Creative Data Mining

Lecture 01: Introduction
22 February 2016

Danielle Griego, griegod@ethz.ch
Matthias Standfes, standfest@arch.ethz.ch

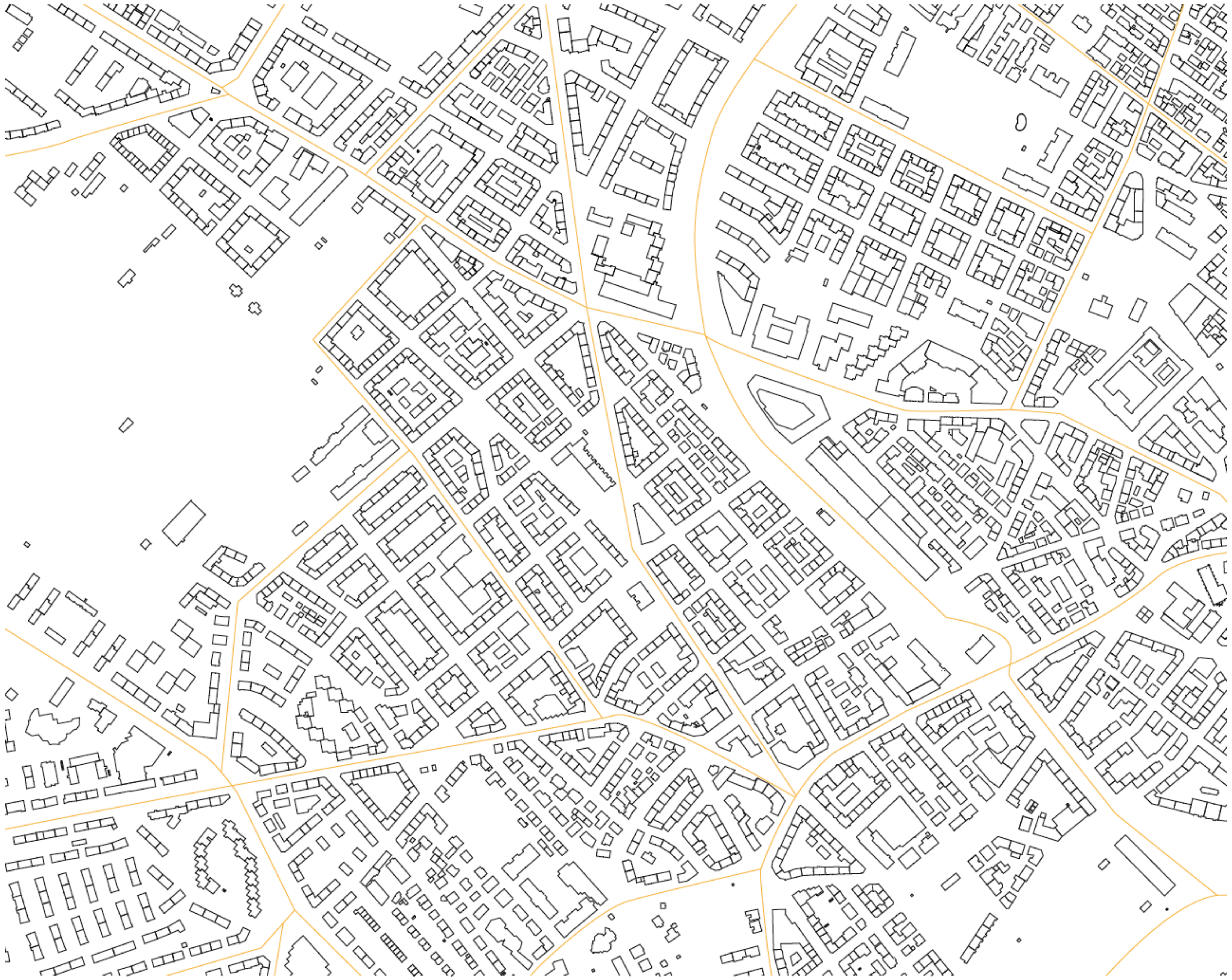
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair of
Information
Architecture

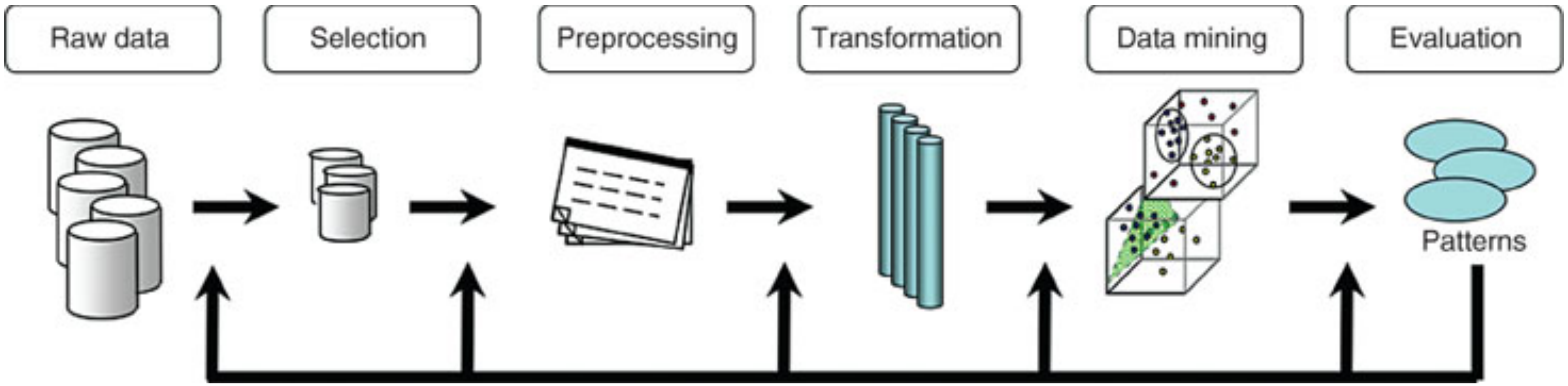
What we'll cover today:

- Course structure & background info
- State of the art for Urban planning
- Related Research at the chair (ESUM)
- Semester project

- Learning objectives
- Course schedule
- Homework/grading
- Discussion
- Install RStudio



What is Data Mining and how can it be Creative?

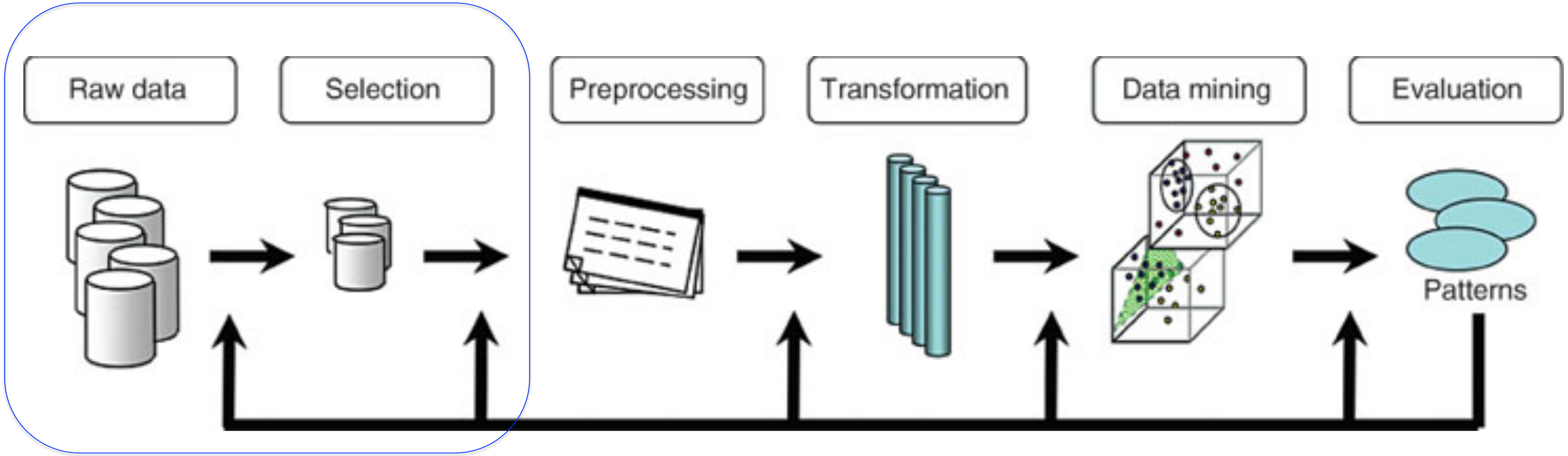


Knowledge Discovery Diagram (KDD)

Encyclopedia of Database Systems, http://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1134

What is Data Mining and how can it be Creative?

Data Source

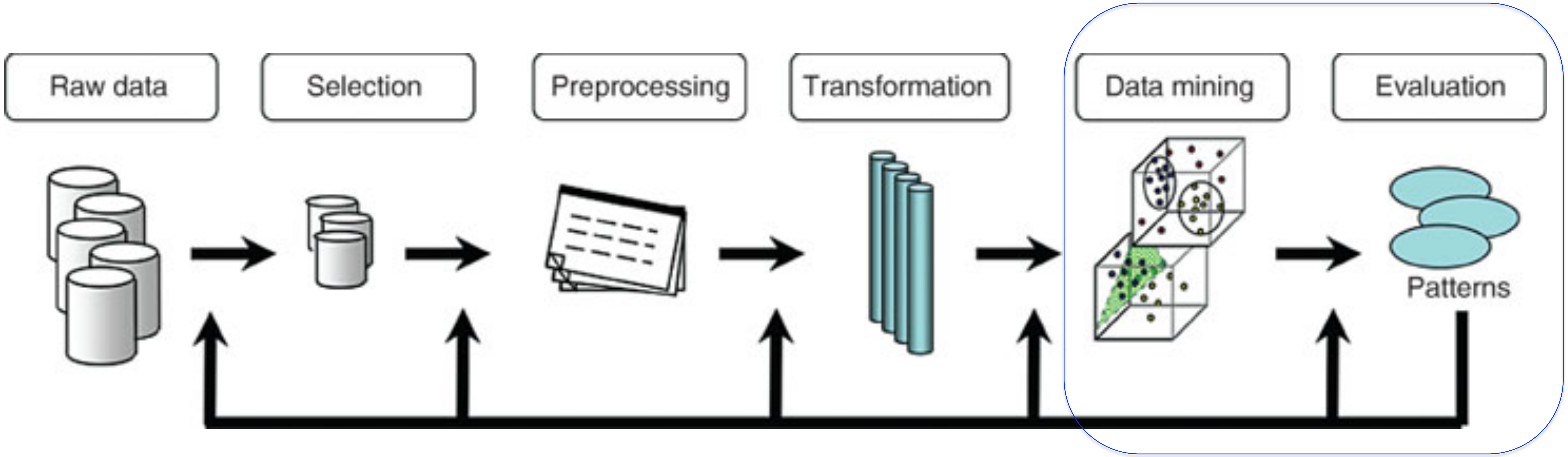


Knowledge Discovery Diagram (KDD)

Encyclopedia of Database Systems,
http://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1134

What is Data Mining and how can it be Creative?

Analysis, visualization, interpretation

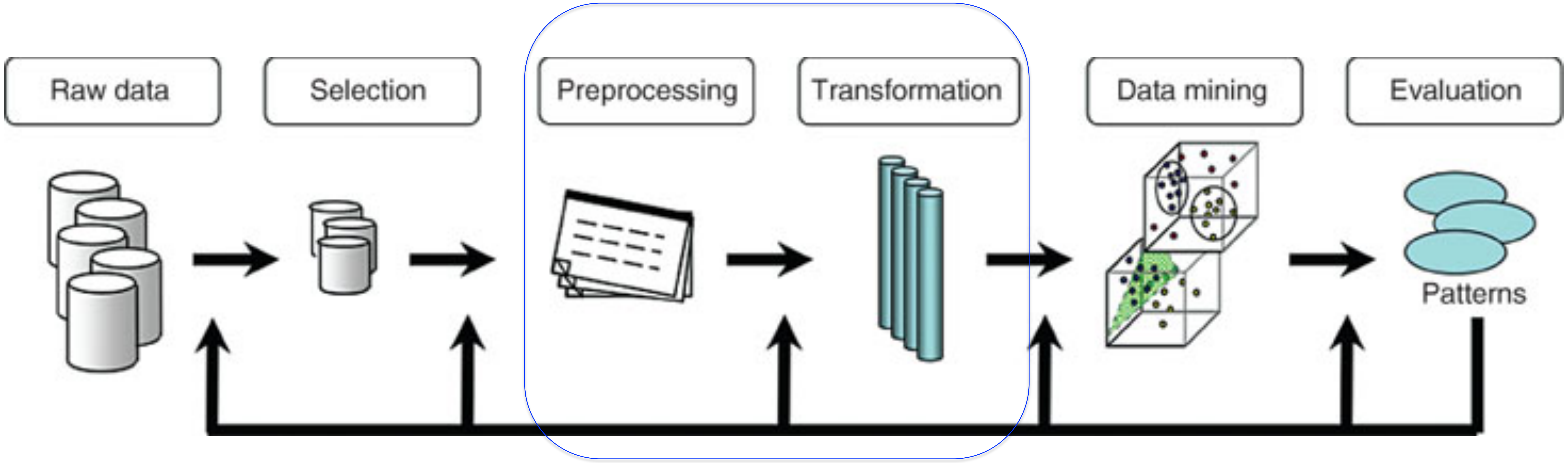


Knowledge Discovery Diagram (KDD)

Encyclopedia of Database Systems,
http://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1134

What is Data Mining and how can it be Creative?

What we'll prepare ahead of time



Knowledge Discovery Diagram (KDD)

Encyclopedia of Database Systems, http://link.springer.com/referenceworkentry/10.1007%2F978-0-387-39940-9_1134

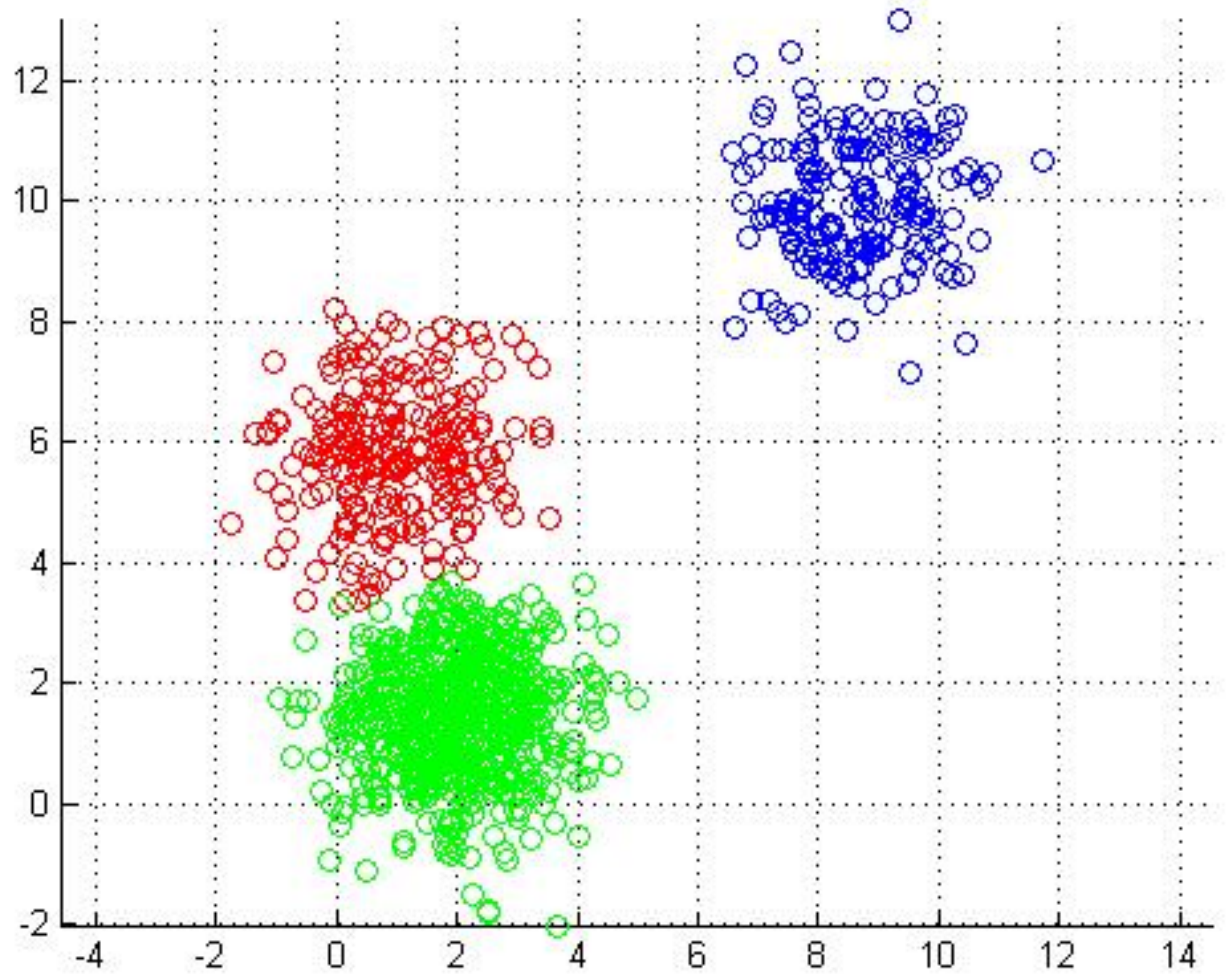
Types of Data

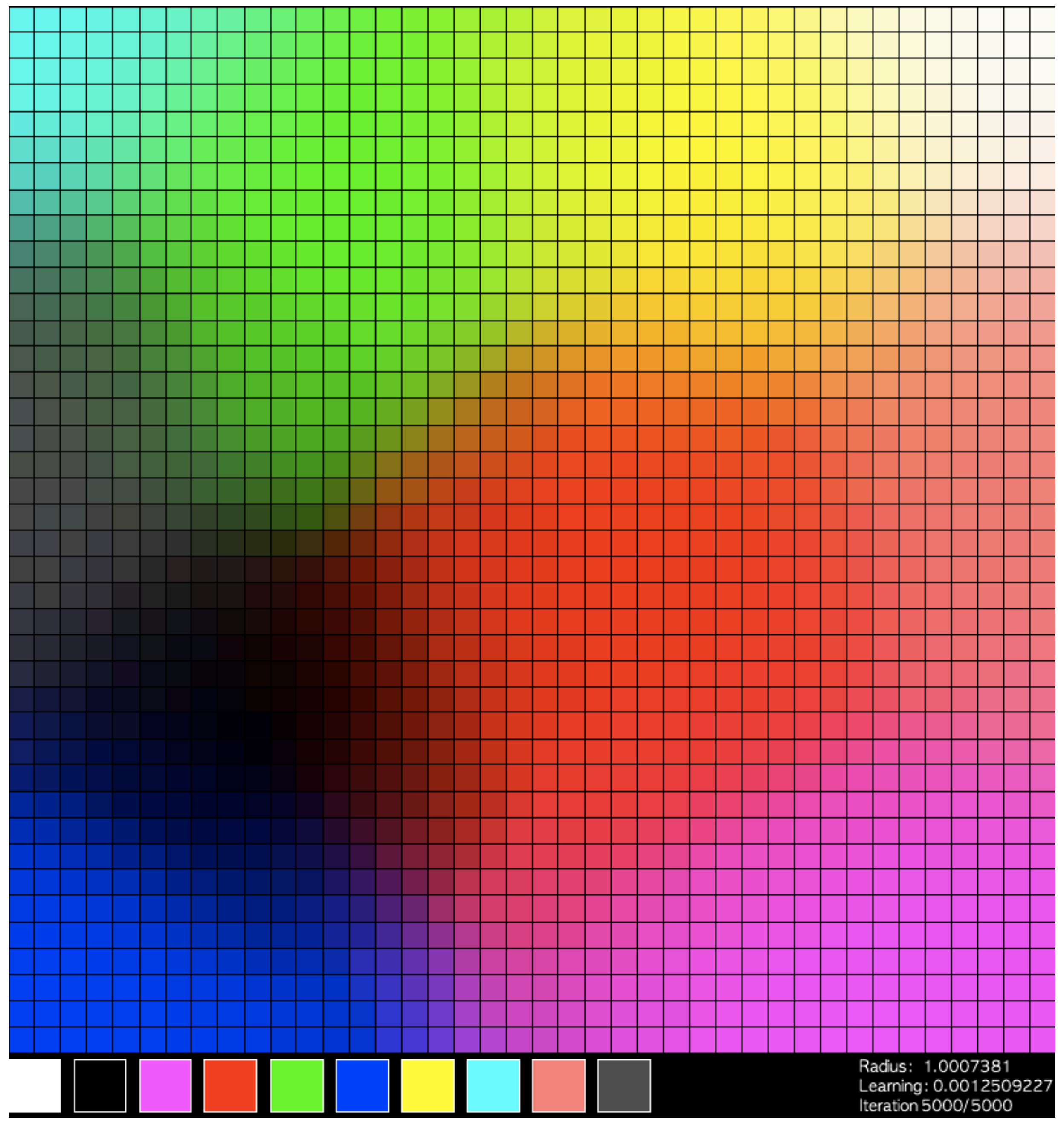
- Pictures
- Plans
- Text
- Categorical
- Numerical
- Time-series
- Multi-dimensional

- Encoding For Machine Learning

Types of Analysis, visualization, interpretation

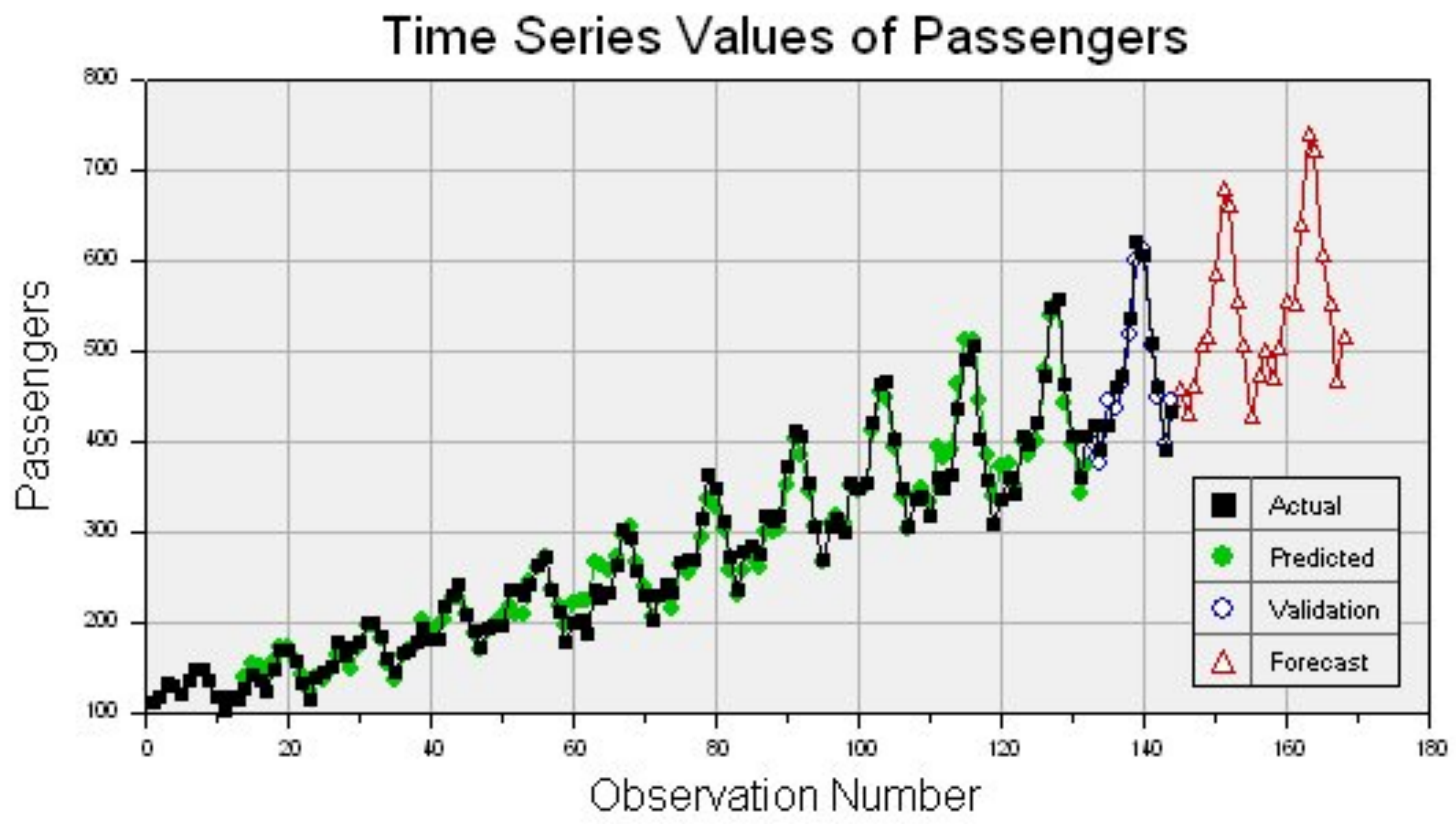
Data Mining and Self Organizing Maps

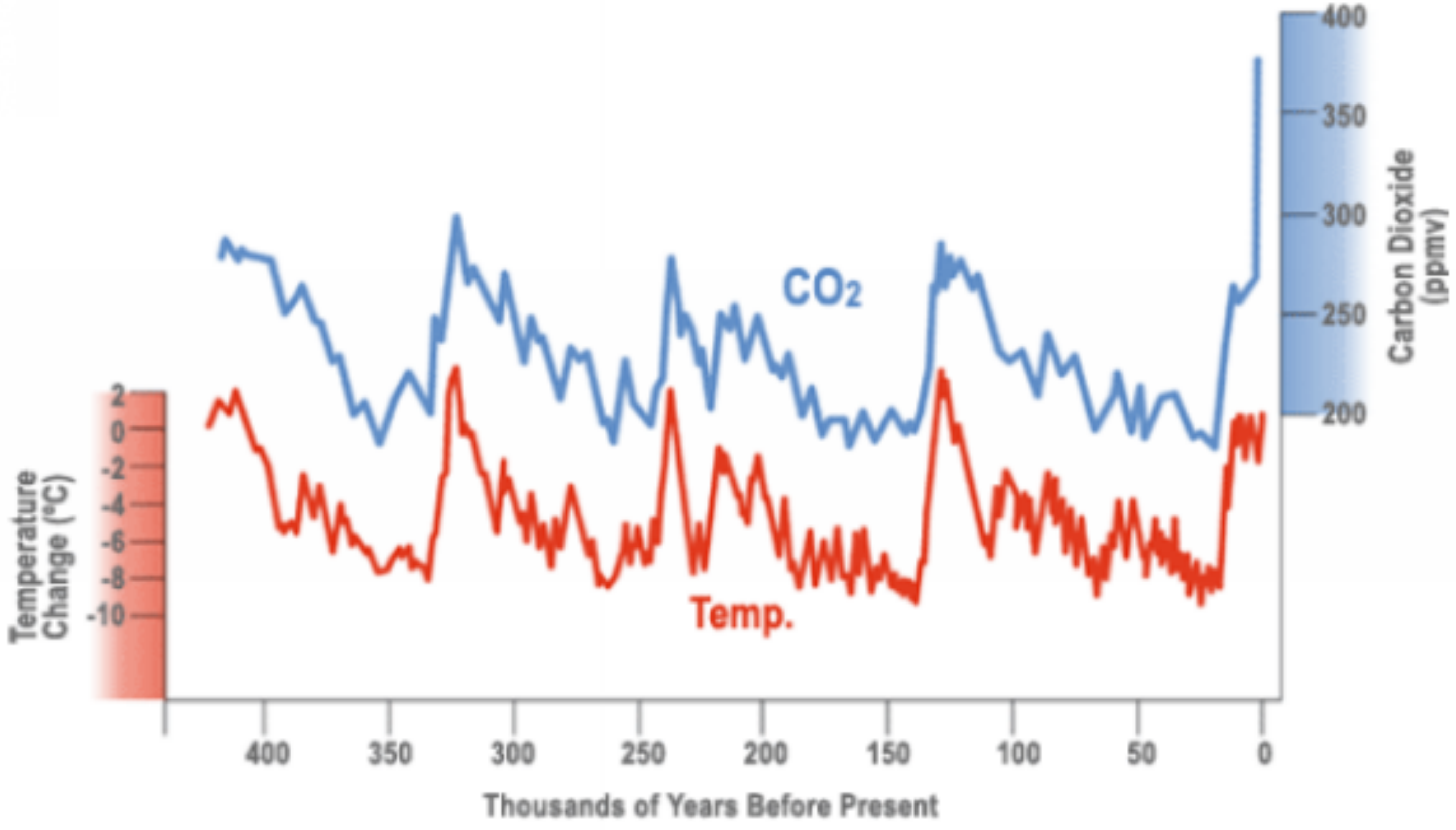




Types of Analysis, visualization, interpretation

Time Series Analysis





Shading analysis using heatmaps

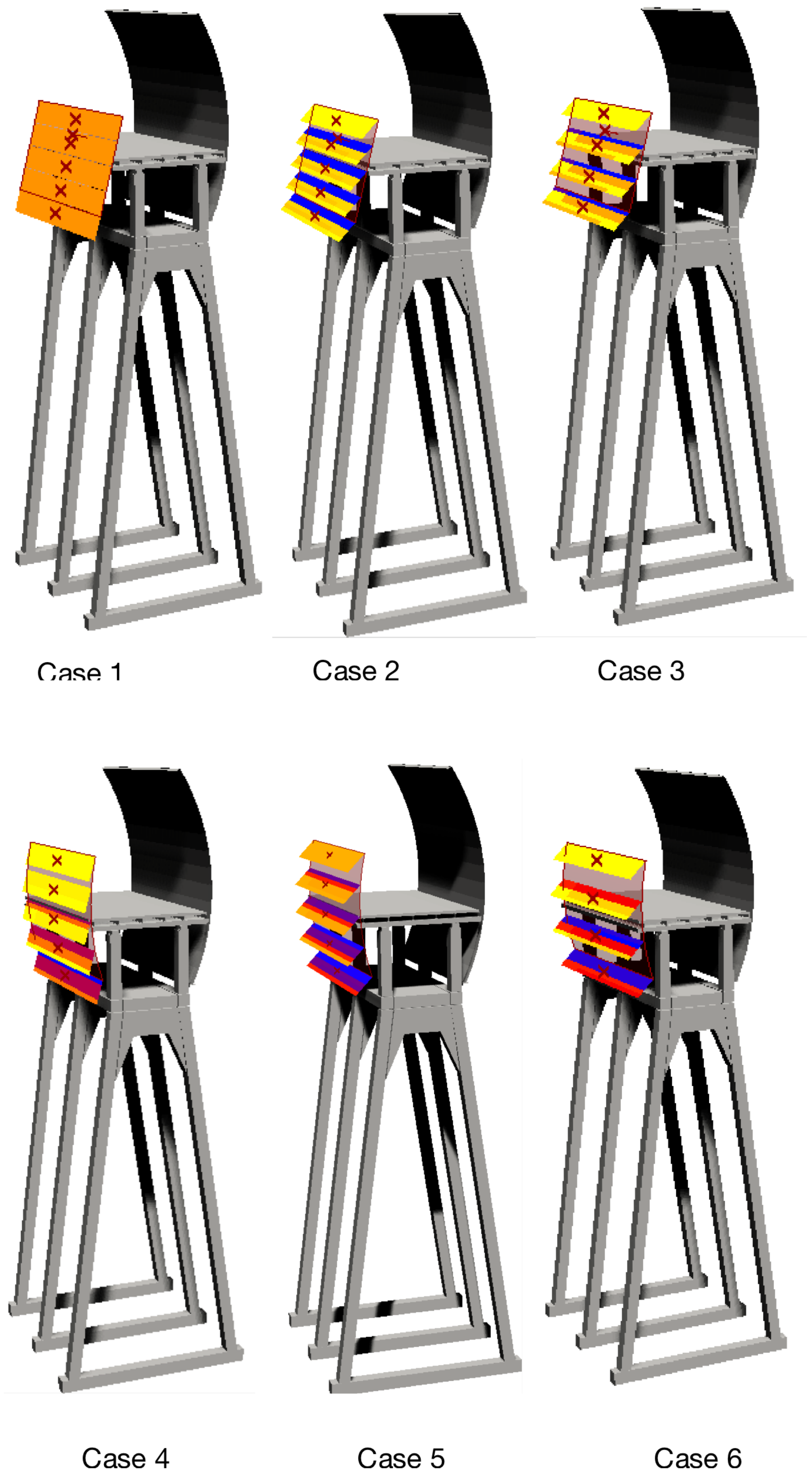


Figure 1: Bridge Configurations Case1-6

Case 1-3 (Top to bottom)

Case 4-6 (Top to bottom)

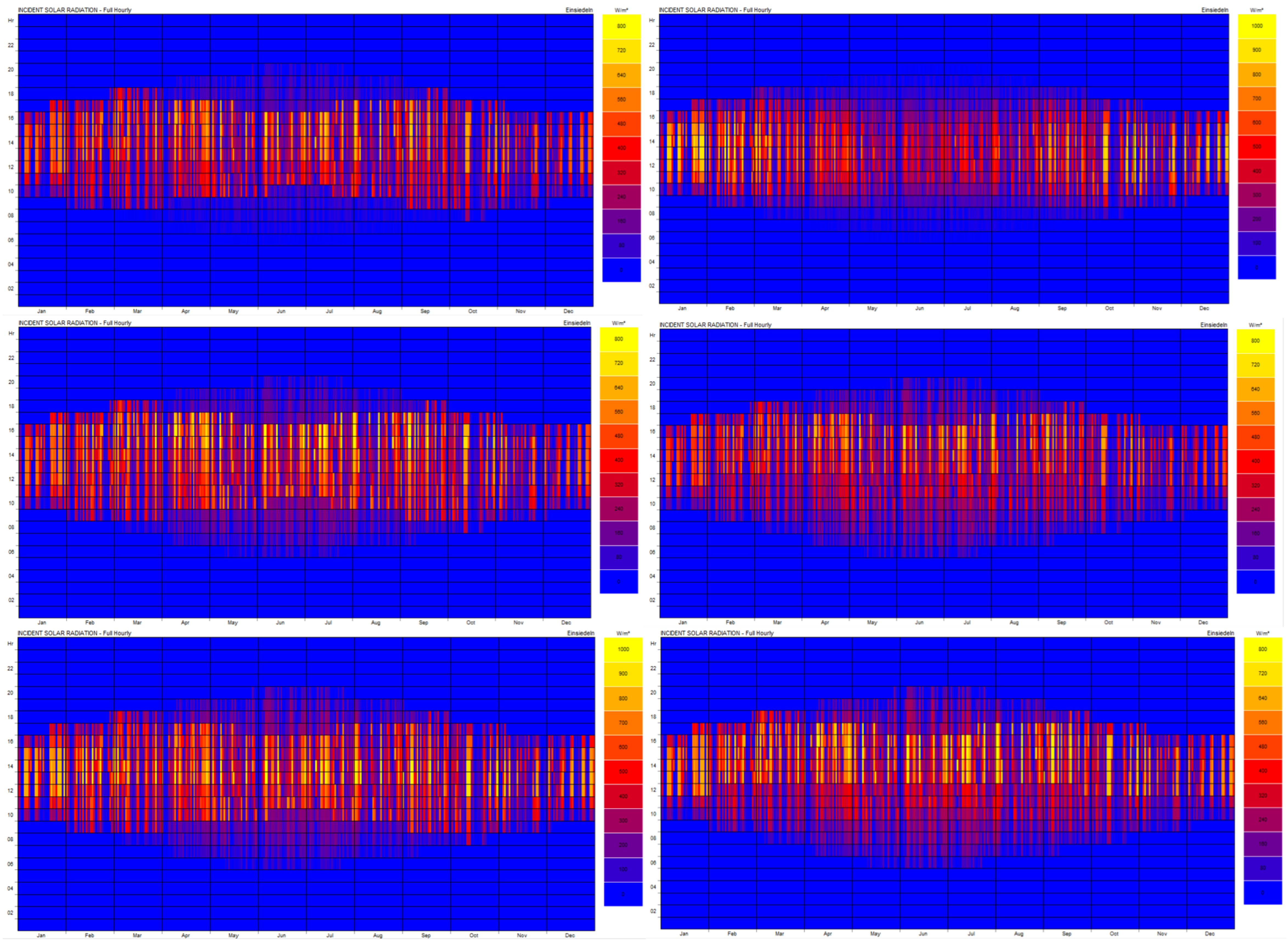


Figure 1: Heatmap Diagrams of the Hourly Incident Solar Radiation Case 1-6

Monthly performance evaluation using line graphs

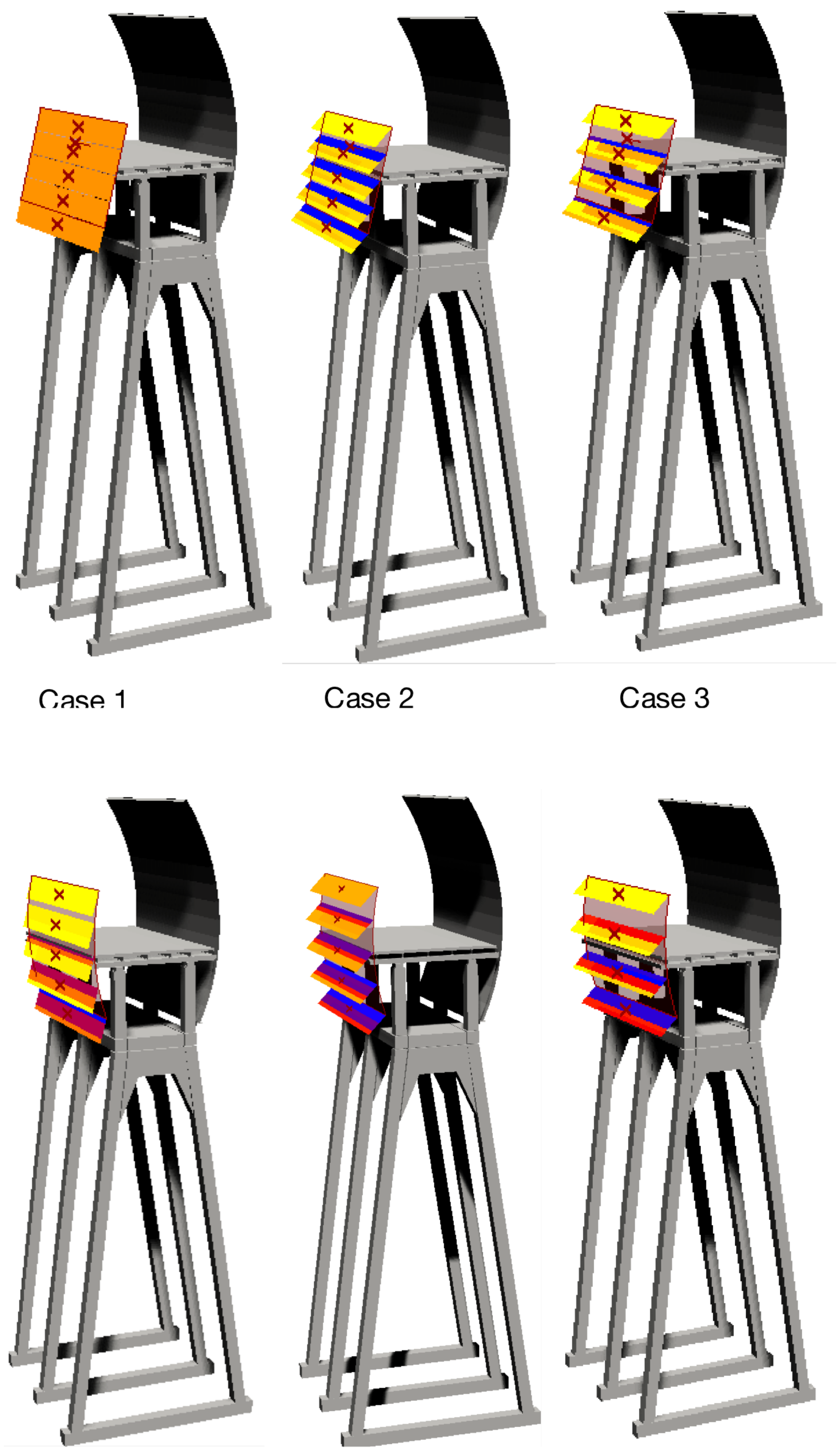


Figure 1: Bridge Configurations Case1-6

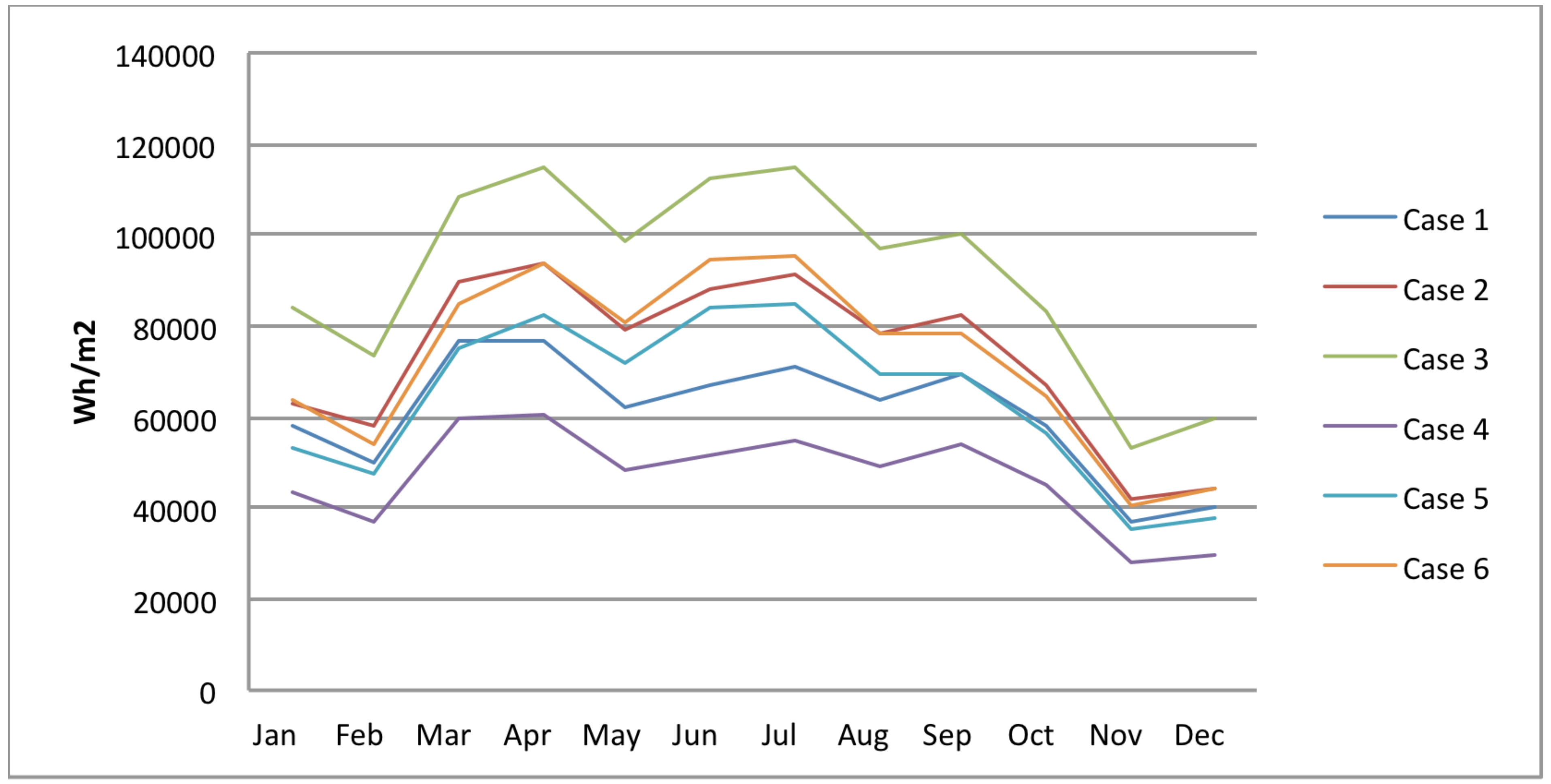


Figure 1: Monthly Total Incident Solar Radiation

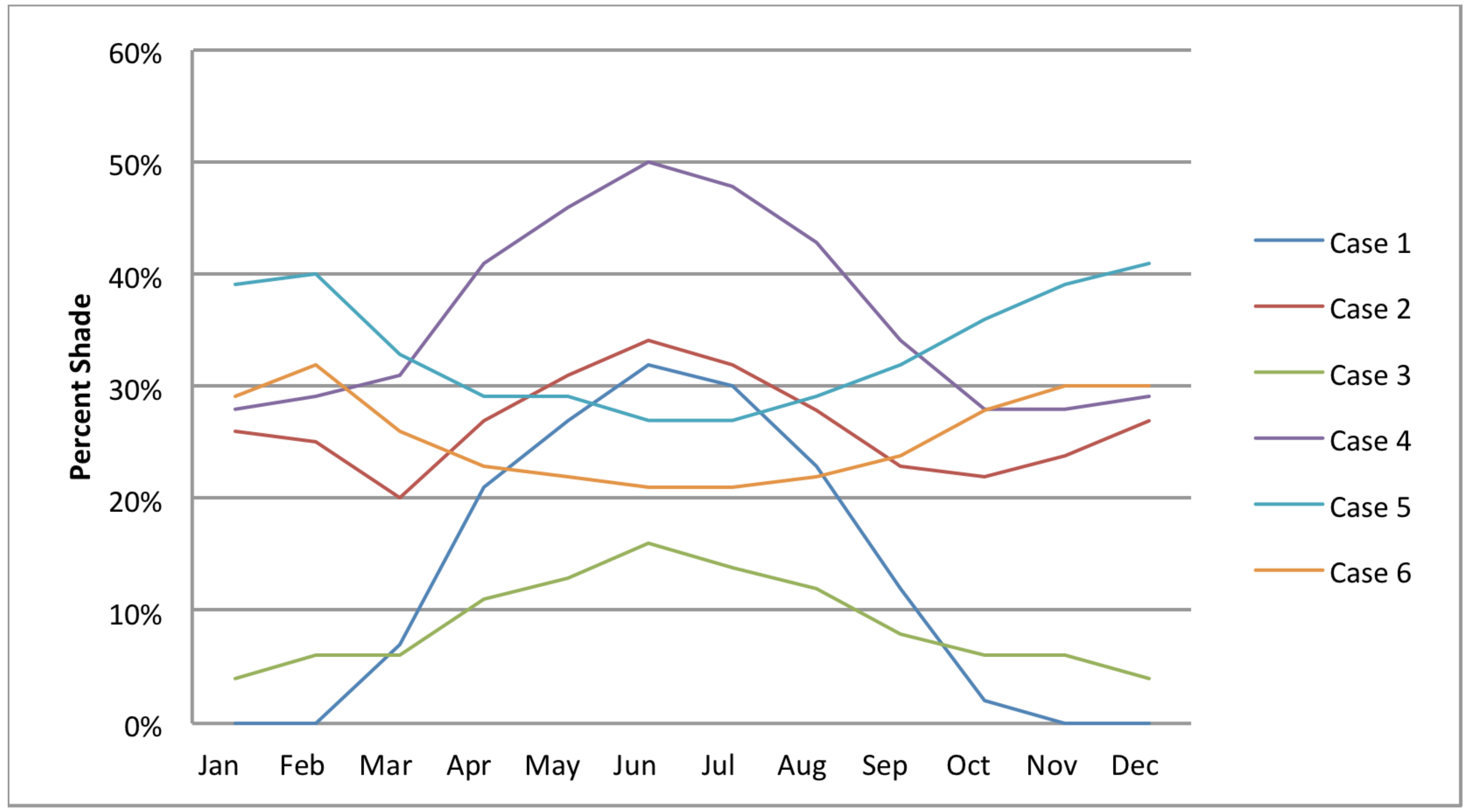


Figure 2: Average Monthly Shading

Sensor measurements from residential buildings

(a) (b)

(c) (d)

Figure 4-1: Construction materials observed during residential site visits

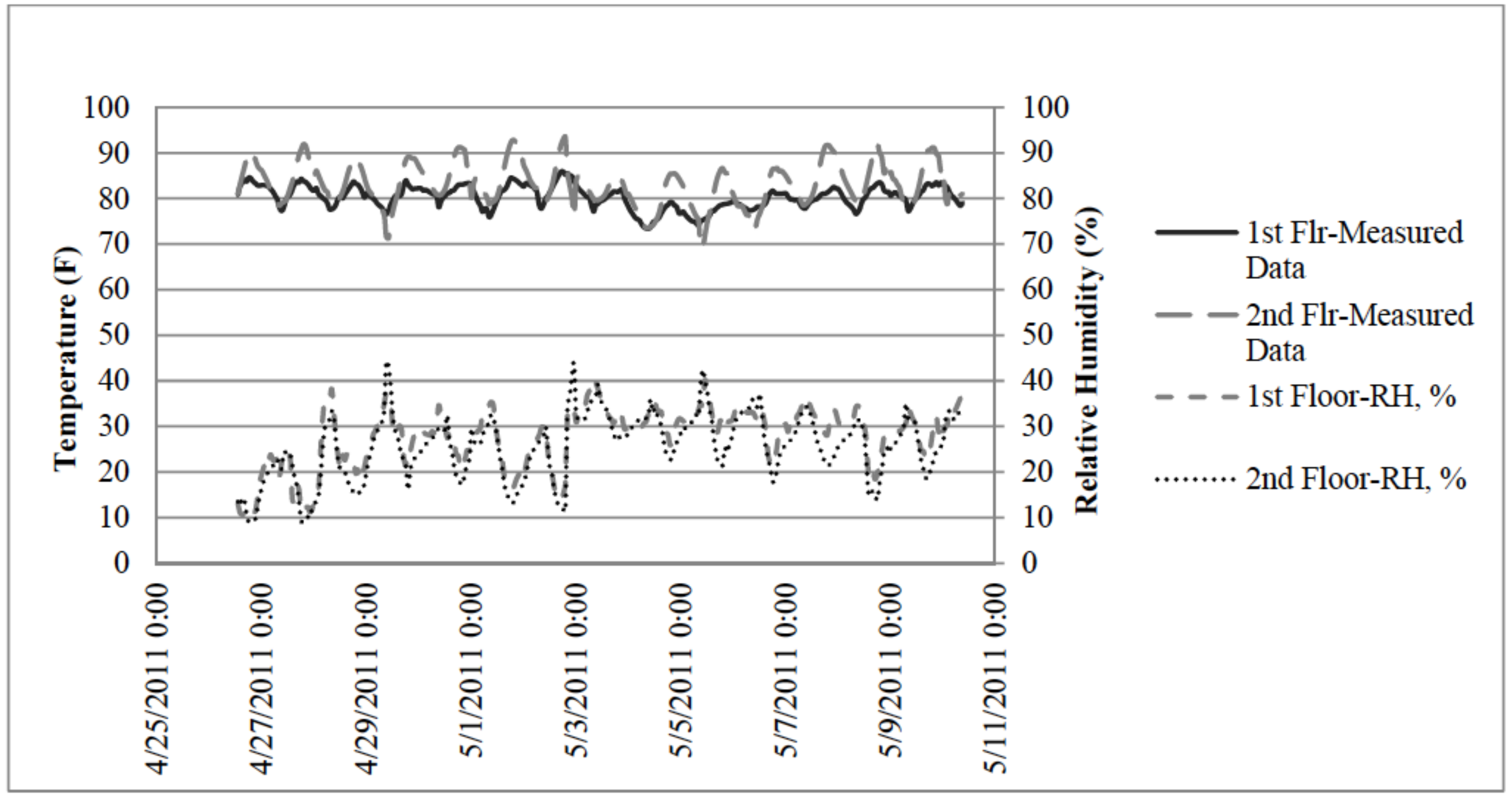


Figure 4-5: Two-week indoor temperature and relative humidity measurements

Model validation with sensor measurements

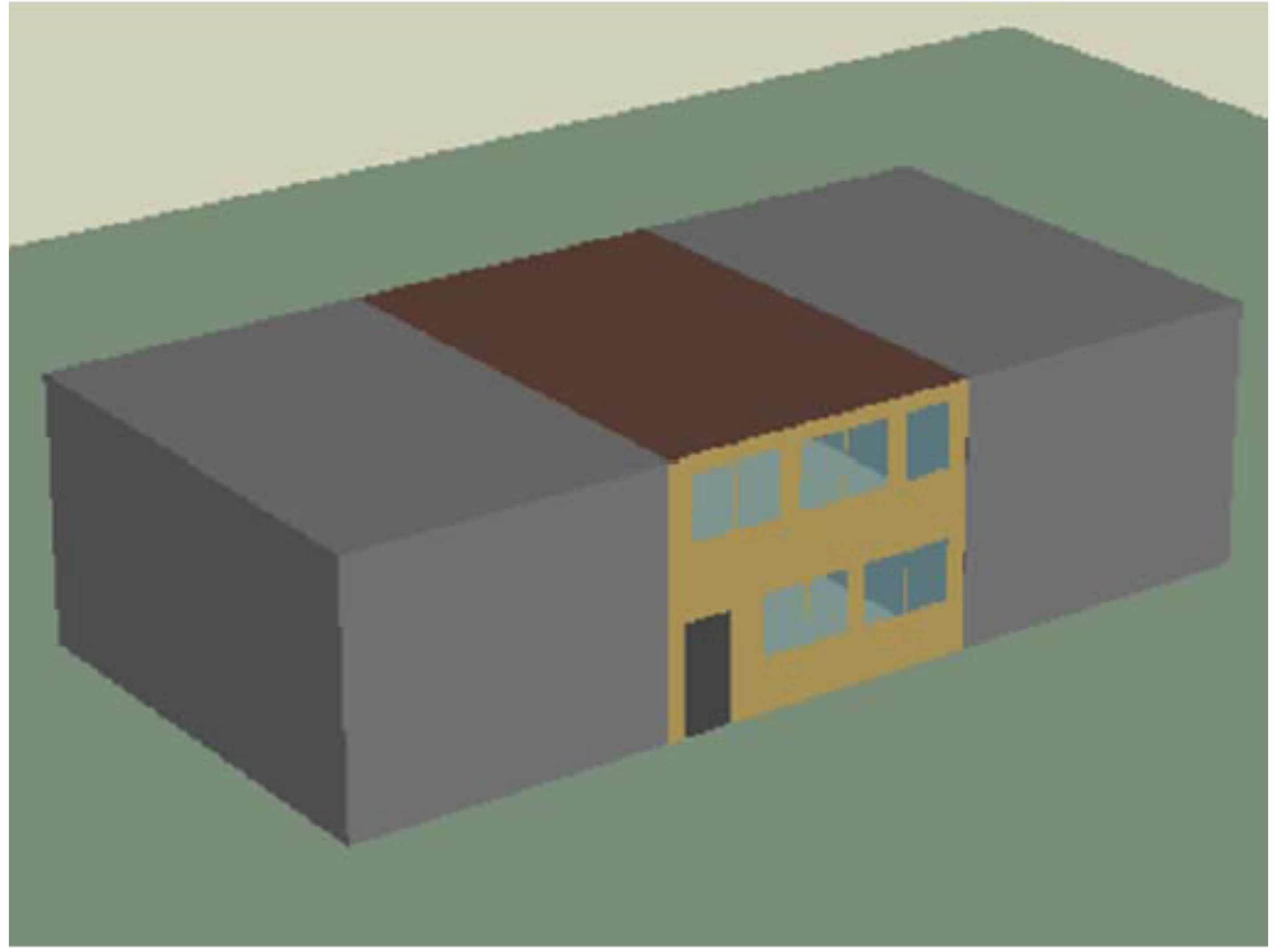


Figure 4-3: Residential building energy model

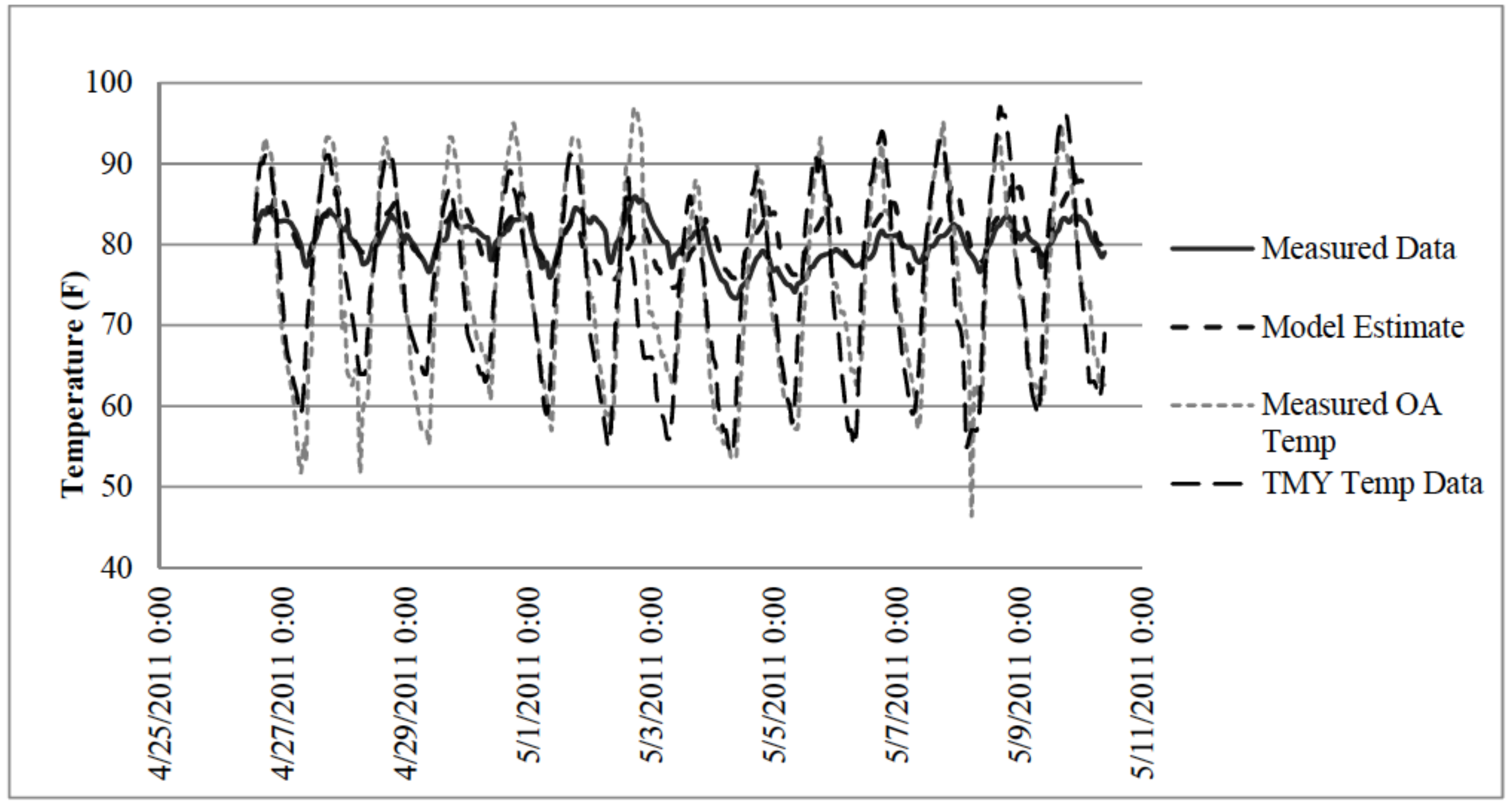


Figure 4-6: First floor temperature validation

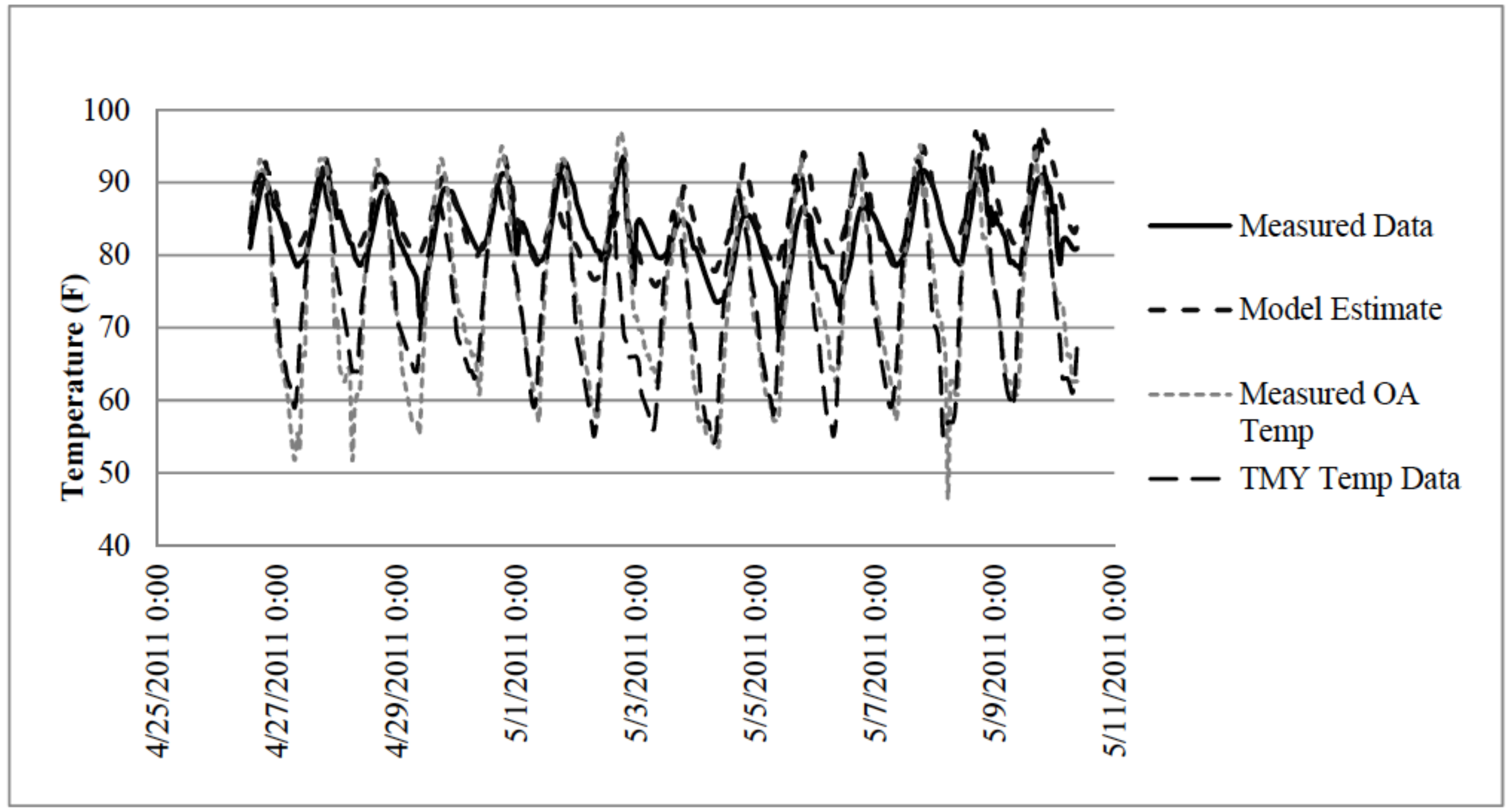


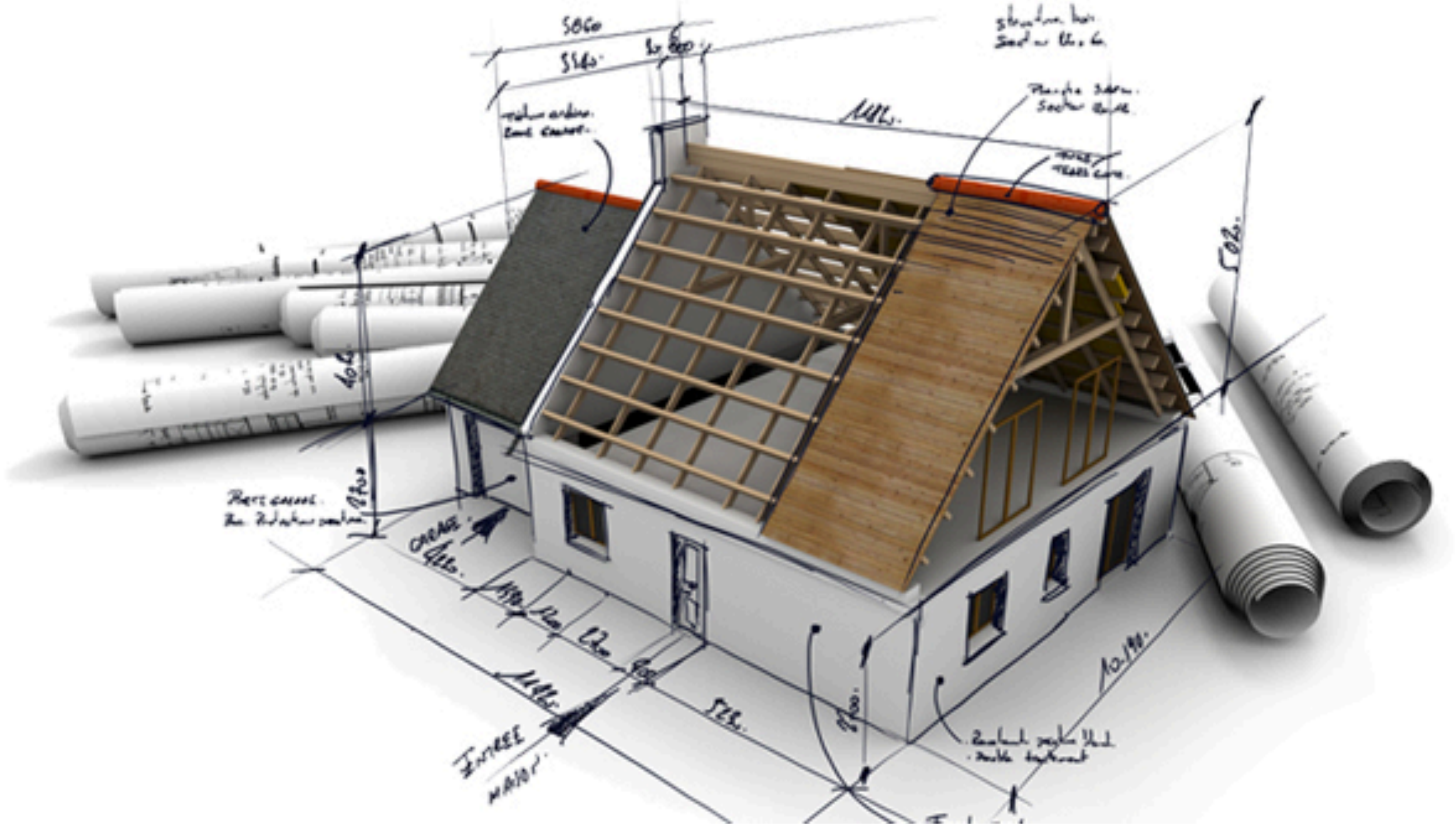
Figure 4-7: Second floor temperature validation

Overview

COURSE STRUCTURE		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

Typical Architectural Process

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Analysis, visualization, interpretation</td></tr>
<tr><td>Manual</td><td>Automated</td></tr>
<tr><td rowspan="2">Data Source</td><td>Manual</td><td></td><td></td></tr>
<tr><td>Automated</td><td></td><td></td></tr>
</table>



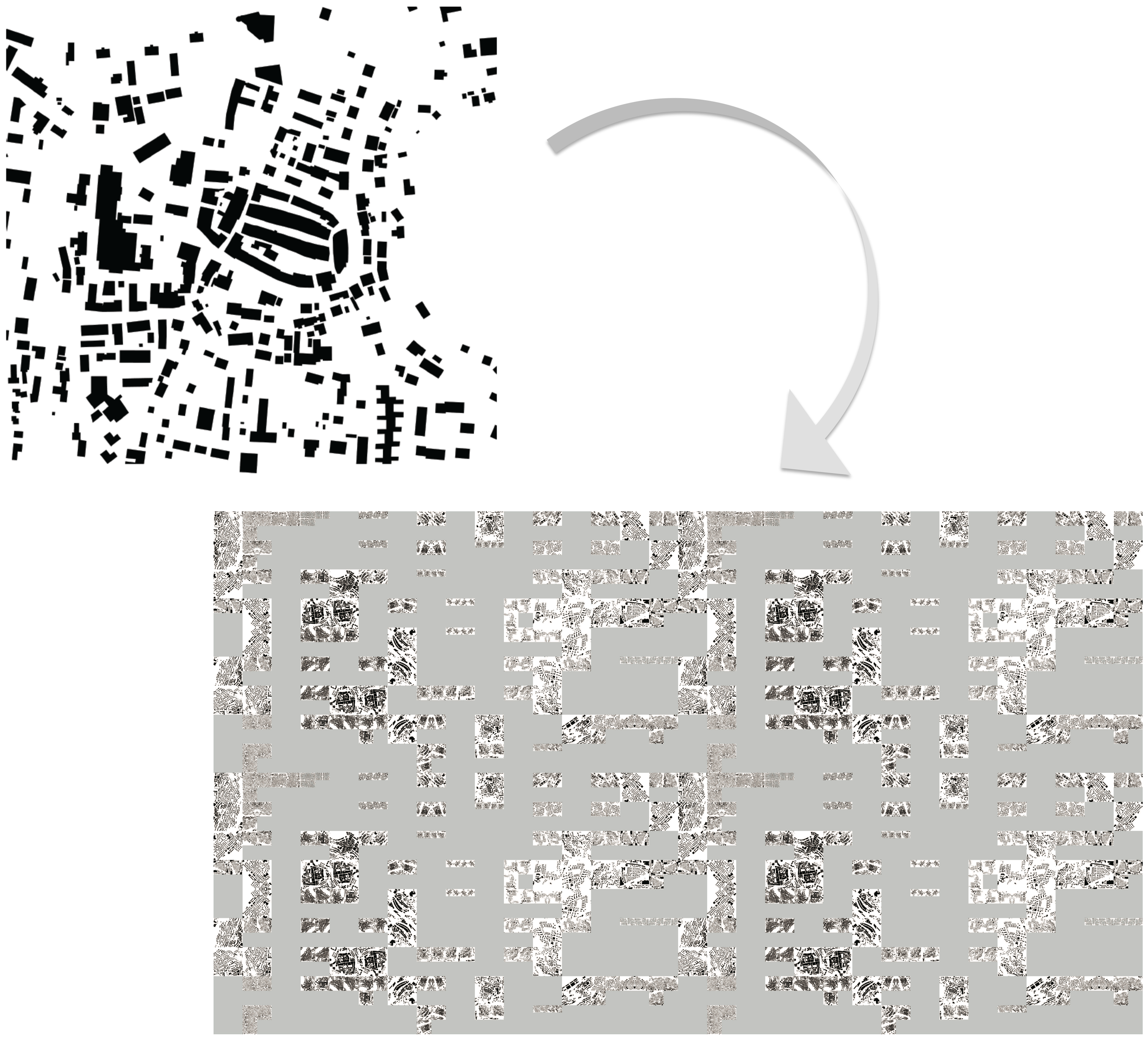
http://www.stamfordbuildingandconstruction.co.uk/our-services/architectural-drawings

Manual data input, automated analysis

<table>
<tr><td colspan="2" rowspan="2">BLOCK 1</td><td colspan="2">Analysis, visualization, interpretation</td></tr>
<tr><td>Manual</td><td>Automated

Data Mining/ machine learning</td></tr>
<tr><td rowspan="2">Data Source</td><td>Manual

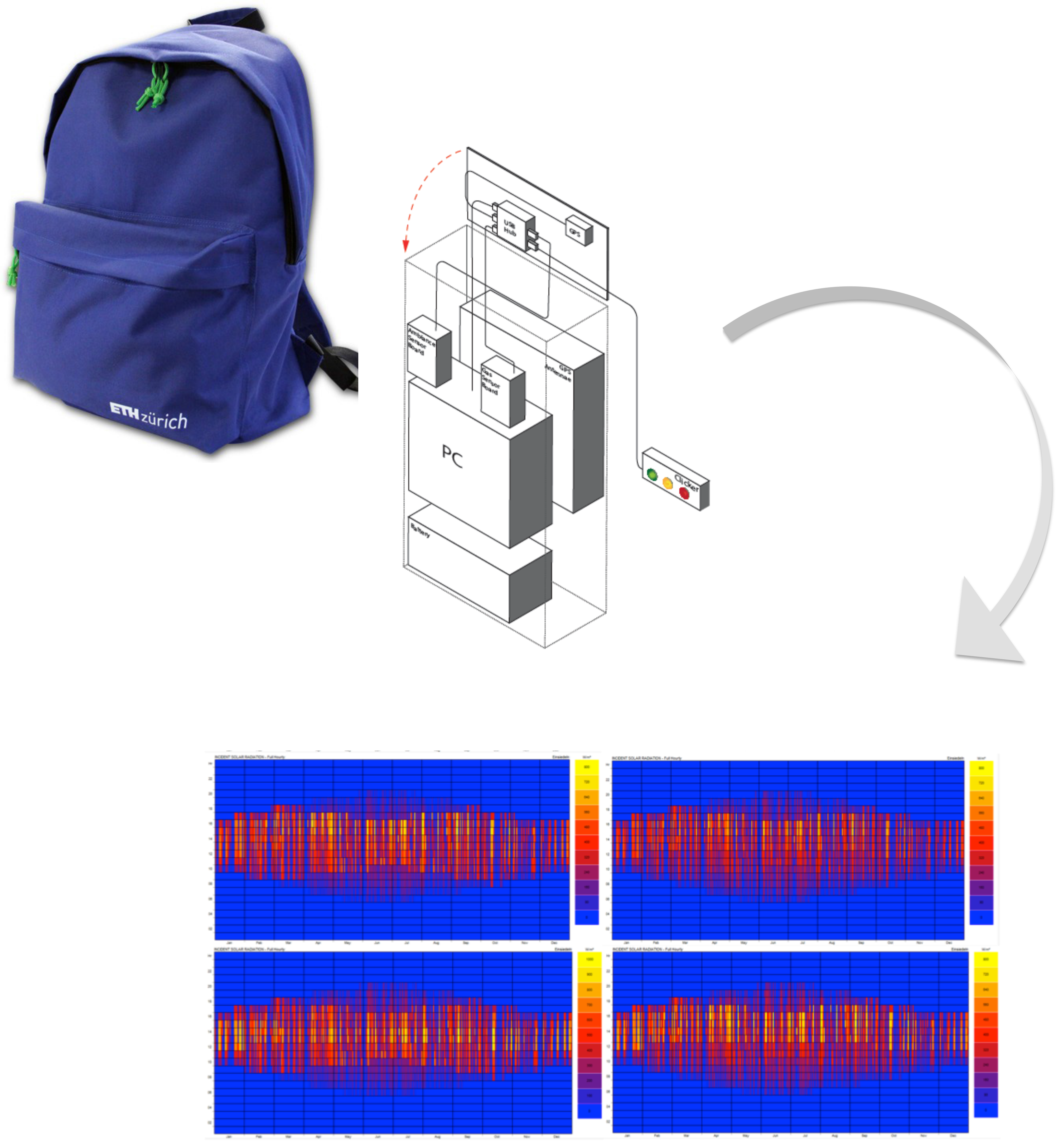
Hand-drawn sketches</td><td></td><td></td></tr>
<tr><td>Automated</td><td></td><td></td></tr>
</table>



Final Project from Moritz Berchtold, Creative Data Mining FS2015

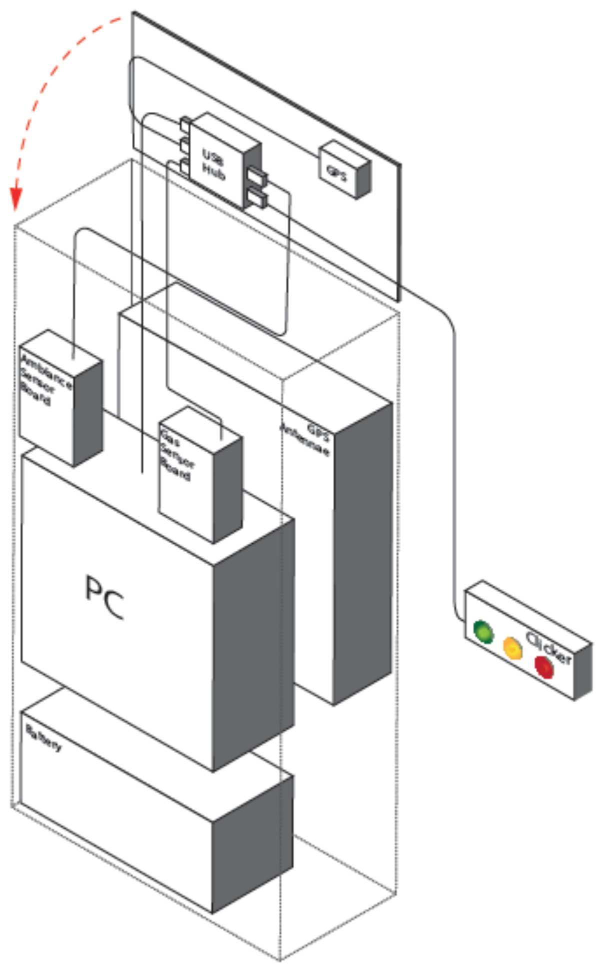
Automated data collection, manual data analysis

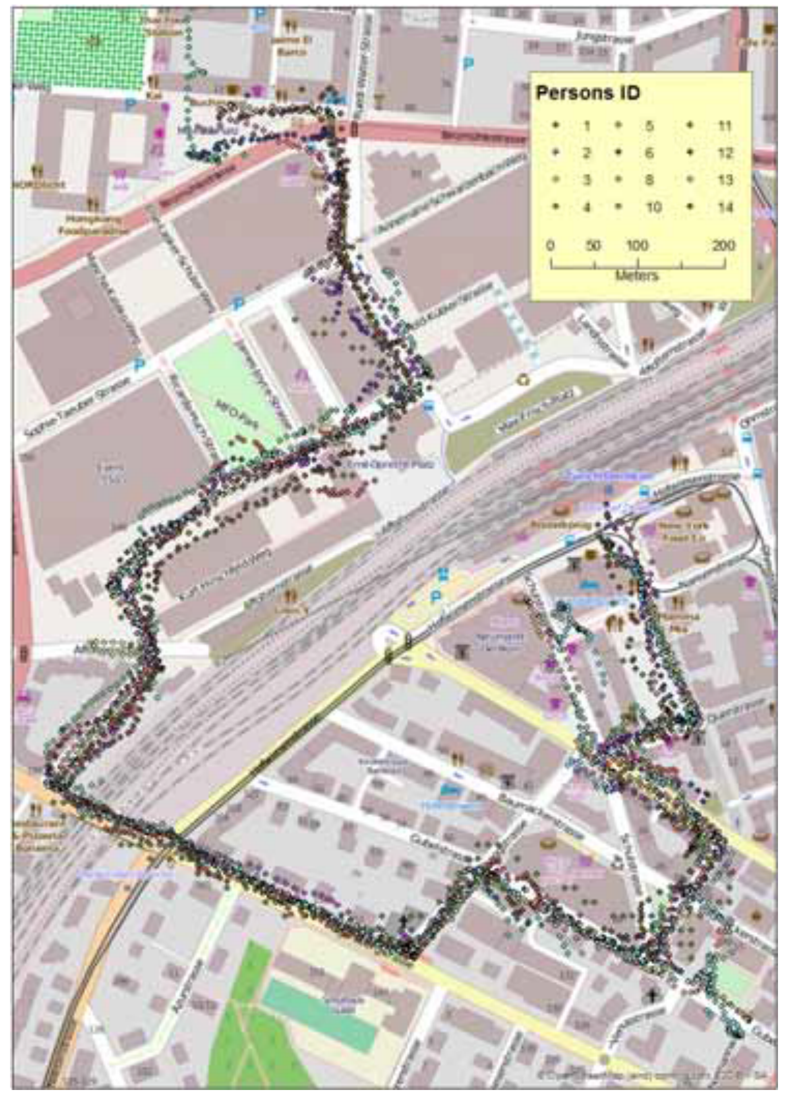
BLOCK 2		Analysis, visualization, interpretation	
		Manual Time-series and geo-referenced analysis	Automated
Data Source	Manual		
	Automated Sensor data		



Automated data collection, automated analysis & visualization

FUTURE WORK		Analysis, visualization, interpretation	
		Manual	Automated Data Mining/ Machine Learning
Data Source	Manual		
	Automated Sensor data		





Data Mining for Architects and Urban Planners?

Examples

National data collection project

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

FCL Lunch Talk

(FCL) FUTURE CITIES LABORATORY 未来城市实验室

Singapore's National Science Experiment

43,000 Students Exploring Their Personal Data

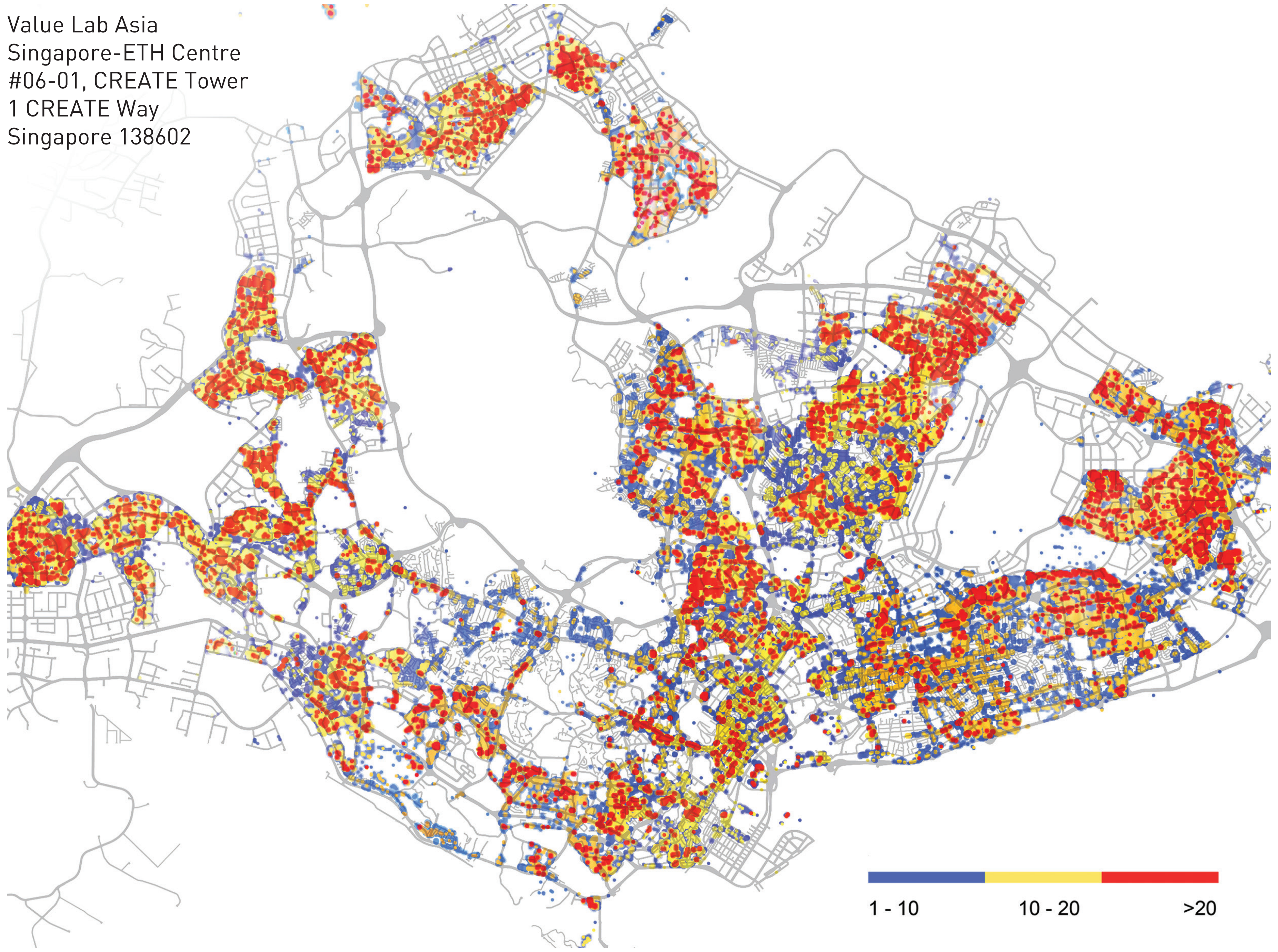
The Singapore University of Technology and Design facilitated an ambitious large-scale science experiment in September and November 2015 which saw over 43,000 students carrying sensors designed to measure temperature, humidity, pressure, light, noise, among other physical parameters in a project supported 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



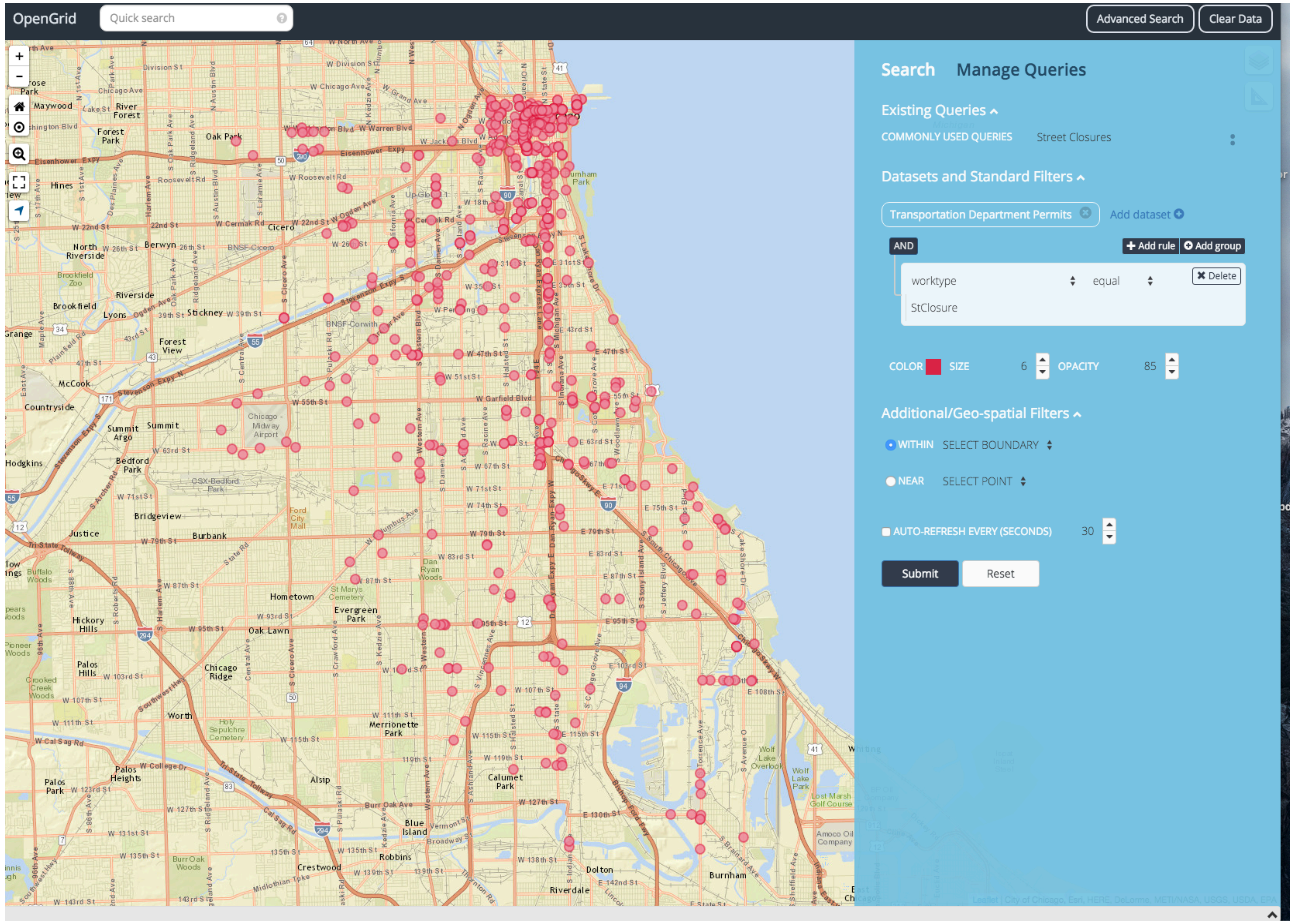
(SEC) SINGAPORE-ETH CENTRE 新加坡-ETH 研究中心

ETH zürich

CREATE
Campus for Research Excellence And Technological Enterprise

Chicago OpenGrid

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		



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

Newcastle University Urban Observatory

Data Source		Analysis, visualization, interpretation: Manual	Analysis, visualization, interpretation: Automated
	Manual		
	Automated		

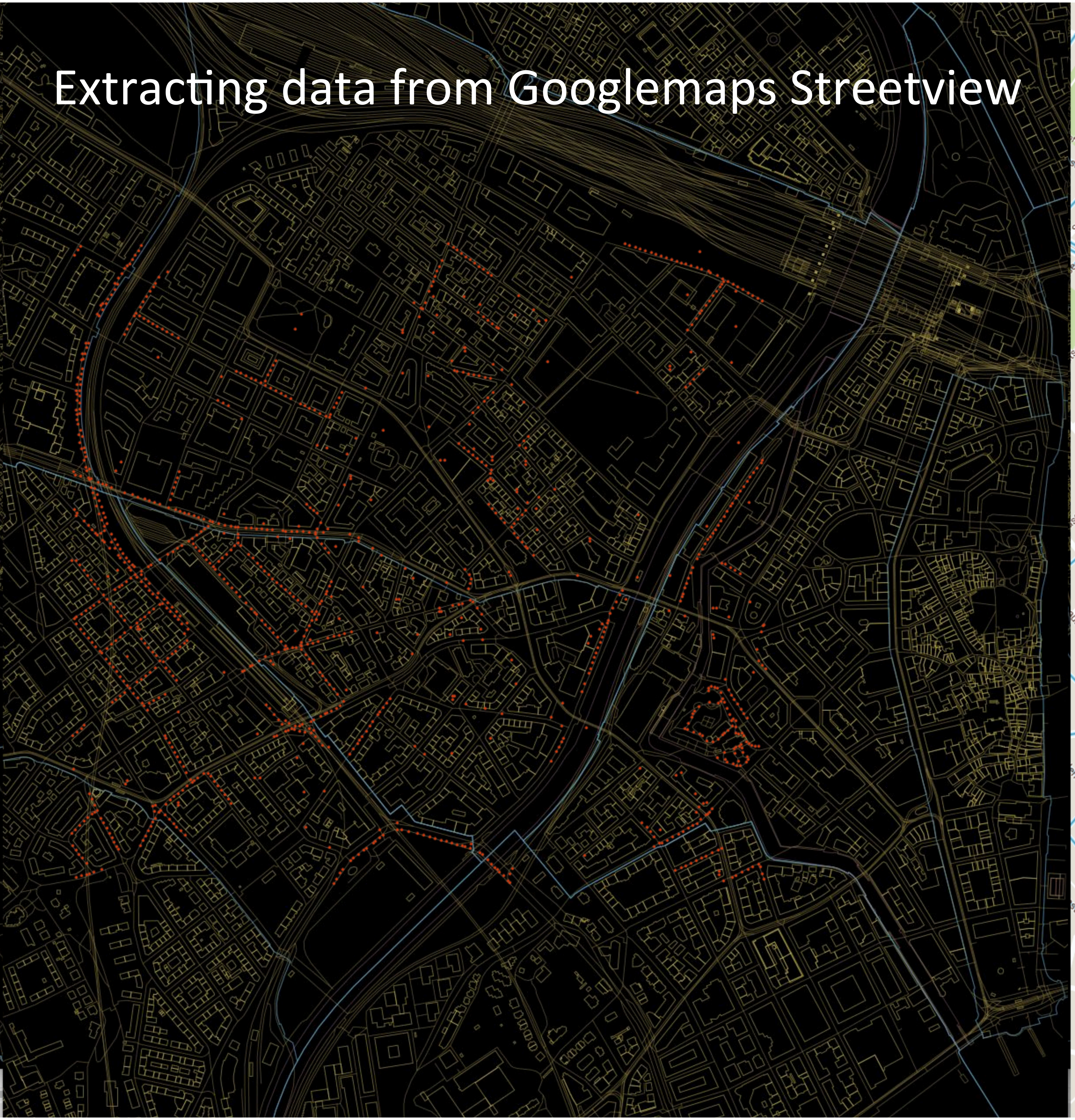
Data Source		Analysis, visualization, interpretation: Manual	Analysis, visualization, interpretation: Automated
	Manual		
	Automated		

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

http://map.datacanvas.org/

Data Canvas project output

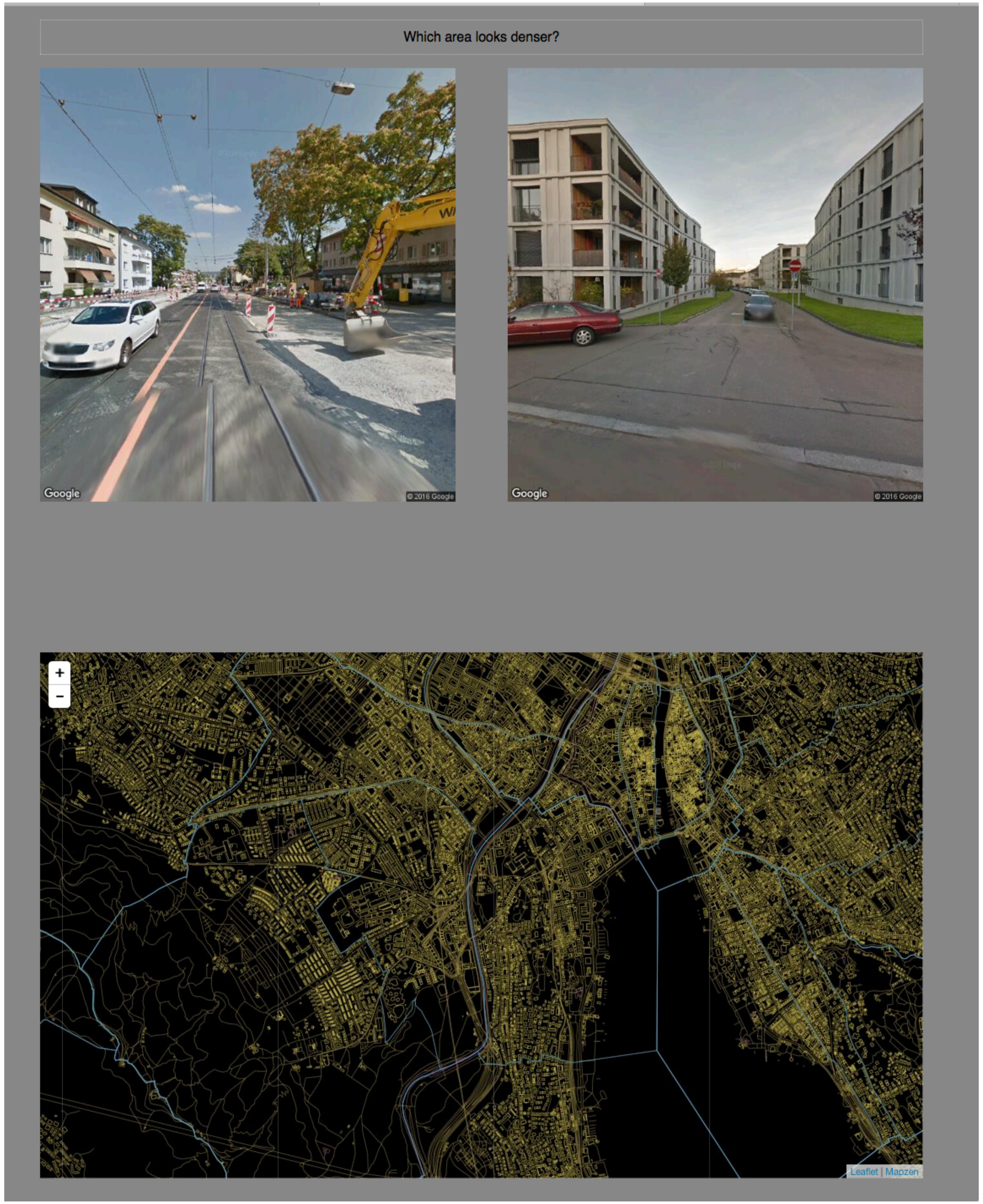
Extracting data from Googlemaps Streetview



		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

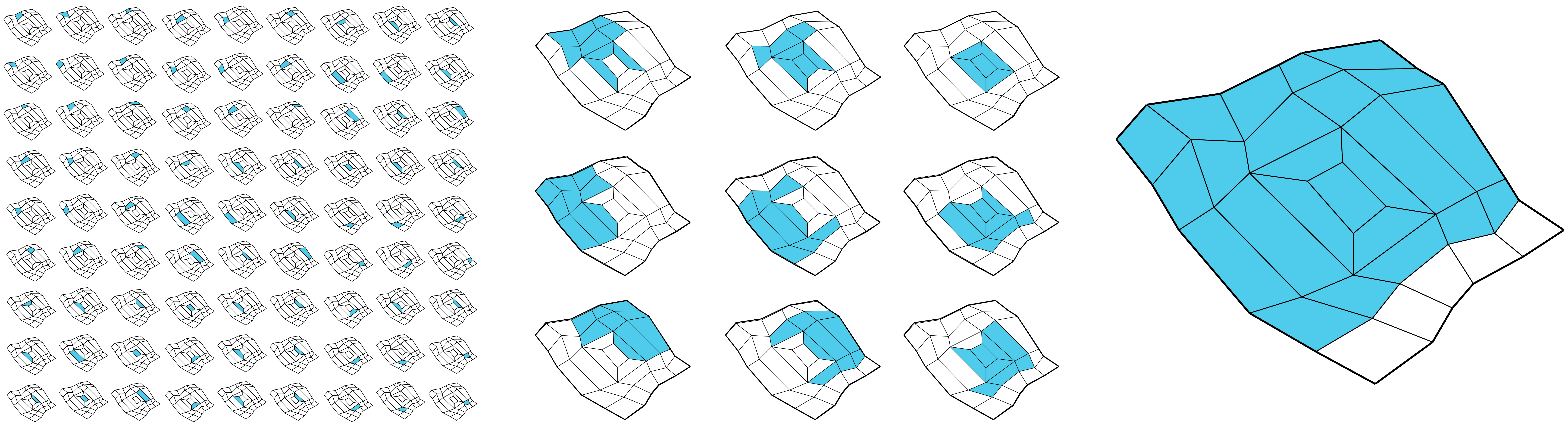
Evaluate density perception

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

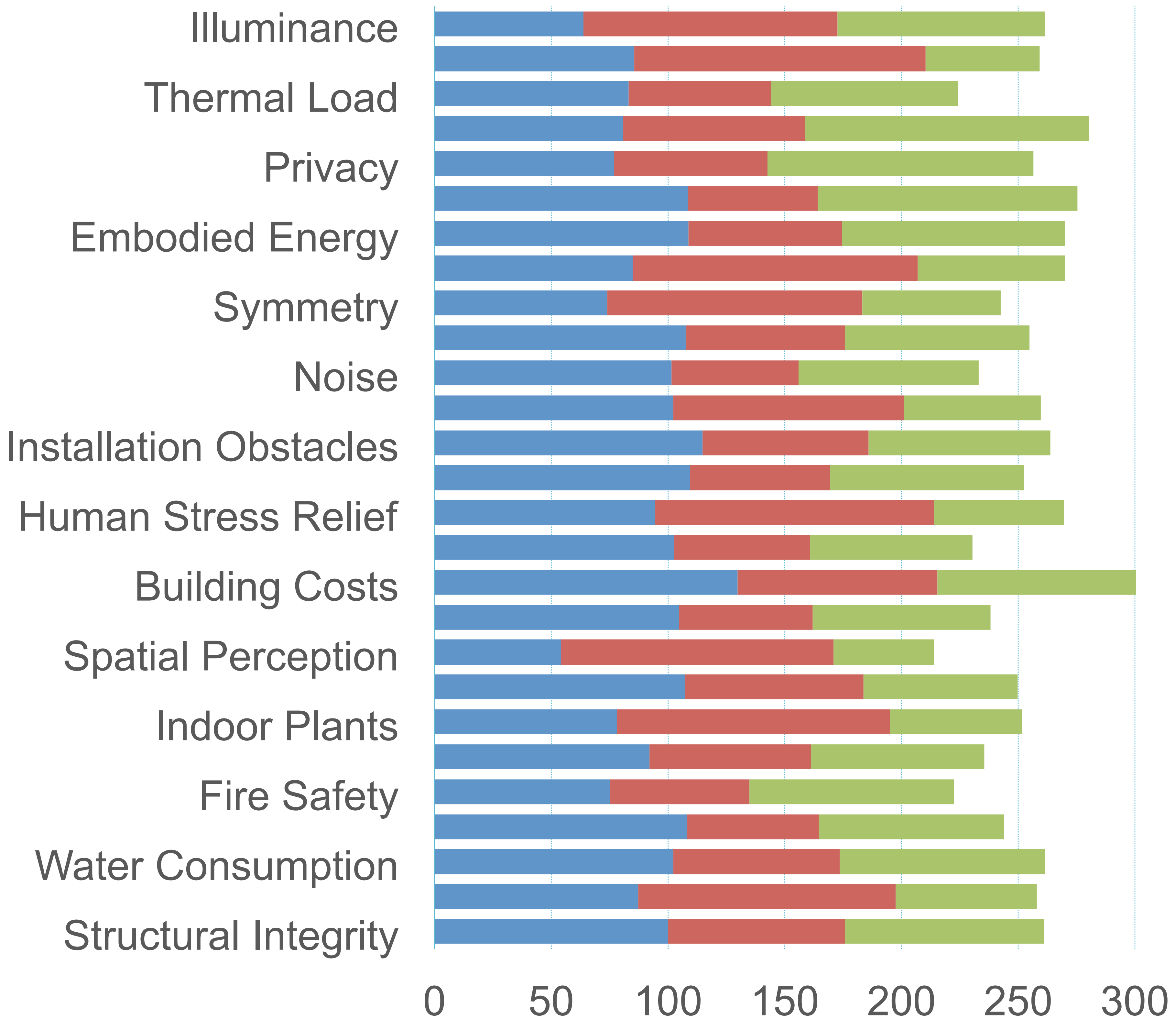


ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
DARCH
iA Chair of Information Architecture

Comparing Geometries – Meaning of Form

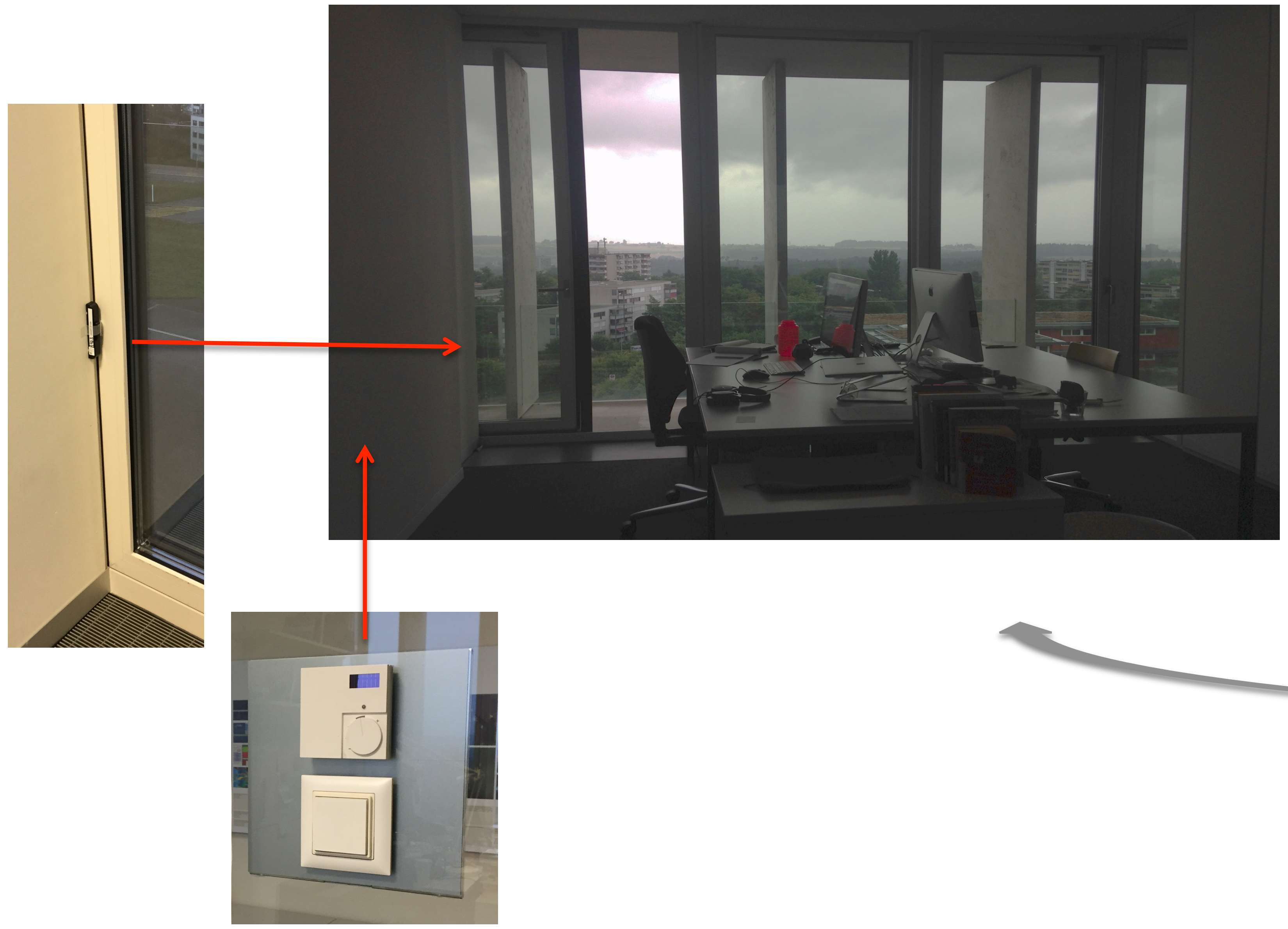


Comparing Geometries – Meaning of Form



What can we do with building data?

"user-behavior" sensors



Building Management System

Database

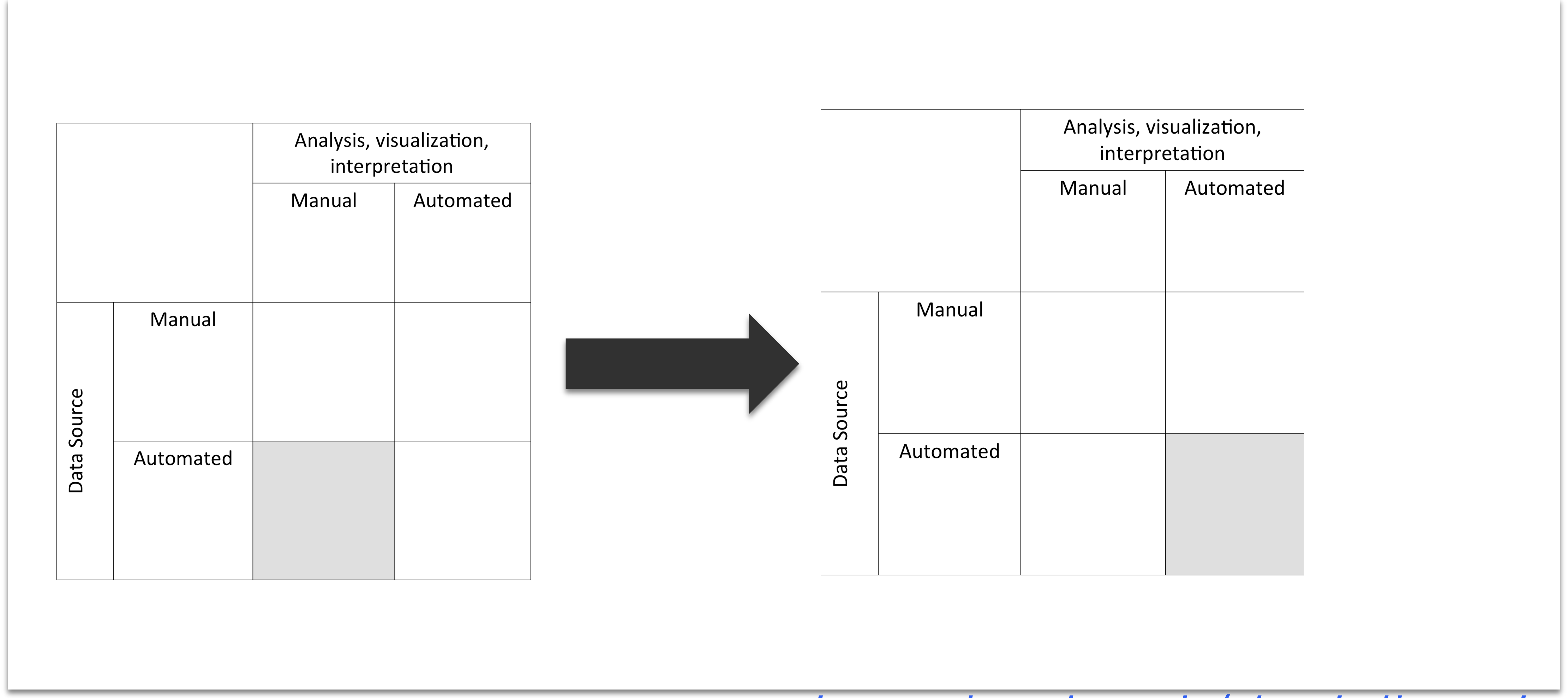
Benchmarking

Prediction

Demand Control

User-engagement

Visualization



		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

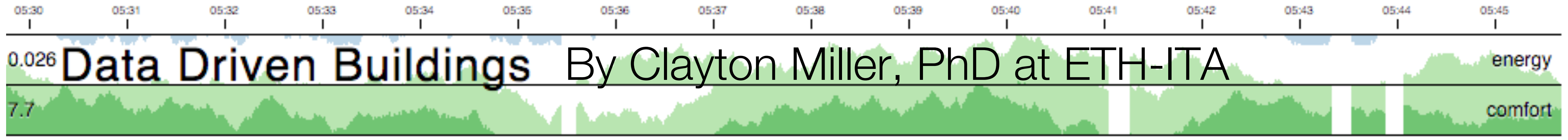
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

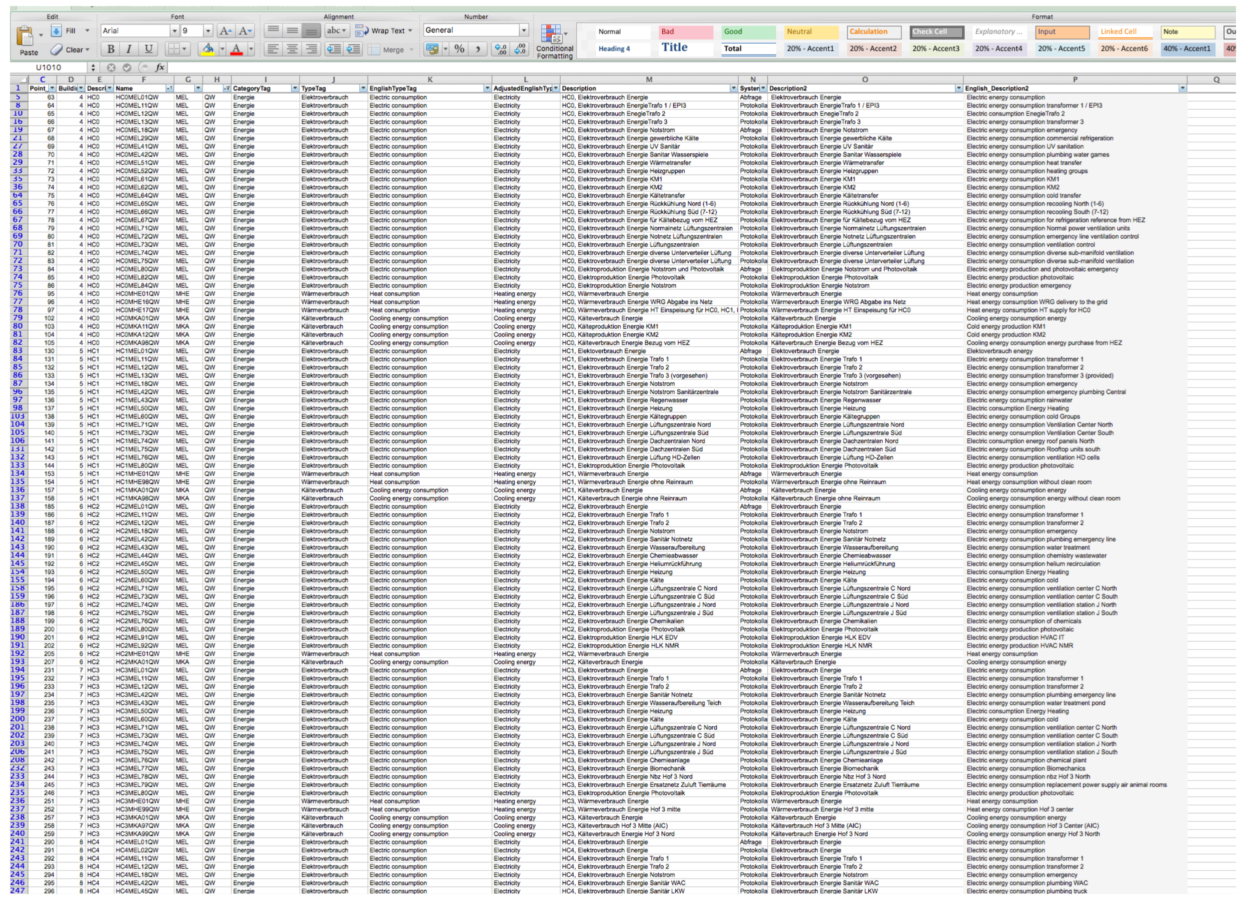
DARCH

iA Chair of Information Architecture

ia.arch.ethz.ch/danielle-griego

A source of inspiration





-Prediction
-Diagnostics
-Fault Detection

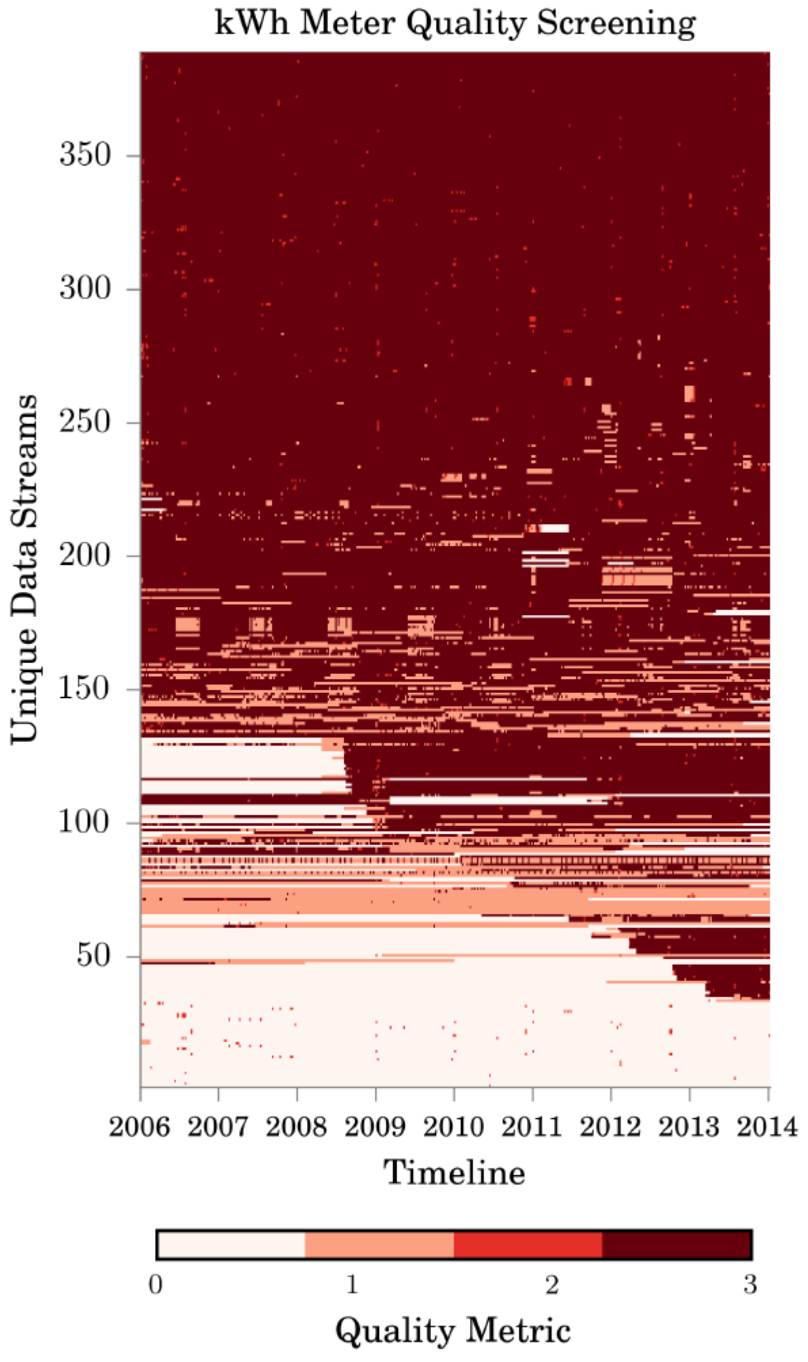


Figure 6. Data quality metrics map for campus sorted (bottom-to-top) according to increasing quality metric

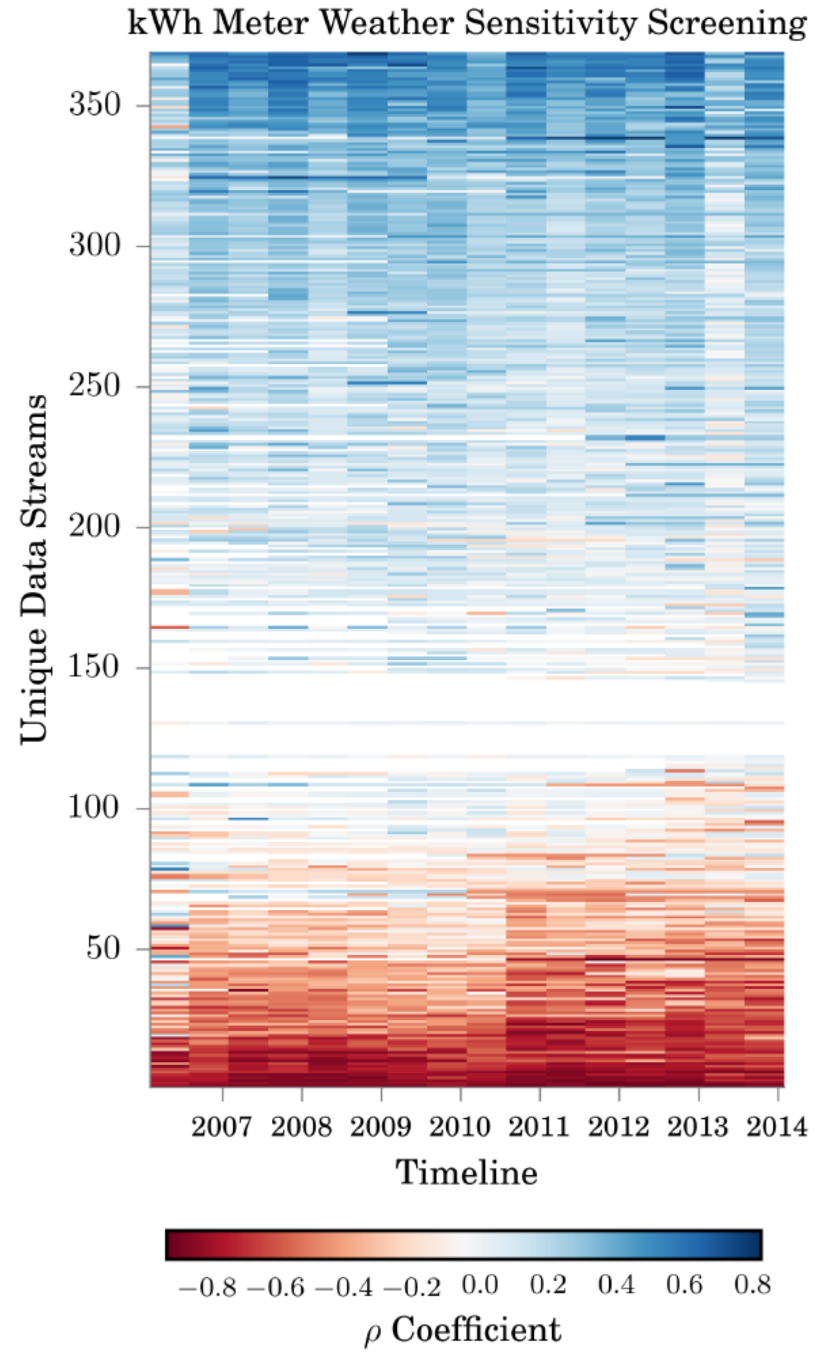
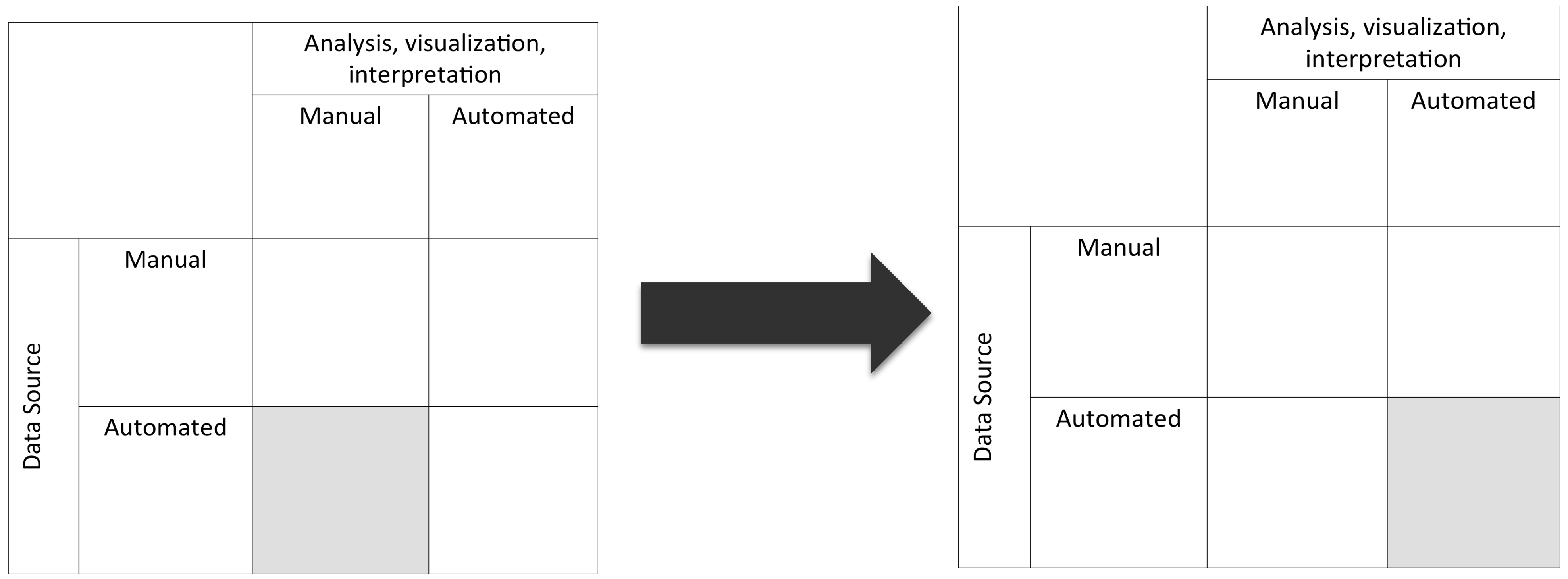


Figure 7. Weather sensitivity map sorted (bottom-to-top) from high negative to high positive ρ coefficient values



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

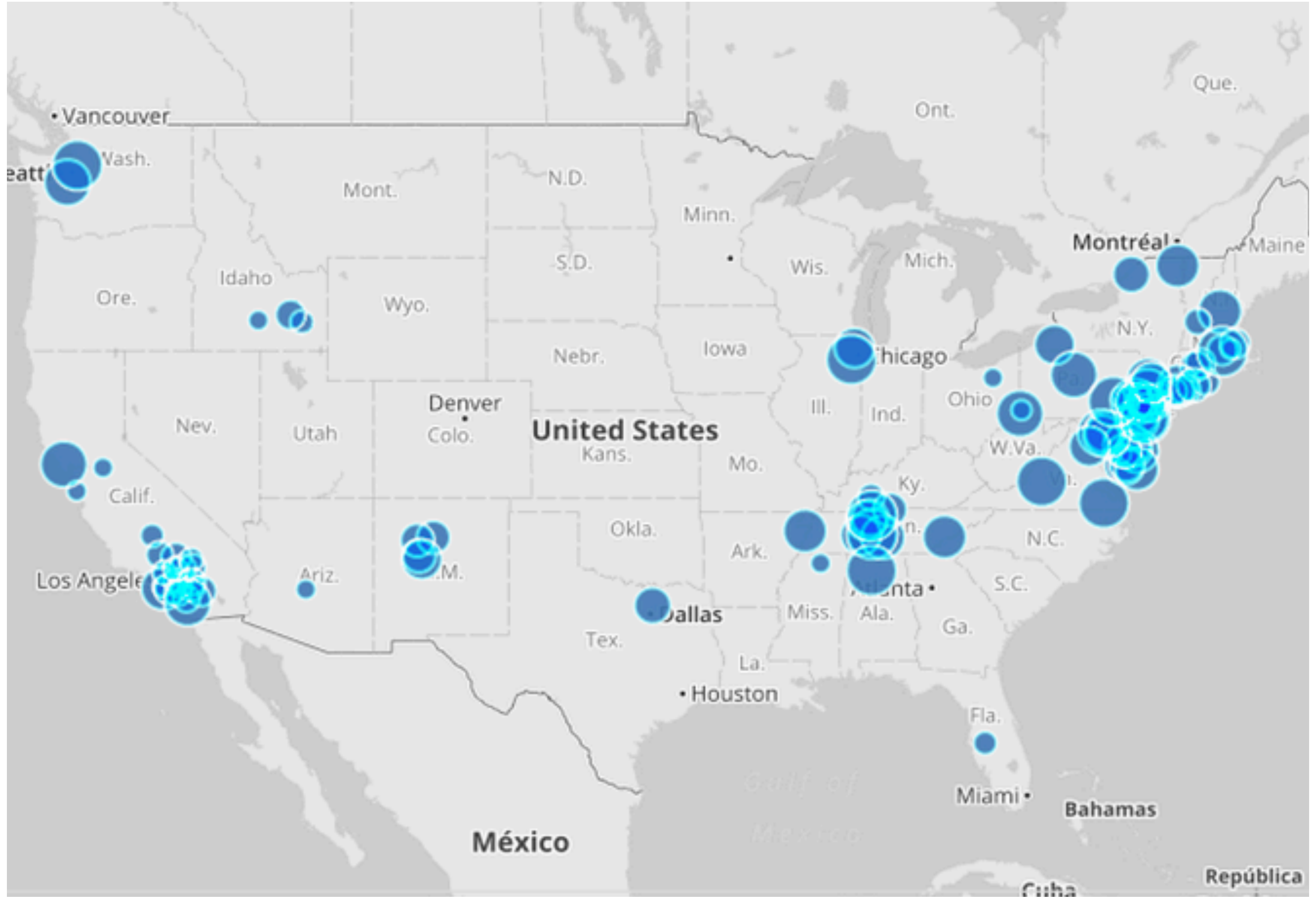
DARCH

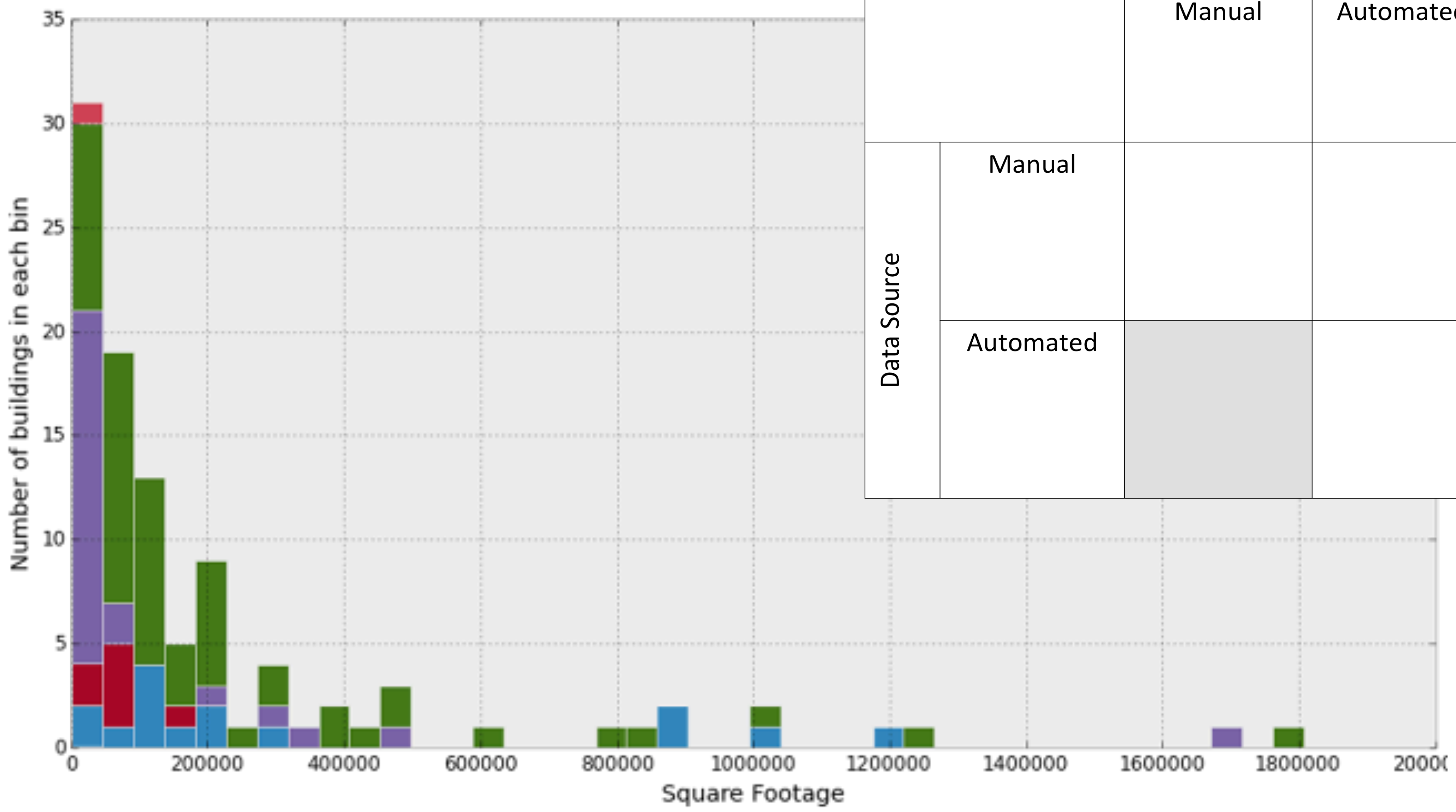
Chair of Information Architecture

datadrivenbuilding.org

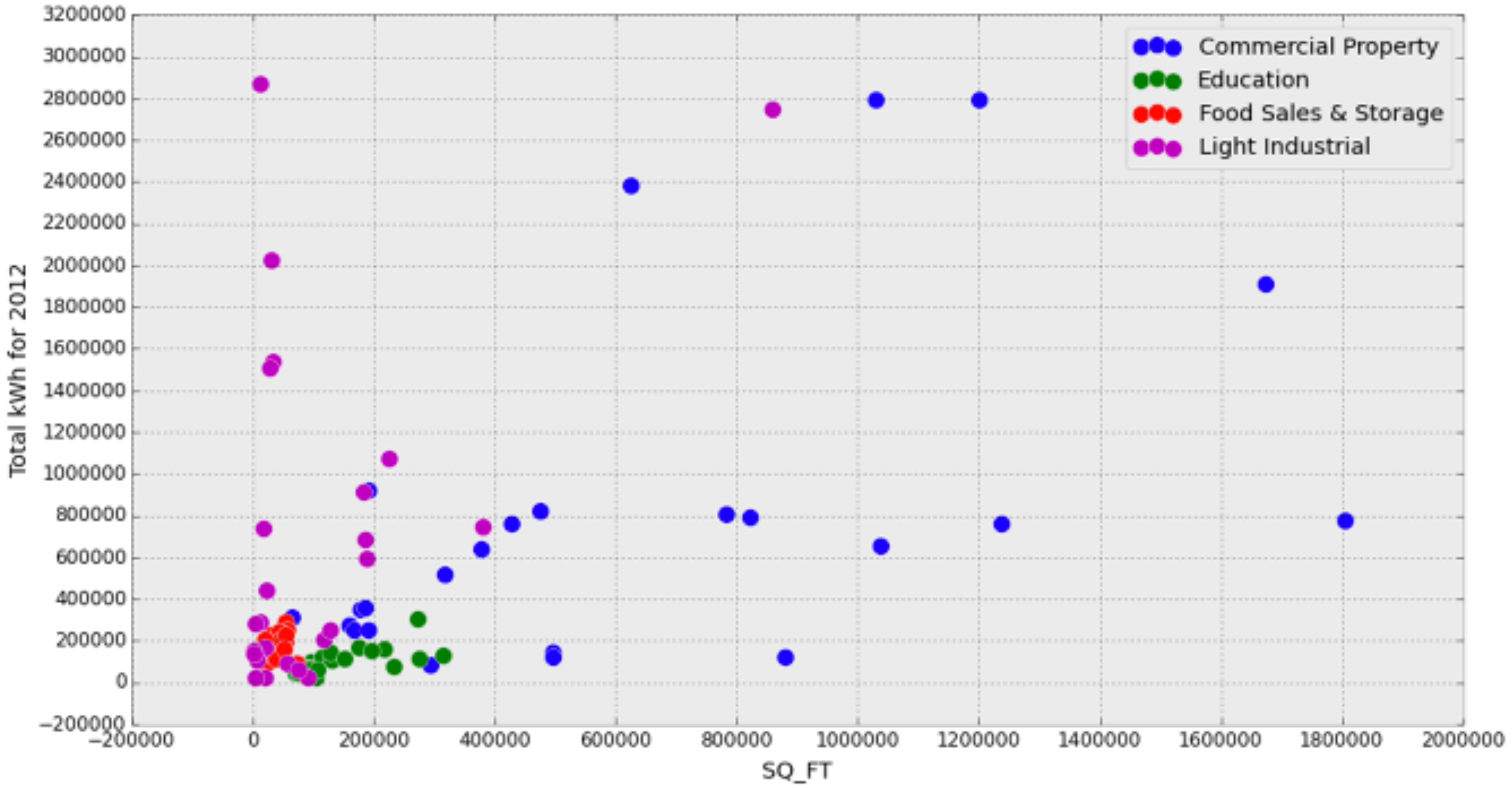
Building Energy Database

EnerNOC dataset of 100 buildings from 2012





		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		



ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH

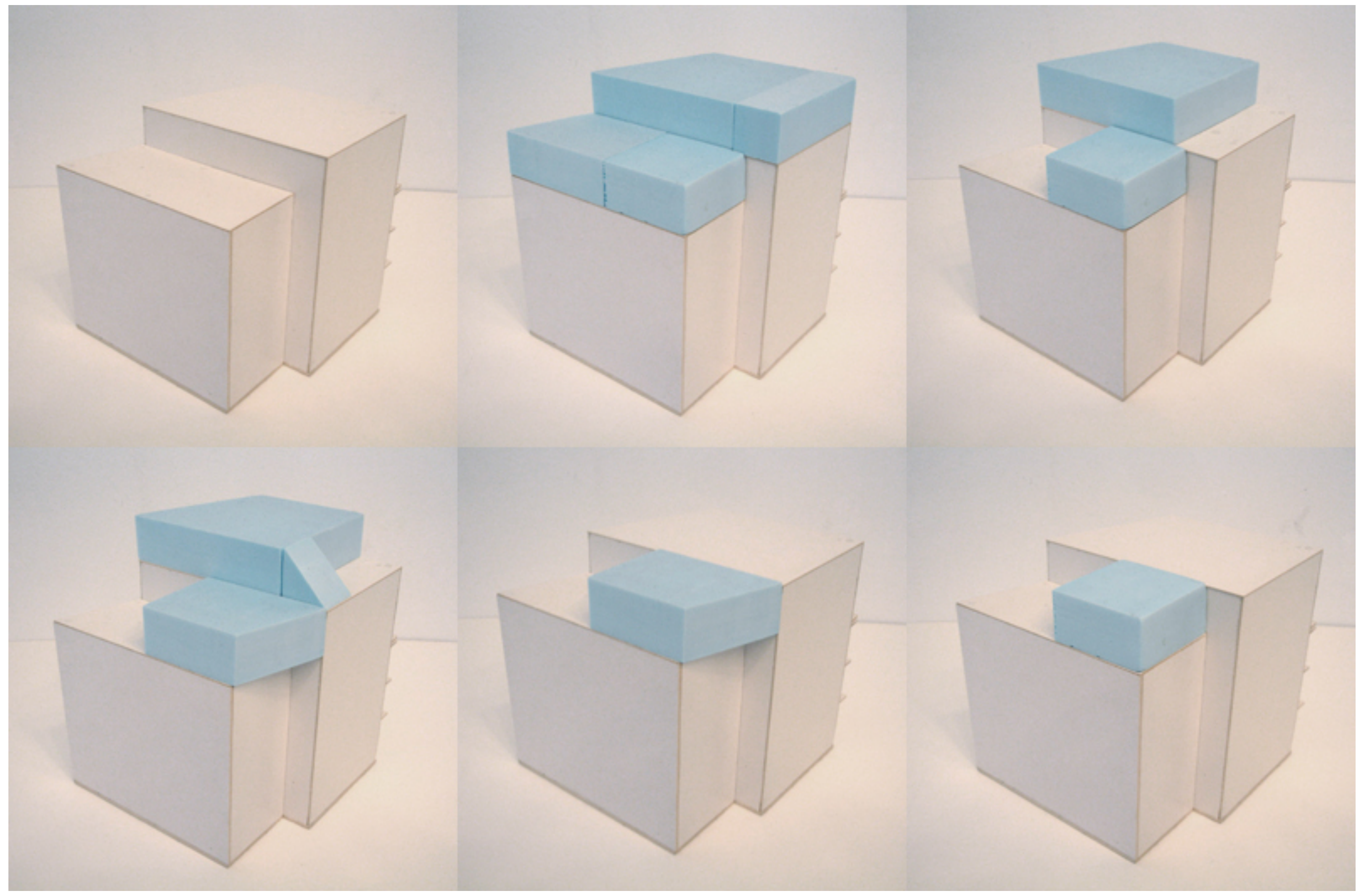
iA Chair of Information Architecture

Source: http://www.datadrivenbuilding.org/100-EnerNOC-Commercial-Buildings

Other examples or projects?

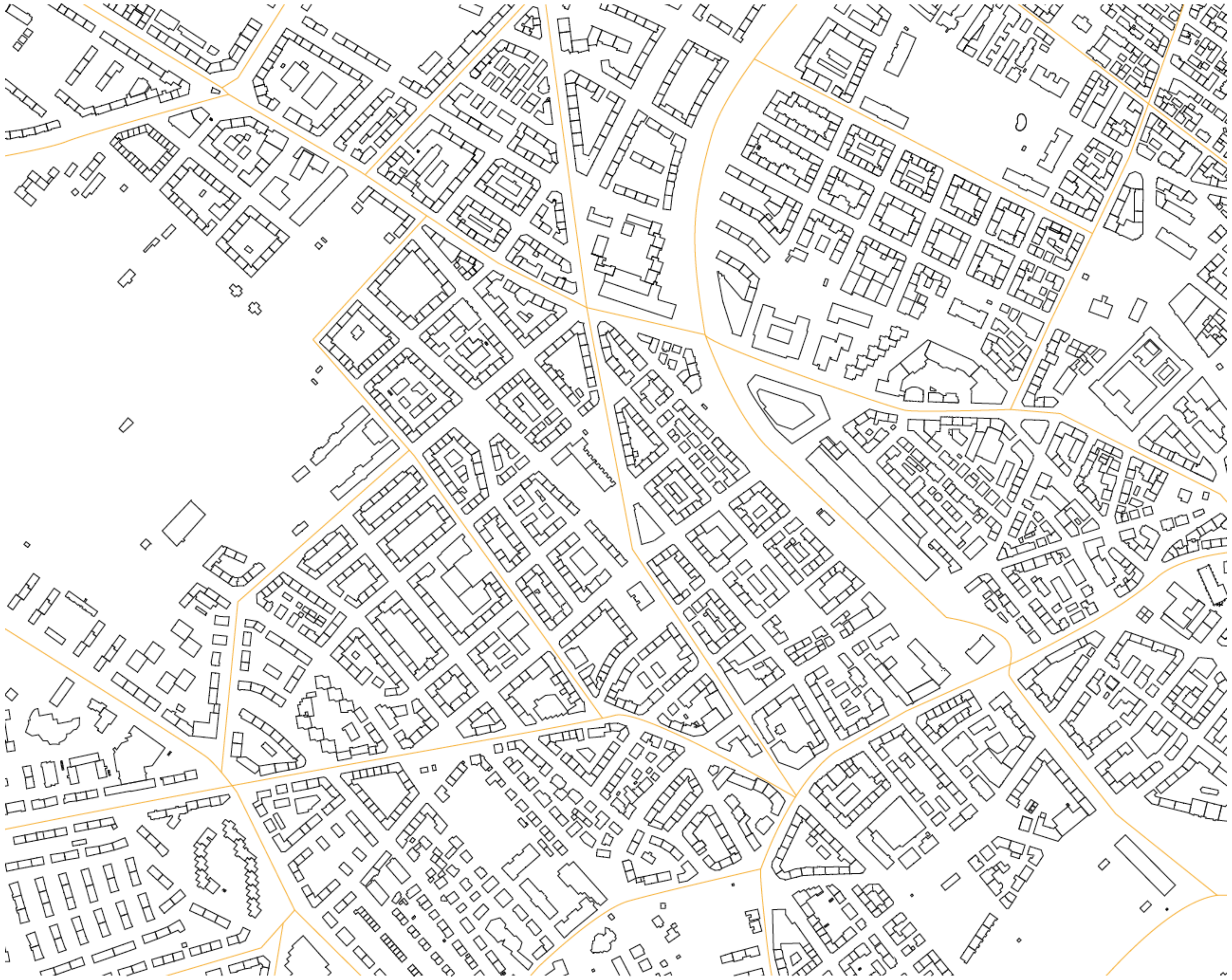
EXAMPLES		Analysis, visualization, interpretation	
		Manual	Automated
Data Source	Manual		
	Automated		

Modeling methods – Other Perspectives, Abstraction and Highlights



ESUM: Analyzing trade-offs between the *Energy* and Social performance of Urban Morphologies

- Investigate relationships between energy performance of urban structures and the perception of a city by its inhabitants
- Case studies of urban morphology in Zurich and Weimar



DARCH

Chair of Information Architecture

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ESUM: Initial analysis and results

USING GEO STATISTICAL ANALYSIS TO DETECT SIMILARITIES IN EMOTIONAL RESPONSES OF URBAN WALKERS TO URBAN SPACE

REINHARD KÖNIG[1], SVEN SCHNEIDER[2]
and
IHAB HAMZI[3], XIN LI[4], MARTIN BIELIK[2], GERHARD SCHMITT[1], DIRK DONATH[2]
[1]ETH Zürich Switzerland, [2]Bauhaus-University Weimar Germany, [3]An-Najah National University Palestine, [4]Wuhan University China,

SUMMARY

- Empirical pre-study on measuring emotional arousal in public space
- Study time: 14 - 22 October 2013, Place: Oerlikon, Zurich, Sitzerland
- N= 14 Participants
- Method of measurement: sensor-wristband Smartband
- **Significant similarities of emotional response to same spatial situation among different participants**

Introduction

In the following we investigate the impact of urban form on emotional response. The presented examination is a preliminary-study aimed at developing a method which allows us to measure human emotions in various spatial configurations and to define the spatial statistical methods needed to analyze the collected data.

Data Collection

To directly measure human emotional responses to real world environmental stimuli we use a sensor-wristband (Smartband), developed by Bodymonitor in combination with a GPS-tracker.
The Smartband [Papastefanou 2013] offers a mobile means of measuring skin conductivity (SC) as well as skin temperature. The collected data are analyzed with respect to four activations: Negative arousal, positive arousal, balanced, and retraction.

Statistical Analysis

The aim of the analysis is to answer following questions: What is the probability that the distribution of the emotion values occurs by chance? If the emotion values occur not by chance, where are the areas of clustering? Where is the clustering of high and low values?
Based on Geospatial analysis methods (Getis-Ord General G, and Getis-Ord hot-spot analysis Gi* (Getis 1991, Mitchell 2005) significant clustering of positive and negative arousal values at specific distances was proved and the null rejected

Conclusion

The statistical analysis has proven that different participants had similar emotional response to the same spatial situation regardless factors such as weather or the individual's respective mood or personal background.

The allocation of positive and negative arousal clusters could by characterized by sequence of spaces with high emotional arousal (positive, negative or both) followed be spaces with balanced emotional activation. A possible explanation is that the emotional response is not caused primarily by a certain spatial situation but rather by a changing sequence of spaces. If there are no changes in the environment our emotions are more or less in a kind of balanced or retrieving mode.

Fig 1
Locations of emotional arousal are colour-coded by Z-scores. Locations surrounded by locations with similarly positive or negative values are shown in red or blue.

Fig 2
(left) 3d model of study environment + walk path and analysis point (every 5 meter), (middle) 2d Isovist set on analysis points looking in walk directioin at restricted angle 60°, (right) 3d isovist

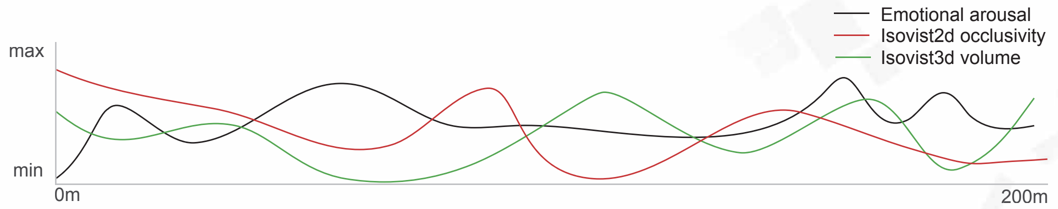


Fig 3
shows a sequence of changes of measures for simple 2D, 3D-Isovists and emotional arousal

Outlook

Assuming that further studies validate our findings, we will then search for correlations between emotional response and spatial measures derived from computational spatial analysis like isovists (2D and 3D Fig 2), visibility graphs (Hillier 1996), daylight and street-network analysis as well as combinations of these.
As the results of this pre-study suggest, changes in a sequence of spaces along a path may be a reason for positive or negative arousal (Fig 3). Consequently, it would be valuable not only to test static values for certain points of view but to investigate how the measured responses change along the path.

References

Getis, A: 1991, Spatial interaction and spatial autocorrelation: a cross product approach, Environment and Planning A 23:1269 – 1277.
Hillier, B: 1996, Space is the machine: a configurational theory of architecture, Cambridge University Press
Mitchell, A: 2005, The ESRI Guide to GIS Analysis, Volume 2, ESRI Press, Redlands, Ca, USA
Papastefanou J Bodymonitor: Sensing emotions. http://www.bodymonitor.de/Projekte %28neu%29. Accessed 26 Nov 2013

Initial Case Study: Zürich Orelikon

Figure 3: The participants were given this map showing the path (red dotted line) they were asked to follow (base map from OpenStreetMap). At places marked with an arrow they took a photo and waited for 5 seconds to enable the Smartband to take adequate measurements. The photographs on the right side show some characteristic street views.

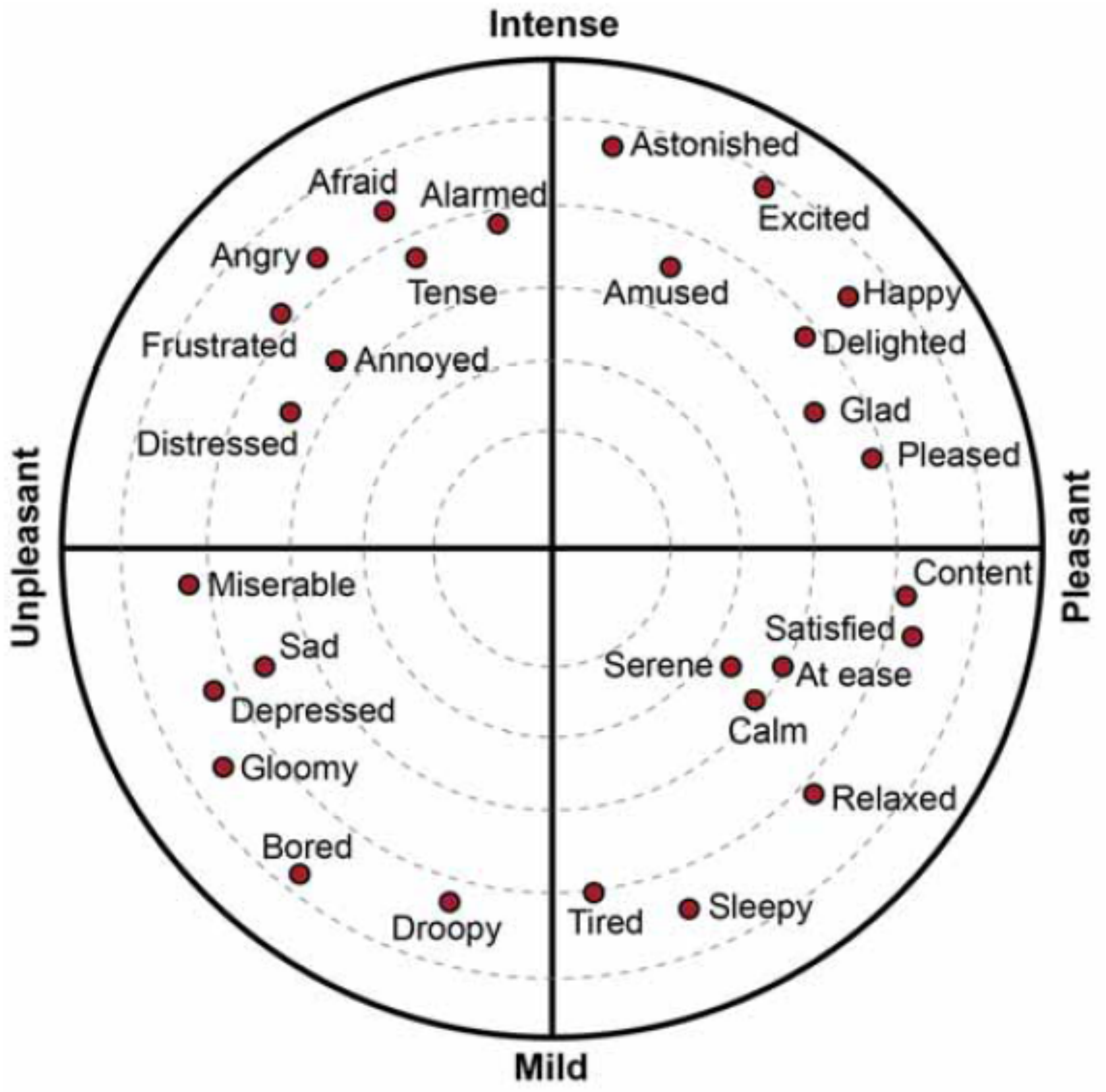
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

“Measuring” Perception


EMPATICA
www.empatica.com
info@empatica.com

Offices:
ITA: Via Alessi 8 - 20122 Milano (MI)
US: 41 Casa Way - San Francisco 94123 (CA)

E3 Technical Specifications
Revision 006 23/12/2013



Initial Findings: perception qualities

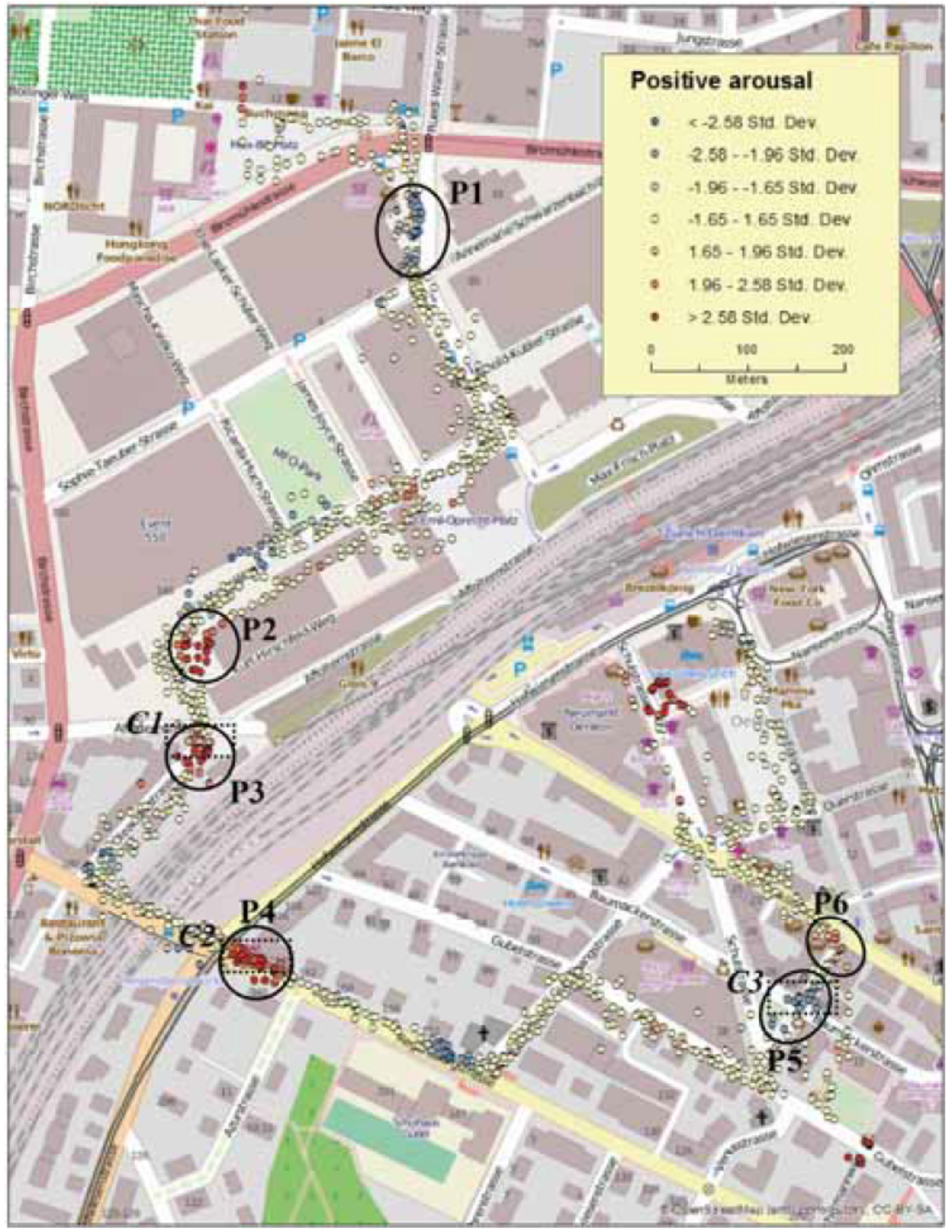


Figure 6: Locations of positive emotional arousal are colour-coded by Z-scores. Locations surrounded by locations with similar high or low values are shown in red or blue. The locations P1 – P6 are used for a photo comparison below. The areas C1 – C3 are used to investigate the number of different test persons inside a cluster.

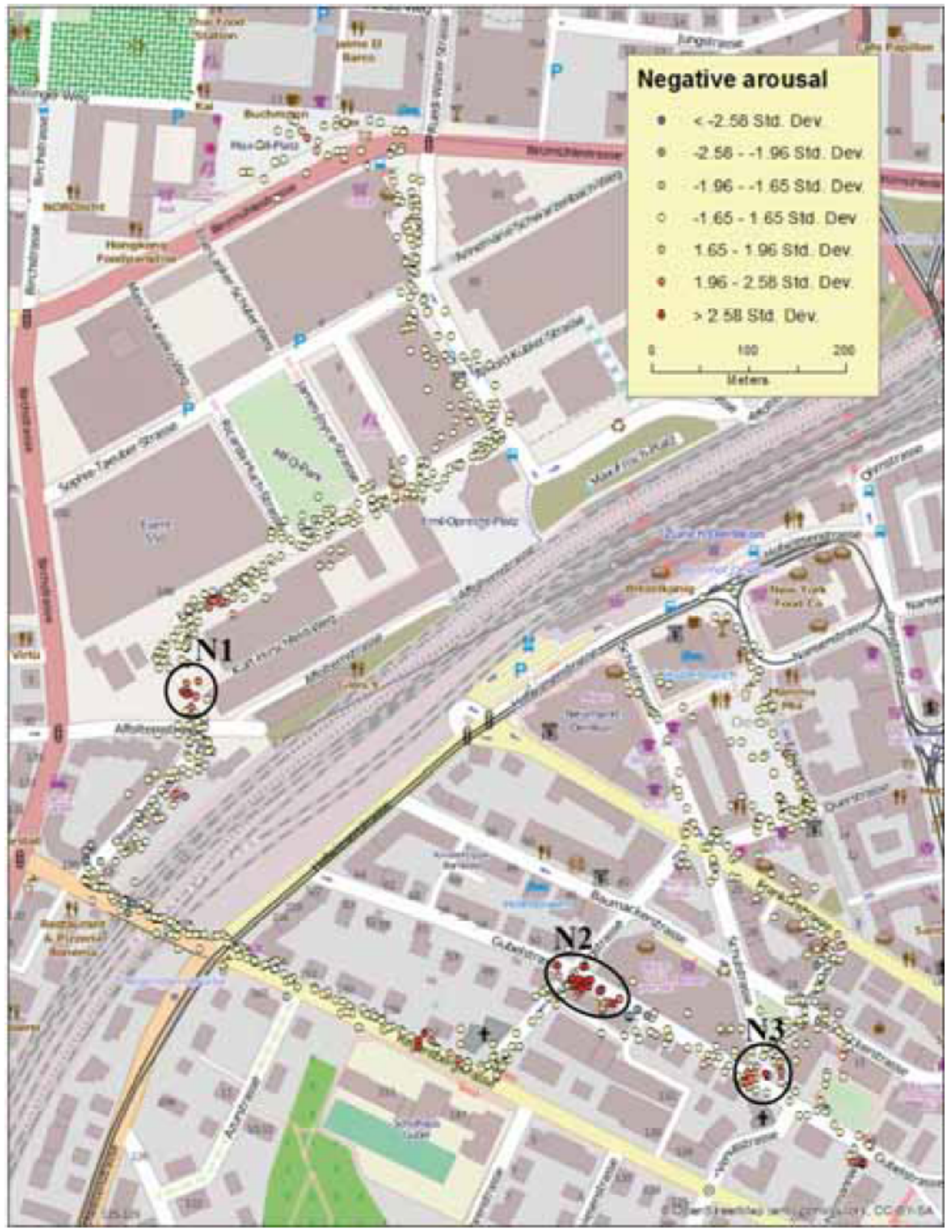


Figure 7: Locations of negative emotional arousal are colour-coded by Z-scores. Locations surrounded by locations with similarly high or low values are shown in red or blue. The locations N1 – N3 are used for a photo comparison below.

Initial Findings: 2D Isovist Analysis

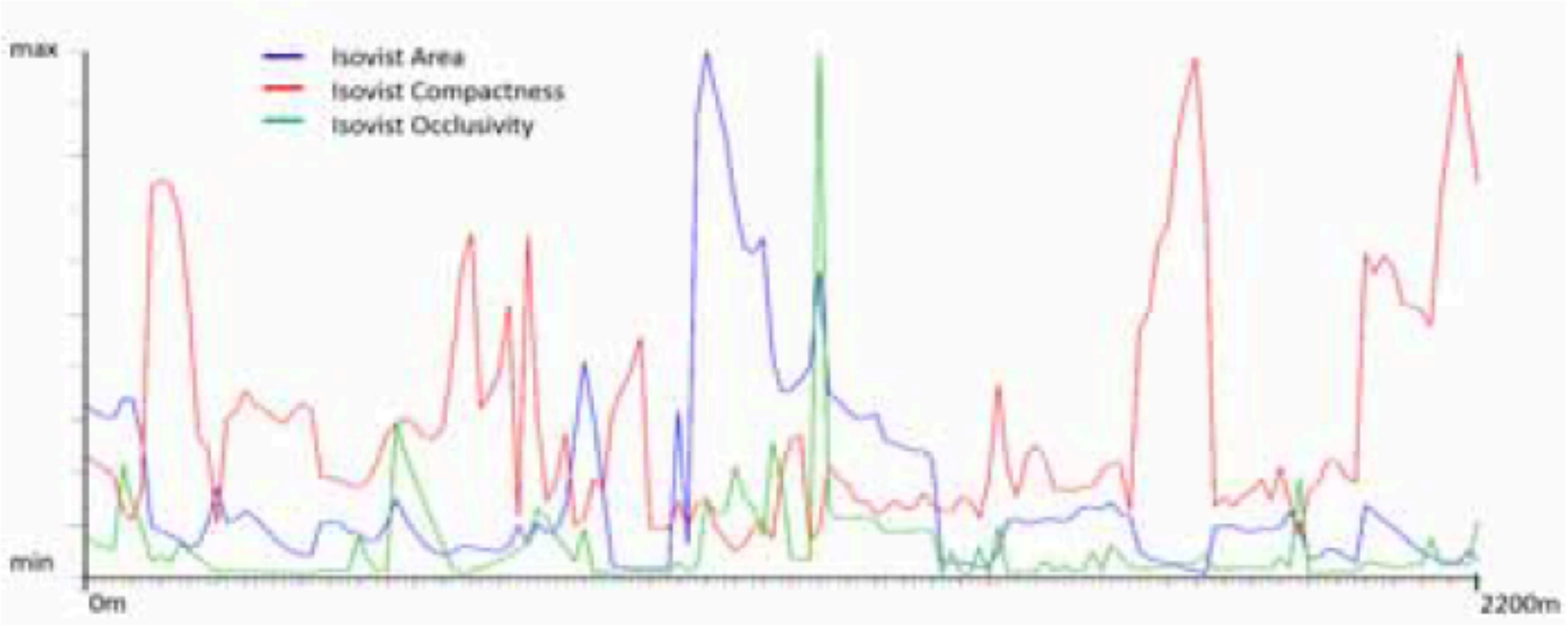


Figure 11, top: Isovists along the path. Bottom: sequence of isovist properties (area, compactness, occlusivity) along the path.

Intermediate Conclusions

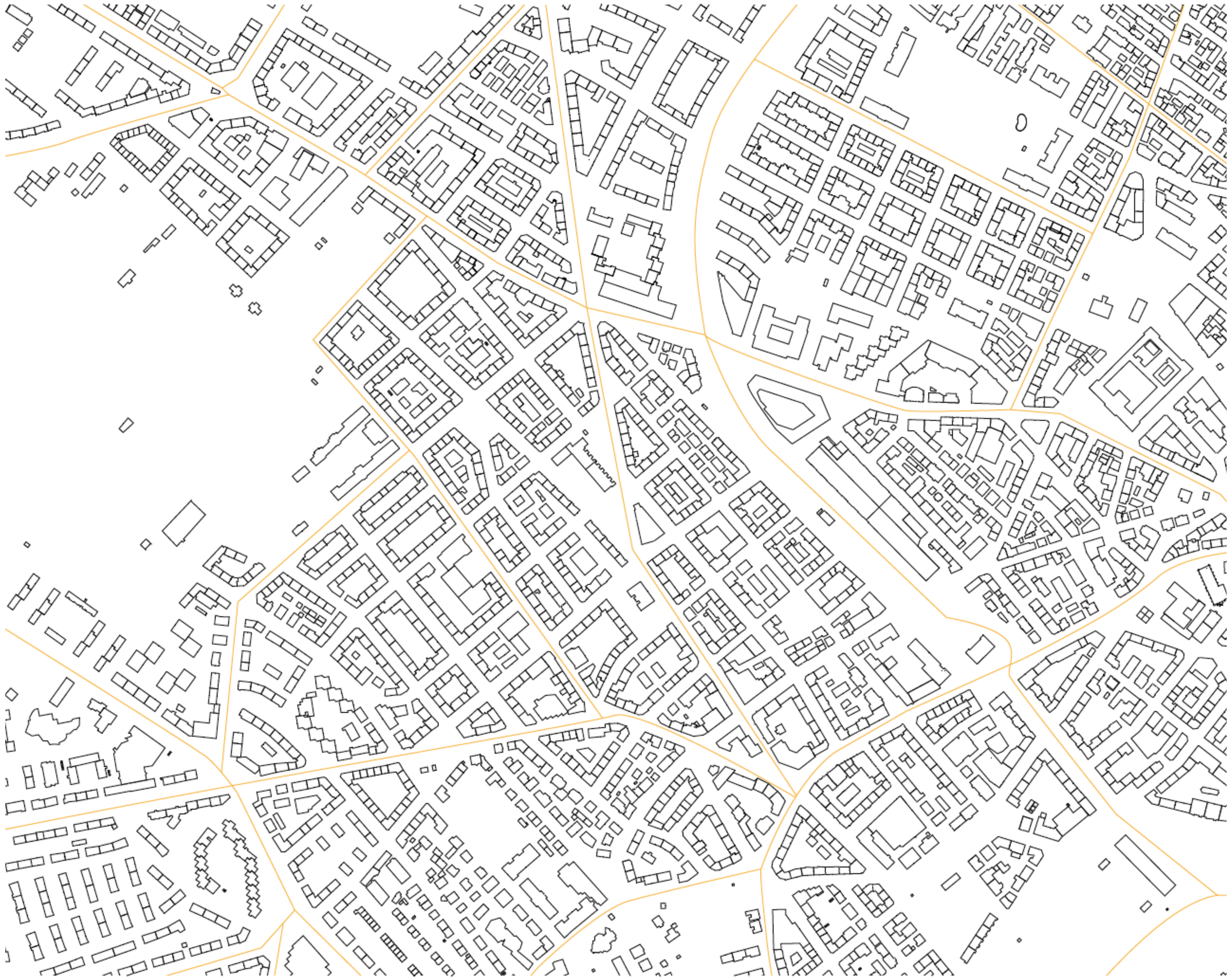
- A Geographically Weighted Regression (GWR) was used to correlate the measured perception and the 2D isovist analysis
- The GWR showed an initial indication for negative response
- However we need a larger sample size, and additional measurements to determine other externalities which may impact perception in the urban environment

Negative Response

Positive Response

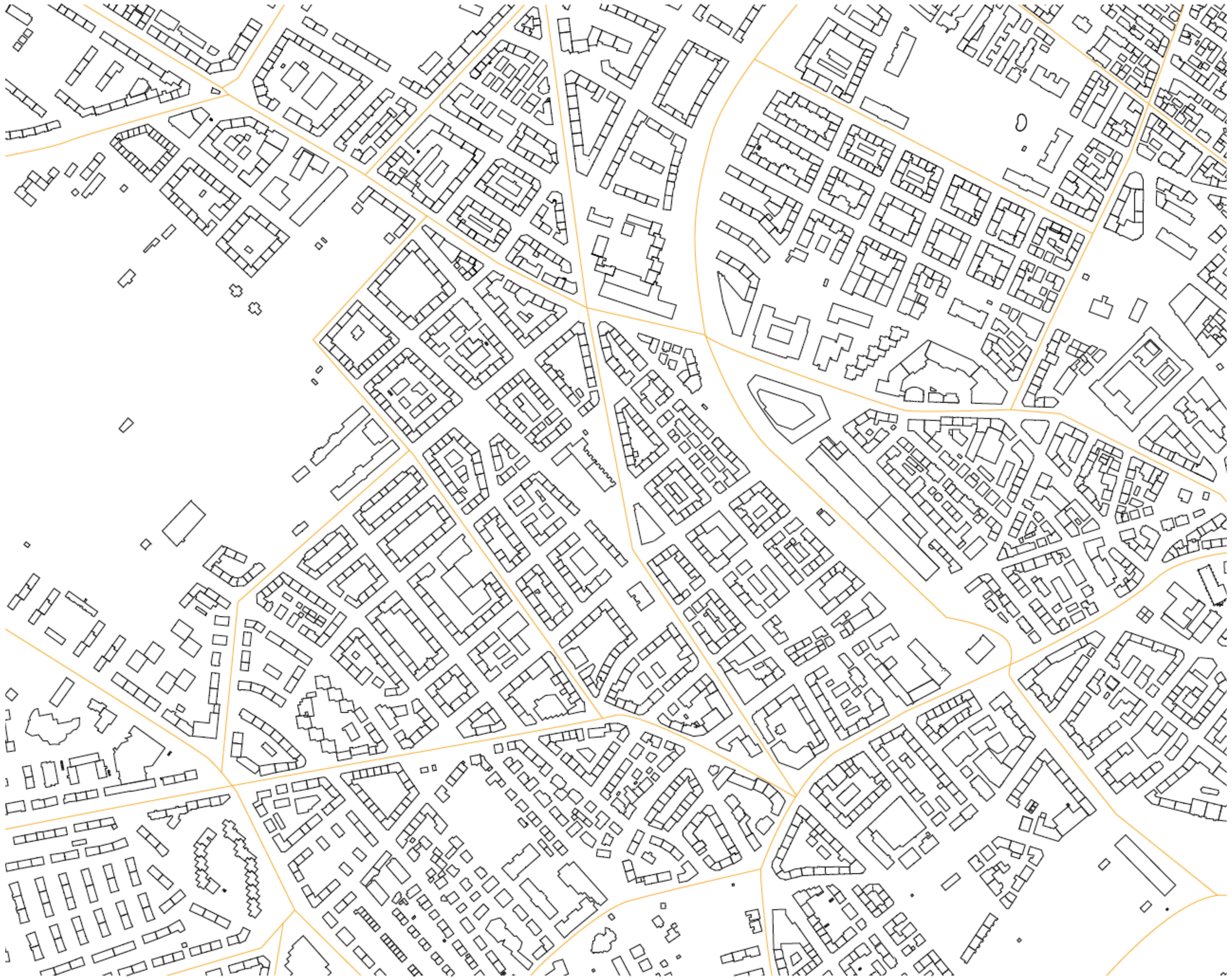
Next steps

- Change case study location to Alt-Wiedikon
- Perform a 3D Isovist Analysis
- Include additional dimensions in the correlation model
  - Sound, light, dust, temp., etc.
- Gather more participants
- Add the energy analysis of the built environment



How will you be involved?

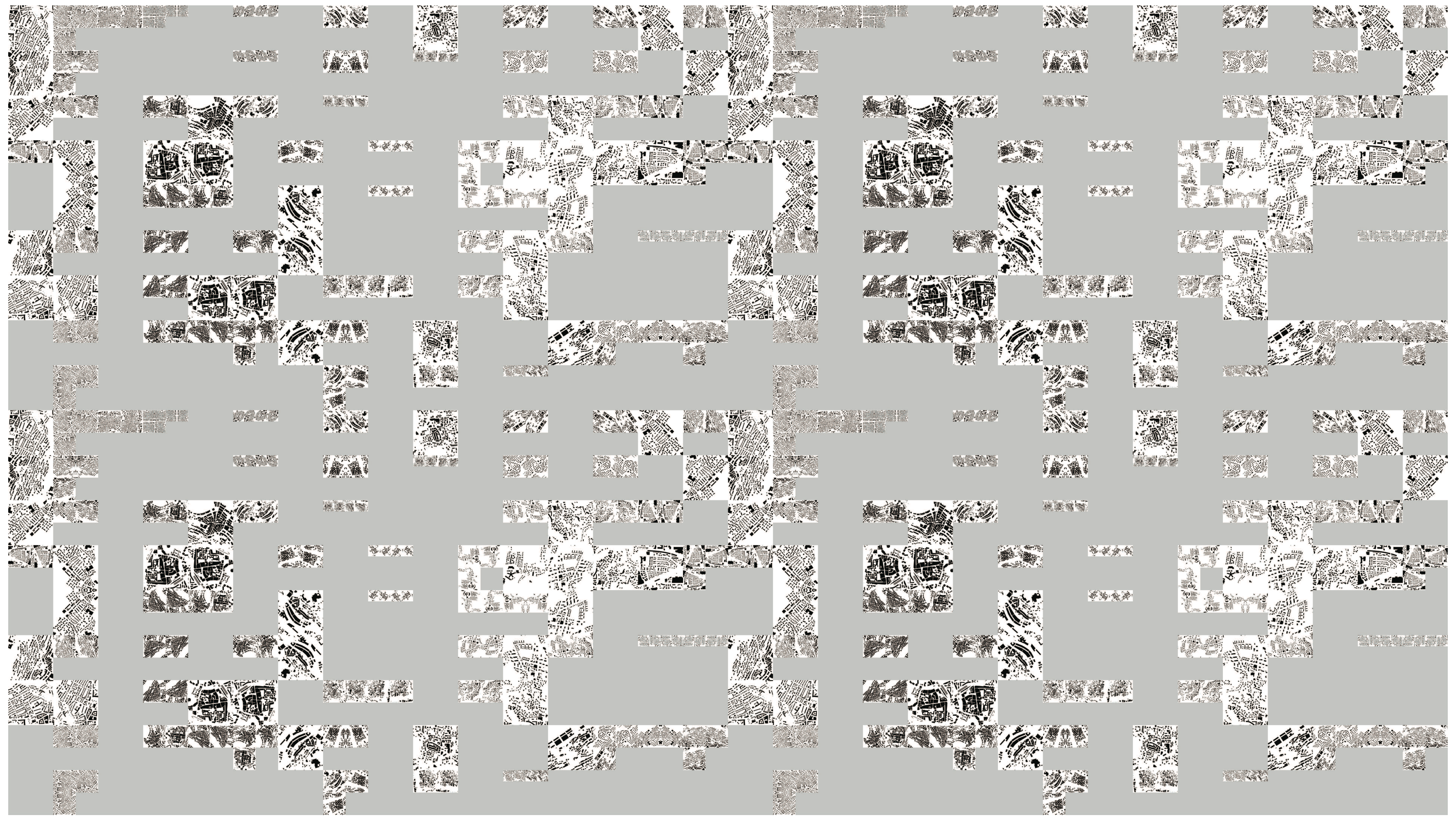
Creative Data Mining Semester Project!



Manual data input, automated analysis

BLOCK 1		Analysis, visualization, interpretation	
		Manual	Automated Data Mining/ machine learning
Data Source	Manual Hand-drawn sketches		
	Automated		

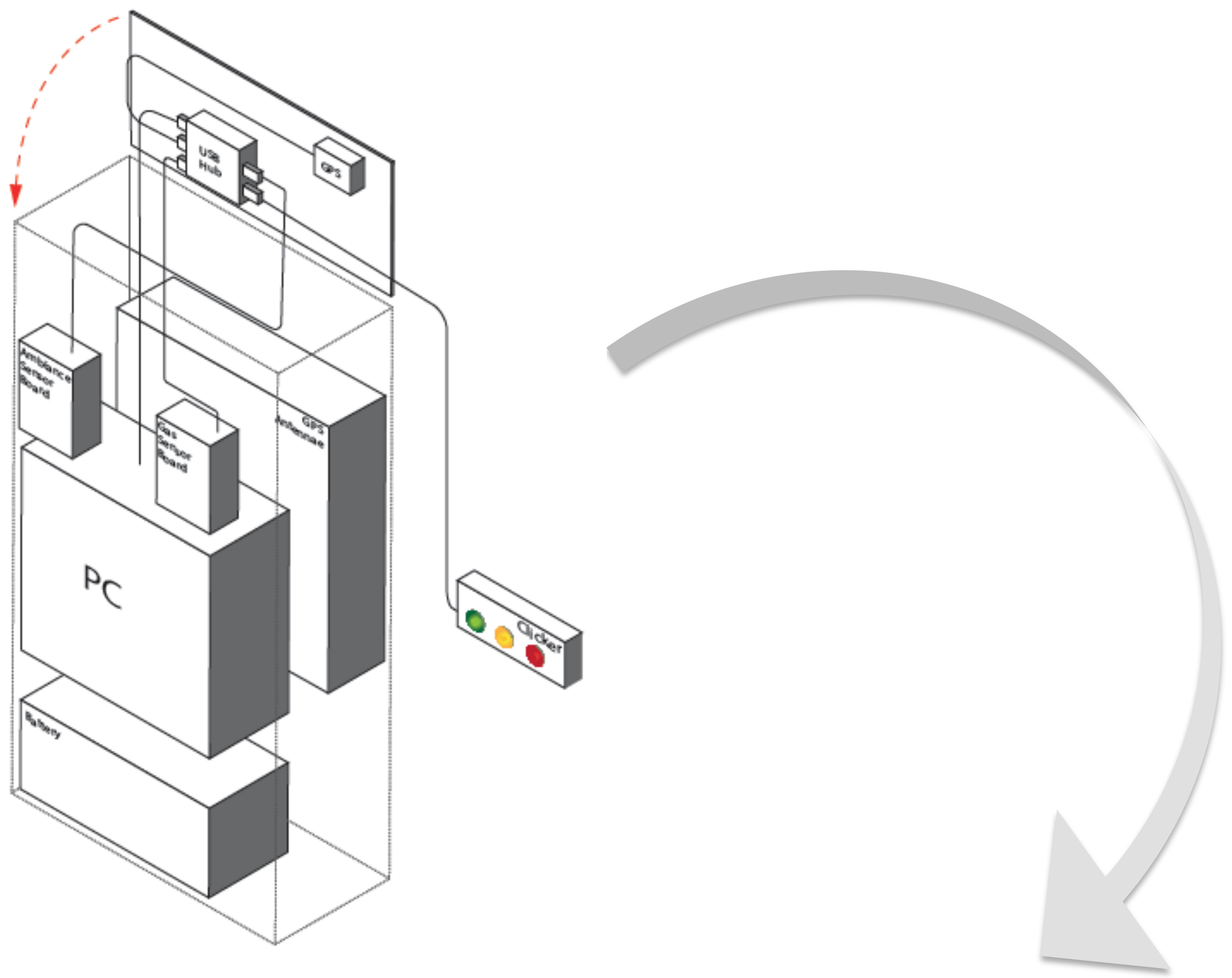
Studio

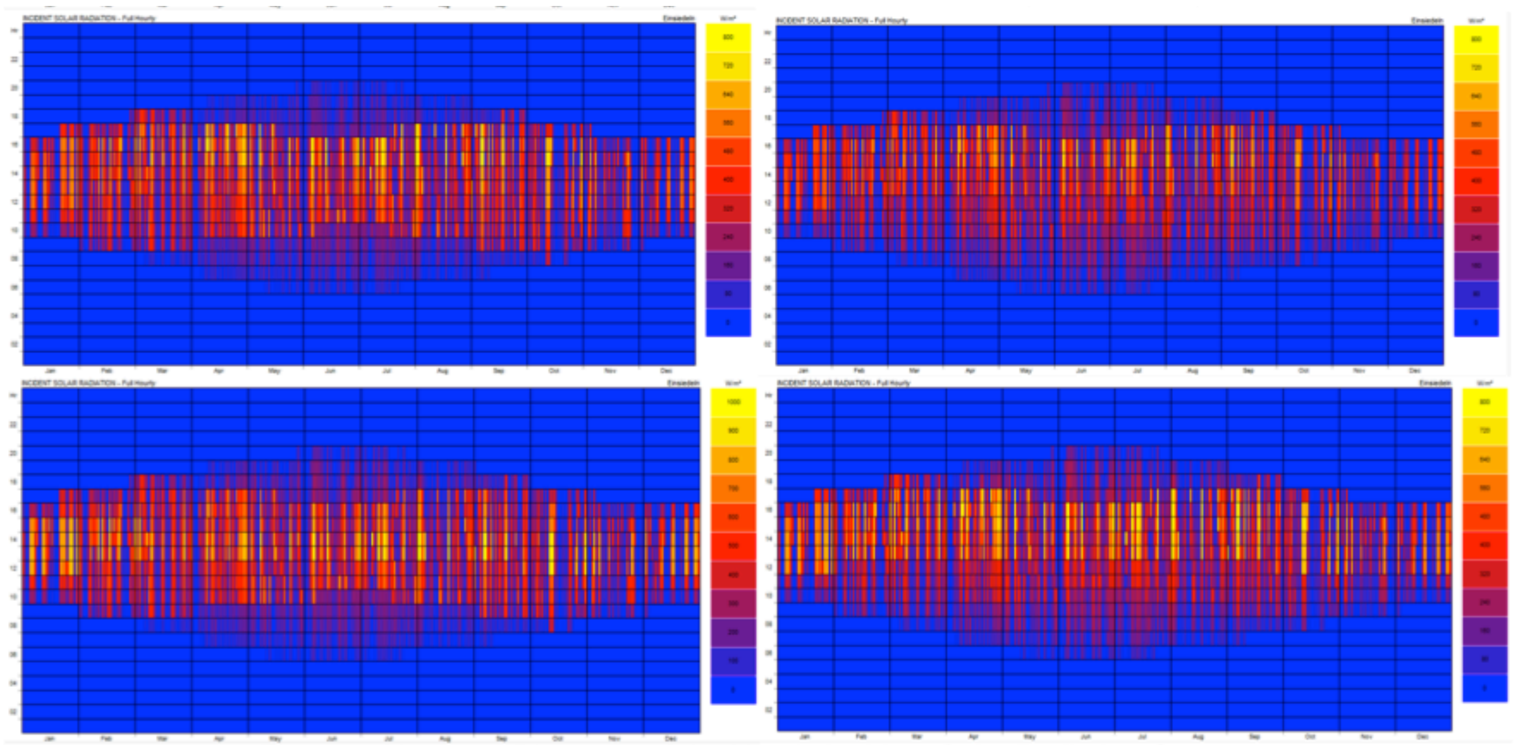


Final Project from Moritz Berchtold, Creative Data Mining FS2015

Automated data collection, manual data analysis

BLOCK 2		Analysis, visualization, interpretation	
		Manual Time-series and geo-referenced analysis	Automated
Data Source	Manual		
	Automated Sensor data		





Automated data collection, automated analysis & visualization

<table>
<tr><td colspan="2" rowspan="2">POTENTIALLY</td><td colspan="2">Analysis, visualization, interpretation</td></tr>
<tr><td>Manual

Geo-referenced Analysis</td><td>Automated</td></tr>
<tr><td rowspan="2">Data Source</td><td>Manual</td><td></td><td></td></tr>
<tr><td>Automated

Sensor data</td><td></td><td></td></tr>
</table>

You are Encouraged to Explore Additional Techniques!

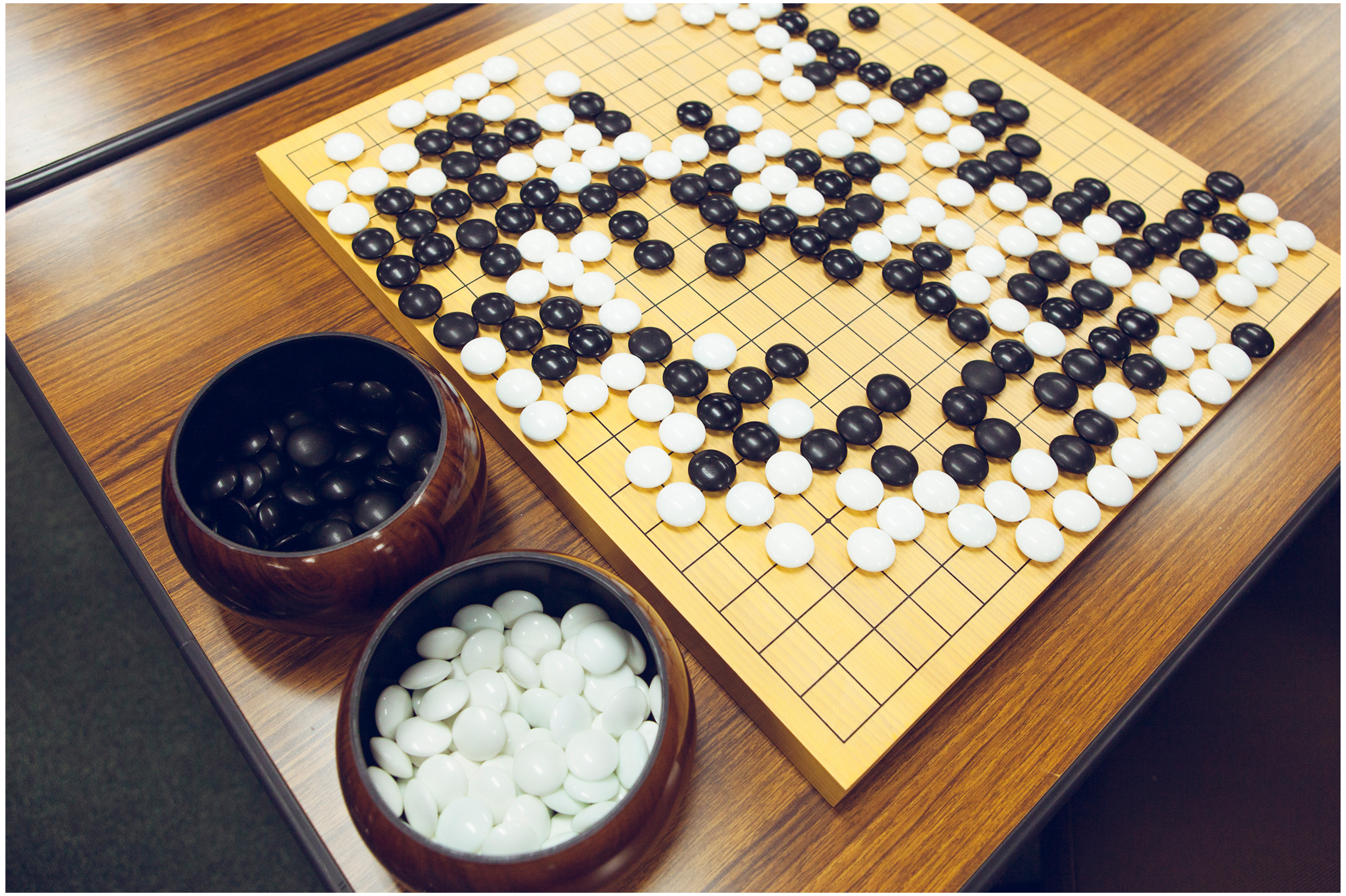
"I expect you all to be independent, innovative, critical thinkers who will do exactly as I say!"

Learning Objectives

- Gain confidence to work with different types of data and relevant tools
- Learn how to select appropriate data sources for your projects
- Learn relevant analysis and interpretation techniques
- Set up and run an urban sensing "experiment"

"I expect you all to be independent, innovative, critical thinkers who will do exactly as I say!"

YOUR EXPECTATIONS?



Schedule

Creative Data Mining
Intuitively Analysing Design Ideas

The goal of this course is to introduce various data mining techniques for design and urban planning applications. Students will learn how to select relevant data sources and collect their own data using a “sensor backpack”. Various methods will be applied to a common project to evaluate the predominant influencing factors of the urban environment on our perceptual experiences. A select neighborhood in the city will be used as a case study. Final results will be presented in the last class.

The course will start with an initial overview to data mining and the relevant mathematics as well as an introduction to the programming tool (RStudio). Then students will learn how to use and interpret results from a machine-learning tool to cluster self-made design sketches, which automatically generate qualitative collages. Finally, students will collect data using a “sensor backpack” with environmental sensors such as noise, temperature, illuminance, and air particulates. Students will also generate the data for perceptual quality in this neighborhood through time-stamped and geo-referenced surveys and biofeedback wristbands. Students will be given a work-flow to collect, process, analyze and interpret this data which may be used in their final projects.

Where
HIT H 12

Supervision
Danielle Griego griego@arch.ethz.ch
Matthias Standfest standfest@arch.ethz.ch

22.02.2016	**Course Introduction** Introduce data-mining techniques and case study
29.02.2016	**Introduction to the Environment** Introduction to R Studio and clustering
07.03.2016	**From analog to digital analysis** Use hand-drawn sketched to auto-generated collages
14.03.2016	**Seminar week (No lecture)**
21.03.2016	**Analysis and interpretation I** Evaluate auto-generated collages
28.03.2016	**Holiday (No lecture)**
04.04.2016	**Time-series data analysis and Urban Planning** Introduction to time-series analysis
11.04.2016	**Data collection with sensor backpack** Collect data and introduce workflows
18.04.2016	**Holiday (No lecture)**
25.04.2016	**Analysis and interpretation II** Evaluate sensor backpack data
02.05.2016	**Q&A Feedback Workshop** Finalise semester projects
09.05.2016	**Final iA critique** Combined critique with the other iA courses (14:00 - 16:00)

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

Exercises 40% (documentations)
Final Presentation 40% (Final project)
Attendance 20%

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

Requirements

Creative Data Mining
Intuitively Analysing Design Ideas

The goal of this course is to introduce various data mining techniques for design and urban planning applications. Students will learn how to select relevant data sources and collect their own data using a "sensor backpack". Various methods will be applied to a common project to evaluate the predominant influencing factors of the urban environment on our perceptual experiences. A select neighborhood in the city will be used as a case study. Final results will be presented in the last class.

The course will start with an initial overview to data mining and the relevant mathematics as well as an introduction to the programming tool (RStudio). Then students will learn how to use and interpret results from a machine-learning tool to cluster self-made design sketches, which automatically generate qualitative collages. Finally, students will collect data using a "sensor backpack" with environmental sensors such as noise, temperature, illuminance, and air particulates. Students will also generate the data for perceptual quality in this neighborhood through time-stamped and geo-referenced surveys and biofeedback wristbands. Students will be given a work-flow to collect, process, analyze and interpret this data which may be used in their final projects.

Where
HIT H 12

Supervision

Danielle Griego griego@arch.ethz.ch
Matthias Standfest standfest@arch.ethz.ch

22.02.2016	**Course Introduction** Introduce data-mining techniques and case study
29.02.2016	**Introduction to the Environment** Introduction to R Studio and clustering
07.03.2016	**From analog to digital analysis** Use hand-drawn sketched to auto-generated collages
14.03.2016	**Seminar week (No lecture)**
21.03.2016	**Analysis and interpretation I** Evaluate auto-generated collages
28.03.2016	**Holiday (No lecture)**
04.04.2016	**Time-series data analysis and Urban Planning** Introduction to time-series analysis
11.04.2016	**Data collection with sensor backpack** Collect data and introduce workflows
18.04.2016	**Holiday (No lecture)**
25.04.2016	**Analysis and interpretation II** Evaluate sensor backpack data
02.05.2016	**Q&A Feedback Workshop** Finalise semester projects
09.05.2016	**Final iA critique** Combined critique with the other iA courses (14:00 - 16:00)

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

Exercises 40% (documentations)
Final Presentation 40% (Final project)
Attendance 20%

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

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DARCH iA Chair of Information Architecture

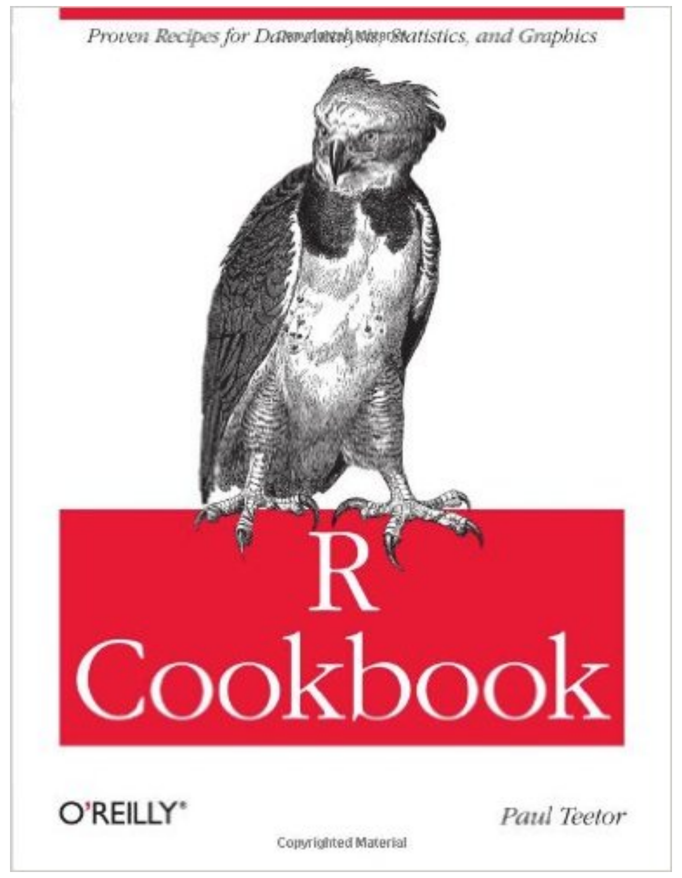
Homework

1. Install R from http://cran.r-project.org/

2. Install RStudio from http://www.rstudio.com/products/rstudio/download/

3. Research other examples of urban data mining and make 3 slides about the most interesting project/application/resarch group(s) that you find. Will be presented at the beginning of next lecture.

Resources for the course



Video Tutorial

- https://www.youtube.com/watch?v=aB7MVpUSHjo

Interactive (try the 1st one!):

- http://tryr.codeschool.com/levels/1/challenges/1
- https://www.datacamp.com/courses/free-introduction-to-r

References:

- http://www.tutorialspoint.com/r/
- http://www.cyclismo.org/tutorial/R/

Additional:

- http://shiny.rstudio.com/tutorial/
- http://sape.inf.usi.ch/quick-reference/ggplot2
- http://tutorials.iq.harvard.edu/R/Rgraphics/Rgraphics.html

“Science without philosophy is blind, and philosophy without science is paralyzed”

(Paul Cilliers, Complexity and Postmodernism)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich