DATA VISUALIZATION – WEEK 1

INTRODUCTION TO VIZUALIZATION

Data visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see **analytics presented visually**, so they can grasp difficult concepts or identify new patterns. With interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed.

Why is data visualization important?

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

A primary goal of data visualization is to **communicate information clearly and efficiently via statistical graphics, plots and information graphics**. 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. It makes complex data more accessible, understandable and usable.

Information graphics: Information graphics are visual representation of information or data quickly and clearly. They utilize graphics to enhance human visual system ability to see patterns and trends. In a broad sense, to what infographics are, and what they do—which is to condense large amounts of information into a form where it will be more easily absorbed by the reader.

In newspapers, infographics are commonly used to show the weather, as well as maps, site plans, and graphs for summaries of data. Modern maps, especially route maps for transit systems, use infographic techniques to integrate a variety of information, such as the conceptual layout of the transit network, transfer points, and local landmarks.

Statistical graphics and plots: Statistical graphics, also known as graphical techniques, are graphics in the field of statistics used to visualize quantitative data. Whereas statistics and data analysis procedures generally yield their output in numeric or tabular form, graphical techniques allow such results to be displayed in some sort of pictorial form.

A plot is a graphical technique for representing a data set, usually as a graph showing the relationship between two or more variables.

Visualization Design

Author Stephen Few described eight types of quantitative messages that users may attempt to understand or communicate from a set of data and the associated graphs used to help communicate the message.

Selecting the appropriate visualization method will be influenced by the definition work you undertook to clarify the intention of your visualization communication. It is about starting the journey towards identifying the most suitable way to answer your main data questions: how are you going to show, what it is you want to say.

Method classification	Communication method
Time-series	A single variable is captured over a period of time, such as the unemployment rate over a 10-year period. A line chart may be used to demonstrate the trend.
Ranking	Categorical subdivisions are ranked in ascending or descending order, such as a ranking of sales performance (the measure) by sales persons (the category, with each sales person a categorical subdivision) during a single period. A bar chart may be used to show the comparison across the sales persons.
Part-to-whole	Categorical subdivisions are measured as a ratio to the whole (i.e., a percentage out of 100%). A pie chart or bar chart can show the comparison of ratios, such as the market share represented by competitors in a market.
Deviation	Categorical subdivisions are compared against a reference, such as a comparison of actual vs. budget expenses for several departments of a business for a given time period. A bar chart can show comparison of the actual versus the reference amount.
Frequency distribution	Shows the number of observations of a particular variable for given interval, such as the number of years in which the stock market return is between intervals such as 0-10%, 11-20%, etc. A histogram, a type of bar chart, may be used for this analysis. A boxplot helps visualize key statistics about the distribution, such as median, quartiles, outliers, etc.
Correlation	Comparison between observations represented

	by two variables (X, Y) to determine if they tend to move in the same or opposite directions. For example, plotting unemployment (X) and inflation (Y) for a sample of months. A scatter plot is typically used for this message.
Nominal comparison	Comparing categorical subdivisions in no particular order, such as the sales volume by product code. A bar chart may be used for this comparison.
Geographic or geospatial	Comparison of a variable across a map or layout, such as the unemployment rate by state or the number of persons on the various floors of a building. A cartogram is a typical graphic used.

In his book Visualizing Data (O'Reilly), Ben Fry identifies seven stages of creating information visualization: **acquire, parse, filter, mine, represent, refine, and interact**. Each stage requires a certain level of technical or artistic talent, and information visualization necessitates the close integration of these talents.

Establishing intent – visualization's function – The intended function of data visualization concerns the functional experience you create between your design, the data, and the reader/user.

- Convey an explanatory portrayal of data to a reader.
- Provide an interface to data in order to facilitate visual exploration.
- Use data as an exhibition of self-expression.

Key factors surrounding a visualization project

- The aim
- Time pressures
- Costs
- Client pressures
- Format
- Technical capabilities

The eight hats of data visualization design

- **The initiator** - The initiator is the leader, the person who is seeking a solution to the task as per

the brief or self-initiated curiosity. The hat is that of an explorer; they want to explore data and different design avenues to find answers to problems or evidence to serve their researcher mindset.

- **The data scientist** - The data scientist is characterized as the data miner, wearing the miner's hat. They are responsible for sourcing, acquiring, handling, and preparing the data.
- **The journalist** - The journalist is the storyteller, the person who establishes the narrative approach to the visualization's problem context.
- **The computer scientist** - The computer scientist is the executor, the person who brings the project alive. With their critical technical capability, they are ultimately the ones who will construct the solution.
- **The designer** - The designer is the creative, the one, who, in harmony with the computer scientist, will deliver the solution. They have the eye for visual detail, a flair for innovation and style and are fully appreciative of the potential possibilities that exist.
- **The cognitive scientist** - The cognitive scientist is the thinker in terms of appreciating the science behind the effectiveness of the technical and designed solutions.
- **The communicator** - The communicator is, naturally, concerned with the communication side of the project. With their hard hat on, they act as the negotiator and presenter, operating at the client-customer-designer gateway, helping to inform all those who are involved on progress, requirements, problems, and solutions.
- **The project manager** - This final role is essentially that of the manager or coordinator, the person who does much to pick up many of the unpopular duties to help bring the whole project together.

Preparing and familiarizing yourself with your data - Data is our raw material, the principle ingredient in our creative recipe. Irrespective of what we intend or hope to show through our visualization design, the data will ultimately do the talking.

- Acquisition
- Examination
- Completeness
- Quality
- Data types
- Transforming of quality

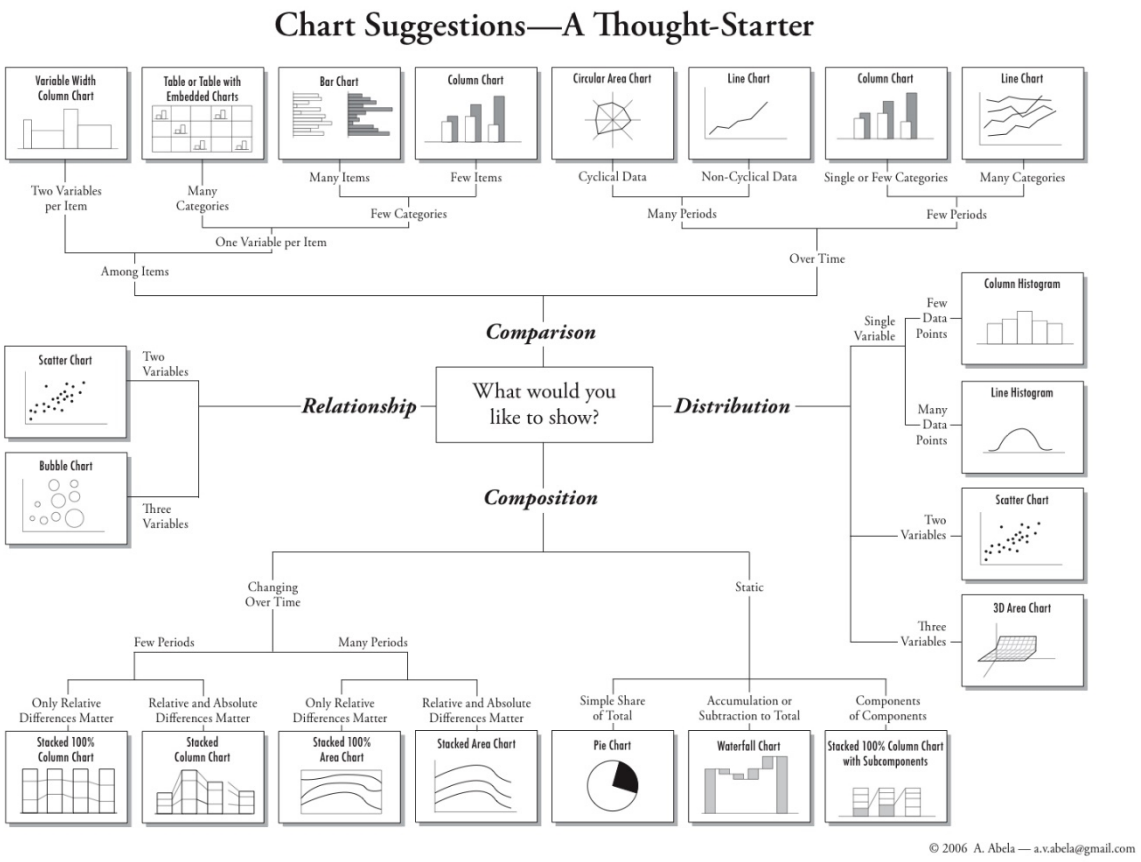
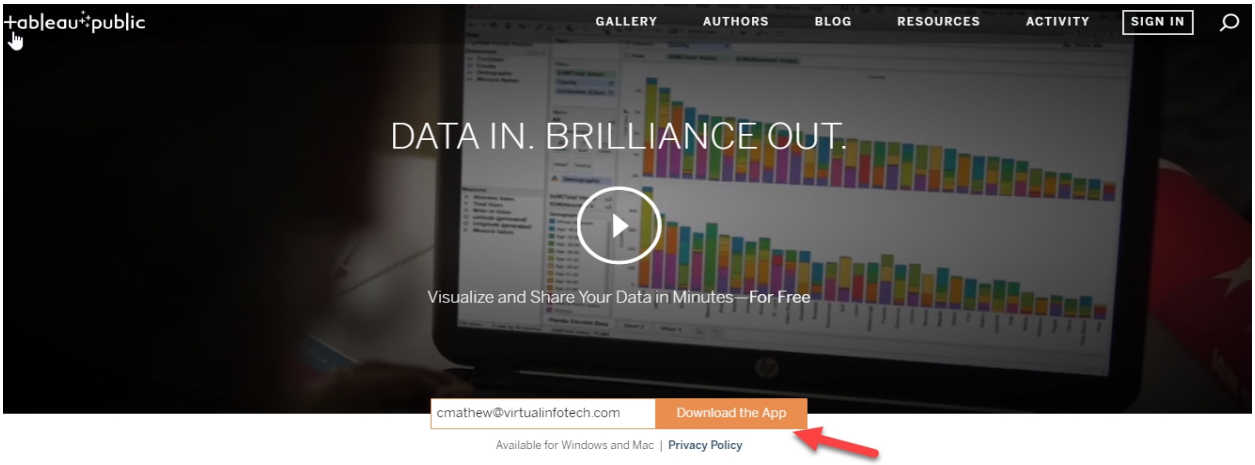
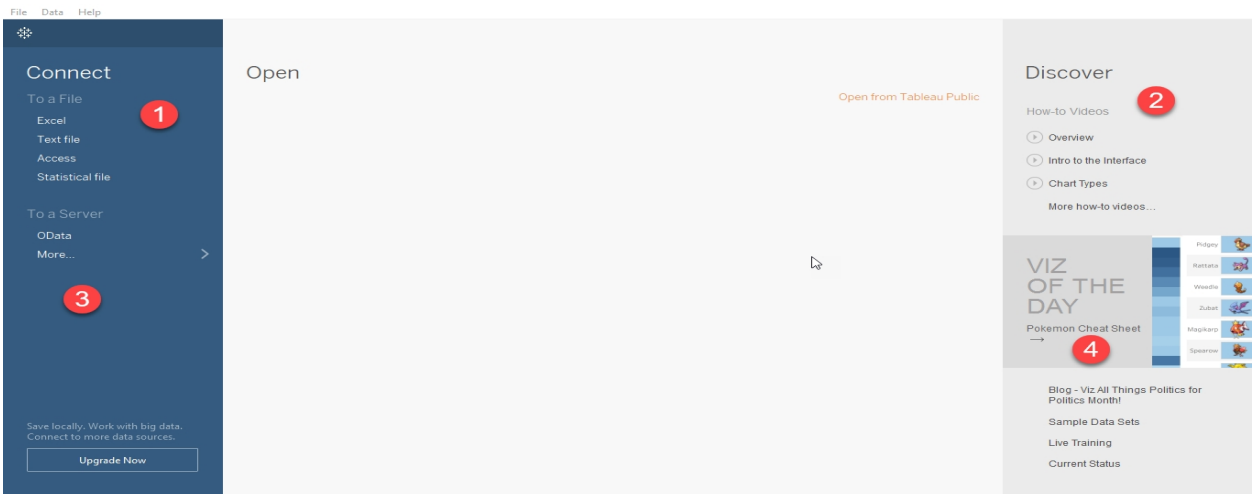


Tableau – One of the visualization tools

- Download Tableau, which is one of the tools we will be using to do visualization from [https://public.tableau.com/s/.](https://public.tableau.com/s/)





- Run the application file to install tableau to your computer.
- **We will look into the tableau user interface** – The Data source interface, the side bar, including the Data window and the Analytics pane, the toolbars and menus, the Columns, Rows, and Filters shelves, the Marks card, the Filters and Pages shelves, the ShowMe card.

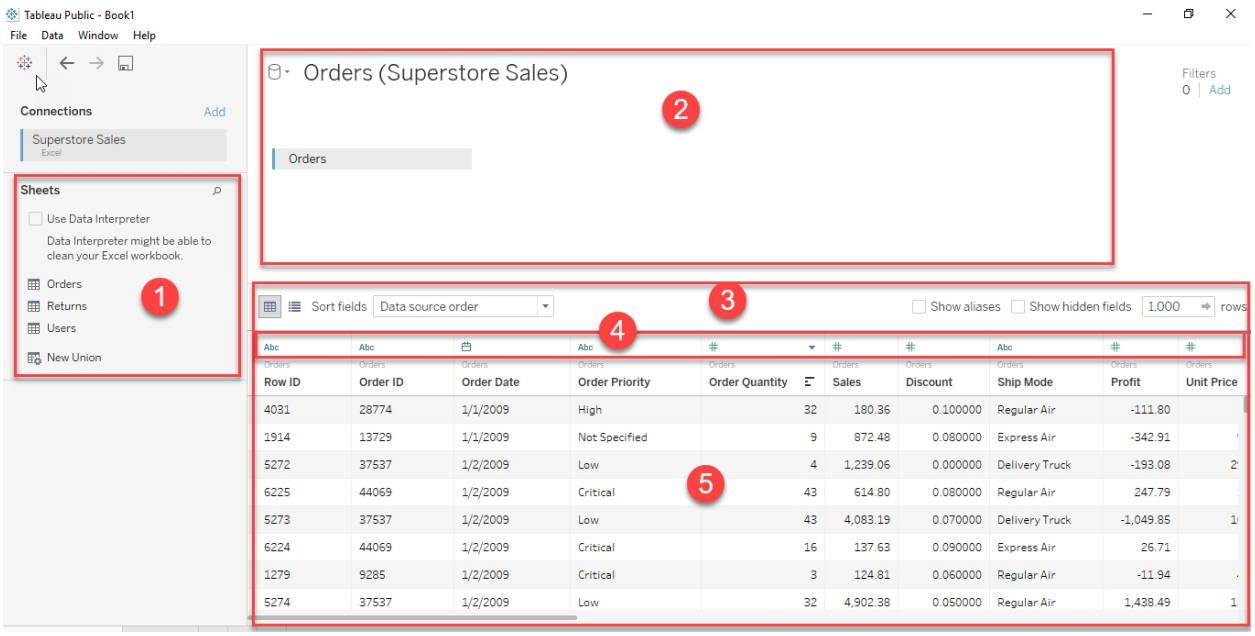


Figure 1

The following are the parts of the data source panel that are shown in Figure 1

- **Data Sheets (1)**: These are the tables within the data source, which can be dragged to the workspace.
- **Tables (2)**: This is the table data you will be working with.
- **Data view grid (3)**: View grid where you can change the sort order, show aliases, hidden fields, and number of rows to load.
- **Data types (4)**: It shows the data type selected for the data in the data table. The default data

types can be changed.

- **Data grid (5)**: Shows the actual data loaded from the data source.

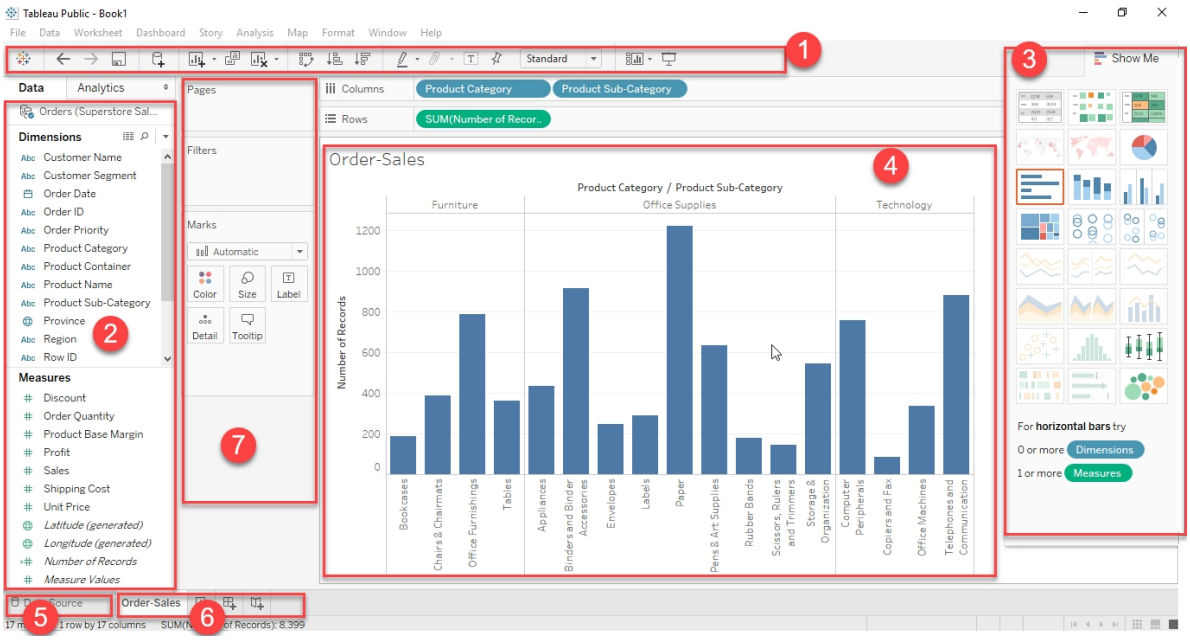


Figure 2

The following are the parts of the user interface that are shown in Figure 2

- **Toolbar (1):** This is where you can save your work, among other functions.
- **Data sidebar (2)**: Data window shows the dimensions and measures of your data.
- **ShowMe Card (3)**: ShowMe panel lets you pick up different chart type for visualization.
- **The View (4)**: This is where the graph or visualization is shown.
- **Data Source (5)**: The link to the data source page where you select the data source.
- **Sheet tabs (6)**: This allows you to create, rename, or duplicate sheets and dashboards.
- **Cards and shelves (7)**: These are the areas where you can add fields or filters to the visualization.

Readings: Some of the useful links you can go over to understand visualization, used in the course

material and just see some of the visualizations being done.

1. http://onlinehelp.tableau.com/current/pro/desktop/en-us/help.html#default.html%3FTocPath%3D_____1
2. https://github.com/d3/d3/wiki/Gallery
3. https://developers.google.com/chart/interactive/docs/gallery
4. https://www.visualcapitalist.com
5. http://www.viswiz.com
6. http://extremepresentation.com/visualization-taxonomies/
7. https://en.wikipedia.org/wiki/Visualization_(computer_graphics)