Mar. 28, 2025

Interactive Data Science: Detecting the Expected and Discovering the Unexpected

Chad Steed, Ph.D.

Oak Ridge National Laboratory

U.S. DEPARTMENT OF ENERGY

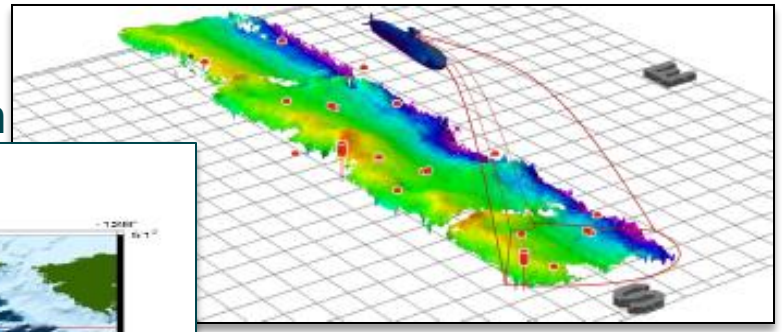
ORNL IS MANAGED BY UT-BATTELLE LLC
FOR THE US DEPARTMENT OF ENERGY

*All images are my own unless otherwise noted.

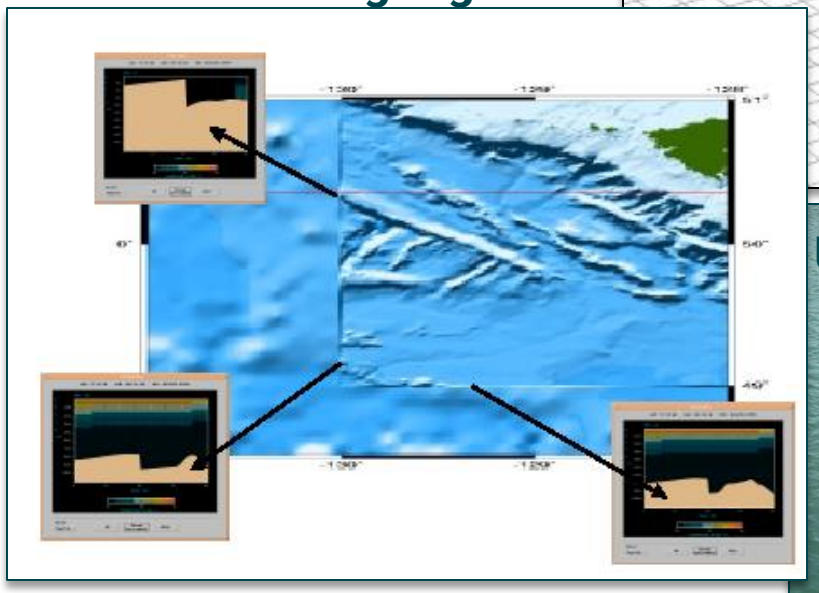
Lockheed Martin and Naval Research Laboratory

- Led international efforts to develop new oceanographic and acoustics data products
- US Navy Through-the-Sensors
- Led some of the first autonomous underwater vehicle efforts
- Hydrographic Science
- Introduced visual analytics to U.S. Navy oceanographic and acoustics operations

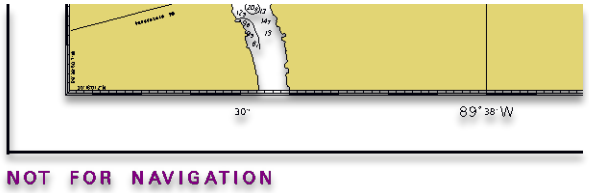
Underwater Mapping

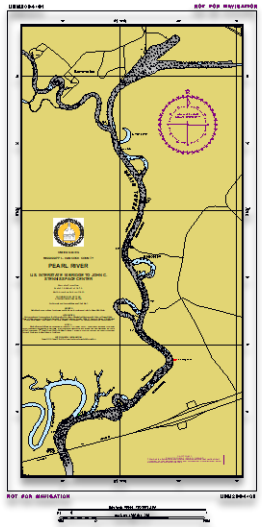


OAML Feathering Algorithm

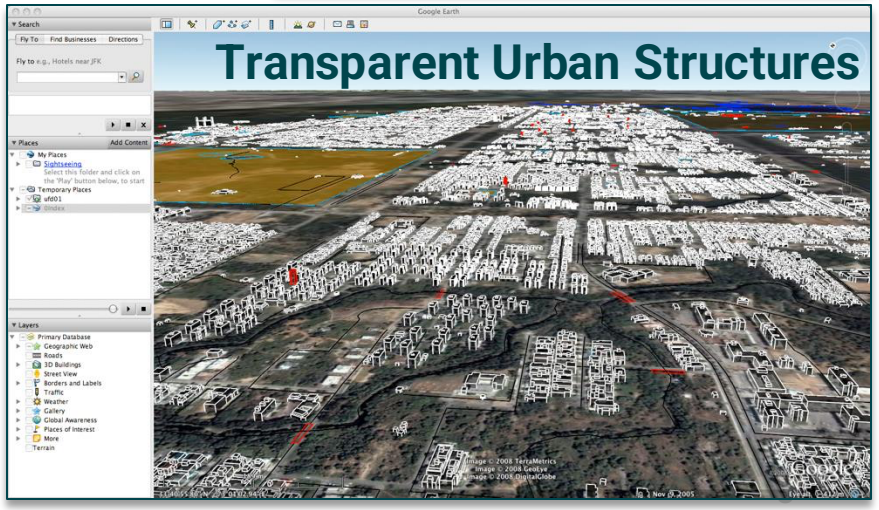


Unmanned Survey Vehicle



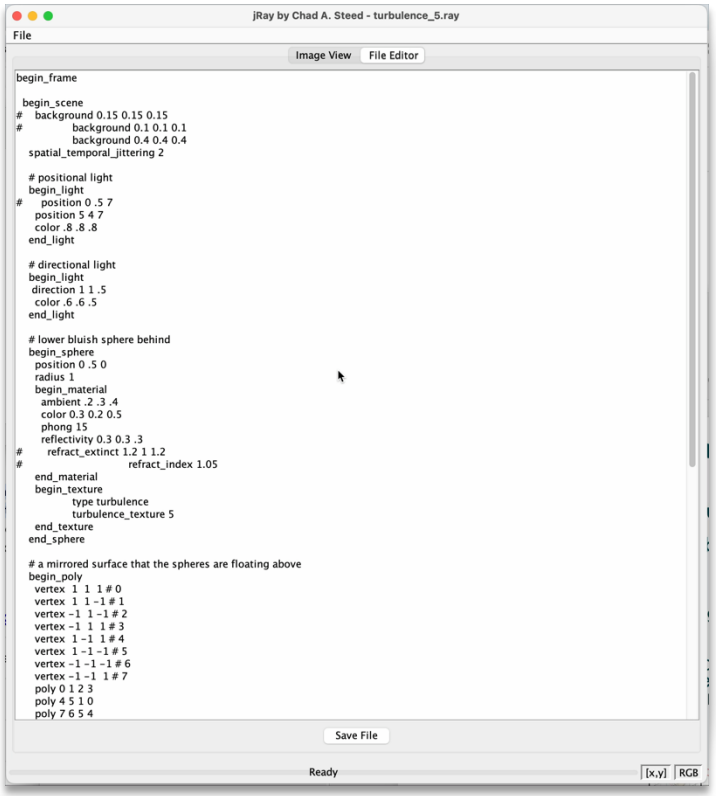


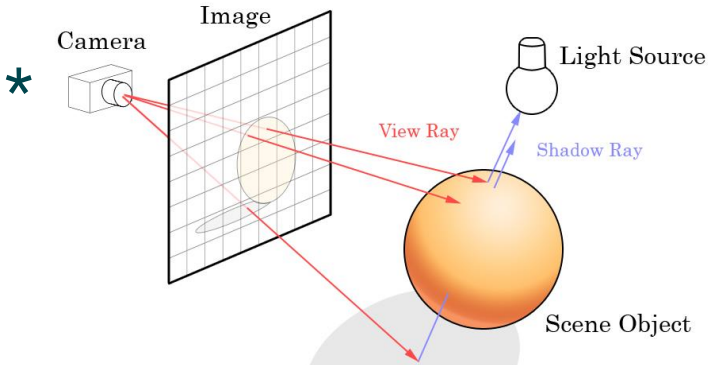
Transparent Urban Structures



jRay: A Ray Tracing Framework and GUI

- Developed during Ph.D. studies at MSU
- Initially developed in C++ but later ported to Java
- Implemented a GUI for editing scene parameters and monitoring the ray tracing process.
- Multi-threaded with several advanced options to render complex surfaces, antialiasing, reflections, textures, Perlin noise, etc.

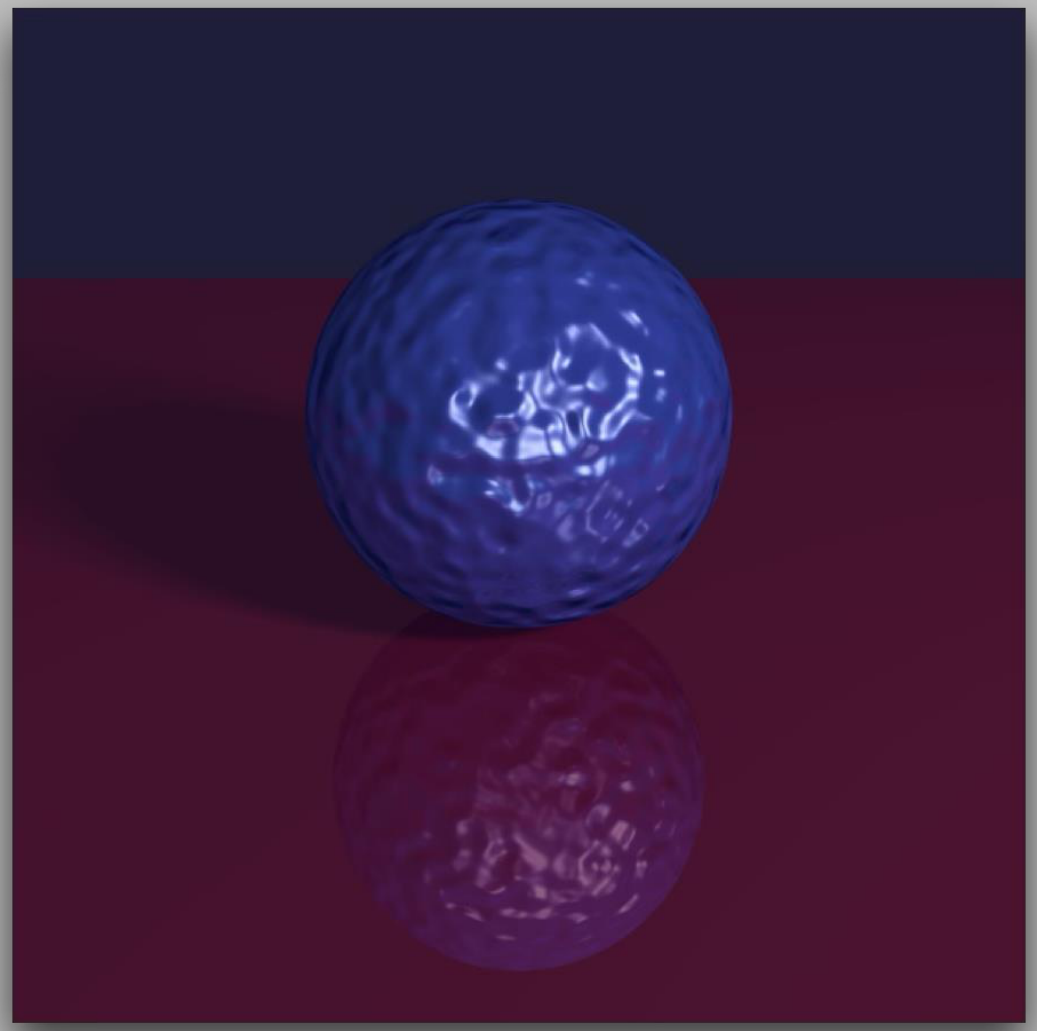




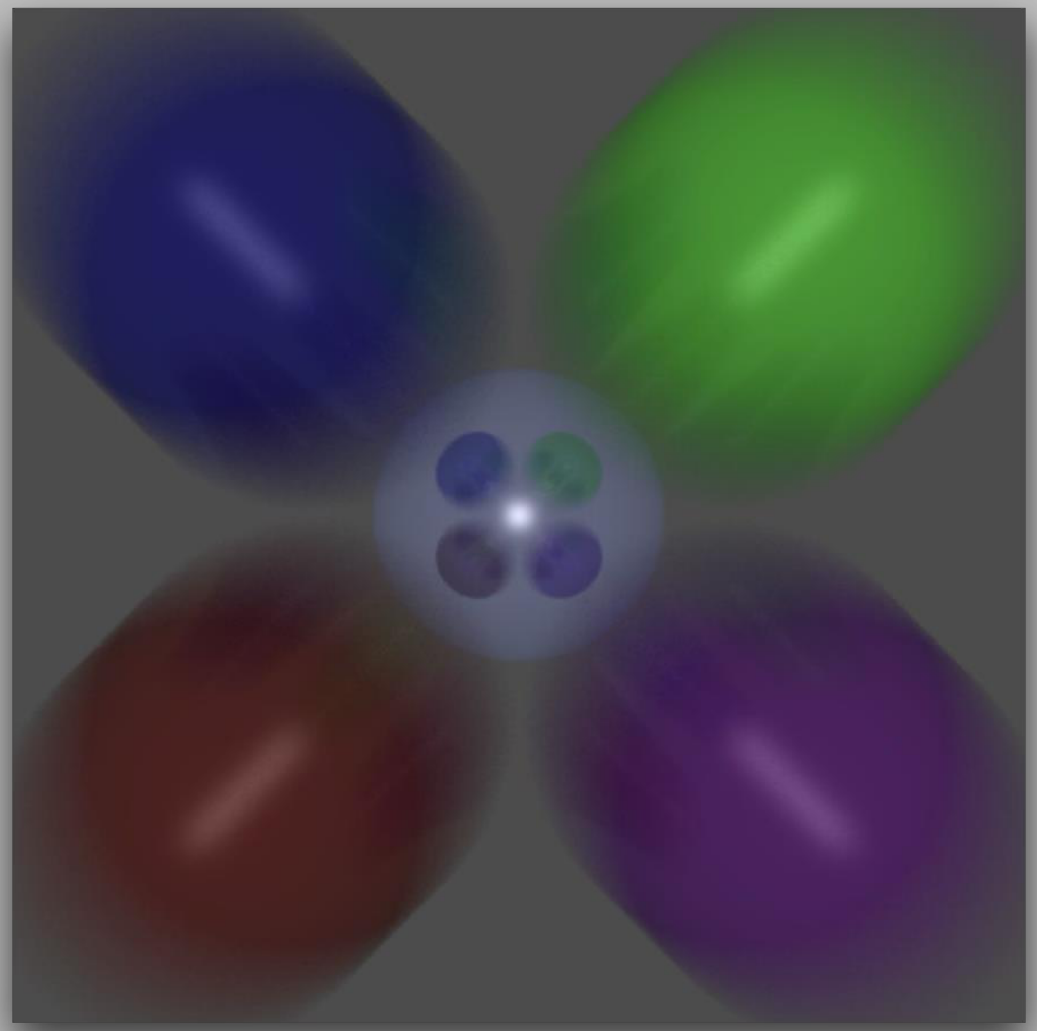
*by Timrb - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=3553478

jRay Outputs

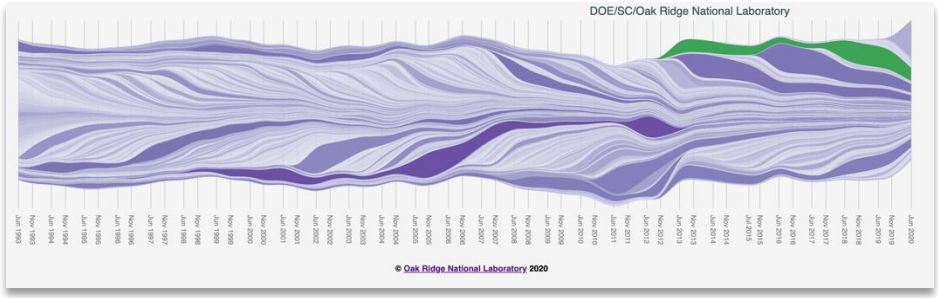
Perlin Noise



Motion Blur with Reflections

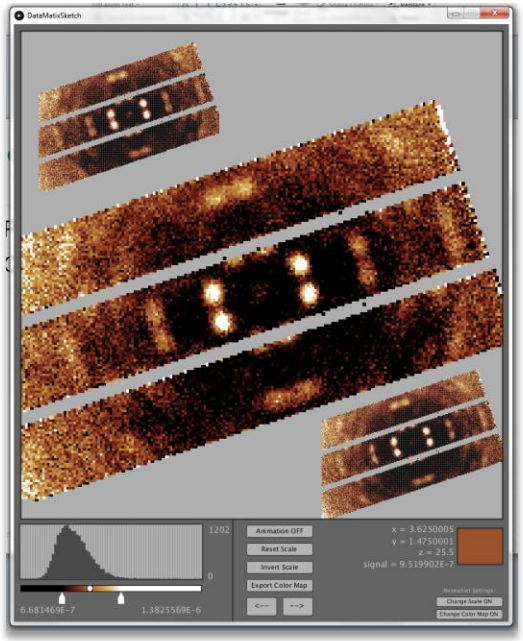


ORNL: Broad Multidisciplinary Applications with Domain Experts

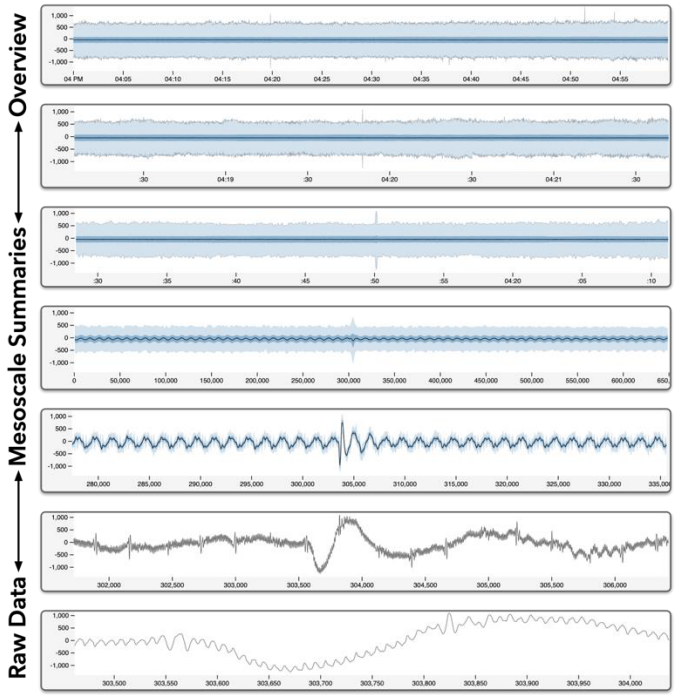


Top500 Trend Analysis

Nuclear Nonproliferation

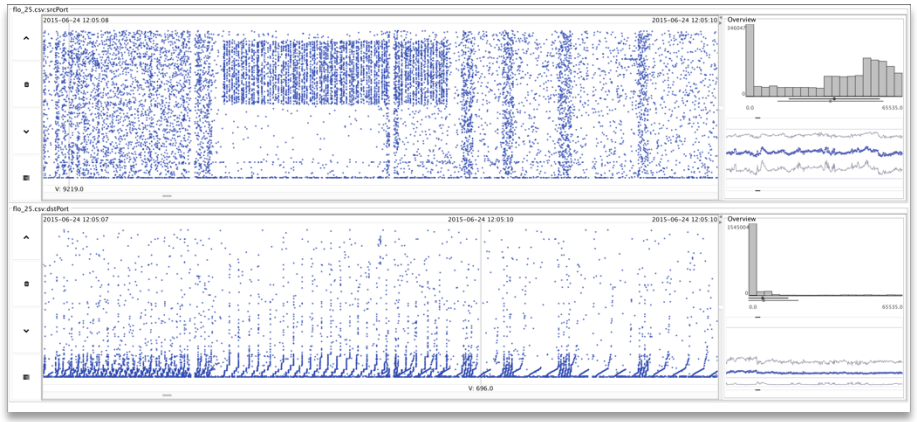


Neutron Science



Scalable Senor Analysis

AR/VR Immersive Analytics



Cyber Security & Cyber Physical

My Industry Sabbatical

- **Regions Financial Corp. (VP, Enterprise Data Analytics)**
  - Introduced advanced multivariate visual analytics to the enterprise
  - Applications in finance, fraud, wealth, and customer support
  - On-premise challenges, vendor interactions, and innovating in a corporation

- **Lirio, LLC (Sr. Director of Analytics)**
  - Startup focused on integrating reinforcement learning with behavioral science for healthcare messaging
  - Reported to CTO and executive team, led data scientists, agile development
  - Moved analytics / reporting from a reactive to proactive maturity level
  - Established internal analytics platform and external customer app
    - Addressed all the legal, technical, and user experience challenges
  - Entrepreneurial processes, applied data science/ML, formal leadership training, 1:1 direct reports

Interactive Data Science

GOAL: Help people explore and understand data

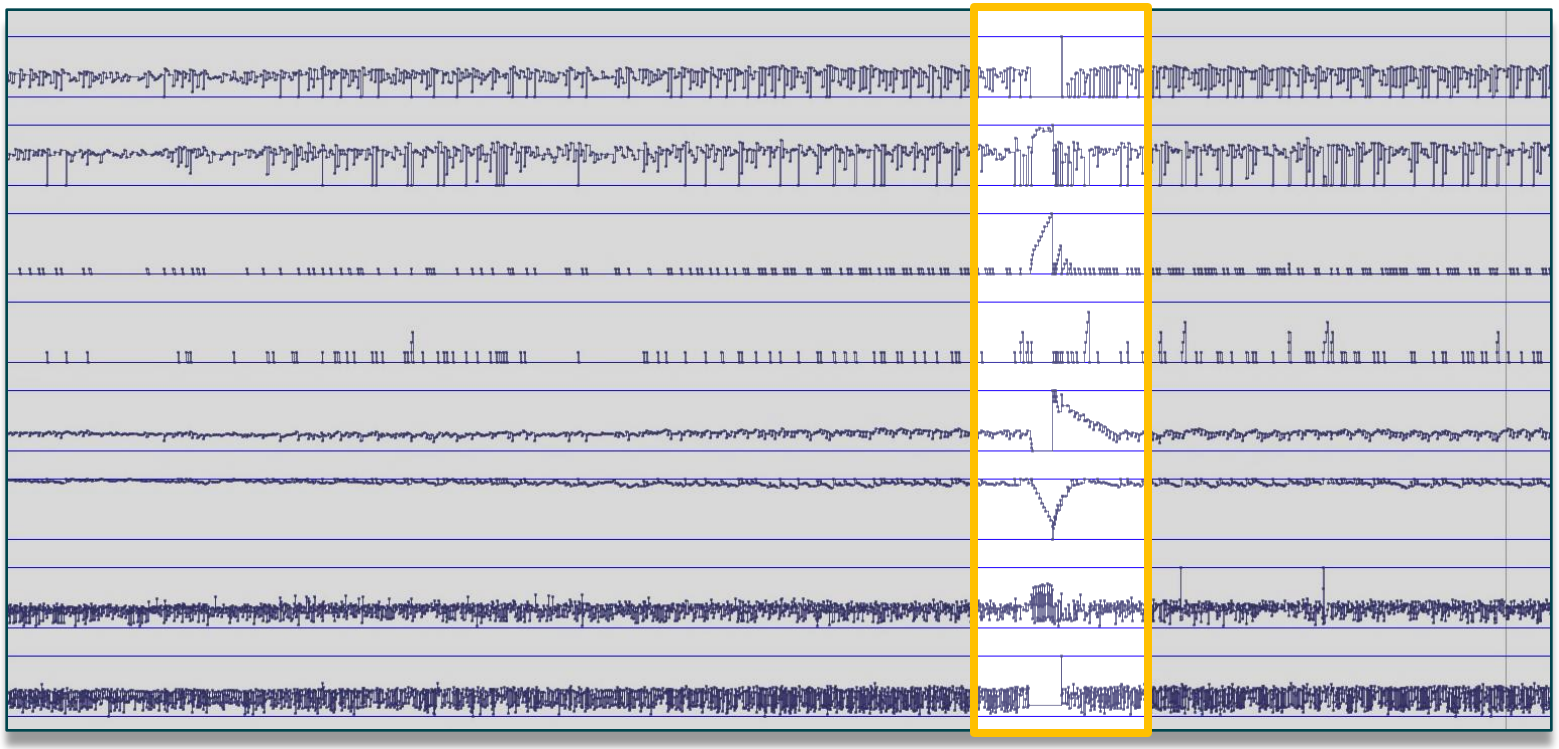
Joseph Berkson calls this
"interocular traumatic impact"

"A conclusion that hits us between the eyes."
[Cleveland, 1993]

DATA

```
492258  2016-06-17 21:07:03.466|OPC.Temperature.ExtraTemp1|SuperUser (OPC)|9526586|618
492259  2016-06-17 21:07:03.466|OPC.Temperature.BottomTemperatureValidation|SuperUser
        (OPC)|9526586|618.19999999999982
492260  2016-06-17 21:07:03.486|Process.CathodeTuningControl.CathodePower|[OnChange(
        OPC.PowerSupply.Filament.VoltageFB)] Arcam.EBMControl.Process.
        CathodeTuningControl.OnCathodPowerChange() (Logic)|9526586|5.929202
492261  2016-06-17 21:07:03.516|Process.CathodeTuningControl.MeanCathodePower|[
        OnChange(Process.CathodeTuningControl.CathodePower)] Arcam.EBMControl.Process.
        CathodeTuningControl.MeasureMeanPower() (Logic)|9526587|5.905572
492262  2016-06-17 21:07:03.516|OPC.PowerSupply.HighVoltage.SafetySignal|[
        OnPositiveFlank(SafetySignalTimer.Timeout)] Arcam.EBMControl.Process.
        HighVoltageControl.OnTimeToSendSafetySignal() (Logic)|9526587|True
```

Insight



“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a **poverty of attention**, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

Herb Simon
Quoted by Hal Varian
Scientific American
September 1995

How Many V's?

MTHIVLWYADCEQGHKILKMTWYN
ARDCAIREQGHLVKMFPSTWYARND
GFPSVCEILQGKMFPSNDRCEKQPDK
SGMDKAKDKKVPSTWYACEQTWRS

How Many V’s?

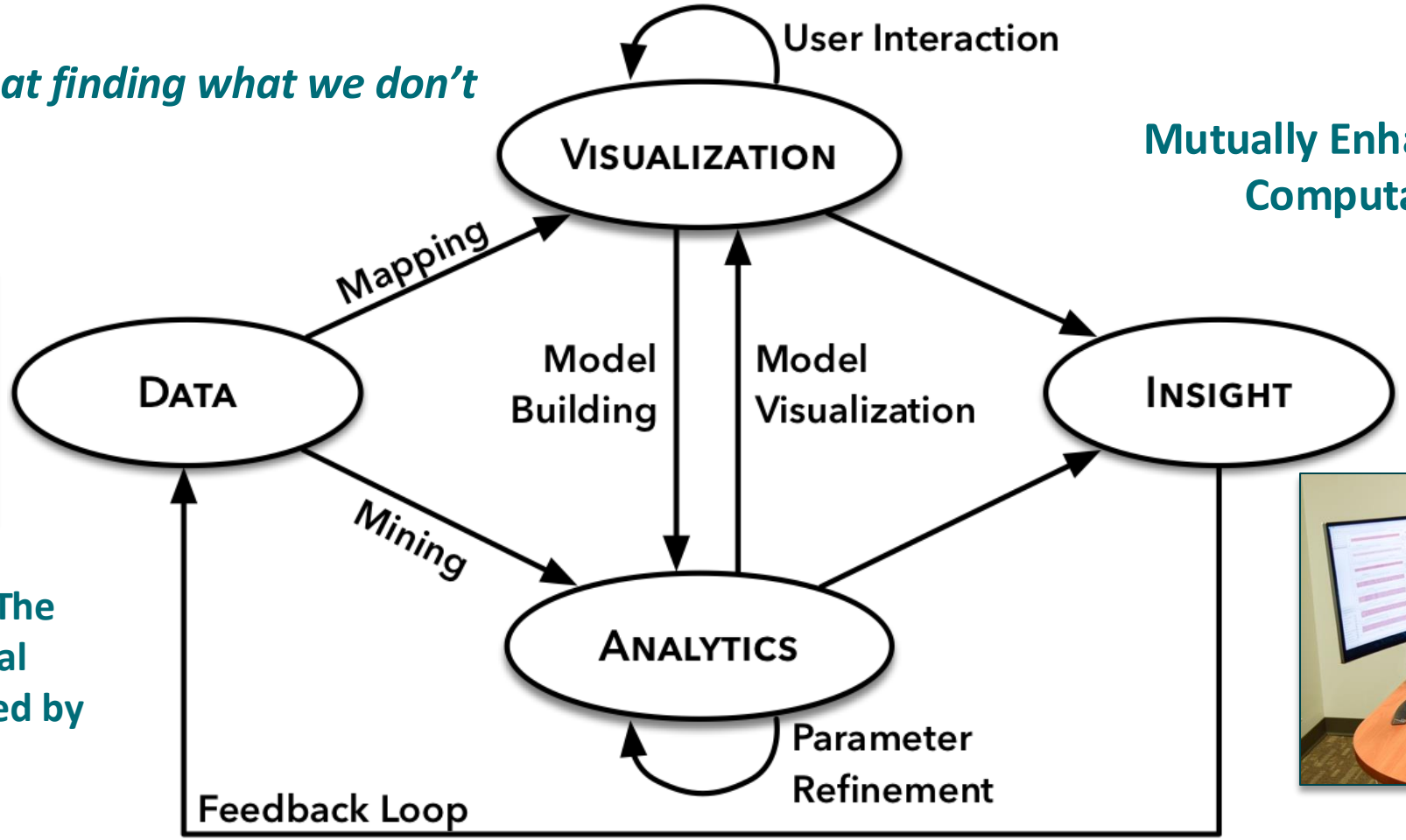
MTHI**V**LWYADCEQGHKILKMTWYN
ARDCAIREQGHL**V**KMFPSTWYARND
GFPS**V**CEILQGKMFPSNDRCEKQPDK
SGMDKAKDKK**V**PSTWYACEQTWRS

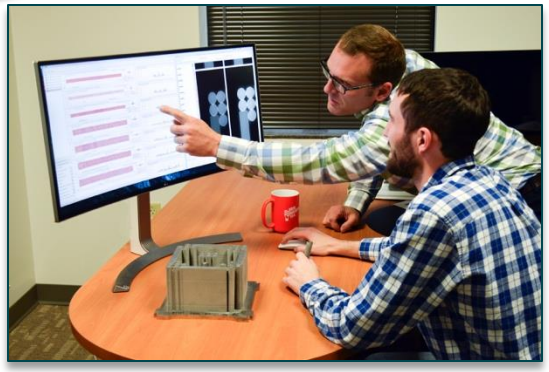
Visual Analytics: Detect the Expected, Discover the Unexpected

"We don't look at enough of our data."

"We're not good at finding what we don't already know."

Mutually Enhance Human and Computational Strengths

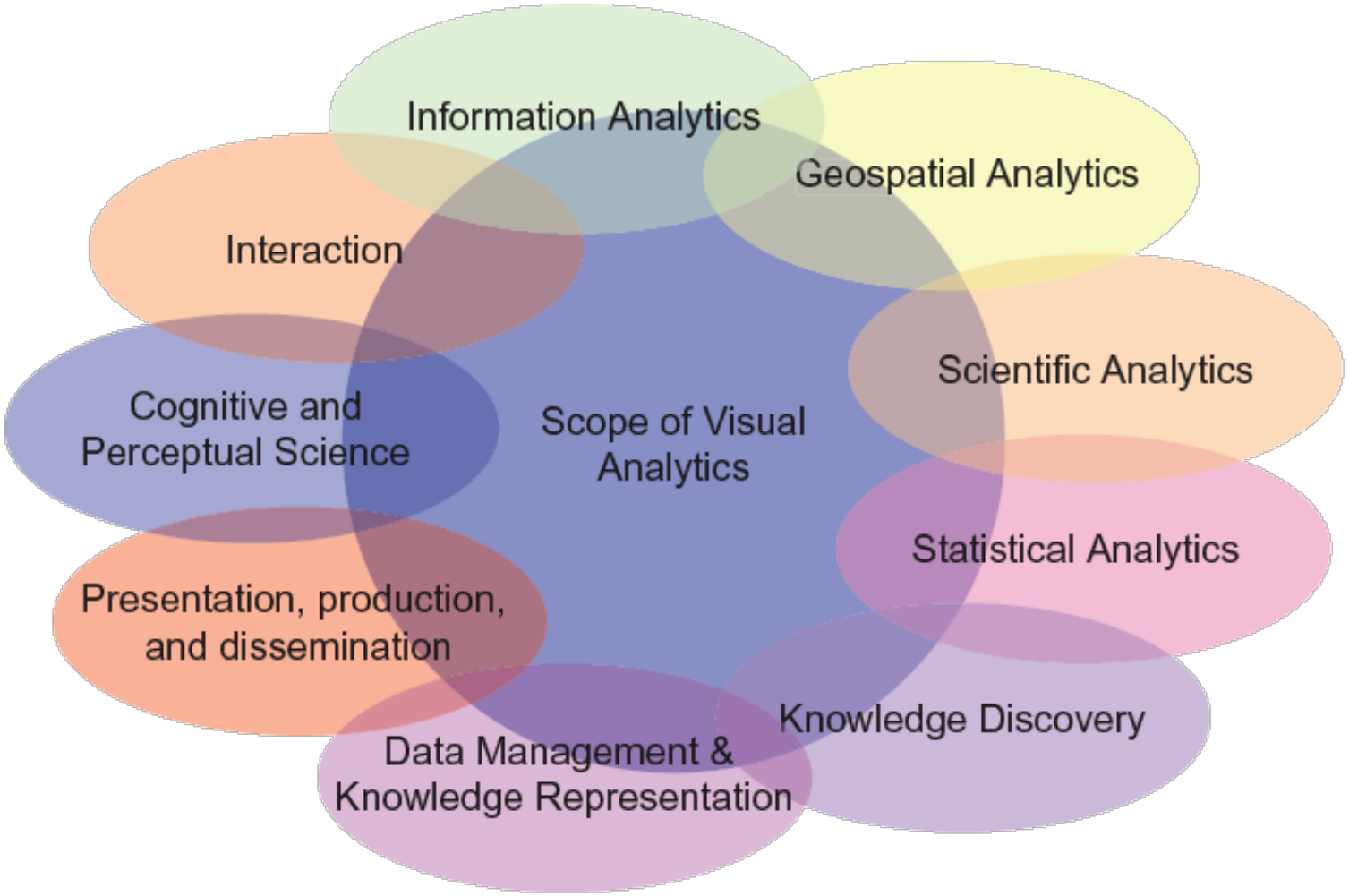




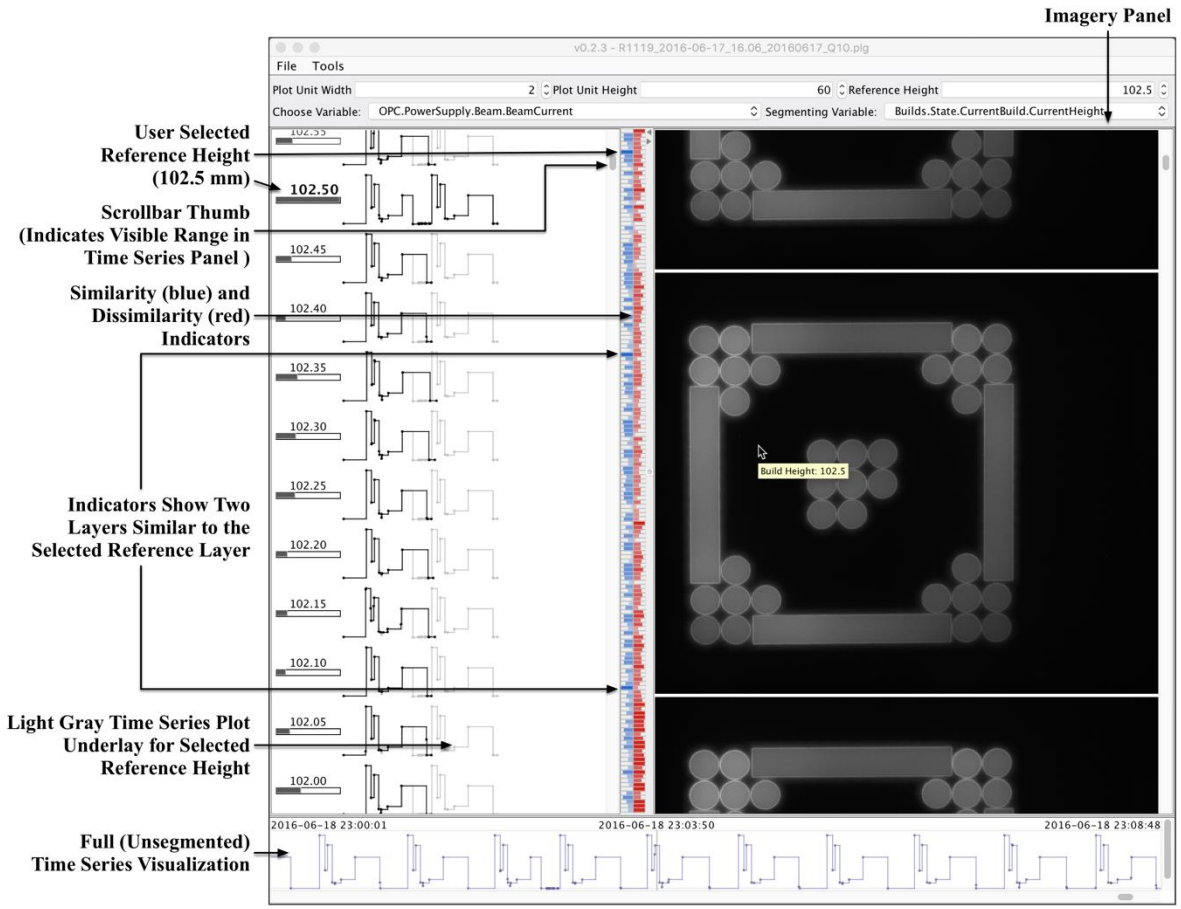
Visual Analytics: "The science of analytical reasoning facilitated by interactive visual interfaces."

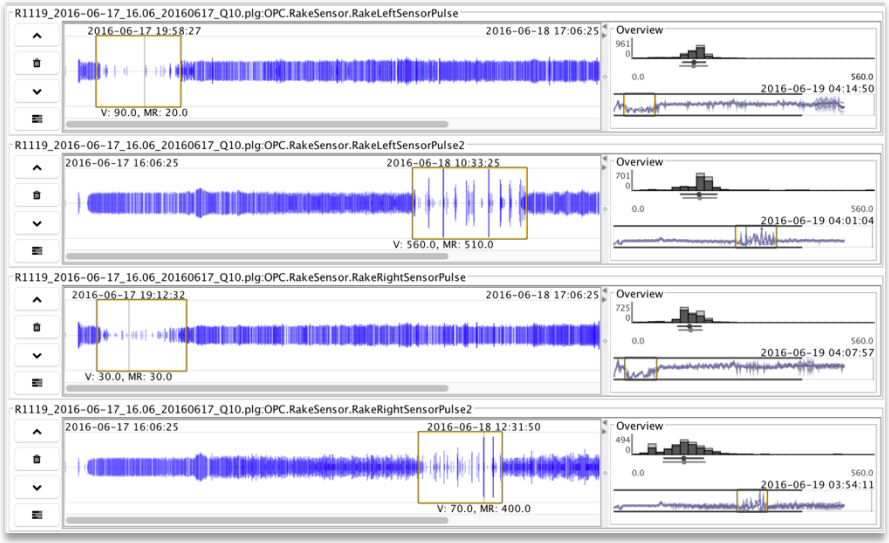
Adapted from *Keim et al., Mastering the Information Age: Solving Problems with Visual Analytics, 2011*

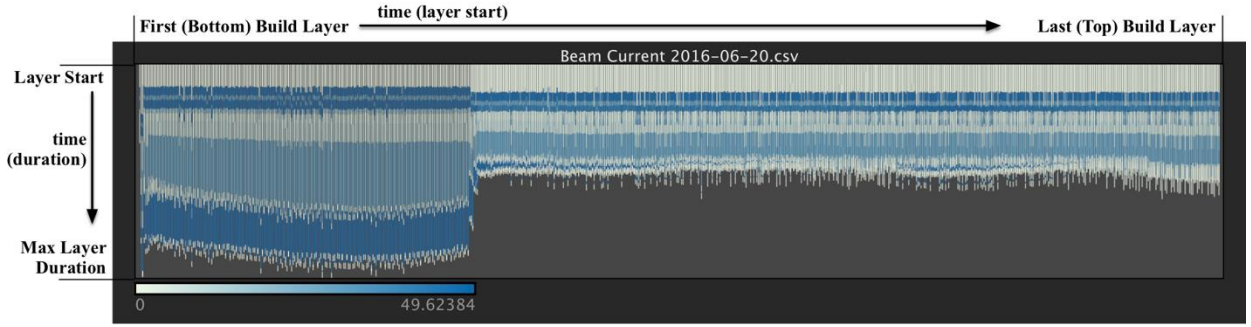
Visual Analytics is Multidisciplinary



Falcon: Guided Exploration of Temporal Sensor Data

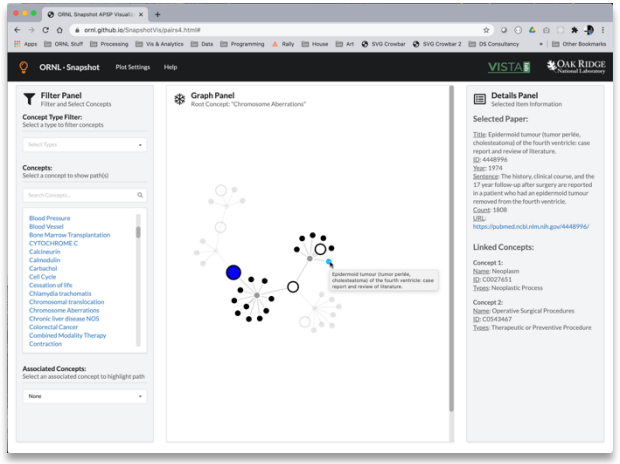




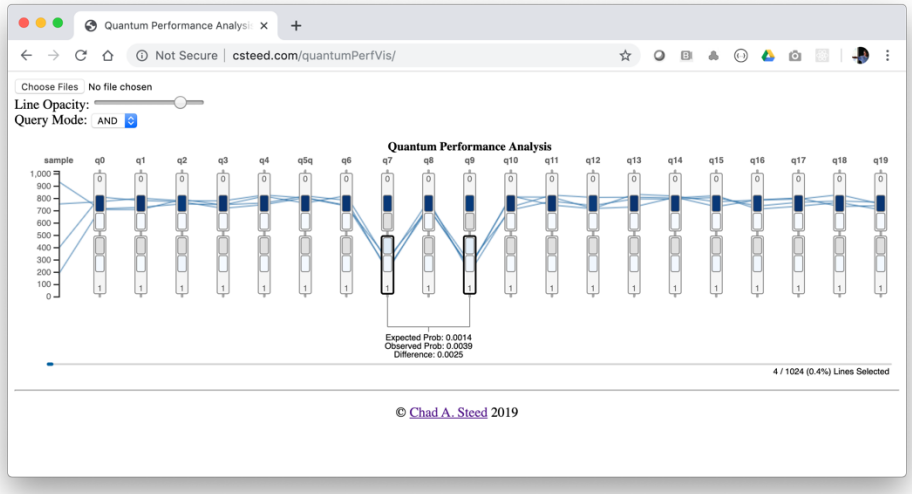


Citation: "Falcon: Visual Analysis of Large, Irregularly Sampled, and Multivariate Time Series Data in Additive Manufacturing." C. A. Steed, W. Halsey, R. Dehoff, S. Yoder, V. Paquit, and S. Powers. *Computers & Graphics*, vol. 63, 2017, pp. 50-64.

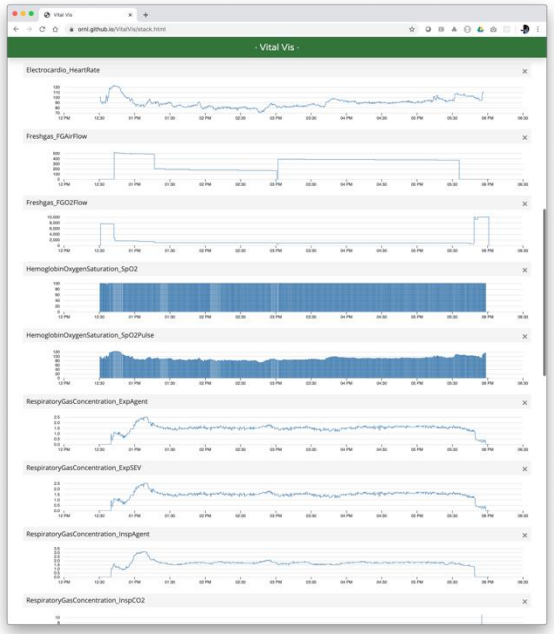
Broad Multidisciplinary Applications of Visual Analytics



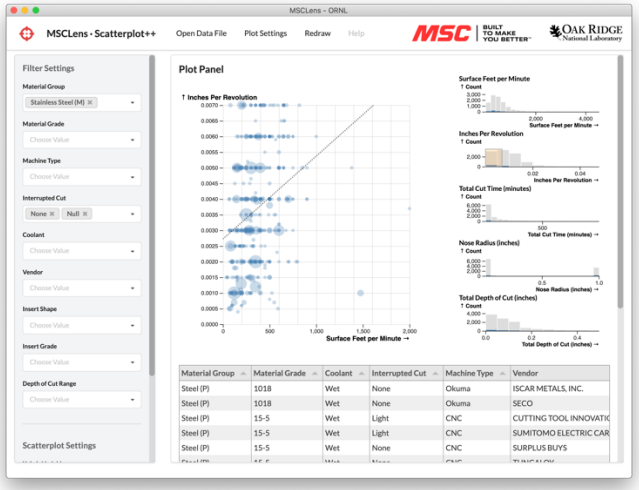
Graph Analytics and Publication Mining



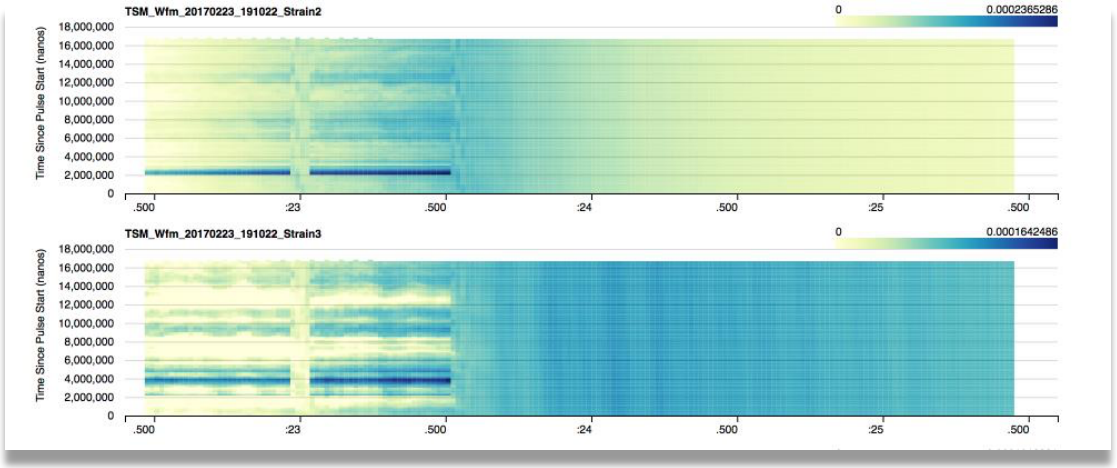
Exploring Noise in Quantum Computing Devices



Healthcare Decision Support



Business Intelligence for Tooling Companies



Strain Sensors from SNS Neutron Pulse Activities

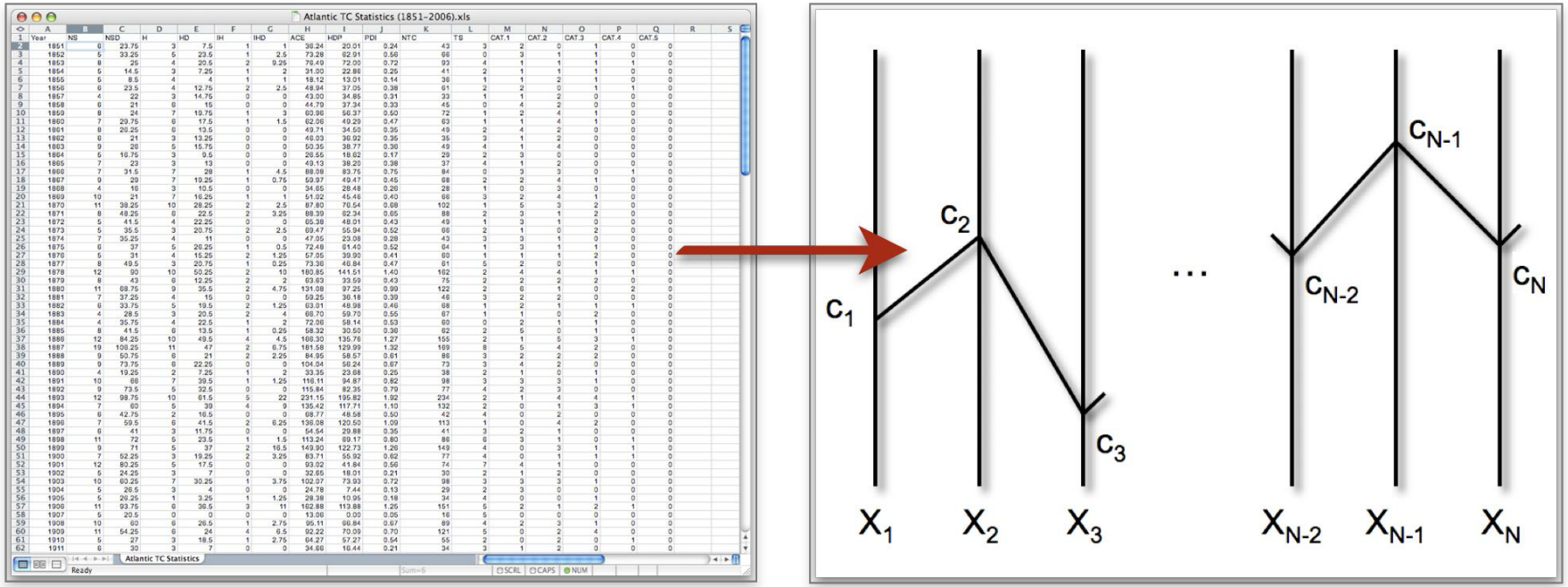
Multivariate Visual Analytics

Multivariate Visual Analytics with Parallel Coordinates

Parallel Coordinates

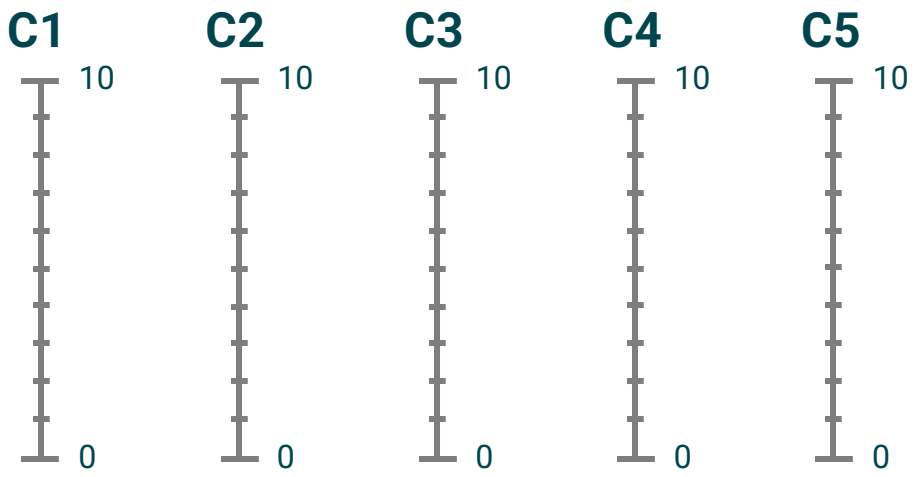
Two-dimensional representation of multidimensional data sets by representing the N-dimensional data tuple C with coordinates $(c_1, c_2, \ldots, c_N)$ by points on N parallel axes which are joined with a polyline.

Inselberg, 1985 | Wegman, 1990



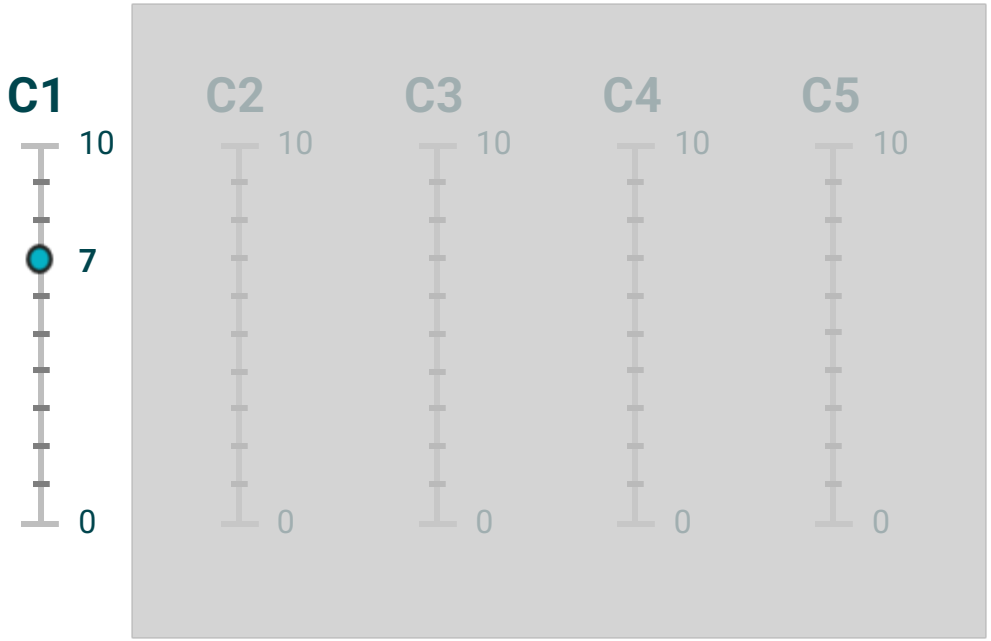
Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



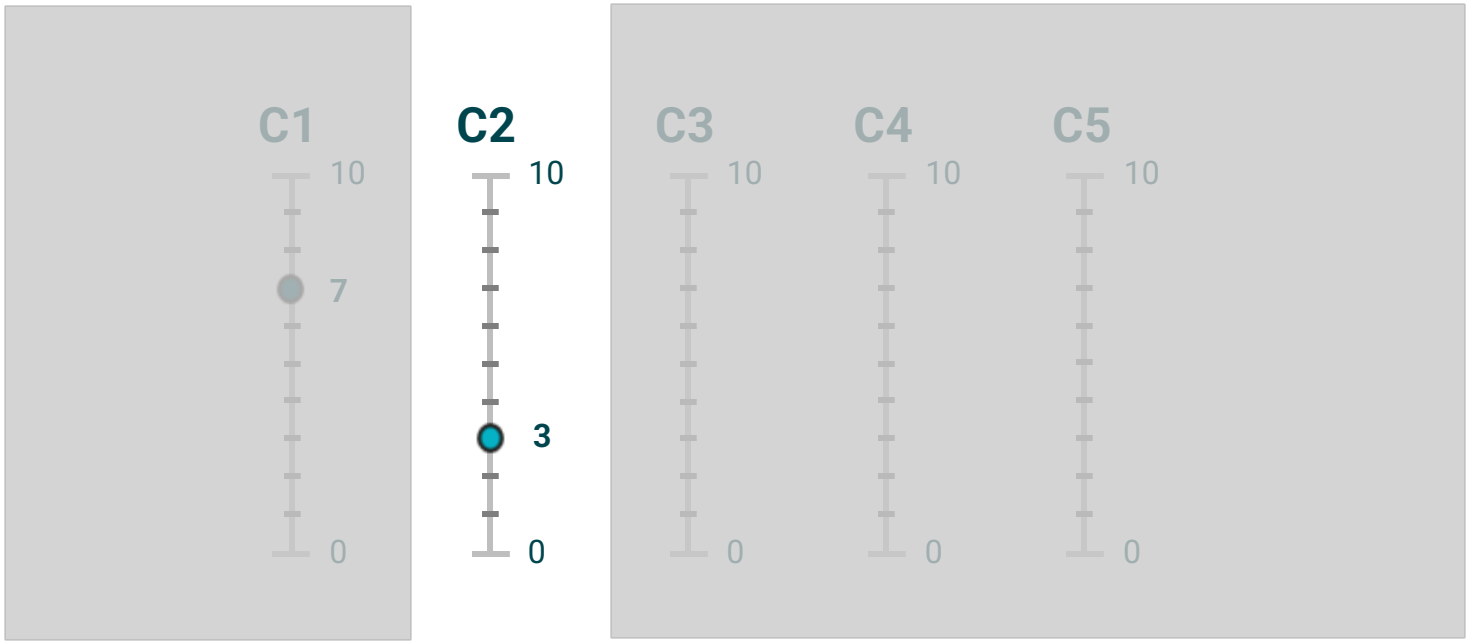
Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



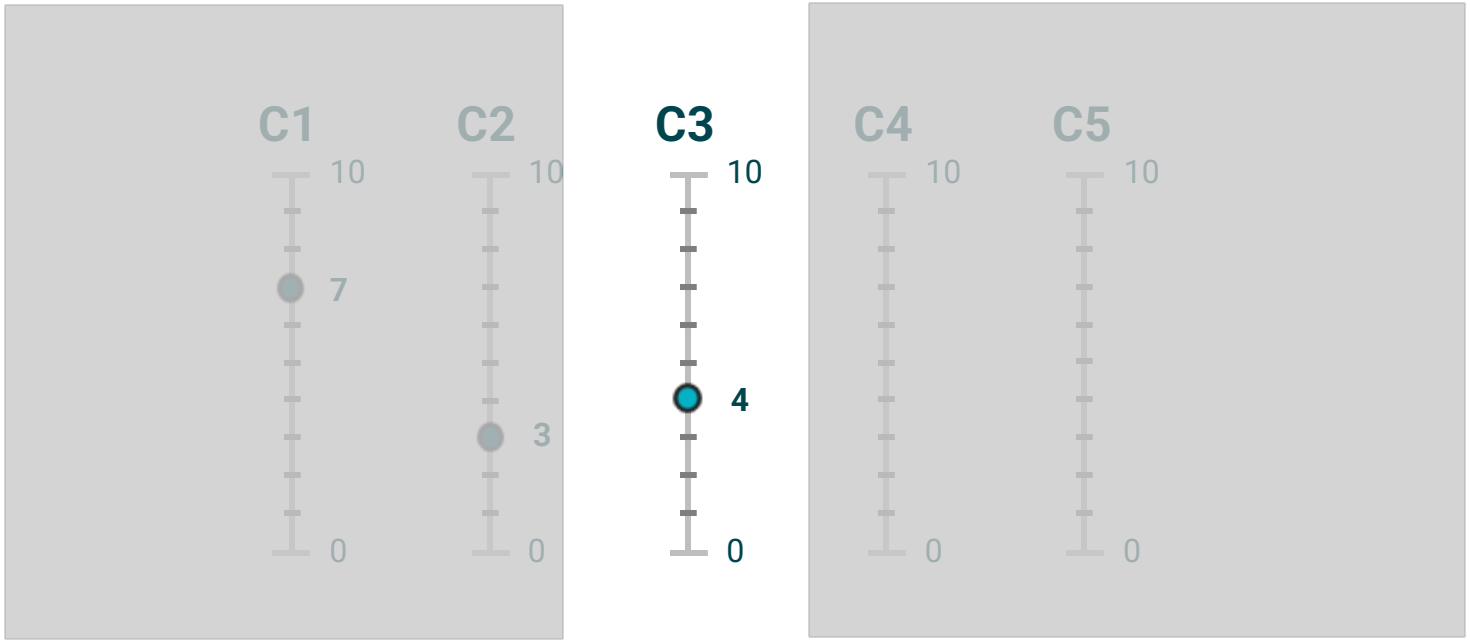
Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



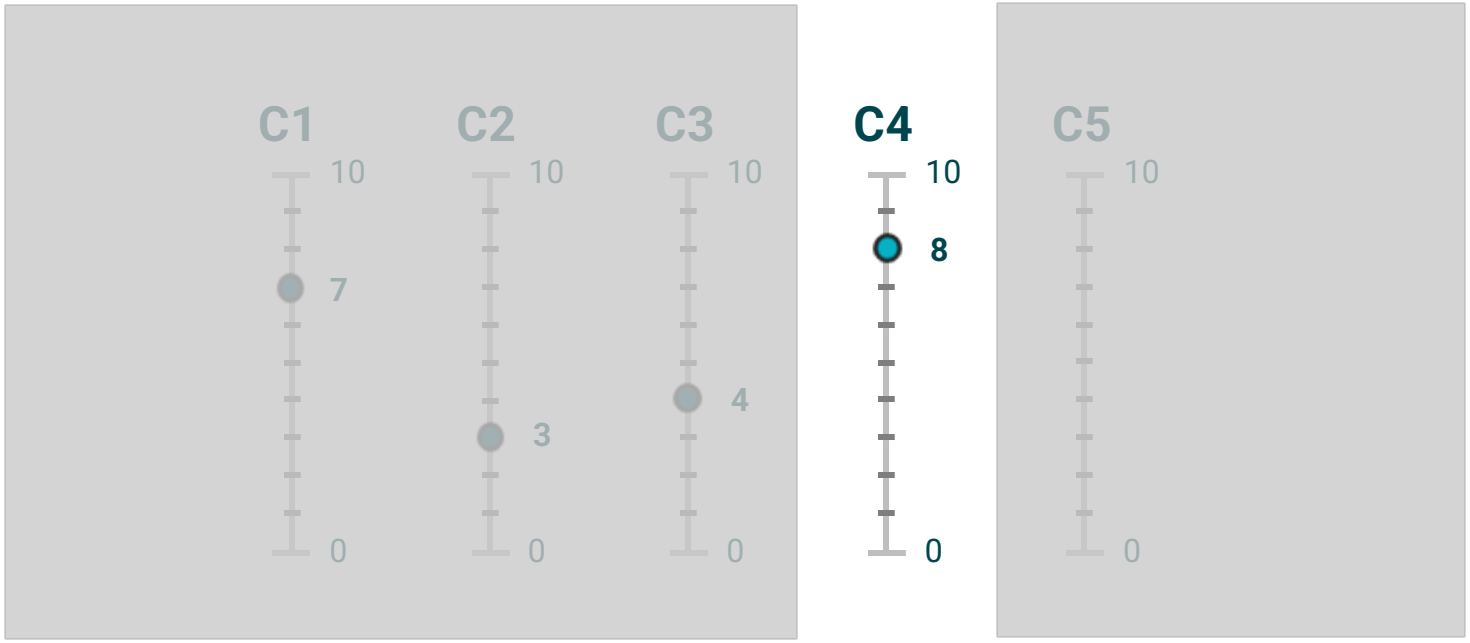
Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



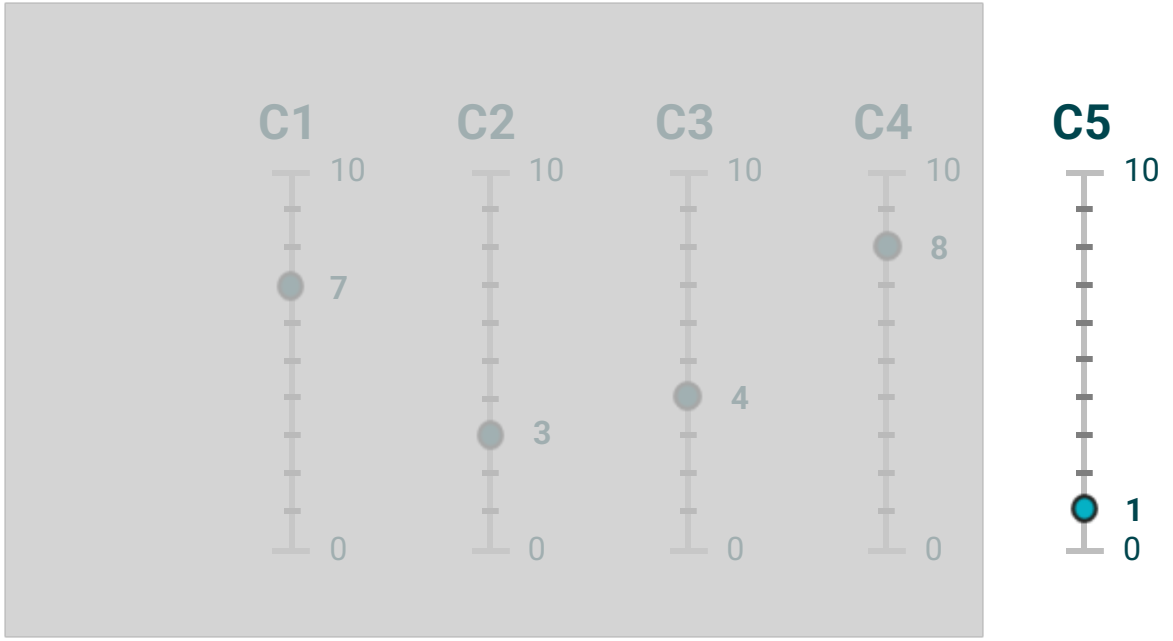
Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



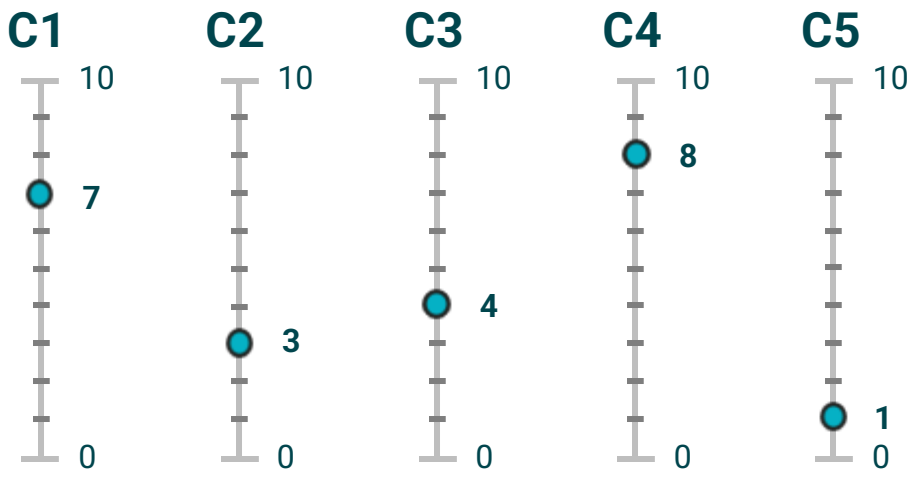
Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



Algorithm to Construct a Parallel Coordinates Plot

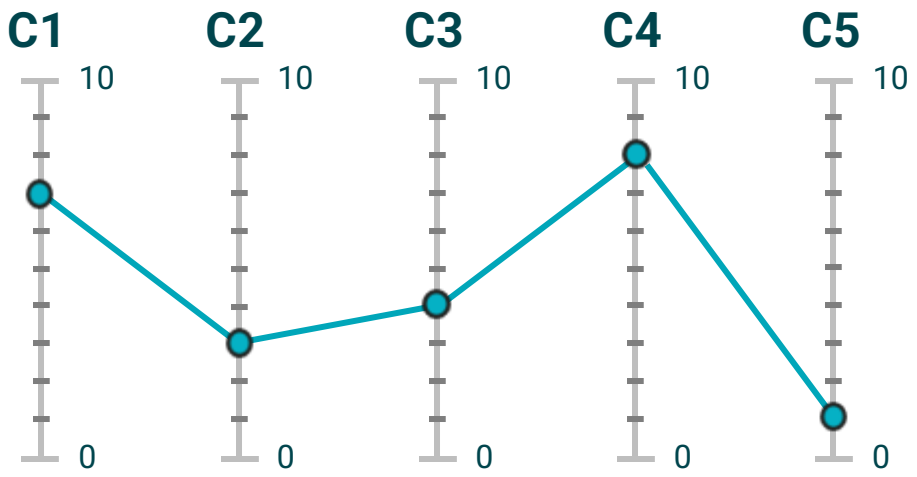
	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2



Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2

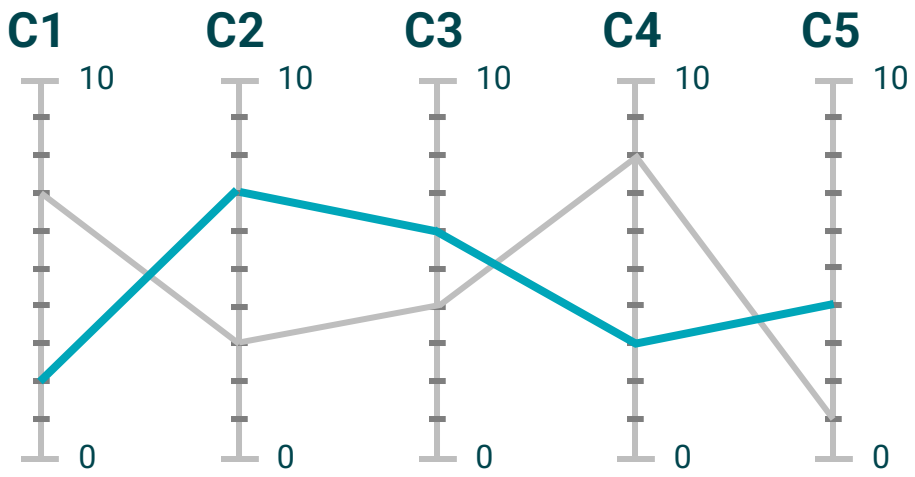
Polyline Representing R1



Algorithm to Construct a Parallel Coordinates Plot

	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2

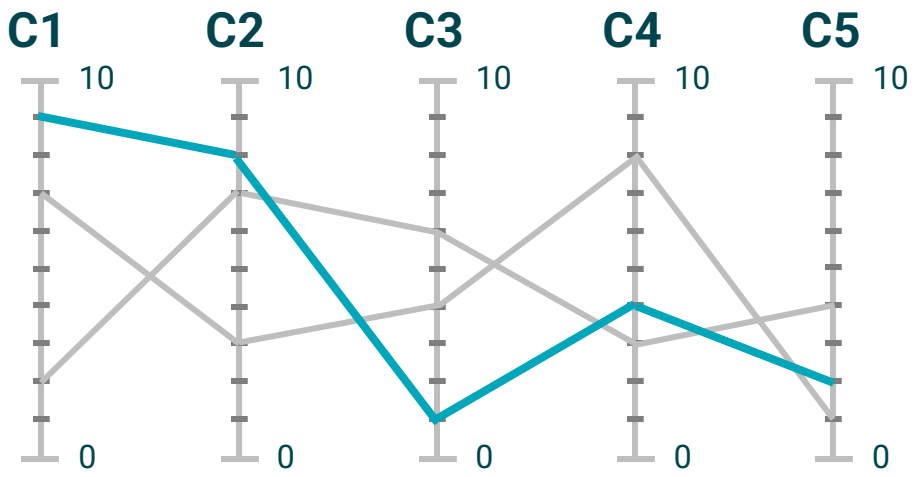
Polyline Representing R2



Algorithm to Construct a Parallel Coordinates Plot

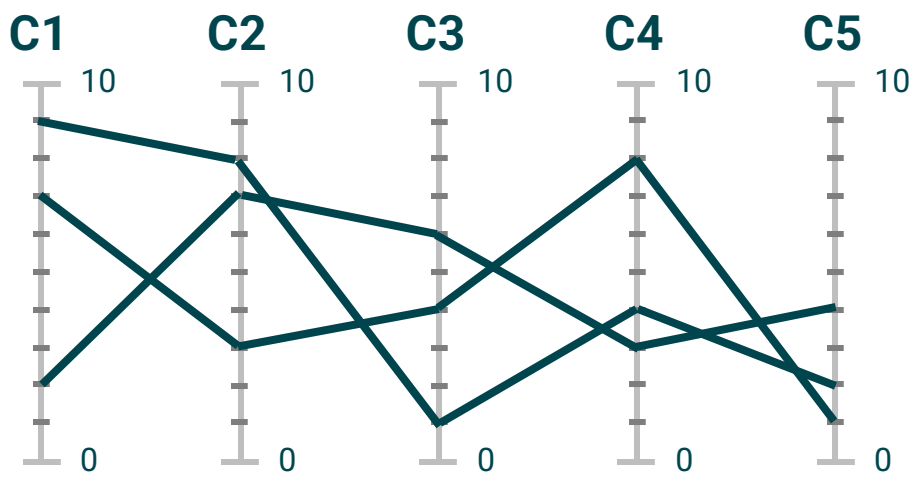
	C1	C2	C3	C4	C5
R1	7	3	4	8	1
R2	2	7	6	3	4
R3	9	8	1	4	2

Polyline Representing R3



Algorithm to Construct a Parallel Coordinates Plot

- Number of variables only limited by display resolution
- Two-dimensional view of high-dimensional data
- Avoids information loss -- no data reduction

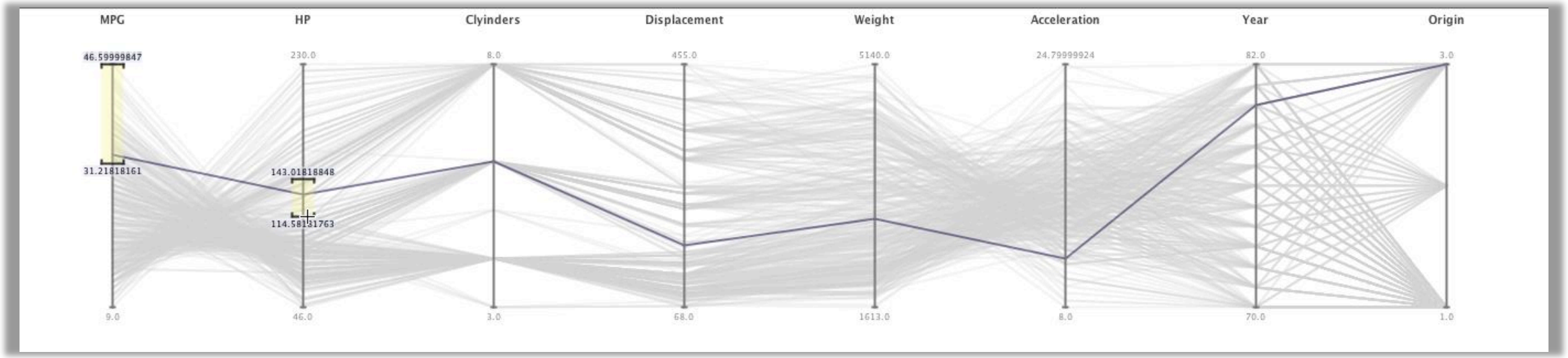


The Exploratory Power of Parallel Coordinates

Visual queries for asking “what if” questions of the data

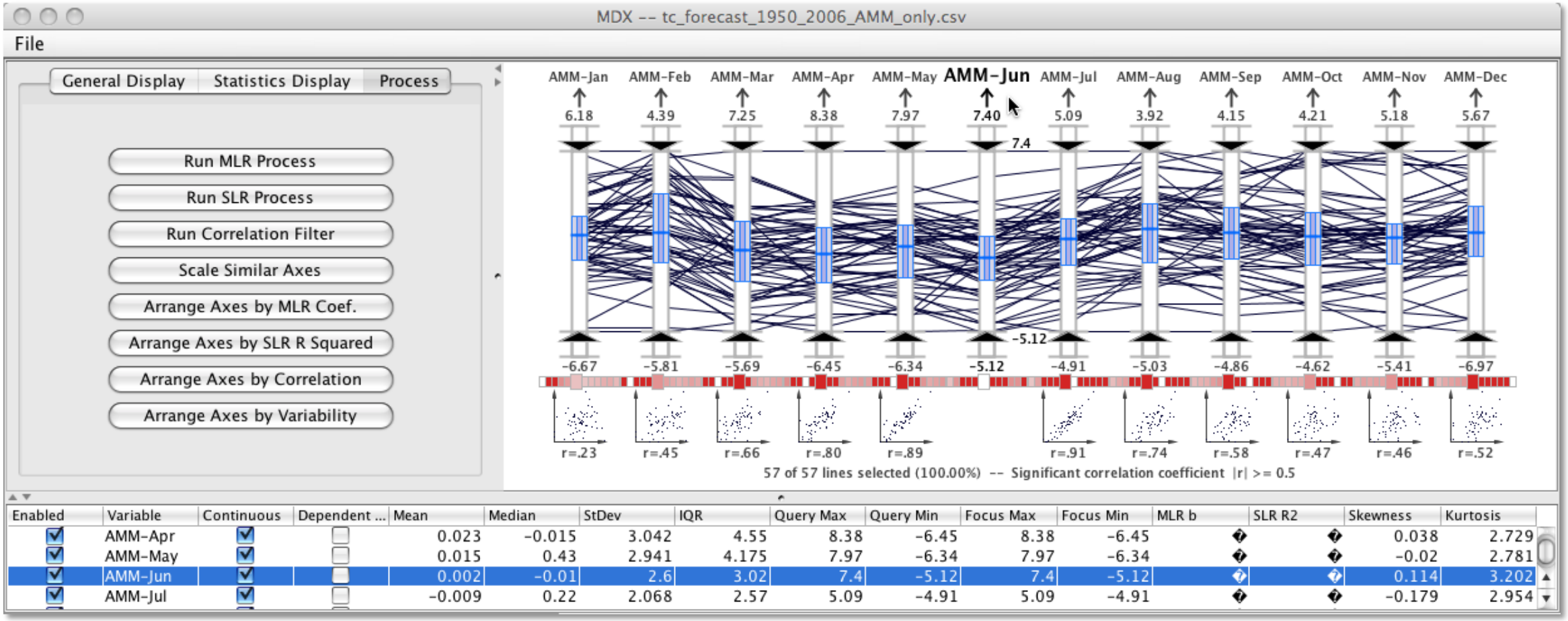
Correlations (line slopes), clusters, and trends are visually revealed

But we have a poverty of attention that would benefit from computational guidance



Classical Parallel Coordinates Plot of the ‘83 ASA Cars Dataset
(Highlighted polyline shows Datsun 280ZX, 32.7 MPG & 132 HP)

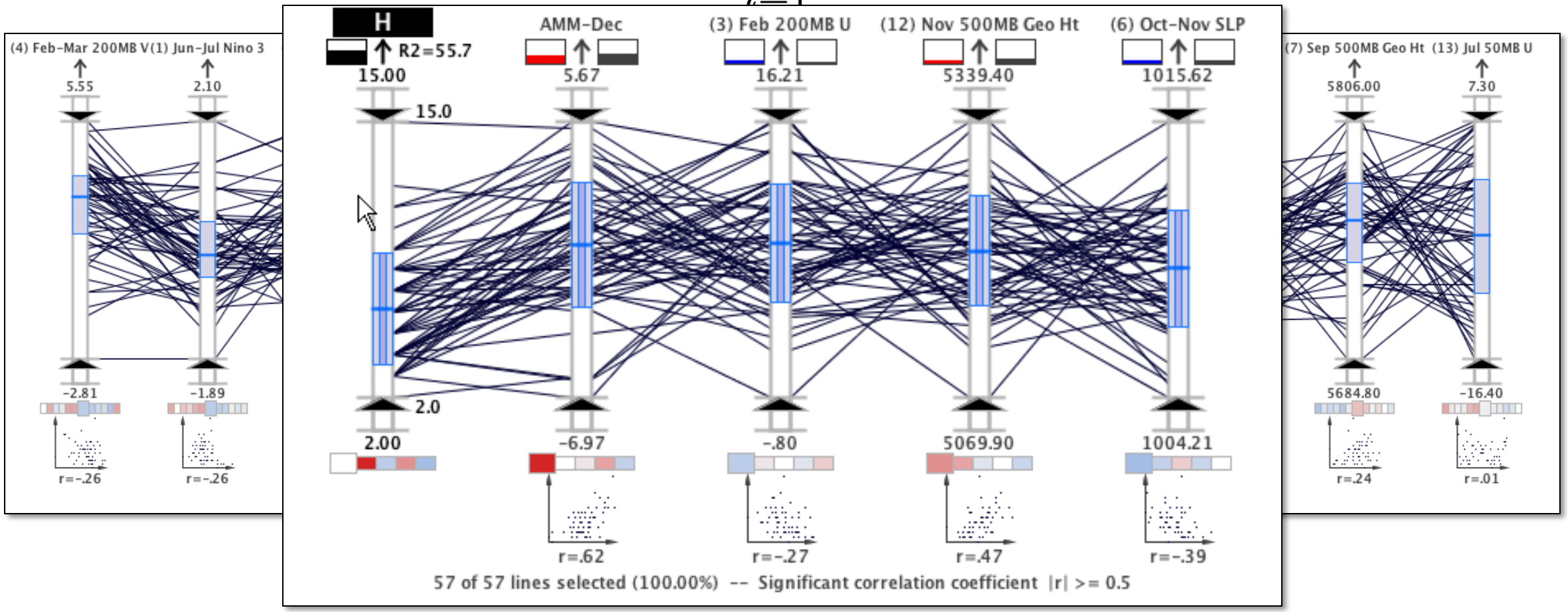
Multidimensional Data eXplorer (MDX)



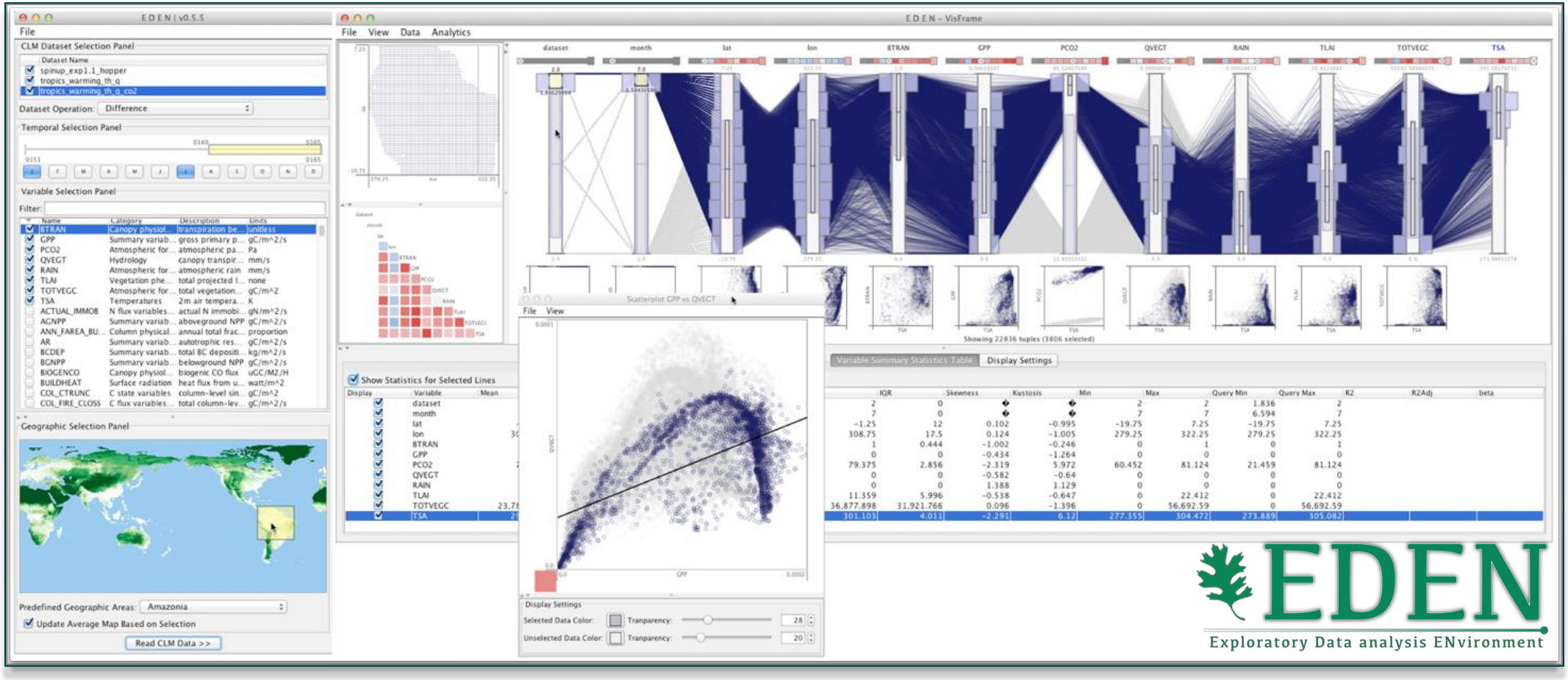
Citation: "Guided Analysis of Hurricane Trends using Statistical Processes Integrated with Interactive Parallel Coordinates". C. A. Steed, J. E. Swan II, T.J. Jankun-Kelly, and P. J. Fitzpatrick. Proceedings of the *Symposium on Visual Analytics Science and Technology (VAST)*, pp. 19–26, Oct. 2009.

MDX: Guided Multivariate Data Analysis & Explainable AI

$$(y - \overline{y})/\sigma_y = \sum_{i=1}^{k} b_i (x_i - \overline{x}_i)/\sigma_i$$

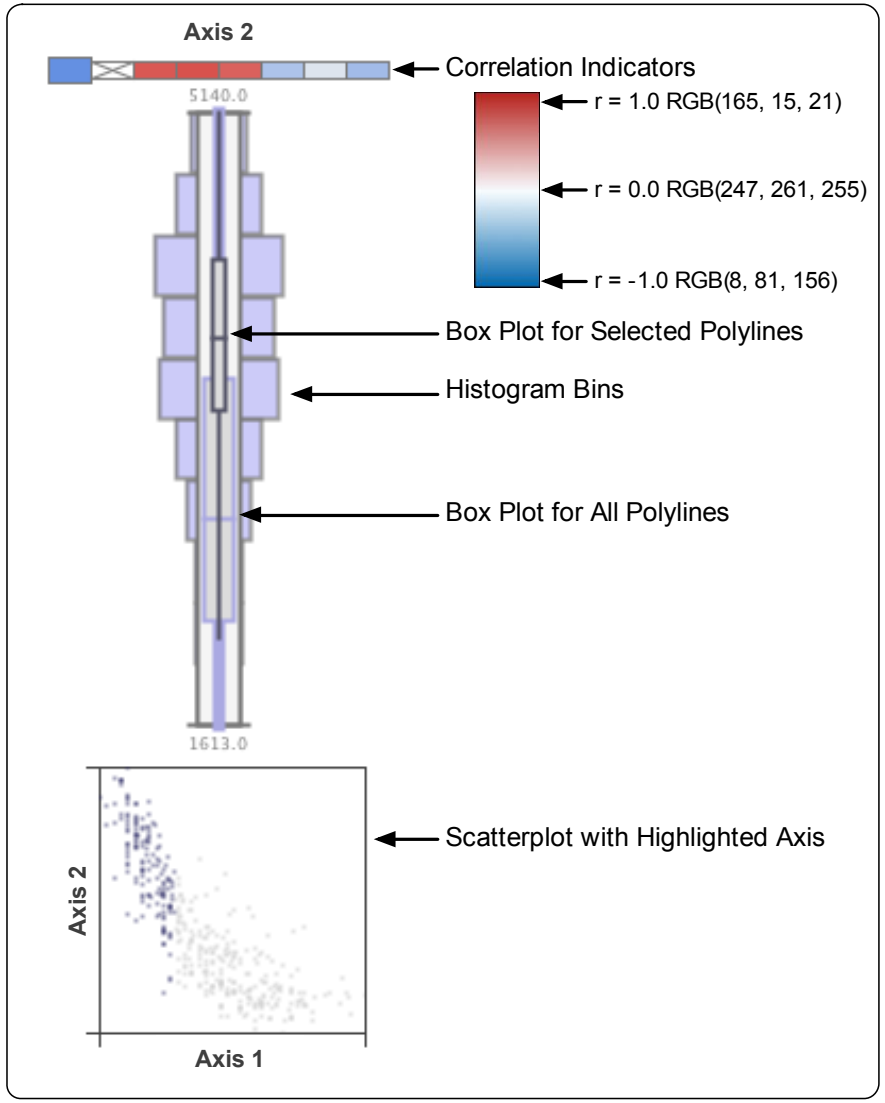


Exploratory Data analysis Environment (EDEN)
Scalable Multivariate Sensitivity Analysis for Climate Models



Citation: "Big Data Visual Analytics for Exploratory Earth System Simulation Analysis". C.A. Steed, D. M. Ricciuto, G. Shipman, B. Smith, P. E. Thornton, D. Wang, and D. N. Williams. *Computers & Geosciences*, 61:71-82, 2013.

Guided Analysis Using Correlation Mining

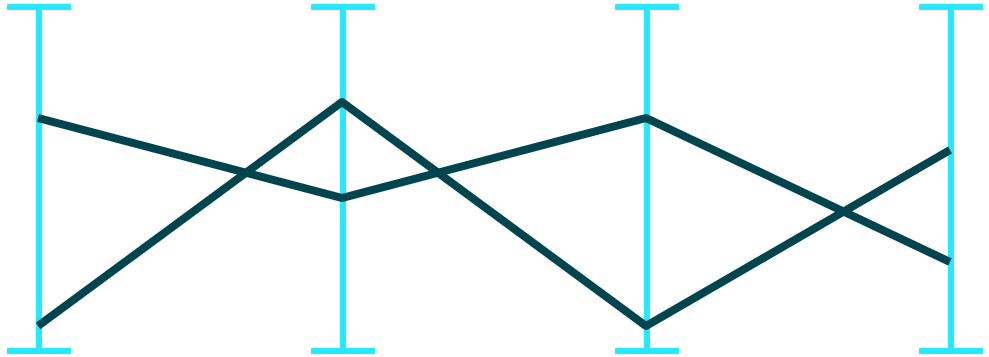


Graphical Correlation Indicators

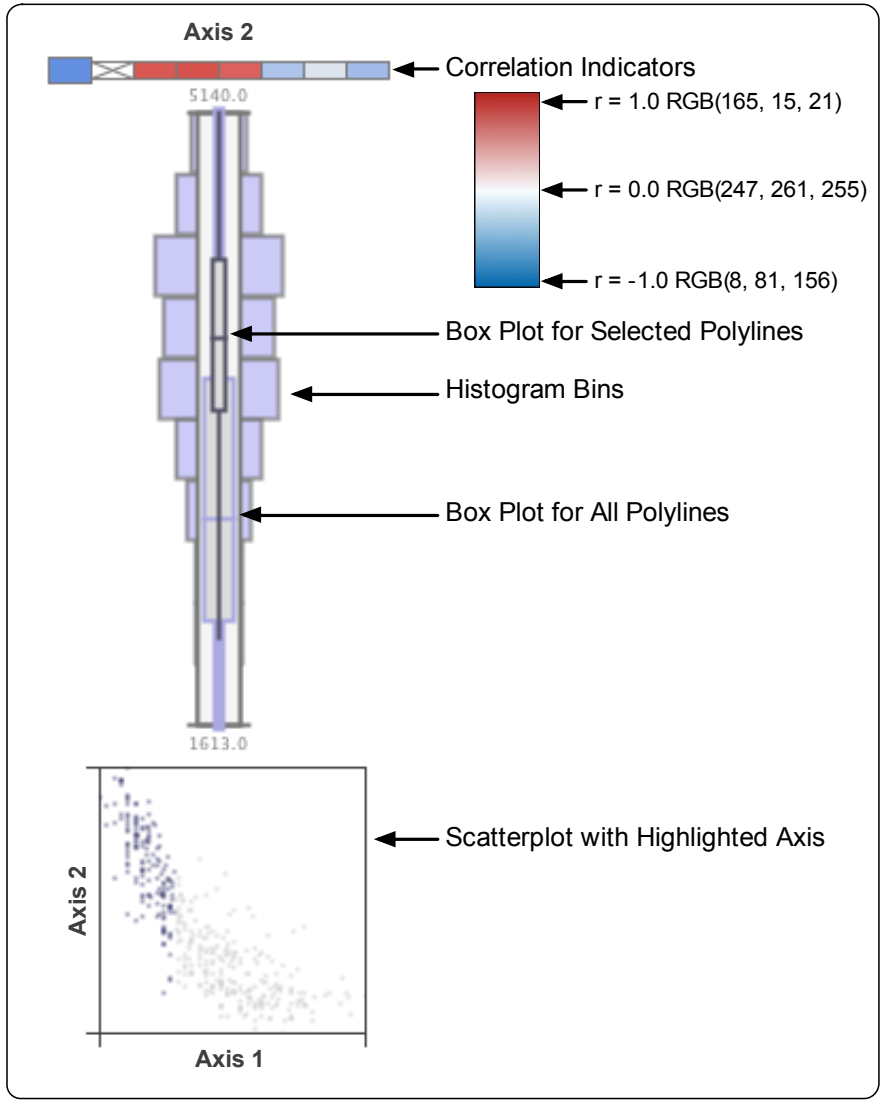
$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum {x_i}^2 - (\sum x_i)^2][n\sum {y_i}^2 - (\sum y_i)^2]}}$$

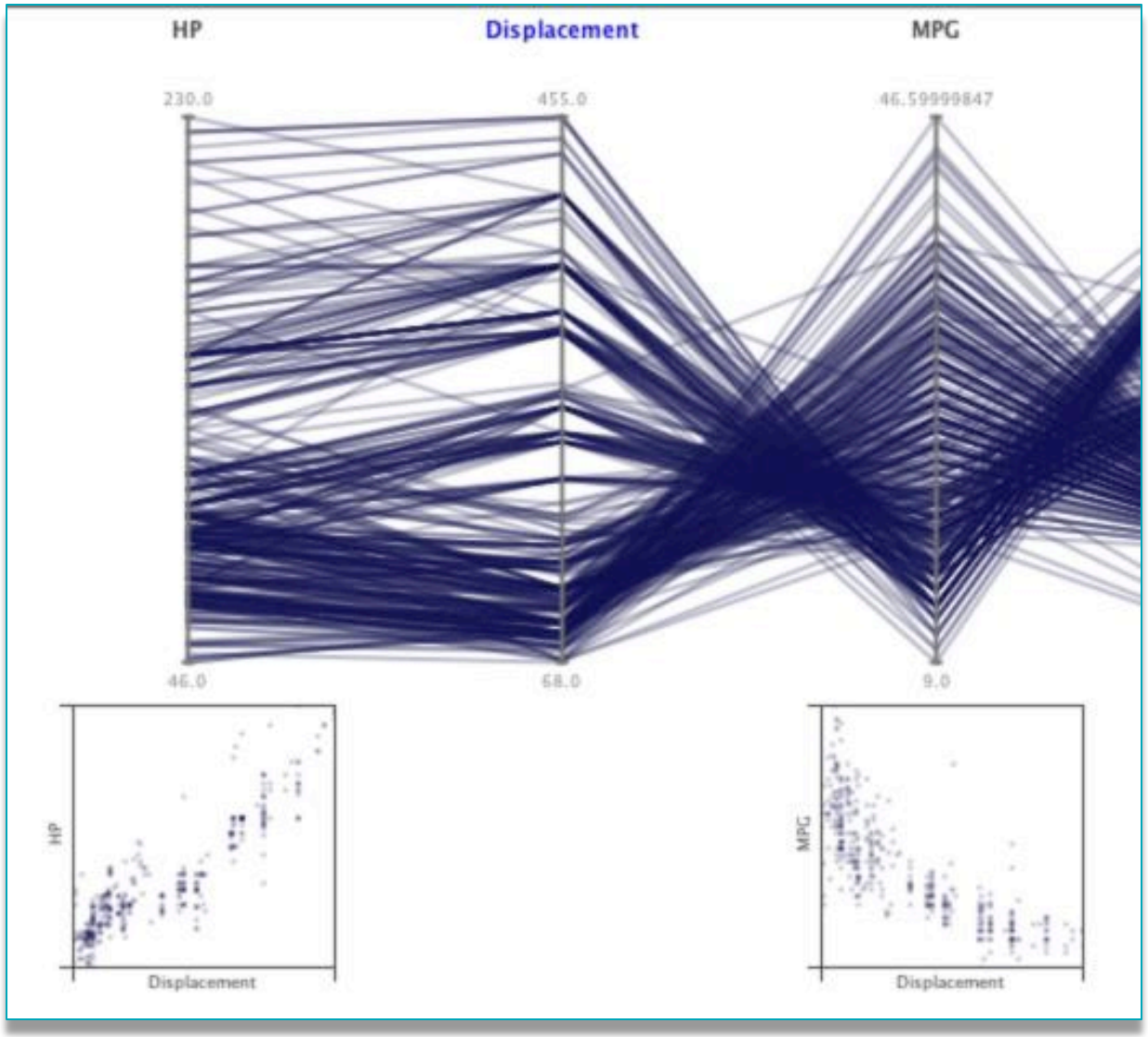
axis	1	2	3	4
1	1.0	0.4	0.6	0.2
2	0.4	1.0	0.1	0.8
3	0.6	0.1	1.0	0.5
4	0.2	0.8	0.5	1.0

axis 1	axis 2	axis 3	axis 4
1.0	0.4	0.6	0.2
0.4	1.0	0.1	0.8
0.6	0.1	1.0	0.5
0.2	0.8	0.5	1.0



Guided Analysis Using Information “Scent”

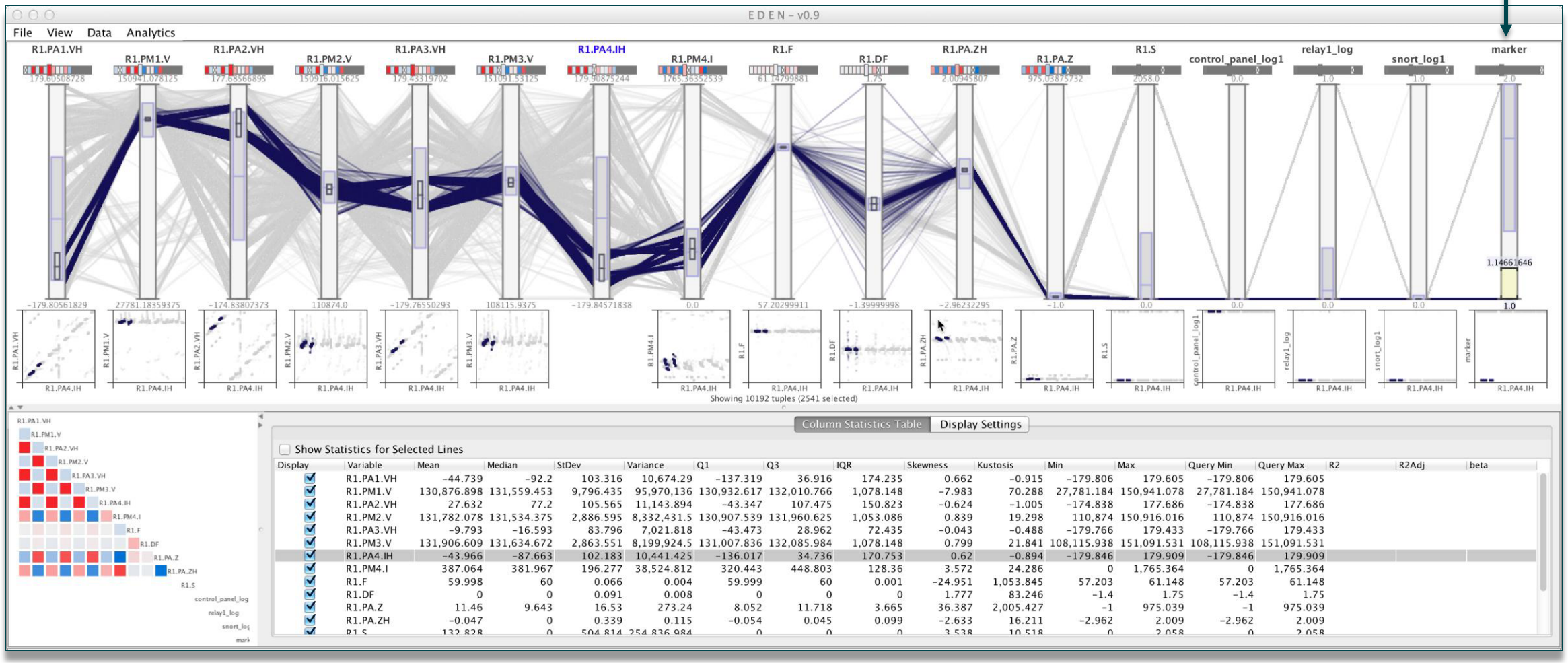




EDEN: Cyber Physical Data Analysis

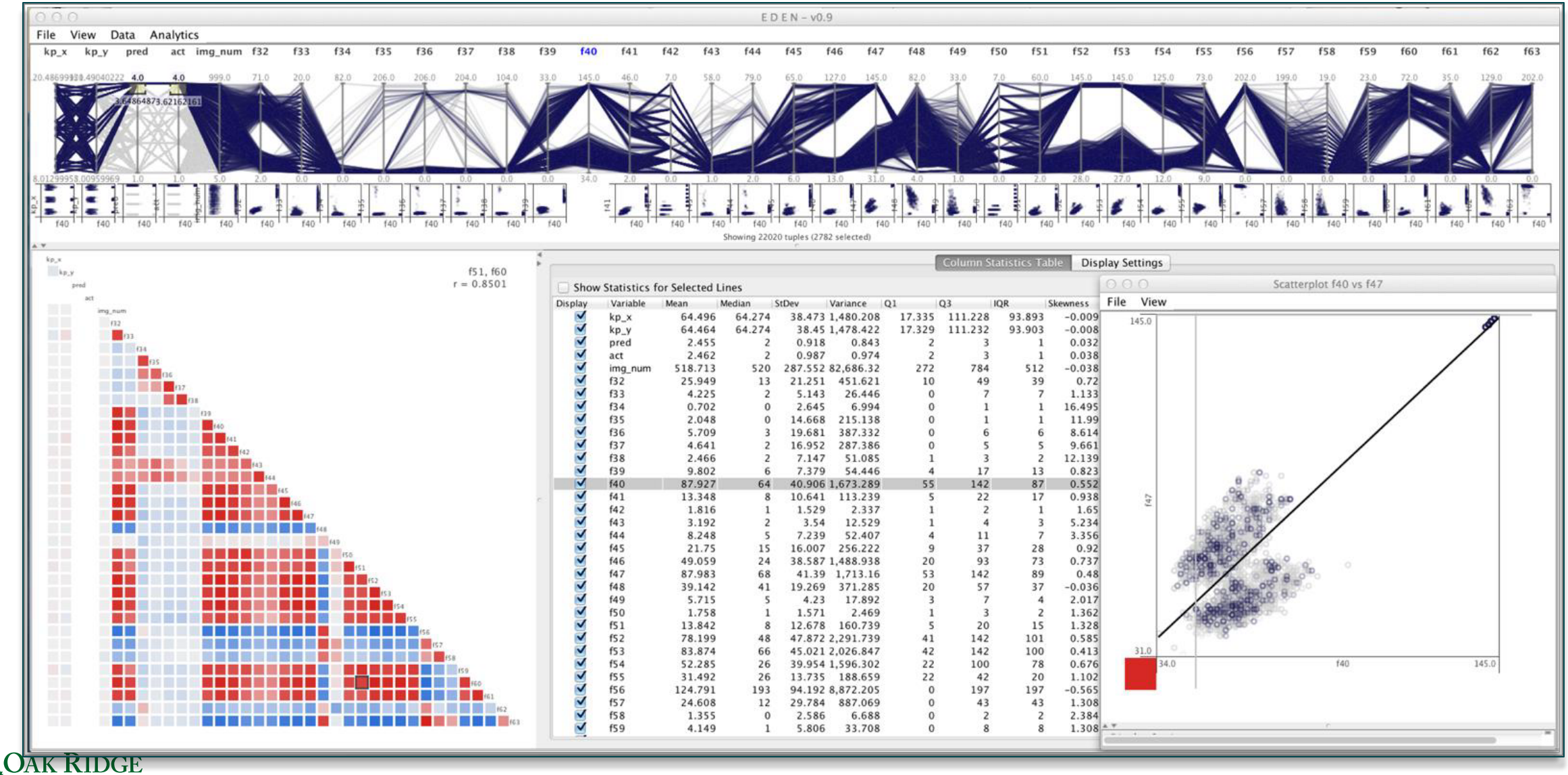
ORCA flag for normal vs. malicious activity

Sensor readings for normal activity highlighted



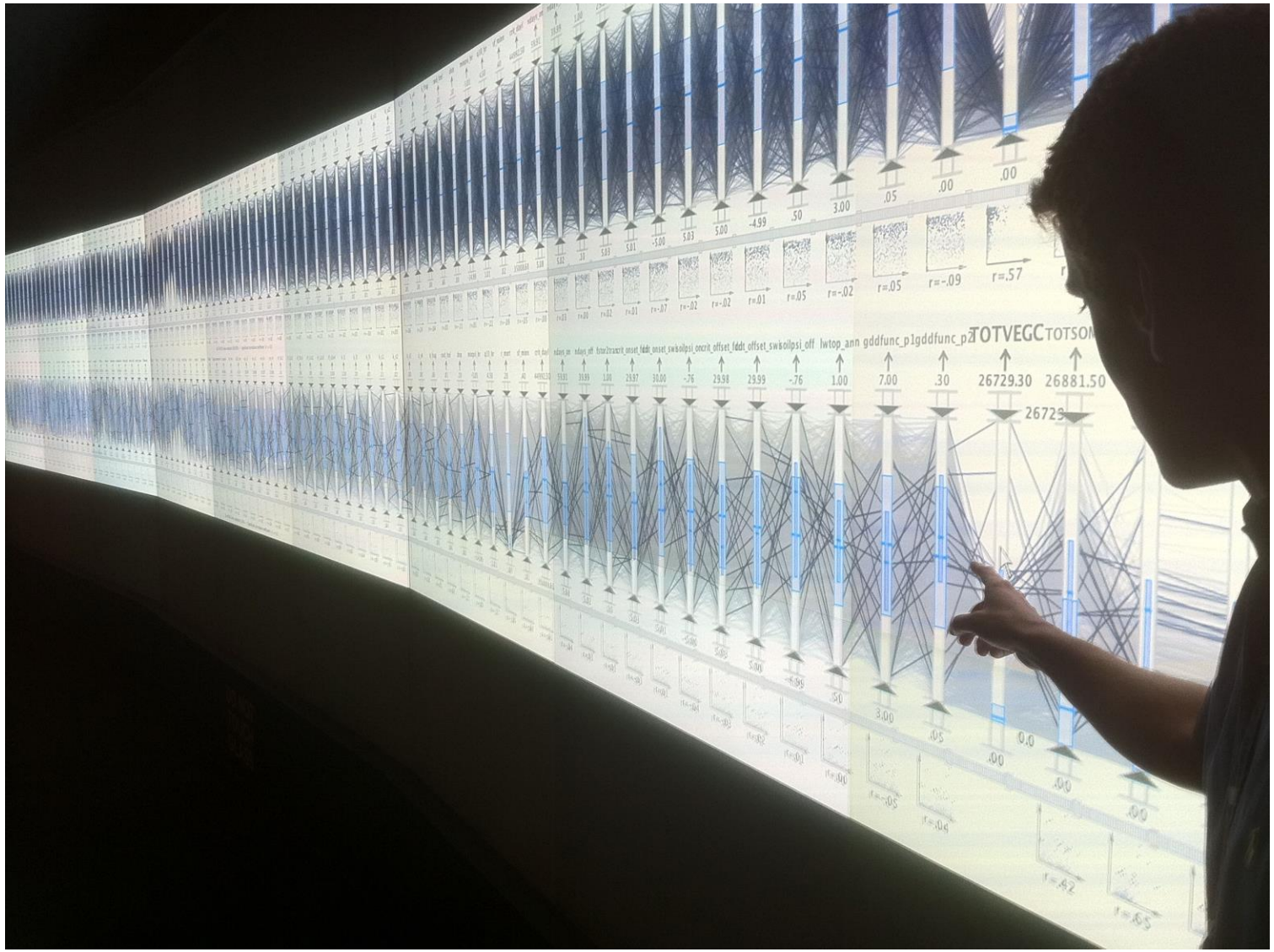
Display	Variable	Mean	Median	StDev	Variance	Q1	Q3	IQR	Skewness	Kustosis	Min	Max	Query Min	Query Max	R2	R2Adj	beta
☑	R1.PA1.VH	-44.739	-92.2	103.316	10,674.29	-137.319	36.916	174.235	0.662	-0.915	-179.806	179.605	-179.806	179.605			
☑	R1.PM1.V	130,876.898	131,559.453	9,796.435	95,970,136	130,932.617	132,010.766	1,078.148	-7.983	70.288	27,781.184	150,941.078	27,781.184	150,941.078			
☑	R1.PA2.VH	27.632	77.2	105.565	11,143.894	-43.347	107.475	150.823	-0.624	-1.005	-174.838	177.686	-174.838	177.686			
☑	R1.PM2.V	131,782.078	131,534.375	2,886.595	8,332,431.5	130,907.539	131,960.625	1,053.086	0.839	19.298	110,874	150,916.016	110,874	150,916.016			
☑	R1.PA3.VH	-9.793	-16.593	83.796	7,021.818	-43.473	28.962	72.435	-0.043	-0.488	-179.766	179.433	-179.766	179.433			
☑	R1.PM3.V	131,906.609	131,634.672	2,863.551	8,199,924.5	131,007.836	132,085.984	1,078.148	0.799	21.841	108,115.938	151,091.531	108,115.938	151,091.531			
☑	R1.PA4.IH	-43.966	-87.663	102.183	10,441.425	-136.017	34.736	170.753	0.62	-0.894	-179.846	179.909	-179.846	179.909			
☑	R1.PM4.I	387.064	381.967	196.277	38,524.812	320.443	448.803	128.36	3.572	24.286	0	1,765.364	0	1,765.364			
☑	R1.F	59.998	60	0.066	0.004	59.999	60	0.001	-24.951	1,053.845	57.203	61.148	57.203	61.148			
☑	R1.DF	0	0	0.091	0.008	0	0	0	1.777	83.246	-1.4	1.75	-1.4	1.75			
☑	R1.PA.Z	11.46	9.643	16.53	273.24	8.052	11.718	3.665	36.387	2,005.427	-1	975.039	-1	975.039			
☑	R1.PA.ZH	-0.047	0	0.339	0.115	-0.054	0.045	0.099	-2.633	16.211	-2.962	2.009	-2.962	2.009			
☑	R1.S	132.828	0	504.814	254,836.984	0	0	0	3.538	10.518	0	2.058	0	2.058			

EDEN: Machine Learning Feature Space

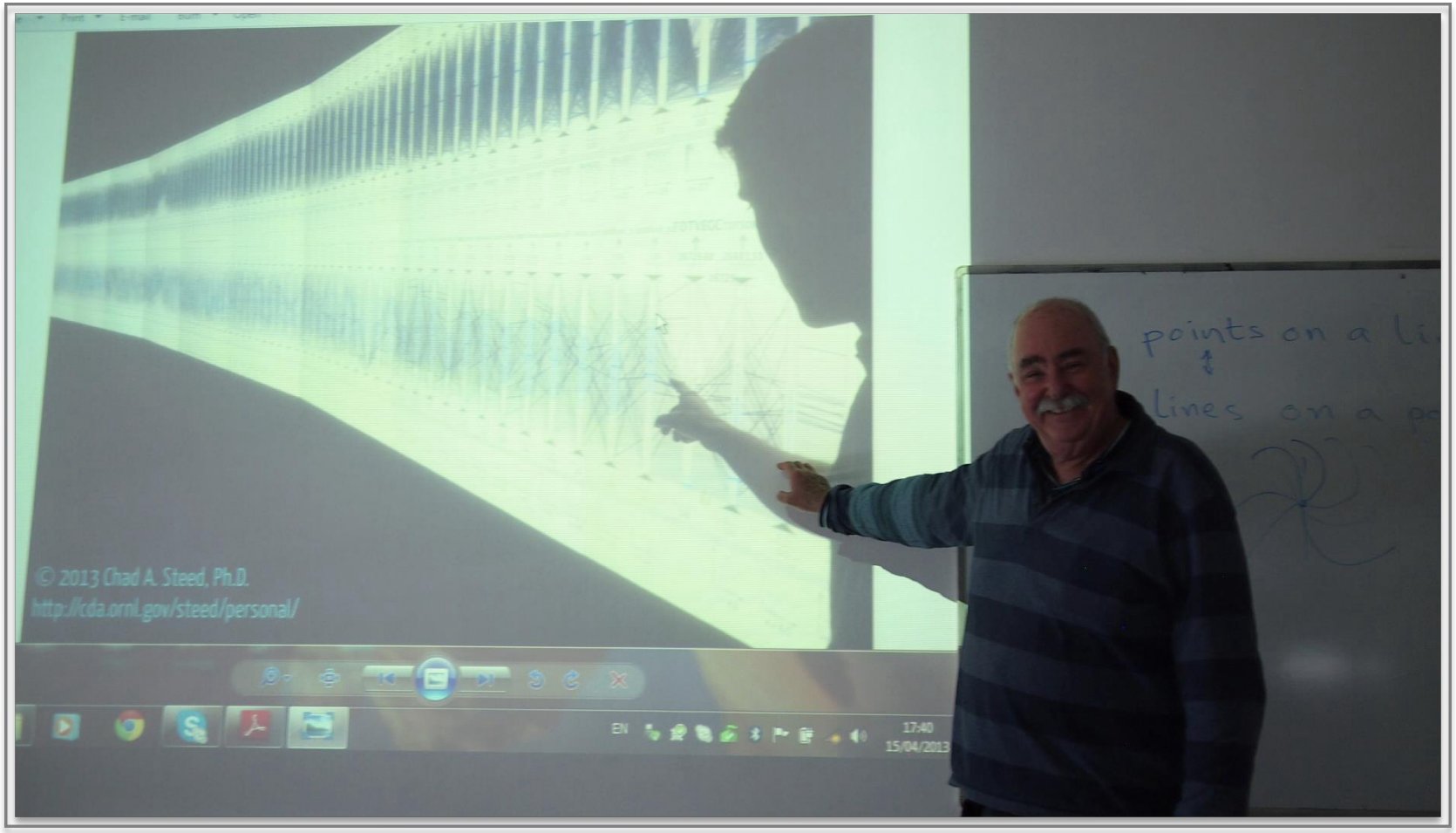


Display	Variable	Mean	Median	StDev	Variance	Q1	Q3	IQR	Skewness
✓	kp_x	64.496	64.274	38.473	1,480.208	17.335	111.228	93.893	-0.009
✓	kp_y	64.464	64.274	38.45	1,478.422	17.329	111.232	93.903	-0.008
✓	pred	2.455	2	0.918	0.843	2	3	1	0.032
✓	act	2.462	2	0.987	0.974	2	3	1	0.038
✓	img_num	518.713	520	287.552	82,686.32	272	784	512	-0.038
✓	f32	25.949	13	21.251	451.621	10	49	39	0.72
✓	f33	4.225	2	5.143	26.446	0	7	7	1.133
✓	f34	0.702	0	2.645	6.994	0	1	1	16.495
✓	f35	2.048	0	14.668	215.138	0	1	1	11.99
✓	f36	5.709	3	19.681	387.332	0	6	6	8.614
✓	f37	4.641	2	16.952	287.386	0	5	5	9.661
✓	f38	2.466	2	7.147	51.085	1	3	2	12.139
✓	f39	9.802	6	7.379	54.446	4	17	13	0.823
✓	f40	87.927	64	40.906	1,673.289	55	142	87	0.552
✓	f41	13.348	8	10.641	113.239	5	22	17	0.938
✓	f42	1.816	1	1.529	2.337	1	2	1	1.65
✓	f43	3.192	2	3.54	12.529	1	4	3	5.234
✓	f44	8.248	5	7.239	52.407	4	11	7	3.356
✓	f45	21.75	15	16.007	256.222	9	37	28	0.92
✓	f46	49.059	24	38.587	1,488.938	20	93	73	0.737
✓	f47	87.983	68	41.39	1,713.16	53	142	89	0.48
✓	f48	39.142	41	19.269	371.285	20	57	37	-0.036
✓	f49	5.715	5	4.23	17.892	3	7	4	2.017
✓	f50	1.758	1	1.571	2.469	1	3	2	1.362
✓	f51	13.842	8	12.678	160.739	5	20	15	1.328
✓	f52	78.199	48	47.872	2,291.739	41	142	101	0.585
✓	f53	83.874	66	45.021	2,026.847	42	142	100	0.413
✓	f54	52.285	26	39.954	1,596.302	22	100	78	0.676
✓	f55	31.492	26	13.735	188.659	22	42	20	1.102
✓	f56	124.791	193	94.192	8,872.205	0	197	197	-0.565
✓	f57	24.608	12	29.784	887.069	0	43	43	1.308
✓	f58	1.355	0	2.586	6.688	0	2	2	2.384
✓	f59	4.149	1	5.806	33.708	0	8	8	1.308

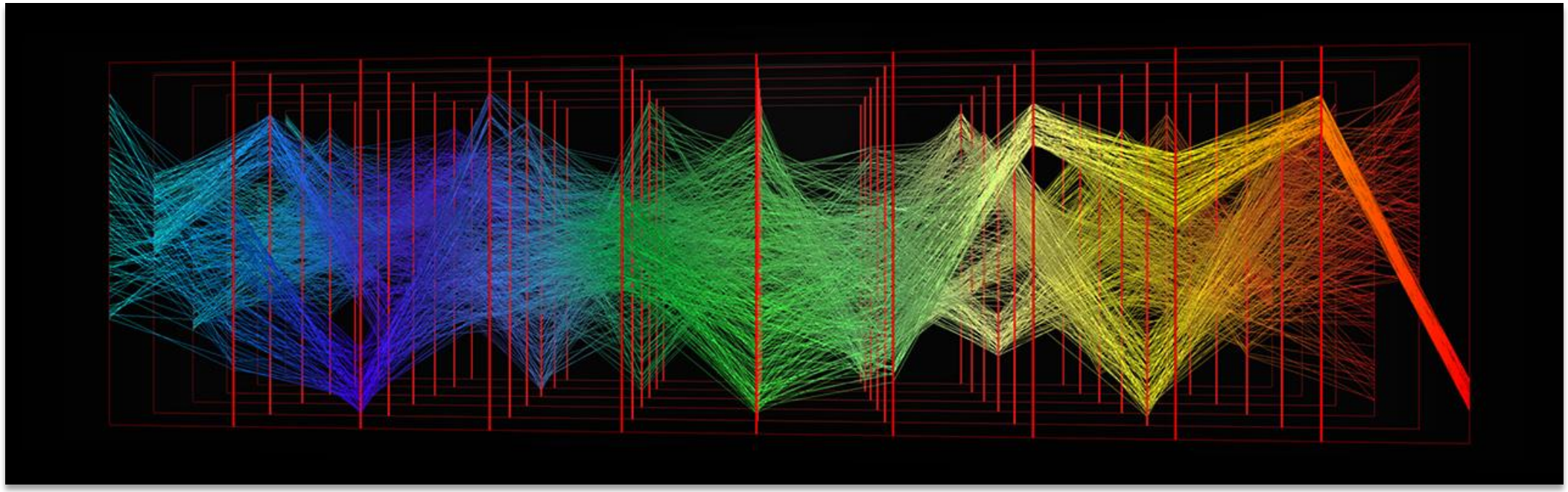
EDEN in EVEREST

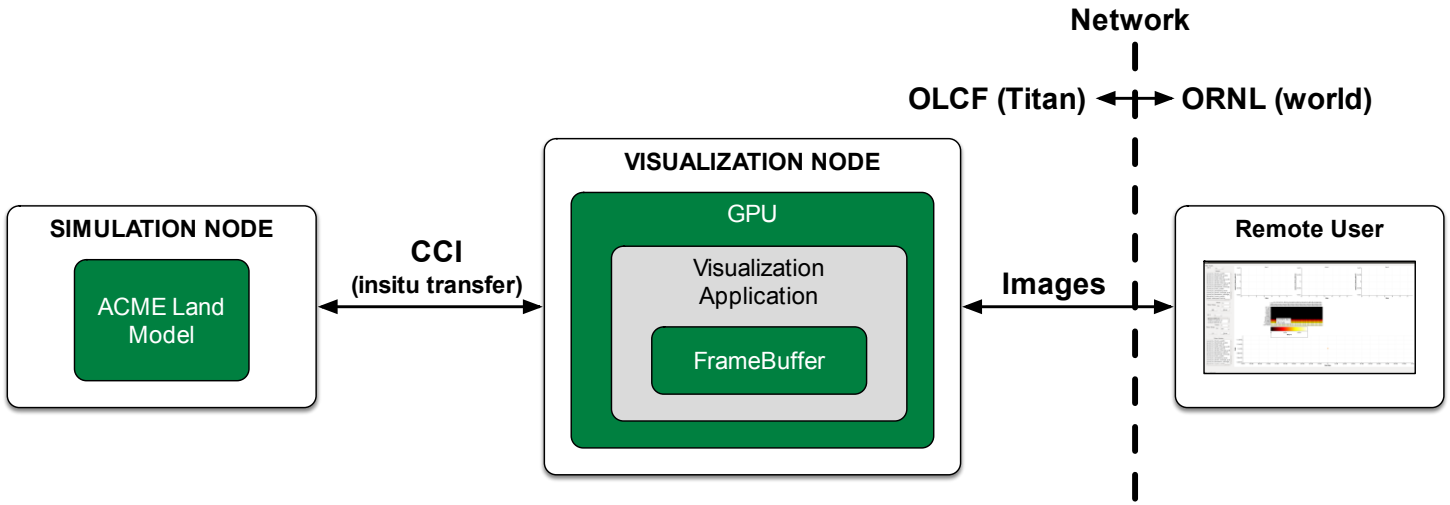


How I Met Alfred Inselberg

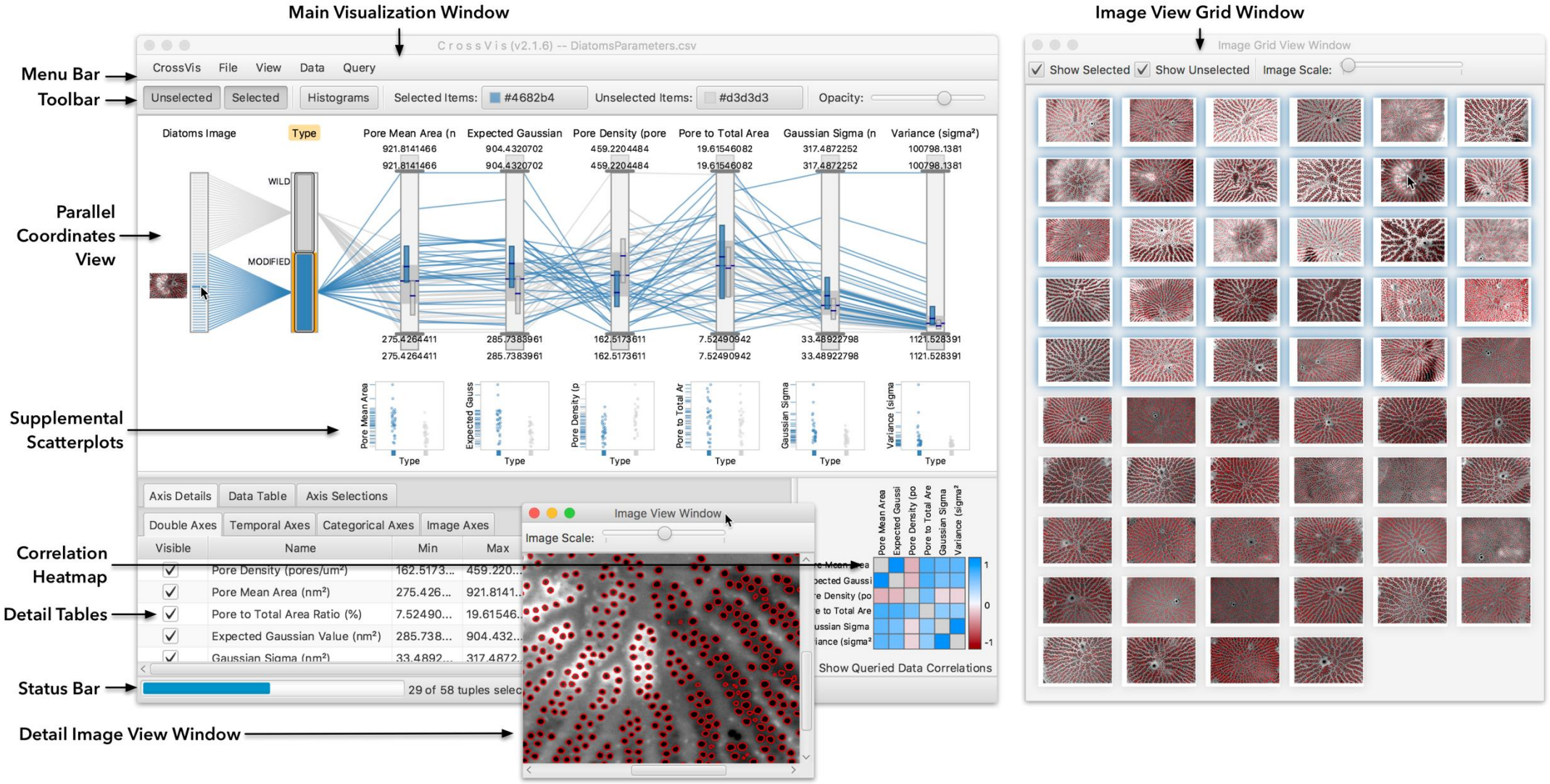


In Situ Visual Analytics using OLCF (Titan)



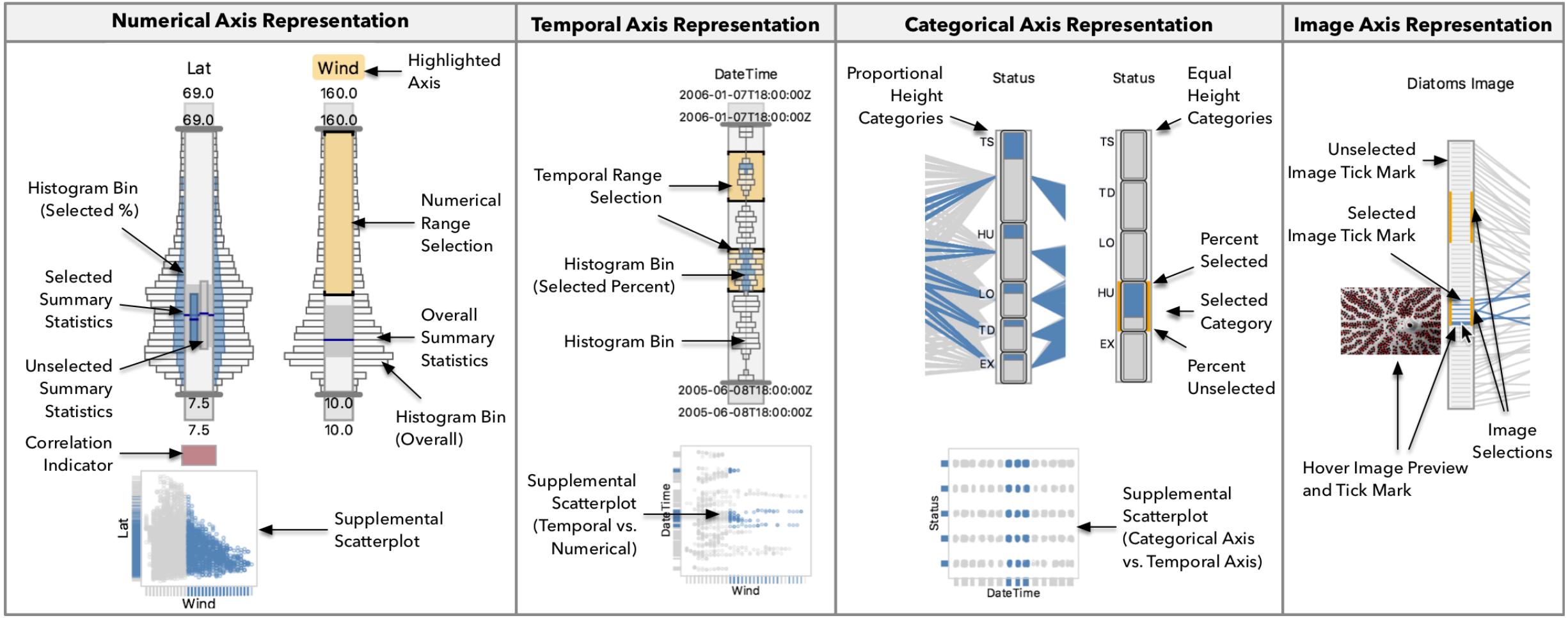


CrossVis: Multi-modal Data Exploration and Explainable AI



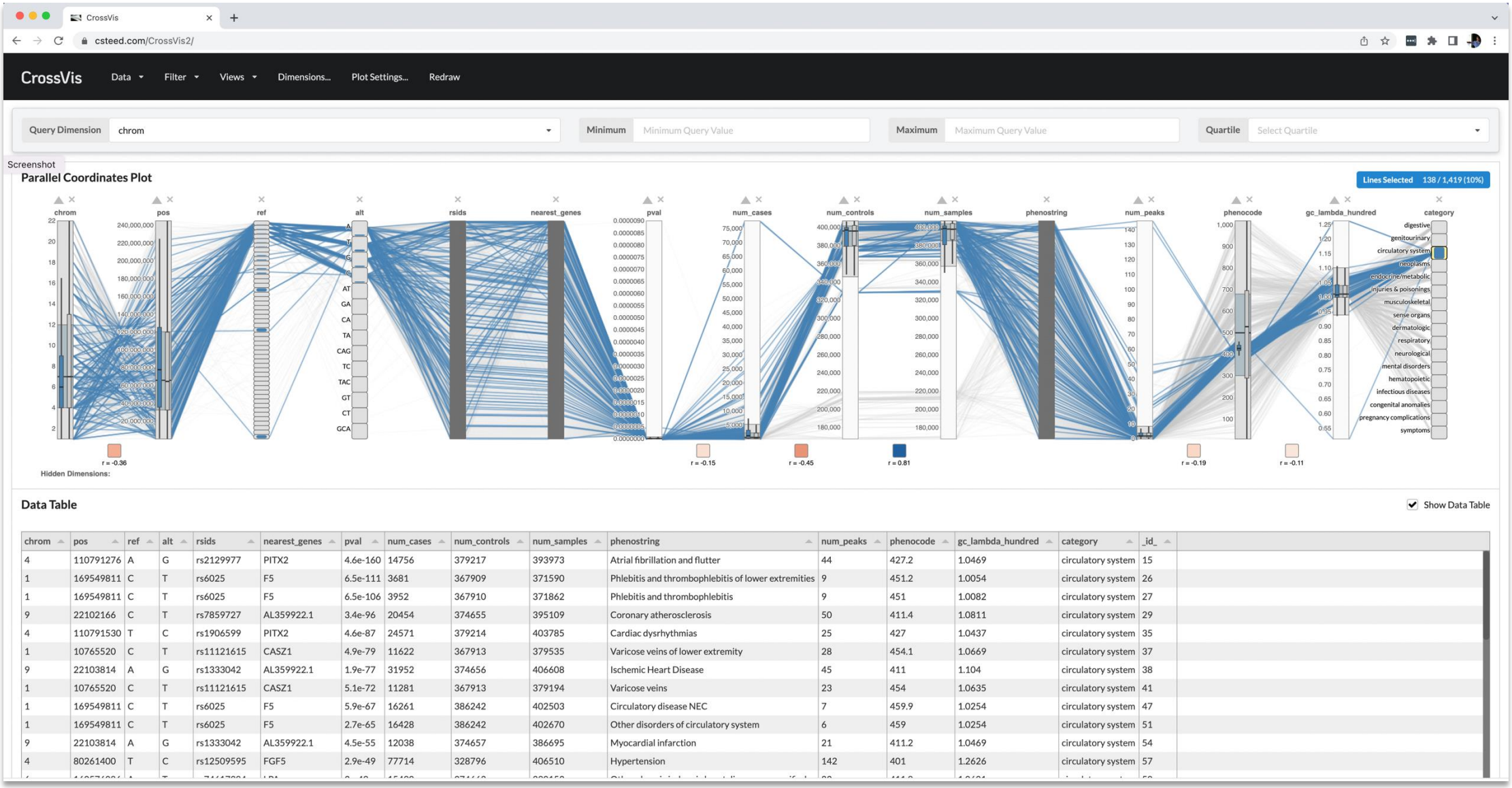
Citation: "CrossVis: A Visual Analytics System for Exploring Heterogeneous Multivariate Data with Applications to Materials and Climate Sciences". C. A. Steed, J. Goodall, J. Chae, and A. Trofimov. *Graphics & Visual Computing*, 3:20013, 2020.

CrossVIS: Supporting Heterogeneous Datatypes



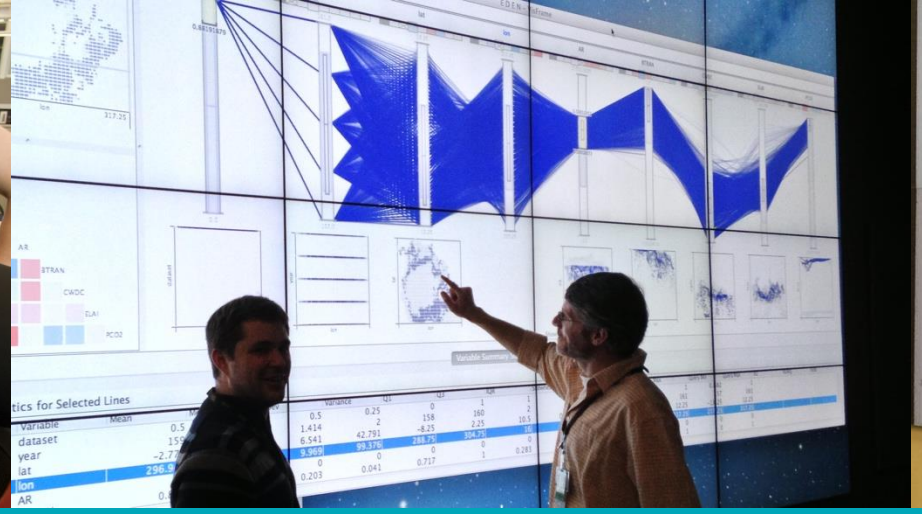
CrossVis2 on the Web

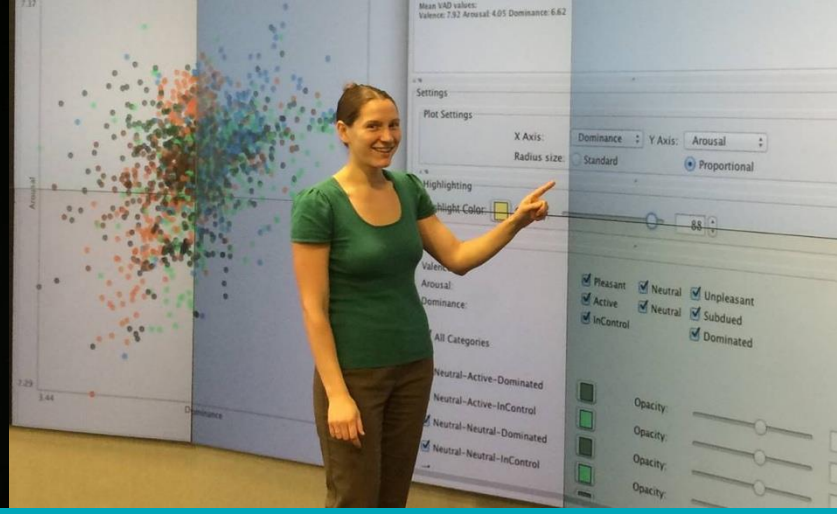
https://ornl.github.io/CrossVis2



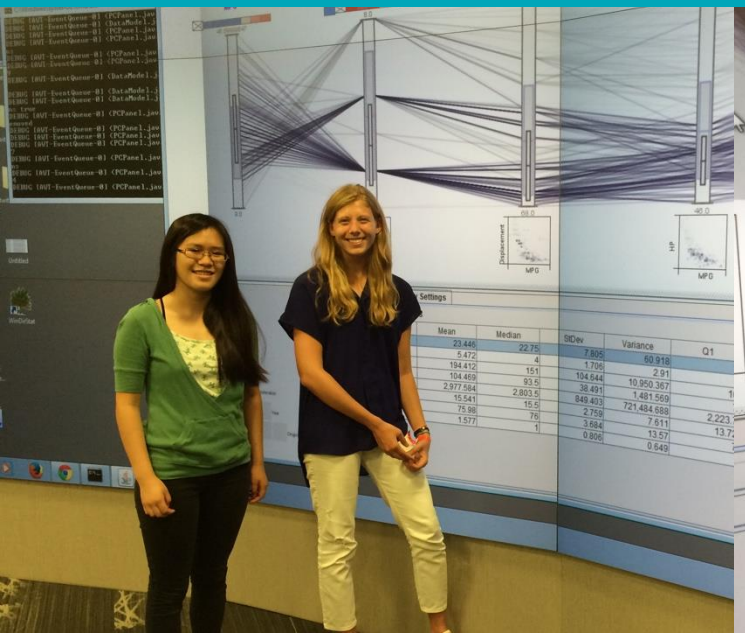
chrom	pos	ref	alt	rsids	nearest_genes	pval	num_cases	num_controls	num_samples	phenostring	num_peaks	phenocode	gc_lambda_hundred	category	_id_
4	110791276	A	G	rs2129977	PITX2	4.6e-160	14756	379217	393973	Atrial fibrillation and flutter	44	427.2	1.0469	circulatory system	15
1	169549811	C	T	rs6025	F5	6.5e-111	3681	367909	371590	Phlebitis and thrombophlebitis of lower extremities	9	451.2	1.0054	circulatory system	26
1	169549811	C	T	rs6025	F5	6.5e-106	3952	367910	371862	Phlebitis and thrombophlebitis	9	451	1.0082	circulatory system	27
9	22102166	C	T	rs7859727	AL359922.1	3.4e-96	20454	374655	395109	Coronary atherosclerosis	50	411.4	1.0811	circulatory system	29
4	110791530	T	C	rs1906599	PITX2	4.6e-87	24571	379214	403785	Cardiac dysrhythmias	25	427	1.0437	circulatory system	35
1	10765520	C	T	rs11121615	CASZ1	4.9e-79	11622	367913	379535	Varicose veins of lower extremity	28	454.1	1.0669	circulatory system	37
9	22103814	A	G	rs1333042	AL359922.1	1.9e-77	31952	374656	406608	Ischemic Heart Disease	45	411	1.104	circulatory system	38
1	10765520	C	T	rs11121615	CASZ1	5.1e-72	11281	367913	379194	Varicose veins	23	454	1.0635	circulatory system	41
1	169549811	C	T	rs6025	F5	5.9e-67	16261	386242	402503	Circulatory disease NEC	7	459.9	1.0254	circulatory system	47
1	169549811	C	T	rs6025	F5	2.7e-65	16428	386242	402670	Other disorders of circulatory system	6	459	1.0254	circulatory system	51
9	22103814	A	G	rs1333042	AL359922.1	4.5e-55	12038	374657	386695	Myocardial infarction	21	411.2	1.0469	circulatory system	54
4	80261400	T	C	rs12509595	FGF5	2.9e-49	77714	328796	406510	Hypertension	142	401	1.2626	circulatory system	57

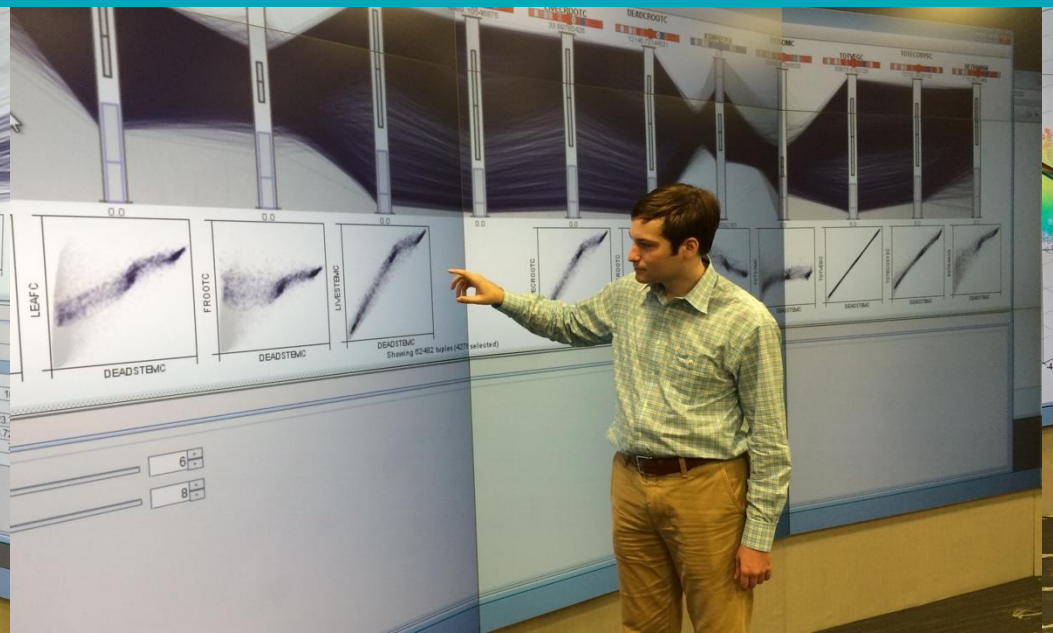
Oculus
3D
Data Visualization
Station

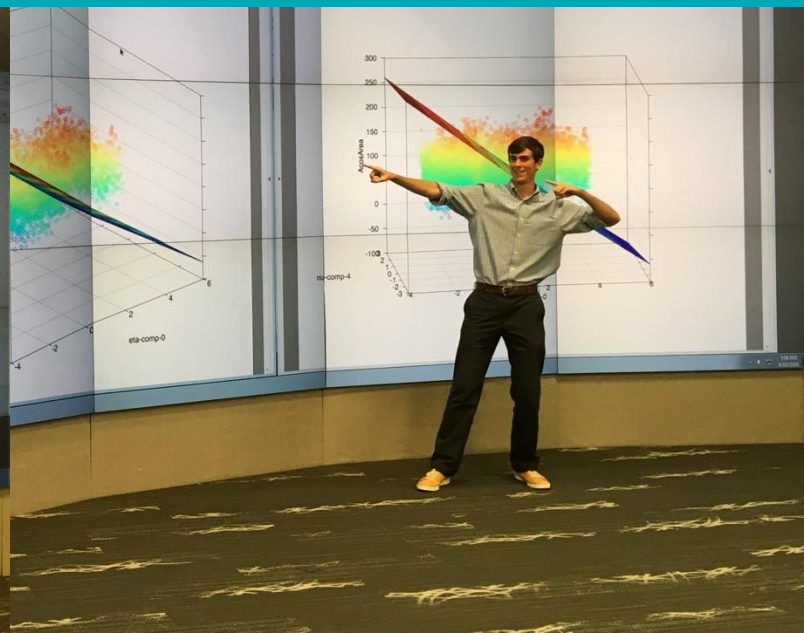



Settings
Plot Settings
X Axis:
Dominance
Y Axis:
Arousal
Radius size:
Standard
Proportional
Highlighting
Pleasant
Neutral
Unpleasant
Active
Neutral
Subdued
InControl
Dominated
All Categories
Neutral-Active-Dominated
Neutral-Active-InControl
Neutral-Neutral-Dominated
Neutral-Neutral-InControl
Opacity

Thank You!




LEAFC
FROOTC
DEADSTEMC
DEADSTEMC
LIVESTEMC
DEADSTEMC


eta-comp-0
nu-comp-4