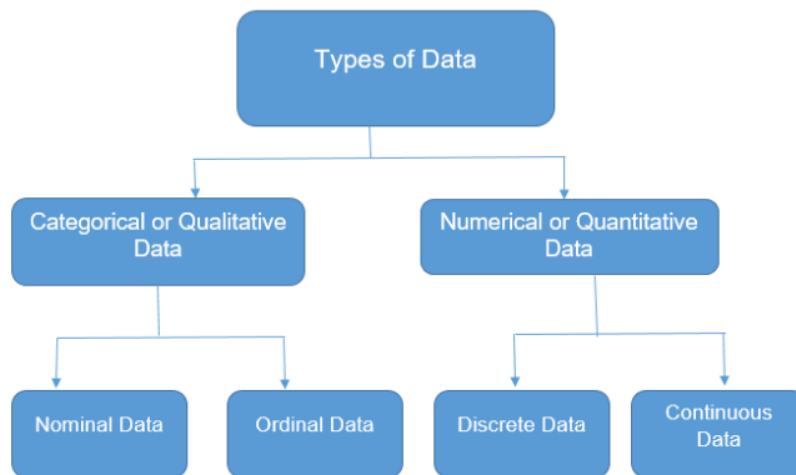# Assignment 1

Submitted to: Dr. Ankur Jain

Ajay Singh
20BAI10300
CSA3004
C11+C12+ …

Q1. Classify the data types in the context of data visualization.

Solution:

The data types are classified as follows in data visualization:

1. **Numerical data**
   These data have meaning as a measurement such as height, weight, IQ or blood pressure or count such as number of stocks, number of teeth, number of pages etc. It can be further classified as:
   - Discrete data: represent items that can be counted.
   - Continuous data: represent measurements; their possible values cannot be counted and can only be described using intervals on the real number line.

2. **Categorical Data**
   Categorical data represent characteristics such as a person's gender, marital status, hometown, or the types of movies they like. Categorical data can take on numerical values (such as "1" indicating male and "2" indicating female), but those numbers don't have mathematical meaning. It can be further classified as:

- **Ordinal Data**

  Ordinal data represent qualitative information without order, indicates that the measurement classifications are different and can be ranked.

  For example, suppose you receive a survey from your favourite restaurant that asks you to provide feedback on the service you received. You can rank the quality of service as "1" for poor, "2" for below average, "3" for average, "4" for very good and "5" for excellent. The data collected by this survey are examples of ordinal data. Here the numbers assigned have an order or rank; that is, a ranking of "4" is better than a ranking of "2."

- **Nominal Data**

  Nominal data represent qualitative information without order. It gives a label to the data. whereas the classification describes the data. An example of nominal data might be a "pass" or "fail" classification for each student's test result. Nominal data provides some information about a group or set of events, even if that information is limited to mere counts.

Q2. Classify the data types for the following snippets.

Solution: Here is a snippet of the data set that we are considering.

Classify the data types for the following set:

Row id: Discrete
Order id: Nominal
Order date: Discrete
Ship Date: Discrete
Customer Id: Nominal
Customer Name: Nominal

Segment: Nominal
City: Nominal
State: Nominal
Country: Nominal
Postal Code: Discrete
Market: Nominal
Region: Nominal
Product Id: Nominal
Category: Nominal
Sub Category: Nominal
Product Name: Nominal
Sales: Continuous
Quantity: Discrete
Discount: Continuous
Profit: Continuous

Question 3. Convince why data visualization is necessary. What is the various significance of data visualization?

Solution:

Data visualization provides a quick and effective way to communicate information in a universal manner using visual information. The practice can also help businesses identify which factors affect customer behavior; pinpoint areas that need to be improved or need more attention; make data more memorable for stakeholders; understand when and where to place specific products; and predict sales volumes.

The importance of data visualization, which are:

1. Analysing the Data in a Better Way

Analysing reports helps business stakeholders focus on the areas that require attention. The visual mediums help analysts understand the key points needed for their business. Whether it is a sales report or a marketing strategy, a visual representation of data helps companies increase their profits through better analysis and better business decisions.

2. Faster Decision Making

Humans process visuals better than any tedious tabular forms or reports. If the data communicates well, decision-makers can quickly take action based on the new data insights, accelerating decision-making, and business growth simultaneously.

3. Making Sense of Complicated Data

Data visualization allows business users to gain insight into their vast amounts of data. It benefits them to recognize new patterns and errors in the data. Making sense of these patterns helps the users pay attention to areas that indicate red flags or progress. This process, in turn, drives the business ahead.

Various significance of data visualization are:

1. Effective Data Visualization is the key to unlock Big Data. It can solve any data inefficiencies and easily and instantly absorb vast amounts of data presented in visual formats.

2. By enabling users to understand data rapidly, visualization can quickly increase the speed of decision making as well. Any business must make fast decisions and not get bogged down by inefficiencies. Timely actions result in averting any losses and benefit from any market condition.

3. A big reveal for any differences in the trends and patterns is vital for any business's survival. It is critical to know what is causing increased losses or what is required to maximize gains.

4. Visualization helps identify errors and inaccuracies in data quickly.

5. Companies can utilize visualization to access real-time information and assist in management functions in a significant manner. Decision-makers can benefit from on-demand data and use visualization to increase the effectiveness of operations and improve productivity.

6. It promotes storytelling in the most compelling way. Visuals are used in the most meaningful way to convey the right message to the audience.

7. It enables enterprises to stay on top of their game by discovering the latest trends through data visualization tools.

Question 4. Differentiate dimensions and measures in Tableau.

Solutions:

Data fields are made from the columns in your data source. Each field is automatically assigned a data type (such as integer, string, date), and a role: Discrete Dimension or Continuous Measure (more common), or Continuous Dimension or Discrete Measure (less common).

Dimensions	Measures
Dimensions contain qualitative values (such as names, dates, or geographical data). You can use dimensions to categorize, segment, and reveal the details in your data. Dimensions affect the level of detail in the view.	*Measures* contain numeric, quantitative values that you can measure. Measures can be aggregated. When you drag a measure into the view, Tableau applies an aggregation to that measure (by default)
According to Tableau's Knowledge Base, a measure is a field that is a dependent variable; that is, its value is a function of one or more dimensions Tableau treats any field containing	According to Tableau's Knowledge Base, a dimension is a field that can be considered an independent variable. By default, Tableau treats any field containing qualitative, categorical

numeric (quantitative) information as a measure.	information as a dimension.
It is an independent variable	It is a dependent variable
It is not dependent on the measure	It is dependent on the dimension
We cannot aggregate it.	We can aggregate it.
It is used to compare the data	It is a metric value that we will use to compare the dimension
It contains a duplication of the data.	It does not contains a duplication of the data.
It contains qualitative and categorial information	It contains quantitative data.
It can be continuous and discrete.	It cannot discrete and continuous.
If the numeric field is the dimension, it can be treated as the measure.	The numeric field can be treated from measure to dimension.
It is not possible to get a number of records because aggregation does not apply to it.	Due to the aggregation feature, we can get the number of records present for the database no matter how huge the dataset is.
It does not have Longitude and Latitude measure present.	Longitude and Latitude measures are present for the geographic dimension.

Consider the following bar chart, created in Tableau with the Sales measure from the Sample – Superstore data set:

Sales is quantitative, so by default, Tableau will guess that the field is a measure. It can be considered a dependent variable, because a measure by itself does not provide much value.

$8,951,931
Sales
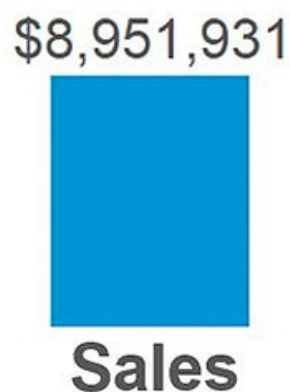
Here is the same Sales measure from above, broken down by the dimension of Region:
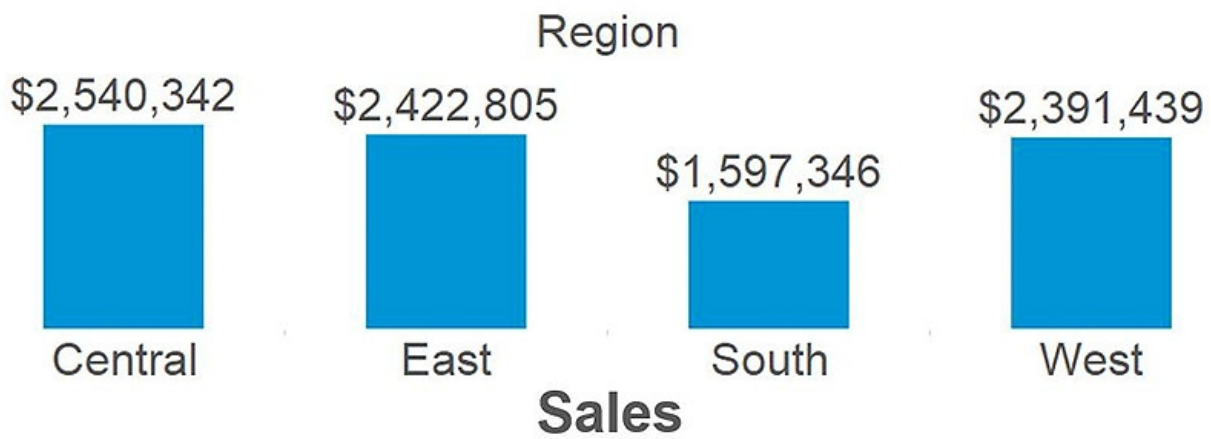
Question 6: What is the various state of art tools available in the market for data visualization? Describe anyone.

Solution:

Various state of art tools available in the market for data visualization:-

- Tableau (and Tableau Public)
- Infogram
- ChartBlocks
- Datawrapper
- D3.js
- Google Charts
- FusionCharts
- Chart.js
- Grafana
- Chartist.js
- Sigmajs
- Polymaps

Tableau (and Tableau Public)

Tableau has a variety of options available, including a desktop app, server and hosted online versions, and a free public option. There are hundreds of data import options available, from CSV files to Google Ads and Analytics data to Salesforce data.

Output options include multiple chart formats as well as mapping capability. That means designers can create color-coded maps that showcase geographically important data in a format that's much easier to digest than a table or chart could ever be.

The public version of Tableau is free to use for anyone looking for a powerful way to create data visualizations that can be used in a variety of settings. From journalists to political junkies to those

who just want to quantify the data of their own lives, there are tons of potential uses for Tableau Public. They have an extensive gallery of infographics and visualizations that have been created with the public version to serve as inspiration for those who are interested in creating their own.

Pros

- Hundreds of data import options
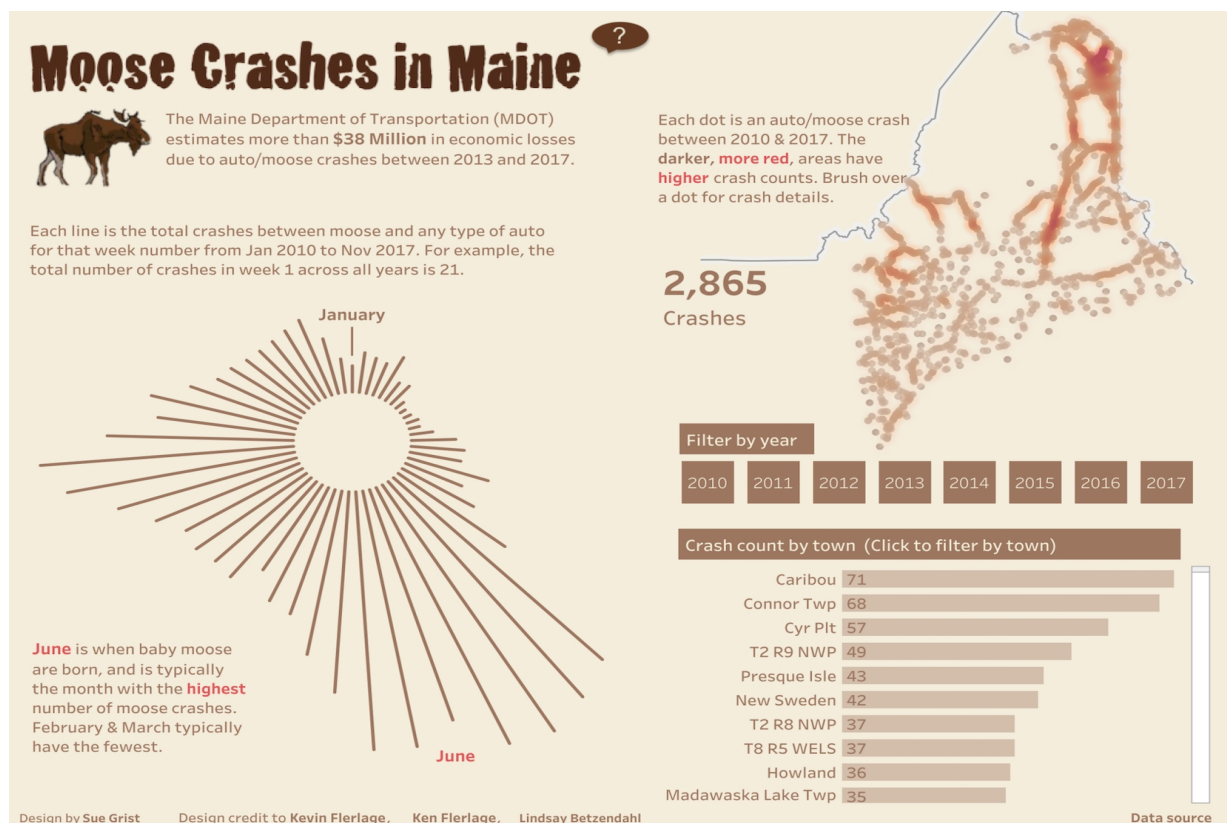
- Mapping capability

- Free public version available

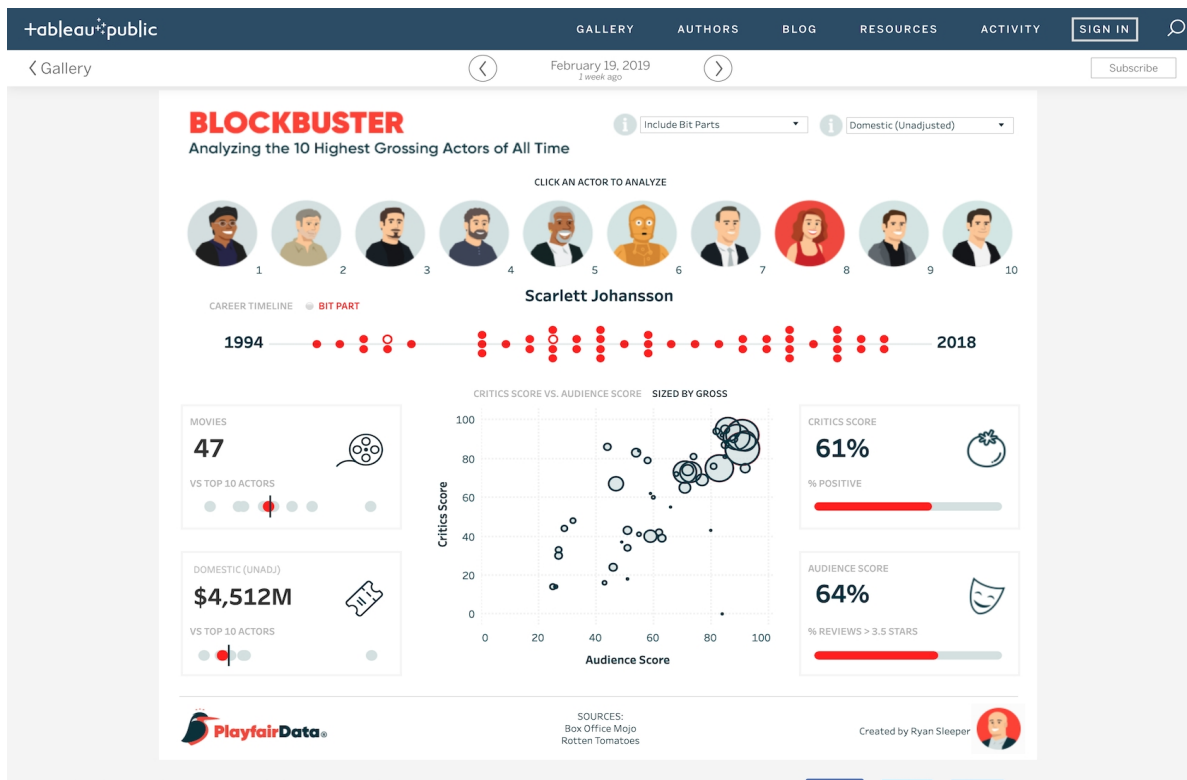- Lots of video tutorials to walk you through how to use Tableau

Cons

- Non-free versions are expensive ($70/month/user for the Tableau Creator software)

- Public version doesn't allow you to keep data analyses private

Data Visualization Examples

Data visualizations can make public safety data easier to digest.

An interactive visualization of the highest-grossing actors of all time.

Bottom Line

Tableau is a great option for those who need to create maps in addition to other types of charts.

Tableau Public is also a great option for anyone who wants to create public-facing visualizations.

Question 7. How does the human mind work for pre-attentive processing? What are the different properties of pre-attentive processing? Illustrate it using an example.

Pre-Attentive Processing

Preattentive processing plays a significant role in human vision. Pre-attentive processing is the subconscious accumulation of information from the environment. All available information is pre-attentively processed. The experience is automated and it takes between 50 to 500 milliseconds for the eye and the brain to process a pre-attentive property of any image.

Then, the brain filters and processes what is important. Information that has the highest salience (a stimulus that stands out the most) or relevance to what a person is thinking about is selected for further and more complete analysis by conscious (attentive) processing. Understanding how pre-attentive processing works is useful in designing, advertising, marketing, education, and prediction of cognitive ability.

As humans, irrespective of our cultures, genders, and age groups, our visual system is wired to look at very much the same aspect of any design/image in the first few seconds of viewing. Visual tasks include reading this sentence, viewing a webpage, ads, packaging looking for an exit sign in a public space, a company logo on a piece of direct mail, or an employee at a retail store. But, for our vision system, these tasks translate into looking for colors, shapes, etc.

There are four basic visual properties (color, form, movement, and spatial position) that can be defined as pre-attentive (e.g. they are processed in our sensory memory without our conscious thought). Information visualization designers can take advantage of these properties in their designs to help users better understand the information they are being presented with.

The Four Pre-attentive Visual Properties

    1.**Color**

Color is one of the most common properties used to call attention. Color can be expressed in many different ways. From the RGB (Red, Green, Blue) scale to the CMYK (Cyan, Magenta, Yellow, and Key) scale to the HSL (Hue, Saturation, and Lightness) scale
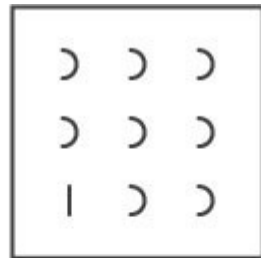
in terms of pre-attentiveness, the HSL scale is useful to us when we examine color. The reason being is that intensities and hues are subjected to preattentive processing. Hue is a measure of the color that we normally give a name to such as "Red". Hue is the correct word to use to refer to just the pure spectrum colors. Any given color can be described in terms of its value and hue. Saturation and lightness are a measure of the intensity of colors.

Hues and intensities are pre-attentively processed and can be used to separate visual elements from their surroundings. Leveraging colors and drawing the eyes can save time and reduce the need to sift through the information.
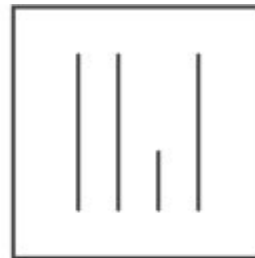
2. Form

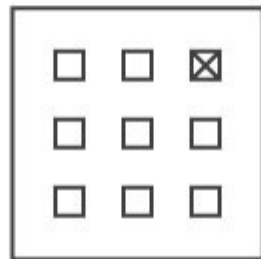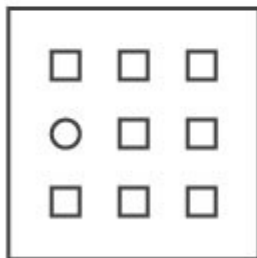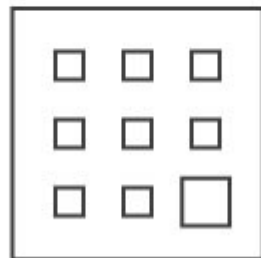|		
Orientation and collinearity	Curvature	Length
Width	Added marks	Numerosity
Shapes	Size	Spatial grouping
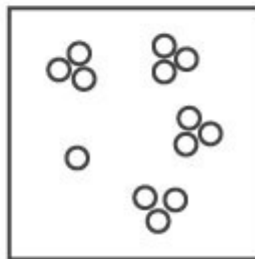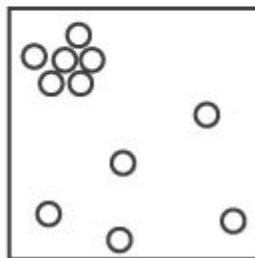
The form applies to various attributes listed below. In design, the form can be used either to increase attention to specific elements or to reduce attention to it. For example, if you would like to show an important text with an authoritative statement of want users to read that text first then using the width and size of the text by increasing as a proportion to other text in design or webpage, an emphasis is put on that element. Using size to indicate its importance in a design by making it larger. Conversely, you could make it smaller and reduce its importance.

Form attributes include:

· Collinearity

· Curvature

· length, breadth, and width

· Marks added to objects

· Numerosity

· Shape

· Size

· Spatial grouping

· Spatial orientation

You might not realize it but you didn't even have to think to see these differences — it was all done in your sensory memory without conscious effort.

3. Movement

Movement can be used very effectively to call someone's attention to a design or image. There are 2 attributes to movement 1. Flicker and 2. Motion. While these attributes are most attention-grabbing, they have some negative effects too. Motion or flicker elements in design sometimes become annoying and distracting for users from the information presented. Digital banner ads have been using these attributes over the years but It has given rise to ad blocking software too. A designer should carefully use these elements in design or image.

4. Spatial Position

Is the ability of the human visual system to perceive two or more object's position in space relative to oneself and in relation to each other. The gestalt principle that applies most to space is that of figure-ground. Everything in a design of yours will be seen as one or the other, and the relationship between them is mutually exclusive. Neither can be perceived except in relation to the other and changing one is impossible without changing the other as well. It sets a context for how your design communicates and how it will be interpreted.

The figure to ground/Gestalt psychology provides categories such as proximity, closure, continuity, connectedness, and similarity.

Example — Closer — This makes use of space as gaps between elements. Viewers fill in the gaps with their own information to complete a whole from the parts. Too much space and no closure occurs. Too little space and no closure is needed. Only the correct balance between space and filled-in space will activate the space and lead to closure.