Introduction to Data Visualization

Srijith Rajamohan

Advanced Research Computing

Monday 10th October, 2016
This week:

- Data Visualization Taxonomy
- 2D/Geospatial
- Temporal Data
- Tree Data
- N-dimensional Data
- Network Data
- Visualization tools
Section 1

1. Data Visualization Taxonomy

2. 2D/Geospatial

3. Temporal data

4. N-dimensional Data

5. Tree Data

6. Network Data

7. Visualization tools
Visualization can be broadly classified as: (Card et al.)

- **Scientific Visualization**
  - Use of interactive visual representation of scientific data to amplify cognition
  - Typically has spatial components

- **Information Visualization**
  - Use of interactive visual representations of abstract, non-physical based data to amplify cognition
  - No spatial components typically
Section 2

1. Data Visualization Taxonomy
2. 2D/Geospatial
3. Temporal data
4. N-dimensional Data
5. Tree Data
6. Network Data
7. Visualization tools
Choropleth

Prevalence of Food Stamp Use, by County (June 2009)
Dot distribution map
Proportional symbol map
Contour map/isopleth
Section 3

1. Data Visualization Taxonomy
2. 2D/Geospatial
3. Temporal data
4. N-dimensional Data
5. Tree Data
6. Network Data
7. Visualization tools
Timeline
Time series
Connected scatter plot
Gantt chart
Polar area

http://www.excelcharts.com/blog/the-consultants-chart-revisited/
Sankey diagram

- Smoke
- Alternator
- Friction
- Condensor
- Pump
- E_in
- E_out
Section 4

1. Data Visualization Taxonomy
2. 2D/Geospatial
3. Temporal data
4. N-dimensional Data
5. Tree Data
6. Network Data
7. Visualization tools
Pie chart

Colors from www.ColorBrewer.org by Cynthia A. Brewer, Penn State Univ.
Histogram

Heights of Black Cherry Trees

<table>
<thead>
<tr>
<th>Height (feet)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>60-65</td>
<td>2</td>
</tr>
<tr>
<td>65-70</td>
<td>4</td>
</tr>
<tr>
<td>70-75</td>
<td>8</td>
</tr>
<tr>
<td>75-80</td>
<td>10</td>
</tr>
<tr>
<td>80-85</td>
<td>2</td>
</tr>
<tr>
<td>85-90</td>
<td>1</td>
</tr>
</tbody>
</table>
State of the Union Address, 2002 vs. 2011

President Bush, January 29, 2002

President Obama, January 25, 2011
Introduction to Data Visualization

Srijith Rajamohan

Data Visualization Taxonomy

2D/Geospatial
Temporal data
N-dimensional Data
Tree Data
Network Data
Visualization tools

Bubble cloud
Tree map
Area chart
Parallel coordinates

- 2D/Geospatial
- Temporal data
- N-dimensional Data
- Tree Data
- Network Data
- Visualization tools
Waterfall chart
Section 5

1. Data Visualization Taxonomy
2. 2D/Geospatial
3. Temporal data
4. N-dimensional Data
5. Tree Data
6. Network Data
7. Visualization tools
Dendrogram
Section 6

1. Data Visualization Taxonomy
2. 2D/Geospatial
3. Temporal data
4. N-dimensional Data
5. Tree Data
6. Network Data
7. Visualization tools
# Introduction to Data Visualization

Srijith Rajamohan

## Data Visualization Taxonomy

- **2D/Geospatial**
- **Temporal data**
- **N-dimensional Data**
- **Tree Data**
- **Network Data**

## Visualization tools

- Matrix

---

![Matrix Diagram]
Node-link diagram
Section 7

1. Data Visualization Taxonomy
2. 2D/Geospatial
3. Temporal data
4. N-dimensional Data
5. Tree Data
6. Network Data
7. Visualization tools
Matplotlib

- Matplotlib is the entire library
- Pyplot - a module within Matplotlib that provides access to the underlying plotting library
- Pylab - a convenience module that combines the functionality of Pyplot with Numpy
- Pylab interface convenient for interactive plotting
Example

```python
>>> import pylab as pl
>>> pl.ion()
>>> pl.isinteractive()
False
>>> x = [1, 3, 7]
>>> pl.plot(x)  # if interactive mode is off use show() after the plot command
[<matplotlib.lines.Line2D object at 0x10437a190>]
>>> pl.savefig('fig_test.pdf', dpi=600, format='pdf')
>>> pl.show()
```
Simple Pylab plot
Example

```python
>>> X = np.linspace(-np.pi, np.pi, 256, endpoint=True)
>>> C, S = np.cos(X), np.sin(X)
# Plot cosine with a blue continuous line of width 1 (pixels)
>>> pl.plot(X, C, color="blue", linewidth=1.0, linestyle="-")
>>> pl.xlabel("X") ; pl.ylabel("Y")
>>> pl.title("Sine and Cosine waves")
# Plot sine with a green continuous line of width 1 (pixels)
>>> pl.plot(X, S, color="green", linewidth=1.0, linestyle="-")
>>> pl.show()
```
Example

```python
>>> pl.figure(figsize=(8, 6), dpi=80)
>>> pl.subplot(1, 2, 1)
>>> C, S = np.cos(X), np.sin(X)
>>> pl.plot(X, C, color="blue", linewidth =1.0, linestyle="-"")
>>> pl.subplot(1, 2, 2)
>>> pl.plot(X, S, color="green", linewidth =1.0, linestyle="-"")
>>> pl.show()
```
Pylab - subplots
Bokeh

- Bokeh is a Python interactive visualization library
- Intended for modern web browsers
- Intended for high-performance
Bokeh exposes three interface levels to users

- a low-level bokeh.models interface that provides the most flexibility to application developers.
- an intermediate-level bokeh.plotting interface centered around composing visual glyphs.
- a high-level bokeh.charts interface to build complex statistical plots quickly and simply
Please visit http://bokeh.pydata.org/en/latest/docs/gallery.html for more examples