Non-Regulated Residential Lead Testing Program in Toronto*

Lead Contamination in Water

Amy Jin

September 24, 2024

The City of Toronto offers a free, non-regulated Residential Lead Testing Program, enabling residents to collect and submit water samples for lead testing. Quality of drinking water is an important public health issue. This paper examines how time and district affects the lead concentration with respect to safe lead concentration guidelines. It also highlights ongoing steps being taken to mitigate lead contamination in the water supply.

1 Introduction

Lead contamination in drinking water poses significant public health risks, particularly for vulnerable populations such as children and pregnant women. Even low levels of lead exposure can have serious health consequences, including developmental delays and cognitive impairments. In response to growing concerns about lead exposure, the City of Toronto has implemented a non-regulated Residential Lead Testing Program. This initiative allows residents to access free water testing services, providing them with the means to assess the lead content in their tap water.

The testing process is straightforward: residents can obtain a testing kit from any of six Toronto Public Health locations, collect a water sample from their home, and return it for analysis. Toronto Water processes the samples and provides results directly to the residents. Although the City cannot control how the samples are collected, this service empowers residents to take action in ensuring the safety of their drinking water. This paper will explore the benefits of the program, its limitations, and the broader efforts undertaken by the City of Toronto to address lead contamination in its water supply system.

We use R Core Team (2023), Gelfand (2022), and Wickham et al. (2019).

*Code and data are available at: https://github.com/aj3616/Lead-Testing-Program-in-Toronto

The remainder of this paper is structured as follows. Section 2

2 Data

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.000050 0.000125 0.000473 0.006401 0.001480 8.440000
```

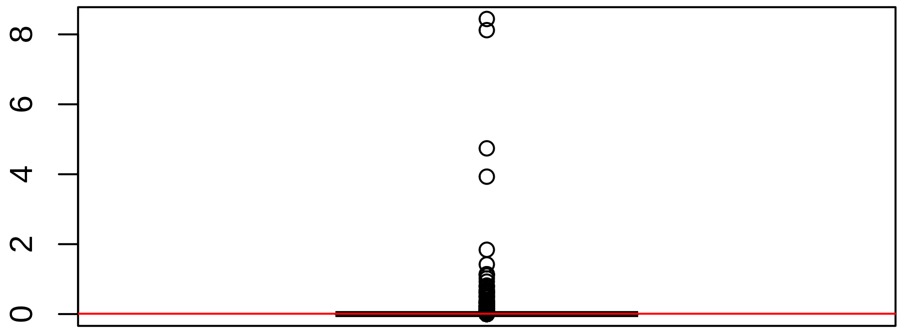

Figure 1: Figure 1: summary of the lead amount in ppm

```
[1] 0.9725893
```

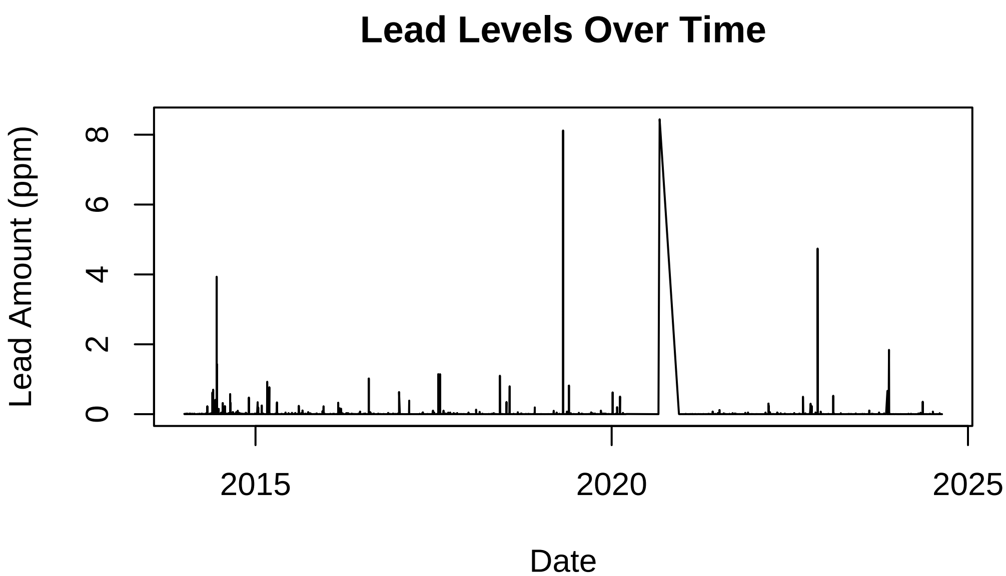

Figure 2: Figure 2: Lead Levels changed over time for times seriese analysis

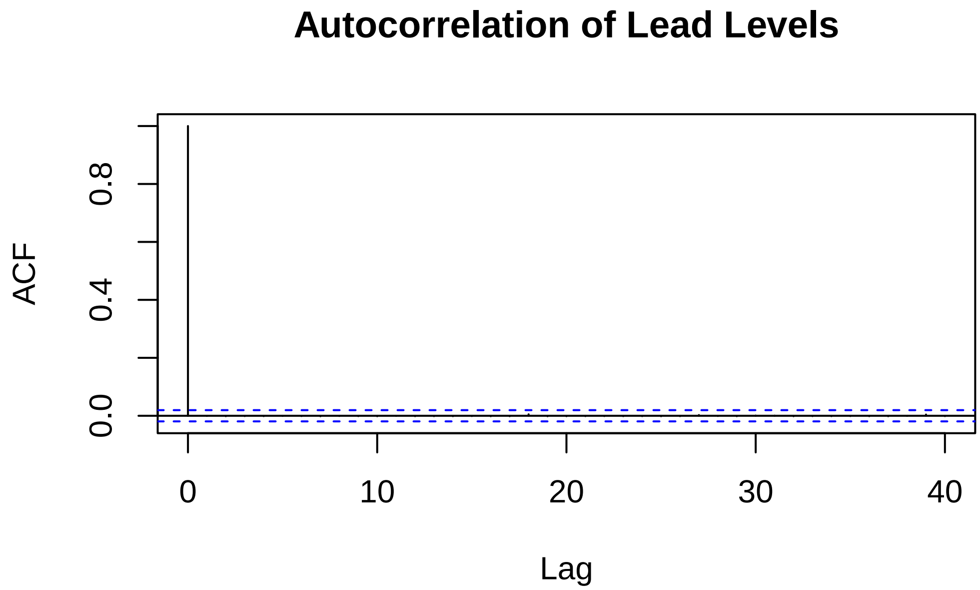

Figure 3: Figure 3: Autocorrelation function for lead concentrations

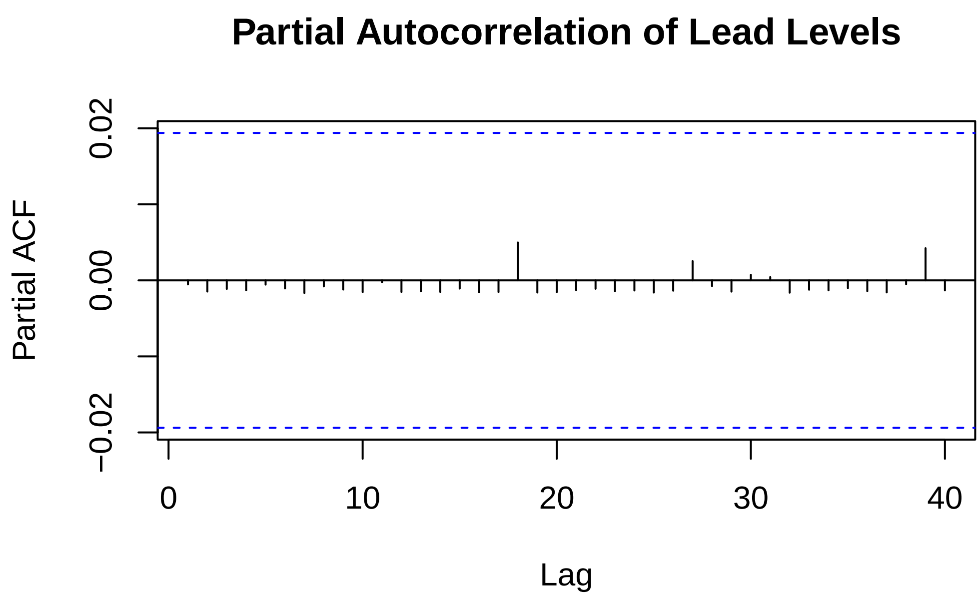

Figure 4: Figure 4: Partial Autocorrelation function for lead concentrations

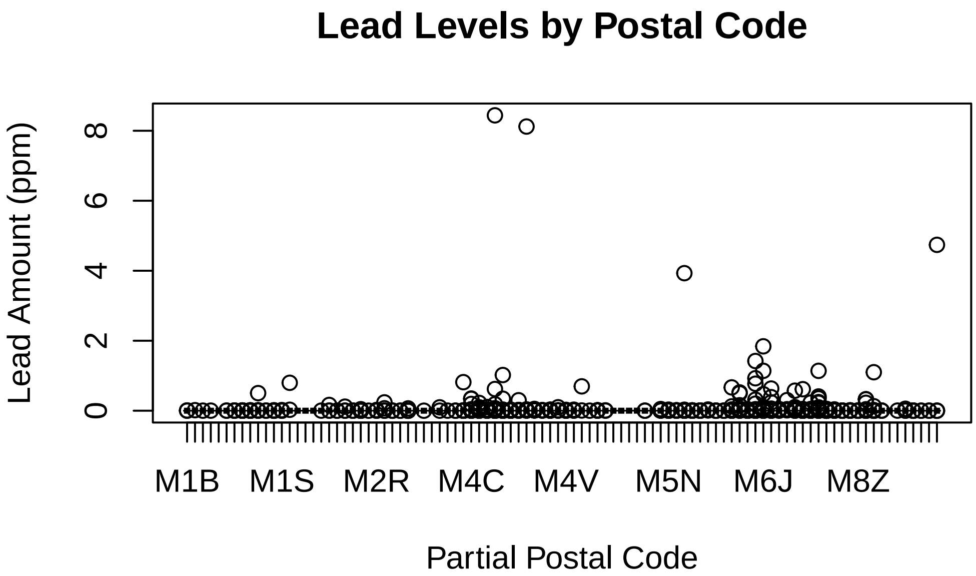

Figure 5: Figure 5: Outliers categorized by district

	District	Lead Amount (ppm)
1	M1	0.003428667
2	M2	0.001996157
3	M3	0.003143738
4	M4	0.003040474
5	M5	0.005496381
6	M6	0.005842934
7	M8	0.000753639
8	M9	0.007450080

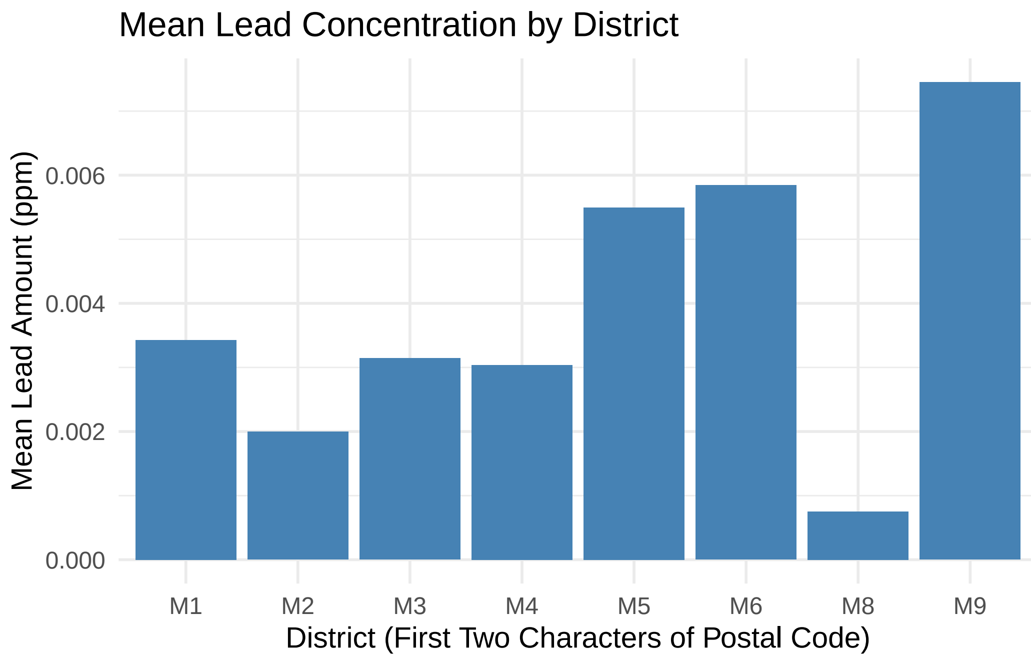

Figure 6: Figure 6: Mean lead concentration categorized by first two characters in partial postal code

```r
threshold <- 0.010
high_lead_samples <- data[data$`Lead Amount (ppm)` > threshold, ]

# Count of samples above the threshold
nrow(high_lead_samples)
```

```
[1] 280
```

```r
# Percentage of samples above threshold
(nrow(high_lead_samples) / nrow(data)) * 100
```

```
[1] 2.741067
```

3 Discussion

3.1 Lead Concentration influencing Health

- WHO and Canadian federal guidance on lead concentration in water

Lead amount The main variable we are examining is the Lead Amount in ppm. The distribution of the data is highly right skewed with most of its data being very small, as shown in figure 1((**figure1boxplot?**)), most data are almost 0 with some outliers which are mostly less than 2 ppm and 4 of them ranging from 3 to 8.44 which is the maximum value as shown in the summary table((**table1summary?**)). We can also verify our observation since the summary((**table1summary?**)) shows that 75% of the data is below 0.001480. Government of Canada had specified that the MAC of drinking water is 0.010mg/L(Government Canada), which converts to 0.010011423 part/million(unitconverters.net), this is shown by the red line in the figure 1((**figure1boxplot?**)).

- percentage of data above this threshold

From calculations, there are 97.26% of the data collected indicatin that the lead amount in the water is safe, and it is safe to assume that 75% of the areas in toronto is safe of lead in water.

- plot of the data above this threshold

3.2 How districts affect Lead Concentration

- graphs of postal code vs lead concentrations

Postal Code Figure 5((**figure5?**)) is a scatterplot of each specific partial postal code(first 3 code), we can see that again, most data points are around 0, but there are a few outliers with very high lead concentration. Figure 5 clarily indicated that which district are those outliers in. We can see that the "M4" has two datapoints aroun 8 ppm, as well as some between 0 and 2ppm. "M6" also have many data between 0 and 2ppm. It is safe to assume that those two district might have a higher overall lead concentration level.

- extract outliers above the threshold

We will verify the assumption with calculating the mean lead concentration when categorized the partial postal codes by only the first two numbers. Since outliers might influence the mean, especially when in our dataset we are dealing with median 0.000473 and outliers as high as 8ppm, we can omit the outliers above 4ppm and perform the data analysis.

- mean value of lead concentration for different districts

Figure 6((**figure6?**)) is the bargraph comparing the mean values of lead concentration without outliers above 4ppm and cetegorized by only the first two characters of the partial postal code. We can see that the top 3 district of the highest lead concentration in water is M9, M6 and M5 respectively. This shows that without the data above 4ppm,

- toronto map

3.3 How months and year affect Lead Concentration

- times series analysis for autocorrelation

Time and Lead In this dataset, we have variables time and lead amount, therefore, we can perform times series analysis(STA457). As shown in figure 2((**figure2ggplot?**)), we can see that except for some sharp increase in 2014, 2019, 2021, 2023, the overall trend is stays around 0 and remains positive.

From the ACF and PACF figures, we can conclude that the previous values of lead concentrations does not impact future values. For ACF, when lag is 0, ACF is always 1 but for all other lags in ACF((**figure3ACF?**)) and PACF((**figure4PACF?**)) they are all within the dotted blue lines which is the evidence that there are no significant autocorrelation nor autoregressive realtionship present in the data.

This observation is reasonable because the samples are collected in different district in Toronto, and it is safe to assume that one incident of a very high lead concentration might again only be outliers instead of some factors that will effect the overall water quality in Toronto.

- graphs of months vs mean lead concentration
- graphs of years vs mean lead concentration
- extract outliers above the threshold

3.4 Weaknesses and next steps

- the data is from many different water source, they are not necessarily direct drinking water since how the water sample collected was not controled. However, the guides are for drinking water.
- importance of pre-treatment such as filter

Appendix

A Additional data details

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.