**Lecture0:** Information Visualization/Scientific Visualization/Data Visualization/InfoGraphics/Visual Analytics

InfoVis vs SciVis : SciVis [Direct Vlumne Rnderding, Isosurfaces, Line Integral Convolution, Streamlines] InfoVis [Parallel coordinates, Scatter Plots, Node-link Diagrams] -> Both have Glyphs

InfoVis – Spatialization chosen[Munzner] / Spatialization chosen and you think of data as collection of discredit items [Tory]

SciVis – spatialization given [Munzner] / Spaialization given and you think of data as samples from a continuous entity [Tory]

**Data Visualization** is the study of the visual representation of data, meaning “information which has been abstracted in some schematic form, including attributes or variables for the units of information.”

**Infographics** is visual representation of information, data or knowledge. These graphics are used where complex information needs to be explained quickly and clearly, such as in signs, maps, journalism, technical writing, and education. They are also used extensively as tools by computer scientists, mathematicians, and statisticians to ease the process of developing and communicating conceptual information.

**Visual Analytics** is the science of reasoning with visual information; pairs machine intelligence (computing, bit-representations) with human intelligence (creativity, visual representations)

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

**Golden Age of Visulaization –** Increasing the representation of everything is in a digital form. Explosion of capture of digital information about everything. Digital data can easily be transformed into many kindsOfVisu

**Astronomy Data Growth – [**From Glass plates to CCDs : Detectors follow Moore’s law], [The result: a data tsunami : available data doubles every two years], [Telescope Growth : 30X glass (concentration), 3000X in pixels (resolution)], [Sigle images : 16kx16k pixels], [Large Synoptic survey Telescope : wide field imaging at 5 terabytes/night]

**Visualizing the Universe: Insights from Telescope –** f**alse-color imaging** can enhance features in celestial objects, which allows for a more detailed and visually striking representations. / **Spectra** obtained by space telescopes enables scientists to identify the chemical composition. / **Spectral line plots and heatmaps** can effectively represent this data, revealing patterns and characteristics of distant / **Visualizing the light curves** generated by Kepler telescope can detect subtle variations in brightness **/ Infrared Imaging** infrared data from JWST helps reveal structures and objects obscured by dust. / **volumetric rendering** can create third dimensional visualizations of nebulae

**Data Vis. In Subatomic collider Experiments** : [Feynman Diagram – representations depict the interactions between particles, illustrating the fundamental processes occurring in collider experiments. It aids in understanding the underlying physics and assists in interpreting the complex data generated by particles collision], [Event Displays – representations of individual collision events provide a comprehensive view of the particles produced during high-energy collisions. Helps scientists identify patterns and understand the dynamics of quark-glouon], [Energy Deposition – Heaptmaps and contour plots visualize the energy deposition patterns of particles in the detector, aiding in the identification of specific particles types and enabling scientists to study the properties of quark-gluon]

**Exploring Medical Data Through Vis.** : [Genonme Browser / Slice – based Vis. (such as CT, MRI) / 3D Volume Rendering (CT MIR in 3D) / Chemical Structures / Activity Heatmaps (heatmaps, dendrograms finding activity compounds)

**Technical Challenges : The Data Tsunami –** Many sources, many causes and enablers ( increased detector resolution, increased storage capability, increased number of sensors) -> Extracting insight

**How Vis. Help?** : [Data Exploration and understanding – vis. Provides a means to explore and understanding large, complex datasets. Enable users to gain insights and make informed decisions / Data Filtering and Reduction – allows users to interactively filter and reduce data, focusing on specific subsets. (brushing, linking, filtering, aggreg. Can navigate through vast data and extract relevant info. / Data Integration – integrate and represent data tsunami to unified manner / Real-time Monitoring and Analysis – timely insight / Communication and Collaboration – easier for non technical stakeholders to understand and interpret data)

**Two Primary Goals :** [Explain, illustrate, communicate] / [Analyze, Explore, Discover, Decide]

**Presentation of Data :** [Characteristics : involves the use of charts, graph, tables, and textual content to convey info. Focuses summarizing and organizing data in a structured manner] / [Benefits : provide an overview of data, highlight key points, trends and insights. Useful for conveying complex info. In a concise format. Used in boardrooms, meetings] / [Considerations : limited by their static nature making it challenging to explore data interactively or to represent multi dimensional info. Require additional explanation or narration to facilitate understanding]

**Visualization of Data** : [Characteristics : involves the use of interactive dynamic visual representations to explore and analyze complex datasets. Uncover patterns, relationships, trends that my not apparent in raw data] / [Benefit: allow users to interact with data, enabling exploration, filtering, drilling down into specific details. Used in data analytics, dashboards, interactive reports] / [Consideration: require design principles – choosing appropriate visual encodings, scales, interactivity. Can be subjective, misinterpretation can occur if not properly designed]

**Charles Joseph Minard** : French Civil engineer, recognized for contribution in field of infographics. (Minard Chart of Napoleon’s Russian Campaign)

**Graphical vocabulary(1967)**: Marks – [Points, lines, areas] / Position – [vertical, horizaontal, ] / Retinal – [Color, size, shape, gray, orientation, texture]

**Semiology –** defined by Bertin, is the study of signs and symbols and their use in communication. In the Context of data visualization, it refers to the study of graphical elements and ability to convey meaning. [Bertin introduced concept of **Visual Variables** which are the fundamental graphical elements that encode data. (position, size, color, shape, orientation, texture etc) -> rentinal variables.] / [**Visual Encoding** involves slecting appropriate visual variables.] / [**Visual Hierarchy**: guide viewer’s attention and comprehension. Highlight important information and relationships.] / [**Graphic Structures** involve arranging and organizing visual elements to reveal patterns, trends, relationshipts in data]

**Cleveland & McGill : Quantitative perception –** (More accurate) Position > Length > Angle, Slope > Area > Volume > Color, Density (Less accurate)

**Bertin’s Three levels of reading :** [Elementary: Single Value] / [Intermediate: relationships btw values] / [Global: relationships of the whole]

**3D Graphics Does Not Break the Barrier** – Only adds a single dimension / Creates occlusions / adds orientation complexities / easy to get lost / suggests a physical metaphor

**Telling Effective Stories –** [Trust : a key design issue] / [Expressive: convey the data accurately] / [Effective: exploit human perception – use the graphical vocab appropriately, utilize white space, avoid extraneous material] / [Context: titles, captions, units, annotations..]

**Stories Involve More Than Data** – [Aesthetics: What is effective is often affective] / [Style: Include information about who you are] / [Playful: Allow people to interact with the data views]/ [Vivid: Make data views memorable]

**Summary**

Visualization & Presentation

Human perception is powerful & limited

Coping with Bertin’s Barrier

- Composition

- Interactivity (Sorting, Filtering, Aggregation, Brushing, Linking)

Telling Stories with data

- Trust is a key design issue

- Always question data

**Deduction vs Induction**

In logic, there are two distinct methods of reasoning namely the deductive and the inductive approaches.  
**Deductive reasoning** works from the "general" to the "specific". This is also called a "top-down" approach.  
The deductive reasoning works as follows: think of a theory about topic and then narrow it down to specific hypothesis (hypothesis that we test or can test). Narrow down further if we would like to collect observations for hypothesis (note that we collect observations to accept or reject hypothesis and the reason we do that is to confirm or refute our original theory). In a conclusion, when we use deduction we reason from general principles to specific cases, as in applying a mathematical theorem to a particular problem or in citing a law of physics to predict the outcome of an experiment.

**Inductive reasoning** works the other way around, it works from observation (or observations) works toward generalizations and theories. This is also called a “bottom-up” approach. Inductive reason starts from specific observations (or measurement if you are mathematician or more precisely statistician), look for patterns (or no patterns), regularities (or irregularities), formulate hypothesis that we could work with and finally ended up developing general theories or drawing conclusion.  
Note: that that is how Newton reached to "Law of Gravitation" from "apple and his head” observation“  
In a conclusion, when we use Induction we observe a number of specific instances and from them infer a general principle or law. Sherlock Holmes never uses deductive reasoning to assist him in solving a crime. Instead, he uses inductive reasoning. ... Deductive reasoning starts with a hypothesis that examines facts and then reaches a logical conclusion. In math terms, think of it this way: A=B, B=C, therefore A=C For deductive reasoning to work, the hypothesis must be correct. Inductive reasoning starts with observations that produce generalizations and theories.

Deduction: Theory->Hypothesis->Observation->Confirmation

Induction: Observation->Pattern->Hypothesis->Theory

**Lecture1** : Tableau Background – [Powerful data visualization software] / [Capable of creating various interactive visualizations from a multitude of data sources] / [Commercial Software, but students free] / [primarily a drag-and-drop]

Gateway – Handles processes such as load balancing, traffic routing, URL rewriting, serving static files to clients, serving multi-thread processes etc. The gateweay sever used by Tableau Is Apache Tomcat

ApplicationServer – Handles login processes, domain authentication, data authorization, user or group permission management, content searches

Repository – Server metadata related to users, permissions, assignments, groups, projects / Stores vis. In flat files (TWS, TDS) and performance data for auditing

VizQL Server – Loads all the Vis. / Cahcing for performance improvement and editing tasks

Data Engine – Handles processes related to Tableau data Extract (TDE). Invoked only when a query is shot, which involves data from TDE. / Stores multiple TDE and can run on multiple servers

Backgrounder – Manages Schedules for info. Refreshing and ensures proper functioning of the tableau sever and DE

Data Server – Central metadata management, driver deployment, extract management / Access control and serves as a proxy to the data sources / Hosts user queries and requests to prevent users from directly accessing the data source

Search and License – Manages the search indexing for the data in the repository. Whereas, the license component is responsible for the licensing and configuration

**Data Sources and Types of Vis.** – Can Connect to variety data sources : [Local files (excel, text, access)] / [Traditional databases (SQL server, MySQL, Oracle, PostgreSQL, DB2)] / [Cloud Technologies (Amazon Aurora, EMR, Redshift, BigQuery)] / [Big Data Technologies (Hadoop, Hive, SparkSQL)

Can create variety of Vis. : [Basic bar and line charts / Geospatial analysis / Word Clouds / Treemaps / Network analysis]

These Vis. Can be combined into interactive dashboards (can be published and shared easily)

텍스트, 도표, 소프트웨어, 스크린샷이(가) 표시된 사진

자동 생성된 설명

**Advanced Analytics –** [A grouping of analytic techniques used to predict future outcomes (IBM)] / [Provides algorithms for complex analysis of either structured or unstructured data. Include sophisticated statistical models, ML, etc (tdwi)] / [The fastest growing segment of the BI and analytics software market (Gartner)]

**Segmentation and Cohort Analysis** : What Tableau can do – [Power and Simplicity] / [Drag Drop segment.] / [Explore distributions] What Tableau can’t do – [Simple set operations – sets] / [Seamless updates when data change-live conn.]

**What-if Analysis** – Simple controls (Parameters) / Full platform integration / Snapshot interesting results (Stories)

**Sophisticated Calculation** – Expressibility / Flexible aggregations (LOD) / Result set computations (Table Calc.)

**Time Sereies Analysis** – Seasonality Exploration / Flexible Sampling / Intuitive Aggregations / Windowed Calculation / Relative date filters (moving average of stock price)

**Predictive Analysis** – Integrated analytics objects (Trendlines) / Simple Quality Metrics / Advanced predictive (Forecasts)

**R Integration** – Virtually unlimited choice of methods / Leverage prior work / Visualize and interrogate model results

**Lecture2** : **Splunk Overview**

**What is Splunk?** – Aggregate, analyze, and get answers from machine data (Application management, Operations Management, Security & Compliance..)

**What Data?** – Index Any data from Any source (computers, network devices, virtual machine, internet device, communication device, sensor, database, logs, messages, config, alerts, call detail, clickstream, metrics..)

**How Splunk work? –** [Splunk Forwarders] -> [Splunk Indexer] <- [Splunk Search Head]

스플렁크는 나중에 채워넣기

**Lecture3:ELK (Elasticsearch / Logstash / Kibana)**

**E:** an opensource, RESTful, distributed search and analytics engine built on Apache Lucene. Support for various languages, high performance, and schema-free JSON documents makes Elasticsearch an ideal choice for various log analytics and search use cases.

**L:** open source data ingestion tool that allows to collect data from variety of sources, transform it, and send it to your desired destination. With pre-built filters and support for over 200 plugins, Logstash allows users to  
easily ingest data regardless of the data source or type.

**K:** open-source data visualization and exploration tool for reviewing logs and events. Kibana offers easy-to-use, interactive charts, pre-built aggregations and filters, and geospatial support and making it the preferred choice for visualizing data stored in Elasticsearch.

텍스트, 스크린샷, 폰트, 웹 페이지이(가) 표시된 사진

자동 생성된 설명

**Why ELK? –** [Link to Google Trends] / [Only open source search-engine with an official Python API]

ELK에 관한 내용 나중에 채워넣기

**Lecture4**