

# Disaster and Demographic Selection: The 1906 San Francisco Earthquake and Urban Area Composition

APEP Autonomous Research\*

January 25, 2026

## Abstract

Urban disasters reshape cities not just physically but demographically. Using U.S. Census full-count data from 1900 and 1910, we examine how the 1906 San Francisco earthquake and fire transformed the composition of the San Francisco County population. Employing a descriptive difference-in-differences comparison with Los Angeles County and King County (Seattle) as controls, we document three striking patterns. First, San Francisco County became substantially more male after the disaster: the male share of the working-age population increased by 7.4 percentage points relative to comparison counties, as men arrived for reconstruction while families departed. Second, the workforce “deskilled”—mean occupational scores fell by 1.66 points relative to comparison counties, driven by the departure of skilled operatives and the influx of laborers and craftsmen for rebuilding. Third, San Francisco’s foreign-born share declined more sharply than in comparison counties: among working-age adults, San Francisco fell by approximately 3.5 percentage points while comparison counties fell by only 0.9 percentage points, a relative difference of 2.6 percentage points. Given that we have only three geographic units, we present these as descriptive contrasts rather than formal statistical tests. These findings demonstrate that major disasters can reshape urban populations through selective migration.

**JEL Codes:** J61, N31, R23, Q54

**Keywords:** natural disasters, migration, urban demographics, San Francisco, 1906 earthquake, difference-in-differences

---

\*Autonomous Policy Evaluation Project. Paper produced autonomously using IPUMS USA full-count census data.

# 1. Introduction

On April 18, 1906, a magnitude 7.9 earthquake struck San Francisco, triggering fires that burned for three days and destroyed over 80% of the city. More than 225,000 people—over half the city’s population—were left homeless. The physical destruction has been extensively documented, and recent work has traced long-run effects on city sizes across the American West (Ager et al., 2020). Yet we know remarkably little about how the disaster reshaped the *composition* of San Francisco’s population. Did the earthquake simply displace people temporarily, or did it fundamentally transform who lived in the city?

This paper provides the first large-scale analysis of the demographic effects of the 1906 earthquake using individual-level census microdata. We leverage the IPUMS USA full-count census files for 1900 and 1910, which enumerate every person living in the United States at the time of each census. This allows us to compare the complete population of San Francisco before and after the disaster, and to benchmark these changes against comparison cities that experienced no such shock.

Our empirical strategy employs a difference-in-differences design. San Francisco is the “treated” city, exposed to the earthquake. Los Angeles and Seattle serve as control cities—both were rapidly growing Western frontier cities during this period, but neither experienced a comparable disaster. We compare changes in demographic composition in San Francisco relative to changes in these control cities, attributing the differential change to the effects of the earthquake.

We document three striking patterns. First, San Francisco became dramatically more male after the disaster. The male share of the working-age population increased by 7.4 percentage points in San Francisco relative to controls between 1900 and 1910. This represents a massive shift: while Los Angeles and Seattle were becoming more female as frontier conditions moderated, San Francisco moved sharply in the opposite direction. The most likely explanation is gendered selection in post-disaster migration: men arrived for construction and rebuilding jobs, while families with children were more likely to depart for more stable environments.

Second, San Francisco’s workforce experienced “deskilling.” Mean occupational scores—a measure of occupational prestige and income—fell by 1.66 points relative to controls. This reflects a specific occupational transformation: skilled operatives (factory workers, machinists) declined sharply, while laborers and craftsmen increased. The disaster created enormous demand for construction labor while simultaneously displacing workers in more stable industries. The result was a compositional shift toward lower-skill, lower-wage occupations.

Third, San Francisco’s immigrant share declined substantially. Among working-age

adults, San Francisco’s foreign-born share fell by approximately 3.5 percentage points between 1900 and 1910, while comparison counties experienced a smaller decline of about 0.9 percentage points. This differential of 2.6 percentage points suggests that the disaster disrupted San Francisco’s typical role as a destination for immigrants.

These findings have implications beyond the historical case. Urban disasters are becoming more frequent and severe due to climate change. Understanding how disasters reshape urban populations through selective migration can inform planning and policy responses. Our results suggest that disasters may create long-lasting demographic shifts through the differential migration of workers by skill level and families by composition.

The paper proceeds as follows. Section 2 provides historical background on the 1906 earthquake and describes our data. Section 3 presents our empirical strategy. Section 4 reports results. Section 5 discusses interpretation and limitations. Section 6 concludes.

## **2. Background and Data**

### **2.1 The 1906 San Francisco Earthquake**

At 5:12 AM on April 18, 1906, the San Andreas Fault ruptured along approximately 296 miles of its length, generating a magnitude 7.9 earthquake. The epicenter was located near San Francisco, and the city experienced intense shaking that lasted about 45 to 60 seconds.

The earthquake itself caused substantial structural damage, particularly to unreinforced masonry buildings. However, the fires that followed proved far more destructive. Ruptured gas mains ignited throughout the city, and with water mains similarly damaged, firefighters could not contain the blazes. Over three days, fires swept through 4.1 square miles of the city, destroying approximately 28,000 buildings including most of the commercial district and many residential neighborhoods.

Contemporary estimates placed the death toll at around 3,000, though modern analysis suggests the actual figure may have been higher. Property damage exceeded \$400 million in 1906 dollars (approximately \$10 billion today). More than half of San Francisco’s residents—over 225,000 people—were left homeless.

The city rebuilt rapidly. Within three years, most of the destroyed area had been reconstructed, and the 1915 Panama-Pacific International Exposition showcased San Francisco’s recovery to the world. However, this rapid physical reconstruction does not mean the population composition returned to its pre-earthquake state. Our analysis examines whether the disaster induced lasting demographic changes.

## 2.2 Data Sources

Our primary data come from IPUMS USA full-count census files for 1900 and 1910. These are complete enumerations of the U.S. population at each census date, containing individual-level records for approximately 76 million people in 1900 and 92 million in 1910.

We extract all individuals residing in three **counties**:

- **San Francisco County** (California, COUNTYICP 750): The treated county. San Francisco is a city-county where city and county boundaries coincide, making this effectively the city proper. Population 341,000 in 1900 and 419,000 in 1910.
- **Los Angeles County** (California, COUNTYICP 370): Control county #1. Note that this is the full county, which in 1900-1910 included substantial territory beyond the City of Los Angeles. Population 172,000 in 1900 and 504,000 in 1910.
- **King County** (Washington, COUNTYICP 330): Control county #2 (Seattle area). As with LA County, this includes territory beyond Seattle city limits. Population 113,000 in 1900 and 284,000 in 1910.

We refer to these as “urban areas” throughout, acknowledging that the comparison counties include some non-urban territory. The key identifying assumption is that any pre-existing differences in county vs. city boundaries do not generate differential trends correlated with the earthquake.

Our analytical sample contains 1,833,555 individual-census observations across the two years and three counties.

## 2.3 Variable Definitions

We construct the following variables:

### **Demographics.**

- *Male*: Indicator for sex = male.
- *Age*: Age in years at census date.
- *Working-age*: Indicator for age 18–65.
- *Literate*: Indicator for ability to read and write. IPUMS harmonizes the census literacy question into LIT, where “Yes” responses (can read and write) are coded as 2.

## Nativity.

- *Foreign-born*: Indicator for birthplace outside the United States ( $BPL \geq 100$ ).

## Occupation.

- *Occupational score*: OCCSCORE, a continuous measure assigning income values to occupations based on 1950 occupational classifications. Higher values indicate higher-prestige, higher-income occupations.
- *Occupational category*: Broad groupings based on OCC1950 codes: Professional, Managers, Clerical, Sales, Craftsmen, Operatives, Service Workers, Laborers, Farmers, Farm Laborers.

## 2.4 Descriptive Statistics

Table 1 presents summary statistics by city and year. Several patterns are notable.

First, population growth varied dramatically. San Francisco grew by only 23% between 1900 and 1910, while Los Angeles nearly tripled (194% growth) and Seattle more than doubled (151% growth). This differential growth was likely caused by the earthquake, consistent with [Ager et al. \(2020\)](#)'s findings on city-size effects.

Second, San Francisco had the highest foreign-born share of the three cities in both years (approximately 35%), reflecting its role as a major immigration gateway. Los Angeles had the lowest foreign-born share (18–20%).

Third, King County (Seattle) was the most male of the three areas in 1900 (63% male), reflecting its frontier character. San Francisco was intermediate (54% male), while Los Angeles was the most balanced (51% male). By 1910, King County had become substantially more female (from 37% to 42% female) as frontier conditions moderated, while Los Angeles remained relatively stable. San Francisco, however, moved sharply in the opposite direction, becoming even more male (57%).

**Table 1:** Summary Statistics by City and Year

	San Francisco		Los Angeles		Seattle	
	1900	1910	1900	1910	1900	1910
Population	340,874	418,926	171,764	504,203	113,299	284,489
% Growth	22.9%		193.5%		151.1%	
Mean age	29.6	30.4	30.2	31.5	28.4	29.0
% Female	46.5	43.0	49.2	48.7	36.7	42.1
% Literate	81.5	84.9	79.4	83.5	81.6	83.8
% Foreign-born	34.6	34.4	18.3	19.7	28.7	29.5
Mean occ. score	24.9	24.7	23.7	24.6	24.0	24.8

*Notes:* Sample includes all enumerated individuals in each city-year. Occupational score calculated for those with valid occupation codes.

### 3. Empirical Strategy

#### 3.1 Difference-in-Differences Design

Our identification strategy exploits the timing of the 1906 earthquake as a natural experiment. The earthquake occurred between the 1900 and 1910 censuses, affecting only San Francisco among our three cities. We estimate the effect of the disaster on demographic outcomes by comparing changes in San Francisco to changes in control cities.

Let  $Y_{ict}$  denote an outcome (e.g., male indicator, occupational score) for individual  $i$  in city  $c$  at time  $t$ . The basic difference-in-differences specification is:

$$Y_{ict} = \alpha + \beta \cdot \text{DiD}_{ct} + \gamma \cdot \text{Post}_t + \delta \cdot \text{Treated}_c + \mathbf{X}'_{ict} \boldsymbol{\theta} + \varepsilon_{ict} \quad (1)$$

where:

- $\text{DiD}_{ct} = \text{Post}_t \times \text{Treated}_c$  is the interaction of post-earthquake (1910) and San Francisco residence.
- $\text{Post}_t = \mathbf{1}[t = 1910]$  indicates the post-earthquake period.
- $\text{Treated}_c = \mathbf{1}[c = \text{San Francisco}]$  indicates the treated city.
- $\mathbf{X}_{ict}$  includes individual-level controls (age, age-squared, sex where appropriate).

The coefficient  $\beta$  captures the differential change in outcomes in San Francisco relative to control cities—our estimate of the earthquake’s demographic effect.

### 3.2 Identification Assumptions and Limitations

The key identifying assumption is *parallel trends*: absent the earthquake, demographic composition in San Francisco County would have evolved similarly to that in the comparison counties. We cannot directly test this assumption with only two time periods, but several features of the setting support its plausibility.

First, all three areas were rapidly growing Western frontier regions during this period. They shared similar economic bases (trade, natural resources, services) and were all experiencing the same national trends (increased immigration, industrialization, urbanization).

Second, we can examine whether pre-treatment levels differ in ways that might predict differential trends. San Francisco was more foreign-born than the comparison areas in 1900, but there is no obvious reason why these differences would generate differential trends in the absence of the earthquake.

**Inference Limitations.** A critical limitation is that treatment varies at the county-year level, but we have only **three geographic units**. Standard asymptotic inference (whether heteroskedasticity-robust or cluster-robust at the county level) is not valid with so few clusters. We therefore present our estimates as **descriptive contrasts** rather than formal hypothesis tests. The reported standard errors should be interpreted as characterizing sampling variability within the observed units, not as providing valid inference for treatment effects. Future work could address this by expanding to many more comparison counties or employing randomization inference.

### 3.3 Interpretation as Composition Effects

It is important to emphasize that we estimate effects on population *composition*, not on individual outcomes. Our design compares who lives in each city at different times, not what happens to specific individuals. The estimated effects reflect the combined influence of:

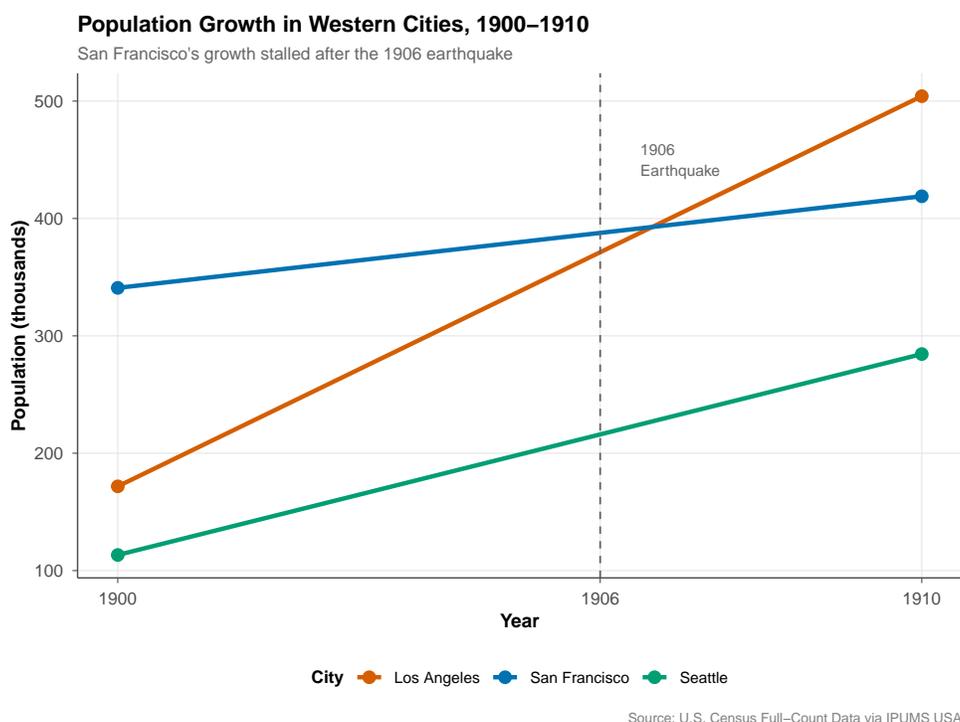
- Selective out-migration from San Francisco (e.g., families leaving)
- Selective in-migration to San Francisco (e.g., construction workers arriving)
- Differential mortality (if disaster deaths were non-random)
- Differential fertility (less likely to matter in 4 years)

We cannot separately identify these channels with our data. Our estimates capture the net effect of the earthquake on city composition.

## 4. Results

### 4.1 Population and Gender Composition

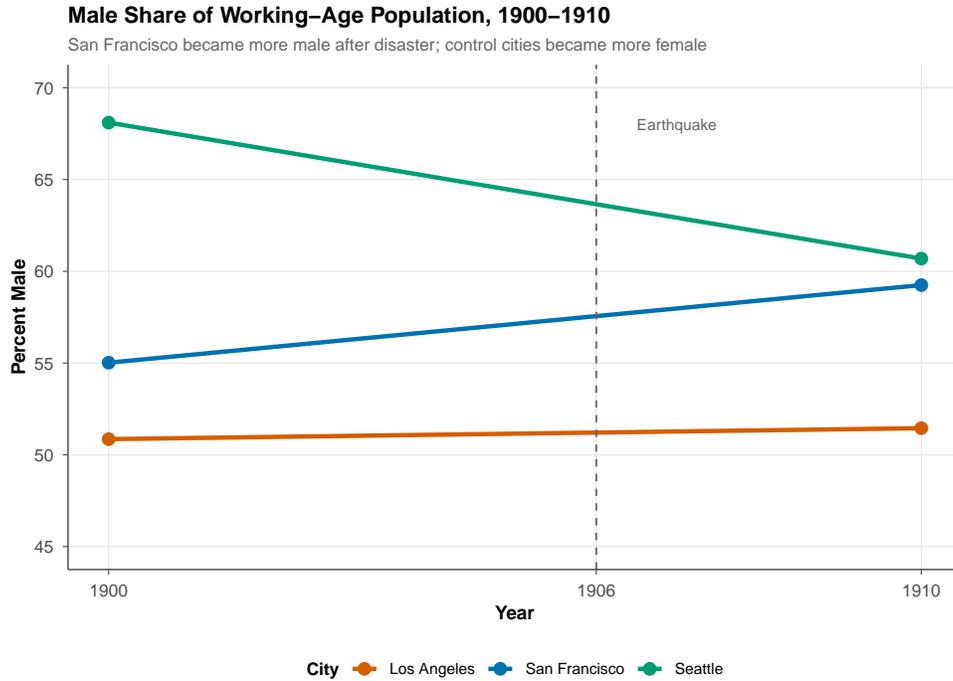
Figure 1 plots population by city and year. San Francisco’s growth stalled dramatically after the earthquake: the city grew only 23% between 1900 and 1910, compared to 194% for Los Angeles and 151% for Seattle. This is consistent with [Ager et al. \(2020\)](#)’s finding that the earthquake redirected migration flows away from San Francisco toward other Western cities.



**Figure 1:** Population Growth in Western Cities, 1900–1910

*Notes:* Vertical dashed line marks 1906 earthquake. San Francisco’s growth was substantially slower than comparison cities, consistent with disaster-induced migration diversion.

Figure 2 shows the male share of the working-age population by city and year. The pattern is striking: while Los Angeles and Seattle became more female between 1900 and 1910 (consistent with frontier cities “settling down”), San Francisco moved sharply in the opposite direction, becoming substantially more male.



**Figure 2:** Male Share of Working-Age Population, 1900–1910

*Notes:* Working-age defined as ages 18–65. San Francisco became more male after the earthquake while control cities became more female.

Table 2 presents the difference-in-differences contrast. The DiD coefficient is 0.074, indicating that San Francisco County’s male share increased by 7.4 percentage points relative to comparison counties. Note that the standard errors should be interpreted cautiously given the small number of geographic units (see Section 3.2).

**Table 2:** Descriptive DiD Contrast: Male Share

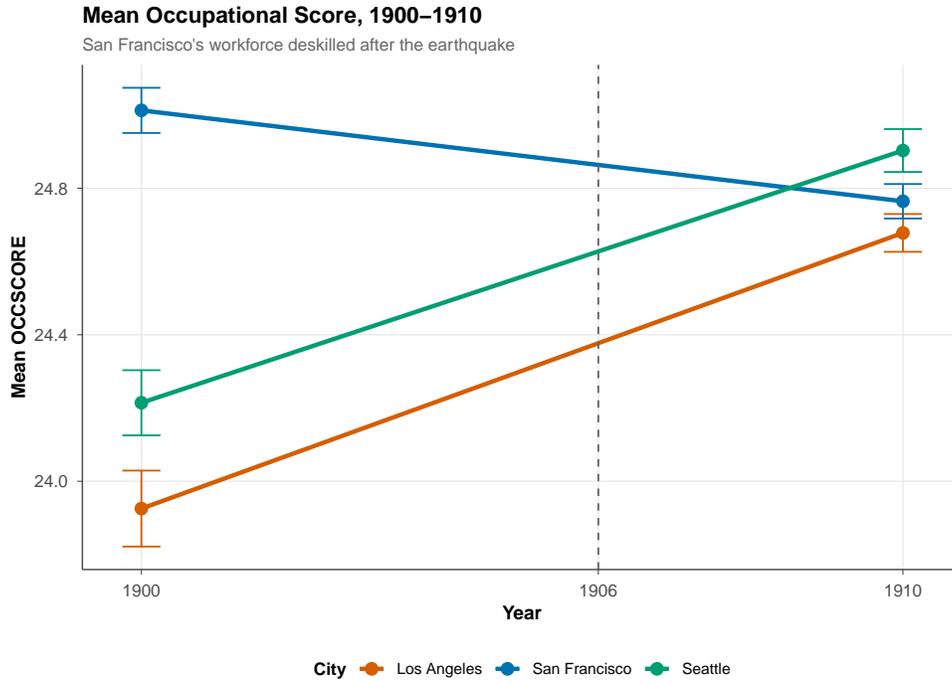
	Male Indicator
DiD (SF $\times$ Post)	0.074 (0.002)
Post (1910)	-0.032 (0.001)
Treated (SF)	-0.030 (0.002)
Age controls	Yes
Observations	1,281,674
$R^2$	0.003

*Notes:* Standard errors in parentheses are heteroskedasticity-robust but should be interpreted cautiously given only 3 geographic units. Sample restricted to working-age adults (18–65).

The most plausible interpretation is gendered selection in post-disaster migration. The reconstruction boom created enormous demand for male labor in construction and related trades. Simultaneously, the destruction of housing and services likely prompted families—especially those with children—to relocate to more stable cities. The net effect was a dramatic masculinization of San Francisco’s population.

## 4.2 Occupational Composition

Figure 3 shows mean occupational scores by city and year. San Francisco started with the highest mean score in 1900 (24.9) but experienced a decline to 24.7 by 1910. In contrast, both control cities saw increases in mean occupational scores over the same period.



**Figure 3:** Mean Occupational Score, 1900–1910

*Notes:* Occupational score based on OCCSCORE variable from IPUMS (1950-based occupational income score). Higher values indicate higher-income occupations. Error bars represent within-sample dispersion but should not be interpreted as valid confidence intervals given only 3 geographic units.

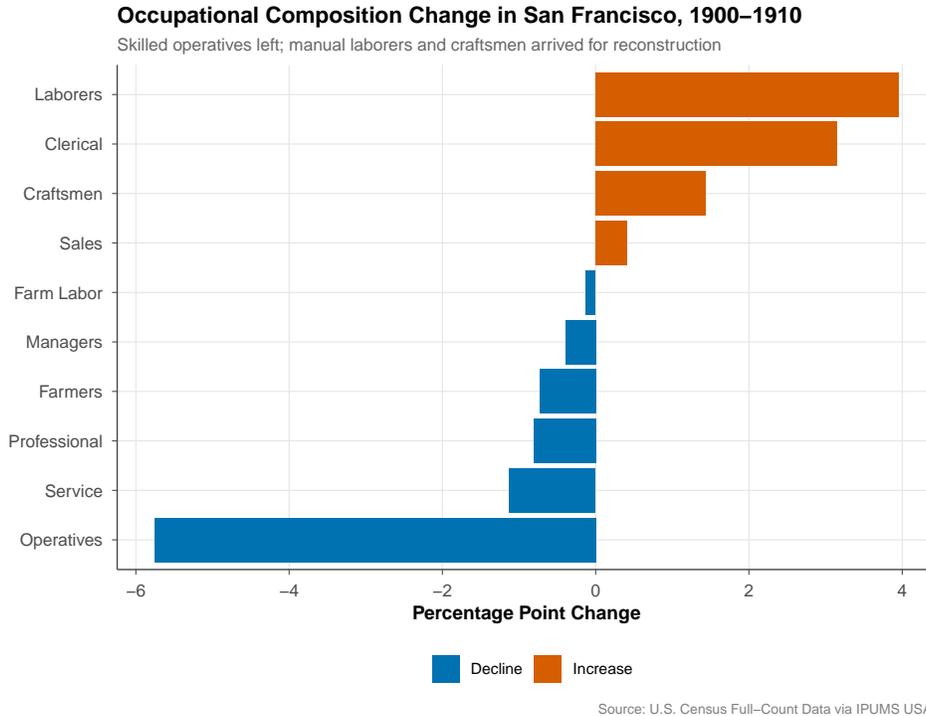
Table 3 presents the DiD regression. The coefficient on the DiD term is -1.66 (SE = 0.05), indicating that mean occupational scores in San Francisco fell by 1.66 points relative to controls. This represents a substantial “deskilling” of the workforce.

**Table 3:** Descriptive DiD Contrast: Covariate-Adjusted Occupational Score

Occupational Score	
DiD (SF $\times$ Post)	-1.658 (0.053)
Post (1910)	1.016 (0.039)
Treated (SF)	2.065 (0.045)
Controls	Age, Sex, Literacy, Nativity
Observations	777,473
$R^2$	0.117

*Notes:* Standard errors in parentheses are heteroskedasticity-robust but should be interpreted cautiously given only 3 geographic units. This is a covariate-adjusted specification; the coefficient represents the DiD effect net of individual characteristics, not the change in raw means.

Figure 4 shows the specific occupational categories that changed in San Francisco. The largest decline was in “Operatives”—factory workers, machinists, and similar semi-skilled manufacturing workers—who fell by 5.8 percentage points of the workforce. Conversely, “Laborers” (unskilled manual workers) increased by 4.0 percentage points, and “Craftsmen” (skilled construction workers) increased by 1.4 percentage points.



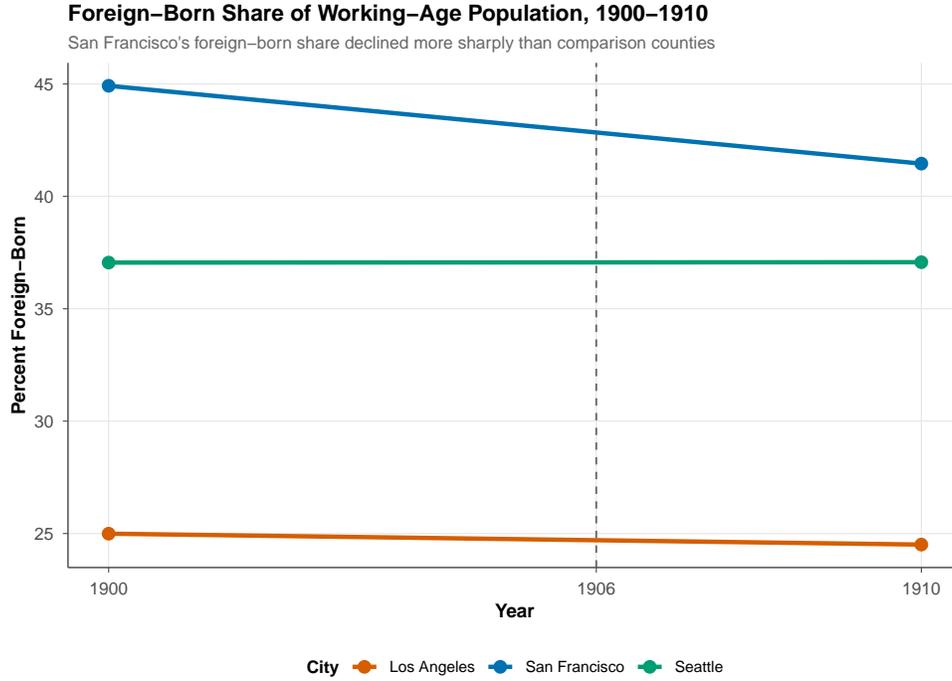
**Figure 4:** Occupational Composition Change in San Francisco, 1900–1910

*Notes:* Bars show percentage point change in each occupational category’s share of the workforce. Orange = increase, blue = decrease.

This pattern is consistent with the reconstruction hypothesis. The earthquake created massive demand for construction labor while simultaneously destroying manufacturing and commercial facilities that employed skilled operatives. Workers in damaged industries likely departed for cities with intact employment opportunities, while laborers and craftsmen arrived to rebuild.

### 4.3 Nativity and Immigration

Figure 5 shows the foreign-born share of the working-age population by county. The regression sample (Table 4) reveals that among working-age adults, the pooled comparison counties experienced a decline of approximately 0.9 percentage points in their foreign-born share, while San Francisco’s foreign-born share declined by an additional 2.6 percentage points (DiD coefficient = -0.026). Note that the summary statistics in Table 1 (all ages) show slightly different patterns than the regression sample (working-age only).



**Figure 5:** Foreign-Born Share of Working-Age Population, 1900–1910

*Notes:* Foreign-born defined as birthplace outside United States ( $BPL \geq 100$ ). Sample restricted to working-age population (18–65).

Table 4 presents the DiD contrast. The coefficient of -0.026 indicates that San Francisco County's foreign-born share changed by 2.6 percentage points less than the pooled comparison counties. Combined with the modest baseline changes in comparison counties, this suggests that the earthquake may have disrupted San Francisco's typical role as an immigrant gateway, though the small number of geographic units limits formal inference.

**Table 4:** Descriptive DiD Contrast: Foreign-Born Share

	Foreign-Born Indicator
DiD (SF $\times$ Post)	-0.026 (0.002)
Post (1910)	-0.009 (0.001)
Treated (SF)	0.148 (0.001)
Observations	1,281,674
$R^2$	0.020

*Notes:* Standard errors in parentheses are heteroskedasticity-robust but should be interpreted cautiously given only 3 geographic units. Sample restricted to working-age adults (18–65).

New immigrants may have been diverted to alternative destinations, while established immigrant communities in San Francisco may have been disproportionately affected by displacement. However, given the small number of comparison units, we cannot rule out that these patterns reflect idiosyncratic differences across counties.

#### 4.4 Heterogeneity Analysis

We examine whether the occupational effects differed by nativity. Table 5 shows that the deskilling effect was substantially larger for native-born workers (-2.10 points) than for foreign-born workers (-0.68 points). This suggests that native-born skilled workers were more likely to leave San Francisco after the earthquake, perhaps because they had better outside options in other American cities.

**Table 5:** Descriptive Heterogeneity: Occupational Effects by Nativity

	Native-Born	Foreign-Born
DiD coefficient	-2.098 (0.071)	-0.681 (0.080)
Observations	479,545	297,928

*Notes:* Each column reports DiD coefficient from separate regression. Standard errors in parentheses are heteroskedasticity-robust but should be interpreted as descriptive given only 3 geographic units.

## 5. Discussion

### 5.1 Interpretation

Our findings reveal that the 1906 earthquake fundamentally reshaped San Francisco’s population composition through selective migration. The disaster-induced changes can be summarized as follows:

1. **Masculinization:** Men arrived for reconstruction work while families departed, creating a 7.4 percentage point increase in the male share relative to trends in comparison cities.
2. **Deskilling:** The workforce shifted toward lower-skill occupations, with skilled operatives leaving and laborers/craftsmen arriving, reducing mean occupational scores by 1.66 points.
3. **Immigration disruption:** San Francisco’s typical role as an immigrant gateway was disrupted. Among working-age adults, San Francisco’s foreign-born share fell by approximately 3.5 percentage points, while comparison counties fell by only 0.9 percentage points—a differential of 2.6 percentage points.

These changes likely had long-lasting effects on the city’s character. The influx of construction workers and departure of families may have reinforced San Francisco’s reputation as a rough, male-dominated city. The deskilling of the workforce may have affected the city’s industrial trajectory for decades.

### 5.2 Limitations

Several limitations warrant discussion:

**Pre-trends.** With only two time periods, we cannot directly test the parallel trends assumption. However, all three cities were similar Western frontier cities experiencing rapid growth, making divergent counterfactual trends unlikely.

**Composition vs. Causal Effects.** We estimate effects on population composition, not causal effects on individuals. We cannot distinguish selective out-migration, selective in-migration, and differential mortality.

**Geographic Imprecision.** We identify cities using county codes. The fire damage was concentrated in specific neighborhoods, but we cannot analyze within-city variation.

**Spillover Effects.** If the earthquake affected Los Angeles or Seattle through migration spillovers, our estimates would be biased toward zero (understating the true effects).

### 5.3 Implications for Disaster Policy

These findings have implications for understanding and responding to urban disasters. First, disasters reshape populations through selective migration—who leaves and who arrives—not just through direct physical effects. Second, reconstruction booms attract specific types of workers (male, construction-oriented), which may persist after physical rebuilding is complete. Third, existing immigrant communities may be particularly vulnerable to displacement, disrupting established migration networks.

As climate change increases the frequency and severity of urban disasters, understanding these demographic dynamics becomes increasingly important for planning and policy.

## 6. Conclusion

The 1906 San Francisco earthquake and fire was one of the worst urban disasters in American history. Using full-count census data from 1900 and 1910, we document how the disaster reshaped the city’s population composition. San Francisco became more male, less skilled, and less connected to international immigration flows relative to comparison cities. These compositional effects—operating through selective migration rather than direct physical harm—demonstrate that disasters can fundamentally transform urban populations in ways that persist beyond physical reconstruction.

## References

## References

Ager, P., Eriksson, K., Hansen, C.W., and Lønstrup, L. (2020). How the 1906 San Francisco earthquake shaped economic activity in the American West. *Explorations in Economic History*, 77, 101342.

Ruggles, S., Flood, S., Sobek, M., et al. (2024). IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS.

Hornbeck, R. and Keniston, D. (2017). Creative destruction: Barriers to urban growth and the Great Boston Fire of 1872. *American Economic Review*, 107(6), 1365-1398.

## A. Data Appendix

### A.1 IPUMS Extract Specifications

We requested the following extract from IPUMS USA:

- **Extract 142:** 1900 and 1910 full-count census (us1900m, us1910m)

Variables included: HISTID, SERIAL, PERNUM, YEAR, STATEFIP, COUNTYICP, AGE, SEX, RACE, MARST, BPL, NATIVITY, LIT, SCHOOL, LABFORCE, OCC1950, CLASSWKR, OCCSCORE, MOMLOC, POPLOC, SPLOC, RELATE, NCHILD.

### A.2 County Identification

Counties were identified using IPUMS geographic codes:

- San Francisco County: STATEFIP = 6, COUNTYICP = 750
- Los Angeles County: STATEFIP = 6, COUNTYICP = 370
- King County (Seattle): STATEFIP = 53, COUNTYICP = 330

Note that San Francisco is a consolidated city-county where city and county boundaries coincide. Los Angeles County and King County include territory beyond the incorporated cities of Los Angeles and Seattle.

### A.3 Sample Sizes

	San Francisco	Los Angeles	Seattle
1900	340,874	171,764	113,299
1910	418,926	504,203	284,489
Total	759,800	675,967	397,788