

## Unit – 4

### 1. Define Quantitative and Qualitative data? Give examples?

#### **IF 2 MARKS THEN WRITE THIS**

##### **Data types**

- There are 2 basic types of data: something you can count and something you can just differentiate.

##### **1. Quantitative**

##### **2. Qualitative**

##### **1. Quantitative**

- *Quantitative* data deals with numbers and things you can measure objectively:
  - Dimensions such as height, width, and length. – Temperature and humidity.
  - Prices.
  - Area and volume.
- Anything that has exact numbers.
  - For example, Effort in points: 0, 1, 2, 3, 5, 8, 13. – Duration in days: 1, 4, 666.
- There are two types of quantitative data, which is also referred to as numeric data: **discrete and continuous**.
  - As a general rule, *counts* are discrete and *measurements* are continuous.
- **Discrete data** is a count that can't be made more precise. Typically it involves integers.
  - For instance, the number of children (or pets) in your family is discrete data, because you are counting whole entities: you can't have 2.5 kids, or 1.3 pets.
- **Continuous data**, on the other hand, could be divided and reduced to finer and finer levels. – For example, you can measure the height of your kids at progressively more precise scales—meters, centimeters, millimeters, and beyond—so height is continuous data.

##### **2. Qualitative**

- *Qualitative* data deals with characteristics and descriptors that can't be easily measured, but can be observed subjectively, such as:
  - smells, tastes, textures, attractiveness, and color.
- Anything that can be compared and ordered. – User Story Priority: Must Have, Great, Good, Not Sure.
- Bug Severity: Blocking, Average, Who Cares.

#### **IF 8 MARKS, WRITE THIS**

##### **TYPES OF DATA**

**Quantitative** – Quantitative data is defined as the value of data in the form of counts or numbers where each data-set has an unique numerical value associated with it. Ex:-Numbers, tests, counting, measuring

##### ➤ Types of Quantitative Data:

- Counter
- Measurement of physical objects
- Sensory calculation
- Projection of data
- Quantification of qualitative entities

Quantitative Data: Collection Methods:

- Surveys : Longitudinal Studies, Cross-sectional Studies. Fundamental Levels of Measurement, Use of Different Question Types,
- Survey Distribution and Survey Data Collection :- Email, Buy respondents, Embed survey in a website, Social distribution, QR code, SMS survey, QuestionPro app, API integration

- One-on-one Interviews:- Face-to-Face Interviews, Online/Telephonic Interviews, Computer Assisted Personal Interview

#### **Steps to conduct Quantitative Data Analysis**

- Relate measurement scales with variables
- Connect descriptive statistics with data
- Decide a measurement scale
- Select appropriate tables to represent data and analyze collected data

#### **Advantages of Quantitative Data**

- Conduct in-depth research
- Minimum bias
- Accurate results

#### **Disadvantages of Quantitative Data**

- Restricted information
- Depends on question types

**Qualitative** – Qualitative data is defined as the data that approximates and characterizes. Ex:-Words, images, observations, conversations, photographs. Qualitative data is defined as the data that approximates and characterizes. Qualitative data can be observed and recorded. This data type is non-numerical in nature. This type of data is collected through methods of observations, one-to-one interviews, conducting focus groups, and similar methods. Qualitative data in statistics is also known as categorical data – data that can be arranged categorically based on the attributes and properties of a thing or a phenomenon. **Importance**- Qualitative data is important in determining the particular frequency of traits or characteristics. It allows the statistician or the researchers to form parameters through which larger data sets can be observed. Qualitative data provides the means by which observers can quantify the world around them.

#### **Qualitative Data Collection Methods-**

- One-to-One Interviews
- Focus groups
- Record keeping
- Process of observation
- Longitudinal studies
- Case studies

**Deductive Approach:** The deductive approach involves analyzing qualitative data based on a structure that is predetermined by the researcher. A researcher can use the questions as a guide for analyzing the data.

**Inductive Approach:** The inductive approach, on the contrary, is not based on a predetermined structure or set ground rules/framework. It is more time consuming and a thorough approach to qualitative data analysis.

#### **Steps to Qualitative Data Analysis**

- Step 1: Arrange your Data
- Step 2: Organize all your Data
- Step 3: Set a Code to the Data Collected
- Step 4: Validate your Data
  - 1.Accuracy of your research design or methods.
  - 2.Reliability, which is the extent to which the methods produce accurate data consistently.
- Step 5: Concluding the Analysis Process

#### **Advantages of Qualitative Data:**

- It helps in-depth analysis
- Understand what customers think
- Rich data

#### **Disadvantages of Qualitative Data:**

- Time-consuming
- Not easy to generalize
- Dependent on the researcher's skills

**Ex-**

The cake is orange, blue, and black in colour (qualitative).

Females have brown, black, blonde, and red hair (qualitative).

Quantitative Data:- There are four cakes and three muffins kept in the basket (quantitative).

One glass of fizzy drink has 97.5 calories (quantitative).

	Qualitative	Quantitative
Conceptual	<p>Concerned with understanding human behaviour from the informant's perspective</p> <p>Assumes a dynamic and negotiated reality</p>	<p>Concerned with discovering facts about social phenomena</p> <p>Assumes a fixed and measurable reality</p>
Methodological	<p>Data are collected through participant observation and interviews</p> <p>Data are analysed by themes from descriptions by informants</p> <p>Data are reported in the language of the informant</p>	<p>Data are collected through measuring things</p> <p>Data are analysed through numerical comparisons and statistical inferences</p> <p>Data are reported through statistical analyses</p>

**REFERENCE :** DPA qn 7,8 & DS SLIDE

## 2. Define few lists of ways you can encode data?

**Data Encoding :** Encoding in data viz basically means translating the data into a visual element on a chart/map/whatever you're making. You need to do it right, because doing it right will mean that other people looking at your visualizations can understand what you're trying to say or show. Another way to think of encoding is as a set of rules to follow. So when you're making them, you need to think logically about how you set up those rules, otherwise people will get totally confused. If you're doing something complicated, a good way of helping yourself to think about how you encode, or set up the rules, is: – Every time <data changes in some way>, do <something visual>. This helps you to be consistent in how you apply the rules. Consider this standard column chart: Most people wouldn't have trouble understanding what it's showing us — but when you break it down, this is encoded in a few different ways: • Colour: Every time <category is bears/dolphins/whales>, change <colour of the column to be blue/orange/grey> • Size: Every time <number goes up>, increase <column height> • Grouping: Every time <month changes>, create <new cluster of columns>

**Here's a (not limited) list of ways you can encode data: {IF NO TIME WRITE THESE ONLY}**

- ✓ Size
- ✓ Shape
- ✓ Color
- ✓ Grouping
- ✓ Area
- ✓ Position
- ✓ Saturation

- ✓ Line pattern
- ✓ Line weight
- ✓ Angle
- ✓ Connections

**Visual Encoding:** Visual encoding is the way in which data is constructed into visual structures. They are the building blocks of graphics (images on a screen).

1. **Planar Encoding** – Planar encoding is as simple as the laying of axis, like the X & Y axis in a simple line chart.
2. **Retinal Encoding** – To represent data in 3 or more variables, retinal encoding comes into the picture. Size, texture, shape, orientation, color gradient and color hue are some examples. Humans are sensitive to the retinal variables. They easily differentiate between various colors, shapes, sizes and other properties.

### 3. Mention when to use bar chart in data visualization?

**Use a bar chart for the following reasons:**

- You want to compare two or more values in the same category
- You want to compare parts of a whole
- You don't have too many groups (less than 10 works best)
- You want to understand how multiple similar data sets relate to each other

**Don't use a bar chart for the following reasons:**

- The category you're visualizing only has one value associated with it
- You want to visualize continuous data

### 4. What are heat maps?

A heat map is a data visualization technique that shows magnitude of a phenomenon as color in two dimensions. The variation in color may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space. By definition, Heat Maps are graphical representations of data that utilize color-coded systems. The primary purpose of Heat Maps is to better visualize the volume of locations/events within a dataset and assist in directing viewers towards areas on data visualizations that matter most. But they're much more than that.

### 5. When to not apply, Retinal variables to data?

• It is quite clear that we can't use all variables to present any data types. • For example, – It is wrong to use color to represent numbers (1, 2, 3) – It is bad to use size to represent various currencies (€, £, ¥). [KODIGA PEDDAGA CHESI RAAAYU IDI]

• Note: That "Planar variables" can be applied to all the data types. • Indeed, we can use the X-axis for categories, ordered variables or numbers.

### 6. What is Data Visualization? Why use data visualization?

**SAME AS QN 11**

### 7. What do you mean by Data Encoding?

**Data Encoding :** Encoding in data viz basically means translating the data into a visual element on a chart/map/whatever you're making. You need to do it right, because doing it right will mean that other people looking at your visualizations can understand what you're trying to say or show. Another way to think of encoding is as a set of rules to follow. So when you're making them, you need to think logically about how you set up those rules, otherwise people will get totally confused. If you're doing something complicated, a good way of helping yourself to think about how you encode, or set up the rules, is: – Every time <data changes in some way>, do <something visual>. This helps you to be consistent in how you apply the rules. Consider this standard column chart:

Most people wouldn't have trouble understanding what it's showing us — but when you break it down, this is encoded in a few different ways: • Colour: Every time <category is bears/dolphins/whales>, change <colour of the column to be blue/orange/grey> • Size: Every time <number goes up>, increase <column height> • Grouping: Every time <month changes>, create <new cluster of columns>

**Here's a (not limited) list of ways you can encode data:**

- ✓ Size
  - ✓ Shape
  - ✓ Color
  - ✓ Grouping
  - ✓ Area
  - ✓ Position
  - ✓ Saturation
  - ✓ Line pattern
  - ✓ Line weight
  - ✓ Angle
  - ✓ Connections
- ~~~~~

## **8. Why is text added to data visualization and mention few texts that can be added?**

### **Role of Text in Data Visualization**

**Label – Text** (words and numbers) is often used to label information. Labels offer critical information to help readers interpret visualizations. **Introduce – Text** is often required to set the reader on a clear path to understanding.

**Introduce – Text** is often required to set the reader on a clear path to understanding. An introduction should give readers a preview of what the graphic will show.

**Explain – Text** can be used to clarify something about the message, right where it is needed in the graphic. This is for information too specific to be included in the introduction.

**Reinforce –** Some information is so important you need to say it more than once. Text can help increase the likelihood your graphic will be fully understood.

**Highlight – Text** can be used to visually highlight important data. This is different from reinforcement because you aren't repeating yourself, you're simply calling the reader's attention to it.

**Sequence –** Showing the order in which your reader should examine the contents of your visualizations can be tricky. Text can help instruct your readers to navigate your graphic in a specific way.

**Recommend –** Recommendations for action are best communicated in words. Sometimes you aren't simply informing the reader – you may be recommending what could or should be done.

**Inquire –** The visual display of quantitative information frequently invites questions. Don't be afraid to ask them. Sometimes one well-timed question is more useful than a series of answers.

~~~~~

## **9. What are the key questions for right visualization?**

**1. What story do you want to tell?** - At its core, online data visualization is about taking data and transforming it into actionable insight by using it to tell a story. Data-driven storytelling is a powerful force as it takes stats and metrics and puts them into context through a narrative that everyone inside or outside of the organization can understand.

**2. Who do you want to tell it to?** - Another key element of choosing the right data visualization types is gaining a clear understanding of who you want to tell your story to – or in other words, asking yourself the question, “Who is my audience?”

**3. Are you looking to analyze particular trends?** Every data visualization project or initiative is slightly different, which means that different data visualization chart types will suit varying goals, aims, or topics.

**4. Do you want to demonstrate the composition of your data?** If your primary aim is to showcase the composition of your data – in other words, show how individual segments of data make up the whole of something – choosing the right types of data visualizations is crucial in preventing your message from becoming lost or diluted.

**5. Do you want to compare two or more sets of values?** While most types of data visualizations will allow you to compare two or more trends or data sets, there are certain graphs or charts that will make your message all the more powerful.

**6. Is timeline a factor?** By understanding whether the data you’re looking to extract value from is time-based or time-sensitive, you’ll be able to select a graph or chart that will provide you with an instant overview of figures or comparative trends over a specific period.

**7. How do you want to show your KPIs?** - It’s important to ask yourself how you want to showcase your key performance indicators as not only will this dictate the success of your analytical activities but it will also determine how clear your visualizations or data-driven stories resonate with your audience.

**[KODIGA SONTAM GA SOLLU ADD CHESE]**

~~~~~

## **10. What are the charts used to display comparison variables and relationship variables?**

Scatter charts are primarily used for correlation and distribution analysis. Good for showing the relationship between two different variables where one correlates to another (or doesn’t). Scatter charts can also show the data distribution or clustering trends and help you spot anomalies or outliers. A good example of scatter charts would be a chart showing marketing spending vs. revenue.

**REFERENCE:** <https://eazybi.com/blog/data-visualization-and-chart-types#:~:text=Scatter%20charts%20are%20primarily%20used,you%20spot%20anomalies%20or%20outliers>

~~~~~

## **11. What is data visualisation? What are the uses of Data Visualisation?**

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

### **USES**

- **Better analysis:** Data visualization helps business stakeholders analyze reports regarding sales, marketing strategies, and product interest. Based on the analysis, they can focus on the areas that require attention to increase profits, which in turn makes the business more productive.
- **Quick action:** As mentioned previously, the human brain grasps visuals more easily than table reports. Data visualizations allow decision makers to be notified quickly of new data insights and take necessary actions for business growth.
- **Identifying patterns:** Large amounts of complicated data can provide many opportunities for insights when we visualize them. Visualization allows business users to recognize relationships between the data, providing greater meaning to it. Exploring these patterns helps users focus on specific areas that require attention in the data, so that they can identify the significance of those areas to drive their business forward.
- **Finding errors:** Visualizing your data helps quickly identify any errors in the data. If the data tends to suggest the wrong actions, visualizations help identify erroneous data sooner so that it can be removed from analysis.
- **Understanding the story:** Storytelling is the purpose of your dashboard. By designing your visuals in a meaningful way, you help the target audience grasp the story in a single glance. Always be sure to convey the story in the simplest way, without excessive complicated visuals.

- Exploring business insights: In the current competitive business environment, finding data correlations using visual representations is key to identifying business insights. Exploring these insights is important for business users or executives to set the right path to achieving the business' goals.
- Grasping the latest trends: Using data visualization, you can discover the latest trends in your business to provide quality products and identify problems before they arise. Staying on top of trends, you can put more effort into increasing profits for your business

## 12. Explain about different visualisation methods in detail.

**SAME AS QN 18**

## 13. Explain about visual encoding.

**REFER QN 20**

## 14. Explain about Geospatial data visualization?

GeoMapping used to develop complex visualizations of large geographically related data. Maps are only the ways of visualizing data. By visualizing geospatial data, show and correlate different variables to geographical locations by layer all these variables over maps.

Types: <https://www.kdnuggets.com/2017/10/7-techniques-visualize-geospatial-data.html>

## 15. Explain Quantitative data and Qualitative data with suitable examples.

**REFER QN1**

## 16. Explain any 4 Retinal Variables in detail.

- We can use the retinal variables! Like:

- Size
- Shape
- Texture
- Orientation
- Color Saturation
- Color Hue

### 2a. Size

- We know that size does matter. You can see the difference right away. Size is a good visualizer for the **quantitative** data.



### 2b. Shape

- Round circles ○
- Stars ☆
- Solid rectangles ■.
- We can easily distinguish dozens of shapes. They do work well sometimes for the visual encoding of **categories**.



### 2c. Texture

- Texture is **less common**. You can't touch it on screen, and it's usually less catchy than color.
- So, in theory texture can be used for soft encoding, but in practice it's better to pass on it.



### 2d. Orientation

- Orientation is **tricky**.
- While we're able to clearly identify vertical vs. horizontal lines.
- It is harder to use it properly for visual encoding.



### 2e. Color Saturation

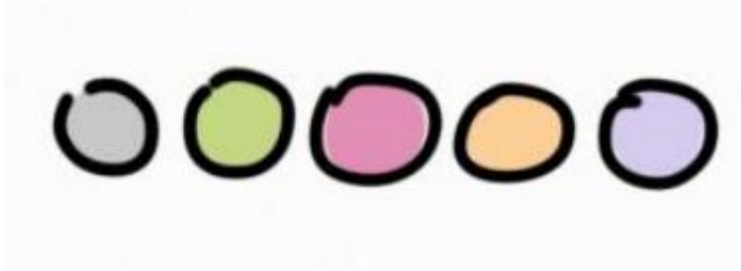
- Any color value can be moved over a scale.
- Greyscale is a good example.
- While we can't be certain that #999 color is lighter than #888, still it's a helpful technique to visualize the **ordered data**.



### 2f. Color Hue

- Colors are great to **separate categories**.
- Red color is alarming.
- Green color is calm.
- Blue color is peaceful.





### How to Apply the Retinal Variables to Data?

- It is quite clear that we **can't use all** variables to present any data types.
- For example,
  - It is wrong to use **color** to represent numbers (1, 2, 3) – It is bad to use **size** to represent various currencies (€, £, ¥).

### 17. What is data encoding?

Same as QN 2 & 7

### 18. Explain about 4 most common graphs used in data visualization?

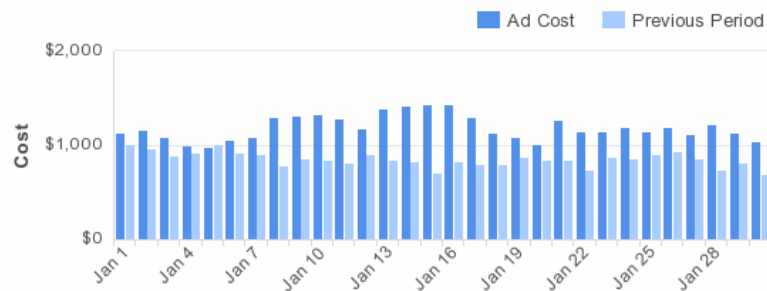
Some of the most common types of data visualization chart and graph formats include: **[write any 4]**

- **Bar Graph** - At some point or another, you've either seen, interacted with, or built a bar chart before. Bar charts are such a popular graph visualization because of how easy you can scan them for quick information. Bar charts organize data into rectangular bars that make it a breeze to compare related data sets.

#### NO NEED TO DRAW PERFECTLY [ROUGH GA DRAW CHEYU]

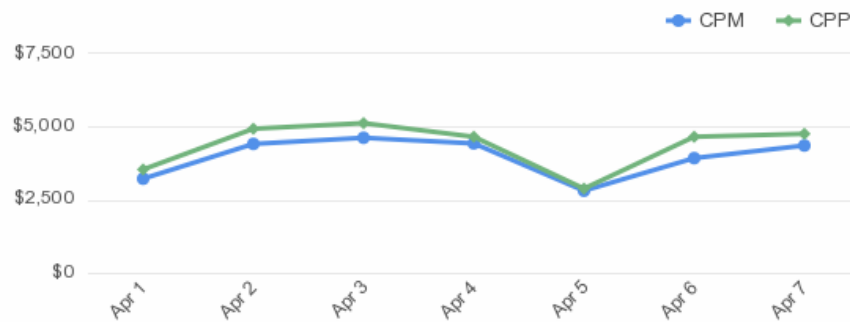
AdWords Ad Cost over Time (Last 30 Days)

All Campaigns

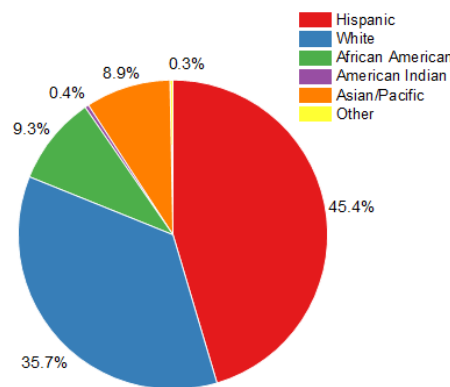


- **Line Graph** - Like bar charts, line charts help to visualize data in a compact and precise format which makes it easy to rapidly scan information in order to understand trends. Line charts are used to show resulting data relative to a continuous variable - most commonly time or money. The proper use of color in this visualization is necessary because different colored lines can make it even easier for users to analyze information.

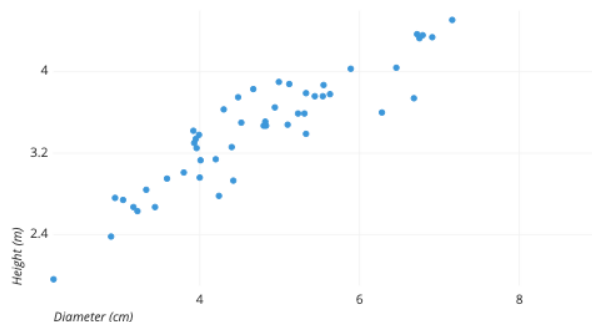
Facebook Ads CPM & CPP (Last 7 Days)



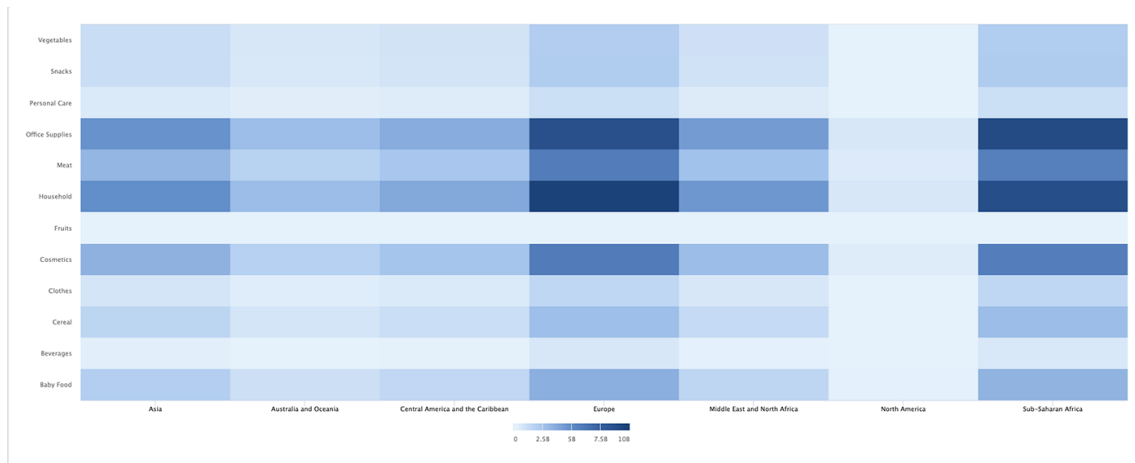
- Pie Chart** – Pie charts are an interesting graph visualization. At a high-level, they're easy to read and understand because the parts-of-a-whole relationship is made very obvious. But top data visual experts agree that one of their disadvantages is that the percentage of each section isn't obvious without adding numerical values to each slice of the pie.



- Scatter Plot Chart** - Scatterplots are the right data visualizations to use when there are many different data points, and you want to highlight similarities in the data set. This is useful when looking for outliers or for understanding the distribution of your data. If the data forms a band extending from lower left to upper right, there most likely a positive correlation between the two variables. If the band runs from upper left to lower right, a negative correlation is probable. If it is hard to see a pattern, there is probably no correlation.



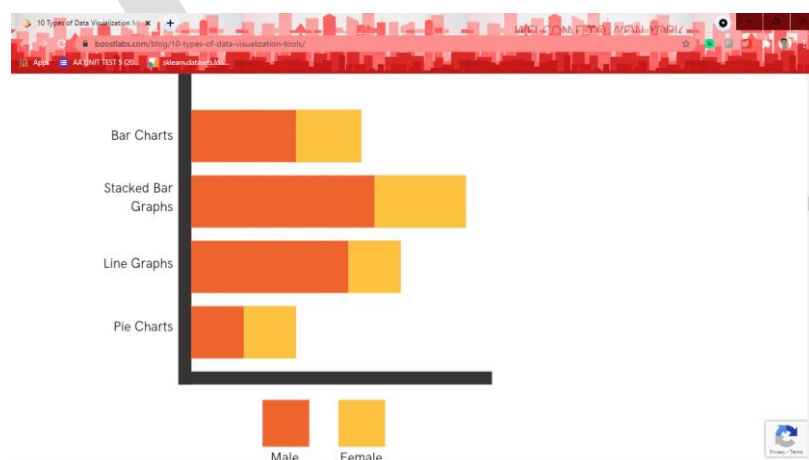
- Heat Map** - A heat map or choropleth map is a data visualization that shows the relationship between two measures and provides rating information. The rating information is displayed using varying colors or saturation and can exhibit ratings such as high to low or bad to awesome, and needs improvement to working well. It can also be a thematic map in which the area inside recognized boundaries is shaded in proportion to the data being represented.



- **Column Chart** - A column chart will include data labels along the horizontal (X) axis with measured metrics or values presented on the vertical (Y) axis, also known as the left side of the chart. The Y-axis will normally start at 0 and go as high as the largest measurement you're tracking. You can use column charts to track monthly sales figures, revenue per landing page, or similar measurements. Consistent colors help keep the focus on the data itself, though you can introduce accent colors to emphasize important data points or to track changes over time.

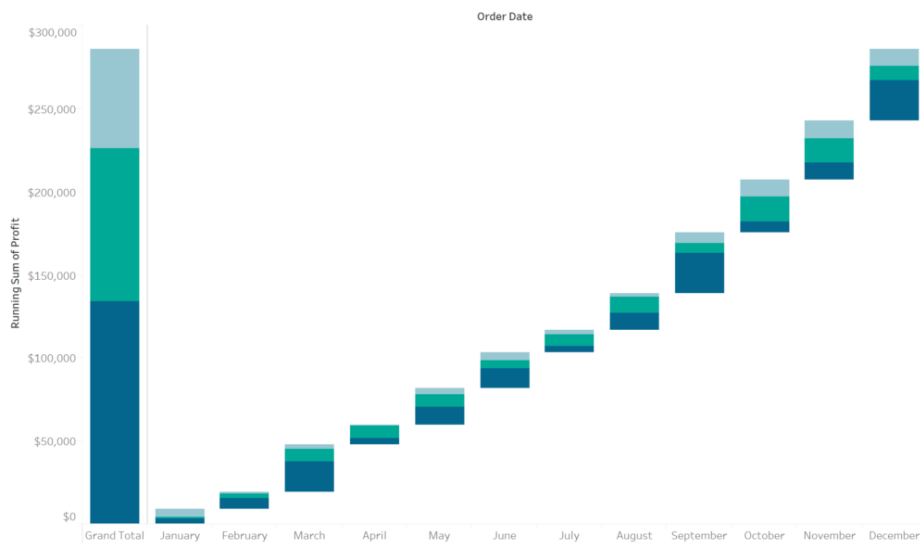


- **Stacked Bar Graph** - If you removed the color from this chart, it would look similar to a standard bar chart. The "stacked" layout represents this chart's contrasting color scheme. These colors map back to a legend that accompanies your map.

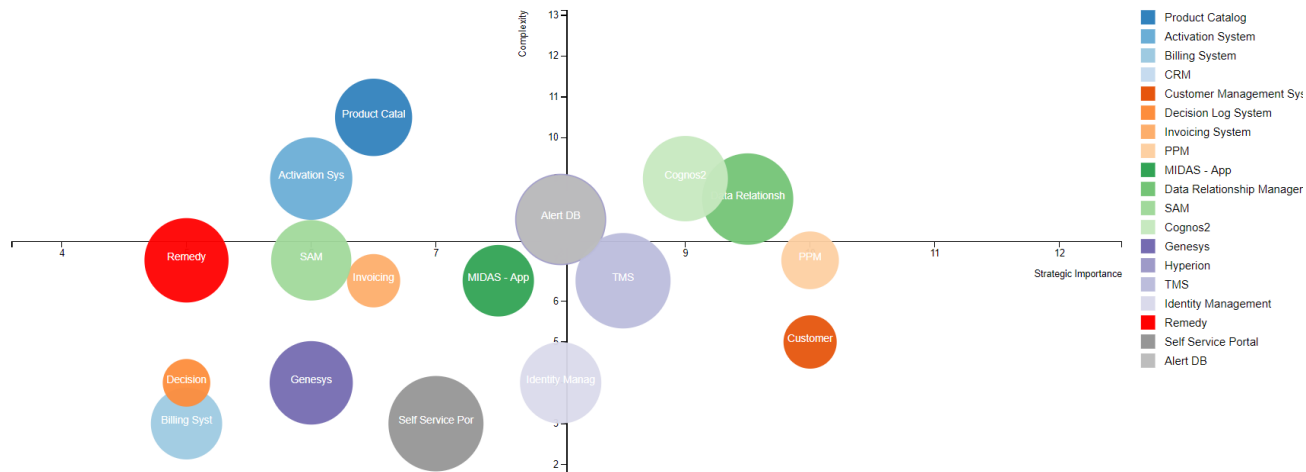


- **Stacked Column Chart**
- **Area Chart**
- **Dual Axis Chart**
- **Mekko Chart**

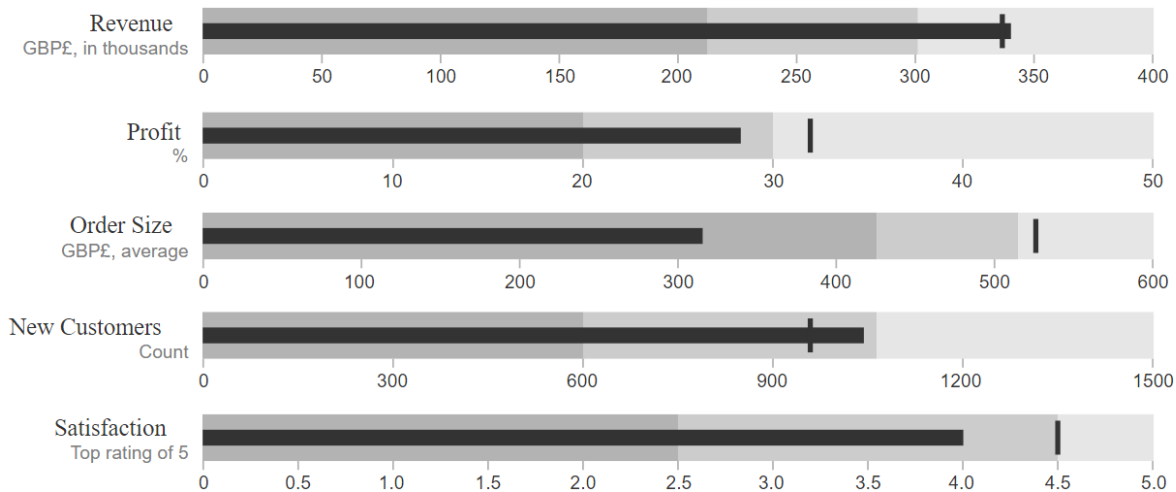
• Waterfall Chart -



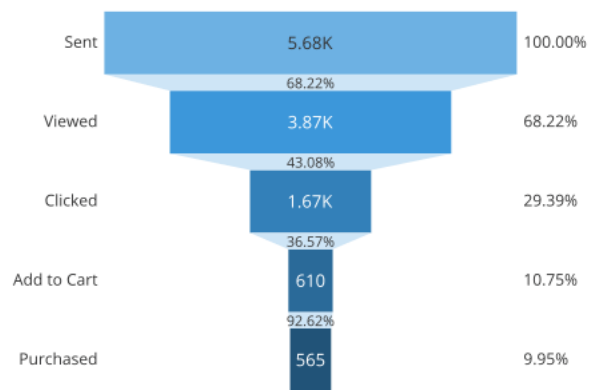
• Bubble Chart



• Bullet Graph



## • Funnel Chart



19. Explain any 4 Types of Data Visualization Categories in detail.

**SAME AS QN 18**

20. Explain in detail Planar and Retinal encoding.

## Visual Encoding

• Visual encoding is the way in which data is constructed into **visual structures**. They are the **building blocks** of graphics (images on a screen).

### 1. Planar Encoding

– Planar encoding is as simple as the laying of **axis**, like the X & Y axis in a simple line chart.

### 2. Retinal Encoding

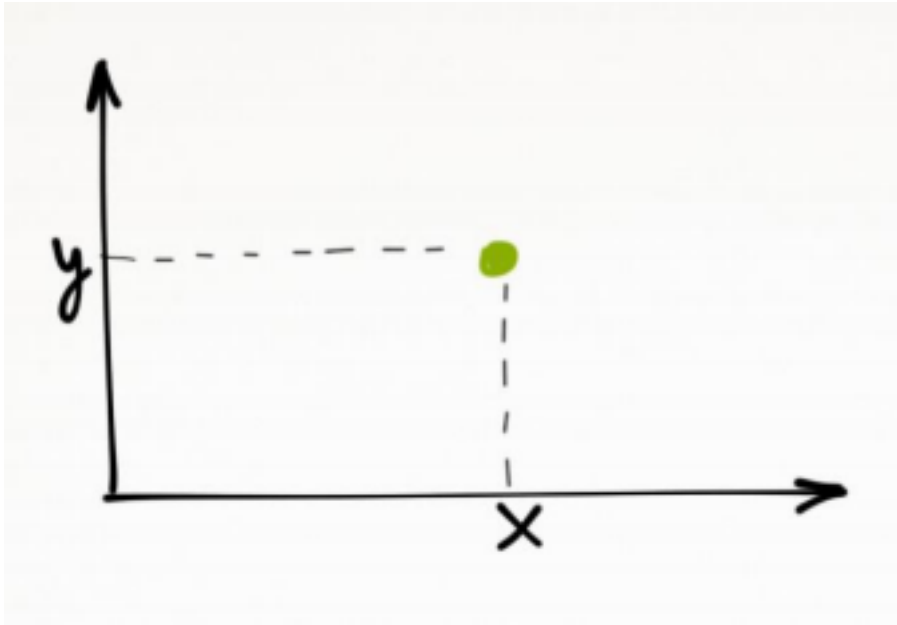
– To represent data in **3 or more variables**, retinal encoding comes into the picture.

– Size, texture, shape, orientation, color gradient and color hue are some examples.

– Humans are sensitive to the retinal variables. They easily differentiate between various colors, shapes, sizes and other properties.

### 1. Planar Variables

- Planar variables are known to everybody.
  - Planar variables work for any data type.
  - They work great to present any **quantitative data**.
  - If you've studied math's, you've been drawing graphs across the X- and Y-axis.
- It's a pity that we have to deal with the flat screens and just two planar variables. Well, we can try to use Z-axis, but 3D charts look horrible on screen in 95.8% of cases.



### 2. Retinal Variables

- So what should we do then to present 3 or more variables?
- We can use the retinal variables! Like: – **Size**
- **Shape**
- **Texture**
- **Orientation**
- **Color Saturation**
- **Color Hue**

#### 2a. Size

- We know that size does matter. You can see the difference right away. Size is a good visualizer for the **quantitative data**.



#### 2b. Shape

- Round circles ○

- Stars ☆
- Solid rectangles ■.
- We can easily distinguish dozens of shapes. They do work well sometimes for the visual encoding of **categories**.



## 2c. Texture

- Texture is **less common**. You can't touch it on screen, and it's usually less catchy than color.
- So, in theory texture can be used for soft encoding, but in practice it's better to pass on it.



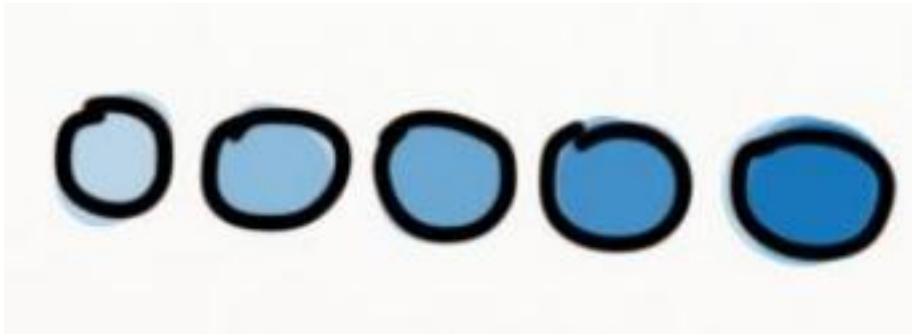
## 2d. Orientation

- Orientation is **tricky**.
- While we're able to clearly identify vertical vs. horizontal lines.
- It is harder to use it properly for visual encoding.



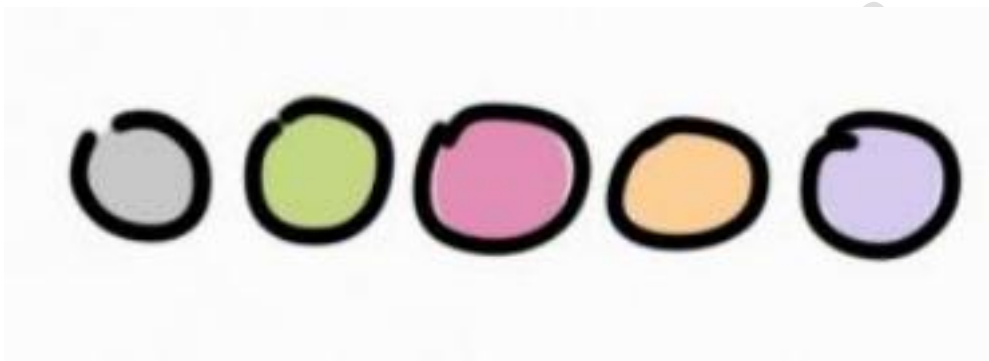
## 2e. Color Saturation

- Any color value can be moved over a scale.
- Greyscale is a good example.
- While we can't be certain that #999 color is lighter than #888, still it's a helpful technique to visualize the **ordered** data.



## 2f. Color Hue

- Colors are great to **separate categories**. – **Red** color is alarming.
- **Green** color is calm.
- **Blue** color is peaceful.



## How to Apply the Retinal Variables to Data?

- It is quite clear that we **can't use all** variables to present any data types.
- For example,
  - It is wrong to use **color** to represent numbers (1, 2, 3) – It is bad to use **size** to represent various currencies (€, £, ¥).
- **Note:** That “Planar variables” can be applied to all the data types.
- Indeed, we can use the X-axis for categories, ordered variables or numbers.

## 21. Explain about Distribution, Comparison, Relationship, Composition?

As mentioned, asking the right questions will form the foundations of choosing the right types of visualization charts for your project, strategy, or business goals. The fundamental categories that differentiate these questions are based on:

- Relationship
- Distribution
- Composition
- Comparison of data

Distribution (or place) is one of the four elements of the marketing mix. Distribution is the process of making a product or service available for the consumer or business user who needs it. This can be done directly by the producer or service provider, or using indirect channels with distributors or intermediaries.

Comparison of data points is probably the most common and easy-to-understand method for data analysis. As the name suggests, we use comparison to evaluate and compare values between two or more data points. With comparison you can also easily find the lowest and highest values in the chart.



The strength of a relationship refers to the extent to which data points on one variable correspond to the data points on the other variable. Often this is determined by the extent to which the observed values on one variable can be predicted from the corresponding values on the other variable.

Compositional Data Analysis (CoDA) refers to the analysis of compositional data (CoDa), which have been defined historically as random vectors with strictly positive components whose sum is constant (e.g., 100, one, a million). Compositional analysis generally refers to the measurement of the distribution of hydrocarbons and other components present in oil and gas samples. ... Compositional analysis can be used to assess the quality of samples, it can also help determine resevoir continuity and investigate leakage.

~~~~~

ADP