Crime in San francisco repeats consistently itself every day

In this document we are working with the criminal incident data from San fransisco, our task is to document a finding about some fact concerning this data.
We use **R** programming language for analyzing and processing the data and also for visualizing (using the library **ggplot2**).
The first thing we need to do is to load data into **R** using the following command:

```
sanfrancisco<-
read.csv("sanfrancisco_incidents_summer_2014.csv",header=T)
```

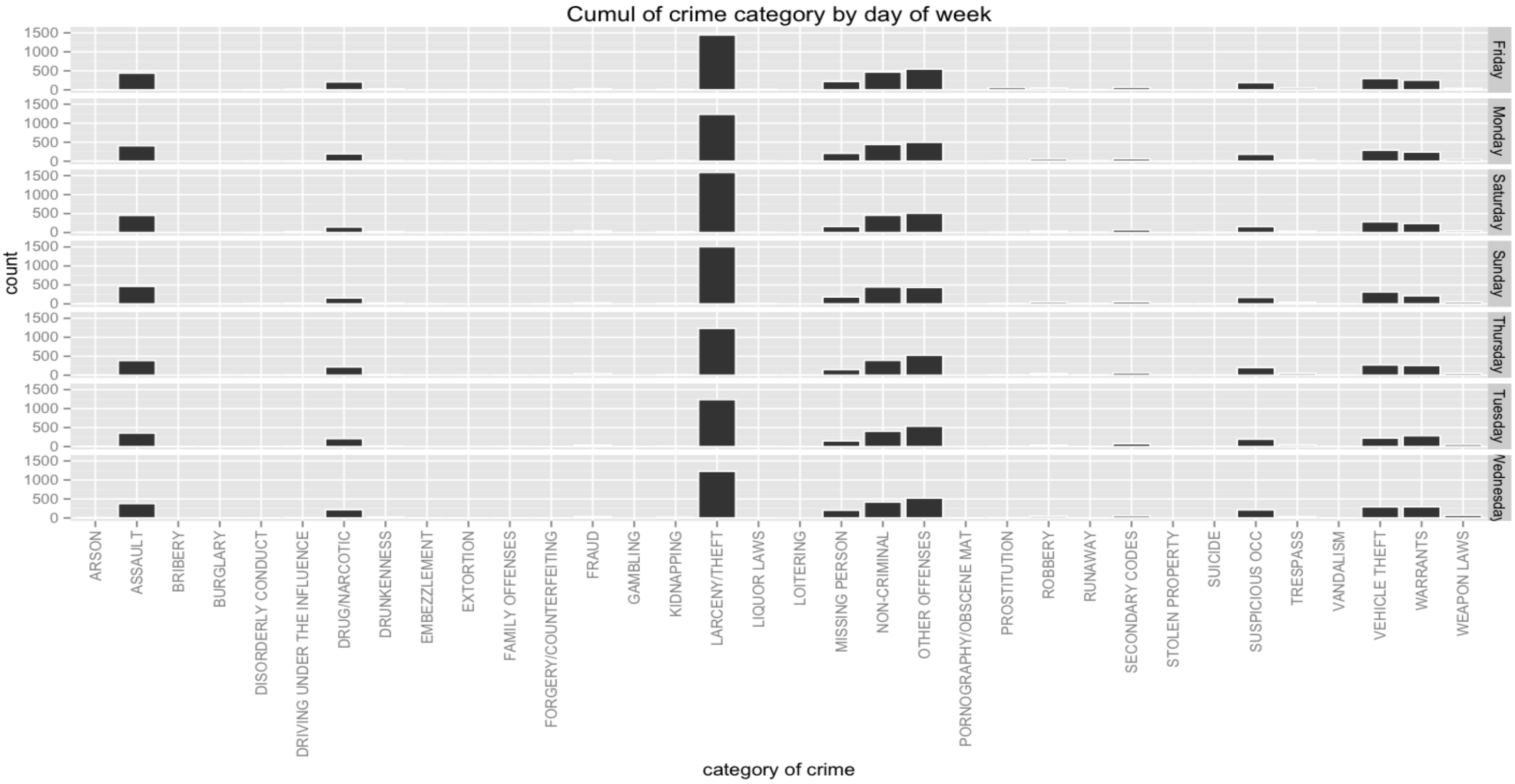
after that we issue commands like: `str(sanfrancisco)` and `summary(sanfrancisco)` in order to understand and get a sense about this data.
Searching about a pattern in some unknown data to you take a lot of time and effort and it is not guaranteed than you will get a result ;-)
Trying to view some correlation about the columns of this data I issued the command `pairs(sanfrancisco)` which gave me a graph showing some correlation between the different variable of this data.
After some trial and error I come up with the following observation: If history repeats itself (as it is always alluded to) can crime also exhibits this property? In other words is the crime the same in all days of the week in San fransisco? Is there a safest day of the week?
What I found is that the crime repeats itself consistently every day of the week in San fransisco, and so there is no safest day.
I will try to convince you using some visualization from this data using the ggplot library of R.

Cumul of crime category by day of week

Issuing the following command:

```
ggplot(sanfrancisco, aes(Category) ) + geom_histogram(color =
"white") + facet_grid(DayOfWeek ~ .) +
theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
ggtitle("Cumul of crime category by day of week" ) +
labs(x="category of crime",y="count")
```

will display this graph:



which display the total count of crimes by category for every day of the week and we see clearly that the number of each category of the crime is roughly the same for every day of the week.

Total Crime Per Hour of Day

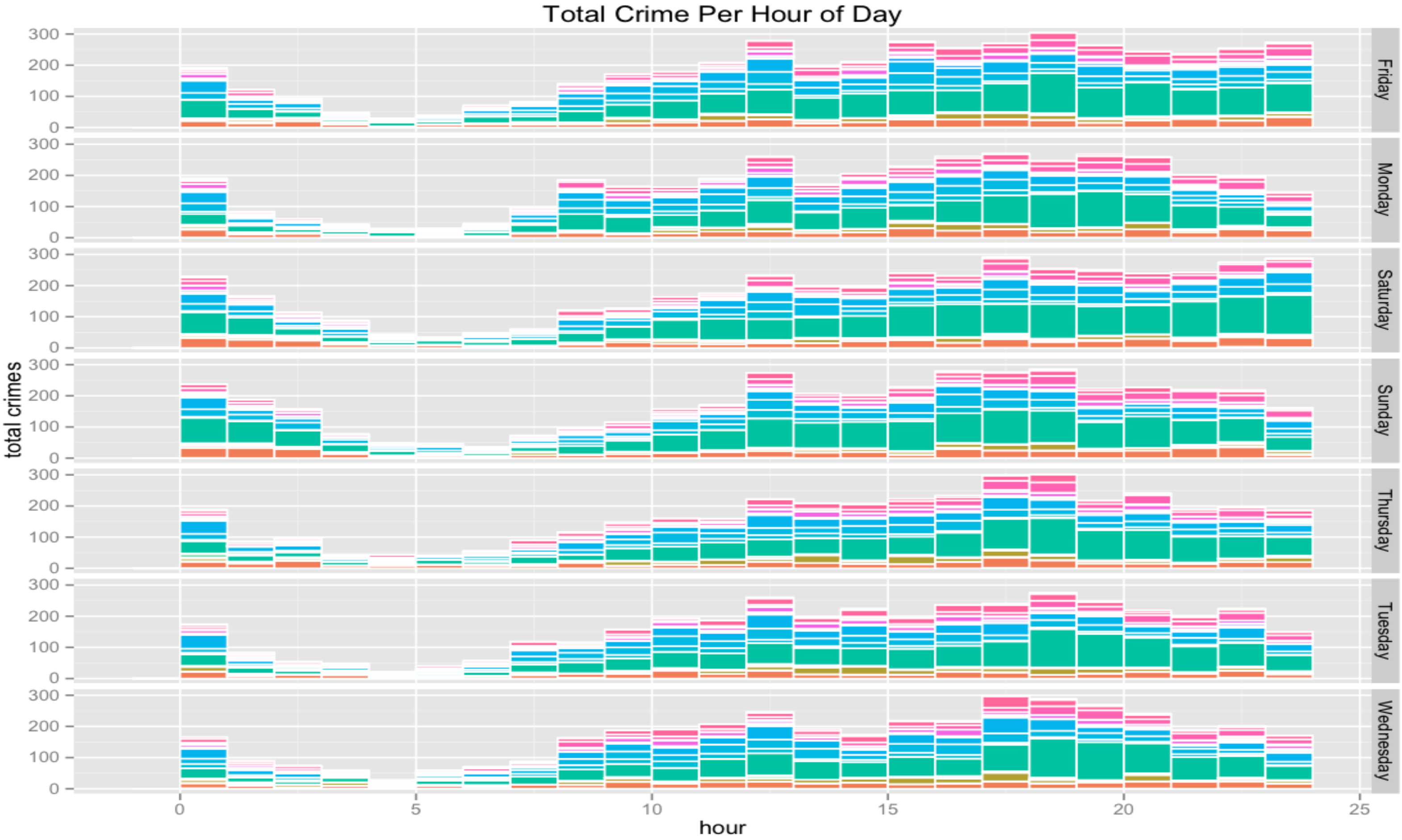
Let's also analyze the distribution of the crimes for every hour of the day for every day in the week.
For this we need to introduce a new variable HourOfDay derived from the Time variable where we simply ignore the minutes and keep only the hours.

```
sanfrancisco$HourOfDay<-
as.numeric(gsub(":.*","",sanfrancisco$Time))
```

Now we execute the following command:

```
ggplot(sanfrancisco, aes(HourOfDay, fill=Category)) +
geom_histogram(binwidth=1,color = "white") + facet_grid(DayOfWeek
~ .) +ggtitle("Total Crime Per Hour of Day") + labs(x="hour",
y="total crimes")
```

which will display the following graph:



which also show us that the distribution of the crimes by hours of the day is also similar for every day of the week.

Conclusion

I hope that you are now convinced that the crime is the same for every day in San fransisco and that there is no safest day in the city.