GeoSocial - Readme

Thank you for using GeoSocial solution. We have developed an R script to execute the data linkage. This document provides a detailed explanation of the script and explains how to execute it. You chose the following elements:

The round (2011) of the TSP (2021) are going to linked with the wave Wave 3 - 2011 of the Longitudinal survey used in the linkage.

The Geographical unit used to do the data linkage was SA3 2021. More information about this: https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/latest-release

The Census information that are going linked to the longitudinal survey are:

- Age By Sex

The longitudinal topic that are going linked are:

- Student

The R script starts cleaning the environment and installs the GeoSocial library with the necessary dependencies.

```r
rm(list=ls())
##### ------ Install GeoSocial ----- #####
install.packages('geosocial_1.0.tar.gz',repos = NULL, type = 'source',dependencies=TRUE)
```

Once the library is installed, the code will load both the library and the parameters file created from the previous questionnaire. Furthermore, the code will generate the required folders where any outputs will be saved.

```r
##### ------ Load GeoSocial ----- #####
library(geosocial)
########## ------- Load parameters --------- ######
LoadParameters(file = "parameters.json")
##### ------ Create folders ------ #####
CreateFolders()
```

A log file is created to record each step of the linkage, flag any error or warning messages to assist with debugging and ensure transparency and reproducibility of the results.

```r
##### ------ Create log file ------ #####
logNm <- 'test.log'
GenerateLog(name = logNm)
```

The following chunk loads the parameters into the global environment.

```
######### ---------- 0. Read parameters ------- ###########
LSAY_cohort = ParsingParameter('LSAY_cohort')
LSAY_waves = as.numeric(ParsingParameter('LSAY_waves'))
LSAY_topics = ParsingParameter('LSAY_topics')
SurveyFiles = ParsingParameter('SurveyFiles')
TSP_year = ParsingParameter('TSP_year')
TSP_variables = ParsingParameter('TSP_variables')
```

After defining the parameters, we proceed to load the longitudinal survey.

```
####### -------------- 1. Loading data ----------- #########
###### -------- Load LSAY ------- ####
LSAY_2009 = LoadLSAY(wave = LSAY_waves,cohort =
                       LSAY_cohort, LSAY_topics = LSAY_topics)
##### ---- Extract survey responses --- #####
SurveyResponses = LSAY_2009[['SurveyResponses']]
##### ---- Exctract GeospatialResponses --- #####
GeospatialResponses = LSAY_2009[['GeospatialResponses']]
```

Now we read the geographical data and the metadata, in this case, the Time series profile 2021.

```
### -------- Load TSP 2021  -------- #######
TSP_2021 = LoadTSP2021(year=LSAY_waves,variables = TSP_variables)
### -------- Extract data ------ #####
TSP_2021_data = TSP_2021[['data']]
### -------- Extract metada ------ #####
TSP_2021_metadata = TSP_2021[['metadata']]
```

We start to prepare the longitudinal survey data for the data linkage. The first step is to transform the postcode to SA3. We store all the information about the linkage, to export at the end.

```
####### -------------- 2. Geospatial data linkage ----------- #########
####### ----------- 2.1 POAS TO SA3 ----------- ######
SurveyResponses_SA3 = LSAY_POA_SA3(data=GeospatialResponses,
                                   concordances=concordances)
###### --------- SA3 ------- ########
SurveyResponses_SA3_IN = SurveyResponses_SA3[['sa3']]
###### --------- ABS Metric ------- ########
SurveyResponses_SA3_metric = SurveyResponses_SA3[['metric']]
###### --------- ABS Concordances ------- ########
SurveyResponses_SA3_concordances = as.data.frame(SurveyResponses_SA3[['concordances']])
###### --------- ABS Ratio ------- ########
SurveyResponses_SA3_ratio = as.data.frame(SurveyResponses_SA3[['ratio']])
```

After having all in terms of SA3, we need to transform all to SA3 2021 to link with TSP 2021.

```
####### ----------- STAGE 2.2: SA3 TO SA3 year ----------- ######
SurveyResponses_SA3_TSP = LSAY_PSA3_SA3(data=SurveyResponses_SA3_IN,
                                        concordances=concordances,
                                        year_out = TSP_year)
###### --------- SA3 ------- ########
SurveyResponses_SA3_TSP_OUT =
        SurveyResponses_SA3_TSP[['sa3']]
###### --------- ABS Metric ------- ########
SurveyResponses_SA3_TSP_metric =
         SurveyResponses_SA3_TSP[['metric']]
###### --------- ABS Concordances ------- ########
SurveyResponses_SA3_TSP_concordances =
         as.data.frame(SurveyResponses_SA3_TSP[['concordances']])
###### --------- ABS Ratio ------- ########
SurveyResponses_SA3_TSP_ratio =
         as.data.frame(SurveyResponses_SA3_TSP['ratio'])
```

After transforming the longitudinal survey in the same terms as the Time series profile, we performance the data linkage between the two datasets.

```
####### ------- GeoSpatial join ------ #####
DataJoined = GeoSpatialJoin(year=LSAY_waves,
                            GeospatialResponses=SurveyResponses_SA3_TSP_OUT,
                            SurveyResponses=SurveyResponses,
                            TSP_data = TSP_2021_data,
                            TSP_metadata = TSP_2021_metadata)
```

We calculated consolidated the metrics, and concordances used in the process and storage, and stored it in an xlsx file.

```
############ --------- Geospatial report --------- ############
summaryResults = SummaryReport(concordances_POA= SurveyResponses_SA3_concordances,
                               concordances_SA3 = SurveyResponses_SA3_TSP_concordances,
                               metrics_POA = SurveyResponses_SA3_metric,
                               metrics_SA3 = SurveyResponses_SA3_TSP_metric)


######## -------- Write the summary report --------- #######
openxlsx::write.xlsx(x=summaryResults,
                     file='outputs/Summary.xlsx')
```

Subsequently, we export the data in Stata, preserving the structure and metadata.

```
####### ------- Write stata ------ #####
WriteStata(path = 'outputs/', waves =LSAY_waves,
           DataJoined = DataJoined,
           SurveyResponses = SurveyResponses)
```

Finally, we write a summary of the data linkage in the log, and storage the metrics.

```
####### -------- Write final log ------ #######
SummaryLog(summaryResults = summaryResults)
```