Flagler Data Review

Amar Sarvepalli

September 27, 2016

# Introduction

This memo is divided into following sections:

1. Airsage Data Files
2. Review of Data
3. AirSage OD Flows
4. Review of ACS / CTPP JTW Flows
5. Census socio-economic data (2000, 2010, 2015 and 2040)
6. Review of APC data for Flagler
7. Model validation (compare CTPP Trn Flows and AirSage to Model)

knitr::opts\_chunk$set(echo = TRUE, fig.width = 9.5)  
# User inputs:  
projDir <- "/Users/amarsarvepalli/Desktop/github/Flagler"  
  
# Airsage data files  
airsage\_data <- c("airsage/trip\_leg\_matrix\_cusWDH.csv",  
 "airsage/trip\_leg\_matrix\_cusWDDP.csv",  
 "airsage/trip\_leg\_submatrix\_cusWDDP.csv",  
 "airsage/trip\_leg\_submatrix\_cusWDH.csv")  
airsage\_shapeFile <- "airsage/airsarge\_districts/TAZs.shp"  
airsage\_Districts <- "airsage/airsarge\_districts/district\_equivalency.csv"  
  
# Define corridor districts  
Study\_districts <- c(20:40)  
  
# Gtfs network  
gtfsShapesFile <- "GTFS/miami-dade-transit\_20150626\_0146/shapes.txt"  
gtfsTripsFile <- "GTFS/miami-dade-transit\_20150626\_0146/trips.txt"  
gtfsRoutesFile <- "GTFS/miami-dade-transit\_20150626\_0146/routes.txt"  
routeName <- "flagler"  
  
# CTPP shape file  
ctpp\_ShapeFile <- "ctpp/ctpp\_centroids\_districts.shp"  
ctpp\_flowFile <- "FLACSJTW/FL\_ctpp3\_sumlv140.ACS"  
  
# Census ACS data files  
acs\_shapeFile <- "acs\_data/Tract\_2014\_Pop\_Emp.shp"  
  
# APC data  
apc\_dir <- "bus\_apc\_data/Flagler St BRTStudy APC Data Oct 2015"  
  
# Output files  
daily\_od\_matrix <- "daily\_od\_trips\_by\_airsage\_districts.csv"  
peak\_od\_matrix <- "peak\_od\_trips\_by\_airsage\_districts.csv"  
daily\_od\_submatrix <- "daily\_od\_sub\_trips\_by\_airsage\_districts.csv"  
peak\_od\_submatrix <- "peak\_od\_sub\_trips\_by\_airsage\_districts.csv"

## 1. AirSage Data Samples

#### 1.1 Airsage Data File Description

This is airsage data review. The following is the description of the data files

1. **trip\_leg\_matrix\_cusWDH.csv** and **trip\_leg\_matrix\_cusWDDP.csv**: reporting trips between study area zones for average weekday 24 hour period and specific AM and PM day parts respectively.
2. **trip\_leg\_submatrix\_cusWDH.csv** and **trip\_leg\_submatrix\_cusWDDP.csv**: reporting trips between study area zones that passed through the sub matrix analysis zone (in the SZCount column) for average weekday 24 hour period and specific AM and PM day parts respectively.

The following are the list of files received from Airsage along with: 1) number of samples 2) number of trips 3) ratio of trips to samples 4) percent of samples 5) percent of trips

##   
##   
## File Samples Trips Trips/Samples % Sample % Trips  
## ------------------------------- -------- ------------ -------------- --------- --------  
## trip\_leg\_matrix\_cusWDDP.csv 9666 2945178.48 30,469 38 19  
## trip\_leg\_matrix\_cusWDH.csv 6052 7457084.91 123,217 24 47  
## trip\_leg\_submatrix\_cusWDDP.csv 5143 1472307.68 28,627 20 9  
## trip\_leg\_submatrix\_cusWDH.csv 4301 4001471.97 93,036 17 25  
## Total 25162 15876043.04 63,095 100 100

Peak data flow shows a total observations of 9,666 whereas total daily flows are only 6,052. Does this mean the daily samples are only for offpeak? 46% of the daily flow are represented from 24% sample?

#### 1.2 Contents of "trip\_leg\_matrix\_cusWDDP.csv" and Contents of "trip\_leg\_matrix\_cusWDDP.csv"

This file contains 9,666 observations and 9 fields. Each sample was geocoded with an *Origin\_Zone* and *Destination\_Zone* pairs and *Count* represents the total flows betweent these OD pairs. The start and end dates define the data collection period, which is April, 2015. The other fields: *Subscriber\_Class*, *Purpose* and *Time\_of\_Day* describe the trip attributes.

data\_wddp <- read.csv("airsage/trip\_leg\_matrix\_cusWDDP.csv")  
 data\_wdh <- read.csv("airsage/trip\_leg\_matrix\_cusWDH.csv")  
 data\_wddp$tod <- "peak"  
 data\_wdh$tod <- "24H"  
 data <- rbind(data\_wddp, data\_wdh)  
 print(head(data))

## Origin\_Zone Destination\_Zone Start\_Date End\_Date Aggregation  
## 1 14 11 20150401 20150430 WD  
## 2 11 34 20150401 20150430 WD  
## 3 24 11 20150401 20150430 WD  
## 4 10 2 20150401 20150430 WD  
## 5 11 20 20150401 20150430 WD  
## 6 2 17 20150401 20150430 WD  
## Subscriber\_Class Purpose Time\_of\_Day Count tod  
## 1 Resident HBO H07:H10 2273.11 peak  
## 2 Resident HBO H07:H10 283.50 peak  
## 3 Resident NHB H16:H19 336.66 peak  
## 4 Resident NHB H16:H19 3715.68 peak  
## 5 Resident HBW H07:H10 1644.29 peak  
## 6 Visitor NHB H16:H19 1363.58 peak

Total trips from both files are shown below.

# Total trips  
 total\_trips <- data %>% group\_by(tod) %>% summarize(sum = sum(Count))  
 kable(total\_trips,format.args = list(big.mark = ","), digits = 0)

|  |  |
| --- | --- |
| tod | sum |
| 24H | 7,457,085 |
| peak | 2,945,178 |

The daily file contains more aggregated data with periods combined and thus the number of samples are fewer compared to AM/PM files. However, the total number of trips reveal that daily includes AM/PM trips although this needs to be confirmed from Airsage.

# Break down by purpose and period  
 samples <- data %>% group\_by(tod, Purpose, Time\_of\_Day) %>% tally  
 trips <- data %>% group\_by(tod, Purpose, Time\_of\_Day) %>% summarize(sum = sum(Count))  
 kable(left\_join(samples,trips, by = c("Purpose" , "Time\_of\_Day", "tod")),  
 format.args = list(big.mark = ","), digits = 0)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tod | Purpose | Time\_of\_Day | n | sum |
| 24H | HBO | H00:H24 | 1,701 | 2,722,739 |
| 24H | HBW | H00:H24 | 1,385 | 1,267,599 |
| 24H | NHB | H00:H24 | 2,966 | 3,466,747 |
| peak | HBO | H07:H10 | 1,332 | 482,475 |
| peak | HBO | H16:H19 | 1,429 | 529,311 |
| peak | HBW | H07:H10 | 1,137 | 363,070 |
| peak | HBW | H16:H19 | 1,068 | 261,128 |
| peak | NHB | H07:H10 | 2,286 | 545,530 |
| peak | NHB | H16:H19 | 2,414 | 763,664 |

## 2. Review of AirSage Data

#### 2.1 Contents of Daily file (*trip\_leg\_matrix\_cusWDH.csv*)

#### Trips by purpose and resident type (Subscriber\_Class)

* There are a total of 7,457,084 OD flows in 24H.
* About 83% of these trips (6.2 Million daily flows) are made by resident pouplation and the remaining 17% of trip are made by the visitors.
* Most of the visitor trip purpose was Non-home based (NHB) as expected, however there are few handful of trips that are coded as Home-based other (HBO) trips (0.03%). These trips can be recoded into NHB as it makes less sense to assume a "home end" for visitors.
* Of 6.2 Million resident trips, 17% are Home-based work (HBW), 29% are NHB and 36% are HBO trips. The work share of trips is in acceptable range for HBW trips (typically ~20%).

## [1] "Total trips = 7457084.91"

|  |  |  |
| --- | --- | --- |
| Subscriber\_Class | trips | percent |
| Resident | 6,201,677 | 83 |
| Visitor | 1,255,408 | 17 |

|  |  |  |  |
| --- | --- | --- | --- |
| Subscriber\_Class | Purpose | trips | percent |
| Resident | HBO | 2,720,440 | 36 |
| Resident | HBW | 1,267,599 | 17 |
| Resident | NHB | 2,213,638 | 30 |
| Visitor | HBO | 2,298 | 0 |
| Visitor | NHB | 1,253,109 | 17 |

#### 2.2 Contents of Peak period files (*trip\_leg\_matrix\_cusWDDP.csv*)

#### Trips by purpose, resident type (Subscriber\_Class) and time of day

* There are a total of 2,945,178 OD flows in the peak two periods.
* About 85% of these trips (2.5 Million flows) are made by resident pouplation and the remaining 15% of trip are made by the visitors.
* Most of the visitor trip purpose was Non-home based (NHB) as expected, however there are few handful of trips that are coded as Home-based other (HBO) trips (0.02%). These trips can be recoded into NHB as it makes less sense to assume a "home end" for visitors.
* Of 2.5 Million resident trips, 21% are Home-based work (HBW), 29% are NHB and 34% are HBO trips. The non-work share of trips seems too high given this file contains data only for the 6 hour peak periods.
* Trips by time of day shows almost equal split between AM and PM peak periods with slightly more trips in PM period.

## [1] "Total trips = 2945178.48"

|  |  |  |
| --- | --- | --- |
| Subscriber\_Class | trips | percent |
| Resident | 2,508,516 | 85 |
| Visitor | 436,663 | 15 |

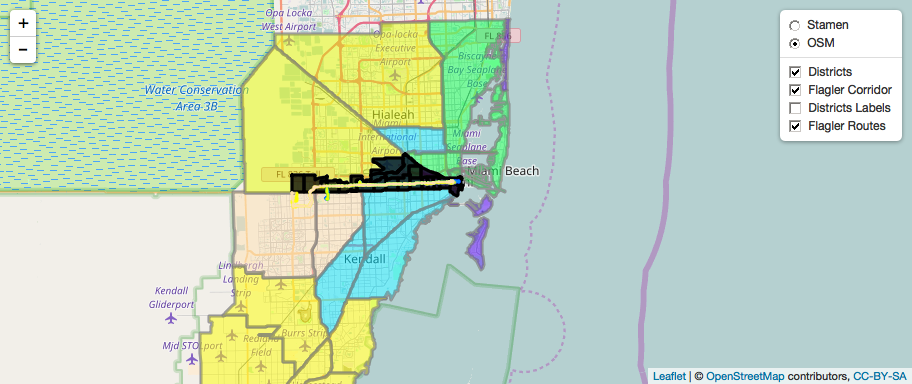
|  |  |  |  |
| --- | --- | --- | --- |
| Subscriber\_Class | Purpose | trips | percent |
| Resident | HBO | 1,011,037 | 34 |
| Resident | HBW | 624,198 | 21 |
| Resident | NHB | 873,281 | 30 |
| Visitor | HBO | 750 | 0 |
| Visitor | NHB | 435,913 | 15 |

|  |  |  |
| --- | --- | --- |
| Time\_of\_Day | trips | percent |
| H07:H10 | 1,391,075 | 47 |
| H16:H19 | 1,554,103 | 53 |

#### 2.3 Airsage districts

The airsage data was provided for about 40 districts as shown in the below interactive map.

# Add airsage district file  
shape <- readOGR(airsage\_shapeFile, layer = "TAZs", verbose = FALSE)  
  
# Read District equivqlency file  
taz\_dist <- read.csv(airsage\_Districts)  
  
# Append districts to 40 Zones  
shape@data <- left\_join(shape@data, taz\_dist, by = "TAZ")



## 3) AirSage OD Flows

#### Study Area Flows

Districts 20 - 40 represent Flagler Corridor study area. The entire study area was defined as one super district "StudyArea" and the rest of the districts as "Outside" super district. The OD matrix generated for this super districts list the External-External (EE), Internal\_External (IE), External\_Internal (EI) and Internal-Internal (II) trips to the study area. The study area serves about 11% of the daily trips in the Miami-Dade county (sum of IE, EI and II trips).

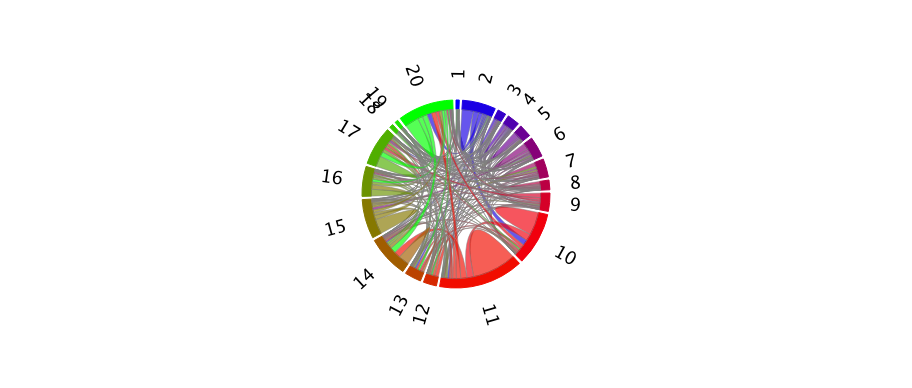
#### 3.1 Daily Study Area Flows

Tables below show the daily OD flows. Inorder to better understand the travel patterns through the corridor, OD tables are developed to & from the studyarea super district to all other districts, where district 20 is the aggregated study area district.

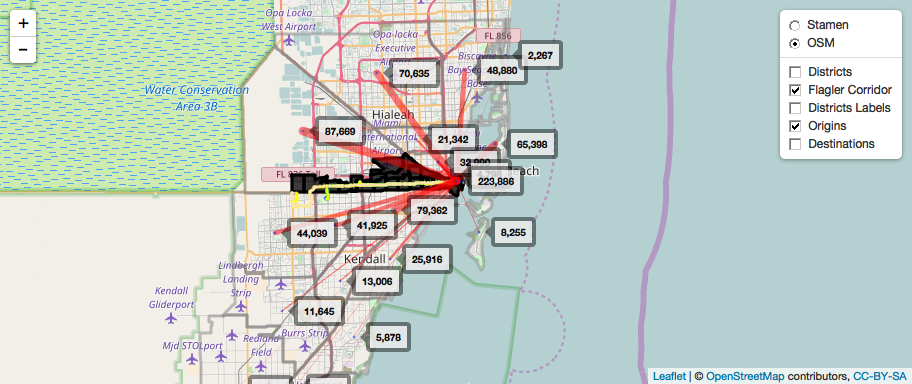
|  |  |  |  |
| --- | --- | --- | --- |
|  | Outside | StudyArea | total |
| Outside | 6,079,863 | 575,832 | 6,655,695 |
| StudyArea | 577,503 | 223,886 | 801,390 |
| total | 6,657,367 | 799,718 | 7,457,085 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | Outside | StudyArea | total |
| Outside | 81.53 | 7.72 | 89.25 |
| StudyArea | 7.74 | 3.00 | 10.75 |
| total | 89.28 | 10.72 | 100.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Origins | Origin\_Pct | Destination | Dest\_Pct |
| 1 | 7,155.59 | 0.89 | 8,255.22 | 1.03 |
| 2 | 74,591.54 | 9.31 | 65,398.40 | 8.18 |
| 3 | 4,332.24 | 0.54 | 3,457.41 | 0.43 |
| 4 | 6,055.95 | 0.76 | 5,162.41 | 0.65 |
| 5 | 3,464.71 | 0.43 | 3,207.18 | 0.40 |
| 6 | 12,026.33 | 1.50 | 11,644.73 | 1.46 |
| 7 | 13,688.95 | 1.71 | 13,006.25 | 1.63 |
| 8 | 6,120.03 | 0.76 | 5,877.74 | 0.73 |
| 9 | 26,166.67 | 3.27 | 25,915.70 | 3.24 |
| 10 | 50,962.80 | 6.36 | 48,879.78 | 6.11 |
| 11 | 70,945.94 | 8.85 | 70,634.60 | 8.83 |
| 12 | 20,345.92 | 2.54 | 21,342.21 | 2.67 |
| 13 | 31,732.68 | 3.96 | 32,990.42 | 4.13 |
| 14 | 76,594.30 | 9.56 | 87,669.36 | 10.96 |
| 15 | 46,132.33 | 5.76 | 44,039.42 | 5.51 |
| 16 | 41,626.60 | 5.19 | 41,924.58 | 5.24 |
| 17 | 79,137.40 | 9.88 | 79,362.06 | 9.92 |
| 18 | 2,100.69 | 0.26 | 2,266.80 | 0.28 |
| 19 | 4,322.68 | 0.54 | 4,797.60 | 0.60 |
| 20 | 223,886.22 | 27.94 | 223,886.22 | 28.00 |
| total | 801,389.57 | 100.00 | 799,718.09 | 100.00 |



Plot of total daily Origins and Destinations to/from Study Area.



#### 3.2 Peak Study Area Flows

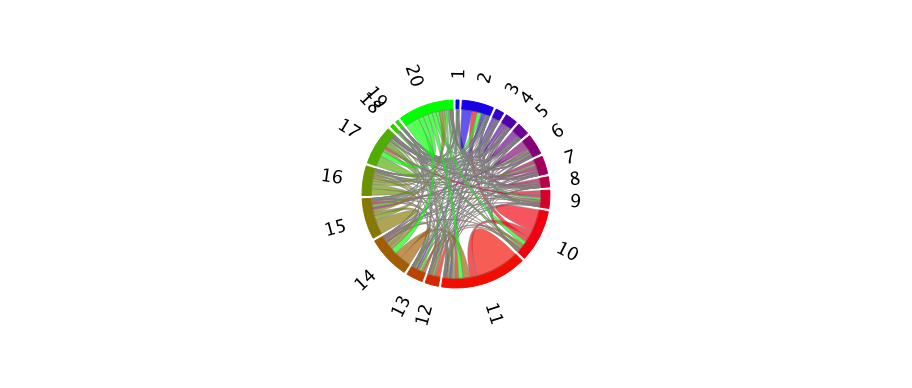
The following are the tables showing the peak trips. Inorder to better understand the travel patterns through the corridor, OD tables are developed to & from the studyarea super district to all other districts, where district 20 is the aggregated study area district.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Outside | StudyArea | total |
| Outside | 2,394,496 | 236,047 | 2,630,544 |
| StudyArea | 225,789 | 88,845 | 314,635 |
| total | 2,620,286 | 324,893 | 2,945,178 |

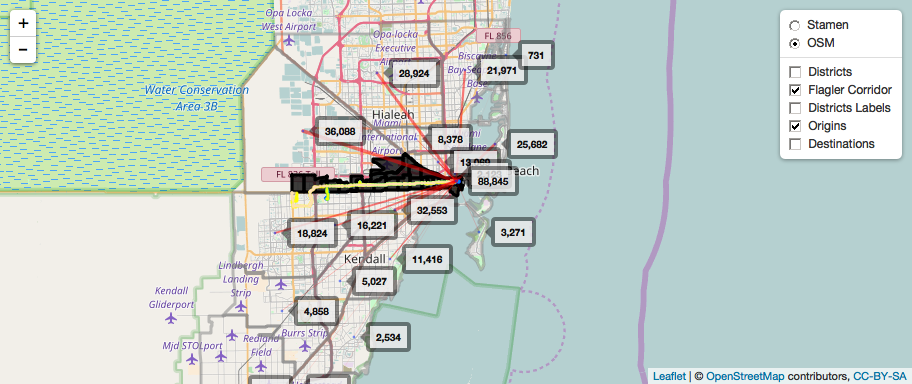
|  |  |  |  |
| --- | --- | --- | --- |
|  | Outside | StudyArea | total |
| Outside | 81.30 | 8.01 | 89.32 |
| StudyArea | 7.67 | 3.02 | 10.68 |
| total | 88.97 | 11.03 | 100.00 |

## [1] "Percent of Peak Trips"

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Origins | Origin\_Pct | Destination | Dest\_Pct |
| 1 | 2,773.64 | 0.88 | 3,271.16 | 1.01 |
| 2 | 24,727.01 | 7.86 | 25,681.84 | 7.90 |
| 3 | 1,530.64 | 0.49 | 1,374.62 | 0.42 |
| 4 | 2,494.50 | 0.79 | 1,980.04 | 0.61 |
| 5 | 1,130.60 | 0.36 | 1,024.44 | 0.32 |
| 6 | 4,675.66 | 1.49 | 4,857.67 | 1.50 |
| 7 | 5,759.00 | 1.83 | 5,027.08 | 1.55 |
| 8 | 2,550.64 | 0.81 | 2,533.91 | 0.78 |
| 9 | 10,519.29 | 3.34 | 11,416.13 | 3.51 |
| 10 | 19,737.23 | 6.27 | 21,970.56 | 6.76 |
| 11 | 27,082.15 | 8.61 | 28,923.66 | 8.90 |
| 12 | 7,409.24 | 2.35 | 8,378.04 | 2.58 |
| 13 | 12,131.86 | 3.86 | 13,068.61 | 4.02 |
| 14 | 33,247.60 | 10.57 | 36,088.36 | 11.11 |
| 15 | 18,975.80 | 6.03 | 18,823.92 | 5.79 |
| 16 | 16,630.44 | 5.29 | 16,220.58 | 4.99 |
| 17 | 32,171.48 | 10.23 | 32,553.28 | 10.02 |
| 18 | 625.63 | 0.20 | 730.59 | 0.22 |
| 19 | 1,617.08 | 0.51 | 2,122.96 | 0.65 |
| 20 | 88,845.32 | 28.24 | 88,845.32 | 27.35 |
| total | 314,634.81 | 100.00 | 324,892.77 | 100.00 |



Plot of total peak Origins and Destinations to/from Study Area.



#### 3.3 Contents of Peak subarea files (*trip\_leg\_submatrix\_cusWDDP.csv*)

#### Trips by purpose, resident type (Subscriber\_Class) and time of day

* There are a total of 1,472,307 OD flows in this file which represent EE flows through the region (Not sure I fully flow this file - Need some explaination from Airsage)
* About 83% of these trips (1.2 Million flows) are made by resident pouplation and the remaining 17% of trip are made by the visitors.
* 10% of these trips are External-External through trips (ee). This is computed based on *SZCount* field in the data.
* The share of ee trips by purpose is consistent across all resident purposes and time of day.
* Most of the visitor trip purpose was Non-home based (NHB) as expected.

## [1] "Total trips = 1472307.68"

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subscriber\_Class | trips | ee | pct\_trips | pct\_ee | ee\_share |
| Resident | 1229302.0 | 127292.81 | 83.49491 | 83.05297 | 10.35489 |
| Visitor | 243005.7 | 25974.21 | 16.50509 | 16.94703 | 10.68872 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subscriber\_Class | Purpose | trips | ee | pct\_trips | pct\_ee | ee\_share |
| Resident | HBO | 379186.23 | 40185.010 | 25.7545508 | 26.2189545 | 10.597698 |
| Resident | HBW | 370170.93 | 30528.632 | 25.1422264 | 19.9185915 | 8.247172 |
| Resident | NHB | 479944.84 | 56579.171 | 32.5981347 | 36.9154243 | 11.788682 |
| Visitor | HBO | 51.36 | 41.595 | 0.0034884 | 0.0271389 | 80.987150 |
| Visitor | NHB | 242954.32 | 25932.612 | 16.5015997 | 16.9198907 | 10.673863 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Time\_of\_Day | trips | ee | pct\_trips | pct\_ee | ee\_share |
| H07:H10 | 674933.9 | 65511.15 | 45.8419 | 42.74315 | 9.706307 |
| H16:H19 | 797373.8 | 87755.87 | 54.1581 | 57.25685 | 11.005612 |

#### 3.4 Contents of Peak subarea files (*trip\_leg\_submatrix\_cusWDH.csv*)

#### Trips by purpose, and resident type (Subscriber\_Class)

* There are a total of 4,001,471 OD flows in this file which represent EE flows through the region (Not sure I fully flow this file - Need some explaination from Airsage)
* About 80% of these trips (2.9 Million flows) are made by resident pouplation and the remaining 20% of trip are made by the visitors.
* About 9% of these trips are External-External through trips (ee). This is computed based on *SZCount* field in the data.
* The share of ee trips by purpose is consistent across all resident purposes and time of day.
* Most of the visitor trip purpose was Non-home based (NHB) as expected.

## [1] "Total trips = 4001471.97"

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Subscriber\_Class | trips | ee | pct\_trips | pct\_ee | ee\_share |
| Resident | 3217870.5 | 298634.30 | 80.41717 | 80.1565 | 9.280495 |
| Visitor | 783601.5 | 73929.74 | 19.58283 | 19.8435 | 9.434609 |

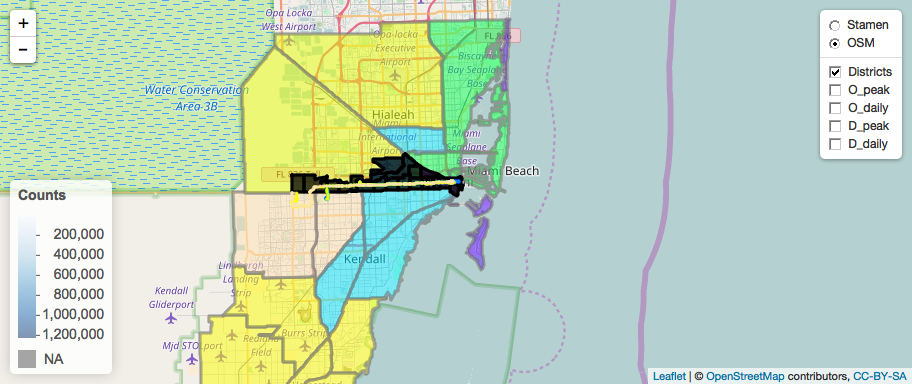
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subscriber\_Class | Purpose | trips | ee | pct\_trips | pct\_ee | ee\_share |
| Resident | HBO | 1154883.54 | 105315.88 | 28.8614677 | 28.2678597 | 9.119177 |
| Resident | HBW | 816881.18 | 52825.53 | 20.4145171 | 14.1789127 | 6.466734 |
| Resident | NHB | 1246105.76 | 140492.89 | 31.1411843 | 37.7097290 | 11.274556 |
| Visitor | HBO | 150.76 | 58.13 | 0.0037676 | 0.0156027 | 38.557973 |
| Visitor | NHB | 783450.73 | 73871.61 | 19.5790633 | 19.8278959 | 9.429005 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Time\_of\_Day | trips | ee | pct\_trips | pct\_ee | ee\_share |
| H00:H24 | 4001472 | 372564 | 100 | 100 | 9.310675 |

## 4. Review of ACS / CTPP JTW Flows

Add ACS/CTPP JTW data to the map and compute:

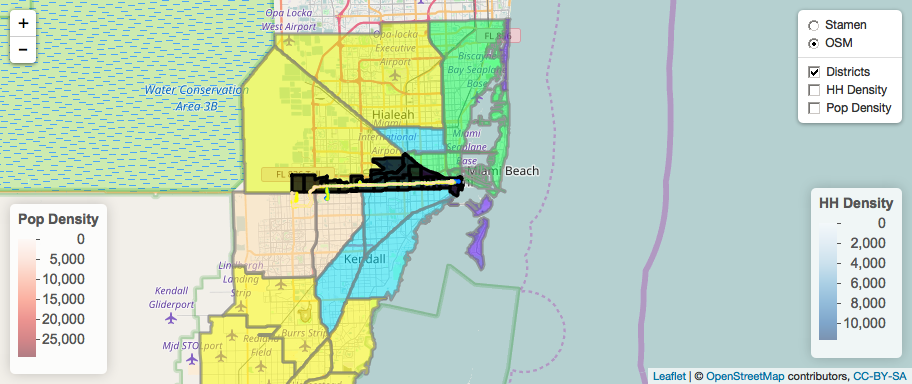
1. Samples origins and destinations by distrcits
2. Trips origins and destinations by area
3. Trips by prupose (table)
4. Trip flows by purpose (table, d3chord, flowmap)



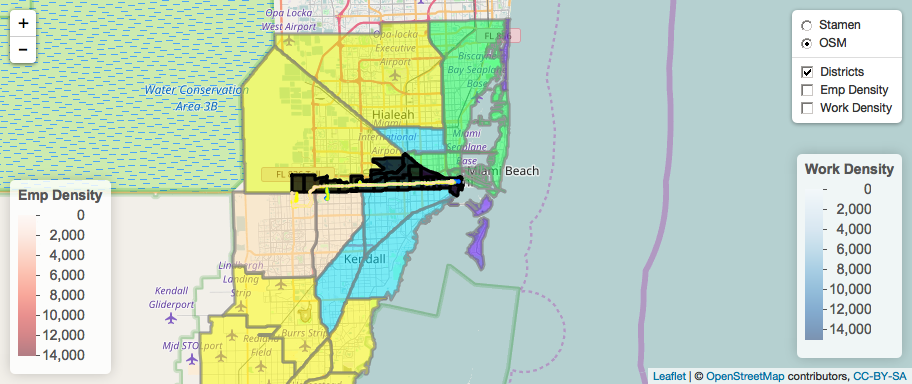
## 5. Review of Census Data

### Review of Household, Population, Worker and Employment data

The map below displays population, and household and employment densities.



The map below displays employment and worker densities. THe worker densities are the household workers which is different the employment data.



### 3. Read Miami Dade Zones

### 6. Review of APC Data

xlsFiles <- dir(apc\_dir, pattern = ".xls")  
  
# for (i in 1:length(xlsFiles)){  
# apc\_data <- read\_excel(paste0(projDir,"/",apc\_dir,"/",xlsFiles[i]))  
#   
# }  
  
# Get layer name from shape file  
layerName <- strsplit(strsplit(acs\_shapeFile,"[.]")[[1]],"/")[[1]][2]  
# Add census shape files  
acsShape <- readOGR(paste0(projDir,"/",acs\_shapeFile), layer = layerName, verbose = FALSE)  
  
map <- leaflet() %>%  
 # Base Map  
 addProviderTiles("Stamen.Toner",group = "Stamen") %>%  
 addTiles(group = "OSM") %>%  
 setView(lng = -80.1918, lat = 25.7617, zoom = 10) %>%  
 addTiles %>%  
 addPolygons(data = acsShape, group = "acs\_data")  
   
map

