Data Science and Predictive Analytics Academic - Industry Partnering Forum

April 27, 2018

Data Analysis and Visualization

Wayne Oldford

Statistics & Actuarial Science

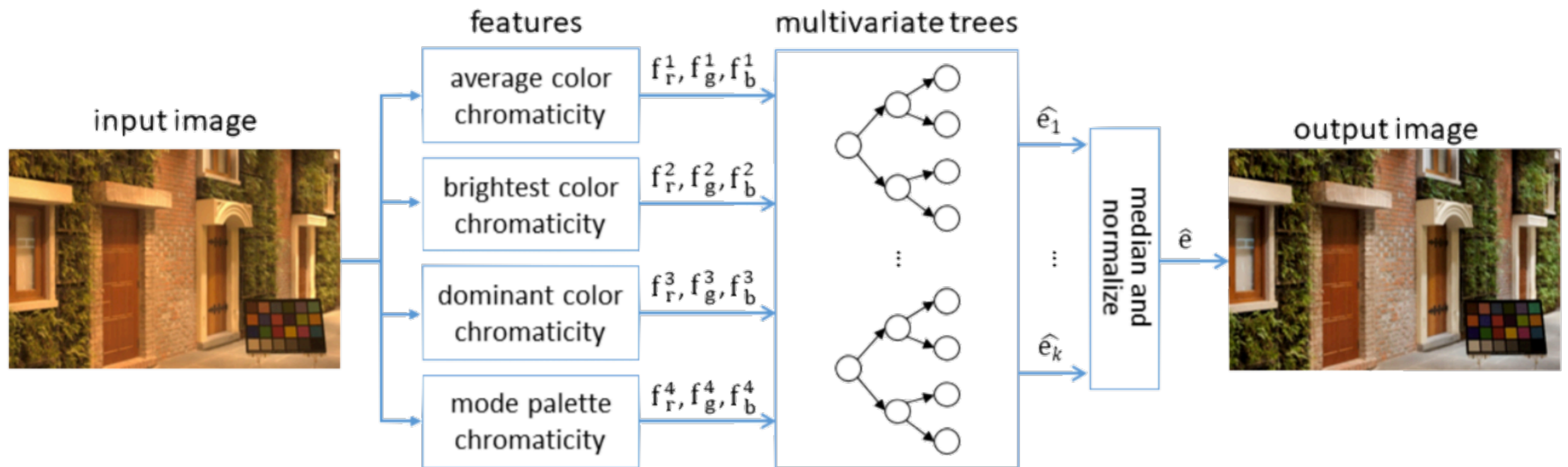UNIVERSITY OF
WATERLOO

Research interests

- ## Data analysis and predictive analytics

man - machine tradeoff:

from enabled to automated analytics

AUTOMATED: Computational photography — illuminant estimation

accurate (competitive with deep neural nets)

compact and efficient (can be implemented on board the camera)

meaningful (models well understood colour features and human colour discrimination)

Research interests

- ## Data analysis and predictive analytics

  man - machine tradeoff:

  from enabled to automated analytics

ENABLING THE ANALYST: Browser-based dashboard for Hillary Clinton's emails
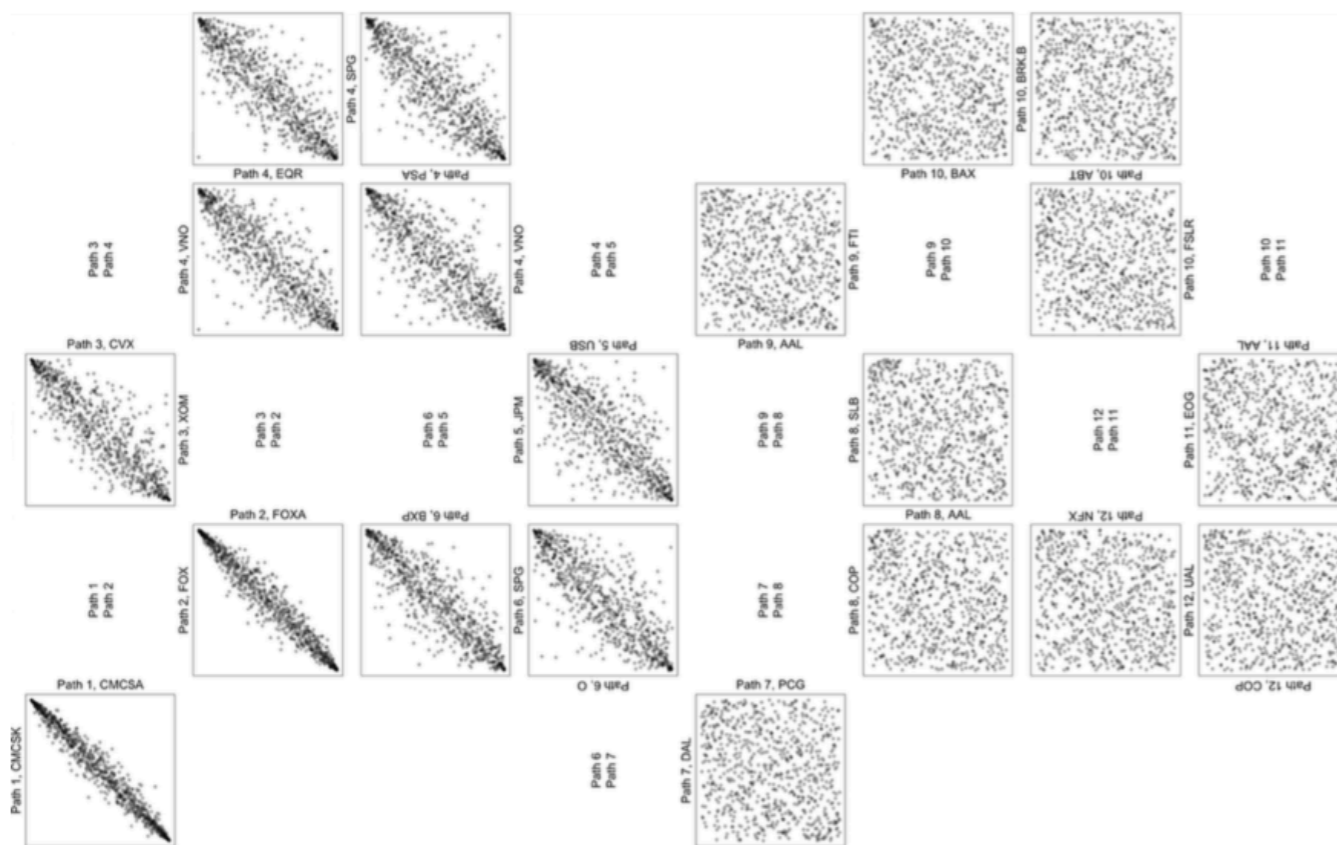
rshiny.math.uwaterloo.ca/clinton
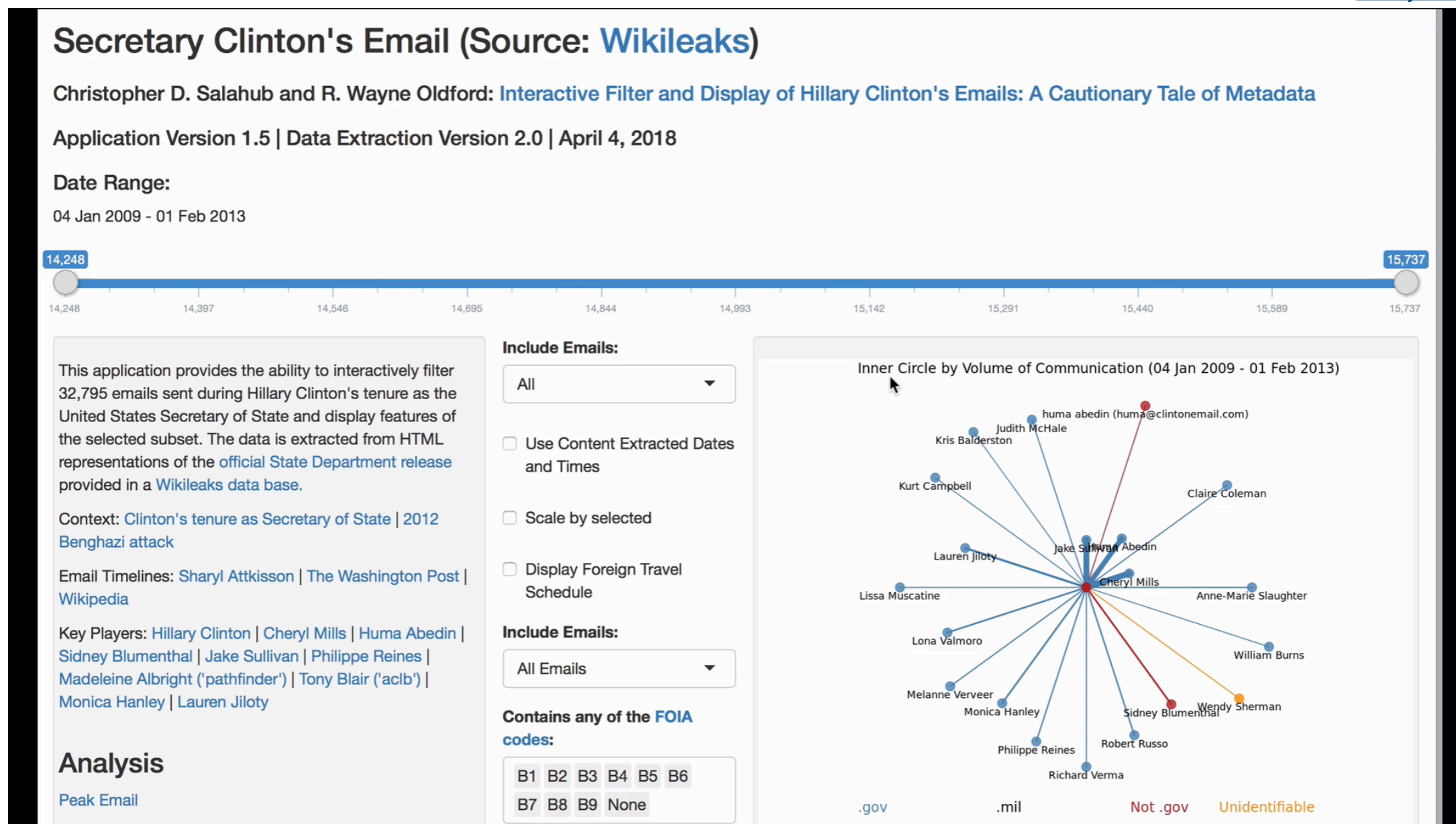
Research interests
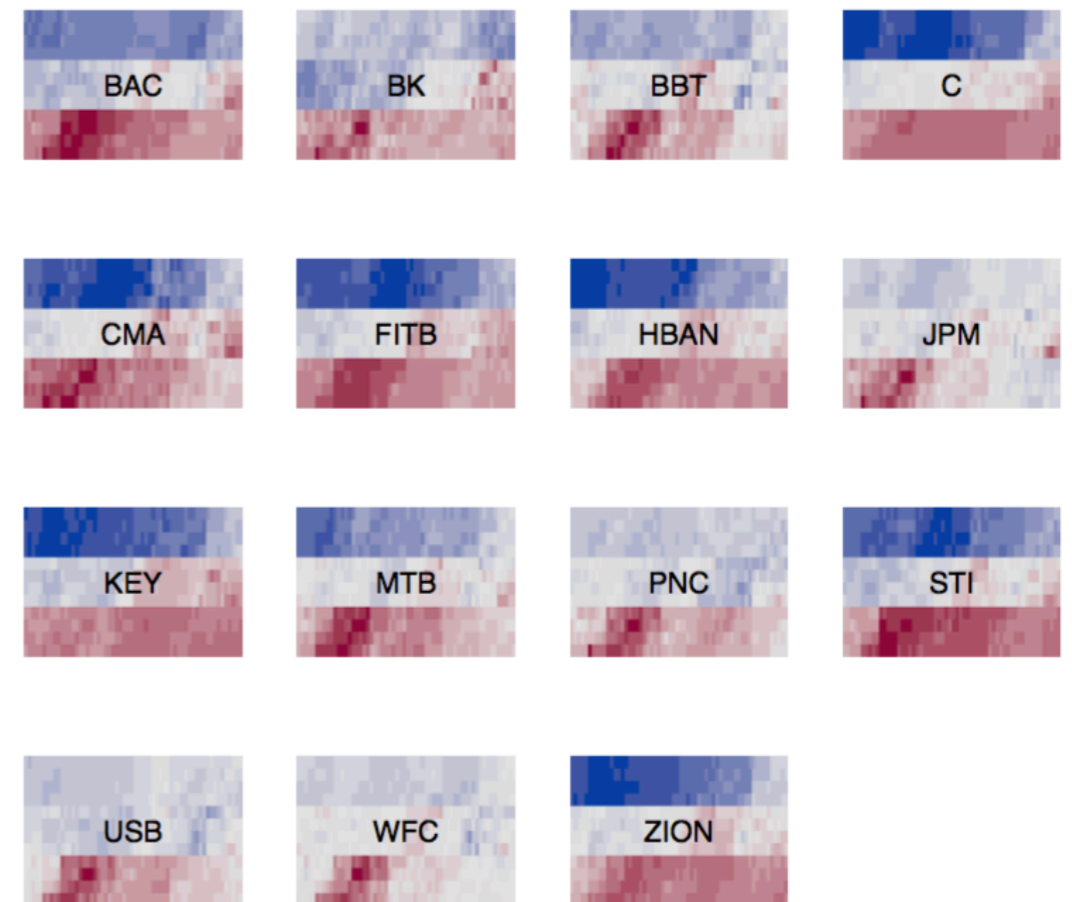
- ## Guided analytics and visualization

Search for interesting patterns:

- pre-compute, sort, and search geometric and statistical measures
- present meaningful displays to show unanticipated pattern

S&P 500 daily returns 2007-2009

10 most and least dependent pairs of stocks

Banks: three years of daily returns

Research interests

- ● Guided analytics and visualization

Data reduction:

- − numbers of observations, numbers of variates/attributes

- − preserve patterns (e.g. distribution and geometry)

560 dimensions -> 5

Research interests

- Guided analytics and visualization

Data reduction:

- numbers of observations, numbers of variates/attributes
- preserve patterns (e.g. distribution and geometry)

560 dimensions -> 5

Research interests

- Guided analytics and visualization

  Data reduction:

  - numbers of observations, numbers of variates/attributes

  - preserve patterns (e.g. distribution and geometry)

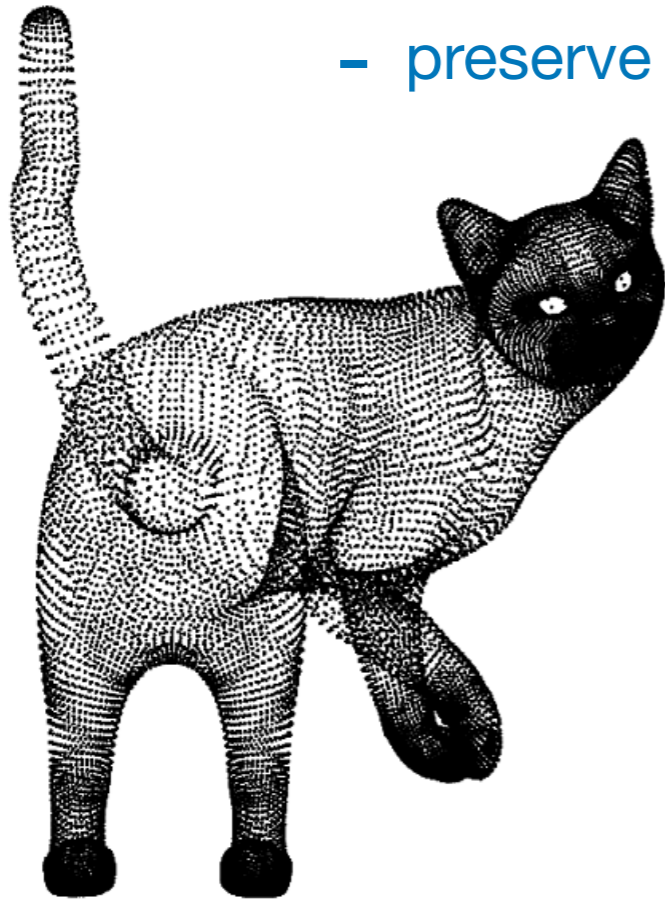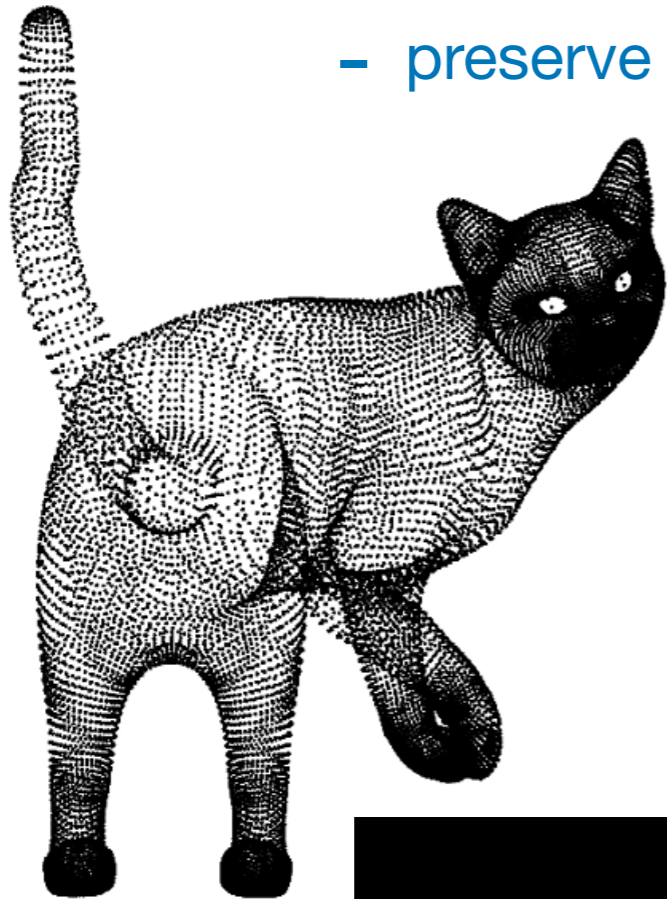 about 28,000 points -> 5,000 points

Research interests

- Guided analytics and visualization

  Data reduction:
  - numbers of observations, numbers of variates/attributes
  - preserve patterns (e.g. distribution and geometry)

about 28,000 points -> 5,000 points

random sampling   our method

Research interests

- ## Data analysis and predictive analytics

  man - machine tradeoff:

  from enabled to automated analytics

- ## Guided analytics and visualization

  Search for interesting patterns:
  - pre-compute, sort, and search geometric and statistical measures
  - present meaningful displays to show unanticipated pattern

  Data reduction:
  - numbers of observations, numbers of variates/attributes
  - preserve patterns (e.g. distribution and geometry)

- ## Power tools for interactive data analysis

  Software tools:
  - LOON: interactive data visualization in R
  - ZENPLOTS: automated layout of hundreds of plots
  - numerous other R packages and statistical analysis software