

Statistics Data Project: Eviction Notices in SF

Brian Liu

21 May 2023

1 Introduction

One of the biggest problems that is beginning to plague America is affordable housing, and San Francisco is a notable city with significant history about the struggle between tenants and landlords with housing. It not only has a known problem with unaffordable housing from the tech booms bringing significant wage gaps, but it also has a rich amount of data telling the stories for hundreds of thousands tenants that have been evicted from their homes. In this paper, I will explore the eviction notices published by SF Rent Board in San Francisco. I will look at the different types of evictions, the neighborhoods that are most affected, and the time periods that have the most evictions. I will also look at the different types of evictions and how they change over time.

1.1 Data

The data is made up of roughly 222,000 eviction notices from 1997 to 2023. The most important variables that I analyzed are the eviction filing date (FD), longitude (LG), latitude (LT), and eviction type (ET), which includes 18 different types of evictions. I will introduce ETs as they come in this paper.

2 Exploratory Data Analysis

2.1 Mapping Evictions

I first plotted the evictions onto a map of San Francisco by year (FD), to both get a sense of the changes in frequencies of evictions and the actual locations of evictions. The density of evictions in the late 1990's was much higher than subsequent years and they were mostly concentrated in the Mission District and Tenderloin, both of which are neighborhoods that suffered from historic redlining (a discriminatory practice that denies services such as a loan or mortgages to minority communities marked as inadequate or risks).

The effect of redlining and historic discriminatory practices is highlighted in this graph, and the evictions continue staying concentrated in those areas onward past the 1990s. One of the devastating consequences of these historic practices is that there is a disproportionate amount of minority groups that are evicted because they have effectively been pushed into these neighborhoods.

One other thing to note is that the periodic clusters of eviction notices in the bottom left corner in the neighborhood, Parkmerced. This large, single-owner neighborhood complex has been under scrutiny for sudden evictions as it prepares for new development projects and renovations.

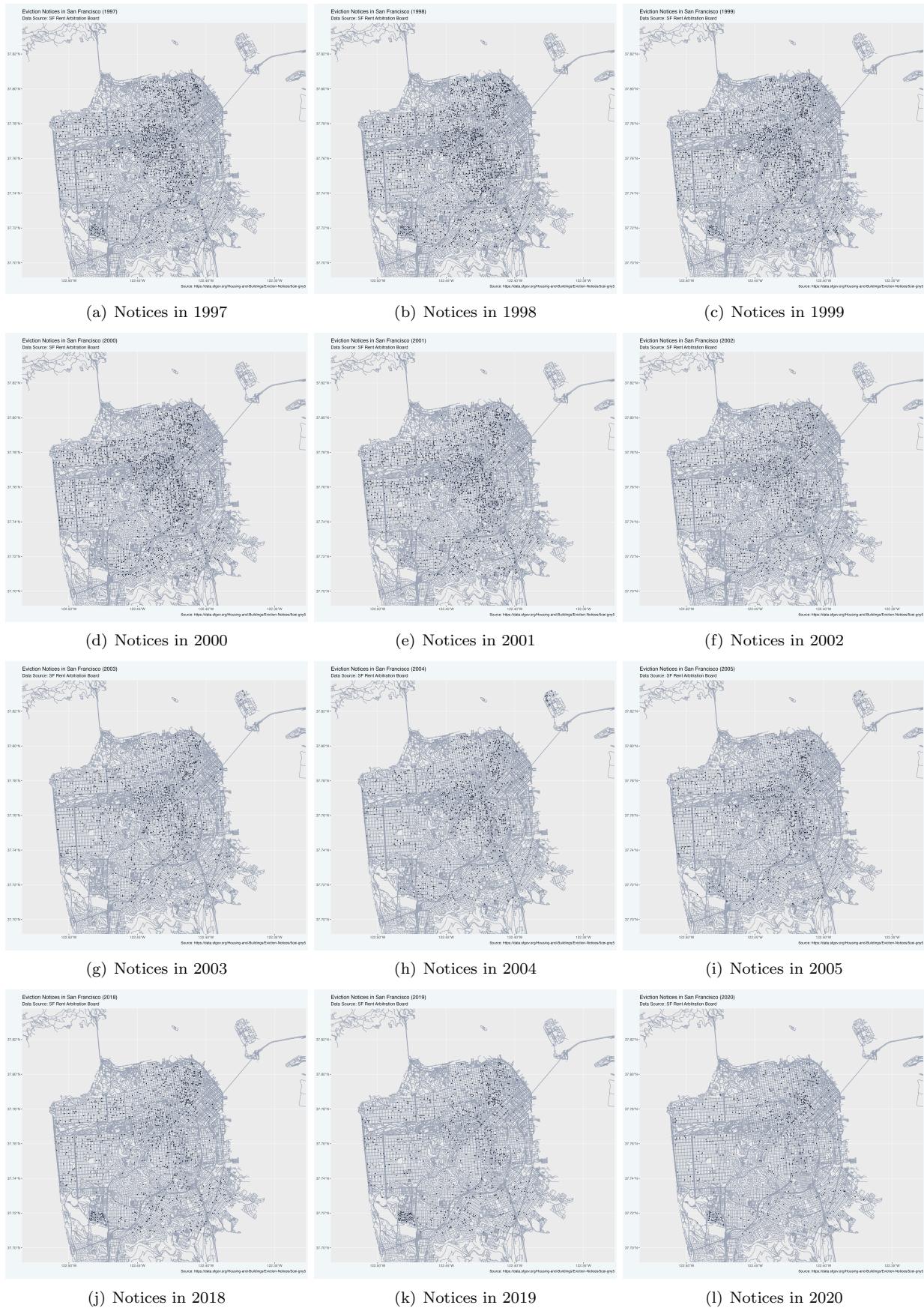


Figure 1: Eviction notices by year (1997-2005, 2018-2020)

2.2 Seasonal Changes

Organizing evictions by seasons using FD allows us to see if there are any seasonal trends in evictions. Winter evictions are significantly lower than the other seasons. I have not found a solid reason from research but it can be a combination of increased rent protections.

Evictions by season:

1. Spring: 45872
2. Summer: 47528
3. Fall: 44048
4. Winter: 12126

2.3 Plotting Counts by Year and Month

Organizing the frequencies of counts by year and months allows for a clear picture on how evictions have been trending. There are two distinct spikes: the period around the "dot-com boom/bubble burst" and the beginning of the 2010s, when the real estate market was growing. It is interesting to note that during the "Internet boom" (1998-99), evictions were at the all time high in this data set because of a frenzy to capture the "'new money' that arrived during the 1990s. The fabled boom-time economy, fueled by the dot.com explosion, cannibalized the housing stock and the last blue-collar jobs."

The second spike in the early 2010s is a result of the real estate market growing after the 2008 recession. One interesting finding I found as a result of research into this is that landlords seek to evict more people when they know they can profit the most, which would mean a growing market.

For the plot of counts by month, I added a LOESS regression curve to show a general trend of evictions. I used GGPlot2's loess curve from the *geom_smooth* function.

2.4 Zooming into the Spikes: Owner Move-In and Ellis Act Evictions

I then looked at two common ETs: owner move-in (OMI) and Ellis Act evictions (EAW). Owner move-in evictions are when the landlord evicts the tenant to move themself or relatives into the unit. Ellis Act Withdrawal evictions are when the landlord evicts the tenant to remove the unit from the rental market. I plotted the evictions by year and by neighborhood to see if there were any trends in the data.

OMI evictions make up a significant amount of the evictions, and it peaks in the late 1990's before significantly decreasing starting in 1998, when EAW evictions began increasing shortly afterwards. This trend comes from the abuse of OMI evictions as a way to clear multifamily buildings, and it was rampantly used until a reform effort passed by Sue Bierman to restrict the number of OMI evictions in a building. EAW evictions then began rising as a result of the tech bubble burst and maintained consistent even after that period. Tied with the mapping of evictions above, the neighborhoods that are most affected by OMI and EAW evictions are the Mission District and the Tenderloin.

3 A Great Recession, a Tech Boom, and a Pandemic

The time period between 2009-2023 is interesting because while the LOESS curve generally followed the evictions before well, it overpredicts the evictions at the start of this time period, underpredicts the evictions in the middle, and does a mix of both at the end. I now bound the scope of this paper to this time period to explore the changes in eviction types and strategies.

3.1 Using a Simple Model to Predict Proportions of ETs

I used a binomial distribution as a model to calculate the expected proportion of each ET.

Let $R_{t,i}$ be the number of evictions with type t in time period i and n_i be the total number of evictions in time period i . Suppose that $R_{t,i} \sim \text{Binomial}(n_i, p_{t,i})$ where $p_{t,i}$ is the proportion of eviction type t in time period i . We can construct a likelihood function and find the maximum likelihood equation (MLE) for $p_{t,i}$.

$$\begin{aligned}
\mathcal{L}(p) &= \prod_{t,i} \text{Binomial}(R_{t,i}|n_i, p_{t,i}) \\
\mathcal{L}(p) &= \prod_{t,i} \binom{n_i}{R_{t,i}} p^{R_{t,i}} (1-p)^{n_i - R_{t,i}} \\
\ln \mathcal{L}(p) &= \sum_{t,i} \left(\ln \binom{n_i}{R_{t,i}} + R_{t,i} \ln p + (n_i - R_{t,i}) \ln(1-p) \right) \\
\frac{d}{dp} \ell(p) &= \sum_{i=1} \frac{R_{t,i}}{p} - \frac{n_i - R_{t,i}}{1-p} \\
0 &= \sum_{i=1} \frac{R_{t,i}}{p} - \frac{n_i - R_{t,i}}{1-p} \\
0 &= \sum_{i=1} R_{t,i} - p_{t,i} \sum_{i=1} n_i \\
p &= \frac{\sum_{t,i} R_{t,i}}{\sum_{t,i} n_i}
\end{aligned}$$

Using this derived MLE for \hat{p} I can calculate the expected proportion for each ET:

```

# Calculate the expected proportion for each eviction type
calc_proportions <- function(data) {
  data %>%
    count(eviction_type) %>%
    mutate(prop = n / sum_evictions_09_23) %>%
    arrange(eviction_type)
}

# Pass to function then bind data_id which is the period
e_prop <- lapply(evictions_09_23, calc_proportions)
e_prop <- bind_rows(e_prop_tp, .id = "data_id")

expected_prop <- sum(e_prop_tp %>% pull(prop)) / 48

```

Where e_prop represents the subsetted data set

$$\hat{p} = \frac{\sum_{t,i} R_{t,i}}{\sum_{t,i} n_i} = 0.02083333$$

Plotting this expected value with the observed proportions, we can see that the observed proportions do not agree with the expected proportions almost all of the time; there is a lot of unexplained variation in the data.

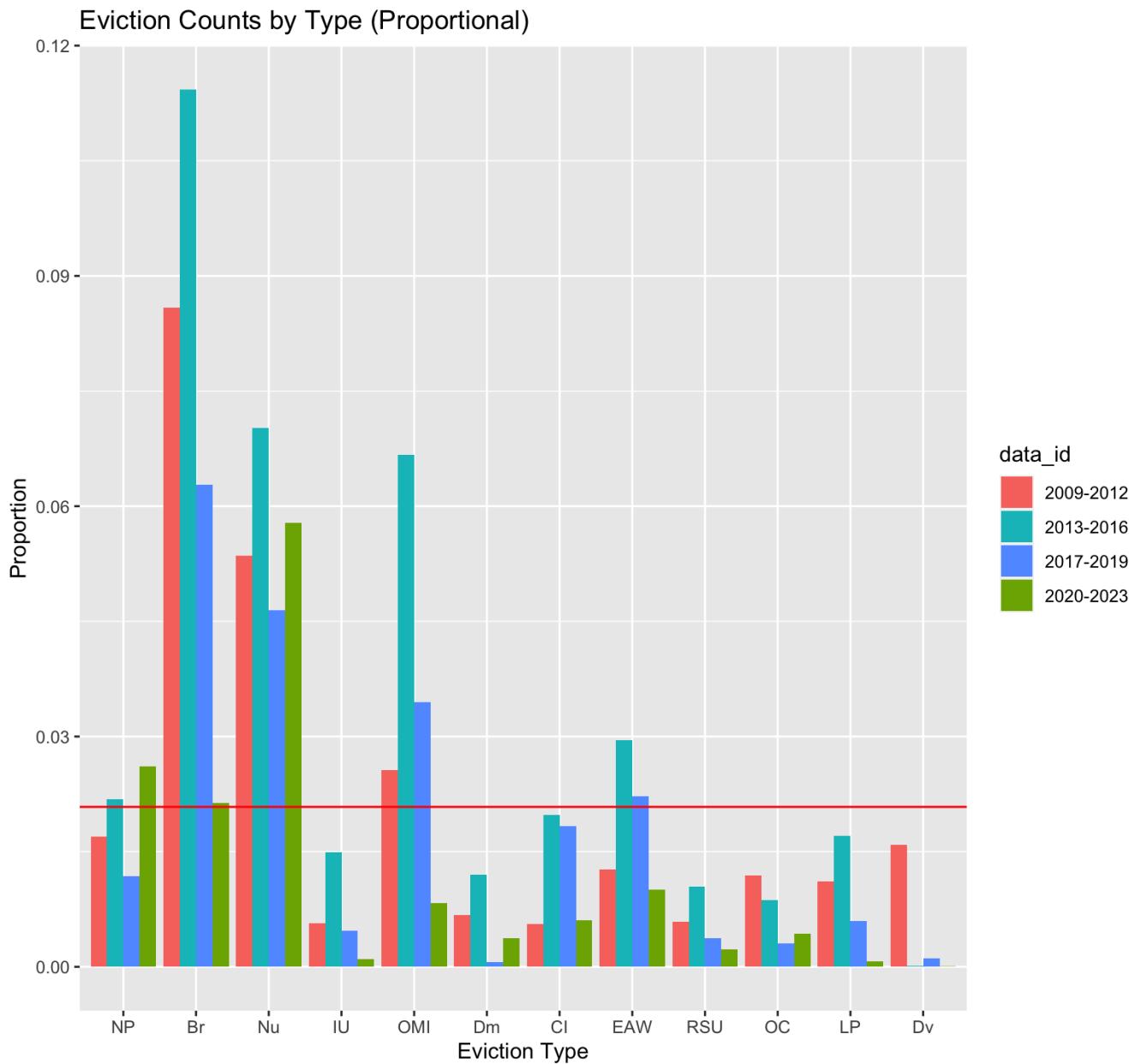


Figure 2: Proportions of ETs with Expected Value (2009-23)

3.2 Examining each ET

I then normalized the proportions of each ET to sum to 1:

```
e_prop_by_et <- e_prop %>%
  group_by(eviction_type) %>%
  summarise(prop = sum(prop))
```

and then plotted the proportions of each ET by time period. The data is then bootstrapped, resampling with replacement 10000 times. I then calculated the 99% confidence intervals for each ET and plotted all of this on a graph.

```
size <- 10000

get_bs <- function(size, type, data) {
  # 4 by size matrix, 4 rows for each time period
  bs_matrix <- array(NA, dim = c(4, size))

  for (i in 1:size) {
    bootstrap_indices <- sample(seq_len(nrow(data)), nrow(data), replace = TRUE)
    bootstrap_sample <- data[bootstrap_indices, ]

    bs <- bootstrap_sample %>%
      filter(eviction_type == type)

    e_09_12_bs <- bs[bs$file.Date >= as.Date("2009-01-01") &
      bs$file.Date <= as.Date("2012-12-31"), ]
    e_13_16_bs <- bs[bs$file.Date >= as.Date("2013-01-01") &
      bs$file.Date <= as.Date("2016-12-31"), ]
    e_17_19_bs <- bs[bs$file.Date >= as.Date("2017-01-01") &
      bs$file.Date <= as.Date("2019-12-31"), ]
    e_20_23_bs <- bs[bs$file.Date >= as.Date("2020-01-01") &
      bs$file.Date <= as.Date("2023-12-31"), ]

    prop_e_09_12_bs <- nrow(e_09_12_bs) / nrow(bs)
    prop_e_13_16_bs <- nrow(e_13_16_bs) / nrow(bs)
    prop_e_17_19_bs <- nrow(e_17_19_bs) / nrow(bs)
    prop_e_20_23_bs <- nrow(e_20_23_bs) / nrow(bs)

    bs_matrix[1, i] <- prop_e_09_12_bs
    bs_matrix[2, i] <- prop_e_13_16_bs
    bs_matrix[3, i] <- prop_e_17_19_bs
    bs_matrix[4, i] <- prop_e_20_23_bs
  }
  return(bs_matrix)
}

bs_matrices <- list()
for (i in 1:12) {
  bs_matrices[[i]] <- get_bs(size, eviction_type[i], e_09_23)
}

# calculate 95% confidence interval for each eviction type
ci_bs_matrices <- list()
for (i in 1:12) {
  ci_bs_matrices[[i]] <- array(NA, dim = c(4, 2))
  for (j in 1:4) {
    ci_bs_matrices[[i]][j, 1] <- quantile(bs_matrices[[i]][j, ], 0.005)
    ci_bs_matrices[[i]][j, 2] <- quantile(bs_matrices[[i]][j, ], 0.995)
```

}

I set out to use these confidence intervals to find if there are any time periods where the observed expected proportion of a given ET is not included in those confidence intervals. Plotting all of the ETs with its proportions per time period, the confidence intervals, and the expected proportion of each ET, almost all of the time periods do not include the expected proportion of each ET.

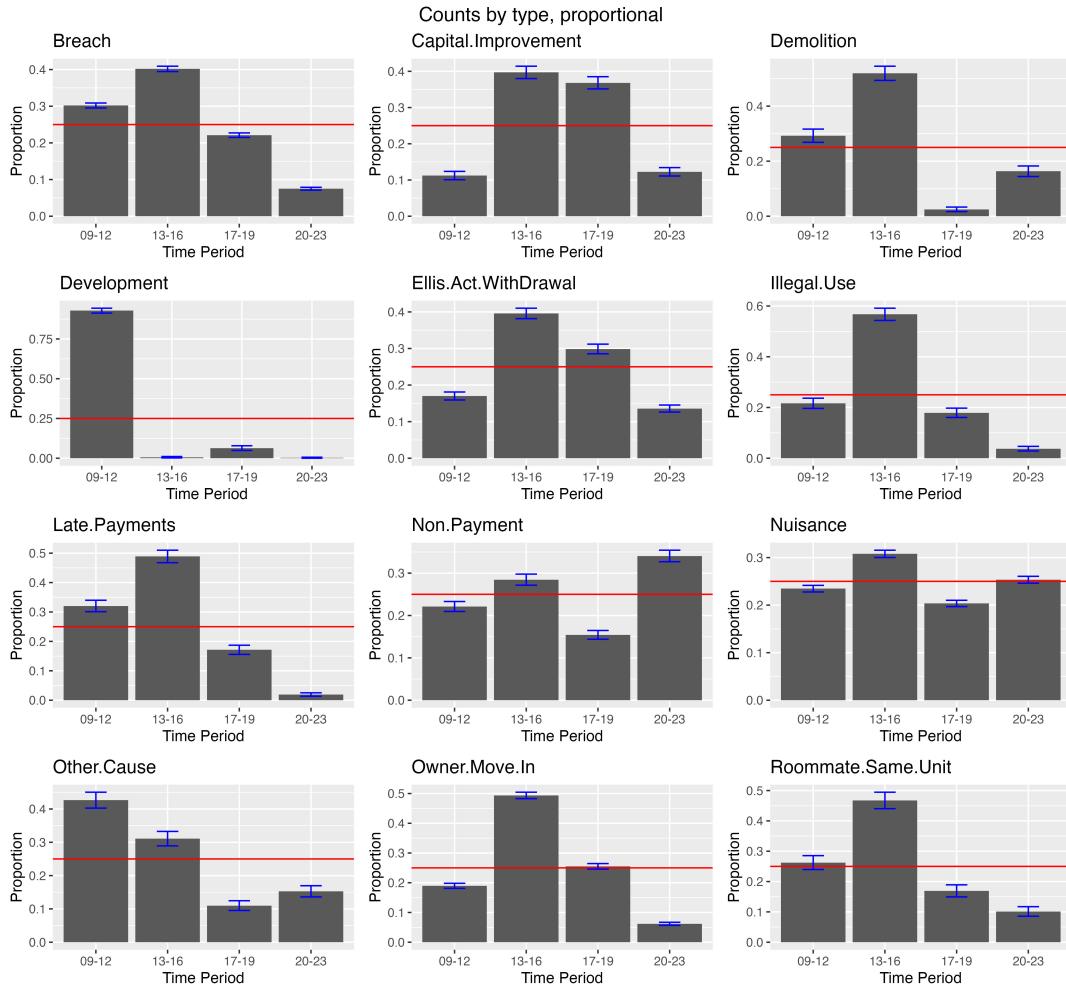


Figure 3: Evictions by Type and Time Period (2009-23) with Expected Proportion and 99% Confidence Intervals

3.3 Calculating z -scores

More importantly though, some periods have a greater deviation from the expected proportion than others. To be able to visualize and quantify this, I calculated the z -score for each ET in each time period.

$$z_{t,i} = \frac{\hat{p}_t - p_{t,i}}{\sigma_{t,i}}$$

where \hat{p}_t is the expected proportion of ET t and $p_{t,i}$ is the observed proportion of ET t in time period i . $\sigma_{t,i}$ is the standard deviation of the bootstrapped samples of ET t in time period i .

```
signed_z_scores <- lapply(1:12, function(i) {
  mean_prop <- get_mean_prop(eviction_type[i])
  ci_low <- ci_bs_matrices[[i]][, 1]
```

```

ci_high <- ci_bs_matrices[[i]][, 2]
z_scores <- numeric(4)
for (j in 1:4) {
  if (mean_prop > ci_high[j]) {
    z_scores[j] <- (ci_high[j] - mean_prop) / sd(bs_matrices[[i]][j, ])
  } else if (mean_prop < ci_low[j]) {
    z_scores[j] <- (ci_low[j] - mean_prop) / sd(bs_matrices[[i]][j, ])
  } else {
    z_scores[j] <- 0
  }
}
return(z_scores)
})

```

Plotting the z -scores allows us to see significant and varying amounts of variation for ETs and time periods. I graphed and organized them by ET and time period. These graphs allow me to perform visual hypothesis testing and find where there are significant deviations from the expected proportion. I note some observations below:

1. Development evictions are significantly higher in the 2009-12 period than any other time period. There may be a correlation with the recovery from the Great Recession and the increase in development evictions.
2. OMI evictions are significantly higher in the 2013-16 period than any other time period. This may be a result of the tech boom in the early 2010s.
3. 2017-19 is the time period with the lesser deviation and an overall decrease in evictions. This may be a result of the local real estate market stumbling during that time.
4. 2020-23 is the time period with the most deviation and decrease in evictions. This is most definitely a result of the pandemic.

4 Conclusion

It is not like San Francisco's population has varied as significantly as the eviction rates have. From this data we can see that eviction comes in waves, whether that be from seasons, the real estate market, a pandemic, a tech boom, or any other factor that affects the balance of capitalism. With each of these events comes different strategies landlords use to evict tenants.

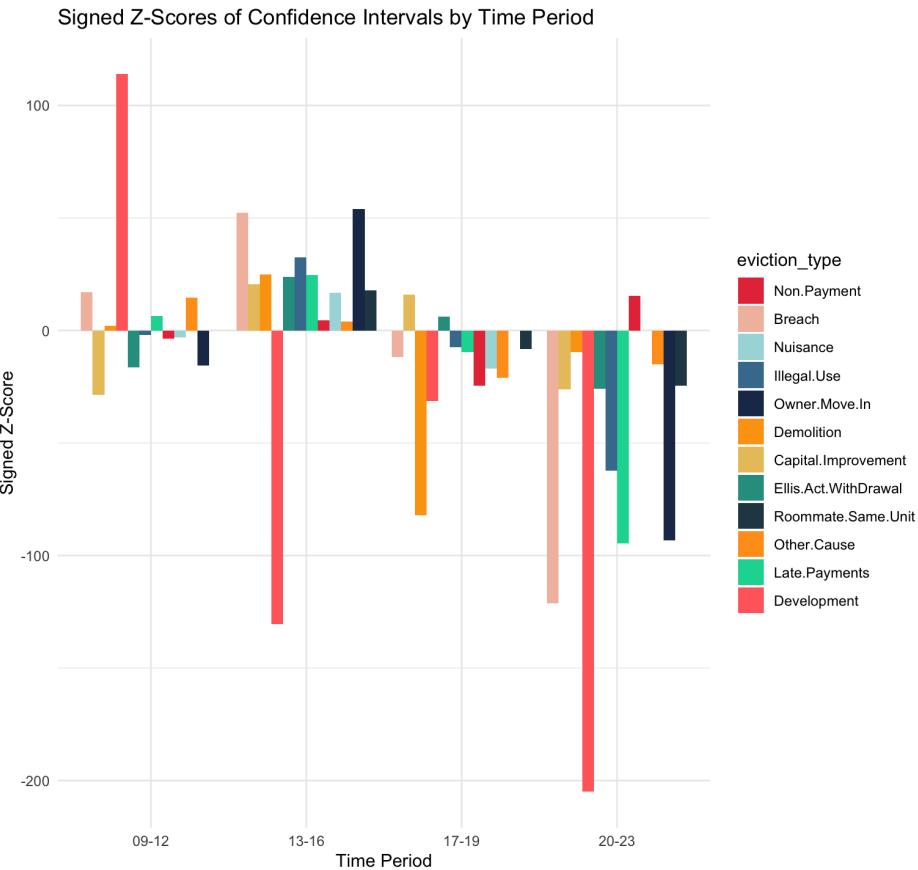
My exploration of the subsetted data from 2009-23 shows that there is a lot of variation in evictions types and filing patterns, and it is important to understand the context of the time period to understand the eviction rates. The map plots also show that there are certain neighborhoods that are more affected by evictions than others, and these neighborhoods are the ones that have been historically redlined.

4.1 Limitations and Future Work

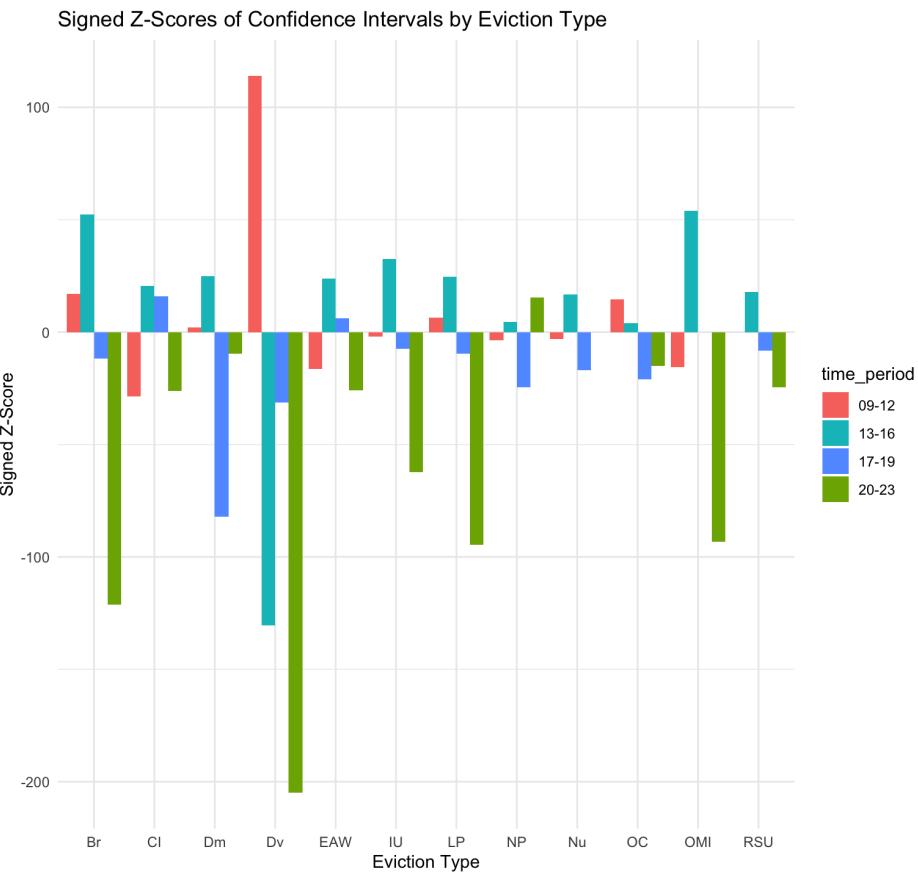
The housing crisis is an incredibly complex topic, and it would be a disservice to not go over at least some of the limitations of this paper. This data set does not include eviction notices from single-room occupancies (SROs), which naturally have a lot of movement in housing and evictions. If one examines the evictions in here closely, the reasons for evictions and intentions are completely different. A lot of people coming to SROs have dealt with trauma from homelessness or other events, and they usually are not getting the support they need. In fact, coming into SROs often entail conforming to strict regulations including zero-tolerance for visitors and curfews.

Evictions are also not the only way people are displaced from their homes. There are other ways that people are displaced from their homes, such as buyouts, harassment, and other strategies that landlords use to get tenants to leave. This data set does not include these other strategies, and it would be interesting to see how these strategies have changed over time.

Finally, I have not fully explored this data set. There is a lot more to explore with the spatial aspect of this data set, and I would like to explore the spatial patterns of evictions in San Francisco. I would also like to explore the relationship between evictions and other variables such as housing prices, rents, and demographics.



(a) z-scores organized by Time Period



(b) z-scores organized by Time Period

Figure 4: Signed z-scores by Time Period and ET

References

1. Eviction Notices in SF Data Set: DataSF and SF Rent Board. Retrieved from <https://data.sfgov.org/Housing-andBuildings/Eviction-Notices/5cei-gny5>
2. Capps, K. (2014). Peak Eviction Came to San Francisco Back in the 1990s. *Bloomberg*. Retrieved from <https://www.bloomberg.com/news/articles/2014-06-04/peak-eviction-came-to-san-francisco-back-in-the-1990s>
3. Central Market/Tenderloin Poverty Portal. Retrieved from <https://cmtldata.org/data/poverty>
4. Brinklow, Eviction notices in San Francisco down for the first time since 2010. Retrieved from <https://cdn2.hubspot.net/hubfs/4408380/PDF/Eviction-Reports-Articles-Cities-States/eviction-sf-decline-tec.pdf>
5. Tracy, A Decade of Displacement. Retrieved from https://www.foundsf.org/index.php?title=A_Decade_of_Displacement