EXPLORING DATA WRANGLING PROTOCOLS IN MAPPING APPLICATIONS

Greice C. Mariano, Sara Diamond

iCity

Urban Informatics for Sustainable Metropolitan Growth

1. Urban Informatics for transportation operations, planning & decision-making

2. Urban Mobility & Integrated urban systems design

3. Visualization & Decision-support

iCity

Urban Informatics for Sustainable Metropolitan Growth

1. Urban Informatics for transportation operations, planning & decision-making
   - Ontology
   - Software Platform
   - Data Collection

2. Urban Mobility & Integrated urban systems design
   - Complete Streets
   - Parking Management
   - Transit Management
   - Urban Modelling
   - Economic Analysis

3. Visualization & Decision-support

iCity

Urban Informatics for Sustainable Metropolitan Growth

1. Urban Informatics for transportation operations, planning & decision-making

- Ontology
- Software Platform
- Data Collection

3. Visualization & Decision-support

2. Urban Mobility & Integrated urban systems design

- Complete Streets
- Parking Management
- Transit Management
- Urban Modelling
- Economic Analysis

iCity

Urban Informatics for Sustainable Metropolitan Growth

1. Urban Informatics for transportation operations, planning & decision-making
   - Ontology
   - Software Platform
   - Data Collection

2. Urban Mobility & Integrated urban systems design
   - Complete Streets
   - Parking Management
   - Transit Management
   - Urban Modelling
   - Economic Analysis

3. Visualization & Decision-support
   - Data Visualizations
   - Integrative Website

iCity

Urban Informatics for Sustainable Metropolitan Growth

1. Urban Informatics for transportation operations, planning & decision-making
   - Ontology
   - Software Platform
   - Data Collection

2. Urban Mobility & Integrated urban systems design
   - Complete Streets
   - Parking Management
   - Transit Management
   - Urban Modelling
   - Economic Analysis

3. Visualization & Decision-support
   - Data Visualizations
   - Integrative Website

Research Approach

Reference Model for Visualization

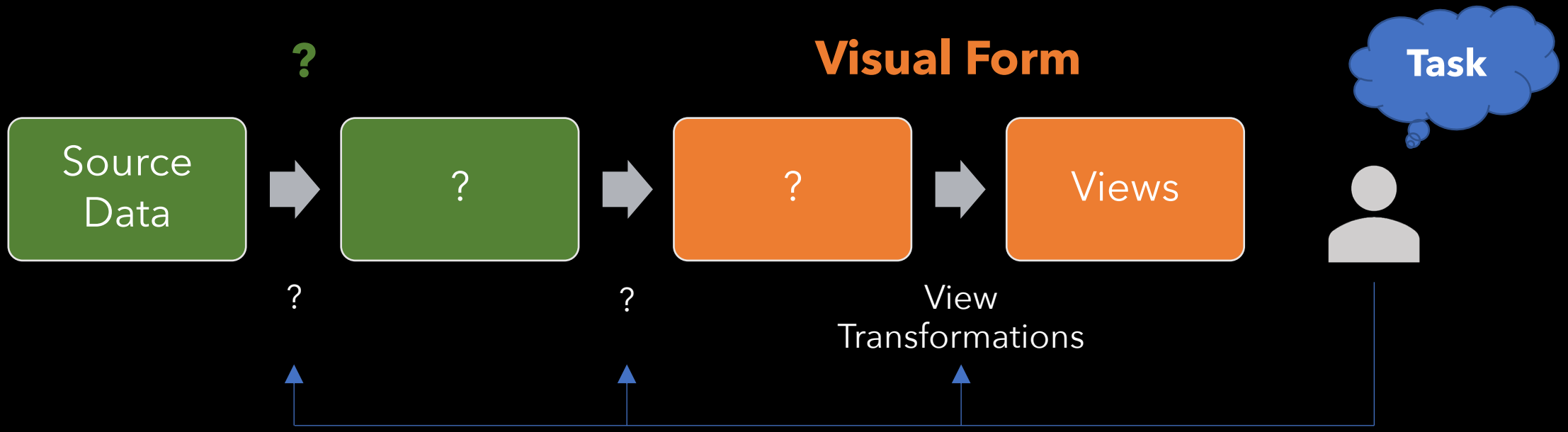
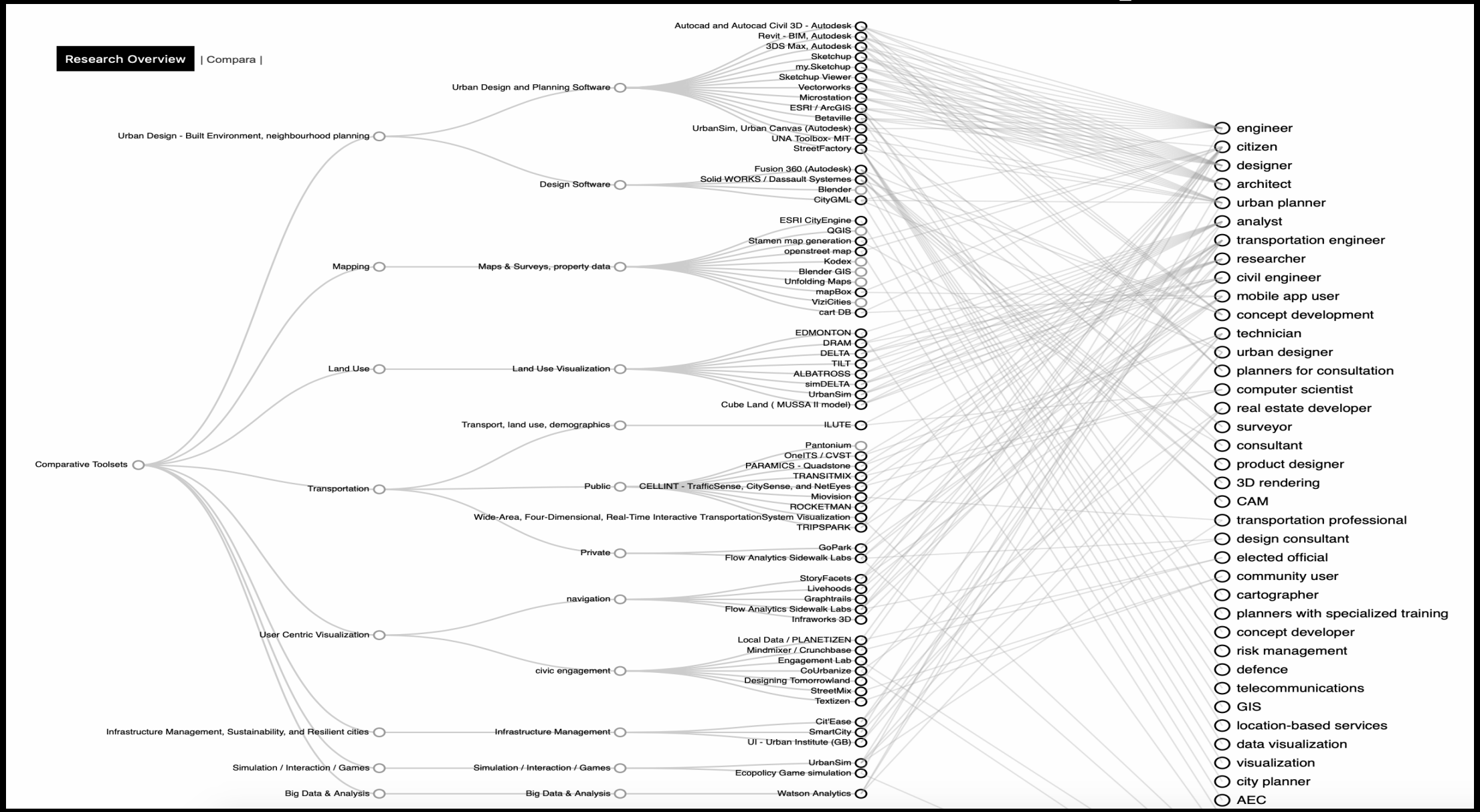


Image adapted from: Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman, eds. Readings in information visualization: using vision to think.

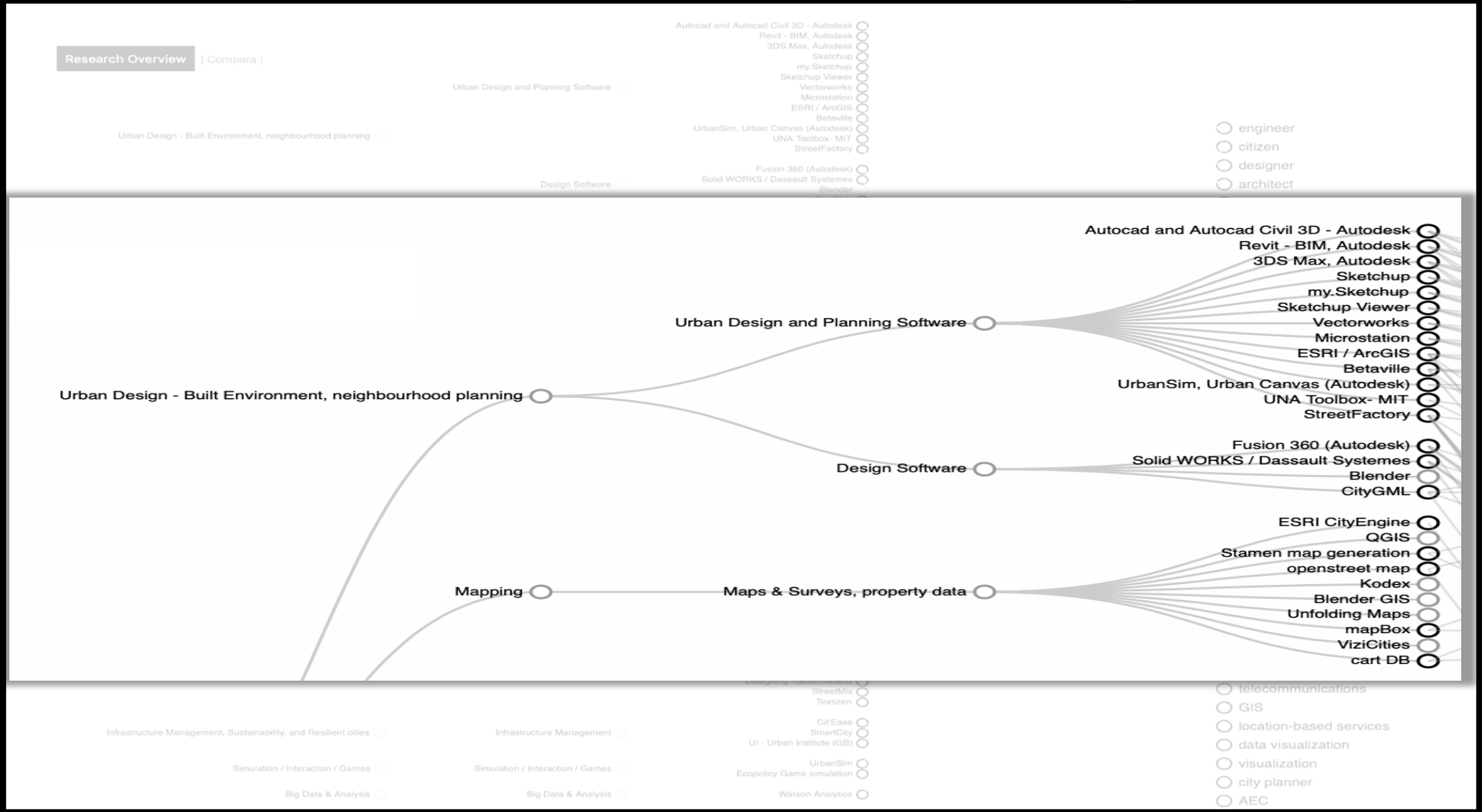
Geographic Information System (GIS)

- Is a computer system for capturing, storing, checking, and displaying data related to positions on Earth's surface.
- Can help the individuals and organizations better understand spatial patterns and relationships
- Requires a georeferenced data:
  - Can use any information that includes location, which can be expressed in many different ways, such as latitude and longitude, address, or zip code.

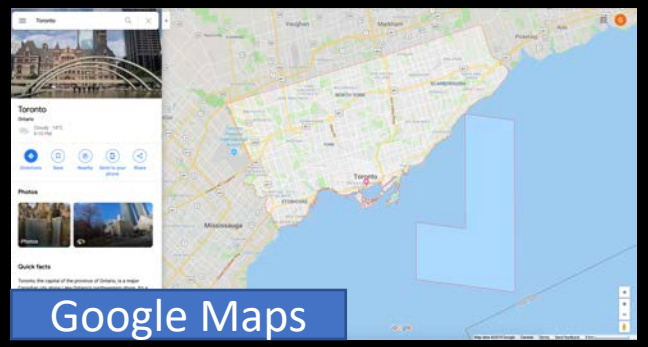
Views for Urban Informatics - Compara

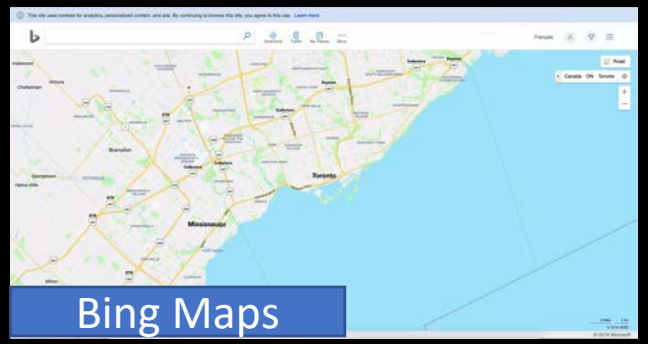


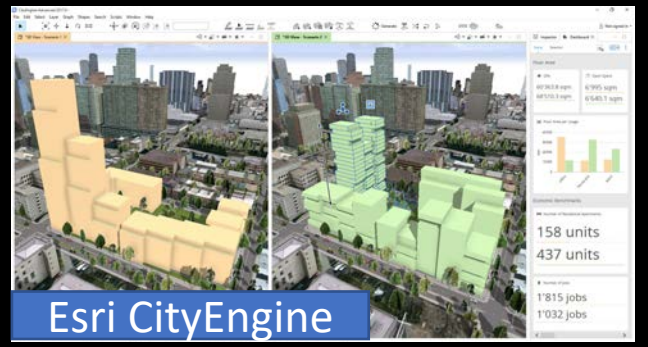
Views for Urban Informatics - Compara

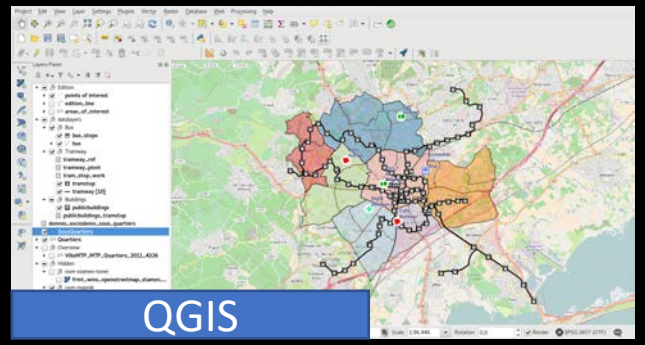


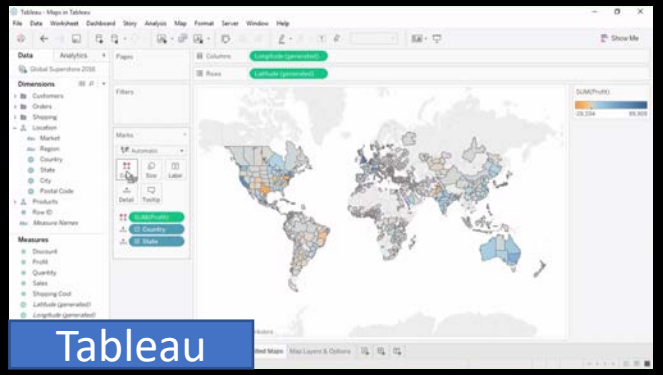
Mapping Approaches

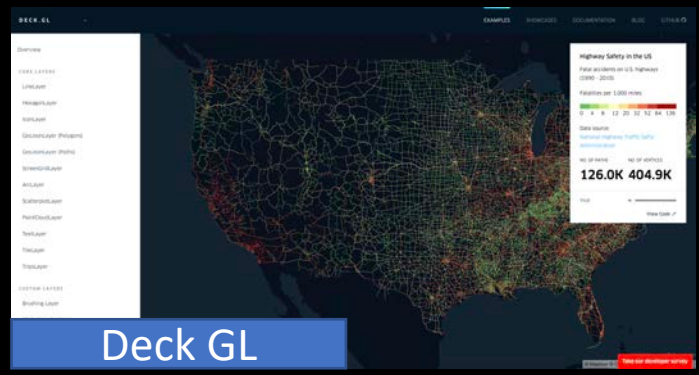




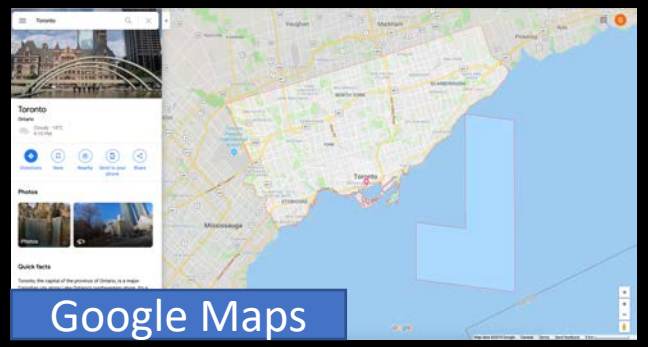


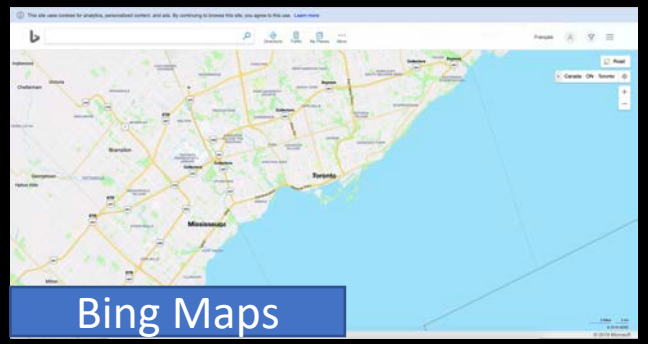


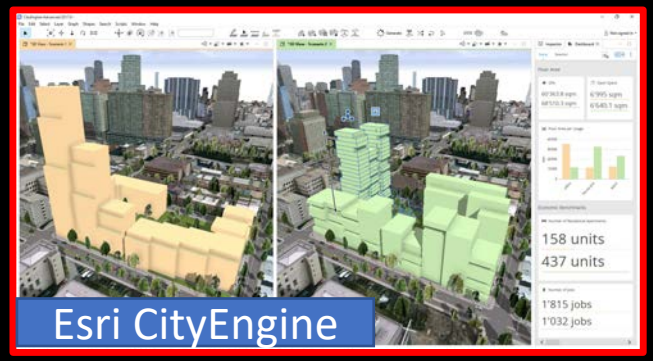


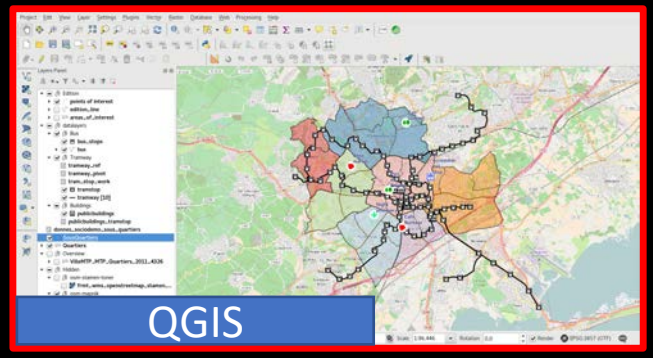


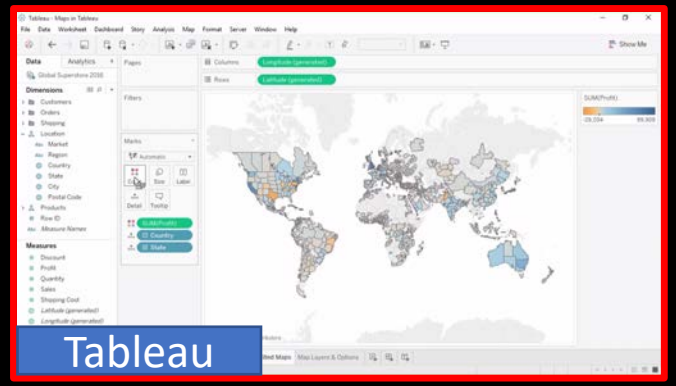
Mapping Approaches

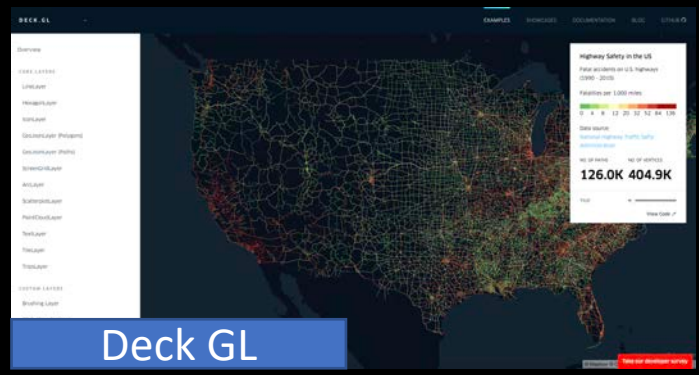




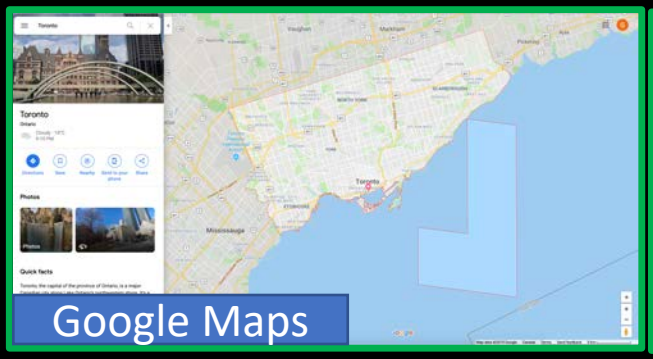


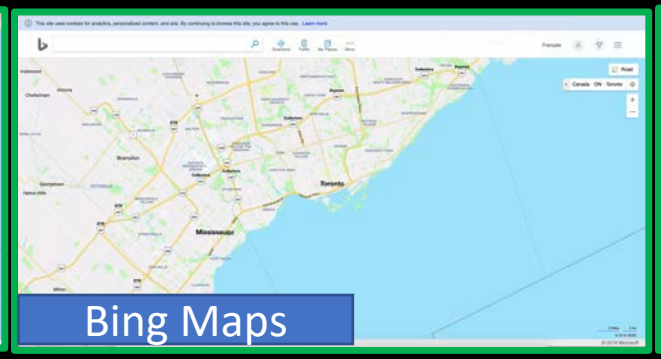


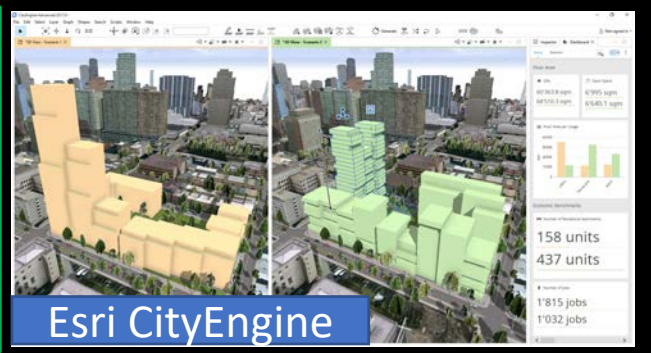


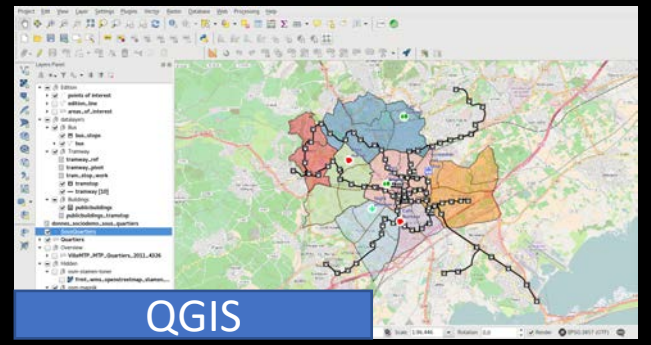


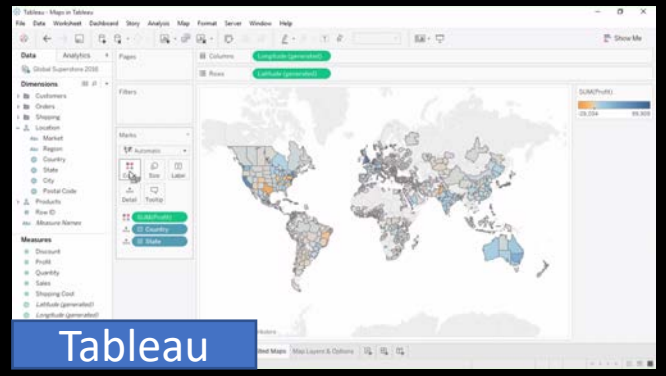
Mapping Approaches

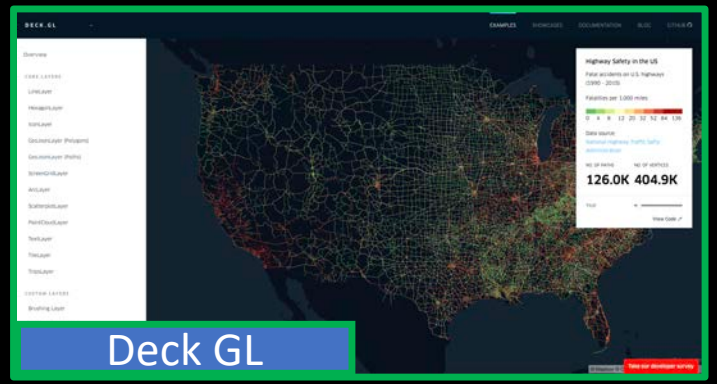












GIS Representation

- Vector:
  - Data is represented using points, lines and polygons

- Raster:
  - Data is represented as a surface modeled by a matrix of values (pixels)
  - Useful for continuous data, such as satellite imagery, aerial photographs, pollution, population, etc.

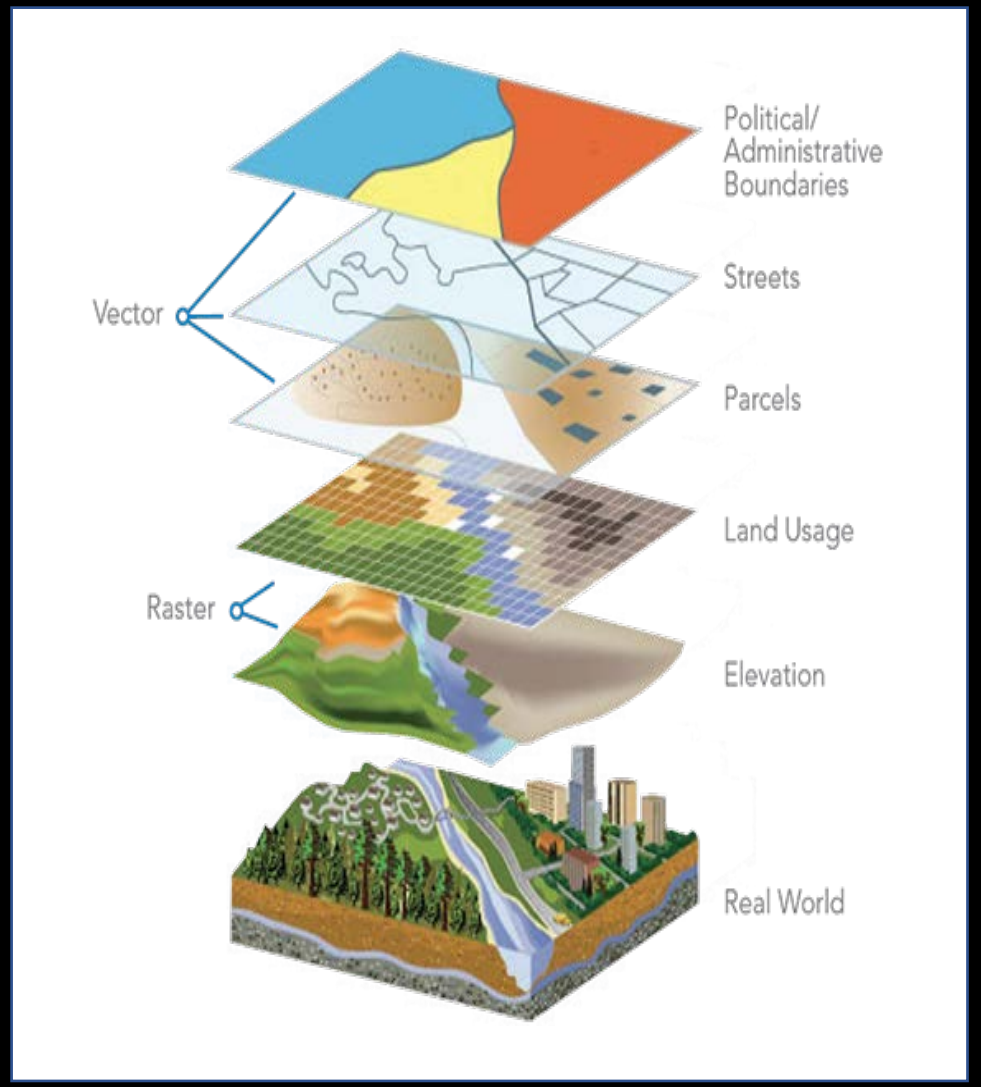


Image source: http://3.bp.blogspot.com/-A140pKBwSXU/VHvp73TcocI/AAAAAAAAAK8/xoP1KQI5L-Y/s1600/Raster%2Band%2BVector%2BData.jpg

GIS File Formats

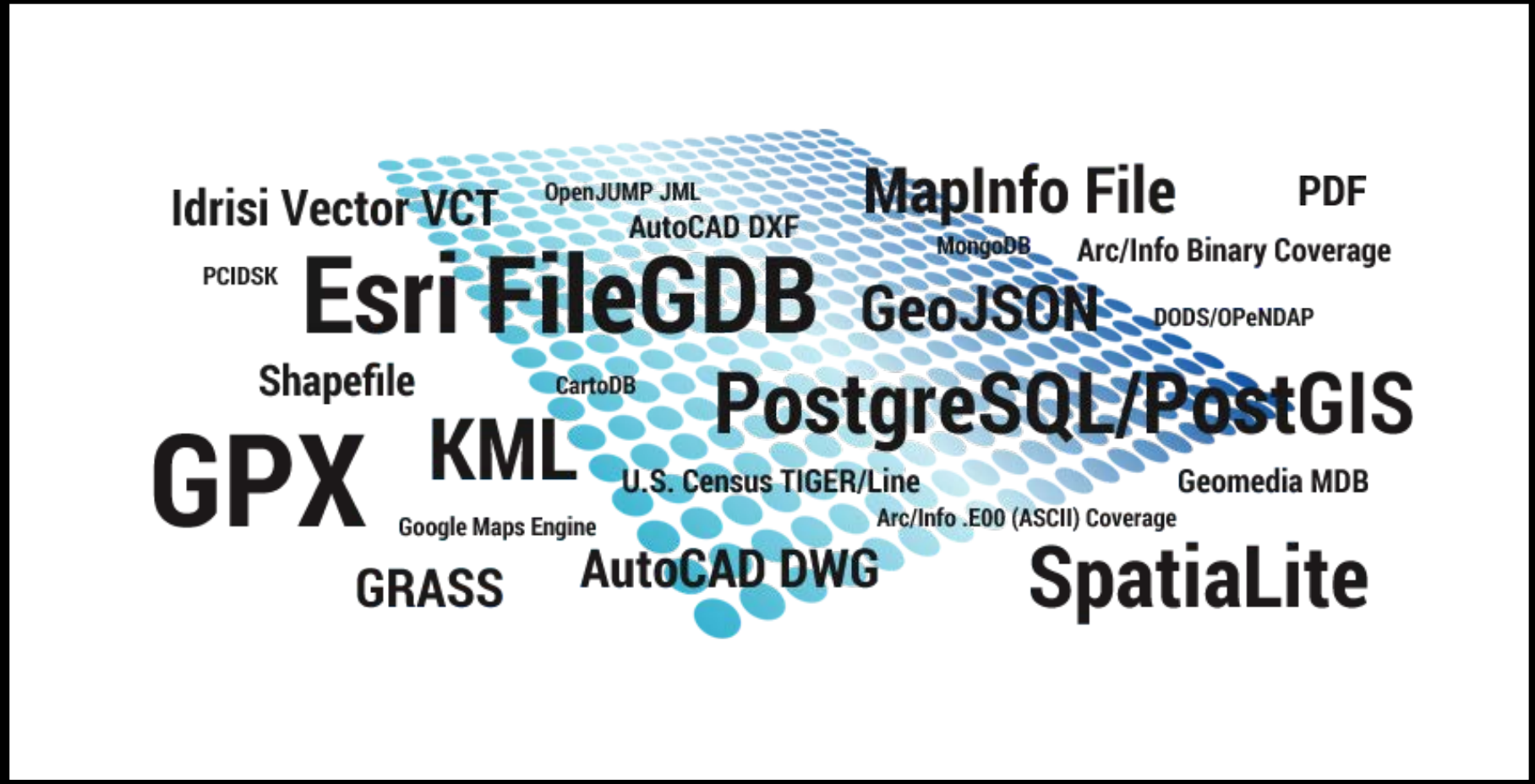


Image source: https://gisgeography.com/qgis-arcgis-differences/

GIS File Formats

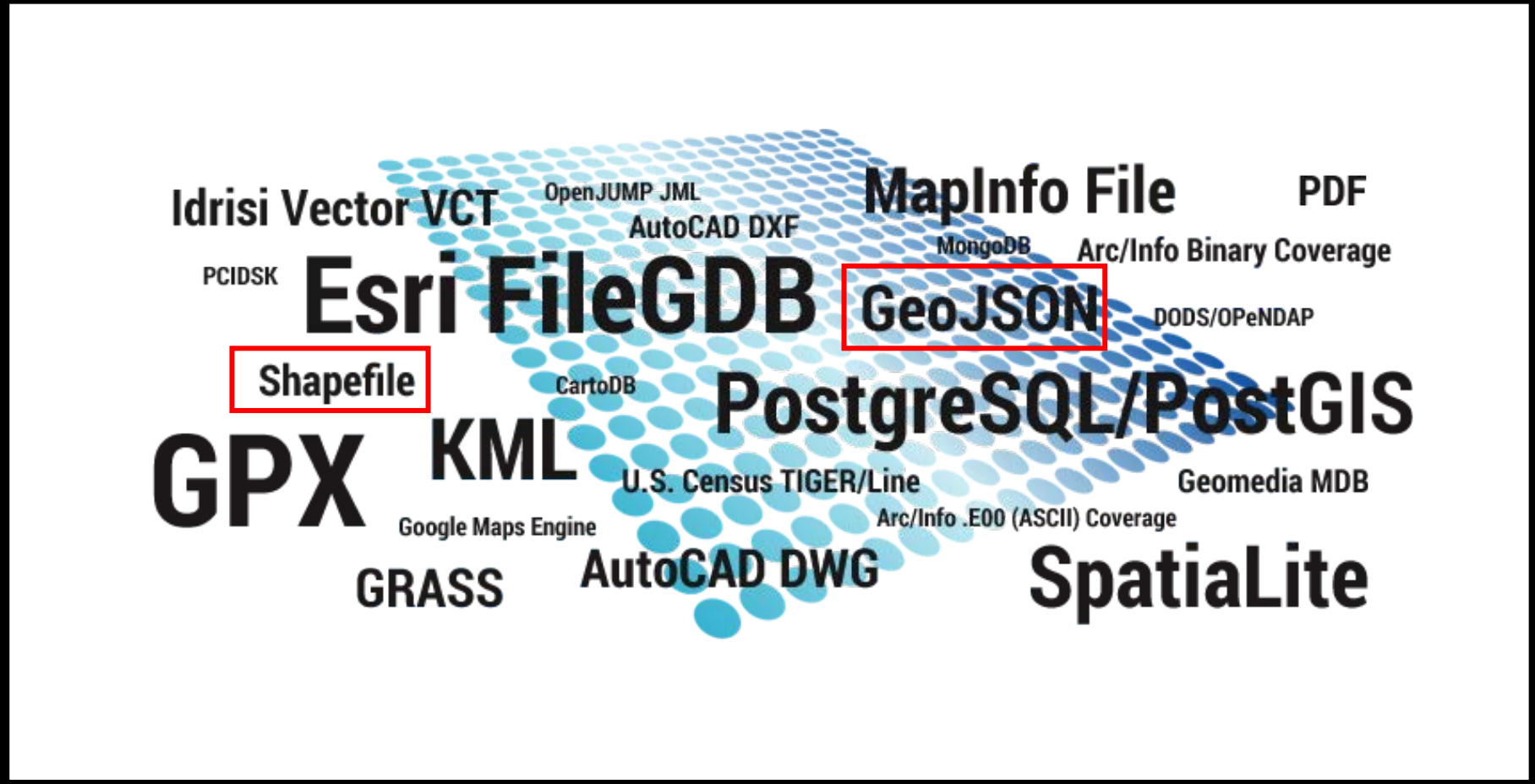
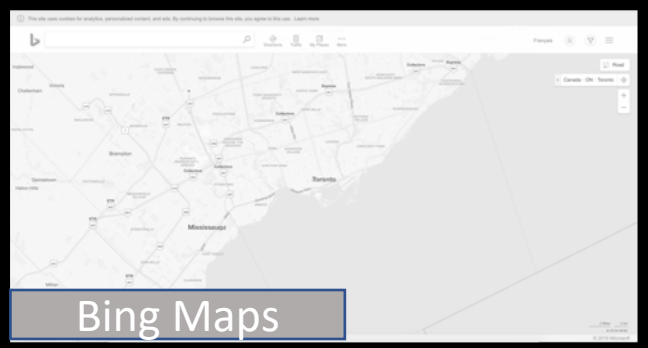
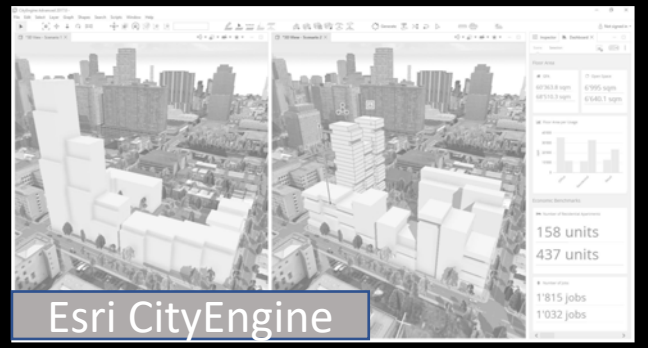
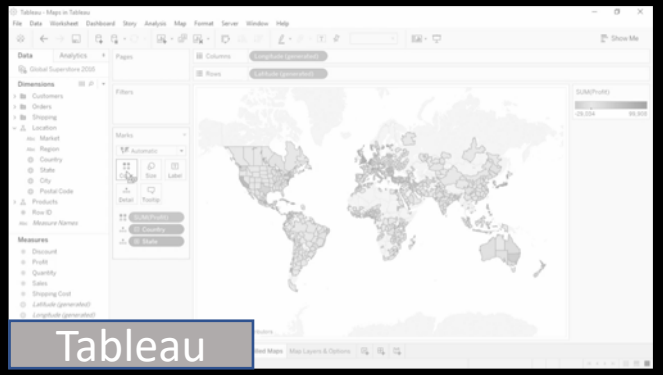


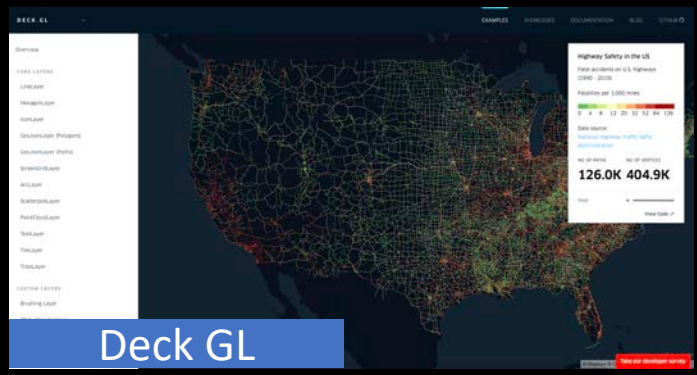
Image source: https://gisgeography.com/qgis-arcgis-differences/

Mapping Approaches









Research Approach

Reference Model for Visualization

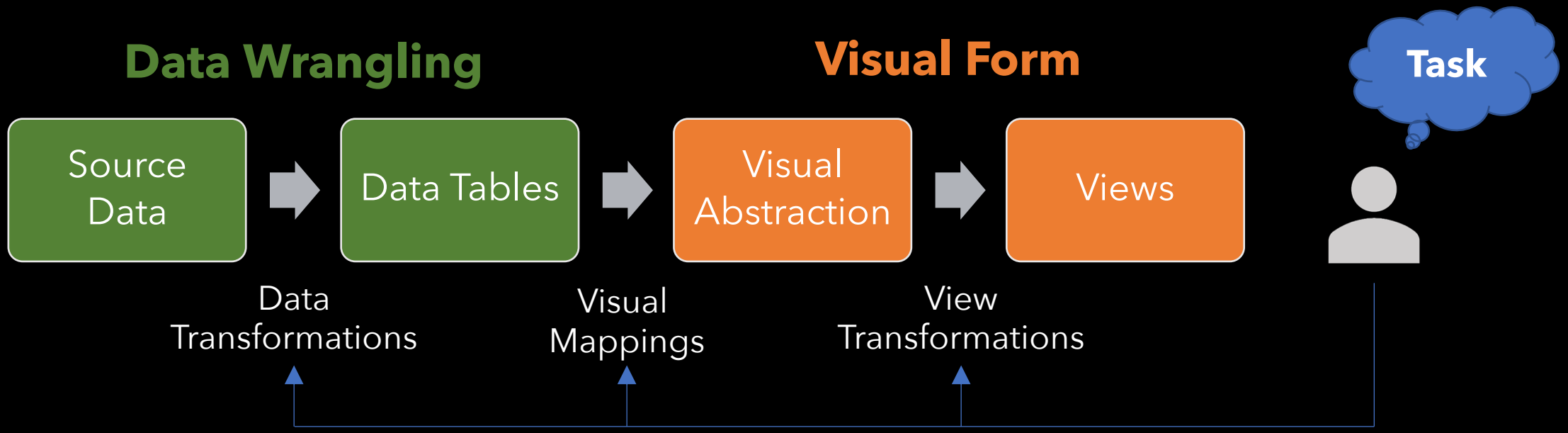


Image adapted from: Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman, eds. Readings in information visualization: using vision to think.

Data Wrangling

"It is the process of ***converting*** and ***mapping data*** from its ***raw*** form to ***another format*** with the purpose of ***making it more valuable*** and ***appropriate*** for advanced tasks in Data Analytics and Machine Learning."

Use Case

iCity Working In Progress

- Transportation Tomorrow Survey (TTS)
  - Data 1: Origin-Destination Matrices
  - File format: set of CSVs files
    - Each file refers to a different variables, such as cost, time wait, volume of people…
  - Data 2: TTC zones Shapefile
- Visualization using Kepler.gl and Deck.gl
  - Steps:
    1. Transformation of shapefile to GeoJson
    2. Data transformation from OD-Matrices to Data Tables
    3. Integration between the different variables in the Data Tables
    4. Mapping the new data tables with the TTC zones.

Transforming Data

Shapefile

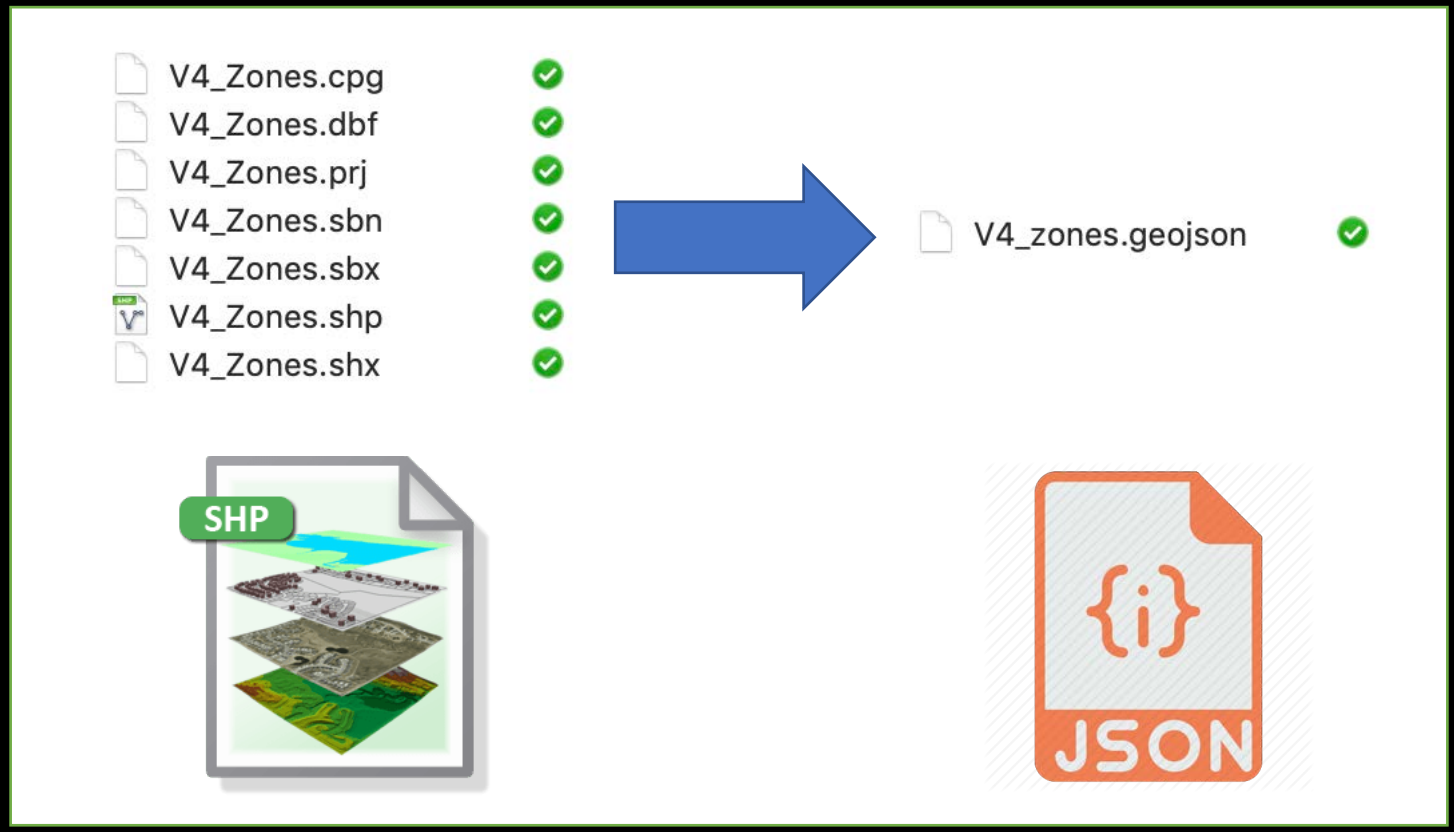
- ESRI standard vector file format
- Enconding features as: point, lines and polygons
- Stored as a set of related files and contains one feature class.

GeoJson

- Open standard vector file format based on JavaScript Object Notation (JSON)
- Encoding features as: point, lineString, polygon, MultiPoint, MultiLineString, MultiPolygon, and GeometryCollection.
- Features contains: geometry object and additional properties, and a FeatureCollection contains a list of Features.

Transforming Data

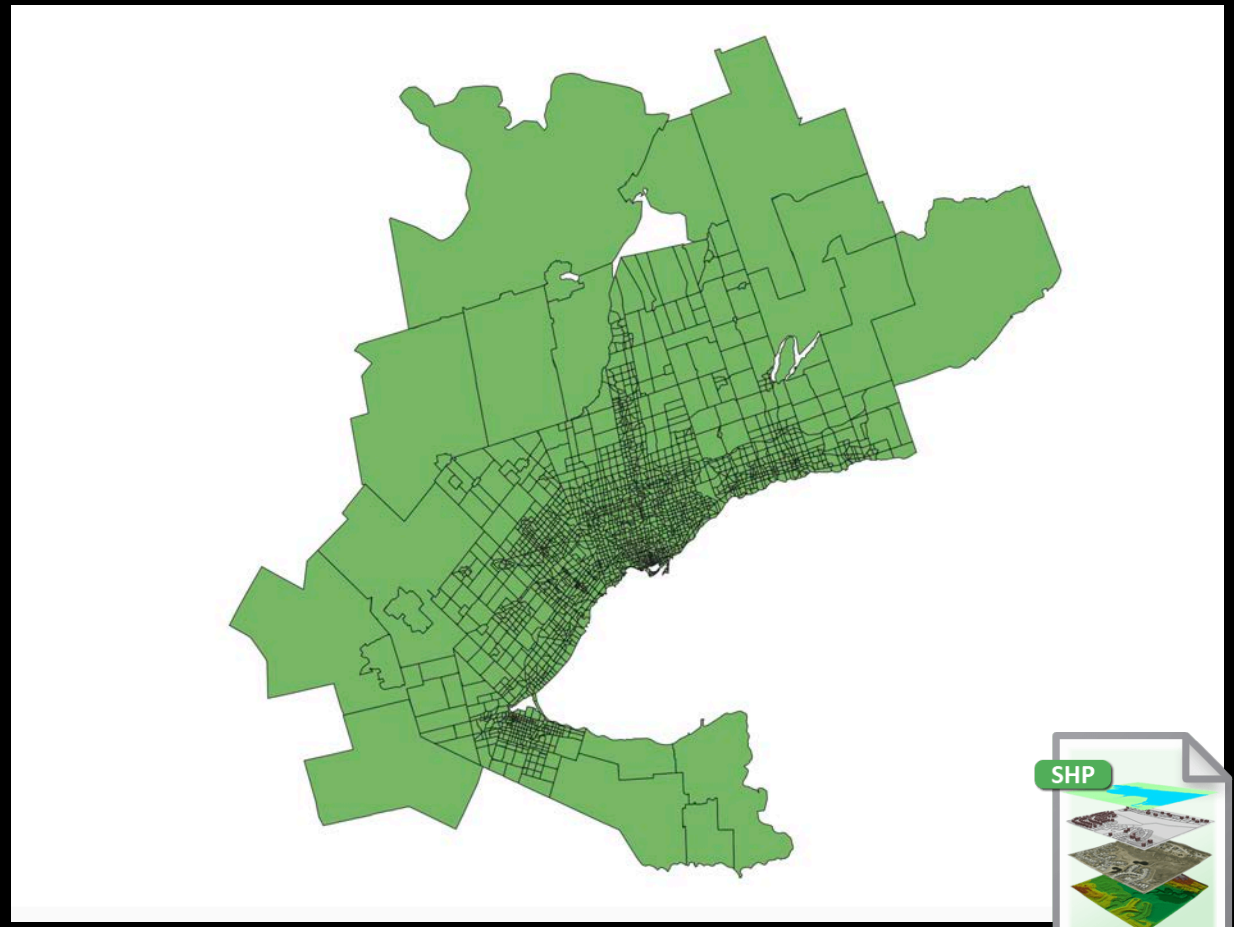
Shapefile → GeoJson



Images source: http://www.gisresources.com/understanding-shapefile-shp-file-format/ (shp) and http://chittagongit.com/icon/json-file-icon-12.html (json)

Transforming Data

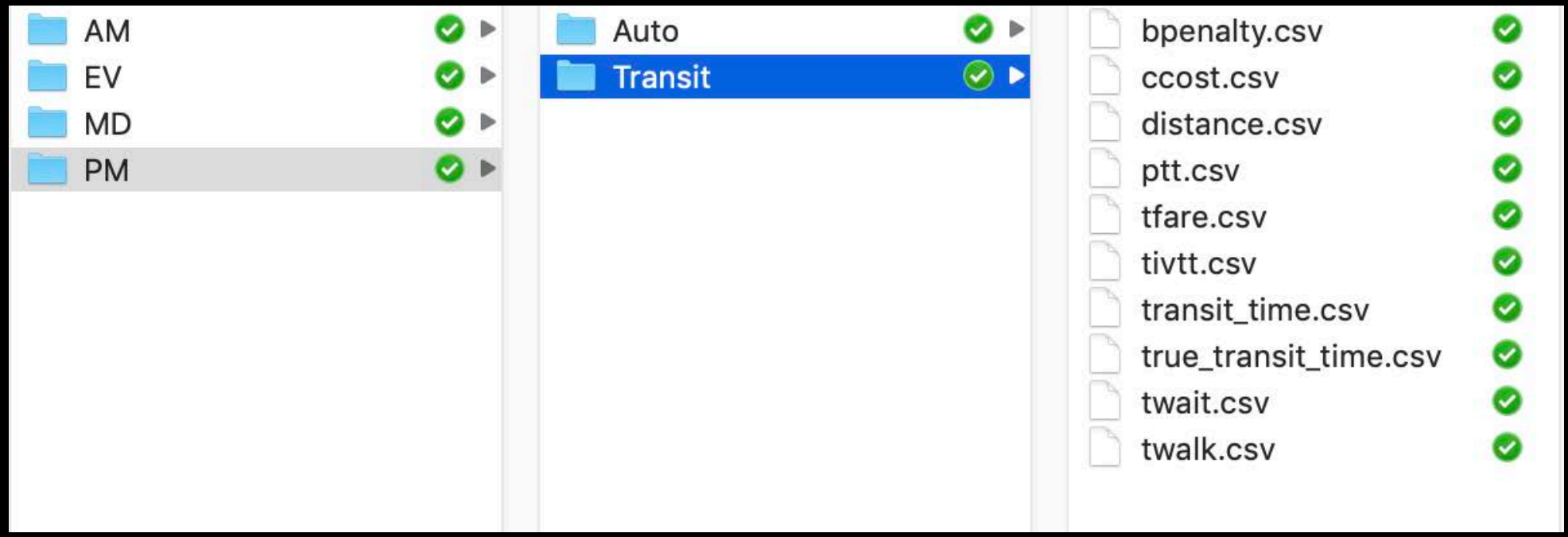
Shapefile x Geojson



Transforming Data

OD-Matrices to Data Tables and integrated with shapefile

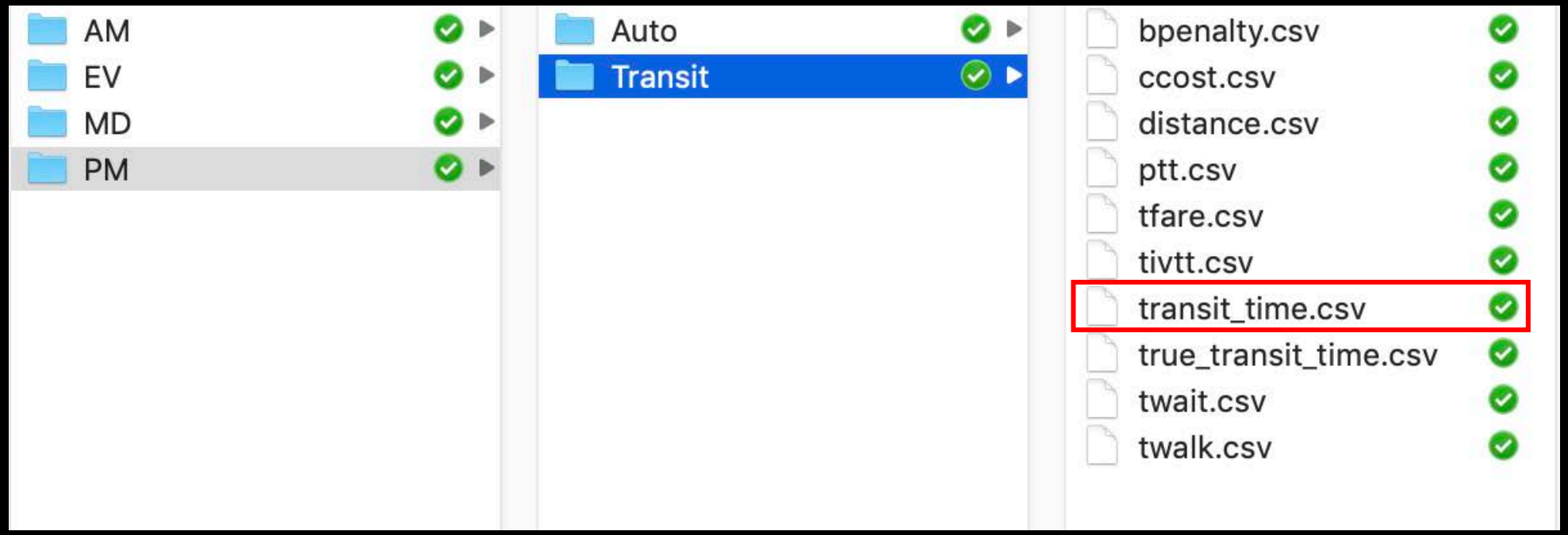
OD-Matrices - Directory



Transforming Data

OD-Matrices to Data Tables and integrated with shapefile

OD-Matrices - Directory



Transforming Data

OD-Matrices to Data Tables and integrated with shapefile

transit_time.csv

Zones O\D	1	2	3	4	5	6
1	0	29.7761	21.25672	24.78801	29.80301	38.22379
2	30.92986	0	22.35746	30.90995	35.72341	36.07895
3	20.17063	22.80784	0	20.66859	25.51105	32.99121
4	23.45448	27.36932	18.84467	0	21.38223	22.12184
5	28.57211	34.18865	24.28002	20.88825	0	22.06426

V4_zones.geojson

OBJECTID (int)	gta06_net0 (int)	ORIG_FID (int)	Shape_Leng (float)	Shape_Area (float)	FID_ (int)	NUM (int)	GTA06 (int)
1	1	1	3240.40115051	523608.711076	2151	1	1
2	2	2	10264.6422225	1179813.05658	1826	2	2
3	3	3	2149.07481928	220460.93084	2153	3	3
4	4	4	2439.97969026	253035.243478	1961	4	4
5	5	5	3050.44455033	501072.324501	1960	5	5
6	6	6	5489.01453565	1054622.94564	1939	6	6

origin	destiny	values	long.dest	lat.dest	long.orig	lat.orig
1	1	0	-79.31193845	43.66501279	-79.31193845	43.66501279
1	2	29.7761	-79.31670115	43.65815388	-79.31193845	43.66501279
1	3	21.25672	-79.32418316	43.66146457	-79.31193845	43.66501279
1	4	24.78801	-79.33393307	43.65813576	-79.31193845	43.66501279
1	5	29.80301	-79.34421793	43.65486016	-79.31193845	43.66501279
1	6	38.22379	-79.33648769	43.65227635	-79.31193845	43.66501279

Input CSV for Kepler and Deck

Next Steps

iCity Working In Progress

- Visualizations in Kepler.gl and Deck.gl
- New design and prototypes for:
  - Complete Streets dashboard
  - Transit Management dashboard
  - iCity ORF Integrated Website

Acknowledgements

- Ontario Research Funding (ORF)
- NSERC
- Esri Canada.

References

- https://journals.sagepub.com/doi/full/10.1177/1473871611415994
- https://www.digitalvidya.com/blog/what-is-data-wrangling/

Thank You!

Find out more about research at OCAD U at:
http://www.ocadu.ca/research