_hour<-as.data.frame(table(unlist(drugFilter$HOURL)))
larceny_hour<-as.data.frame(table(unlist(larcenyFilter$HOURL)))

# Tabela de quantidade de crimes de homicídio, drogas e roubo por dia
da semana
homicide_day_of_week<-as.data.frame(table(unlist(homicideFilter$DAY_OF_WEEK)))
drug_day_of_week<-as.data.frame(table(unlist(drugFilter$DAY_OF_WEEK)))
larceny_day_of_week<-as.data.frame(table(unlist(larcenyFilter$DAY_OF_WEEK)))

# Criando dataframes estruturado por horas e dias da semana para os
crimes filtrados
df_crimes_hour <- data.frame(Hour = homicide_hour$Var1, Homocide =
homicide_hour$Freq, Drug_Violation = drug_hour$Freq, Larceny =
larceny_hour$Freq )
df_crimes_day_of_week <- data.frame(Hour = homicide_day_of_week$Var1,
Homocide = homicide_day_of_week$Freq, Drug_Violation =
drug_day_of_week$Freq, Larceny = larceny_day_of_week$Freq )

View(df_crimes_hour)

#when you melt essentially you create only one column with the value

```



```

#and one column with the variable i.e. your others columns
df_hour <- melt(df_crimes_hour, id.vars='Hour')
df_day_of_week <- melt(df_crimes_day_of_week, id.vars='Hour')

View(df_hour)

# Gráfico de barras crimes filtrados por hora e por dia da semana
ggplot(df_hour, aes(x=Hour, y=value, fill=variable)) +
  geom_bar(stat='identity', position='dodge')

ggplot(df_day_of_week, aes(x=Hour, y=value, fill=variable)) +
  geom_bar(stat='identity', position='dodge')

# -----
# Insight procurado -> Onde os crimes ocorrem?

# Filtro de todos os crimes relacionado a homicidio, drogas e roubo
homicideFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Homicide")
drugFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Drug Violation")
larcenyFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Larceny")
geralFilter <- filter(crimes, OFFENSE_CODE_GROUP == "Homicide" |
OFFENSE_CODE_GROUP == "Drug Violation" | OFFENSE_CODE_GROUP ==
"Larceny")

# Mapa dos crimes agrupados por crimes de roubo, assassinato e drogas
# apresentação de forma unida no mesmo ambiente
grafico <- ggplot(geralFilter, aes(x = Long, y = Lat,
group=OFFENSE_CODE_GROUP, color = OFFENSE_CODE_GROUP)) +
  geom_point(size = 1.2, alpha = 0.9) +
  xlim(-71.15, -70.90) +
  ylim(42.2, 42.4)

grafico + scale_color_manual(values=c("#E69F00", "red", "#56B4E9"))

# Mapa de crimes de roubo, assassinato e drogas plotadas de forma
separada
grafico <- ggplot(geralFilter, aes(x = Long, y = Lat,
group=OFFENSE_CODE_GROUP, color = OFFENSE_CODE_GROUP)) +
  geom_point(size = 1.2, alpha = 0.9) +
  xlim(-71.15, -70.90) +
  ylim(42.2, 42.4) +

```

```

facet_grid(.~OFFENSE_CODE_GROUP)

grafico + scale_color_manual(values=c("#E69F00", "red", "#56B4E9"))
barplot(convertedDistrict$Freq, names.arg = convertedDistrict$distritos)

#Apresente este mesmo gráfico com os pontos coloridos pela região.
ggplot(geralFilter, aes(x = Long, y = Lat, group=DISTRICT, colour = DISTRICT)) +
  geom_point(size = 1, alpha = 1) +
  xlim(-71.15, -70.90) +
  ylim(42.2, 42.4) +
  facet_grid(.~OFFENSE_CODE_GROUP)

# -----
# Explorando os dados temporais

#Para os crimes tratados anteriormente, plote séries temporais que
#apresentem como estes crimes tem evoluído ao longo dos anos, no geral.

homicide_year<-as.data.frame(table(unlist(homicideFilter$YEAR)))
drug_year<-as.data.frame(table(unlist(drugFilter$YEAR)))
larceny_year<-as.data.frame(table(unlist(larcenyFilter$YEAR)))

crimesYears <- data.frame(Year = homicide_year$Var1, Homocide = homicide_year$Freq, Drug_Violation = drug_year$Freq, Larceny = larceny_year$Freq )
View(crimesYears)
df_years <- melt(crimesYears, id.vars='Year')
View(df_years)
ggplot(df_years, aes(x=Year, y=value, fill=variable)) +
  geom_bar(stat='identity', position='dodge')

ggplot(data=df_years, aes(x=Year, y=value, group=variable, color=variable)) +
  geom_line() +
  geom_point()

#Para os crimes tratados anteriormente, plote séries temporais,
considerando
#apenas as 3 regiões mais violentas.

```

```

#Preparação dos dados para todos os crimes
geral_distric<-as.data.frame(table(unlist(geralFilter$DISTRICT)))
distOrder <- geral_distric[order(- geral_distric$Freq), ]

#3 Distritos mais violentos
firstViolent <- distOrder$Var1[1]
secondViolent <- distOrder[2,1]
thirdViolent <- distOrder[3,1]
print(thirdViolent)

#Função que gera o grafico temporal dado um distrito
plot_temporal_grap <- function(dist, geralFilter) {

  #Gerando os filtros especificos
  lacerny_filter <- filter(geralFilter, DISTRICT == dist &
  OFFENSE_CODE_GROUP == "Larceny")
  drug_filter <- filter(geralFilter, DISTRICT == dist &
  OFFENSE_CODE_GROUP == "Drug Violation")
  homicide_filter <- filter(geralFilter, DISTRICT == dist &
  OFFENSE_CODE_GROUP == "Homicide")

  #Gerando tabelas
  lacerny_year <- as.data.frame(table(unlist(lacerny_filter$YEAR)))

  drug_year <- as.data.frame(table(unlist(drug_filter$YEAR)))
  homicide_year <- as.data.frame(table(unlist(homicide_filter$YEAR)))

  #Gerando a tabela principal que sera usada para gerar os graficos
  main_table <- data.frame(YEAR = lacerny_year$Var1, LACERNY =
  lacerny_year$Freq, DRUG = drug_year$Freq,
  HOMICIDE = homicide_year$Freq)

  #Gerando o gráfico temporal que será retornado
  df_main <- melt(main_table, id.vars='YEAR')
  result <- ggplot(df_main, aes(x=YEAR, y=value, fill=variable)) +
  geom_bar(stat='identity', position='dodge') +
  labs(subtitle = "Ocorrencias ao longos do anos na regioao", title =
  dist) +
  xlab("Anos") +
  ylab("Ocorrências")
  return(result)
}

firstViolent_graph <- plot_temporal_grap(firstViolent, geralFilter)

```

```

secondViolent_graph <- plot_temporal_grap(secondViolent, geralFilter)
thirdViolent_graph <- plot_temporal_grap(thirdViolent, geralFilter)

firstViolent_graph
secondViolent_graph
thirdViolent_graph

# Implementação adicional onde é criado um data frame onde relaciona
crimes tratados anteriormente,
# para plote de séries temporais com gráfico de linha, seprando por
grupos das 3 regiões mais violentas

# Preparação dos dados para todos os crimes
geral_distrib <- as.data.frame(table(unlist(geralFilter$DISTRICT)))
View(geral_distrib)

# D4, A1 e B2 são os 3 distritos com mais ocorrências
more_dist <- filter(geralFilter, DISTRICT == "D4" | DISTRICT == "A1" |
DISTRICT == "B2")

distrib_D4_homicide <- filter(homicideFilter, DISTRICT == "D4")
distrib_D4_drug <- filter(drugFilter, DISTRICT == "D4")
distrib_D4_larceny <- filter(larcenyFilter, DISTRICT == "D4")

distrib_C11_homicide <- filter(homicideFilter, DISTRICT == "A1")
distrib_C11_drug <- filter(drugFilter, DISTRICT == "A1")
distrib_C11_larceny <- filter(larcenyFilter, DISTRICT == "A1")

distrib_B2_homicide <- filter(homicideFilter, DISTRICT == "B2")
distrib_B2_drug <- filter(drugFilter, DISTRICT == "B2")
distrib_B2_larceny <- filter(larcenyFilter, DISTRICT == "B2")

# Tabela de quantidade de crimes de homicídio, drogas e roubo por ano
more_distrib <- as.data.frame(table(unlist(more_dist$YEAR)))
View(more_distrib)

# Tabela de quantidade de crimes de homicídio, drogas e roubo por ano
separado nos 3 distritos mais violentos
dist_D4_homicide_year <-
as.data.frame(table(unlist(distrib_D4_homicide$YEAR)))
dist_D4_drug_year <- as.data.frame(table(unlist(distrib_D4_drug$YEAR)))
dist_D4_larceny_year <-
as.data.frame(table(unlist(distrib_D4_larceny$YEAR)))

```

```

dist_A1_homicide_year <-
as.data.frame(table(unlist(district_A1_homicide$YEAR)))
dist_A1_drug_year <- as.data.frame(table(unlist(district_A1_drug$YEAR)))
dist_A1_larceny_year <-
as.data.frame(table(unlist(district_A1_larceny$YEAR)))

dist_B2_homicide_year <-
as.data.frame(table(unlist(district_B2_homicide$YEAR)))
dist_B2_drug_year <- as.data.frame(table(unlist(district_B2_drug$YEAR)))
dist_B2_larceny_year <-
as.data.frame(table(unlist(district_B2_larceny$YEAR)))

# Criando dataframes estruturado por ano para os crimes filtrados de
acordo com o distrito especificado
df_crimes_D4 <- data.frame(Year = dist_D4_homicide_year$Var1, Homocide
= dist_D4_homicide_year$Freq, Drug_Violation = dist_D4_drug_year$Freq,
Larceny = dist_D4_larceny_year$Freq )
df_crimes_A1 <- data.frame(Year = dist_A1_homicide_year$Var1, Homocide
= dist_A1_homicide_year$Freq, Drug_Violation = dist_A1_drug_year$Freq,
Larceny = dist_A1_larceny_year$Freq )
df_crimes_B2 <- data.frame(Year = dist_B2_homicide_year$Var1, Homocide
= dist_B2_homicide_year$Freq, Drug_Violation = dist_B2_drug_year$Freq,
Larceny = dist_B2_larceny_year$Freq )

# Adicionando coluna referente ao distritos aos data frames
nrow(df_crimes_D4)

total <- rbind(df_crimes_D4, df_crimes_A1)
total <- rbind(total, df_crimes_B2)

new_df <- melt(total, id.vars='Year')

new_df$District<-
c('D4','D4','D4','D4','A1','A1','A1','A1','B2','B2','B2','B2',
'D4','D4','D4','D4','A1','A1','A1','A1','B2','B2','B2','B2',
'D4','D4','D4','D4','A1','A1','A1','A1','B2','B2','B2','B2')
View(new_df)

# Gráfico de série temporal do tipo linha separado em grid dos 3
distritos mais violentos
ggplot(data=new_df,aes(x=Year, y=value, group=variable,
color=variable)) +

```

```
geom_line() +  
geom_point() +  
facet_grid(.~District)
```

7. Referências

<https://www.rdocumentation.org/>

<https://www.kaggle.com/AnalyzeBoston/crimes-in-boston>