YRBSS Data & Documentation

This data set is presented by CDC based on their survey response, which is conducted once in two years.

Data is available at <https://www.cdc.gov/healthyyouth/data/yrbs/data.htm>

Youth Risk Behavior Survey (YRBS) data are available in two file formats: Access® and ASCII.

The YRBS Combined Datasets include YRBS survey data from all national and the majority of weighted national, state, and large urban school district surveys conducted from 1991 to 2017.

Note:

1. All surveys are not in the combined dataset. State and district YRBS datasets are owned and controlled by the health and education agencies that conducted the surveys. Some of these agencies did not give CDC permission to include their data in the Combined Datasets. Therefore, some analyses conducted using the Combined Datasets may produce results that do not match results previously published by CDC. Refer to the “YRBSS Participation History & Data Quality” table on the “YRBSS Participation Maps & History” page at [www.cdc.gov/yrbss](http://www.cdc.gov/yrbss).
2. The Combined Datasets are based on the 2017 YRBS standard questionnaire and include data from selected optional YRBS questions.
3. Many state and district surveys do not include all of the 2017 standard YRBS questions. Some surveys delete some standard questions from their questionnaires. Data will be missing from some surveys for some variables due to these deletions.
4. The datasets are split into four groups due to large file sizes. The split files make it easier to download just the file(s) of interest. The datasets are split as follows:
   1. National data
   2. Data from states with names beginning with A through M
   3. Data from states with names beginning with N through Z
   4. District data

Why work on State based survey dataset?

* National Survey has 14,765 responses for the year 2017 and total of 203,663 responses in their combined dataset from 1991 to 2017.
* State based survey dataset contains 107,664 responses for the year 2017 and total of 962,925 responses in their combined dataset from 1991 to 2017.
* District based survey dataset contains 36,777 responses for the year 2017 and total of 357, 925 responses in their combined dataset from 1991 to 2017.

Based on above facts we decided to work on the largest dataset i.e. State based survey dataset. Note:

There is not a single state that has provided data in the combined dataset from beginning of the survey year 1991 to 2017 in state based survey dataset.

Data for following states has at least one missing survey from 2009 to 2017 for trend analysis for a decade study: Alabama(AL), California(CA), Colorado(CO), Iowa(IA), Kansas(KS), Louisiana(LA), Mississippi(MS), Missouri(MO), Nebraska(NE), Nevada(NE), New Jersey(NJ), Pennsylvania(PA), South Dakota(SD), Utah(UT), Virginia(VA), Wisconsin(WI) and Wyoming(WY)

Data for following states is available for at least a decade i.e. 2009 to 2017 for trend analysis: Alaska(AK), Arizona(AZB), Arkansas(AR), Delaware(DE), Florida(FL), Hawaii(HI), Idaho(ID), Illinois(IL), Kentucky(KY), Maine(ME), Michigan(MI), Montana(MT), New Hampshire(NH), New York(NY), North Carolina(NC), North Dakota(ND), Oklahoma(OK), Rhode Island(RI), South Carolina(SC), Tennessee(TN) and West Virginia(WV).

Using SAS Files

1. Save the SAS Format Program, SAS Input Program, and the ASCII data files to a folder on your computer.
2. Open the SAS Format Program in SAS and edit it according to the instructions included in the comments in the program.
3. Run the program. This will create a permanent format library in the folder specified in the SAS program.
4. Open the SAS Input Program in SAS and edit it according to the instructions included in the comments in the program.
5. Run the program. This will read the ASCII data file and convert it into a permanent SAS dataset for the particular year in the folder specified in the SAS program.

SAS Format Library:

1. The SAS format library contains the formats used to make SAS output more readable. Formats are linked to the data so that results are displayed as words (“Male” or “Female”, for instance) instead of numbers (1 or 2). The SAS YRBS data file is designed to use its companion format library.
2. The following example SAS program shows how to use the format library. It assumes that both the data file and the format library are in “c:data”. Note that the program contains two libname statements. The first libname statement indicates where the data file is located; the second libname statement indicates where the format library is located.

libname mydata ‘c:data’; /\* tells SAS where the data are \*/  
libname library ‘c:data’; /\* tells SAS where the formats are \*/

proc freq data=mydata.yrbs2005;  
tables q2;

run;

1. Using the format library is recommended but technically is optional. If you do not want to use the format library, include the following statement at the start of your SAS program: options nofmterr; /\* tells SAS to not look for formats \*/
2. Please note that each year of YRBS data has its own format library. Format libraries are not the same across years of data.

**Appendix:**

Programs to convert ASCII format to SAS format and format the data. Note: we did not format the data since we wanted the data to be in easily identifiable nominal values rather than with its descriptions. When need description of the values we can refer to the mapping document attached.

Format Program:



Input Program



Mapping Study of the Dataset:



Dataset Documentation/User Guide:

