```
In [1]:  %load_ext autoreload
         %autoreload 2
         import pandas as pd
         import geopandas as gpd

         # import utility functions
         import sys
         sys.path.append('../')
         from src.utils import *
```

```
In [2]:  %matplotlib inline
         import matplotlib as mpl
         mpl.rcParams['figure.dpi']= 300
```

```
In [3]:  df_2022 = get_data(2022)
         df_2023 = get_data(2023)
```

```
Breakdown by Outlier Condition:
        Outlier Rents: 6335 (34%)
        Outlier Increase vs Base: 777 ( 4%)
        Outlier Increase vs Previous: 609 ( 3%)
        Overall: 6682 (36%)

Breakdown by Subset:
        6682 outliers (36%)
        11890 non-outliers (64%)


        4142 rent increase (22%)
        14430 no rent increase (78%)


        7036 exempt (38%)
        11536 not exempt (62%)
Breakdown by Outlier Condition:
        Outlier Rents: 5461 (30%)
        Outlier Increase vs Base: 608 ( 3%)
        Outlier Increase vs Previous: 328 ( 2%)
        Overall: 5905 (33%)

Breakdown by Subset:
        5905 outliers (33%)
        12081 non-outliers (67%)


        8567 rent increase (48%)
        9419 no rent increase (52%)


        7146 exempt (40%)
        10840 not exempt (60%)
```

```
In [4]:  print(f"2022 Registrations: {df_2022.shape[0]}")
         print(f"2023 Registrations: {df_2023.shape[0]}")
```

```
2022 Registrations: 18572
2023 Registrations: 17986
```

```
In [5]:  # count number of License Numbers and Parcels present in both years
         print(f"License Numbers in both years: {len(set(df_2022['LICENSENUMBER']) & set(df_2023[
         print(f"Parcels in both years: {len(set(df_2022['PARCELNUMBER']) & set(df_2023['PARCELNU
```

```
License Numbers in both years: 3868
Parcels in both years: 3974
```