Visualization Highlights

Tamara Munzner
Department of Computer Science
University of British Columbia

Amore Reading Group, Mar 24 2017

http://www.cs.ubc.ca/~tmm/talks.html#amore17

Visualization highlights

- terrain of visualization venues
  - names, scopes, relative strengths
- a few recent interesting papers
  - dimensionality reduction for visual data analysis
    - Probing Projections *https://uclab.fh-potsdam.de/projects/probing-projections/*
  - visualization to understand deep learning
    - Towards Better Analysis of Deep Convolutional Neural Networks *http://www.shixialiu.com/publications/cnnvis/paper.pdf*
    - Visualizing the Hidden Activity of Artificial Neural Networks *http://www.cs.rug.nl/~alext/PAPERS/VAST16/paper.pdf*
  - visualization incorporating ideas from ML
    - Surprise! Bayesian Weighting for De-Biasing Thematic Maps *https://idl.cs.washington.edu/papers/surprise-maps/*
  - scalable algorithms
    - Nanocubes *http://www.nanocubes.net/*
    - Hashedcubes *https://cscheid.net/static/papers/infovis_hashed_cubes_2016.pdf*

Dimensionality reduction: Background, our past work

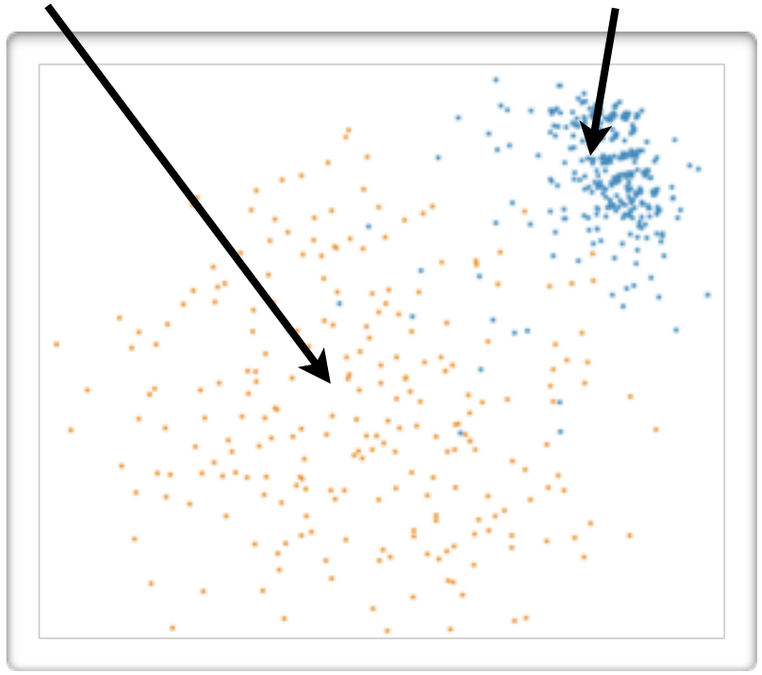
Dimensionality Reduction

- what is it?
  - –map data from high-dimensional measured space into low-dimensional target space
- when to use it?
  - –when you can't directly measure what you care about
    - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
    - latent factors, hidden variables

DR Example: Tumor Malignancy

Tumor
Measurement
Data

→ DR →



9 Dimensional
Measured Space

2 Dimensional
Target Space

DR Example: Large Document Collections

Task 1

	Dim 1	Dim ...	Dim n
Item 1			
Item ...			
Item n			

	Dim 1	Dim 2
Item 1		
Item ...		
Item n		

In
HD data

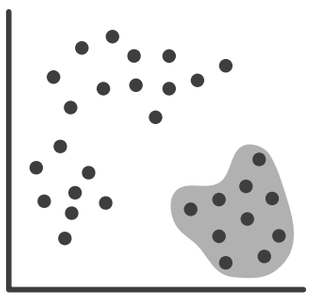
Out
2D data

What?
- In High-dimensional data
- Out 2D data

Why?
- Produce
- Derive

Task 2

	Dim 1	Dim 2
Item 1		
Item ...		
Item n		



In
2D data

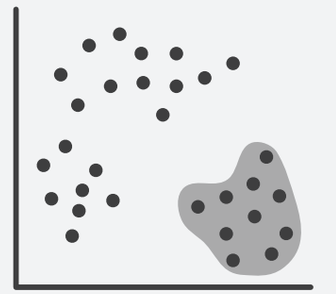
Out
Scatterplot
Clusters & points

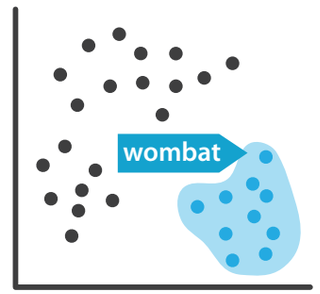
What?
- In 2D data
- Out Scatterplot
- Out Clusters & points

Why?
- Discover
- Explore
- Identify

How?
- Encode
- Navigate
- Select

Task 3





In
Scatterplot
Clusters & points

Out
Labels for clusters

What?
- In Scatterplot
- In Clusters & points
- Out Labels for clusters

Why?
- Produce
- Annotate

Dimensionality Reduction

- why do people do DR?
  - –improve performance of downstream algorithm
    - avoid curse of dimensionality
  - –data analysis
    - if look at the output: visual data analysis

Visualizing Dimensionally-Reduced Data:

Interviews with Analysts and a Characterization of Task Sequences

joint work with:
Michael Sedlmair, Matthew Brehmer, Stephen Ingram

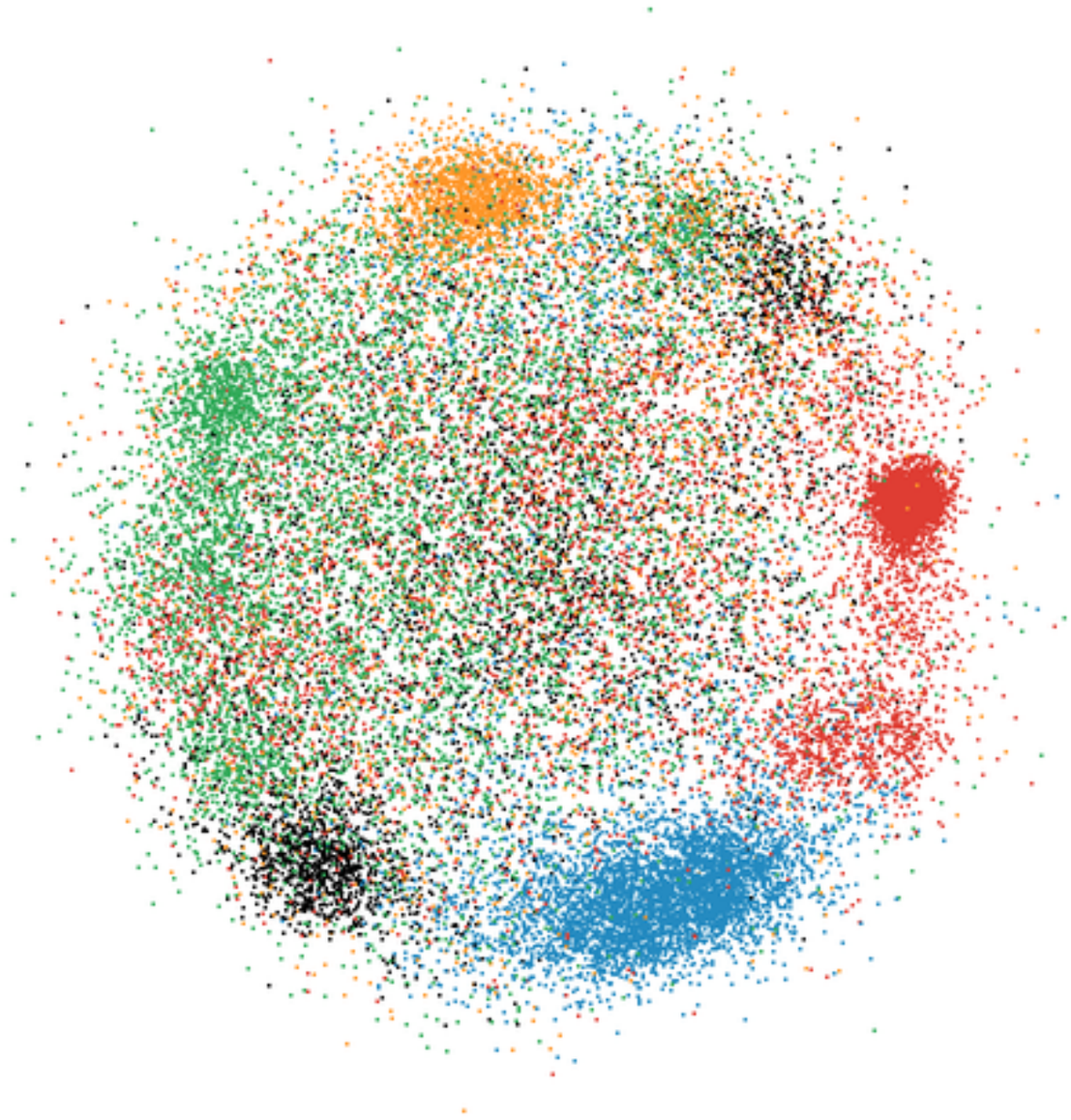
http://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/

Two-Year Cross-Domain Qualitative Study

- interviewed two dozen high-dim data analysts
  - how are they using DR?
    - does it match up with assumptions?
  - in the wild: HCI term for field work with real users
- five abstract tasks
  - naming synthesized dimensions
  - mapping synthesized dimension to original dimensions
  - verifying clusters
  - naming clusters
  - matching clusters and classes

Glimmer

Multilevel MDS on the GPU



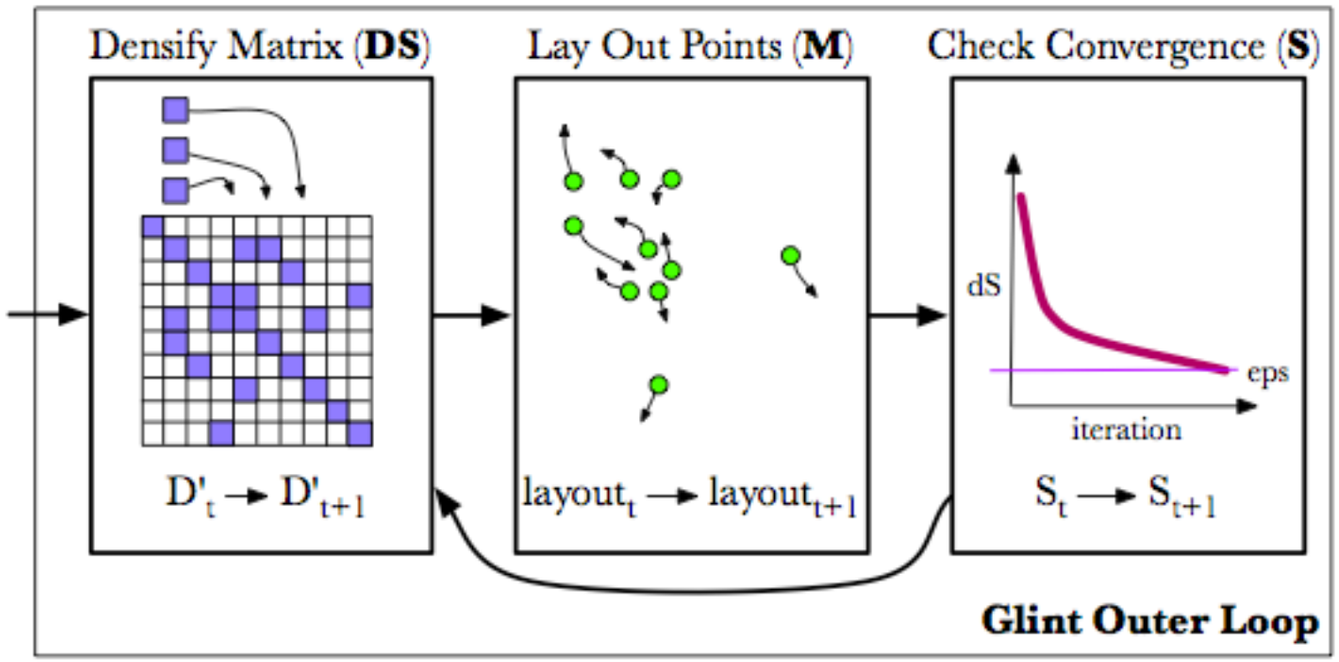
joint work with:
Stephen Ingram, Marc Olano

http://www.cs.ubc.ca/labs/imager/tr/2008/glimmer/

Glimmer: Multilevel MDS on the GPU.
Ingram, Munzner, Olano. IEEE TVCG 15(2):249-261, 2009.

MDS: Multidimensional Scaling

- entire family of methods, linear and nonlinear
- classical scaling: minimize strain
  - Nystrom/spectral methods: O(N)
    - Landmark MDS [de Silva 2004], PivotMDS [Brandes & Pich 2006]
  - limitations: quality for very high dimensional sparse data
- distance scaling: minimize stress
  - nonlinear optimization: $O(N^2)$
    - SMACOF [de Leeuw 1977]
  - force-directed placement: $O(N^2)$
    - Stochastic Force [Chalmers 1996]
    - limitations: quality problems from local minima
- Glimmer goal: O(N) speed and high quality



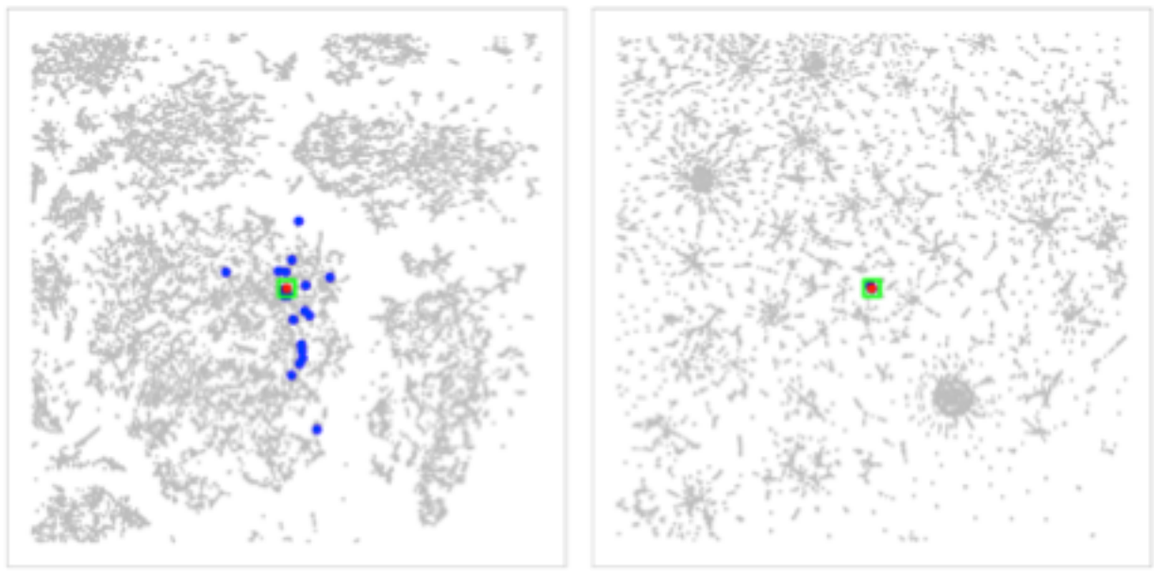
Glint

An MDS Framework for Costly Distance Functions

joint work with:
Stephen Ingram

http://www.cs.ubc.ca/labs/imager/tr/2012/Glint/

Glint: An MDS Framework for Costly Distance Functions.
Ingram, Munzner. Proc. SIGRAD 2012.



Dimensionality Reduction for Documents with

Nearest Neighbour Queries

joint work with:
Stephen Ingram

http://www.cs.ubc.ca/labs/imager/tr/2014/QSNE

Dimensionality Reduction for Documents with Nearest Neighbor Queries. *Ingram, Munzner. Neurocomputing (Special Issue for Workshop on Visual Analytics using Multidimensional Projections (VAMP) held at EuroVis 2013), Volume 150 Part B, p 557-569, 2015.*