

3-D modelling of multidimensional & disparate data in visual analytics

Oshin Rawlley

Dept. of Computer Science & Engineering
Amity University, Noida
Uttar Pradesh, India
oshinrawlley1@gmail.com

Dr. Shailendra Narayan Singh

Dept. of Computer Science & Engineering
Amity University, Noida
Uttar Pradesh, India
snsingh36@amity.edu

Abstract—Great amount of emerging data is creating many opportunities with data analysis, visual analytics etc on the top. Here I have explored the angle of how can we put big data into use. Big data can turn into trash if no significant information can be extracted out. So it needs to be presented & visualized in a manner that visuals can milk out most of it. Already in my previous accepted paper, I have discussed the conventional & novel approaches of data visualization & also the future scope of it. In this paper I am going to work upon the future scope proposed by me in my previous paper i.e. “interactive visualization”. Only blindly showing a diagram sometimes may not be useful as it doesn’t reveal much of the facts. A proper analyzable pattern or an insight can be concluded if the visuals interact with the user. Also it is important that one should choose the right chart to the data to be analyzed. A set of right visualizations therefore helps in seeing various angles of the same topic & hence deriving various trends & results unknown to us. So a set of visuals used commonly & in latest corporate world has also been shown in this paper. An attempt has been made to propose a new feature of storytelling in this paper which is discussed in brief.

Keywords—*story telling, holographs, interactive visualization.*

I. INTRODUCTION

Behind big statistics how do we understand the story behind it. Data will increase exponentially. It increase 40% compound annual rate peaking nearly 45 ZB by 2020. In 2007 data has tripled the amount of it generated in 2007. Consumers are new generators of data by updating adding ,every 60 sec data is generate, a fundamental premise. Discovering trends & drawing insights from the data is what is being done in all the organizations & in disparate fields. Here interactive visualization comes into play where it is needed to make sense out of data. There are many visual analytics like SAS, R, Python which are so fast that you start from interrogating questions about the data set to

getting into an actual chit-chat with the set. The best possible way to create visual include SAS Visual Analytics which implements intelligent self charting based on selection of data. For visualizing results for whatever you are including an analyst or a business user, or a scientist either one of a statistician they have to share the above. For complete sense of data, visualization comes into play. Discovering of data in an interactive way lets you look at every angle to explore quickly the happenings & critical things get identified. The prime relations of clusters formed are highlighted, outliers detected, springing trends and much more are found. It guides you to insights which are critical in nature & also inspires action. Dazzling reports are created in quick fashion. Then they are shared through phones & applications of Microsoft.

II. MOTIVATION

Descriptive & predictive analyses are the two kinds through which we comprehend about the past & the future respectively. To know the happenings in the past we need a descriptive analysis of an organization, about the trends & patterns of its growth, sales, profit & loss statements. But to predict a particular growth or insight of a result or predicting future results of the company one needs to do a predictive analysis. This can be best done by interactive visuals.

III. NEW TECHNIQUES OF INTERACTIVITY

A. *3-d holographs spatio temporal visualization*

The way of interacting is changing tremendously[1]. Many revolutionised techniques have come up which does the interaction with PCs, starting from mouse, keyboard, 2-D touch to multi touch & rising up to the level of 3-D touch and gesturing. We are

having holographic views with a screen topped with 3-d views. We can interact with the views by touching them zooming & panning. This type of dynamic interaction helps in viewing the models at different angles. We can make the model sit above or below the screen virtually in accordance of the need to be justified. It is easier to gain insights in these holographic views as not that in 2-d views[2].

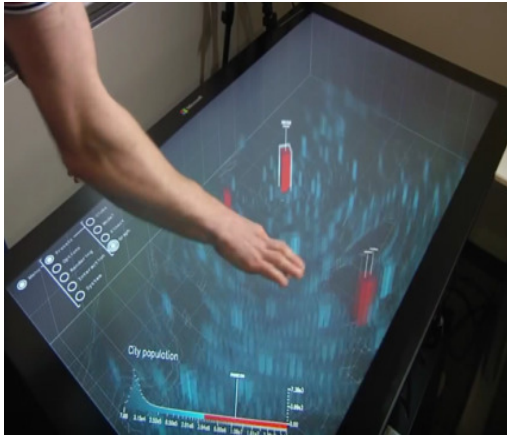


Fig.1 City population in 3-d image can be zoomed marked etc.

- B. Also recordings using 3-dimensional can be done using connect in which a virtual view of the recording can be shown.



Fig.2 Holographic image of a moving saint in a 3-d space

- C. The other ways with which the data can be represented are additional glyphs including lines markers, patches.

Virtually every field is turning to big data for which the challenge is to mine these vast stores of information to inform decision making, to improve medical diagnosis & otherwise to answer others questions & desires of tomorrow's society ways that are unimagined today.

IV. METHODOLOGY

We used python as a data visualization tool. Pandas also termed as python analysing of data library is also an open source library. Common python library for data science is pandas. It is meant for tabular data sets & time series. It provides performance which is very high, data structures which are easier to use and analysing tools of data for programming in Python language. Tools are provided for the purpose of read-write of data between data structures which are in-memory and in different formats: txt, CSV, databases like SQL, Microsoft Excel, and the fast HDF5 format and also alignment of data in an intelligent way, also handling of missing data: attaining alignment which is self labelled in computations and it easily manipulates scattered data into an ordered way which is done in python.

There are various data visualization libraries in python some are listed like:-

- Altair- declarative statistical visualization library
- Bokeh- interactive web plotting for python
- Matplotlib- A 2-d python library

We used matplotlib & bokeh library to show the comparisons & various visualization ways between these 2 libraries. For that we installed ANACONDA PYTHON DISTRIBUTION. It's a world's most popular & trusted data science ecosystem.

V. IMPLEMENTATION

First we implemented the visualization using matplotlib library of python.[15] Here only 2-dimensional visuals can be produced like lines, histograms etc. Also customization can be done by labelling the axes, adding titles.

```
In [6]: import matplotlib.pyplot as plt
year=[1950, 1951,1952,2100]
pop=[2.538,2.57,2.62,10.85]
year=[1800,1850,1900]+year
pop=[1.2,1.262,1.650]+pop
plt.plot(year,pop)
plt.xlabel('year')
plt.ylabel('population')
plt.title('world stats')
plt.xticks([0,2,4,6,8,10],['0','20','40','60','80','100'])
plt.show()
```

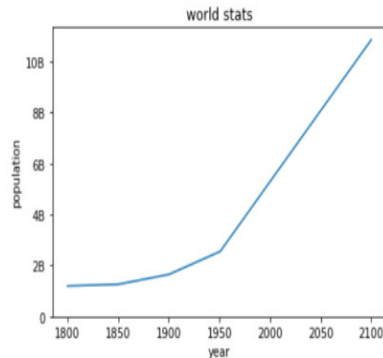


Fig.3 Customized Matplotlib code of depicting a line graph

Since only 2-d figures can only be implemented with this library, in this paper we chose a 3-dimensional enabled library as follows.

BOKEH

The other library we use is the bokeh library. A bokeh plotting interface is used where glyphs are adopted for visualization. Glyphs are visual shapes like circles, triangles, squares, rectangles, lines, wedges etc. They have properties attached to data like:-

- Coordinates(x,y)
- Size,color,transparency

A. Data structure used in bokeh: Column data source

- It is a commonly used data structure which maps string column names to sequences of data.
- It is often created automatically for you.
- It can be shared between glyphs to link selections
- Tooltips like hover are used with extra columns.

B. Import Column data source from bokeh models

Now we create an instance of column data source by passing a data dictionary to the initializer. This dictionary should have string keys & sequences such as arrays or lists as their values.

All columns in column data source must be of the same length.

```
In [1]: from bokeh.models import ColumnDataSource

In [2]: source = ColumnDataSource(data={
...:     'x': [1,2,3,4,5],
...:     'y': [8,6,5,2,3]})

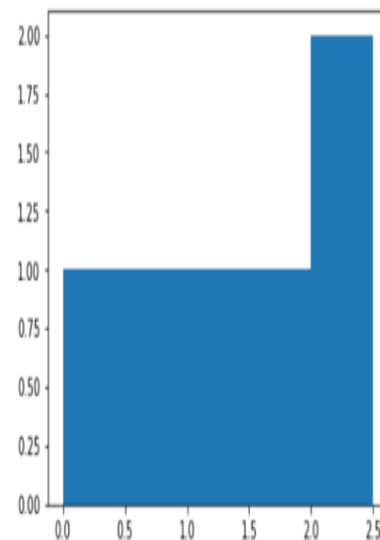
In [3]: source.data
Out[3]: {'x': [1, 2, 3, 4, 5], 'y': [8, 6, 5, 2, 3]}
```

Fig.4 Set of data stored in array for implementation

Importing flower data set from bokeh.sampledata.iris. The data is stored in pandas dataframe which v examine with doc.head method. To create a column data source in this data frame v simply pass this dataframe to the column data source initialize. Now the data source can be passed to any glyph method.

Fig.5 For distribution data plot histogram

```
In [2]: import matplotlib.pyplot as plt
values=[0, 0.6,1.4,1.6,2.2,2.5]
plt.hist(values, bins=5)
plt.show()
```



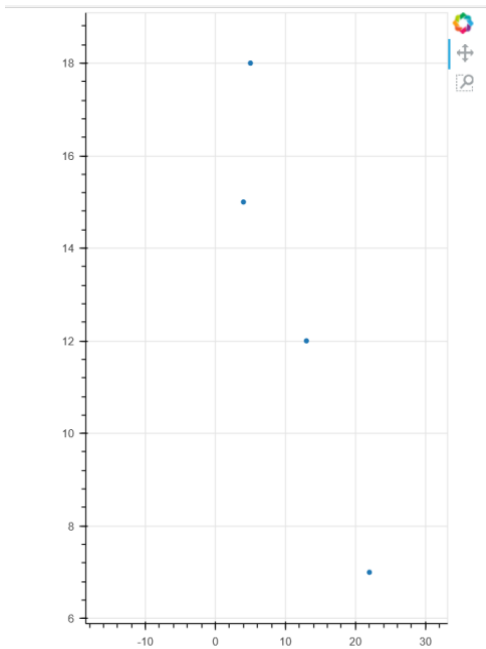


Fig.6 Bokeh plotting in circles & scatter charts.

1) Lines & markers together

- a) They can show trends, patterns over growing & lowering rates.

```
In [1]: from bokeh.io import output_file, show
In [2]: from bokeh.plotting import figure
In [3]: x = [1,2,3,4,5]
In [4]: y = [8,6,5,2,3]
In [5]: plot = figure()
In [6]: plot.line(x, y, line_width=2)
In [7]: plot.circle(x, y, fill_color='white', size=10)
In [8]: output_file('line.html')
In [9]: show(plot)
```

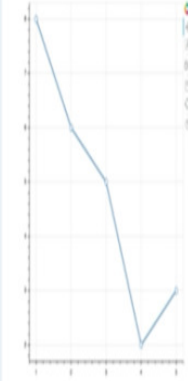


Fig.9 Bokeh implementation of a line graph

2) Patches

- a) Useful for showing geographic regions
b) Data given as “list of lists”

```
In [5]: from bokeh.io import output_file, show
from bokeh.plotting import figure
plot=figure()
plot.circle(x=10, y=[2,5,8,12],size=[10,20,30,40])
output_file('circle.html')
show(plot)
```

Fig.10 Bokeh implementation of a patch graph

```
In [1]: from bokeh.io import output_file, show
In [2]: from bokeh.plotting import figure
In [3]: xs = [ [1,1,2,2], [2,2,4], [2,2,3,3] ]
In [4]: ys = [ [2,5,5,2], [3,5,5], [2,3,4,2] ]
In [5]: plot = figure()
In [6]: plot.patches(xs, ys,
...:               fill_color=
...:               ['red', 'blue', 'green'],
...:               line_color='white')
In [7]: output_file('patches.html')
In [8]: show(plot)
```

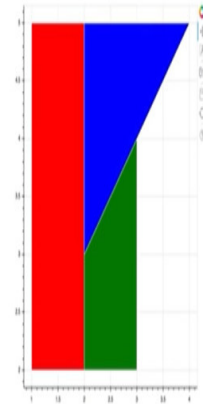


Fig.7 Snippet of plotting a circle with input data in bokeh

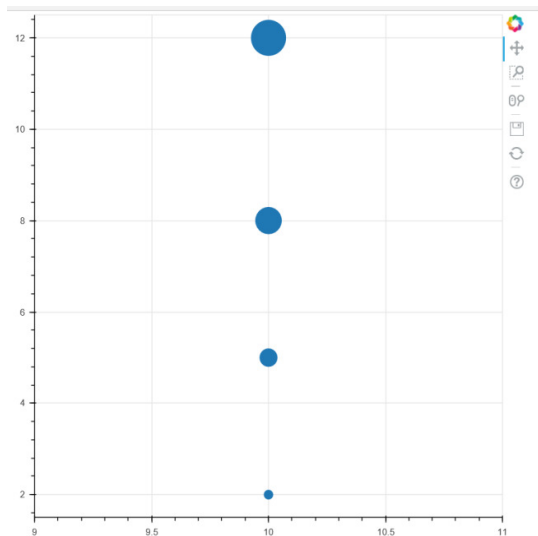


Fig.8 A scatter chart plotted as the execution of the above code

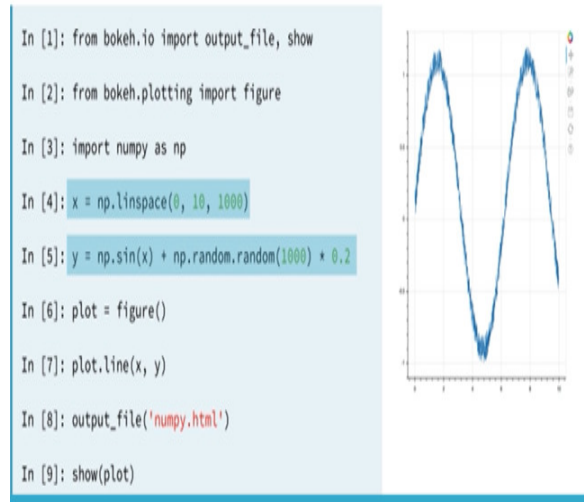
C. Additional glyphs

3) NumPy arrays

- a) For dealing with the multidimensional arrays we use a python library called NumPy arrays. It is the foundation of scientific python stack used for data science data analytics to rely upon directly or indirectly. Here we use a linspace, a NumPy function to create an array of 0 to

1000 with equal spacing of 10. These are the x values. Then v use a random function to compute a noisy sine curve based on x-values

Fig.11 NumPy array from bokeh library



VI. CONCLUSION AND FUTURE SCOPE

We can conclude that matplotlib is capable of projecting 2-D figures & bokeh visualization library enables 3-D image which helps in viewing the facts from different horizons & draw out different perspective of the insights derived. Also a feature of storytelling stands a chance of being explored in further research. As it is said that a picture is worth thousand words unless you see it. So what if a voice over a picture narrating it painting a picture into your mind & making it feel alive in your mind is an add-on feature in which contributions can be done.

REFERENCES

- [1] Syed Mohd Ali, Noopur Gupta, Gopal Krishna Nayak, Rakesh Kumar Lenka 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I) "Big data visualization: Tools and challenges" Year: 2016.
- [2] R.S. Raghav, Sujatha Pothula, T. Vengattaraman, Dhavachelvan Ponnurangam 2016 International Conference on Communication and Electronics Systems (ICCES) "A survey of data visualization tools for analyzing large volume of data in big data platform" Year: 2016.
- [3] Carson K. Leung; Vadim V. Kononov; Adam G. M. Pazdor; Fan Jiang 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology "PyramidViz: Visual Analytics and Big Data Visualization for Frequent Patterns"
- [4] Viktor PEKAR CLG, University of Wolverhampton MB 114, Stafford Road Wolverhampton, UK "Linguistic Preprocessing for Distributional Classification of Words".
- [5] Lidong Wang, Guanghui Wang, Cheryl Ann Alexander, Department of Engineering Technology, Mississippi Valley State University, USA "conventional & new methods of visualization".
- [6] Fraunhofer Center for Exp. Software Eng, MD, USA - "getting an intuition of the data".
- [7] Evgeniy Yur'evich Gorodov and Vasilii Vasil'evich Gubarev Novosibirsk State Technical University, St. Karla Marksa, 20-630073 Novosibirsk, Russia "Analytical Review of Data Visualization Methods in Application to Big Data".
- [8] <http://analytics-magazine.org/data-visualization-the-future-of-data-visualization/multidimensional-view/>.
- [9] Lidong Wang, Guanghui Wang, Cheryl Ann Alexander, Department of Engineering Technology, Mississippi Valley State University, USA "conventional & new methods of visualization "benefits of data visualization tools".
- [10] Omesh Kumar; Abhishek Goyal 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT) "Visualization: A Novel Approach for Big Data Analytics" Year: 2016.
- [11] Shujaat Hussain; Sungyoung Lee 2015 Tenth International Conference on Digital Information Management (ICDIM) "Visualization and descriptive analytics of wellness data through Big Data" Year: 2015.
- [12] Yuriy O. Koval; Hlib O. Mendrul; Andrii O. Salnikov; Ievgen A. Sliusar; Olexandr O. Sudakov 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) "Interactive dynamical visualization of big data arrays in grid" Year: 2015.
- [13] Daniel Keim; Huamin Qu; Kwan-Liu Ma IEEE Computer Graphics and Applications Year: 2013 "Big-Data Visualization".
- [14] Rishikesan Kamaleswaran; Carolyn McGregor 2014 IEEE 27th International Symposium on Computer-Based Medical Systems Year: 2014, "A Real-Time Multi dimensional Visualization Framework for Critical and Complex Environments".
- [15] Jaegul Choo; Haesun Park IEEE Computer Graphics and Applications Year: 2013, "Customizing Computational Methods for Visual Analytics with Big Data".