

SESSIONS 5

Introduction to Data Visualization

Data Science Developer

Outline

- Basic Concept of Visualization
- Tools for visualization
- Type of Visualization

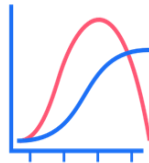
What is Data Visualization?



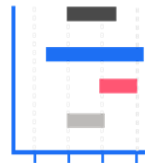
Bar chart



Stacked bar chart



Line graph



Gantt chart



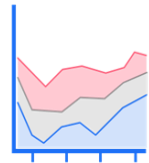
Polar area diagram



Scatter plot



Calendar heatmap



Stacked area chart



Sparkline



Column sparkline

- Data Visualization is the presentation of data in pictorial or graphical format
- Data Visualization is a discipline to understand data by serving it visually so any **pattern or trends, composition, comparison** and **relationship** can be exposed

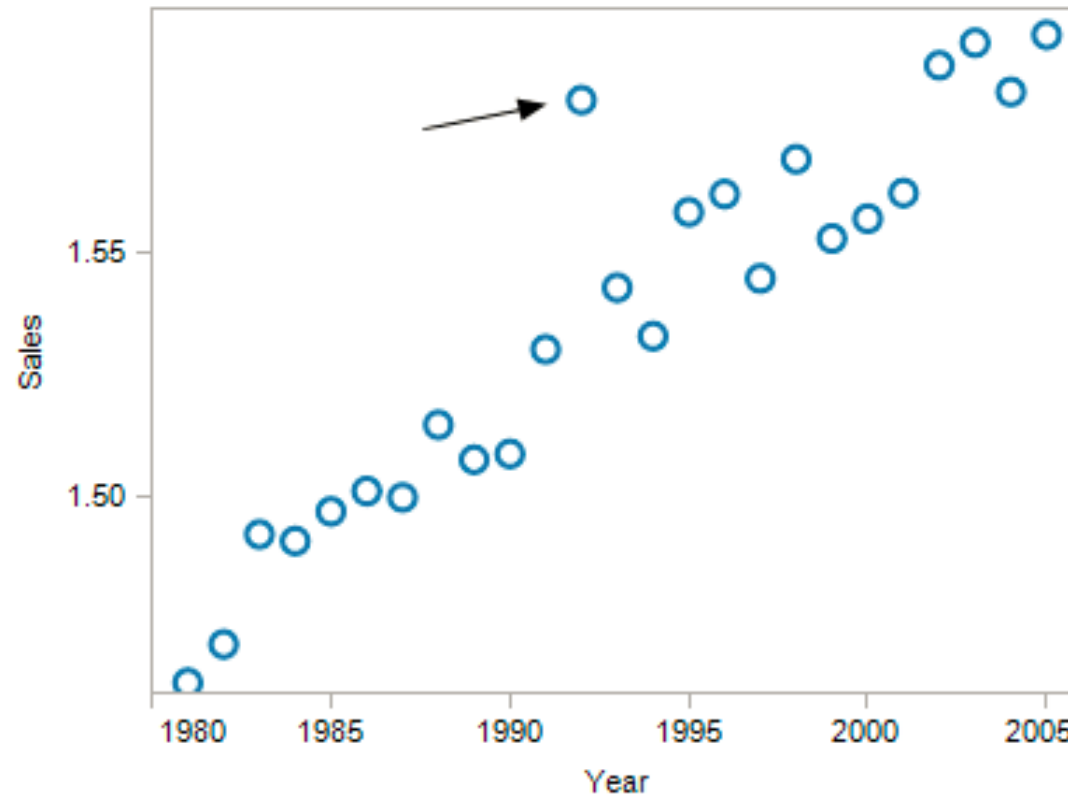
Why is Data Visualization Very Important ?

- You can utilize data visualization to **explore your data**. Exploring your data can improve your chance to find insight.
- You can also **share your findings** with other by using data visualization. Human brain is easier to process any information using visualization than spreadsheet or report
- Quick and easy way to convey concept in a universal manner.



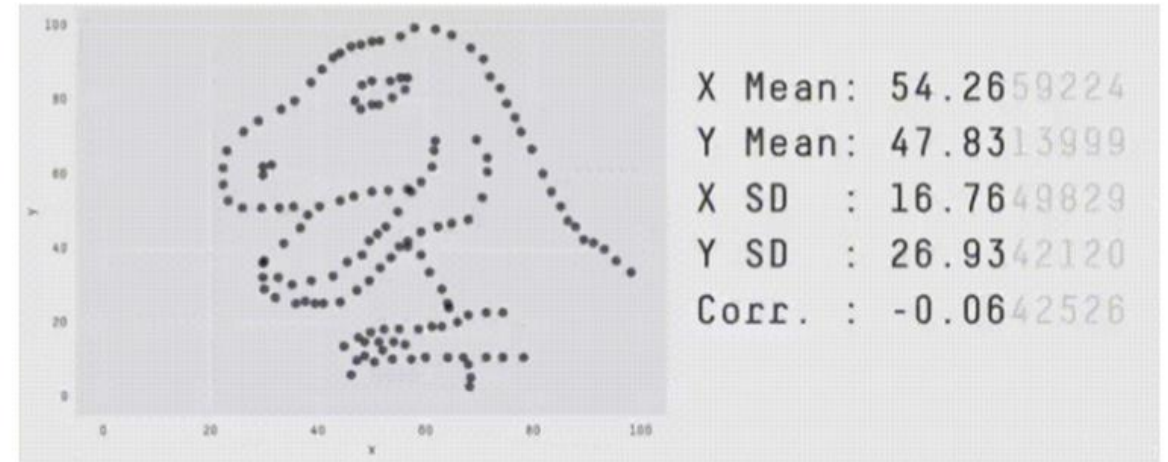
Which method is easier to identify uncommon event (anomaly) between year and sales?

	A	B
1	Year	Sales
2	1981	1.4622
3	1982	1.47004
4	1983	1.49253
5	1984	1.49118
6	1985	1.49722
7	1986	1.50138
8	1987	1.50008
9	1988	1.51493
10	1989	1.50781
11	1990	1.50899
12	1991	1.53037
13	1992	1.58137
14	1993	1.54299
15	1994	1.53307
16	1995	1.55845
17	1996	1.56213
18	1997	1.54488
19	1998	1.56927
20	1999	1.55305
21	2000	1.5571
22	2001	1.56235
23	2002	1.58847
24	2003	1.59309
25	2004	1.58303
26	2005	1.5947



Visualization Can Be More Meaningful

- It's hard to find patterns & derive insights from raw data
- Statistics can summarize data, but may hide patterns in how the data is spread
- We use visual encoding techniques to map data to visual attributes

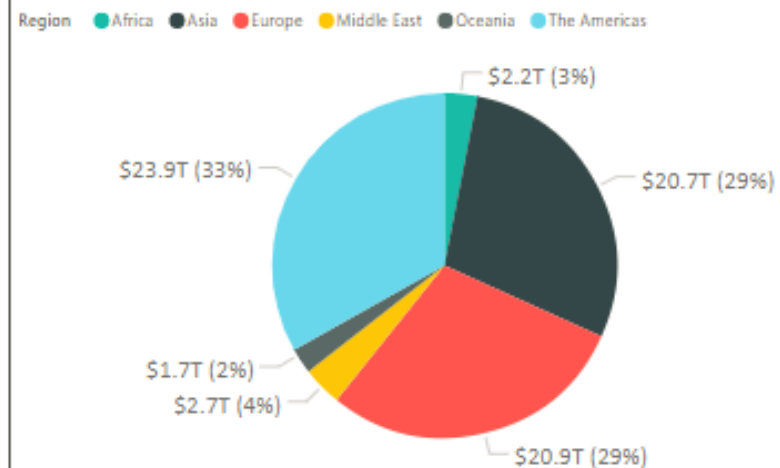


Example : World GDP Indicator 2012

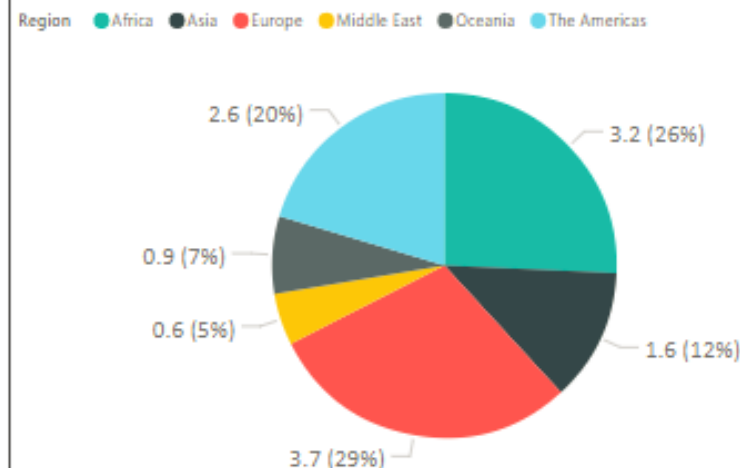
Region	Population 0-14	Population 15-64	Population 65+	Population Urban	GDP	Health Exp % GDP	Health Exp/Capita
Africa	21.30	29.84	1.86	22.04	\$2,232,736,028,370	3.24	\$7,941
Asia	9.01	22.66	2.33	17.10	\$20,689,754,411,401	1.57	\$14,214
Europe	6.72	27.84	6.44	32.79	\$20,930,307,966,112	3.69	\$120,452
Middle East	3.50	9.01	0.49	10.22	\$2,681,972,907,100	0.61	\$11,424
Oceania	4.00	8.13	0.87	7.75	\$1,727,175,612,930	0.92	\$11,652
The Americas	9.88	24.11	3.01	29.48	\$23,926,023,856,769	2.57	\$33,654
Total	54.41	121.58	15.00	119.38	\$72,187,970,782,682	12.59	\$199,337

QUESTION:
What region has
the highest and
lowest GDP and
Health Exp %
GDP?

Example : World GDP Indicator 2012 (GDP)



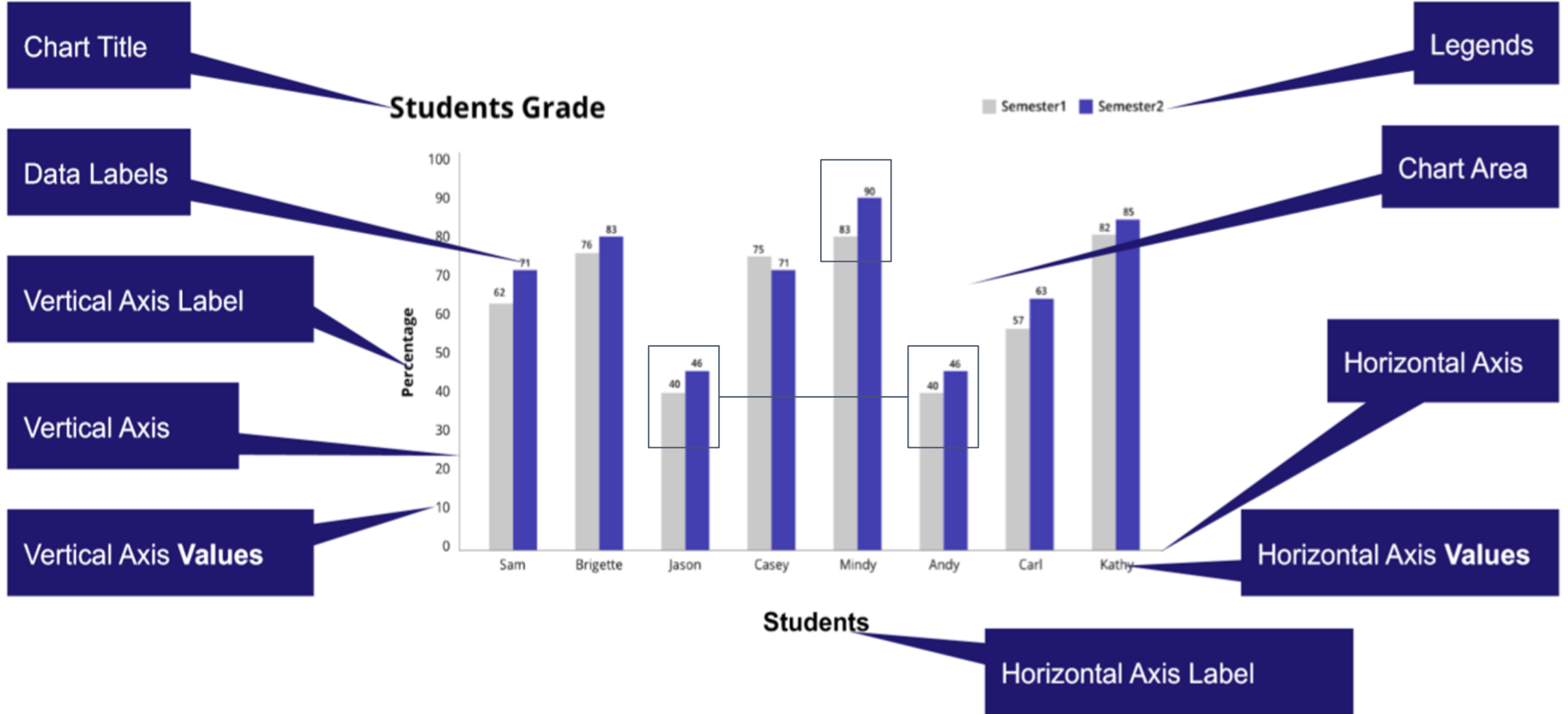
Example : World GDP Indicator 2012 (Health Exp % GDP)



Source data : World Indicator (Power BI)

How to do Data Visualization?

- Understand the context of your data
- Making some questions for you data
- Choose appropriate type of visualization and identify the message of each visualization you made
- From technical perspective:
 - ✓ Add title to your data visualization to make it more informative
 - ✓ Label your axis (x-axis and y-axis)
 - ✓ Label your graph (legend)
 - ✓ Mark interesting data points
 - ✓ Play with color and size
- Get conclusion



Some questions that can be answered through data visualization

How is the trend of the sales this past 3 months ?

Which factor influence customer behavior ?

How to place your product to improve sales ?

Which products has the highest sales ?

Tool for Visualization

Python Libraries:

- Matplotlib
- Seaborn
- Pandas
- Plotly
- Bokeh
- Etc

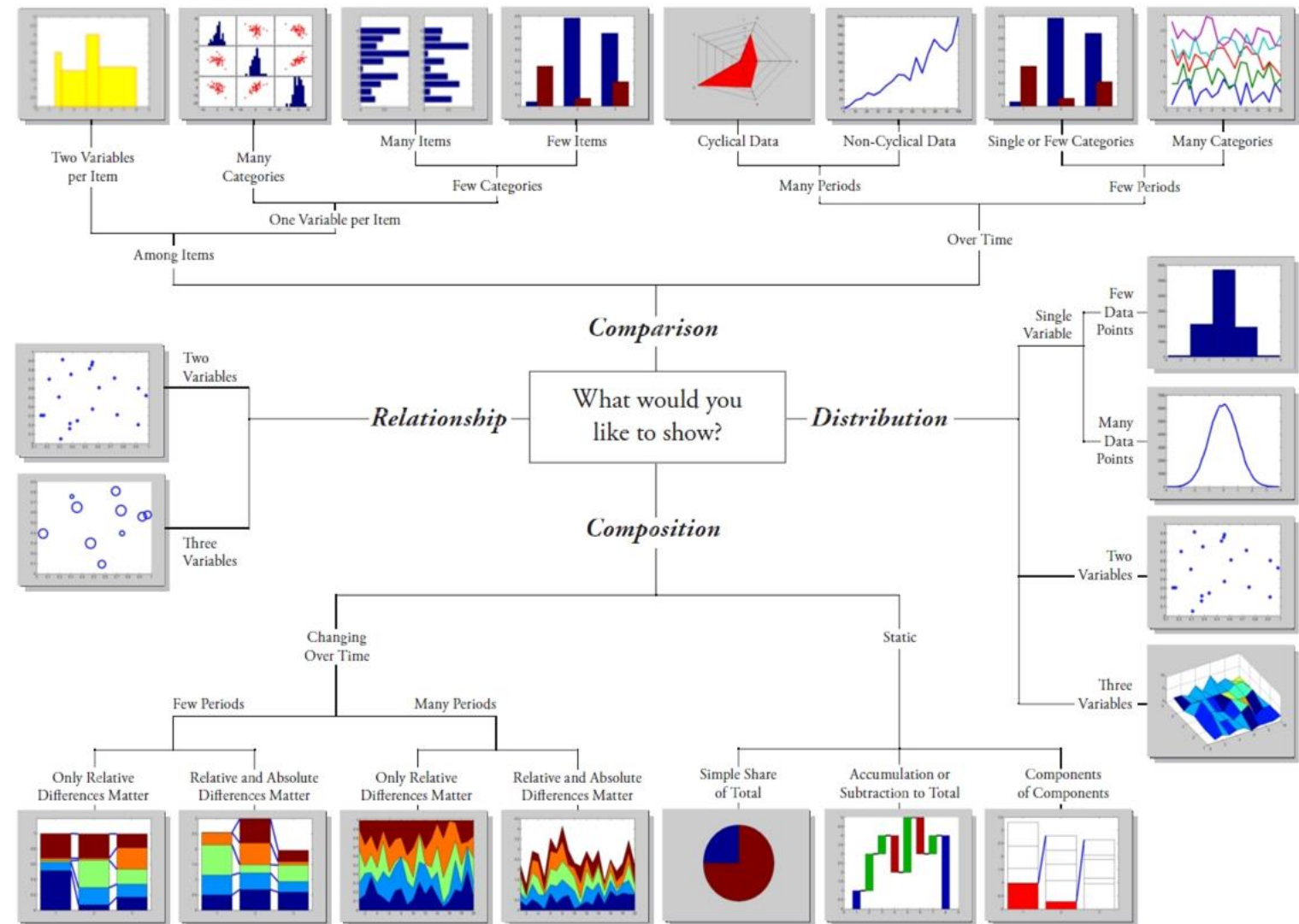
Commonly used:

- Power BI
- Tableau
- Microstrategy
- Qlik
- Etc

Chart Suggestions—A Thought-Starter

Data Visualization Types:

1. Comparison
2. Relationship
3. Distribution
4. Composition



Modified with permission -Doug Hull
 blogs.mathworks.com/videos
 hull@mathworks.com 2009

www.ExtremePresentation.com
 © 2009 A. Abela — a.v.abela@gmail.com

Data Visualization Types

In this course, we will learn how to create and interpret:

1. Histogram
2. Boxplot
3. Lineplot
4. Scatterplot
5. Barplot
6. Pie Chart
7. Heatmap
8. Folium (Geographical Map)