

# Group Project:

## Analyse SA Social Atlas data

Establish how the aspect of **Early childhood development** assigned to your group below (**expressed as percentage**) is related to the social indicators listed as dot-points.

For example for Group 1

The aspect of Early childhood development:

AEDC, Language and cognitive skills (schools-based) - developmentally at risk)

Social Indicators are:

- Housing/ Transport
- Income support

Import data, clean it up, do exploratory data analysis and correlation analysis; plot the data on geographic maps, perform linear regression analysis and summarise your findings.

The tasks and the marking criteria are outlined below in detail.

Assignment tasks and marking criteria

1. Get the data from

<http://www.phidu.torrens.edu.au/social-health-atlases/data#social-health-atlases-of-australia-local-government-areas>

Task	Cost in points of tasks done incorrectly	Total marks
<b>Import data from the provided link. Data preparation and clean-up</b>		20
Import data from the specified worksheets. Establish data structure, number of rows and columns. Combine data into one data set using dplyr package Convert factors in the data into character variables Establish any empty columns using summary functions and delete them using dplyr package functions Rename columns - create short, informative names. Identify and address missing values in the data. Create a character variable showing Local Government Area type (e.g. C, DC etc). Delete Local Government Area information from the name of LGAs	Import data from the specified worksheets. -2 Establish data structure, number of rows and columns. - 2 Combine data into one data set using dplyr package - 2 Convert factors in the data into character variables - 2 Establish any empty columns using summary functions - 2 Delete them using dplyr package functions - 2 Rename columns - create short, informative names. - 2 Identify and address missing values in the data for predictors - 2 Identify and address missing values in the data for response - 2 Create a character variable showing Local Government Area type (e.g. C, DC etc). Delete Local Government Area information from the name of LGAs - 2	
<b>Exploratory Data Analysis.</b>		

<p>Create:  Frequency tables for categorical predictors. Bar charts for categorical predictors.  Summaries for numeric predictors. Histograms for numeric predictors.  Summary tables of predictors and response by LGA type (C, DC etc)  Scatterplots with the response variable for all numeric predictors  Boxplots with the response variable for all categorical predictors  All graphs should be formatted so that each graph has a title, axis titles and a black-and-white background.  Well summarised graph and tables interpretation</p>	<p>Frequency tables for categorical predictors. -1  Bar charts for categorical predictors. - 2  Summaries for numeric predictors. Histograms for numeric predictors. -1  Summary tables of predictors and response by LGA type -1  Scatterplots with the response variable for all numeric predictors - 2  Boxplots with the response variable for all categorical predictors - 2  All graphs should be formatted so that each graph has a title, axis titles and a black-and-white background. -3  Well summarised graph and tables interpretation - 2</p>	20
<p><b>Create Geographic Maps of childhood development per LGA using:</b>  Leaflet  Ggmap  tmap</p>	<p>For each map lack of mapping technique shown in the lecture -2</p>	10
<p><b>Perform correlation analysis on the resulting data</b></p>		15
<p>Do Correlation analysis  prior to regression;  on predictors only;  include corrplot; The corrplot should be well-formatted  Summarise the analysis findings</p>	<p>Correlation analysis should be done prior to regression; -2  Should be done on predictors only; -2  Should include corrplot; - 2  The corrplot should be well-formatted -2  Analysis findings are well-summarised and explained -2</p>	
<p><b>Regression model specification and refining.</b></p>		5
<p>From the data assigned to your group choose 2 numeric variables that have the lowest correlation with each other and the LGA type (character variable) and use them as predictors of the response variable that is specified for your group.  Make sure that the initial model is specified correctly;</p>	<p>The 2 variables with the lowest correlations are chosen correctly -1  initial model specified correctly; -1  fit, R-sq and predictors significance are commented on correctly -1  Interactions are correctly introduced and their significance assessed and commented on -1  Findings are not well-summarised -1</p>	

fit, R-sq and predictors significance are commented on correctly; Interactions are correctly introduced and their significance assessed and commented on Findings are well-summarised		
<b>Regression residual diagnostics, plots and analysis</b>		5
All regression diagnostics plots using techniques from the lecture materials Comments on regression assumptions tested by the plots, identified outliers and influential observations	All regression diagnostics plots using techniques from the lecture materials (-1 point if a plot is incorrect or not included) Comments on regression assumptions tested by the plots, identified outliers and influential observations (-1 point per incorrect interpretation of a plot, outliers or influential observations )	
<b>Outlier and Influential observation resolution</b>	Outlier and Influential observation resolution - 1 point per each incorrect handling of outliers and influential observations)	5
<b>Final model explanation (depth of approach; fit; issues; resolution; business implications)</b>	Final model explanation (depth of approach; fit; issues; resolution; business implications) -1 point for any incorrect statement	5
<b>Can the Rmd document supplied be knitted into HTML</b>	If cannot be knitted, then -10	10
Formatting and professionalism of presentation	No Table of contents -2.5 No Structure (eg, introduction, Methodology, Findings, Conclusion etc) -2.5	5
<b>Bonus marks</b>	For extra effort, additional techniques etc	As needed

# Early childhood development aspects and social indicators for each group for this assignment

## Group 1 South Australia

Early childhood development: AEDC, **Language and cognitive skills** (schools-based) - % developmentally at risk

- Housing/ Transport
- Income support

## Group 2. South Australia

Early childhood development: AEDC, Early childhood development: AEDC, % Developmentally vulnerable on two or more domains

- Internet access at home
- Labour force

## Group 3. South Australia

Early childhood development: AEDC, Language and cognitive skills (schools-based) - % developmentally at risk

- Internet access at home
- Indigenous status

## Group 4. South Australia

Early childhood development: AEDC, % Developmentally vulnerable on two or more domains

- Housing/ Transport
- Education

## Group 5. South Australia

Early childhood development: AEDC, Communication skills and general knowledge - % developmentally vulnerable

- Labour force
- Summary measure of disadvantage

Group 6. South Australia

Early childhood development: AEDC, Language and cognitive skills (schools-based) - % developmentally at risk

- Labour force
- Education

Group 7. South Australia

Early childhood development: AEDC, % Developmentally vulnerable on two or more domains

- Internet access at home
- Education

Group 8. South Australia

Early childhood development: AEDC, Communication skills and general knowledge - % developmentally vulnerable

- Indigenous status
- Education

Group 9. South Australia

Early childhood development: AEDC, Language and cognitive skills (schools-based) - % developmentally at risk

- Families
- Education

Group 10. South Australia

Early childhood development: AEDC, Communication skills and general knowledge - % developmentally vulnerable

- Housing/ Transport
- Families