

**Web Appendix: GeoReddit: Geolocating Social Media Users to Enable Spatio-Temporal
Marketing Insights**

Web Appendix A: Robustness Check for State Assignment	2
Web Appendix B: Visualizations for Validation of Geolocation	5
Web Appendix C: Robustness Analysis for Vignette 1 (Pepsi's Brand Crisis).....	21
Web Appendix D: Robustness Analysis for Vignette 2 (The Legalization of Marijuana)	26
Web Appendix E: Robustness Analyses for Vignette 3 (Sanders Primaries).....	40

These materials have been supplied by the authors to aid in the understanding of their paper. The AMA is sharing these materials at the request of the authors.

Web Appendix A: Robustness Check for State Assignment

To evaluate the robustness of our rule-based geolocation procedure, we compared user assignments based on year-specific data with those based on the full-period assignment. Table W1 reports the decomposition of users into several categories: (1) *Total Year* is the number of users in a given year meeting the $\text{Ratio}_2 > 1$ threshold, which serves as the effective sample size; (2) *Only Year* are users that appear exclusively in that year's set but not in the full-period set; (3) *Only All* are users assigned in the global set but absent from that year; (4) *Mismatch* counts users present in both sets whose year-specific state assignment differs from their global state assignment; and (5) *Match* counts users consistently assigned to the same state in both sets.

Additionally, we report two summary indicators. The first, *Single Year Mismatches Global*, refers to the share of users in a given year whose assigned state differs from their global assignment, relative to the total number of users that can be allocated in that year. This captures the extent to which year-specific assignments deviate from the global baseline. The second, *Percentage of Final if Limited to Year*, refers to the share of users that could be covered if we relied only on year-specific information, relative to the full set of users identified in the global assignment. While the first indicator reflects the consistency of year-specific assignments, the second reflects their coverage.

In the early years of Reddit (2006–2009), the number of users allocable on a year-specific basis was very small (fewer than 55,000 per year), reflecting both the platform's limited user base and sparse geographic signals. From 2010 onward, the sample expanded rapidly, surpassing 100,000 users that year and reaching over 4.5 million by 2023. The mismatch rate, as shown in *Single Year Mismatches Global*, remained consistently low across the period, averaging 2.5% and peaking just above 3% in 2011–2013 before stabilizing around 2.2–2.4% in recent years.

This indicates that year-specific and global assignments are highly consistent. At the same time, *Percentage of Final if Limited to Year* grew steadily as Reddit scaled, rising from negligible levels before 2010 to nearly 10% in 2012, 34% in 2018, and more than 60% by 2023. This trend shows that year-specific coverage becomes increasingly comprehensive over time. Accordingly, our vignette analyses (Vignettes 1 and 2) begin in 2014, when coverage first exceeds 15%, to ensure that results are not unduly affected by the scarcity of geolocated users in earlier years.

Table W1: Comparison of Year-Specific and Global State Assignments.

Year	Total Year	Only Year	Only All	Mismatch	Match	Single Year Mismatches Global	Percentage of Final if Limited to Year
2006	2,816	-	7,424,603	-	2,816	-	0.0%
2007	9,318	-	7,418,101	-	9,318	-	0.1%
2008	21,846	16	7,405,589	39	21,791	0.2%	0.3%
2009	52,524	146	7,375,041	477	51,901	0.9%	0.7%
2010	137,674	1,379	7,291,124	3,474	132,821	2.5%	1.9%
2011	367,597	5,669	7,065,491	12,295	349,633	3.3%	4.9%
2012	717,364	11,425	6,721,480	23,049	682,890	3.2%	9.7%
2013	1,043,498	17,479	6,401,400	33,421	992,598	3.2%	14.0%
2014	1,310,488	20,625	6,137,556	38,454	1,251,409	2.9%	17.6%
2015	1,560,867	24,708	5,891,260	44,725	1,491,434	2.9%	21.0%
2016	1,822,368	27,847	5,632,898	48,505	1,746,016	2.7%	24.5%
2017	2,114,497	33,670	5,346,592	57,382	2,023,445	2.7%	28.5%
2018	2,533,205	39,037	4,933,251	63,556	2,430,612	2.5%	34.1%
2019	3,001,528	42,786	4,468,677	69,699	2,889,043	2.3%	40.4%
2020	3,641,711	58,412	3,844,120	86,191	3,497,108	2.4%	49.0%
2021	4,200,847	64,207	3,290,779	93,681	4,042,959	2.2%	56.6%
2022	4,524,447	75,583	2,978,555	102,323	4,346,541	2.3%	60.9%
2023	4,559,157	70,624	2,938,886	104,794	4,383,739	2.3%	61.4%
Average	-	-	-	-	-	2.5%	23.7%

Notes: Single Year Mismatches Global = Mismatch ÷ Total Year. Percentage of Final if Limited to Year = Total Year ÷ (Only All + Mismatch + Match).

In sum, these results confirm two points. First, the global assignment provides highly consistent geolocation, as the mismatch rate is negligible relative to the large user base. Second, year-specific data increasingly provide good coverage as Reddit's scale grows, but relying exclusively on year-specific assignment would substantially reduce the usable sample in earlier periods. Therefore, for consistency and comprehensiveness, we base all subsequent analyses on the global assignment, while recognizing that the results are robust to year-specific allocations.

Web Appendix B: Visualizations for Validation of Geolocation

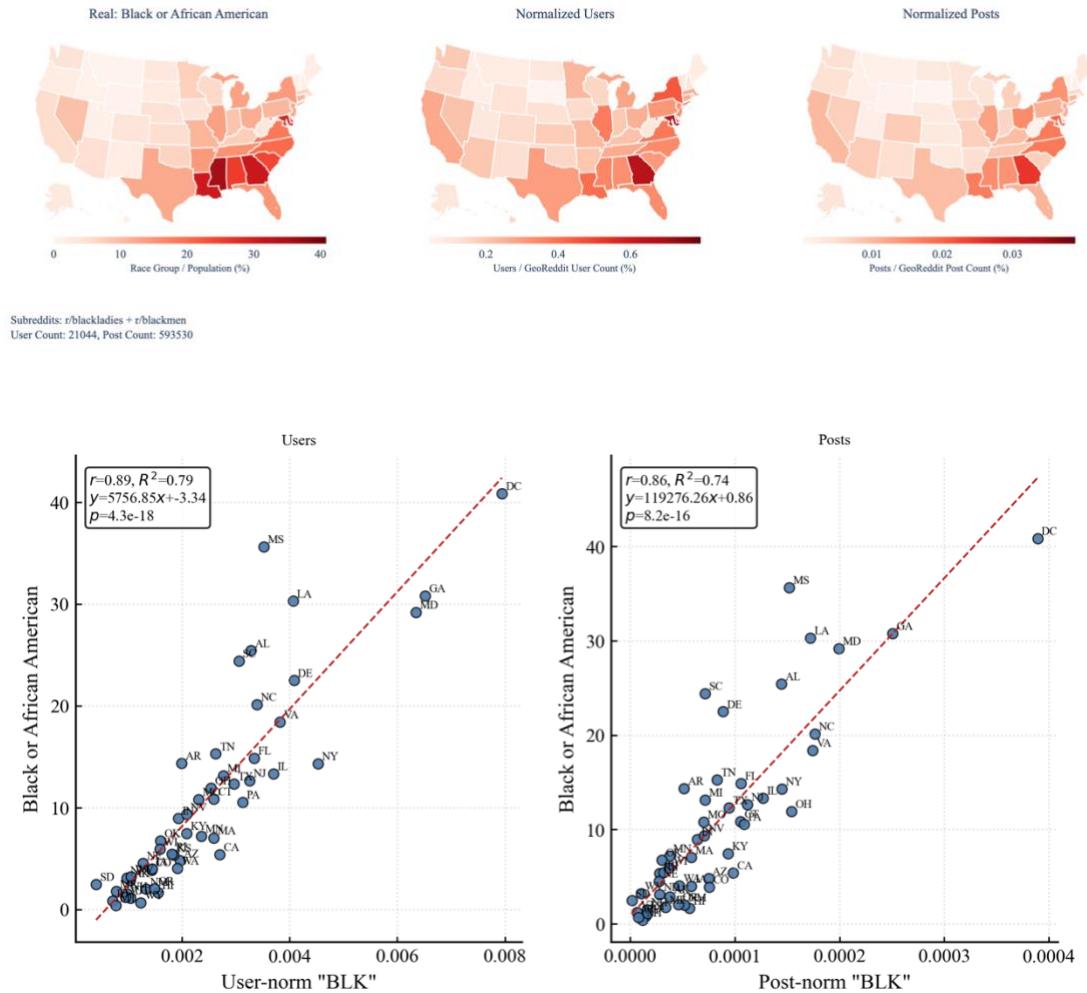
This appendix provides supplementary visualizations for the validation of our geolocation procedure. Specifically, we include: (1) State-level distribution maps, which present side-by-side comparisons of real-world benchmarks and GeoReddit estimates. For each selected subgroup (demographic, religious, occupational or gun ownership), we plotted three maps: the benchmark distribution from external data (e.g., ACS, BLS, survey-based gun ownership rates), the normalized distribution of GeoReddit users, and the normalized distribution of their posts. This tripartite visualization highlights the geographic alignment between official statistics and GeoReddit-based measures. (2) Scatterplots of Pearson correlation analyses, which compare state-level GeoReddit user and post distributions against external benchmarks. These scatterplots provide a statistical complement to the maps and further illustrate the strength of the correspondence between GeoReddit-based estimates and ground-truth data. Overall, these figures complement the results reported in the main text and facilitate interpretation of state-wise coverage and representativeness.

Demographic Validation

Racial and Ethnic Distributions

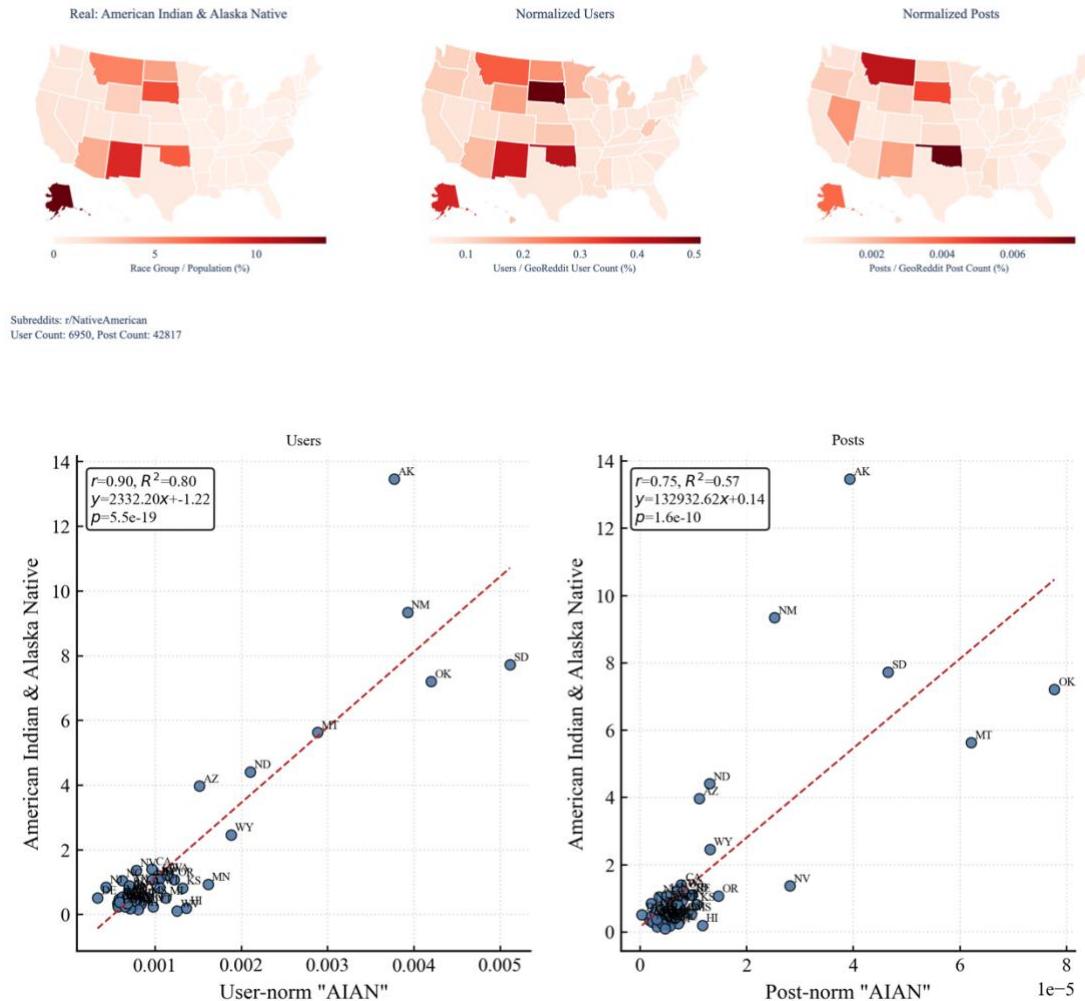
This section presents state-level distribution maps and scatterplots for racial and ethnic subgroups, comparing GeoReddit-based geolocation with 2023 American Community Survey (ACS) benchmarks. Specifically, we use ACS codes B02001_003E (Black or African American alone), B02001_004E (American Indian and Alaska Native alone), B02001_005E (Asian alone), B02001_006E (Native Hawaiian and Other Pacific Islander alone), and B03003_003E (Hispanic or Latino of any race). Each of these subgroup measures is normalized by B02001_001E (Total population) to ensure comparability across states.

Figure W1: Black or African American — GeoReddit vs. ACS 2023.



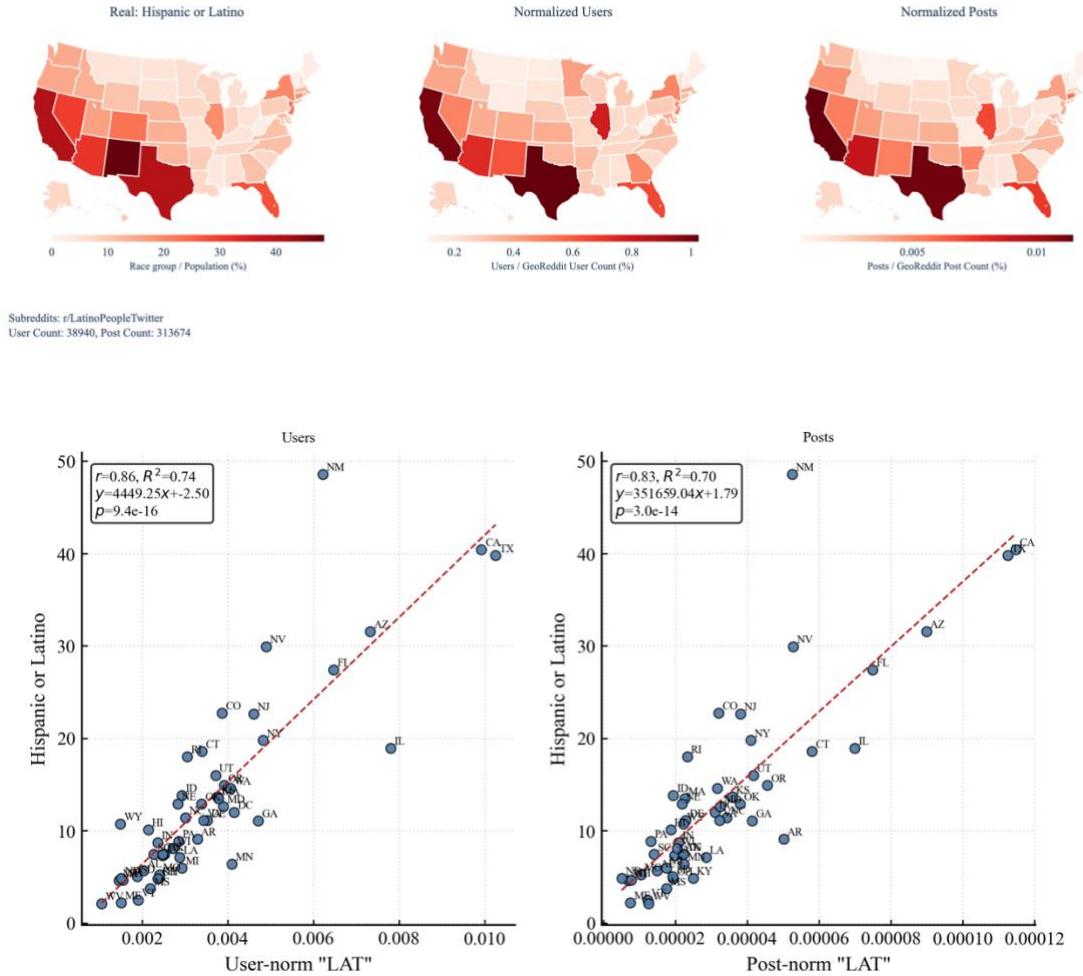
Notes: (top) State-level distribution maps (ACS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ACS.

Figure W2: American Indian and Alaska Native — GeoReddit vs. ACS 2023.



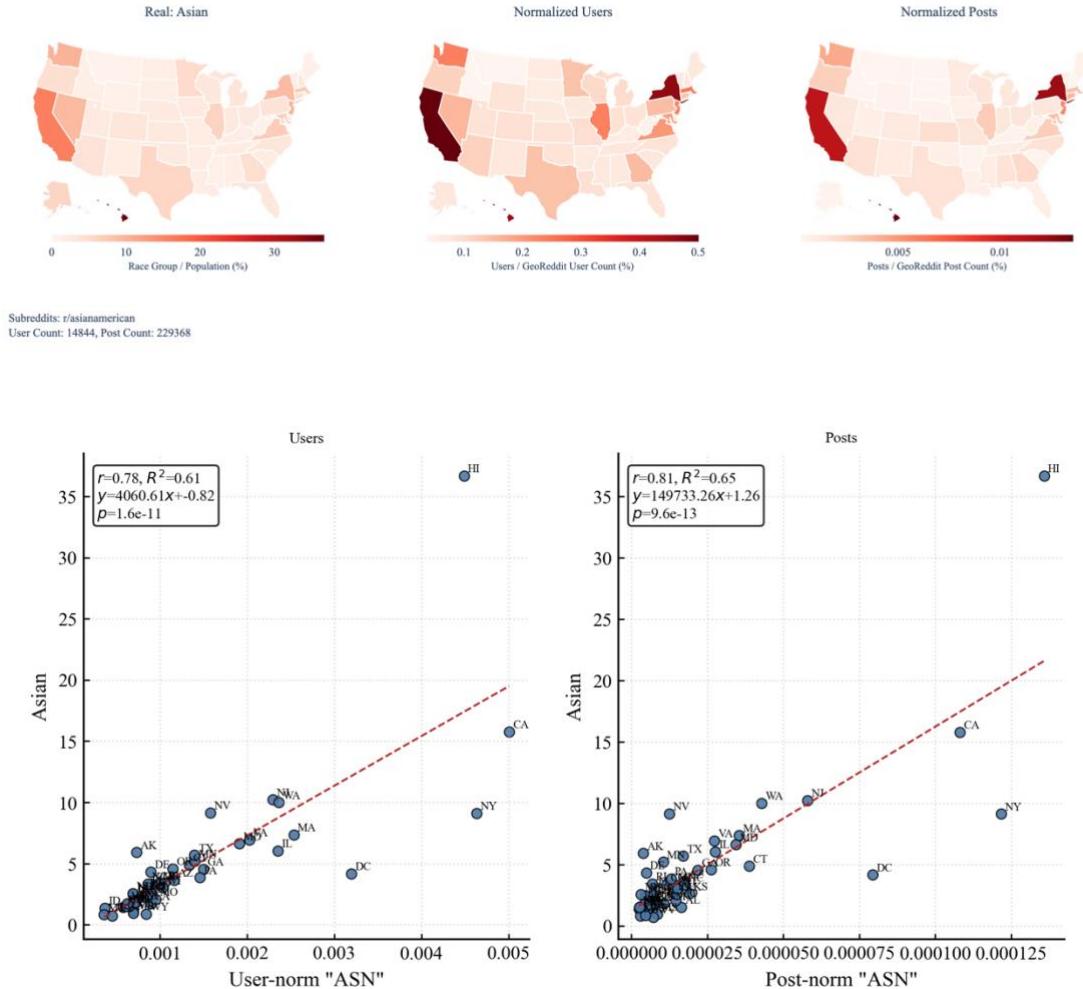
Notes: (top) State-level distribution maps (ACS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ACS.

Figure W3: Hispanic or Latino — GeoReddit vs. ACS 2023.



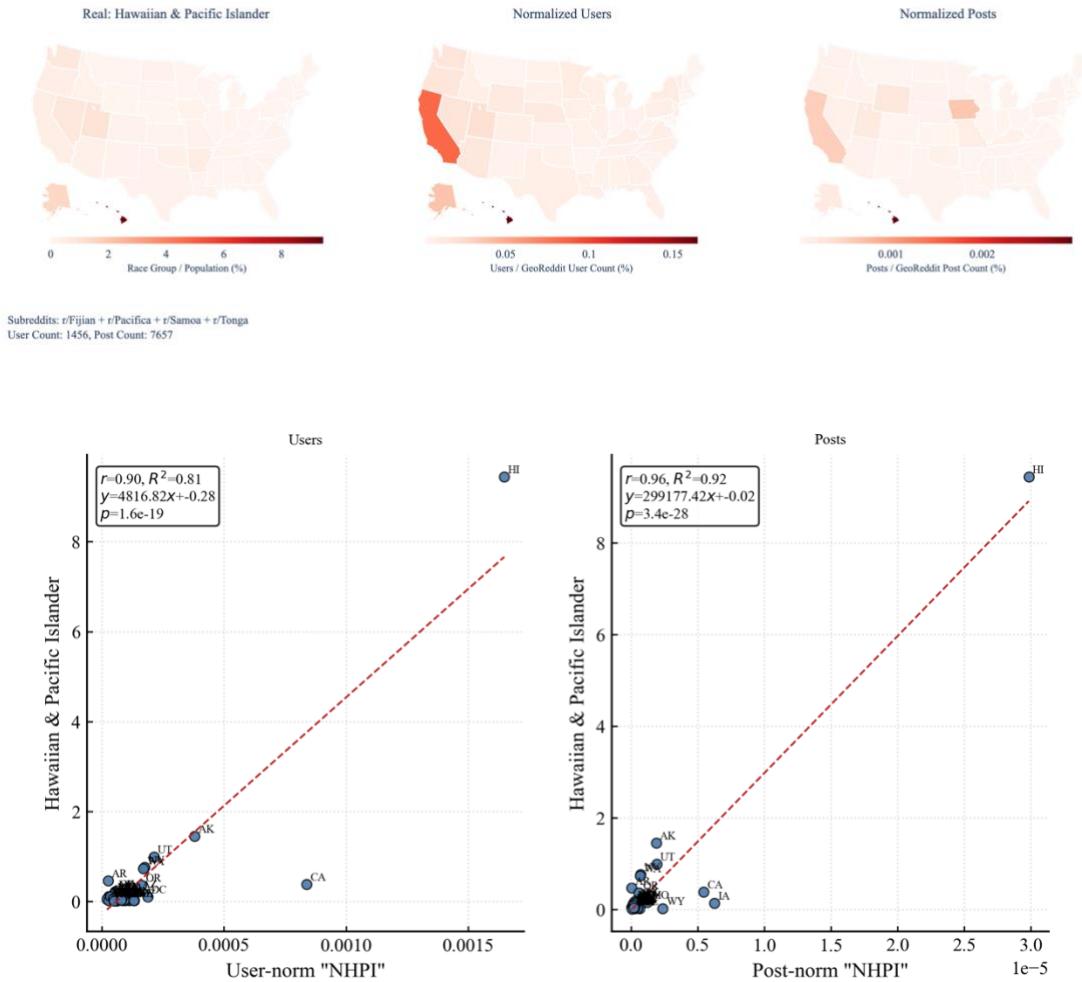
Notes: (top) State-level distribution maps (ACS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ACS.

Figure W4: Asian American — GeoReddit vs. ACS 2023.



Notes: (top) State-level distribution maps (ACS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ACS.

Figure W5: Native Hawaiian and Pacific Islander — GeoReddit vs. ACS 2023.

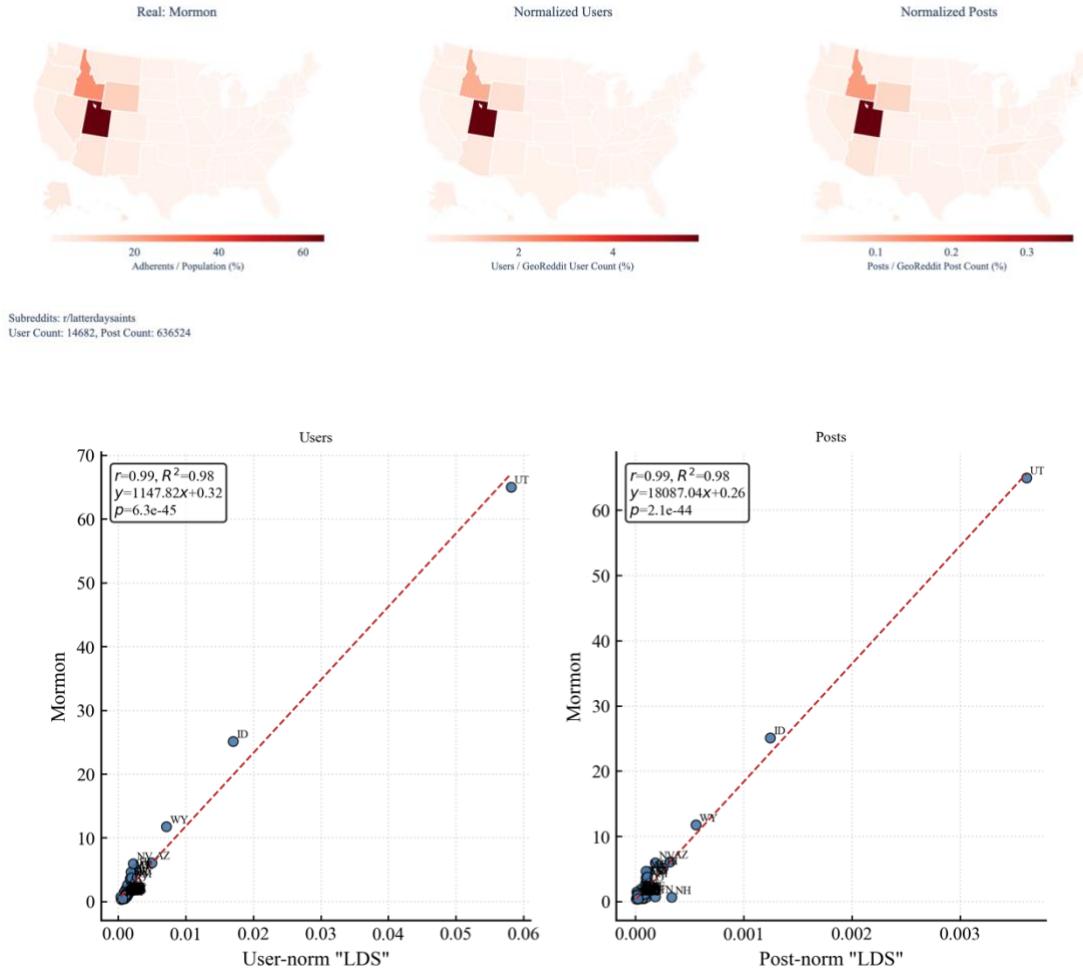


Notes: (top) State-level distribution maps (ACS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ACS.

Religious Identity Distributions

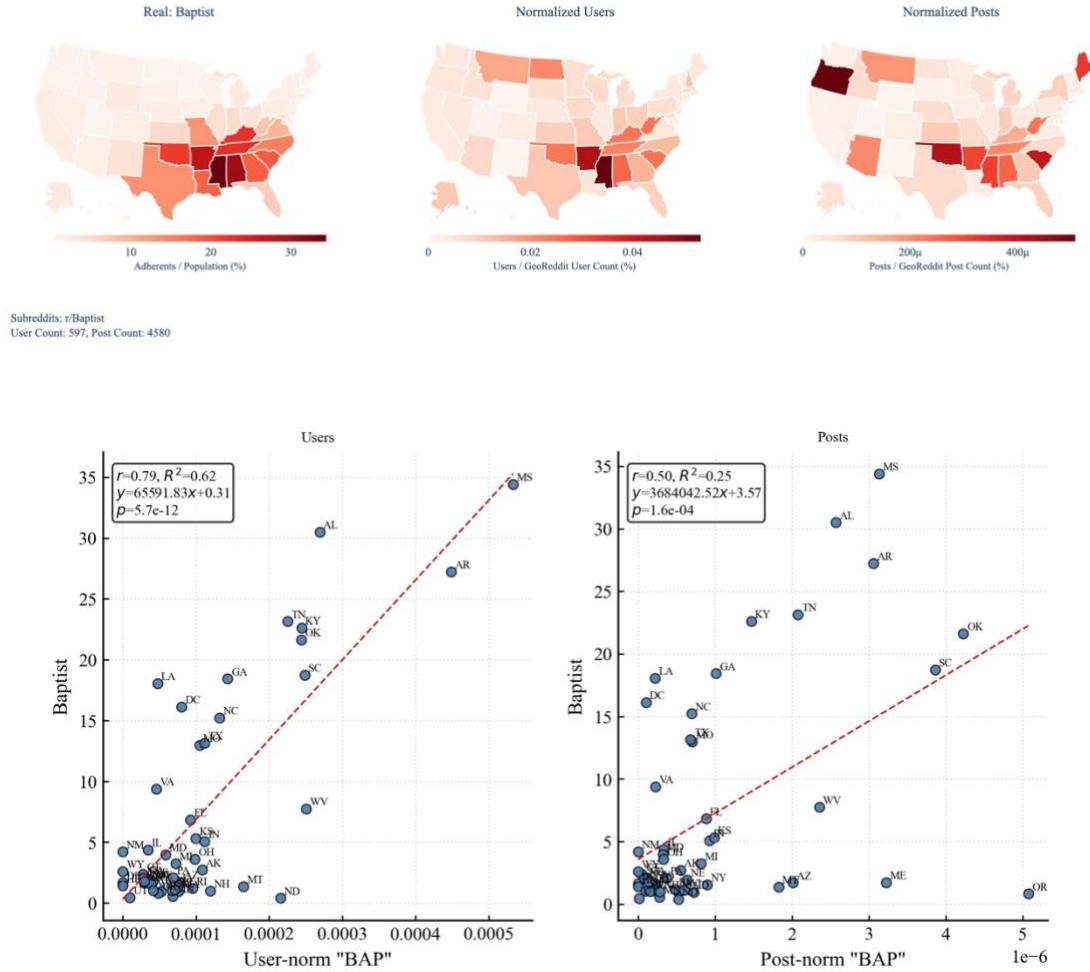
This section provides supplementary figures for religious identity validation. We compare state-level distributions of GeoReddit users in faith-based subreddits with 2020 The Association of Religion Data Archives (ARDA) that capture the known geographic concentration of each group. The subgroups examined are Latter-day Saints, Baptists, Lutherans, Muslims, and Jews.

Figure W6: Latter-day Saints — GeoReddit vs. ARDA 2020.



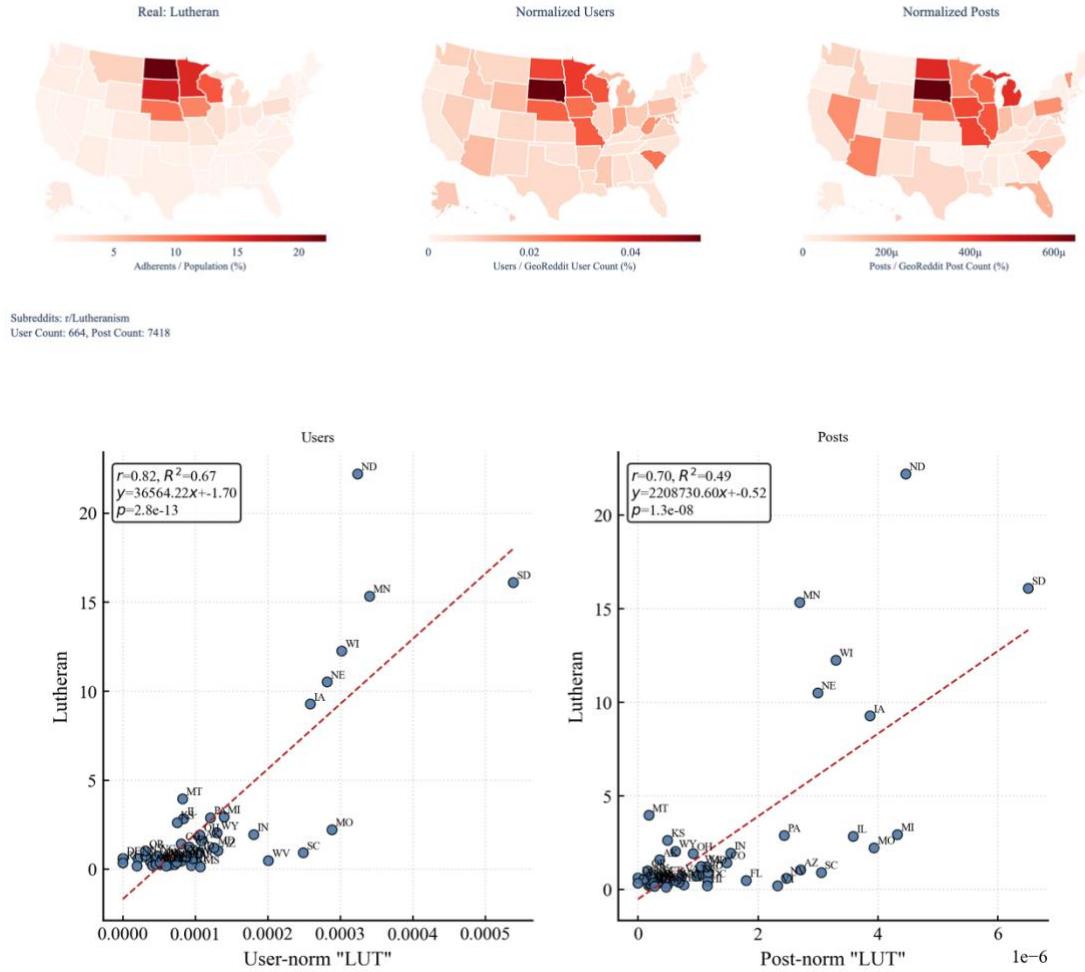
Notes: (top) State-level distribution maps (ARDA benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ARDA.

Figure W7: Baptists — GeoReddit vs. ARDA 2020.



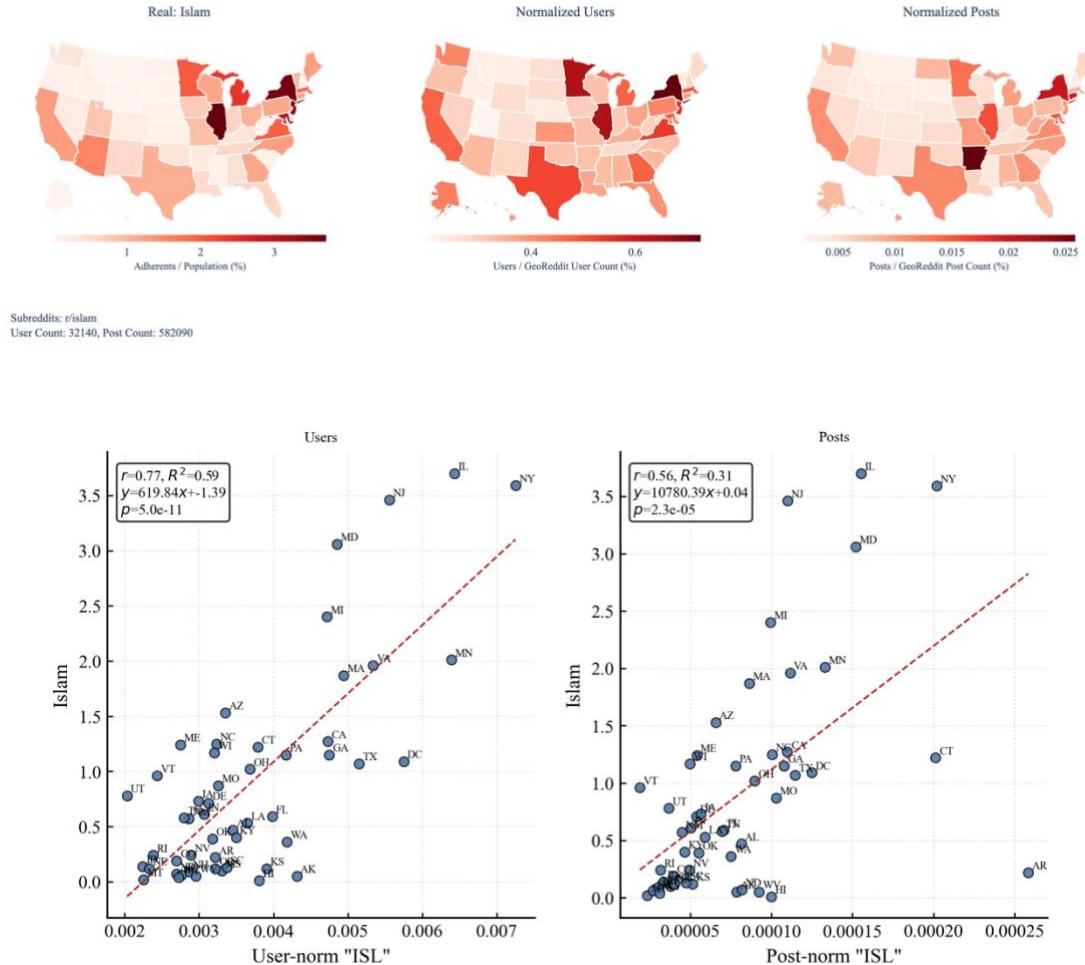
Notes: (top) State-level distribution maps (ARDA benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ARDA.

Figure W8: Lutheran — GeoReddit vs. ARDA 2020.



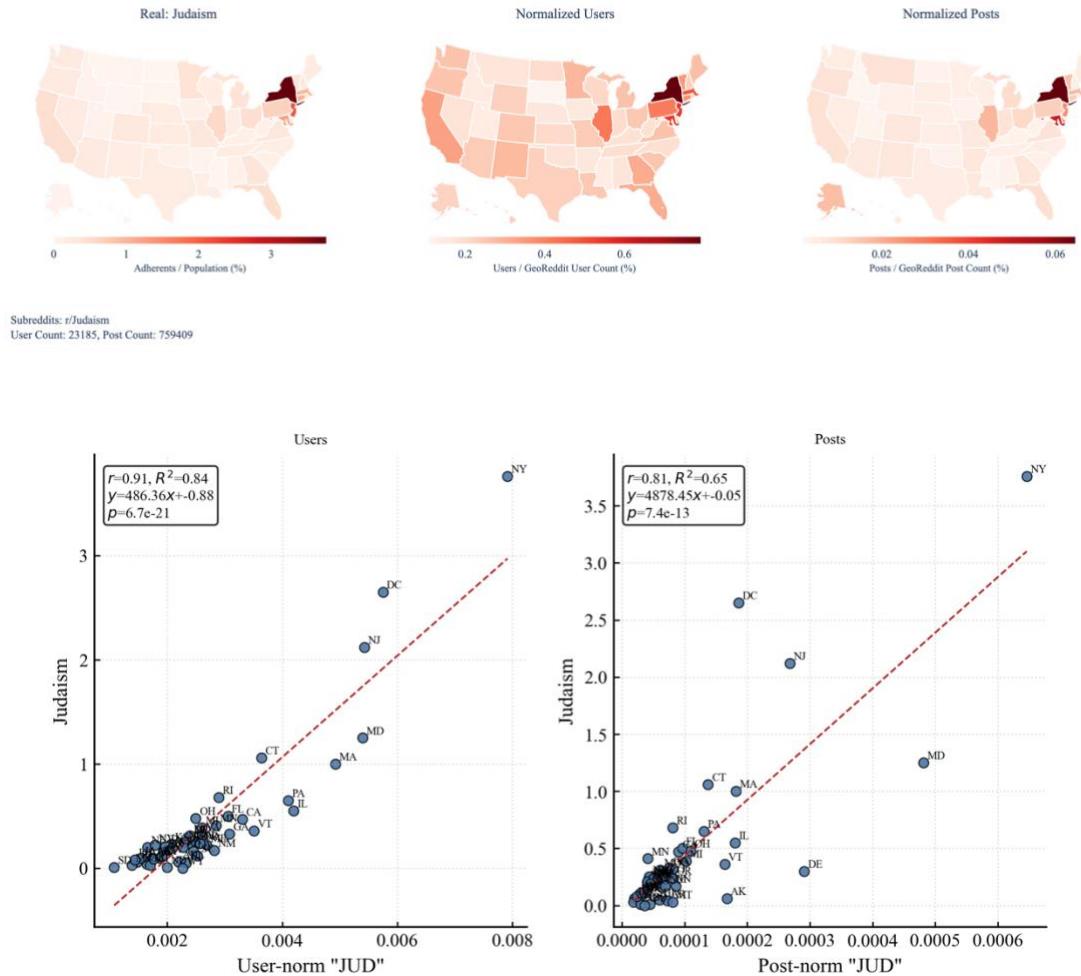
Notes: (top) State-level distribution maps (ARDA benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ARDA.

Figure W9: Muslims — GeoReddit vs. ARDA 2020.



Notes: (top) State-level distribution maps (ARDA benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ARDA.

Figure W10: Judaism — GeoReddit vs. ARDA 2020.



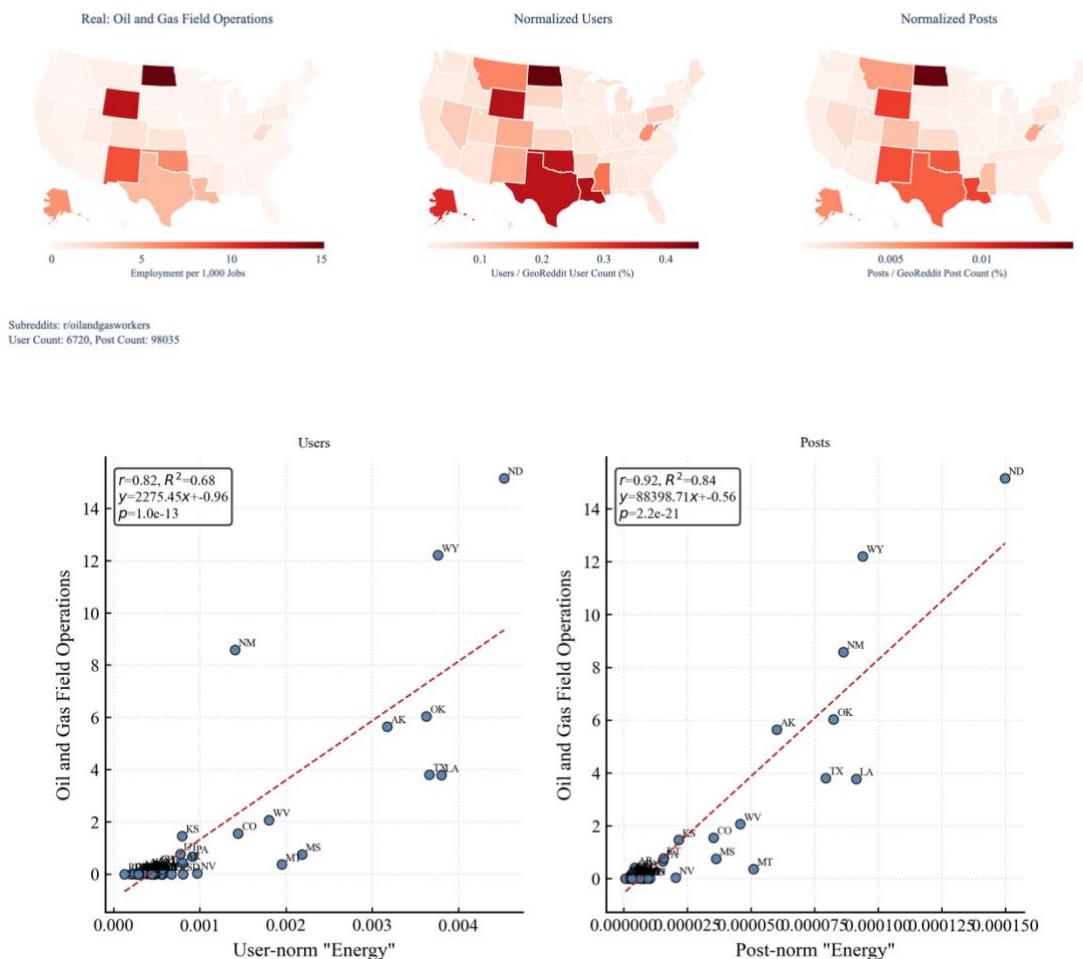
Notes: (top) State-level distribution maps (ARDA benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with ARDA.

Occupational Validation

This section presents state-level distribution maps and scatterplots for selected occupations, comparing GeoReddit-based geolocation with 2023 Bureau of Labor Statistics (BLS) benchmarks. Specifically, we use BLS Standard Occupational Classification (SOC) codes 47-5011 (Derrick Operators, Oil and Gas), 47-5012 (Rotary Drill Operators, Oil and Gas), 47-5013 (Service Unit Operators, Oil and Gas), and 47-5071 (Roustabouts, Oil and Gas) to capture

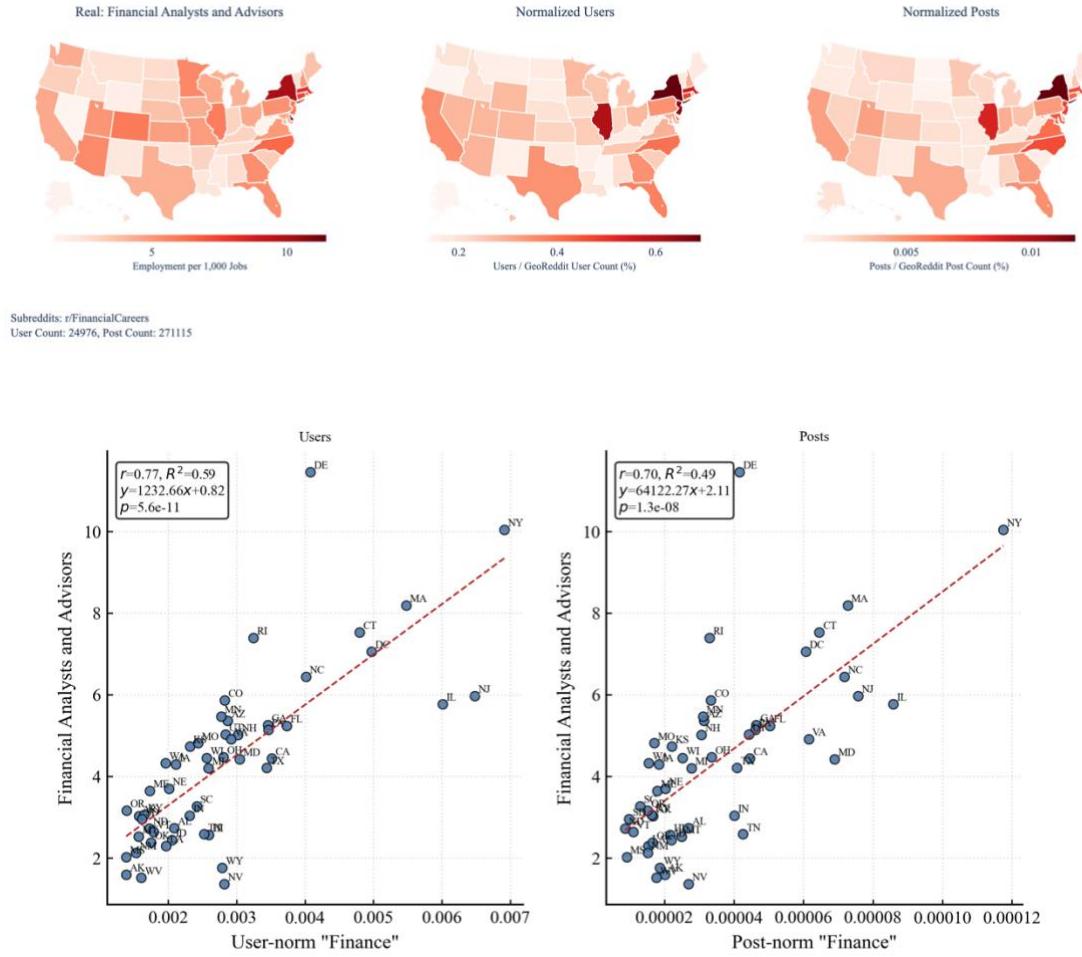
oil and gas field operations; 13-2050 (Financial Analysts and Advisors) for finance professionals; 41-3011 (Advertising Sales Agents) for advertising professionals; and 15-1252 (Software Developers) and 15-1254 (Web Developers) for software and web development. The benchmark measure is employment per 1,000 jobs, which expresses how many workers in a given occupation are employed for every 1,000 employed persons in a state, thereby standardizing occupational presence relative to overall state employment.

Figure W11: Oil and Gas Workers — GeoReddit vs. BLS 2023.



Notes: (top) State-level distribution maps (BLS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with BLS.

Figure W12: Finance Professionals — GeoReddit vs. BLS 2023.



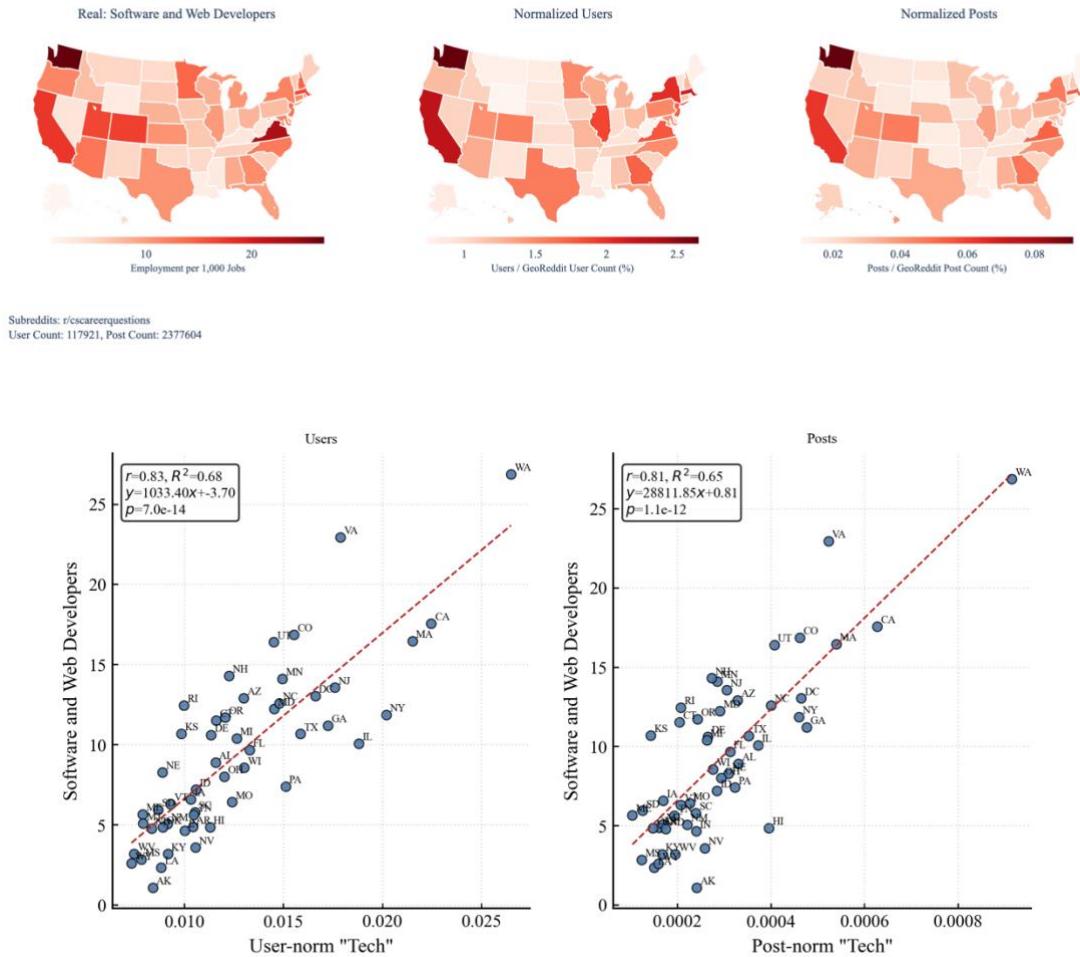
Notes: (top) State-level distribution maps (BLS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with BLS.

Figure W13: Advertising Professionals — GeoReddit vs. BLS 2023.



Notes: (top) State-level distribution maps (BLS benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with BLS.

Figure W14: Software and Web Developers—GeoReddit vs. BLS 2023.



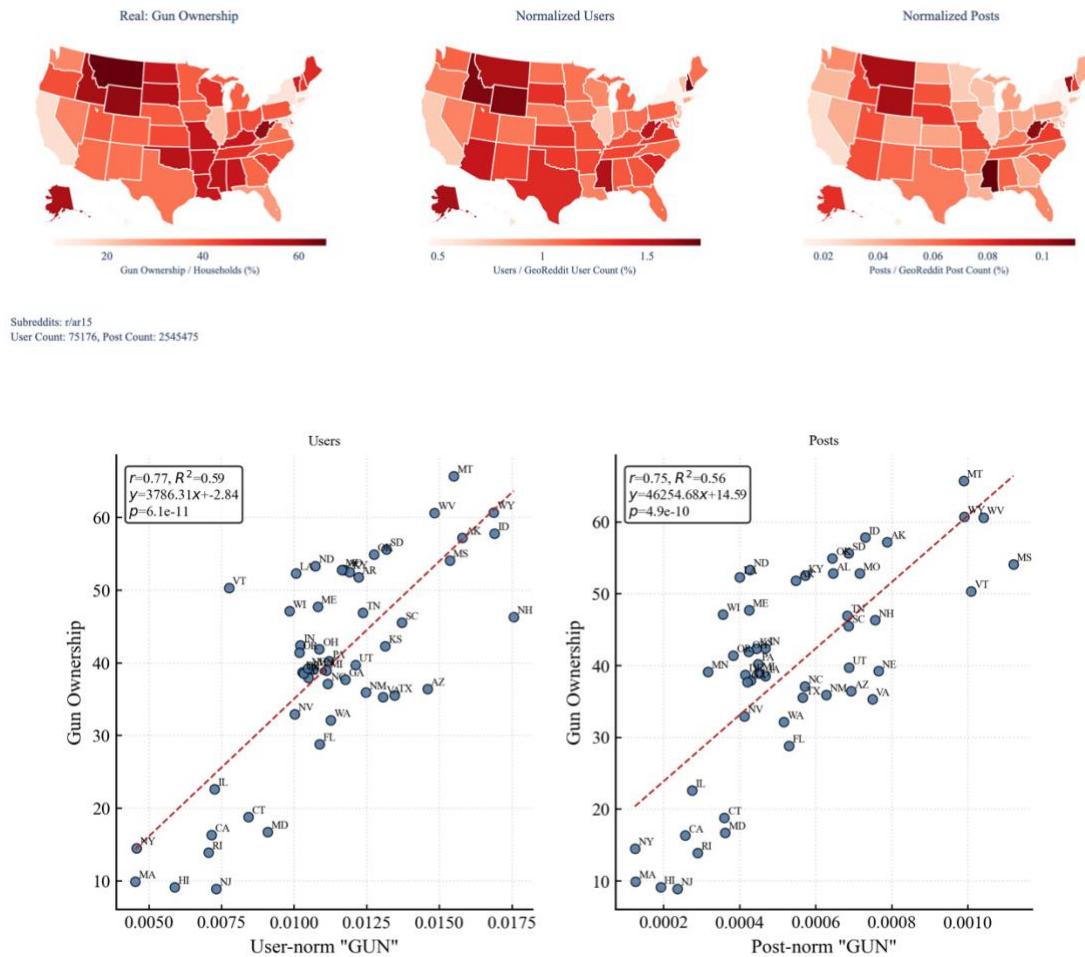
Notes: (top) State-level distribution maps (BLS benchmark, GeoReddit users, GeoReddit

posts); (bottom) Scatterplot comparing GeoReddit with BLS.

Behavioural Validation

To extend validation beyond census- and labor-based indicators, we compare GeoReddit participation in firearms-related discussions with survey-based state-level household gun ownership estimates. Focusing on r/ar15 (75,325 users; 2,550,165 posts)—a subreddit devoted to one of the most discussed firearms in the U.S.—we observe a strong correlation, $r = 0.77$, $p < .0001$.

Figure W15: Household Gun Ownership—GeoReddit vs. Wisevoter 2023.



Notes: (top) State-level distribution maps (Wisevoter benchmark, GeoReddit users, GeoReddit posts); (bottom) Scatterplot comparing GeoReddit with Wisevoter.

Across demographic, religious, occupational, and firearms-related validations, the supplementary figures consistently demonstrate that GeoReddit-based state-level distributions align closely with external benchmarks. Correlations are generally high, often exceeding 0.75, and visual inspections of the maps confirm that geographic patterns observed in official statistics are well reproduced in GeoReddit, supports the robustness of our rule-based geolocation procedure and its ability to capture meaningful variation in state-level user representation.

Web Appendix C: Robustness Analysis for Vignette 1 (Pepsi's Brand Crisis)

We additionally conducted the analysis with different base-rates for word share (proportion [prop]: the number of emotional words per comment; frequency [freq]: the total number of words per comment), for different cut-off criteria of observations per state ($n = 30$, and $n = 50$), and for different correlation estimation methods (Pearson and Spearman). Results remained the same for different word share base-rates and correlation estimation methods. However, imposing a more conservative threshold of $n = 50$ per state showed descriptively the same pattern, but failed to reach significance due to a lower number of analyzed states. In the following we list the results for the different permutations of word-share base-rate, cut-off thresholds, and correlation estimation method.

Figure W16: Pearson correlation of frequency of anger related words for at least $n \geq 30$ observations per state.

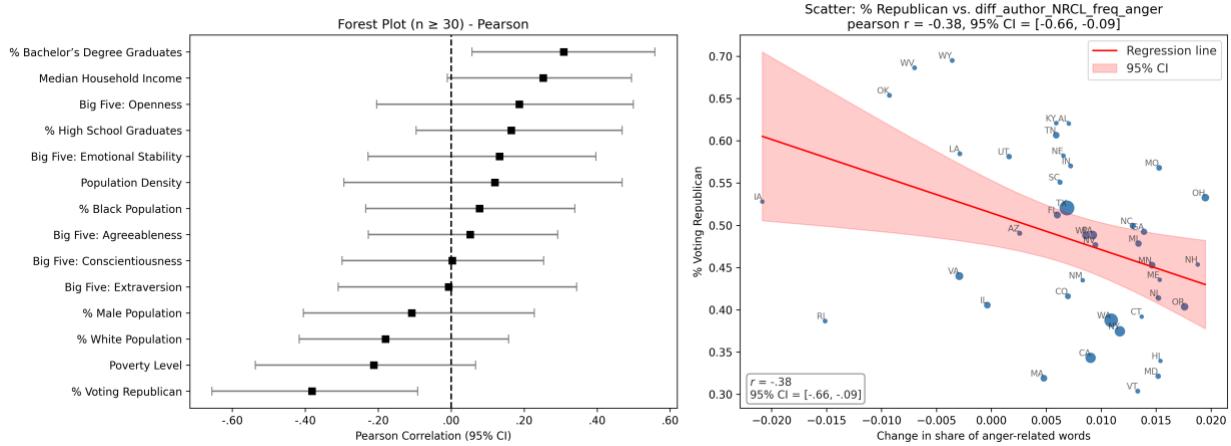


Figure W17: Pearson correlation of frequency of anger related words for at least $n \geq 50$ observations per state.

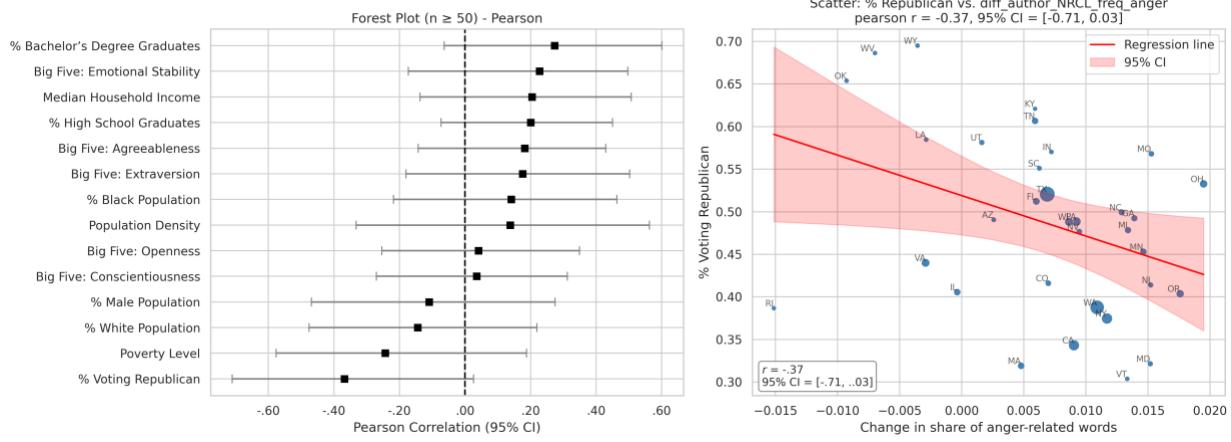


Figure W18: Spearman correlation of frequency of anger related words for at least $n \geq 30$ observations per state.

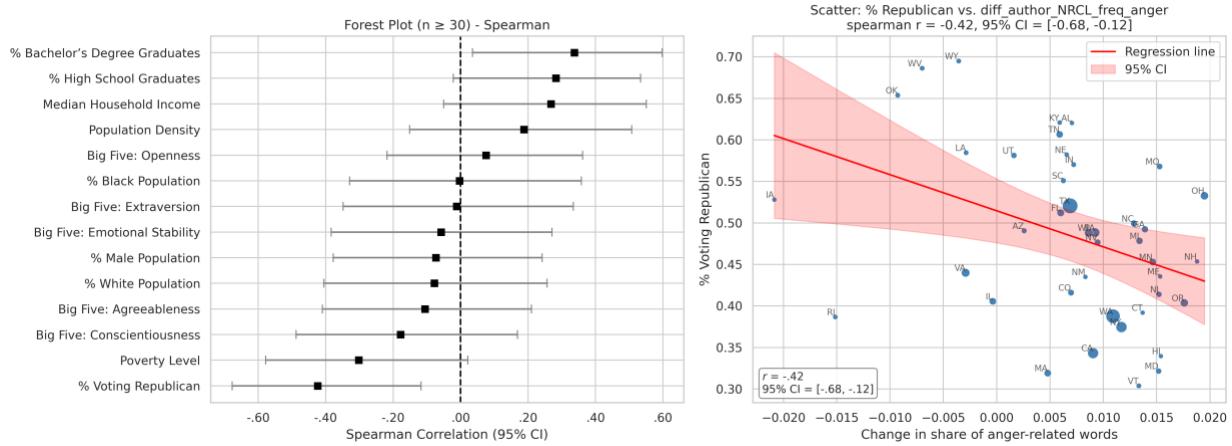


Figure W19: Spearman correlation of frequency of anger related words for at least $n \geq 50$ observations per state.

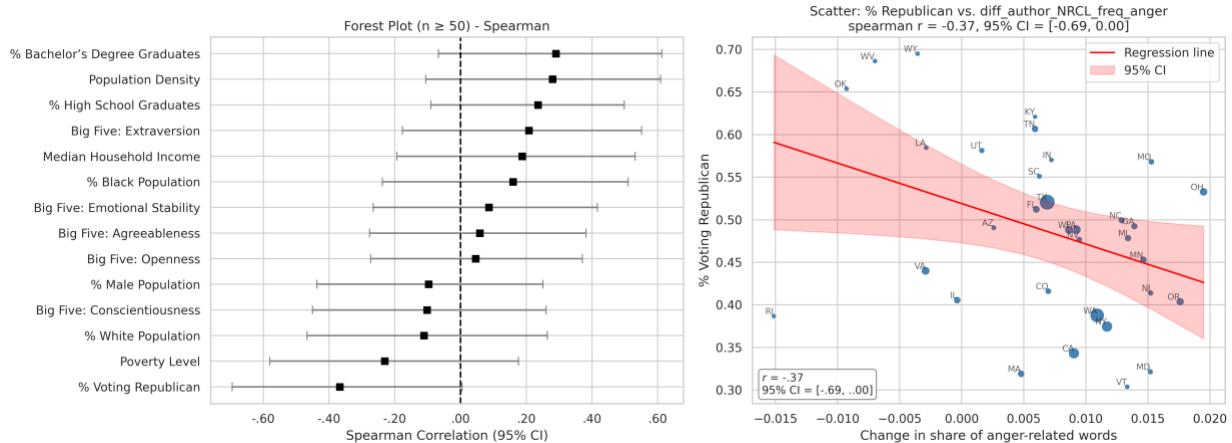


Figure W20: Pearson correlation of proportion of anger related words for at least $n \geq 30$ observations per state.

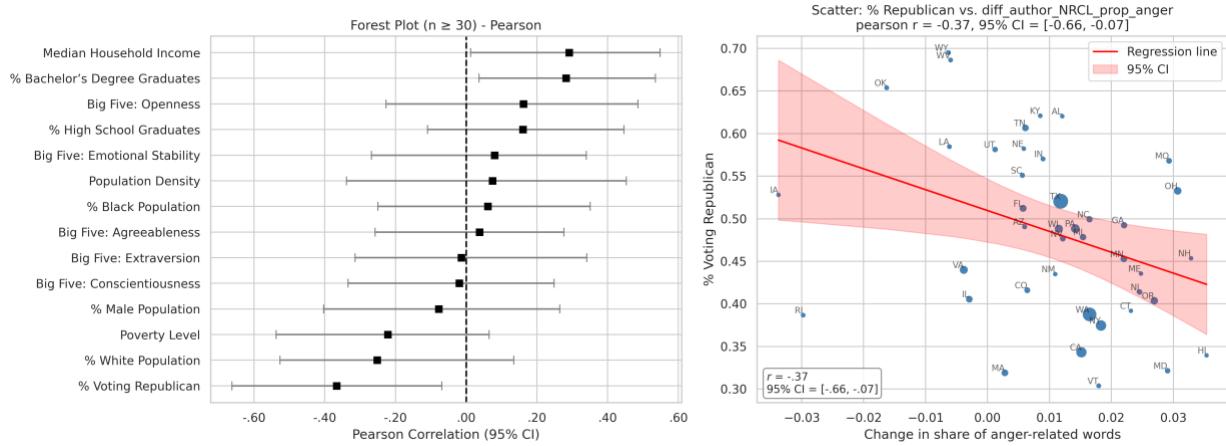


Figure W21: Pearson correlation of proportion of anger related words for at least $n \geq 50$ observations per state.

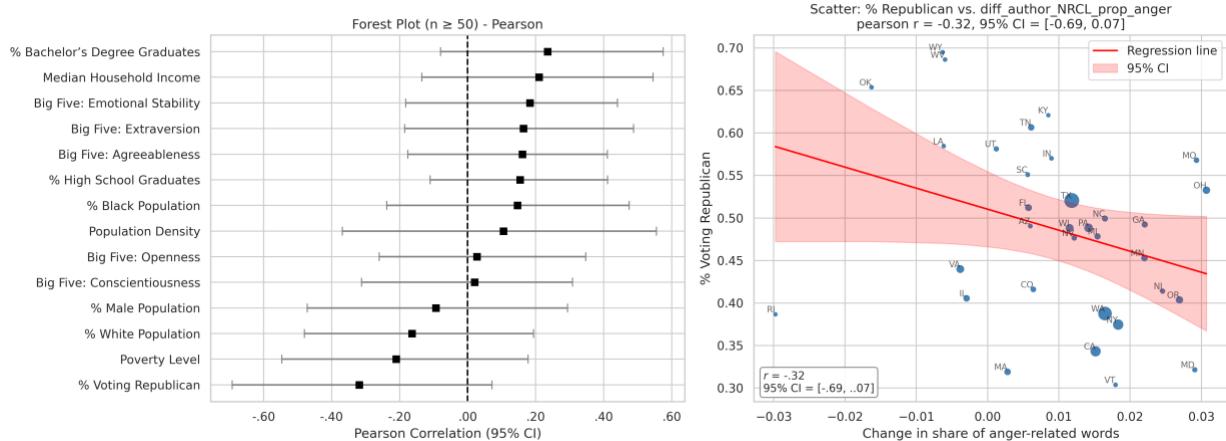


Figure W22: Spearman correlation of proportion of anger related words for at least $n \geq 30$ observations per state.

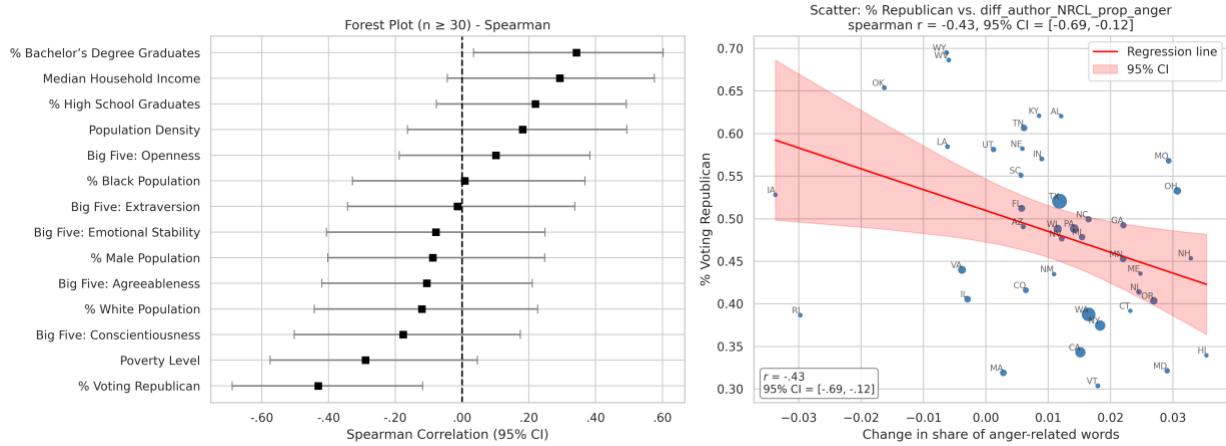
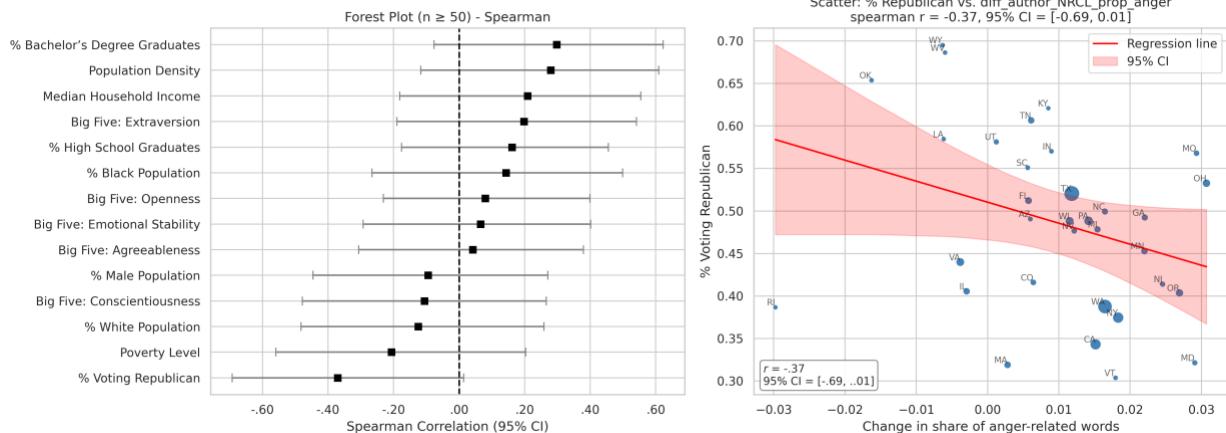


Figure W23: Spearman correlation of proportion of anger related words for at least $n \geq 50$ observations per state.



Web Appendix D: Robustness Analysis for Vignette 2 (The Legalization of Marijuana)

We additionally conducted the analysis of Vignette 2 with different base-rates for word share (proportion [prop]: the number of emotional words per comment; frequency [freq]: the total number of words per comment), for different cut-off criteria of observations per state ($n = 30$, $n = 50$, and $n = 100$), for different event variables `vote_012` (months before, month at, and months after voting to legalize marijuana) and `poss_012` (months before, month at, and months after possession was officially legal), and across different sentiments (i.e., positive and negative). Results remained similar across different word share base-rates cut-off criteria, and event variables. Overall positive sentiment shows an inverse pattern as expected.

In the following we list the results for the different permutations of word-share base-rate, cut-off thresholds, event variables, and sentiments. The figures show on the upper left side the standardized parameter estimates of the linear-mixed model (see below) and on the bottom left side a barplot of the predicted unstandardized parameter estimates. The right side shows the z-standardized time trend of states which legalized marijuana ± 12 months before and after the event variable.

The linear-mixed model estimates the sentiment value by event-variable level. The linear-mixed model includes random intercepts for states and accounts for the general time trend:

$$Y_{ij} = \beta_0 + \sum_{k=1}^K \beta_k D_{ijk} + \gamma T_j + b_i + \varepsilon_{ij}$$

Y_{ij} : value of sentiment for month j in state i

T_j : value of time for month j

D_{ijk} : event variable with k -th level (i.e., 0,1,2) indicating before, event month, and after event

β_0 : global intercept

β_k : fixed effect for the k -th level of the event variable

γ : fixed effect for time

b_i : random intercept for state i with normal distribution

ε_{ij} : residual error term with normal distribution

Figure W24: Influence of voting to legalize marijuana on frequency of sentiment for at least $n \geq 30$ observations per state.

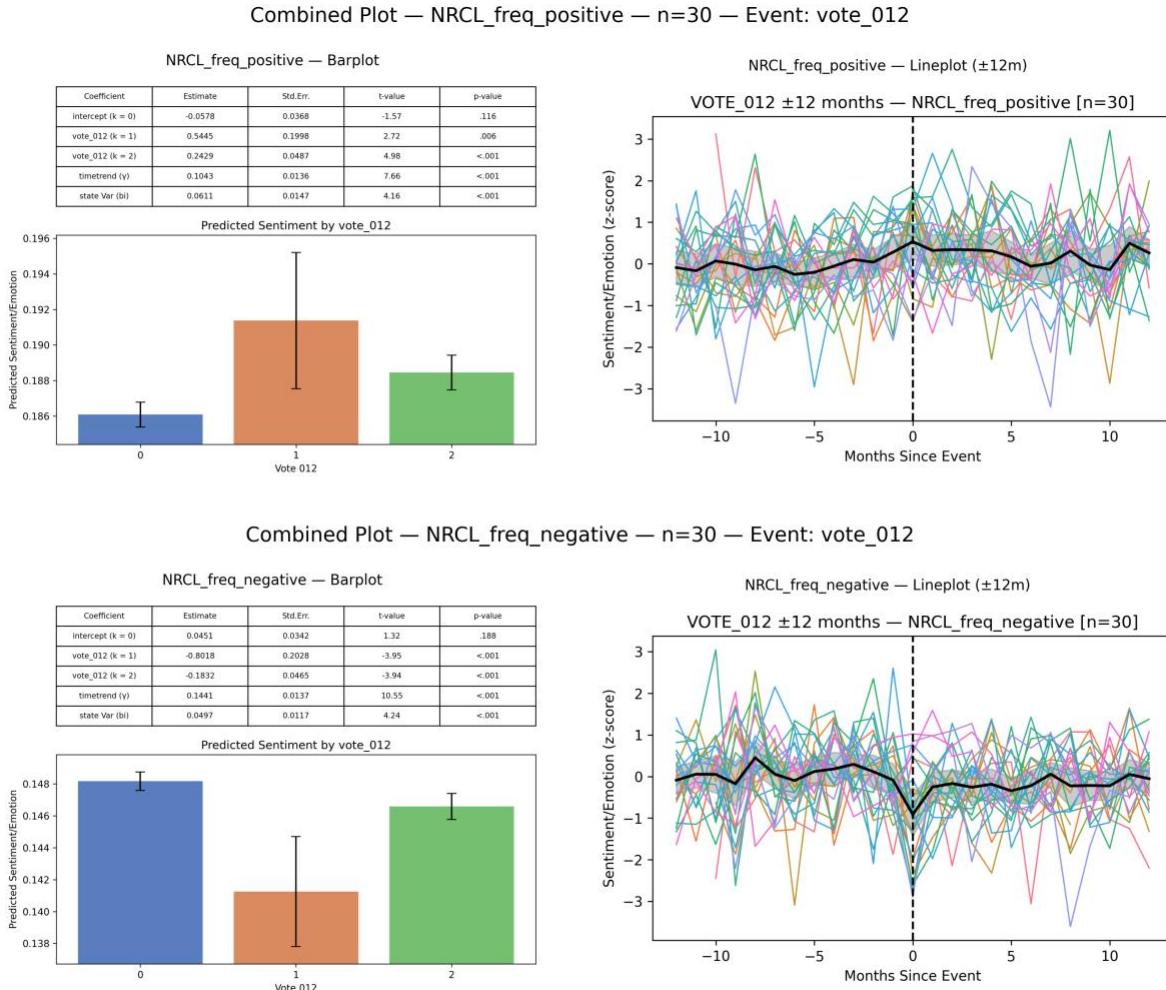


Figure W25: Influence of voting to legalize marijuana on frequency of sentiment for at least $n \geq 50$ observations per state.

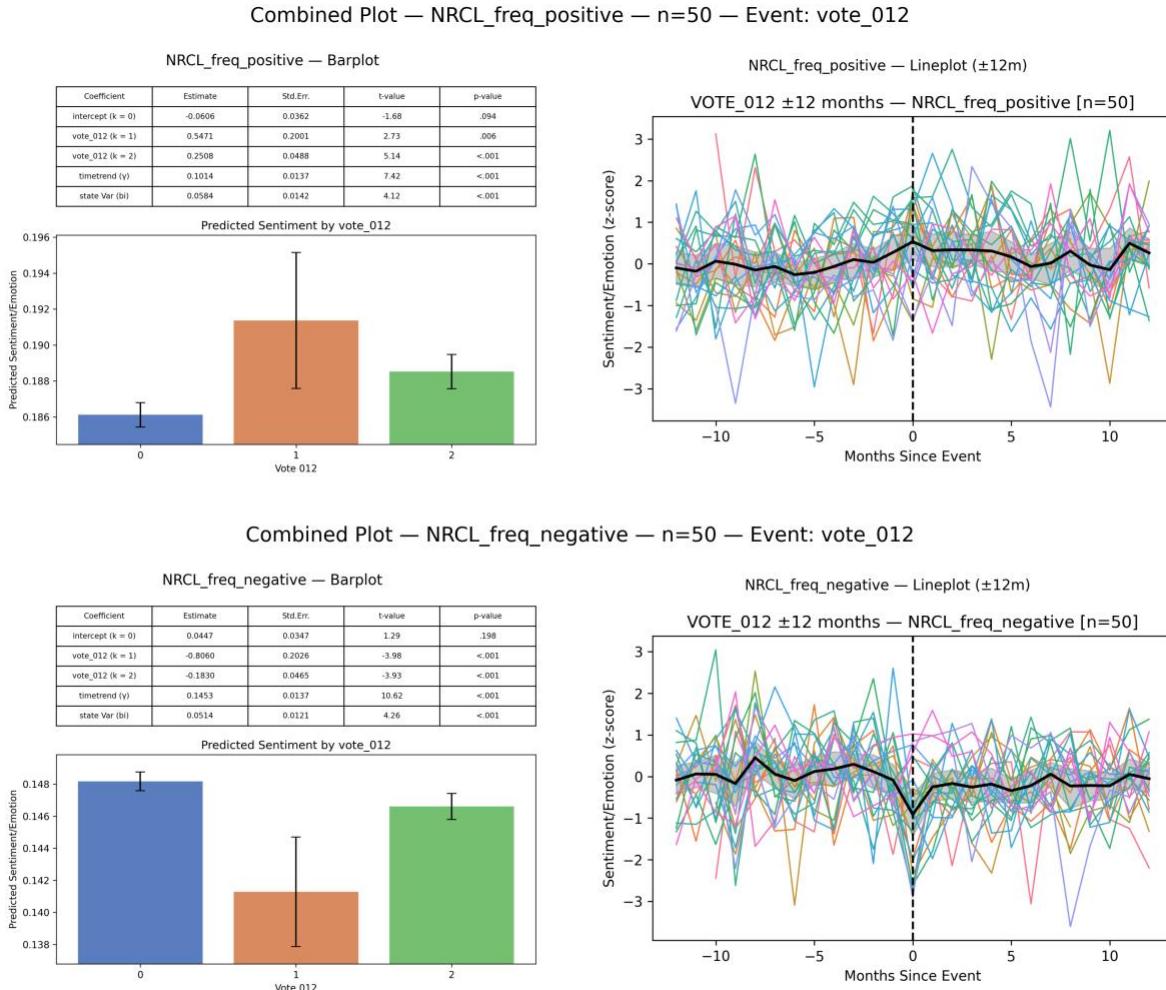


Figure W26: Influence of voting to legalize marijuana on frequency of sentiment for at least $n \geq 100$ observations per state.

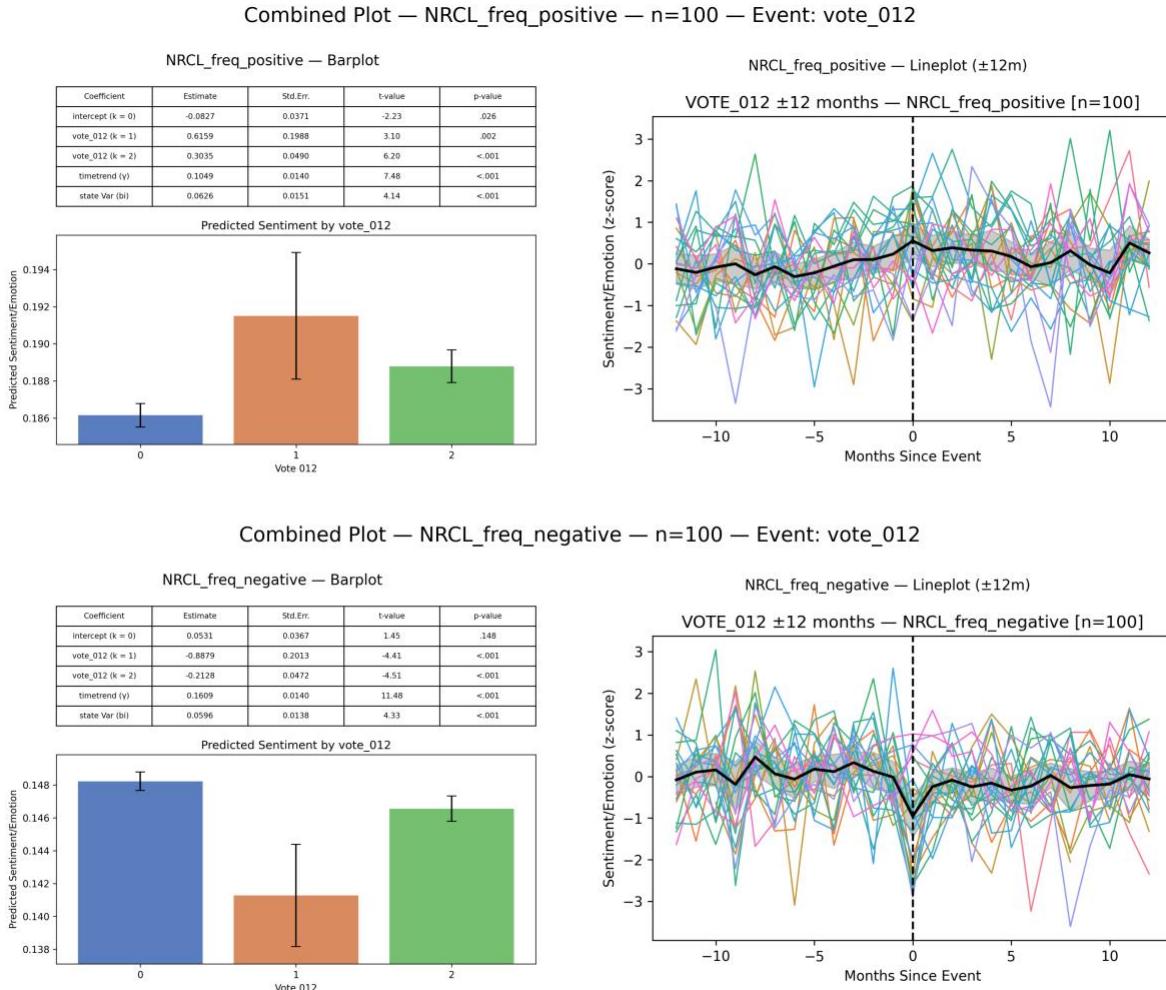


Figure W27: Influence of legal possession of marijuana on frequency of sentiment for at least $n \geq 30$ observations per state.

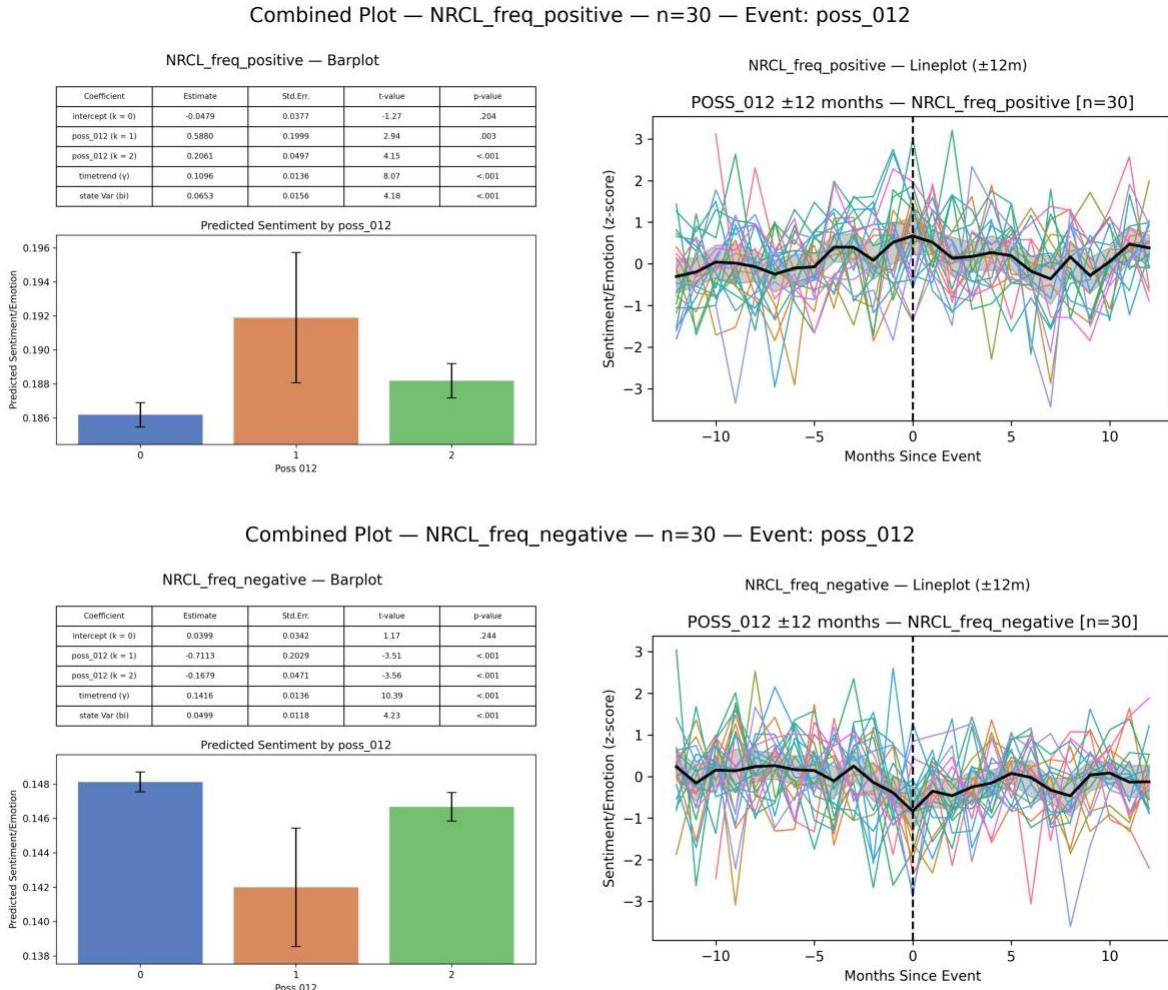


Figure W28: Influence of legal possession of marijuana on frequency of sentiment for at least $n \geq 50$ observations per state.

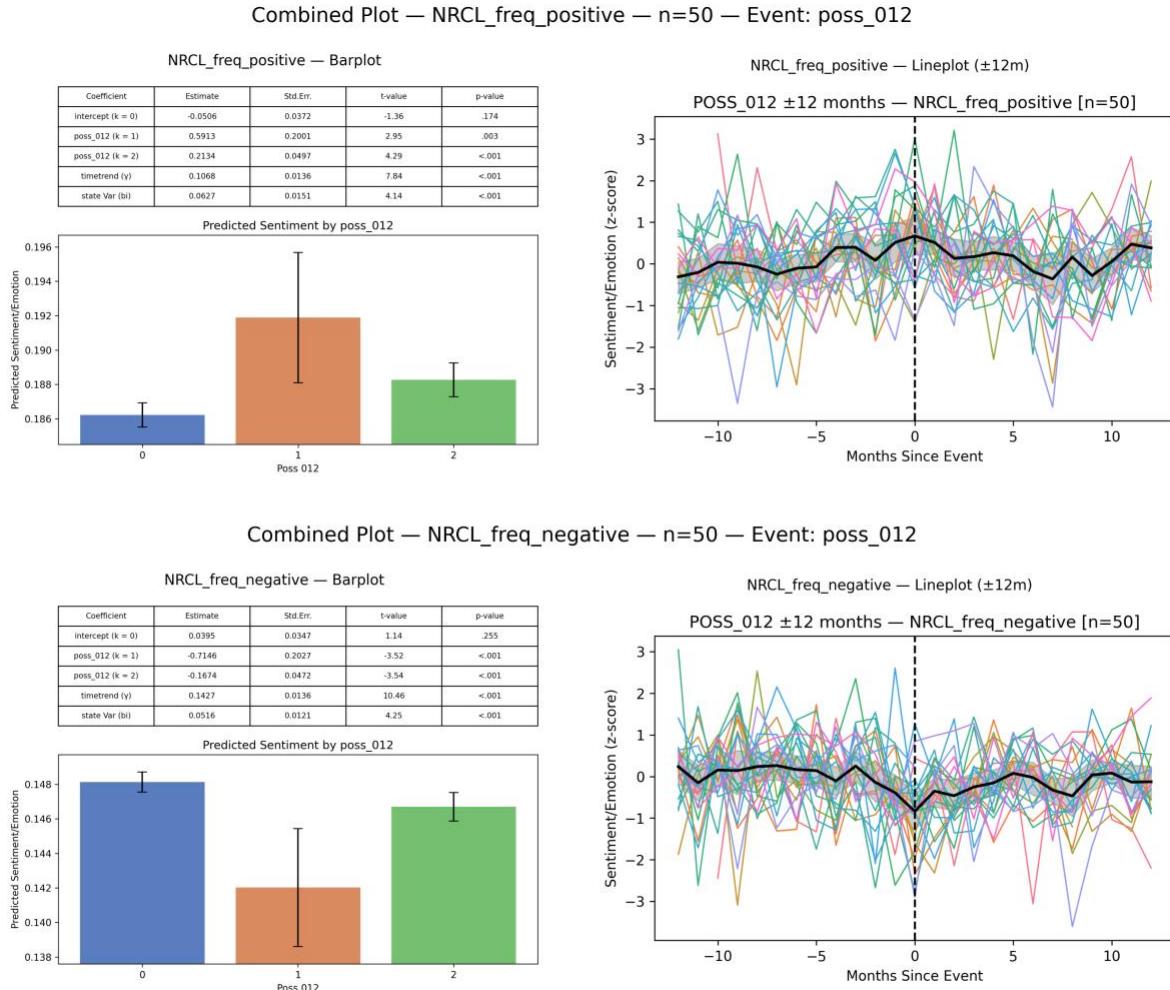


Figure W29: Influence of legal possession of marijuana on frequency of sentiment for at least $n \geq 100$ observations per state.

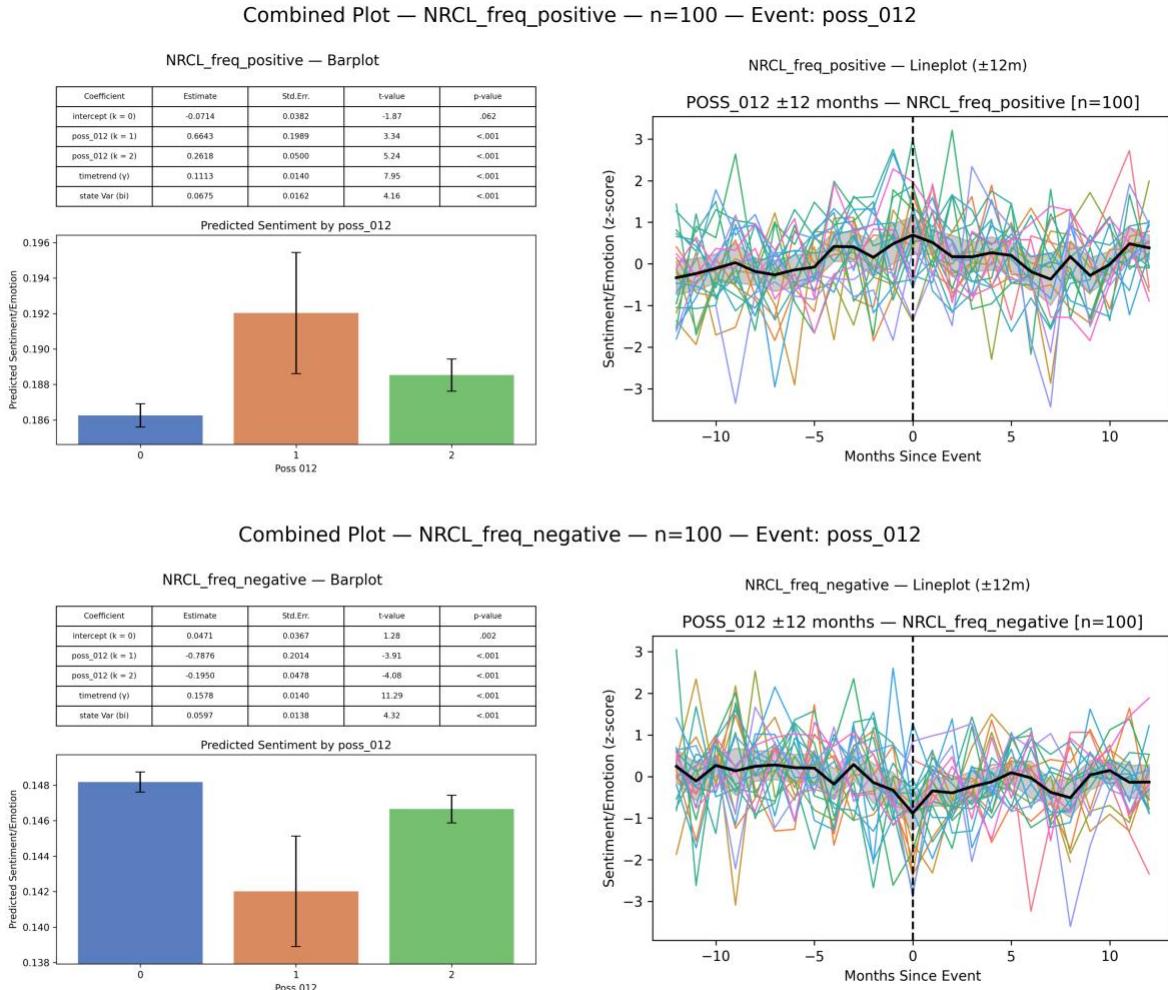


Figure W30: Influence of voting to legalize marijuana on proportion of sentiment for at least $n \geq 30$ observations per state.

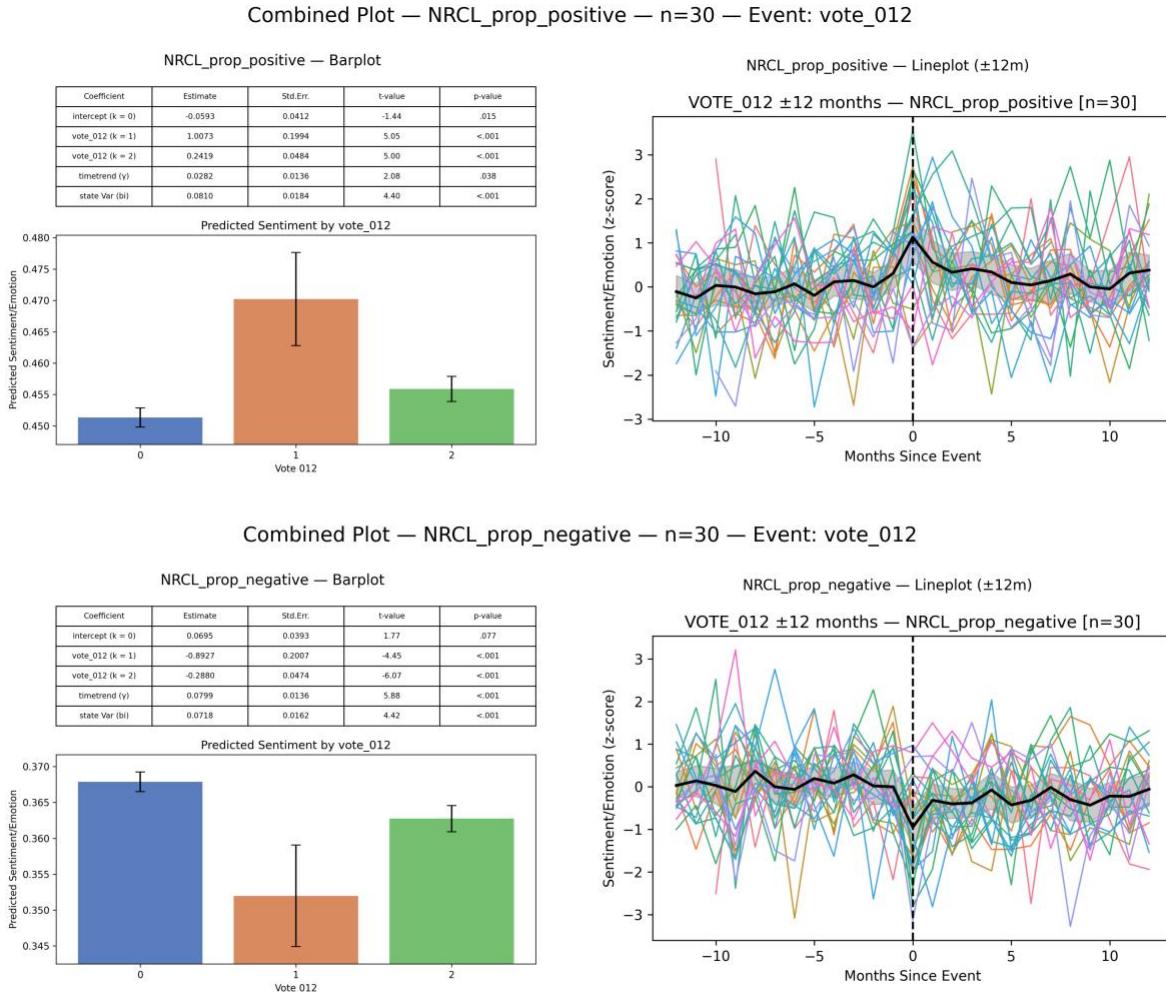


Figure W31: Influence of voting to legalize marijuana on proportion of sentiment for at least $n \geq 50$ observations per state.

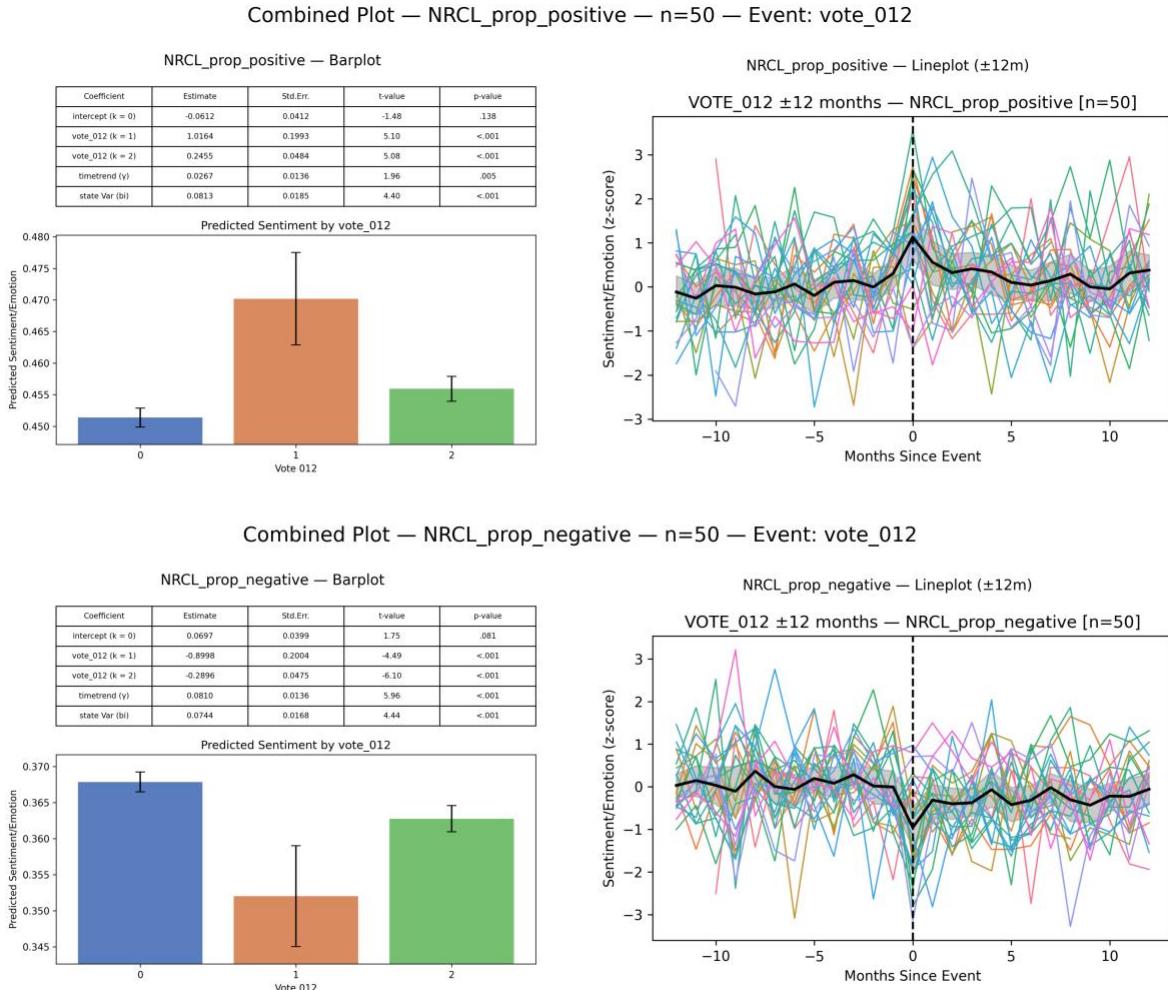


Figure W32: Influence of voting to legalize marijuana on proportion of sentiment for at least $n \geq 100$ observations per state.

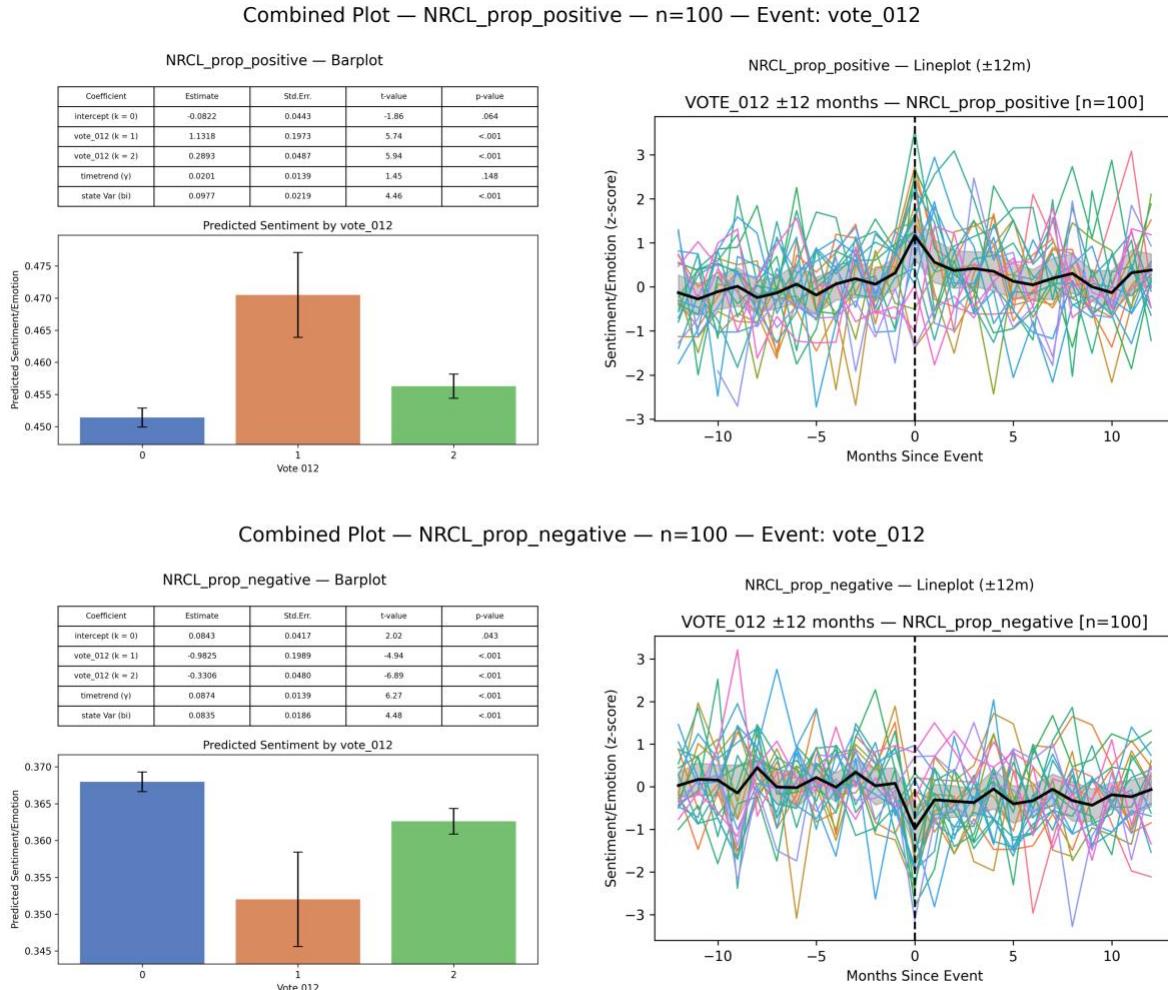


Figure W33: Influence of legal possession of marijuana on proportion of sentiment for at least n ≥ 30 observations per state.

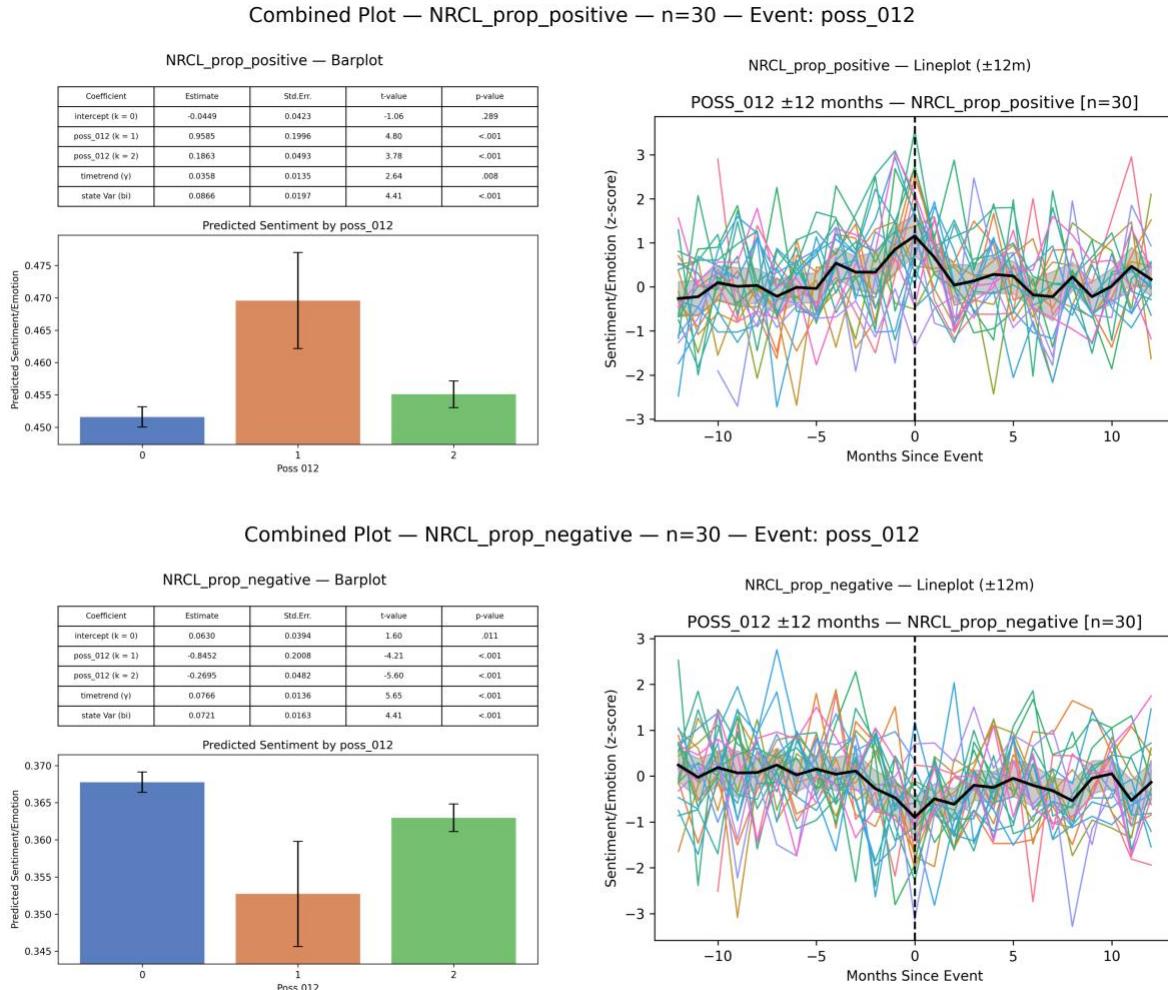


Figure W34: Influence of legal possession of marijuana on proportion of sentiment for at least n ≥ 50 observations per state.

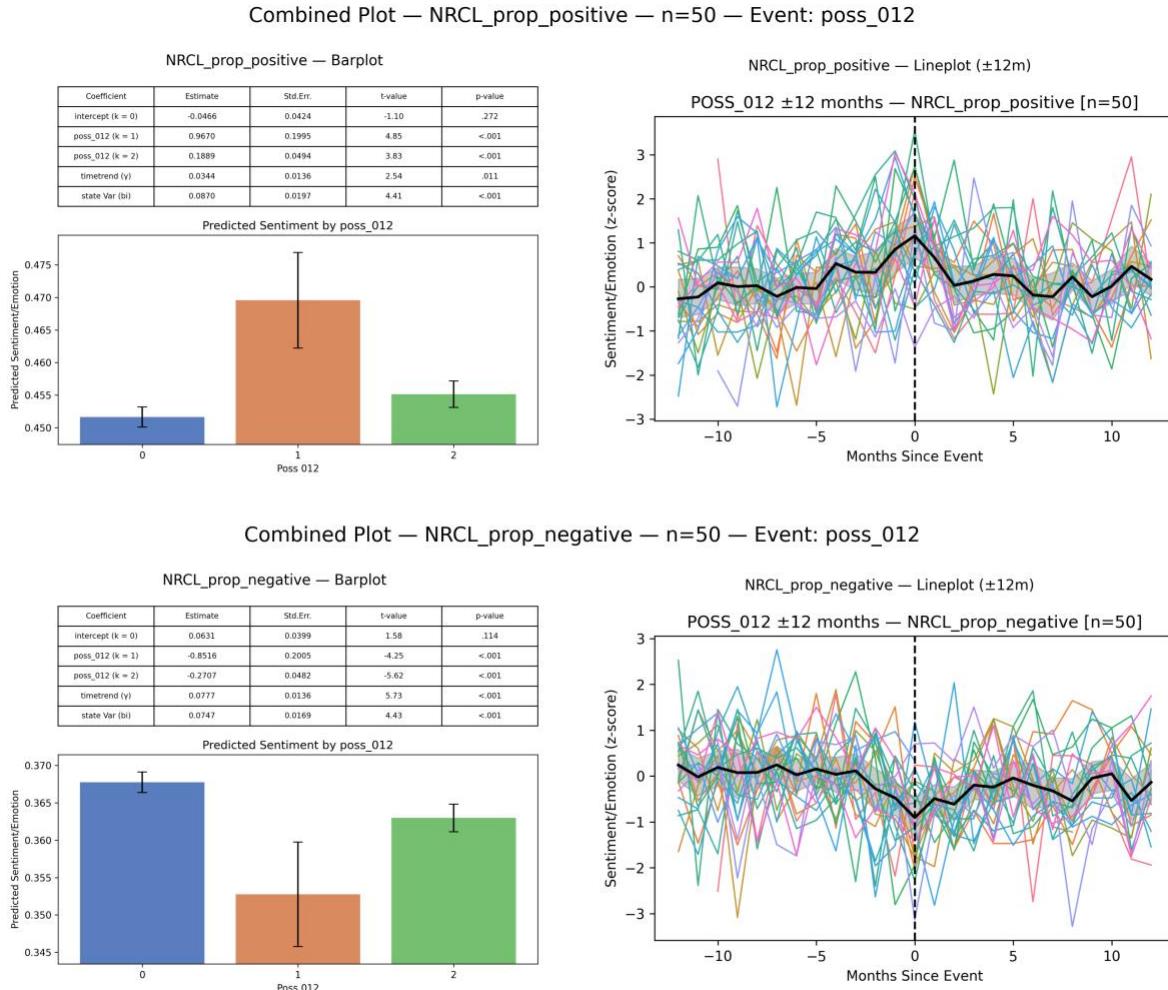
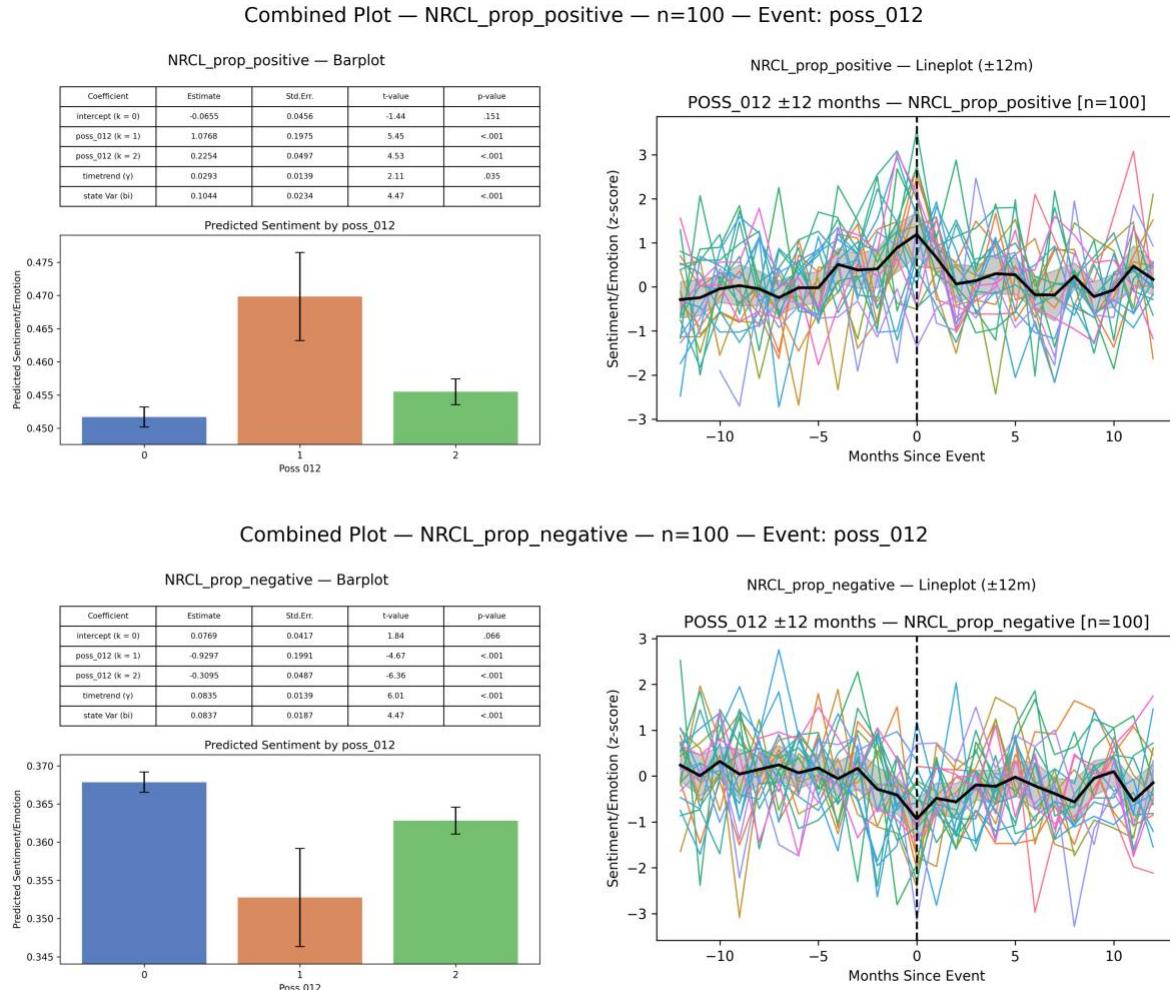


Figure W35: Influence of legal possession of marijuana on proportion of sentiment for at least n ≥ 100 observations per state.



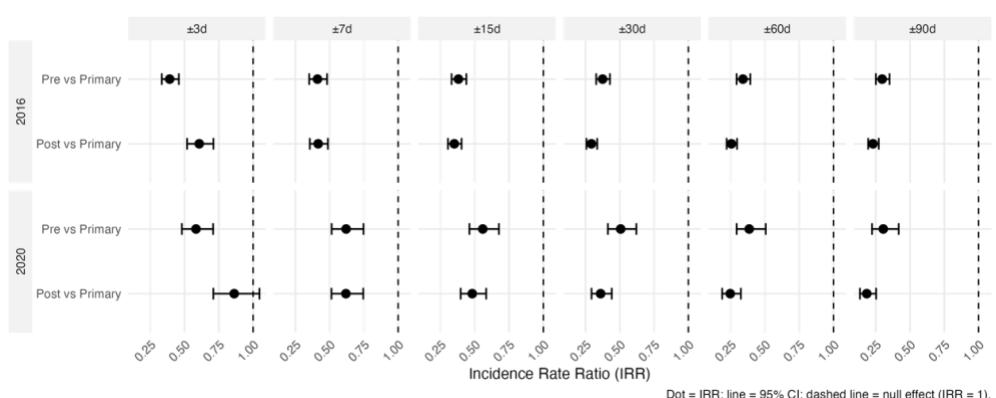
Web Appendix E: Robustness Analyses for Vignette 3 (Sanders Primaries)

This appendix reports additional results complementing the main analyses of r/SandersForPresident activity during the 2016 and 2020 Democratic primaries. We report regression estimates across alternative event windows (± 3 , ± 15 , ± 30 , ± 60 , and ± 90 days), all of which converge successfully and confirm the primary-day spike documented in the main text. We also include corresponding forest plots and event-aligned time series, standardized to a pre-primary baseline, which consistently show elevated participation on primary days followed by sharp declines.

Unique-user Results

Table W2 reports NB-GLMM estimates of unique-user counts in r/SandersForPresident relative to the primary day across alternative event windows. In 2016, user participation was sharply elevated on primary days across all windows. Daily specifications yield Pre vs. Primary IRRs between 0.30 and 0.41 and Post vs. Primary IRRs between 0.23 and 0.60, with all 95% confidence intervals fully below one ($p < .001$), implying that activity on the primary day was roughly two to three times higher than in surrounding periods.

Figure W36: Forest plot of incidence rate ratios (IRRs) for unique-user count around state primary days (r/SandersForPresident), with 95% confidence intervals across alternative time windows.



In 2020, event-triggered spikes are again evident. In short-term windows (± 3 – 30 d), the contrasts are no larger than in 2016—Pre vs. Primary IRRs span 0.51–0.62 and Post vs. Primary 0.36–0.62, indicating less pronounced short-run drops than 2016. By contrast, in the longer windows, the declines are steeper in 2020 for the Post vs. Primary contrasts, falling to 0.18 (± 60 d) and 0.25 (± 90 d), compared to 0.23 and 0.26 in 2016. Pre vs. Primary estimates are similar at ± 90 d (both ≈ 0.30) and slightly higher in 2020 at ± 60 d (0.39 vs. 0.34 in 2016). Only one contrast is not statistically significant—Post (± 3 d) (IRR = 0.86, 95% CI [0.71, 1.05], $p = .133$)—as indicated in the forest plot (Figure W36) where its confidence interval crossed the reference line at IRR = 1. The sharper long-run post-primary decline in 2020 can be explained by two factors: the primaries were compressed into a shorter calendar, and Sanders ended his campaign early in April. As a result, users had fewer opportunities and less motivation to stay active after his dropout, leading to a faster drop in participation compared to 2016.

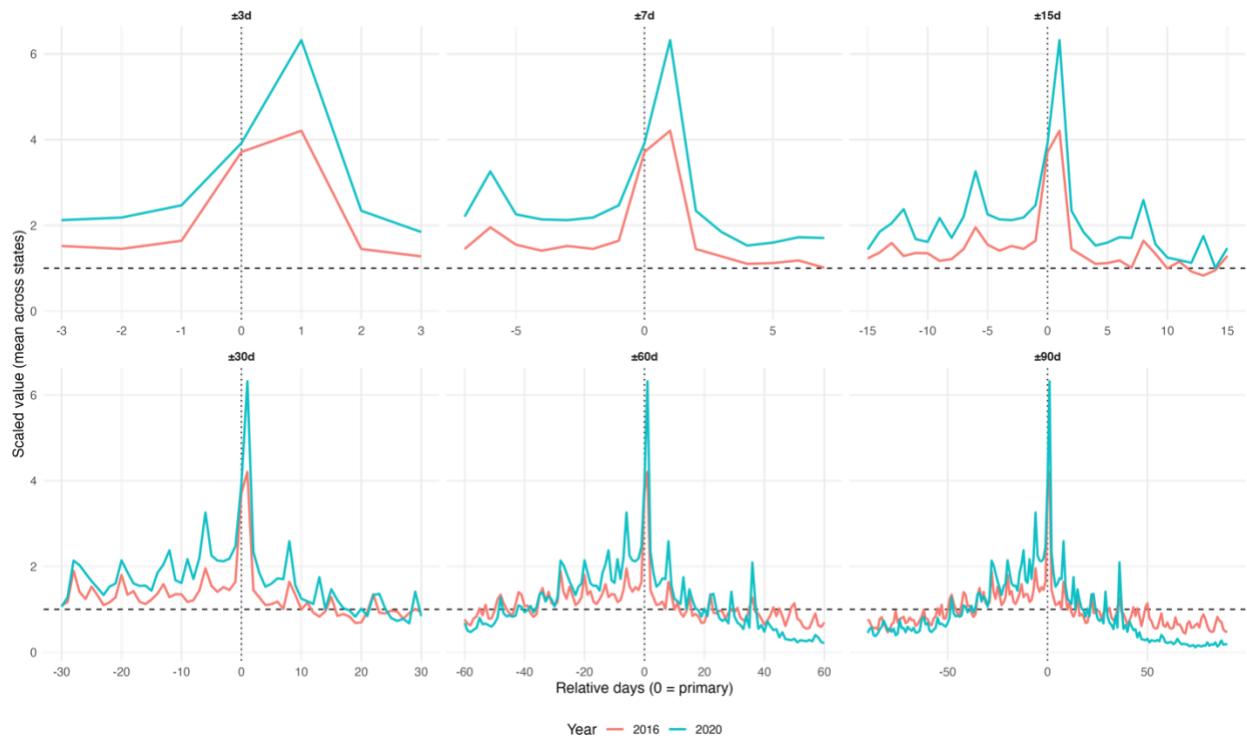
Table W2: NB-GLMM estimates of pre- and post-primary unique-user count relative to primary day (IRR, 95% CI, p-values) by year and window.

Year	Window	Pre vs. Primary (IRR/95% CI)	Post vs. Primary (IRR/95% CI)
2016	± 3 d	0.39 / [0.33, 0.46]***	0.60 / [0.52, 0.71]***
2016	± 7 d	0.41 / [0.35, 0.48]***	0.42 / [0.36, 0.49]***
2016	± 15 d	0.38 / [0.33, 0.44]***	0.35 / [0.31, 0.40]***
2016	± 30 d	0.37 / [0.33, 0.43]***	0.29 / [0.26, 0.33]***
2016	± 60 d	0.34 / [0.29, 0.39]***	0.26 / [0.22, 0.30]***
2016	± 90 d	0.30 / [0.25, 0.35]***	0.23 / [0.19, 0.27]***
2020	± 3 d	0.58 / [0.48, 0.71]***	0.86 / [0.71, 1.05]
2020	± 7 d	0.62 / [0.51, 0.75]***	0.62 / [0.51, 0.75]***
2020	± 15 d	0.56 / [0.46, 0.68]***	0.48 / [0.40, 0.58]***
2020	± 30 d	0.51 / [0.41, 0.62]***	0.36 / [0.29, 0.44]***
2020	± 60 d	0.39 / [0.29, 0.51]***	0.25 / [0.19, 0.32]***
2020	± 90 d	0.31 / [0.22, 0.42]***	0.18 / [0.14, 0.25]***

Notes: Stars denote significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$. All NB-GLMMs converged.

Also, these differences in magnitude show that Sanders supporters in 2016 were disproportionately concentrated on primary days, producing two- to threefold jumps in participation, whereas in 2020 the spikes, though still evident, represented a more moderate 1.5- to 2-fold increase over surrounding activity levels. This divergence reflects the contrasting campaign contexts: Sanders's 2016 insurgent run was novel, grassroots-driven, and heavily event-focused, making each state primary a focal point of digital mobilization, while the 2020 campaign was more institutionalized and ended with an earlier withdrawal, reducing the extent to which participation was concentrated on single-day events.

Figure W37: Event-aligned time series of unique-users in r/SandersForPresident, standardized to a pre-primary baseline, showing primary-day spikes and surrounding declines.



Moving beyond the regression framework, Figure W37 presents event-aligned time series standardized to a pre-primary baseline. These curves offer a complementary view of participation

dynamics, clearly illustrating the sharp surge on the primary day and the subsequent return to baseline in both election cycles.

Post Result

In addition to unique-user counts, we analyze post counts to capture the intensity of participation. While unique users provide a measure of how many individuals engaged on primary days, posts reflect how actively these individuals contributed. Examining posts therefore allows us to assess whether the primary-day surges observed at the user level were also accompanied by higher volumes of content creation. This distinction is important: a campaign might mobilize many supporters who each contribute minimally, or fewer supporters who post intensively. By analyzing both dimensions, we ensure that our conclusions about primary-day spikes are robust to whether participation is defined in terms of breadth (users) or depth (posts).

Table W3: NB-GLMM estimates of pre- and post-primary post count relative to primary day (IRR, 95% CI, p-values) by year and window.

Year	Window	Pre vs. Primary (IRR / 95% CI)	Post vs. Primary (IRR / 95% CI)
2016	$\pm 3d$	0.36/ [0.30, 0.44]***	0.70/ [0.57, 0.85]***
2016	$\pm 7d$	0.40/ [0.32, 0.49]***	0.46/ [0.37, 0.56]***
2016	$\pm 15d$	0.36/ [0.30, 0.44]***	0.38/ [0.31, 0.45]***
2016	$\pm 30d$	0.37/ [0.31, 0.45]***	0.30/ [0.25, 0.37]***
2016	$\pm 60d$	0.35/ [0.28, 0.42]***	0.25/ [0.21, 0.31]***
2016	$\pm 90d$	0.30/ [0.24, 0.38]***	0.23/ [0.18, 0.28]***
2020	$\pm 3d$	0.59/ [0.46, 0.75]***	0.95/ [0.75, 1.20]
2020	$\pm 7d$	0.56/ [0.44, 0.71]***	0.65/ [0.51, 0.83]***
2020	$\pm 15d$	0.58/ [0.45, 0.74]***	0.50/ [0.39, 0.64]***
2020	$\pm 30d$	0.53/ [0.41, 0.70]***	0.35/ [0.27, 0.45]***
2020	$\pm 60d$	0.40/ [0.29, 0.55]***	0.23/ [0.17, 0.32]***
2020	$\pm 90d$	0.31/ [0.21, 0.45]***	0.17/ [0.12, 0.24]***

Notes: Stars denote significance levels: * $p < .05$, ** $p < .01$, *** $p < .001$. All NB-GLMMs converged.

Similar to the unique-user results, the results of post counts (Table W3) in *r/SandersForPresident* likewise reveal sharp primary-day spikes across both election cycles. Across all specifications, the results consistently show sharp primary-day surges in distinct participation. In 2016, pre-primary IRRs ranged from 0.30 to 0.40 and post-primary IRRs from 0.23 to 0.70, all significant at $p < .001$; this implies primary-day activity was about 2.4–3.3× the pre-primary level and 1.7–4.3× the post-primary level. In 2020, the same pattern emerged but with attenuated magnitudes: pre-primary IRRs ranged from 0.31 to 0.59 and post-primary IRRs from 0.17 to 0.95, with all but the narrow ± 3 -day post-primary window significant at $p < .001$.

Figure W38: Forest plot of incidence rate ratios (IRRs) for post count around state primary days (*r/SandersForPresident*), with 95% confidence intervals across alternative time windows.

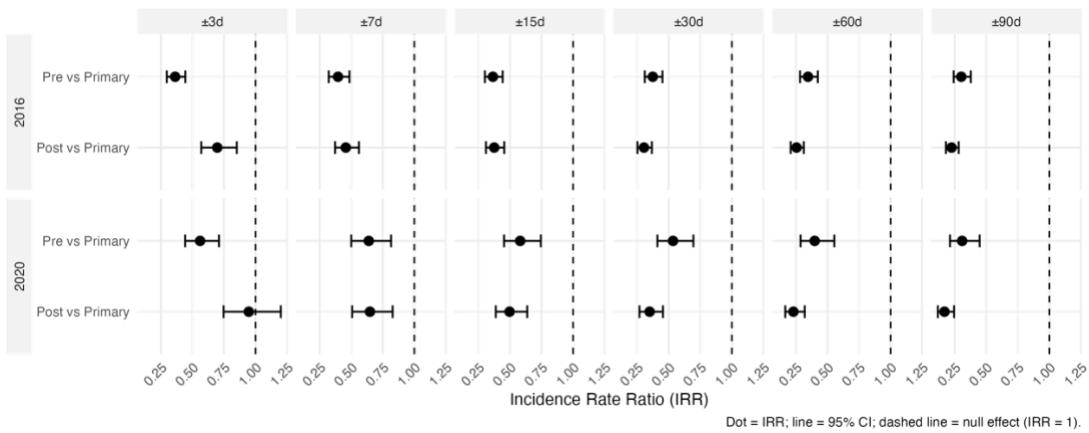
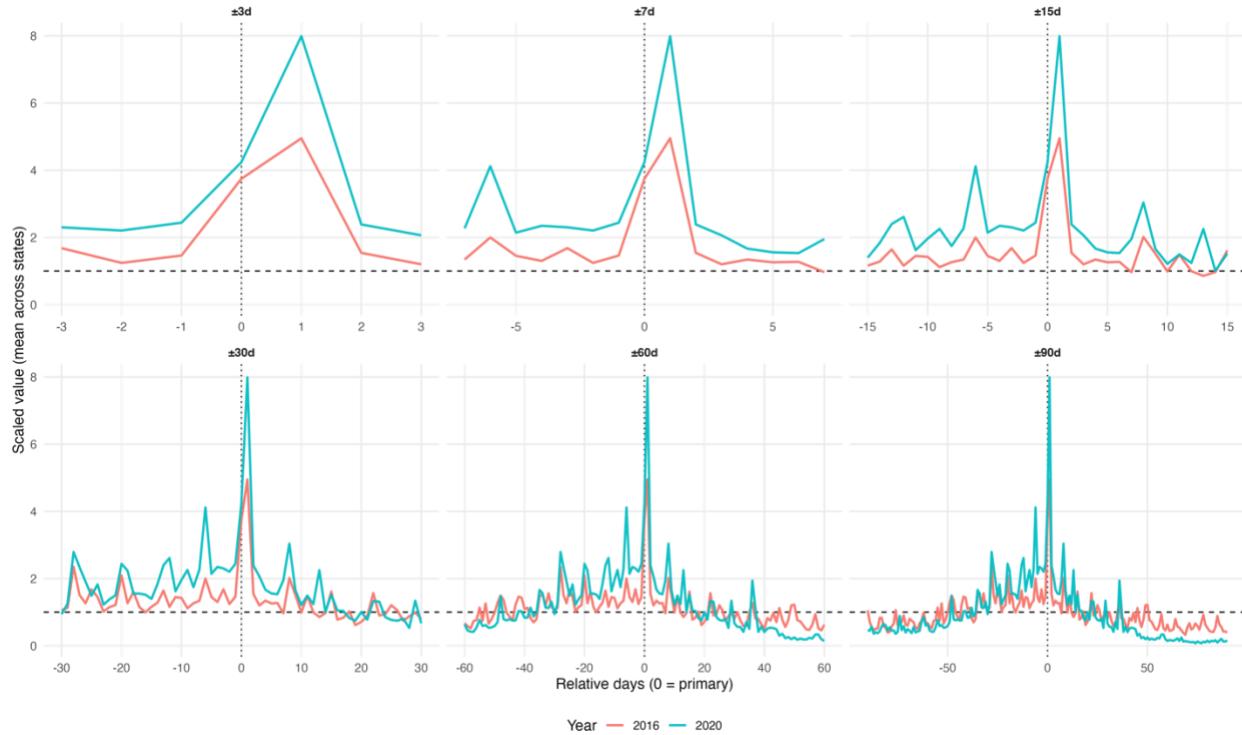


Figure W38 provides a visual summary of these estimates. The forest plots display IRRs and 95% confidence intervals across alternative time windows. In nearly all cases, the confidence intervals lie entirely below 1, reinforcing the conclusion that primary-day participation significantly exceeded surrounding activity. The one exception—the 2020 ± 3 -day post-primary window—shows an interval crossing the reference line at 1, consistent with the non-significant regression result reported above. Figures W38 and W39 corroborate these

regression-based findings, showing consistent primary-day peaks followed by rapid declines across states.

Figure W39: Event-aligned time series of posts in r/SandersForPresident, standardized to a pre-primary baseline, showing primary-day spikes and surrounding declines.



In sum, the dual evidence from unique users and posts shows that primary-day mobilization was both extensive and intensive: Sanders supporters were not only more likely to show up on the subreddit but also more likely to contribute content when they did. Importantly, the replication of this temporal concentration across two electoral cycles and multiple event windows underscores that the observed surges are not artifacts of sampling or model specification. Rather, they highlight the capacity of GeoReddit to recover meaningful patterns of geographically anchored collective action in real time. By demonstrating sensitivity to politically salient events, these findings validate the use of GeoReddit as a lens through which broader questions of spatial and temporal variation in participation can be investigated.