

A Dataset to Study Transportation, Residential Context, and Well-being in Santiago, Chile

Abstract

Nowadays commuting as a daily travel mostly between work and home is considered as an inevitable part of modern lifestyle. Understanding commuting patterns and travel behavior is important for analyzing stress-related issues, consequences and coping strategies. This dataset is part of a thesis aiming to identify how motorized and non-motorized commuters interact with experiencing stress during their travels. The dataset is based on a paper-based survey conducted face-to-face during November and December 2016 in Santiago, Chile in 2016. A large body of data has made inroads investigating the psychological impact on travellers ranging from positive feelings of enjoyment in some to the sensation of stress in many others that can affect the effectiveness of policy measures and are known to affect health outcomes. The data set contributes to the psychological impact on travellers on both active and motorized modes of transportation to examine not only the feeling of stress but also how these effects are experienced by travellers and investigate the importance that travellers attach to their feelings of stress which makes it valuable for researchers who focused on public sector development and health-related policies. Furthermore, the dataset lies in its potential contributions to transportation policies, public sector development, and health-related policies, as it addresses commuting stress and diverse travel-related issues, encouraging further exploration and fostering reproducibility through transparent methodologies, comprehensive documentation, and open data practices, thereby laying a foundation for future research.

Introduction

Understanding the multi-dimensional nature of urban living in Santiago is crucial for effective planning and policymaking.

As Gainza and Livert (2013) have mentioned that travel has influenced the urban form of Santiago as the city spreads out because of the monocentric nature of that. In terms of urban development and sprawling land uses, Hall, Hall, Zegras, and Rojas (1994) indicated that Santiago has experienced highly deregulated urban growth on the urban periphery at the expense of the conventional business district and more central comunas.

Furthermore, Basso et al. (2020) focused on public transport accessibility and its role in mitigating inequalities across Santiago. Living and traveling in Santiago can be influenced by social exclusion and income level as Ureta (2008) analyzed the mobility patterns of low-income people in everyday life and their situation of social exclusion.

As Graells-Garrido, Peredo, and García (2016) mentioned, urban policy takes several inputs into account. Travel surveys are one of them, because they provide rich information about the city: where people travel to (and from), the purpose of the trip (e.g., commuting to work), the mode (e.g., metro), the time spent traveling, as well as other socio-demographic variables. Large-scale travel surveys are invaluable sources of information to understand travel behaviour and other aspects of the urban experience, such as the residential context. By their nature, they often shy away from overloading respondents with additional questions.

This paper introduces a comprehensive dataset derived from a meticulously designed survey that captures various aspects of the experience of living and moving in a major city in the Global South. The dataset includes self-assessed health questions, respondents' feelings and emotions during commuting experience,

characteristics of the built environment, individual demographics, commuting behaviors, the social experience of a variety of transportation modes, and patterns of use of information and telecommunication technologies. Moreover, it encompasses respondents' attitudes towards the effectiveness of the local transportation system, as well as their perceptions of the city's sustainability and natural environment.

The combination of all these aspects makes this data set a novel, multifaceted and significant as it allows for a holistic analysis of how different modes of transportation influence daily life and individual well-being in an urban setting. By integrating traditional urban study variables with emerging factors like technology use and environmental attitudes, this dataset provides a unique lens through which the urban experience in Santiago can be understood. Furthermore, the inclusion of geographical variables offers critical insights into the spatial patterns of urban living, highlighting how location influences lifestyle and transportation choices in Santiago.

Data Collection, Materials and Methods

The study is based on a paper-based survey conducted face-to-face in Santiago during November and December 2016 by Dr. Beatriz Mella-Lira. The survey collected information on a wide range of travel-related issues. The data collection considered a quota-sampling method based on the information from the Pre-census of 2012 and with a total of 451 persons. More information about data specifications can be seen in Table_1.

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 behaviour, Global South
Type of data	R Data Package
How data was acquired	The survey was conducted using a 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 the Pre-census of 2012, and in total, 451 persons validly completed the face-to-face survey in Santiago, Chile in 2016. The survey collected information on a wide range of travel-related issues (socio-demographics, health-related, perceptions and travel behaviour, travel choices and planning, social interaction factors, built environment, and so on)

Items	Explanation
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/

How does data preprocessing contribute to reproducibility?

As mentioned earlier, the current study is based on a survey. In the data preprocessing step we went through addressing some issues such as splitting multiple answers in a single cell of the survey. Having the answers separated into distinct columns, allows future researchers to gain flexibility in analyzing the data and generating tables tailored to their research questions or goals. The specific analyses and tables created will depend on the nature of the data and the research objectives.

When starting from scratch after splitting and reshaping the data, researchers can use various techniques for a more granular analysis of individual responses and choosing the specific answer among multiple ones in a single cell.

One idea is using frequency tables as they can observe the occurrences of each response for each split variable useful in understanding the distribution of the response.

The second approach way is to take advantage of cross-tabulations for split variables against each other, as it gives insights into relationships and patterns in the data.

Also calculating summary statistics for each split variable allows us to understand central tendency and variability. There are many other ways, such as visualization, comparisons, tests, and data exploration, to specify the appropriate answer out of multiple ones.

When starting from the raw data and following the same preprocessing and analysis steps, researchers could generally end up with similar tables and results. However, several factors can influence whether the results are exactly the same such as differences in data cleaning and handling missing data, parameters or thresholds according to the analysis, and random sampling (if applicable).

In summary, while exact replication is not always guaranteed, careful documentation, consistency in preprocessing, and transparent reporting of analysis steps contribute to the reproducibility of results. Additionally, using the bSantiago GitHub repository and shared codes can facilitate collaboration and make it easier for future researchers to reproduce the results.

Brief overview of the tables in the data Package

The dataset contains multiple tables, each focusing on different aspects of transportation, residential context, and well-being of urban residents in Santiago. Below is a summarized overview of these tables:

Santiago_IC (Individual Characteristics):

This section includes a wide range of socio-economic and demographic attributes such as age, gender, education, occupation, and household context.

Santiago_TW (Commuting Behavior):

Commuting behavior and job-related variables including job accessibility, job satisfaction, commute duration, and transportation expenditure can be found in this section. In terms of the level of satisfaction with their current job, statistics shows 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 of respondents spending 1 h or more travelling 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).

Santiago_SI (Social Interaction):

This section is focused on social interaction variables such as: levels of social interaction and discrimination experienced during commuting, including interaction levels and discrimination by mode of transport.

Social interaction is a topic of interest for mode-related choices given earlier evidence that for some commuters privacy is an important consideration and/or a way to manage social stressors (Páez and Whalen 2010); (Lowe and Mosby 2016); (Gardner and Abraham 2007).

Santiago_IS (Information and Telecommunication, Mode Shifting):

This section will give insights on designing a strong transportation network to facilitate people's different activities such as work, school, grocery shopping and so forth. Also, experts can keep users informed about traffic circumstances, road information and the cost of transportation by using various modes.

Following this, car-use deduction has been turned into the political agenda around the world due to the increasing negative effects of motorized modes of travel on the environment and public health. Therefore, improvements in the public transportation sector are needed to achieve this goal regarding previous studies that have shown participants are more likely to use buses with increased frequencies, shorter travel time, and high access to bus stops (Ettema et al. 2011).

Santiago_BE (Built Environment):

Perceptions and importance of built environment attributes such as space for autos, parking, highways, pedestrian space, and cleanliness of bus stops are included in Santiago_BE section.

The theme of table Santiago_BE is perceptions of the built environment. The way environments are perceived, and not just their objective attributes have been shown to correlate with travel behaviour (Jamal, Mohiuddin, and Paez 2020). Surveying perceptions can be a valuable remedy for the lack of standardized or widely accepted canonical datasets serving as references, especially in regions of the world where attributes of the built environment are not systematically collected.

Santiago_H (Health):

Health-related information including stress, physical effort, proximity to other travelers, environmental pollution, safety, and comfort during commutes have been stored in health section.

Health information could be useful in investigating the effects of transportation policy decisions on public health. Also being aware of specific factors making modes stressful will help transportation and public health experts make commuting a safer, more enjoyable and less stressful activity. Consequently, planners and transportation experts could mitigate the potentially serious health outcomes of a stressful commute (Legrain, Eluru, and El-Geneidy 2015). Having such a detailed data set and understanding the relationship

between health and mode choice in commuting would help them to adopt reform policies and management strategies in accordance with healthy modes of travel as active and public transportation (Mattisson et al. 2018).

Santiago_FE (Feelings and Emotions):

Emotional responses and feelings associated with different modes of transportation, including freedom, safety, luxury, environmental care, and social interaction can be found in feeling and emotions section.

Feelings and emotional information can be used to map and understand travel behaviour and would lead to a more sustainable transportation network. Additionally, this information can make policymakers aware to make targeted choices about where to make physical amelioration as well as on which characteristics to attempt to influence perceptions of alternative modes (Anable and Gatersleben 2005).

Santiago_RPD (Decision-making and Planning): Variables related to decision-making and planning, are namely as: accessibility to employment, social and recreational activities, importance of transportation options, and quality of public transport.

This data set gives us a wide range of variables helpful for identifying people's travel patterns and what level of importance people assign. So it would be useful for transport-related experts to have an insight into people's movement to provide network systems with an appropriate level of performance to increase efficiency and promote urban transportation systems.

Santiago_NS (Nature and Sustainability):

Perspectives on sustainable transportation, willingness to pay for sustainable modes, and the importance of environmental attributes like trees and parks are grouped as nature and sustainability section.

Nature and sustainability information will help planners to provide people with a safe, convenient and environment-friendly transportation system. This will also boost the quality of life, and people's mobility and goods and enhance economic growth through efficient transportation services. Moreover, policymakers would put in place environmental laws and excessive charges for personal car use and ownership (Alyavina, Nikitas, and Njoya 2020).

comuna_1 (Geographical Variables):

Geographical information related to the respondents' location, including commune and macro zone identifiers, providing region-specific insights can be found in comuna section.

Each table is linked by a common unique identifier (ID), allowing for multi-theme analysis of the data. The dataset provides a comprehensive view of how transportation and built environment influence daily life and well-being in Santiago.

Discussion and Conclusion

The current summary of the Santiago dataset showcases a comprehensive commitment to ensuring reproducibility through detailed stages including data collection, cleaning, preprocessing, documentation, storage, and version control. The inclusion of a specifications table in the documentation provides a quick and concise reference, summarizing key details such as the subject area, data format, collection parameters, and accessibility. This not only aids in reproducibility by offering a high-level overview of the dataset's characteristics but also assists researchers in quickly understanding the pertinent details of the dataset.

The data collection process, employing a quota-sampling method, is crucial for the dataset's reproducibility. The detailed explanation of how the sample was obtained allows future researchers to understand the representativeness of the dataset in relation to the population of Santiago in 2016. This level of transparency

makes it feasible to replicate the sampling method in future studies or to compare it with alternative strategies, thus contributing to the reproducibility of the research.

Additionally, the thorough description of the paper-based survey, including its content focusing on areas such as demographics, health, and emotional states, enhances the ability to reproduce the study. Researchers aiming to replicate or extend this work can gain valuable insights into the survey structure, question types, and the scope of topics covered, ensuring that future studies can align with or build upon the original research framework. The data cleaning and preprocessing steps are meticulously documented, including the tools and packages used, such as the `dplyr` package in R. This documentation provides a clear roadmap for future researchers, ensuring that the processed dataset meets the quality standards intended by the original researchers. Moreover, the use of a Quarto Markdown file for data documentation helps in providing a clear and concise reference for understanding the dataset's structure and content, including informative tables and graphs generated with various R packages. Making the dataset publicly available on a dedicated GitHub repository enhances reproducibility by fostering transparency and openness. The use of version control with Git ensures that any changes to the dataset are meticulously tracked over time, allowing for a clear understanding of its evolution and facilitating reproducibility in subsequent analyses.

In summary, the Santiago dataset is not just a collection of data but a well-documented and robustly managed resource that encourages reproducibility through transparent methodologies, comprehensive documentation, and open data practices. This approach not only secures the reliability of the current study but also sets a strong foundation for future research endeavors and collaborations within the scientific community.

Acknowledgements

Profound gratitude is extended to the esteemed supervisor, Dr. Antonio Pérez, for invaluable guidance and support throughout the research endeavor and to Dr. Beatriz Mella-Lira for generous support and provision of survey materials.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research has been supported through a M.Sc. research grant of the McMaster university.

Footnotes

GitHub repository: `bSantiago`

ORCID IDS

Niloofer Nalaee

References

Alayavina, Elena, Alexandros Nikitas, and Eric Tchouamou Njoya. 2020. "Mobility as a Service and Sustainable Travel Behaviour: A Thematic Analysis Study." *Transportation Research Part F: Traffic Psychology and Behaviour* 73: 362–81.

- Anable, Jillian, and Birgitta Gatersleben. 2005. "All Work and No Play? The Role of Instrumental and Affective Factors in Work and Leisure Journeys by Different Travel Modes." *Transportation Research Part A: Policy and Practice* 39 (2-3): 163–81.
- Basso, Franco, Jonathan Frez, Luis Martínez, Raúl Pezoa, and Mauricio Varas. 2020. "Accessibility to Opportunities Based on Public Transport Gps-Monitored Data: The Case of Santiago, Chile." *Travel Behaviour and Society* 21: 140–53.
- 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* 10 (1): 13–24.
- Ettema, Dick, Tommy Gärling, Lars Eriksson, Margareta Friman, Lars E Olsson, and Satoshi Fujii. 2011. "Satisfaction with Travel and Subjective Well-Being: Development and Test of a Measurement Tool." *Transportation Research Part F: Traffic Psychology and Behaviour* 14 (3): 167–75.
- Gainza, Xabier, and Felipe Livert. 2013. "Urban Form and the Environmental Impact of Commuting in a Segregated City, Santiago de Chile." *Environment and Planning B: Planning and Design* 40 (3): 507–22.
- Gardner, Benjamin, and Charles Abraham. 2007. "What Drives Car Use? A Grounded Theory Analysis of Commuters' Reasons for Driving." *Transportation Research Part F: Traffic Psychology and Behaviour* 10 (3): 187–200.
- Graells-Garrido, Eduardo, Oscar Peredo, and José García. 2016. "Sensing Urban Patterns with Antenna Mappings: The Case of Santiago, Chile." *Sensors* 16 (7): 1098.
- Hall, Stephen, Christopher Zegras, and Henry Malbrán Rojas. 1994. "Transportation and Energy in Santiago, Chile." *Transport Policy* 1 (4): 233–43.
- Jamal, Shaila, Hossain Mohiuddin, and Antonio Paez. 2020. "How Do the Perceptions of Neighborhood Conditions Impact Active Transportation? A Study in Rajshahi, Bangladesh." *Transportation Research Part D: Transport and Environment* 87: 102525.
- Legrain, Alexander, Naveen Eluru, and Ahmed M El-Geneidy. 2015. "Am Stressed, Must Travel: The Relationship Between Mode Choice and Commuting Stress." *Transportation Research Part F: Traffic Psychology and Behaviour* 34: 141–51.
- Lowe, Kate, and Kim Mosby. 2016. "The Conceptual Mismatch: A Qualitative Analysis of Transportation Costs and Stressors for Low-Income Adults." *Transport Policy* 49: 1–8.
- Mattisson, Kristoffer, Ahmed Osman Idris, Ellen Cromley, Carita Håkansson, Per-Olof Östergren, and Kristina Jakobsson. 2018. "Modelling the Association Between Health Indicators and Commute Mode Choice: A Cross-Sectional Study in Southern Sweden." *Journal of Transport & Health* 11: 110–21.
- Páez, Antonio, and Kate Whalen. 2010. "Enjoyment of Commute: A Comparison of Different Transportation Modes." *Transportation Research Part A: Policy and Practice* 44 (7): 537–49.
- Ureta, Sebastian. 2008. "To Move or Not to Move? Social Exclusion, Accessibility and Daily Mobility Among the Low-Income Population in Santiago, Chile." *Mobilities* 3 (2): 269–89.