

Public Health Summer School Wellington



University
of Otago
ŌTĀKOU WHAKAIHU WAKA

Pōneke
Wellington
Campus

GIS techniques in health research and policy — **DAY 1**

Te whakamahi GIS (Pūnaha Mōhiohio
Matawhenua) i roto i te rangahau
hauora

Day 1, 16 February 2026

Welcome announcements

- Useful information (in-person courses)
 - Orientation – bathrooms, lunch & refreshments
 - Website materials
(www.spacehealthandcommunitylab.org)

Emergency Evacuation

Mein Street

Link Block Levels 4, 5, 6, 7, 8 and 9 (D, E, F, G, H & J) through Vibe (The Health Deli) Café.

WSB levels 5,7,8 and 9 (E, G & J) Deli) Café.

Lecture theatre (Nordmeyer)

Ambulance

Ask Otago and Level 3 (C) link carpark

Exit through to WSB, use the stairs down to level 3 and exit

Use the stairs down to level 3 and exit through Vibe (The Health

Exit down the stairs and use the fire exit opening out to the exterior of the building and report to carpark near bay.

Exit through the main entrance by Askotago and report to near Ambulance bay.

Prefabs

Exit through doors and gather in the main car park at front of building.

Hanson Street

Exit through doors and gather on street at front of building.

Welcome announcements

- Keeping to time
- Evaluation forms
- Certificates of attendance
- Special room requirements - computer lab
- Introductions

Considering postgraduate study?

- Talk to us about the options available at our 3 campuses: Wellington, Christchurch, Dunedin
 - Certificate in Public Health
 - Diploma in Public Health
 - Master of Public Health
 - PhD
- To find out more, ask at the registration desk or visit otago.ac.nz/publichealth



Access UO_GUEST

UO_GUEST is the network available for use by visitors to the University of Otago who do not have eduroam access, and provides internet access.

How do I access the UO_GUEST network

To access the UO_GUEST network:

1. Select the **UO_GUEST** network from the list of available networks on your device. A webpage should open automatically with the login screen. If not, open a web browser manually and the page should open.
2. Click on the *Don't Have an Account?* link on the login screen and follow the prompts to create a new account. If you already have an existing UO_GUEST account, enter your login details here. You will be required to provide a working email address and contact phone number.
3. Your UO_GUEST credentials will be emailed to the account provided. You will need to be able to access your email to collect your password. Once you have received your username and password, click the *Sign On* button.
4. You are required to read and accept the University network regulations. You will be unable to access the internet until this is done.
5. After you click *Accept* you will be redirected to the ITS Visitor Network Access page, and you should now be able to access the internet.

Academic Journey

From Pacific Northwest to New Zealand to the Great Lakes

From foundational training in spatial analysis to pioneering novel exposure assessment methods

Methodological innovation addressing critical gaps in environmental health research

Personal Journey

1
PhD



University of
Laid the foundation in geospatial analysis and environmental health, developing expertise in spatial statistics and exposure assessment methodologies.

2
Postdoc



University of
Canterbury
Developed the **Vertical Visibility Index (VVI)**, began pioneering work with Google Street View for neighborhood neighborhood assessment, focus on urban resilience.

3
Research Fellow



University of Otago
Expanded research program integrating GPS, wearables, and population-wide analyses. Advanced multi-sensor approaches for comprehensive exposure assessment.

4
Professor



Michigan State University
Lead large teams, longitudinal and experimental studies, developed best practices for integrating GPS, wearables, and soundscape mapping. Funded research on neighborhood exposome across Michigan (100 years).



Met husband and transcontinental move to Michigan, where our daughter was born on the 10-year anniversary of the Canterbury earthquake—a poignant reminder of how place, health, and intersect.

Measuring Geospatial Exposures

Move through environments and exposures



Movement Through Activity Space & Environmental Exposures

Viewshed and soundscape mapping to assign exposures

Develop and validate algorithm for determining time (and locations) spent outdoors using GPS + accelerometer data

Measuring Geospatial Exposures

Understand and quantify behaviors using wearables



Refine best practices for use of GPS in health research
Develop tool for linkage of wearable data [in progress]

Measuring Geospatial Exposures

Measure changing environments and neighborhoods over time – not static!



Compare static (residential location) versus activity space exposures and health
Understand and quantify historical neighborhood processes

Programme

Time	Session
9.00 am	S1: Welcome & Overview
9.30 am	S2: Data types, sources, units, map basics, storage, geodatabases
10.30 am	<i>Morning break</i>
11.00 am	S3: Overview of ArcGIS Pro
11.45 pm	Tutorial 1 – making a map
12.30 pm	<i>Lunch break</i>
1.30 pm	S4: Creating spatial data
1.45 pm	Tutorial 2 – geocoding
2.15 pm	S5: Geographic definitions & techniques
3.00 pm	<i>Afternoon break</i>
3.30 pm	S6: Defining neighborhoods, applications & health ethics
4.45 pm	Discussion/questions
5.00 pm	Finish

Welcome! Haere mai!

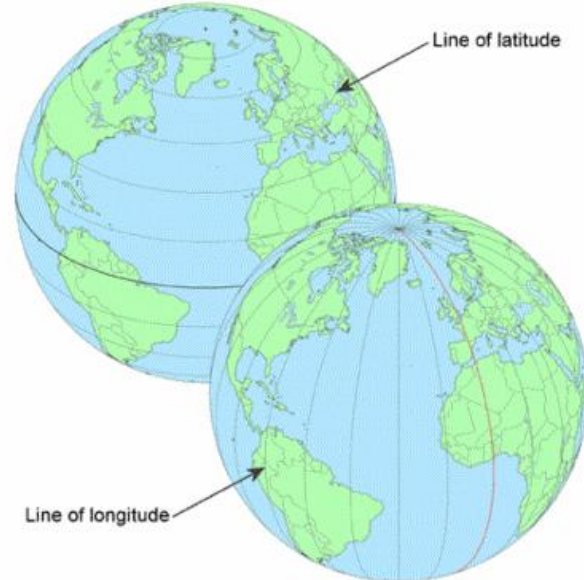
- Goals for today, to:
 - Understand what spatial data are and where to find them
 - Explore map making
 - Think of ways to apply these tools and spatial thinking to the work you already do
 - Consider the power of maps and ethics in our health research
 - Provide resources for further learning and data

Software used in the course

- ArcGIS Pro
- Excel
- R coding (optional)

Section 1: Overview

- Fixed, measurable point on the earth's surface
 - Census area unit, address, coordinates of lat & long
 - 23 Mein Street, Wellington 6242
 - 4042'42"N, 7400'45"W

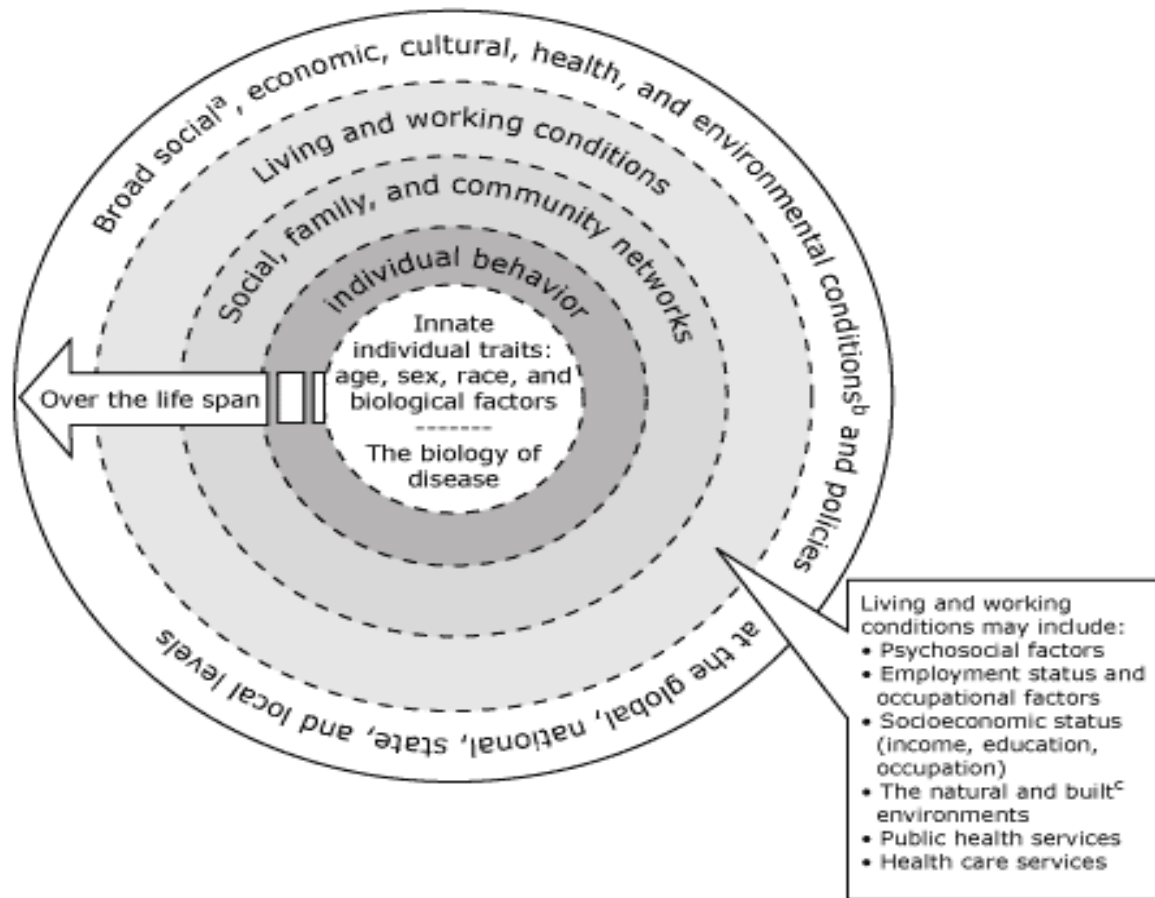


Scale

- The level of resolution with which we view the world
- Definitions of these vary widely
 - Body
 - Household
 - Neighbourhood
 - Community
 - Village
 - City
 - Region
 - Nation
 - Globe



Scale – socio-ecological model of health



Importance of scale

- You might want to know about children's exposure to gambling outlets in their neighbourhood.
- How do we define 'neighbourhood'?
- What is the appropriate scale?
 - Meshblock?
 - Census area unit?
 - SA1?
 - SA2?
 - 500m around home? As the crow flies? Along the footpath?

What is GIS?

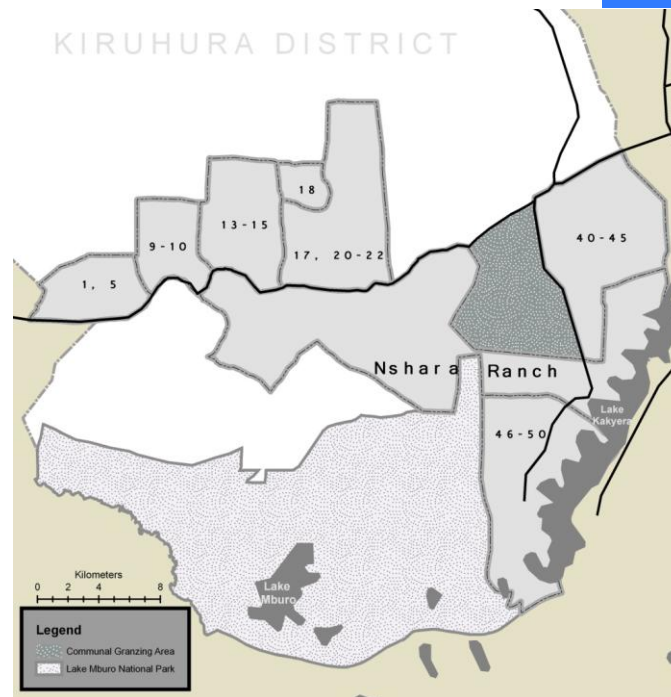
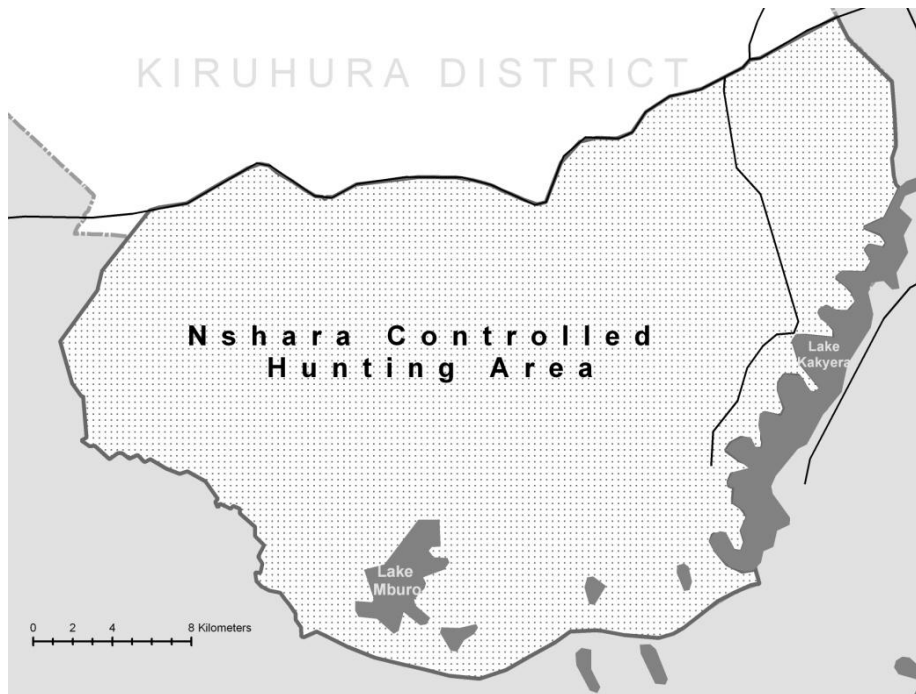
- Geographic Information System
- Both graphical display of spatial features and underlying tabular data about those features
- Most commonly – ArcGIS/ArcPro software (ESRI)
 - But other free software (TBD)

What is GIS?

- A way of using location information to:
 - Record and store information
 - Display
 - Present information and communicate findings
 - Measure spatial relationships
 - locational distributions and spatial patterns
 - geographic covariates

- Examine trends in influenza over time
- Early warning signs of water-borne outbreak; or predictions for the future
- Air pollution exposure estimation
- Demonstrate health inequalities across Aotearoa/New Zealand

Record and store information



Record and store information

Dearborn

Michigan

Group(s) Excluded

Black
Jewish

Comments

Orville Hubbard, former mayor of Dearborn, was vocal and well-known in his successful attempts to keep Dearborn "clean", meaning "white". In a 1956 interview with an Alabama newspaper Hubbard stated "They can't get in here. Every time we hear of a Negro moving in... we respond quicker than you do to a fire." A 1956 US News & World Report article reported that the city shut off the gas of a black man who moved to Dearborn and refused to collect his garbage.

Blacks were allowed to reside in Dearborn as live-in domestic workers, but not as independent households. Dearborn's black population through 1970 is mostly female.

In the mid 1960s, the NAACP staged a series of fair housing marches in suburban Detroit, with the first march taking place in Dearborn. Marchers were pelted with rotten eggs and rocks and called numerous racial slurs.

In 1985, Dearborn passed an ordinance limiting use of the public parks to Dearborn residents only. This was widely seen as a measure to keep the parks all white.

Basic Information

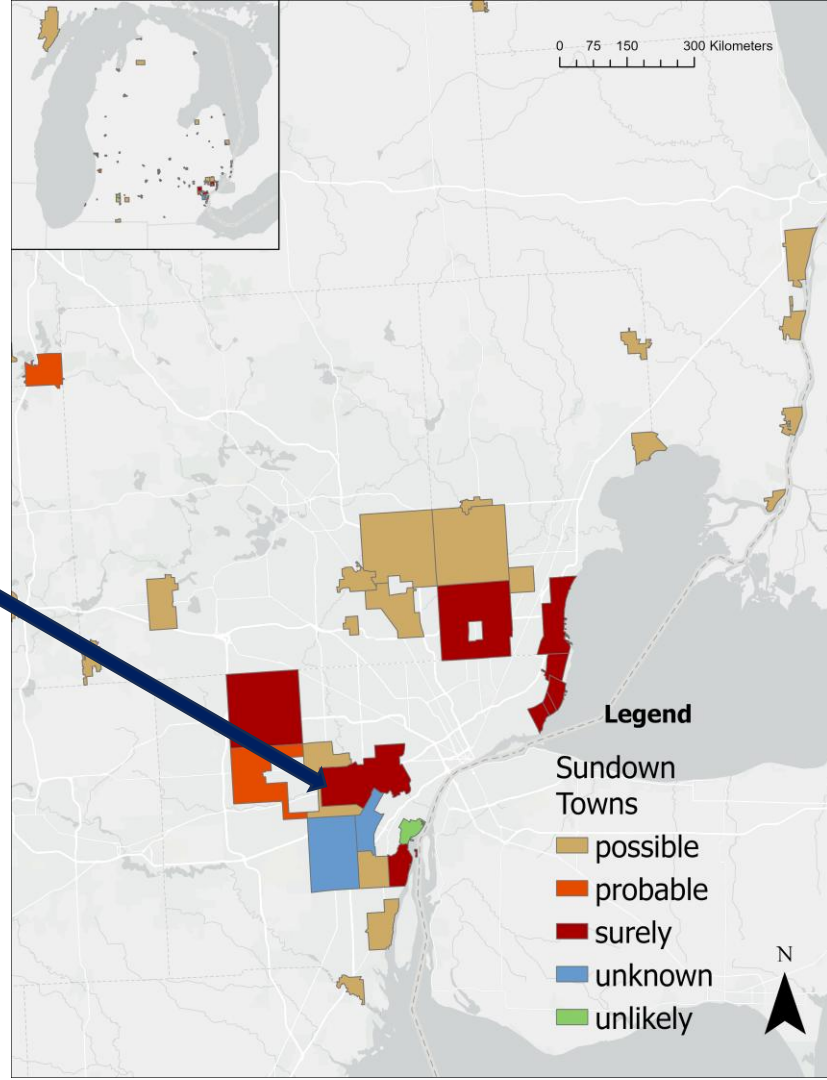
Type of Place
Metro Area
Politics c. 1860?
Unions, Organized Labor?

Sundown Town Status

Sundown Town in the Past?
Was there an ordinance?
Sign?

Suburb
Detroit

Surely
Yes, Strong Oral Tradition
Don't Know



Analyse location and patterns - Day 2

- 1854 - London
- Cholera outbreak
- Unknown source
- Mapped locations of patients & water pumps
- Intervention



Generation of geographic covariates -

Day 2

- Key concept: where we live, work and play affect our lives and our health
- We often extract information from GIS to be used in other software programs for statistical analyses (eg, regression modeling)
- This gives us information about places or spatial patterns
- Helps us characterise exposures or 'dose'

Generation of geographic covariates

- As epidemiologists or public health practitioners
 - We often think about exposures at the individual-level (eg, genetics, behaviour)
 - GIS helps us assess exposure for individuals at larger scales (eg, air pollution)
 - Based on where they live, work, attend school
 - Based on where they spend their time (GPS tracking data)

Generation of geographic covariates

- Generate variables which describe:
 - Neighbourhood context
 - Work environment
 - Exposure to environments (built, physical, social)
- Variables are then used to:
 - Assess exposure (assign to individuals or to areas)
 - Assess health (usually to areas)
- Variables used in regression analyses, for example:
 - Land use regression to predict air pollution
 - Access to greenspace and depression
 - Neighbourhood walkability and CVD

Geocovariate example

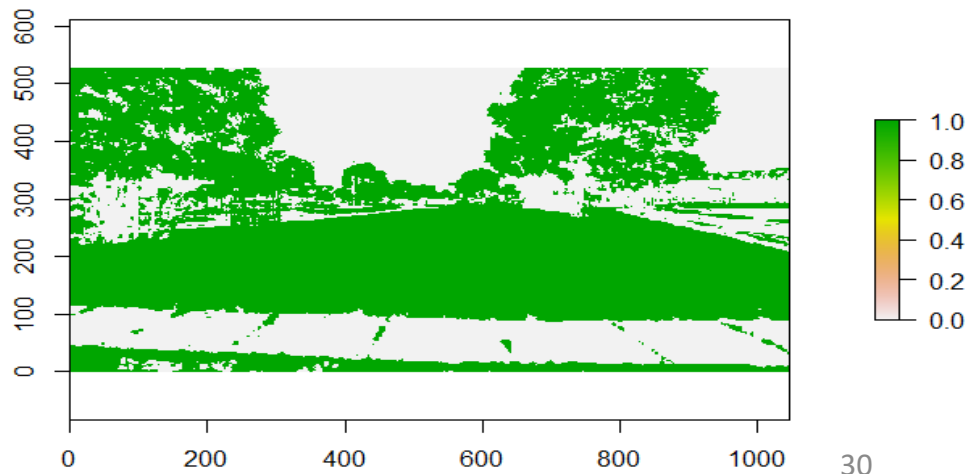
- Exposure to "greenery"
- Is it beneficial for mental health?
- How do we measure exposure to greenery?



-
- Average NDVI for buffer: 0.129**
- Legend**
NDVI
-0.99 0.99
- Copyright 2010, Center for Global Change Science, Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA. GeoCensus.org is a project of the Center for Global Change Science, Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA. GeoCensus.org is a project of the Center for Global Change Science, Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA. GeoCensus.org is a project of the Center for Global Change Science, Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA.

Geocovariate example

- Dynamic
 - Use GPS points to assign exposure
 - Google Street view image for each GPS point
 - Calculate green pixels







300 City parks

5,000 acres of parkland

5 parks >100 acres

>25 sq. miles vacant land



Approved by Detroit City Council, January 24, 2017



2017 PARKS AND RECREATION IMPROVEMENT PLAN

EXPANDING RECREATION OPPORTUNITIES IN DETROIT NEIGHBORHOODS

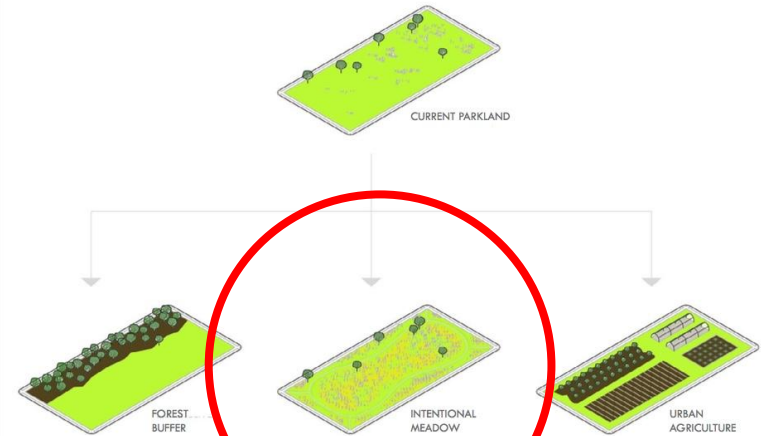


Figure 3.11 Parks that will no longer have traditional recreation spaces but will be open spaces for green infrastructure with public benefit

Nature-based intervention



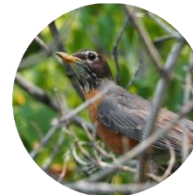
Pre-intervention park



Ecological restoration



To benefit



Intervention involved:

Removal of turf grass, litter,
and broken cement



Public Health
Summer School
Wellington



Intervention involved:

Planting native species



Intervention involved:

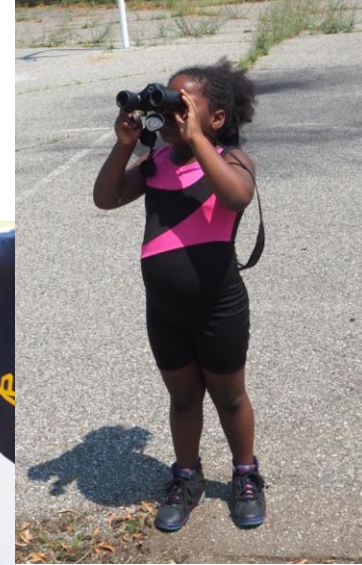
Installing trails, benches,
and signage

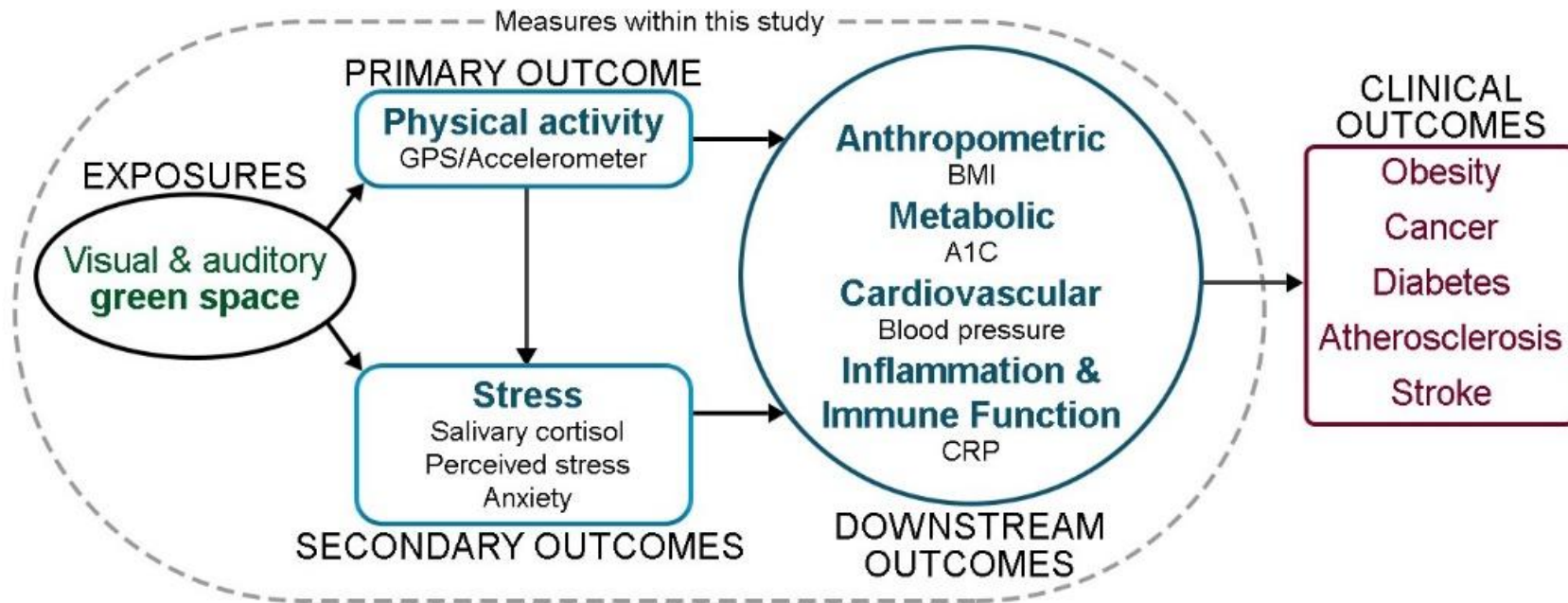


Intervention involved:

Engagement activities

Public Health
Summer School
Wellington

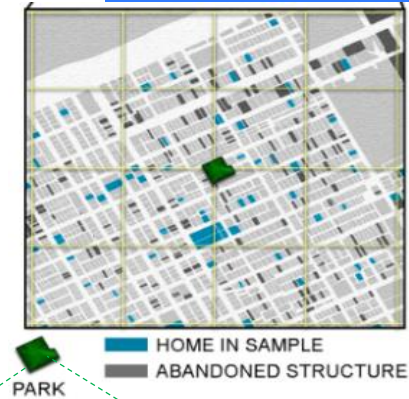




We measured how **park restoration** increases **physical activity**, lowers **stress** and improves **cardiometabolic outcomes** in Detroit.

Participant recruitment

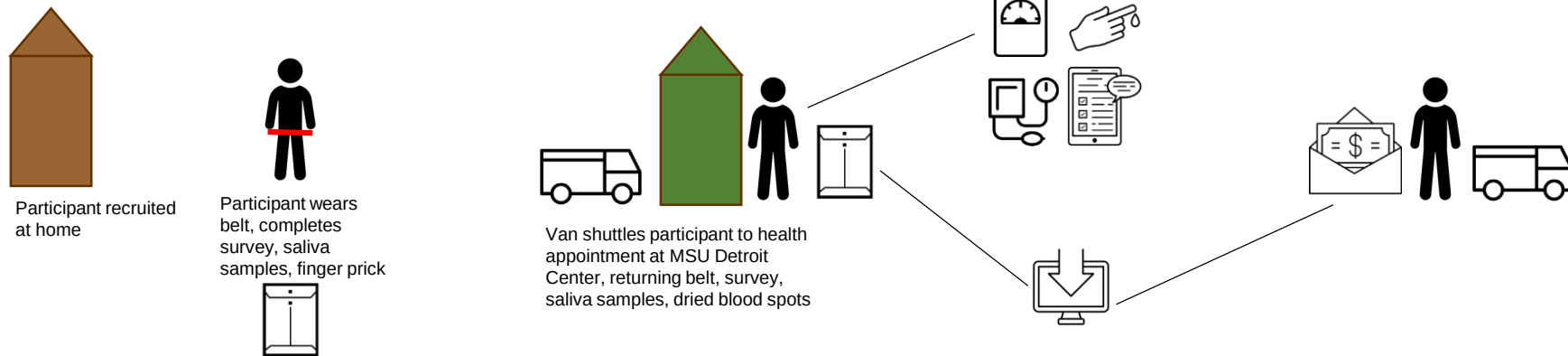
- Selected all ‘community open space’ parks – located in higher vacancy areas of Detroit
- “**Restored park neighborhoods**”, n=4
- “**Other neighborhoods**”, n=7
- Recruited participants that lived within the ‘neighborhood’ (500 meters)



District	Number of neighborhoods
3	1
4	1
5	7
6	2



Data collection



Day 1

Day 7

May to October 2019-2023
2020 and 2023 no recruitment



How we used GIS

- Static measures
- Dynamic measures
- GPS unit on people



Dynamic geocovariates (rely on GPS data)

Time spent in/outdoor by activity mode (e.g., sedentary, light, moderate, vigorous)

Time spent in park

Time spent in StAND park

Points in vacant lots

of vacant lots encountered

of abandoned houses encountered

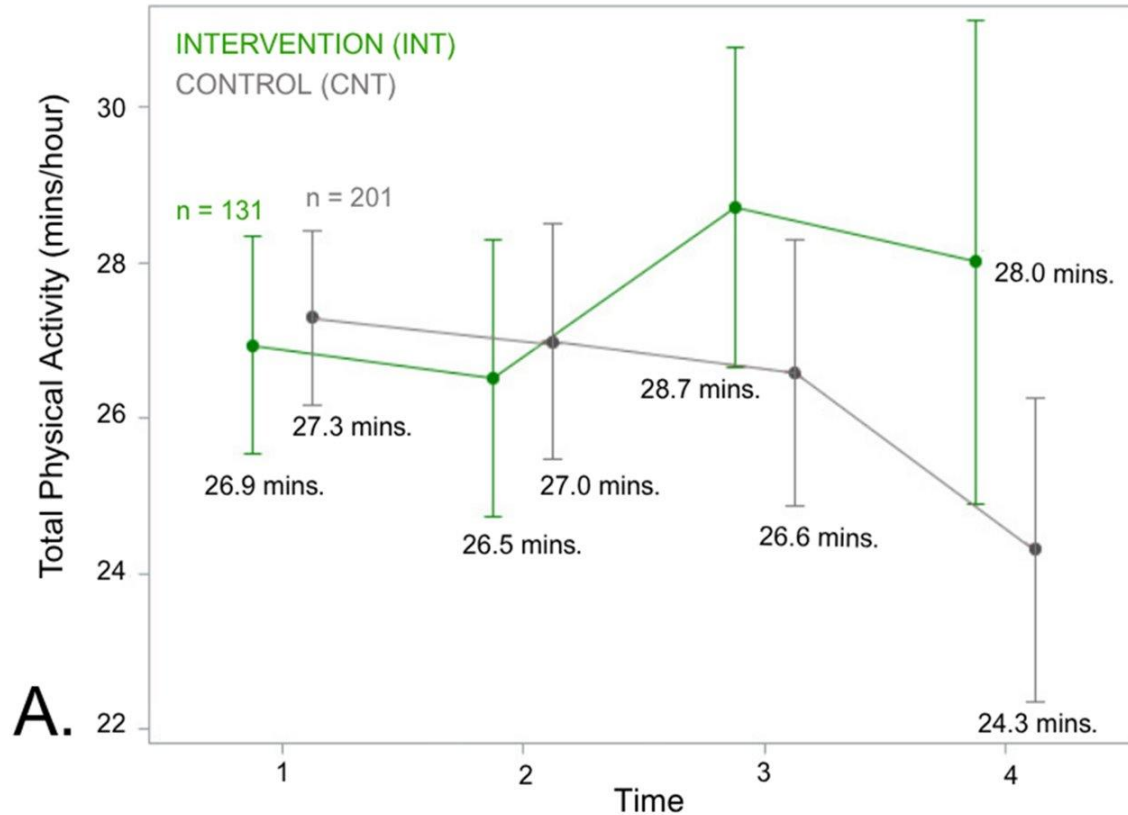
Tree coverage encountered

How we used GIS



What we found

Mean Estimates of Total Physical Activity (Minutes)



Rehabilitated Park:

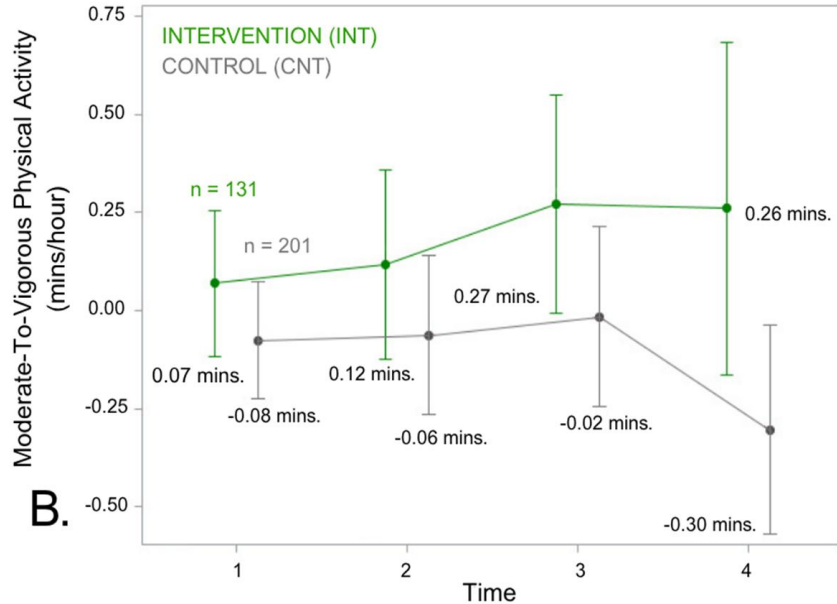


Unrehabilitated Park:

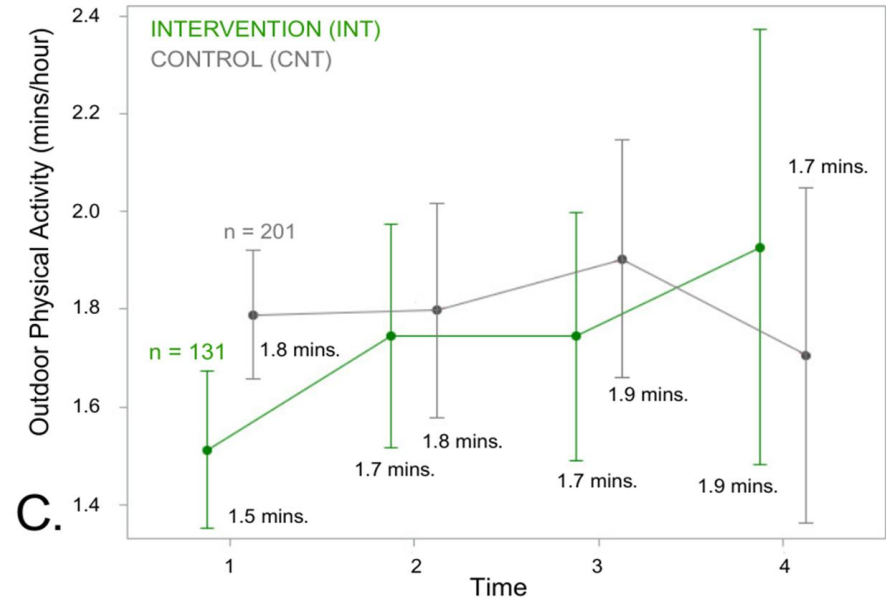


What we found

Mean Estimates of Moderate-To-Vigorous Physical Activity (Log Scale)



Mean Estimates of Outdoor Physical Activity (Log Scale)



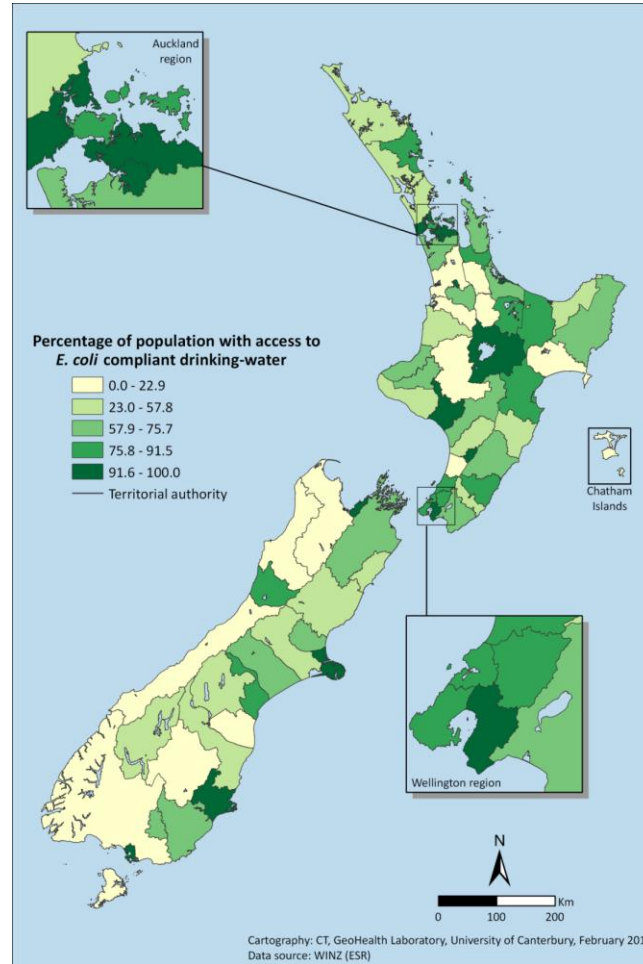
- 4.5 more minutes/day in MVPA
- Insignificant, but trending toward benefits for “Green PA”

What we found

- On average, participants in neighborhoods with restored parks spent ~**30 more minutes** per day (8 hours) in physical activity!
- This was not found in other neighborhoods.
- We will learn about opportunities for action in your neighborhood today

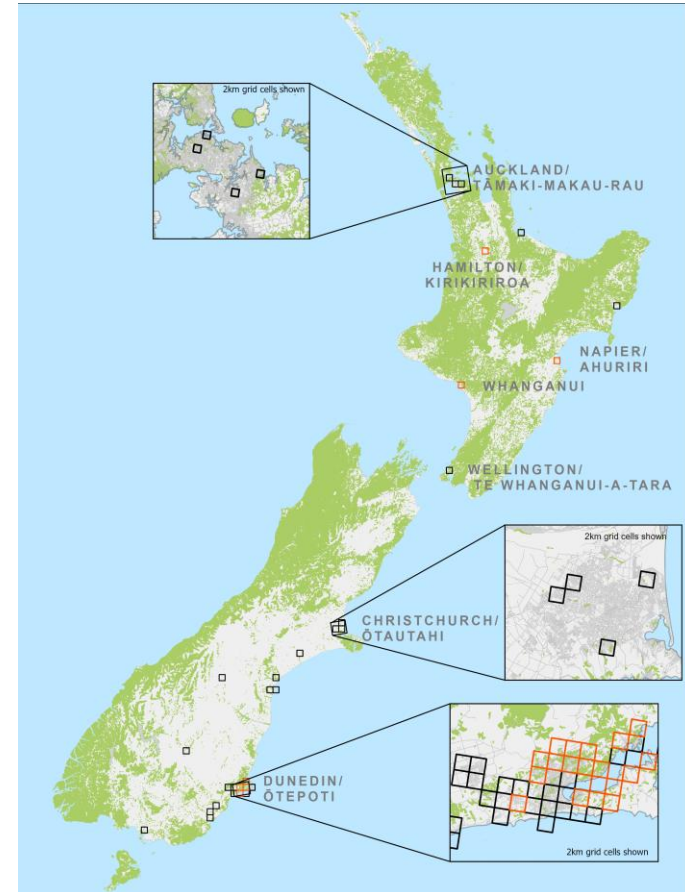
Present information and communicate findings - Today

- Percentage pop.
- Hues to show varying levels
- Geographic inequality in access to safe water



Present information and communicate findings

- Dunedin Study
- Does childhood greenery influence adult mental health?
 - Satellite derived greenery
 - Anonymized home locations via grid cells
 - Participants exposed to differing levels of greenery; those in rural/remote areas tend to be exposed to more vegetation



Cartography as communication

- Maps - a form of visual communication
 - Using resources: colour, shape and pattern to communication spatial info and relationships
 - Must pay special attention to:
 - Coordinate systems
 - Projections
 - Scale and direction
 - *Sensitive data*
- Graphical abstracts

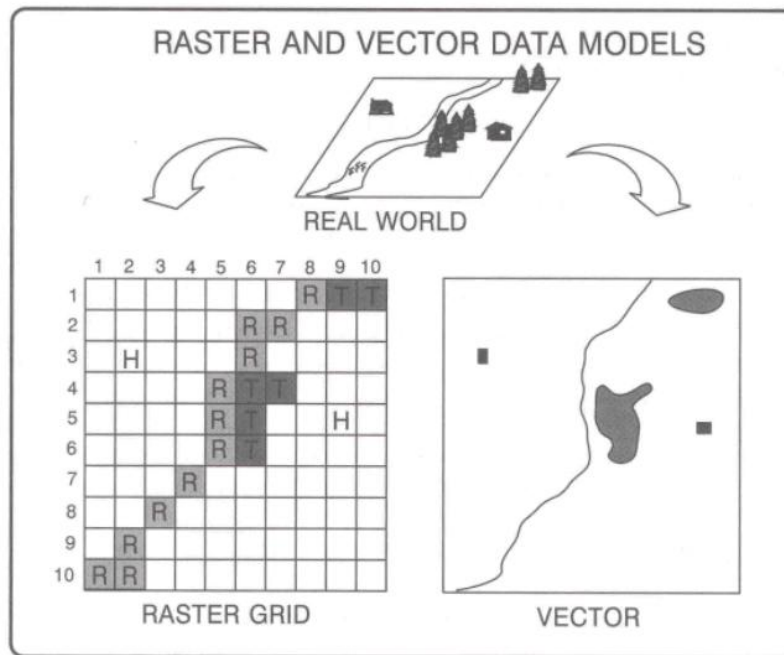
Section 2: Data types, sources and units, storage, geodatabases, map basics

Potential spatial data

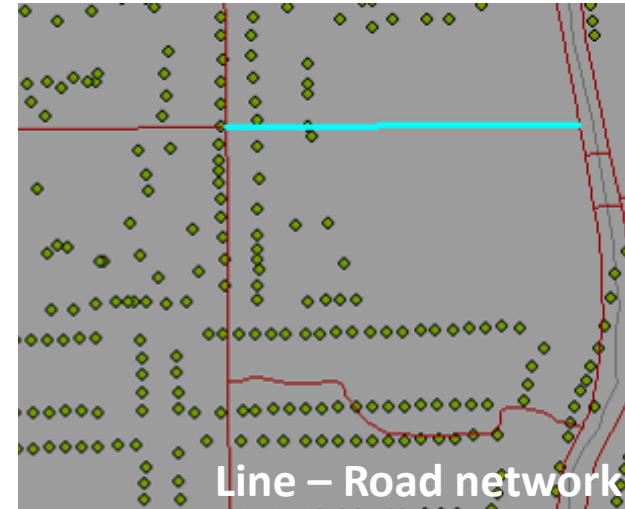
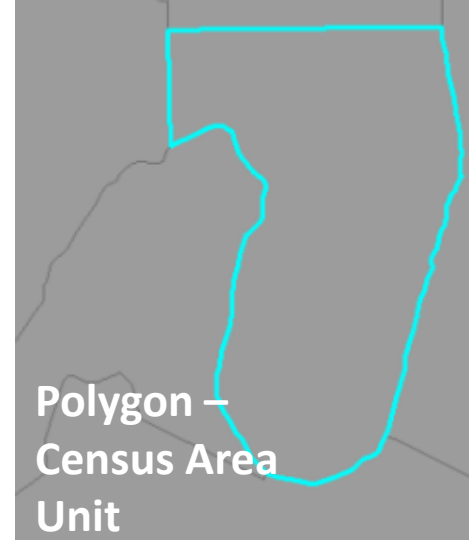
- Health - eg, cancer incidence, COVID infection rate, clinic locations
- Social - eg, volunteerism, deprivation of neighbourhoods, demographics
- Environmental - eg, UV radiation, air pollution, greenery
- Built - eg, airports, fast food outlets, sidewalks

Data types

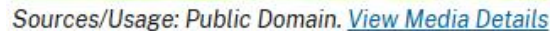
1. Vector data - point, polygon or line
2. Raster data – satellite images, pixels containing data



Data types – Vector data



Earth Data Analysis
 Source: Earth Data Analysis
 "Opportunity for the future"
 available from the National
 Earth Information Center



Source: EROS for USGS

Data sources

1. You can create spatial data
 - GPS unit/mobile phone/Field Maps
 - Google earth or Google Street View
 - Geocoding address data TODAY
2. You can obtain existing data BONUS
Shapefiles or raster files
3. You can link tabular data to spatial data BONUS

GPS devices



- Global positioning system (GPS)
 - Satellite navigation system
 - Started for military purposes
 - Primarily maintained by governments
 - Private market is growing: Starlink and Blue Origin
- GPS receiver calculates position by timing signals sent by satellites
- Uses distance between satellites – usually 4+
- Satellites continually transmit
 - the time the message was transmitted
 - satellite position at time of message transmission

GPS devices

- Common types – often worn on a belt, or wrist



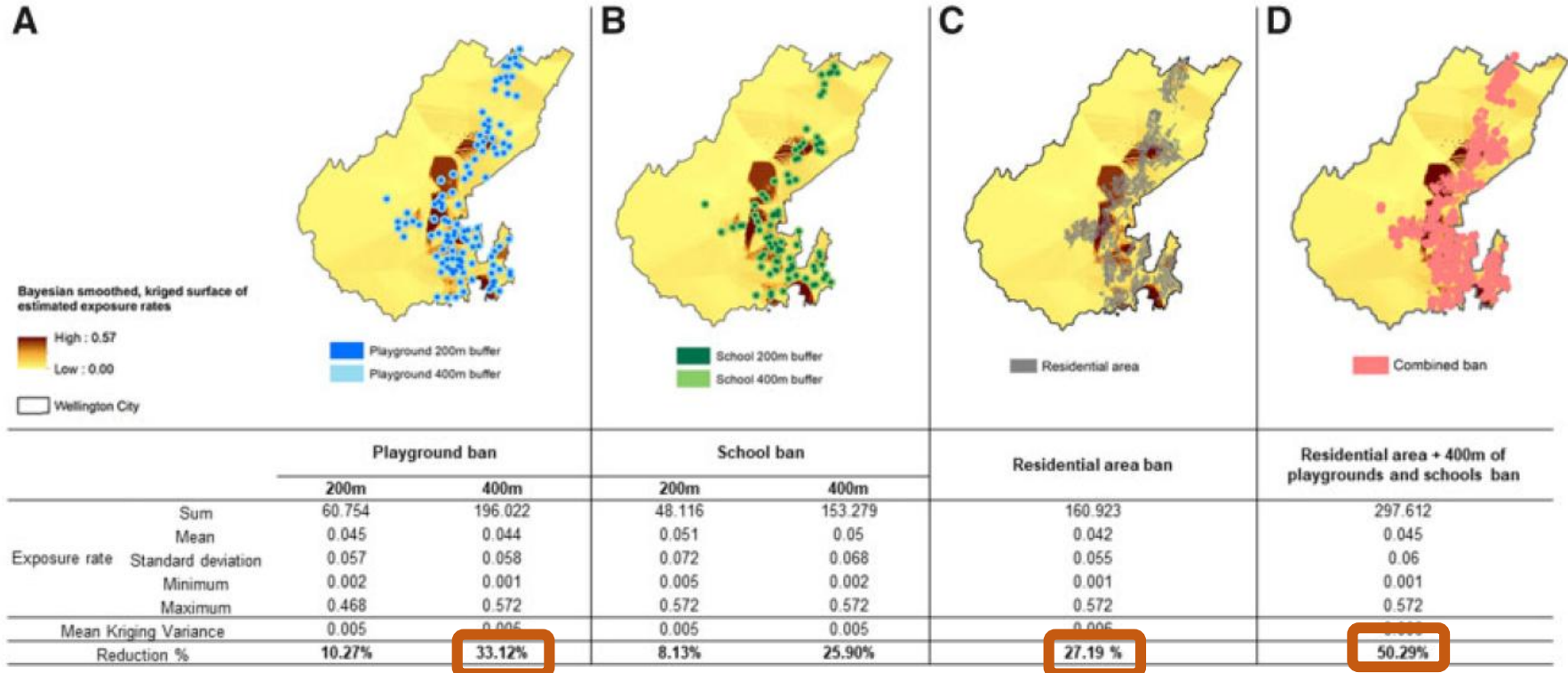
- Children may be influenced by harmful advertising
 - Gambling, alcohol, unhealthy foods
- Where and how frequently they are exposed is unknown
- If we know more about the contexts where children are exposed, we may be able to develop policies to reduce that exposure

Research example – KIDS CAM

- Automatic camera + GPS device
- Children frequently exposed through:
 - product packaging, signs, merchandise, vending machines, TV and electronic media and in-store marketing
- In school:
 - journey to and from school, in the home, and at convenience, fast food stores, and supermarkets



Research example – KIDS CAM



Google earth – GSV/Mapillary

- Can see features from above (satellites)
- Or at street-level
- And mark the locations of features or conduct audits remotely
- Can then be used in GIS to conduct analyses or generate maps

Google earth – GSV Example

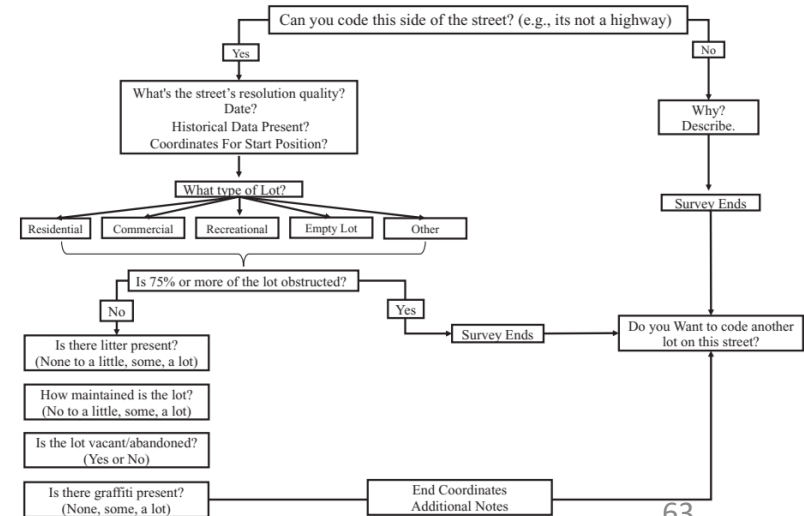
- Loads of recent uses of GSV in health research
 - Systematic review of GSV in health research (n=54):
 - 7% used GSV for sampling
 - 82% for exposure assessment
 - 11% for health policy monitoring
 - 65% utilized a checklist approach to virtually auditing the built environment for features linked to health
 - injury, alcohol and tobacco use, physical activity and mental health
 - Most studies reported considerable benefits of GSV, when compared to non-virtual methods, via the reduction of research time and costs
- Reading

GSV example- LAND

- Goal: assess level of maintenance and care of lots/parcels within neighborhood – signs of disorder = mental health
- Use GSV to conduct virtual audit
 - "Walk" street segments and code each lot

Code Structure

Code category	Question	Response
Lot type	What type of lot?	1 = residential 2 = commercial 3 = recreational 4 = empty lot 5 = other
Obstruction	Is 75% or more of the image obstructed?	1 = yes or 0 = no
Litter	Is there litter present?	1 = none to a litter (zero to five pieces) 2 = some litter (about six to 20 pieces) 3 = a lot of litter (over 20 pieces)
Maintenance	How maintained is the lot?	1 = no to little deterioration 2 = some deterioration 3 = a lot of deterioration
Vacant	Is the lot vacant or abandoned?	1 = yes or 0 = no
Graffiti	Is there graffiti present?	0 = no graffiti (zero tags) 1 = some graffiti (one to three tags) 2 = a lot of graffiti (four or more tags)





Composite	ICC	<i>M (SD)</i>	Range
Block face			
Litter	.98	8.22 (6.90)	1–41
Maintenance	.96	11.20 (10.30)	1–55
Vacant	.94	3.22 (3.07)	0–18
Graffiti	.56	0.31 (0.65)	0–4
Empty lot boundary determination	.79	NA	NA
LAND neighborhood disorder score	.97	22.95 (19.84)	2–116
Street segment			
Litter	.98	15.92 (13.42)	1–73
Maintenance	.98	21.72 (19.46)	1–101
Vacant	.96	6.32 (5.75)	0–24
Graffiti	.57	0.59 (0.92)	0–5
LAND neighborhood disorder score	.98	44.55 (37.98)	2–211

Note. NA = not applicable; ICC = intraclass correlation; LAND = Lot Assessment of Neighborhood Disorder.

GSV restrictions & alternatives

- GSV has (as of 2024) restrictions/limitations on use of their data

Additional restrictions

While we're excited to see many uses of Street View, there are some use cases that are prohibited, including but not limited to:

DON'T DO THIS

- ✗ Creating data from Street View images, such as digitizing or tracing information from the imagery
- ✗ Using applications to analyze and extract information from the Street View imagery
- ✗ Downloading Street View images to use separately from Google services (such as an offline copy)
- ✗ Merging or stitching together multiple Street View images into a larger image

These restrictions apply to all academic, nonprofit, and commercial projects.



To use or not to use proprietary street view images in (health and place) research? That is the question

Marco Helbich ^{a,1}, Matthew Danish ^{a,1}, S.M. Labib ^a, Britta Ricker ^b

Show more

+ Add to Mendeley Share Cite

Global Streetscapes—A comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics

Y. Hou, M. Quintana, M. Khomiakov, W. Yap... - ISPRS Journal of ..., 2024 - Elsevier

Street view imagery (SVI) is instrumental for sensing urban environments, benefiting numerous domains such as urban morphology, health, greenery, and accessibility. Billions ...

☆ Save Cite Cited by 86 Related articles All 5 versions

From space to street: A systematic review of the associations between visible greenery and bluespace in street view imagery and mental health

M. Bardhan, F. Li, M. H. M. Browning, J. Dong... - Environmental ..., 2024 - Elsevier

A large body of literature shows that living near greenery supports healthy lifestyles and improves mental health. Much of this research has used greenery measured from a bird's ...

☆ Save Cite Cited by 13 Related articles All 8 versions

Coverage and bias of street view imagery in mapping the urban environment

Z. Fan, C. C. Feng, F. Biljecki - Computers, Environment and Urban Systems, 2025 - Elsevier

Abstract Street View Imagery (SVI) has emerged as a valuable data form in urban studies, enabling new ways to map and sense urban environments. However, fundamental concerns ...

☆ Save Cite Cited by 26 Related articles All 6 versions

ZenSVI: An open-source software for the integrated acquisition, processing and analysis of street view imagery towards scalable urban science

K. Ito, Y. Zhu, M. Abdelrahman, X. Liang, Z. Fan... - Environment and Urban ..., 2025 - Elsevier

Street view imagery (SVI) has been instrumental in many studies in the past decade to understand and characterize street features and the built environment. Researchers across ...

☆ Save Cite Cited by 21 Related articles All 8 versions

[HTML] A citizen science toolkit to collect human perceptions of urban environments using open street view images

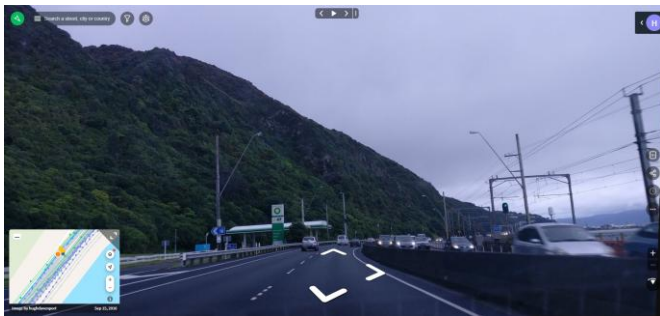
M. Danish, S. M. Labib, B. Ricker, M. Helbich - Computers, Environment and ..., 2025 - Elsevier

Abstract Street View Imagery (SVI) is a valuable data source for studies (eg, environmental assessments, green space identification or land cover classification). While commercial SVI ...

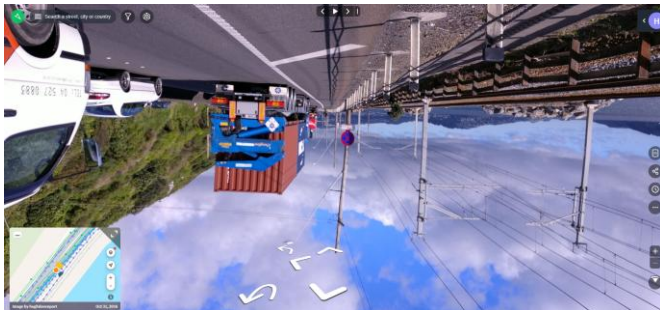
☆ Save Cite Cited by 22 Related articles All 8 versions

GSV restrictions & alternatives

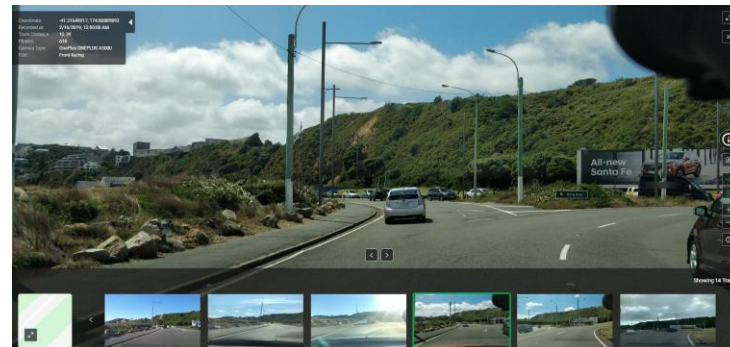
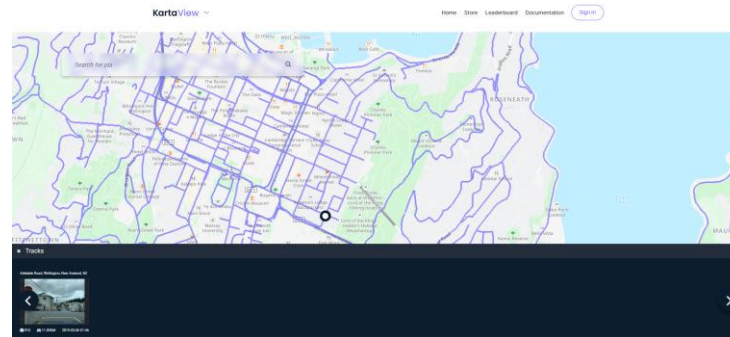
- Alternatives do exist and are predominately crowd-sourced imagery



Left: Examples
of images
available on
Mapillary

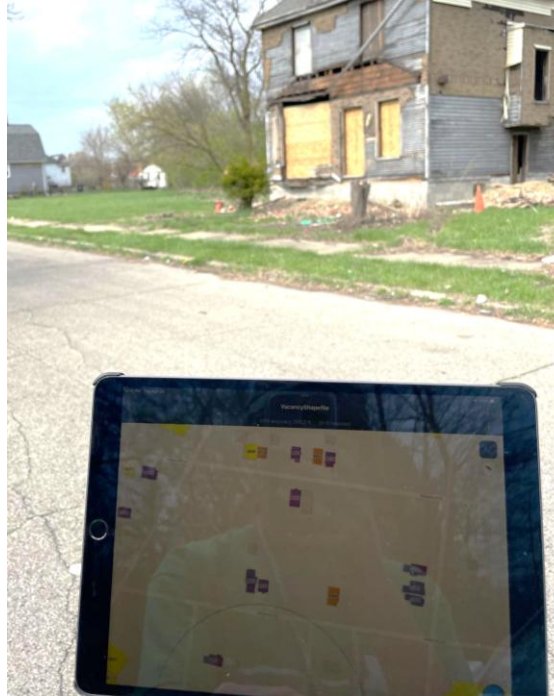
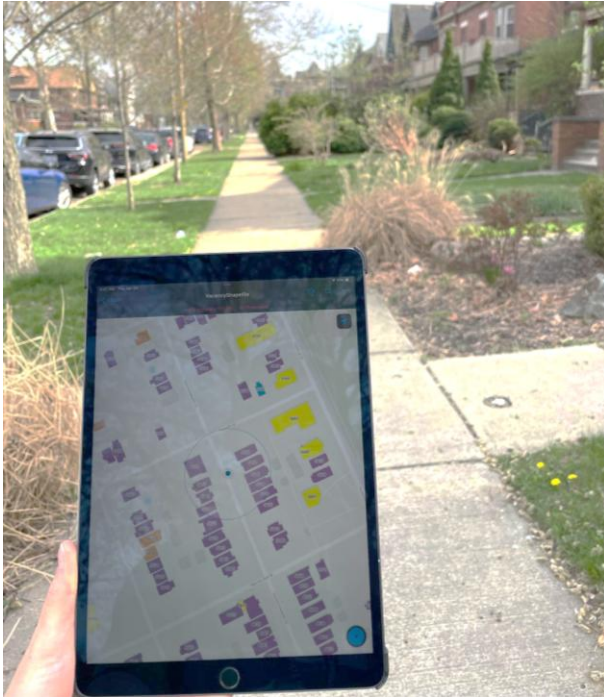


Right: Examples
of images
available on
KartaView



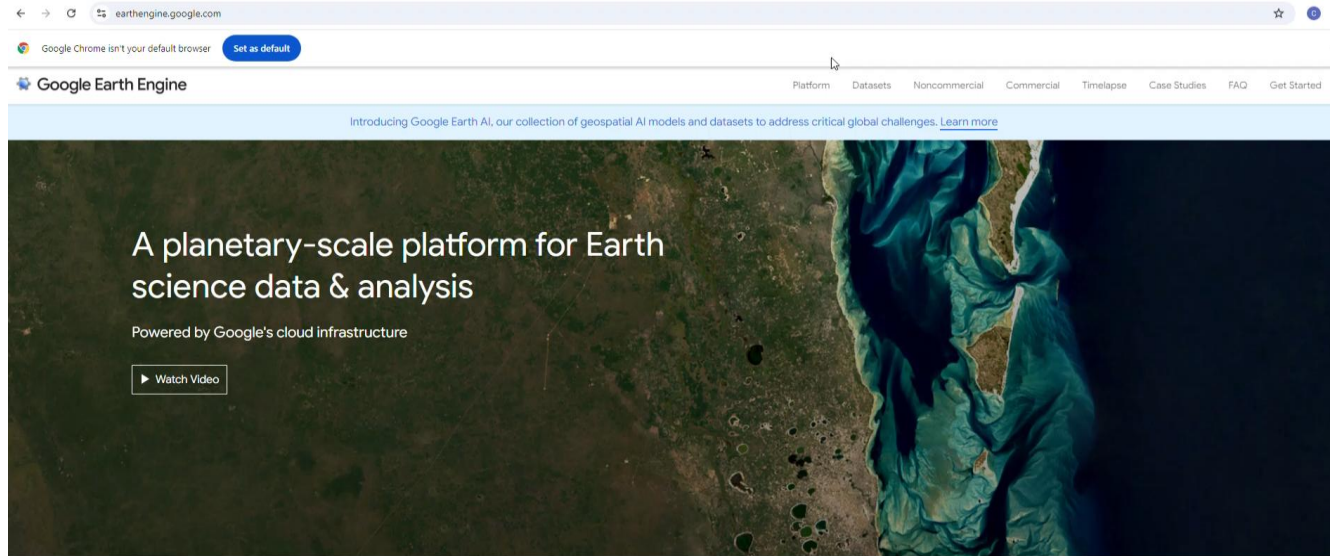
Research example – Neighborhood audits

- Audit neighborhood built conditions in FieldMaps



Example - Google Earth Engine

- Repository for large spatial datasets



Meet Earth Engine

Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. Scientists, researchers, and developers use Earth Engine to detect changes, map trends, and quantify differences on the Earth's surface. Earth Engine is now available for commercial use, and remains free for academic and research use.

Data sources and software overview

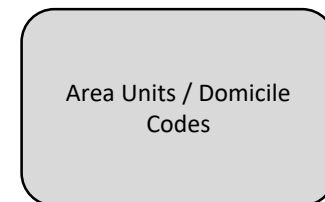
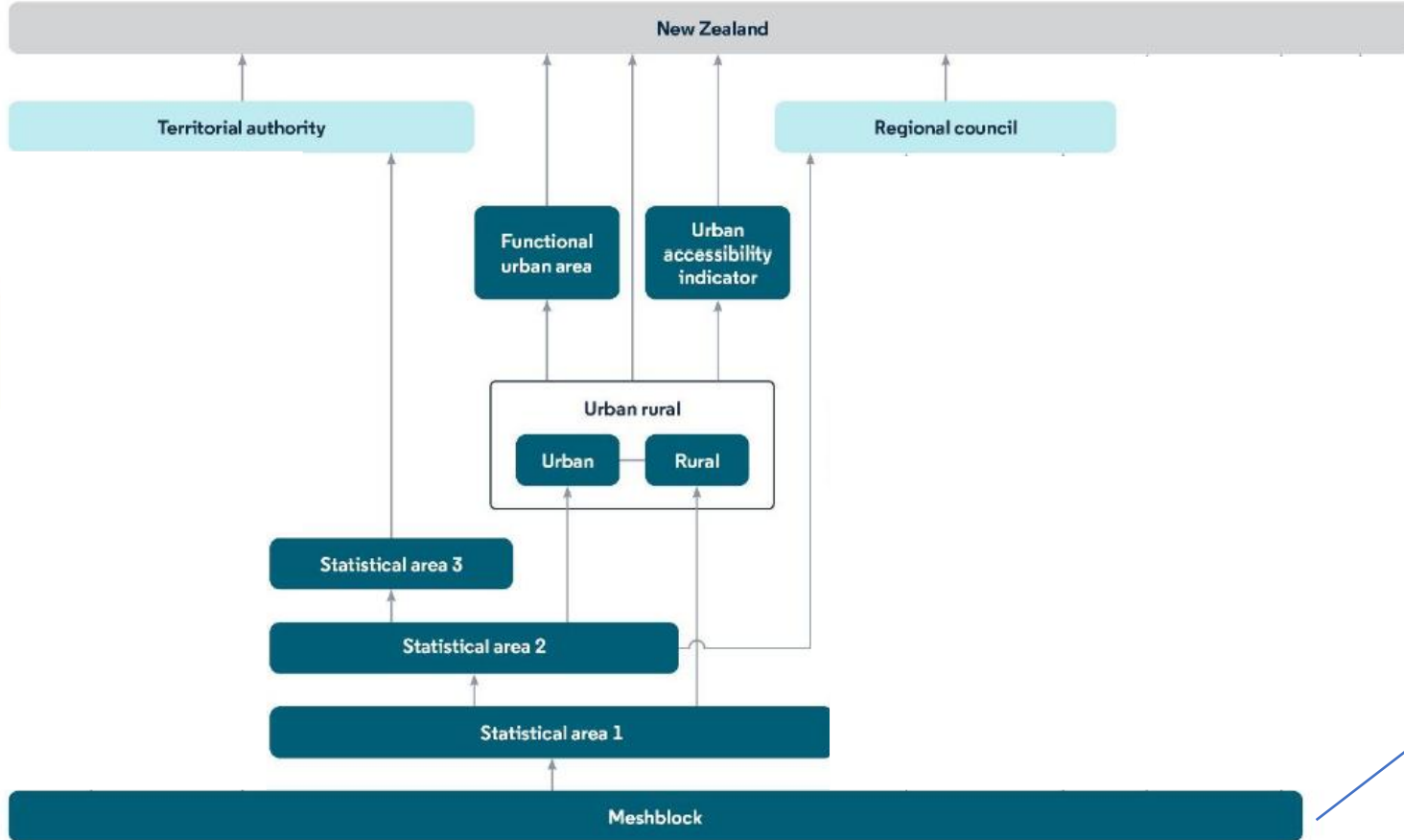
***refer to course book

- **Statistics New Zealand**
- **Interactive boundary maps (koordinates.com)**
- **Various files (in ESRI format) from other websites**
- **Software choices**
 - **ESRI ArcGIS:** industry standard for display but also many important analytical capabilities (license fee)
 - **SaTScan:** Cluster analysis software (free)
 - **R:** has packages for spatial analysis and geocoding (uses google earth)
 - **MapInfo:** (license fee)
 - **QGIS:** (free)
 - **GeoDa:** good for geographic regression, autocorrelation analyses (free)
 - **Stata 13 onwards:** programs/commands designed for mapping. [A worked example with data on my website.](#)
 - **Google:** for GSV images (currently free) or geocoding (free up to 10,000 callouts IP address/day – but have to register a credit card with Google :/)

Units and scale

- Point
 - Polygon
 - Line
-
- In New Zealand....used to be
 - Meshblock
 - Census area unit (CAU)
 - Territorial authority (TA)
 - District Health Board (DHB)
 - Updated with 2018 Census

- SSGA18 has been updated with new concepts and definitions, international compatibility
- Now SSGA23: <https://www.stats.govt.nz/methods/statistical-standard-for-geographic-areas-2023/>
 - No census 2030 (last one 2025)
- Meshblocks: 30–60 dwellings (containing approximately 60–120 residents).
 - still fundamental ‘building block’
- New statistical geographies:
 - Statistical area 1 (SA1): ideal size range of 100–200 residents
 - Statistical area 2 (SA2): 2,000–4,000 residents (city) / 1,000–3,000 residents (district)
 - Statistical area 3 (SA3): range of population sizes
 - Urban rural (UR), including
 - Functional urban areas (FUA)
 - Urban accessibility (UA)



■ Statistical geography
 ■ Electoral geography
 ■ Administrative geography
 ■ Analytical geography

■ Not all territorial authorities or regional councils contain these
 ■ Ward and community board/local board can nest within each other

Note: If a council has Māori wards/constituencies then wards and constituencies become **General ward** **General constituency**

Longitudinal work

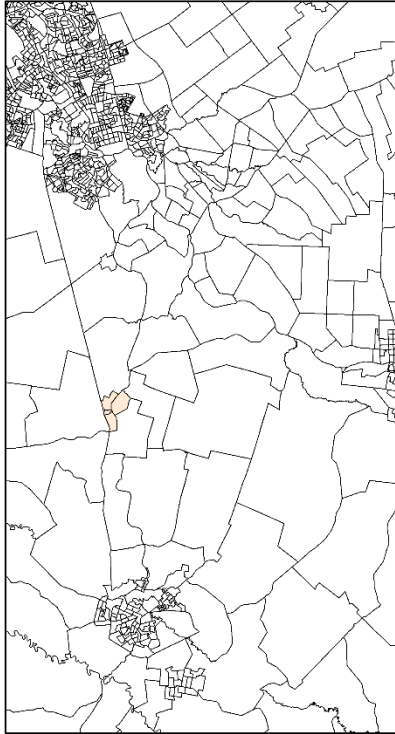
- Even before this change...
- Need to harmonize units (to larger unit)
- Or possibly use concordance files:
- http://archive.stats.govt.nz/browse_for_stats/Maps_and_geography/Geographic-areas/geographic-area-files.aspx

2018 geographic areas

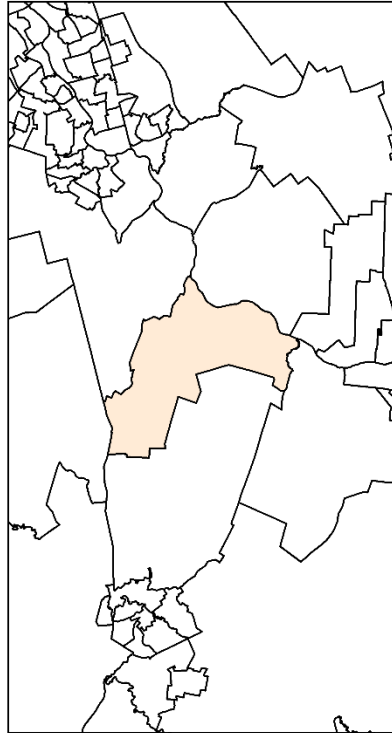
Download 2018 Annual Areas file (WinZip text file, 854kb)
(This concordance file relates to the 2018 geographic boundaries)

Contents of 2018 annual areas file	Column heading	Field size	Example
Meshblock 2001 code	MB2001_code	7	1461500
Meshblock 2006 code	MB2006_code	7	1461501
Meshblock 2010 code	MB2010_code	7	1461501
Meshblock 2011 code	MB2011_code	7	1461501
Meshblock 2013 code	MB2013_code	7	1461501
Meshblock 2017 code	MB2017_code	7	4000023
Meshblock 2018 code	MB2018_code	7	4000023
Statistical area 1 2018 code	SA12018_code	7	7016339
Statistical area 2 2018 code	SA22018_code	6	213200
Statistical area 2 2018 descriptor	SA22018_name	254	Taradale South
Urban rural 2018 code	UR2018_code	4	1296
Urban rural 2018 descriptor	UR2018_name	254	Napier
Urban rural indicator 2018 code	IUR2018_code	2	12
Urban rural indicator 2018 descriptor	IUR2018_name	254	Large urban area
Territorial authority 2018 code	TA2018_code	3	031
Territorial authority 2018 descriptor	TA2018_label	254	Napier city
Ward 2018 code	WARD2018_code	5	03104
Ward 2018 descriptor	WARD2018_label	254	Taradale ward
Community board 2018 code	CB2018_code	5	03199
Community board 2018 descriptor	CB2018_label	254	Area outside community
Subdivision 2018 code	TASUB2018_code	5	03199
Subdivision 2018 descriptor	TASUB2018_label	254	Area outside subdivision
Regional council 2018 code	REGC2018_code	2	06

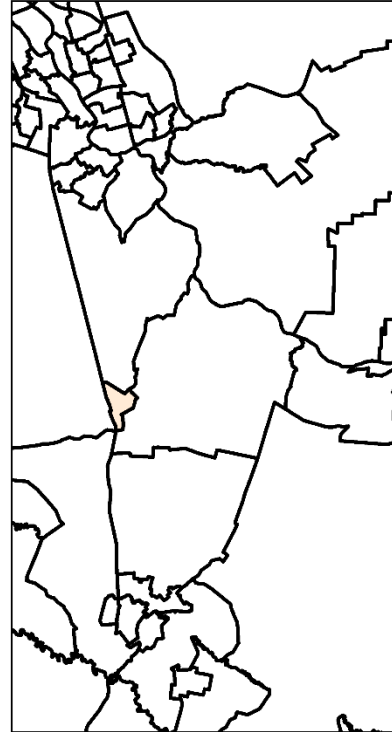
Statistical Area 1



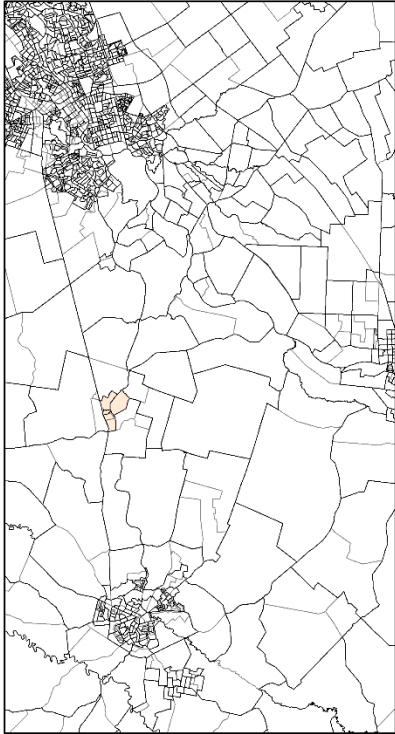
Statistical Area 2



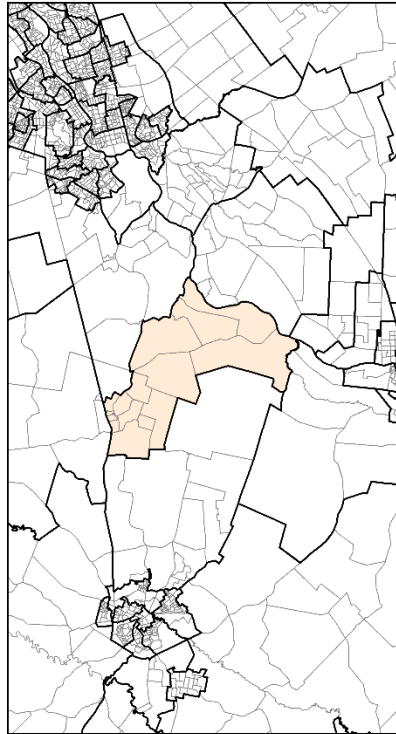
Domicile



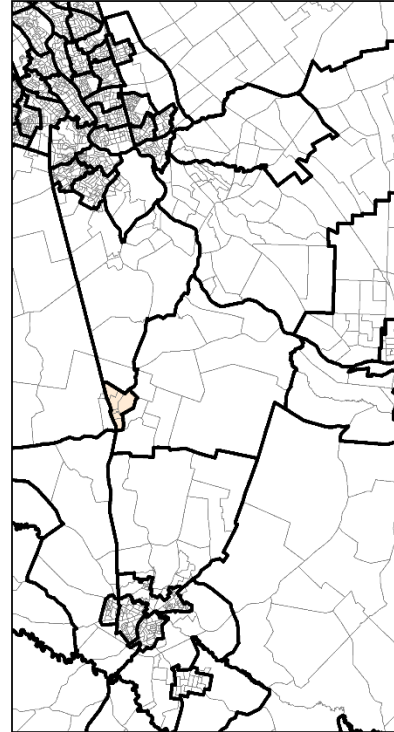
Statistical Area 1



Statistical Area 2

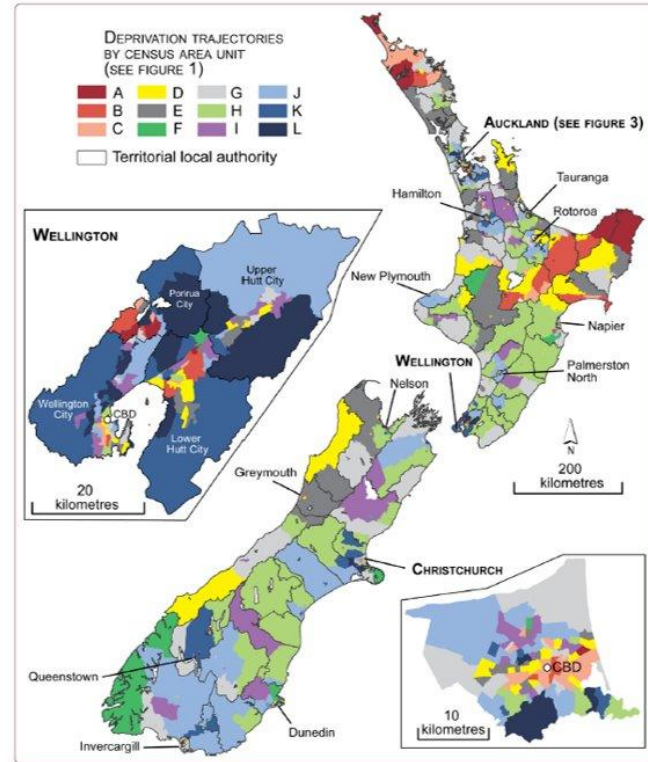


Domicile



Example boundary harmonization over time

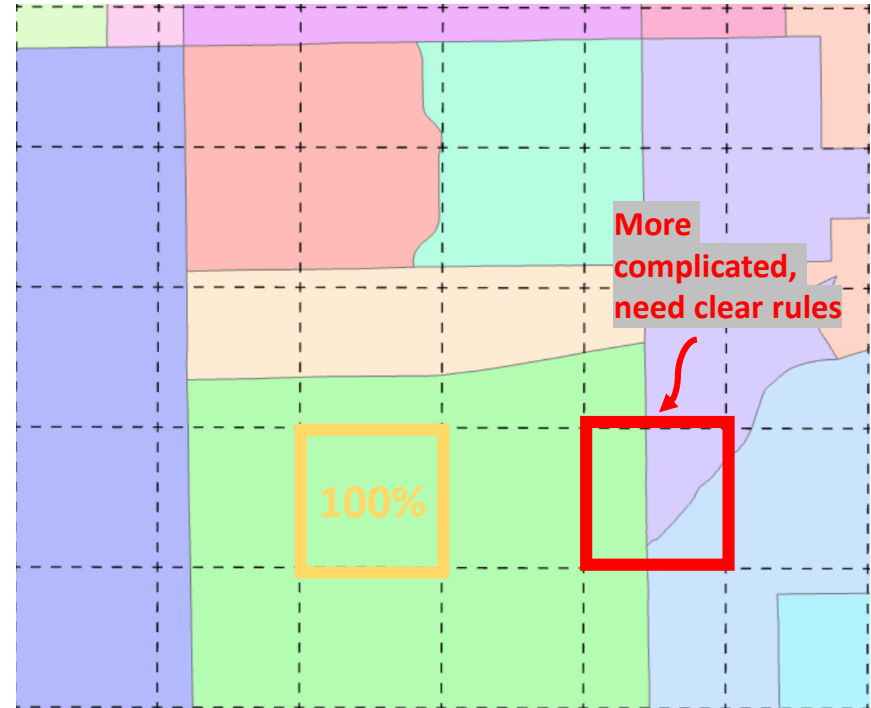
- Geographic boundaries of CAUs changed 1991-2006
- N CAUs 1637 (1991) to 1784 (2006)
- Some CAUs were split into two (due to population increases)
- To obtain a repeated cross-sectional dataset with the same spatial units throughout the period, we
 - harmonised geographic boundaries by aggregating contiguous CAUs which were split at some point after 1991



Cumulative disadvantage? Exploring relationships between neighbourhood deprivation trends (1991 to 2006) and mortality in New Zealand

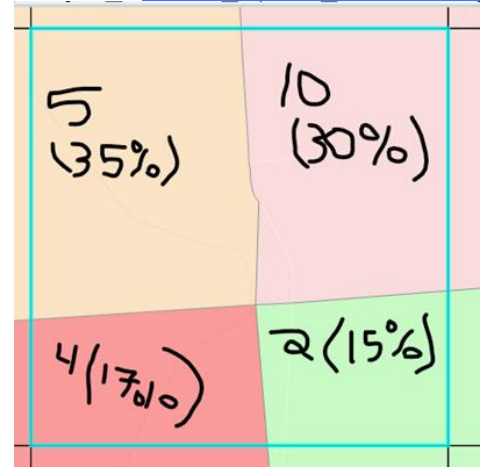
Boundary harmonization workaround

- To avoid challenges with changing boundaries, can choose to use new geography – grid cells
- Select a cell size (e.g., 1km), overlay data, then set rules for how to transfer values from original geography to grid



Boundary harmonization workaround

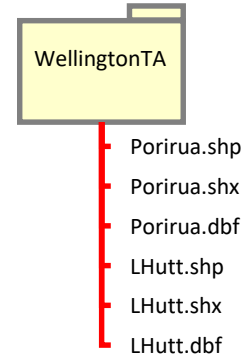
- Example: assigning RUCA codes to grid cells
- Establish rules:
 - If >50% of cell is one RUCA code— assign that value
 - If none >50%, and the cell HAS a code 1,4, or 7— assign that value
 - If none >50% and no code 1,4, or 7, assign value of RUCA code with maximum land area



Storage of geographic data

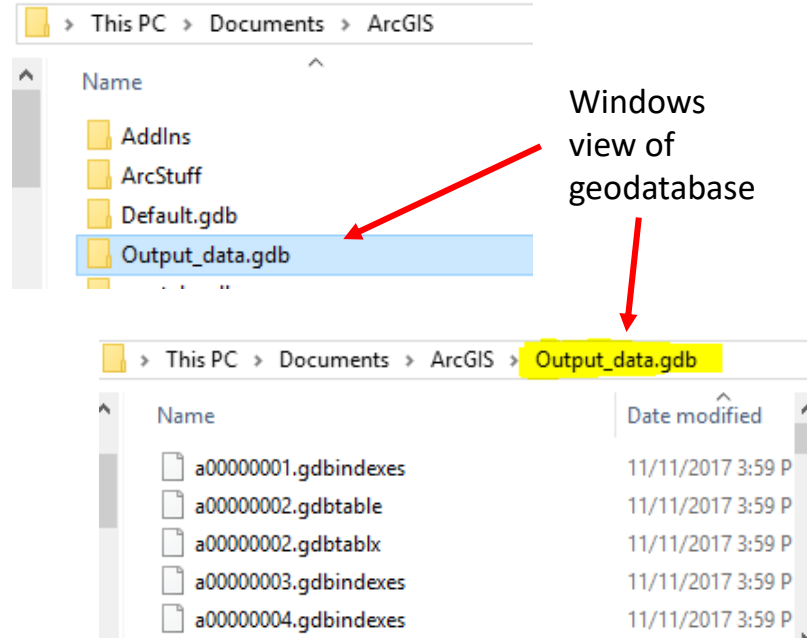
- Shapefiles are stored in three to five files (with extensions .shp, .shx, .dbf, .sbx and .sbn)
- Shapefiles allow features to have multiple, disconnected, intersecting and overlapping components
- **Vector data** is stored as individual shapefiles or feature classes within a geodatabase
- **Raster data** is stored as .tiffs, .jpgs, or other image formats, and can be stored in a geodatabase

Shapefiles



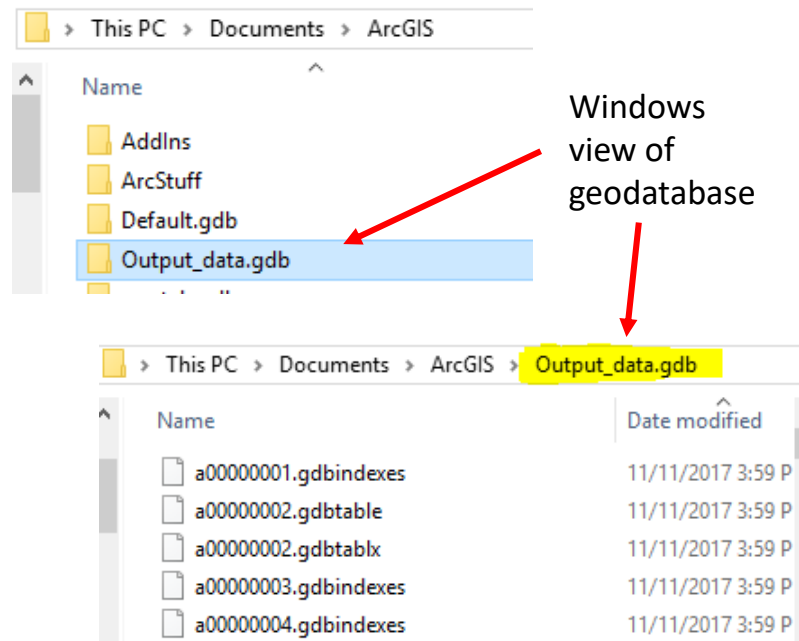
Geodatabase (.gdb)

- Three types: Personal, File and Enterprise
- Container for multiple data types
 - Raster
 - Vector
 - Tables
 - Tools and more



Geodatabase (.gdb)

- Replaces shapefiles
- GDB as primary storage
- Scaleable- single user to multi-user/editor
- Allows for data validation (tools for validation analyses in geodatabase)



ArcPro Geodatabase example

The screenshot displays the ArcGIS Pro interface. The top ribbon includes tabs for Project, Catalog, Insert, Analysis, View, Imagery, Share, and Manage. The Catalog pane on the left shows a tree view of the project structure, with 'Auck_CAUs' selected under 'Data_forTutorial'. The main view shows a table of data with columns: FID, Shape, AU06, UA06, TA06, AU06_NAME, TA06_NAME, UA06_NAME, and DHBID06. The table contains 25 rows of data, each representing a different Auckland Council ward.

ArcCatalog:
Organise, manage,
examine, edit, find
geographic datasets

View data
(like Windows Explorer)

	FID	Shape *	AU06	UA06	TA06	AU06_NAME	TA06_NAME	UA06_NAME	DHBID06
1	0	Polygon	514000	4	7	Freemans Bay	Auckland City	Central Auckland Zone	3
2	1	Polygon	514101	4	7	Auckland Harbourside	Auckland City	Central Auckland Zone	3
3	2	Polygon	514102	4	7	Auckland Central West	Auckland City	Central Auckland Zone	3
4	3	Polygon	514103	4	7	Auckland Central East	Auckland City	Central Auckland Zone	3
5	4	Polygon	514301	4	7	Grafton West	Auckland City	Central Auckland Zone	3
6	5	Polygon	514302	4	7	Grafton East	Auckland City	Central Auckland Zone	3
7	6	Polygon	514401	4	7	Roberton	Auckland City	Central Auckland Zone	3
8	7	Polygon	514402	4	7	Glenavon	Auckland City	Central Auckland Zone	3
9	8	Polygon	514500	4	7	New Windsor	Auckland City	Central Auckland Zone	3
10	9	Polygon	514600	4	7	Avondale South	Auckland City	Central Auckland Zone	3
11	10	Polygon	514700	4	7	Blockhouse Bay	Auckland City	Central Auckland Zone	3
12	11	Polygon	514801	4	7	Rosebank	Auckland City	Central Auckland Zone	3
13	12	Polygon	514802	4	7	Avondale West	Auckland City	Central Auckland Zone	3
14	13	Polygon	514900	4	7	Waterview	Auckland City	Central Auckland Zone	3
15	14	Polygon	515001	4	7	Point Chevalier West	Auckland City	Central Auckland Zone	3
16	15	Polygon	515002	4	7	Point Chevalier East	Auckland City	Central Auckland Zone	3
17	16	Polygon	515003	4	7	Point Chevalier South	Auckland City	Central Auckland Zone	3
18	17	Polygon	515100	4	7	Westmere	Auckland City	Central Auckland Zone	3
19	18	Polygon	515201	4	7	Herne Bay	Auckland City	Central Auckland Zone	3
20	19	Polygon	515202	4	7	St Marys	Auckland City	Central Auckland Zone	3
21	20	Polygon	515301	4	7	Ponsonby West	Auckland City	Central Auckland Zone	3
22	21	Polygon	515302	4	7	Ponsonby East	Auckland City	Central Auckland Zone	3
23	22	Polygon	515410	4	7	Grey Lynn West	Auckland City	Central Auckland Zone	3
24	23	Polygon	515420	4	7	Grey Lynn East	Auckland City	Central Auckland Zone	3
25	24	Polygon	515431	4	7	Surrey Crescent	Auckland City	Central Auckland Zone	3

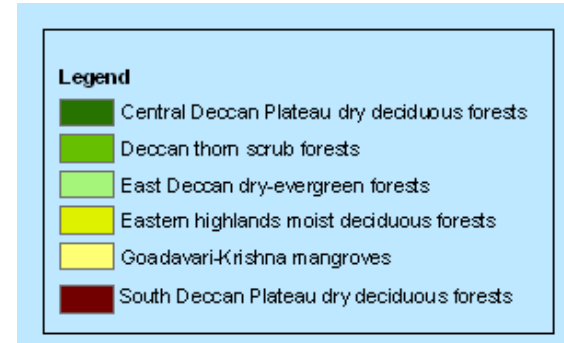
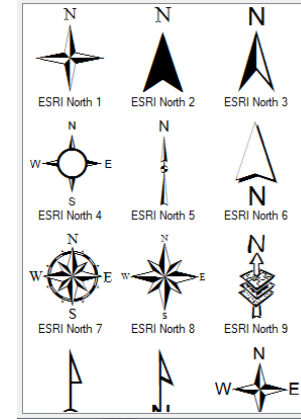
- What is the goal of the map?
- Who will read the map?
- Where will the map be used?
- What data are available for the map composition?
- What resources are available?

Basic elements of map composition

- Generalisation & simplification
 - How much to simplify situation depicted?
 - How to symbolise the relationships represented?
 - Variable representation – hue, shape, size, orientation, spacing
 - May depend on data - nominal, ordinal, Interval, ratio
- Defensible reason for each element
- Map layouts, printing, format of final product

Basic elements of most maps

- Distance or scale
- Direction
- Legend
- Sources of info/processing method



Elements which depend on purpose

- Title
- Labels
- Projection
- Cartographer
- Date of production
- Locator or inset maps

Coordinate systems and projections

- We will not cover all aspects.
- Geographic coordinates help us assign a location to a point (e.g., latitude, longitude)
- This is like seeing the earth as a 2-D flat piece of paper
- If we want to account for the earth's curvature - projected coordinate system.
- However, all projections distort map properties (which is why there are so many; to reduce bias in some places, bias in other areas are increased).

Checking a coordinate system

- Checking/setting coordinate system
- You can always right click a dataset title in the Contents window
- **Properties**
 - *Source* → *Spatial Reference*

Projected Coordinate System		NAD 1983 (2011) Michigan Georef (Meters)
Projection	Hotine Oblique Mercator Azimuth Natural Origin	
WKID	6497	
Previous WKID	102968	
Authority	EPSG	
Linear Unit	Meters (1.0)	
False Easting	2546731.496	
False Northing	-4354009.816	
Scale Factor	0.9996	
Azimuth	337.25556	
Longitude Of Center	-86.0	
Latitude Of Center	45.30916666666666	
Geographic Coordinate System		NAD 1983 (2011)
WKID	6318	
Previous WKID	104145	
Authority	EPSG	
Angular Unit	Degree (0.0174532925199433)	
Prime Meridian	Greenwich (0.0)	
Datum	D NAD 1983 2011	

- *You will be re-projecting in Tutorial 1 – detailed instructions*

Base maps

- You have data of interest to your research question or public health issue (fast food outlet locations or brain cancer incidence)
- But the audience needs background spatial reference points or features
- Base maps (in ArcPro tutorial coming up)
- Morning tea!

Add Basemap



Imagery



Imagery with Labels



Streets



Topographic



Dark Gray Canvas



Light Gray Canvas



National Geographic



Terrain with Labels



Oceans

Section 3: Overview of ArcGIS Pro

Overview of ArcGIS Pro

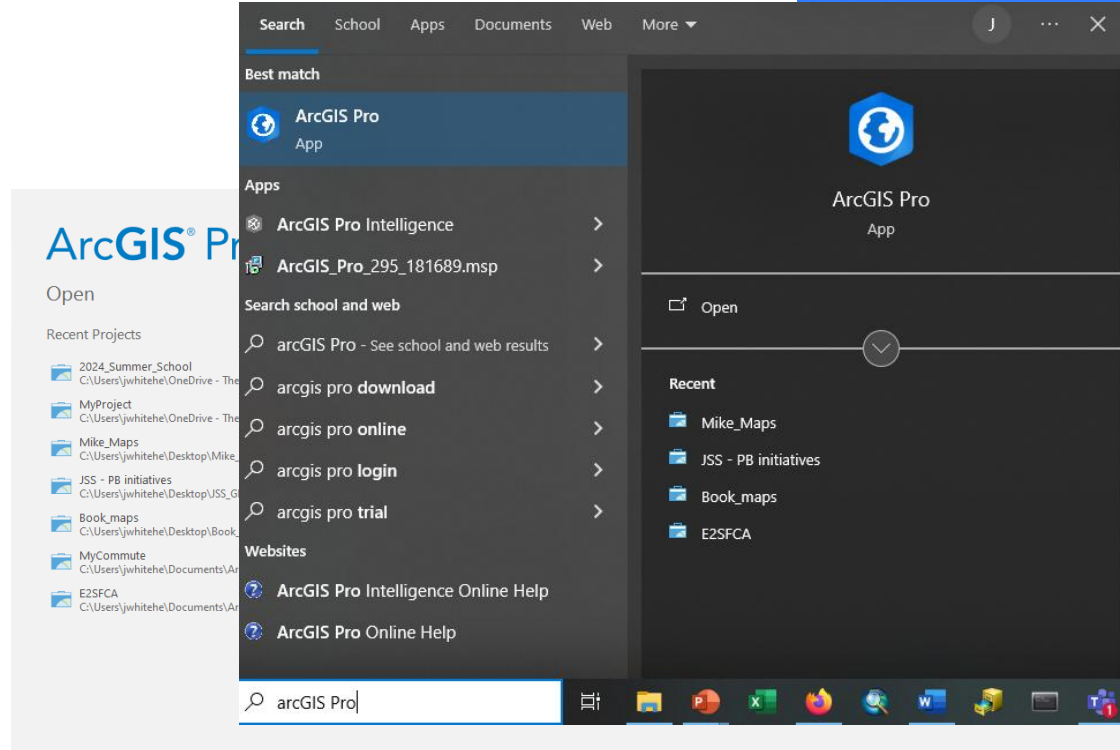
- Opening program
- Key features
- Adding data
- Opening attribute tables
- Changing symbology
- Exporting map

Getting set up

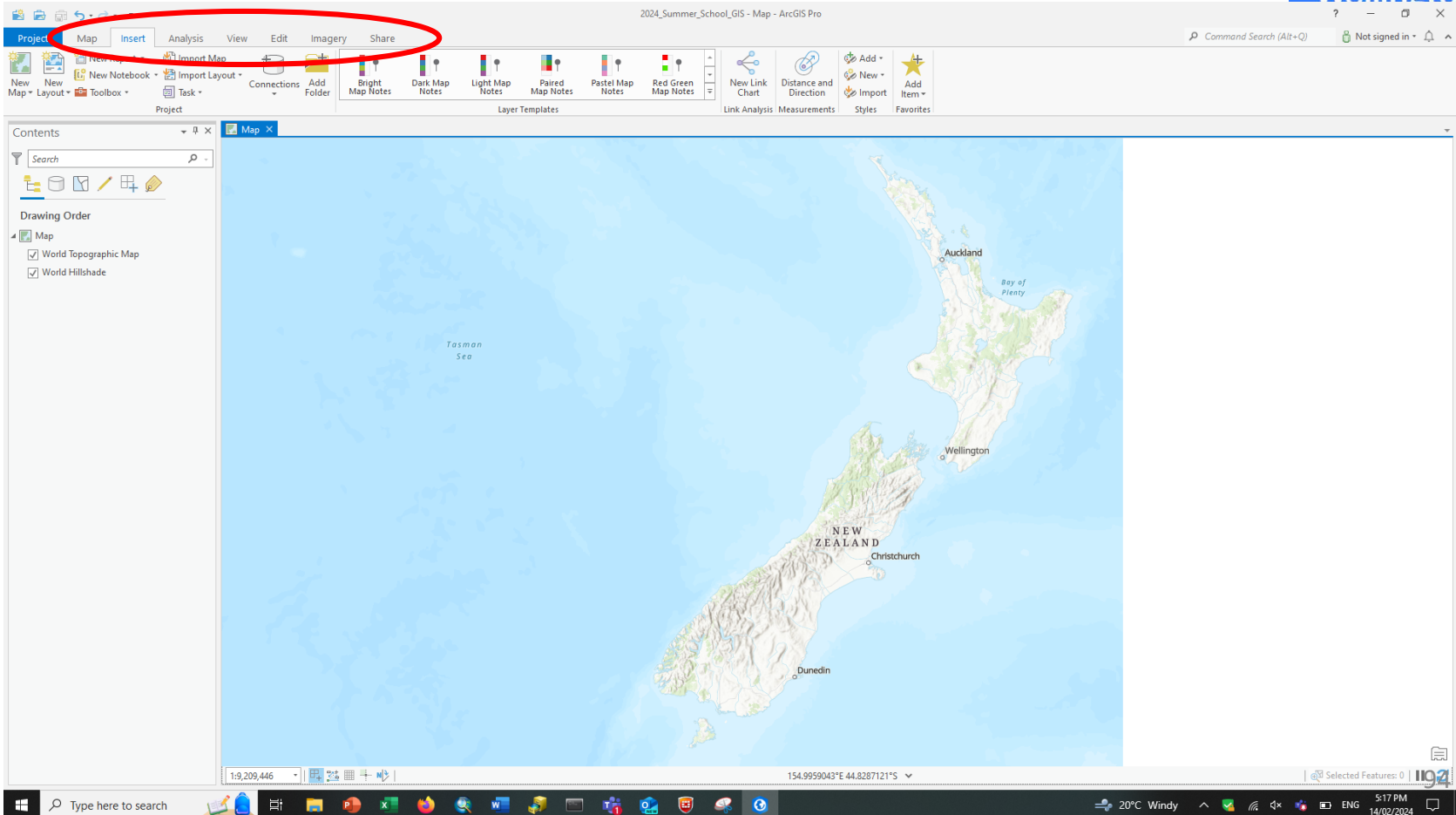
- Directories – data location
- Data access after the course – my website

Opening ArcGIS

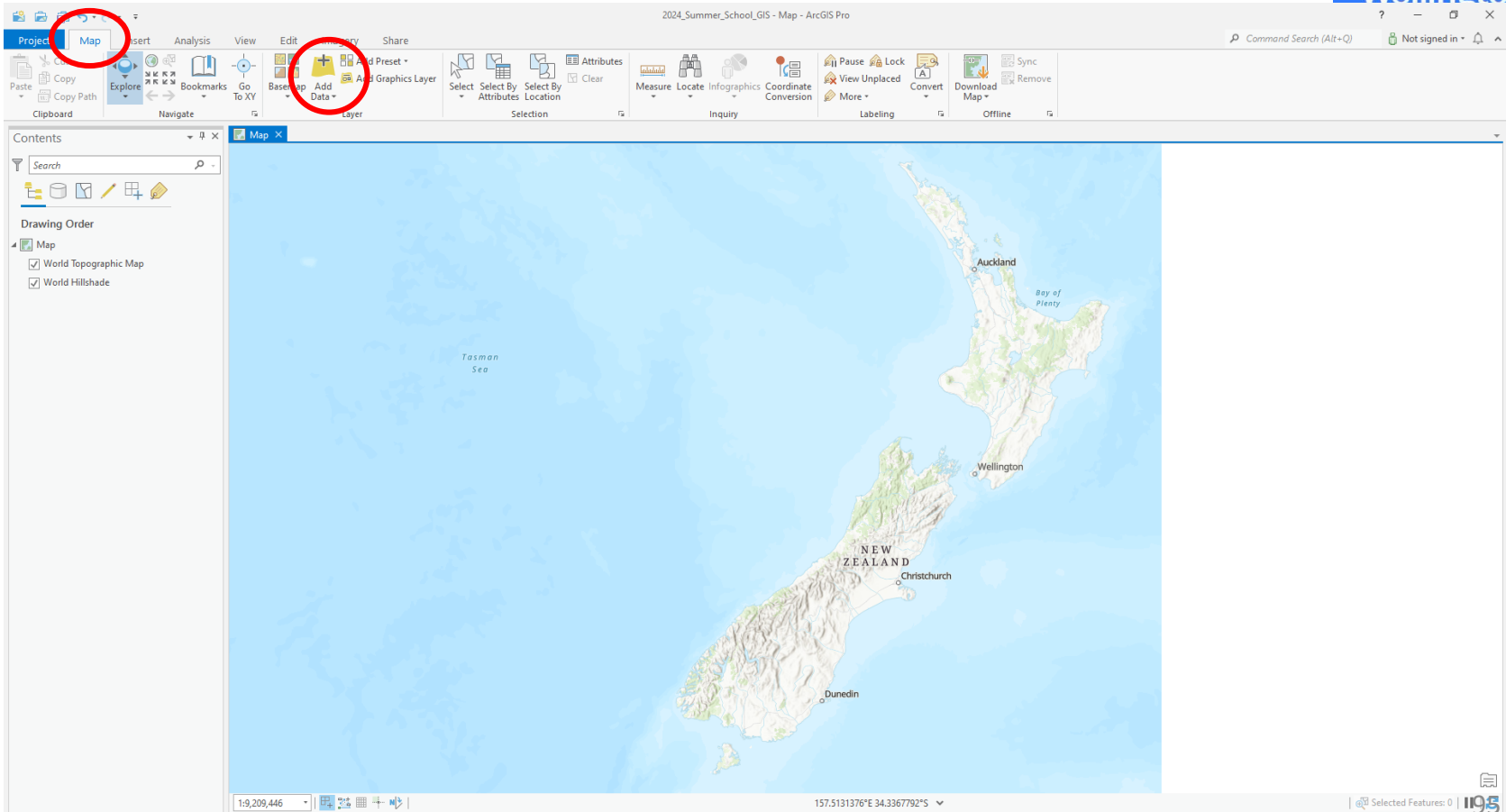
- Go to Programs
 - Select ArcGIS Pro
- Click “Map”
 - Choose name and place to save it



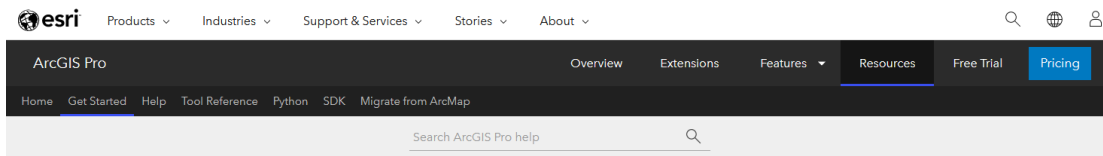
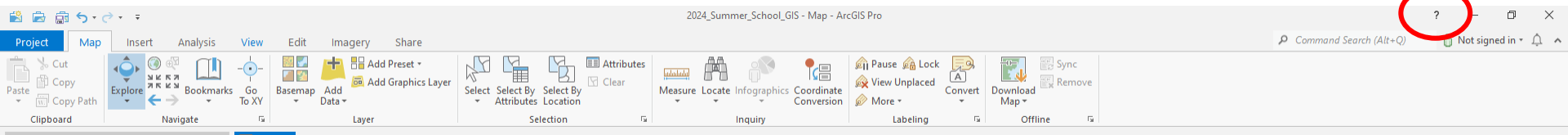
User Interface



User Interface



Help files



- > Set up
- > Licensing
- > Quick-start tutorials
- > Extensions
- > General
- > What's new

About ArcGIS Pro

This ArcGIS 2.9 documentation has been [archived](#) and is no longer updated. Content and links may be outdated. [See the latest documentation.](#)

ArcGIS Pro 2.9: All patches and release dates are listed on [the Technical Support site](#). For the best experience, update to the latest patch.

ArcGIS Pro is the latest professional desktop GIS application from Esri. With ArcGIS Pro, you can explore, visualize, and analyze data; create 2D maps and 3D scenes; and share your work to your ArcGIS Online or ArcGIS Enterprise portal. The sections below introduce basic elements of the user interface. You will build on this knowledge as you use the application, read the help, and complete the [tutorials](#).

Sign in

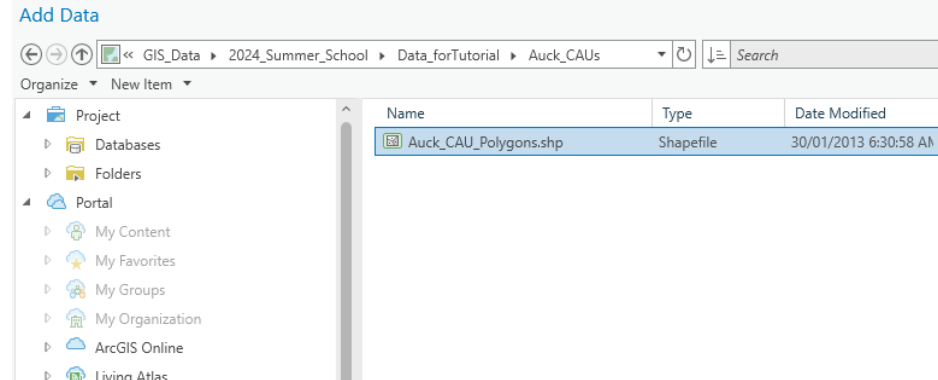
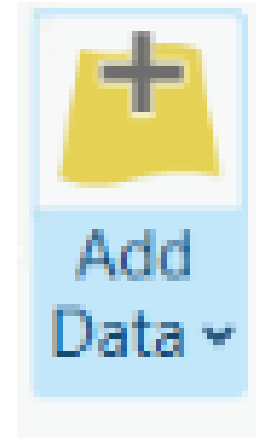
To start using ArcGIS Pro, you typically [sign in](#) with the credentials of your ArcGIS Online or ArcGIS Enterprise organization. Signing in is required if your [software license](#) is managed by your ArcGIS organization, which it is by default. Signing in also enables you to access web content, such as web maps, stored in your organization and other online repositories.

In this topic

- [Sign in](#)
- [Projects](#)
- [ArcGIS Pro user interface](#)
- [Learning resources](#)

Adding data - .shp files only

- Shapefiles are actually sets of 5-7 files with different suffixes
- Click on the Map tab, then click “Add Data
- Navigate to working folder
- Select Auck_CAU_polygons.shp
- Data are added using default symbols, which you can modify

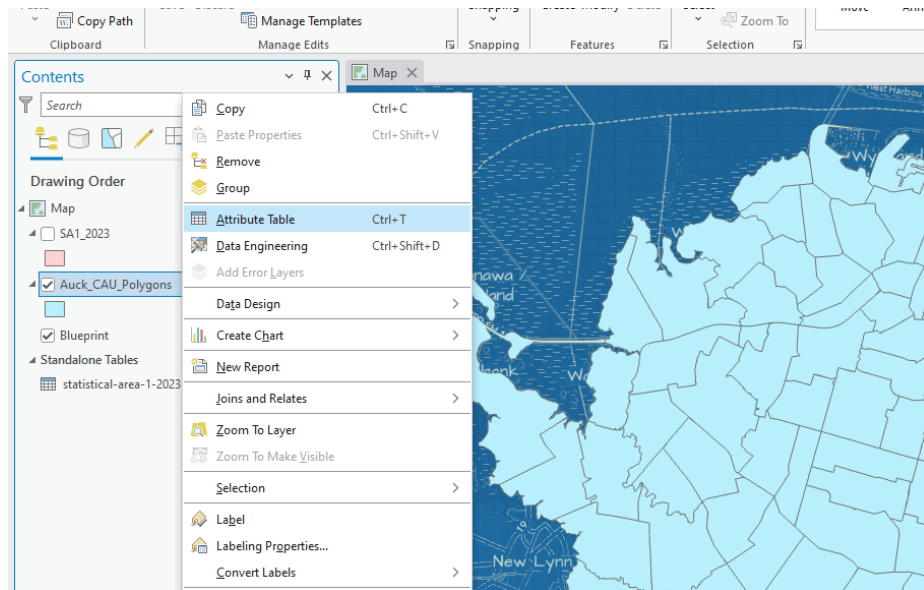


Adding data

- Connecting to the right folder can sometimes be tricky....
- Folder on desktop or flashdrive
- Watch as I do it now...

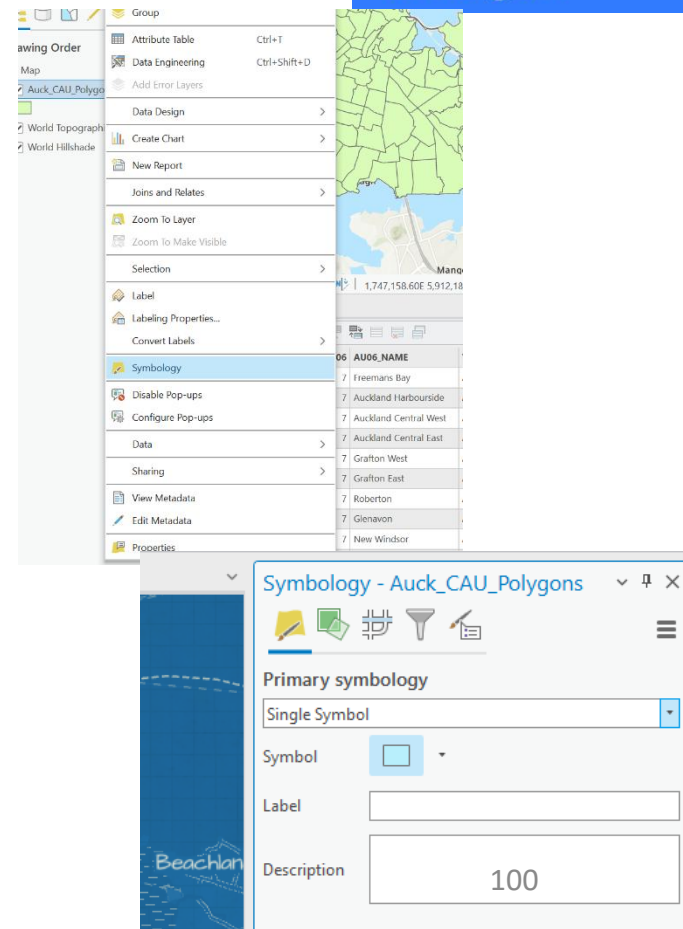
Looking at the attribute table

- Right click .shp name
- Click Open Attribute Table
 - Now you can see all the data for each geographic unit (in this case polygons)
 - You can highlight a row and the map will highlight the relevant CAU
 - Here you can also add a column or make new variables and calculations



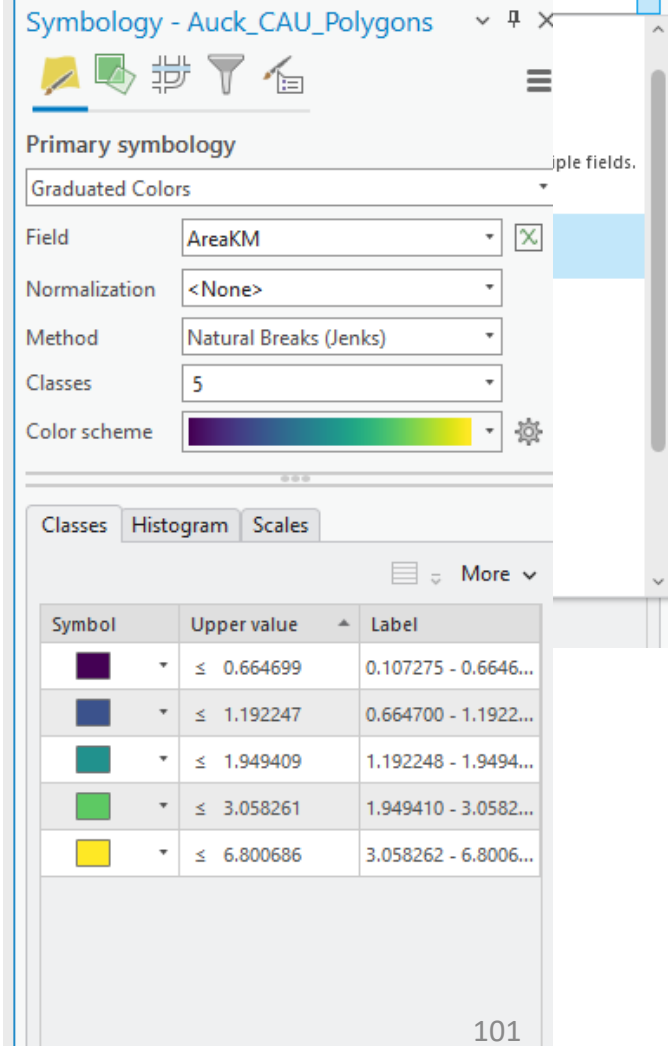
Changing symbology

- Right click the shapefile name in the Contents window
- Select Symbology
- In the Symbology window click on the coloured box next to “Symbol”
- Change the colour and click “Apply” when you’re done



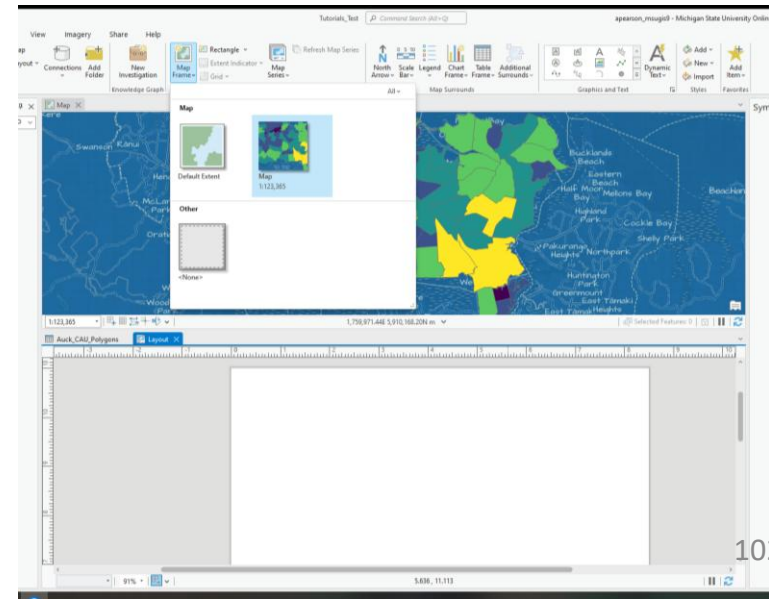
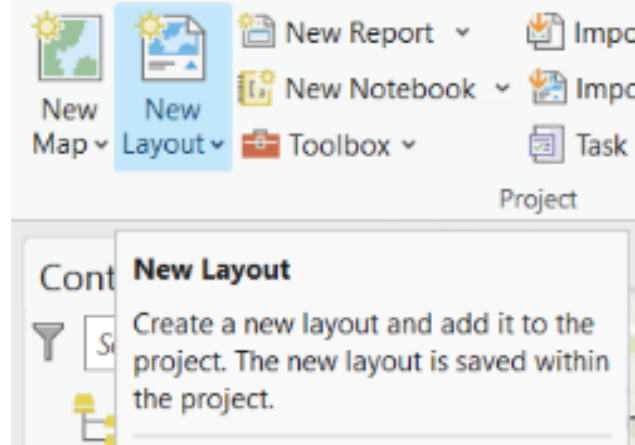
Changing symbology

- Display census area units in varying colours, by their area
 - Re-open the “Symbology” window.
 - Click on “Primary Symbology”
 - Choose “Graduated Colors” from the dropdown menu
 - Click the “Field” dropdown menu
 - Change to “AreaKM”
 - Have a play with:
 - number of groups
 - method of choosing groups
 - colour scheme graduated colours
 - Click Apply and Ok



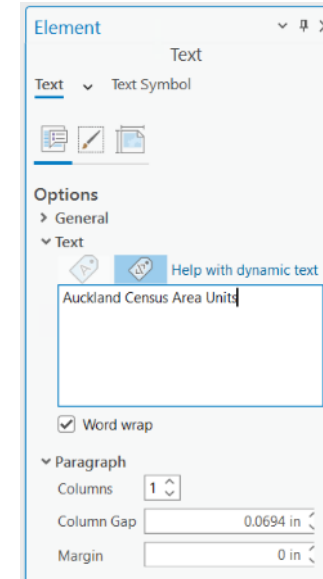
Making a map

- Change to layout view
- Select “Map Frame”
- Select generated map scale option
- Click “Zoom to Layer”
- Use mouse to draw map frame



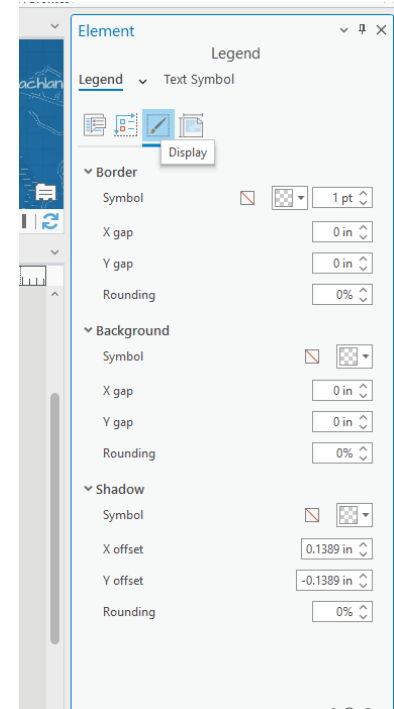
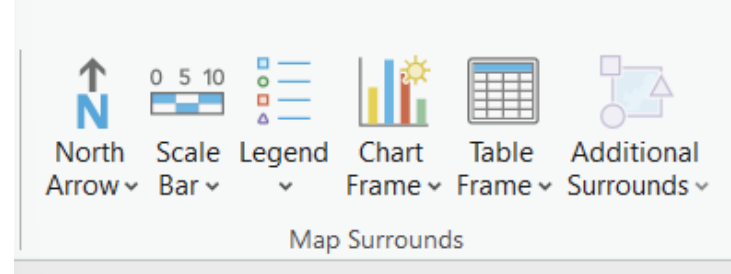
Making a map

- Add a title, scale bar and legend
 - Go to Insert, “Dynamic text”
 - Select “Name (*Map Frame*)”
 - Use mouse to click & drag to draw a text box
 - Double click “Text 1”
 - Name the map “Auckland Census area units”



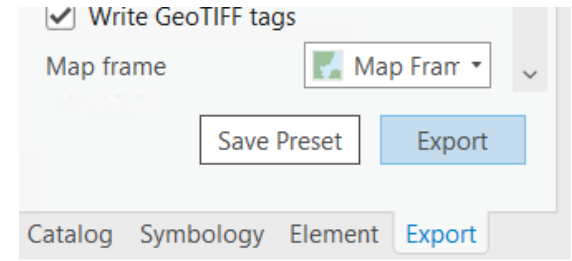
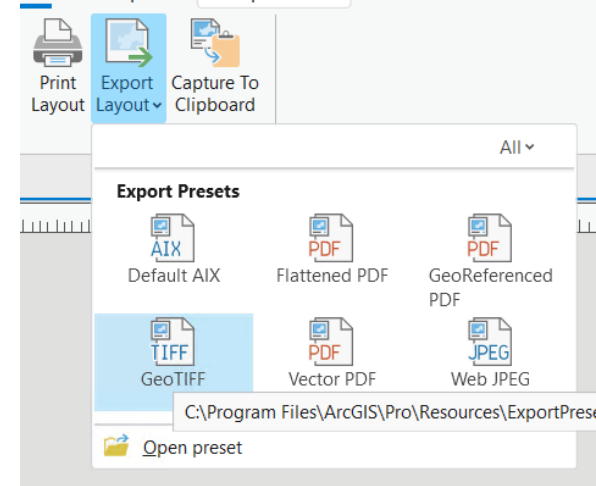
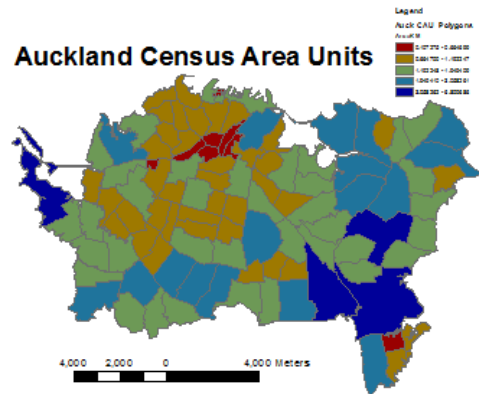
Making a map

- Add a scale bar and legend
 - Go to “Insert” and select “Scale Bar”
 - Go to “Insert” and select “Legend”
 - Start with option “A” and use mouse to draw it on the layout map
 - Use the Element tab to make a solid background



Exporting map

- Export map as a jpeg or tiff
 - Go to “Share”, select “Export layout”
 - Choose GeoTIFF
 - Choose a file name, location
 - Notice options for dpi, etc
 - Click “Export”



Tutorial 1

A. What is the largest area out of all the CAUs (km²)?

Tutorial 1

A. What is the largest area of all the CAUs (km²)?

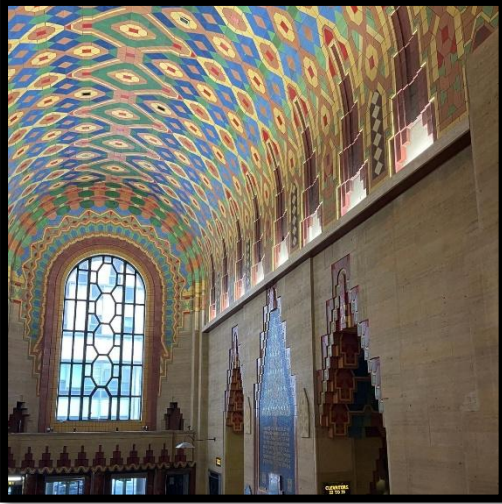
Mount Wellington South 6.800686 km²

How did you solve this?

Lunch

Public Health
Summer School
Wellington

Detroit,
MI



Section 4: Creating spatial data – Geocoding

Geocoding

- Assigning a geographic location (x, y cartesian coordinates) to an address location
 - e.g. convert 133 Molesworth Street, Wellington to 748894, 5429415
- Not error free!

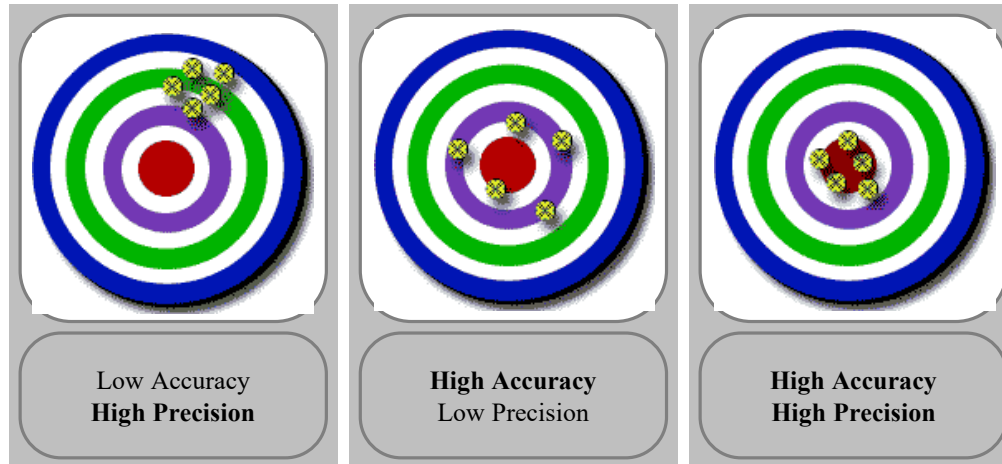
- What types of errors?
 - Did not geocode at all – shows up as a 0 lat 0 long
 - Geocoded to clearly the wrong place – northern hemisphere
 - Geocoded to wrong place within the city
 - Geocoded to the right place but was marginally off (a few metres)

- How to correct errors?
 - Sort latitude (or long) and assess there are no obvious problems
 - Delete rows that did not geocode at all or were completely wrong
 - Select a random subset of geocoded locations and check spatial accuracy
 - Put the name of the business or the address into google earth
 - Copy and paste the X,Y of the correct location into Excel
 - Map those 'gold standard' locations in ArcMap
 - Measure the distance between the geocoded and the gold standard locations
 - Report this!
 - Find failed locations in Google Earth and perform the steps above

Geocoding

Accuracy - degree to which information on a map or in a digital database matches true or accepted values.

Precision - the level of measurement and exactness, may measure position to a fraction of a unit.





Geocoding



Sl. No	Area	Latitude	Longitude
1	Wardha, Maharashtra, India	20.427012814257385	78.7060546875
2	Tumkur, Karnataka, India	13.239945499286312	76.5966796875
3	Orissa, India	21.69826549685252	84.0673828125
4	Gadchiroli, Chhattisgarh, India	19.932041306115536	80.15625



- **Geographic Coordinate System** - uses an angular unit of measure (e.g. latitude and longitude), a prime meridian and a datum based on a spheroid
- **Projected Coordinate Systems** converts the latitude and longitude that fall on the 3-D surface of the earth onto a 2-D surface for use in GIS (important for distances)
- In NZ, the two most common projections are :
 - New Zealand Map Grid
 - New Zealand Transverse Mercator
 - * used in tutorial

- ArcPro: needs an address locator file (need credits)
 - Garbage in, garbage out
 - Road files can be proprietary \$\$\$
- Html, using Google maps
- R, using Google maps
 - Need Google API key = credit card on account
 - Google allows each IP address to geocode 10,000 addresses/day for free
 - Benefit: free, pretty up-to-date

Soon we will work Tutorial 2

- Bonus material: Downloading spatial data from Stats NZ
- Appending a data table to spatial data
- *This is optional. You can always come back and do this later.*

- Often we have spatial data
 - SA2 boundaries in 2018
- And we have tabular data
 - Mortality rates for Northland SA2s in 2010
- Need to combine the two to make a map or conduct spatial analysis
 - What is the spatial pattern of low mortality SA2s in this region?

Other examples of appending data

- We have addresses of all bottle stores across Auckland – we geocoded these
- Then, a group of students went to each one and recorded whether or not the store sells tobacco
- We want to add the tabular data we collected to the spatial point data to make a map of bottle stores that sell tobacco

How to...

- Open ArcPro and add your downloaded SA2.shp
- Now, add the ASM_Rate_bySA2.csv file
- Join csv to the shp, based on attribute (SA2 id in both files)

How to...

- This makes a new column of data in the SA2.shp
 - You can make a map by changing the symbology to reflect mortality (e.g. quintiles)
 - Make a map with a legend, scale bar, and title
 - Export the jpeg or tiff to work folder
-
- **Remember that joining information is not permanent. Shapefiles must be exported to save joined data. Also note <<NULL>> missing data in a join will be saved as a 0.**

Tutorial 2 - geocoding

- A. How many points did you geocode (stroke case locations)?
- B. Do you see any errors? If so, how many and what type of errors?

Tutorial 2 - geocoding

A. How many points did you geocode (stroke case locations)?

287

B. Do you see any errors? If so, how many and what type of errors?

7 ties...anything else?

How to remove errors

- Use the Editor dropdown menu to edit your geocoded (exported) shapefile
- Open the attribute table
- Highlight the rows to be deleted
- Delete using the X button in the attribute table or pressing “Delete” on keyboard
- Save edits and stop editing

Future work

- If you have extra time....
- Try the Google Earth options
- Repeat steps you performed – excel, display X,Y

Best practices

- Data quality checks for correct address
- Data quality checks for spatial accuracy
- Decisions about assigning point location (e.g. stadium)
- Reporting geocoding methods and results (e.g., percent to rooftop, spatial precision)

Section 5: Geographic definitions and techniques

Definitions and techniques — DAY 2 but overview today

- Centroid
 - Geometric
 - Population-weighted
- Buffer
- Proximity
 - Euclidean
 - Network

**refer to course book (Thornton et al. 2011)*

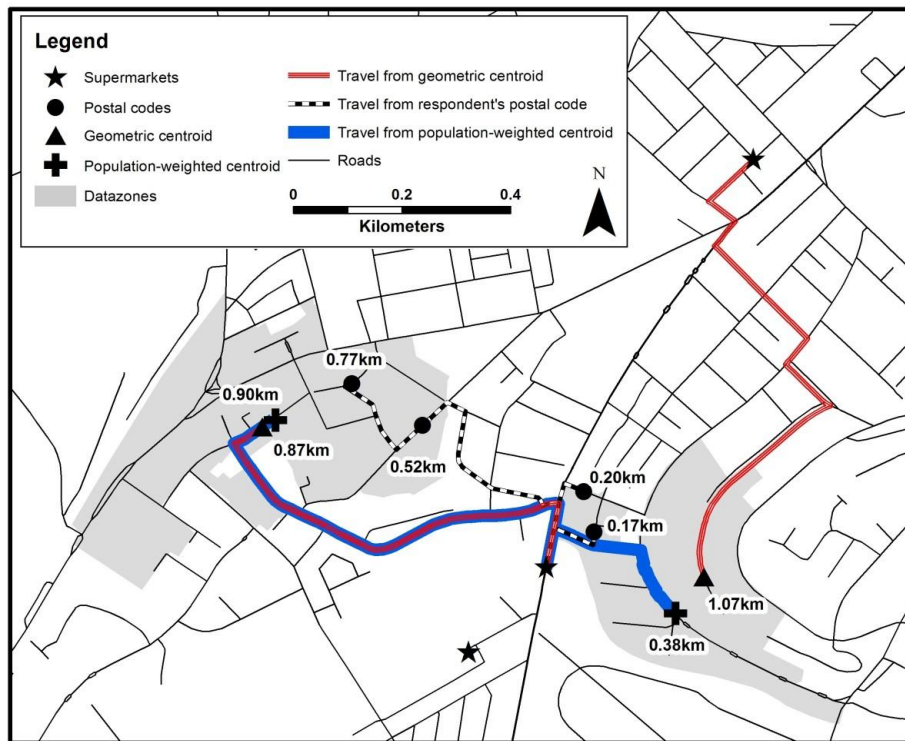
Definitions and techniques — DAY 2 but overview today

- Density
 - Kernel density estimation
- Accessibility
- Connectivity
- Walkability
- Clustering

**refer to course book (Thornton et al. 2011)*

Centroid

- Single point representing the **centre** of a spatial unit
- Used to measure proximity, or may serve as the centre of a buffer
 - Geometric
 - Population-weighted



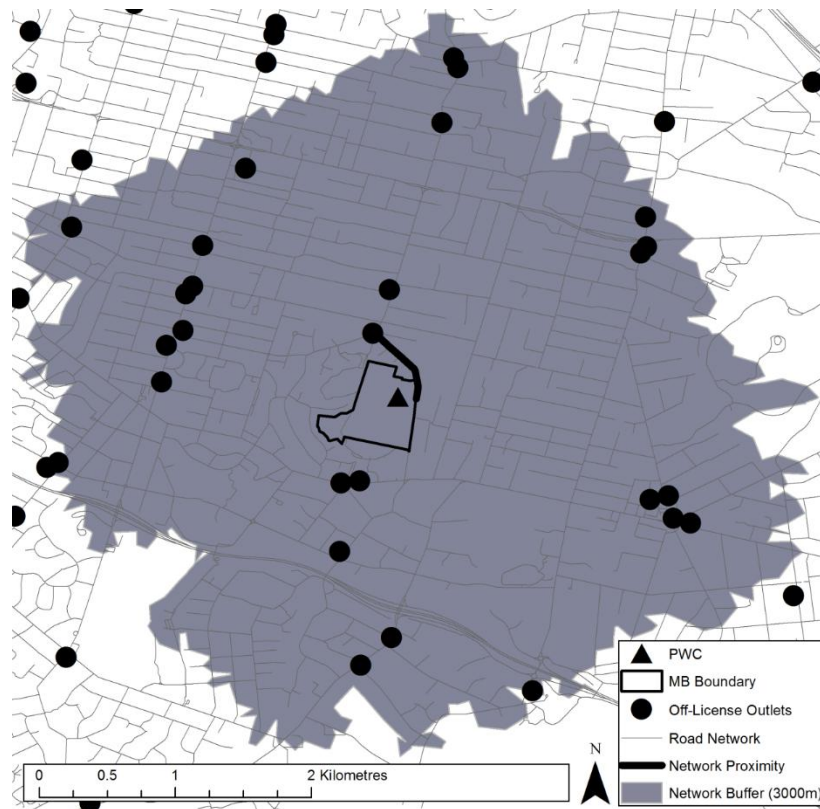
Source: Thornton *et al.* *International Journal of Health Geographics* 2012

11:29

130

Proximity

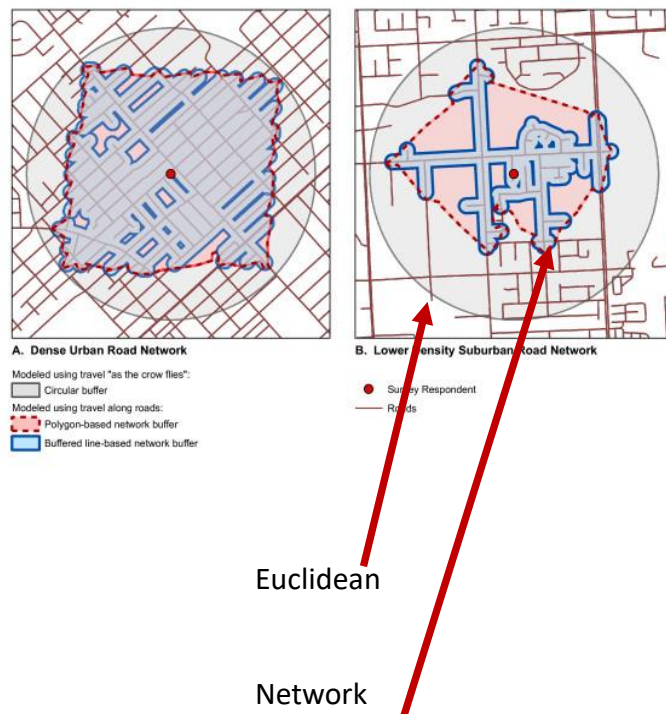
- Distance between points, point to line/polygon
- Travel
- Euclidean distance
- Network distance
- Social
 - Techniques of social distancing through language, imagery



Buffer

- Boundaries placed around areas (polygons), lines or points
- May use a centroid or point as centre
- Set radius
- Then euclidean or network buffering technique used to assign buffer at given distance from centre point

Source: Oliver et al. *International Journal of Health Geographics* 2007 6:41.



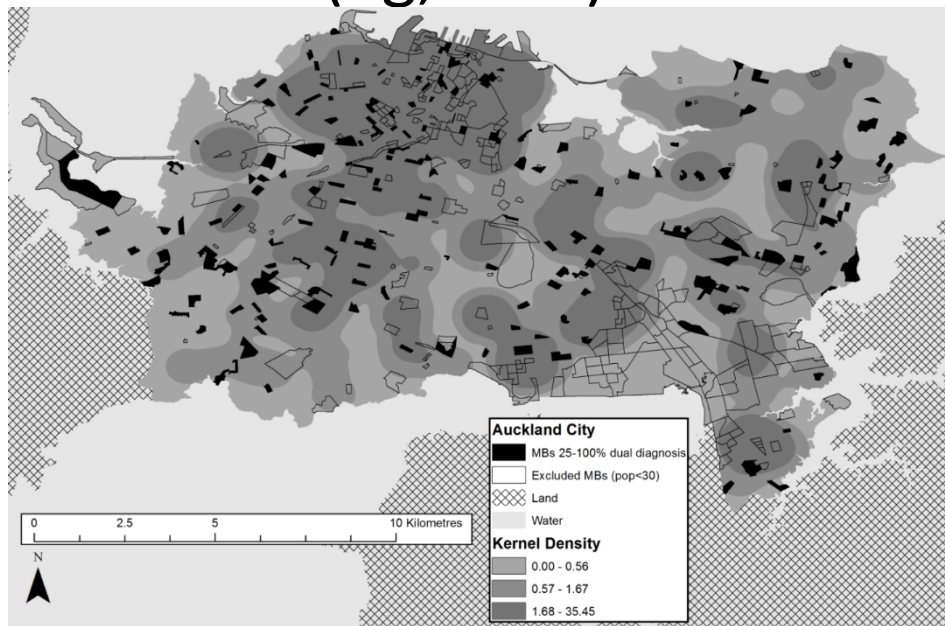
- A measure of intensity – exposure to features of the built environment
- Relates to accessibility
 - Sum (count) of features within specified area (buffer)
 - Count per population/geographic area
 - Kernel density
- Continuous measures assume: linear relationship between feature and health outcome

Density

$$p(x) \cong \frac{k}{NV} \text{ where } \begin{cases} V & \text{volume surrounding } x \\ N & \text{total \#examples} \\ k & \text{\#examples inside } V \end{cases}$$

- Kernel density estimation
 - Different from distance - relies on defining the interpolation interval (eg, 1km)

Thus, a zero indicates that there were no outlets within one kilometre.

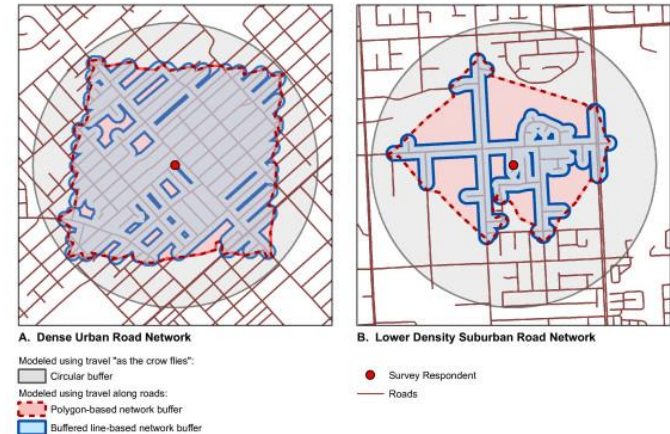


Accessibility

- Ease of access to a feature
 - Better access = lower travel costs
 - \$
 - Time
 - Distance
 - Resources
- Built environment



- Availability and directness of travel routes to move through network from origin to destination
 - Spacing between roads
 - Number intersections
 - One-way streets



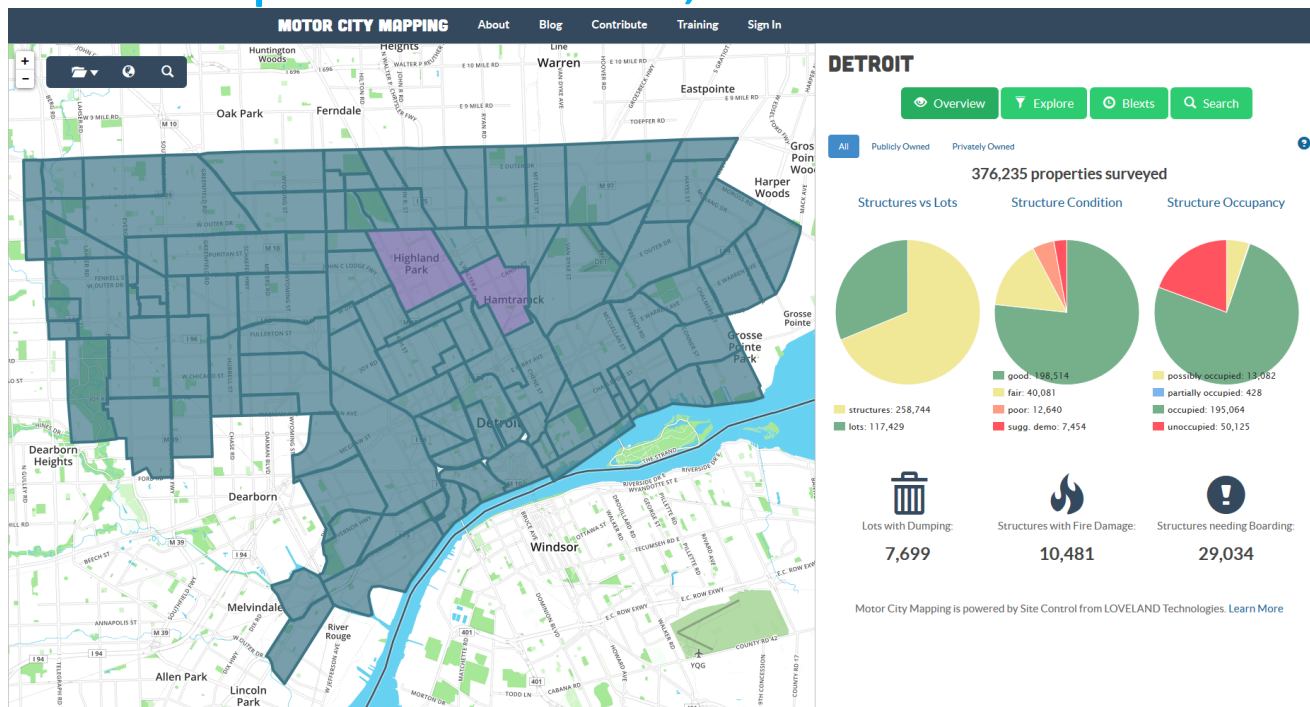
- Functionality
- Safety
- Aesthetics
- Destinations
 - Footpaths
 - Lighting
 - Crime

Can you think of an application of any of these techniques in work you already conduct?

Section 6: Defining neighborhoods, applications and ethical health geography

At what scale does this phenomenon of interest operate?

- Example 1 – Detroit, MI



At what scale does this phenomenon of interest operate?

- I wanted to know if there was a relationship between the human microbiome diversity and neighborhood conditions in Detroit
- Do deteriorating built environment features (abandoned buildings) or features of urban greening (tree plantings) = lower/higher microbial diversity

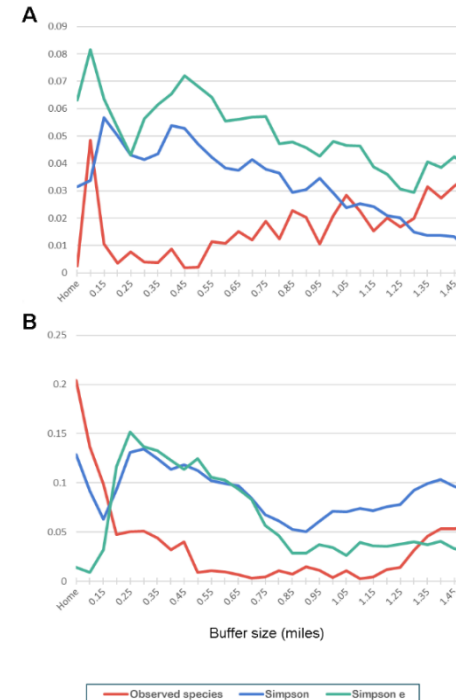
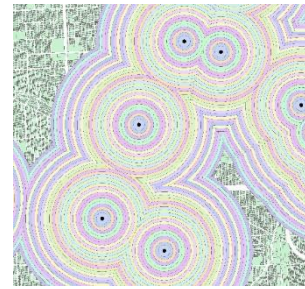


At what scale does this phenomenon of interest operate?

- However, since there has not been existing research in this area, I was not sure which scale would be appropriate for the neighborhood influence on the microbiome.
- *So, I decided to conduct a sensitivity analyses to determine 'neighborhood size'*

Scale – Example 1

- Generated concentric rings around home locations, assigned 'blight' and 'greening' values based on parcels within each circle
- Tested correlation between blight and diversity for each circle size
- Tested correlation between greening and diversity for each circle size



Scale – Example 2

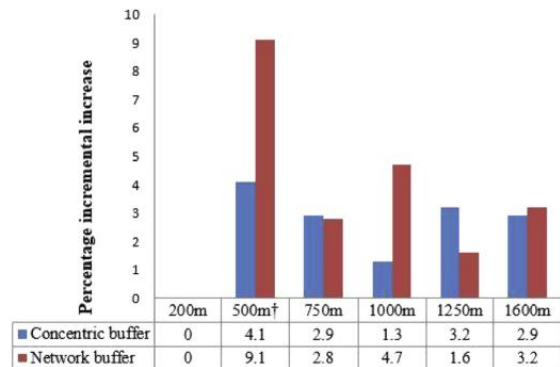
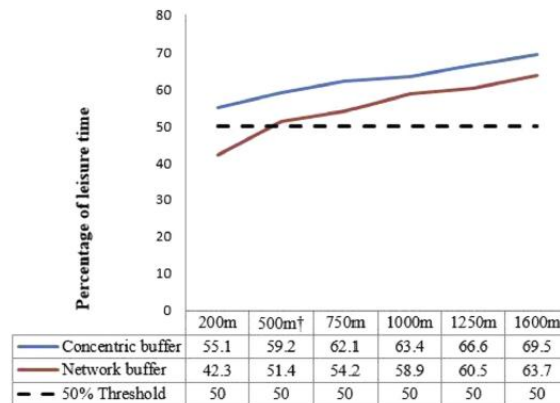
- Example 2 – Wellington, NZ, Kids'Cam data
- Lots of research on children evaluates their exposure to things in their neighborhood
 - E.g. Children's exposure to alcohol marketing
- Most of this research defines neighborhood:
 - based on conventions, other research or an estimated walking time
- I wanted to know how we should define a child's neighbourhood extent based on actual activity spaces of children observed in a study

Defining neighbourhood extent



Creating criteria for neighbourhood research

- Determined 500m best representation
- Over 50% of time
- Largest incremental difference



†The 500m buffer was chosen as children's neighbourhood extent based on having the greatest percentage difference between mean values of adjacent buffer sizes and the point at which participants spent the majority of their time in both concentric and network buffers.

- We use GIS
 - To record spatial information (property records)
 - To display spatial information (maps)
 - To evaluate spatial relationships
- How could we use GIS to inform health policy?
- How could we use GIS to inform health care services decisions?
- How might you use GIS in your own work?

1) Urban Revitalization and Mental Health

- Gentrification-like change near home
 - Subjective and objective measures
- Value of amenities near home

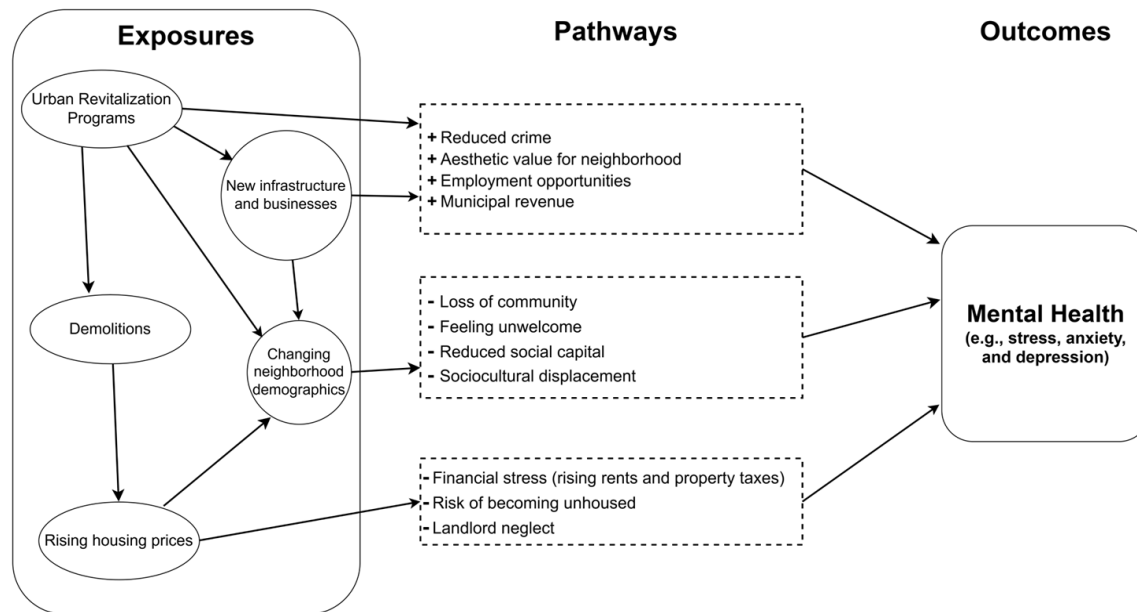
2) Student nature exposure

- What type of nature is available on college campuses?
- Assess available nature outside classrooms
 - More nature, specifically trees, associated with lower behaviour problems

3) GPS processing

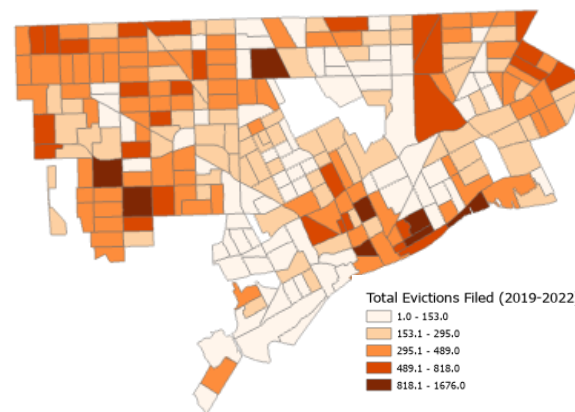
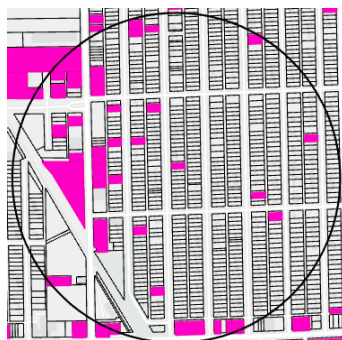
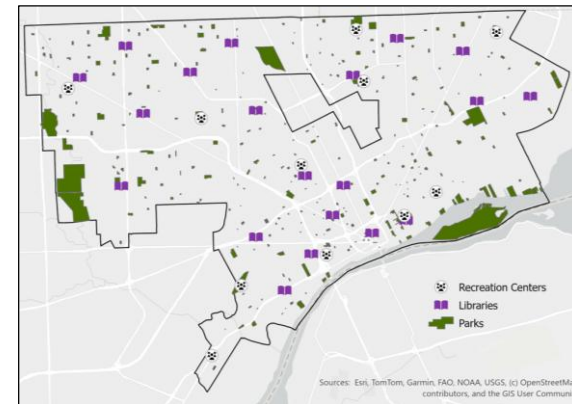
- Enhance techniques used to process GPS data
 - Consider indoor/outdoor, driving/non-driving, etc.

- Goal: Explore how urban revitalization is associated with mental health?



Gentrification, neighbourhood amenities, and mental health

- Rely on neighborhood level data from secondary sources
- Link to participants based on home location

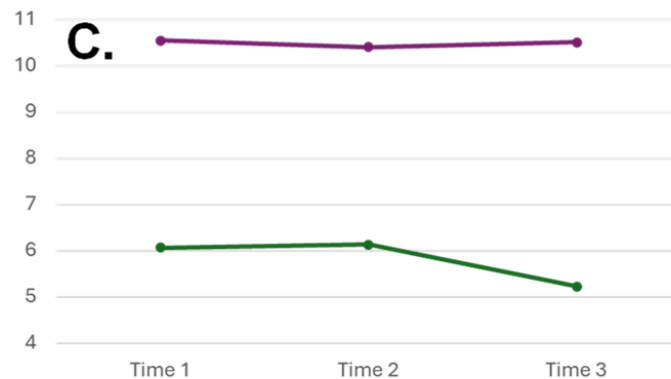
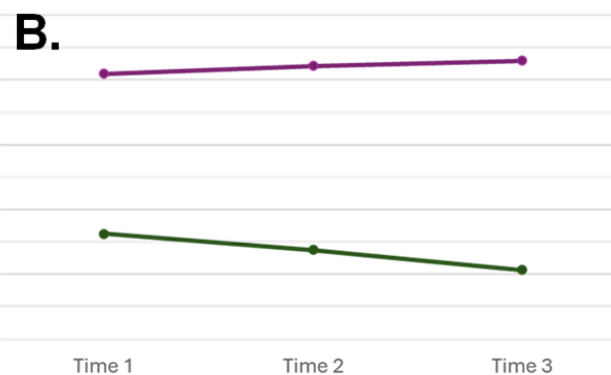
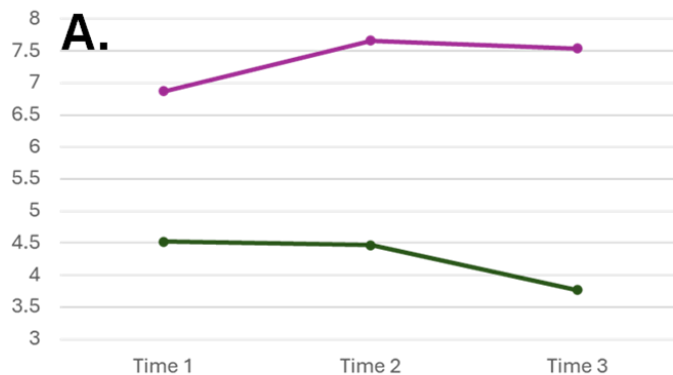


Gentrification, neighbourhood amenities, and mental health

- Participants with high amenities available have significantly better mental health

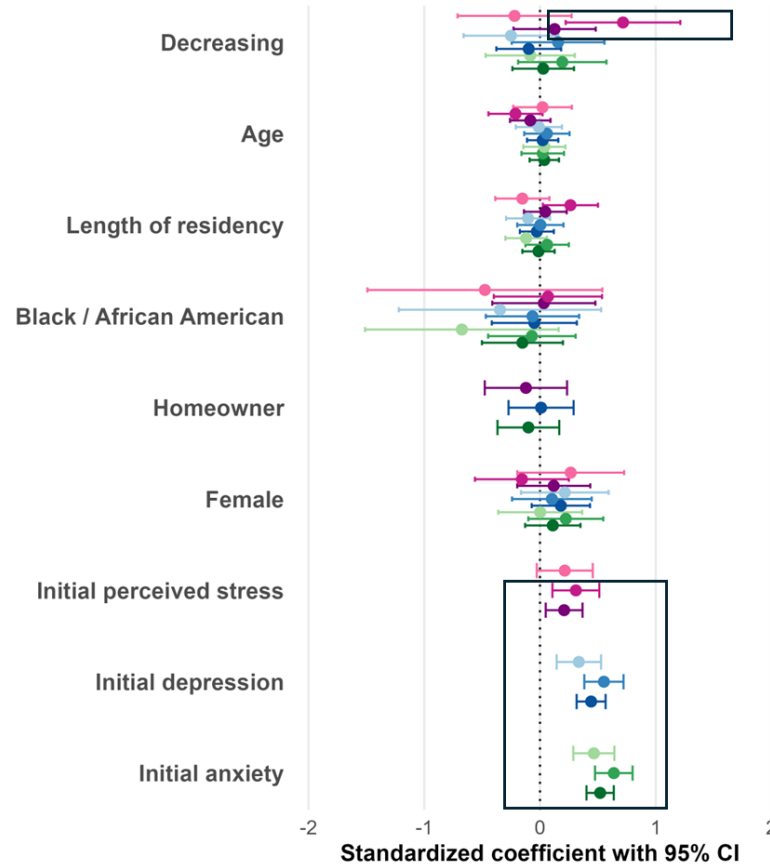
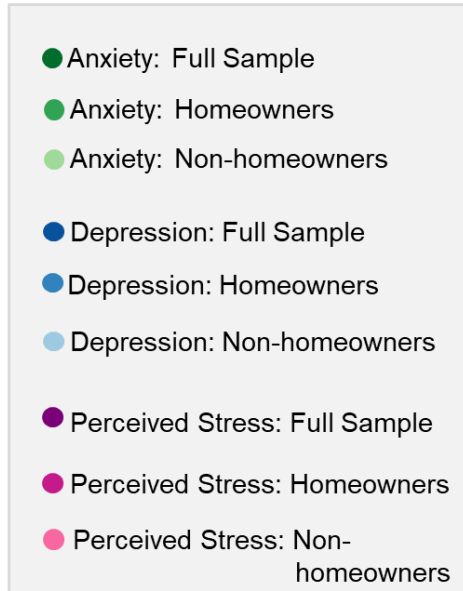
	Anxiety			Depression			Perceived Stress			SWLS		
	coeff	p value	SE	Coeff	p value	SE	coeff	p value	SE	coeff	p value	SE
No gentrification	ref			ref			ref			ref		
Gentrifiable	-2.07	0.473	2.88	-0.95	0.733	2.77	-2.08	0.204	1.62	3.58	0.304	3.39
Gentrifying	-9.10	0.040	4.35	-5.96	0.169	4.28	-3.02	0.196	2.30	11.63	0.021	4.62
Female	3.49	0.114	2.18	2.78	0.179	2.05	1.10	0.349	1.16	4.86	0.130	3.07
Length of residency	0.04	0.588	0.07	0.03	0.728	0.07	0.04	0.278	0.04	0.01	0.928	0.08
Homeowner	-0.93	0.743	2.82	0.56	0.839	2.73	-0.54	0.737	1.60	-4.22	0.196	3.15
Black/African American	-8.48	0.011	3.24	-4.52	0.158	3.17	-3.32	0.042	1.59	7.12	0.042	3.26
Age	0.12	0.170	0.09	0.05	0.542	0.09	-0.03	0.502	0.05	-0.23	0.062	0.11
Initial mental health	0.38	<0.001	0.10	0.27	0.008	0.10	0.02	0.572	0.04	0.30	0.221	0.24
Adjusted R-squared	0.26			0.11			0.04			0.37		

Subjective alternatives



Subjective alternatives

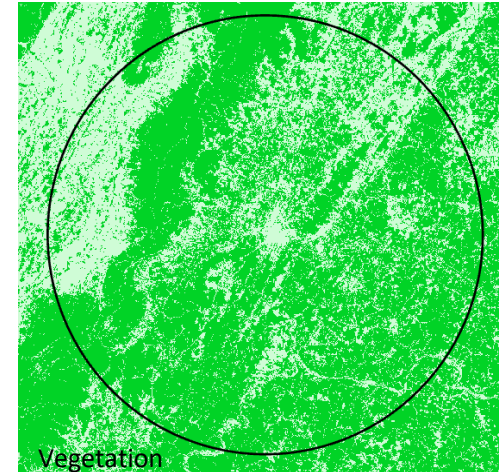
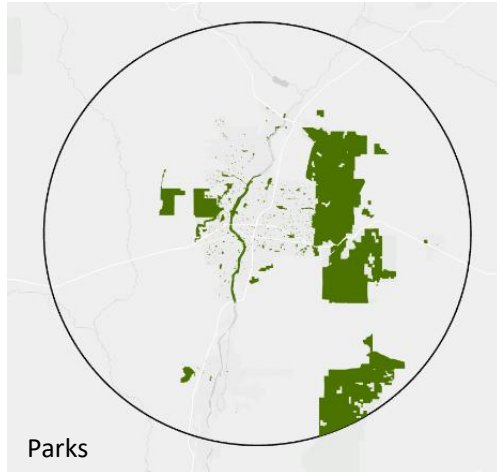
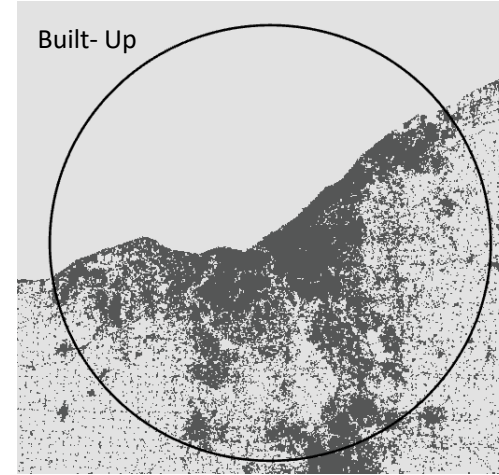
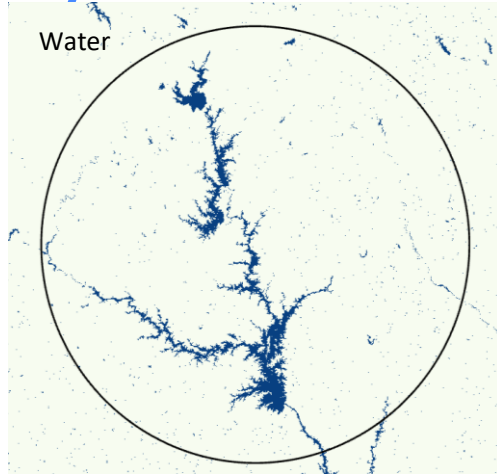
Changes in resident empowerment



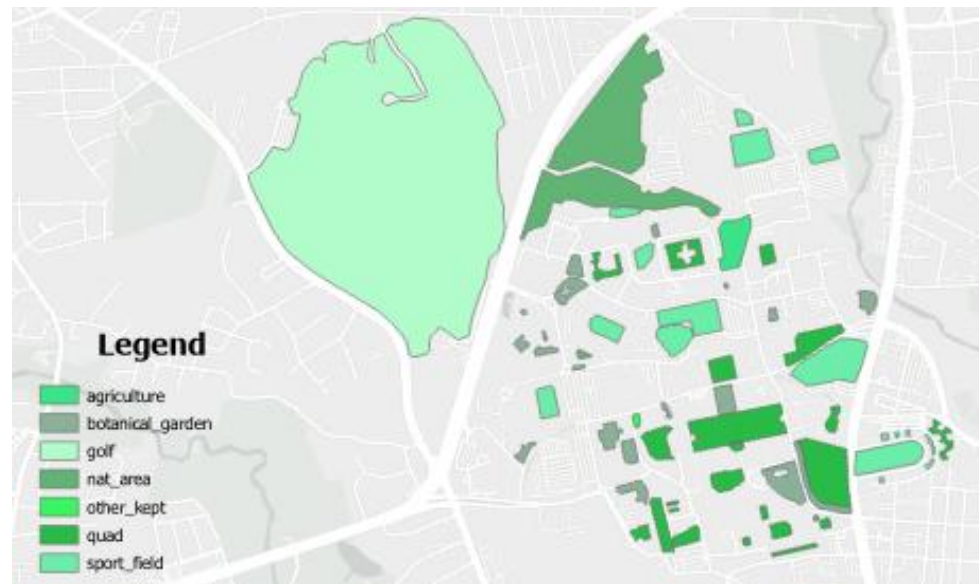
- NatureIDEAS project
- *Goal:* understand how much time students are spending outdoors, or in nature
- *Sample:* 9 universities across the United States
- *Data:* satellite imagery, secondary sources, novel spatial collected/constructed



University nature



- Identify relevant nature on campus
 - Set definitions
 - Agriculture
 - Botanical garden
 - Golf
 - Natural area
 - Other kept grounds
 - Quad
 - Sports field
 - Create shapefile for each campus
 - Validate with campus grounds keeper



- *Goal:* clean and classify GPS data, link to other devices

Open accessOriginal research

BMJ Open

Systematic review of best practices for GPS data usage, processing, and linkage in health, exposure science and environmental context research

Amber L Pearson ¹ Calvin Tribby² Catherine D Brown ³ Jiue-An Yang² Karin Pfeiffer⁴ Marta M Jankowska²

To cite: Pearson AL, Tribby C, Brown CD, et al. Systematic review of best practices for GPS data usage, processing and linkage in health, exposure science and environmental context research. *BMJ Open* 2024;14:e077036. doi:10.1136/bmjopen-2023-077036

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2023-077036>).

Received 26 June 2023
Accepted 16 January 2024

ABSTRACT
Global Positioning System (GPS) technology is increasingly used in health research to capture individual mobility and contextual and environmental exposures. However, the tools, techniques and decisions for using GPS data vary from study to study, making comparisons and reproducibility challenging.
Objectives: The objectives of this systematic review were to: (1) identify best practices for GPS data collection and processing; (2) quantify reporting of best practices in published studies; and (3) discuss examples found in reviewed manuscripts that future researchers may employ for reporting GPS data usage, processing and linkage of GPS data in health studies.
Design: A systematic review.
Data sources: Electronic databases searched (24 October 2023) were PubMed, Scopus and Web of Science (PROSPERO ID: CRD42022321166).
Eligibility criteria: Included peer-reviewed studies published in English met at least one of the criteria: (1) protocols involving GPS for exposure/context and human health research purposes

STRENGTHS AND LIMITATIONS OF THIS STUDY
→ This systematic review used standard database search to find articles as well as citation assessment of review articles to encompass a comprehensive set of articles.
→ Article types included association focused, methodological development, feasibility studies and tracking tool comparisons providing a broad scope of Global Positioning System (GPS) applications in human health research.
→ We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses reporting guidelines to report on results.
→ We did not consider the necessity of differential reporting for certain health outcomes or subpopulations studied.
→ A potential limitation of this review was the omission of studies using mobile phone apps to collect GPS data.

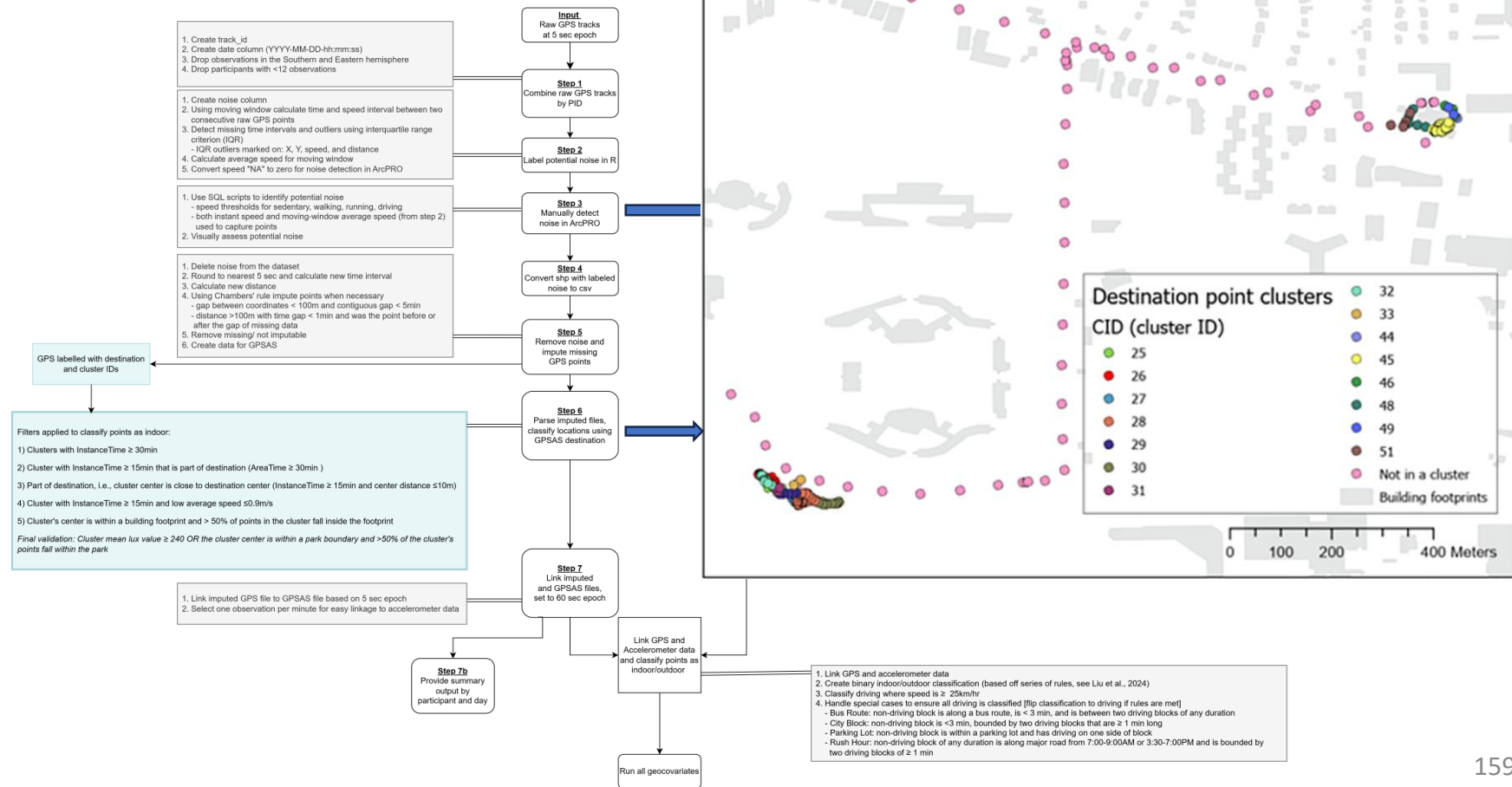
BMJ Open: first published as 10.1136/bmjopen-2023-077036 on 1 February 2024. Downloaded from <https://bmjopen.bmj.com/> by <https://www.gpi.com/> on 04 February 2024. Protected by copyright.

Table 1 Themes and specific best practices discussed as relevant for GPS data collection and processing, identified in best practice manuscripts (n=8)		
Best practice theme	Discussed in n (%) manuscripts	Specific best practices
P1—GPS device	4 (50) ^{12 13 20 186}	► Brand of GPS device used ► Model of GPS device used
P2—sampling frequency	6 (75) ^{12 13 20 22 23 186}	► GPS sampling frequency
P3—wear time	6 (75) ^{12 13 19 20 22 186}	► Days/periods of wear time
P4—GPS missing data	7 (88) ^{12 13 19–21 23 186}	► Signal loss %
P5—noise	6 (75) ^{12 13 19–21 23 186}	► Identification method of noise ► Noise identification threshold ► Noise % ► Tool/method used for dealing with noise
P6—imputation	4 (50) ^{12 13 19 20}	► Imputation performed ► Imputation method ► Imputation threshold ► Tool used for imputation ► Points imputed %
P7—linkage	6 (75) ^{12 19–22 186}	► Data type linked ► Linkage epoch ► Tool used for linkage ► Data lost in linkage process %
P8—data inclusion	6 (75) ^{12 13 19 20 23 186}	► Criteria for data inclusion specified ► Minutes of data to be a valid wear day/period ► Valid wear days required ► % participants lost due to compliance criteria
GPS, Global Positioning System.		

- Primarily automated, with some manual checks
- Relies on GPS + accelerometer (any brands)
- Accounts for speed and duration
- Removes noise
- Imputes under certain conditions
- Identifies destinations (typically indoor)
- Also accounts for: bus routes, rush hour, repeated stop signs in residential areas

GPS Processing

Red points captured as "noise"



	adequate GPS data	inadequate GPS data	did not receive GPS unit	Total	p-value
	<i>n=438</i>	<i>n=402</i>	<i>n=331</i>	<i>n=1171</i>	
age	56.1 (14.8)	53.8 (14.6)	53.1 (14.2)	54.5 (14.6)	0.001
female	57.8	63.6	53.1	58.2	0.021
income <\$10,000	43.8	51	57.9	50.4	0.004
is employed	29.9	29.6	26.3	28.8	0.507
is married	19.2	14.3	13.3	15.9	0.055
owns home	50.4	35.5	33.2	40.7	<0.001
Black	85.6	88.7	91.4	88.3	0.046
number of kids at home	0.4 (1.1)	0.5 (1.1)	0.4 (1.0)	0.4 (1.1)	0.594
years resident	14.5 (15.5)	12.3 (14.8)	13.2 (16.0)	13.4 (15.4)	0.144
perceived stress score	22.6 (4.6)	23.2 (5.2)	24.0 (6.5)	23.2 (5.4)	0.006
anxiety	54.9 (9.8)	56.7 (9.9)	57.2 (11.3)	56.1 (10.3)	0.005
heart rate	77.3 (13.4)	77.1 (12.9)	80.9 (14.8)	77.9 (13.6)	0.008
c-reactive protein	2.3 (9.3)	2.4 (3.5)	2.0 (3.2)	2.3 (6.6)	0.849
BMI	29.5 (8.2)	30.5 (8.5)	28.6 (7.8)	29.6 (8.2)	0.015
total physical activity	27.8 (8.2)	26.1 (8.7)	29.1 (11.6)	27.1 (8.5)	0.022
NDVI	0.2 (0.1)	0.2 (0.1)	0.3 (0.1)	0.3 (0.1)	<0.001
tree cover	12.3 (3.2)	12.6 (2.9)	12.1 (3.1)	12.3 (3.1)	0.107
noise	166.6 (1.6)	167.3 (1.6)	166.8 (1.6)	166.9 (1.6)	<0.001
poverty	31.4 (10.6)	31.5 (9.7)	31.3 (12.6)	31.2 (10.9)	0.9687
violent crime rate	82.1 (57.2)	81.9 (50.3)	97.4 (47.5)	86.5 (52.8)	<0.001
property crime rate	95.0 (60.4)	98.2 (60.1)	124.3 (67.1)	104.5 (63.7)	<0.001
vacant lots near home	27.1 (21.3)	23.2 (19.5)	27.8 (20.8)	26.1 (20.7)	0.007

A bit about outdoor time in our sample

	Low time outdoors in PA	High time outdoors in PA	Total	p-value
	<i>n</i> = 56	<i>n</i> = 56	<i>n</i> = 112	
Age, mean (sd)	58.4 (15.2)	57.4 (13.2)	57.9 (14.2)	0.415
Female [§] , %	58.9	53.6	56.3	0.568
Income [§] , % <\$10,000	37.0	52.1	44.7	0.140
Caretaker/unpaid employment [§] , %	1.9	3.7	2.8	
Employed 40+ hours/week [§] , %	16.7	18.5	17.6	
Employed 1-39 hours/week [§] , %	13.0	5.6	9.3	0.670
Unemployed [§] , %	37.0	42.6	39.8	
Retired [§] , %	29.6	29.6	29.6	
Student [§] , %	1.9	0	0.9	
Married/partnered [§] , %	23.2	19.6	21.4	0.645
Own home [§] , %	56.6	46.8	52.0	0.328
Black/African American [§] , %	82.1	87.5	84.8	0.430
Number of children <18y living in home, mean (sd)	0.3 (0.6)	0.4 (0.9)	0.3 (0.8)	0.800
Household size, mean (sd)	2.4 (2.7)	2.1 (1.3)	2.3 (2.1)	0.209
Years resident, mean (sd)	14.8 (15.5)	14.8 (18.0)	14.8 (16.7)	0.716
PSS score, mean (sd)	22.2 (3.6)	24.2 (5.2)	23.2 (4.5)	0.038
Anxiety, mean (sd)	52.2 (8.5)	56.3 (10.2)	54.3 (9.6)	0.029
Depressive symptoms, mean (sd)	48.2 (7.8)	53.0 (8.9)	50.6 (8.7)	0.004
mean NDVI, mean (sd)	0.29 (0.1)	0.27 (0.1)	0.28 (0.1)	0.384
% NDVI ≥0.2, mean (sd)	67.7 (11.7)	64.3 (12.5)	66.0 (12.2)	0.142
Greenness, mean (sd)	0.25 (0.1)	0.21 (0.1)	0.23 (0.1)	0.007
Vacant lots per minute, mean (sd)	0.47 (0.4)	0.21 (0.2)	0.34 (0.4)	<0.001
Vacant houses per minute, mean (sd)	0.10 (0.12)	0.03 (0.03)	0.06 (0.09)	0.0001
Perceived aesthetics, mean (sd)	14.4 (5.2)	14.9 (5.5)	14.6 (5.3)	0.628
Minutes spent in parks, mean (sd)	10.6 (27.0)	27.6 (49.3)	19.1 (40.5)	<0.001

Bold font: $p \leq 0.05$

[§] *p*-values calculated using chi-squared test; otherwise *t* test or nonparametric test as appropriate was used.



Using GPS data to assign exposures

- *Moving from static residential exposures to activity space exposures throughout the day*
- Urban heat – does exposure to greenery buffer effects of heat on stress?
- Visual exposures – greenness, disorder
- Noise
- Birdsong
- Gun violence
- Air pollution

- Maps can make people and places vulnerable!
- Consider a map of:
 - The “most contaminated” neighbourhoods in Wellington
 - HIV cases in New Zealand
 - Rates of “preventable” hospital deaths at New Zealand hospitals
- The media can use maps for sensational purposes
- Map data can be ‘reverse’ geocoded

02 Whakapapa | Relationships

2.1 Context. All data has a whakapapa (genealogy). Accurate **metadata** should, at minimum, provide information about the provenance of the data, the purpose(s) for its collection, the context of its collection, and the parties involved.

2.2 Data disaggregation. The ability to **disaggregate** Māori data increases its relevance for Māori communities and iwi. Māori data shall be collected and coded using categories that prioritise Māori needs and aspirations.

2.3 Future use. Current decision-making over data can have long-term consequences, good and bad, for future generations of Māori. A key goal of Māori data governance should be to protect against future harm.

03 Whanaungatanga | Obligations

3.1 Balancing rights. Individuals' rights (including privacy rights), risks and benefits in relation to data need to be balanced with those of the groups of which they are a part. In some contexts, collective Māori rights will prevail over those of individuals.

3.2 Accountabilities. Individuals and organisations responsible for the creation, collection, analysis, management, access, security or dissemination of Māori data are accountable to the communities, groups and individuals from whom the data derive.

05 Manaakitanga | Reciprocity

5.1 Respect. The collection, use and interpretation of data shall uphold the dignity of Māori communities, groups and individuals. Data analysis that stigmatises or blames Māori can result in collective and individual harm and should be actively avoided.

5.2 Consent. Free, prior and informed consent (FPIC)² shall underpin the collection and use of all data from or about Māori. Less defined types of consent shall be balanced by stronger governance arrangements.

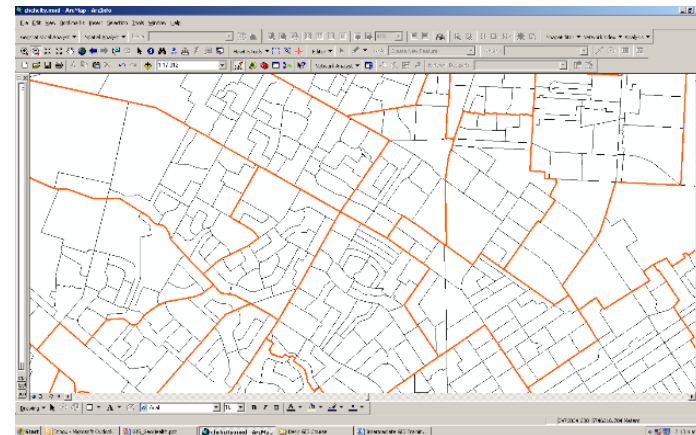
Special considerations – health or sensitive data

- When working with health data, we may be required by data providers (and ethics) to consider:
 - Preserving anonymity of survey participants
 - Rarity of event (infantile astrocytoma)
 - Stigmatism of places or population groups
- We can effectively and ethically do this:
 - Jitter individual point locations
 - Aggregate to larger population/geographic unit

Considerations

Spatial resolution of health data

- Health data collected for individuals BUT **rarely released** (Privacy and confidentiality)
- So aggregated to predefined geographical units e.g. Statistics NZ
- May bear little relationship to areas defined by socioeconomic, demographic criteria (What is a neighbourhood?)
- May not be rational units for GIS analysis
- NZ mortality data only at CAU (**red**) not meshblock (**black**) level



Data issues

Individual level data

- Advantages
 - More 'honest'
 - People live in exact locations, not artificial areas
 - Overcomes scale and aggregation problems
- Disadvantages
 - Data confidentiality: too much detail
 - Data availability: often hard to obtain
 - Techniques more complex

Options for maintaining anonymity

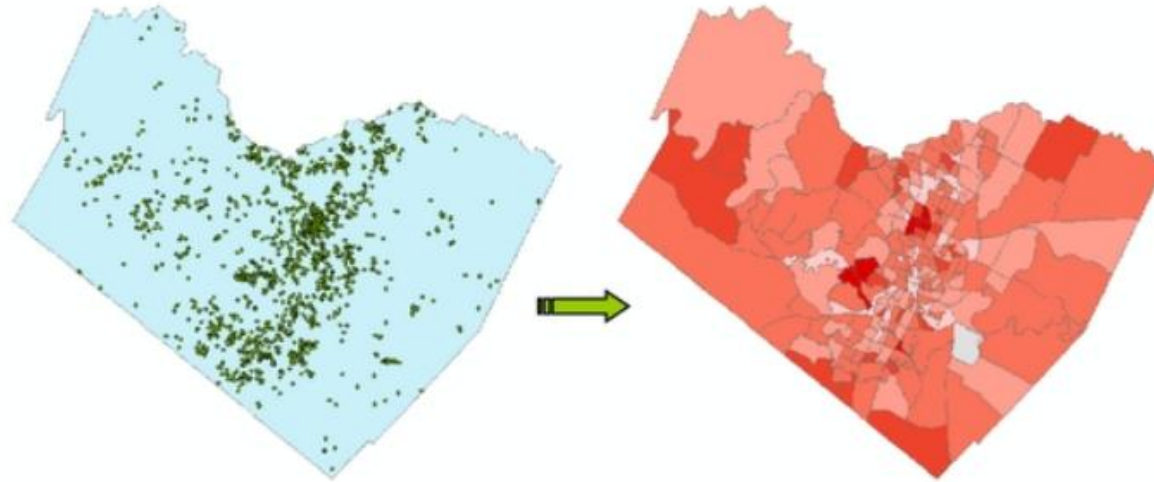
1) Don't make a map!

- Is it essential?
- Use spatial covariates for analysis instead
 - ie, show regression results only

Options for maintaining anonymity

2) Don't show individual-level data

- Aggregate to polygons (CAUs, SA2s, TAs, DHBs)
- Often a requirement of data providers



Locations of cases

Number of cases per 10,000 residents

Options for maintaining anonymity

3) Jitter point data for display (geomasking)

- Maintains general pattern without actual location display
- Don't use jittered locations for spatial analyses!



Jittering

- Lots of different options (refer to course book)
 - Paul A. Zandbergen, "Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data," *Advances in Medicine*, vol. 2014, Article ID 567049, 14 pages, 2014. doi:10.1155/2014/567049
- Complex algorithms
- More important if then using points for analysis or sharing
- Simple methods available
 - Should not use new locations for spatial analysis

Options for maintaining anonymity

4) Do not use basemaps that show road networks or other identifying features – only spatial pattern of data

Geospatial health research

- Has made contributions to our understandings of:
 - Environmental pollutants (setting policies)
 - The importance of neighbourhoods and equitable opportunities for health (for different pops)
 - Most important areas to target for interventions
 - Inequalities and the spatial drivers of inequalities
- DAY 2: More advanced techniques...

Free opportunities to continue learning

- <https://www.gislounge.com/learn-gis-for-free/>
- Website for more tutorials (Stata mapping)

Questions and discussion

More time to work on tutorials