US Traffic Accident Report

PERIOD (2016-2023)

Contents

Table of figures

US Road Incidents — Introduction

This report explores patterns in recorded roadway incidents across the dataset period, focusing on when events happen (by month and hour), the surrounding conditions (day vs night and weather), and how disruptive they are (severity 1–4). The goal is to give operations, planning, and safety teams a shared baseline: where the volume comes from, when resources are most strained, and which conditions are worth targeted interventions.

What you'll find inside

- Incidents by month (seasonality)
- Incidents by hour (daily rhythm)
- Day vs Night split
- Weather at the time of incident (grouped conditions)
- Incident severity distribution (traffic impact 1–4)

Each view presents simple counts to keep the story comparable across pages. Where helpful, we also include long-format series (e.g., incidents by month and severity) for stacked or grouped charts.

Quick takeaways

- Seasonality: Counts dip through summer and rise toward year-end, typically peaking in late fall and December. Winter weather and holiday travel likely add exposure and complexity.
- Daily rhythm: The curve is bimodal—morning and evening commute peaks (around 7–8 a.m. and 4–6 p.m.), with a midday plateau and very low overnight volumes.
- Light conditions: Most incidents occur in daylight, reflecting exposure (more vehicles on the road). Nighttime shares are smaller by count but often warrant separate risk analyses.
- Weather: The majority of incidents occur in ordinary conditions (cloudy/fair), not during severe storms. Rain is a consistent amplifier; snow/hail and thunderstorms are less frequent but operationally significant.
- Severity: Mid-range impact (Severity 2–3) dominates by count. Highest-impact events (Severity 4) are rare but can drive disproportionate delay and response effort.

How to read the charts

- Metric: All primary charts display incident counts (not rates). Use them to understand volume and workload; use rates (per mile, per hour of exposure, per vehicle-mile traveled) for risk benchmarking.
- Time handling:
    - "Incidents by month" can be shown as totals by month or as the average per calendar month across years (12 rows) to smooth year-to-year noise.
    - "Incidents by hour" aggregates clock hours 00–23 over the dataset period.
- Categories:
    - Day vs Night comes from Sunrise_Sunset (values mapped to Day, Night, Unknown).
    - Weather groups are derived from Weather_Condition via keyword matching: Clear; Clouds/Overcast; Rain/Drizzle; Snow/Sleet/Hail; Low visibility (Fog/Mist/Haze/Smoke); Thunderstorm; Other; Unknown.
    - Severity represents traffic impact on a 1–4 scale, where 1 is least impactful and 4 is most impactful.

Using this report

- For operations: Align staffing and patrol focus with the two daily peaks; use weather outlooks to pre-position resources for rain, fog, and the first commute after snowfall.
- For planning: Pair these counts with exposure (traffic volumes, VMT, daylight hours) to assess rates and target interventions where risk is truly elevated.
- For communications: Time messaging for peak windows; emphasize speed management and headway in wet or low-visibility periods.

Caveats and data notes

- Coverage and reporting intensity vary by location and over time; counts reflect detection as well as true incident occurrence.
- Counts are not normalized—high-volume regions and times will dominate. Use rate metrics for cross-region or cross-season comparisons.
- Weather and category mappings are heuristic; small classification differences won't change the overall story but may shift small slices.
- Severity here describes traffic impact, not injury severity.

Color and accessibility

- A consistent palette is used across pages (e.g., warm hues for higher severity, cool hues for weather groups). See "Suggested colors" in the appendix for hex values.
- Labels avoid relying solely on color; legends and value callouts are included. If presenting on dark backgrounds, use the provided dark-theme variants.

Next steps and deeper dives

- Add rate views: incidents per million vehicle miles, per hour of daylight/night.
- Cross-tabs: Severity by weather and light condition; Hourly patterns by season.
- Regional lenses: State/metro comparisons to surface local seasonality and weather effects.
- Outcome metrics: Clearance time distribution by severity and condition.

Data source: `us_accidents` (fields used include Start_Time, End_Time, Severity, Sunrise_Sunset, Weather_Condition, State).
Timeframe: dataset period as loaded in your environment. Apply dashboard filters (date range, geography) to tailor the views to your scope.

Incident Severity Distribution

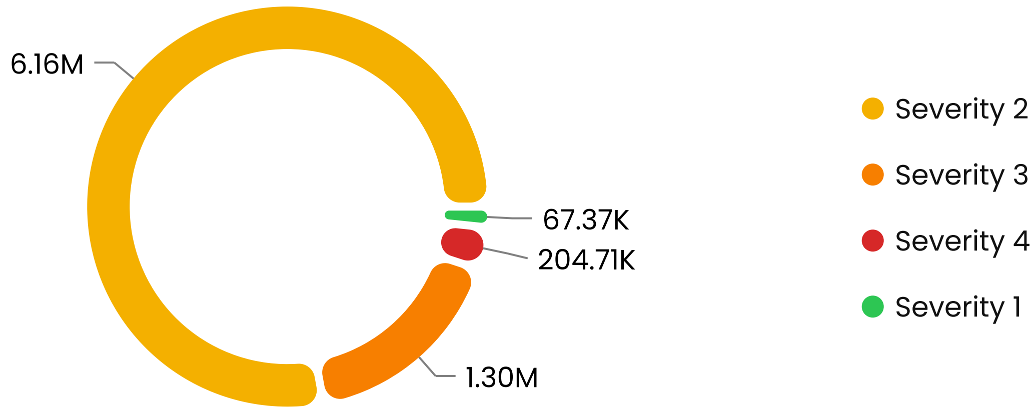


Figure 1 Incident Severity Distribution

Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic

This chart shows the mix of incident severities across the dataset. Most records fall into mid-range impact: Severity 2 dominates at roughly 6.16M incidents, followed by Severity 3 at about 1.30M. High-impact events (Severity 4) are relatively rare at ~205K, and very low-impact events (Severity 1) are the smallest slice at ~67K.
Interpretation: the system captures many incidents that slow traffic but stop short of the most disruptive scenarios. Counts reflect reporting and detection as much as risk—minor fender-benders may be underreported, while major events are more consistently logged. Severity here refers to the impact on traffic (e.g., disruption and duration), not injury severity. For deeper insight, compare severity by time of day, weather, and region, and examine median duration within each severity band.

Incidents by Light Condition

The donut chart shows a clear split between daytime and nighttime incidents. Roughly 5.33M incidents occurred during the day versus 2.37M at night, with only about 23.25K records lacking a light-condition tag. That translates to about 69% of incidents in daylight, 31% at night, and less than 1% unknown.
This pattern is largely driven by exposure: more vehicles are on the road during daytime and commuting windows, which raises the opportunity for low-severity, congestion-related collisions. Nighttime volumes are lower, so counts drop—even though risk per mile can be higher after dark due to reduced visibility and fatigue. The small "Unknown" slice suggests the source data generally records light conditions consistently.

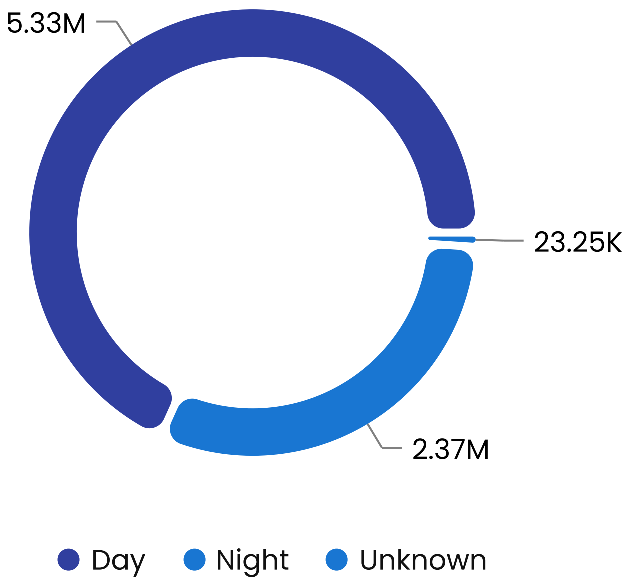


Figure 2 Incidents by Light Condition

Weather at the Time of Incident

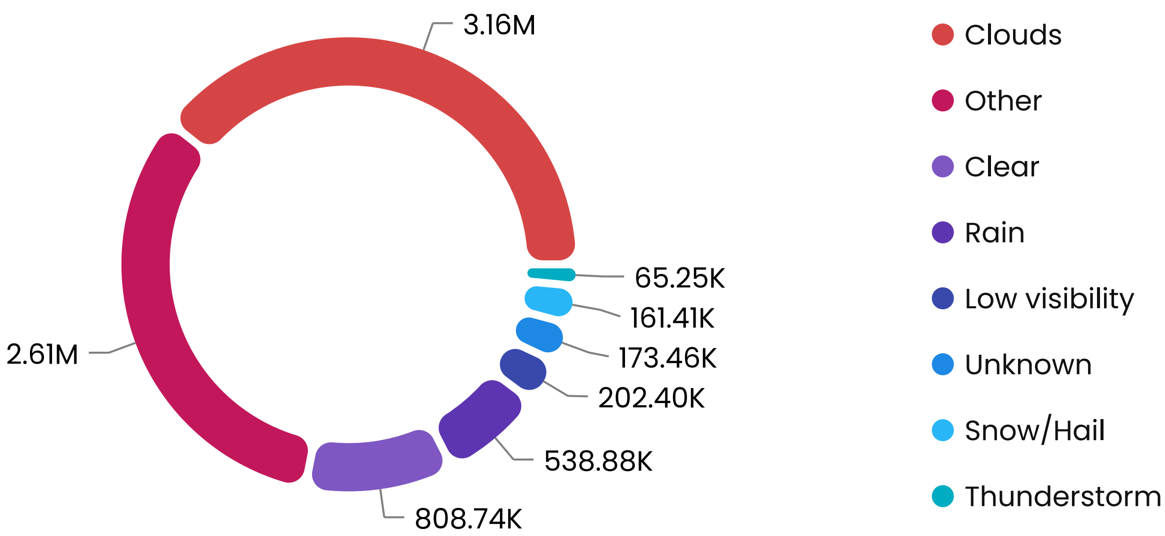


Figure 3 Weather at the Time of Incident

This view shows how often crashes occur under different weather conditions. Most incidents happen in ordinary conditions rather than in heavy precipitation. From the chart, approximately:

- Clouds: ~3.16M (~41%)
- Other: ~2.61M (~34%) — largely "Fair," "Partly Cloudy," and mixed descriptors not matched to the core groups
- Clear: ~0.81M (~11%)
- Rain: ~0.54M (~7%)
- Low visibility (fog/mist/haze/smoke): ~0.20M (~3%)
- Unknown: ~0.17M (~2%)
- Snow/Hail: ~0.16M (~2%)
- Thunderstorm: ~0.07M (<1%)

Total incidents represented are roughly 7.7M across the dataset period.

What this means

- Exposure dominates counts. Drivers spend far more time under cloudy/fair conditions than in active storms, so most incidents occur in these commonplace environments.
- Precipitation matters even at small shares. Rain represents a modest slice of time on the road but still accounts for a meaningful portion of incidents; it's a consistent risk amplifier.
- Severe but infrequent. Snow/hail and thunderstorms contribute relatively few incidents by count, yet they often coincide with higher severity and longer clearance times. They deserve disproportionate attention in preparedness and response planning.

Operational takeaways

- Prevention: Prioritize wet-pavement messaging and speed management during the first 15–30 minutes of rainfall, when friction and driver adaptation are worst.
- Staffing and staging: Pre-position resources for forecast rain bands and fog events; plan surge capacity for the first commute after fresh snow.
- Targeted analysis: Compare median duration and severity by weather group to identify where limited resources have the biggest payoff (e.g., Night × Rain, or Morning peak × Low visibility).

Clearance performance

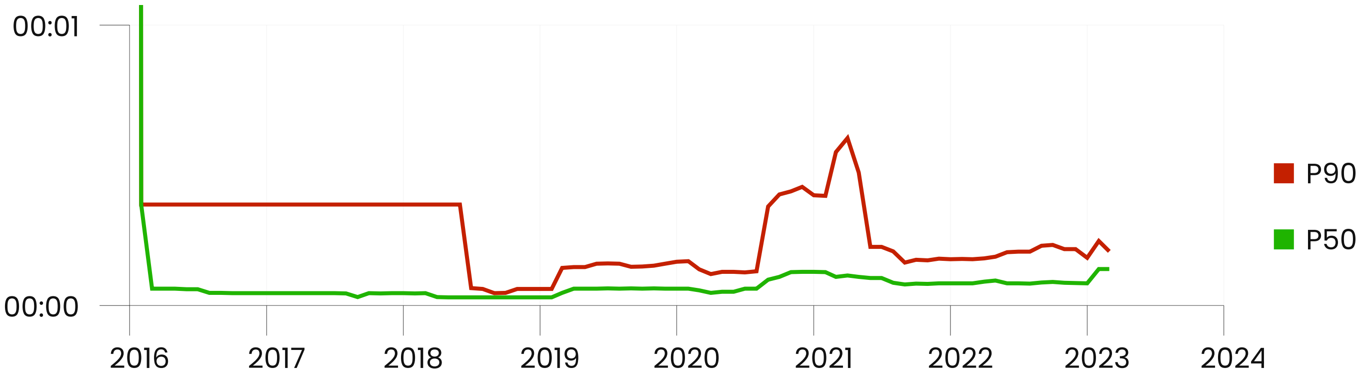


Figure 4 Clearance performance

Clearance performance (p50 and p90)

Most incidents are cleared quickly and consistently over time, as shown by the median (p50, green) staying near the lower bound throughout the period. The 90th percentile (p90, red) tells a different story: there is a long right-tail of slower clearances that varies by year.

Key observations

- 2016–2018: Stable, low median and relatively contained p90 — a tight distribution with few long events.
- 2019–2020: Gradual lift in p90 while the median remains flat, indicating more frequent long-duration incidents without affecting the typical case.
- 2021 spike: A pronounced, temporary surge in p90 suggests clusters of prolonged incidents or a measurement/process change. The median barely moves, reinforcing that this is a tail effect rather than a broad slowdown.
- 2022–2023: Tail risk eases from the 2021 peak but stays somewhat elevated versus pre-2019, with small upticks late in the period.

Interpretation

- Operationally, the "typical" incident clears quickly (stable median), but a subset of cases occasionally takes much longer, driving variability in the tail (p90).
- The widening gap between p50 and p90 in certain periods implies increased dispersion and greater volatility in long-running incidents.

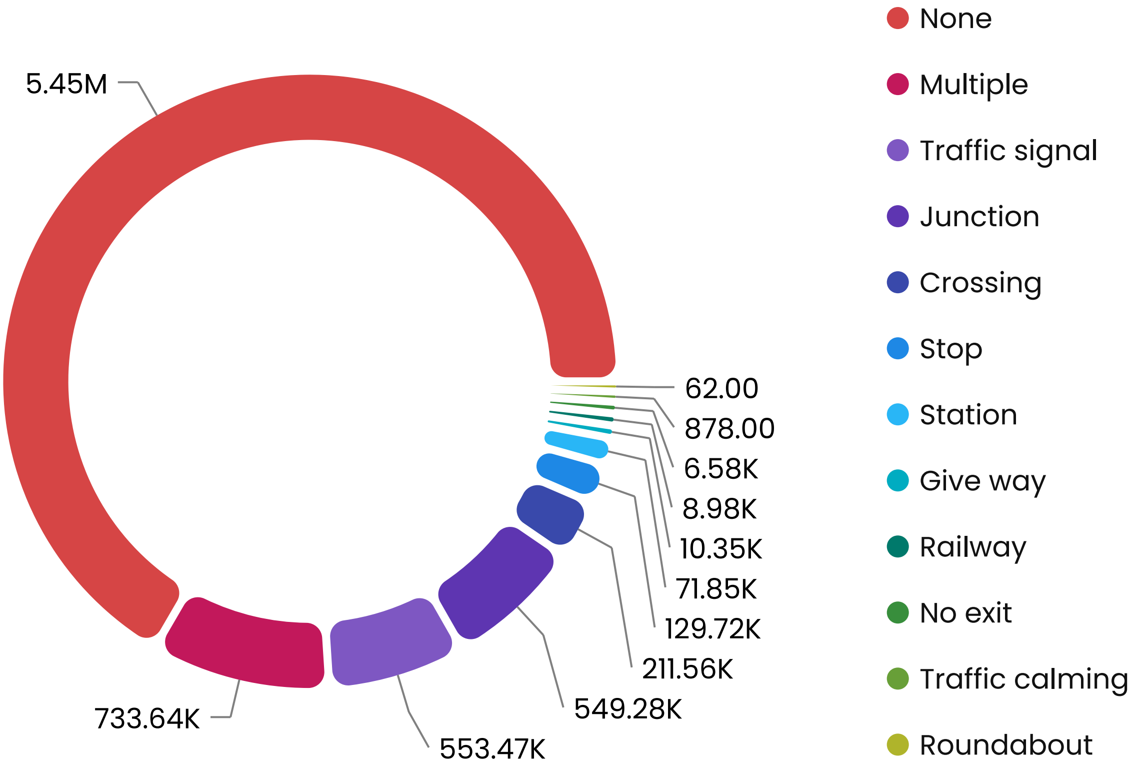
Actions to consider

- Segment tail performance (p90/p95) by state, weather, and severity to isolate drivers of prolonged events.
- Review data quality and workflow changes around the 2021 spike (e.g., End_Time capture, system migrations, timezone handling).

Method note

- Each point reflects monthly percentiles of incident duration in minutes. If you're using the "cleaned" series, extreme durations are capped at a robust threshold (e.g., global p99) or filtered with a maximum plausible duration to limit the impact of data errors.

Incidents by road feature



Most incidents occur away from the listed features: the “None” segment dominates the distribution, indicating that either (a) many events happen on general-purpose roadway without a specific feature nearby, or (b) feature flags are often not present in source data for those locations. Among specific contexts, traffic-controlled environments lead: “Traffic signal” and “Junction” together account for the largest identifiable share, followed by “Crossing” and “Stop.” Less common contexts such as “Station,” “Give way,” and “Railway” contribute small slices, while “Traffic calming,” “Roundabout,” and “No exit” are negligible in this dataset. A notable portion of records fall under “Multiple,” meaning incidents occurred where two or more features co-exist (e.g., signalized junction near a crossing), which is typical in denser urban networks.

Key takeaways

- Typical context: A large majority of incidents do not have a single standout feature flagged (“None”).
- Urban control points: “Traffic signal” and “Junction” form the largest specific contexts, consistent with urban stop-and-go conditions and turning movements.
- Pedestrian interfaces: “Crossing” and “Stop” contribute meaningful but smaller shares, highlighting multi-modal points of interaction.
- Rare contexts: “Railway,” “Give way,” “Traffic calming,” and “Roundabout” appear infrequently in this dataset; treat small segments cautiously due to sample size.
- Complexity: The “Multiple” category is sizable; co-located controls (signals at junctions, crossings near stations) are common and may concentrate risk.

How to use this

- Prioritize countermeasures at signals and junctions (turn-phase timing, red-light enforcement, protected turns, conflict detection).
- For “Crossing” and “Stop,” review sight lines, signage compliance, and lighting; consider pedestrian refuge or advanced stop lines.
- Investigate “Multiple” hotspots separately—co-located features often need holistic redesign rather than one-off fixes.

Method note

- Each incident is assigned to exactly one bucket: “None,” exactly one feature (e.g., “Traffic signal”), or “Multiple” (≥2 features true). This avoids double counting and ensures pie segments sum to 100%. Sparse or missing flags naturally flow into “None,” so the size of that segment reflects both true “no feature” sites and gaps in feature annotation.

State distribution

STATE	TIMEZONE	TIMEZONE SHARE PCT	TIMEZONES COVERED	INCIDENTS	SHARE PCT	RANK OVERALL	CUMULATIVE SHARE PCT
CA	US/Pacific	99.95	3	1,741,433	22.5329	1	22.5329
FL	US/Eastern	96.06	3	880,192	11.3891	2	33.922
TX	US/Central	97.11	3	582,837	7.5415	3	41.4635
SC	US/Eastern	99.08	2	382,557	4.95	4	46.4135
NY	US/Eastern	100	1	347,960	4.5024	5	50.9159
NC	US/Eastern	100	2	338,199	4.3761	6	55.2919
VA	US/Eastern	100	2	303,301	3.9245	7	59.2164
PA	US/Eastern	100	1	296,620	3.8381	8	63.0545
MN	US/Central	99.98	2	192,084	2.4854	9	65.5399
OR	US/Pacific	96.27	3	179,660	2.3247	10	67.8646
AZ	US/Mountain	98.95	3	170,609	2.2076	11	70.0722
GA	US/Eastern	99.93	2	169,234	2.1898	12	72.2619
IL	US/Central	99.95	3	168,958	2.1862	13	74.4481
TN	US/Central	75.96	2	167,388	2.1659	14	76.614
MI	US/Eastern	99.8	3	162,191	2.0986	15	78.7126
LA	US/Central	88.61	3	149,701	1.937	16	80.6497
NJ	US/Eastern	100	2	140,719	1.8208	17	82.4705
MD	US/Eastern	99.81	2	140,417	1.8169	18	84.2874
OH	US/Eastern	100	2	118,115	1.5283	19	85.8157
WA	US/Pacific	99.97	2	108,221	1.4003	20	87.216
AL	US/Central	99.19	3	101,044	1.3074	21	88.5234
UT	US/Mountain	99.48	3	97,079	1.2561	22	89.7796
CO	US/Mountain	99.99	3	90,885	1.176	23	90.9556
OK	US/Central	100	1	83,647	1.0823	24	92.0379
MO	US/Central	99.99	2	77,323	1.0005	25	93.0384
CT	US/Eastern	100	1	71,005	0.9188	26	93.9572
IN	US/Eastern	84.15	2	67,224	0.8698	27	94.827
MA	US/Eastern	100	1	61,996	0.8022	28	95.6292
WI	US/Central	99.98	2	34,688	0.4488	29	96.078
KY	US/Eastern	97.93	2	32,254	0.4173	30	96.4954
NE	US/Central	99.24	2	28,870	0.3736	31	96.8689

STATE	TIMEZONE	TIMEZONE SHARE PCT	TIMEZONES COVERED	INCIDENTS	SHARE PCT	RANK OVERALL	CUMULATIVE SHARE PCT
MT	US/Mountain	98.63	3	28,496	0.3687	32	97.2376
IA	US/Central	100	1	26,307	0.3404	33	97.578
AR	US/Central	99.87	2	22,780	0.2948	34	97.8728
NV	US/Pacific	97.78	3	21,665	0.2803	35	98.1531
KS	US/Central	99.92	2	20,992	0.2716	36	98.4247
DC	US/Eastern	100	1	18,630	0.2411	37	98.6658
RI	US/Eastern	100	1	16,971	0.2196	38	98.8854
MS	US/Central	99.93	2	15,181	0.1964	39	99.0818
DE	US/Eastern	100	1	14,097	0.1824	40	99.2642
WV	US/Eastern	100	1	13,793	0.1785	41	99.4427
ID	US/Mountain	90.91	2	11,376	0.1472	42	99.5899
NM	US/Mountain	99.17	2	10,325	0.1336	43	99.7235
NH	US/Eastern	99.95	2	10,213	0.1321	44	99.8556
WY	US/Mountain	99.95	2	3,757	0.0486	45	99.9042
ND	US/Central	88.84	2	3,487	0.0451	46	99.9494
ME	US/Eastern	99.74	2	2,698	0.0349	47	99.9843
VT	US/Eastern	100	1	926	0.012	48	99.9963
SD	US/Central	84.43	2	289	0.0037	49	100

State severity table

STATE	INCIDENTS	SEVERITY 3 4 SHARE PCT
CA	1,741,433	16.38
FL	880,192	13.32
TX	582,837	21.9
SC	382,557	11.91
NY	347,960	23.22
NC	338,199	11.83
VA	303,301	22.92
PA	296,620	15.81
MN	192,084	16.01
OR	179,660	8.7
AZ	170,609	13.29
GA	169,234	43.74
IL	168,958	36.73
TN	167,388	19.12
MI	162,191	29.67
LA	149,701	13.04
NJ	140,719	19.33
MD	140,417	25.88
OH	118,115	31.36
WA	108,221	31.43
AL	101,044	23.13
UT	97,079	17.15
CO	90,885	37.2
OK	83,647	8.71
MO	77,323	36.58
CT	71,005	28.93
IN	67,224	31.96
MA	61,996	30.92
WI	34,688	37.67
KY	32,254	42.5
NE	28,870	16.38
MT	28,496	2.11

STATE	INCIDENTS	SEVERITY 3 4 SHARE PCT
IA	26,307	34.3
AR	22,780	15.38
NV	21,665	20.36
KS	20,992	29.16
DC	18,630	10.72
RI	16,971	46.24
MS	15,181	29.07
DE	14,097	16.08
WV	13,793	9.77
ID	11,376	6.78
NM	10,325	33.14
NH	10,213	22.65
WY	3,757	19.4
ND	3,487	0.89
ME	2,698	16.12
VT	926	33.15
SD	289	23.53

References

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. “A Countrywide Traffic Accident Dataset.”, 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.