

# Data Visualization and Big Data Analytics

# Unit-1 INTRODUCTION TO VISUALIZATION

- CONCEPT : Importance of data visualization
- choosing appropriate visual encodings
- ordering of items
- number of distinct values
- Structure of visualization

## POSTIONING :

- Placement and proximity
- Graphs and layouts
- Colors
- Size
- Text and typography
- Shape
- Lines

# Data visualization

- Data visualization is **the process of translating large data sets and metrics into charts, graphs and other visuals**. The resulting visual representation of data makes it easier to identify and share real-time trends, outliers, and new insights about the information represented in the data.
- To communicate information clearly and efficiently, data visualization uses statistical graphics, plots, information graphics and other tools.
- Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message.
- Effective visualization helps users analyze and reason about data and evidence.

# WHAT IS DATA VISUALIZATION?



- **Data visualization** refers to the techniques used to communicate data or information by encoding it as visual objects (e.g., points, lines, or bars) contained in graphics.
- The **goal** is to communicate information clearly and efficiently to users

# Data visualization

- WHAT?
- WHY?
- HOW?

## Data Visualization



You are a Sales Manager in a leading global organization. The organization plans to study the sales details of each product across all regions and countries. This is to identify the product which has the highest sales in a particular region and up the production. This research will enable the organization to increase the manufacturing of that product in the particular region.

# Data Visualization



## **Data Visualization Considerations**

Three major considerations for data visualization are:



Clarity



Accuracy



Efficiency

Ensure the dataset is complete and relevant. This enables the Data Scientist to use the new patterns yield from the data in the relevant places.



## **Data Visualization Considerations**

Three major considerations for data visualization are:



Clarity



Accuracy



Efficiency

Ensure using appropriate graphical representation to convey the right message.

## **Data Visualization Considerations**

Three major considerations for data visualization are:



Clarity



Accuracy



Efficiency

Use efficient visualization technique which highlights all the data points.

# Why is Data Visualization Important?

- What business or career you've chosen, data visualization can help by delivering data in the most efficient way possible.
- As one of the essential steps in the business intelligence process, data visualization takes the raw data, models it, and delivers the data so that conclusions can be reached.
- In advanced analytics, data scientists are creating machine learning algorithms to better compile essential data into visualizations that are easier to understand and interpret.
- **Data visualization uses visual data to communicate information in a manner that is universal, fast, and effective.**
- Visualized data gives stakeholders, business owners, and decision-makers a better prediction of sales volumes and future growth.

# What Are The Benefits of Data Visualization?

- Data visualization **positively affects an organization's decision-making process with interactive visual representations of data.**
- Businesses can now recognize patterns more quickly because they can interpret data in **graphical or pictorial forms.**
- Here are some more specific ways that data visualization can benefit an organization:
  - 1. **Correlations in Relationships:**
  - 2. **Trends Over Time:**
  - 3. **Frequency:**
  - 4. **Examining the Market:**
  - 5. **Risk and Reward:**
  - 6. **Reacting to the Market:**

# Which Data Visualization Techniques are Used?

- There are many different methods of putting together information in a way that the data can be visualized.
- Depending on the data being modeled, and what its intended purpose is, **a variety of different graphs and tables may be utilized to create an easy to interpret dashboard.**
- Some visualizations are manually created, while others are automated.
- Either way, there are many types to meet your visualization needs.
- **1. Infographics:**
- **2. Heatmap Visualization:**
- **3. Fever Charts:**
- **4. Area Chart (or Graph):**
- **5. Histogram:**

- **RESULT :**
- **We need data visualization because the human brain is not well equipped to devour so much raw, unorganized information and turn it into something usable and understandable.**
- We need graphs and charts to communicate data findings so that we can identify patterns and trends to gain insight and make better decisions faster.

# Visual Encoding

- The visual encoding is the way in which data is mapped into visual structures, upon which we build the images on a screen.
- There are two types of visual encoding variables:
  - planar and retinal.
  - Humans are sensitive to the retinal variables.
  - They easily differentiate between various colors, shapes, sizes and other properties.
- Retinal variables were introduced by Bertin (→) about 40 years ago, and this concept has become quite popular recently. While there's some critique about the effectiveness of retinal variables (→), most specialists find them useful.

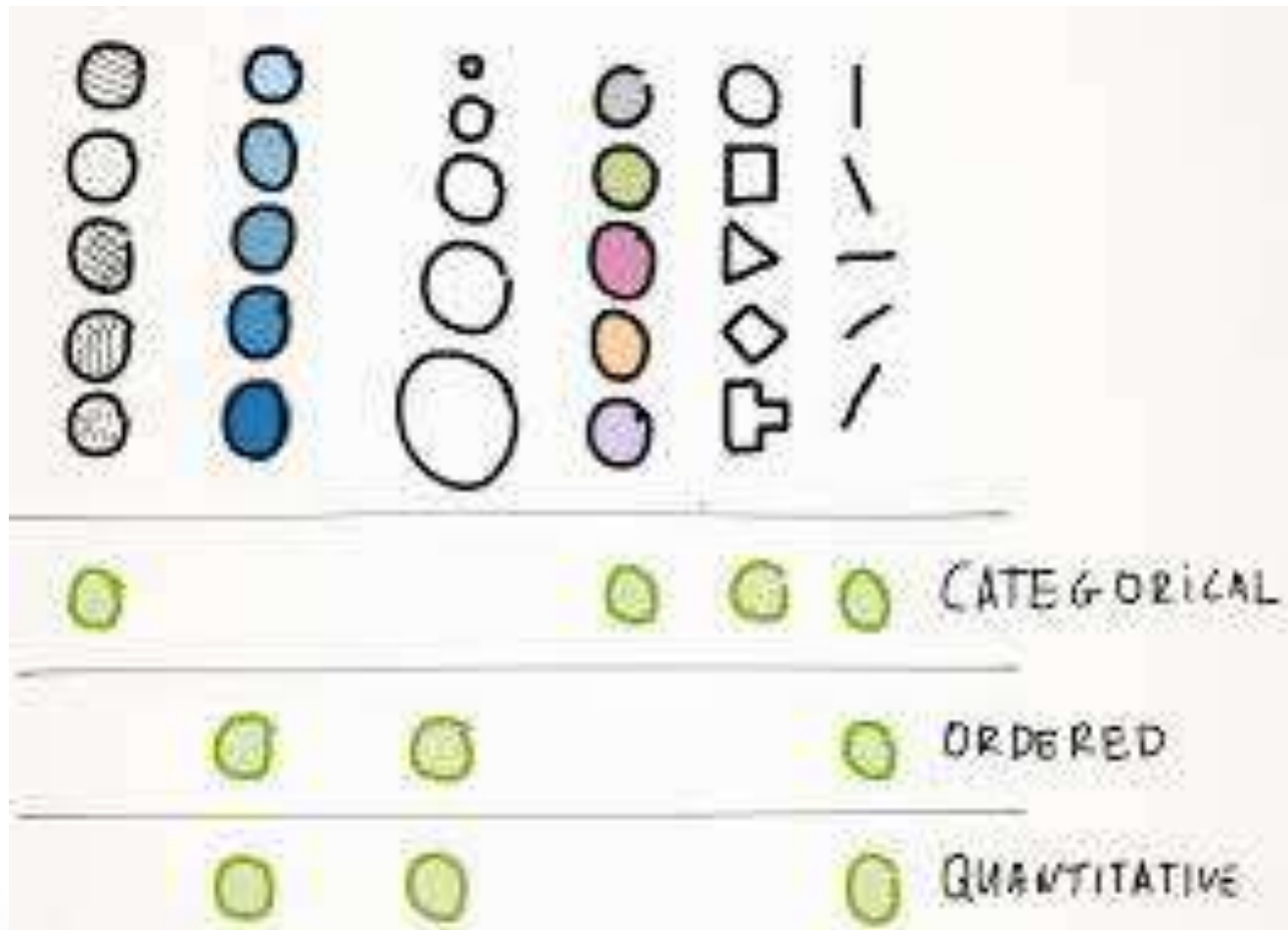
- Data, once you know the “shape” of your data, you can encode its various dimensions with appropriate visual properties. Different visual properties vary—or may be modified—in different ways, which makes them good for encoding different types of data.
- Two key factors are whether a visual property is naturally ordered, and how many distinct values of this property the reader can easily differentiate.
- **Natural ordering and number of distinct values** will indicate whether a visual property is best suited to one of the main data types: quantitative, ordinal, categorical, or relational data. (Spatial data is another common data type, and is usually best represented with some kind of map.)



# Natural Ordering

- A visual property has a natural ordering is determined by whether the mechanics of our visual system and the “software” in our brains automatically—unintentionally—assign an order, or ranking, to different values of that property.
- The “software” that makes these judgments is deeply embedded in our brains and evaluates relative order independent of language, culture, convention, or other learned factors; it’s not optional and you can’t design around it.

# Example visual encoding



[illegible]

# Examples

- For example:
- 1. position has a natural ordering; but shape doesn't.
- 2. Length has a natural ordering; texture doesn't (but pattern density does).
- 3. Line thickness or weight has a natural ordering; line style (solid, dotted, dashed) doesn't.
- Depending on the specifics of the visual property, its natural ordering may be well suited to representing quantitative differences (27, 33, 41), or ordinal differences (small, medium, large, enormous).

# Color is not ordered

- Color (hue) is not naturally ordered in our brains but the Brightness (lightness or luminance, sometimes called tint) and intensity (saturation) are, but color itself is not.

- **Distinct Values**

- The second main factor to consider when choosing a visual property is how many distinct values it has that your reader will be able to perceive, differentiate, and possibly remember.
- For example, there are a lot of colors in the world, but we can't tell them apart if they're too similar and We can more easily differentiate a large number of shapes, a huge number of positions, and an infinite number of numbers while choosing a visual property, select one that has a number of useful differentiable values and an ordering similar to that of your data

## ordering items:

### VARIABLES OF THE IMAGE

**XY** 2 DIMENSIONS OF THE PLANE

**Z**

SIZE

VALUE





























### DIFFERENTIAL VARIABLES

TEXTURE













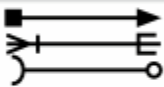

COLOR

ORIENTATION

SHAPE

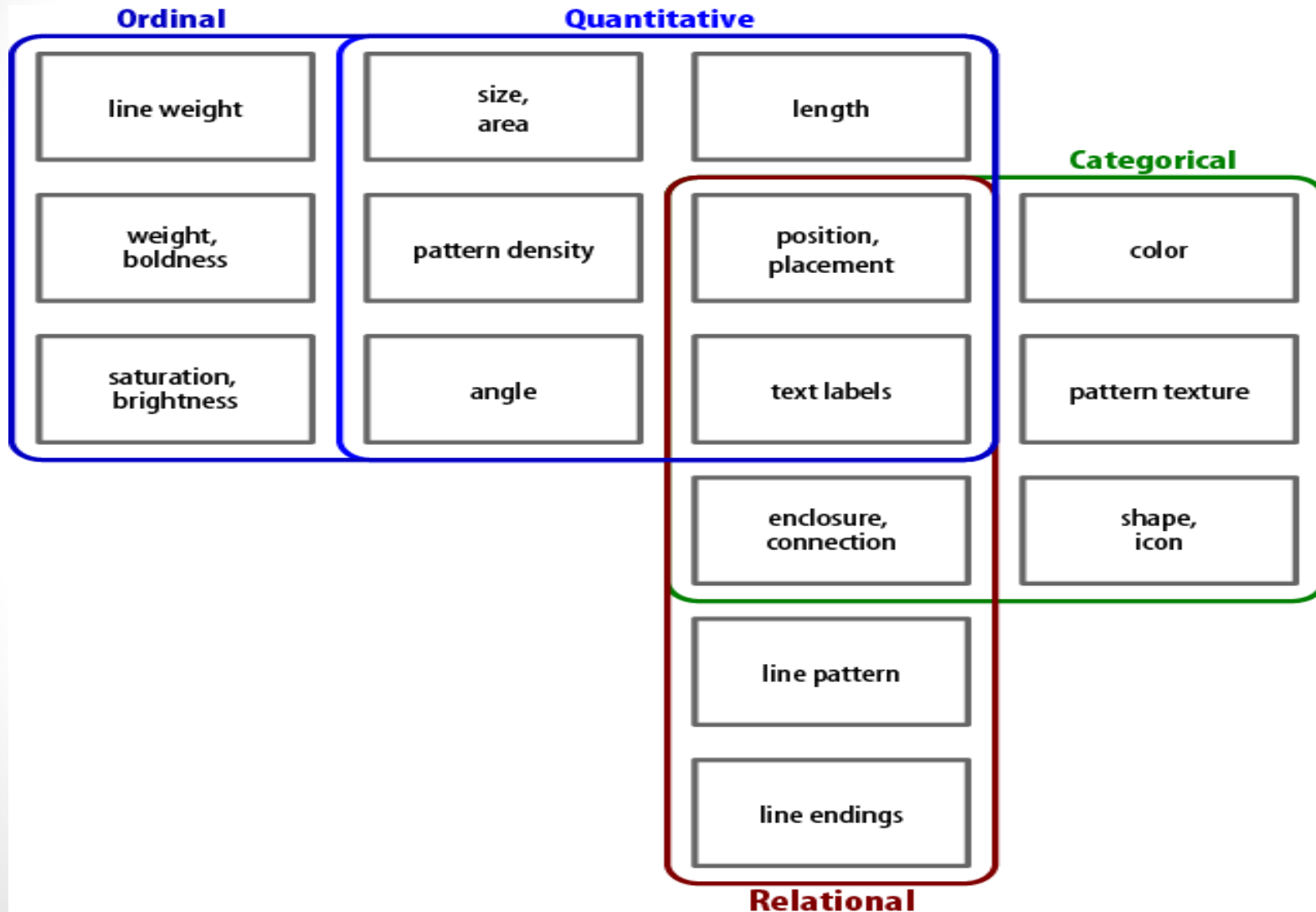
	POINT	LINE	AREA (ZONE)
			
			
			
			
			
			
			

# Properties and Best Uses of Visual Encodings

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional (alphabetical or numbered)	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium/few	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		



- Visual properties grouped by the types of data they can be used to encode.





# positioning

- Many visual properties may be used to encode multiple data types.
- Position and placement, text, can be used to encode any type of data—which is refereeing that why every visualization you design needs to begin with careful considerations.

# postioning

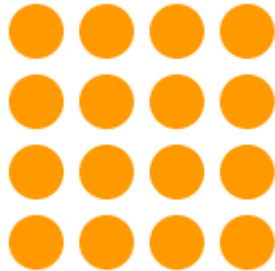
To further illustrate this process, check out the GIF below. Your brain sees a dog walking, but it's nothing more than a series of moving dots.



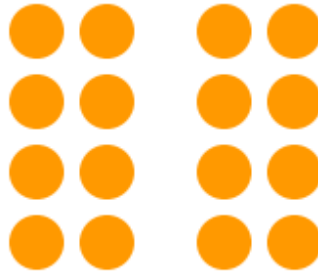
# proximity

The principle of proximity states that things that are close together appear to be more related than things that are spaced farther apart.

*This is perceived to be one group and the components somehow related to each other.*



*We perceive two groups here, and understand that there are differences between them.*



# proximity

Proximity is so powerful that it overrides similarity of color, shape, and other factors that might differentiate a group of objects.

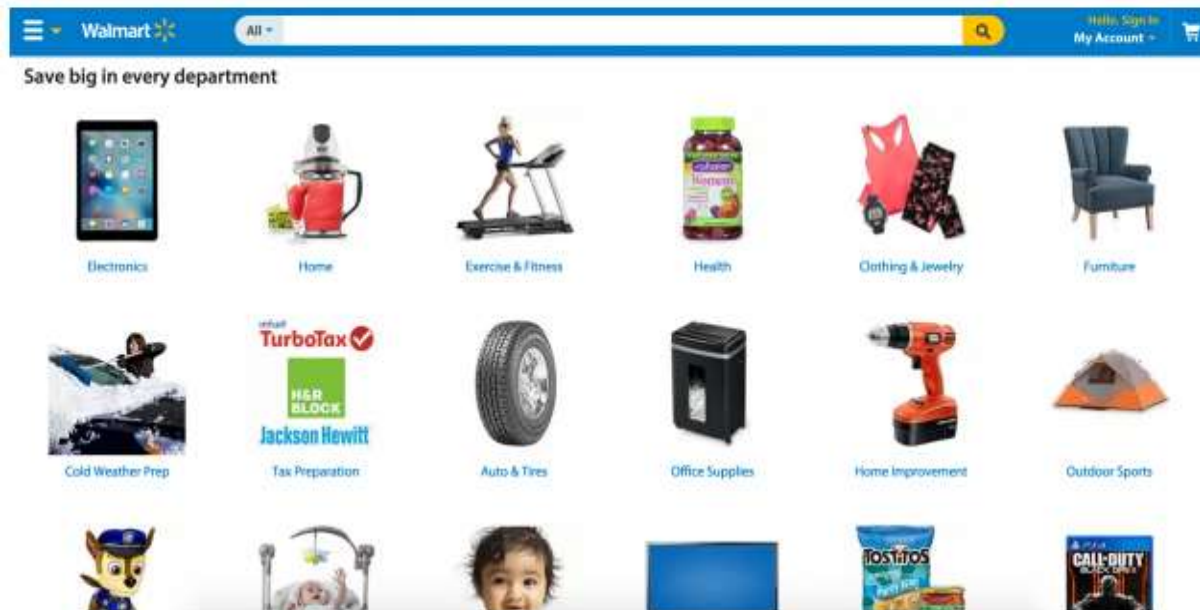


*Source: Steven Bradley*

Notice the three groups of black and red dots above? The relative nearness of the objects has an even stronger influence on grouping than color does.

# Example of proximity

The nearness of each image and its corresponding text communicates that they're related to one another.



And Vice uses it to distinguish between the images, headlines, descriptions, and other information for each of its stories.

# HTML5 & CSS3 READINESS

2008 / 2009 / 2010 / 2011

Many of these features are ready to implement today.

We don't have to wait for entire specs to be completed, we can start using some hawtness now.

data from [caniuse.com](http://caniuse.com)

- IE 7
- IE 8
- IE 9
- FIREFOX 3.6
- FIREFOX 4.0
- OPERA 11
- SAFARI 5
- CHROME

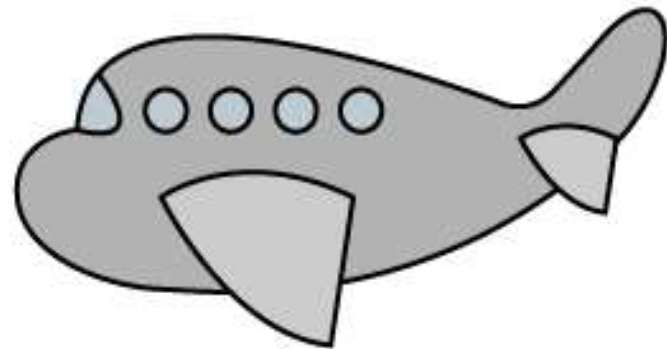
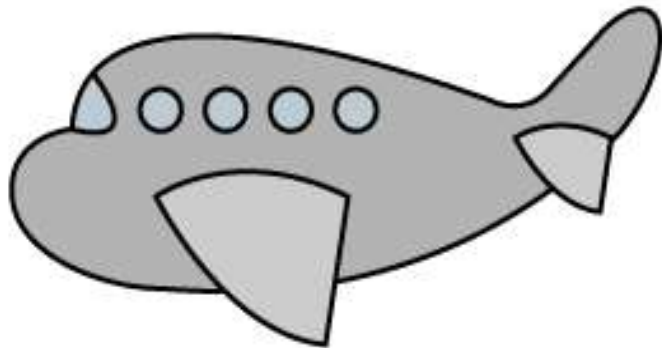


FROM THE ASYLUM RUN BY **PAUL IRISH** AND **DIVYA MANIAN** | [ABOUT THIS SITE](#)
















HTML



# layouts



# shapes and colors examples

Browser	Existing color	Better color
		
		
		
		
		



# color -size-text-lines

## Periodic Table of Elements

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
1 <b>H</b> Hydrogen 1.00794	Atomic # Symbol Name Atomic Mass																2 <b>He</b> Helium 4.002602		
3 <b>Li</b> Lithium 6.941	4 <b>Be</b> Beryllium 9.012182	<div><div>C Solid</div><div>Hg Liquid</div><div>H Gas</div><div>Rf Unknown</div></div>																10 <b>Ne</b> Neon 20.1797	
11 <b>Na</b> Sodium 22.98976928	12 <b>Mg</b> Magnesium 24.3050	<div>Alkali metals</div> <div>Alkaline earth metals</div> <div>Lanthanoids</div> <div>Actinoids</div> <div>Transition metals</div> <div>Poor metals</div> <div>Other nonmetals</div> <div>Noble gases</div>																18 <b>Ar</b> Argon 39.948	
19 <b>K</b> Potassium 39.0983	20 <b>Ca</b> Calcium 40.078	21 <b>Sc</b> Scandium 44.955912	22 <b>Ti</b> Titanium 47.867	23 <b>V</b> Vanadium 50.9415	24 <b>Cr</b> Chromium 51.9961	25 <b>Mn</b> Manganese 54.938045	26 <b>Fe</b> Iron 55.845	27 <b>Co</b> Cobalt 58.933195	28 <b>Ni</b> Nickel 58.6934	29 <b>Cu</b> Copper 63.546	30 <b>Zn</b> Zinc 65.38	31 <b>Ga</b> Gallium 69.723	32 <b>Ge</b> Germanium 72.64	33 <b>As</b> Arsenic 74.92160	34 <b>Se</b> Selenium 78.96	35 <b>Br</b> Bromine 79.904	36 <b>Kr</b> Krypton 83.796		
37 <b>Rb</b> Rubidium 85.4678	38 <b>Sr</b> Strontium 87.62	39 <b>Y</b> Yttrium 88.90585	40 <b>Zr</b> Zirconium 91.224	41 <b>Nb</b> Niobium 92.90538	42 <b>Mo</b> Molybdenum 95.95	43 <b>Tc</b> Technetium (97.9072)	44 <b>Ru</b> Ruthenium 101.07	45 <b>Rh</b> Rhodium 102.90550	46 <b>Pd</b> Palladium 106.42	47 <b>Ag</b> Silver 107.8682	48 <b>Cd</b> Cadmium 112.411	49 <b>In</b> Indium 114.818	50 <b>Sn</b> Tin 118.710	51 <b>Sb</b> Antimony 121.760	52 <b>Te</b> Tellurium 127.60	53 <b>I</b> Iodine 126.90447	54 <b>Xe</b> Xenon 131.293		
55 <b>Cs</b> Caesium 132.9054519	56 <b>Ba</b> Barium 137.327	57–71										79 <b>Au</b> Gold 196.966569	80 <b>Hg</b> Mercury 200.59	81 <b>Tl</b> Thallium 204.3833	82 <b>Pb</b> Lead 207.2	83 <b>Bi</b> Bismuth 208.98040	84 <b>Po</b> Polonium (209, 9824)	85 <b>At</b> Astatine (209, 9671)	86 <b>Rn</b> Radon (222, 0176)
87 <b>Fr</b> Francium (223)	88 <b>Ra</b> Radium (226)	89–103										111 <b>Rg</b> Roentgenium (272)	112 <b>Uub</b> Ununbium (285)	113 <b>Uut</b> Ununtrium (284)	114 <b>Uuq</b> Ununquadium (289)	115 <b>Uup</b> Ununpentium (288)	116 <b>Uuh</b> Ununhexium (292)	117 <b>Uus</b> Ununseptium (294)	118 <b>Uuo</b> Ununoctium (294)

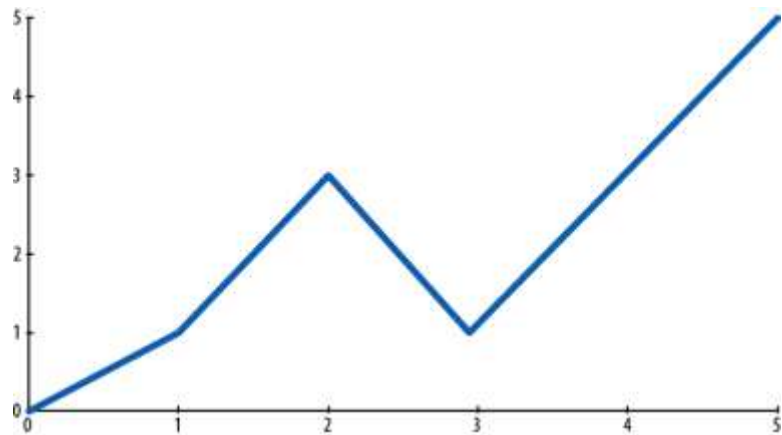
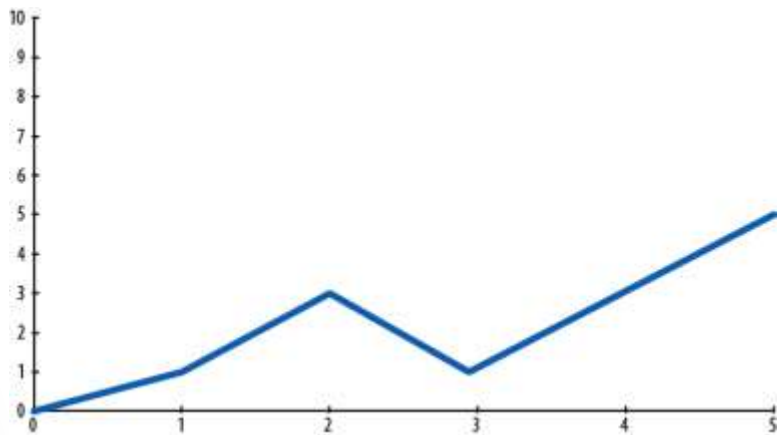
For elements with no stable isotopes, the mass number of the isotope with the longest half-life is in parentheses.

Periodic Table Design and Interface Copyright © 1997 Michael Dayah. <http://www.ptable.com/> Last updated: May 27, 2008

Ptable.com

57 <b>La</b> Lanthanum 138.90547	58 <b>Ce</b> Cerium 140.116	59 <b>Pr</b> Praseodymium 140.90765	60 <b>Nd</b> Neodymium 144.242	61 <b>Pm</b> Promethium (145)	62 <b>Sm</b> Samarium 150.36	63 <b>Eu</b> Europium 151.964	64 <b>Gd</b> Gadolinium 157.25	65 <b>Tb</b> Terbium 158.92535	66 <b>Dy</b> Dysprosium 162.500	67 <b>Ho</b> Holmium 164.93032	68 <b>Er</b> Erbium 167.259	69 <b>Tm</b> Thulium 168.93421	70 <b>Yb</b> Ytterbium 173.054	71 <b>Lu</b> Lutetium 174.9668
89 <b>Ac</b> Actinium (227)	90 <b>Th</b> Thorium 232.03806	91 <b>Pa</b> Protactinium 231.03588	92 <b>U</b> Uranium 238.02891	93 <b>Np</b> Neptunium (237)	94 <b>Pu</b> Plutonium (244)	95 <b>Am</b> Americium (243)	96 <b>Cm</b> Curium (247)	97 <b>Bk</b> Berkelium (247)	98 <b>Cf</b> Californium (251)	99 <b>Es</b> Einsteinium (252)	100 <b>Fm</b> Fermium (257)	101 <b>Md</b> Mendelevium (258)	102 <b>No</b> Nobelium (259)	103 <b>Lr</b> Lawrencium (262)

# Graphs and layout differentiation



# Typograpghy

- Typography is one of the most effective tools for communicating the meaning of data to users.
- When building data visualizations, establishing a clear hierarchy of typographic elements and using easy-to-read font families will help users better understand and interact with your data.
- The term **typography** is also applied to the style, arrangement, and appearance of the letters, numbers, and symbols created by the process.

- Typography include a broad range, covering all aspects of letter design and application, both mechanical (typesetting, type design, and typefaces) and manual (handwriting and calligraphy). Typographical elements may appear in a wide variety of situations, including:
  - Documents
  - Presentations
  - Display typography (described below)
  - Clothing
  - Maps and labels
  - Vehicle instrument panels
  - As a component of industrial design—type on household appliances, pens, and wristwatches, for example
  - As a component in modern poetry (see, for example, the poetry of e. e. cummings)
- Since digitization, typographical uses have spread to a wider range of applications, appearing on web pages, LCD mobile phone screens, and hand-held video games.





# Display graphics

- Type may be combined with negative space and images, forming relationships and dialog between the words and images for special effects. Display designs are a potent element in graphic design

- Display typography encompasses:
  - Advertisements in publications, such as newspapers and magazines
  - Magazine and newspaper headline type
  - Signs and other large-scale-letter designs, such as information signs and billboards
  - Posters
  - Brochures and flyers
  - Packaging and labeling
  - Business communications and advertising
  - Book covers
  - Typographic logos, trademarks, and word marks



- Typography has long been a vital part of promotional material and advertising. Designers often use typefaces to set a theme and mood in an advertisement (for example, using bold, large text to convey a particular message to the reader).
- Choice of typeface is often used to draw attention to a particular advertisement, combined with efficient use of color, shapes, and images.[50] Today, typography in advertising often reflects a company's brand.



# typography adverting video

- <https://motioncue.com/best-kinetic-typography-videos-to-fire-up-your-artistry/>