

Replicate Ba et al. 2021 - Cleaning and Processing

Merge stops to officer shift 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 stops and officer assignments using officer_id and date. I.e. keep every assignment and only stops which occurred during the start date of the shift.
2. Keep any stops which occurred during the hours of the shift.
 - Round the stop 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 stops and assignments using officer_id and date (incremented by one day). This captures stops which occurred the **following day** during an overnight shift.
4. Keep any stops which occurred during the hours of the shift using the same logic as in step 2.
5. Merge the data sets (stops merged by date and stops merged by the next day) by row.
 - Keep only distinct entries. Shift assignments which had no stops will have duplicate entries since they will appear as having had no stops 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 one date but not the other. For example, a non-overnight shift will obviously not match with any stops the following day. As a result, there will be an entry for that shift assignment with no stops associated with it. However, there will be other entries for that shift assignment with the appropriate stops. As a result, the shift assignment entry in question is erroneous.

Merge results - 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 unique officers who had at least one stop during a shift: 6947
 - Percentage of officers: 100%
- Number of shift assignments retained: 2928751
 - Percentage of shift assignments: 99.866027%
 - Why aren't all shift assignments retained? Ba et al. (2021) have it so that any shift assignment which had a stop on the same day **but** did not occur during the shift times is counted as missing and not retained.
- 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.
 - Ba et al. (2021) implicitly assume the matched stops occurred during these shifts. The **date** matches at the very least (i.e. the date of the stop and the date of the shift).
- 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.

- How can a stop match to more than one shift assignment?
 - * The officer in question could have had overlapping shift assignments, and the stop occurred during the intersection of their shifts.
 - * One of the shifts had a missing start time or end time, and the stop matched with that shift (in addition to matching with at least one another shift assignment).

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 results - Risi

- Number of resulting rows: 3650620
- Number of stops which occurred during a shift: 1314118
 - Percentage of stops (1703158): 77.157727%
 - Number of fewer stops matched than Ba et al. 2021: 24637
 - Percent reduction from Ba et al. 2021 (1338755): 1.8402919%
- Number of unique officers who had at least one stop during a shift: 6947
 - Percentage of officers: 100%
- 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 0.0031131% of all stop-shift observations.
 - As noted above, **all these entries are dropped**.
 - How can a stop match to more than one shift assignment?
 - * The officer in question could have had overlapping shift assignments, and the stop occurred during the intersection of their shifts.

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.