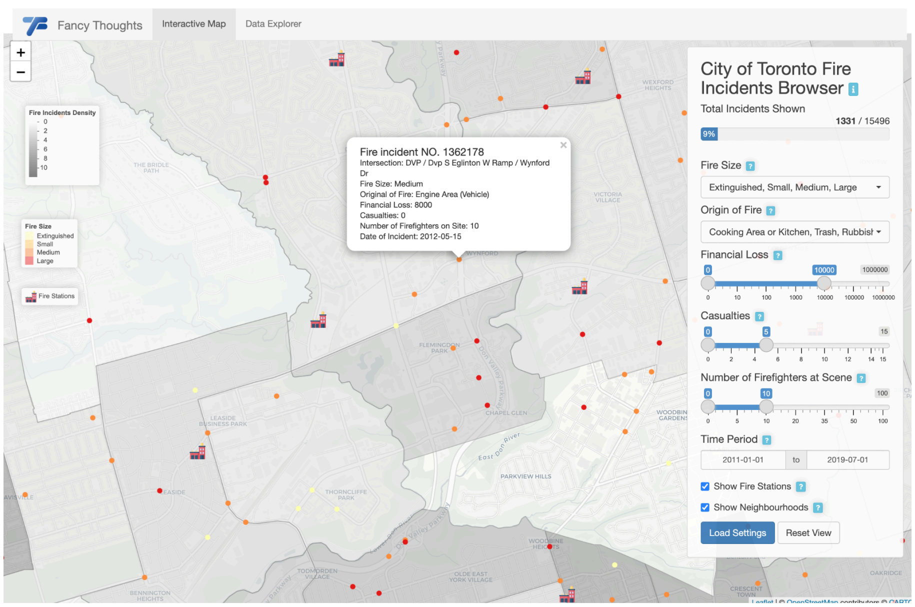
[Project] City of Toronto Fire Incidents Browser

[Presented by] The Fancy Thoughts Squad

Yiheng Bian ; Yilin Huang ; Zijing Huang ; Sijie Xu



[*Link to the Product*](https://sijiexu.shinyapps.io/City_of_Toronto_fire_incidents_browser/) : https://sijiexu.shinyapps.io/City_of_Toronto_fire_incidents_browser/

[*Link to the Source Code*](https://github.com/JackXu2333/STA313_Final_Project) : https://github.com/JackXu2333/STA313_Final_Project

If one merely runs the keywords “Toronto Fire” through Google search, one will see that the first 5 pages of the news section will detail every fire incident that has occurred in Toronto within the past month. Our group chose the topic of fire incidents in the City of Toronto, in order to identify the fire incident distribution in Toronto’s neighbourhoods and provide possible solutions for preventative measures that could be taken to minimize both financial loss and life loss. Included as our main audiences are both fire department and home insurance companies in the City of Toronto, with our ultimate goal to inform them about the pertinent details of fire incidents in the City of Toronto. More specifically, this project aims to enable the fire department in strategically deciding where additional fire stations should be built, as well as how firefighting resources should be distributed within the city. This will ensure that they can provide superior public safety services, and inform insurance companies about how to offer better home insurance plans to homeowners. This topic is related to social good due to its close connection to social and public safety. If the distribution or pattern could be isolated from the data, the fire department will be able to decide exactly where and which neighbourhoods they could invest more in, for the goal of prevention and firefighting capacity. Also, insurance companies will be able to identify high-risk areas to provide tailored plans that benefit every homeowner. This topic is thus critical for our targeted audience as it will be used as an opportunity to educate, and ultimately help them to plan and design policies, for additional fire stations, designing new fire education policies or designing new insurance policies.

The data that are used in the visualization is from *The City of Toronto’s Open Data Portal*. The three datasets used included fire incidents data, fire station locations, and neighbourhood profiles (Toronto Census 2016). All explanations for each variable term will be featured on the interactive design.

For the fire incidents data, the relevant data dimensions utilized include the following: Status_of_Fire_On_Arrival, Area_of_Origin, Estimated_Dollar_Loss, Civilian_Casualties, Number_of_responding_personnel, TFS_Alarm_Time, and Intersection. On the design, the Status_of_Fire_On_Arrival was represented by Fire Size. Fire sizes are represented by four different types, Extinguished: fire was extinguished prior to arrival; Small: there was no evidence of fire from the street; Medium: fire with smoke only or flames coming from a small area; Large: flames were showing from a large area or that and explosion was involved; Area_of_Origin was represented by Origin of Fire, which is the most common origin of fire (top four origins, the rest of the origins are in the “Other” category); Estimated_Dollar_Loss was represented by Financial Loss; Civilian_Casualties was represented by Casualties, which is Civilian casualties observed at scene, including both injured and dead; Number_of_responding_personnel was represented by Number of Firefighters at Scene; TFS_Alarm_Time (which is timestamp of when TFS(Toronto Fire Service) was notified of the incident, the date was extracted from here to represent the date of incident) was represented by Date of Incident (yyyy-mm-dd), which was also used in variable Time Period; Lastly, Intersection, is the nearest major or minor intersection (in the format of Road A/Road B) in the ward of the incident. For the fire station locations data, the relevant data dimensions used include address. This refers to the exact address of the fire station, including accurate street number and street name. Finally, for the neighbourhood profiles data, the relevant data dimensions used in each neighbourhood include the name of the neighbourhood, its population, its land area in square kilometres (which was represented by Total Area), and its Population density per square kilometre (which was represented by Population Density).

Other data dimensions that did not exist in the above data sets but were nonetheless generated by ourselves include the total number of incidents in the city or in each of the neighbourhoods (which was represented by Total Incidents); the ratio of the number of fire incidents in the neighbourhood over the area of the neighbourhood (which was represented by Fire Incident Density); the average casualty, the average financial loss, the average response time, the average household sizes, average personal income in each of the neighbourhoods; the number of fire stations, the total number of fire incidents, and dwelling density (which refers to the number of dwelling occupied by usual residents divided by the area of land in square kilometres), the percentage low-income population, and the percentage of the population without a diploma in each neighbourhood.

The main analytical questions that our design aims to provide a solution for, are elaborated in the following: How are fire incidents spread in Toronto? What factors affect the average response time, the average casualty, and the average financial losses in each neighbourhood?

By initially generating and developing a sufficient overview of the fire incidents, and then delving into further pertinent details, we hope to provide useful insights to readers to achieve our purpose. For example, if turns out that the average response time is particularly long in one neighbourhood, or if certain factors are affecting the average response time – this might necessitate the neighbourhood's need for more fire stations to be built, or that perhaps some policy should be made to help eliminate the limiting factors. Similarly, if there are higher casualties or financial loss in one of the neighbourhoods, or if a certain factor is affecting those, then it can be argued that some safety plans need to be offered by the government, such as special home insurance plans from insurance companies to accommodate people living within those limiting factors.

There are two main facets of the visualization: 1) a map on the first tab and 2) a scatterplot on the second tab, with two extra tabs of data tables (within the second tab) to provide supporting information. When the website is first opened, an information box will pop up to the reader. This information includes the purpose of the visualization, a short description of each of the two views, and some technical tips to obtain an optimal view of the visualization in the reader's browser. The readers can close it whenever it is most convenient, and if they are interested in reading it again, they can click on a small button on the top right corner of the selection bar on the first tab (map) to open the info box again.

The first part of the visualization is a map of Toronto, which comprises the first tab on the visualization website, with each neighbourhood indicated within its borders. The reason for choosing a map for the design is that the medium provides perhaps the easiest way to view the overall fire incidents in Toronto along with their full details (fire size, financial loss etc), and to compare cases among areas. With this necessary context in mind, it is evident that our topic of interest is geographic by nature. There are a total of 140 neighbourhoods on the map; when the reader clicks on a neighbourhood, the neighbourhood area will appear larger and pop up in the center of the screen, with an information box about this neighbourhood linking to it. This information box includes geographical information about the name of the neighbourhood, the population, its total area, its population density, the number of fire stations, and the total number of fire incidents that have occurred (2011-2019) in this neighbourhood. Going back to the aforementioned pop-up neighbourhood map, there will be many small dots and a few red icons detailed on it. Each dot represents a fire incident. The location of the dot necessarily depends on the location of the fire incident, which is on which intersection that the fire occurred. The colour or hue of the dot depends on the size of the fire, which is the fire situation recorded when the firefighters first arrived at the scene. For the dots, yellow will represent extinguished fire, orange will represent a small fire size, orange-red for medium fire size and deep red for large fire size. The hues chosen are those which naturally mimic the different stages of fire. When a cursor is placed on a dot within the enlarged neighbourhood map, a small message window will pop up. This message window will include information on the specific incident, consisting of the following data: the intersection of occurrence, the size, the origin of the fire, the financial loss, the number of firefighters on site, and the time of occurrence of the incident. In addition to the small dots, there will also be some small red icons of fire department buildings featured on the map, indicating the location of fire stations. When the readers click on the fire station symbol, there will be an information box detailing the address of the fire station and the neighbourhood that it is in.

On the right side of the map, there will be a selection bar. This area has six variables, namely fire size, area of origin, financial loss, casualties, number of personnel present and the time period. There will also be two display buttons within the selection bar, including fire station display and neighbourhood background display, which are the options to choose for showing the fire stations and the neighbourhood background colour in the map, respectively. The neighbourhood background colour is represented by the ratio of the incidents that fit the criteria of the reader's choice over the area of the neighbourhood, which is fire incidents density by area. Colour luminance was used here; the higher the ratio, the deeper the neighbourhood background colour, the lower the ratio, the lighter the colour will be. In addition, there will be a ratio of the number of total incidents shown currently over the total number of (all) incidents on the top of the selection bar. For each variable, a small question mark will appear next to it. When the reader clicks on the question mark, the definition of the variable will appear. There will either be multiple categories to choose from, or there will be a slider to choose the range for each of the variables. Categories were assigned to categorical variables and sliders were assigned to quantitative variables. Sliders were designed for some variables so that the reader can choose which range to look at. Readers can choose the variable(s) and the category(s) that they are most interested in, then the fire incidents that meet their criteria will appear on the map. Single or multiple variables and categories can be chosen. If all options are left empty, all fire incidents that happened from 2011 to 2019 will appear on the neighbourhood map that the reader chose. For example, if the reader clicks on neighbourhood named “Runnymede-Bloor West Village”, and then chooses “Medium” for the fire size, and “Cooking Area” in the area of origin in the design, then those fire incidents that happened from 2011 to 2019 that were medium fires, which occurred in the cooking area in neighbourhood “Runnymede-Bloor West Village" will appear on the map. At the same time, at the top of the selection bar, the ratio of the fire incidents that fit the criteria would be shown as, for example, “7%, 700/10000”. This means that the fire incidents that fit the criteria take up seven percent of all the fire incidents, and the exact number of incidents that belong to this criterion is 700. Also, there is a zoom in/out button at the top left corner of the map, enabling the readers to zoom in and choose the smaller neighbourhoods that are packed with dots, which would be difficult and very crowded to look at on the original map, and they can similarly zoom out for larger neighbourhoods.

The second part of the visualization, which will be featured in the second tab, serves the purpose of providing additional information about the trends in the variables. From the first tab, the data is compared vertically, so the readers can focus on the geographic feature of a specific neighbourhood. While for the second tab, the data is compared horizontally, meaning that we are looking at data from Toronto as a whole and trying to find some meaningful insight from this angle. In this part, there will be an interactive scatterplot. A scatterplot was chosen for use because if there is a trend between the x and y variables, it can be easily depicted and displayed on the scatterplot. On the left side of this tab, there will be a short description detailing how to use the scatterplot and the additional tables for the reader’s benefit (which will be mentioned later). There will be a selection bar for readers to choose the x (independent variable) and y (dependent variable) variables as well as three additional options to check if they want to show the moving average, linear model, or x-axis in log scale (which will give the trend more clarity) on the plot. On the y-axis, the readers will be able to choose such variables as average response time, average casualty, average financial loss or the total number of incidents as the y variable. On the x-axis, the readers will be able to choose twelve relevant variables, such as dwelling density, average household sizes, population density, and average household income. Readers will also be able to visualize the relationship between variables by selecting the pair of their choice. For example, if the readers choose the average response time as the y variable, and the population density as the x variable, then the scatterplot of the relationship between the average response time and the population density will appear. Also, the calculated fit and explanation of the fit of the plot would be shown beneath the scatterplot (for example, Average Response Time = 0.95* Population Density + 2.5, the model shows that for the increase in 1 unit of population density, the average response time will change by 0.95). The readers will be able to determine from the plot and from the actual fitted model, whether the average response time increases as the population density increases.

The additional two tabs within the second tab will be there for supporting information. Data used as variables in the scatterplots will be depicted as full tables, with sorting features. There are two tables, one in each tab, the Neighborhood table and the fire incidents table. To elaborate, the total number of incidents (in each neighbourhood), which is a variable from the second tab, will be shown as a column with 140 entries (140 neighbourhoods) within the neighbourhood table. The readers will be able to sort these from largest to smallest (or vice versa) which will enable them to see the neighbourhood ranking of the total number of fire incidents.

The visualization and interactive design serve to support answering the proposed analytical questions. Our first question was “How are the fire incidents spread in Toronto?”. To begin with, from the map, if we choose to look at the fires by fire sizes, and select the financial loss of 1000+ and the casualties of 1+, we would see that high incident density (by area) areas with extinguished fires are mostly in Neighborhood Danforth East York, North St.James Town and Moss Park. For small fires with the same conditions (financial loss 1000+ and casualties 1+), the high incident density neighbourhood is North St.James Town. For medium fires, the high-density neighbourhood is Church-Yonge Corridor. Lastly, for large fires, the high-density neighbourhoods are Wychwood and Greenwood-Coxwell. Within these high incident density neighbourhoods involving losses, Moss Park, Church-Yonge Corridor and Greenwood-Coxwell do not have any fire stations within neighbourhoods. This suggests that the fire department might want to build a fire station within each of the neighbourhoods to decrease the number of losses. Also, if we choose to look at the incidents by the origins of fires with 1000+ financial loss and 1+ casualties, we will see North St. James Town has the highest incident density of kitchen fires, Regent Park has the highest incident density of trash fires, Weston and Moss Park with porch fires, and Church-Yonge Corridor with vehicle fires. Community lessons might be offered by the fire department within community centers in these neighbourhoods to provide specific instructions and preventative measures including how to prevent fires of these origins and what to do to prevent losses when there is a fire of these origins. Also, there are some repetitive neighbourhoods from sorting by fire sizes and sorting by origins of fires with losses: Neighborhood North St. James Town with extinguished fires and origins from kitchen or cooking areas, Moss park with extinguished fires and origins from porch or balcony, and Church-Yonge Corridor with medium fires and origins from vehicle engines. Insurance companies, then, might want to offer tailored life insurance and car insurance plans to these neighbourhoods, taking these factors into consideration to encourage the prevention of losses. For example, insurance companies can offer lower prices plans if residents in these areas have firefighting equipment prepared in their kitchen, garages, balconies, or vehicles.

Our second question was “What factors affect the average response time, the average casualty, and the average financial losses in each neighbourhood?”. From the scatterplot, we will be able to directly and accurately answer this question. The average response time is affected by the dwelling density and the average household sizes. For instance, as the dwelling density decreases, as the average household size increases, or as the population density decreases, it will be shown that the average response time increases. This seems to suggest that neighbourhoods with larger household sizes or a lower dwelling density should have more fire stations built to support them. Also, it may benefit insurance companies to offer plans in these areas, for example, offering cheaper insurance plans for larger families who purchase firefighting equipment at home. This, in turn, acts as an incentive to keep the population in these neighbourhoods prepared in case of a fire. In addition, we can see that the average casualty is affected by the population density, the average household income, and the percentage of population without a diploma. As the population density increases, as the average household income decreases, or as the percentage of population without a diploma increases, it can be noted that the average casualties increase. This suggests that insurance companies might want to provide higher-value life insurance plans to lower-income families to encourage them to take community classes about firefighting knowledge. Besides, to help with the high population density problems associated with fires, the fire department might want to create policies regarding limited height buildings in the city (to avoid too high a population in a building) or to create stricter requirements for firefighting equipment in high-density buildings. Moreover, targeting the no-diploma population, the fire department should offer more educational resources on related topics to those populations or organize a fire department open day for adults to visit and learn everything that they need to know. Instead of investing all money in building more fire stations, investing in education and preventative safety knowledge might help save more lives. Lastly, we could see that the average financial loss is affected by the population density, the average household income, and the average household size. As the average household income or the average household size increases, or as the population density decreases, the average financial loss increases. Similarly, tailored plans should be offered to these families for the purpose of preventing further losses. Moreover, firefighting equipment should be required to be installed in the homes of all families and offered for free by the fire department, with annual indoor assessments (with fire fighting equipment) by the community or management offices in buildings.

Our project has evolved substantially from the initial proposal to our current product. Initially, our group selected a completely different topic and received some negative feedback for it, after which we decided to change the topic into what we have now: the fire incidents in the city of Toronto. Our first mockup design involved one tab only, with an interactive map and a scatterplot (of the total number of fire incidents over the years in each neighbourhood) that changed together as the cursor clicked on a neighbourhood. However, we discovered that the scatter plot was not very useful as there were not any obvious trends in any of the neighbourhoods, and the plot was not helpful with answering the analytic questions. Therefore, we decided to split the view into two, with an interactive map and an interactive scatterplot, in conjunction with an interactive table for supplementary information. The questions were evolved by adding social-economic variables into the design.

The challenges we faced included both designing the product and its resulting technical difficulties. From the perspective of designing, we received feedback detailing that the design was not effectively clear enough about how to achieve the purpose of the design or how to answer the questions. Hence, we decided to fragment the questions into multiple angles, from both the view of individual neighbourhoods and from the view of the entire city of Toronto. We attempted to allow readers to be able to gather some useful information regarding fire incidents in each of the neighbourhoods, to enable them to be educated on trends in the general city. When looking for trends, we investigate both social and economic factors, considering every conceivable choice including the average income, average household size, average level of education and so on. We hope to provide this information from multiple perspectives. From the technique perspective, we failed to have the “leaflet” map working on the shiny dashboard layout the way we anticipated due to compatibility issues. Eventually, we chose to use the shiny “absolutepanel” for loading the controls and implemented the leaflet map as background. We also implemented custom styling, using an embedded css file. However, this orientation has a limitation, and we were unable to dynamically fit other components on the page, such as infoBox or data tables. Hence, we must create another tab, “data explorer,” and put our scatterplot and data table there.

There are multiple limitations in the design. To start with, the views are visually disparate. Each view is offered in a separate tab and the readers must connect the overall information from the map to the trends on the scatterplot together by themselves. Also, the map legends were not thoughtfully planned and needed a more efficient way to visualize them. Besides, many variables from the data set were not used.

One improvement would be to put all of the map, the scatterplot, and the table together in one interactive website, and the scatterplot and table changes as the view of the map changes. Or there could be an animation, depicting fire incidents from 2011-2019 in the time series, which would connect the view together in this way as well. Also, some dots overlap on the map, which might be difficult to view on the general map. One improvement would be to represent each neighbourhood by a big dot and have the readers click into the neighbourhood to be able to see the small dots (incidents) within the neighbourhood. Lastly, many variables from the data source were not used in the design, such as sex or race-based data. One improvement to remedy this could be using APIs directly from the open Toronto data portal.

During the shiny app development, the project was split into three main parts, namely data processing, plotting and analyzing graphs, and building the shiny framework. This arrangement allowed each member to work simultaneously without engendering too many distractions:

- Yilin Huang was responsible for adding descriptive documentation in visualization, such as the text on welcome popups and other explanatory texts. She was also responsible for leading the data analysis section of our project and found almost every insight from our dataset.

- Sijie Xu is responsible for building the shiny app's primary framework and tuning the final product. He built the shiny front-end with "shinywidgets," "shinybs" and "DT," implemented the reactive components, wrote the GitHub README page, and designed our team logo.

- Yiheng Bian worked on the data cleaning and processing step. He cleaned the fire incidents datasets from "Open data Toronto," mapped to neighbourhoods, and annotated the file with population datasets from "Statistics Canada."

- Zijing Huang worked on constructing the plots. He created our interactive map with a leaflet package and the interactive scatter plots using ggplot2. He also finished the plot render functions in the shiny server and worked on debugging.

[Reference]

Eric Bailey (2015). shinyBS: Twitter Bootstrap Components for Shiny. R package version 0.61. https://CRAN.R-project.org/package=shinyBS

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Joe Cheng, Bhaskar Karambelkar and Yihui Xie (2021). leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 2.0.4.1. https://CRAN.R-project.org/package=leaflet

Victor Perrier, Fanny Meyer and David Granjon (2021). shinyWidgets: Custom Inputs Widgets for Shiny. R package version 0.6.0. https://CRAN.R-project.org/package=shinyWidgets

Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny

Statistics Canada. 2017. Toronto [Census metropolitan area], Ontario and Ontario [Province] (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E.

Sharla Gelfand (2020). opendatatoronto: Access the City of Toronto Open Data Portal. R package version 0.1.4. https://CRAN.R-project.org/package=opendatatoronto

Yihui Xie, Joe Cheng and Xianying Tan (2021). DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.18. https://CRAN.R-project.org/package=DT