

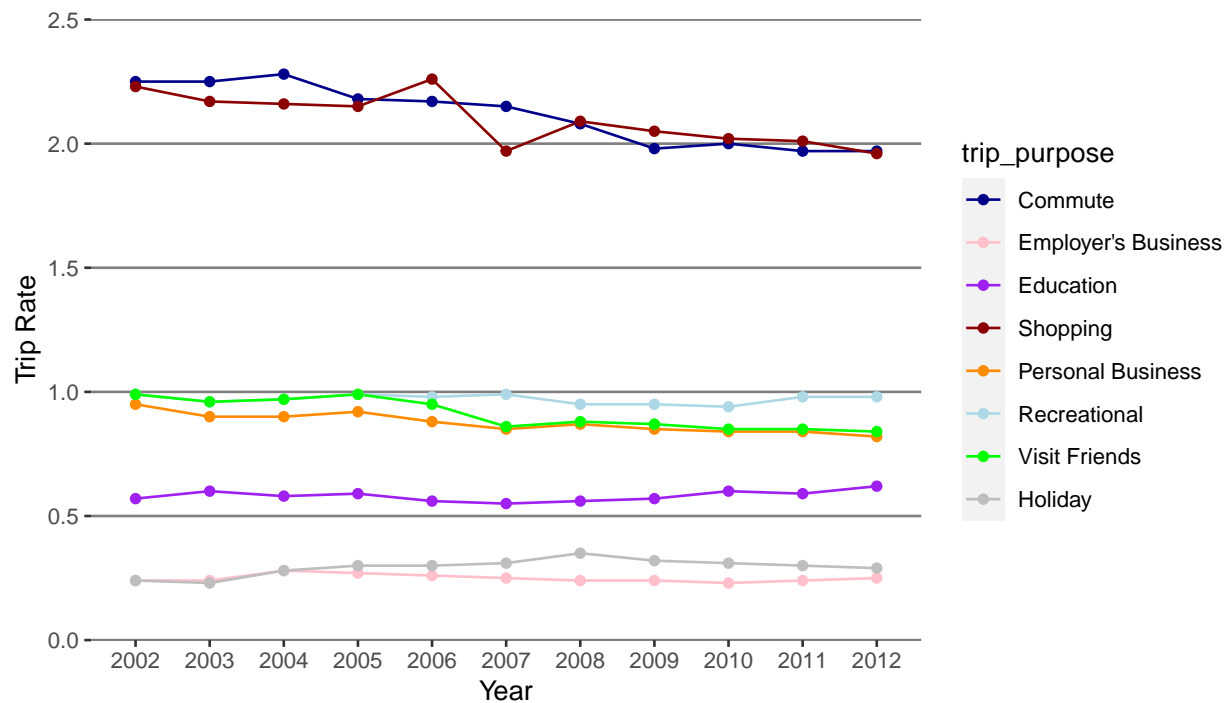
# TfN HB Trip Rates Data Assurance

Humza Ali

15/03/2021

## Summary

This documents walks through the methodology implemented to produce a processed data set ready for trip rates model building. This is a minimum reproducible example, where the aim is to reproduce Figure 3-1 in the AECOM/ICL/ATKINS trip rates report<sup>1</sup>. This will be referred to as the 2016 trips rates report in this document.



By reproducing the data and Figure from the report above, TfN can ascertain any discrepancies found from TfN's trip rate model to NTEM outputs is due to TfN's trip rates model and not the underlying data inserted into the model. This issue has been a cause for concern as outputs for education and recreational purposes do not align well with control outputs from NTEM, such as trip productions for the UK population etc.

Potential causes for any discrepancies have been identified as incorrect classifications from the NTS variable 'TripPurpTo\_B01ID' to trip purpose or incorrect weighting method.

The output at the end of this document is the closest we have been able to obtain by following the methodology in the above report and another report by the DfT in 2009 by Ian Williams et al<sup>2</sup>. This will be referred to as the 2009 trip rates report in this document.

<sup>1</sup>1. Provision of Travel Trends Analysis and Forecasting Model Research (page 17)

<sup>2</sup>2. Research into Changing Trip Rates over Time and Implications for the National Trip End Model: Final Report - Department for Transport

The first section details simple processing of the data to replicate the number of individual records stated in the 2016 trip rates report. This is followed by defining the classifications of trip purpose.

## Loading Dependencies and Processing data

The code below is self-explanatory. Packages are loaded and the csv file containing an unclassified build of NTS Raw data is read.

Variables of interest are selected and filtered to contain only data from 2002 to and including 2012. The diary sample is also selected.

NOTE: You will need to install the packages if not already installed to run the source code.

```
library(dplyr)
library(readr)
library(stringr)
library(tidyr)
library(ggplot2)
library(kableExtra)
library(magrittr)
library(htmltools)

# Read in unclassified Build
df <- read_csv("Y:/NTS/import/tfn_unclassified_build19.csv")

# Select necessary columns to reproduce plot
df <- df %>%
  select(IndividualID, TripPurpFrom_B01ID, TripPurpTo_B01ID, MainMode_B04ID, TripDisIncSW,
         SurveyYear, W1, W2, W5, W5xHH)

# Filter for same number of years as report and select the diary sample
df <- df %>%
  filter(between(SurveyYear, 2002, 2012),
         W1 == 1)

head(df)
```

```
## # A tibble: 6 x 10
##   IndividualID TripPurpFrom_B0~ TripPurpTo_B01ID MainMode_B04ID TripDisIncSW
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1  2002000005         23             6             12             3
## 2  2002000005          6            23            12             4
## 3  2002000005         23            14            11            72.5
## 4  2002000005         23            14            11            72.5
## 5  2002000005         23            14            11            72.5
## 6  2002000005         23            14            11            72.5
## # ... with 5 more variables: SurveyYear <dbl>, W1 <dbl>, W2 <dbl>, W5 <dbl>,
## #   W5xHH <dbl>
```

The number of Individuals recorded is:

```
## [1] 209873
```

which matches the figure stated in the 2016 report. Therefore, up until now, the data extracted is correct. The sample sizes of individuals, households and trips by year has been confirmed by cross-checking with NTS reports and table 6.2 in the 2009 report.

## Classifications of Trip Purpose

The definition of trip purpose is based on Table 7.1 from the 2009 report. To match NTEM purpose's like for like, Purposes 4 & 5 are switched and Purposes 7 & 8 are switched. Therefore, the correct classifications of trip purposes are defined as:

No	NTEM Purpose	Composition	NTS Classification
1	Commute	Work	TripPurpTo_B01ID == 1
		Escort to Work	TripPurpTo_B01ID == 18
2	EB	In course of work	TripPurpTo_B01ID == 2
		Escort in course of work	TripPurpTo_B01ID == 19
3	Education	Education	TripPurpTo_B01ID == 3
		Escort Education	TripPurpTo_B01ID == 20
4	Shopping	Food Shopping	TripPurpTo_B01ID == 4
		Non-Food Shopping	TripPurpTo_B01ID == 5
		Escort PB/Shopping	TripPurpTo_B01ID == 21
5	PB	PB Medical	TripPurpTo_B01ID == 6
		PB eat/drink	TripPurpTo_B01ID == 7
		PB Other	TripPurpTo_B01ID == 8
6	Recreation	Eat/drink W friends	TripPurpTo_B01ID == 9
		Other Social	TripPurpTo_B01ID == 11
		Entertainment	TripPurpTo_B01ID == 12
		Sport	TripPurpTo_B01ID == 13
		Day trip (just walk trips)	TripPurpTo_B01ID == 15 & MainMode_B04ID == 1
		Other non-escort	TripPurpTo_B01ID == 16
		Other escort	TripPurpTo_B01ID == 22
7	Visit Friend	Visit friends	TripPurpTo_B01ID == 10
		Escort Home	TripPurpTo_B01ID == 17
8	Holiday	Holiday	TripPurpTo_B01ID == 14
		Day Trip (Excluding Just Walk Trips)	TripPurpTo_B01ID == 15 & MainMode_B04ID != 1

Classify trip purposes based on the above table:

```
# Classify hb trip purposes for outbound only based on above:
df <- df %>%
  filter(TripPurpFrom_B01ID == 23) %>% # Trips from Home only
  mutate(trip_purpose = case_when(
    TripPurpTo_B01ID == 1 ~ 1,
    TripPurpTo_B01ID == 2 ~ 2,
    TripPurpTo_B01ID == 3 ~ 3,
    TripPurpTo_B01ID == 4 ~ 4,
    TripPurpTo_B01ID == 5 ~ 4,
    TripPurpTo_B01ID == 6 ~ 5,
    TripPurpTo_B01ID == 7 ~ 5,
    TripPurpTo_B01ID == 8 ~ 5,
```

```

TripPurpTo_B01ID == 9 ~ 6,
TripPurpTo_B01ID == 10 ~ 7,
TripPurpTo_B01ID == 11 ~ 6,
TripPurpTo_B01ID == 12 ~ 6,
TripPurpTo_B01ID == 13 ~ 6,
TripPurpTo_B01ID == 14 ~ 8,
TripPurpTo_B01ID == 15 & MainMode_B04ID == 1 ~ 6,
TripPurpTo_B01ID == 15 & MainMode_B04ID != 1 ~ 8,
TripPurpTo_B01ID == 16 ~ 6,
TripPurpTo_B01ID == 17 ~ 7,
TripPurpTo_B01ID == 18 ~ 1,
TripPurpTo_B01ID == 19 ~ 2,
TripPurpTo_B01ID == 20 ~ 3,
TripPurpTo_B01ID == 21 ~ 4,
TripPurpTo_B01ID == 22 ~ 6))

```

## Short Walk weighting

Although it is best practice to use the specified weights in the NTS for short walking trips, for this minimum reproducible example, we have simply put an if else function to calculate weights. The logic is, if a trip is under 1 mile and is a walking trip then we assign a short walk weight of 7, otherwise assign a default 1.

```
df <- mutate(df, sw_weight = ifelse(TripDisIncSW < 1 & MainMode_B04ID == 1, 7, 1))
```

## Apply trip weighting and prepare data set for model building

The dataset has been filtered and trip purposes classified for home-based outbound trips only. The next step is to calculate weekly weighted trip rates. Two attempts are shown in this document to calculate a correct weighting for trips.

1) Trip weights calculation - 2 attempts

a. Ian Williams 2009 methodology:  $W5xHH \times W2$

There is a concern this weighting methodology is incorrect. The rationale behind this is as follows:

$W5xHH = \frac{W5}{W2}$ , therefore  $W5 = W5xHH \times W2$ . Thus, it is unclear why  $W5$  cannot be substituted on behalf of this multiplication. If this is indeed correct, then we can simply use  $W5$ .

b. NTS recommended methodology:  $W5 \times W2$

Therefore, this report will output 2 plots. One for each of the above scenarios.

```

df <- df %>%
  mutate(IW_weight = W5xHH * W2 * sw_weight,
         nts_weight = W5 * W2 * sw_weight)

```

where *sw\_weight* is the short walk weight.

2) Aggregate to obtain weekly trip rates for an individual by trip purpose

```
df <- df %>%
  group_by(IndividualID, trip_purpose, SurveyYear, W2) %>%
  summarise(IW_weight = sum(IW_weight),
            nts_weight = sum(nts_weight)) %>%
  ungroup()

head(df)
```

```
## # A tibble: 6 x 6
##   IndividualID trip_purpose SurveyYear    W2 IW_weight nts_weight
##       <dbl>      <dbl>    <dbl> <dbl>   <dbl>    <dbl>
## 1  2002000001         3      2002 0.988     7.73     7.63
## 2  2002000001         4      2002 0.988     1.11     1.10
## 3  2002000001         6      2002 0.988     4.26     4.21
## 4  2002000001         7      2002 0.988     1.99     1.96
## 5  2002000002         1      2002 0.988     5.08     5.02
## 6  2002000002         3      2002 0.988     1.01     0.994
```

- 3) Each individual must have a record for the total number of trips by trip purpose. If an individual has not recorded a trip for a certain trip purpose, we infill this trip purpose with 0 trips. Therefore, the total number of weekly trip records is: The number of Individuals \* 8

```
df <- df %>%
  complete(nesting(IndividualID, SurveyYear, W2),
            trip_purpose = 1:8,
            fill = list(IW_weight = 0,
                        nts_weight = 0))

head(df)
```

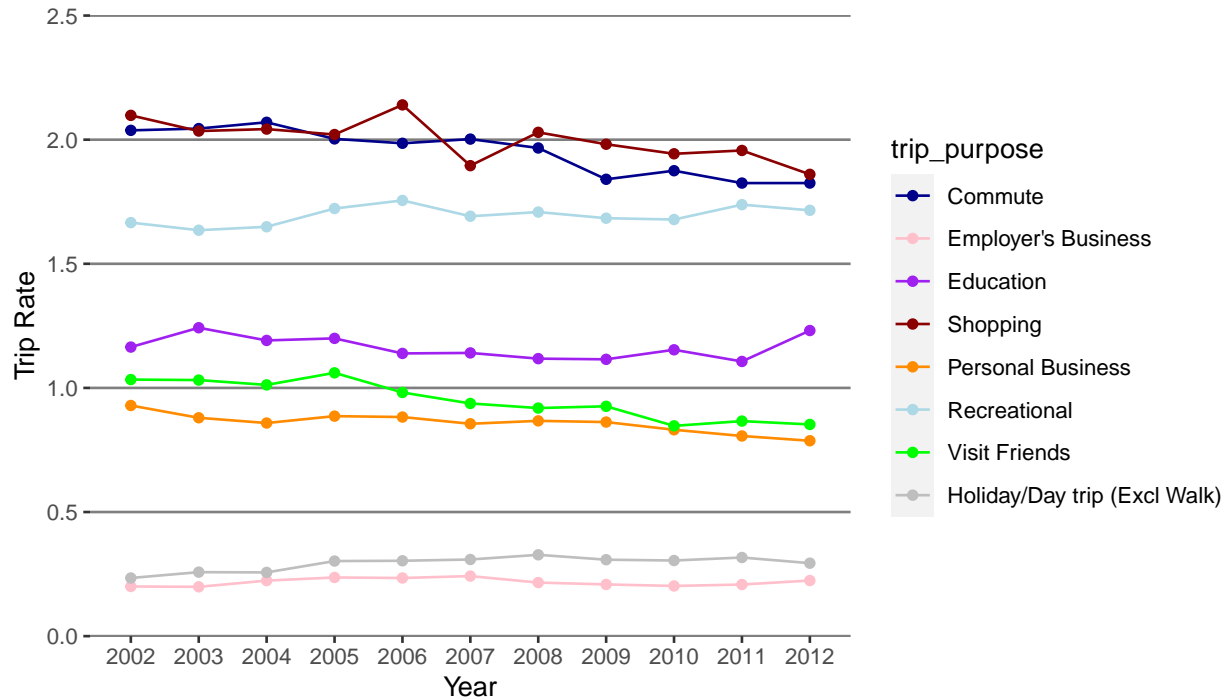
```
## # A tibble: 6 x 6
##   IndividualID SurveyYear    W2 trip_purpose IW_weight nts_weight
##       <dbl>      <dbl> <dbl>      <dbl>   <dbl>    <dbl>
## 1  2002000001      2002 0.988         1       0       0
## 2  2002000001      2002 0.988         2       0       0
## 3  2002000001      2002 0.988         3     7.73     7.63
## 4  2002000001      2002 0.988         4     1.11     1.10
## 5  2002000001      2002 0.988         5       0       0
## 6  2002000001      2002 0.988         6     4.26     4.21
```

- 4) To reproduce Figure 3-1, aggregate by year and trip purpose, and calculate the average trip rate = sum of trips/sum of individuals weighted by household weight.

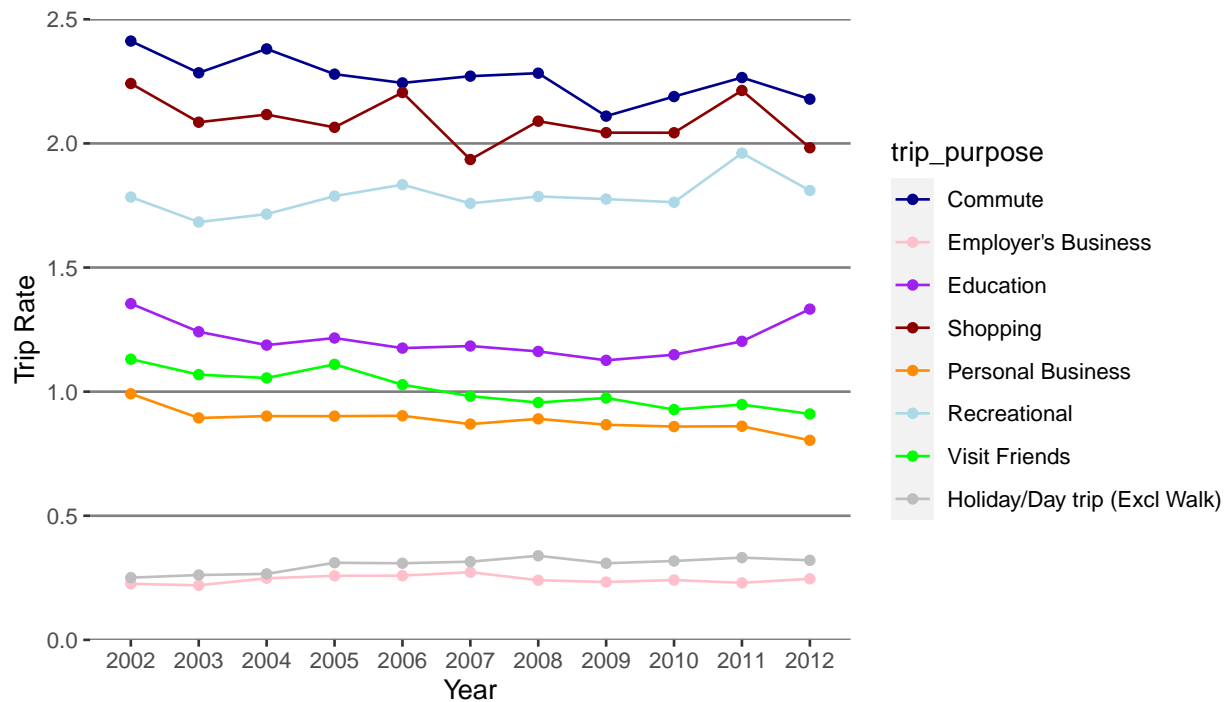
```
df <- df %>%
  group_by(trip_purpose, SurveyYear) %>%
  summarise(IW_trip_rate = sum(IW_weight)/sum(W2),
            nts_trip_rate = sum(nts_weight)/sum(W2)) %>%
  ungroup()
```

## Output

A: 2009 report methodology - Weighted trips =  $W5xHH \times W2$



B: NTS recommendation - Weighted trips =  $W5 \times W2$



## Discussion

Both outputs clearly do not match Figure 3-1 in the 2016 AECOM/ICL/ATKINS report. Output B seems to be the worst, whereas A exhibits signs of some association with the control figure. The control plot presents lower trip rates for recreational, visiting friends and education purposes. This is evident as the control figure

has no purpose completely lie in the range of 1-2 trip rates, whereas the 2 outputs above have the suspected purposes within this range.

It is not clear where the discrepancies arise from. The NTS raw data has proven to be from the same extract as the 2016 and 2009 report. Therefore there needs to be further assurance on the procedure above to process the raw dataset to a dataset ready for hb trip rates model building.

There are many talking points to go through, which would be more beneficial to discuss formally, however there are clear signs the data loosely follows the control but educational, recreational and visiting friends purposes are both far off from the control. These three purposes do not align well with TfN's trip rate model outputs to NTEM, therefore we hope to obtain some guidance on what is causing the discrepancy which in turn will converge our model outputs to NTEM outputs. We suspect the trip purpose classifications and weighting methodology requires revisiting. Therefore our questions are:

- 1) Are the classifications for trip purposes correct? Although some purposes in the 3 outputs above match the control figure well, they still do not match like-for-like which is expected as the same data is used.
- 2) What is the correct weighting methodology?
- 3) What is required to be adjusted or added in the process above to replicate the figure?