COLORADO

Governor's Office of Information Technology

Serving people serving Colorado

Data Tigers
Crime Analytics and Mapping

Smart Data Analytics Challenge 2018

1. Introduction

Problem Statement

The current state of tools and methodologies for analyzing and mapping crime data is lagging behind the capabilities of currently available technologies. While reasonable datasets are available in many geographic areas throughout Colorado, the methodology and tools for effectively leveraging this data for analyses and geographic mapping is lacking.

Our team located several online crime analysis and mapping tools (from DenverGov.org, the Denver Post, Spot Crime, Trulia, etc.). However, these tools either had limited analysis functionality or failed to perform properly even when tested on multiple browsers (we tested them on Chrome, Firefox and I.E). The following site failed on all 3 browsers: denvergov.org/content/denvergov/en/police-department/crime-information/crime-map.html

Problem Solved

Our team analyzed the past 5 years of crime data to understand the trends from the different aspects of crimes, such as year-over-year growth trends and crime rate by neighborhood. In addition, to further understand the underlying drivers of the different categories of crime, we conducted detailed statistical analyses to identify the association between the crime rates and the Census demographic data.

We then developed an easy to use mapping interface to provide instant visual data, analytical comparisons between target neighborhoods, and an online citizen-centric crime reporting crowdsourcing interface. We have the ability to view crime counts in the last five years or any user-selectable date range, and to visually pan/zoom in on specific geographic areas. The system is also extensible and can incorporate real-time data feeds by city, state or IoT devices.

2. Impact

By observing the crime trends of the last 5 years (2013 to 2017) we ranked the top 10 neighborhoods by Total Crimes and Non-traffic Crimes. We also looked at the growth trends of crime by category. This tracking could assist the law enforcement agencies to place more focus on the neighborhoods with the higher crime occurrences and provide preventive actions, such as planning educational programs in specific neighborhoods as appropriate. Furthermore, the correlation analyses with the Census data can help local judiciaries anticipate certain neighborhood inclinations to have higher crimes.

One compelling area of research that we addressed was investigating how adjacent Denver neighborhoods relate in terms of household income, population and crime. We used past demographic and crime statistics data, and applied statistical methods to analyze the relationships between these variables.

Geographic Information Systems (GIS) is a platform by which data can be analyzed quickly and viewed graphically thus providing a user understanding of data in a more robust and intelligent way. We used the “Operation Dashboard” - an extension to the ArcGIS platform – to allow users to see all Denver Neighborhoods, zoom automatically and interactively to selected neighborhoods, and provide related neighborhood data such as crime statistics, demographic data (e.g., population and household income) and other data. Through the lens of an interactive dashboard a user can visualize geographically all data components as they relate to one another.

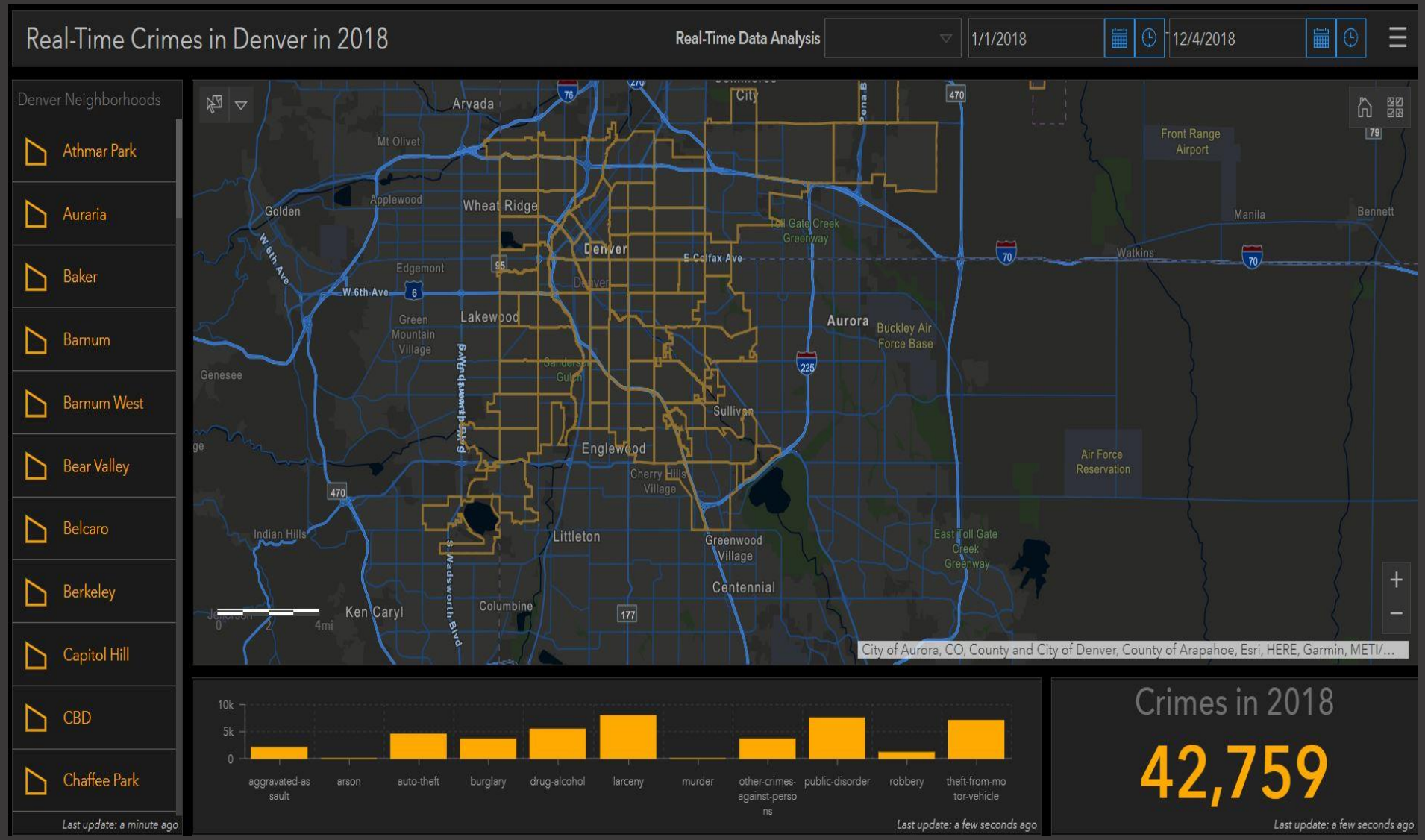
As an added bonus, real-time data such as a real-time crime data feed from the City of Denver, or live citizen crowdsourcing of reported crime incidents, can enhance the experience of a user by providing a mechanism to analyze data real-time. It allows for faster and more efficient decision making for a first responder for example. As an aspect of IoT (Internet of Things) and how it relates to a Smart City, crowdsourcing is becoming a hot topic. By adding a crowdsourcing element to the crime/demographic GIS interface we can create a valuable real-time IoT component. The ingestion of the crowdsourcing data adds to the Smart City element and provides real-time analytics in the dashboard.

Along with crime data, real-time crowdsourcing data can be used to monitor crime and traffic incidents instantly and answer such queries as “Is there a noticeable uptick in crime incidents in the Montbello neighborhood today?”. The GIS dashboard with its real-time component provokes many questions of “Why?” Some answers can be found in the real-time correlation analysis tool within the dashboard.

Using our real-time GIS, real-time IoT crowdsourcing, and interactive spatial correlation analysis, we provide a better way to explore, analyze, decide and react to real-time information in real-world scenarios. Studying the use case of crime and how it relates to geographic neighborhoods - both spatially and temporally - gives us a sense of how real world relationships correlate, and provides an opportunity to make better decisions at the state, city and neighborhood levels. By taking into account real-time spatial data and real-time analytics, Smart City crime analysis becomes a reality!

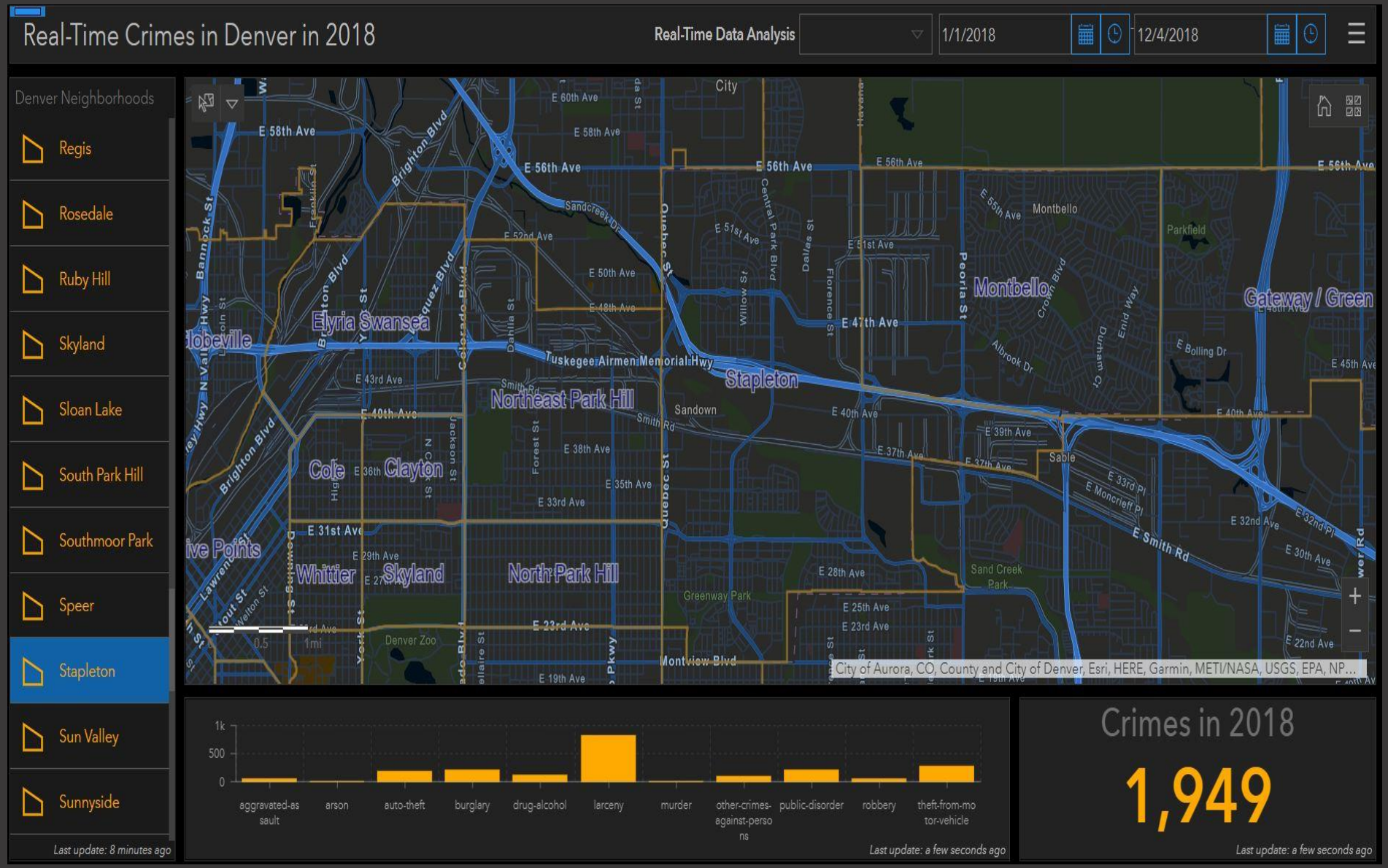
3. Methodology

All analyses, calculations and data manipulations are conducted using R version 3.5.1 which is an open source Statistical Computing package. Data is sourced from the reliable public data websites, reviewed and cleaned to ensure it is ready for analyses. Before conducting statistical analyses, we fully surveyed all variables and the structure of the data sets in order to prevent misinterpretation of the relationship of the data. We conduct multiple regression analysis on the 2017 crime data and 2011-2015 census data on the Denver neighborhoods, which are the latest data we could find from a reliable data source. The results were carefully reviewed to ensure the integrity of the analyses and the accurate interpretation of the summary statistics.



Web-based Dashboard

To better visualize target data, we then used Web GIS, an intuitive, vibrant, interactive real-time tool. We used a web extension called the Operations Dashboard. Its interface provides an interactive and dynamic user experience unparalleled by any web experience we've seen for the City of Denver or the State of Colorado.

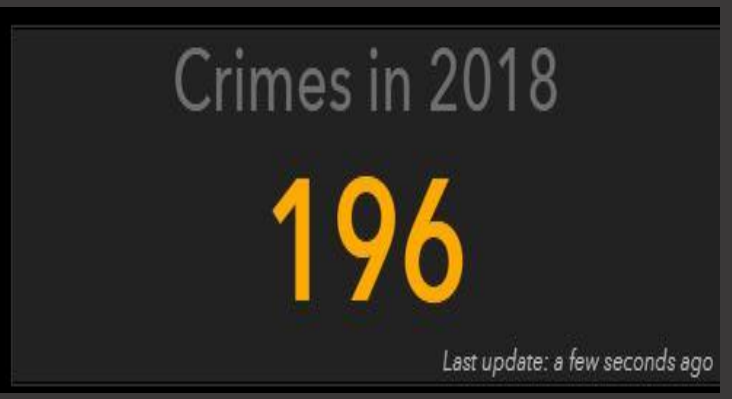


Neighborhood Crime

Within the dashboard you can report on crime neighborhood numbers by clicking the specific neighborhood (left). The dashboard then shows crimes for the selected neighborhood.

The bar chart (bottom) shows crime distribution by type.

Neighborhood labels appear on the map once a zoom action zooms to the map extent of the selected neighborhood.

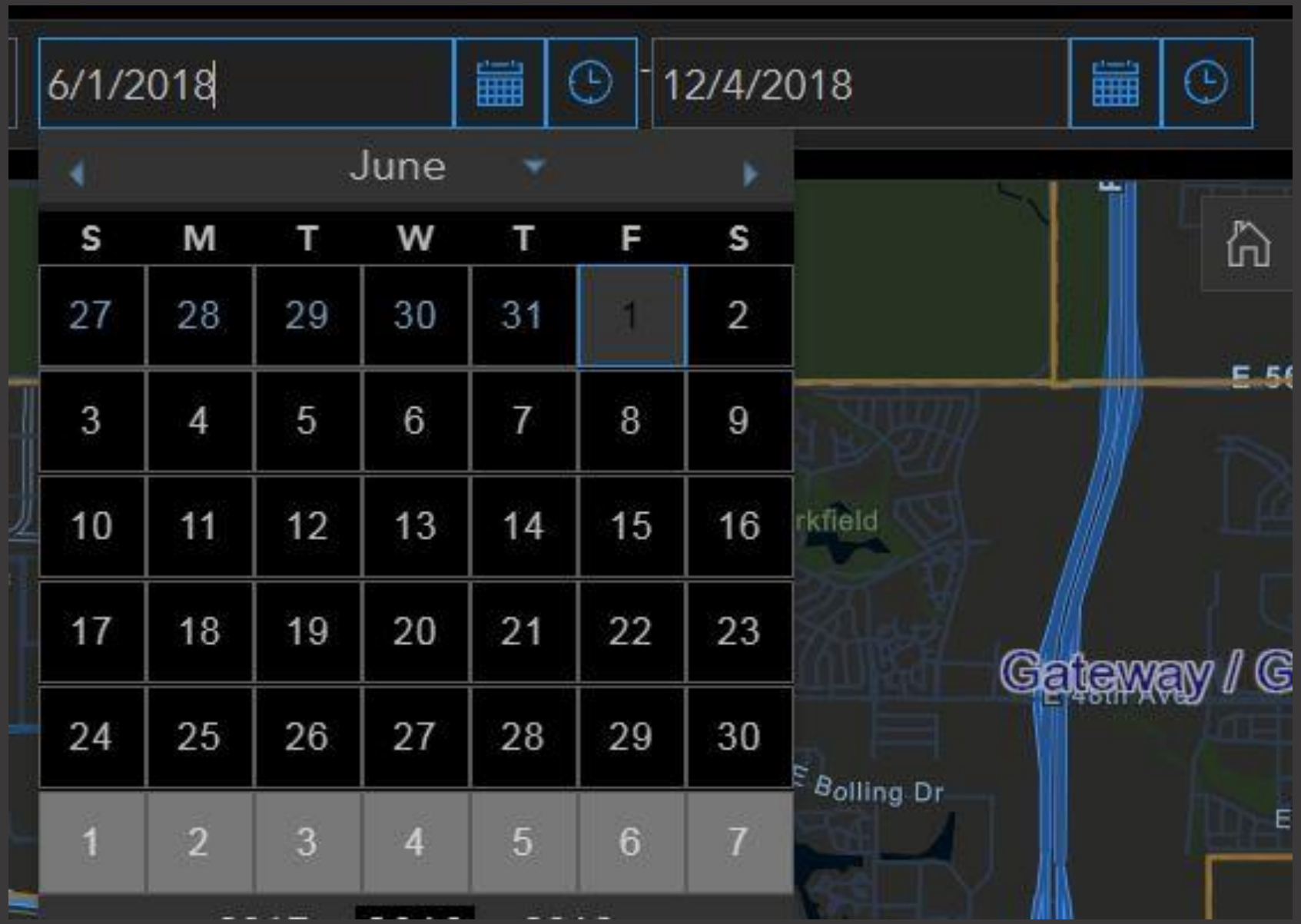


Crimes by Type

The bar chart in the dashboard displays the different types of crime within the selected neighborhood

You can focus on a specific crime type by clicking its respective bar graph element.

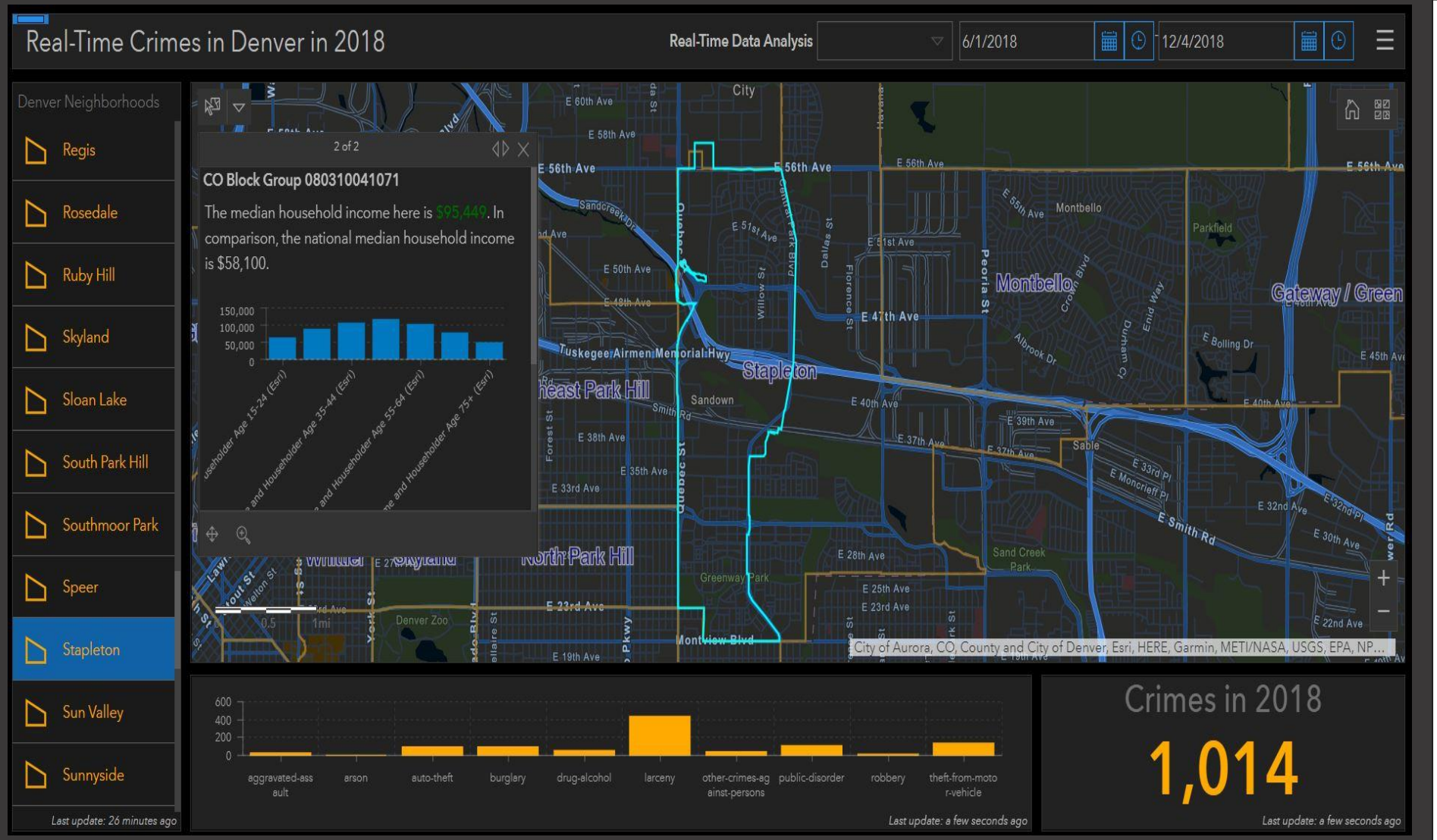
In this example we are showing the crime count (in this case, burglary) for the Stapleton neighborhood.



Time Range Indicator

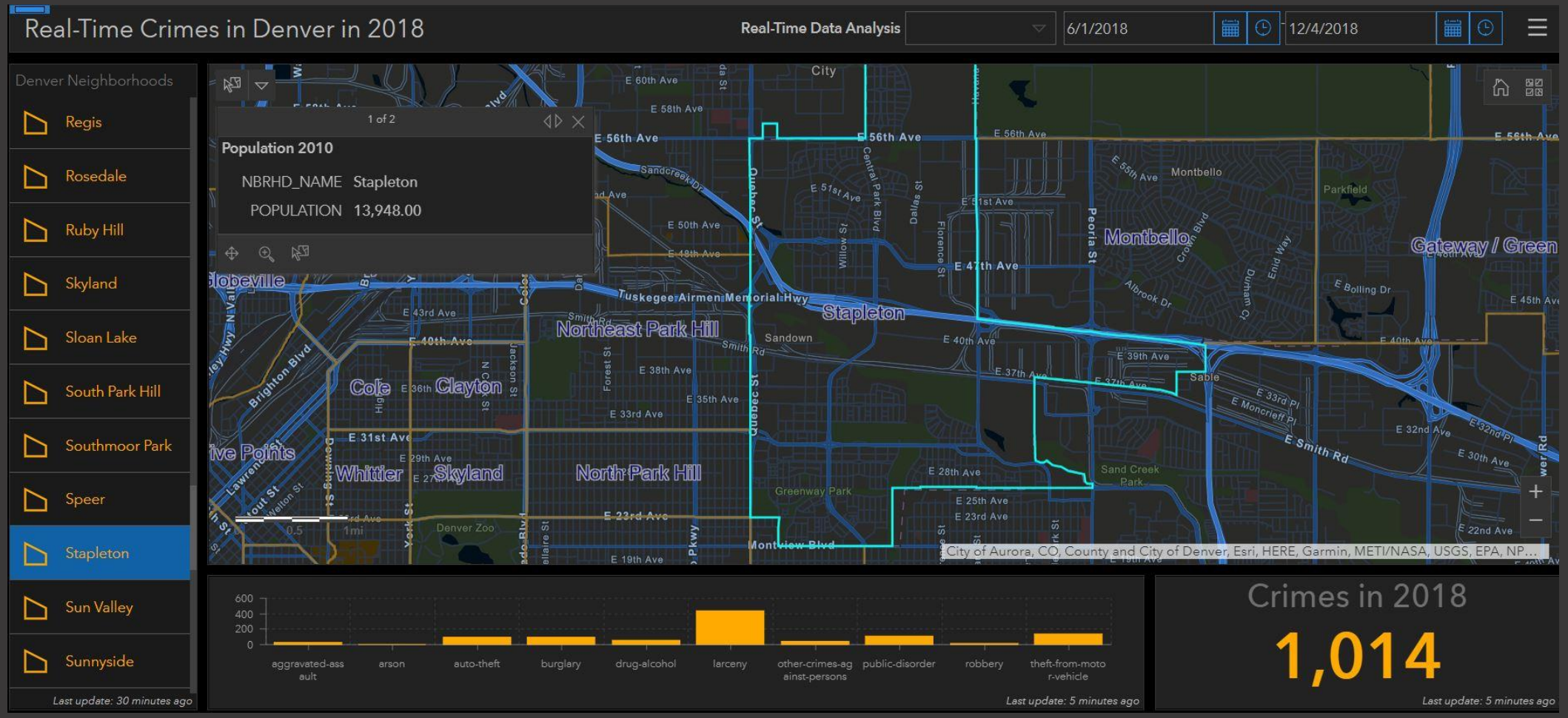
The dashboard provides a time-range indicator to refine a crime count by a date range. The interface is interactive and dynamic.

When a date range is selected, the bar chart for total crimes will reflect the selected dates (This is done through a date range query in the GIS).



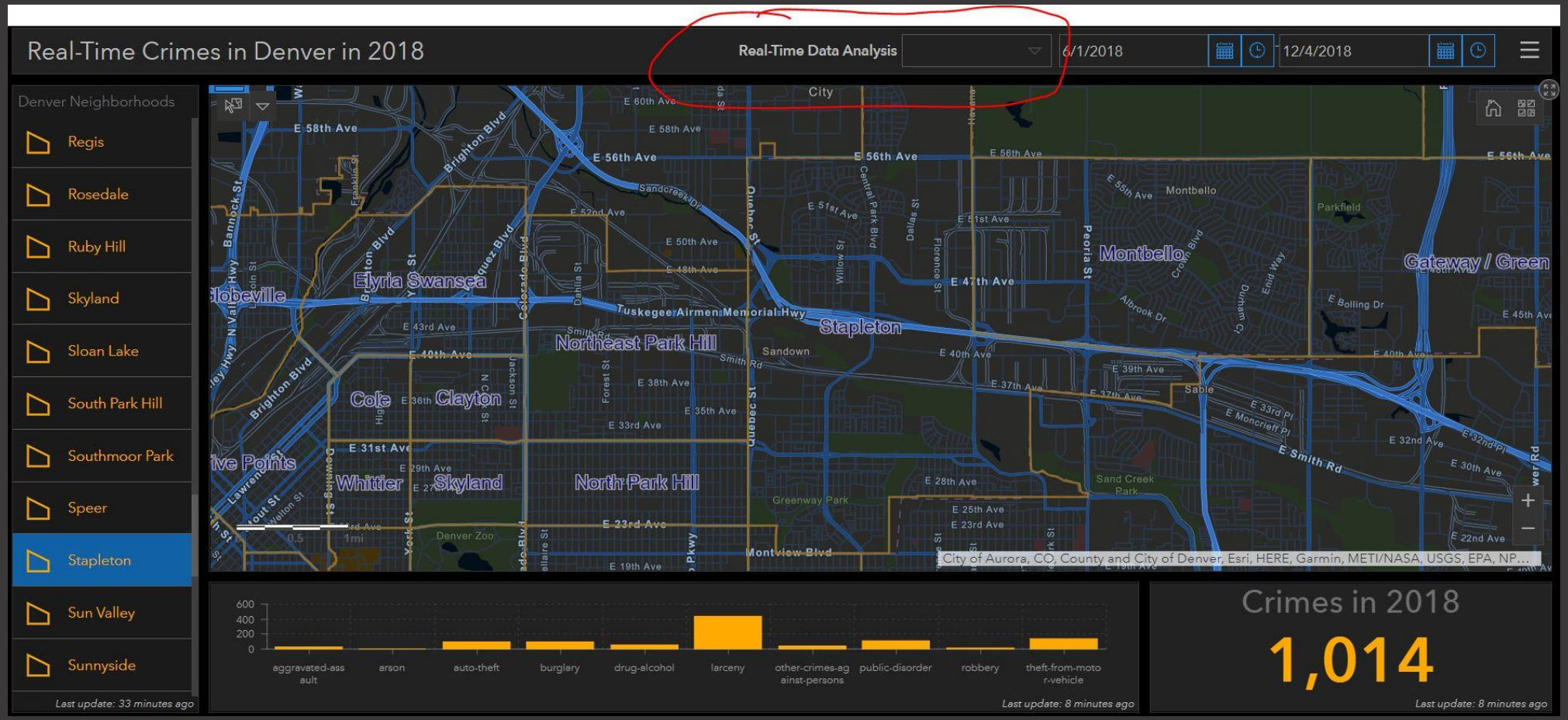
Household Income by Neighborhood

By clicking on the neighborhood polygon on the interactive map, a popup menu will appear and present data that shows the information for household income ranges by census block group.



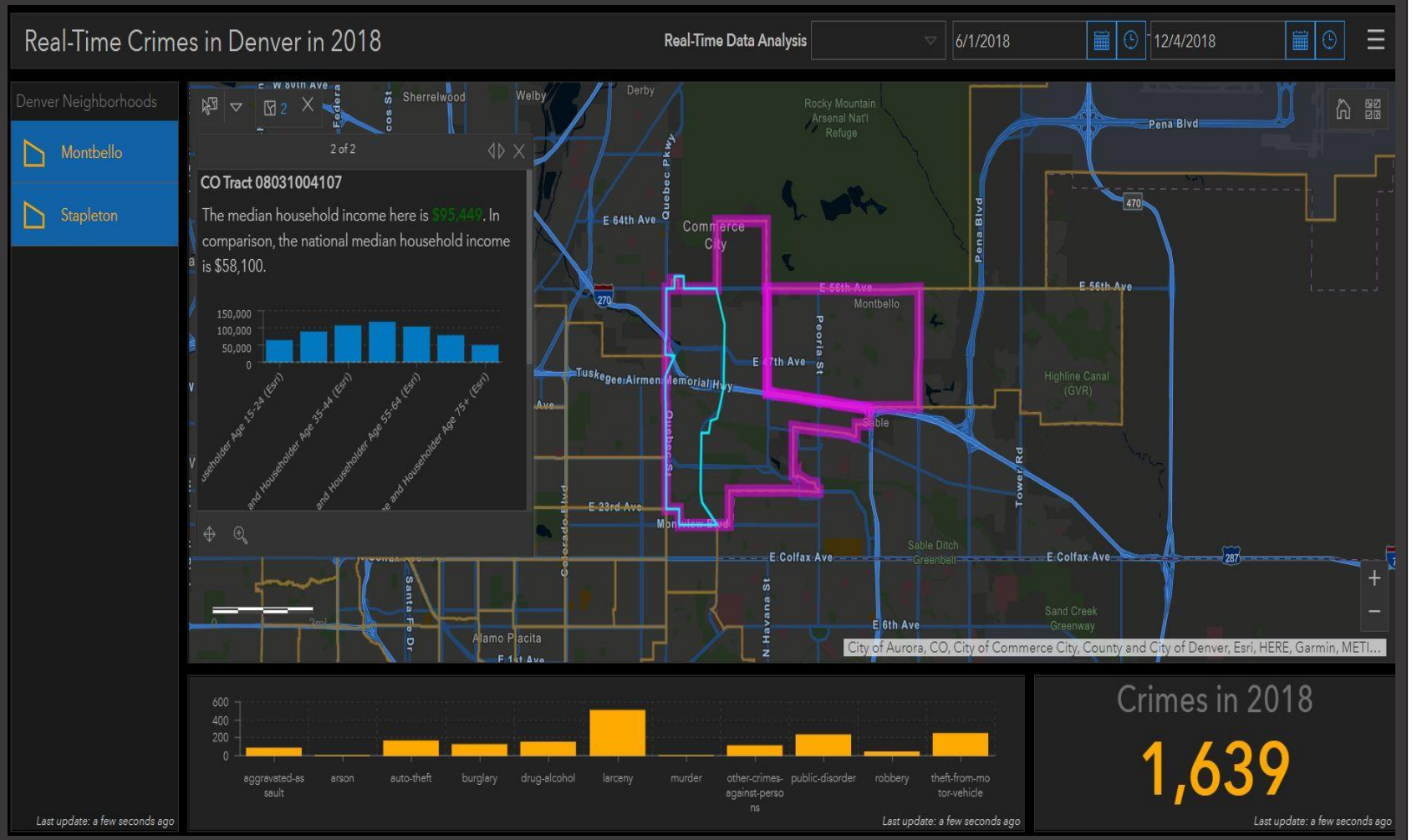
Population by neighborhood

By clicking on the neighborhood polygon on the interactive map, a popup menu will appear and present data that shows the information for population



Integrate Realtime Analytics through a Widget

This analysis can be done by accessing the R analysis on the fly through an API into the interface. We would suggest integrating any R related analysis to be done in ArcGIS Enterprise extension GeoEvent server with GIS database residing on a cloud platform.



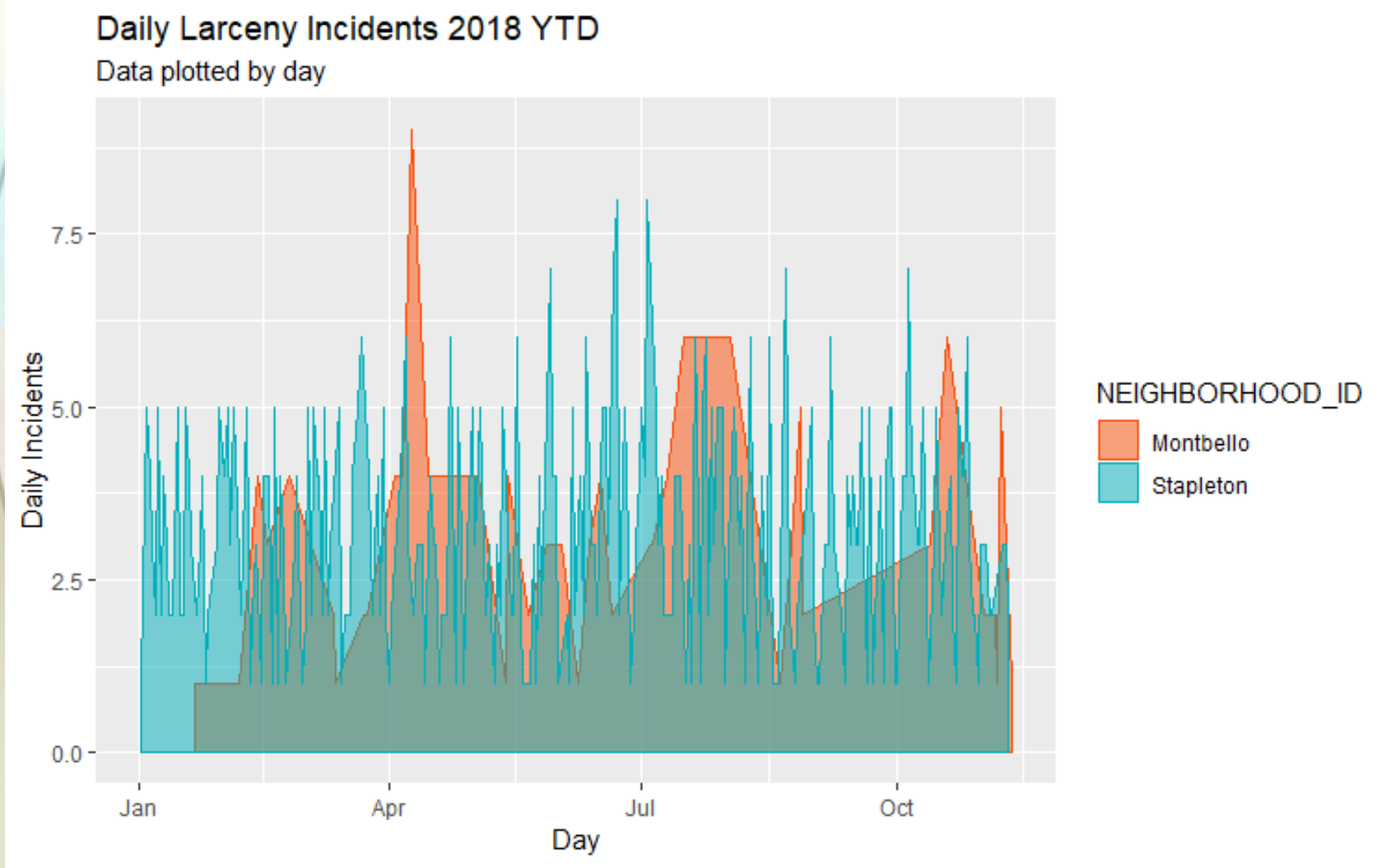
Additional Features

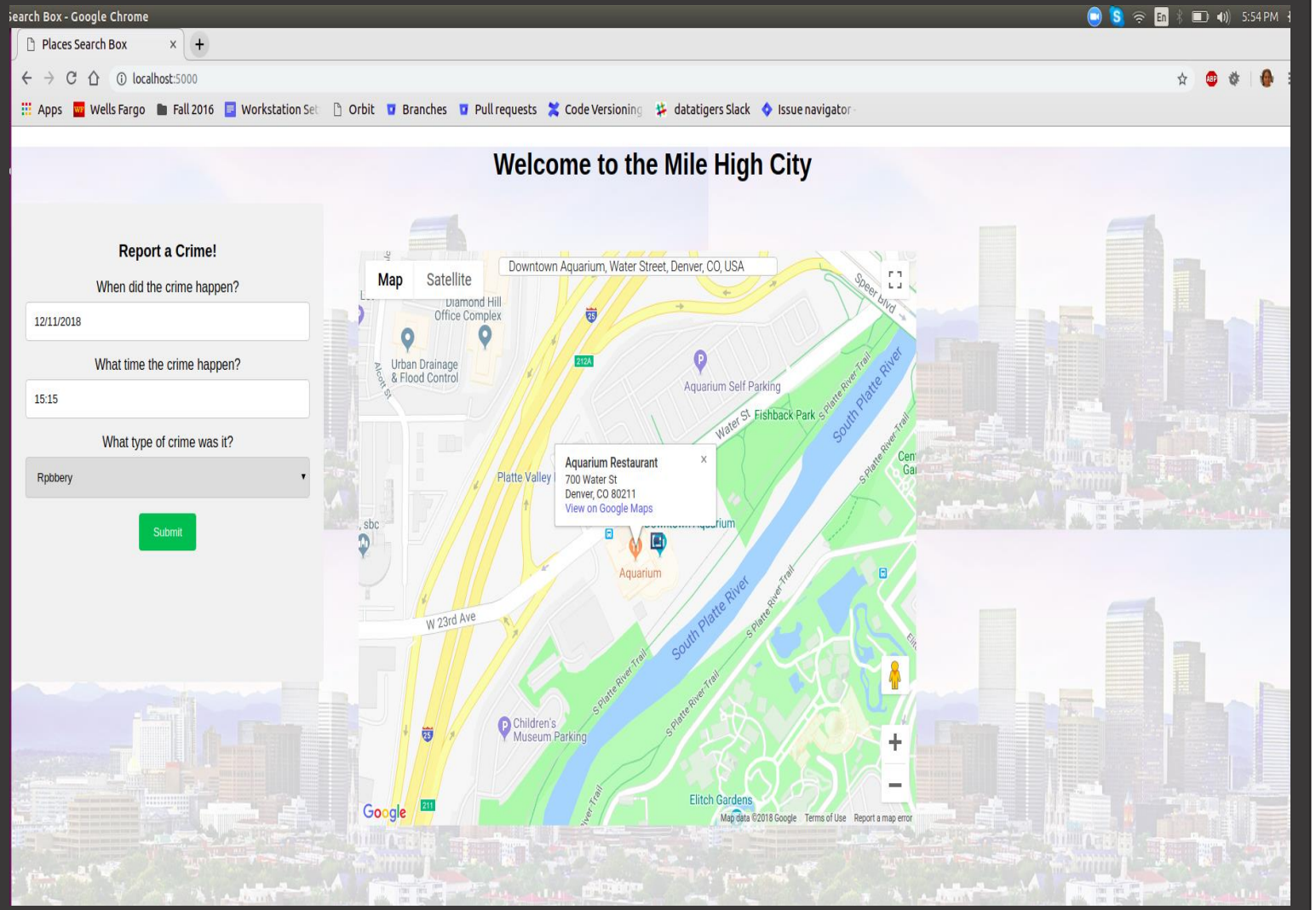
You can do side-by-side neighborhood comparisons through crime x population x income statistical analysis.

Further statistical analysis can be automatically performed by selecting the "Real-Time Data Analysis" dropdown menu (top center of screen).

The next slide demonstrates what this detailed analysis looks like (graphically) when comparing the Montbello and Stapleton neighborhoods.

Two Adjacent Neighborhood Comparison – Montbello vs Stapleton





Real-time data and Crowdsourcing

In addition to using the historical data currently available from the City of Denver, our tool is capable of taking in a real-time data feeds from outside sources such as an IOT link or our crowdsourcing application. The real-time data can then be ingested and pushed directly into the dashboard or GIS.

Our crowdsourcing application allows citizens to provide real-time reporting of local crimes, either using an exact address or an approximate location, such as an intersection, general vicinity, etc.

4. The Data

Crime Data Set:

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime

Obtaining and preparing: download the .csv file and imported into R Studio. Convert the timestamp data type to a format that R language can understand so the date related functions can be applied.

Census Data Set:

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-american-community-survey-nbrhd-2011-2015

Obtaining and preparing: download the .csv file from the above web site and imported into R Studio. Clean the data by converting some values (remove “-”, convert to capital letters and etc.). Calculate the ratio/percentage to prepare for the regression analyses.

Denver Census Population: https://www.denvergov.org/opendata/search?q=population

Denver Neighborhood Boundaries: https://www.denvergov.org/opendata/search?tag=boundaries&page=3

City of Denver Crime Data: https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime

Shapefiles: The following shapefiles, in GIS file format, were loaded into a web GIS called ArcGIS Online (AGOL) as feature layers.

- *Household Income* - 2018 USA Median Household Income - Source ESRI https://www.arcgis.com/home/item.html?id=3e5f8ebe5a114a61b7f350e7a1203761
- *Aerial Background* - Source ESRI, Digital Globe 2018 Default background layer in AGOL.

5. Feasibility

The analyses and capabilities are feasible as described. Additional analyses can be conducted with additional time and resources. For example, we can slice and dice the data to yield many varied and useful perspectives. Using the R code we developed, we can conduct additional analyses as needed. Some interesting possibilities include the seasonality of the crimes, crime reporting cadence, comparable neighborhood crime rates. This is only the beginning…

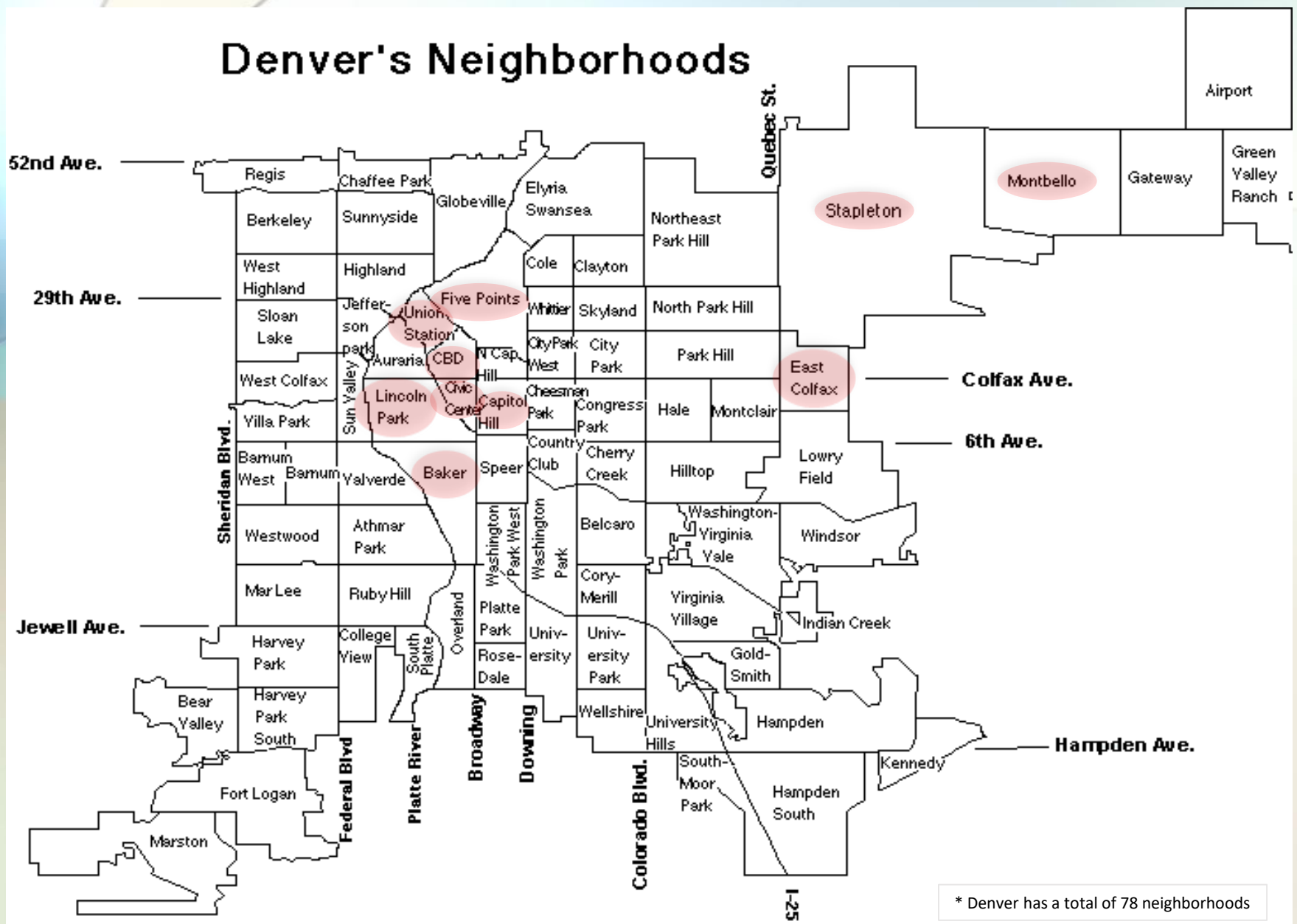
A desktop-based R package can handle the current size of data, but over time the data will continue to grow, and there will likely be requirements for real-time analytics. In anticipation of this need we are planning the implementation of a cloud-based database with an R front end. ArcGIS and Google Maps are already cloud-based and will be integrated and scaled appropriately for the implementation.

6. Results

The following slides demonstrate the detailed analysis that our tools can provide. Additional analysis using other variables and data are available as needed.

We are planning to present our work at the upcoming GIS in the Rockies premier geo-spatial conference in Denver and the ESRI International GIS conference in San Diego, CA.

Top 10 neighborhoods of Total Incidents (2017 All Incidents)



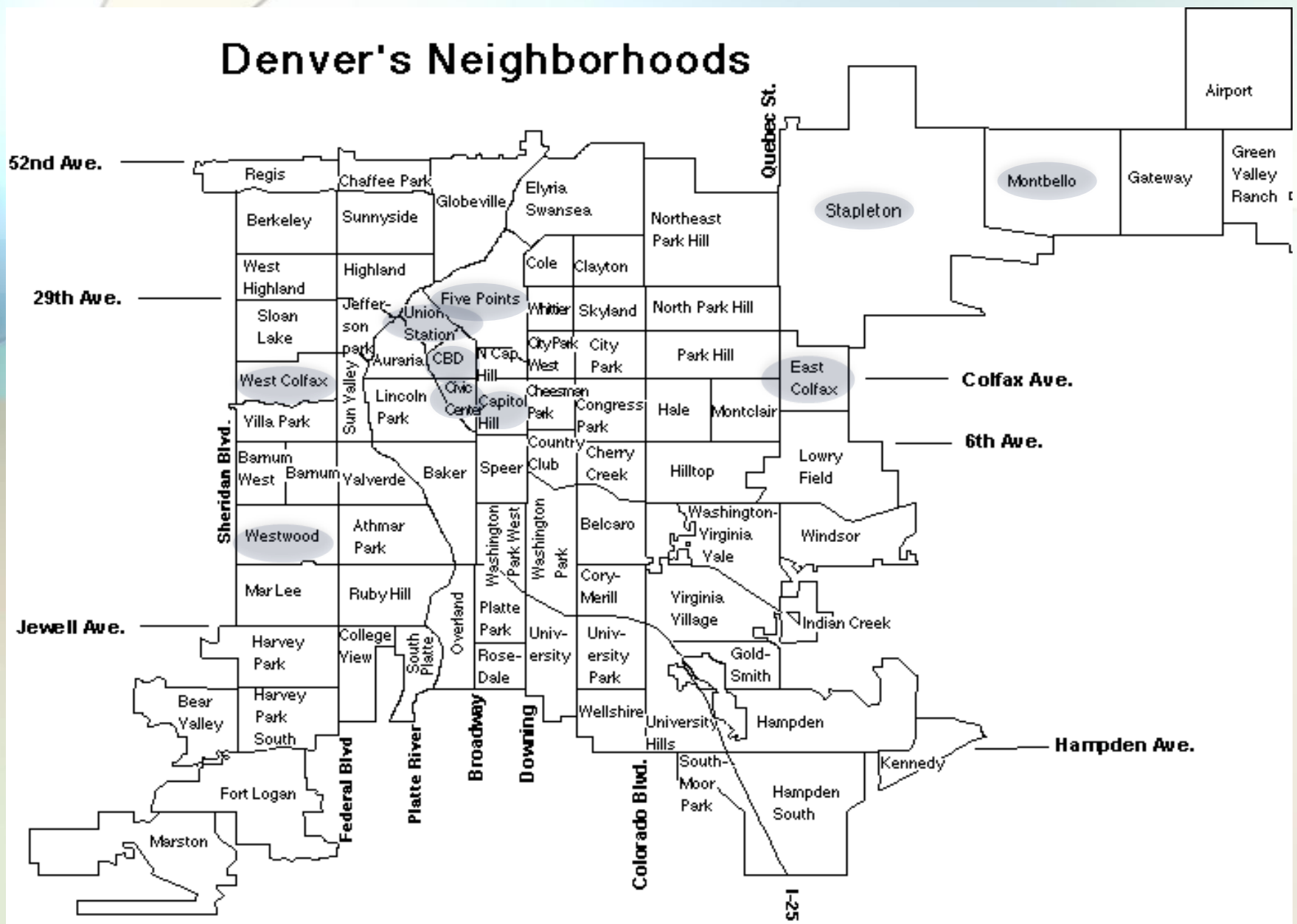
* Denver has a total of 78 neighborhoods

Neighborhood Name	2017	Growth Rate 2013-2017
Five Points	5,052	**63%**
Stapleton	3,659	42%
Capitol Hill	3,253	41%
CBD	3,037	15%
Montbello	2,956	46%
East Colfax	2,541	52%
Baker	2,480	31%
Civic Center	2,308	53%
Lincoln Park	2,253	26%
Union Station	2,196	**70%**

From 2013 to 2017 the total number of crime incident of Denver neighborhoods grew **36% (2013: 47,896; 2017: 66,434)**

- Union Station grew 70%
- Five Points grew 63%
- Civic Center grew 53%
- East Colfax grew 52%

Top 10 neighborhoods of Non-traffic Incidents (2017)

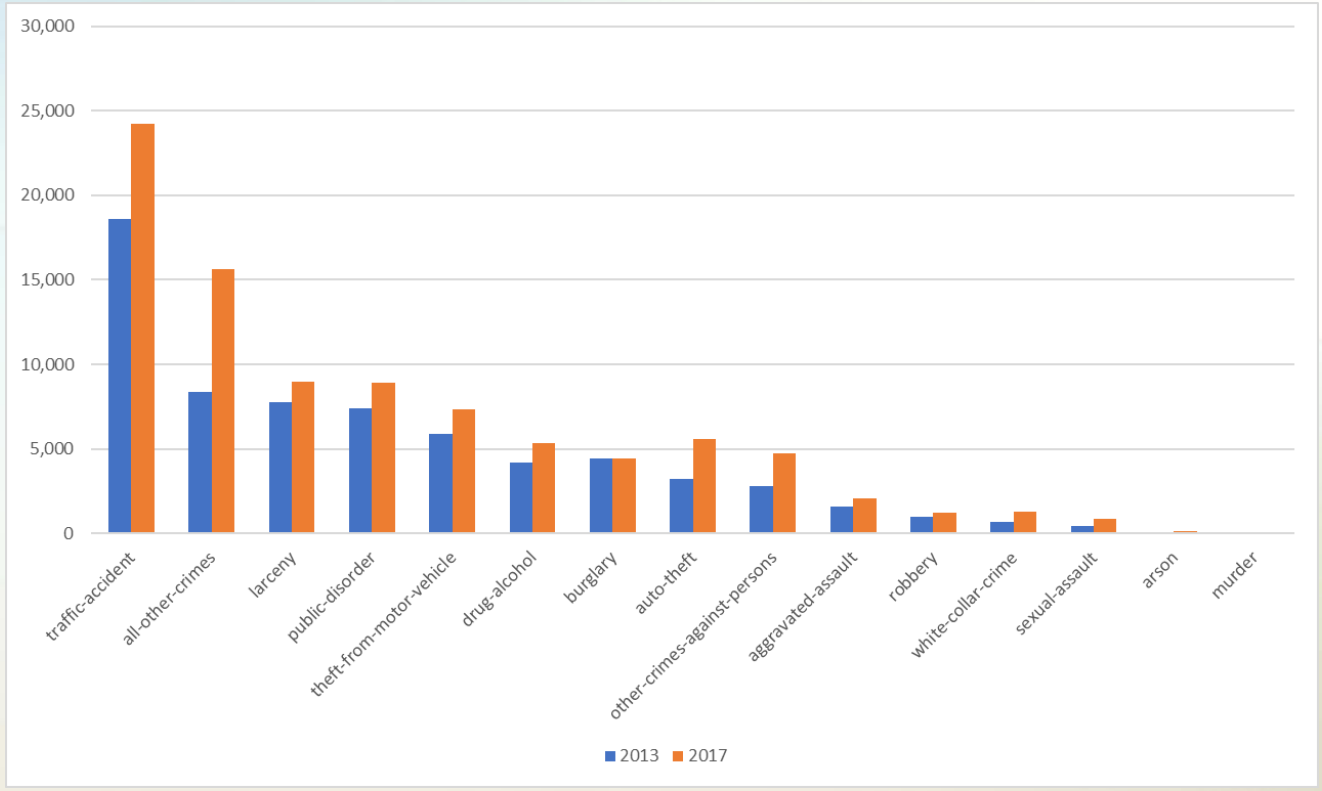


Row Labels	2017	Growth Rate 2013 to 2017
Five Points	**4,197**	**68%**
CBD	2,571	15%
Capitol Hill	2,552	49%
Stapleton	2,485	54%
Montbello	2,337	45%
East Colfax	2,246	54%
Civic Center	**1,939**	**61%**
Union Station	**1,758**	**75%**
West Colfax	**1,724**	**91%**
Westwood	1,606	17%

From 2013 to 2017 the total number of non-traffic crime of Denver neighborhoods grew **39%** - West Colfax, Union Station, Five Points and Civic Center among the highest.

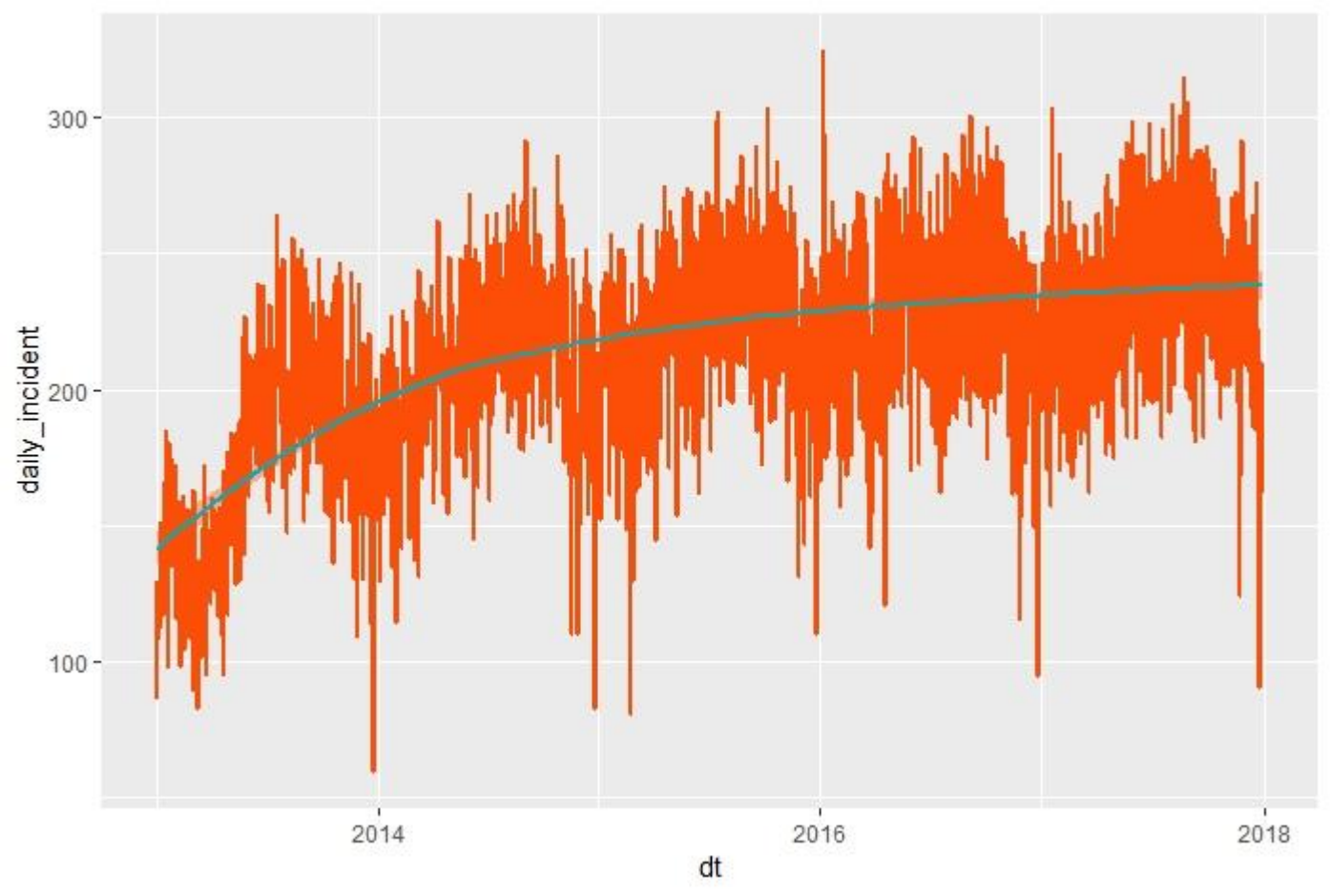
Incident Rank

Crime Category	2013	2017	Growth Rate
traffic-accident	18,600	24,225	30%
all-other-crimes	8,377	15,608	86%
larceny	7,753	8,957	16%
public-disorder	7,424	8,890	20%
theft-from-motor-vehicle	5,892	7,329	24%
drug-alcohol	4,190	5,312	27%
burglary	4,460	4,439	0%
auto-theft	3,199	5,559	74%
other-crimes-against-persons	2,789	4,731	70%
aggravated-assault	1,555	2,076	34%
robbery	1,009	1,206	20%
white-collar-crime	666	1,295	94%
sexual-assault	450	853	90%
arson	91	123	35%
murder	41	59	44%
Grand Total	**66,496**	**90,662**	**36%**



From 2013 to 2017 almost all crimes increased except Burglary remained flat. White-collar-crime, Sexual-assault, All-other-crimes and Auto-theft were among the highest grow categories. All-other-crimes include Criminal trespassing, traffic other, public-order-crimes-others, vehicular-eluding-no-chase and etc.

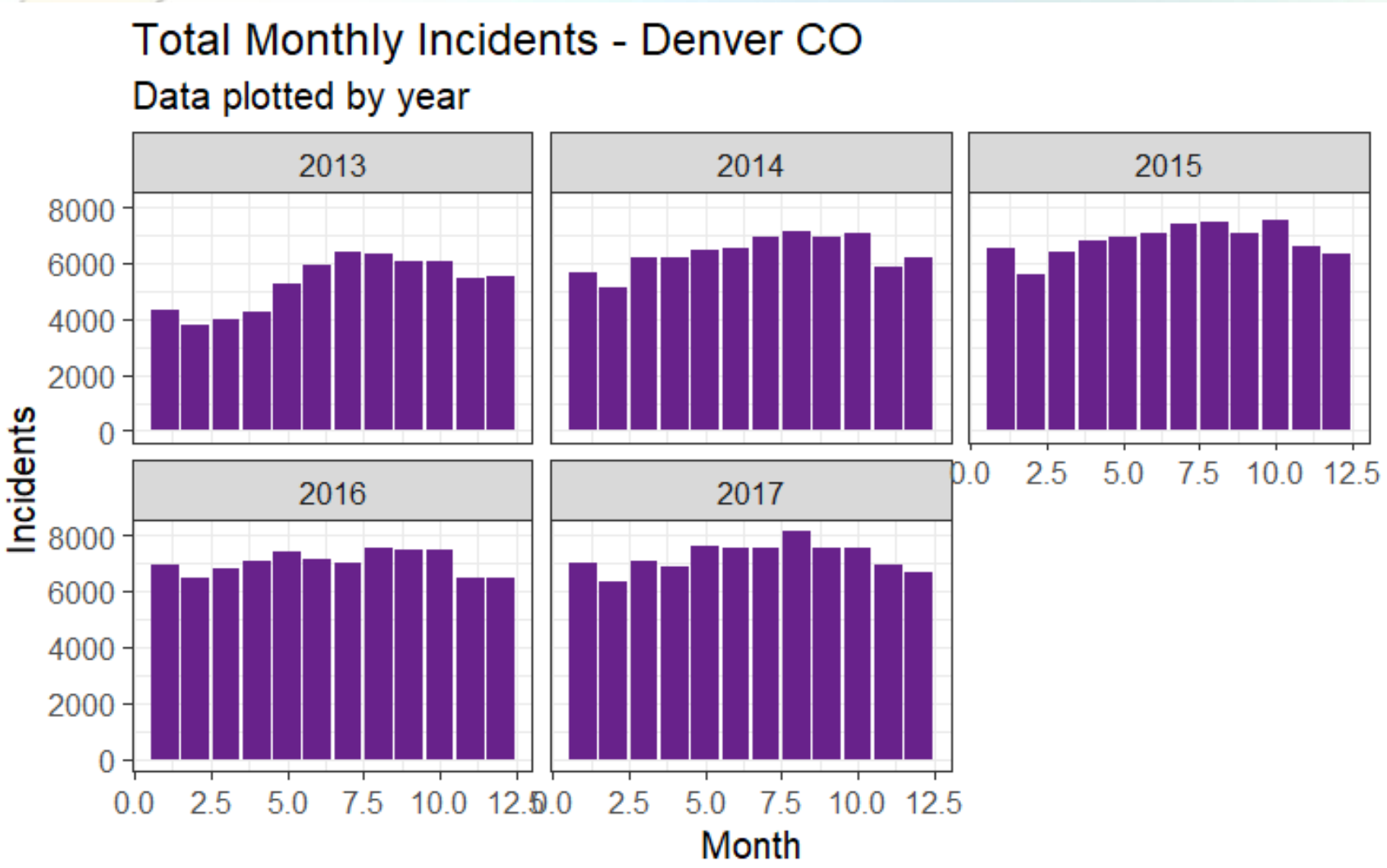
Year over Year Trends



Year	2013	2014	2015	2016	2017
Total Incidents	**47,896**	**59,417**	**63,157**	**64,958**	**66,437**
YoY Increase (%)		24%	6%	3%	2%
Cumulative Increase (%)		24%	32%	36%	39%

From 2013 to 2017, although over the five-year period total incidents increased 39%, the increasing pace is definitely slowed down over the last 2 years.

Monthly trends by Year



February, November and December were the months with lower incidents. The middle of the year seems to have higher incidents

Correlation analysis with Census Data

Larceny:

- Positive correlated with HOUSEHOLDER_ALONE, BUILD_2010_OR_LATER, FRENCH_LNG
- Negative correlated with BUILD_2000_2009
- Multiple linear regression model can explain 52% (Adjusted R Squared 0.52) of the incidents at 90% confident level

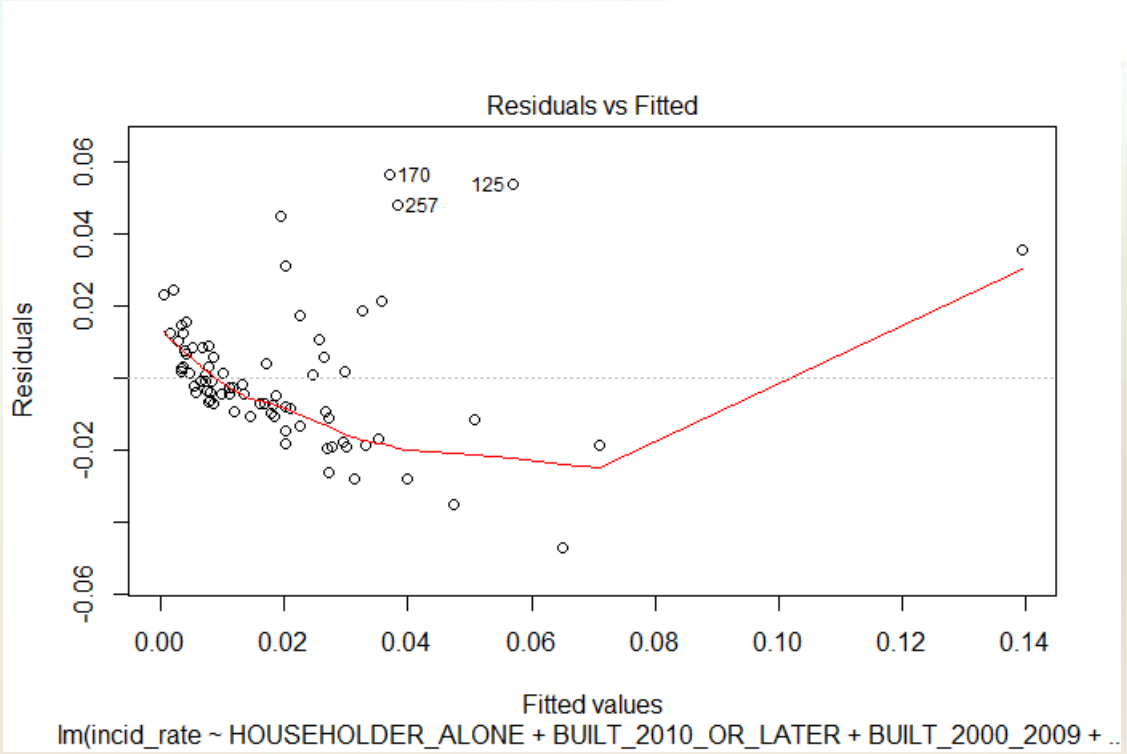
Outputs:

```
Call:
lm(formula = incid_rate ~ HOUSEHOLDER_ALONE + BUILT_2010_OR_LATER +
    BUILT_2000_2009 + FRENCH_LNG, data = crime_hh)

Residuals:
      Min        1Q    Median        3Q       Max
-0.047303 -0.009881 -0.002599  0.008259  0.056450

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.570e-03  3.459e-03   1.032 0.305345
HOUSEHOLDER_ALONE  4.862e-06  2.081e-06   2.337 0.022206 *
BUILT_2010_OR_LATER  1.389e-04  2.022e-05   6.867 1.83e-09 ***
BUILT_2000_2009    -1.483e-05  4.078e-06  -3.638 0.000509 ***
FRENCH_LNG         1.132e-04  5.696e-05   1.987 0.050673 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01906 on 73 degrees of freedom
Multiple R-squared:  0.5432,  Adjusted R-squared:  0.5182
F-statistic:  21.7 on 4 and 73 DF,  p-value: 7.91e-121
```




```
(Intercept)   HOUSEHOLDER_ALONE BUILT_2010_OR_LATER     BUILT_2000_2009
 3.570201e-03       4.862299e-06       1.388578e-04       -1.483418e-05
       FRENCH_LNG
 1.131878e-04
```

COLORADO
Governor's Office of Information Technology
Serving people serving Colorado

Data Tiger Team Members

• Kevin Smerdell	Kevin.Smerdell@Aerovine.com	303-522-4520
• Ed Vigil	EdVigil@gmail.com	303-619-8500
• Helen Wang	Helen.Wang@Aerovine.com	312-898-1949
• Apoorva Bapat	AndBapat@gmail.com	720-421-8579
• George Dinakar	George.Dinakar@state.co.us	720-560-7676

Aerovine's Capabilities Statement has been included in the repository.
See https://github.com/Colorado-Data-Analytics-Challenge/Data-Tigers


Q & A

COLORADO

Governor's Office of Information Technology

Serving people serving Colorado

Thank you!