Introduction
Visualization is an integral part of data analysis [

LiuWangThiagarajan2015]
Who am I?

Prof. Bei Wang Phillips

Data Analysis and Data Visualization

beiwang@sci.utah.edu
http://www.sci.utah.edu/~beiwang/
Structural inference of point cloud data & stratification learning

Source: Bei Wang
Robust Feature Extraction of vector field data

[SkrabaWangChen2015]
Visual Analytics of high-dim data

[LiuWangThiagarajan2014]
Visualizing brain networks & social networks

[WongPlandeWang2016]
Topology Inspired machine learning & statistics

[PalandeJoseZielinski2017]
Visualization is the secret weapon for Machine learning
Course Objective

Enable the students to become familiar with new and innovative techniques that combine data analysis with data visualization, from algorithmic and implementation perspectives.
Hot Market, Cool Startups

- AtScale: multidimensional analysis, calculation engine to run against any BI visualization tool, prediction-defined aggregates.
- Noodle.ai: maths, algorithms that learn, glass box (not black box) algorithms that allow executives to understand risks (probabilities) and causality, and make informed decisions.
- Periscope: helps data scientists quickly build customized, highly detailed visualizations of their data.
- Ayasdi: analyze and build predictive models using big data or high-dimensional data sets; hypothesis-free, automated analytics at scale; topological data analysis.
- Gluent: data virtualization technology that makes possible what it calls "hybrid data" computing.

New, Cutting-Edge & Emerging

- Visualization research venues: recent publications, conferences
- Emerging research topics
- Known and recent techniques employed by industry
Goal

Successful completion of the course will enable the students to:

- Obtain a deeper understanding of visualization as a powerful tool for data analysis, in particular, machine learning [User]

- Apply emerging and innovative techniques to data in various application domains [Expert User]

- Pursue new research directions in data analysis and data visualization [Developer, Researcher]
Prerequisite

Students are expected to have basic knowledge of data structures and algorithmic techniques, bachelor-level knowledge in mathematics or computer science, and working knowledge of programming, ideally with Python and/or C++.

Targeted audience: PhD students, master students and very-motivated upper level undergraduate students.

The students are not required to be majoring in CS, but it is preferable that the students have some background in algorithms and/or other data science related courses, and have working knowledge of programming, ideally with Python and/or C++.
Assignments and Grading

- 4 assignments in the form of mini projects (60 points, 60%)
- Final project (40 points, 40%)
  - Final project proposal (10 points, 10%)
  - Final project report (20 points, 20%; 5 points are for progress report; 15 points for final report)
  - Final project presentation (10 points, 10%)
- Additional 10 bonus points may be available in the form of bonus assignment questions.

Grading:
- A 100-93 A- 93-90
- B+ 90-87 B 87-83 B- 83-80
- C+ 80-77 C 77-73 C- 73-70
- D+ 70-67 D 67-63 D- 63-60
- E 60-0
Course Communications

Website:
  Primary source for course information, schedule, etc.

Canvas:
- Communication from instructor via course announcement
- Secondary source for course information
- Homework submission portal
- Check to make sure you receive class announcement daily

Email: beiwang@sci.utah.edu for questions on the course
Study large and complex data

- high-dim data
- Network data
- Personal data
- Scalar, Vector field
Mutually Inclusive Modules

HD
High-dim data visualization

Topo
Topological data abstraction and summarization

NV
Network Visualization

PV
Personalized Visualization
Machine Learning At a Glance

Supervised Learning
- Problems: Classification, Regression
- Algorithms: Logistic Regression, Back Propagation Neural Network

Unsupervised Learning
- Problems: Clustering, Dimensionality Reduction
- Algorithms: k-means, Data Mining, Topological Data Analysis

Semi-supervised Learning
- Problems: Classification, Regression
- Algorithms: extensions to flexible algorithms, model unlabelled data

Source: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/
Vis+ML

- high-dim data
- Network data
- Personal data
- Scalar, Vector field
Obtain insight from high-dimensional data through ML and interactive VIS
A square grid with equal spacing between points. Try convergence at different sizes.
Each dot is a 2D data sample: real samples, fake samples.

Background colors of grid cells represent discriminator’s classifications. Samples in green regions are likely to be real; those in purple regions likely fake.

Manifold represents generator’s transformation results from noise space. Opacity encodes density: darker purple means more samples in smaller area.

Pink lines from fake samples represent gradients for generator. This sample needs to move upper right to decrease generator’s loss.

Source: https://poloclub.github.io/ganlab/
2: TOPO
Topological abstraction & summarization
Topological data analysis and visualization capture the shape of complex data.
Scalar & vector field data

Topology ToolKit
Efficient, generic and easy
Topological data analysis and visualization
3:NV
Network Vis
A picture is worth a 1000 words, but
Static vs time-varying networks

[Haji]WangScheideggerRosen2018
Multilayer & multivariate networks

[DomenicoPorterArenas2015]
Scalability

Source: Bei Wang and Sravankuma Neerati
4:PV
Personalized Vis
Visualizing for individuals

Source: Giorgia Lupi, Accurat
Visualizing personal data
Class Syllabus and Final Project

- Final project key dates:
  - Project team creation: due September 12.
  - Project proposal description: due October 15.
  - Project progress report: due November 12.
  - Project final report: due December 10.
  - Project presentations: on December 5 (9:10 - 10:30 a.m.) and December 9 (8:00 - 10:00 a.m.)


How to succeed in class

- Attend lectures
- Start thinking about final project early
- Ask questions in class
- Getting help: office hour, Tuesday 10:30 to 11:30 a.m. or by appointment, by email with title “CS 6965”
- Learning programming along the way: D3.js, TTK, Python, etc.
Mandatory readings

- Scikit-learn tutorial:
Getting ready for mini-projects
Python, D3.js, etc.

- Interactive Data Visualization for the Web, 2nd Ed.
Slide Deck References


Thanks!

Any questions?

You can find me at: beiwang@sci.utah.edu
CREDITS

Special thanks to all people who made and share these awesome resources for free:

▷ Presentation template designed by Slidesmash
▷ Photographs by unsplash.com and pexels.com
▷ Vector Icons by Matthew Skiles
Presentation Design

This presentation uses the following typographies and colors:

Free Fonts used:
https://www.fontsquirrel.com/fonts/open-sans

Colors used

[Color swatches]