

# Narrative Visualization Project Essay

## CS 416 Data Visualization, Summer 2023

David Burrus ([dburrus3@illinois.edu](mailto:dburrus3@illinois.edu))

### Introduction

My visualization focuses on criminal activity in the city of Chicago during the last seven years. It may be viewed at <https://dburrus3.github.io/>.

The data set was exported from the “Crimes – 2001 to Present” database available through the City of Chicago’s data portal at <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2>. Each record in the dataset corresponds to a reported crime, and includes:

- **date and time** of occurrence
- **geographic location** of crime, described in terms of street address, latitude and longitude, zip code, census tract, aldermanic ward, and named community
- **location type**, corresponding to one of 192 categories such as “street”, “pawn shop”, “Chicago Transit Authority train”, “parking lot”,
- **police district** and beat responsible or assigned
- **primary type of crime**, corresponding to one of 26 FBI codes such as “Homicide 1st & 2nd Degree”, “Forgery & Counterfeiting”, and “Disorderly Conduct”
- **detailed type of crime**, corresponding to one of 339 IUCR (Illinois Uniform Crime Reporting) codes such as “Second Degree Murder”, “Purse-Snatching”, “Insurance Fraud”, and “Impersonation”

The data portal yielded 1.73 million records for the period of interest – January 1, 2016 to December 31, 2022.

To focus and simplify this visualization, the data were preprocessed as follows:

- **date and time** data were summarized by year and calendar month, day of week (Monday through Sunday), and hour of day (from 00:00 to 23:00)
- **geographic location** data were summarized into Chicago’s nine recognized regions, such as “North Side”, “West Side”, and similar. This was done by mapping each named community to its region.
- **location type** data were first analyzed for significance, then summarized into nine groups, such as “business” (which includes 36 categories of business), “vehicle” (which includes private and commercial cars and trucks, taxicabs, boats), and “Other” (which includes 54 categories such as schools and colleges, airport buildings, public housing, public transit, highways, lakefront, and vacant lots).
- **type of crime** data were summarized into four fundamental categories: Violent (which includes homicide, assault, battery, sexual abuse, and robbery), Property (which includes burglary, larceny, vehicle theft, vandalism, and arson), Financial (which includes fraud, embezzlement, and forgery), and Statutory (which includes prostitution, drug abuse, gambling, liquor license issues, and similar activities).

For this visualization, the data were distilled into approximately 20,000 records. The file containing those records may be viewed at <https://github.com/dburrus3/dburrus3.github.io/blob/main/data/LocationTime.csv>. An additional distillation, trends.csv, was prepared for Scene 2 and is in the same directory.

A separate data set was obtained to compare Chicago with the 100 most populous U.S. cities. These data include population (2019 estimated and 2020 census) and per capita crime rates based on FBI statistics for 2019. Source data may be found at [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_crime\\_rate](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_crime_rate) and [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_population](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population). These data were distilled into two CSV files, CityPopul.csv and CityCrimes.csv, available in the same directory as LocationTime.csv.

## Messaging

The primary message of this visualization is that crime in Chicago is far better than stories in national and regional media imply. Relative to the 100 largest cities in America, Chicago's crime rate is merely average. Furthermore, Chicago's crime rate dropped significantly since that comparison was made, primarily due to COVID, and for the most part has remained well below pre-COVID levels.

The second message of this visualization is that crime within Chicago is concentrated in certain areas. A significant portion of Chicago residents experience even lower crime levels than the citywide per capita statistic.

The intended audience for this visualization is anyone who lives in Chicago, commutes into or visits Chicago, or is considering relocating to the Chicago area.

## Narrative structure

My visualization follows the "martini glass" structure. It comprises three scenes, each of which provides the user with more opportunity for exploration than its predecessor.

Scene 1, "A tale of two cities", presents two charts that establish the primary message. The only interaction in this scene is allowing the user to see the underlying data through tooltips.

Scene 2, "Recent Trends", presents a single chart reinforcing the primary message. It encourages the user to explore trends in different types of crime. Annotations enhance the user's understanding of the data he or she chooses to explore.

Scene 3, "Where and When", presents a single chart and encourages the user to interact with data by changing the X-axis, Y-axis, and data filter. This scene also provides a map for the user's reference. (Linking this map to the user's choice of X-axis, Y-axis, and filter could be a worthwhile future enhancement to this visualization, but doing so is well beyond my current skills.)

Additional detail is in the Scenes section, below.

## Visual structure

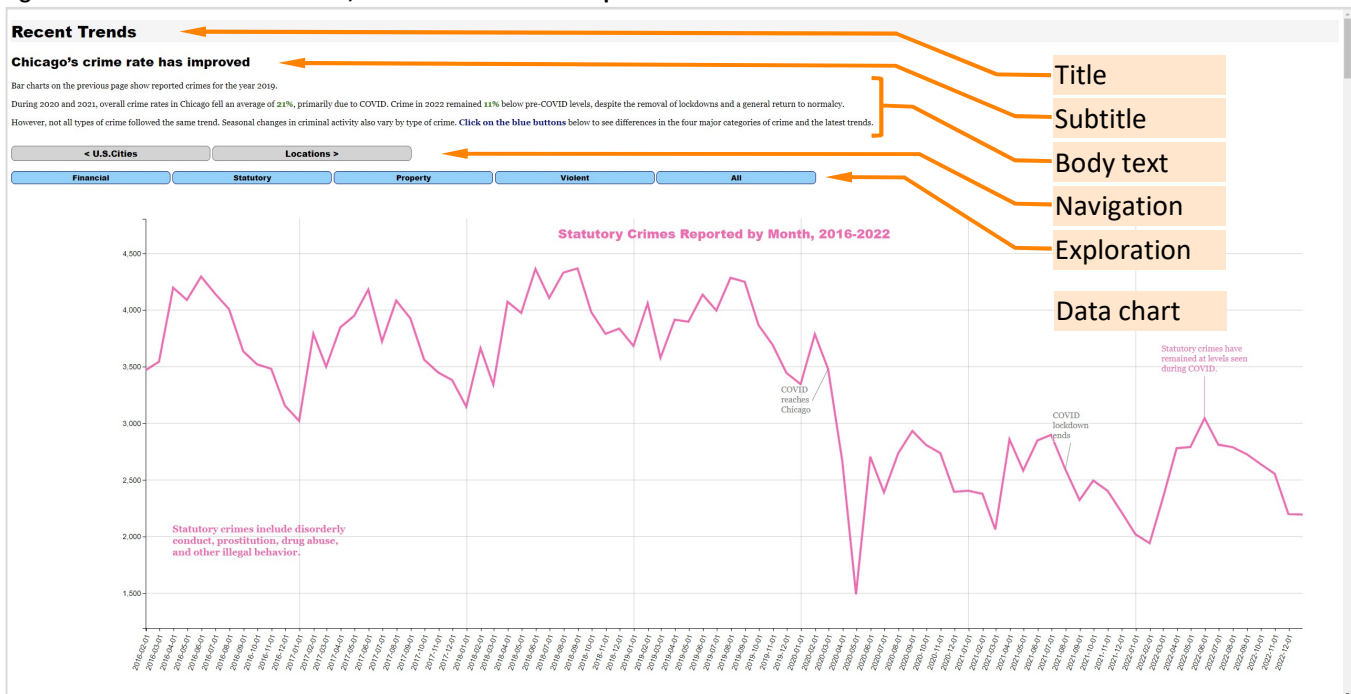
All three scenes follow a consistent visual structure, exemplified in Figure 1 below. From top to bottom:

- **Title** – a headline to orient the user within the three-scene visualization. Scene 1's title, "A Tale of Two Cities", invites the user to contemplate several meanings of Dickens' familiar title in this context, such as:
  - Chicago's crime level is both better and worse than one might expect.
  - Comparisons of two cities, such as Chicago vs. Memphis or Chicago vs. Indianapolis, may be surprising.
  - Foreshadowed, but unstated in Scene 1, is the disparity between lower-crime (generally wealthier) regions of Chicago and higher-crime (generally poorer) regions.
- **Subtitle** – summarizes the primary message or takeaway of the scene, and is intended to be self-explanatory.
- **Body** – summarizes, in only 100 to 200 words, (a) what the user should look for within the scene and (b) what actions the user can perform within the scene. Instructions for possible user actions (e.g., clicking, hovering) are bolded to attract the user's attention.
- **Navigation buttons** – show which scenes a user can reach from the current scene. As the narrative structure is "martini glass", these buttons are deliberately restricted to the equivalent of "next page" and "previous page". The buttons are labeled with "<" and ">" to suggest that. Rather than "< Prev" and "Next >", however, each button's text is a headline for its destination. The user thus is spared having to recall where she has been, and can anticipate what she will see next. The color and style of the navigation buttons is intentionally consistent across all scenes.

- **Exploration buttons** or dropdowns – allow the user to explore the data by altering the scene’s chart. The color and style of these buttons also is consistent across all scenes, but visually distinct from the navigation buttons. Scene 1 has no exploration buttons, because it is the beginning of the martini glass.
- **Data** – visualizes the actual data pertinent to each scene. Different colors are employed to emphasize distinctions in the data being charted. Scene 1, for example, shows two rankings of cities, one by population and one by crimes per capita. Consistency and comparability are maintained by showing both data sets as horizontal bar charts. Different color schemes make immediately visible that the two charts differ in some significant way – which the chart title and annotations then explains.

Contrasting colors and bold text are used in the body and data sections to draw the user’s attention to important elements of the scene.

Figure 1. Visual structure of scenes, with Scene 2 as an example



## Scenes

As mentioned, the primary message of this visualization is that crime in Chicago is far better than stories in national and regional media imply.

**Scene 1, “A Tale of Two Cities”**, provides an initial frame of reference by comparing Chicago to other American cities. Because people absorb information differently, this point is made both in the body text and graphically. Skeptical users are invited to hover over the charts and see the numbers for themselves. Highlighting (and annotating) Chicago’s position within each chart makes the difference immediately visible, even to users who are not “into the details”. Clicking on the “Trends >” navigation button takes the user to Scene 2.

**Scene 2, “Recent Trends”**, builds on Scene 1 by showing visually how the crime in Chicago has dropped since the data in Scene 1 were collected. The primary message is emphasized by Scene 2’s subtitle, “Chicago’s crime rate has improved”.

Scene 2 also introduces the idea (without saying it) that “average crime rate” is an incomplete and possibly misleading measure, by showing how year-over-year trends as well as seasonality in criminal activity. Following the martini-glass structure, most of Scene 2’s content is author-driven. But, starting in Scene 2, the user may

follow his or her interests by exploring the different patterns exhibited by different types of crime. Those most concerned about online fraud and scam artists will see that such crime occurs at a relatively low level that is not seasonal. Those most concerned about being a victim of robbery or assault can click on the “Violent” exploration button to see that category of criminal activity, and may be comforted to see that it remains roughly 10% lower than the timeframe captured in Scene 1. Users can return to Scene 1 by clicking on the “U.S. Cities” navigation button, or advance to Scene 3 by clicking on “Locations >”. As mentioned, navigation buttons are labeled to help users anticipate what they will see next.

**Scene 3**, “Where and When”, keeps the dimension of time introduced in Scene 2, and introduces the dimension of geographic location to further the user’s understanding (see Figure 2). As in the previous scenes, the subtitle “Risk of becoming a victim varies by time and location” introduces the main takeaway from this scene. Unlike the previous scenes, the user is afforded an opportunity to freely explore Chicago’s crime activity, as well as a third dimension, type of crime, that also was present in Scene 2. To orient users, a map is provided that shows Chicago’s location within Illinois, the location of regions (e.g., North Side, Far Southeast Side) within Chicago, and how each region is divided into a total of 77 named communities.

Scene 3’s interactive chart is a heat map with location on the Y-axis and time on the X-axis. The user may change freely each axis, as well as filter by type of crime, using the exploration buttons. This is explained in the brief body text. The exploration buttons follow the same style and color scheme as the exploration buttons in Scene 2, so that the user may easily recognize that they serve a similar function. In Scene 2, however, each exploration button is a drop-down menu of options.

The sole navigation button, “< Trends”, indicates that the user has reached the final scene.

## Annotations

Annotations (as opposed to tooltips) appear in Scene 2. Three different templates are employed across the five charts in Scene 2 to convey three different kinds of information.

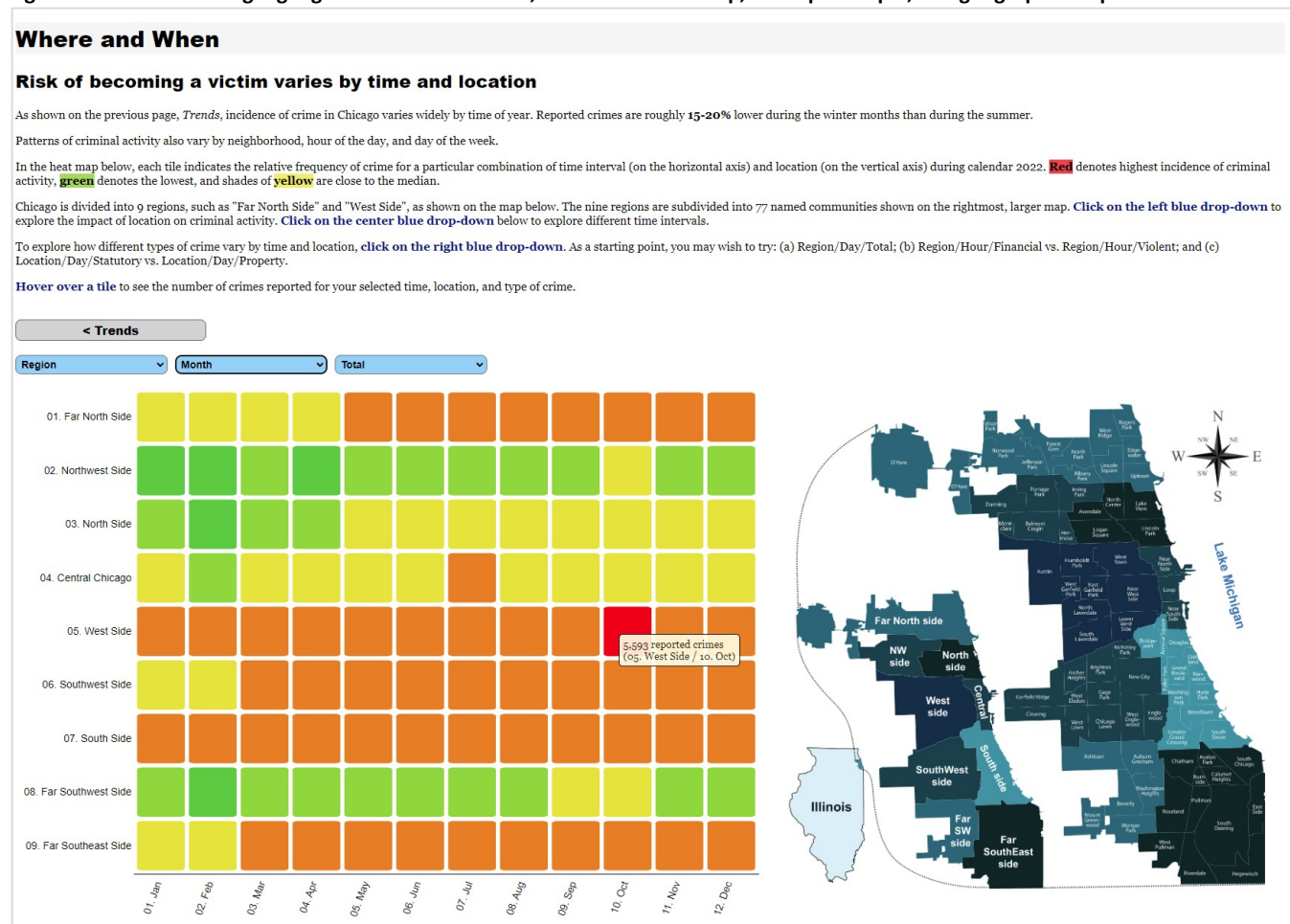
1. **Gray text with drop line**, e.g., “COVID reaches Chicago” and “COVID lockdown end” – These annotations are the same across all five types of crime. Once the user has seen these annotation for one type of crime, it’s not necessary to emphasize them when the user looks at other types of crime. These annotations always appear in the same months, although they change vertical location to match the charted data.
2. **Color text with drop line**, e.g., “Statutory crimes have remained at levels seen during COVID” – These annotations change location and wording according to the type of crime. This is emphasized by matching the text color to the line color, which changes by type of crime.
3. **Bold color text with no drop line** – These annotations provide a brief explanation of each type of crime. This is emphasized by again matching the text color to the line color. These annotations (one per type of crime) are not tied to particular date, so they appear without a call-out line. Because these explanations help understand the chart and English is read left-to-right, these annotations are positioned horizontally near the left edge of each chart. Their vertical location is set to avoid overlapping the charted line and to avoid competing visually with the chart title.

Annotation types 2 and 3, because they are specific to type of crime, change according to the user’s selection. For the same reason, annotation types 2 and 3 do not appear in the chart for “All Crimes”. All available annotations appear immediately when each chart is presented.

Annotations in Scene 1 follow the type 2 (“call-out in color”) template described above. In scene 1, the text color matches the bar to which each annotation points.

Scene 3 does not use annotations, because the heat map and Chicago map already are visually dense. Tooltips are used instead, to allow users to see the underlying data. Scene 1 also uses tooltips, for the same purpose. Tooltips throughout this visualization use the same template.

**Figure 2. Scene 3 showing highlighted user instructions, interactive heat map, tool tip example, and geographic map**



## Parameters

Both Scene 2 and Scene 3 have a total of four parameters.

In **Scene 2**, each of the five blue exploration buttons (e.g., **Financial**) causes an integer parameter to be passed to the drawChart function, as shown in the code excerpt below. This parameter corresponds to type of crime. The integer parameter triggers changes in the data that are displayed, as well as the color and chart title (by using the parameter as an array index).

**Figure 3. Scene 2 code excerpt showing parameter origination, passing, and effect on line chart**

```

<button class="dataSelect" onClick="clearChart(); drawChart(2);">Financial</button>
:
function drawChart(crimeType) {
  var yValues = trends[0, crimeType];
  :
  var path = svg.append("path")
    .attr("transform", "translate(" + chartOriginX + "," + chartOriginY + ")")
    .attr("d", line(xValues, yValues))
  :
}

```

In **Scene 3**, each of the three blue exploration drop-downs (reproduced at right) causes a text parameter to be passed to the `getData` function, which then generates a subset of data from the combination of those three parameters. The data subset is passed to the `drawChart` function, triggering changes in the chart's X-axis (time interval), Y-axis (location), or data filter (type of crime), depending on which drop-down option was selected by the user. Here is a code excerpt for the time interval drop-down:



**Figure 4. Scene 3 code excerpt showing parameter origination, passing to `getData`, and effect on `drawChart`**

```
var selectTime = d3.select("#chartButtons")
  .append("select")
  .attr("class", "dropdown")
  .on("change", function(d) {
    time = this.value;
    chartData = getData(locTimeRaw, time, place, type);
    drawChart(chartData);
  })
```

**Scene 1** has no parameters, as it is the beginning of the martini glass and its content is entirely author-driven.

## Triggers

Each of the three scenes has one or more triggers that connect a user action, such as clicking or hovering, to a visible change in the scene. Affordances (hints to inform the user how to use the interface) always appear in body text near the top of each scene, and are highlighted with bolding and color.

**Scene 1** includes pointer location triggers for `mouseover`, `mousemove`, and `mouseout`. These triggers cause a tooltip to appear when the user hovers over a horizontal bar in either chart, and to disappear when the user moves the pointer away from a bar.

**Scene 2** includes triggers linked to the five exploration buttons. Clicking one of those buttons calls function `clearChart()` to erase the chart currently displayed, then calls function `drawChart()` with a parameter corresponding to the button that was clicked. Figure 3 shows some of the related code.

**Scene 3** includes both location triggers like Scene 1 and exploration button triggers like Scene 2. The location triggers cause the appearance and disappearance of tooltips that show underlying data. The button triggers causes the heat map to be redrawn with new parameters.

## Future enhancements

This project is the first time I have worked with any aspect of HTML or JavaScript, which limited how far I could take the assignment within the allotted time. Opportunities to improve this visualization include:

- Responsive maps in Scene 3 – The visualization could be even more powerful if the regional and community maps changed color in a way that corresponds with the user's exploration choices.
- Maps as an additional filter in Scene 3 – The user would be allowed to click on a region (e.g., North Side) or a named community (e.g., Lincoln Park). In response, the heat map's Y-axis would change to "Location of crime" and crime data would be filtered to show only the selected region or community. (Similar functionality might be achievable with additional drop-down buttons, but a map could more intuitive for many users. Such a feature would allow users to explore, for example, how the risk of living in a Loop apartment compares with the risk of living in a Roger Parks residence.
- X-axis in Scene 2 – For unknown reasons, the first month (2016-01) does not display on the bottom axis, as though the labels of the line chart are shifted by one. The last X-axis value (2022-12) is correct. Multiple attempts were made to cure this, without success. A deeper knowledge of JavaScript is required.

- Axis labels in Scene 3 – It is desirable that heat map X and Y values appear in a logical sequence, such as Monday to Sunday, January to December, regions from northernmost to southernmost, and communities grouped by region. To guarantee this, some value sets were given a numeric prefix, for example, 01.Mon to 07.Sun, or 01.Jan to 12.Dec. This makes the displayed values less immediately familiar to users. It would be better to achieve the desired sort order without resorting to this expedient.