



PhUSE US Connect 2019

Paper DV04

Data Visualization (DV)

Anand Phulphagar, PathKonnnect Consulting Inc, Parsippany, NJ, USA

ABSTRACT

Data Visualization (DV) is a branch of Data Science used for presenting data in a visual format. While it has become a hot topic in many industries, it is not a new subject and has been around for couple of centuries. According to Wikipedia "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". DV is one of the most discussed & searched topics today following Artificial Intelligence, Machine Learning & Big Data.

DV is an art and science to present data in a visual format. There are many factors that need to be put in place to create a great DV. DV has a single aim and that is to tell its audience an interesting story which would compel them to take action, enlighten them with new information or raise new questions. It helps them to see what the data has to say.

INTRODUCTION

Data Visualization - before we dive into this subject let's look at these two words separately.

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.

IT IS SAID HUMANS HAVE THE CAPACITY TO PROCESS IMAGES 60000 TIMES FASTER THAN TEXT.

WHAT IS DV?

At its core DV is a form of visual communication, encoding aggregations and representing data visually to gain insight from the data. While the primary goal of DV is to communicate information clearly and efficiently with visuals, effective visualization not only helps users analyze data but makes data more accessible, understandable and usable.

With such huge amounts of data being generated every day, it is practically impossible to make sense of this data without the help of technology. While there are many software's, programming packages to display the data visually, all these tools are limited in some sense or the other. Either through the changing technology or the skills and knowledge of the professional using these tools.

Finally, most DV will eventually be used in **Data Analytics** and will be classified into 4 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

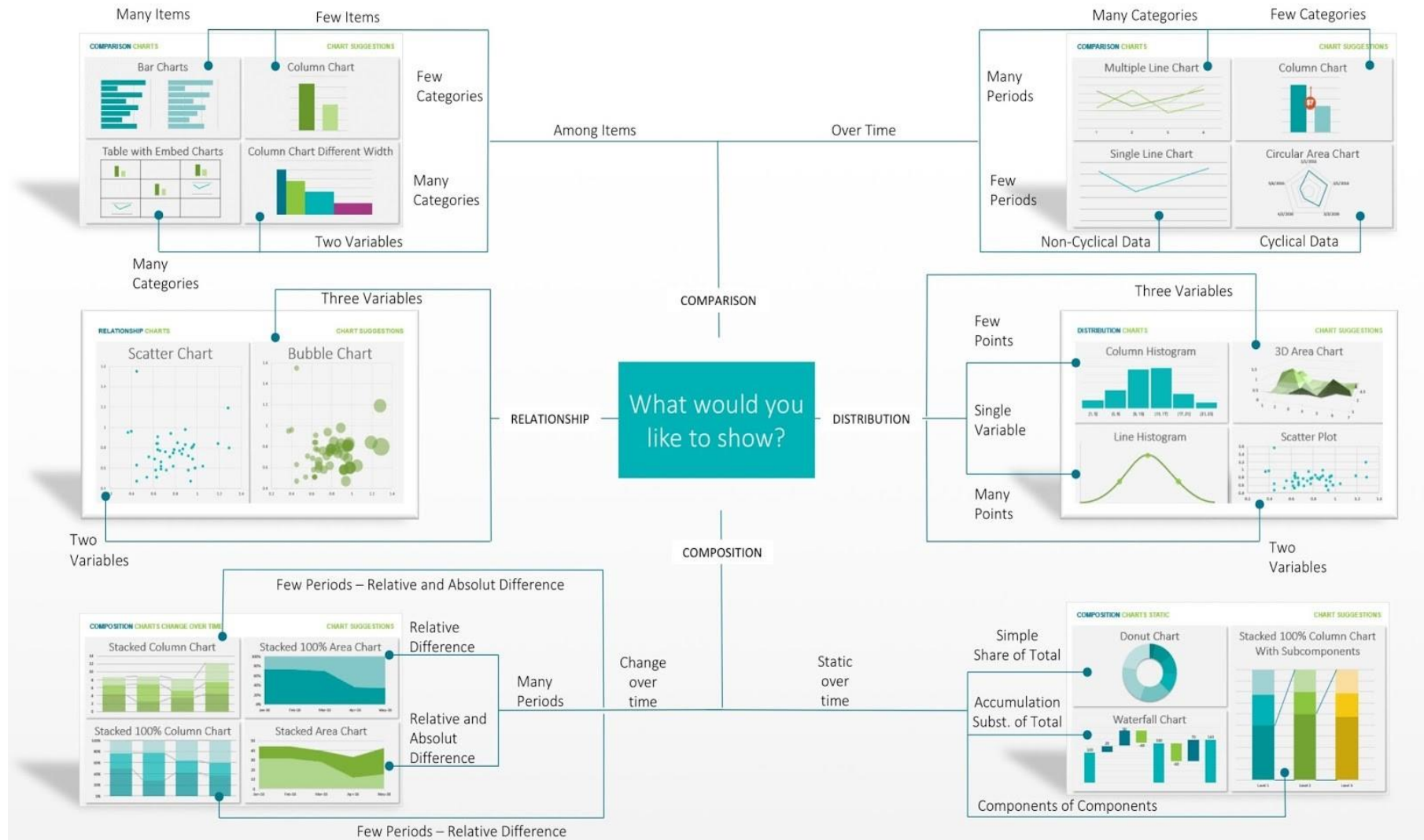
While the field is still evolving, there are few practices that have been accepted as standards, which makes an ordinary DV into a great one. Let's see a few recommended Standard Practices to generate a great DV.

- It is very important to know your audience.
- Remember not everyone one in your audience is a Data Analyst.
- Display size, what is the output medium of your DV.
- Limit the of number of views and colors.
- Understand different types of charts and their purpose and use them consistently.
- Highlight important stuff, remove clutter, clearly label titles, legends etc.
- Add interactivity to encourage exploration.
- Limit eye travel.

DV revolves around four data relationships aspects (**Comparison, Composition, Distribution & Relationship**).

Data Relationship	Meaning
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.

Image below by **Dr. Andrew Abela** guides us in selecting a correct chart type for our visualization.





CLINICAL TRIAL STORY TELLING

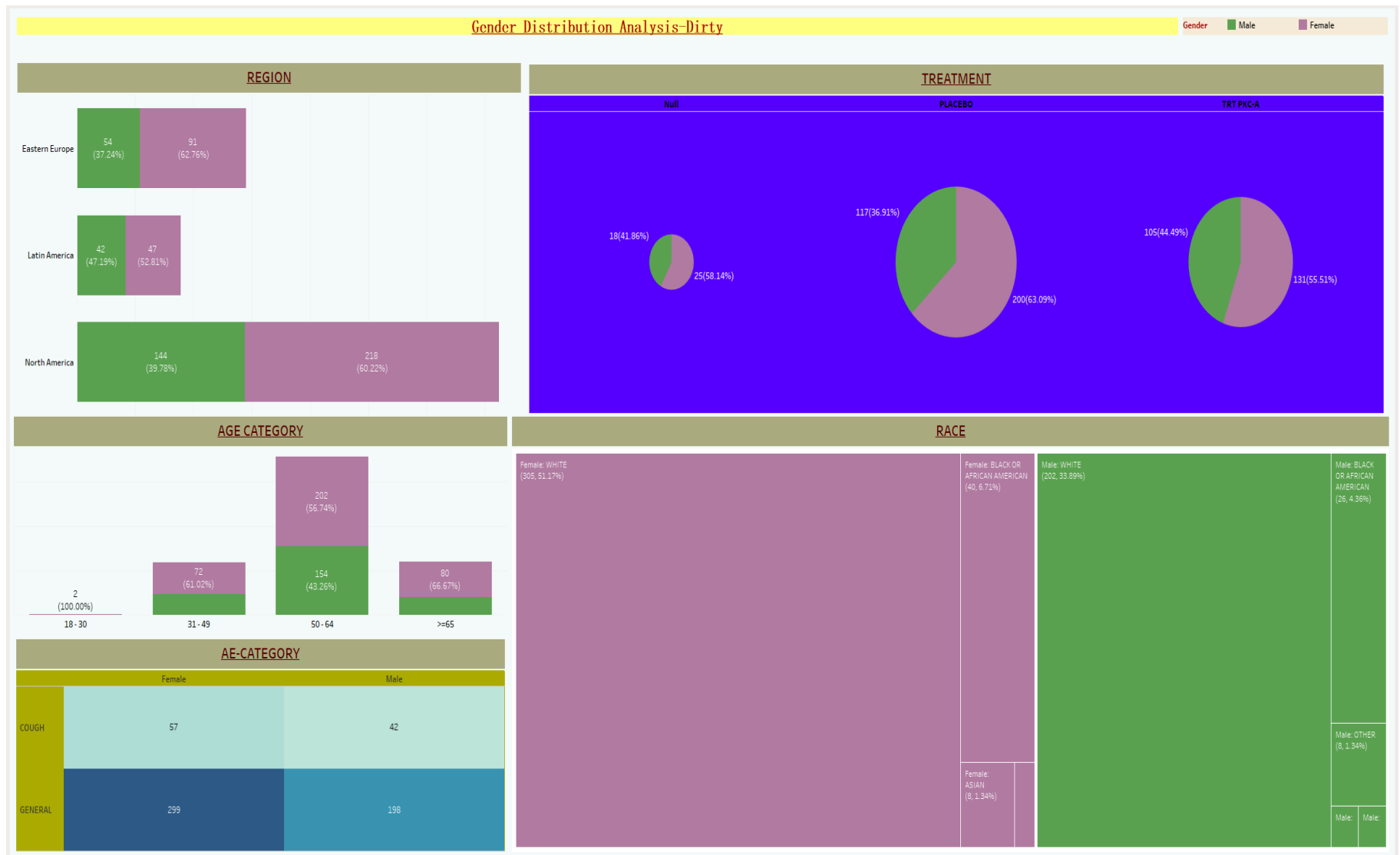
Every DV tells a story, let's look at a few examples below.

Clean Dashboard Example: The DV below uses consistent Chart Types, color shades etc.





Dirty Dashboard Example: The DV below has different colors schemes, different chart types etc. Making it difficult for its audience to focus and easily understand the DV.





STUDY OVERVIEW WITH AE INFORMATION:

THE DV SHOWN BELOW HAS **22** DIFFERENT VIEWS PRESENTED ON A SINGLE DASHBOARD, ALLOWING INTERACTIVE BY REGION & AE START DATE. CONSISTENT CHART-TYPES, SIMILAR COLORING, CLEAR TITLES, LABELS, LEGENDS AND LIMITED EYE TRAVEL MAKE IT AN EASY TO UNDERSTAND DV.





UNDERLYING DATA:

SCREENSHOT OF THE UNDERLYING DATASET USED TO GENERATE THE ABOVE SHOWN DATA VISUALIZATIONS.

#	A	C	Q	T	V	W	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AH	AO	AP	AQ	AR	AS	AX	BD	BE	BF	BH	BI	BJ
1	STUDYID	SUBJID	AESTDTC	AENDTDC	TEAEFL	RUNINFL	AESR	AESCONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESMIE	AESSEV	AEREL	RELPROC	AEDVREL	AECN	AESBWD	AESOUT	AESTRF	SITEID	REGION	SEX	AGE	AGECAT	RACE	TRTTP	RANFL	ELIGFL	SAFFL	FASFL	PPFL	RIPOPFL
531	PKC-IN-007	4105	12/13/2012	12/13/2012	Y	N	Y				Y	Y		SEVERE	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	483	Latin America	Female	53	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	Y	
532	PKC-IN-007	4105	2/15/2013	2/19/2013	Y	N	N							MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	483	Latin America	Female	53	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	Y	
533	PKC-IN-007	4107	11/20/2012	2/26/2013	Y	N	N							MILD	RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	70	North America	Male	41	31 - 49	BLACK OR AFRICAN AMERICAN	PLACEBO	Y	Y	Y	Y	Y	
534	PKC-IN-007	4107	10/9/2012		N	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	NOT RESOLVED	Before Run-In	70	North America	Male	41	31 - 49	BLACK OR AFRICAN AMERICAN	PLACEBO	Y	Y	Y	Y	Y	
535	PKC-IN-007	4109	5/3/2013		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	Placebo	484	Latin America	Male	57	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	Y	
536	PKC-IN-007	4119	1/10/2013	2/6/2013	Y	N	N							MILD	POSSIBLY RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	483	Latin America	Male	59	50 - 64	BLACK OR AFRICAN AMERICAN	PLACEBO	Y	Y	Y	Y	Y	
537	PKC-IN-007	4119	5/20/2013	5/23/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	RESOLVED	Placebo	483	Latin America	Male	59	50 - 64	BLACK OR AFRICAN AMERICAN	PLACEBO	Y	Y	Y	Y	Y	
538	PKC-IN-007	4119	5/20/2013	5/27/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	483	Latin America	Male	59	50 - 64	BLACK OR AFRICAN AMERICAN	PLACEBO	Y	Y	Y	Y	Y	
539	PKC-IN-007	4120	2/8/2013	2/26/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	484	Latin America	Female	59	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
540	PKC-IN-007	4120	3/26/2013		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	484	Latin America	Female	59	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
541	PKC-IN-007	4134	5/7/2013		Y	N	N							MILD	UNLIKELY RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	382	Eastern Europe	Female	63	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
542	PKC-IN-007	4134	5/7/2013		Y	N	N							MILD	UNLIKELY RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	382	Eastern Europe	Female	63	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
543	PKC-IN-007	4134	5/7/2013		Y	N	N							MILD	UNLIKELY RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	382	Eastern Europe	Female	63	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
544	PKC-IN-007	4142	11/22/2012		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	484	Latin America	Male	55	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	N	Y	
545	PKC-IN-007	4143	11/22/2012	2/7/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	484	Latin America	Female	59	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	Y	
546	PKC-IN-007	4147	11/26/2012		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	Placebo	484	Latin America	Male	40	31 - 49	WHITE	PLACEBO	Y	Y	Y	N	Y	
547	PKC-IN-007	4154	5/7/2013		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	853	Eastern Europe	Female	59	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
548	PKC-IN-007	4159	5/7/2013		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	Placebo	484	Latin America	Male	60	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	Y	
549	PKC-IN-007	4159	5/28/2013		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	Placebo	484	Latin America	Male	60	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	Y	
550	PKC-IN-007	4160	11/20/2012	4/30/2013	Y	N	N							MILD	POSSIBLY RELATED	POSSIBLY RELATED	POSSIBLY RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	59	50 - 64	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
551	PKC-IN-007	4160	12/28/2012	1/2/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	59	50 - 64	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
552	PKC-IN-007	4160	2/22/2013	2/24/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	59	50 - 64	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
553	PKC-IN-007	4160	2/24/2013	2/24/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	59	50 - 64	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
554	PKC-IN-007	4162	11/22/2012	2/20/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	484	Latin America	Male	39	31 - 49	WHITE	TRT PKC-A	Y	Y	Y	N	Y	
555	PKC-IN-007	4167	4/2/2013	4/9/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	852	Eastern Europe	Female	65	>=65	WHITE	PLACEBO	Y	Y	Y	Y	Y	
556	PKC-IN-007	4202	2/18/2013	2/20/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	559	Eastern Europe	Male	69	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
557	PKC-IN-007	4211	11/27/2012		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	Placebo	484	Latin America	Male	65	>=65	WHITE	PLACEBO	Y	Y	Y	Y	Y	
558	PKC-IN-007	4211	2/19/2013	5/10/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Placebo	484	Latin America	Male	65	>=65	WHITE	PLACEBO	Y	Y	Y	Y	Y	
559	PKC-IN-007	4214	11/16/2012		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	TRT PKC-A	484	Latin America	Male	67	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
560	PKC-IN-007	4218	10/23/2012		N	Y	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	NOT RESOLVED	Run-In	183	North America	Male	68	>=65	WHITE		N	N	N	N		
561	PKC-IN-007	4236	12/26/2012		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVING	TRT PKC-A	388	North America	Male	79	>=65	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
562	PKC-IN-007	4236	12/26/2012		Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVING	TRT PKC-A	388	North America	Male	79	>=65	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
563	PKC-IN-007	4236	3/20/2013	4/2/2013	Y	N	N							MILD	UNLIKELY RELATED	UNLIKELY RELATED	UNLIKELY RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	388	North America	Male	79	>=65	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
564	PKC-IN-007	4236	3/26/2013	4/6/2013	Y	N	N							MILD	UNLIKELY RELATED	UNLIKELY RELATED	UNLIKELY RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	388	North America	Male	79	>=65	BLACK OR AFRICAN AMERICAN	TRT PKC-A	Y	Y	Y	Y	Y	
565	PKC-IN-007	4275	10/16/2012	10/16/2012	N	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	Before Run-In	141	North America	Male	53	50 - 64	WHITE		N	Y	N	N	N	
566	PKC-IN-007	4277	11/10/2012	1/22/2013	N	Y	Y				Y			MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	Y	RESOLVED	Run-In	272	North America	Male	54	50 - 64	WHITE		N	Y	N	N	N	
567	PKC-IN-007	4280	3/23/2013		Y	N	Y				Y			SEVERE	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVING	TRT PKC-A	852	Eastern Europe	Male	53	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
568	PKC-IN-007	4290	10/25/2012	10/26/2012	Y	N	N							MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	RESOLVED	Run-In	382	Eastern Europe	Female	60	50 - 64	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
569	PKC-IN-007	4293	12/17/2012	12/25/2012	Y	N	N							MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	RESOLVED	Placebo	555	Eastern Europe	Female	63	50 - 64	WHITE	PLACEBO	Y	Y	Y	N	Y	
570	PKC-IN-007	4293	12/25/2012	1/18/2013	Y	N	N							MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	RESOLVED	Placebo	555	Eastern Europe	Female	63	50 - 64	WHITE	PLACEBO	Y	Y	Y	Y	N	Y
571	PKC-IN-007	4298	10/16/2012	10/21/2012	N	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	RESOLVED	TRT PKC-A	483	Latin America	Female	61	50 - 64	WHITE		N	N	N	N		
572	PKC-IN-007	4304	12/8/2012	1/19/2013	Y	N	N							MILD	POSSIBLY RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	73	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
573	PKC-IN-007	4304	1/30/2013	3/15/2013	Y	N	N							MODERATE	POSSIBLY RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	73	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
574	PKC-IN-007	4304	11/7/2012	11/19/2012	N	Y	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	NOT APPLICABLE	N	RESOLVED	Run-In	483	Latin America	Female	73	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
575	PKC-IN-007	4304	12/27/2012	1/3/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	73	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
576	PKC-IN-007	4304	1/30/2013	3/15/2013	Y	N	N							MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	73	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
577	PKC-IN-007	4304	2/25/2013	2/25/2013	Y	N	N							MODERATE	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	73	>=65	WHITE	TRT PKC-A	Y	Y	Y	Y	Y	
578	PKC-IN-007	4304	3/1/2013	3/4/2013	Y	N	N							MILD	NOT RELATED	NOT RELATED	NOT RELATED	DOSE NOT CHANGED	N	RESOLVED	TRT PKC-A	483	Latin America	Female	73	>=65</								



CONCLUSION

Data Visualization is a very intriguing field and one could get carried away very easily. 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.

REFERENCES

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://extremerepresentation.typepad.com/blog/2006/09/choosing_a_good.html

RECOMMENDED READING

History of Data Visualization: <https://medium.com/@Infogram/key-figures-in-the-history-of-data-visualization-30486681844c>

Tableau: <https://www.tableau.com/learn/whitepapers>

CONTACT INFORMATION

Author Name: Anand Phulphagar
Company: PathKonnnect Consulting Inc
Address: 130 New Road, L20
City/Postcode: Parsippany, NJ 07054
Work Phone: 001-973.978.6558
Fax: 001-856.210.7274
Email: aphulphagar@pathkonnnect.com
Web: www.pathkonnnect.com

Brand and product names are trademarks of their respective companies.