RSG
the science of insight

Gen3 Model Development Project

Travel Forecasting Subcommittee Meeting

November 20, 2020

IN PARTNERSHIP WITH

BASELINE MOBILITY

Metropolitan Washington Council of Governments

Discussion Topics

- Synthetic population update
- Household travel survey data coding and visualization tool update

Synthetic Population Update

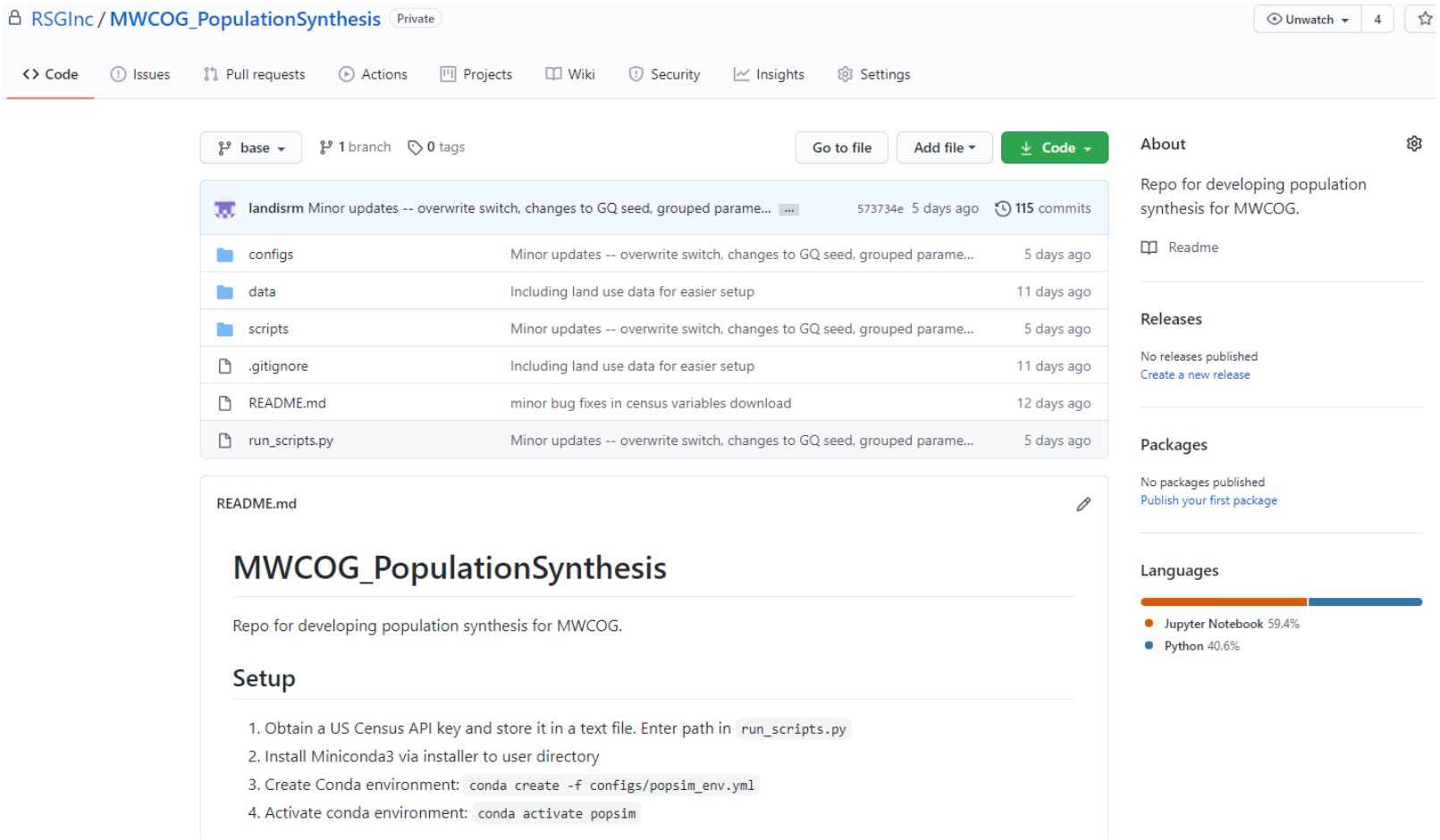
Refresher

- See last TFS meeting slide deck for explanation of
  - PopulationSim software and algorithm
  - Seed data
  - Marginal controls (revisited later in this presentation)

Implementation Features

- Python based implementation
- Fully automated Census data download
- Runs multiple years in a single run
- Generates residential and group quarters population
- Auto generation of validation charts and summaries

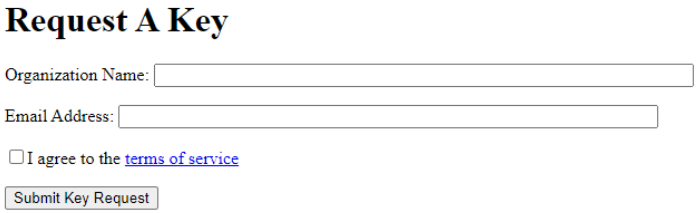
GitHub Repository



Software Requirements & Installation

- Get and install **Anaconda** 64bit Python 3
  - Pre-built collection of Python libraries (and underlying C/C++)

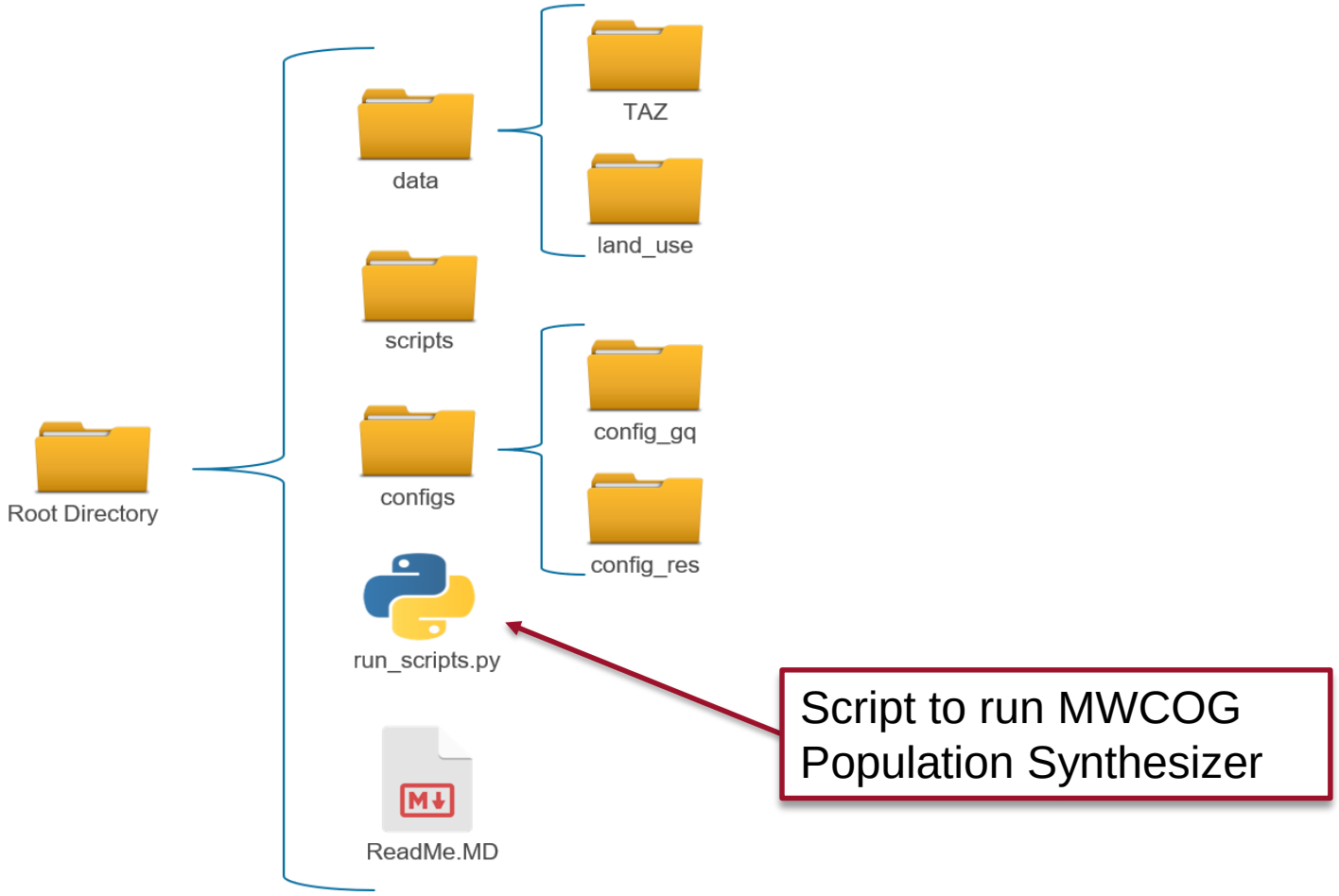
- Obtain a **US Census API Key**
  - For auto download of Census data



- Install **PopulationSim package** and dependencies
  - Custom PopulationSim environment

- Download and unzip **MWCOG Population Synthesizer Package**

Directory Setup



Run Steps

Data Download
- 01_get_pums.py
- 02_get_census_geography.py
- 03_get_census.py

Data preparation
- 04_create_crosswalk.py
- 05_create_seed_sample.py
- 06_create_controls.py

Run PopulationSim
- 07_run_populationsim.py – launches residential and GQ runs

Post-processing
- 08_postprocessing.py

Validation
- validation.py

Python-based implementation

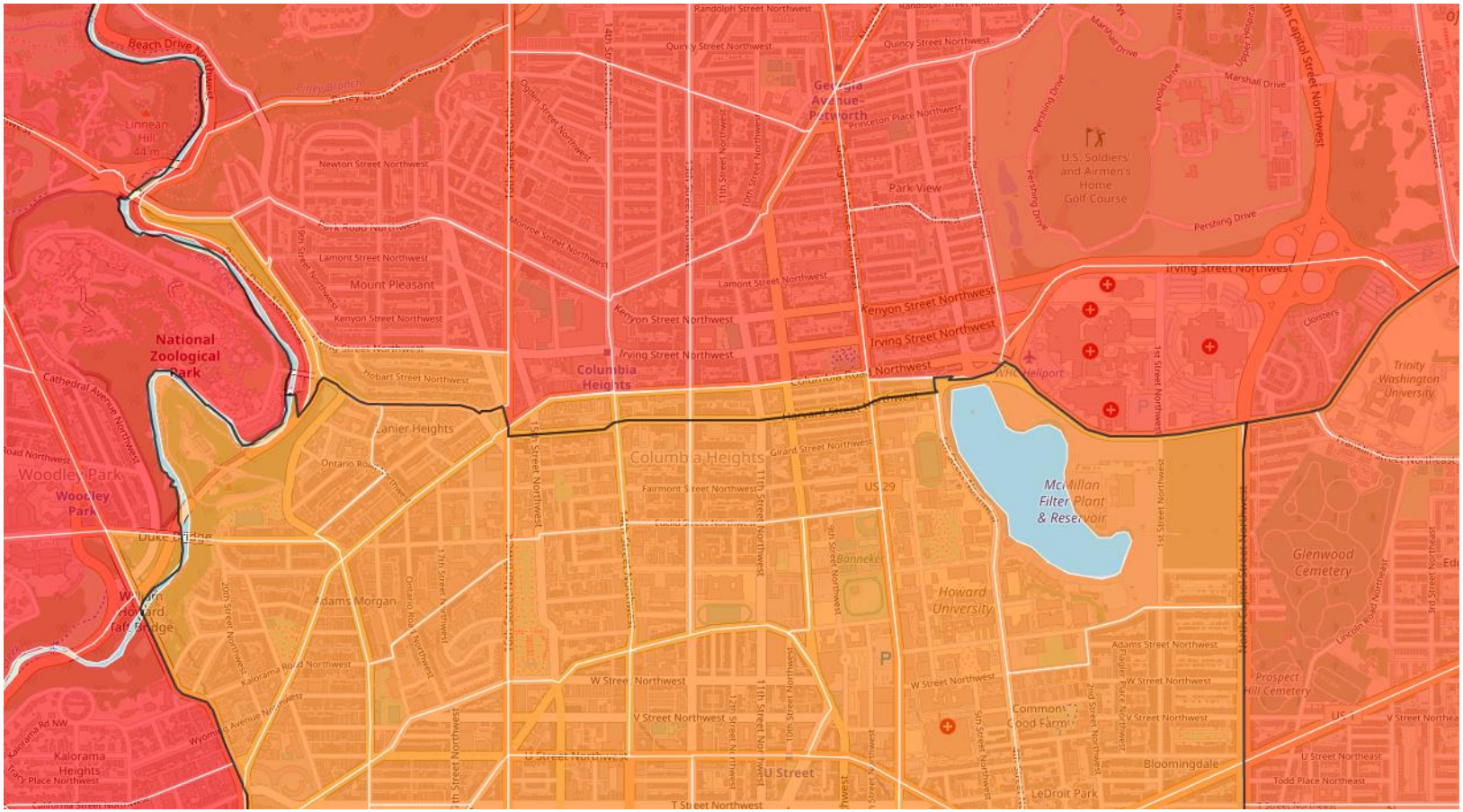
A Python script for each step

Data Preparation

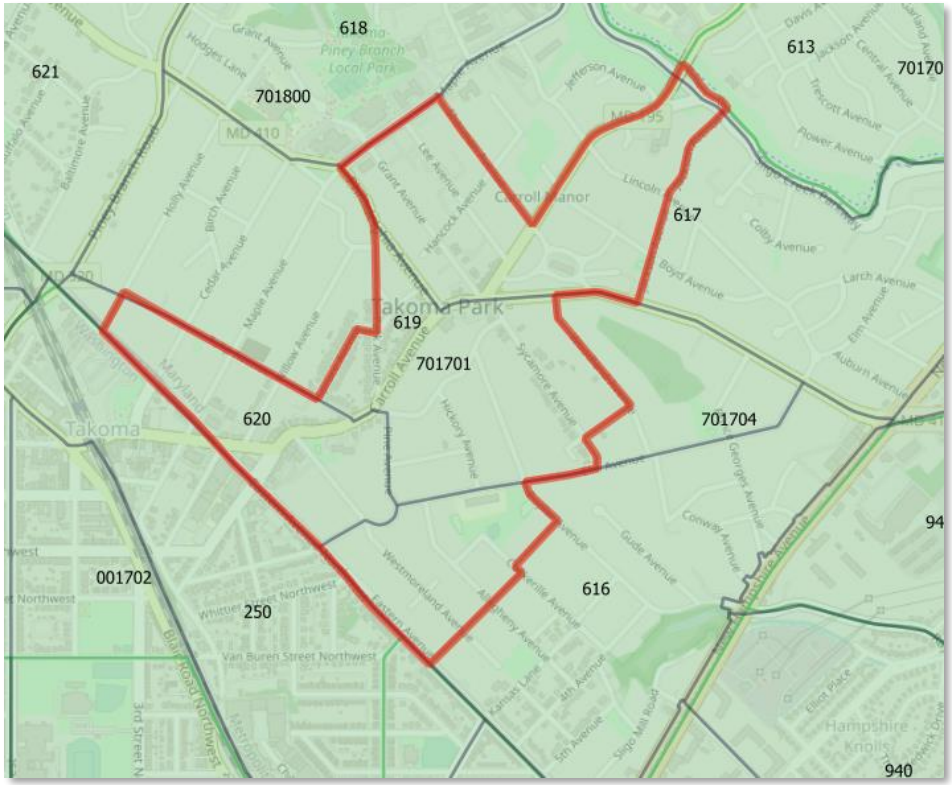
- Seed Sample
  - 2014-18 5-year ACS PUMS
  - Variable transformations
    - Adjust income to 2018 $
    - Number of workers
- Geographic crosswalk
  - PUMA, Tract, and TAZ shape files
  - Tract-to-TAZ, TAZ-to-PUMA
- Marginal Controls
  - Tract level Census distributions (2018 ACS)
  - Tract to TAZ disaggregation
  - Rescaling at TAZ level using Round 9.1a Forecasts

TAZ-PUMA Crosswalk

TAZs are assigned to PUMAs with maximum overlap



Census Distributions
Tract-to-TAZ Disaggregation



TRACTCE	TAZ	Fraction of Tract in TAZ
701701	250	0.00129
701701	613	0.00411
701701	616	0.15714
701701	617	0.31864
701701	618	0.00014
701701	619	0.31679
701701	620	0.20189

Tract to TAZ Disaggregation

Round 9.1a CF

Tracts to TAZs

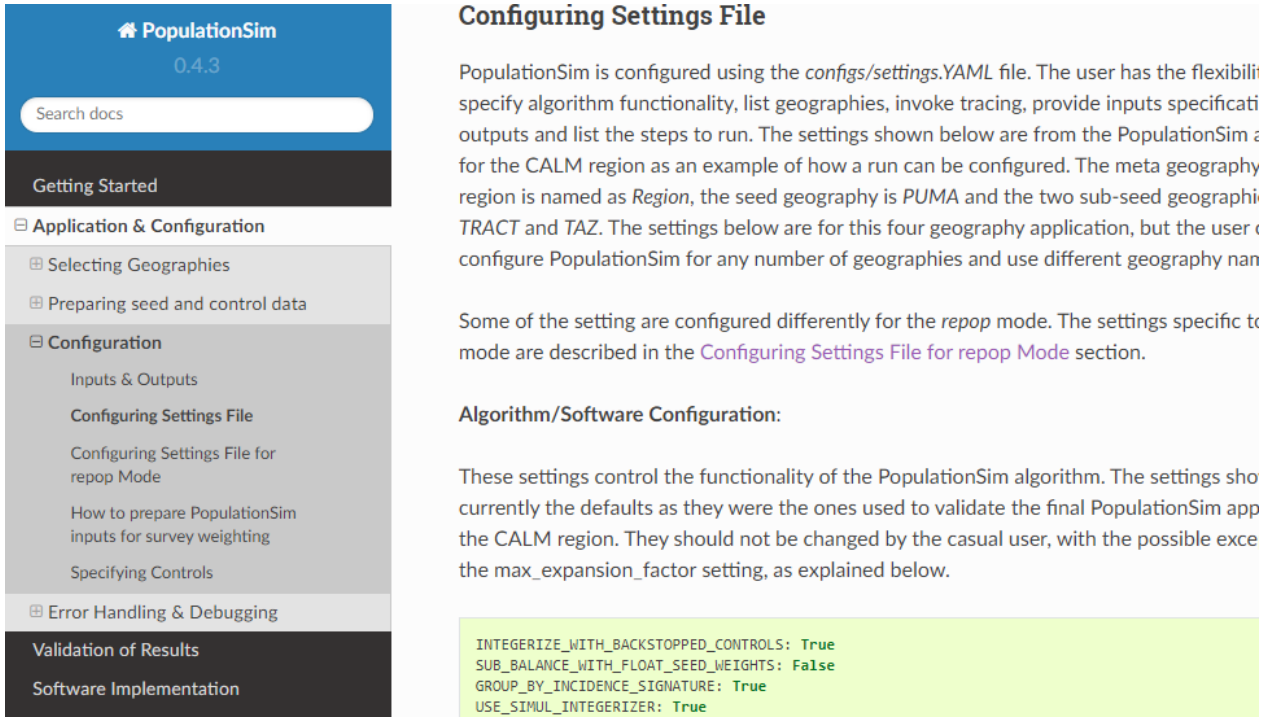
Sum TAZs

Rescale TAZs

TAZ-level Controls

Configuration – core PopulationSim

- Settings.YAML and controls.csv for residential and GQ runs
- Online wiki: https://activitysim.github.io/populationsim/



Configuration – run_scripts.py

User Inputs

```
# %% User parameters------------------------------------------------------------

# Years for which to synthesize data
sim_years = [2018, 2030, 2045]

# Should existing PopSim output be overwritten?
# Does not affect downloaded input data.
overwrite_output = False

# Enter the path to your census API key here.
census_key_path = os.path.join(os.getenv('HOMEDRIVE'), os.getenv('HOMEPATH'),
                               'Documents', 'census-api-key.txt')
```

Add a new forecast year

- Requires TAZ level land use data for the forecast year
- Add the land use data to the ***data/land_use*** directory
- Use same data and filename format
- Update ***sim_years*** in ***run_scripts.py*** to include the new forecast year

```
# Years for which to synthesize data
sim_years = [2018, 2030, 2045]
```

```
land_use/
    ├── LU_taz3722_rnd91a_2015_adj.dbf
    ├── LU_taz3722_rnd91a_2018_adj.dbf
    ├── LU_taz3722_rnd91a_2020_adj.dbf
    ├── LU_taz3722_rnd91a_2025_adj.dbf
    ├── LU_taz3722_rnd91a_2030_adj.dbf
    ├── LU_taz3722_rnd91a_2035_adj.dbf
    ├── LU_taz3722_rnd91a_2040_adj.dbf
    └── LU_taz3722_rnd91a_2045_adj.dbf
```

Input Data Updates

- To select a different vintage of PUMS data
  - Configure the following settings in run_scripts.py

```
# PUMS data download settings
pums_dir = os.path.join(data_dir, 'PUMS') # Path to save the PUMS data
pums_year = 2018
pums_period = '5-Year'  # '5-Year' or '1-Year'
```

- To select a different source of control data
  - Configure configs/census_variables_needed.csv file

- To change TAZ level HH totals or population
  - Update the land use data for the forecast year in the data/land_use directory

Example – changing HH Income Thresholds

- New thresholds should be consistent with Census income groups
- Update the 06_create_controls.py script to generate appropriate marginal control fields
- Update the expressions and control_field names in the configs/controls.csv

Census income groups
Less than $10,000
$10,000 to $14,999
$15,000 to $19,999
$20,000 to $24,999
$25,000 to $29,999
$30,000 to $34,999
$35,000 to $39,999
$40,000 to $44,999
$45,000 to $49,999
$50,000 to $59,999
$60,000 to $74,999
$75,000 to $99,999
$100,000 to $124,999
$125,000 to $149,999
$150,000 to $199,999
$200,000 or more

Controls.csv file

target	geography	seed_table	importance	control_field	expression
hh_inc_0_25	TAZ	households	1000	income_0to25	(households.hhincadj > -999999999) & (households.hhincadj <= 25000)
hh_inc_25_50	TAZ	households	1000	income_25to50	(households.hhincadj >25000) & (households.hhincadj <= 50000)
hh_inc_50_100	TAZ	households	1000	income_50to100	(households.hhincadj > 50000) & (households.hhincadj <= 100000)
hh_inc_100_150	TAZ	households	1000	income_100to150	(households.hhincadj > 100000) & (households.hhincadj <= 150000)
hh_inc_150_200	TAZ	households	1000	income_150to200	(households.hhincadj > 150000) & (households.hhincadj <= 200000)
hh_inc_200_plus	TAZ	households	1000	income_200plus	(households.hhincadj > 200000) & (households.hhincadj <= 999999999)

PopulationSim wiki instructions on specifying controls
https://activitysim.github.io/populationsim/application_configuration.html#specifying-controls

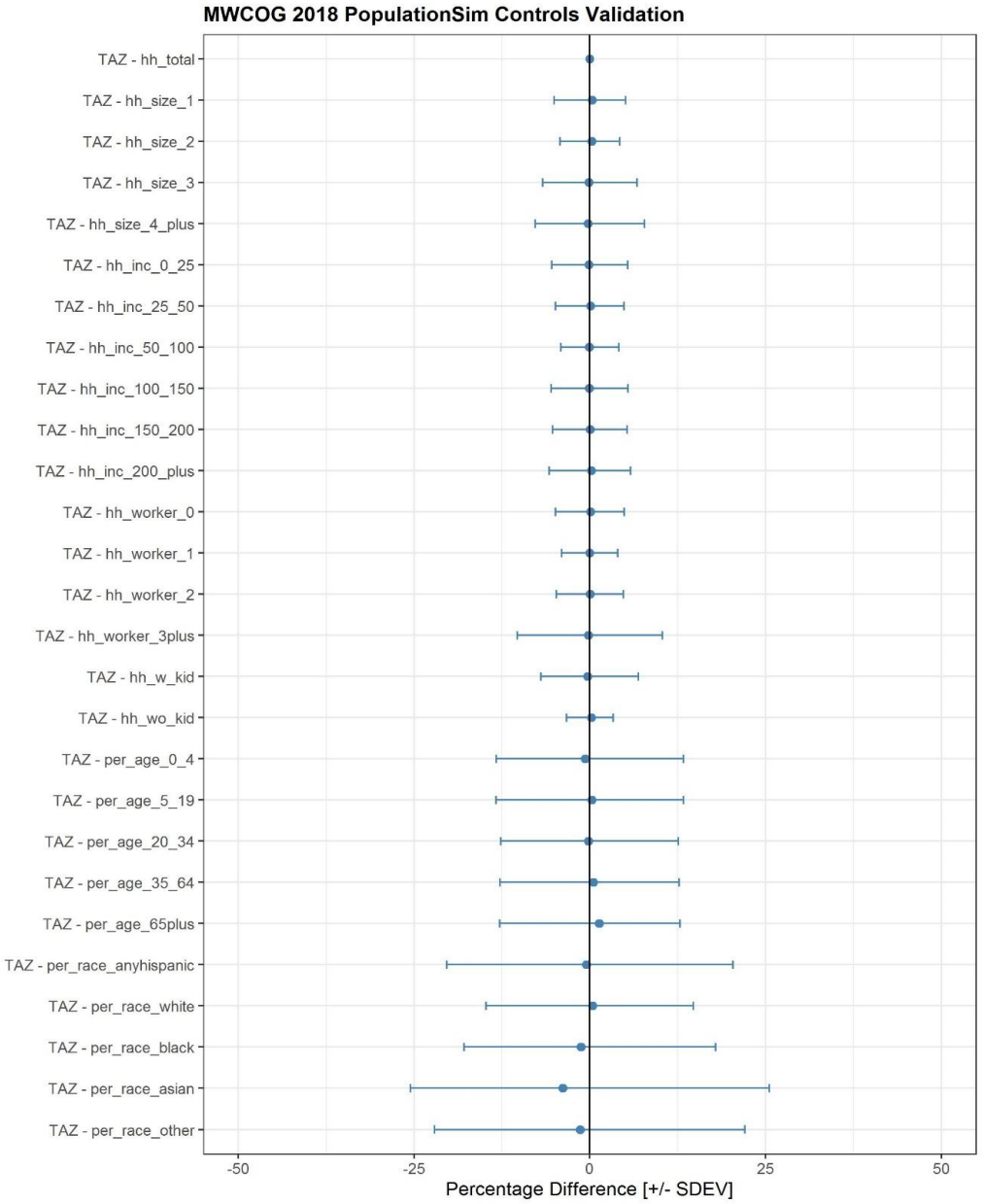
MWCOG PopulationSim Controls

CONTROL VARIABLE	CATEGORIES	CONTROL SOURCE	GEOG
Household Variables			
Total number of households		Round 9.1 Cooperative Forecast	TAZ
Household Size	1, 2, 3, 4+	2018 ACS 5-year. Census Tract [Table S2501]	TAZ
Household Income	0-$25K, $25K-$50K, $50K-$100K, $100k-$150K, $150K-$200K, $200K+	2018 ACS 5-year. Census Tract [Table B19001]	TAZ
Number of Workers	0, 1, 2, 3+	2018 ACS 5-year. Census Tract [Table B08202]	TAZ
Presence of Children	0, 1	2018 ACS 5-year. Census Tract [Table S1101]	TAZ
Person Variables			
Person Age	0-4, 5-19, 20-34, 35-64, 65+	2018 ACS 5-year. Census Tract [Table S0101]	TAZ
Person Race	White, Hispanic, Black, Asian, Other	2018 ACS 5-year. Census Tract [Table DP05]	TAZ

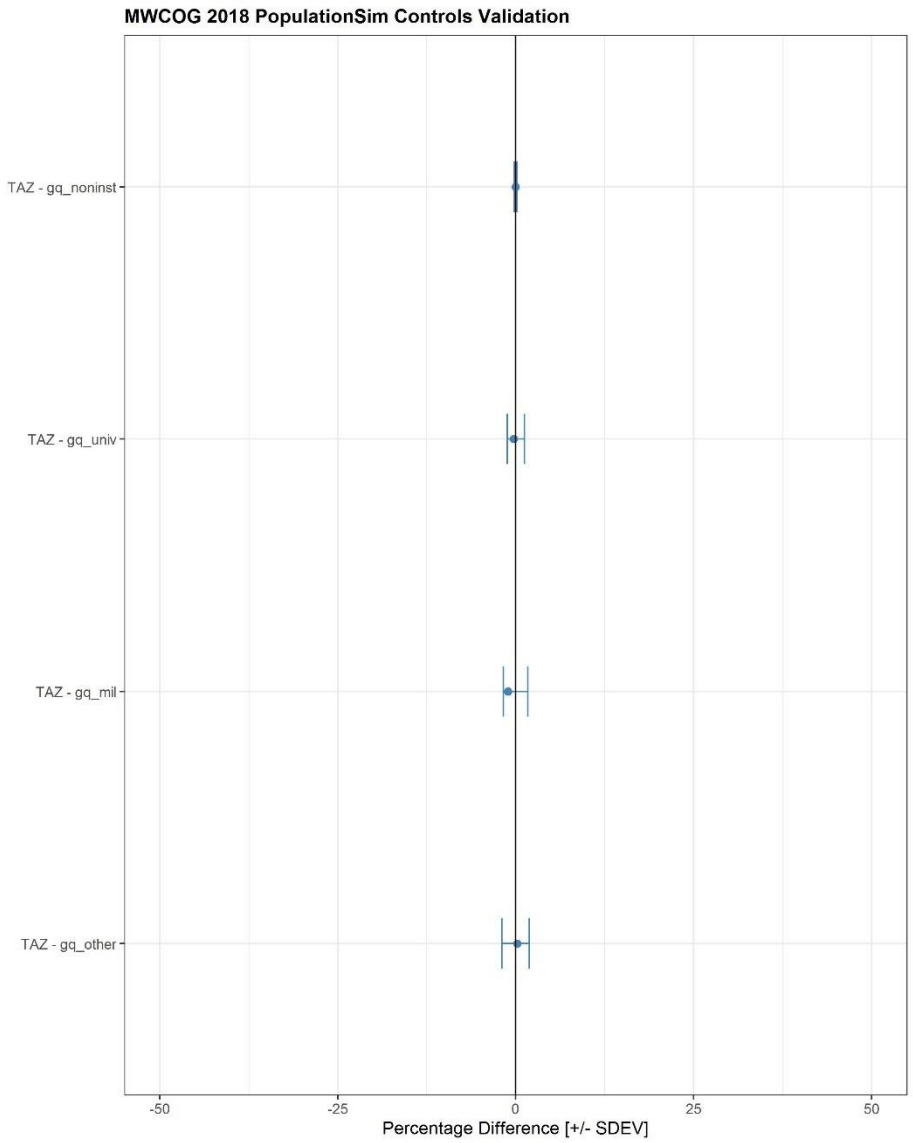
GQ Controls

CONTROL VARIABLE	CATEGORIES	SOURCE	GEOG
Total GQ units		Round 9.1 Cooperative Forecast, and 2010 SF1	TAZ
GQ Type	University, Military, Other Non-Institutional	Round 9.1 Cooperative Forecast, and 2010 SF1	TAZ

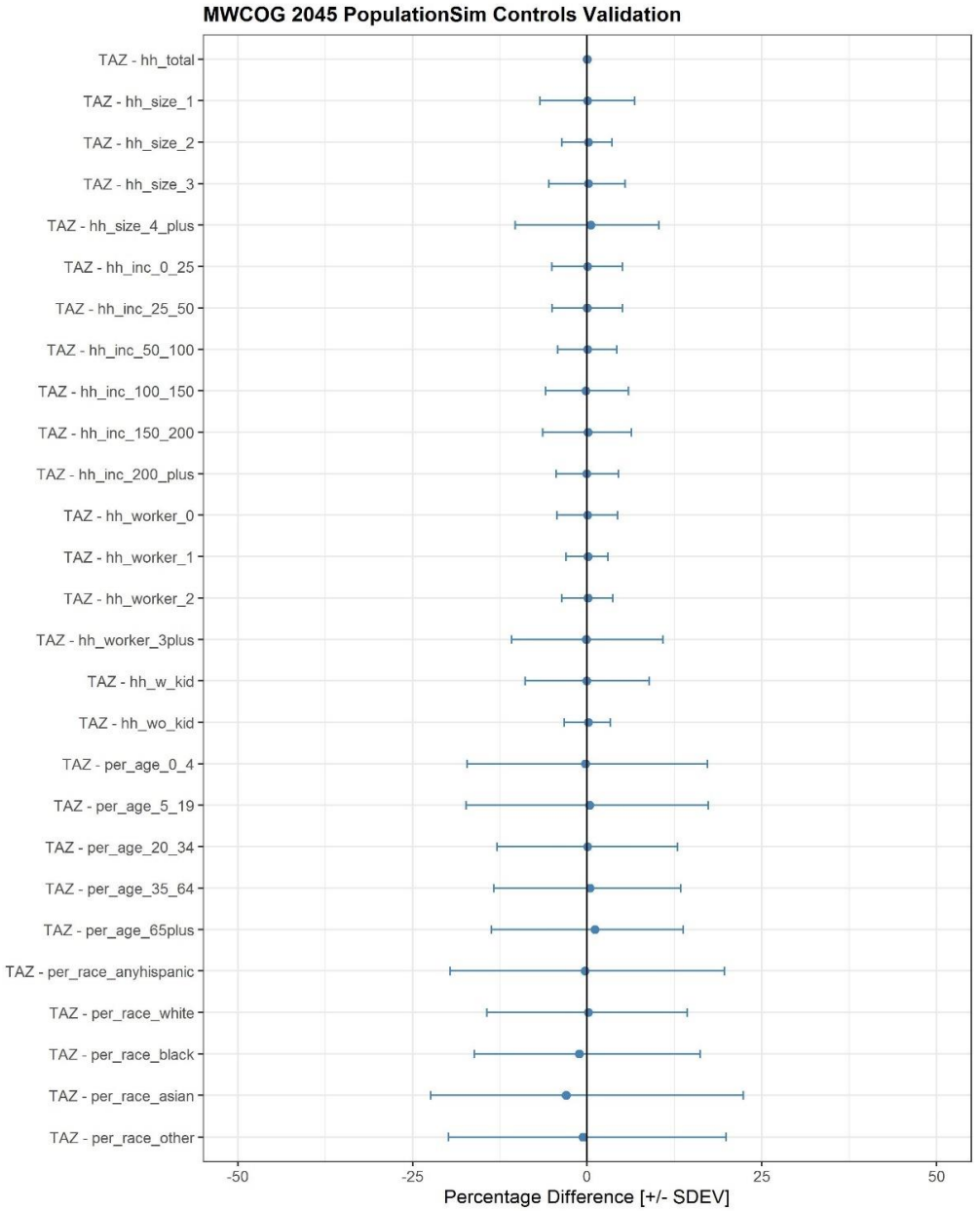
2018 Residential



2018 GQ


MWCOG 2018 PopulationSim Controls Validation
TAZ - gq_noninst
TAZ - gq_univ
TAZ - gq_mil
TAZ - gq_other
-50
-25
0
25
50
Percentage Difference [+/- SDEV]

2045 Residential



Documentation

RSG
the science of insight

METROPOLITAN WASHINGTON COUNCIL OF GOVERNMENTS (MWCOG)

MWCOG POPULATION SYNTHESIZER

Final Report | November 18, 2020

PREPARED FOR:
METROPOLITAN WASHINGTON COUNCIL OF GOVERNMENTS (MWCOG)

55 Railroad Row
White River Junction, VT 05001
802.295.4999
www.rsginc.com

SUBMITTED BY:
RSG

CONTENTS

LIST OF FIGURES

Household Survey Coding Update

Household Survey Data Processing Overview

- Combined Regional Travel Survey/Maryland Travel Survey
- Sample of 18,022 households containing 39,251 persons
- Includes 126,900 trips
- Survey Processing Application transforms data into *ActivitySim* format

ActivitySim Data Format

Households:

- Size
- Income
- Workers
- Vehicles
- Number of children
- Home TAZ

Persons:

- Age
- Employment status
- Student Status
- Work TAZ
- School TAZ
- Person Type

Tours:

- Purpose
- Mode
- Type
- Origin TAZ
- Destination TAZ
- Start Time
- End Time
- Joint Status

Joint Tours:

- Household members on tour

Trips:

- Purpose
- Mode
- Origin TAZ
- Destination TAZ
- Start Time
- End Time

Survey Processing Requires:

- Person Type Coding
- Grouping trips into tours
- Mode and purposes to match those in ActivitySim

Generated by Survey Processing Application

Person Type Coding

Output	Input Condition		
Person type [PERSONTYPE]	AGE [AGE]	Employment Category [EMP_CAT]	Student Category [STU_CAT]
1: Full-time worker	>=16	Full-time (1)	any
2: Part-time worker	>=16	Part-time (2)	Not attending (3)
3: University student	>=17	Not full-time (2,3)	College+ (2)
4: Non-worker	>=16 and <=64	No (3)	No (3)
5: Retired	>=65	No (3)	No (3)
6: Student of driving age	>=16 and <=19	Not full-time 2,3	K-12 (1)
7: Student of non-driving age	>=6 and <=15	Not in labor force (4)	K-12 (1)
8: Child too young for school	>=0 and <=5	Not in labor force (4)	K-12, Not attending (1,3)

No part-time worker information in Household Travel Survey!

- Neither work hours per week or part-time work status was asked
- Part-time work status needs to be imputed

Part Time Status Imputation

SEMCOG Part-Time Status Imputation Model

- Trained Machine Learning (Random Classification) model on 2005 dataset
- Household, person, activity, travel variables
- Applied the 2005 model to impute part-time status in 2015 dataset

MWCOG Part-Time Status Imputation Steps:

- Process relevant variables in the format required for SEMCOG model
- Deploy SEMCOG model to workers in MWCOG RTS

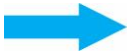
Survey Processing Application (SPA)

The SPA tool cleans and processes HTS's into CT-RAMP format

SPA Algorithm:

- For each observed household
- Create a HOUSEHOLD object
- For each observed person
  - Create a PERSON object for the HOUSEHOLD
  - Compute person type
  - For each series of observed place records that start and end at home

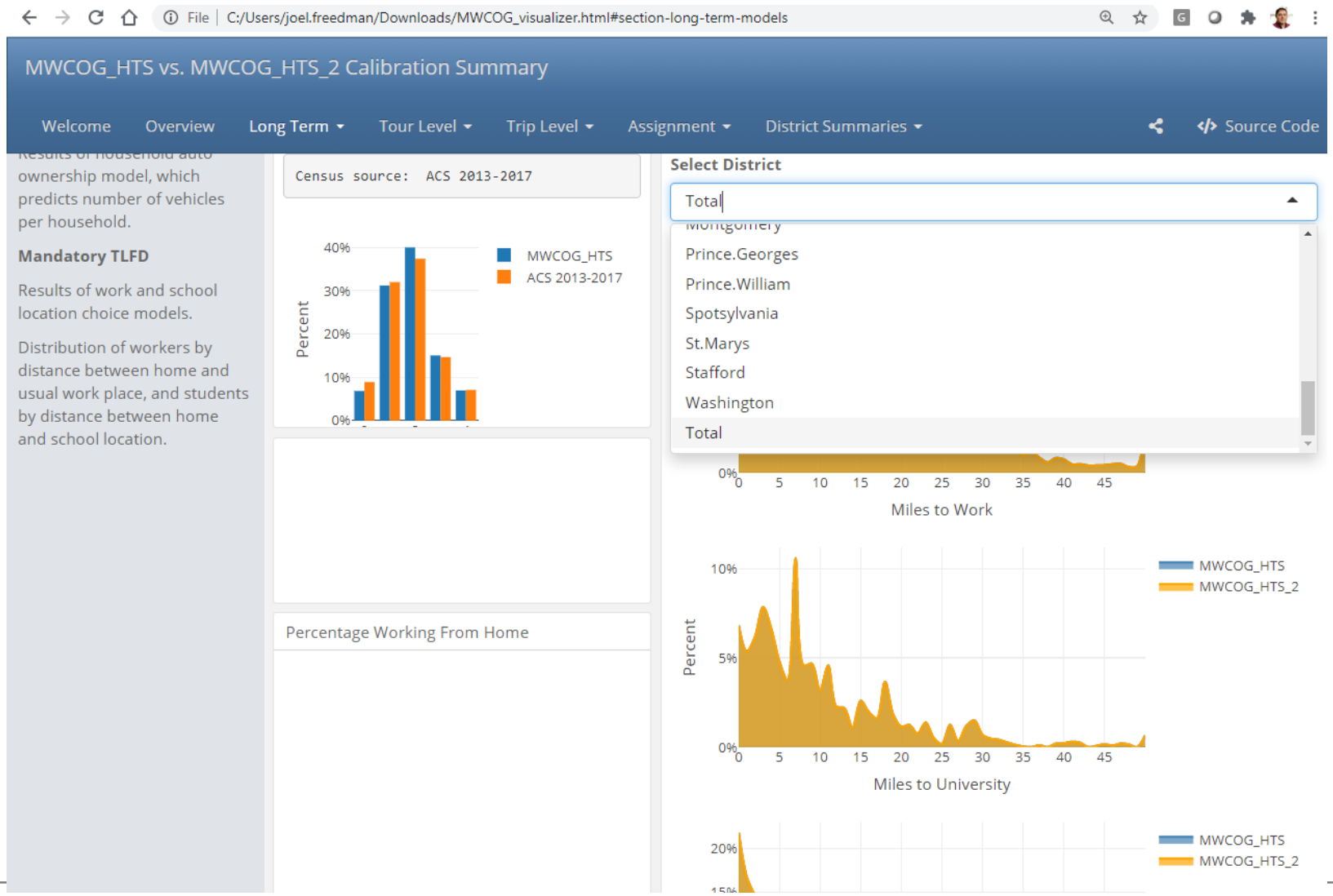
    - Create a TOUR object for the PERSON
    - For each series of observed place records that make up a linked trip within the tour
      - Create and attribute a TRIP object
      - For each leg of the trip involving joint travel
        - Create a JOINT_ULTRIP object for the HOUSEHOLD



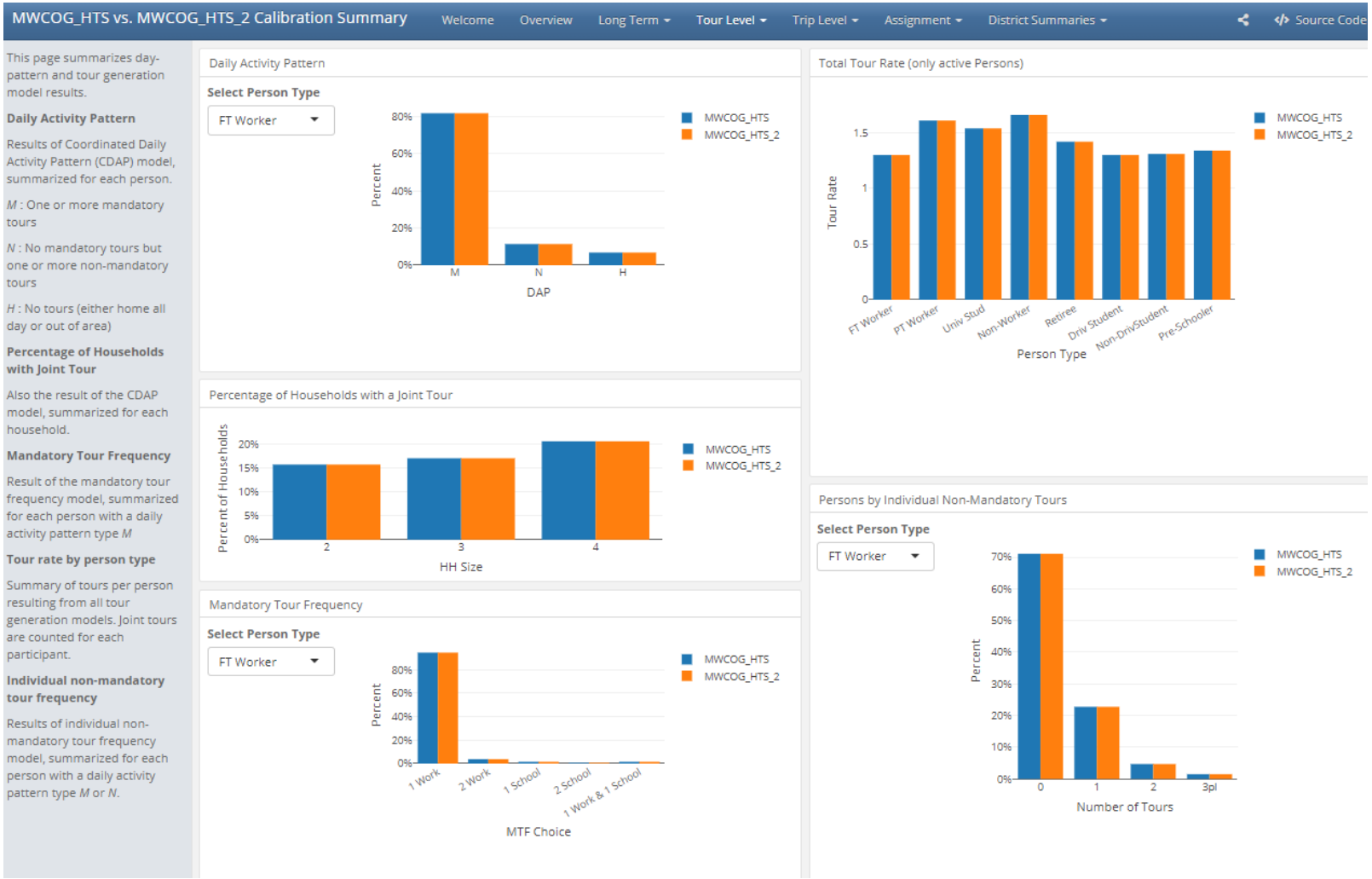
  - Populate tour attributes
  - Match JOINT_ULTRIP objects into joint travel episodes

- Compute escort related attributes for TRIPs
- Create and attribute a JOINT_TOUR object for each fully joint tour
- Compute escort related attributes for partially joint tours

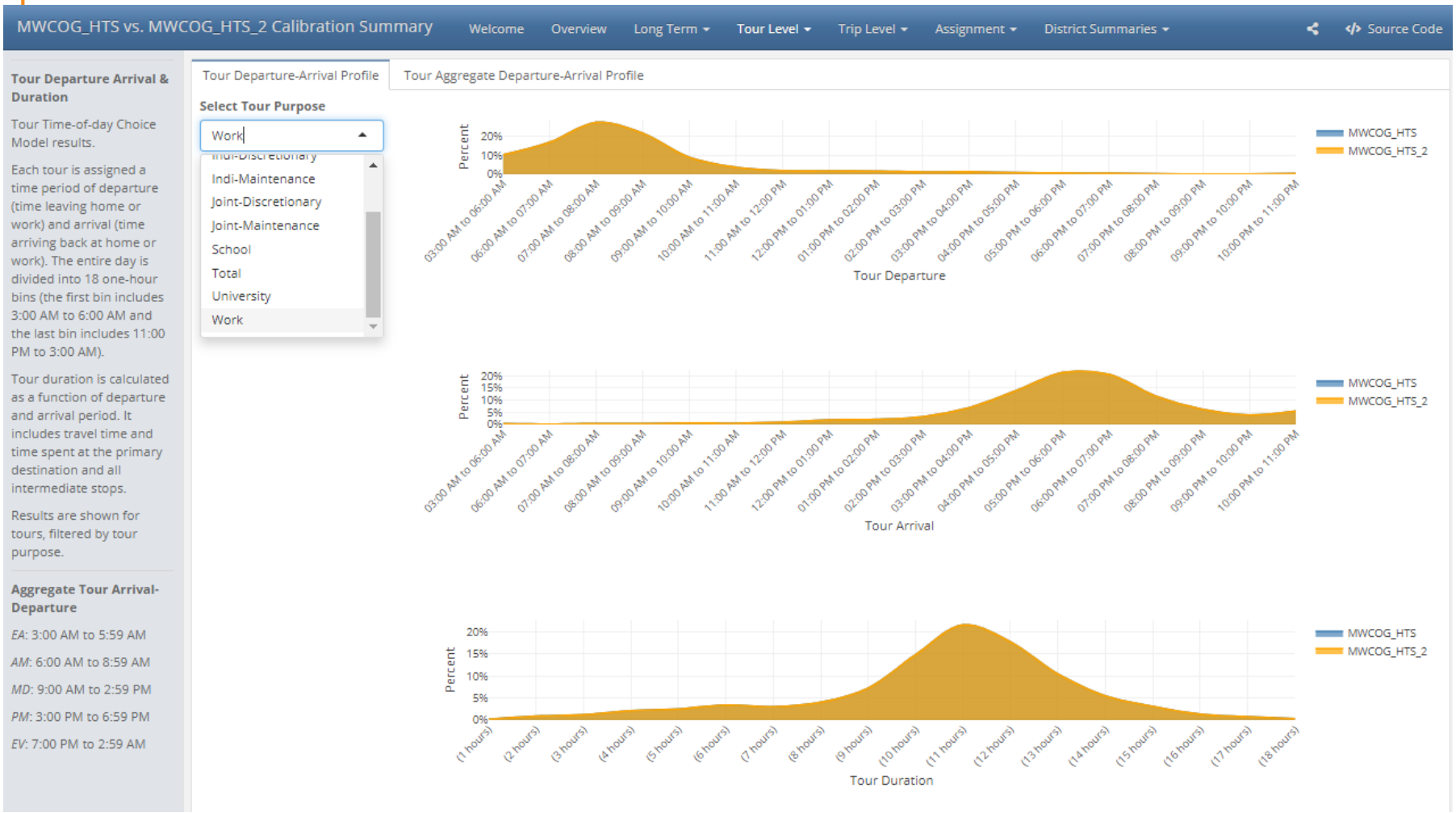
Visualizer


File | C:/Users/joel.freedman/Downloads/MWCOG_visualizer.html#section-long-term-models
MWCOG_HTS vs. MWCOG_HTS_2 Calibration Summary
Welcome
Overview
Long Term
Tour Level
Trip Level
Assignment
District Summaries
Source Code
ownership model, which predicts number of vehicles per household.
Mandatory TLFD
Results of work and school location choice models.
Distribution of workers by distance between home and usual work place, and students by distance between home and school location.
Census source: ACS 2013-2017
Percent
MWCOG_HTS
ACS 2013-2017
Percentage Working From Home
Select District
Total
Prince.Georges
Prince.William
Spotsylvania
St.Marys
Stafford
Washington
Total
Miles to Work
Miles to University
MWCOG_HTS
MWCOG_HTS_2

Visualizer

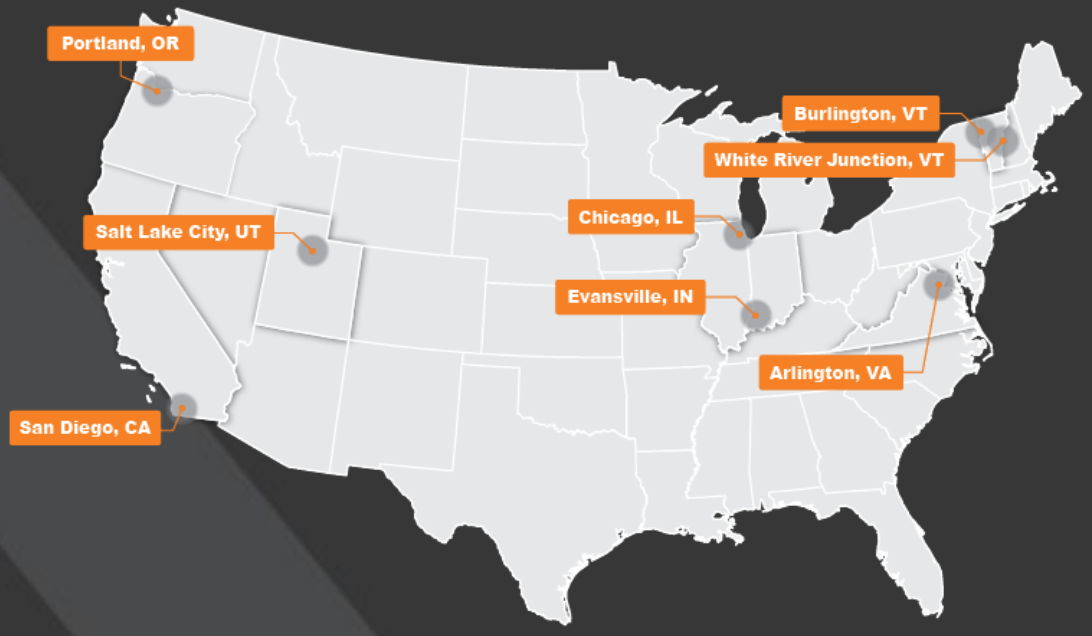

MWCOG_HTS vs. MWCOG_HTS_2 Calibration Summary
Welcome
Overview
Long Term
Tour Level
Trip Level
Assignment
District Summaries
Source Code
This page summarizes day-pattern and tour generation model results.
Daily Activity Pattern
Results of Coordinated Daily Activity Pattern (CDAP) model, summarized for each person.
M : One or more mandatory tours
N : No mandatory tours but one or more non-mandatory tours
H : No tours (either home all day or out of area)
Percentage of Households with Joint Tour
Also the result of the CDAP model, summarized for each household.
Mandatory Tour Frequency
Result of the mandatory tour frequency model, summarized for each person with a daily activity pattern type M
Tour rate by person type
Summary of tours per person resulting from all tour generation models. Joint tours are counted for each participant.
Individual non-mandatory tour frequency
Results of individual non-mandatory tour frequency model, summarized for each person with a daily activity pattern type M or N.
Daily Activity Pattern
Select Person Type
FT Worker
MWCOG_HTS
MWCOG_HTS_2
Percent
DAP
Total Tour Rate (only active Persons)
Tour Rate
Person Type
Percentage of Households with a Joint Tour
Percent of Households
HH Size
Mandatory Tour Frequency
MTF Choice
Persons by Individual Non-Mandatory Tours
Number of Tours

Visualizer


MWCOG_HTS vs. MWCOG_HTS_2 Calibration Summary
Welcome
Overview
Long Term
Tour Level
Trip Level
Assignment
District Summaries
Source Code
Tour Departure Arrival & Duration
Tour Time-of-day Choice Model results.
Each tour is assigned a time period of departure (time leaving home or work) and arrival (time arriving back at home or work). The entire day is divided into 18 one-hour bins (the first bin includes 3:00 AM to 6:00 AM and the last bin includes 11:00 PM to 3:00 AM).
Tour duration is calculated as a function of departure and arrival period. It includes travel time and time spent at the primary destination and all intermediate stops.
Results are shown for tours, filtered by tour purpose.
Aggregate Tour Arrival-Departure
EA: 3:00 AM to 5:59 AM
AM: 6:00 AM to 8:59 AM
MD: 9:00 AM to 2:59 PM
PM: 3:00 PM to 6:59 PM
EV: 7:00 PM to 2:59 AM
Tour Departure-Arrival Profile
Tour Aggregate Departure-Arrival Profile
Select Tour Purpose
Work
Indi-Maintenance
Joint-Discretionary
Joint-Maintenance
School
Total
University
Work
Percent
Tour Departure
Tour Arrival
Tour Duration
MWCOG_HTS
MWCOG_HTS_2

Immediate Next/Ongoing Activities

- Household survey re-expansion
  - With MTS data
- Transit on-board data coding
  - Internal/external trip tables
- ActivitySim implementation



RSG
the science of insight

Joel Freedman
SENIOR DIRECTOR
Joel.Freedman@rsginc.com