


**Towards Comprehensible Modeling**

**Michael Gleicher**  
 Department of Computer Sciences  
 University of Wisconsin Madison  
 on sabbatical:  
 University of Grenoble and INRIA Rhone-Alpes, team Imagine



**Acknowledgements**

This work is a collaboration with a great set of people!

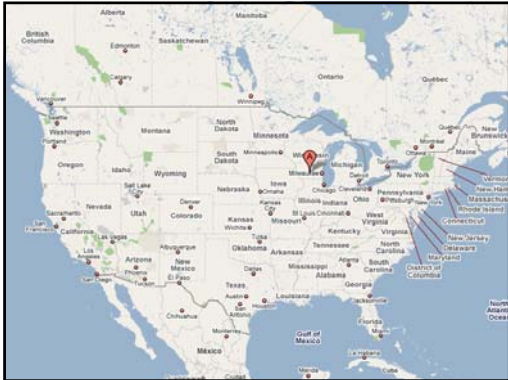
**Students:**  
 Danielle Albers, Eric Alexander, Michael Correll, Alper Sarikaya, Adrian Mayorga, Aaron Bryden, ...

**Domain Collaborators:**  
 Jonathan Hope, Robin Valenza, Mike Witmore, Dave O'Connor, ...

**Vis & HCI Collaborators:**  
 Steven Franconeri, Chuck Hansen, Bilge Mutlu, ...


**INRIA**  
 Remi Ronfard, Marie-Paule Cani, Francois Faure, Martin Guay, ...

This research is funded in part by the National Science Foundation and the Andrew Mellon Foundation. At INRIA I am supported by ERC "Expressive".



**Towards Comprehensible Modeling**

**Michael Gleicher**  
 Department of Computer Sciences  
 University of Wisconsin Madison  
 on sabbatical:  
 University of Grenoble and INRIA Rhone-Alpes, team Imagine




**Next Big Thing alert!**

My last "next big thing" (and current "big thing")

**Visualizing Comparisons**

My next big thing at Dagstuhl 2010...

### Visualizing Comparisons:

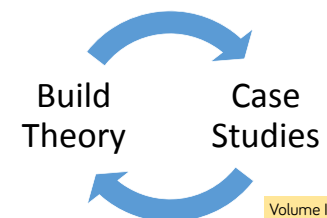
There are general principles that apply across domains, data types, ...

And if we can figure them out, it'll be easier to crank out the comparison tools/techniques quickly

### Mike's theory of visual comparison 0.2\*



Circa 2010 – hasn't changed much/  
Hasn't been written up (yet) either.



Volume Images  
Molecular Motions  
Small Graphs  
Gene Sequences  
Collections of Text  
...  
(your data here)

### Models in Comparison

**Descriptive** – reduce complexity so can see differences

**Statistical** – understand if differences are significant

**Discriminative** – are there differences between groups?

**Generative** – trendy path to discrimination

**Predictive** – can you generalize to the unseen?

(and others, and for other tasks)



Warning: beware of conflating  
model type and model application

### What is a Good model?

### What is a Good model?

Accuracy


### What is a Good model?

Descriptive Accuracy  
 Predictive Accuracy

### What is a Good model?

Descriptive Accuracy	Pre-scriptive power
Predictive Accuracy	Self-consistency
Generalizability	Validatability
Efficiency in use	Verifyability
Efficiency in construction	Simplicity
Robustness	Parsimony
Data required to build	
Conciseness	<b>Understandability</b>
Compactness	<b>Comprehensibility</b>

### Tradeoffs



Give the **user** control over the tradeoffs

### Understandability / Comprehensibility

What? Why?	To whom?
Build/assess causal theories?	To the model creator
Observe salient patterns?	To the model user
Trust?	To the model user's readers
<b>?</b>	To the general public

### Kinds of Understandability

- Understand the internals
- Understand the outputs
- Understand the meaning
- Understand the rationale
- Understand the implementation
- Understand the application
- Understand the validation

Understanding **models** vs. understanding **modeling**

### Measuring model quality

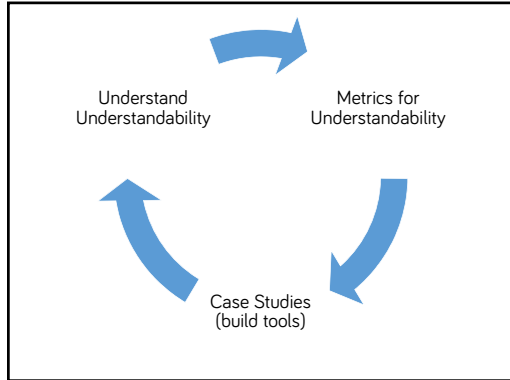
In order to make tradeoffs, you need to measure

- Accuracy  
(RMS, ROC, F1, Mathews correlation, ...)
- Generalizability  
(cross-fold, holdout, bootstrap, ...)
- Efficiency, ...

Understandability ??????

### Agenda

- Understand Understandability
- Metrics for Measuring
- Tools for user control over tradeoffs
  - tunable methods
  - different methods to give options



### 3 Paths to Understandability

- Easier to understand models
- Better explanations/visualizations of existing models
- Co-design of models and visualizations

### 3 Paths to Understandability

- Easier to understand models
  - Explainers (VAST 2013)
- Better explanations/visualizations of existing models
  - Protein Surface Classifier Validation (EuroVis14)
- Co-design of models and visualizations
  - Molecular Motion Illustration (TVCG12)
  - Splatterplots (TVCG13)

### Explainers

An approach to explore high dimensional data:

- Organize data according to **user**-defined concepts
- Explain **user**-defined concepts according to the data
- Give the **user** control over tradeoffs

Gleicher. Explainers: Expert Explorations with Crafted Projections. IEEE TVCG 19(12), Dec 2013. Proceedings VAST 2013, Best Paper Honorable Mention.

Texts

→

Vectors

```
4, 0, 3.6, 4.7, 0, 3.4, ...
3, 2.4, 0, 4.2, 4.7, 5, ...
1.5, 2.3, 0, 1.2, 6.2, ...
...
```

36 Plays = 36 Vectors

115 "Measurements" of each text = 115 dimensions

### comedic-ness

A measure of how much of a comedy something is

It's the "stuff" comedies have more of  
Where "stuff" has to be in the data

Tragedy

History

Late Plays

<

Comedy

**Organization:**  
What is most/least comedic?

**Explanation:**  
How is the word usage (measured stuff) different in comedies?

M - I (5 wrong)  
 C - B - I (4 wrong)  
 C - I - 10 M (1 wrong)  
 31 D - 100 M - 3 A (none wrong)

"standard" L1 SVM (none wrong, reasonable margin)  
 25.3698 Q + 11.8823 U + 6.9492 F + 5.4897 A + 4.1489 P -  
 3.3765 N + 2.6392 D + 2.0172 F - 1.5404 I + 1.1864 R -  
 0.7958 C + 0.7272 D

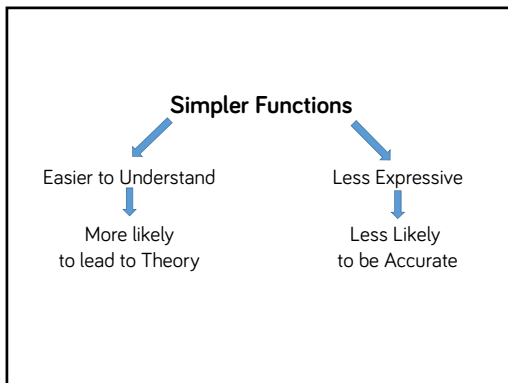
### What's Understandable\*?

Simple form (linear vs. non-linear,...)  
**A+B** vs.  $e^{-\omega k(A,B)}$


Parsimony (few variables)  
**A+B** vs.  $W+X+Y+Z$

Simple Coefficients (small integers)  
**A - 2B** vs.  $1.235 A - 4.327 B$

Familiar Variables  
**A + B** vs  $Q + W$

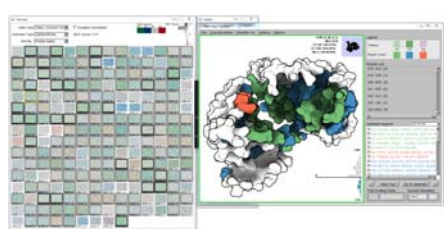


### Tradeoffs



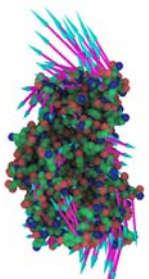
Give the **user** control over the tradeoffs

### Visual validation of modeling experiments

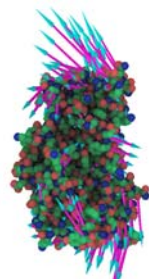


Sarikaya, Albers, Mitchel and Gleicher. *Visualizing Validation of Protein Surface Classifiers*. Computer Graphics Forum. Proceedings EuroVis 2014 (next week!)

### Molecular Motions

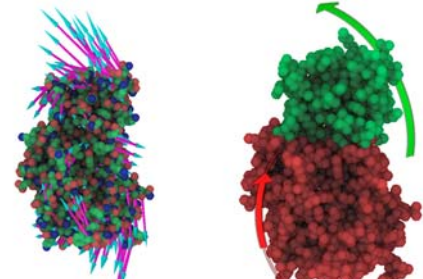


Bryden, Phillips and Gleicher. *Automated Illustration of Molecular Flexibility*. IEEE TVCG 18(1), Jan 2012.



Coarse-grained models  
Normal-mode Analysis (NMA)

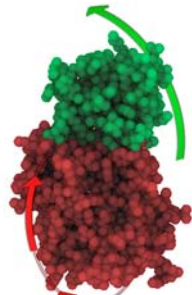
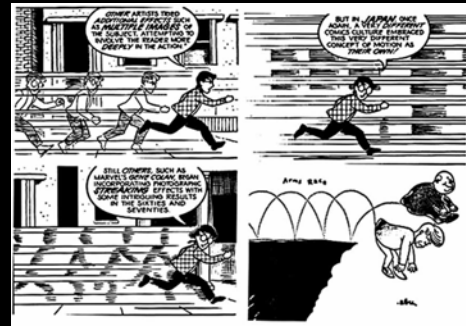
### Motion Illustration



### Motion Illustration

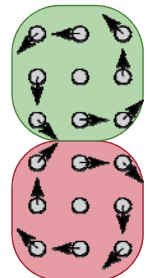
Artistic Inspirations:  
Comic Books / Diagrams

Abstract  
Model  
Illustrate

Originally from Scott McCloud's *Understanding Comics*.  
Reprinted from Cutting, *Representing motion in a static image*, 2002

Abstract  
Group to fit models  
Model  
Illustrate



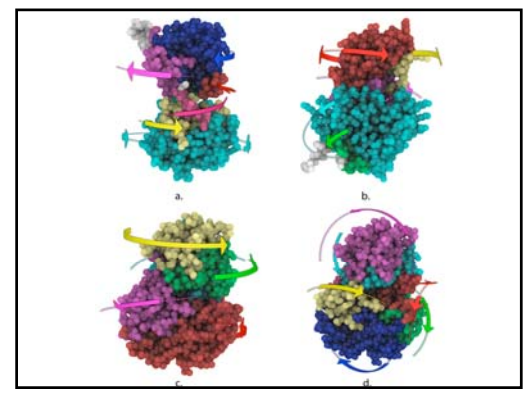
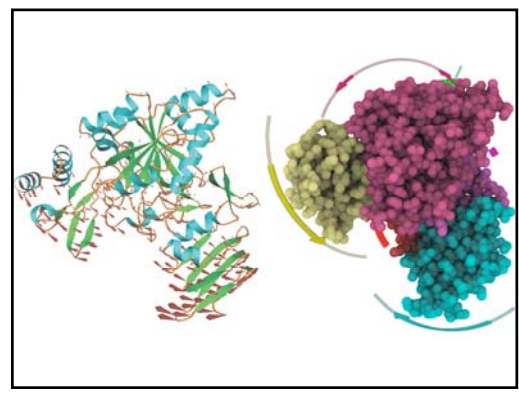
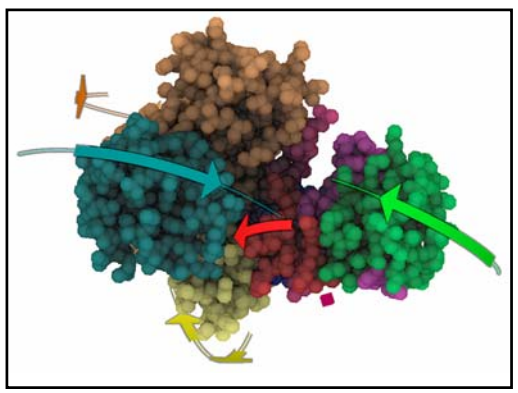
Abstract  
 Group to fit models  
**Model**  
 Affine model per group  
 Illustrate

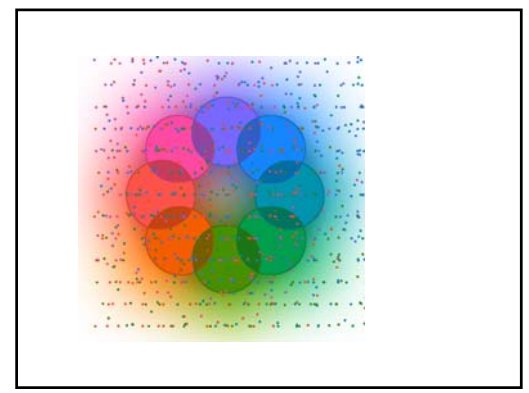
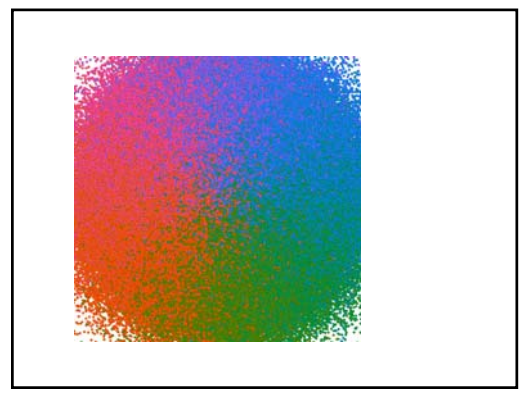
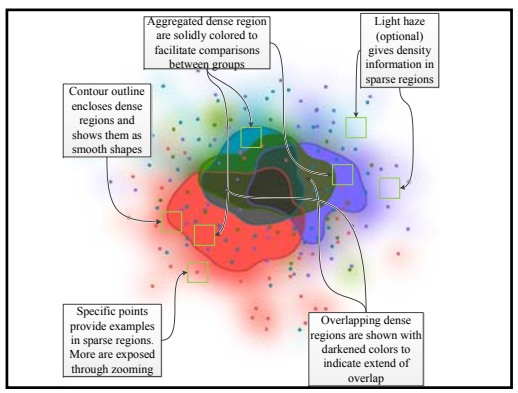
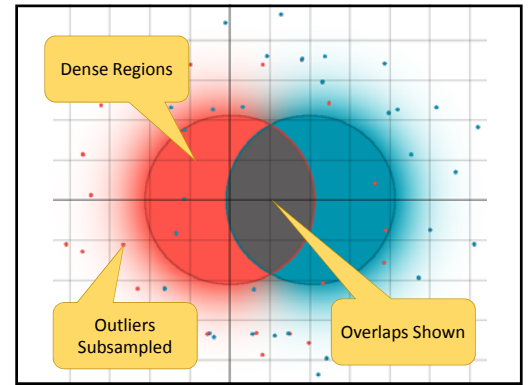
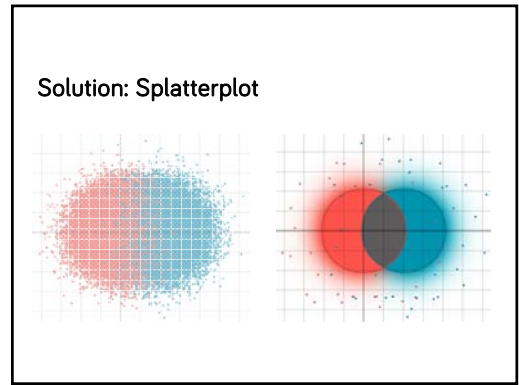
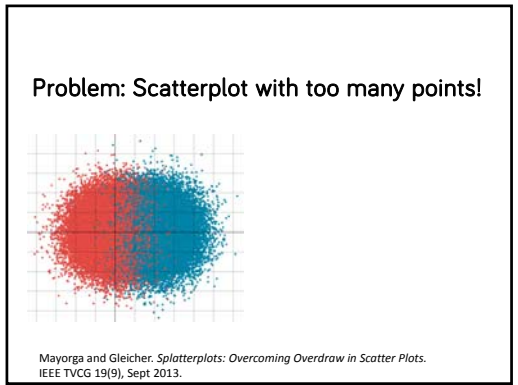
$\dot{x} = A x$   
 $\dot{x} = A x$

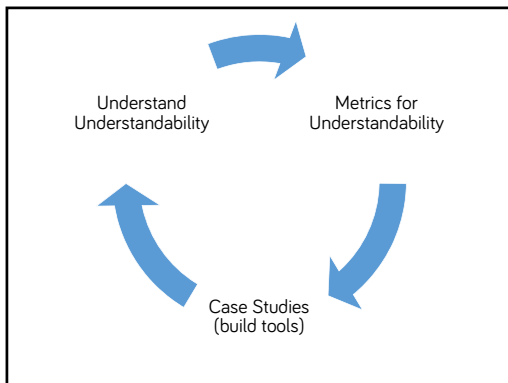
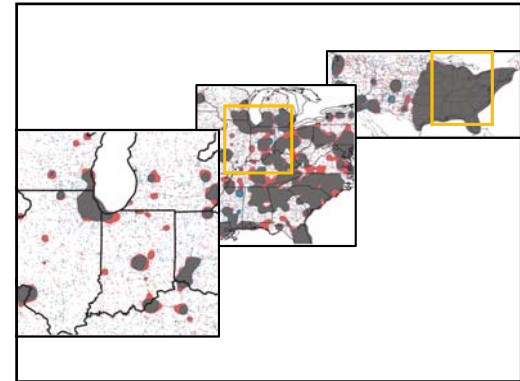
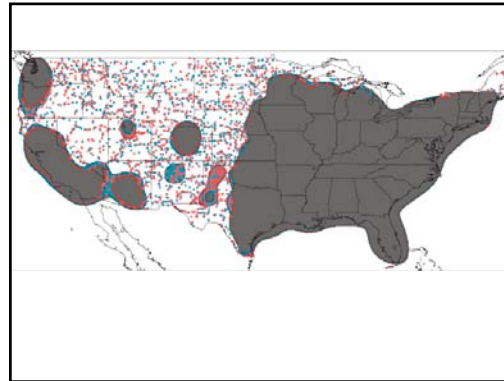
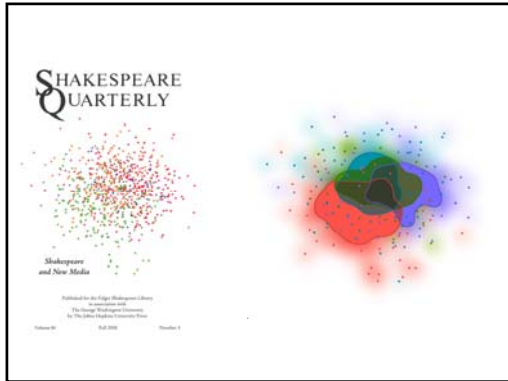
Abstract  
 Group to fit models  
 Model  
 Affine model per group  
**Illustrate**  
 Affine Exponentials  
 Glyph Design

$\dot{x} = A x$   
 $p(t) = e^{At}$   
 $\dot{x} = A x$   
 $p(t) = e^{At}$

$\dot{x} = A x$   
 $p(t) = e^{At}$   
 $\dot{x} = A x$   
 $p(t) = e^{At}$







**Thanks!**  
To you for listening.  
To the organizers for inviting me  
To my students and collaborators.  
To the NSF and Mellon Foundation for funding.

**Towards Comprehensible Modeling**

**Michael Gleicher**  
University of Wisconsin Madison  
(on sabbatical at INRIA Rhone-Alpes)  
gleicher@cs.wisc.edu

