

SafeCity Design Document

Eric He (ehe)
Kavin Sivakumar (ks16)
Stefan Keselj (skeselj) – Liason

March 14, 2016

1 OVERVIEW

2 REQUIREMENTS AND TARGET AUDIENCES

There is security in information. People can avoid danger and authorities can better combat it if key information is known. The police departments of most major American cities are continuously releasing vast amounts of crime data in hopes that people will use it to be better informed, but the average person does not gain insight from such raw and disorganized data dumps. SafeCity is a website which will bridge this knowledge gap by clearly and comprehensively displaying information about crimes in the US as they happen.

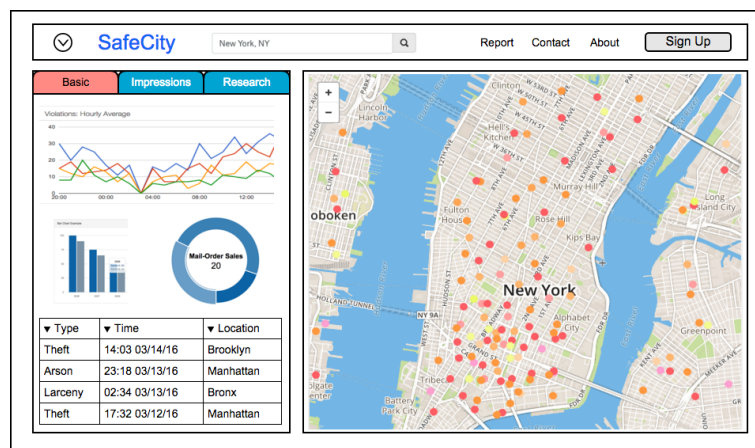
There have been multiple attempts at a tool like this, but none have managed to achieve the two attributes we think will differentiate our product: clear design and comprehensive information. CrimeMapping has sparse data, almost exclusively in Los Angeles (there are two crimes in the last month in New York). CrimeReports covers many cities, but the vast majority of crimes it reports are sexual assaults. SpotCrime has passable density over a fair number of cities, but the user can only see the crimes within a circle of a few miles in radius around a reference point, and zooming out does not display any more information.

SafeCity will be different from all these attempts in that it will display a large amount carefully curated crimes in an intuitive way according to type and time. This is a viable place for us to add value because the current resources are irresponsible about data completeness and careless with their user interfaces. Our vision will enable people with safety concerns about areas

of their city, so all people at some point or another, to get much more meaningful insights than they would on any of these other websites.

3 FUNCTIONALITY

The main focus of this application is to connect users with information about crimes happening in their cities. Each of the possible scenarios will be similar: a user will pan and select different areas on the map to the right and then interact with one of three different data representations on the control panel to the left. What differentiates these users is the type of data they will be interacting with, which comes in three forms: graphical, verbal, and quantitative.



3.1 CASUAL PASSERBY

This user represents the people who stumble upon our site after having a transient interest in crime rates while surfing the web. The user will load the front page to find a large map taking up the rightmost two thirds of the screen, displaying all the crimes that happened in Manhattan in the last year as color coded bubbles. There will be a slider on the bottom of the map which will allow the user to play around with the time frame displayed on the map (maximum last ten years, minimum last week). The user can mouse over the map to find out details of each crime: time, type, and any people involved. To the left there will be a control panel summarizing the information in the map in different ways, which we hope will catch the user's eye for a few seconds. It will contain a line graph displaying the total number of crimes over time, a pie chart displaying all the different types of crimes, and a scrollable table containing all the crimes sorted chronologically. Once the user gets bored with Manhattan, he or she will type a different place, most likely his or her hometown, up in the search bar at the top of the screen and go through the same process and then leave.

3.2 ACTIVE MEMBER

This user represents the people who have more lasting involvement with our application because they care about how the world views their community. This person will want the full benefits of being a member, so at some point they will click on the "Sign up" button on the top right and enter their username, password, full name, email address, and postal code (the last two will be verified). From then on they will either sign in through a login page or automatically sign in through their browser. This type of user has already explored the basic summaries mentioned above, and will spend more time on the second tab of the control panel: the impressions page. This page will be something like a Twitter feed, containing geo-tagged tweets involving safety or crime in an area, and containing a stack of 200 character testimonies people wrote about the safety of an area. Active members will be the primary contributors and curators of this information. They will write the testimonies about the areas in their city that they know, and they will upvote or downvote content as they feel fit to ensure that the most representative snippets are noticed. Periodically, these members will receive an email digest of what has been said about their area to invite them back.

3.3 IN-DEPTH RESEARCHER

This user represents the people who are familiar to our site and come to it for more detailed and technical forms of information than those mentioned above. This user will occasionally check out the two tabs mentioned previously and be a member like the case above, but will spend most of their time viewing and downloading data from the third section of the control panel: the research tab. In this tab, the user will be able to download plain text files containing all the data used to make our application: all the times, types, and locations of crimes in an area and all the testimonials written about an area. The hope is that this user will use our information to gather further insights into how and why crime is happening. We will offer the option to upload documents in this tab, so that If the researcher uses our data to write a paper or report it can be made available to any other researchers looking to do something similar. Additionally, we will display our own take on using our data to research crime: a k-nearest-neighbors crime prediction algorithm which estimates the likelihood that certain crimes will happen in certain areas. This data will be overlaid to the right as a heatmap.

4 DESIGN

SafeCity will be implemented as a classic three-tier system, with the data tier (a MariaDB MySQL database with bash update scripts), logic tier (querying through PHP and Python preprocessing modules), and presentation tier (Bootstrap layout, KeenIO charts, and Google Maps API) will all be mutually independent. In general, the user will engage in a small action that prompts our business logic to request crime or user information from the database, which will be processed and then displayed for the user in a responsive format.

4.1 DATA

This is a data driven application, so the data tier is arguably the most essential to its success. These days NoSQL databases like MongoDB are incredibly popular, but we have chosen to keep it simple with MariaDB relational database because we are certain our data will have a simple and uniform structure. There will be two tables: the first will store the type, time, longitude, and latitude of each crime; the second will store the name, username, password (md5 hashed), email, and zip code of each user. We will initialize the crime tables using data dumps of historical data, and then update them every day by scraping the webpages of different police and news stations with bash scripts that use wget or possibly BeautifulSoup. The users table will start out empty but every time a user signs up a row of information will be added. No values in the database will ever be deleted. We are not worried about running out of space because each entry in our databases takes only on the order of 10 bytes.

4.2 LOGIC

The role of the logic tier is to respond to requests from the presentation tier by retrieving information from the data tier and then process it before giving it to the presentation tier. A crime request will come as four floats: the minimum and maximum longitude and latitude of a rectangular area of the map. Using PHP select requests, we will query the database to find pull all crimes with longitude and latitude between the minimum and maximum (all the crimes in the area). It is not yet clear how we will make this operation quick when the amount of data we have gets very large, but for now we will likely allow the select operation to traverse all the crimes. Once these crimes are obtained, a python script will package them as a list of four-tuples (type, time, latitude, longitude) and then send it to the presentation layer. A user request will come as a username and password pair. We will again use a PHP select statement to see if there are any users with that username and password and then return a pair to the presentation layer (username, zipcode) if the user exists and nothing otherwise.

4.3 PRESENTATION

The presentation tier is where the magic happens, our quick and intuitive way of presenting information will be an important differentiator from our competition. The overall layout will be a KeenIO template for an analytics dashboard, which we chose to keep the different sections organized. KeenIO is compatible with Bootstrap (in fact it was implemented with it), which will allow us to use Bootstrap's elegant buttons, search bar, and tables. Most importantly, both Bootstrap and KeenIO are responsive frameworks, so our website can be viewed on screens of different sizes (such as phone screens, our competitors cannot do this). Finally, the driver of the application is the map, which will be a Google Maps window displaying all the crimes given to it by the logic layer. Every time the user enters a new address, pans across the map, or selects a region of it, the maps API will return the center coordinates, length, and width of the current window (we think), which will then be sent to the logic layer in four tuple form for information retrieval.

5 TIMELINE

5.1

5.1.1

6 RISKS AND OUTCOMES

6.1

6.1.1

7 ROUGH WORK

Most users that come to our site will be people who had a spurious interest in the crime data of a particular city or area while surfing the web. These types of users have limited patience and interest in the subject, they need information displayed clearly without effort on their part. This type of user will open up our homepage to find a large map taking up two thirds of the screen zoomed in Manhattan by default, but there will be a search bar on top for navigation. The map will display crimes as basic bubbles color coded by time, and to the left there will be the three simplest ways to summarize the data for the user: a line graph displaying the crimes in the area over time, a pie chart displaying the types of crimes in the area, and a scrollable table containing the most recent crimes in that area.

Some users will have more lasting involvement with the site. For example, a community member concerned about buying or selling houses or a local politician who cares about his or her riding might want to get a sense of the character of an area from the perspective of people who have been there. For this purpose, there will be a second tab in the control panel featuring written "impressions" people have of the area (we label them as impressions and not reviews or reports to make them less formal). Each impression will be limited to 200 characters to ensure people get their point across effectively and to ensure compatibility with Twitter. Tweets about the area featuring key words like "crime" or "safe", and their relatives, will be displayed in the impressions tab as well. To contribute to the impressions, a user will have by scrolling through and contributing to the impressions tab, an active member of SafeCity will be able to define an area by more than just the cold numbers.