San Jose State University

Department of Applied Science

One Washington Square, 95112

DATA 230: Sec-22 Data Visualization

DATA 230 – Final Project: Report - Team 08

The Relationship Between Parenting, Trauma with Delinquency: A Cross-National Study Using ISRD-2 Data

Course Instructor :	Seungjoon Lee
Team Members :	Akshat Patel (016800121)
	Ashritha Kumari (016798860)
	Dhrumil Shah (016781934)
	Prudhvi Chirumamilla (016803410)
	Shashank Reddy (016452501)

Submission Date: 05.09.2023

Table of Contents

1. Motivation

Understanding the complex interactions between trauma, family factors, demographics, and delinquency is crucial to effectively addressing the issue of juvenile delinquency. Visualizing these relationships through various graphs and charts can provide critical insights and inform interventions and policies aimed at supporting at-risk youth. By using visual representations such as scatterplots, bar graphs, and pie charts, we can gain a deeper understanding of the factors that contribute to delinquent behavior and identify potential risk and protective factors.

In conclusion, visualizing the relationships between trauma, family factors, demographics, and delinquency can provide critical insights into the complex interactions underlying juvenile delinquent behavior. By using various graphs and charts, we can identify potential risk and protective factors, guide targeted interventions, and inform policies aimed at supporting at-risk youth. This approach can lead to more effective and comprehensive strategies for addressing juvenile delinquency and supporting the well-being of youth in our communities.

2. Data Visualization

1: Bar chart of trauma and victim across nations.

2:Choropleth Map: Criminal behaviour and Closeness across nations

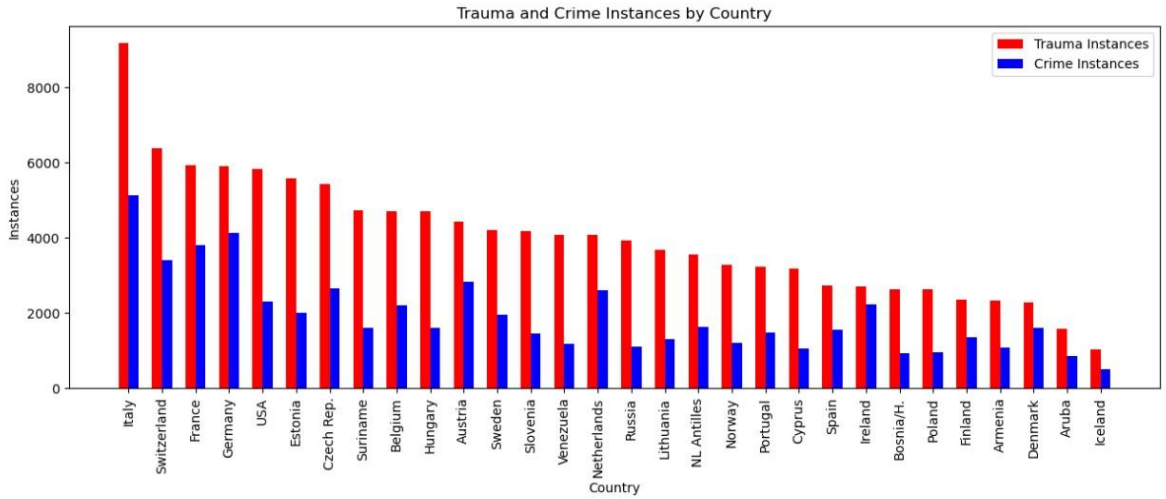
3: Heatmap of different types of victims engaging in different delinquent activities.

4: Nested pie chart of Genders committing delinquent activities based on their age.

5: 2- Dimensional PCA reduction of trauma and crime.

2.1 Double Bar chart of trauma and victim across nations

The bar chart map below shows the correlation between criminal behaviour and childhood trauma across nations (see Fig .1).



Correlation between criminal behavior and childhood trauma across nations.

The data processing involved finding the correlation between criminal behavior and childhood trauma across nations within the dataset that has either used drugs/alcohol or committed some sort of delinquency. Thus, the criminal behavior and Childhood trauma were calculated using the formula below.

$$r = (n\Sigma XY - \Sigma X\Sigma Y) / \text{sqrt}[(n\Sigma X^2 - (\Sigma X)^2) * (n\Sigma Y^2 - (\Sigma Y)^2)]$$

r : the correlation coefficient between criminal behavior and childhood trauma is the number of observations (or nations in this case)
ΣXY : sum of the product of the values for criminal behavior and childhood trauma for each nation
ΣX: sum of the values for criminal behavior across all nations
ΣY: sum of the values for childhood trauma across all nations
ΣX^2: sum of the squared values for criminal behavior across all nations
ΣY^2: sum of the squared values for childhood trauma across all nations

Here we're trying to find a correlation between the purpose of this visualization is to show the relationship, if any, between childhood trauma and criminal behavior across nations. Specifically, to highlight whether there is a correlation between these two variables and whether this correlation is positive or negative.

To achieve this, we are using ***Line(s)*** as **Mark**. The visualization is clearly displayed by the values for each nation's variables (childhood trauma and criminal behavior), using a side-by-side bar chart. In this case, each line in the chart represents a different country's data. The line chart is a useful mark to show how data changes over time or in this case, how the sum of traumatic and criminal cases varies across countries.

This visualization uses two **channels** to represent the data. The ***horizontal X-axis*** displays the categorical data of country names, with each line corresponding to a specific country's data and aligned with its name. The ***vertical Y-axis*** represents the quantitative data of the sum of traumatic and criminal cases, with the ***size(length)*** of each line indicating the magnitude of

this data. The longer the line, the larger the sum of traumatic and criminal cases for that country, allowing for easy comparison of the magnitude of data for different countries.

So, what are we trying to show with this visualization? We want to effectively communicate the complex relationships between delinquency per traumatic and criminal cases in different countries.

By providing a clear visualization of the data, we can identify which countries have higher or lower levels of these cases. This chart aims to highlight the variation in traumatic and criminal cases across different countries.

2.2 Choropleth Map: Criminal behaviour and Closeness across nations

The choropleth map below show the correlation between criminal behaviour and Closeness across nations (see Fig. 2).

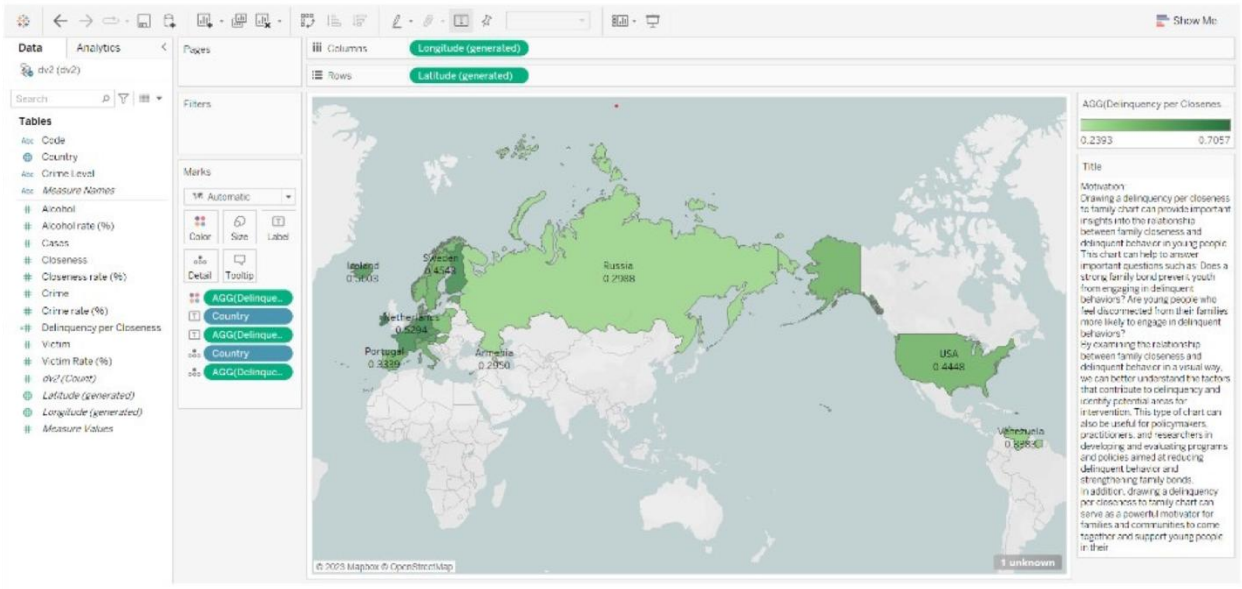


Fig. 2 :correlation between criminal behaviour and Closeness across nations

The data processing involved finding the correlation between criminal behaviour and Closeness across nations within the dataset that has either used drugs/alcohol or committed some sort of delinquency. Thus, the criminal behaviour and Closeness were calculated using the formula below.

$$choropleth(x, c) = N(X \ni x \cap X \ni c) \times 100\% \ (1)\ N(X \ni c)$$

X: The number of crimes committed
x: The level of closeness to family
$X \ni x$: The number of crimes committed by people who are close to their families
$X \ni c$: people who are close to their families
$N(X \ni c)$:The total number of people who are close to their families

We're trying to find a correlation between criminal behaviour and closeness to family across different countries. To do this, we've chosen some specific marks and channels that help us effectively communicate the data.

First off, we're using a two-dimensional area **mark** to represent each country. This is a great way to show spatial patterns and trends, making it easy to see how the Delinquency per Closeness variable changes across different regions.

Now, let's talk about **channels**. We've decided to use colour, specifically hue, to represent the Delinquency per Closeness variable. This allows us to easily spot variations in the data across nations. The darker the colour, the higher the ratio, which means lower closeness to family and greater delinquent behaviour in that country.

So, what are we trying to show with this visualization? We want to effectively communicate the complex relationships between delinquency per closeness and regions. This thematic map helps us explore these relationships by showing that as closeness to family diminishes delinquent behaviour becomes more prevalent.

With this visualization, we can easily identify regions with high or low values of the Delinquency per Closeness parameter, which could be useful for understanding the correlation between criminal behaviour and family closeness on a global scale.

2.3 A Heatmap Representing the Distribution of Delinquency Types among Juvenile Victims

The heatmap presented below illustrates the percentage distribution of victim types in relation to delinquency types among juvenile offenders, as depicted in Figure 3.

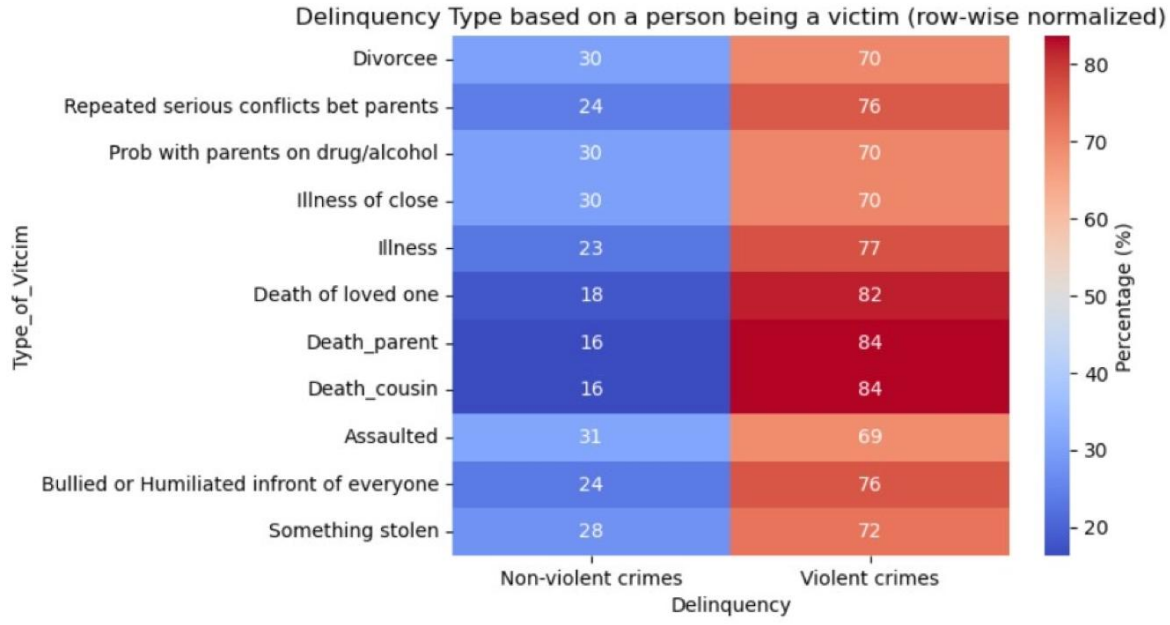


Figure 3: Type of victims corresponding to delinquency type (row-wise normalized)

Victim	Violent	Non-violent
VICTHEFP	ASLTLTP	BURGLTP
VICBULLP	GFIGLTP	HACKLTP
VICASSAP	SNATLTP	SHOPLTP
LIFEEV01	VANDLTP	BICTLTP
LIFEEV02	EXTOLTP	CARTLTP
LIFEEV03		
LIFEEV04		
LIFEEV05		
LIFEEV06		
LIFEEV07		
LIFEEV08		

Table: The columns utilized to classify a student as either a victim, a perpetrator of violent crime, or a perpetrator of non-violent crime**.**

During the data processing stage for the presented graphs, a significant challenge was encountered due to considerable overlap among the subcategories. For example, labeling a student who was bullied at school and experienced assault posed a labeling dilemma, as did the labeling of students who engaged in both violent and non-violent delinquent activities. To address this issue, students who committed both types of delinquencies were classified as violently delinquent students. Given the highly imbalanced distribution of traumatic experiences, the heatmaps were normalized row-wise, such that each row sums up to 100%. To calculate the percentage value in each heatmap cell, the formula below was utilized:

$$Heatmap(\text{vic}, del) = N(X \ni \text{vic} \cap X \ni del)N(X \ni \text{vic})\} * 100\%$$

vic: The most severe type of traumatic experience faced by a student (y-axis)

del: The most severe type of delinquency committed by a student (x-axis)

The presented heatmap displays two categorical variables, namely the type of victim and the delinquency type, and one quantitative variable derived from the data, which is the percentage of crimes committed by different victims. The heatmap consists of a 2D grid of rectangles, with each tile representing a data point. The position and color of each tile are used as visual channels. The x-axis categorizes the severity of delinquency as violent and

non-violent crimes, while the y-axis groups the types of victims according to the severity of the traumatic experience they faced. These include theft, bullying, public humiliation, assault, death of family members, illness of self or loved ones, drug/alcohol-related problems with parents, and parental conflicts leading to divorce. The position of each data point in the grid is mapped using these channels. The frequency of crimes committed by different victims is indicated using a color-coded system.

The rationale behind the selection of these marks and channels lies in the fact that the intensity of the color reflects the severity of the delinquent type being committed. Heatmaps are a useful tool for detecting patterns and trends in data sets. In this case, the heatmaps effectively communicate the percentage of violent and non-violent delinquent activities for different types of victims. The heatmap displays numerical data in the form of a color-coded matrix, where each color represents a different value. The color gradient helps to quickly identify areas of high and low values, allowing for the detection of patterns and trends.

The heatmap was utilized to verify the correlation between delinquency and type of victim, as well as to analyze the degree of violence in students for each traumatic experience. The plot aids in estimating the probability of a student with a particular traumatic experience being a certain type of delinquent. The results indicate that individuals who have experienced traumatic events are more likely to commit violent crimes than nonviolent ones.

By effectively communicating complex spatial patterns and trends, the thematic map examines the relationships between delinquency types (violent/nonviolent) and the victims. Based on the heatmap, it can be concluded that victimized individuals are more likely to commit violent crimes rather than nonviolent ones.

2.4 Nested pie chart of Genders committing delinquent activities based on their age

When tasked to find who has committed the most crimes, we classified students as young males and females committing violent or non-violent crimes and reclassified them on the basis of their age.

To showcase this effectively we have chosen to plot a nested pie chart

here we calculated individually from the dataset the count of males and females from 6th to 8th grades of students committing crimes (violent/non-violent) and then plotted them

Here the delinquency factor takes into consideration - violent or non-violent. Violent delinquency and Non-Violent Delinquency overlook all the factors – damaging, burglary, or threatening someone.

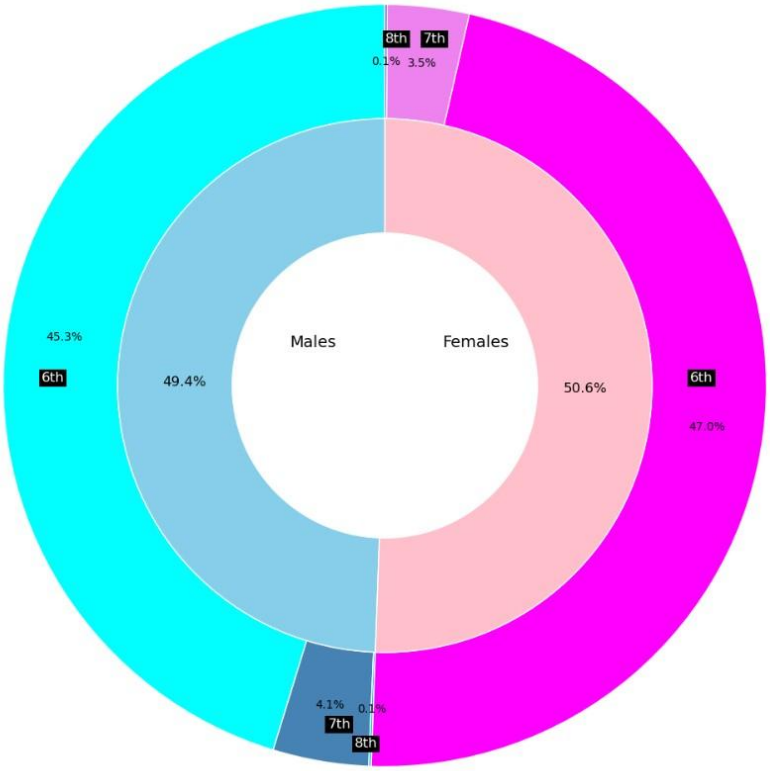


Figure 4: Percentage of Gender-wise and grade-wise delinquency amongst youth.

$pie(xi) = N(X \ni xi) / \sum i \in all\ classes\ N(X \ni xi) \times 100\%$

xi : Delinquency (Violent / Non-Violent) i
$X \ni xi$: Set of students that have done xi before on the basis of their gender and grades.
$all\ classes$: Set of all classes possible for x

Here the Data is 2 Categorical (Gender, Grade), and 1 Quantitative (Derived Data Grade, Gender Percent).

The **Marks** are - Area radially separated with uniform height.
The **Channels** used are - Hue (Color) and Angle (2d Area).

Task - Part of the whole judgment of young males and females committing Crimes.

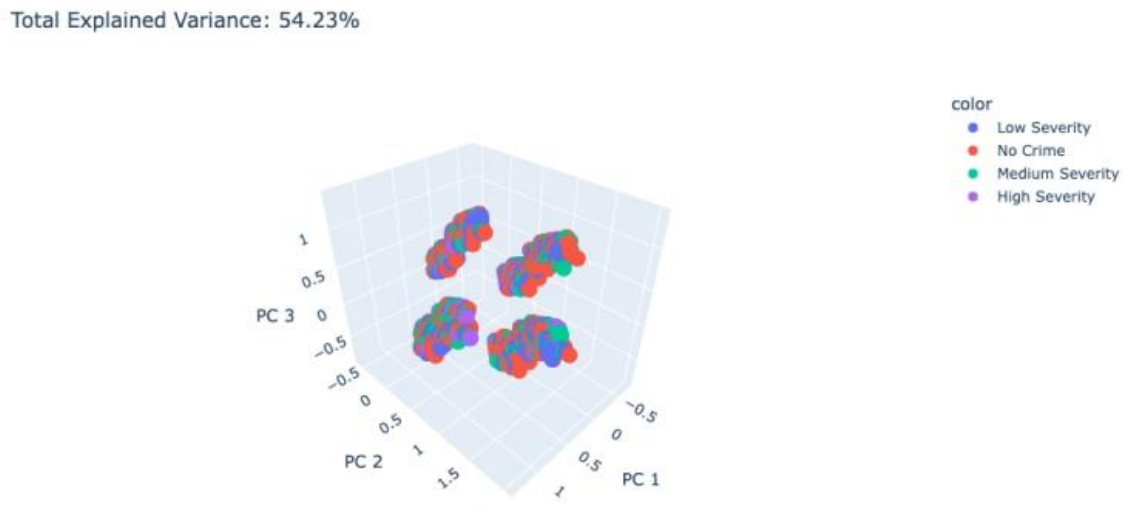
The main motive behind plotting the plot is to delve deeper into the statistical analysis and distinguish percent young males and females committing crimes. From the plot, we see that the majority of the crimes committed are by females who are in their 6th grade ~47% followed by males in their 6th grade ~45%

We can use this data can help identify which age groups have higher rates of delinquency. The chart can be particularly useful for policy-makers, law enforcement agencies, and social workers who are working to prevent and address juvenile delinquency. By identifying the specific age groups and gender categories that are most affected by delinquency, interventions can be better targeted to address the root causes and reduce the prevalence of delinquency, and development of healthy community

This type of data visualization that displays hierarchical data using concentric circles helps an individual to effectively, and efficiently fathom and makes it visually appealing in conveying the percentage of males and females individually committing crimes.

2.5 3D Scatter plot: Principal Component Analysis for Delinquency Severity Levels

The scatter plot below shows the first three principal components of PCA on variables related to trauma grouped by delinquency severity level. To perform the analysis, variables were created based on questions such as "Have you been assaulted?", "Have you been bullied?", "Do you live with your mother or father?", "Have you experienced the death of a family member or loved one?", "Have you experienced illness in your family or among loved ones?", and "Have you experienced problems or separation with your parents?" The delinquency score was based on the number of instances of involvement in various types of crime, ranging from zero (no crime) to 1-4 (low severity crimes), 5-8 (medium severity crimes), and 9 or higher (high severity crimes).



The 3D scatter plot uses **points** as **marks**, with **position and color** serving as **channels**. The color helps to categorize the severity levels of delinquency of the students.

The PCA was used during **exploratory data analysis** to identify the most important features that explain the variability in the data when classified by delinquency severity. Results indicated that the variables related to the death of loved ones had the highest contribution, followed by illness of parents. However, as can be observed from the plot and the fact that the explained variance ratio is 0.54 even with the third principal component, the issue cannot be attributed solely to trauma, given the other features included in the analysis. Therefore, while we have found important correlations between trauma and delinquency severity among juveniles, this analysis suggests that other factors may also be contributing to the problem.

3. Arrangement

- The order of the visualization techniques mentioned can be based on a logical progression of analyzing the data.
- Discovery - A bar chart showing the frequency of trauma and victim cases across nations provides a quick overview of the incidence rates, enabling identification of countries with the highest and lowest rates.
- Exploration -A choropleth map can show the distribution of delinquency cases by location, aiding in identifying patterns and hotspots globally.
- In-Depth Analysis - A victim-activity heatmap can uncover correlations and vulnerable victim types by examining patterns of delinquent activities among different victim categories.
- Data Distribution -A nested pie chart displaying genders and ages of delinquent activity can identify common age and gender groups involved.
- Statistical Analysis - A 2D PCA reduction of trauma and crime can identify patterns and relationships, providing insights into underlying factors and potential interventions.
- Therefore, the order of these visualization techniques can be based on a logical progression of analyzing the data, starting with a high-level overview and then gradually delving deeper into the relationships between different variables.

4. Summary/Result

1. Discovered a strong correlation between the traumatic instances and crimes committed (Bar Graphs)
   The greater the traumatic cases (Parenting and Victim) in a particular nation, the more likely is a delinquent act committed.
2. The thematic map provided an efficient way to plot the geographic distribution of the Delinquency per closeness ratio (Choropleth Graph).
   Found an inverse relationship between closeness and delinquency.
   Hence, the greater the closeness in the family, the lesser the number of delinquent acts committed.
3. The color coding system in a 2D grid represents how likely a person victim of a particular scenario is to commit a crime efficaciously. (Heatmap)
   The map clearly depicts a victimized person committing more violent crimes than nonviolent crimes.
4. The concentric circles provide a coherent way to clearly distinguish a male or a female in a particular grade (7th - 9th) committing a crime. (Nested Pie Chart)
5. Reduction in dimensions and retaining the necessary information effectively helped denote patterns and relationships within the data to become more clearer. (3-dimensional Principal Component Analysis)

References

[1] Examining the link between traumatic events and delinquency among juvenile delinquent girls: A longitudinal study.

[2] Profiles and behavioral consequences of child abuse among adolescent girls and boys from Barbados and Grenada.

[3] Parenting practices and adolescent delinquency: COVID-19 impact in the United States.

Appendices : Summary of the Visualizations

Plot	Bar Chart	Choropleth Map	HeatMap	Nested Pie Chart	3d Scatter Plot
Data Processing	Comparing the quantitative variables for different Countries via plotting a side-by-side (double)bar graph.	Communicate information about the distribution of a Delinquency per Closeness variable across different regions.	The ordering of labels was performed based on the severity of violence to categorize the data.	Categorizing based on genders and grades to find who commits most crimes and at what age.	Categorizing based on violence severity level and under-sampling for visualization discriminability.
Marks	Line(s)	Item (2 Dimensional) AREA	2D grid of rectangles	Area radially separated with uniform height.	0D Point
Channels	Position and Size(Length)	Hue (Color)	Position and Hue (Color)	Hue (Color) and Angle (2d Area)	Position and Color
Why?	Double bar graph make it easier to comprehend traumatic and criminal instances for a	It uses color to show the values of a variable. In this case, the variable is delinquency	Verify the correlation between traumatic experience faced and delinquency	To delve deeper into the statistical analysis and distinguish percent young males and	Observe the key factors for predicting delinquency violence levels.

	particular nation. Observing the relation between criminal instances and trauma across nations	per closeness. The darker the color, the higher the level of delinquency per closeness. This allows us to easily see the spatial patterns of delinquency per closeness across different regions.	using different data type	females committing crimes	Demonstrating that the data cannot fully account for the variation in the delinquent kinds.