These are the four variables for which I have information about named contacts that can help to identify people who are named more than once. If I were working my way through this manually, this is how I would be thinking about it:

**Name**

Problem: Most people only know other people’s first names, so that is usually all we have. When we have a last name (the last in the sample I sent you), it’s rarely helpful because it can’t link to anything else. However, as you’ll see from the sample I sent there is often two first names where the first-first name (e.g. Ab, Md, Abdul, Mohammad) is not helpful because it is a very common name or abbreviation. The other problems are spelling differences (see “Aesha” and “Aeyasha”) and spelling mistakes.

Function need: I need a function that can search both the first and second name and perform some kind of “fuzzy” match so that it recognizes all instances of a name and even when it is spelled differently, within reason. Thus something that can pick up both differences in length and differences within letters.

**Age**

Problem: People really don’t know their friends’ exact ages so I know I can’t just look for exact matches by name and age; I have to be able to find people of the same name with similar ages. For example, looking only at name and age (for the moment), Ab Rahim who is 60 years old is almost certainly not Ab Rahim who is 45 years old. That’s easy. But look at Ayesha who is 26 and Aesha who is 25—there is a good chance (without looking at other information) these are the same person.

Function need: I need a function that can search ages of all previous dataframe entries and identify those who are within, say, 5 years (absolute value) of entry [x]. I see how this will automatically get us into the same problem that the old code was supposed to create.

Is there a way to match people within age ranges by also checking if they also have matched names?

**Job**

Problem: No problems with this one. This should be an exact match.

Function need: Just identify other people with the same job code. This is from a standard list of <100 jobs, so there will be many many matches.

**Location**

Problem: No problems with this one. I spent two years(!) cleaning this variable and it is not down to a list of ~215 locations and I’ve confirmed there are no spelling mistakes. This should be an exact match.

Function need: Just identify other people with the same location.