Chapter 5

Student Exercise

In Chapter 2 of *The Data Journalist,* we introduced you to reporter Steve Rennie's story on parking tickets in Toronto. Through analysis of parking ticket data available for download from Toronto's open data site, Rennie was able to identify the city's most lucrative fire hydrant, at least in terms of parking tickets issued for people who parked in front of it.

You can read the story at this URL: http://globalnews.ca/news/1502051/toronto-fire-hydrants-a-cash-cow-for-the-city/

A slightly different version is available at: http://www.theglobeandmail.com/globe-drive/culture/commuting/this-fire-hydrant-cost-toronto-drivers-nearly-300000-in-parking-tickets/article19985642/

Based on your reading of the story, please answer these questions:

1. What analysis did Rennie do?
2. How did he incorporate the analysis into his reporting?
3. What other reporting did he do to provide context to his data findings?
4. What other reporting could be done to bring other perspectives to the story?
5. Why do you think the story effective? Ineffective?
6. What would you do differently?

Now, try it yourself:

The data: There are two ways you can obtain the data. You can download a zip file of tickets for fire hydrant offences from the companion site to *The Data Journalist*. These are the same offences analyzed by Steve Rennie. If you prefer, and you would like to explore other offences, you can download one or more years of data from the City of Toronto open data site (http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=ca20256c54ea4310VgnVCM1000003dd60f89RCRD). The tickets for each year are divided into several CSV files, each usually more than 50 MB.

Whichever data set you use, prior to importing the data into a database program, you may wish to open one of the files in a text editor (e.g. Notepad ++, TextEdit) or spreadsheet program to examine the table structure and the contents of each field.

Cleaning: You may also wish to clean the data to eliminate inconsistencies in the descriptions of offences, and of locations. See the online Appendix A to *The Data Journalist* for details on data cleaning. The section on cleaning with Open Refine explains how to consolidate entries that are very similar, to eliminate inconsistencies. As explained in the appendix, you can also use a spreadsheet such as Excel for the task.

To give an example, the data contains varying infraction descriptions such as PARK 3 M OF FIRE HYDRANT and PARK -3M OF FIRE HYDRANT. They are almost identical, but to the computer, they are as different as apples and oranges. Cleaning can eliminate such inconsistencies.

Failing to clean the data will not invalidate your analysis, but you will need to account for descriptions that use slightly different wording, but refer to the same type of ticket or the same place. This can be done by consolidating inconsistencies after you have finished your analysis.

You may wish to use parking ticket data from another city, obtained on an open-date site if available, or through a formal or informal request. The analysis you can do will depend on the fields provided to you.

Importing: If you are unsure how to import the data, see the companion tutorials **Making Tables and Importing Data into Access** and **Making Tables and Importing Data into MySQL**.

The analysis: Using the SQL query language covered in Chapter 5 of *The Data Journalist*, analyze the data to find:

1. The tickets are given out the most often. The least often.
2. The most ticketed places, for different types of offences.
3. How much money was raised at different locations? For different offences?
4. Any patterns for dates when tickets are given out? For example, by month. By day of the month.

Challenge 1: Create a new field and populate it with the hour of the day each ticket was given. You will have to account for different lengths of the TIME_OF_INFRACTION field. Can you find any patterns in the time of day tickets are given out.

Challenge 2: Do your own analysis different from that suggested above.

Assignment: Based on the analysis you did above, develop your own story, different from Steve Rennie's story, about parking enforcement in Toronto.