# Package: tripaccess (via r-universe)

October 27, 2026

**Title** American Travel Behavior and Access Datasets

**Version** 0.2.0.9000

**Description** Subsets of data from the National Household Travel Survey 2017. It includes personal trips, mobility, demographic, and household information. It is suitable for data visualization, data wrangling, joining datasets, exploratory data analysis, group comparisons, simple linear regression, categorical data analysis, and data ethics discussion in data science and statistics classes.

**License** CC0 | file LICENSE

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.3

**URL** https://github.com/scao53/tripaccess, https://scao53.github.io/tripaccess/

**BugReports** https://github.com/scao53/tripaccess/issues

**Depends** R (>= 4.1.0)

**LazyData** true

**LazyDataCompression** xz

**Suggests** knitr, rmarkdown (>= 2.31), testthat (>= 3.0.0), tidyverse, tidytree, dplyr

**Config/testthat/edition** 3

**VignetteBuilder** knitr

**Repository** https://scao53.r-universe.dev

**Date/Publication** 2026-08-28 18:51:38 UTC

**RemoteUrl** https://github.com/scao53/tripaccess

**RemoteRef** HEAD

**RemoteSha** 25033477d55b99ff95612342c37ac4081eb4b5cd

# Contents

house . . . . . 2
person . . . . . 3
trip . . . . . 5
tripaccess . . . . . 6

**Index** **9**

---

`house` *House data*

---

## Description

Include household characteristics categorical and numeric variables.

## Usage

```
house
```

## Format

A data frame with 129695 rows (each row is a household) and 9 columns

**household_id** Household identifier. Use this variable to join the house dataset and the person dataset as well as join the house dataset and the trip dataset.

**region** 2010 Census division classification for the respondent's home address.

**number_drivers** Number of drivers in household.

**count_household_members** Count of household members.

**number_vehicles** Count of household vehicles.

**household_life_cycle** Life Cycle classification for the household, derived by attributes pertaining to age, relationship, and work status.

**count_adult_household_members** Count of adult household members at least 18 years old.

**number_workers** Number of workers in household.

**count_young_child** Count of persons with an age between 0 and 4 in household.

## Source

https://nhts.ornl.gov/

## Examples

```
if (require("tidyverse")) {
# Filtered to households with at least one driver
house_with_drivers <- house |>
  filter(number_drivers > 0)

# Filtered to households with at least one vehicle
house_with_vehicles <- house_with_drivers |>
  filter(number_vehicles > 0)

# Plot household vehicles by number of drivers
ggplot(data = house_with_vehicles,
       aes(x = number_drivers,
           y = number_vehicles)) +
  geom_jitter(alpha = 0.08, width = 0.15, height = 0.15) +
  geom_smooth(method = lm, se = FALSE, formula = y ~ x, color = "blue") +
  labs(title = "Household Vehicles versus Number of Drivers",
       x = "Number of Drivers in Household",
       y = "Number of Household Vehicles") +
  theme_bw()
}
```

---

# person *Person data*

---

## Description

Include disability, mobility, and other demographic categorical and numeric variables.

## Usage

```
person
```

## Format

A data frame with 99564 rows (each row is a person) and 32 columns

**household_id** Household identifier. Use this variable to join the person dataset and the house dataset. Use both this variable and person_id to join the person dataset and the trip dataset.

**person_id** Person identifier. Use both this variable and household_id to join the person dataset and the trip dataset.

**travel_disability** How long the respondent has had a medical condition that makes it difficult to travel outside of home. Values include 6_months_or_less_disability, More_than_6_months_of_disability, Lifelong_disability, and No_disability.

**sex** Sex of the respondent. Values include Male and Female.

**race** Race of the respondent. Values include White, Black, Asian, American Indian, Hawaiian/Pacific Islander, Multiracial, and Other.

**hispanic_ethnicity** Hispanic or Latino origin. Values include Hispanic and Non-Hispanic.

**nativity** Born in United States. Values include Yes and No.

**age** Age of the respondent. Filtered to ages 18 to 61.

**education** Educational attainment. Values include Less than a high school graduate, High school graduate or GED, Some college or associates degree, Bachelor's degree, and Graduate degree or professional degree.

**self_rated_health** Opinion of health. Values include Excellent, Very good, Good, Fair, and Poor.

**employment_status** Primary activity in previous week. Values include Employed and Unemployed.

**household_income** Household income. Values include Under $10,000, $10,000 to $34,999, $35,000 to $74,999, $75,000 to $149,999, and $150,000 and over.

**household_structure** Count of household members. Values include Lives alone and Does not live alone.

**population_density** Category of population density (persons per square mile) in the census block group of the household's home location. Values include 0-99, 100-499, 500-999, 1,000-1,999, 2,000-3,999, 4,000-9,999, 10,000-24,999, and 25,000 and over.

**urban_rural** Household in urban or rural area. Values include Urban and Rural.

**state** Household state. Includes the 50 states and Washington, DC.

**driver_status** Driver status. Values include Drives and Does not drive.

**cane** Does the respondent use a cane to aid their travel?

**manual_wheelchair** Does the respondent use a manual wheelchair to aid their travel?

**crutches** Does the respondent use a crutch to aid their travel?

**dog** Does the respondent use a dog to aid their travel?

**motorized_wheelchair** Does the respondent use a motorized wheelchair to aid their travel?

**scooter** Does the respondent use a scooter to aid their travel?

**white_cane** Does the respondent use a white cane to aid their travel?

**walker** Does the respondent use a walker to aid their travel?

**other_accommodation** Derived from the original W_NONE variable. Indicates whether no listed mobility aid was reported.

**yearly_miles_personally_driven** Miles personally driven in all vehicles. Values range from 0 to 200000.

**count_of_public_transit_usage** Count of public transit usage in last month. Values range from 0 to 30.

**count_of_rideshare_app_usage** Count of rideshare app usage in last month. Values range from 0 to 99.

**count_of_bike_trips** Count of bike trips in past week. Values range from 0 to 99.

**count_of_walk_trips** Count of walk trips in past week. Values range from 0 to 200.

**count_of_online_delivery** Count of times purchased online for delivery in last 30 days. Values range from 0 to 99.

## Source

https://nhts.ornl.gov/

## Examples

```
if (require("tidyverse")) {
# Summary statistics of public transit use by travel disability status
transit_summary <- person |>
  group_by(travel_disability) |>
  summarize(
    people = n(),
    public_transit_users = sum(count_of_public_transit_usage > 0),
    public_transit_use_prop = mean(count_of_public_transit_usage > 0),
    public_transit_usage_median = median(count_of_public_transit_usage),
    public_transit_usage_mean = mean(count_of_public_transit_usage),
    public_transit_usage_sd = sd(count_of_public_transit_usage)
  )
# Test whether public transit use differs by travel disability status
prop.test(
  x = transit_summary$public_transit_users,
  n = transit_summary$people
)
}
```

---

# trip *Trip data*

---

## Description

Include trip related categorical and numeric variables.

## Usage

```
trip
```

## Format

A data frame with 921590 rows (each row is a trip) and 8 columns

**household_id** Household identifier. Use both this variable and person_id to join the trip dataset and the person dataset. Use this variable to join the trip dataset and the house dataset.

**person_id** Person identifier. Use both this variable and household_id to join the trip dataset and the person dataset.

**trip_purpose** Generalized purpose of trip on travel day. A travel day is a 24-hour day that starts at 4:00 a.m. (local time) of the assigned travel day and ends at 3:59 a.m. of the following day. The NHTS randomly assigns the travel days for one-seventh of the sample addresses to each day of the week and the remaining six-sevenths of the households to evenly across weekdays (Monday-Friday).

**gas_price** Price of gasoline, in cents, on respondent's travel day.

**num_of_people_on_trip** Number of people on trip including respondent on respondent's travel day.

**trip_miles** Trip distance in miles on respondent's travel day, derived from route geometry returned. Google Maps was used for routing the shortest path for motorized travel on the road network. Non-motorized modes, like walk and bike, had the shortest path calculated using network routes paths.

**trip_duration** Trip duration in minutes on respondent's travel day.

**trip_miles_personally_driven_vehicle** Trip distance in miles for personally driven vehicle trips on respondent's travel day. -1 = Appropriate skip.

## Source

https://nhts.ornl.gov/

## Examples

```
if (require("tidyverse")) {
# Filtered to shorter trips for a clearer introductory visualization
short_trips <- trip |>
  filter(trip_miles <= 50,
         trip_duration <= 180)

# Filtered to trips with positive distance and duration
positive_distance_trips <- short_trips |>
  filter(trip_miles > 0,
         trip_duration > 0)

# Fit a simple linear regression model
duration_miles_model <- lm(trip_duration ~ trip_miles,
                           data = positive_distance_trips)
summary(duration_miles_model)

# Correlation between trip distance and trip duration
cor(positive_distance_trips$trip_miles, positive_distance_trips$trip_duration)
}
```

---

# tripaccess *Tripaccess data*

---

## Description

Include disability and other demographic as well as mobility and trip categorical and numeric variables for data visualization and single table analysis.

## Usage

```
tripaccess
```

## Format

A data frame with 86521 rows (each row is a person) and 41 columns

**household_id** Household identifier.

**person_id** Person identifier.

**travel_disability** How long does the respondent have a medical condition that makes it difficult to travel outside of home?

**sex** Sex of the respondent.

**race** Race of the respondent.

**hispanic_ethnicity** Hispanic or Latino origin.

**nativity** Born in United States.

**age** Age of the respondent.

**education** Educational attainment.

**self_rated_health** Opinion of health.

**employment_status** Primary activity in previous week.

**household_income** Household income of the respondent.

**household_structure** Count of household members.

**population_density** Category of population density (persons per square mile) in the census block group of the households home location.

**urban_rural** Household in urban or rural area.

**state** Household state.

**driver_status** Driver status.

**cane** Does the respondent use a cane to aid their travel?

**manual_wheelchair** Does the respondent use a manual wheelchair to aid their travel?

**crutches** Does the respondent use a crutch to aid their travel?

**dog** Does the respondent use a dog to aid their travel?

**motorized_wheelchair** Does the respondent use a motorized wheelchair to aid their travel?

**scooter** Does the respondent use a scooter to aid their travel?

**white_cane** Does the respondent use a white cane to aid their travel?

**walker** Does the respondent use a walker to aid their travel?

**other_accommodation** Does the respondent use any other accommodation rather than those listed?

**yearly_miles_personally_driven** Miles personally drive in all vehicles.

**count_of_public_transit_usage** Count of public transit usage in last month.

**count_of_rideshare_app_usage** Count of rideshare app usage in last month.

**count_of_bike_trips** Count of bike trips in past week.

**count_of_walk_trips** Count of walk trips in past week.

**count_of_online_delivery** Count of times purchased online for delivery in last 30 days.

**avg_num_of_people_on_trip** Average number of people on the respondent's trip including the respondent on the respondent's travel day. A travel day is a 24 hour day that starts at 4 am (local time) of the assigned travel day and ends at 3:59 am of the following day. The NHTS randomly assigns the travel days for one-seventh of the sample addresses to each day of the week and the remaining six-sevenths of the households to evenly across weekdays (Monday to Friday). The following variables are related to trip information on travel day.

**avg_trip_distance_in_miles** Average trip distance in miles on the respondent's travel day.

**avg_trip_duration_in_minutes** Average trip duration in minutes on the respondent's travel day.

**shopping_trip** Whether having shopping trips on the respondent's travel day.

**social_recreational_trip** Whether having social or recreational trips on the respondent's travel day.

**other_home_based_trip** Whether having other home-based trips on the respondent's travel day.

**work_trip** Whether having work trips on the respondent's travel day.

**other_non_home_based_trip** Whether having other non home-based trips on the respondent's travel day.

**person_weight** Person weights that can be used for person-level estimates of non-household and non-travel day items of interest, for example workers by gender, drivers by annual miles estimated, etc.

## Source

https://nhts.ornl.gov/

## Examples

```
if (require("tidyverse")) {
# Filtered to people who have a travel disability
tripaccess_disabled <- tripaccess |>
  filter(travel_disability != "No_disability")

# Summary statistics of public transit usage by disabled people who use a walker
tripaccess_disabled |>
  filter(walker == "True") |>
  group_by(travel_disability) |>
  summarize(public_transit_usage_median = median(count_of_public_transit_usage),
  public_transit_usage_mean = mean(count_of_public_transit_usage),
  public_transit_usage_sd = sd(count_of_public_transit_usage))
}
```

# Index

∗ **datasets**
house, 2
person, 3
trip, 5
tripaccess, 6

house, 2

person, 3

trip, 5
tripaccess, 6