1. VISUAL ANALYTICS

INTRO

S. Rinzivillo – rinzivillo@isti.cnr.it
WHO I AM?

• Salvatore Rinzivillo
  • rinzivillo@isti.cnr.it

• Page course: http://didawiki.cli.di.unipi.it/
  • Visual Analytics

• Github page:
  • https://github.com/va602aa-master

• Telegram channel:
  • https://t.me/va602aa
SCHEDULE

• On Monday
  • 16:15 to 18:00
  • Fibonacci Room C

• On Tuesday
  • 14:15 to 16:00
  • Fibonacci Room A1
**GRADING**

- Project (50%)
  - Up to 2 persons per group (!)

- Project discussion (50%)

- Project topic
  - Multidimensional exploration of a dataset
  - One (or two) dataset(s) assigned for all
  - Specific proposal may be discussed
PROJECT FEATURES

• A project should have the following requirements:
  • The application should contain **multiple visual widgets**, each providing insights on a selection of dimensions of the original data
  • It is possible to use state-of-the-art charts (bar charts, line charts, etc.) and libraries (plotly, vega, etc). It is should implement a **novel, original visualization** to present the data in a creative, non-trivial way. (see examples on Vast Challenge 2008 developed in class)
  • **Interactivity** should be implemented, providing toolbars, selections and filters for the data.
  • The visual widget should interact among them, realising a set of **linked display** to browse the data across multiple dimensions
EXAMPLE: SCHOOL DISTRICTS

Where is the best place to send your child to school in New York City?

http://itisaasta.com/nycs/
Service starts at 5 AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch “jump over” the Ashmont branch.

Train frequency increases around 6:30 AM as morning rush hour begins.

Location of stations in blue. Orange and yellow dots represent mid stations/buses.

Trains are divided into different colors. See the evening return.

BOSTON SUBWAY SYSTEM

http://mbtaviz.github.io/
TEXTBOOKS

Design for Information
Isabel Meirelles

Visual Analytics for Data Scientists
Andrienko et al.

http://alignedleft.com/tutorials
INTERESTING READINGS

Information Visualization
Colin Ware

The Visual Display of Visual Information
Edward R. Tufte
DATA VISUALIZATION AND VISUAL ANALYTICS

INTRODUCTION
VA - CRASH COURSE

• Effective Visual Representation
  • Vision System
  • Visual Variables
• Toolbox – Bootstrap, Node.js, Vue.js, crossfilter.js
• Toolbox – Base visualizations (Plotly.js, Vega, DC.js)
• Toolbox – D3.js
  • Basics
  • Charts
  • Advanced Visualization
• Scientific Visualization
  • Plotting
  • Geography
• Storytelling
DATA VISUALIZATION

Convey Information through graphical representation of data
MOTIVATIONS

• Data everywhere
• No value for raw data
  • Need to extract valuable information
• Information overload:
  • Irrelevant for current task
  • Processed in an inappropriate way
  • Presented in an inappropriate way
VISUALIZATION GOAL

• Record Information
  • Sketches, photographs, ECG,…

• Analyze data to support decisions (exploration)
  • Create and verify hypotheses
  • Identify Patterns
  • Identify Outliers

• Communicate (explanation)
  • Share or highlight insights on data
  • Persuade
## ANALYZE: ANSCOMBE’S QUARTET - DATASETS

<table>
<thead>
<tr>
<th></th>
<th>Data Set A</th>
<th></th>
<th>Data Set B</th>
<th></th>
<th>Data Set C</th>
<th></th>
<th>Data Set D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X</strong></td>
<td></td>
<td><strong>X</strong></td>
<td></td>
<td><strong>X</strong></td>
<td></td>
<td><strong>X</strong></td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
<td>6.58</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>5.76</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
<td>7.71</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
<td>8.84</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
<td>8.47</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
<td>7.04</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
<td>5.25</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
<td>5.39</td>
<td>8.0</td>
<td>5.56</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
<td>7.91</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
<td>6.42</td>
<td>8.0</td>
<td>6.89</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
<td>5.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Analyze: Anscombe's Quartet - Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of $x$ in each case</td>
<td>9 (exact)</td>
</tr>
<tr>
<td>Sample variance of $x$ in each case</td>
<td>11 (exact)</td>
</tr>
<tr>
<td>Mean of $y$ in each case</td>
<td>7.50 (to 2 decimal places)</td>
</tr>
<tr>
<td>Sample variance of $y$ in each case</td>
<td>4.122 or 4.127 (to 3 decimal places)</td>
</tr>
<tr>
<td>Correlation between $x$ and $y$ in each case</td>
<td>0.816 (to 3 decimal places)</td>
</tr>
<tr>
<td>Linear regression line in each case</td>
<td>$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)</td>
</tr>
</tbody>
</table>
ANALYZE: ANSCOMBE'S QUARTET - GRAPHICS
COMMUNICATE: HIERARCHICAL STRUCTURES

http://www.stefanieposavec.co.uk/entangled-word-bank/

https://atlas.cid.harvard.edu/
COMMUNICATE: NETWORKS

https://www.flickr.com/photos/blprnt sets/72157614008027965/
https://atlas.cid.harvard.edu/
COMMUNICATE: TEMPORAL STRUCTURES

Cubism And Abstract Art (Alfred H. Barr 1936)

Over the Decades, How States Have Shifted

Recent elections have placed a heavy emphasis on "swing states"—Ohio, Florida and the other competitive states. Yet in the past, many more states shifted between the Democratic and Republican parties. A look at how the states shifted up to the year 2010 and how they have shifted over past elections.

COMMUNICATE: MAPS

https://www.flickr.com/photos/walkingsf/sets/72157624209158632/

COMMUNICATE: SPATIO-TEMPORAL DATA

"Minard" by Charles Minard (1781-1870) - see upload log. Licensed under Public Domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Minard.png#mediaviewer/File:Minard.png

Visual Analytics of Movement.
G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Wrobel
Springer, 2013
COMMUNICATE: TEXT
VISUALIZATION AND VISUAL ANALYTICS

• Make data and information processing transparent
• Combine strengths of humans and computers
VISUAL ANALYTICAL PROCESS

Exploration  Explanation

Adapted from:
Mastering the Information Age
Keim, Kohlhammer, Ellis, Mansmann
ELEMENTS OF GOOD VISUALIZATION

- Inviting Visualization
  - Rich content
  - Sophisticated Execution

Information Graphics
IMPORTANCE OF VALID DATA

Bad Data + Good Visualization = Bad Visualization
OTHER RESOURCES

Observe how others resolved design problems

datavisualization.ch

https://flowingdata.com/

infosthetics.com