Crime Data Visualization

Babloo Markose
College of Information
University of North Texas
Denton,USA
babloomarkose@my.unt.edu

***Abstract*—The first section of this article investigates the emergence of data visualization as a critical tool in decision making. As the world moves from the information age to knowledge age, data which was earlier seen as a resource, is now a business enabler and we are recounting how data visualization has aided that transition from historical times to present. This section describes various concepts and techniques of contemporary data visualization that can aid in a deeper understanding of the data available to us. First part of this paper attempts to give a bird's eye view of far-reaching applications of data visualization by reviewing recent literatures available on the topic. The cognitive amplification bestowed by visualizing data can pave the way to knowledge creation and helps take humanity to the next level of intelligence. The second part of this paper is a domain specific study of an application of visualization. In this part visualization is used as an analytic tool to carry out a descriptive data analysis of Denton county crime data. This part validates the fact that, for the right kind of data visualization can be used to carry out entire data analysis and is not a secondary or supporting tool but the primary tool for data analysis.**

Keywords—decision making, visualization concepts, cognitive amplification, history of visualization, visualization techniques, applications of visualization, crime data visualization, data analysis, time series analysis, geographic location-based analysis.

I. Introduction

Visualization is an important data analytics tool and classical definition of visualization is as follows: the formation of mental visual images [1] or in other words a process of explaining the meaning of data in visual terms. With the advent of computer-generated visualization, this definition can be modified as 'a tool or method for interpreting image data fed into a computer and for generating images from complex multi-dimensional data sets' [2]. The purpose of visualization is not limited to conveying information to others, and it also serves in recording information and supporting reasoning. So, a more apt definition will be: Scientific Visualization is concerned with exploring data and information in such a way as to gain understanding and insight into the data. The goal of scientific visualization is to promote a deeper level of understanding of the data under investigation and to foster new insight into the underlying processes, relying on the humans' powerful ability to visualize. In many instances, the tools and techniques of visualization have been used to analyze and display large volumes of, often time-varying, multidimensional data in such a way as to allow the user to extract significant features and results quickly and efficiently [3].

The human brain process information iteratively and that process is augmented using a graphical and pictorial representation of data [4]. Data are the information on which some level of analysis is done to reveal meaning, which is presented to the viewer as picture using visualization. It helps decision making more comfortable by changing the way information is perceived and processed. The design criteria should revolve around expressiveness and effectiveness in the sense that the visualization should depict all the facts and only the facts and the depiction should be in a way that the data can be readily perceived [5]. Clarity, precision, and efficiency make the visualization excellent [6]. More data, in less space, using less ink makes visualization useful. An effective visualization increases support for analysis and make visualization a significant step in the knowledge discovery.

It should be noted here that visualization can be used from EDA to present results. But for some kind of studies that pertains to spatial studies or time series studies it is more appropriate to conduct data analysis using only visualization. One such study is crime data analysis. Data analysis is even proposed as the most critical weapon that our intelligence and law enforcement agencies should carry [7]. In the present scenario visualization and for that matter even data analysis is only employed by high profile law enforcement agencies dealing with terrorism and organized crimes. The effort of tis paper is to establish that data analysis, particularly visualization can be used to effectively track and describe criminal behavior even at a county level.

The main challenges, in an effort to implement data analysis in such a low-level jurisdiction is the lack of access to sophisticated tools. This paper attempts to break the dilemma that arise due to having lots of data and not having access to high end tools [7]. In this research easily accessible and low-cost tools like excel and tableau were used to clean process and draw insights from relatively medium sized dataset. This is a revolutionary technique that can be implemented in small scale law enforcement precincts which was done manually till date. It should be noted that there is very limited study on crime data in scholarly and academic arena, even those present looks at big data with sophisticated data mining techniques. There is a lack of studies covering crime data analysis at a precinct or county level and even less studies conducted using open access, freely available tools and techniques. Hence this paper aims to establish that one need not have access to high end tools to conduct crime data analysis.

Furthermore, there is lack of studies that predicts crime type and crime locations [8] and this paper attempts to concentrate on that aspect of crime data analysis. Several studies have focused on identifying patterns among criminal incidents. These studies found spatial and temporal crime occurrence patterns. They additionally demonstrated the relationship between the occurrence of a crime and information about the surrounding area [8]. But these patters dealt with correlation of crime spot to location and didn't study stand-alone effect of location on crime count and time on crime rate which will be the focus of this study.

This paper is organized into two subsections namely, Literature Review and Data Analysis. Literature review includes the current knowledge including substantive findings, as well as theoretical and methodological contributions of different researches in the field of data visualization. This section is organized in such a way as to establish the concept of data visualization first, in Section II. It will then give the readers a high-level view of the historical emergence of data visualization in Section III. Data visualization creation and presentation techniques are covered in sections IV and V respectively. Section VI takes the readers through different domains that visualization can be applied effectively. The second subsection Data Analysis pertains to Crime data analysis of Denton county crime dataset. This section establishes how visualization technique can be used to effectively conduct data analysis of raw data collected through law enforcement agencies at county level using very easily accessible tools and techniques against all established norms. This is covered in Section VII onwards.

Literature Review

II. Basic concepts of data Visualization

Envisioning information, that the world around us has in multi-dimension, on a two-dimensional space is the major challenge of data visualization. However, that is not a challenge that stands alone. Understanding the data and drawing knowledge from it adds to this challenge. Hence, enhancing the understanding using details and enhancing information using layering and separation, along with usage of a series of images forms the basic principles of information visualization. Multivariate nature of all the information in the multidimensional space coupled with the vastness of data available, due to the data explosion that happened in the information age has challenged data visualization to evolve.

The evolution of data visualization imparted the ability to represent more data in a restricted space and reveal hidden meanings in data. The process came at the cost of great compromise of loss of data, but that can be justified with the cognitive amplification that the visualization process imparts. The basic steps necessary for scientific visualization is depicted as a reference model in Fig. 1 [9].

A. Reference Model

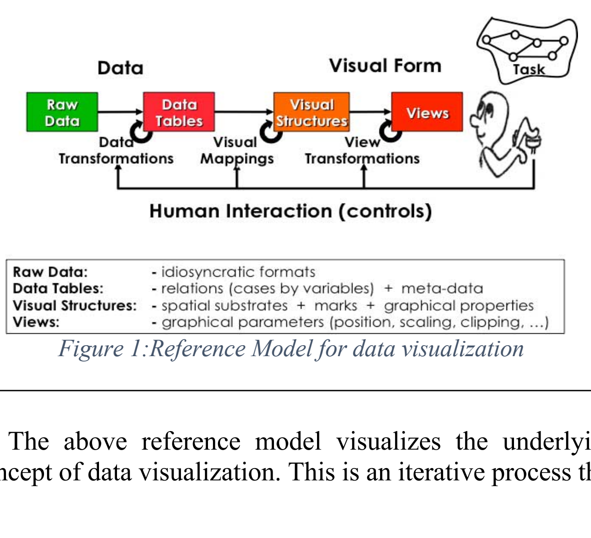

Figure 1:Reference Model for data visualization

The above reference model visualizes the underlying concept of data visualization. This is an iterative process that perfects the visualization through multiple cycles. Here the raw data undergoes a series of transformation, that makes it suitable to be represented in a visual format.

1) Data Transformation

The raw data gets transformed into data tables in this step. The data transformation includes transforming values to derived values and derived structures, and structures to derived values and derived structures. Errors and missing values in raw data gets addressed through these transformations. These include the following transformations.

a) Data Normalisation

b) Data rezoning

c) Data filtering and smoothing

d) Coordinate transformation

The result of the transformation is a data table which should not be confused with a table containing data. The data table is characterized by metadata and defining characteristics of data and has a mathematical function defining it. Metadata is the data about data, and that describes other data, and characteristics of data are defined by data types.

e) Data characteristics/Types

Data can be categorized as qualitative and quantitative. Qualitative variable gives information on quality that cannot be measured. It can be divided into nominal and ordinal data. Nominal data refers to unordered symbolic names whereas ordinal data refers to rank-ordered data. On the other hand, quantitative data can be scalars, vectors or tensors on which mathematical operation can be done. Spatial, geospatial and temporal variables come under this category.

2) Visual Mapping

Mapping data to useful visual structures that cater to the properties of human perception is the next step in visualization. To understand the selection of useful visual structures it is essential to understand the functioning of human perception

a) Human Perception

The human eye system maintains a constant computationally parallel surveillance system over the entire visual field. At the same time, it is continually moving the position of the foveola, sampling from the visual field to build up a percept or to attend to areas of high information content, such as moving objects. Visual perception is an active process in which head, eye, and attention are all employed to amplify information per unit time from the visual world [10]. This visual information gets processed in two different ways as controlled and automatic processing. Controlled processing happens through conscious effort and helps in processing textual details whereas automatic perception is loaded independent and does not require conscious effort [11]. Data visualization should appeal to both controlled and automatic perception to effectively communicate data. This processing assigns meaning to dimension forming a spatial external working memory and working on that can improve cognitive functions. This space is used by visual structures like marks, graphical properties of maps and spatial substrates to represent data.

b) Visual Structures

Spatial structures: They include a physical and geospatial coordinate system that is the most effective perceptual tool. It can be used as a frame that can hold other visual structures.
Marks: Points, lines, areas, and volume are the optical structure that can hold information in the spatial structure.
Graphs and Trees: These are used to visualize the structure of data with a focal point.
Retinal Variables: Position, size, orientation, color, hue, texture and shape appeal to the retina of human eyes. Hence these are called retinal variables.

c) Temporal Encoding

Human brain perceives data iteratively and time is an effective baseline to appeal to this property. Change of data on a time series can be perceived effectively, and hence such a representation is the most used one in visualization.

d) Visual Encoding of Data

Choice of visual structures to encode data depends on the data type and relative importance of data that is being presented. Some structures are more effective than others for specific data types. It is important to choose efficient encodings that express the data adequately.

The different types of visual mapping are concept maps, mind maps, conceptual diagrams, and visual metaphors [12]. These rendering techniques have their advantages and disadvantages which are depicted in Fig.2 [13]and Fig.3 [13].

Format	*Concept map*	*Mind maps*	*Conceptual diagram*	*Visual metaphor*
Main advantages	1. Rapid information provision[15] 2. Systematic, proven approach to provide overview[15] 3. Emphasizes relation-ships and connections among concepts[17] 4. Ability to assess quality of concept map through evaluation rules[2]	1. Easy to learn and apply[6] 2. Encourages creativity and self-expression[5] 3. Provides a concise hi-erarchic overview[6] 4. Easy to extend and add further content[6]	1. Provides a concise overview[24] 2. Structures a topic into systematic building blocks 3. Assures that main aspects are con-sidered 4. Can be applied to a variety of situa-tions in the same manner	1. Serves as a mnemonic aid (method loci) 2. Draws attention and in spires curiosity 3. Activates prior knowl-edge about metaphor domain[13,25–27] 4. Facilitates understand-ing by triggering func-tional associations[13]

Figure 2 :Advantages of different visualization concepts

Main disadvantages	1. Not easy to apply by novices; requires exten-sive training[17] 2. Concept maps tend to be idiosyncratic[17] 3. Time consuming evalu-ation through tutors[17] 4. The overall pattern does not necessarily assist memorability	1. Idiosyncratic, hard to read for others 2. Represents mostly hierarchic relation-ships[6] 3. Can be inconsistent 4. Can become overly complex (loss of big picture)	1. Can be difficult to understand with-out knowledge of category mean-ings 2. May not be appli-cable to the topic at hand 3. Does not provide mnemonic help 4. Does not foster creativity or self-expression	1. Cannot easily be extended or modified 2. May be misunderstood, may trigger wrong asso-ciations 3. Can be difficult to draw (rapidly) 4. Can be manipulative or misunderstood[23]

Figure 3:Disadvantages of different visualization concepts

Concept maps: They are graphical tools used to depict suggested a relationship between concepts. Ideas and information are represented as circles or squares and relationship with linking verbs. It represents a hierarchical model of related concepts. A concept map is a top-down diagram showing the relationships between concepts, including cross-connections among concepts, and their manifestations.

Mind maps: Visually organize concepts around a single main idea. A mind map is a multicolored and image centered, radial diagram that represents semantic or other connections between portions of learned material hierarchically.

Conceptual diagrams: A conceptual diagram is a systematic depiction of an abstract concept in pre-defined category boxes with specified relationships, typically based on a theory or mode
Visual metaphors: A visual metaphor is a graphic structure that uses the shape and elements of a typical natural or humanmade artifact or of an easily recognizable activity or story to organize content meaningfully and use the associations with the metaphor to convey additional meaning about the content.

3) View Transformation

Creating views of visual structures by manipulating graphical attributes helps in instilling ideas to the viewer. These transformations include distortion, location and viewpoint controls. These transformations pertain to pinpointing important data from less important one. Moreover, scaling and labeling data help to augment cognition. It deals with zooming panning and clipping on the most relevant data. Creating focus and context for the data also comes in view transformation.

B. Concepts supporting and automating visualization

Human-interactions support and control visualization techniques in a variety of ways which is imperative to ensure detecting and pruning misadvised and unworkable data visualization. According to Gitta Domik of University of Paderborn, Germany, concepts that describe possible support or even automation of the visualization process in addition to the Reference model given in the previous section are given below [14].

API: A presentation tool is based on generating and test the concept and primarily uses two-dimensional presentations.
BOZ: Boz is a task-based automation tool that takes into account the intended purpose of visualization.

AUTO VISUAL: Auto visual helps in generating a three-dimensional visual world to direct focus on the relationship between data.

METADATA: Rogowitz and Treinish identified two classes of rules which are essential for the visualization process: Class I rules are designed to ensure that the structure of the data is faithfully represented in the visualization. These are either rules for providing an isomorphic mapping from data to perceptual dimensions or rules for ensuring the perceptual invariance of higher-level visual features, such as color, size, and shape. Class II rules intentionally transform the structure of the data to highlight features in the data to attract attention, scale dimensions to exaggerate details, or segment the image into regions [15].

GLYPHMAKER: Glyphmaker is a system for data analysis and visualization [16].

VISUALIZATION IDIOMS: According to Haber and McNabb the three significant transformations in conceptual visualization are data enrichment, visualization mapping and rendering [17]. These form the idioms for visualization.

MATRIX: Wehrend mapped the visualization techniques, and the task needs to be accomplished with the technique in a matrix format as objects against operations cataloguing the techniques [18].

NSP: Natural scene paradigm is based on the capability of the human brain to interpret natural scenes. Identifying patterns and distinguishing between the object in foreground and background is easier for the brain when represented in a natural set up [19].

VISTA: Visualization tool assistant can be best described as a knowledge-based system that helps scientists design visualization techniques. It generates a technique for a given data set and lets users modify the design interactively [20].

SAGE: A system for automatic and graphical explanation is an iteration of APT which adds a semantic description of data in the visualization [21].

VIS-AD: Visualization for algorithm development is used to describe an algorithm in graphical terms [22].

III. History of Data visualization

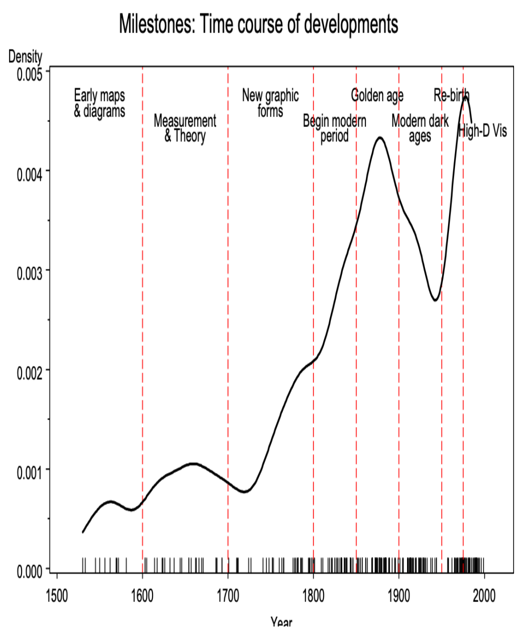

Figure 4: The time distribution of events considered milestones in the history of data visualization

The history of data visualization is the history of data. From cave paintings to modern databases, data was rendered and scaled down catering to the size of data available. Statistics has progressed by gaining in accuracy and has already broken away from the shackles that restricted one to consider only aggregate statistics. More exceptional control of data paved a way to the emergence of data visualization techniques that project more excellent details in the massive amount of data. This section dwells deep into this transition.

Fig. 4 [12]visualizes the history of modern visualization. Before 17th-century visualization was in the infant stage, but it was not new. The use of graphical depiction of data dates back to the Pleistocene era, where cavemen used to visualize stellar data on walls of a cave [23]. Physical artefacts such as Mesopotamian clay tokens (5500 BC), Inca quipus (2600 BC) and the Marshall Islands stick charts (n.d.) can also be considered as visualizing quantitative information [24]. The thematic cartography was the most dominant form of data visualization until the 16th century which visualized information using geographical illustrations. Developments in statistics and probability theory helped William Playfair to generate and develop graphical methods of statistics [25]. He is considered the father of modern scientific visualization. This caused the transition of visualization into measurement and theory era.

The next era in visualization was in the 18th century as isolines and contours were invented, and visualization entered new domains of economic and medical data representation. From 1800 to 1850, the advancement made by data visualization was substantial. Modern forms of data depiction like line graphs bar charts pie charts were invented and novel forms of symbolism were invented during this time [26].

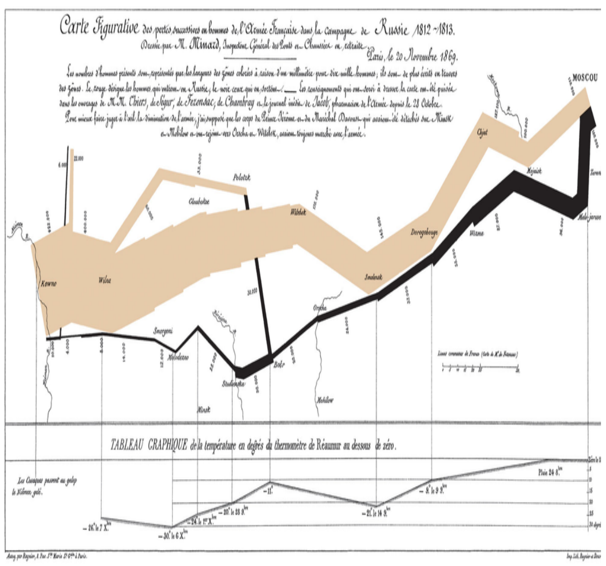

Figure 5 The famous "Best Graphic ever made"

The latter half of the 19th century was the golden age of data visualization. This is the time when the visualization indeed escaped flatlands and was widely represented in multi-dimensions. Graphical innovations by Minard in using flow lines on maps of width proportional to quantities (people, goods, imports, exports) to show movement and transport geographically helped him in visualizing [27] Napoleon's March to Russia, a visualization called as "best graphics ever produced" which is given in Fig. 5 [28]. Following the golden age, visualization was met with a dormancy stage as statistical models took the main stage forcing visualization to the backstage. In this dark age, only a few innovations happened, and formal quantitative models gained importance over graphics.

From 1950 visualization gained momentum with data analysis being adopted as a mainstream branch of statistics. Computer science, input and display technology advancements, helped it to have a rebirth. However, after 1975, it had a fast-paced growth with software tools and applications augmenting large scale analysis and visualization of “Big data.” This was when interactive visualization gained popularity, and the cost of interactive visualization was bought down by the massive increase in computing power and programming capabilities.

Figure 6 Data Visualization History at a glance

Given Fig. 6 [29] is an image of an interactive map from infowetrust.com which takes us on a journey of all of history’s best graphs, charts, maps, and diagrams. Beginning with Christoph Scheiner’s *Rosa Ursina sive Sol* (1630), which mapped the motion of sunspots over time, and ending with Edward Walter Maunder’s *Distribution of Spot Centres in Latitude* (1904), a butterfly diagram used to study the variation of sunspots over time, this whimsical map explores 75 different visuals in a fun and easy to understand way [30].

IV. Data visualization Presentation techniques

Data visualization sets a platform for relaying the data and only the data, effectively without distortion and distracting audience. This requires extensive planning, knowing one’s data and making the right decision about the tools and techniques. The different steps included making a visualization effective can be summarized as below.

A. Cater to the audience

An important aspect most presenters overlook is the fact that the audience may not be as technically or statistically inclined. When data visualization is being prepared, one should keep the audience in mind and present it in such a way that the audience can follow along.

B. Know one’s data and domain

The validity and authenticity of data need to be assured before making visualization of it. This is because a visualization can to render lies as truths and bend facts. It is the responsibility of the data analyst and visualization expert not to lie using visualization. However, that is not enough. One should know what one wants to achieve using visualization. This knowledge empowers the visualization expert to visualize the data better.

C. Cluster and categorize data

Extensive exploratory analysis of data should go into the analysis phase to understand and compartmentalize or generalize the data. This can help in making effective presentations using the information derived from data.

D. Prioritize and label data

As important as it is to know one’s data, it is imperative that the one has extensive knowledge about visualization technique. Prioritizing the different levels of information, one wants to convey, can help in choosing what to present in which format. It is as important as this to label the data so that there is no ambiguity.

E. Choose the right form of presentation

Choosing the right form of presentation can take the cognitive amplification provided by visualization to the next level. On the contrary, failing to do so will extensively affect the power of visualization.

F. Make use of digital aids/tools

An array of visualization tools is available and making use of these tools will reduce the time and effort needed to create an effective visualization. These tools come in free and paid versions which can be used according to the level of visualization expertise and data needs

G. Select Colors with care

Color plays a vital role in relaying information to the audience. This is evident from the fact that when one sees a red color one immediately thinks of danger. There is also a psychological aspect to this. Green and blue are said to be soothing and black to arouse fear. Hence choosing colors in visualization is as essential as the data one is presenting.

H. Use Interactive visualization

With the use of digital tools, it is far easier to create interactive visualization. However, the cost of this process is high. If the cognitive amplification provided by the visualization can justify the cost, then it is better to use this form. The downside of this is the fact that creating interactive visualization needs extensive domain knowledge.

I. Structure one's story

Apart from this, one need to consider the structure of the story one wants to tell using the visualization. Presenting the data linearly, iterating and adding to the knowledge previously presented makes the audience comfortable following the story one wants to tell.

These are the techniques for presenting data visualization. The following section elaborates on the techniques used in creating data visualization.

V. Data-visualization Creation techniques

Various use cases call for various techniques, and the properties of data and properties of the function of data decide which visualization technique is appropriate for the task.

1. Two-dimensional-Area Data Visualization
   The historical forms of visualization were depicting data on geographic maps. With the advent of big data, the volume of data needed to be presented using visualization exploded and data maps were found to be the most appropriate tool for realizing this goal. Different types of two-dimensional area data visualization include dot maps, area or distance cartograms and choropleth.

   Dot maps: Dot maps or dot distribution maps or dot density map is a type of map that uses a dot to show the presence of a feature or phenomenon. Dot maps use visual scatter to visualize the spatial pattern.

   Area or distance cartograms: Land area or distance is substituted for some thematic mapping variable like travel time, or population in a cartogram. Depending upon which is used land parameter is used the two common types of cartograms are area and distance cartograms.

   Choropleth: The average values or quantities of factors, like the value of a property or quantity of variable, within predefined areas are visualized using differences in shading, coloring, or the placing of symbols in those areas in a map. This form of visualization is called choropleth.

2. Multi-dimensional Data Visualization
   Easy to understand visuals are created combining one or more dimensions. The most common and widely used form of multi-dimensional data visualization is a Pie chart. Other multidimensional data visualizations are histograms and scatterplots.

   Pie charts: A visualization in which a circle is divided into sectors to represent a portion of the whole is called a pie chart.

   Histograms: Histograms are used to visualize univariate variables as a function of the variable's frequency distribution. Data are grouped into ranges and depicted as rectangles.

   Scatterplots: A scatter plot is the best way to compare two variables to detect the correlation between them. Variables are plotted on to dimensional graph along two axes.

3. Hierarchical Data Visualization
   The techniques stated above are useful when visualizing a single data set. However, hierarchical visualization comes in handy when one needs to visualize data from multiple data sets. It helps in comparing one or more datasets. Graphs, tree diagrams, dendrograms, and sunburst diagrams are potent visualization techniques that can be scaled even to visualize big data.

   Hierarchical Graphs: An abstraction of graphs and maps are combined to form hierarchical data visualization. Graphs which are too large are augmented with maps providing indices for easier navigation.

   Tree diagrams: Tree diagrams are used to visualize not only data elements but also the relation between them. Linear processes are best visualized using tree diagrams.

   Dendrograms: These are particular forms of tree diagrams that incorporate more details to show taxonomic relations.

   Sunburst diagrams: Concentric circles are used to represent hierarchical data where the center of the circle represents the root node; the child nodes grow outward towards the outer circles.

4. Network Data Models
   When the volume of data available increased it become a necessity to compare and relate different data sets. The data models prevailed until now fail short of this requirement, and visualization moved to network models. Different types of network models are matrix diagrams, alluvial diagrams, and node-link diagrams

   Matrix diagrams: Matric diagrams where grid-like visualization is used to depict relationships between two or more variables in rows and columns where marks are used instead of texts is another form of visualization.

   Alluvial diagrams: These are flow diagrams that represent changes over a given factor like time. They got their names from the visual similarity to the alluvial soil deposit pattern.

   Node-link diagrams: Nodes representing data elements are interconnected to other nodes forming a node-link diagram. These are used when the interconnection between data elements are not directional but spans in multiple directions.

5. Temporal data visualization
   Most businesses and researches base their studies of factors over time, and thus visualizations based on timelines emerged as a separate technique called temporal data visualization. Time series connected

scatter plots, and polar area diagrams are very handy in visualizing large data sets over varying timelines.

Time series: Simplest of temporal data visualization is time series where data points are indexed in time order. Equally spaced points in time are used to track the change in data.

Connected scatter plot: It is a form of time series where data is not continuous and is represented by points. These points are joined using straight lines to visualize a trend.

Polar area diagrams: Cyclic phenomenon that repeats itself after a period in time is best visualized using polar diagrams. These are similar to pie charts, but the data volume is represented using the area of the sector and not using the angle of the sector.

All these are extensively used in visualization, but this is not an exhaustive list of all techniques. The techniques detailed here are emerging ones that are relevant in analyzing big data. That does not mean traditional techniques like box and whisker plots or bubble clouds and other techniques are impertinent.

VI. Applications of data Visualization

Applications of data visualization are so vast that it is a topic in itself for another literature review. With the world in an information age moving towards knowledge age visualization has emerged as a business enabler. The data explosion paved the way to this transformation. The most critical application domains are scientific research, business, organizational communication. The purpose of visualization can be broadly classified as knowledge creation, Cognition amplification, and communication.

Knowledge Creation:
Researchers rely greatly on visualization to make sense of large volume of data in a limited time frame. Any analysis of data aims to create new knowledge and visualization is an important technique. Even prediction using data analysis is easily portrayed using effective visualization.
The applications of visualization in research includes exploring the literature, data and code analysis, connecting and communicating with others, processing data, writing, publishing, and evaluating research.

- Exploring the literature
  Exploration of literature can help in finding knowledge gaps and identifying areas in the interested domain that are candidates for the research. Effective visualization of data can decrease this exploration time.
- Data and code analysis
  EDA or exploratory data analysis is the primary step in any research. It aims at finding the best analysis for the data, finding assumptions for which analysis the data set satisfies and to understand the nature of dataset to determine if the sample can be generalized to describe a population. Histograms can verify the normal distribution and find data anomalies, outliers which can pose problems for data analysis.
- Processing Data
  Data processing using different transformations will be more understandable when visualized. To understand if a transformation on a curvilinear relation between two variables yield a linear relation, it is advisable to plot the transformed variables on a graph to analyze it.
- Evaluating Research
  The validity and statistical soundness of a model after data analysis can be easily estimated using visualization. For example, a residual plot of linear regression is more informative than a table of residuals to understand the validity of the regression model. If the predicted value deviates from the regression-line, then the model can be deemed as not efficient.
- Data Communication
  Data communication is the most important and the most frequent road blocker, which brings down the decision inducing capability of a statistically sound, highly productive results of data analysis. This is due to the fact that the data analysis team fail to efficiently communicate and convince the value of the results for their research to the stakeholders. This is an area which can be easily overcome using data visualization. Images speaks for themselves and will be easily perceived by the decision makers.

Cognition amplification
Creation of knowledge, in turn, amplify the cognition. From a lay man's perspective, the knowledge created in research is easily understood using visualization. This aspect of data visualization can be exploited in academia. Difficult scientific processes can be easily described using data visualization. A waterfall model or agile model of SDLC in computer science or a periodic table of elements in chemistry can be easily explained to students using images rather than oral lectures. Academic community also use visualization for publishing new ideas and researches.

Communication
Organizations use visualization for internal communication, client reporting, and marketing. Using data visualization, data becomes business enabler and gives a competitive edge and proof to the decisions made by decision makers. Data can be used to reflect on the factors that cause employee attrition or factors that help in retaining employees. Data visualization can help a manager to focus on the areas that needs improvement for employees to work efficiently. Effective communication using visualization can add credibility to the claims one presents and increases the impact the ideas presented have on the audience.

VII. Retrospection on Data Visualization

From the review, it is clear what Data visualization is not. It is not producing an excellent plot; on the contrary, it "should be pervasive and embedded in the environment of a good company. They are integrated into products and processes. They should enable action. ML (machine learning) and data visualization together augment human intelligence." We have limited cognitive abilities as human beings. However, by leveraging data, we can have a more sophisticated understanding of the facts that the data present and this is facilitated by data visualization.

When it comes to automating visualization, a variety of tools are available in the market. This range from pivot charts in excel to great interactive visualization using Tableau. A hurdle in the visualization arena in this regard is the hardness of using these tools available. The expertise-oriented tools will make it hard for non-specialists to adopt the use of these tools. It works well with visualization experts but in the changing world, data visualization is an integral part of any domain, and a more natural approach needs to be adopted to generalize the use of the tools.

Even though the concept of data visualization has drastically changed over the years, the basics remain the same, and only rendering got evolved. Hence to visualize any data one need not have access to high technical and domain knowledge and the following data analysis section establishes this fact. Everyday software applications like excel and tools like tableau are used in the research reported in the subsequent sections. This verifies that fact that descriptive or predictive study of data can take visualization analysis and render valuable insights.

Data Analysis

VIII. Executive Summary of Crime Data Analysis

Various studies have been conducted on the prediction of crime occurrences. This predictive capability is intended to assist in crime prevention by facilitating effective implementation of police patrols. Previous studies have used data from multiple domains such as demographics, economics, and education. Their prediction models treat data from different domains equally. These methods have problems in crime occurrence prediction, such as difficulty in discovering highly nonlinear relationships, redundancies, and dependencies between multiple datasets. [8]

Studies on crime history of a region can help the law enforcement agencies in decision making process and in policy making. It helps in identifying problem areas as well as the times of the day which need stricter policing. It also helps in validating the effectiveness of policies and procedures in place, and which policies are effective and which ones are obsolete.

The dataset used for this study was collected by the law enforcement agency of Denton and has over 84000 records. The data set has the crimes with location, time stamp, location type and accuracy of the location. The location address was partially suppressed to safeguard the privacy of victims and fully suppressed for victims of sexual crimes. Fig. 7 depicts a small portion of the initial dataset.

_id	crime	locname	incidentdatetime	publicadress	agency	accuracy
1	DRUG/NARCOTIC VIOLATIONS	COMMERCIAL/OFFICE BUILDING	2009-12-31T00:26:00	30XX COLORADO BLVD	Denton Police Department	Address
2	ALL OTHER OFFENSES	HIGHWAY/ROAD/ALLEY	2009-12-31T00:59:00	39XX E MCKINNEY ST	Denton Police Department	Address
3	DRIVING UNDER THE INFLUENCE	HIGHWAY/ROAD/ALLEY	2009-12-31T02:08:14	LINDSEY ST // FORT WORTH DR	Denton Police Department	Intersection
4	DRUNKENNESS	BAR/NIGHT CLUB	2009-12-31T02:51:17	10XX S AVE C	Denton Police Department	Address
5	DRUNKENNESS	HIGHWAY/ROAD/ALLEY	2009-12-31T02:44:49	13XX W HICKORY ST	Denton Police Department	Address
6	DRUNKENNESS	HIGHWAY/ROAD/ALLEY	2009-12-31T03:32:31	1XX W OAK ST	Denton Police Department	Address
7	ALL OTHER OFFENSES	RESIDENCE/HOME	2009-12-31T04:24:00	14XX BERNARD ST	Denton Police Department	Address
8	DRUG/NARCOTIC VIOLATIONS	RESIDENCE/HOME	2009-12-31T04:24:00	14XX BERNARD ST	Denton Police Department	Address

Figure 7:Initial Data Set

The goal of this study is to answer the following questions specific to location and trend.

- How safe is Denton?
- Which are the most dangerous streets in the Denton county?
- How Denton crime rates changed over years?
- Is there a prime time for the crimes?
- Is there any correlation between Crime rate and Population growth?
- What are your chances of being a victim?

IX. Exploratory Data Analysis

A. Pruning DataSet

The data set was pruned to remove data fields that was not adding value to the research. Initial analysis of data revealed fields in the data set that are of very limited use in data analysis. The fields 'agency' and 'accuracy' were removed from the dataset based on the following reasons.

The field 'agency' just had two distinct entries for the 84000 records as 'Denton Police Department' and 'TX0610200'. Inference was made that the entry recorded the agency which entered the particular record in the data set. Denton Police Department was easy to understand but which agency was indicated using TX0610200 was not clear as the data dictionary or metadata failed to describe it. Hence this field was not used in the data analysis.

The second field that was identified to be problematic in the initial analysis was the field accuracy. This filed also had just two distinct values for 84000 records and they were 'Address' and 'Intersection'. Since this field also had only limited power in a data analysis, this filed also was omitted from further analysis.

For year 2009, only 36 records were available and for the date Dec 31st. Hence these records were not considered for further studies. For year 2019 data was available for only 2 months and this was also omitted from further analysis.

B. Data Transformations

Fields present in the dataset underwent transformation to make it suitable for visualization analysis. The fields were address field and incidentdatetime.

To safeguard the privacy of victims the addresses were partially suppressed. Even though exact address was not available to facilitate location-based studies, street level addresses were obtainable. Excel built in functions were used to trim the address to street address and remove the suppressed part of the address. The field incidentdatetime was separated out into three fields time, date and year.

C. Adding Derived Fields

The following new fields were added which was derived from existing fields in the dataset.

1) Categorizing Crime

Based on the severity of crimes, the field was categorized into high, medium and low. Felonies were classified high, misdemeanors medium and infractions as low.

2) Population Statistics

To facilitate comparison of population to crime and calculate crime rates, yearly population of Denton county was added to the dataset.

3) Geocoding

To facilitate location-based studies, it was decided to add geo-mapping for the street addresses. Macros were used to geocode street addresses and obtain latitude and longitude. This data was visualized using Tableau resulting in Fig 8.

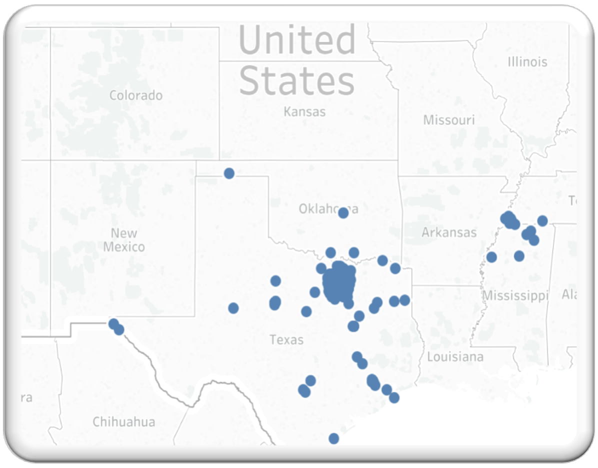

Figure 8:Data Discrepancy-Geocoded locations outside Texas

D. Correcting Data Discrepancies

Initial visualization of location-based data brought to light some data discrepancies like locations outside the state of Texas when the data was collected only for the Denton county.

Manual geocoding using tableau was done to correct the data discrepancies due to automatic geocoding.

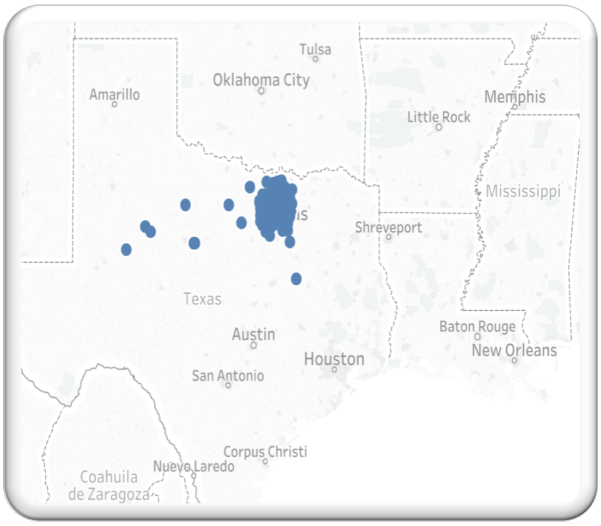

Figure 9: Manually corrected geocoded locations

But still few records showed locations outside Denton county as shown in Fig.9, which had to be removed, as the address where strictly from inside Denton county and those records were considered anomalies.

The clean data set in Fig. 10 was obtained after the above-mentioned EDA.

	A	B	C	G	H	I	L	M	N
1	crime	crimelevel	TypeofLocation	Year	incidentdate	Time	Address	Latitude	Longitude
2	SIMPLE ASSAULT	Low	RESIDENCE/HOME	2010	1/25/10	18:20:00	SPRINGFIELD ,Denton,Tx	31.6620407	-96.49918
3	DRIVING UNDER THE INFLUENCE	Medium	RESIDENCE/HOME	2011	10/18/11	23:40:27	SPRINGFIELD ,Denton,Tx	31.6620407	-96.49918
4	LIQUOR LAW VIOLATIONS	Medium	HIGHWAY/ROAD/ALLEY	2016	4/2/16	2:53:46	FANNIN MAP 511C ,Denton,Tx	31.6710475	-96.49170
5	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	CONVENIENCE STORE	2017	9/21/17	23:38:00	US-87 ,Denton,Tx	32.2630303	-101.4901
6	DRIVING UNDER THE INFLUENCE	Medium	HIGHWAY/ROAD/ALLEY	2017	9/16/17	17:11:29	ORIOLE ST ,Denton,Tx	32.39953	-99.4850
7	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	OTHER/UNKNOWN	2018	3/9/18	14:55:00	FOREST CEDARS DR ,Denton,Tx	32.40455	-99.5135
8	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	COMMERCIAL/OFFICE BUILDING	2016	2/12/16	14:20:00	FLINT HILL ST ,Denton,Tx	32.405003	-99.5087
9	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	DEPARTMENT/DISCOUNT STORE	2015	12/3/15	0:01:00	ELAM RD ,Denton,Tx	32.406728	-99.4975
10	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	BANK/SAVINGS AND LOAN	2017	8/25/17	15:40:00	FM 813 W. ,Denton,Tx	32.42916	-96.679
11	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	OTHER/UNKNOWN	2018	11/20/18	18:03:00	JOHN HENRY DR ,Denton,Tx	32.5579692	-97.266
12	FALSE PRETENSE/SWINDLE	Low	BANK/SAVINGS AND LOAN	2014	3/17/14	13:33:00	S HULEN RD ,Denton,Tx	32.6315232	-97.38801
13	ALL OTHER OFFENSES	Low	SCHOOL/COLLEGE	2012	5/24/12	13:20:00	W WILLIS AVE ,Denton,Tx	32.635972	-100.762
14	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	AUTO DEALERSHIP NEW/USED	2017	1/8/17	17:25:00	I20 HWY ,Denton,Tx	32.6567616	-96.74379
15	FALSE PRETENSE/SWINDLE/CONFIDENCE GAME	Low	RESIDENCE/HOME	2014	10/1/14	16:10:00	E CAMP WISDOM RD ,Denton,Tx	32.66197	-96.8994

Figure 10: Clean data set after EDA

Once the cleaner data set started taking shape, it was evident that this data set is a very good candidate for location-based studies as well as time series analysis. This realization paved way for refining the initial question of "How safe is Denton" to "How Denton crime rates changed over years?" and "Is there a prime time for the crimes?". Also "Which are the safest and riskier parts of the county" was another question that was identified at this stage.

E. *Frequency Analysis*

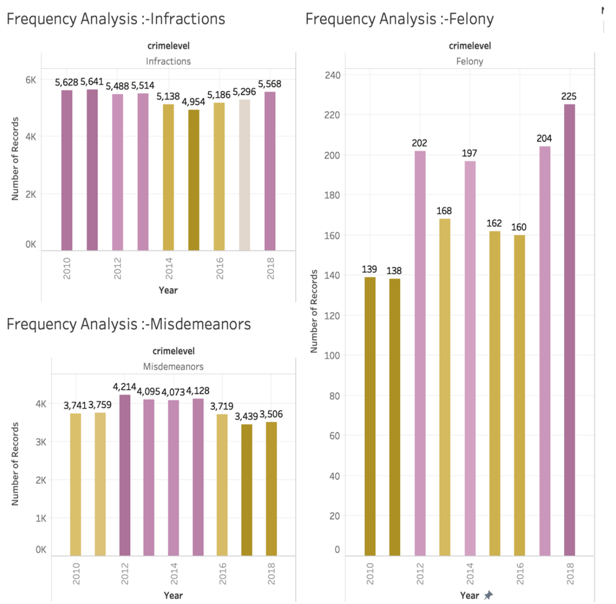

Figure 11Crime across years segregated on crime level

The data frequency analysis was done to assess the distribution of data across each year as shown in Fig. 11. This exploratory data analysis of frequency of crimes across years indicated a dip in crime rates for year 2015 except for misdemeanors.

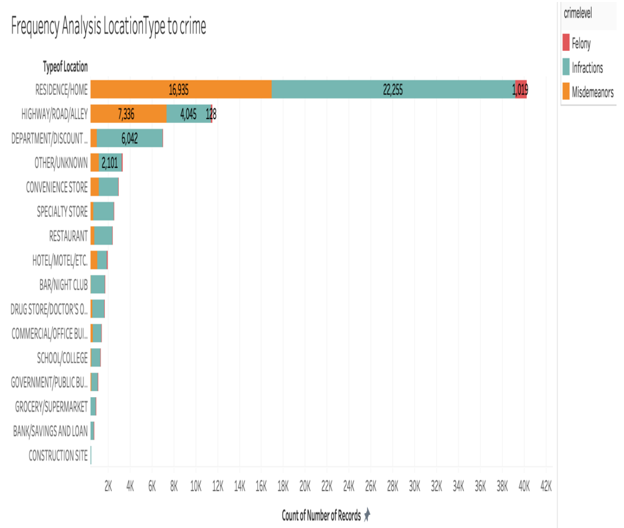

Figure 12 LocationType Crime concentration

Frequency analysis of location type revealed an interesting fact as shown in Fig.12 that all locations has count of felonies less than count of misdemeanors less than count of infractions except for Highway/road/alleys.

According to the reference model for visualization given in Figure 1, after cleaning and making necessary transformation to initial data set now, we have data tables. It is now time to perform visual mapping to produce visual structures. This process helps to determine the nature of data and map it to the best visual structure for that data.

Visual Mapping

These questions either rely on analysis of data over time or analysis of data in regard to location. Hence, data was mapped to time series using year and time for analysis on time and to geographic location for location-based studies.

View Transformation

In view transformation, the visual structures are manipulated to best describe the data by zoning, zooming, using color, labels, indexes and many other properties of visualization.

X. Denton Crime data Analysis

A. Time series Analysis

Different crime levels were plotted against year and the graph in Fig.13 was obtained. This graph gives the change in number of crimes over years. From this time series analysis, we can infer that less serious crimes(infractions) and serious crimes(misdemeanors) showed a decrease over time but felonies are on rise.

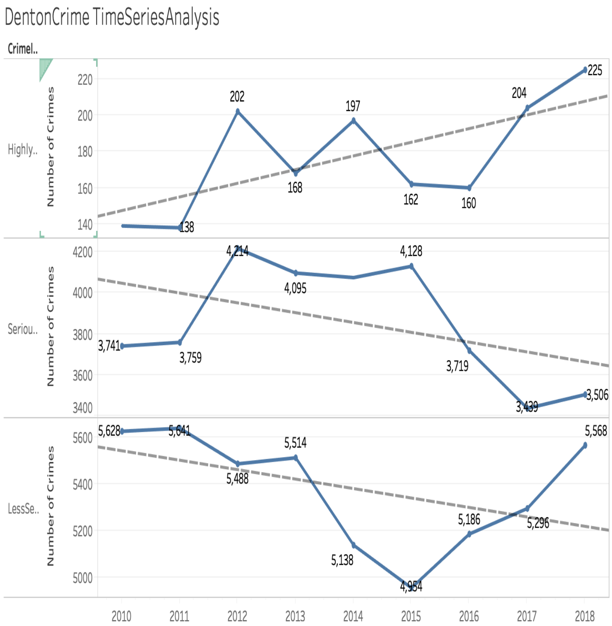

Figure 13 How number of crimes changes over years

The prime time for crimes were analyzed by plotting the crimes at different crime levels were plotted against hour of the day as depicted in the graph in Fig 14.

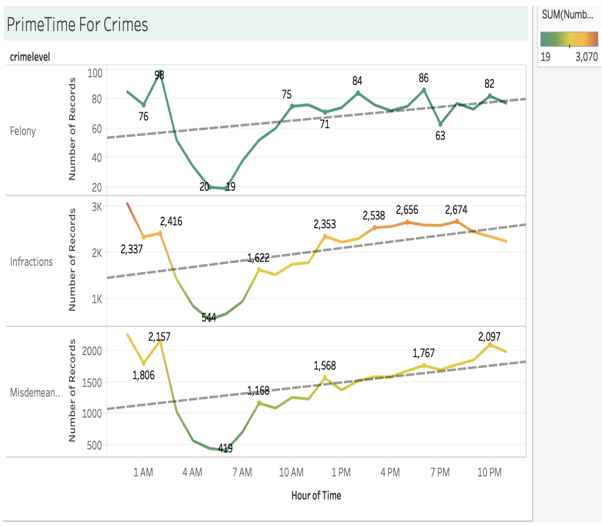

Figure 14 : Crime count at different time of the day

From the graph, we can see that highest number of crimes takes place at around 2am in the morning and then there is a comparatively safe time till 6am. From there the crime count increases throughout the day.

Year	Chance of being victim	Chance of being victim to felony	Chance of being victim to Misdemeanor	Chance of being victim to Infractions
2018	14.65404882	605.6355556	38.867085	24.47341954
2017	15.24421076	667.9803922	39.62430939	25.73036254
2016	14.82460011	839.90625	36.13471363	25.91303509
2015	14.2012116	810.345679	31.80135659	26.49899071
2014	13.64827806	651.7918782	31.52541124	24.99085247
2013	12.80392759	745.1428571	30.56996337	22.70293798
2012	12.4858643	612.1782178	29.34504034	22.53279883
2011	12.602013	871	31.97605746	21.30792413
2010	12.31089609	842.1007194	31.28896017	20.7981521

Figure 15: Chance of victimization to be read as 1 in N

Going back to Fig. 13, there is a decline in count of crimes around 2015. To understand the reason behind it, population change across years was also taken into consideration and the table in fig. 15 was obtained.

Year	Population	Crime count total	Felonies	Misdemeanors	Infractions
2018	136,268	9299	225	3506	5568
2017	136,268	8939	204	3439	5296
2016	134,385	9065	160	3719	5186
2015	131,276	9244	162	4128	4954
2014	128,403	9408	197	4073	5138
2013	125,184	9777	168	4095	5514
2012	123,660	9904	202	4214	5488
2011	120,198	9538	138	3759	5641
2010	117,052	9508	139	3741	5628

Figure 16 Population to crime

From Fig.16, we can see a flat region of where the population of Denton was in a phase of limited growth from 2014 till 2016.

The reason for declining crime rates can be attributed to this decline in population growth. This is an area which needs further studies.

When we factor in population to the crime, a person's chances of becoming a victim in Denton county can be assessed as well by using the function 'calculated fields' in Tableau. From the table in Fig.17, it is clear that you have 1 in 14 chance of becoming a victim to crime, 1 in 605 chance of becoming a victim to felony, 1 in 38 chance of becoming a victim to misdemeanor and 1 in 24 chance of becoming a victim to infractions in Denton county according to 2018 data.

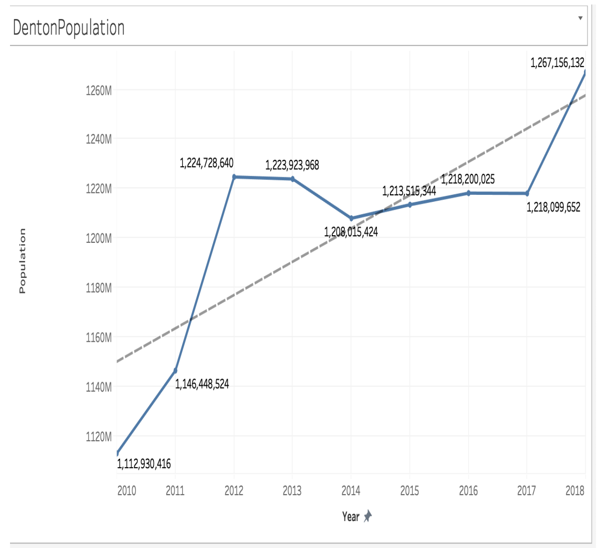

Figure 17 Denton Population growth

There is a decline in victimization chance over years. To analyze if the crime rate also has a same trend, further analysis was done. Crime rate is calculated by dividing the number of reported **crimes** by the total population; the result is multiplied by 100,000. This crime rate is plotted against year as shown in Fig. 18.

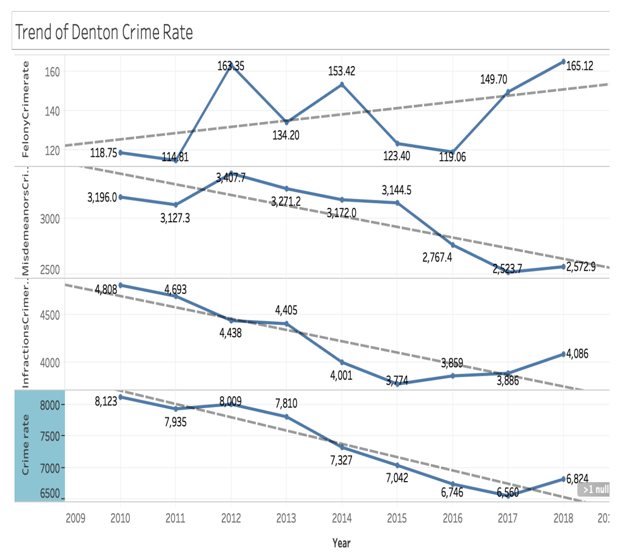

Figure 18:Crime Rate over time

This graph shows a decreasing trend in all crimes except felonies which is on the rise. This is a concerning finding and needs closer look by the law enforcement agencies. To analyze the statistical significance of these trends Tableau was used to describe the trend lines.

1) Trend model for overall Crime rate

A linear trend model is computed for Crimerate.
The models may be significant at p <= 0.05.
P-value:< 0.0001
Equation:
Crimerate = -0.210282*Year1 + 430.883

Model formula:	**(Year1+ intercept)**
Number of modeled observations:	9
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	7
SSE (sum squared error):	0.289061
MSE (mean squared error):	0.0412944
R-Squared:	0.901753
Standard error:	0.20321
p-value (significance):	< 0.0001

Figure 19 Model Summary for total crime rate trend

A time series analysis was done using tableau for years 2010 to 2018. It uses exponential smoothing to predict or forecast dependent variable crime rate using independent variable year.

In this analysis alpha was selected to be 0.05 and model had a p-value less than 0.0001 making the model statistically significant. The model explained 90% of variance in crime rate over year which is a statistically strong model and the model summary is given in Fig.19.

By using the model for year 2019 the crime rate can be computed as shown in the subsequent section.
Crime rate for 2019 = -0.210282*Year1 + 430.883
= -0.210282*2019 + 430.883 = 6.323642

The crime rates forecasted for year 2019 is 6323 which is less than previous years.

2) Trend model for Misdemeanors rate

A linear trend model is computed for MisdemeanorsCrimerate.
The models may be significant at p <= 0.05.
P-value:0.0065607
Equation:
MisdemeanorsCrimerate = -0.0951795*Year1 + 194.712

Model formula:	**(Year1 + intercept)**
Number of modeled observations:	9
Number of filtered observations:	0
Model degrees of freedom:	2

Residual degrees of freedom (DF):	7
SSE (sum squared error):	0.261036
MSE (mean squared error):	0.0372909
R-Squared:	0.675564
Standard error:	0.193108
p-value (significance):	0.0065607

Figure 20 Model Summary for Misdemeanor rate trend

In this analysis alpha was selected to be 0.05 and model had a p-value less than 0.007 making the model statistically significant. The model explained 67.5% of variance in crime rate over year which is a statistically strong model and the model summary is given in Fig. 20.

By using the model for year 2019 the misdemeanors rate can be computed as shown in the subsequent section.
Misdemeanors rate for 2019 = -0.0951795*2019 + 194.712 = -192.1674105 + 194.712 = 2.5445895

The misdemeanors rates forecasted for year 2019 is 2544 which is less than previous years.

3) Trend model for Infractions rate

A linear trend model is computed for sum of InfractionsCrimerate.
The models may be significant at p <= 0.05.
P-value: 0.0035779
Equation:
InfractionsCrimerate = -0.118282*Year1 + 242.437

Model formula:	**(Year1 + intercept)**
Number of modeled observations:	9
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	7
SSE (sum squared error):	0.318178
MSE (mean squared error):	0.045454
R-Squared:	0.725144
Standard error:	0.2132
p-value (significance):	0.0035779

Figure 21 Model Summary for Infraction rate trend

In this analysis alpha was selected to be 0.05 and model had a p-value less than 0.004 making the model statistically significant. The model explained 72.5% of variance in infractions rate over year which is a statistically strong model and the model summary is given in Fig. 21.

By using the model for year 2019 the infractions rate can be computed as shown in the subsequent section.
Infractions rate for 2019 = -0.118282*Year1 + 242.437 = -238.811358 + 242.437= 3.625642

The misdemeanors rates forecasted for year 2019 is 3625 which is less than previous years.

4) Trend model for Felony rate

A linear trend model is computed for Felony Crime rate.
The models may be significant at p <= 0.05.
P-value:0.8
The model was found to be not statistically significant as p-value was more than 0.05.

B. Location Based Analysis

Tableau was used to map crime locations on the Denton county map which gave a clear knowledge about the location of each type of crime. When highly serious crimes(felonies) were concentrated in the downtown area covered by zip code 75201 other type of crimes were spread across the county. The map in Fig. 22 was used to see which region of Denton county has risk of witnessing a greater number of crimes, it plots the location of crimes and color code it into three different categories.

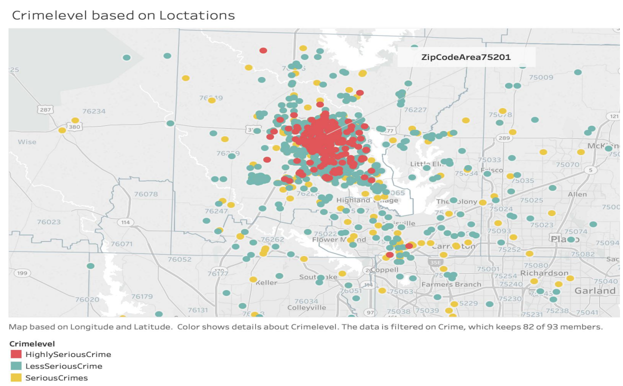

Figure 22:Prime Location of crimes

To drill down deep and identify locations/street that are at high risk, a street level view of map was plotted, and count of crimes were plotted using size of points as in Fig.23. This gave a detailed idea about locations that are prone to crimes.

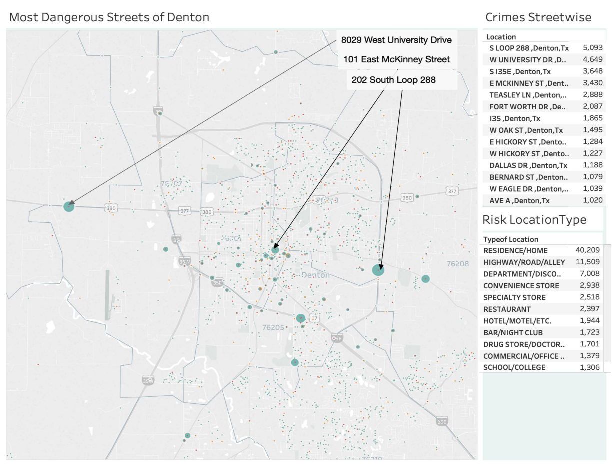

Figure 23:Street level view of crime locations

Three locations were identified to be at high risk and they are labeled in the map. Also, if you look closely at the Crimes streetwise table S 135E is the third, in regard to number of crimes but is not showing up on the map. This is because this is a long road with crime locations scattered all along it. A close inspection of Risk location type table reveal highways/roads/alleys stands second in number of crimes, first being residences.

But since the number of felonies are less compared to misdemeanors and infractions a more drilled down visualization was needed to identify streetwise distribution of crimes at different levels. The following visualization helped in inferring details.

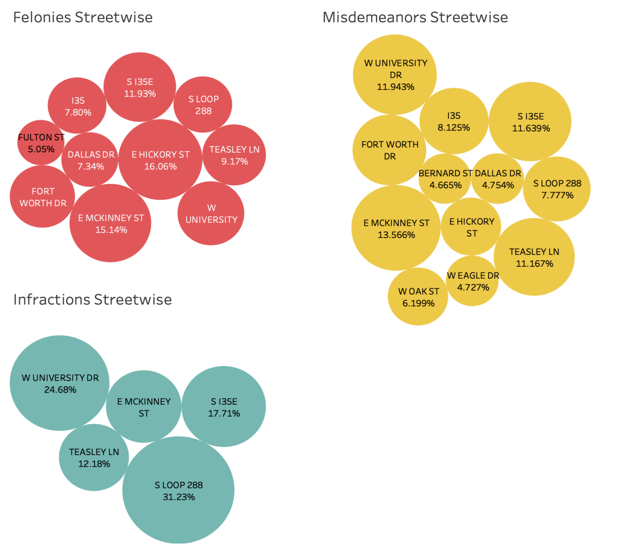

Figure 24:High Risk streets categorized into crime levels

The street identified to be riskier from previous visualization emerged as riskier in this one also. But an interesting development in this visualization is that, in felonies street wise the E Hickory St emerged as the street where the highest number of felonies takes place which was not shown in any other study.

XI. Conclusion

To sum up, this paper is consistent with many of its predecessors on data visualization research in that "a picture is worth a thousand words" - often more - but only when the story is best told graphically rather than verbally, and the picture is well designed [10]. I would like to direct readers attention to an interesting exert from an article. “Visualization is a way of telling stories and finding patterns in the data; it existed before computers, before technology, and even before language. The prehistoric humans would look up at the sky and connect the stars by invisible lines to imagine various constellations. Instead of the names of the stars, it was easier to remember which constellations they belonged to, the dots make more sense when they are connected. Among countless success stories of visualization, we discuss one [31]” Same is true regarding this paper, among countless anecdotes on data visualization, mine is a humble effort to shed some light on less addressed and discussed areas of the scientific data visualization. I dared to venture on this voyage in an attempt to find possible knowledge gaps. In course of the literature review, I was able to see how efficient it will be to use data visualization, not as a support tool, but as the primary data analysis tool but how rarely visualization is used. The academic and research community restrict the use of visualization to assumption verification and presenting results. But for some research questions, visualization is the only tool that can lead to implementable insights. Such questions are answered through mostly geospatial data analysis and time-based data visualization. There is no better tool than data visualization for this purpose. In this fact lies how this study deviates from and defies the previous studies or sets a new path for research. An example of such a problem that can prove this is, a study on bringing down the crime rate and predicting possible crime locations from crime data analysis which was done in the second part of this paper.

An interesting fact emerged as a result of this research is that, visualization can be effectively administered, using tools available to every layman, who lack access and technical know-how in using sophisticated data analysis tools, techniques and technology, who do not understand linear regression and Manova, who think statistics is above them. This paper was also helpful in establishing less researched domain of county level crime data analysis. Every crime data analysis available till date, pertains to organized crimes and terrorist activities which is not very helpful to local law enforcement agencies in bringing down the crime rate locally. Here lies the importance of a study of the nature as the one conducted in this paper. It can help in identifying locally active crime spots and timings which can lead to enforcing stricter policing by the law enforcement agencies. This task was done most effectively and efficiently using visualization.

A point to keep in mind is the ability of visualization to bend and even break facts. As easy as it is to present data to amplify cognition using visualization, it is natural or easier to project a lie as truth using visualization [32]. An intuitive eye and a suspicious mind are all one need to break this mirage of lie presented through visualization. This fact should not impede one from the importance of visualization. With the hype, exaggeration and over expectation created by the big data one should approach any techniques dealing with it with caution especially data visualization due to its power to instill decision making.

A possible future research of this study is the leading indicator study of crime data. This study focuses on predicting the crime location and time based on leading indicators. This is a time series analysis that study the indicators that possibly lead to a crime. It is time lagged study which observes less serious crimes to predict felonies. For example, a research of this nature studies if an increase in infractions for a time period will lead to increase in felonies in the subsequent time period. As stated earlier, among countless applications of visualization, we discuss one and there are many left unexplored.

XII. Limitations of this Study

This study is strictly based on available crime reports and cannot be implemented on regions that lack historic crime data. Missing and insufficient data can lead to wrong insights and this fact needs to be accommodated in future studies using this method.

ACKNOWLEDGMENT

I thank Fariba Sadeghinaeenifard for her guidance and timely help in completing this literature review. I thank my fellow course mates who have engaged in healthy debates and conversation leading to more profound insights on the topic.

XIII. BIBLIOGRAPHY

[1] Webtser's New World Dictionary, Collins, 1970.

[2] B. H. McCormick, T. A. DeFanti and M. D. Brown, Visualization in Scientific Computing Computer Graphics Vol. 21, No. 6,, November 1987.

[3] K. W. Brodlie, L. A. Carpenter and R. A. Earnshaw, Scientific Visualization Techniques and Applications, Verlag: Springer, 1992.

[4] H. Eisenberg, "Humans Process Visual Data Better," 15 September 2014. [Online]. Available: http://www.t-sciences.com/news/humans-process-visual-data-better. [Accessed 17 April 2019].

[5] H. Senay and E. Ignatius, "Rules and Principles of Scientific Data Visualization," Department of Electrical Engineering and Computer Science The George Washington University, Washington, D. C., February 11, 1999.

[6] A. I. Globus, "Principles of Information Display for Visualization Practitioners," NASA Ames Research Center, 28 November 1994.

[7] H. C. Jennifer Xu, "Criminal Network Analysis and Visualization," *COMMUNICATIONS OF THE ACM,* vol. 48, no. 6, June 2005.

[8] H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," 10.1371/journal.pone.0176244, Online, 2017 Apr 24.

[9] M. Card, Readings in Information Visualization: Using vision to think, Morgan Kaufmann, 1999, pp. 1-34.

[10] S. Few, "Data Visualization for Human Perception," in *The Encyclopedia of Human-Computer Interaction*, The Interaction Design Foundation, 2015-11-23.

[11] R. M. Shiffrin and R. A. Atkinson, Stevens' handbook of experimental psychology, 1988.

[12] M. Friendly, A Brief History of Data Visualization, Springer-Verlag, 19 November 2017.

[13] M. J. Eppler, "A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing," in *Information visualization* , 2006, pp. 202-210.

[14] G. Domic, "Tutorial on Visualization," [Online]. Available: www.cs.rit.edu/usr/local/pub/ncs/hypervis/concepts/concepts.htm.

[15] G. G. Robertson, S. K. Card and D. M. Jock, Information Visualization Using 3D Interactive Animation, CACM, Vol. 36, No. 4, April 1993.

[16] J. Foley and W. Ribarsky, Next-Generation Data Visualization Tools. In Scientific Visualization, Academic Press Ltd, 1994.

[17] R. B. Haber and D. A. McNabb, "Visualization Idioms: A Conceptual Model for Scientific Visualization Systems," *Visualization in Scientific Computing.*

[18] S. Lewis and C. Wehrend, "A problem-oriented classification of visualization techniques," *In Proceedings IEEE Visualization,* no. 90, pp. 139-14.

[19] P. K. Robertson, "A Methodology for Choosing Data Representations," *IEEE Computer Graphics and Applications, No. 3,,* Vols. Vol. 11,, pp. 56-68, 1991.

[20] H. Senay and E. Ignatius, "A Knowledge-Based System for Visualization Design," *IEEE Computer Graphics and Applications,,* pp. 36-47, November 1994.

[21] S. Roth and J. Mattis, "Data characteristization for intelligent graphics presentation," *In Proceedings CHI,ACM Press,* no. 90, pp. 193-200., April, 1900.

[22] W. Hibbard, C. R. Dyer and B. Paul, "Display of Scientific Data Structures for Algorithm Visualization," *IEEE Proceedings in Visualization,* no. 92, pp. 139-146.

[23] M. Friendly, "Milestones in the history of data visualization: A case study in statistical historiography," in *Classification—the Ubiquitous Challenge*, Berlin, Heidelberg, Springer, 2005, pp. 34-52.

[24] P. Dragicevic and Y. Jansen, List of Physical Visualizations and Related Artefacts, 2012.

[25] M. Friendly, Milestones in the history of thematic cartography, statistical graphics, and data visualization., 7 July 2008.

[26] B. E. Rogowitz and L. A. Treinish, "An Architecture for Perceptual Rule-Based Visualization Conference edited by Gregory M. Nielson and Dan Bergeron," *Proceedings of the IEEE Computer Society Visualization,* no. 93, 1993.

[27] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian and K. Hornbaek, "Opportunities and challenges for data physicalization.," *Proceedings of the 33rd Annual ACM Conference,* 2015.

[28] E. R. Tufte, The Visual Display of Quantitative Information, Cheshire, Conn: Graphic Press, 1983.

[29] R. J. Andrews, "Interacting with DataViz History," [Online]. Available: www.infowetrust.com.

[30] D. Huff, How to Lie with Statistics., New York: Norton, 1954.

[31] H.Bhatia, "Enabling Discovery through Visual Exploration," *U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract .*

[32] B. E. Rogowitz and L. A. Treinish, How not to Lie with Visualization.

MATERIALS REFERED

Owen,GS."Definitions and Rationale for Visualization." Retreived https://www.siggraph.org//education/materials/HyperVis/visgoals/visgoal2.htm.

"Visual Mapping Best Practices for Use." Retrieved from http://www.ctl.gatech.edu/node/582113

Data Visualization retrieved from https://en.wikipedia.org/wiki/Data_visualization

Horne, J. "The History of the History of Data Visualization" Retrieved from https://www.idashboards.com/blog/2017/11/29/the-history-of-the-history-of-data-visualization/