SUPPLEMENTAL REFERENCE DOCUMENT FOR REVIEWING DATASETS

PUGET SOUND REGIONAL TRAVEL STUDY 2014

This document provides additional explanation of the Puget Sound Regional Travel Study datasets beyond the variable labels in the data dictionaries, for the datasets with the following release date: **December 5, 2014**. A comprehensive list of data variables can be found in the data codebooks . This document seeks to expand on those descriptions in cases where additional explanation may be helpful.

1.1 | PRIVACY

This dataset guide accompanies versions of the datasets that exclude certain data that can be considered sensitive or confidential. Specifically, home and trip (location) coordinates are not provided.

1.2 | OVERVIEW

5,996 households completed the main study in April–June of 2014, and 175 households (HHs) completed the pilot study in March 2014. Both pilot and main study households are included in the dataset.

After initial review of the dataset, 77 households were excluded from the dataset. An additional 52 households were removed as unusable from the dataset during the data clean-up in fall 2014 The final number of households in the dataset is 6,042.

The data deliverable includes five distinct datasets:

1. Household-level data
2. Person-level data
3. Trip-level data
4. Vehicle-level data

In the final dataset, so-called "meta" variable, concerning how the survey was taken by the survey participants, were moved to separate tabs in the Household and Person files.

All data in these datasets are from households that completed the entire survey, meaning they filled out the household information/recruit survey and every travel diary for household members age 5 and over. Partially complete households are not included in these datasets.

MISSING DATA

Blank cells are intentionally missing data, e.g. a question was not asked, or an answer choice was not shown to the respondent. Questions not asked in the pilot but asked in the main are coded -99. The household income questions offered respondents the option to select "Prefer not to answer", and has been coded 98 in the dataset. If data are missing because of a logic error, the value is coded to -9999 (unintentionally missing). This dataset does not have any unintentionally missing data.

1.3 | WEIGHTS

Initial household-level weights were developed in a two-step process as described in the memo titled "2014 Household Survey weighting process.pdf". In the first step, the number of survey households was expanded to the number of households in each sampling segment, by assigning an expansion factor to each household based on the sampling rate. The control totals for this step were ACS (American Community Survey) 2008-2012 number of households.

In the second step, these initial weights were adjusted to match demographic control data targets from the ACS PUMS (Public Use Microdata Sample) 2008-2012, within the four counties and 31 PUMA (Public Use Microdata Area) geographies. An iterative proportional fitting (IPF) procedure looped over the target dimensions and survey data with initial expansion weights, and gradually adjusted the weights to match the control totals, along the following dimensions:

- Household size (1, 2, 3, 4, 5+)
- Number of workers (0, 1, 2, 3+)
- Income group (nine categories, same as the detailed income question, but with the two highest income categories collapsed into one)
- Number of vehicles (0, 1, 2, 3+)
- Lifecycle (eight categories, a combination of presence of children age 0-4, presence of children age 5-17, number of adults 1 or 2+, and age of householder-under 35, 35-64, 65 or older)

The dataset can be used with or without weights. To create a frequency table of the number of household vehicles, for example, using sum of expwt_final will give the weighted distribution, whereas a count will give the un-weighted distribution.

The weighting process will be further described in the 2014 Household Survey Tech Memo.

1.4 | HOUSEHOLD-LEVEL DATASET

The household-level dataset has 6,042 rows- one row per household.

Unique identifier: HHID

Example = 14101093. All HHIDs start with '14', marking the year of the study.

Household weight (expwt_final)

Household-level expansion weight. The sum of household weights equals the number of households in the PSRC region based on ACS 2008-2012 county level control totals.

Sample segment (sample_segnum, sample_segname, h_segname, h_segname_wgt)

"Sample_segnum" refers to the household's assigned segment based on the purchased sample. "H_segname" is based on the final home address. Pilot segments were recoded to match the main study sample segments based on the block group. "H_segname_wgt" is the sample segment used in weighting, where the urban village oversample are separate groups.

Household income variables

Households had the option of reporting income in ten categories or select "prefer not to answer" ("hh_income_detailed"). A follow-up question offered the option of reporting a broad income category.

For weighting purposes, ten-category household income was imputed for households not responding to the ten-category income question. Income was imputed with a multinomial logit model with input variables being household and person demographics, and ACS income distribution in the home block group. Income was imputed for 704 households, including 233 households that answered the follow-up income question ("hh_income_followup", reported in five broad categories "hh_income_broad"). In "hh_income_detailed_imp", all households have one of the ten income categories, either self-reported or imputed.

Note that the imputed income data are estimates with some degree of uncertainty, and as such, any analysis performed with the imputed data should be interpreted with caution. To use only households with a reported income, use "hh_income_detailed" or "hh_income_broad".

In addition, the income estimate provided by the sample provider is included in the dataset ("hh_income_samp_est").

For full documentation on income imputation, see the Puget Sound Regional Travel Study Report.

Previous residence address information

Participants who reported living in their current residence for five years or less were asked whether their previous home was in Washington. If yes, participants were asked to locate their previous home address on a map. If the previous home was outside of Washington, participants had the option to either input the city and state of their previous residence, or both. This is why the numbers for previous city, state, and ZIP code are different. A few households reported their previous home was outside of Washington, and then selected "WA". These have not been recoded.

Household lifecycle stage (lifecycle)

A variable derived for use in weighting. Derived from the presence and ages of children, and the number and ages of adults in the household. Households are initially classified by the presence of children, first by the presence of any young children (under 5 years old) and then by the presence of only school children (ages 5-17). If there are no children in the household, it is then classified by the household size (either 1-person or 2-or-more-person household) and the age of the oldest person in the household (under 35, 35-64, and 65 years and over). Note: This variable uses the same lifecycle definition as PSRC 2006, with the slight difference that the age category is "under 5".

Household number of trips on travel day (hh_num_trips)

The number of trips reported by each household is the count of trip records associated with each household's password. This includes trip records for children under 5 (see Trip-level dataset).

[Note: the following fields, along with other fields concerning the survey-taking process by the household, have been moved to separate meta-data file.]

Recruit survey duration (hh_info_dur)

The recruit survey duration is calculated as the difference between the timestamp recorded for the last survey page ("participate") and the timestamp for the first survey page ("intro"). Exclude extremely long durations when interpreting survey duration as it is possible the respondent left their web browser open for periods of time.

Call center completed recruit survey (call_center)

If the user's IP address upon exiting the survey was determined to be located in Olathe Kansas or Blue Springs Missouri, the household is recorded as using the call center to complete the info survey. The call center's location is near the border of Kansas and Missouri. Note: To protect privacy, the IP address is not provided with the dataset.

Foreign language household (foreign_language)

A flag to indicate if the household was a foreign language participant, if the household used a web browser with a non-English language setting, called the PSRC language line, or called ETC and participated in Spanish.

1.5 | VEHICLE-LEVEL DATASET

The vehicle-level dataset has 9,464 rows – one row per vehicle reported by the 5,355 households (out of 6,042) that own a vehicle.

Unique identifier: hhid + vehnum

To get a unique identifier for the vehicle-level dataset, concatenate hhid and vehnum.

1.6 | PERSON-LEVEL DATASET

These are the person-level variables from the travel diary, for which there is one row per person for a total of 12,211 persons, including adults and children.

Unique identifier: PersonID

Example = 1410109301, where each PersonID is the household's HHID (14101093) with a unique ID appended for each household member (01, 02, etc.)

Person number (pernum)

Person number (1 through hhsize). Unique within each household. Person number 1 completed the recruit survey.

Type of household respondent (resptype)

This variable identifies a "primary" person (the person who completed the recruit survey), "other" household members (any household members aged 18 and up), and children (any household members under the age of 18).

Number of trips made on travel day (numtrips)

The number of trips reported by each respondent is the count of trip records associated with each person's personID. This includes trip records for children under 5 (see Trip-level dataset).

Travel diary duration (diary_duration_min)

The full diary duration, or the amount of time participants spent taking the travel diary survey, is calculated as the difference between the timestamp recorded for the last survey page including the stated preference section ("comments") and the first survey page ("loc_start"). Use care when interpreting survey duration, as it is possible the respondent left their web browser open for long periods of time.

Proxy variable (proxy)

The "proxy" variable indicates if respondents took the diary for themselves, if other people filled out the answers while they were present, or if people filled out the answers and they were not present. The call center was given instructions to answer this question as if they were the respondent.

Previous workplace

Participants who commute to their current workplace for five years or less were asked whether their previous workplace was in Washington. If yes, participants were asked to locate their previous work address on a map. If the previous workplace was outside of Washington, participants had the option to either input the city and state of their previous workplace, or both. This is why the numbers for previous city, state, and ZIP code are different. A few respondents reported their previous home was outside of Washington, and then selected "WA". These have not been recoded.

Had additional trips to report (added_trip_flag)

Indicates if the respondent went back and added more trips in the roster after seeing the prompt asking whether they had made any additional trips not already reported. Also, see related variables with prefix "added_" for the types of trips that were added.

1.7 | TRIP-LEVEL DATASET

These are the trip-level variables, from the trips made on the assigned travel date. The Trip-level dataset has 50,103 trips – one row per (one-way) trip.

The travel date starts at 3 AM on the assigned travel date and ends at 3 AM the following day.

Unique identifier: recordID

This is a permanent ID attached to each original trip record at the time the data was received from the consultant. It is in the form "A*nnnnn*" where "*nnnnn*" can range from 00001 to 50856 (the original number of trip records. If a record is removed from the file, that recordID is not reused.

When a new record is added, that record is assigned a recordID in the form "A9*nnnn*", where "*nnnn*" begins at 0001 and goes up to 0231, the maximum number of new records added in this release. There is no specific

order in which new records were added, such that in a diary with more than one new record added, the recordIDs may not be sequential.

Unique identifier: tripID

Example = 141010930101, where each tripID is the household's HHID (14101093) with a unique ID appended for each household member (01, 02, etc.) and a unique ID appended for each unique trip (01, 02, etc.)

This identifier is updated whenever a trip is added or removed from the diary.

Person number (pernum)

Person number (1 through hhsize). Unique within each household. Person number 1 completed the recruit survey.

Trip number (tripnum)

Values 1 through number of trips for a respondent, sorted by departure time (the first trip of the day is trip number 1). Trip number is unique within each respondent.

This number is updated whenever a trip is added or removed from the diary.

Origin trip purpose and destination trip purpose (o_purpose and d_purpose)

Respondents report the destination trip purpose. The origin purpose is derived from the destination purpose of the previous trip, except for first trip in the day. For the first trip in the day, origin purpose is instead coded based origin-location description in place_start, and is typically home.

Start and end time: Minutes after midnight (time_start_mam and time_end_mam)

Format is minutes after midnight, from 3 AM on the travel date to 3 AM on the following day.

Start and end time: HH:MM (time_start_hhmm and time_end_hhmm)

Format is HH:MM (24-hour clock). Included for convenience.

Reported trip duration (trip_duration)

Travel time is derived as the difference between respondent-reported start and end time of the trip.

Driving distance and time (gdist, gtime)

The RSG survey instrument estimated travel time and duration for each trip in addition to the user-reported travel time. The estimates of time and duration were calculated using the Google Maps API Distance Matrix Service. These estimates indicate the distance and duration of a trip for "standard driving directions using the road network". The estimates do not account for traffic, thus representing free flow conditions on the roadway. Google estimated drive time and distance can be used to validate the reported trip durations. Time and distance estimates were not available for 18 of the ferry and airplane trips.

Activity duration

This is computed as the difference between the arrival time of the current trip and the departure time of the following trip. For the last trip of the day, ADUR is calculated at the remaining time to 3:00 AM the following day.

Implied speed in miles per hour (implied_speed)

Driving distance over reported travel time. Can be used in trip validation to detect trip records with issues (extremely high or low speeds).

Driver or passenger (driver)

Asked of respondents whose mode was drive with others (household members or not) or vanpool. Derived as driver for drive-alone trips.

Trip copied from other household member (prepop)

Flag to indicate that a respondent copied this trip from another household member who had already reported them on their trip. This option is available to all household members regardless of age, and reduces the respondent burden of repeating trip details.

Trip record derived from trip reported by household member age 5 or over (child_under5)

Children under the age of 5 were not required to complete a diary, but could be reported on trips made by household members 5 or older. During data processing, trip records were created for the children under 5 by copying trip records from other household members and editing relevant details such as creating unique trip ids and recoding instances of 'driver' to 'passenger' for vehicle trips. The variable "child_under5_originaltripid" indicates the trip the record was created from.

Household member on trip (member1 through member8)

PersonIDs of household members on the trip.