#### Creative Data Mining Spring 2018





# What we'll cover today

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

# ETHZUICH DARCH IA Chair of Information Architecture



What is Data Mining?





It is an exploratory and iterative process



# ETHZUICH DARCH iA Chair of Information Architecture



#### What is machine learning?





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





How can data mining be creative?



# ETHZUICH DARCH iA Chair of Information Architecture

#### What do we want to know?





How can data mining be creative?

#### Domain specific data source(s)





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

#### Is the data usable?





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

# ETHZÜRICH DARCH IA Chair of Information Architecture



Types of analysis, visualization & interpretation: Time Series and georeferenced data visualization



# ETHZUICH DARCH iA Chair of Information Architecture



Types of analysis, visualization & interpretation: Hierarchical clustering



# ETHZUICCH DARCH IA Chair of Information Architecture





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



#### Types of analysis, visualization & interpretation: SOM- Self organizing Maps



SOM clustering map of participants (indicated by numbers)





Integrating the "creative aspects" of data mining

		Analysis visualization & interpretation	
		Manual	Automated
Source	Manual		
Data S	Automated		

### ETHZUICH DARCH iA Chair of Information Architecture

#### Domain specific data source(s)



Elaborating on the traditional architectural process

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

# ETHZÜRICH DARCH IA Chair of Information Architecture



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



Process taught in previous semesters

		Analysis visualization & interpretation	
		Manual	Automated
			Machine Learning: SOM
	Manual		
ource	Hand-drawn sketches		
Data S	Automated		

# ETHZUICCH DARCH IA Chair of Information Architecture



Final Project from Moritz Berchtold, Creative Data Mining FS2015





Time-series & geo-referenced data visualizations

		Analysis visualization & interpretation		
		Manual	Automated	
		Time-series & geo- referenced data visualization		
	Manual			
Source				
ata S	Automated			
Ď	Sensor data			

# ETHZUICCH DARCH IA Chair of Information Architecture



ESUM project experimental equipment set up and data analysis techniques



#### Machine Learning

		Analysis, visualization & interpretation	
		Manual	Automated
			Machine Learning Techniques
	Manual		
Source			
ata S	Automated		
Õ	Sensor data		

# ETHZÜRICH DARCH 1A Chair of Information Architecture





ESUM project experimental equipment set up and data analysis techniques





#### Data Mining for Architects and Urban Planners?

A few examples

# ETHZUICH DARCH iA Chair of Information Architecture



# National data collection project

Geo-referenced sensor data visualization

		Analysis, visualization, interpretation	
		Manual	Automated
ource	Manual		
Data S	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





# Chicago OpenGrid

#### Geo-referenced data visualization

		Analysis, visualization, interpretation	
		Manual	Automated
ource	Manual		
Data S	Automated		



# ETHZÜRICH DARCH 1A Chair of Information Architecture



#### Newcastle University Urban Observatory

Geo-referenced and time-series data visualization



### ETHZÜRICH DARCH 1A Chair of Information Architecture





# 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





View by: HOUR DAY WEEK



		Analysis, visua interpreta	
		Manual	A
source	Manual		
Data S	Automated		

Cities BANGALORE BOSTON GENEVA RIO DE JANEIRO SAN FRANCISCO SHANGHAI SINGAPORE

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





# Data Canvas project output

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



# ETHZÜRICH DARCH 1A Chair of Information Architecture

		Analysis, visualization,						Analysis, vi interpr	sualization etation
	_	Manual	Automated					Manual	Automa
a	Manual					9	Manual		
Data Source	Automated					Data Sourc	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  $\rho$  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

# ETHZUICICH DARCH IA Chair of Information Architecture



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 Simulaiton International.





### Other Examples?

		Analysis visualization & interpretation	
		Manual Automated	
Source	Manual		
Data S	Automated		

### ETHZUICCH DARCH 1A Chair of Information Architecture



### Course Structure



ETHZUICH DARCH iA Chair of Information Architecture





#### Course Schedule

What to Expect

#### Creative Data Mining

The Creative data mining course aims to provide aspirants a handson experience on machine learning (ML) tools and techniques for data processing and analysis. Since future technologies increasingly rely upon the ML, urban systems and architecture shall adopt it and aspirant should learn creative ways to apply ML to better understand urban systems. The course covers a wider range ML techniques including supervised and unsupervised learning methods for data analysis and pattern recognition that help to better understand urban system for improving urban life.

All methods taught in the course will be applied to a common project to evaluate various dynamics of the urban environment. 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. A selected neighborhood in the city of Zurich will be used as the case study and each student will present the findings of their research question in a final project.

Additionally, there are two of non-architectural skills the participants can develop during this course. First is an introduction to programming where at a minimum they can successfully copy and paste code-snippets to customize the computational tools presented in the course. Second, how clustering methods like PCA or K-Means could be applied in an architectural context.

ETHZURICH DARCH IA Chair of Information Architecture



19.02.2018	Introduction to the course
26.02.2018	Concepts of python & programming
05.03.2018	Fundamentals of supervised learning
12.03.2018	Concept of python & programming II
19.03.2018	Seminar Week
26.03.2018	Supervised learning problem solving in python
09.04.2018	Fundamentals of unsupervised learning
23.04.2018	Unsupervised learning problem solving in python
30.04.2018	Review data and initial proposals for final projects
07.05.2018	Project proposal discussions
14.05.2018	Final Critique

### Semester Project

Something to start thinking about

- Formulate 1-2 specific question(s) of interest to you 1.
- 2.
- 3.

  - Include a time series and/or clustering analysis
- Summarize your results 4.
- **Conclusions & lessons learned** 5.
- Include motivation and references 6.

# ETHZÜRICH DARCH 1A Chair of Information Architecture

State your hypothesis/expected outcome based on supporting literature (minimum one source) your expertise, and intuition

Answer that question through your analysis, for this:

Select the best available data sources for your question (min. of 2 data sources)

Show a clear conclusion, does your analysis answer your question(s)?



# 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







### Short discussion

#### Your expectations?







### Homework

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

- 1.
- 2.
- 3. make 2 slides about the most interesting





Install Spyder from https://pythonhosted.org/spyder/

Research other examples of urban data mining and project/application/research group(s) that you find. This will be presented at the beginning of next lecture



e python



#### Resources for the course

**Course Material Posted to:** 

#### Tutorials:

- http://www.informatics.indiana.edu/rocha/academics/ibic/lab1/Python%20review.pdf

References:

- "A Byte of Python" https://python.swaroopch.com/

# ETHZUICH DARCH 1A Chair of Information Architecture



#### "Science without philosophy is blind, and philosophy without science is paralyzed"

Questions?

(Paul Cilliers, Complexity and Postmodernism)

# ETHZÜRICH DARCH IA Chair of Information Architecture

Lecture 1: Introduction

