InterConnect

Global data for diabetes and obesity research

Prototype geospatial analysis in DataSHIELD

Tom Bishop,
Technical Lead, MRC Epidemiology Unit, University of Cambridge, UK

This project is funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 602068.

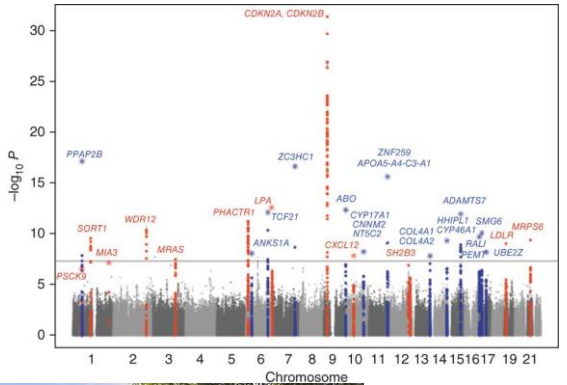
MRC Epidemiology Unit researches causes and prevention of diabetes and obesity

MRC | Epidemiology Unit

UNIVERSITY OF CAMBRIDGE

Prevention

Causes



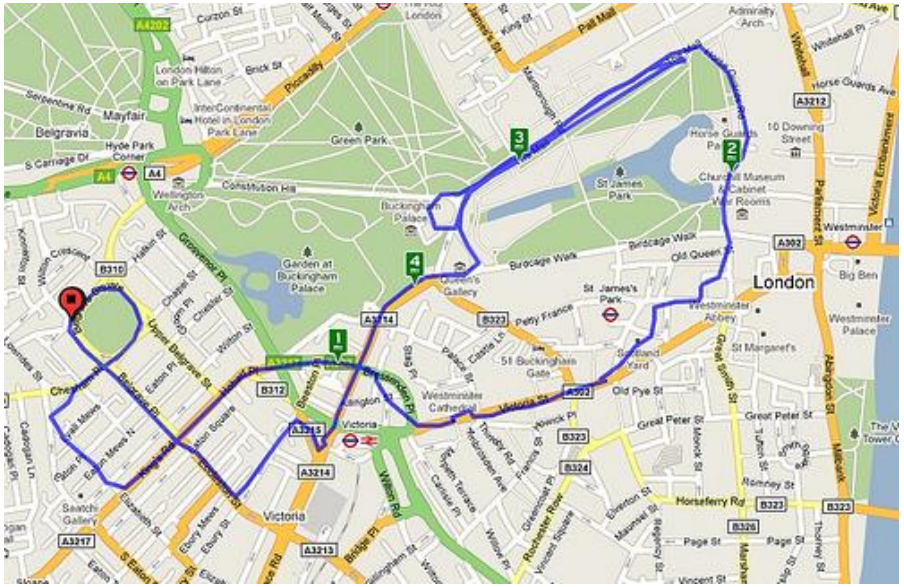
What is InterConnect and how does it use DataSHIELD?

InterConnect – analyse between-population differences in diabetes & obesity across many studies without sharing data

DataSHIELD takes the analysis to the data:

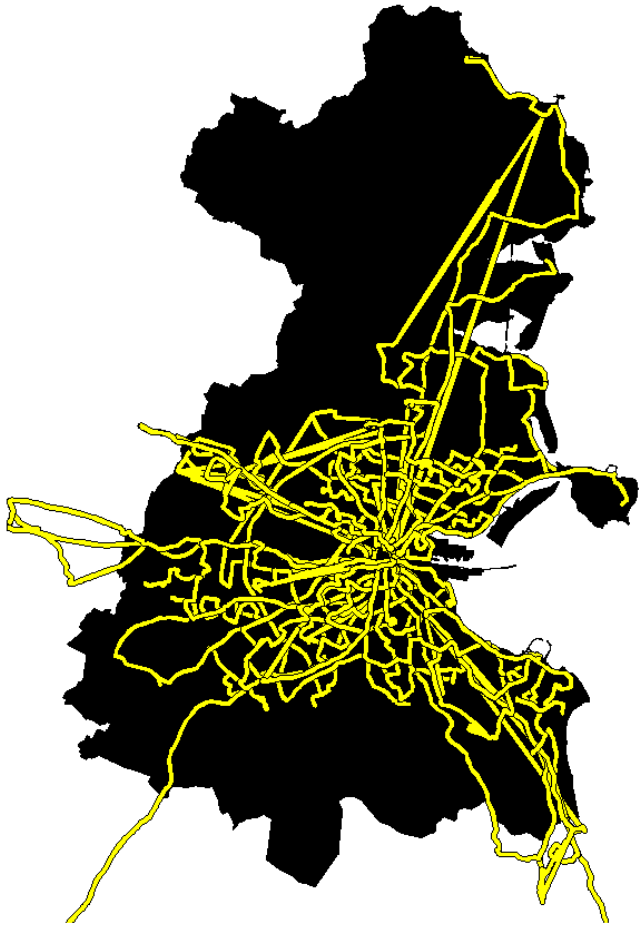
- Retain control of data
- No data sharing legalities
- Reduced work for collaborators

The environments that people encounter may affect risk of diabetes or obesity



- This information is very sensitive, so you might not be able to share it with other researchers
- Ideal use case for DataSHIELD although new geospatial functions needed

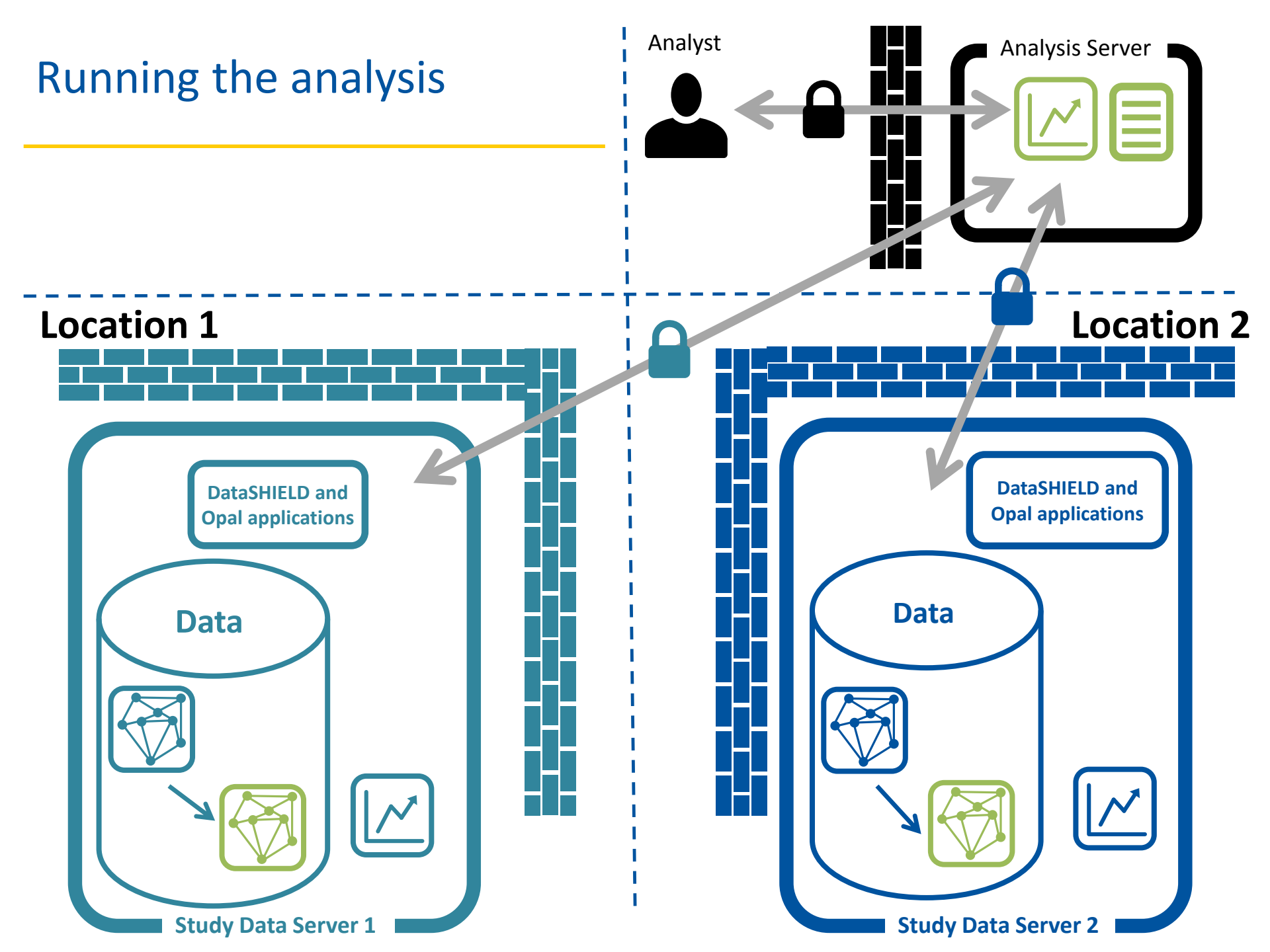
Does the number of fast food outlets passed on an individual's commute affect BMI?



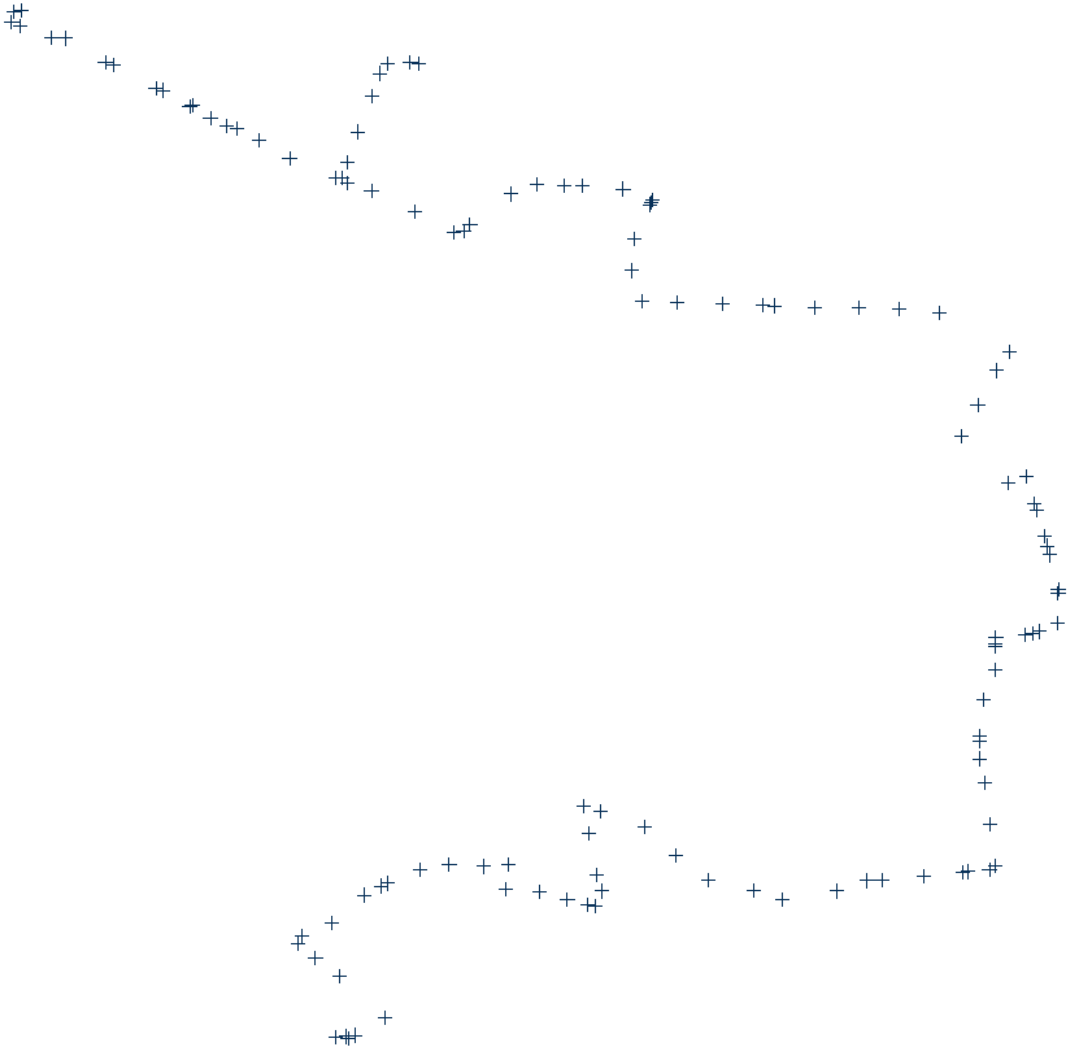
Proof of concept:

1. Use a sample data set from Dublin
2. Split the data set in half to simulate 2 different studies from 2 separate research groups
3. Write new DataSHIELD functions that wrap & reuse existing rgeos and rgdal geospatial functions
4. Run the analysis with DataSHIELD AND with the data pooled to compare results

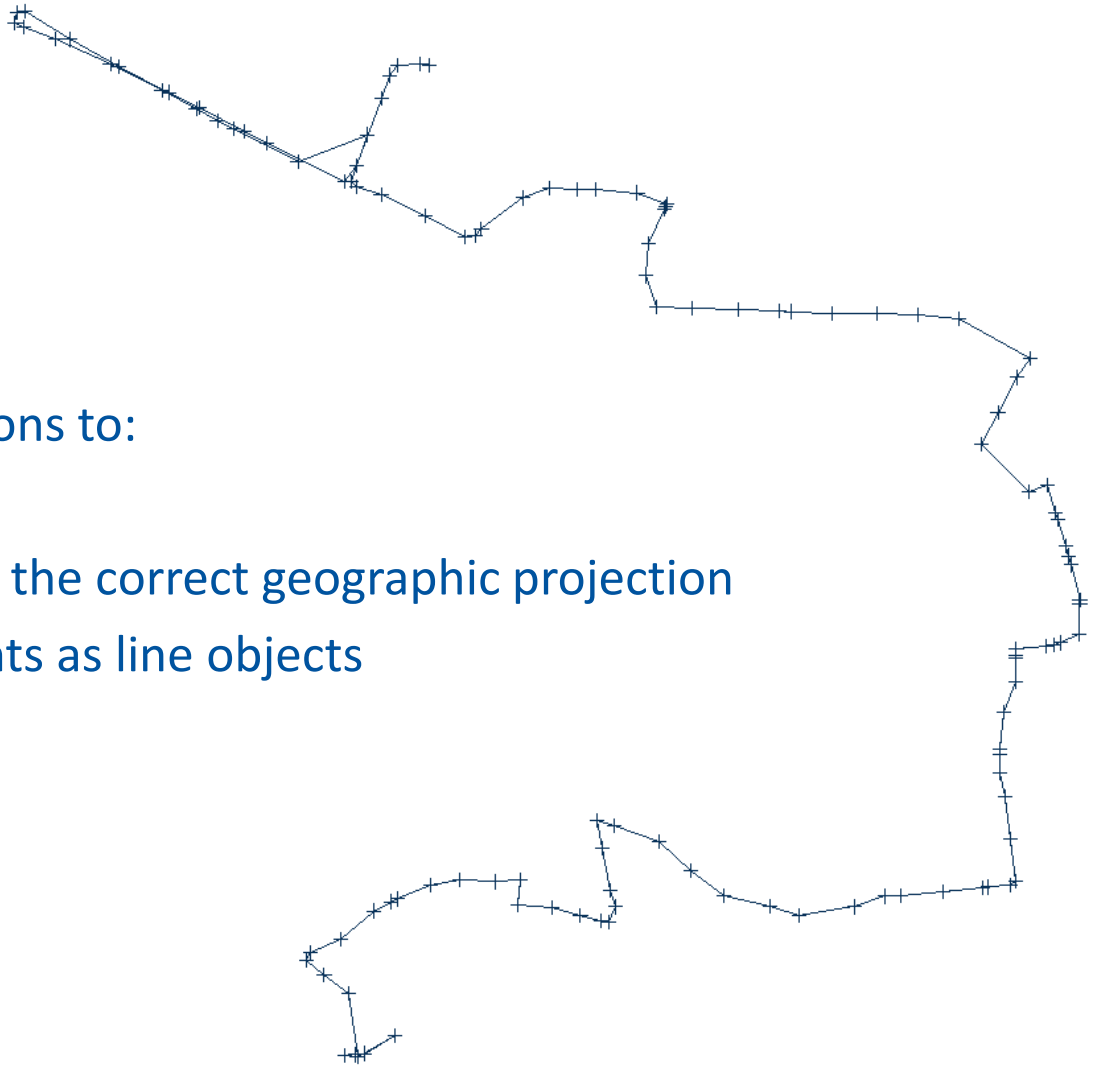
Running the analysis



Analysis: raw GPS points for 1 individual



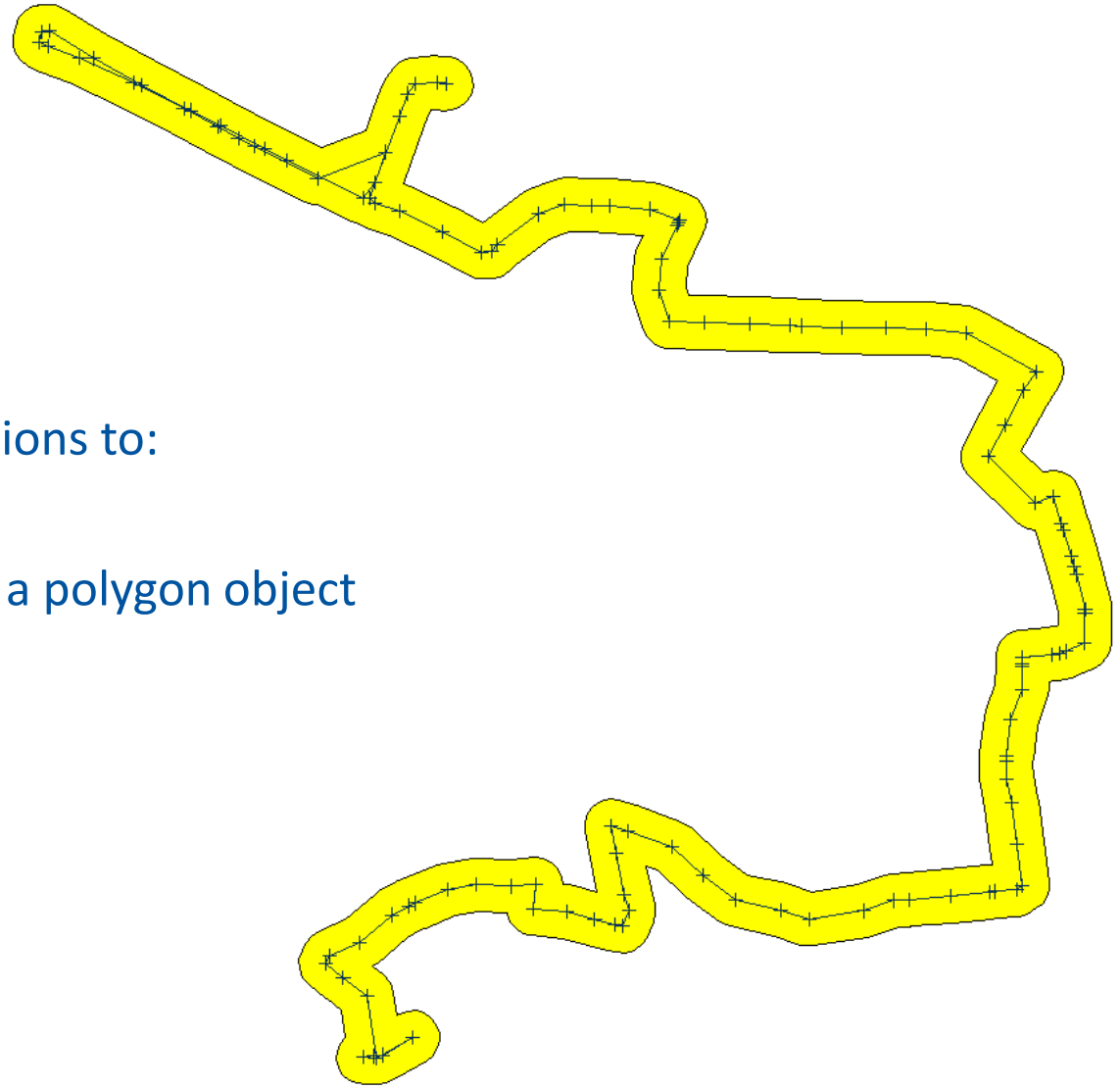
Analysis: define journey as a line



New DataSHIELD functions to:

1. Convert the points to the correct geographic projection
2. Define groups of points as line objects

Analysis: draw a 25m buffer



New DataSHIELD functions to:

1. Define the buffer as a polygon object

Analysis: plot fast food outlets & count how many fall in the buffer

New DataSHIELD functions to:

1. Plot the location of the fast food outlets
2. Count the points falling in the buffer to give exposure

And finally… regress the exposure to fast food outlets against BMI, using existing DataSHIELD functionality

Results and conclusions

- Results were the same for pooled analysis and DataSHIELD analysis where data was distributed and not accessed directly
- Existing rgeos, rgdal functions can be wrapped and used within DataSHIELD
- DataSHIELD is well set up for adding new functions

Questions