

# Freedom on the Rails: Immigration and Abolitionist Voting in Historical Brazil

HAZEL QUANTITATIVE ANALYSIS CENTER



## Overview

This project introduces new station-level data tracing Brazil’s railway expansion over time. By harmonizing 2,527 station records with historical boundaries, census data, and abolition-era roll-call votes, it creates comparable measures of railway access and links them to and political outcomes. The analysis asks whether whether railways help predict immigrants' sorting across geographical locations. (Immigrants sorting predict voting outcomes.) Preliminary evidence shows that districts with greater railway exposure had higher foreign-born population shares in 1890, and the regression analysis further tests whether this railway-linked immigrant exposure was more politically important where slavery was locally salient. More broadly, the dataset provides a reusable foundation for studying the long-term economic and political effects of infrastructure development.

## 01. QUALITY CONTROL

- For each imperfectly matched station information from webscraping:
- Trace its source through URL of Brazil official website
- Validate each Provincia–Distrito–Município combination
- Trace post-1872 municipalities to a documented 1872 ancestor
- Record the correction, evidence, and final verdict

## DATA. MEASURES & COVERAGE

|                 |                                |                        |
|-----------------|--------------------------------|------------------------|
| 2,527           | 1,152                          | 2,461                  |
| Station records | Records needed for manually QC | Valid records after QC |

**Limitations:** Some station records could not be verified due to outdated or incomplete online sources. We used Google Maps to identify original station names, but unresolved cases were excluded from the dataset.

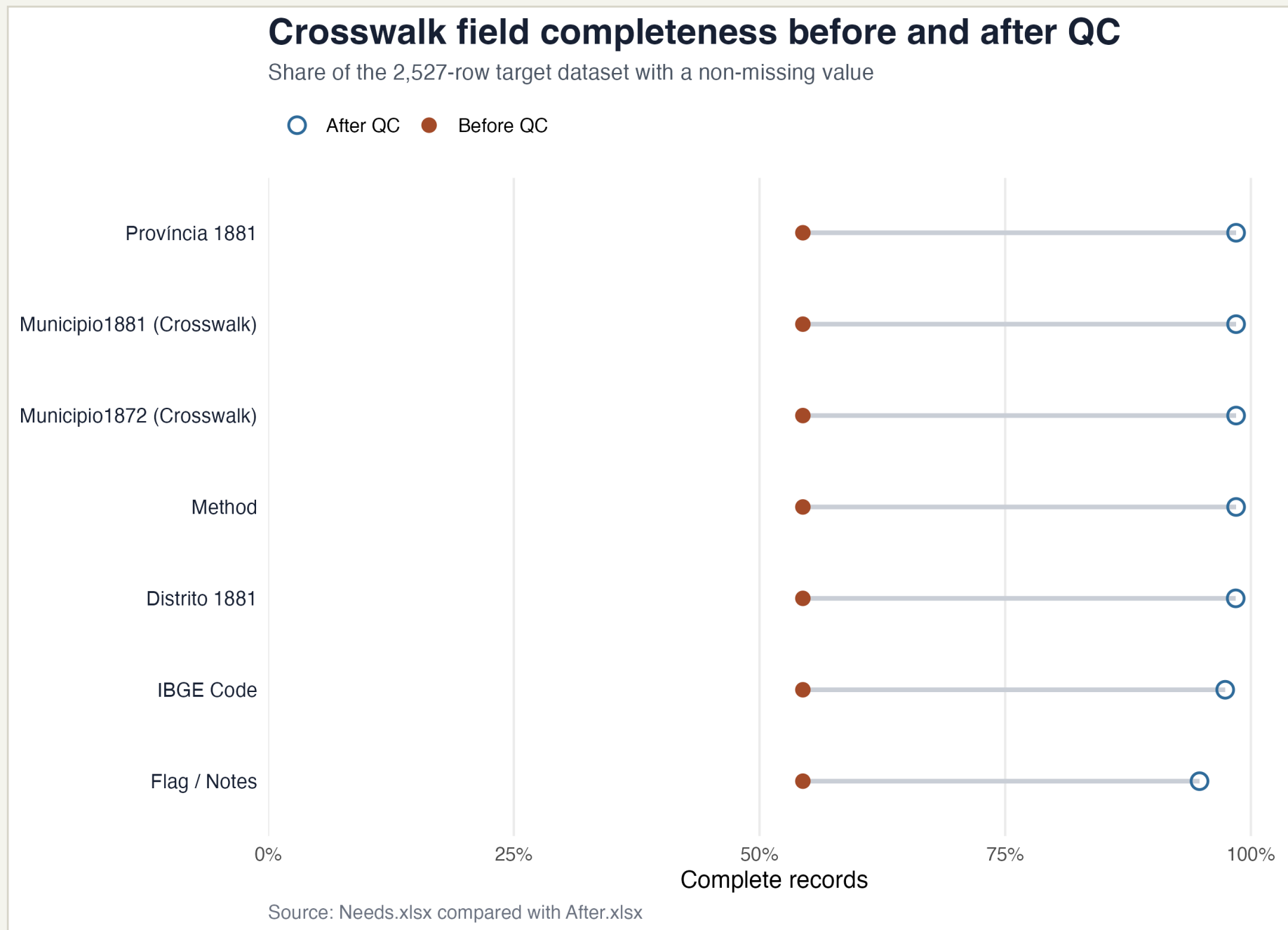
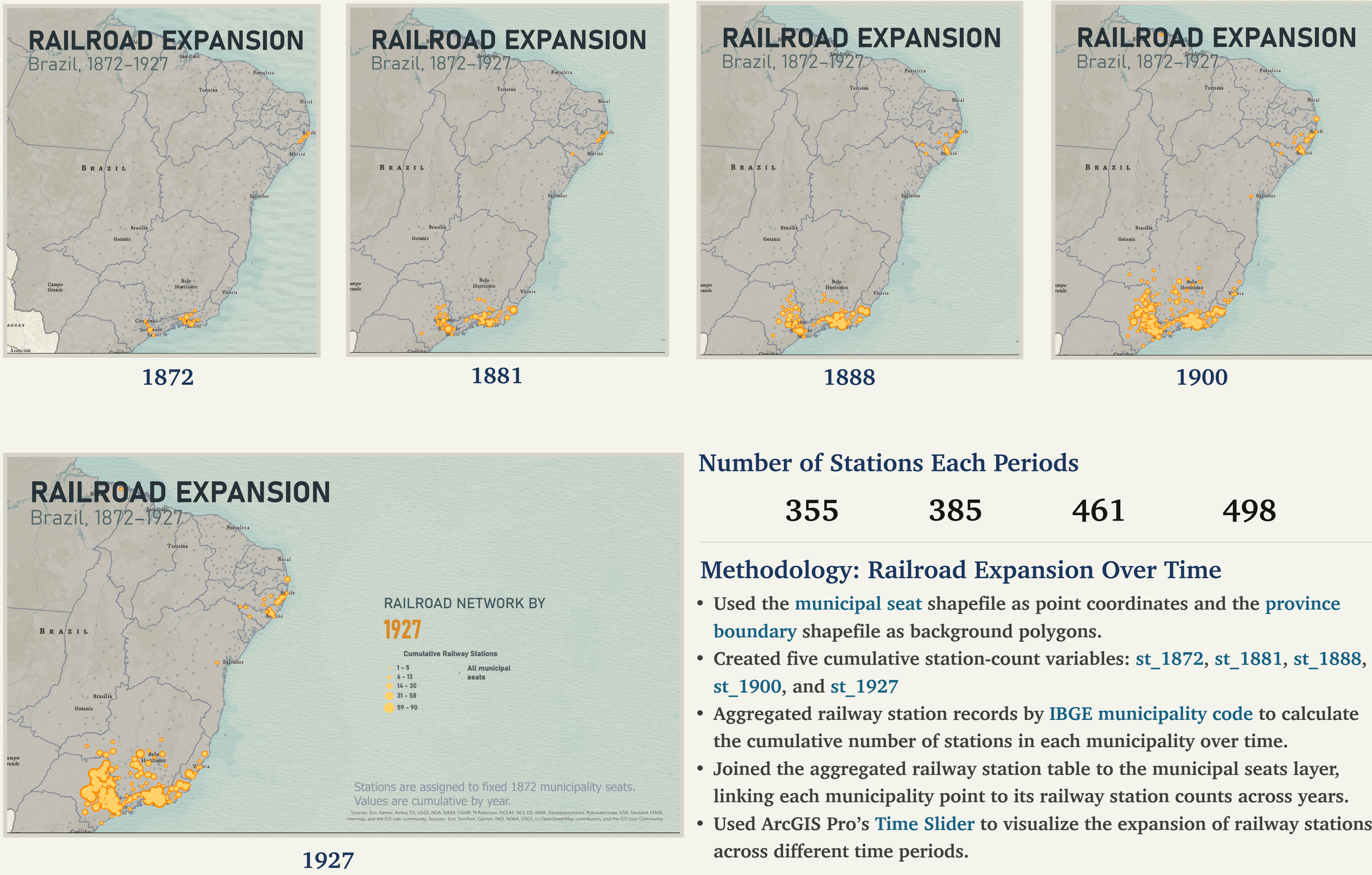


Figure 1.1 Comparison between before and after QC

## 02. DATA VISUALIZATION

### CUMULATIVE RAILWAY EXPANSION · THE 1872–1927 SERIES



### Number of Stations Each Periods

|     |     |     |     |
|-----|-----|-----|-----|
| 355 | 385 | 461 | 498 |
|-----|-----|-----|-----|

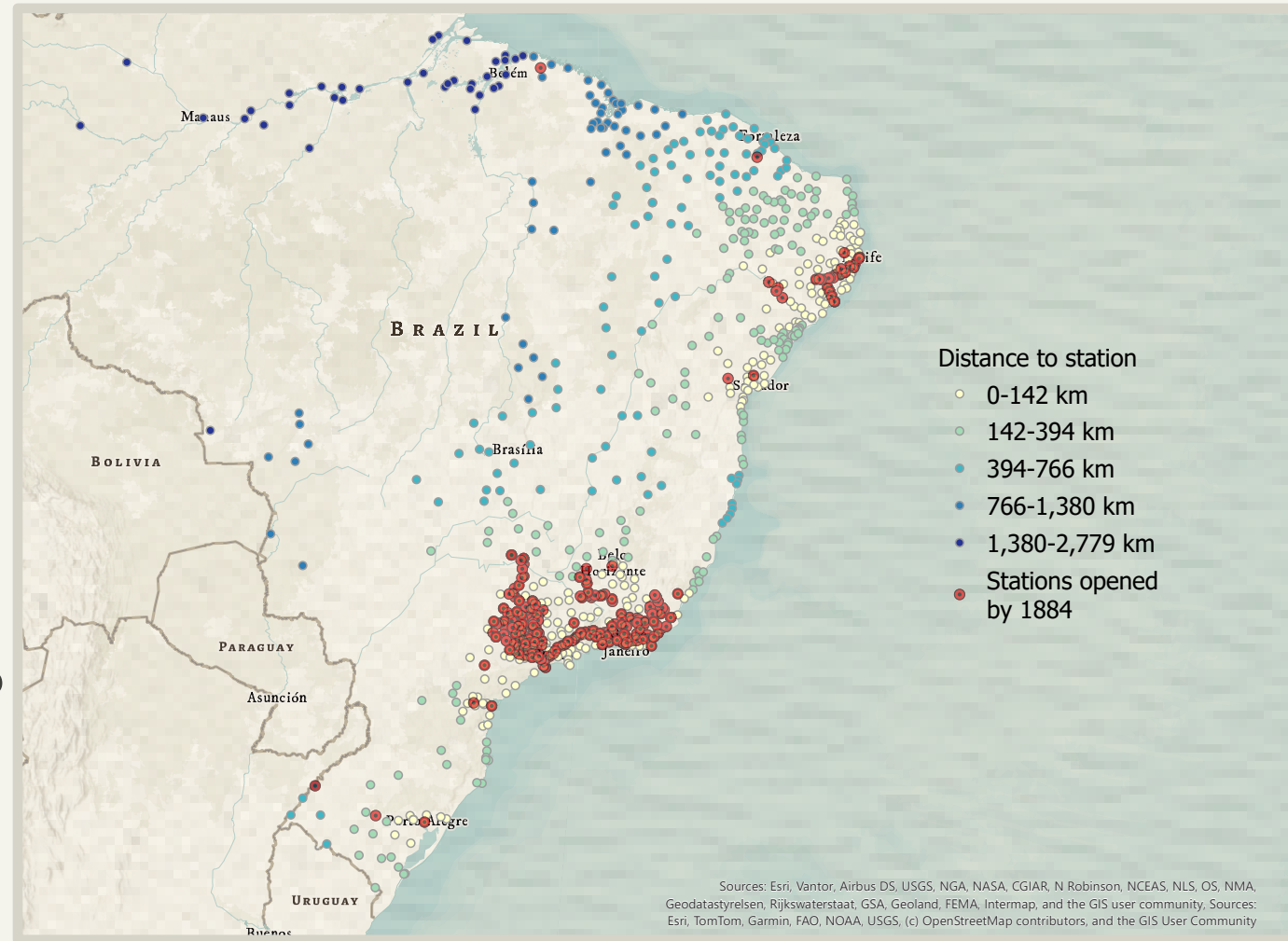
### Methodology: Railroad Expansion Over Time

- Used the [municipal seat](#) shapefile as point coordinates and the [province boundary](#) shapefile as background polygons.
- Created five cumulative station-count variables: `st_1872`, `st_1881`, `st_1888`, `st_1900`, and `st_1927`
- Aggregated railway station records by [IBGE municipality code](#) to calculate the cumulative number of stations in each municipality over time.
- Joined the aggregated railway station table to the municipal seats layer, linking each municipality point to its railway station counts across years.
- Used ArcGIS Pro’s [Time Slider](#) to visualize the expansion of railway stations across different time periods.

### DISTANCE BETWEEN STATIONS AND MUNICIPALITIES: FROM STATIONS TO PANEL

- Station records → Opening-year filters → Nearest distance → District-year mean → Session panel
- Cleaned the railway station data by keeping stations with valid coordinates and opening years.
  - Imported 1872 municipality locations and railway stations into GIS as point layers.
  - Projected both layers into a metric coordinate system to calculate distances accurately.
  - For each voting year, kept only railway stations that had already opened by that year.
  - Calculated each municipality’s distance to the nearest open railway station.
  - Aggregated municipality-level distances to the district-period level and merged them into the main voting panel by Ind and Period.
  - Created the map by symbolizing municipality points by distance bins and overlaying railway stations opened by 1884.

Figure 2.1 The Development of railway stations in Brazil across periods (1872–1927)



## 03.1 Regression: Methodology

### Central Question:

Can railway-linked immigrant exposure predict abolitionist political outcomes?

### Data.

- The dataset combines Imperial parliamentary roll calls, legislator biographies, the 1872 and 1890 Brazilian censuses, historical maps, and georeferenced records.
- It contains 1,586 district-vote observations, covering 122 electoral districts and 13 emancipation-related votes between 1884 and 1888. Because of legislator absences, the baseline voting outcome is available for 1,284 observations
- Key variables include abolition voting, immigration and slavery shares, agricultural and geographic conditions, legislators’ characteristics, and railway exposure measured using historical station locations and opening years.

### Model Constructions: 5 Stages Methodologies

#### Railways and Immigrant Exposure

- Test whether railway access predicts the foreign- population share in 1890, establishing the demographic channel from railways to immigration.

#### Baseline OLS Political Regression

- Examine whether immigrant exposure predicts abolitionist political outcomes and whether this relationship is stronger in areas with larger enslaved populations.

#### Bartik IV Models

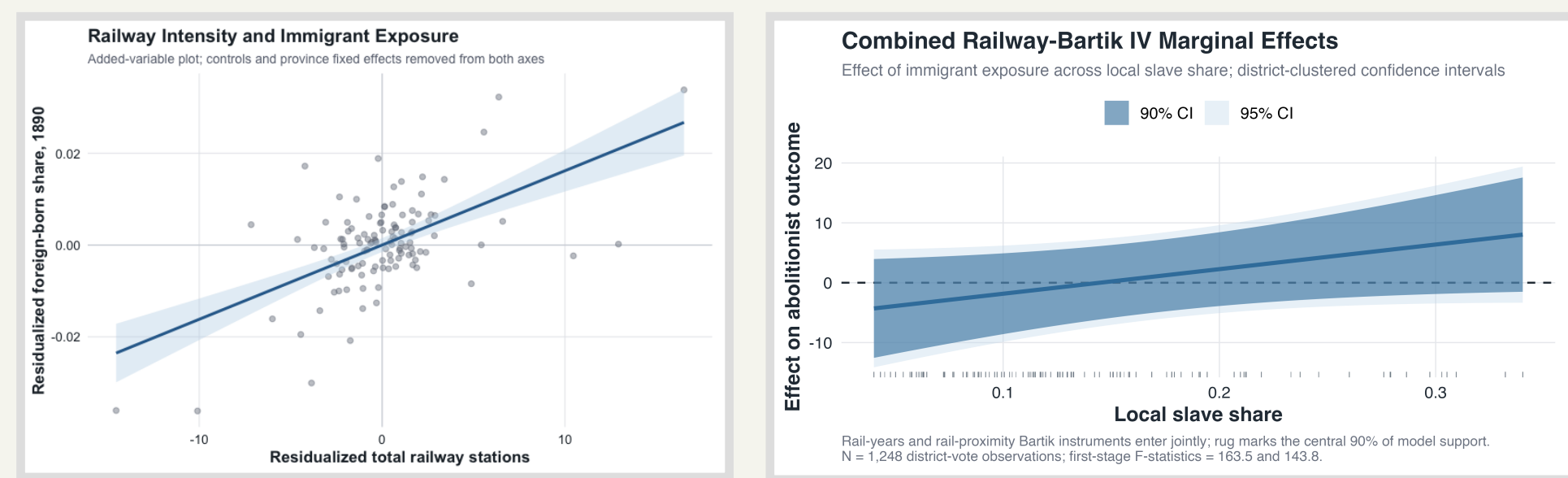
- Test whether the IV effect of immigrant exposure remains after controlling for railway access.
- Examine whether railway exposure strengthens the political effect of immigration.
- Railways in the IV Framework
  - Test whether immigrant exposure predicted by the Bartik instrument directly predicts political outcomes, providing a sanity check for the IV results.
  - A positive ZIV\_country\_x\_slaves indicates stronger abolitionist outcomes in high-slavery areas with greater predicted immigrant exposure.

#### Robustness and Inference: Weak-IV & AR Interval & Conley HAC SE

## 03.2 REGRESSION: Results

### Railway Exposure and Baseline Political Association

- Railway station intensity, measured by `n_stations_total`, is positive and statistically significant, even after district-level controls.
- This suggests that railway network density predicts foreign-born population share in 1890 better than a simple railway-access dummy.
- In the baseline OLS model, Foreign Share × Slave Share is positive, with a coefficient around 32.26 and p-value around 0.121.
- This means immigrant exposure is more strongly associated with abolitionist voting in districts where slavery was more locally salient.
- However, the OLS result is only suggestive and should not be interpreted as causal evidence.

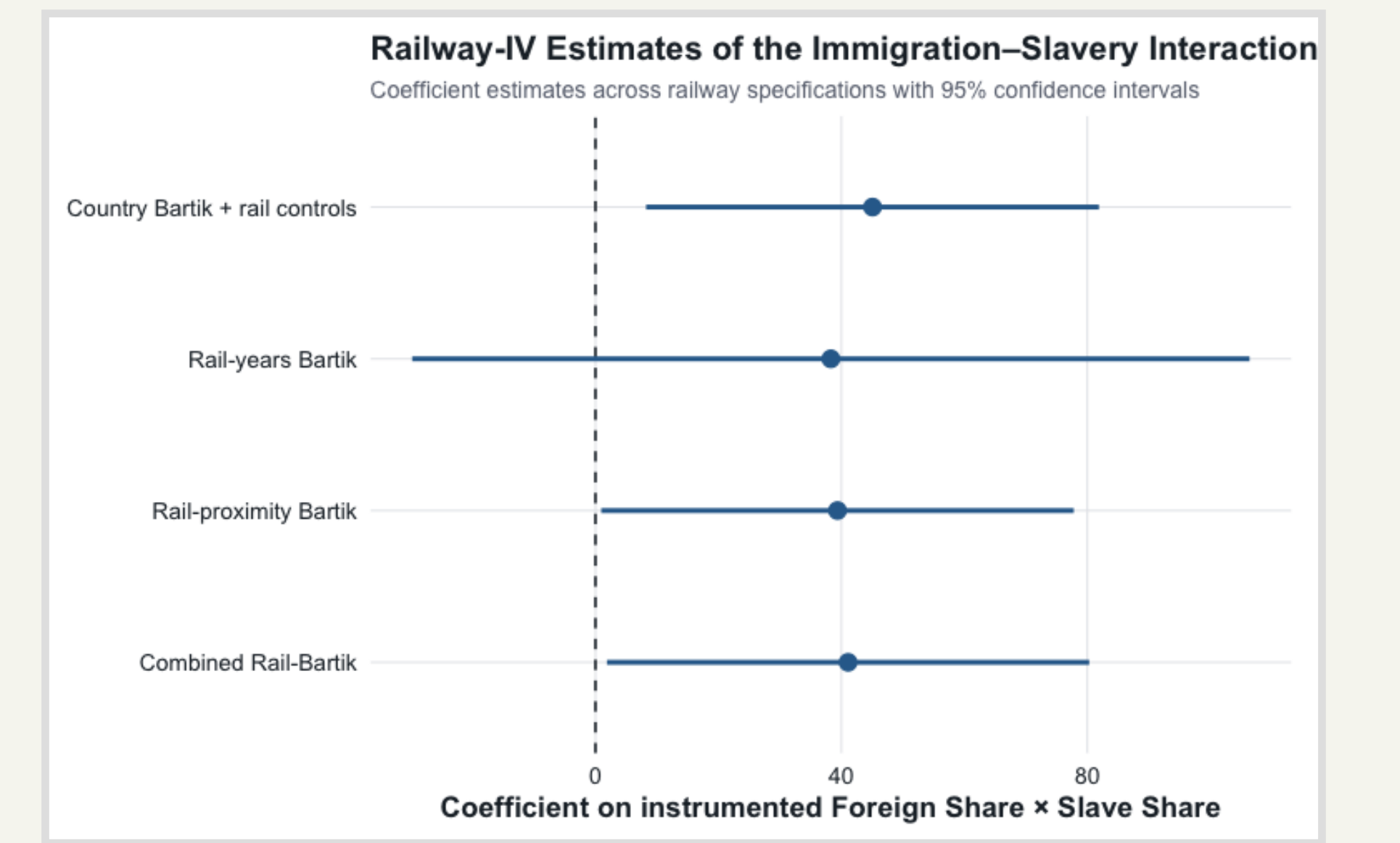


### Bartik IV Strategy

- Bartik IV addresses the endogeneity of immigrant settlement, since immigrants did not randomly choose where to live.
- The Country Bartik model gives the clearest IV result: Foreign Share × Slave Share is positive, with coefficient around 43.84 and p-value around 0.087.
- First-stage F-statistics are very strong, with both Country Bartik first stages above 200, reducing weak-IV concerns.
- The Sargan-Hansen test does not reject instrument consistency, so the IV evidence supports the immigration-slavery channel as suggestive evidence.

### Rail-Bartik Extension

- This stage reconnects the analysis to the project’s central railway question.
- After adding direct railway controls, instrumented Foreign Share × Slave Share remains positive and significant.
- In the rail-control IV model, the coefficient is about 45.06, with p-value around 0.017.
- Rail-proximity Bartik and combined Rail-Bartik models also produce positive and significant interaction effects.
- In the combined Rail-Bartik model, the interaction coefficient is about 41.09, with p-value around 0.040.
- Direct railway variables, such as `rail_access` and `rail_years`, are not statistically significant, suggesting that railways mattered mainly through the immigrant exposure channel.



### Inference Checks

- Conservative inference checks weaken some statistical significance: district clustering gives  $p \approx 0.123$ , while two-way clustering gives  $p \approx 0.077$ .
- The Anderson-Rubin style grid does not overturn the positive interpretation, with an accepted beta2 range of about 3.84 to 97.84.
- Conley spatial HAC is the most conservative check; results become weaker, so the final claim should be framed as suggestive evidence rather than definitive causal proof.