

Analyses on Mobility Metrics for Central Valley and California Counties

Jeffrey Pan

*Central Valley Community Foundation Summer 2026 Internship
Program*

Introduction

Regional economic disparities remain a persistent challenge in the United States, with local clusters of economies often diverging from general macroeconomic trends on the state or national level¹. We attempt to model such a disparity in California, in particular the Central Valley, using a four element statistical analysis. We compare Central Valley counties (Fresno, Kern, Kings, Madera, Merced, San Joaquin, Tulare, Mariposa) with each other and with other similar counties in California and the U.S., initially using debt rate as an indicator for financial distress and pivoting towards other data points such as employment rate, median income, and share of digital access. Ultimately, we apply geospatial autocorrelation to examine the trajectories of all counties in California over time relative to employment and income.

Research Questions and Hypotheses

Our research is driven by the following research questions:

1. What are some of the socioeconomic indicators of regional financial distress, and how might these predictors be impacted by other relevant metrics?
2. How do Central Valley counties cluster (for a given year) compared to each other, other counties in California, and other counties in the U.S.?
3. Are common economic indicators of prosperity (such as employment and income) significantly impacted by geographic region, and are there outlier counties relative to their neighbors and California as a whole?

We hypothesized that the employment rate would have a strong, negative correlation with debt. Furthermore, we expected that Central Valley counties would see greater similarities within each other than with other peer counties (CA, or otherwise), and inter-Central Valley clusters would favor geographically neighboring counties. Finally, we hypothesized that the Central Valley forms an outlier cluster, with its own economic stability (mostly driven by agriculture) refusing spillover economic effects from neighboring counties due to the Central Valley's unique geographic landscape.

Methods

We conducted four connected, exploratory analyses using county-level upward-mobility data. Unless otherwise noted, analyses used the Urban Institute's national mobility dataset². The first analysis used a cleaned analytic file derived from the source data documented in the Fresno Mobility Metrics Data Dashboard³. Because several measures were not available for every year, each analysis was limited to the years with sufficient data coverage.

1. **Multiple linear regression of debt in collections.** We estimated a cross-sectional multiple linear regression model to examine the adjusted associations of the living-wage

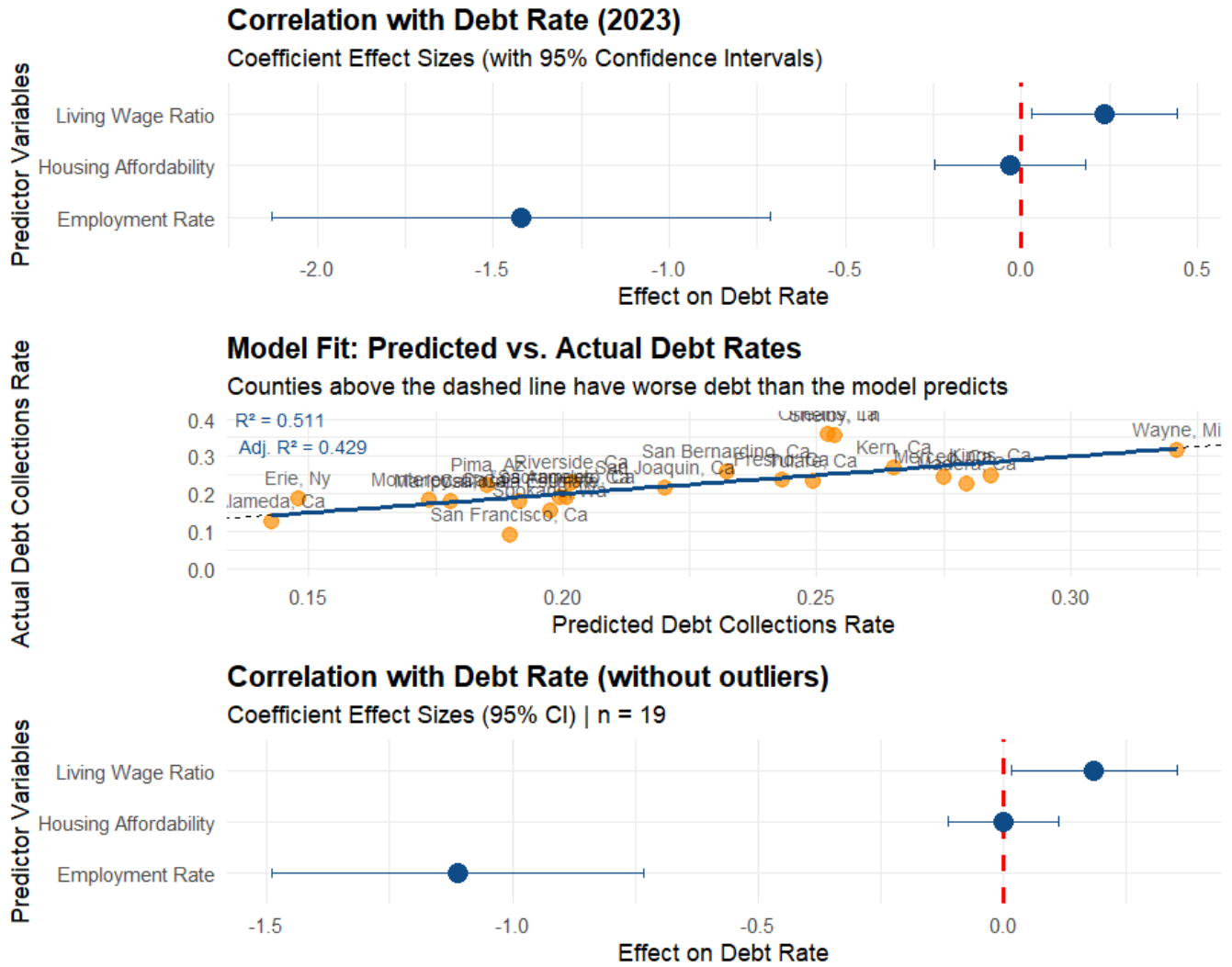
ratio⁴, housing affordability⁵, and employment rate⁶ with the county debt-in-collections rate⁷ in 2023. The analytic sample included nine Central Valley counties, five California peer counties, and six national peer counties selected for the Fresno Mobility Metrics Data Dashboard³. The living-wage ratio⁴ is defined as the ratio of the average wage to the local cost of living; housing affordability⁵ is operationalized as the average of affordability-and-availability ratios for households at 30%, 50%, and 80% of area median income, and employment rate⁶ is defined as the employment-to-population ratio. Statistical significance was evaluated using two-sided tests with an a priori threshold of $p < 0.05$. We report regression coefficients, 95% confidence intervals, p-values, and model-fit statistics.

2. **Descriptive K-means choropleth mapping.** To examine whether additional dashboard mobility measures differentiated Central Valley counties from their peer counties, we created choropleth maps for nine key mobility metrics². Maps were constructed at three geographic scopes: (a) Central Valley counties only, (b) Central Valley counties and California peer counties, and (c) Central Valley counties and national peer counties. Within each geographic scope, counties were grouped into three clusters using K-means clustering based on similarity across the nine metrics⁸. This classification was descriptive; cluster membership was not interpreted as statistically significant. Missing or substandard values were replaced with the overall dataset mean before clustering to prevent unavailable data from disproportionately determining cluster assignment; thus, counties with unavailable or substandard data are displayed in white on maps.
3. **National cross-sectional regression analyses of employment.** Using the full national county dataset, we estimated separate cross-sectional multiple linear regression models for 2021 and 2023. Models examined the adjusted associations between eight key mobility metrics² and county employment rate. Comparing coefficient estimates and model fit across the two year-specific models enabled us to assess whether the observed associations were consistent or changed between 2021 and 2023. Statistical significance was evaluated using two-sided tests with an a priori threshold of $p < 0.05$, and results are reported as regression coefficients, 95% confidence intervals, p-values, and model-fit statistics.
4. **Local spatial autocorrelation analysis.** We assessed local spatial patterns in employment and income among California counties from 2021 to 2023 using Local Moran's I (LISA). Feature values were identified through linear regression against time from 2021 to 2023. Spatial relationships were defined using queen-contiguity weights, whereby counties sharing either a boundary or vertex were treated as neighbors⁹. We also estimated models using K-nearest-neighbor (KNN) weights (for $k = 4, 5$, and 6) as a sensitivity analysis¹⁰. These methodologies are the best suited for determining the correlation between counties on jurisdiction-based values such as employment and income. For each measure and year, Local Moran's I identified high-high and low-low spatial clusters, as well as high-low and low-high spatial outliers. Statistical inference

was based on 9,999 conditional random permutations, producing location-specific pseudo-p-values under the null hypothesis of spatial randomness. Locations with permutation-based pseudo-p-values of $p < 0.05$ were considered statistically significant; these results should be interpreted cautiously because local tests involve multiple comparisons.

Results

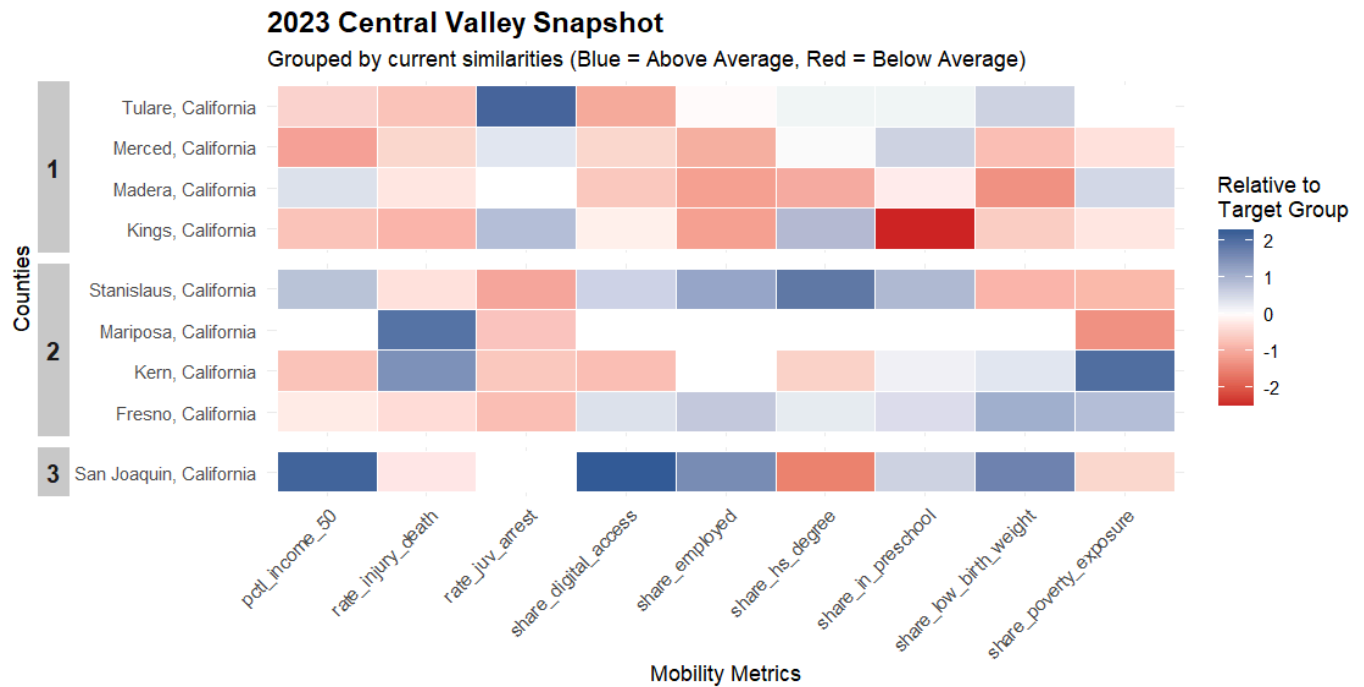
We began with an analysis on factors associated with debt rate and found that, in line with expectations, *employment rate*⁶ (the employment-to-population ratio for adults aged 25 to 54) had the strongest negative association with debt rate⁷ (*financial security* metric: share of households with debt in collections) in 2023. This was a strong, statistically significant association ($p = 0.0005$). Surprisingly, *housing affordability*⁵ proves statistically insignificant ($p = 0.76$) and the living wage ratio⁴ had a weak but positive correlation ($p = 0.026$). Running the same model without the outliers of San Francisco, CA, Orleans, LA, and Shelby, TN, we see similar general results but decreased scale of effects throughout.



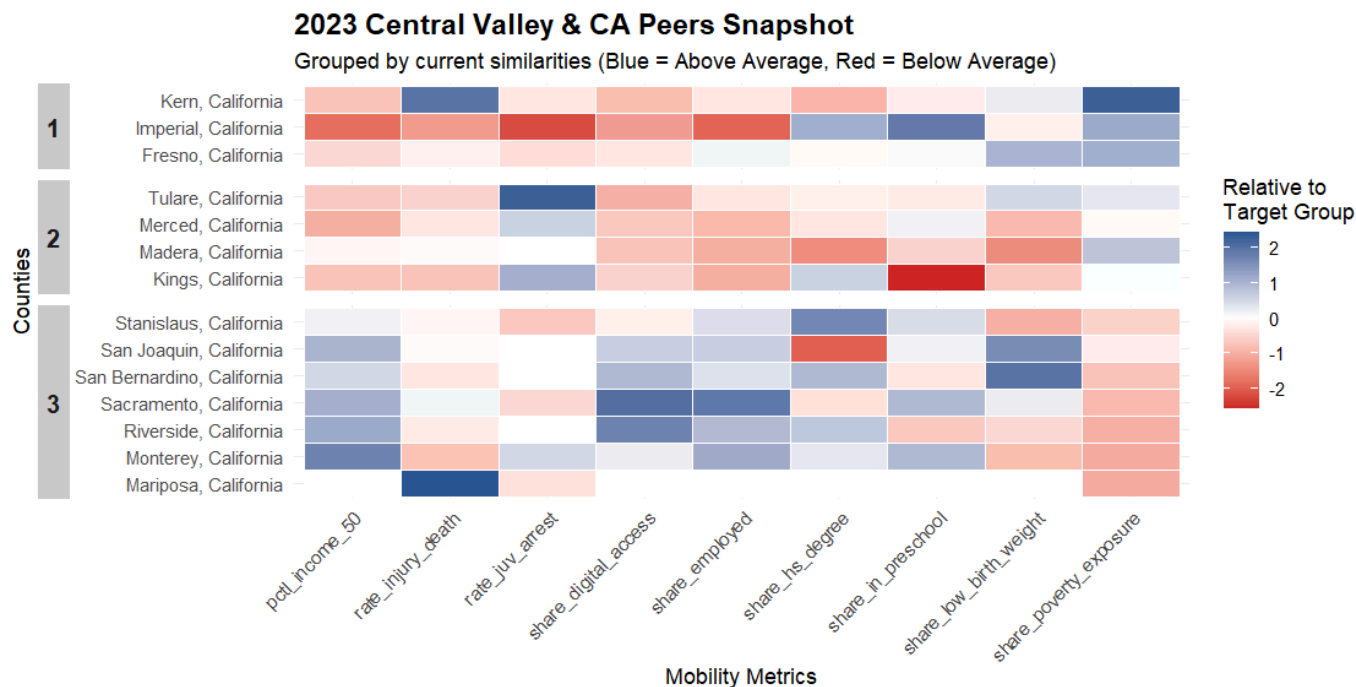
We then expand the scope of this analysis by including additional metrics (share_hs_degree, pctl_income_50, rate_juv_arrest, share_in_preschool, share_digital_access, share_low_birth_weight, rate_injury_death, share_poverty_exposure)¹¹ and varying the counties included in the study (with the “target group” being the average of all counties in the study). Help for interpreting the metrics can be found in the Mobility Metrics Guide¹¹. Note that since Mariposa is missing six out of nine total metrics, we disregard its data due to data quality and missingness.

Considering strictly Central Valley counties, we might expect neighboring counties to share similar statistics, but the counties surrounding Fresno (Tulare, Merced, Madera, and Kings) are in one cluster, while Fresno is in the same cluster as Kern and Stanislaus. San Joaquin is an outlier, existing in its own cluster and sits above the average in five out of nine metrics (50th percentile income, digital access, employment rate, share in preschool, and share poverty

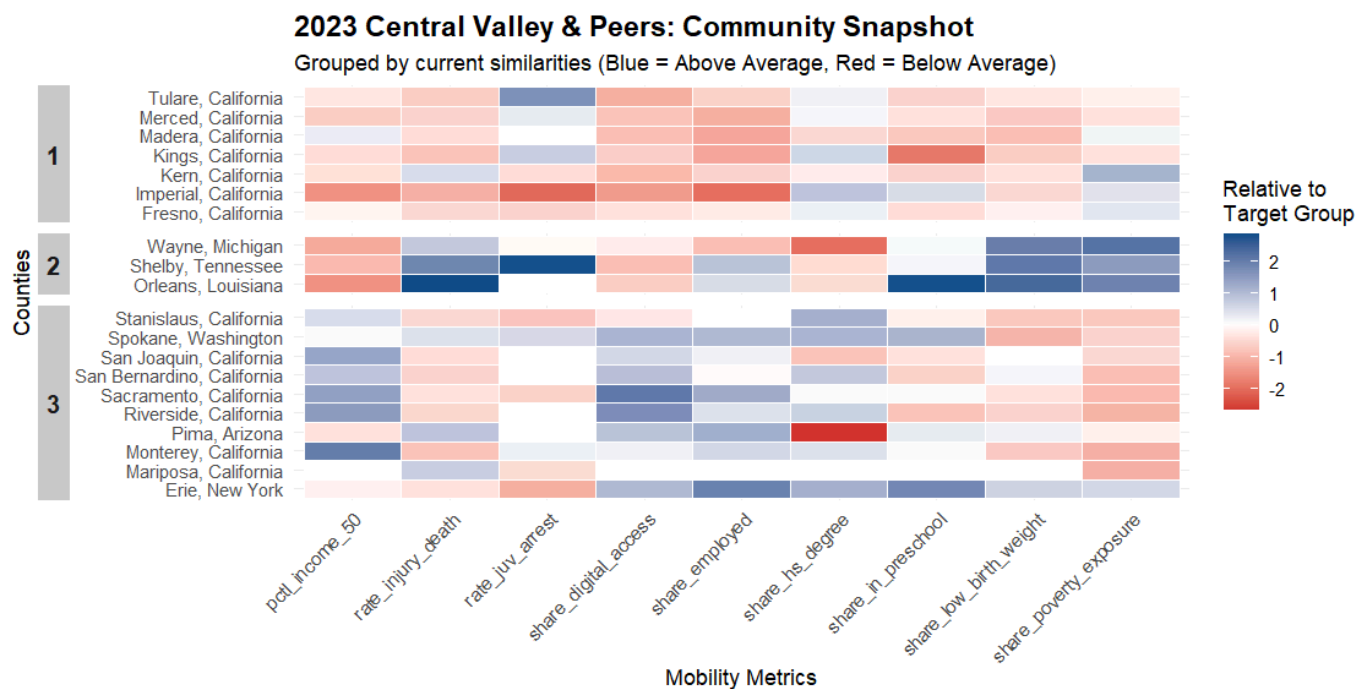
exposure)¹¹. San Joaquin’s status as an outlier might be due to its proximity to the Bay Area, where spillover in productivity and infrastructure improve its prosperity.



The core group of Tulare, Merced, Madera, and Kings exist in the same cluster for every snapshot (those including California peer counties and national peer counties), suggesting they may be heavily influenced by Fresno. This group also tends to underperform in nearly all nine metrics when compared to both (1) other Central Valley Counties and (2) other California and national peers.

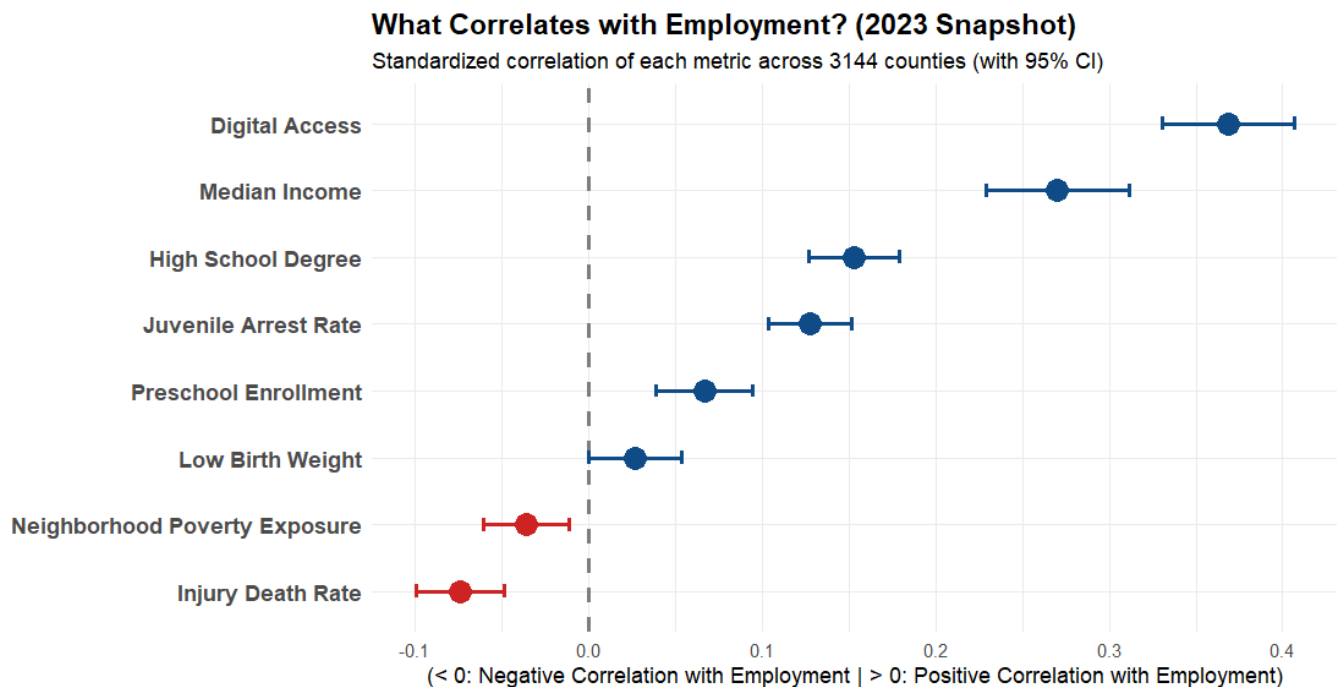


Fresno, Kern, and Imperial counties are grouped in the same cluster, which may be explained by their rural geographic makeup and agricultural similarities¹². Fresno and Kern, as shown, are more similar (both agricultural powerhouses), while Imperial underperforms in almost all metrics across the board (except education), likely due to its lower agricultural output in comparison.

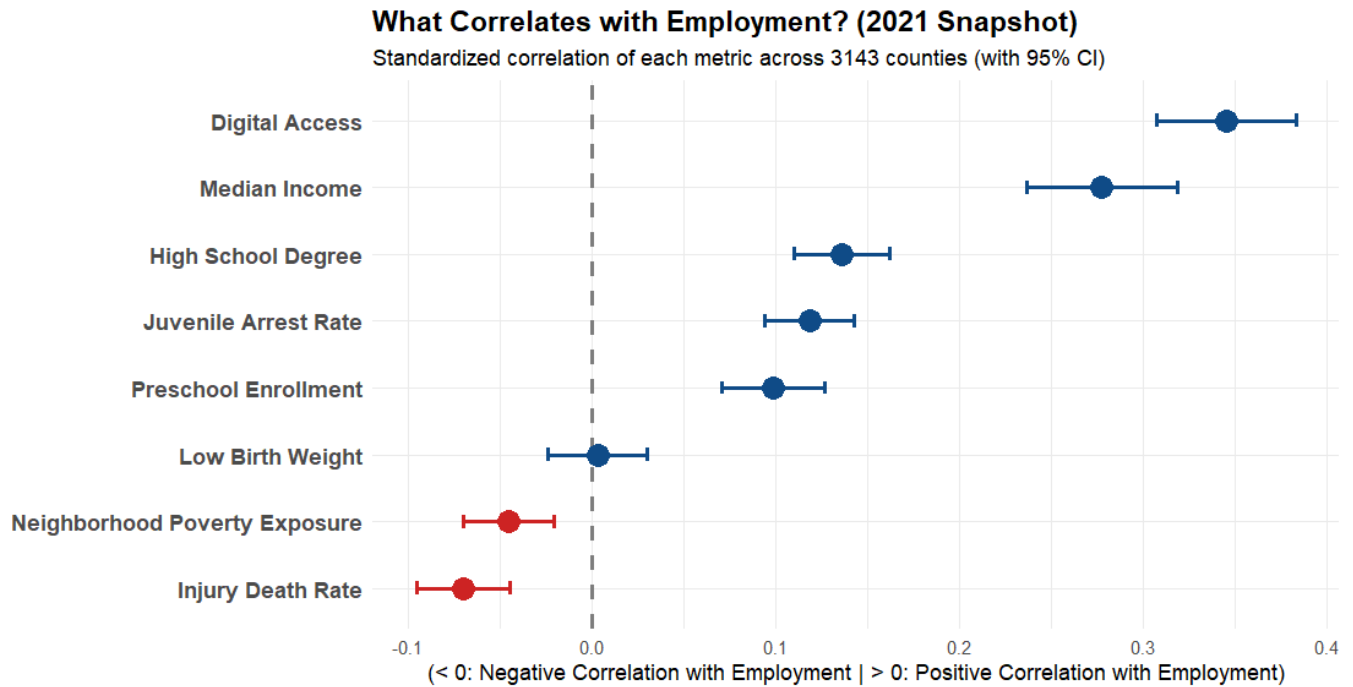


Most Central Valley counties, with the exceptions of Stanislaus and San Joaquin, which we once again suspect is due to their proximity to the Bay Area, tend to underperform in nearly all metrics in a dataset that includes national peer counties. Even the agricultural powerhouses of

Fresno, Kern, and Tulare fall behind, suggesting that a strong economic output does not necessarily result in prosperity for its people, and that counties that rely on agriculture may need to consider other means of promoting prosperity for its residents.



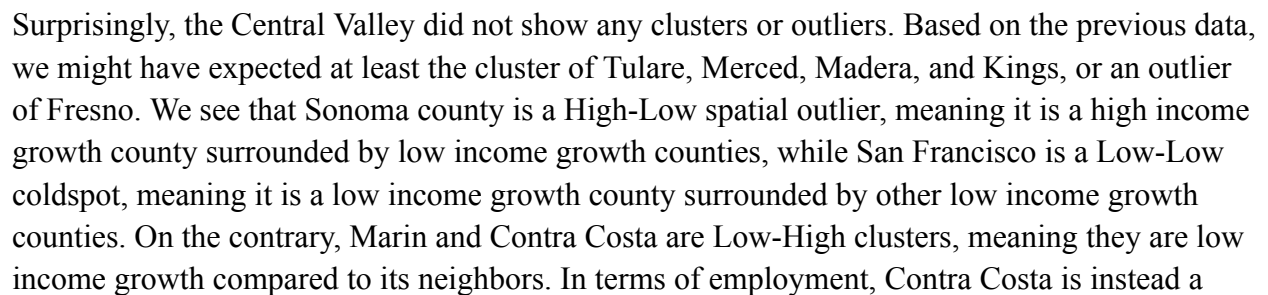
Combining this with the first analysis, we pivot to focus on these same metrics, but instead regressing the remaining eight metrics (depicted in the Y axis above) against employment in both 2021 and 2023. Injury death rate (deaths caused by injury per 100,000) and neighborhood poverty exposure (share of people experiencing poverty living in high-poverty neighborhoods) have a negative correlation with employment, while all other metrics have a positive correlation. They have p-values of $7.38e^{-9}$ and 0.00386, respectively, in 2023, showing weak, but statistically significant correlations. The correlation with low birth weight is likely spurious, given that it is also statistically insignificant in 2021 ($p = 0.0526$ and 0.81 in 2023 and 2021, respectively). Studies also have shown that the correlation between juvenile arrest rate and employment are spurious or caused by other variables as well¹³. We suspect that the high correlation between employment and median income are mostly due to many of the same factors contributing towards both lower unemployment and higher median income (quality of education, industry access, etc).



Unsurprisingly, digital access has the strongest positive correlation ($p < 2e^{-16}$ in 2021 & 2023), as the benefits of digital products are pervasive at every age and massively impacts the quality of education. We expected that the two education related metrics (high school degree & preschool enrollment) would be similarly correlated with employment, but overall expected both to have a higher correlation.

Our final step was to evaluate income and employment trajectories in California using the above metrics. Due to data constraints, we were only able to use a portion of the metrics (rate_injury_death, rate_juv_arrest, pctl_income_20, pctl_income_50, pctl_income_80, share_poverty_exposure, share_employed, share_low_birth_weight)¹¹ and were limited to the years 2021 - 2023 to examine the most recent trends. The following two figures use linear regression and Local Moran's I with queen's contiguity⁹ (meaning the spatial weights matrix is constructed based on the assumption that a given county and its "neighbor" share a border) to identify outliers and clusters.

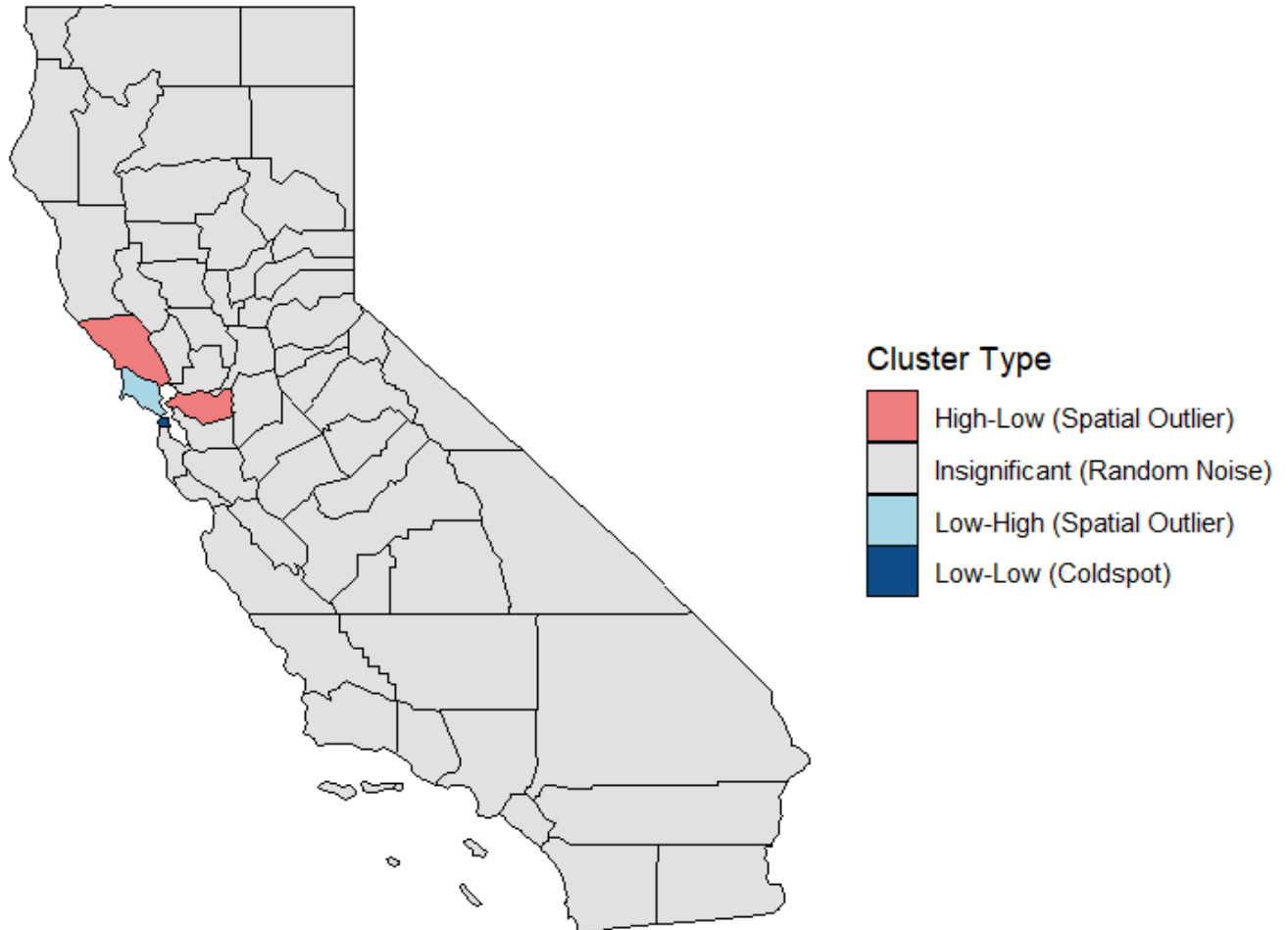
Local Moran's I (Spatial Autocorrelation of Income Trajectories)



High-Low cluster, with all other counties remaining the same.

California Employment Share Trajectory Clusters (2021-2023)

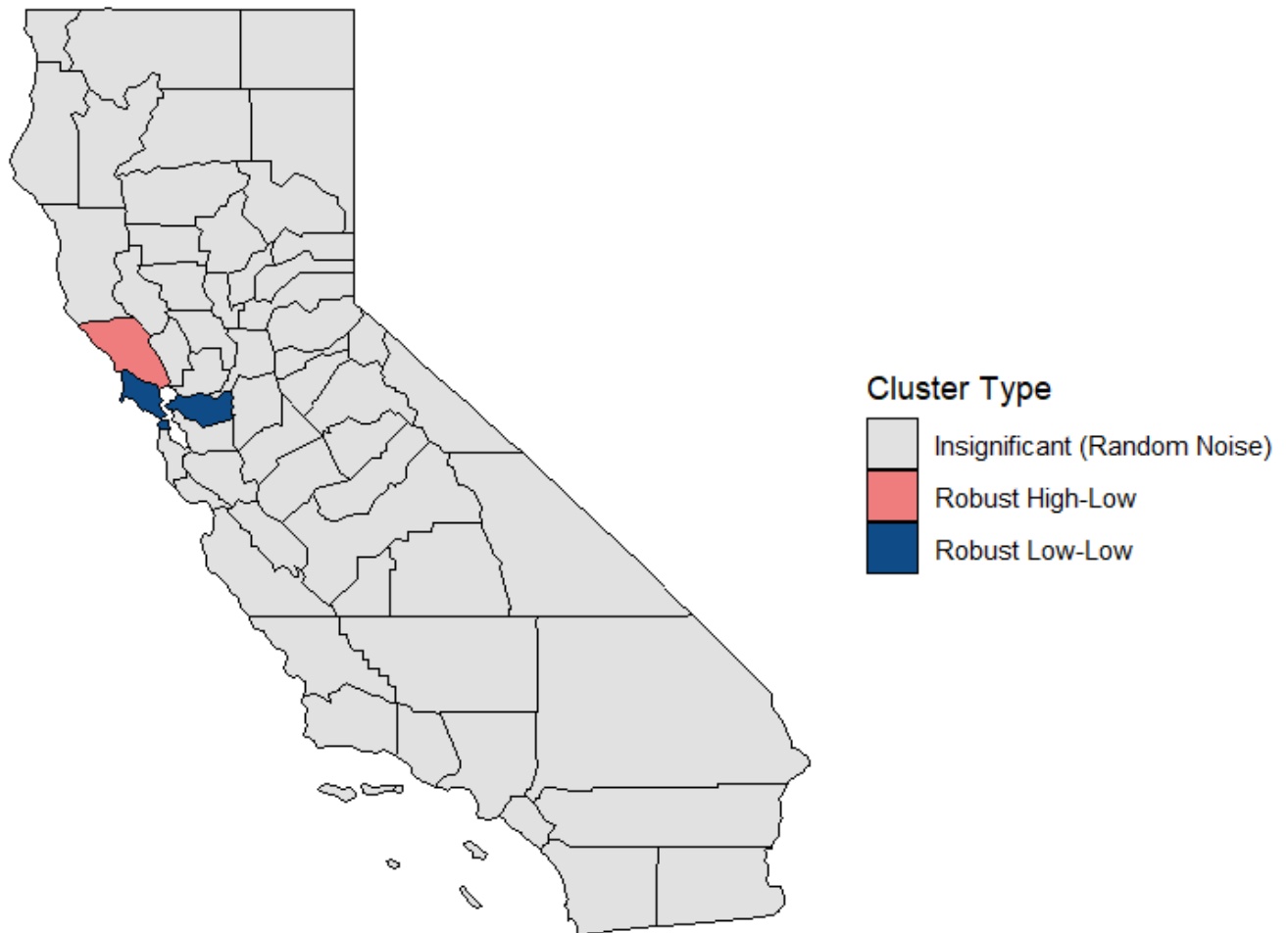
Local Moran's I (Spatial Autocorrelation of Employment Growth)



For the following two graphs, we use the same methodology, except the spatial weights are constructed using K Nearest Neighbors¹⁰ (KNN), where a county's "neighbors" are just the k nearest counties. For consistency, we run the analysis using $k = 4, 5$, and 6 . Robust clusters are determined when the cluster is significant ($p > 0.05$ for a two-tailed test) for all k, while fragile clusters are determined when the cluster is significant for some k. Once again, in both employment and income, Sonoma sees higher growth over time compared to its neighbors. Almost the entire Bay Area suffers low income growth, with 3 robust counties (Marin, San Francisco, and Contra Costa).

Robust California Median Income Trajectories (2021-2023)

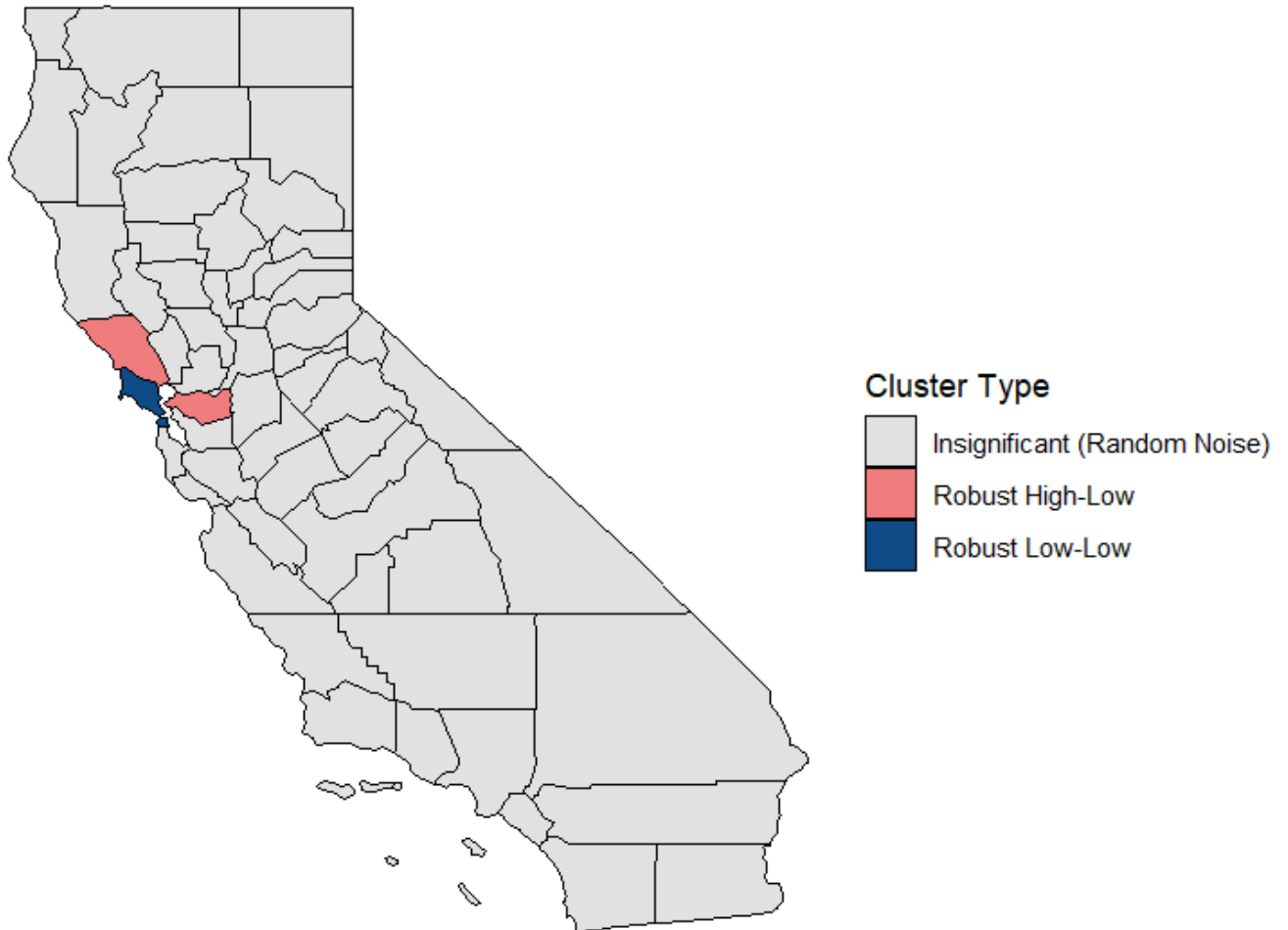
Survivor Clusters Across KNN Topologies (k = 4, 5, 6)



In terms of employment, Marin and San Francisco remain Robust Low-Low coldspots, while Contra Costa is now a Robust High-Low outlier. Examining both types of spatial weights jointly (queen's contiguity and KNN), several patterns emerge: (1) Sonoma county is a High-Low outlier in all four analyses, (2) San Francisco is a Low-Low coldspot in all four analyses, (3) and in both methodologies, Contra Costa is a Low-Low coldspot in terms of income and a High-Low outlier in terms of employment.

Robust California Employment Trajectories (2021-2023)

Survivor Clusters Across KNN Topologies ($k = 4, 5, 6$)



The primary difference between the two methodologies is how they weigh spatial influence. Queen's contiguity treats any bordering counties as neighbors, which causes it to understate the impact of individual neighbors for counties with many borders, and overstates it for counties with few. For example, San Francisco only technically borders San Mateo county. Only using queen's contiguity⁹ may not accurately represent the influence counties have on each other, which is why KNN¹⁰ was also used for sensitivity analysis. In this case, the results were largely the same.

Discussion

The Central Valley struggles with unemployment and several other upward mobility metrics relative to the rest of California and national peer counties. The low performance of metrics such as digital access, share in preschool, share in high school, and 50th percentile income relative to

California and national peers, reveal structural vulnerabilities that represent underdeveloped infrastructure and an agricultural core that do not translate into much else. The Urban Institute's Upward Mobility Initiative identifies five key infrastructures key to advancing upward mobility¹⁴ (rewarding work, high quality education, opportunity rich & inclusive neighborhoods, healthy environment & access to good health care, and responsive & just governance), of which our metrics roughly translate into three (rewarding work, high quality education, and ~responsive & just governance). Of these three, the Central Valley consistently falls short on rewarding work and high quality education. If we had access to a more complete dataset, we hypothesize that the Central Valley may fall short on more.

We initially intended to examine the Central Valley for clustering, and we strongly expected influence from these strong agricultural counties to bleed into each other, especially based on our heat maps. However, the data provided unexpected insights. Our geospatial trajectory analysis reveals that influence between counties in the Central Valley is statistically insignificant. Instead, Sonoma county repeatedly presents as an outlier with higher growth than its neighboring counties, with other Bay Area counties reporting lower growth. Since none of the other counties near the Bay Area show higher growth (besides Contra Costa, which shows higher employment growth), employers and capital seem to be rotating into Sonoma county in favor of traditional tech-heavy areas such as San Francisco and Silicon Valley. Big tech has seen significant layoffs and transitions during and after the pandemic¹⁵, which is supported by our analysis. Sonoma county may be a geography where economic transitions are taking place.

References:

1. U.S. Department of Commerce. (2023). Geographic Inequality on the Rise in the U.S. In *U.S. Department of Commerce*.
<https://www.commerce.gov/news/blog/2023/06/geographic-inequality-rise-us>
2. Urban Institute. (2025). Mobility Metrics Data for the Upward Mobility Framework (v2025.01). Accessible from
<https://datacatalog.urban.org/dataset/mobility-metrics-data-upward-mobility-framework>. Data originally sourced from, developed at the Urban Institute, and made available under the ODC-BY 1.0 Attribution License.
3. Fresno Mobility Metrics Data Dashboard (2026): Retrieved August 9, 2026, from
<https://www.fresnodrive.org/data/data-dashboard>. See original source: US Census Bureau 2023 1-year American Community Survey, Public Use Microdata Sample (via [IPUMS](#)); Missouri Census Data Center Geocorr 2022: [Geographic Correspondence Engine](#).

4. Fresno Mobility Metrics Data Dashboard: Jobs paying living wages metric. Retrieved August 10, 2026, from <https://www.fresnodrive.org/data/data-dashboard>. See original source: US Bureau of Labor Statistics, Quarterly Census of Employment and Wages data, 2023; Living Wage Institute, Inc. "Benchmark Living Wage Data Series," 2024.
5. Fresno Mobility Metrics Data Dashboard: Housing Affordability metric. Retrieved August 10, 2026, from <https://www.fresnodrive.org/data/data-dashboard>. See original source: US Department of Housing and Urban Development, Office of Policy Development and Research, Fair Market Rents and Income Limits, FY 2023; US Census Bureau 2023 1-Year American Community Survey, Public Use Microdata Sample (via IPUMS); Missouri Census Data Center Geocorr 2022: Geographic Correspondence Engine.
6. Fresno Mobility Metrics Data Dashboard: Employment opportunities metric. Retrieved August 10, 2026, from <https://www.fresnodrive.org/data/data-dashboard>. See original source: US Census Bureau 2023 1-year American Community Survey, Public Use Microdata Sample (via IPUMS); Missouri Census Data Center Geocorr 2022: Geographic Correspondence Engine.
7. Fresno Mobility Metrics Data Dashboard: Financial Security metric. Retrieved August 10, 2026, from <https://www.fresnodrive.org/data/data-dashboard>. See original source: 2023 credit bureau data from Urban Institute's Debt in America feature.
8. Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295.
9. Getis, A. and Aldstadt, J. (2004), Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis*, 36: 90-104. <https://doi.org/10.1111/j.1538-4632.2004.tb01127.x>
10. Zhang Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11), 218. <https://doi.org/10.21037/atm.2016.03.37>
11. *Mobility Metrics Data for the Upward Mobility Framework*. (2026). Github.io. <https://ui-research.github.io/mobility-from-poverty-dictionary/>
12. Office of Public Affairs. (2024, February 13). *Nine California Counties Make Top-10 List for Ag Sales in the U.S.* CDFA's Planting Seeds Blog. <https://plantingseedsblog.cdfa.ca.gov/index.php/2024/02/13/nine-california-counties-make-top-10-list-for-ag-sales-in-the-u-s/>

13. Staff, J., Osgood, D. W., Schulenberg, J. E., Bachman, J. G., & Messersmith, E. E. (2010). Explaining the Relationship between Employment and Juvenile Delinquency. *Criminology : an interdisciplinary journal*, 48(4), 1101–1131. <https://doi.org/10.1111/j.1745-9125.2010.00213.x>
14. Upward Mobility Framework. (2024). In Urban Institute | Upward Mobility Initiative. <https://upward-mobility.urban.org/framework>
15. Van Dijke, David and Gunsilius, Florian and Wright, Austin L., Return to Office and the Tenure Distribution (May 7, 2024). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2024-56, Available at SSRN: <https://ssrn.com/abstract=4820135> or <http://dx.doi.org/10.2139/ssrn.4820135>