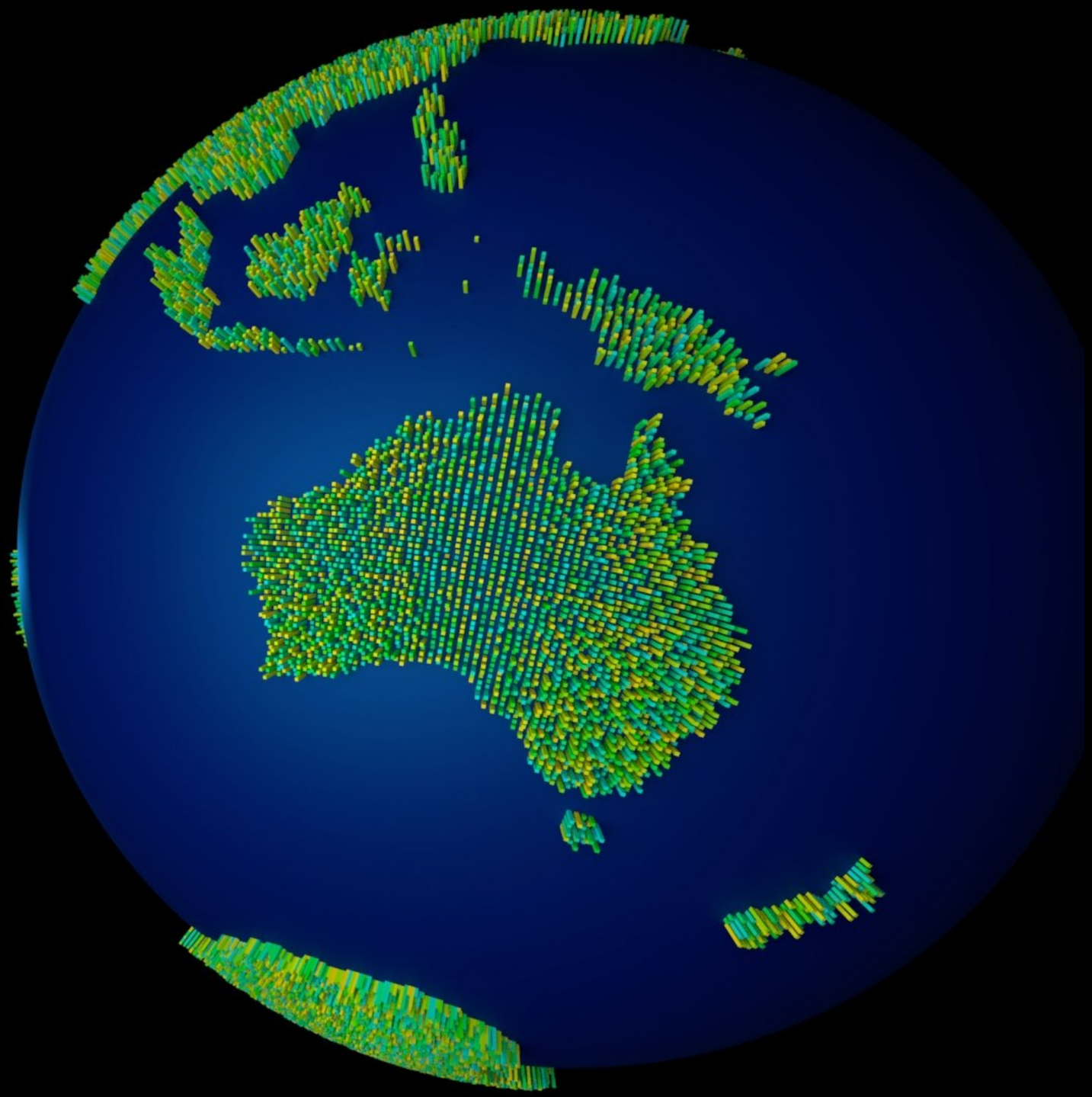
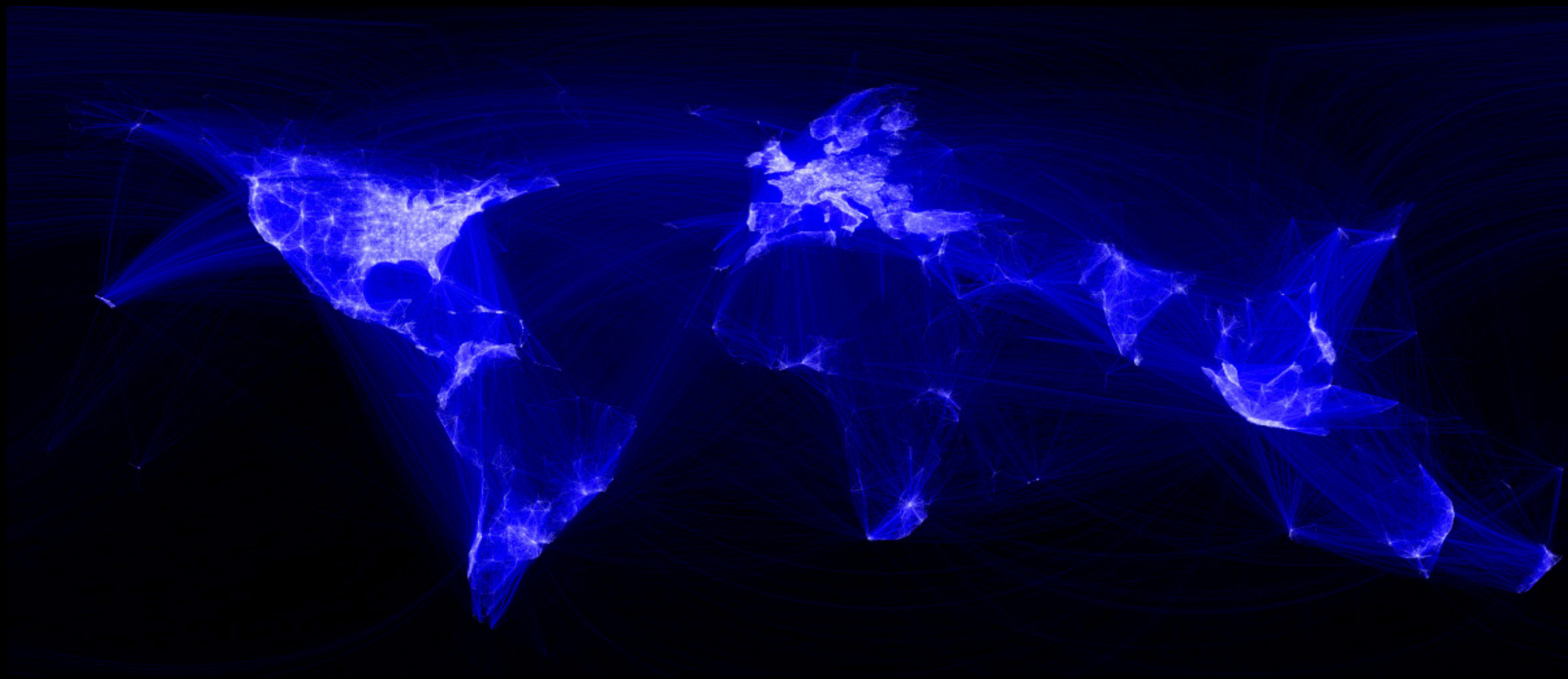
R AS A GEOGRAPHIC INFORMATION SYSTEM



Produced using Autodesk Maya — MASH procedural graphics

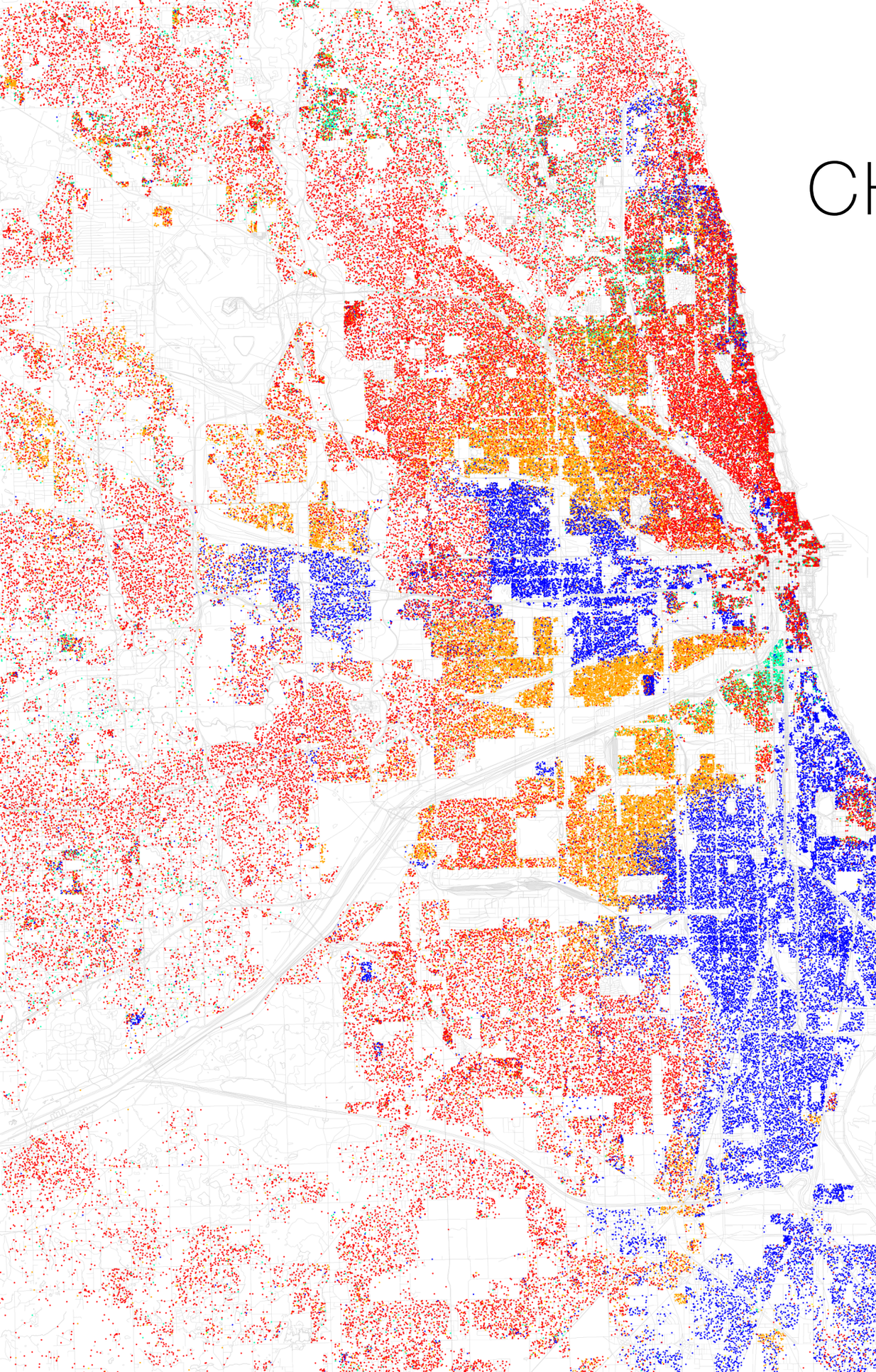


Visualizing Friendships Paul Butler http://fbmap.bitaesthetics.com/

http://topography.geotheory.co.uk/ Robin Edwards

CHICAGO ETHNICITY 2010

Bill Rankin



Negro
White
Asian
Hispanic

Each dot represents 25 people

http://www.radicalcartography.net/

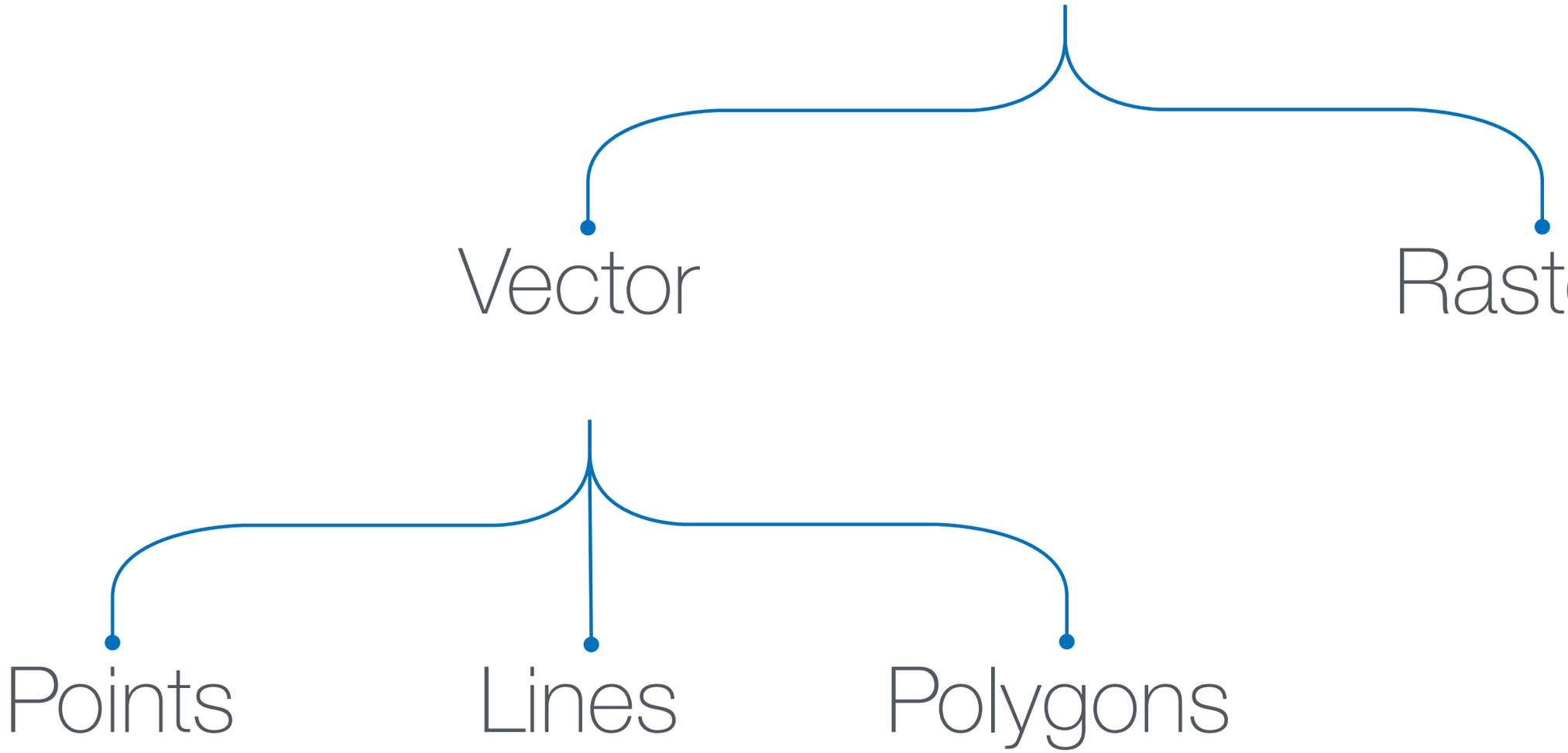
SPATIAL DATA

Everything happens somewhere

SPATIAL BASICS

CLASSIFYING SPATIAL DATA

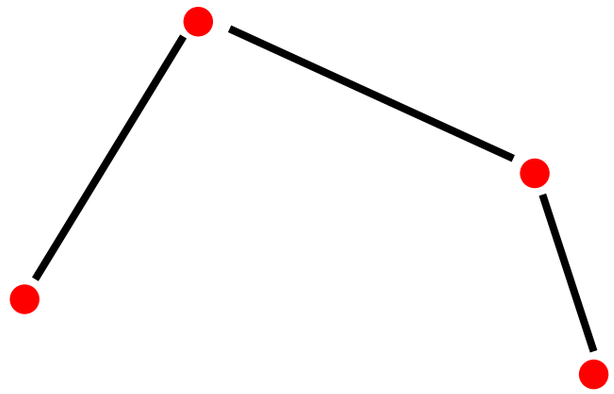
Simple spatial taxonomy



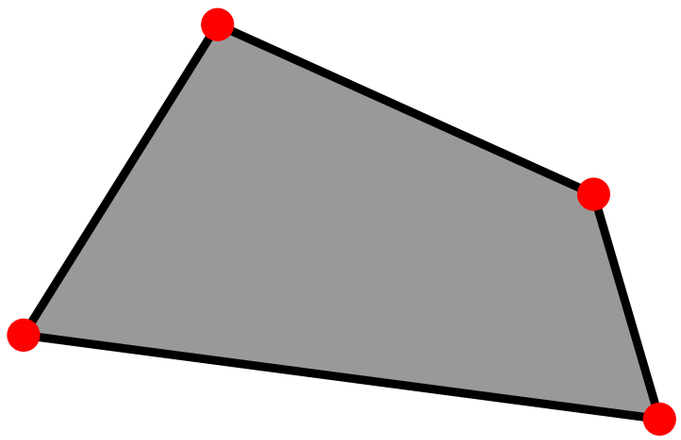
VECTOR DATA TYPES

Point

Line

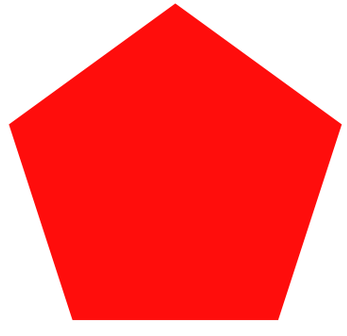


Polygon

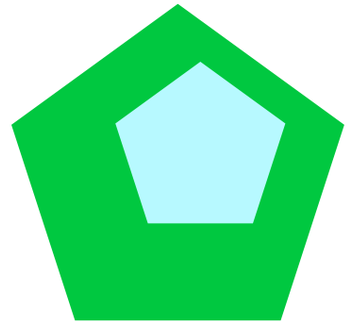


POLYGON TYPES

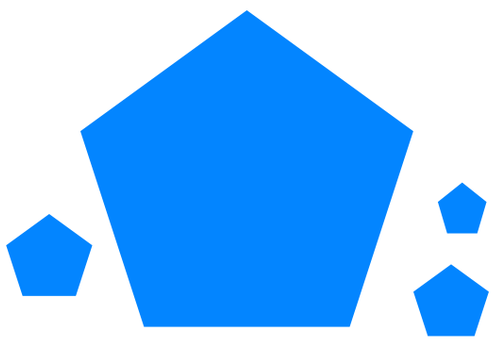
Simple



Interior ring (i.e. lake)



Multipolygon (i.e islands)



GEODESY

Measurement and representation of the earth

Geographic Coordinate Systems (GCS)

Spherically based. Uses polar coordinates.

Projected Coordinate Systems (PCS)

Projected onto a 2d plane. Uses Cartesian coordinates.
Based on Geographic Coordinate Systems. Examples: Mercator / Albers.

Reference: http://resources.esri.com/help/9.3/arcgisengine/dotnet/89b720a5-7339-44b0-8b58-0f5bf2843393.htm

GCS SYSTEMS

Geographic Coordinate Systems

Codified systems

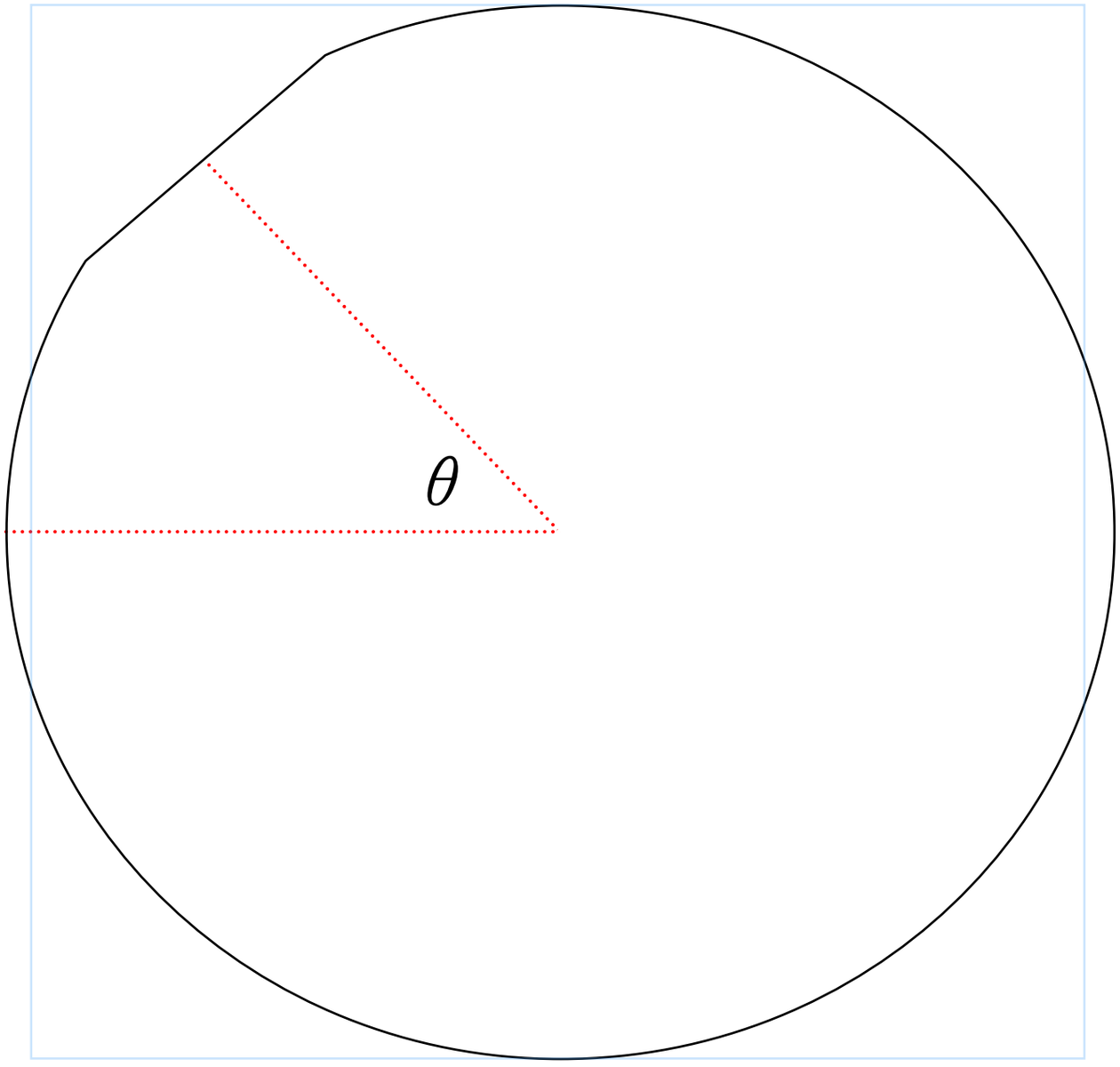
Various GCS systems have been codified making them easy to refer to.
The World Geodetic System and EPSG Geodetic Parameter Dataset are examples.
These codes are available as metadata or are embedded into the spatial data.

Ellipsoid | prime meridian | geocentre | units

The GCS determines the specific ellipsoid, prime meridian, geocentre and the units used.

Reference: https://en.wikipedia.org/wiki/World_Geodetic_System https://www.epsg-registry.org/

ELLIPSOID EARTH



The earth is wider than it is tall — it is flatter at the poles.

It is also not regular. Some places are bumpy or indented.

Depending on the location and purpose, GCS systems use different ellipsoids and different assumptions for the location of the earth's centre.

Reference: http://resources.esri.com/help/9.3/arcgisengine/dotnet/89b720a5-7339-44b0-8b58-0f5bf2843393.htm

CENSUS DATA

OVERVIEW

1996, 2001, 2006, 2011, 2016 data sets are publicly available

Geographic boundaries and CSV data

Geographic boundaries (i.e. geometries) are released in two formats: ESRI shapefiles and Mapinfo files. To be useful the geometry needs to be joined to the CSV data.

SA1 to SA4 hierarchy

Non-overlapping and exhaustive hierarchical partitions of Australia.

Additional geometries

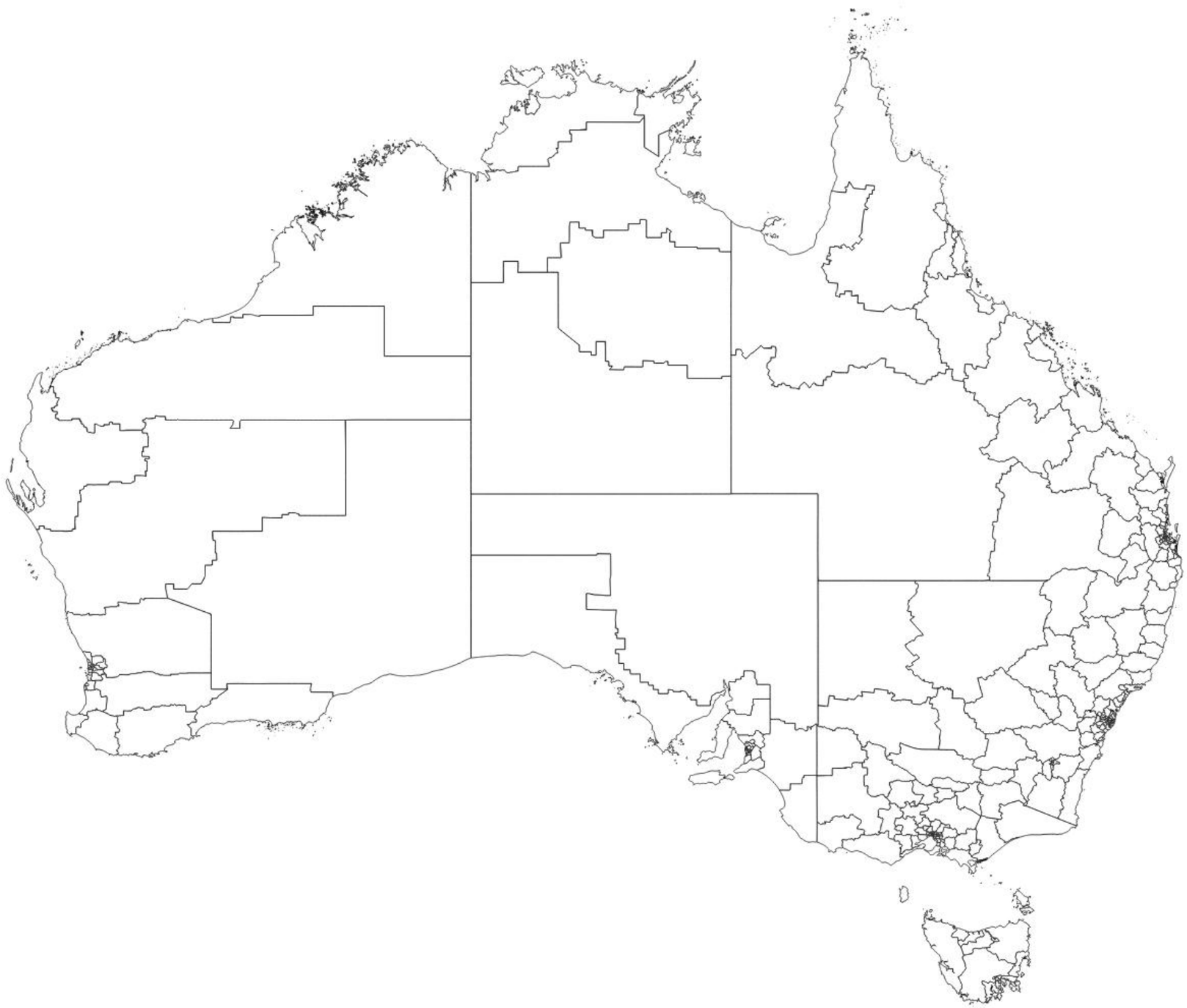
Indigenous areas, Commonwealth / state / local electoral boundaries, state boundaries, urban localities, greater city areas, significant urban areas.

CENSUS REGIONS —SA4

106 polygons, 39 Mb

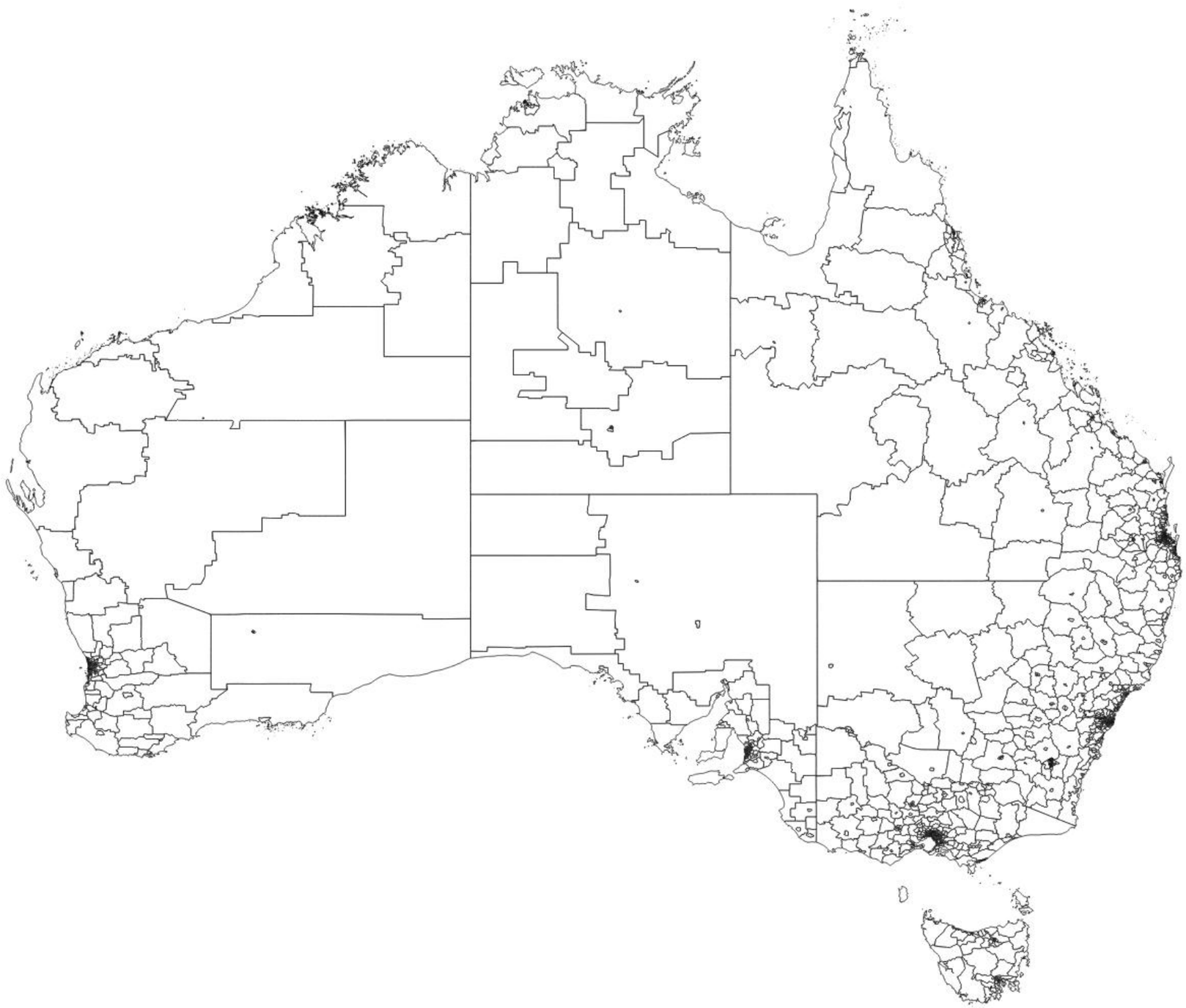
CENSUS REGIONS —SA3

351 polygons, 46 Mb



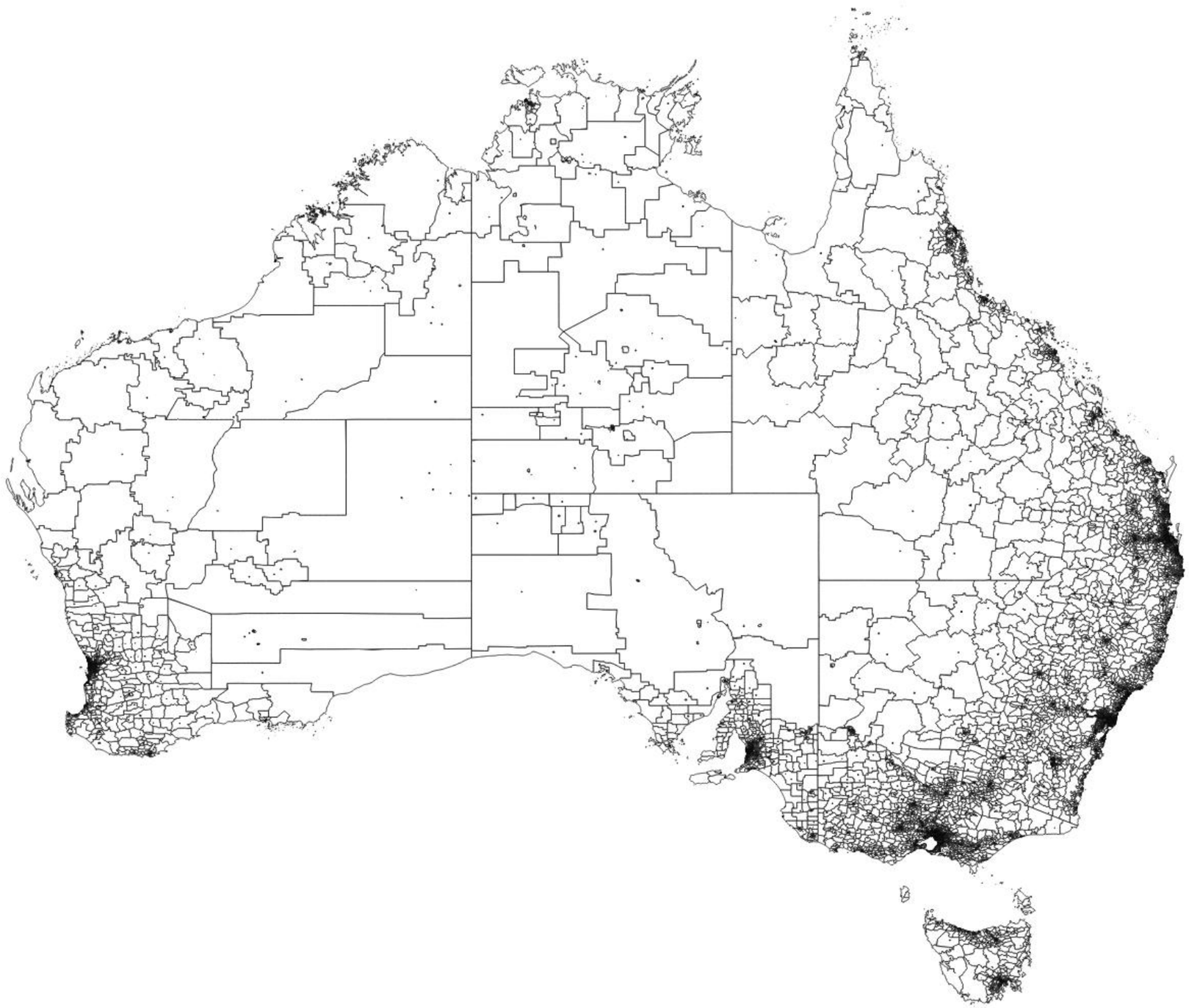
CENSUS REGIONS —SA2

2,214 polygons, 66 Mb



CENSUS REGIONS —SA1

54,805 polygons, 145 Mb



ADDITIONAL INFORMATION

Data examples

languages spoken, income, gender, age, work type, country born, married status, work location, indigenous status, work location, time series.

Mesh blocks

Approximately 347,00. Only number of people and number of dwellings is publicly available. Each mesh block contains about 30 to 60 dwellings.

Reference: http://www.abs.gov.au/websitedbs/censushome.nsf/home/meshblockcounts

LOADING DATA

ESRI SHAPEFILES

An industry standard

Developed by ESRI in the 1980s. Used by the ABS (and many others) to disseminate digital boundaries / spatial data. Vector based data.

One shapefile consists of many files

Confusing naming. A single 'shapefile' consists of usually four (minimum is 3) files. The four files are located in the same folder.

The files will have the same name but different suffixes.

**.prj contains projection information. *.dbf contains the actual geometry. The *.dbf is the largest file and can be read in using Microsoft Access.*

LOAD AND SET CRS

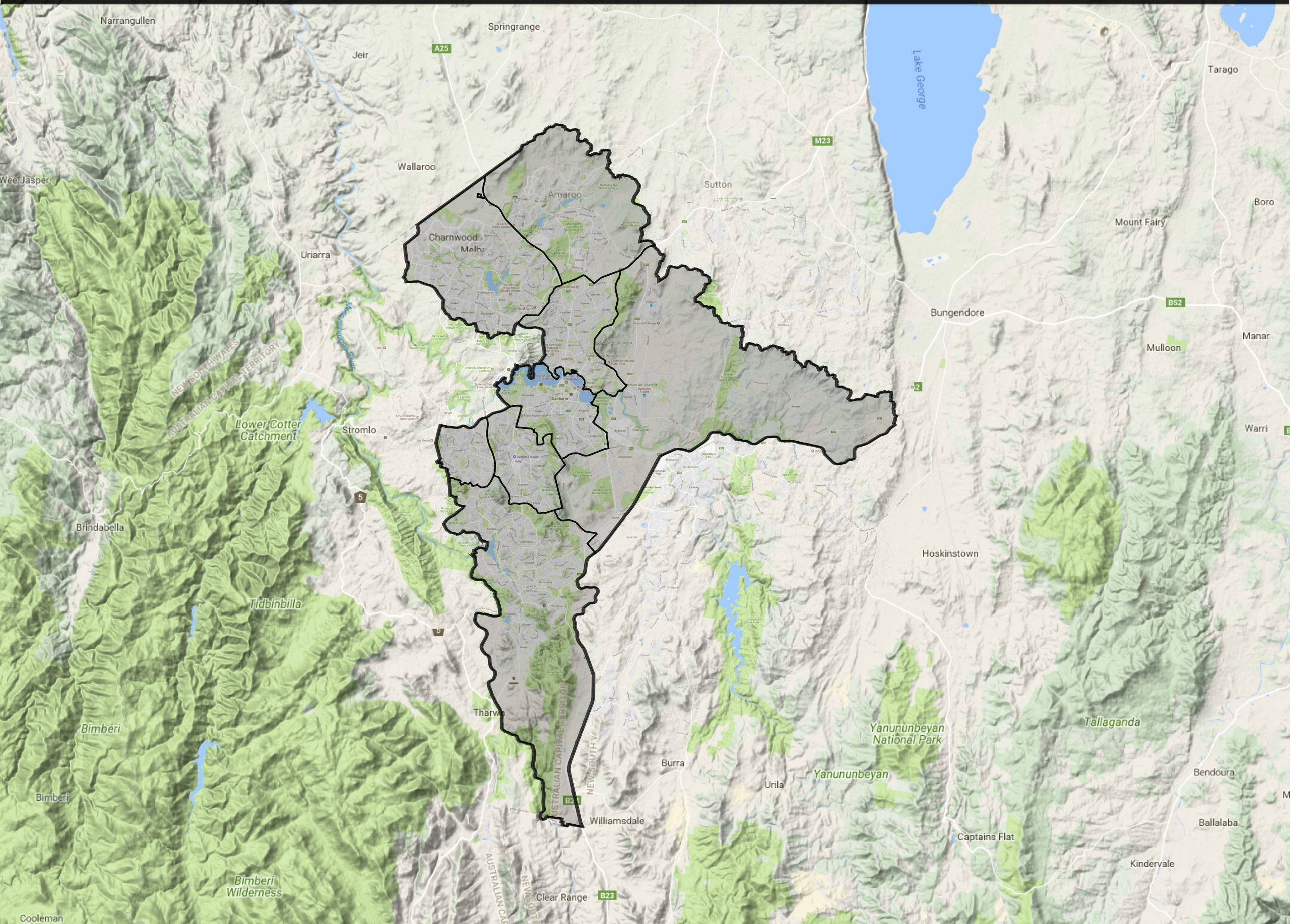
```
# 1) LOAD in shapefile from disk
# dsn = the location of the folder
# layer = the prefix for the four files.

This means that: SA3_2011_AUST.dbf, SA3_2011_AUST.prj,
SA3_2011_AUST.shp, SA3_2011_AUST.shx are read, processed and the
result is assigned to “sp_poly_df_AUST_SA3”
sp_poly_df_AUST_SA3 <- rgdal::readOGR(
                       dsn = "input_data/2011_SA3_shape",
                       layer = "SA3_2011_AUST")


# 2) SET CRS (Should be the same as in the *.proj file)
# the metadata stated that the file was “GDA94” found the
# EPSG by searching here: http://prj2epsg.org/search
proj4string(zp_poly_df_AUST_SA3) <- sp::CRS("+init = epsg:4283")
```

FILTER AND PLOT

```r
# 1) FILTER — get ACT data only (State Code == 8) and EXCLUDE the
# the large & irrelevant Cotter region (SA3_CODE != # 80102)

# 1.1) Create a logical vector
vct_act_only <- sp_poly_df_AUST_SA3@data$STATE_CODE == "8" &
                    sp_poly_df_AUST_SA3@data$SA3_CODE != "80102"

# 1.2) Apply the logical Vektor.
# We can use base R subsetting on spatial objects – many options!!
sp_poly_df_act <- sp_poly_df_AUST_SA3[vct_act_only, ]



# 2) PLOT
# 2.1) zap any previous plots
if (length(dev.list()) > 0) dev.off(dev.list()["RStudioGD"])

# 2.2) finally we can plot
plot(sp_poly_df_act)
```

RESULTS WITH BASE MAP

source: Google maps 'Terrain'. Layers exported from R and styled using Adobe Illustrator



SPATIAL ATTRIBUTES

Attributes of R's SpatialPolygonDataFrame

```r
# returns: "SpatialPolygonsDataFrame"
sp_poly_df_act %>% class()


# returns: 8 (number of polygons)
nrow(sp_poly_df_act)


# returns: data, polygons, plotOrder, bbox, proj4string
slotNames(sp_poly_df_act)


# returns: data.frame
sp_poly_df_act@data %>% class()


# returns: list - very complicated list of line segments
sp_poly_df_act@polygons %>% class()
```

DATA SLOT

The following was created with "sp_poly_df_act@data".

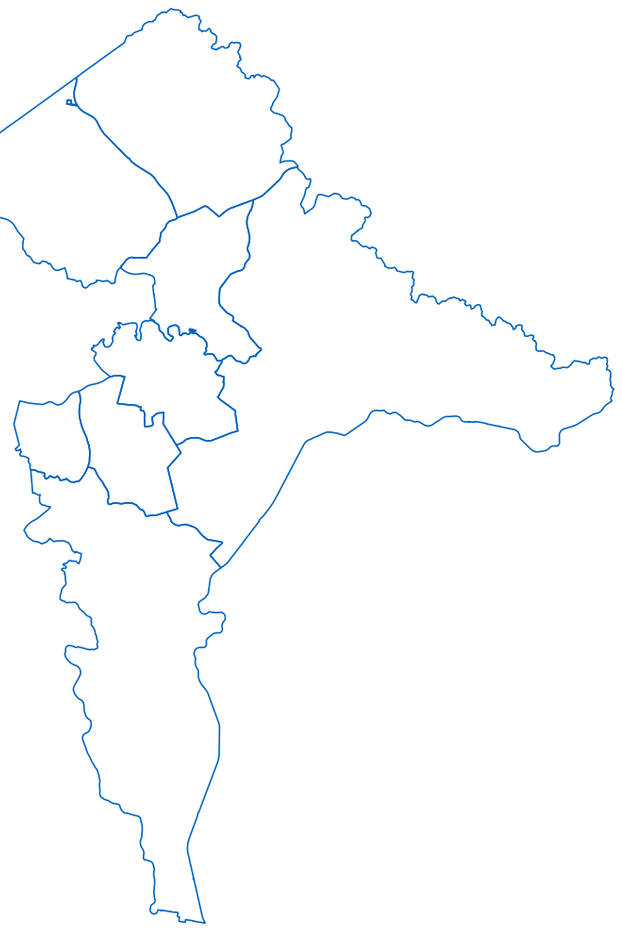
There are eight rows — one for each polygon. The left most column (i.e. SA3_Code is unique and is used to join to additional data usually obtained from CSV files.

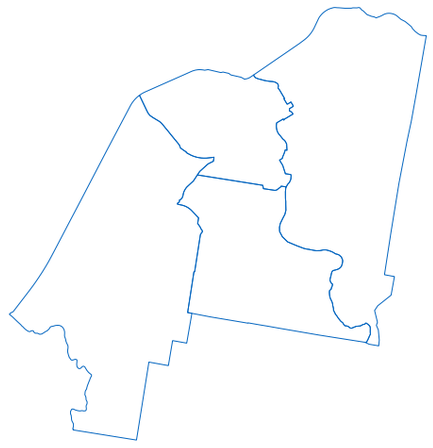
SA3_CODE	SA3_NAME	STATE_CODE	STATE_NAME	AREA_SQK
80101	Belconnen	8	Australian Capital Territory	76.98870249
80103	Fyshwick - Pialligo - Hume	8	Australian Capital Territory	221.6065111
80104	Gungahlin	8	Australian Capital Territory	90.57614952
80105	North Canberra	8	Australian Capital Territory	37.68480395
80106	South Canberra	8	Australian Capital Territory	34.64295179
80107	Tuggeranong	8	Australian Capital Territory	150.4578637
80108	Weston Creek	8	Australian Capital Territory	15.82016282
80109	Woden	8	Australian Capital Territory	28.60022906

SPATIAL VERBS

APPEND

Combine Canberra and Queenbeyan to form Greater Canberra using base::rbind()



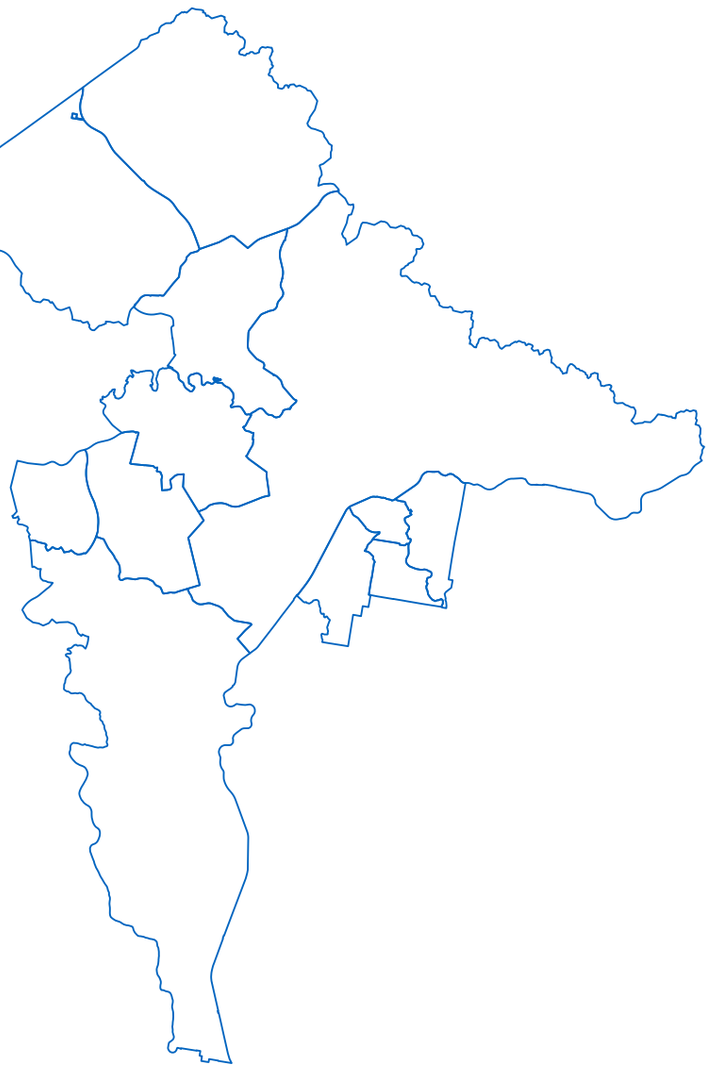


Canberra + Queenbeyan = Greater Canberra

Eight SA3 polygons *Four SA2 polygons* *8 + 4 = 12 polygons*

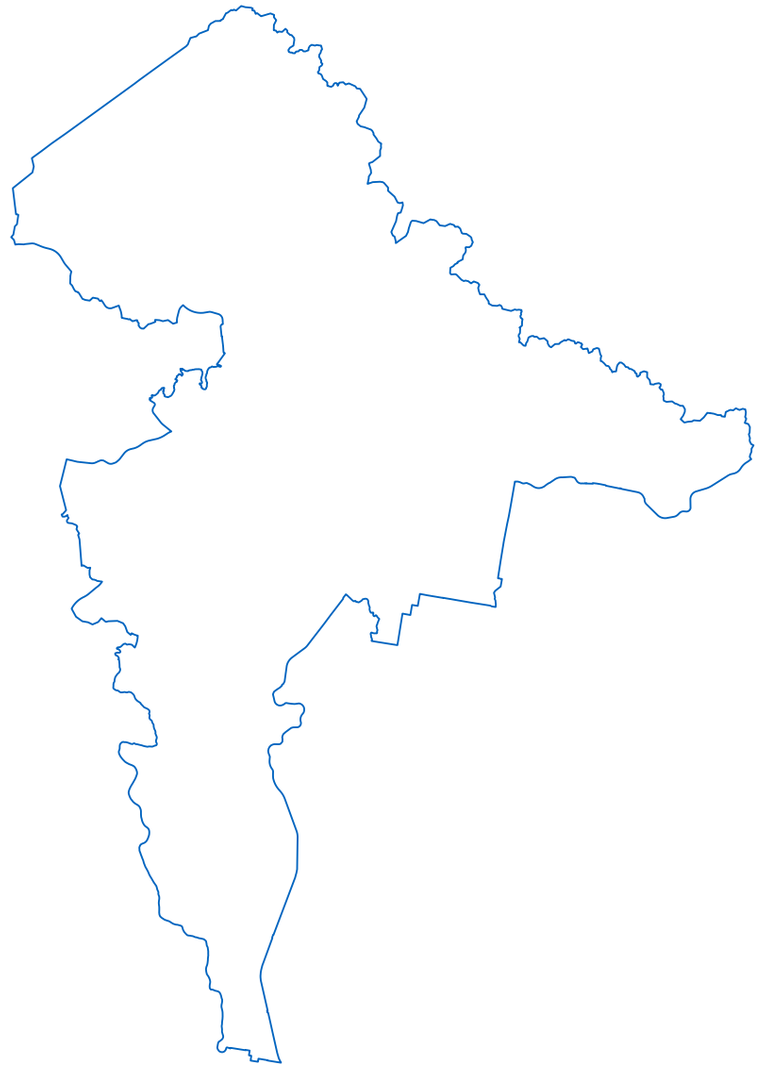
DISSOLVING POLYGONS

Dissolve 12 polygons into 1 using rgeos::gUnaryUnion



Original

12 polygons



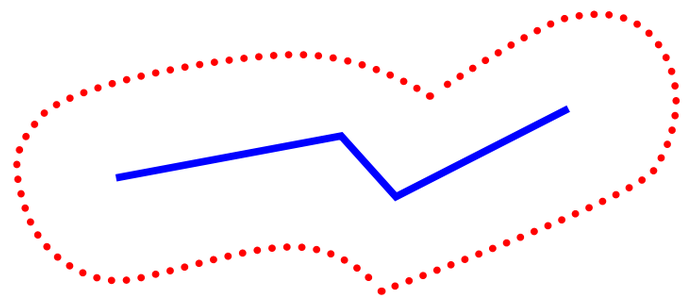
New

One polygon

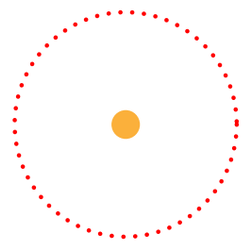
BUFFERING

Creates a polygon that surrounds existing features at a specified distance using rgeos::gBuffer. The original features need to be converted into a Projected Coordinate System for the specified distances to be meaningful.

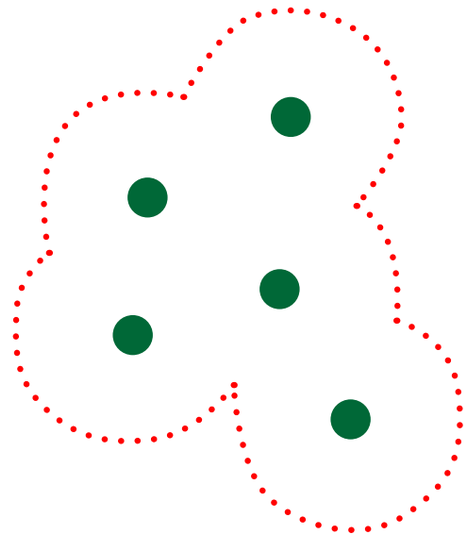
Line



Point



Multipoint

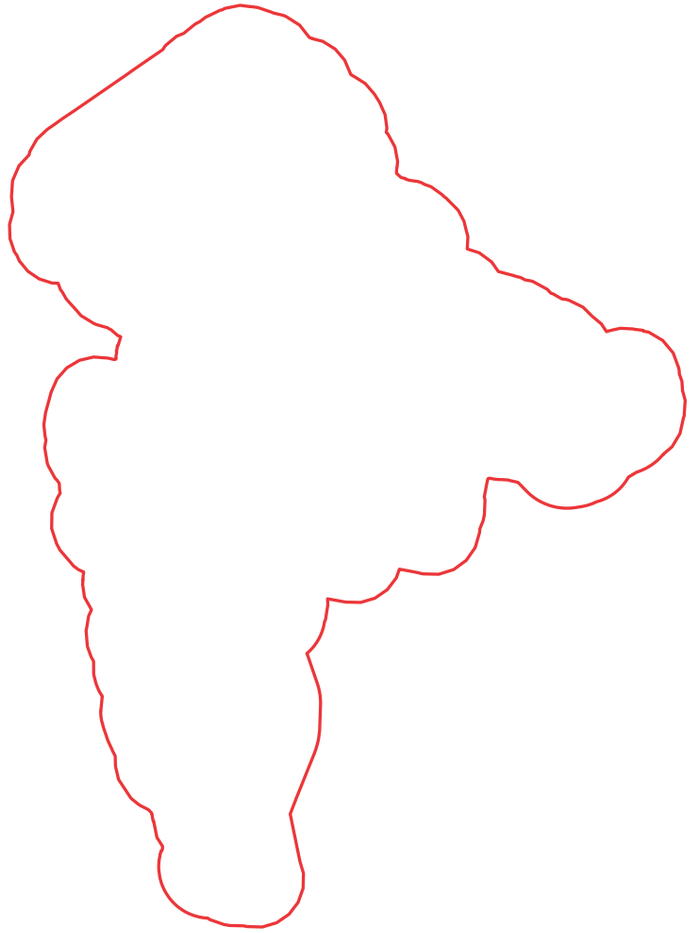


BUFFERING CANBERRA

To create a buffer the geometry needs to be projected to a Projected Geometry System (PGS) using features at a specified distance using sp::spTransform(). A buffer was then created using rgeos::gBuffer().

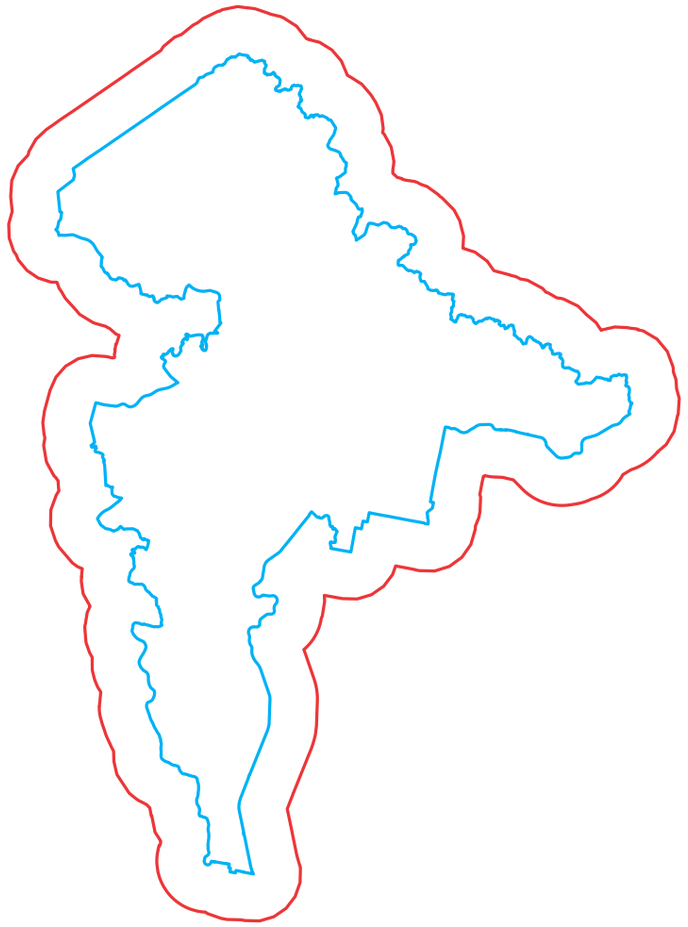
Original

Needs to be transformed to PGS



Buffer

3,000 metre buffer applied

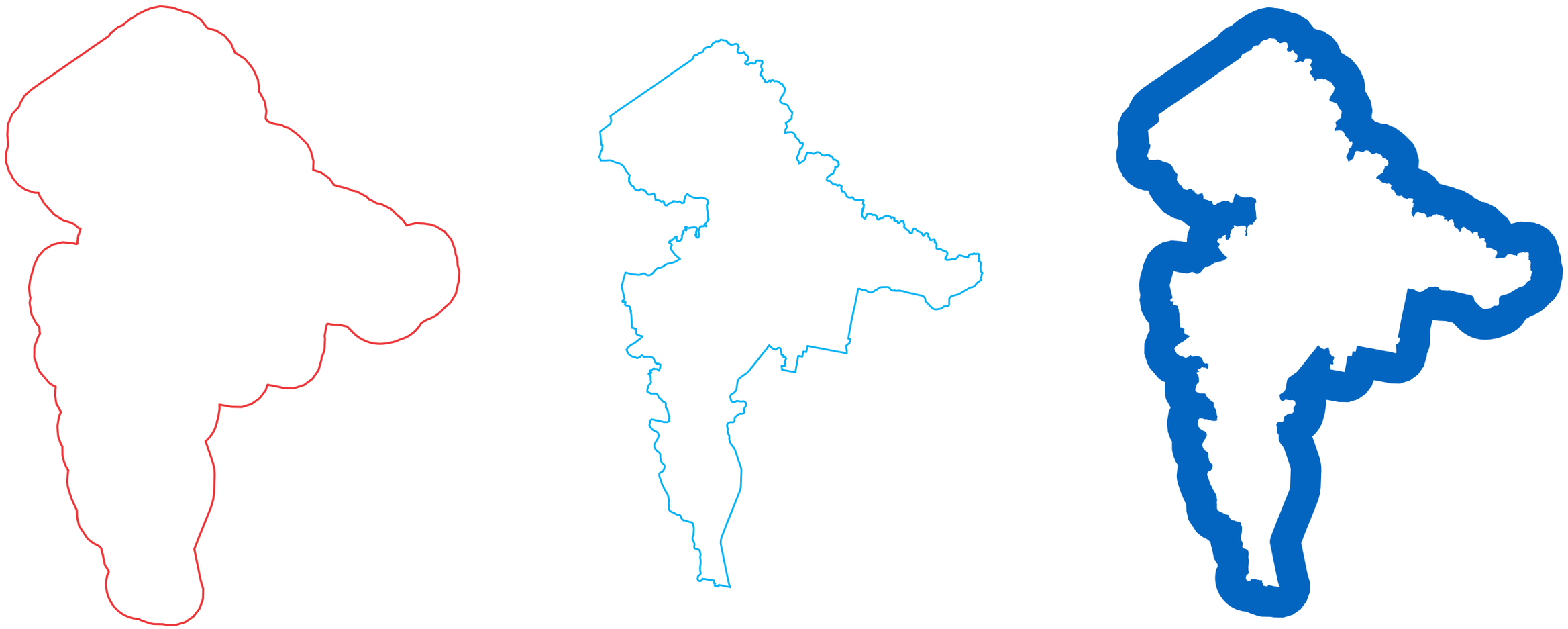


Original + Buffer

Notes:Used PGS http://spatialservices.finance.nsw.gov.au/surveying/geodesy/projections http://prj2epsg.org/epsg/3308

DIFFERENCE

Difference between the buffer and the original was created using en created using rgeos::gDifference::().

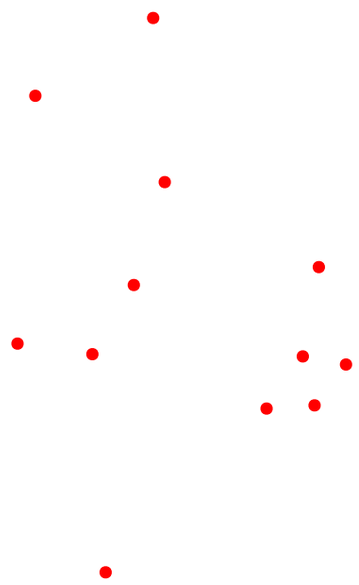


Buffer *less* Original = Difference

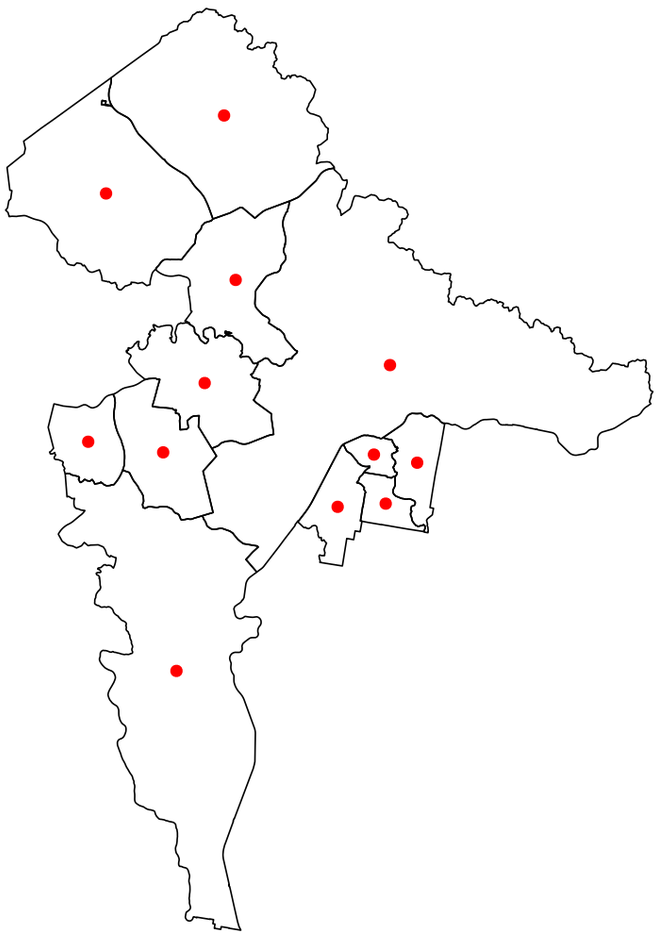
CENTROIDS

From polygons, centroids were created. Centroid coordinates were created using rgeos::gCentroid(). The results of this function were passed to the sp::SpatialPointsDataFrame() constructor.

Polygons



Centroids



Both

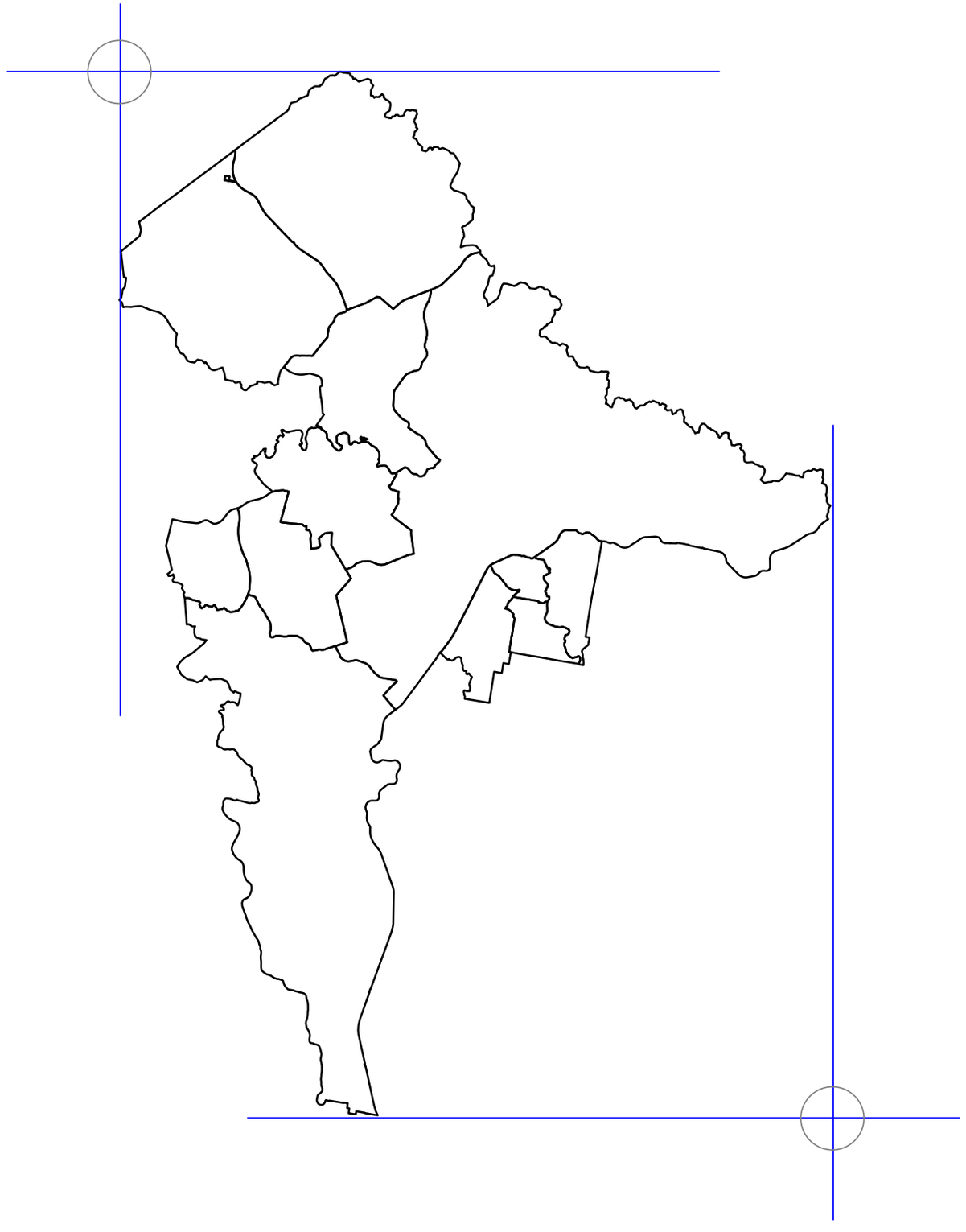
EXTENT

The spatial extent is a bounding box which is the smallest rectangular shape which totally encloses all of the referenced data.

The rectangle can can be defined by two pairs of latitude and longitude coordinates.

This rectangle can then be used to find the centre of the referenced data.

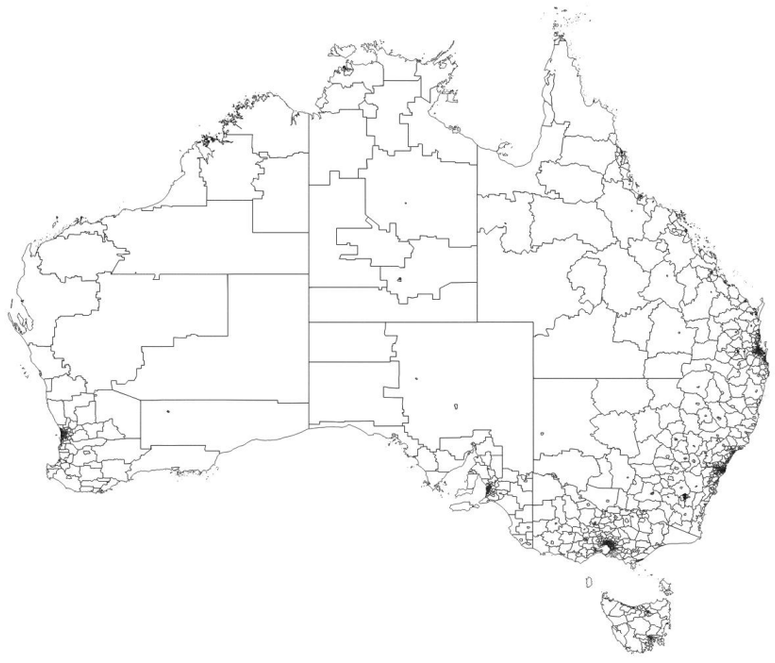
The spatial extent can be get or set using the @ operator. For example:
variable <- spatial_object@bbox



SPATIAL JOINS

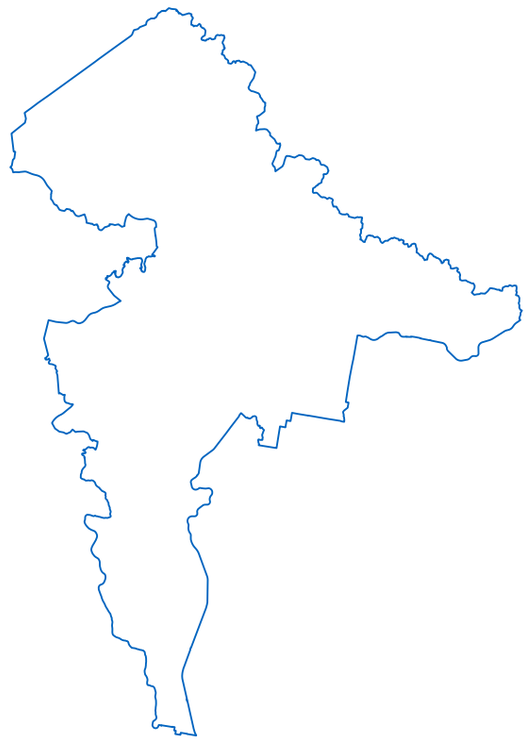
INTERSECTING POLYGONS

The Canberra region polygon (a single polygon) was used to 'cut out' 111 polygons from the 2,214 Australian SA2 polygons using the rgeos::gContains() function.



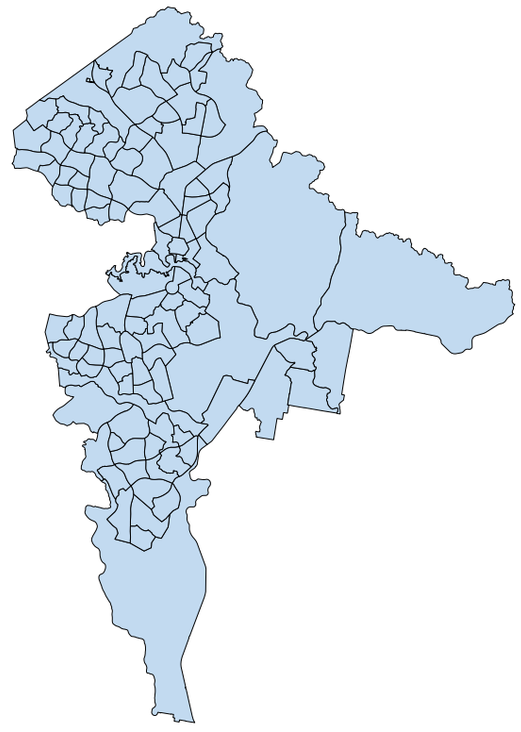
Australia *(SA2)*
2,214 polygons

intersected with



Canberra
1 polygon

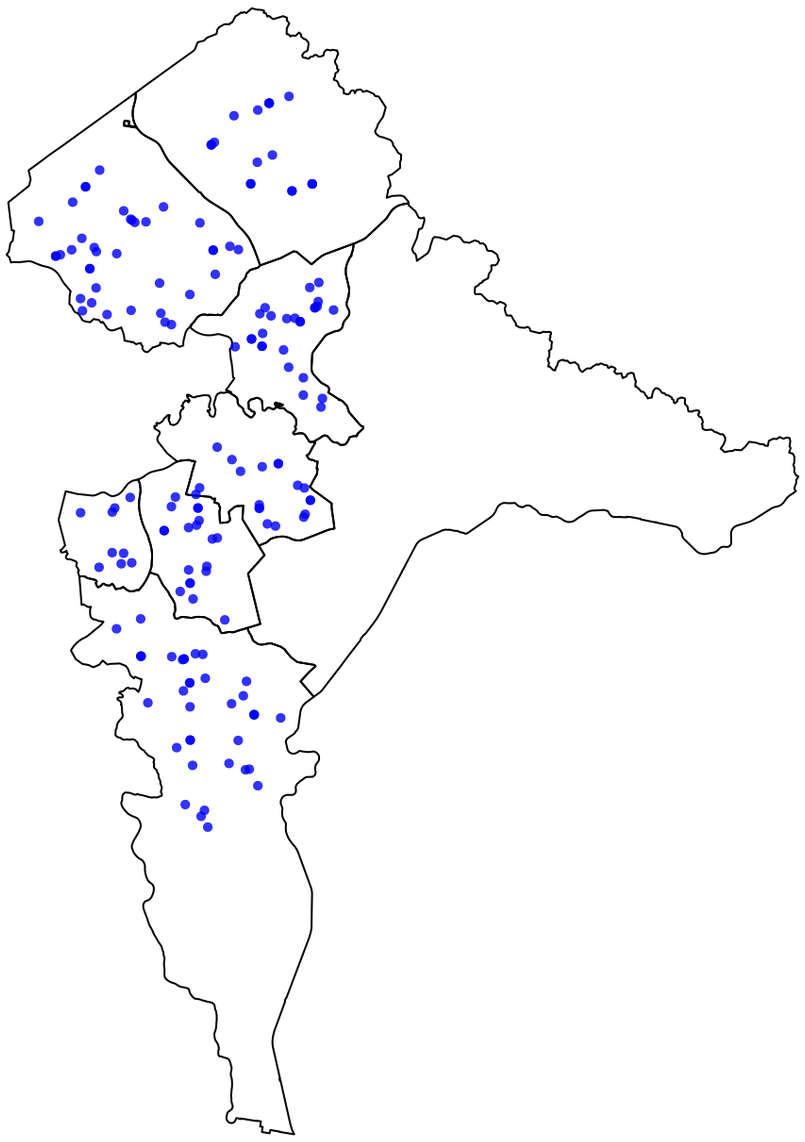
=



Canberra *(SA2)*
111 polygons

POINTS WITH POLYGONS

The point data is the location of ACT schools (i.e. point data). This was 167 observations. This point data was intersected with the SA3 census polygon data using sp::over(). Counts were then obtained using the dplyr set of functions.



Region	Count
Belconnen	38
Tuggeranong	37
Gungahlin	26
North Canberra	19
Woden	19
South Canberra	19
Weston Creek	9
Fyshwick / Pialligo	0
Total	**167**

Data source: https://www.data.act.gov.au/Education/Schools-in-the-ACT/9v9g-n5ht

GEOMETRY WITH DATA

The original data is data contained in the original shape file geometry. Additional data, presented in csv format, was joined (using the SA2_Main) primary key using dplyr::inner_join(). Only two fields were added (population 2006 and population 2011). The percentage change and five quantiles were calculated in R

SA2_MAIN	SA2_NAME	AREA_SQKM	population_06	population_11	pc_change	quantile
801061062	Deakin	3.571223	2,722	2,853	4.59%	q4
801051054	Dickson	1.577911	2,207	2,554	13.59%	q5
801051055	Downer	1.633967	3,351	3,547	5.53%	q4
801081092	Duffy	2.794046	2,907	3,182	8.64%	q4
801011006	Dunlop	3.578086	5,794	7,077	18.13%	q5
801011007	Evatt	3.060405	5,434	5,296	-2.61%	q1
801071076	Fadden	3.106416	3,124	3,005	-3.96%	q1
801091101	Farrer	2.068837	3,290	3,322	0.96%	q2
801081093	Fisher	1.582517	2,885	2,956	2.4%	q3
801011008	Florey	2.755330	5,047	4,981	-1.33%	q2

Original data Joined data Calculations

BASIC VISUALISATION

CHOROPLETH MAP

Introduction

From Greek: khôros ('area/region') + plêthos ('multitude'). Assigns polygons different colours based on the data values. To use ggplot, the ggplot::fortify() function is used to convert the polygons to line segments. For example, a 10 hexagon polygon, would become (6 x 10) 60 rows

Presentation issues

Discrete colour schemes require that data is discretised. This is achieved with evenly spaced, k - means, or quantiles. There are compromises associated with all discretisation methods.

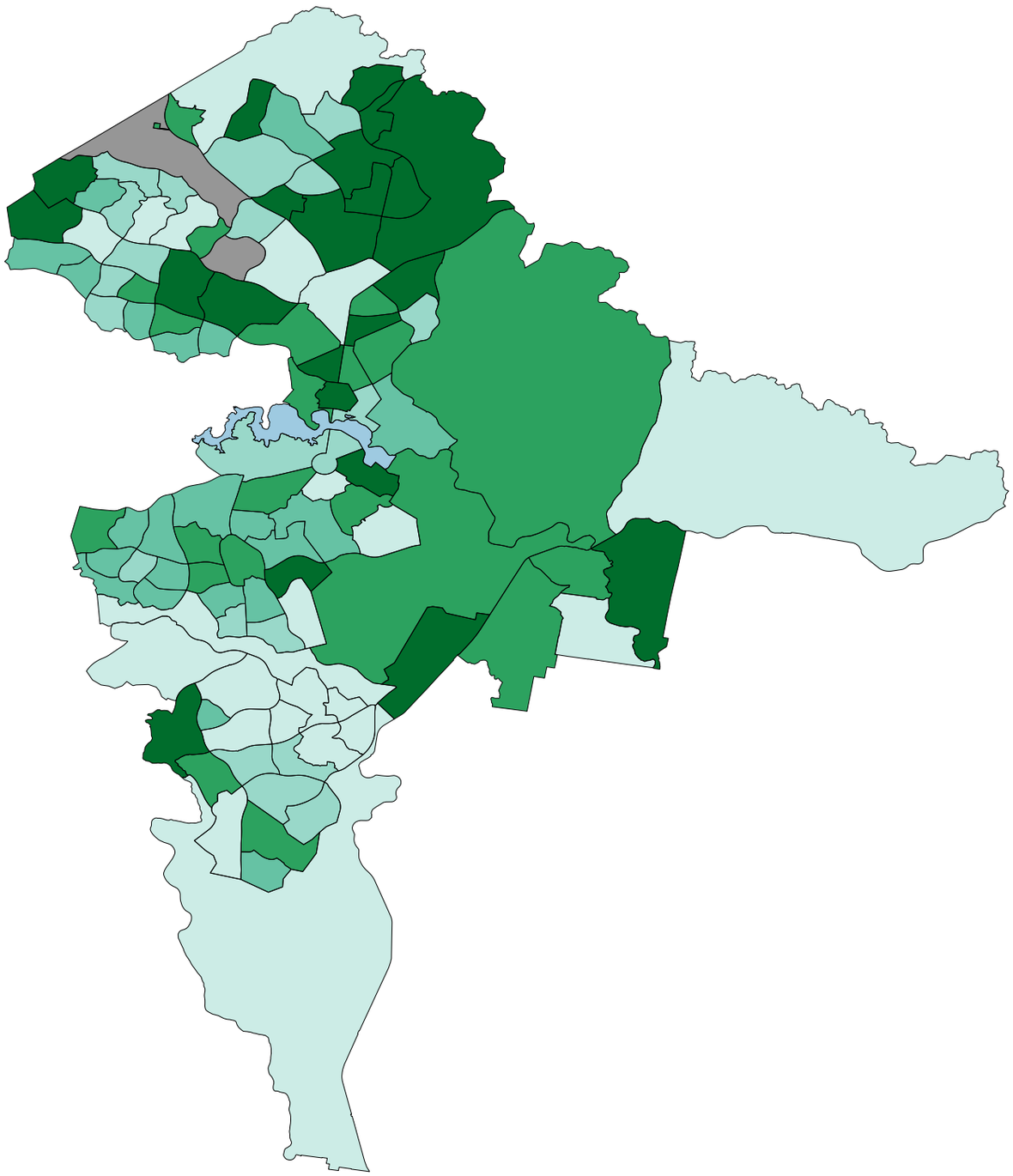
For ordinal data, a specific hue is usually used. Humans can only resolve about six shades of a specific hue. Map specific colour scales are produced such that the human eye perceives the difference from shade 2 to shade 3 as being the same as from shade 4 to shade 5.

Large polygons may represent the largest land area but not the most data. When these are assigned a specific colour, that colour may dominate the map.

Useful links: https://medium.com/@mbostock/command-line-cartography-part-4-82d0d26df0cf http://colorbrewer2.org/

CHOROPLETH MAP

Growth in greater Canberra region from 2006 to 2011. ABS SA2 aggregation level.
Plot created using the package ggplot2. Colour scale from www.colorbrewer.org



0 - 20 %

20 - 40 %

40 - 60 %

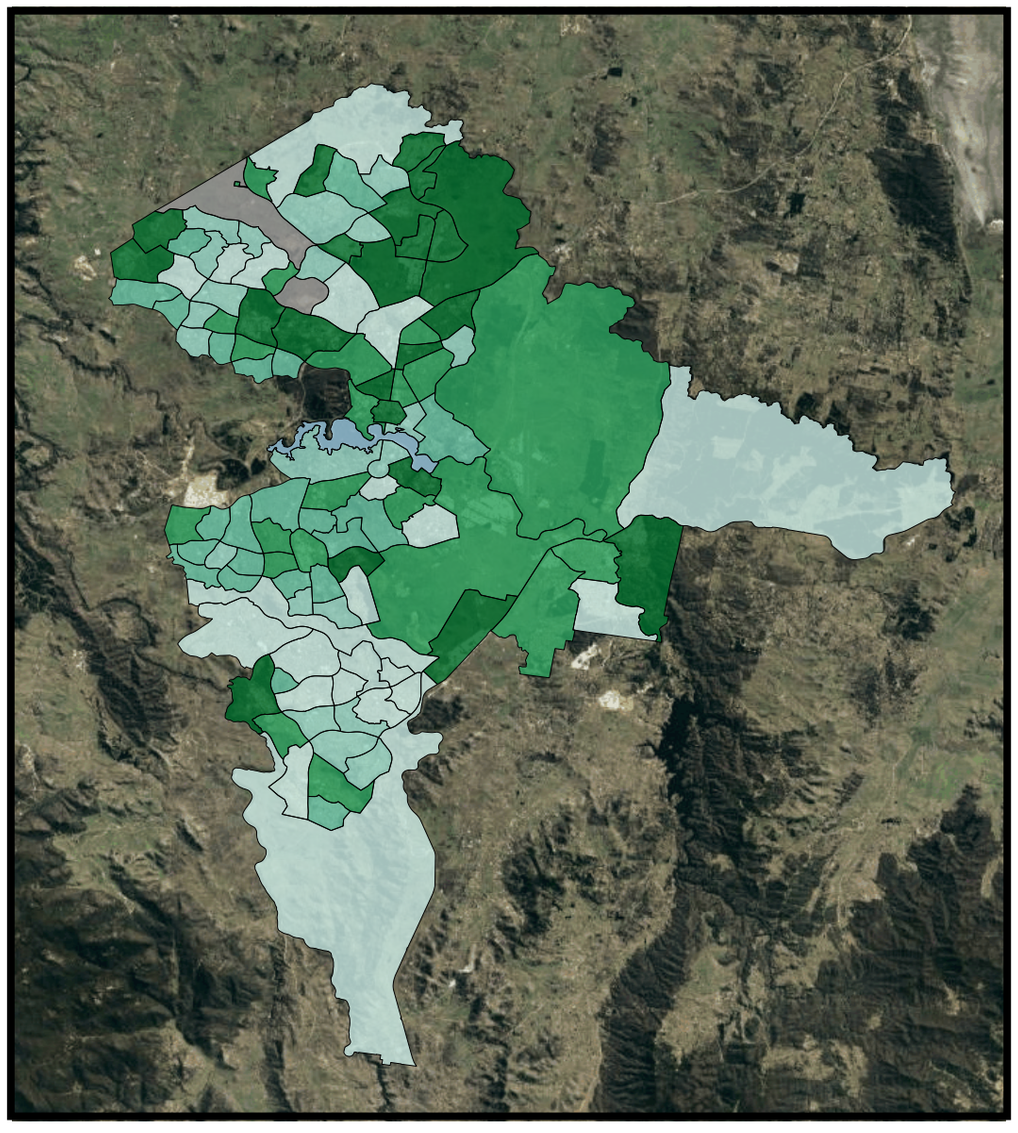
60 - 80 %

80 - 100 % *(fastest growth)*

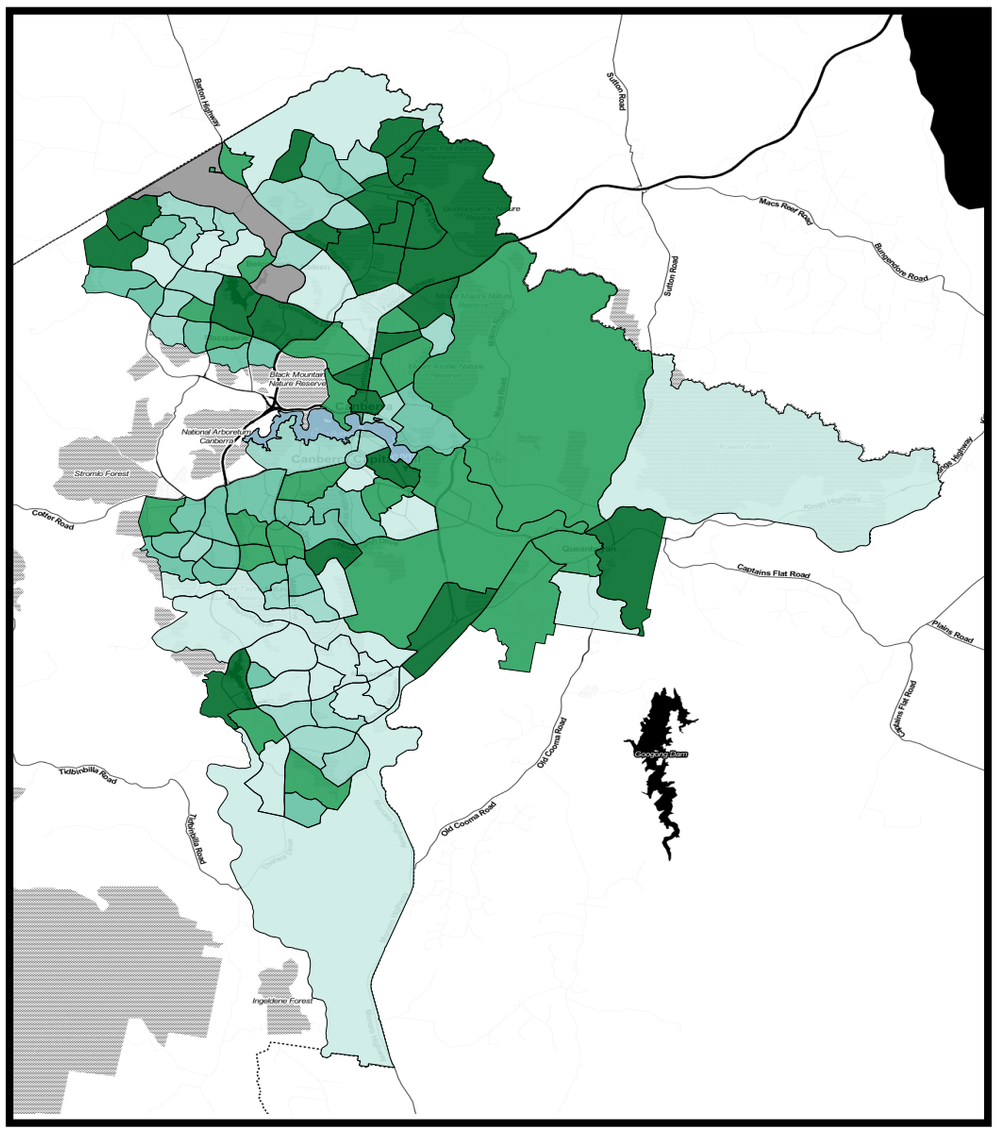
ADDING A BASE MAP

Base maps added using ggmap(). Various cloud-based base maps are available. Need to pass the map's extent (Stamen Toner); for Google maps need to pass the centre of the bounding box and a zoom level. Cloud based maps usually use WGS 84, so need to convert using sp::Transform()

Google Satellite



Stamen Toner



DATA & COMPUTER CODE AVAILABLE FROM

https://github.com/thefactmachine/spatial_presentation

END