Merge Arrests to Shift Assignments

Arrests

1. I drop the **statute_description** column because there are too many unique values for it to be of use.
2. **civ.race**: I combine Asian American/Pacific Islanders and Native Americans/Alaskan Natives into an their own category.
3. I drop the **civilian_race_short** column.
4. I rename **civ.gender** to **civ.sex**.

n	prcnt	name	value
214048	0.6650097	civ.race	BLACK
70132	0.2178879	civ.race	HISPANIC
34528	0.1072725	civ.race	WHITE
2494	0.0077484	civ.race	aapi_native
670	0.0020816	civ.race	NA

Merge arrests - Risi using Ba et al. 2021

- **Number of resulting rows**: 3450072
- **Number of arrests which occurred during a shift**: 264032

    - **Percentage of arrests** (321872): 82.0301238%

- **Number of shift assignments retained:** 3417221

    - **Percentage of shift assignments**: 99.9704818%

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

    - This represents 0% of all shift assignments.
    - 0 arrests are affected or 0% of all arrests.

- There are 10 arrests which match to more than one shift assignment.

- These **duplicate** arrests represent 0.0031068% of all arrests (321872).
- 20 shift assignments are affected or 5.8509814×10^{-4}% of all shift assignments.
- In total, there are 20 **duplicate stop-shift observations** representing 5.7969805×10^{-4}% of all arrest-shift observations.

Merge arrests - Risi

- **Number of resulting arrests**: 3450822
- **Number of arrests which occurred during a shift**: 262558

  - **Percentage of arrests** (321872): 81.5721778%
  - **Number of fewer arrests matched than Ba et al. 2021**: 1474
  - **Percent reduction from Ba et al. 2021**: 0.5582657%

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

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

- 0 shift assignments matched with at least one arrest **without having** a start time and an end time. **NOTE** that this should be 0.
- There are 10 arrests which match to more than one shift assignment.

  - These **duplicate** arrests represent 0.0031068% of all arrests (321872).
  - 20 shift assignments are affected or 5.8509814×10^{-4}% of all shift assignments.
  - In total, there are 20 **duplicate arrest-shift observations** representing 5.7957206×10^{-4}% of all arrest-shift observations.
  - As noted above, **all these entries are dropped**.