Creating an Australian geocoded synthetic Census to calculate access to services for different groups of the population

October 29, 2026

Robert Tanton

Robert.tanton@deakin.edu.au

Outline

- Project
- Literature Review
- Method
- Validation
- Results

Project

- LIEF Grant
  - Calculate geographic access to services and facilities to help plan an equitable resource allocation for Health services
    - Create a synthetic population
    - Geocode to dwellings
    - Calculate time to services/facilities
  - Recognise that don’t have this data but essential when locating services
  - Also essential for geographic analysis using own geographies
  - Synthetic – anyone can use it with no privacy concerns
- Long term project over 2024 and 2025

Literature Review

- Synthetic Populations

Literature Review – Countries using synthetic populations

- Germany (Munnich et al, 2021 - Mikrosim)
- UK (Wu et al, 2022)
- US (Whitworth, 2022)
- Canada (Spielauer et al, 2013)
- American Samoa (Xu et al, 2017)
- Australia (Mokhtrain et al, 2014; Tanton et al, 2011)
- Bangladesh (Rubinyi et al, 2022)
- Namibia (Thomson et al, 2018)
- Austria (Templ et al, 2017)

Literature Review - Methods

- Sample based and sample free synthetic populations
    - Sample based needs a starting sample
        - Synthetic population contains people from the sample
    - Sample free is purely synthetic
        - No confidentiality risk from sample
- Decided on sample free so anyone can use

Method

- Stages
  - Derive full synthetic population for SA2 based on limited aggregate data from Census
    - Separate Household, Family and Person files
    - Assign people to families based on family type and age
      - Use rules, eg, not more than 10 years apart initially for marriage
    - Assign families to households
  - Assign to dwelling register
    - Enhanced G-NAF using ABS address register
    - Currently random by SA2 but can add rules when have more information

Validation

- Against aggregate SA2 data
  - This is the data we started with
- ABS is partner on grant and talking about individual level validation against Census data
  - Waiting for University PLIDA access

Validation – Total Population

- Absolute value of difference to Census data by SA2
- Because have at SA2 level, can show distribution in boxplot
    - Median (50%), 25%, 75%, extremes

Validation – by Age and Gender

Validation – Number of households and families

Validation – Family type and number of people in family

Number of dependent and non-dependent children in family

Number of people in household

Overall Validation

- Absolute difference as a % for each SA2
- Averaged to get difference across Australia

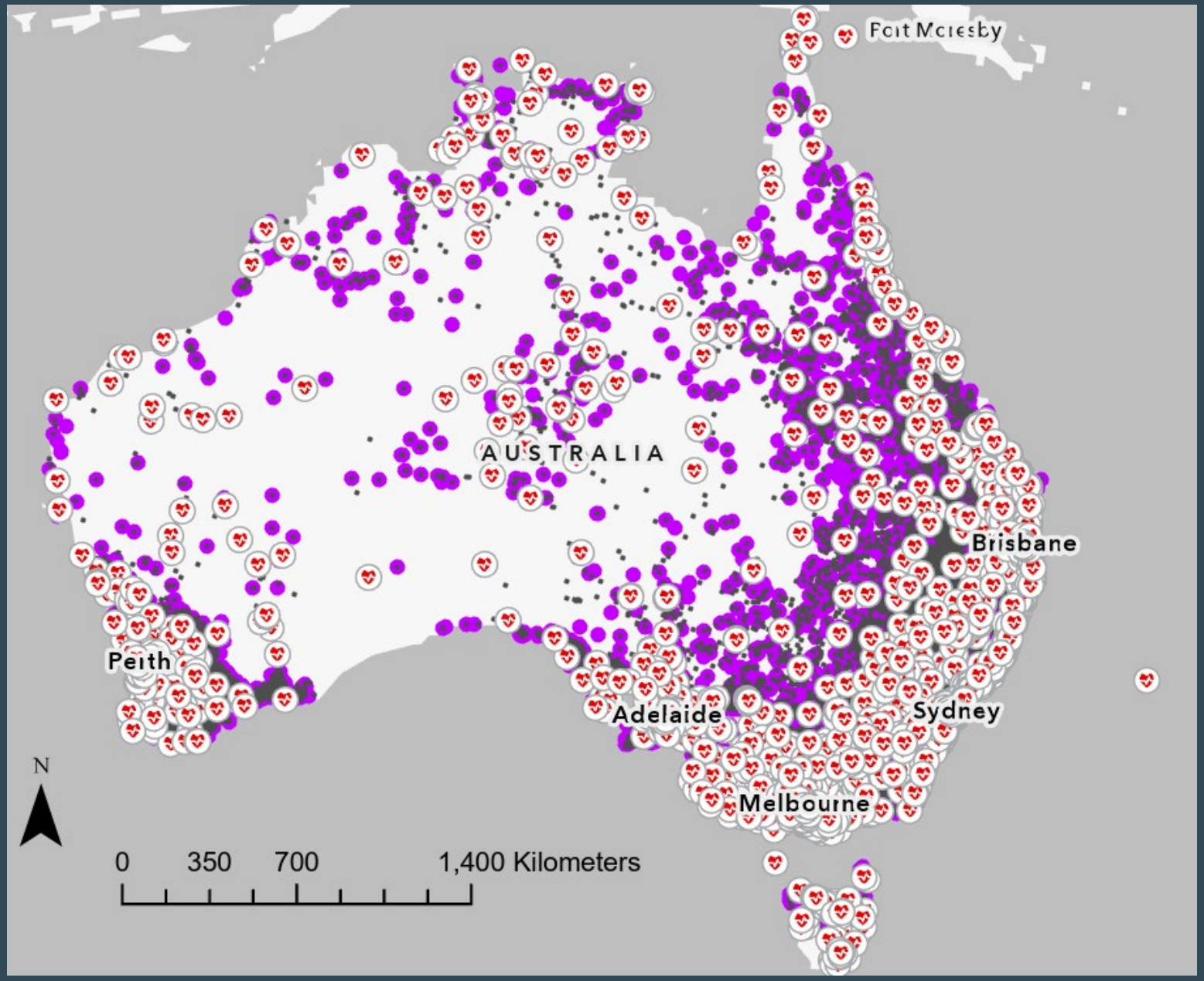
Benchmarking

- Know total number of households in SA2
- Geocoding to G-NAF better if HH total in SynPop matches G-NAF
- Force number of HH in SynPop to known SA2 HH counts

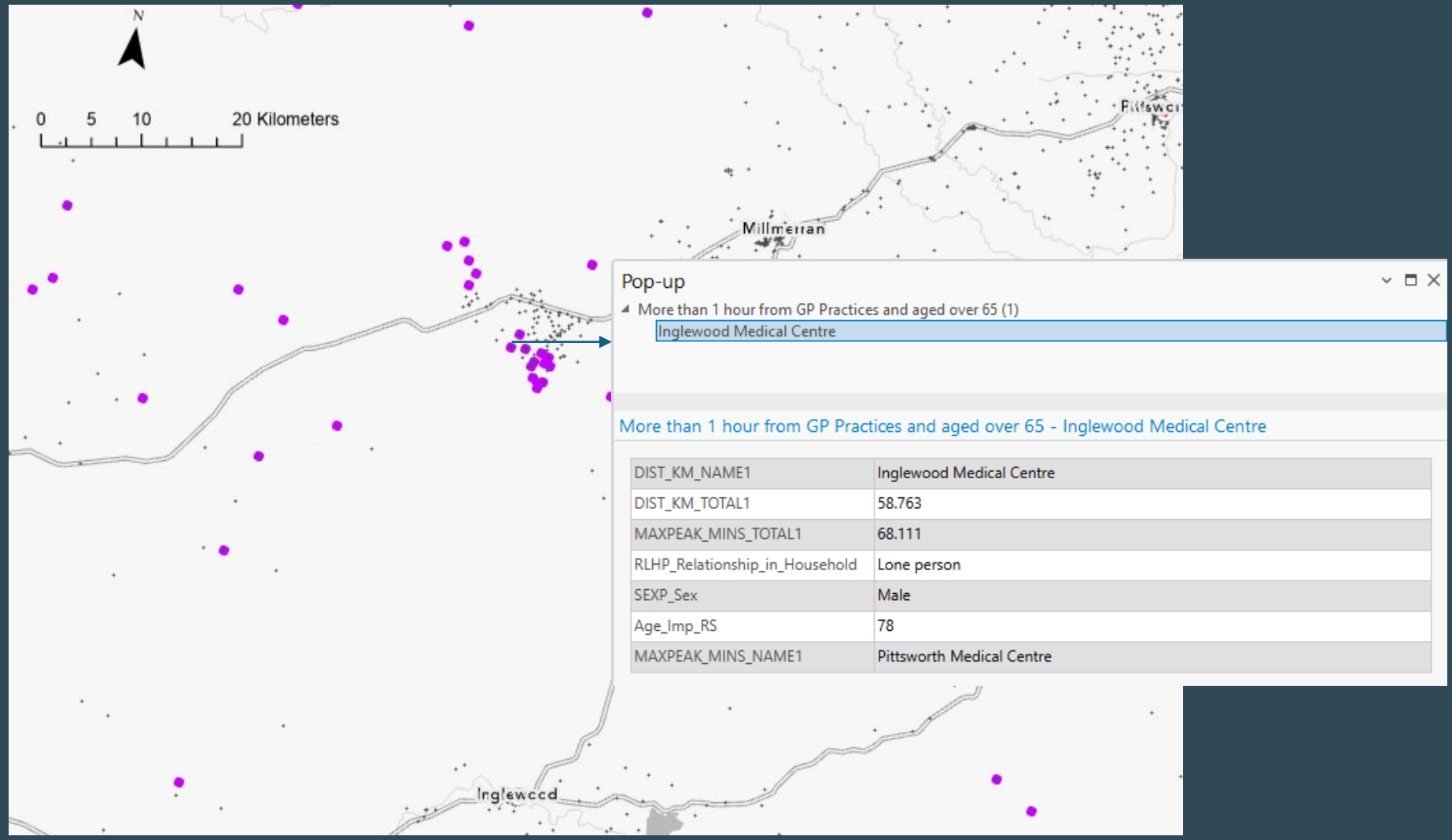
Time to services

- Travel time on road network
- Conservative assumptions about speed if don’t have speed

Time to services



Results - Millmerran



Further work

- Non Private Dwellings
  - Prisons/nursing homes have 700 people/families/households in 1 building
  - Not a problem for hospitals, hotels as using usual residence data
  - Is a problem for nursing homes, prisons, etc
- Same sex couples
  - ABS has the data on Census but need to incorporate into SynPop
- Impute other variables using microsimulation probabilistic methods
- Working with ABS to validate and refine
- Add policy microsimulation models – need unit record data

Policy Modelling

- Microsimulation models need unit record data
- We have geocoded unit record data
  - Tax/Transfer model
    - Impact of changes in tax rates at household level
  - Housing model
    - Impact of changes in CGT at dwelling level

Lessons

- Is possible to get a synthetic census with good results
- Won't be exactly the same as Census – don't want it to be
  - We have the Census already
  - If want SA2 level data, use the Census
- Is useful for calculating distance and time to residences of particular groups of the population
  - Need to aggregate as do have model error and isn't actual Census data
- Can be used to link to other microsimulation models, eg, tax/transfer
  - Get policy change impacts to households and aggregate to small areas
- Spatial analysis where need smallest area possible
  - Cluster analysis, impact of pollution, reducing MAUP
- Other uses?

Thanks

- LIEF Grant - LE220100028, “A national e-infrastructure for high resolution population spatial modelling”
- Centre for Australian Research into Access, Deakin University
  - Vin Versace, Neil Coffee, Marcus Blake, Sam Quinsey

Literature

- Mokhtarian, P., & Perez, P. (2014). *Generating a Dynamic Synthetic Population – Using an Age-Structured Two-Sex Model for Household Dynamics*. *9*(4). https://doi.org/10.1371/journal.pone.0094761
- Münnich, R., Schnell, R., Brenzel, H., Dieckmann, H., Dräger, S., Emmenegger, J., Höcker, P., Kopp, J., Merkle, H., Neufang, K., Obersneider, M., Reinhold, J., Schaller, J., Schmaus, S., & Stein, P. (2021). A Population Based Regional Dynamic Microsimulation of Germany: The MikroSim Model. *Methods, Data, Analyses*, *15*(2), 241–264. https://doi.org/10.12758/mda.2021.03
- Rubinyi, S., Verschuur, J., Goldblatt, R., Gussenbauer, J., Kowarik, A., Mannix, J., Bottoms, B., & Hall, J. (2022). High-resolution synthetic population mapping for quantifying disparities in disaster impacts: An application in the Bangladesh Coastal Zone. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.1033579
- Spielauer, M., Hicks, C., Gribble, S., Rowe, G., Lin, X., Moore, K., & Plager, L. (2013). *The LifePaths Microsimulation Model: An Overview* (pp. 1–33). Stats Canada.
- Tanton, R., Vidyattama, Y., Nepal, B., & McNamara, J. (2011). Small area estimation using a reweighting algorithm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(4), 931–951. https://doi.org/10.1111/j.1467-985X.2011.00690.x
- Templ, M., Meindl, B., Kowarik, A., & Dupriez, O. (2017). Simulation of synthetic complex data: The R package simPop. *Journal of Statistical Software*, *79*(10). https://doi.org/10.18637/jss.v079.i10
- Thomson, D. R., Kools, L., & Jochem, W. C. (2018). Linking synthetic populations to household geolocations: A demonstration in Namibia. *Data*, *3*(3), 1–19. https://doi.org/10.3390/data3030030
- Whitworth, A. (n.d.). *synthACS: Spatial MicroSimulation Modeling with Synthetic American Community Survey Data*.
- Wu, G., Heppenstall, A., Meier, P., Purshouse, R., & Lomax, N. (2022). A synthetic population dataset for estimating small area health and socio-economic outcomes in Great Britain. *Scientific Data*, *9*(1), 1–11. https://doi.org/10.1038/s41597-022-01124-9
- Xu, Z., Glass, K., Lau, C. L., Geard, N., Graves, P., & Clements, A. (2017). A Synthetic Population for Modelling the Dynamics of Infectious Disease Transmission in American Samoa. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-17093-8

Questions

- Robert Tanton
- Robert.tanton@deakin.edu.au