

## BIOL/STATS 2244 Assignments 2 and 3 Metadata

### Purpose of this file

---

Normally, you—the researcher—would be familiar with how your data were collected and your choice of measurements. That isn't quite the case for our Assignments. So, this file is meant to provide:

- a brief description of where the data (that you will use for Assignment 2 and 3) came from
- a description of the measurements taken
- an explanation of the column names in the data file, assignment.csv
- some quality check characteristics

Essentially, this is an elaborate, text-based metadata file to accompany the data file we are working with for our Assignments.

### Sampling and Study Design

---

A sample of public playgrounds that had at least two material types (i.e. soil, wood chips/mulch, rubber) were selected from the city of Boston, Massachusetts' list of public parks and playgrounds was selected. The neighbourhoods that the playgrounds on the list belong to were separated by socioeconomic status (SES) as follows:

- low SES = average median household income below the Boston 2006-2010 average median household income
- medium SES = average median household income within the Boston 2006-2010 average median household income
- high SES = average median household income above the Boston 2006-2010 average median household income

then 9 playgrounds from each of these neighbourhood SES levels were randomly selected. Each of these selected playgrounds—except for one—had both soil and rubber as material types, so an additional one playground (from the relevant neighbourhood SES level) was selected to ensure at least 9 playgrounds with both rubber and soil surfacing present were selected from each SES level.

At each playground (and/or surrounding play area), two (2) quantities of any non-rubber surfacing material (i.e. soil, sand, and mulch—whatever was present) were collected, and 1 to 3 quantities of rubber surfacing was/were collected.

### *Information on measurements*

Data for several variables were collected for each playground. See the table on page for the list of all variables included. Specific details on the measurements for some of those variables are provided here:

- lead and other potential toxic metal measurements (e.g. avgPb, avgSr, etc.): quantities of lead and other potential contaminants (i.e. Strontium, Cadmium, Arsenic, and Zinc) were measured using a handheld device. The device was calibrated repeatedly using standard reference samples (i.e. with known quantities of the substance being measured) to ensure that the device was recording accurately). Where more than one measurement was made on a particular

quantity of surface material (i.e. for technical replicates), the average value of all quantities measured was reported.

- distance to major roadways: the distance of the playground to nearest major roadways (e.g. interstate highways, etc) was measured using data from the 2013 Massachusetts Department of Transportation Inventory Data. The distance was calculated in GIS software, based on GPS coordinates of the sampling locations and the nearest roadway.
- **environmental justice characterization**: the neighbourhood in which each playground is situated was assessed on several criteria used to evaluate Environmental Justice (EJ) characteristics, to form an EJ index. The criteria were assessed using data from the United States 2010 Census for the neighbourhood. The index used in this study was calculated as the number of criteria met by the neighbourhood; these were the criteria evaluated:
  - whether the neighbourhood had a low median income (i.e. less than or equal to \$40,673)
  - whether the neighbourhood had a high percent of minorities (greater than or equal to 25% minorities)
  - whether the neighbourhood had high non-English speaking population (greater than or equal to 25% of households)

An EJ index is used—in combination with data on environmental and socioeconomic information to identify communities that may be affected by pollution, and to identify where vulnerable groups are located.

## Modification of Dataset for 2244 use

---

The actual dataset study from which the data we are using came included a few more variables that were deemed unnecessary for the purposes of 2244 or, where a suitable description of what the variables represented could not be identified. These were simply deleted from the dataset for 2244.

In addition, some ‘cleaning’ of the dataset was conducted for the purpose of 2244. Variable names were adjusted to make them more user- and R-“friendly”. Any missing values in the original dataframe that were written as ‘NA’ (or otherwise) were removed so the missing values are just empty cells (which is easier to work with in R).

## Variable descriptions

---

The following variables are present in the assignment.csv file:

Variable name	Description
playground	a code that identifies the specific playground from which data were collected, e.g. h15
type	the type of surface related to the observation from a playground. Values are mulch, rubber, soil, or sand. For any particular playground, only one observation exists for each type of surface, although not all surface types are present at each playground.
SES	the socioeconomic status (SES) of the neighbourhood in which the playground is located. Values are high, medium, or, low as described in sampling and study design.

avgPb	mean amount of Lead (Pb) in sampled surface; mean is calculated based on the 1-3 quantities of the surface type for the sampled playground (i.e. for sand, soil, and mulch, the two quantities were averaged; for rubber, the 1 to 3 quantities collected were averaged). Values are in $\mu\text{g/g}$ .
abgAs	mean amount of Arsenic (As) in sampled surface; mean is calculated based on the 1-3 quantities of the surface type for the sampled playground (i.e. for sand, soil, and mulch, the two quantities were averaged; for rubber, the 1 to 3 quantities collected were averaged). Values are in $\mu\text{g/g}$ .
avgZn	mean amount of Zinc (Zn) in sampled surface; mean is calculated based on the 1-3 quantities of the surface type for the sampled playground (i.e. for sand, soil, and mulch, the two quantities were averaged; for rubber, the 1 to 3 quantities collected were averaged). Values are in $\mu\text{g/g}$ .
avgSr	mean amount of Strontium (Sr) in sampled surface; mean is calculated based on the 1-3 quantities of the surface type for the sampled playground (i.e. for sand, soil, and mulch, the two quantities were averaged; for rubber, the 1 to 3 quantities collected were averaged). Values are in $\mu\text{g/g}$ .
avgCd	mean amount of Cadmium (Cd) in sampled surface; mean is calculated based on the 1-3 quantities of the surface type for the sampled playground (i.e. for sand, soil, and mulch, the two quantities were averaged; for rubber, the 1 to 3 quantities collected were averaged). Values are in $\mu\text{g/g}$ .
peeling_paint	Indication of the presence (value of 1) or absence (value of 0) of peeling paint—which may be a source of lead—on playground structures.
rubber_condition	Visual characterization of rubber surfacing deterioration; values are: 0 = none, 1 = low, 2 = high
num_homes	The number of buildings (i.e. homes) directly surrounding the playground
num_homes_paint	The number of buildings (i.e. homes) directly surrounding the playground that had visibly peeling exterior paint (as a potential lead dust contribution to the playground).
distance	The distance of the playground to the nearest major roadway (e.g. interstate highway, state route, major roads), measured in metres.
EJ_criteria	Number of criteria relevant to the environmental justice (EJ) index that are met by the playground's neighbourhood.
notes	Additional observations made about condition of playground and/or surfacing and/or surrounding environment.

## Access to data

The complete data set was accessed from the website for the journal, PLOS One at the following URL:  
<https://doi.org/10.1371/journal.pone.0216156.s001>

The datafile used for Biology/Statistics 2244 Assignments is available on our OWL course site, through the "Assignments" tool, posted for Assignment 2.

## Quality check characteristics

---

The file provided through OWL Assignments contains 15 columns and 86 rows (INCLUDING the row of column names). The values in the column named *avgCd* sum to 669.6453. The column named *peeling\_paint* contains thirty-seven (37) 1s and forty-eight (48) 0s.