

Hints for Cleaning “us_foreignaid_greenbook.xlsx”

- 1) Import (specify cell range)
- 2) Destrting variables that should be numbers
- 3) Correct country names to match WDI (using the “namecheck” command below)
- 4) Deal with 2 problematic cases (North Korea and Serbia)

Notes on using namecheck:

- 1) The code at the top of `Data Tutorial.do` defines the new command; the code must be run during that Stata session for the command to work.

- 2) The command syntax is:

```
namecheck name
```

where *name* is the name of a variable that is supposed to match between files.

- 3) `namecheck` compares values in the current variable *name* against values in the `names.dta` file created earlier in your `.do` file. The values in `names.dta` are the ones we are standardizing on (in our case draw from the WDI database). That variable needs to match across files. So rename it as needed. In our example, we are using the variable name `Country` so in the WDI data, you need to do this:

```
rename countryname Country
```

- 4) The Excel file `namelist.xls` is created (or replaced) if `namecheck` finds a mismatch. `namelist.xls` must not be open when you run `namecheck` or you get an error message.
- 5) ALL country names in the data you are checking against `names.dta` must match. That means if there are any countries (or territories, etc.) in your data that are not covered in the WDI, you need to drop them (or not use the `namecheck` command in your last run of the `.do` file).
- 6) Once `namecheck` has created `namelist.xls` you go through the file (columns A & B) copying pairs of country names that should match into columns D & E. I usually turn on “Freeze Panes” (found under the “View” menu) and put titles on columns I’m pasting into so I don’t forget where to paste. You may need to move names around (if the “match” is a long way from its partner) before pasting into the new columns. The key is that the order in columns D & E match each other so that when you delete duplicates, the country names line up. Once all the names are cut and pasted in D & E, insert a new column in between D & E. Then delete duplicates (found under the “Data” menu). Then insert code before, in between and after these columns. Suppose the names are in D3 and F3. Then that code is:

```
In C3: replace Country=""
```

```
In E3: " if Country=="
```

```
In G3: "
```

```
In H3: =C3&D3&E3&F3&G3
```

Copy all these values down their respective columns. Copy created code in column H into your `.do` file (above the `namecheck` command). Finally, copy all the remaining unmatched names in Column B to a fresh Excel column. Drop duplicates. Put Stata code before and after this column to generate commands to drop each of these observations. If the “fresh column” starts at M3:

```
In L3: drop if Country=="
```

```
In N3: "
```

```
In O3: =L3&M3&N3
```

Copy this code into your `.do` file, too.