

# Exploratory Data Analysis

DSAN 5300

Marion Bauman

## Table of contents

Household Data . . . . .	2
Long Distance Data . . . . .	6
Person Data . . . . .	12
Trip Data . . . . .	24
Vehicle Data . . . . .	32

In this notebook, we will explore the NHTS data set in order to understand the data and determine what data modeling steps to take.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.0      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(arrow)
```

Attaching package: 'arrow'

The following object is masked from 'package:lubridate':

duration

The following object is masked from 'package:utils':

timestamp

```
household <- read_parquet("../data/clean/household.parquet")
long_distance <- read_parquet("../data/clean/long_distance.parquet")
person <- read_parquet("../data/clean/person.parquet")
trip <- read_parquet("../data/clean/trip.parquet")
vehicle <- read_parquet("../data/clean/vehicle.parquet")
```

## Household Data

```
household |> glimpse()
```

Rows: 7,893

Columns: 35

\$ household_id	<chr> "9000013002", "9000013016", "9000013026~
\$ household_weight_7_day	<dbl> 4621.758, 2982.998, 9979.573, 10355.891~
\$ household_weight_5_day	<dbl> 0.000, 3878.814, 14700.797, 13150.679, ~
\$ household_weight_2_day	<dbl> 13946.19, 0.00, 0.00, 0.00, 0.00, 0.00,~
\$ num_adults_in_household	<dbl> 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2, ~
\$ num_drivers_in_household	<dbl> 2, 2, 1, 2, 1, 2, 1, 2, 1, 2, 0, 2, 2, ~
\$ count_household_trips	<dbl> 7, 6, 0, 2, 0, 4, 1, 4, 8, 0, 0, 6, 7, ~
\$ household_size	<dbl> 4, 2, 1, 4, 1, 3, 1, 2, 2, 2, 1, 5, 2, ~
\$ num_vehicles_in_household	<dbl> 2, 1, 0, 1, 1, 2, 1, 2, 1, 2, 0, 2, 3, ~
\$ num_children_5_17	<dbl> 2, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 1, 0, ~
\$ num_children_under_5	<dbl> 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 2, 0, ~
\$ num_respondents_in_household	<dbl> 4, 2, 1, 4, 1, 2, 1, 2, 2, 2, 1, 3, 2, ~
\$ travel_day_date	<chr> "202202", "202202", "202202", "202201",~
\$ WRKCOUNT	<dbl> 1, 2, 1, 1, 1, 2, 0, 2, 2, 0, 0, 1, 1, ~
\$ household_stratum_id	<chr> "1021", "1021", "1021", "1021", "1021",~
\$ household_grouping	<chr> "New England MSA/CMSA of 1 million+ wit~
\$ census_division	<chr> "New England", "New England", "New Engl~
\$ census_region	<chr> "Northeast", "Northeast", "Northeast", ~
\$ all_household_completed_survey	<chr> "All eligible household members complet~
\$ household_income	<chr> "\$200,000 or more", "\$75,000 to \$99,999~

```

$ household_income_imputed      <chr> "$200,000 or more", "$75,000 to $99,999~
$ household_related             <chr> "At least two persons in hh are related~
$ hispanic                     <chr> "Not Hispanic or Latino", "Not Hispanic~
$ race                         <chr> "White", "White", "White", "White", "Wh~
$ own_or_rent_home             <chr> "Owned by hh member with mortgage or lo~
$ home_type                    <chr> "One-family detached", "Building with 2~
$ LIF_CYC                      <chr> "2+ adults, youngest child 0-5", "2+ ad~
$ metro_area_category           <chr> "MSA of 1 million or more, with rail", ~
$ metro_area_size               <chr> "In an MSA or CMSA of 3 million or more~
$ metro_rail_status             <chr> "Yes", "No", "Yes", "Yes", "No", "Yes",~
$ travel_day_of_week           <chr> "Sunday", "Wednesday", "Tuesday", "Thur~
$ household_urban_classification <chr> "In an urban area", "In an urban area",~
$ household_urban_size          <chr> "1,000,000 or more with heavy rail", "2~
$ household_urban_rural         <chr> "Urban", "Urban", "Urban", "Urban", "Ur~
$ household_urban_rural_2010    <chr> "Urban", "Urban", "Urban", "Urban", "Ur~

```

```
household |> summary()
```

```

household_id      household_weight_7_day household_weight_5_day
Length:7893      Min.   : 100.8      Min.   : 0
Class :character  1st Qu.: 6571.7      1st Qu.: 0
Mode  :character  Median :11529.1      Median : 10534
                  Mean   :16159.2      Mean   : 16159
                  3rd Qu.:20636.0      3rd Qu.: 22549
                  Max.   :81832.0      Max.   :112970

household_weight_2_day num_adults_in_household num_drivers_in_household
Min.   : 0      Min.   :1.000      Min.   :0.000
1st Qu.: 0      1st Qu.:1.000      1st Qu.:1.000
Median : 0      Median :2.000      Median :2.000
Mean   : 16159   Mean   :1.839      Mean   :1.714
3rd Qu.: 15082   3rd Qu.:2.000      3rd Qu.:2.000
Max.   :264588   Max.   :8.000      Max.   :7.000

count_household_trips household_size num_vehicles_in_household
Min.   : 0.000      Min.   : 1.000      Min.   : 0.00
1st Qu.: 2.000      1st Qu.: 1.000      1st Qu.: 1.00
Median : 3.000      Median : 2.000      Median : 2.00
Mean   : 3.937      Mean   : 2.272      Mean   : 1.86
3rd Qu.: 6.000      3rd Qu.: 3.000      3rd Qu.: 2.00
Max.   :45.000      Max.   :10.000      Max.   :17.00

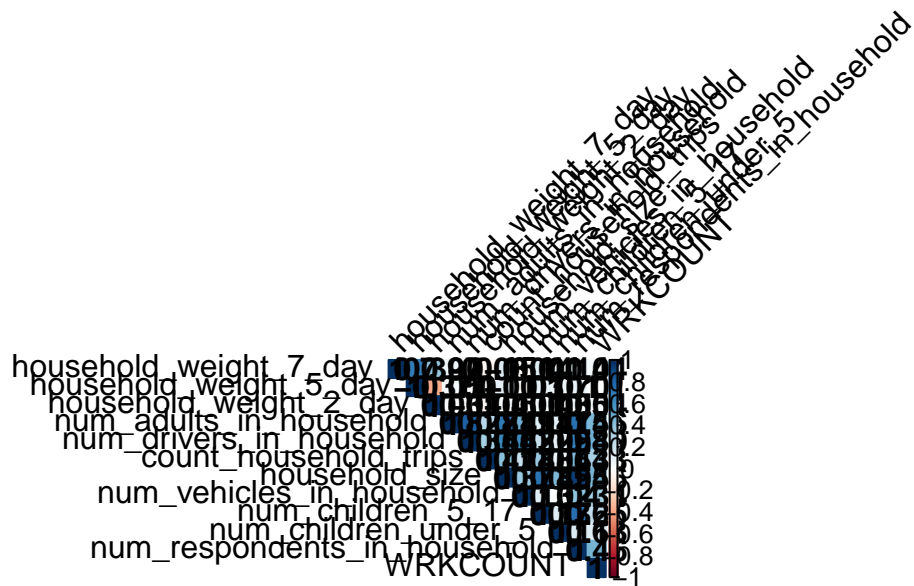
num_children_5_17 num_children_under_5 num_respondents_in_household
Min.   :0.000      Min.   :0.0000      Min.   : 1.000
1st Qu.:0.000      1st Qu.:0.0000      1st Qu.: 1.000

```

Median :0.000	Median :0.0000	Median : 2.000	
Mean :0.328	Mean :0.1045	Mean : 2.153	
3rd Qu.:0.000	3rd Qu.:0.0000	3rd Qu.: 3.000	
Max. :6.000	Max. :4.0000	Max. :10.000	
travel_day_date	WRKCOUNT	household_stratum_id	household_grouping
Length:7893	Min. :0.0000	Length:7893	Length:7893
Class :character	1st Qu.:0.0000	Class :character	Class :character
Mode :character	Median :1.0000	Mode :character	Mode :character
	Mean :0.9966		
	3rd Qu.:2.0000		
	Max. :6.0000		
census_division	census_region	all_household_completed_survey	
Length:7893	Length:7893	Length:7893	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
household_income	household_income_imputed	household_related	
Length:7893	Length:7893	Length:7893	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
hispanic	race	own_or_rent_home	home_type
Length:7893	Length:7893	Length:7893	Length:7893
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
LIF_CYC	metro_area_category	metro_area_size	metro_rail_status
Length:7893	Length:7893	Length:7893	Length:7893
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
travel_day_of_week	household_urban_classification	household_urban_size	
Length:7893	Length:7893	Length:7893	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	

```
household_urban_rural household_urban_rural_2010
Length:7893          Length:7893
Class :character     Class :character
Mode :character      Mode :character
```

```
# Corr plot
household |>
  select(-household_id) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot(
    method = "color",
    type = "upper",
    tl.col = "black",
    tl.srt = 45,
    addCoef.col = "black",
  )
```



## Long Distance Data

```
long_distance |> glimpse()
```

Rows: 16,997

Columns: 66

```
$ household_id      <chr> "9000013002", "9000013002", ~
$ person_id        <chr> "01", "02", "03", "04", "01~
$ LONGDIST         <dbl> 1, 1, 0, 0, 0, 0, 1, 0, 0, ~
$ num_people_on_long_distance_trip <dbl> 2, 2, 5, 5, 2, 2, 1, 4, 4, ~
$ num_used_amtrak_last_year      <dbl> 0, 0, 0, 0, 0, 0, 2, 0, 0, ~
$ num_used_intracity_bus_last_year <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
$ num_nights_away_from_home      <dbl> 3, 3, -9, -9, -9, -9, 8, -9~
$ great_circle_distance_traveled <dbl> 61.86880, 61.86880, 382.211~
$ num_adults_in_household        <dbl> 2, 2, 2, 2, 2, 2, 1, 2, 2, ~
$ num_drivers_in_household       <dbl> 2, 2, 2, 2, 2, 2, 1, 2, 2, ~
$ household_size                 <dbl> 4, 4, 4, 4, 2, 2, 1, 4, 4, ~
$ num_vehicles_in_household      <dbl> 2, 2, 2, 2, 1, 1, 0, 1, 1, ~
$ travel_day_date               <chr> "202202", "202202", "202202~
$ WRKCOUNT                   <dbl> 1, 1, 1, 1, 2, 2, 1, 1, 1, ~
$ household_stratum_id         <chr> "1021", "1021", "1021", "10~
$ person_weight_7_day          <dbl> 3938.689, 5409.515, 3013.58~
$ person_weight_5_day          <dbl> 0.000, 0.000, 0.000, 0.000,~
$ person_weight_2_day          <dbl> 13363.809, 19360.095, 7285.~
$ respondent_age               <dbl> 39, 42, 8, 5, 32, 41, 44, 3~
$ airport_size                 <chr> "Valid Skip", "Valid Skip",~
$ begin_trip_date              <chr> NA, NA, "Not ascertained", ~
$ household_grouping           <chr> "New England MSA/CMSA of 1 ~
$ census_division              <chr> "New England", "New England~
$ census_region                <chr> "Northeast", "Northeast", "~
$ DRIVER                       <chr> "Yes", "Yes", "Valid skip",~
$ education_level              <chr> "Master's degree", "Master'~
$ ENDTRIP                      <chr> NA, NA, "Not ascertained", ~
$ exit_census_division         <chr> "Valid Skip", "Valid Skip",~
$ farthest_census_division_traveled <chr> "New England", "New England~
$ farthest_census_region_traveled <chr> "Northeast", "Northeast", "~
$ reason_for_travel            <chr> "Personal vacation", "Perso~
$ great_circle_distance_longer_than_50_miles <chr> "GCD 50 miles or more", "GC~
$ household_income             <chr> "$200,000 or more", "$200,0~
$ household_income_imputed     <chr> "$200,000 or more", "$200,0~
$ hispanic                    <chr> "Not Hispanic or Latino", "~
```

```

$ race <chr> "White", "White", "White", ~
$ own_or_rent_home <chr> "Owned by hh member with mo~
$ international_trip <chr> "Inside the US", "Inside th~
$ long_distance_trip <chr> "Long distance info asked o~
$ LIF_CYC <chr> "2+ adults, youngest child ~
$ main_mode_of_transport <chr> "Car", "SUV/Crossover", "SU~
$ date_of_most_recent_trip <chr> NA, NA, "Not ascertained", ~
$ metro_area_category <chr> "MSA of 1 million or more, ~
$ metro_area_size <chr> "In an MSA or CMSA of 3 mil~
$ person_1_on_long_distance_trip <chr> "On trip", "On trip", "On t~
$ person_10_on_long_distance_trip <chr> "Not ascertained", "Not asc~
$ person_2_on_long_distance_trip <chr> "On trip", "On trip", "On t~
$ person_3_on_long_distance_trip <chr> "Not on trip", "Not on trip~
$ person_4_on_long_distance_trip <chr> "Not on trip", "Not on trip~
$ person_5_on_long_distance_trip <chr> "Not ascertained", "Not asc~
$ person_6_on_long_distance_trip <chr> "Not ascertained", "Not asc~
$ person_7_on_long_distance_trip <chr> "Not ascertained", "Not asc~
$ person_8_on_long_distance_trip <chr> "Not ascertained", "Not asc~
$ person_9_on_long_distance_trip <chr> "Not ascertained", "Not asc~
$ proxy_respondent <chr> "Self", "Someone else", "So~
$ metro_rail_status <chr> "Yes", "Yes", "Yes", "Yes",~
$ respondent_hispanic <chr> "Not Hispanic", "Not Hispan~
$ respondent_race <chr> "White", "White", "White", ~
$ respondent_sex <chr> "Female", "Male", "Refuse",~
$ respondent_sex_imputed <chr> "Female", "Male", "Male", "~
$ travel_day_of_week <chr> "Sunday", "Sunday", "Sunday~
$ household_urban_classification <chr> "In an urban area", "In an ~
$ household_urban_size <chr> "1,000,000 or more with hea~
$ household_urban_rural <chr> "Urban", "Urban", "Urban", ~
$ trip_includes_weekend <chr> "Long weekend (4 to 6 days ~
$ worker_status <chr> "Not worker", "Worker", "Va~

```

```
long_distance |> summary()
```

```

household_id      person_id      LONGDIST
Length:16997      Length:16997      Min.   : 0.000
Class :character  Class :character  1st Qu.: 0.000
Mode  :character  Mode  :character  Median : 0.000
                                Mean  : 1.701
                                3rd Qu.: 1.000
                                Max.   :207.000
num_people_on_long_distance_trip num_used_amtrak_last_year

```

Min. : -9.000	Min. : 0.00000		
1st Qu.: 1.000	1st Qu.: 0.00000		
Median : 2.000	Median : 0.00000		
Mean : 1.932	Mean : 0.06642		
3rd Qu.: 3.000	3rd Qu.: 0.00000		
Max. : 99.000	Max. : 35.00000		
num_used_intracity_bus_last_year	num_nights_away_from_home		
Min. : 0.00000	Min. : -9.000		
1st Qu.: 0.00000	1st Qu.: -9.000		
Median : 0.00000	Median : -1.000		
Mean : 0.08131	Mean : -3.037		
3rd Qu.: 0.00000	3rd Qu.: 0.000		
Max. : 100.00000	Max. : 31.000		
great_circle_distance_traveled	num_adults_in_household		
Min. : -9.00	Min. : 1.000		
1st Qu.: 11.26	1st Qu.: 2.000		
Median : 79.33	Median : 2.000		
Mean : 258.11	Mean : 2.135		
3rd Qu.: 241.79	3rd Qu.: 2.000		
Max. : 5005.86	Max. : 8.000		
num_drivers_in_household	household_size	num_vehicles_in_household	
Min. : 0.000	Min. : 1.00	Min. : 0.000	
1st Qu.: 2.000	1st Qu.: 2.00	1st Qu.: 1.000	
Median : 2.000	Median : 2.00	Median : 2.000	
Mean : 1.998	Mean : 2.91	Mean : 2.098	
3rd Qu.: 2.000	3rd Qu.: 4.00	3rd Qu.: 3.000	
Max. : 7.000	Max. : 10.00	Max. : 17.000	
travel_day_date	WRKCOUNT	household_stratum_id	person_weight_7_day
Length: 16997	Min. : 0.000	Length: 16997	Min. : 30.9
Class : character	1st Qu.: 0.000	Class : character	1st Qu.: 5614.8
Mode : character	Median : 1.000	Mode : character	Median : 10798.7
	Mean : 1.216		Mean : 17977.3
	3rd Qu.: 2.000		3rd Qu.: 21084.0
	Max. : 6.000		Max. : 781700.0
person_weight_5_day	person_weight_2_day	respondent_age	airport_size
Min. : 0	Min. : 0	Min. : 5.00	Length: 16997
1st Qu.: 0	1st Qu.: 0	1st Qu.: 28.00	Class : character
Median : 9142	Median : 0	Median : 48.00	Mode : character
Mean : 17977	Mean : 17977	Mean : 45.92	
3rd Qu.: 22271	3rd Qu.: 11093	3rd Qu.: 65.00	
Max. : 972351	Max. : 473082	Max. : 92.00	
begin_trip_date	household_grouping	census_division	census_region
Length: 16997	Length: 16997	Length: 16997	Length: 16997



Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

DRIVER	education_level	ENDTRIP	exit_census_division
Length:16997	Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

farthest_census_division_traveled	farthest_census_region_traveled
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

reason_for_travel	great_circle_distance_longer_than_50_miles
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

household_income	household_income_imputed	hispanic
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

race	own_or_rent_home	international_trip	long_distance_trip
Length:16997	Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

LIF_CYC	main_mode_of_transport	date_of_most_recent_trip
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character

Mode :character Mode :character Mode :character

metro_area_category	metro_area_size	person_1_on_long_distance_trip
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

person_10_on_long_distance_trip	person_2_on_long_distance_trip
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

person_3_on_long_distance_trip	person_4_on_long_distance_trip
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

person_5_on_long_distance_trip	person_6_on_long_distance_trip
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

person_7_on_long_distance_trip	person_8_on_long_distance_trip
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

person_9_on_long_distance_trip	proxy_respondent	metro_rail_status
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

```
respondent_hispanic respondent_race respondent_sex
Length:16997      Length:16997      Length:16997
Class :character   Class :character   Class :character
Mode :character    Mode :character    Mode :character
```

```
respondent_sex_imputed travel_day_of_week household_urban_classification
Length:16997           Length:16997           Length:16997
Class :character        Class :character        Class :character
Mode :character         Mode :character         Mode :character
```

```
household_urban_size household_urban_rural trip_includes_weekend
Length:16997           Length:16997           Length:16997
Class :character        Class :character        Class :character
Mode :character         Mode :character         Mode :character
```

```
worker_status
Length:16997
Class :character
Mode :character
```

```
# Corr plot
long_distance |>
  select(-household_id) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot(
    method = "color",
    type = "upper",
    tl.col = "black",
    tl.srt = 45,
```

```
addCoef.col = "black",
)
```



## Person Data

```
person |> glimpse()
```

```
Rows: 16,997
Columns: 129
$ household_id      <chr> "9000013002", "90000130~
$ person_id         <chr> "01", "02", "03", "04",~
$ person_weight_7_day <dbl> 3938.689, 5409.515, 301~
$ person_weight_5_day <dbl> 0.000, 0.000, 0.000, 0.~
$ person_weight_2_day <dbl> 13363.809, 19360.095, 7~
$ respondent_age     <dbl> 39, 42, 8, 5, 32, 41, 4~
$ great_circle_distance_work <dbl> -1.000000, -9.000000, --
$ days_used_taxi_last_30 <dbl> -1, -1, -1, -1, -1, -1,~
$ days_used_ride_share_last_30 <dbl> -1, -1, -1, -1, -1, -1,~
$ days_used_scooter_last_30 <dbl> -1, -1, -1, -1, -1, -1,~
$ days_used_public_transit_last_30 <dbl> -1, -1, -1, -1, -1, -1,~
```

\$ days_used_motorcycle_last_30	<dbl> -1, -1, -1, -1, -1, -1,~
\$ days_walked_as_transit_last_30	<dbl> 10, 3, -1, -1, 3, 4, 15~
\$ days_cycled_last_30	<dbl> -1, -1, -1, -1, -1, -1,~
\$ days_used_bikeshare_last_30	<dbl> -1, -1, -1, -1, -1, -1,~
\$ hours_worked_per_week	<dbl> -1, 55, -1, -1, 30, 40,~
\$ num_online_deliveries_last_30	<dbl> 10, 12, -1, -1, 10, 10,~
\$ num_goods_deliveries_last_30	<dbl> 10, 12, -1, -1, 8, 8, 6~
\$ num_food_deliveries_last_30	<dbl> 0, 0, -1, -1, 2, 2, 4, ~
\$ num_grocery_deliveries_last_30	<dbl> 0, 0, -1, -1, -9, -9, 1~
\$ num_services_deliveries_last_30	<dbl> 0, 0, -1, -1, -9, -9, --
\$ num_returned_home_pickup	<dbl> 0, 0, -1, -1, -9, -9, --
\$ num_returned_ups_usps_fedex	<dbl> 2, 2, -1, -1, -9, 3, -9~
\$ num_returned_amazon_dropoff_center	<dbl> 1, 0, -1, -1, 2, -9, 1,~
\$ num_returned_store	<dbl> 2, 1, -1, -1, 1, -9, -9~
\$ home_parking_cost	<dbl> -1, -1, -1, -1, -1, -1,~
\$ count_person_trips	<dbl> 2, 3, 0, 2, 3, 3, 0, 2,~
\$ num_adults_in_household	<dbl> 2, 2, 2, 2, 2, 2, 1, 2,~
\$ num_drivers_in_household	<dbl> 2, 2, 2, 2, 2, 2, 1, 2,~
\$ household_size	<dbl> 4, 4, 4, 4, 2, 2, 1, 4,~
\$ num_vehicles_in_household	<dbl> 2, 2, 2, 2, 1, 1, 0, 1,~
\$ travel_day_date	<chr> "202202", "202202", "20~
\$ WRKCOUNT	<dbl> 1, 1, 1, 1, 2, 2, 1, 1,~
\$ household_stratum_id	<chr> "1021", "1021", "1021",~
\$ household_grouping	<chr> "New England MSA/CMSA o~
\$ census_division	<chr> "New England", "New Eng~
\$ census_region	<chr> "Northeast", "Northeast~
\$ daytime_limit_driving	<chr> "Valid skip", "Valid sk~
\$ unaffected_travel	<chr> "Valid skip", "Valid sk~
\$ limited_transit_use	<chr> "Valid skip", "Valid sk~
\$ limited_transit_use_noresponse	<chr> "Valid skip", "Valid sk~
\$ requested_rides	<chr> "Valid skip", "Valid sk~
\$ limited_driving	<chr> "Valid skip", "Valid sk~
\$ requested_ride_shares	<chr> "Valid skip", "Valid sk~
\$ requires_special_transport	<chr> "Valid skip", "Valid sk~
\$ reduce_travel_disability	<chr> "Valid skip", "Valid sk~
\$ covid_online_delivery_impact	<chr> "Do the same as before"~
\$ covid_public_transit_impact	<chr> "Do less often than bef~
\$ covid_school_travel_impact	<chr> "Valid skip", "Valid sk~
\$ covid_work_travel_impact	<chr> "Valid skip", "Do less ~
\$ covid_online_delivery_impact_permant	<chr> "Valid skip", "Valid sk~
\$ covid_public_transit_impact_permant	<chr> "Temporary", "Don't Kno~
\$ covid_school_travel_impact_permant	<chr> "Valid skip", "Valid sk~
\$ covid_work_travel_impact_permant	<chr> "Valid skip", "Permanen~

\$ DRIVER  
 \$ drive\_for\_work  
 \$ DRIVINGVEHICLE  
 \$ education\_level  
 \$ employer\_pays\_transit  
 \$ started\_travel\_at\_home  
 \$ household\_income  
 \$ household\_income\_imputed  
 \$ hispanic  
 \$ race  
 \$ own\_or\_rent\_home  
 \$ used\_bike\_last\_30  
 \$ used\_bikeshare\_last\_30  
 \$ used\_e\_scooter\_last\_30  
 \$ used\_motorcycle\_last\_30  
 \$ used\_public\_transit\_last\_30  
 \$ used\_ride\_share\_last\_30  
 \$ used\_taxi\_last\_30  
 \$ walked\_last\_30  
 \$ LIF\_CYC  
 \$ medical\_condition  
 \$ medical\_condition\_duration  
 \$ metro\_area\_category  
 \$ metro\_area\_size  
 \$ away\_entire\_day  
 \$ paid\_for\_parking\_travel\_day  
 \$ pay\_home\_parking  
 \$ whether\_pay\_home\_parking  
 \$ home\_parking\_payment\_periodicity  
 \$ worked\_last\_week  
 \$ primary\_activity\_not\_work  
 \$ proxy\_respondent  
 \$ non\_english\_language\_spoken  
 \$ how\_well\_english\_spoken  
 \$ metro\_rail\_status  
 \$ respondent\_hispanic  
 \$ respondent\_race  
 \$ respondent\_race\_imputed  
 \$ respondent\_relationship  
 \$ respondent\_sex  
 \$ respondent\_sex\_imputed  
 \$ reason\_no\_travel  
 \$ enrolled\_in\_school

<chr> "Yes", "Yes", "Valid skip"  
 <chr> "Valid skip", "No", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "Master's degree", "Master's degree"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "Yes", "Yes", "Yes", "Yes"  
 <chr> "\$200,000 or more", "\$200,000 or more"  
 <chr> "\$200,000 or more", "\$200,000 or more"  
 <chr> "Not Hispanic or Latino"  
 <chr> "White", "White", "White"  
 <chr> "Owned by hh member with access"  
 <chr> "No", "No", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "No", "No", "Valid skip"  
 <chr> "No", "No", "Valid skip"  
 <chr> "No", "No", "Valid skip"  
 <chr> "No", "No", "Valid skip"  
 <chr> "No", "No", "Valid skip"  
 <chr> "Yes", "Yes", "Valid skip"  
 <chr> "2+ adults, youngest child"  
 <chr> "No", "No", "No", "No", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "MSA of 1 million or more"  
 <chr> "In an MSA or CMSA of 3-5 million"  
 <chr> "No", "No", "No", "No", "Valid skip"  
 <chr> "No", "Valid skip", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "No", "Yes", "Valid skip"  
 <chr> "A homemaker", "Valid skip"  
 <chr> "Self", "Someone else", "Valid skip"  
 <chr> "No", "No", "No", "No", "Valid skip"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "Yes", "Yes", "Yes", "Yes"  
 <chr> "Not Hispanic", "Not Hispanic"  
 <chr> "White", "White", "White"  
 <chr> "White", "White", "White"  
 <chr> "Self", "Spouse/Partner"  
 <chr> "Female", "Male", "Refused"  
 <chr> "Female", "Male", "Male"  
 <chr> "Valid skip", "Valid skip"  
 <chr> "No", "No", "Yes", "Yes"

```

$ school_type_nontraditional      <chr> "Valid skip", "Valid sk~
$ school_transportation          <chr> "Valid skip", "Valid sk~
$ school_type                    <chr> "Valid skip", "Valid sk~
$ school_program_description     <chr> "Valid skip", "Valid sk~
$ travel_day_of_week             <chr> "Sunday", "Sunday", "Su~
$ days_used_discounted_transit_last_30 <chr> "Valid skip", "Valid sk~
$ household_urban_classification <chr> "In an urban area", "In~
$ household_urban_size           <chr> "1,000,000 or more with~
$ household_urban_rural          <chr> "Urban", "Urban", "Urba~
$ fewer_trips_last_30            <chr> "No, did not take fewer~
$ fewer_trips_last_30_reason_more_deliveries <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_covid <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_not_feel_safe <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_not_feel_clean <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_not_reliable <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_not_go_where_needed <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_unaffordable <chr> "valid skip", "valid sk~
$ fewer_trips_last_30_reason_health <chr> "valid skip", "valid sk~
$ fewer_trips_last_30_reason_time <chr> "Valid skip", "Valid sk~
$ fewer_trips_last_30_reason_other <chr> "Valid skip", "Valid sk~
$ USEPUBTR                      <chr> "No", "No", "No", "No",~
$ who_completed_survey          <chr> "Valid skip", "Person1"~
$ days_per_week_work_from_home <chr> "Valid skip", "Five or ~
$ worker_status                 <chr> "Not worker", "Worker",~
$ work_location                 <chr> "Valid skip", "Only one~
$ work_transportation           <chr> "Valid skip", "Valid sk~
$ used_cane                     <chr> "Valid skip", "Valid sk~
$ used_wheelchair               <chr> "Valid skip", "Valid sk~
$ no_assistance                 <chr> "Valid skip", "Valid sk~
$ used_blind_deaf_assistance    <chr> "Valid skip", "Valid sk~
$ W_VISIMP                     <chr> "Valid skip", "Valid sk~
$ used_walker_crutches          <chr> "Valid skip", "Valid sk~

```

```
person |> summary()
```

household_id	person_id	person_weight_7_day	person_weight_5_day
Length:16997	Length:16997	Min. : 30.9	Min. : 0
Class :character	Class :character	1st Qu.: 5614.8	1st Qu.: 0
Mode :character	Mode :character	Median : 10798.7	Median : 9142
		Mean : 17977.3	Mean : 17977
		3rd Qu.: 21084.0	3rd Qu.: 22271
		Max. : 781700.0	Max. : 972351

person_weight_2_day	respondent_age	great_circle_distance_work
Min. : 0	Min. : 5.00	Min. : -9.000
1st Qu.: 0	1st Qu.:28.00	1st Qu.: -1.000
Median : 0	Median :48.00	Median : -1.000
Mean : 17977	Mean :45.92	Mean : 3.101
3rd Qu.: 11093	3rd Qu.:65.00	3rd Qu.: 2.717
Max. :473082	Max. :92.00	Max. :2543.954

days_used_taxi_last_30	days_used_ride_share_last_30	days_used_scooter_last_30
Min. :-1.0000	Min. :-1.0000	Min. :-1.0000
1st Qu.: -1.0000	1st Qu.: -1.0000	1st Qu.: -1.0000
Median : -1.0000	Median : -1.0000	Median : -1.0000
Mean : -0.8781	Mean : -0.4042	Mean : -0.9355
3rd Qu.: -1.0000	3rd Qu.: -1.0000	3rd Qu.: -1.0000
Max. :30.0000	Max. :30.0000	Max. :30.0000

days_used_public_transit_last_30	days_used_motorcycle_last_30
Min. :-1.0000	Min. :-1.0000
1st Qu.: -1.0000	1st Qu.: -1.0000
Median : -1.0000	Median : -1.0000
Mean : -0.2196	Mean : -0.8399
3rd Qu.: -1.0000	3rd Qu.: -1.0000
Max. :30.0000	Max. :30.0000

days_walked_as_transit_last_30	days_cycled_last_30	days_used_bikeshare_last_30
Min. :-1.000	Min. :-1.0000	Min. :-1.0000
1st Qu.: -1.000	1st Qu.: -1.0000	1st Qu.: -1.0000
Median : -1.000	Median : -1.0000	Median : -1.0000
Mean : 3.016	Mean : -0.3464	Mean : -0.9715
3rd Qu.: 3.000	3rd Qu.: -1.0000	3rd Qu.: -1.0000
Max. :30.000	Max. :30.0000	Max. :30.0000

hours_worked_per_week	num_online_deliveries_last_30
Min. :-1.00	Min. :-1.000
1st Qu.: -1.00	1st Qu.: 0.000
Median : -1.00	Median : 2.000
Mean :16.89	Mean : 4.758
3rd Qu.:40.00	3rd Qu.: 6.000
Max. :99.00	Max. :99.000

num_goods_deliveries_last_30	num_food_deliveries_last_30
Min. :-9.000	Min. :-9.000
1st Qu.: -1.000	1st Qu.: -9.000
Median : 2.000	Median : -1.000
Mean : 3.581	Mean : -2.213
3rd Qu.: 5.000	3rd Qu.: 0.000
Max. :99.000	Max. :59.000

num_grocery_deliveries_last_30	num_services_deliveries_last_30
--------------------------------	---------------------------------



Min. :-9.000	Min. :-9.000		
1st Qu.:-9.000	1st Qu.:-9.000		
Median :-1.000	Median :-1.000		
Mean :-2.842	Mean :-3.033		
3rd Qu.: 0.000	3rd Qu.: 0.000		
Max. :30.000	Max. :60.000		
num_returned_home_pickup	num_returned_ups_usps_fedex		
Min. :-9.000	Min. :-9.00		
1st Qu.:-1.000	1st Qu.:-1.00		
Median :-1.000	Median : 0.00		
Mean :-1.977	Mean :-1.31		
3rd Qu.: 0.000	3rd Qu.: 0.00		
Max. :60.000	Max. :47.00		
num_returned_amazon_dropoff_center	num_returned_store	home_parking_cost	
Min. :-9.000	Min. :-9.000	Min. : -1.0000	
1st Qu.:-1.000	1st Qu.:-1.000	1st Qu.: -1.0000	
Median :-1.000	Median :-1.000	Median : -1.0000	
Mean :-1.779	Mean :-1.922	Mean : -0.9344	
3rd Qu.: 0.000	3rd Qu.: 0.000	3rd Qu.: -1.0000	
Max. :31.000	Max. :17.000	Max. :200.0000	
count_person_trips	num_adults_in_household	num_drivers_in_household	
Min. : 0.000	Min. :1.000	Min. :0.000	
1st Qu.: 0.000	1st Qu.:2.000	1st Qu.:2.000	
Median : 2.000	Median :2.000	Median :2.000	
Mean : 1.828	Mean :2.135	Mean :1.998	
3rd Qu.: 3.000	3rd Qu.:2.000	3rd Qu.:2.000	
Max. :36.000	Max. :8.000	Max. :7.000	
household_size	num_vehicles_in_household	travel_day_date	WRKCOUNT
Min. : 1.00	Min. : 0.000	Length:16997	Min. :0.000
1st Qu.: 2.00	1st Qu.: 1.000	Class :character	1st Qu.:0.000
Median : 2.00	Median : 2.000	Mode :character	Median :1.000
Mean : 2.91	Mean : 2.098		Mean :1.216
3rd Qu.: 4.00	3rd Qu.: 3.000		3rd Qu.:2.000
Max. :10.00	Max. :17.000		Max. :6.000
household_stratum_id	household_grouping	census_division	census_region
Length:16997	Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
daytime_limit_driving	unaffected_travel	limited_transit_use	
Length:16997	Length:16997	Length:16997	

Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

limited_transit_use_noresponse	requested_rides	limited_driving
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

requested_ride_shares	requires_special_transport	reduce_travel_disability
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

covid_online_delivery_impact	covid_public_transit_impact
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

covid_school_travel_impact	covid_work_travel_impact
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

covid_online_delivery_impact_permanent	covid_public_transit_impact_permanent
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

covid_school_travel_impact_permanent	covid_work_travel_impact_permanent
Length:16997	Length:16997
Class :character	Class :character

Mode :character

Mode :character

DRIVER	drive_for_work	DRIVINGVEHICLE	education_level
Length:16997	Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

employer_pays_transit	started_travel_at_home	household_income
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

household_income_imputed	hispanic	race
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

own_or_rent_home	used_bike_last_30	used_bikeshare_last_30
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

used_e_scooter_last_30	used_motorcycle_last_30	used_public_transit_last_30
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

used_ride_share_last_30	used_taxi_last_30	walked_last_30
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

LIF_CYC	medical_condition	medical_condition_duration
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

metro_area_category	metro_area_size	away_entire_day
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

paid_for_parking_travel_day	pay_home_parking	whether_pay_home_parking
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

home_parking_payment_periodicity	worked_last_week	primary_activity_not_work
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

proxy_respondent	non_english_language_spoken	how_well_english_spoken
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

metro_rail_status	respondent_hispanic	respondent_race
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

respondent_race_imputed	respondent_relationship	respondent_sex
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

respondent_sex_imputed	reason_no_travel	enrolled_in_school
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

school_type_nontraditional	school_transportation	school_type
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

school_program_description	travel_day_of_week
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

days_used_discounted_transit_last_30	household_urban_classification
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

household_urban_size	household_urban_rural	fewer_trips_last_30
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

fewer_trips_last_30_reason_more_deliveries	fewer_trips_last_30_reason_covid
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

fewer\_trips\_last\_30\_reason\_not\_feel\_safe  
Length:16997  
Class :character  
Mode :character

fewer\_trips\_last\_30\_reason\_not\_feel\_clean  
Length:16997  
Class :character  
Mode :character

fewer\_trips\_last\_30\_reason\_not\_reliable  
Length:16997  
Class :character  
Mode :character

fewer\_trips\_last\_30\_reason\_not\_go\_where\_needed  
Length:16997  
Class :character  
Mode :character

fewer_trips_last_30_reason_unaffordable	fewer_trips_last_30_reason_health
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

fewer_trips_last_30_reason_time	fewer_trips_last_30_reason_other
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

USEPUBTR	who_completed_survey	days_per_week_work_from_home
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

worker_status	work_location	work_transportation	used_cane
Length:16997	Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

used_wheelchair	no_assistance	used_blind_deaf_assistance
Length:16997	Length:16997	Length:16997
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

W_VISIMP	used_walker_crutches
Length:16997	Length:16997
Class :character	Class :character
Mode :character	Mode :character

```
# Corr plot
person |>
  select(-household_id) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot(
```

```

method = "color",
type = "upper",
tl.col = "black",
tl.srt = 45,
addCoef.col = "black",
)

```

Warning in `corrplot::corrplot(cor(select(select(person, -household_id), : Not been able to calculate text margin, please try again with a clean new empty window using {plot.new(); dev.off()} or reduce tl.cex`



## Trip Data

```
trip |> glimpse()
```

Rows: 31,074

Columns: 85

\$ household\_id

<chr> "9000013002", "9000013002", "9000013~

\$ person\_id

<chr> "01", "01", "02", "02", "02", "04", ~



\$ trip_id	<chr> "01", "02", "01", "02", "03", "01", ~
\$ sequential_trip_id	<chr> "01", "02", "01", "02", "03", "01", ~
\$ time_at_destination	<dbl> 75, -9, 10, 10, -9, 75, -9, 260, 105~
\$ trip_duration_minutes	<dbl> 15, 10, 30, 10, 30, 15, 10, 10, 15, ~
\$ num_people_on_trip	<dbl> 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 2, ~
\$ num_non_household_members_on_trip	<dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
\$ num_household_members_on_trip	<chr> "02", "02", "01", "01", "01", "02", ~
\$ parking_cost	<dbl> -1, -1, -1, -1, -1, -1, -1, -1, -1, ~
\$ trip_miles	<dbl> 3.9024239, 3.9024239, 17.0770665, 4.~
\$ trip_weight_7_day	<dbl> 1608362, 1608362, 2208973, 2208973, ~
\$ trip_weight_5_day	<dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1~
\$ trip_weight_2_day	<dbl> 1569852, 1569852, 2274239, 2274239, ~
\$ trip_case_id	<chr> "90000130020101", "90000130020102", ~
\$ vehicle_miles_traveled	<dbl> 3.9024239, 3.9024239, 17.0770665, 4.~
\$ gas_price_at_travel_time	<chr> "355.1", "355.1", "355.1", "355.1", ~
\$ num_adults_in_household	<dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
\$ num_drivers_in_household	<dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
\$ household_size	<dbl> 4, 4, 4, 4, 4, 4, 4, 2, 2, 2, 2, ~
\$ num_vehicles_in_household	<dbl> 2, 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, ~
\$ travel_day_date	<chr> "202202", "202202", "202202", "20220~
\$ WRKCOUNT	<dbl> 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, ~
\$ household_stratum_id	<chr> "1021", "1021", "1021", "1021", "102~
\$ respondent_age	<dbl> 39, 39, 42, 42, 42, 5, 5, 32, 32, 32~
\$ household_grouping	<chr> "New England MSA/CMSA of 1 million+ ~
\$ census_division	<chr> "New England", "New England", "New E~
\$ census_region	<chr> "Northeast", "Northeast", "Northeast~
\$ DRIVER	<chr> "Yes", "Yes", "Yes", "Yes", "Yes", "~
\$ driver_flag	<chr> "Driver on trip", "Driver on trip", ~
\$ education_level	<chr> "Master's degree", "Master's degree"~
\$ end_trip_time	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ started_travel_at_home	<chr> "Yes", "Yes", "Yes", "Yes", "Yes", "~
\$ household_income	<chr> "\$200,000 or more", "\$200,000 or mor~
\$ household_income_imputed	<chr> "\$200,000 or more", "\$200,000 or mor~
\$ drove_on_trip	<chr> "Household member drove", "Household~
\$ hispanic	<chr> "Not Hispanic or Latino", "Not Hispa~
\$ race	<chr> "White", "White", "White", "White", ~
\$ own_or_rent_home	<chr> "Owned by hh member with mortgage or~
\$ LIF_CYC	<chr> "2+ adults, youngest child 0-5", "2+~
\$ loop_trip	<chr> "Not a loop trip", "Not a loop trip"~
\$ metro_area_category	<chr> "MSA of 1 million or more, with rail~
\$ metro_area_size	<chr> "In an MSA or CMSA of 3 million or m~
\$ person_1_on_trip	<chr> "Selected", "Selected", "Not Selecte~
\$ person_10_on_trip	<chr> "Valid skip", "Valid skip", "Valid s~

\$ person_2_on_trip	<chr> "Not Selected", "Not Selected", "Sel~
\$ person_3_on_trip	<chr> "Not Selected", "Not Selected", "Not~
\$ person_4_on_trip	<chr> "Selected", "Selected", "Not Selecte~
\$ person_5_on_trip	<chr> "Valid skip", "Valid skip", "Valid s~
\$ person_6_on_trip	<chr> "Valid skip", "Valid skip", "Valid s~
\$ person_7_on_trip	<chr> "Valid skip", "Valid skip", "Valid s~
\$ person_8_on_trip	<chr> "Valid skip", "Valid skip", "Valid s~
\$ person_9_on_trip	<chr> "Valid skip", "Valid skip", "Valid s~
\$ paid_for_parking_travel_day	<chr> "No", "No", "Valid skip", "Valid ski~
\$ paid_for_parking_trip	<chr> "Valid skip", "Valid skip", "Valid s~
\$ parking_payment_periodicity	<chr> "Valid skip", "Valid skip", "Valid s~
\$ primary_activity_not_work	<chr> "A homemaker", "A homemaker", "Valid~
\$ proxy_respondent	<chr> "Self", "Self", "Someone else", "Som~
\$ PSGR_FLG	<chr> "Not passenger on trip", "Not passen~
\$ public_transit_used_on_trip	<chr> "Did not use public transit", "Did n~
\$ metro_rail_status	<chr> "Yes", "Yes", "Yes", "Yes", "Yes", "~
\$ respondent_hispanic	<chr> "Not Hispanic", "Not Hispanic", "Not~
\$ respondent_race	<chr> "White", "White", "White", "White", ~
\$ respondent_sex	<chr> "Female", "Female", "Male", "Male", ~
\$ respondent_sex_imputed	<chr> "Female", "Female", "Male", "Male", ~
\$ start_trip_time	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ weekend_trip	<chr> "Weekend", "Weekend", "Weekend", "We~
\$ travel_day_of_week	<chr> "Sunday", "Sunday", "Sunday", "Sunda~
\$ trip_purpose	<chr> "Home-based social/recreational (HBS~
\$ vehicle_used_on_trip	<chr> "Yes, used a household vehicle", "Ye~
\$ trip_mode	<chr> "SUV/Crossover", "SUV/Crossover", "S~
\$ household_urban_classification	<chr> "In an urban area", "In an urban are~
\$ household_urban_size	<chr> "1,000,000 or more with heavy rail",~
\$ household_urban_rural	<chr> "Urban", "Urban", "Urban", "Urban", ~
\$ vehicle_case_id	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ vehicle_id	<chr> "", "", "", "", "", "", "", "", "", ~
\$ vehicle_type	<chr> "SUV (Santa Fe, Tahoe, Jeep, etc.)",~
\$ minutes_walked_from_parking	<chr> "1 to less than 5 minutes", "0 minut~
\$ who_drove	<chr> "Person 1", "Person 1", "Person 2", ~
\$ who_drove_imputed	<chr> "Person 1", "Person 1", "Person 2", ~
\$ reason_for_travel	<chr> "Regular activities at home", "Recre~
\$ reason_for_travel_to	<chr> "Recreational activities", "Regular ~
\$ why_trip	<chr> "Social/Recreational", "Home", "Meal~
\$ trip_purpose_old_schema	<chr> "Other Social/Recreational", "Other ~
\$ worker_status	<chr> "Not worker", "Not worker", "Worker"~

```
trip |> summary()
```

household_id	person_id	trip_id	sequential_trip_id
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

time_at_destination	trip_duration_minutes	num_people_on_trip
Min. : -9.00	Min. : -9.00	Min. : 1.000
1st Qu.: -9.00	1st Qu.: 10.00	1st Qu.: 1.000
Median : 20.00	Median : 15.00	Median : 1.000
Mean : 95.18	Mean : 24.55	Mean : 1.997
3rd Qu.: 110.00	3rd Qu.: 30.00	3rd Qu.: 2.000
Max. :1050.00	Max. :1425.00	Max. :99.000

num_non_household_members_on_trip	num_household_members_on_trip
Min. : 0.0000	Length:31074
1st Qu.: 0.0000	Class :character
Median : 0.0000	Mode :character
Mean : 0.4141	
3rd Qu.: 0.0000	
Max. :98.0000	

parking_cost	trip_miles	trip_weight_7_day	trip_weight_5_day
Min. : -1.0000	Min. : -9.000	Min. : 33695	Min. : 0
1st Qu.: -1.0000	1st Qu.: 1.759	1st Qu.: 2694005	1st Qu.: 0
Median : -1.0000	Median : 4.280	Median : 4919086	Median : 3173995
Mean : -0.8029	Mean : 13.965	Mean : 8166137	Mean : 5815804
3rd Qu.: -1.0000	3rd Qu.: 10.422	3rd Qu.: 9468638	3rd Qu.: 7071783
Max. :900.0000	Max. :4859.477	Max. :322721156	Max. :285950601

trip_weight_2_day	trip_case_id	vehicle_miles_traveled
Min. : 0	Length:31074	Min. : -9.000
1st Qu.: 0	Class :character	1st Qu.: -1.000
Median : 0	Mode :character	Median : 2.094
Mean : 2302368		Mean : 7.527
3rd Qu.: 797350		3rd Qu.: 7.129
Max. :53643494		Max. :1682.699

gas_price_at_travel_time	num_adults_in_household	num_drivers_in_household
Length:31074	Min. :1.000	Min. :0.000
Class :character	1st Qu.:2.000	1st Qu.:2.000
Mode :character	Median :2.000	Median :2.000
	Mean :2.059	Mean :1.993

	3rd Qu.:2.000	3rd Qu.:2.000	
	Max. :8.000	Max. :7.000	
household_size	num_vehicles_in_household	travel_day_date	WRKCOUNT
Min. : 1.000	Min. : 0.000	Length:31074	Min. :0.000
1st Qu.: 2.000	1st Qu.: 1.000	Class :character	1st Qu.:1.000
Median : 2.000	Median : 2.000	Mode :character	Median :1.000
Mean : 2.822	Mean : 2.134		Mean :1.304
3rd Qu.: 4.000	3rd Qu.: 3.000		3rd Qu.:2.000
Max. :10.000	Max. :11.000		Max. :6.000
household_stratum_id	respondent_age	household_grouping	census_division
Length:31074	Min. : 5.0	Length:31074	Length:31074
Class :character	1st Qu.:32.0	Class :character	Class :character
Mode :character	Median :49.0	Mode :character	Mode :character
	Mean :46.8		
	3rd Qu.:64.0		
	Max. :92.0		
census_region	DRIVER	driver_flag	education_level
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
end_trip_time	started_travel_at_home	household_income	
Length:31074	Length:31074	Length:31074	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
household_income_imputed	drove_on_trip	hispanic	
Length:31074	Length:31074	Length:31074	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
race	own_or_rent_home	LIF_CYC	loop_trip
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

metro_area_category	metro_area_size	person_1_on_trip	person_10_on_trip
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

person_2_on_trip	person_3_on_trip	person_4_on_trip	person_5_on_trip
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

person_6_on_trip	person_7_on_trip	person_8_on_trip	person_9_on_trip
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

paid_for_parking_travel_day	paid_for_parking_trip	parking_payment_periodicity
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

primary_activity_not_work	proxy_respondent	PSGR_FLG
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

public_transit_used_on_trip	metro_rail_status	respondent_hispanic
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

respondent_race	respondent_sex	respondent_sex_imputed
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

start_trip_time	weekend_trip	travel_day_of_week	trip_purpose
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

vehicle_used_on_trip	trip_mode	household_urban_classification
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

household_urban_size	household_urban_rural	vehicle_case_id
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

vehicle_id	vehicle_type	minutes_walked_from_parking
Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

who_drove	who_drove_imputed	reason_for_travel	reason_for_travel_to
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

why_trip	trip_purpose_old_schema	worker_status
----------	-------------------------	---------------

Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

```
# Corr plot
trip |>
  select(-household_id) |>
  select(where(is.numeric)) |>
  cor() |>
  ## ggplot style
  corrplot::corrplot(
    method = "color",
    type = "upper",
    tl.col = "black",
    tl.srt = 45,
    addCoef.col = "black",
  )
```

Warning in corrplot::corrplot(cor(select(select(trip, -household\_id),  
where(is.numeric))), : Not been able to calculate text margin, please try again  
with a clean new empty window using {plot.new(); dev.off()} or reduce tl.cex





```

$ household_weight_2_day      <dbl> 13946.19, 13946.19, 0.00, 0.00, ~
$ household_grouping          <chr> "New England MSA/CMSA of 1 milli~
$ census_division              <chr> "New England", "New England", "N~
$ census_region                <chr> "Northeast", "Northeast", "North~
$ commercial_frequency_of_veh_last_30 <chr> "Appropriate skip", "Appropriate~
$ household_income             <chr> "$200,000 or more", "$200,000 or~
$ household_income_imputed     <chr> "$200,000 or more", "$200,000 or~
$ days_vehicle_used_deliveries_last_30 <chr> "Appropriate skip", "Appropriate~
$ days_vehicle_used_other_last_30   <chr> "Appropriate skip", "Appropriate~
$ days_vehicle_used_ride_shares_last_30 <chr> "Appropriate skip", "Appropriate~
$ hispanic                     <chr> "Not Hispanic or Latino", "Not H~
$ race                         <chr> "White", "White", "White", "Whit~
$ own_or_rent_home             <chr> "Owned by hh member with mortgag~
$ hybrid_vehicle               <chr> "No", "No", "No", "No", "No", "N~
$ LIF_CYC                     <chr> "2+ adults, youngest child 0-5",~
$ vehicle_make                 <chr> "Subaru", "Subaru", "Toyota", "N~
$ metro_area_category          <chr> "MSA of 1 million or more, with ~
$ metro_area_size              <chr> "In an MSA or CMSA of 3 million ~
$ metro_rail_status            <chr> "Yes", "Yes", "No", "Yes", "No",~
$ travel_day_of_week           <chr> "Sunday", "Sunday", "Wednesday",~
$ household_urban_classification <chr> "In an urban area", "In an urban~
$ household_urban_size         <chr> "1,000,000 or more with heavy ra~
$ household_urban_rural        <chr> "Urban", "Urban", "Urban", "Urba~
$ vehicle_used_for_commercial_purposes <chr> "No", "No", "No", "No", "No", "N~
$ vehicle_used_for_deliveries   <chr> "Valid skip", "Valid skip", "Val~
$ vehicle_used_for_other        <chr> "Valid skip", "Valid skip", "Val~
$ vehicle_used_for_ride_shares  <chr> "Valid skip", "Valid skip", "Val~
$ vehicle_fuel_type             <chr> "Gas", "Gas", "Gas", "Gas", "Gas~
$ vehicle_owned_more_1_year     <chr> "Yes", "Yes", "Yes", "Yes", "Yes~
$ vehicle_owned_months         <chr> "Appropriate skip", "Appropriate~
$ vehicle_type                  <chr> "SUV (Santa Fe, Tahoe, Jeep, etc~
$ main_driver                   <chr> "Person1", "Person 2", "Person1"~

```

```
vehicle |> summary()
```

household_id	vehicle_id	vehicle_year	num_vehicles_in_household
Length:14684	Length:14684	Min. :1982	Min. : 1.000
Class :character	Class :character	1st Qu.:2007	1st Qu.: 2.000
Mode :character	Mode :character	Median :2014	Median : 2.000
		Mean :2012	Mean : 2.534
		3rd Qu.:2018	3rd Qu.: 3.000
		Max. :2023	Max. :17.000

vehicle_case_id	annual_miles	vehicle_age	num_adults_in_household
Length:14684	Min. : -9	Min. : 1.00	Min. :1.000
Class :character	1st Qu.: 2000	1st Qu.: 4.00	1st Qu.:2.000
Mode :character	Median : 6000	Median : 8.00	Median :2.000
	Mean : 12821	Mean :10.17	Mean :2.061
	3rd Qu.: 12000	3rd Qu.:15.00	3rd Qu.:2.000
	Max. :200000	Max. :40.00	Max. :8.000

num_drivers_in_household	household_size	travel_day_date	WRKCOUNT
Min. :0.000	Min. : 1.000	Length:14684	Min. :0.000
1st Qu.:2.000	1st Qu.: 2.000	Class :character	1st Qu.:0.000
Median :2.000	Median : 2.000	Mode :character	Median :1.000
Mean :2.014	Mean : 2.558		Mean :1.179
3rd Qu.:2.000	3rd Qu.: 3.000		3rd Qu.:2.000
Max. :7.000	Max. :10.000		Max. :6.000

household_stratum_id	household_weight_7_day	household_weight_5_day
Length:14684	Min. : 100.8	Min. : 0
Class :character	1st Qu.: 6611.2	1st Qu.: 0
Mode :character	Median :11437.4	Median : 10593
	Mean :15856.5	Mean : 15821
	3rd Qu.:20122.9	3rd Qu.: 22007
	Max. :81832.0	Max. :112970

household_weight_2_day	household_grouping	census_division
Min. : 0	Length:14684	Length:14684
1st Qu.: 0	Class :character	Class :character
Median : 0	Mode :character	Mode :character
Mean : 15955		
3rd Qu.: 15189		
Max. :264588		

census_region	commercial_frequency_of_veh_last_30	household_income
Length:14684	Length:14684	Length:14684
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

household_income_imputed	days_vehicle_used_deliveries_last_30
Length:14684	Length:14684
Class :character	Class :character
Mode :character	Mode :character

days_vehicle_used_other_last_30	days_vehicle_used_ride_shares_last_30
---------------------------------	---------------------------------------

Length:14684  
Class :character  
Mode :character

Length:14684  
Class :character  
Mode :character

hispanic	race	own_or_rent_home	hybrid_vehicle
Length:14684	Length:14684	Length:14684	Length:14684
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

LIF_CYC	vehicle_make	metro_area_category	metro_area_size
Length:14684	Length:14684	Length:14684	Length:14684
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

metro_rail_status	travel_day_of_week	household_urban_classification
Length:14684	Length:14684	Length:14684
Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character

household_urban_size	household_urban_rural
Length:14684	Length:14684
Class :character	Class :character
Mode :character	Mode :character

vehicle_used_for_commercial_purposes	vehicle_used_for_deliveries
Length:14684	Length:14684
Class :character	Class :character
Mode :character	Mode :character

vehicle_used_for_other	vehicle_used_for_ride_shares	vehicle_fuel_type
Length:14684	Length:14684	Length:14684

```
Class :character      Class :character      Class :character
Mode  :character      Mode  :character      Mode  :character
```

```
vehicle_owned_more_1_year vehicle_owned_months vehicle_type
Length:14684              Length:14684          Length:14684
Class :character          Class :character    Class :character
Mode  :character          Mode  :character    Mode  :character
```

```
main_driver
Length:14684
Class :character
Mode  :character
```

```
# Corr plot
vehicle |>
  select(-household_id) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot(
    method = "color",
    type = "upper",
    tl.col = "black",
    tl.srt = 45,
    addCoef.col = "black",
  )
```

