Market names and village names cleaning and cross referencing:

Markt Survey A:

* Use as the key (lookup)
* Clean market names variable:
* Clean village names variable:
  + Tolower, removeWS, replace space with “.”
* Export as csv files: clean market names (distinct), clean village names (distinct) and market & village lookup (unique combinations of market-village)
* Use these csv as lookups for market survey c

Markt Survey C:

* Clean market and village names variable (as above)
* Export csv files
  + clean market names (distinct),
  + clean village names (distinct) and
  + market & village lookup (unique combinations of market-village)

compare village&market names from market survey c with market survey A list

* village names:
  + first compare survC-village names
  + for all entries from survey c- villages, compare with list of cvillages from survey A
  + If clearly the same as, or a typo of a name contained in survey A then add this to the corrected column
  + If the name doesn’t appear in survey-A village list then mark with ? (even if something similar is present)
  + Repeat for markets

Cross reference sweep 2:

* Populate missing names in eithr the market or village column based on a completed market or village name (i.e. backfilling)

Cross reference sweep 3:

* Populate remaining missing entries based on original entry.

Check markets on Tanzania Map:

* Use names cleaning script and plot locations of corrected market names
* Check that markets under same names are in the same location – corrections listed on Tanzania data presentation.

Additional cleaning steps:

* Check ward names
* Clean gps coordinates – some have mistakes.