Evaluation of LinDA and Comparison with other tools for visualization of Linked Data

User Manual

Authors:

Rose-Mary Owusuaa Mensah

Nishananth Baskaran

Ugochukwu Chimbo Ejikeme

Nikhil Patra

Mentor:

Dr. Fabrizio Orlandi

Chapter

1

Introduction

1.1. Overview

The goal of this project is to provide an evaluation and comparison report about the LinDA visualization tool. The evaluation of LinDA is mainly the usability testing. Comparison of LinDA is with the other available visualization tools. For comparing some parameters are chosen in common between LinDA and other tools. The reports will help the LinDA developers to know how users think about the tool. It will also help them to know the pros and cons of LinDA compared to other visualization tools already available.

1.2. Definitions, Acronyms, Abbreviations

- SME: Small and Medium Enterprises.
- LinDA: Linked Data Analytics.
- RDF: Resource Description framework is a family of World Wide Web Consortium (W3C) specifications used as a general method for conceptual description or modelling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats.
- Evaluators: A group of people who are involved in understanding the product quality which are: usability, efficiency and scalability and compare the new tool with the already existing tool.
- Users: A group of people who are involved in experiencing the tool and give a feedback on the same.

1.3. About the project

The project mainly aims in evaluating the new visualization tool called LinDA. There are some visualization tools available already and they can be used mostly by technical people. LinDA also focuses on people who are novice. In order to see, whether the developers of LinDA have achieved what they want this evaluation is being carried out. The developers wanted to see how the users feel about the interface. Whether the user is able to access the interface without any assistance and whether he is able to visualize the data are the main interest for the developers of LinDA. In addition to this, tasks like how long a user takes to generate visualization, how long the tool takes to generate the visualization will help to know the efficiency of the tool. Finally, the comparison of LinDA with the other visualization tools will help to know the pros and cons of LinDA visualization tool. To summarize, this

project mainly focus on the product quality which is usability, efficiency, scalability and comparison with other tools.

1.4. Architecture

The architecture of this evaluation system has two modules or blocks. The one block is called the evaluation module which has blocks related to the evaluation activities. The other block is the software module. This module gives the working of the LinDA software and the analysis of the evaluation results. Both the modules will be interacted by evaluators and users. Below diagram shows the architecture diagram of the evaluation system.

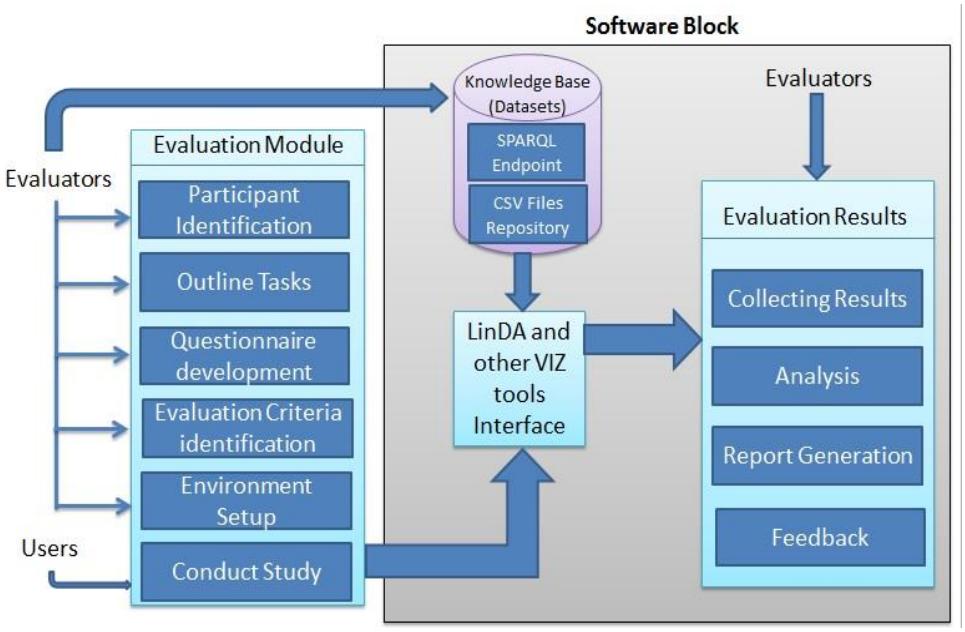

Figure 1: Architecture Diagram

In the evaluation module, the evaluators are associated with the following blocks: participant identification, outline tasks, questionnaire development, evaluation criteria identification and environment setup. With the name of the block, the activities behind the block are understandable. Users are associated to the conduct study block of the evaluation module. This conduct study block is associated to the LinDA software in the software module. The users will perform the tasks in the LinDA software and will provide a feedback and rating for the LinDA software to the evaluators. Later, these feedbacks from the users will be analysed by the evaluators and will generate reports from it. These blocks are present in the software module and they are associated with the evaluators again. At the end, the evaluators will provide a feedback on LinDA based on the evaluations. One more block which is present in the software module is the comparison of other visualization tools with LinDA. From this evaluation also, evaluators will analyse and provide reports about LinDA when compared to other visualization tools which are already available based on linked data.

Chapter 2

LinDA Visualization Tool

2.1. About LinDA

The aim of the LinDA project is to make the benefits of Linked Open Data accessible to SMEs and data providers by providing libraries for Open Data consumption. One of the main tasks in this context is to build an ecosystem of tools for visualizing Linked Data to assist SMEs in their daily tasks by hiding complexity through automation and an intuitive user interface. To complete this task, a generic visualization workflow is being implemented based on state-of-the-art Linked Data visualization approaches. Most existing approaches are only usable by a technical audience or limited to certain domains or data representations. LinDA proposes a generic approach for visualization selection in form of a faceted browser that imposes on the user the task of describing the visualization at an unfamiliar level of abstraction.

LinDA Visualization Workflow:

- *Select Data*: The user starts with the selection of the dataset which is intended to be visualized.
- *Select Visualization*: Based on the content and format of the selected dataset, the possible visualizations are computed and presented to the user.
- *Configure Visualization*: After choosing the necessary visualization, the user needs to provide necessary input for the application in order to map the data for the visualization selected.
- *Visualize*: Finally, the input is performed and the visualization of the data happens. Along with this, the user can export or share the visualization.

Chapter

3

Basic Configuration

3.1. Technological Specifications

The following are the software requirements for LinDA visualization tool:

- Git
- Nodejs
- Virtuoso 7
- Compass (Ruby)
- LinDA
- Ubuntu

3.2. Installing Git

The git can be installed by the following command in Ubuntu:

sudo apt-get install git-core

Git is mainly used to download the LinDA software from the GitHub repository.

3.3. Installing Nodejs

The following commands are used to install Nodejs in Ubuntu:

sudo apt-get update

sudo apt-get install nodejs

sudo apt-get install npm

npm install –g nodemon

3.4. Installing Virtuoso 7

- Virtuoso 7 can be downloaded from the below link:
  http://sourceforge.net/projects/virtuoso/files/latest/download?source=files

- Navigate to the downloaded folder.
- Extract the package.
- Run the following command in Ubuntu:
  sudo apt-get install libssl-dev

- Go the extracted package location. It is better to run autogen.sh by typing ./autogen.sh, which checks for the presence and right version of some of the required components, and if it reports any missing package then, install that package.
- Set following environmental variable by typing
  CFLAGS=”-O2”
  export CFLAGS
- Execute ‘./configure’ command to configure the package for your system.
- Execute ‘sudo make’ command to compile the package.
- Type `sudo make install’ to install the programs and any data files and documentation.
- Start the Virtuoso Server by the following two steps:
  1. Go to ‘cd /usr/local/virtuoso-opensource/var/lib/virtuoso/db’
  2. Then execute the command ‘sudo /usr/local/virtuoso-opensource/bin/virtuoso-t -f &’.

3.5. Installing Compass (Ruby)

For LinDA Ruby 2 and Compass 0.12.3 are necessary. Old versions of Ruby and Compass may create problems. Hence it is better to remove them first before installing the required versions. These can be removed with the following commands:

ruby -v

compass –v

which compass

which ruby

sudo apt-get remove ruby_version

Now install Ruby Version manager and Compass:

rvm install 2.0.0
rvm use 2.0.0 (as root/sudo)
gem install compass --version="0.12.3"

3.6. Installing LinDA Visualization

The LinDA Visualization tool can be installed in Ubuntu by the following commands:

git clone https://github.com/LinDA-tools/visualisation.git

cd visualisation/frontend

npm install

bower install

cd ../backend

npm install

The LinDA can be started by the following commands:

nodemon &
sudo service virtuoso-opensource-6.1 start
cd *visualisation/frontend*
grunt serve

Chapter 4

Alternate Visualization Tools

4.1. Tableau Public

Tableau Public is a free service that lets anyone publish interactive data to the web. Once on the web, anyone can interact with the data, download it, or create their own visualizations of it. No programming skills are required. Be sure to look at the gallery to see some of the things people have been doing with it. Tableau Public includes a free desktop product that you can download and use to publish interactive data visualizations to the web. The Tableau Public desktop saves work to the Tableau Public web servers nothing is saved locally on your computer. All data saved to Tableau Public will be accessible by everyone on the internet, so be sure to work only with publically available (and appropriate) data.

4.2. OntoWiki CubeViz

Since the amount of available statistical data in the web is increasing in the past years, user-friendly accessibility to such data is becoming equally important. There are also numerous statistical datasets published within the Data Portal by the European Commission. In dependence of the data format of these datasets, the containing data can be browsed by users with the help of different tools such as Google spreadsheet, Libre Office or MS Office. One of the upcoming data formats is the Resource Description Format (RDF) which is designed to represent data not on a document/text level but on the level of information and facts. Furthermore RDF enables publishers to encode not only data items but also underlying semantics. In order to represent such semantics, domain specific vocabularies can be reused.

The RDF DataCube vocabulary is one of the vocabularies used to encode statistics. There is an increasing amount of tools dealing with RDF in order to facilitate users to browse datasets encoded in this way. One of these tools is OntoWiki that facilitates users to browse RDF data in a generic way. CubeViz was developed based on *OntoWiki* with the scope of offering user-friendly exploration possibilities for statistical data represented in RDF with the RDF DataCube vocabulary. CubeViz represents the statistical dataset to be visualized as a facetted based browsing component. This component enables the users to select interesting parts of the dataset. After selection the user can proceed while clicking on the button *Update Selection / Update Chart*. One can instantly start using CubeViz while clicking on the button Start CubeViz above. As a result CubeViz processes a chart according to user's selection. The current version of CubeViz processes basic chart types - such as *line, bar* and *pie chart* facilitating the exploration of up to two statistical dimensions in a data structure.

4.3. CODE: Query and VIZ Wizard

The amount of Linked Data available on the Web is growing continually, due largely to an influx of new data from research and open government activities. However, it is still quite difficult to directly access this wealth of semantically enriched data without having in-depth knowledge of semantic technologies. Therefore, one of the goals of the EU-funded CODE project has been to develop a web-based visual analytics platform that enables non-expert users to easily perform exploration and analysis tasks on Linked Data. CODE's vision is to establish a toolchain for the extraction of knowledge encapsulated in scientific research papers along with its release as Linked Data. A web-based visual analytics interface should empower the end user to analyse, integrate, and organize the data. The CODE Query Wizard and the CODE Vis Wizard fulfill this role.

4.4. LodLive

LodLive is an experimental project that was set-up to spread and promote the linked-open data philosophy and to create a tool that can be used for connecting RDF browser capabilities with the effectiveness of data graph representation. LodLive is the first navigator to use RDF resources based solely on SPARQL endpoints. LodLive aims to demonstrate how resources published by W3C standards for the semantic web can be made easily accessible and legible with a few viable tools. LodLive is capable of connecting the resources existing in its configured endpoints, allowing the user to pass from one endpoint to another by making use of LOD interconnection capacities. It is a web based application and provides visualization for linked open data. It has options to use numerous SPARQL endpoints like DBpedia and freebase. Also provides the option to use other resource address. It simply asks to select the resource and with a simple start button you can start working on the dataset.

4.5. RDF: SynopsViz

RDF: SynopsViz is a tool for hierarchical charting and visual exploration of RDF & Linked Open Data. Hierarchical RDF exploration is based on the creation of multiple levels of hierarchically related groups of resources based on the values of one or more properties. The adopted hierarchical model provides effective information abstraction and summarization. Also, it allows efficient -on the fly- statistic computations, using aggregations over the hierarchy levels.

4.6. Lumira

Lumira is a Visualization tool from SAP. Lumira helps to create stunning and interactive visualizations. There is a rich library available for visualization types. The library includes: scatter plots, heat and geo maps to tag clouds, box plots and network charts. We can share our findings with others, both internal and external, through engaging infographics and story board.

4.7. Supported File Formats, Download Link and License Information

The below table gives the information about the file formats supported by each tools, the download links for the tools, the license information for the tools if any and also link to the academic paper if any. This will help in getting the overall picture of the tools and help us in comparing the tools.

P.S. If the tool does not have information then the value entered is NA.

Tool Name	File Formats Supported	License Information	Academic Paper Information	Download Link
Tableau Public	MS Excel, MS Access and multiple text file formats. Limit 1 million rows and 1GB of data.	Free Version available for Students. For commercial purpose needs a license.	NA	http://www.tableausoftware.com/public/download
OntoWiki CubeViz	RDF	Free Software	NA	http://cubeviz.aksw.org/
CODE: Query and VIZ Wizard	RDF	Free Software	Yes	http://code.know-center.tugraz.at/search
LodLive	RDF	Free Software	NA	http://en.lodlive.it/
RDF:Synops Viz	RDF/XML, N-Triple, Turtle, N3	Free Software	NA	http://83.212.97.83:8084/
Lumira	MS Excel, CSV	Licence	NA	http://saplumira.com

Table 1: Other Visualization Tools Overview

Chapter 5

Datasets Information

5.1. List of Datasets

1. Elvis Impersonator dataset
   Description: Extensive metadata about 81 Elvis impersonators (from a movie), some with scary video clips.

   Type: rdf

   Source: http://www.rdfdata.org/

2. National Science Foundation RKB Explorer
   Description: 11,822,283 triples put together by National Science Foundation

   Type: rdf

   Source: http://datahub.io/dataset/rkb-explorer-nsf/resource/fcf9502f-b9c5-4544-8052-440b9440076b?inner_span=True

3. LinDA Visualization tool , Sample datasets
   Type: rdf, csv

   Source: https://github.com/LinDA-tools/Visualization/tree/master/backend/testsets

4. US Security Data
   Description: About 1.8 million triples, corporate ownership dataset on US Security

   Type: rdf

   Source: http://www.w3.org/wiki/DataSetRDFDumps

5. World Bank Linked data
   Description: It contains World Development Indicators, World Bank Finances, World Bank Projects and Operations, and World Bank Climate Change data.

   Type: rdf

   Source: http://www.w3.org/wiki/DataSetRDFDumps

6. Causes of death in France 2001 to 2008

Description: Causes of death in France from 2001-2008. Variables include year, gender, cause of death, and number of deaths.
Type: csv
Source: http://perso.telecom-paristech.fr/~eagan/class/as2013/inf229/labs/datasets

7. Sample insurance portfolio
   Description: The sample insurance file contains 36,634 records in Florida for 2012 from a sample company that implemented an agressive growth plan in 2012.
   Type: csv

   Source: http://support.spatialkey.com/spatialkey-sample-csv-data/

Chapter 6

Evaluation of LinDA

6.1. Usability Testing

Participants:

20 users including students, researchers and working professionals were participated in the usability testing. Students were from different field of study mainly from computer science, physics, biology and the like. Among 20 users, there were 16 students, 2 researchers and 1 working professional. Students were from Master degree and Bachelor degree. All the users had experience in using computers and well versed in using web browsers. Many users had experience in other visualization tools before. They had used it for their study or for work purpose. At the end of the usability testing, we gave small gifts to the users as a token of appreciation for participating in the testing activity.

Apparatus:

The experiment was conducted in a laptop computer with the following configurations:

- Processor : Intel Core i5 CPU @ 2.50GHz
- RAM : 4GB
- Operating System : Windows 7
- System Type : 64-bit Operating System
- VM Software : Oracle VirtualBox
- VM Processor : 1 CPU
- VM RAM : 4GB
- VM Video Memory : 64MB

Tasks:

In order for the user to evaluate the LinDA tool, we just formulated tasks which will help the user to go through the features available on LinDA. This will also help them feel how easy or difficult LinDA is to use. Also, they will be aware whether the user interface of LinDA is guiding them in a correct way to achieve the tasks or misleading them. We have created a background scenario for the users who are participating. Based on the background scenario, the tasks had been set up. Below are the background scenario and the tasks that were involved in the task sheets handed over to the users.

Background Scenario

You are Perry Johnson, a sales representative working for PC4U selling computer. In your company you are responsible for creating dashboards of sales results.

You are at your office and have opened your dashboard software.

Task 1

It's Monday morning, you get to the office at 9:00 am. As a part of your daily task, you want to review sales figures.
LinDA has started; you want to choose the appropriate dataset.

Task 2

You want to create a chart showing the gross profit over time (month). What is the 50th data entry in the selected data?

Task 3

You want to see the profit trend over the months as bar diagram. How has the profit been in the last 4 months of the data we have?

Task 4

You want to add one more column called brand for a new visualization. Keep the already selected two columns in addition to the new column.

Task 5

You want to see the brand wise contribution to profit as pie chart. Which brands give the highest and lowest profit? You want to change the label display of "gross profit" to "profit in dollars"

Task 6

Now you want to access the visualization in future. You want to share these visualizations with your manager who does not use LinDA. Save the current visualization as an image file.

Task 7

Visualize another dataset of your own and save the chart.

The tasks from 1 to 6 will help the user to know all the functionalities that are present in the LinDA visualization tool. By the end of task 6, the user will be able to select the dataset and can visualize the columns he need in particular. So, in the task 7 we made user to select any dataset of his own and then visualize anything of his like.

Measurements:

We tried to measure the design level of the user interface of LinDA, the difficulty level of usage of LinDA and the overall impression of the LinDA. There were follow up questions which we had asked to the user at the end of each task. These follow up questions are mainly about the design of the user interface for that particular task and the difficulty level of that particular task. This helped us to identify how the users feel about the functionality of using LinDA tool. In the end, there was post-test survey. In that, the overall experience, the likes about LinDA, the dislikes about LinDA, missing functionalities and many more were addressed. In order to measure all these things, we had a grading system for each tasks. The grading system is from 1 to 7. 1 is the lowest and 7 is the highest. The results of these measurements can be found in the next section.

Results:

Impression on design of LinDA based on tasks:

	Remarks	Good	Average	Bad
Number of people	**Task 1**	11	9	
	Task 2	5	13	2
	Task 3	6	10	4
	Task 4	9	8	3
	Task 5	11	5	4
	Task 6	10	9	1

Table 2: Design of LinDA based on tasks overview

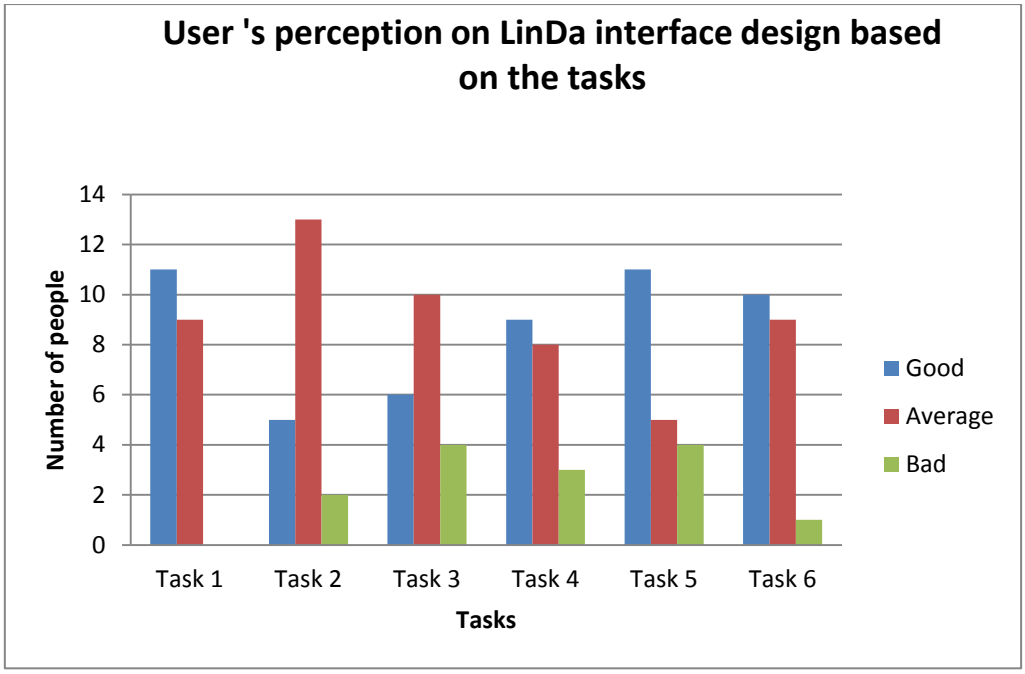

Figure 2: Bar Diagram showing design impression based on tasks

Overall Impression about the design of LinDA:

Design remarks	Good	Average	Bad
Scores	52	54	14

Table 3: Overall Design of LinDA overview

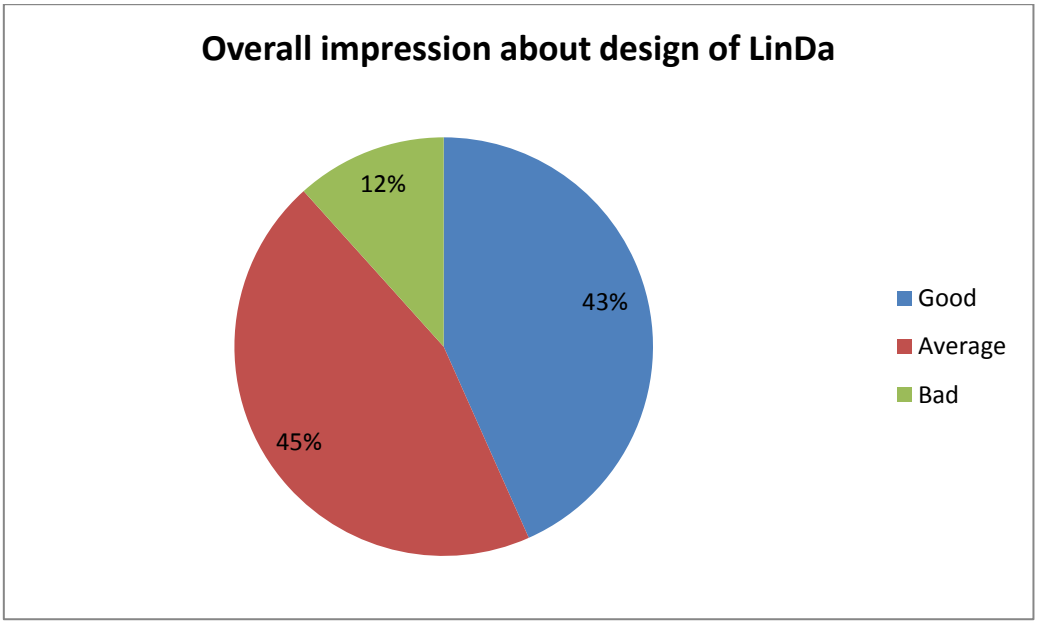

Figure 3: Pie Chart showing overall impression on design

Level of Difficulty based on tasks:

	Remarks	Easy	Medium	Difficult
Number of people	**Task 1**	11	9	
	Task 2	12	7	1
	Task 3	8	8	4
	Task 4	14	4	2
	Task 5	10	7	3
	Task 6	12	6	2
	Task 7	14	5	1

Table 4: Level of Difficulty of LinDA based on tasks overview

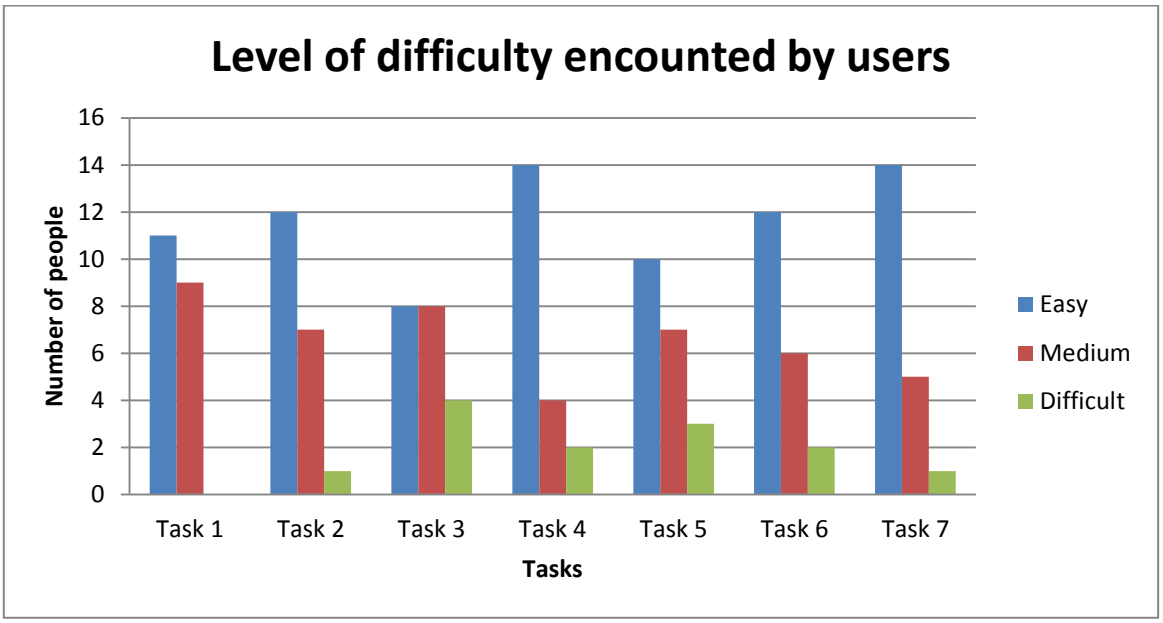

Figure 4: Bar Diagram showing Level of Difficulty based on Tasks

Overall Level of Difficulty:

Remarks	Easy	Medium	Difficult
Scores	81	46	13

Table 5: Overall Level of Difficulty of LinDA based overview

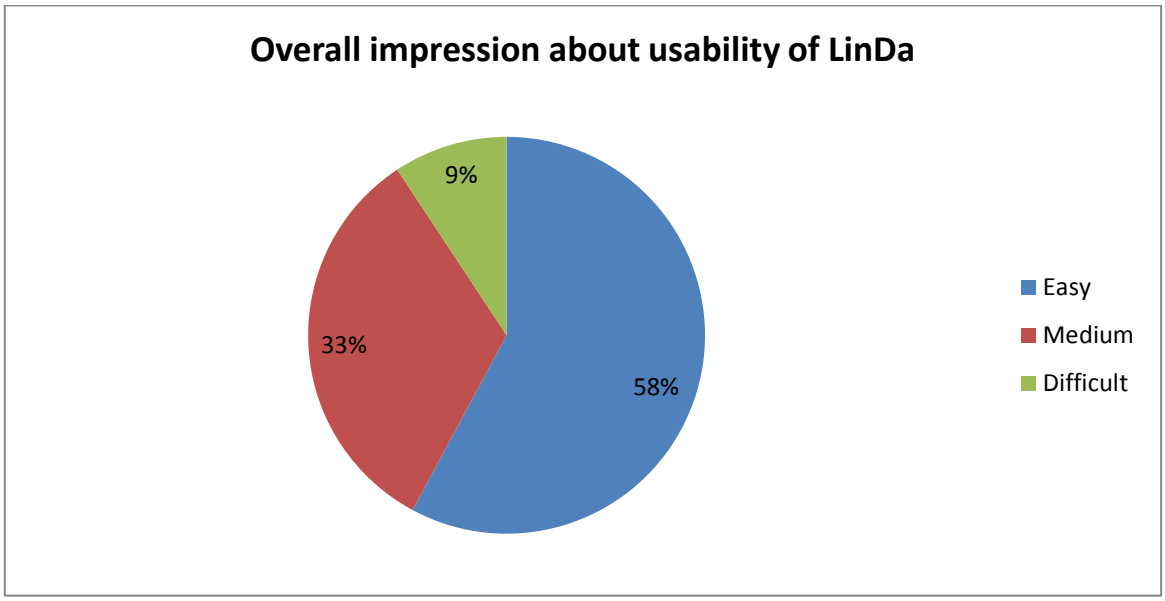

Figure 5: Pie chart showing overall level of difficulty

Satisfaction of Users with LinDA:

Remarks	Satisfied	Average
Number of users	7	12

Table 6: Satisfaction of users with LinDA overview

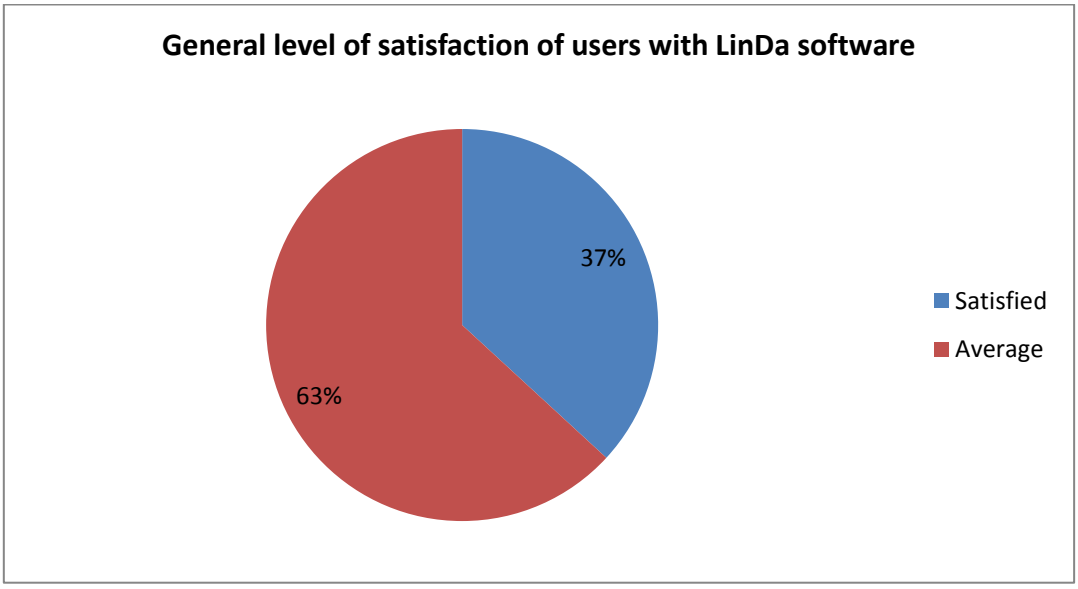

Figure 6: Pie chart showing satisfaction of users with LinDA

Improvements or Suggestions from Users:

The users were exited to carry out the tasks which were provided to them. They also did not hesitate to answer all the questions which were asked in the form of feedback. In every tasks, we had asked why they feel this task is easy or difficult. In addition to this, many users were very interested to give comments on suggestions or improvements for every task. Some of those are as follows:

As a part of usability testing, we imagined the user in a situation and asked them to select the appropriate dataset based on the situation. All the users had selected the correct dataset. However, they feel if the datasets are displayed in order or group in one name will be more helpful when there are thousands of datasets to be displayed in the LinDA tool homepage. In addition to this if the font size is bigger; it will help the user to identify the datasets more easily. When the user goes to the next page, he can see the columns that are present in the dataset. With the help of the expand button he can see the columns of the dataset. There is a check box in front of each column in order to select those columns for visualizing the same. Users felt if the columns names were arranged in alphabetical order then it will be easy for them to identify the things they need. Also, they suggested check boxes in before the column names that can be expanded misleading for the selection. Because, when we select those columns they don't have any data.

Many users were satisfied with the preview section of the selected columns. While some felt it will be nice to improve the design of the preview section as it was kind of boring to look into the data. And they also felt, if they wanted to see some 43rd data entry they need to count in order find the 43rd data entry. So adding a serial number for the data displayed will help to access to any number. When it comes to the visualization, users were surprised to see the selection of the charts based on the columns they had selected. They were able to see only few charts were enabled and others were disabled. Since there was no selection mark in the available charts users tried to select it once again even after the chart had been drawn. Because of this, users felt that there should be some selection color or some design change in order for them to identify which chart has been selected. One task for the users is to add a new row in addition to the already selected previous rows. Many users tried to press back button in order to navigate to the column selection page. However they took time to identify the button in the visualization page which takes them to the column selection page. Because of this users felt that the design of this button can be much better in order to identify it in the first look rather than searching for it.

When it comes to the visualized chart many users were satisfied with the image what they had obtained. Few users who were using some other visualization tool compared the charts with the charts of their used visualization tools. They felt they should be able to change the design of the charts and they should be able to select the colours of the chart. This is because if they can select the colours then they can differentiate between the slices of the chart. The users also felt that the axis label name was not visible if the data labels were long. As a part of a task, we asked the users to download the chart for future purpose. Some users got confused with the Export and Save Settings. They selected the file type and started clicking the button in the Save Settings section. They wanted these two can be differentiable

and also suggested it would be very nice if they can enter the name of the file that can be exported. Also, the users feel if the downloaded image is transparent then it will be very easy for them to navigate in the future.

Best of LinDA according to users:

In our wrap up questions to the users we asked them which feature or design of LinDA you like the most. We found that the users gave us the honest answers and they are below:

- Simple User Interface which helps to understand most of the features.
- Usability increases with time. Users gained more experience after the first try.
- Algorithm that selects the chart types based on the columns selection.
- Available of different chart types including the standard and advanced ones.
- Providing information while navigating the chart with the mouse pointer.
- Clear Visualization of data.
- Downloading of the chart.
- Drag and drop feature for visualizing.
- Search in the preview section.

Least likes of LinDA according to users:

We also asked the users the least feature or design you like in LinDA. The users answered for that as well and they are as follows:

- Arrangement of datasets.
- Choosing columns sometimes misleading because of the checkbox and expand buttons for the same column.
- Visualize button design can be enhanced more.
- Export section and Save Settings section are misleading.
- Scaling of the axis manually should be possible.
- Design enhancement for the preview data section.
- Downloaded chart as image file was not presentable.
- No Tool tips available.
- Unable to select different colours for charts.

6.2. Installability Testing

Installability testing as the name suggests it is used to test installation of the tool. The testing involves whether the user can successfully install the tool or whether he has any difficulties in the installation of the tool. In this section we will describe how we installed the LinDA visualization tool. The prerequisites needed for the LinDA installation has been mentioned clearly in the GitHub repository page of the LinDA Visualization tool except for the Virtuoso server version. The procedure and the command to install the dependencies have been mentioned well. It is a step by step instruction which made us to understand the procedure clearly.

However, we experienced the problem while installing Virtuoso server. We were able to install Virtuoso Server 6 without any problem. And we were able to start LinDA Visualization tool without any problem. However, later we identified the datasets are displayed in the LinDA tool homepage but when clicked the columns of the datasets were not loading. This is because Virtuoso server 6 doesn't support SPARQL endpoint which is necessary to load the columns in the web interface of the tool. Later from the mentors we came to know that installing Virtuoso server 7 will solve this problem.

In the GitHub repository there are no steps given for installing Virtuoso server 7 as it is different from installing Virtuoso server 6. However, our mentor sent us the steps for installing Virtuoso server 7. We successfully installed the Virtuoso server 7. However, the same problem existed. The columns of the datasets were not loading. This problem was new to our mentors as well and could not solve it. Finally, we got the working image file of the LinDA Visualization tool and we imported the image file in the Oracle VirtualBox. Then we found the LinDA visualization tool was working perfectly fine and able to see the dataset columns and visualize the selected columns.

As a conclusion from our Installability testing, we could only partially install the LinDA Visualization tool. We also recommend mentioning the steps more in detail and some workarounds if some error occurs while installing the LinDA visualization tool. It will also be nice to add some screenshots of the installation so that new users can install the tool easily and can be sure the steps they have followed is correct.

6.3. Effectivity Testing

Effectivity Testing is testing mainly the files supported by LinDA. The files here are referred to as datasets. In order to visualize, the datasets has to be loaded in the virtuoso server. Then these datasets will be available in the web interface of the tool for the purpose of visualization. LinDA visualization tool is mainly based on the RDF data. RDF is mainly designed as a metadata data model. The RDF datasets will contain columns with the values as well as the metadata information. The user can use the columns which has data for the visualization purpose. In LinDA, there are a number of RDF files loaded as examples. These RDF files are datasets. The visualization works fine with these RDF files by selecting the necessary columns.

In order to make LinDA visualization to be used by other users who are not familiar with RDF datasets, it is made to support CSV files as well. These CSV files have data in the form of tables. This is one of the common ways of representing the data by many people. There is a dataset available in CSV file in LinDA Visualization tool. This dataset was also used in the usability testing. This dataset is more generic and every user was able to associate the background scenario to this dataset. By this way, it helped the user to go through the functionality of the LinDA tool and to experience the user interface of the LinDA tool in much detailed manner.

We collected sample datasets from various sources, uploaded them into virtuoso and tried to visualise them with LinDA software. The table below describes the results obtained from trying to visualise the datasets. Partial visualisation used in the table implies that the

visualisation LinDA gave was not readable or interpretable. This could be as a result of a badly organised dataset or the chart recommendation algorithm from LinDA.

Dataset format	Name	Domain	Size	Data preview (sorting into columns and showing a preview)	Successful Visualization
.rdf	Elvis impersonators	Movies	118KB	Yes	No
.rdf	NSF projects standard grant	Finance	16KB	Yes	Yes
.nt	Service expenditure	Finance	364.3MB	Yes	Yes
.n3	US Data Sec	Corporate ownership	57.8MB	Yes	Yes (partial)
.ttl	World bank	Finance, operations	52.5KB	Yes	Yes (partial)
.csv	Causes of death in France 2001 to 2008	Health	100.1KB	Yes	Yes
.csv	Sample Insurance portfolio	Insurance	4.1MB	Yes	Yes

Table 7: Effectivity testing overview

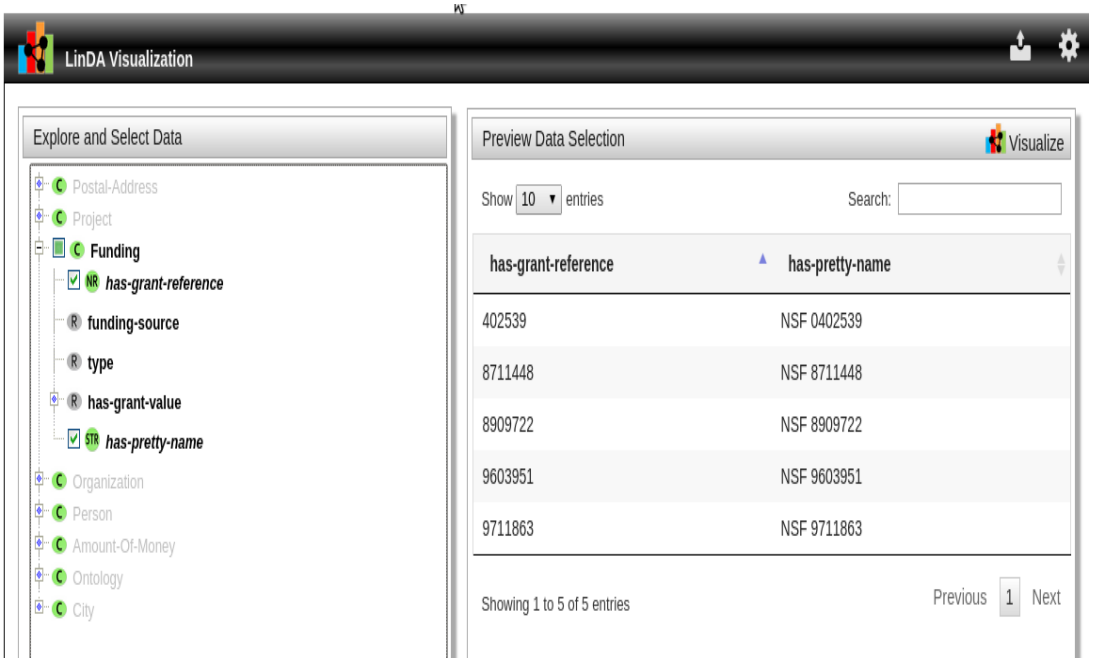

Figure 7: NFS Standard Grant dataset loaded in LinDA and its preview section

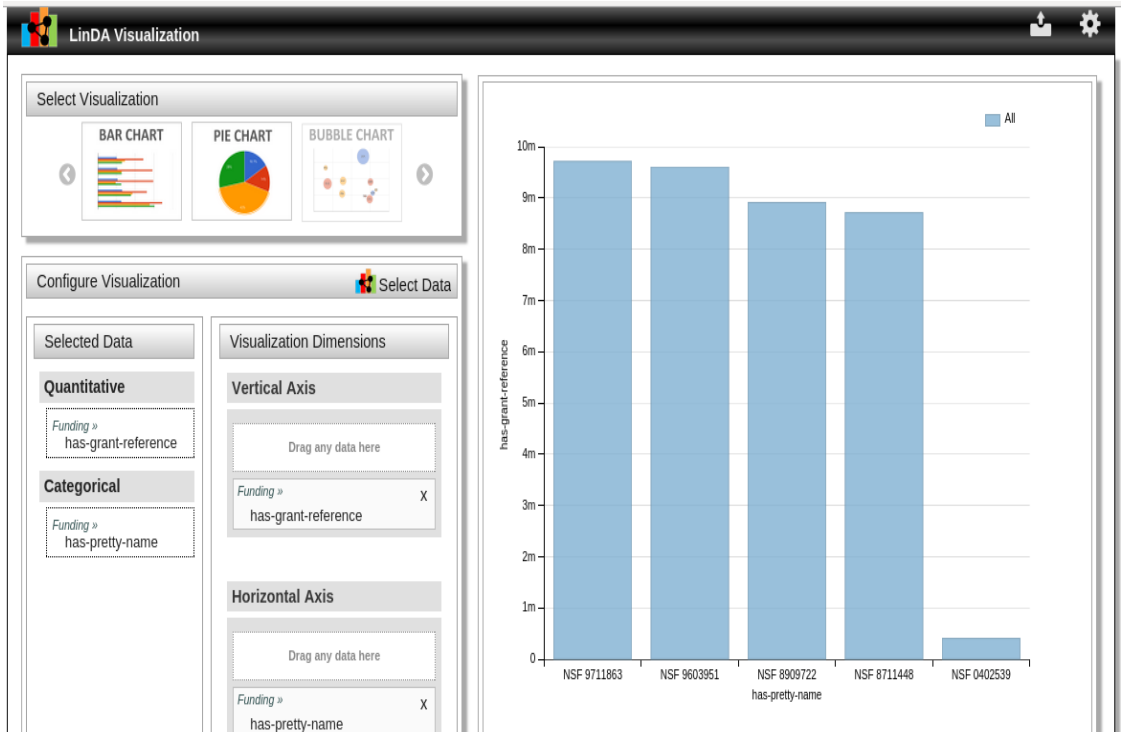

Figure 8: NFS Standard Grant dataset visualized as bar chat in LinDA

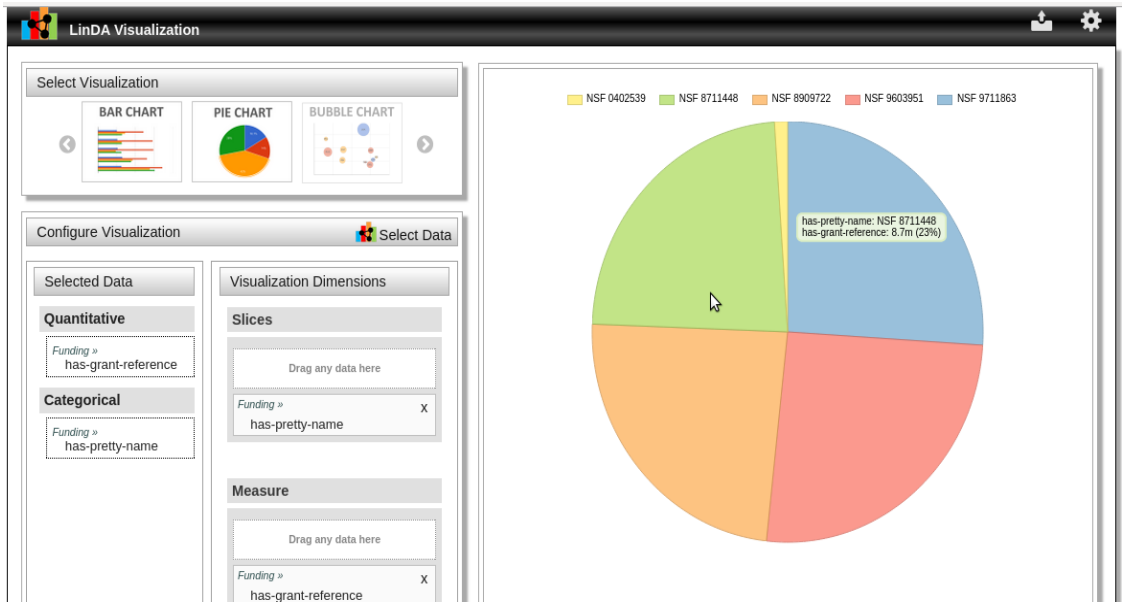

Figure 9: NFS Standard Grant dataset visualized as pie chat in LinDA

As a conclusion, LinDA Visualization tool can successfully visualise datasets of file formats .rdf, .n3, .ttl, .nt, .csv. It is also observed that LinDA software is able to better visualise data with numerical values as compared with those with text and symbols from other domains.

6.4. Scalability and Efficiency Testing

In order to examine LinDA's ability to handle increasing sizes of datasets, we gathered some datasets whose sizes ranged from 2.4MB to 364.3MB. We setup virtual machines on two systems (a personal computer and a workstation), uploaded the datasets and visualised in LinDA. Described below are the characteristics of the computers used and the results obtained.

	Virtual machine (VM) A	**Virtual Machine (VM) B**
Host	**Personal computer**	**EIS Workstation**
Memory of host computer	4096MB	16384
Memory of virtual machine	2885MB	7642

Table 8: Characteristics of computers used

	Dataset size	Load dataset in Virtuoso	Data preview	Visualisation (Chart suggestion)	Visualisation (Chart display)
VM A	2.4 MB	✓	✓	✓	✓
VM B		✓	✓	✓	✓
VM A	364.3MB	✓	✓	✓	x
VM B		✓	✓	✓	✓

Table 9: Scalability and Efficiency testing overview

Results from the table above and the observations made during the testing indicates that, for large datasets, the software may not perform very well on computers with less memory capacity as compared to systems with higher specifications, hence not very efficient. Sometimes, the software could preview the data and even suggest charts but it is unable to display the charts due to insufficient system requirements. See the two figures below for a pictorial description on the two interesting scenarios (chart suggestion without display and chart suggestion with display). We therefore suggest that, developers should have another look at their algorithms and optimise them in order to make the system run efficiently on personal computers as these are the most common systems available in various places.

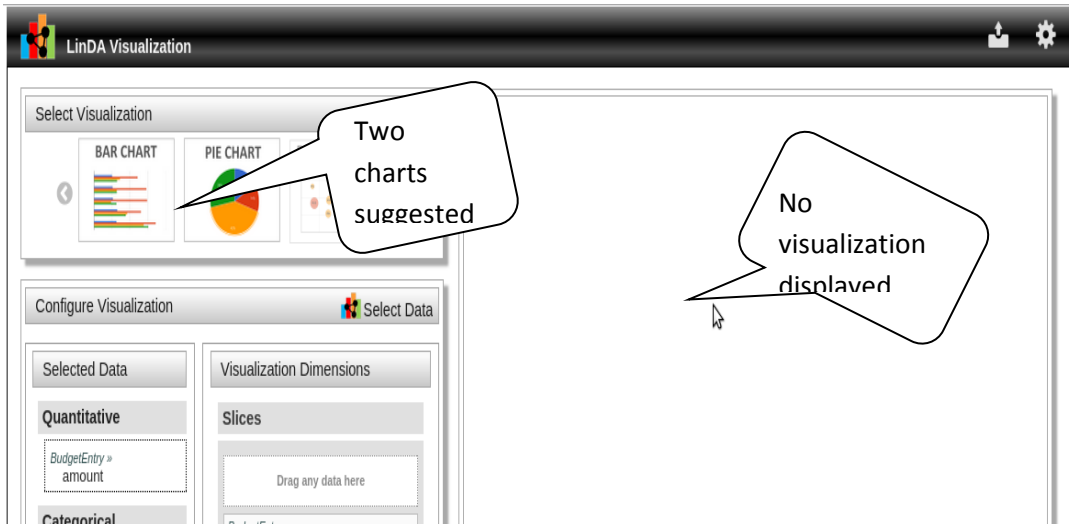

Figure 10: Incomplete Scenario. Charts are suggested but not displayed due to system memory limit

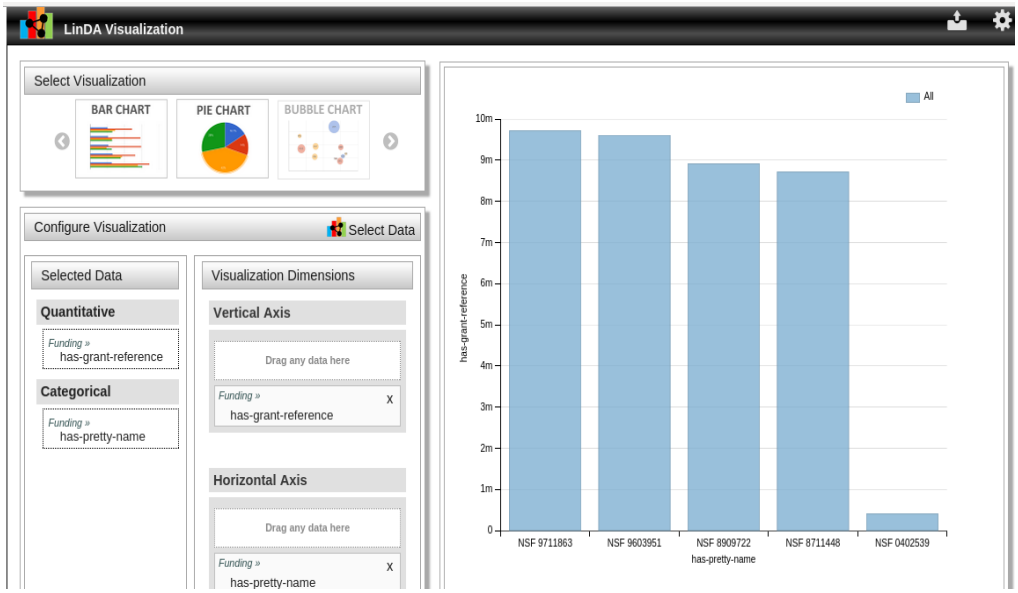

Figure 11: Complete data visualization in the machine where it has full capacity to run the software

6.5. Comparison with other Visualization Tools

The below comparison table shows the comparison of the other Visualization tools with the LinDA Visualization tool. The comparison is mainly done with the features of the tools. The features are mainly classified into four main types. They are Data Access, User Interface, Usage/Operation and Licensing. A short description of those features is explained after the comparison table.

								
		LinDA	**Tableau Public**	**CODE**	**Cube Viz**	**LodLive**	**RDF:Synopsviz**	**Lumira**
Data Access	**Local data files**	Yes	Yes	No	Yes	No	Yes	Yes
	Customization	No	Yes	No	No	No	No	Yes
	Data overview	Yes	Yes	Yes	Yes	No	Yes	Yes
	Detail on demand	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Online data sources	No	Yes	Yes	Yes	Yes	Yes	No
	Multi Table Access	No	Yes	Yes	No	No	No	Yes
	Faceted search	No	Yes	Yes	Yes	No	Yes	Yes
	In-memory engine	Server	Desktop, Server	Server	Server	Server		Server, Desktop
	Data Access Model	Menu-based	Menu Based	Menu Based, Wizard	Menu Based	Menu Based	Menu-Based	Menu Based, Wizard
	Supported File Format	RDF, CSV	CSV,XLS,TXT	RDF	RDF,XLS,CSV	RDF	RDF/XML, N-Tripple, Turtle, N3	CSV, TXT, EXCEL,LOG, PRN
User Interface	**Visual Representation**	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Drag-and-Drop	Partial	Yes	No	Yes	Yes	No	Yes
	Standard Visualizations	Yes	Yes	Yes	Yes	Graph only	Yes	Yes
	Advanced Visualizations	Maps	Maps, Circle, Gantt, Tree	Maps, Stream Graph	Polar Chart	Maps	Timeline, TreeMap	Map, Heat map, Tree,
	Filtering	No	Yes	Yes	Yes	No	Yes	Yes
	Geo-Mapping	Yes	Yes	Yes	No	Yes	No	Yes
Usage / Operation	**Dashboard**	No	Yes	No	No	No	No	Yes
	Run from Desktop	No	Yes	No	No	No	No	Yes
	Run from Server	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	User/Group security	No	No	No	No	No	No	Yes

Licensing	**License**	Free Software	Limited	Free Software	Limited	Free Software	Free Software	Limited

Table 10: Comparison of Tools

Feature Name	Description
Local Data Files	Ability to import our own datasets into the tool.
Customization	Ability to change the parameters based on user needs.
Data Overview	Ability to provide high level overview on the selected dataset.
Detail on Demand	Ability to provide a preview of the selected columns.
Online Data Access	Ability to use the datasets that are available online.
Multi Table Access	Ability to access multiple datasets at the same time for visualization.
Faceted Search	Ability to filter the columns based on the user needs and then visualizing the same.
In-Memory Engine	Whether the tool is desktop or server based.
Data Access Model	Whether the tool is Menu or Wizard based.
Supported File Formats	Describes what are the file formats the tool is supported.
Visual Representation	The use of images, pictures and animations to aid understanding of the system.
Drag and Drop	Ability to drag and drop the columns or some other thing the user needs.
Standard Visualizations	Availability of standard visualization charts.
Advanced Visualizations	Availability of advanced visualization charts.
Filtering	Enables sorting of items according to a specified criteria.
Geo Mapping	Availability of maps chart for the geographical dataset.
Dashboard	Availability of Dashboard.
Run from Desktop	Tool is runnable like a Desktop application.
Run from Server	Tool is runnable like a web based application.
User/Group security	Security for a user or a group
License	About the license information of the tool.

Table 11: Comparison parameters description