# Merge Stops to Shift Assignments

## Changes to the data - Risi

### Stops

1. I decided to use the **contact_type** variable for categorizing stop types rather than **stop_type** in line with Ba et al. 2021.
2. **civ.race**: I combine Asian American/Pacific Islanders and Native Americans/Alaskan Natives into their own category.
3. I drop the **civilian_race_short** column.
4. **civ.gender**: Any civilians whose gender is categorized as **X** gets recoded as missing. Also I rename the column to be **civ.sex**.

| n | prcnt | name | value |
|---|---|---|---|
| 1124118 | 0.6600198 | civ.race | BLACK |
| 346558 | 0.2034797 | civ.race | HISPANIC |
| 210545 | 0.1236204 | civ.race | WHITE |
| 21937 | 0.0128802 | civ.race | aapi_native |

| n | prcnt | name | value |
|---|---|---|---|
| 1322801 | 0.7766754 | civ.sex | M |
| 379401 | 0.2227632 | civ.sex | F |
| 956 | 0.0005613 | civ.sex | NA |

## Merge process on Risi data using Ba et al. 2021 merging method

### Merge stops - Risi using Ba et al. 2021

- **Number of resulting rows**: 4203693
- **Number of stops which occurred during a shift**: 1442337

    - **Percentage of stops** (1703158): 84.6860362%

- **Number of shift assignments retained:** 3413954

– **Percentage of shift assignments**: 99.874906%

- 0 shift assignments matched with at least one stop **without having** a start time and an end time.

  – This represents 0% of all shift assignments.
  – 0 stops are affected or 0% of all stops.

- There are 71 stops which match to more than one shift assignment.

  – These **duplicate** stops represent 0.0041687% of all stops (1703158).
  – 80 shift assignments are affected or 0.0023404% of all shift assignments.
  – In total, there are 142 **duplicate stop-shift observations** representing 0.003378% of all stop-shift observations.

## Merge stops - Risi

- **Number of resulting rows**: 4203139
- **Number of stops which occurred during a shift**: 1435503

  – **Percentage of stops** (1703158): 84.2847816%
  – **Number of fewer stops matched than Ba et al. 2021**: 6834
  – **Percent reduction from Ba et al. 2021**: 0.4738144%

- **Number of shift assignments retained:** 3418230

  – **Percentage of shift assignments**: 100% **NOTE** this should be 100%.

- 0 shift assignments matched with at least one stop **without having** a start time and an end time. **NOTE** that this should be 0.
- There are 59 stops which match to more than one shift assignment.

  – These **duplicate** stops represent 0.0034642% of all stops (1703158).
  – 64 shift assignments are affected or 0.0018723% of all shift assignments.
  – In total, there are 118 **duplicate stop-shift observations** representing 0.0028074% of all stop-shift observations.
  – As noted above, **all these entries are dropped**.