

LAPORAN TUGAS 2

SELEKSI ASISTEN BASIS DATA 2016

oleh

Steffi Indrayani 13514063

Scarletta Julia Yapfrine 13514074



PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
BANDUNG
2016

I. Penjelasan Singkat *Dataset*

Dataset yang penulis dapatkan merupakan *dataset* yang berisi data mengenai Kriminal San Francisco (*Criminality on San Francisco*). Beberapa sumber mengatakan bahwa tingkat kriminal di San Francisco relatif tinggi, terutama dibandingkan dengan kota-kota lain di Amerika Serikat. Akan tetapi, dengan *dataset* yang tersedia, penulis harus dapat memastikan validitas pernyataan sumber tersebut. Analisis harus dilakukan berdasarkan *dataset* yang tersedia. *Dataset* yang penulis terima dibagi menjadi dua, yakni *test* dan *train*. Pada berkas data *test*, relasi yang ada memiliki atribut ID kejadian kriminal, tanggal kejadian kriminal, hari kejadian kriminal, distrik kejadian kriminal, alamat kejadian kriminal, absis lokasi kejadian, dan ordinat lokasi kejadian. Pada berkas data *train*, relasi yang ada memiliki atribut tanggal kejadian kriminal, kategori kriminalitas, deskripsi kejadian kriminal, hari kejadian kriminal, distrik kejadian kriminal, resolusi kejadian kriminal, alamat kejadian kriminal, absis lokasi kejadian, dan ordinat lokasi kejadian.

II. *Tools* yang Digunakan

Tools atau kakas yang penulis gunakan dalam melakukan analisis dan visualisasi data dari *dataset* yang telah diberikan adalah Bahasa R. Perangkat lunak yang mendukung analisis dan visualisasi data dengan menggunakan Bahasa R adalah RStudio beserta *packages* yang terkait, yaitu *ggplot2*, *leaflet*, *RColorBrewer*, *readr*, *lubridate*, dan *dplyr*.

III. Langkah-Langkah Analisis

1. Memasukkan *dataset train* dan *test* ke dalam variabel pada RStudio
2. Menambahkan kolom *StreetCorner* pada kedua *dataset*. Kolom *StreetCorner* memiliki nilai “Yes” apabila alamat kejadian kriminal terjadi di ujung jalan dan “No” jika tidak. Penulis mengidentifikasi apakah suatu kejahatan terjadi di ujung jalan dengan cara melihat alamat tempat terjadinya kecelakaan (kolom *Address*). Jika pada kolom *Address* suatu kejahatan terdapat karakter “/” berarti kejahatan terjadi di antara kedua jalan (di ujung dari kedua jalan tersebut)
3. Mengambil data kejahatan yang terjadi di ujung jalan. Dari data-data tersebut, dilihat kategori kejahatan apa saja yang terjadi di ujung jalan. Kejahatan-kejahatan yang terjadi di ujung jalan tersebut kemudian divisualisasikan pada peta San Francisco dan diberi warna sesuai dengan kategori kejahatannya.
4. Menunjukkan tren kejahatan di ujung jalan di San Francisco secara keseluruhan dengan cara menampilkan di *console* kategori dengan jumlah kejahatan paling besar ke paling kecil. Tren kejahatan tersebut juga ditampilkan ke histogram.
5. Bagian keempat dilakukan juga untuk kejahatan perdistrik, pertahun, dan perhari dalam satu minggu.

6. Menggolongkan kategori kejahatan menjadi tiga golongan, yaitu *white-collar crime*, *blue-collar crime*, dan *other crime*. Kejahatan yang sudah digolongkan ke dalam tiga golongan tersebut kemudian dibuat dalam bentuk histogram.

IV. Hasil Analisis dan Visualisasi Hasil Analisis

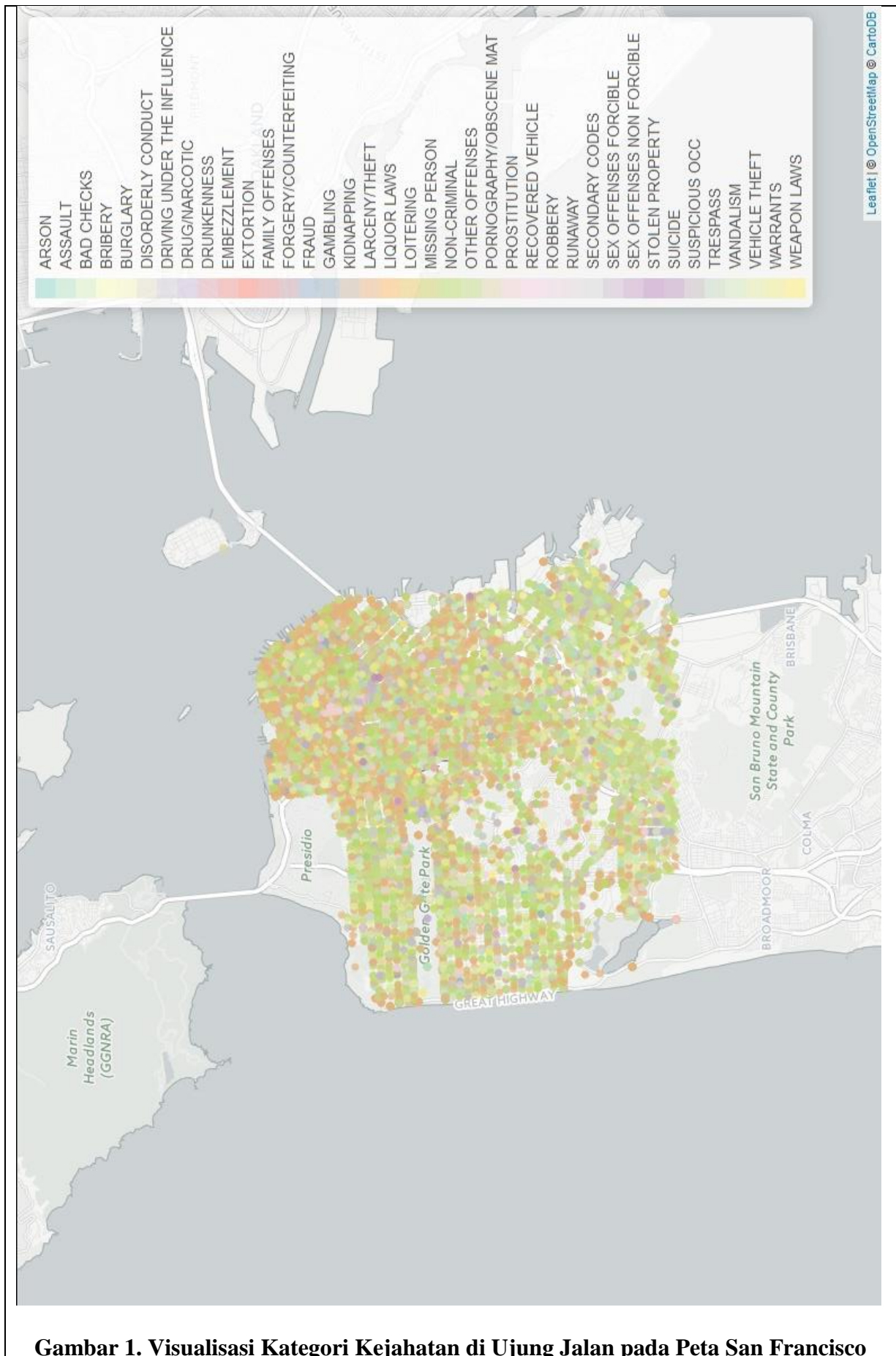
1. Memprediksi kategori-kategori kriminalitas yang terjadi di ujung jalan (*street corner*)

Penulis memunculkan kategori-kategori kriminalitas yang terjadi di ujung jalan pada *console* R-Studio. Definisi ujung jalan adalah alamat yang terdapat lebih dari satu jalan dan jalan-jalan tersebut dituliskan dengan dibatasi tanda garis miring '/'. Dari 37 kategori kejahatan yang terjadi di San Francisco, 36 kategori terjadi di ujung jalan. Kategori TREA tidak terjadi di ujung jalan.

Berikut kategori-kategori kriminalitas yang terjadi di San Francisco yang dimunculkan pada *console* R-Studio

[1] WARRANTS	OTHER OFFENSES
[3] VEHICLE THEFT	LARCENY/THEFT
[5] VANDALISM	ROBBERY
[7] ASSAULT	NON-CRIMINAL
[9] SUSPICIOUS OCC	DRUG/NARCOTIC
[11] SECONDARY CODES	BURGLARY
[13] STOLEN PROPERTY	DRIVING UNDER THE INFLUENCE
[15] SEX OFFENSES FORCIBLE	WEAPON LAWS
[17] MISSING PERSON	FRAUD
[19] DRUNKENNESS	KIDNAPPING
[21] LIQUOR LAWS	DISORDERLY CONDUCT
[23] FORGERY/COUNTERFEITING	SUICIDE
[25] BRIBERY	TRESPASS
[27] ARSON	SEX OFFENSES NON FORCIBLE
[29] PROSTITUTION	EXTORTION
[31] EMBEZZLEMENT	GAMBLING
[33] LOITERING	BAD CHECKS
[35] RUNAWAY	FAMILY OFFENSES
[37] RECOVERED VEHICLE	PORNOGRAPHY/OBSCENE MAT

Kategori-kategori tersebut di atas juga divisualisasikan pada peta San Francisco.



Gambar 1. Visualisasi Kategori Kejahatan di Ujung Jalan pada Peta San Francisco

2. Menampilkan tren kategori-kategori kriminalitas yang ditemukan pada bagian 1

Dari soal yang diberikan, penulis menarik kesimpulan bahwa tren kategori merupakan kategori yang umumnya atau marak terjadi pada satu kondisi tertentu. Penulis menganalisis tren kategori yang ditemukan pada bagian 1 berdasarkan 3 kondisi berbeda, yakni secara keseluruhan di setiap ujung jalan di San Francisco, perhari, dan perdistrik kepolisian. Dari analisis di bawah ini, dapat dilihat tren kategori kriminalitas di ujung jalan.

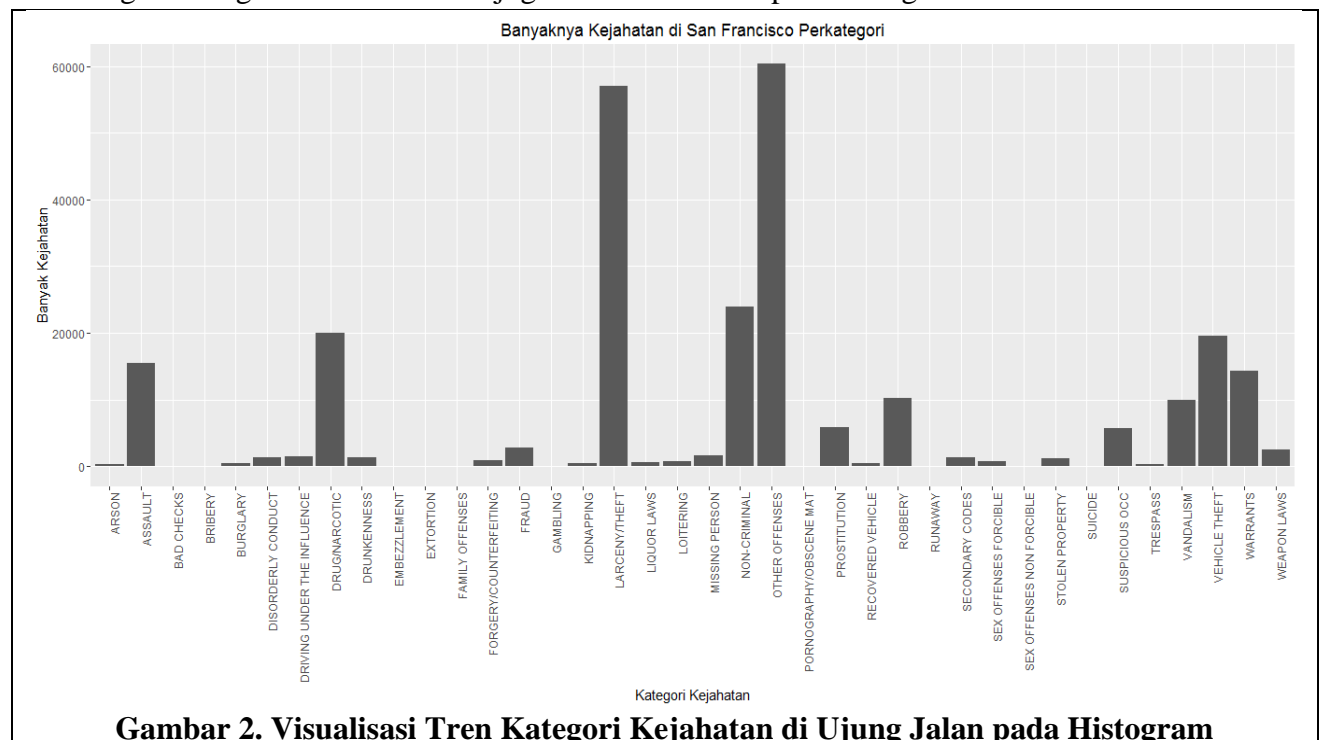
- Tren kategori-kategori kriminalitas di ujung jalan di San Francisco

Penulis menampilkan di *console* R-Studio beberapa kategori-kategori kejahatan yang terjadi di San Francisco beserta jumlah kejadiannya dari kejadian paling besar ke kejadian paling kecil. Analisis tersebut menunjukkan bahwa kategori *Other Offenses* paling banyak terjadi di ujung jalan di San Francisco.

Berikut tren kategori-kategori kriminalitas yang terjadi di San Francisco yang dimunculkan pada *console* R-Studio. Hanya enam kategori terbanyak yang ditampilkan.

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	DRUG/NARCOTIC	VEHICLE	THEFT
60327	56987	23927	19919		19585
ASSAULT					
15430					

Kategori-kategori tersebut di atas juga divisualisasikan pada histogram.



- Tren kategori-kategori kriminalitas di ujung jalan di San Fransico perhari

Penulis menampilkan di *console* R-Studio kategori-kategori kejahatan yang terjadi di San Francisco beserta jumlah kejadiannya dari kejadian paling besar ke kejadian paling kecil perharinya dalam satu minggu. Analisis tersebut menunjukkan bahwa dalam satu minggu, hari Jumat merupakan hari dengan kejadian kriminal paling banyak. Perharinya, penulis menjabarkan kategori kejahatan yang paling banyak dilakukan.

Berikut tren kategori-kategori kriminalitas yang terjadi di San Francisco perhari yang dimunculkan pada *console* R-Studio beserta histogramnya. Kategori yang dimunculkan di *console* hanya enam kategori kejahatan dengan kejadian terbanyak.

- Hari Senin

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	DRUG/NARCOTIC	VEHICLE THEFT	
8288	7460	3201	2910	2598	
ASSAULT					
2128					

- Hari Selasa

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	DRUG/NARCOTIC	VEHICLE THEFT	
8936	7551	3215	3075	2566	
WARRANTS					
2041					

- Hari Rabu

OTHER OFFENSES	LARCENY/THEFT	DRUG/NARCOTIC	NON-CRIMINAL	VEHICLE THEFT	
9659	7682	3396	3171	2544	
WARRANTS					
2419					

- Hari Kamis

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	DRUG/NARCOTIC	VEHICLE THEFT	
8747	7842	3179	3164	2699	
WARRANTS					
2220					

- Hari Jumat

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	VEHICLE THEFT	DRUG/NARCOTIC	
8940	8886	3656	3240	2568	
ASSAULT					
2311					

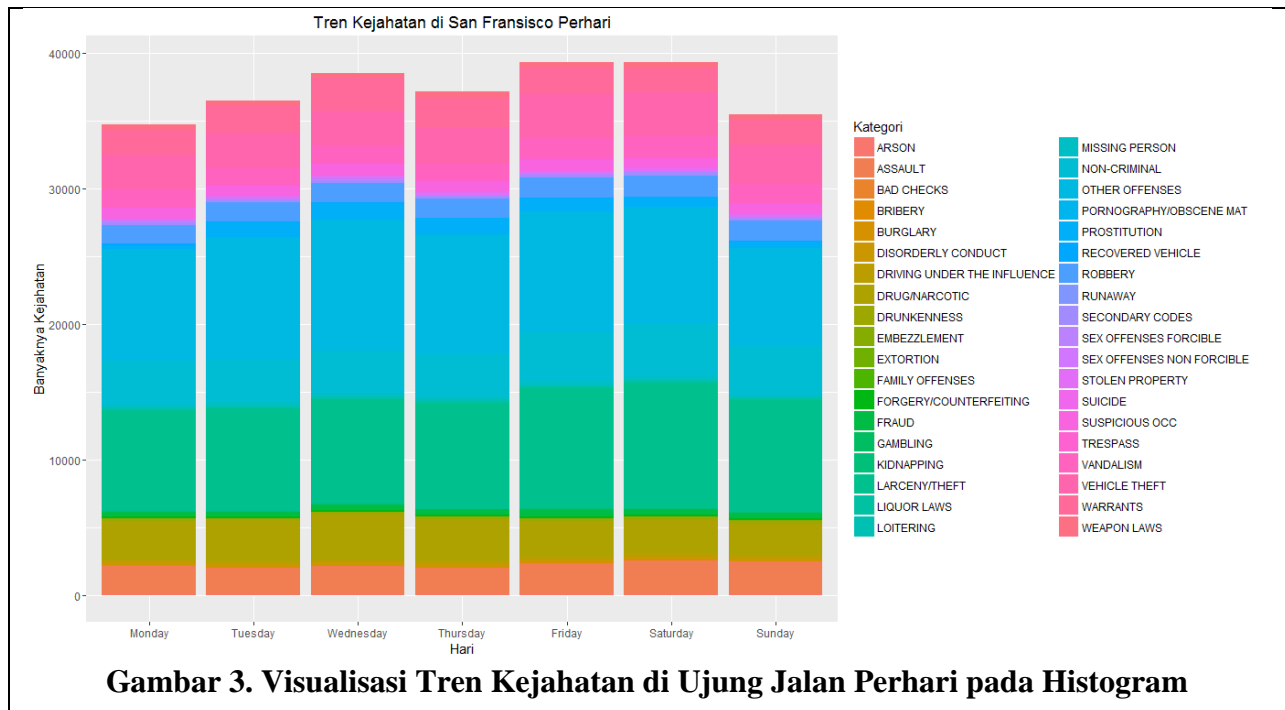
- Hari Sabtu

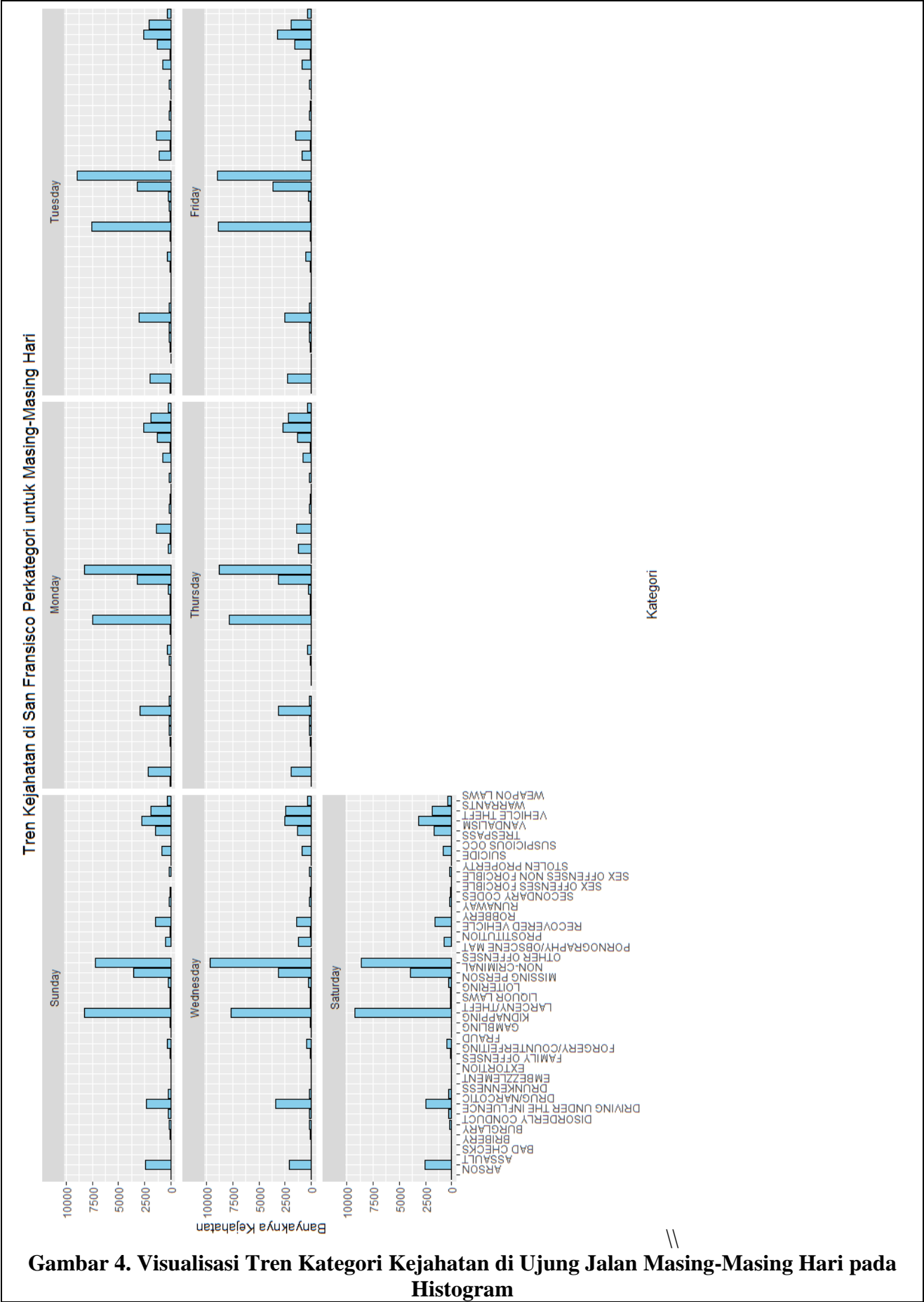
LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	VEHICLE THEFT	ASSAULT	
9279	8588	3913	3134	2510	
DRUG/NARCOTIC					
2453					

- Hari Miinggu

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	VEHICLE THEFT	ASSAULT	
8287	7169	3592	2804	2434	
DRUG/NARCOTIC					
2353					

Tren di atas divisualisasikan pada sebuah histogram tren kategori perhari dan sebuah histogram tren kategori perkategori pada masing-masing hari.





Gambar 4. Visualisasi Tren Kategori Kejahatan di Ujung Jalan Masing-Masing Hari pada Histogram

- Tren kategori-kategori kriminalitas di ujung jalan di San Fransico perdistrik kepolisian

Penulis menampilkan di *console* R-Studio kategori-kategori kejahatan yang terjadi di San Francisco beserta jumlah kejadiannya dari kejadian paling besar ke kejadian paling kecil perdistriknya. Dapat dilihat bahwa distrik dengan kejahatan terbanyak adalah distrik Southern.

Berikut tren kategori-kategori kriminalitas yang terjadi di Ujung Jalan di San Francisco perdistriknya yang dimunculkan pada *console* R-Studio beserta histogramnya. Kategori yang dimunculkan di *console* hanya enam kategori kejahatan dengan kejadian terbanyak.

- Distrik Tenderloin

DRUG/NARCOTIC	OTHER OFFENSES	WARRANTS	LARCENY/THEFT	ASSAULT
6367	5697	2358	2322	1687
NON-CRIMINAL				
1600				

- Distrik Bayview

OTHER OFFENSES	LARCENY/THEFT	DRUG/NARCOTIC	VEHICLE THEFT	ASSAULT
9884	3159	1931	1912	1783
WARRANTS				
1739				

- Distrik Central

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	VEHICLE THEFT	ASSAULT
8250	3922	3184	1951	1434
VANDALISM				
1057				

- Distrik Ingleside

OTHER OFFENSES	VEHICLE THEFT	LARCENY/THEFT	ROBBERY	NON-CRIMINAL
5883	2137	1991	1299	1194
ASSAULT				
1179				

- Distrik Mission

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	VEHICLE THEFT	DRUG/NARCOTIC
9984	6237	3929	3582	3040
PROSTITUTION				
2903				

- Distrik Northern

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	VEHICLE THEFT	DRUG/NARCOTIC
12825	6520	3122	2906	2340
WARRANTS				
2206				

- Distrik Park

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	VEHICLE THEFT	DRUG/NARCOTIC
3159	2945	1822	1585	952
WARRANTS				
805				

- Distrik Richmond

LARCENY/THEFT	OTHER OFFENSES	VEHICLE THEFT	NON-CRIMINAL	VANDALISM
2829	2413	1551	1509	644
ASSAULT				
511				

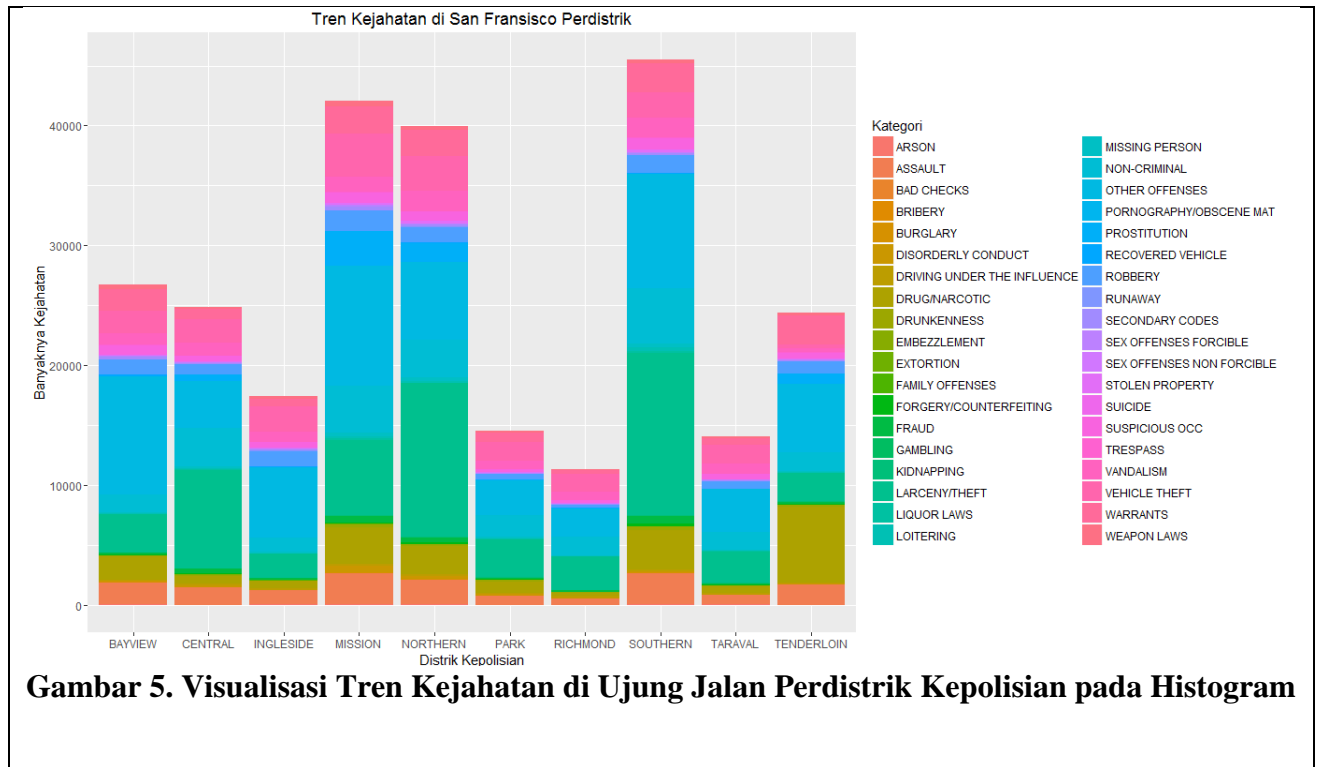
- Distrik Southern

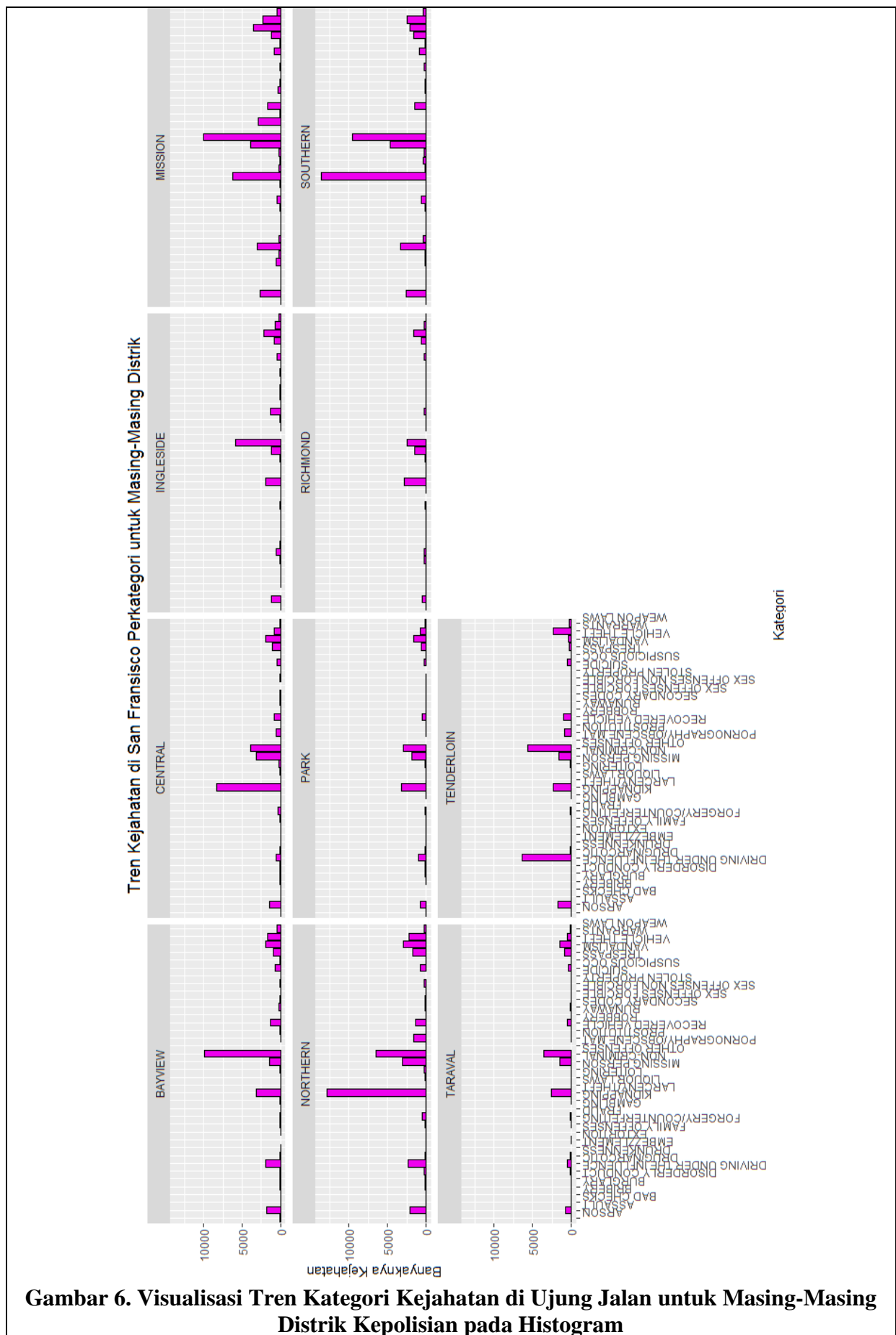
LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	DRUG/NARCOTIC	ASSAULT
13588	9517	4634	3283	2593
WARRANTS				
2452				

- Distrik Taraval

OTHER OFFENSES	LARCENY/THEFT	VEHICLE THEFT	NON-CRIMINAL	VANDALISM
3562	2627	1523	1516	867
ASSAULT				
791				

Tren di atas divisualisasikan pada sebuah histogram tren kategori perdistrik dan sebuah histogram tren kategori perkategori pada masing-masing distrik.





- Tren kategori-kategori kriminalitas di ujung jalan di San Fransico pertahun

Penulis menampilkan di *console* R-Studio kategori-kategori kejahatan yang terjadi di San Francisco beserta jumlah kejadiannya dari kejadian paling besar ke kejadian paling kecil pertahunnya

Berikut tren kategori-kategori kriminalitas yang terjadi di Ujung Jalan di San Francisco pertahun dari tahun 2003 sampai 2015 yang dimunculkan pada *console* R-Studio beserta histogramnya. Kategori yang dimunculkan di *console* hanya enam kategori kejahatan dengan kejadian terbanyak.

- Tahun 2003

OTHER OFFENSES	LARCENY/THEFT	VEHICLE THEFT	NON-CRIMINAL	DRUG/NARCOTIC
5092	4590	2847	2061	1856
WARRANTS				
1580				

- Tahun 2004

OTHER OFFENSES	LARCENY/THEFT	VEHICLE THEFT	NON-CRIMINAL	DRUG/NARCOTIC
5115	4096	3158	2112	1738
WARRANTS				
1407				

- Tahun 2005

OTHER OFFENSES	LARCENY/THEFT	VEHICLE THEFT	NON-CRIMINAL	DRUG/NARCOTIC
4378	4335	3385	2064	1600
WARRANTS				
1233				

- Tahun 2006

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	DRUG/NARCOTIC	VEHICLE THEFT
5027	4502	1876	1794	1428
ASSAULT				
1377				

- Tahun 2007

OTHER OFFENSES	LARCENY/THEFT	DRUG/NARCOTIC	NON-CRIMINAL	WARRANTS
4757	4206	1975	1692	1326
VEHICLE THEFT				
1312				

- Tahun 2008

OTHER OFFENSES	LARCENY/THEFT	DRUG/NARCOTIC	NON-CRIMINAL	ASSAULT
6314	3921	2184	1511	1361
VEHICLE THEFT				
1246				

- Tahun 2009

OTHER OFFENSES	LARCENY/THEFT	DRUG/NARCOTIC	NON-CRIMINAL	ASSAULT
6351	3911	2427	1565	1340
WARRANTS				
1051				

- Tahun 2010

OTHER OFFENSES	LARCENY/THEFT	DRUG/NARCOTIC	NON-CRIMINAL	ASSAULT
4688	3442	1559	1511	1020
WARRANTS				
903				

- Tahun 2011

OTHER OFFENSES	LARCENY/THEFT	NON-CRIMINAL	DRUG/NARCOTIC	ASSAULT
4412	3551	1845	1090	955
WARRANTS				
887				

- Tahun 2012

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	DRUG/NARCOTIC	ASSAULT
4737	4016	1921	1152	1047
VEHICLE THEFT				
979				

- Tahun 2013

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	ASSAULT	DRUG/NARCOTIC
5972	4115	2539	1271	1214
VEHICLE THEFT				
1132				

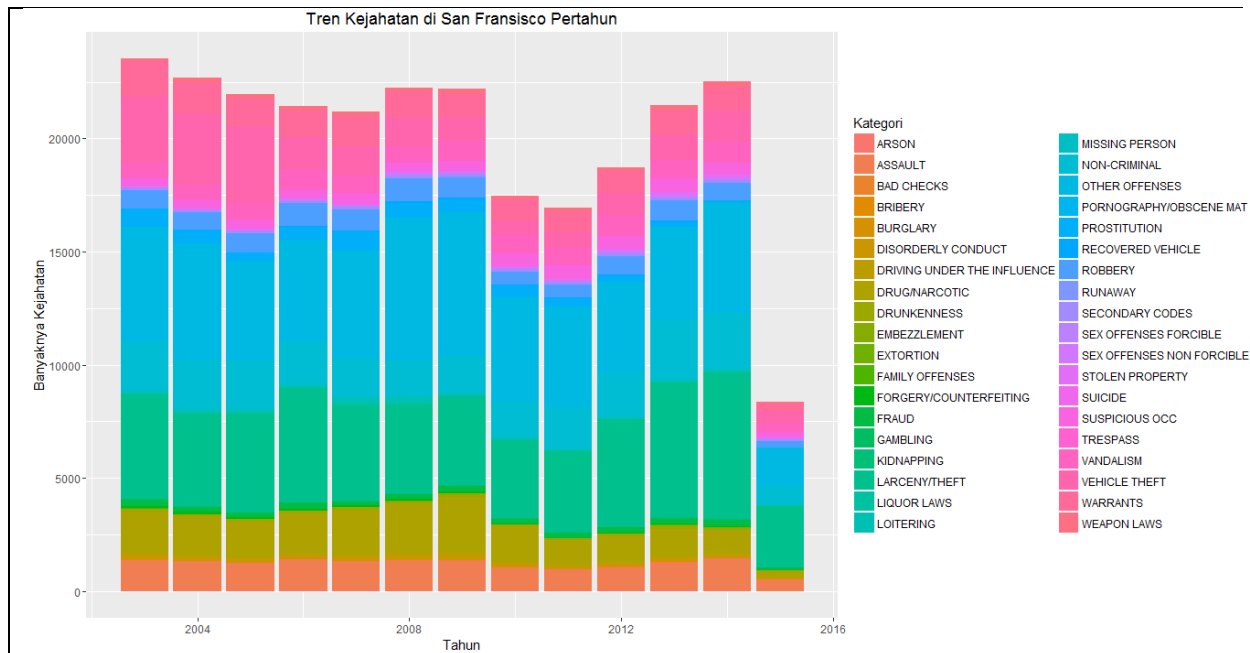
- Tahun 2014

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	ASSAULT	VEHICLE THEFT
6484	4869	2435	1407	1318
WARRANTS				
1094				

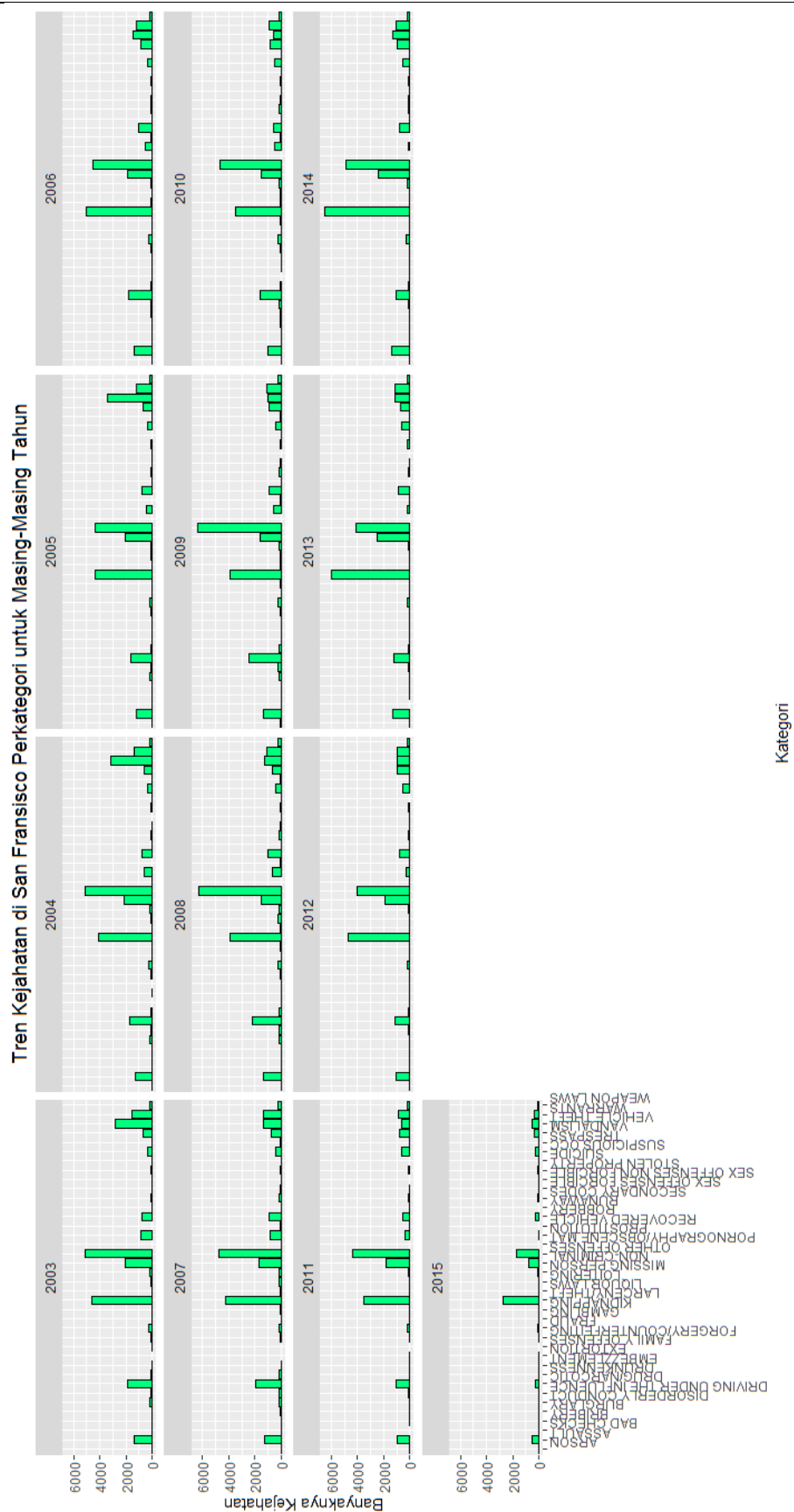
- Tahun 2015

LARCENY/THEFT	OTHER OFFENSES	NON-CRIMINAL	VEHICLE THEFT	ASSAULT
2715	1718	795	525	487
WARRANTS				
378				

Tren di atas divisualisasikan pada sebuah histogram tren kategori pertahun dan sebuah histogram tren kategori perkategori pada masing-masing tahun.



Gambar 7. Visualisasi Tren Kejahatan di Ujung Jalan Pertahun pada Histogram



Gambar 8. Visualisasi Tren Kategori Kejahatan di Ujung Jalan untuk Masing-Masing Tahun Kepolisian pada Histogram

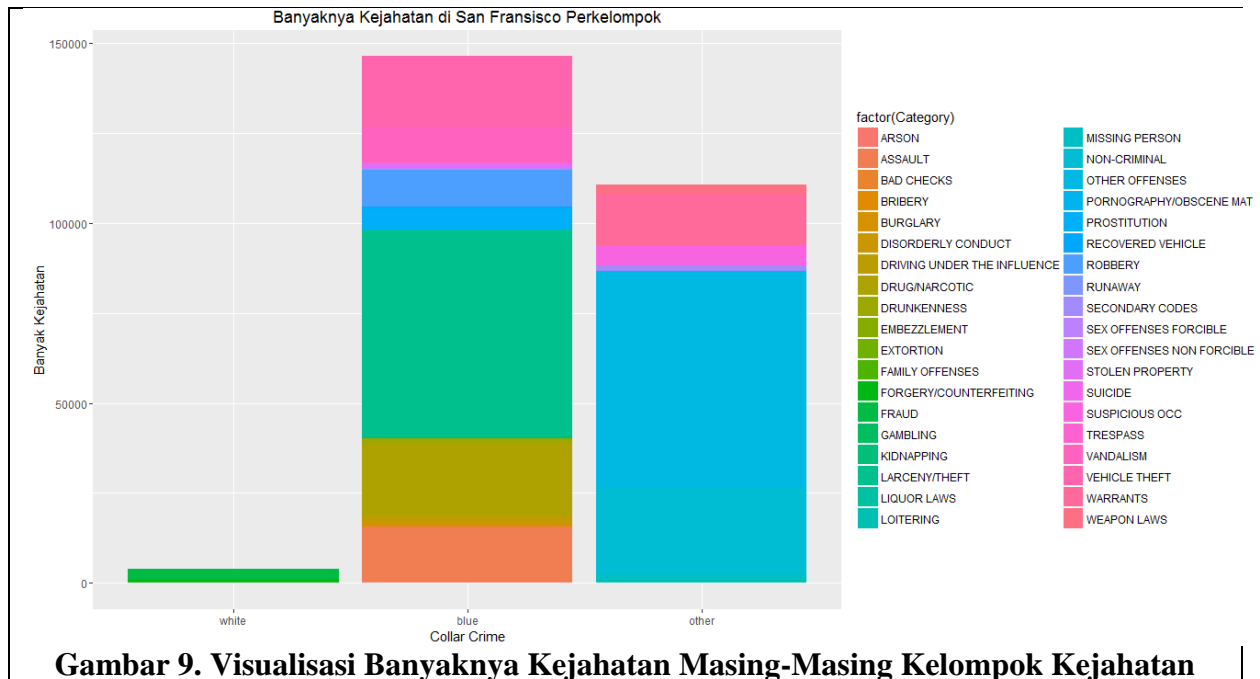
3. Mengategorikan kategori-kategori kriminalitas pada bagian 1 menjadi 3 kategori kejahatan yang dapat digunakan untuk pengambilan keputusan kejahatan mana yang harus ditindak terlebih dahulu.

Berdasarkan soal yang diberikan, kategori kejahatan dibagi menjadi tiga, yakni *White-Collar Crime*, *Blue-Collar Crime*, dan *Other Crime*. Menurut situs resmi FBI (<https://www.fbi.gov/>), *White-Collar Crime* adalah suatu bentuk kejahatan yang dilakukan oleh para profesional dalam bisnis dan pemerintahan. Dilihat dari bentuknya, kejahatan ini biasanya dilakukan dalam bentuk kejahatan ekonomi, kejahatan terhadap kepemilikan, dan kejahatan lain yang melanggar aturan keselamatan, lingkungan, dan kesehatan. *White-collar crime* merupakan kejahatan yang bersifat sistematis. Berdasarkan definisi tersebut, maka penulis mengelompokkan kategori-kategori kriminalitas FRAUD, BAD CHECKS, FORGERY/COUNTERFEITING, EXTORTION, EMBEZZLEMENT, SUSPICIOUS OCC, dan BRIBERY sebagai *White-Collar Crime*. *Blue-Collar Crime* adalah suatu bentuk kejahatan yang dilakukan oleh masyarakat kelas bawah. Bentuk kejahatan ini biasanya tidak terlalu sistematis jika dibandingkan dengan *White-Collar Crime* dikarenakan perbedaan sarana yang cukup besar antara kalangan kelas bawah dan profesional. Berdasarkan definisi tersebut, maka penulis mengelompokkan kategori-kategori kriminalitas VANDALISM, LARCENY/THEFT, STOLEN PROPERTY, ROBBERY, DRIVING UNDER THE INFLUENCE, DISORDERLY CONDUCT, LIQUOR LAWS, VEHICLE THEFT, ASSAULT, KIDNAPPING, TRESPASS, ARSON, BURGLARY, DRUG/NARCOTIC, DRUNKENNESS, PROSTITUTION, GAMBLING, SEX OFFENSES FORCIBLE, SEX OFFENSES NON FORCIBLE dan RECOVERED VEHICLE sebagai *Blue-Collar Crime*. Pembagian kategori kriminalitas dapat dilihat pada tabel berikut.

White-Collar Crime	Blue-Collar Crime	Other Crime
FRAUD	VANDALISM	WARRANTS
BAD CHECKS	LARCENY/THEFT	OTHER OFFENSES
FORGERY/COUNTERFEITING	STOLEN PROPERTY	NON-CRIMINAL
EXTORTION	ROBBERY	MISSING PERSON
EMBEZZLEMENT	DRIVING UNDER THE INFLUENCE	SECONDARY CODES
SUSPICIOUS OCC	DISORDERLY CONDUCT	PORNOGRAPHY/OBSCENE MAT
BRIBERY	LIQUOR LAWS	FAMILY OFFENSES
	VEHICLE THEFT	WEAPON LAWS
	ASSAULT	SUICIDE
	KIDNAPPING	RUNAWAY
	TRESPASS	LOITERING
	ARSON	
	RECOVERED VEHICLE	
	BURGLARY	
	DRUG/NARCOTIC	
	DRUNKENNESS	

	PROSTITUTION	
	GAMBLING	
	SEX OFFENSES FORCIBLE	
	SEX OFFENSES NON FORCIBLE	

Berikut visualisasi dari banyaknya kejahatan pada masing-masing kelompok kejahatan.



Gambar 9. Visualisasi Banyaknya Kejahatan Masing-Masing Kelompok Kejahatan

V. Script

```
#SCRIPT DATA ANALYSIS DAN DATA VISUALIZATION KRIMINAL SAN FRANCISCO
```

```
#Library yang dibutuhkan
```

```
library(ggplot2)
```

```
library(readr)
```

```
library(dplyr)
```

```
library(RColorBrewer)
```

```
library(leaflet)
```

```
library(lubridate)
```

```
#Memasukan data ke dalam
```

```
train <- read.csv("train.csv", header = TRUE)
```

```
test <- read.csv("test.csv", header = TRUE)
```

```
#PREDIKSI KATEGORI KEJAHATAN DI UJUNG JALAN
```

```
#Menambahkan kolom street corner pada kedua dataset
```

```
train$StreetCorner = "No"
```

```

train[which(grepl( '/',train$Address)), 'StreetCorner']="Yes"

test$StreetCorner = "No"
test[which(grepl( '/',train$Address)), 'StreetCorner']="Yes"

#Memprediksi kategori kejahatan yang ada dengan melihat isi kategori pada
dataset train
unique((train$Category[which(train$StreetCorner == "Yes"))])

#Mengambil Data Kejahatan yang Terjadi di Ujung Jalan
train.streetcorner <- train[which(train$StreetCorner == "Yes"),]
test.streetcorner <- test[which(test$StreetCorner == "Yes"),]

#Menampilkan peta wilayah semua kejahatan di San Francisco
trains.map <- function(categories, n) {

new.train <- filter(train.streetcorner, Category %in% categories) %>%
droplevels()

pal <- colorFactor(brewer.pal(length(unique(new.train$Category)), "Set3"),
domain = new.train$Category)

leaflet(new.train[1:n,]) %>%
addProviderTiles("CartoDB.Positron") %>%
addCircleMarkers (lng = ~X, lat = ~Y,
color = ~pal(Category),
opacity = .7, radius = 1) %>%
addLegend(pal = pal, values = new.train$Category)
}

trains.map(c("WARRANTS", "OTHER OFFENSES", "LARCENY/THEFT", "VEHICLE
THEFT",
"VANDALISM", "NON-CRIMINAL", "ROBBERY", "ASSAULT", "WEAPON LAWS",
"BURGLARY", "SUSPICIOUS OCC", "DRUNKENNESS",
"FORGERY/COUNTERFEITING", "DRUG/NARCOTIC", "STOLEN PROPERTY",
"SECONDARY CODES", "TRESPASS", "MISSING PERSON", "FRAUD",
"KIDNAPPING", "RUNAWAY", "DRIVING UNDER THE INFLUENCE",
"SEX OFFENSES FORCIBLE", "PROSTITUTION", "DISORDERLY CONDUCT",
"ARSON", "FAMILY OFFENSES", "LIQUOR LAWS", "BRIBERY",
"EMBEZZLEMENT", "SUICIDE", "LOITERING", "SEX OFFENSES NON FORCIBLE",
"EXTORTION", "GAMBLING", "BAD CHECKS", "RECOVERED VEHICLE"
, "PORNOGRAPHY/OBSCENE MAT"), n = nrow(train.streetcorner))

#TREN KEJAHATAN DI UJUNG JALAN

#Menampilkan banyaknya kejahatan di San Francisco perkategori pada console
head(sort(table(train.streetcorner$Category), decreasing = TRUE))

#Menampilkan histogram kejahatan di San Francisco
train.streetcorner %>%
ggplot(aes(x = Category)) +
geom_bar() +
ggtitle("Banyaknya Kejahatan di San Francisco Perkategori") +
ylab("Banyak Kejahatan") +

```

```

xlab("Kategori Kejahatan") +
theme(axis.text.x = element_text(angle = 90, hjust = 1))

#Menampilkan histogram banyaknya kejahatan di San Francisco perhari dalam
satu minggu
train.streetcorner %>%
ggplot(aes(x = factor(DayOfWeek, levels=c('Monday', 'Tuesday',
'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday')))) +
geom_bar(aes(fill=factor(Category))) +
ggtitle("Tren Kejahatan di San Fransisco Perhari") +
xlab("Hari") +
ylab("Banyaknya Kejahatan") +
labs(fill = "Kategori")

#Menampilkan histogram banyaknya kejahatan di San Francisco perkategori
untuk masing-masing hari
train.streetcorner$DayOfWeek <- factor(train.streetcorner$DayOfWeek,
levels = c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday",
"Friday", "Saturday"))
ggplot(data=train.streetcorner, aes(x=Category)) +
geom_bar(colour="black", fill="skyblue") +
xlab('Kategori') +
ylab('Banyaknya Kejahatan') +
ggtitle("Tren Kejahatan di San Fransisco Perkategori untuk Masing-Masing
Hari") +
facet_wrap(~DayOfWeek) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))

#Menampilkan tren kejahatan di San Francisco perkategori perhari dalam
satu minggu pada console
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Monday")])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Tuesday")])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Wednesday")])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Thursday")])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Friday")])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Saturday")])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$DayOf
Week == "Sunday")])), decreasing = TRUE))

#Menampilkan histogram banyaknya kejahatan di San Francisco perdistribusi
train.streetcorner %>%
ggplot(aes(x = PdDistrict)) +
geom_bar(aes(fill=factor(Category))) +
ggtitle("Tren Kejahatan di San Fransisco Perdistribusi") +
xlab("Distrik Kepolisian") +
ylab("Banyaknya Kejahatan") +
labs(fill = "Kategori")

```

```

#Menampilkan histogram banyaknya kejahatan di San Francisco perkategori
untuk masing-masing distrik
ggplot(data=train.streetcorner, aes(x=Category)) +
geom_bar(colour="black", fill="magenta2") +
xlab('Kategori') +
ggtitle("Tren Kejahatan di San Fransisco Perkategori untuk Masing-Masing
Distrik") +
ylab('Banyaknya Kejahatan') +
facet_wrap(~PdDistrict) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))

#Menampilkan banyaknya kejahatan di San Francisco perdistrik kepolisian
pada console
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "TENDERLOIN"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "BAYVIEW"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "CENTRAL"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "INGLESIDE"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "MISSION"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "NORTHERN"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "PARK"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "RICHMOND"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "SOUTHERN"])), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$PdDis
trict == "TARAVAL"])), decreasing = TRUE))

#Membuat kolom baru berupa tahun yang diambil dari tanggal
train.streetcorner$Year = year(train.streetcorner$Dates)
test.streetcorner$Year = year(test.streetcorner$Dates)

#Menampilkan histogram banyaknya kejahatan di San Francisco perkategori
untuk masing-masing tahun
ggplot(data=train.streetcorner, aes(x=Category)) +
geom_bar(colour="black", fill="springgreen") +
xlab('Kategori') +
ylab('Banyaknya Kejahatan') +
ggtitle("Tren Kejahatan di San Fransisco Perkategori untuk Masing-Masing
Tahun") +
facet_wrap(~Year) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))

#Menampilkan histogram banyaknya kejahatan di San Francisco pertahun
train.streetcorner %>%
ggplot(aes(x = Year)) +
geom_bar(aes(fill=factor(Category))) +
ggtitle("Tren Kejahatan di San Fransisco Pertahun") +

```

```

xlab("Tahun") +
ylab("Banyaknya Kejahatan") +
labs(fill = "Kategori")

#Menampilkan banyaknya kejahatan di San Francisco perdistrik kepolisian
pada console
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2003)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2004)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2005)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2006)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2007)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2008)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2009)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2010)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2011)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2012)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2013)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2014)]), decreasing = TRUE))
head(sort(table(train.streetcorner$Category[which(train.streetcorner$Year
== 2015)]), decreasing = TRUE))

#MENGATEGORIKAN KATEGORI-KATEGORI KEJAHATAN (COLLAR-CRIME)

#Menambahkan kolom Collar Crime berdasarkan kategori yang ada

train.streetcorner$CollarCrime = "other"
train.streetcorner[which(train.streetcorner$Category=="FRAUD"), 'CollarCrim
e'] = "white"
train.streetcorner[which(train.streetcorner$Category=="BAD
CHECKS"), 'CollarCrime'] = "white"
train.streetcorner[which(train.streetcorner$Category=="FORGERY/COUNTERFEIT
ING"), 'CollarCrime'] = "white"
train.streetcorner[which(train.streetcorner$Category=="EXTORTION"), 'Collar
Crime'] = "white"
train.streetcorner[which(train.streetcorner$Category=="EMBEZZLEMENT"), 'Col
larCrime'] = "white"
train.streetcorner[which(train.streetcorner$Category=="SUSPICIOUS
ACC"), 'CollarCrime'] = "white"
train.streetcorner[which(train.streetcorner$Category=="BRIBERY"), 'CollarCr
ime'] = "white"

train.streetcorner[which(train.streetcorner$Category=="VANDALISM"), 'Collar

```

```

Crime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="LARCENY/THEFT"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="STOLEN PROPERTY"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="ROBBERY"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="DRIVING UNDER THE INFLUENCE"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="DISORDERLY CONDUCT"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="LIQUOR LAWS"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="VEHICLE THEFT"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="ASSAULT"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="KIDNAPPING"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="TRESPASS"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="ARSON"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="RECOVERED VEHICLE"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="BURGLARY"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="DRUG/NARCOTIC"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="DRUNKENNESS"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="PROSTITUTION"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="GAMBLING"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="SEX OFFENSES FORCIBLE"), 'CollarCrime'] = "blue"
train.streetcorner[which(train.streetcorner$Category=="SEX OFFENSES NON FORCIBLES"), 'CollarCrime'] = "blue"

#Menampilkan histogram kelompok kejahatan di San Francisco
train.streetcorner %>%
ggplot(aes(x = factor(CollarCrime, levels=c('white', 'blue', 'other')))) +
geom_bar(aes(fill=factor(Category))) +
ggtitle("Banyaknya Kejahatan di San Fransisco Perkelompok") +
ylab("Banyak Kejahatan") +
xlab("Collar Crime")

#MENULIS TABEL KE FILE CSV
write.csv(train.streetcorner, file = "train_new.csv")
write.csv(test.streetcorner, file = "test_new.csv")

```