

Summarizing Rulesy Error Codes Fixie

PSRC

2023-10-04

Contents

function to summarize a variable across flags	4
Error flags by mode	4
function to summarize avariable across a single flag	5

```
library(data.table) # general data manipulation
library(srvyr)      # survey statistics
```

```
##
## Attaching package: 'srvyr'

## The following object is masked from 'package:stats':
##
##   filter
```

```
library(dplyr)      # helpers for srvyr (mutate, group_by, and %>% pipe),
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':
##
##   between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(stringr)    # plus function near(), used in qa/qc checks
library(flextable)  # formatted tables
library(ggplot2)    # plotting
library(scales)     # ggplot2 helpers
library(plotly)     # interactive plots
```

```

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following objects are masked from 'package:flextable':
##
##     highlight, style

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout

library(rstudioapi) # for setting working directory
library(psrclmer)
library(tidyr)

# Point to where the 2023 codebook lives:
codebook_path = 'J:/Projects/Surveys/HHTravel/Survey2023/Data/data_deliverable_81823/codebook_guide/PSR

# Read Data =====

hh_rulesy<-get_table(db_name = "HouseholdTravelSurvey2023", "combined_data", "v_household")
person_rulesy<-get_table(db_name = "HouseholdTravelSurvey2023", "combined_data", "v_person")
day_rulesy<-get_table(db_name = "HouseholdTravelSurvey2023", "combined_data", "v_day")
trip_raw<-get_table(db_name = "HouseholdTravelSurvey2023", "combined_data", "v_trip")
trip_rulesy_before<-get_table(db_name = "hhts_cleaning", "HHSurvey", "Trip")
trip_error_flags<-get_table(db_name = "hhts_cleaning", "HHSurvey", "trip_error_flags")
error_types<-get_table(db_name = "hhts_cleaning", "HHSurvey", "error_types")

## Codebook -----

variables = readxl::read_xlsx(
  codebook_path,
  sheet = 'variable_list_2023'
)

values = readxl::read_xlsx(
  codebook_path,
  sheet = 'value_labels_2023'
)

```

```
## Warning: Expecting logical in H1461 / R1461C8: got 'could be imputed'

## Warning: Expecting logical in H1468 / R1468C8: got 'new meaning for "can drive"
## labels'

## Warning: Expecting logical in H1469 / R1469C8: got 'new meaning for "can drive"
## labels'

## Warning: Expecting logical in H1470 / R1470C8: got 'new meaning for "can drive"
## labels'

## Warning: Expecting logical in H1535 / R1535C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1545 / R1545C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1590 / R1590C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1600 / R1600C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1645 / R1645C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1655 / R1655C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1700 / R1700C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1710 / R1710C8: got 'does PSRC want one of these
## to map to the new value, "Other motorcycle in household"?'

## Warning: Expecting logical in H1811 / R1811C8: got 'can be condensed with mode
## "car" for compatibility between 2021 & 2023'
```

```
setDT(variables)
setDT(values)
```

Create visualizable data tables, before and after the changes. Also show the errors

```
trip_raw<-trip_raw%>%mutate(tripid=as.character(tripid))
trip_rulesy<-trip_rulesy_before%>%mutate(tripid=as.character(tripid))
trip_rulesy<-left_join(trip_rulesy, trip_error_flags, by = c('person_id', 'tripnum'), suffix=c('_rulesy'
#trip_rulesy<-trip_rulesy%>%mutate(person_id=as.character(person_id))
#person_rulesy<-person_rulesy%>%mutate(person_id=as.character(person_id))
#trip_rulesy<-left_join(trip_rulesy, person_rulesy, by = 'person_id')

trip_rulesy_subset<-trip_rulesy%>%
  select(error_flag, person_id, day_id, tripnum, tripid, depart_time_timestamp, arrival_time_timestamp,

trip_rulesy_raw<-left_join(trip_raw, trip_rulesy, by='tripid', suffix=c('_before', '_after'))%>%
  select(error_flag,person_id_before, day_id_before, day_id_after, tripnum_after, tripid,origin_purpose,
```

select all records of a person who has a particular error mode

```
# first find the person ids of trips with the error code  
##trip_rulesy%>%s
```

function to summarize a variable across flags

```
var_error_compare<-function(df, dfname, id, var_name){  
  #id=!!ensym(id)  
  #var_name=!!ensym(var_name)  
  df<-df%>%filter(!is.na(error_flag))  
  df_cols<-df%>%select(!ensym(id), !ensym(var_name), error_flag)  
  vals<- values%>%filter(variable==sym(var_name))  
  df_cols<-df_cols%>%  
    left_join(vals, by=join_by(!ensym(var_name)==value))  
  df_summary<-df_cols%>%  
    group_by(final_label, error_flag)%>%  
    count()  
  df_summary<-df_summary%>%  
    rename(!quo_name(dfname):=n)  
}
```

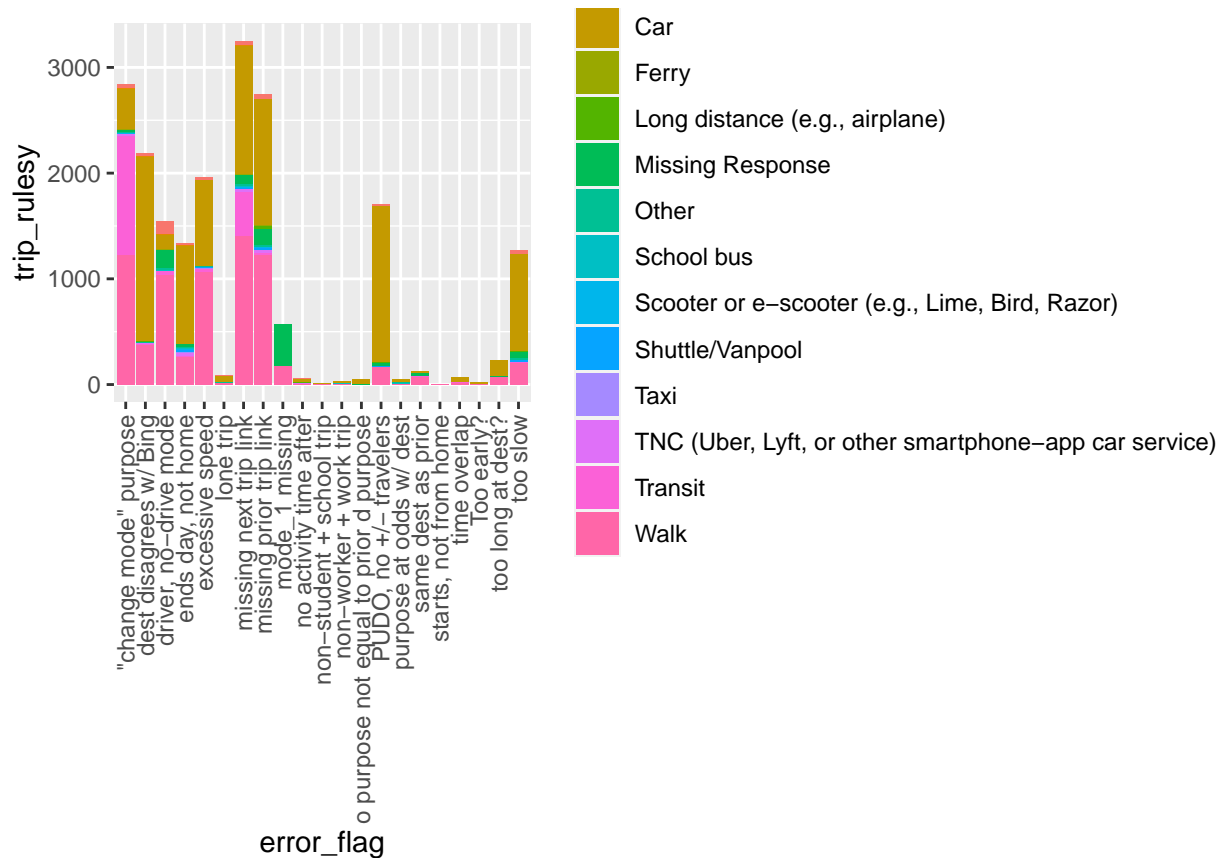
```
trip_mode_data_rulesy<-var_error_compare(df=trip_rulesy,dfname='trip_rulesy',id='tripid', var_name='mode')
```

Adding missing grouping variables: 'person_id', 'tripnum'

Error flags by mode

There are very high numbers of walk, transit trips with the change mode purpose and/or missing next trip link or missing prior trip link. This is why we think there is some problem with trip linking.

```
ggplot(data=trip_mode_data_rulesy, aes(x=error_flag, y=trip_rulesy, fill= final_label))+  
  geom_bar(stat="identity")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



#Summaries of change mode purpose trips only, post-rulesy

```
example_recs_change_mode<-trip_rulesy_raw%>%filter(error_flag=="change mode" purpose')
example_recs_missing_trip_link<-trip_rulesy_raw%>%filter(error_flag=='missing next trip link')
```

function to summarize variable across a single flag

```
var_compare<-function(df, id, var_name){
  #id=!!ensym(id)
  #var_name=!!ensym(var_name)

  df_cols<-df%>%select(!!ensym(id), !!ensym(var_name))
  vals<- values%>%filter(variable==sym(var_name))
  df_cols<-df_cols%>%
    left_join(vals, by=join_by(!!ensym(var_name)==value))
  df_summary<-df_cols%>%
    group_by(final_label)%>%
    count()
}
```