

Data Wrangling

This project utilizes data about the American prison system published by the U.S. Bureau of Justice Statistics. The data summarizes the yearly statistics of the prison population and occupancy rate from the 1970s to 2010s.

1. Prisoner population

The U.S. Bureau of Justice Statistics surveys the inmates annually and publishes the *Prisoners In XXXX* series, reporting the demographics of the U.S. prisoner population. This project congregates the data in both federal and states' male and female prison system. Table 1 exemplifies the structure of the data collection process.

Jurisdiction	1978	1979	1980	...	2011	2012	2013	2014	2015	2016
Federal	29803	26371	24363	...	216362	217815	215866	210567	196455	189192
Alabama	5625	5464	6543	...	32270	32431	32381	31771	30810	28883
Alaska	712	760	822	...	5597	5633	5081	5794	5338	4434
Arizona	3456	3749	4372	...	40020	40080	41177	42259	42719	42320
Arkansas	2654	3042	2911	...	16108	14654	17235	17874	17707	17537
California	21325	22632	24569	...	149569	134534	135981	136085	129593	130390
Colorado	2486	2668	2629	...	21978	20462	20371	20646	20041	19981
Connecticut	3189	4061	4308	...	18324	17530	17563	16636	15816	14957
Delaware	1325	1419	1474	...	6739	6914	7004	6955	6654	6585
Florida	21436	19748	20735	...	103055	101930	103028	102870	101424	99974
...

Table 1. Total Population of Prisoners in the U.S.

The data table was then melted to an analysis-ready and clean format with rows representing observation while columns indicates variables (Table 2).

Jurisdiction	Jurisdiction Abbreviation	Year	Population
Alabama	AL	1978	5625
Alabama	AL	1979	5464
Alabama	AL	1980	6543
Alabama	AL	1981	7657
Alabama	AL	1982	9233
Alabama	AL	1983	9856
Alabama	AL	1984	10482
Alabama	AL	1985	11015
Alabama	AL	1986	11710
Alabama	AL	1987	12827
...

Table 2. Total Population of Prisoners in the U.S. (cleaned)

The GeoPandas-compatible geodata of U.S. states (Table 3) were merged with the population data to make possible plotting the choropleth map of the prisoner population (Figure 1)

GEO_ID	STATE	NAME	LSAD	CENSUSAREA	geometry
0400000US04	4	Arizona		113594.084	POLYGON (...)
0400000US05	5	Arkansas		52035.477	POLYGON (...)

Table 3. Geodata of the U.S. states

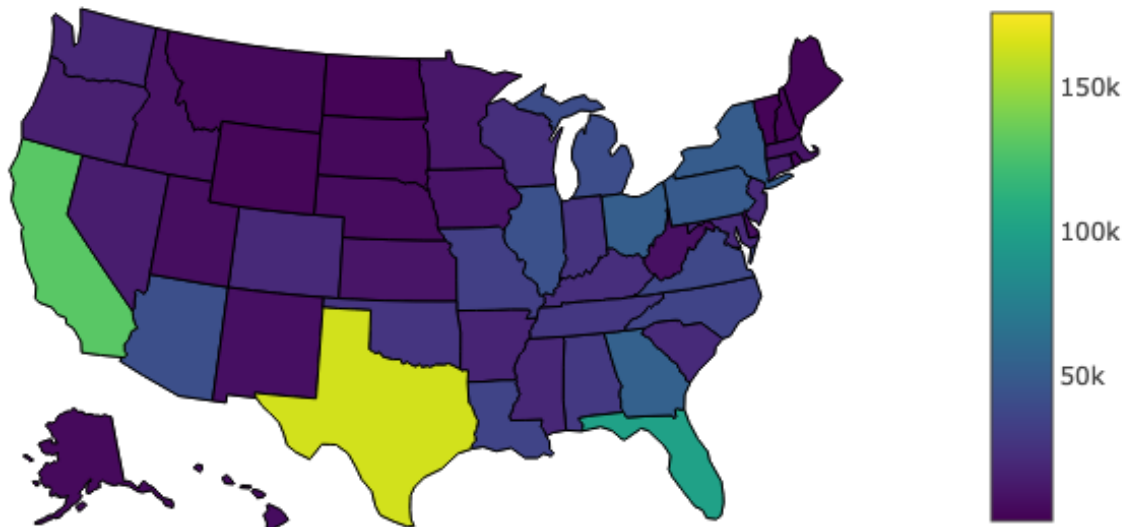


Figure 1. Prisoner population by states in 2016

The population data and geodata were then aggregated into different regions and divisions defined by the U.S. Census Bureau (Figure 2) to generate regional statistics.

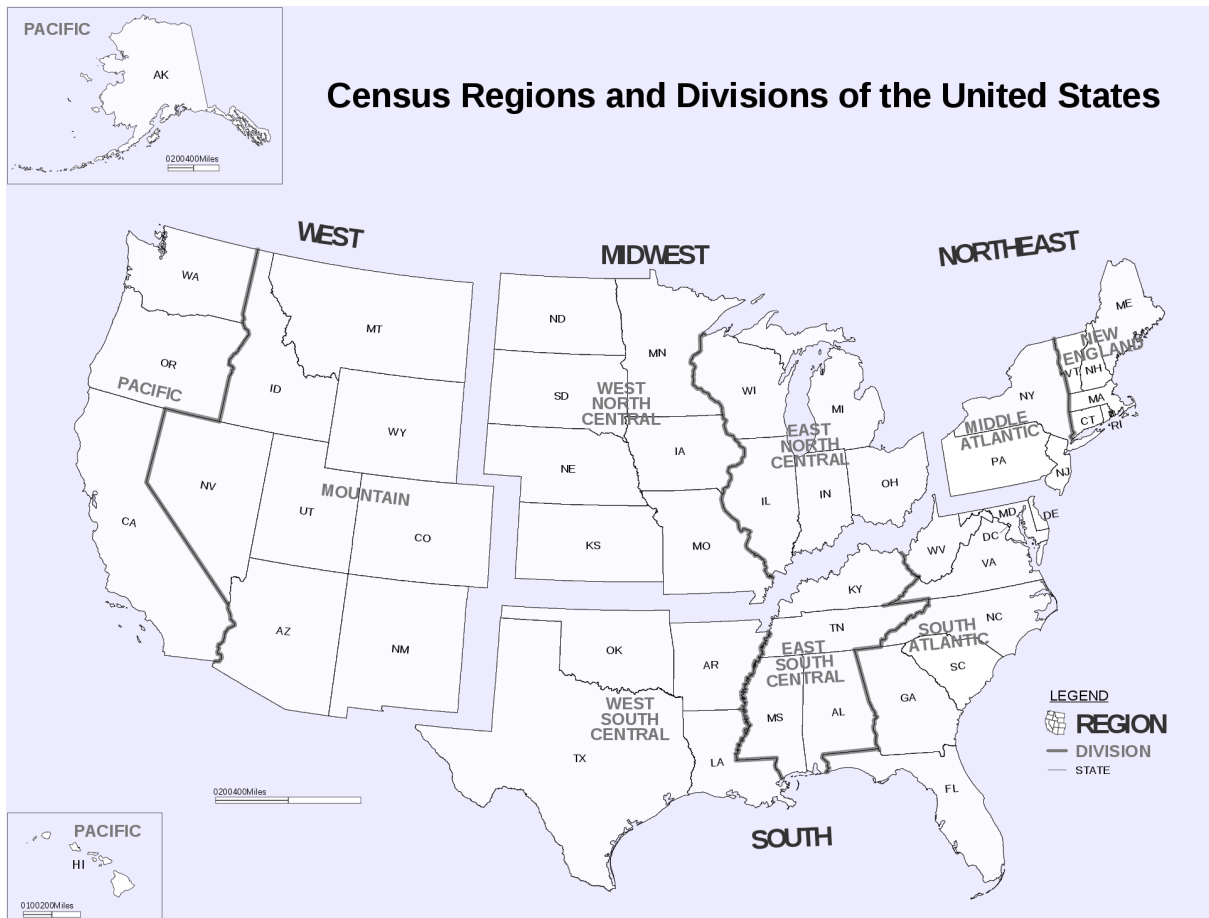


Figure 2. Census regions and divisions of the U.S.

2. Facility occupancy rate

The Bureau of Justice Statistics also provides reports about the correction facility occupancy for the 2010s (Table 4). Occupancy is defined as the ratio of prisoner population to the capacity of the facilities. Some jurisdictions failed to report their data, which could be identified as *not a number* (NaN) in the table. The data table was consequentially reformatted and merged with the geodata in the same fashion as mentioned above. The missing data was represented as grey area in the choropleth maps illustrating the state-wise prison occupancy rate (Figure 3).

Jurisdiction	2011	2012	2013	2015	2016
Federal	138	137.3	133	128	119.7
Alabama	100	99.6	100	98.2	96.3
Alaska	116	142.7	NaN	96.9	98
Arizona	79	80.4	82.4	81.9	81.7
Arkansas	100	97.1	98.7	98.2	97.2
...

Table 4. Prison occupancy in percentage

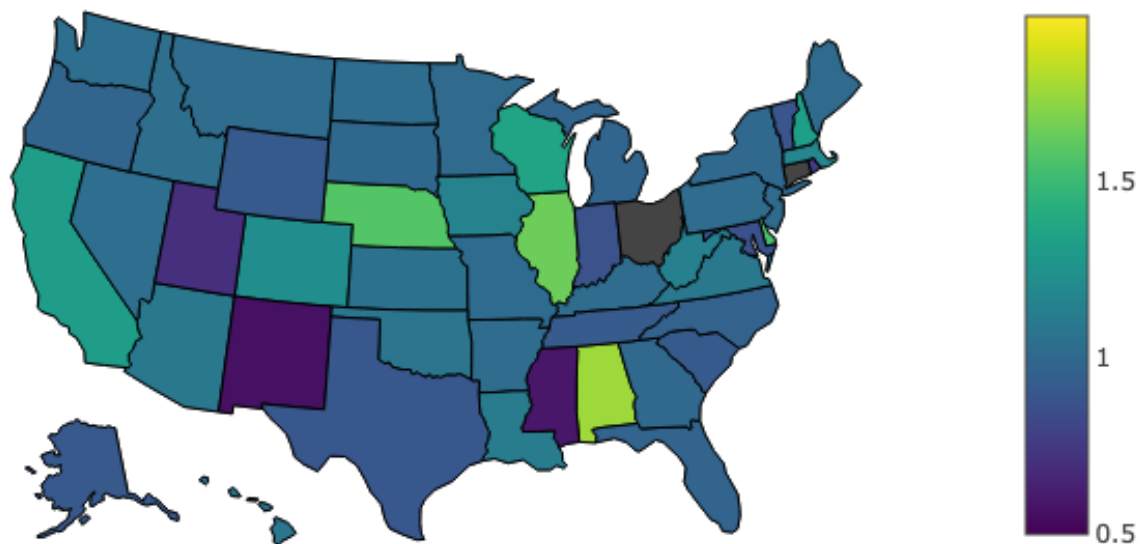


Figure 3. Prison occupancy by states in 2016

Given that the Bureau of Justice Statistics also publishes the prisoner population and facility capacity by states, aggregated regional and divisional occupancy were also computed. Jurisdiction with missing data were omitted in the aggregation calculation.