This study included 400 participants.

* Participant ID numbers go from 1 to 204 (season 1) and then from 301 to 497 (season 2). **I skipped numbers 205 to 300** to make clear that samples were collected at different times.

In this study I applied 4 different surveys/questionnaires:

* **Screening**: age, neighborhood, health status, household characteristics. etc
* **SES**: instrument developed by a Mexican agency to measure socioeconomic status based on the current characteristics of the population. Each answer has a specific score and summing up all answers gives us a final score that is associated with a specific socioeconomic group (A/B, C+, C, C-, D+, D-, E)

|  |  |
| --- | --- |
| A/B | 205 or more |
| C+ | 166-204 |
| C | 135-165 |
| C- | 112 a 135 |
| D+ | **90-111** |
| D- | **48 a 89** |
| E | **less than 47** |

* **HWISE**: An internationally used 12-item questionnaire to measure water insecurity. Higher scores mean higher water insecurity. Some people calculate subscores from this survey, by selecting a few questions and adding up their scores.
  + Water worry subscores
* Questions used: worry, anger, shame
* Hygiene subsores
  + Questions used: clothes, hands, body
* **PSS:** Perceived stress scores, an internationally used scale to measure perceived stress. Scores are calculated by adding up questions 1,2,3, 8, 11, 12, 14, and subtracting 4,5,6,7,9,10, and 13
* **MEX questionnaire:** a set of questions selected and modified by me to better capture the local reality in Mexico City. All files that begin with “Q” contain answers for questions included in this questionnaire. Most questions were open ended, and I am in the process of transforming some answers into categorical variables.

Merging:

* Some observations were entered twice to test inter-coding error
  + Please flag IDs that present inconsistencies
* Some files have information for only one season: this is why it seems that some questions are in more than one file.

Variable names:

* They have a prefix, root and a suffix.
  + Prefix represent the group or type of question
    - e.g. all Health related questions start with “HLTH\_”
  + Root represents what the question is about
    - E.g. all information related to location or neighborhood have the root “LOC”
  + Suffixes indicate if a variable is categorical (CAT), a score (SC), or if it contains comments or other types of info (INFO). Suffixes are not present in all variables.

Missing information:

* Empty cells represent missing information.
* Some questions have an important portion of the answers missing.
  + In those cases, the questions may have been added later to the survey, or we stopped asking them if they stopped making sense.
* Please code all missing info as N/A (no response).
* DK means “does not know”.
* There might be some answers with “pending” or the letter “P”, which stands for “Pending”. This info needs to be corroborated by looking at the original, hard-copy surveys.
* There are some participant IDs that do not exist, because they were removed or skipped. One of these IDs is ID-164.
* 493 should not be considered, as it does not meet inclusion criteria