

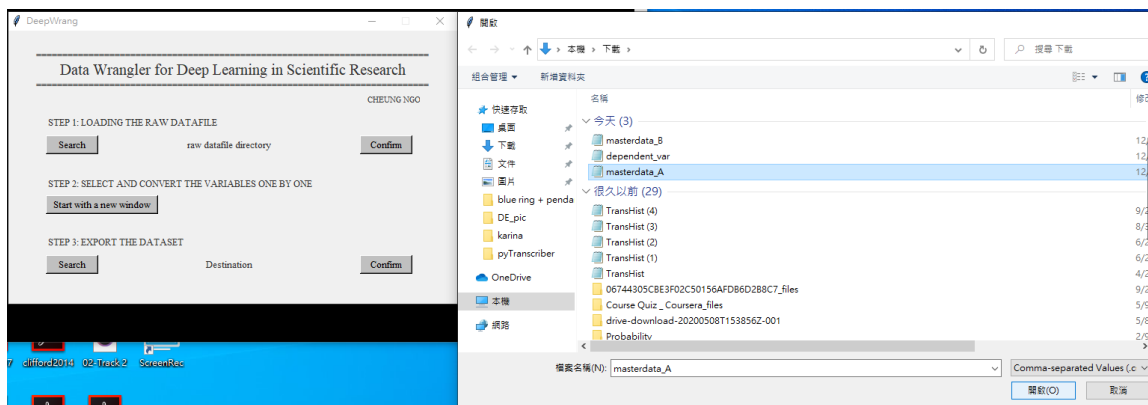
# Data Wrangler for Deep Learning in Scientific Research v1.0

## A Manual

Download link: (only for PC)

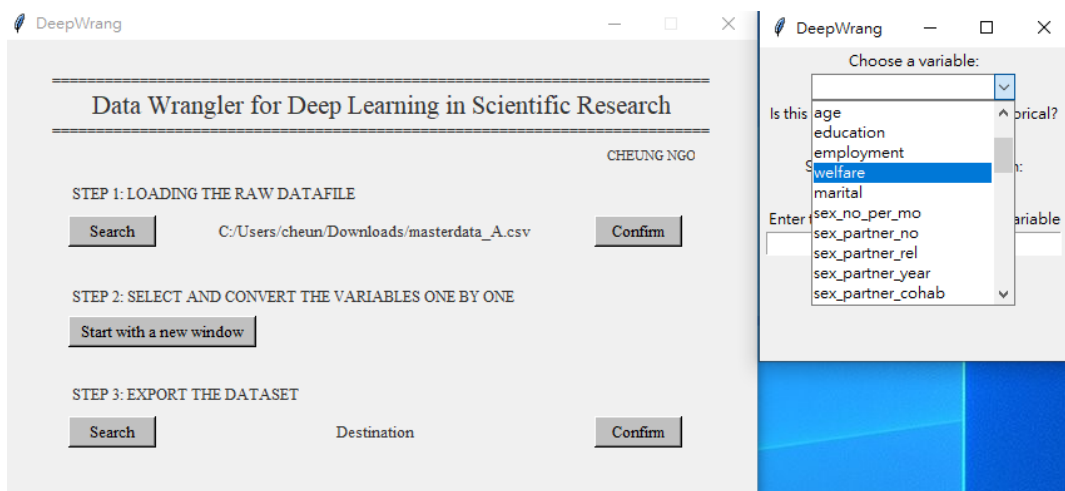
<https://drive.google.com/file/d/1XtcP971PHynPfflcZLMxsnL9oOj22nNF/view?usp=sharing>

### Step 1:

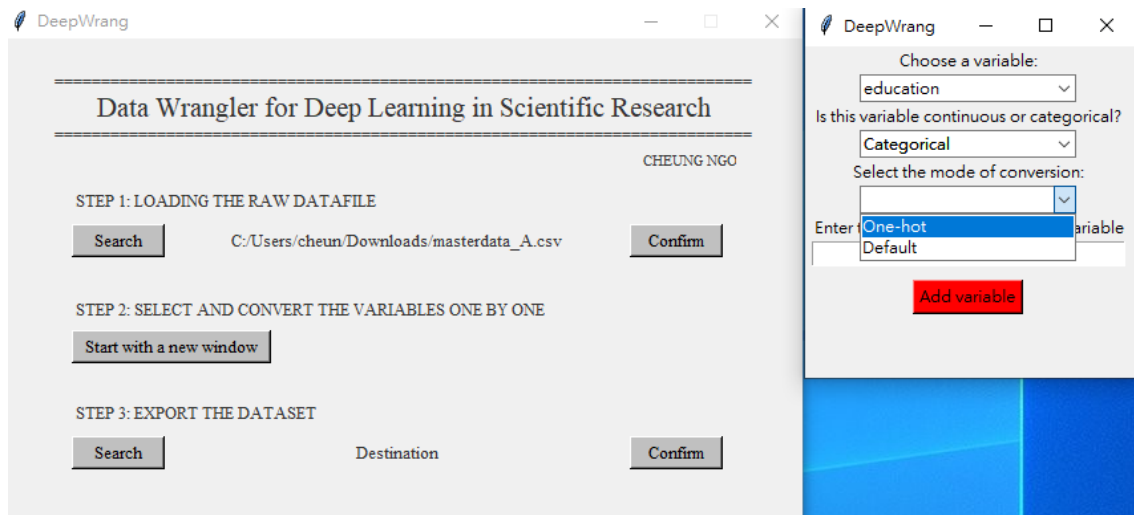


- Just double click the deepwrang.exe file will do
- Click the “Search” button under “STEP 1”, then a search window will appear
- Find the file then open it, then click “Confirm”
- For now it could just open a \*.csv file; if you only have an \*.xls file, you will have to convert it into a \*.csv file in excel

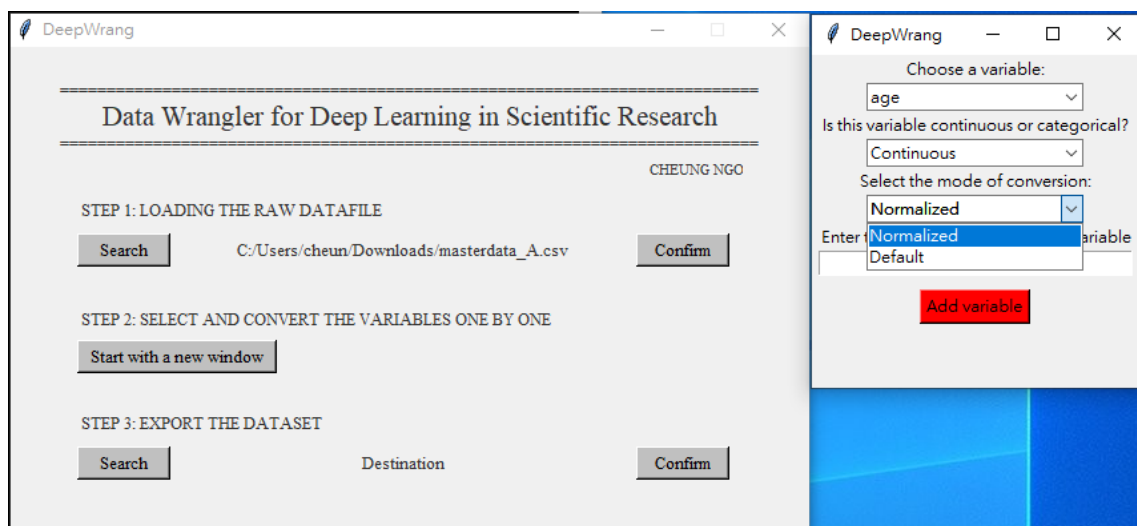
### Step 2:



- Click the button under “STEP 2”, then a new small window will appear
- The software will automatically load the variables you have
- You will have to choose what kind of variable it is, i.e. “Continuous” or “Categorical”



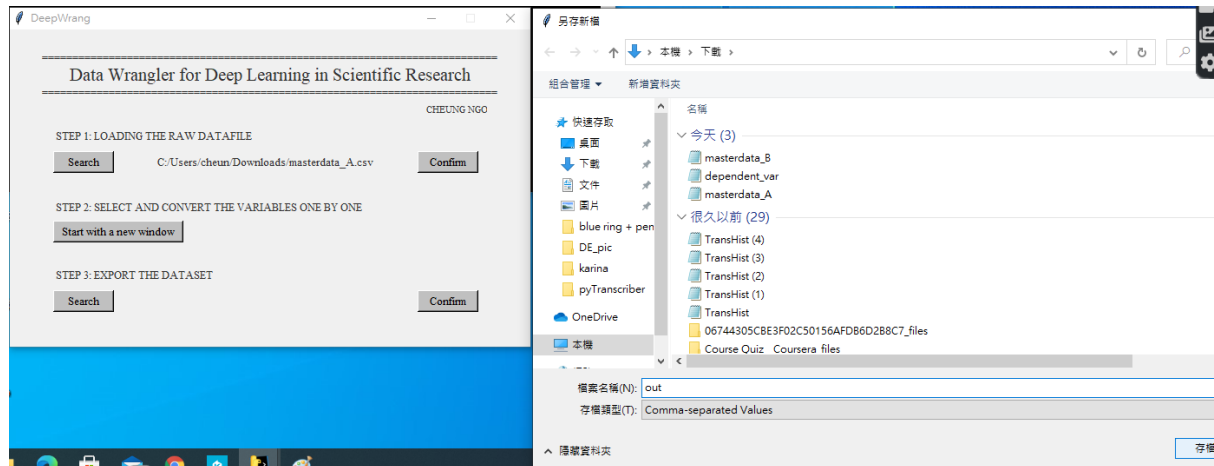
- For “Categorical” factors, there are two mode of conversions: “One-hot” and “Default”
- “Default” conversions will automatically convert your classes of variables into digits (0,1,2,etc..); for example, if a variable contains F and M it will now be 0.0 and 0.1; the key (which class to which number) will be recorded in the log (\*\_logs.csv); This function is particular useful for part 3 candidates
- “One-hot” conversions are useful for multivariate stepwise regressions (esp. for the ones using statistical softwares like R and Python) and deep learning



- “Continuous” factors have two modes of conversions: “Normalized” and “Default”
- “Normalized” are specifically for building deep learning models, preventing gradient explosions during propagations and back-propagations
- “Default” basically does nothing except for casting integers into floats, which is necessary for building deep learning model
- After selecting and converting all the variables, you can just close the small window or leave it there

- If you have more than one raw datasets, you can just load the new dataset in “STEP 1”, and convert the variables in “STEP 2”. The previously converted variables will still be here as long as you do not close the program

### Step 3:



- After all the hard work, you can now export the data
- You can search the directory you want, and type the name you want the export file to be
- There would be 4 files written: \*.csv, \*\_vartype.csv, \*\_stats.csv, and \*\_log.txt
- \*.csv will be the converted data
- \*\_vartype.csv will be a dataframe recording the variable type (continuous or categorical) of each of the variables; this is very important for some other programs that I wrote which can perform descriptive or regression analysis for all variables in the data at one go at the same time (available in R, will be released later)
- \*\_stats.csv recorded the mean and SD of all the “Normalized” “Continuous” variables, which will be useful for predictions using the trained Deep Learning model
- \*\_logs.csv recorded all the activities