**CS 590V: Data Visualization and Exploration**

**Homework 4: April 11th, 2017**

**Kriti Shrivastava (Student ID: 31041848)**

**Dataset: Crime in Context, 1975-2015**

<https://www.kaggle.com/marshallproject/crime-rates>

**Metadata:**The table contains around 2.8k records and 15 columns (planning to add up to 15k records).  
File size: 257.75 KB  
Column Description:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Description** | **Data Type** |
| report\_year | Year the data was reported (1975-2015) | Numeric |
| agency\_code | Code of the agency which reported the data | String |
| agency\_jurisdiction | Jurisdiction area of the agency (City, State code) | String |
| population | Population of the area | Numeric |
| violent\_crimes | Total number of violent crimes in the area | Numeric |
| homicides | Number of homicides | Numeric |
| rapes | Number of rapes | Numeric |
| assaults | Number of assaults | Numeric |
| robberies | Number of robberies | Numeric |
| months\_reported | Month of the year for which the data was reported | Numeric |
| crimes\_percapita | Number of crimes per 100,000 residents | Numeric |
| homicides\_percapita | Number of homicides per 100,000 residents | Numeric |
| rapes\_percapita | Number of rapes per 100,000 residents | Numeric |
| assaults\_percapita | Number of assaults per 100,000 residents | Numeric |
| robberies\_percapita | Number of robberies per 100,000 residents | Numeric |

Data preprocessing tasks:

1. To learn how to work with larger datasets, I will write a report to create additional records (up to 15k) and combine it with the existing dataset.
2. Missing values (technique to be decided)
3. Separate the city and state code values from *agency\_jurisdiction* column.
4. Find the latitude and longitude values for the location (city, state code).

**Interest**  
By analyzing and visualizing this data, I am trying to find the answers to the following questions-

* + Is crime in America rising or falling?
  + Which cities have a higher crime rate? Which cities are safer?
  + Is there a relation between the time of the year and the crime rate?
  + What type of crime are you more susceptible to in a given area?
  + Is the crime rate dependent on the population of the area? If yes, how?

**Analytics**

1. Aggregation:
   * + For every year,
       - For every state,
         1. calculate the total and average number of violent crimes
         2. calculate the crime per capita
         3. calculate the total and average number of rapes, assaults, robberies and homicides
       - For every city,
2. calculate the total and average number of violent crimes
3. calculate the crime per capita
4. calculate the total and average number of rapes, assaults, robberies and homicides
5. Clustering: clustering the cities into 2 groups *safe* and *unsafe* based on
   * + Total violent crimes
     + Crimes per capita
     + Total rapes/assaults/homicides/robberies

The user will be able to select one of the above-mentioned criteria to cluster the cities. The threshold for clustering will vary with each criterion and still needs to be decided.

**Visualizations**   
The following visualizations will be used:

1. Geo-spatial (US map) heatmap: To show the crime rate in different regions of the US
2. Time series chart: To select a range of years between 1975 to 2015
3. Scatter plot: To show the cluster of *safe* and *unsafe* cities
4. Pie chart: To show the distribution of crimes (rape, assault, homicide and robberies)
5. Line Chart: To show the trend in different type of crimes over the years (zoom to months)
6. Histogram: To show how the population of an area affects the crime rate
7. Table: To show detailed crime records of an area

**Interactions**   
All visualizations will be linked with each other.

1. Geo-spatial (US map) heatmap: Zoom (view details of cities in the selected area), selection and probing (display the crime statistics of the area)
2. Time series chart: Selection (a range of years)
3. Scatterplot: Selection (select a group of cities) and probing (display the details of a city)
4. Pie Chart: Selection (select specific types of crime) and probing (display the numbers and percentage of the chosen crime type)
5. Line Chart: Zoom (view the crime rate trend over a selected year) and probing (display number of each crime type for the selected year)
6. Histogram: Selection (select population ranges) and probing (display crime statistics for the selected population range)
7. Table: Selection (select specific crime records of interest) and sorting (sort table based on different columns)

*Other features:*

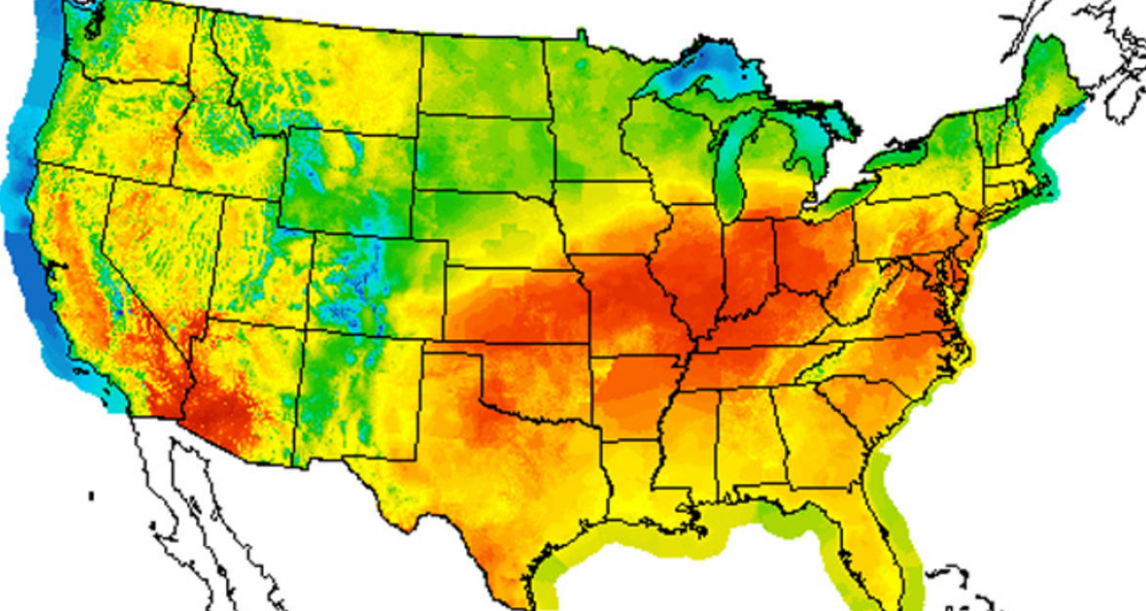
1. Parameter Selection for heatmap: View visualizations on Total Crimes (default) or Crimes per capita
2. Parameter Selection for scatterplot: Select criterion for clustering cities
3. Reset all graphs (clear all filters)

**CRIME IN CONTEXT: 1975-2015**

Descriptive text goes here --- Is crime in America rising or falling? Which cities have a higher crime rate? Which cities are safer? Is there a relation between the time of the year and the crime rate? What type of crime are you more susceptible to in a given area? I s the crime rate dependent on the population of the area? If yes, how? -- Descriptive text goes here --- --- Descriptive text goes here ------ Descriptive text goes here ------ Descriptive text goes here ----- Descriptive text goes here --- Descriptive text goes here --- Descriptive text goes here --- Descriptive text goes here --- Descriptive text goes here ---

Reset All

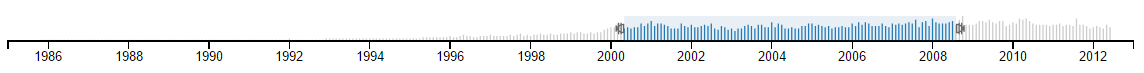
Change Color



---Color Legend for heatmap goes here---

Select Parameter:

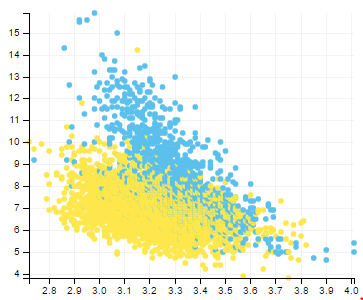
* Total Crimes (default)
* Crimes per capita



**Crime distribution across the US over the years (color: Total number of violent crime by default)**

Select criterion for clustering:

* Total Crimes (default)
* Crimes per capita
* Rapes
* Assaults
* Homicides
* Robberies

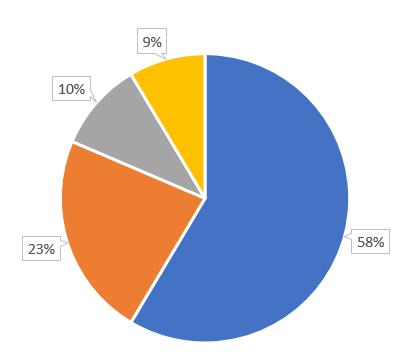


City

Year

---Color Legend for scatterplot goes here---

**Crime distribution across the US over the years (color: Total number of violent crime by default)**

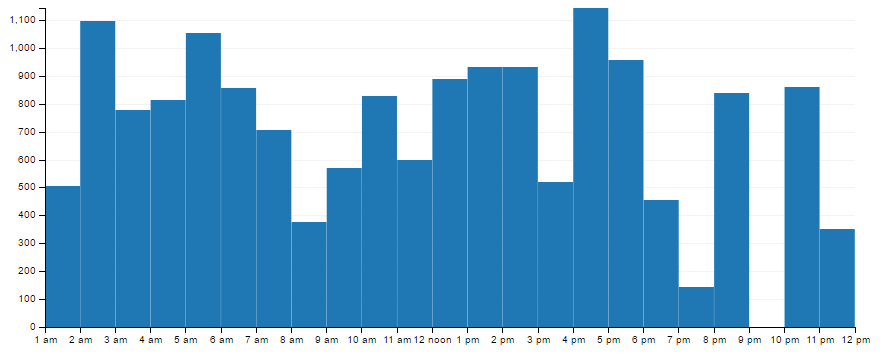
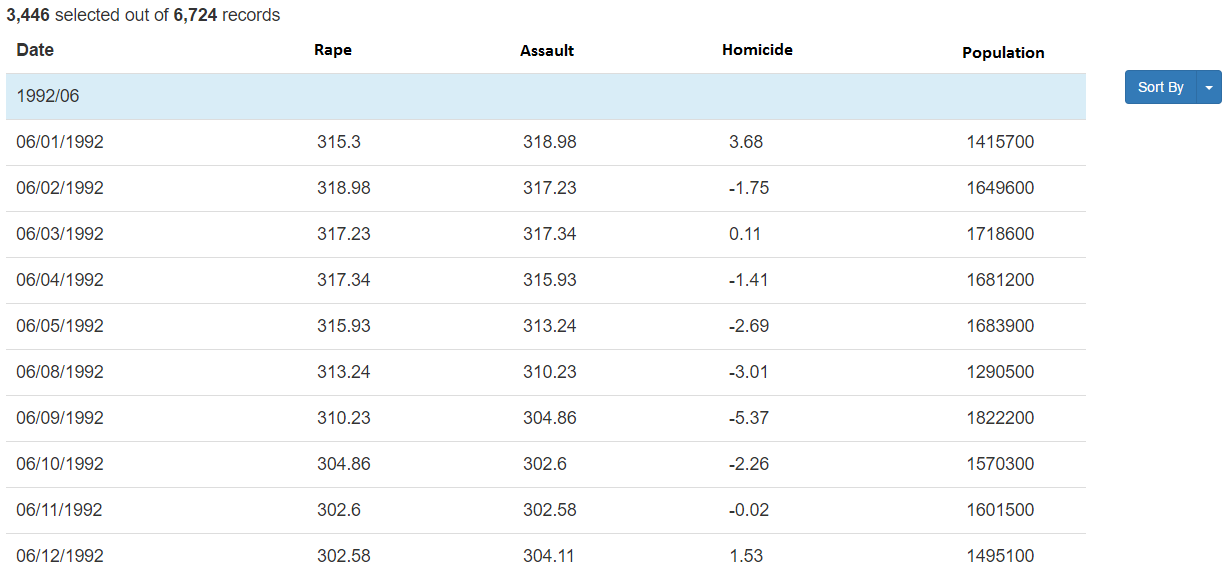


---Color Legend for pie-chart and line graph goes here---



**Type of Crime**

**Trend in crime rate over the years (Dotted line shows total crime trend)**



**Effect of population on crime rate (population vs the average number of crimes)**

Population

Crime rate

