Python for Open Data Lovers: Explore It, Analyze It, Map It

Dana Bauer

spatial analyst & mapmaker / Azavea

@geography76

Childcare capacity

500 200 100 50 10

Percent of persons living below the poverty level

- Less than 10%
- 10% to 25%
- 25% to 50%
- 50% to 75%
- 75% to 100%
- Insufficient data

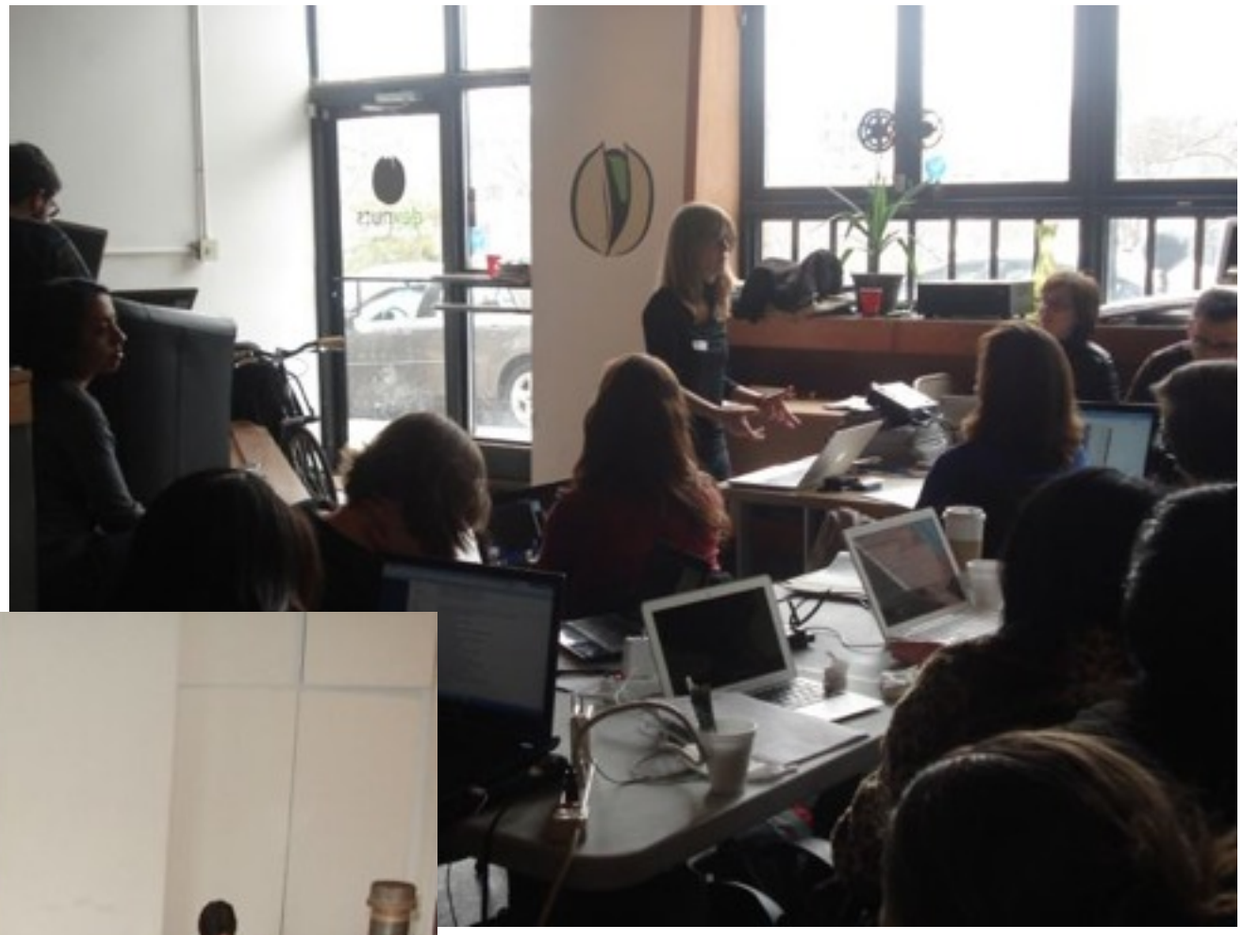
19151

19139

19104

19143

19142



python

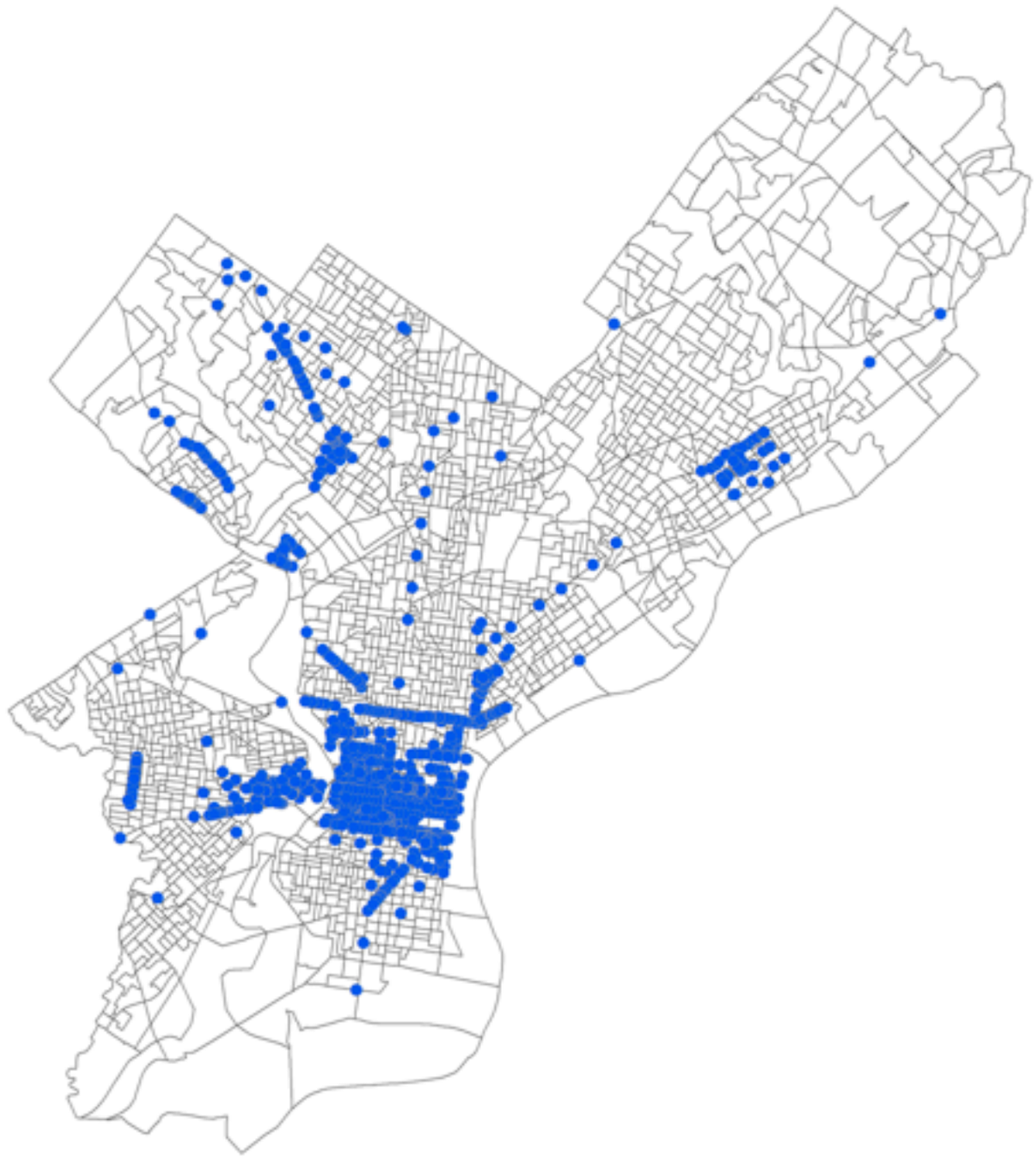
Where are we going?

- Data, data, data
- Spatial is special
- Finding patterns

- Data.gov
- OpenDataPhilly
- DC Data Catalog
- DataSF
- Chicago Data Portal
- NYC Open Data
- London Datastore

assembly member expenses
bicycle lanes
city purchase orders
dialysis centers
elevation data
filming locations
Google Transit Feed Specification (GTFS)
historical photos
influenza rates
judicial districts
Key Stage 2 test results by free school meal eligibility
land cover
monthly calls to Human Services Agency switchboard operators
neighborhood health clinics
Oyster ticket stop locations
political districts
quality of life indicators
restaurant inspections
sewer lines
traffic counts
utility excavation and paving five-year plan
violent crime incidents
ward offices
youth centers
zoning

****real-time parking availability and pricing****



swiss army knife, part I

- csvkit: http://csvkit.readthedocs.org/
- a set of Python utilities for working with CSV
- meant to replace csv module

```
$ csvcut -n variablesR4.csv
  1: MEDINC09
  2: PERYOUTH
  3: PERBLACK
  4: PERASIAN
  5: POPDEN
  6: PERHIS
  7: PERENRLSCH
  8: PERHSUP
  9: HOMEINTRT
 10: ANYINTRT
...
```

```
$ csvcut -c 11,46 variablesR4.csv |
csvstat
  1. PERHSUP
   <type 'float'>
   Nulls: No
   Min: 27.272727
   Max: 100.0
   Sum: 28605.585196
   Mean: 78.5867725165
   Median: 80.522477
2. PERPOVU200
   <type 'float'>
   Nulls: No
   Min: 0.0
   Max: 100.0
   Sum: 16202.698932
   Mean: 44.5129091538
   Median: 44.217176
```

```
$ csvcut -c 46,61 variablesR4.csv | csvsort -r -l | csvlook
-------------------------------------
|  line_number | PERPOVU200 | TRACT2  |
-------------------------------------
|  1           | 100.0      | 32800   |
|  2           | 100.0      | 12400   |
|  3           | 95.073891  | 32700   |
|  4           | 94.480519  | 36400   |
|  5           | 89.816671  | 17500   |
|  6           | 88.47768   | 17601   |
|  7           | 87.362018  | 17602   |
|  8           | 86.896762  | 8800    |
|  9           | 86.876355  | 16600   |
|  10          | 85.426081  | 15200   |
...

$ createdb OTIsocio
$ csvsql --db postgresql:///OTIsocio --table fy09 --insert
variablesR4.csv
```

swiss army knife, part 2

- GDAL/OGR: http://pypi.python.org/pypi/GDAL
- GDAL/OGR: Easy to read and write spatial data formats, change projections, translate between format, combine data
- shapely, descartes-matplotlib

Spatial is special

- spatial data = attributes, location, time
- mappable!
- spatial data must be referenced in space
- Tobler's First Law of Geography

Spatial analysis

- large data sets → a smaller amount of meaningful information
- exploratory (ESDA)
- spatial statistics
- mathematical modeling and prediction of spatial processes

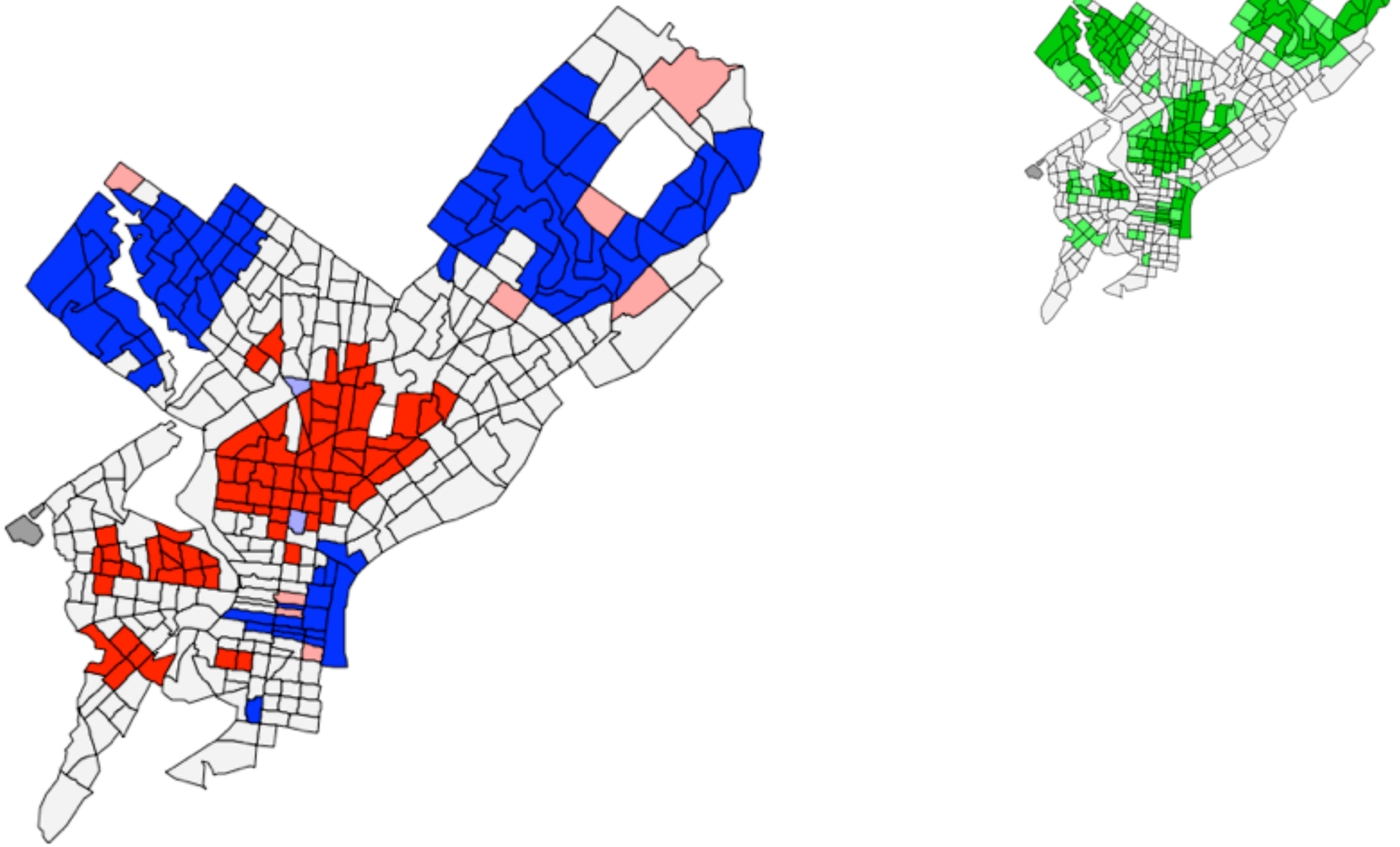
Techniques

- point pattern analysis -- hot spots, k density, nearest neighbor
- spatial interpolation -- kriging
- spatial regression -- ordinary least squares, geographically weighted regression

PySAL

- developers looking for spatial analytical methods to incorporate in application development
- analysts working on projects that require custom scripting
- looking for a user-friendly GUI? Try STARS, GeoDA, GeoDASpace.
- want to integrate into a powerful GIS? Look for plug-ins for ArcGIS & QGIS.


MST
Voronoi
Computational Geometry
Hulls
GWT
GAL
MAT
Weights
Estimation
Testing
Spatial Econometrics
Diagnostics
Simulation
ARISEL
Clustering
AZP
max-p
PySAL
ESDA
Smoothing
LISA
Spatial Dynamics
Spatial θ
Spatial τ
Spatial Markov



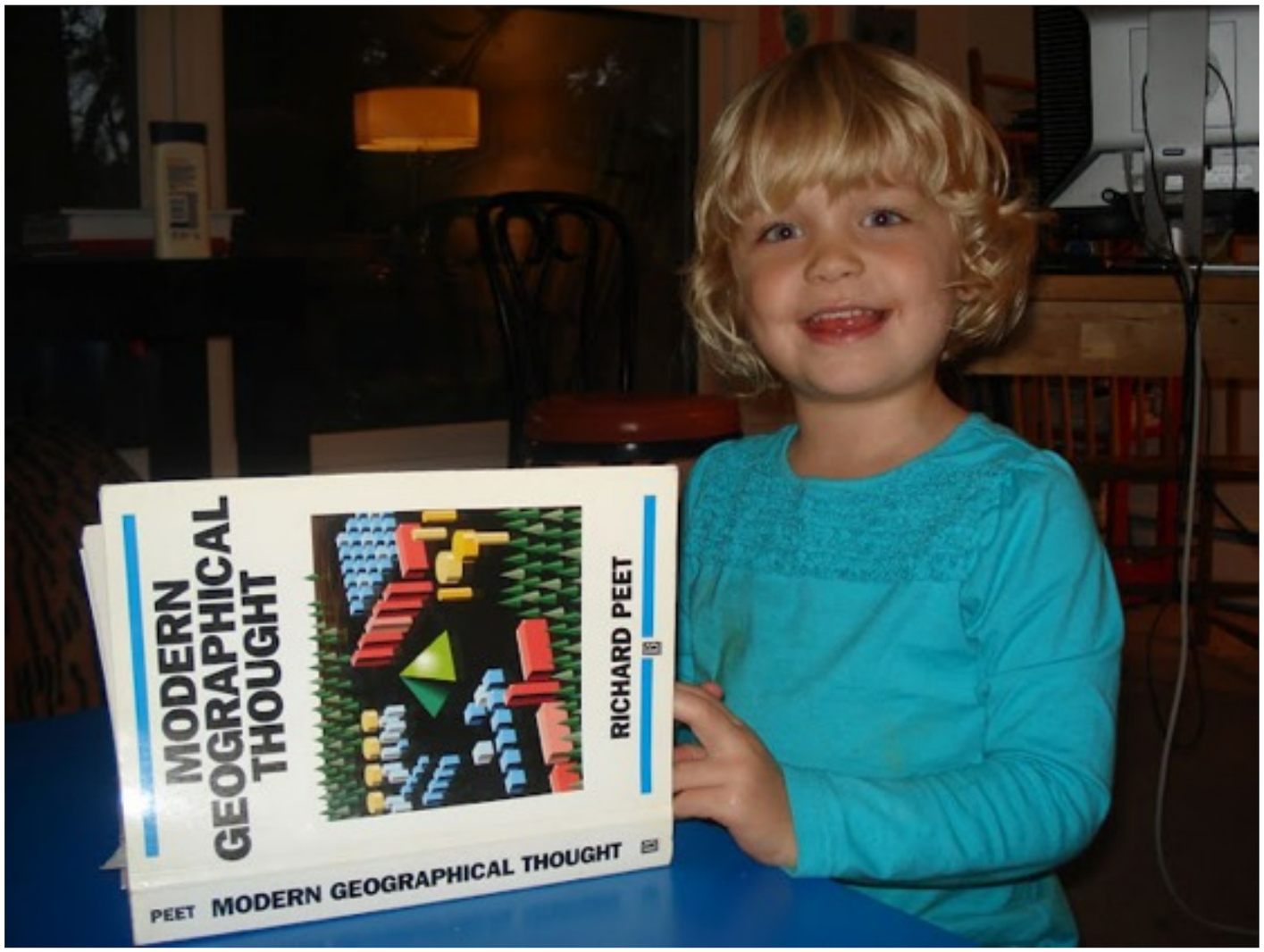
Want to learn more?

The SAGE Handbook of Spatial Analysis
eds. A. Stewart Fotheringham and Peter A. Rogerson

Interactive Spatial Data Analysis
Trevor Bailey and Tony Gatrell

Geographic Information Analysis
David O'Sullivan and David Unwin

PySAL
Luc Anselin, GeoDA Center
Arizona State University



Mia, age 3, geographer in training