



Creation of visualizations based on Linked Data

Alvaro Graves

gravea3@rpi.edu - [@alvarograves](https://twitter.com/alvarograves)

Tetherless World Constellation
Rensselaer Polytechnic Institute

Agenda

- **Challenges in Linked Data usage**
- **How to create an LOD visualization?**
- **User study**
- **Conclusions**



Challenges in Linked Data usage

Good and bad news about Linked Data

- **Good:**
 - Billions of triples available on the Web
 - Lots of information about people, events, places, etc.
- **Not so good:**
 - Not a lot people outside the Semantic Web community find it usable

Who are the users?

- People with domain expertise, not technical
- Journalists
- Scientists (Bio-, Chemo-, X-Informaticians)
- Novice Semantic Web students
- People who want an easy/rapid way to understand certain aspects of LOD datasets

Challenges

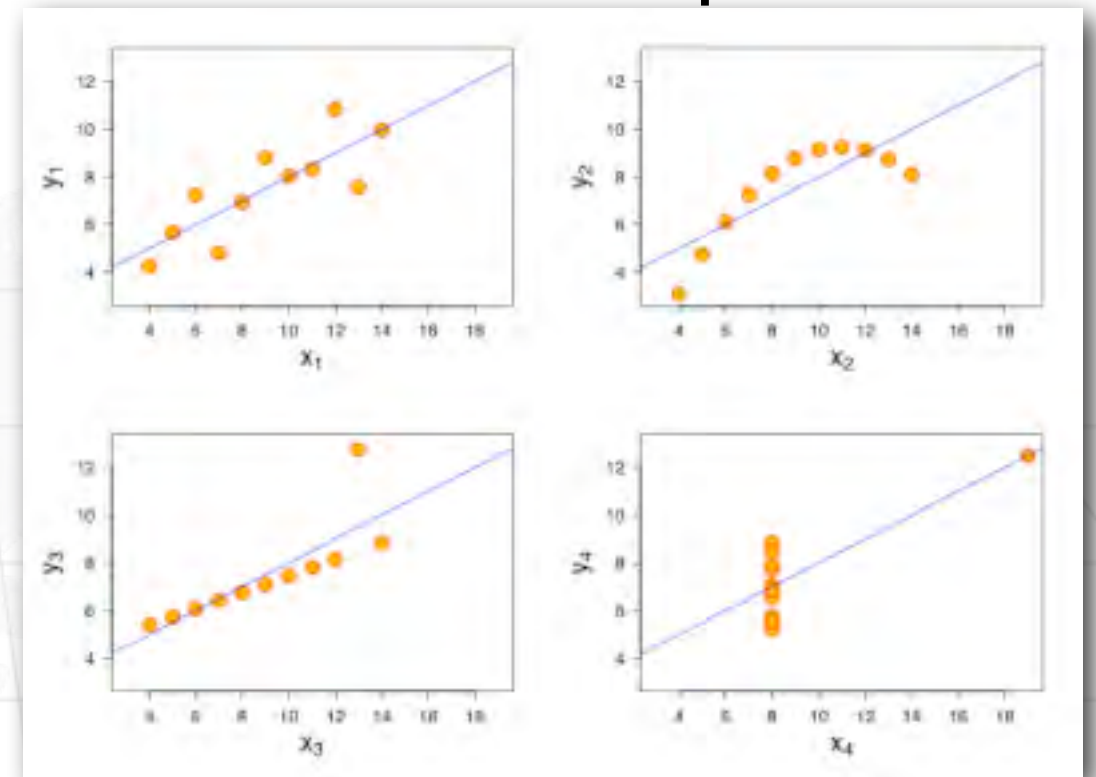
- What is needed for non-technical users to consume and explore Linked Data?
- How can we empower domain experts that lack experience with Linked Data?

Idea: Use of visualizations can simplify consumption of Linked Data

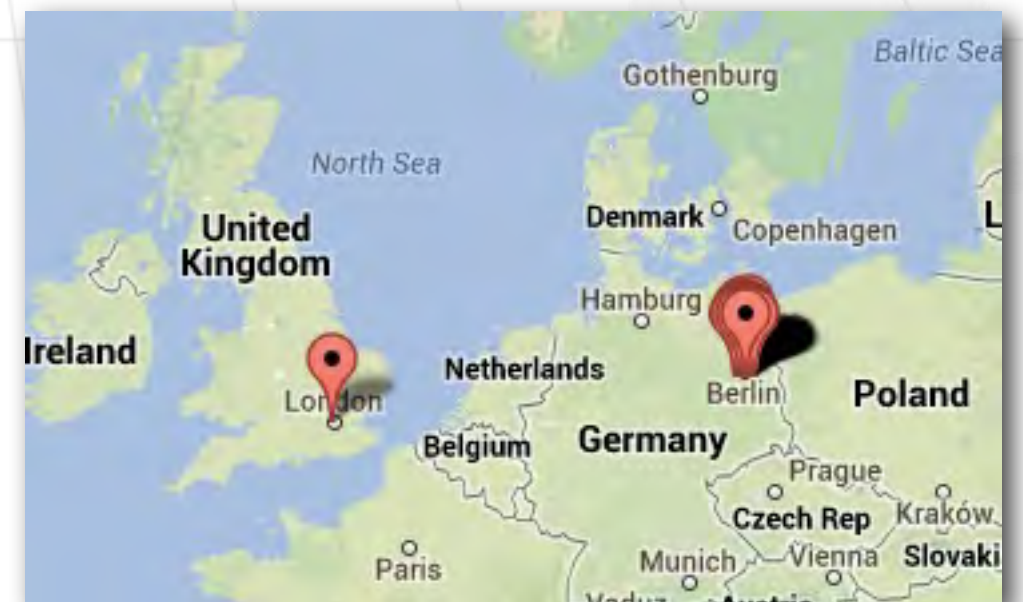
Why Visualizations?

- Visual inspection of data is helpful
- Can help to detect trends and outliers
- It is easy to share and consume for non data experts

Anscombe's quartet



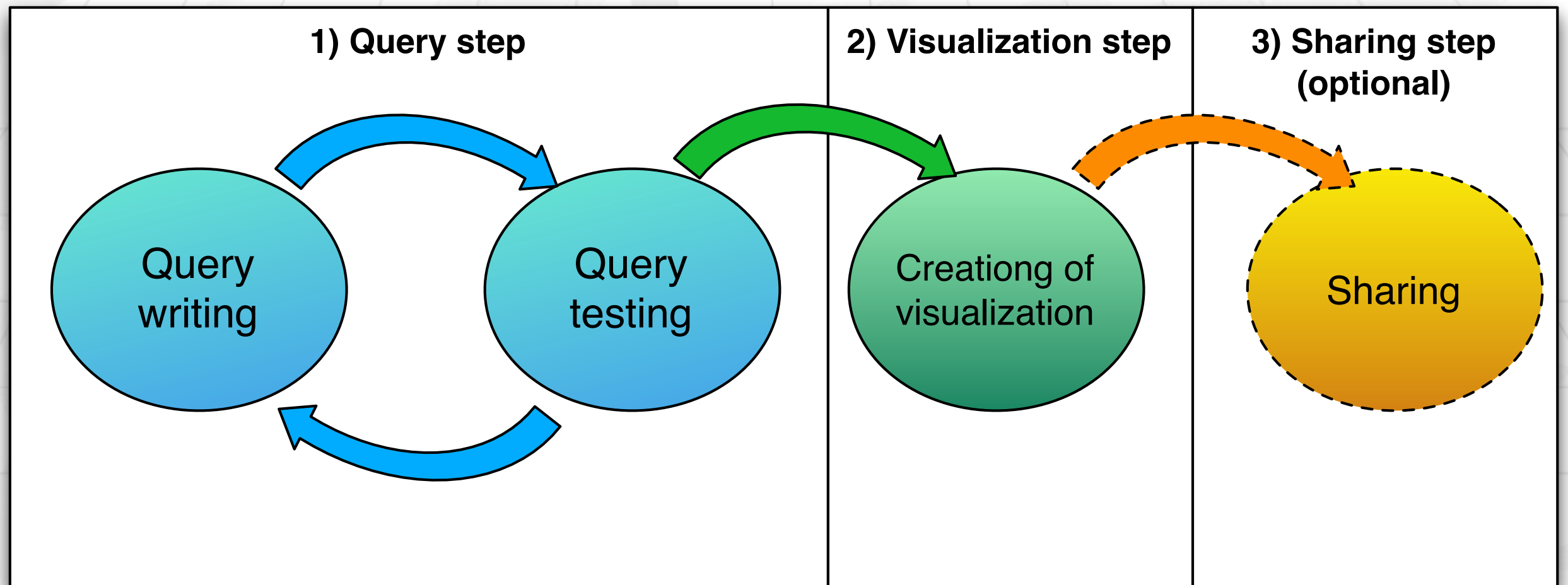
“Attractions in Berlin”





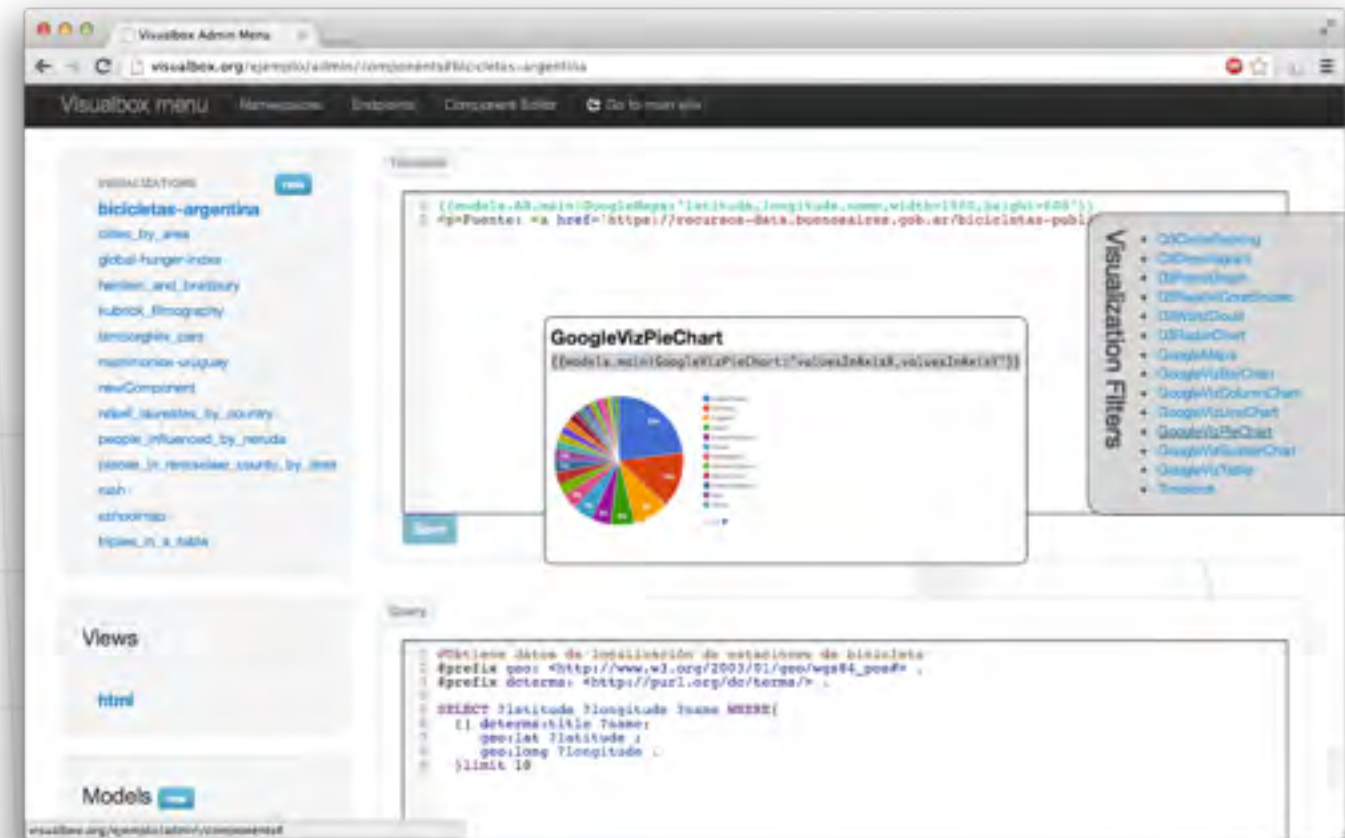
How to create an LOD visualization?

Visualization workflow



Visualbox

- “Easy Linked Data visualizations”
- Integrated environment to create visualizations based on LOD
- Makes use of available SPARQL endpoints
- Provides 14 types of visualizations out-of-the-box



Visualbox

The screenshot shows the Visualbox web application interface. It features a top navigation bar with 'Visualbox menu', 'Namespaces', 'Endpoints', 'Component Editor', and 'Go to main site'. On the left, there are three panels: 'Visualizations' with a list of visualization types and a 'new' button (labeled 1), 'Views' with a 'html' button, and 'Models' with a 'main' button. The main area is divided into two sections: 'Visualization Editor' (labeled 3) and 'Query Editor' (labeled 2). The 'Visualization Editor' contains a code editor with a query: `{{models.main|GoogleMaps:"latitude,longitude,capital,width=700,height=700"}}`. The 'Query Editor' contains a query:

```
1 PREFIX dcterms: <http://purl.org/dc/terms/>
2 PREFIX dbp: <http://dbpedia.org/property/>
3 PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
4
5 SELECT * WHERE {
6   ?country dcterms:subject <http://dbpedia.org/resource/Category:South
7   dbp:capital ?capital .
8   ?capital geo:lat ?latitude ;
9   geo:long ?longitude .
10 }
11
```

 Below the query editor are buttons for 'Save', 'Test the query against', and 'local (http://dbpedia.org/local)'. At the bottom, there is a 'Query results preview' section (labeled 6) with columns for 'country', 'capital', 'latitude', and 'longitude'. On the right side of the interface, there is a 'Visualization Filters' panel (labeled 4) and a 'View visualization' button (labeled 7). A large red arrow points from the interface towards the map on the right.



- 1 - Existing visualizations
- 2 - Query editor
- 3 - Visualization Editor
- 4 - Visualization helper

- 5 - Buttons for query testing
- 6 - Query testing results
- 7 - Sharing buttons

Mozilla Festival

- Tested Visualbox against real users
- Most of them never heard about Linked Data



Feedback from Mozilla Festival

- **Positive:** Most people evaluated it as a great experience for learning about Linked Data and visualizations
- **Negative:** Most people found hard to even start working with SPARQL
 - Unsure about the syntax
 - Unsure “what’s in there” in an endpoint



TetherlessWorld



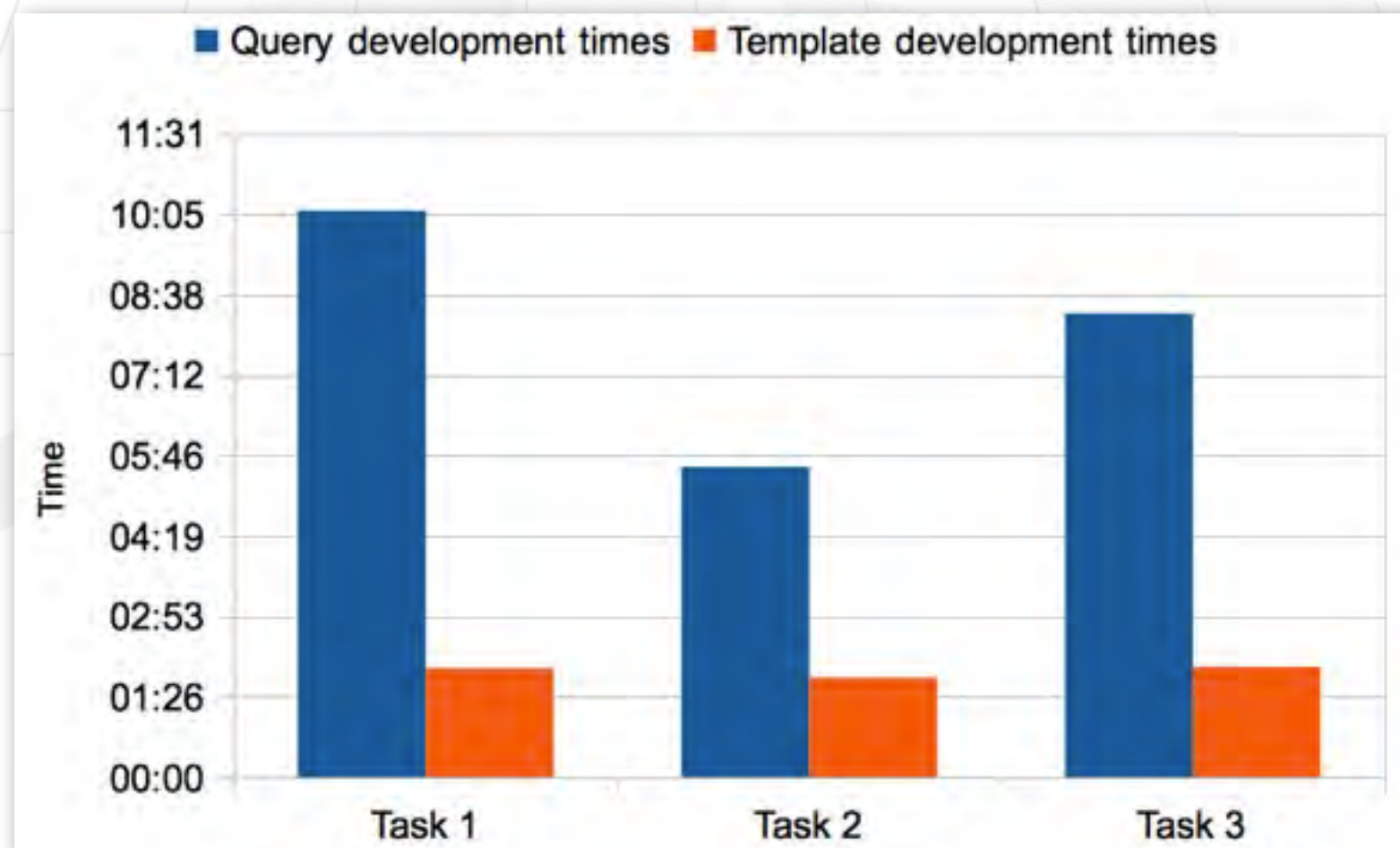
User study

User study

- People with basic to medium knowledge of SPARQL and Linked Data
- Participants watched an introductory video
- Used *think aloud* approach
- Three tasks, visualizations based on DBpedia
- Ex-post subjective evaluation

Results

- Most of the time (5-10 min.) was focused on creating and debugging SPARQL query
- Once the query was ready, the rest of the process was very short (1-2 min.)



Participant's feedback

- Positive: Visualbox makes it easy to create visualizations
 - *Much easier than doing it from scratch*
 - Add features to power users
- Negative: Creation of SPARQL query
 - One size does not fit all

Possible solutions

- Syntax assistance
 - Namespaces and prefixes
 - Improved syntax highlighting
- Recommendations/suggestions based on:
 - Endpoint's content
 - What others have queried
 - Other existing visualizations

Conclusions

- Creating LOD visualizations is a two-step process:
 - Query creation
 - Visualization development
- Visualbox provides an integrated environment for the first step
- It reduces the effort for the second step
- There are open challenges in helping users to create SPARQL queries