

Data Nutrition Label

51 Rows (50 states and DC), 12 Columns

<u>COLUMN NAME</u>	<u>TYPE</u>	<u>What is this variable?</u>	<u>MAX</u>	<u>MIN</u>	<u>ISSUES</u>
<u>State</u>	Categorical	The State the data is from	N/A	N/A	N/A
<u>prevalence</u>	Quantitative	Percent of those who said they get low sleep	42.8	29.3	Could be named better
X95..Confidence.Interval	Categorical	Confidence interval for prevalence	41.0-44.6	28.2-30.4	Could be named better
Quartile	Quantitative	Quartile values for prevalence	37.0-42.8	29.3-32.8	N/A
Outcome	Categorical	What the variable estimate is	N/A	N/A	N/A
age_group	Categorical	The age group being sampled	N/A	N/A	N/A
year_pair	Categorical	Year in which data was gathered	N/A	N/A	N/A
<u>estimate</u>	Quantitative	Estimate of those who are abusing substances	0.04008	0.02386	Could be named better
ci_lower	Quantitative	Lower Confidence interval of estimate	0.02822	0.01669	N/A
ci_upper	Quantitative	Upper Confidence interval of estimate	0.05662	0.03280	N/A

estimate_confidence_interval	Categorical	Confidence interval in the same format as X95..Confidence.Interval, which will help for interactive graphs	2.823-5.662	1.67-3.399	N/A
ci_median	Quantitative	Confidence interval median for estimate	4.242	2.535	Not sure whether this value will be helpful or not

Data Usage

- This data will be used to find whether there is a relationship between sleep quality and substance abuse
- Additionally, there are similar datasets at the sources, where for sleep there is information at the county level instead of the state level, and for substance abuse, there is more data on specific types of abuse and for mental illness as well
- Possible bias in the data collection which could lead to faulty analysis, so it is important to keep in mind that any analysis can only be applied to the surveyed population

Populations Present in Data

- Americans aged 12 and older
- No additional information about how the data was gathered, but it was likely that it was gathered at hospitals
- Data gathering could possibly leave out low-income groups or those without access to hospitals where data was being gathered
- This bias could lead to some issues in determining whether there is a relationship between sleep quality and substance abuse

Sources

- <https://www.cdc.gov/sleep/data-and-statistics/Adults.html>
 - Sleep dataset from the CDC

- <https://datatools.samhsa.gov/saes/state>
 - Substance abuse dataset from the Substance Abuse and Mental Health Services Administration