

## **MODULE - I**

### **Introduction to Data Visualization**

Python Pandas Data Science Library- Python Lambdas: md the NumPy Library - Data Visualization for Business Decision Making- Defining Data Visualization Technique- Data visualization Methodology- Visualization design objective- Refining Editorial focus- Using Visual Analysis to find Stories- An Example of finding and Telling Stories.

## **Introduction to Data Visualization**

### **What is Data Visualization?**

**Data visualization is the practice of turning data into graphics.**

Good graphics are more easily interpreted than the raw data.

A well-designed visualization is faithful to the original data and does not mislead the intended audience.

Every data visualization is a simplification and approximation of a raw dataset, and involves a perspective — the goals and biases of the person producing the visualization.

### **What is the purpose of data visualization?**

Data visualization is used to help tell stories with data.

The goal is to communicate

an interpretation of a data set to a particular audience,  
to make an argument worked out from an analysis of the data.

Visualizations are best when accompanied by written explanations. (*Even if a picture is worth a thousand words*), (*visualizations do not usually stand on their own.*)

Once the message is understood and internalized, a good visualization can sometimes tell the story “By itself”.

### **Benefits of data visualization**

Data visualization can be used in many contexts in nearly every field.

#### **Storytelling:**

People are drawn to colours and patterns in clothing, arts and culture, architecture, and more. Data is no different—colours and patterns allow us to visualize the story within the data.

#### **Accessibility:**

Information is shared in an accessible, easy-to-understand manner for a variety of audiences.

#### **Visualize relationships:**

It's easier to spot the relationships and patterns within a data set when the information is presented in a graph or chart.

## **Exploration:**

More accessible data means more opportunities to explore, collaborate, and inform actionable decisions.

## **Roles in data visualization**

### **Data visualization analyst:**

As a **data visualization analyst** (or specialist), you'd be responsible for **creating and editing visual content** such as maps, charts, and infographics from large data sets.

### **Data visualization engineer:**

Data visualization **engineers and developers** are experts in both **manoeuvring data** with SQL, as well as assisting product teams in creating **user-friendly dashboards** that enable storytelling.

### **Data analyst:**

A data analyst **collects, cleans, and interprets data** sets to **answer questions or solve business problems**.

## **Data Visualization for Business Decision Making**

### **Shifting from Input to Output**

A decision-maker for any business wants highly visual business intelligence (BI) tools that can help to make the right decisions quickly.

Business intelligence (**BI**) has become more mainstream; hence,

Vendors are focusing on both ends of the pipeline

- ▶ Improve the **quality of data** input
- ▶ Ensuring that the output is **well-structured and clearly presented**.

BI dashboards have become tool to project **different data visualizations** to provide an **at-a-glance overview of business performance and areas for improvement**.

### **Why Is Data Visualization Important?**

A picture is worth a thousand words,

Humans just understand data better through pictures rather than by reading numbers

data presented in a graphical format, enable people to effectively find correlations and raise important questions.

**Data visualization helps the business to achieve numerous goals.**

- Converting the business data into interactive graphs for dynamic interpretation to serve the business goals
- Transforming data into visually appealing, interactive dashboards of various data sources to serve the business with the insights
- Creating more attractive and informative dashboards of various graphical data representations
- Making appropriate decisions by drilling into the data and finding the insights
- Figuring out the patterns, trends, and correlations in the data being analysed to determine where they must improve their operational processes and thereby grow their business
- Giving a fuller picture of the data under analysis
- Organizing and presenting massive data intuitively to present important findings from the data
- Making better, quick, and informed decisions with data visualization

### **Visualization Help in Making Business Decisions**

In order to get the most benefits out of data visualization, the following criteria must be met:

- 1) The visualizations used in a given presentation must be the most relevant and befitting to the body of information in question, usually the source data.
- 2) The data must be combined with background information that puts the charts, figures, and symbols into context for easy comprehension by all interested parties.
- 3) The visualizations must present key data in a manner that points to sound courses of action.

As long as data visualization is implemented with all three of the above factors in place, analysts can quickly put the insights they glean into action, all with great confidence that the next step will bring about newfound levels of success.

## **Why Do Modern Businesses Need Data Visualization?**

With the huge volume of data collected about business

Business leaders need proper techniques to easily drill down into the data to see where they can improve operational processes and grow their business.

Data visualization brings business intelligence to reality.

Data visualization is the need for modern businesses

- Data visualization helps companies to analyse its different processes so the management can focus on the areas for improvement to generate more revenue and improve productivity.
- It brings business intelligence to life.
- It applies a creative approach to understanding the hidden information within the business data.
- It provides a better and faster way to identify patterns, trends, and correlation in the data sets that would remain undetected with just text.
- It identifies new business opportunities by predicting upcoming trends or sales volumes and the revenue they will generate.
- It supplies managers with information they need to make more effective comparisons between data sets by plotting them on the same visualization.
- It enables managers to understand the correlations between the operating conditions and the business performance.
- It helps businesses to discover the grey areas of the business and make the right decisions for improvement.
- Data visualization helps managers to understand customers' behaviours and interests and hence retains customers and market share.

## **Advantages of Data Visualization**

### **Accelerated response speeds.**

Data visualization tools enhance the information at hand and allow analysts to absorb all the relevant info

- distinguish positives from negatives
- draw conclusions
- Take logical actions with improved speed and efficiency.

Data visualization saves tremendous amounts of effort and time

### **Improved, simplified focus on what matters most.**

The combination of **visualizations and contextual info** allows analysts to see everything at once and instantly zero in (focus) on the most important details

- Data becomes simplified because everything is simultaneously presented
- Key facts take centre stage

### **Easier viewing of trends and patterns.**

Traditional data formats make it difficult to distinguish patterns within large bodies of information

With visualization tools the most important details are brought immediately to the fore

**it's far easier to understand trends and patterns when laid out in charts, maps, and trees, and varying graphics.**

### **Ability to modify and interrelate with data.**

data visualization can present actionable ideas in ways that can be easily understood by all interested parties

While traditional charts and graphs are merely meant to be viewed

with modern tools of visualization analysts can interact with data.

### **Ability to develop a new language for business.**

data visualization tells stories through image representation instead of merely presenting static facts.

### **Improved collaboration among work teams.**

Data visualization makes it easier to collaborate (*Cooperate*).

- key info across all areas of a study are presented with equal clarity
- Everyone learns what they need to know within seconds.

## **The Future of Data Visualization**

Data visualization is moving from being an art to being a science field.

Data science technologies impose the need to move from relatively simple graphs to multifaceted relational maps.

- Multidimensional visualizations will boost the role of data visualizations

- Data visualization will play a vital role, now and in the future, in many concepts such as
  - Internet of Things
  - network and complexity theories
  - nanoscience
  - social science research
  - education systems
  - conative science,
  - space, and much more.
- Data visualization will be used intensively to analyse and visualize data streams
  - collected from billions of interconnected devices,
  - smart appliances
  - wearables
  - automobile sensors
  - environmental monitors
  - smart cities monitors
  - Internet of Things device etc...
- visualization of data will provide extraordinary insight into what's happening around the globe

In this context,

- data visualization will
- improve safety levels
- drive operational efficiencies
- help to better understand several worldwide phenomena
- improve and customize intercontinental services.

## **How Data Visualization Is Used for Business Decision-Making**

Data visualization is a real asset for any business to help make real-time business decisions.

It visualizes extracted information into logical and meaningful parts and helps users avoid information overload by keeping things simple, relevant, and clear.

There are many ways in which visualizations help a business to improve its decision-making.

## Faster Responses

Quick response to customers' or users' requirements is important for any company to retain their clients, as well as to keep their loyalty. With the massive amount of data collected daily via social networks or via companies' systems, it becomes incredibly useful to put useful interpretations of the collected data into the hands of managers and decision-makers so they can quickly identify issues and improve response times.

## Simplicity

It is impossible to make efficient decisions based on large amounts of raw data. Therefore, data visualization gives the full picture of the scoped parameters and simplifies the data by enabling decision-makers to cherry-pick the relevant data they need and dive into a detailed view wherever is needed.

## Easier Pattern Visualization

Data visualization provides easier approaches to identifying upcoming trends and patterns within data sets and hence enables businesses to make efficient decisions and prepare strategies in advance.

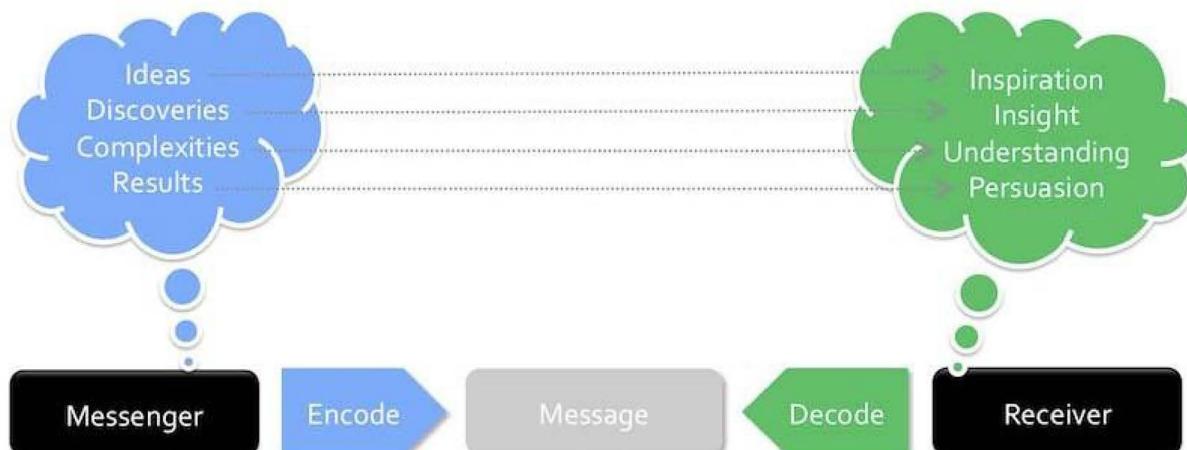
## Team Involvement

Data visualizations process not only historical data but also real-time data. Different organization units gain the benefit of having direct access to the extracted information displayed by data visualization tools. This increases the levels of collaboration between departments to help them achieve strategic goals.

## Unify Interpretation

Data visualizations can produce charts and graphics that lead to the same interpretations by all who use the extracted information for decision-making.

## Defining Data Visualization Technique



## **Defining data visualization**

The main agents involved in the exchange of information

- The messenger
- The receiver
- The message

**Messenger => Imparts => Results, Analysis, And Stories**

**Receiver / Readers => Users Of Visualization**

**Message => Channel Of Communication**



To ensure that our message is conveyed in the most effective and efficient form, one that will serve the requirements of the receiver, we need to make sure we design our message in a way that actively exploits how the receiver will most effectively interpret the message through their visual perception capabilities.



*The representation and presentation of data that exploits our visual perception abilities in order to amplify cognition.*

### **The key elements of this definition**

#### **Representation**

The representation of data is the way you decide to depict data through a choice of physical forms.

Data is the raw material and creating a representation to best portray its attributes

#### **Presentation**

The presentation of data concerns how to integrate data into the overall communicated work, including the choice of colours, annotations, and interactive features

#### **Visual perception**

Exploiting visual perception abilities relates to the scientific understanding of how our eyes and brains process information most effectively

#### **Cognition**

Amplify cognition is about maximizing how efficiently and effectively to process the information into thoughts, insights, and knowledge

## **Defining Data Visualization Technique**

Data visualization aims to understand data by

extracting and graphing information to show patterns, spot trends, and identify outliers.

There are two basic types of data visualization.

#### **Exploration**

helps to extract information from the collected data.

#### **Explanation**

demonstrates the extracted information.

## Data visualization Methodology

Data visualization methodology is about recognizing the key stages, considerations, and tactics to navigate smoothly through the visualization project.

*Visualization is fundamentally an iterative, creative design process following a process might be uncomfortable. But an organized and sequenced way might help to reduce inefficiency and wasted resource.*

**key point is, “Data Visualization is not an exact science”**

There is rarely, a single right answer or single best solution.

Mostly it is about using heuristic methods to determine the most satisfactory solutions.

The methodology is supposedly flexible, based on self-judgment and discretion

- Laying out all the important things needed to be taken into account
- Proposing some potential solutions for different scenarios.

## Visualization design objective

The **key objectives** of Visualization design are

- Strive for form and function
- Justifying the selection of everything
- Accessibility through intuitive design
- Never deceive the receiver

### Strive for form and function

This has been the age-old debate of

Form Vs function or

Style Vs substance

Aspects of design that should be combined and brought together in harmony, not at the sacrifice of one or the other

There's a need for both.

*(It is a very difficult balancing act)*

**Aim should be to hit that sweet-spot where something is aesthetically inviting and functionally effective.**

Good aesthetics

Can create favourable emotional and mental responses

Good usability

Can also create favourable emotional from the experience

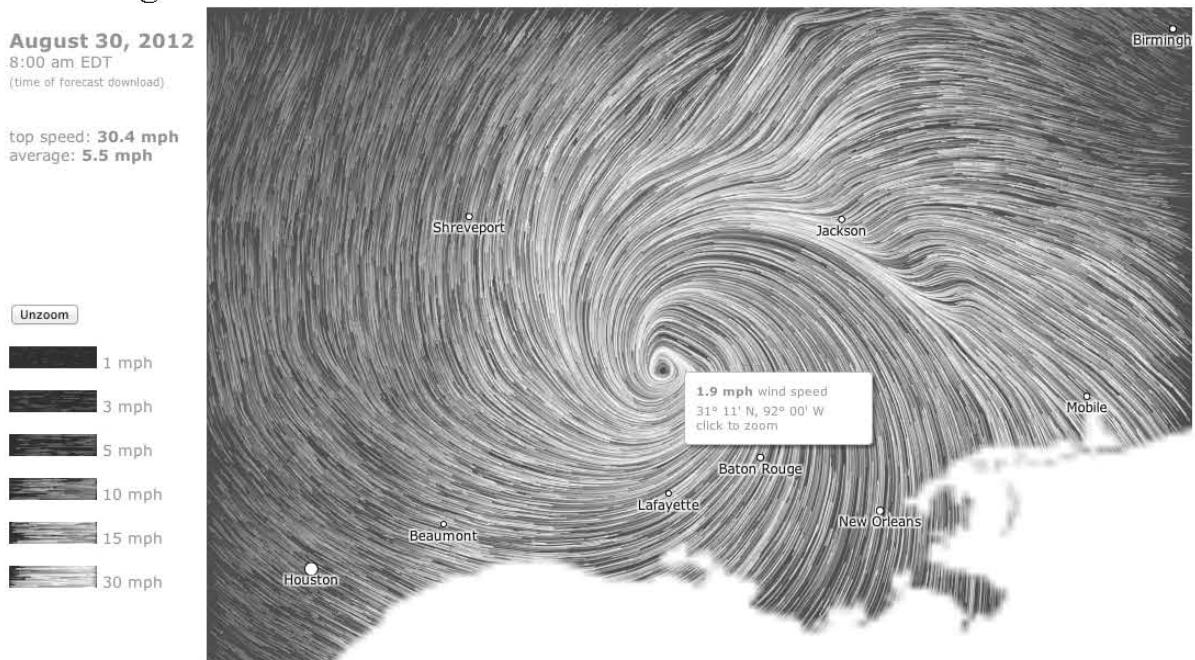
**Attractive form enhances function and the function portrays beauty through its effect.**

**Advice for beginners,**

First, achieve the foundation of something that informs

Second, explore the ways of enhancing its form

Eg..  
**elegance in form and in function**



an animated wind map developed by Fernanda Viégas and Martin Wattenberg.

exceptionally well designed and executed but it also serves its purpose as a way of informing users about the wind patterns, strength, and directions occurring across the United States

### Justifying the selection of everything

The central idea behind the methodology is to determine that everything is □

- Thoroughly planned
- Understood, &
- Reasoned.

Design methodology (steps)

- Gather the information
- Understand the scope of the information
- Ensure that everything undertaken thereafter serves
  - The purpose of our work &
  - The needs of the audience.

Deliberate design, means that the inclusion, exclusion, and execution of every single mark, characteristic, and design feature is done for a reason.

Example, when displaying a section of a tree-hierarchy work by data illustrator, Stefanie Posavec, every visible property presented is used to communicate data, whether it be the use of colour, the arc lengths of the petals, the position and sequence of stems; nothing is redundant and everything is deliberate.



## Accessibility through intuitive design

We don't want people to have to spend unnecessary time thinking about how to use or how to read and interpret something.

**visualization, => visual design + subject matter's data**

The design is the window into the data

=> the design and execution create's the accessibility of data.

The speed with which user is able to read or interpret a visualization should be determined by the complexity of the subject and the purpose of the project, not by the ineffectiveness of design.

If Data is **fundamentally simple** and the **portrayal of the data is straightforward** => reader's task of interpreting the data should be **relatively easy**

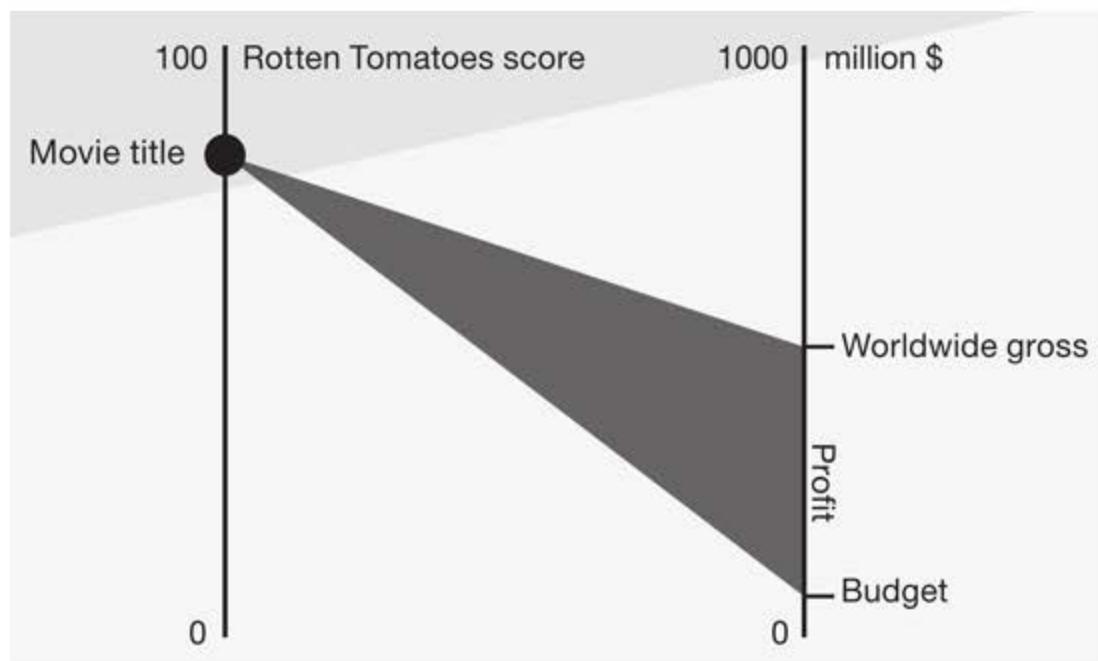
If **data is more complex**. challenge will be to respect the complexity and **avoid simplifying, diluting, or reducing the essence** of this subject. => **Visualization is not immediately easy to interpret. => reader needs to put in effort practice and experience to read the display**

Eg

This is an image of a legend that was used to explain how to read an innovative visualization to portray three separate indicators of a movie's success.

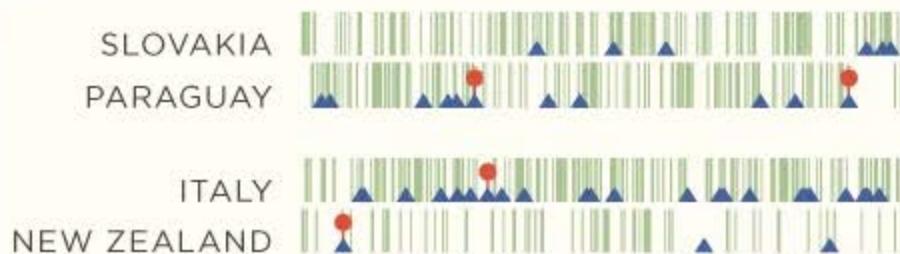
On the left-hand side of the image is the aggregate reviews (the higher the value, the better)

On the right-hand side of the image are both the budget and gross takings (the bigger the gap, the better)



### Example

Which portrays the key events in a couple of soccer matches showing completed passes (green lines), shots (blue triangles), and goals (red dots) as shown in the following image:



Once the reader has mastered the understanding of what each shape and its position means, these displays provide a powerful and rewarding insight into the key incidents and the general ebb and flow of each game.

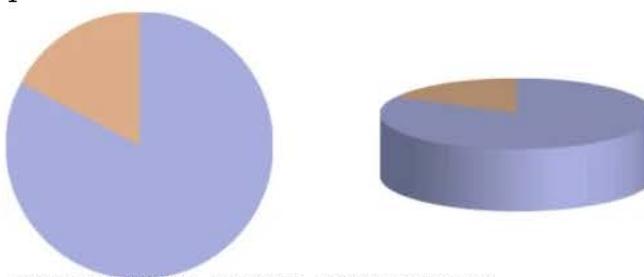
Make sure that the efforts needed from the reader or user to understand how to use and interpret a visualization are ultimately rewarded with a worthy amount of insight gained.

### Never deceive the receiver

Visualization ethics relates to the potential deception that can be created, intentionally or otherwise, from an ineffective and inappropriate representation of data.

Eg: Eye perceiving the proportion of visible pixels

2D pie chart and a 3D version



On the left-hand side of the diagram

We see a

Blue segment 82 percent &

Orange segment representing 18 percent.

On the right-hand side of the diagram

We perceive

Blue segment 91 percent of the visible area

Orange segment as only 9 percent visible area

This is clearly a hugely distorted reading of the values

Eg **Aesthetics & Style completely hijack a visualization**



3D bar chart of football results between two teams

Yellow Drawn bar representing value 1 appears to be more than half the length of a red Aston Villa bar representing 4.

Designers chose to include the category labels within the bar's length, thus completely distorting the values being represented

Clearly objective for any project should be Obeying visualization ethics

## Refining Editorial focus

To avoid being prematurely into diving into the construction of a visualization design.

we first need to do more work to fine-tune our analysis of what are the important messages.

there will be many ways of telling data stories.

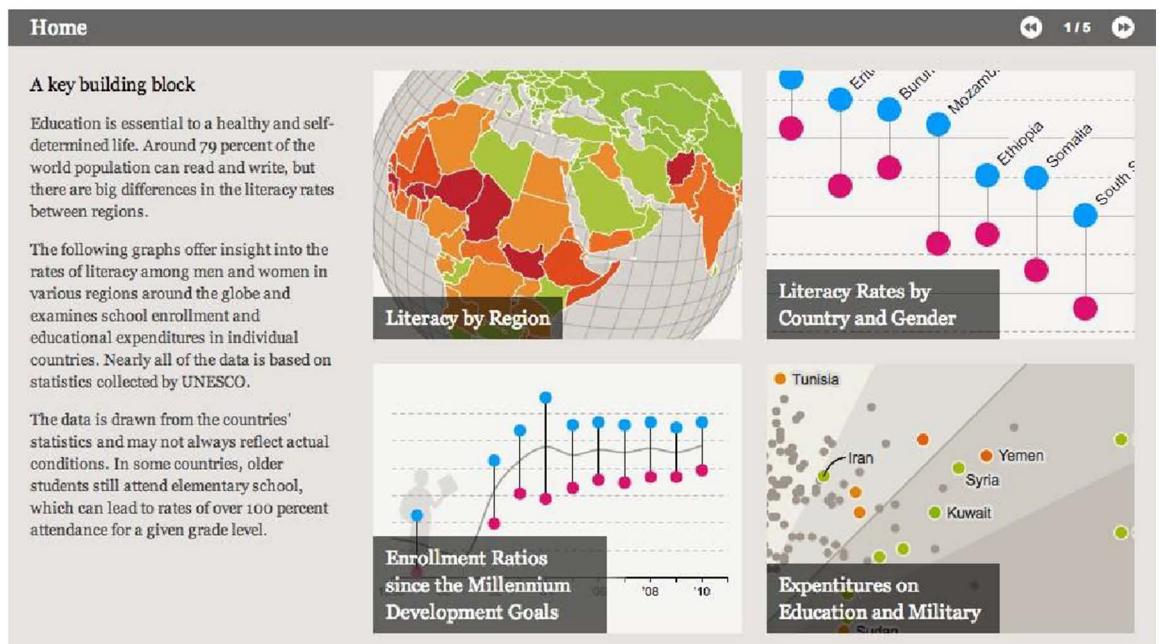
a designer is faced with the challenge of rationalizing so many different potential dimensions

The project becomes stronger from the scoping and definition of the chosen narrative and slices of analysis. Instead of bombarding the reader with endless pages of facts and figures

**Example To enlighten people about the matter of education around the world**

### In Numbers: Education Around the World

HOME GLOBAL LITERACY LITERACY BY COUNTRIES ENROLLMENT RATIOS EXPENDITURES EDUCATION & MILITARY



The story is framed around a small number of interesting angles about education: literacy by region, literacy rates by country/ gender, enrolment ratios, and expenditure on education versus military.

As we navigate through each story panel we are presented with a series of explanatory visualizations.

Considering the relationship between our editorial focus and the potential visualization design options

**The way the data is represented—the form given to it (selection of chart type)—should be influenced by the questions you are trying to answer.**

## **Induction / Deduction**

For a proper design we need to know what questions we're trying to answer

An effective approach to tackling this can be drawn from the practice of logical reasoning, specifically induction and deduction.

### **Deductive**

Deductive reasoning involves confirming or finding evidence to support specific ideas

It is a targeted and quite narrow approach concerned with validating certain hypotheses

A deductive approach to defining your data questions

What stories might be

Interesting,

Relevant, and

Potentially available within your data

□ pursuing a curiosity by interrogating your dataset in order to substantiate your ideas of what may be the key story dimensions.

### **Inductive**

It is much more open-ended and exploratory

Not sure of what the interesting stories might be

Apply analytical and visualization techniques to try and unearth potentially interesting discoveries

Thus, forming different and evolving combinations of data questions

May end up with nothing

May find plenty

Fundamentally, this is about using visual analysis to find stories.

## **Using Visual Analysis to find Stories**

When intending to discover insights about a subject. We should consider the application of visual analysis.

**Visually analysing a dataset implies applying both inductive and deductive reasoning, enables us to learn more about our subject by exploring a dataset from all directions.**

This activity also known as data sketching or preproduction visualization. This visualization techniques helps to understand the data (raw material) and form an understanding of what we might portray to others and how we might accomplish that.

Prepared to **observe the following characteristics that will lead to the identification of our key stories**

## Comparisons and proportions

### Range and distribution:

Discovering the range of values and the shape of their distribution within each variable and across combinations of variables

### Ranking:

Learning about the order of data in terms of general magnitude, identifying the big, medium, and small values.

### Measurements:

Looking beyond just the order of magnitude to learn about the significance of absolute values

### Context:

Judging values against the context of averages, standard deviations, targets, and forecasts.

### Example



## Trends and patterns:

### Direction:

Are values changing in an upward, downward, or flat motion?

### Rate of change:

How steep or flat do pattern changes occur? Do we see a consistent, linear pattern, or is it much more exponential in shape?

### Fluctuation:

Do we see evidence of consistent patterns or is there significant fluctuation? Maybe there is a certain rhythm, such as seasonality, or perhaps patterns are more random

### Significance:

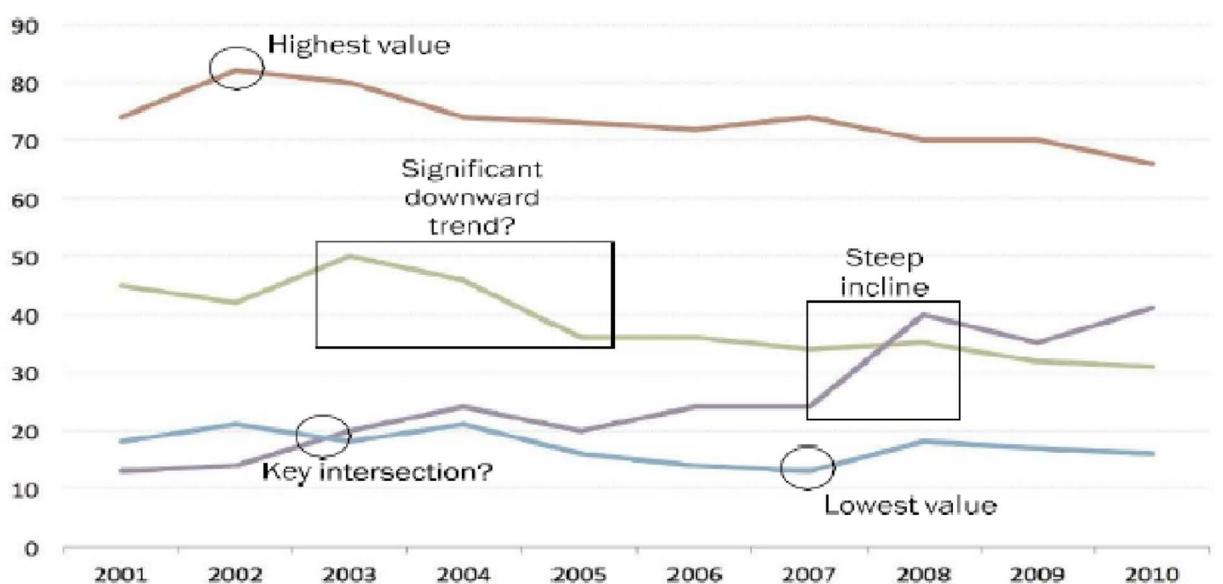
Can we determine if the patterns we see are meaningful signals

Or simply represent the noise within the data?

### Intersections:

Do we observe any important intersections or overlaps between variables, crossover points that indicate a significant change in relationship?

### Example



## **Relationships and connections:**

### **Exceptions:**

Can we identify any significant values that sit outside of the norm, such as outliers that change the dynamics of a given variable's range?

### **Correlations:**

Is there evidence of strong or weak correlations between variable combinations?

### **Associations:**

Can we identify any important connections between different combinations of variables or values?

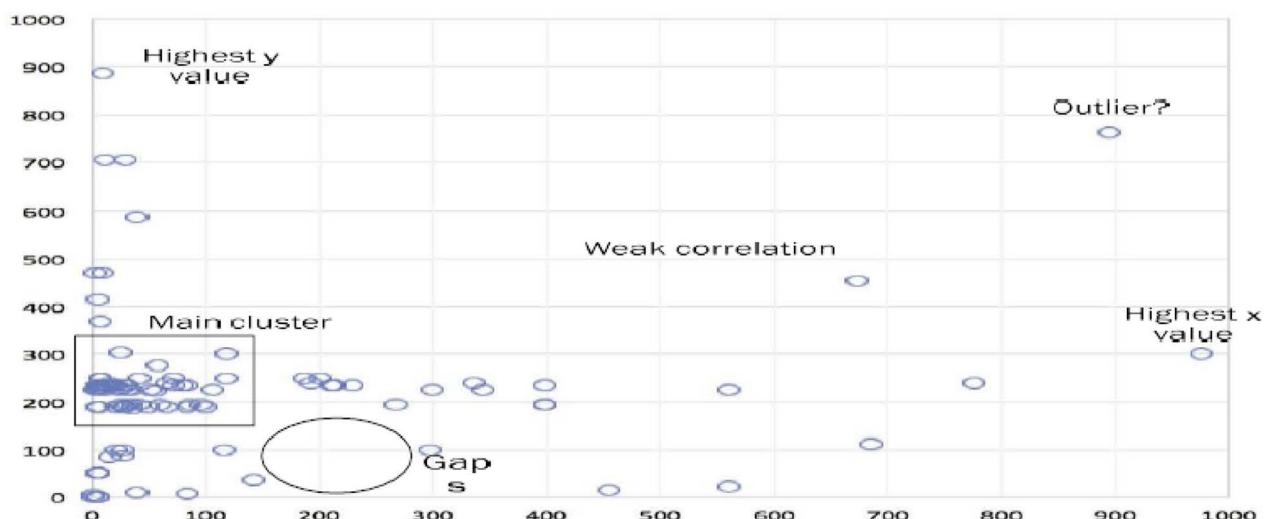
### **Clusters and gaps:**

Where is there evidence of data being "bunched"? Where are there gaps in values and data points?

### **Hierarchical relationships:**

Determining the composition, distribution, and relevance of the data's categories and subcategories.

## **Example**



In-depth visual analysis, achieve high level of data intimacy required to refine editorial focus

The visual interrogations on the data will unearth evidence from the features and the stories.

The process of visual analysis can potentially go on endlessly, with seemingly infinite combinations of variables to explore, deploying a disciplined and sensible balance between deductive and inductive enquiry should be able to efficiently and effectively navigate towards the source of the most compelling stories.

## An Example of finding and Telling Stories.

Project: to unearth some key stories around Olympics medal winning trends in recent years,

Country	Total medals won in the Summer Olympics				
	2008	2004	2000	1996	1992
United States of America	110	103	92	101	108
People's Republic of China	100	63	59	50	54
Russian Federation	72	92	88	63	112*
Great Britain	47	30	28	15	20
Australia	46	49	58	41	27
Germany	41	49	56	65	82
France	40	33	38	37	29
Republic of Korea	31	30	28	27	29
ALL	951	929	925	842	815

\* When part of former Soviet Union. Data from <http://www.databaseolympics.com/index.htm>

scanning the data with our eyes to find anything that stands out.

Russian Federation medals total for 1992 was actually when it was known as the Soviet Union. It will be sensible to just ignore this value from our analysis. Otherwise, it will skew our interpretations

**continuing examination** of the dataset's **descriptive and statistical** properties to develop an increased level of familiarity:

- Two variables: Country and event year
- Country is a categorical nominal variable with nine values (each country and the aggregate)
- Event year is a quantitative (interval-scale) variable with five values
- The maximum country medal count value is 110 medals, the minimum is 15
- The maximum aggregate value is 951 and the minimum is 815 (but that includes the Russian Federation contribution)
- Each event year is spaced 4 years apart
- The longest country name is People's Republic of China, the shortest is France

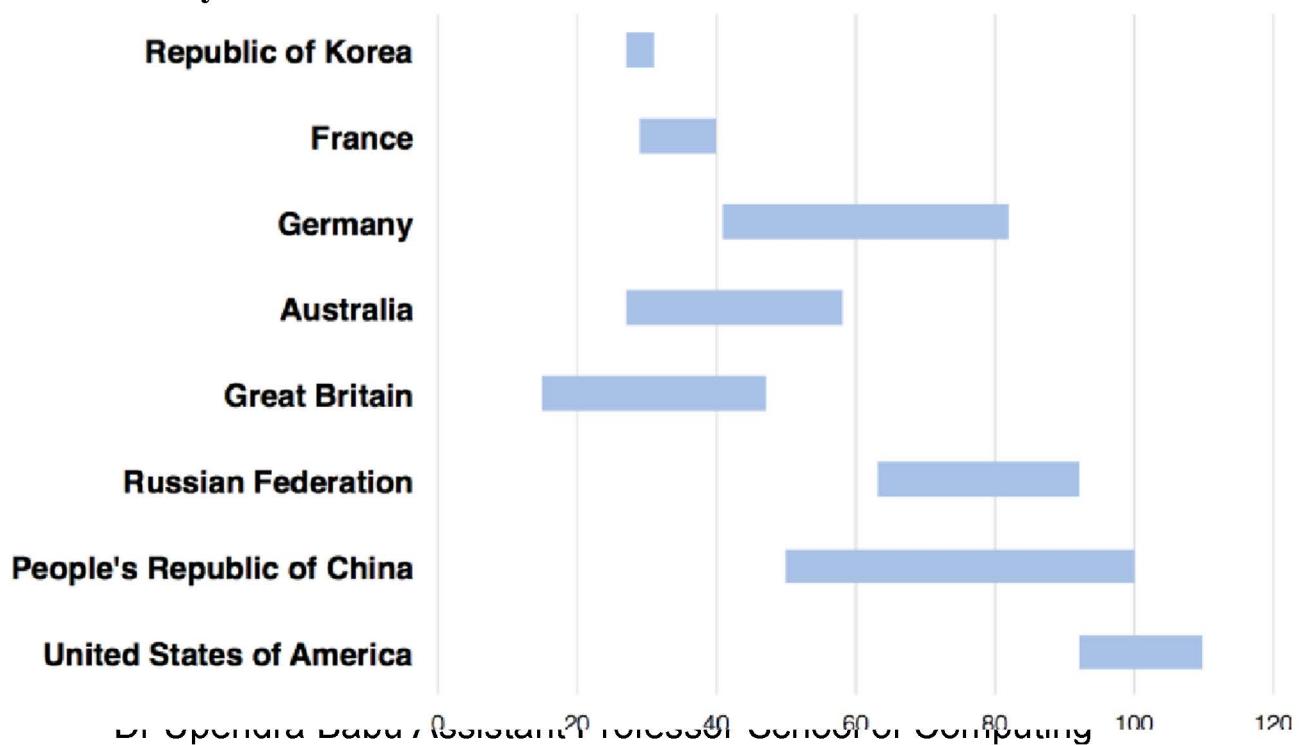
**Many more options can be included to enhance and contextualize this subject matter.**

- The details behind the medal totals of how many golds, silvers, and bronzes each country has won
- The full dataset of medal statistics for all the other countries who have competed, not just the recent top eight
- The full dataset of medal statistics for every Olympic games
- The number of competitors who were taking part in the games for each country, in order to understand the percentage of success of each team
- The split of performances between the different sporting events
- Population figures to contextualize the achievements, maybe even sporting participation figures if they were recorded
- Historical milestones of socio-political and geo-political issues to help us appreciate the status of the different countries at these key points in time
- You might look to bolster the ingredients of your visualization design resources with national flags' image files or URL links to national Olympic associations

**Next, develop editorial focus,**

- What initial sparks of curiosity crossed our minds when we were given the brief and initially saw the data?
- What dimensions of analysis do we think might be of interest or relevance about this subject matter?
- What data questions will we seek to answer in our visualization design?

**The first graphic takes a look at the variation of medal winning across the years**



## Question Answer

Which countries have experienced a significant change in their medal-winning performance levels?

We're looking for the widest bars to show the variability, this could be improvement, decline or inconsistency. We would identify the spread of Germany and China as being particularly interesting.

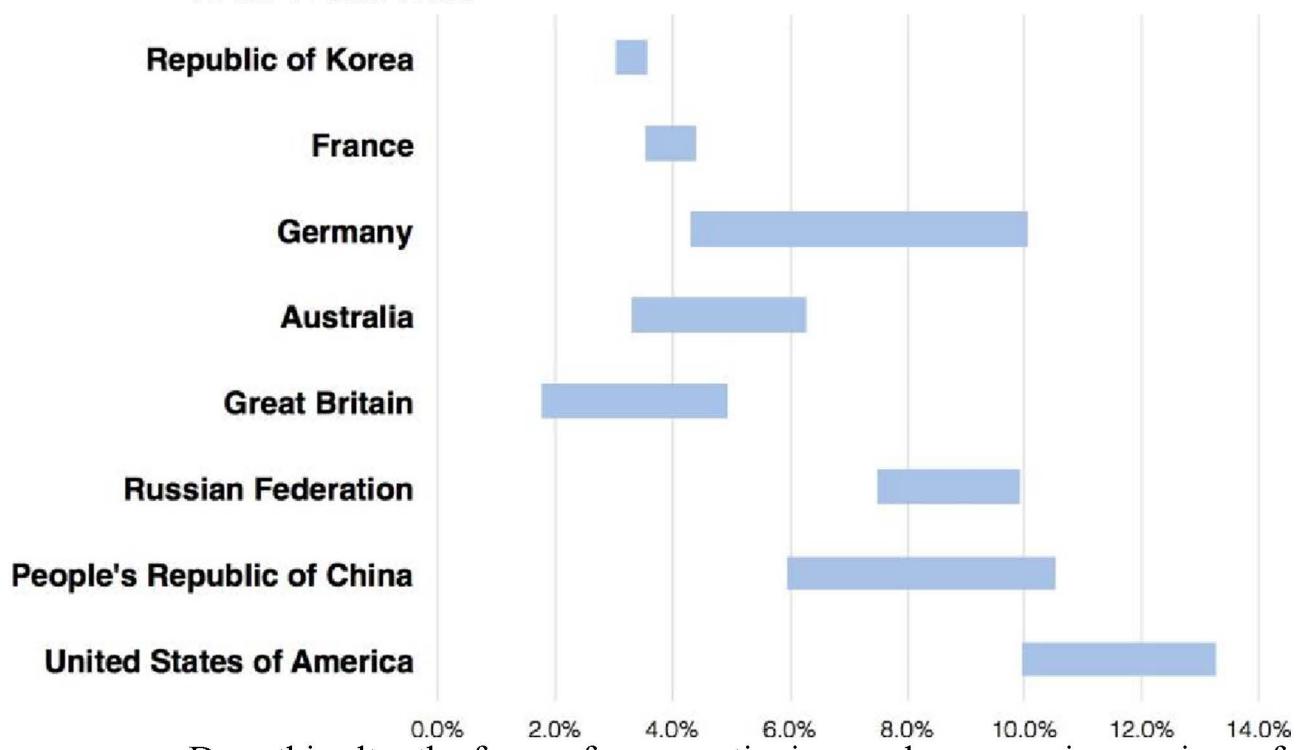
Which countries have maintained consistency in their performance levels?

Now we're looking for the narrowest bars, the tightest of value ranges. This leads to noticing the USA, France, and especially Republic of Korea

What have been the most interesting country stories in terms of the transition of their performance and rankings?

Possibly too hard to see with this chart, but there is potentially something going on with the bars that intersect and exceed the lengths of others. At this stage, the story of China seems to stand out as being something to look out for.

Let's now repeat the same chart to show the medals won as a percentage of the overall total:



Does this alter the focus of our questioning or change our impressions of the main insights?

It reinforces our interest in the varied performance levels for Germany and China. It also emphasizes the remarkable consistency of Republic of Korea and France.

At this point, we have established a scent for the story.

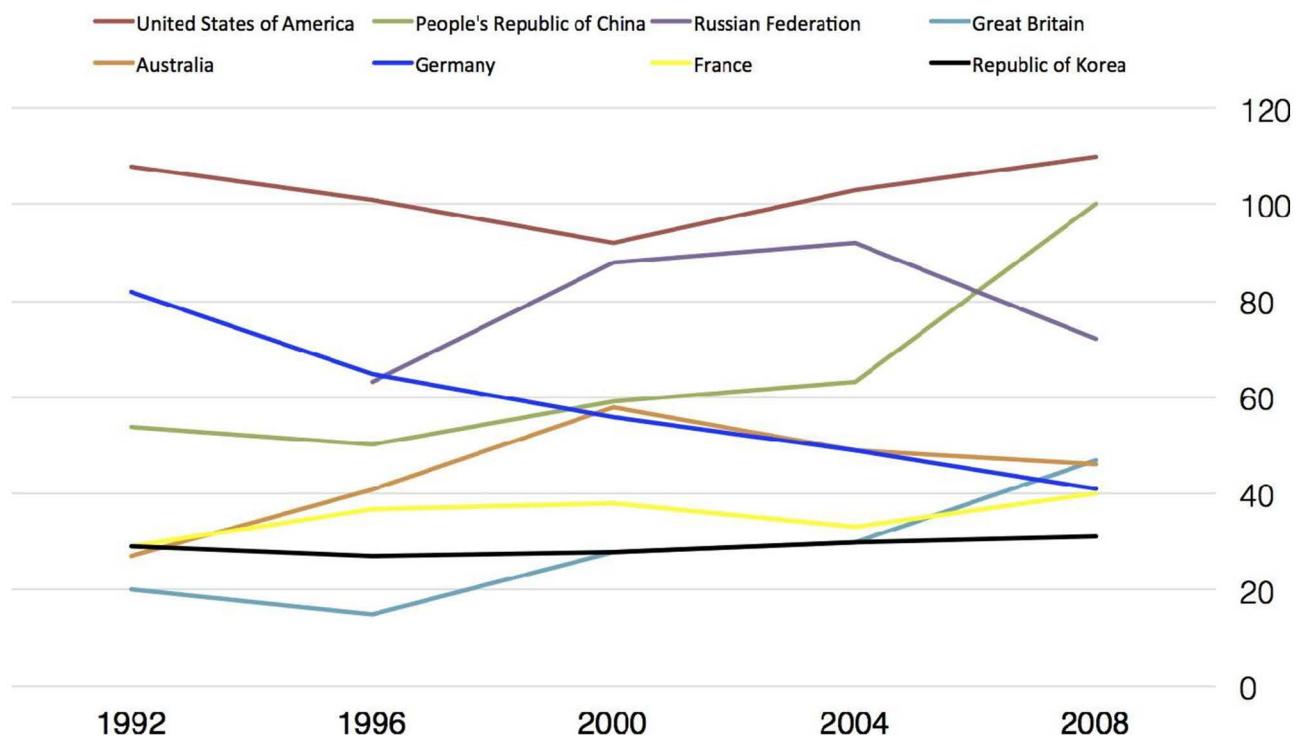
We have started to articulate the data questions that best interrogate this data and most likely reflect what the readers of a visualization about this subject will wish to learn.

Need a different visual representation

Need to switch our perspective to the other main variable and curiosities about

- the transition of medal-winning performances and
- the transition in ranking of the individual countries across the five Olympic Games

Visual sketch - line chart. On a single chart we plot the eight countries, differentiated by colour, showing the absolute medal wins from left to right across the five Olympic events:



This is an exploratory visualization for ourselves

See if we can discover some interesting physical properties about the data in this display

The chart shows Germany's (blue) results reflects their general decline

China's distribution shows a rise over the past four games

This depicts a fascinating trend

"Different forms do better jobs and answering different questions"

Let's reduce the story to a simple contrast between China and Germany. Our main data question will be something like "how have the medal-winning performances of China and Germany compared over the past five events?"

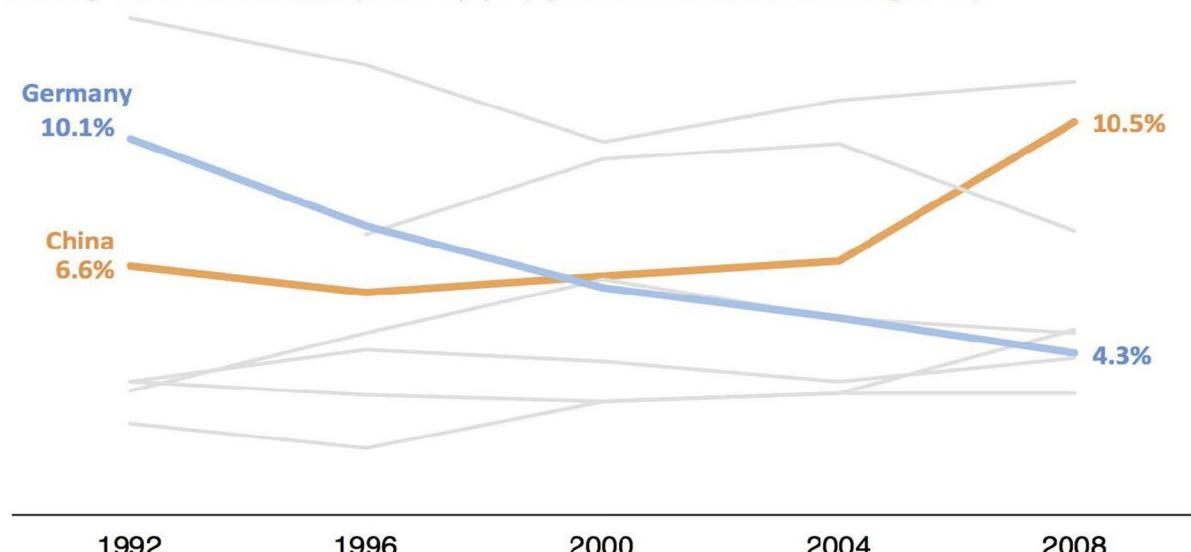
most suitable method for giving form to and answering this question will be a line chart

This time we're conveying the story to others, so we need to refine the visuals in order to make it explanatory:

- We need to elevate the important features of the main story and relegate any background context and secondary content.
- We need to ensure that there are annotations for labels, values, and captions so the reader is entirely clear about what is being communicated.

### The Contrasting Fortunes of German and Chinese Olympic Success

Percentage of total medals won across past five Olympics (eight countries selected based on ranking at 2008)



Things to point out

Calculated data for medals won as a percentage of the total

**Standardize and contextualize the performance across all events**

Aim is to

Provide a clear **visual hierarchy emphasizing the two main countries in our story**

Diminishing the contextualizing other six nations into the background

A subtle shade of grey for other six countries, provides sense of the overall rankings (context).

#### Title

**Provide a title that neatly frames the story**

Subheading that describes the chart and the data

Labels to help the reader compare the two countries' relative trajectory.

#### Colour

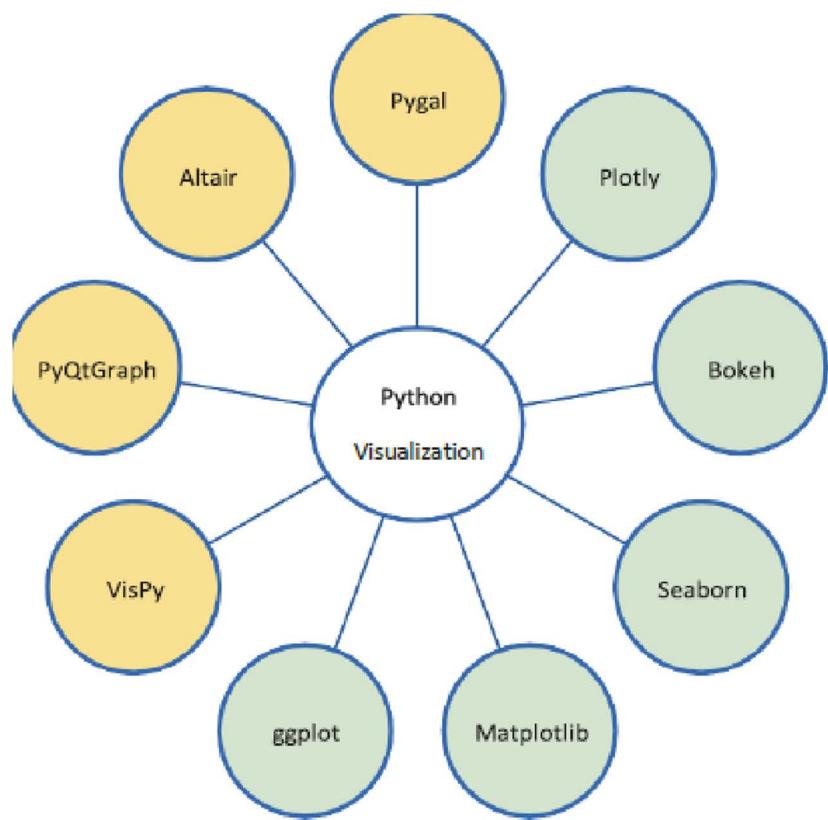
**Selective use of colour** to help imply the positive improvement (orange = hot = good) of China and the negative decline (blue = cold = bad) of Germany.

## Popular Libraries for Data Visualization in Python

The Python language provides numerous data visualization libraries for plotting data.

The most used and common data visualization libraries are

- Pygal
- Altair
- VisPy
- PyQtGraph
- Matplotlib
- Bokeh
- Seaborn
- Plotly, &
- ggplot,



**In the current scope we shall briefly study**

- Pandas
- Lambdas &
- NumPy