A dataset to study transportation, residential context, and well-being in Santiago, Chile

Niloofar Nalaee, Beatriz Mella Lira, Antonio Paez

# Abstract

Large-scale travel surveys are invaluable sources of information to understand travel behavior and other aspects of the urban experience, such as residential context. By their nature, they often shy away from overloading respondents with additional questions. The data presented in this article result from an effort to purposefully collect data on various aspects of the experience of living and moving in a major city in the Global South. The data set contains essential socio-economic and demographic information about the respondents, as well as their built environment and behaviors commuting to work. In addition, the survey (conducted between DATE-DATE, 2016) includes information about the respondents’ feelings and emotions in relation to their commuting experience, the social experience of a variety of transportation modes, various self-assessed health questions, patterns of use of information and telecommunication technologies, and questions about sustainability and the environment. The survey was adopted a quota-sampling method that used the Pre-Census of 2012 as a frame, and in total includes 451 validated questionnaires.

## Key words

* Transportation
* Commuting to work
* Built environment
* Well-being
* Equity
* Santiago
* Chile

## Specifications Table

See [Table 1](#tbl-specifications) for details of the data set.

Table 1: Specifications table

| Items | Explanation |
| --- | --- |
| Subject area | Transportation, Geography, Public Health and Health Policy, Urban development |
| More specific subject area | Transport inequalities, Stress and limited horizons, Travel behavior, Global South |
| Type of data | R Data Package |
| How data was acquired | The survey was conducted using a (pen-and-paper??) questionnaire. The instrument contains some quantitative variables regarding the individual characteristics of respondents and mostly 5-point Likert scale responses in the rest of the questionnaire |
| Data format | Thematic tables and documentation in native R format. The thematic tables can be linked by means of a common ID field |
| Parameters for data collection | The survey was collected using a quota-sampling method based on the information from Pre-Census of 2012, and in total, 451 persons validly completed the survey and face-to-face in Santiago, Chile in 2016. The survey collected information on a wide range of travel-related issues (socio-demographics, health-related, perceptions and travel behavior, travel choices and planning, social interaction factors, built environment, among others) |
| Description of data collection | Data was acquired through the 5-Likert scale questionnaire regarding most sections of the questionnaire, using a face-to-face and quota-sampling method for individual characteristics |
| Data source location | Santiago, Chile |
| Data accessibility | https://paezha.github.io/bSantiago/ |

## Value of the data

A large body of the data has made inroads investigating psychological impact on travelers ranging from positive feelings of enjoyment in some to the sensation of stress in many others that can affect the effectiveness of policy measure (in the case of positive feelings) and are known to affect health outcomes (in the case of stress).This would be interesting for those with transport policies concerns as public feelings like happiness and satisfaction rating of travel mode switching has been studied previously (Abou-Zeid et al. 2012). Data set contribute to psychological impact on travelers both active and motorized modes of transportation to examines not only the feeling of stress, but also how these effects are experienced by travelers and investigates the importance that travelers attach to their feelings of stress which makes it valuable for researchers who focused on public sector development and health-related policies. It allows us to realize the concept of “limited horizons” and focus on developing people’s capability to select what they have their reason behind it. Furthermore, it would be advantageous for decision makers to evaluate their attempts to flourish well-being by expanding capabilities in further research regarding transport inequalities, index of stress, travel behavior in the region or even as a representative for other areas with similar attributes(Walker 2005). The dataset provides a wide range of travel-related issues such as socio-demographics, health-related, perceptions and travel behavior, travel choices and planning, social interaction factors, built environment, among others.

## Data

The data package described in this paper contains 9 data objects in tabular format, which are listed in [Table 2](#tbl-list-of-tables). Each object corresponds to a theme and includes information about the individual characteristics, travel-to-work related information, and perceptions of the built environment. Furthermore, other themes include social interactions by mode of transportation, mode shifting and use of information technologies, a number of self-assessed health responses, feelings and emotions as they relate to the commute experience, attitudes with respect to how well transportation systems satisfy the needs of respondents, and attitudes about nature and sustainability. As seen in [Table 2](#tbl-list-of-tables), the number of variables in these tables ranges from 5 to 24. The tables share a common unique identifier (ID) that can be used to link several tables in a consistent way for multi-theme analysis of the data.

Table 2: List of tables in data package

| Table | Theme |
| --- | --- |
| Santiago\_IC | A table with 22 variables that describe the individual characteristics of respondents |
| Santiago\_TW | A table with 8 variables regarding the commuting behavior of respondents |
| Santiago\_SI | A table with 5 variables about social interaction by mode of transportation |
| Santiago\_IS | A table with 10 variables related to mode-shifting and use of information technologies |
| Santiago\_BE | A table with 23 variables about built environment at the place of residence of respondents |
| Santiago\_H | A table with 13 variables about health attributes of respondents |
| Santiago\_FE | A table with 23 variables with information about the feelings and emotions of respondents |
| Santiago\_RPD | A table with 24 variables with information about attitudes with respect to the effectiveness of the transportation system |
| Santiago\_NS | A table with 11 attitudinal variables that refer to nature and sustainability aspects of city |

The theme of table Santiago\_IC is the individual characteristics of respondents and includes several essential socio-economic and demographic attributes. [Table 3](#X2eb449a934c2f8b9182ef914af54518f65500b3)) presents a summary of these variables, including information about the neighborhood of residence of respondents. All respondents were successfully geolocated to the level of Communes (large administrative partitions equivalent to municipalities) and Macro Zones (larger aggregations of communes used for descriptive purposes). The street address information is less complete. The place of work (geocoded at the level of Communes) is missing only in a small number of cases (complete rate 0.96) and few respondents report a *second* place of work/studies where they travel (r01\_WORK\_COM\_2).

As seen in the table, most respondents’ information is quite complete. Only 16 respondents reported facing some disability, and of these only 11 completed the section asking them about the type (r0O\_TYPE\_DIS). In terms of gender, the sample includes approximately 15.7% more women than men.

**BEA:** PLEASE COMMENT ON THE DISTRIBUTION OF AGES, EDUCATION, OCCUPATION, AND INCOME; ALSO, IS THE VARIABLE MONTHLY OR ANNUAL INCOME? HOW WHERE THE LEVELS/CUTS FOR INCOME SELECTED? ARE THE UNITS IN PESOS OR 1,000 PESOS?

**BEA:** r0P\_MODE\_1 IS PRIMARY MODE? AND r0Q\_MODE\_2 AND r0R\_MODE\_3 ARE SUBSEQUENT MODES FOR THE SAME TRIP? OR MODES USED IN TERMS OF DECREASING FREQUENCY?

Other variables regarding the respondents include their height (cm) and weight (kg), and the household context in terms of number of adults (between 1 and 7) and minors in the family (75% of respondents report no children, and of the remainder the number of children varies between 1 and 6).

These variables provide insights about different segments of the population. In terms of place of residence (**BEA:** ARE ALL COMMUNES URBAN? IS THERE AN URBAN-RURAL INTERFACE?).

[1] "977: 90, 238: 83, Les: 71, 155: 68"

(a) Data summary

|  |  |
| --- | --- |
| Name | select(Santiago\_IC, -ID) |
| Number of rows | 451 |
| Number of columns | 21 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 7 |
| factor | 10 |
| numeric | 4 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| r0A\_COMMUNE | 0 | 1.00 | 4 | 16 | 0 | 43 | 0 |
| r0B\_MACR\_ZONE | 0 | 1.00 | 3 | 12 | 0 | 7 | 0 |
| r0B\_STREET\_1 | 38 | 0.92 | 4 | 24 | 0 | 200 | 0 |
| r0B\_STREET\_2 | 108 | 0.76 | 1 | 23 | 0 | 211 | 0 |
| r0I\_WORK\_COM\_1 | 18 | 0.96 | 4 | 16 | 0 | 30 | 0 |
| r0I\_WORK\_COM\_2 | 371 | 0.18 | 1 | 16 | 0 | 25 | 0 |
| r0O\_TYPE\_DIS | 440 | 0.02 | 5 | 19 | 0 | 10 | 0 |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| r0C\_SEX | 0 | 1.00 | FALSE | 2 | WOM: 242, MAN: 209 |
| r0D\_AGE | 1 | 1.00 | TRUE | 6 | 35 to 54: 135, 18 to 24: 134, 25 to 34: 125, 55 to 64: 44 |
| r0G\_EDUCATION | 1 | 1.00 | TRUE | 5 | COL: 223, SEC: 104, PRO: 84, POS: 29 |
| r0H\_OCCUPATION | 1 | 1.00 | FALSE | 9 | FUL: 257, STU: 101, SEL: 26, HOM: 19 |
| r0J\_INCOME | 6 | 0.99 | TRUE | 6 | 977 to 1,550: 90, 2,380 or more: 83, Less than 423: 71, 1,550 to 2,380: 68 |
| r0N\_DRIVING\_LICENSE | 0 | 1.00 | FALSE | 2 | YES: 260, NO: 191 |
| r0M\_DISABILITY | 0 | 1.00 | FALSE | 2 | NO: 435, YES: 16 |
| r0P\_MODE1 | 0 | 1.00 | FALSE | 8 | MET: 160, CAR: 115, BUS: 110, WAL: 22 |
| r0Q\_MODE2 | 15 | 0.97 | FALSE | 8 | MET: 136, BUS: 131, CAR: 66, WAL: 32 |
| r0R\_MODE3 | 92 | 0.80 | FALSE | 8 | CAR: 74, WAL: 62, BUS: 59, TAX: 51 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 | hist |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r0E\_WEIGHT | 29 | 0.94 | 69.60 | 13.09 | 42 | 60 | 68 | 79.75 | 120 | ▃▇▅▂▁ |
| r0F\_HEIGHT | 20 | 0.96 | 167.21 | 9.46 | 146 | 160 | 167 | 174.00 | 197 | ▃▇▇▃▁ |
| r0K\_ADULTS | 5 | 0.99 | 2.81 | 1.29 | 1 | 2 | 3 | 4.00 | 7 | ▇▃▃▁▁ |
| r0L\_CHILDREN | 6 | 0.99 | 0.66 | 0.92 | 0 | 0 | 0 | 1.00 | 6 | ▇▂▁▁▁ |

Table 3: Variables regarding individual characteristics of respondents

The theme of the next table, Santiago\_TW, deals with commuting and work variables (see [Table 4](#tbl-Travel-work-Descriptives)) and consists of seven ordinal categorical variables (factors). Variable r8A\_ACCESSJOB refers to the impact that respondents feel the transportation network has affected their chances of having better jobs. The most common responses were SOME IMPACT and NO IMPACT, but we see that approximately 14.2% of respondents feel that the network has had a major impact. This variable tracks to some extent with the responses to job opportunities in the commune of residence (r8B\_JOBOPP), suggesting a possible correlation between local opportunities and the impact of the transportation network on job outcomes. When asked about their ideal level of accessibility in the commune of the residence (r8C\_ACC\_COM), a majority respondents opt for excellent and very good (**BEA:** I AM A LITTLE UNCLEAR ABOUT THE MEANING OF THE QUESTION THAT PRODUCES THIS VARIABLE).

In terms of the level of satisfaction with their current job, we see that almost 60% of respondents are at least highly satisfied. We also see that long commutes are frequent in this sample, with about one third or respondents spending 1 h or more travelling (**BEA:** ONE WAY OR TWO WAYS?) and about one quarter of respondents spending between 40 minutes and one hour in their daily commute. This distribution is noteworthy because time spent commuting has been recognized as a factor that can affect physical and mental health and well-being in particular in association with motorized transportation (Brutus, Javadian, and Panaccio 2017).

The most common time of the day for commuting is between 7 am and 9 am, but there are also 171 missing responses in this variable, so not much can be read from it. Finally, we note that many people spend 35,000-75,000 (**BEA:** WHAT ARE THE UNITS OF THIS? IS IT 35 PESOS OR 35,000 PESOS) monthly on their transportation expenditure. (**BEA:** MAYBE COMMENT ON THE TYPICAL RATIO OF TRANSPORT EXPENDITURE TO INCOME).

(a) Data summary

|  |  |
| --- | --- |
| Name | select(Santiago\_TW, -ID) |
| Number of rows | 451 |
| Number of columns | 7 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| factor | 7 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| r8A\_ACCESSJOB | 15 | 0.97 | TRUE | 5 | SOME IMPACT: 124, NO IMPACT : 123, MINOR IMPACT: 66, MAJOR IMPACT: 64 |
| r8B\_JOBOPP | 11 | 0.98 | TRUE | 5 | GOOD: 169, POOR: 86, FAIR: 85, VERY GOOD: 64 |
| r8C\_ACC\_COM | 15 | 0.97 | TRUE | 5 | EXCELLENT: 184, VERY GOOD: 109, GOOD: 93, FAIR: 27 |
| r8D\_EMPLSATISF | 37 | 0.92 | TRUE | 5 | HIGH SATISF: 135, VERY HIGH SATISF: 130, MEDIUM SATISF: 116, LOW SATISF: 20 |
| r8E\_TIMECOMMUT | 10 | 0.98 | TRUE | 4 | 1h and more : 133, 40-60 min: 113, 20-40 min: 103, 0-20 min: 92 |
| r8F\_SCHEDULE | 171 | 0.62 | TRUE | 6 | 7:00 - 9:00: 168, Others: 47, 9:00 - 13:00: 39, 18:00 - 21:00: 14 |
| r8G\_SPENDING | 7 | 0.98 | TRUE | 4 | 35,000-75,000: 199, LESS THAN 35,000: 166, 75,000-125,000: 46, MORE THAN 125,000: 33 |

Table 4: Variables regarding the commuting behavior of respondents

**NOTE FROM ANTONIO:** SEVERAL TABLES NEED MY ATTENTION BEFORE THEY ARE READY TO BE SEEN BY BEA

## Experimental Design, Materials and Methods

The study is based on a paper-based survey conducted face-to-face in Santiago in 2016. The survey collected information on a wide range of travel-related issues (socio-demographics, health-related, perceptions and travel behavior, travel choices and planning, social interaction factors, built environment, among others). The data collection considered a quota-sampling method based on the information from Pre-Census of 2012, and in total, 451 persons validly completed the survey. This paper considers the first part of the survey, with information about the basic socio-economic data, travel choices, activities and commuting information, and the question related to the levels of stress experienced in while traveling.

## Acknowledgements

## References

Abou-Zeid, Maya, Regina Witter, Michel Bierlaire, Vincent Kaufmann, and Moshe Ben-Akiva. 2012. “Happiness and Travel Mode Switching: Findings from a Swiss Public Transportation Experiment.” *Transport Policy* 19 (1): 93–104.

Brutus, Stéphane, Roshan Javadian, and Alexandra Joelle Panaccio. 2017. “Cycling, Car, or Public Transit: A Study of Stress and Mood Upon Arrival at Work.” *International Journal of Workplace Health Management*.

Walker, Melanie. 2005. “Amartya Sen’s Capability Approach and Education.” *Educational Action Research* 13 (1): 103–10.