Burglary, Income, and Border Proximity: A Municipal-Level Analysis of Canton Zurich

Thilo Holstein, Hans Joseph Thalathara

2026-02-20

Contents

1 Introduction

Burglary patterns across Swiss municipalities have long been a subject of practical interest for law enforcement agencies and local communities alike. In the Canton of Basel-Landschaft, anecdotal observations from police practitioners suggest that burglaries tend to cluster in wealthier municipalities located near Switzerland's international borders. The reasoning behind this observation is intuitive: affluent households may present more attractive targets, and proximity to a border could facilitate quick escape across national jurisdictions. If true, this pattern would have direct implications for resource allocation in crime prevention.

However, anecdotal observations, no matter how consistent, are not the same as empirical evidence. What appears as a clear pattern in one canton may not hold elsewhere, and apparent correlations can dissolve once confounding factors are accounted for. This analysis takes those observations as a starting point and asks: do they generalise?

Specifically, we examine whether the same relationship between income, border proximity, and burglary rates exists in the Canton of Zurich – a canton with a different economic structure, geographic configuration, and population density than Basel-Landschaft. By combining burglary statistics, income data, and geographic boundary data at the municipal level, we test these assumptions empirically using both exploratory visualisation and formal count regression models. The goal is not to confirm the Basel-Landschaft narrative, but to evaluate it critically in a new context.

1.1 Research Question

The central question guiding this analysis is:

> *How does income level relate to burglary rates across municipalities in Canton Zurich, and does proximity to the national border act as an independent predictor of crime?*

This question has two dimensions. First, we investigate whether wealthier municipalities systematically experience more burglaries – a relationship that would suggest property crime follows economic incentives. Second, we test whether geographic proximity to the Swiss national border adds explanatory power beyond income, which would point to spatial or cross-border dynamics as a contributing factor.

1.2 Approach

We analyse open data from Canton Zurich to investigate these questions at the municipal level. By joining income statistics, crime records, and geographic boundary data, we examine whether income predicts burglary rates, whether border proximity adds explanatory power, and how these patterns evolve over time. The analysis proceeds in three stages: data preparation and cleaning, exploratory visualisation to identify patterns and anomalies, and formal count regression modelling to test our hypotheses statistically.

1.3 Hypotheses

Based on the anecdotal observations described above, we formulate three testable hypotheses. These range from the individual effects of income and border proximity to their potential interaction.

H1 – Income and Burglaries: Municipalities with higher median income exhibit higher burglary rates, as wealthier areas may present more attractive targets.

H2 – National Border Proximity and Burglaries: Municipalities located closer to national borders may experience higher burglary rates due to reduced distances to potential escape routes across international boundaries.

H3 – Combined Effect: Income and border proximity jointly predict burglary rates better than either variable alone.

2 Data Sources

Our analysis draws on five publicly available datasets from three different sources, provided in two distinct formats (CSV and GeoPackage). All data were downloaded locally and are included in the project repository.

2.1 Burglary Data

Burglary statistics by municipality and city district for Canton Zurich (2009–2024), published by the Canton of Zurich on the Open Data portal (opendata.swiss). The dataset contains approximately 15,000 records with variables including municipality identifiers, year, offence type, number of offences, population, and burglary rate per 1,000 inhabitants. Format: CSV.

2.2 Income Data

Two separate income datasets were used:

- **Canton level:** Median taxable income by municipality (1999–2022), published by the Statistical Office of Canton Zurich (data.zh.ch). Format: CSV.
- **City district level:** Median taxable income by city district for the city of Zurich (1999–2023), published by the City of Zurich (data.stadt-zuerich.ch). Format: CSV.

The two datasets were joined at different spatial levels to achieve complete income coverage across both municipalities and city districts.

2.3 Geographic Data

- **Municipality boundaries:** GeoPackage of municipal boundaries for Canton Zurich (geo.zh.ch), used for spatial visualisation and as join keys for the burglary and income data.
- **City district boundaries:** GeoPackage of Zurich city district boundaries (geo.zh.ch), providing geometries for the 12 Stadtkreise.
- **Swiss national borders:** GeoPackage from swisstopo (swissBOUNDARIES3D), used to calculate each municipality's distance to the nearest national border.

3 Cleaning and Formatting the Dataset

3.1 Data Formatting

Burglary data for the canton of Zurich were restricted to records representing total burglaries ("Einbrüche insgesamt"), which was the sum across different burglary types. Observations with unknown municipality names or unknown city districts were excluded. Variables related to legal references were removed, as they were not required for the analysis.

Median income data were integrated at two seperate spatial levels. At the municipality level, median income data from Canton Zurich were filtered to exclude cantonal aggregates and the city of Zurich itself. The dataset was reduced to essential variables: municipality identifier, name, year, and income value. At the city district level, median income data for the city of Zurich were filtered to include only the standard tax rate, and converted from thousands to actual values by multiplying the 50th percentile income by 1000.

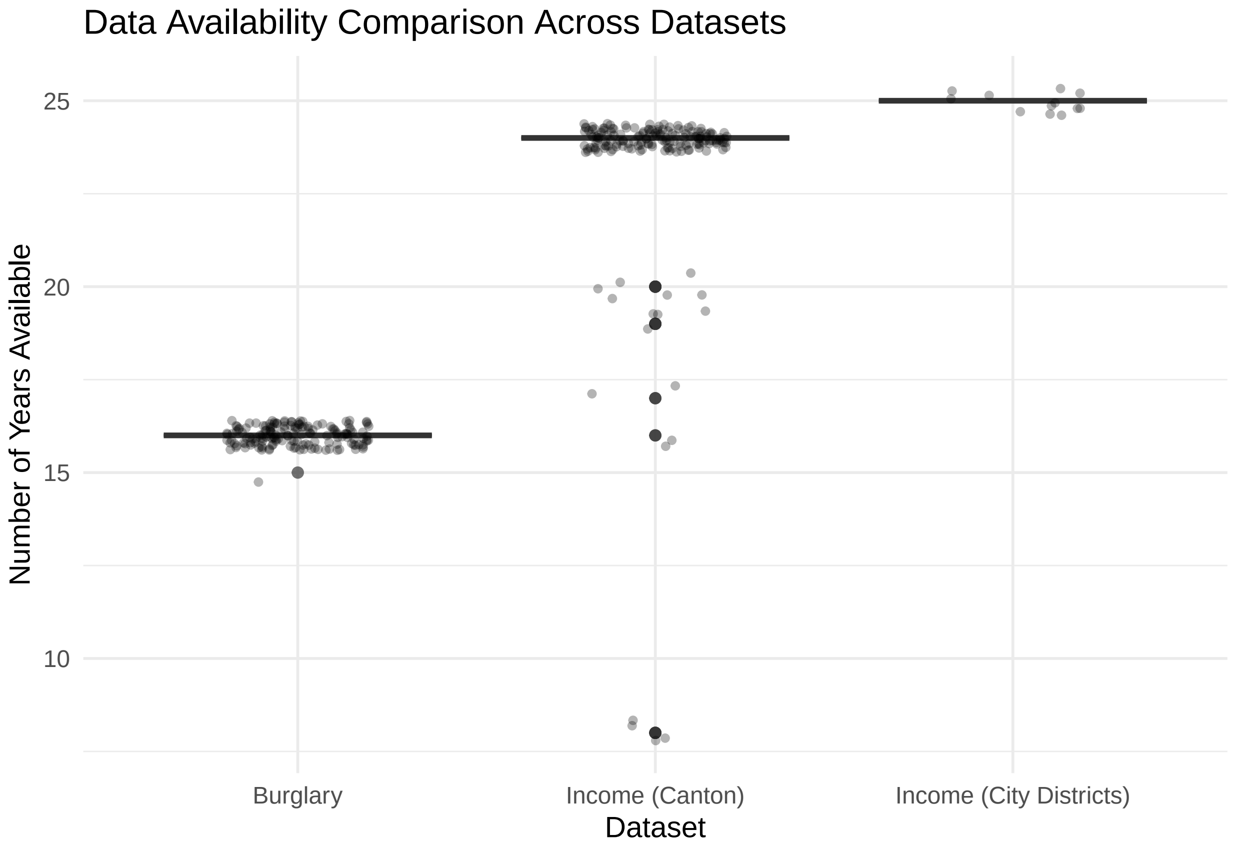
Spatial boundary data for municipalities in the canton of Zurich were obtained from an official GeoPackage. Only geometries corresponding to municipalities were retained, while non-municipality and city-level geometries were excluded. The dataset was reduced to the municipality identifier and the corresponding geometry.

Swiss national border data were sourced from swisstopo and imported from a GeoPackage file. The data were filtered to retain only the Swiss national boundary. The polygon representation of the national territory was converted into a boundary line to enable distance calculations.

Prior to spatial analysis, all municipality geometries were transformed to a common coordinate reference system matching that of the Swiss border data. For each municipality, the centroid of the polygon geometry was computed. Euclidean distances from these centroids to the Swiss national border line were then calculated. Distances were stored in meters and additionally converted into kilometers for interpretability.

City district geometries for the city of Zurich were processed separately. Because the original dataset contained curved polygon geometries, the data were converted into standard multipolygon geometries using GDAL utilities. The converted geometries were subsequently imported, restricted to district identifiers and geometries, and transformed to the same coordinate reference system as the Swiss border data. Centroids were calculated for each city district, and distances to the Swiss national border were computed analogously to the municipality-level analysis, again expressed in both meters and kilometers.

3.2 Investigating Missing Values



The datasets employed in this analysis vary in their temporal completeness across the observation period from 1999 to 2023. Among the three primary data sources, only the income data for city districts (Stadt Zurich) provide complete coverage for all geographic units across the entire timeframe, with no missing observations. The burglary dataset exhibits minimal missingness, with one municipality showing absent data for reasons that could not be determined from the available documentation. The limited scope of this gap suggests negligible impact on the overall analysis. Income data at the cantonal level (municipal aggregation) demonstrate more substantial incompleteness, with several municipalities missing data for specific years. Table 1 provides a detailed breakdown of missing observations by municipality and year, allowing for transparent assessment of potential limitations in temporal coverage.

Table 1: Municipalities with Incomplete Coverage

Dataset	Municipality	BFS Nr	Years Available	Missing Years	Start Year	End Year
Burglary	Stammheim	292	15	1	2009	2024
Income (Canton)	Stammheim	292	8	16	2015	2022
Income (Canton)	Wädenswil	293	8	16	2015	2022
Income (Canton)	Elgg	294	8	16	2015	2022
Income (Canton)	Horgen	295	8	16	2015	2022
Income (Canton)	Bauma (bis 2014)	171	16	8	1999	2014
Income (Canton)	Sternenberg (bis 2014)	179	16	8	1999	2014
Income (Canton)	Illnau-Effretikon (bis 2015)	174	17	7	1999	2015
Income (Canton)	Kyburg (bis 2015)	175	17	7	1999	2015
Income (Canton)	Hirzel (bis 2017)	132	19	5	1999	2017
Income (Canton)	Horgen (bis 2017)	133	19	5	1999	2017
Income (Canton)	Elgg (bis 2017)	217	19	5	1999	2017
Income (Canton)	Hofstetten (bis 2017)	222	19	5	1999	2017
Income (Canton)	Oberstammheim (bis 2018)	36	20	4	1999	2018
Income (Canton)	Unterstammheim (bis 2018)	42	20	4	1999	2018
Income (Canton)	Waltalingen (bis 2018)	44	20	4	1999	2018
Income (Canton)	Hütten (bis 2018)	134	20	4	1999	2018
Income (Canton)	Schönenberg (bis 2018)	140	20	4	1999	2018
Income (Canton)	Wädenswil (bis 2018)	142	20	4	1999	2018

Further investigation revealed that the missing income data were not random but rather systematic, arising from administrative boundary changes. Between 2015 and 2019, several municipalities in Canton Zurich underwent mergers, resulting in the creation of new administrative units with updated BFS (Federal Statistical Office) identifiers. Specifically:

- On 1 January 2019, the former municipalities of Ober-, Unterstammheim, and Waltalingen merged to form the new municipality of Stammheim
- On 1 January 2019, the former municipalities of Hütten and Schönenberg merged into the municipality of Wädenswil
- On 1 January 2018, the former municipality of Hofstetten merged into the municipality of Elgg
- On 1 January 2018, the former municipality of Hirzel merged into the municipality of Horgen
- On 1 January 2016, Kyburg and Illnau-Effretikon merged to form the municipality of Illnau-Effretikon

These mergers resulted in discontinued BFS identifiers for the former municipalities, with new identifiers assigned to the merged entities. Consequently, income data appeared "missing" for the merged municipalities in pre-merger years, as these records remained associated with the former municipal identifiers. To address this issue, we implemented a systematic reconciliation procedure. For each merged municipality, we first assessed the completeness of income data under the new identifier. Where gaps existed for pre-merger years, we reconstructed the median income values by averaging data from the constituent former municipalities. This approach ensured temporal continuity in the income variable while maintaining geographic consistency with the current administrative boundaries used in the burglary and spatial datasets.

After the merger reconciliation above, 329 of 2,735 observations (12.0%) still had missing values in the income variable (`INCOME_VALUE`). These remaining gaps arise from two causes:

- **2023 (158 missing):** Municipal income statistics for Canton Zurich had not yet been published at the time of analysis. Only the 12 city districts of Zurich have income data for this year, sourced from the separate city-level dataset.
- **2024 (171 missing):** No income data are available for any municipality or city district.

All other variables – burglary counts, population, burglary rates, and border distance – are complete across the entire observation period. Since the missingness is entirely explained by data availability rather than by any systematic relationship with the outcome variable, the missing-at-random assumption is satisfied.

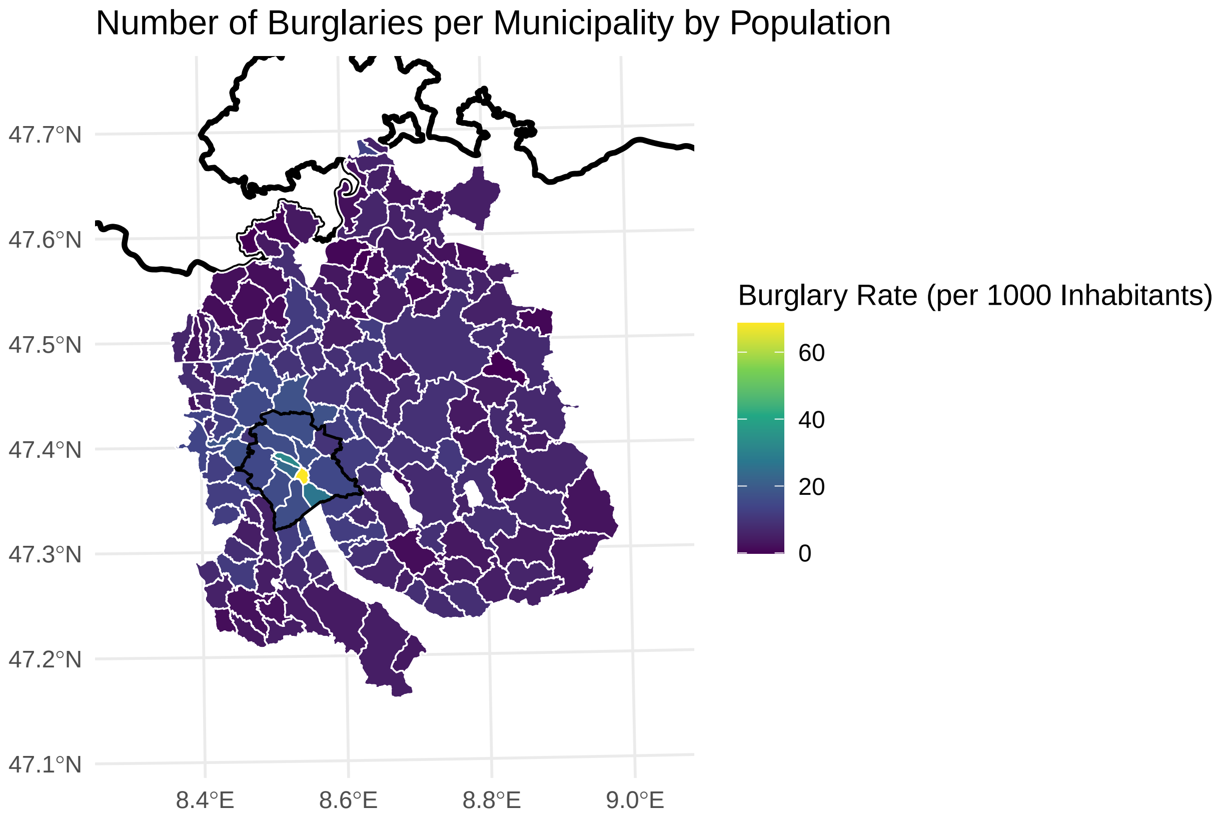
We retain the full dataset for descriptive and spatial analyses (maps, time series, distributions), where income is not required. For all analyses involving income as a variable – including the correlation matrix, scatterplots, and regression models – observations with missing income values are excluded. This approach maximises the use of available data while ensuring that statistical models are estimated on complete cases only.

4 Exploratory Data Analysis

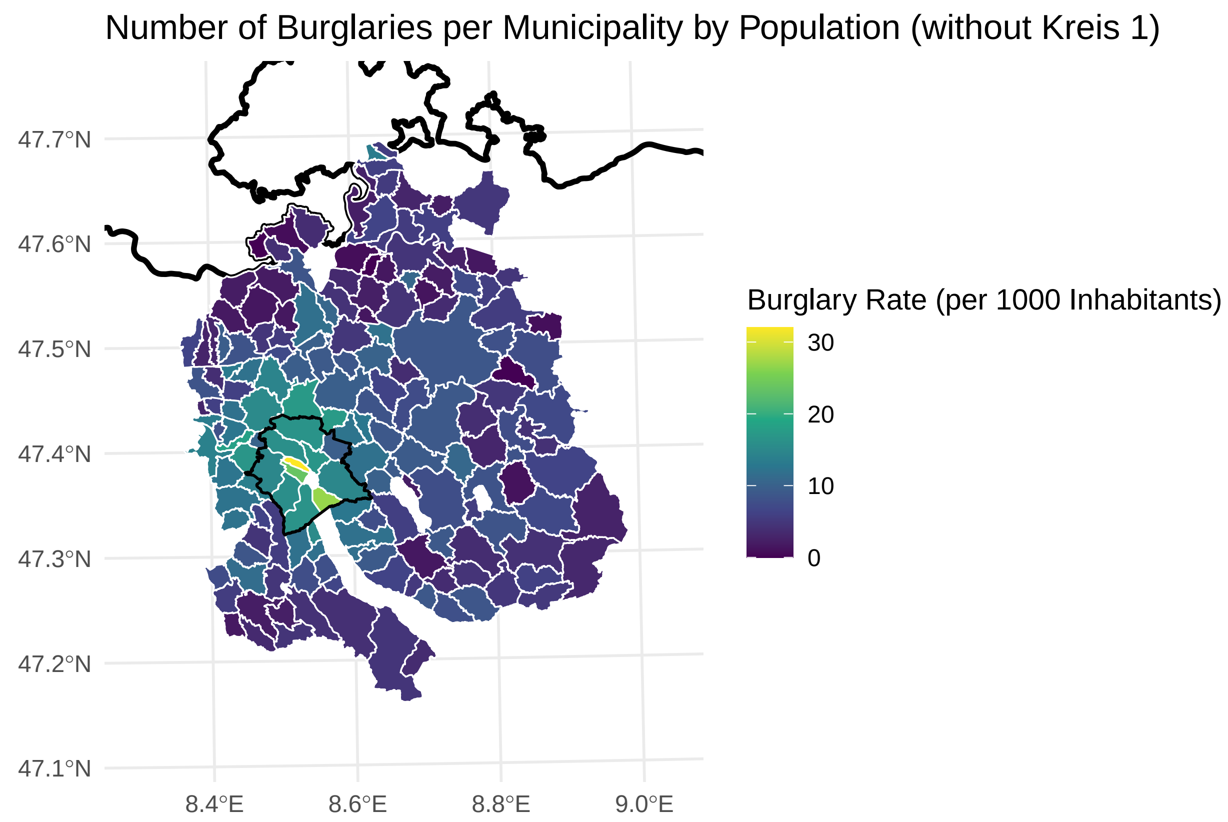
Before testing our hypotheses through formal modelling, we explore the data visually. This section begins with a spatial overview of burglary patterns across the canton, then follows a narrative from the macro level (canton-wide trends over time) down to the micro level (individual municipality patterns), building an increasingly detailed picture of how income, burglaries, and geography interact across Canton Zurich.

4.1 Spatial Overview: Burglary Rates by Municipality

The following maps show the total number of burglaries per municipality, normalised by average population, across the full observation period. This provides a first visual impression of how burglary risk is distributed geographically.


Number of Burglaries per Municipality by Population
47.7°N
47.6°N
47.5°N
47.4°N
47.3°N
47.2°N
47.1°N
8.4°E
8.6°E
8.8°E
9.0°E
Burglary Rate (per 1000 Inhabitants)
60
40
20
0

The city district Kreis 1 seems to be a strong outlier in the data. To have a more balanced output of the burglary rate per municipality and population, this is how the map of the canton of Zurich looks without the outlier:


Number of Burglaries per Municipality by Population (without Kreis 1)
47.7°N
47.6°N
47.5°N
47.4°N
47.3°N
47.2°N
47.1°N
8.4°E
8.6°E
8.8°E
9.0°E
Burglary Rate (per 1000 Inhabitants)
30
20
10
0

4.2 Descriptive Overview

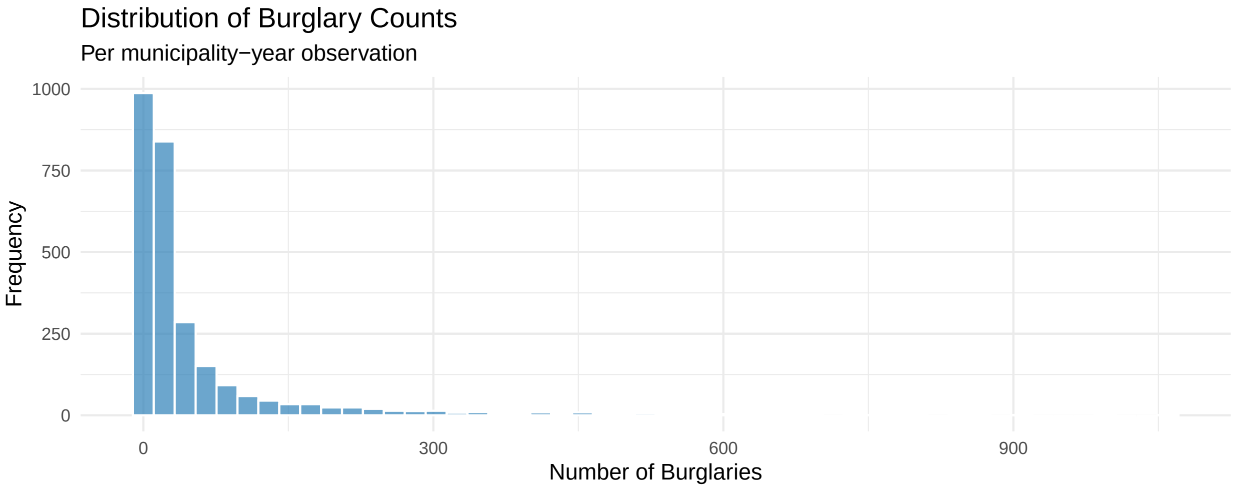
We begin with a summary of the dataset's structure and the distributions of our key variables. This provides the necessary context for all subsequent analyses.

The dataset contains **2'735 observations** across **171 municipalities and city districts** over **16 years** (2009–2024). Each observation represents one municipality in one year, recording the number of burglaries, population size, burglary rate, median income, and distance to the national border.

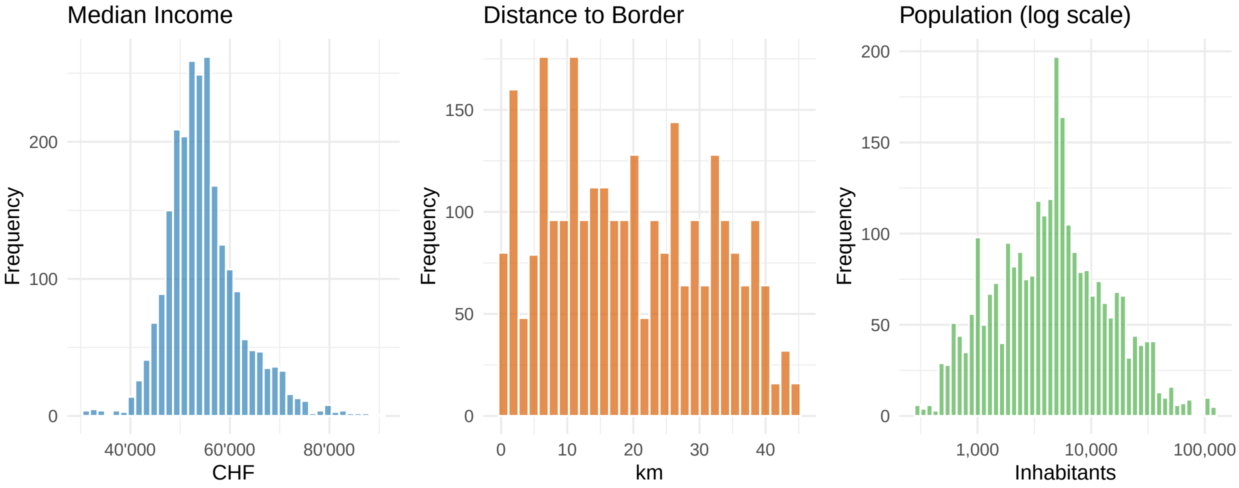
Table 2: Summary statistics for key variables

Variable	Mean	Median	SD	Min	Max	Missing
Burglary Count	55.1	18.0	109.0	0.0	1060.0	0
Population	8612.2	4699.0	12759.9	296.0	119315.0	0
Burglary Rate (per 1k)	5.0	4.0	4.7	0.0	74.6	0
Median Income (CHF)	54716.2	53800.0	7420.4	31100.0	90300.0	329
Border Distance (km)	19.6	18.5	12.0	0.3	44.5	0

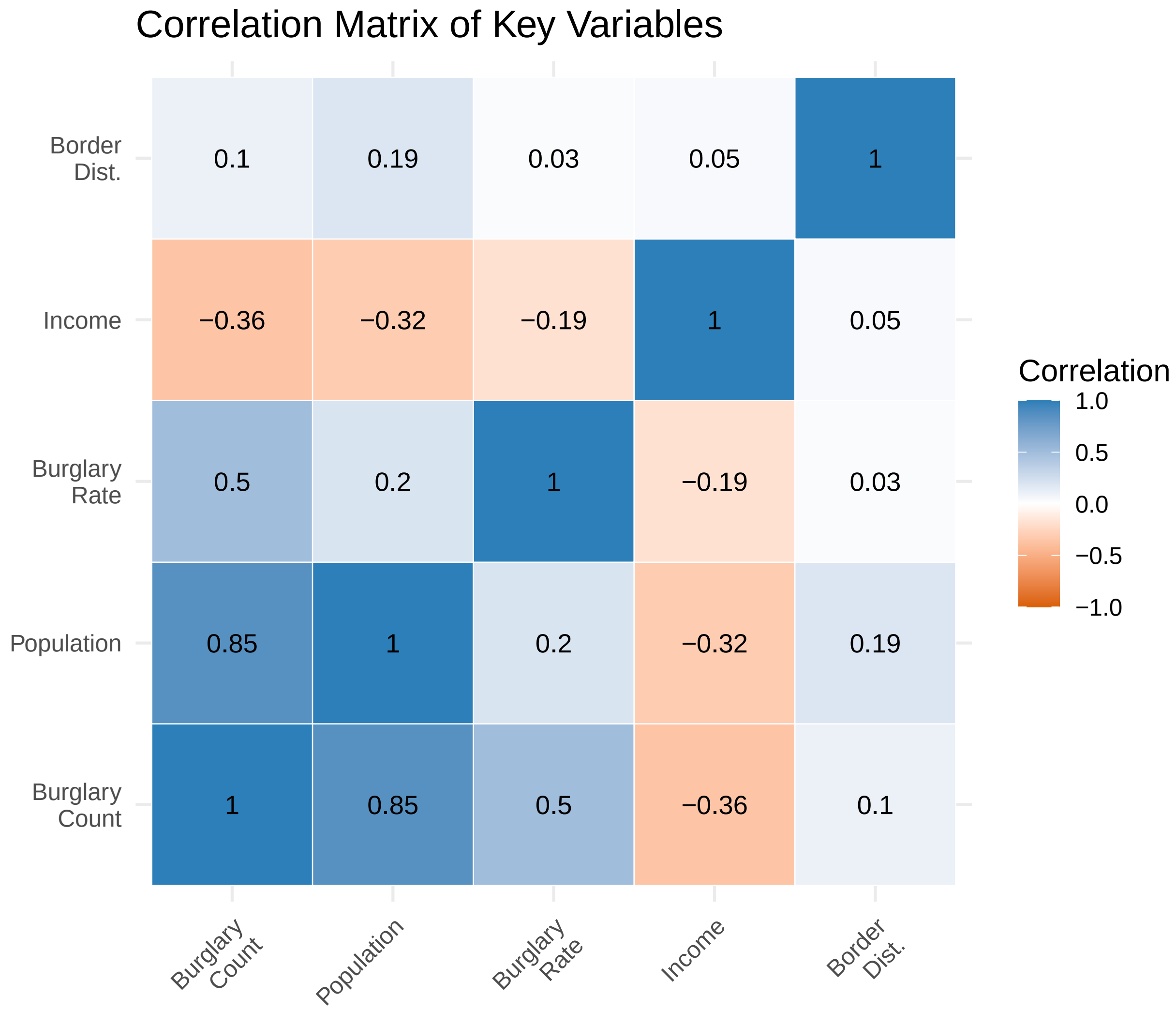
The distributions of our key variables reveal important characteristics of the data:



The burglary count distribution is heavily right-skewed: most municipality-year observations record relatively few burglaries, while a small number of large municipalities account for the extreme values. This skew is a hallmark of count data and one reason why standard linear regression is inappropriate here — a point we return to in the modelling section.



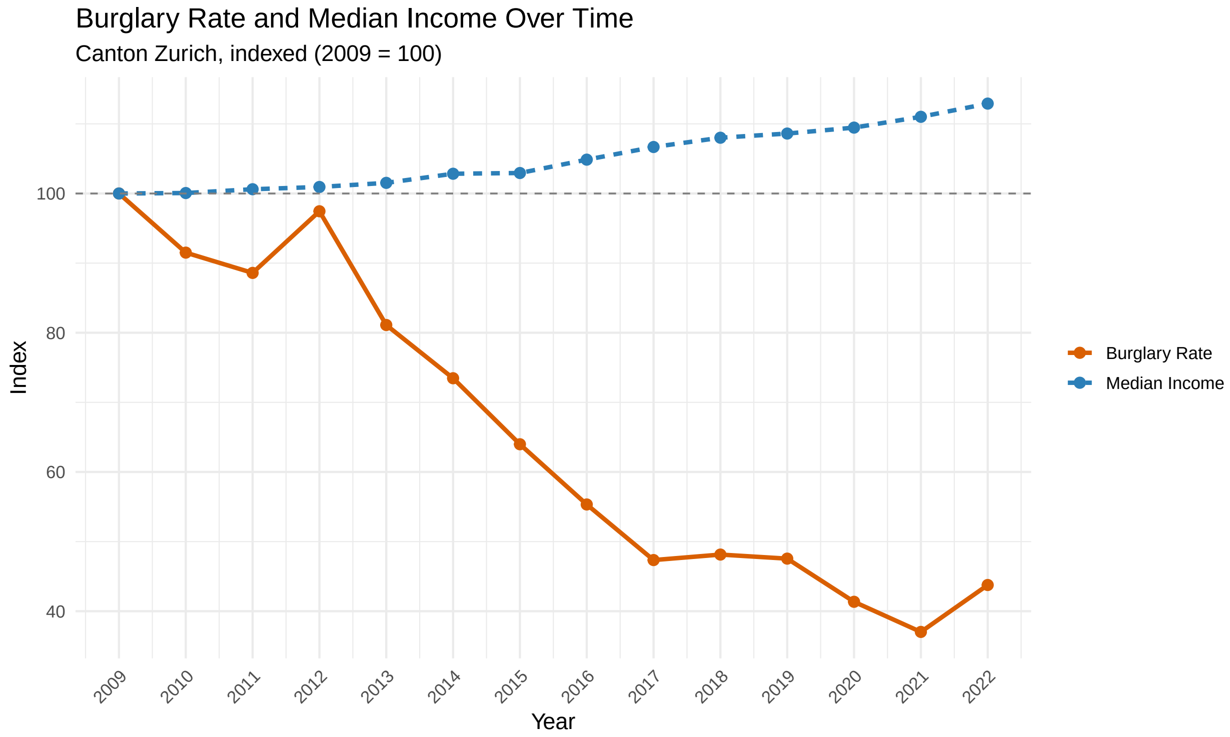
Income is approximately normally distributed across municipalities, with most values clustered between 30,000 and 50,000 CHF. Border distance shows a right-skewed distribution — most municipalities lie within 20–30 km of the border, with few exceeding 40 km. Population varies enormously (note the log scale): from small rural communes with fewer than 2,000 residents to Zurich city districts with over 50,000.



The correlation matrix provides a first hint at the relationships we will explore in detail. Raw burglary counts correlate strongly with population — unsurprisingly, since larger municipalities have more of everything. This is precisely why we use the burglary *rate* (or count models with a population offset) rather than raw counts in our analysis. The correlations between burglary rate, income, and border distance are the ones relevant to our hypotheses and will be examined in the following sections.

4.3 Temporal Trends: Income and Burglary Rates

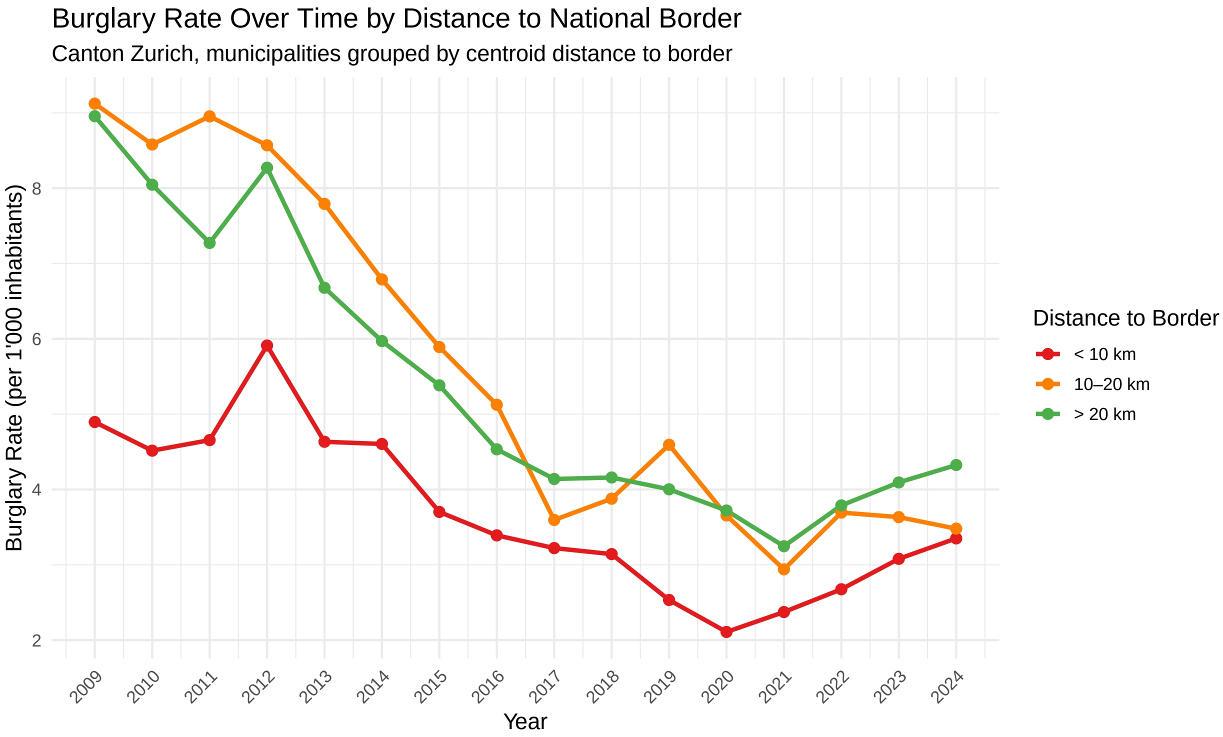
We begin by examining how burglary rates and median income have evolved across Canton Zurich over time. Since the two variables operate on fundamentally different scales, we normalise both to an index where the first available year equals 100. This allows direct comparison of their relative trajectories.



The indexed timeline reveals a striking divergence: while median income has risen steadily over the observation period, burglary rates have declined substantially. This inverse macro-trend already raises questions about the assumed positive relationship in H1. If higher income attracted more burglaries, we would expect both lines to move in the same direction — yet they clearly do not. This motivates a closer look at whether the pattern holds at the municipal level, or whether spatial factors such as border proximity might better explain variation in burglary rates.

4.4 Spatial Dimension: Burglary Rates by Border Proximity

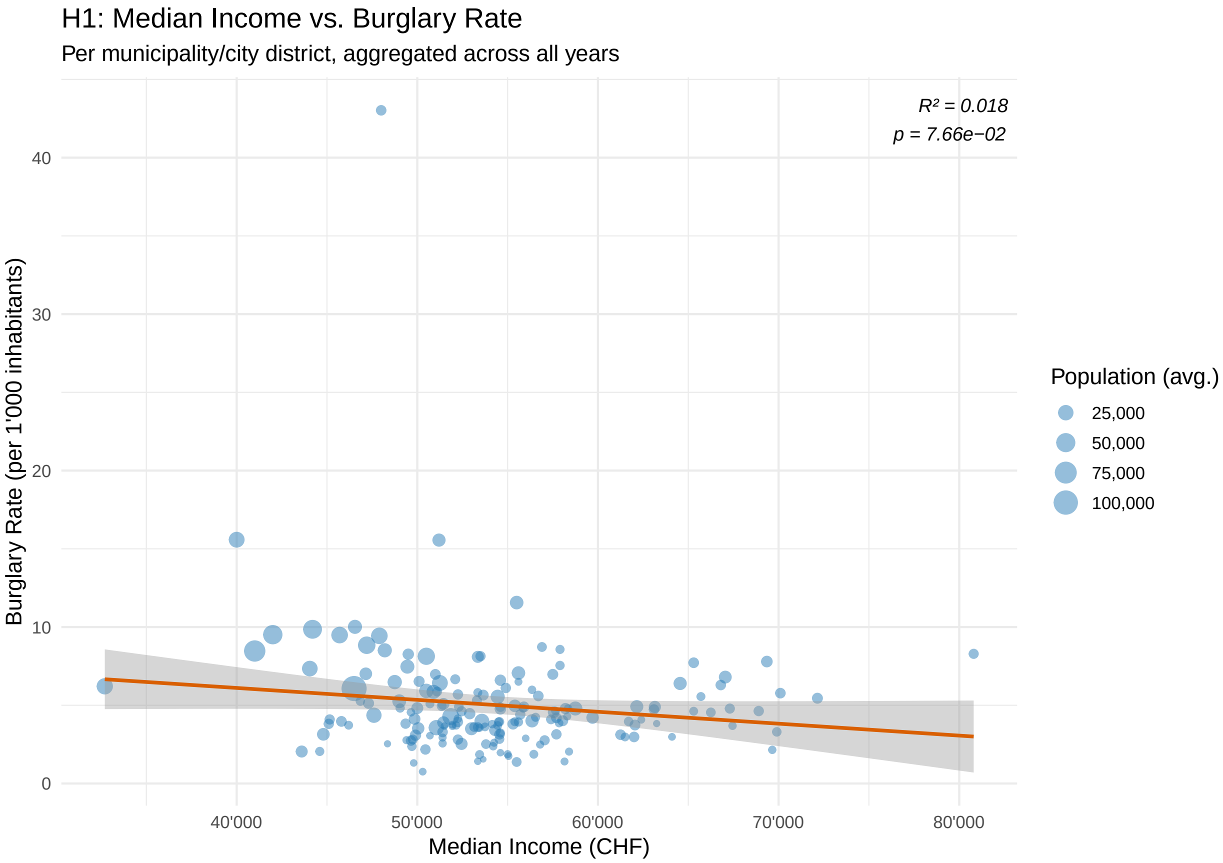
The macro-level decline in burglaries may mask important spatial heterogeneity. To investigate whether geography matters, we group municipalities into three categories based on their centroid distance to the Swiss national border and track their burglary rates over time separately.



This chart reveals a clear spatial gradient: municipalities within 10 km of the national border consistently exhibit higher burglary rates than those further inland. While all three groups follow the general downward trend observed in Section 4.1, the gap between border-proximate and interior municipalities persists across the entire observation period. This suggests that border proximity is a structurally relevant factor — not merely a transient pattern — and provides initial support for H2.

4.5 Cross-Sectional Analysis: Income vs. Burglary Rate

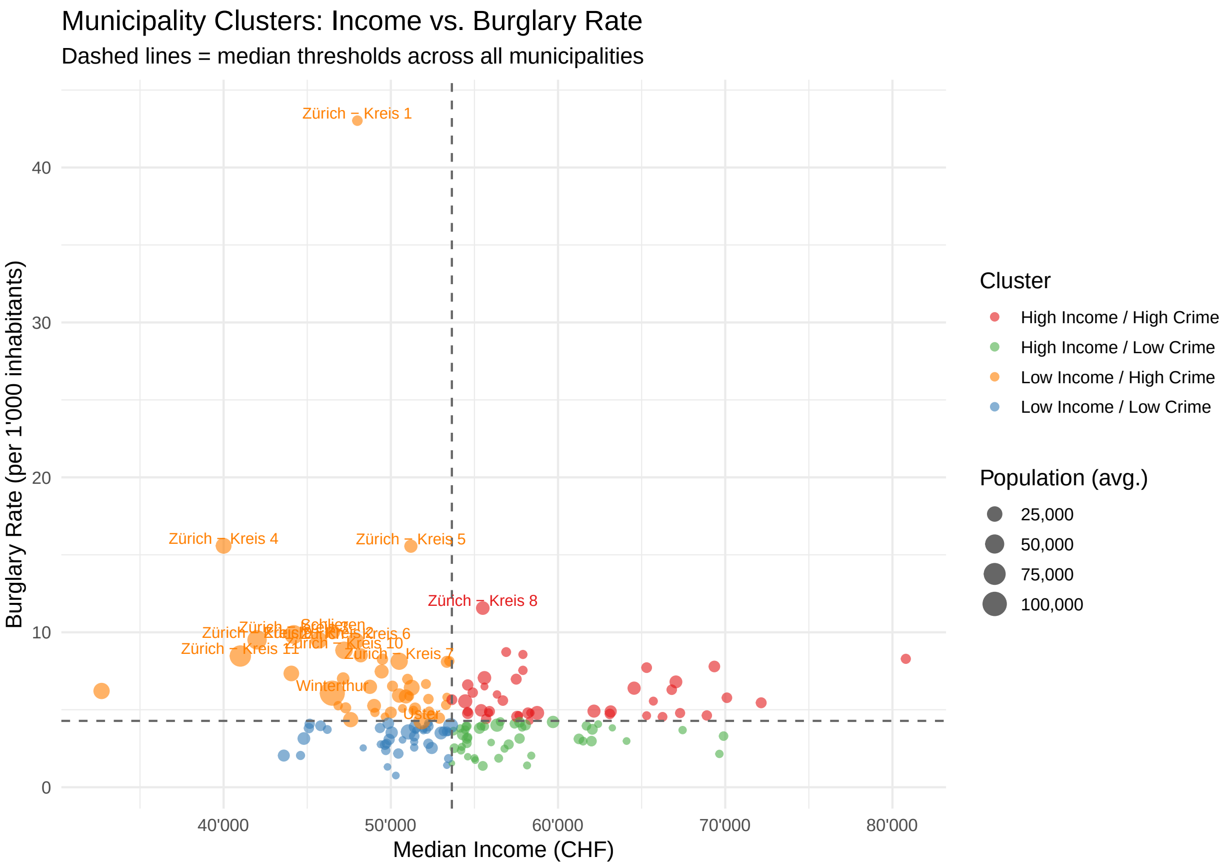
Having established the temporal and spatial context, we now examine the core relationship hypothesised in H1: does higher municipal income predict higher burglary rates? We aggregate each municipality across all available years and fit a simple linear regression.



The scatterplot does not support H1. The regression line slopes slightly downward, suggesting that wealthier municipalities tend to have marginally lower — not higher — burglary rates. However, the effect is weak: income explains only about 2% of the variance in burglary rates, and the p-value does not reach conventional significance at the 5% level. A few outliers with very high burglary rates (e.g., smaller municipalities like Kreis 1 of Zurich City, where a handful of incidents inflate the rate) are visible in the upper portion of the plot. Given these findings, income alone is a poor predictor of burglary rates. This motivates the inclusion of border proximity as an additional predictor.

4.6 Municipality Typology: Income–Crime Clusters

We classify all municipalities into four quadrants based on whether their median income and average burglary rate fall above or below the canton-wide median. This typology provides an intuitive summary of the landscape and highlights which municipalities defy the expected patterns.


Municipality Clusters: Income vs. Burglary Rate

Dashed lines = median thresholds across all municipalities

The quadrant plot reveals that municipalities are distributed relatively evenly across all four clusters, with no strong concentration in the "High Income / High Crime" quadrant that H1 would predict. The labelled outliers — municipalities with exceptionally high burglary rates or large populations — provide concrete cases for further investigation. Notably, the distribution of municipalities across clusters supports the earlier regression findings: income alone does not systematically sort municipalities into high- or low-crime categories. Other factors, particularly geographic location, appear to play a more decisive role.

5 Modelling

The exploratory analysis revealed visual patterns – a spatial gradient near the border and no clear income–crime link – but visual impressions can be misleading. To rigorously test our three hypotheses, we now turn to formal statistical modelling. We move beyond simple linear regression (which treats burglary *rates* as continuous) and instead model the raw **burglary counts** directly. This is more appropriate because:

- Burglary counts are non-negative integers — classic **count data**.
- Municipalities differ vastly in population size, so we need to account for this.
- The variance in crime counts often exceeds the mean (**overdispersion**), which standard linear models cannot handle.

We fit two types of count regression models — **Poisson** and **Negative Binomial** — and compare their performance.

5.1 Why Count Regression?

A standard linear regression on rates (Häufigkeitszahl) treats every municipality equally, regardless of whether it has 2,000 or 50,000 residents. A municipality with 2 burglaries and 1,000 residents gets the same weight as one with 200 burglaries and 100,000 residents — even though the latter estimate is far more precise. Count regression models solve this by working with the raw counts and using a so-called **offset variable** to account for population size.

5.1.1 The Offset Variable

The key idea: we model `Straftaten_total` (raw burglary count) as the outcome, but include `log(Einwohner)` as an **offset**. Mathematically:

$$\log(\text{count}) = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 \cdot \text{border distance} + \log(\text{population})$$

Rearranging:

$$\log\left(\frac{\text{count}}{\text{population}}\right) = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 \cdot \text{border distance}$$

So the model effectively predicts the **rate** (burglaries per capita), but it correctly weights each observation by its population size. The offset is not estimated — it enters the model with a fixed coefficient of 1.

5.1.2 Covariates

We include three covariates (predictor variables):

- **Income** (in 10,000 CHF): Tests whether wealthier municipalities experience more burglaries (H1).
- **Distance to border** (km): Tests whether proximity to the Swiss national border increases burglaries (H2).
- **Year** (centered): Controls for the canton-wide declining trend in burglaries over time.

5.2 Poisson Regression

The Poisson model is the natural starting point for count data. It assumes that the outcome follows a Poisson distribution, where the mean equals the variance (**equidispersion**).

Poisson Full Model: Burglary Count ~ Income + Border Distance + Year + offset(log Population)

Variable	Estimate	Std. Error	p-value
(Intercept)	-3.3233	0.0189	0.00e+00
Income (10k CHF)	-0.2146	0.0038	0.00e+00
Distance to Border (km)	-0.0079	0.0003	1.62e-156
Year	-0.0696	0.0007	0.00e+00

Under the Poisson model, all three predictors appear highly significant. However, before interpreting these results, we must verify the model's core assumption — that the variance of burglary counts equals the mean.

5.2.1 Checking for Overdispersion

The critical Poisson assumption is that the variance equals the mean. We check this with the **dispersion statistic**: the sum of squared Pearson residuals divided by the degrees of freedom. A value near 1 indicates the assumption holds; values much greater than 1 signal **overdispersion**.

Dispersion statistic: 18.99

The dispersion statistic is substantially greater than 1, confirming that the data are **overdispersed**. This means the Poisson model underestimates the true variability in burglary counts. Consequently, the standard errors are too small, and the p-values are artificially low — the model appears more confident in its estimates than it should be. This motivates fitting a Negative Binomial model, which explicitly accounts for overdispersion.

5.3 Negative Binomial Regression

The Negative Binomial (NB) model extends the Poisson by adding an extra parameter **theta** (θ) that captures the overdispersion. When $\theta \to \infty$, the NB reduces to the Poisson. When θ is small, there is substantial extra variability beyond what the Poisson allows.

This extra variability reflects real-world factors not captured by our covariates: differences in local policing, urban vs. rural character, socioeconomic conditions, and other unobserved heterogeneity between municipalities.

5.3.1 NB Model Results

NB Full Model: Burglary Count ~ Income + Border Distance + Year + offset(log Population)

Variable	Estimate	Std. Error	p-value
(Intercept)	-4.1671	0.0988	0.00e+00
Income (10k CHF)	-0.1249	0.0183	9.76e-12
Distance to Border (km)	0.0011	0.0011	3.20e-01
Year	-0.0678	0.0034	2.42e-89

Theta (overdispersion parameter): 2.821

Compared to the Poisson model, the NB standard errors are noticeably larger and several p-values increase substantially. This is the expected consequence of accounting for overdispersion: the model no longer overstates its confidence. The theta parameter quantifies the degree of extra-Poisson variability — a small theta indicates substantial overdispersion.

5.3.2 Incidence Rate Ratios

The exponentiated coefficients give **Incidence Rate Ratios (IRR)**. An IRR of 0.95 for border distance means: each additional kilometre away from the border *multiplies* the expected burglary rate by 0.95 — i.e., a 5% decrease per kilometre.

Variable	IRR	CI lower (2.5%)	CI upper (97.5%)
(Intercept)	0.0155	0.0130	0.0184
Income (10k CHF)	0.8826	0.8548	0.9113
Distance to Border (km)	1.0011	0.9986	1.0036

Variable	IRR	CI lower (2.5%)	CI upper (97.5%)
Year	0.9344	0.9285	0.9404

The IRR table translates the model coefficients into a more intuitive scale. An IRR close to 1 indicates no meaningful effect. The confidence interval for border distance spans 1, reinforcing that it is not a reliable predictor. Income, by contrast, has an IRR and confidence interval entirely below 1 (CI: 0.855–0.911), confirming a significant protective association. The year variable similarly has an IRR consistently below 1, confirming the steady annual decline in burglary rates.

5.3.3 Interaction Model: Does Income Matter More Near the Border?

To test H3, we add an interaction term between income and border distance. If the interaction is significant, it would mean that the effect of income on burglaries *depends* on how close a municipality is to the border.

Variable	Estimate	Std. Error	p-value
(Intercept)	-2.6080	0.2605	1.33e-23
Income (10k CHF)	-0.4124	0.0486	2.06e-17
Distance to Border (km)	-0.0663	0.0104	1.76e-10
Year	-0.0668	0.0034	9.99e-88
Income x Border Distance	0.0123	0.0019	1.19e-10

The interaction term (Income × Border Distance) is statistically significant (p = 1.19e-10), and notably, including it renders both main effects significant as well — a pattern that was absent in the additive NB model. This suggests that the relationship between income and burglary rates is modulated by geographic proximity to the border.

However, interpreting this result requires caution. The significant interaction indicates a statistical dependency between income and border distance, but the practical relevance is limited. The positive interaction coefficient (+0.0123) implies that the negative income effect weakens as distance to the border increases — in other words, the protective association of higher income diminishes further from the border. Yet the AIC improvement over the additive NB model is marginal (19,097 vs. 19,131.1), suggesting that the interaction, while statistically detectable, adds only modest explanatory power. Given the large sample size (n = 2,406 complete observations), even small effects can reach significance without being substantively meaningful. We therefore interpret H3 with caution: there is evidence of a statistical interaction, but it does not substantially improve prediction of burglary rates.

5.4 Model Comparison: Poisson vs. Negative Binomial

Having fitted both model families, we now systematically compare them to determine which provides the more reliable basis for inference. This comparison proceeds through four complementary lenses: a formal statistical test, an information criterion, a visual comparison of coefficient uncertainty, and diagnostic checks of the residuals.

5.4.1 Likelihood Ratio Test

Since the Poisson model is nested within the NB (the Poisson is a special case when $\theta \to \infty$), we can use a **likelihood ratio test** to determine whether the extra parameter θ significantly improves the fit.

- **LR statistic:** 28038.44

- **p-value:** 0.000e+00

The test is highly significant: the NB model fits substantially better than the Poisson. This confirms that overdispersion is present and must be accounted for.

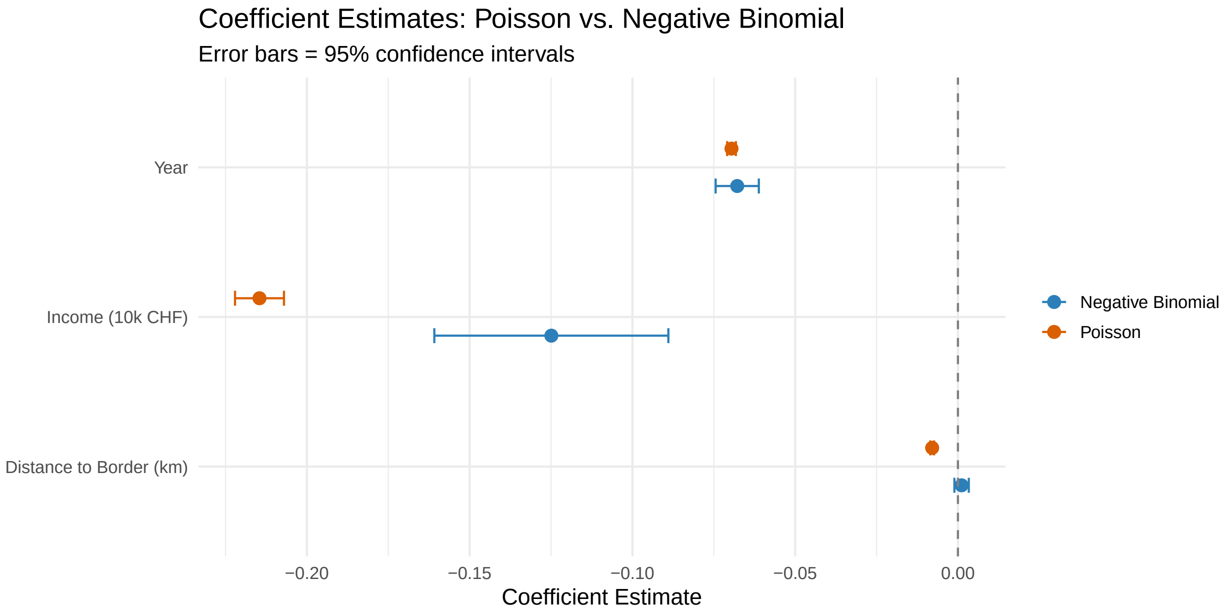
5.4.2 AIC Comparison

Model	AIC	Log-Likelihood
Poisson (full)	47167.6	-23579.8
Negative Binomial (full)	19131.1	-9560.6
NB with interaction	19097.0	-9542.5

Lower AIC indicates better fit, penalised for model complexity.

5.4.3 Coefficient Comparison

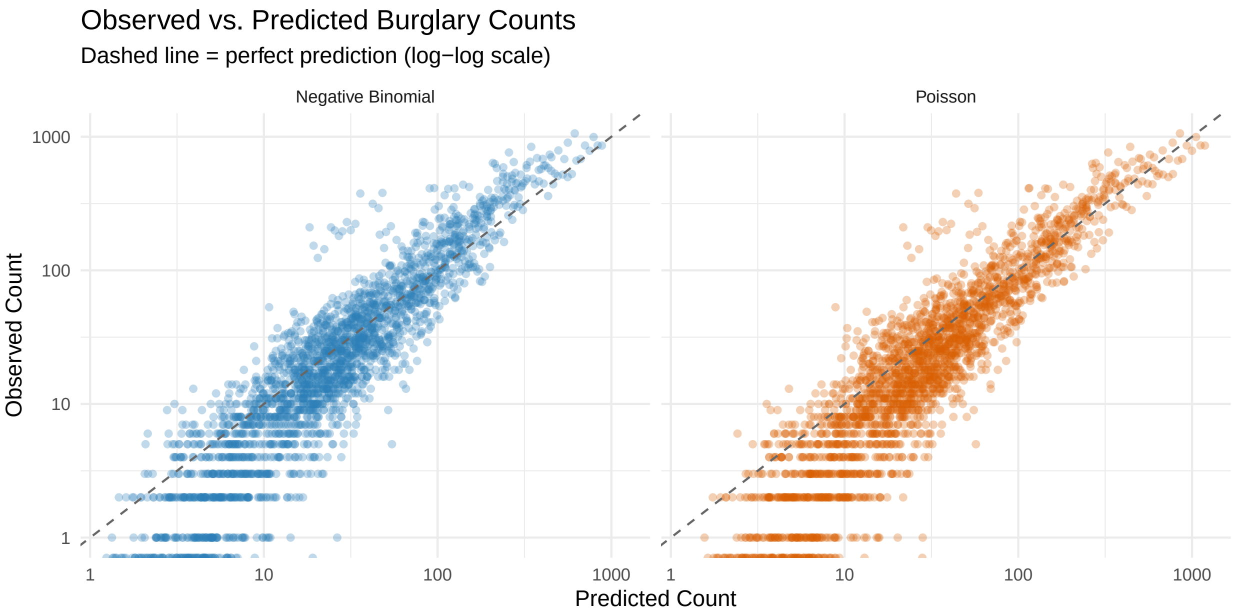
A key consequence of overdispersion: the Poisson model produces **artificially small standard errors**, making predictors appear more significant than they truly are. The NB model provides more honest uncertainty estimates.



The coefficient estimates are broadly similar between both models for income and year, but the NB confidence intervals are substantially wider — reflecting the true uncertainty in the data. The most striking difference concerns border distance: in the Poisson model, the coefficient is negative and highly significant ($= -0.007$, p 0), suggesting that municipalities closer to the border experience higher burglary rates. In the Negative Binomial model, the coefficient flips sign ($= +0.001$) and becomes non-significant ($p = 0.32$). This reversal is a direct consequence of accounting for overdispersion: the Poisson model's artificially narrow standard errors produced a spurious result that dissolves under more appropriate modelling assumptions. This finding serves as a cautionary example — had we relied solely on the Poisson model, we would have drawn the opposite conclusion regarding border proximity.

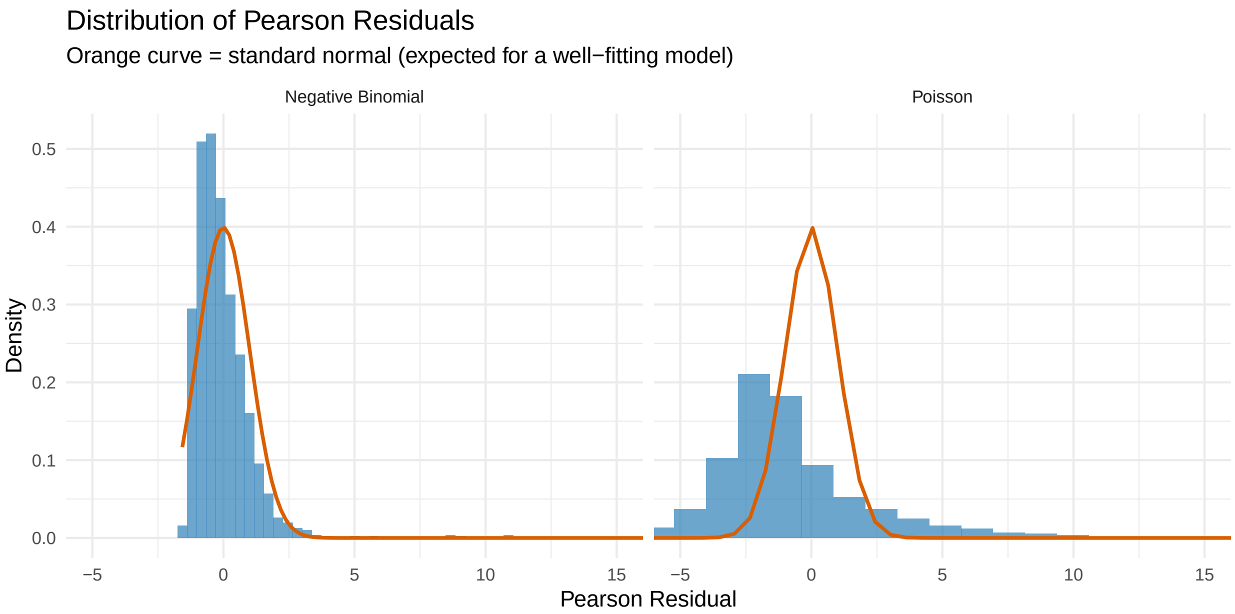
Any predictor that remains significant under both models — in this case, income and year — can be considered robustly supported.

5.4.4 Observed vs. Predicted



If the models predicted perfectly, all points would fall on the diagonal dashed line. Both models show a similar pattern: predictions cluster along the diagonal for moderate counts, but diverge for very small and very large values. This is expected — municipalities with extreme burglary counts (either very few or very many) are inherently harder to predict with only three covariates. The key observation is that both models produce comparable predictions; the advantage of the NB lies not in better point predictions, but in more honest uncertainty estimates.

5.4.5 Residual Diagnostics



Model	Mean	SD
Poisson	-0.453	4.331
Negative Binomial	-0.025	1.251

For a well-fitting model: mean 0, SD 1.

The residual distributions confirm the earlier findings. The Poisson residuals have a standard deviation far exceeding 1, indicating systematic underfitting of the variance — a direct consequence of the equidispersion assumption being violated. The NB residuals are much closer to the expected standard normal shape, with a standard deviation nearer to 1, indicating that the extra theta parameter successfully absorbs the excess variability.

Across all comparison criteria – the likelihood ratio test, AIC, coefficient uncertainty, and residual diagnostics – the Negative Binomial model consistently outperforms the Poisson. The Poisson model's assumption of equidispersion is clearly violated, leading to artificially narrow confidence intervals and overly optimistic p-values. We therefore adopt the **Negative Binomial full model** as our primary model for all subsequent hypothesis tests.

5.5 Hypothesis Assessment

With the Negative Binomial model established as the appropriate framework, we now formally evaluate each of our three hypotheses based on its coefficient estimates and significance levels.

H1 — Income and Burglaries: Not supported (direction reversed): The NB model yields a statistically significant effect of income (p = 9.76e-12), but in the opposite direction to what H1 predicted. Wealthier municipalities show lower, not higher, burglary rates (IRR = 0.88 per 10,000 CHF). The "wealthy municipalities attract burglars" hypothesis is thus contradicted by the data — income, if anything, has a protective association.

H2 — Border Proximity and Burglaries: Not supported: In the NB model, distance to the national border is not statistically significant (p = 0.32). While the exploratory analysis revealed a visually compelling spatial gradient, this pattern does not survive formal modelling once overdispersion is accounted for. The Poisson model had found a significant effect, but as demonstrated in Section 5.4.3, this was an artefact of artificially narrow standard errors. H2 is not supported.

H3 — Combined Effect: Partially supported (statistically significant, practically limited): The interaction between income and border distance is statistically significant (p = 1.19e-10), indicating that the effect of income on burglaries depends on border proximity. However, the AIC improvement over the additive model is marginal, and the interaction does not substantially increase the model's predictive power. Given the large sample size, statistical significance alone does not imply practical importance. We therefore characterise H3 as partially supported: there is evidence of a statistical interaction, but it does not meaningfully improve our ability to predict burglary rates.

6 Chapter of Choice

6.1 Interactive Web Application with Shiny

As the Chapter of Choice, we developed an interactive web application using the `shiny` package — a framework for building browser-based applications directly from R, without requiring any knowledge of web development. The application serves as a companion to this report, but with a deliberately different emphasis: where the report focuses on a cross-sectional statistical analysis, the application prioritises **interactivity, geographic exploration, and the development of key variables over time**.

6.1.1 Motivation

The datasets used in this analysis are well-suited to longitudinal exploration. Burglary data spans 2009 to 2024, income data reaches back to 1999, and all variables are available at a fine spatial resolution — individual municipalities and, for the city of Zurich, individual city districts. A static report cannot fully capture this temporal richness. An interactive application allows the reader to explore patterns at their own pace, select specific locations, and observe how income and burglary rates have evolved over 15 years.

6.1.2 Structure

The application consists of two main tabs:

Map — A choropleth map of Canton Zurich rendered with the `leaflet` package. The user can select any of four variables (median income, population, total burglaries, or burglary rate per 1,000 inhabitants), choose a year via a slider, and optionally exclude known outliers. Polygons are coloured using a continuous viridis palette, and hovering over a municipality displays the exact value. The map updates dynamically without reloading the page.

Data Explorer — A municipality- and district-level panel showing four time-series plots side by side: population, median income, burglary trends (broken down by type), and burglary rate per 1,000 inhabitants. All locations in the canton are accessible via a dropdown. For municipalities that underwent administrative mergers between 2015 and 2019, the income plot additionally visualises the pre-merger constituent municipalities, making the data continuity transparent to the user.

→ Link to Shiny Webapp

7 Conclusions

7.1 Key Findings

This analysis set out to test whether anecdotal observations from Basel-Landschaft – that burglaries concentrate in wealthier, border-proximate municipalities – hold in the Canton of Zurich. The results paint a more nuanced picture than expected.

Income is associated with lower burglary rates, contradicting H1. Contrary to the intuition that affluent households attract property crime, the Negative Binomial regression shows a significant negative relationship between median income and burglary rates (IRR = 0.88 per 10,000 CHF, $p < 0.001$). The exploratory scatterplot already foreshadowed this result, with an R^2 of just 0.02 and a slightly downward-sloping regression line. Wealthier municipalities in Canton Zurich do not experience systematically higher burglary rates — if anything, the opposite is true.

Border proximity is not a reliable predictor of burglary rates (H2 not supported). The exploratory analysis revealed a visually striking spatial gradient: municipalities within 10 km of the national border consistently showed higher burglary rates than those further inland. However, this pattern does not survive formal modelling. In the Negative Binomial regression, border distance is not statistically significant ($p = 0.32$). Notably, the Poisson model did find a significant effect for border distance, but this result reversed sign and lost significance once overdispersion was properly accounted for — illustrating how model misspecification can produce misleading conclusions. The visual border gradient observed in the EDA may be confounded with other spatial factors such as urbanisation, transport connectivity, or policing intensity, none of which are included in the model.

The interaction of income and border proximity is statistically detectable but practically limited (H3 partially supported). The interaction model yields a significant interaction term ($p = 1.19e-10$), suggesting that the income–burglary relationship varies with border proximity. However, the improvement in model fit over the additive NB model is marginal ($\Delta AIC = 34.2$), and the large sample size means that even small effects can reach statistical significance. We therefore characterise this as a partial finding: there is evidence of an interaction, but it does not substantially improve prediction.

Burglary rates have declined substantially over time. The year variable is the strongest and most robust predictor across all models. Burglary rates have fallen markedly since 2009 across all municipalities and all border distance categories, likely reflecting broader factors such as improved security technology, changes in policing strategies, or shifts in criminal behaviour. This temporal decline is the clearest and most consistent finding of the analysis.

7.2 Limitations

Several limitations should be considered when interpreting these results.

Commercial area density and building-use composition. Kreis 1 emerges as a clear outlier in burglary rates per 1,000 inhabitants. A likely contributing factor is its exceptionally high concentration of commercial properties. Particularly, high-value retail establishments may attract a qualitatively different type of burglary than residential break-ins. The burglary data used in this analysis does not distinguish between household and commercial premises, meaning that municipalities or districts with a high density of shops, offices, or other non-residential buildings cannot be meaningfully compared to predominantly residential areas on the basis of raw burglary rates alone. Incorporating a measure of commercial land-use density or building-use composition into the model would likely improve both its explanatory power and the interpretability of spatial patterns, particularly for urban centres such as Kreis 1.

Median income as an incomplete measure of wealth distribution. The income predictor used in this analysis reflects the median i.e., the 50th percentile of municipal income. While the median is robust to extreme values, it provides no information about the spread or skewness of the income distribution within a municipality. This is a meaningful limitation in the context of burglary research, where it may not be average wealth that attracts offenders, but the presence of high-value targets at the upper end of the distribution. A city district such as Kreis 1, for example, may have a deceptively moderate median income while simultaneously containing a disproportionate share of very high-income households and businesses. Without access to measures such as the Gini coefficient, the 90th income percentile, or the interquartile range, the model cannot capture this within-municipality heterogeneity. As a result, the income predictor may systematically underperform in economically polarised areas, and its null finding in the regression model should be interpreted with this caveat in mind.

Ecological fallacy. All analyses operate at the municipal level. Even where visual patterns suggest spatial gradients (e.g., higher burglary rates near the border), we cannot draw conclusions about individual-level behaviour from aggregate data.

Limited model explanatory power. While the Negative Binomial model is statistically appropriate for the data, the included covariates — income, border distance, and year — explain only a modest share of the variation in burglary counts. A formal goodness-of-fit measure such as McFadden's pseudo-R^2 would likely confirm that much of the municipality-level variation remains unexplained. This is consistent with the observation that important predictors are missing from the model.

Omitted variables. Neither income nor border distance proved to be robust predictors in the formal model, yet the exploratory analysis showed clear visual patterns. This suggests that important explanatory variables — such as urbanisation, transport connectivity, policing intensity, or local socioeconomic structure — are missing from the model and may confound the observed spatial patterns.

Centroid distance as a proxy. Border distance is measured from each municipality's geographic centroid to the nearest national border. This is a simplification — municipalities with irregular shapes or large areas may have actual border distances that differ substantially from the centroid measure. A more granular measure (e.g., minimum distance from any point within the municipality boundary) might yield different results.

Incomplete income data. Income data are unavailable for 2023–2024, limiting the temporal scope of income-related analyses to 2009–2022. While the missingness mechanism is well understood and documented in Section 3.2, it means our models cannot capture the most recent developments.

Single canton. The analysis is restricted to Canton Zurich. The patterns observed here may not generalise to other cantons, particularly those with different border configurations (e.g., Basel-Landschaft, Geneva, or Ticino) or different economic structures.

8 Generative AI Reflection

AI proved to be a powerful tool that definitely supported us in crafting this project. While it was important to us to understand the concepts of R and exciting to dive into the statistics and explore what stories the data could tell, we used available tools such as ChatGPT and Claude to debug code, improve syntax and writing, and broaden our understanding of what we might research with the given data. In all cases, we tried to validate outputs against course materials rather than accepting them uncritically.

That said, especially agentic AI left us stunned by the potential and automation it brings. Watching an agent autonomously iterate on code, diagnose issues, and propose fixes was both impressive and unsettling. The temptation to delegate entire analytical steps was real, and resisting it required conscious effort — we deliberately wrote core model specifications, data joins, and visualisations ourselves, using AI only to accelerate where we were already confident in the direction.

This leaves us with a broader reflection. AI, and especially the upcoming agents, have the potential to change the way we work in a lasting manner. But even as we develop tools that make it easy to outsource tasks, we shouldn't forget that every step builds on the one before it — and that we, whether as a society or individually, need to keep investing in our own progress and our desire to learn. The Applied Information and Data Science programme definitely sits at a useful intersection here: building the foundation needed to understand how AI tools actually work, while learning to apply them effectively, feels increasingly valuable. Finally, the question shouldn't only be what we'll occupy ourselves with, but how we find a way to coexist with algorithms and machines in this disruptive age — following our curiosity and ultimately not losing the spark that brought us here in the first place.