

Creative Data Mining

Uncover and evaluate

Lecture 1: Introduction

20 | 02 | 2017

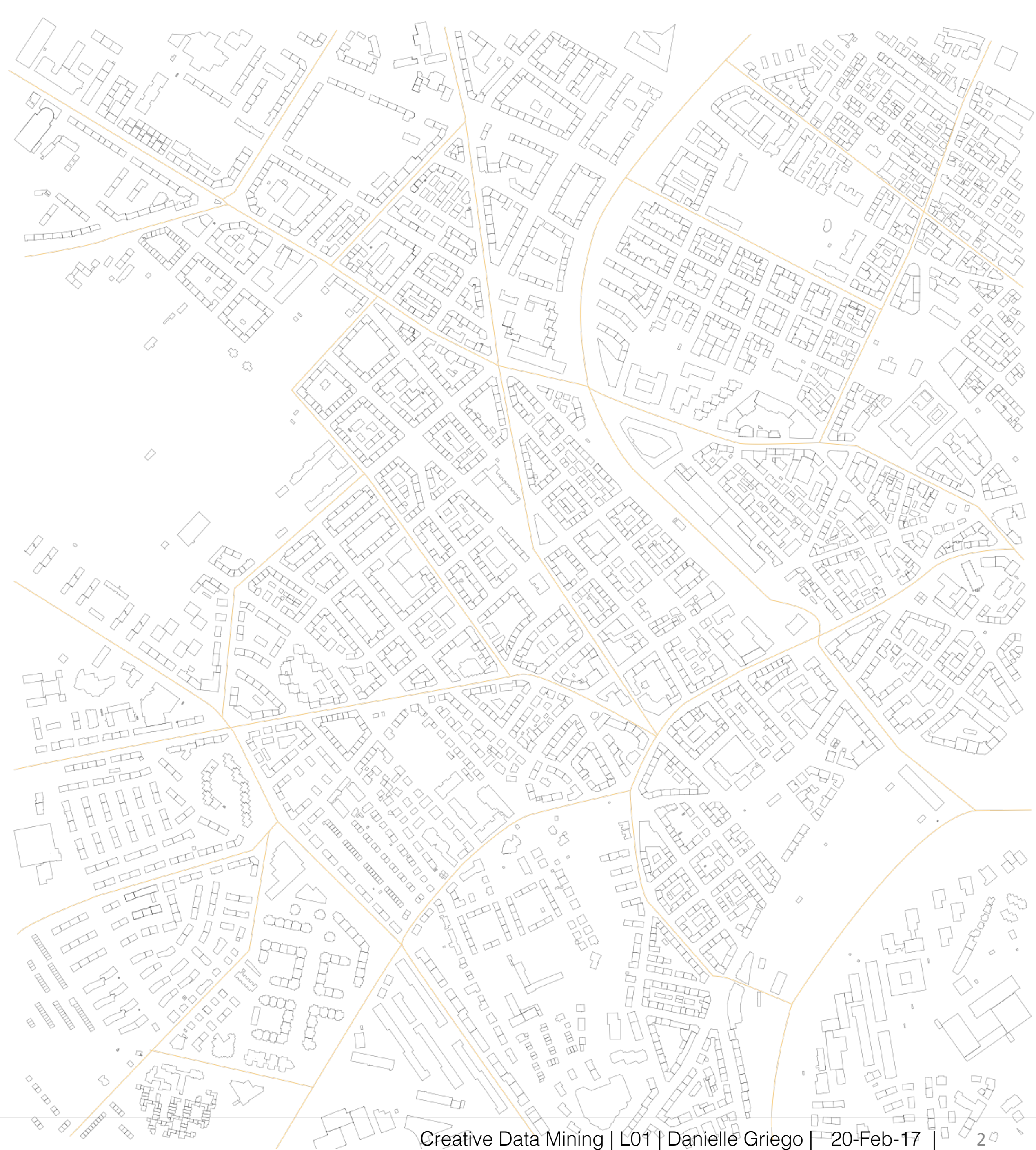
Dr. Daniel Zünd, zuend@arch.ethz.ch

Danielle Griego, griego@arch.ethz.ch

Artem Chirkin, chirkin@arch.ethz.ch

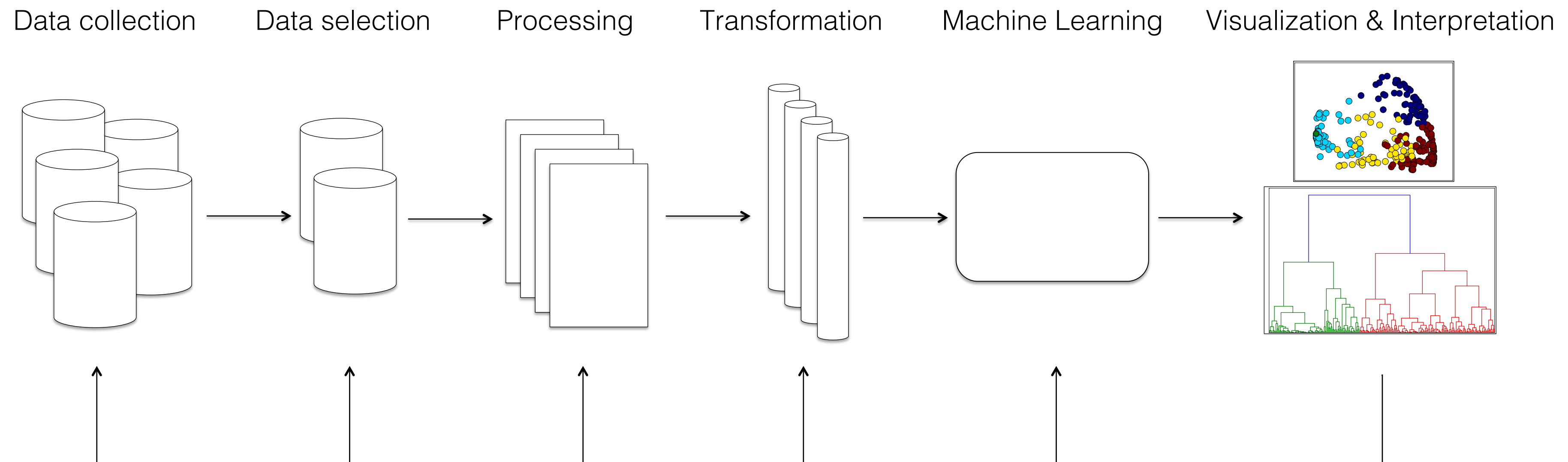
What we'll cover today

- Background
- Course Structure
- Data Mining for Architects and Urban Planners
- Semester project
- Learning objectives & Course schedule
- Discussion
- Homework
- Install Python and PyCharm



Background

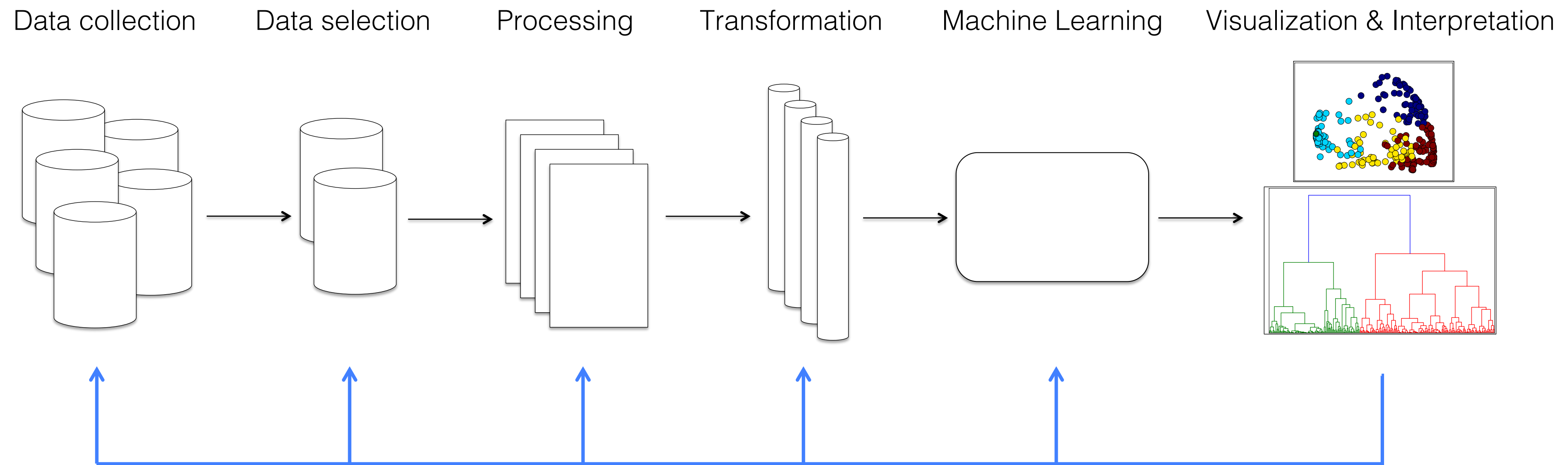
What is Data Mining?



Typical Knowledge Discovery Diagram (KDD)

Background

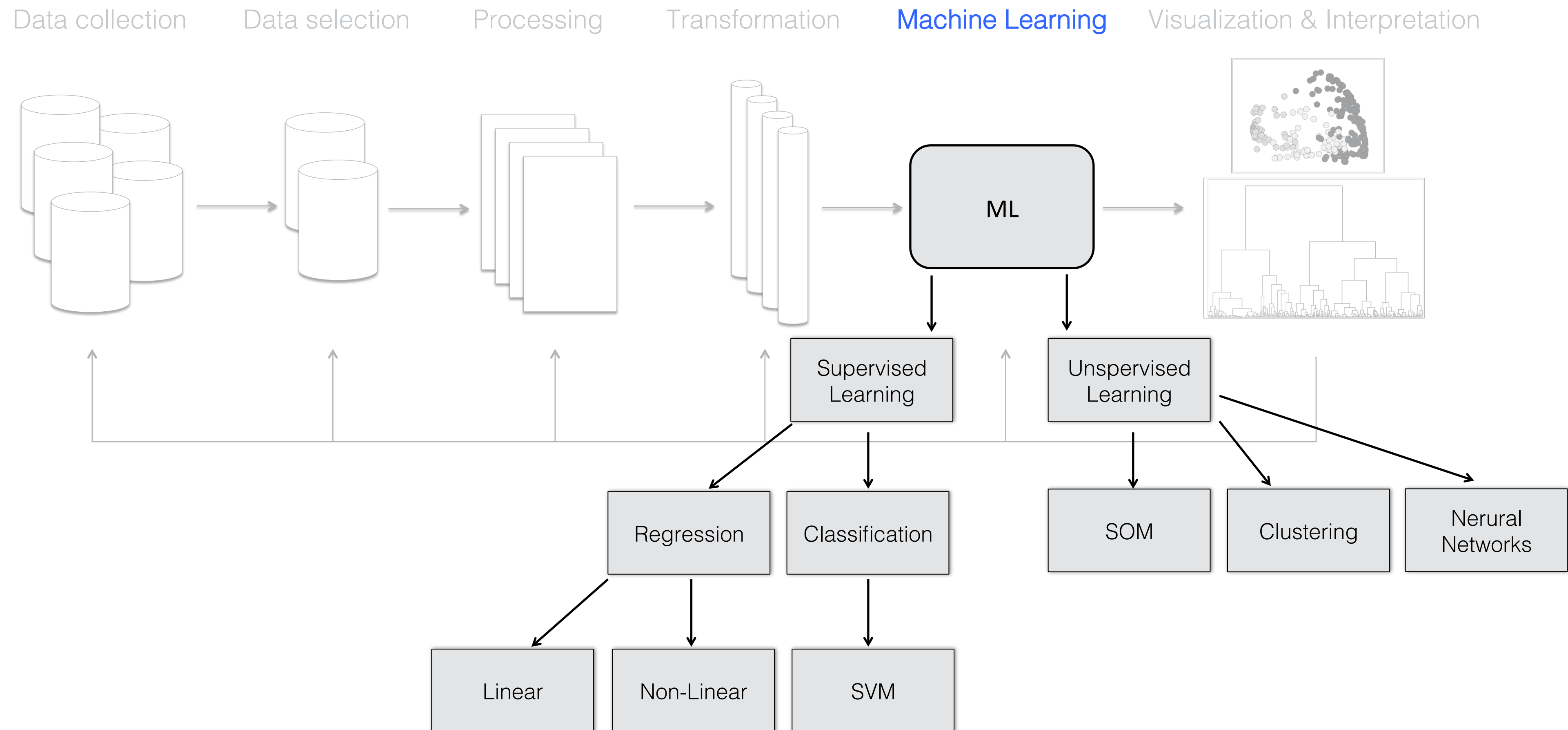
It is an exploratory and iterative process



Typical Knowledge Discovery Diagram (KDD)

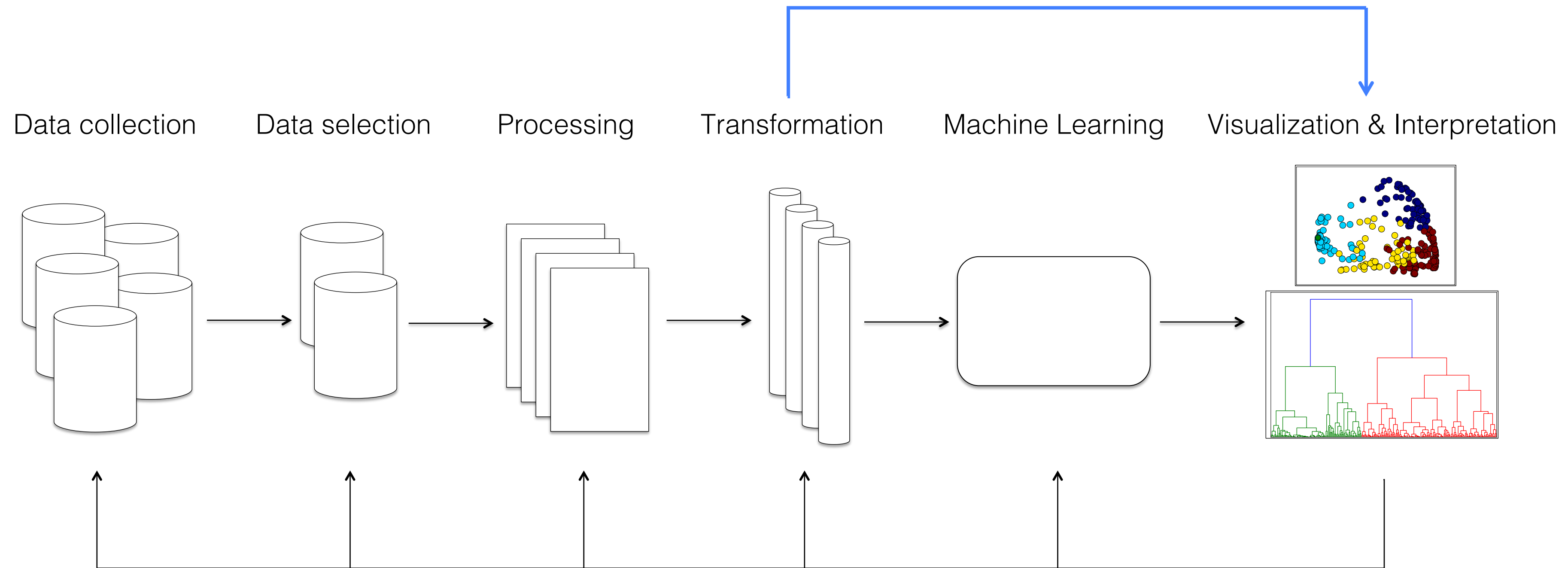
Background

What is machine learning?



Background

Data mining does not always include machine learning, for example in many time-series analysis and geo-referenced data visualization

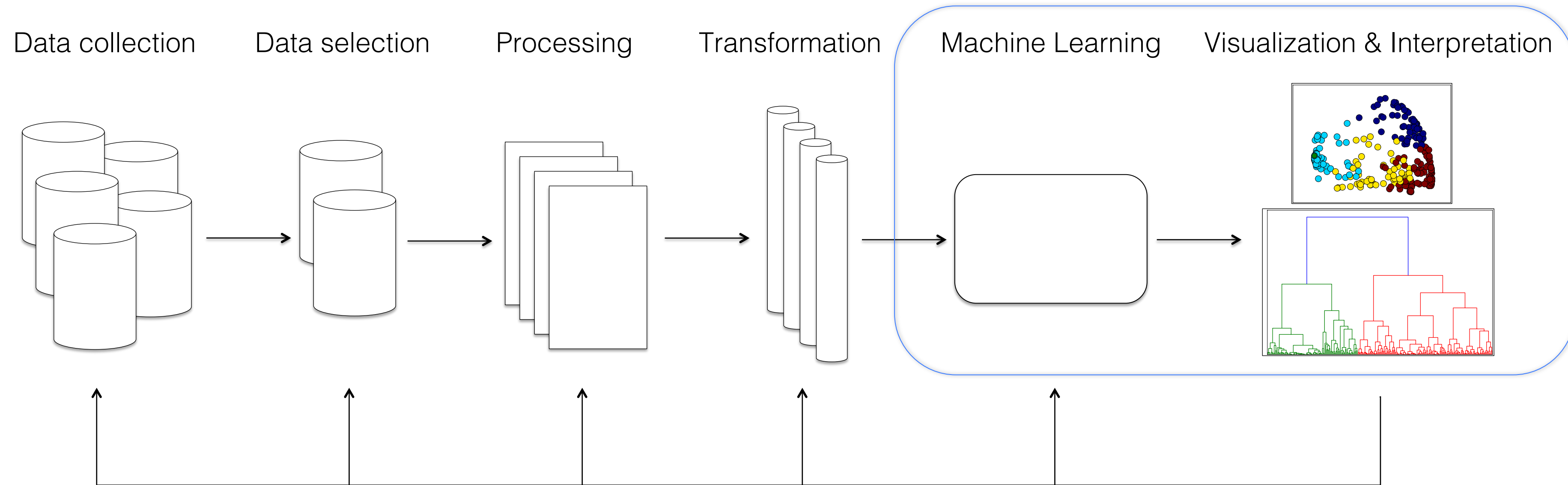


Typical Knowledge Discovery Diagram (KDD)

Background

How can data mining be creative?

What do we want to know?

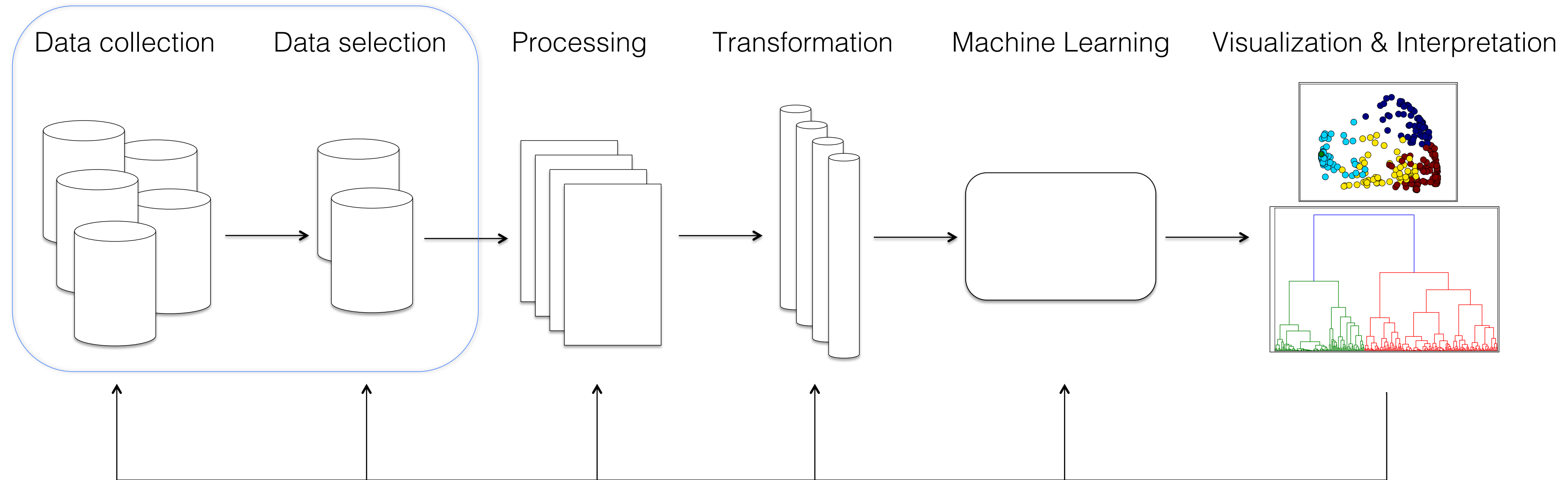


Typical Knowledge Discovery Diagram (KDD)

Background

How can data mining be creative?

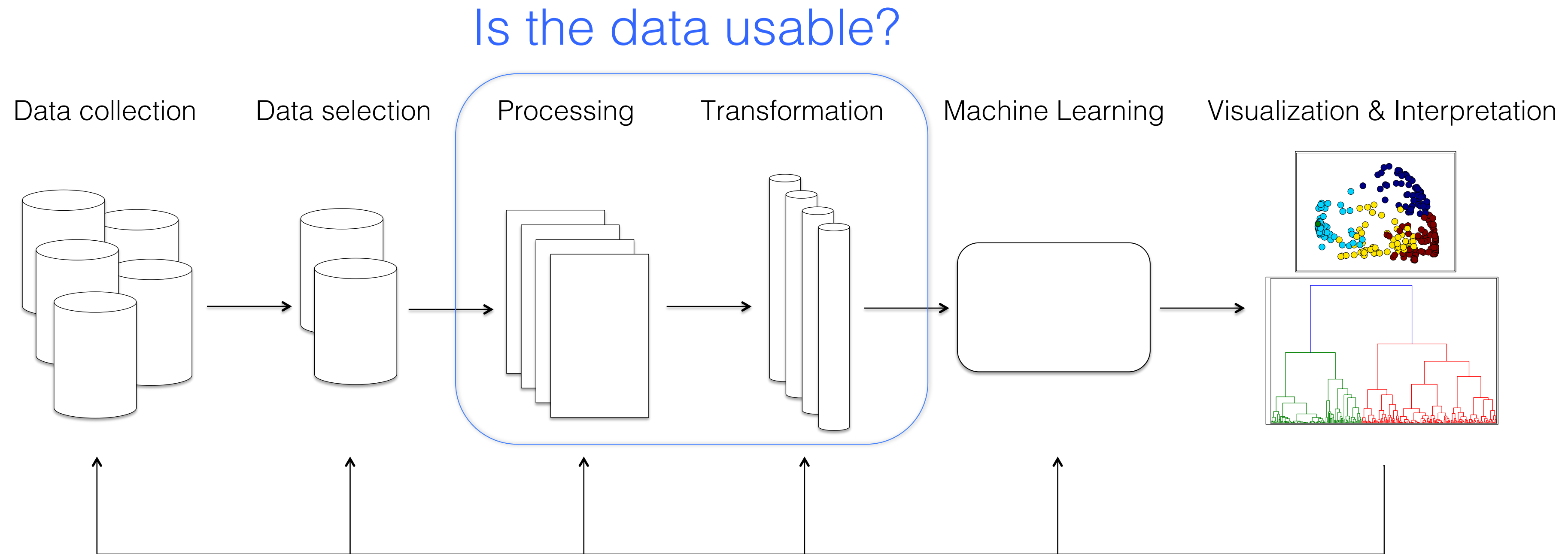
Domain specific data source(s)



Typical Knowledge Discovery Diagram (KDD)

Background

The not-so creative, but essential part of data mining



Typical Knowledge Discovery Diagram (KDD)

Background

Types of data

Original data sources:

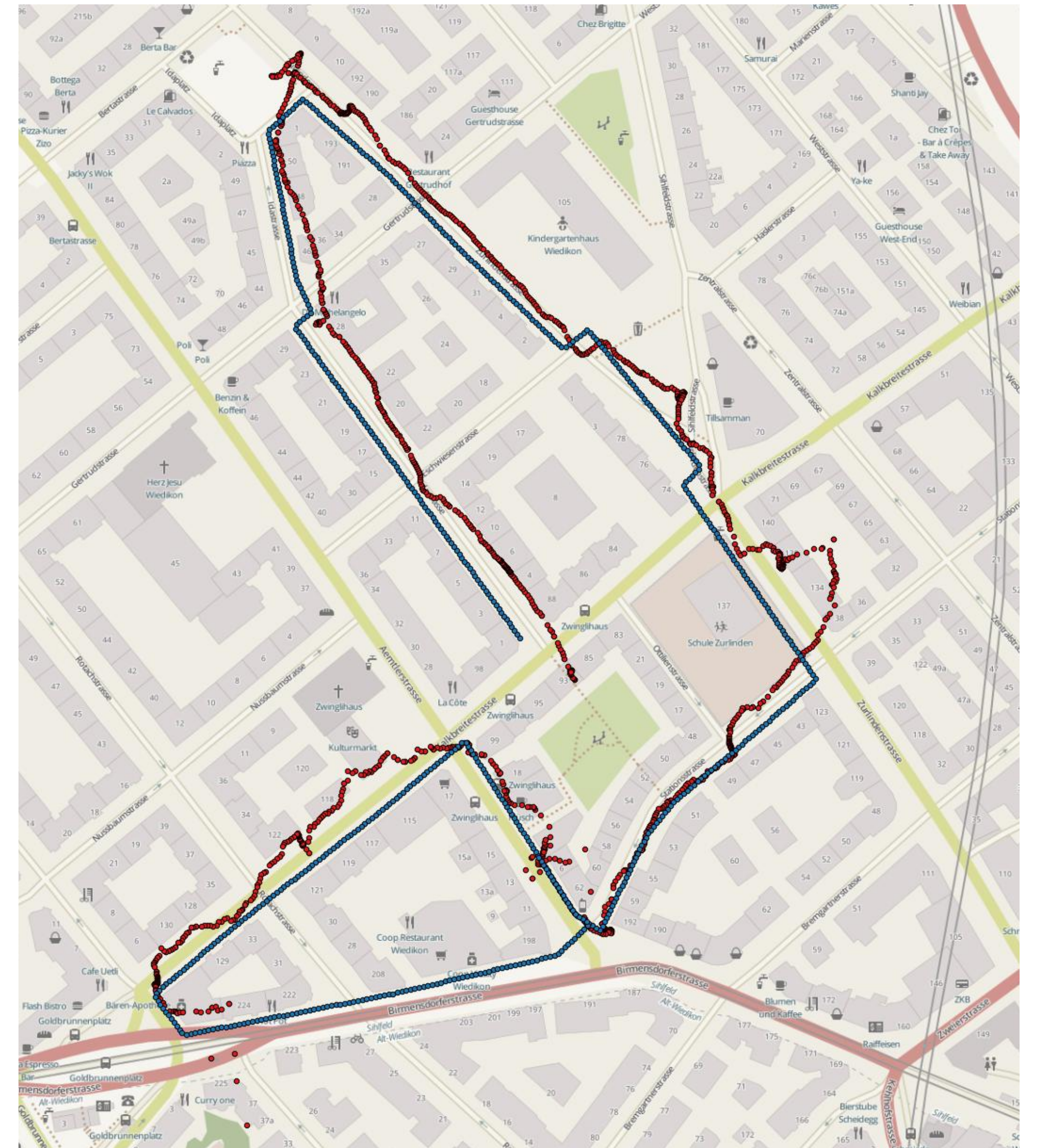
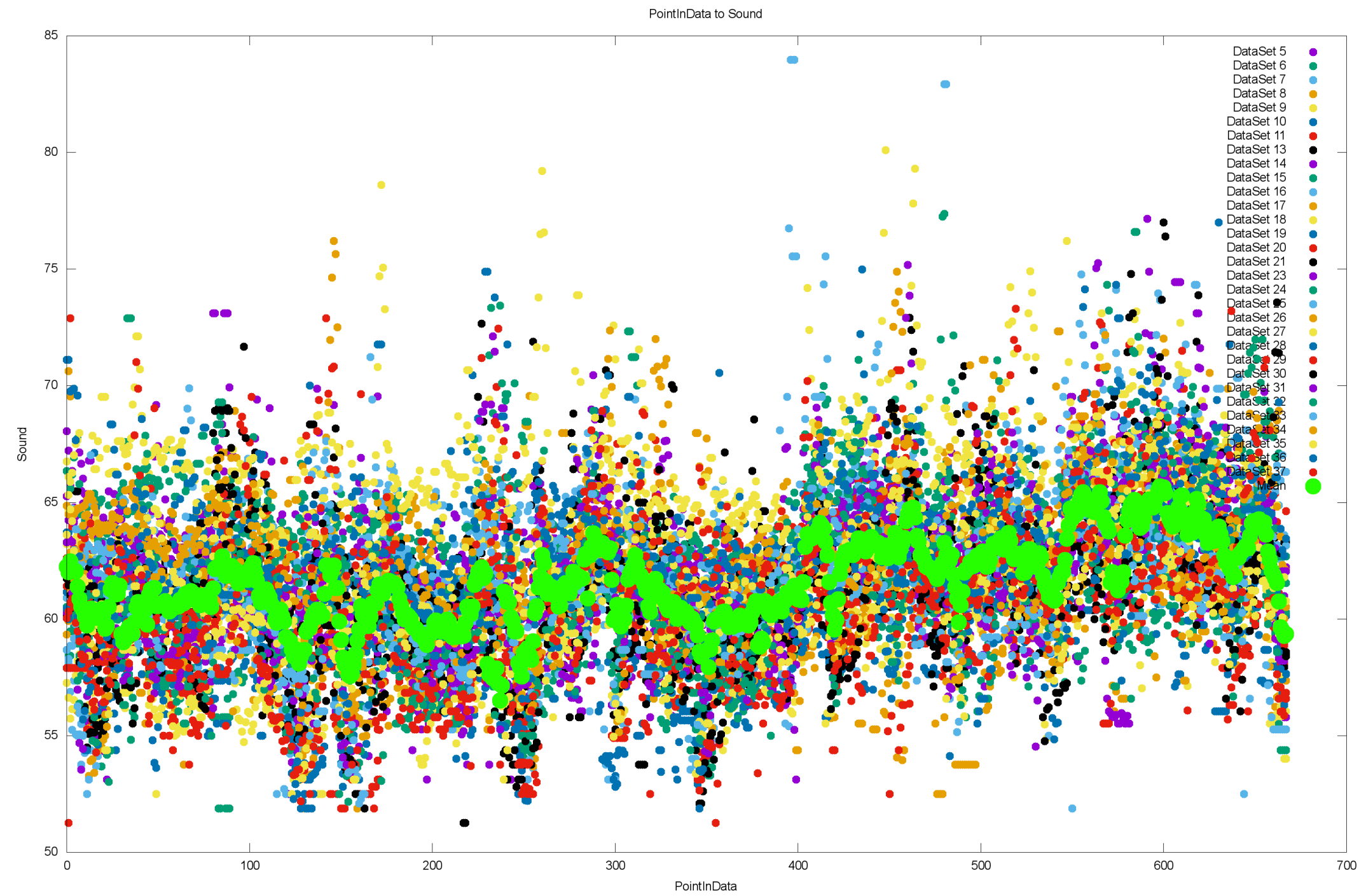
- Images (pixels)
- Categorical (labels)
- Numeric (integers and floats)
- Binary (0/1) - useful for yes/no, true/false...
- Metadata - data descriptors for multi-dimensional data sets
-

Processed for analysis



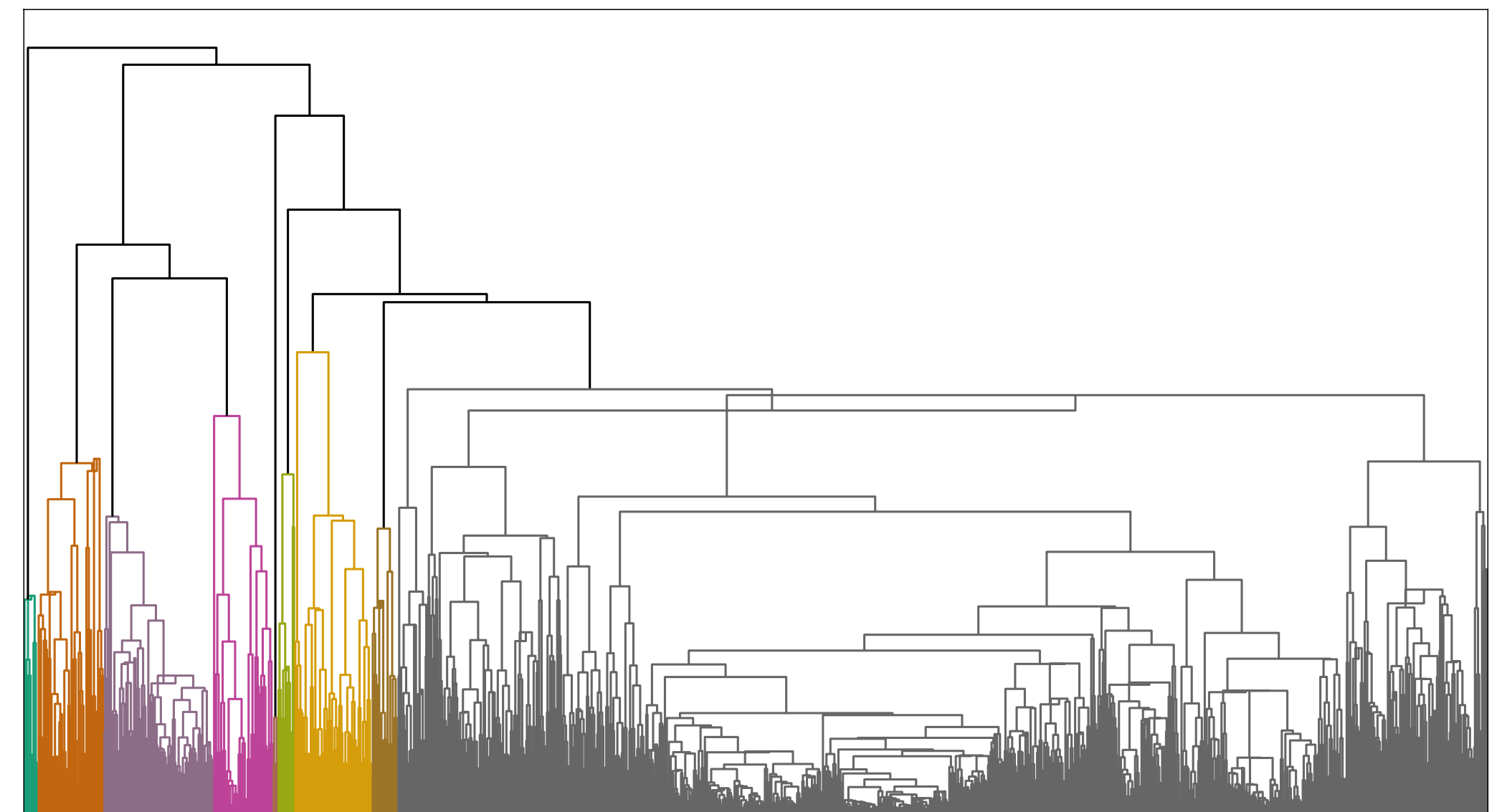
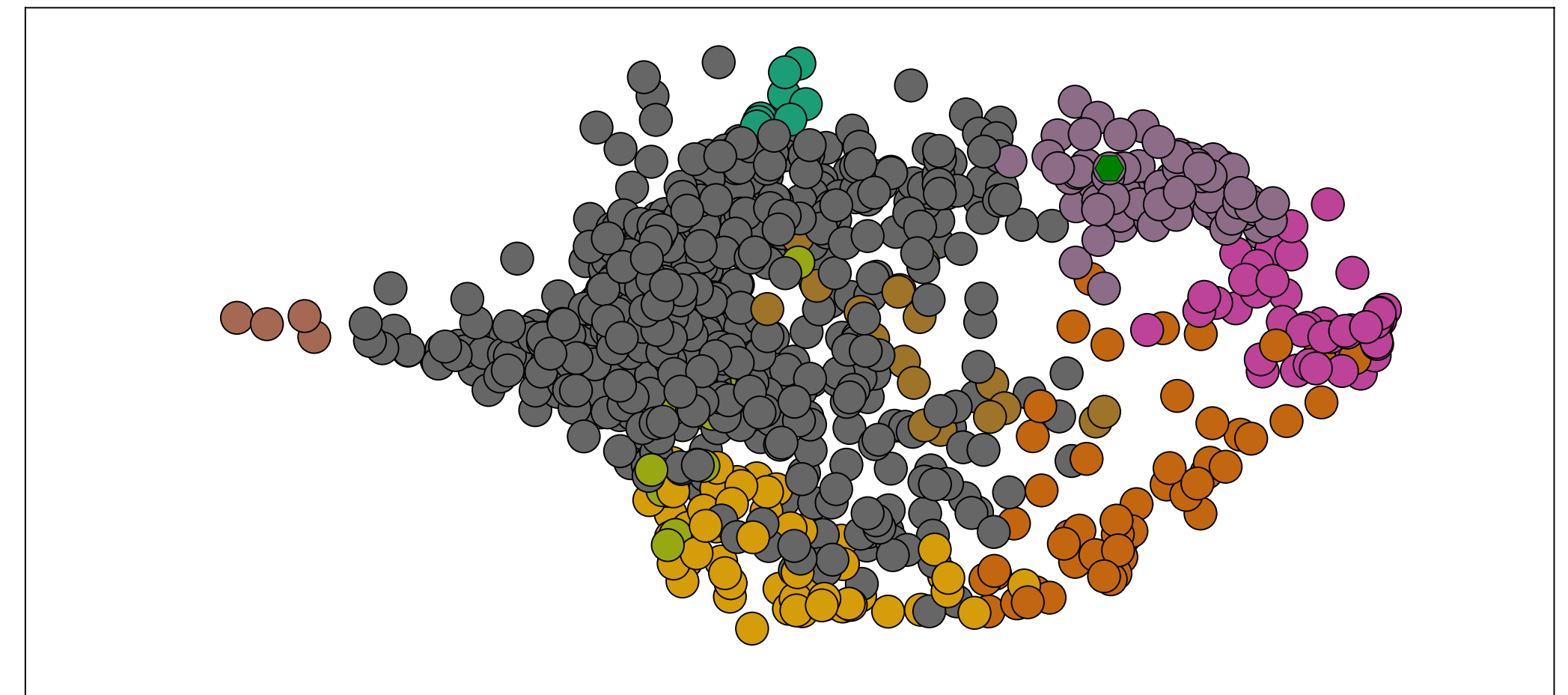
Background

Types of analysis, visualization & interpretation: [Time Series](#) and [geo-referenced data visualization](#)



Background

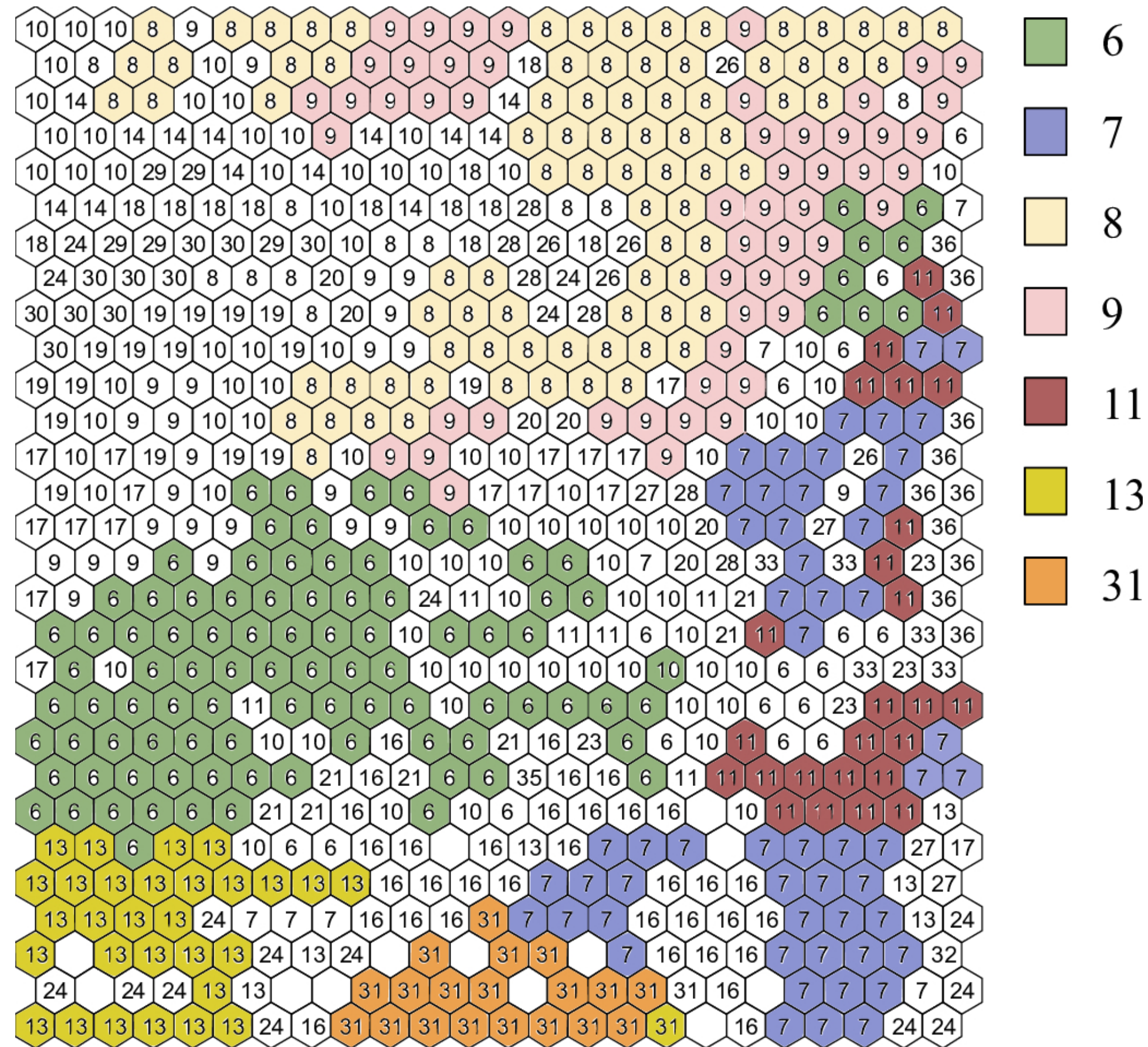
Types of analysis, visualization & interpretation: [Hierarchical clustering](#)



Zünd D. (2016). A Meso-Scale Framework to Support Urban Planning (Doctoral dissertation)

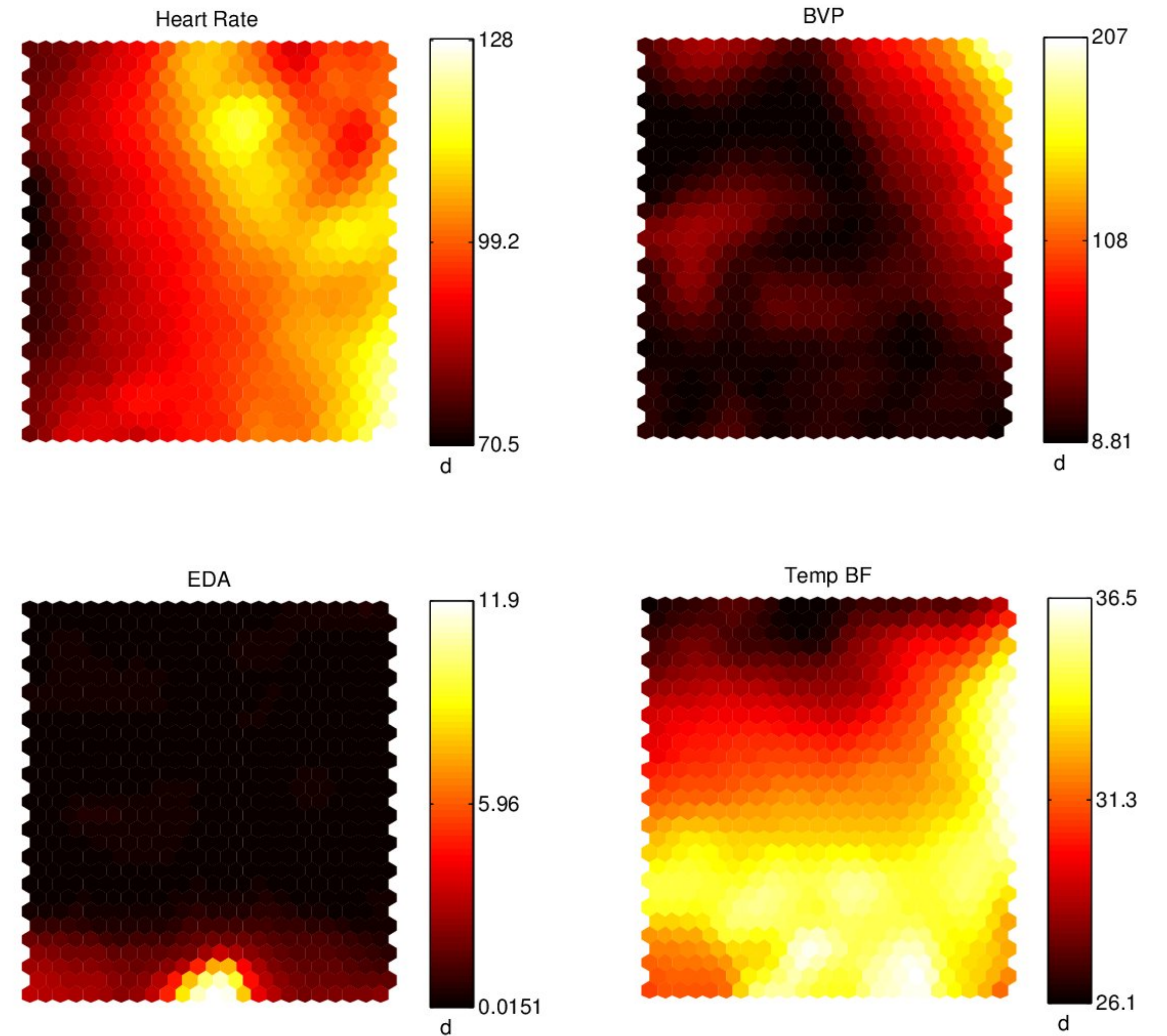
Background

Types of analysis, visualization & interpretation: [SOM- Self organizing Maps](#)



Preliminary data analysis:

SOM clustering map of participants (indicated by numbers)



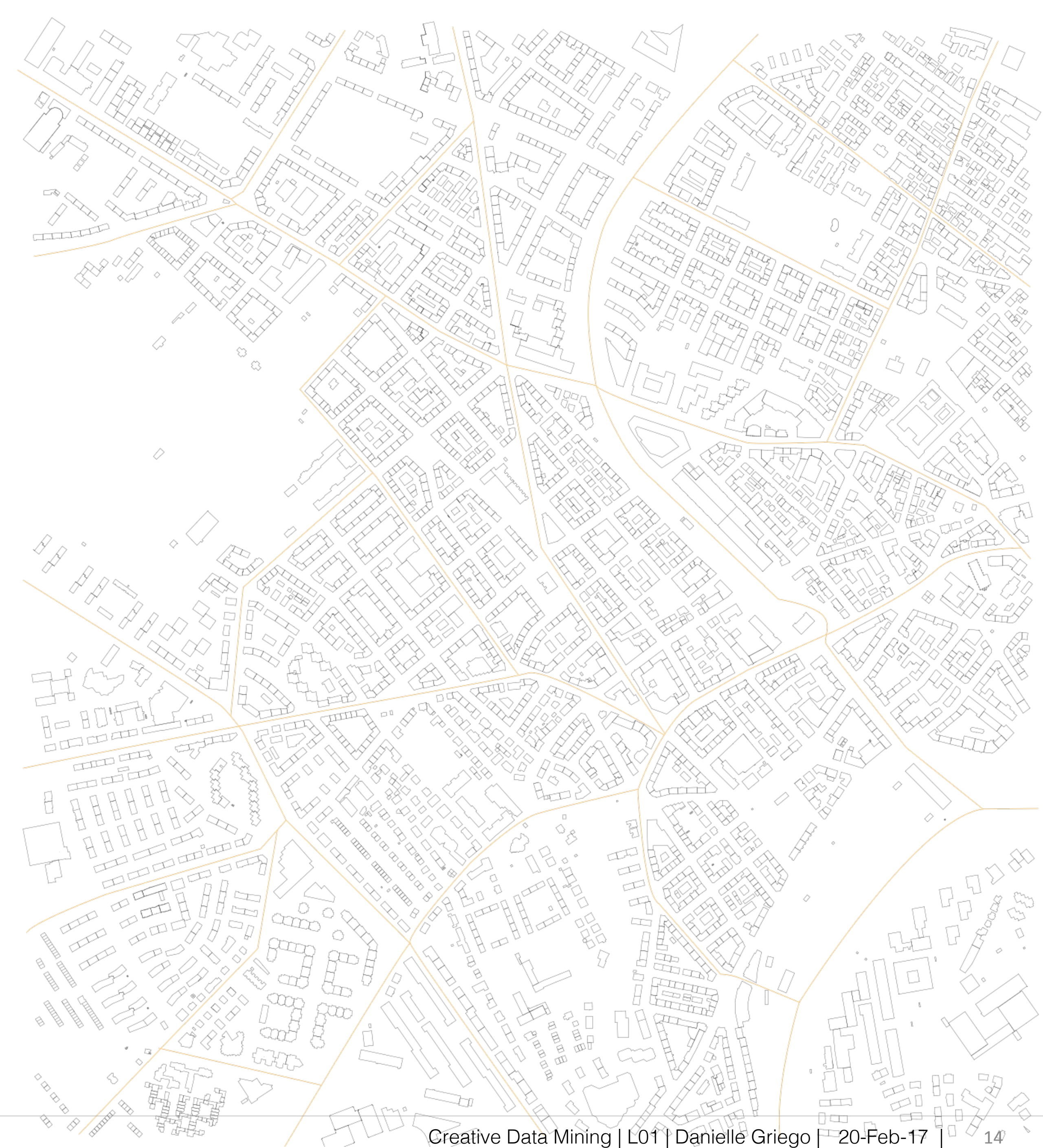
Changing of participants behavior – biofeedback responses

Ojha V. ESUM-Analyzing Tradeoffs between Energy and Social Performance of Urban Morphology

Course Structure

Integrating the “creative aspects” of data mining

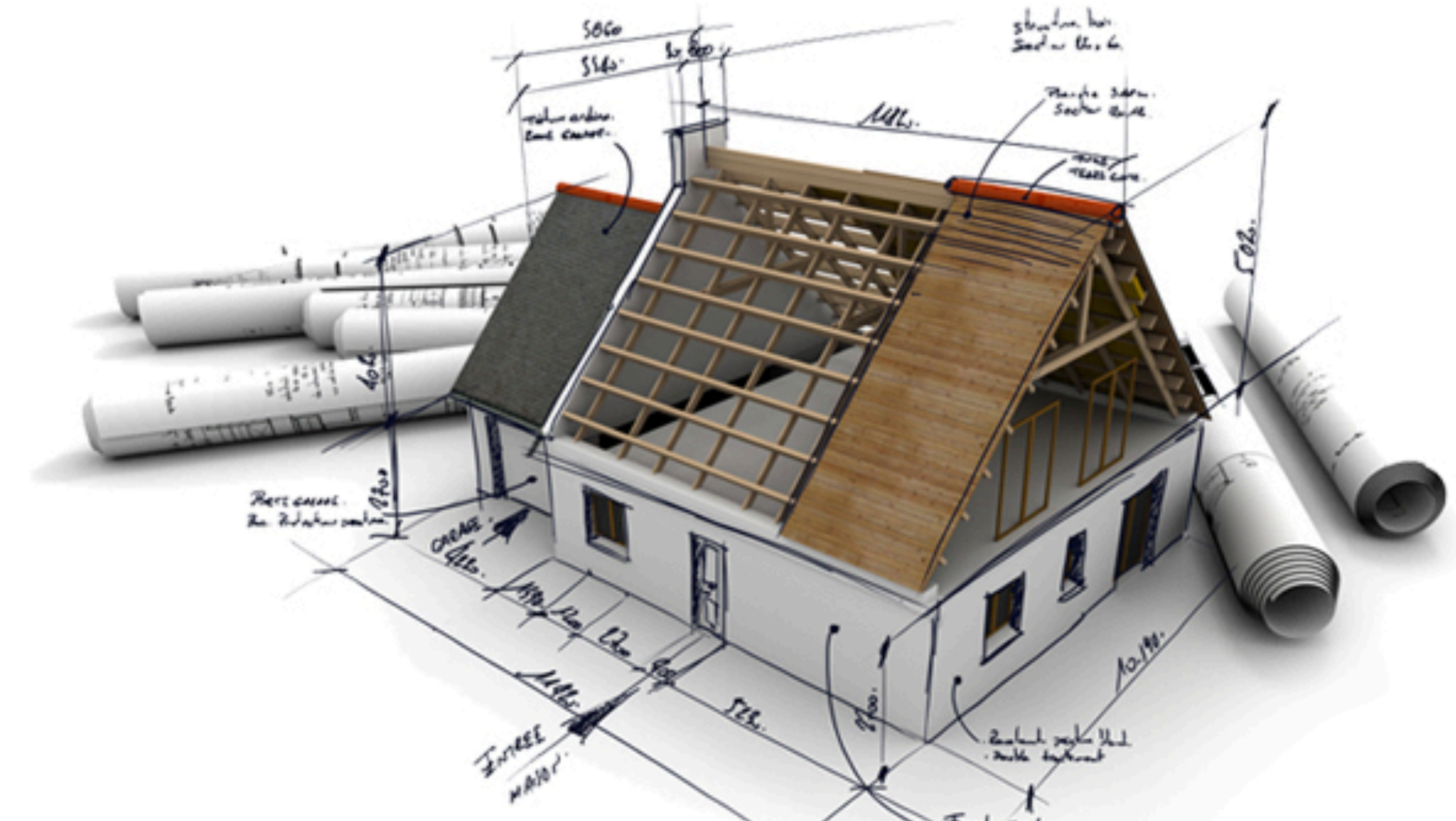
		Analysis visualization & interpretation	
		Manual	Automated
Data Source	Manual		(OPTIONAL)
	Automated	BLOCK 1	BLOCK 2



Course Structure

Elaborating on the traditional architectural process

		Analysis visualization & interpretation	
		Manual	Automated
		Iterative evaluations	
Data Source	Manual		
	Automated		
	Hand-drawn sketches		

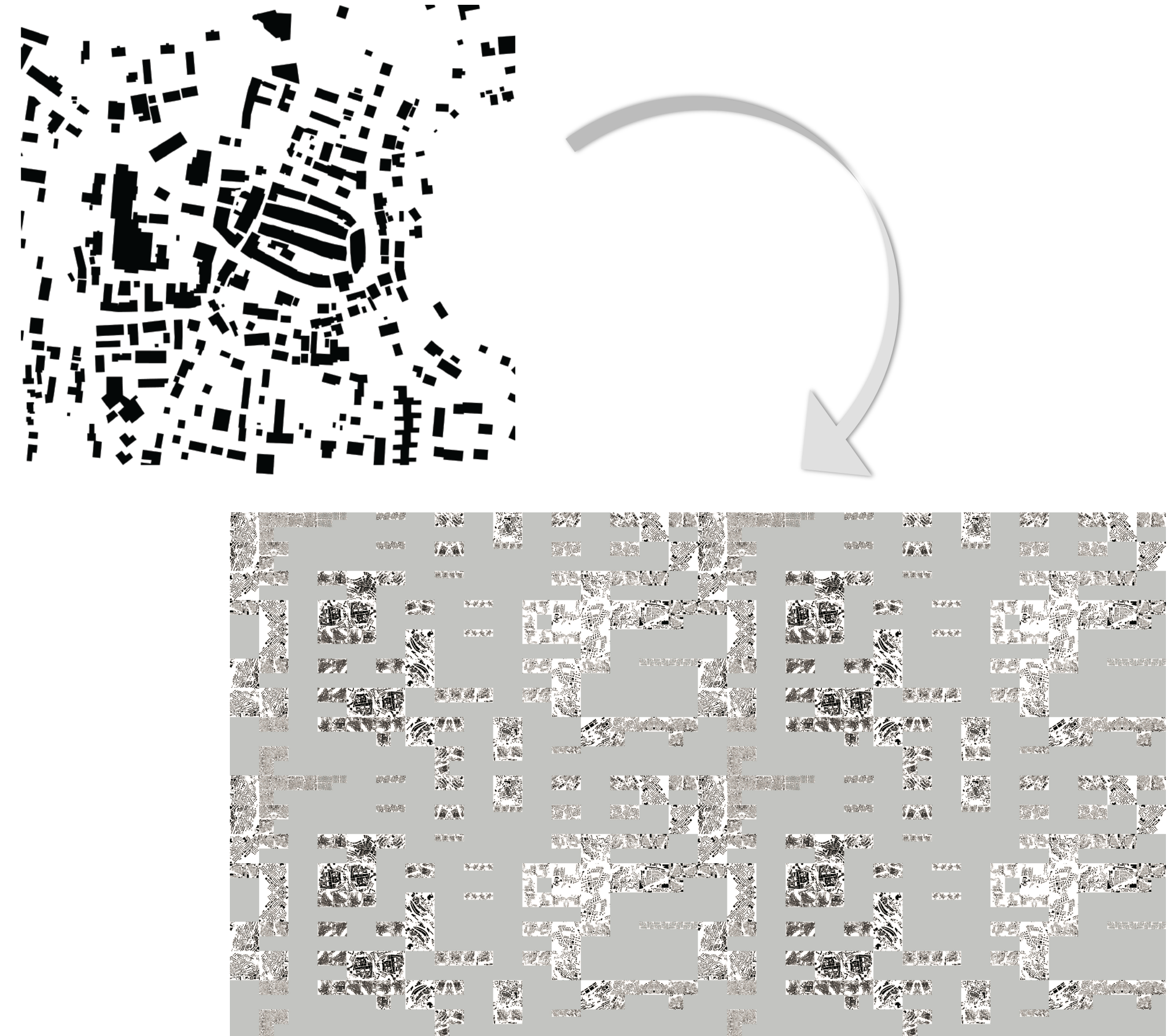


<http://www.stamfordbuildingandconstruction.co.uk/our-services/architectural-drawings>

Course Structure

Process taught in previous semesters

		Analysis visualization & interpretation	
		Manual	Automated Machine Learning: SOM
Data Source	Manual Hand-drawn sketches		(OPTIONAL)
	Automated		

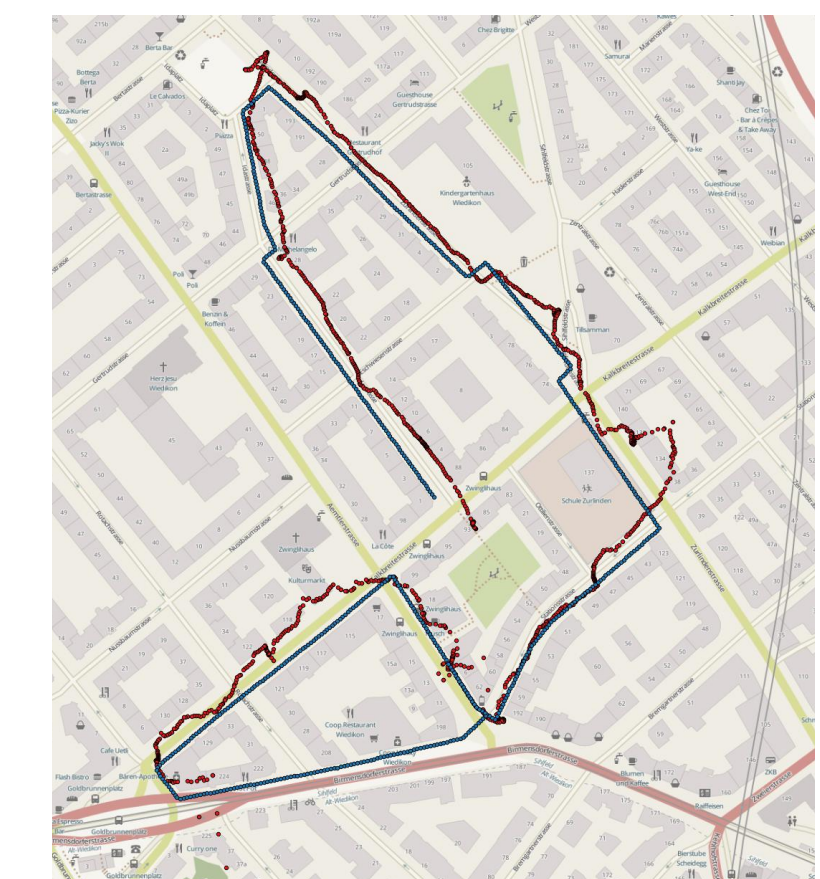
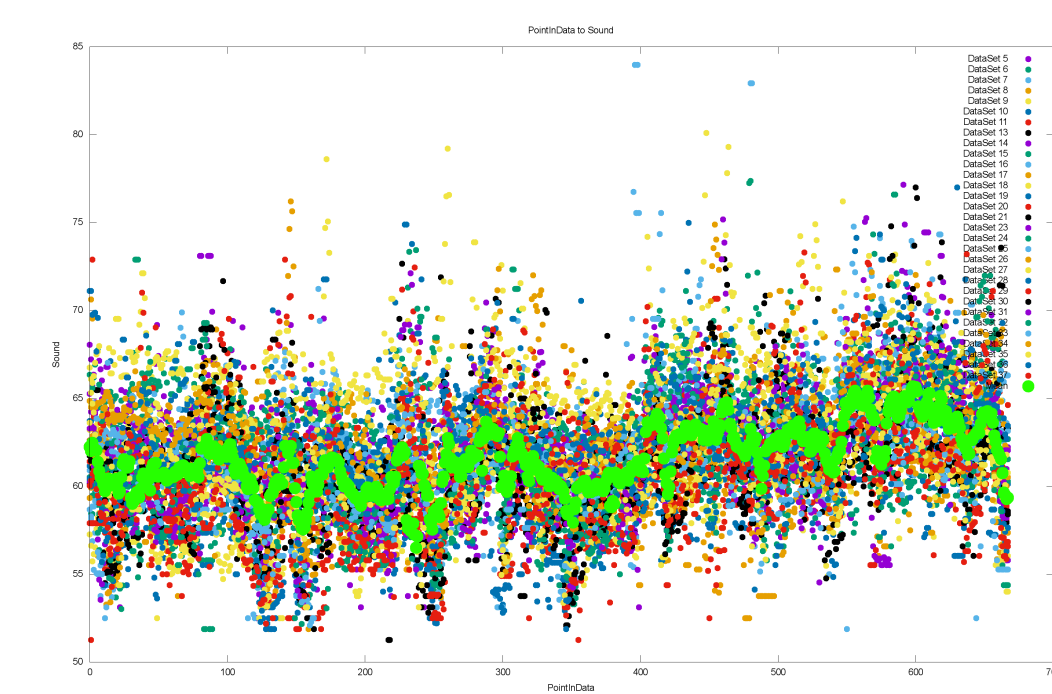
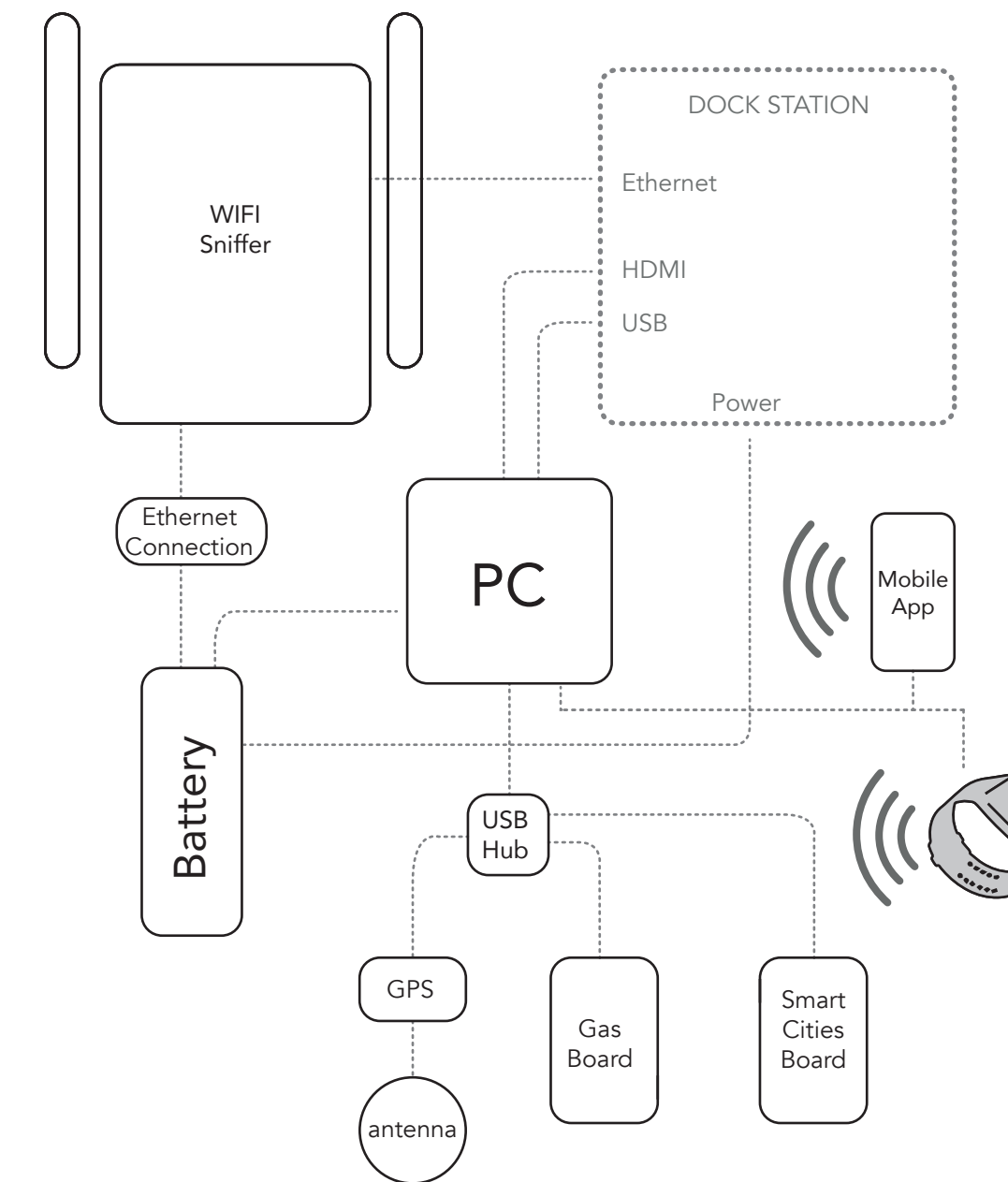


Final Project from Moritz Berchtold, Creative Data Mining FS2015

Course Structure

Block 1: Time-series & geo-referenced data visualizations

		Analysis visualization & interpretation	
		Manual	Automated
		Time-series & geo-referenced data visualization	
Data Source	Manual		
	Automated	BLOCK 1	
		Sensor data	

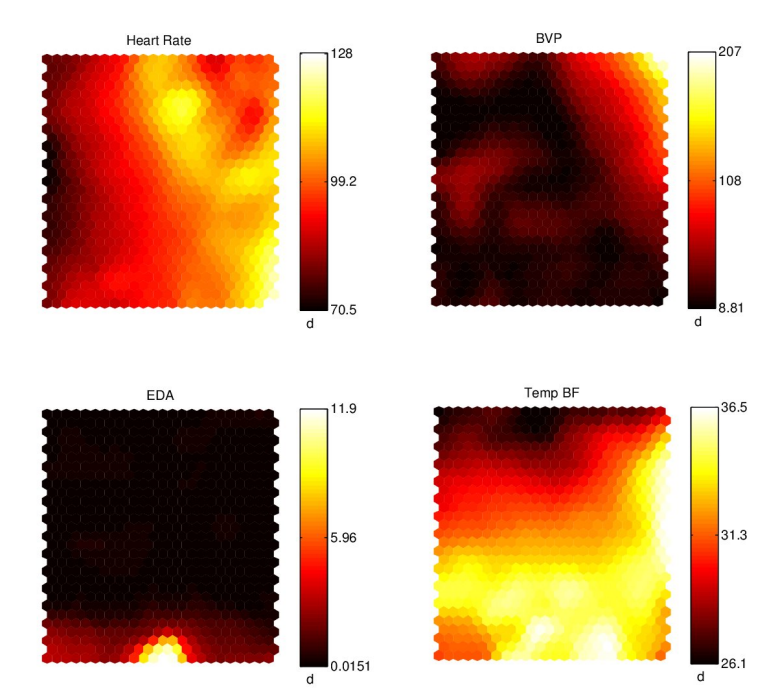
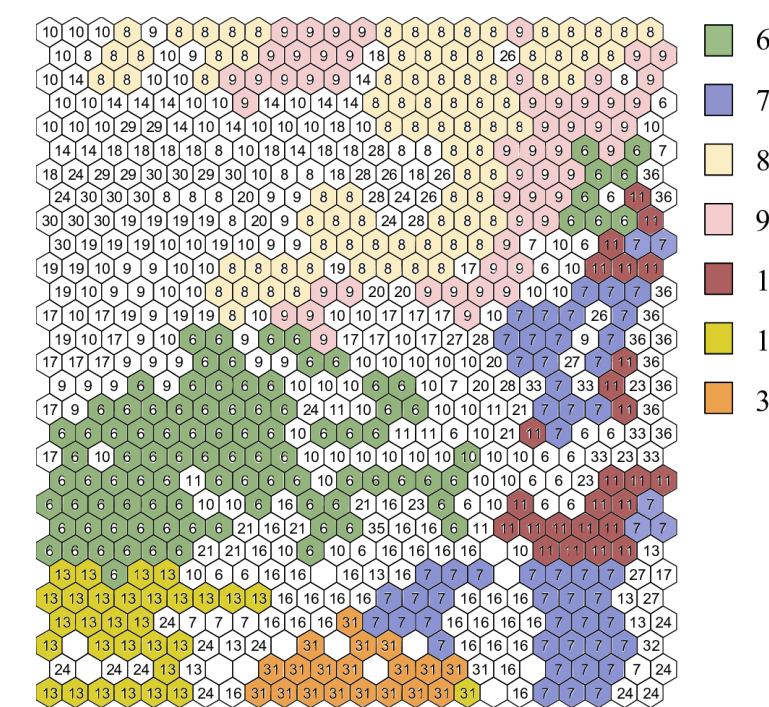
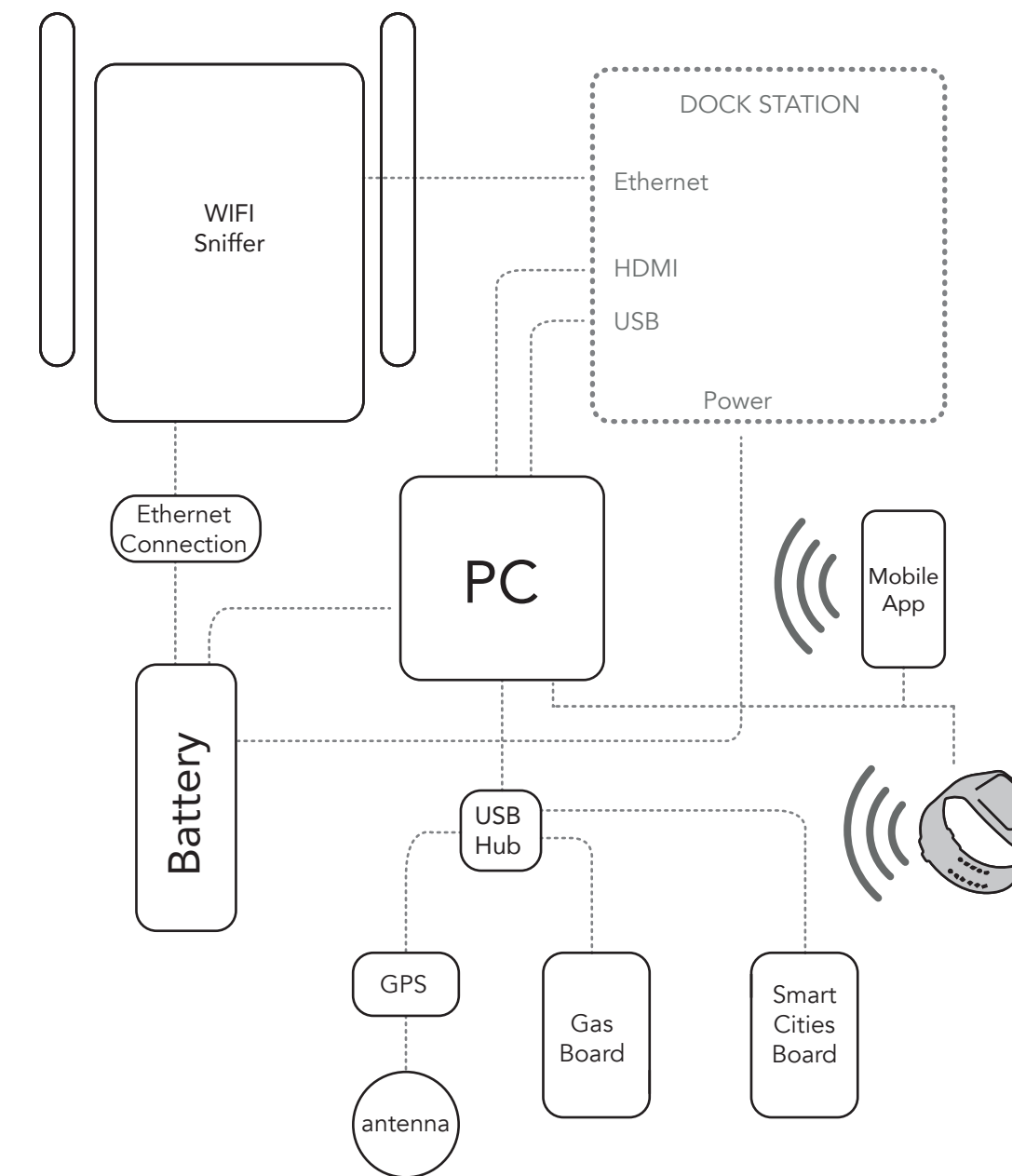


ESUM project experimental equipment set up and data analysis techniques

Course Structure

Block 2: Machine Learning

		Analysis, visualization & interpretation	
		Manual	Automated Machine Learning Techniques
Data Source	Manual		
	Automated Sensor data		BLOCK 2



ESUM project experimental equipment set up and data analysis techniques

Data Mining for Architects and Urban Planners?

A few examples



National data collection project

Geo-referenced sensor data visualization

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

Singapore's National Science Experiment 43,000 Students Exploring Their Personal Data

The Singapore University of Technology and Design facilitated an ambitious large-scale science experiment in September and November 2015 which saw over 43,000 students carrying sensors designed to measure temperature, humidity, pressure, light, noise, among other physical parameters in a project supported by the National Research Foundation and carried out with partners from the Ministry of Education and the Singapore Science Center. The sensors were designed to be able to localize themselves in their environments using a radio-map of Singapore, and to be able to identify which transportation mode was being used during the participant's daily travels. This talk will center on the massive data set which the SUTD is in the process of analysing and sharing, and how it can be leveraged to learn things about Singapore's built environment.

Erik Wilhelm

Erik Wilhelm is an assistant professor in the Engineering Product Development Pillar at the Singapore University of Technology and Design. He earned his PhD from the ETH-Zurich while studying multi-criteria vehicle design, data analytics, and control optimization. While in Zürich, Dr. Wilhelm co-founded a start-up in the area of vehicle telematics for reducing on-road energy use. His post-doctoral research was performed at the Massachusetts Institute of Technology in the Field Intelligence Lab.

Wednesday, 3 February 2016
12 (noon) – 1 pm

Value Lab Asia
Singapore-ETH Centre
#06-01, CREATE Tower
1 CREATE Way
Singapore 138602

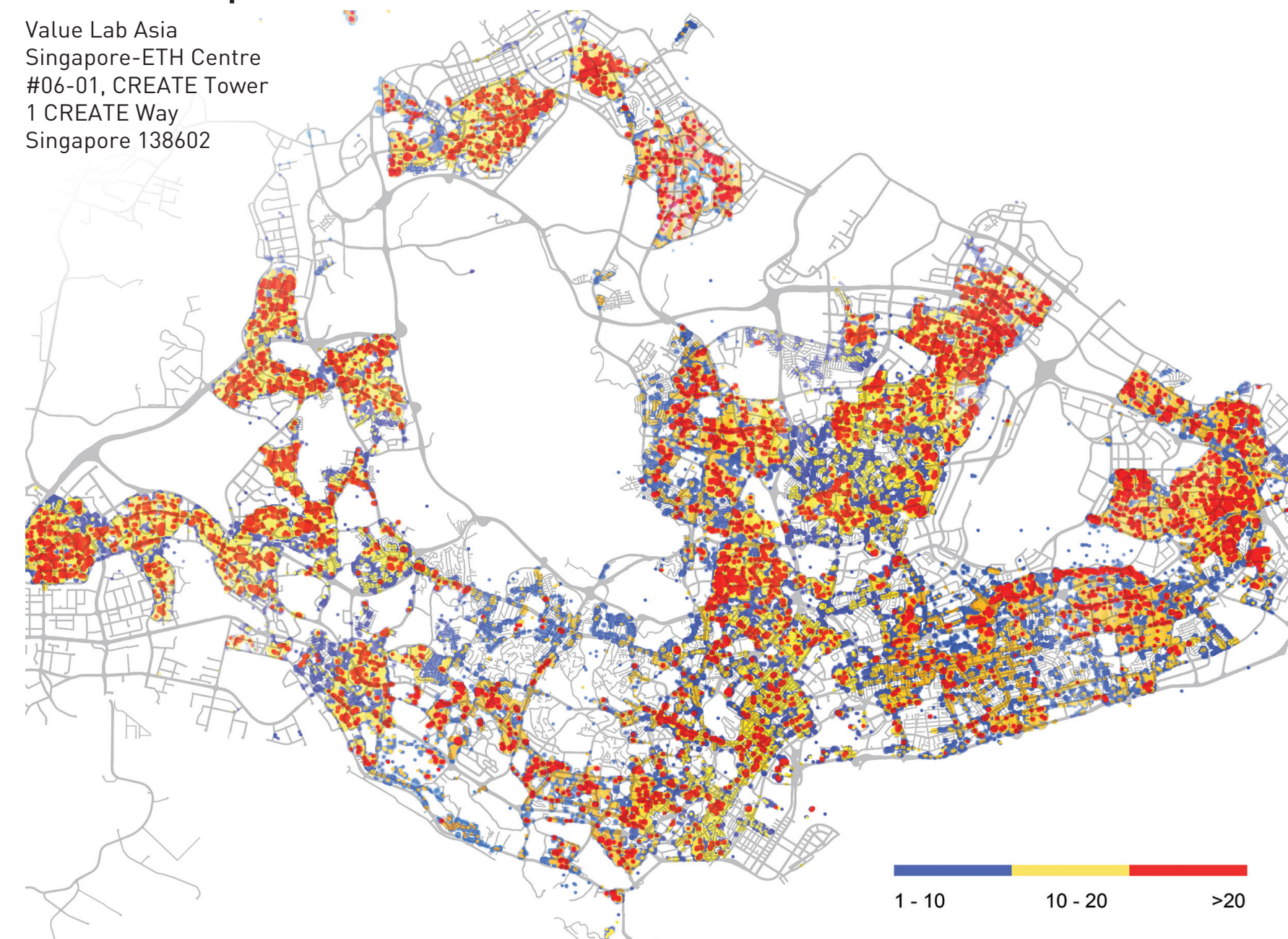
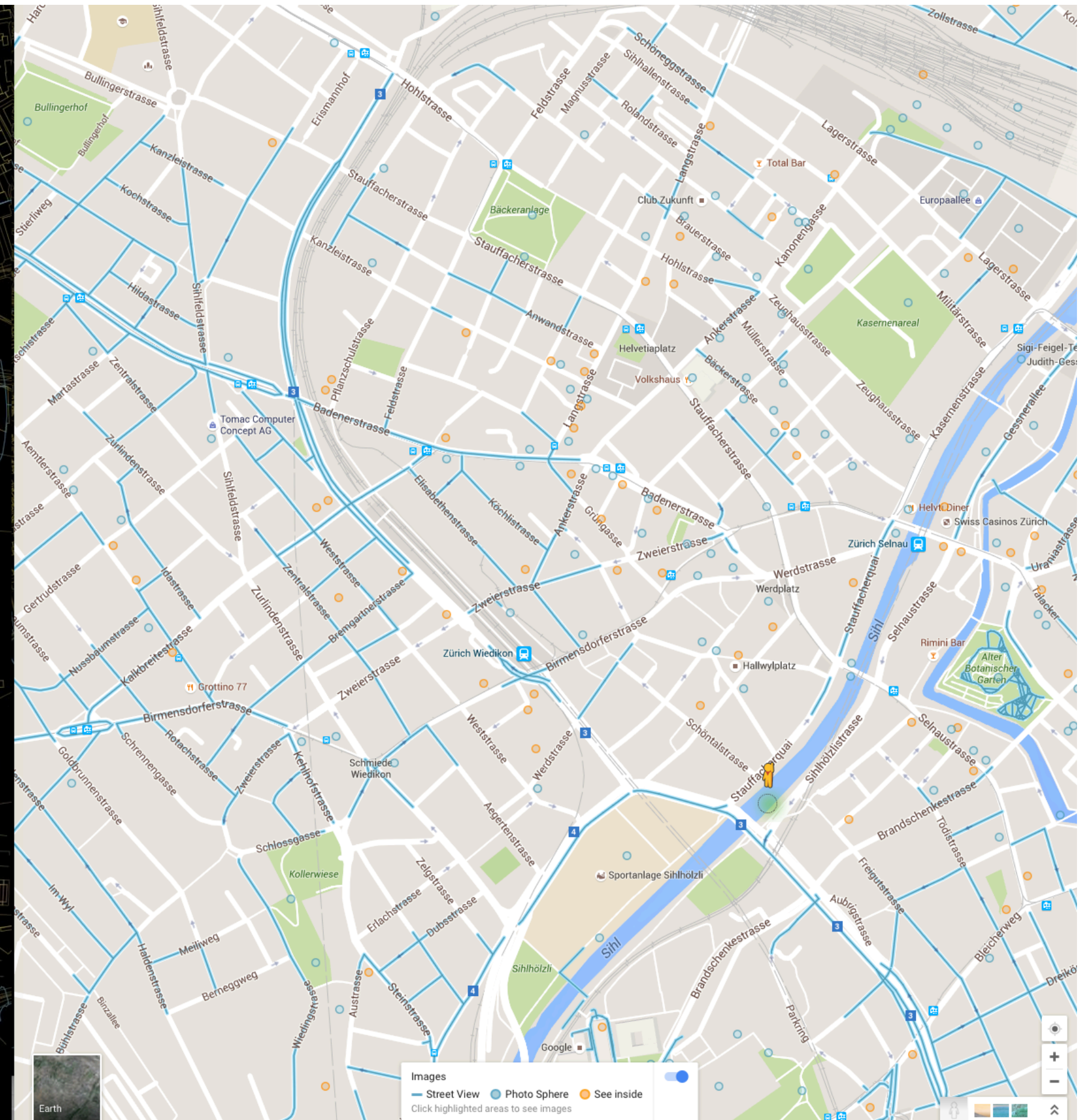
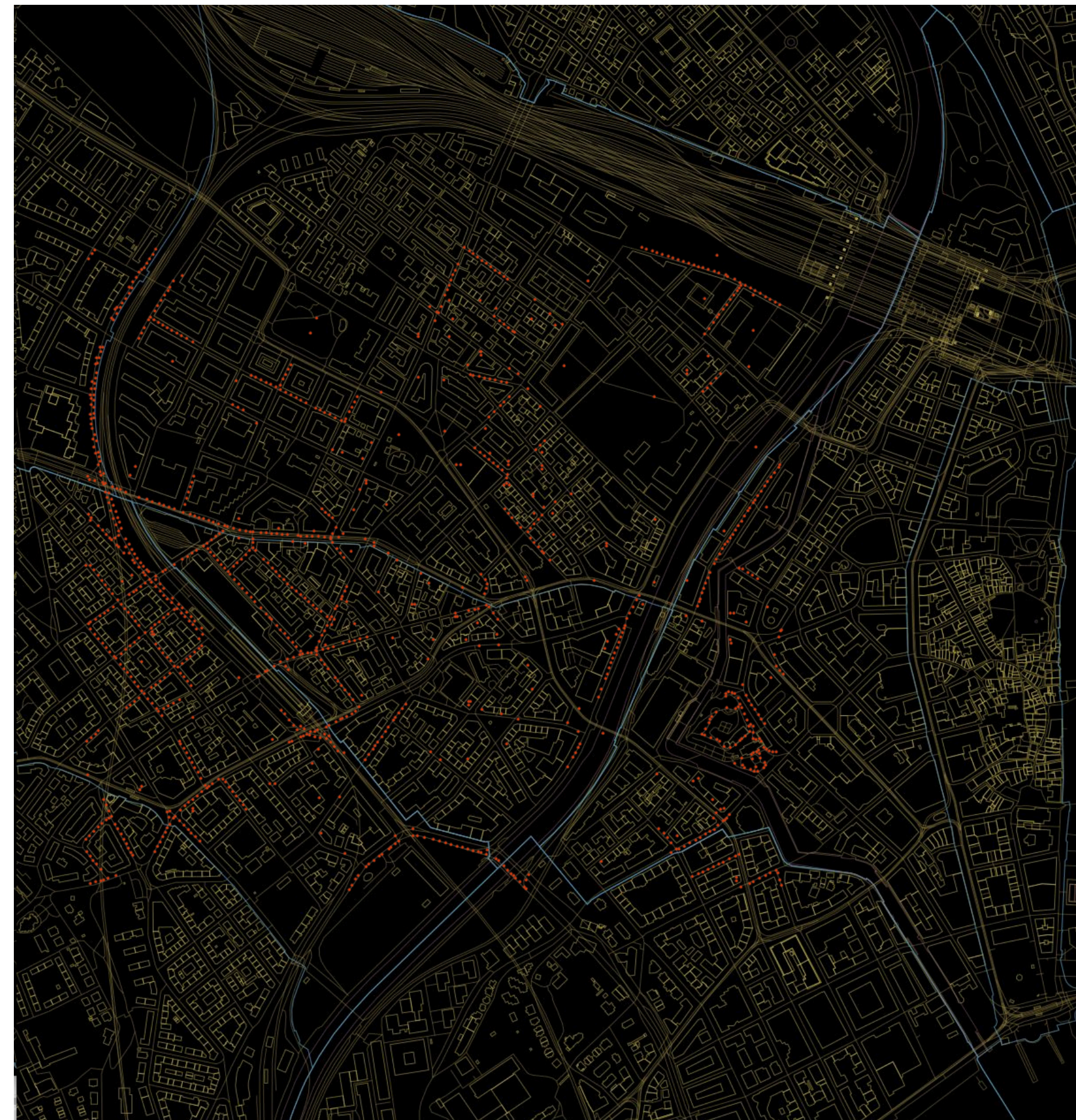


Image extraction from Googlemaps Streetview

Geo-referenced data extraction and visualization

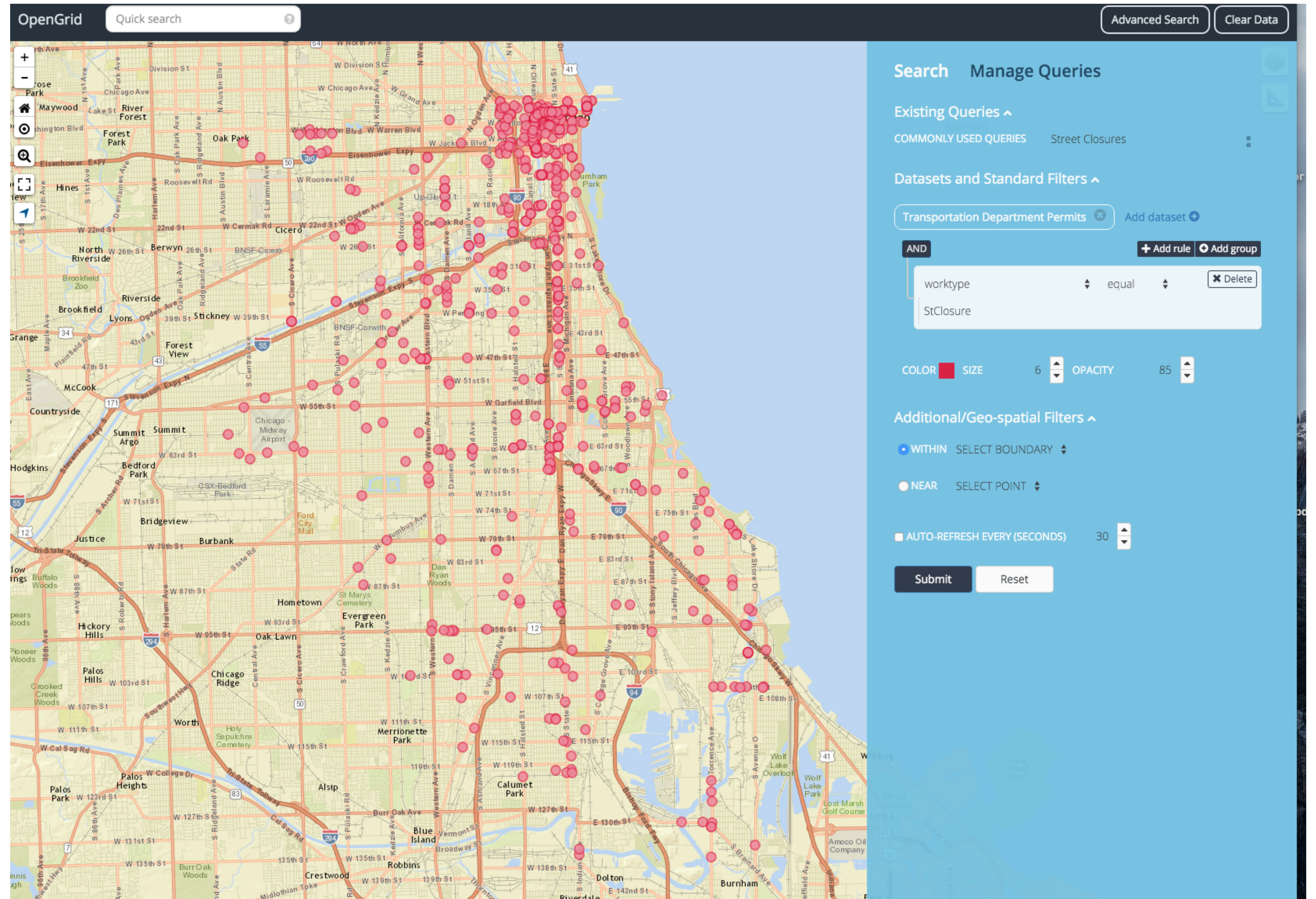
		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		



Tryer L. LUCI application at the Chair of Information Architecture

Chicago OpenGrid

Geo-referenced data visualization

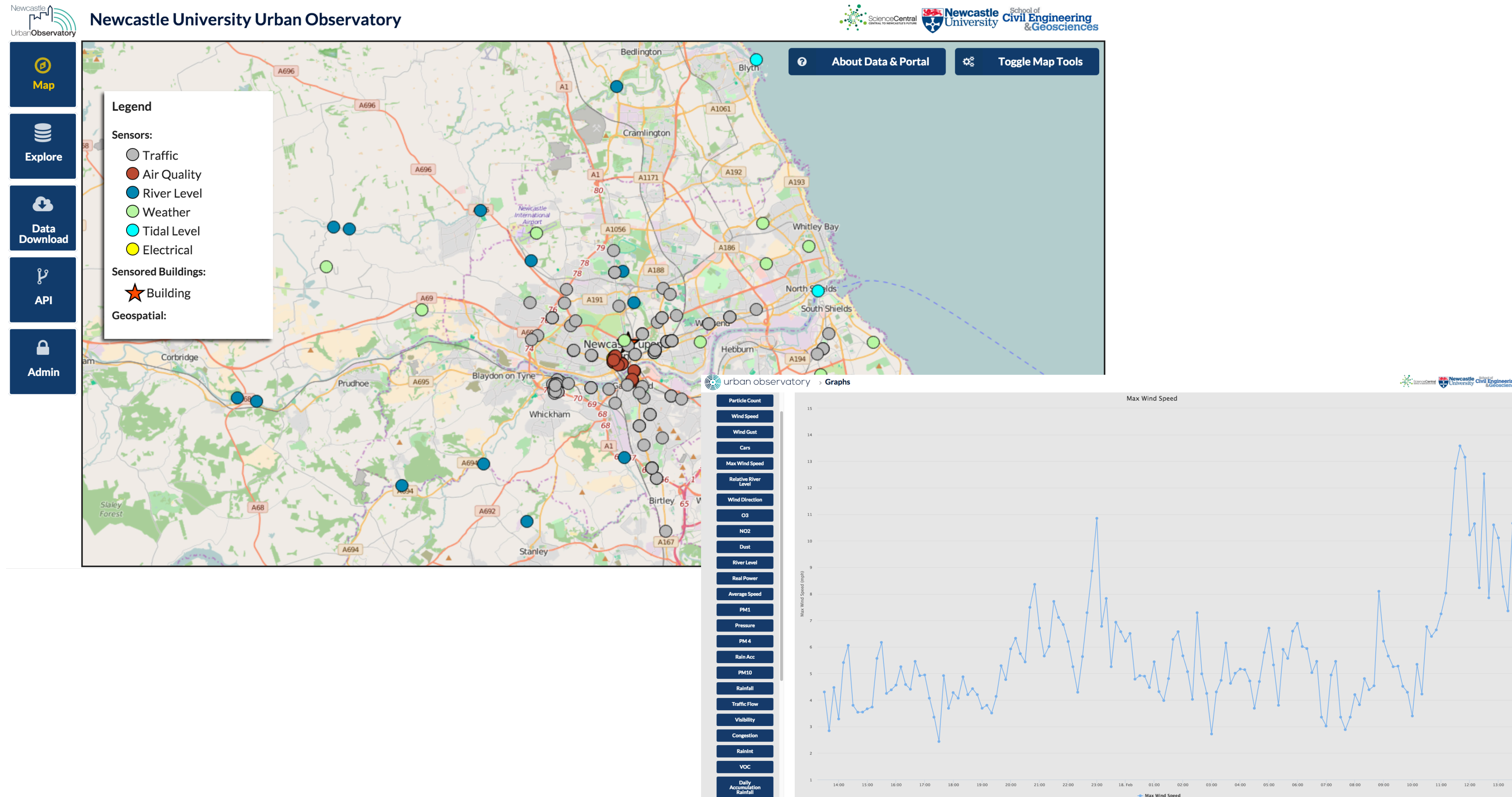


<http://chicago.opengrid.io/opengrid/#>

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

Newcastle University Urban Observatory

Geo-referenced and time-series data visualization

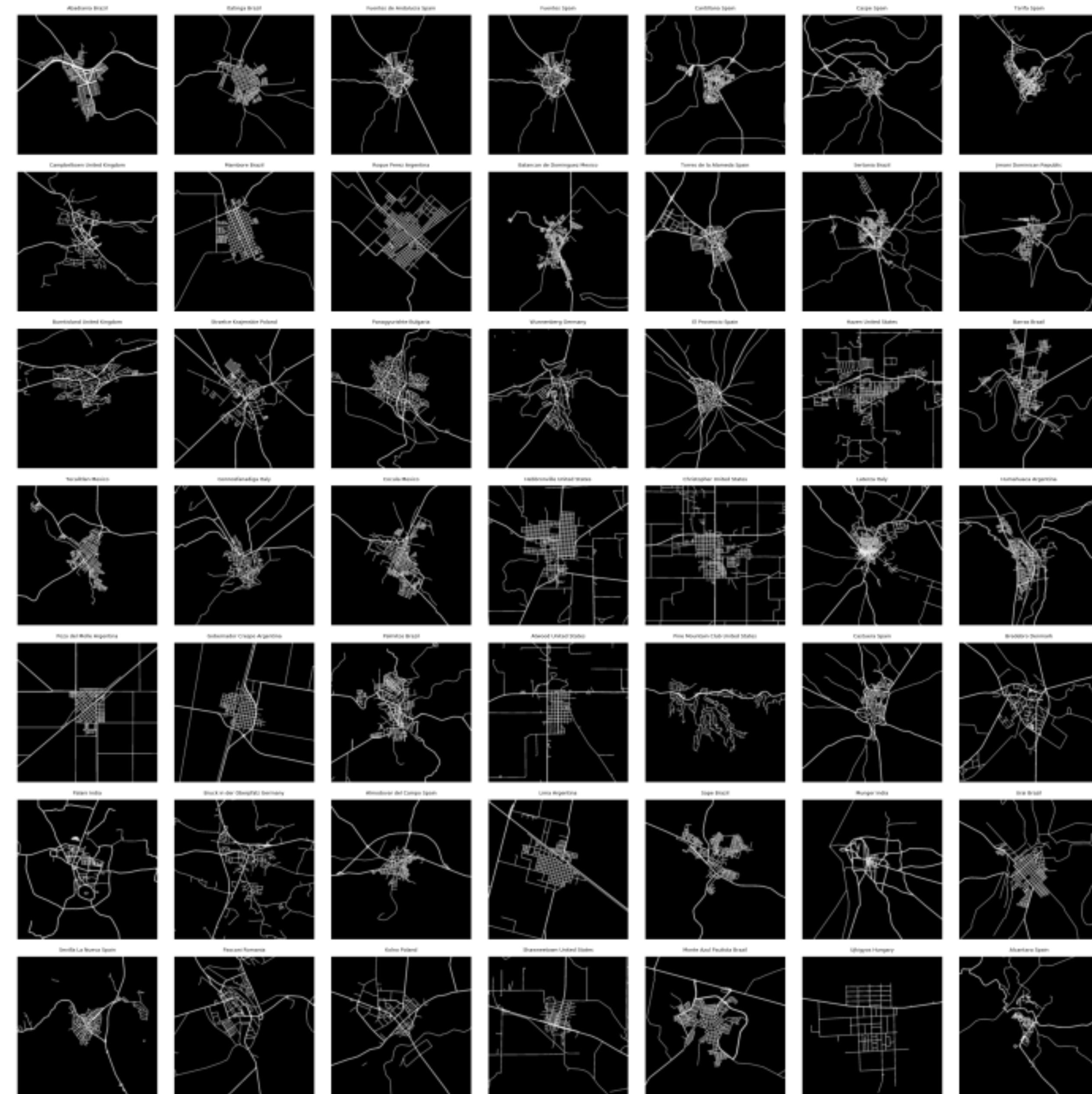
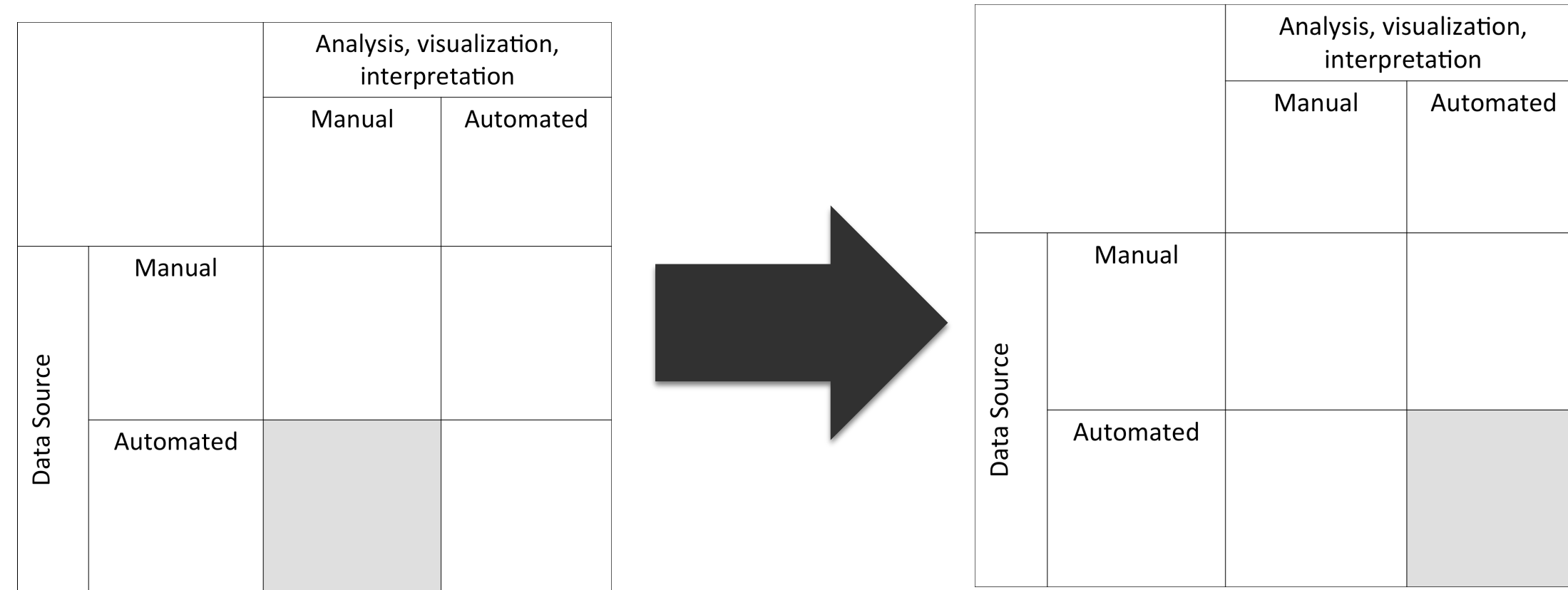


		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

<http://uoweb1.ncl.ac.uk/>

Urban Morphology meets big data

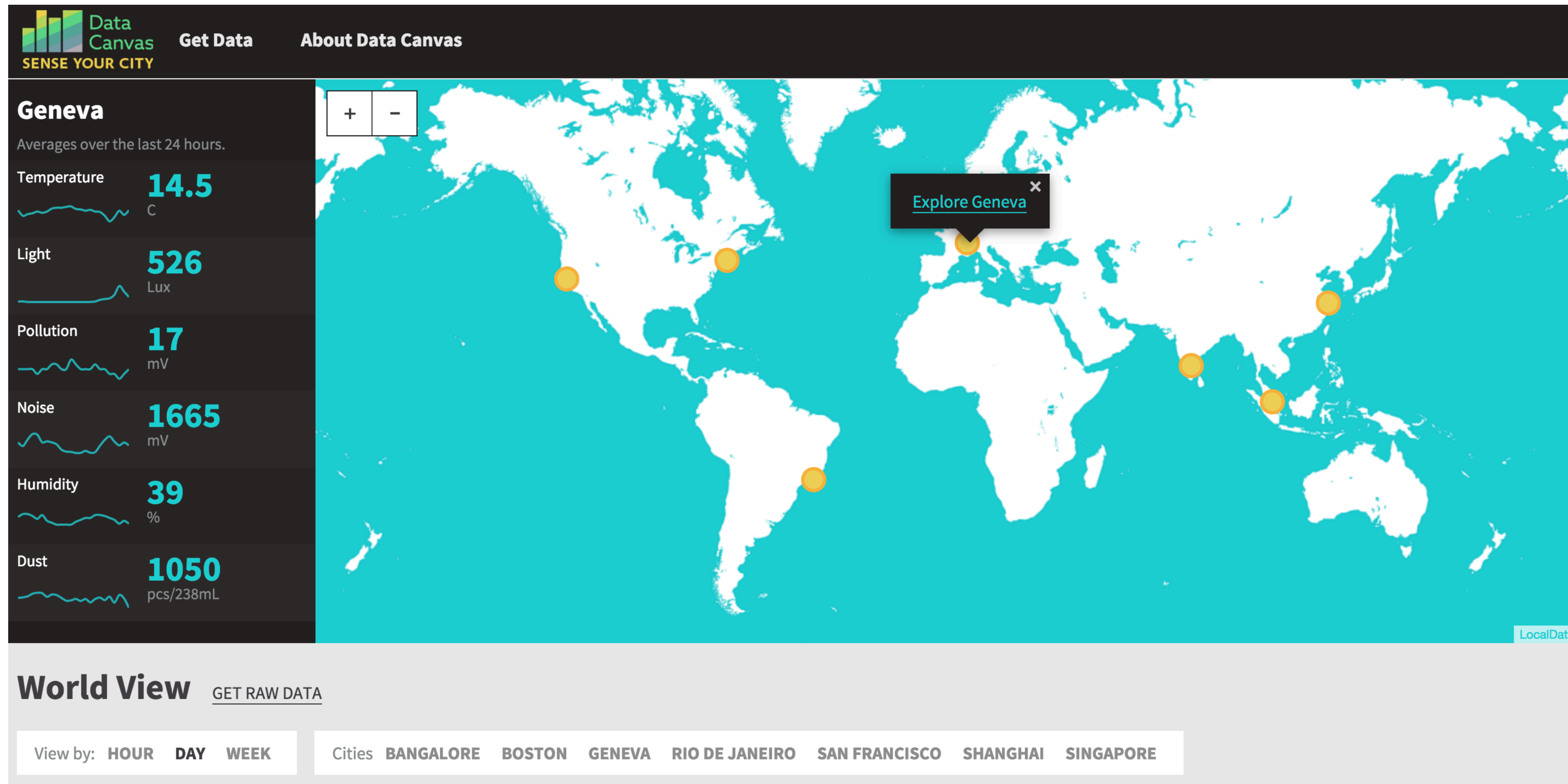
Urban network classification using nearest neighbor clustering



<https://vahidmoosavi.com/2017/01/20/gitpitch-sevamooroadsarereadmaster/>

Data canvas project: Sense your city

Geo-referenced and time-series data visualization



		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

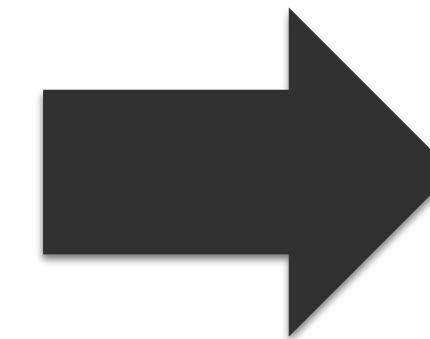
<http://datacanvas.org/sense-your-city/>

Data Canvas project output

Nearest neighbor clustering with images and time-series/geo-referenced weather



		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		



		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

Data driven buildings

Clustering and anomaly detection

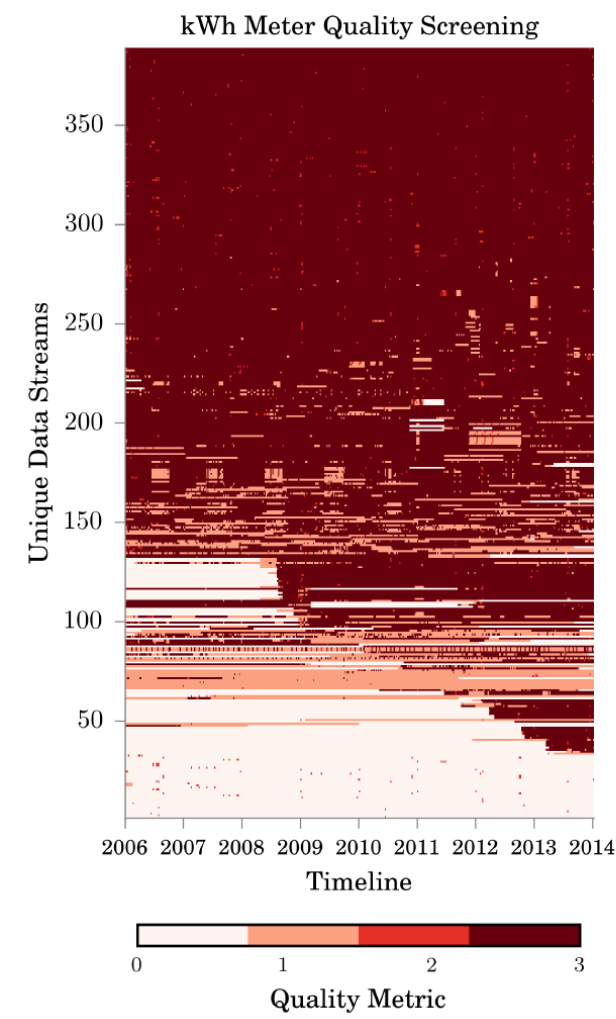


Figure 6. Data quality metrics map for campus sorted (bottom-to-top) according to increasing quality metric

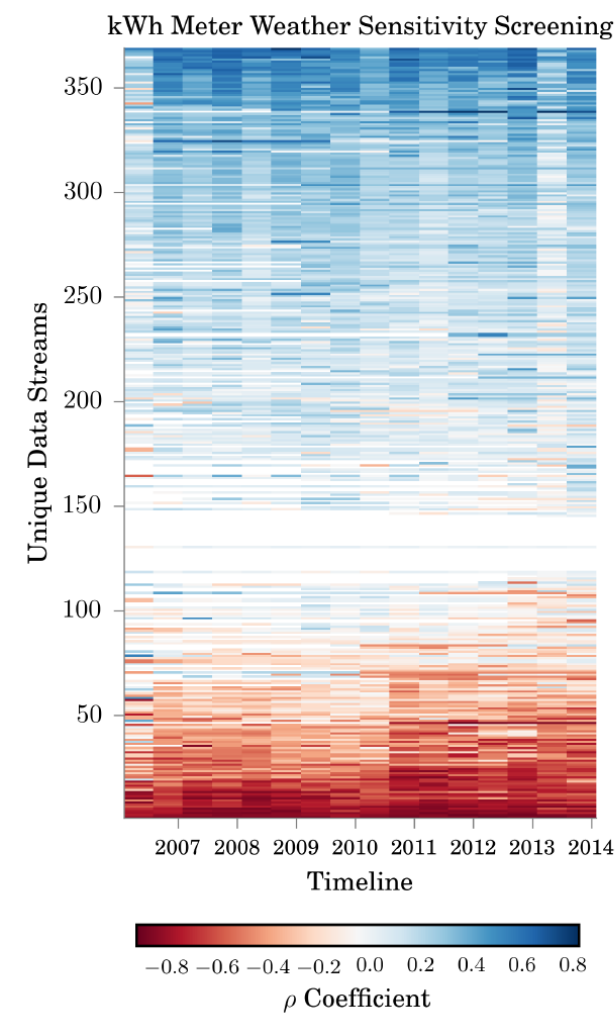


Figure 7. Weather sensitivity map sorted (bottom-to-top) from high negative to high positive ρ coefficient values

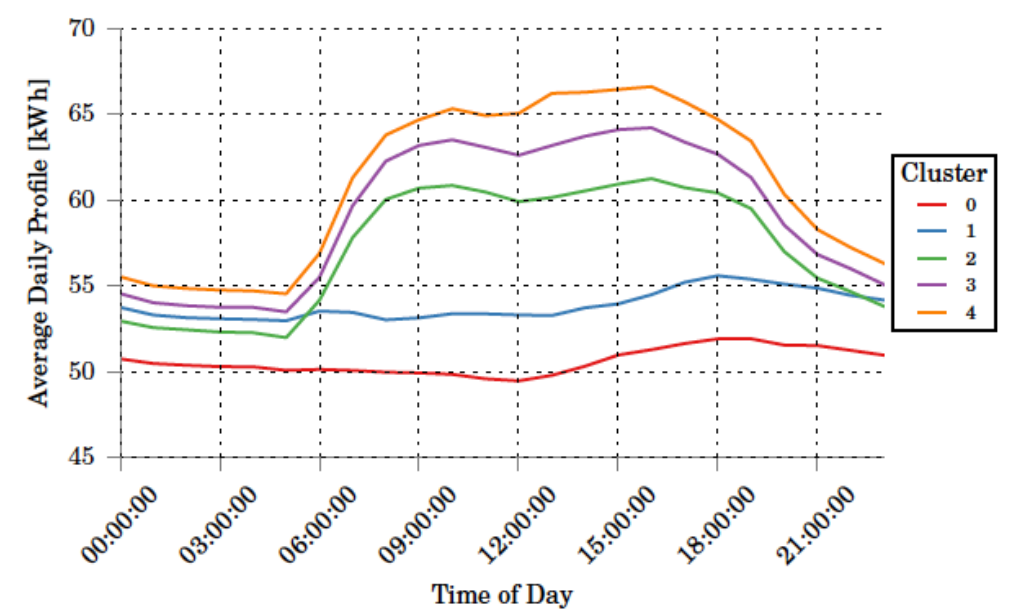


Figure 10. Example of a performance cluster profiles created by the typical profile creation process for a selected kWh meter from Region 4

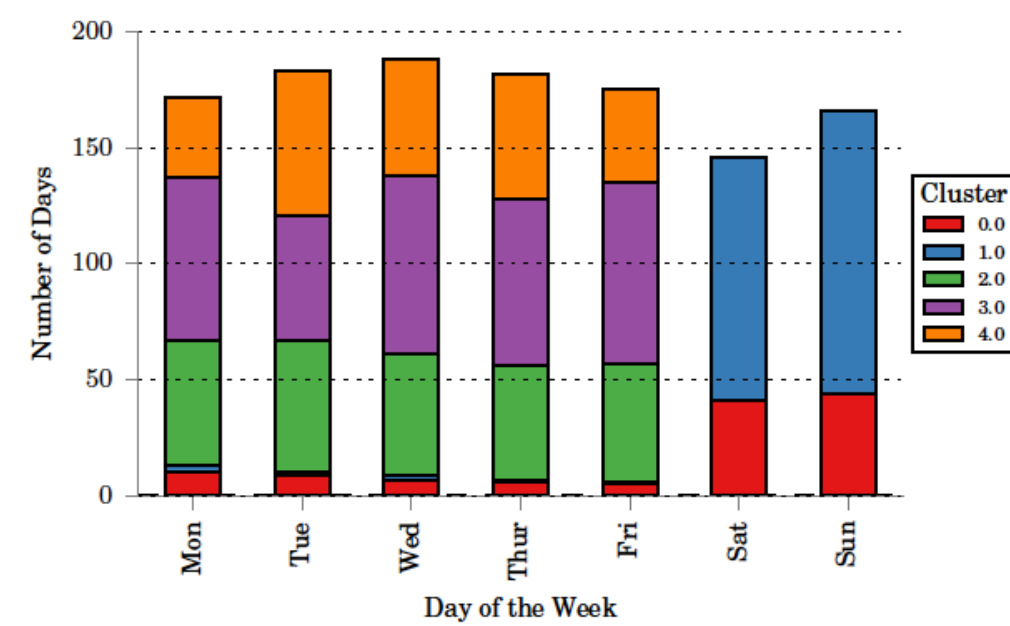
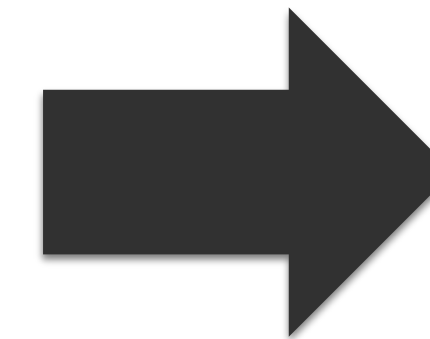


Figure 12. Number of instances of each performance cluster across the days of the week

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		



		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

Miller C., & Schlueter A. (2015, April). Forensically Discovering Simulation Feedback Knowledge from a Campus Energy Information System. In Proceedings of the *Symposium on Simulation for Architecture and Urban Design (SimAUD)*. (pp. 136-143). Society for Computer Simulation International.

Other Examples?

		Analysis visualization & interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		



Semester Project

Something to start thinking about

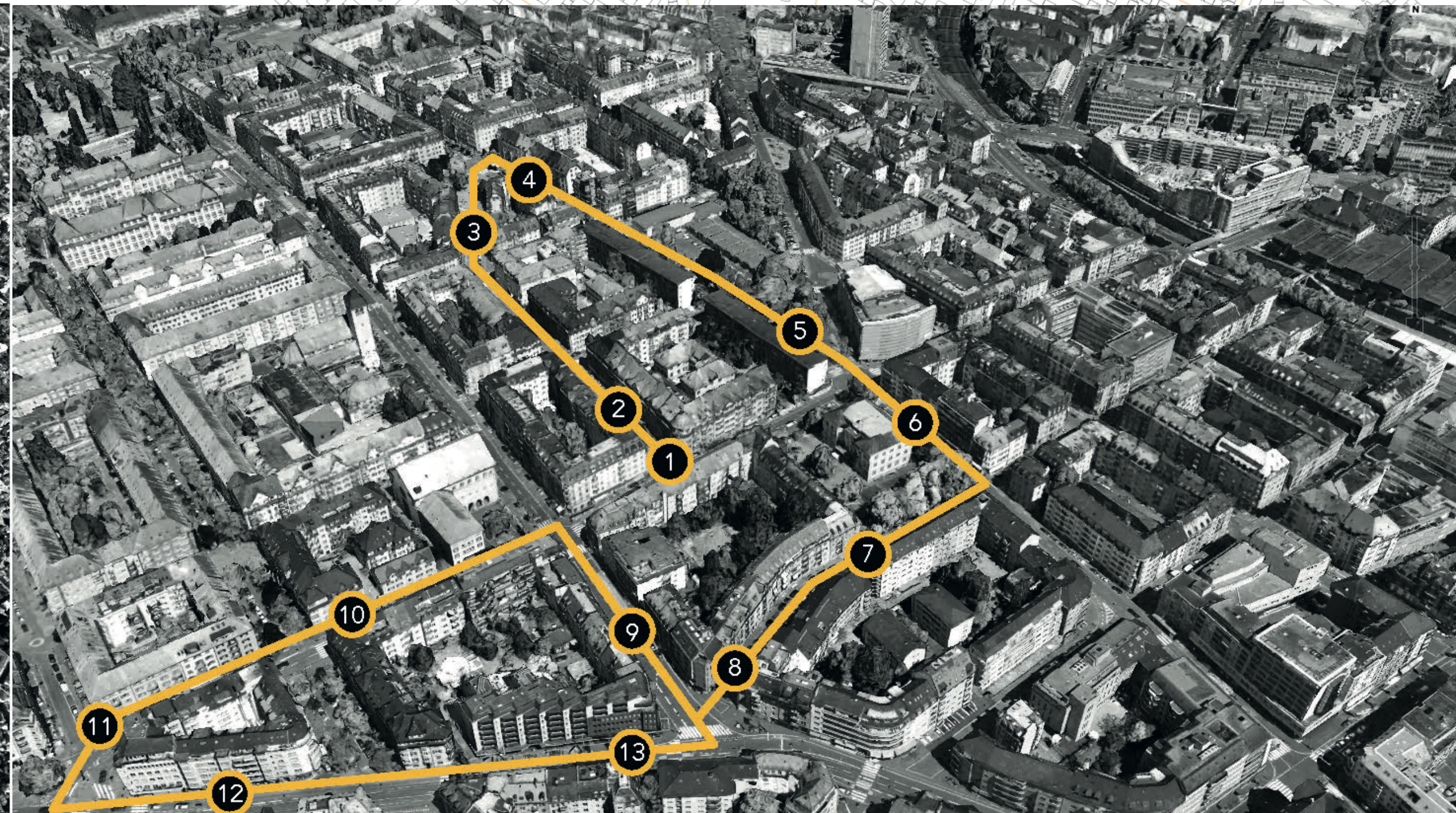
1. Formulate 1-2 specific question(s) of interest to you
2. State your hypothesis/expected outcome based on supporting literature (minimum one source) your expertise, and intuition
3. Answer that question through your analysis, for this:
 - Select the best available data sources for your question (min. of 2 data sources)
 - Include a time series and/or clustering analysis
4. Summarize your results
 - Show a clear conclusion, does your analysis answer your question(s)?
5. Conclusions & lessons learned
6. Include motivation and references

Semester Project

Potential Case Study: ESUM- Analyzing trade-offs between Energy and Social performance of Urban Morphologies



Location Wiedikon Zürich



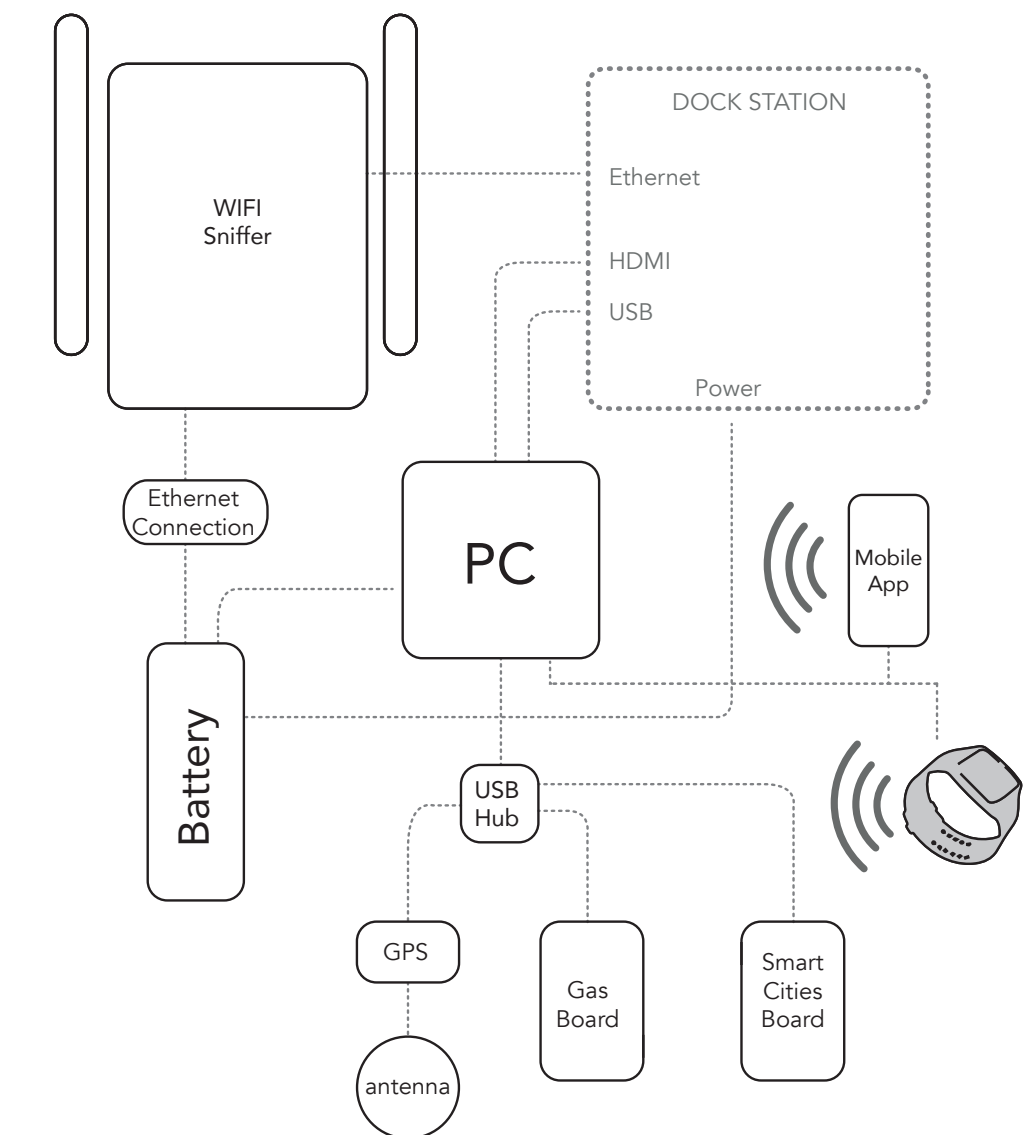
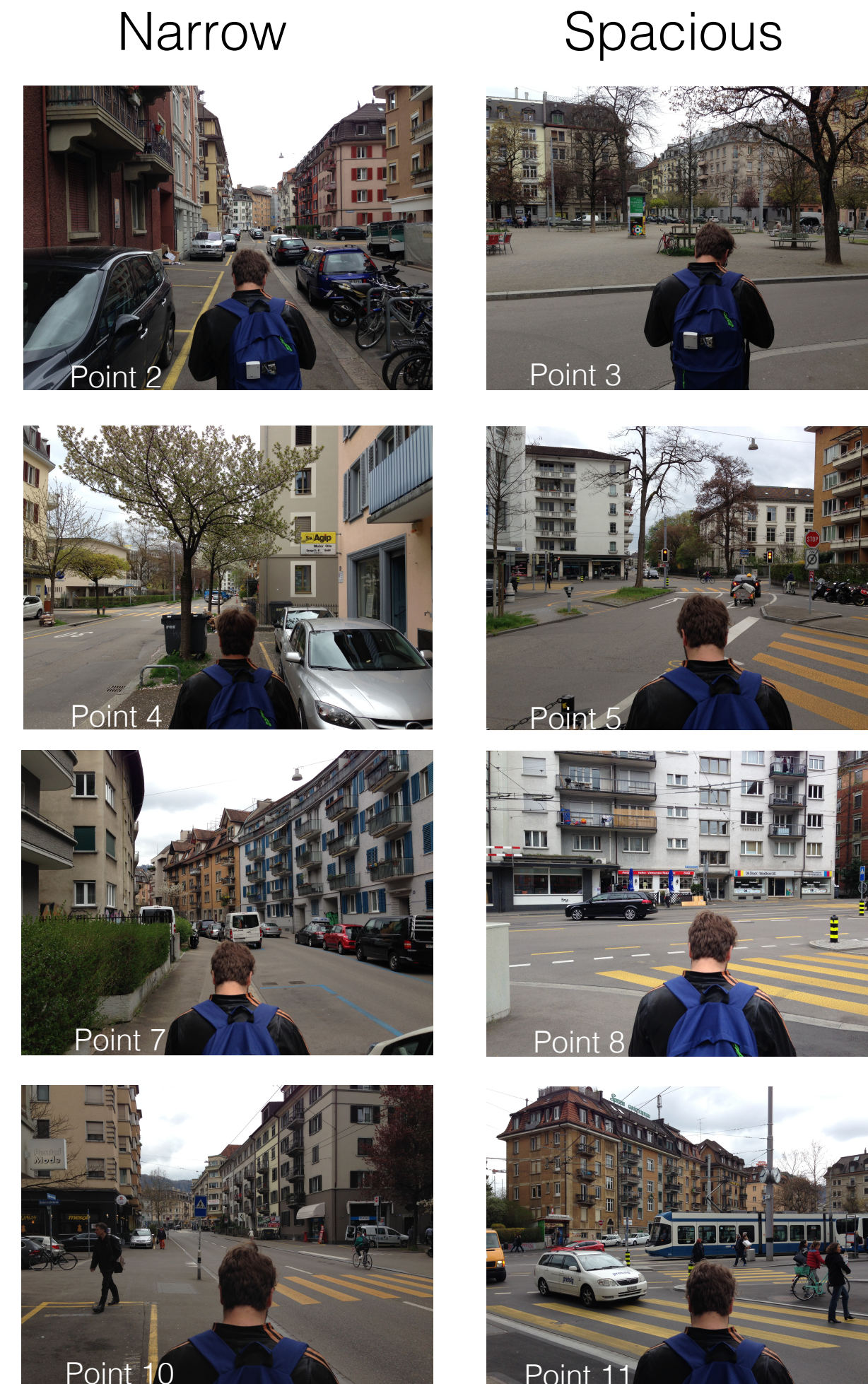
14 survey checkpoints along experimental path

Semester Project

Potential Case Study: ESUM- Analyzing trade-offs between Energy and Social performance of Urban Morphologies

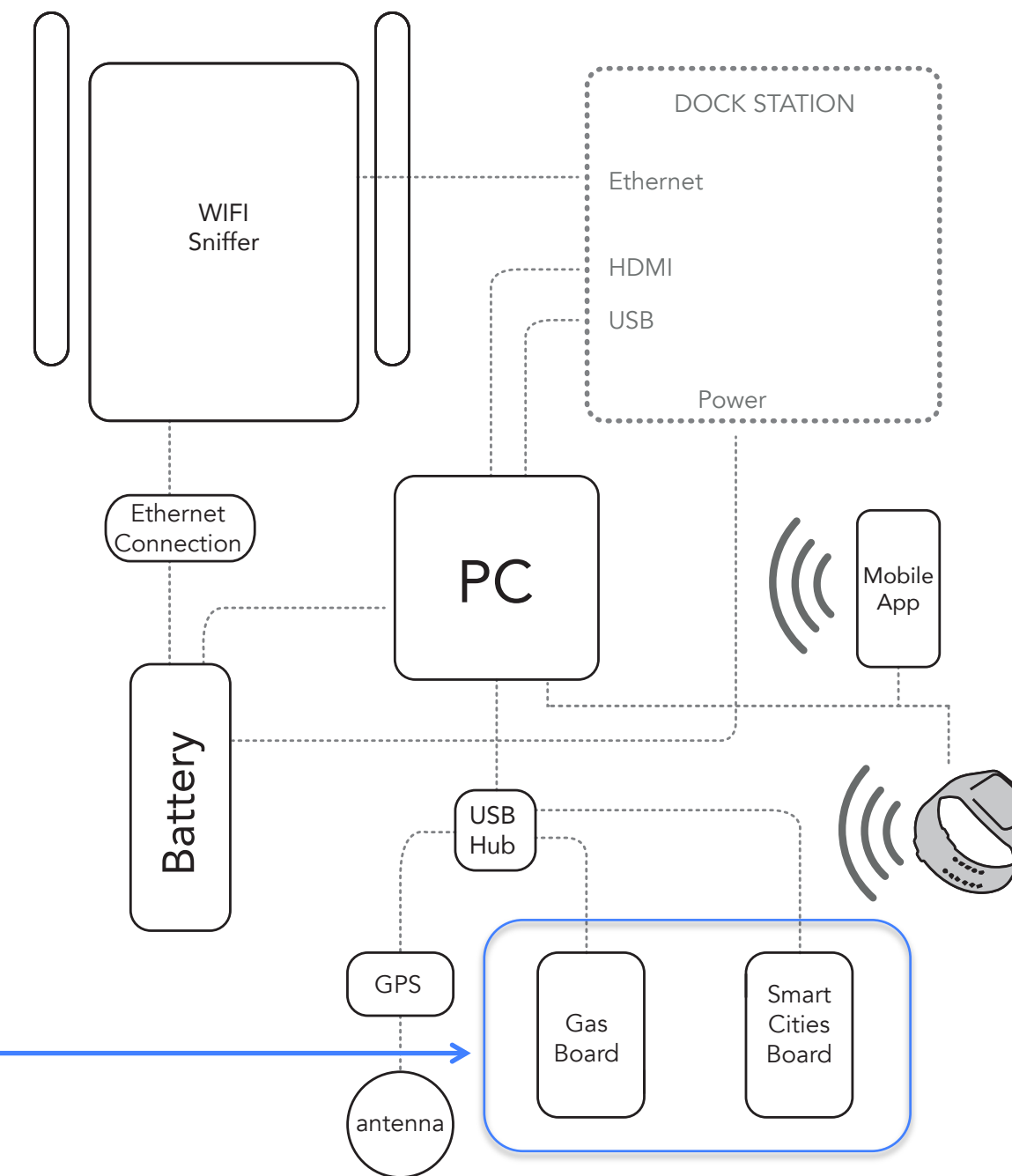
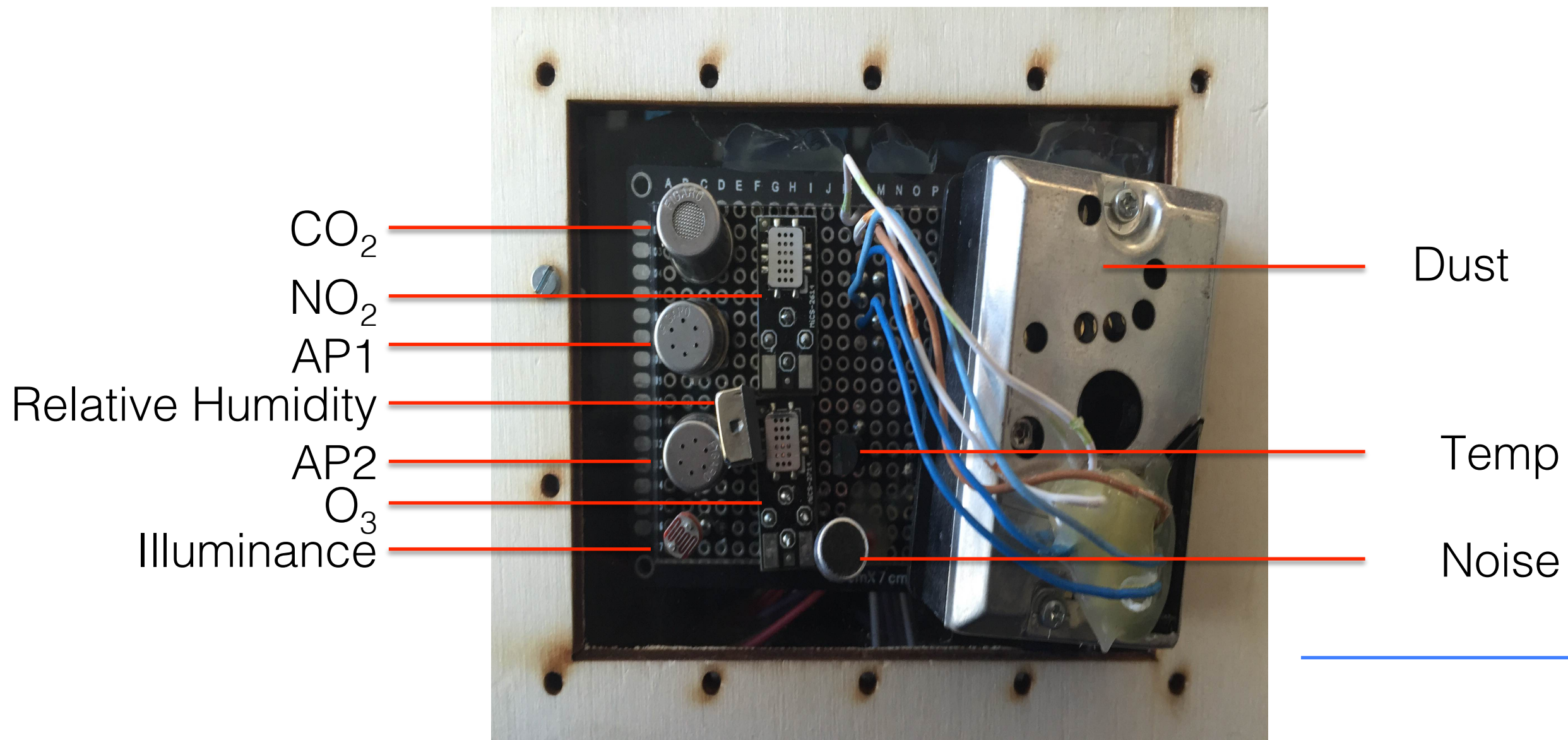
Use data set from a case study of 37 participants in Zurich to:

- Investigate impact of static (urban morphology) and dynamic features (environmental sensors) of the built environment on perception (using surveys and biofeedback data)



Mobile sensor equipment

Sensorbackpack with environmental and position sensors



Mobile Sensor equipment

Biofeedback wristband

E4 Sensors



PPG Sensor

Photoplethysmography Sensor - Measures Blood Volume Pulse (BVP), from which heart rate, heart rate variability (HRV), and other cardiovascular features may be derived



3-axis Accelerometer

Captures motion-based activity



Event Mark Button

Tags events and correlate them with physiological signals



EDA Sensor (GSR Sensor)

Electrodermal Activity Sensor - Used to measure sympathetic nervous system arousal and to derive features related to stress, engagement, and excitement.



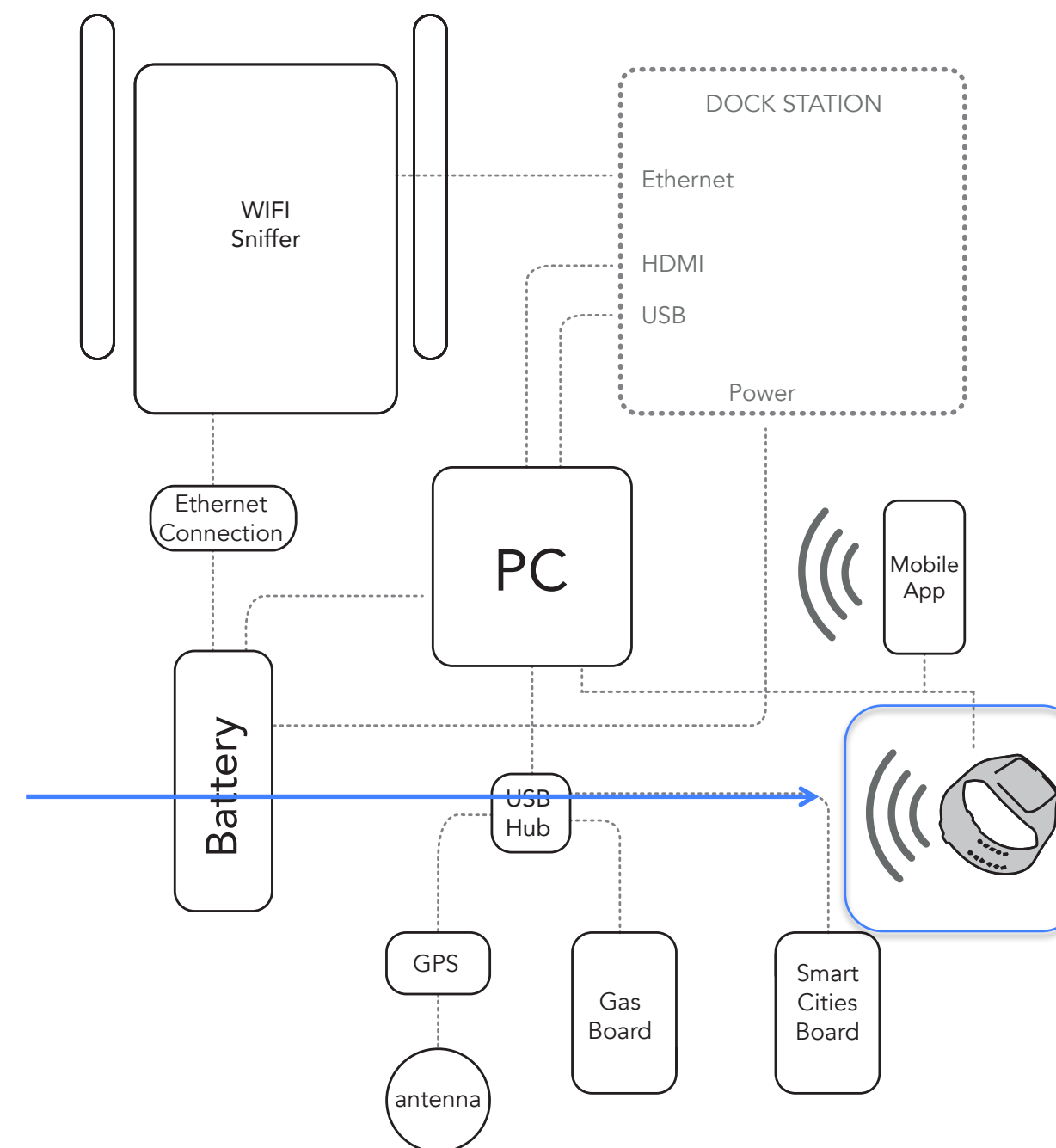
Infrared Thermopile

Reads peripheral skin temperature



Internal Real-Time Clock

Temporal resolution up to 0.2 seconds in streaming mode



<https://www.empatica.com/e4-wristband>

Mobile Sensor equipment

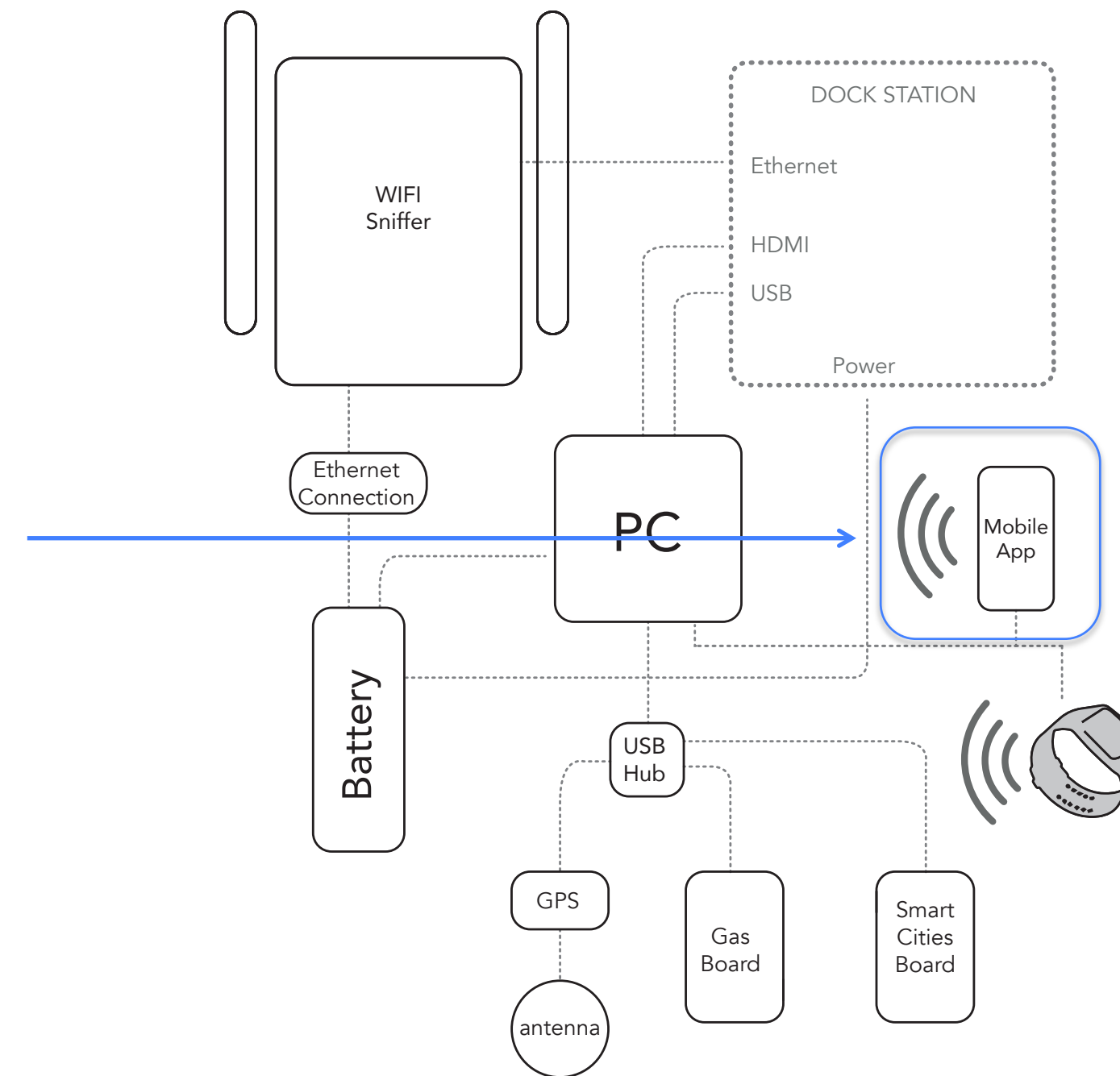
Biofeedback wristband

Checkpoint ID:.....
 Atmosphere

dislike like

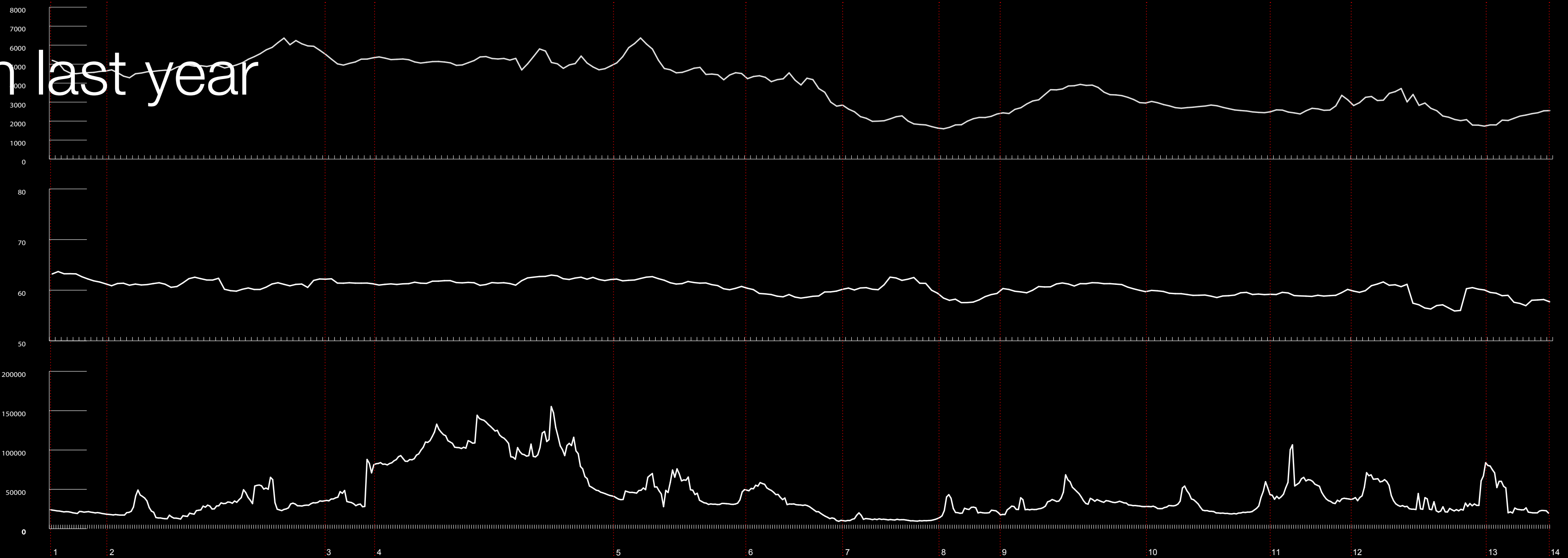
chaotic	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	ordered
noisy	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	quiet
private	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	public
boring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	interesting
crowded	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	empty
insecure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	secure
ugly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	beautiful
narrow	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	spacious
enclosed	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	open
dark	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	light

Unfamiliar Familiar





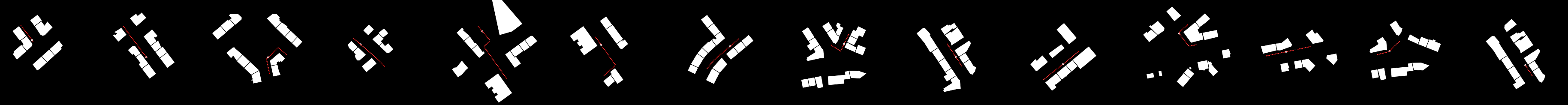
One example from last year



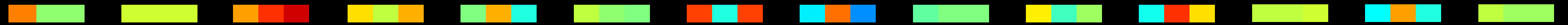
SURVEY POINT ONE SURVEY POINT TWO SURVEY POINT THREE SURVEY POINT FOUR SURVEY POINT FIVE SURVEY POINT SIX SURVEY POINT SEVEN SURVEY POINT EIGHT SURVEY POINT NINE SURVEY POINT TEN SURVEY POINT ELEVEN SURVEY POINT TWELVE SURVEY POINT THIRTEEN SURVEY POINT FOURTEEN



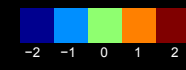
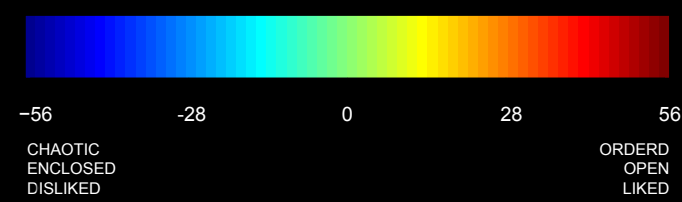
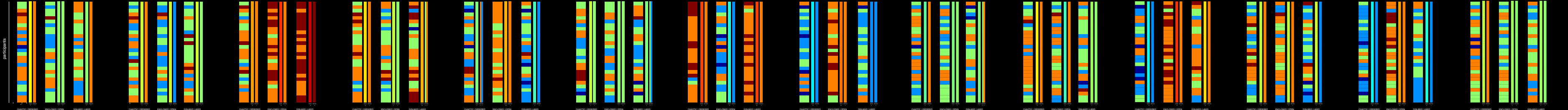
SURVEY POINT TYPOLOGIES



STRUCTURE | SPACIOUSNESS | PREFERENCE



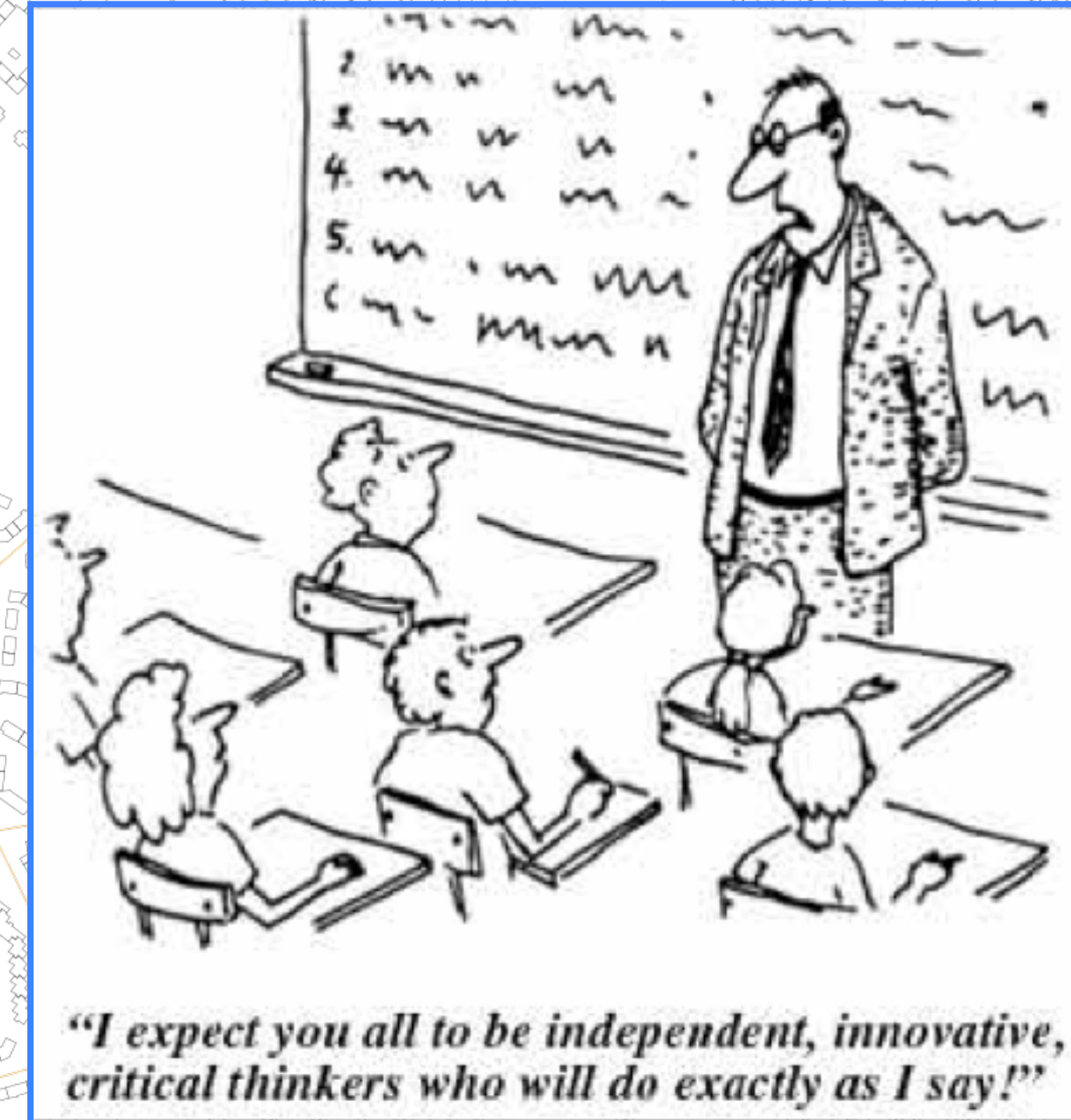
SURVEY POINT ONE SURVEY POINT TWO SURVEY POINT THREE SURVEY POINT FOUR SURVEY POINT FIVE SURVEY POINT SIX SURVEY POINT SEVEN SURVEY POINT EIGHT SURVEY POINT NINE SURVEY POINT TEN SURVEY POINT ELEVEN SURVEY POINT TWELVE SURVEY POINT THIRTEEN SURVEY POINT FOURTEEN



Learning objectives

We encourage you to be creative!

1. Become familiar with programming and integrating new tools in your work
2. Come up with an interesting “research” question and learn how to answer it by:
 - Selecting appropriate data source(s)
 - Applying the relevant analysis and visualization techniques
 - Interpreting and refining your results



Course Schedule

What to Expect

Mondays 10:00 - 12:00
051-0726-17U | 2 ECTS*

Creative Data Mining Uncover and Evaluate

The participants of this course learn how to collect, process, analyze and interpret real spatial and temporal data in order to work with quantifiable qualities in urban planning. This is achieved by using actual data from a recent study and analysing it with different data processing and machine learning techniques.

The goal of the course is to explore a specific research question about the urban environment and test the stated hypothesis using different techniques presented in the course, thus preparing students with a skill-set to further support their design and decision making processes.

The course focuses on creating deeper insights to critically evaluate design alternatives for urban planning projects. Students will work with time-series and geo-referenced data including temperature, relative humidity, illuminance, noise, people density, and dust particulate matter. Subjective impression survey data will also be integrated into the student projects to further explore influencing factors of the urban environment on our perceptual experiences. Non-architectural skills the participants can develop during this course are 1) an introduction to programming 2) how clustering methods like PCA or K-Means could be applied in an architectural context.

Where
HIT H 34.1 (Video Wall)

Supervision
Danielle Griego griego@arch.ethz.ch
Daniel Zünd zuend@arch.ethz.ch
Artem Chirkin chirkin@arch.ethz.ch

Prof. Dr. Gerhard Schmitt
Chair of Information Architecture
Information Science Lab
Wolfgang-Pauli-Strasse 27, 8093 Zurich
www.ia.arch.ethz.ch

- 20.02.2017 **Course Introduction**
- 27.02.2017 **Introduction to Python & Programming I**
- 06.03.2017 **Introduction to Python & Programming II**
- 13.03.2017 **Data Processing**
- 20.03.2017 **Seminar week (No lecture)**
- 27.03.2017 **Intro to time-series data analysis**
- 03.04.2017 **Time series data analysis ctd. & Machine learning**
- 10.04.2017 **Machine learning ctd.**
- 17.04.2017 **Holiday (No lecture)**
- 24.04.2017 **Programming tutorial applications**
- 01.05.2017 **Holiday (No lecture)**
- 08.05.2017 **Q&A Feedback Workshop I**
- 15.05.2017 **Final iA critique**
Combined critique with the other iA courses
(13:00 - 18:00)

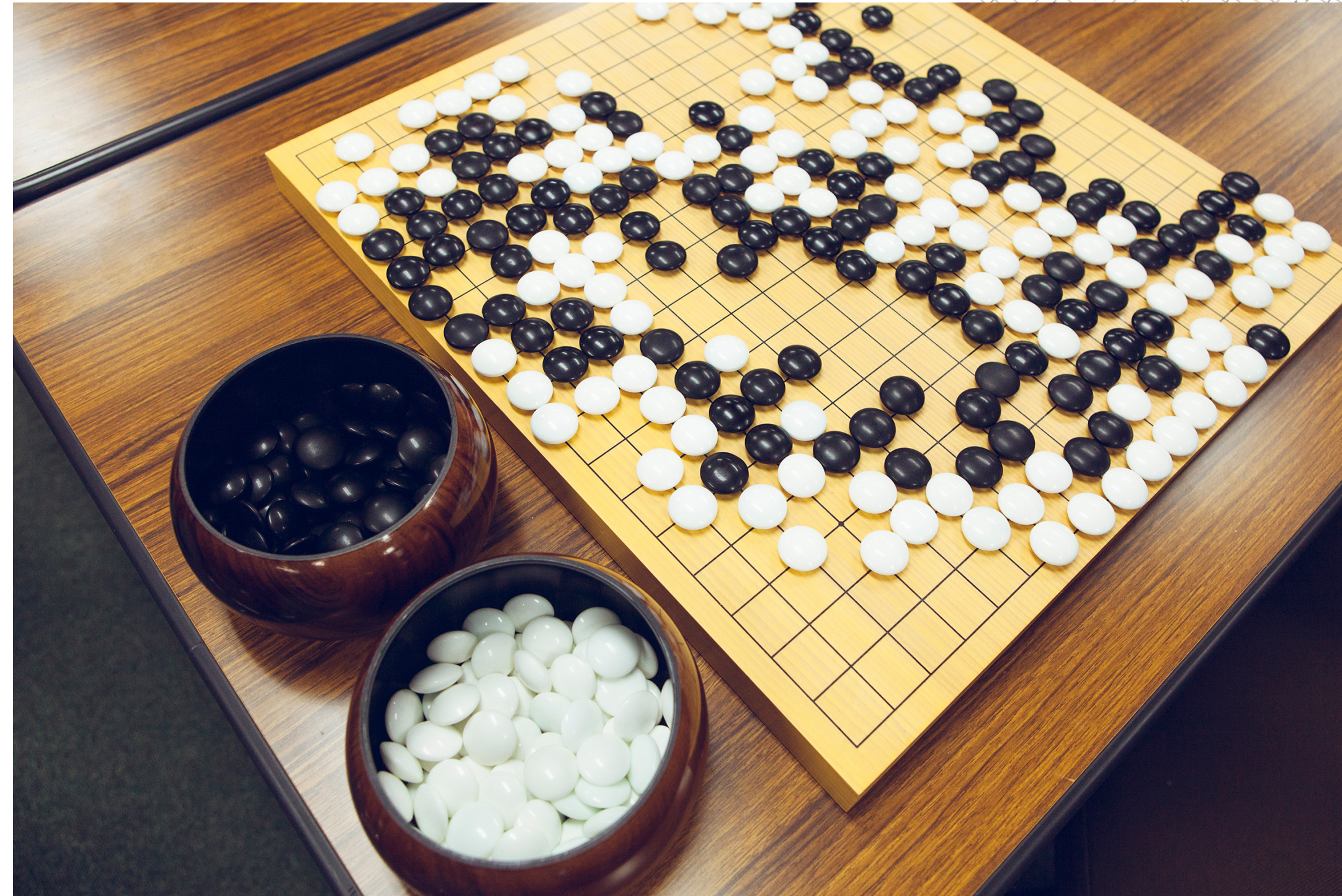
Requirement Former knowledge of any digital tool or coding language is most welcome but NOT required. You only need to provide a reasonable amount of motivation and of course a notebook.

* Total 60 h = 2 ECTS
Ungraded Semester Performance

The most recent outline will be found on www.ia.arch.ethz.ch

Short discussion

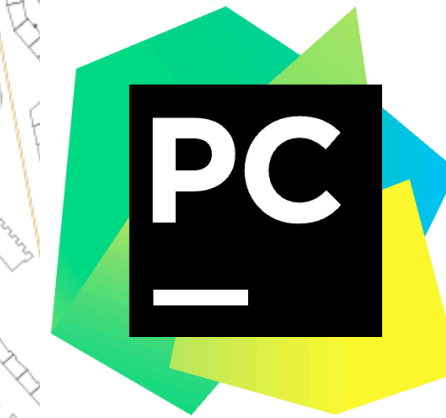
Your expectations?



Homework

You can stick around and install the programs now if you'd like

1. Install Python from <https://www.python.org/downloads/>
2. Install **PyCharm** from <https://www.jetbrains.com/pycharm/download/>
3. Research other examples of urban data mining and make 2 slides about the most interesting project/application/research group(s) that you find. This will be presented at the beginning of next lecture



Resources for the course

Course Material Posted to:

- <http://www.ia.arch.ethz.ch/category/fs2017-creative-data-mining/>

Tutorials:

- https://www.tutorialspoint.com/python/python_basic_operators.htm
- <http://www.informatics.indiana.edu/rocha/academics/i-bic/lab1/Python%20review.pdf>

References:

- “A Byte of Python” <https://python.swaroopch.com/>
- Coelho, Luis Pedro; Richard, Will. Building Machine Learning Systems with Python, Packt Publishing (Adobe Editions Library)

“Science without philosophy is blind, and philosophy without science is paralyzed”
(Paul Cilliers, Complexity and Postmodernism)

THANK YOU

?

Lecture 1: Introduction

20 | 02 | 2017

Dr. Daniel Zünd, zuend@arch.ethz.ch

Danielle Griego, griego@arch.ethz.ch

Artem Chirkin, chirkin@arch.ethz.ch