

Analysis of wage and hour case data from US state labor agencies

NOTE: notes in red were added on January 11, 2023 and document how we're handling new information from the Texas Workforce Commission. See notebooks/texas/texas.ipynb for more details.

Data import and preparation

```
(658452, 18)
```

Data completeness

The data states provided in response to my public records requests was wildly inconsistent. I requested a long list of fields -- case statuses, start, end and paid dates, amounts claimed, assessed by the agency and paid -- but not all states possessed that data. Many states exported data directly from their case management systems, so lots of those fields were missing. This section will examine the completeness of various data points and explain how I am handling missing data.

NOTE: the `get_coverage_df` function outputs the completeness (% non-null) of key columns in the analysis. You can find it at `shared/src/shared_functions.py`.

The data overall is very incomplete and has a ton of missing values. This is largely because many of the cases sent from case management systems are not completed.

```
case_status      0.426
amount_claimed   0.521
amount_assessed  0.482
amount_paid      0.382
date_opened      0.985
date_closed      0.231
dtype: float64
```

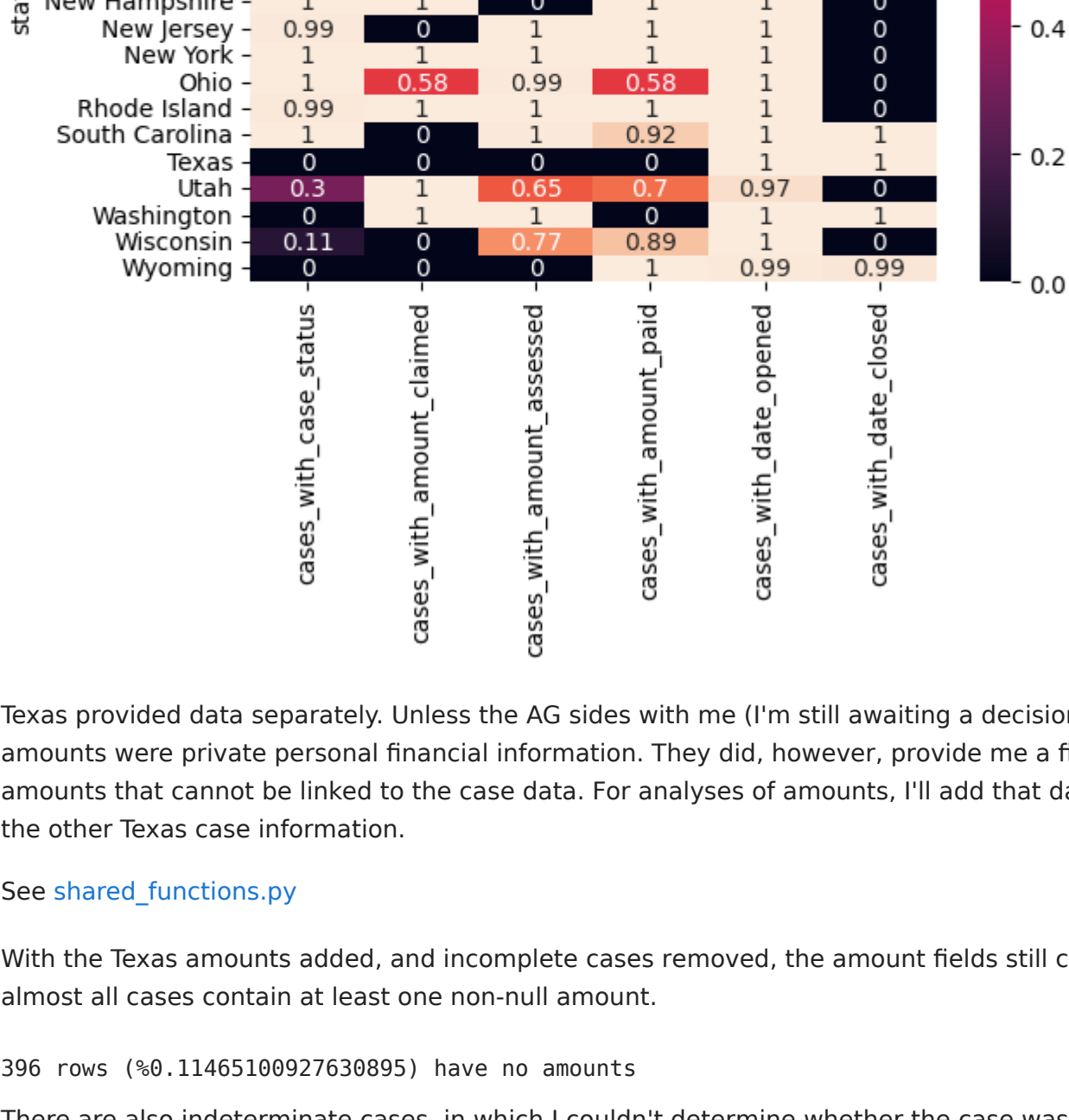
This state breakdown heatmap shows which states have which data.



The `determine_case_outcome` task of this repo assigns a `case_decided_in_favor_of_claimant` column to the dataframe, as well as the reason why that task determined the case outcome.

When you select only cases marked completed, you get much better completeness.

```
case_status      0.442
amount_claimed   0.636
amount_assessed  0.613
amount_paid      0.687
date_opened      0.975
date_closed      0.286
dtype: float64
```



Texas provided data separately. Unless the AG sides with me (I'm still awaiting a decision), they determined that the amounts were private personal financial information. They did, however, provide me a file containing only the amounts that cannot be linked to the case data. For analyses of amounts, I'll add that data separately and remove the other Texas case information.

See [shared_functions.py](#)

With the Texas amounts added, and incomplete cases removed, the amount fields still contain missing data but almost all cases contain at least one non-null amount.

396 rows (≈0.11465180927630895) have no amounts

There are also indeterminate cases, in which I couldn't determine whether the case was completed

| | case_uid | violation_uid | state_name | employer_name | employer_dba_name | employer_city | violation_category |
|--------|--------------------------------------|---------------------------------------|------------|--------------------------------|-------------------|---------------|--------------------|
| 626545 | d6fe9235-f7c9-4fbb-9528-33c8ca3a793 | b501eb08-01fd-465d-8654-4c327212abb9 | Indiana | Mission Complete Transport LLC | NaN | NaN | NaN |
| 626713 | 6a4a2fa3-ac76-4c2c-a15d-495729a8d88d | 497a40ca-7237-4b21-848b-9ddb3f65c0df | Indiana | Diverse Net | NaN | NaN | NaN |
| 626749 | 1e7b6072-d656-4a7b-9953-52b1e7499e1b | 511f4bda-137e-41e3-850c-7629f3a644c4 | Indiana | Code red safety | NaN | NaN | NaN |
| 626774 | 42ab3179-cf6b-4e40-8b6d-42bc665c902 | f3c7eb96-8649-4b73-ab01-7ba9b77727b0 | Indiana | Morningview Assisted Living | NaN | NaN | NaN |
| 626785 | ac4726c1-fa3f-4639-9a1d-ac0b12350c52 | 5778756e-3b11-4e2e-96e7-af832d22b721 | Indiana | Hidden Valley Inn, Inc | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 647666 | 34e4776f-0caf-4bb2-a91d-b7e2f0d4a9a8 | 371e2ce6-571e-44d7-9748-140b4b261838 | Indiana | Creative Technic | NaN | NaN | NaN |
| 647714 | ca1e349f-efee-431a-a913-4687af3f523f | 2fd21a9-2266-4a1c-b7b3-2506fcb16335 | Indiana | Aaa Mattress & Furniture | NaN | NaN | NaN |
| 647742 | cc52bc3c-466b-4355-836f-8b364a354b73 | 8d6bb2f2-2193-4282-a58e-f18879e5cb35 | Indiana | Trugreen Lawn-care | NaN | NaN | NaN |
| 647848 | 7548ead1-c13a-4936-8c9a-4595137add69 | 0380a7c0-cf58-4131-aad0-367857787878a | Indiana | XY Tool & Die | NaN | NaN | NaN |
| 647940 | 9259a97e-ec49-411d-86d4-5b69a305e8f6 | 1c7841ba-c6f1-43c8-9936-ed0a8fa72cf7 | Indiana | Real Alloy | NaN | NaN | NaN |

325 rows x 18 columns

| | case_uid | violation_uid | state_name | employer_name | employer_dba_name | employer_city | violation_category |
|--------|--------------------------------------|--------------------------------------|------------|-------------------|-------------------|---------------|--------------------|
| 307280 | 6f557c25-809d-438c-b468-e34d3501a748 | 40323844-8df7-4b83-b1ce-30895c271ef | Illinois | The Lucky Hot dog | NaN | NaN | NaN |
| 307281 | 864217e1-e07a-4f1f-a6ed-469318eb3e51 | a29ab7cd-ce08-439b-b2c5-cd2a7778c0d4 | Illinois | The Lucky Hot dog | NaN | NaN | NaN |
| 307282 | d4f625c1-c357-4331-e956-cde19961c9f7 | 525ba461-2d07-408b-a72a-03291c03bafc | Illinois | The Lucky Hot dog | NaN | NaN | NaN |
| 307283 | 0fde0606-1108-4f4e-a237-ce2501da787a | 78c31a0d-0e66-4014-b5d8-f1c35138640a | Illinois | The Lucky Hot dog | NaN | NaN | NaN |
| 307284 | cc40d977-616b-4592-b745-f3eaf798ce36 | 320ba338-262a-404e-8f39-11ce9d4f5ad8 | Illinois | The Lucky Hot dog | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 331062 | 0c6b83dc-8160-4813-ae59-ad609f13e457 | 46e9a477-defd-4398-b2c5-11a3912cca9e | Illinois | BLUE CHIP 2000 | NaN | NaN | NaN |
| 331063 | 2ca36515-8dd0-4893-9ac3-ec14d416a1b8 | 888e73ff-e1be-4b0fb1b4-b581-c6fd9629 | Illinois | TRIPLE S SERVICES | NaN | NaN | NaN |
| 331069 | 4929b2df-c2ef-47e5-a56b-a526657e7f98 | 931df107-7aa5-4520-9424-20abb2a549e4 | Illinois | TOTAL EXPRESS | NaN | NaN | NaN |
| 331070 | 437b5ba7-fac0-47bf-a546-06fd5a912491 | a0e610fe-4c18-4e19-89f5-53a4920a8324 | Illinois | EATALY | NaN | NaN | NaN |
| 331071 | e1c17d36-2541-4028-8f4a-bbdbc2e3dfe9 | c94fa6e6-4b50-4eb1-b725-163bc87869d1 | Illinois | CLASSIC SNACK | NaN | NaN | NaN |

11141 rows x 18 columns

They contain very incomplete data, but all for one single reason: they are closed cases with no assessed amount and no paid amount in a state that didn't provide either.

If the state provided either an assessed or paid amount, I can use those to determine the case outcome. But if the state provided only claim amounts or no amounts at all AND no clear case disposition other than "closed," I can't be confident that those cases are all decided in favor of the claimants.

closed case with no assessed amount and no paid amount in a state that provided neither 11141
Name: case_decided_in_favor_of_claimant_reason, dtype: int64
case_status 0.999
amount_claimed 1.000
amount_assessed 0.000
amount_paid 0.000
date_opened 1.000
date_closed 0.000
dtype: float64

Those cases come only from three states -- Illinois, West Virginia and Colorado. All of those provided insufficient data to analyze

Illinois 11141
Name: state_name, dtype: int64

Analysis

Below you will find basic descriptive statistics and state breakdowns. I will expand on this in the coming days, including by making comparisons to average household income and consumer expenses. 1

I will focus on the following data points:

- overall amount assessed and the impacts that can have
- amount claimed or assessed by the state vs. the amount paid by the company
- the amount of time these cases take to resolve

All cases used in this analysis are decided in favor of the claimant unless otherwise specified.

Overall case amounts

```
count      346165.000
mean       4581.853
std        30530.137
min         0.000
25%         200.000
50%        715.000
75%        2400.000
max       5912348.830
Name: overall_case_amount, dtype: float64
```

Total amount across all states: 1,586,077,186

| | total_cases | pct_of_total_cases |
|---------------------|-------------|--------------------|
| overall_case_amount | | |
| 0–100 | 24610 | 0.080 |
| 100–500 | 83237 | 0.270 |
| 500–1,000 | 55690 | 0.181 |
| 1,000–2,500 | 61080 | 0.198 |
| 2,500–5,000 | 33460 | 0.109 |
| 5,000–10,000 | 23038 | 0.075 |
| 10,000–50,000 | 22247 | 0.072 |
| 50,000–100,000 | 2689 | 0.009 |
| 100,000–500,000 | 1961 | 0.006 |
| 500,000–1,000,000 | 99 | 0.000 |
| \$1,000,000+ | 39 | 0.000 |

| | overall_case_amount | count | mean | std | min | 25% | 50% | 75% | max |
|------------|---------------------|-------|------|-----|-----|-----|-----|-----|-----|
| state_name | | | | | | | | | |

| | | | | | | | | | |
|----------------|---------------|-----------|-----------|-----------|--------|---------|----------|----------|-------------|
| California | 558617653.636 | 63442.000 | 8805.171 | 44730.691 | 0.000 | 295.025 | 2070.010 | 6955.500 | 5912348.830 |
| Illinois | 4213957.400 | 769.000 | 5479.789 | 11667.095 | 0.000 | 882.260 | 2300.000 | 5804.840 | 156500.000 |
| Indiana | 33940254.013 | 7832.000 | 5031.187 | 51612.102 | 0.000 | 245.642 | 590.550 | 1524.862 | 2297378.000 |
| Iowa | 853899.000 | 7370.000 | 115.861 | 625.433 | 0.000 | 0.000 | 0.000 | 0.000 | 7625.000 |
| Kansas | 10508991.876 | 4931.000 | 2131.209 | 29528.906 | 0.440 | 250.000 | 560.000 | 1280.500 | 2019015.000 |
| Kentucky | 2383093.630 | 7020.000 | 3395.740 | 16168.527 | 25.380 | 207.765 | 398.645 | 1164.223 | 125810.820 |
| Maine | 1977300.070 | 792.000 | 2496.591 | 8676.580 | 2.750 | 222.040 | 585.750 | 1551.312 | 153993.430 |
| Maryland | 20140568.600 | 8290.000 | 2429.502 | 6628.096 | 0.000 | 384.000 | 934.470 | 2200.245 | 287701.580 |
| Massachusetts | 48474294.560 | 2419.000 | 20038.981 | 75475.087 | 0.000 | 857.320 | 2560.000 | 9826.160 | 1325827.250 |
| Michigan | 61287422.210 | 41752.000 | 1467.892 | 8299.871 | 0.000 | 200.248 | 482.580 | 1137.010 | 1000000.000 |
| Minnesota | 6559247.360 | 744.000 | 8816.193 | 47355.421 | 0.000 | 289.175 | 886.500 | 4312.220 | 1115426.880 |
| Montana | 11437534.690 | 5361.000 | 2133.470 | 5969.388 | 0.230 | 280.000 | 720.000 | 1767.870 | 167750.180 |
| Nebraska | 508919.000 | 563.000 | 903.941 | 2406.379 | 50.000 | 500.000 | 500.000 | 500.000 | 51000.000 |
| New Hampshire | 42787166.000 | 8200.000 | 5217.947 | 26228.588 | 0.000 | 342.000 | 920.000 | 2631.500 | 658292.000 |
| New Jersey | 146034669.760 | 25725.000 | 5676.761 | 25955.542 | 0.200 | 393.750 | 1101.190 | 3515.000 | 1083752.300 |
| New York | 363780038.260 | 52662.000 | 6907.828 | 41563.036 | 0.000 | 0.000 | 600.000 | 2775.630 | 4130507.510 |
| Ohio | 41043963.840 | 9596.000 | 4277.195 | 23145.961 | 0.000 | 0.000 | 390.790 | 1851.245 | 1225315.440 |
| Rhode Island | 9288775.110 | 2292.000 | 4052.694 | 31058.217 | 0.000 | 0.000 | 0.000 | 414.430 | 703623.260 |
| South Carolina | 8358985.600 | 4810.000 | 1737.835 | 21497.200 | 0.000 | 200.000 | 497.815 | 1171.410 | 1424044.000 |
| Texas | 119781779.020 | 63158.000 | 1896.542 | 7704.863 | 0.980 | 300.000 | 710.500 | 1602.105 | 628000.000 |
| Utah | 18597232.020 | 10103.000 | 1840.763 | 3290.260 | 0.000 | 83.190 | 622.100 | 2014.460 | 30411.000 |
| Washington | 13663941.420 | 6365.000 | 2146.731 | 6924.469 | 0.010 | 280.000 | 723.000 | 1716.660 | 23785.980 |
| Wisconsin | 33833405.810 | 11005.000 | 3074.367 | 41164.701 | 0.000 | 204.310 | 540.000 | 1365.000 | 2031142.800 |
| Wyoming | 1085092.920 | 964.000 | 1125.615 | 1629.319 | 3.360 | 252.750 | 578.025 | 1216.500 | 13703.660 |

Amounts recovered

For Texas paid amounts, we have been told by the Texas Workforce Commission that they do not reliably track the amounts paid, but rather use a numeric status code that was included in the original data to designate if payment was made. For calculations of percentages of cases paid, we will use that status code to designate whether a case was paid, and for total amounts recovered we will use the actual paid amount field, even though it is incomplete.

| | total_cases | pct_of_total_cases |
|-------|-------------|--------------------|
| True | 201068 | 0.612 |
| False | 127564 | 0.388 |

Total amount paid: 631,838,076

| | total_cases | cases_with_paid_amounts | total_amount_paid | pct_of_total_cases |
|----------------|-------------|-------------------------|-------------------|--------------------|
| state_name | | | | |
| California | 63442.000 | 26509.000 | 126853456.268 | 0.418 |
| Indiana | 8175.000 | 5541.000 | 3645489.950 | 0.678 |
| Iowa | 7370.000 | 3199.000 | 2617419.000 | 0.434 |
| Kansas | 4931.000 | 4931.000 | 10508991.876 | 1.000 |
| Maine | 792.000 | 690.000 | 1451891.010 | 0.871 |
| Maryland | 8290.000 | 5102.000 | 8346019.710 | 0.615 |
| Massachusetts | 2419.000 | 2408.000 | 48474294.560 | 0.995 |
| Michigan | 41787.000 | 37796.000 | 36834248.610 | 0.904 |
| Minnesota | 747.000 | 606.000 | 5319233.840 | 0.813 |
| Montana | 5361.000 | 4041.000 | 7312386.590 | 0.754 |
| Nebraska | 563.000 | 80.000 | 37324.000 | 0.142 |
| New Hampshire | 8200.000 | 3227.000 | 7670262.000 | 0.394 |
| New Jersey | 25725.000 | 20677.000 | 90118620.150 | 0.804 |
| New York | 52662.000 | 31109.000 | 219924360.450 | 0.591 |
| Ohio | 9651.000 | 967.000 | 1481679.000 | 0.100 |
| Rhode Island | 2292.000 | 1749.000 | 4571119.010 | 0.763 |
| South Carolina | 4818.000 | 2053.000 | 307040.600 | 0.426 |
| Utah | 10103.000 | 7036.000 | 7911044.940 | 0.696 |
| Wisconsin | 11011.000 | 9807.000 | 25047136.630 | 0.891 |
| Texas | 63158.000 | 33540.000 | 23406057.880 | 0.556 |

Case duration

```
count      91217
mean      310 days 14:02:53.448449204
std       412 days 07:59:42.409025324
min         0 days 00:00:00
25%        105 days 00:00:00
50%        195 days 00:00:00
75%        337 days 00:00:00
max       4585 days 00:00:00
Name: case_duration, dtype: object
```

| | total_cases | pct_of_total_cases |
|---------------|-------------|--------------------|
| duration_days | | |
| 0-30 days | 4594 | 0.050 |
| 30-60 days | 5102 | 0.056 |
| 60-90 days | 8393 | 0.092 |
| 90-180 days | 24256 | 0.266 |
| 180-360 days | 28515 | 0.313 |
| 360-720 days | 12934 | 0.142 |
| 720+ days | 7323 | 0.080 |

| | count | mean | std | min | 25% | 50% | 75% | max |
|------------|-------|------|-----|-----|-----|-----|-----|-----|
| state_name | | | | | | | | |

| | | | | | | | | |
|----------------|-----------|---------|---------|--------|---------|---------|---------|----------|
| California | 12366.000 | 523.836 | 306.061 | 26.000 | 300.000 | 439.000 | 676.750 | 1951.000 |
| Kansas | 4917.000 | 158.355 | 306.790 | 0.000 | 47.000 | 99.000 | 168.000 | 3869.000 |
| South Carolina | 4807.000 | 65.342 | 48.877 | 0.000 | 35.000 | 57.000 | 84.000 | 826.000 |
| Texas | 60683.000 | 332.816 | 450.520 | 29.000 | 133.000 | 208.000 | 323.000 | 4585.000 |
| Washington | 6364.000 | 55.863 | 55.484 | 0.000 | 22.000 | 43.000 | 69.000 | 1197.000 |
| Wyoming | 2080.000 | 99.992 | 137.802 | 0.000 | 24.000 | 48.000 | 124.000 | 1114.000 |

Comparison to state felony theft thresholds

| | total_cases | pct_of_total_cases |
|--------------|-------------|--------------------|
| felony_theft | | |
| False | 483145 | 0.743 |
| True | 167307 | 0.257 |

| felony theft | False | True | total | pct |
|----------------|------------|------------|------------|-------|
| California | 150618.000 | 39609.000 | 190227.000 | 0.208 |
| Colorado | 454.000 | NaN | NaN | NaN |
| Illinois | 7470.000 | 16339.000 | 23809.000 | 0.686 |
| Indiana | 12821.000 | 8690.000 | 21511.000 | 0.404 |
| Iowa | 7090.000 | 280.000 | 7370.000 | 0.038 |
| Kansas | 7171.000 | 3080.000 | 10251.000 | 0.300 |
| Kentucky | 3978.000 | 3159.000 | 7137.000 | 0.443 |
| Maine | 520.000 | 272.000 | 792.000 | 0.343 |
| Maryland | 5364.000 | 2926.000 | 8290.000 | 0.353 |
| Massachusetts | 802.000 | 1618.000 | 2420.000 | 0.669 |
| Michigan | 58677.000 | 24347.000 | 83024.000 | 0.293 |
| Minnesota | 773.000 | 354.000 | 1127.000 | 0.314 |
| Montana | 5336.000 | 2722.000 | 8058.000 | 0.338 |
| Nebraska | 530.000 | 49.000 | 579.000 | 0.085 |
| New Hampshire | 4319.000 | 4107.000 | 8426.000 | 0.487 |
| New Jersey | 3606.000 | 22119.000 | 25725.000 | 0.860 |
| New York | 40212.000 | 21648.000 | 61860.000 | 0.350 |
| Ohio | 14468.000 | 3328.000 | 17796.000 | 0.187 |
| Rhode Island | 1955.000 | 341.000 | 2296.000 | 0.149 |
| South Carolina | 13944.000 | 680.000 | 14624.000 | 0.046 |
| Texas | 60683.000 | NaN | NaN | NaN |
| Utah | 15258.000 | 6418.000 | 21676.000 | 0.296 |
| Washington | 34523.000 | 3127.000 | 37650.000 | 0.083 |
| West Virginia | 3778.000 | NaN | NaN | NaN |
| Wisconsin | 26880.000 | 1787.000 | 28667.000 | 0.062 |
| Wyoming | 1915.000 | 307.000 | 2222.000 | 0.138 |
| total | 483145.000 | 167307.000 | 650452.000 | 0.257 |