



**The Longitudinal Study of Australian Children:
An Australian Government initiative**

Distance to Coast Data Information

April 2018



Australian Government

Australian Institute of Family Studies

1 Introduction

Geographic data offers significant potential to enhance analytic utility by adding new and informative dimensions to data. Distance to coast data can be used to analyse the impact of access to services and amenities such as climate, environmental and urban conditions on well-being and health.

2 Distance to Coast Data

Distance to coast data has been generated for every residential address in waves 1-7 that was successfully geocoded to latitude and longitude. A total of 57,993 distances were calculated, providing distance to the coast measures for 10,0089 study participants over the 7 waves. The distance measure is the straight-line distance for the residential to closest point on the Australian coastline. While this distance measure provides a readily useable and comparative measure of coastal accessibility, it does not reflect access to transport infrastructure, such as access to roads, bridges or railways, or geographic impediments, such as mountains or rivers.

3 Distance to Coast Calculation

The separation principle¹ was maintained during the distance calculation process. The latitude and longitude coordinates were provided to the LSAC Data Linkage and Integrating Authority² unit (DLIA) with a dummy ID by the ABS. The resulting distance data did **not** include the latitude and longitude, only the calculated distance to the coast with the dummy ID. An ID concordance file mapping the dummy ID to the LSAC study ID was provided to the LSAC Data Manager by the ABS to join the distance data to the LSAC content data.

Distances to the coast were calculated using the QGIS v2.18.2 graphical information system (GIS)³. Australia's coastline was defined by a spatial point layer with 100m resolution that was provided by Deakin University. The latitude and longitude data were projected to a spatial point layer using the EPSG 3577 coordinate system (GDA 94, Australian Albers). The coastline line was also projected to the same coordinate system. The distance matrix analysis tool in QGIS was used to determine the distance from each residential data point to the closest coastal data point.

4 Confidentialisation

The addition of geographic can raise the risk of re-identification of individuals. The distribution and the geographic spread of distance data were assessed to determine the appropriate treatments required to mitigate disclosure risk. AIFS consulted with GIS data users to ensure that the proposed treatments did not impact detrimentally on the analytic utility of the data. To control disclosure risk while maintaining a high degree of discrimination in the distance measure a progressive rounding approach with a top code was applied. All distances over 350km were truncated and were assigned a top coded value of "350km+". This approach maintains distance as a continuous measure while controlling risk of re-identification. The following rounding and top coding treatment was applied:

- Rounding distances <1km to the nearest 100m
- Rounding distances <10km to the nearest 1km
- Rounding distances <300km to the nearest 10km
- Top coding distances ≥350km to "350km+"

¹ Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data – a best practice protocol. Aust N Z J Public Health. 2002;26(3):251-5

² <https://aifs.gov.au/our-work/resources/data-linkage>

³ QGIS Development Team (2016). QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

5 Data Structure

The distance to the coast data for each cohort (B and K) are each stored in a separate file. The file contents are detailed on table 1.

Table 1: Distance to coast data variable details				
File	Cohort	Variable Name	Child's age	Variable Label
lsacbgeodtc	B	ageodtc	0/1	0/1 - Distance to coast
		bgeodtc	2/3	2/3 - Distance to coast
		cgeodtc	4/5	4/5 - Distance to coast
		dgeodtc	6/7	6/7 - Distance to coast
		egeodtc	8/9	8/9 - Distance to coast
		fgeodtc	10/11	10/11 - Distance to coast
		ggeodtc	12/13	12/13 - Distance to coast
lsackgeodtc	K	cgeodtc	4/5	4/5 - Distance to coast
		dgeodtc	6/7	6/7 - Distance to coast
		egeodtc	8/9	8/9 - Distance to coast
		fgeodtc	10/11	10/11 - Distance to coast
		ggeodtc	12/13	12/13 - Distance to coast
		hgeodtc	14/15	14/15 - Distance to coast
		igeodtc	16/17	16/17 - Distance to coast

The variable names use the standard LSAC naming convention (see section 6.1 of the LSAC Data User Guide). The topic is “geo” and the specific data item identifier for the distance to the coast measure is “dtc”.