Imputing Missing Start and End Times for Officer Shift Assignments

The Logic Behind Imputation

The key assumption being made is that shift assignments with a missing start time and end time will have the same or similar start and end times as other shift assignments on the same beat, on the same shift, in the same month, on the same day of the week. If there was a systematic reason for the shift assignments to be missing their start and end times though (e.g. because they are unlike other similar shift assignments), this assumption is not valid and the imputation will not be valid.

Some of the missing shift assignments will be unique in some sense (e.g. they may represent the only shift assignment for a particular beat, shift, month, day of the week combination) in which case no values can be imputed for that shift assignment.

Similarity of shift assignments is systematically varied (e.g shift and beat **or** shift, beat, and month **or** shift, beat, month, and day of week) and the most common start and end times for shifts in those categories are found. These values are then imputed for the missing start and end times.

Calibration Plot
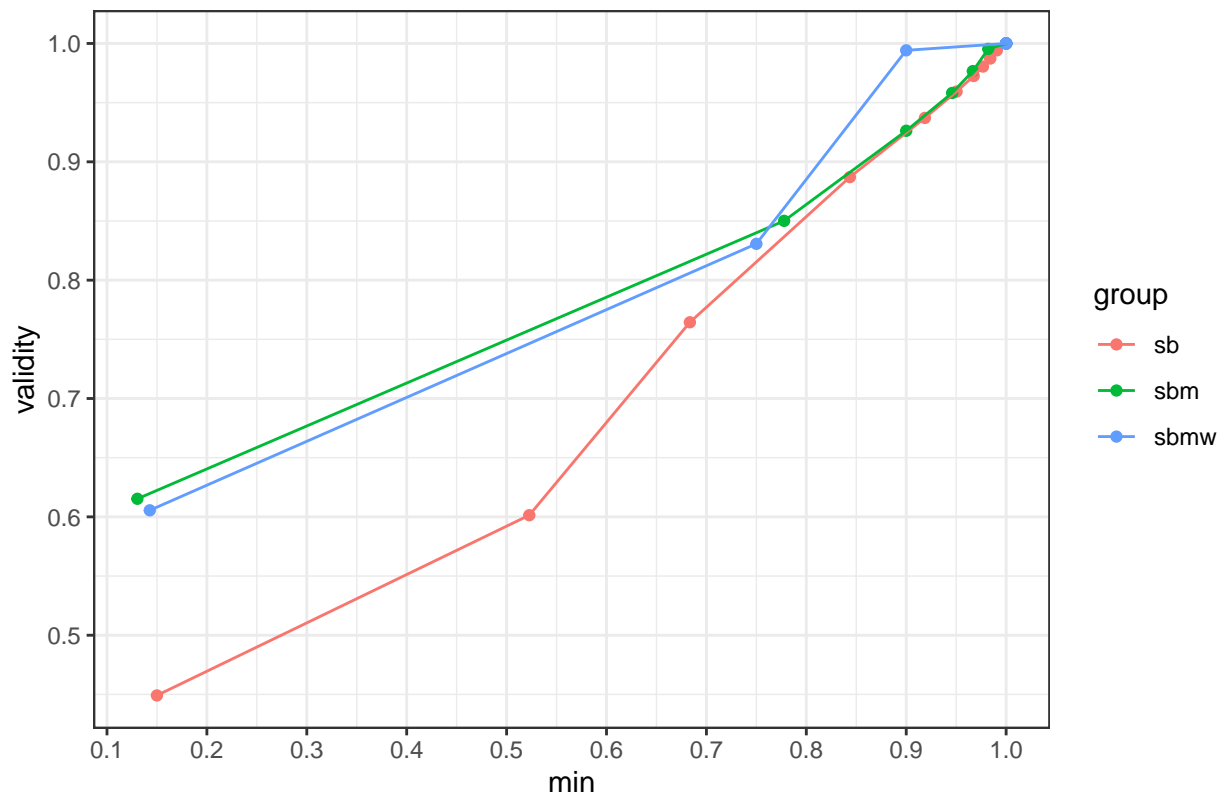
Percentage of Like Assignments vs. Number of Like Assignments
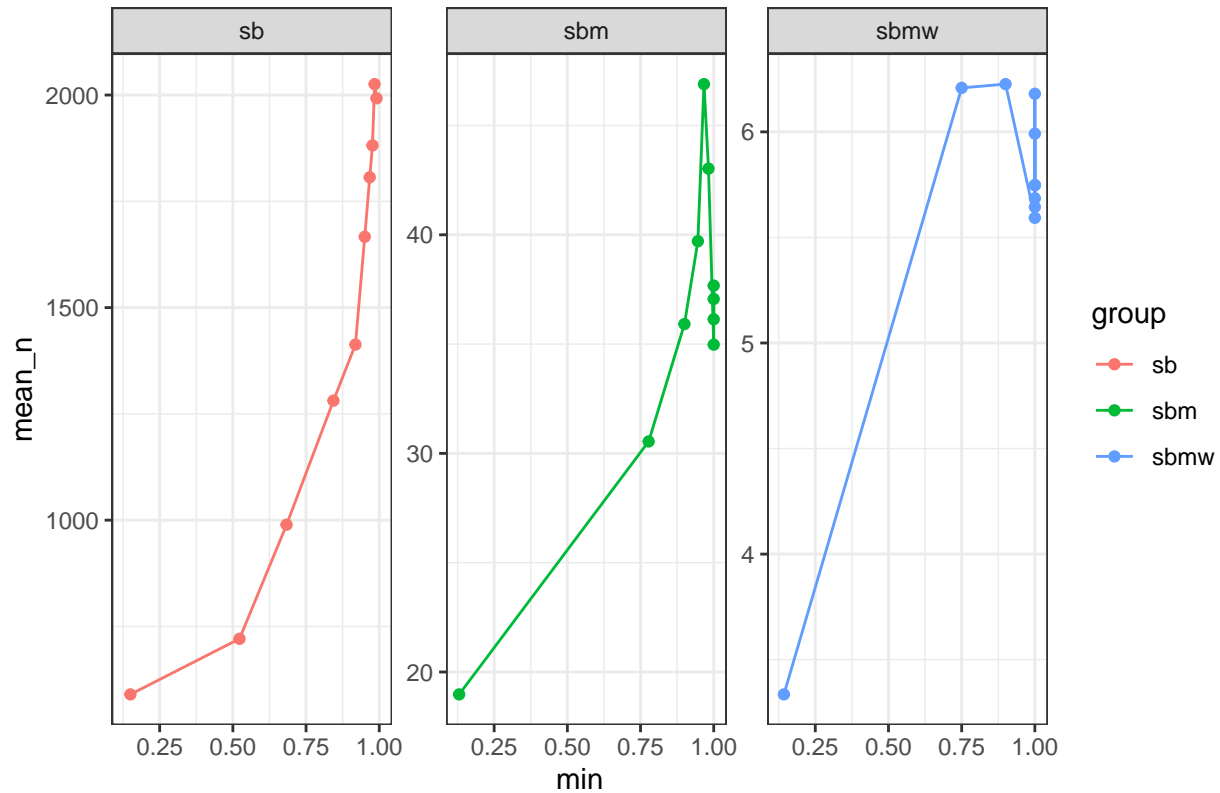
Evaluation Of Imputation Procedures

To evaluate how well the imputation procedures might be performing, I find the most common start and end times for each type of similarity of shift assignments. I then replace the **real** shift start and end times with the **the most common** start and end times for all alike shifts. I then calculate what percentage of start and end time pairs are correct using this procedure.

- Percentage correct for **shifts and beats**: 85.328737%
- Percentage correct for **shifts, beats, and months**: 93.2185544%
- Percentage correct for **shifts, beats, months, and weekdays**: 94.3044557%

The first graph is a calibration plot of sorts. For each shift assignment, a new start and end time are given. Accompanying the new start and end time is a percentage indicating how many other alike shift assignments have the same start and end time. These **match percentages** are sliced into deciles and the overall number of correct imputations is calculated for each decile group.

An example interpretation... if a given imputed start and end time is shared by 90% of all other alike shift assignments (**x-axis**), then about 93% of all imputed shift assignments in those categories are actually correct for **similar shifts and beats** and **similar shift, beat, and month**, and about 99% of all imputed shifts shift assignments in those categories are actually correct for **similar shifts, beats, months, and weekdays**.

The validation plot could be used for determining when someone would want to impute or not. For example, unless 90% of alike shift assignments shared times, a user might think the accuracy is not high enough to conduct imputation. This threshold could also be applied with an absolute number. For example, unless there are at least 10 other alike shift assignments, and at least 90% of them share the same start and end times, imputation should not happen. As seen in the 2nd plot, even though shifts similar on **shifts, beats, months, and weekdays** typically have the highest percentage correct, they also have the smallest number

of alike shift assignments. Whatever one values more (i.e. more absolute number of alike shifts vs. more relative number of alike shifts) will guide ones decision making.
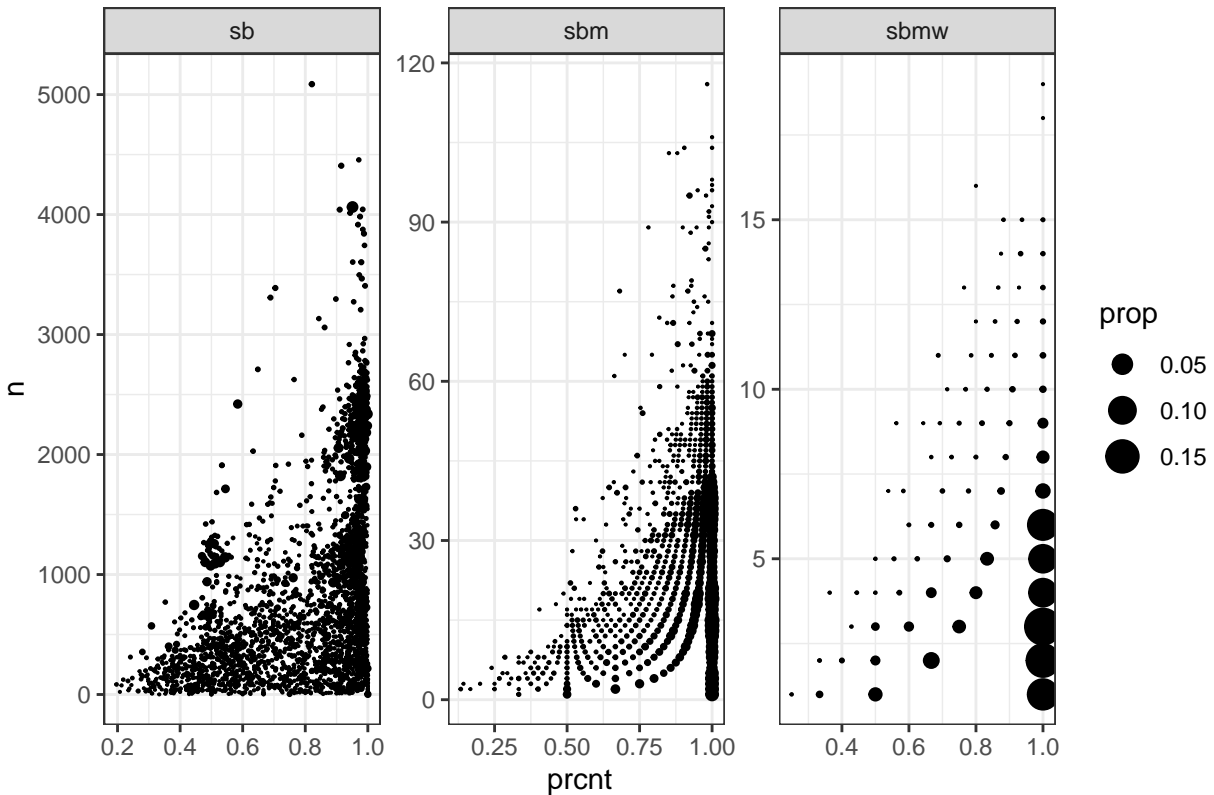
Graphs demonstrating what imputation process might actually look like

Number of missing observations which cannot be imputed

Percentage of alike shift assignments
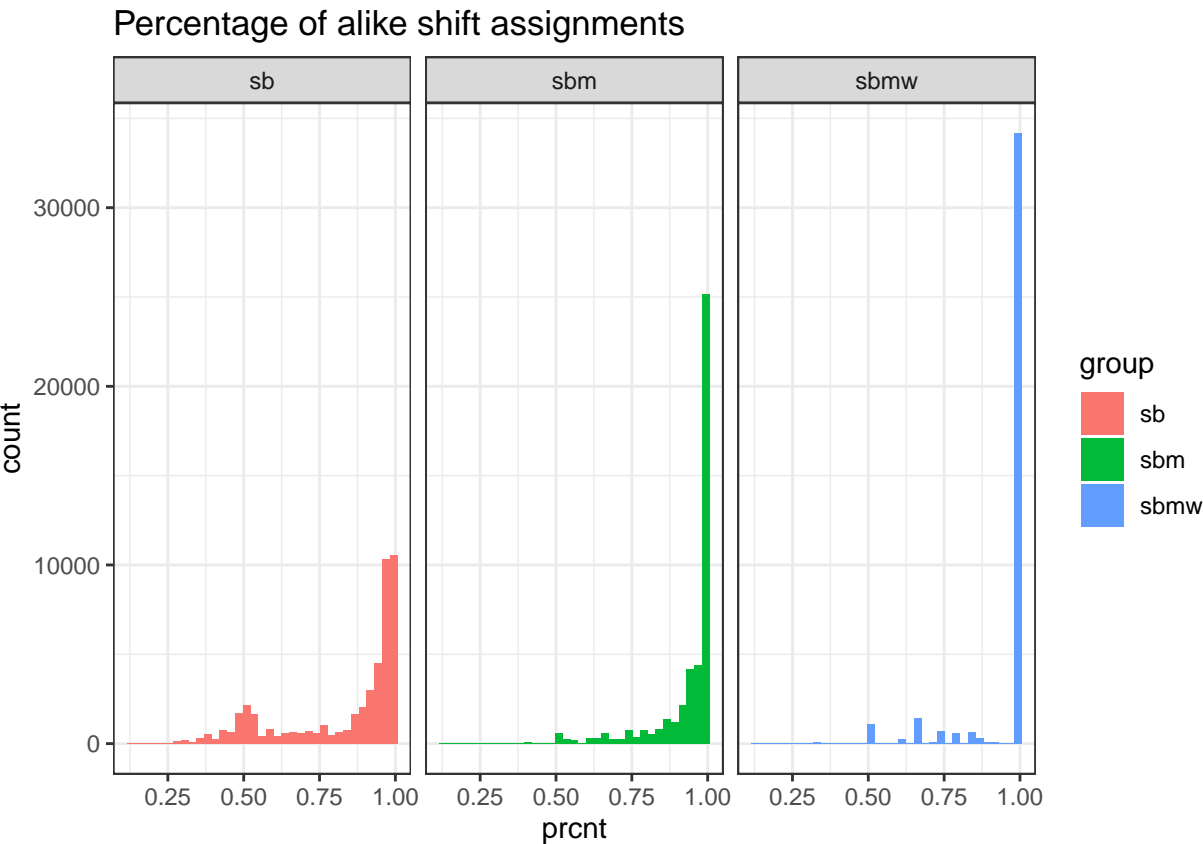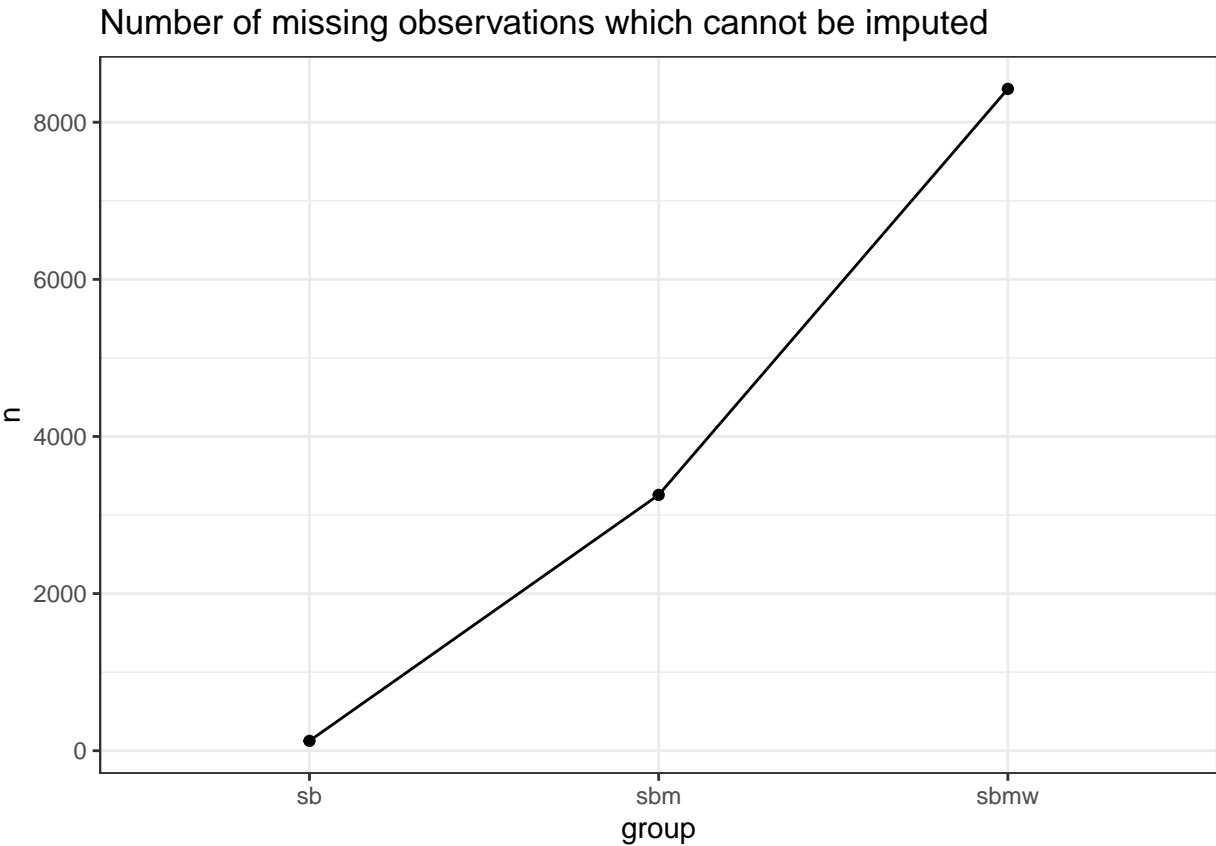
Absolute number of alike shift assignments
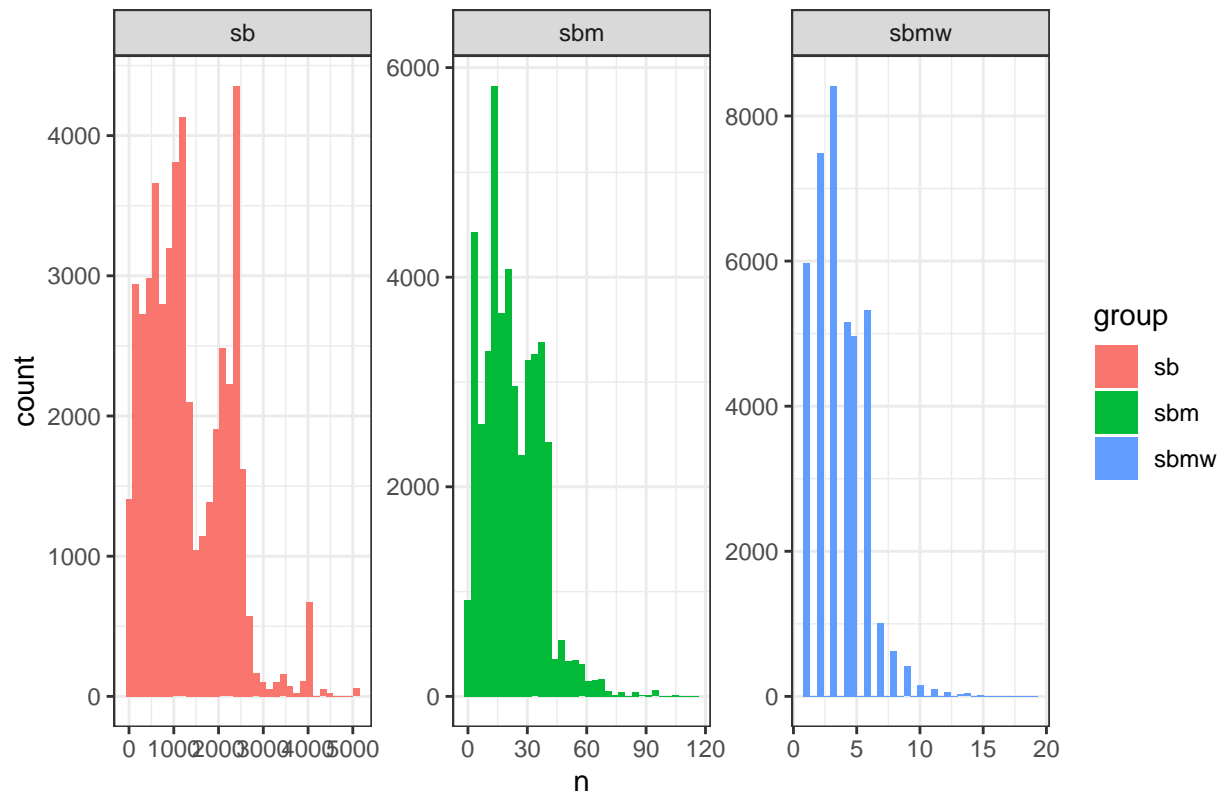
Number vs. percentage of alike assignments

Do shift assignments with missing start/end times look like non-missing shift assignments?

- Race largely looks the same. Percentage-wise, it would appear as if slightly more Black officers have missing shift assignments.
- Gender looks the same.
- Age looks the same.
- Spanish-speakers look the same.
- Disproportionately, units 3 and 8 have the most shift assignments with missing times.
- Weekday distributions don't completely align, but I am not concerned.
- Missing shift times occur almost entirely in 2012 which is an intriguing finding.
- Shift distribution looks roughly the same with the 1st shift being slightly underrepresented amongst shift assignments.
- Rank distribution looks largely the same.