



muHVT: Case based computational geometry modeling toolkit using R

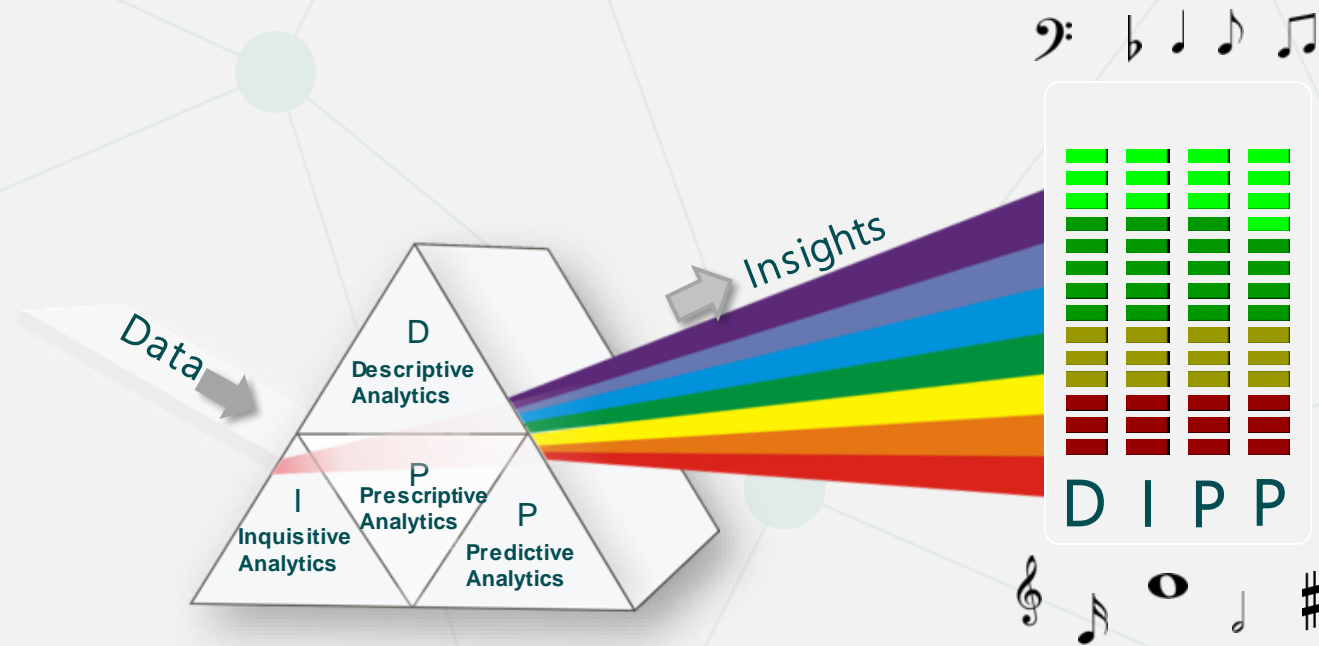
Zubin Dowlaty, Sangeet Moy Das, Shubhra Prakash
July 2020

Agenda

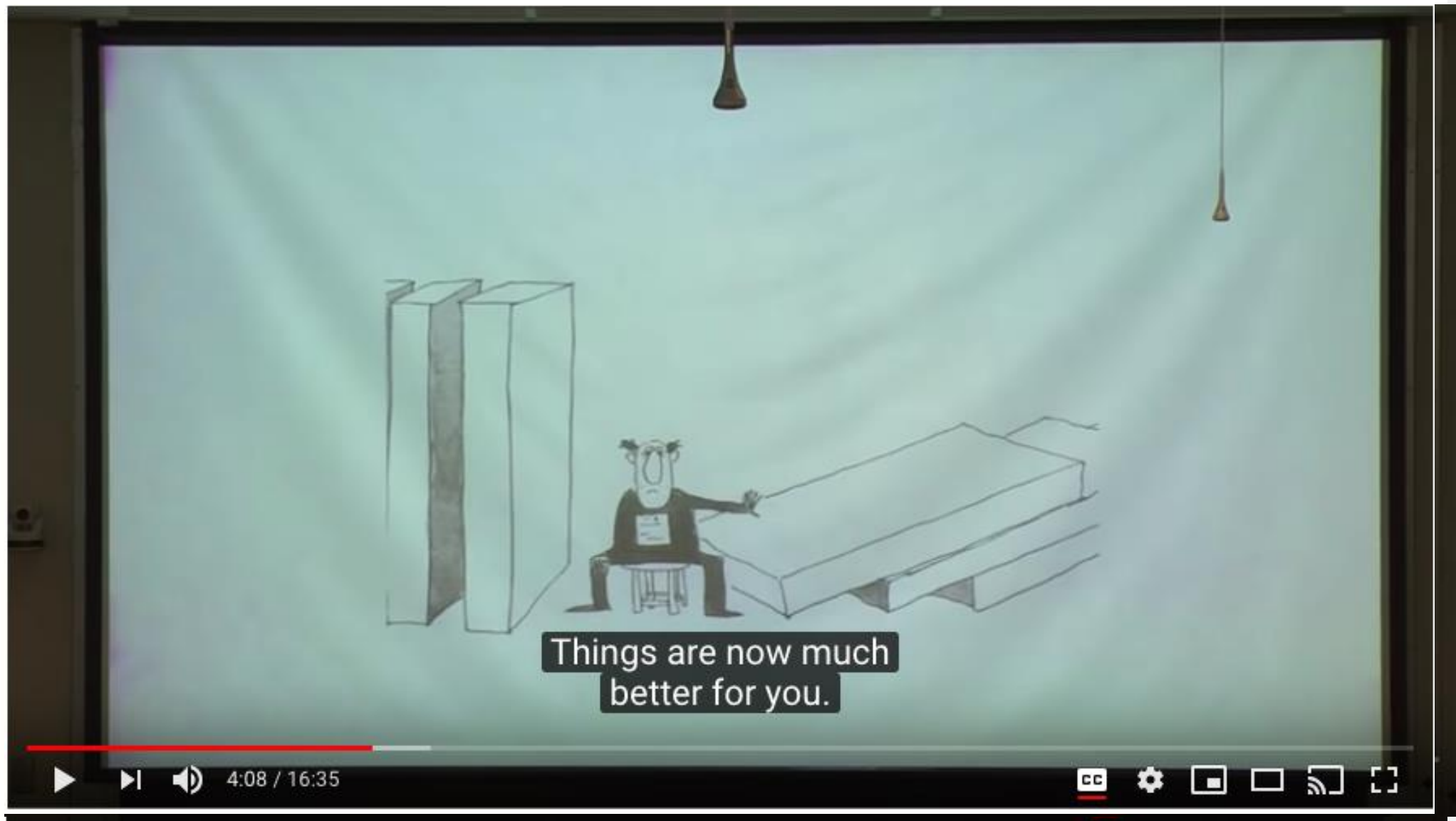
- ▶ **The Enterprise Requirement for Complexity Science**
- ▶ A Point of View - Representation of a Complex System
- ▶ Use Case: Intelligent Reporting
- ▶ Demo

DIPP: Prescriptive Analytics, “what should we do”, “our best action”, “evaluate this policy change”

Correct Perspective



Why Prescriptive Analytics for the Enterprise?



Things are now much
better for you.

Default view in Decision Sciences: Comfortable.. But Deceiving

- **Mechanical View / Complicated Mindset**
 - "Assembly Line", rules and recipes on components
 - Complicated: Computer, Phone, Car, Rocket, Clocks
 - The classic **Additive** regression model: $Y = a + B_{x1} + B_{x2} + e$
- Are the systems we manage so **tame**?
 - **Mandelbrot**, "Clouds are not spheres, mountains are not cones"
- The world is complex whether we like it or not – **Get Real**

1,090 views | Apr 20, 2019, 01:09pm

Our Entire AI Revolution Is Built On A Correlation House Of Cards



Kalev Leetaru Contributor

AI & Big Data

I write about the broad intersection of data and society.

f

t

in



Getty Images GETTY



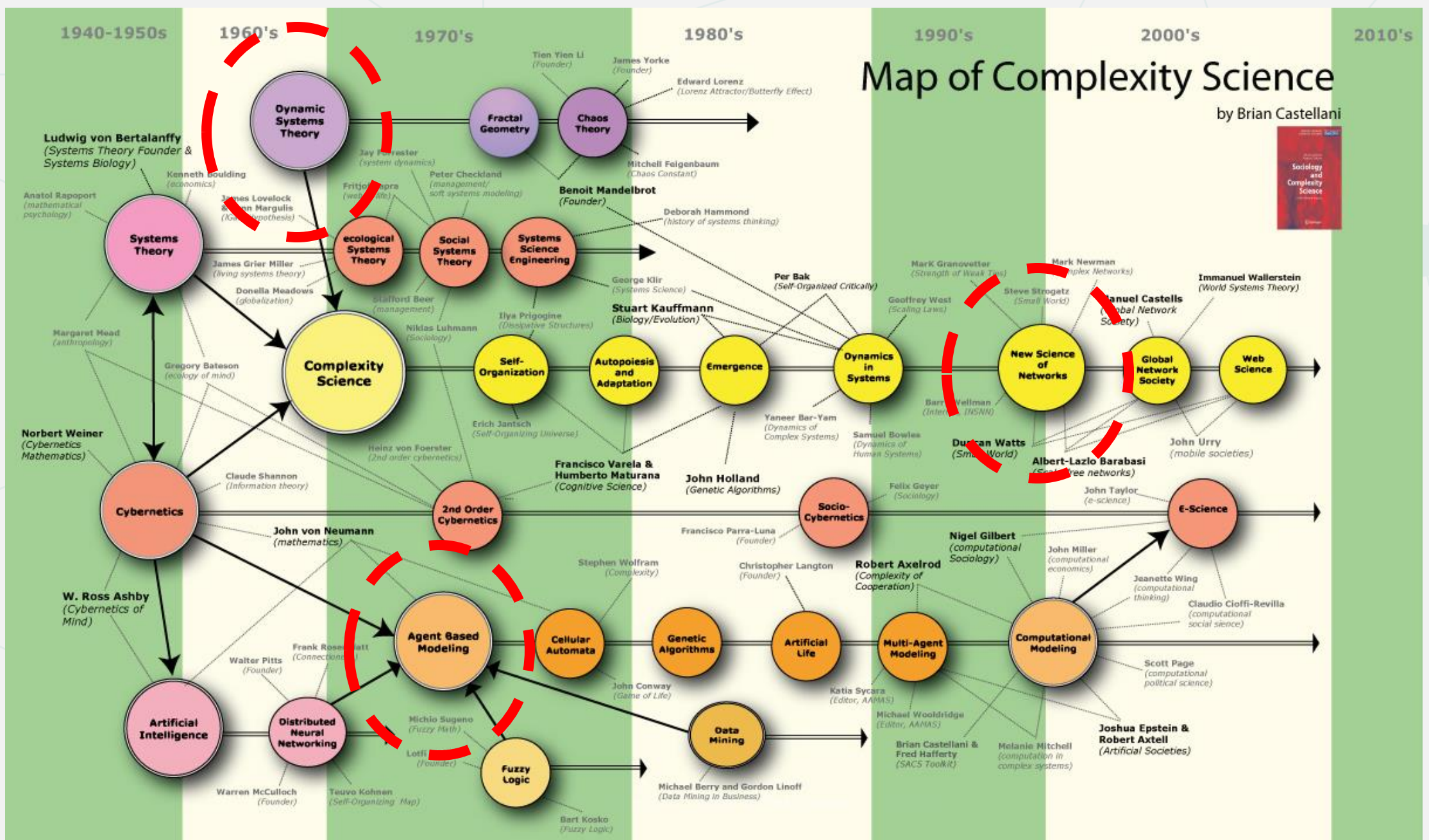
Examples of Complex Systems

Climate, Animals, Brains, Cells, Cities, Ecosystems, Supply Chains,
Traffic, Economies, Firms & Customers

Agents, Feedback, No Central Coordinator, Emergent, Self Organize, Learning



Complexity Science



First Principles: Complexity Science

- **Properties**

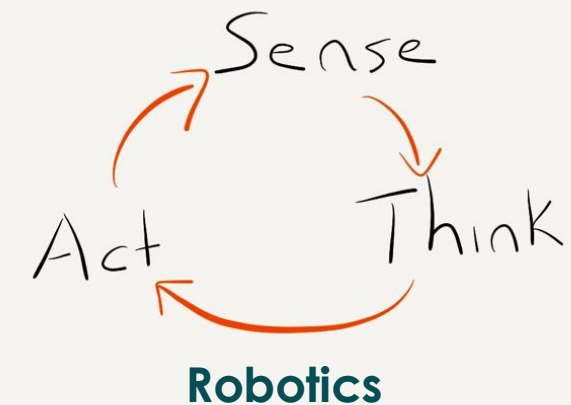
- Simple components or **Agents** (simple relative to the whole system)
- **Nonlinear Interactions** among components usually caused by **Feedback Loops**
- **No central control**
- **Emergent** behaviors
 - Hierarchical(**Layers**) & **Network** design patterns
 - **Information Processing**
 - Dynamics: Study of continuing changing structure and behavior – **STATES & Self Organization**
 - **Learning** / Evolutionary / Adaptation

- **Study of Organisms (Ecology)**

- Computation model inspiration = **Robotics** (Sense, Plan, Act)

Use Case: Intelligent Reporting

- Dealing with Networks (**organism**)
 - Possible **Unanticipated** Consequences and difficult to predict
 - **Novelty Detection** key characteristic
 - Shannon Entropy “Surprise”
- **Monitor** at the case level, **states and trajectories**
 - **Case** = profile of a contextual set of inter-dependent variables, a situation of the case-object
 - Start with a Robust **Sensing** Apparatus
 - Visualize states and trajectories of the case-object

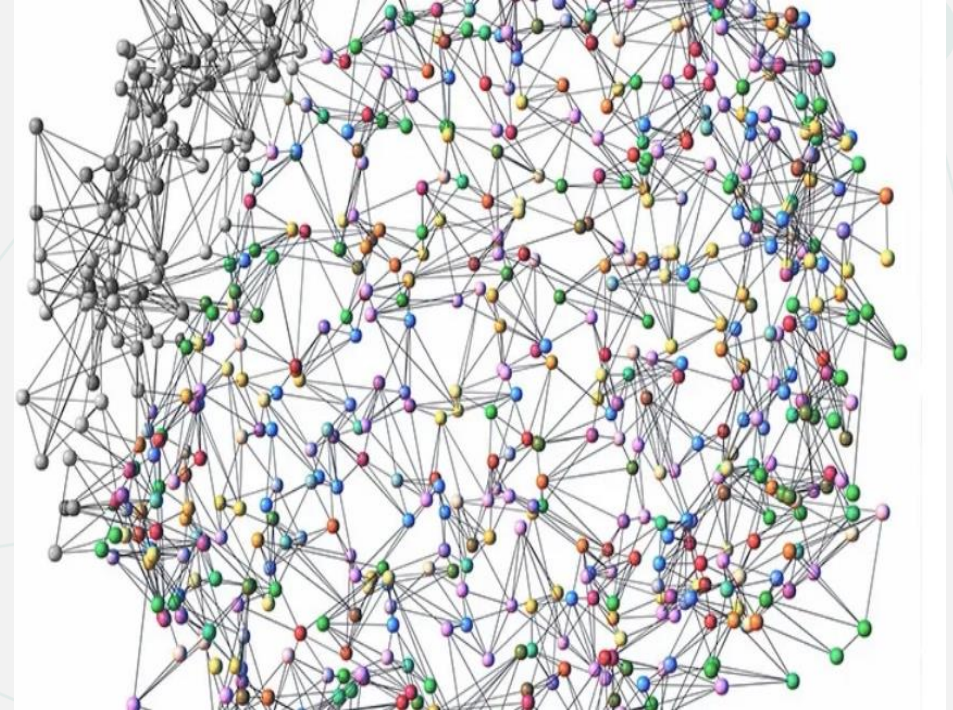


Agenda

- ▶ The Enterprise Requirement for Complexity Science
- ▶ **A Point of View - Representation of a Complex System**
- ▶ Use Case: Intelligent Reporting
- ▶ Demo

Why complex system **representation** an issue?

- The intricate nature of complex systems currently poses great difficulty in an unsupervised representation across different domains
- Some of these existing problems include:
 - Organized Simplicity
 - Disorganized Complexity
 - Organized Complexity
- Now, how do we tackle these problems?



Emergence in a Complex System

"These new problems and the future of the world depends on many of them, requires science to make a third great advance, an advance that must be even greater than the nineteenth-century conquest of problems of organized simplicity of the twentieth-century victory over problems of disorganized complexity. Science must, over the next 50 years, learn to deal with these problems of organized complexity."

- Warren Weaver, 1948

- This is what **emergence** is all about; the overall effect of the interactions between all the components of the "organic whole"
- These changing interaction leads to a **dynamic** emergent behavior over time

Case Based Modeling for a Complex System

- Multiple behaviors emerge in a complex system
- Behaviors are divided into discrete **cases** based on structural similarity giving a condensed representation of the system
- The above process are repeated in a **hierarchy** for a microscopic view into the sub-behaviors of the system
- This is achieved using the following techniques from our [muHVT](#) package on CRAN:
 - **Unsupervised learning** (Vector Quantization),
 - **Computational geometry** (Voronoi regions) and
 - **Multi-dimensional scaling** (Sammon's Projection)

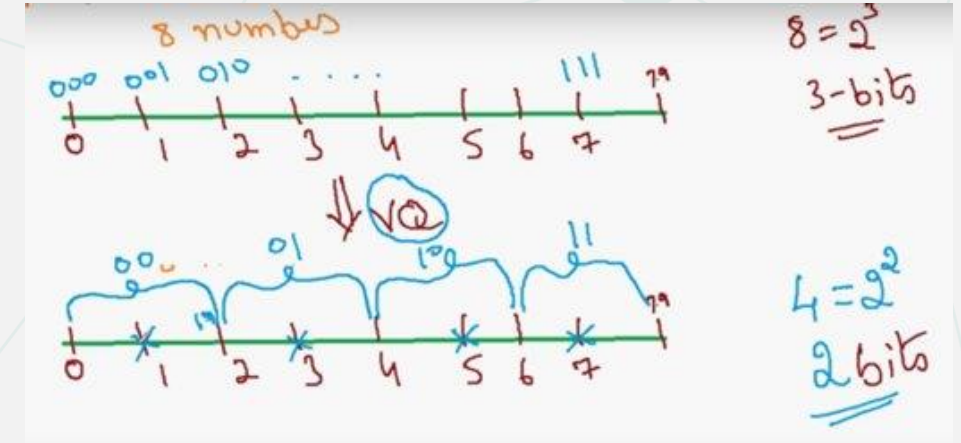
Agenda

- ▶ The Enterprise Requirement for Complexity Science
- ▶ A Point of View - Representation of a Complex System
- ▶ **Use Case: Intelligent Reporting**
- ▶ Demo

Vector Quantization

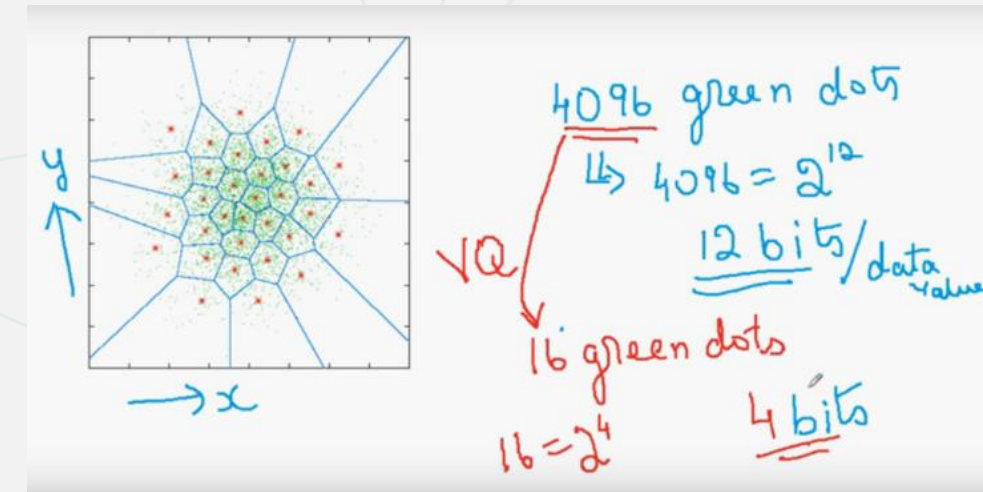
- VQ on a 1D Number Line

- Compression technique, tries to reduce the number of bits being used to encode the numbers



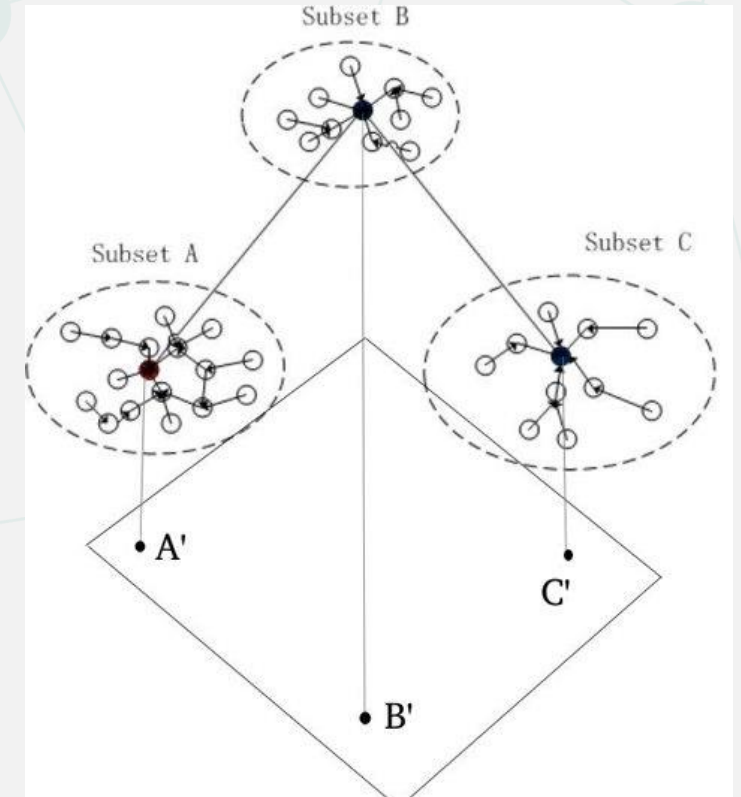
- VQ on a 2D Plane

- 1. The plane above contains 4096 green dots which can be represented in 12 bits per data value
 - 2. These 4096 dots have been replaced by 16 green dots which can be represented in 4 bits per data value

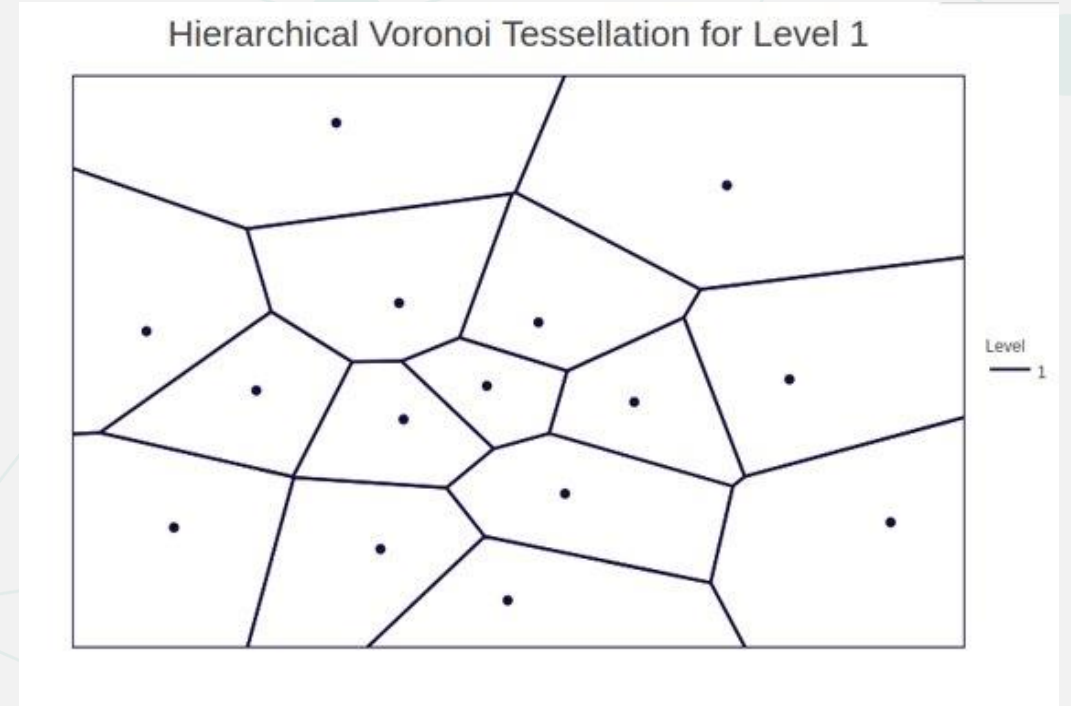
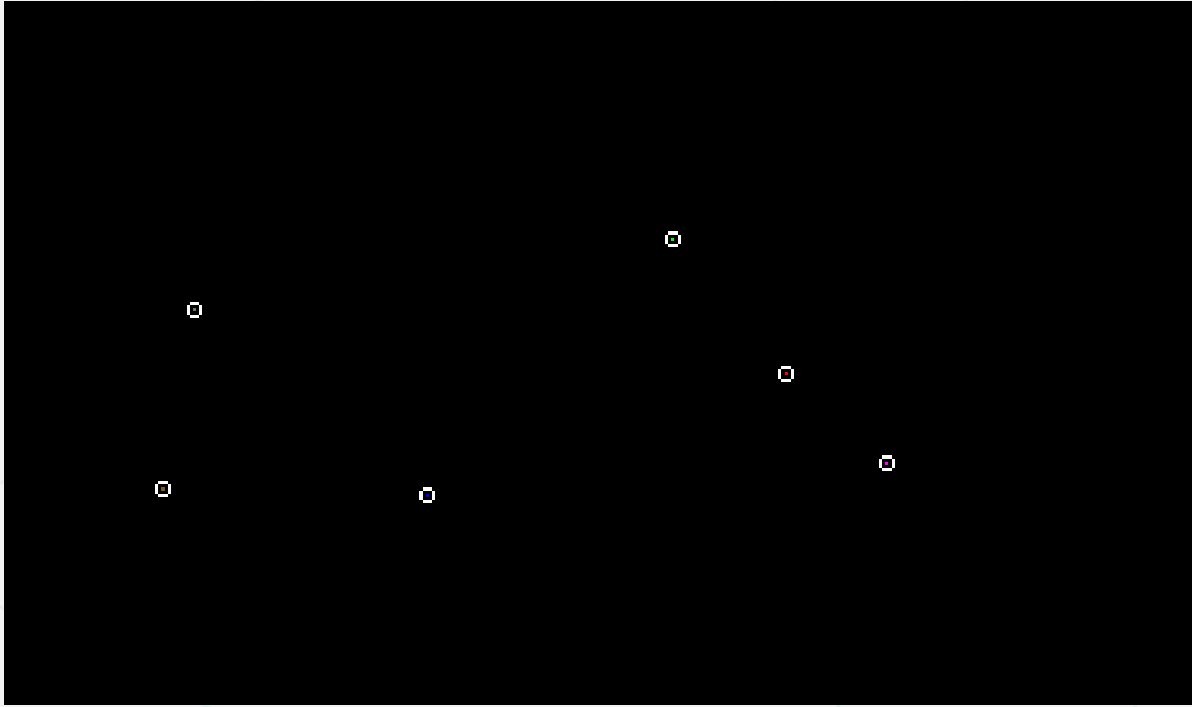


Dimensionality Reduction

- Sammons projection
 - Projects points from a space of higher to lower dimensionality
 - Preserving the structure of inter-point distances in n-dimensional space.
 - Minimization of the error function involves distance between the points in original space and corresponding distance after projection



Voronoi Tessellations

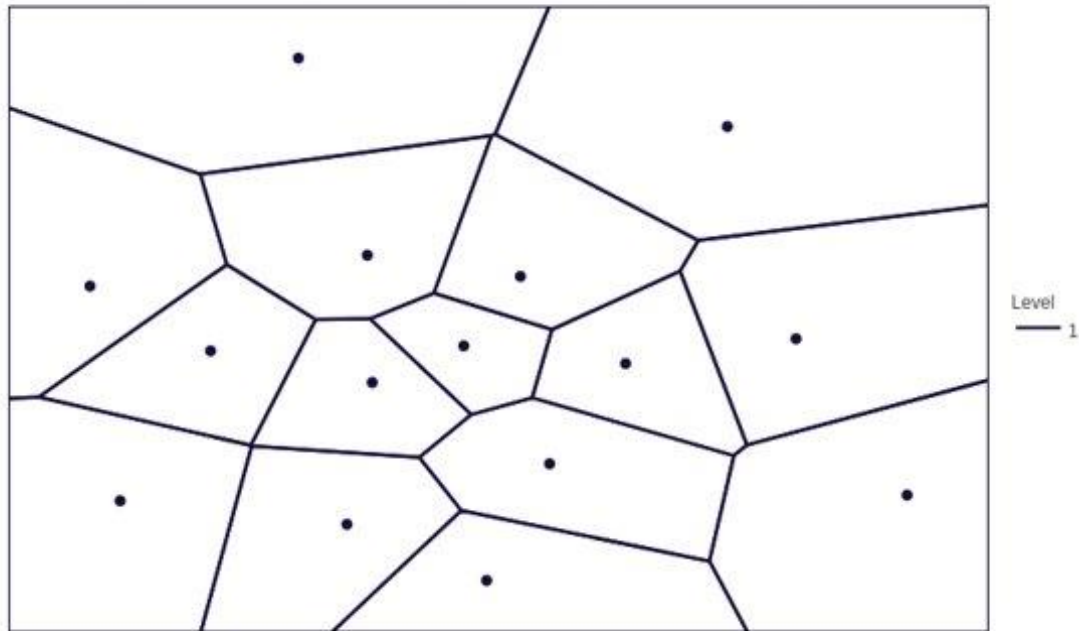


Presence of geometric arrangements in nature

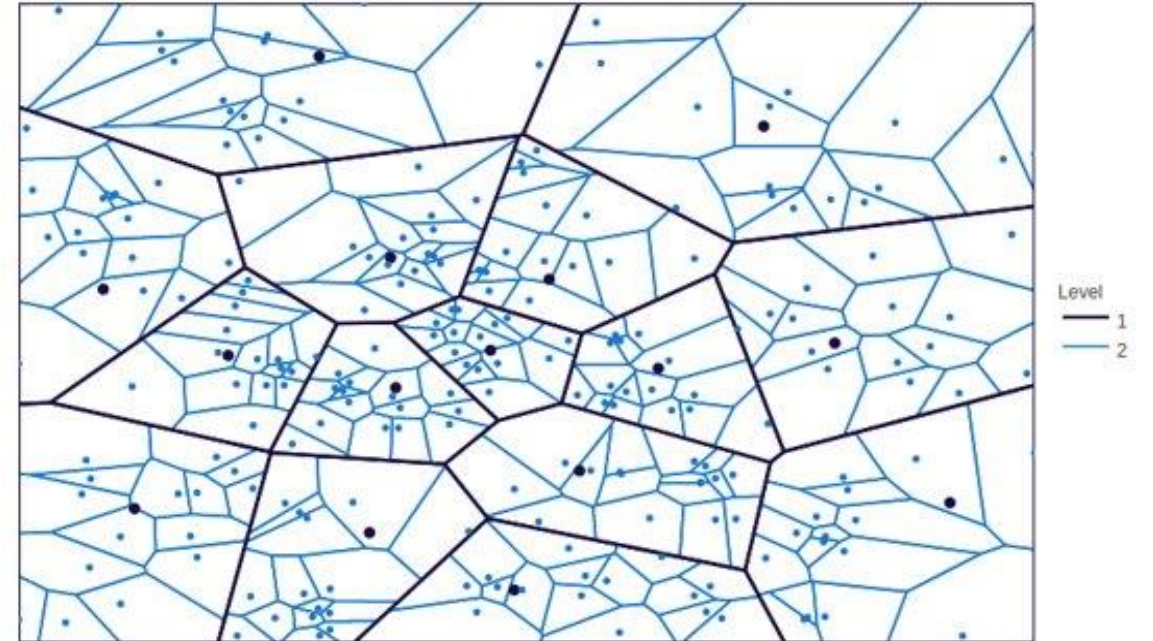


Hierarchical Voronoi Tessellations

Hierarchical Voronoi Tessellation for Level 1



Hierarchical Voronoi Tessellation for Level 2



Agenda

- ▶ The Enterprise Requirement for Complexity Science
- ▶ A Point of View - Representation of a Complex System
- ▶ Use Case: Intelligent Reporting
- ▶ **Demo**

Shiny interface for **muHVT**

Mu Sigma

v20.06.03

Hierarchical Voronoi Tessellations Brick

Torus Demo

Choose CSV File

Browse... may_19_pair_data.csv

Upload complete

☒ unchecked to use raw data, z-score normalized by default

Maximum Number of Cells at each Level

15

Maximum Number of Cells at each Level

15

Maximum Depth

1

Quantization Error Threshold

0.2

Note: Mean absolute deviation is used for error calculation. Non-numeric fields are ignored.

Select Error Metric

☒ Mean ☐ Max

Select Distance Metric

☒ Manhattan Distance ☐ Euclidean Distance

Submit

Press "Submit" to refresh plots with given values
Max Depth: 2 , Number of Clusters: 15 , Quant. Error Threshold: 0.2 , Error Metric: Mean , Distance Metric: Manhattan Distance

Dataset

Hierarchical Voronoi Tessellation

Heatmap

Explorer

3D Surface Plot

Quantization Error Summary

Predict

Note: Ignored, Dependent or ID columns won't be used while building the model.

Select Columns to be Ignored

STD RATIO

Select Dependent Variable

PERCENTPOSITIVETRADES

Select Identifier Column

EVENTTIME NAME

Note: Below section is to choose between time series and non-time series data.

Data Type

☐ Non Time Series ☒ Time Series

Select Time Column

EVENTTIME

Select Time Format

%Y-%m-%d %H:%M:%S

Select Panel Column

NAME

Update

Input Dataset

	COEFFVARIANCE	CORRELATION	HURST	MEAN	PERCENTPOSITIVETRADES	PVALUE	ZEROCROSSOVERRATE	ZSC
	All	All	All	All	All	All	All	All
1	0.4755	0.9873	0.4383	1.7234	0	0.4527	0	0.9
2	0.4679	-0.7745	0.5443	8.0596	0	0.1612	0	-0.1
3	0.8726	-0.8509	0.5301	15.3503	0	0.3204	0	-0.2
4	0.4879	0.843	0.4754	1.3563	0	0.0759	0	0.1
5	0.495	0.7377	0.3596	1.9046	0	0.0209	0	-0.3

Showing 1 to 5 of 376,014 entries

Previous12345...75203Next

Dependent (Y) Feature Summary

	variable	hist	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
1	PERCENTPOSITIVETRADES		0	1	28.1945	40.2494	0	0	0	50	100



Thank you

