DATA2001 – Data Science, Big Data and Data Variety
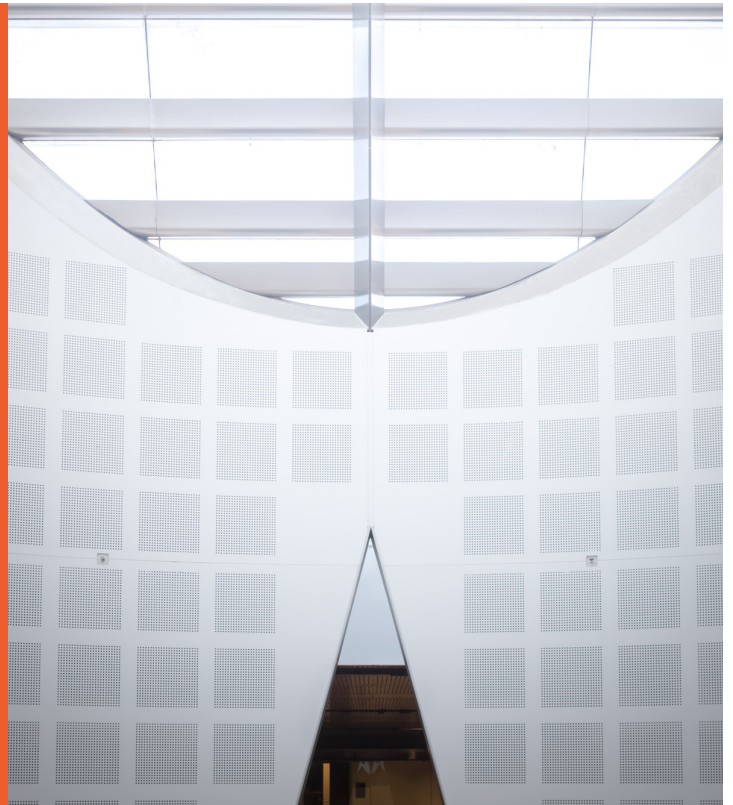# Week 12: Assignment Tips

Presented by

A/Prof Uwe Roehm

School of Information Technologies

THE UNIVERSITY OF SYDNEY

Practical Assignment: Walkability Analysis

– Assignment published end of last week (Canvas: Modules –> Assignment)

    – **Worth 20% pf the final grade in DATA2001**

    – **Due in tutorial of Week 13**

[https://www.walkscore.com/AU-NSW/Sydney]

Practical Assignment: Walkability Analysis

- Goal: Practical experience with data variety, data analysis, and presentation
    - Technologies as covered in this course: Python, Jupyter notebooks, and SQL
- Three tasks:
    - Data import and integration
        - We provided census data and some transport data (carsharing pods)
        - Needs to be combined, eg. via spatial join
        - Feel free to extend with own datasets
    - Walkability Analysis
        - Computation of walkability score; example formula given
        - When adding other datasets, feel free to adjust formula
        - Correlation of your score with some ABS statistics
    - Documentation and (brief) Report

Provided Datasets (cf. Canvas)

- ABS Data
    - Census data on *neighbourhoods* (SA2-level areas) in Greater Sydney such as population, landarea, number of dwellings
    - Business statistics per SA2-area
    - Income and rent statistics to check for correlation with
- Carsharing "Pods"
    - One example of transport data: names and locations of dedicated carsharing parking bays ('pods').
- Note that SA2-level data from the ABS does not always match suburbs, and that the carsharing pods have a GPS location, but not the neighbourhoods
    - cf. tutorial this week on how to retrieve boundary data for neighbourhoods too
- Adding more datasets from your side is explicitly encouraged.
    - Try different types and forms, not just CSV…

Assignment Rules

- Groupwork
  - teams of 2 (unless odd-size class or other good reasons)
  - All team members should be in the same tutorial
- Deliverables see handout, page 3
- Due on Tuesday of Week 13
  - Submission page and marking rubric have been published in Canvas
  - Only one member per team needs to submit for the whole group; she should submit both a ZIP archive under "Walkability Analysis Assignment" <u>and</u> also the PDF of your report in the separate "TurnItIn Dropbox - Walkability Analysis"
  - Late submissions: -20% of achieved mark per day late
- Demo in Week 13
  - There will be a short demo during the tutorials to tutors
  - Individual grades can be scaled based on participation in project or demo

News

- We have PostGIS installed on soit-db-pro-1.ucc.usyd.edu.au
  - Spatial extension for PostgreSQL supporting geographic objects (OGC)
  - Same logins (y18s1d2001_*abcd1234*) than on the existing server, but you need to re-create your schema there to use it.
  - http://postgis.net

- We have pyshp installed on all Jupyter servers
  - Parser for ESRI shapefiles (as can be downloaded from ABS)
  - https://pypi.org/project/pyshp/

PostGIS

[http://postgis.net/documentation/]

- – Spatial database extension for PostgreSQL supporting geographic objects (OGC)
  - – **Geometry types** for Points, LineStrings, Polygons, MultiPoints, etc.
    - • including import/export from standard formats such as GeoJSON or KML
  - – Support for **spatial reference systems** and transformations between
  - – **Spatial predicates** on geometries using the 3x3 nine-intersection model
  - – Spatial operators for determining **geospatial measurements** like area, distance, length and perimeter, and **geospatial set operations**, like union, difference etc.
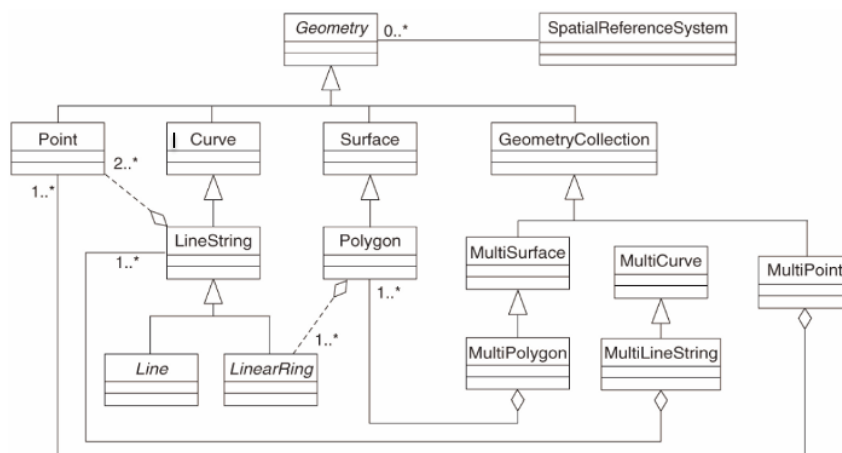 - – **R-Tree indexing** (over GiST)
- – Example:

```
INSERT INTO superhero VALUES ('Catwoman', ST_SetSRID(ST_MakePoint(41.87,-87.634), 4326);
SELECT superhero.name
  FROM city, superhero
 WHERE ST_Contains(city.geom, superhero.location)
   AND city.name = 'Gotham';
```

OpenGIS Consortium (OGC) Data Model

[Source: OGC Simple Features, 2016]

WGS84 versus Australian GDA94

- – WGS84 is used by the GPS system
- – The official geodetic datum (coordinate system) for Australia is GDA94 ("Geocentric Datum of Australia")
  - – Based on IERS Terrestrial Reference Frame (ITRF), but fixed to a number of reference points in Australia.
- – Difference between WGS84 and GDA94:
  - – "The spheroids used for WGS84 and GDA84 are also almost identical, and both systems are geocentric. Thus for most mapping, exploration and GIS uses, WGS84 and GDA94 coordinates will be the same. [...] For precise surveys, however, the difference between WGS84 and GDA94 may be significant, and changes slowly over time. [...] The difference between GDA94 and WGS84 is approximately 45cms in 2000."

    [http://www.geoproject.com.au/gda.faq.html]