

Notes on the Violence Prevention through Urban Upgrading Dataset

Alex Montgomery

March 24, 2014

1 Data Quality Issues

- Dateofinterview variable seems to have interviews in 2020...
- There is a variable in the excel spreadsheet just called "Column1". I am assuming that this is roof material (it has standard answers for that sort of question) but it is very unclear.
- What is the distinction between an after school activity and a post school activity? Post school = post graduation?
- **breadwinnerwork** seems to cover both businesses and areas/suburbs
- There is a variable simply called **other**. What? It comes straight after the mode of transport question and is always coded 1
- Minutes to get there variables need some attention, they are all in different formats (some as large as weeks, some in hours etc.) There are a bunch of them...
- **specifyname** doesn't have any non-missing values
- The alcohol purchase location variable, **ifathomewherepurchased**, has some answers that fit more than one category. A new code was created to reflect this. The other answers actually included string inputs describing the other place, so I coded those as well.
- Encoded all the use_1 through use_3 variables
- The survey designers left the in-between scores open to the interpretation of interviewee/interviewer in the questions that deal with level of acceptable safety. I've assigned generic codes to the undefined numbers. To clarify, most of the variables like this have values 1, 3 and 5 that correspond with "intolerably unsafe," "acceptable" and "completely safe" respectively. I've taken some license interpreting here so this might be worth cutting out.
- For **adult_risk_gender** variable I am assuming that "B" stands for "Both". The questionnaire doesn't really give much guidance on this.
- actually ended up having to create codes for a number of variables so that they were consistent across. To clarify,