

Exploratory Data Analysis

DSAN 5300

Marion Bauman

Table of contents

Data Exploration	2
Household Data	2
Person Data	23
Trip Data	92
Vehicle Data	136
Long Distance Data	141

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 becom
```

```
library(arrow)
```

Attaching package: 'arrow'

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

```
duration
```

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

```
timestamp
```

```
library(GGally)
```

```
Registered S3 method overwritten by 'GGally':
```

```
method from  
+.gg ggplot2
```

```
# 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 <- 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")
```

Data Exploration

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, ~
```

```

$ num_adults_in_household      <dbl> 2, 2, 1, 2, 1, 2, 2, 2, 1, 2, 2, ~
$ num_drivers_in_household     <dbl> 2, 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, 0, 2, 3, ~
$ num_children_5_17              <dbl> 2, 0, 0, 2, 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                <date> 2022-02-01, 2022-02-01, 2022-02-01, 20~
$ 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
Min. : 2022-01-01	Min. : 0.0000	Length:7893 Length:7893
1st Qu.: 2022-04-01	1st Qu.: 0.0000	Class :character Class :character
Median : 2022-08-01	Median : 1.0000	Mode :character Mode :character
Mean : 2022-07-14	Mean : 0.9966	
3rd Qu.: 2022-11-01	3rd Qu.: 2.0000	
Max. : 2023-01-01	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

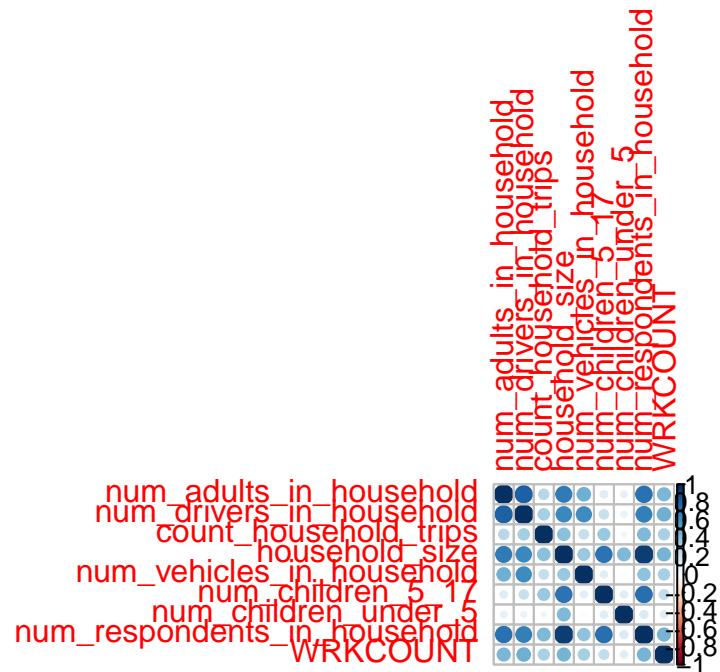
LIF_CYC           metro_area_category metro_area_size    metro_rail_status
Length:7893       Length:7893        Length:7893       Length:7893
Class :character  Class :character  Class :character  Class :character
Mode  :character  Mode  :character  Mode  :character  Mode  :character

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

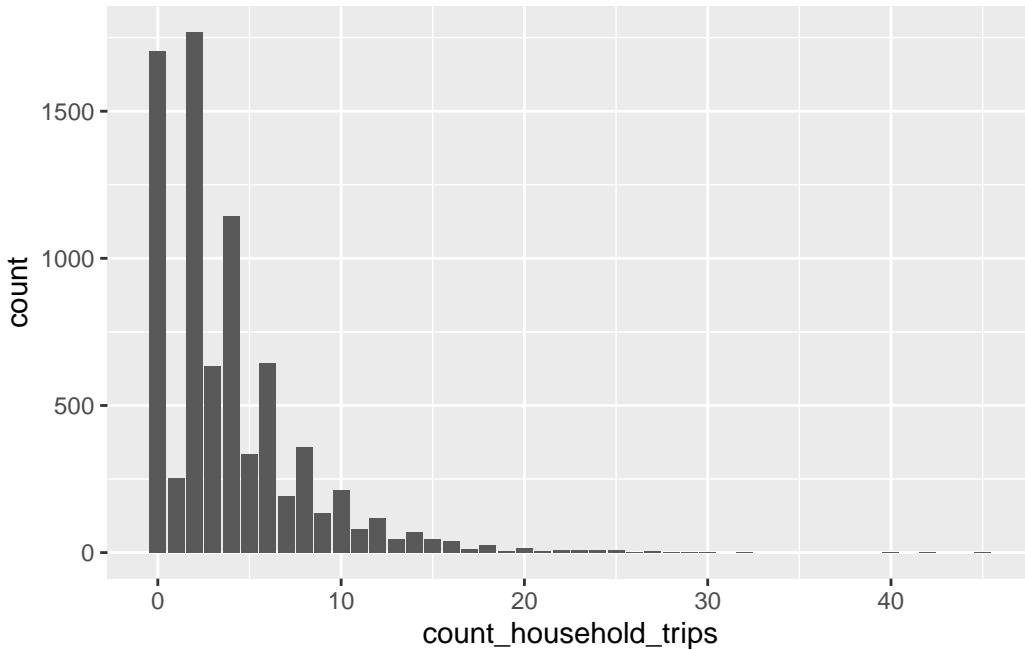
household_urban_rural household_urban_rural_2010
Length:7893          Length:7893
Class :character      Class :character
Mode  :character      Mode  :character

household |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot()

```



```
household |>
  ggplot(aes(count_household_trips)) +
  geom_bar()
```

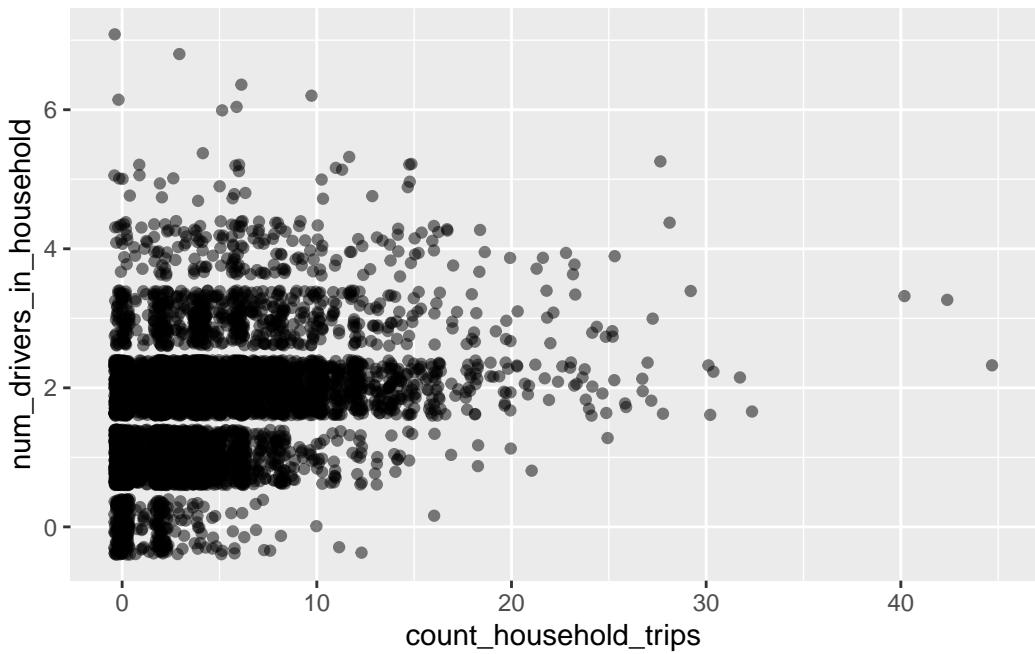
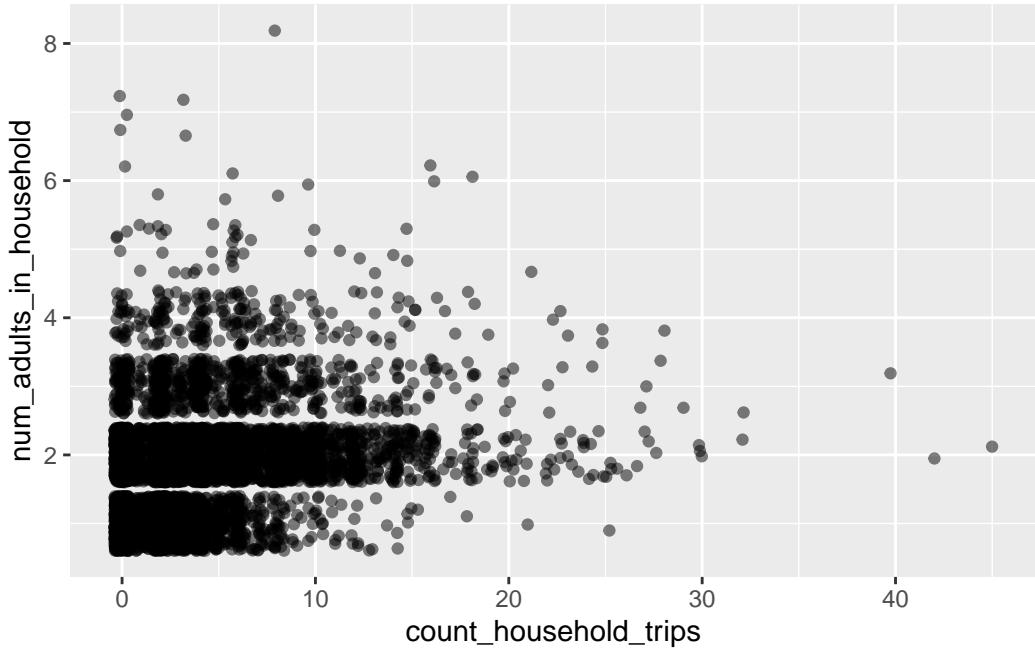


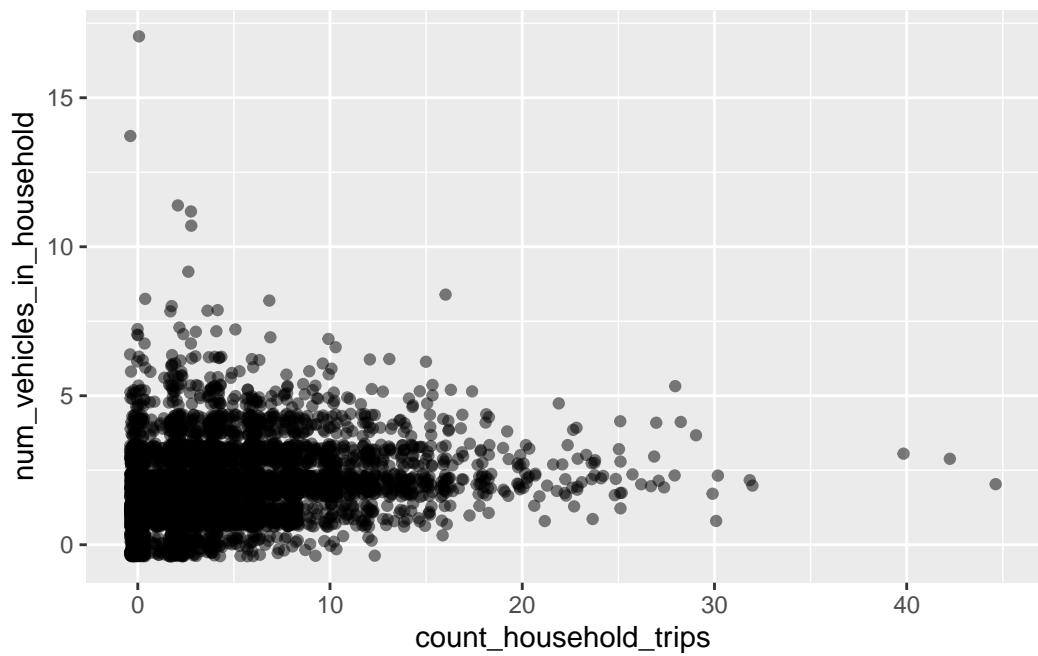
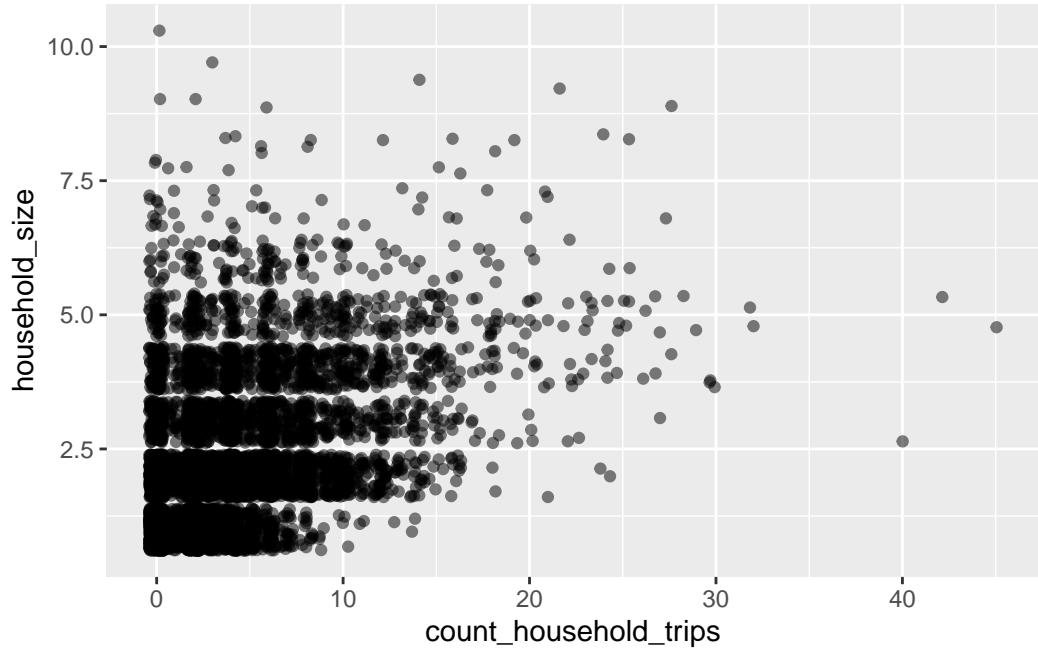
```

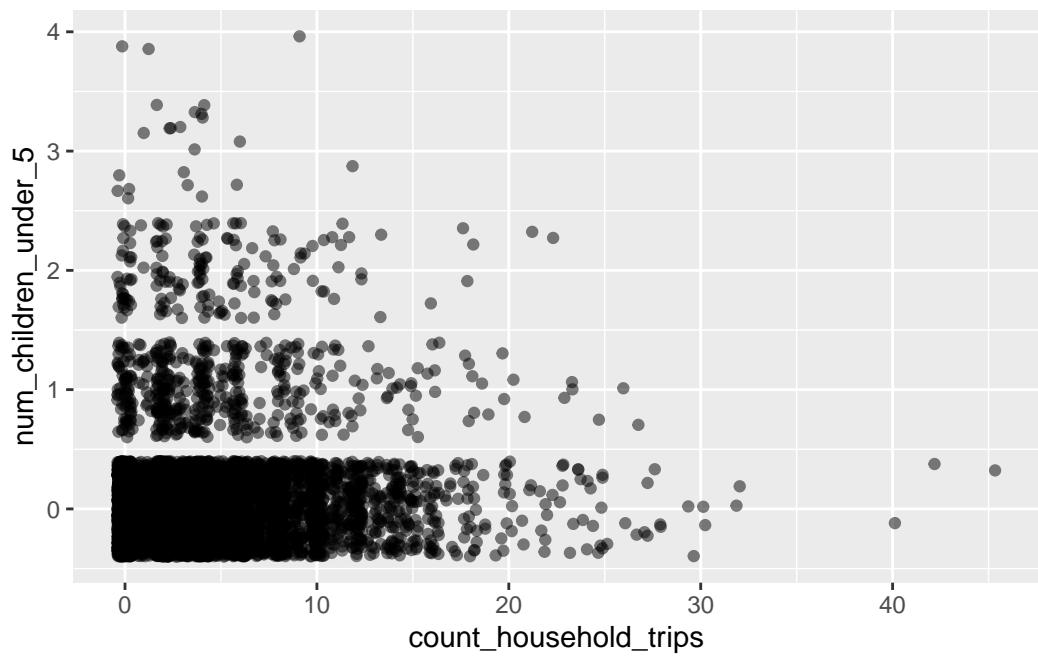
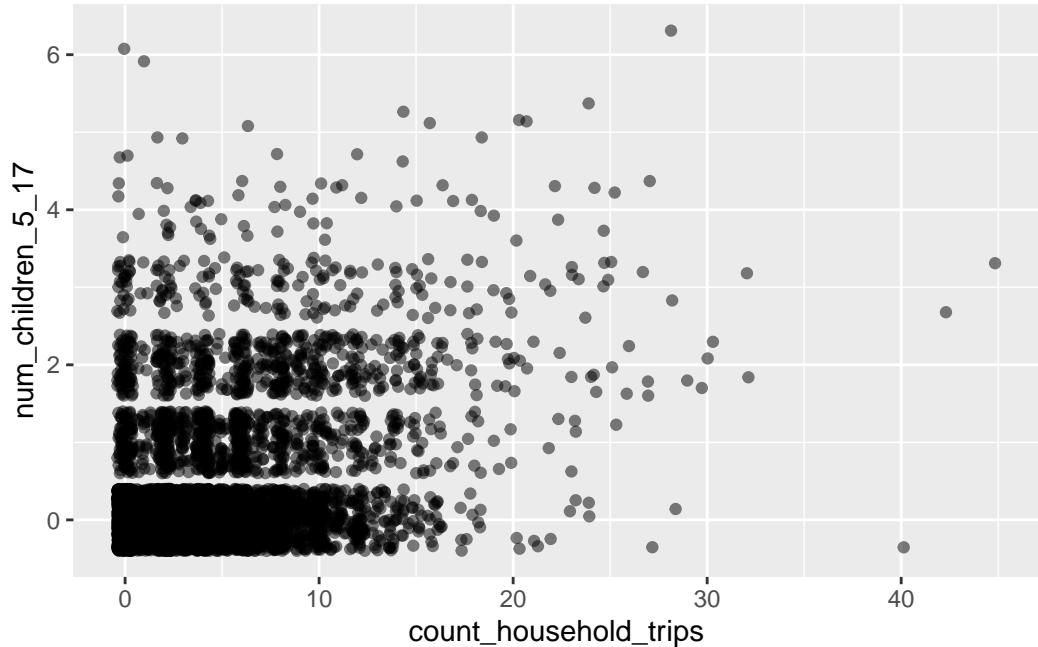
# Plot count_household_trips by all variables
household |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  colnames() -> house_cols

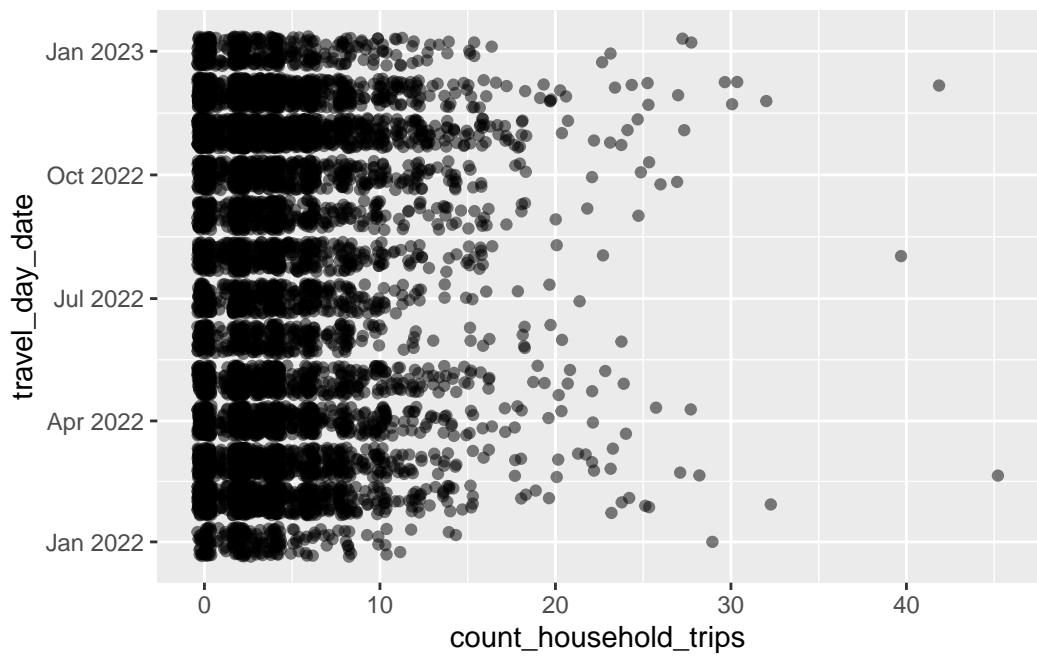
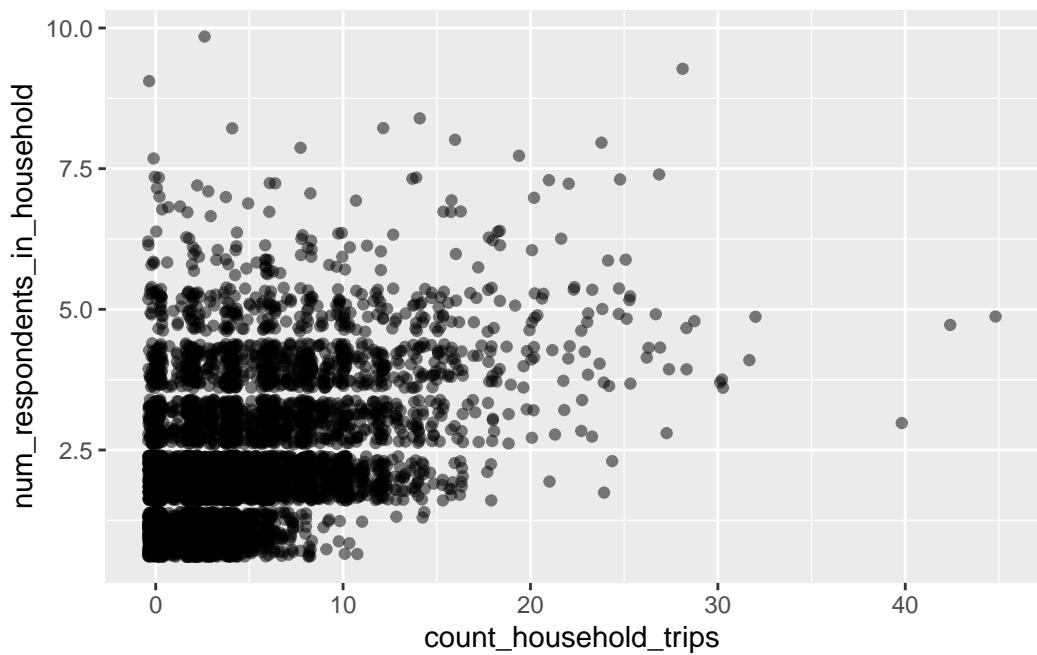
for (col in house_cols) {
  if (is.character(household[[col]])) {
    p <- household |>
      ggplot(aes(x = count_household_trips, fill = .data[[col]])) +
      geom_density() +
      facet_wrap(~.data[[col]])
    print(p)
  } else if (col == "count_household_trips") {
    next
  } else {
    p <- household |>
      ggplot(aes(x = count_household_trips, y = .data[[col]])) +
      geom_jitter(alpha = 0.5)
    print(p)
  }
}

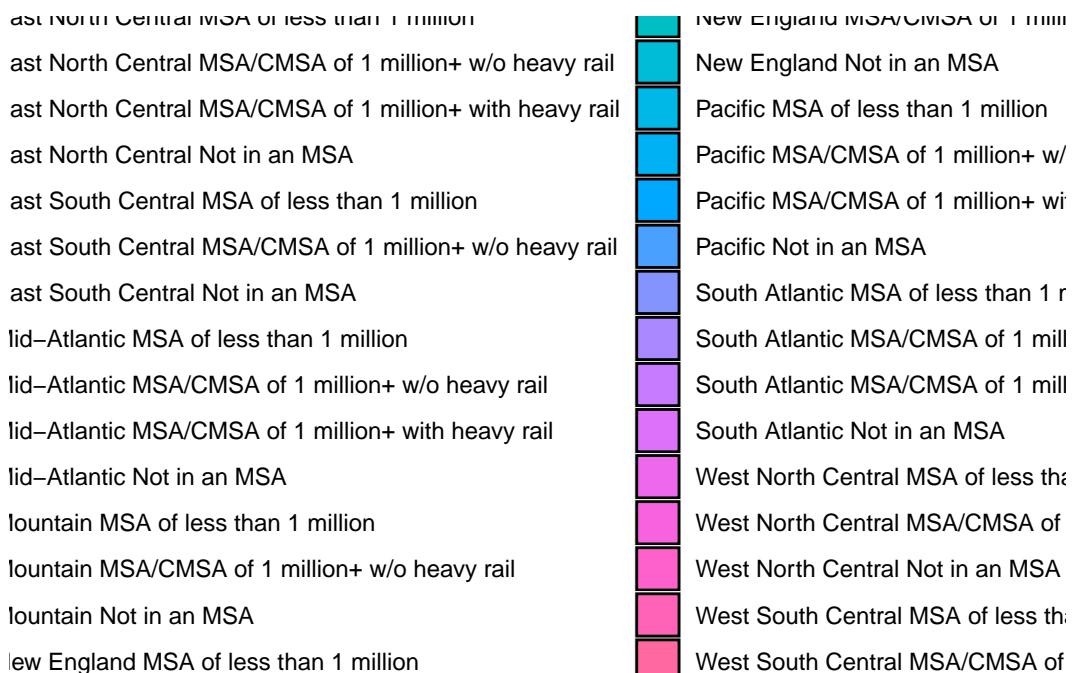
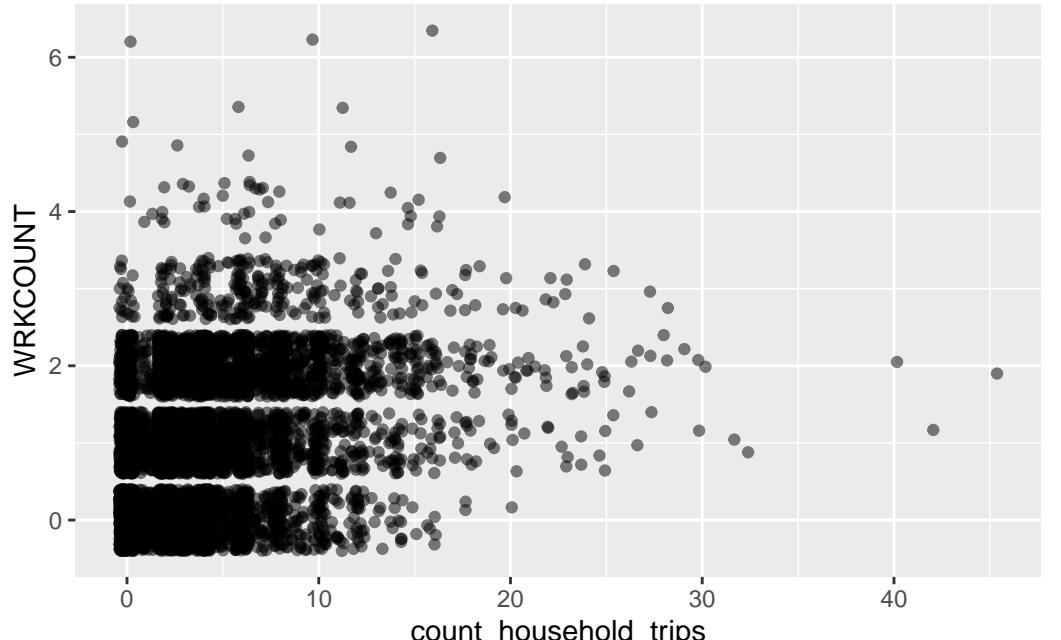
```

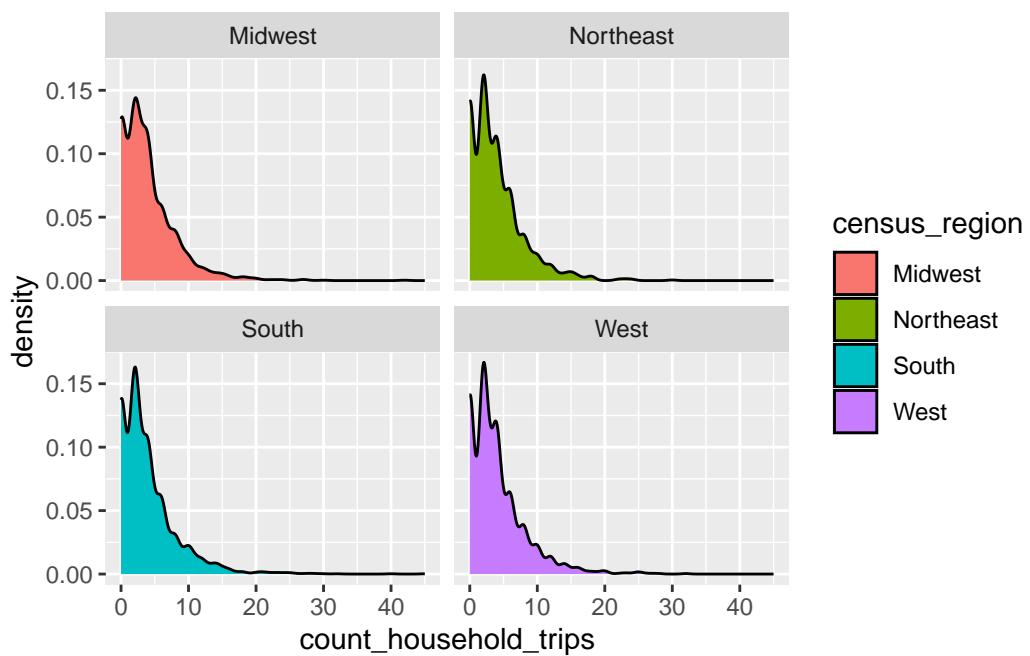
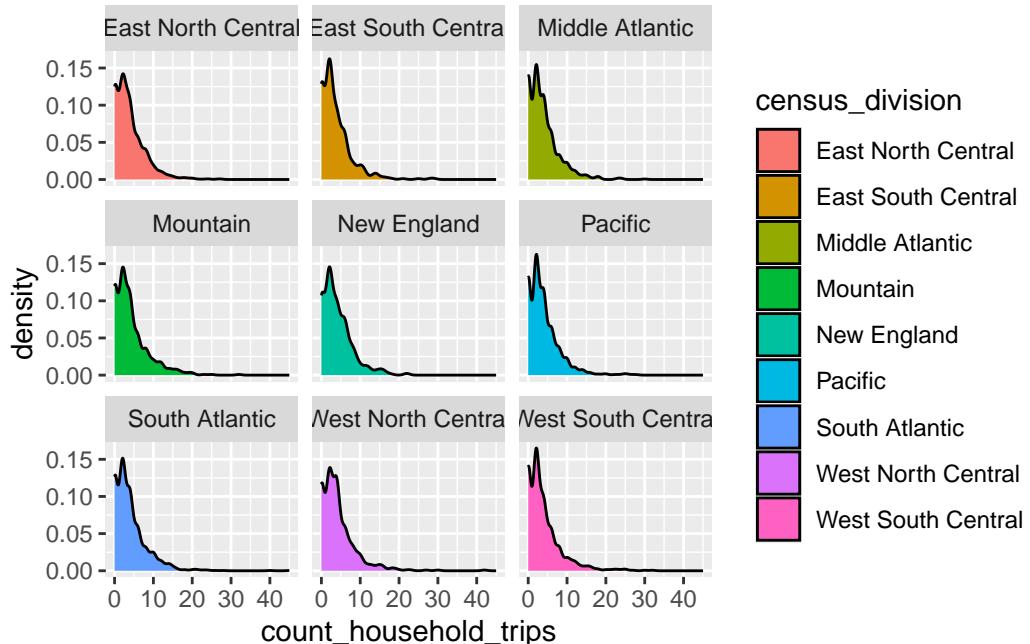


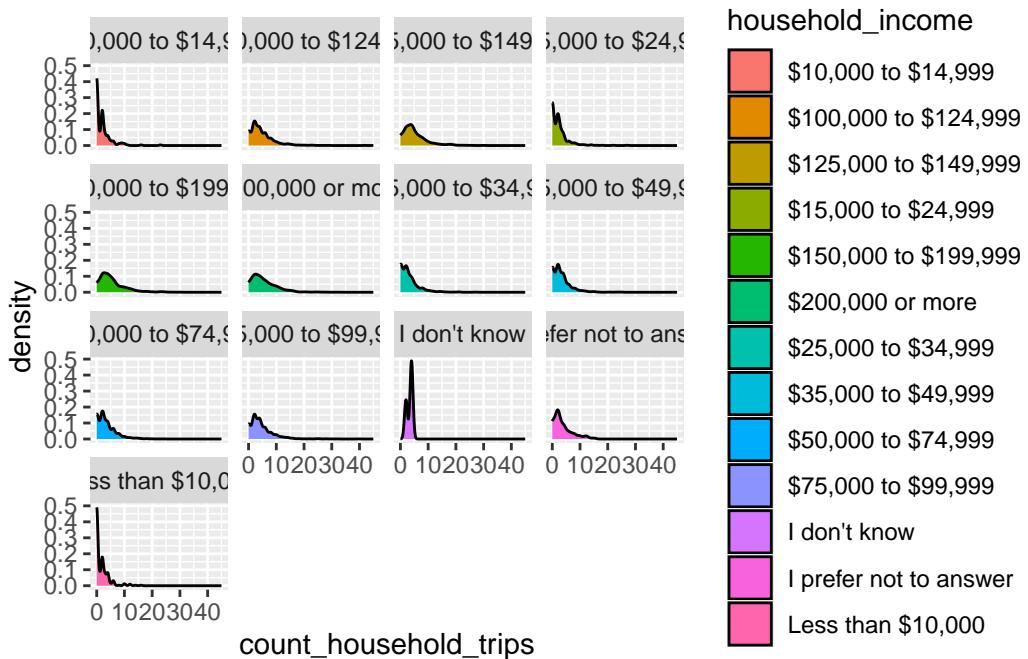
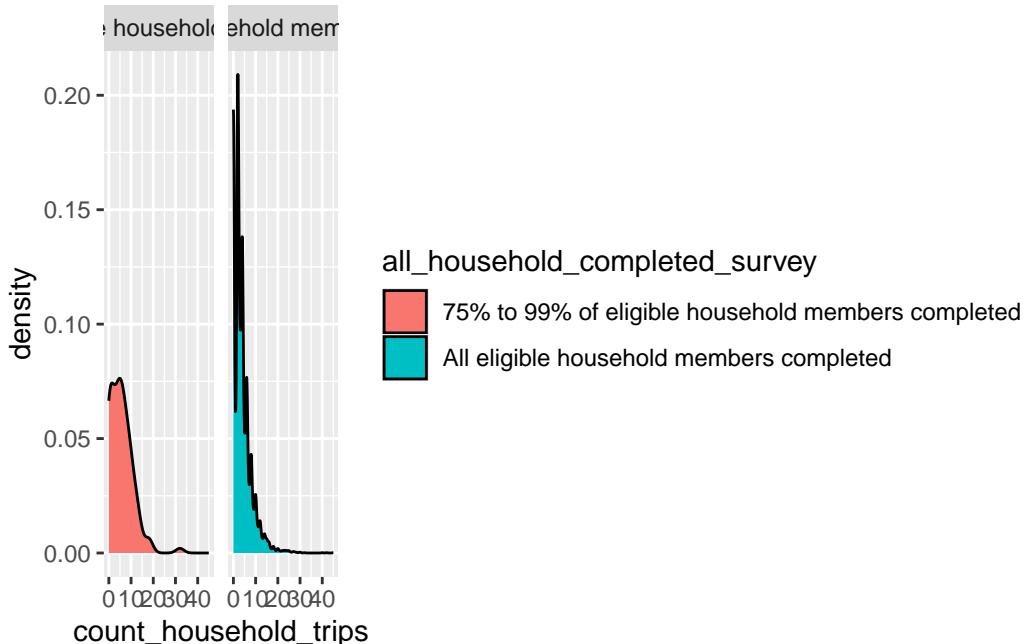


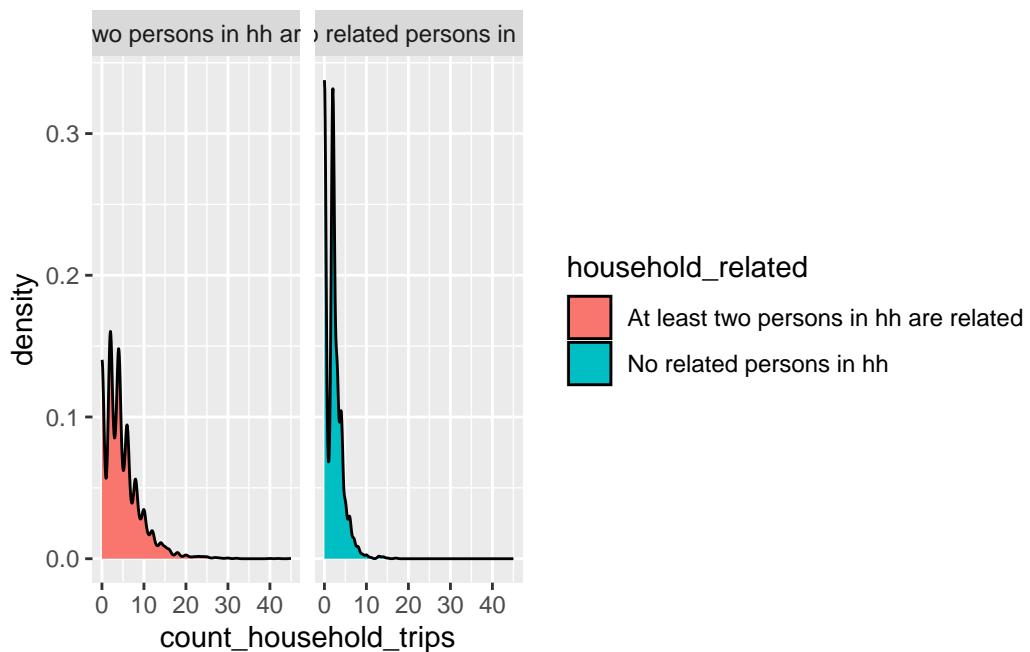
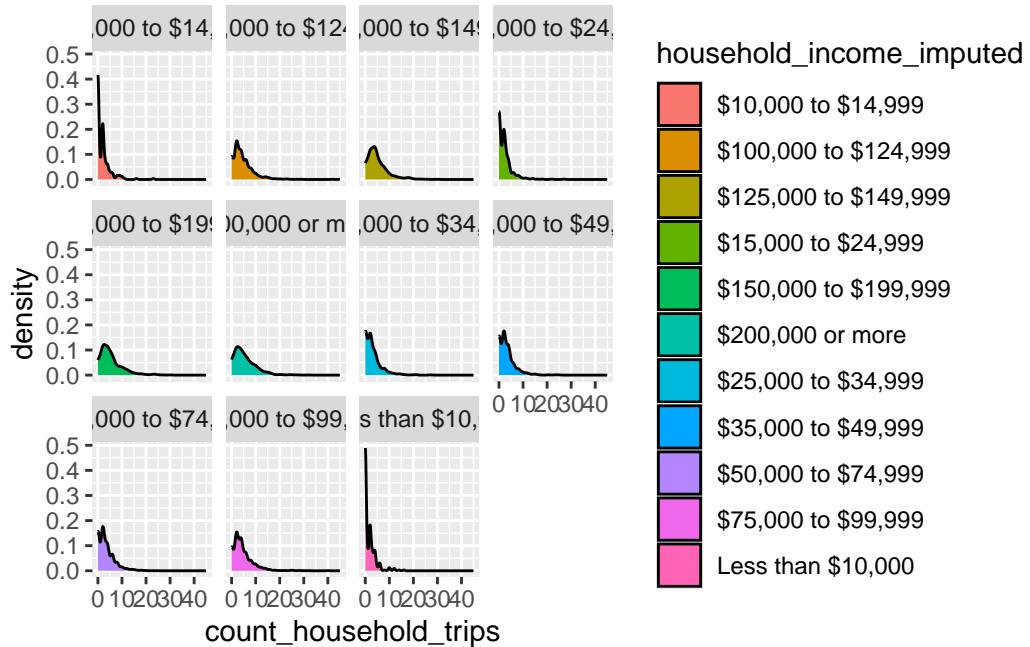


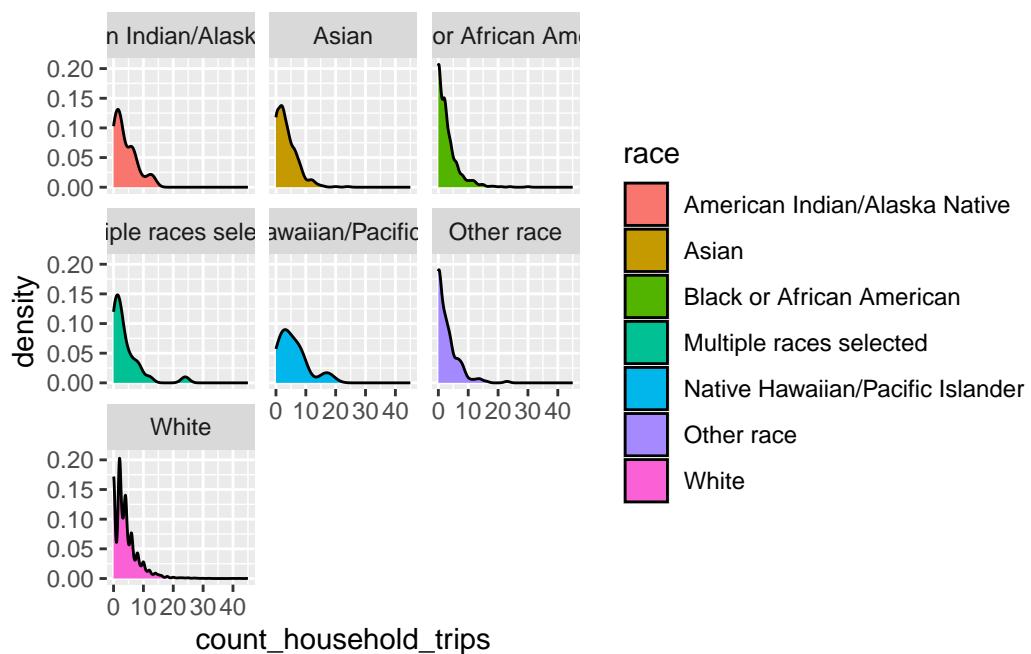
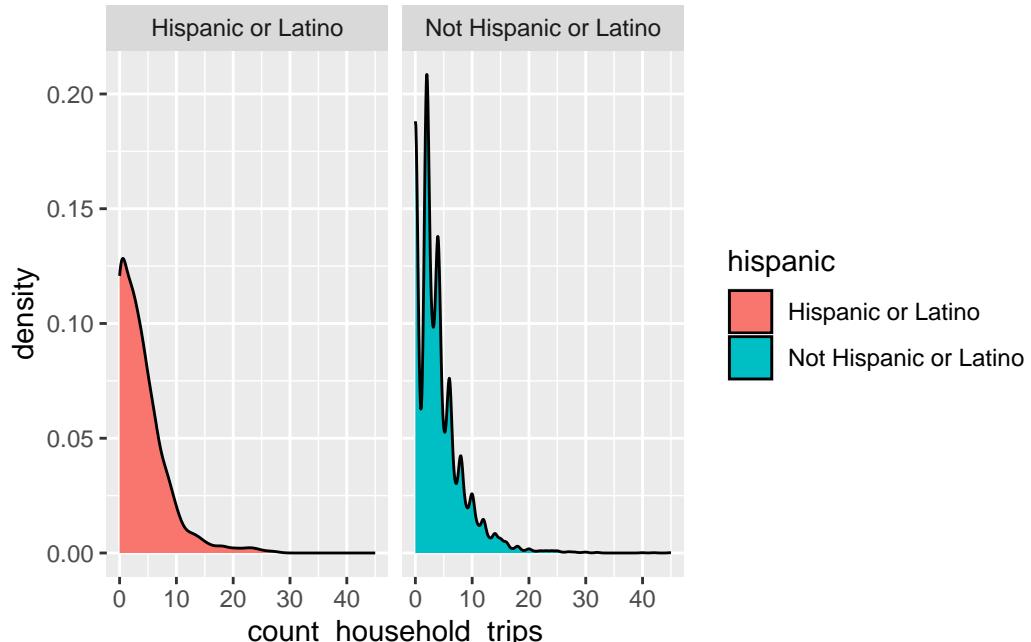


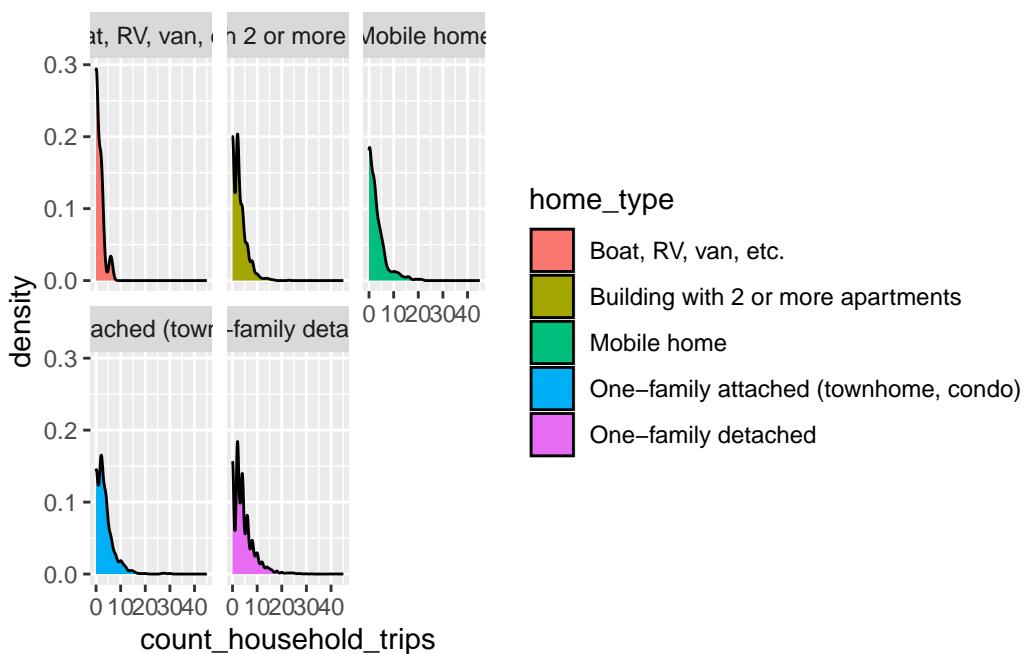
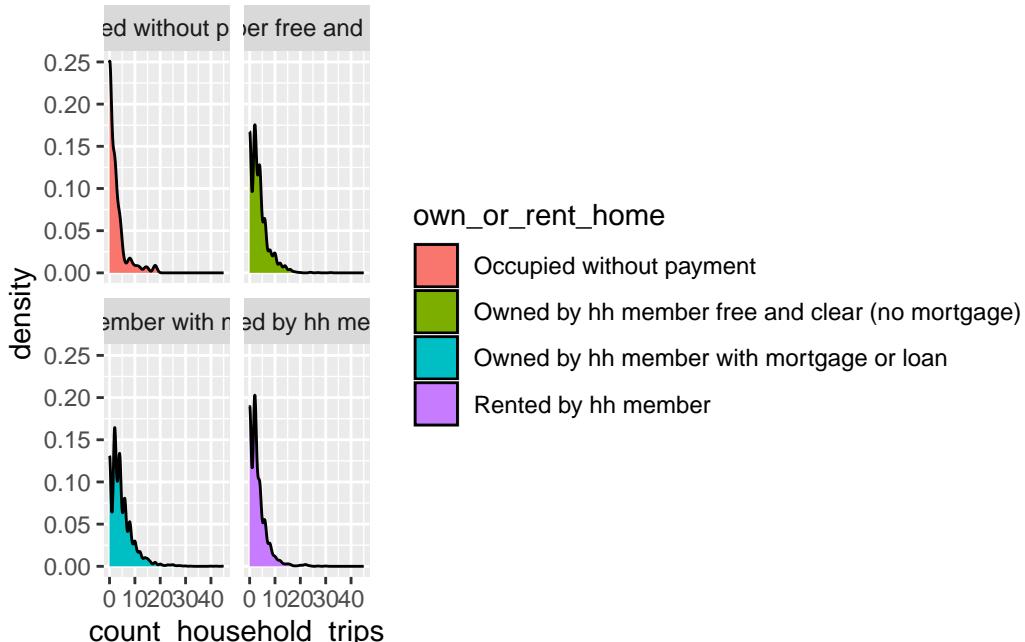


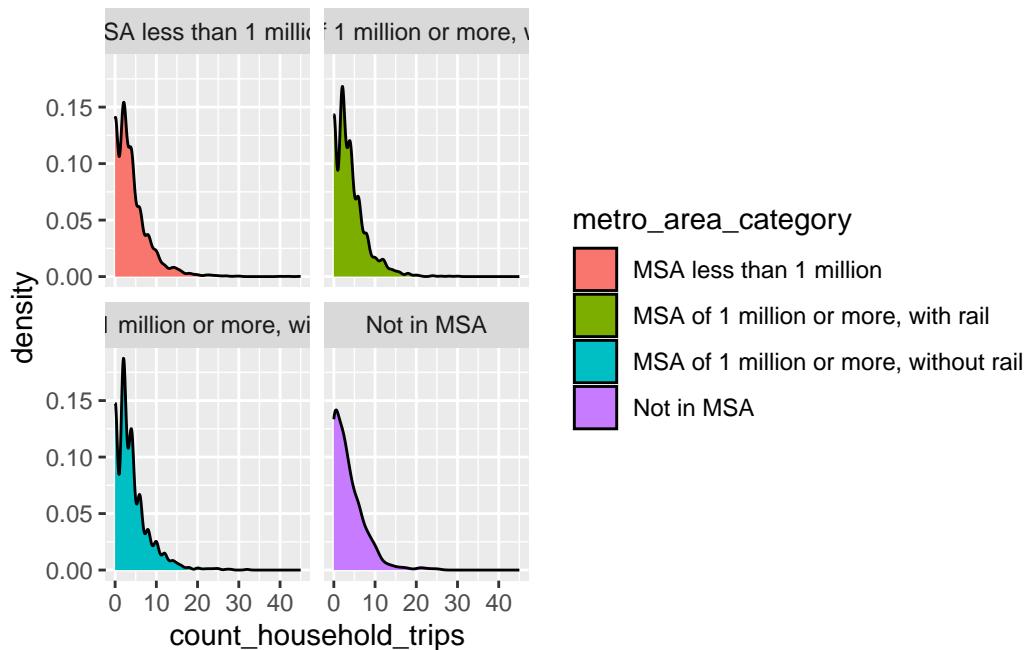
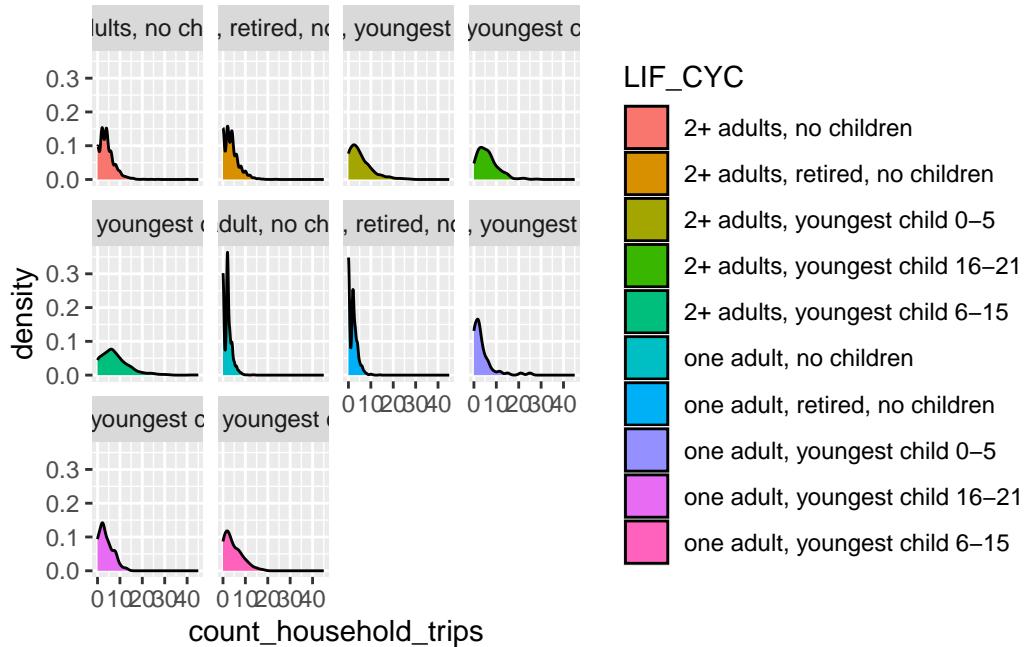


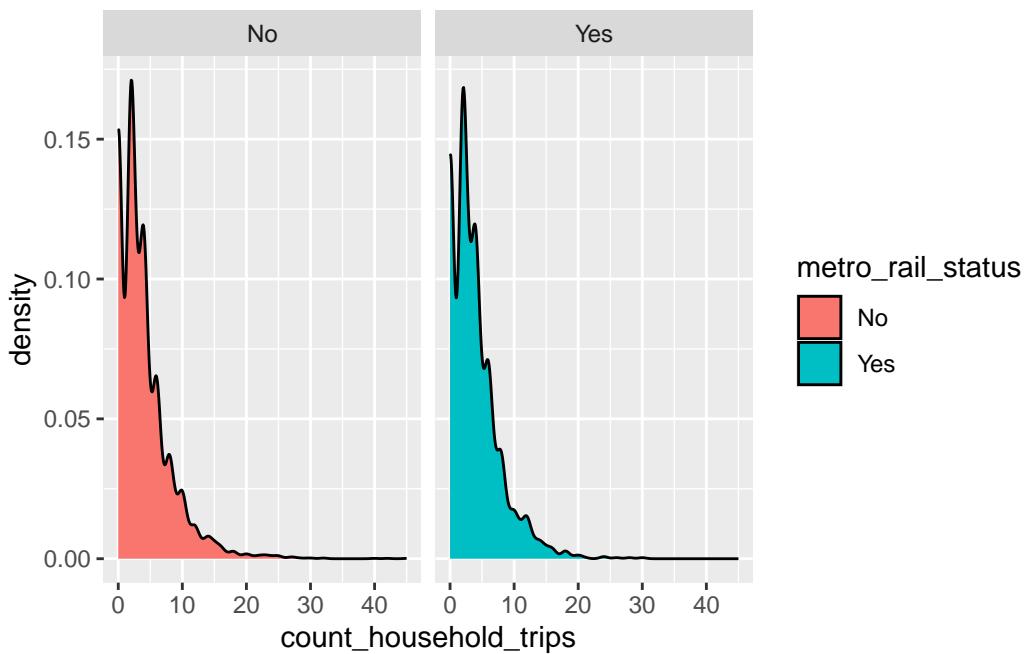
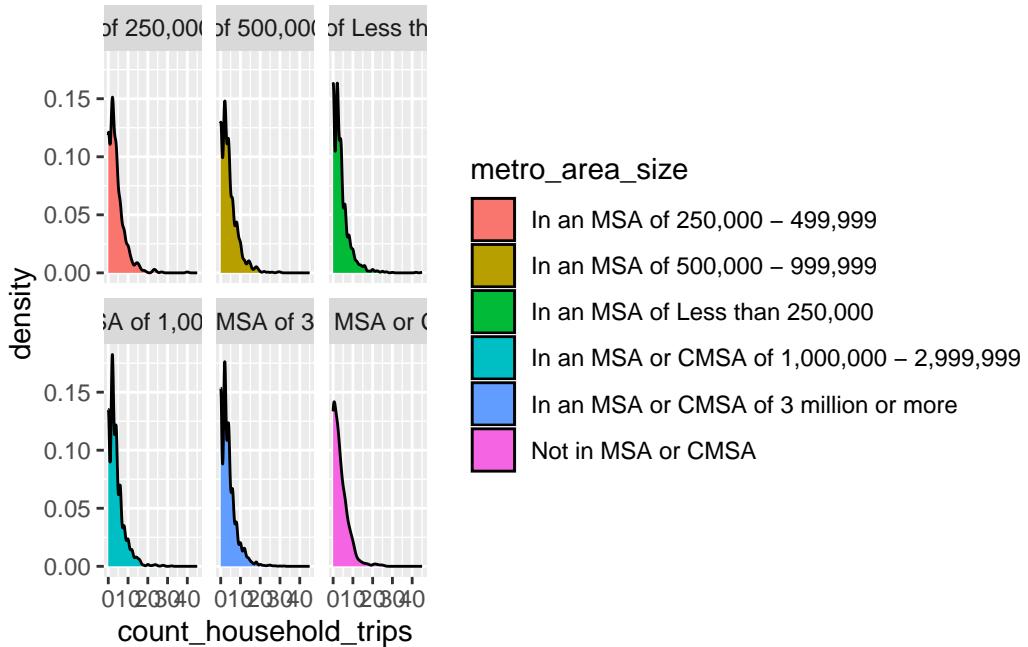


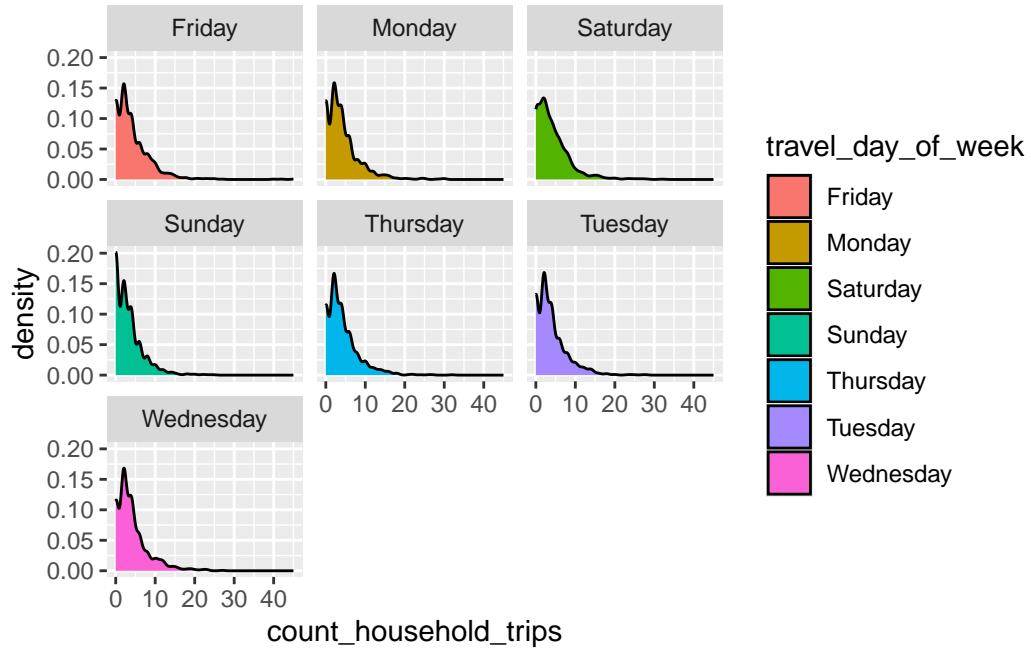






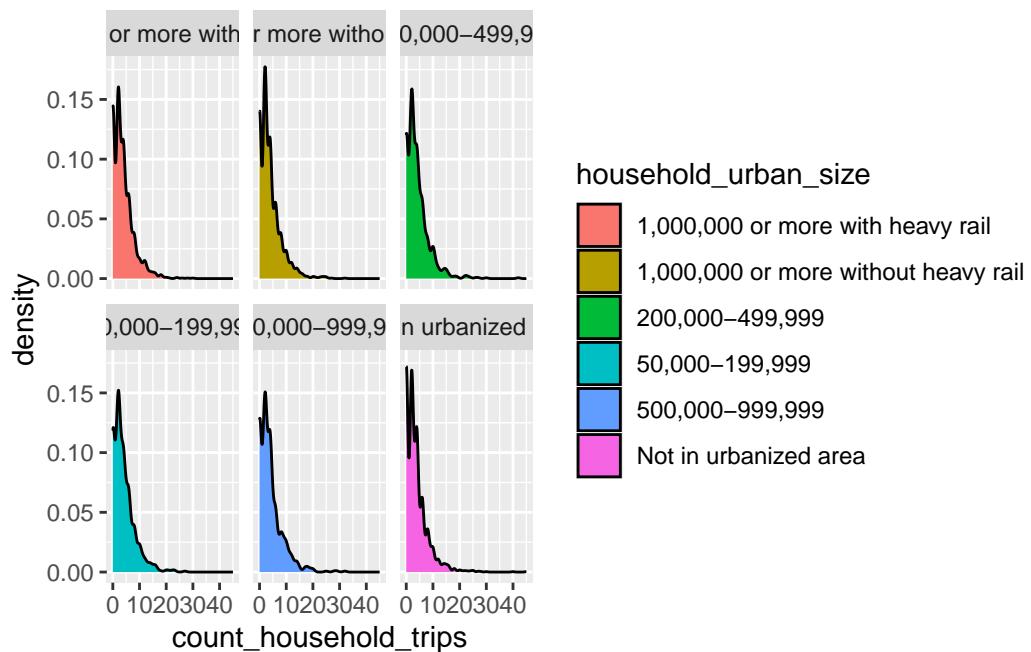
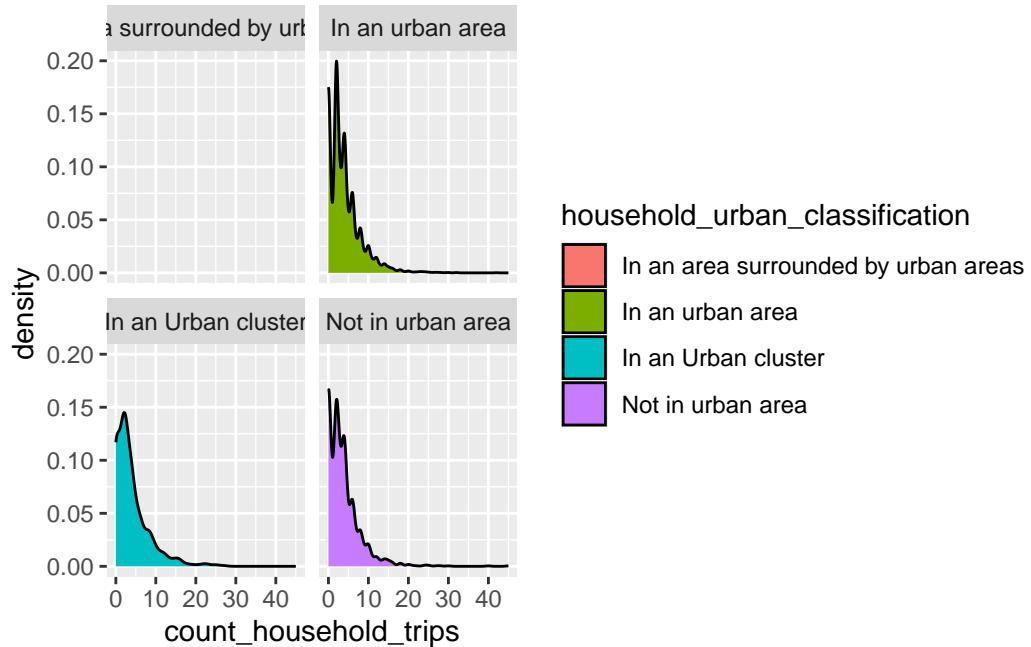


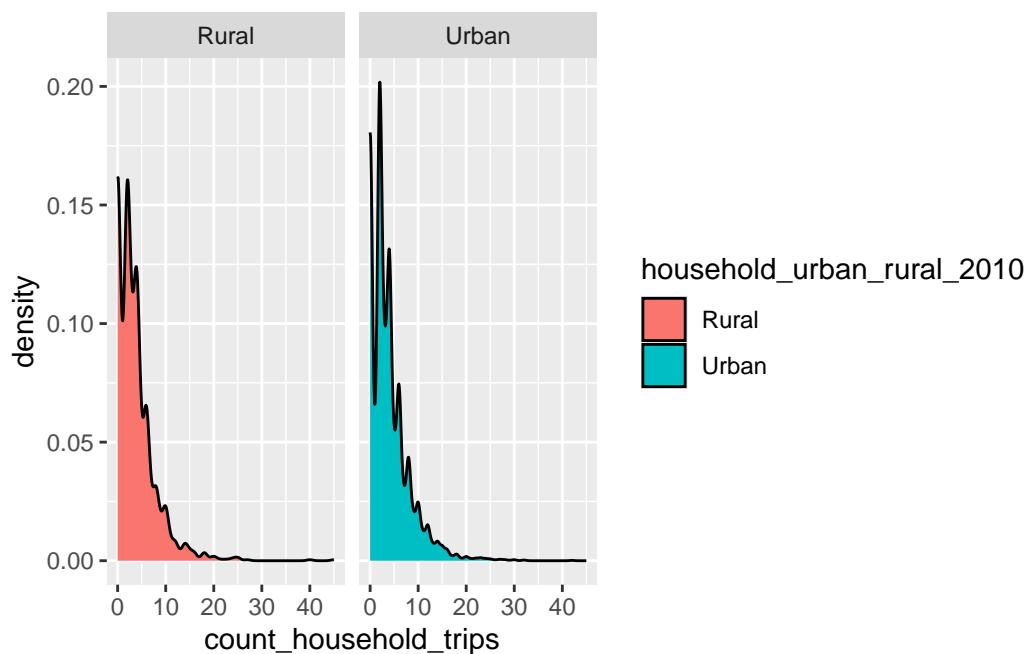
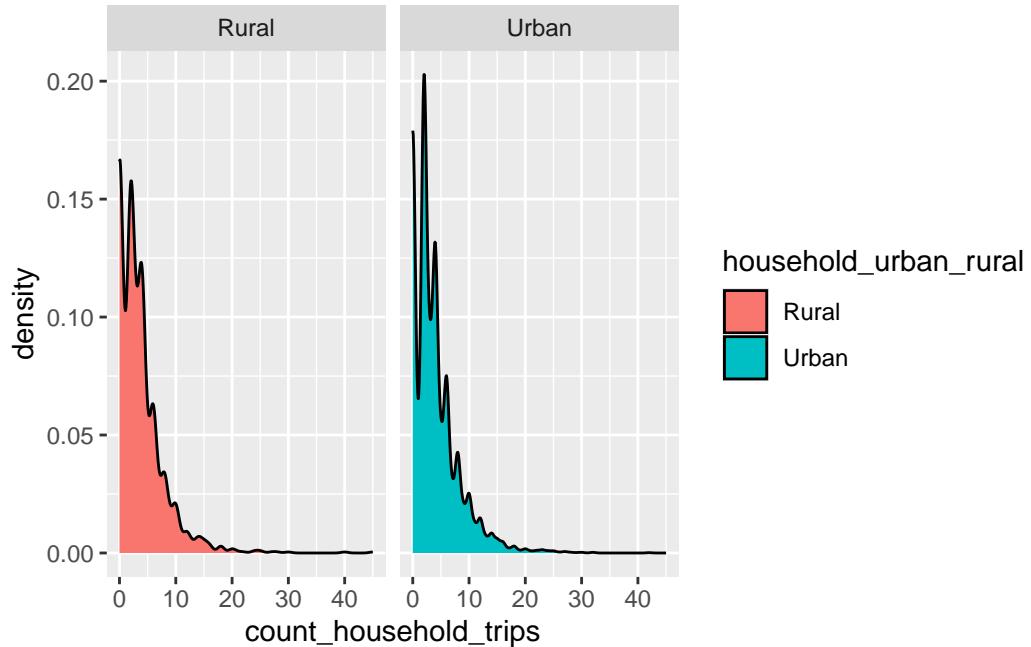




Warning: Groups with fewer than two data points have been dropped.

Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning -Inf





Person Data

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

```
Rows: 16,997
Columns: 129

$ household_id
$ person_id
$ person_weight_7_day
$ person_weight_5_day
$ person_weight_2_day
$ respondent_age
$ great_circle_distance_work
$ days_used_taxi_last_30
$ days_used_ride_share_last_30
$ days_used_scooter_last_30
$ days_used_public_transit_last_30
$ days_used_motorcycle_last_30
$ days_walked_as_transit_last_30
$ days_cycled_last_30
$ days_used_bikeshare_last_30
$ hours_worked_per_week
$ num_online_deliveries_last_30
$ num_goods_deliveries_last_30
$ num_food_deliveries_last_30
$ num_grocery_deliveries_last_30
$ num_services_deliveries_last_30
$ num_returned_home_pickup
$ num_returned_ups_usps_fedex
$ num_returned_amazon_dropoff_center
$ num_returned_store
$ home_parking_cost
$ count_person_trips
$ num_adults_in_household
$ num_drivers_in_household
$ household_size
$ num_vehicles_in_household
$ travel_day_date
$ WRKCOUNT
$ household_stratum_id
$ household_grouping
$ household_grouping_2
$ household_grouping_3
$ household_grouping_4
$ household_grouping_5
$ household_grouping_6
$ household_grouping_7
$ household_grouping_8
$ household_grouping_9
$ household_grouping_10
$ household_grouping_11
$ household_grouping_12
$ household_grouping_13
$ household_grouping_14
$ household_grouping_15
$ household_grouping_16
$ household_grouping_17
$ household_grouping_18
$ household_grouping_19
$ household_grouping_20
$ household_grouping_21
$ household_grouping_22
$ household_grouping_23
$ household_grouping_24
$ household_grouping_25
$ household_grouping_26
$ household_grouping_27
$ household_grouping_28
$ household_grouping_29
$ household_grouping_30
$ household_grouping_31
$ household_grouping_32
$ household_grouping_33
$ household_grouping_34
$ household_grouping_35
$ household_grouping_36
$ household_grouping_37
$ household_grouping_38
$ household_grouping_39
$ household_grouping_40
$ household_grouping_41
$ household_grouping_42
$ household_grouping_43
$ household_grouping_44
$ household_grouping_45
$ household_grouping_46
$ household_grouping_47
$ household_grouping_48
$ household_grouping_49
$ household_grouping_50
$ household_grouping_51
$ household_grouping_52
$ household_grouping_53
$ household_grouping_54
$ household_grouping_55
$ household_grouping_56
$ household_grouping_57
$ household_grouping_58
$ household_grouping_59
$ household_grouping_60
$ household_grouping_61
$ household_grouping_62
$ household_grouping_63
$ household_grouping_64
$ household_grouping_65
$ household_grouping_66
$ household_grouping_67
$ household_grouping_68
$ household_grouping_69
$ household_grouping_70
$ household_grouping_71
$ household_grouping_72
$ household_grouping_73
$ household_grouping_74
$ household_grouping_75
$ household_grouping_76
$ household_grouping_77
$ household_grouping_78
$ household_grouping_79
$ household_grouping_80
$ household_grouping_81
$ household_grouping_82
$ household_grouping_83
$ household_grouping_84
$ household_grouping_85
$ household_grouping_86
$ household_grouping_87
$ household_grouping_88
$ household_grouping_89
$ household_grouping_90
$ household_grouping_91
$ household_grouping_92
$ household_grouping_93
$ household_grouping_94
$ household_grouping_95
$ household_grouping_96
$ household_grouping_97
$ household_grouping_98
$ household_grouping_99
$ household_grouping_100
$ household_grouping_101
$ household_grouping_102
$ household_grouping_103
$ household_grouping_104
$ household_grouping_105
$ household_grouping_106
$ household_grouping_107
$ household_grouping_108
$ household_grouping_109
$ household_grouping_110
$ household_grouping_111
$ household_grouping_112
$ household_grouping_113
$ household_grouping_114
$ household_grouping_115
$ household_grouping_116
$ household_grouping_117
$ household_grouping_118
$ household_grouping_119
$ household_grouping_120
$ household_grouping_121
$ household_grouping_122
$ household_grouping_123
$ household_grouping_124
$ household_grouping_125
$ household_grouping_126
$ household_grouping_127
$ household_grouping_128
$ household_grouping_129
```

```
$ census_division
$ census_region
$ daytime_limit_driving
$ unaffected_travel
$ limited_transit_use
$ limited_transit_use_noresponse
$ requested_rides
$ limited_driving
$ requested_ride_shares
$ requires_special_transport
$ reduce_travel_disability
$ covid_online_delivery_impact
$ covid_public_transit_impact
$ covid_school_travel_impact
$ covid_work_travel_impact
$ covid_online_delivery_impact_permant
$ covid_public_transit_impact_permant
$ covid_school_travel_impact_permant
$ covid_work_travel_impact_permant
$ 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
<chr> "New England", "New Eng-
<chr> "Northeast", "Northeast-
<chr> "Valid skip", "Valid sk-
<chr> "Do the same as before"-
<chr> "Do less often than bef-
<chr> "Valid skip", "Valid sk-
<chr> "Valid skip", "Do less ~
<chr> "Valid skip", "Valid sk-
<chr> "Temporary", "Don't Kno-
<chr> "Valid skip", "Valid sk-
<chr> "Valid skip", "Permanen-
<chr> "Yes", "Yes", "Valid sk-
<chr> "Valid skip", "No", "Va-
<chr> "Valid skip", "Valid sk-
<chr> "Master's degree", "Mas-
<chr> "Valid skip", "Valid sk-
<chr> "Yes", "Yes", "Yes", "Y-
<chr> "$200,000 or more", "$2-
<chr> "$200,000 or more", "$2-
<chr> "Not Hispanic or Latino-
<chr> "White", "White", "Whit-
<chr> "Owned by hh member wit-
<chr> "No", "No", "Valid skip-
<chr> "Valid skip", "Valid sk-
<chr> "No", "No", "Valid skip-
<chr> "Yes", "Yes", "Valid sk-
<chr> "2+ adults, youngest ch-
<chr> "No", "No", "No", "No", ~
<chr> "Valid skip", "Valid sk-
<chr> "MSA of 1 million or mo-
<chr> "In an MSA or CMSA of 3~
```

```

$ 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
$ school_type_nontraditional
$ school_transportation
$ school_type
$ school_program_description
$ travel_day_of_week
$ days_used_discounted_transit_last_30
$ household_urban_classification
$ household_urban_size
$ household_urban_rural
$ fewer_trips_last_30
$ fewer_trips_last_30_reason_more_deliveries
$ fewer_trips_last_30_reason_covid
$ fewer_trips_last_30_reason_not_feel_safe
$ fewer_trips_last_30_reason_not_feel_clean
$ fewer_trips_last_30_reason_not_reliable
$ fewer_trips_last_30_reason_not_go_where_needed
$ fewer_trips_last_30_reason_unaffordable
$ fewer_trips_last_30_reason_health
$ fewer_trips_last_30_reason_time
$ fewer_trips_last_30_reason_other
$ USEPUBTR
$ who_completed_survey
$ days_per_week_work_from_home
$ worker_status

<chr> "No", "No", "No", "No", ~
<chr> "No", "Valid skip", "Va~
<chr> "Valid skip", "Valid sk~
<chr> "Valid skip", "Valid sk~
<chr> "Valid skip", "Valid sk~
<chr> "No", "Yes", "Valid ski~
<chr> "A homemaker", "Valid s~
<chr> "Self", "Someone else", ~
<chr> "No", "No", "No", "No", ~
<chr> "Valid skip", "Valid sk~
<chr> "Yes", "Yes", "Yes", "Y~
<chr> "Not Hispanic", "Not Hi~
<chr> "White", "White", "Whit~
<chr> "White", "White", "Whit~
<chr> "Self", "Spouse/Partner~
<chr> "Female", "Male", "Refu~
<chr> "Female", "Male", "Male~
<chr> "Valid skip", "Valid sk~
<chr> "No", "No", "Yes", "Yes~
<chr> "Valid skip", "Valid sk~
<chr> "Sunday", "Sunday", "Su~
<chr> "Valid skip", "Valid sk~
<chr> "In an urban area", "In~
<chr> "1,000,000 or more with~
<chr> "Urban", "Urban", "Urba~
<chr> "No, did not take fewer~
<chr> "Valid skip", "Valid sk~
<chr> "No", "No", "No", "No", ~
<chr> "Valid skip", "Person1"~
<chr> "Valid skip", "Five or ~
<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	Min. :2022-01-01	Min. :0.000
1st Qu.: 2.00	1st Qu.: 1.000	1st Qu.:2022-04-01	1st Qu.:0.000
Median : 2.00	Median : 2.000	Median :2022-08-01	Median :1.000
Mean : 2.91	Mean : 2.098	Mean :2022-07-16	Mean :1.216
3rd Qu.: 4.00	3rd Qu.: 3.000	3rd Qu.:2022-11-01	3rd Qu.:2.000
Max. :10.00	Max. :17.000	Max. :2023-01-01	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_permant covid_public_transit_impact_permant
Length:16997           Length:16997
Class :character        Class :character
Mode  :character        Mode  :character

covid_school_travel_impact_permant covid_work_travel_impact_permant
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_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

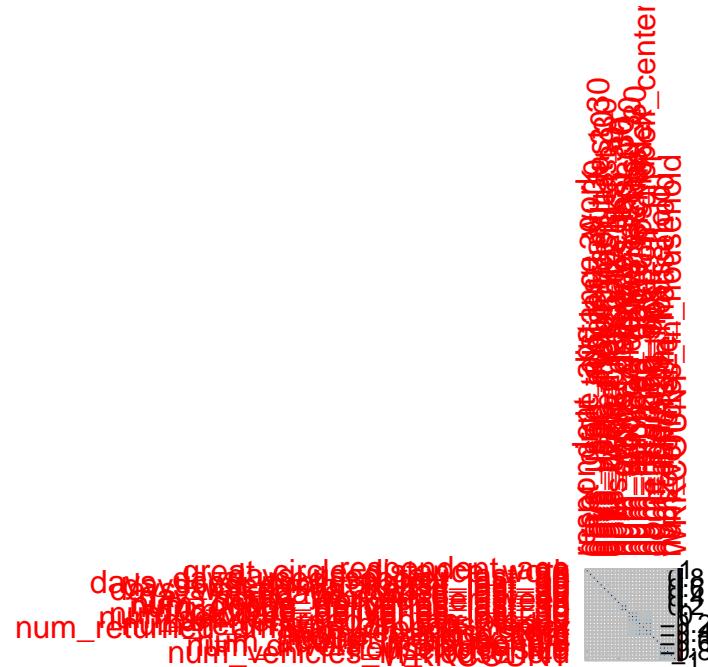
```

```

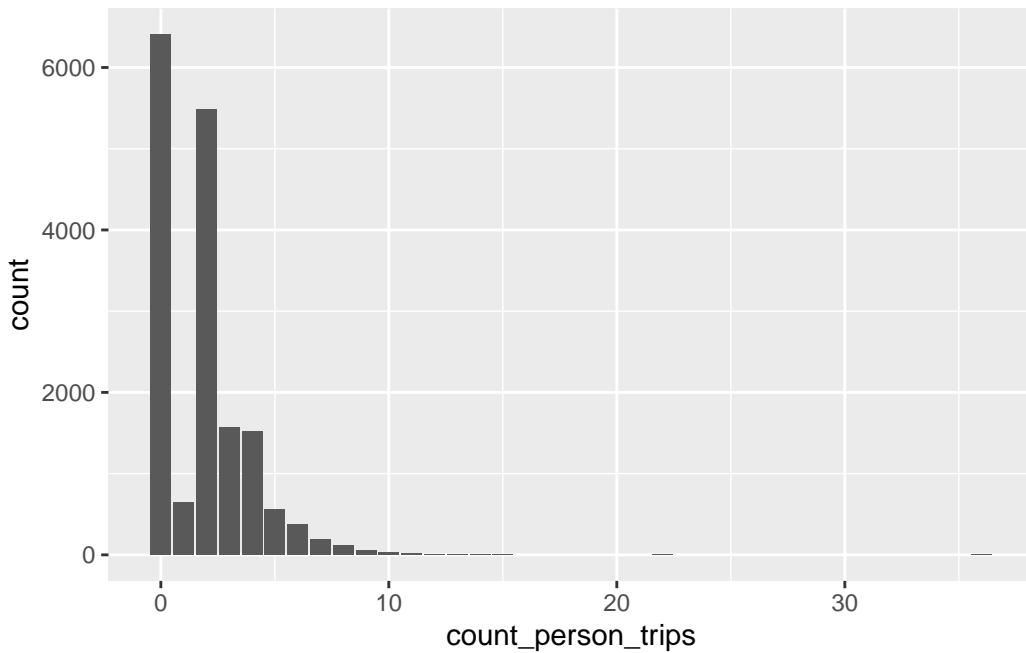
person |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot()

```

Warning in corrplot::corrplot(cor(select(select(select(person,
 -contains("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



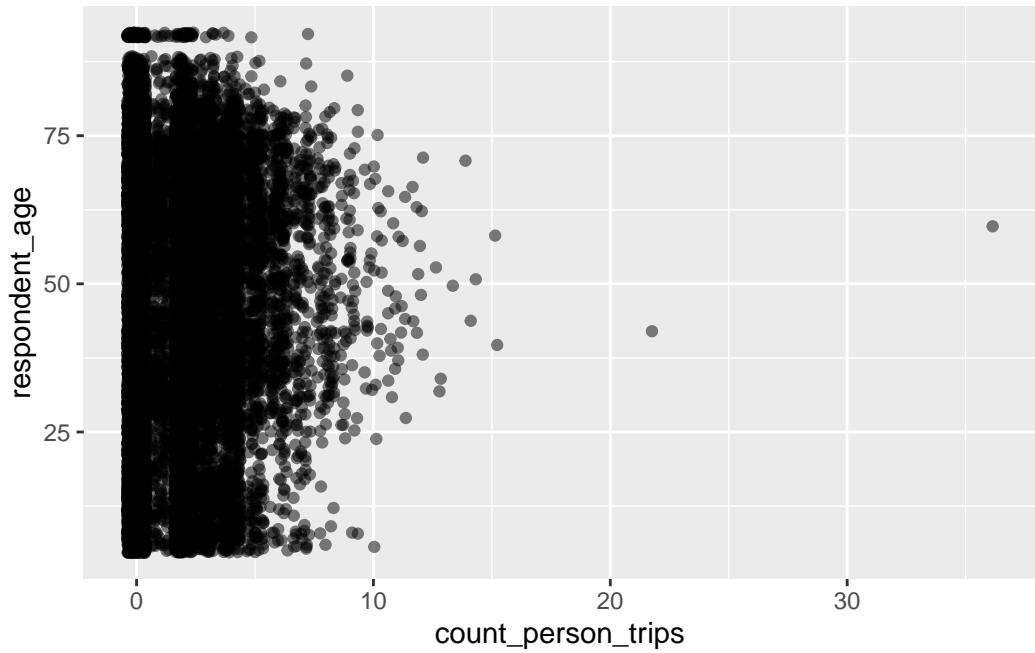
```
person |>
  ggplot(aes(count_person_trips)) +
  geom_bar()
```

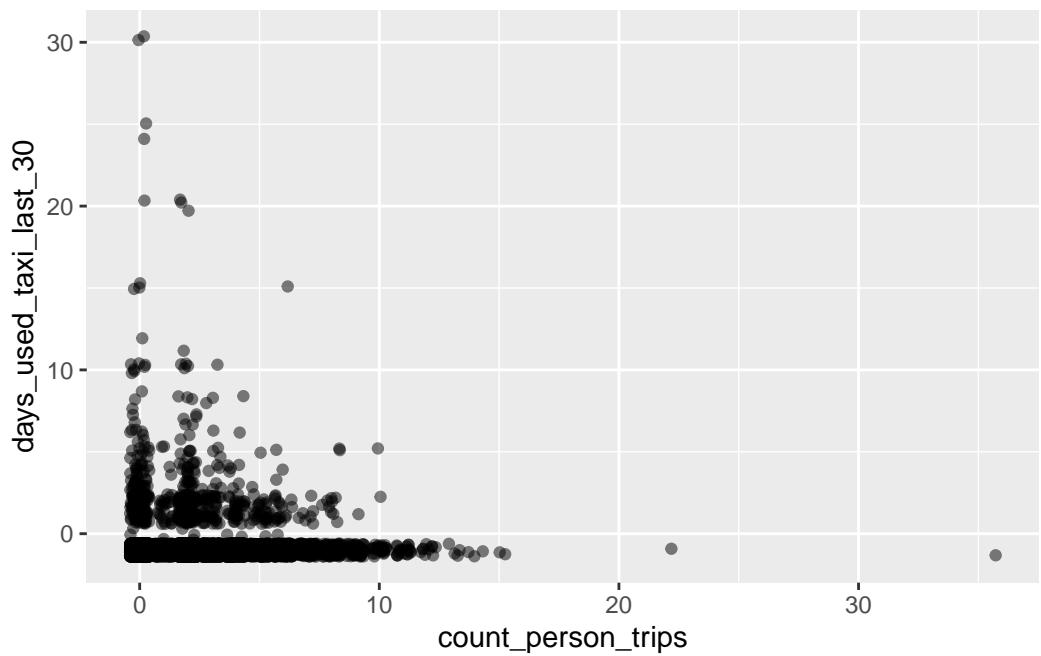
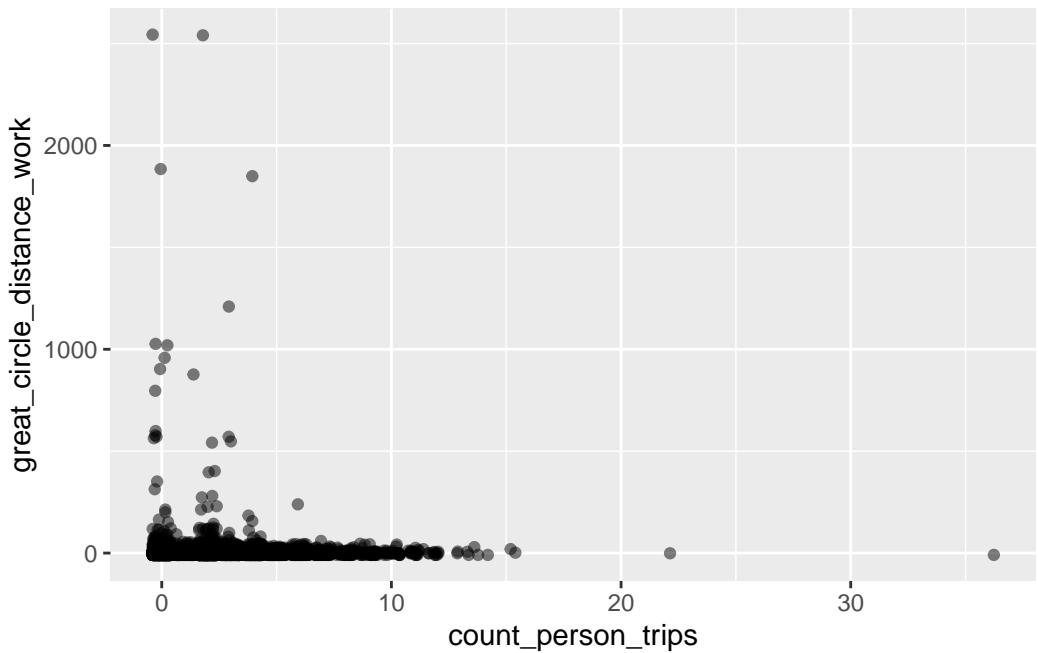


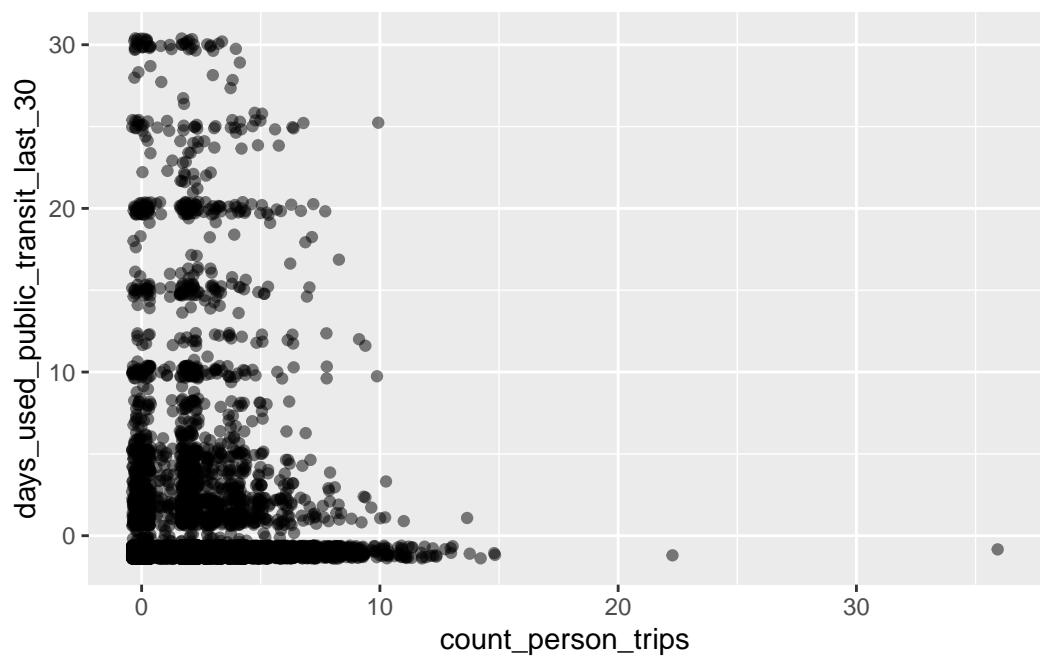
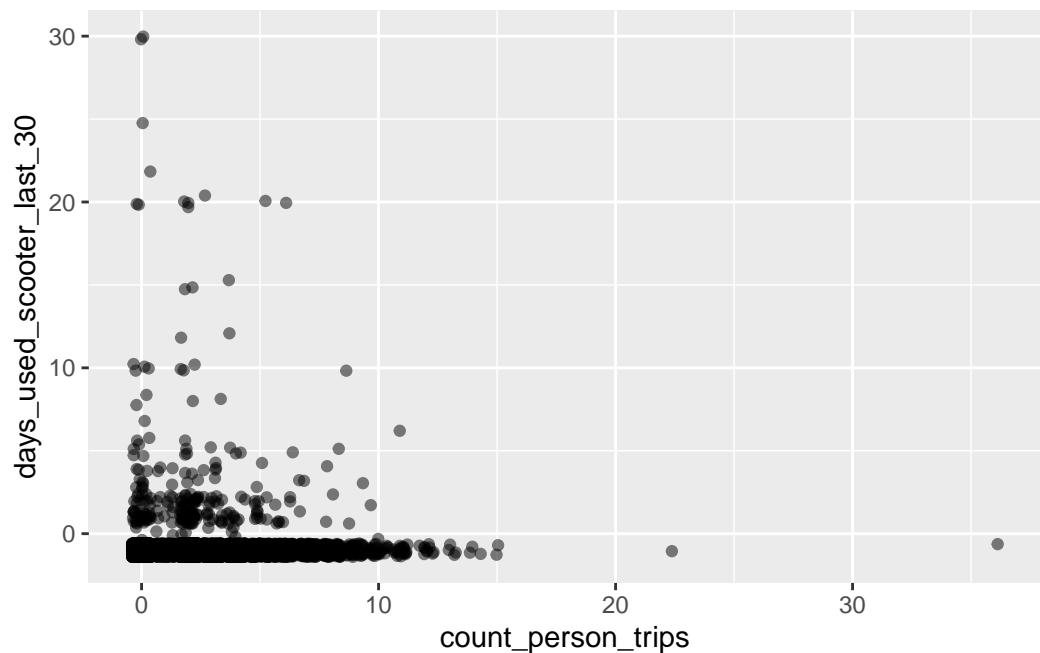
```
# Plot count_person_trips by all variables
person |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  colnames() -> person_cols

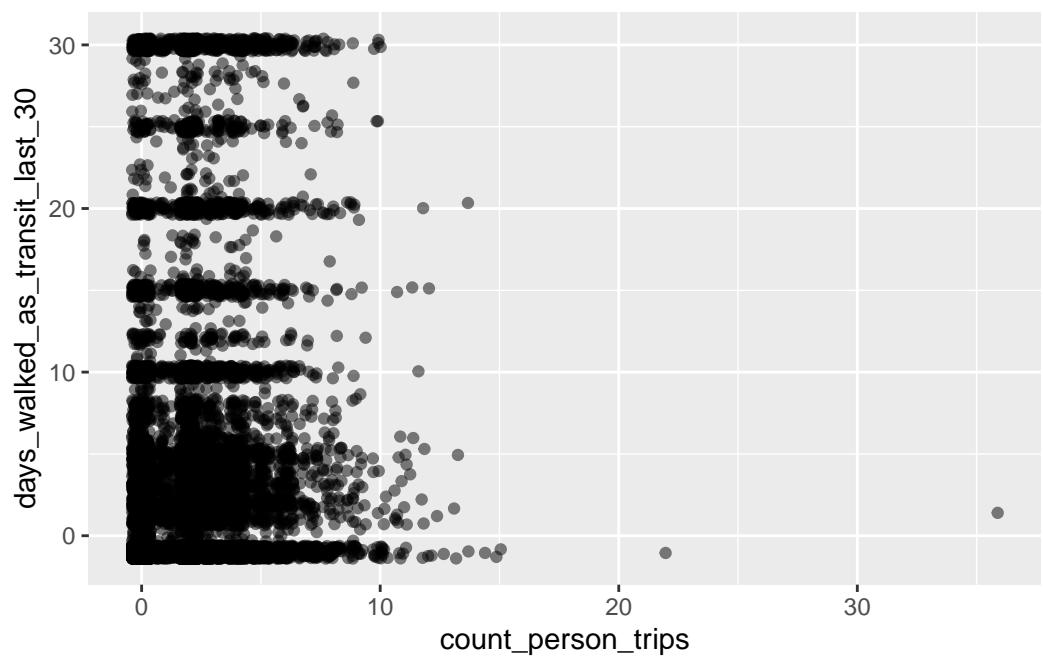
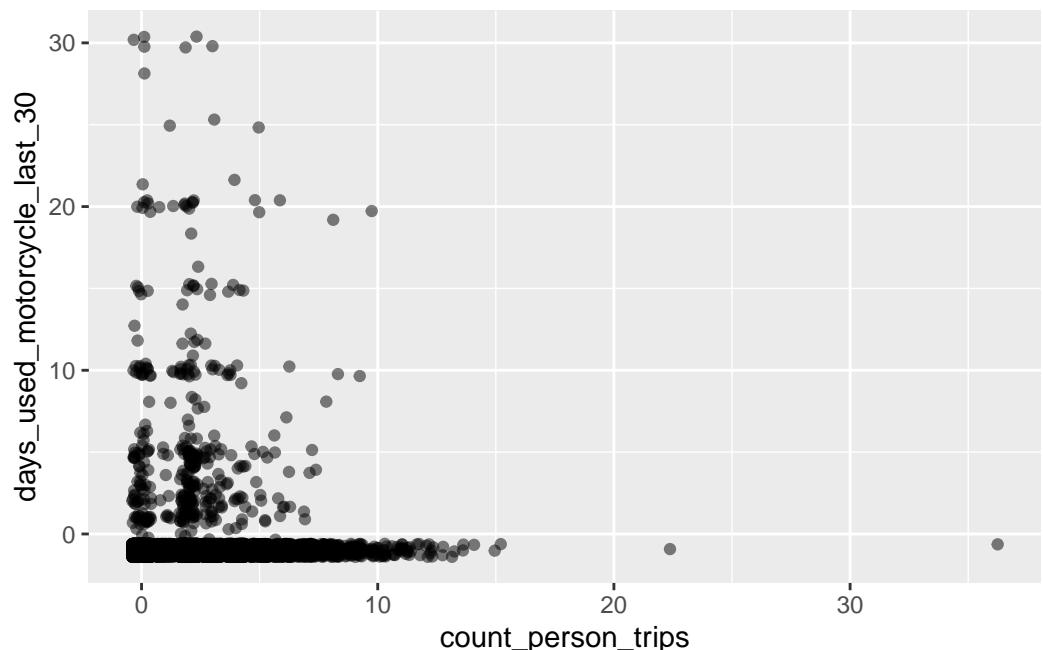
for (col in person_cols) {
  if (is.character(person[[col]])) {
    p <- person |>
      ggplot(aes(x = count_person_trips, fill = .data[[col]])) +
      geom_density() +
      facet_wrap(~.data[[col]])
    print(p)
  } else if (col == "count_person_trips") {
    next
  } else {
    p <- person |>
```

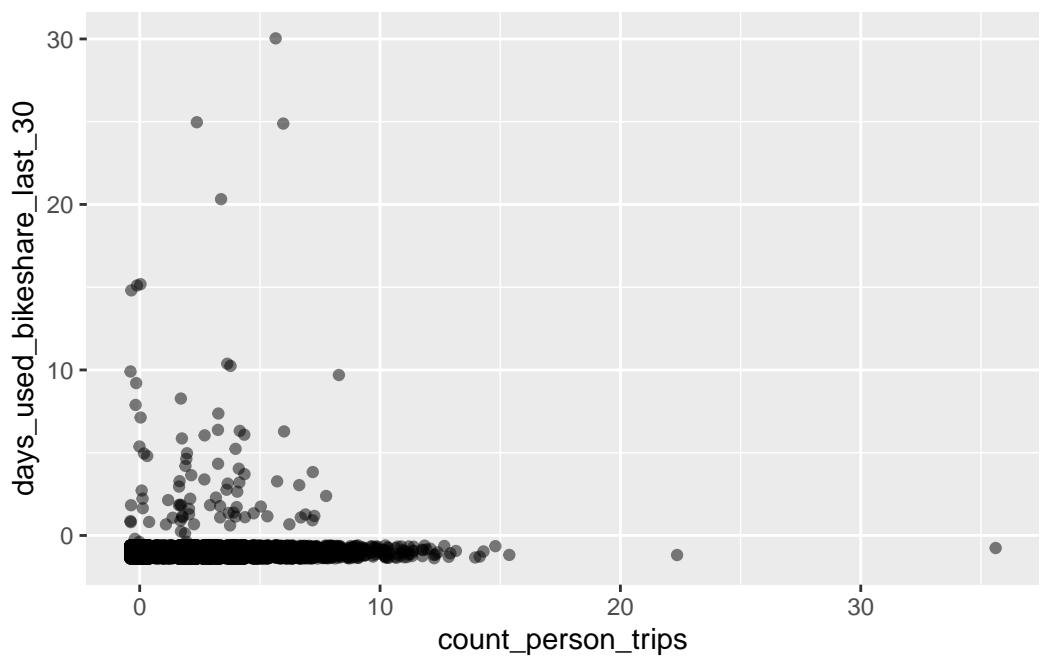
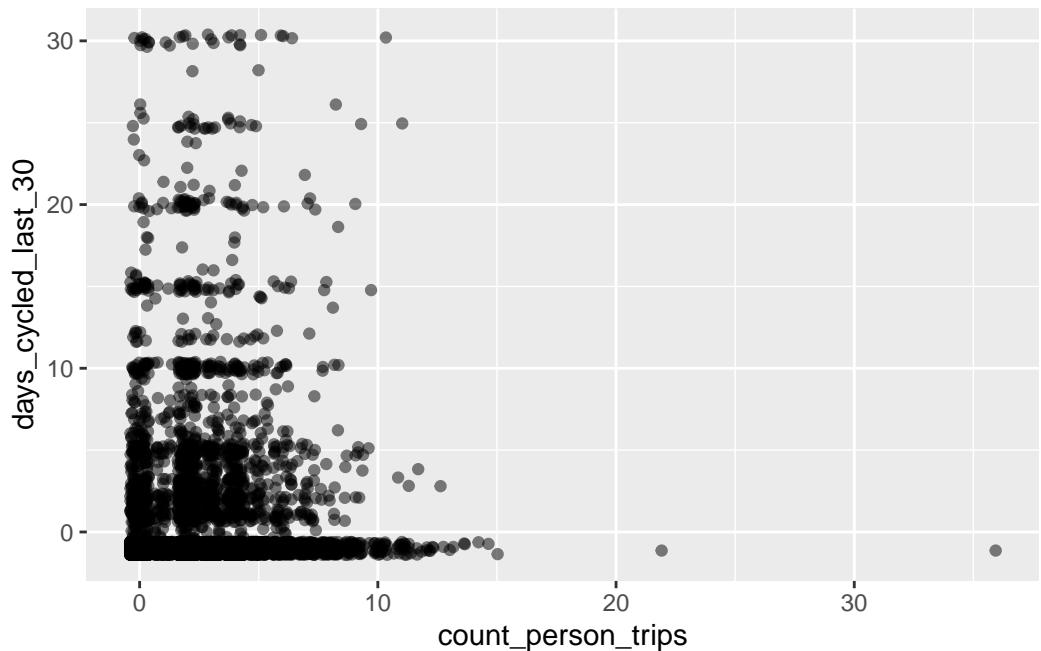
```
        ggplot(aes(x = count_person_trips, y = .data[[col]])) +
          geom_jitter(alpha = 0.5)
      print(p)
    }
}
```

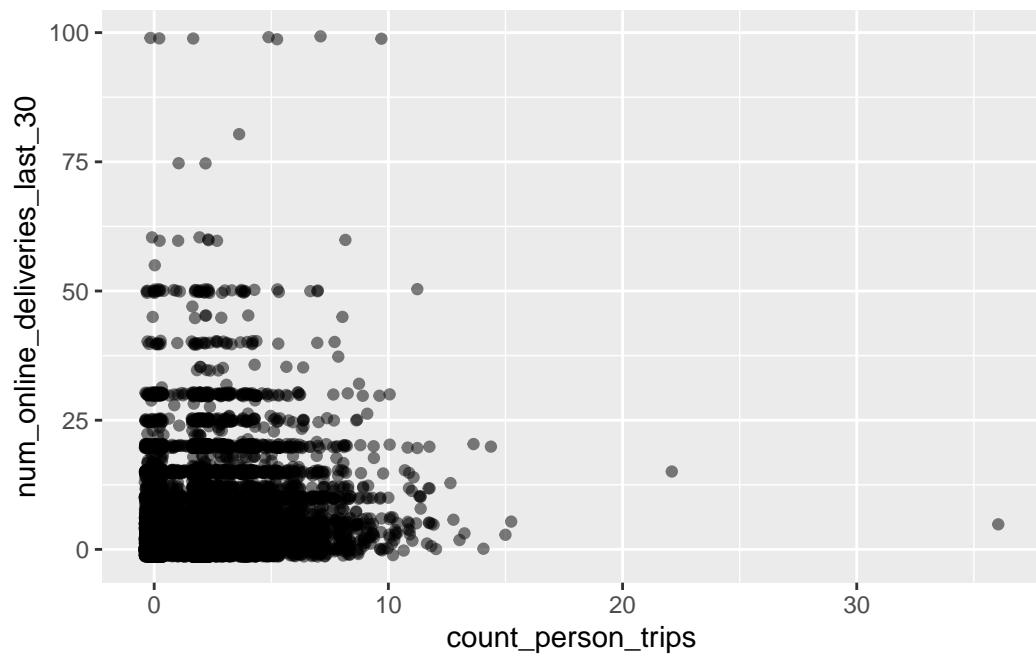
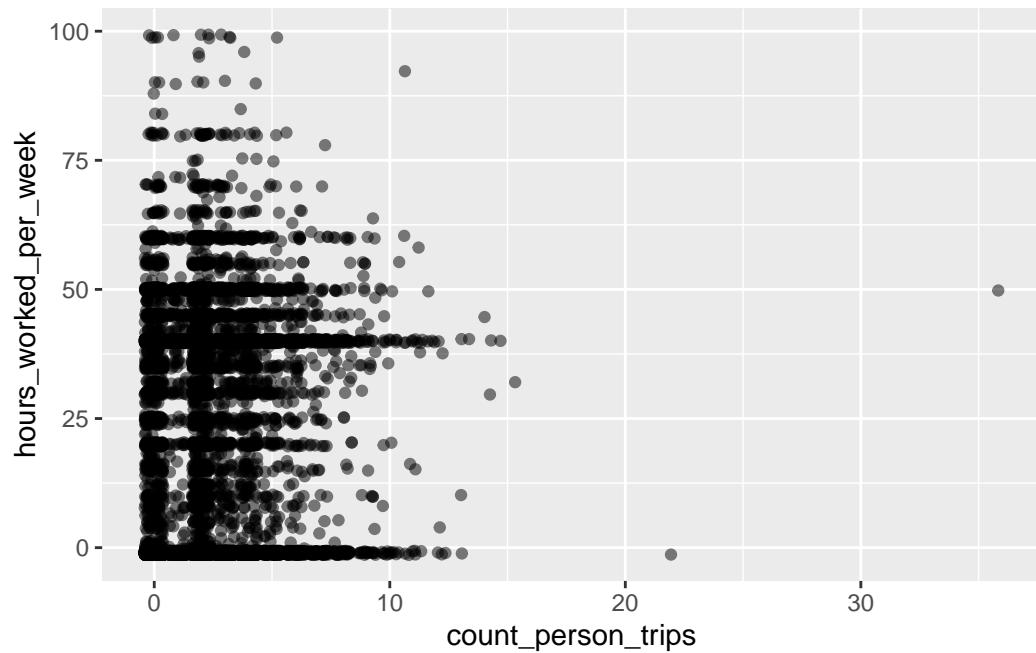


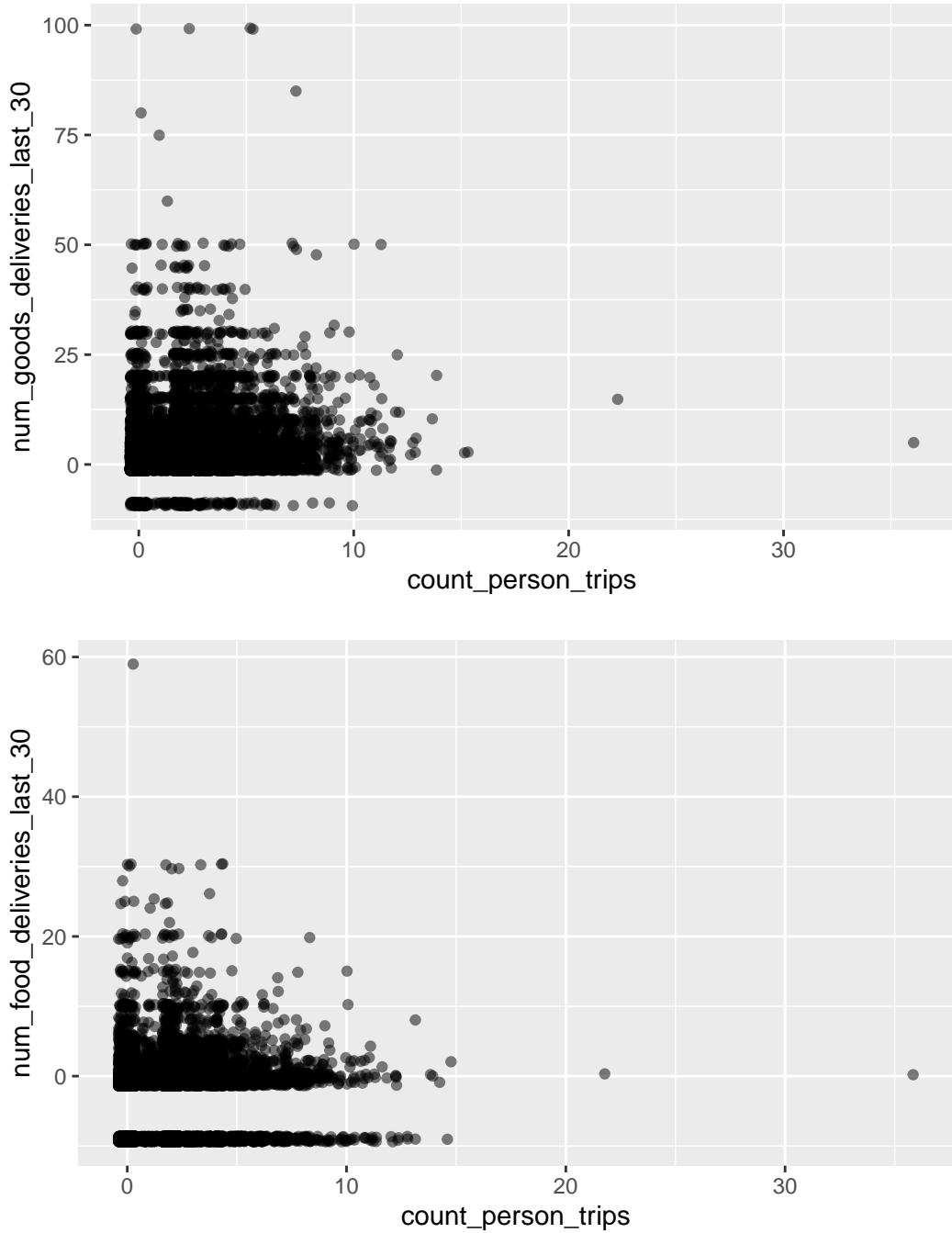


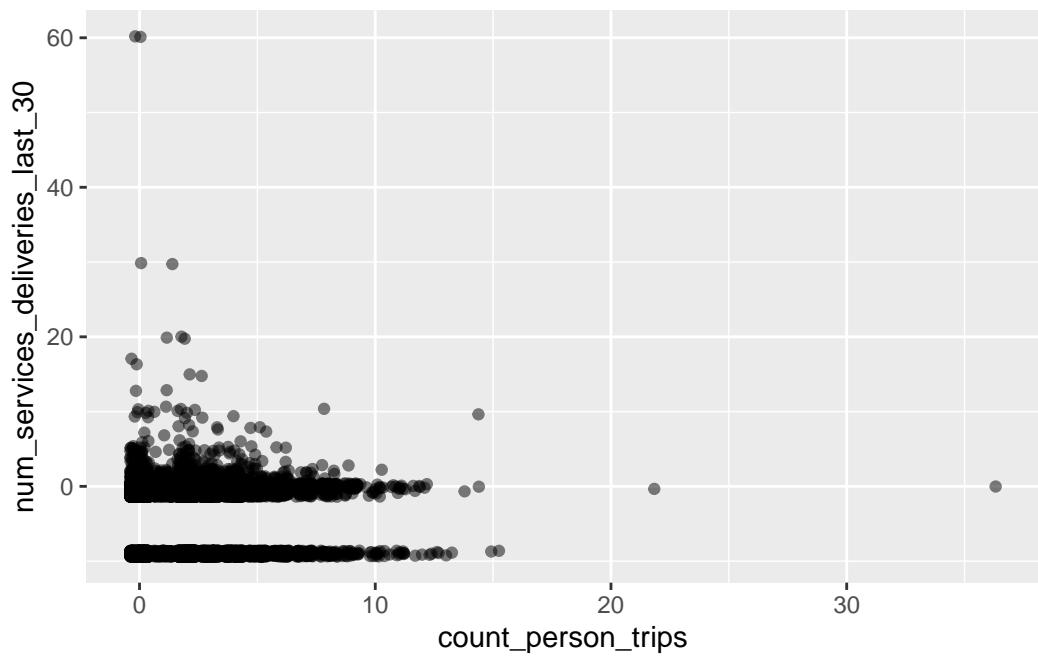
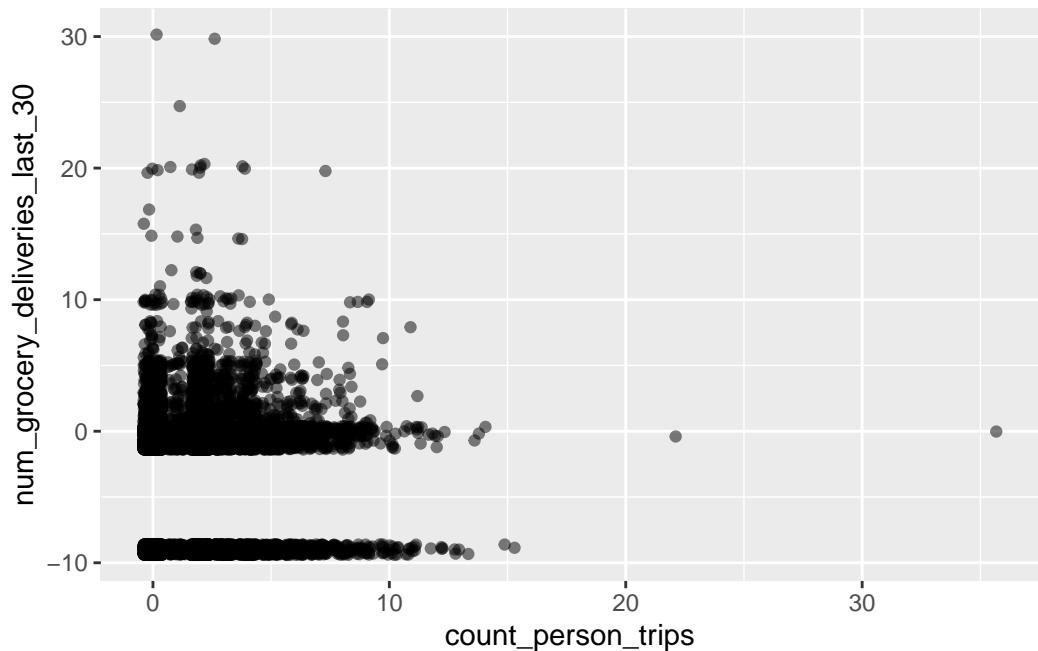


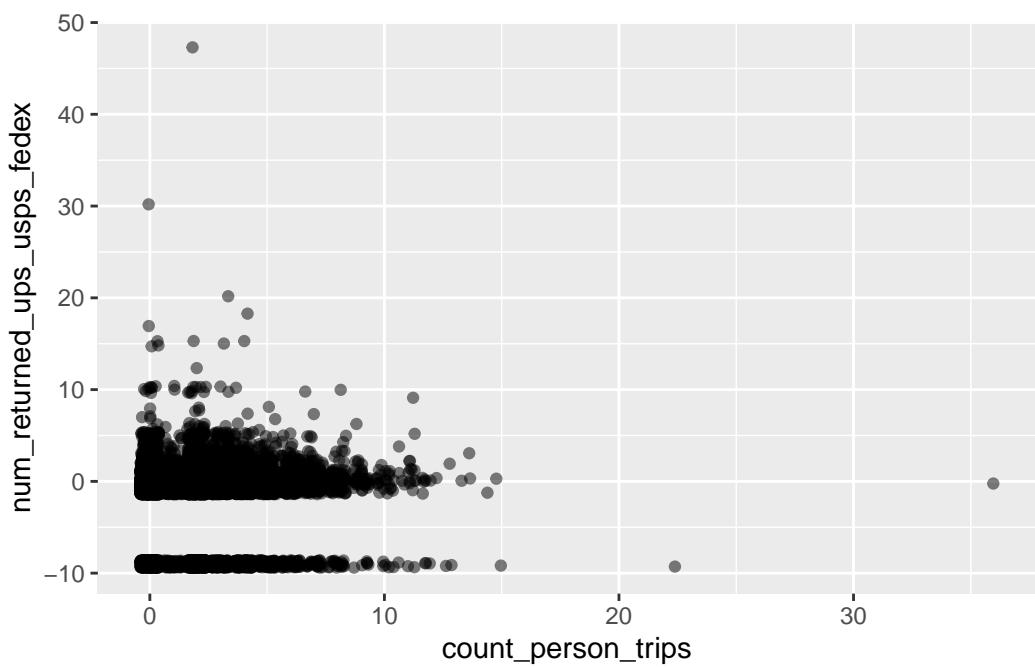
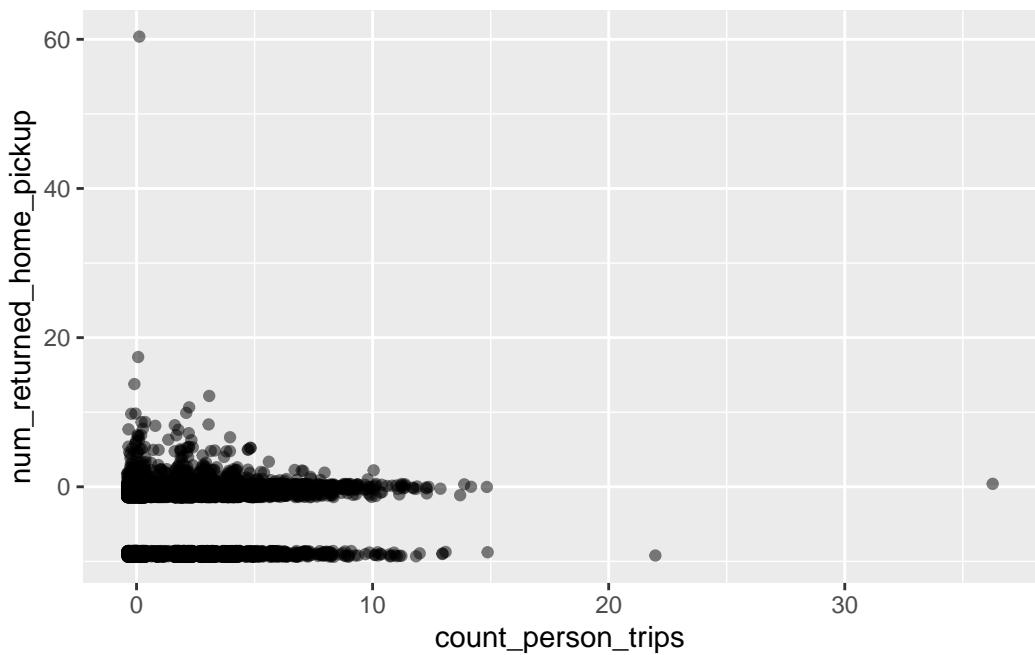


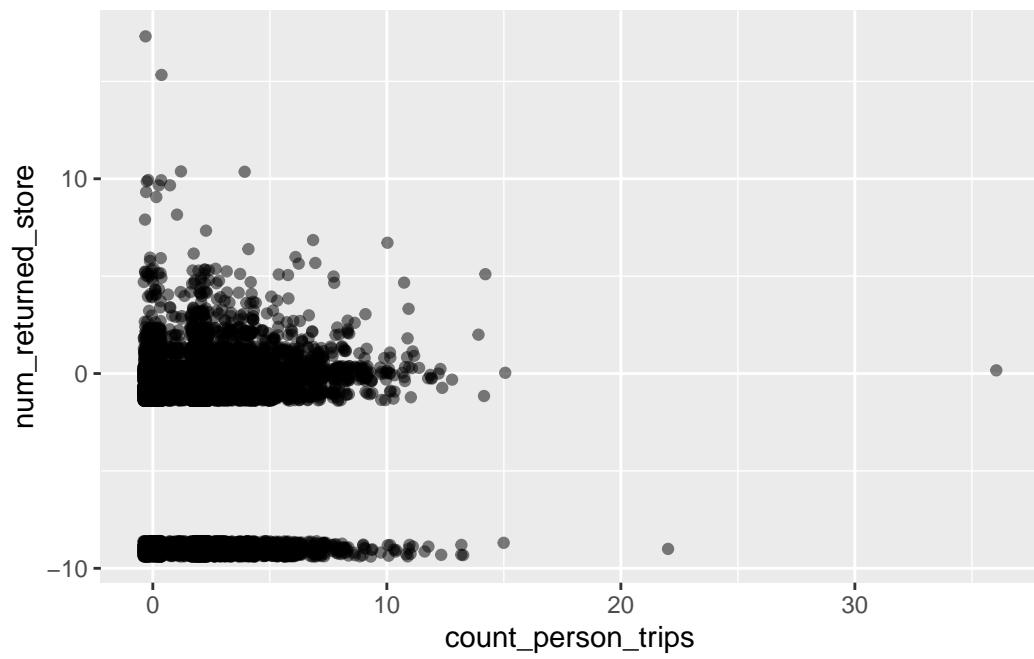
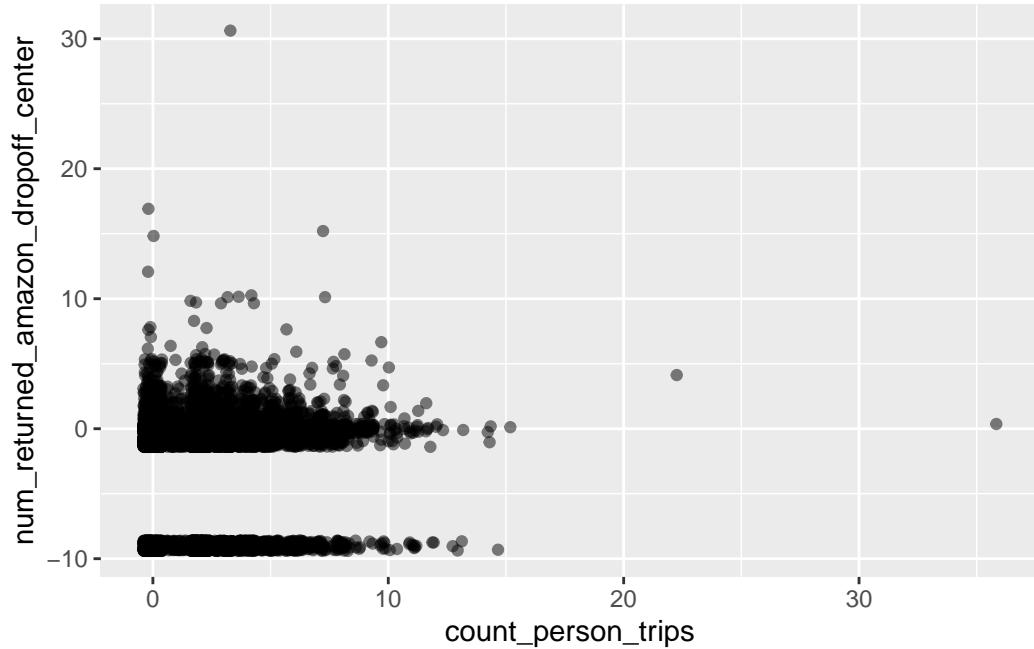


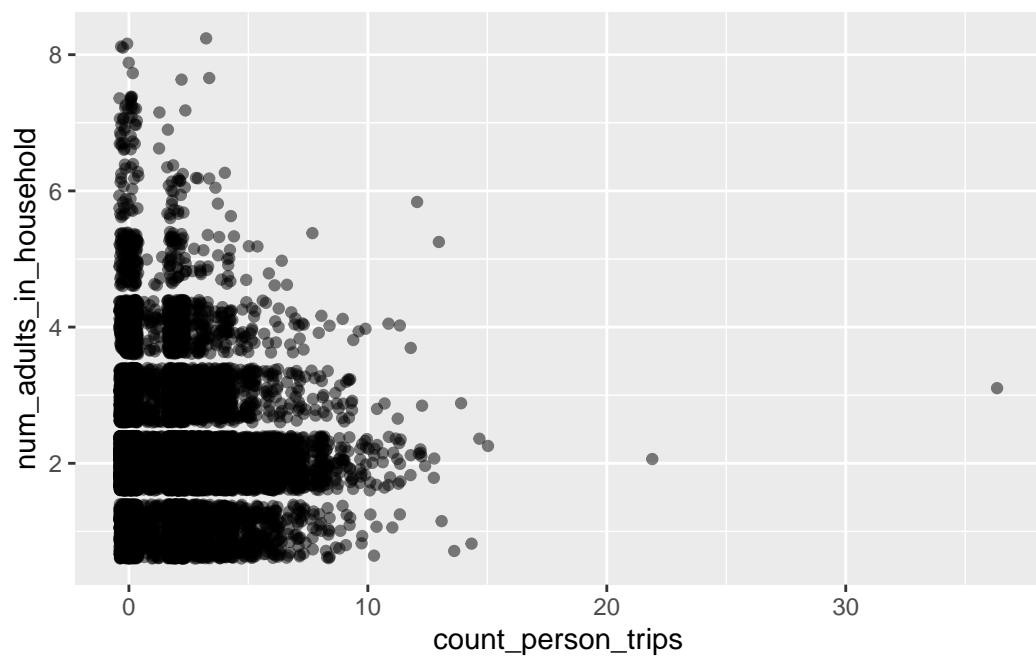
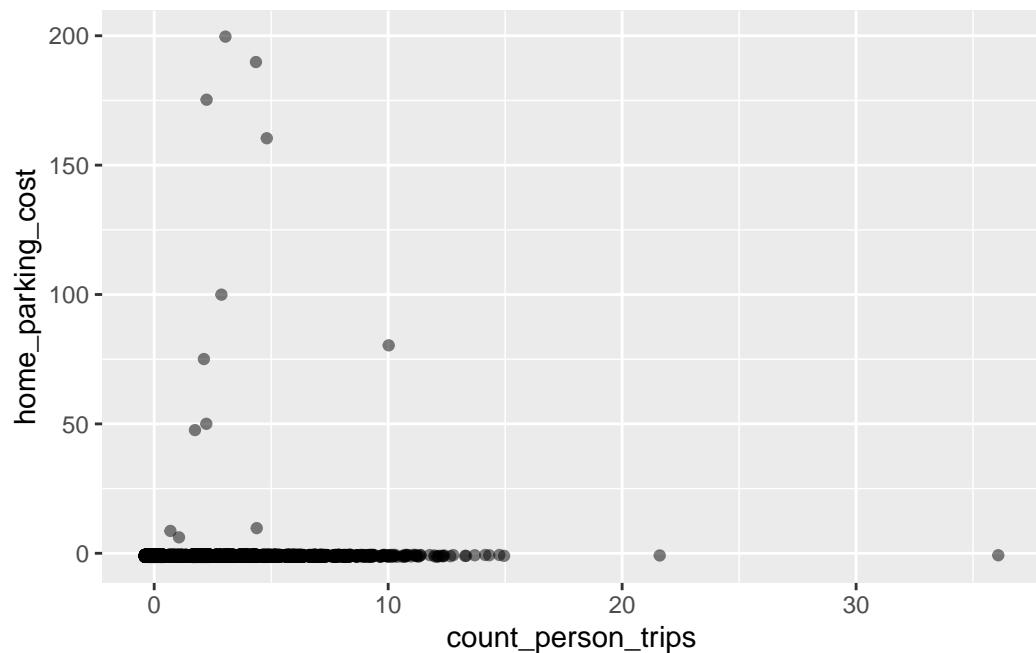


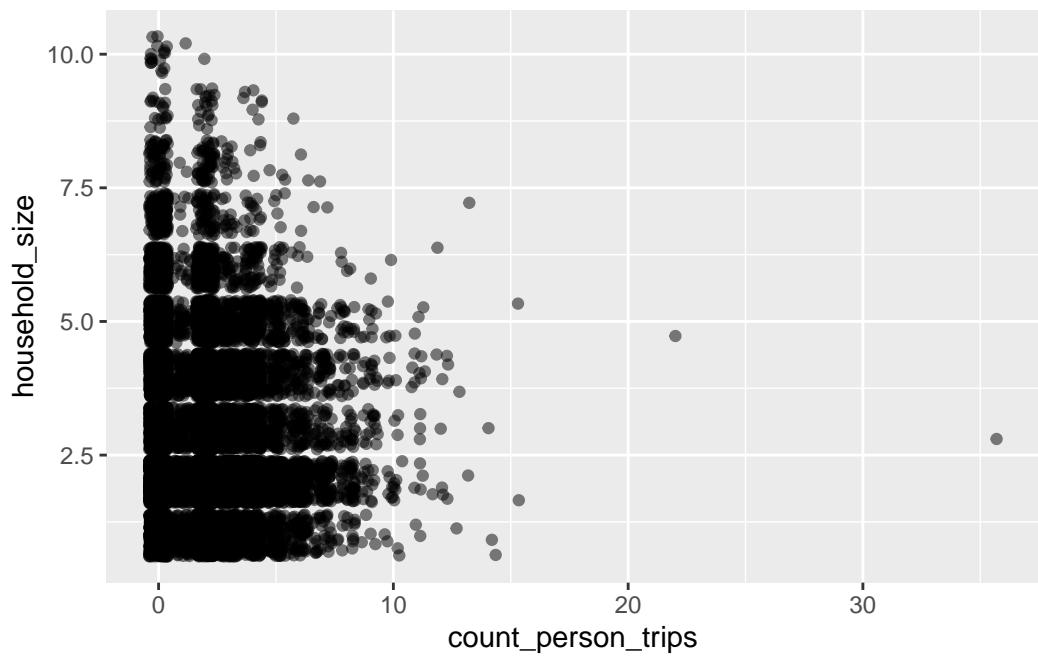
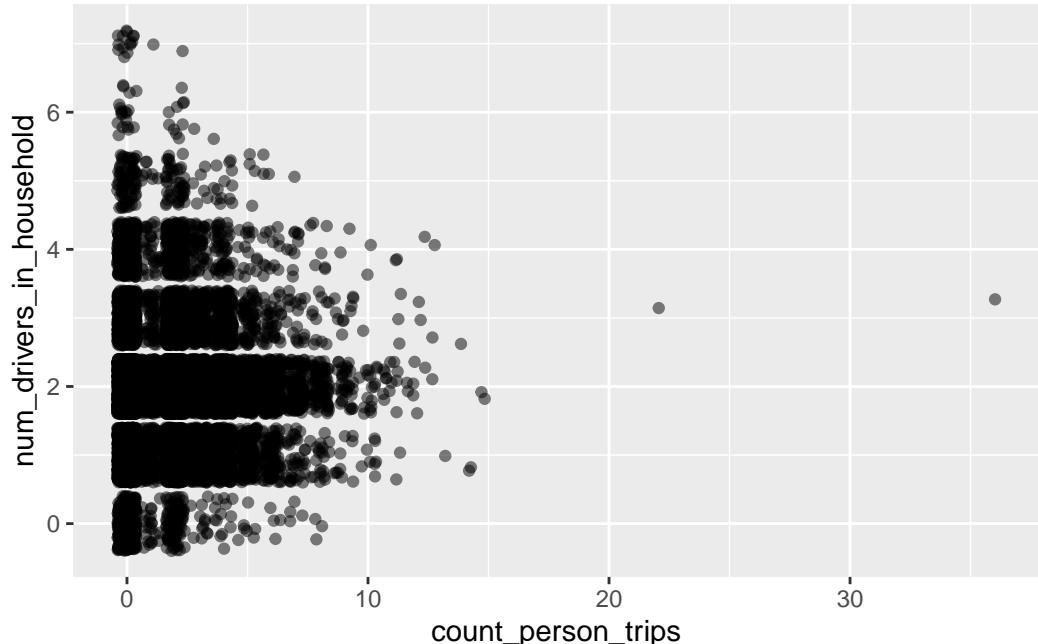


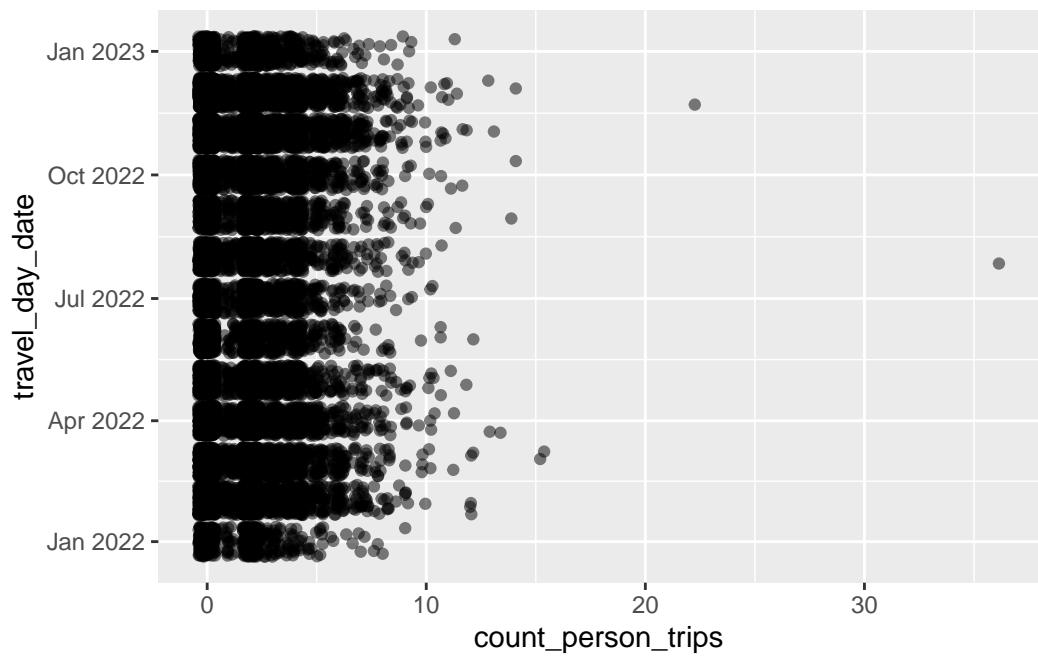
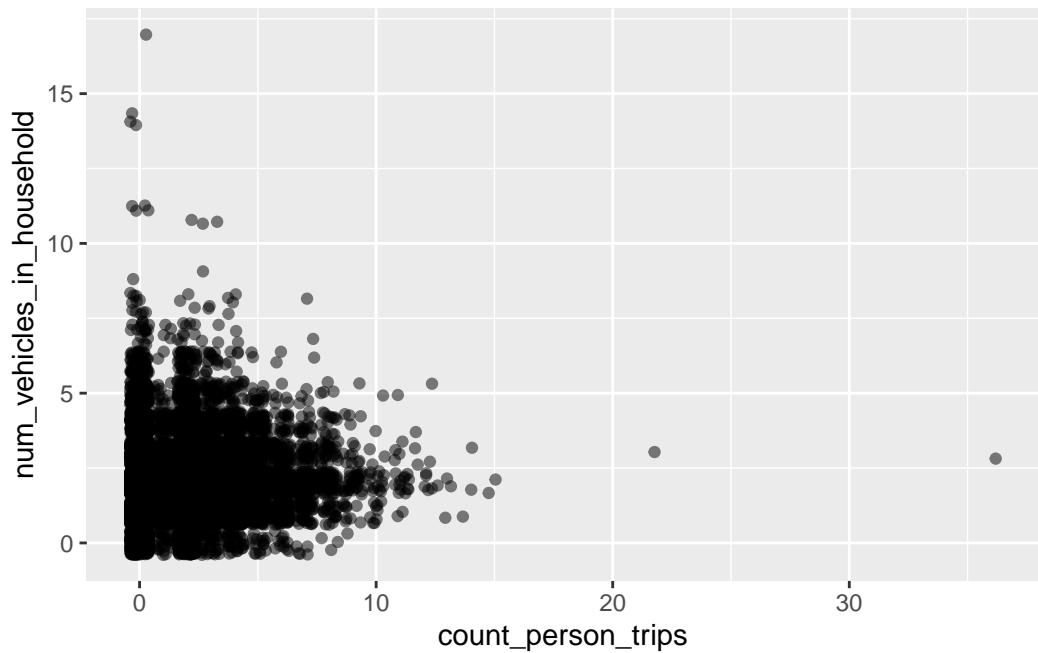


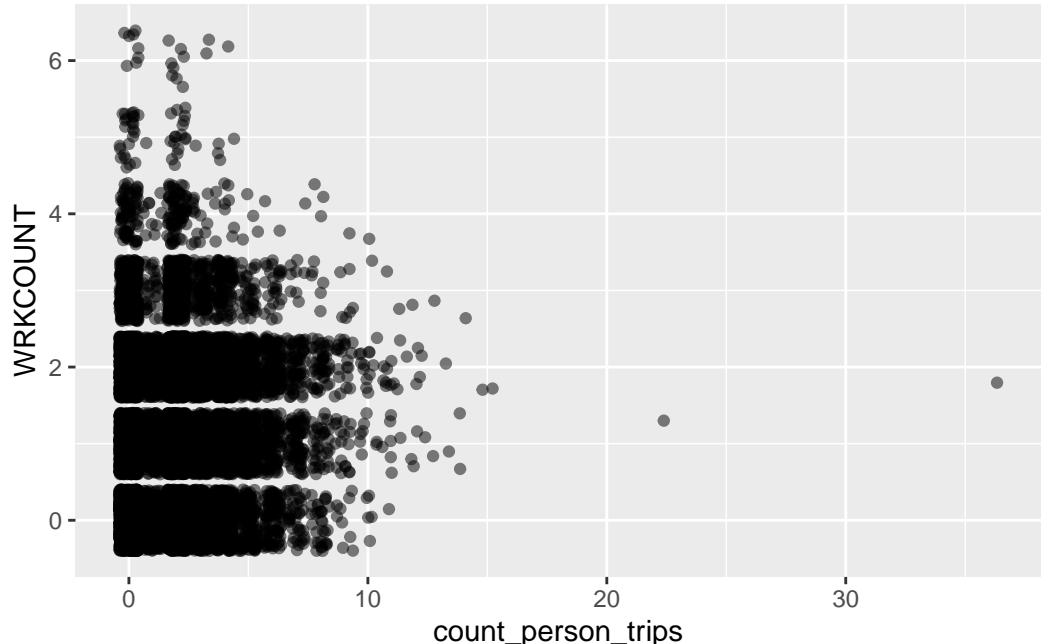




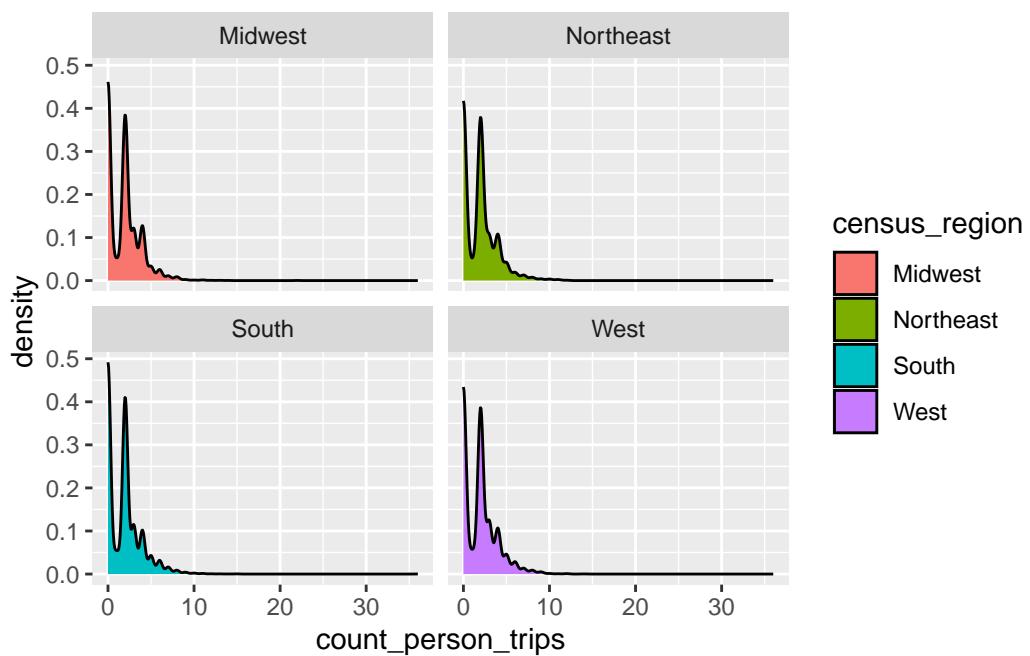
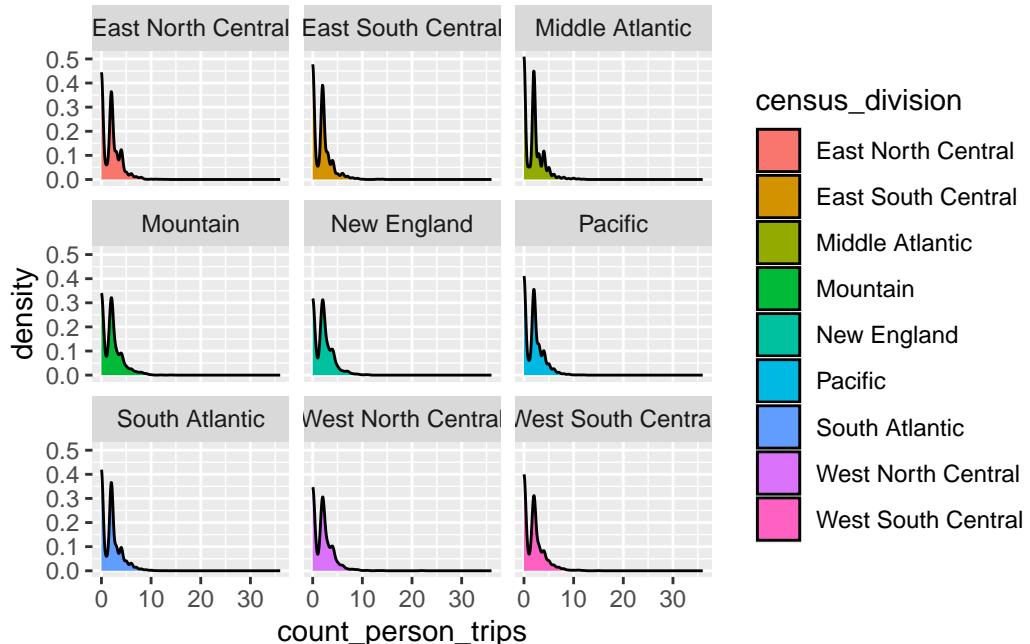


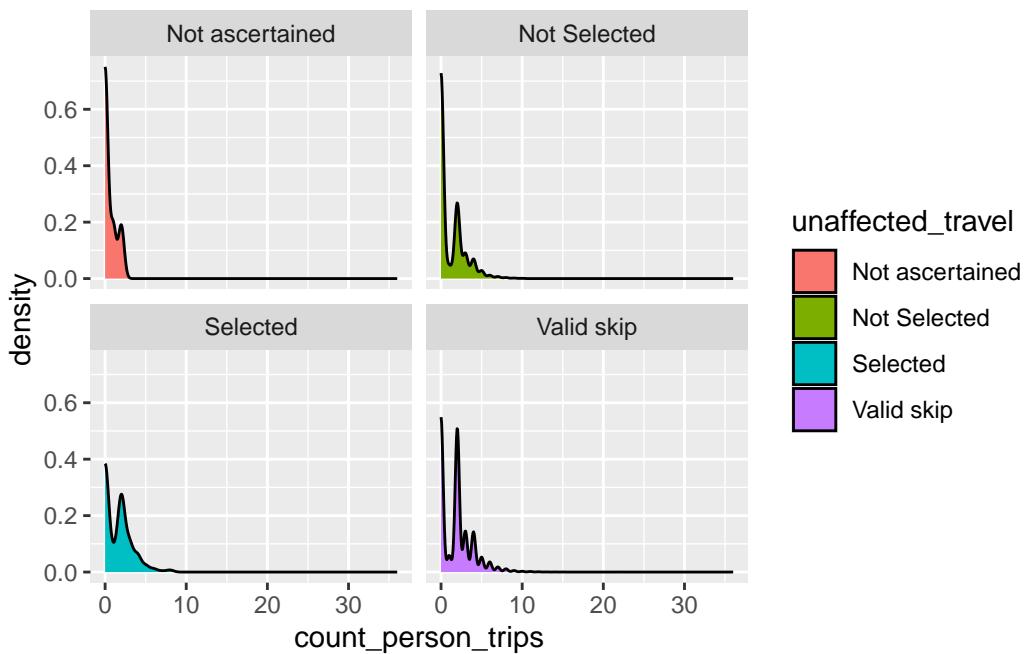
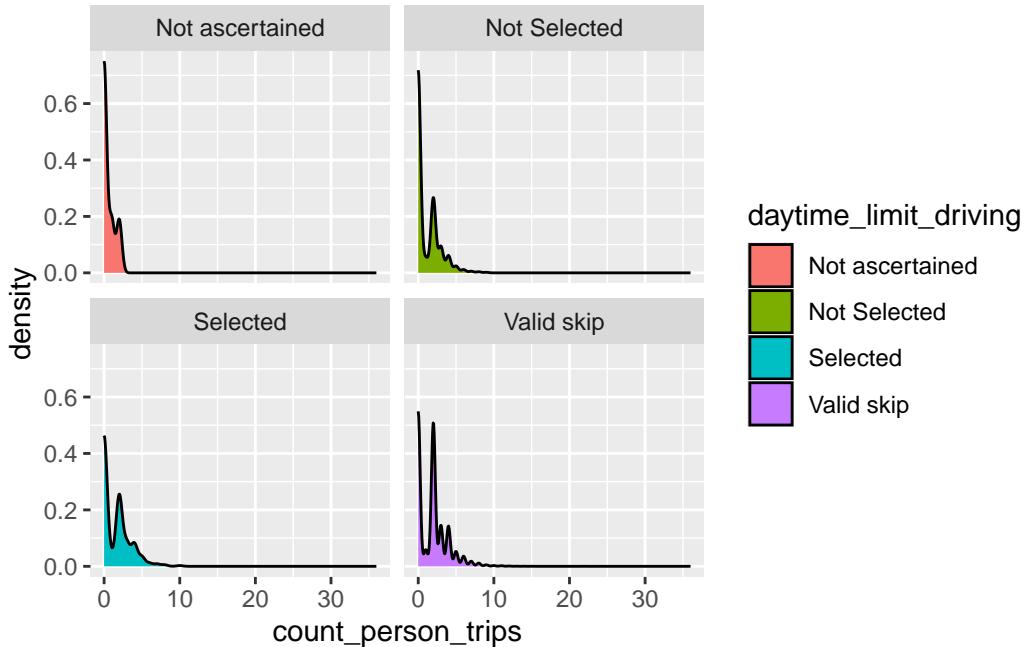


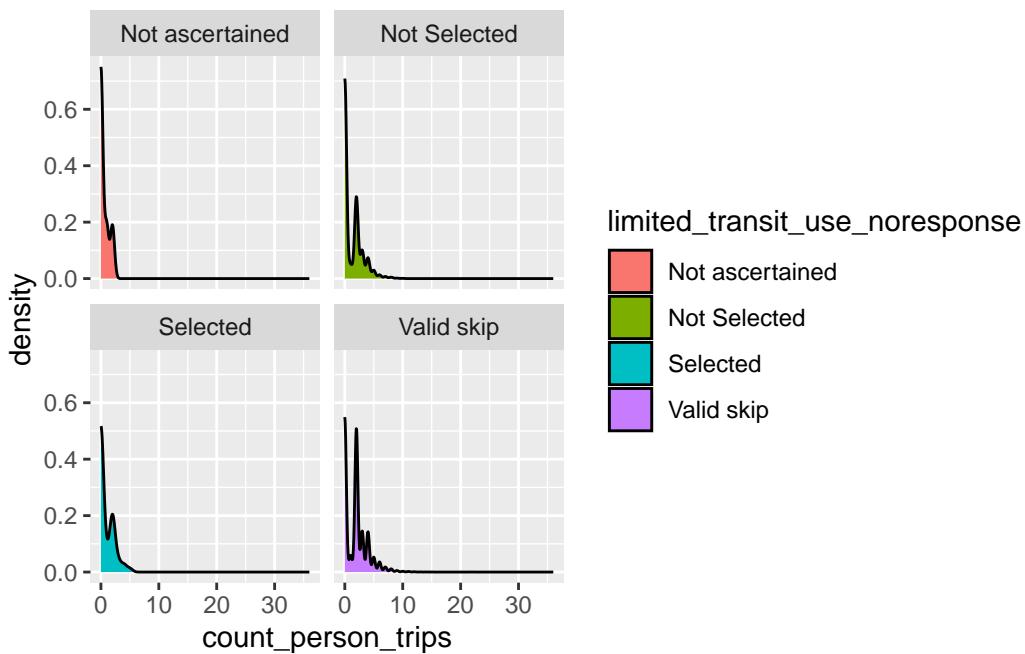
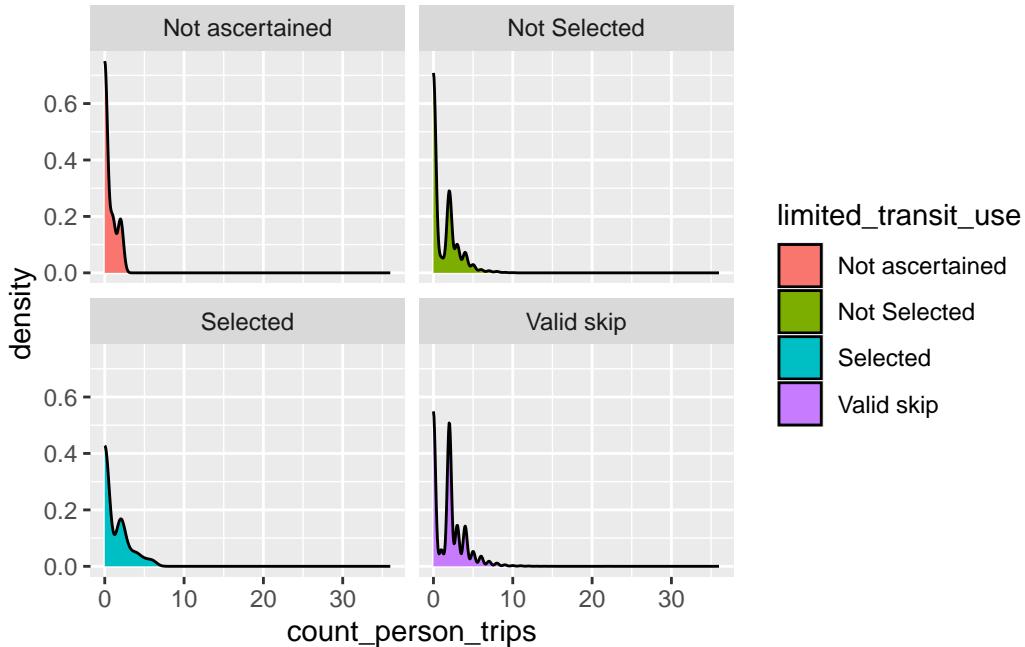


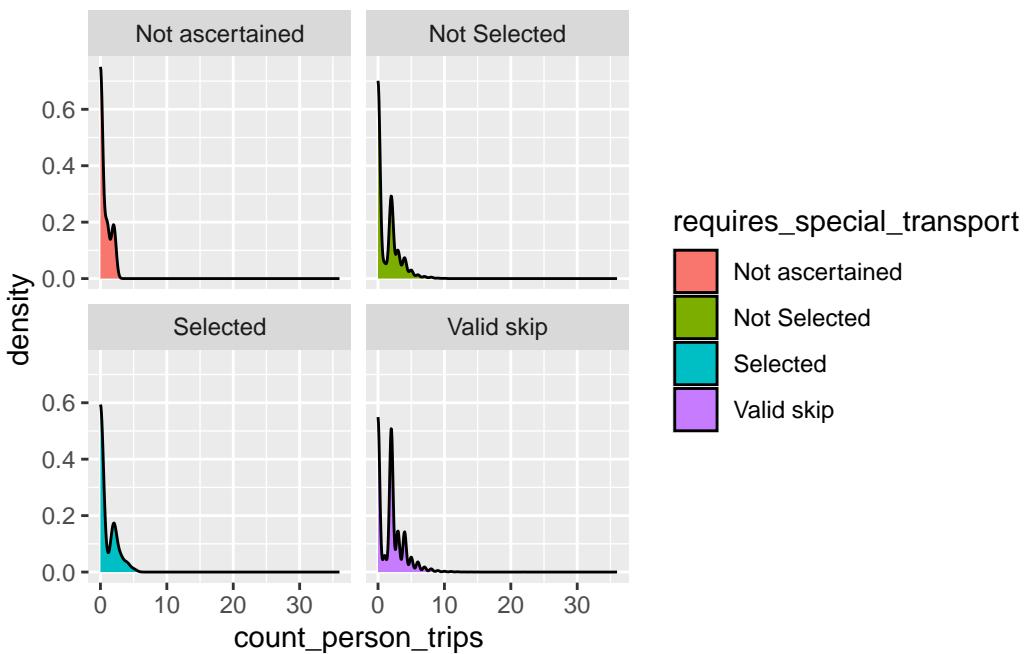
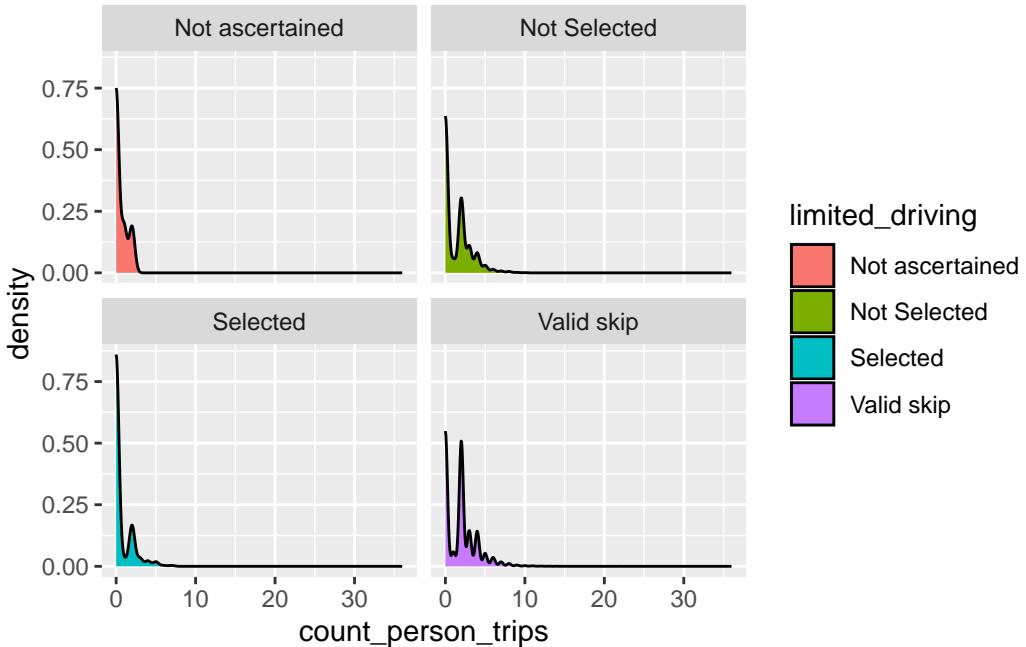


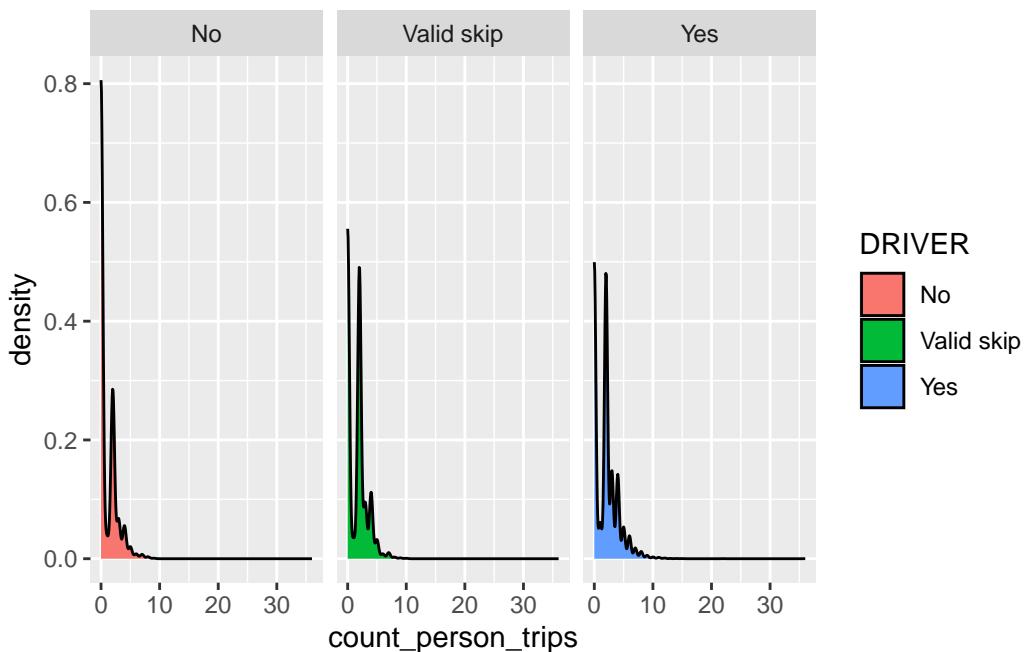
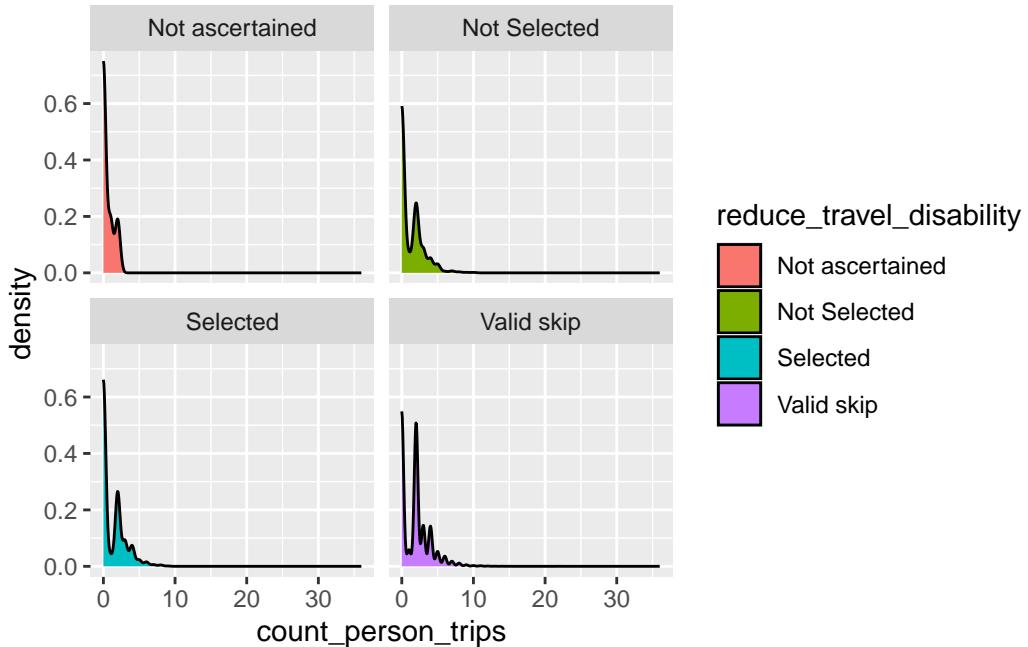
dst North Central MSA or CMSA of less than 1 million	new England MSA/CMSA of 1 million+
ast North Central MSA/CMSA of 1 million+ w/o heavy rail	New England Not in an MSA
ast North Central MSA/CMSA of 1 million+ with heavy rail	Pacific MSA of less than 1 million
ast North Central Not in an MSA	Pacific MSA/CMSA of 1 million+ w/o heavy rail
ast South Central MSA of less than 1 million	Pacific MSA/CMSA of 1 million+ with heavy rail
ast South Central MSA/CMSA of 1 million+ w/o heavy rail	Pacific Not in an MSA
ast South Central Not in an MSA	South Atlantic MSA of less than 1 million
lid-Atlantic MSA of less than 1 million	South Atlantic MSA/CMSA of 1 million
lid-Atlantic MSA/CMSA of 1 million+ w/o heavy rail	South Atlantic MSA/CMSA of 1 million+ w/o heavy rail
lid-Atlantic MSA/CMSA of 1 million+ with heavy rail	South Atlantic MSA/CMSA of 1 million+ with heavy rail
lid-Atlantic Not in an MSA	South Atlantic Not in an MSA
lountain MSA of less than 1 million	West North Central MSA of less than 1 million
lountain MSA/CMSA of 1 million+ w/o heavy rail	West North Central MSA/CMSA of 1 million+ w/o heavy rail
lountain Not in an MSA	West North Central Not in an MSA
ew England MSA of less than 1 million	West South Central MSA of less than 1 million
	West South Central MSA/CMSA of 1 million

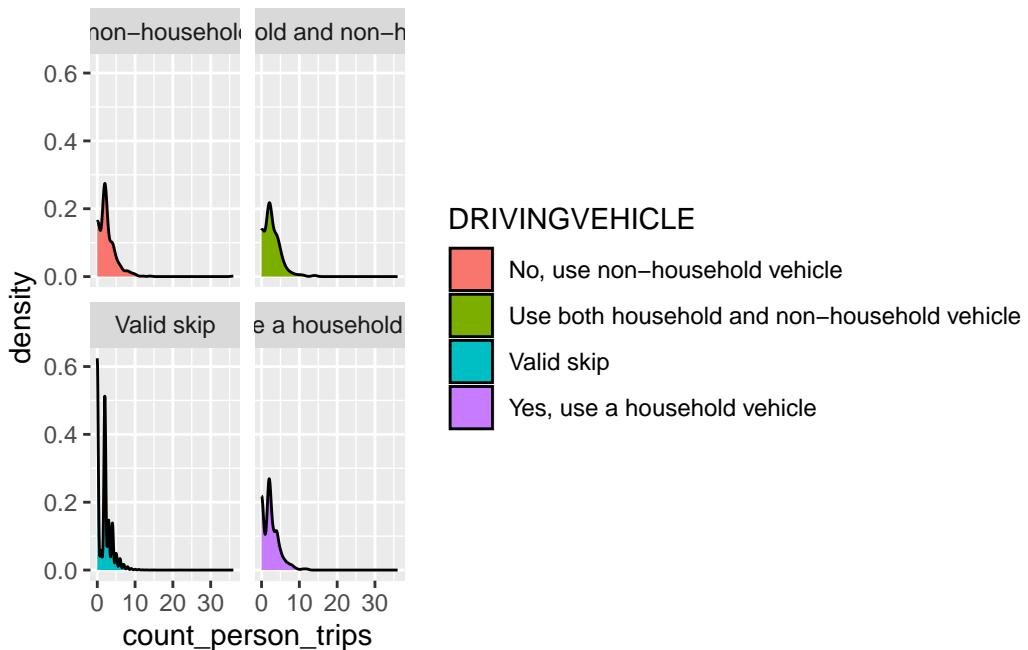
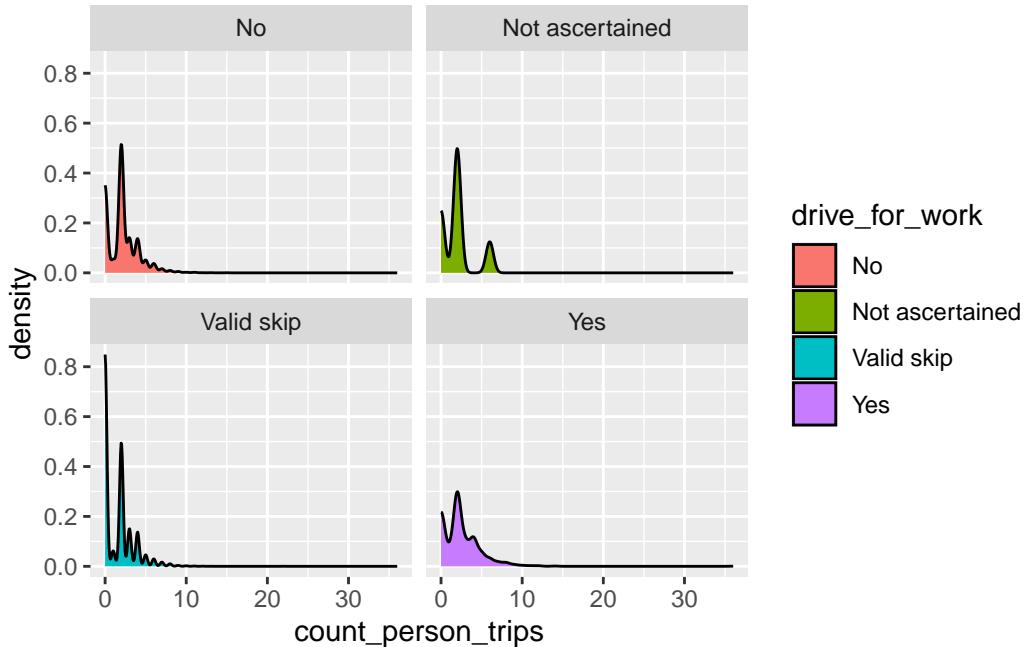


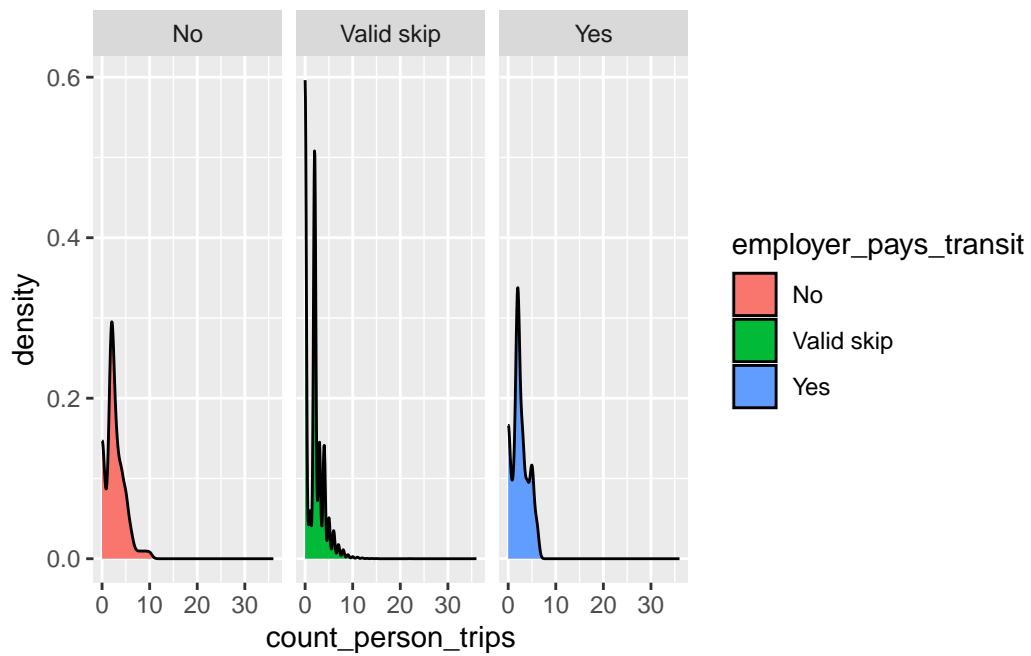
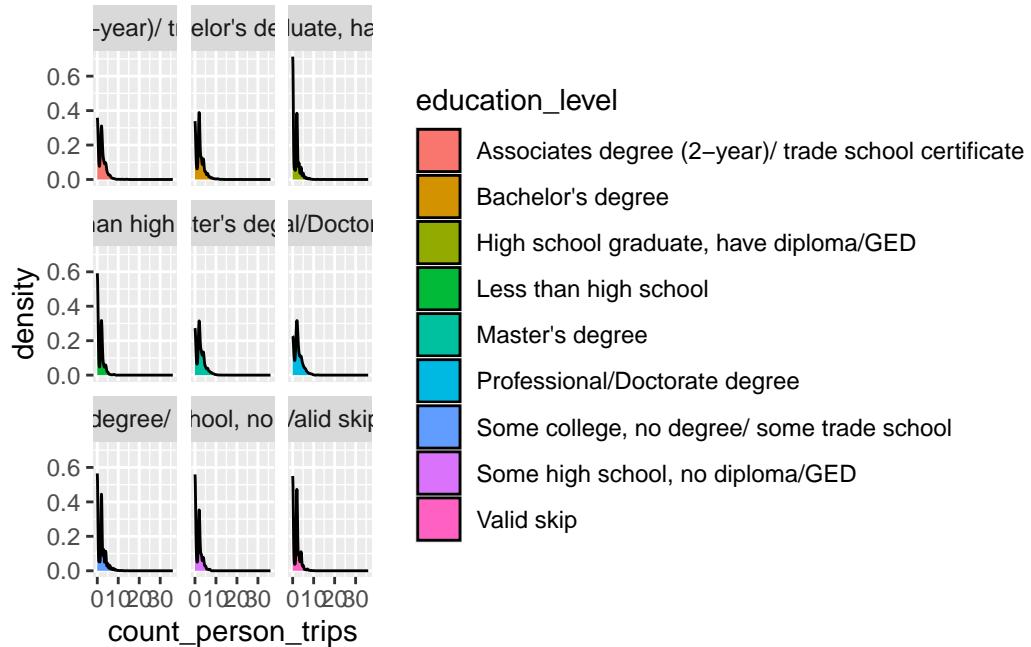


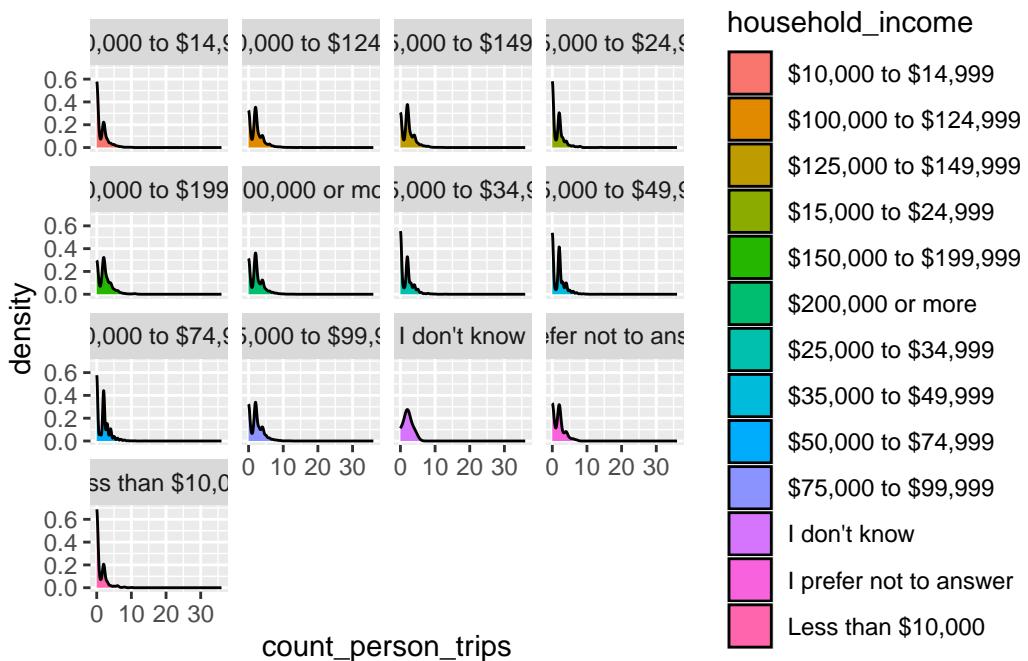
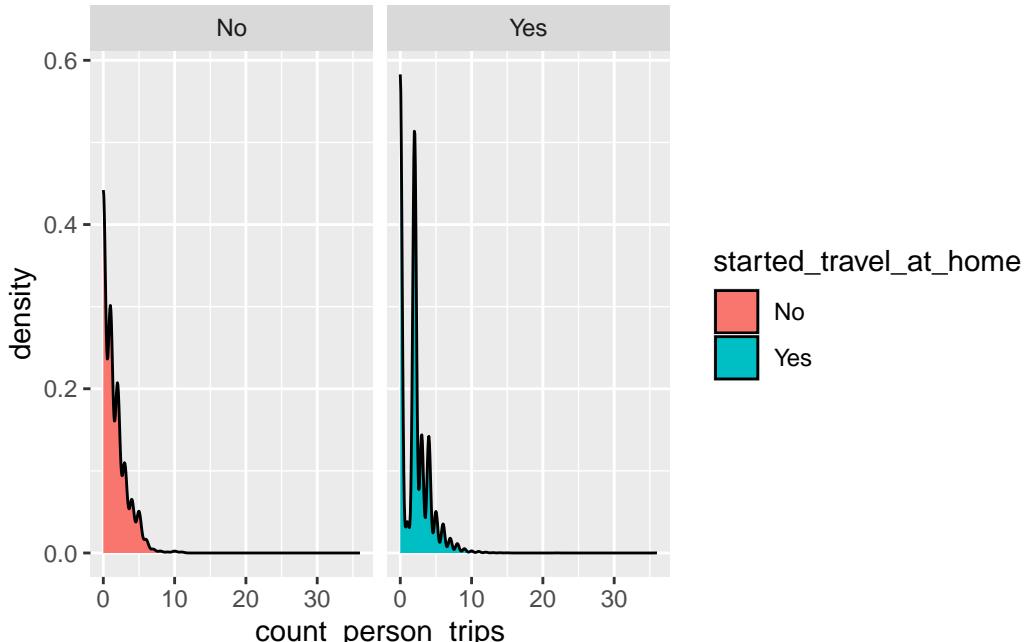


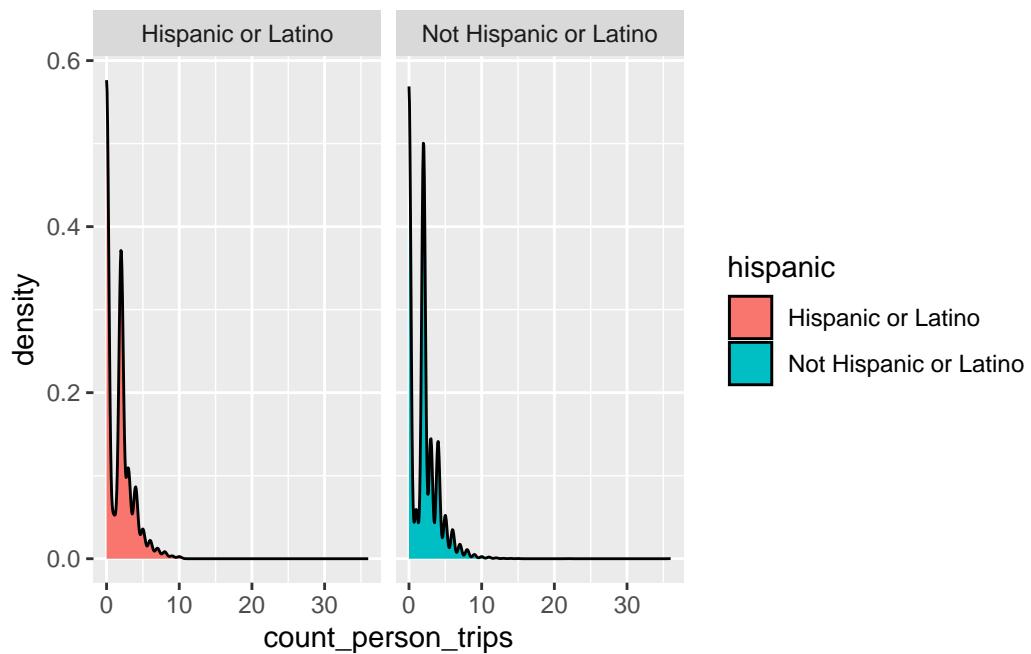
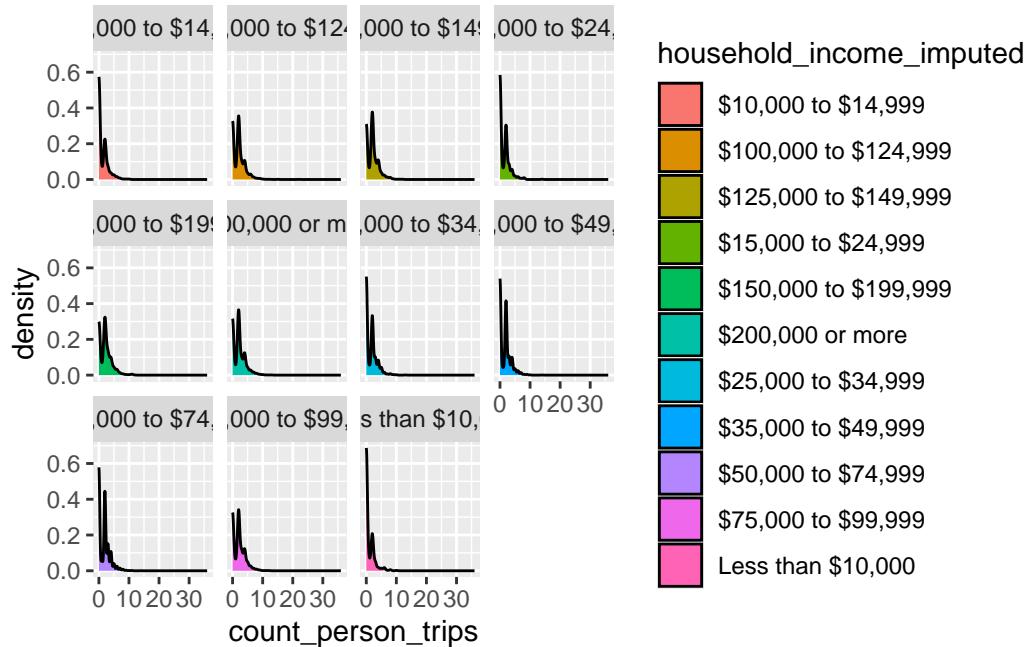


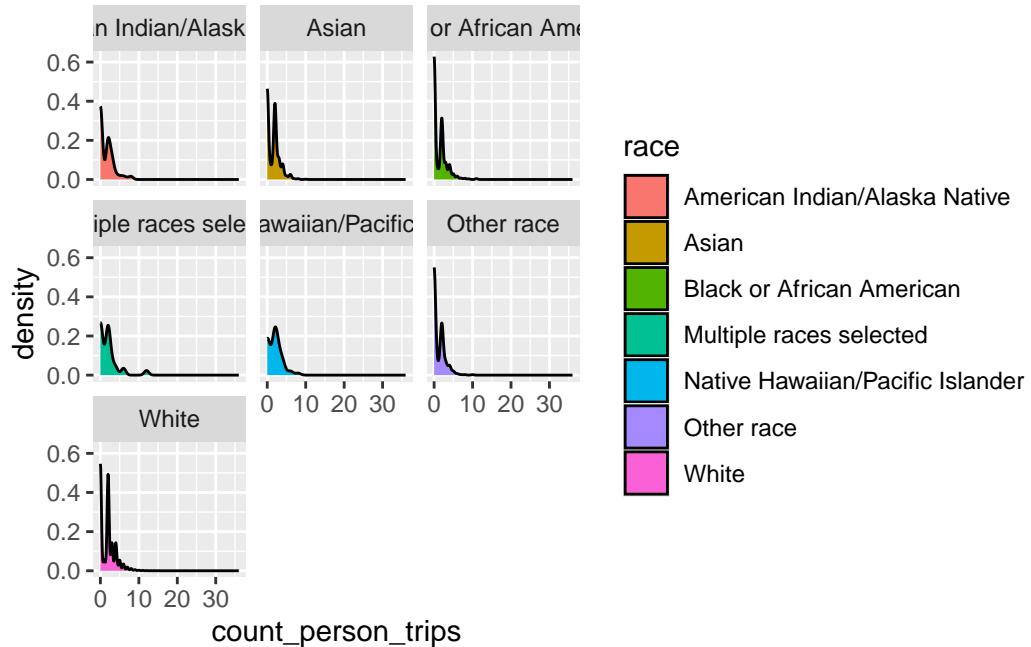


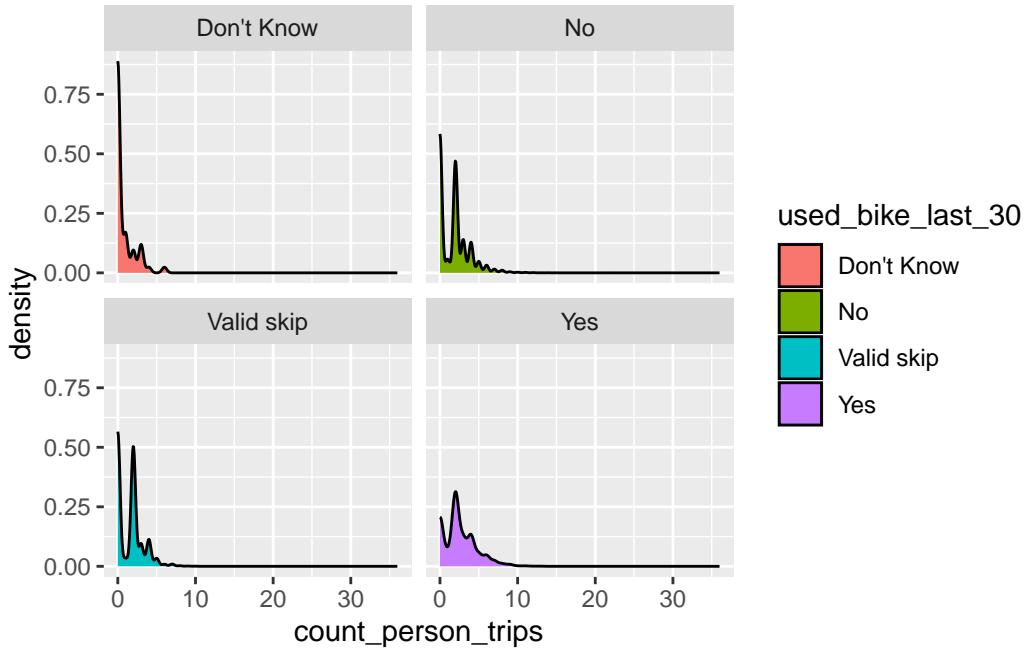






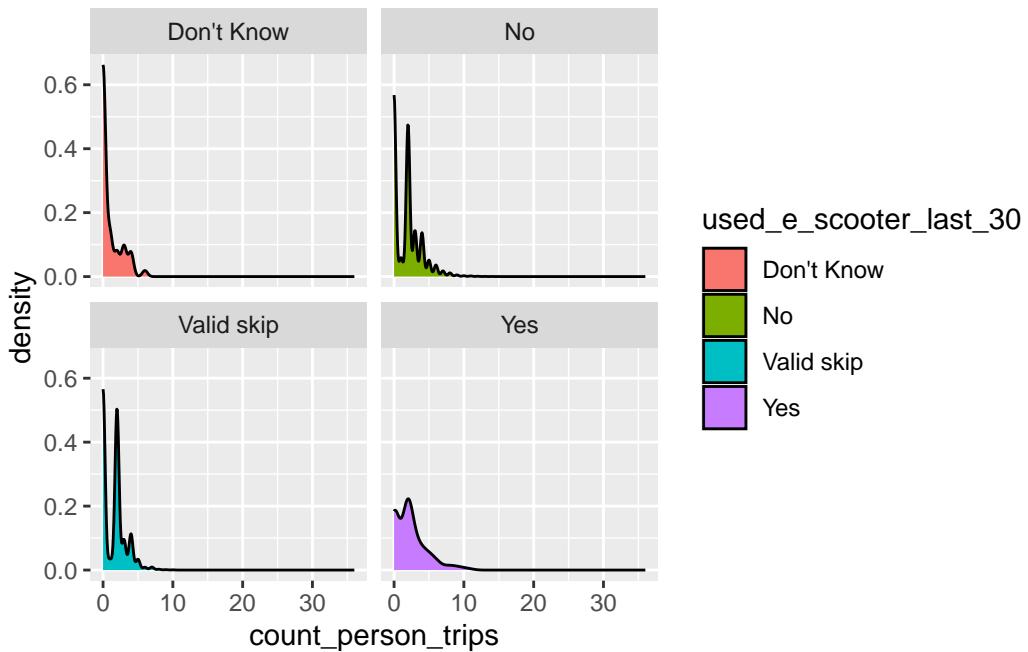
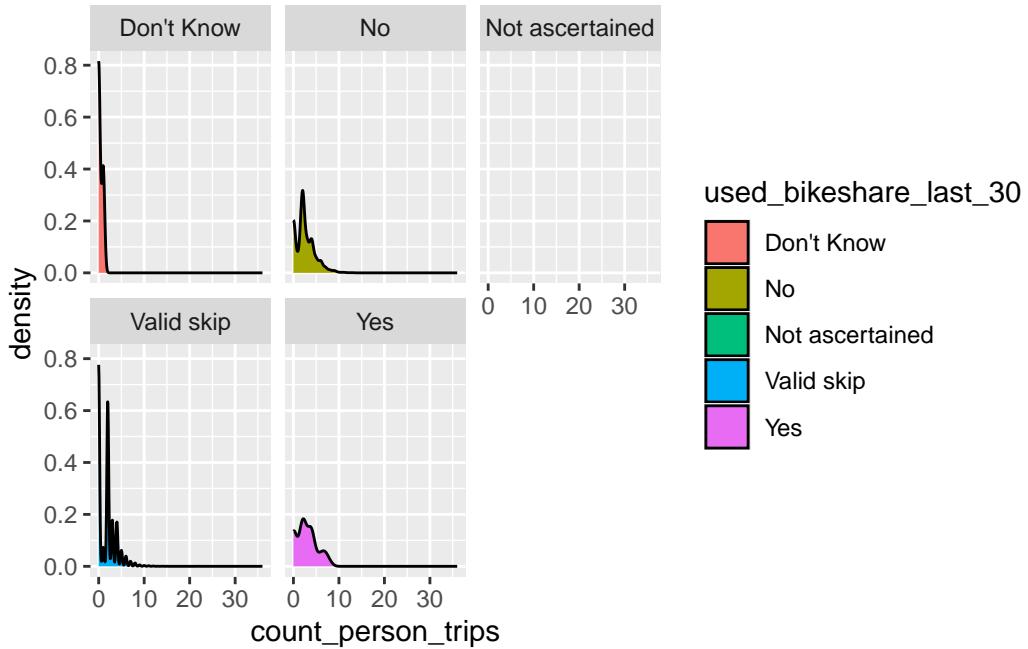


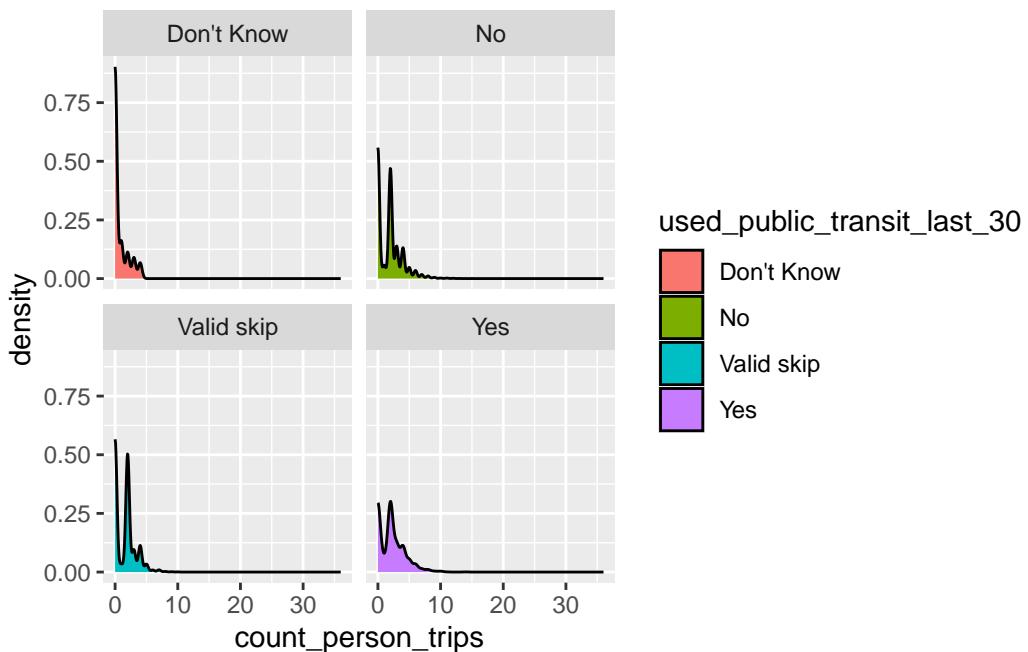
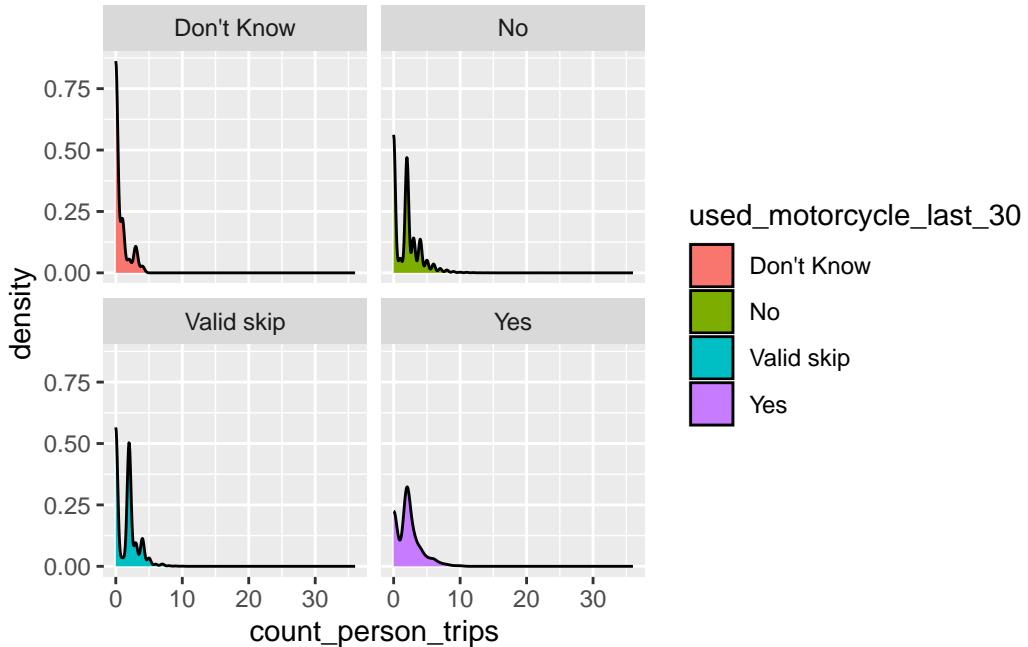


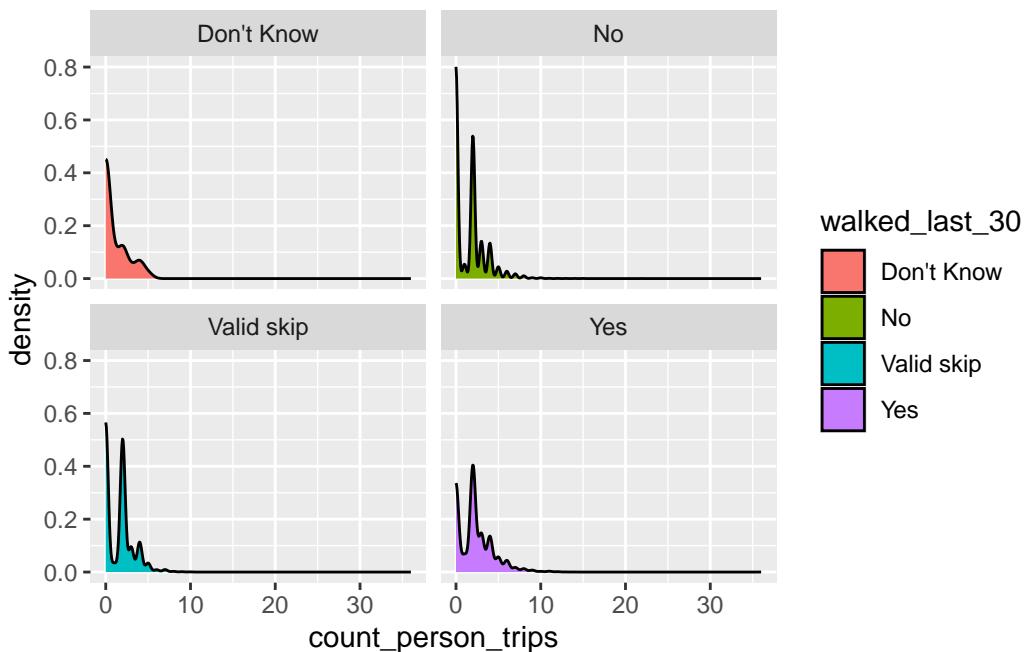
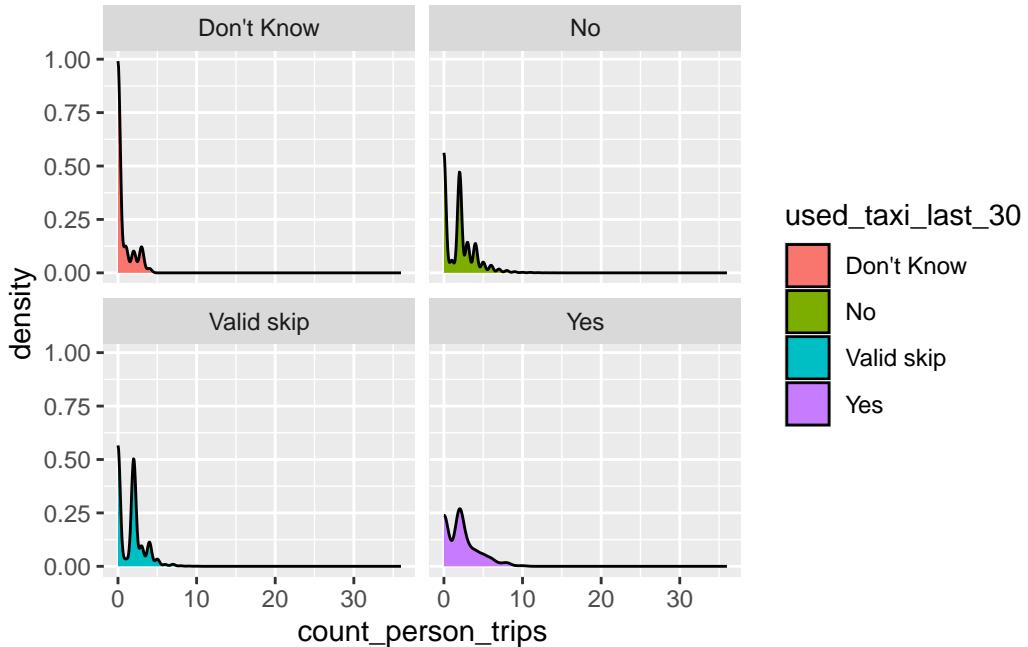


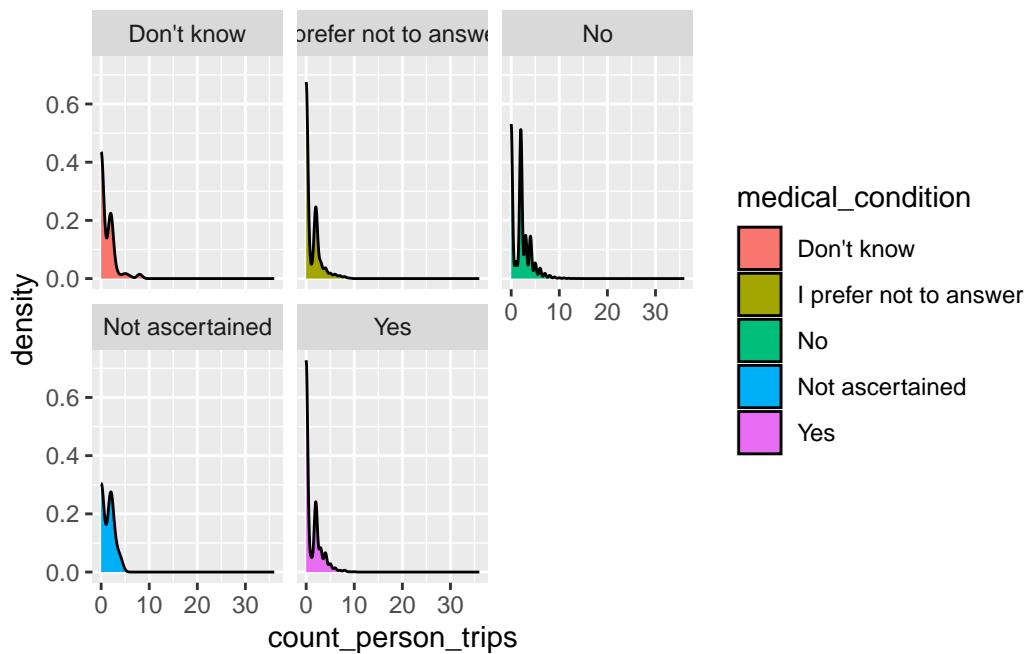
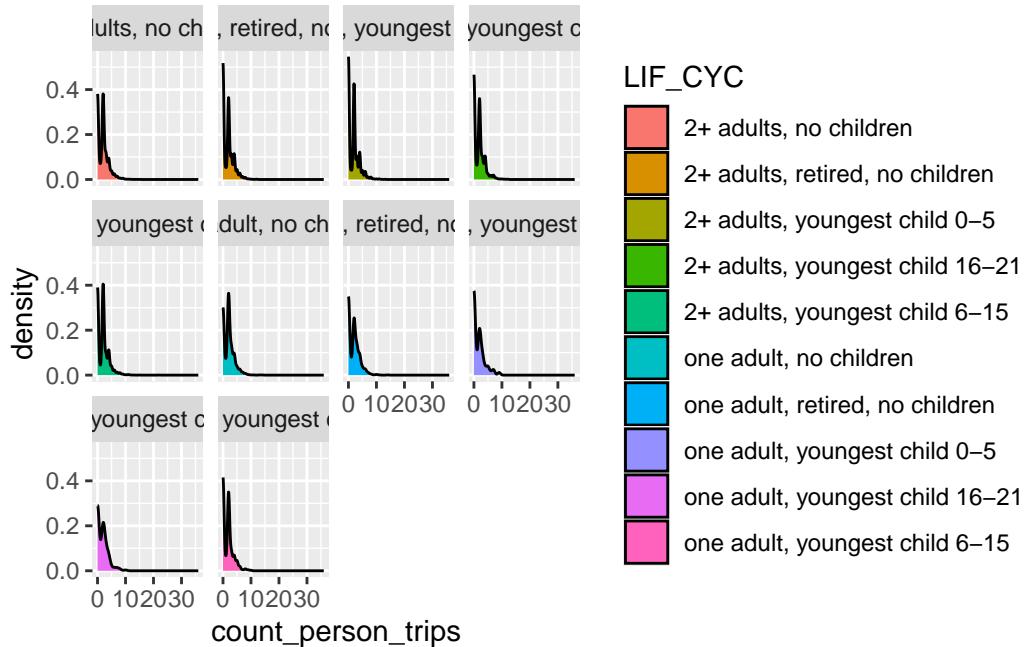
Warning: Groups with fewer than two data points have been dropped.

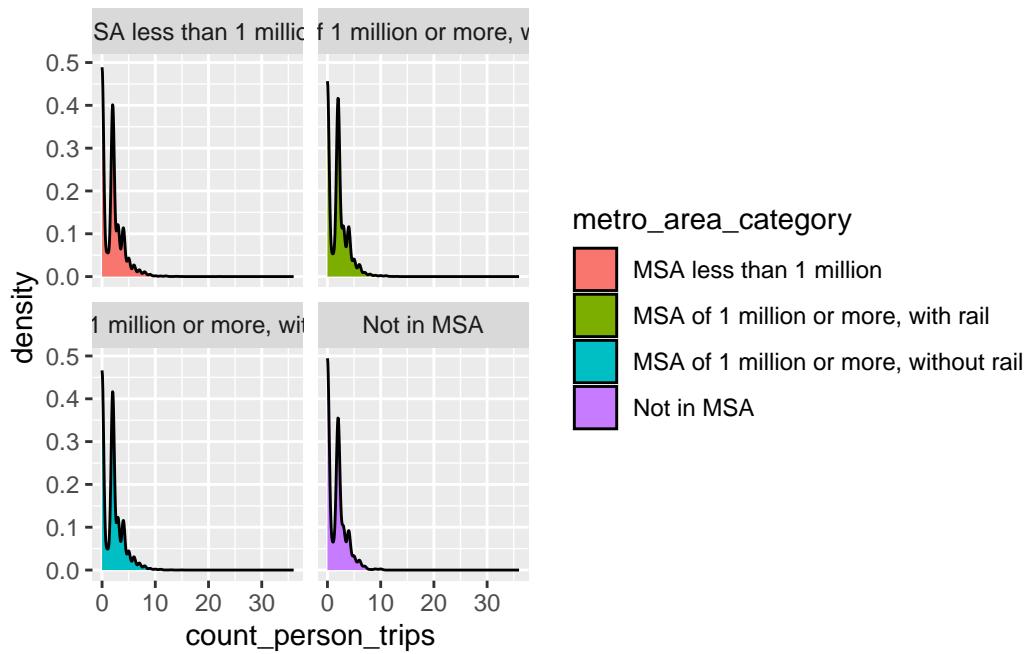
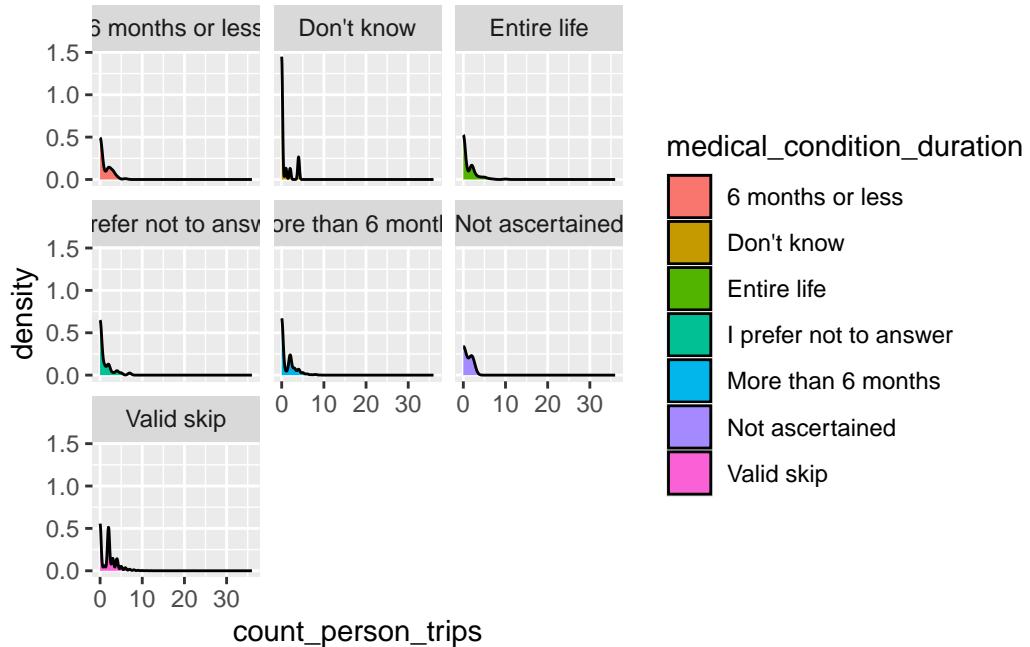
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning -Inf

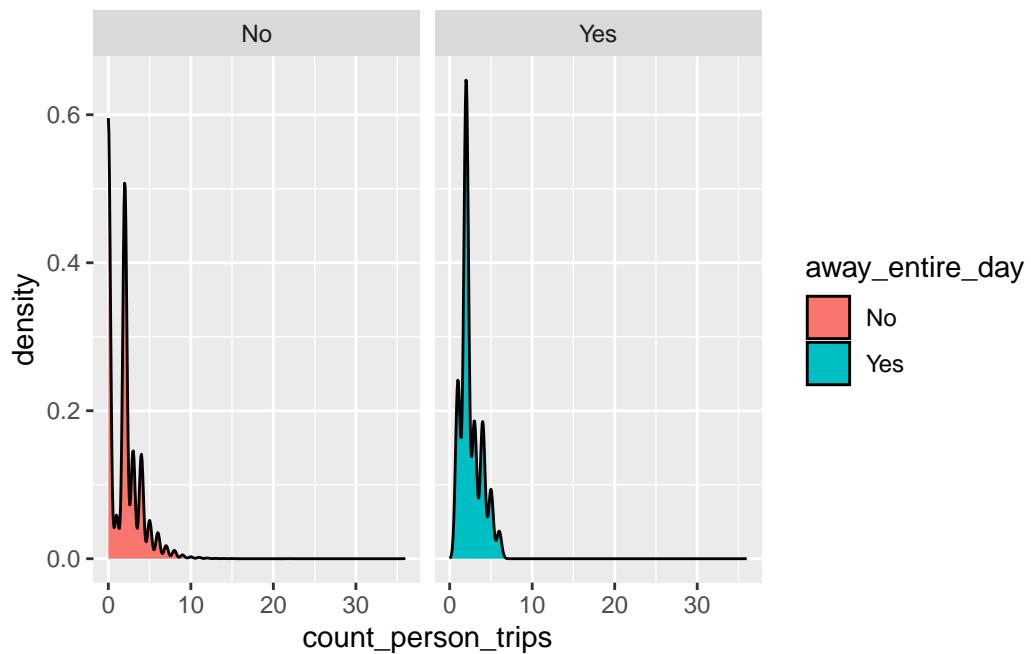
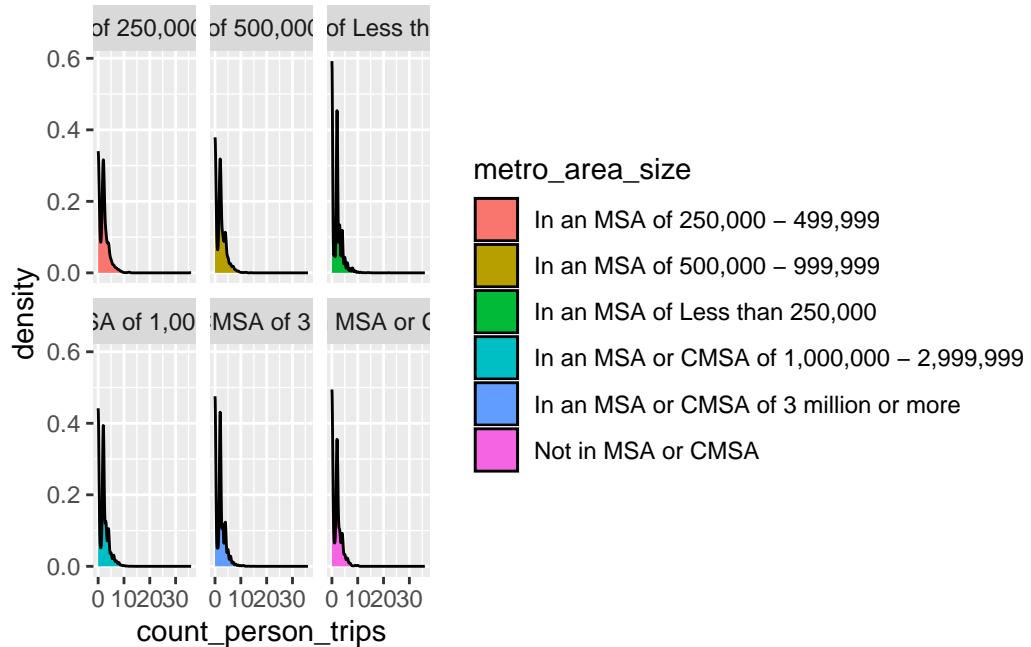


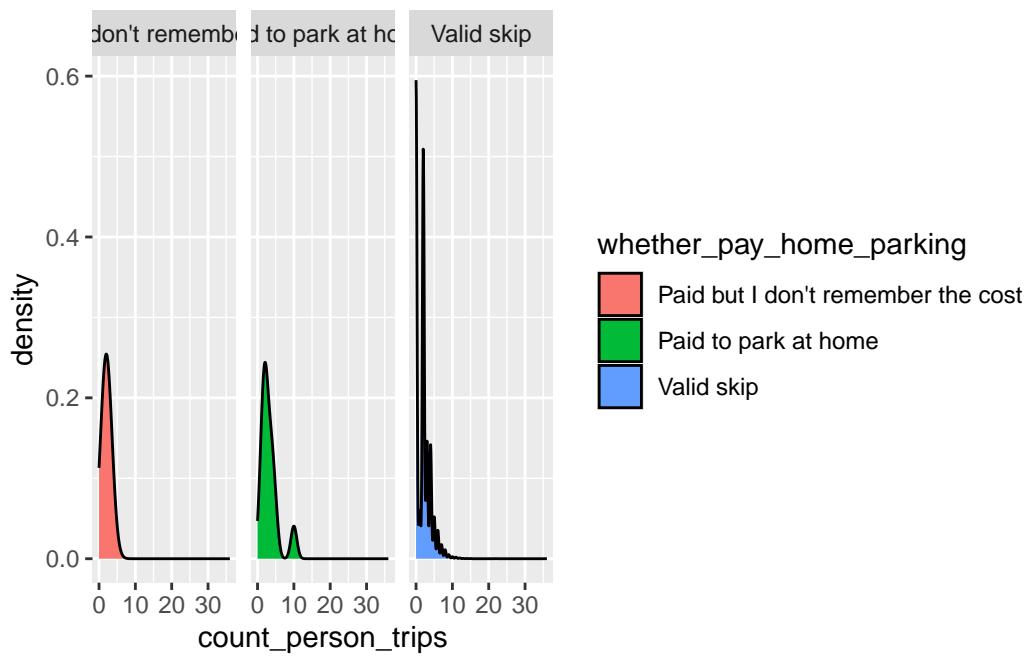
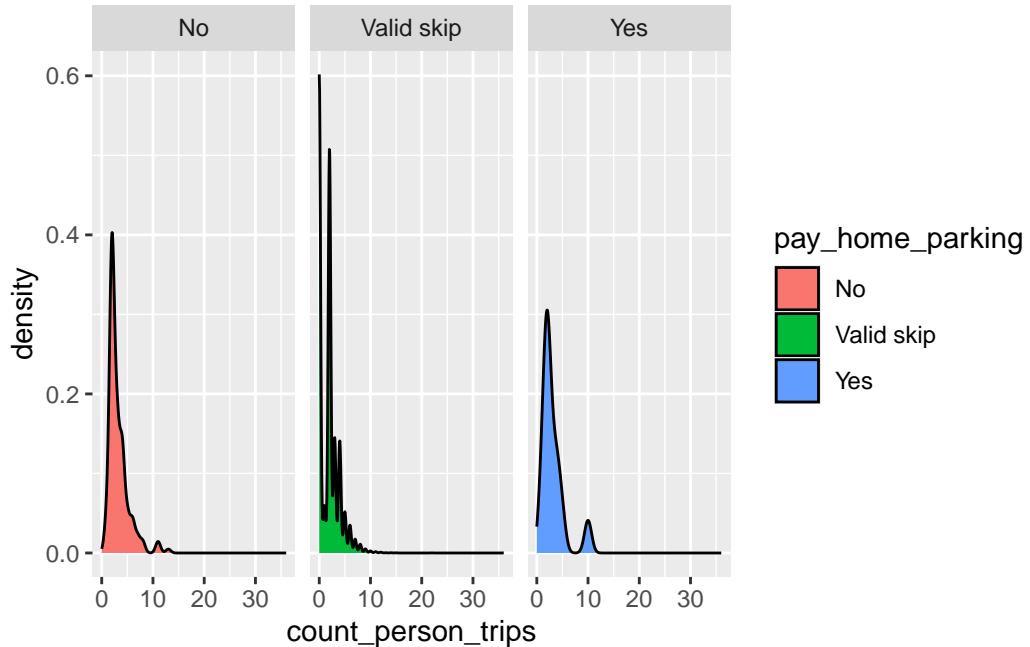


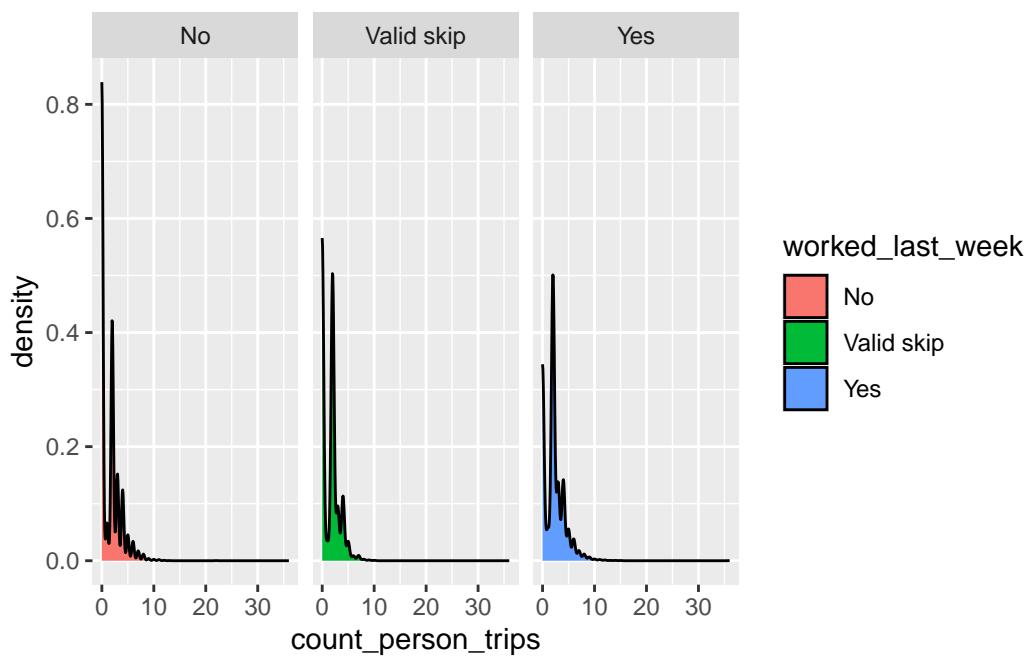
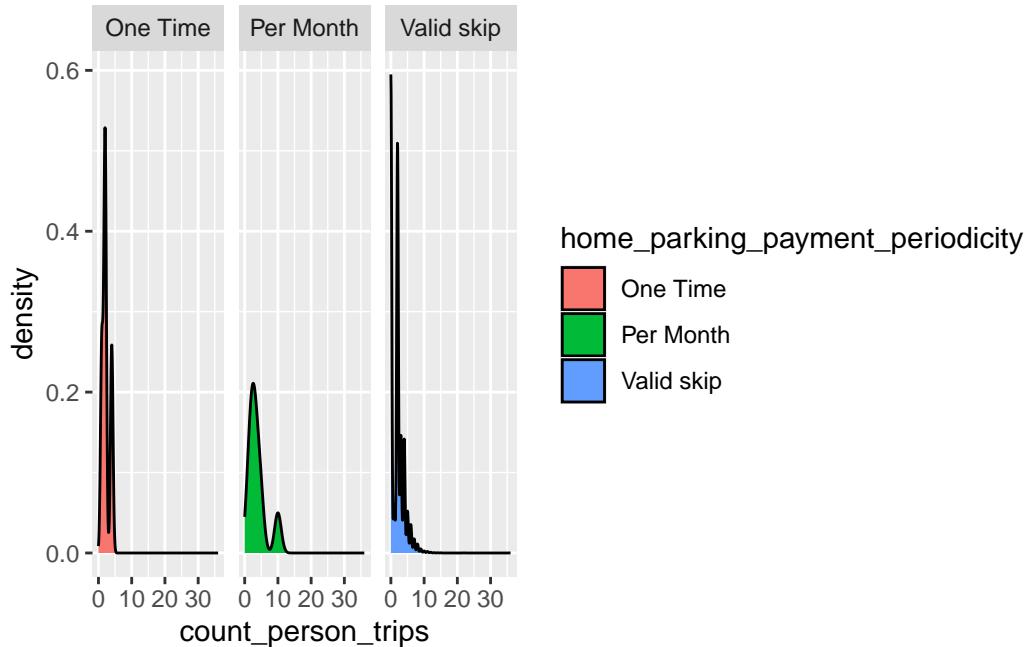


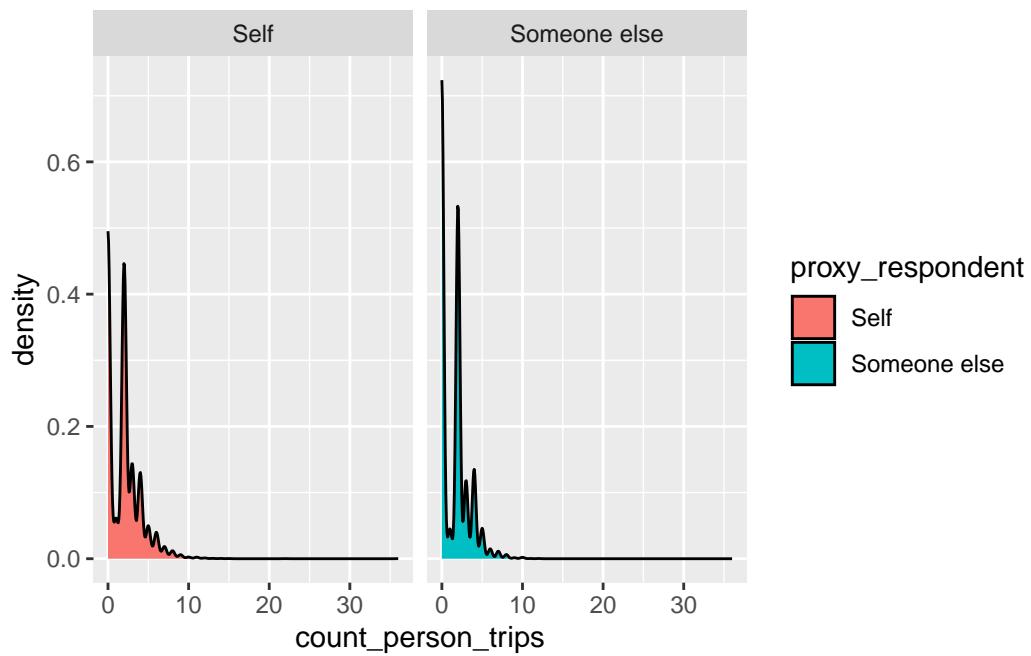
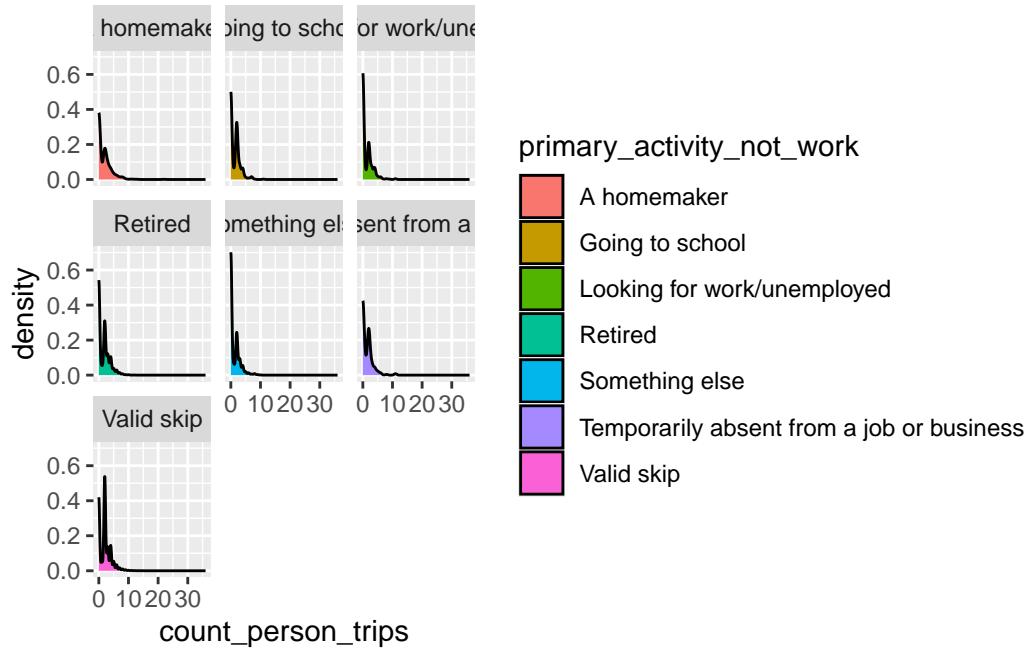


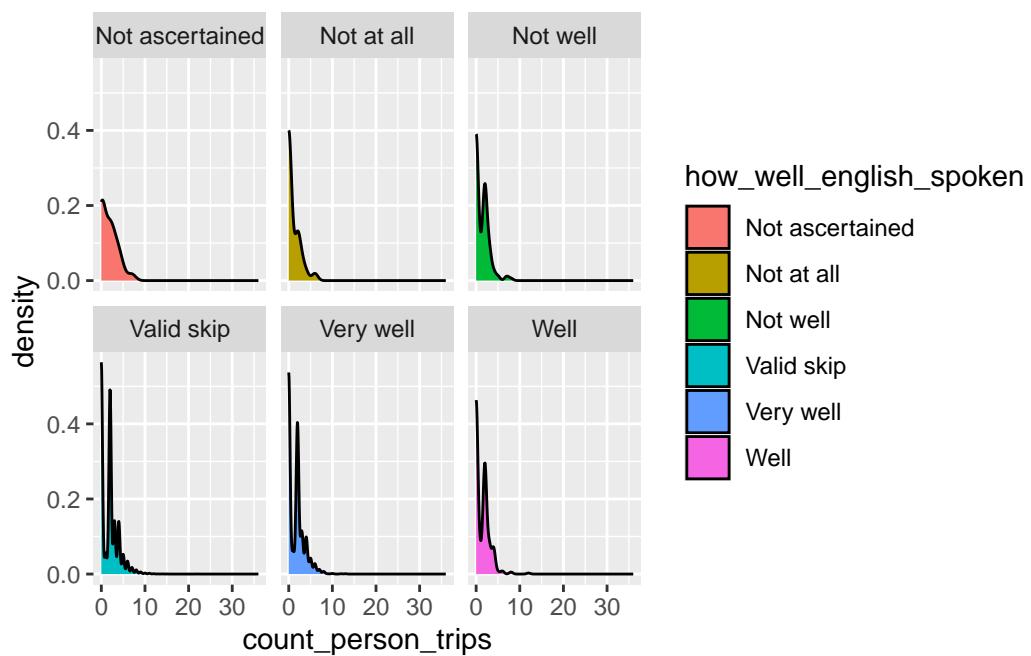
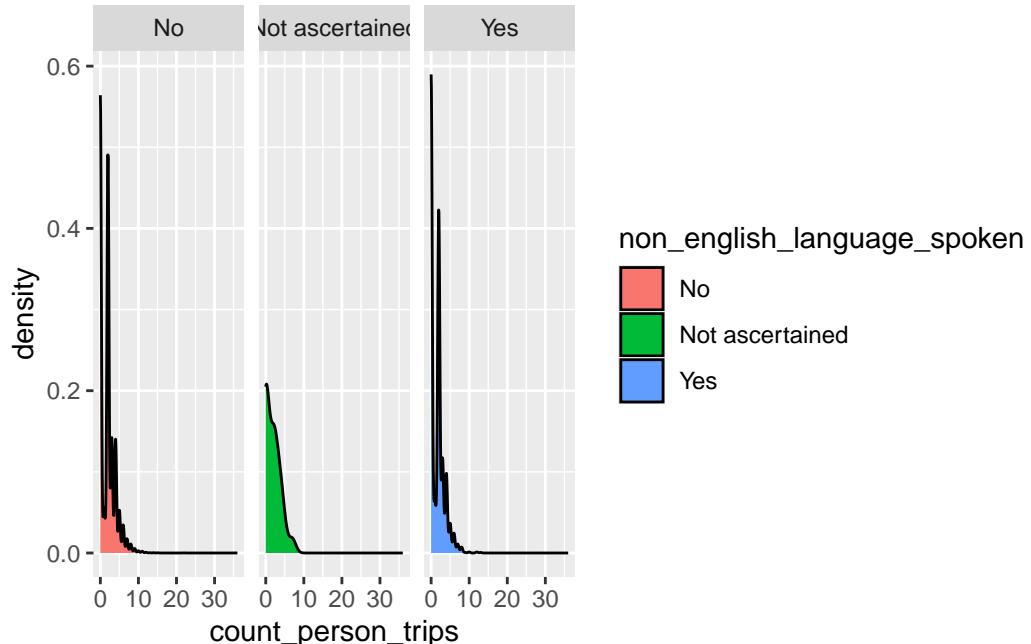


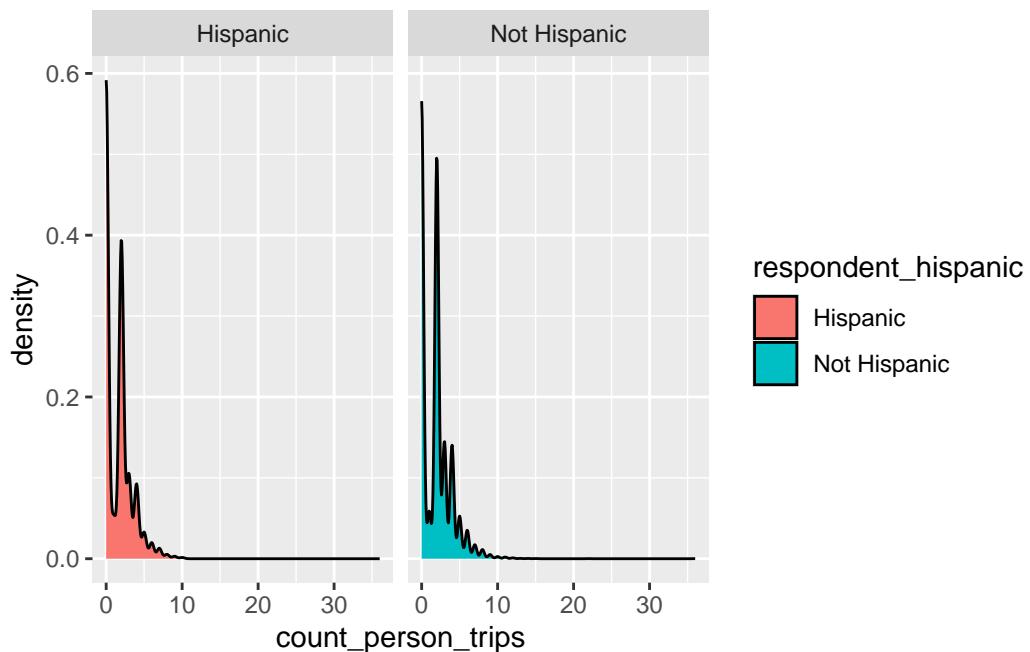
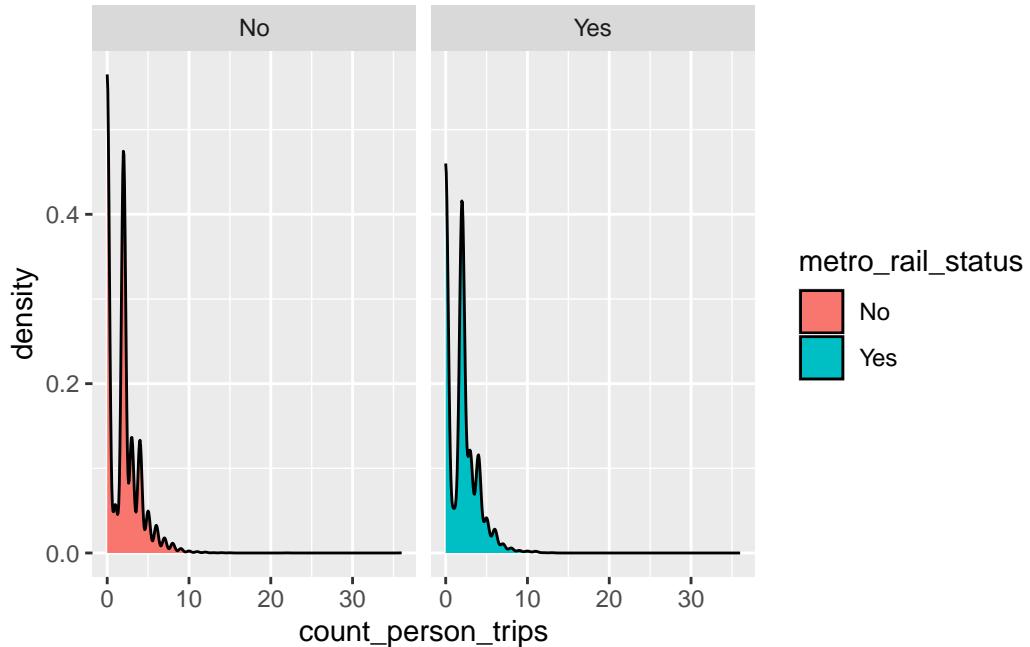


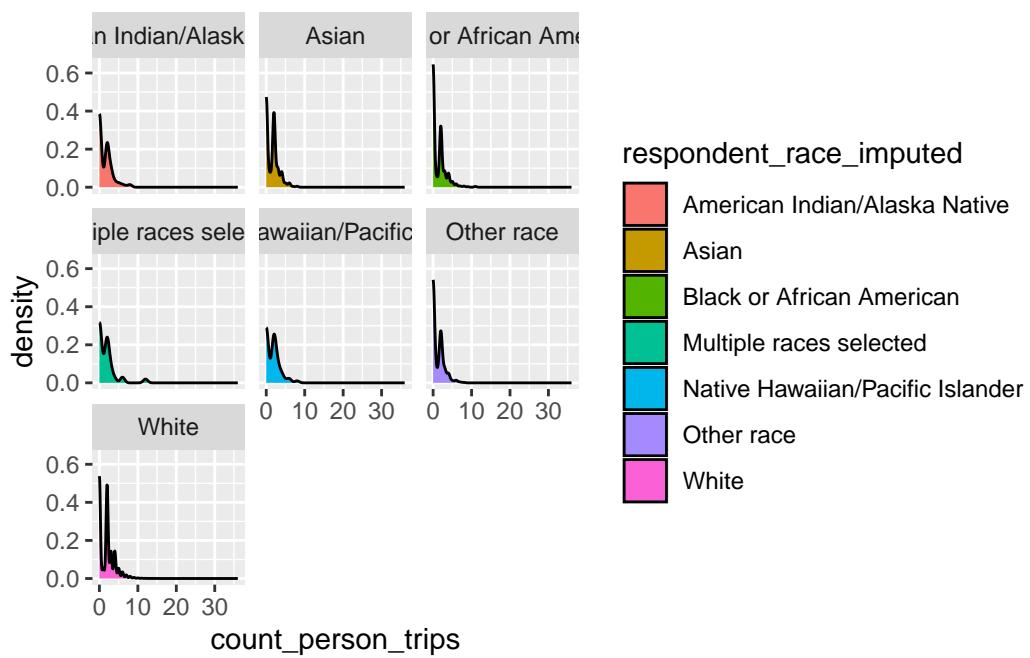
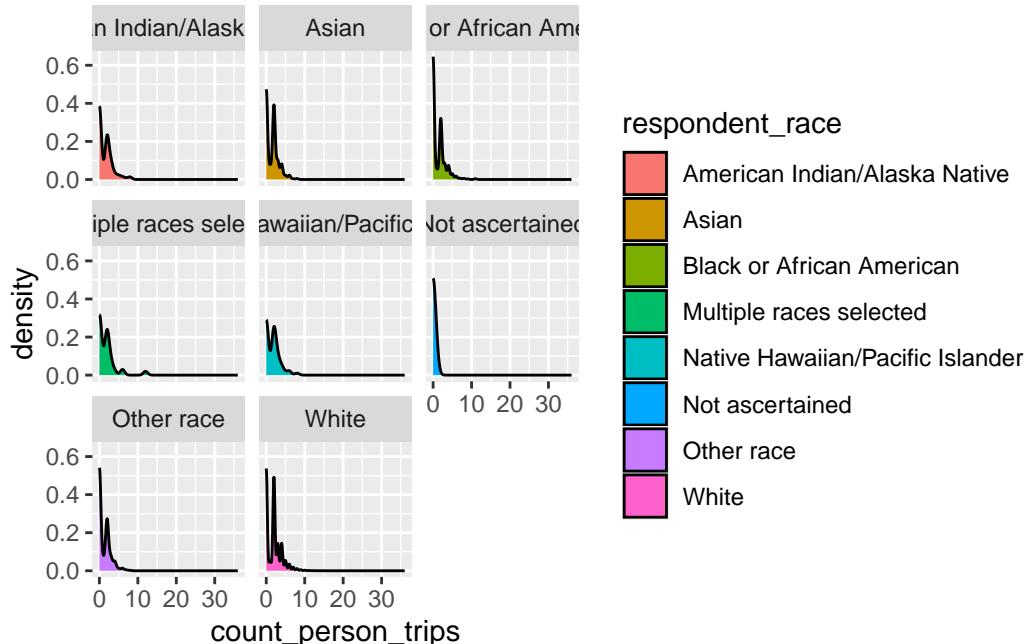


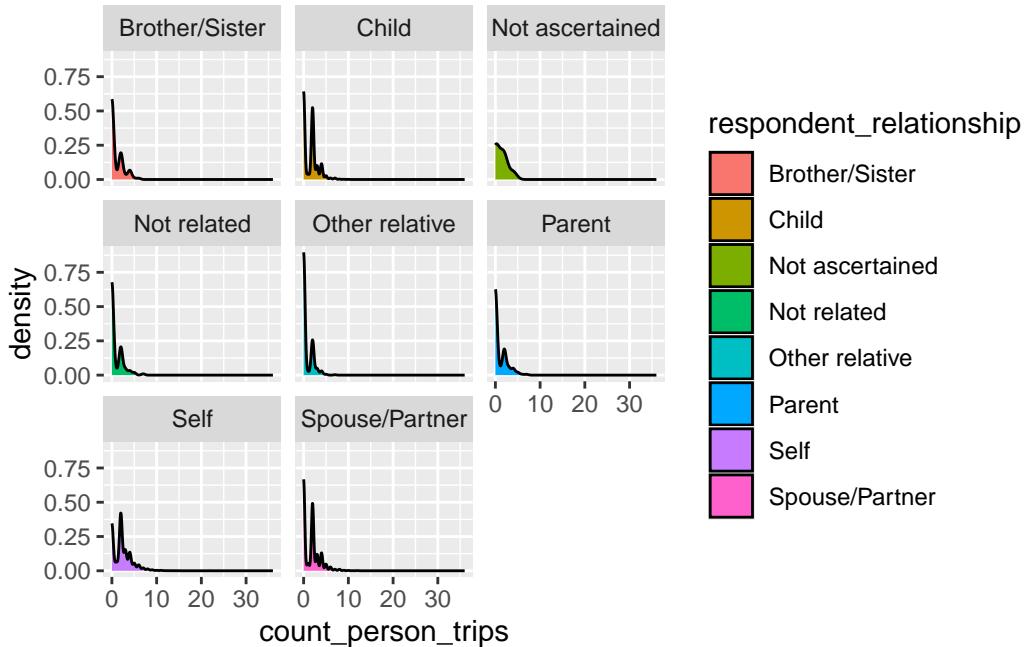




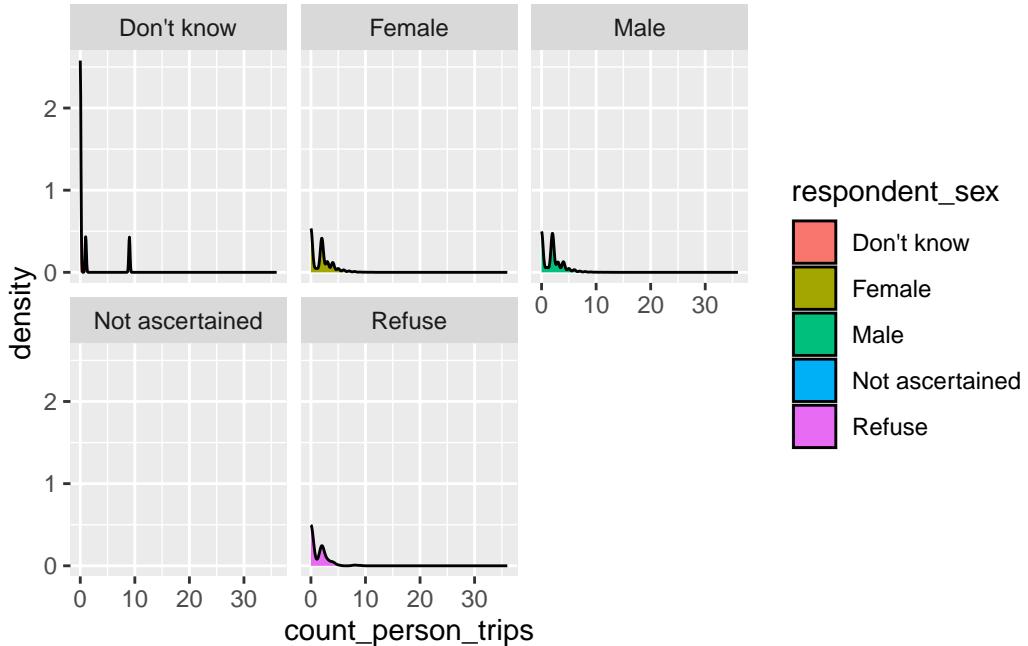


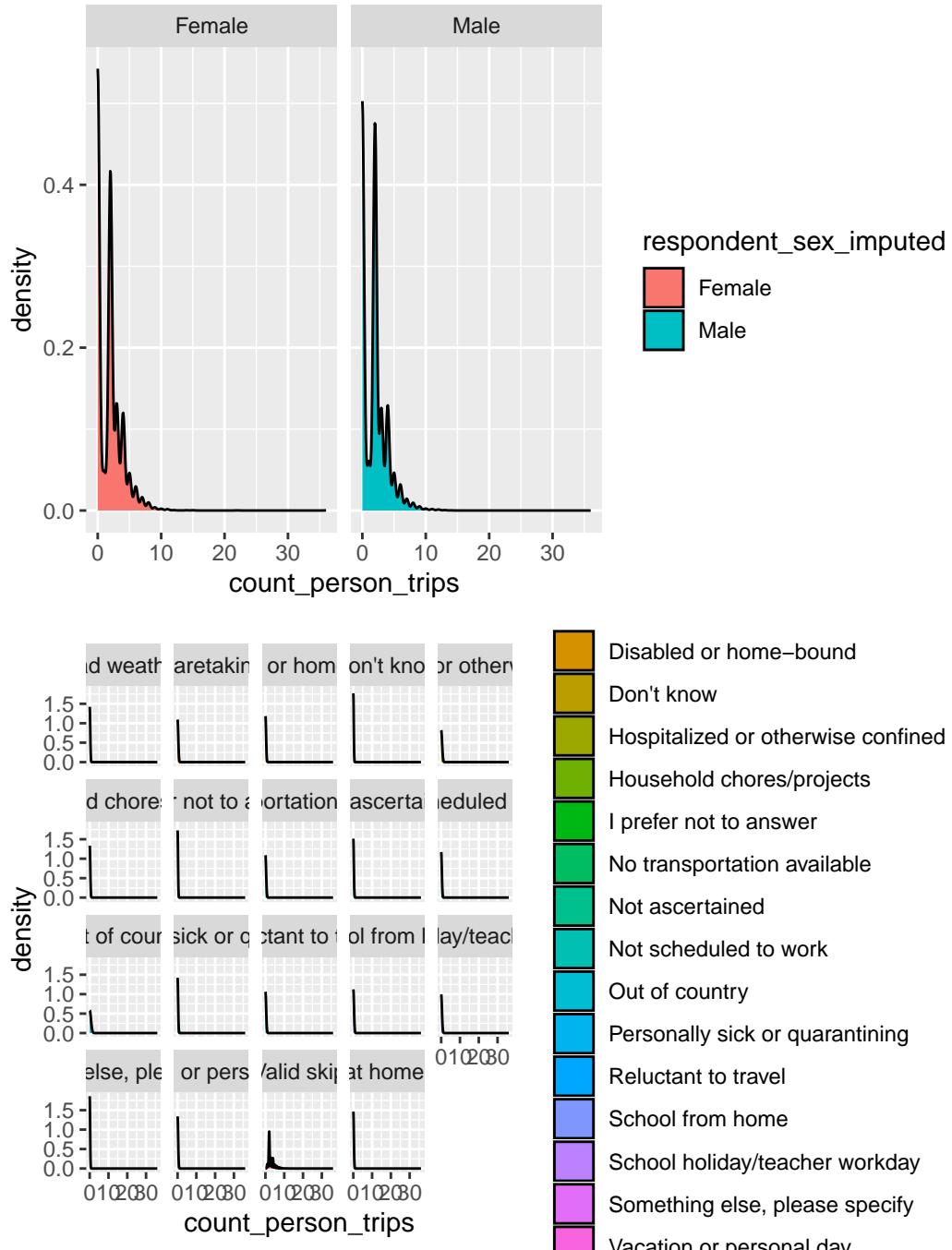


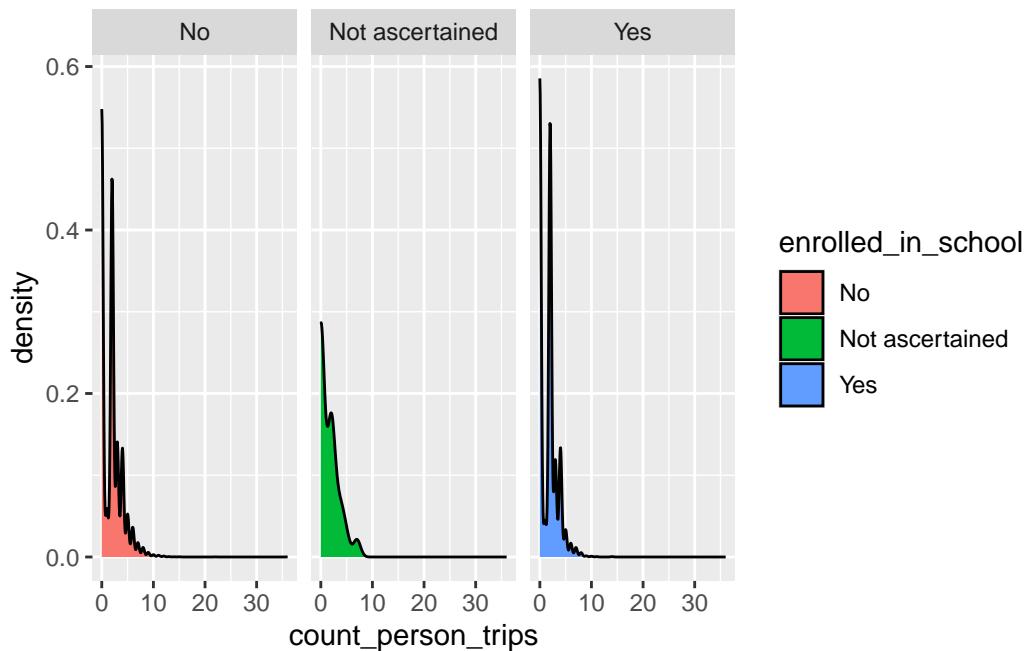




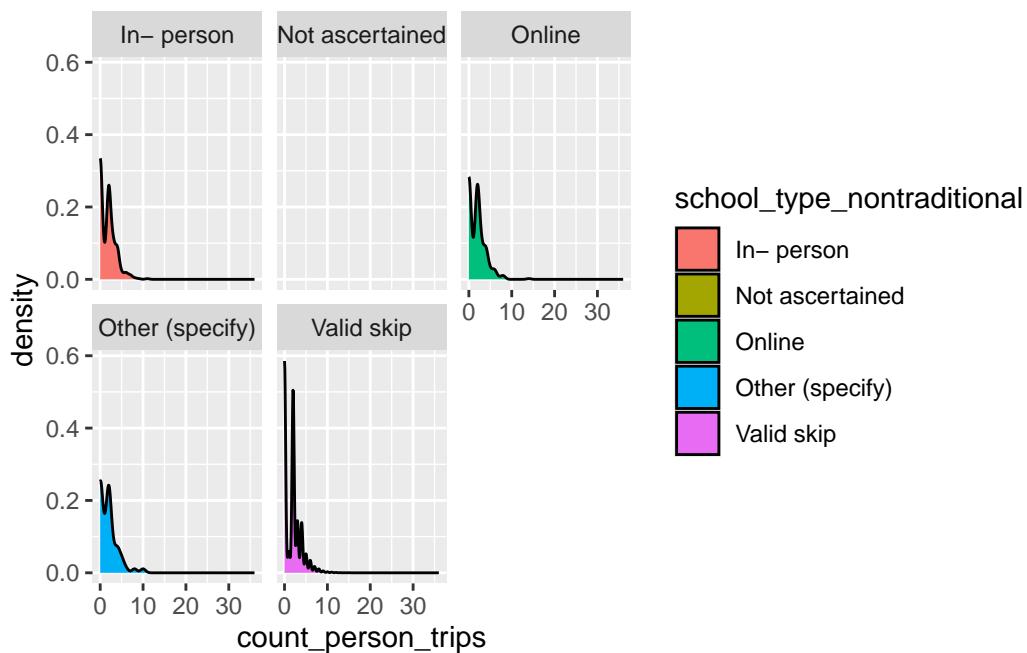
Warning: Groups with fewer than two data points have been dropped.
no non-missing arguments to max; returning -Inf







Warning: Groups with fewer than two data points have been dropped.
no non-missing arguments to max; returning -Inf



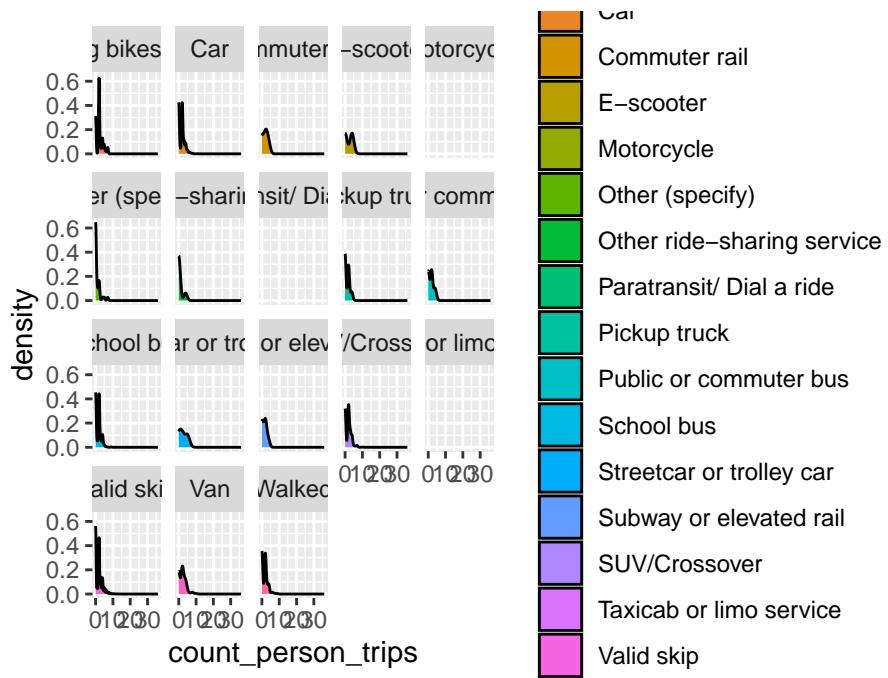
Warning: Groups with fewer than two data points have been dropped.

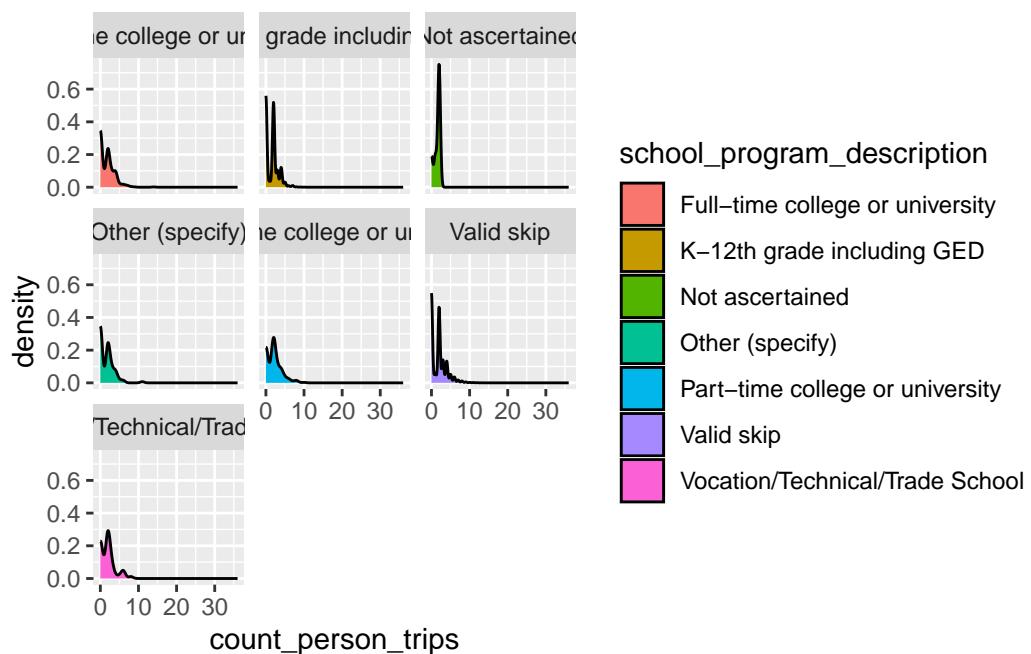
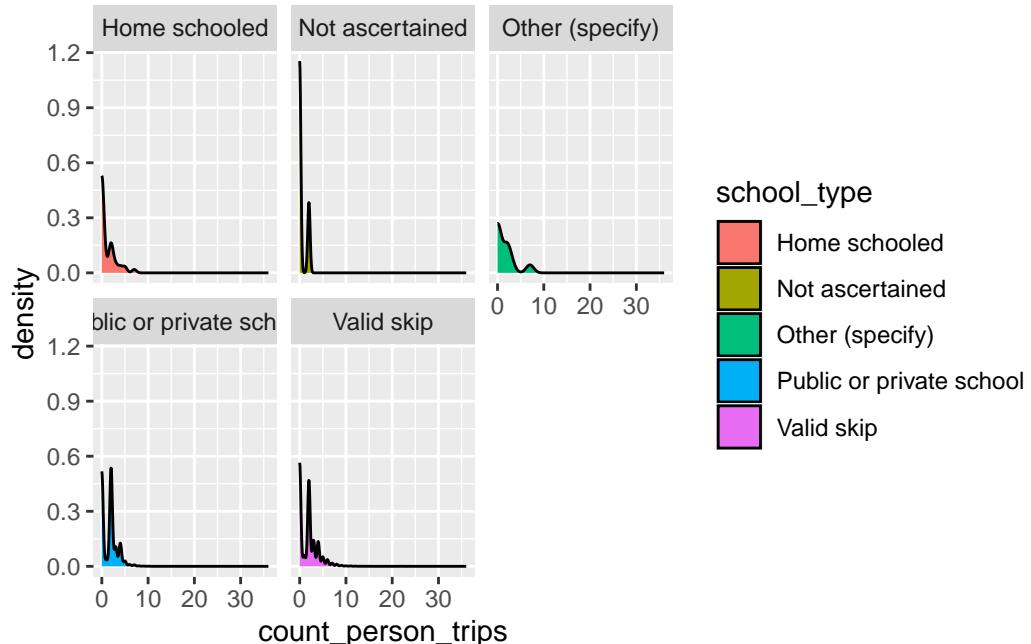
Warning: Groups with fewer than two data points have been dropped.
Groups with fewer than two data points have been dropped.

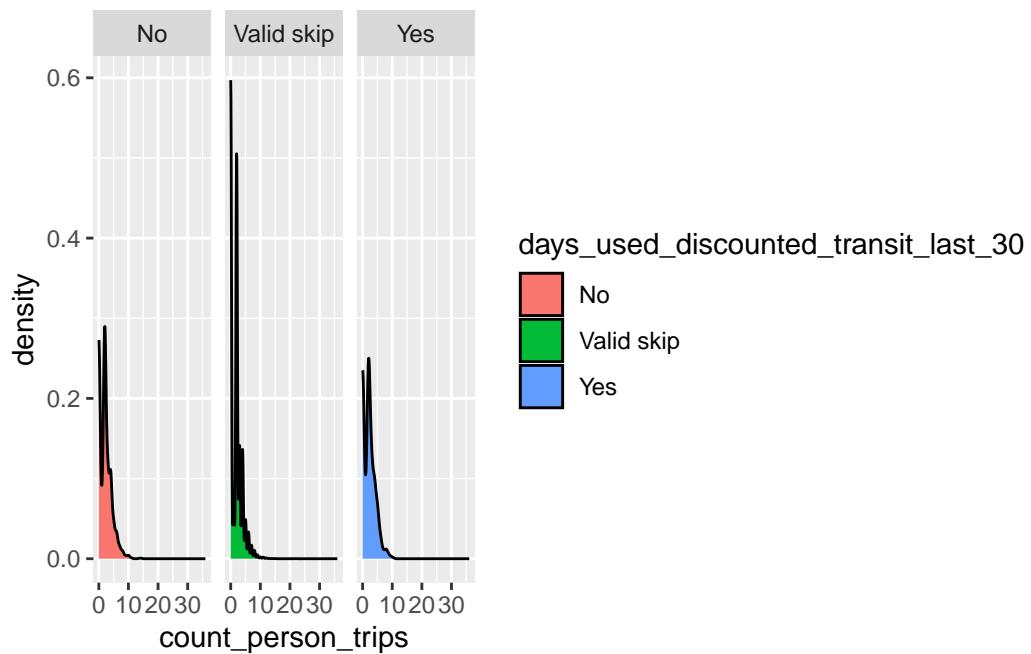
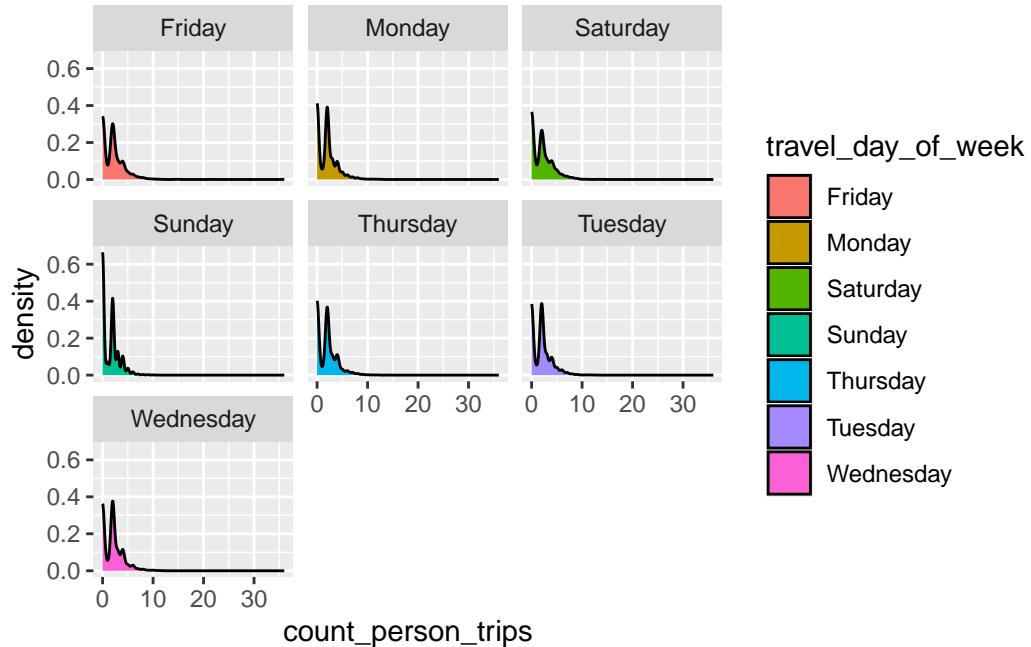
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning -Inf

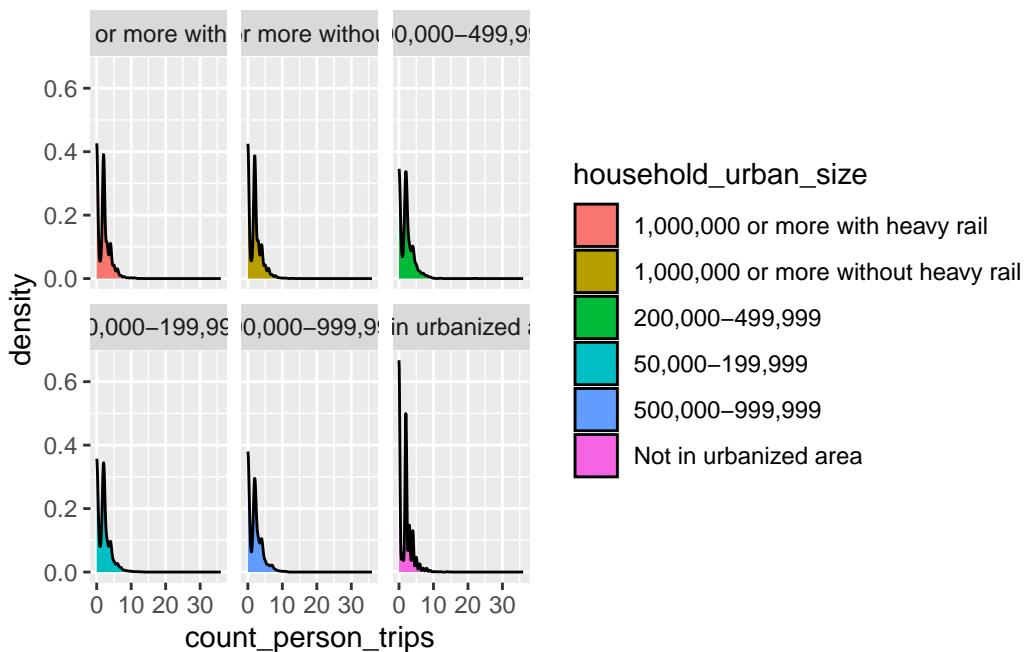
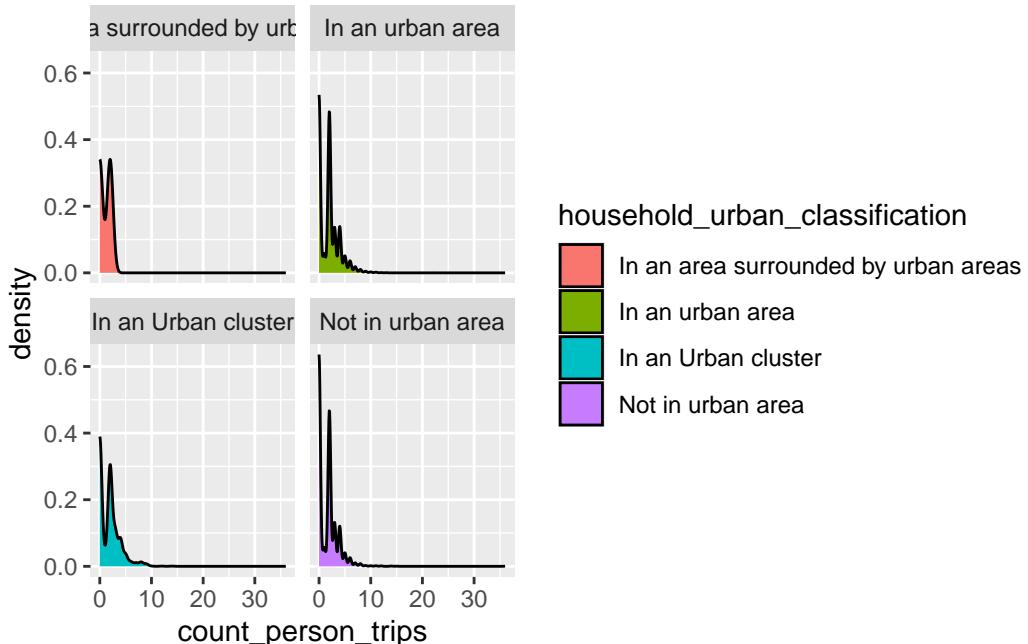
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning -Inf

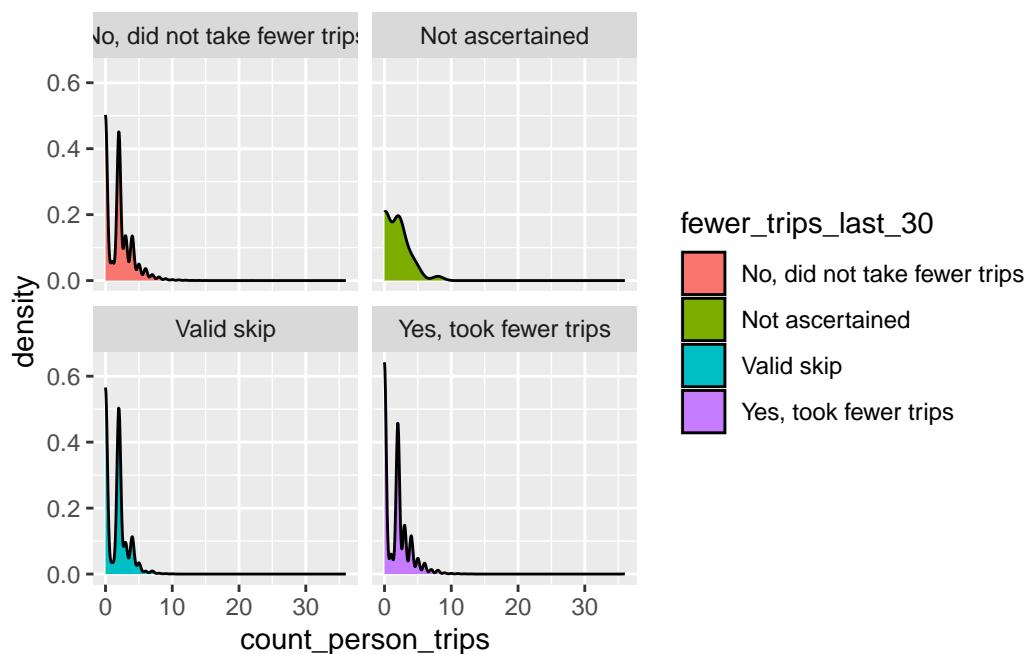
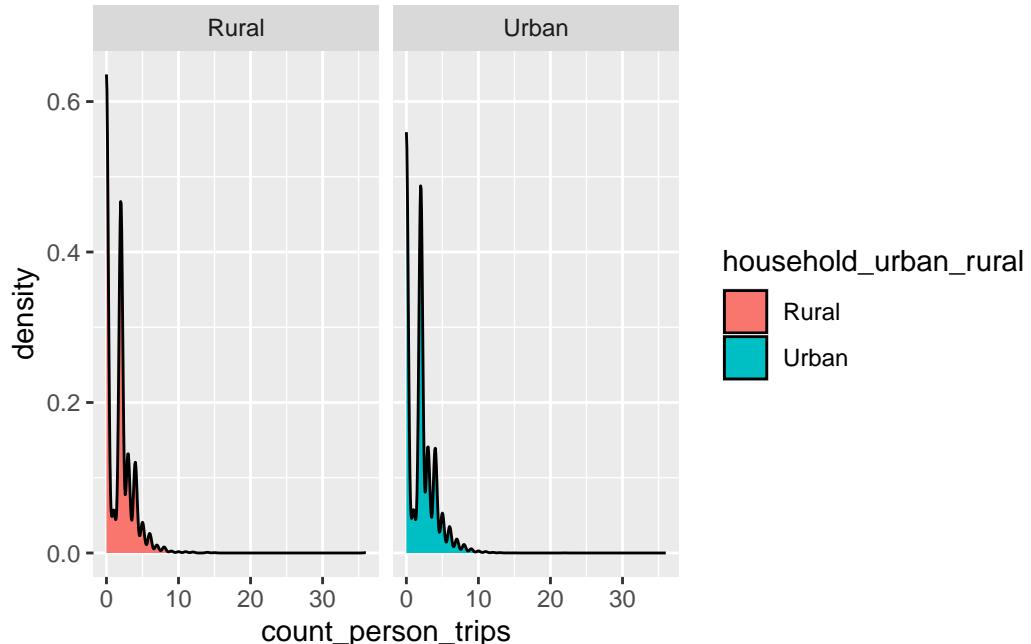
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning -Inf

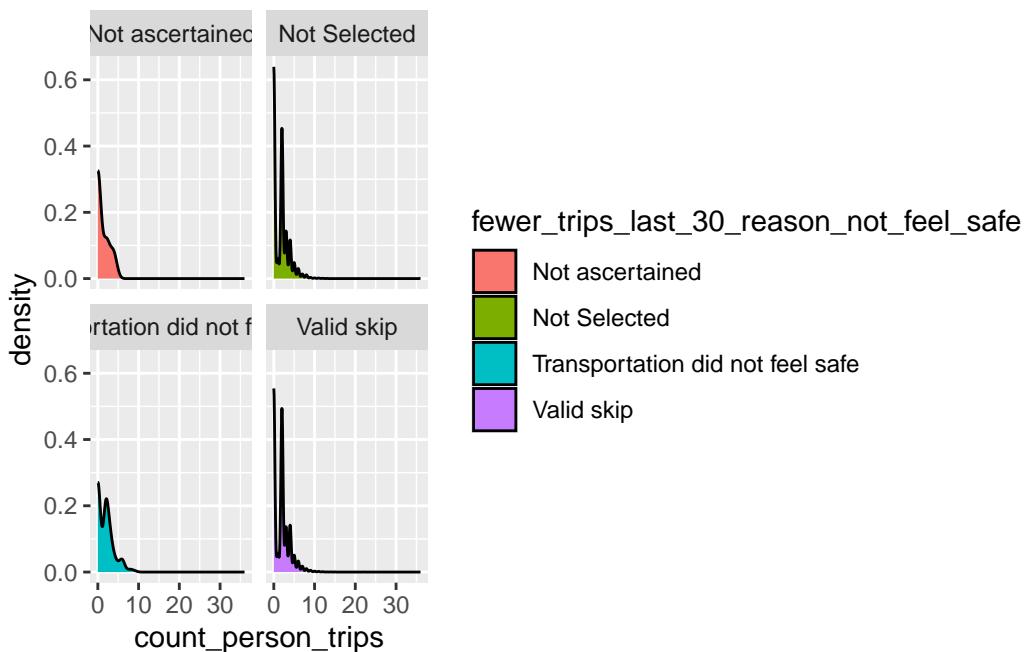
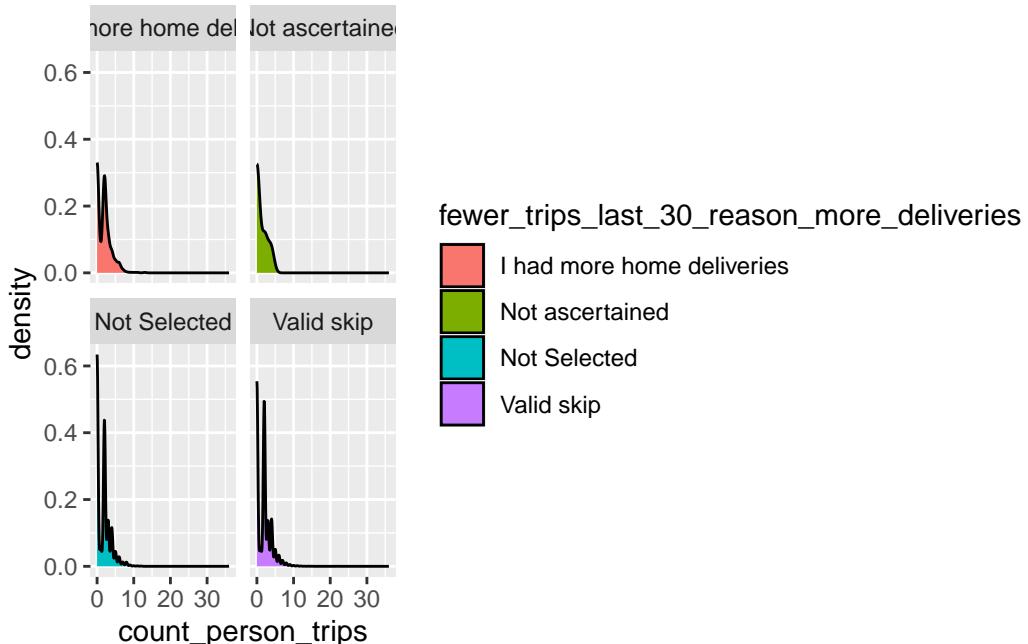


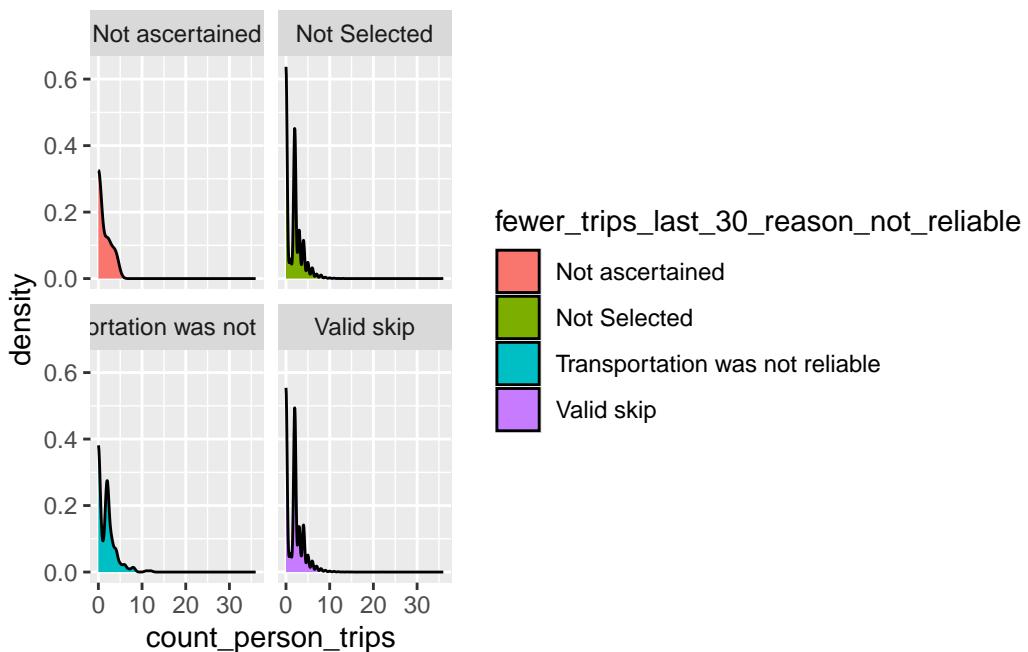
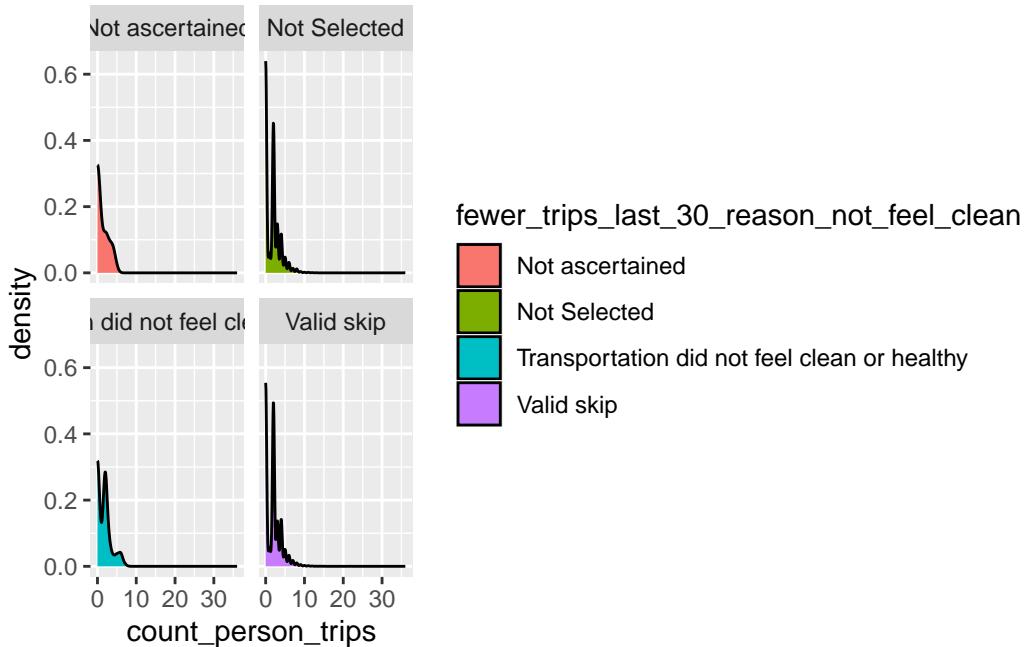


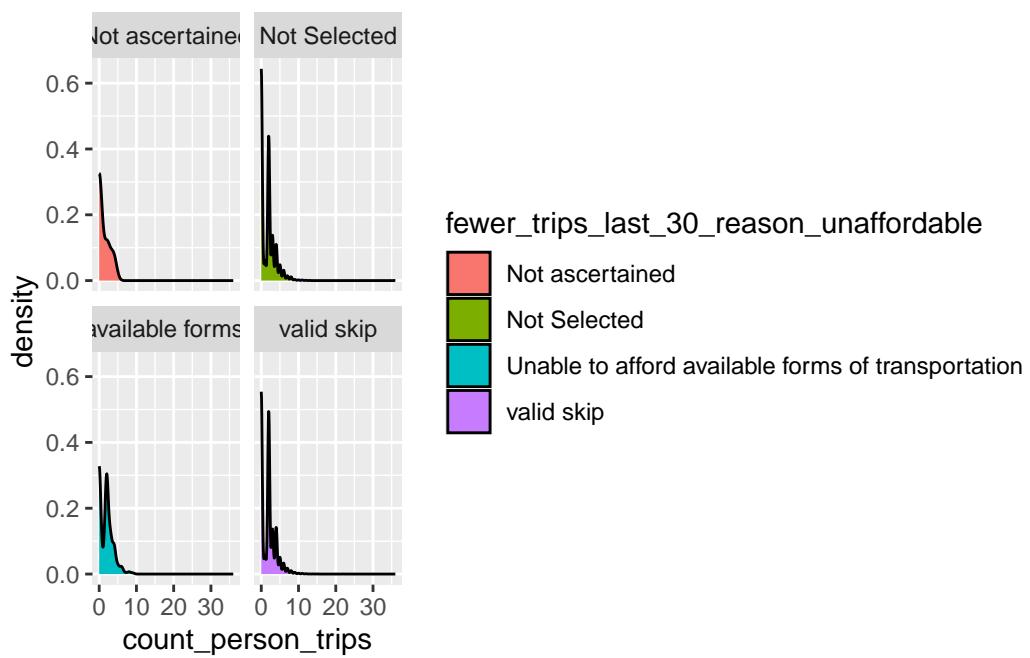
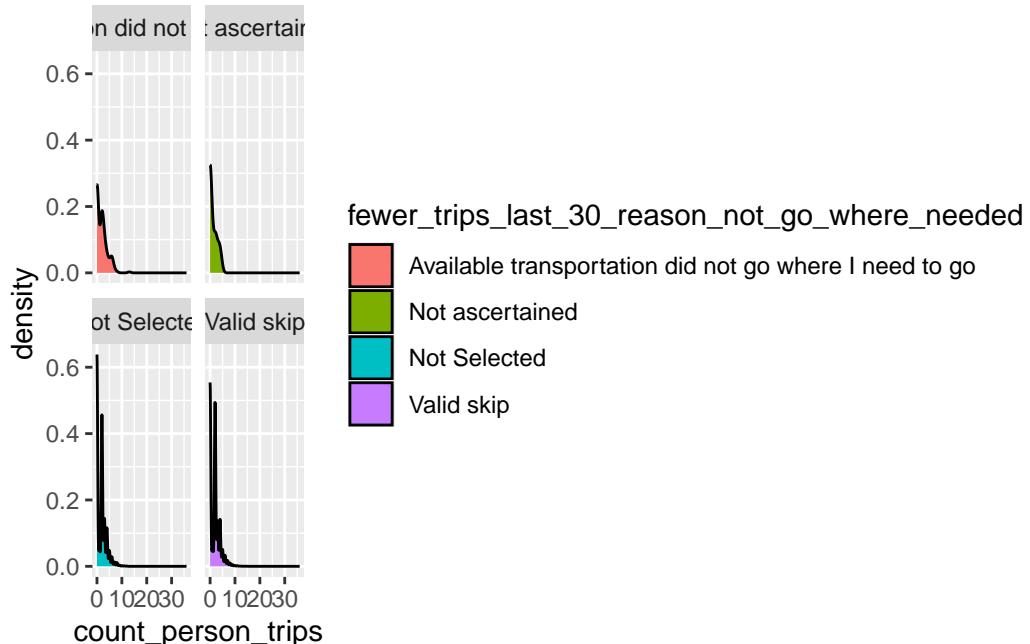


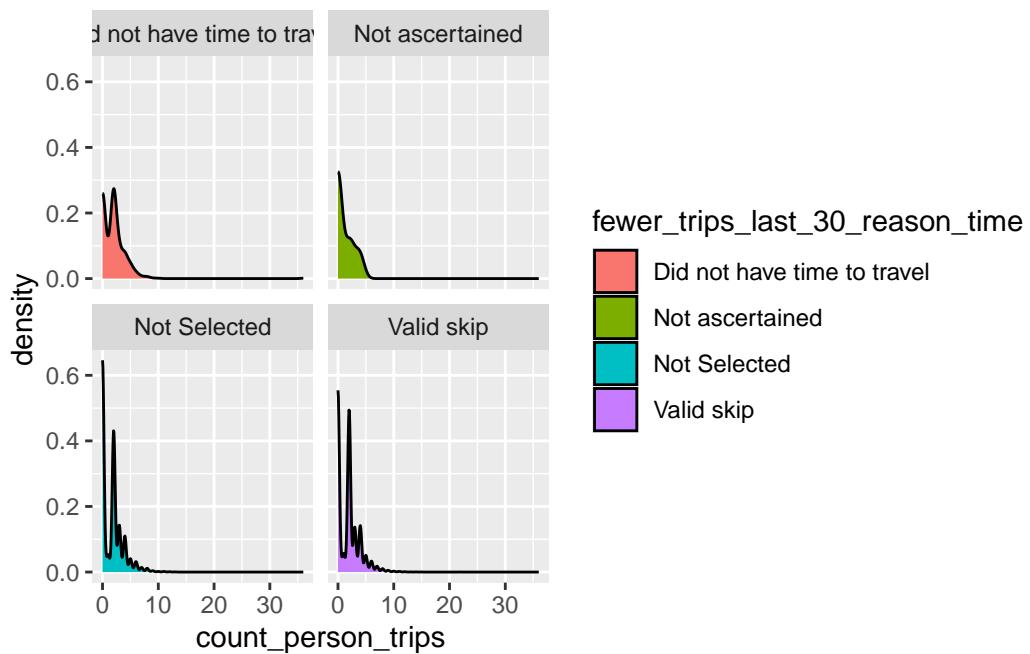
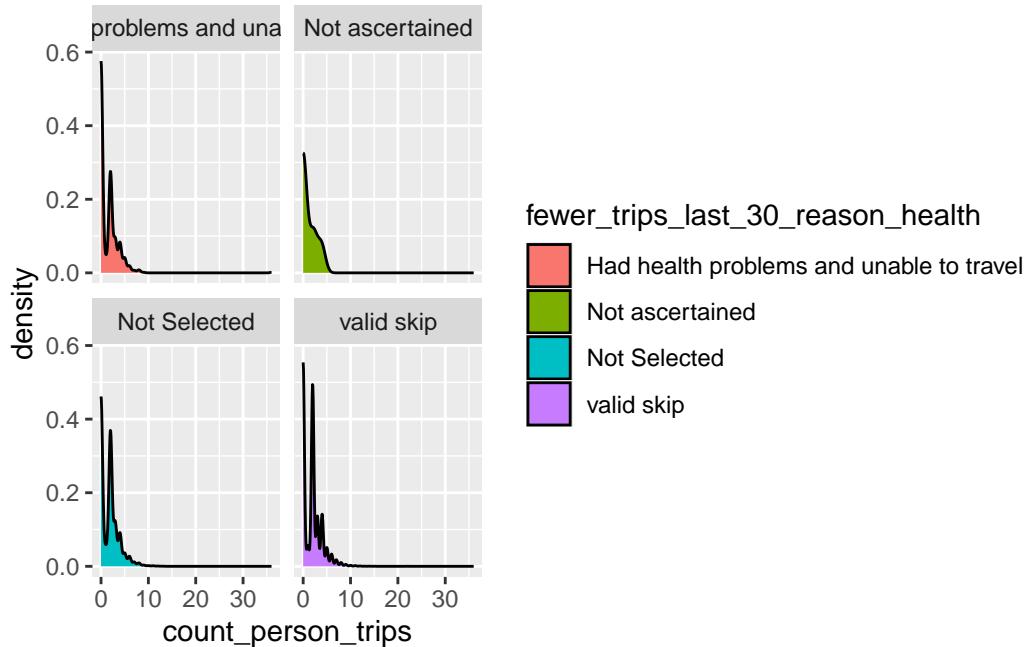


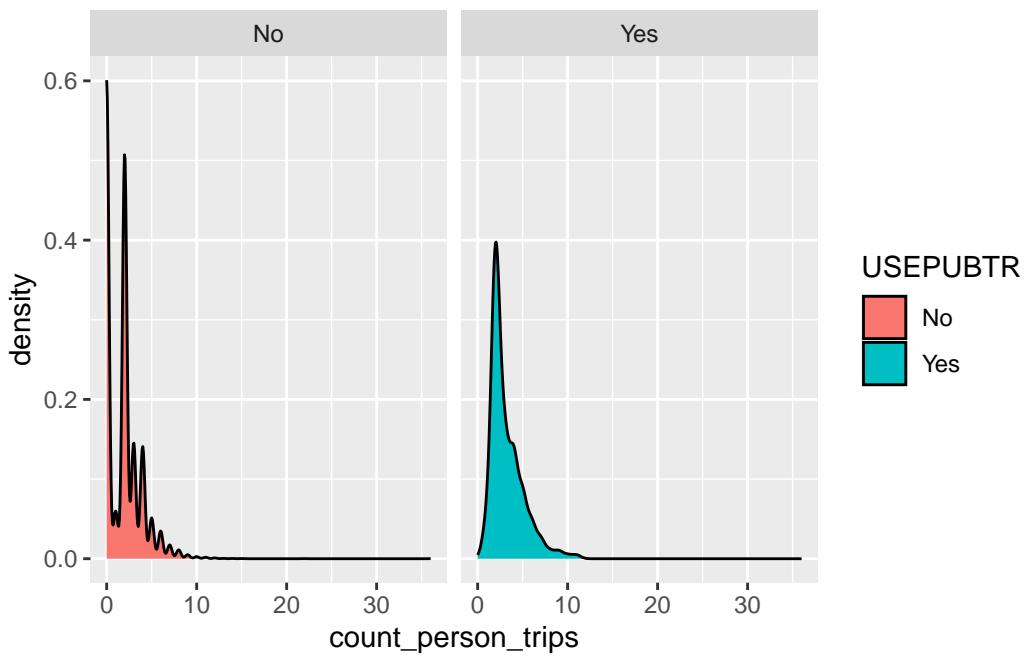
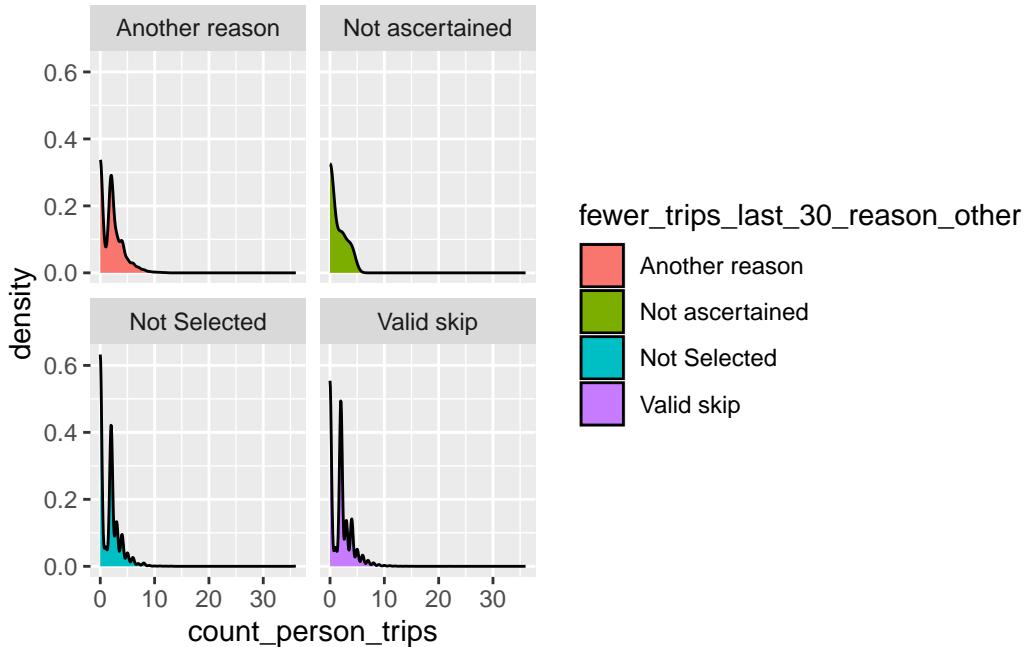


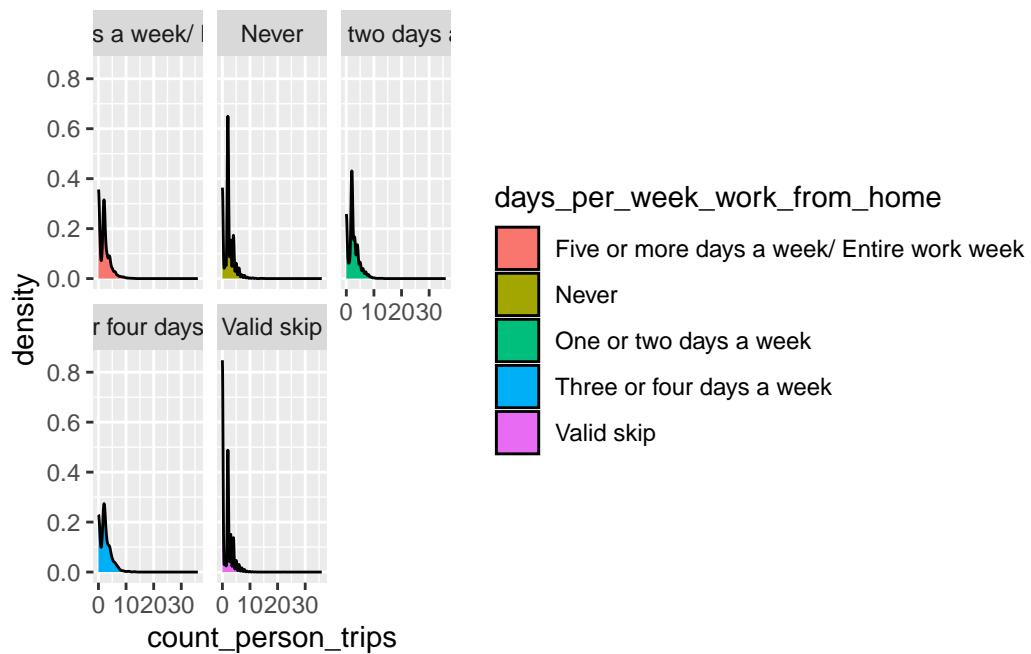
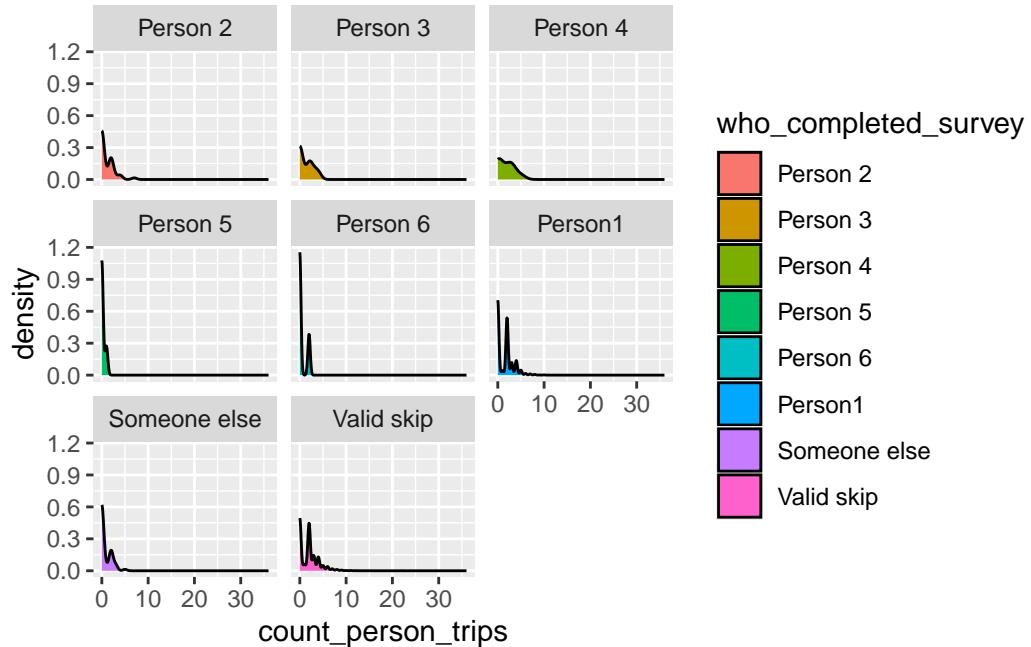


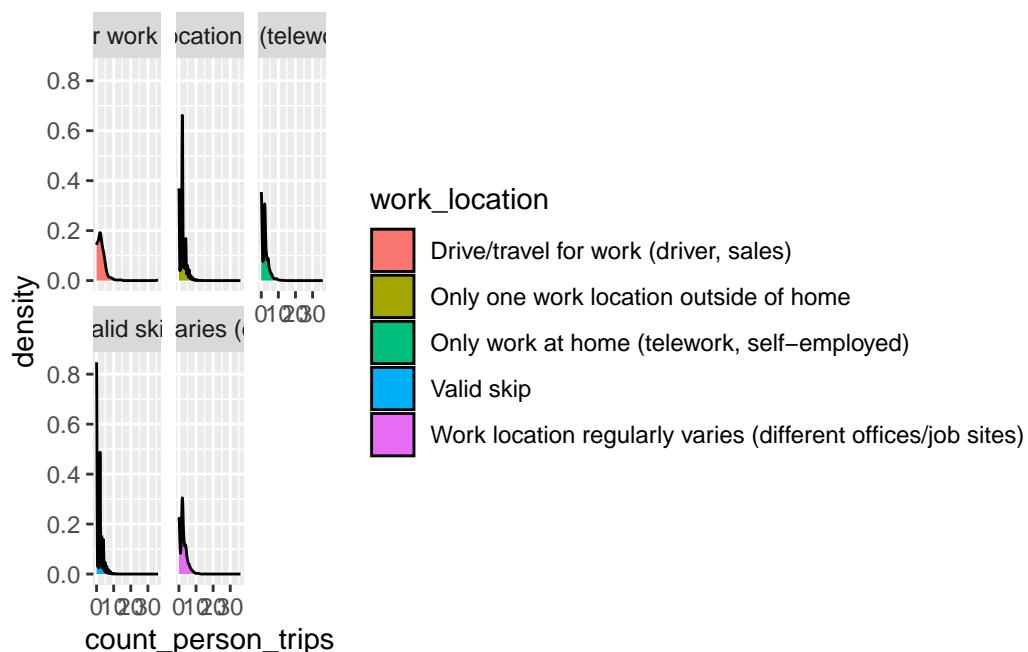
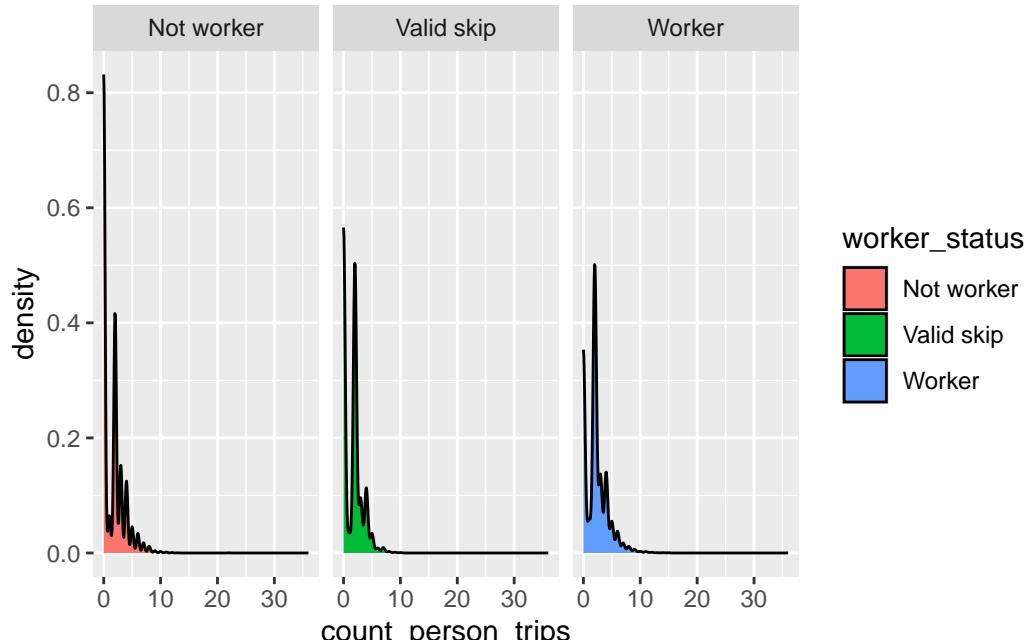










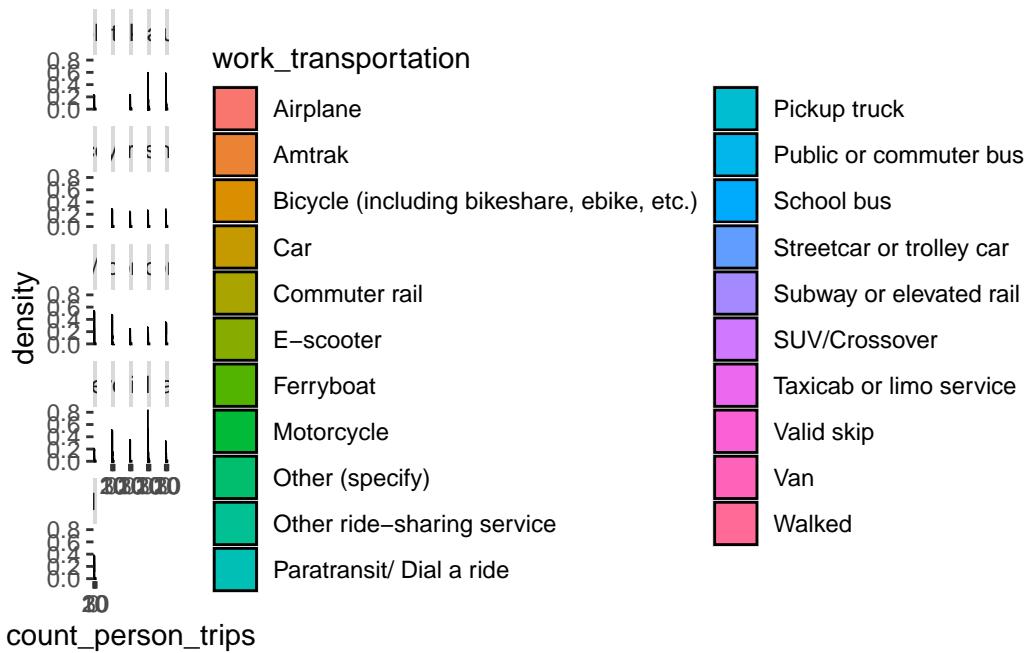


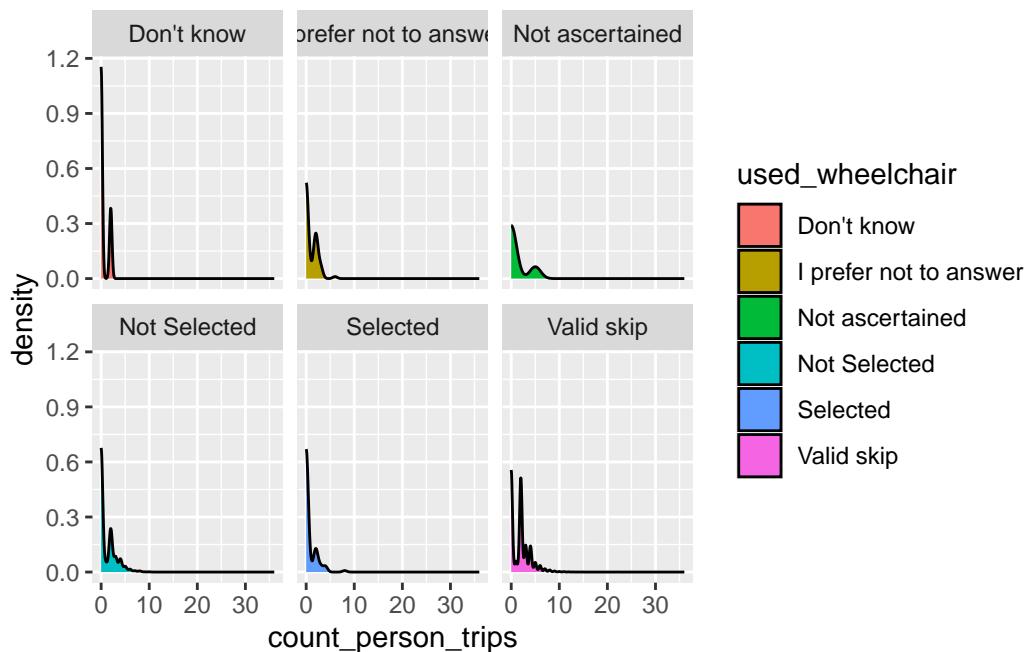
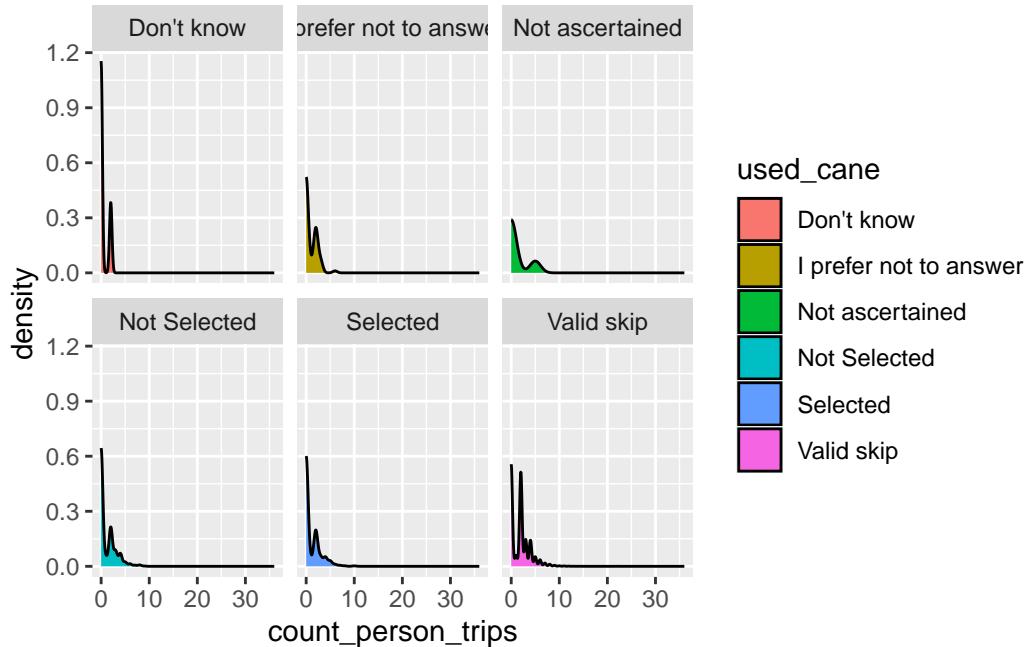
Warning: Groups with fewer than two data points have been dropped.

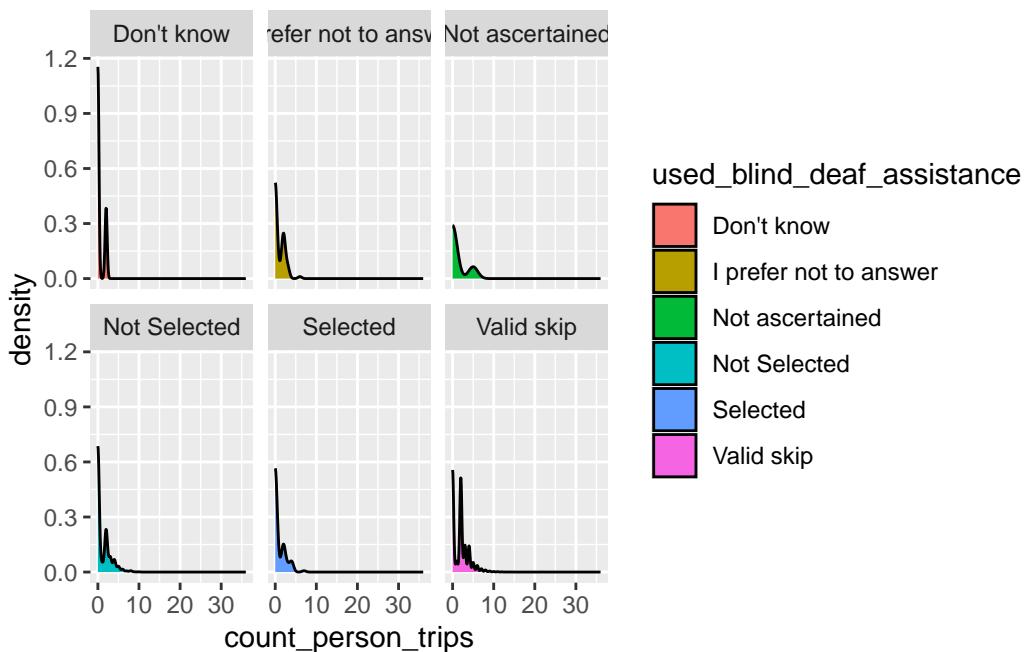
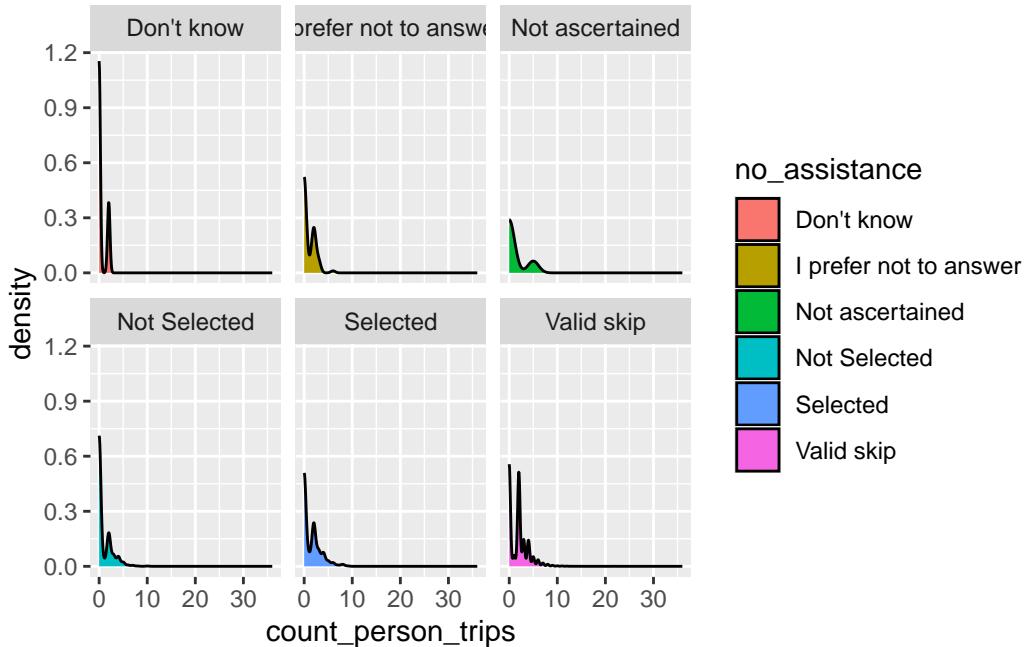
Warning: Groups with fewer than two data points have been dropped.

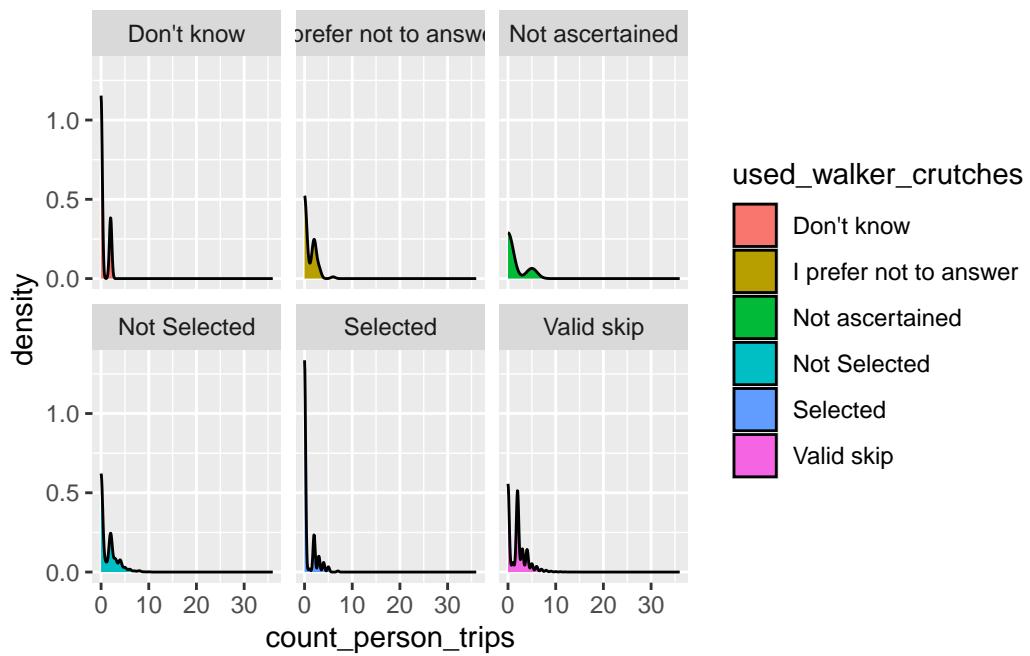
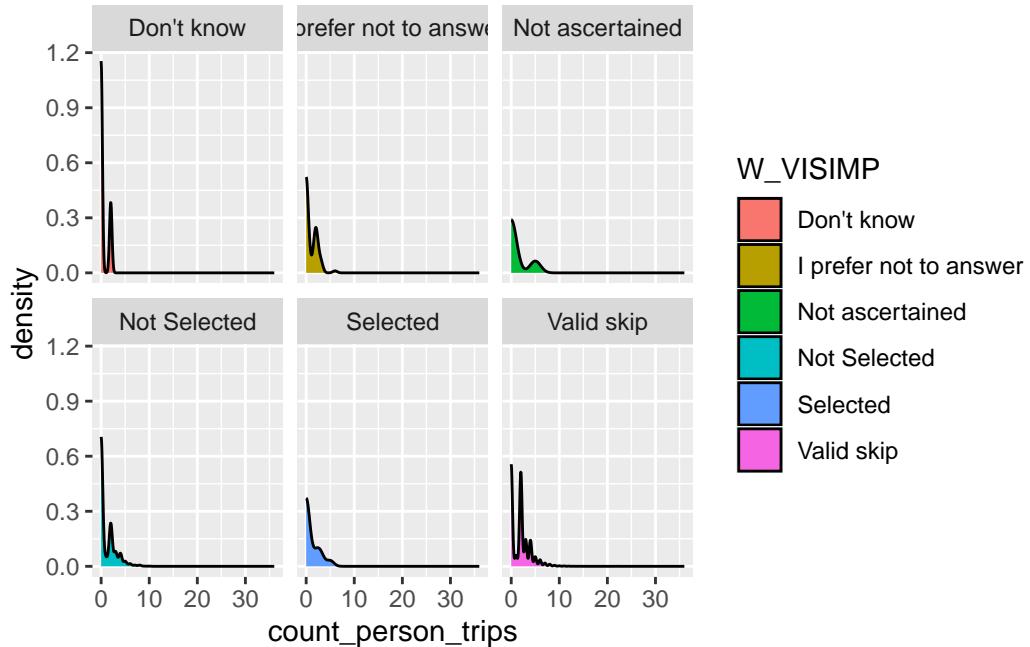
```
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning  
-Inf
```

```
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning  
-Inf
```









Trip Data

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

```
Rows: 31,074
Columns: 85
#> #> $ household_id
#> #> $ person_id
#> #> $ trip_id
#> #> $ sequential_trip_id
#> #> $ time_at_destination
#> #> $ trip_duration_minutes
#> #> $ num_people_on_trip
#> #> $ num_non_household_members_on_trip
#> #> $ num_household_members_on_trip
#> #> $ parking_cost
#> #> $ trip_miles
#> #> $ trip_weight_7_day
#> #> $ trip_weight_5_day
#> #> $ trip_weight_2_day
#> #> $ trip_case_id
#> #> $ vehicle_miles_traveled
#> #> $ gas_price_at_travel_time
#> #> $ num_adults_in_household
#> #> $ num_drivers_in_household
#> #> $ household_size
#> #> $ num_vehicles_in_household
#> #> $ travel_day_date
#> #> $ WRKCOUNT
#> #> $ household_stratum_id
#> #> $ respondent_age
#> #> $ household_grouping
#> #> $ census_division
#> #> $ census_region
#> #> $ DRIVER
#> #> $ driver_flag
#> #> $ education_level
#> #> $ end_trip_time
#> #> $ started_travel_at_home
#> #> $ household_income
#> #> $ household_income_imputed
#> #> <chr> "9000013002", "9000013002", "9000013~
#> #> <chr> "01", "01", "02", "02", "02", "04", ~
#> #> <chr> "01", "02", "01", "02", "03", "01", ~
#> #> <chr> "01", "02", "01", "02", "03", "01", ~
#> #> <dbl> 75, -9, 10, 10, -9, 75, -9, 260, 105~
#> #> <dbl> 15, 10, 30, 10, 30, 15, 10, 10, 15, ~
#> #> <dbl> 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 2, ~
#> #> <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
#> #> <chr> "02", "02", "01", "01", "01", "02", ~
#> #> <dbl> -1, -1, -1, -1, -1, -1, -1, -1, ~
#> #> <dbl> 3.9024239, 3.9024239, 17.0770665, 4.~
#> #> <dbl> 1608362, 1608362, 2208973, 2208973, ~
#> #> <dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1~
#> #> <dbl> 1569852, 1569852, 2274239, 2274239, ~
#> #> <chr> "90000130020101", "90000130020102", ~
#> #> <dbl> 3.9024239, 3.9024239, 17.0770665, 4.~
#> #> <chr> "355.1", "355.1", "355.1", "355.1", ~
#> #> <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
#> #> <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ~
#> #> <dbl> 4, 4, 4, 4, 4, 4, 2, 2, 2, 2, ~
#> #> <dbl> 2, 2, 2, 2, 2, 2, 1, 1, 1, 1, ~
#> #> <date> 2022-02-01, 2022-02-01, 2022-02-01, ~
#> #> <dbl> 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, ~
#> #> <chr> "1021", "1021", "1021", "1021", "102~
#> #> <dbl> 39, 39, 42, 42, 42, 5, 5, 32, 32, 32~
#> #> <chr> "New England MSA/CMSA of 1 million+ ~
#> #> <chr> "New England", "New England", "New E~
#> #> <chr> "Northeast", "Northeast", "Northeast~
#> #> <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", ~
#> #> <chr> "Driver on trip", "Driver on trip", ~
#> #> <chr> "Master's degree", "Master's degree", ~
#> #> <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
#> #> <chr> "Yes", "Yes", "Yes", "Yes", "Yes", "Yes", ~
#> #> <chr> "$200,000 or more", "$200,000 or mor~
#> #> <chr> "$200,000 or more", "$200,000 or mor~
```

```
$ drove_on_trip
$ hispanic
$ race
$ own_or_rent_home
$ LIF_CYC
$ loop_trip
$ metro_area_category
$ metro_area_size
$ person_1_on_trip
$ person_10_on_trip
$ person_2_on_trip
$ person_3_on_trip
$ person_4_on_trip
$ person_5_on_trip
$ person_6_on_trip
$ person_7_on_trip
$ person_8_on_trip
$ person_9_on_trip
$ paid_for_parking_travel_day
$ paid_for_parking_trip
$ parking_payment_periodicity
$ primary_activity_not_work
$ proxy_respondent
$ PSGR_FLG
$ public_transit_used_on_trip
$ metro_rail_status
$ respondent_hispanic
$ respondent_race
$ respondent_sex
$ respondent_sex_imputed
$ start_trip_time
$ weekend_trip
$ travel_day_of_week
$ trip_purpose
$ vehicle_used_on_trip
$ trip_mode
$ household_urban_classification
$ household_urban_size
$ household_urban_rural
$ vehicle_case_id
$ vehicle_id
$ vehicle_type
$ minutes_walked_from_parking

<chr> "Household member drove", "Household
<chr> "Not Hispanic or Latino", "Not Hispan-
<chr> "White", "White", "White", "White", ~
<chr> "Owned by hh member with mortgage or-
<chr> "2+ adults, youngest child 0-5", "2+-
<chr> "Not a loop trip", "Not a loop trip"-
<chr> "MSA of 1 million or more, with rail-
<chr> "In an MSA or CMSA of 3 million or m-
<chr> "Selected", "Selected", "Not Selecte-
<chr> "Valid skip", "Valid skip", "Valid s-
<chr> "Not Selected", "Not Selected", "Sel-
<chr> "Not Selected", "Not Selected", "Not-
<chr> "Selected", "Selected", "Not Selecte-
<chr> "Valid skip", "Valid skip", "Valid s-
<chr> "No", "No", "Valid skip", "Valid ski-
<chr> "Valid skip", "Valid skip", "Valid s-
<chr> "Valid skip", "Valid skip", "Valid s-
<chr> "A homemaker", "A homemaker", "Valid-
<chr> "Self", "Self", "Someone else", "Som-
<chr> "Not passenger on trip", "Not passen-
<chr> "Did not use public transit", "Did n-
<chr> "Yes", "Yes", "Yes", "Yes", "Yes", "~
<chr> "Not Hispanic", "Not Hispanic", "Not-
<chr> "White", "White", "White", "White", ~
<chr> "Female", "Female", "Male", "Male", ~
<chr> "Female", "Female", "Male", "Male", ~
<chr> NA, NA, NA, NA, NA, NA, NA, NA, ~
<chr> "Weekend", "Weekend", "Weekend", "We-
<chr> "Sunday", "Sunday", "Sunday", "Sunda-
<chr> "Home-based social/recreational (HBS-
<chr> "Yes, used a household vehicle", "Ye-
<chr> "SUV/Crossover", "SUV/Crossover", "S-
<chr> "In an urban area", "In an urban are-
<chr> "1,000,000 or more with heavy rail", ~
<chr> "Urban", "Urban", "Urban", "Urban", ~
<chr> NA, NA, NA, NA, NA, NA, NA, NA, ~
<chr> "", "", "", "", "", "", "", "", ~
<chr> "SUV (Santa Fe, Tahoe, Jeep, etc.)", ~
<chr> "1 to less than 5 minutes", "0 minut-
```


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	
Min. : 1.000	Min. : 0.000	Min. :2022-01-01	
1st Qu.: 2.000	1st Qu.: 1.000	1st Qu.:2022-04-01	
Median : 2.000	Median : 2.000	Median :2022-08-01	
Mean : 2.822	Mean : 2.134	Mean :2022-07-18	
3rd Qu.: 4.000	3rd Qu.: 3.000	3rd Qu.:2022-11-01	
Max. :10.000	Max. :11.000	Max. :2023-01-01	
WRKCOUNT	household_stratum_id	respondent_age	household_grouping
Min. :0.000	Length:31074	Min. : 5.0	Length:31074
1st Qu.:1.000	Class :character	1st Qu.:32.0	Class :character
Median :1.000	Mode :character	Median :49.0	Mode :character
Mean : 1.304		Mean : 46.8	
3rd Qu.:2.000		3rd Qu.:64.0	
Max. :6.000		Max. :92.0	
census_division	census_region	DRIVER	driver_flag
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
education_level	end_trip_time	started_travel_at_home	
Length:31074	Length:31074	Length:31074	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
household_income	household_income_imputed	drove_on_trip	
Length:31074	Length:31074	Length:31074	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	

hispanic	race	own_or_rent_home	LIF_CYC
Length:31074	Length:31074	Length:31074	Length:31074
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
loop_trip	metro_area_category	metro_area_size	person_1_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_10_on_trip	person_2_on_trip	person_3_on_trip	person_4_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_5_on_trip	person_6_on_trip	person_7_on_trip	person_8_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_9_on_trip	paid_for_parking_travel_day	paid_for_parking_trip	
Length:31074	Length:31074	Length:31074	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
parking_payment_periodicity	primary_activity_not_work	proxy_respondent	
Length:31074	Length:31074	Length:31074	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	

```

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

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

respondent_sex_imputed start_trip_time    weekend_trip
Length:31074         Length:31074          Length:31074
Class :character    Class :character     Class :character
Mode  :character    Mode  :character     Mode  :character

travel_day_of_week trip_purpose        vehicle_used_on_trip  trip_mode
Length:31074         Length:31074          Length:31074        Length:31074
Class :character    Class :character     Class :character   Class :character
Mode  :character    Mode  :character     Mode  :character   Mode  :character

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

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

```

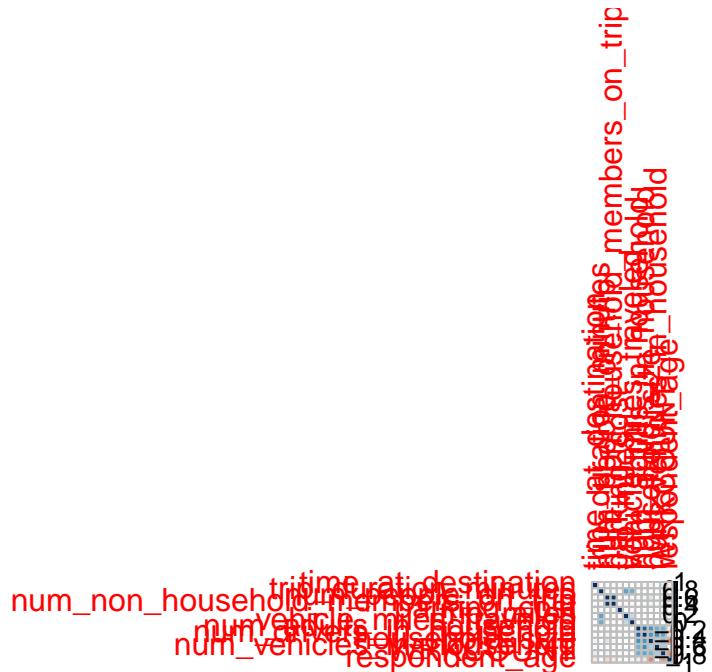
```
minutes_walked_from_parking  who_drove      who_drove_imputed
Length:31074                  Length:31074      Length:31074
Class :character              Class :character  Class :character
Mode  :character              Mode  :character  Mode  :character
```

```
reason_for_travel  reason_for_travel_to  why_trip
Length:31074        Length:31074        Length:31074
Class :character    Class :character    Class :character
Mode  :character    Mode  :character    Mode  :character
```

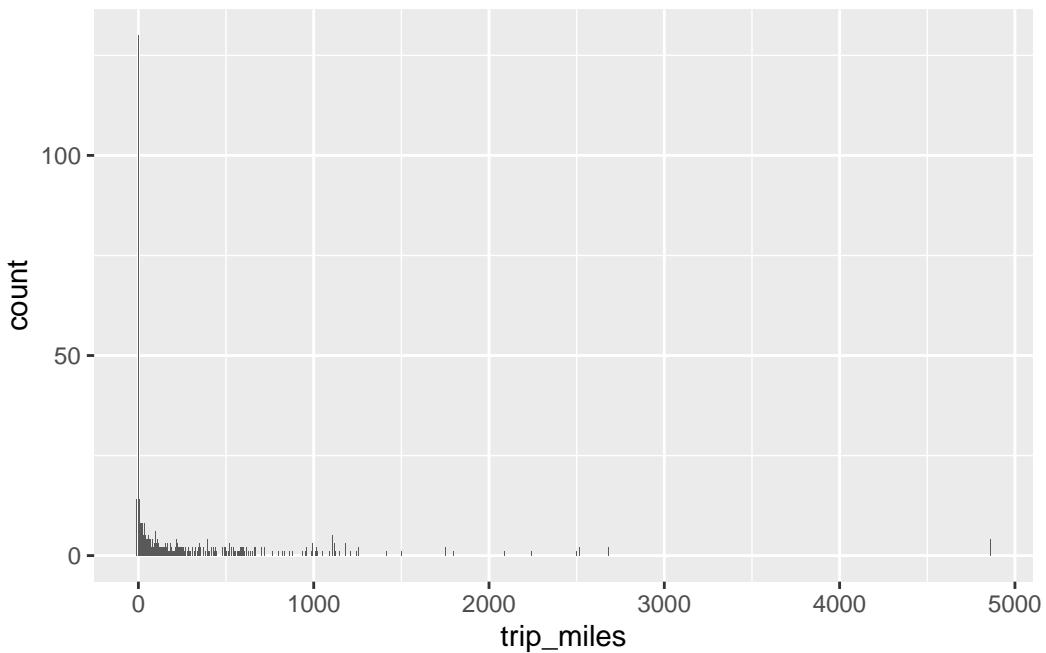
```
trip_purpose_old_schema worker_status
Length:31074          Length:31074
Class :character      Class :character
Mode  :character      Mode  :character
```

```
trip |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot()
```

```
Warning in corrplot::corrplot(cor(select(select(select(select(trip,
-contains("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 |>  
  ggplot(aes(trip_miles)) +  
    geom_bar()
```

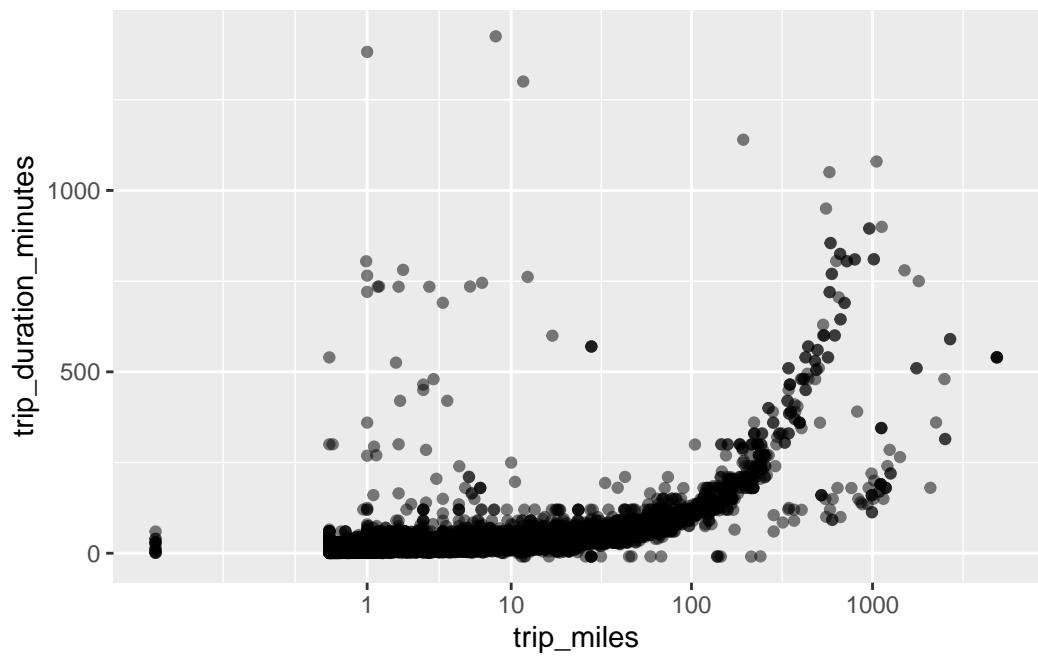
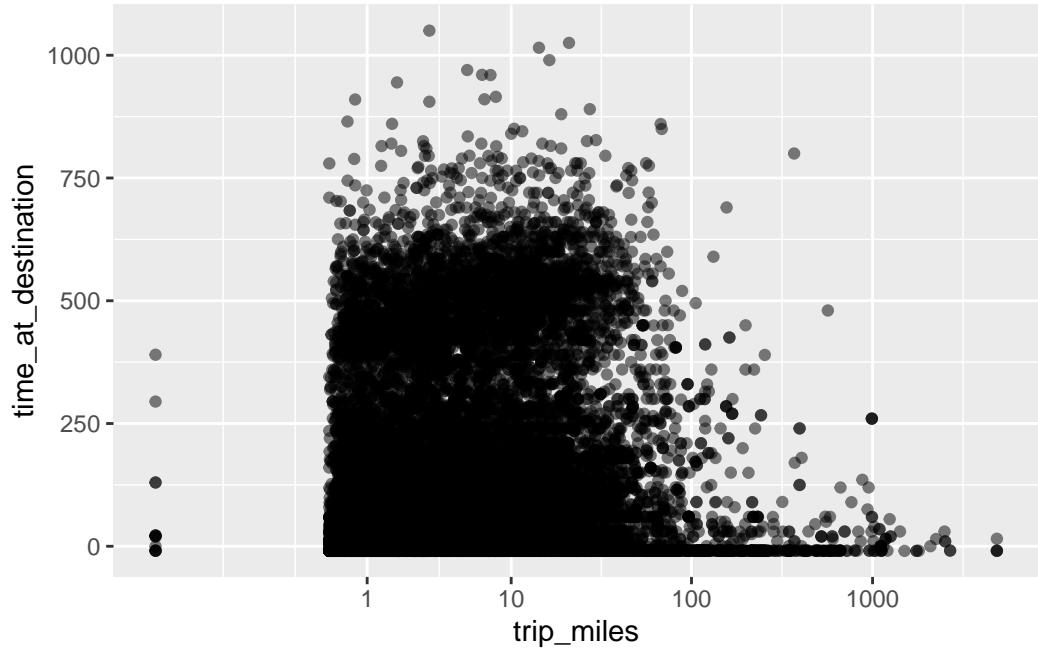


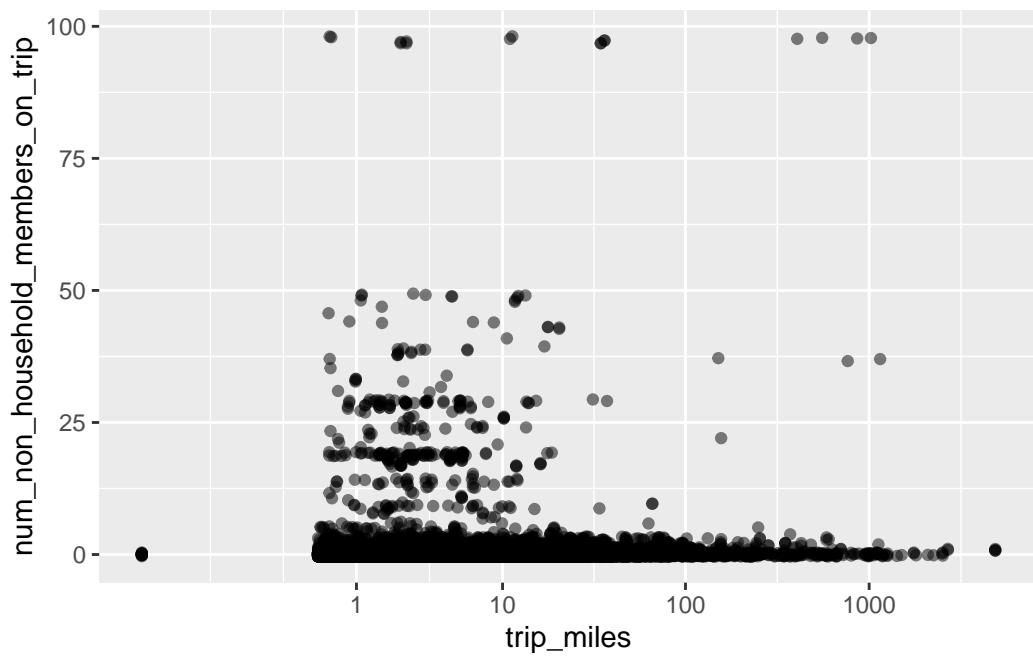
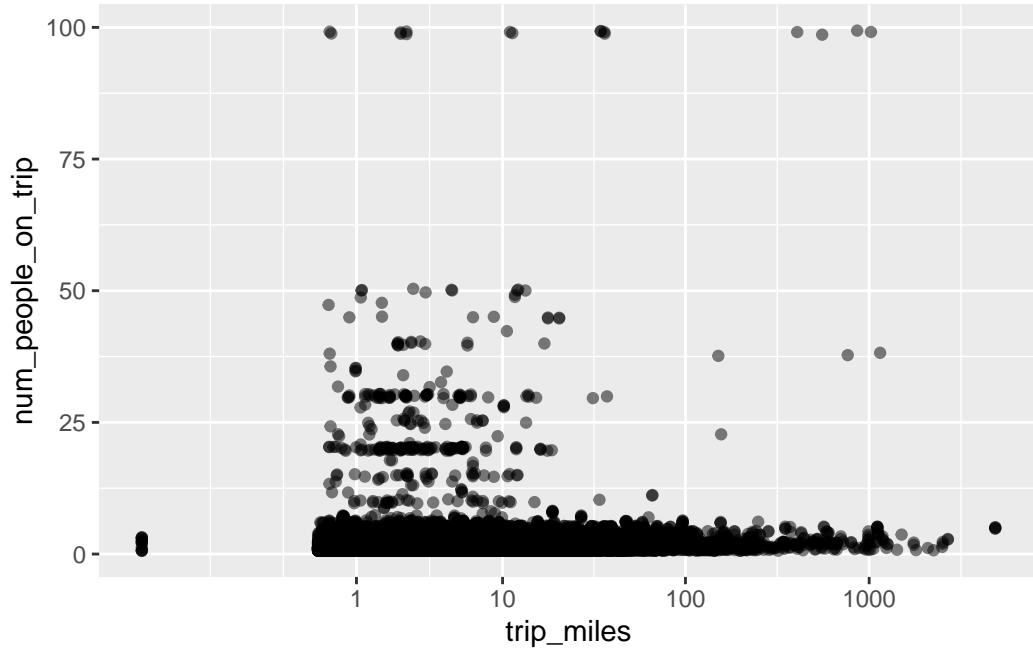
```

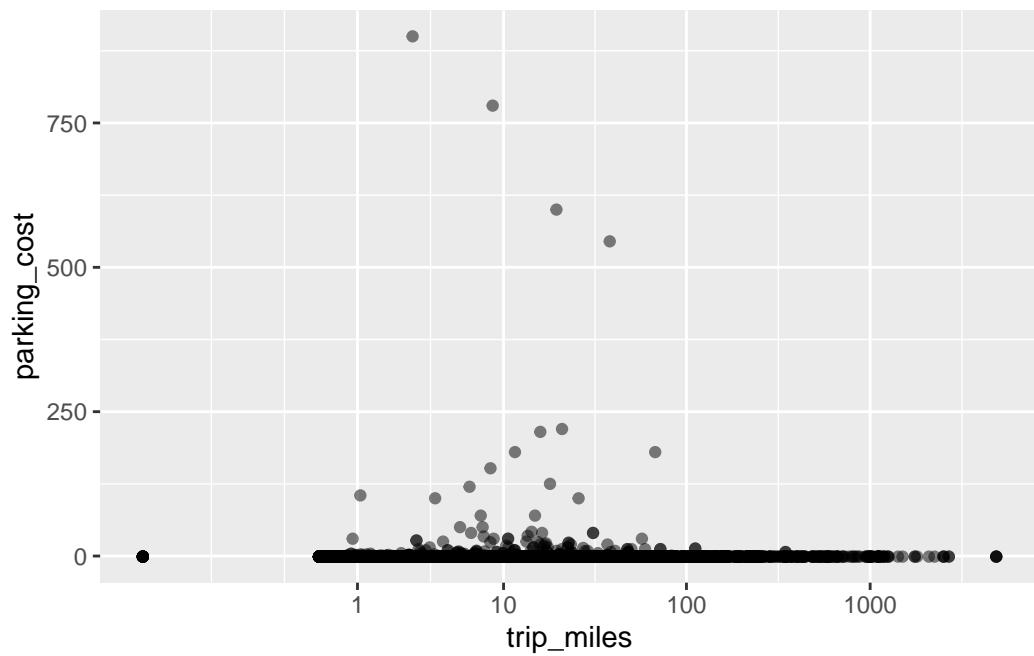
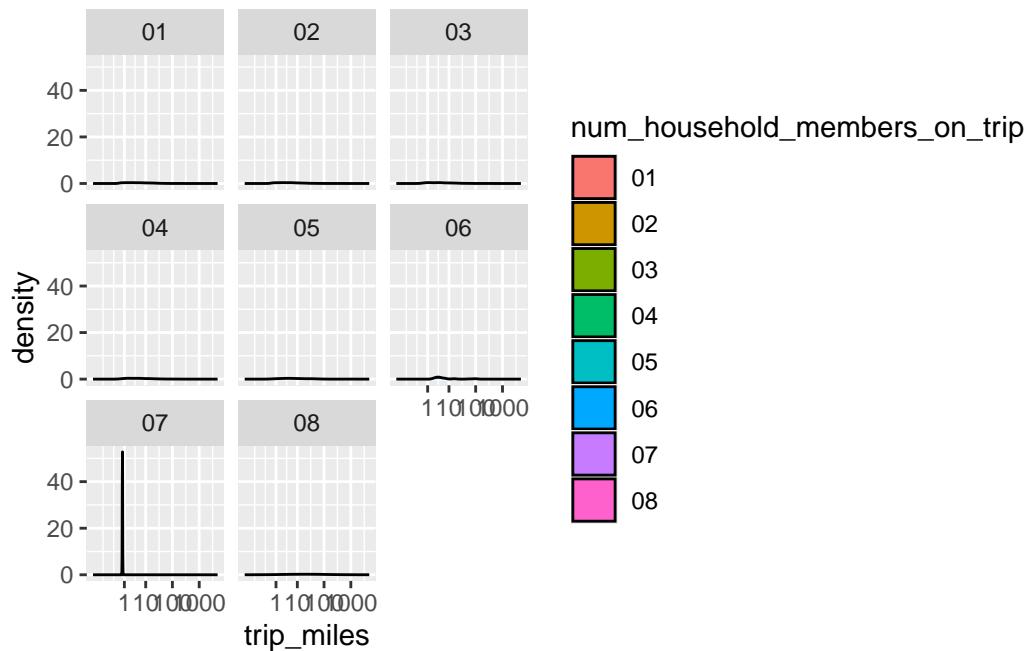
# Plot trip_miles by all variables
trip |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  colnames() -> trip_cols

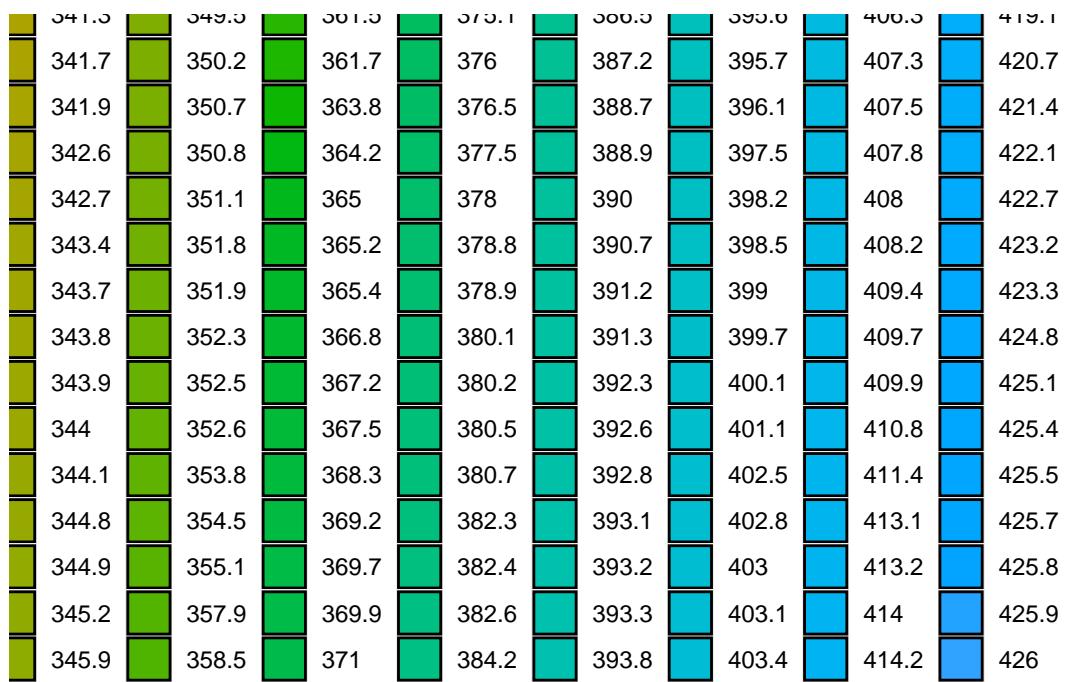
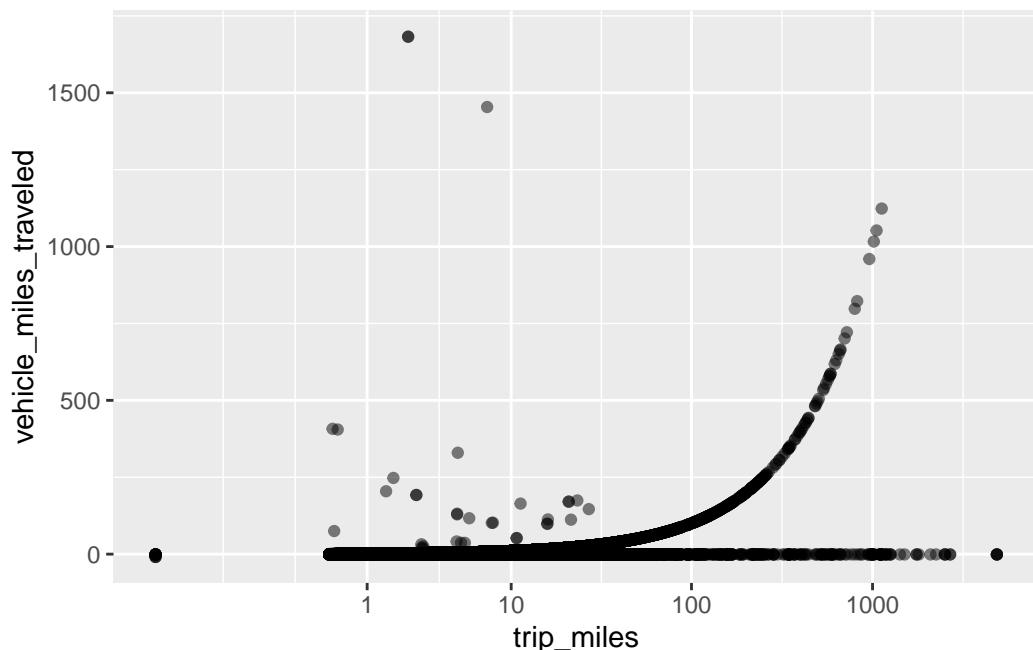
for (col in trip_cols) {
  if (is.character(trip[[col]])) {
    p <- trip |>
      ggplot(aes(x = trip_miles, fill = .data[[col]])) +
      geom_density() +
      facet_wrap(~.data[[col]]) +
      scale_x_continuous(
        trans = "pseudo_log",
        breaks = c(1, 10, 100, 1000, 10000),
        )
    print(p)
  } else if (col == "trip_miles") {
    next
  } else {
    p <- trip |>
      ggplot(aes(x = trip_miles, y = .data[[col]])) +
      geom_jitter(alpha = 0.5) +
      scale_x_continuous(
        trans = "pseudo_log",
        breaks = c(1, 10, 100, 1000, 10000),
        )
    print(p)
  }
}

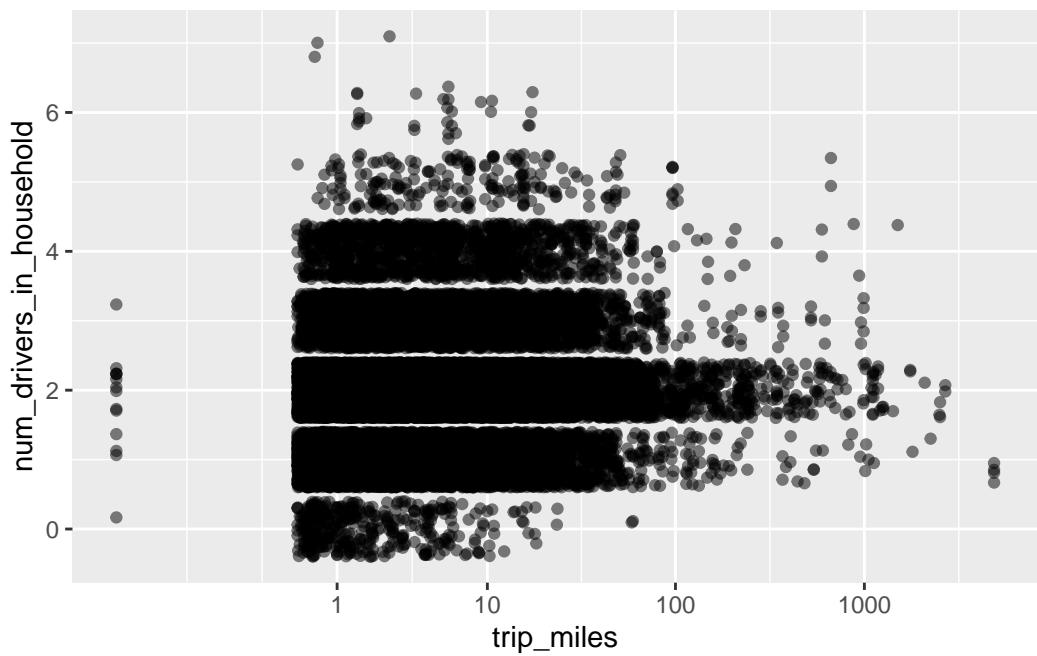
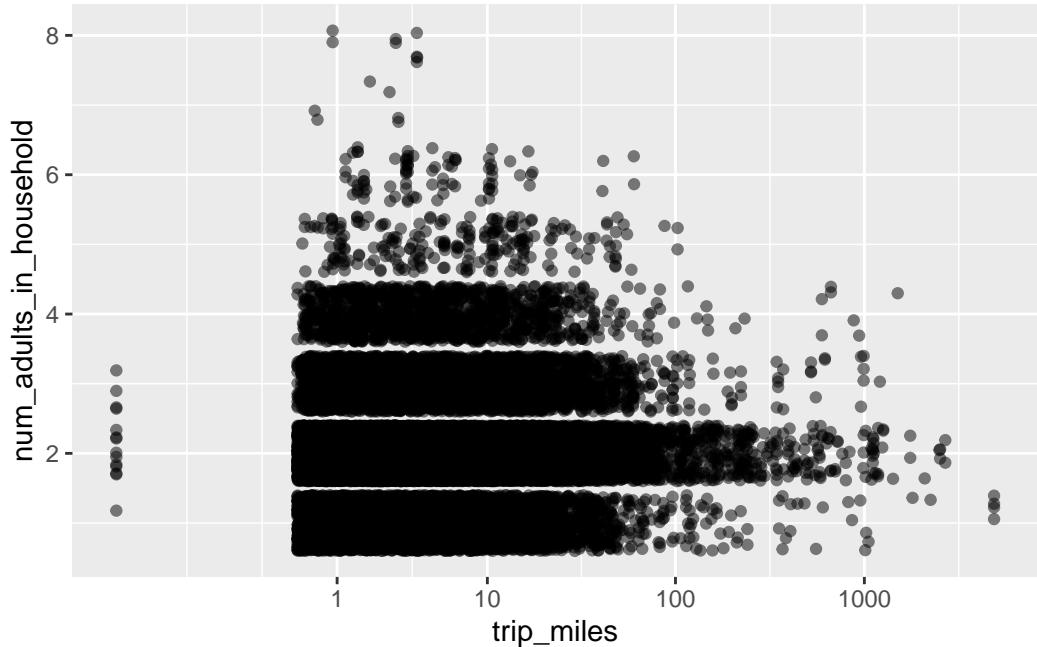
```

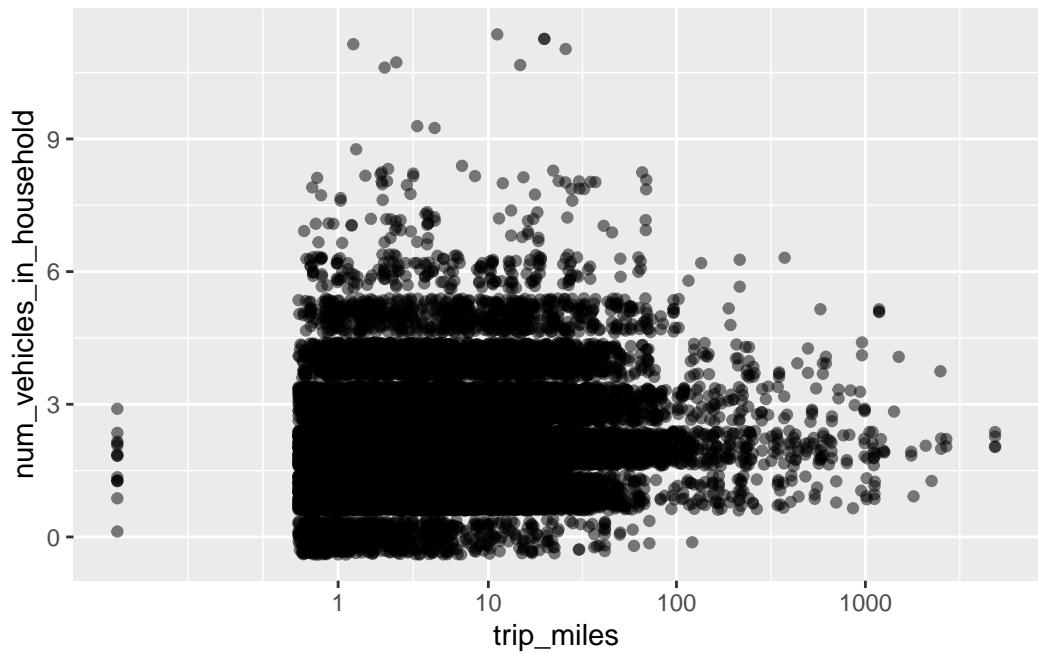
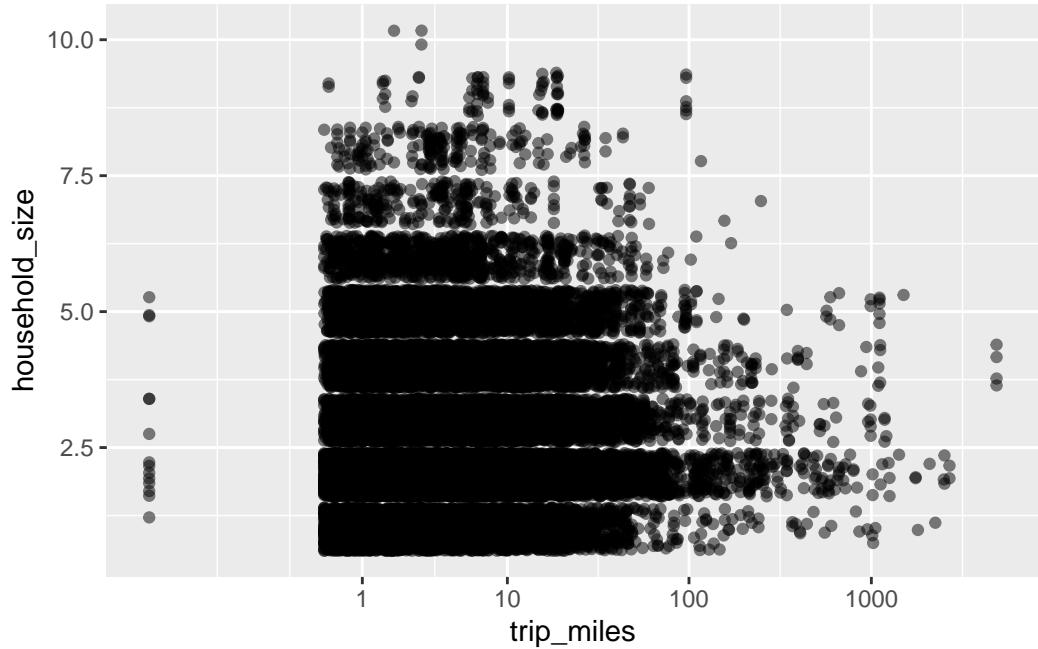


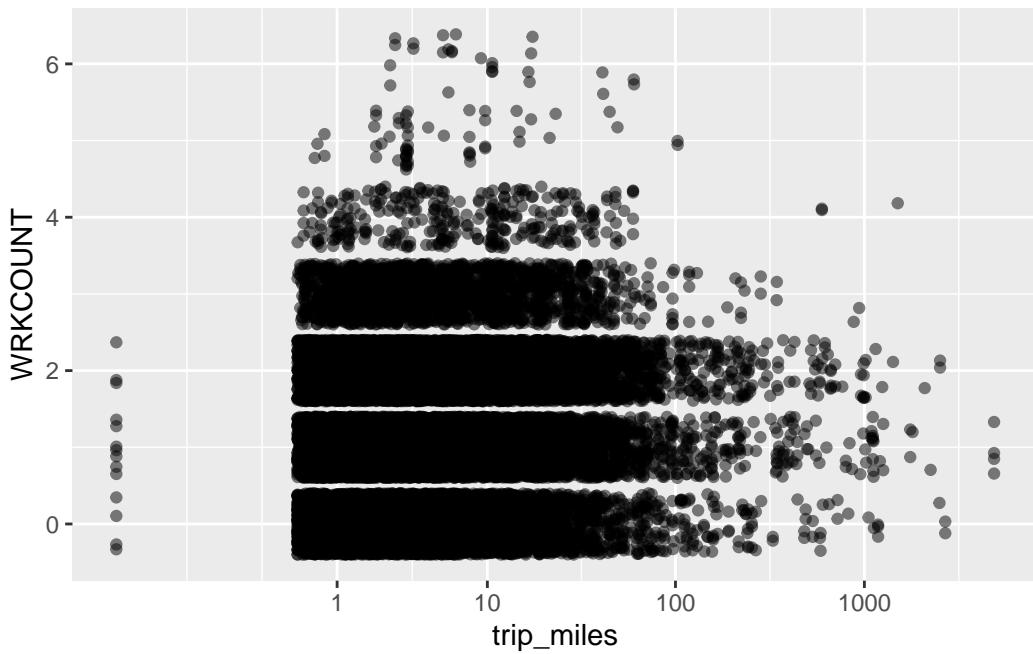
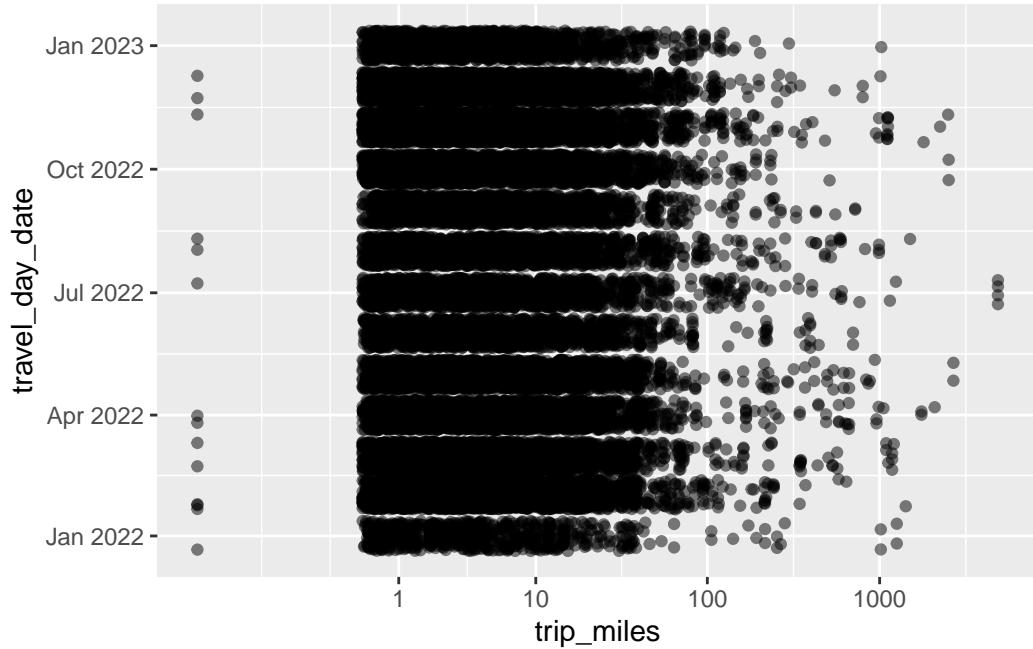


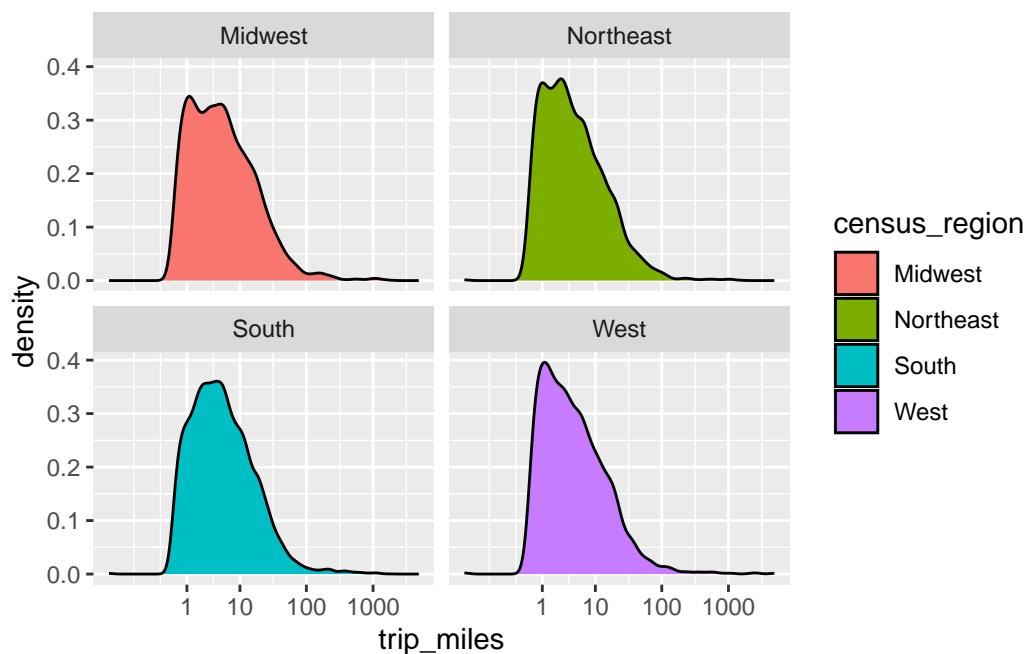
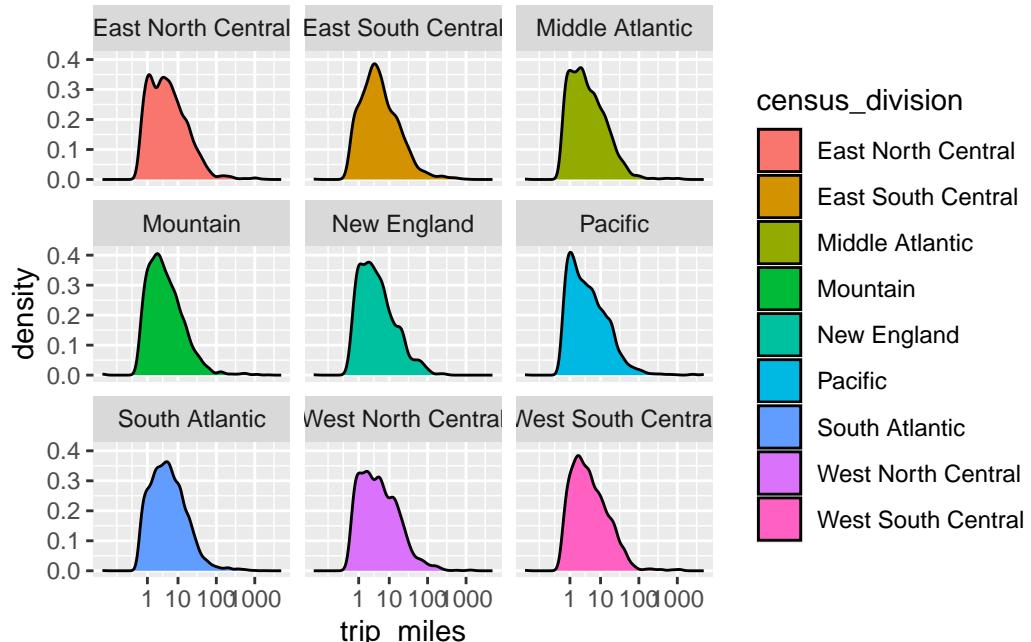


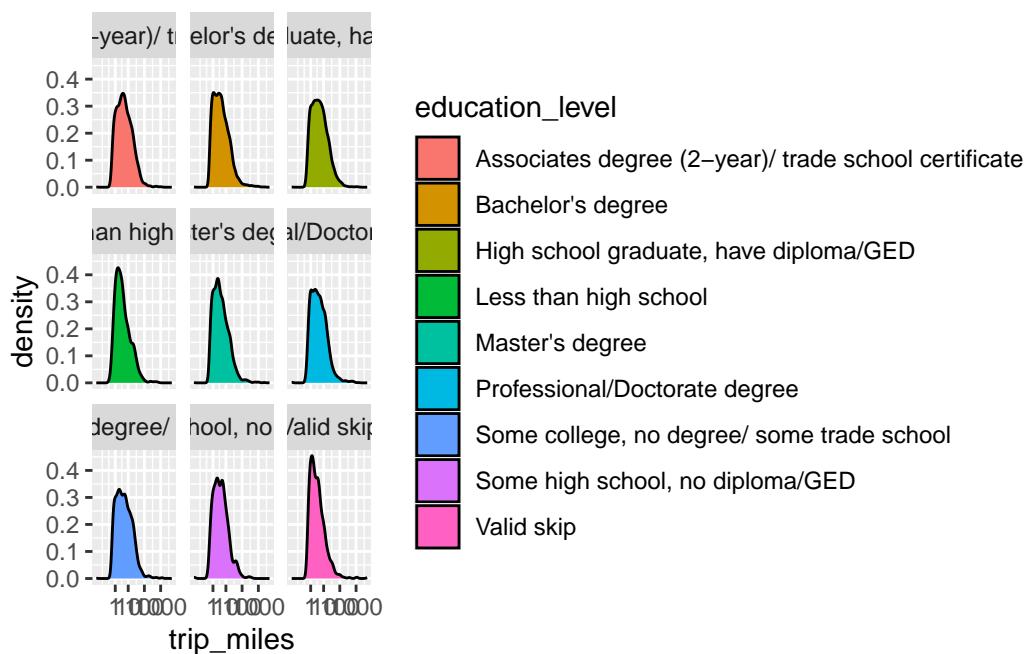
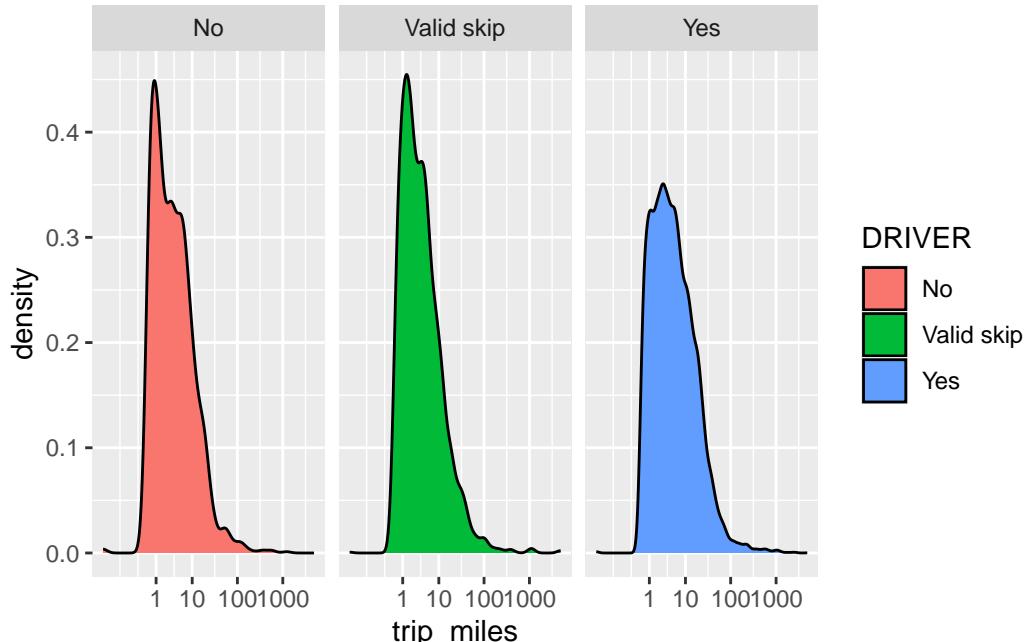


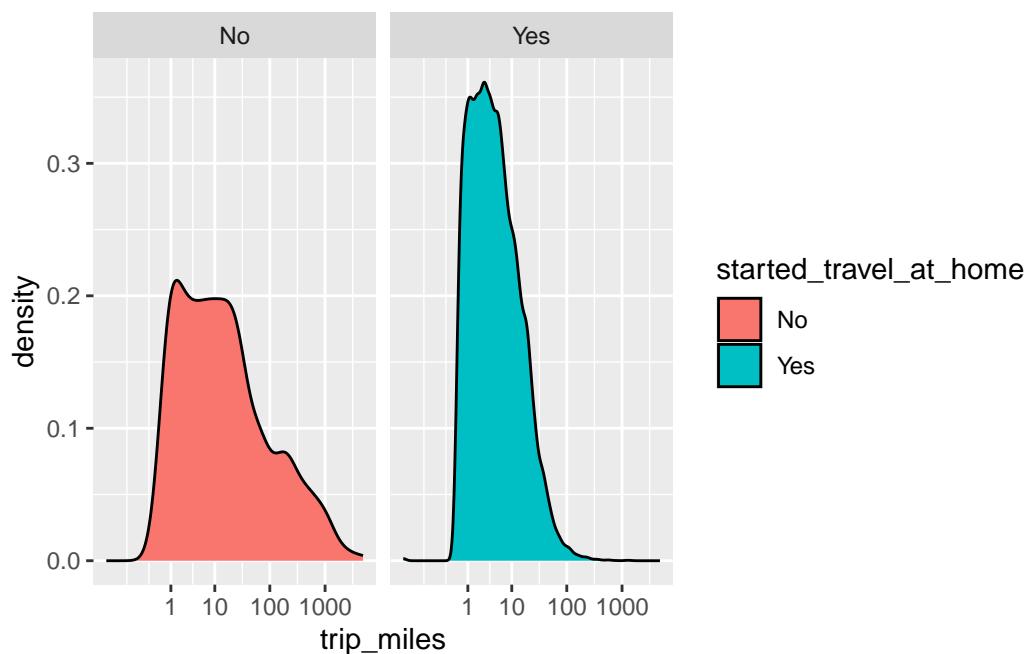
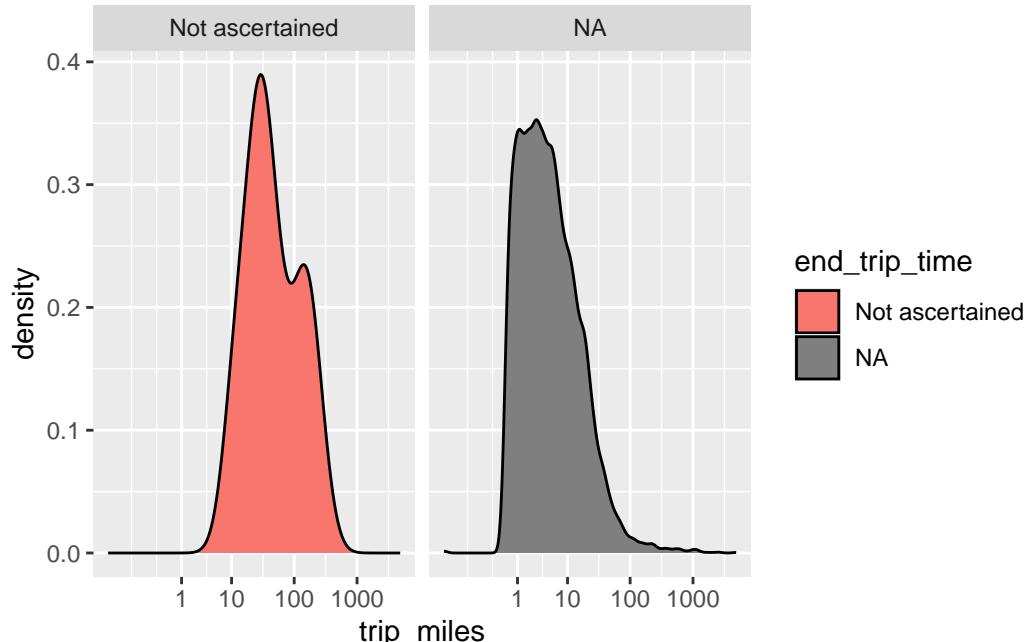


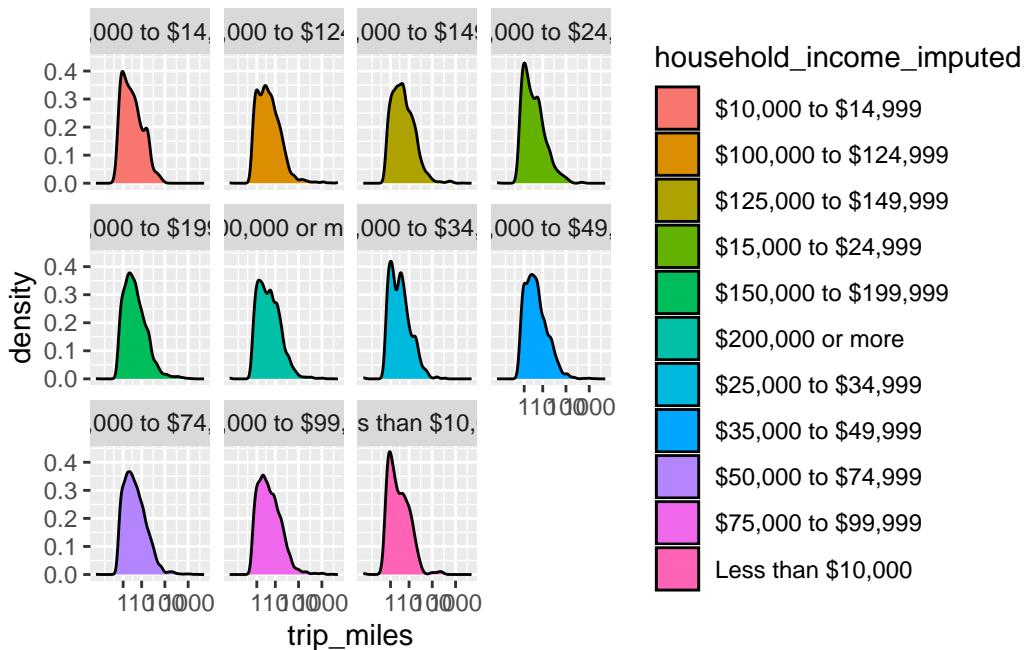
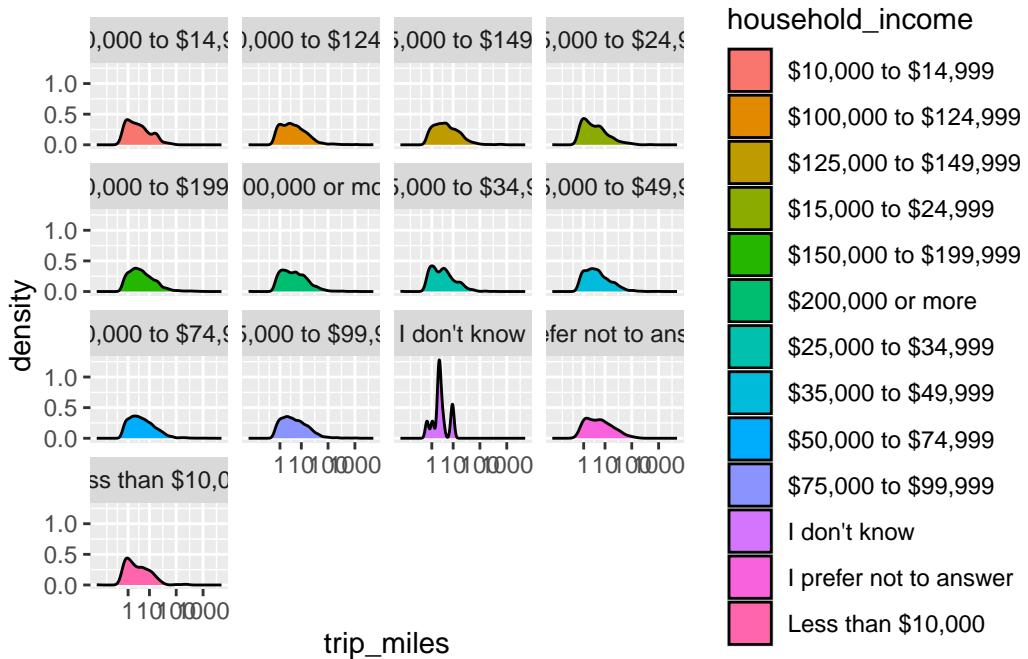


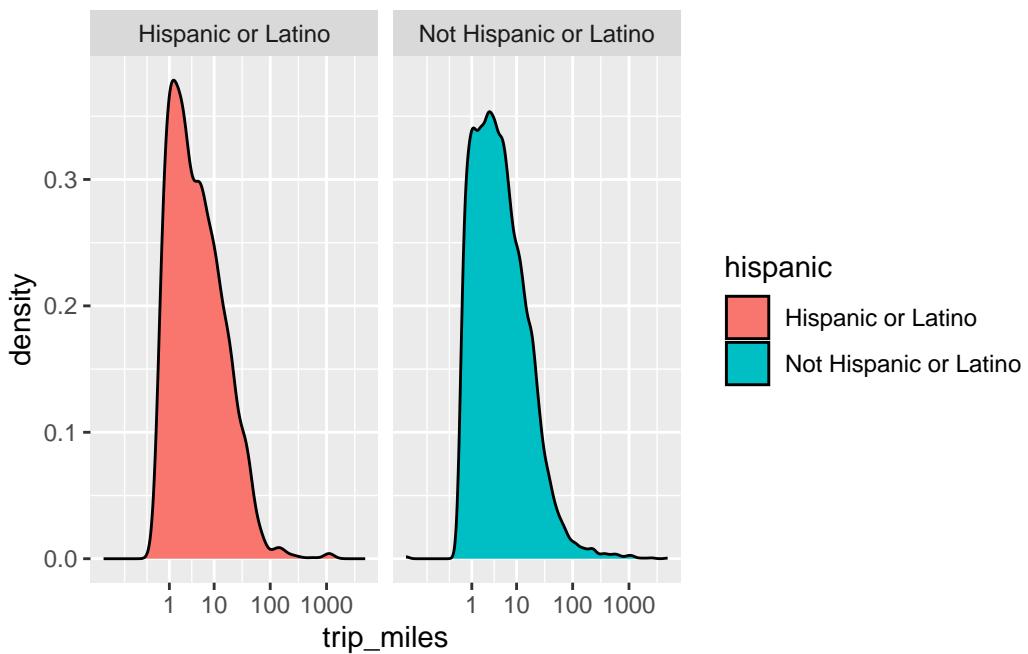
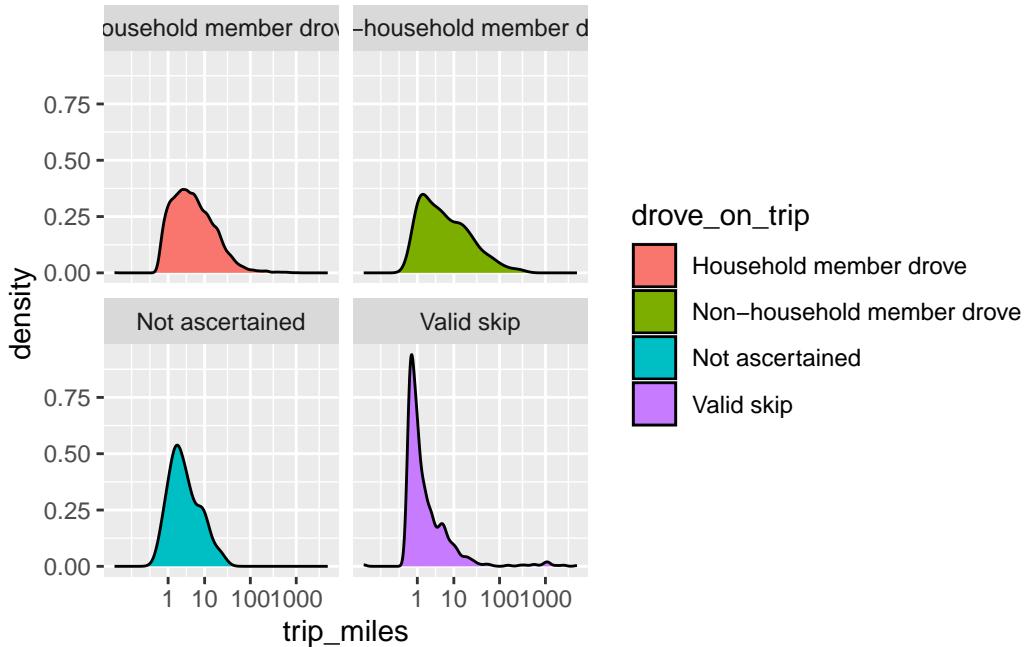


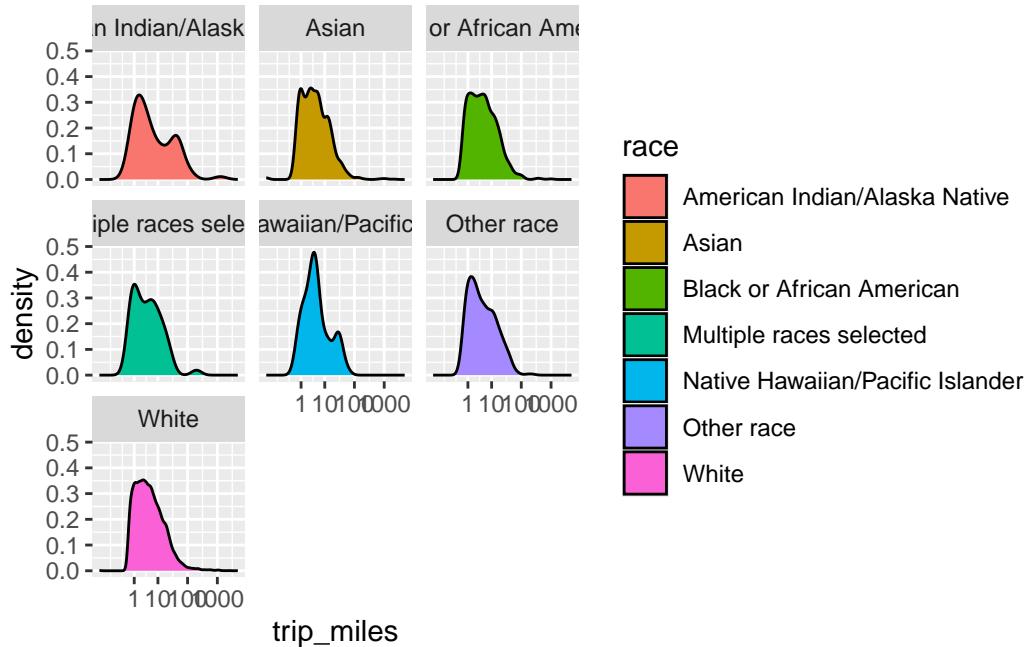


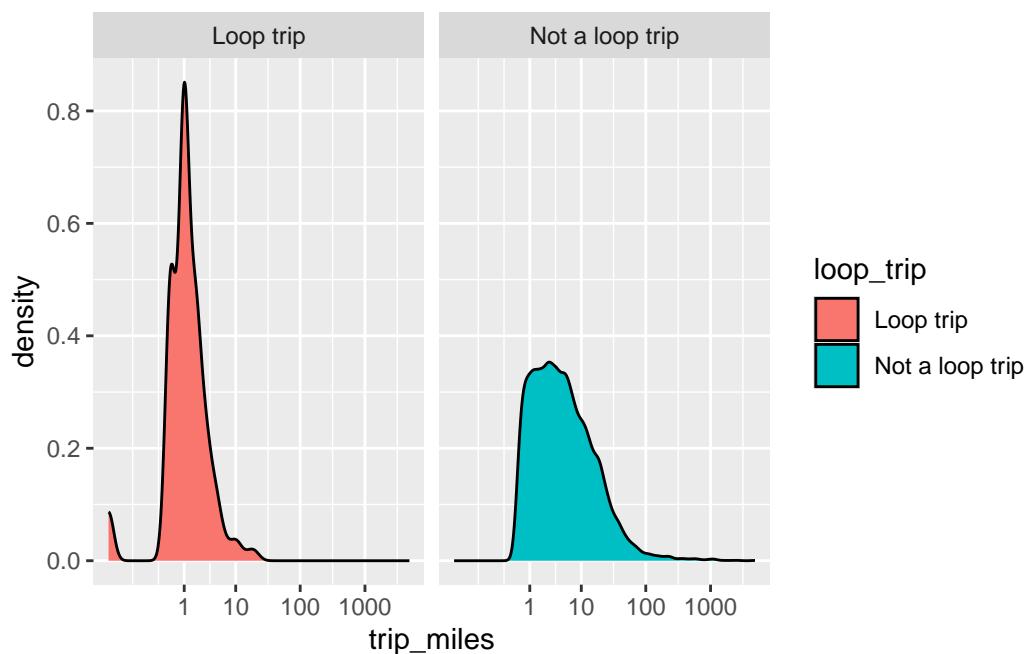
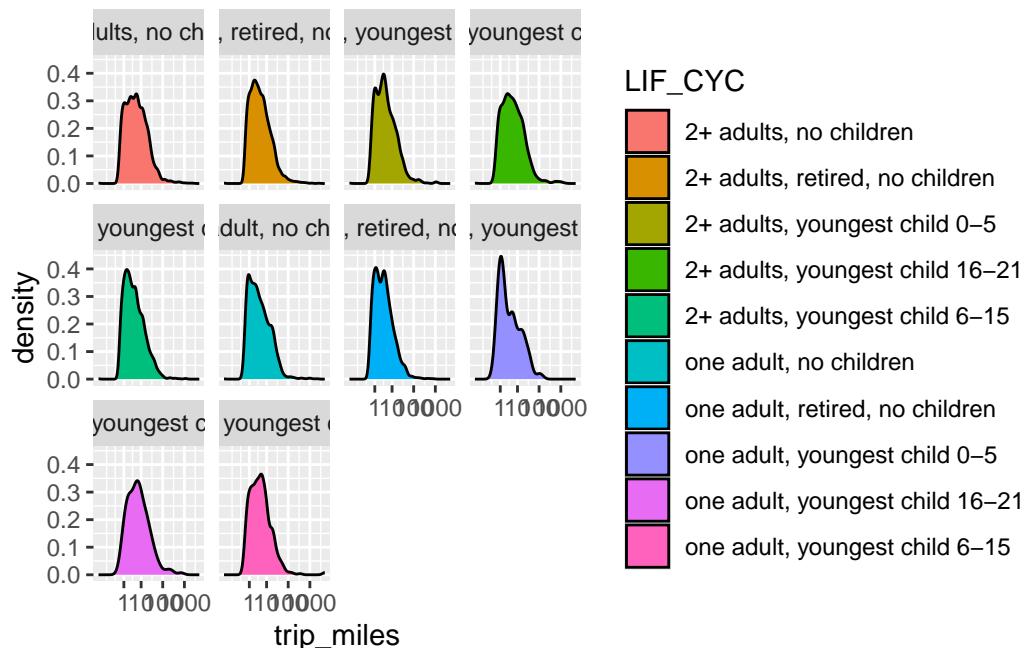


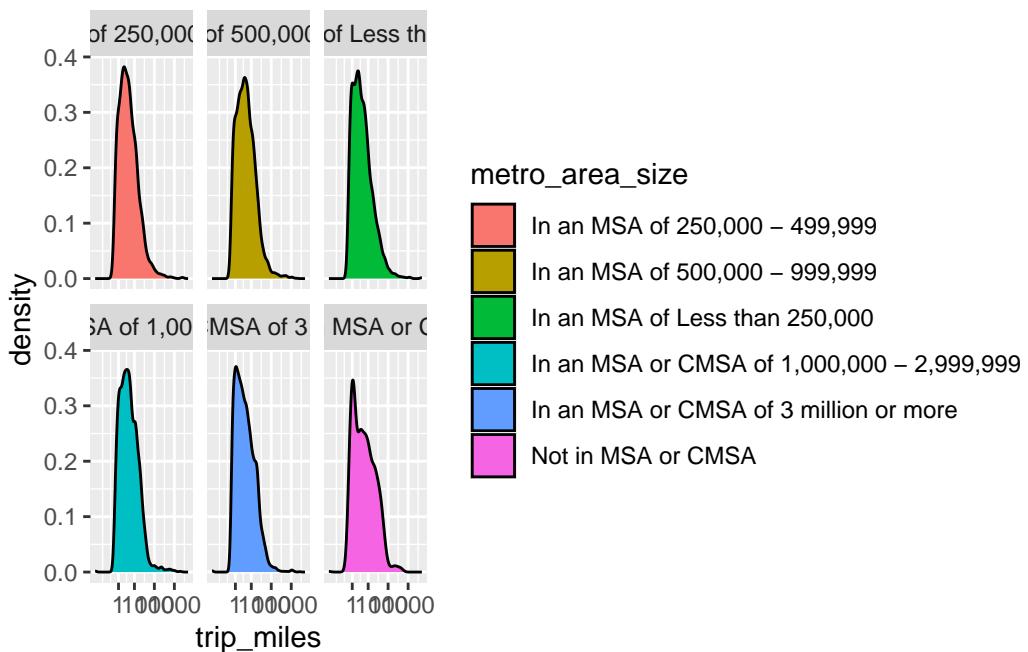
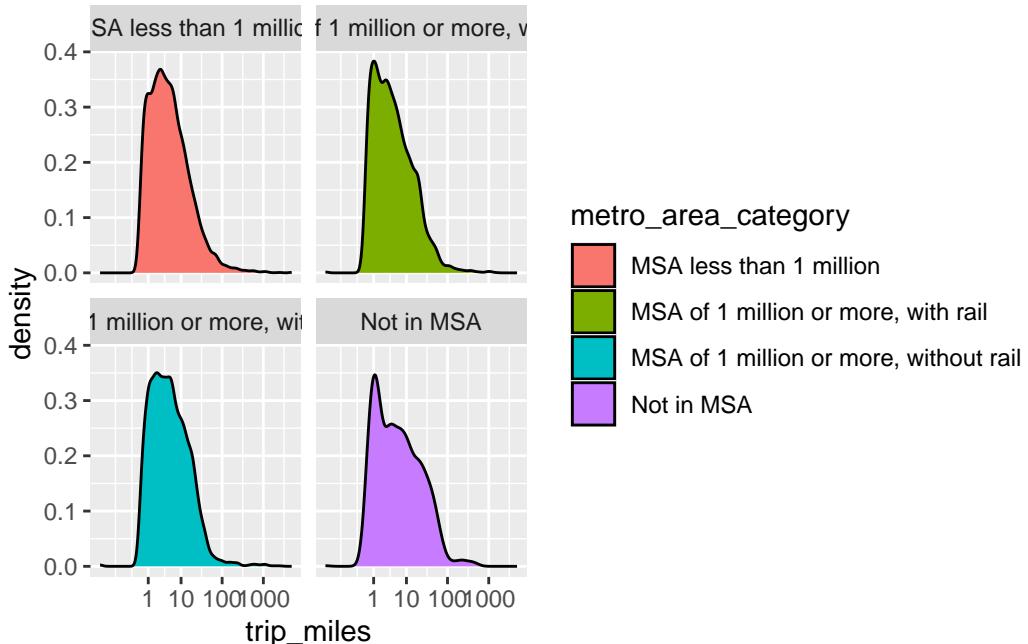


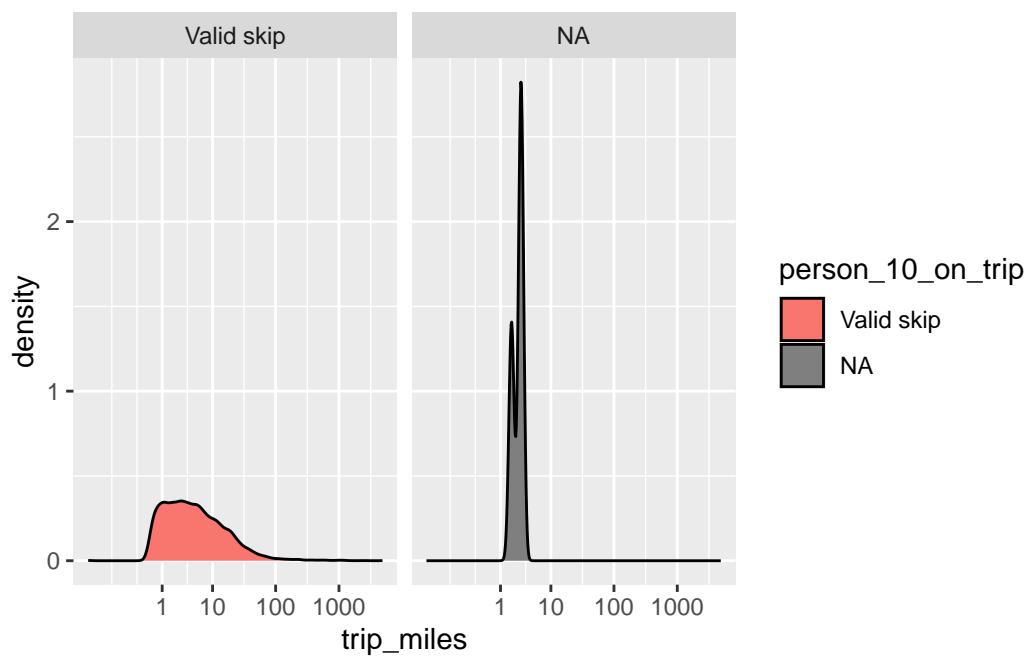
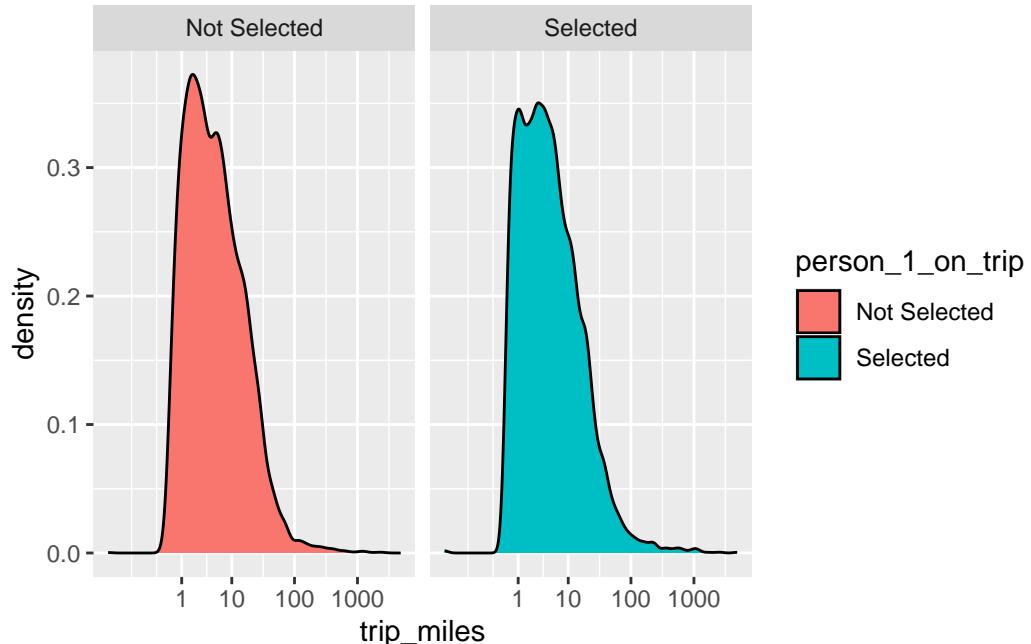


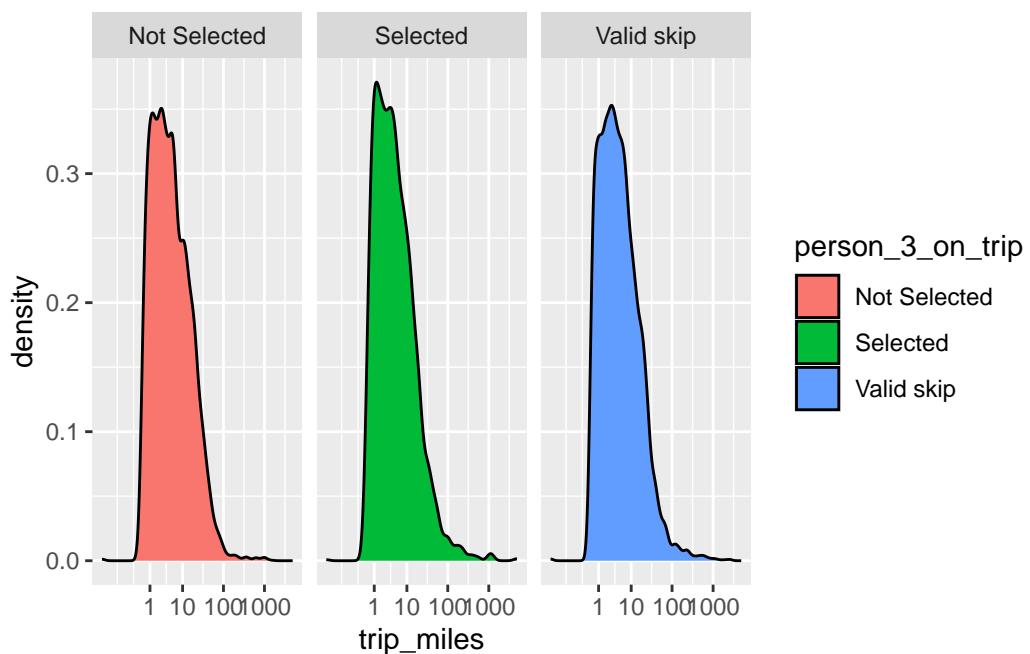
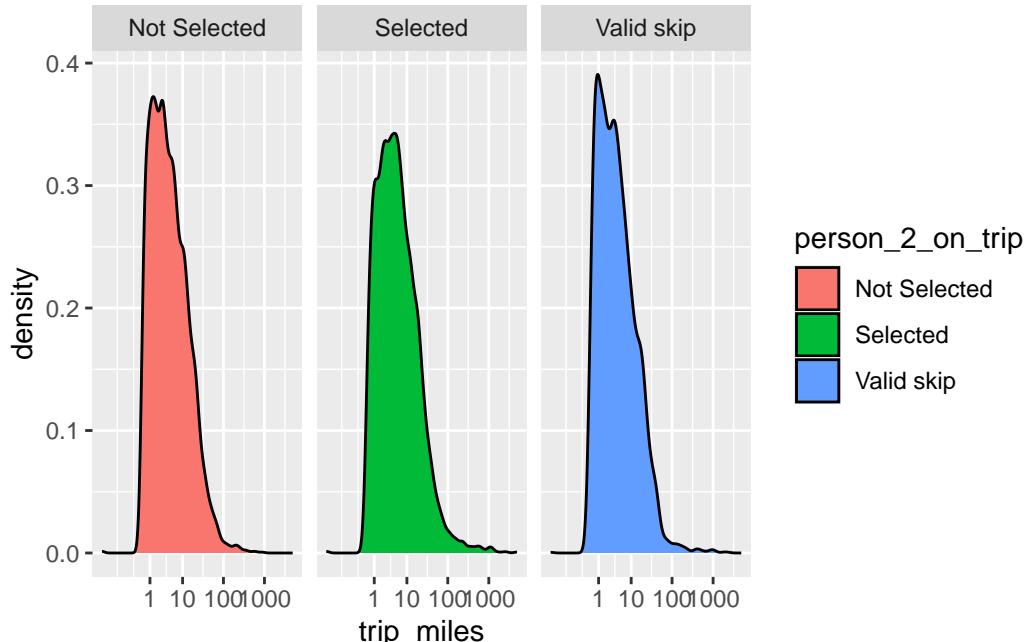


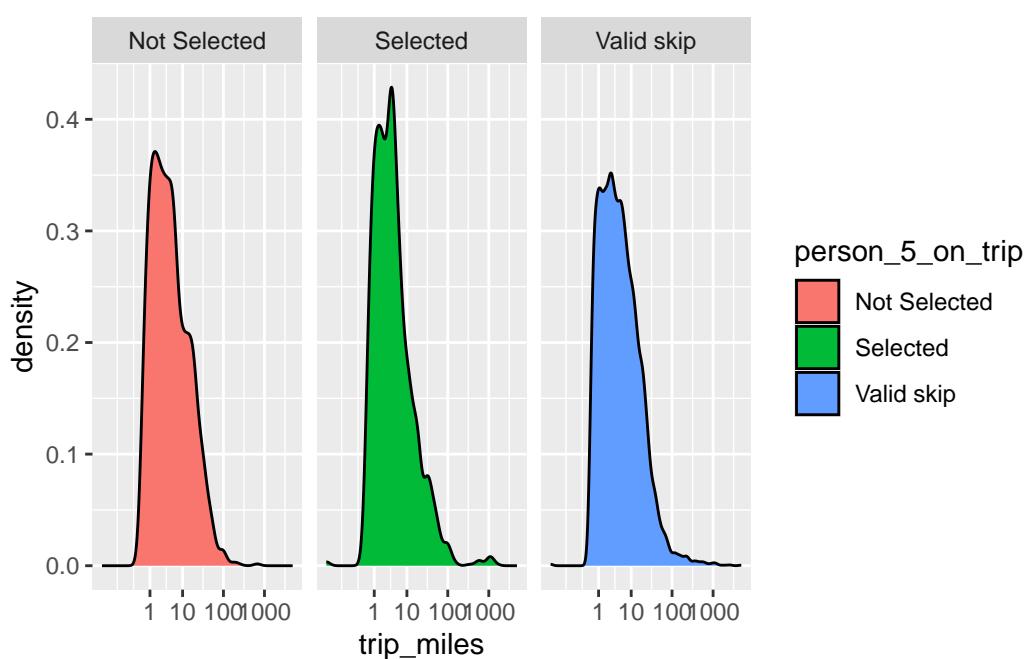
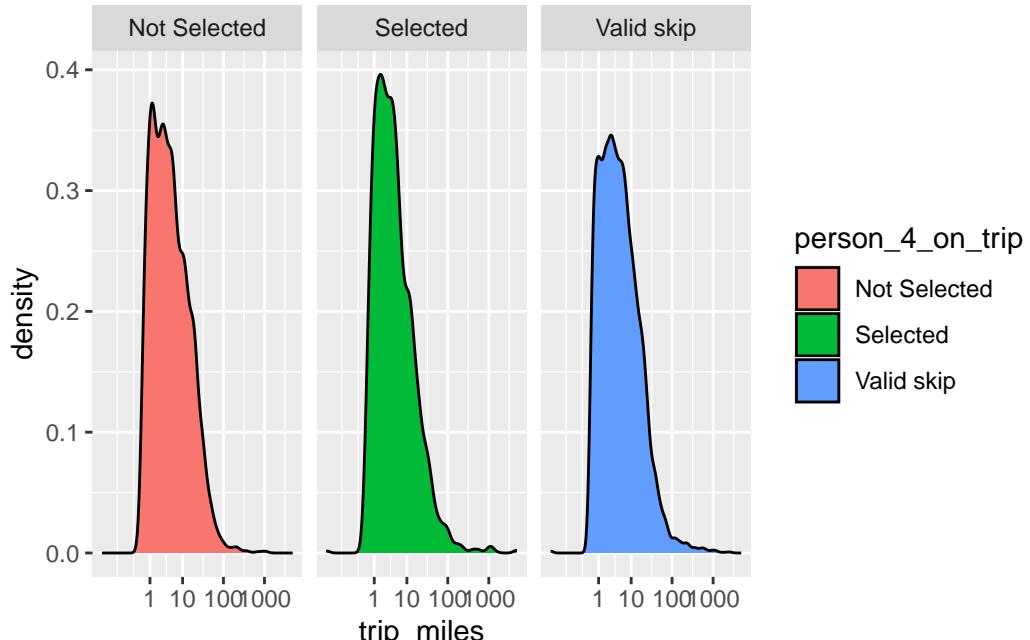


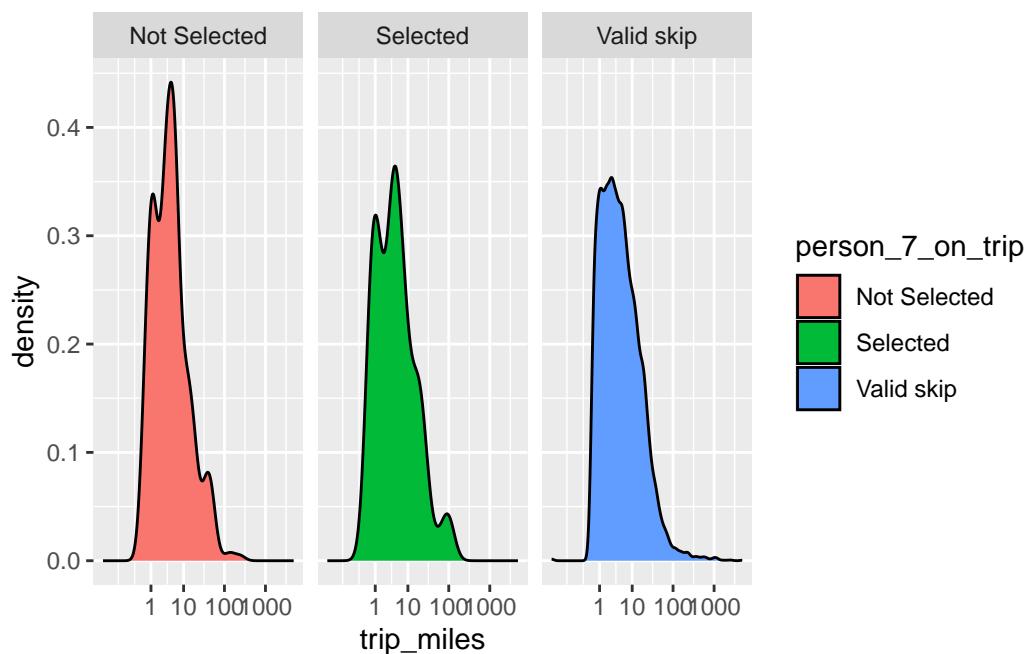
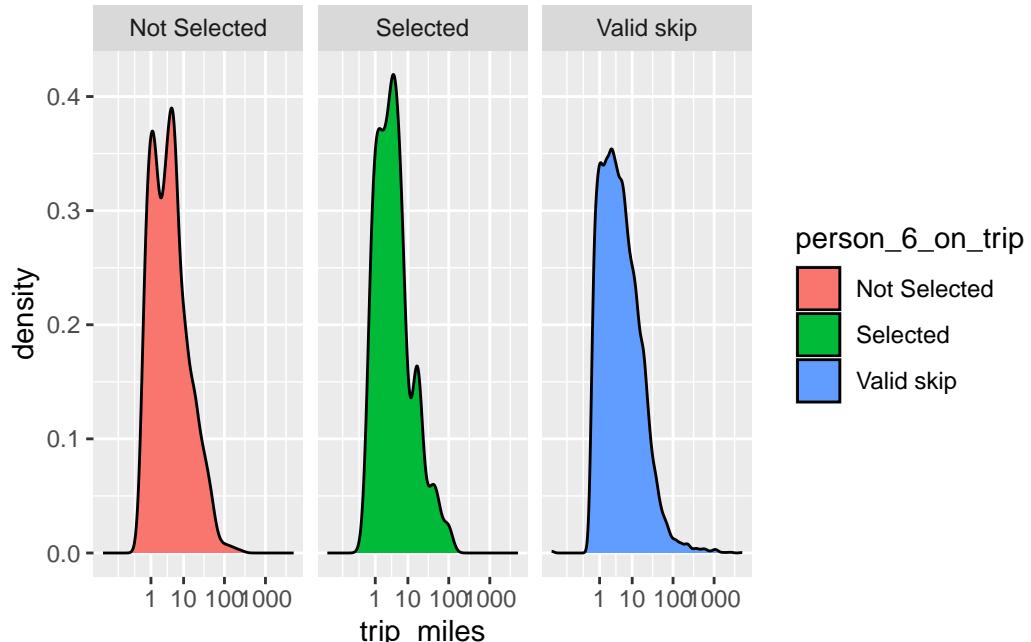


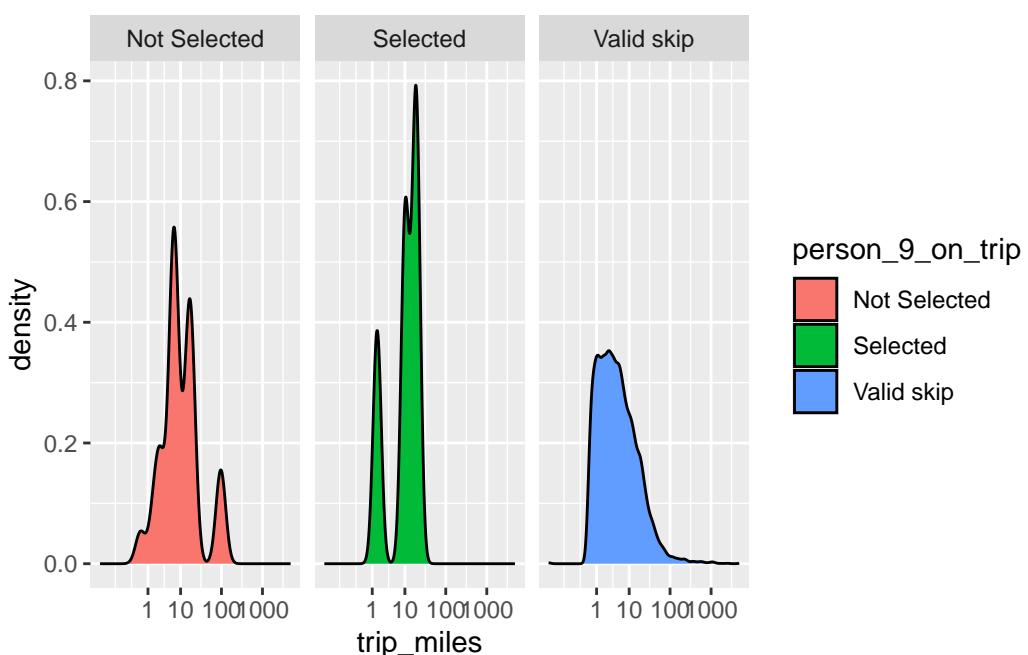
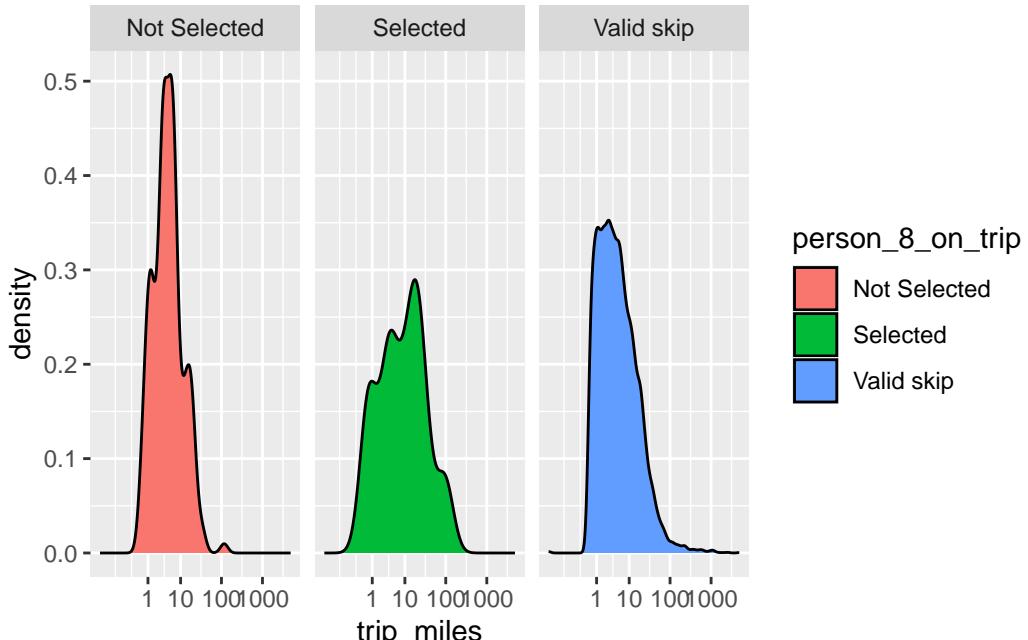








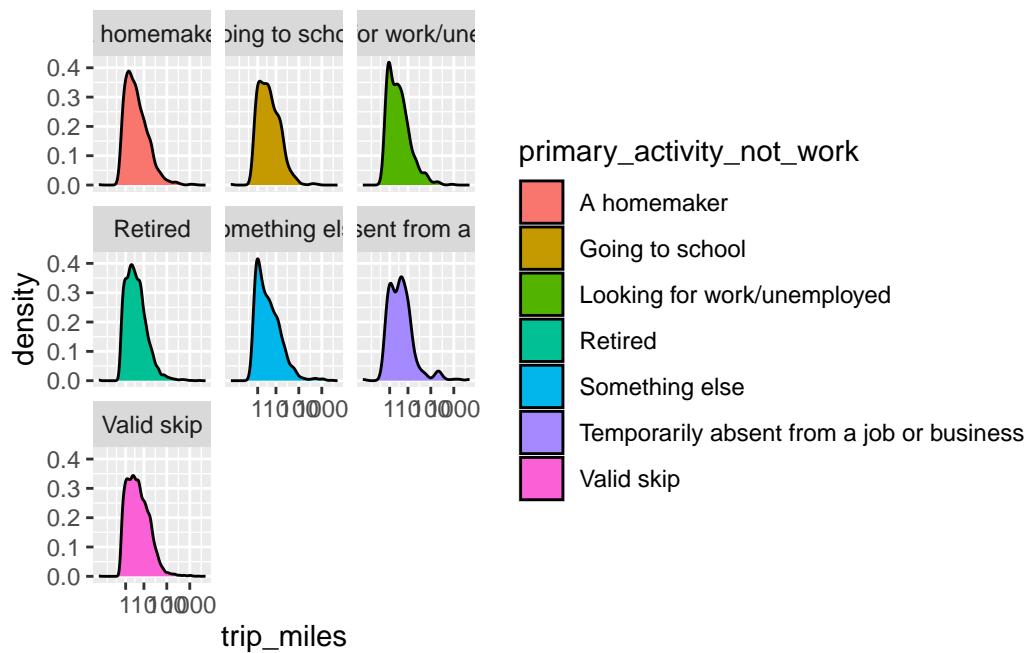
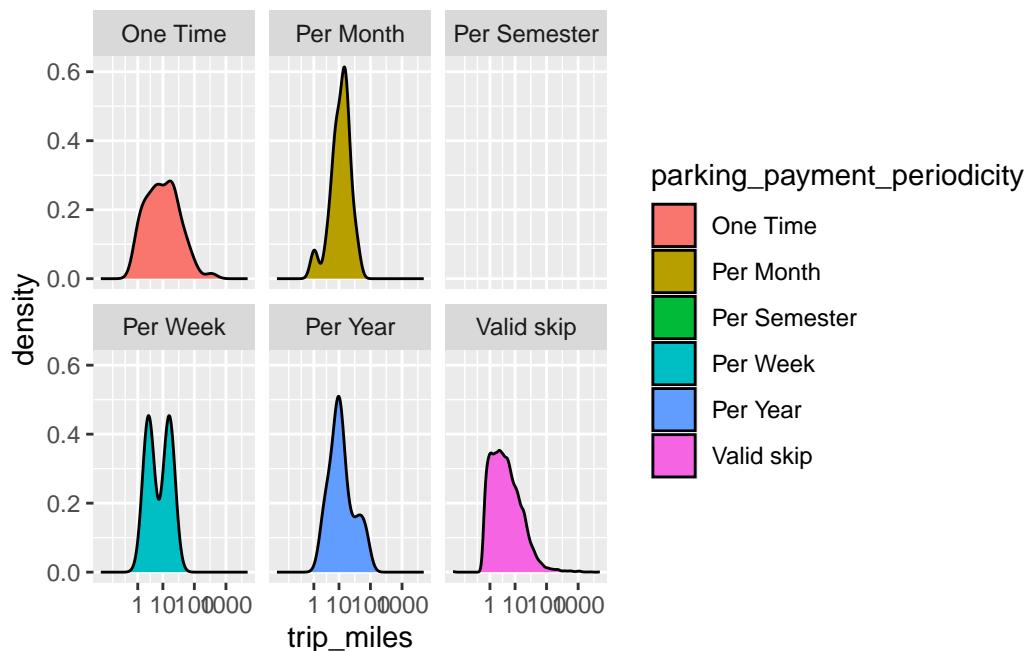


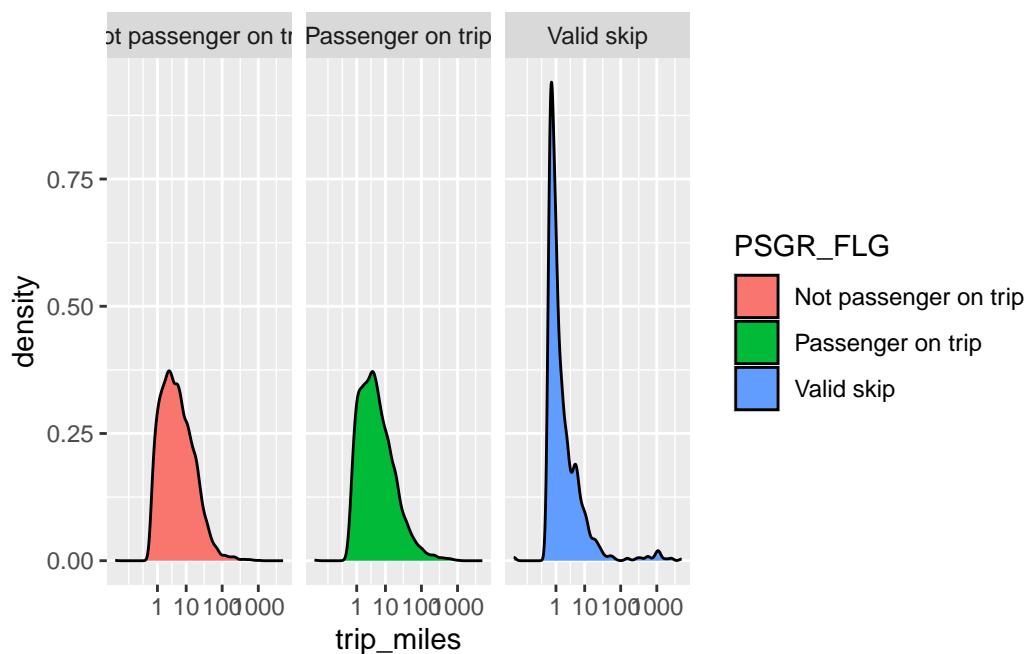
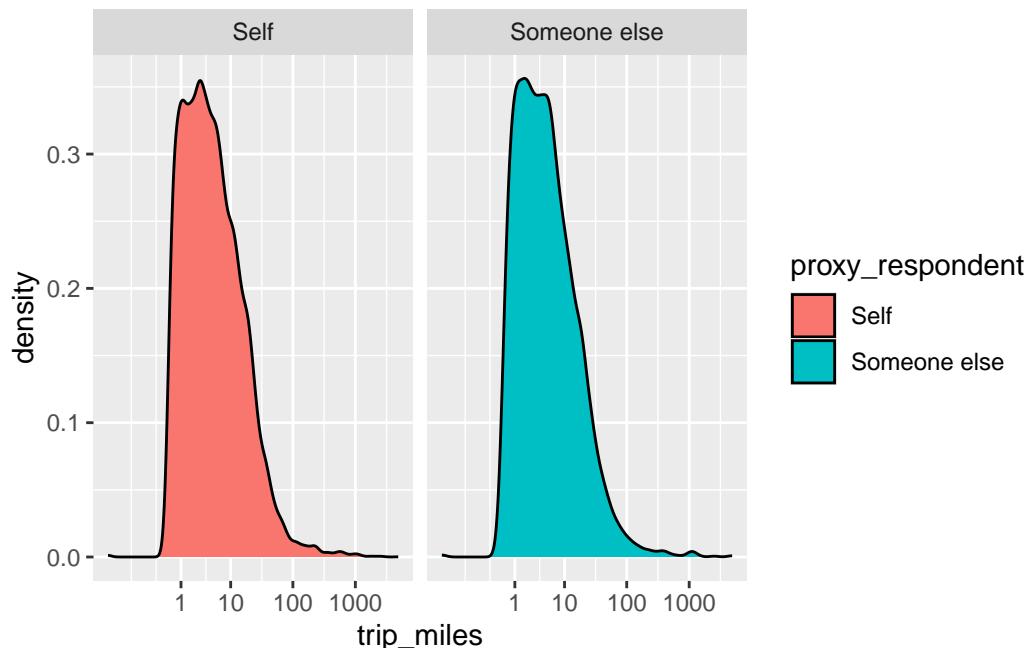


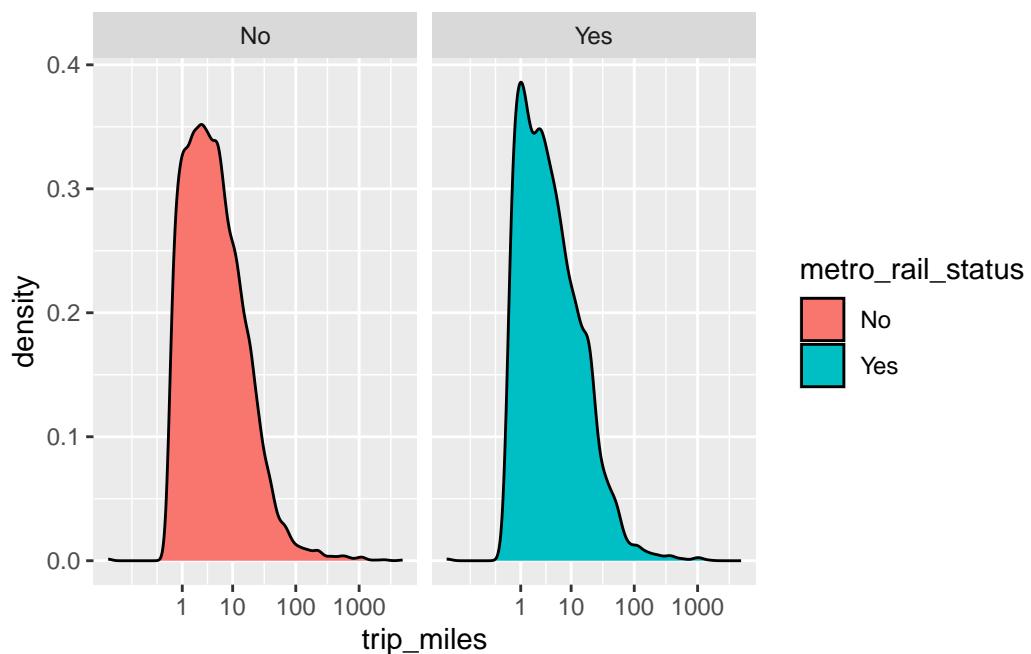
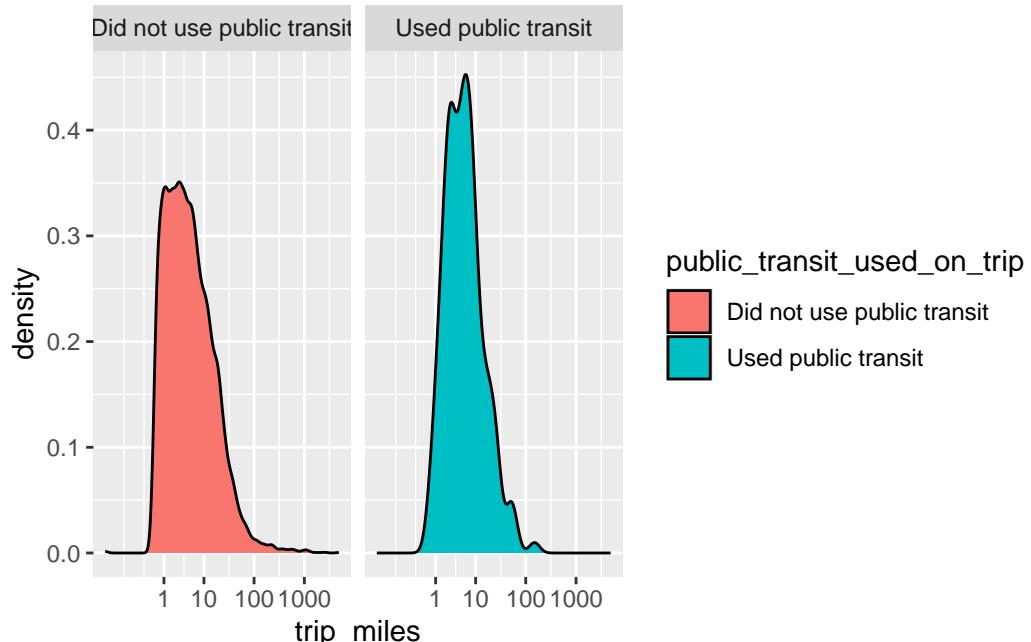
Warning: Groups with fewer than two data points have been dropped.

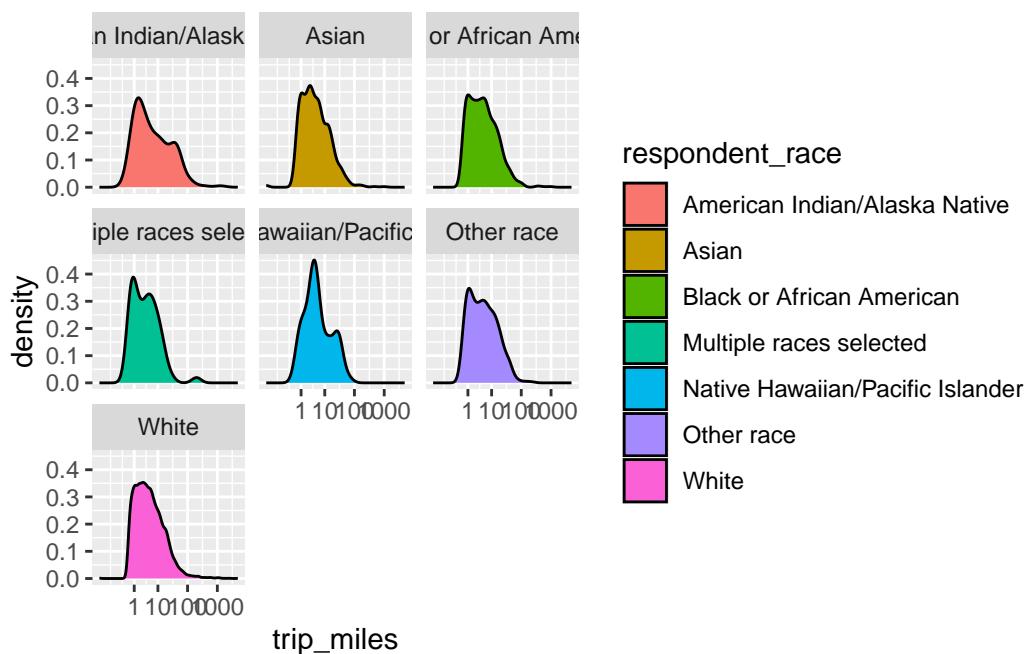
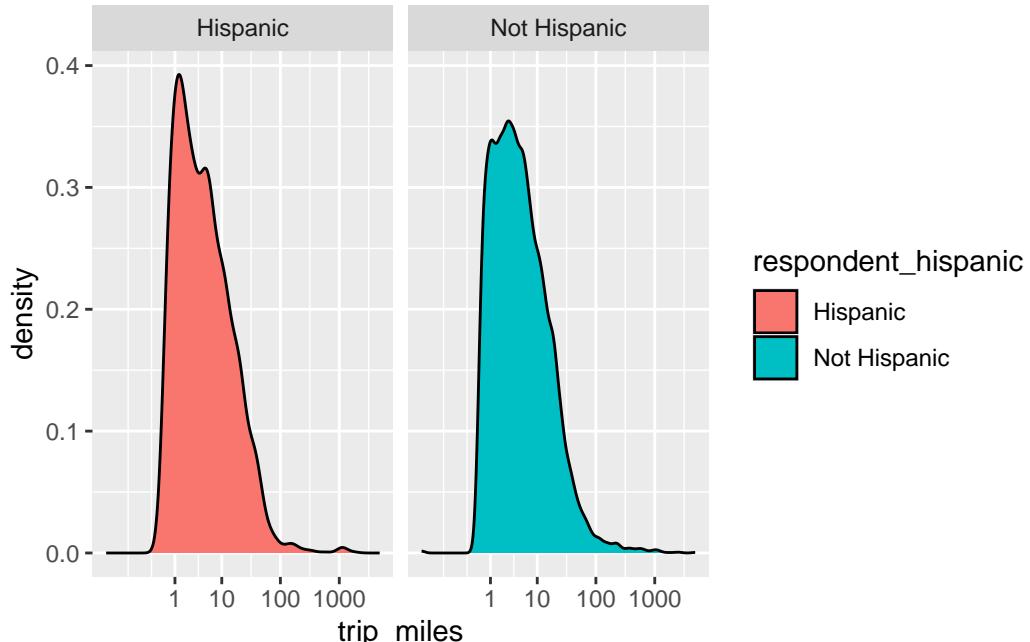
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning

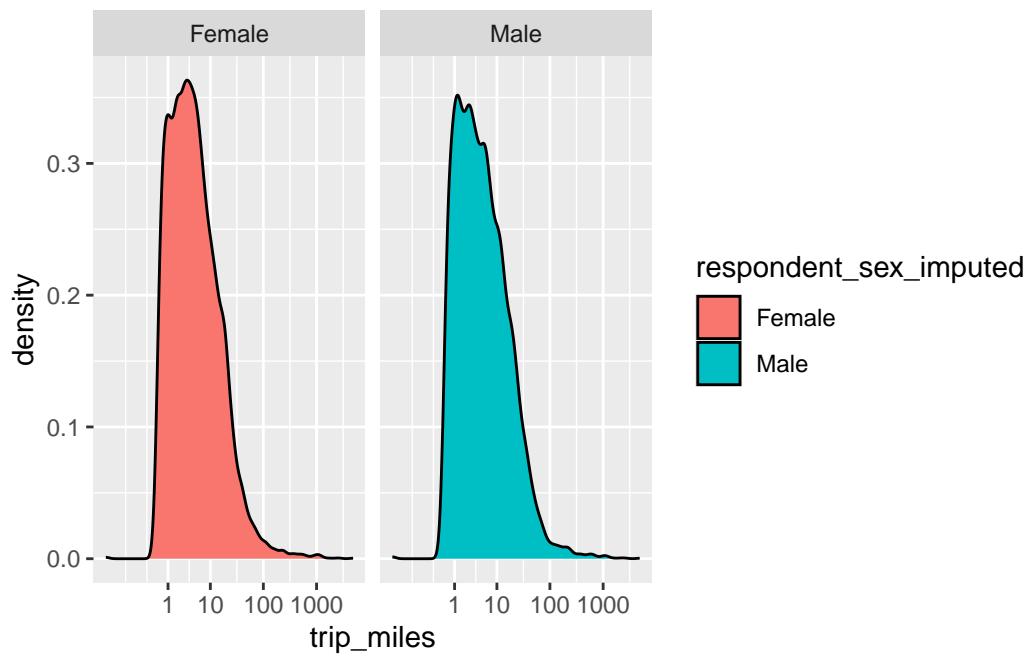
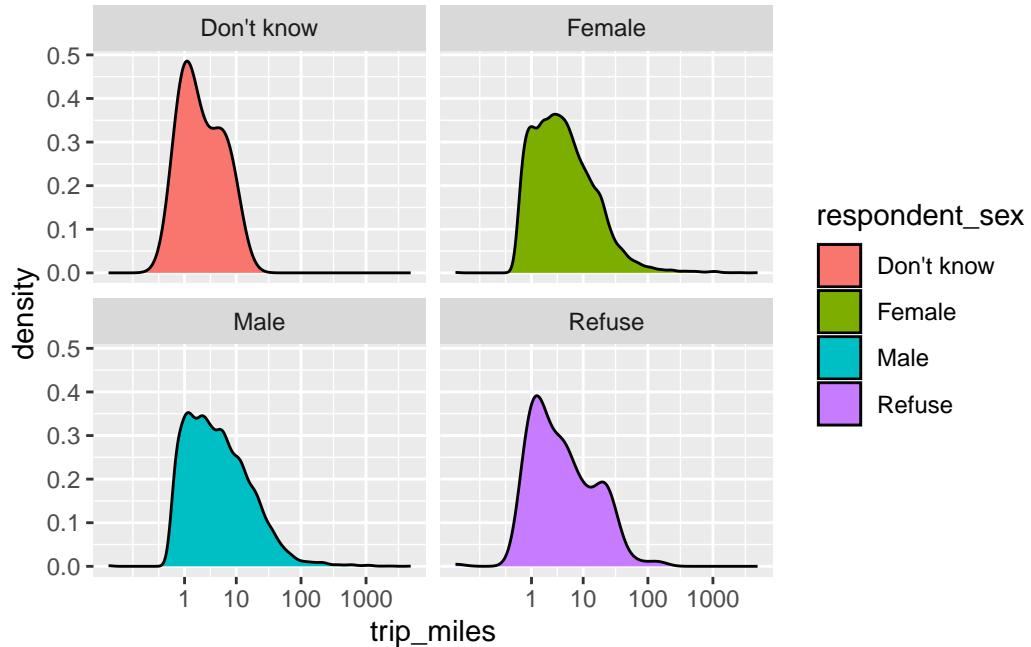
-Inf

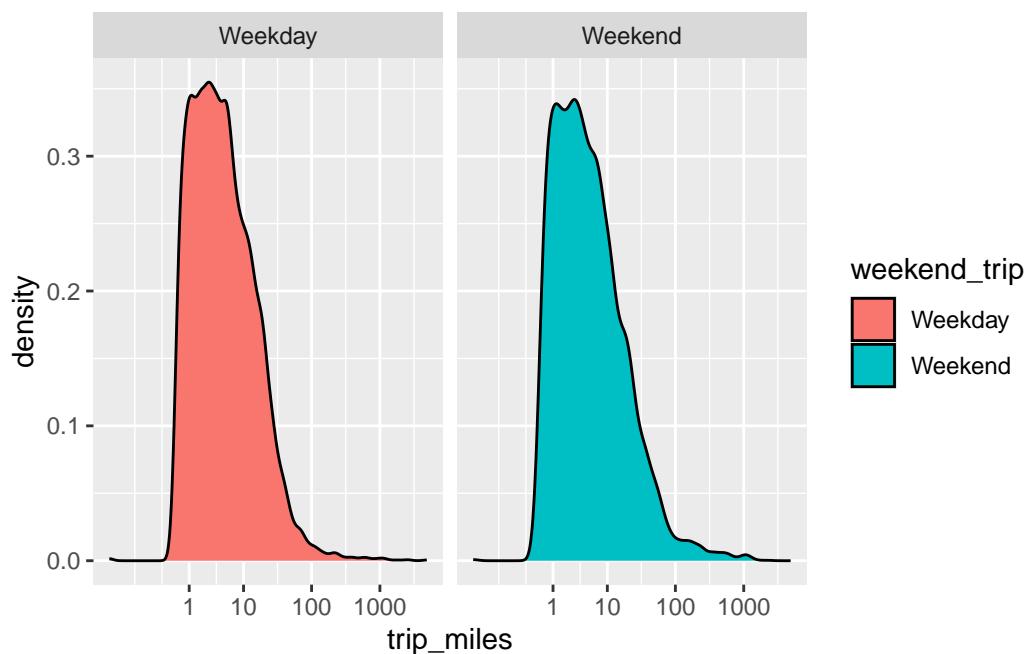
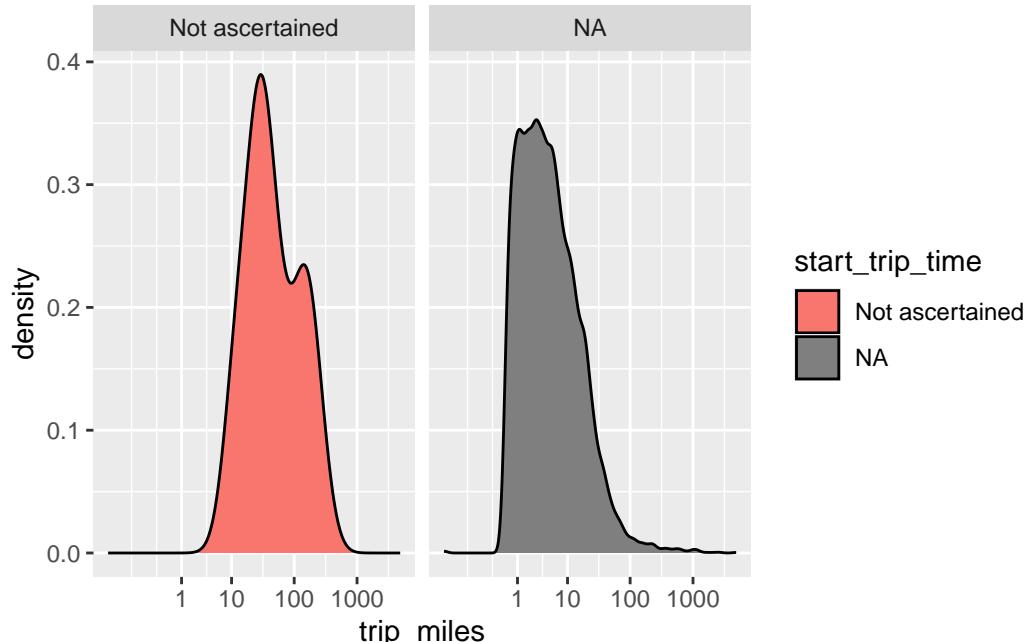


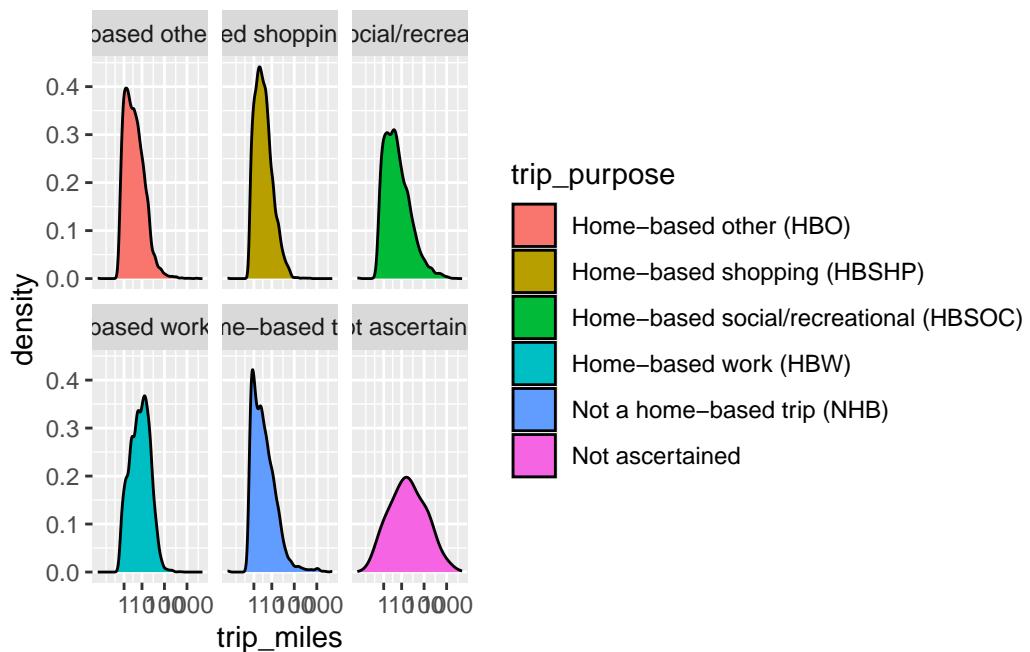
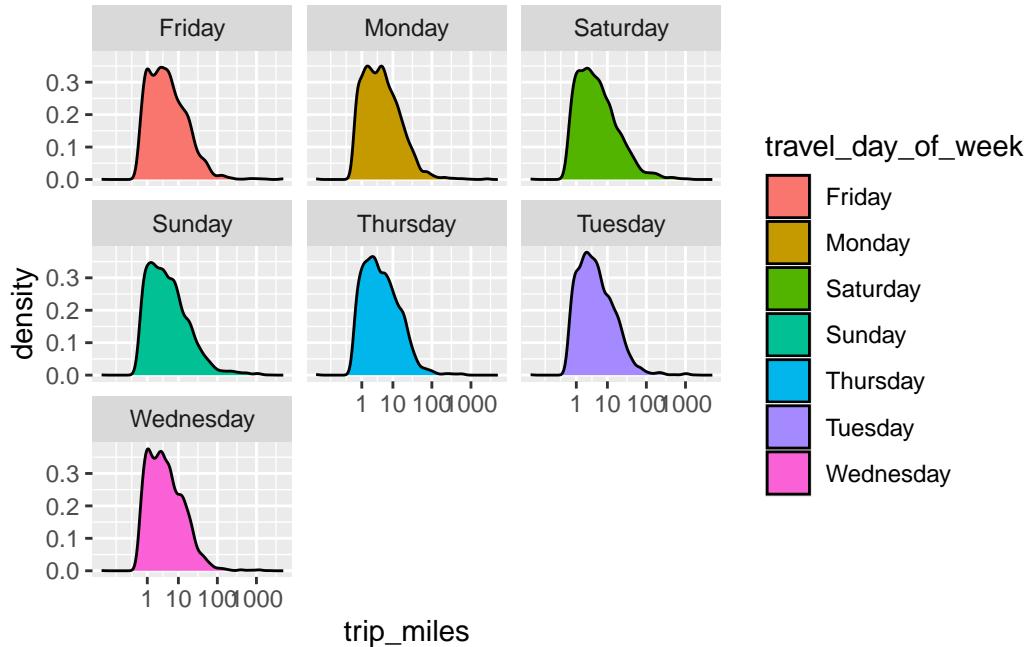


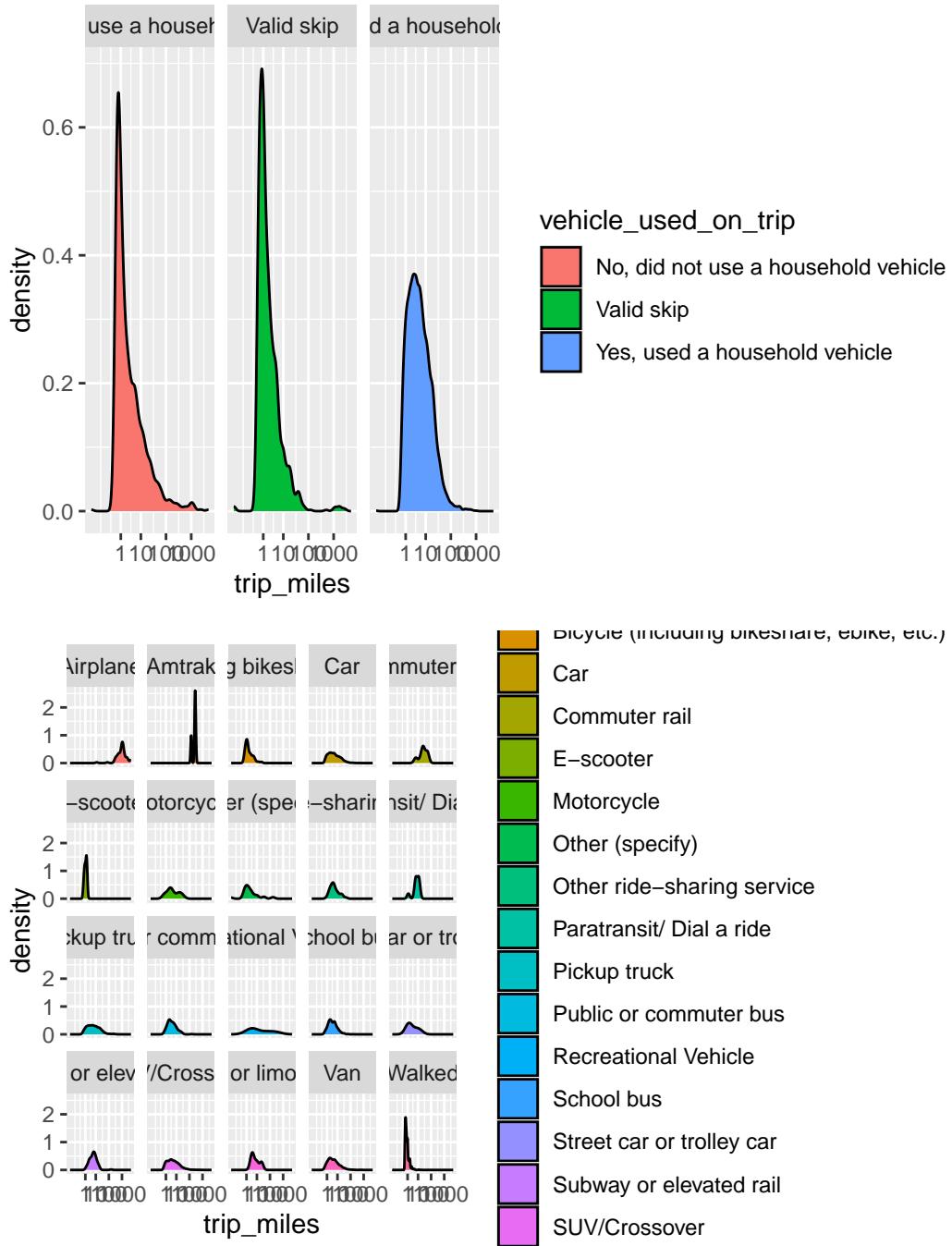


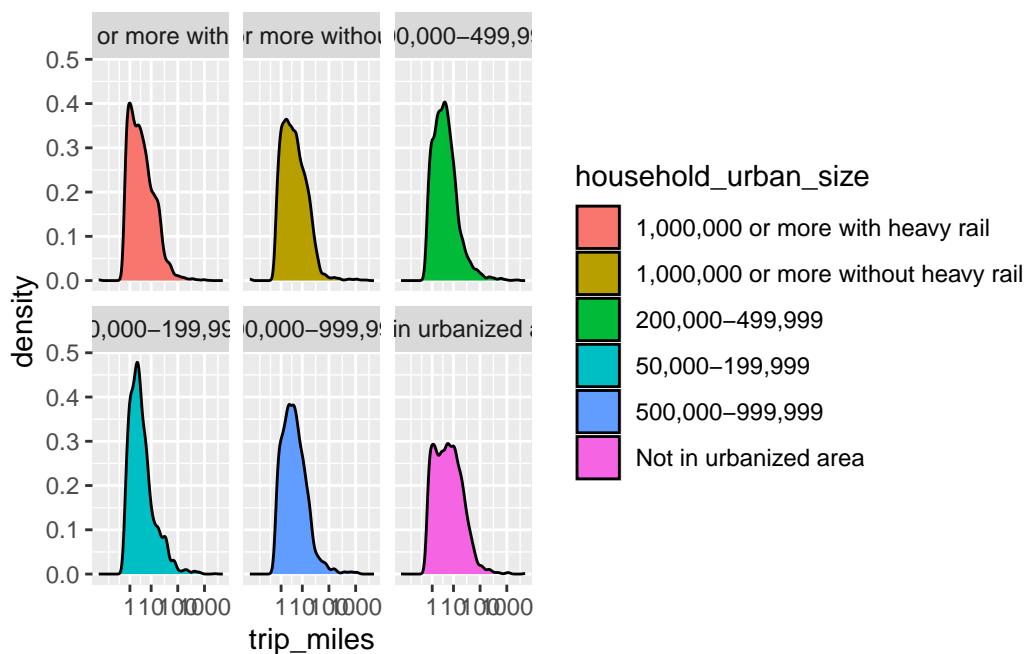
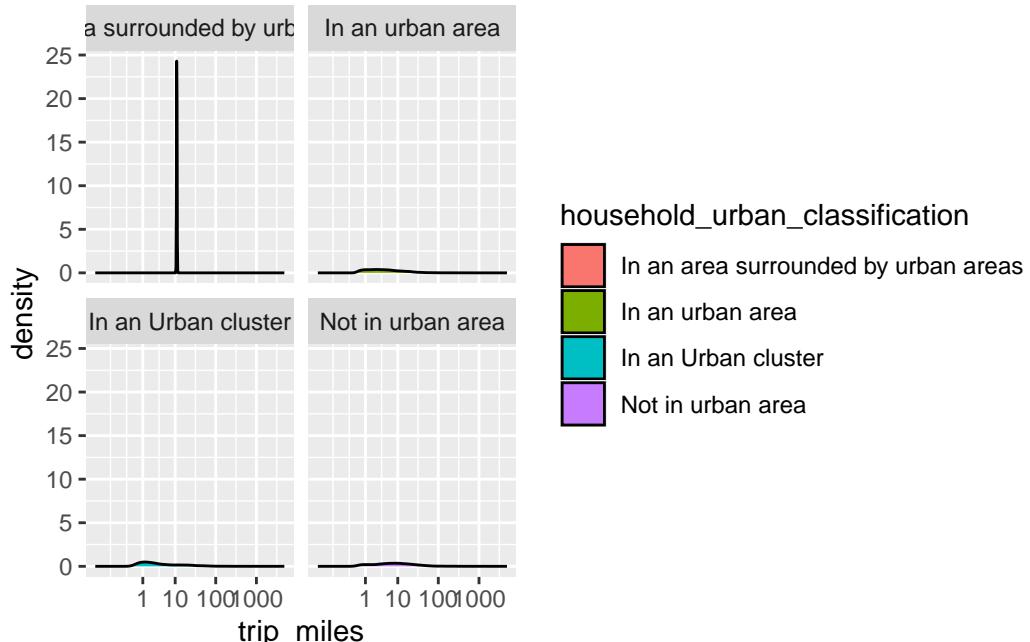


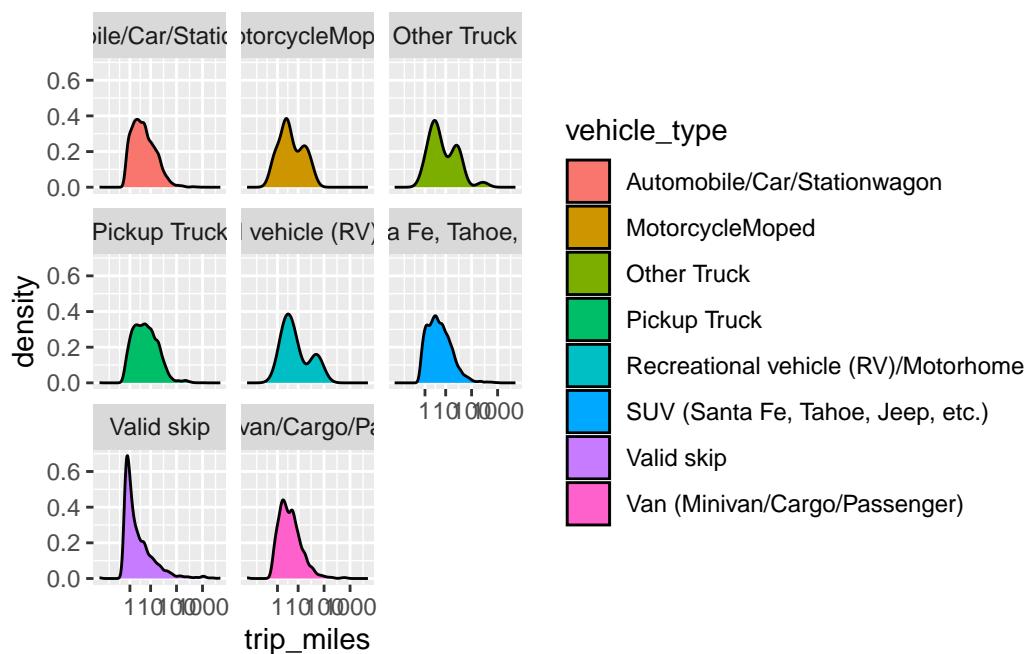
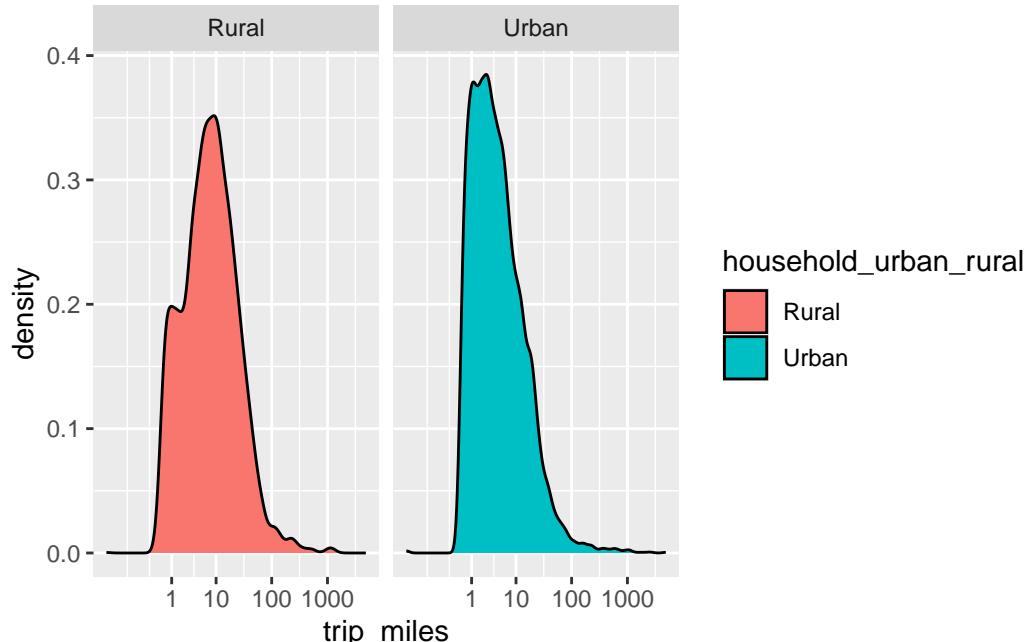


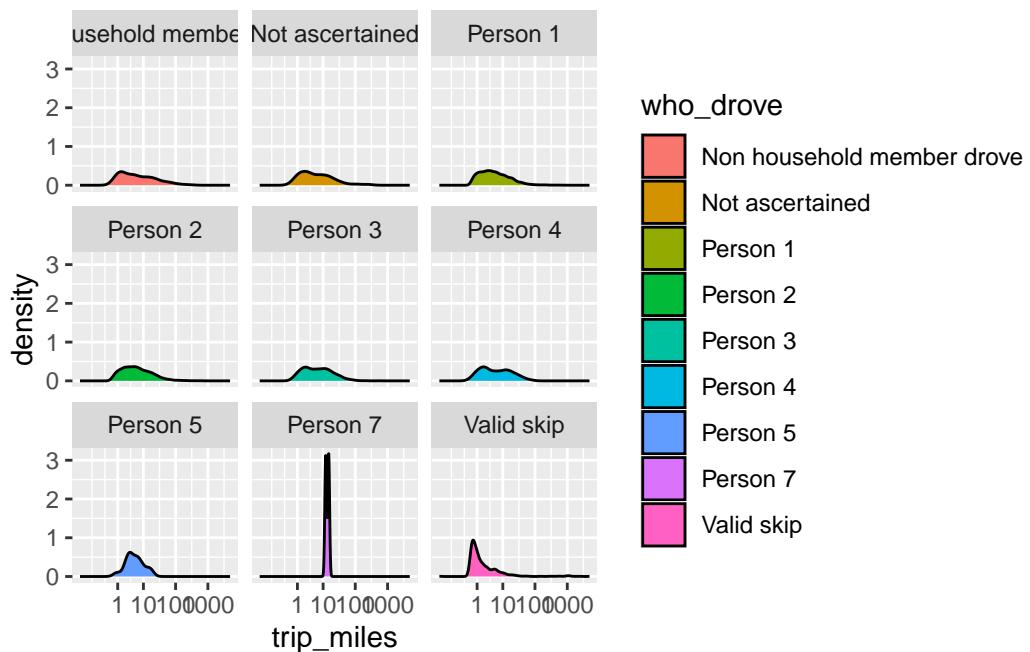
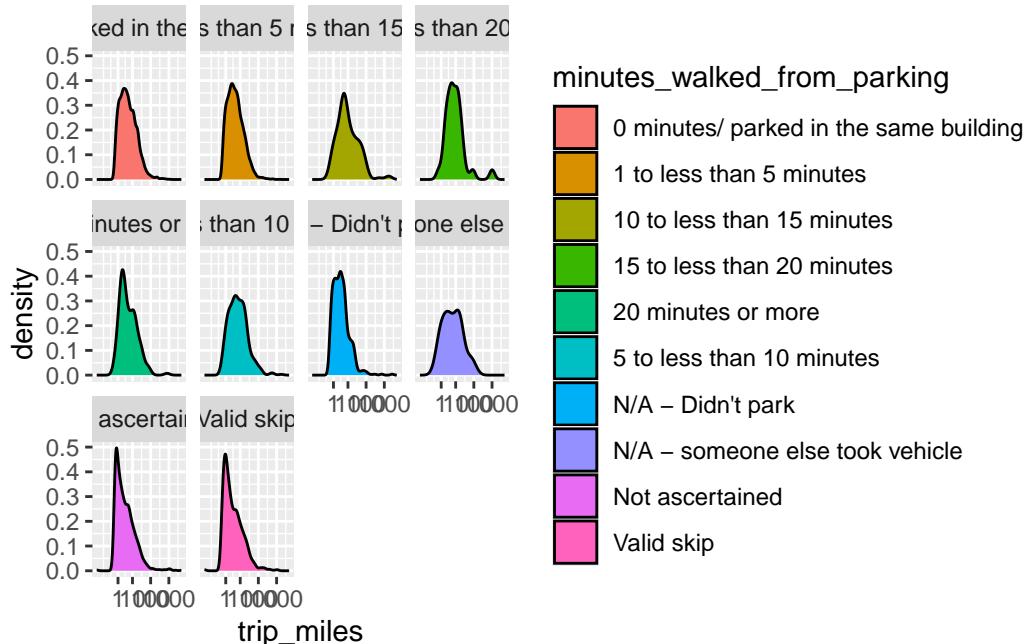


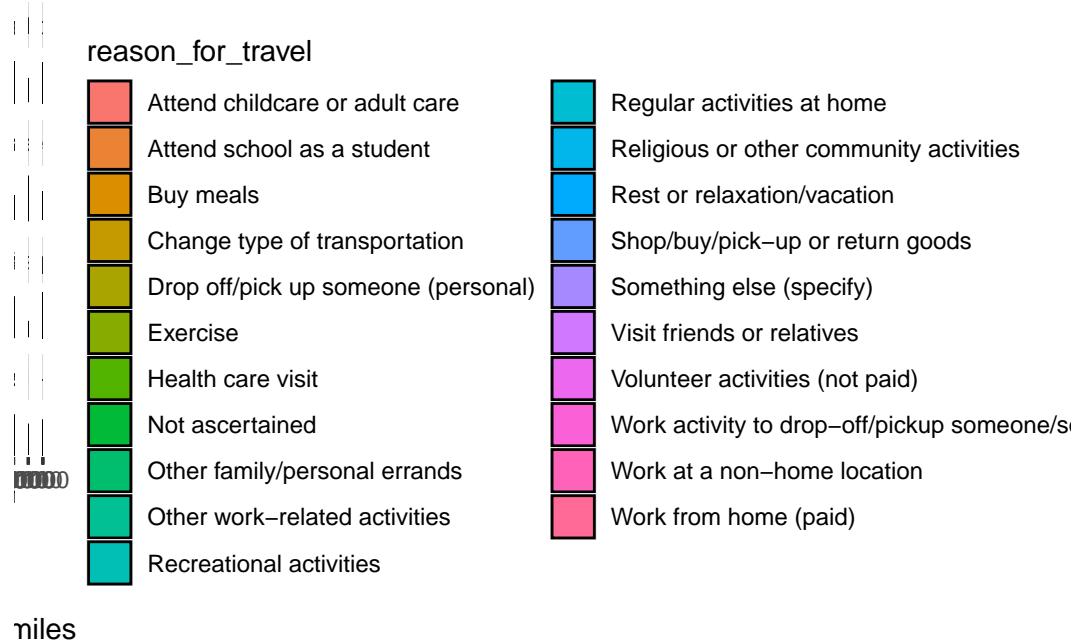
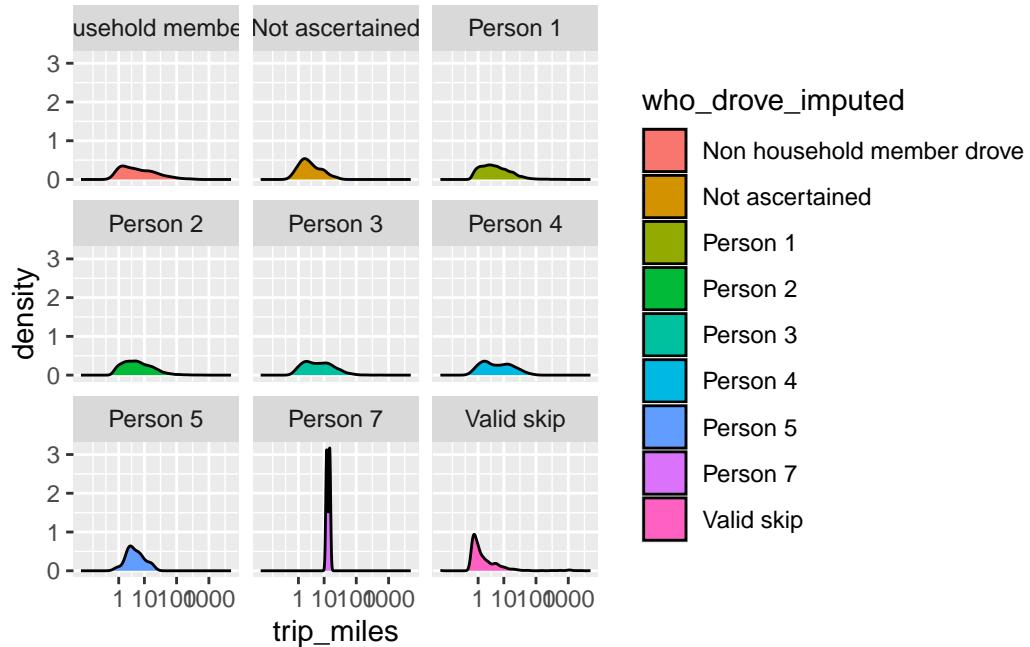


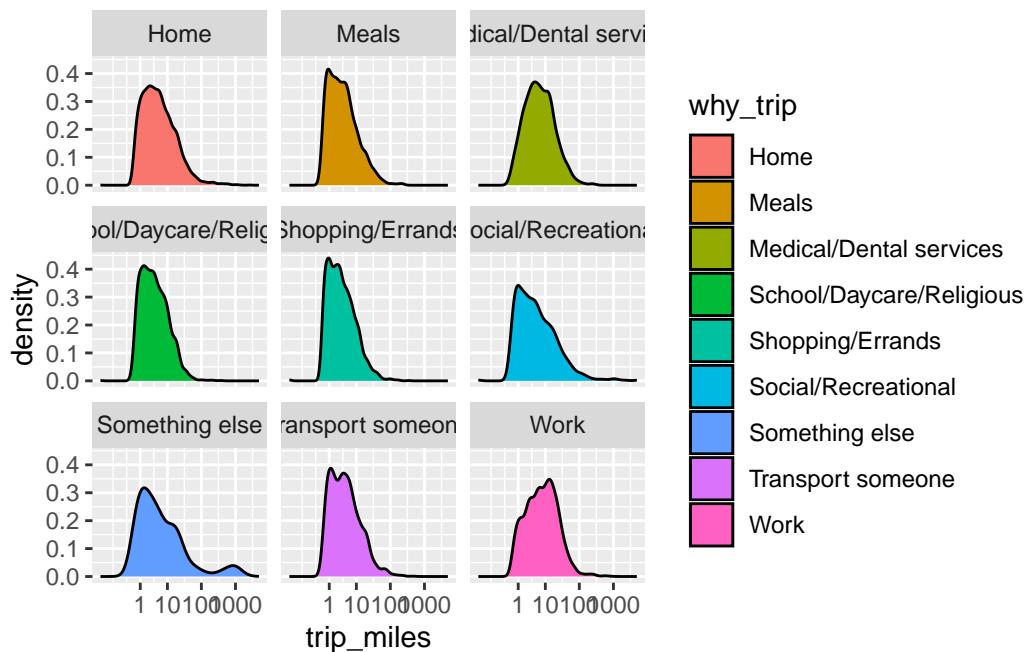
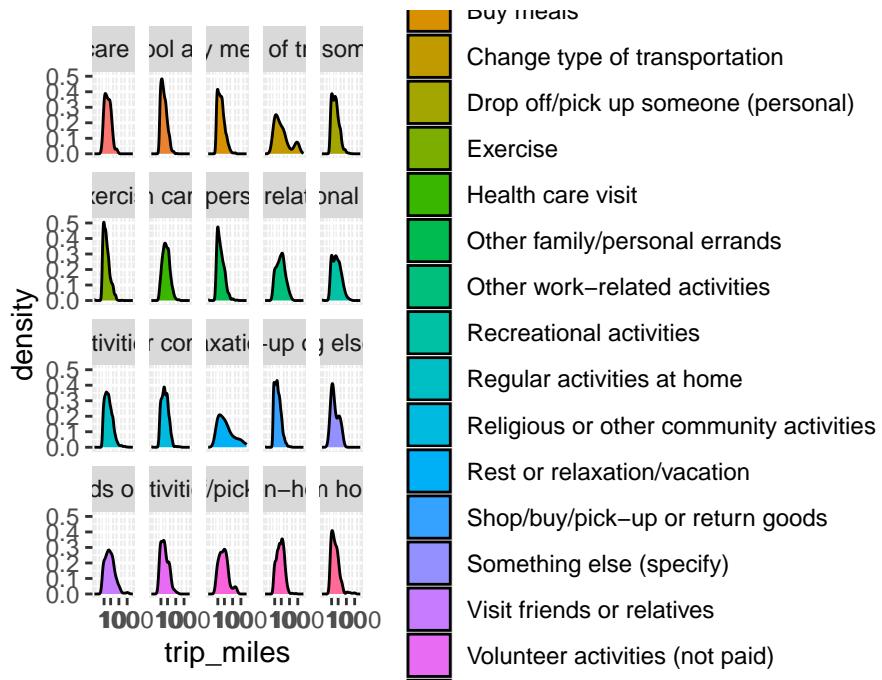


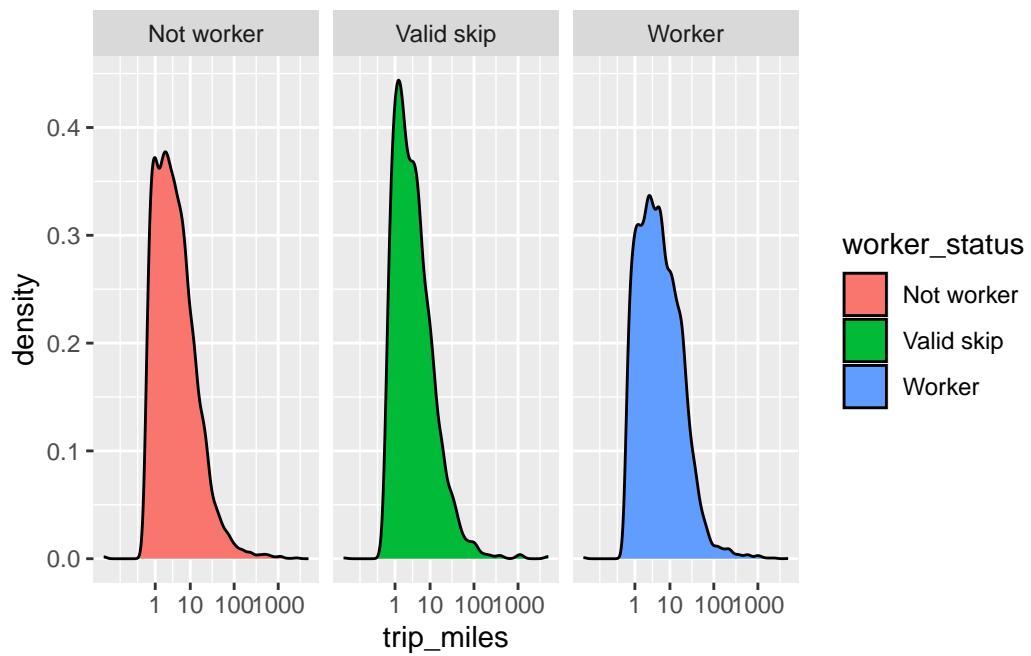
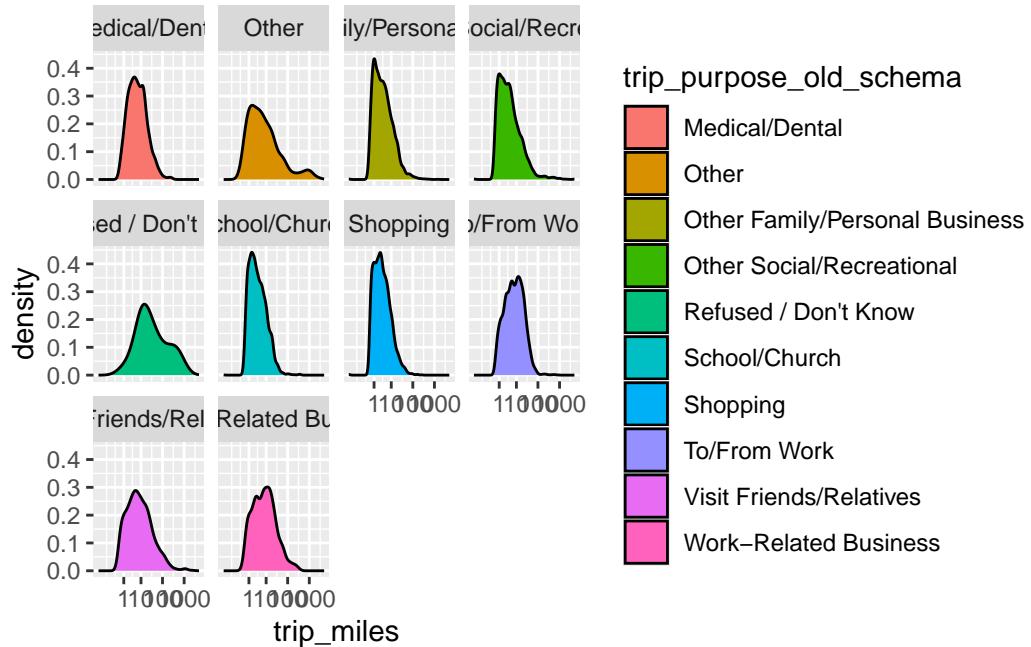












Vehicle Data

```
vehicle |> glimpse()
```

```
Rows: 14,684
Columns: 47
#> #> $ household_id <chr> "9000013002", "9000013002", "900~  
#> #> $ vehicle_id <chr> "01", "02", "01", "01", "01", "0~  
#> #> $ vehicle_year <dbl> 2015, 2019, 2011, 2015, 2014, 20~  
#> #> $ num_vehicles_in_household <dbl> 2, 2, 1, 1, 1, 2, 2, 1, 2, 2, 1, ~  
#> #> $ vehicle_case_id <chr> "900001300201", "900001300202", ~  
#> #> $ annual_miles <dbl> 2000, 10000, 10000, 5000, 5000, ~  
#> #> $ vehicle_age <dbl> 7, 3, 11, 7, 8, 5, 22, 8, 13, 5, ~  
#> #> $ num_adults_in_household <dbl> 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 2, ~  
#> #> $ num_drivers_in_household <dbl> 2, 2, 2, 2, 1, 2, 2, 1, 2, 2, 1, ~  
#> #> $ household_size <dbl> 4, 4, 2, 4, 1, 3, 3, 1, 2, 2, 2, ~  
#> #> $ travel_day_date <date> 2022-02-01, 2022-02-01, 2022-02~  
#> #> $ WRKCOUNT <dbl> 1, 1, 2, 1, 1, 2, 2, 0, 2, 2, 2, ~  
#> #> $ household_stratum_id <chr> "1021", "1021", "1021", "10~  
#> #> $ household_weight_7_day <dbl> 4621.758, 4621.758, 2982.998, 10~  
#> #> $ household_weight_5_day <dbl> 0.000, 0.000, 3878.814, 13150.67~  
#> #> $ 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", "Urb~  

$ 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	Min. :2022-01-01	Min. :0.000
1st Qu.:	2.000	1st Qu.: 2.000	1st Qu.:2022-04-01	1st Qu.:0.000
Median :	2.000	Median : 2.000	Median :2022-08-01	Median :1.000
Mean :	2.014	Mean : 2.558	Mean :2022-07-14	Mean : 1.179
3rd Qu.:	2.000	3rd Qu.: 3.000	3rd Qu.:2022-11-01	3rd Qu.:2.000
Max. :	7.000	Max. :10.000	Max. :2023-01-01	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            Length:14684
Class :character         Class :character
Mode  :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
```

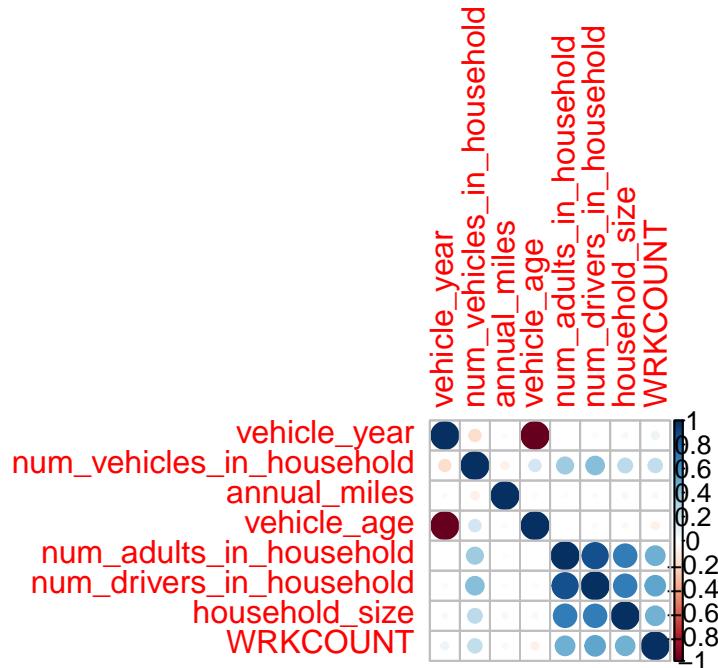
```
vehicle_owned_more_1_year vehicle_owned_months vehicle_type
Length:14684          Length:14684          Length:14684
Class :character      Class :character      Class :character
Mode  :character      Mode  :character      Mode  :character
```

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

```

vehicle |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot()

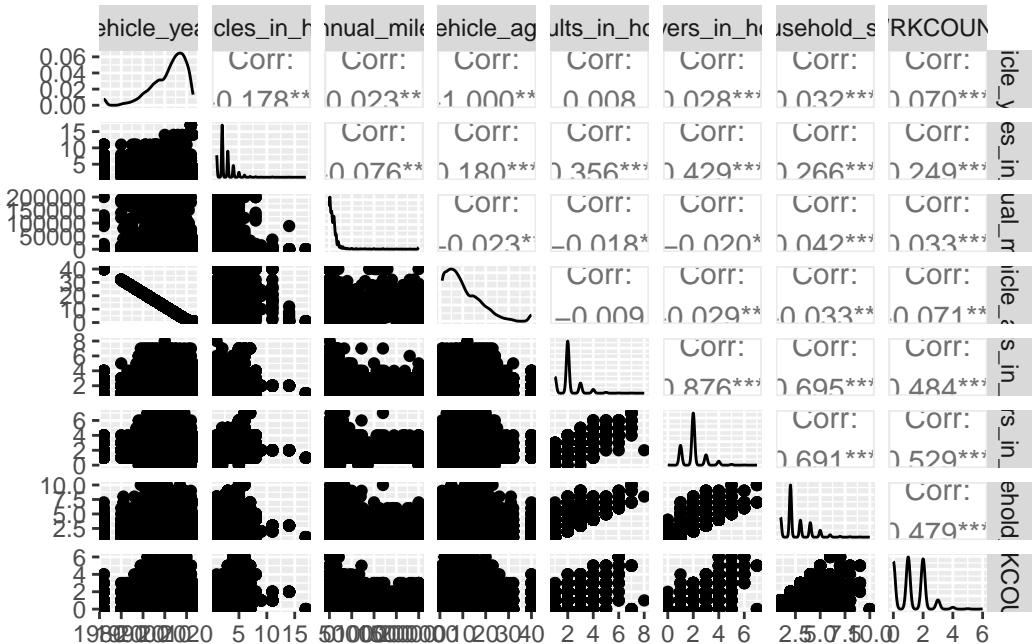
```



```

# Pair plot of vehicle data
vehicle |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  select(where(is.numeric)) |>
  ggpairs()

```



Long Distance Data

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

```
Rows: 16,997
Columns: 66
$ household_id
$ person_id
$ LONGDIST
$ num_people_on_long_distance_trip
$ num_used_amtrak_last_year
$ num_used_intracity_bus_last_year
$ num_nights_away_from_home
$ great_circle_distance_traveled
$ num_adults_in_household
$ num_drivers_in_household
$ household_size
$ num_vehicles_in_household
$ travel_day_date
$ WRKCOUNT
$ household_stratum_id

<chr> "9000013002", "9000013002", ~
<chr> "01", "02", "03", "04", "01~
<dbl> 1, 1, 0, 0, 0, 0, 1, 0, 0, ~
<dbl> 2, 2, 5, 5, 2, 2, 1, 4, 4, ~
<dbl> 0, 0, 0, 0, 0, 0, 2, 0, 0, ~
<dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
<dbl> 3, 3, -9, -9, -9, -9, 8, -9~
<dbl> 61.86880, 61.86880, 382.211~
<dbl> 2, 2, 2, 2, 2, 2, 1, 2, 2, ~
<dbl> 2, 2, 2, 2, 2, 2, 1, 2, 2, ~
<dbl> 4, 4, 4, 4, 2, 2, 1, 4, 4, ~
<dbl> 2, 2, 2, 2, 1, 1, 0, 1, 1, ~
<date> 2022-02-01, 2022-02-01, 20~
<dbl> 1, 1, 1, 1, 2, 2, 1, 1, 1, ~
<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
Min. : 2022-01-01	Min. : 0.000	Length:16997	Min. : 30.9
1st Qu.: 2022-04-01	1st Qu.: 0.000	Class :character	1st Qu.: 5614.8
Median : 2022-08-01	Median : 1.000	Mode :character	Median : 10798.7
Mean : 2022-07-16	Mean : 1.216		Mean : 17977.3
3rd Qu.: 2022-11-01	3rd Qu.: 2.000		3rd Qu.: 21084.0
Max. : 2023-01-01	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
```

```

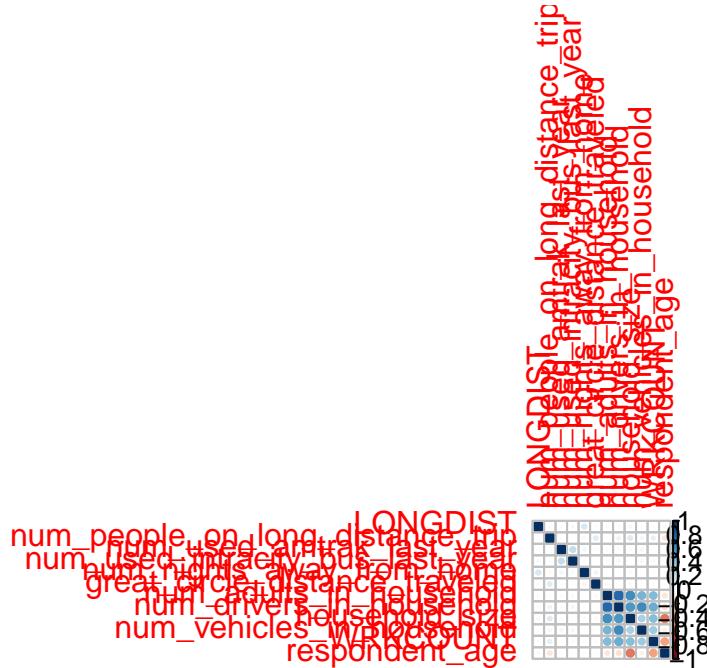
worker_status
Length:16997
Class :character
Mode  :character

```

```

long_distance |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag")) |>
  select(where(is.numeric)) |>
  cor() |>
  corrplot::corrplot()

```



```

# Plot great_circle_distance_traveled by all variables
long_distance |>
  select(-contains("id")) |>
  select(-contains("weight")) |>
  select(-contains("flag"))

```

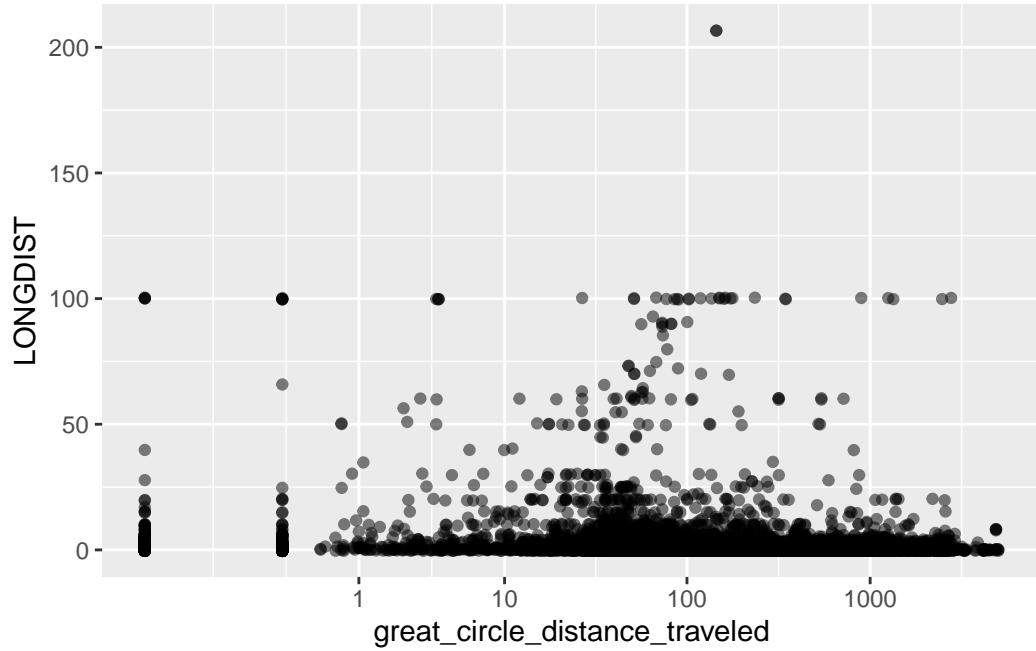
```

select(-contains("on_long_distance_trip")) |>
colnames() -> long_distance_cols

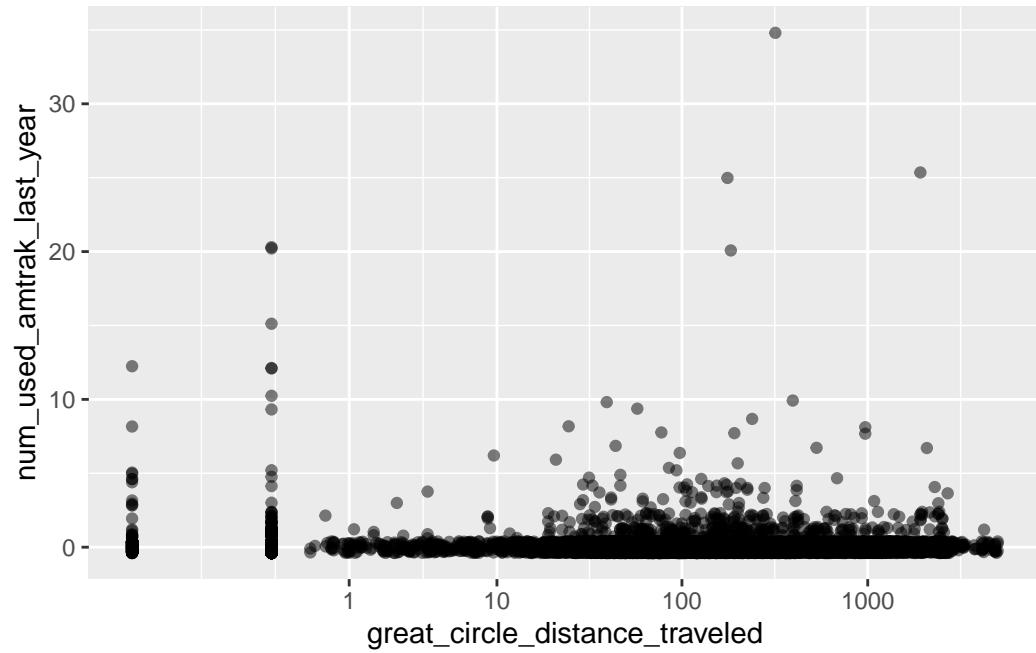
for (col in long_distance_cols) {
  print(col)
  if (is.character(long_distance[[col]])) {
    p <- long_distance |>
      ggplot(aes(x = great_circle_distance_traveled, fill = .data[[col]])) +
      geom_density() +
      facet_wrap(~.data[[col]]) +
      scale_x_continuous(
        trans = "pseudo_log",
        breaks = c(1, 10, 100, 1000, 10000),
        )
    print(p)
  } else if (col == "great_circle_distance_traveled") {
    next
  } else if (inherits(long_distance[[col]], "Date")) {
    p <- long_distance |>
      ggplot(aes(x = .data[[col]], y = great_circle_distance_traveled)) +
      geom_jitter(alpha = 0.5) +
      scale_x_date() +
      scale_y_continuous(
        trans = "pseudo_log",
        breaks = c(1, 10, 100, 1000, 10000),
        )
    print(p)
  } else {
    p <- long_distance |>
      ggplot(aes(x = great_circle_distance_traveled, y = .data[[col]])) +
      geom_jitter(alpha = 0.5) +
      scale_x_continuous(
        trans = "pseudo_log",
        breaks = c(1, 10, 100, 1000, 10000)
        )
    print(p)
  }
}

```

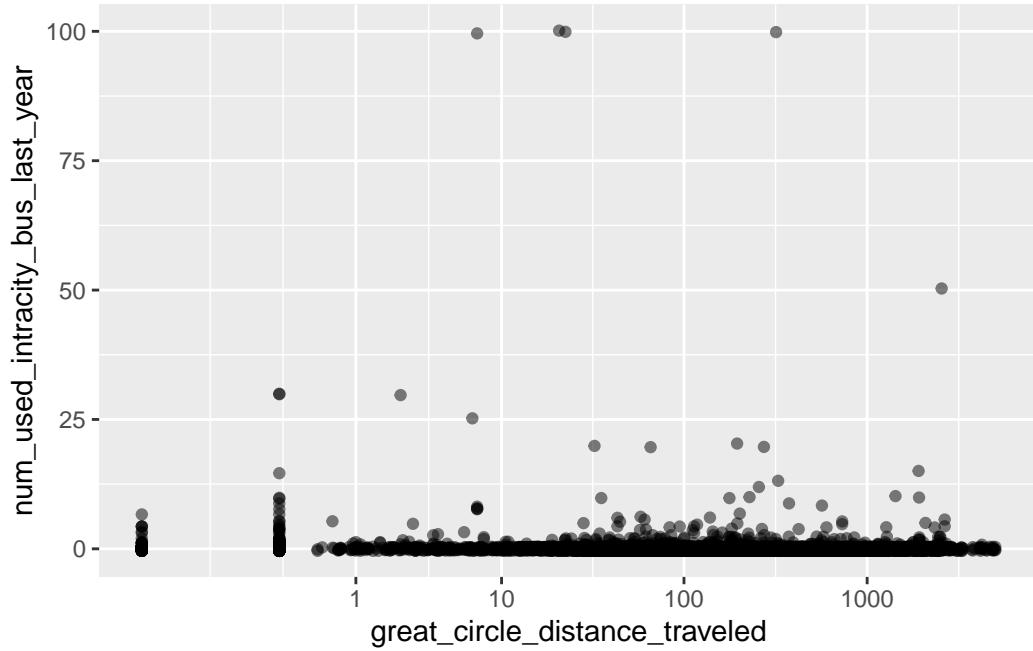
[1] "LONGDIST"



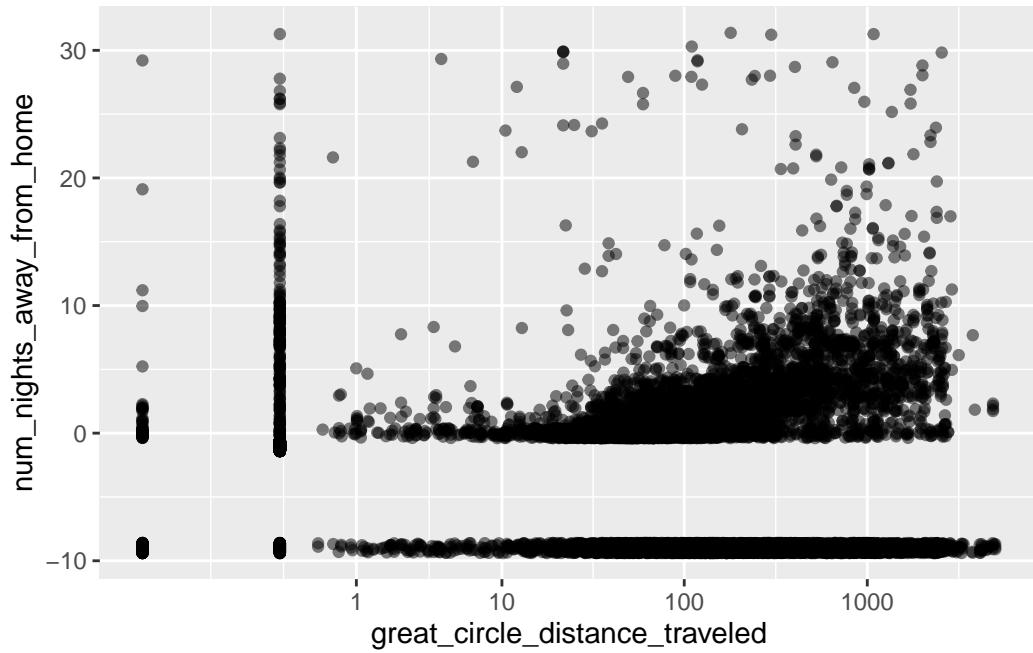
```
[1] "num_used_amtrak_last_year"
```



```
[1] "num_used_intracity_bus_last_year"
```

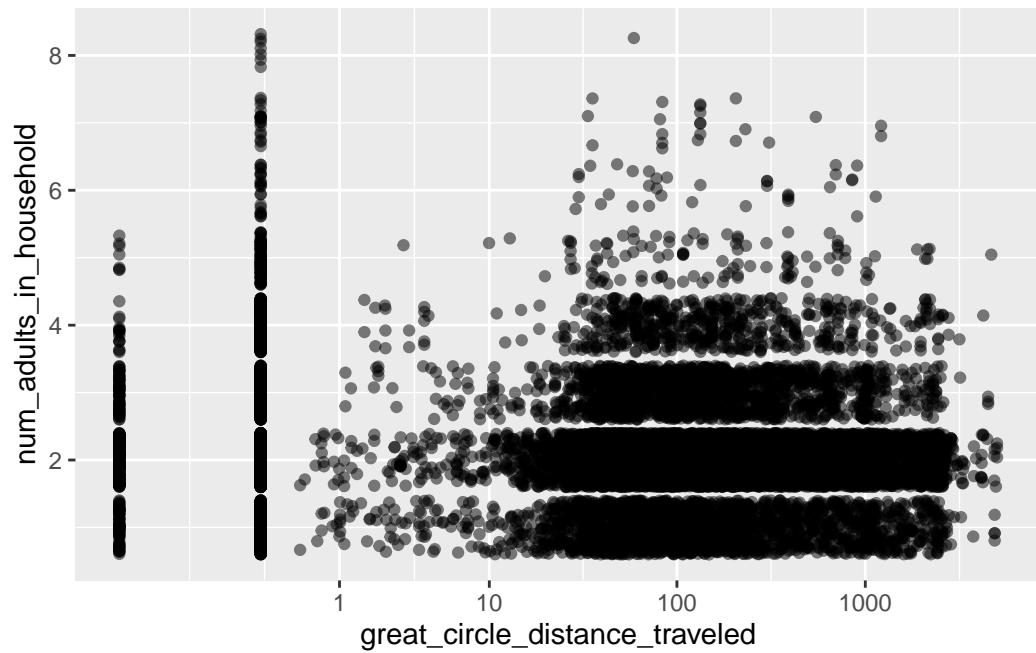


```
[1] "num_nights_away_from_home"
```

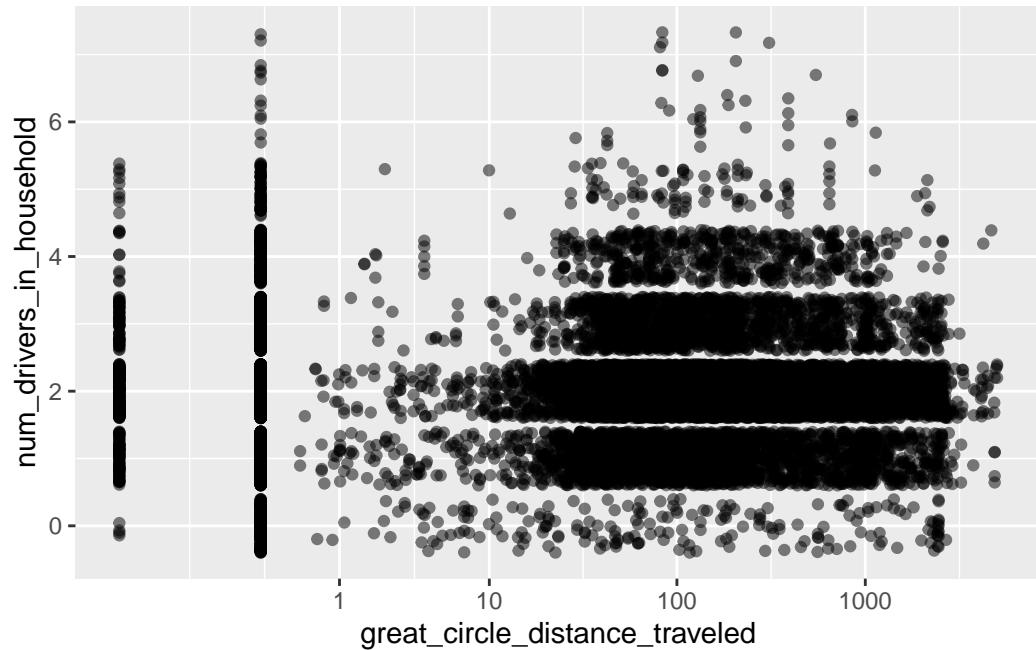


```
[1] "great_circle_distance_traveled"
```

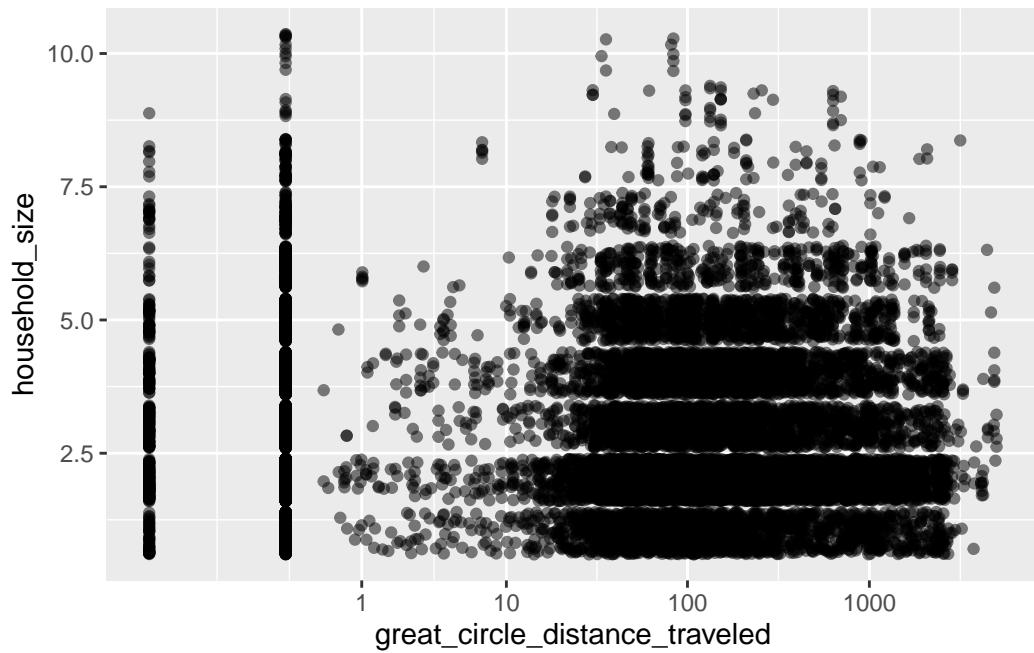
```
[1] "num_adults_in_household"
```



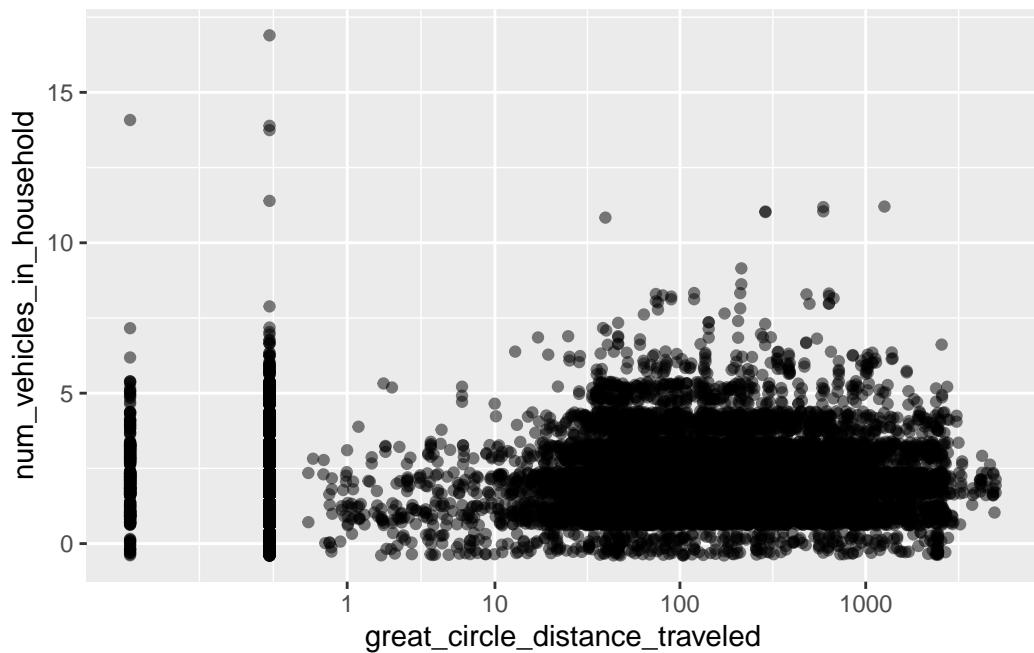
```
[1] "num_drivers_in_household"
```



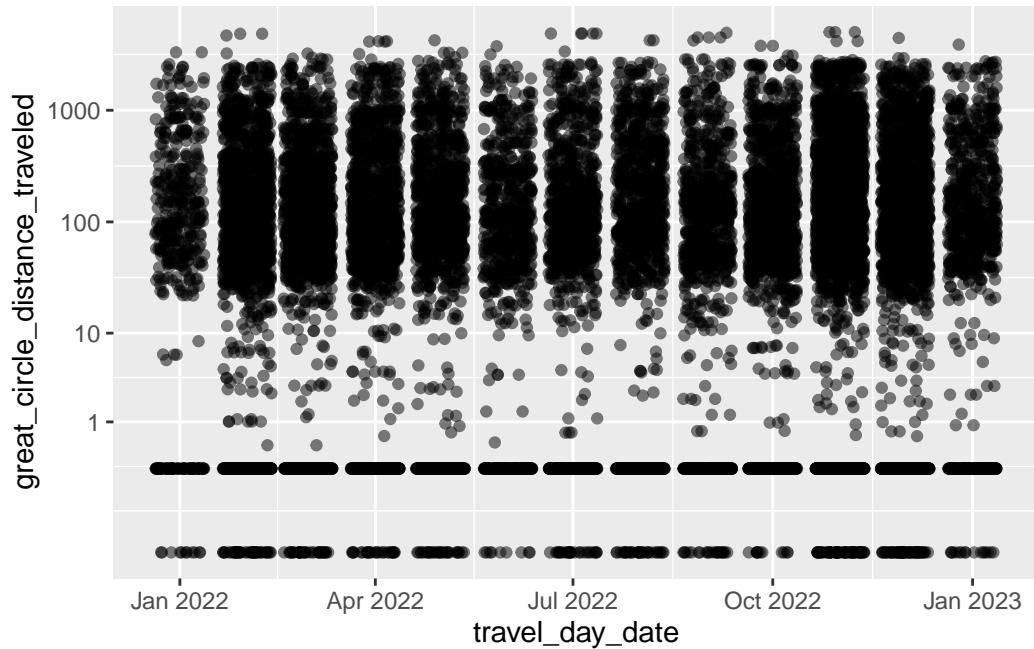
```
[1] "household_size"
```



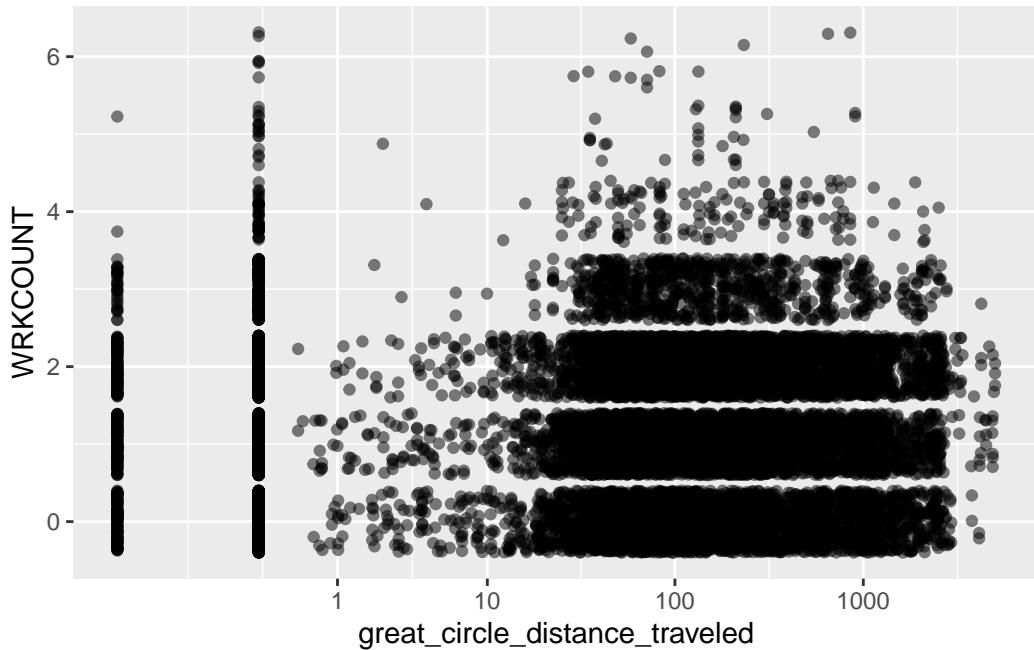
```
[1] "num_vehicles_in_household"
```



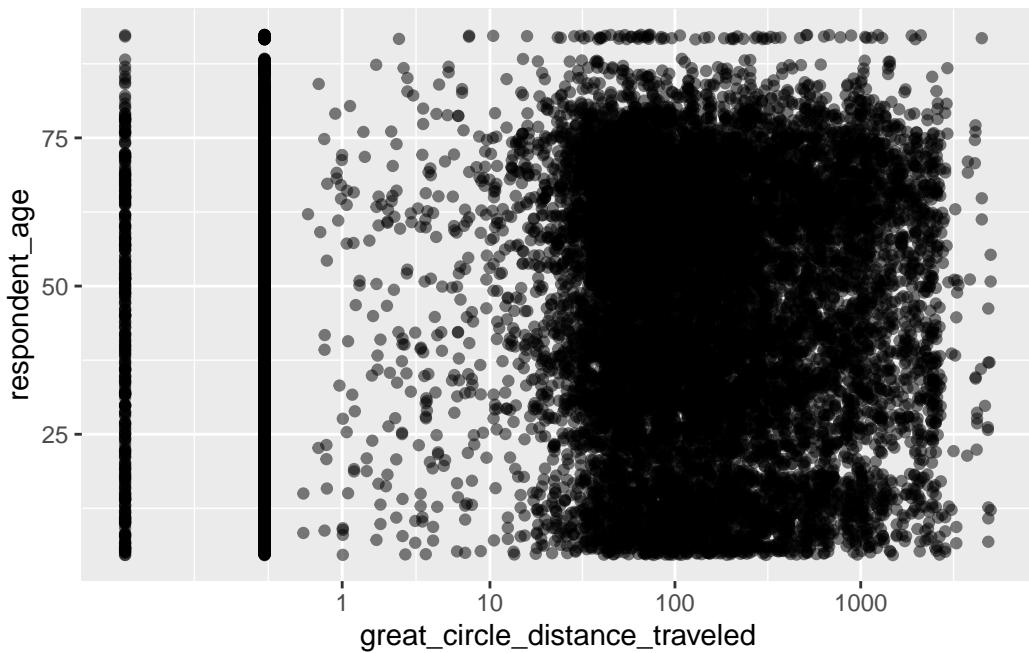
```
[1] "travel_day_date"
```



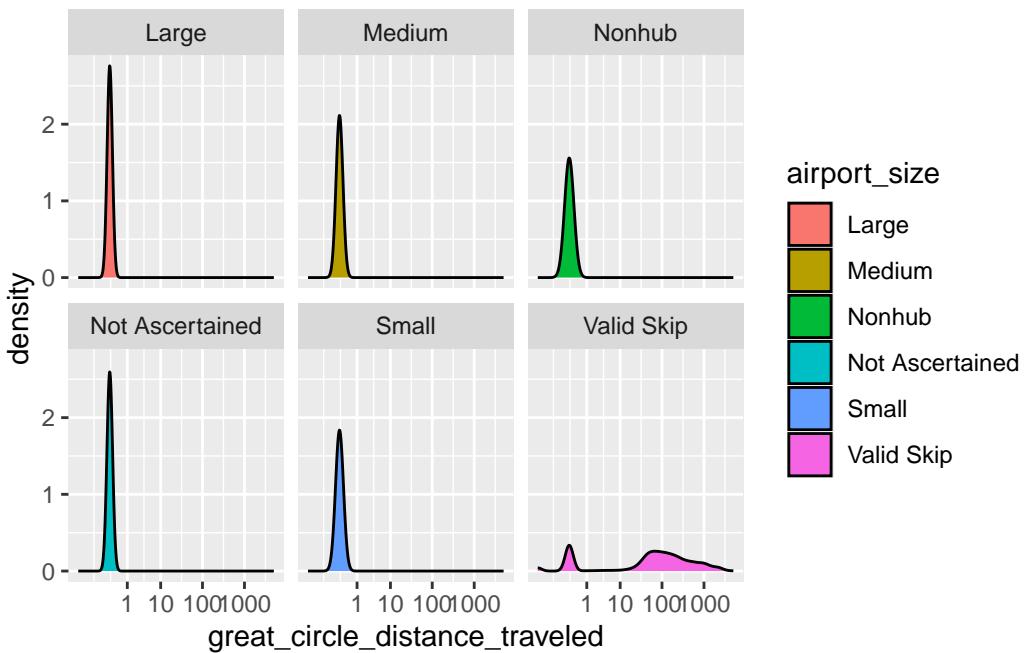
```
[1] "WRKCOUNT"
```



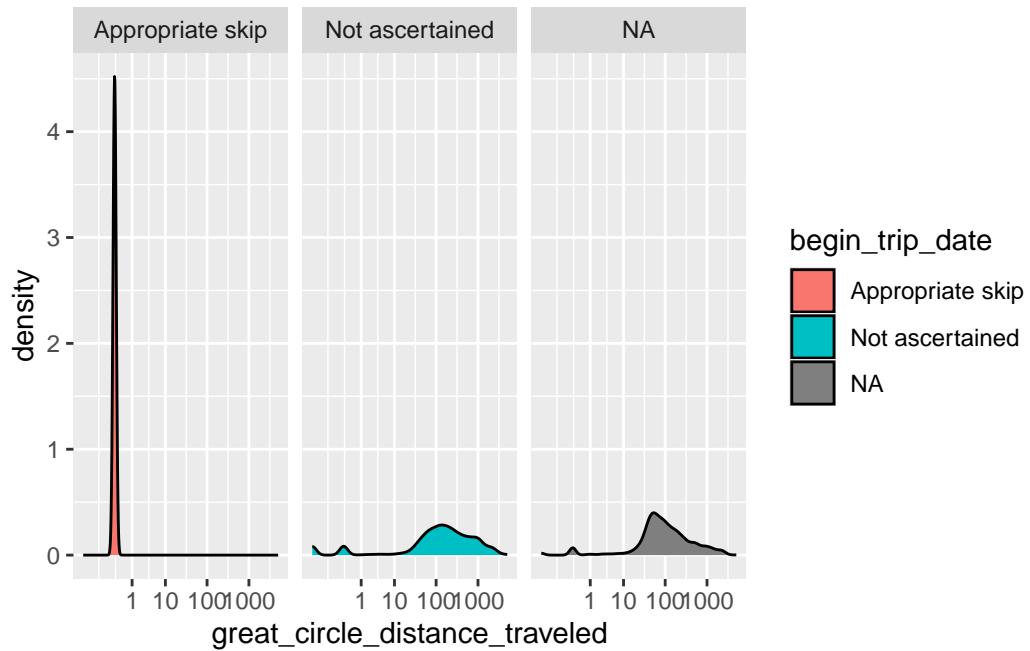
```
[1] "respondent_age"
```



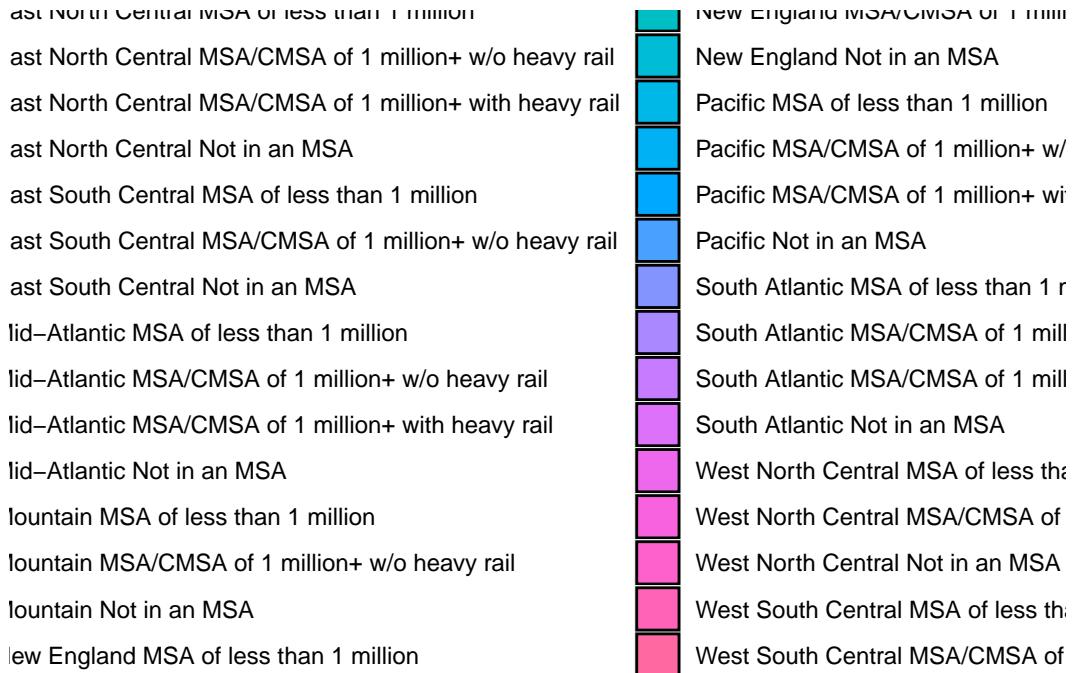
```
[1] "airport_size"
```



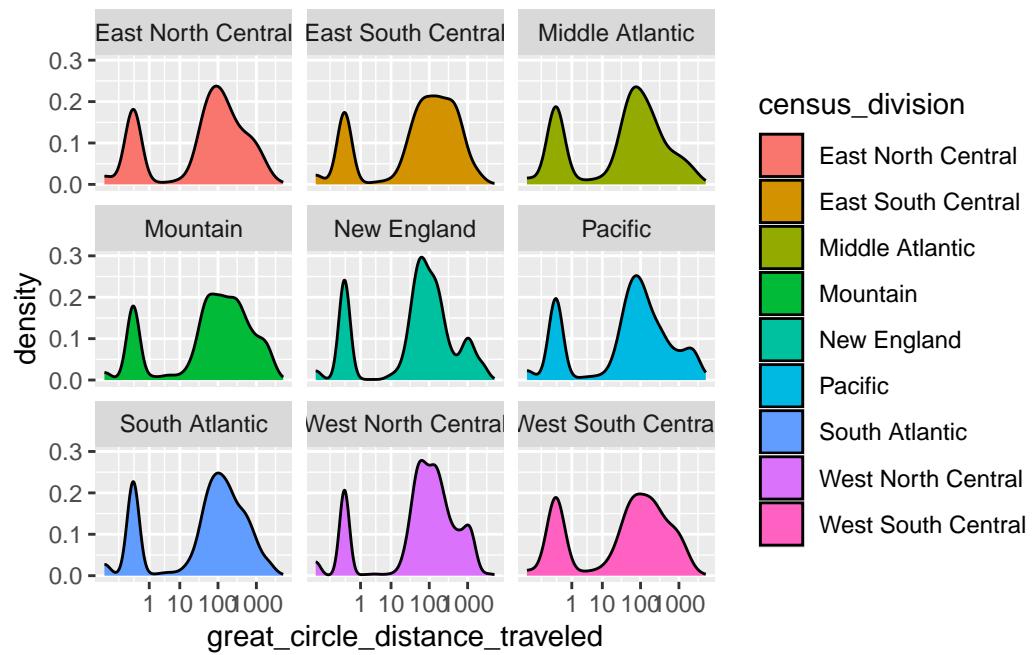
```
[1] "begin_trip_date"
```



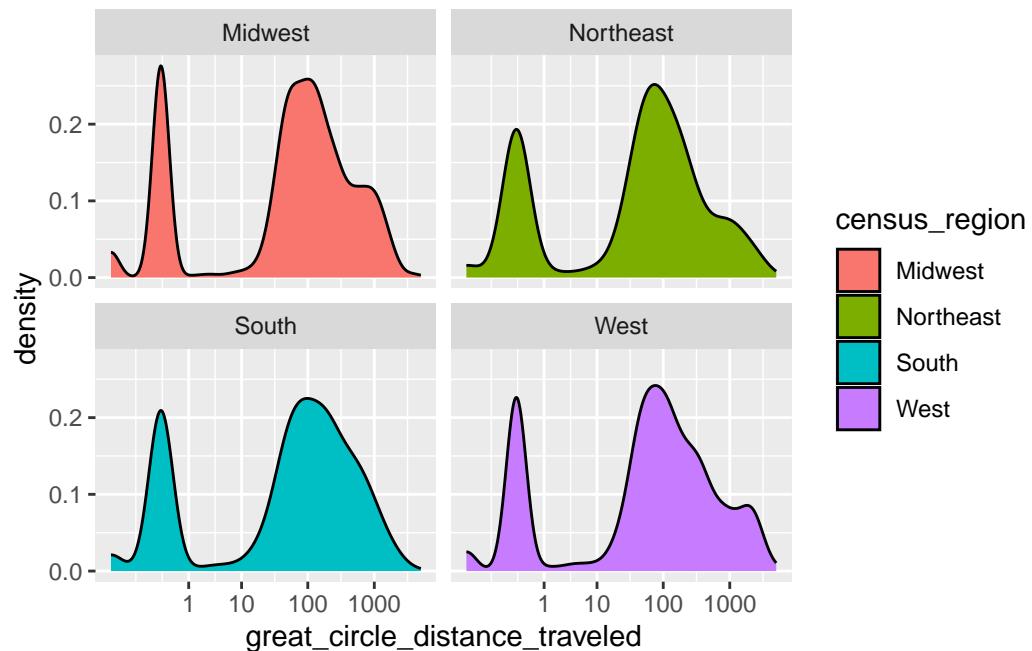
```
[1] "household_grouping"
```



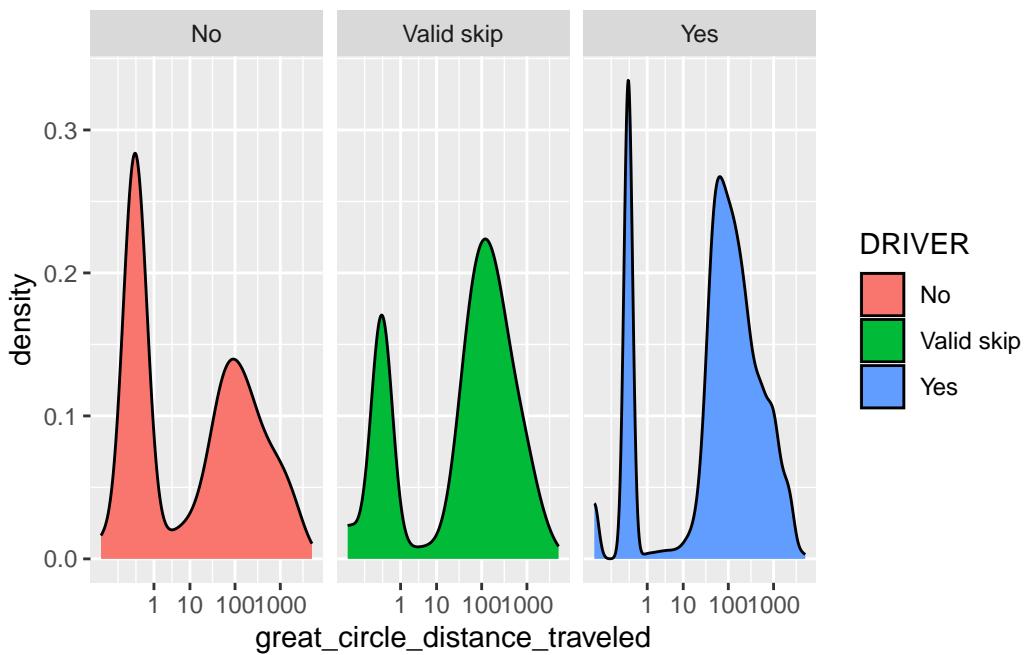
```
[1] "census_division"
```



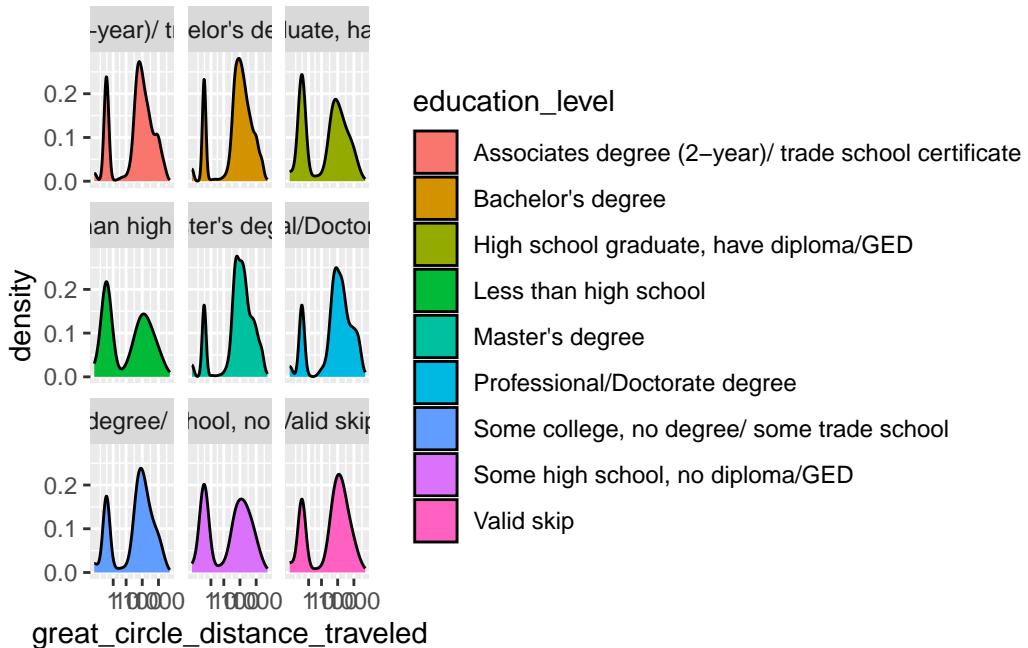
```
[1] "census_region"
```



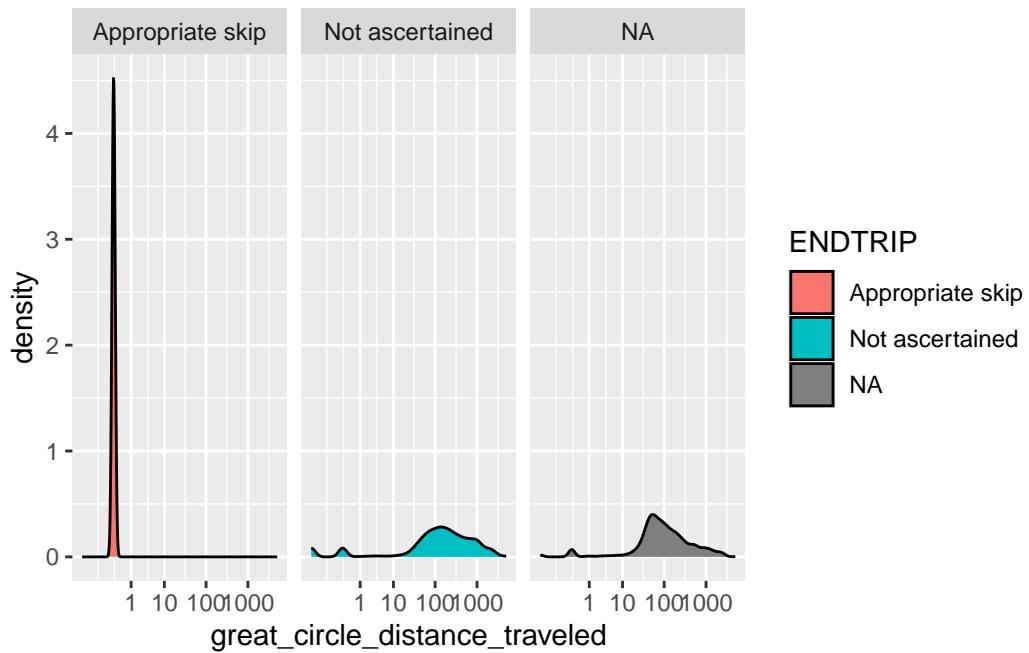
[1] "DRIVER"



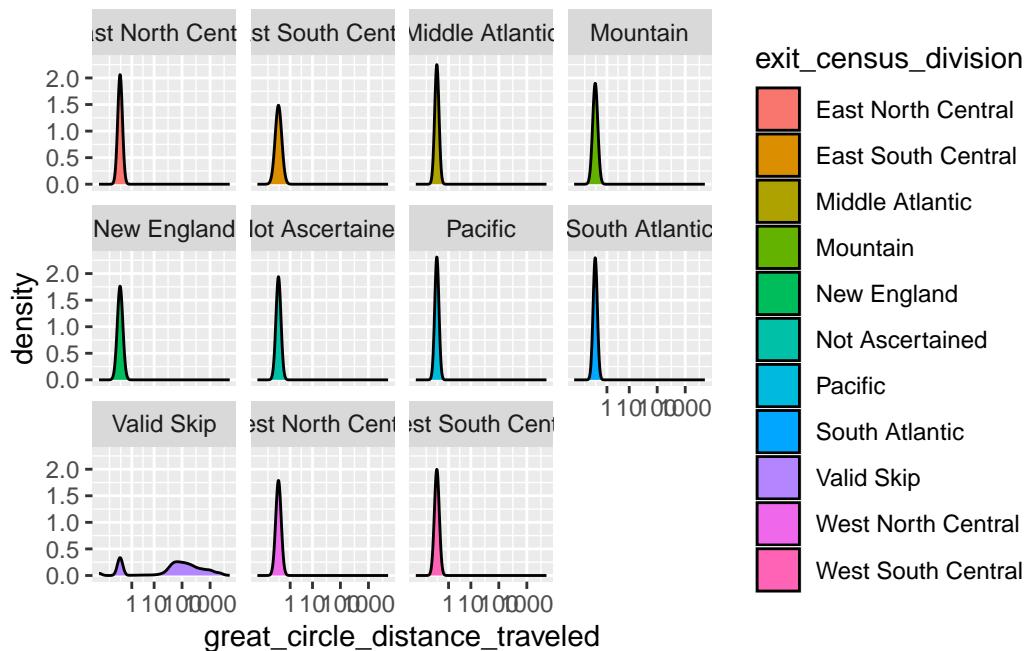
[1] "education_level"



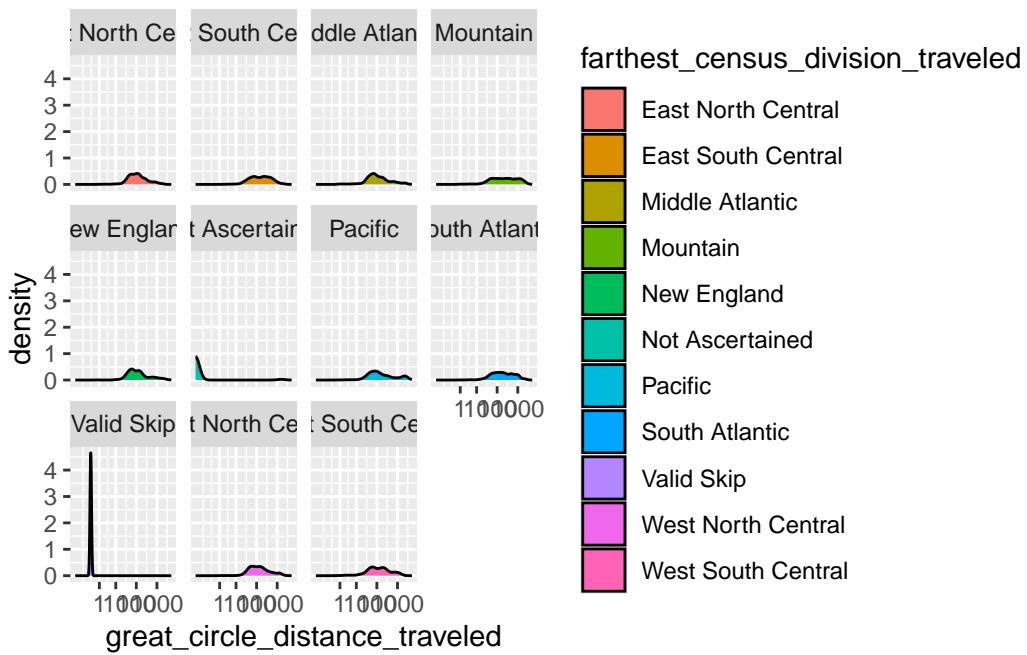
```
[1] "ENDTRIP"
```



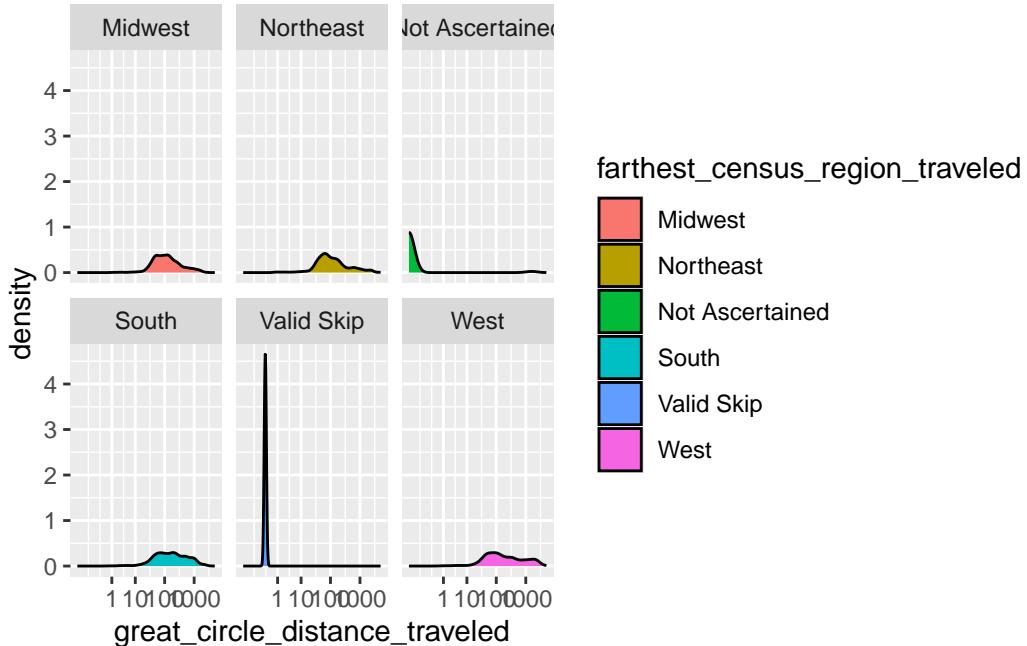
```
[1] "exit_census_division"
```



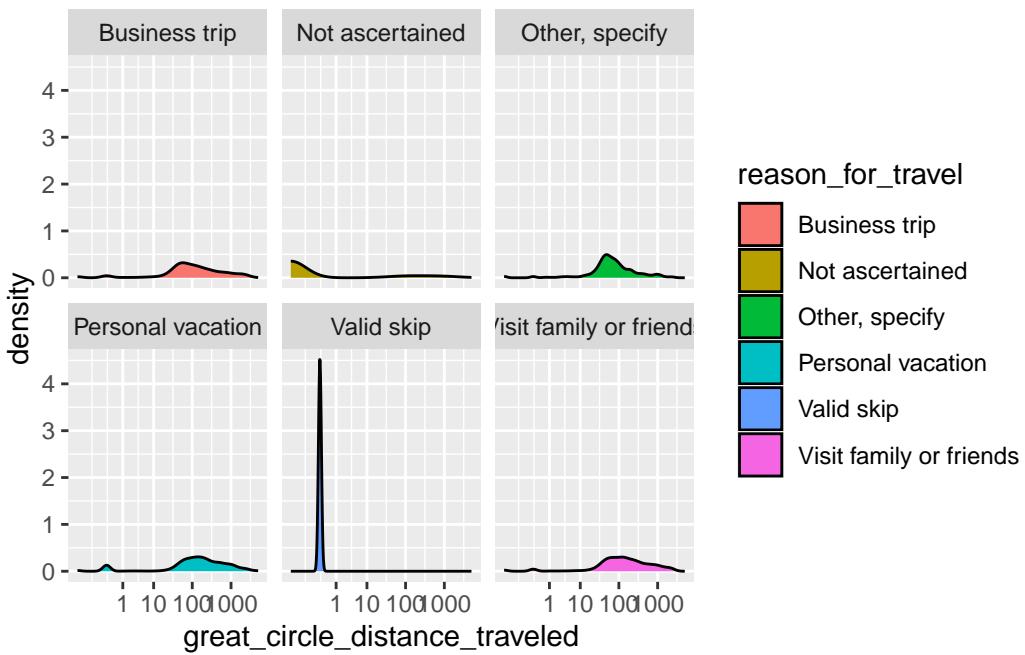
```
[1] "farthest_census_division_traveled"
```



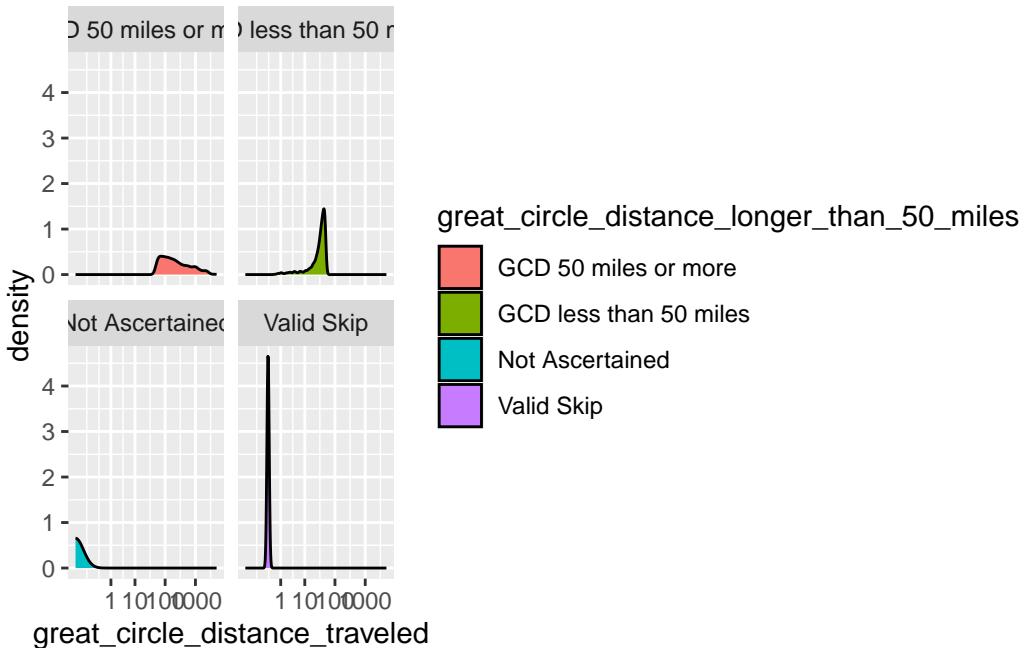
```
[1] "farthest_census_region_traveled"
```



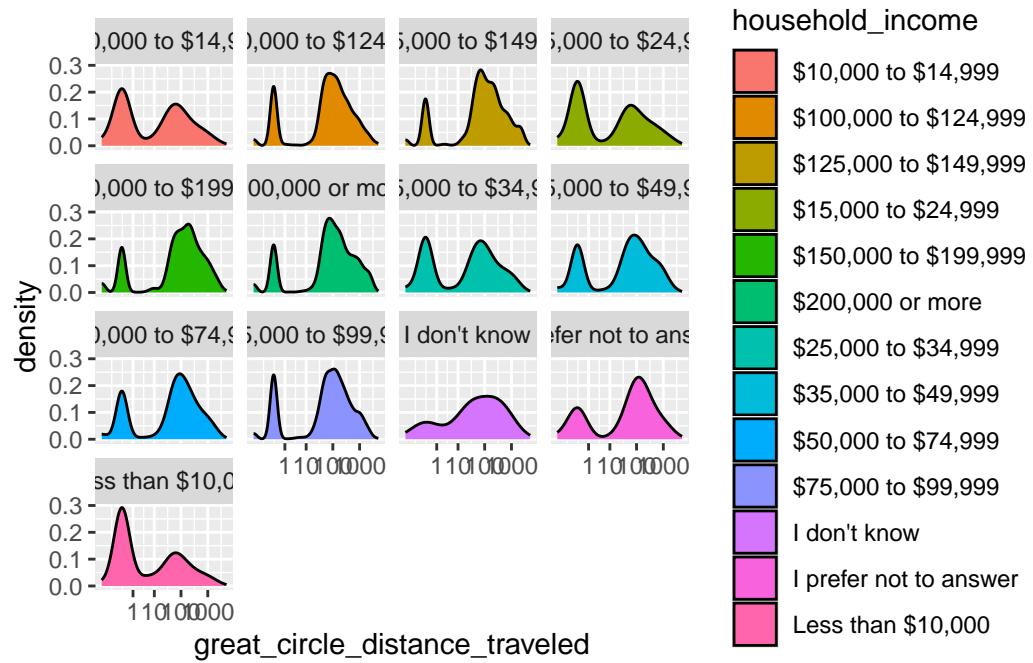
```
[1] "reason_for_travel"
```



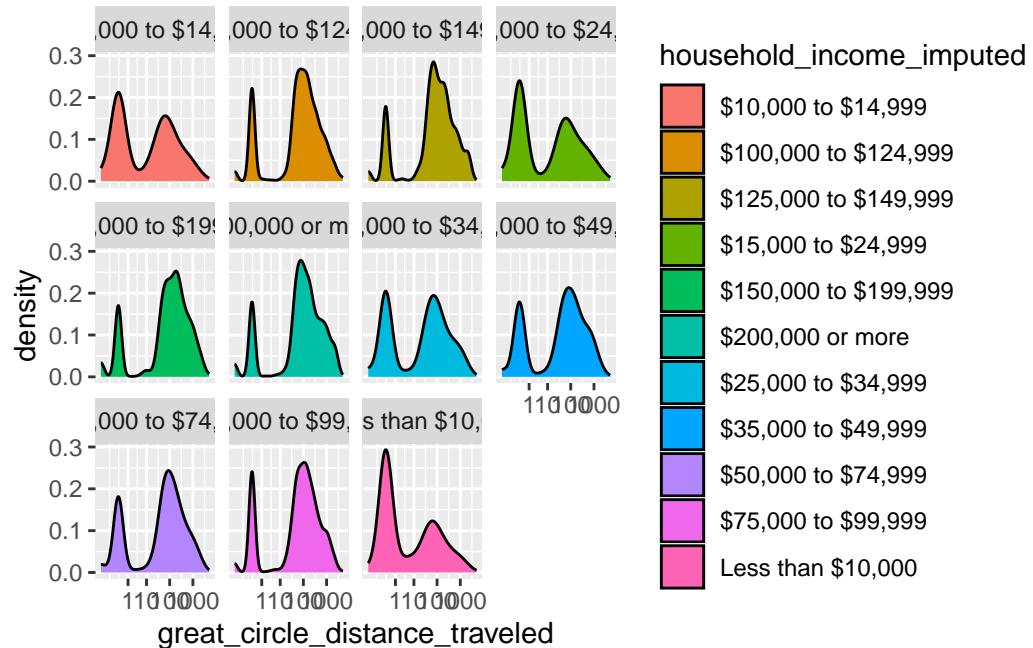
```
[1] "great_circle_distance_longer_than_50_miles"
```



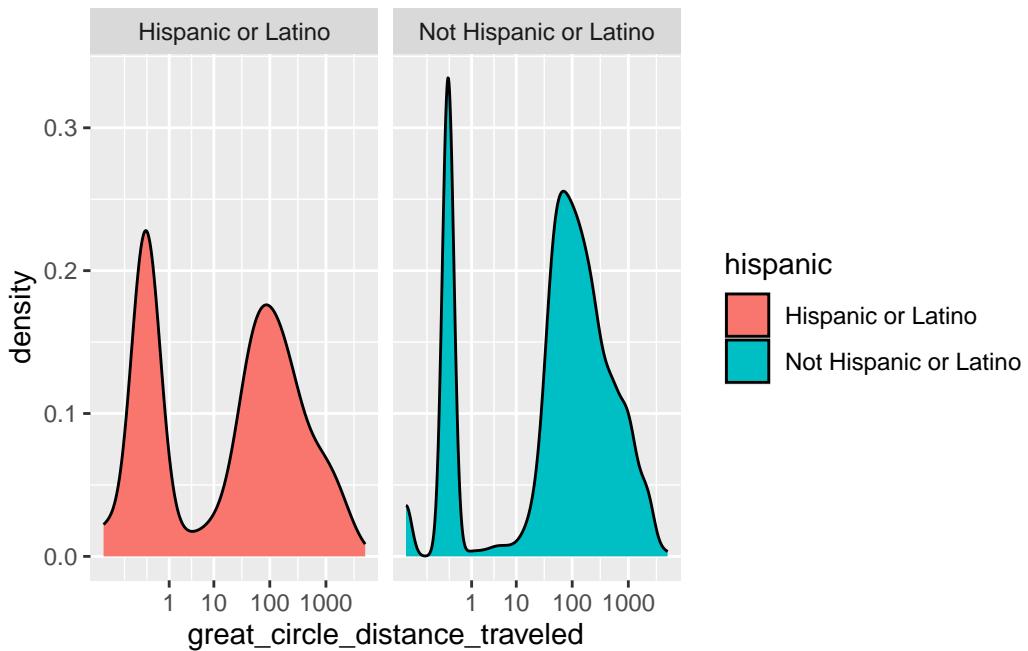
[1] "household_income"



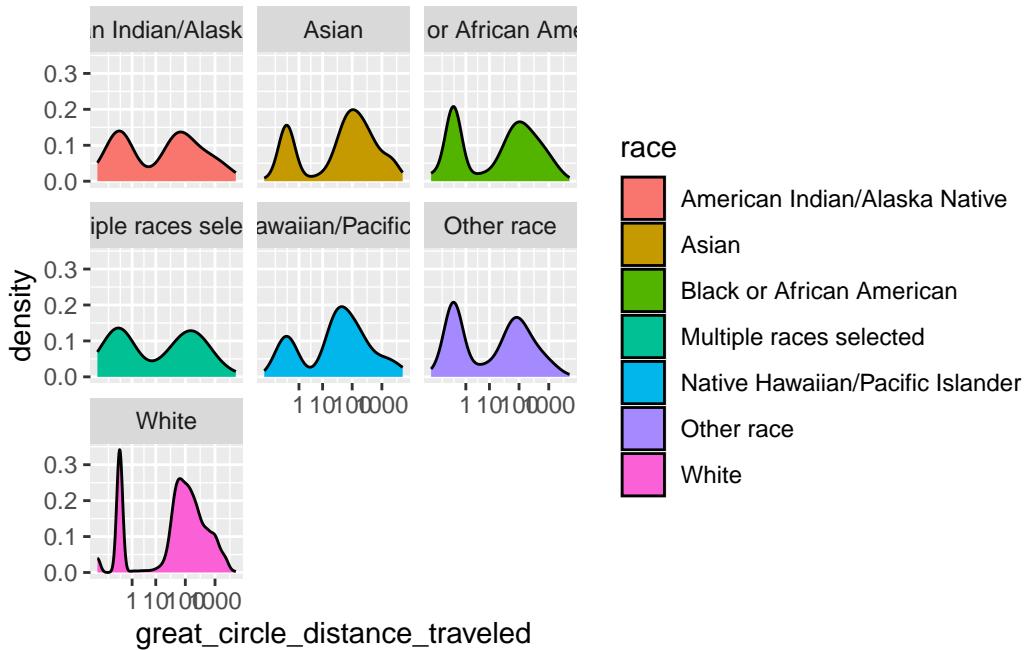
[1] "household_income_imputed"



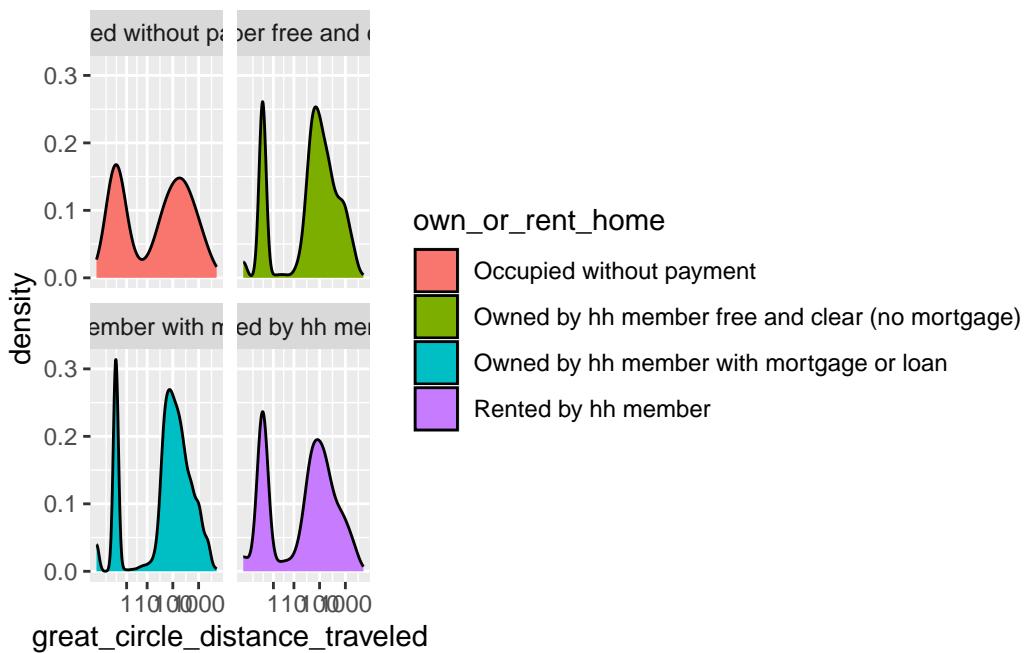
```
[1] "hispanic"
```



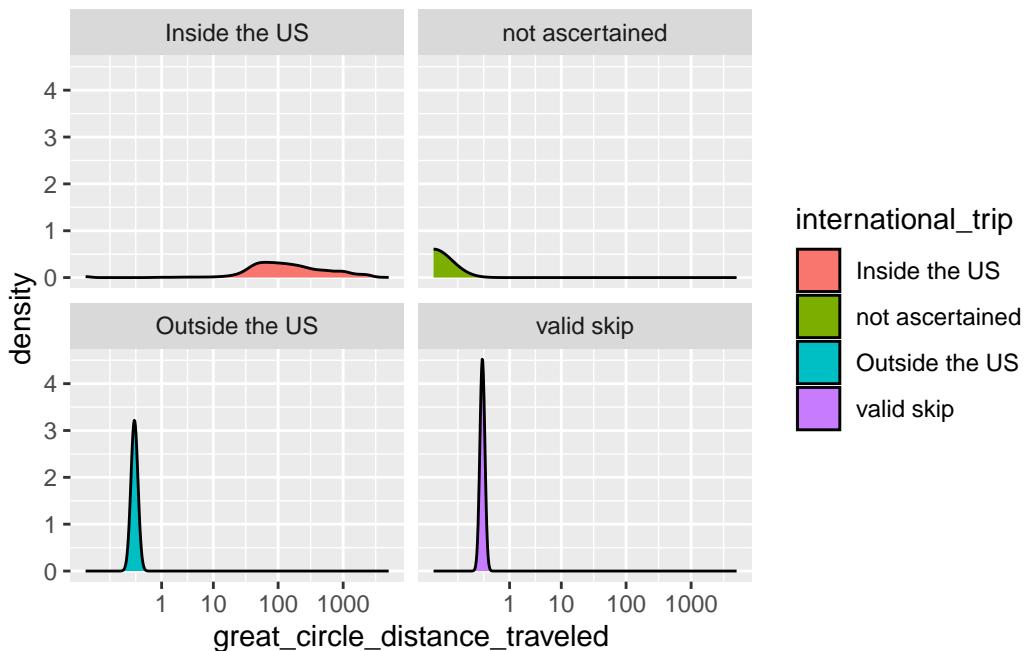
```
[1] "race"
```



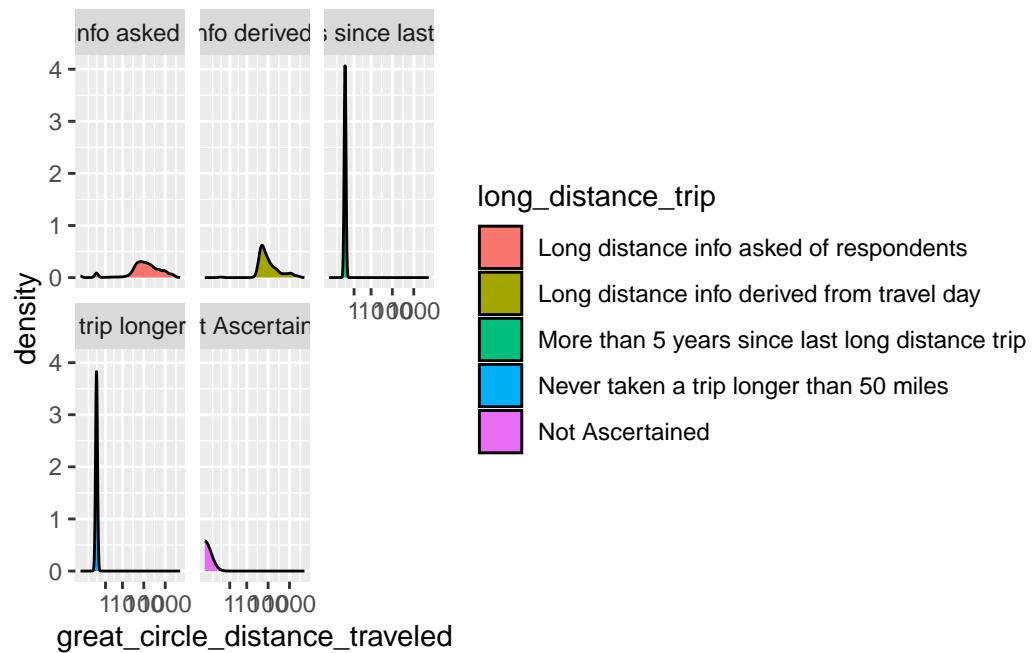
```
[1] "own_or_rent_home"
```



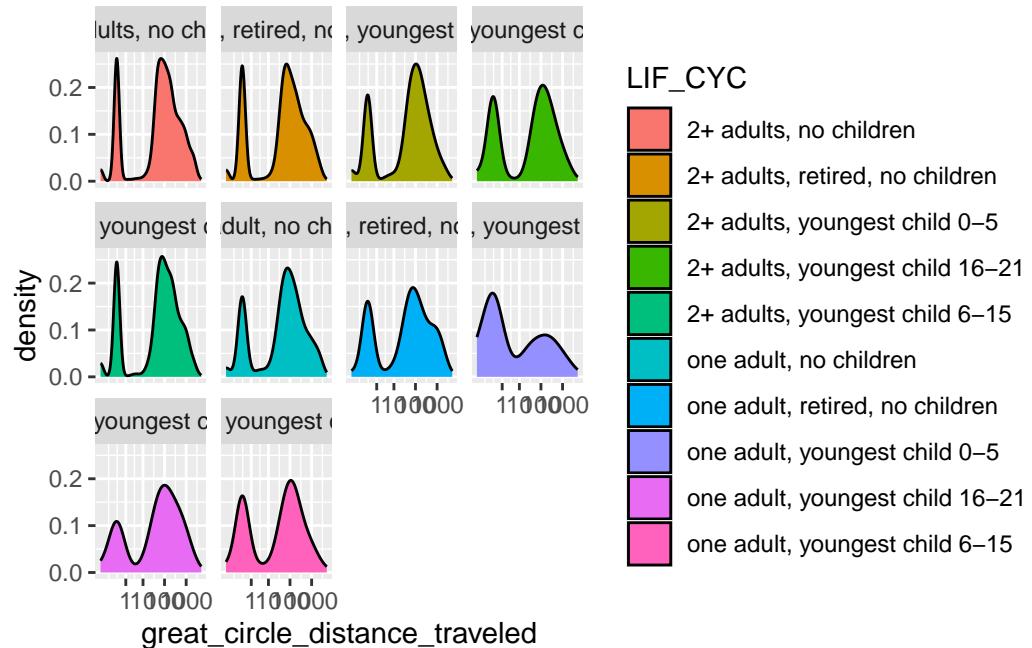
```
[1] "international_trip"
```



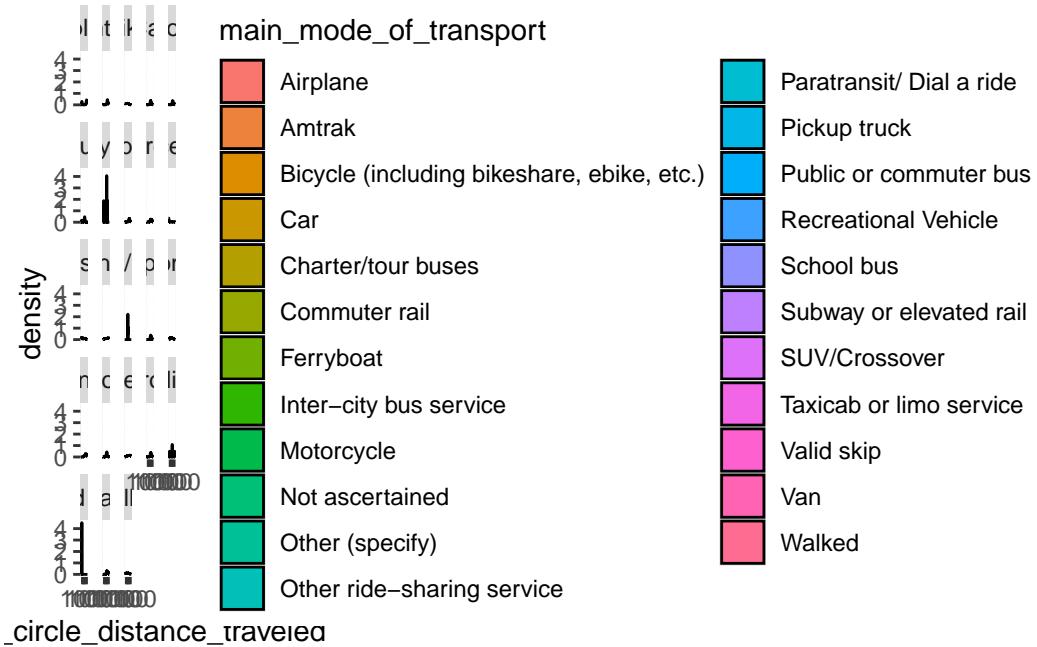
```
[1] "long_distance_trip"
```



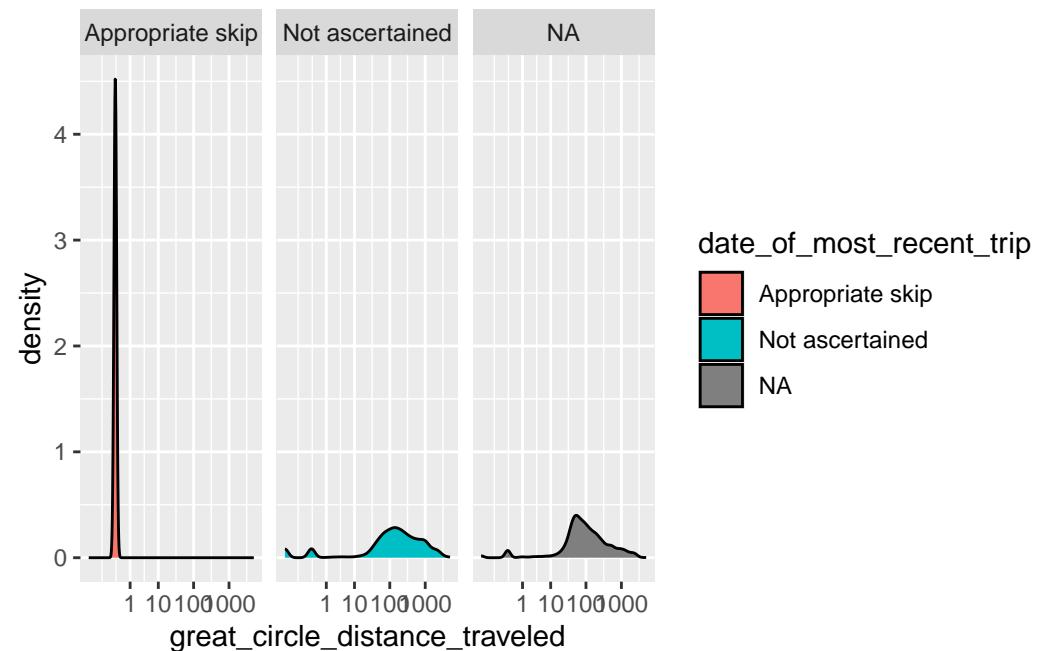
```
[1] "LIF_CYC"
```



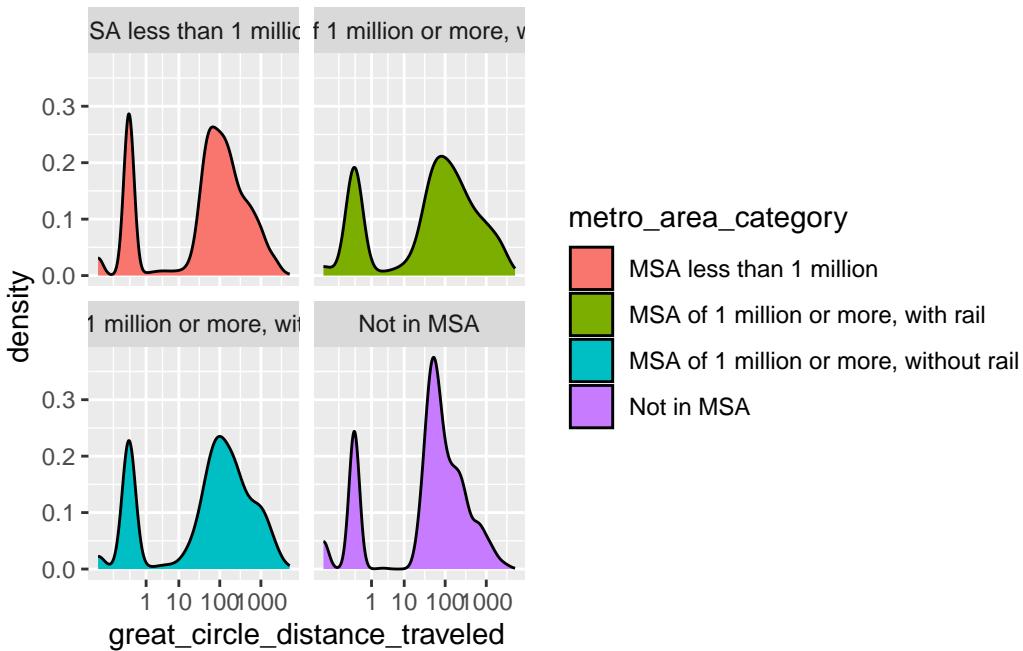
```
[1] "main_mode_of_transport"
```



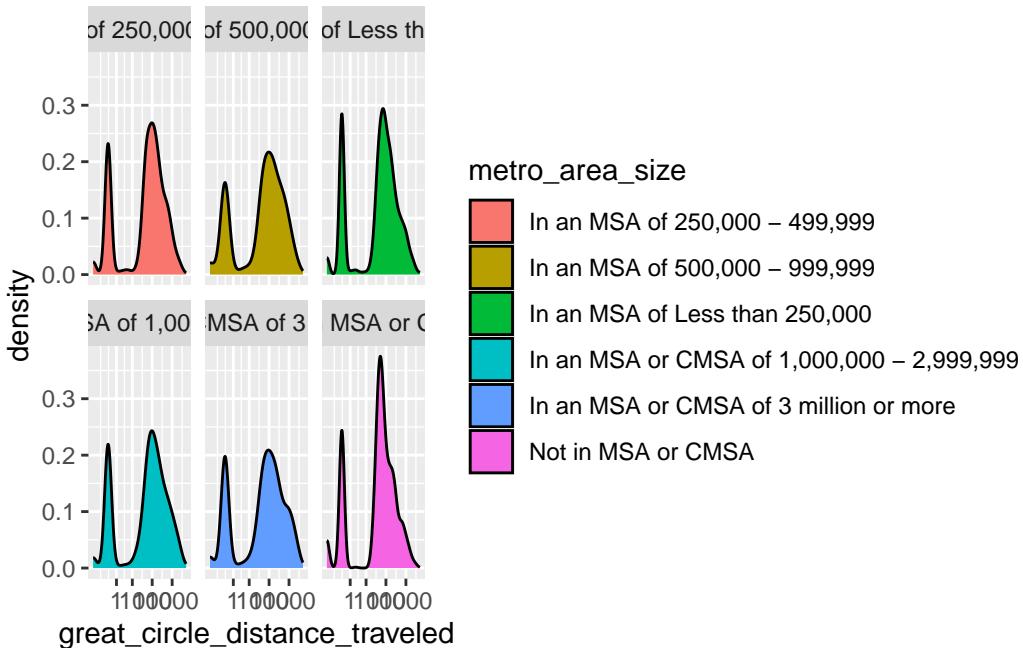
```
[1] "date_of_most_recent_trip"
```



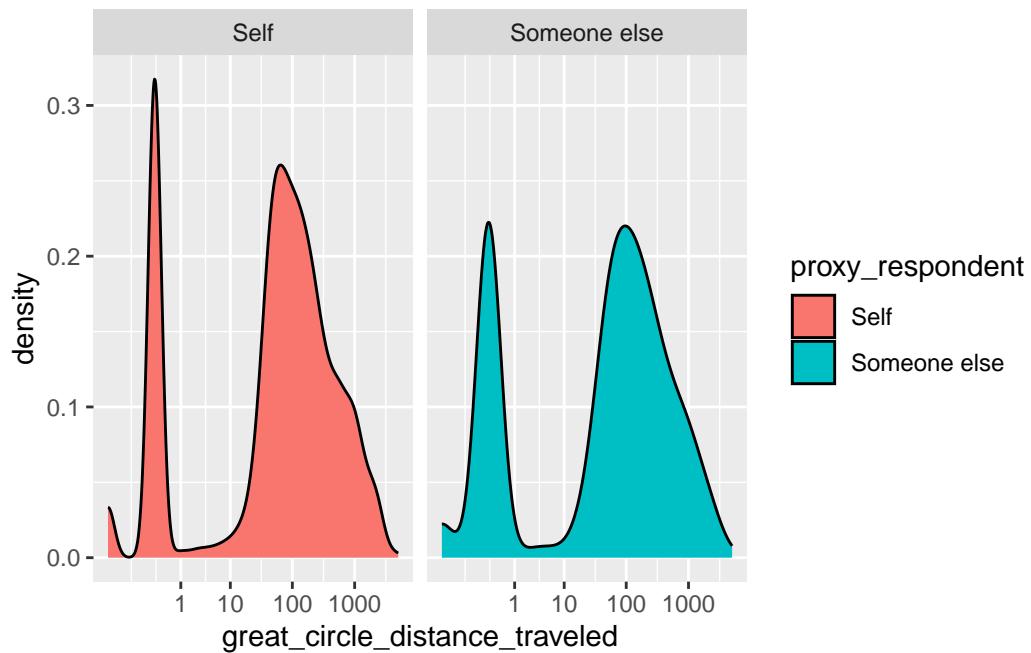
```
[1] "metro_area_category"
```



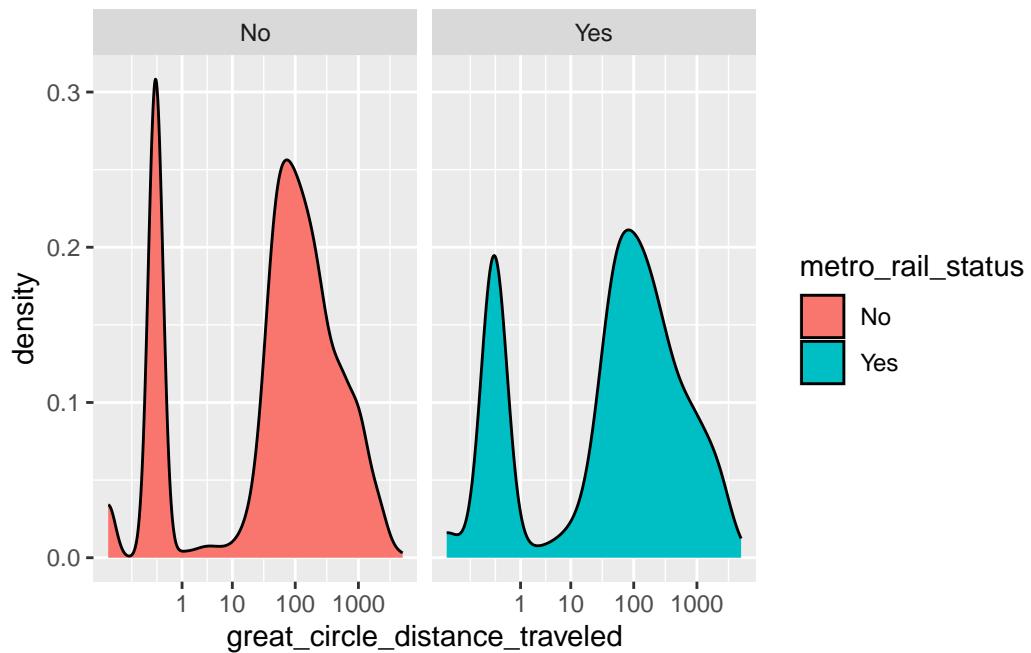
```
[1] "metro_area_size"
```



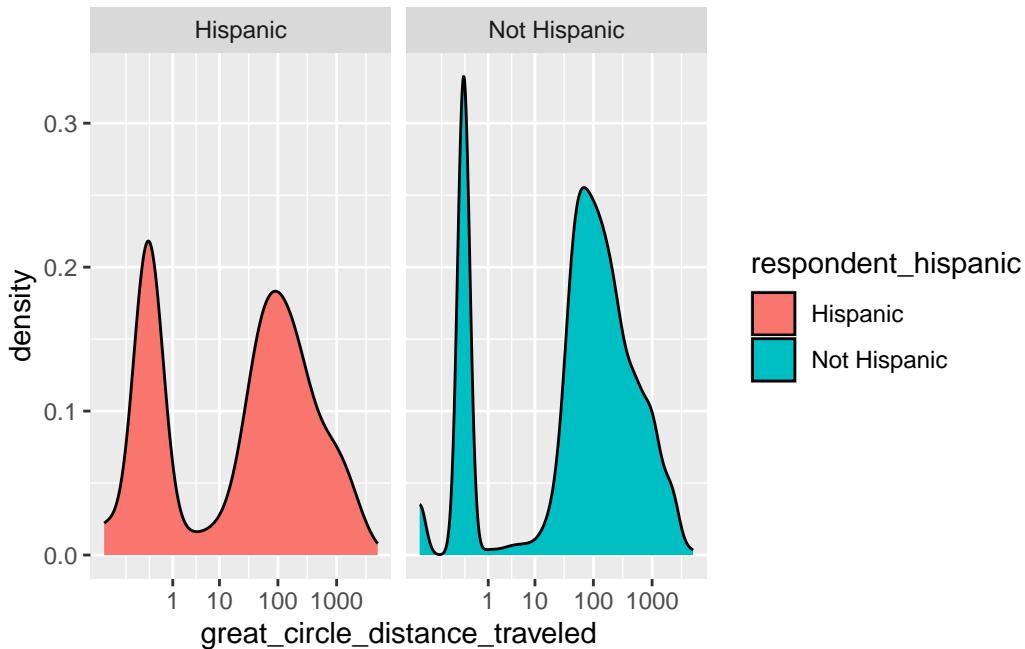
```
[1] "proxy_respondent"
```



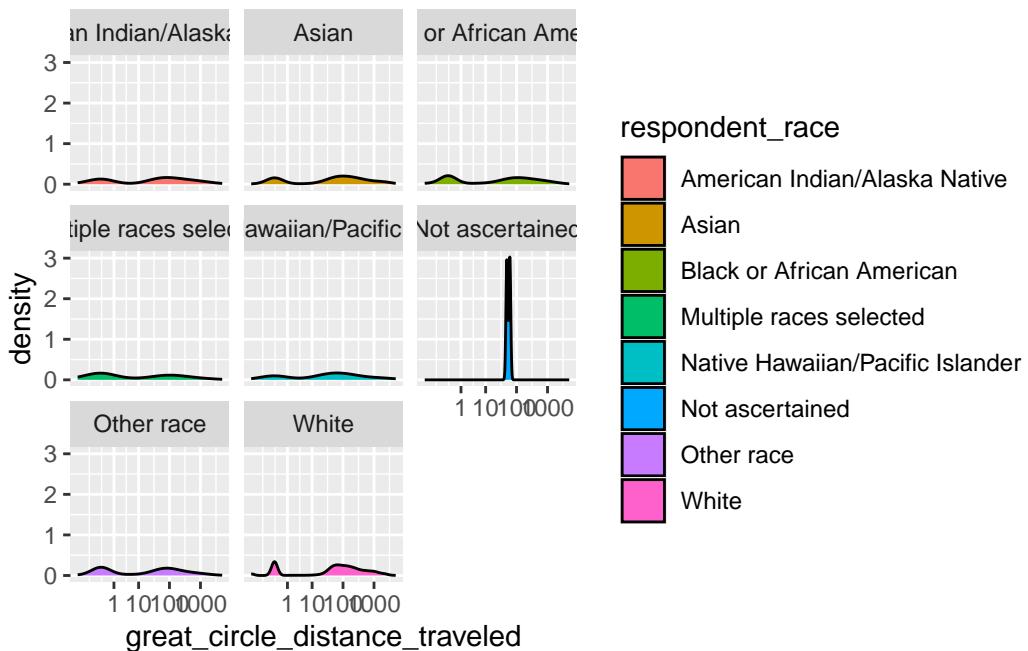
```
[1] "metro_rail_status"
```



```
[1] "respondent_hispanic"
```



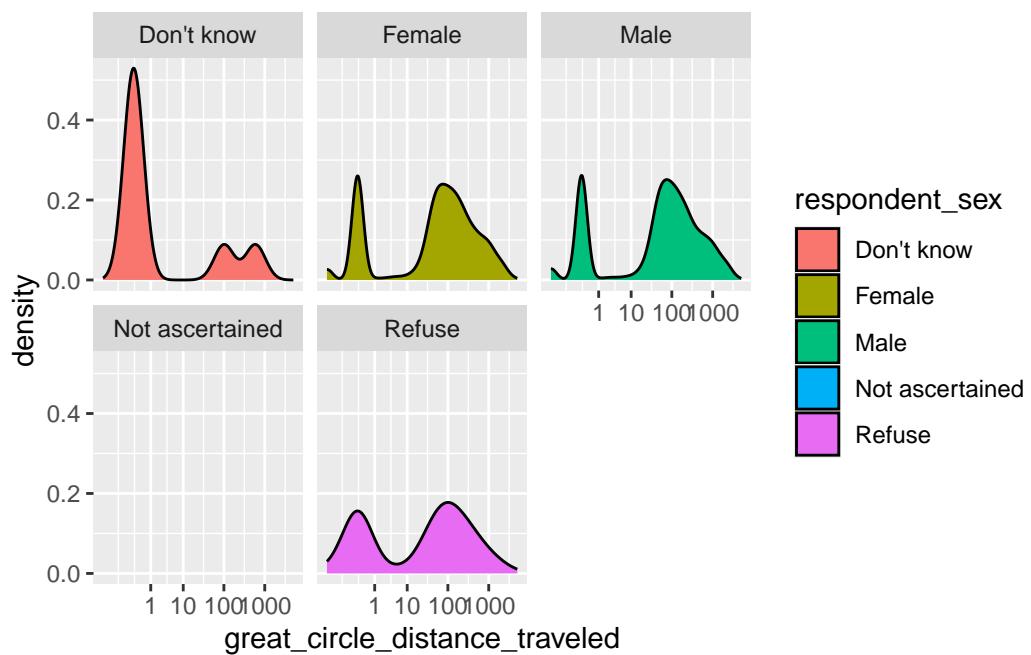
```
[1] "respondent_race"
```



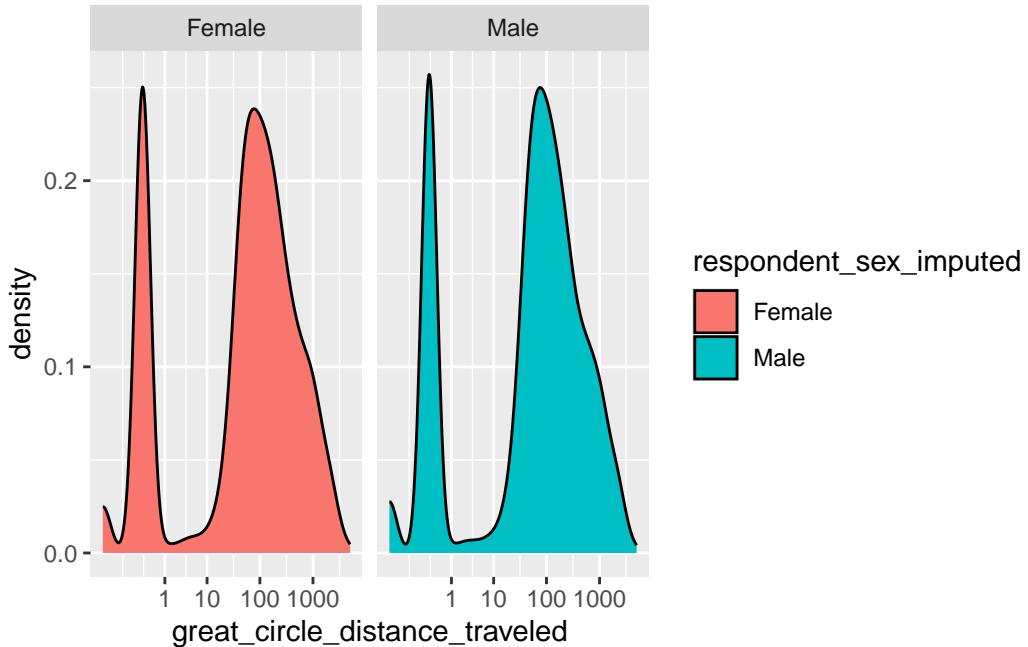
```
[1] "respondent_sex"
```

Warning: Groups with fewer than two data points have been dropped.

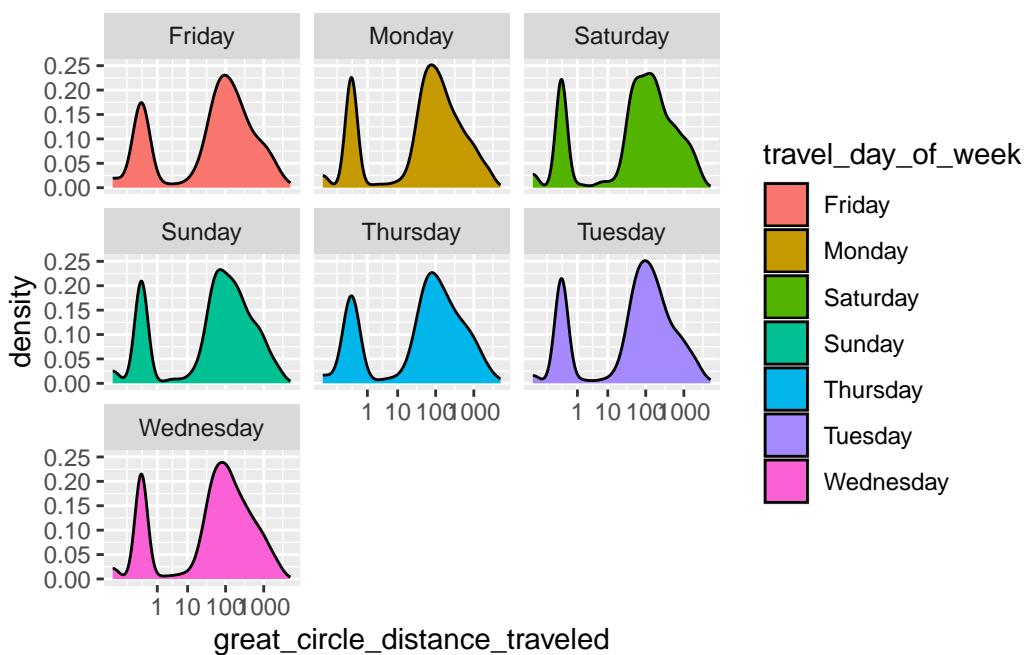
Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning -Inf



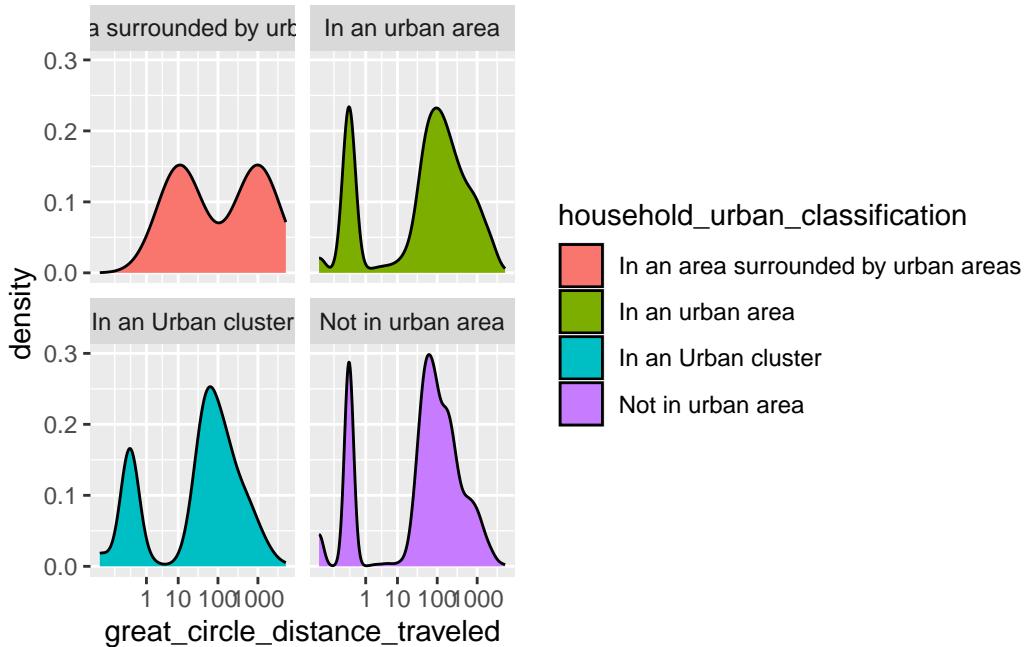
```
[1] "respondent_sex_imputed"
```



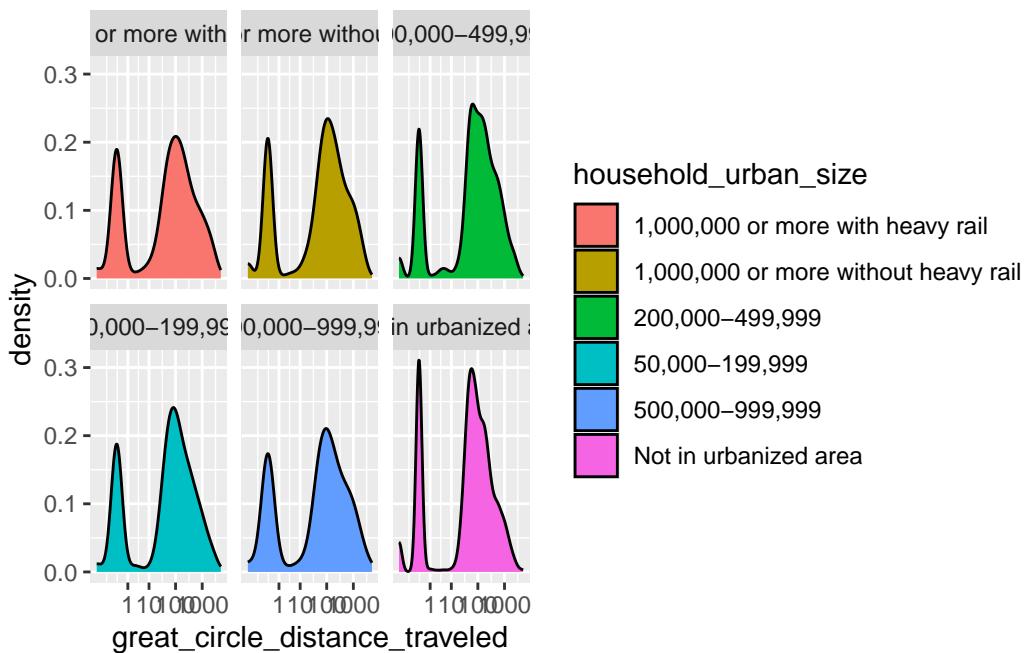
```
[1] "travel_day_of_week"
```



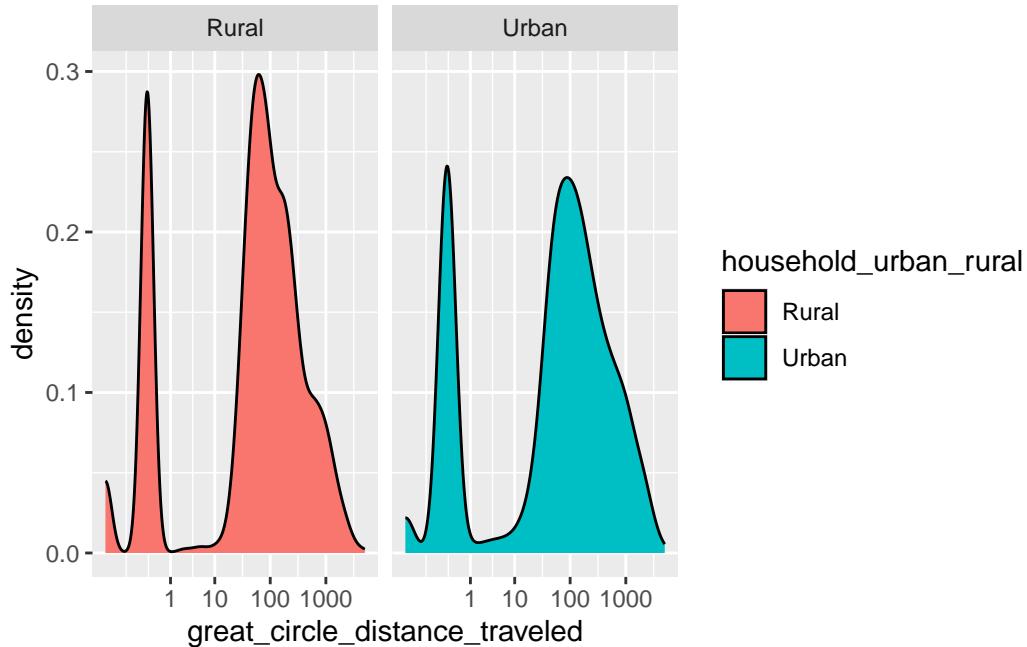
```
[1] "household_urban_classification"
```



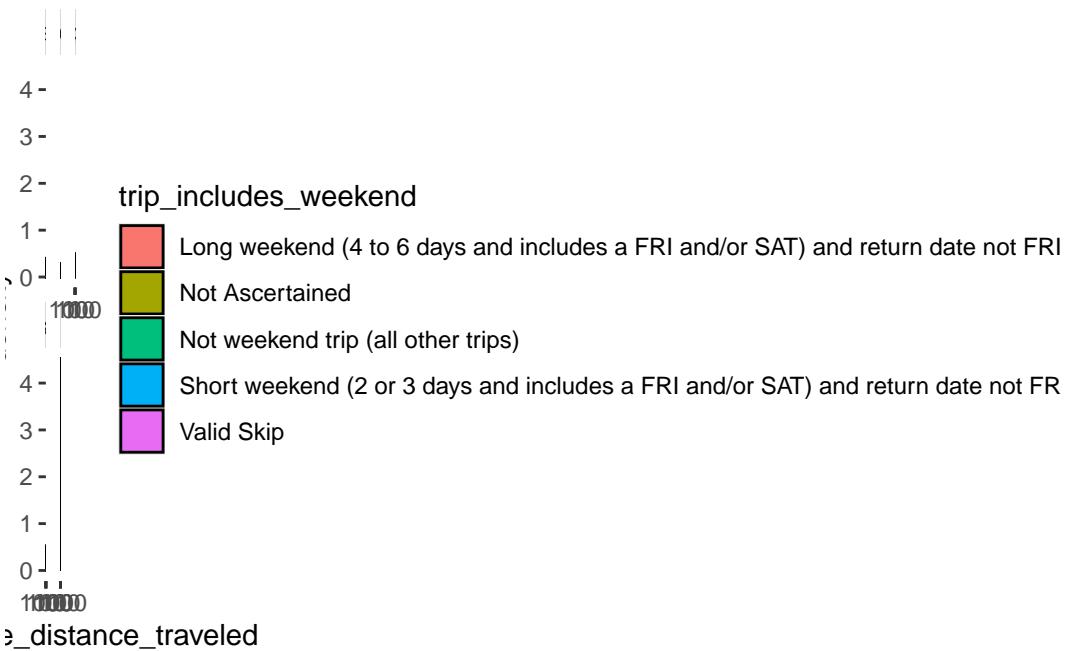
```
[1] "household_urban_size"
```



```
[1] "household_urban_rural"
```



```
[1] "trip_includes_weekend"
```



```
[1] "worker_status"
```

