

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

## Data import and preparation

### Data completeness

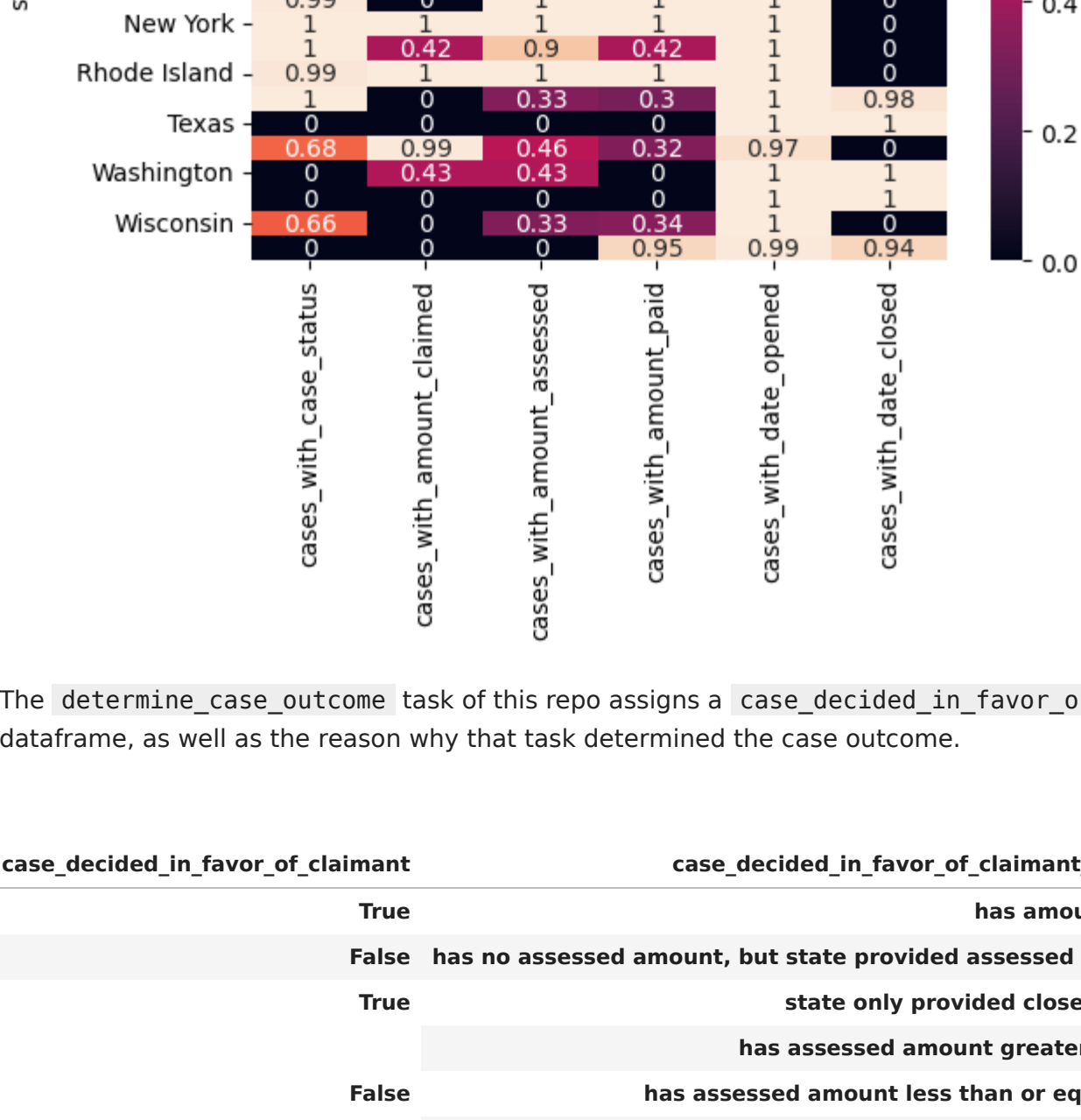
The data states provided in response to my public records requests was wildly inconsistent. I requested a long list of fields -- case statuses, start, end and paid dates, amounts claimed, assessed by the agency and paid -- but not all states possessed that data. Many states exported data directly from 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  0.402
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.

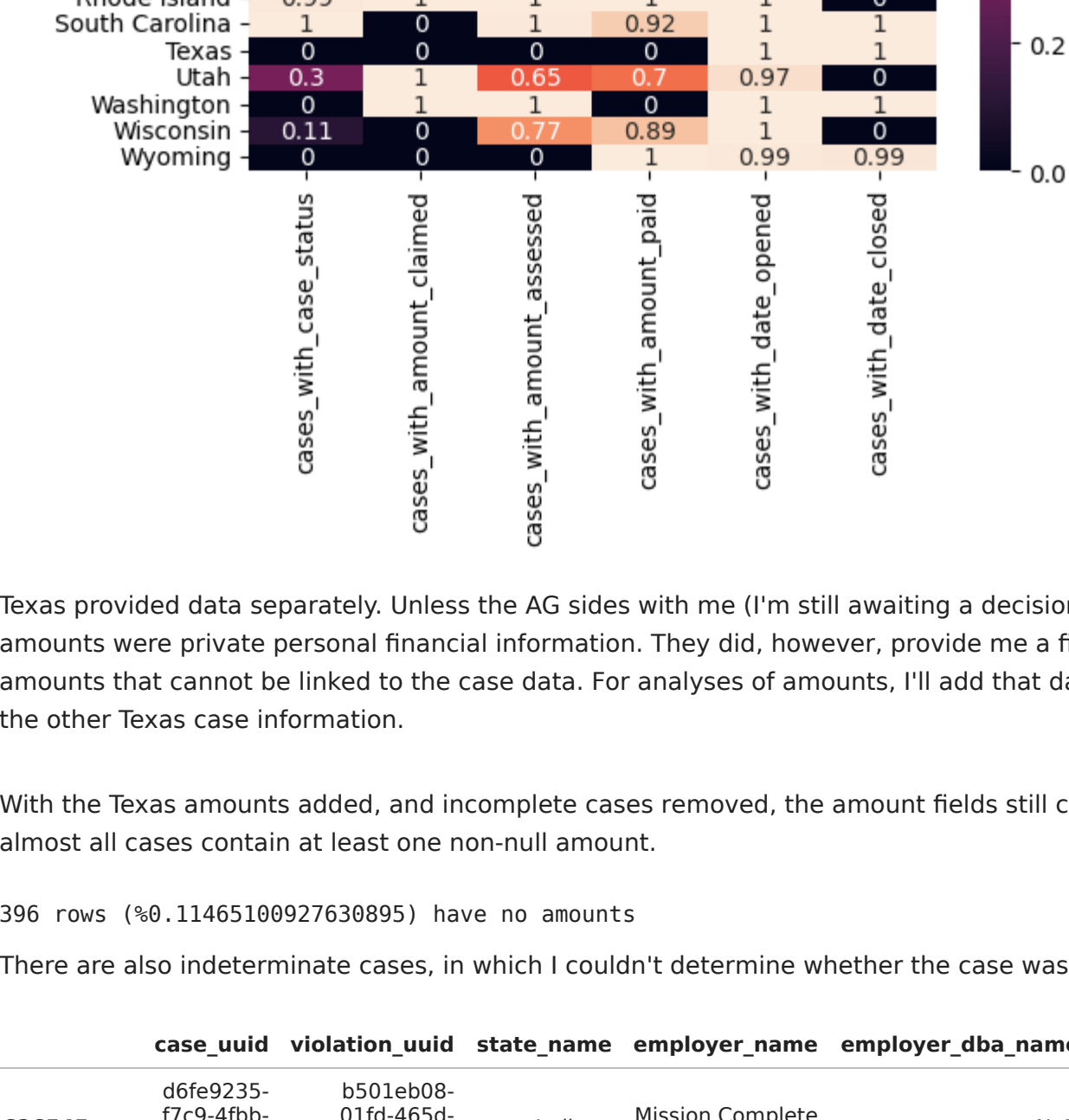
		rows
case_decided_in_favor_of_claimant		
True	has amount paid	237082
False	has no assessed amount, but state provided assessed amount	223131

		rows
True	state only provided closed cases	60683
False	has assessed amount greater than 0	44684
False	has assessed amount less than or equal to 0	37868
True	has open or incomplete case status	28684

		rows
True	has final case status	2947

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.

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 (%0.11465100927630895) have no amounts

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

	case_uuid	violation_uuid	state_name	employer_name	employer_dba_name	employer_city	violation_category
626545	d6fe9235-f7c9-4fbb-9528-33c8c8a33793	b501eb08-01fd-465d-8654-4c327212ab99	Indiana	Mission Complete Transport LLC	NaN	NaN	NaN
626713	6a4a2fa3-ac76-4c2c-a15d-495729a8d88d	497a40ca-7237-4b21-846b-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-42bc665c9f02	f3c7eb96-8649-4873-ab01-7ba9b77727b0	Indiana	Morningview Assisted Living	NaN	NaN	NaN
626785	ac4726c1-fa3b-4639-9a1d-ac0b12350c52	5778756e-3b11-4e2e-96e7-af832d22b721	Indiana	Tidden Valley Inn, Inc	NaN	NaN	NaN
...	...	...	...	...	...	...	...
647666	34e47f6f-0caf-4b2b-a91d-b7e2fd044a9b	371e2ce6-571e-44d7-9748-140b4b261838	Indiana	Creative Technic	NaN	NaN	NaN
647714	ca1e349f-efee-431a-a913-4687af3f523f	2fdc21a9-2266-4a1c-b7b3-2506fcb16335	Indiana	Aaa Mattress & Furniture	NaN	NaN	NaN
647742	cc52bc3c-466b-4355-836f-8b364a354b73	8d6bb2f2-2193-4282-a5ae-a72a-f18879e5c335	Indiana	Trugreen Lawnicare	NaN	NaN	NaN
647848	7548ead1-c13a-4936-8c9a-e4595137add69	0380a7c0-cf58-4131-aad0-36785778f78a	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 16 columns

	case_uuid	violation_uuid	state_name	employer_name	employer_dba_name	employer_city	violation_category
307280	6f557c25-8090-438c-b668-e34d3501a748	40323844-8df7-4b83-b1ce-3089c52f71ef	Illinois	The Lucky Hot dog	NaN	NaN	NaN
307281	864217e1-e07a-4f1f-a6ed-469310eb3e51	a29ab7cd-ced8-439b-81e0-cd2a7778cd04	Illinois	The Lucky Hot dog	NaN	NaN	NaN
307282	d4f635c1-c357-4331-9e56-cde19961c9f7	525ba461-2dd7-408b-a72a-03291c03befc	Illinois	The Lucky Hot dog	NaN	NaN	NaN
307283	0fde0606-1108-44fe-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
...	...	...	...	...	...	...	...
626464	19c728ff-023c-4bde-9a27-f1709c6a2143	2b4c2999-d8b3-41e8-97be-421aebc01bce	West Virginia	KOONTZ CONTRACTING LLC	NaN	NaN	unpaid wage
626465	31235bf7-436f-4b30-9a96-0241afa6c917	9dc29efe-b630-4ee7-b56e-51c1fa86f22e	West Virginia	MLK STAFFING LLC	NaN	NaN	unpaid wage
626466	ee949774-51e5-4e82-8616-f1e63eb0d8eb	48e51629-d9f5-4a9c-92dd-6961434c5206	West Virginia	MAXIM HEALTHCARE SERVICES INC	NaN	NaN	unpaid wage
626467	f77e248b-5931-497a-aa38-3871e5520e7f	71da83ad-4094-4ad1-9ac5-72f8e27ec6fb	West Virginia	WENDYS INTERNATIONAL LLC	NaN	NaN	unpaid wage
626468	8cc50829-8676-449b-ab56-7481d93004d1	1dc99eb1-4b09-402a-9d39-e28280b93c41	West Virginia	SDR FOODS OF WV LLC	NaN	NaN	unpaid wage

15373 rows x 16 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 15373
Name: case_decided_in_favor_of_claimant_reason, dtype: int64
case_status      0.724
amount_claimed   0.725
amount_assessed  0.000
amount_paid      0.000
date_opened      0.970
date_closed      0.275
dtype: float64
```

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

```
Illinois      11141
West Virginia 3778
Colorado      454
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      346117.000
mean       5121.276
std        32016.237
min         0.000
25%        300.000
50%        867.680
75%       2847.700
max       5912348.830
Name: overall_case_amount, dtype: float64
```

overall_case_amount	total_cases	pct_of_total_cases
0–100	26634	0.079
100–500	90611	0.269
500–1,000	60395	0.179
1,000–2,500	66112	0.196
2,500–5,000	37225	0.111
5,000–10,000	25639	0.076
10,000–50,000	24902	0.074
50,000–100,000	3015	0.009
100,000–500,000	2174	0.006
500,000–1,000,000	117	0.000
\$1,000,000+	45	0.000

	count	mean	std	min	25%	50%	75%	max
state_name								
California	63442.000	10692.895	46977.871	0.010	924.820	3401.950	8658.000	5912348.830
Illinois	769.000	5479.789	11667.095	0.000	882.260	2300.000	5804.840	156500.000
Indiana	7832.000	5031.187	51612.102	0.000	245.642	590.550	1524.862	2297378.000
Iowa	7370.000	1174.898	1593.859	0.000	260.000	660.000	1500.000	45416.000
Kansas	4931.000	2131.209	29528.906	0.440	250.000	560.000	1280.500	2019015.000
Kentucky	7020.000	3395.740	16168.527	25.380	207.765	398.645	1164.223	125810.820
Maine	792.000	2496.591	8276.580	2.750	222.040	585.750	1551.312	153993.440
Maryland	8290.000	2429.502	6628.096	0.000	384.000	934.470	2200.000	287701.580
Massachusetts	2408.000	20130.521	75635.165	50.000	874.562	596.640	9919.257	1325287.250
Michigan	41752.000	1468.392	8299.844	0.010	200.632	492.920	1140.000	1000000.000
Minnesota	743.000	8836.480	47384.888	2.000	298.000	900.000	4313.100	1115426.880
Montana	5361.000	2133.470	5969.388	0.230	280.000	720.000	1767.870	167750.180
Nebraska	563.000	903.941	2406.379	50.000	500.000	500.000	500.000	51000.000
New Hampshire	8720.000	5217.947	26258.588	0.000	242.632	920.000	2631.500	658292.000
New Jersey	25725.000	5676.761	25955.542	0.200	393.750	1101.190	3515.000	1083752.300
New York	52662.000	7895.793	45394.639	0.000	238.000	900.000	3502.043	4130507.510
Ohio	9567.000	4321.408	23184.341	0.000	0.000	415.250	1910.330	1225315.440
Rhode Island	2292.000	5943.196	31963.695	2.910	297.803	823.190	2381.997	703623.260
South Carolina	4804.000	1740.089	21510.529	4.000	200.000	498.575	1176.612	1424044.000
Texas	63158.000	1896.542	7704.863	0.980	300.000	710.500	1602.105	628000.000
Utah	10103.000	2038.912	3261.794	0.010	309.985	855.000	2259.655	30411.000
Washington	6365.000	2146.731	6924.469	0.010	280.000	723.000	1716.660	233785.980
Wisconsin	11004.000	3095.354	41166.491	1.000	215.220	554.630	1397.485	2031142.800
Wyoming	964.000	1125.615	1629.319	3.360	252.750	578.025	1216.500	13703.660

## Amounts recovered

	pct_of_claimed_or_assessed_paid	unpaid_amount
count	255766.000	347871.000
mean	inf	1592.339
std	NaN	18142.105
min	0.000	0.000
25%	0.000	0.000
50%	1.000	0.000
75%	1.000	301.500
max	inf	3767316.730

	total_cases	pct_of_total_cases
funds recovered		
False	208808	0.733
True	75905	0.267

553928577.4067786

	sum	count	mean	std	min	25%	50%	75%	max
state_name									
California	71104025.493	63442.000	1120.772	17926.635	0.000	0.000	0.000	0.000	3767316.730
Indiana	951550.794	8175.000	116.398	2594.808	0.000	0.000	0.000	29.830	231039.270
Iowa	596069.000	7370.000	80.878	524.621	0.000	0.000	0.000	0.000	7625.000
Kansas	4596031.530	4931.000	932.069	5669.344	0.000	0.000	0.000	173.725	145000.000
Maine	300382.560	792.000	379.271	5235.251	0.000	0.000	0.000	0.000	143126.310
Maryland	16176739.150	8290.000	1951.356	7979.102	0.000	0.000	296.580	1562.315	408010.100
Michigan	35710253.990	41787.000	854.578	4595.715	0.000	0.000	2.190	362.185	250000.000
Minnesota	988249.230	745.000	1326.509	13495.129	0.000	0.000	0.000	0.000	267256.140
Montana	1510725.360	5361.000	281.799	3299.363	0.000	0.000	0.000	0.000	167673.210
Nebraska	471595.000	563.000	837.647	2421.692	0.000	500.000	500.000	500.000	51000.000
New Hampshire	38291271.000	8200.000	4669.667	25789.510	0.000	0.000	532.500	2100.000	658292.000
New Jersey	55916049.610	25725.000	2173.607	20744.120	0.000	0.000	0.000	0.000	1083752.300
New York	183733408.220	52662.000	3488.918	33241.377	0.000	0.000	0.000	62.377	3257734.000
Ohio	19678929.430	9651.000	2039.056	20448.727	0.000	0.000	0.000	32.695	1225315.440
Rhode Island	6989232.550	2292.000	3049.403	28793.288	0.000	0.000	0.000	0.000	633503.260
South Carolina	7638860.460	4818.000	1585.484	21460.725	0.000	69.812	362.250	1009.523	1424044.000
Texas	96994415.180	63158.000	1535.742	7436.291	0.000	32.000	420.000	1246.378	628000.000
Utah	4509400.470	10103.000	446.343	1543.811	0.000	0.000	0.000	20.000	26205.580
Wisconsin	7771388.380	11013.000	705.784	22237.170	0.000	0.000	0.000	0.000	1987997.000

## 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