2019-winter-transportation-survey\Analysis\Scripts\Describe-scripts.docx

Description of scripts

* **clean\_initial.R, clean\_diary12.R, clean\_diary3.R, clean\_diary4.R, clean\_diary5.R, clean\_diary6.R, clean\_final.R**
  + *Summary*: Process and clean raw CSV data for each survey, with light editing
  + *Inputs*: Raw CSV data for each survey (Data 0 Raw)
  + *Steps*: Format columns and data types. Change column names. Convert to time (POSIXct) formats. Create factors with named levels. Process other text responses into defined levels. Create dummy variables from checkbox columns. Change DATE format in diaries. Check HH ids and PER names match. Fix wrong HH ids and PER names. Make other minor edits.
  + *Outputs*: Separate CSV and RDS files (and CSV files for column names) for initial survey, travel diary survey days 1 & 2, 3, 4, 5, 6, final survey (Data 2 Cleaned)
* **format\_initial.R, format\_diary.R, format\_final.R**
  + *Summary*: Format data and prepare data for use in analyses
  + *Inputs*: Separate RDS files for initial survey, travel diary survey days 1 & 2, 3, 4, 5, 6, final survey (Data 2 Cleaned)
  + *Steps*: Split columns into different groups of data: metadata, contact data, and others (Initial: HH, PER, VEH; Diary: DIARY, PLACE, TRIP; Final: PER\_final). Create new columns (variations on existing columns: collapsing categories, making numeric). Fix any missing data or errors that are noticed. Fix missing or wrong DATE in diaries. Fix missing or wrong PLACE, ACTIVITY, MODE, times in diaries. Edit records based on comments in diaries. Note diaries to remove based on: duplicate, incomplete, missing. Create (or add) anonymous HH code, PER code, VEH code. Add info about # PER and # VEH to HH.
  + *Outputs*: Separate CSV and RDS files for households initial, persons initial, vehicles initial, persons final, diary, place, trip (Data 3a Formatted)
* **Geocoding: geo\_initial.R, geo\_diary.R, geo\_diary\_addl.R, geo\_trip.R**
  + *Summary*: Geocode locations and trip information from initial survey, travel diary surveys
  + *Inputs*: Separate RDS files for households initial, persons initial, places, geocoded places (Data 3a Formatted)
  + *Steps*: Inspect data. Make any necessary edits to home, work, school, place locations. Geocode locations and trips using Google API and ggmap package.
  + *Outputs*: Separate RDS files for geocoded home, work, school locations, places, trips (Data 3a Formatted)
* **finish\_initial.R, finish\_diary.R**
  + *Summary*: Add geocoded location information to initial survey data, travel diary survey data
  + *Inputs*: Separate RDS files for households initial, persons initial, diary, place, trip; and separate RDS files for home, work, school locations (Data 3a Formatted)
  + *Steps*: Add Census block group (GEOID) to geocoded home, work, school, place, trip locations. Add geocoded location information to initial survey data, travel diary survey data.
  + *Outputs*: Separate CSV and RDS files for households initial, persons initial, place, trip (Data 3b Finished)
* **anonymize.R**
  + *Summary*: Anonymize data by removing personally identifiable information
  + *Inputs*: Separate RDS files for households initial, persons initial, place, trip (Data 3b Finished); and separate RDS files for vehicles initial, persons final, diary (Data 3a Formatted)
  + *Steps*: Remove identifying metadata, contact information, names, locations, etc. Remove some rows (duplicates, withdrawn).
  + *Outputs*: Separate CSV and RDS files for households initial, persons initial, vehicles initial, persons final, diary, place, trip (Data 4 Anonymized)