Replicate Ba et al. 2021 - Merging Stops to Officer Assignments

Merge process - Ba et al. 2021

"For stops, arrests, and uses of force, we drop all events that occur outside of the reported patrol start/stop times, eliminating off-duty activity." page 8 of Appendix Section S1.5 in Ba et al. 2021.

1. Right join the outcome (i.e. stops, arrests, or uses of force) and officer assignments using officer_id and date. I.e. keep every assignment and only outcomes which occurred during the start date of the shift.
2. Keep any outcomes which occurred during the hours of the shift.
   - Round the outcome time to the lowest hour.
   - Round the shift start time to the lowest hour.
   - Round the shift end time to the highest hour.
3. Right join outcomes and assignments using officer_id and date (incremented by one day). This captures outcomes which occurred the **following day** during an overnight shift.
4. Keep any outcomes which occurred during the hours of the shift using the same logic as in step 2.
5. Merge the data sets (outcomes merged by date and outcomes merged by the next day) by row.
   - Keep only distinct entries. Shift assignments which had no outcomes will have duplicate entries since they will appear as having had no outcomes on the focal day **and** the next day (since we also check the next day due to overnight shifts).
   - Filter out any shift assignments which matched with outcomes on one date but not the other. For example, a non-overnight shift will obviously not match with any outcomes the following day. As a result, there will be an entry for that shift assignment with no outcomes associated with it. However, there will be other entries for that shift assignment with correctly matched outcomes. As a result, the shift assignment entry in question will be erroneous.

Notes

- It will occasionally be the case that an outcome will match to multiple shift assignments. How can an outcome match to more than one shift assignment?
  - The officer in question could have had overlapping shift assignments, and the outcome occurred during the intersection of their shifts.
  - One of the shifts had a missing start time or end time, and the outcome matched with that shift (in addition to matching with at least one another shift assignment).
- It will occasionally be the case that not all shift assignments are retained in the merging process despite using a right join. Why aren't all shift assignments retained? Ba et al. (2021) have it so that any shift assignment which had an outcome on the same day **but** did not occur during the shift itself is counted as missing and not retained.
- Ba et al. (2021) implicitly assume that if an outcome and a shift assignment match on the **date** but the shift assignment is missing its **start time and end time** then the outcome **should still** match that shift assignment.

Merge stops - Ba et al. 2021

- **Number of resulting rows**: 3661994
- **Number of stops which occurred during a shift**: 1338755
  - **Percentage of stops** (1703158): 78.6042751%
- **Number of shift assignments retained:** 2928751
  - **Percentage of shift assignments**: 99.866027%
- 7850 shift assignments matched with at least one stop **without having** a start time and an end time.

- This represents 0.2676733% of all shift assignments.
- 18665 stops are affected or 1.0959054% of all stops.
- There are 315 stops which match to more than one shift assignment.
  - These **duplicate** stops represent 0.0184951% of all stops (1703158).
  - 326 shift assignments are affected or 0.0111161% of all shift assignments.
  - In total, there are 630 **duplicate stop-shift observations** representing 0.0172037% of all stop-shift observations.

Merge arrests - Ba et al. 2021

- **Number of resulting rows**: 2962392
- **Number of arrests which occurred during a shift**: 245153
  - **Percentage of arrests** (321872): 76.1647487%
- **Number of shift assignments retained:** 2931777
  - **Percentage of shift assignments**: 99.9692091%
- 3373 shift assignments matched with at least one arrest **without having** a start time and an end time.
  - This represents 0.1150143% of all shift assignments.
  - 3842 arrests are affected or 1.1936422% of all arrests.
- There are 97 arrests which match to more than one shift assignment.
  - These **duplicate** arrests represent 0.0301362% of all arrests (321872).
  - 180 shift assignments are affected or 0.0061377% of all shift assignments.
  - In total, there are 194 **duplicate stop-shift observations** representing 0.0065488% of all arrest-shift observations.

Merge uses of force - Ba et al. 2021

- **Number of resulting rows**: 2932837
- **Number of uses of force which occurred during a shift**: 7945
  - **Percentage of uses of force** (9293): 85.4944582%
- **Number of shift assignments retained:** 2932680
  - **Percentage of shift assignments**: 100%
- 132 shift assignments matched with at least one use of force **without having** a start time and an end time.
  - This represents 0.004501% of all shift assignments.
  - 134 uses of force are affected or 1.4419456% of all uses of force.
- There are 5 uses of force which match to more than one shift assignment.
  - These **duplicate** uses of force represent 0.0538039% of all uses of force (9293).
  - 10 shift assignments are affected or 3.4098504×10^{-4}% of all shift assignments.
  - In total, there are 10 **duplicate stop-shift observations** representing 3.4096678×10^{-4}% of all force-shift observations.

Merge process - Risi

1. Conduct a right non-equi join on stops and officer assignments using officer_id, time of the stop, the start time of the shift, and the end time of the shift. I.e. Keep every shift assignment and keep those stops which occurred exactly between the start time and end time for a particular shift.
2. **Differences from Ba et al. 2021**
   - Any shift with a missing start time and end time will not match with any stops. Ba et al. 2021 will match a shift with a missing start time and end time with any stop that occurred on the same day(s) as the shift.
   - The stop must have occurred exactly between the start time and end time of the shift. Ba et al. 2021 use rounding to create a window of time around the shift start and end times in which a stop may have occurred and would still count as happening during the shift.
   - If a shift assignment has a stop which occurs the same day, but the stop does not occur during the shift times, the shift is recorded as having no stops rather being recorded as *missing*.

- Any stops which are matched to multiple shift assignment are dropped. **All duplicate entries are dropped.**

Merge stops - Risi

- **Number of resulting rows**: 3650734
- **Number of stops which occurred during a shift**: 1314175
  - **Percentage of stops** (1703158): 77.1610737%
  - **Number of fewer stops matched than Ba et al. 2021**: 24580
  - **Percent reduction from Ba et al. 2021**: 1.8360342%
- **Number of shift assignments retained:** 2932680
  - **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 57 stops which match to more than one shift assignment.
  - These **duplicate** stops represent 0.0033467% of all stops (1703158).
  - 60 shift assignments are affected or 0.0020459% of all shift assignments.
  - In total, there are 114 **duplicate stop-shift observations** representing 0.0031227% of all stop-shift observations.
  - As noted above, **all these entries are dropped**.

Merge arrests - Risi

- **Number of resulting arrests**: 2962589
- **Number of arrests which occurred during a shift**: 240045
  - **Percentage of arrests** (321872): 74.5777825%
  - **Number of fewer arrests matched than Ba et al. 2021**: 5108
  - **Percent reduction from Ba et al. 2021**: 2.0835968%
- **Number of shift assignments retained:** 2932680
  - **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 9 arrests which match to more than one shift assignment.
  - These **duplicate** arrests represent 0.0027961% of all arrests (321872).
  - 18 shift assignments are affected or 6.1377307×10^{-4}% of all shift assignments.
  - In total, there are 18 **duplicate arrest-shift observations** representing 6.0757668×10^{-4}% of all arrest-shift observations.
  - As noted above, **all these entries are dropped**.

Merge force - Risi

- **Number of resulting rows**: 2932834
- **Number of uses of force which occurred during a shift**: 7773
  - **Percentage of uses of force** (9293): 83.6436027%
  - **Number of fewer uses of force matched than Ba et al. 2021**: 172
  - **Percent reduction from Ba et al. 2021**: 2.1648836%
- **Number of shift assignments retained:** 2932680
  - **Percentage of shift assignments**: 100% **NOTE** this should be 100%.
- 0 shift assignments matched with at least one use of force **without having** a start time and an end time. **NOTE** that this should be 0.
- There are 0 uses of force which match to more than one shift assignment.
  - These **duplicate** uses of force represent 0% of all uses of force (9293).
  - 0 shift assignments are affected or 0% of all shift assignments.
  - In total, there are 0 **duplicate force-shift observations** representing 0% of all force-shift observations.

– As noted above, **all these entries are dropped**.

Final notes

- Shift assignments which are missing their start time and end time could potentially be imputed which is explored elsewhere. The number of records affected is so small though it is not currently a priority.
- Similarly efforts could be made to de-duplicate stops and determine the correct shift assignment they should be matched to. This is explored elsewhere, but ultimately the number of records affected is so small it is not currently a priority.