Creating visualizations using Linked Data

Alvaro Graves
gravea3@rpi.edu
@alvarograves
Agenda

• Visualizations based on Linked Data
  • Why they are useful?
  • Challenges
• What is Visualbox
  • Examples
Linked Data Visualizations

- Why governments should publish Linked Data?
- Structured data on the Web
- Flexible to model multiple domains (health, education, budgets)
- Open, machine-friendly standards to publish data
Linked Data Visualizations

- Publishing is not enough!
  - How do we make easier for users to consume it?
- How do we explore the data?
- How can we discover trends?
- How can we find errors or outliers?
Linked Data Visualizations

• Why visualizations are useful?
  • In many cases, visualizations make it easy to consume data
  • It helps to discover and explore trends in a simply manner
  • Errors that otherwise would be hard to find, can be found easily
Detect anomalies

Attraction in Berlin?

http://visualbox.org/demos/AttractionsInBerlin
Discover trends
Challenges

• Using visualization tools is hard
  • Many technologies
    • JavaScript, Google Maps, SVG, D3.js, etc.
  • Steep learning curve
• It’s not easy to create rapid prototypes
Visualbox

• Environment for creating LD visualizations

• Provides multiple visualizations types
  • Maps, graphs, charts, timelines

• Simple creation process

• Trivial to share (just pass a URL)
  • http://myserver/myVisualization
How do you create a visualization in Visualbox?

Step 1: Create a SPARQL query

```
SELECT DISTINCT ?country ?population WHERE{
  ?country ex:hasPopulation ?population
}
```

Step 2: Add a visualization filter

```
{{models.main|GoogleVizColumnChart:"country, population"}}
```
Example: Trends in UK school system

Step 1: Create a SPARQL query (called main)

```sparql
SELECT DISTINCT (SUM(?c) AS ?schoolCapacity)  
    (COUNT(?school) AS ?schoolsPerDistrict)  
    (SUM(?c)/COUNT(?school) as ?ratio)  
    (MAX(?districtl) AS ?l)
WHERE{
    ?school    school:schoolCapacity        ?c ;  
    school:districtAdministrative         ?district.
}
GROUP BY ?district
ORDER BY ?schoolCapacity
```

Step 2: Add a visualization filter

```
{{models.main|GoogleVizColumnChart:"l,schoolsPerDistrict"}}
```
Number of schools per district

Birmingham, Stockport, Sunderland, Tameside, Greenwich, Bury, East Lindsey, Wealden, South Cambridgeshire and ... Horsham, Merton, West Oxfordshire, Thurrock, Easington, Arun, Essex, Eastleigh, Lincoln, Bromsgrove, Penworth, Rochford, Rutland, Swansea.

{{models.main|GoogleVizColumnChart:"l,schoolsPerDistrict"}}
Average capacity per school in each administrative district
Average school size vs. number of schools per district

{{models.main|GoogleVizColumnChart:"districtLabel,schoolCapacity"}}
Average school size vs. number of schools per district
Further questions

• What if we have
  • Average scores in Math, Language, etc.?
  • Poverty level of each area?
  • Other social metrics

• Compare to other countries?

• Linked Data makes it easy to query and mix data

• Visualbox makes it easy to visualize it
More info

Visualbox:
http://visualbox.org

Demos:
http://visualbox.org/demos

Github:
https://github.com/alangrafu/visualbox

Visualbox Virtual Machine: