

Documentation for SHPPS Vending Machine Datasets

As part of the 2014 School Health Policies and Practices Study, CDC analyzed photographs of a nationally representative sample of school vending machines available to students during the 2013–2014 school year. CDC has created datasets that researchers can use to conduct their own analyses of school vending machines.

There are two vending machine datasets. One dataset provides data at the individual product level. The other dataset provides data at the vending machine level. The latter dataset was used to produce the results included in the *Characteristics of Vending Machines Available to Students in U.S. Schools: Results from the School Health Policies and Practices Study, 2014* report. For both product-level and machine-level datasets, ASCII, Access, SAS, and SPSS data files are available for use.

There are some important understandings and considerations that investigators should be aware of before undergoing analysis using this data. These are indicated below, by dataset. Also noted for each dataset is information regarding how to appropriately account for weights during analysis.

Product-level dataset

- In this study, all product slots (i.e., spaces in the machine in which food or beverage products are placed) were coded, even if the slot was empty.
- Each record in this dataset corresponds to a slot in a vending machine.
- Abstractors were instructed to only use information provided in the photos to code for the variables and NOT to use any other reference material, such as online photos of products. The exception to this was for the product name variable. Abstractors were allowed to research online for full product names.
- Limited editing of text variables (e.g., full product name, text description of marketing/branding) was conducted. The lead abstractor simplified some text (e.g., shortened descriptions of marketing/branding to be more succinct) and corrected observed typos. Such editing was not conducted in a systematic fashion; therefore, additional text errors may remain in the dataset.
- Some corrections to typos for other variables (e.g., a misplaced period in the price of the product) were also corrected when observed.
- Some corrections were made for internal inconsistencies in coding after the fact, but no systematic checks for inconsistencies were performed. Below are some examples of corrections that were made:
 - Simply Chex was categorized different ways – trail mix for some records and chips, crackers, pretzels for some other records. The records coded as trail mix were recoded as chips/crackers/pretzels.
 - For one machine, Vitamin Water was coded as non-diet flavored/enhanced water for 8 records but non-diet soda for another record. The record coded as non-diet soda was recoded as non-diet flavored/enhanced water.
- There is no weight variable that would enable the calculation of national estimates at the product level (i.e., percentage of slots containing milk). We recommend that data be aggregated to the vending machine or school level (note: no school-level dataset was created by CDC).

Machine-level dataset

- Each record in this dataset corresponds to a vending machine in a school.

- Data from the product-level dataset were “flattened” to create summary information for each machine.
- No text variables are included in this dataset.
- The weight variable to use at the machine level is **vm_weight**. If analysts wish to aggregate and report data at the school level, the weight variable to use is **sch_vm_weight**.

Overview of Codebooks

Codebook Introduction

The codebooks document information for each variable in the data files. The columns in the codebook include the following information: column location, variable, variable label, value, formatted value, and count.

Column Location

This column contains the numbers indicating the beginning and ending column position for each variable. The data location is relevant only when using the ASCII formatted data file; for other data files, data should be referenced by variable name, not location.

Variable

This column contains the variable names listed in the order that the variables appear in the data files.

Variable Label

This column contains text that provides a full description of the variable.

Value

This column provides the numerical value of each response option for the variable. For continuous variables, “minimum” and “maximum” will be presented in this column. For non-analytic variables, such as ID variables, this column will be blank.

Formatted Value

This column contains text indicating the meaning of the numerical value, for example “Yes” and “No.” For continuous variables, the minimum and maximum variables will be presented. For non-analytic variables, this column will read “NON-BLANK VALUE.”

Count

This column contains the number of slots (product-level dataset) or vending machines (machine-level dataset) coded for each specific response option.