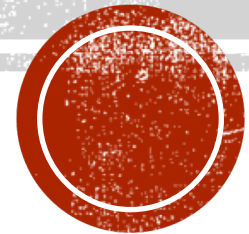


DATA VISUALISATION (DV)

Anand Phulphagar, MBA, PathKonnnect Consulting



ROADMAP

- DV – Definition
- DV – Introduction
- What is DV?
- DV to Data Analytics
- Why we need DV?
- DV Standard Practices
- DV & Data Relationship
- Chart Type Guide
- DV Process
- KPI, KRI & KQI
- Conclusion
- References & Readings



DV- DEFINITIONS

- **Data visualization** is viewed by many disciplines as a modern equivalent of visual communication. It involves the creation and study of the visual representation of data. To communicate information clearly and efficiently, data visualization uses statistical graphics, plots, information graphics and other tools. - **Wikipedia.**
- **Data visualization** is the process of presenting information through some visible means. Its goal is to make it easier to see patterns, and determine **meaning**, from complex **data** sets. – **Edward Tufte**

DV- INTRODUCTION

DATA has been available to mankind since the age of time. Data has been stored in various formats like paintings, carvings, manuscripts etc. These were the ancient forms of data storage trying to tell us a story or convey some message.

Currently, it is estimated **2.5 Million Terabytes** of data is being generated **every day**. Over the last few decades data has grown exponentially and moving forward it will only continue to grow. Currently it is said that more than **80% of the data generated today is unstructured**.

VISUALIZATION is a technique of creating images, diagrams or animations to present structured & unstructured data to communicate a message.

The single aim of DV is to tell its audience an interesting story based on their data that would lead to the following:

- Take action based on the information.
- Enlighten them with new information.
- Question preconceived notions of reality.

HUMANS HAVE THE CAPACITY TO PROCESS IMAGES **60000** TIMES **FASTER** THAN TEXT

WHAT IS DV?

- DV is a branch of Data Science.
- At its core DV is a form of visual communication, encoding aggregations and representing data visually to gain insight from the data.
- **70% - 80% time is spent prepping the data.**
- Provide clear and efficient visuals.
- Ensure data is more accessible, understandable and usable.

DATA VISUALIZATION



Data Analytics

Four Categories

Categories	Time Horizon	Questions Addressed
Descriptive Analytics	Past	What happened?
Diagnostics Analytics	Past	Why did it happen?
Predictive Analytics	Future	What will happen next?
Prescriptive Analytics	Future	How can we make it happen?

WHY WE NEED DV?

There is no question DV can be a good thing. In fact, it is already helping thousands of analysts do their job efficiently. While there could be many reasons for and against DV, if DV are not done right, there could be some serious implications. We will look into both aspects.

Advantages of DV.

- Since data generated today is humongous DV provides a quick overview.
- Absorbing information becomes easier since humans process visuals faster than text.
- Discover relationships and patterns.
- Identify emerging trends faster.
- Share your insights with the target audience.
- Data can be made more interactive.
- Faster decision making OR acting on the findings quickly.

Problems of DV.

- Oversimplification of data.
- Human limitation of algorithms.
- Overdependence on visuals.
- Can be used as manipulative tool.
- Different interpretations depending on the audience.

DV STANDARD PRACTICES

While the field is still evolving, there are few practices that have been accepted as standards, which makes an ordinary DV into a great one. Let's see a few recommended Standard Practices to generate a great DV.

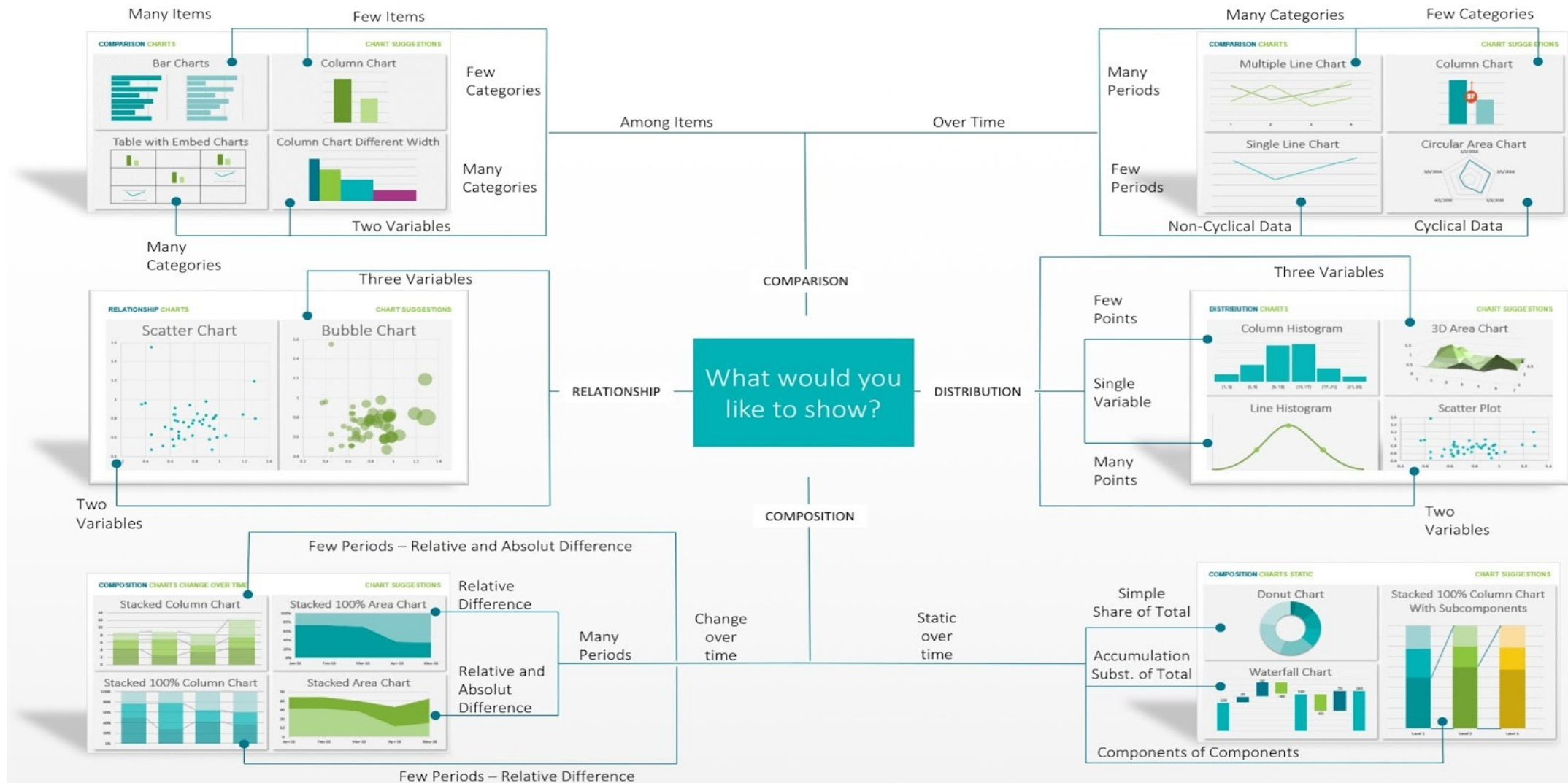
- Requirement gathering template.
 - E.g. Types or report, # of users, data refresh schedule etc.
- Formatting style:
 - E.g. Font, size, logo, color (be mindful of your color blind audience) etc.
- Data Profiling, putting logical data together to ensure DV have a flow.
- Understand different types of charts and their purpose and use them consistently.
- Highlight important stuff, remove clutter, clearly label titles, legends etc.
- Add limited interactivity to encourage exploration.
- Display medium. Customization is required.
 - E.g. TV, laptop, projector & mobile etc.
- Limit the number of views and colors.
- Limit eye travel & ensure ease of learning.

DV & DATA RELATIONSHIP

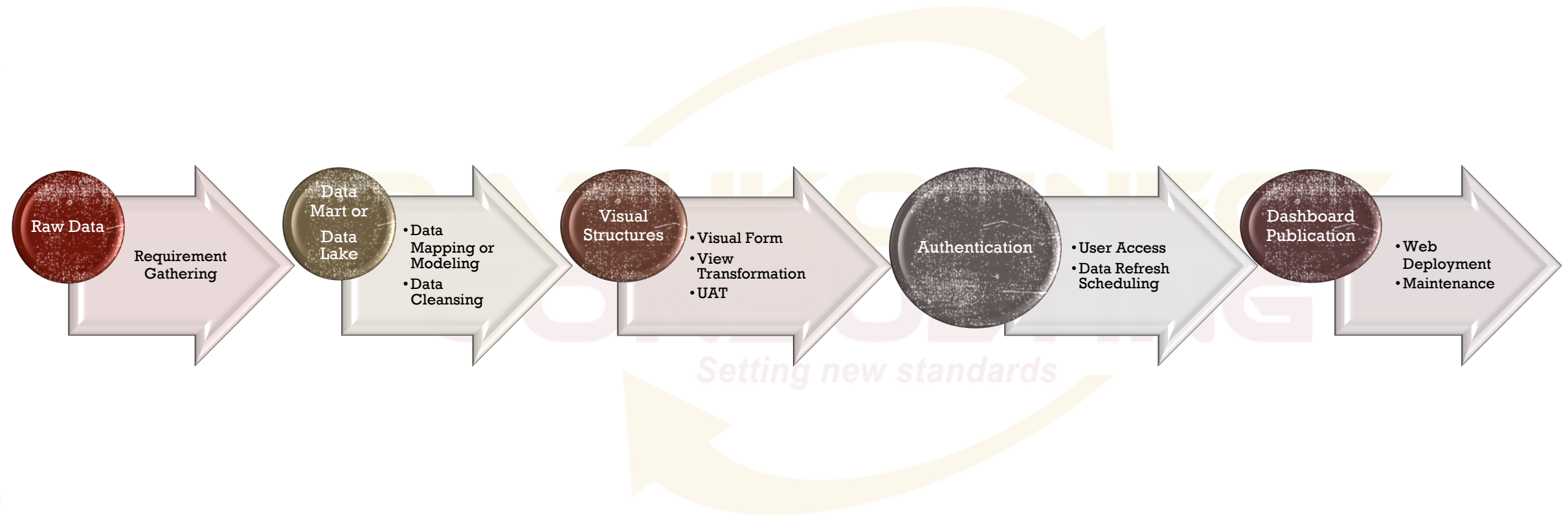
<u>Data Relationship</u>	<u>Meaning</u>
Comparison	Comparing one set of variables with another and display how those two variables interact.
Composition	Collects different types of information that make up a whole and display them together.
Distribution	Displaying of related or un-related information simply to see how it correlates, if at all, and understand if there is any interaction between those variables.
Relationship	Shows connection or correlation between two or more variables through the data presented.

CHART TYPE GUIDE –

DR. ANDREW ABELA



DV PROCESS



KEY PERFORMANCE, RISK & QUALITY INDICATORS (KPI, KRI & KQI)

- **KPI are critical indicators of progress towards an intended result.**
 - E.g.: Number of Patients enrolled, Number of Adverse Events etc.
- **KRI are critical predictors which incorporate additional information about thresholds and certain risk which may take place if no action is taken.**
 - E.g.: Number of Patients enrolled is less than planned, Number of Adverse Events is too high for a particular site in comparison too other site.
- **KQI are quality measures put in place, where quality is degree of excellence.**
 - E.g.: Edit Checks or Data Validation checks against data points.

CONCLUSION

Data Visualization is a very intriguing field and one could get carried away very easily. It can be simple and yet very complex. If done right, it will be a great asset on various fronts. But step back and take a closer look, if your DV is conveying the story you want it to tell. Remember it is telling you a story based on the available data, however it is written/programmed by humans. Make sure you have the necessary tools to test the accuracy of your data and your DV. *Setting new standards*

REFERENCES & READING

<u>Name</u>	<u>Link</u>
Data Science Central	https://www.datasciencecentral.com/profiles/blogs/4-potential-problems-with-data-visualization
Seeing data	http://seeingdata.org
Wikipedia	https://en.wikipedia.org/wiki/Data_visualization
Chart Type	https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html
History of Data Visualization:	https://medium.com/@Infogram/key-figures-in-the-history-of-data-visualization-30486681844c

SPECIAL THANKS

- **PhUSE Data Visualization (Chair & Co-Chairs)**
- **Mentors & Co-Workers**
 - **Mike Williams** (Manager Clinical Programming – Abbott).
 - **Chili Li** (Director Data Management – Medarex).
 - **Qian Ding** (Manager Clinical Programming – Medarex).
 - **Shafi Huda** (Sr. Clinical Programmer - Abbott & Medarex).
 - **Kartik Patel** (Sr. Bio-Statistician – Abbott).
 - **Stanley Coulanges** (Sr. Specialist IT – Abbott & Allergan).
 - **Rajesh Malani** (Manager Statistical Programming – Syneos Health).
 - **Mark Elliott** (Co-Founder, PRM Payments).



Questions ?

CONTACT INFORMATION

Author Name:	Anand Phulphagar
Company:	PathKonnnect Consulting Inc
Address:	130 New Rd, L20
City/Postcode:	Parsippany, NJ 07054
Work Phone:	1.973.978.6558
Fax:	1.856.210.7274
Email:	info@pathkonnnect.com
Web:	www.pathkonnnect.com