Class 10 Data Visualization & Information Graphics

BIG PICTURE: HUMAN DESIRE TO UNDERSTAND, KNOW AND PERHAPS CONTROL LIFE – FROM OUR BODIES TO OUR ENVIRONMENT TO OUR UNIVERSE.

Slide
Tracks of Yu – Stone engraving, 32" x 31" - Extraordinary due to scale and accuracy, nothing like it in Europe until 450 years later – circa 1550, Dieppe maps, France – are a series of world maps produced by the Dieppe school of cartographers.

Slide
Andreas Vesalius (31 December 1514 – 15 October 1564) referred to as the founder of modern human anatomy was an anatomist, physician, and author of one of the most influential books on human anatomy - On the Fabric of the Human Body.

Image, Word, Number combine to present evidence and make an argument.

Data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication. Important stories live in our data and data visualization is a powerful means to discover and understand these stories, and then to present them to others. The information is abstract in that it describes things that are not physical. Statistical information is abstract. Whether it concerns sales, incidences of disease, athletic performance, or anything else, even though it doesn't pertain to the physical world, we can still display it visually, but to do this we must find a way to give form to that which has none. This translation of the abstract into physical attributes of vision (length, position, size, shape, and color, to name a few) can only succeed if we understand a bit about visual perception and cognition. In other words, to visualize data effectively, we must follow design principles that are derived from an understanding of human perception.
   - Interaction Design Foundation

We acquire more information through vision than through all of the other senses combined. The 20 billion or so neurons of the brain devoted to analyzing visual information provide a pattern-finding mechanism that is a fundamental component in much of our cognitive activity… the human visual system is a flexible pattern finder, coupled with an adaptive decision-making mechanism. And today we have vast computational power and information resources through the World Wide Web. Improving the interfaces can substantially improve the performance of the entire system.

Slide
Sports - In the late 1990s, Billy Beane, the General Manager of the Oakland Athletics and his assistant GM used data to discover which statistics really mattered in Baseball. Instead of the flashy statistics that fans usually paid attention to such as

home runs and stolen bases, they focused on batters getting on base, pitch counts, even where the ball should be hit. This sabermetric approach to assembling a team is represented in the book and film Moneyball.

The emphasis of data visualization is to use past and current information to peer in to the future. Is data analysis and visualization today's crystal ball?

Infographics or graphics that present complex information quickly and clearly, such as in signs, maps, journalism, technical writing, and education.

Infotainment is "information-based media content or programming that also includes entertainment content in an effort to enhance popularity with audiences and consumers."

Social Networks and Consumerism - From Targeted Advertising to Brand Growth by Exploiting Information retrieved from Social Network APIs.
PBS Frontline – "Generation Like" -
http://www.pbs.org/wgbh/pages/frontline/generation-like/

Journalism – New York Times employs Multimedia, Data Visualization, Explanatory Graphics, Breaking News, Visual and Interactive Features for Interactive Storytelling. Take a look at the 2015 interactive storytelling features:
http://www.nytimes.com/interactive/2015/us/year-in-interactive-storytelling.html
More recently, we can look at the representation of popular and Electoral College votes for the 2016 Presidential Elections and consider what maps present the clearest picture of the U.S. population's voter choice. The first known mapping of U.S. Presidential Elections onto the geography of the country is from 1880. According to Susan Schulten author of Mapping the Nation:

> A map allowed one to correlate, in a systematic fashion, the possible geographical influences at work, including topography, climate, or historic patterns of settlement. Perhaps some counties drew particular industries that spawned protective tariffs, while others had historic traditions of abolition. All of these influences could shape the spatial pattern of party identity.

Unfortunately, the majority of Election Day mapping of the popular vote do a poor job of representing the reality of the country's population. Over the past 16 years, the large interactive screens have become very popular, but pools of solid color do not reflect the number of votes in any given state. For example, a mapping of this years elections using only solid red and blue would reflect a landslide of the popular vote for Trump when in reality, Hillary Clinton won the popular vote by over two million voters. If one calculates the voting power based on population density vs electoral votes, Arizona has nearly 200x the voting power of NY.

Slides
Steve Ballmer, past CEO of Microsoft and current owner of Clippers decided to investigate our money - the United States government budget and try to

find out where it comes from, where it goes and what are the results. With a small team, Ballmer spent a year searching for this information. Public information that was incredibly hard to find and what was discovered was a mess that needed to be sorted.

The vast amount of the government budget comes from our taxes and is used to fund public education, policing, national defense, social security, medicare, the environment and natural resources, civil rights, elections… Ballmer and his team developed an online visualization tool that helps us understand our government budget: https://www.usafacts.org/

The spending section is divided by the lofty concepts of the preamble to the constitution. Drill down from these ideals and you will see the government spending on our every day realities such as education.

Why is this tool important? What can be done with it? For one thing, we can consider what things we may want to fight for as citizens. For example, if the higher education budget was higher, could it bring down the law enforcement and corrections budget? And the American Dream - how is that represented? What does it mean that the voter turn out is so low? Having this information may stir one to work toward changing this reality.

Slides
William Playfair (1759-1823), Scottish engineer and political economist, invented four types of diagrams: in 1786 the line graph and bar chart of economic data, and in 1801 the pie chart and circle graph, used to show part to whole relations. He is commonly considered the founder of graphical methods of statistics for his book Commercial and Political Atlas published in 1786 that presented data in a graphical format.

Line Graph - Exports and Imports between Denmark /Norway and England over 80 years by 10 thousand pounds.

Bar Chart - Exports and Imports of Scotland to and from different parts for one Year from Christmas 1780 to Christmas 1781.

Pie Chart - From the book - Statistical Breviary, 1801 "Chart Representing the Extent, Population to Revenues of the Principal Nations in Europe"
"Each of these circles has two tangent lines, a yellow one for population and a red one for revenue. At a glance, then, one trained to interpret this visual information can make a comparison of the population and revenue of each nation relative to its geographical extent. It is, in the end, a visually innovative way to make the argument that, so long as wealth is defined quantitatively as revenue, it cannot be equated with the extent of a nation's territory—**an argument, in other words, against those who argued that territory and population constituted a nation's wealth."**
Russian Empire, Turkish Empire Sweden Emperor's Dominion, France, Denmark,

German Empire, Spain, Britain and Ireland, Countries under the Dominion of France, Portugal

Slide
Florence Nightingale (1820–1910).
This "Diagram of the causes of mortality in the army in the East" was published in Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army" sent to Queen Victoria in 1858. The polar chart demonstrates that lives could be saved by having medical aid on the battleground.
Created during the **Crimean War (Oct. 1853-Feb. 1856)** conflict between Russian Empire and alliance of French, British, Ottoman Empires and Kingdom of Sardinia. It was for control of the areas that belonged to the declining Ottoman Empire.
Polar Chart – multiple data sets (three or more quantitative variables represented) in the form of a two-dimensional chart on axes starting from the same point.
This graphic indicates the number of deaths that occurred from preventable diseases (in blue), those that were the results of wounds (in red), and those due to other causes (in black). In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black. The entire areas may be compared by following the blue, the red, & the black lines enclosing them.
Florence Nightingale, OM, RRC (1820 -1910) was a celebrated English social reformer and statistician, and the founder of modern nursing. She came to prominence while serving as a nurse during the Crimean War, where she tended to wounded soldiers. She was dubbed "The Lady with the Lamp" after her habit of making rounds at night. In 1860, Nightingale laid the foundation of professional nursing with the establishment of her nursing school at St Thomas' Hospital in London. - Wikipedia

Slides
Apple Dictionary - **Chart**: a sheet of information in the form of a table, graph, or diagram.
Statistics is the study of the collection, organization, analysis, interpretation, and presentation of data. It deals with all aspects of this, including the planning of data collection in terms of the design of surveys and experiments.

A chart is a graphical representation of data, in which the data is represented by symbols, such as bars in a bar chart, lines in a line chart, or slices in a pie chart. - Wikipedia

Pie Chart is amongst the most commonly used statistical chart in the business world and the mass media.

Column Chart – compare values across categories.

Pie Chart - Display contribution of each value to a total. A pie chart is ideal for a small single dataset, once that data set grows it is not very effective.

Line Chart – displays trends over time. Gross domestic product (GDP) is the market value of all officially recognized final goods and services produced within a country. When put in to the context of the last 80 years, the 2008 flash crash is just a bump. Tells a story

Slides
John Snow (1813-1858) was an English physician and a leader in the adoption of anesthesia and medical hygiene. He is considered to be one of the fathers of modern epidemiology.
Epidemiology is the study of the patterns, causes, and effects of health and disease conditions in defined populations. It is the cornerstone of public health, and informs policy decisions and evidence-based medicine by identifying risk factors for disease and targets for preventive medicine. Epidemiologists help with study design, collection and statistical analysis of data, and interpretation and dissemination of results.

THE BROAD STREET PUMP map, which represents the first use of a spot map in epidemiology. The text contains the substance of all of Snow's articles published since the first edition of 1849, 'together with much new matter' (Preface, p. iii). In this edition Snow provides detailed historical and statistical evidence for his conviction that cholera is a contagious disease that attacks the alimentary canal and is communicated primarily through contaminated water.
On the Mode of Communication of Cholera by John Snow

September 1854 Cholera epidemic breaks out in central London in the area of Broad street - from August 19 to the end of September there were 616 deaths.

John Snow a doctor who had investigated previous epidemics suspected the water from a community pump-well at Broad and Cambridge. He tested the water on Sept. 3rd and no impurities were visible. However, there was no other common element to the area so he continued to test the water and found that the water varies in the amount of organic impurity.

From the General Register Office, Snow obtained a list of 83 deaths from cholera. The data listed the victim's names, residence and date of death. A time-series display or chronology of death certificates would be useless to finding the source. Snow plotted this data to a map as horizontal marks for deceased and a target for pumps in the neighborhood. When plotted on a map, these data showed a close link between cholera and the Broad Street pump. House by house detective work yielded detailed evidence about a possible cause-effect relationship, as Snow made a street corner correlation:

> On proceeding to the spot, I found that nearly all of the deaths had
> taken place within a short distance of the pump. There were only ten
> deaths in houses situated decidedly nearer to another street pump. In

five of these cases the families of the deceased persons informed me that they always sent to the pump in Broad Street, as they preferred the water to that of the pump which was nearer. In three other cases, the deceased were children who went to school near the pump in Broad Street. Two of them were known to drink the water; and the parents of the third think it probable that it did so. The other two deaths, beyond the district which this pump supplies, represent only the amount of mortality from cholera that was occurring before the irruption took place.

With regard to the deaths occurring in the locality belonging to the pump, there were sixty-one instances in which I was informed that the deceased persons used to drink the pump-water from Broad Street, either constantly or occasionally. In six instances I could get no information, owing to the death or departure of every one connected with the deceased individuals; and in six cases I was informed that the deceased persons did not drink the pump-water before their illness.
 - Dr. John Snow

Brewery - not a single death although it has 70 employees. The men were allowed a certain quantity of beer and the brewery has its own deep well. The owner was certain that none of the employees drank from the Broad Street well.

Work House, a prison in which petty offenders are expected to work, had 535 inmates, only five died. The workhouse has its own pump well and they received a water supply from a private water works. The residents never sent for Broad Street water. "If the mortality in the workhouse had been equal to that in the streets immediately surrounding it on three sides, upwards of one hundred persons would have died."

Others who died and lived far from the pump either worked or went to school near the pump and took water from it.

Snow described his findings to the authorities responsible for the community water supply. The board then ordered the pump-handle on Broad Street removed immediately and the epidemic soon ended.

Snow's explanation replaced previously held beliefs that cholera spread through the air or through sexual intercourse or by some other means, including the fantastic theory that cholera vaporously rose out of the burying grounds of plague victims from two centuries earlier. In 1886 the discovery of the bacterium Vibrio cholera confirmed Snow's theory. He is still celebrated for establishing the mode of cholera transmission and consequently the method of prevention: keep drinking water, food and hands clear of infected sewage.
- Edward Tufte

Slide
In the case of the 1854 London cholera epidemic, Dr. John Snow the collection, organization, analysis, interpretation, and presentation of data

Today, when we speak of data visualization, we tend to think of computational data gathering. And data collection has gotten better and better due to the speed and capacity of today's computers as well as the expansion of the internet in to all sectors of our day-to-day lives.

Big data usually includes data sets with sizes beyond the ability of commonly-used software tools to capture, curate, manage, and process the data within a tolerable elapsed time.

Examples include web logs, RFID, sensor networks, social networks, social data (due to the social data revolution), Internet text and documents, Internet search indexing, call detail records, astronomy, atmospheric science, genomics, biogeochemical, biological - Wikipedia

Slide 38
Data Visualization will require some combination of the following:
Acquire
Obtain the data, whether from a file on a disk or a source over a network.
Parse
Provide some structure for the data's meaning, and order it into categories.
Filter
Remove all but the data of interest.
Mine
Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context. For example - identify the maximum and minimum of a data set.
Represent
Choose a basic visual model, such as a bar graph, list, or tree.
Refine (Print or Static Display - Infographics)
Improve the basic representation to make it clearer and more visually engaging.
Interact (Screen-based and Interactive Display)
Add methods for manipulating the data or controlling what features are visible.

Slide
And due to the complexity of data and these various steps, data visualization may require collaboration between diverse fields: statistics, data mining, graphic design, and programming.

Slide – Immersion

Slide/Video - ***Sight*** a short film by Eran May-raz and Daniel Lazo, 2012

Slide - **<u>Information Visualization Perception for Design, Colin Ware</u>**
When the missing pieces are interpreted as foreground objects, continuity between the background letter fragments is easier to infer. Perspective inference by constructing a foreground and background.

We know that the visual system has its own rules. We can easily see patterns presented in certain ways, but if they are presented in other ways, they become invisible. Thus, for example, the word DATA is much more visible in the bottom version than in the one at the top. This is despite the fact that identical parts of the letters are visible in each case and in the lower figure there is more irrelevant noise than in the upper figure. The rule that applies here, apparently, is that when the missing pieces are interpreted as foreground objects, continuity between the background letter fragments is easier to infer. The more general point is that when data is presented in certain ways the patterns can be readily perceived. If we understand how perception works, our knowledge can be translated into rules for displaying information.

Slide
Generally, diagrams are held to be made up of symbols and symbols are based on social interaction. The meaning of a symbol is normally understood to be created by convention which is established by human communication.

A cave drawing of a hunt still conveys much of its meaning across several millennia.

Sensory vs Arbitrary Symbols
Sensory - perception without learning - basis because we share the same visual system, it is likely that we all see in the same way... transcend cultural and racial boundaries
Arbitrary - representation that must be learned
Sensory aspects of visualization derive their expressive power from being well designed to stimulate the visual sensory system. In contrast, arbitrary, conventional aspects of visualizations derive their power from how well they are learned. Sensory is studied through experiments derived by sensory neuroscience while the other from methodologies by structuralist social science. Most visualizations are hybrids, an obvious example is the combination of images and words.

Sensory Immediacy (pre-attentive distinction)
The processing of certain kinds of sensory information is hard-wired and fast.
Orientation, size, curvature, grouping, color, blur, numerosity
Texture segmentation - spatial frequency
Five Regions of Texture
Segmentation - early rapid processing

Slide
Dangers of Potomac River Reprint of a 1985 Washington Post graphic by Johnstone Quinan.

Slide
Technical Questions: Is the visualization appropriate for the data?
How does the visualization fare in terms of usability issues?
How do people respond to one visualization versus another?
Critical Questions: WHO MADE IT? FOR WHOM? AND FOR WHAT PURPOSE?

Today, data provides the means by which science progresses, legislation changes, and society advances. However data is always gathered at a certain time with a certain purpose and to be useful it must be mined, parsed and presented.

Slides
Scientific Practice
Visualization of galaxy formation, predicting weather patterns or oil spill patterns, simulation of electron behavior... Scientific visualization fashions itself as a tool of discovery improved through scientific method. The implicit assumption is that the tool allows us to explore the data without bias. Scientific visualization tends to be specialized, generated and read by experts who understand how to read the data.

Slide/Video
The Powers of Ten films are two short American documentary films written and directed by **Charles and Ray Eames**. Depicts the relative scale of the Universe according to an order of magnitude (or logarithmic scale) based on a factor of ten, first expanding out from the Earth until the entire universe is surveyed, then reducing inward until a single atom and its quarks are observed. The first film: A Rough Sketch for a Proposed Film Dealing with the Powers of Ten and the Relative Size of Things in the Universe[1] — was a prototype and was completed in 1968; the second film: Powers of Ten: A Film Dealing with the Relative Size of Things in the Universe and the Effect of Adding Another Zero[2] — was completed in 1977.
The Powers of Ten films were adaptations of the book Cosmic View (1957) by Dutch educator Kees Boeke.

Mercator (mərˈkātər) **projection** (1569)
cylindrical map projection by the Flemish cartographer Gerardus Mercator
The term "normal cylindrical projection" is used to refer to any projection in which meridians are mapped to equally spaced vertical lines and circles of latitude (parallels) are mapped to horizontal lines.
The mapping of meridians to vertical lines can be visualized by imagining a cylinder whose axis coincides with the Earth's axis of rotation. This cylinder is wrapped around the Earth, projected onto, and then unrolled.
As on all map projections, shapes or sizes are distortions of the true layout of the Earth's surface. The Mercator projection exaggerates areas far from the equator. For example:
Greenland takes as much space on the map as Africa, when in reality Africa's area is 14 times greater and Greenland's is comparable to Algeria's alone.

Alaska takes as much area on the map as Brazil, when Brazil's area is nearly five times that of Alaska.
Finland appears with a greater north-south extent than India, although India's is greater.
Antarctica appears as the biggest continent, although it is actually the fifth in terms of area.
Although the Mercator projection is still used commonly for navigation, due to its unique properties, cartographers agree that it is not suited to general reference world maps due to its distortion of land area. - Wikipedia

Slide
Examples of **Infographics** for Journalism and Education.
Riot Gear's Evolution
NY Times Opinion article by Chi Birmingham and Alex S. Vitale, December 3, 2011
> Technological advances, training innovations and changing attitudes toward the right to assemble have all shaped the way the police handle the challenges of large demonstrations. During the 1960s and '70s, police officers treated many protests as a threat to the social order and responded with brute force. In the 1980s and '90s, demonstrations tended to be less confrontational and the police responded with more accommodating tactics."
> "Following the "Battle in Seattle" protests against the World Trade Organization in 1999, a more restrictive, preemptive and aggressive form of protest policing emerged at the 2003 protests in Miami over the Free Trade Area of the Americas. The current Occupy demonstrations have adopted a defiant and disruptive style of protest that has pushed the boundaries of the First Amendment, but the response has been mixed. While many cities have evicted encampments and blocked marches through enormous shows of force, others have tried to use communication and flexibility to preserve the nonviolent if disruptive character of these protests.

Slides
Art for Advocacy
Street Vendor Guide - INFOGRAPHIC
Candy collaborated with Rosten Woo and John Mangin of the Center for Urban Pedagogy (CUP), Sean Basinski of The Street Vendor Project, and street vendors around NYC to develop this guide so vendors can understand their rights, avoid fines, and earn an honest living.
The team: Sean Basinski, Rosten Woo, John Mangin, and Candy Chang
Hanging out with street vendors like Munnu to understand their experiences
Meeting with street vendors to get feedback on drafts of the guide

Slides
Critical Art
They Rule – DATA VISUALIZATION

Aims to provide a glimpse of some of the relationships of the US ruling class. It takes as its focus the boards of some of the most powerful U.S. companies, which share many of the same directors. Some individuals sit on 5, 6 or 7 of the top 1000 companies. It allows users to browse through these interlocking directories and run searches on the boards and companies. A user can save a map of connections complete with their annotations and email links to these maps to others. They Rule is a starting point for research about these powerful individuals and corporations. http://www.theyrule.net/

Out of Sight, Out of Mind – DATA VISUALIZATION
A visualization of all documented drone strikes in Pakistan since 2004
Since 2004, the US has been practicing in a new kind of clandestine military operation. The justification for using drones to take out enemy targets is appealing because it removes the risk of losing American military, it's much cheaper than deploying soldiers, it's politically much easier to maneuver (i.e. flying a drone within Pakistan vs. sending troops) and it keeps the world in the dark about what is actually happening. It takes the conflict out of sight, out of mind. The success rate is extremely low and the cost on civilian lives and the general well-being of the population is very high. This project helps to bring light on the topic of drones. Not to speak for or against, but to inform and to allow you to see for yourself whether you can support drone usage or not. http://drones.pitchinteractive.com/

Marcela Armas – ART
I-Machinarius " consists of a map of Mexico turned upside down, outlined by moving chains that drip oil down the wall all the way to the museum floor, creating a growing, slippery puddle and evoking saddening thoughts of all the Mexican oil that flows to the United States, like its best citizens. A blood-curling work, visually strong, ugly, harsh, yet at bottom fragile and slow…"
http://www.artnexus.com/Notice_View.aspx?DocumentID=24284

Eve Mosher – DATA VISUALIZATION
The water line is rising along NYC's waterfront, and public works artist Eve Mosher is using her High Water Line project to make sure that we understand the very real effects of climate change along our shared urban coastline. From May to October 2007, Mosher is drawing (by-hand or pushcart) a white chalk line through the waterfront communities of Brooklyn and lower Manhattan in order to illustrate the 10-feet above sea level mark that potentially threatens unsuspecting neighborhoods, commercial zones, city streets, and private residences. High Water Line marks the extent of increased flooding brought on by stronger and more frequent storms as a result of climate change.

Slide 79
Extra: Download and read "Visualizing Information for Advocacy"