CSE 332

Introduction to Visualization

Introduction

Klaus Mueller

Computer Science Department
Stony Brook University
First hit on Google Image:

Creative visualization is a mental technique that uses the imagination to make dreams and goals come true. Used in the right way, creative visualization can improve your life and attract to you success and prosperity. It can alter your circumstances, cause events to happen, and attract money, possessions, work, people, and love into your life.
It shares the “imagination” part
And these days, it can also bring you “success and prosperity”
And it may “attract work, people, and love into your life”
Why Visualization?
Let’s go back some 150 years to 1854, London, England
The most terrible outbreak of cholera which ever occurred in this kingdom, is probably that which is taking place in Broad Street, Golden Square, and adjoining streets.

Within two hundred and fifty yards of the spot where Cambridge Street joins Broad Street, there are upwards of five hundred fatal attacks of cholera in ten days.

The mortality in this limited area probably equals any that was ever caused in this country, even by the plague; and it is much more sudden, as the greater number of cases terminated in a few hours.
What Can We Do?
What Is The Cause?
How Can We Eliminate It?
TIME FOR “IMAGINATION”
Cholera spreads through water
- and not via some other fantastic causes
- one said it rose out of the burying grounds of plague victims from two centuries earlier
- the bacteria was discovered later, in 1886

A real-life experiment
- established the mode of cholera transmission
- and consequently the method of prevention: keep drinking water, food, and hands clear of infected sewage

Visualization provided
- inspiration
- convincing arguments to justify actions
- led to Dr. John Snow’s historic immortality
- a bar near the old Broad Street pump bears his name (safe drinking)
What Is Needed for Visualization?

socrative by MasteryConnect
What’s Socrative?

Classroom app for effective teacher – student interaction

- I ask a question
- you answer it
- we all look at the answers and learn

Platforms (free to students!!!)

- iOS, android → download the student edition from app store
- web browser → log into http://www.socrative.com/ as student

Login procedures

- Room name: CSE332
- Student ID: your SBU netID
WHAT IS NEEDED FOR VISUALIZATION?
What Is Needed for Visualization – Some Appropriate Answers

Data (wide variety)

Algorithms
- data mining
- data analytics

Computer
- run those algorithms
- data storage

Humans
- with a purpose/need to understand their data
- endowed with cognitive faculties, creative thought, intuition
- domain expertise

Understanding of humans
- perception, cognition, HCI issues
- we can gain it through experimentation with humans
What Is Needed for Visualization – Some Appropriate Answers

Data (wide variety)
Algorithms
  ▪ data mining
  ▪ data analytics
Computer
  ▪ run those algorithms
  ▪ data storage
Humans
  ▪ with a purpose/need to understand their data
  ▪ endowed with cognitive faculties, creative thought, intuition
  ▪ domain expertise
Understanding of humans
  ▪ perception, cognition, HCI issues
  ▪ we can gain it through experimentation with humans

= Visual Analytics
Dr. John Snow’s London Cholera Map of 1854

- data collection
- data assimilation
- statistical testing
- visualization
- computational analysis (brain)
- domain knowledge

Very early example of visual analytics
Let’s go back some 30 years to 1986, JFK Space Center, FL

73 SECONDS AFTER LIFT-OFF
What Happened?

What Was The Cause?
The Day of the Launch

36 degrees F on Launch Pad 39
Rubber O-rings, nearly 38 feet (11.6 meters) in circumference; 1/4 inch (6.4 mm) thick.

The field joint that leaked.

Upper segment of rocket casing

Primary O-ring
Secondary O-ring

Upon ignition, smoke leaked from this joint. A flame burned through 59 seconds later.

Exterior wall of rocket

Lower segment of rocket casing

Inside of rocket (filled with 500 tons of propellant)
FAST FORWARD
58 SECONDS AFTER IGNITION
What Happened?

What Was The Cause?

Could It Have Been Prevented?
Two days before launch they presented their concerns
  - created 13 charts to make their case

Slide #1:

Temperature Concern on
SRM Joints
27 Jan 1986

- SRM – Solid Rocket Motor
Teaches about past damages to O-ring

### HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

<table>
<thead>
<tr>
<th>SRM No.</th>
<th>Cross Sectional View</th>
<th>Top View</th>
<th>Clocking Location (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Erosion Depth (in.)</td>
<td>Perimeter Affected (deg)</td>
<td>Nominal Dia. (in.)</td>
</tr>
<tr>
<td><strong>20A</strong></td>
<td>None</td>
<td>None</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>21A</strong></td>
<td>0.010</td>
<td>154.0</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>21B</strong></td>
<td>0.038</td>
<td>130.0</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>30B</strong></td>
<td>None</td>
<td>45.0</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>40B</strong></td>
<td>0.028</td>
<td>110.0</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>41A</strong></td>
<td>None</td>
<td>None</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>41B</strong></td>
<td>0.040</td>
<td>217.0</td>
<td>0.280</td>
</tr>
<tr>
<td><strong>50B</strong></td>
<td>0.053</td>
<td>116.0</td>
<td>0.280</td>
</tr>
</tbody>
</table>

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

**OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.**

**SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.**
Teaches about O-ring damage mechanics and erosion

**Primary Concerns** -

**Field Joint - Highest Concern**

- Erosion penetration of primary seal requires reliable secondary seal for pressure integrity
  - Ignition transient - (0-600 ms)
    - (0-170 ms) high probability of reliable secondary seal
    - (170-330 ms) reduced probability of reliable secondary seal
    - (330-500 ms) high probability of no secondary seal capability

- Steady State - (600 ms - 2 minutes)
  - If erosion penetrates primary O-ring seal - high probability of no secondary seal capability
    - Bench testing showed O-ring not capable of maintaining contact with metal parts gap opening rate to PEOG
    - Bench testing showed capability to maintain O-ring contact during initial phase (0-170 ms) of transient
Lists temperature and blow-by history for two SRMs

**Blow By History**

**SRM-15** worst Blow-By
- 2 Case Joints (80°, 110°) arc
- Much worse visually than SRM-32

**SRM 32** Blow-By
- 2 Case Joints (30-40°)

**SRM-13 A, 15, 16 A, 18, 23 A 24A**
- Nozzle Blow-By

**History of O-Ring Temperatures (Degrees - F)**

<table>
<thead>
<tr>
<th>MOTOR</th>
<th>MBT</th>
<th>AMB</th>
<th>O-Ring</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM-4</td>
<td>68</td>
<td>36</td>
<td>47</td>
<td>10 MPH</td>
</tr>
<tr>
<td>DM-2</td>
<td>76</td>
<td>45</td>
<td>52</td>
<td>10 MPH</td>
</tr>
<tr>
<td>QM-3</td>
<td>72.5</td>
<td>40</td>
<td>48</td>
<td>10 MPH</td>
</tr>
<tr>
<td>QM-4</td>
<td>76</td>
<td>48</td>
<td>51</td>
<td>10 MPH</td>
</tr>
<tr>
<td>SRM-15</td>
<td>52</td>
<td>64</td>
<td>53</td>
<td>10 MPH</td>
</tr>
<tr>
<td>SRM-22</td>
<td>77</td>
<td>78</td>
<td>75</td>
<td>10 MPH</td>
</tr>
<tr>
<td>SRM-25</td>
<td>55</td>
<td>26</td>
<td>29</td>
<td>25 MPH</td>
</tr>
</tbody>
</table>

**Wind Speeds:**
- DM-4: 10 MPH
- DM-2: 10 MPH
- QM-3: 10 MPH
- QM-4: 10 MPH
- SRM-15: 10 MPH
- SRM-22: 10 MPH
- SRM-25: 25 MPH
Given the information provided in the company slides

- would you vote for a launch?
- ignore you know about the consequences

Be keenly aware of the immense PR pressures

- President Reagan’s upcoming State of the Union speech
- the first civilian in space
- NASA’s funding problems

Launch:

- **No**: OK with a PR disaster & possible budget cuts down the road
- **Yes**: the rocket company is too cautious & concerns are unproven
Why The Recommendation Failed

Presentation only has exactly two shuttle flights
  - one with two blow-by’s and high temperature
  - one with two blow-by’s and low temperature
  - ignores all other 22 shuttle flights (SRM)

Statistically weak

Recommendation
  - “O-ring temp must be >53ºF at launch”

  - is only based on a sample size of 1
  - context of other flights is missing
  - no statistical leverage
Lots of numbers and facts

But no causal evidence that could predict

What is needed?
What Is Needed?

Need a measure for damage
# Damage Index

<table>
<thead>
<tr>
<th>Flight</th>
<th>Date</th>
<th>Temperature °F</th>
<th>Erosion incidents</th>
<th>Blow-by incidents</th>
<th>Damage index</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>51-C</td>
<td>01.24.85</td>
<td>53°</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>Most erosion any flight; blow-by; back-up rings heated. Deep, extensive erosion.</td>
</tr>
<tr>
<td>41-B</td>
<td>02.03.84</td>
<td>57°</td>
<td>1</td>
<td></td>
<td>4</td>
<td>O-ring erosion on launch two weeks before Challenger.</td>
</tr>
<tr>
<td>61-C</td>
<td>01.12.86</td>
<td>58°</td>
<td>1</td>
<td></td>
<td>4</td>
<td>O-rings showed signs of heating, but no damage.</td>
</tr>
<tr>
<td>41-C</td>
<td>04.06.84</td>
<td>63°</td>
<td>1</td>
<td></td>
<td>2</td>
<td>Coolest (66°) launch without O-ring problems.</td>
</tr>
<tr>
<td>1</td>
<td>04.12.81</td>
<td>66°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>04.04.83</td>
<td>67°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>51-A</td>
<td>11.08.84</td>
<td>67°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>51-D</td>
<td>04.12.85</td>
<td>67°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>11.11.82</td>
<td>68°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>03.22.82</td>
<td>69°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11.12.81</td>
<td>70°</td>
<td>1</td>
<td></td>
<td>4</td>
<td>Extent of erosion not fully known.</td>
</tr>
<tr>
<td>9</td>
<td>11.28.83</td>
<td>70°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>41-D</td>
<td>08.30.84</td>
<td>70°</td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>51-G</td>
<td>06.17.85</td>
<td>70°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>06.18.83</td>
<td>72°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>08.30.83</td>
<td>73°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>51-B</td>
<td>04.29.85</td>
<td>75°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>61-A</td>
<td>10.30.85</td>
<td>75°</td>
<td>2</td>
<td></td>
<td>4</td>
<td>No erosion. Soot found behind two primary O-rings.</td>
</tr>
<tr>
<td>51-I</td>
<td>08.27.85</td>
<td>76°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>61-B</td>
<td>11.26.85</td>
<td>76°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>41-G</td>
<td>10.05.84</td>
<td>78°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>51-J</td>
<td>10.03.85</td>
<td>79°</td>
<td></td>
<td></td>
<td>0</td>
<td>O-ring condition unknown; rocket casing lost at sea.</td>
</tr>
<tr>
<td>51-F</td>
<td>07.29.85</td>
<td>81°</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
O-ring damage index, each launch

26°–29° range of forecasted temperatures (as of January 27, 1986) for the launch of space shuttle Challenger on January 28

Temperature (°F) of field joints at time of launch
Extrapolation of damage curve to the cold
Challenger launch: 31° forecasted
temperature for January 28, 1986

Dots indicate temperature and O-ring damage for 24
successful launches prior to Challenger. Curve shows
increasing damage is related to cooler temperatures.
Used these charts

All information is there
- but very hard to identify and assimilate
- why?
Four seminal books

- standard literature for every visualization enthusiast

- taught information design at Princeton University
- now a professor at Yale University
Course Topics

CSE 332 Introduction to Visualization

Spatial Data
Non-Spatial Data
Data Mining
Display Technology
Insight
Perception & Cognition
Domain Knowledge
Visualization
Interaction
Knowledge
Large & Big Data
High Performance Computing
SPATIAL DATA

- shock wave
- virtual frog
- spiral flow
- nerve cell
- transparent MRI head
- wind flow
- semi-transparent tomato
- MRI head
SPATIAL DATA

- shock wave
- virtual frog
- nerve cell
- spiral flow
- transparent MRI head
- wind flow
- semi-transparent tomato
- MRI head
Example: Datasets obtained by 3D volumetric scans (CT, MRI)

- what are some questions you might have?
Example: Datasets obtained by 3D Simulations
- what are some questions you might have?

**Spatial Data**

**Hypothesis:** Matter clumps together and attracts more matter.
Spatial Data

Example: Data obtained by observation-supported simulations

- what are some questions you might have?
- One question might be: How did Hurricane Katrina evolve?
The salient features of a car:
- miles per gallon (MPG)
- top speed
- acceleration
- number of cylinders
- horsepower
- weight
- year
- country origin
- brand
- number of seats
- number of doors
- reliability (# of breakdowns)
- and so on...
How are MPG, weight, HP, and reliability related? Are there tradeoffs?

Which car is best for me?
What Is This?
Poll Results From Last Class
WHAT IS NEEDED FOR VISUALIZATION?

You answered:

imagination
data
information
creativity
Big Data

12+ TBs of tweet data every day

30 billion RFID tags today (1.3B in 2005)

100s of millions of GPS enabled devices sold annually

4.6 billion camera phones worldwide

25+ TBs of log data every day

76 million smart meters in 2009... 200M by 2014

2+ billion people on the Web by end 2011

http://www
VISUAL ANALYTICS VS. DATA SCIENCE

- Information Analytics
- Geospatial Analytics
- Scientific Analytics
- Statistical Analytics
- Cognitive and Perceptual Science
- Interaction
- Presentation, production, and dissemination
- Data Management & Knowledge Representation
- Knowledge Discovery

Data Science
- Data Engineering
- Scientific Method
- Domain Expertise
- Math
- Hacker Mindset
- Statistics
- Advanced Computing
- Visualization
MODERN DATA SCIENTIST

MATH & STATISTICS
- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression

PROGRAMMING & DATABASE
- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing packages, e.g., R
- Databases: SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query

DOMAIN KNOWLEDGE & SOFT SKILLS
- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION
- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau
Make decisions based on data

- not purely on intuition and long business experience
- use a combination of these
Visualization can be beautiful
VISUALIZATION CAN BE BEAUTIFUL
Visualization Can Be Interactive

D3 Demo

Data-Driven Documents
Visualization Has a Long History
Visualization Can be Inspired by Art
Visualization Can be Deceptive
Visualization Can be Deceptive
Visualization Can be Deceptive

Count the number of black dots
Visualization Can be Deceptive
Visualization Can be Deceitful

Are the horizontal lines parallel or do they slope?
Visualization Can be Deceptive

How many legs does this elephant have?
Visualization Can be Deceptive

Julian Beever
Which circle in the middle is bigger?
Visualization Can Be Deceptive

Gun deaths in Florida

Number of murders committed using firearms

2005
Florida enacted its ‘Stand Your Ground’ law

873
721

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014
The Power of The Visual System

The human visual system is not perfect, but it’s extremely powerful.

Vision is an integral part of life.

Vision is the gateway to higher-level regions of the brain.

Exploit this fast and powerful processor for
  ▪ complex data analyses, creative tasks, communicating ideas

→ The science of visualization and visual analytics
Text Books

Required

Optional
Tentative Schedule
<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topic</th>
<th>Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intro, schedule, and logistics</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Applications of visual analytics, data, and basic tasks</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Introduction to D3, basic vis techniques for non-spatial data</td>
<td>Project 1 out</td>
</tr>
<tr>
<td>4</td>
<td>Visual perception and cognition</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Data preparation and reduction</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Visual design and aesthetics</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Foundations of statistics</td>
<td>Project 2 out</td>
</tr>
<tr>
<td>8</td>
<td>Data types, notion of similarity and distance</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Data mining techniques: clusters, text, patterns, classifiers</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Data mining techniques: clusters, text, patterns, classifiers</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>High-dimensional data, dimensionality reduction</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Computer graphics and volume rendering</td>
<td>Project 3 out</td>
</tr>
<tr>
<td>13</td>
<td>Techniques to visualize spatial (3D) data</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Scientific and medical visualization</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Scientific and medical visualization</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Non-photorealistic rendering</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Principles of interaction</td>
<td>Project 4 out</td>
</tr>
<tr>
<td>19</td>
<td>Visual analytics and the visual sense making process</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Correlation and causal modeling</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Big data: data reduction, summarization</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Visualization of graphs and hierarchies</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Visualization of text data</td>
<td>Project 5 out</td>
</tr>
<tr>
<td>24</td>
<td>Visualization of time-varying and time-series data</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Memorable visualizations, visual embellishments</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Evaluation and user studies</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Narrative visualization and storytelling</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Data journalism</td>
<td></td>
</tr>
</tbody>
</table>
Midterm (1\textsuperscript{st} part of the course): 35%
Final (2\textsuperscript{nd} part of the course): 35%
Projects (5): 30%
\begin{itemize}
  \item propose a dataset DS and argue why you think it’s interesting (5%)
  \item code up a set of basic interactive D3.js visualizations for DS (5%)
  \item implement a set of suitable data analytics for DS (5%)
  \item interlude: create some spatial visualizations using ImageVis3D (5%)
  \item create an interactive D3.js visual analytics dashboard for DS (10%)
\end{itemize}
all projects are double-blind peer reviewed via web submission

Participation:
\begin{itemize}
  \item taken implicitly by ways of provided Socrative feedback
\end{itemize}

For late submission policy see course website
\begin{itemize}
  \item course website will publish all course materials
\end{itemize}