1. **Plotly - San Francisco crime analysis**

**URL:** [**https://plot.ly/dashboard/panjwani.h:18**](https://plot.ly/dashboard/panjwani.h:18)

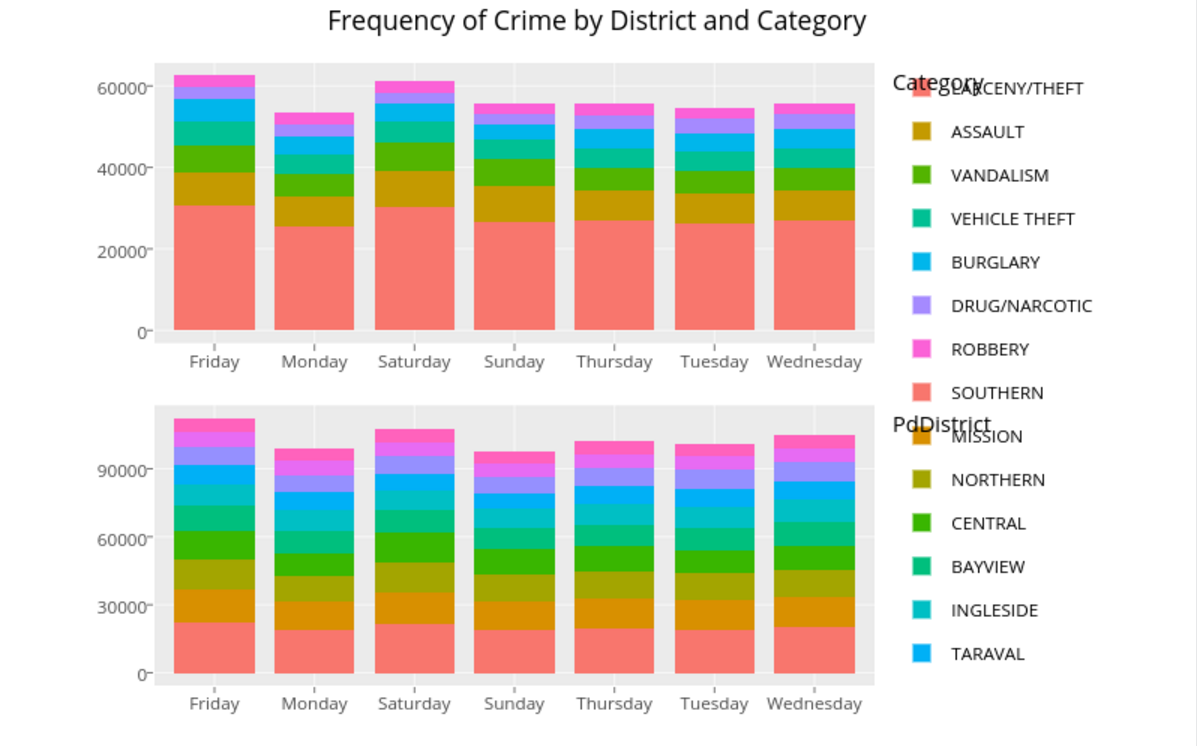
**Description**

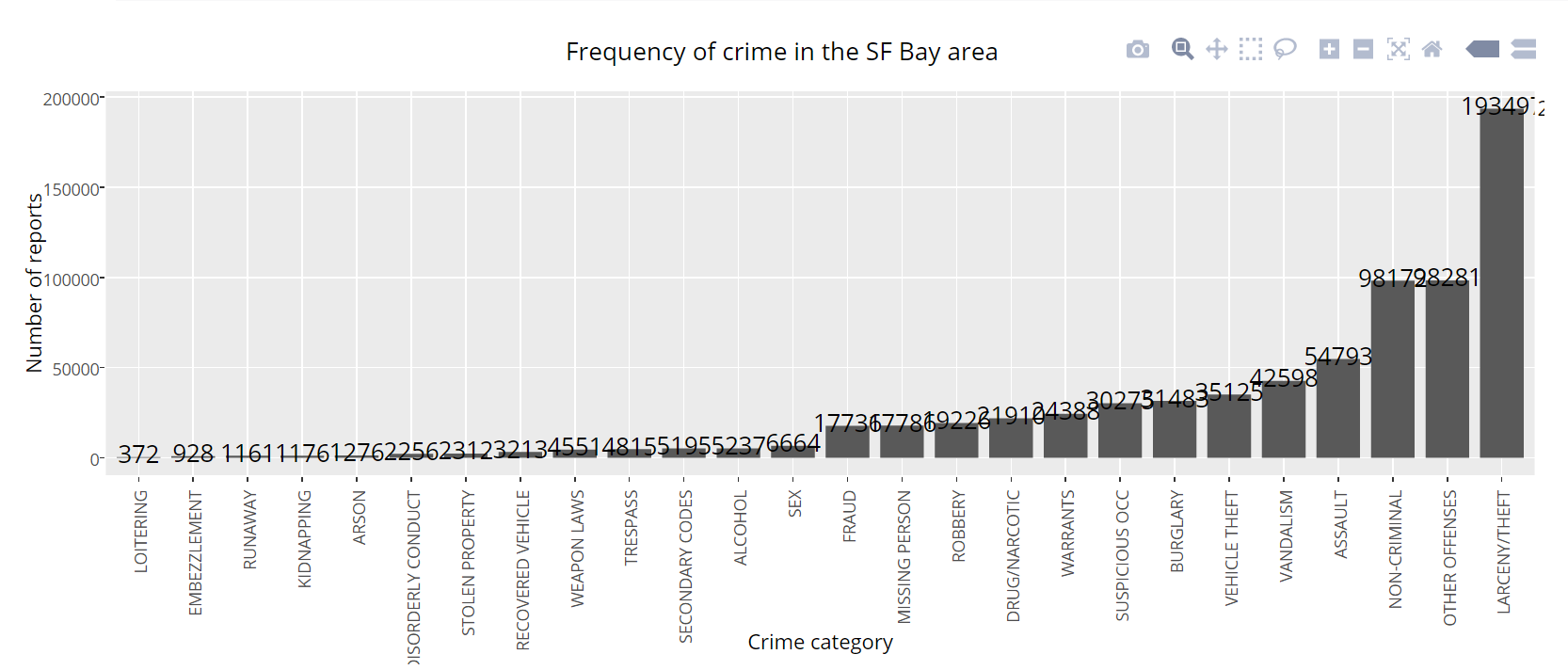
San Francisco is currently the cultural, commercial and financial center of Northern California. Today the city is known more for its tech scene but it has a massive criminal past. The sudden growth in the population has brought an inequality in terms of living, housing shortages leading to no scarcity of crime in the city by the bay.

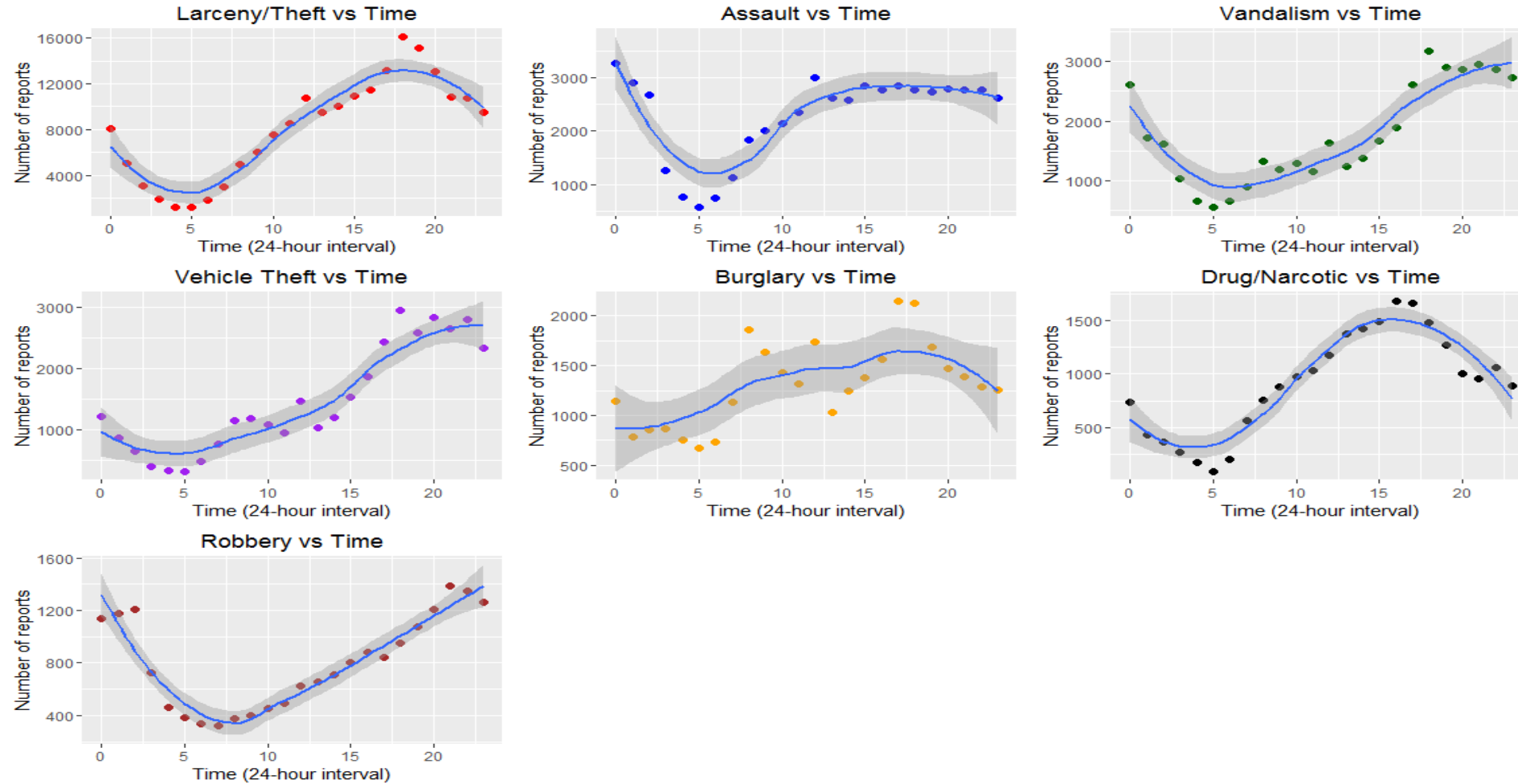
**Business Case:**

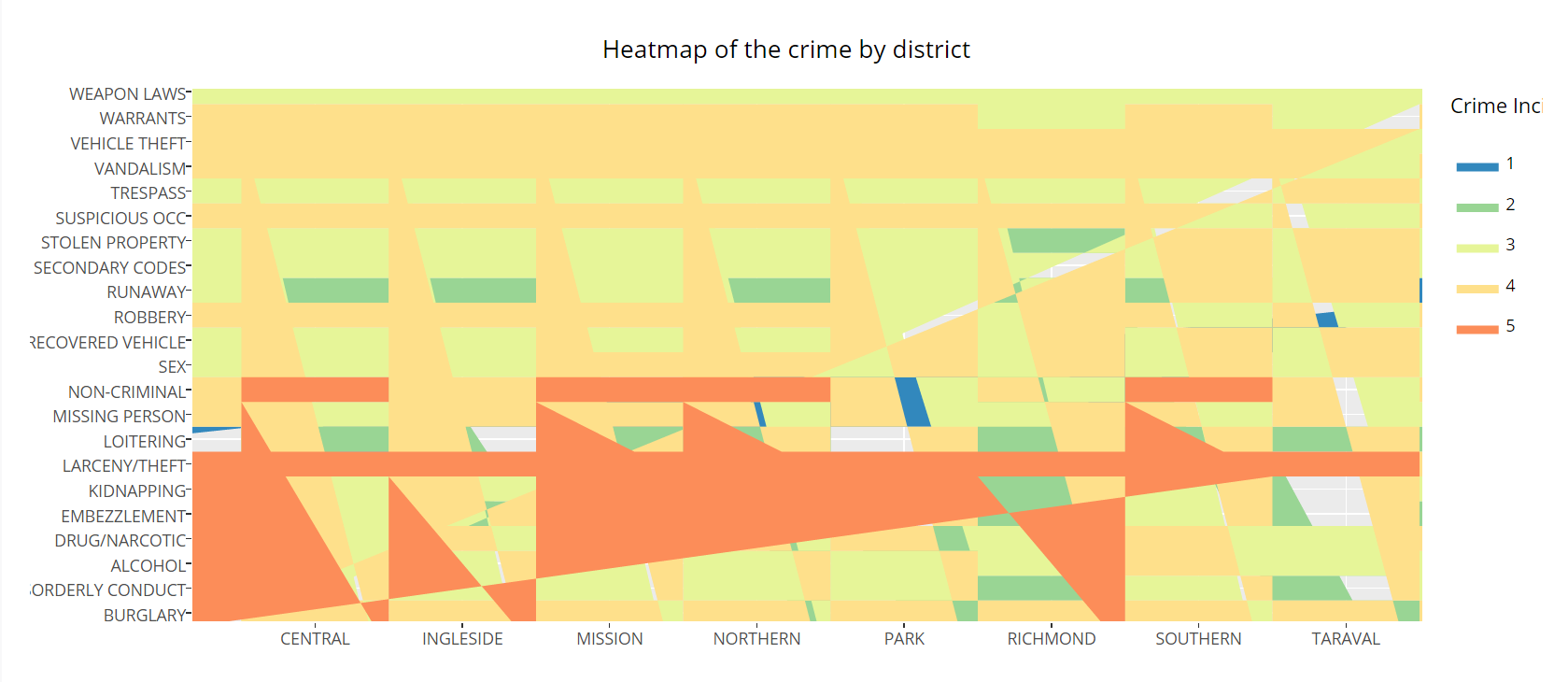
Our analysis could help the police department to get an overall view on the category of crime occurring in a particular area. Based on our analysis the police department could set up extra patrolling/ checks in notorious areas to avoid criminal activities in the city of San Francisco.

**Crime Analysis**









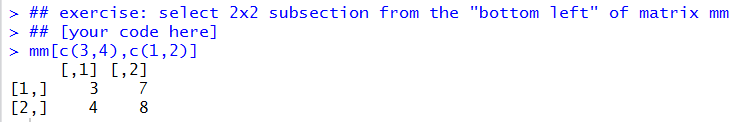
1. **R Lab Code**
2. **0-Intro.R**

**Exercise: select 2x2 subsection from the "bottom left" of matrix mm**

**Solution:**

**mm[c(3,4), c(1,2)]**

**Output**



1. **1-data.R**

**Exercise:**

**obtain this data view from "df":**

**X Grad.Rate**

**1 James Madison University 98**

**2 Incarnate Word College 95**

**3 Johns Hopkins University 90**

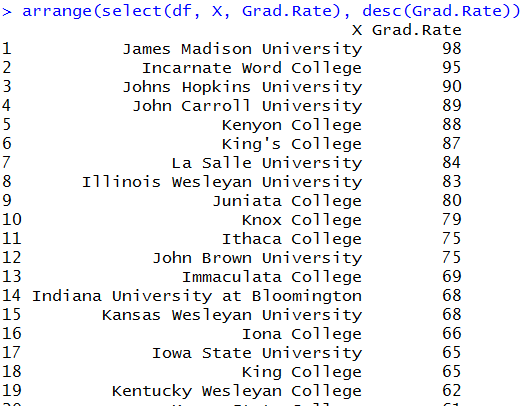
**4 John Carroll University 89**

**5 Kenyon College 88**

**6 King's College 87**

**Solution:**

**arrange(select(df, X, Grad.Rate), desc(Grad.Rate))**

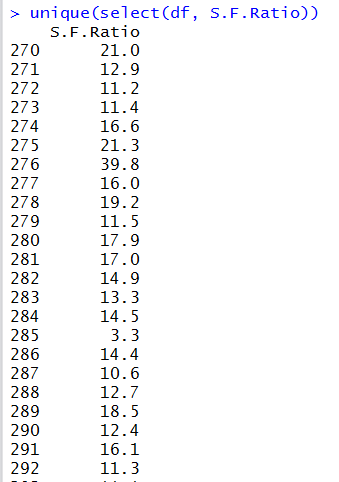


**Exercise: extract distinct (unique) rows**

**select(df, S.F.Ratio)**

**Solution:**

**unique(select(df, S.F.Ratio))**



**Exercise: find max and min tuition ("Outstate") grouped by private/public school, in dataset 'df' and 'college'**

**DF:**

**Private max min**

1. **No 9766 3946**
2. **Yes 19240 6398**

**college:**

**Private max min**

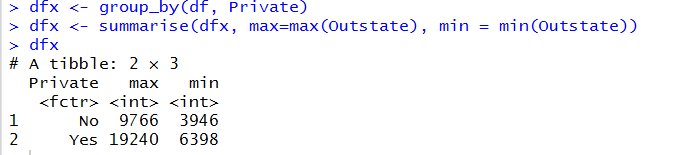
1. **No 15732 2580**
2. **Yes 21700 2340**

**Solution:**

**dfx <- group\_by(df, Private)**

**dfx <- summarise(dfx, max=max(Outstate), min = min(Outstate))**

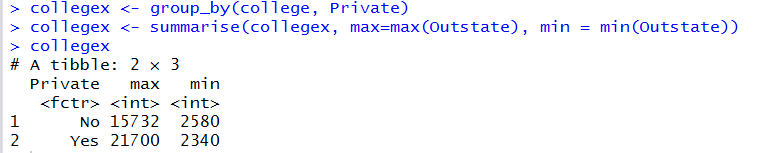
**dfx**



**collegex <- group\_by(college, Private)**

**collegex <- summarise(collegex, max=max(Outstate), min = min(Outstate))**

**college**



1. **Hadoop Lab 3**

**Commands to run the code**

Within the directory Hadoop Workspace/Lab 3

* javac -classpath ../../hadoop-0.20.2/hadoop-0.20.2-core.jar -d ipcount\_classes Runner.java
* jar cvf ipcount.jar -C ipcount\_classes/ .
* ../../hadoop-0.20.2/bin/hadoop jar ipcount.jar Cloud.ApacheLog.Runner input output

