

# Apartment building data analysis\*

Wanghanyue Fang

September 24, 2024

Apartment data analysis

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview . . . . .	1
1.2	Results . . . . .	1
<b>A</b>	<b>Appendix</b>	<b>8</b>
A.1	Dataset and Graph Sketches . . . . .	8
A.2	Data Cleaning . . . . .	8
A.3	Attribution Statement . . . . .	8
	<b>References</b>	<b>9</b>

## 1 Introduction

] # Data {#sec-data}

### 1.1 Overview

### 1.2 Results

```
# A tibble: 1 x 11
  title          id    topics civic_issues publisher excerpt dataset_category
```

---

\*A GitHub Repository containing all data, R code, and other files used in this investigation is located here:  
<https://github.com/bennyrochweg/healthcare-outbreaks-toronto>

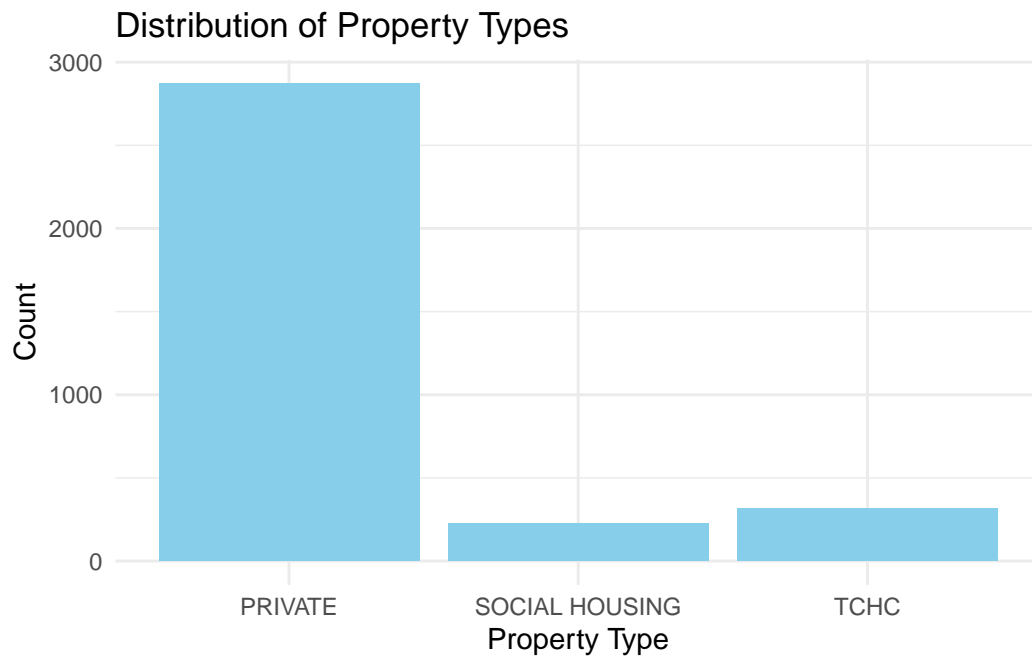
```

      <chr>          <chr> <chr>  <chr>          <chr>      <chr>  <chr>
1 Apartment Buildi~ 4ef8~ City ~ <NA>          Municipa~ This d~ Table
# i 4 more variables: num_resources <int>, formats <chr>, refresh_rate <chr>,
#   last_refreshed <date>

# A tibble: 8 x 4
  name                                     id      format last_modified
  <chr>                                <chr>   <chr>   <date>
1 Apartment Building Evaluations 2023 - current.csv 7fa98~ CSV    2024-09-24
2 Apartment Building Evaluations 2023 - current.xml 73e0b~ XML    2024-09-24
3 Apartment Building Evaluations 2023 - current.json 34a7d~ JSON   2024-09-24
4 Pre-2023 Apartment Building Evaluations            b987b~ CSV    2023-07-26
5 Pre-2023 Apartment Building Evaluations.csv        979fb~ CSV    2023-07-31
6 Pre-2023 Apartment Building Evaluations.xml        c8672~ XML    2023-07-31
7 Pre-2023 Apartment Building Evaluations.json       e5b03~ JSON   2023-07-31
8 Apartment Building Evaluations 2023 - current      244f7~ CSV    NA

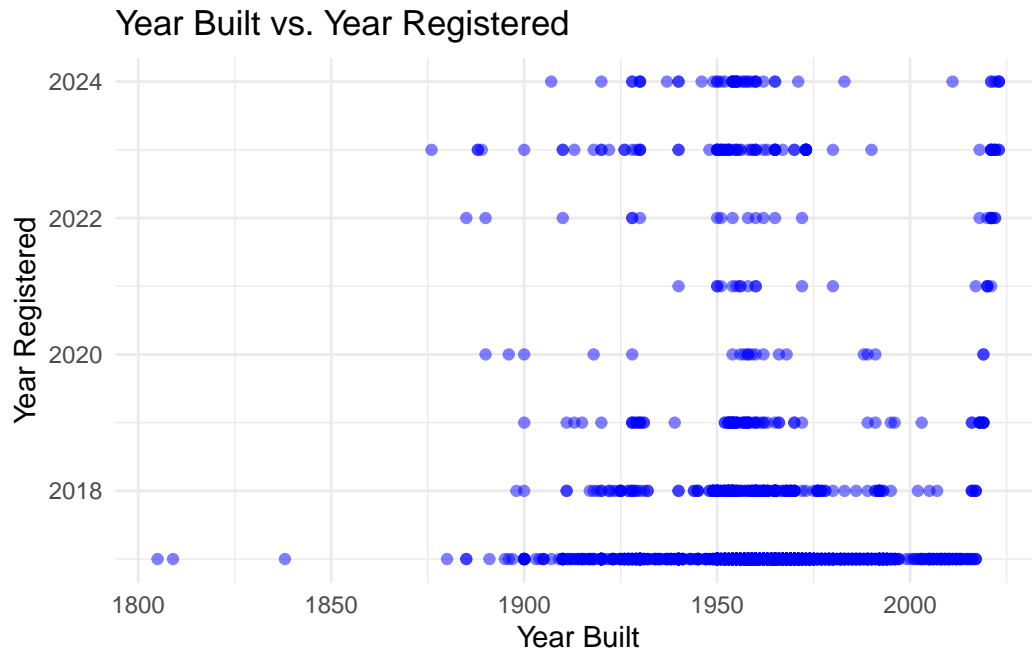
```

After loading the dataset using the R programming language (R Core Team 2022) and the `here` package (Müller 2020), the `tidyverse` (Wickham et al. 2019) package was used to generate graphs. In doing so, R code was adapted from Alexander (2023).



```
# 2. Year of Registration vs. Year Built
ggplot(data, aes(x = `YEAR.BUILT`, y = `YEAR.REGISTERED`)) +
  geom_point(alpha = 0.5, color = "blue") +
  theme_minimal() +
  labs(title = "Year Built vs. Year Registered", x = "Year Built", y = "Year Registered")
```

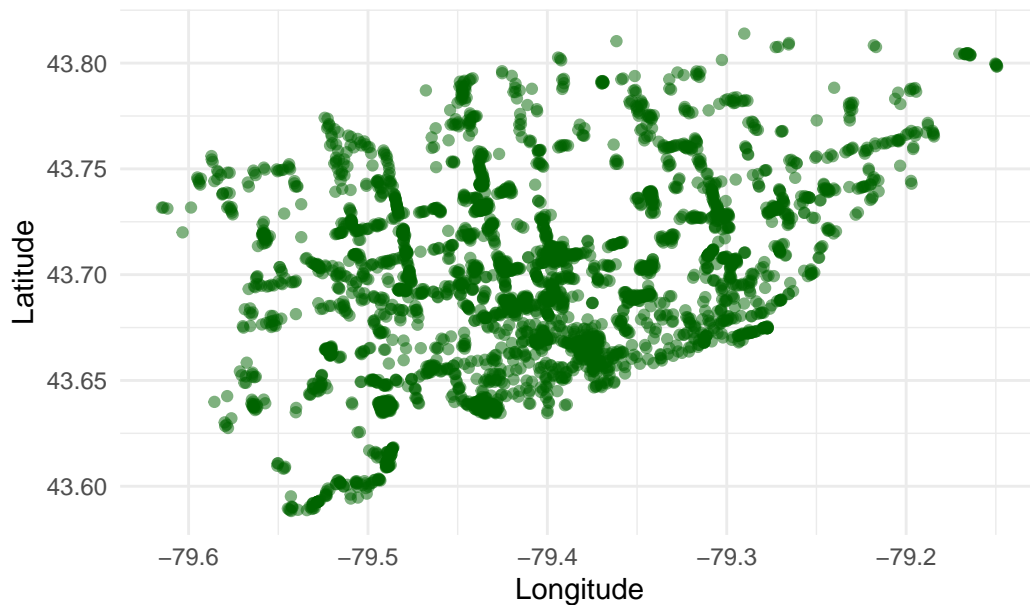
Warning: Removed 57 rows containing missing values or values outside the scale range (``geom_point()``).



```
# 3. Geographic Plot (Latitude vs Longitude)
ggplot(data, aes(x = LONGITUDE, y = LATITUDE)) +
  geom_point(alpha = 0.5, color = "darkgreen") +
  theme_minimal() +
  labs(title = "Geographic Distribution of Properties", x = "Longitude", y = "Latitude")
```

Warning: Removed 97 rows containing missing values or values outside the scale range (`geom\_point()`).

## Geographic Distribution of Properties

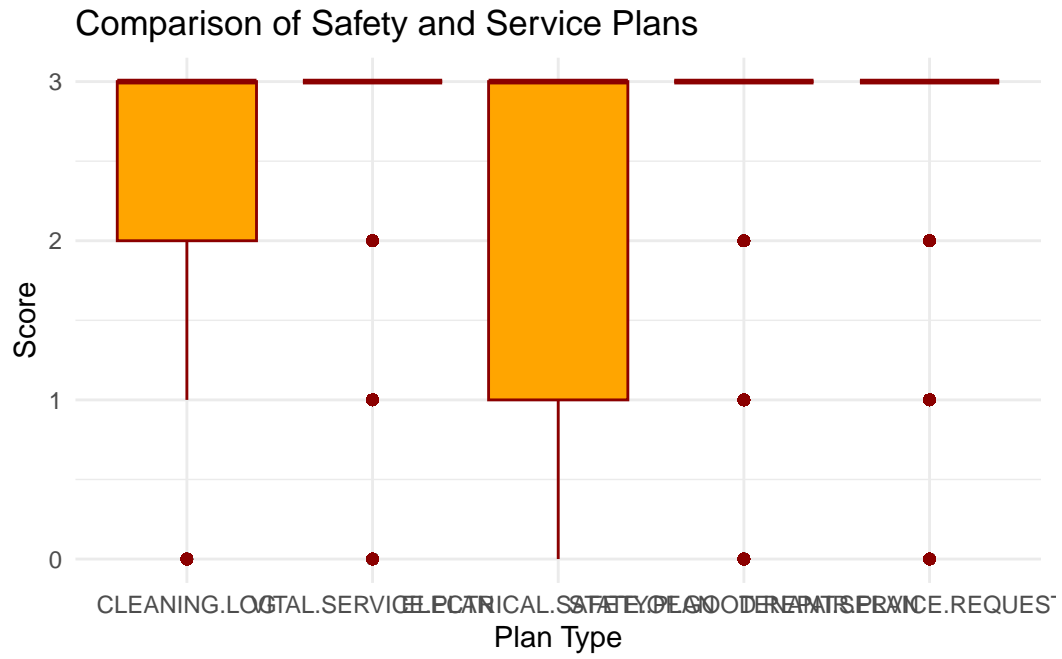


```
# 4. Safety and Service Plan Scores Boxplots
# Combine all safety and service-related columns for comparison
safety_service_columns <- data %>%
  select(CLEANING.LOG, VITAL.SERVICE.PLAN, ELECTRICAL.SAFETY.PLAN,
         STATE.OF.GOOD.REPAIR.PLAN, TENANT.SERVICE.REQUEST.LOG)

safety_service_melt <- melt(safety_service_columns)
```

No id variables; using all as measure variables

```
ggplot(safety_service_melt, aes(x = variable, y = value)) +
  geom_boxplot(fill = "orange", color = "darkred") +
  theme_minimal() +
  labs(title = "Comparison of Safety and Service Plans", x = "Plan Type", y = "Score")
```



“

## **A Appendix**

### **A.1 Dataset and Graph Sketches**

Sketches depicting both the desired dataset and the graphs generated in this analysis are available in the GitHub Repository.

### **A.2 Data Cleaning**

The data cleaning process involved filtering out some of the columns from the raw dataset and renaming some of the data entries for clarity and simplicity.

### **A.3 Attribution Statement**

“Contains information licensed under the Open Government Licence – Toronto” (City of Toronto, n.d.).



## References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. <https://tellingstorieswithdata.com/>.
- City of Toronto. n.d. “Open Data License.” <https://open.toronto.ca/open-data-license/>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2022. *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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.