Assignment 01 - Exploratory Data Analysis

Due Date

This assignment is due at the end of the day January 31, 2024

Learning Goals

- Download, read, and get familiar with an external dataset.
- Step through the EDA "checklist" presented in class
- Practice making exploratory plots

Assignment Description

We will work with Toronto Police Department Public Safety Data Portal, in particular, about break and enters.

The primary question you will answer with these data is: *When and where do most break and entries occur in Toronto?*

Download the data from here.

Your assignment should be completed in RMarkdown.

Steps

Given the formulated question from the assignment description, you will now conduct EDA Checklist items 2-4.

1. First, read in the data. Update the missing data identifiers to `NA`. Check for import issues (dimensions, headers, footers, variable names and variable types). Check for any data issues (import issues, missing values, data errors) particularly in the key variable we are analyzing. Make sure you write up a summary of all of your findings.

2. Clean the data – keep only necessary data columns and change the names of the key variables so that they are easier to identify. Change the type of key variables from string to factor as appropriate. Identify any outliers, and justify how you handle them.

3. Create a variable for season, and calculate summary statistics for each season and then conduct some basic analyses that enable you to compare across winter/spring/summer/fall (e.g. chi-square test, linear regression/anova). Be sure to show the results and write up explanations of what you observe in these data.

4. Create exploratory plots of which neighbourhoods, seasons, and time of day break and enters occurred.

5. Create a basic map of the locations of the break and enters, and then one in leaflet() that shows the counts by neighborhood.