1. Filter the data on the unit section number (consider only Tempe and Avondale) and on the congregate care type (consider only group home and shelter, discard institution/rtc)
2. See if case id matches, if it does build the flat file with the needed variables.

Following variables will be matched:

*Demographics of child and Family*:

* Birth date of child
* Gender of child
* Race/ethnicity of child
* Primary language of caretaker
* Marital status of primary caretaker

*Case Characteristics*:

* Date of first entry into out of home care – Totality of case – MM/DD/YY
* Type of Placement – Shelter/Group home – congregate care as of 7/1/16
* Number of removals for the child
* ICWA
* Placement type – why unsuccessful
* Allegation – difficult placement

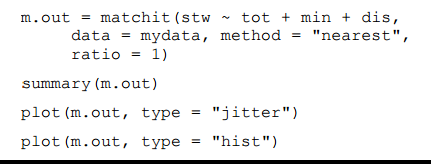
How to carry out propensity score matching?

* To perform PSM, we will need a data set that has cases in rows and variables in columns. We will need a grouping variable and one or more matching variables.
* The grouping variable in our case will be congregate/non-congregate care (Specifying which group a cases belongs to – treatment/non-treatment) and the matching variables are the ones that we want to attempt to equalize the groups on.
* We need to make sure that there is no missing data or R may not be able to perform the analysis.
* The next step is to perform the matching and evaluate the results.

m.out = matchit(stw ~ tot + min + dis, data = mydata, method = "nearest", ratio = 1)

stw - grouping variable

tot,min,dis – matching variables



* The ratio command indicates one-to-one matching - every treatment case will be matched with one control case. You can increase the number of control cases matched to each treatment case by increasing this number; usually this number is between 1 and 5.
* We can try the different methods provided by MatchIt and then select the one that results in lowest mean differences between the groups.

Points to be noted while choosing variables:

1. Children characteristics may be associated with both participation in treatment and outcome.
2. Children characteristics are of two types – measured and unmeasured.
3. Choose variables based on theory, exclude variables that are highly correlated with each other.
4. Identify the research Question – Find children from non-treatment with characteristics similar to kids who have received treatment so as to understand the benefits that the treatment has had.
5. For each child in the treatment group, find at least one untreated child from the comparison group who is identical or as similar as possible on all baseline characteristics.