**MLC Demo with R scripts (C&P): Read Me File**

**Contents and lay-out of the Replication Documentation:**

The replication documentation provided here for the ‘MLC Demo with R Scripts (Copy and Paste Protocol)” contains a group of folders and files, which are organized as below:

**MLC-Demo-RScripts/** (the main folder)

*readme.pdf* (the document you are reading currently)

**OriginalData/** (a sub-folder of the **MLC-Demo-RScripts/** folder)

*OriginalPew.sav* (the original version of the PEW data file, in the *.sav* file format)

*OriginalWDI.xlsx* (the original version of the WDI data downloaded from the World Bank, in Microsoft Excel *.xlsx*format

**Metadata/** (a sub-folder of the **Original-Data/** folder)

*metadata-guide,html* (the guidebook listing variables for the *original-wdi.xlsx* dataset)

*Pew-GAP-Survey-details 1.1.html*  (the guidebook listing variables for the *original-pew.sav* and *importable-pew.dta* dataset)

**CommandFiles/**(sub folder of the **MLC-Demo-RScripts/** Folder, which contains the R Scripts necessary to replicate the Midlife Crisis Demo)

*Master.R* (An R Script with executes the following four files with sequential order)

*1-Processing.R* (the R Script which is responsible for importing both the *OriginalWDI.xlsx* and *OriginalPew.sav* data files, cleaning the data and processing the data by removing irrelevant observations and variables, and saving it in the **Temp/** folder)

*2-Merging.R* (The R Script which is responsible for merging the two data files in the **Temp/** folder created by *1-Processing.R*, and saving this in the **AnalysisData/** folder as *Analysis.csv*)

*3-DataAppendix.R (*The R Script which is responsible for recreating the figure, graphs, and tables in the data appendix)

*4-Analysis.r* (The R Script which is responsible for creating figures, tables, regressions, and graphs in the report)

**AnalayisData/ (**sub-folder of the “**MLC-Demo-RScripts/**” folder)

*Analysis.csv* (A *.csv* file which is the processed data, an output of *2-Merging.R*, and is used by both *3-DataAppendix.R* and *4-Analysis.R* in generating figures, graphs, tables, and regressions)

*DataAppendix.pdf (*The data appendix for the final analysis file, *Analysis.csv*, which contains the summary statistics and basic frequency tabulations of the variables used in the Analysis)

**Temp/ (**sub-folder of the “**MLC-Demo-RScripts/**” folder)

*PEWAnalysis.csv* (a temporary file which is processed by *1-Processing.R’s* effects on *OriginalPEW.sav* but not yet merged with the below data set by *2-Merging.R*)

*WDIAnalysis.csv* (a temporary file which is processed by *1-Processing.R’s* effects on *OriginalWDI.xlsx* but not yet merged with the above file by *2-Merging.R*)

**Output/ (**sub-folder of the “**MLC-Demo-RScripts/**” folder)

**DataAppendix/ (**sub-folder of the “**Output/**” folder, and contains the following files)

*agedistribution.png* (a figure for the data appendix, which is generated by *3-DataAppendix.R* R Script)

*gdp\_pcdistribution.png* (a figure for the data appendix, which is generated by *3-DataAppendix.R* R Script)

*cm\_satisdistribution.png* (a figure for the data appendix, which is generated by *3-DataAppendix.R* R Script)

*gov\_consdistribution.png* (a figure for the data appendix, which is generated by *3-DataAppendix.R* R Script)

*age\_sqdistribution.png* (a figure for the data appendix, which is generated by *3-DataAppendix.R* R Script)

*satisdistribution.png* (a figure for the data appendix, which is generated by *3-DataAppendix.R* R Script)

*agetable.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*analysismeans.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*countrytable.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*gdp\_pctable.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Scriptt)

*gov\_constable.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*satis\_freq.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*satistable.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*agetable.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*Cm\_satistable.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*analysismeans.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*countrytable.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*gdp\_pctable.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Scriptt)

*gov\_constable.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*satis\_freq.txt* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*satistable.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

*Cm\_satistable.csv* (a table for the data appendix, which is created by the *3-DataAppendix.R* R Script)

**Figures/ (**sub-folder of the “**Output/**” folder，and contains the following files)

*figure1.png* (a graph generated by the *4-Analysis.R* R Script, using data from the analysis data file *analysis.csv*）

*figure2.png* (a graph generated by the *4-Analysis.R* R Script, using data from the analysis data file *analysis.csv*）

**InText/ (**sub-folder of the “**Output/**” folder, and contains the following files)

*intext1.txt* (a number generated from the *4-Analysis.R* R Script, using data from the analysis data file *analysis.csv*）

*intext2.txt* (a number generated from the *4-Analysis.R* R Script, using data from the analysis data file *analysis.csv*）

**Tables/ (**sub-folder of the “**Output/**” folder, and contains the following files)

*regression\_table.tex* (a table created by the *4-Analysis.R* R Script, the LaTeX form of the table below)

*regression\_table.txt* (a table created by the *4-Analysis.R* R Script, the plain text form of the table above)

*means\_table.csv* (a table created by the *4-Analysis.R* R Script, with the means of satis, gdp per capita and government consumption by country)

*means\_table.txt* (a table created by the *4-Analysis.R* R Script, with the means of satis, gdp per capita and government consumption by country)

**Using this documentation to replicate the demo**

To replicate the exercise, you will need access to a computer with a copy of R Studio. The R Scripts are designed to be compatible with Version [1.2.1335](https://www.google.com/search?q=1.2.1335&stick=H4sIAAAAAAAAAONgVhLQL9E3MjbPKCkzMiyvKM-qWMTKYahnpGdobGwKAC_nxUwfAAAA&sa=X&ved=2ahUKEwiMuPDF-M3iAhXqUd8KHVPkBPYQmxMoATAlegQIDBAP) (2019, April 8 Release), and were written on Windows 10, however they are backwards compatible with Windows 8 Releases.

Please note that in addition to installing R Studio, R must be installed prior to using R Studio. Both are available as open-source, free softwares. There are also certain packages which are required in addition to Base R, those commands are not recogized by Base R. They will be automatically downloaded, installed, and loaded if one follows the below instructions.

The steps required for the replication are as follows:

* Download the **MLC-Demo-RScripts/** folder, as well as the contents, before unzipping the compressed files Please take care not to change the organization of the files or the contents in any way.
* Install R and then R Studio, in this order, if not done so already. Using the default settings is fine.
* Launch R Studio, and set the working directory of R to the **MLC-Demo-RScripts/** folder from the *sessions* tab.
* Open up the file *1-Processing.R*

*1-Processing.R* will open up the original data files located in the

**OriginalData/** folder (O*riginalWDI.xlsx* and *OriginalPew.sav*) remove the unnecessary variables which are not used in the exercise, clean the data, drop certain observation which are missing, and then save them in the **Temp/** folder as *PewAnalysis.csv* and *WDI*A*nalysis.csv* respectively).

Next, open up the *2-Merging.R* file in the **CommandFiles/** folder, and run the file. Again,please make sure the working directory is set to the main folder of **MLC-Demo-RScripts** Folder. If there are any steps which are unclear, please refer to the comments written in the R script file.

*2-Merging.R* will open up the two files created by *1-Proccessing.R* which are saved in the **Temp/** folder before merging them together, and then then save this in the **AnalysisData/** folder as the csv type file A*nalysis.csv* (Please note that if there exists a previous version of the csv, this will be overwritten). *analysis.csv* will be needed for both the data appendix and analysis portions of the exercise.

Next, open up the *3-DataAppendix.R* file in the **Command-Files/** folder, and run the file. Again,please make sure the working directory is set to the main folder of **MLC-Demo-RScripts** Folder. If there are any steps which are unclear, please refer to the comments written in the R script file.

*3-DataAppendix.R* will:read the data from the analysis data file *Analysis.csv, g*enerate the graphs and tables in the Data Appendix. . Each graph and tables will show the the frequency distribution of one of the variables in the analysis data file. These graphs will be saved in the **DataAppendix/** subfolder of the **Output/** folder. (If previously generated versions of these graphs are already stored in the **Output/** folder when you run *3-DataAppendix.R*, and they are different from the newly generated graphs, they will be overwritten.) It also generates the tables (such as frequency tables) found in the data appendix, as well as the counts of missing and total variables, and saves those in the **DataAppendix/** subfolder as well.

Finally, open up *4-Analysis.R*, ensuring that the working directory (under session tab) is set to **MLC-Demo-RScripts/** folder. This file will open up the *Analysis.csv* file, (which was created and saved in the **Analysis-Data/** folder as an output of *1-Processing.R*). It will generate the figures and regressions needed for the demo and save them in the various subfolders of the **Output/** folder.

If needed, *Master.R* can quickly run all of the above R Scripts in sequential order by calling up all of them: open up the original data files located in the

**OriginalData/** folder process the data by removing variables and unnecessary observations before saving them in **Temp/**, calling them again and merging them, save the results as *analysis.csv* in the **AnalysisData/** folder, then generates the graphs and tables and regressions in both the data appendix and the analysis.