Journal Title
XX(X):1–15

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

ActiveCA: Time Use Data from the General Social Survey of Canada to Study Active Travel

Anon1, Anon2, Anon3

Abstract

This paper describes {ActiveCA}, and open data product with Canadian time use data. {ActiveCA} is an R data package that contains analysis-ready data related to active travel spanning almost 40 years, extracted from cycles 2 (1986), 7 (1992), 12 (1998), 19 (2005), 24 (2010), and 29 (2015) of Canada's General Social Survey. Active travel is characterized by mode, with walking being part of every cycle and bicycling starting in 1992. The attributes of active trips are the types of locations of origins and destinations, the duration of trips, and episode weights for expanding the trips to population-wide estimates. Based on the year of the survey, a variety of locations are coded. In earlier cycles, these include home, work or school, and other's home, whereas in later cycles these are augmented with locations such as grocery stores, restaurants, outdoor destinations, and others. The geographical resolution includes the province and whether the episode was in an urban or rural setting.

Keywords

Active; mobility; walking; cycling; travel time; time-use; General Social Survey

Introduction

The objective of this paper is to introduce {ActiveCA}, an open data product with time use data from Canada's General Social Surveys. Open data products (ODPs) are the outcome of a process that transforms raw data (open or not) into analysis-ready data, following a transparent process in which all stages of development follow open principles (Arribas-Bel et al., 2021). ODPs, while still open, differ from general open data in their

degree of ease of access, their heightened usability, and potentially the value they add to the raw data.

{ActiveCA} provides analysis-ready data concerning active travel in Canada spanning a period of almost 40 years. The source of these data is Canada's program of General Social Surveys (GSSs). This program is designed to provide cross-sectional data on topics of interest to improve the well-being of Canadians. As part of this program, every five to seven years the survey is done on the topic of time use. Concretely, {ActiveCA} covers Cycles 2 (1986), 7 (1992), 12 (1998), 19 (2005), 24 (2010), and 29 (2015) of the GSS. Time use data in these surveys is coded using a very fine grain, from time spent in chores, leisure, and sleeping, to time spent working or at school. These surveys have proved valuable in investigations of mobility and quality of life (Spinney et al., 2009), the relationship between active travel and transit use (Lachapelle and Pinto, 2016), and travel behavior and time poverty (Kim et al., 2024), to name but a few examples.

For {ActiveCA} and using GSS Public Use Microdata Files (PUMFs), we extracted all data needed to characterize active travel in Canada, namely, episodes where the activity was identified as moving between an origin and a destination, either by walking or cycling. Although Statistics Canada offers Public Use Microdata Files and documentation for the GSS program (see Canada, 2024), accessing these files, and preparing them for analysis is not a straightforward matter, given their size and complexity. The process of extracting information of interest from the source files is time-consuming, tedious, and challenging and/or prone to error due to the expertise required to work with these files. To create {ActiveCA} we collected, cleaned, and processed the cycles of the GSS surveys concerning time use to make them ready for analysis.

{ActiveCA} is distributed as an `R` package with a number of data objects and their documentation. `R` packages contain code, data, and documentation in a standardized format that can be installed by `R` users via a software repository, such as CRAN (Comprehensive R Archive Network) or GitHub, which makes them an adroit medium to distribute analysis-ready data.

Given the level of interest in active travel (e.g., McCurdy et al., 2023), reducing the barriers to using data contained in rich, but difficult to access and use surveys, such as Canada's GSSs, is a worthy endeavour that can only improve data-driven decisions in transportation, urban, and health policy.

The rest of this paper discusses the sources of data, and the process implemented to retrieve and package them. Then, we explain how to use the package and show some selected examples of analysis to whet the imagination of potential users. This ODP provides not only data that are easy to use, but also all the code and documentation that make this a reproducible research project. In summary, {ActiveCA} aims to implement and inspire the best principles of open spatial sciences (Páez, 2021; Brunsdon and Comber, 2021).

General Social Survey (GSS) collection

Statistics Canada (2024) conducts GSS surveys to obtain data on social trends to track changes in Canadians' living conditions and well-being over time. The series of survey

on time use are used to understand how Canadian residents spend and manage their time, and what factors contribute to their happiness and stress. The GSS program was created in 1985, and is serialized to provide a collection of annual, representative cross-sectional surveys.

The topics of the survey cycle every few years to cover topics that include family, health, social identity, and every five to seven years time use. The first Canadian time use survey done as part of the GSS program was conducted in 1986, and the most recent was completed in 2015. These Time Use Surveys (Canada, 2022) collect data on respondents' participation and time spent on a wide range of everyday activities using a 24-hour retrospective diary, with information on the location of these activities (e.g. at home, at work, etc.) and, for non-personal activities, the people who were present with the respondent at the time of the activity. In addition, time-use surveys also cover topics related to leisure time, work-life balance, health, commuting, culture and sports, and many others.

The Public Use Microdata are released by Statistics Canada in two files: a *Main File* and an *Episode File*. The files are linked by keys that identify households, individuals, and episodes (i.e., activities) conducted by individuals. We discuss these files in more detail in the following section.

The Main File

The main file of the time use survey compiles a large array of aggregated data, summarizing the answers to the questionnaire that describe households and individuals, as well as derived variables that summarize the respondents' use of time use across different activities, locations, and social interactions. This file documents the time and duration that respondents allocate to each activity and location. The Main File provides a overview of daily routines and social dynamics, not focusing on individual activity episodes. Additionally, this file categorizes activities into bigger groups and subcategories, facilitating the data's analytical utility with additional metrics such as total transit time, time spent with household members, and counts of activities and episodes.

Table 1 shows the first ten rows and first six variables of the GSS PUMF 2015 main file (Cycle 29). Each row in the table correspond to a survey respondent, while the columns refer the following information: record identification (`PUMFID`), the person's weight (`WGHT_PER`), the month the survey data was collected (`SURVMNTH`), the respondent's age group (`AGEGR10`), the respondent's sex (`SEX`), and the respondent's marital status (`MARSTAT`).

The main file of the 2015 GSS surveys includes a total of 17,390 respondents, representing 29,766,399 individuals and 848 variables. For discrete variables, Statistics Canada has assigned specific codes to the possible values, with each code accompanied by a label. For instance, in the case of the variable `SURVMNTH`, a value of `1` means `January 2016`, 2 means `February 2016`, 3 corresponds to `March 2016`, and so on.

As shown in Table 1, the variables are not labeled. Adittionaly, the format of the tables (comma-separated values) does not allow for the specification of variable types (whether

a variable is continuous or discrete), which can lead to mistakes analysts who have limited experience working with PUMFs.

Table 1. Visualization of the first ten lines and first six columns of the main file of the 2015 GSS.

PUMFID	WGHT_PER	SURVMNTH	AGEGR10	SEX	MARSTAT
10000	616.6740	7	5	1	5
10001	8516.6140	7	5	1	1
10002	371.7520	1	4	2	1
10003	1019.3135	3	6	2	5
10004	1916.0708	9	2	1	6
10005	1952.2015	4	1	1	6
10006	5761.5528	8	1	1	6
10007	466.0426	6	5	2	3
10008	2479.2991	2	2	2	1
10009	1436.1641	8	6	1	3

Note:
Legend: 'PUMFID': record identification. 'WGHT_PER': person weight. 'SURVMNTH': survey month of data collection. 'AGEGR10': age group of the respondent. 'SEX': sex of the respondent. 'MARSTAT': marital status of the respondent.

The Episode File

The episode is a much bigger file that records detailed data for each activity episode reported by respondents. Each episode represents a single activity and its duration, and the sum of all episodes throughout the day adds up to 24 hours. Each entry in this file includes the start and end times of the activity, the duration, location, and accompanying social context, informing when and where activities occurred and with whom. The focus of the Episode File is not on the characteristics of the respondents, but on the characteristics of the activities, and the data are structured around the numerous activity instances that compose a day of the respondent. Although respondent-specific characteristics are not included within the episode file, it is possible to link the main file and the episode file by using a key present in both the Main and Episode Files.

Similar to Table 1, which displayed an example of the main file structure, Table 2 presents the first seven episodes for the record identification number `10041` and some variables from the GSS PUMF 2015 episode file (Cycle 29). Each row in the table corresponds to an episode associated with the specified record identification (`PUMFID = 10041`), including the episode's weight (`WGHT_EPI`), episode number (`EPINO`), activity code (`TUI_01`), episode duration (`DURATION`), and episode location (`LOCATION`).

In total, the episode file of the 2015 GSS surveys contains 274,108 records, representing 461,837,622 episodes and 527 variables detailing the episodes. Similar to the main file, Statistics Canada has created codes for the discrete variables, with each value corresponding to a label.

In the case illustrated in Table 2, this respondent began the diary description by sleeping at home (`TUI_01 = 1` and `LOCATION = 300`) for 210 minutes, followed by

40 minutes of personal hygiene (`TUI_01 = 2`). The respondent then spent 15 minutes on personal care activities, such as getting ready for school, supervising homework, reading, playing, reprimanding, or providing educational or emotional support, as indicated by `TUI_01 = 27`. Next, they recorded a travel episode, walking for 15 minutes (`TUI_01 = 7` and `LOCATION = 315`), where both the origin and destination were their home. Such trips, where the journey starts and finishes at home, can be classified as recreational or leisure trips. Next, the respondent spent 3 hours searching for a job (`TUI_01 = 9`), took a 15-minute lunch break (`TUI_01 = 6`), and then cleaned the house (`TUI_01 = 18`) for two hours. Table 2 displays only six variables out of the 527 available. As shown, since the dataset does not label the variable values, decoding them can be both time-consuming and challenging.

Table 2. Visualization of the seven first episodes of the record number '10041'.

PUMFID	EPINO	WGHT_EPI	TUI_01	DURATION	LOCATION
10041	1	1353.818	1	210	300
10041	2	1353.818	2	40	300
10041	3	1353.818	27	15	300
10041	4	1353.818	7	15	315
10041	5	1353.818	9	180	300
10041	6	1353.818	6	15	300
10041	7	1353.818	18	120	300

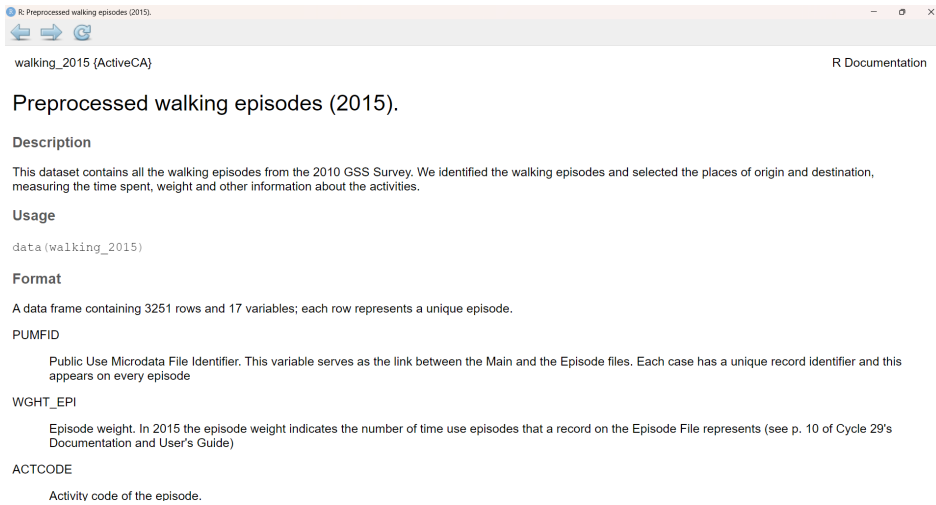
Note:
Legend: 'PUMFID': record identification. 'EPINO': episode number. WGHT_EPI: episode's weight. TUI_01: activity code. DURATION: episode's duration. LOCATION: episode's location.

Data extraction

For each selected cycle of the GSS surveys, we reviewed the episode files to identify episodes of movement that involved walking or cycling. This allowed us to select the activities immediately before and after the movement episode. After that, we labeled the code variables with their appropriate descriptions, identifying each origin and destination, mode of travel, time spent in the active trip, as well the province and urban classification of the episode.

The active trips (walking and cycling) were identified by their corresponding activity codes, accounting for differences between the survey cycles. Identifying the preceding and succeeding activities of each activity episode helps to understand the motivation for the trip and also determine the origin and destination of the trips.

For each GSS main file, we selected socioeconomic variables to help profile the individuals engaged in active trip episodes. We included key socioeconomic variables such as the respondent's age group, sex, marital status, and number of children, among others. After this, as we did with the episode files, we labeled the code variables with their appropriate descriptions.


R: Preprocessed walking episodes (2015).

walking_2015 {ActiveCA} R Documentation

Preprocessed walking episodes (2015).

Description

This dataset contains all the walking episodes from the 2010 GSS Survey. We identified the walking episodes and selected the places of origin and destination, measuring the time spent, weight and other information about the activities.

Usage

```
data(walking_2015)
```

Format

A data frame containing 3251 rows and 17 variables; each row represents a unique episode.

PUMFID

Public Use Microdata File Identifier. This variable serves as the link between the Main and the Episode files. Each case has a unique record identifier and this appears on every episode

WGHT_EPI

Episode weight. In 2015 the episode weight indicates the number of time use episodes that a record on the Episode File represents (see p. 10 of Cycle 29's Documentation and User's Guide)

ACTCODE

Activity code of the episode.

Figure 1. Documentation for the dataset of walking episodes from 2015.

How to use `ActiveCA`

This section presents some potential applications of the `ActiveCA` R package. In fact, we expect that the application of this package to extend beyond our pre-imagined range of uses.

To start the demonstration, it is first necessary to install the package. The code chunk below demonstrates how to install the `ActiveCA` package in R:

```
if(!require("ActiveCA", character.only = TRUE)) {
      remotes::install_github("dias-bruno/ActiveCA")
}
```

After installing the `ActiveCA` R package, you can load it using the following code:

```
library(ActiveCA)
```

By running `?ActiveCA` in `R`, you can access the complete package documentation. The documentation includes a description of the package, information about the authors and maintainers, and an overview of all pre-processed data sets included in the package.

Figure 1 displays the documentation for the walking episodes from the 2015 GSS (`walking_2015.rda`). The documentation provides a brief description of the dataset, instructions on how to load the package, as well as details about the file format and descriptions of all variables (already labeled).

The chunk below shows how to load the walking episodes from the 2015 GSS (Cycle 29):

```
data(walking_2015)
```

The GSS surveys apply a probability sampling methodology, in which each episode or person selected in the sample represents several other episodes or persons not in the sample. The number of episodes and persons represented by a episode or person is determined by the weight or weighting factor. Because of this, every estimates of the number of episodes or persons need to be calculated applying the corresponding weighting factors.

For instance, if someone wants to calculate the percentage of respondents from the 2015 GSS survey with active travel episodes, it is necessary to account for the person's weights. In 2015, the weight variable is represented by `WGHT_PER`. The code chunk below demonstrates how to obtain the percentage of people with active travel episodes by age group. It uses the dplyr package to manipulate the data.

The process begins by creating a dataset that sums the population by age group. Then, it joins the 2015 episodes with the 2015 main file. Note that in both operations, the code sums the person weight variable (`WGHT_PER`) to obtain the correct population values. After this, a new dataset, `Active_percentage`, is created by merging both previous datasets. The percentage of active travel episodes by age group is calculated by dividing the total population by the population with active trip episodes, then multiplying by 100.

The result shows that the age group with the highest share of active trips is those between 15 and 20 years old, with almost 37%, followed by those between 25 and 34 years old with around 33%. There is a significant drop in percentage for the following groups, with the percentage falling to between 15% and 18% for the remaining age groups.

```
# Load the dplyr library for data manipulation
library('dplyr')

# Load the episodes dataset
data(gss_episodes)

# Calculate the total population for each age group
Total_population <- gss_main_2015 %>%
  group_by(AGEGR10) %>%
  summarise(Total_population  = sum(WGHT_PER))

# Calculate the active population for each age group
Active_population <- gss_episodes %>%
      filter(YEAR == 2015) %>%
  left_join(gss_main_2015, by = c("PUMFID"="PUMFID")) %>%
  group_by(AGEGR10) %>%
  summarise(Active_population  = sum(WGHT_PER))

# Combine total and active population and calculate the percentage
Active_percentage <- Total_population %>%
  left_join(Active_population, by = c("AGEGR10" = "AGEGR10")) %>%
```

```
  mutate(Percentage = round(Active_population/Total_population  * 1

# Display the final result
Active_percentage

# A tibble: 7 x 4
  AGEGR10           Total_population Active_population Percentage
  <ord>                        <dbl>             <dbl>      <dbl>
1 15 to 24 years            4511131.          1641468.       36.4
2 25 to 34 years            4956386.          1644033.       33.2
3 35 to 44 years            4734506.           847770.       17.9
4 45 to 54 years            5136125.           790667.       15.4
5 55 to 64 years            4831306.           767064.       15.9
6 65 to 74 years            3283969.           559025.       17.0
7 75 years and over         2312976.           384359.       16.6
```

Processed data set

Table 3 displays the total number of observations processed for main and episode files. For the main files, a total of 89,331 observations were processed, referring to all observations from the Time Use Surveys from 1986 to 2015, that together represents more of 149,389,839 respondents (29,766,399 for 2015, 28,075,610 for 2010, 26,095,819 for 2005, 24,260,137 for 1998, 21,294,313 for 1992, and for 2005, 24,260,137 for 1998, and 21,294,313 for 1986). Table X also presents the total cases of active trips episodes identified. In total 21,748 observations with register of active travel activity. Together, these observations account for NA episodes (0 for 2015, 0 for 2010, 0 for 2005, 0 for 1998, 0 for 1992, and for 2005, 0 for 1998, and 0 for 1986).

Table 3. Total number and weighted sum of observations processed.

		Observations	
Survey	Year	Count	Weighted
Main	2015	17390	29766399
	2010	15390	28075610
	2005	19597	26095819
	1998	10749	24260137
	1992	9815	21294313
	1986	16390	19897562

Episode	2015	3496	6634387
	2010	4615	8516753
	2005	5866	7583838
	1998	1789	3606987
	1992	1635	
	1986	4347	8241196

Table 4 shows the first ten rows and first six variables of the GSS PUMF 2015 main file (Cycle 29), displayed in 1 before our processing. Table 5 presents the walking episodes for the record identification number `10041` from the GSS PUMF 2015 episode file (Cycle 29), previously displayed in Table 2. Only active travel episodes appear in Table 5 since the observations were filtered to select cases with walking or cycling episodes. For both cases, Tables 4 and 5 contain labeled variables, facilitating the interpretation of the data.

Table 4. Visualization of the first ten lines and first six columns of the main file of the 2015 GSS.

PUMFID	WGHT_PER	SURVMNTH	AGEGR10	SEX	MARSTAT
10000	616.6740	July	55 to 64 years	Male	Divorced
10001	8516.6140	July	55 to 64 years	Male	Married
10002	371.7520	January	45 to 54 years	Female	Married
10003	1019.3135	March	65 to 74 years	Female	Divorced
10004	1916.0708	September	25 to 34 years	Male	Single, never married
10005	1952.2015	April	15 to 24 years	Male	Single, never married
10006	5761.5528	August	15 to 24 years	Male	Single, never married
10007	466.0426	June	55 to 64 years	Female	Widowed
10008	2479.2991	February	25 to 34 years	Female	Married
10009	1436.1641	August	65 to 74 years	Male	Widowed

Note:
Legend: ‘PUMFID‘: record identification. ‘WGHT_PER‘: person weight. ‘SURVMNTH‘: survey month of data collection. ‘AGEGR10‘: age group of the respondent. ‘SEX‘: sex of the respondent. ‘MARSTAT‘: marital status of the respondent.

Table 5. Visualization of active travel episodes of the record number ‘10041‘ for the 2015 GSS survey.

PUMFID	WGHT_EPI	Activity	Duration	Origin	Destination	Mode
10041	1353.818	Transport to or from activity	15	Home	Home	Walking

Note:
Legend: ‘PUMFID‘: record identification. ‘EPINO‘: episode number. WGHT_EPI: episode’s weight. TUI_01: activity code. DURATION: episode’s duration. LOCATION: episode’s location.

Descriptive statistics

Considering GSS Cycles analyzed, we identified 21,748 episodes that recorded active travel episodes, with trip duration ranging from 0 to 900 minutes, to twelve different

destinations. `ActiveCA` includes all these episodes ready for analysis. Table 6 presents descriptive statistics on walking and cycling trips between 1986 and 2015, with measures of trip duration in minutes: maximum (max), mean, median, and minimum (min). The 1986 survey did not include bicycle trips.

Table 6. Descriptive statistics for episodes with active transport records

		Year					
Mode	Statistic	1986	1992	1998	2005	2010	2015
Walking	Maximum	660	300	255	515	480	900
	Mean	21		12	12	13	18
	Median	15	10	5	10	10	10
	Minimum	1	1	1	0	0	5
	Standard deviation	31	25	17	16	17	27
Cycling	Maximum		240	90	180	153	120
	Mean		28	24	20	19	25
	Median		15	15	15	10	20
	Minimum		5	2	1	1	5
	Standard deviation		36	18	18	23	20

Table 6 shows that, in general, the median values for walking trips is 10 minutes, except for 1998 when the median was 5 minutes. In the case of cycling trips, the duration fluctuated over the years, ranging from 10 to 20 minutes. The table also highlights very high maximum values, particularly for walking trips, with recorded episodes exceeding 4 hours in all cases.

Table 7 and 8 provide descriptive statistics for the two modes of transportation, split by destination categories, from 1986 to 1998 and from 2005 to 2015, respectively. In Table 7, one can observed that in 1986 and 1992, walking trips destined for `home` had the highest medians. However, by 1998, the highest medians shifted to trips to `work or school`, a transition that also occurred for cycling trips between 1992 and 1998. Table 8 indicates that the median duration for walking trips to `home` and `work or school` remained at 10 minutes.

Table 7. Comparison of travel statistics by mode and destination: 1986, 1992, 1998

		1986				1992				1998			
Destination	**Mode**	**Min**	**Med**	**Max**	**(%)**	**Min**	**Med**	**Max**	**(%)**	**Min**	**Med**	**Max**	**(%)**
Cycling	Home					5	20	240	55.7	2	15	90	51.6
	Other's home					5	10	145	19.7	2	10	80	15.7
	Work or school					5	15	45	24.7	5	25	75	32.8
Walking	Home	1	15	330	46.3	1	15	300		1	5	255	51.9
	Other's home	1	10	660	42.6	1	5	135		1	5	120	27.8
	Work or school	1	10	450	11.1	2	10	60		1	10	75	20.3

Note:
'Min' denotes the minimum time to reach the destination; 'Max' denotes the maximum time to reach the destination; '(%)' indicates a percentage of the total time to reach the destination; 'Med' refers to the median time to reach the destination

Table 8. Comparison of travel statistics by mode and destination: 2005, 2010, 2015

		2005				2010				2015			
Destination	**Mode**	**Min**	**Med**	**Max**	**(%)**	**Min**	**Med**	**Max**	**(%)**	**Min**	**Med**	**Max**	**(%)**
Cycling	Cultural venues	10	10	15	0.3	10	25	30	1.0	15	15	15	0.5
	Grocery store	2	10	30	10.4	5	10	75	7.2	5	15	80	5.5
	Health clinic									10	15	90	1.9
	Home	1	15	180	48.4	1	10	135	48.8	5	20	120	46.4
	Neighbourhood									10	30	45	1.3
	Other's home	1	10	35	9.4	5	10	45	9.7	5	15	40	4.8
	Outdoors	5	15	45	6.0	3	10	115	2.6	15	30	30	1.0
	Place of worship	20	20	20	0.1					15	15	15	0.2
	Restaurant	5	15	35	2.8	10	15	153	1.7	10	20	60	3.0
	Sport area									10	15	15	2.0
	Work or school	1	15	90	22.6	1	15	100	29.0	5	20	120	33.5
Walking	Business									5	10	30	0.2
	Cultural venues	5	10	40	0.6	2	10	40	0.7	5	15	40	1.5
	Grocery store	1	10	90	11.5	1	7	105	13.1	5	10	130	10.3
	Health clinic									5	10	130	0.9
	Home	0	10	515	43.0	0	10	270	41.4	5	10	900	44.0
	Neighbourhood									5	10	60	2.6
	Other's home	1	5	300	10.2	0	5	140	10.2	5	10	120	6.2
	Outdoors	1	5	295	3.4	0	10	480	5.0	5	10	135	3.1
	Place of worship	1	10	30	0.7	1	7	60	0.8	5	15	45	1.1
	Restaurant	0	5	85	9.9	1	7	153	11.0	5	10	120	9.0
	Sport area									5	10	45	3.3
	Work or school	0	10	175	20.7	0	10	150	17.9	5	10	190	17.8

Note:
'Min' denotes the minimum time to reach the destination; 'Max' denotes the maximum time to reach the destination; '(%)' indicates a percentage of the total time to reach the destination; 'Med' refers to the median time to reach the destination

{ActiveCA} also enables visual analysis of active travel in Canada using traditional exploratory data analysis techniques. Figures 2 and 3 show walking and cycling trips from 1992 and 2015 through heat maps. These maps use color gradients to represent the percentage of trips between various origins and destinations, with darker colors indicating higher percentages and lighter colors representing less frequent routes. For conciseness, we omitted the heat maps for the other years analyzed.

In 1992, walking trips with `home` as both the origin and destination made up the majority, accounting for about 30% of all walking trips. These trips often involved leisure activities, like short walks or dog walking. Following this, trips from `home` to `work or school` comprised 18% of walking trips. Overall, `home` emerged as a crucial hub, either as an origin or destination, with only 5% of trips not involving `home`. By 2015, `home` remained a significant node, but new locations distributed the proportion of trips to areas not considered in 1992. In 2015, the highest proportion of trips were from `home` to `work or school` (12%) and vice versa (11%). `home` to `home` accounted for 8% of trips, and `grocery stores` became a notable destination for those leaving `home` (6%), surpassing trips to `other's home` (4%).

For cycling trips, Figure 3, shows that in 1992, when this mode of transportation was first included as an activity, the majority of trips were from `home` to `work or school`, accounting for about 25% of cases. This pattern remained in 2015, with these trips representing 30% of the cases. However, a notable change occurred in `home` to `home` trips, which decreased significantly from 19% in 1992 to 5% in 2015.

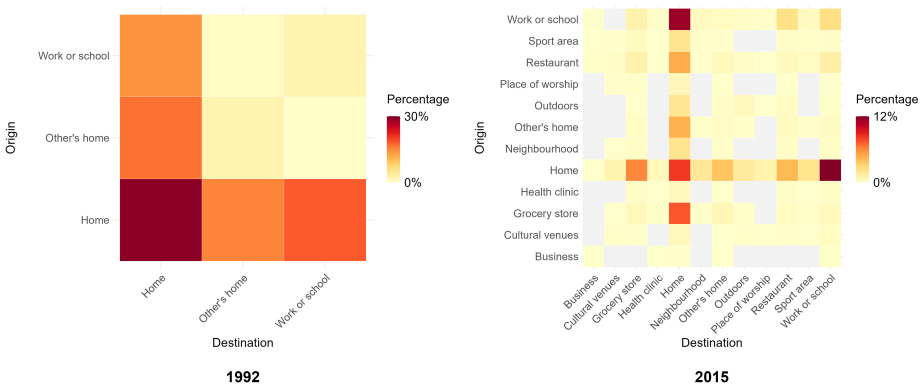


Figure 2. Percentage of walking trips categorized by origin and destination

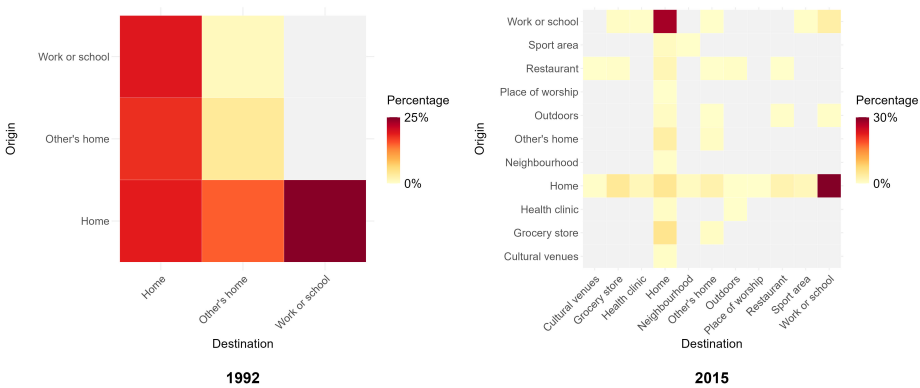


Figure 3. Percentage of cycling trips categorized by origin and destination

`ActiveCA` also enables obtaining insights from the main processed files. Figure 4 present how the level of stress varied among respondents depending on their marital status in 2015. According to this plot, married respondents reported the highest level of stress, with 17% of possible cases.

Python integration

Concluding remarks

This paper presents `ActiveCA`, an open data product that provides analysis-ready data from Cycles 2 (1986), 7 (1992), 12 (1998), 19 (2005), 24 (2010), and 29 (2015)

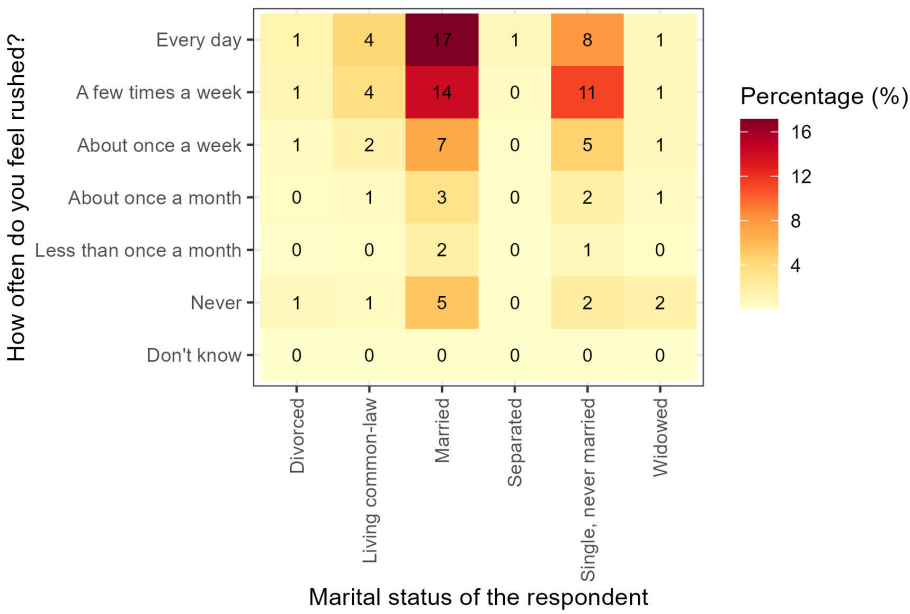


Figure 4. Level of stress among respondents of different marital statuses (2015).

of the GSS surveys on active travel in Canada. In the form of an R data package, {ActiveCA} was developed after collecting, cleaning, and processing the survey data, providing information on origins, destinations, and duration of active travel, as well other information.

It is important to remark that the present version of {ActiveCA} covers all Canadian time use surveys up to 2015. While the most recent time use survey was carried out in 2022 (Wray, 2024), the Public Use Microdata Files are currently unavailable, and at the moment it is estimated that they will only be published in the later part of 2025. The R package will be updated once these files become available.

The value of {ActiveCA} lies in its transparency, accessibility, and ease of use, which facilitates the addition of complementary data sets in the future. R users can seamlessly explore GSS walking and cycling episodes, with the option to suggest enhancements to the package as needed. This article adopts the structure proposed by Anastasia and Páez (2023), whose work provided essential guidance for the creation of this package. Similarly, we aim to contribute to the academic community by promoting transparent research practices that encourage replication and innovation in related fields. We believe that {ActiveCA} will serve as a basis for further research on GSS and for the integration of additional data by the authors or the wider open source community.

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

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Social Sciences and Humanities Research Council of Canada (*More description about the funding source after the review process*).

ORCID

Author 1
Author 2
Author 3

Data availability statement

The {ActiveCA} R data package can be found and installed on Github (*link*).

References

Arribas-Bel D, Green M, Rowe F and Singleton A (2021) Open data products-a framework for creating valuable analysis ready data. *Journal of Geographical Systems* 23(4): 497–514. DOI:10.1007/s10109-021-00363-5. URL https://doi.org/10.1007/s10109-021-00363-5.

Brunsdon C and Comber A (2021) Opening practice: supporting reproducibility and critical spatial data science. *Journal of Geographical Systems* 23(4): 477–496. DOI:10.1007/s10109-020-00334-2.

Canada S (2022) Time use survey. Technical report. URL https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4503. Last Modified: 2024-06-04.

Canada S (2024) Statistics canada: Canada's national statistical agency. Technical report. URL https://www.statcan.gc.ca/en/start.

Kim SO, Palm M, Han S and Klein NJ (2024) Facing a time crunch: Time poverty and travel behaviour in Canada. *Transportation Research Part D: Transport and Environment* 126: 104028. DOI:10.1016/j.trd.2023.104028. URL https://www.sciencedirect.com/science/article/pii/S136192092300425X.

Lachapelle U and Pinto DG (2016) Longer or more frequent walks: Examining the relationship between transit use and active transportation in Canada. *Journal of Transport & Health* 3(2): 173–180. DOI:10.1016/j.jth.2016.02.005. URL https://www.sciencedirect.com/science/article/pii/S2214140516000153.

McCurdy A, Faulkner G, Cameron C, Costas-Bradstreet C and Spence JC (2023) Support for Active Transport Policy Initiatives Among Canadian Adults: The Canadian National Active Transportation Survey. *Active Travel Studies* 3(2). DOI:10.16997/ats.1450. URL https://activetravelstudies.org/article/id/1450/. Number: 2 Publisher: University of Westminster Press.

Páez A (2021) Open spatial sciences: an introduction. *Journal of Geographical Systems* 23(4): 467–476. DOI:10.1007/s10109-021-00364-4. URL https://doi.org/10.1007/s10109-021-00364-4.

Soukhov A and Páez A (2023) Tts2016r: A data set to study population and employment patterns from the 2016 transportation tomorrow survey in the greater golden horseshoe area, ontario, canada. *Environment and Planning B: Urban Analytics and City Science* 50(2): 556–563. DOI:10.1177/23998083221146781. URL https://doi.org/10.1177/23998083221146781. Publisher: SAGE Publications Ltd STM.

Spinney JEL, Scott DM and Newbold KB (2009) Transport mobility benefits and quality of life: A time-use perspective of elderly Canadians. *Transport Policy* 16(1): 1–11. DOI: 10.1016/j.tranpol.2009.01.002.

Wray D (2024) Telework, time use, and well-being: Evidence from the 2022 time use survey. Technical report. URL https://www150.statcan.gc.ca/n1/daily-quotidien/240605/dq240605a-eng.htm.