Uncertainty and Parameter Space Analysis in Visualization | Oct 15, 2012

http://tinyurl.com/8hwtzzr

STATISTICAL UNCERTAINTY

From Quantification to Visualization

Kristi Potter

Scientific Computing and Imaging Institute

University of Utah

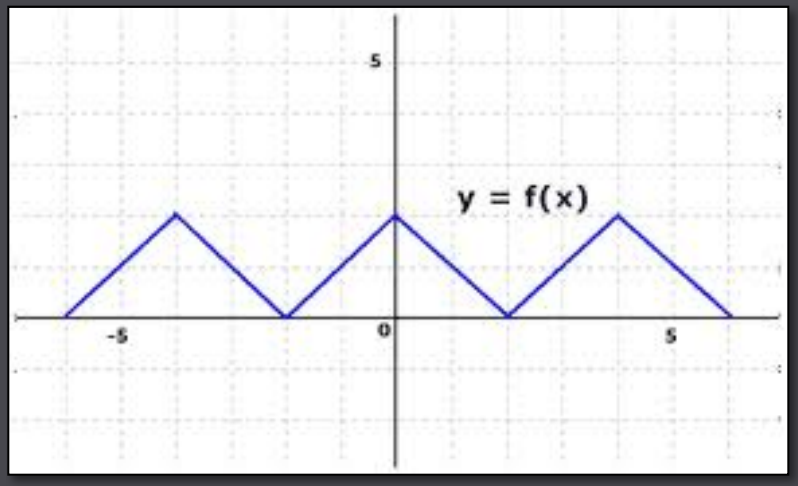
VisWeek 2012 Tutorial

SCI INSTITUTE

Statistical Uncertainties

random fluctuations of measurement

Sources of Uncertainty



Sources

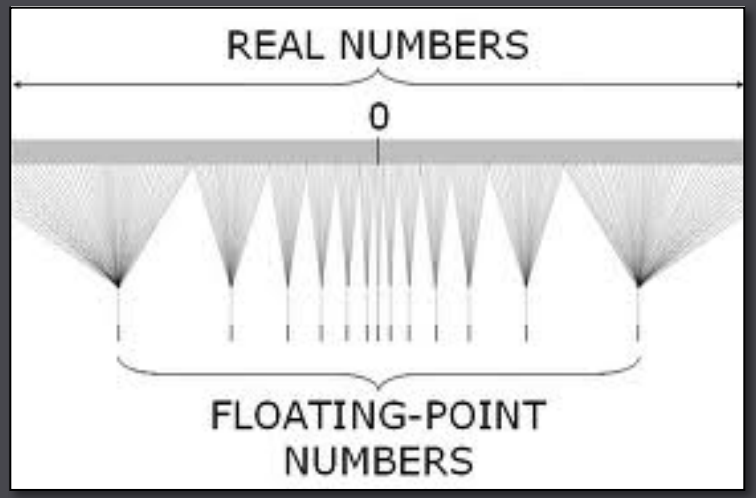
Parameter Uncertainty

- *influence of model parameters*
- *(exact, correct, best) values unknown*
- *cannot be controlled by experiment*

Sources

Model Inadequacy

- *aka model bias, model discrepancy*
- *model is an approximation*
- *lack of knowledge of the underlying problem*
- *accuracy discrepancy of the model to reality*



Sources

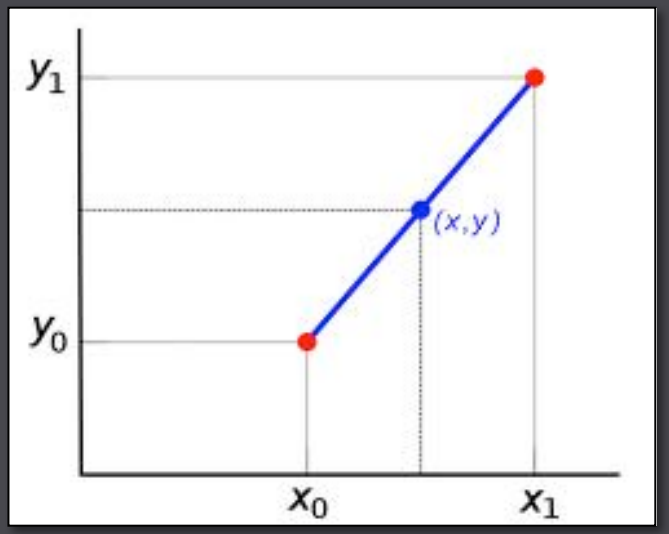
Algorithmic Uncertainty

- *aka numerical uncertainty*
- *numerical errors, approximations*
- *translation of mathematical model to the computer*

Sources

Experimental Uncertainty

- *aka observational error*
- *variability of experimental measurements*
- *non-determinism*



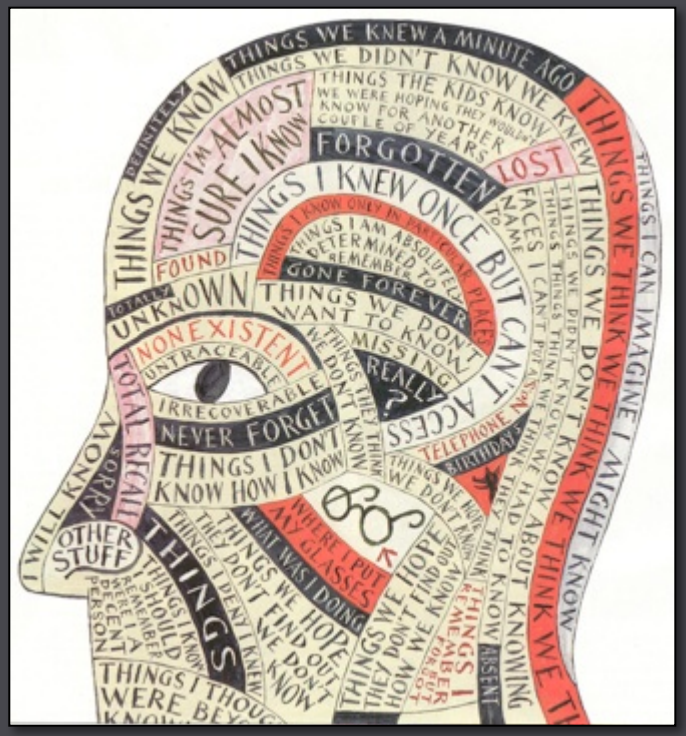
Sources

Interpolation Uncertainty

- *lack of available data*
- *interpolate/extrapolate for desired response*
- *choice of interpolation method*

Categories

Categories



Epistemic Uncertainty

- *aka* ***systematic*** *uncertainty*
- *things we could in principle know but don't in practice*
- *i.e. insufficient measurement or modeling, missing data*

Reducible: *can be alleviated by better models, more accurate measurement*

SCI INSTITUTE

Categories

Aleatoric Uncertainty

- *aka* ***statistical*** *uncertainty*
- *unknowns that differ on each run*
- *i.e. throwing dice*

Irreducible: *cannot be eliminated through improvements in models or measurements*

Quantification

reduce epistemic uncertainties to aleatoric
get to a relatively straightforward quantification

use statistics for quantification

Statistical Uncertainties

common representations for visualization

Probability Distribution Functions (PDFs)

approximate outcome through a probability function

Probability Density

continuous random variables

frequency of outcome values

Probability Distribution Functions (PDFs)

approximate outcome through a probability function

Categorical Distribution

discrete random variables

finite set of outcome values

Statistics on PDFs

Mean

expected value, arithmetic mean

Statistics on PDFs

Median

middlemost value

Statistics on PDFs

Standard Deviation

spread of values

Statistics on PDFs

Mode

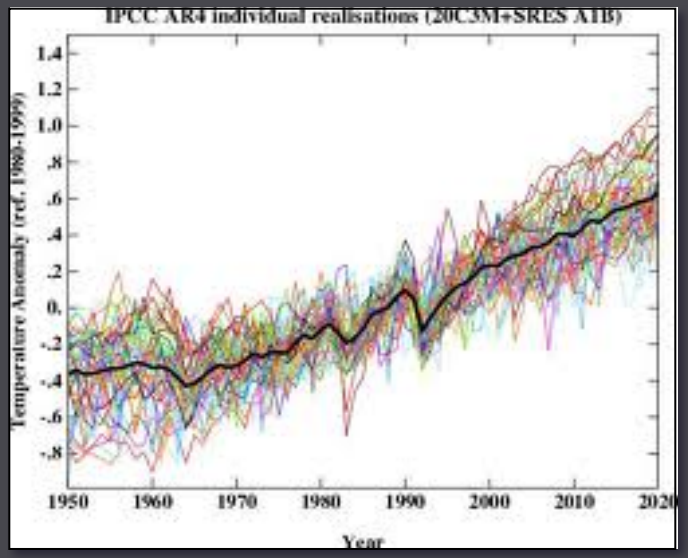
most frequently occurring value

Statistics on PDFs

SUPPORT/RANGE

interval where value probability is not zero

Uncertainty in Data

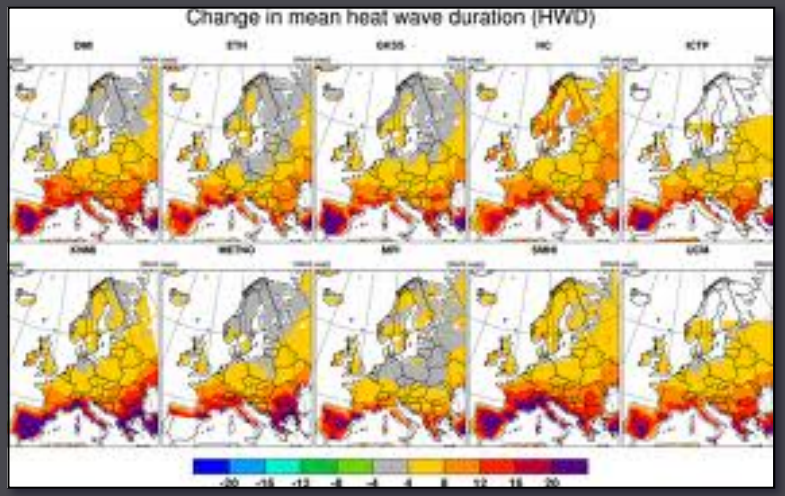


Ensembles

Multi-Run Simulations

- *explore space of parameters*
- *mitigate model error*
- *cover range of initial conditions / outcomes*
- *combine multiple models*

SCI INSTITUTE



Ensembles

Collection of Datasets

- *members, realizations*
- *full simulation run for each parameter set/input condition*

SCI INSTITUTE

Ensembles

MULTI-

- *dimensional*

  spatial domain & time

- *variate*

  simulate over many variables

- *valued*

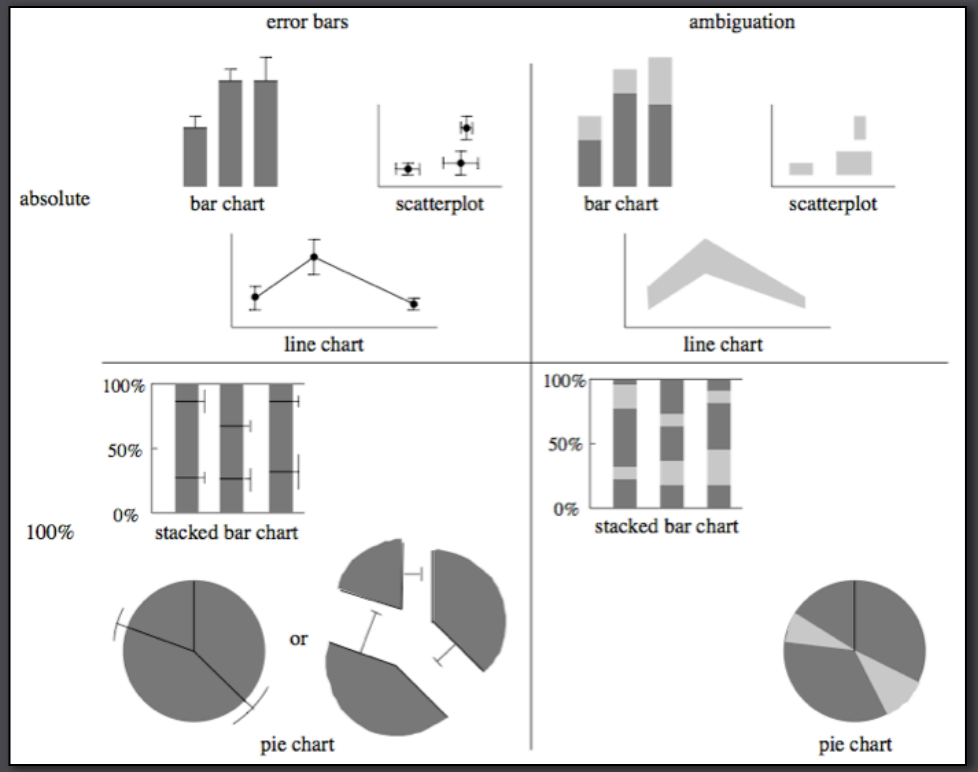
  many vales for each variable/location

Graphical Data Analysis

visual indication of pdf

Ambiguation

- *numeric interval guaranteed to contain data value*
- *no assumptions about the pdf within the interval*

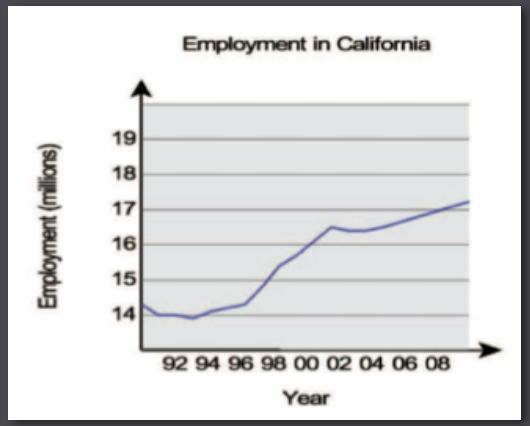


Statistical Uncertainty Bounded Uncertainty

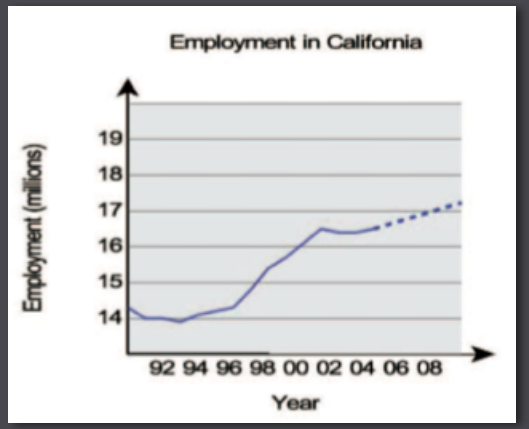
Visualizing Data with Bounded Uncertainty.
C. Olston, J.D. Mackinlay. InfoVis, 2002.

Information Uncertainty

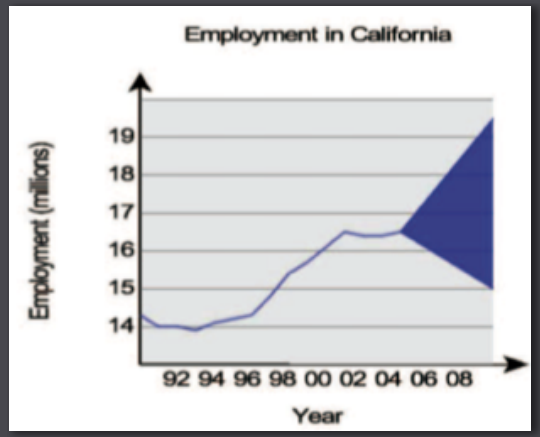
- *indication of known information*
- *qualitative rather than quantitative*
- *spreadsheet interface characterizes the data*



Average Growth



Estimated Growth



Possible Growth

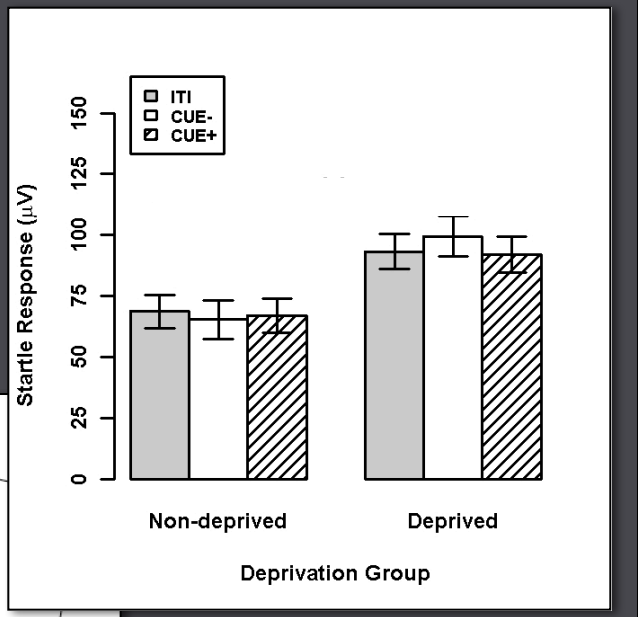
Likely Growth

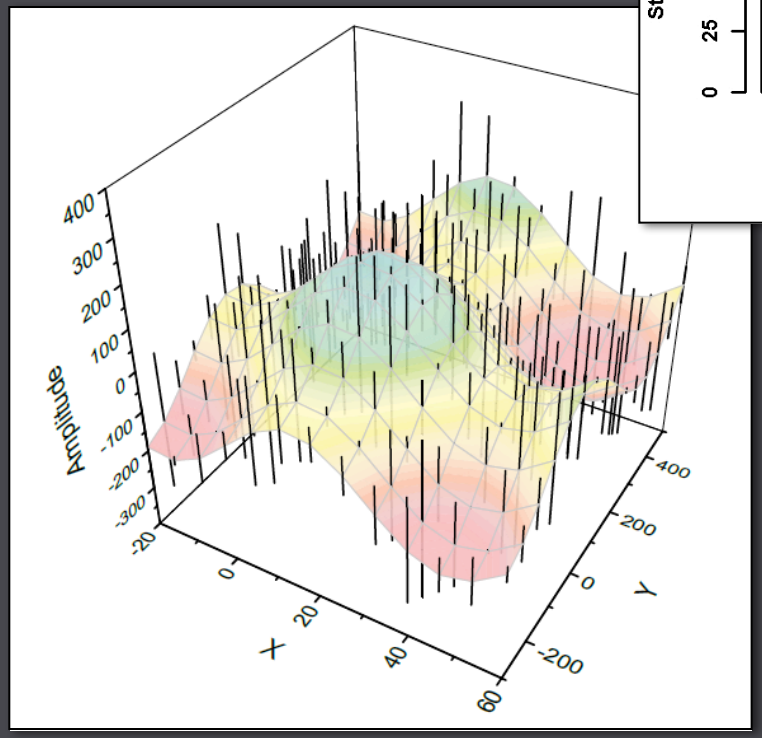
A Spreadsheet Approach to Facilitate Visualization of Uncertainty in Information.
A. Streit, B. Pham, R. Brown. TVCG 14(1), 2008.

SCI INSTITUTE

Error Bars

- *convey accuracy by amount of +/- error*
- *std dev or std error*

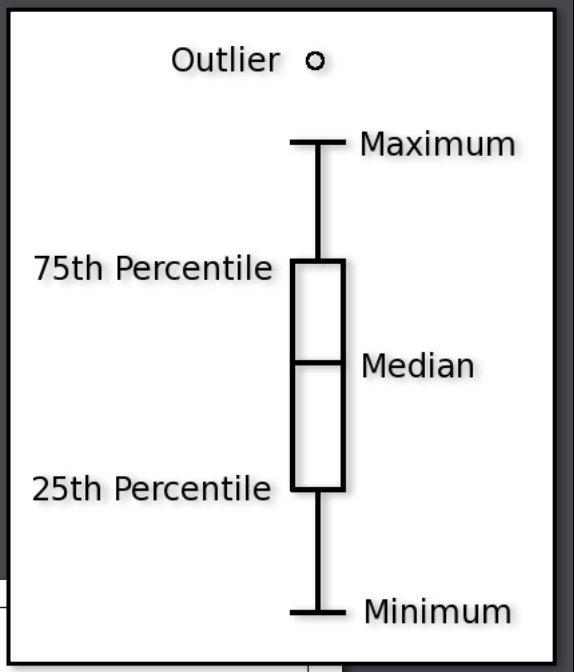


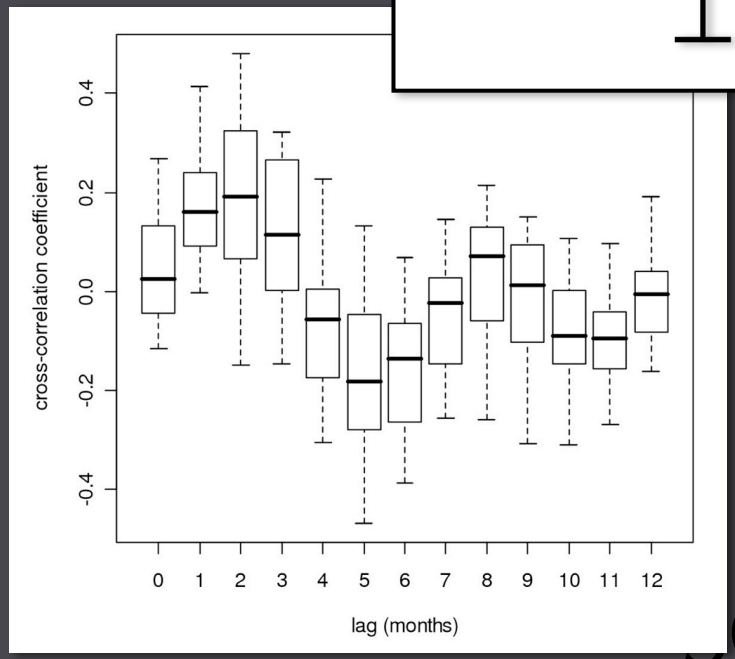


SCI INSTITUTE

Boxplots

- *quartile range including median, outliers*
- *assume Gaussian*



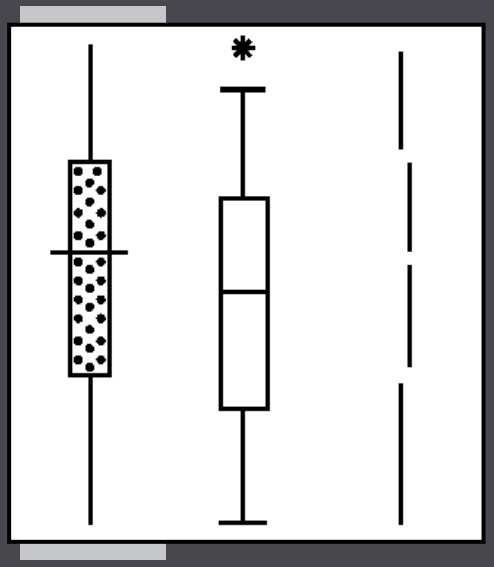


Boxplot Modifications

Visual Modifications

refinement for aesthetic purposes

range bar



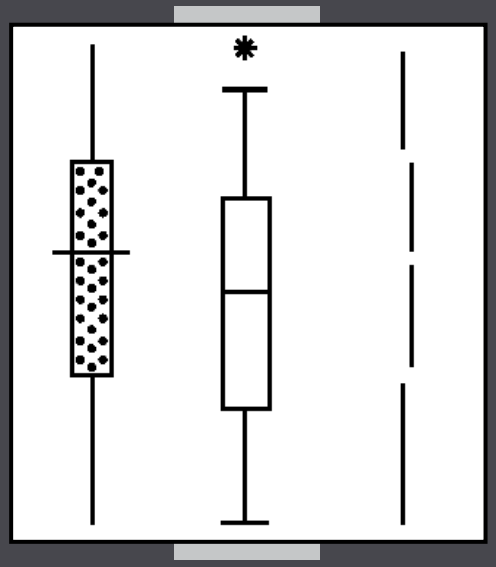
Charting Statistics.
Mary Eleanor Spear. McGraw-Hill, 1952.

SCI INSTITUTE

Boxplot Modifications

Visual Modifications

refinement for aesthetic purposes



boxplot

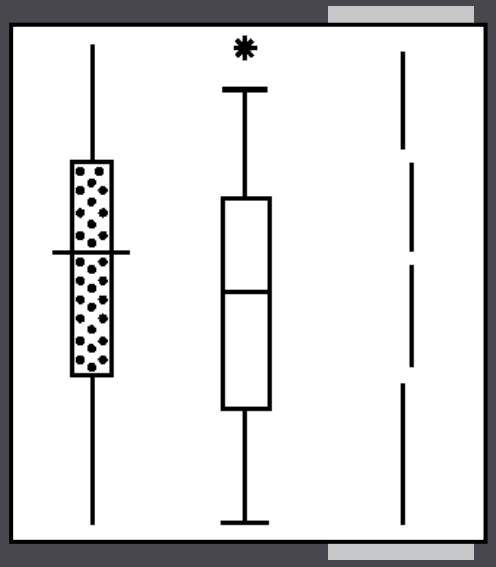
Exploratory Data Analysis.
John W. Tukey. Addison-Wesley, 1977.

SCI INSTITUTE

Boxplot Modifications

Visual Modifications

refinement for aesthetic purposes



inner-quartile plot

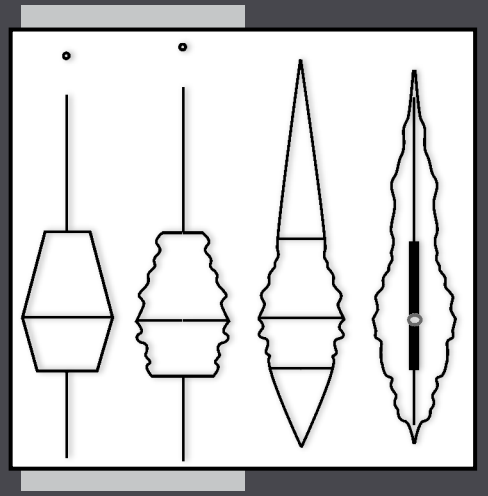
The Visual Display of Quantitative Information.
Edward Tufte. Graphics Press, 1983.

SCI INSTITUTE

Boxplot Modifications

Density Modifications

add indication of value prevalence



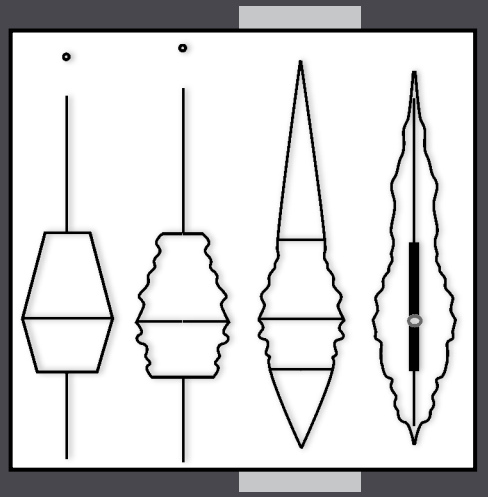
Opening the Box of a Boxplot.
Y. Benjamini. The American Statistician, 42(4), 1988.

SCI INSTITUTE

Boxplot Modifications

Density Modifications

add indication of value prevalence



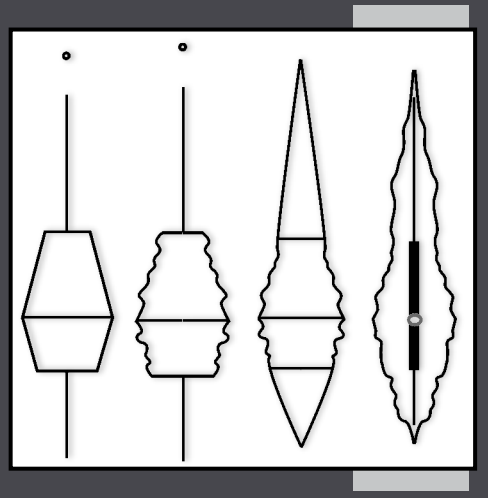
The Box-Percentile Plot.
W. Esty, J. Banfield. Journal of Statistics Software, 8(17), 2003.

SCI INSTITUTE

Boxplot Modifications

Density Modifications

add indication of value prevalence



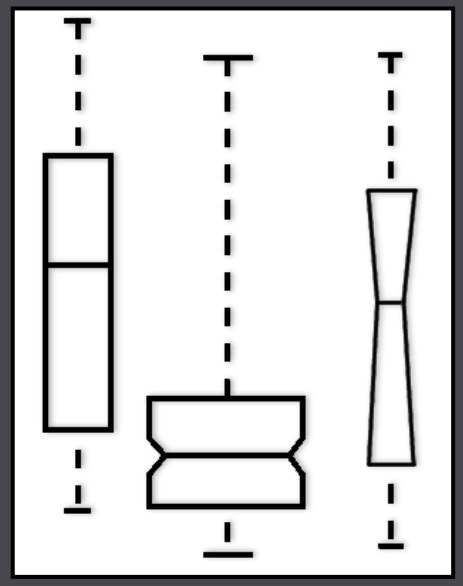
Violin Plots.
J. Hintze, R. Nelson. The American Statistician, 52(2), 1998.

SCI INSTITUTE

Boxplot Modifications

Data Characteristics

sample size, confidence levels



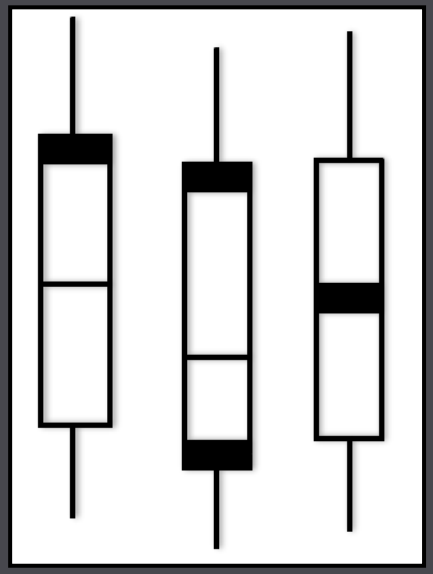
Variations of Box Plots.
R. McGill, J.W. Tukey, W.A Larsen. The American Statistician, 32(1), 1978.

SCI INSTITUTE

Boxplot Modifications

Additional Statistics

- *moments, modality*



Can the Box Plot Be Improved?
C. Choonpradub, D. McNeil. Songklanakarin J Sci Technol, 27(3), 2005,

SCI INSTITUTE

Boxplot Modifications

Summary Plot

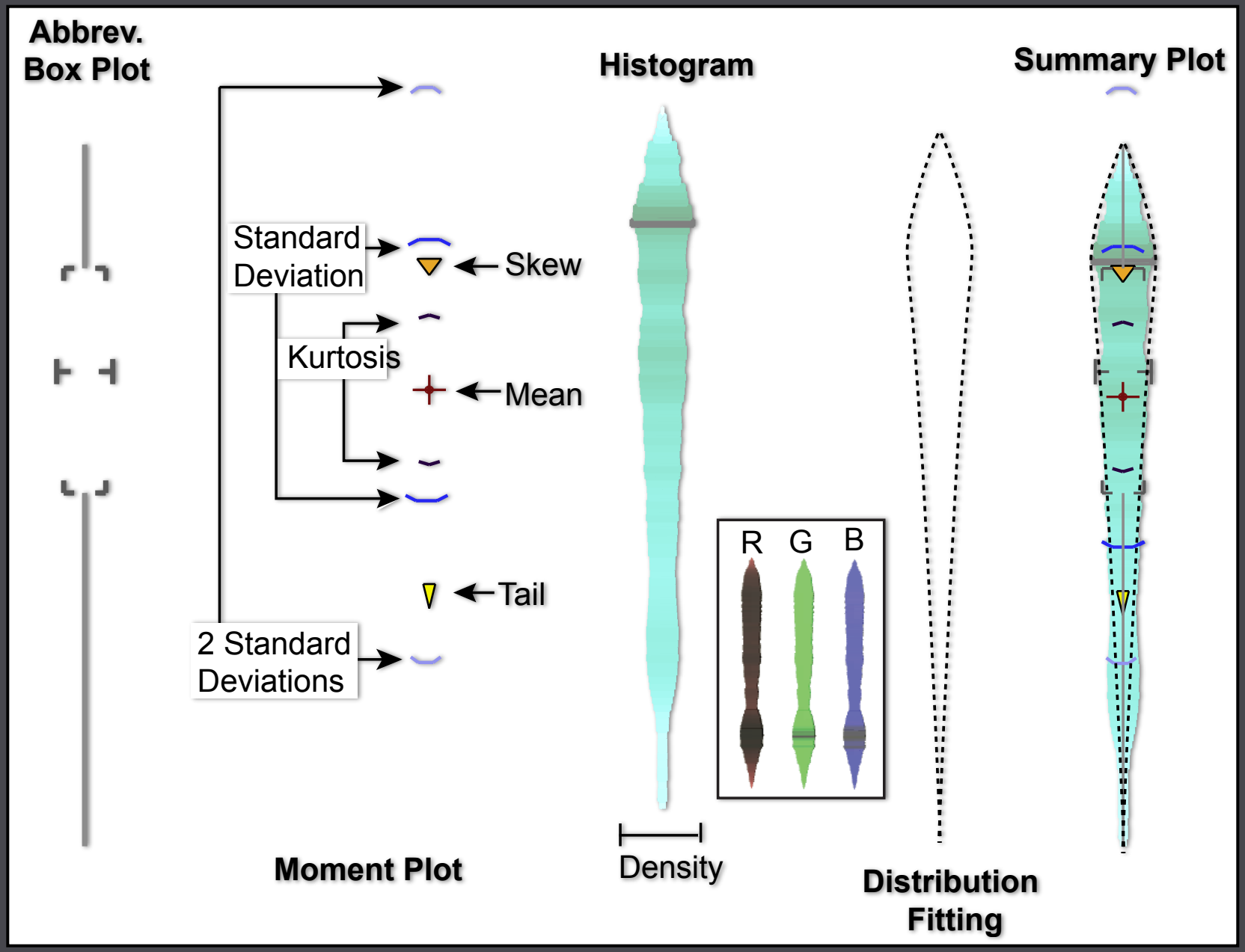
combine 4 plots into one

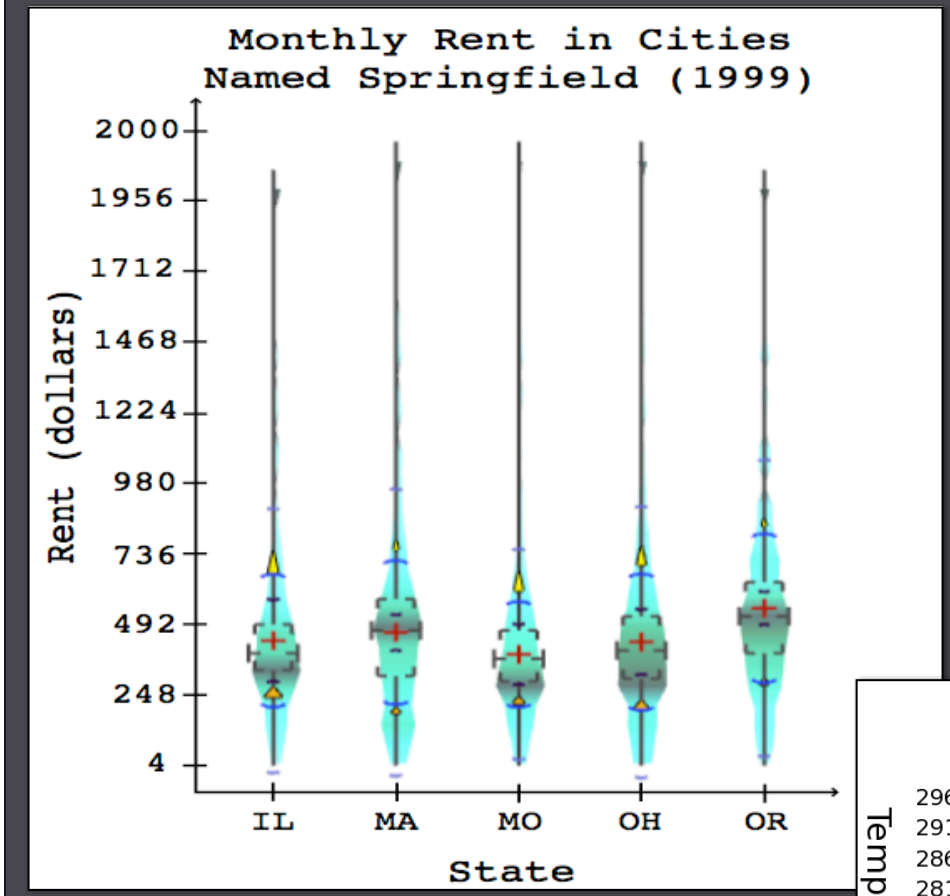
augment with more descriptive statistics

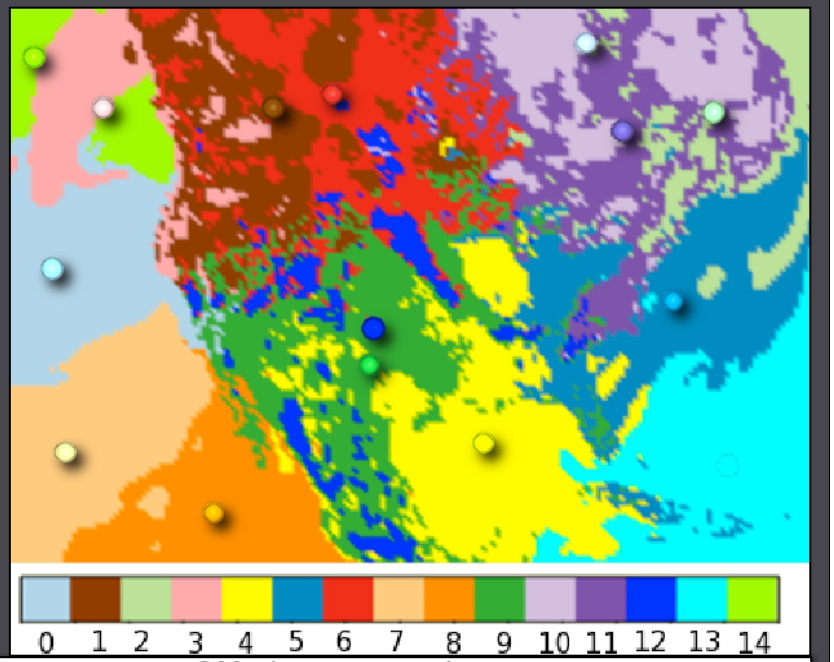
indicate quantity & location of uncertainty

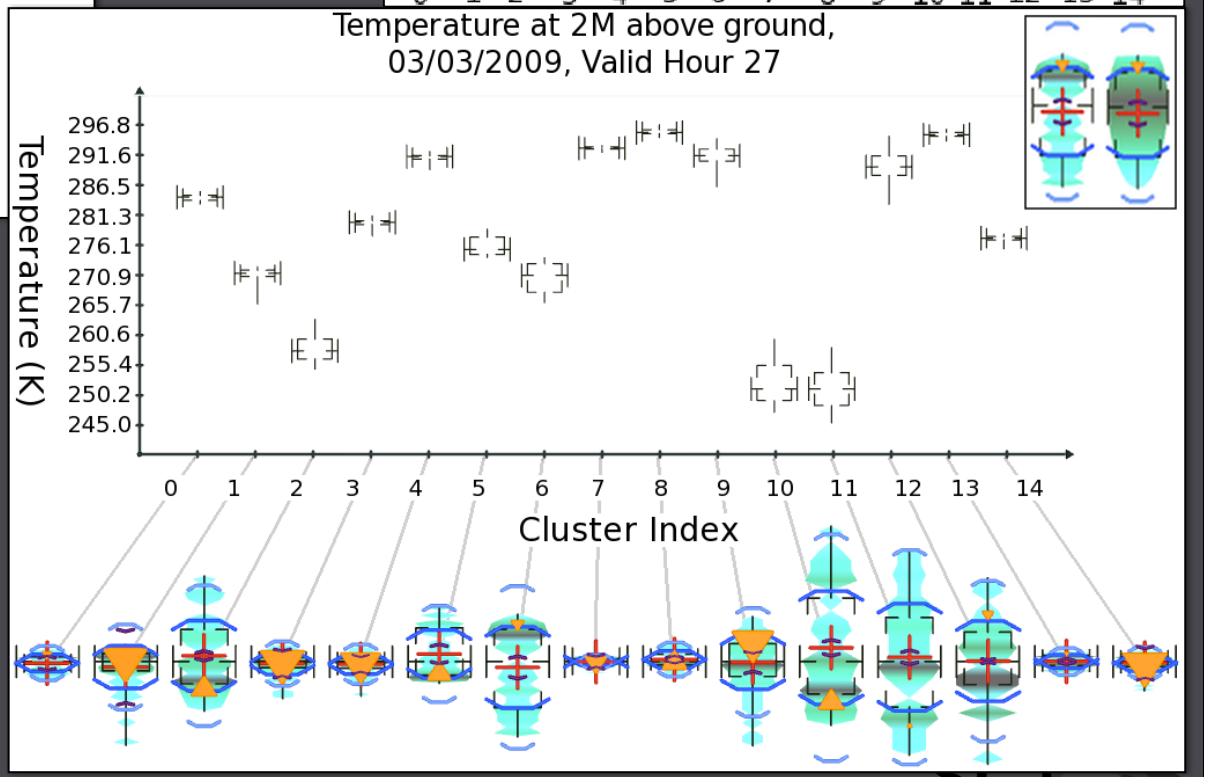
Visualizing Summary Statistics and Uncertainty.
K. Potter, J. Kniss, R. Riesenfeld, C.R. Johnson. CGF 29(3), 2010.

SCI INSTITUTE


Abbrev.
Box Plot
Histogram
Summary Plot
Standard
Deviation
Skew
Kurtosis
Mean
Tail
2 Standard
Deviations
R
G
B
Density
Moment Plot
Distribution
Fitting


Monthly Rent in Cities
Named Springfield (1999)
Rent (dollars)
2000
1956
1712
1468
1224
980
736
492
248
4
IL
MA
MO
OH
OR
State


0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

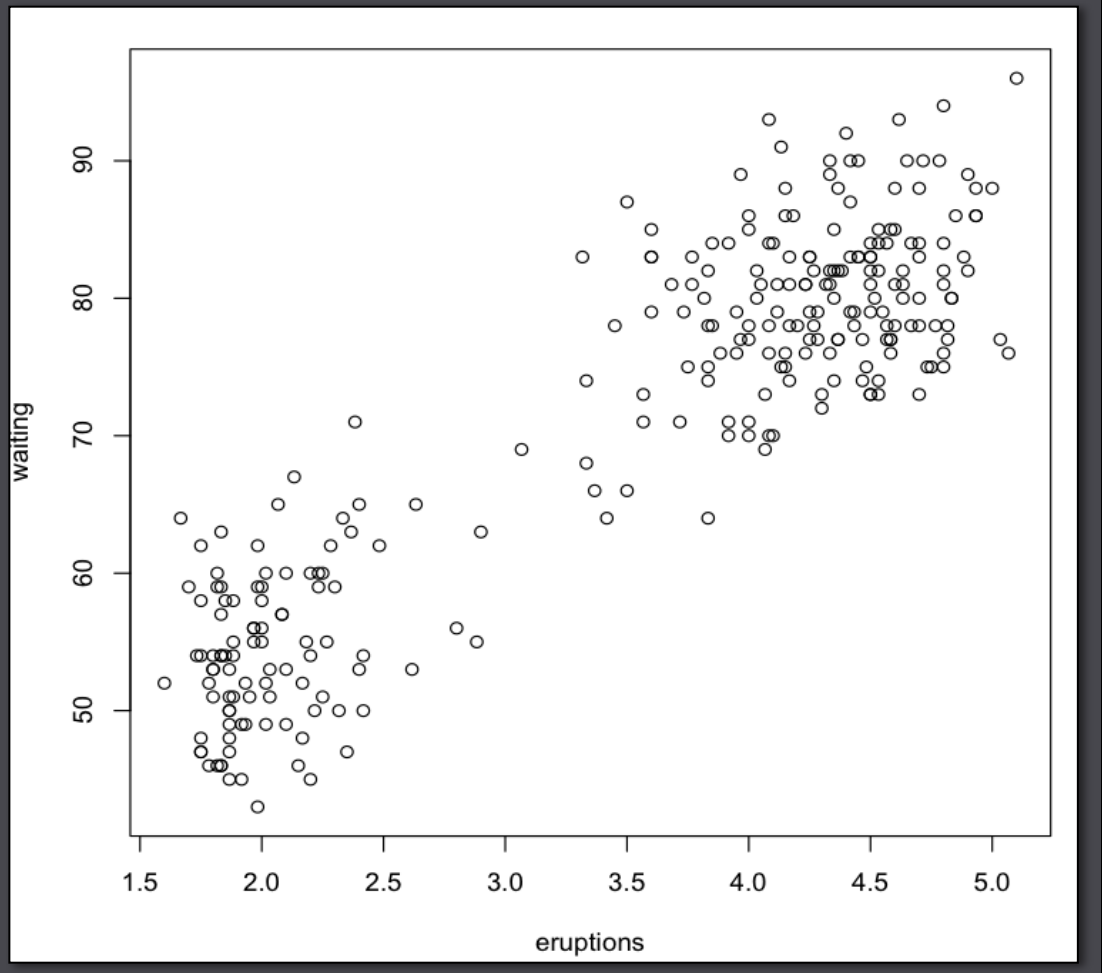

Temperature at 2M above ground,
03/03/2009, Valid Hour 27
Temperature (K)
296.8
291.6
286.5
281.3
276.1
270.9
265.7
260.6
255.4
250.2
245.0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Cluster Index

2D Data

- *Scatter plot of 2D position of samples*

- *William Playfair (1759–1823)*

pie charts
line graphs
bar charts



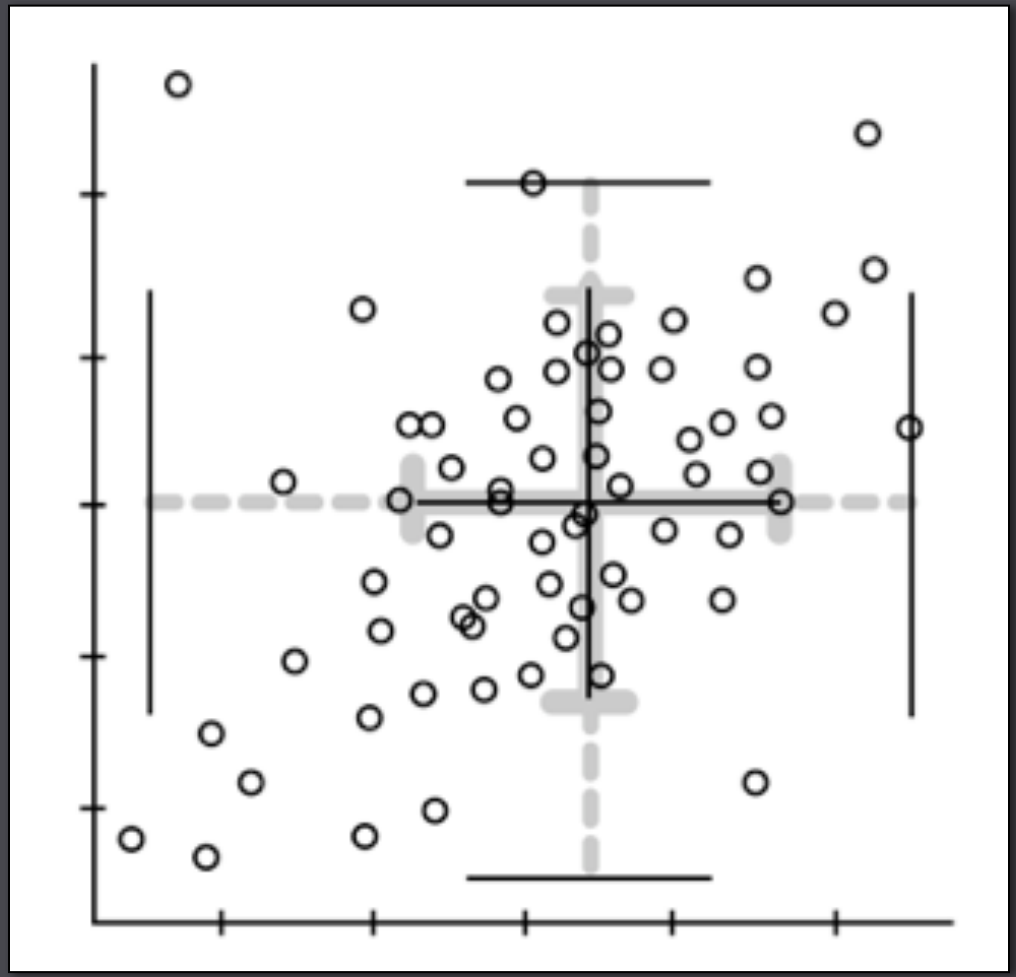
The Early Origins and Development of the Scatterplot
Michael Friendly and Daniel Denis
Journal of the History of the Behavioral Sciences, Vol. 41(2), 103–130 Spring 2005

SCI INSTITUTE

2D Box Plots

RangeFinder Plot

1D boxplot per axis



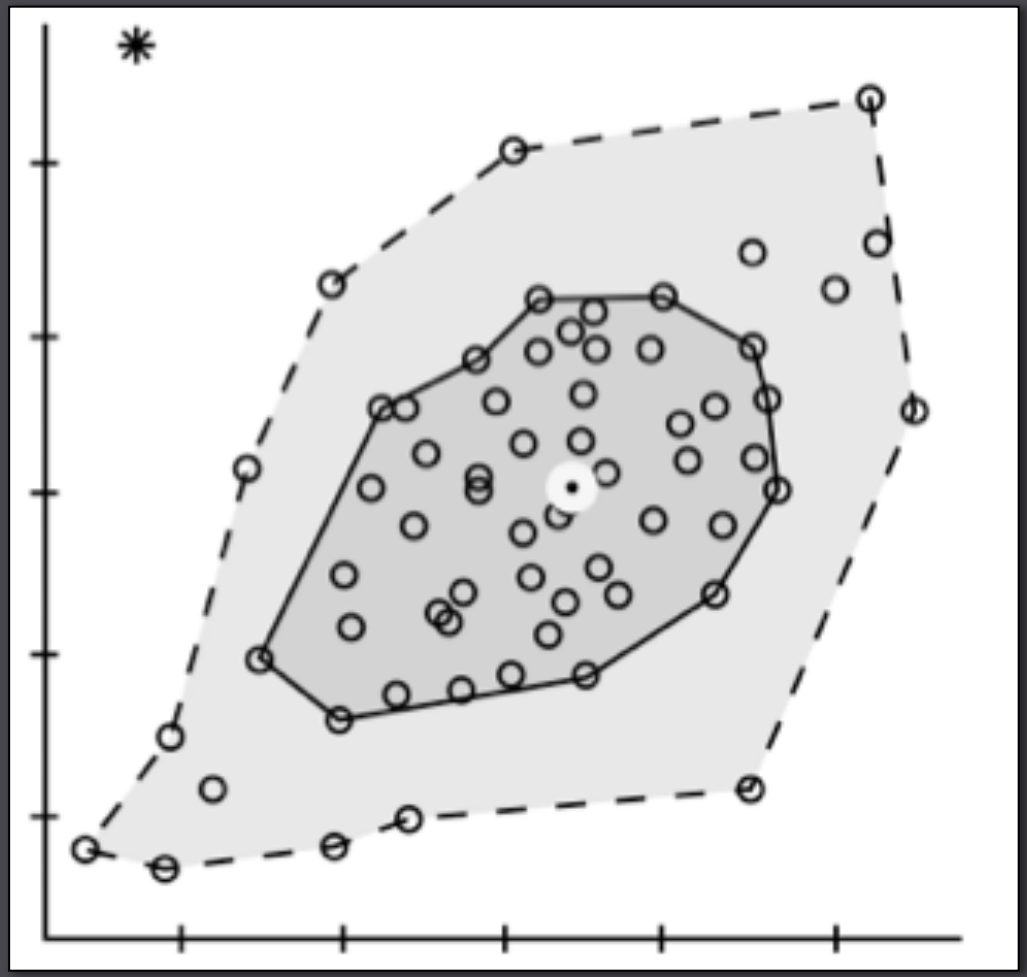
Rangefinder box plots.
S. Becketti, W. Gould, TAS 41(2), 1987.

SCI INSTITUTE

2D Box Plots

Two-Dimensional Boxplot

Robust line partition



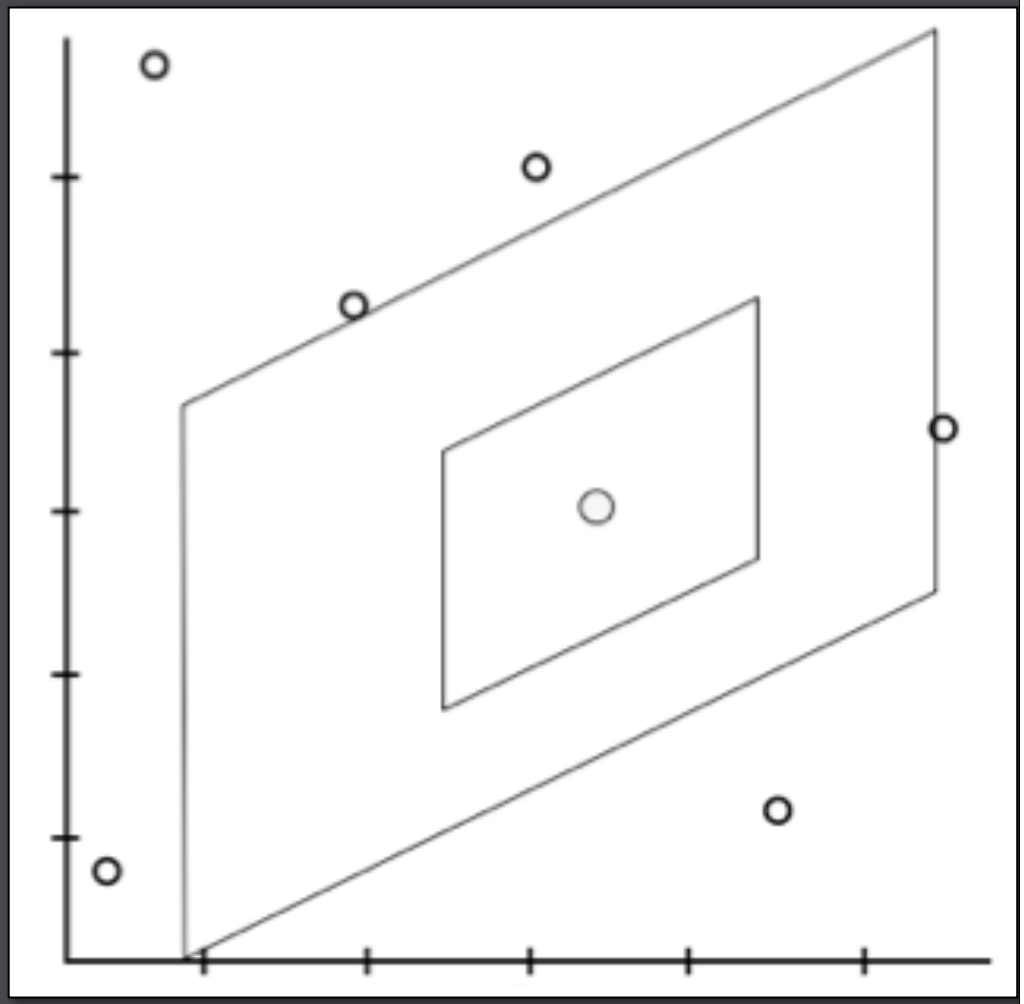
Two-dimensional box plot.
P. Tongkumchum, Songklanakarin J Sci Technol, 27(4),2005.

SCI INSTITUTE

2D Box Plots

Bagplot

Halfspace depth (spatial quartiles)



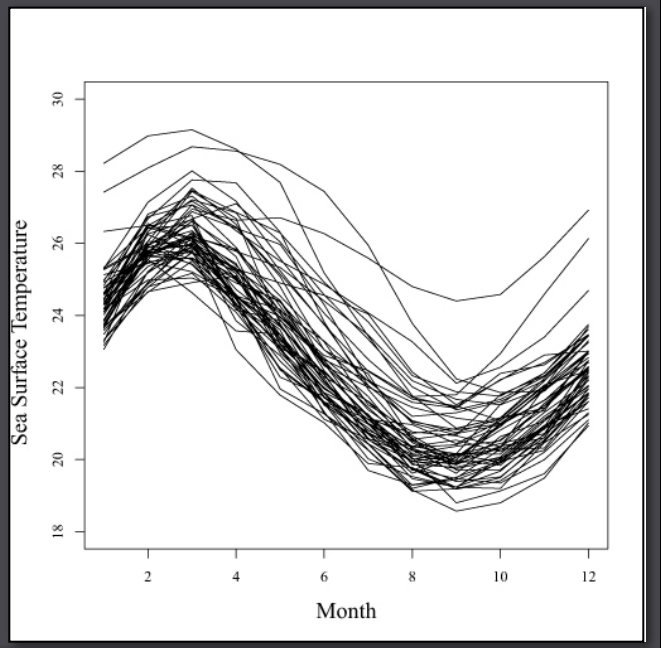
The Bagplot: A Bivariate Boxplot.
P.J. Rousseeuw, I. Ruts, J. Tukey. TAS, 53(4), 1999.

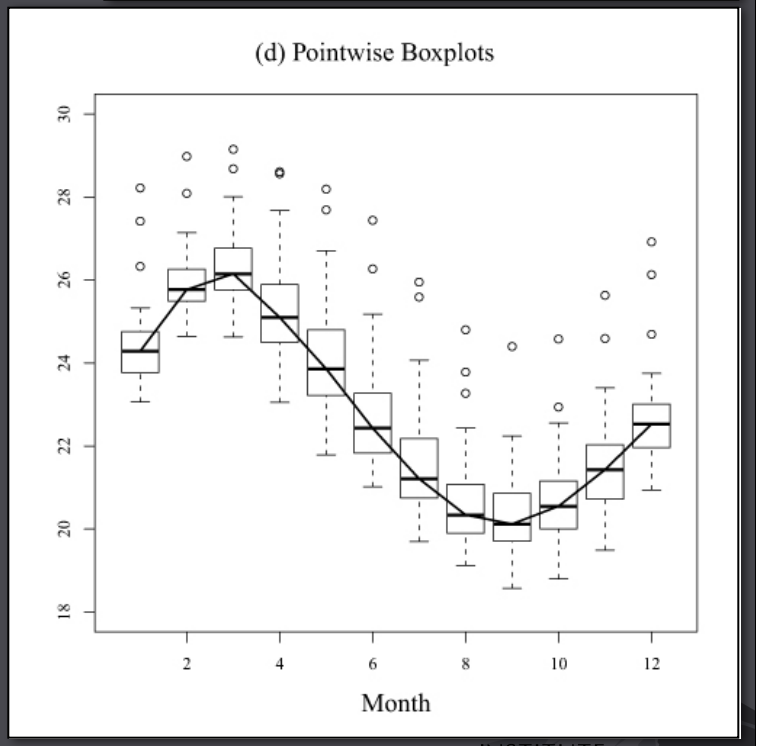
SCI INSTITUTE

Functional Box Plot

Boxplot statistics on 2D functions

defined on the function, rather than point-wise



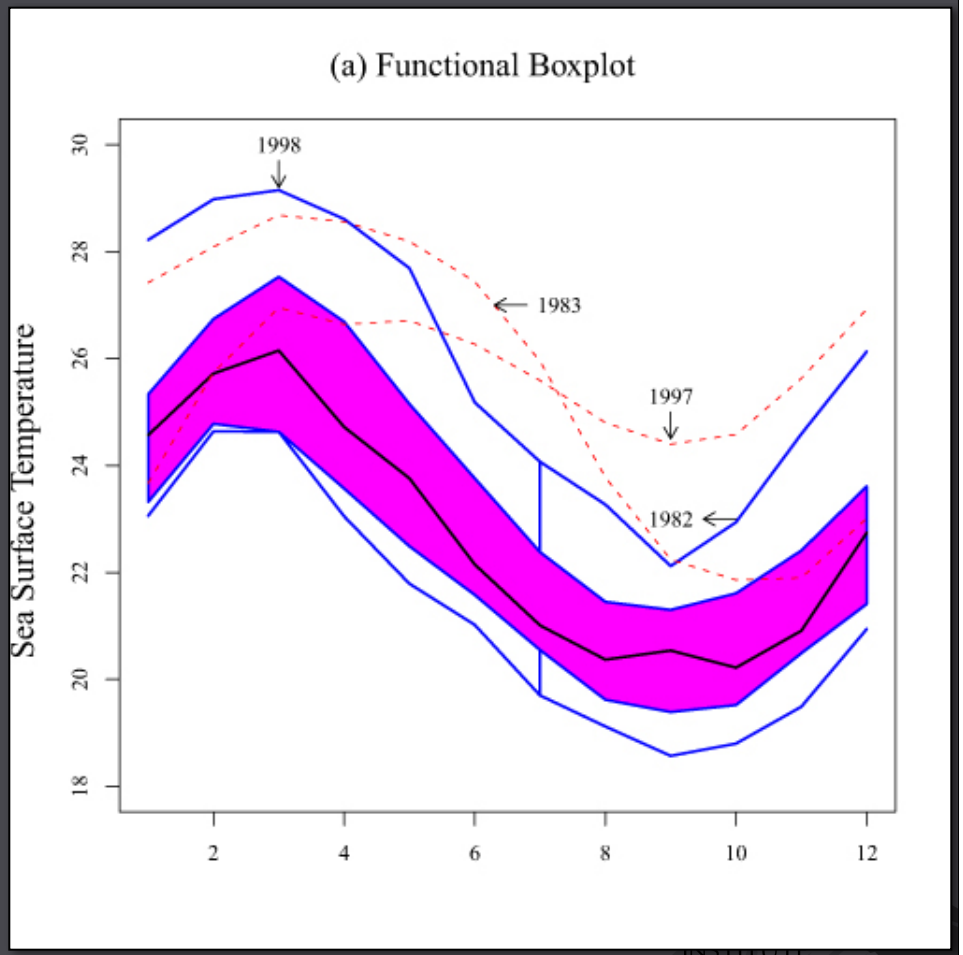


Functional Boxplots.
Ying Sun, Marc G. Genton.
J. of Comp. and Graphical Statistics 20:2, 2011, 316-334.

Functional Box Plot

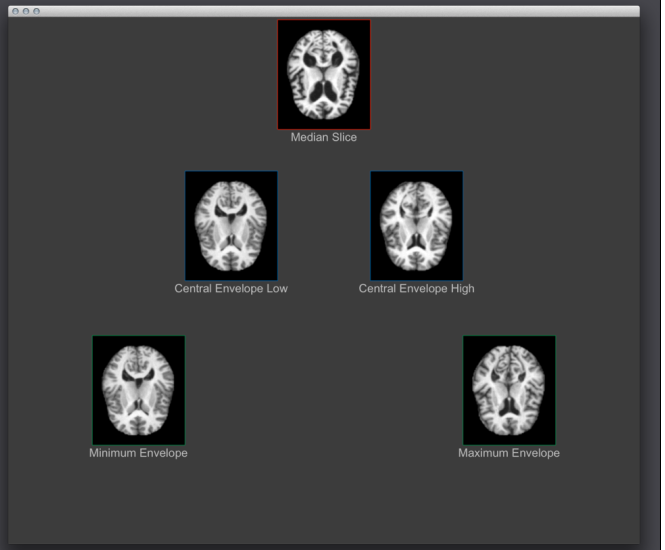
Band Depth

the amount of time a function lies within the set of functions



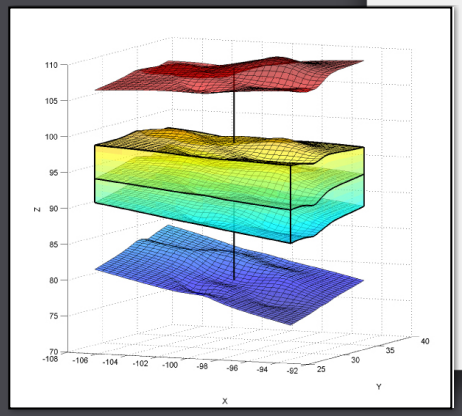
Functional Boxplots.
Ying Sun, Marc G. Genton.
J. of Comp. and Graphical Statistics 20:2, 2011, 316-334.

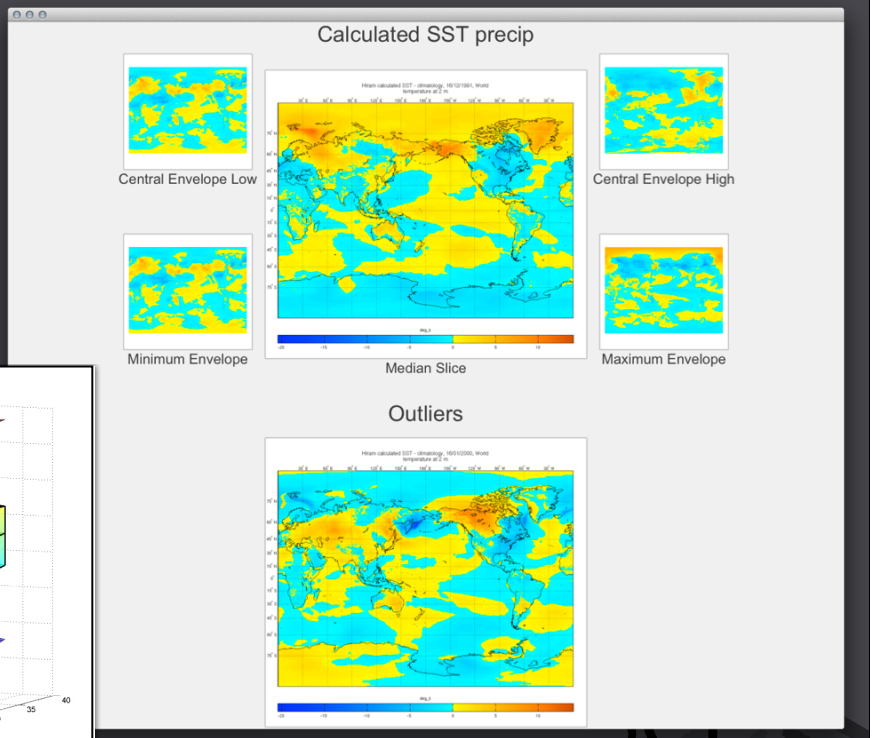
Surface Box Plots



Extension to 3D

- *images rather than curves*
- *volume-based band-depth*





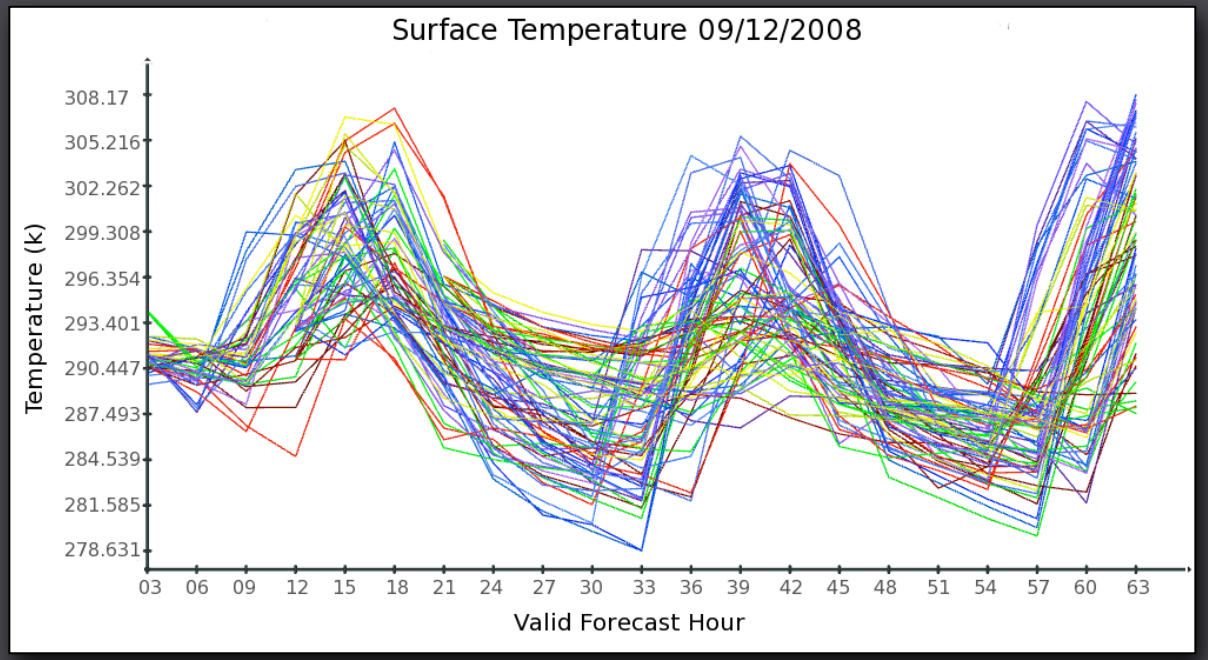
Visualization

incorporate uncertainty information into data display

Challenges

Increased complexity

- *Visual clutter*
- *Data concealment*
- *Confusion*



SCI INSTITUTE

Information-Seeking Mantra

"Overview first, then zoom and filter,
and finally, details on demand."

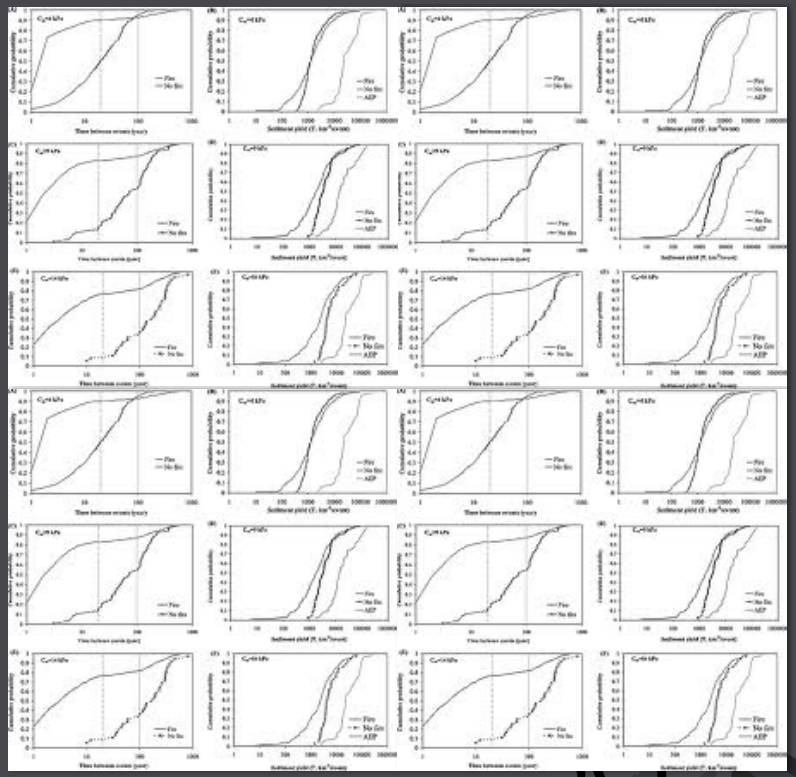
-Ben Shneiderman

Overview → Summary → Details

Overview vs Summary

Overview

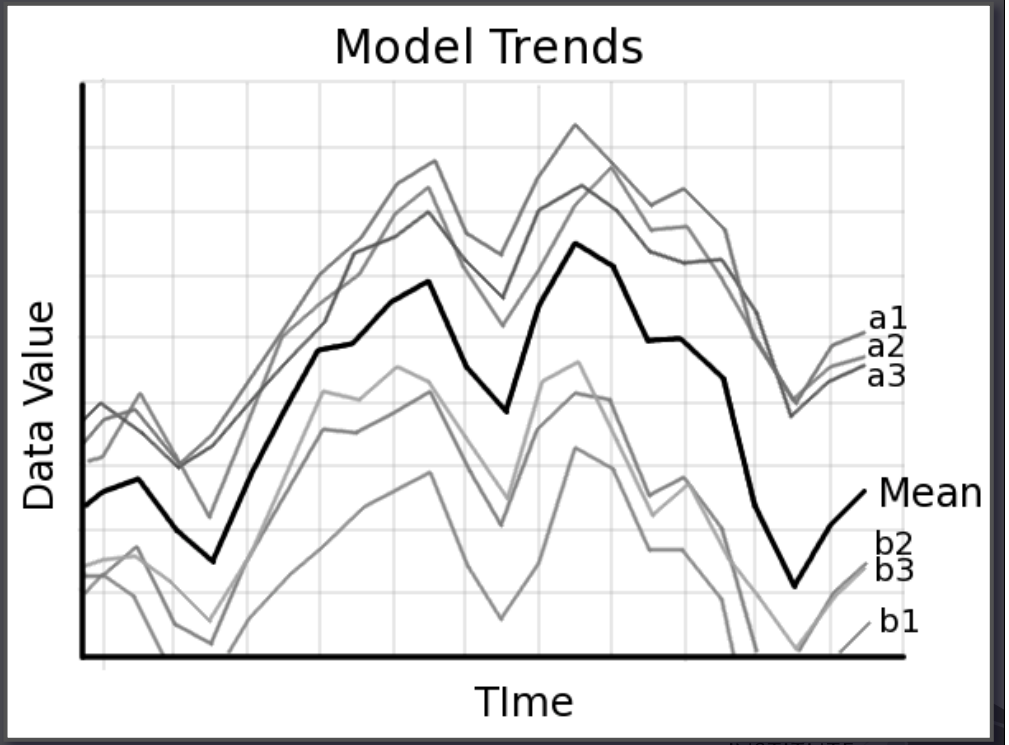
- *Show all data at once through many charts*
- *Manual search for patterns*
- *Finite screen resolution*



Overview vs Summary

Summary

- *Aggregate data along some dimension*
- *Automation summaries*
- *Have an idea of the questions*



Aggregation for Visualization

Reduce Data Using Summaries

- *mimic human visual system*
- *done implicitly*

phenomena modeling
floating pt quantization
limited # of pixels

Aggregation for Visualization

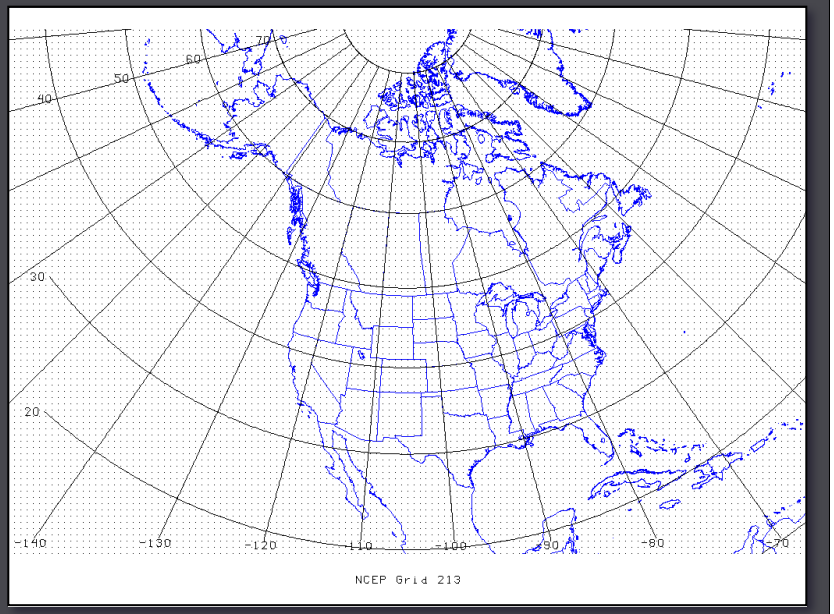
But How?

- *In what dimension do we summarize?*
- *Is mean/standard deviation appropriate?*
- *Do we need multiple summarizations?*

Ex: Weather Forecast Data

Weather Forecasting

Short-Range Ensemble Forecasts (SREF)



NOAA / NCEP

- *Domain across North America*
- *Forecast weather variables out to ~3.5 days*

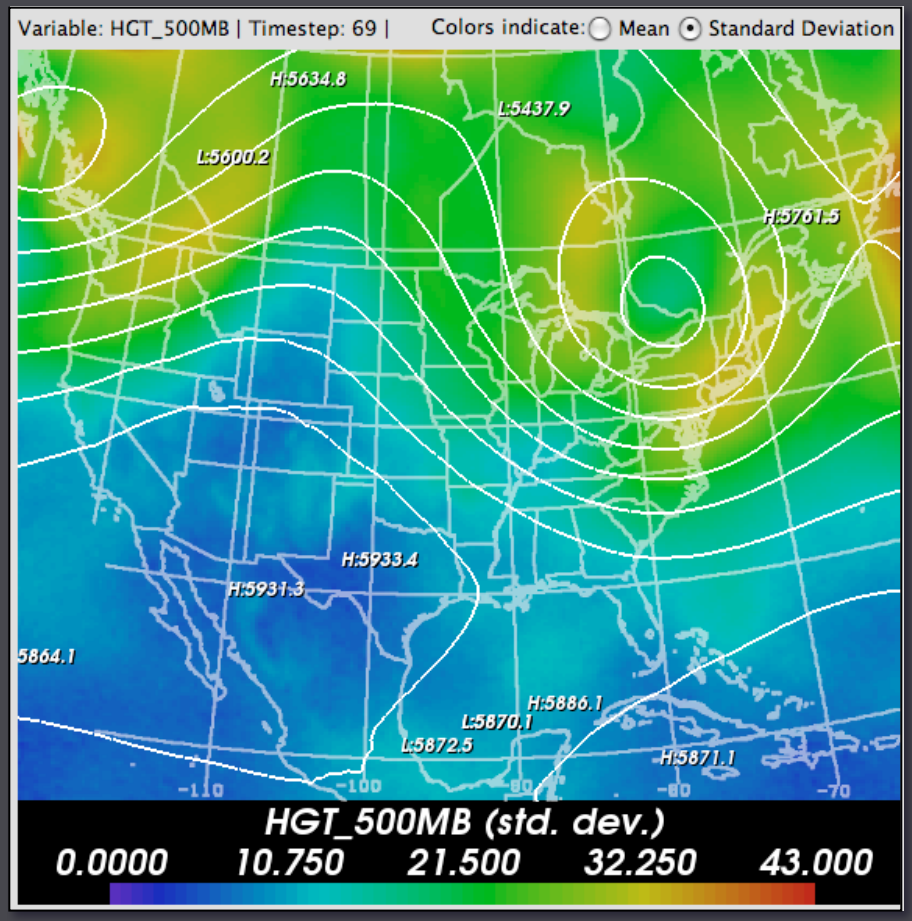
Goal

- *Public notification, warnings, aviation*

SCI INSTITUTE

Data

- *Study the variations of output between models*
- *Vary input parameters, initial conditions*

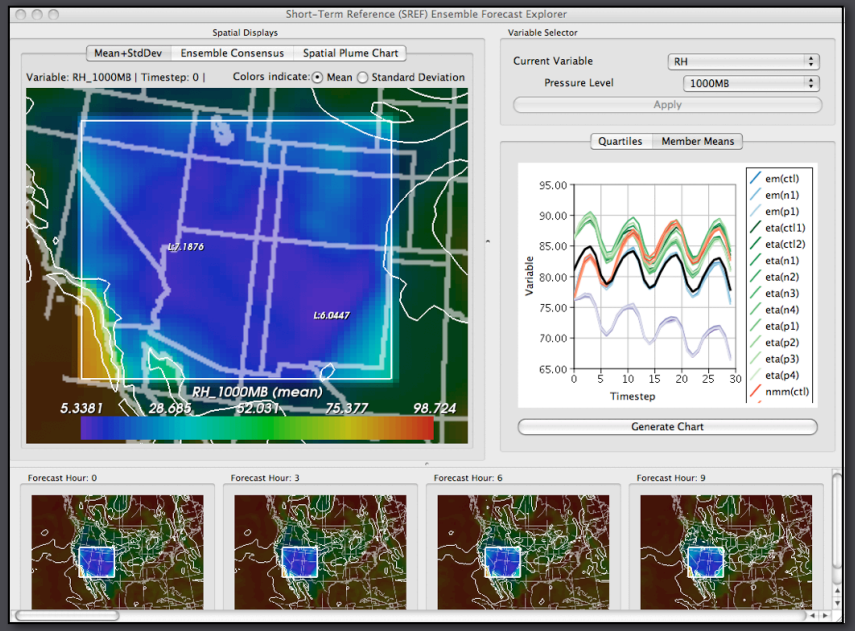


SCI INSTITUTE

Ex: Forecast Data

Questions

- *what is the the weather going to be tomorrow?*
- *what information is available from this data?*



Ensemble-Vis

User-driven, component-based framework

- *Combine multiple summaries*
- *explore range of possible forecasts*
- *reveal probabilities of outcomes*
- *interrogate the ensemble*

Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data.
K. Potter, A. Wilson, P.T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson
In IEEE Workshop on Knowledge Discovery from Climate Data: Prediction, Extremes, and Impacts, pp. 233-240, 2009.

SCI INSTITUTE

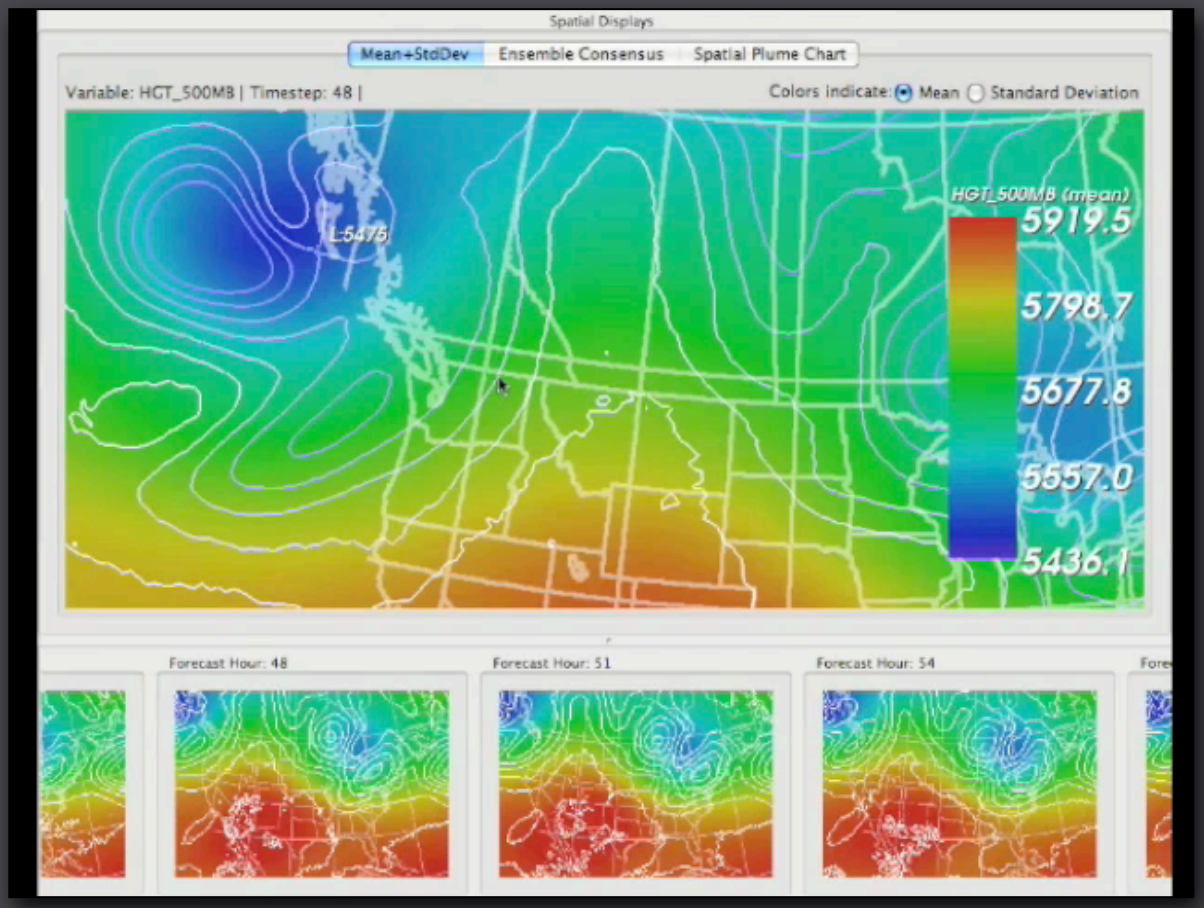
Global Summaries

Questions

- *What is the average model temperature at a given time step?*
- *Where do the models vary?*

Global Summaries

- *Aggregate over models at each grid point*
- *Colormap & contour*
- *Single timestep*



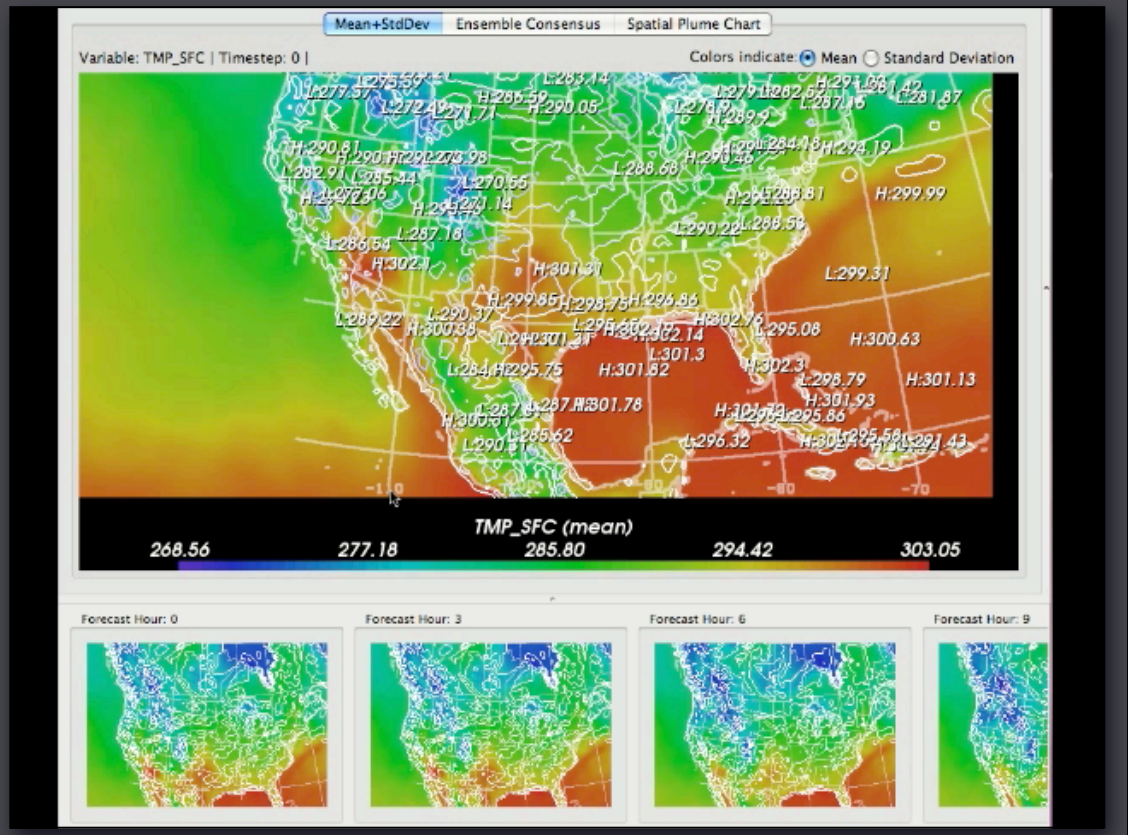
Time Summaries

Questions

- *How does the temperature change across time?*
- *What time step am I most interested in?*

Time Summaries

- *Small multiples*
- *User interaction to scroll through time*
- *Select step of interest, reflect to overview*



SCI INSTITUTE

Contour Summaries

Questions

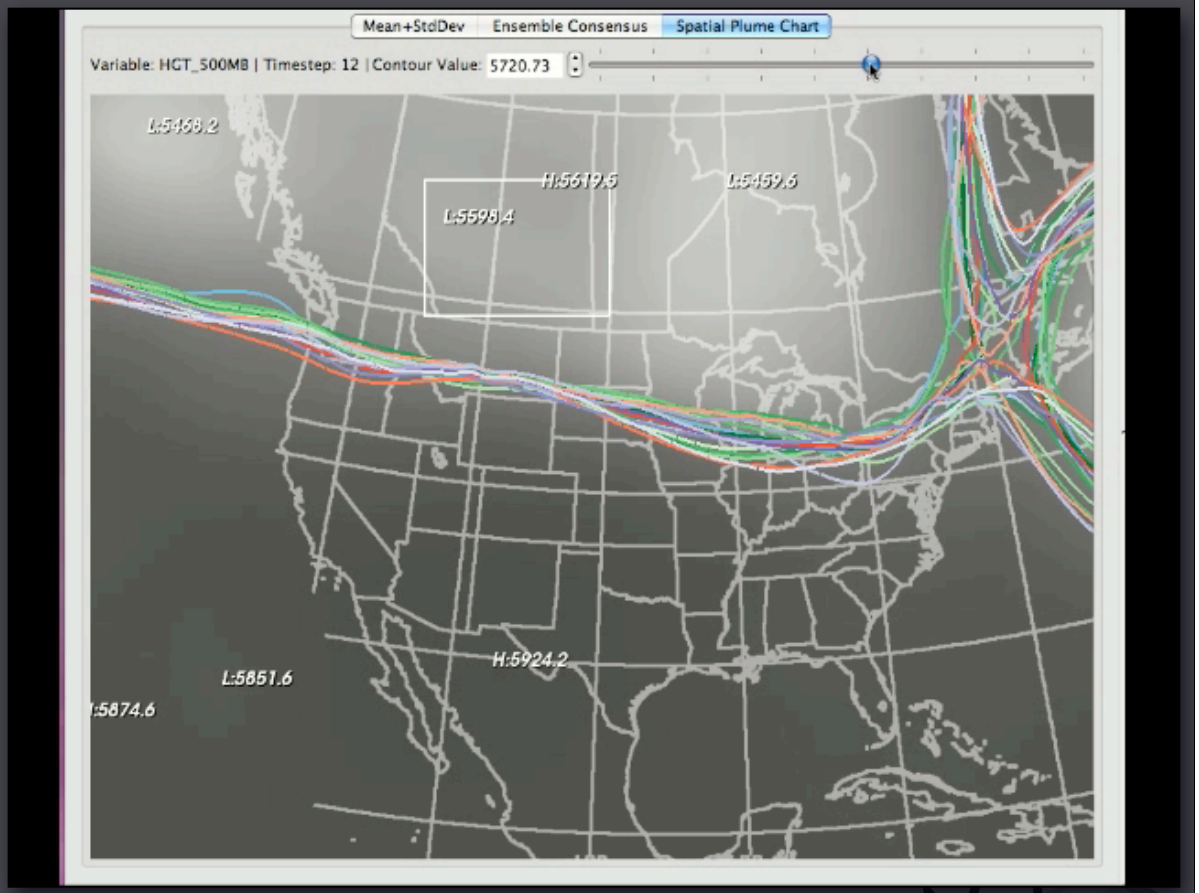
- *Where does a particular value exist for each model?*
- *Where does that value move across time?*

SCI INSTITUTE

Contour Summaries

- *Isocontour of value across spatial domain for each model*

- *Model bundle shows variation, outliers, divergence*



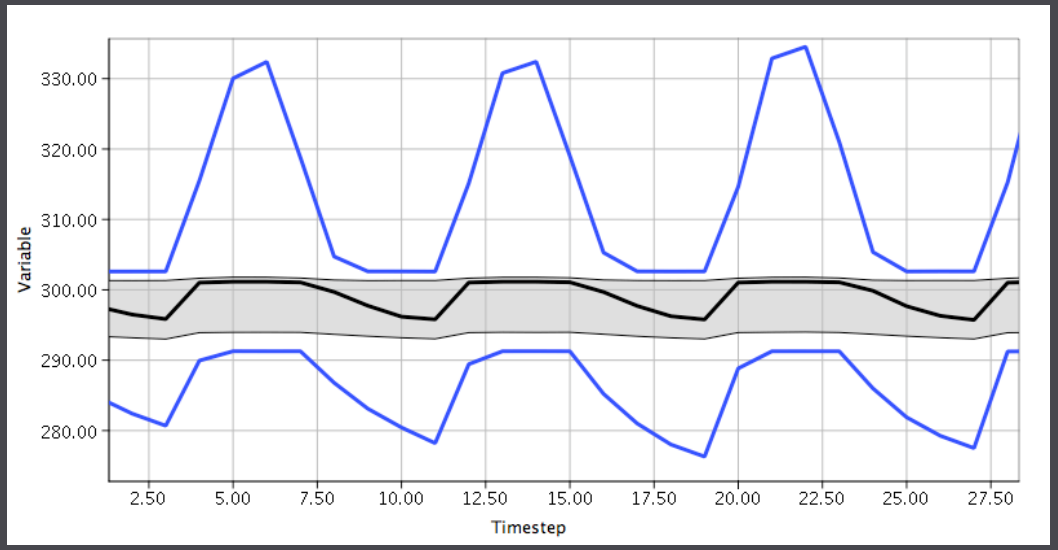
Spatial Summaries

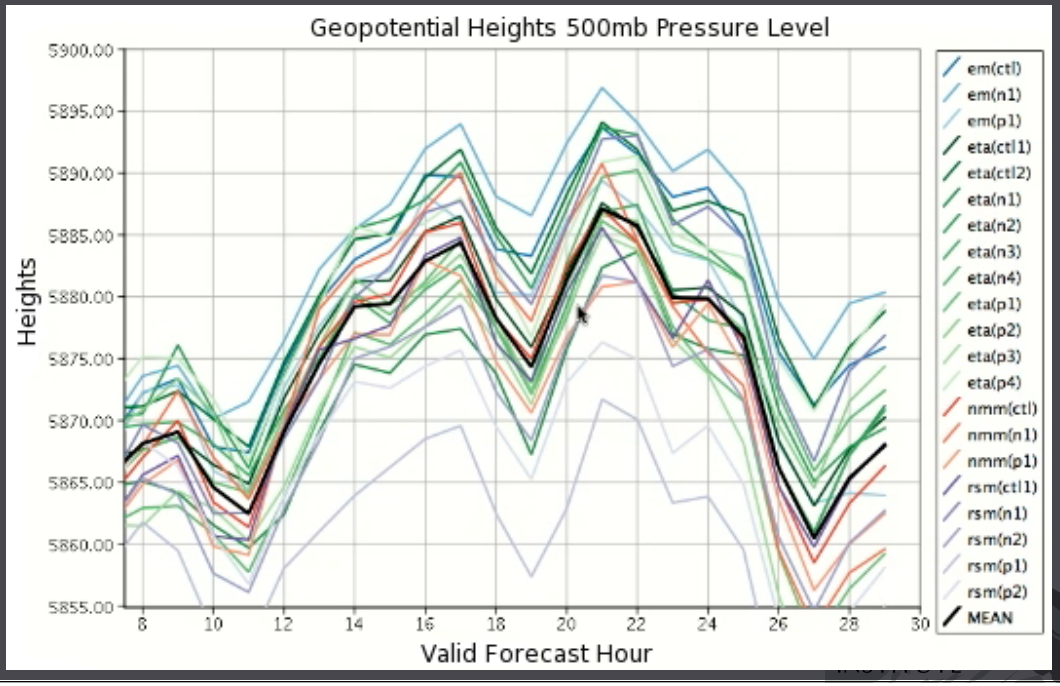
Questions

- *What is the trend of the models over a region?*
- *What is the average trend?*

Spatial Summaries

- *summarize over user selected region*
- *average per model*
- *overall average/ boxplot statistics*





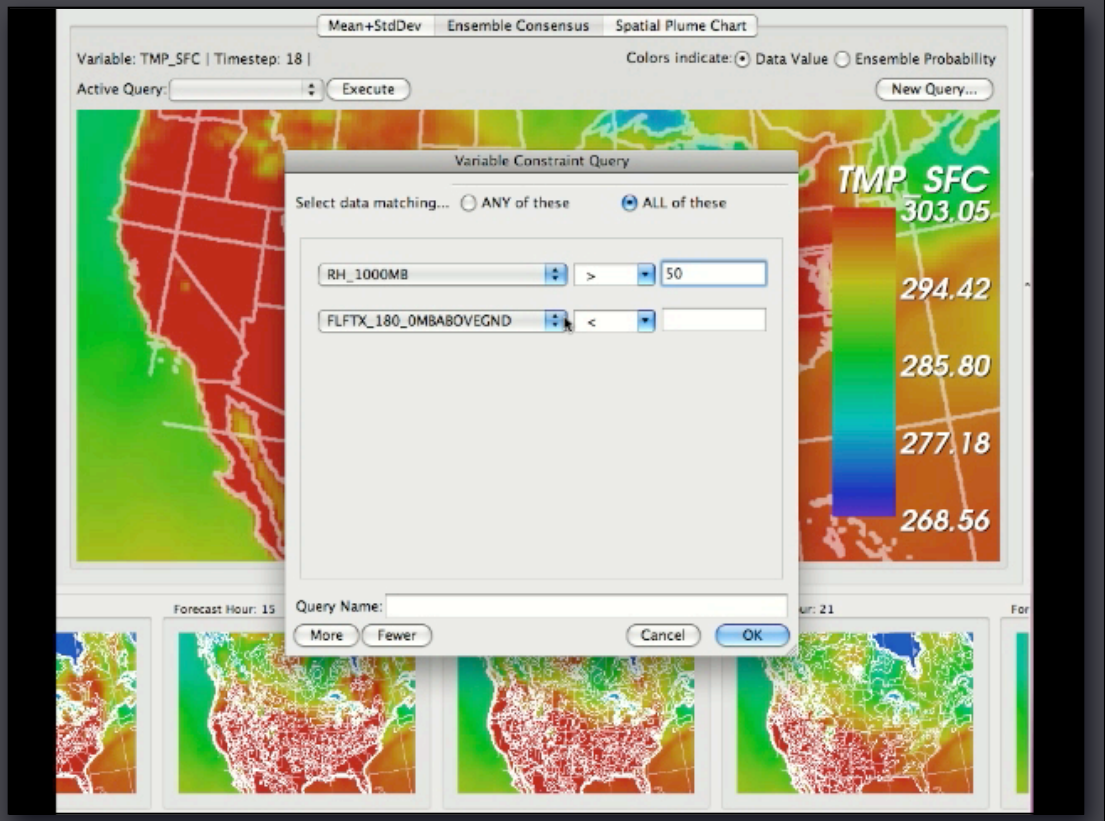
Query Summaries

Questions

- *Where does the data express particular characteristics?*
- *What fraction of the data expresses it?*

Query Summaries

- *SQL type queries to filter the data*
- *contours of ensemble fraction that predicts the condition*



Are we asking the right questions?

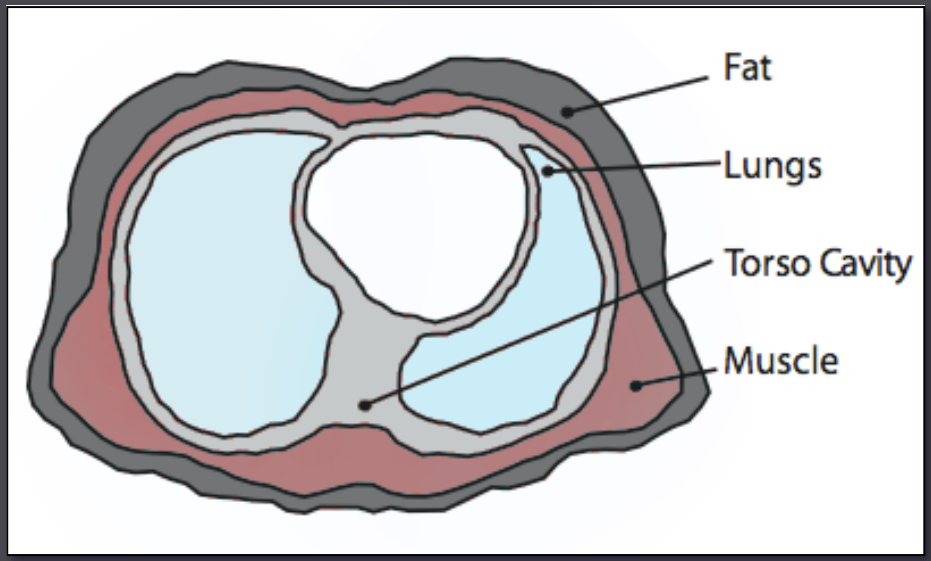
Ex: Torso Potentials

Electrical Conductivity of the Heart

Goal

- *Simulate how signals from the heart propagate across the torso*

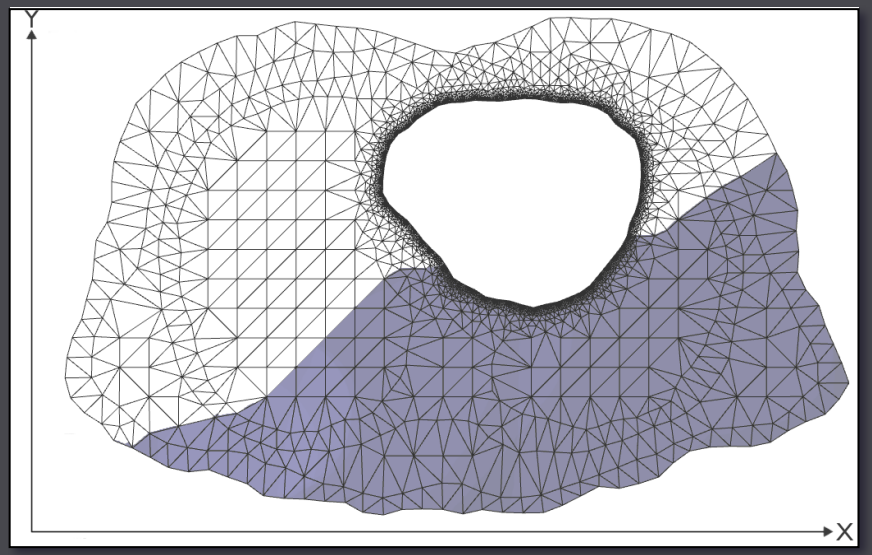
- *Distinguish normal changes (breathing, movement) from abnormal heart function*



SCI INSTITUTE

Data

- *Study the impact of variation on input conductivity*
- *Vary lung conductivity uniformly +/- 50% from the reference*
- *10,000 realizations of potentials*

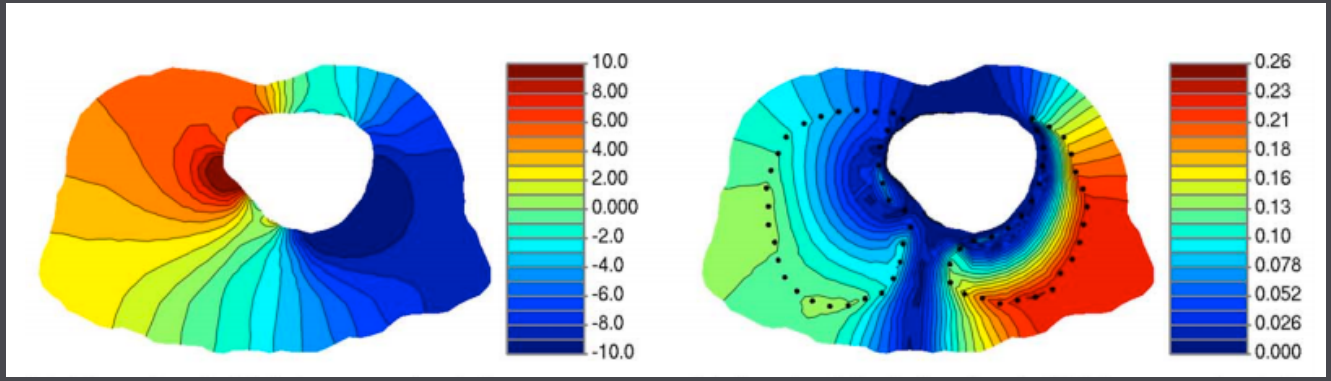


SCI INSTITUTE

Ex: Torso Data

Questions

- *what is the average potential across the domain?*
- *what is the variation of potential across the domain?*



Mean

Standard Deviation

Global Summary

Questions

- *what is the average potential across the domain?*
- *what is the variation of potential across the domain?*

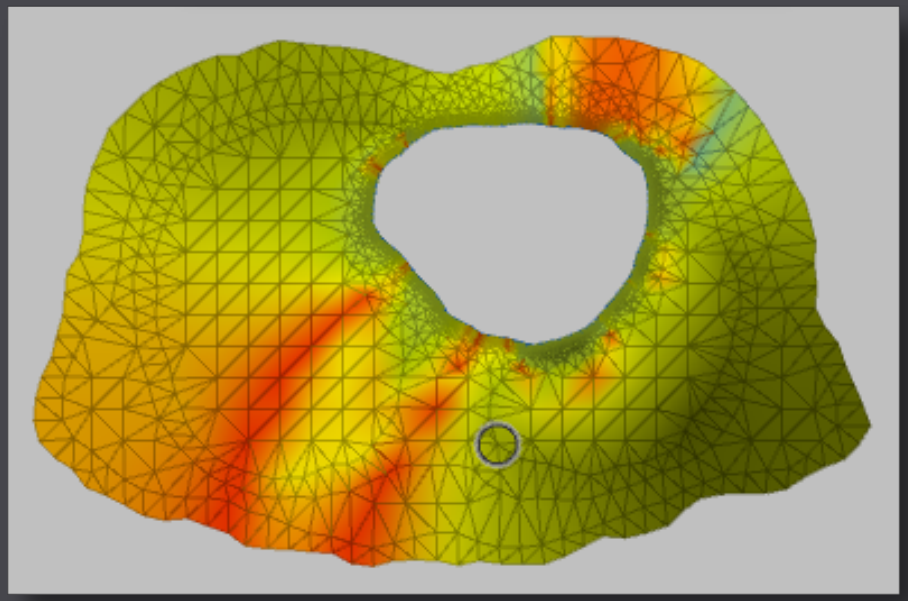
But is it meaningful?

- *are mean & standard deviation appropriate statistics?*
- *can we look at the data without individually inspecting each grid point?*

Torso Data

New Questions

- *what do the PDFs look like across the domain?*
- *where are the PDFs similar or different?*



ProbVis

Visualization & Exploration for Distributions

- *show differences between PDFS*
- *summarize all data in a single view*

http://www.sci.utah.edu/~kpotter/software/ProbVis/

SCI INSTITUTE

Method

Comparison Between Two PDFS

- *compare all data points to a single PDF*
- *get a single metric for each data point*
- *compare between points*

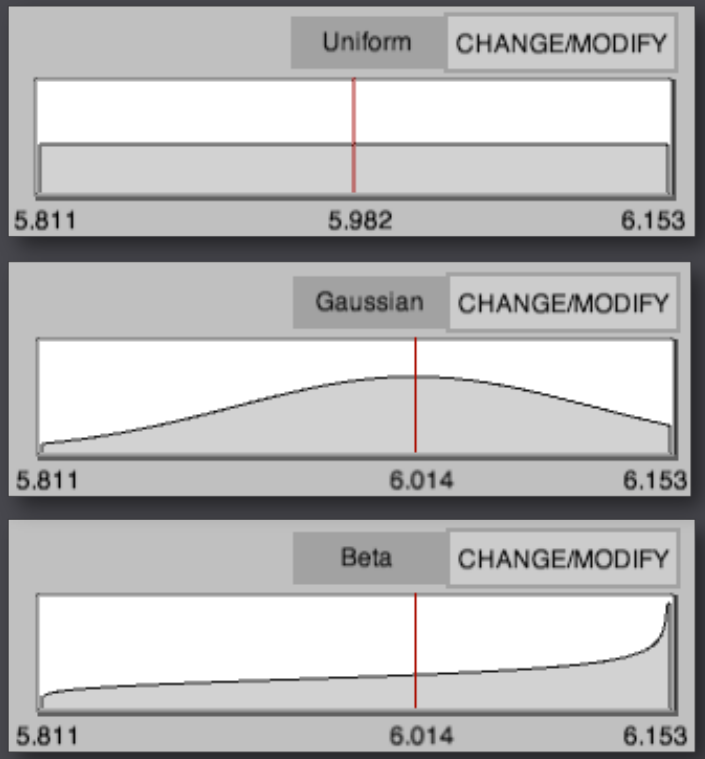
Method

Library of Canonical Distributions

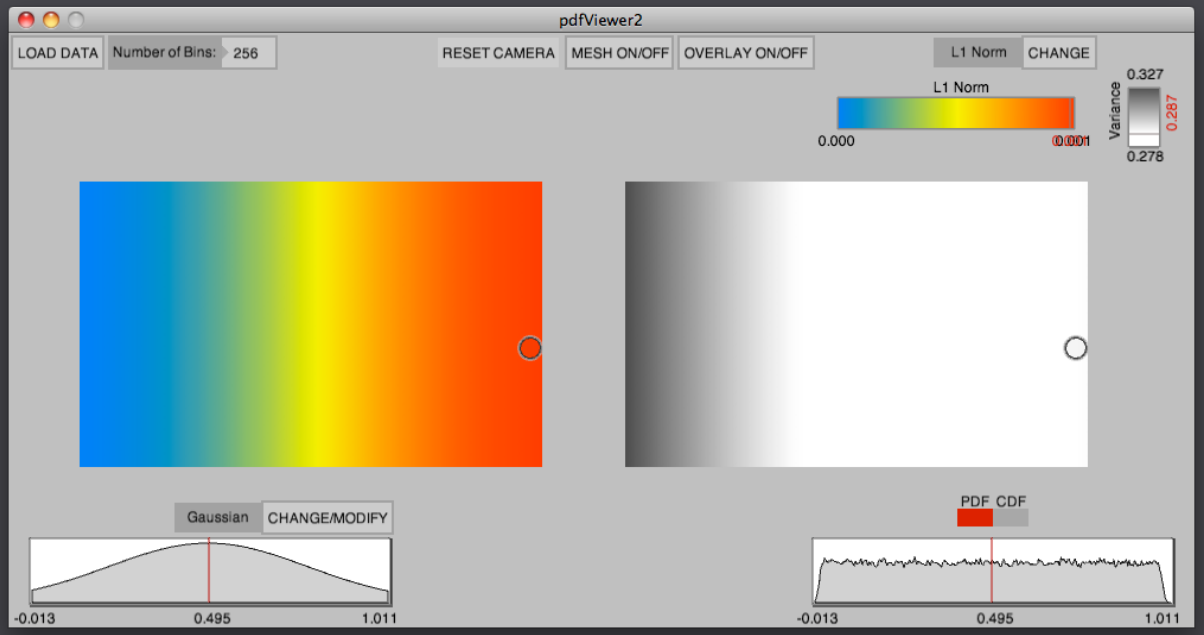
uniform

gaussian

beta



SCI INSTITUTE

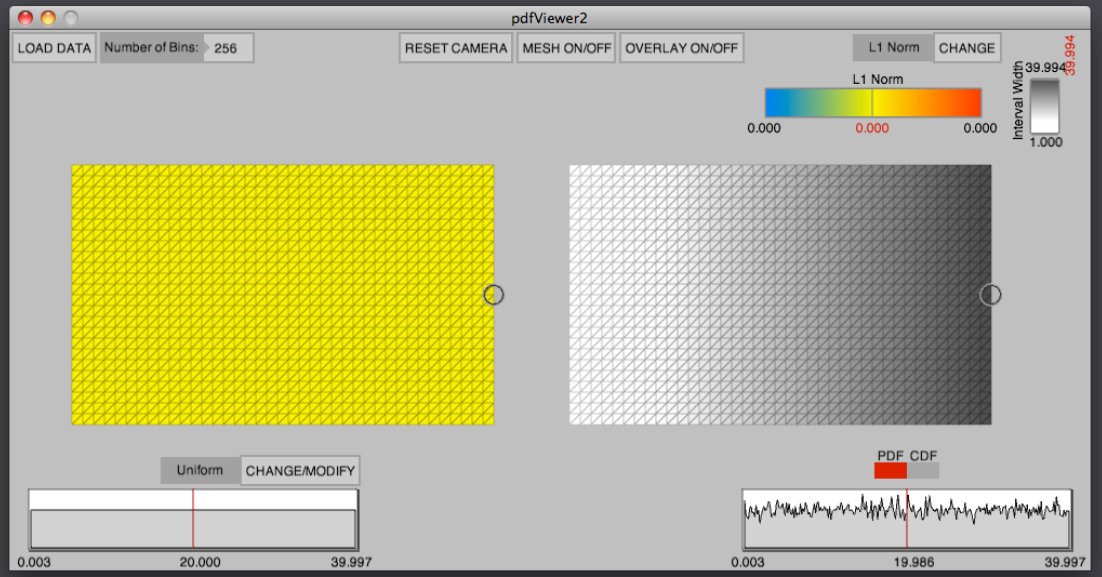


Method

Define a Difference Measure

shape: L1 Norm or Hellinger Distance

SCI INSTITUTE



Method

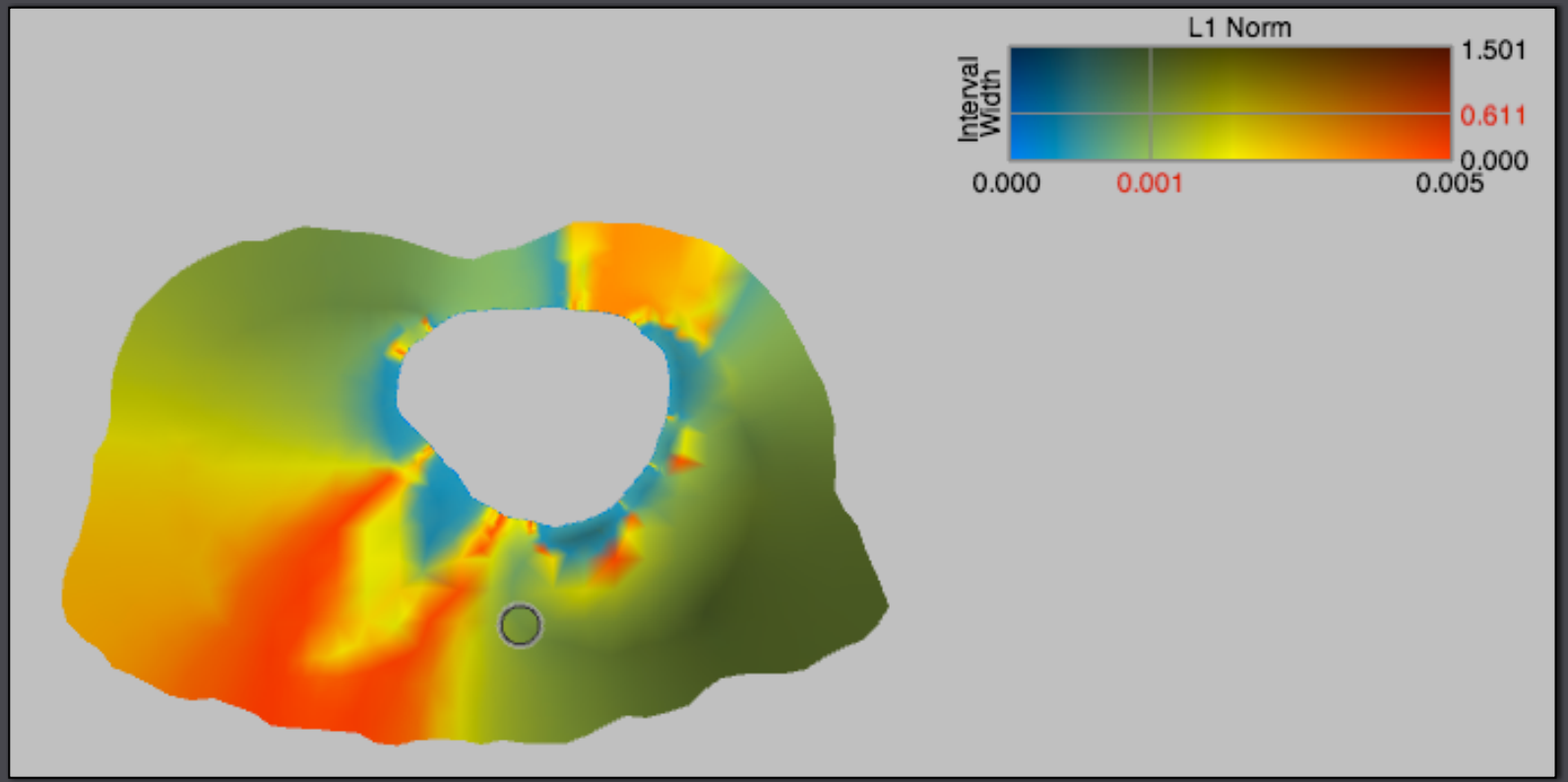
Define a Difference Measure

interval: (range of sample values)

SCI INSTITUTE

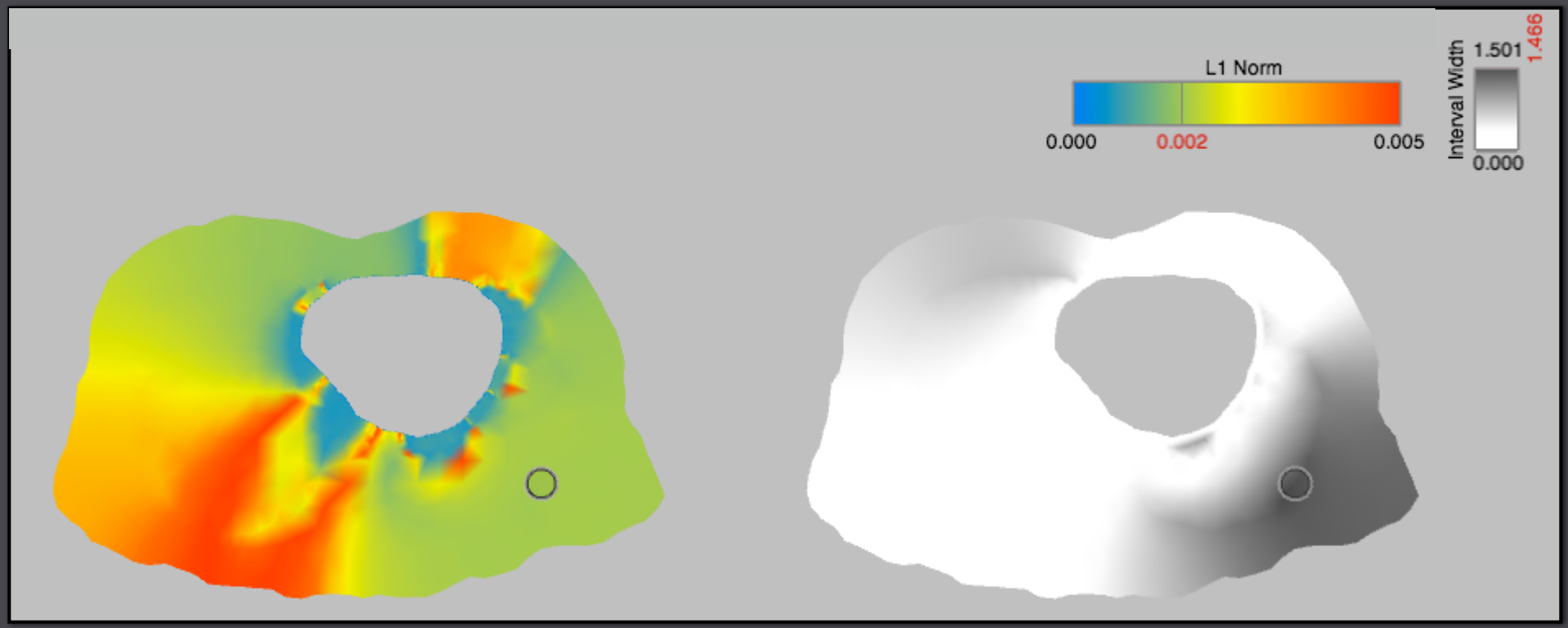
Visualization

Colormap distance measure



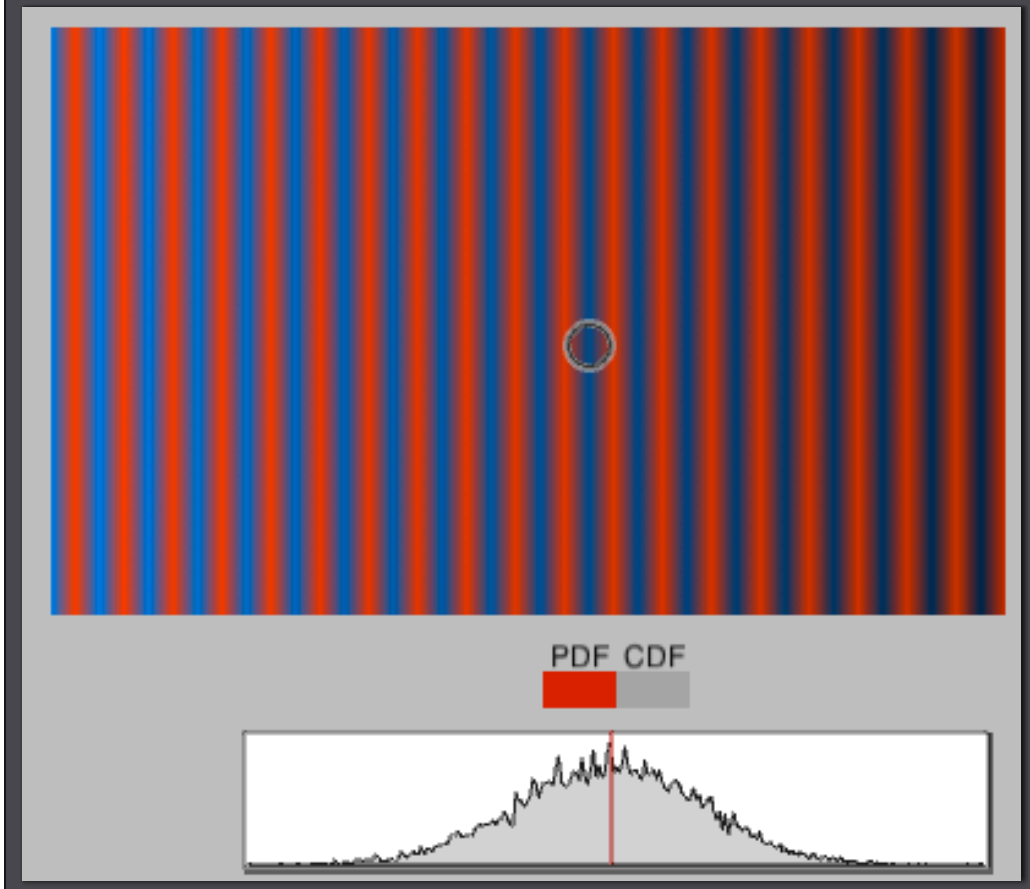
Visualization

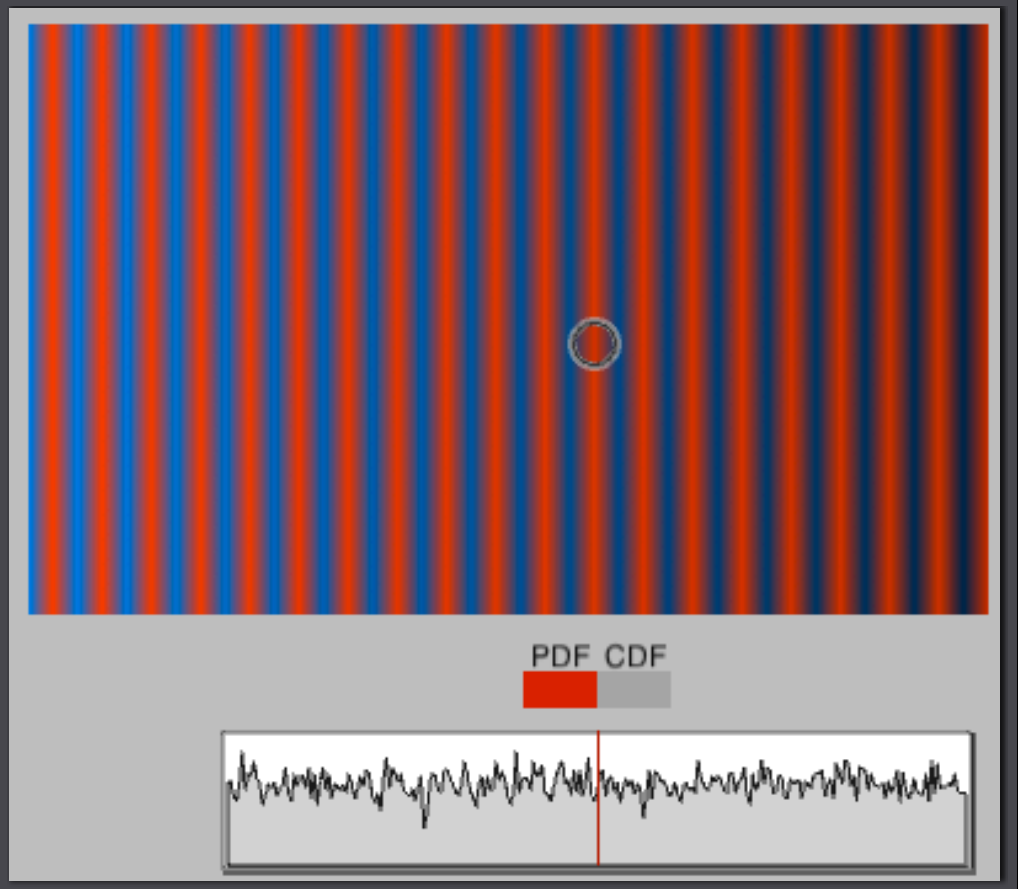
Colormap distance measure

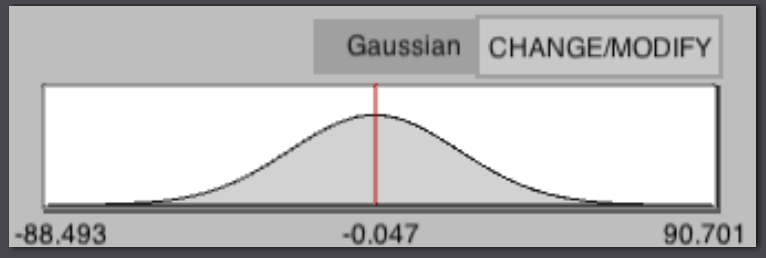


SCI INSTITUTE

Alternating Example

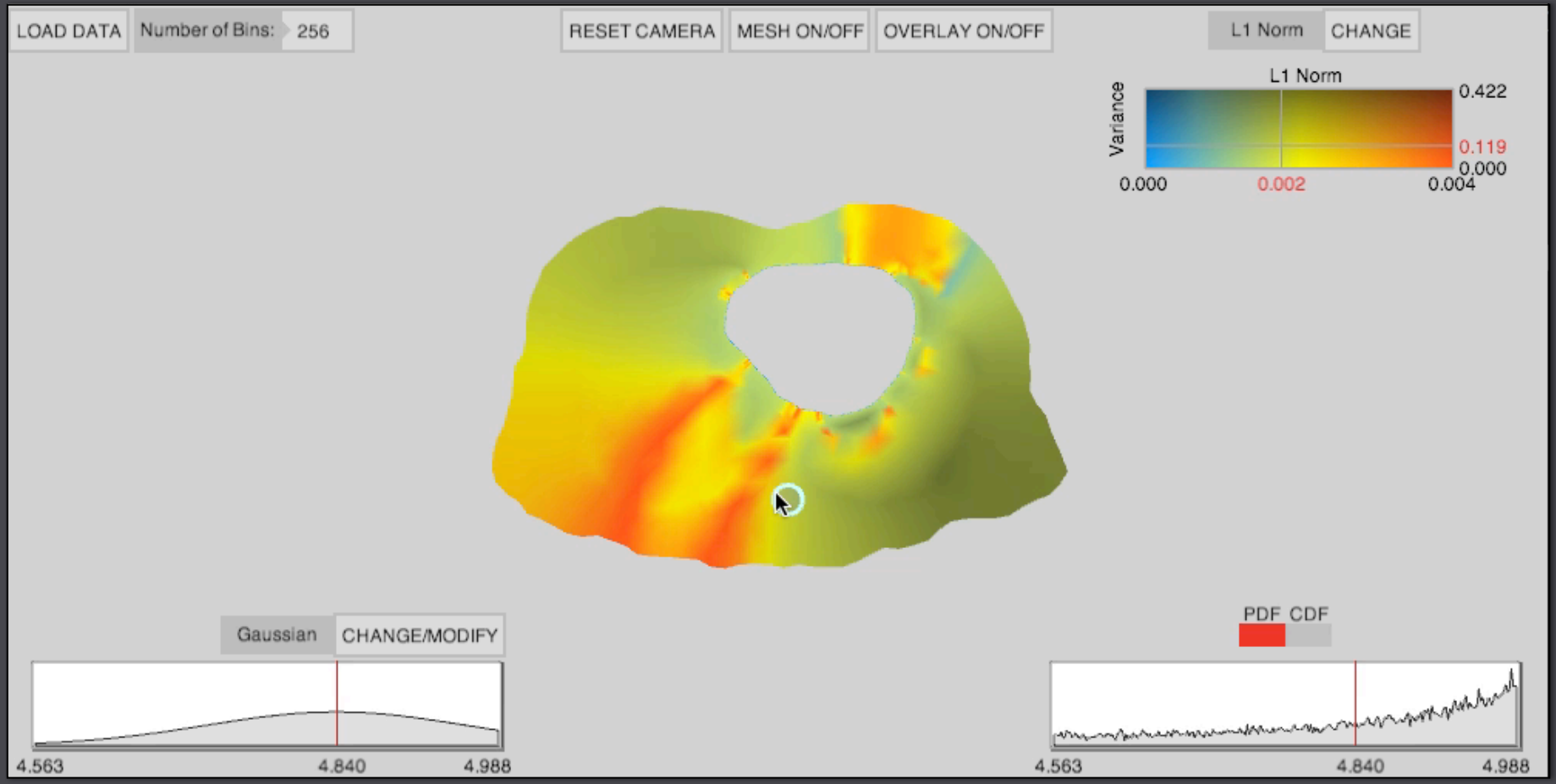






SCI INSTITUTE

Interactivity



What about Categorial Data?

Ex: Brain Segmentation

DISCRETE DISTRIBUTIONS

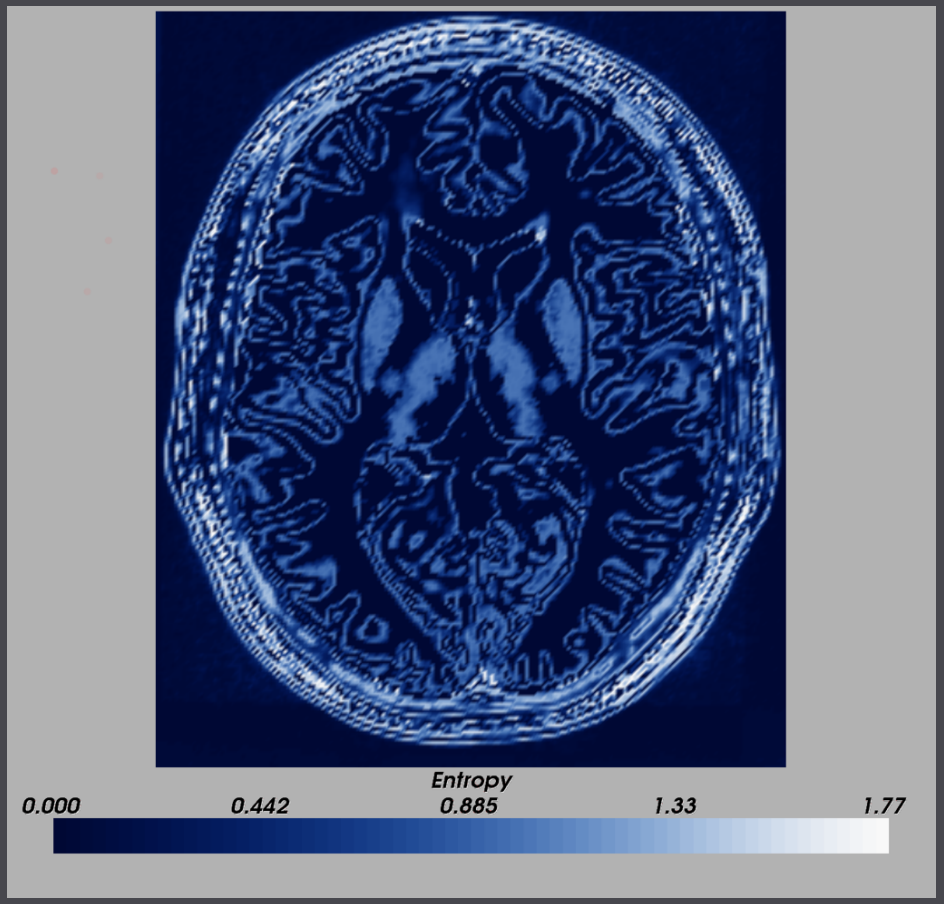
- *11 possible tissues*
- *11 distributions, one for each tissue type, describing every voxel*
- *statistics such a mean are not defined*

BrainWeb

www.bic.mni.mcgill.ca/brainweb.

SCI INSTITUTE

Measures other than mean



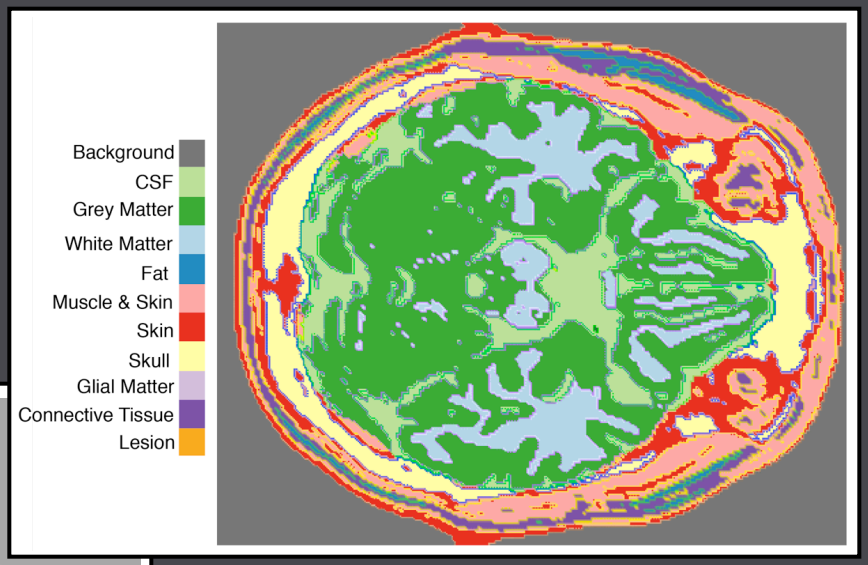
Entropy (information theory)

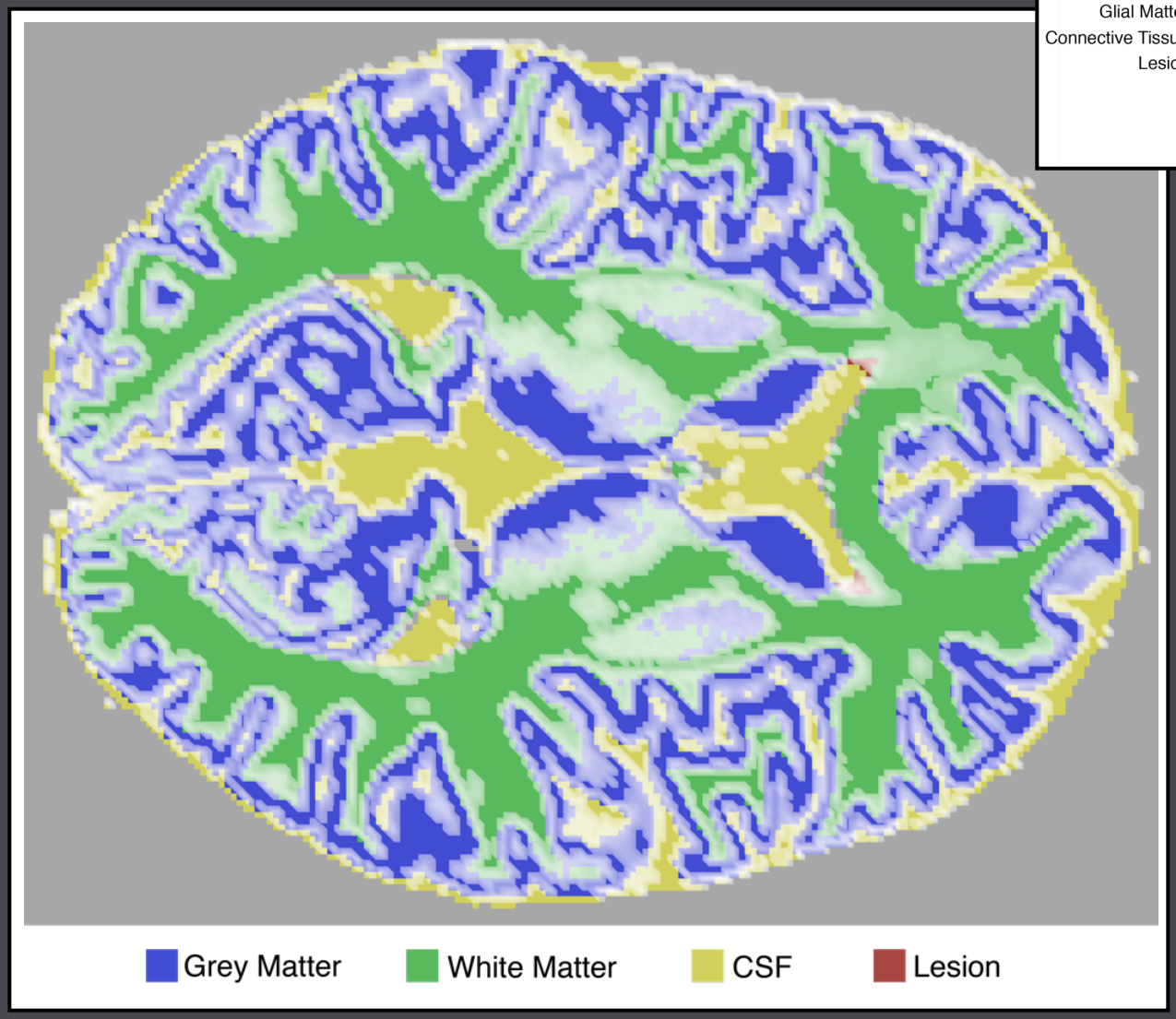
- *describes the randomness of a voxel*
- *i.e. 0 entropy a tissue's type is defined*
- *higher entropy, more uncertainty*

Visualization of Uncertainty without a Mean
K. Potter, S. Gerber, and E. W. Anderson. CG&A Vis Viewpoints, 2013, to appear.

SCI INSTITUTE

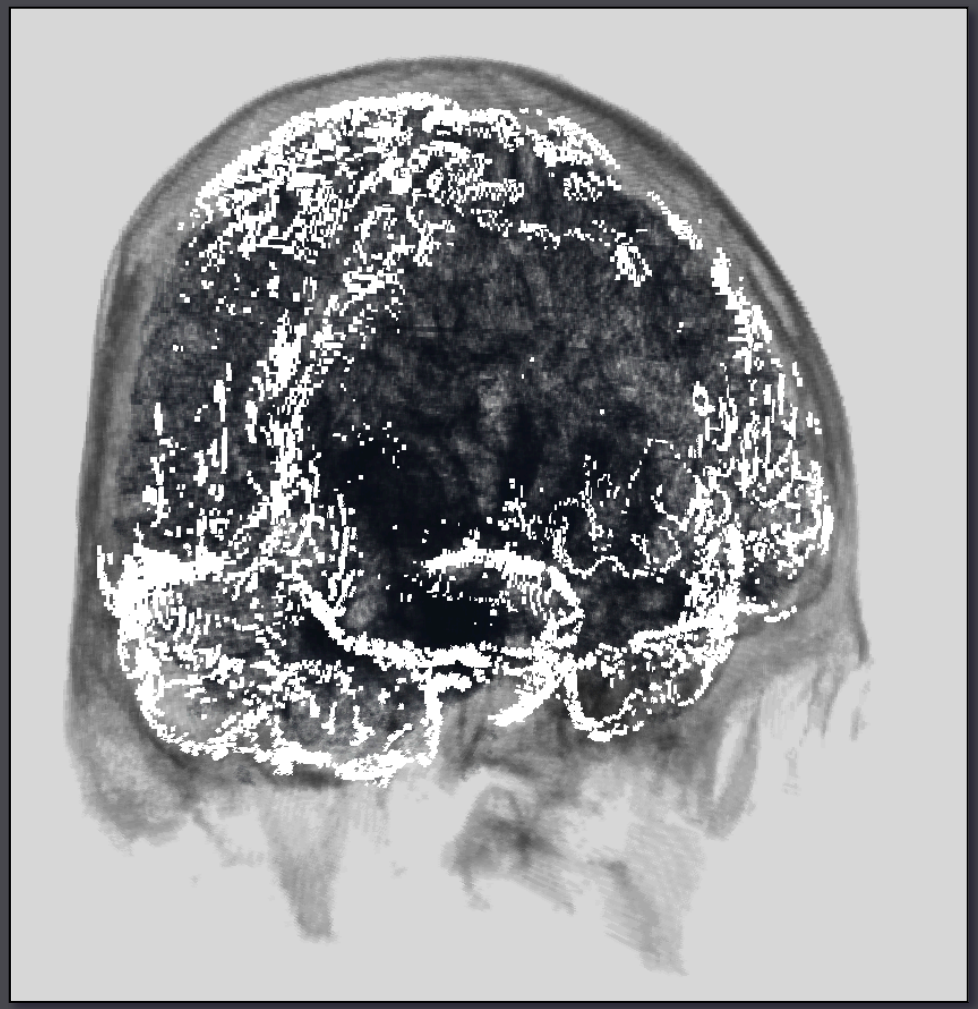
2D Display

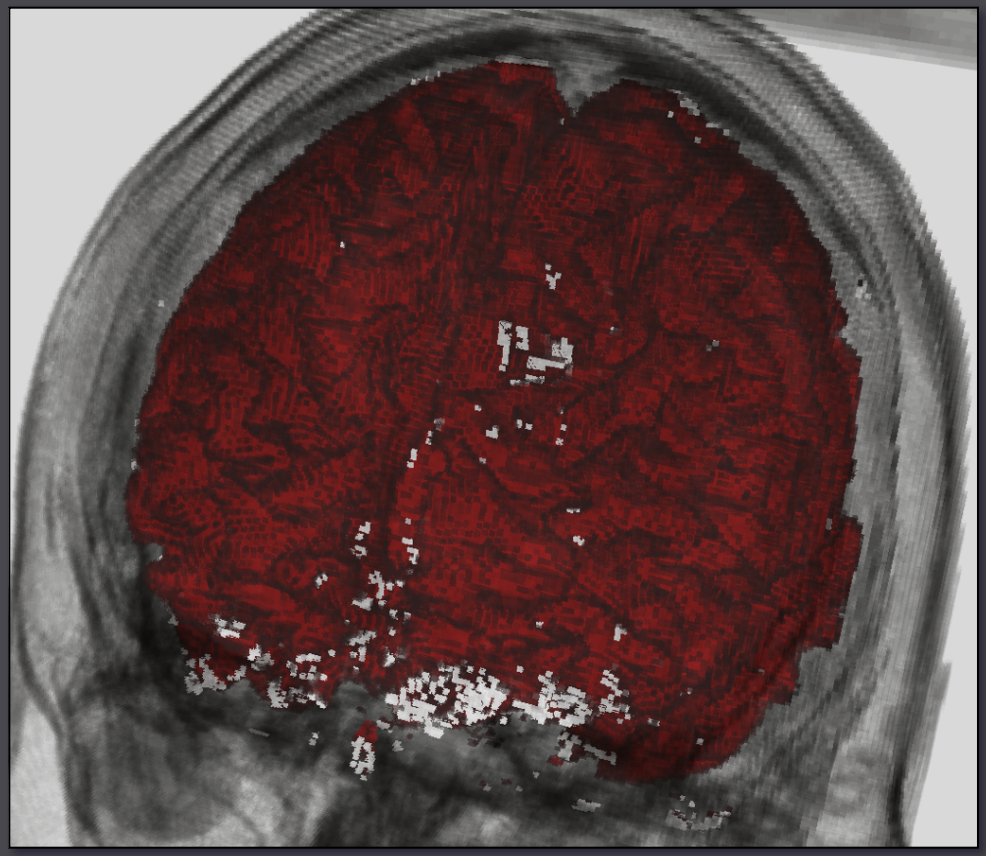




SCI INSTITUTE

3D Display





SCI INSTITUTE

In Conclusion

Aggregation takes place everywhere

- *required for visualization*
- *controllable for certain questions*

Future Directions

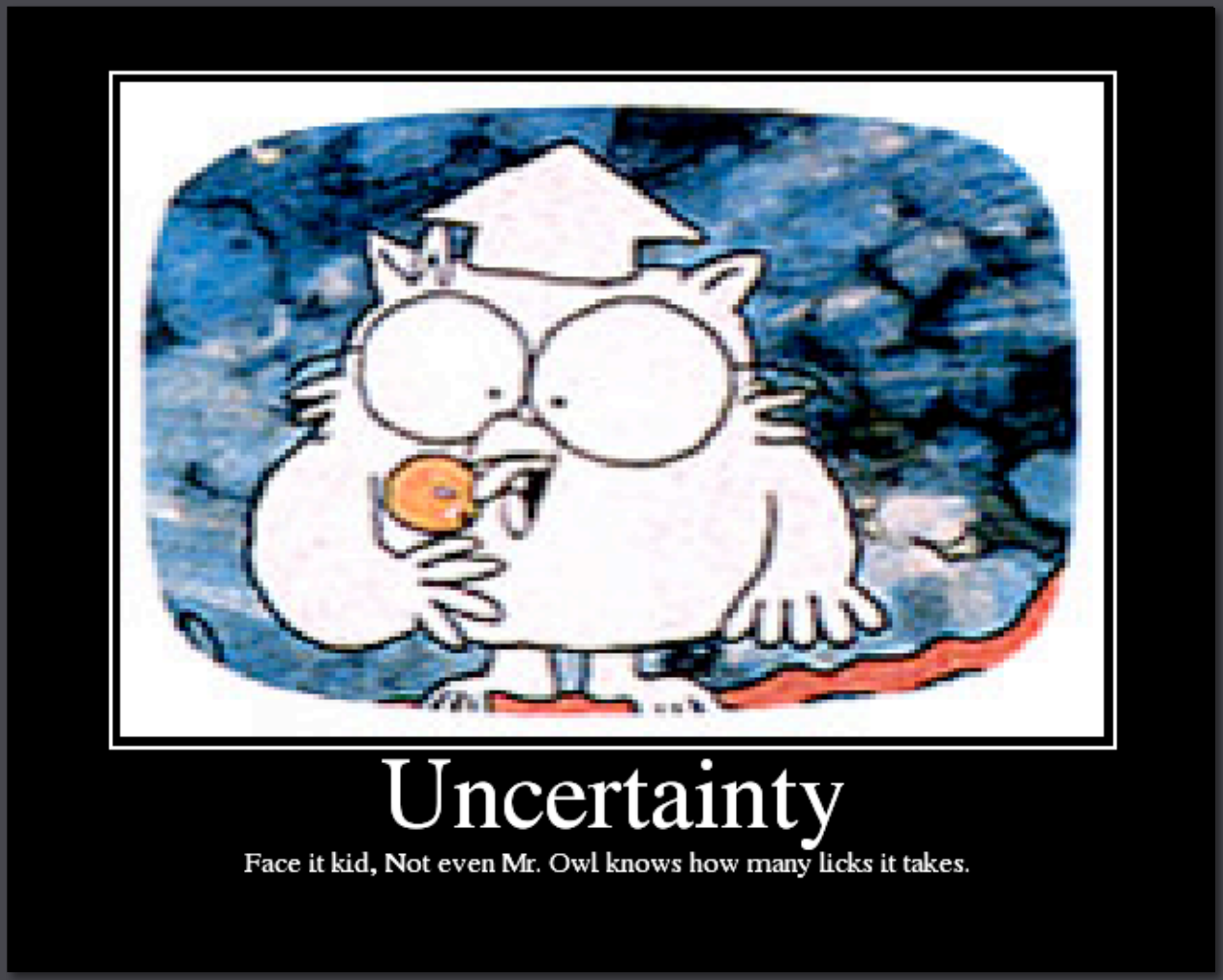
Let's keep working!

- *increase our vocabulary of uncertainty measures*
- *further the visual metaphors used for uncertainty*
- *interaction will be required for most applications*

Thanks!

Questions

- *KAUST No. KUS-C1-016-04*
- *DOE NETL DE-EE0004449*



http://tinyurl.com/8hwtzzr

SCI INSTITUTE