

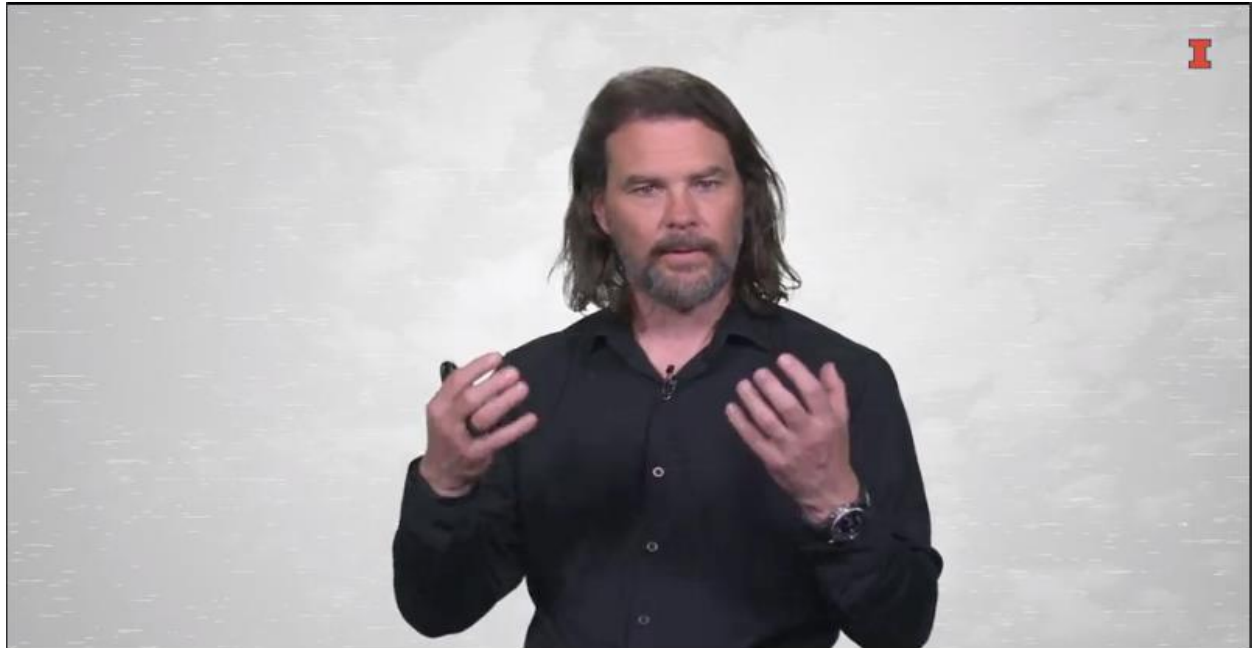


| | |
|--|----|
| About the Course and Your Classmates | 2 |
| Meet Professor Kevin Hartman | 2 |
| Module 1 | 3 |
| Welcome to Introduction to Business Analytics: Communicating with Data | 3 |
| Module 1 Overview | 13 |
| Lecture 1-1: Studying The History of Dataviz | 14 |
| Lecture 1-2: Understanding Today's Dataviz Tools | 25 |
| Lecture 1-3: Making Connections With Visuals | 29 |
| Lecture 1-4: Evaluating The Effectiveness Of Dataviz | 37 |



About the Course and Your Classmates

Meet Professor Kevin Hartman

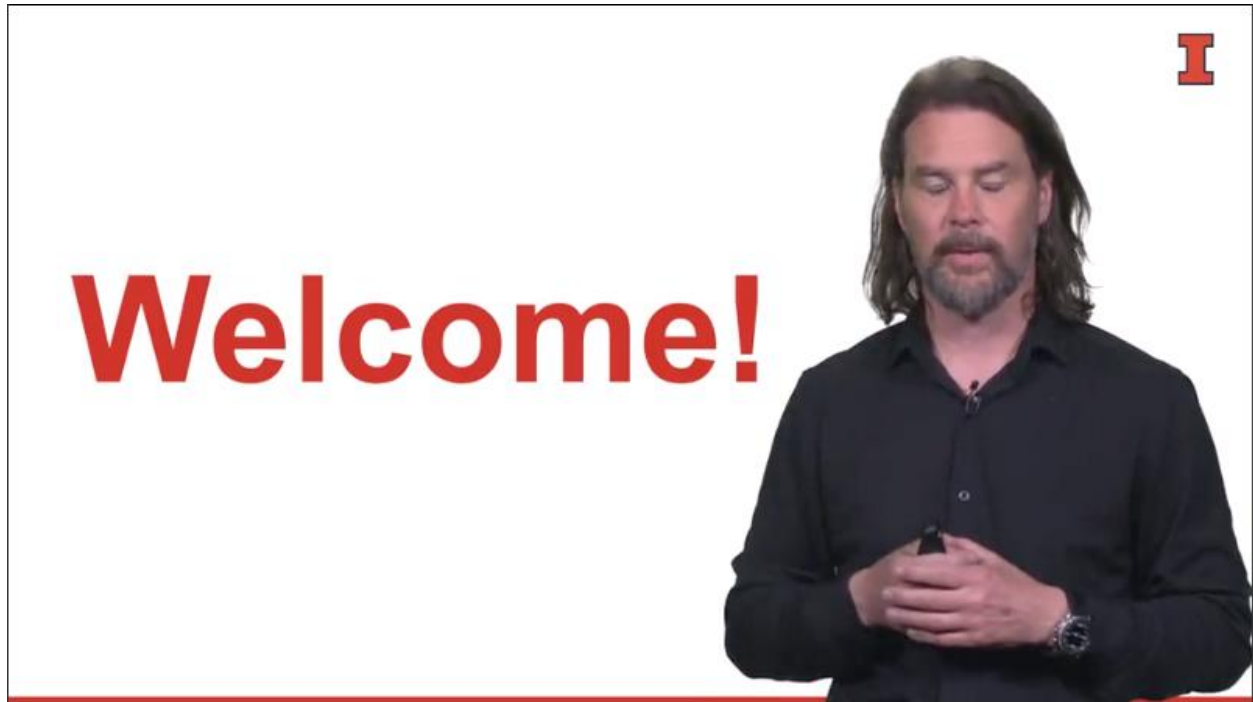


Hello, I am Kevin Hartman, I am the instructor for communicating with data, I am also head of analytics at Google out of our Chicago offices. I've been there for at this point around six years of my career prior to Google I was at an advertising agency for a long time, prior to that I was in consulting before picking up a couple of masters degrees, one an MBA and another master of public policy. The one thing that has kind of marked my professional career, no matter where I was, if it was in consulting at an agency or here with Google now as I am today, it's been this idea of telling stories from data. It's something that I love. I've always felt really passionate about it. I was not always very good at it, it's not an easy thing to do but you learn in the 20 or so years that I've been doing this kind of work, a few things here and there along the way. You learn how to do it the right way, you learn how to do it the wrong way, so that you can you can kind of build on those habits that you should have and avoid those habits that you should not. Teaching the course through the University of Illinois and through Coursera is a wonderful experience. You get to meet so many bright, young as well as seasoned, and experienced people from all over the world. The amount of ideas that are exchanged is tremendous, the quality of the thought and the effort that goes into the courses are tremendously impressive. For all those reasons, I am so happy to be here, so happy to be a part of this course and so happy to be a part of your experience, and really looking forward to over the four modules that I get to teach, watching you learn, watching you develop, and at the end coming out a professional that knows so much more about communicating with data and it's so much more effective in that art.



Module 1

Welcome to Introduction to Business Analytics: Communicating with Data



Hello, and welcome to communicating with data. I am Kevin Hartman, I am head of analytics at Google and I will be your instructor for this course. I am super pleased to be here. Data visualization and the practice of communicating with data is something that I feel very passionate about. We're going to take a very long view of this practice. We are not simply going to focus on creating beautiful images, we're going to talk about the entire process from beginning to end and how you develop data visualization and how you effectively communicate with data. We have broken this course into a number of different modules, four of them.



Syllabus Review



Topics and materials for Module 1

| Module | Discussion Topics | References and Readings |
|--------|---|--|
| 1 | Pictures You See with Your Brain —Frameworks and approaches behind data collection and analysis methods used by business professionals | <p>Berinato: Introduction, Chapters 1, 2, and 9</p> <p>A word from a founding father of web analytics: http://marketinggeek.blogspot.com/2006/02/geeks-in-marketing.html</p> <p>Comparing Data Visualization Software—Here Are the 7 Best Tools for 2018: https://www.forbes.com/sites/bernardmarr/2018/06/20/comparing-data-visualization-software-here-are-the-7-best-tools-for-2018/</p> <p>A Curated Guide to the Best Tools, Resources, and Technologies for Data Visualization: http://dataviz.tools/</p> <p>The "Door" Study—YouTube: https://www.youtube.com/watch?v=FWSxSQsspiQ</p> <p>What Makes a Good Visualization? http://www.informationisbeautiful.net/visualizations/what-makes-a-good-data-visualization/</p> |

In the first one, we're going to give some background on the way the brain actually works. We're going to look at the history of data visualization. We're going to set the context for us as data viz artists in what this practice really is and how we can be effective in the production of data visualizations.

Syllabus Review



Topics and materials for Module 2

| Module | Discussion Topics | References and Readings |
|--------|--|---|
| 2 | Working Fast and Thinking Slow —Data collection, analysis methods, visualization techniques, and tools used by business professionals based on data volume, velocity, and usability | <p>Roles Required for Successful Web Analytics Endeavors: http://marketinggeek.blogspot.com/2006/11/data-visualization-practice.html</p> <p>Big Data, Big Deal: http://www.stephen-few.com/blog/2012/09/19/big-data-big-deal/</p> <p>Big Data, Big Dupe—A Progress Report: http://www.stephen-few.com/blog/2018/02/23/big-data-big-dupe-a-progress-report/</p> <p>Different Tools for Different Tasks: http://www.stephen-few.com/blog/2018/02/19/different-tools-for-different-tasks/</p> |

In the second module, we're going to take a long good look at the entire process of data collection, story development, goal identification, all the component parts that are very



important to completing a final database that allows us to effectively communicate our insight through visuals.

Syllabus Review



Topics and materials for Module 3

| Module | Discussion Topics | References and Readings |
|--------|--|--|
| 3 | Finding Your Data Story —Understanding visual form and how to begin the transition from work product to client-ready graphics | Berinato: Chapters 3 and 4 Wong: Introduction, Chapter 1 The World's Best Data Visualizations (2017): http://www.informationisbeautiful.net/2017/the-best-infographics-and-data-visualizations-2017/ FlowingData's R Visualization Tutorials: https://flowingdata.com/category/tutorials/ |

In the third module, we're going to start looking at images and visuals that communicate meaning and insight. We are going to find those patterns that you'll be looking for in data and we'll show you techniques and tips and how to identify them.

Syllabus Review



Topics and materials for Module 4

| Module | Discussion Topics | References and Readings |
|--------|--|--|
| 4 | Getting Your Story Across —Distinctions between effective and poor data presentation from work product to client-ready graphics | Berinato: Chapters 5–8 and 4 Wong: Chapters 2–5 Chapter 1 Know Your Audience—Good Luck with That: https://www.stephenfew.com/blog/2018/03/15/know-your-audience-good-luck-with-that/ Don't Refrain from Using Uncommon Visualizations—Explain Them: http://www.thefunctionalart.com/2017/12/dont-refrain-from-using-uncommon.html The "Action" Dashboard: http://www.kaushik.net/avinash/the-action-dashboard-an-alternative-to-crappy-dashboards/ |



Then finally, in the fourth module, we will end by bringing our data visualizations from a state of being good in communicating some insight too great and with full attention to detail in a great number of techniques that will get us there. At the end of this, we will know the full process of a communication journey from conceptualizing what we want to talk about to actually on our feet presenting that data visualization in an effective way.

Course Resources



Required reading: Berninato's *Good Charts*



(Berninato, 2016)

In *Good Charts*, Berninato provides an essential guide to how visualization works and how to use the new language of dataviz to impress and persuade. Dataviz today is where spreadsheets and word processors were in the early 1980s—on the cusp of changing how we work.

This book is much more than a set of static rules for making visualizations. It taps into both well-established and cutting-edge research in visual perception and neuroscience, as well as the emerging field of visualization science, to explore why good charts (and bad ones) create “feelings behind our eyes.”

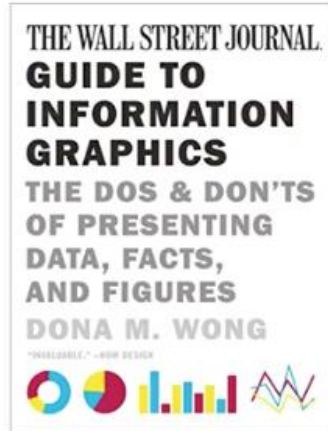
Good Charts will help you turn plain, uninspiring charts that merely present information into smart, effective visualizations that powerfully convey ideas.

There are a few resources that we're going to use in this course. One of them is this book, *Good Charts* by Scott Berninato. This book gives great insight into the process you go through in creating data visualization. Like this course, it does take a very long view of that practice and there are several really important concepts that we're going to pull from this work.

Course Resources



Required reading: Wong's *The WSJ Guide*



(Wong, 2013)

The definitive guide to the graphic presentation of information.

Information graphics is rarely taught in schools or is the focus of on-the-job training. Now, for the first time, Dona M. Wong, a student of the information graphics pioneer Edward Tufte, makes this material available to all of us. In this book, you will learn

- to choose the best chart that fits your data,
- the most effective way to communicate with decision makers when you have five minutes, and
- how to use color effectively.

Organized in a series of mini-workshops backed up with illustrated examples, this is an invaluable reference work for students and professional in all fields.

Another resource that we'll be using is Donna Wong's Guide To Information Graphics. Now, Donna Wong has put a great deal of effort into the guidelines and the real deep attention required to create data visualization. How you take something that is good and okay and might communicate some insight and use tips and techniques to present great graphics, leveraging pre-attentive attributes, using color and font appropriately. All the real detail that we need when we are presenting our visualizations in prime time. Those will be the primary resources that we'll use.

Course Resources



Recommended reading: McCandless's *Knowledge Is Beautiful*



(McCandless, 2014)

We are living in the Information Age, in which we are constantly bombarded with data on television, in print, and online. How can we relate to this mind-numbing overload?

Enter David McCandless and his amazing infographics: simple, elegant ways to understand information too complex or abstract to grasp in any way but visually. In his unique signature style, he creates dazzling displays that blend facts with their connections, contexts, and relationships, making information meaningful, entertaining, and beautiful.

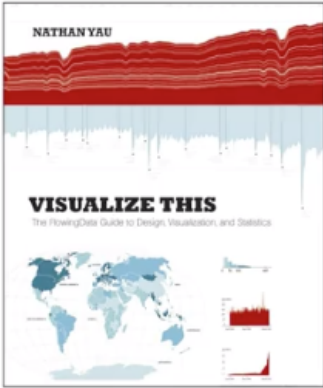
Taking infographics to the next level, his book *Knowledge Is Beautiful* is an endlessly fascinating spin through the world of visualized data, which offers a deeper, more wide-ranging look at the world and its history. Covering everything from dog breeds and movie plots to the origins of life and a timeline of the far future, this stunning book is guaranteed to enrich your understanding of the world.



A couple others that I would recommend any work by David McCandless. David McCandless will play a large role in this course, as we'll reference him quite a bit. Any of the books that he's produced, *Information Is Beautiful*, *Knowledge Is Beautiful*, will serve to inspire us in the creation of data visualization. So, I would recommend any of his works to do just that.

Course Resources

Recommended reading: Yau's *Visualize This*



(Yau, 2011)

This book, by the data visualization expert and the creator of flowingdata.com, Nathan Yau, presents a unique approach to visualizing and telling stories with data.

The book offers step-by-step tutorials and practical design tips for creating statistical graphics, geographical maps, and information design to find meaning in the numbers.

It details tools that can be used to visualize data-native graphics for the web—such as ActionScript, Flash libraries, PHP, and JavaScript—and tools to design graphics for print, such as R and Illustrator.

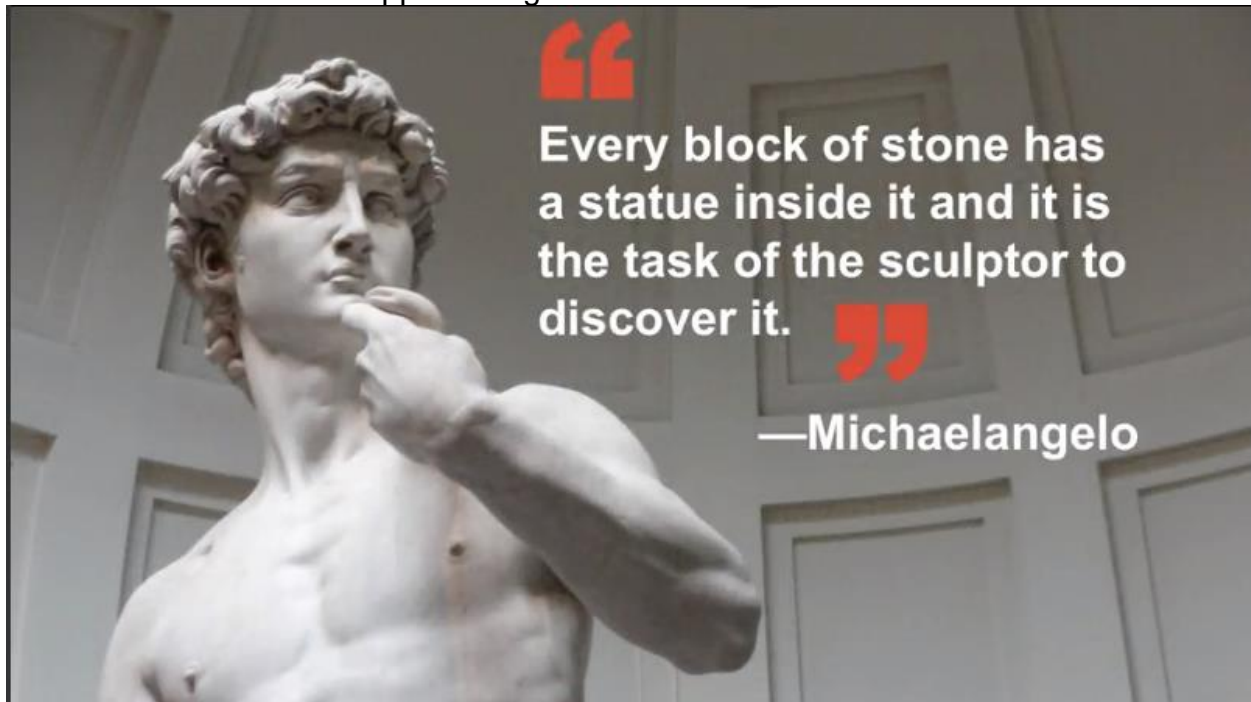
It contains numerous examples and descriptions of patterns and outliers and explains how to show them.

Then a final recommended resource is the book by Nathan Yau called *Visualize This*. Now, Nathan Yau in the flowing data guys do a great job in creating R based case studies in lessons for creating visualization. So, their site is packed with information. But this resource that they produced coupled with that site provides great insight into how you can tactically create visualization using tools and primarily the tool that they'll utilize are. Two required resources for the course, two that I'd recommend that will make you round out your understanding and your experience here.



A Centering Thought ...

Now, before we dive in, I want to provide a bit of a centering thought so that you know the contexts that we'll be approaching data visualization from.



You have probably heard this quote sometime along the way, right? That from Michelangelo stating that, "For every block of stone, there is a statue inside of it and it is a sculptor's job to just chip away what is not required until that statutes revealed." Well, the practice of creating data visualization and communicating effectively with data is much the same way. The only difference being that this is our block of stone. There's so much information that we will be collecting that is unnecessary. Is our job to find the



things that are vital that are required. The guides, the rules that we'll talk about in this course will help direct you there. But this is not intended to be a simple practice and indeed it's not. I think this quote from David McCandless actually addresses that well.

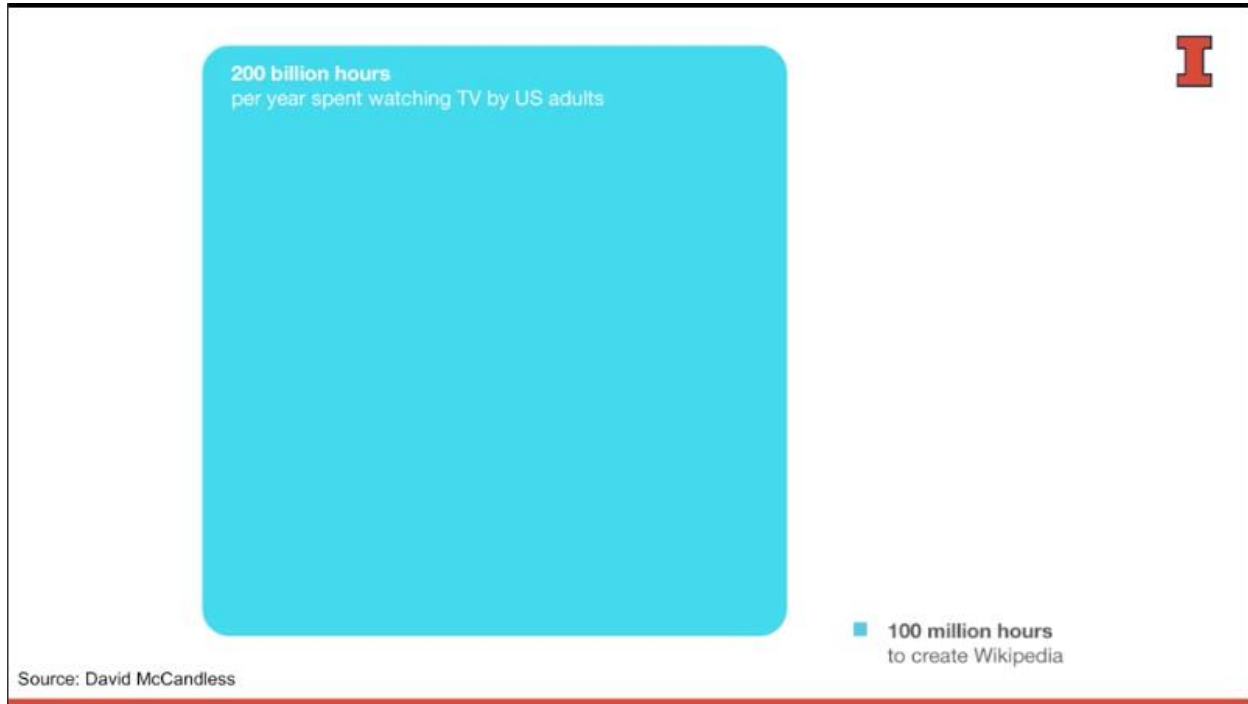


Visualization is hard. I've written books, created software, directed films in my career, but visualization is by far the most challenging discipline I've ever engaged with. It's something about the precision needed at every level, I think. Concept, data, story, design, style—all are precision arts. In visualization, they're stacked one on top of the other. If one sags or slips, the entire edifice can collapse.

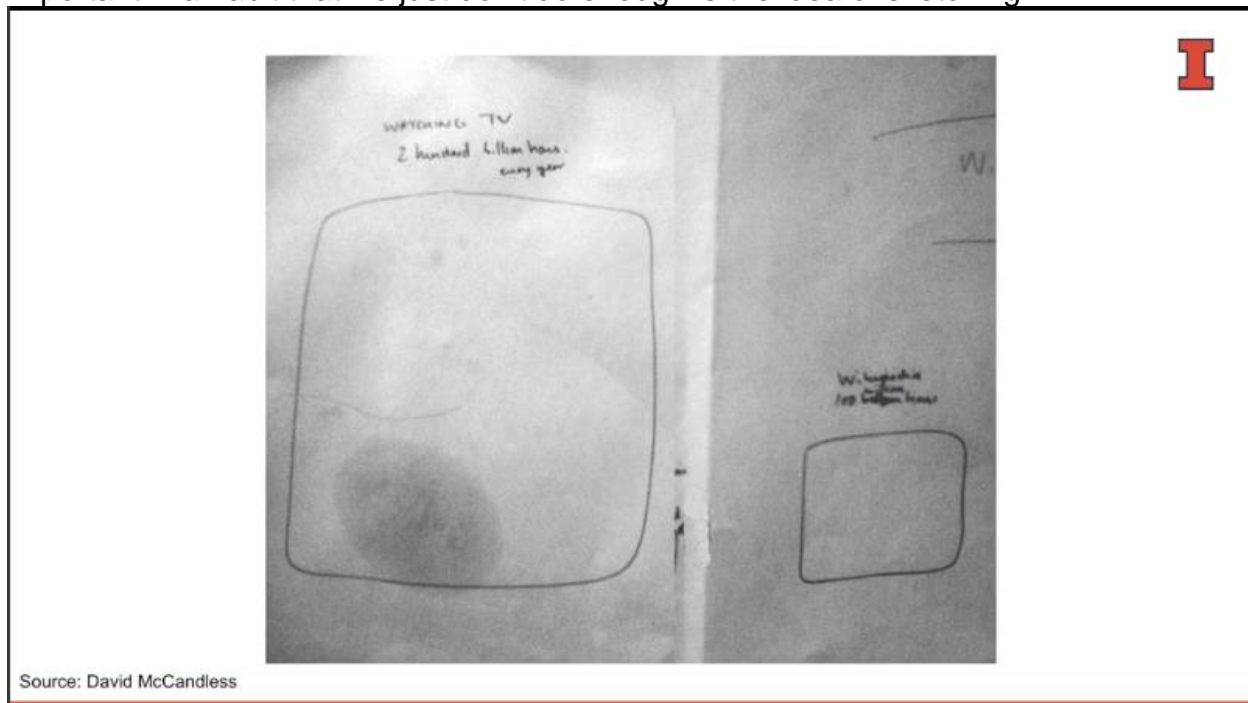


—David McCandless

David McCandless has done a number of things. In his experience, the act of visualizing data is the hardest thing he's had to do, and part of that is because for the reason that there are so many elements that go into a data visualization. Now, in this course, we will spend time on all of them. Again, we're not looking at the pretty picture and this is not a class that will teach you how to produce pretty pictures. We're going to talk about all the context and all the elements around a data visualization to ensure that you are building your visual on solid ground, that you're communicating that visual correctly, and that you are ultimately producing something that allows you to effectively communicate with data to your stakeholders. Now, David McCandless has produced a number of visuals and some of them that are tremendously beautiful and insightful.



One of the ones that I like the most is the example here where he's used contrast to show the amount of time spent by Americans every year watching television. Then conversely in as he introduces contrast, the amount of time required to create a wonderful resource like Wikipedia. Really does put into stark contrast how we spend our time as a country. Now, a visualization like this includes a lot of the elements that I will be talking about and teaching you how to employ as well. One of the things that is also important in a habit that we just don't do enough is the idea of sketching.





This graphic actually comes from a sketchbook that David McCandless carries around and jots down ideas. This is something that we all should be doing. It is the way the masters of this trade actually operate, and it will make us more creative, make us more efficient, save time and ultimately produce a better outcome.

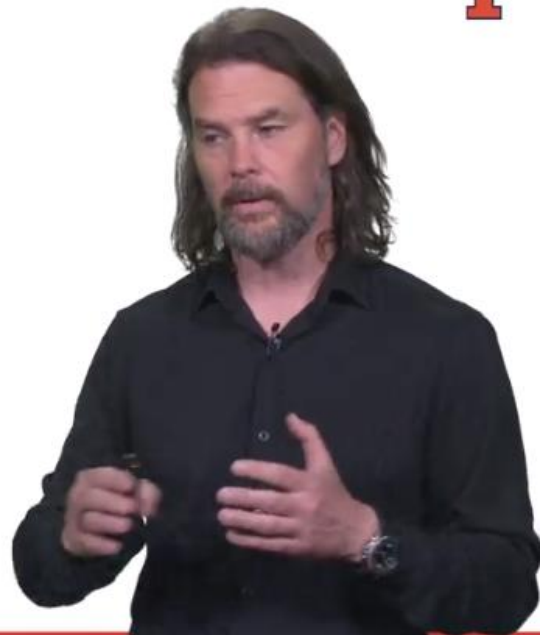
In This Module



Module 1: Pictures You See with Your Brain

Key Concepts

- Studying the history of dataviz
- Understanding today's dataviz tools
- Making connections with visuals
- Evaluating the effectiveness of dataviz



In this first module, there are a number of things that we will talk about, but the primary concepts will be a review of the history of database. We're going to look at 7,000 years of man's experience, humankind's experience expressing themselves through data. There are several things that we can learn from that experience. We're going to take a look at the tools that are used today, a very dynamic and changing market, and I will give you a way to actually evaluate that market and think in terms of the tools that are laid out there. We will talk about how the audience that you will be addressing actually is interpreting physically the visuals that you're putting in front of them. Understanding how their minds work, will help you construct visuals that work in the way that they want those visuals to work. Then finally, we're going to look at a framework that will allow us to answer the question, what makes good data visualization? By understanding that kind of framework and that definition of good, we can build toward that definition. We can construct something that we can be sure is a good final product.



Module 1 Overview

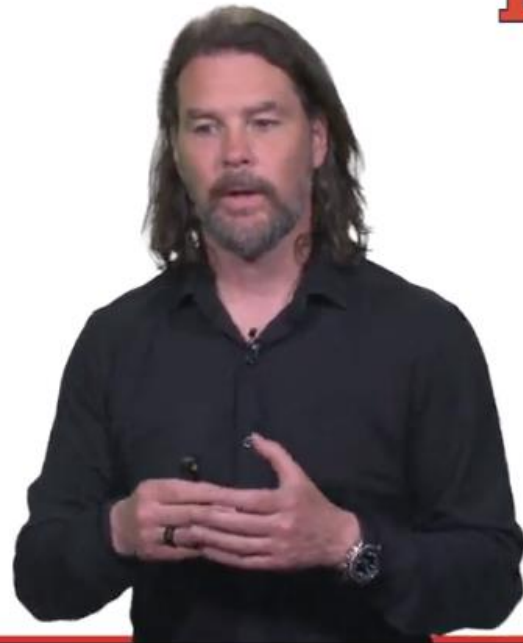
In This Module



Module 1: Pictures You See with Your Brain

Key Concepts

- Studying the history of dataviz
- Understanding today's dataviz tools
- Making connections with visuals
- Evaluating the effectiveness of dataviz



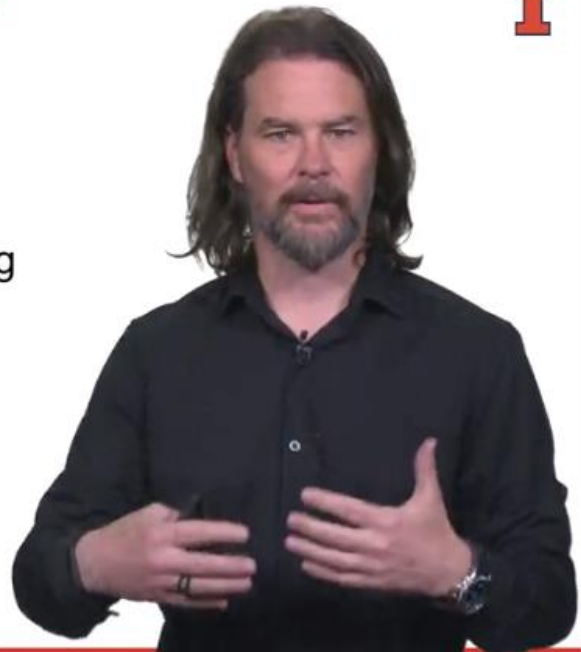
In this first module, there are a number of things that we will talk about, but the primary concepts will be a review of the history of database. We're going to look at 7000 years of man's experience, humankind's experience, expressing themselves through data. There are several things that we can learn from that experience. We're going to take a look at the tools that are used today, a very dynamic and changing market. I will give you a way to actually evaluate that market and kind of think in terms of the tools that are laid out there. We will talk about how the audience that you will be addressing, actually is interpreting physically the visuals that you're putting in front of them. Understanding how their minds work, will help you construct visuals that work in the way that they want those visuals to work. Then finally, we're going to look at a framework that will allow us to answer the question, what makes good data visualization? By understanding that kind of framework and that that definition of good, we can build toward that definition. We can construct something that we can be sure is a good final product.



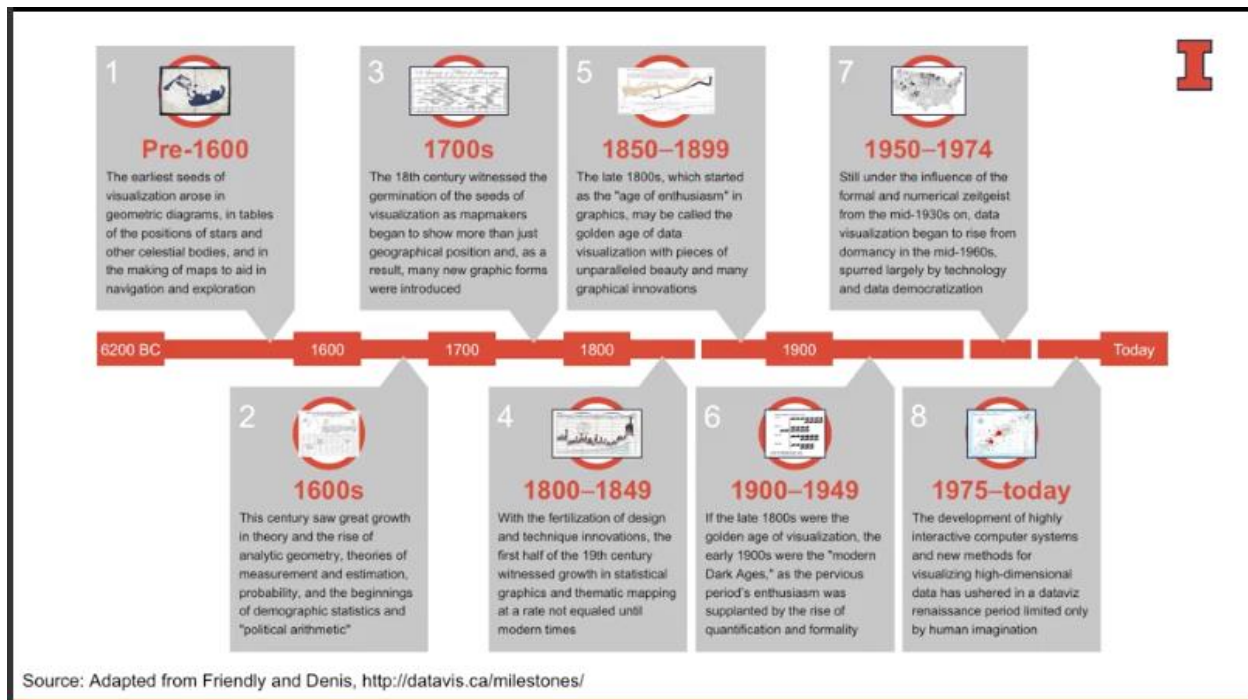
Lecture 1-1: Studying The History of Dataviz

Studying the History of Dataviz

Dataviz has a long history, starting with mapmaking and fanning into statistics, politics, medicine, and other fields.



So, Studying the History of Data Visualization, let's trace that long rich history of humans using data to communicate. What we'll see is this discipline evolving over time, adding capability, confronting a lot of issues that will look very familiar to us today and in the end really emerging as this perfect blend of art and science.





So, this long rich history of humans using data to communicate can really be traced back to nearly the beginning of time in human existence and can best be categorized around eight different epics. We talk about each of them. At each of them there were some inflection point that moves us from one point in this history onto another. In this first epic, what we would call epic one, the early maps, and diagrams phase.

A 6200 BC: Town map, with an erupting volcano (Hasan Daó?) and the Konya Plain, possibly the first map created

B 150: Ptolemy's world map

C 1375: Catalan Atlas

Epoch I: Early Maps And Diagrams (Pre-1600)

A

B

C

Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

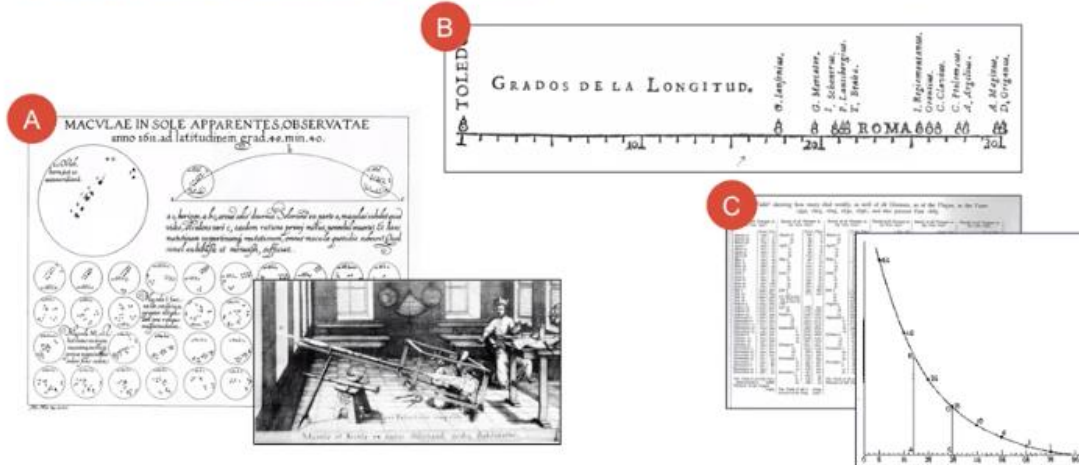
This is pre 1600 and what we're seeing here are humans using data in this case maps to make sense of the world around them and displaying that visually allowing us to share information between each other, to record, and measure things that are in our everyday lives and around us. This is becoming a critical part of who we are, our social interactions, as well as our understanding of the world around us.

A 1626: Christopher Schooner's visual representations used to chart the changes in sunspots over time

B 1644: Michael F. van Langren; variations in determination of longitude between Toledo and Rome

C 1669: Christiaan Huygens; first graph of a continuous distribution function, a graph of Gaunt's life table

Epoch II: Measurement and Theory (1600s)



Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

In the next epic, we move into the 1600's, this epic of measurement in theory. We're actually seeing the capability of measurement grow as we start to make more sense of the world around us and in fact as a great example here Christopher Scooner's visual shows us actually humans developing tools to record data in this case spots on the sun and then translating that into a visual form that communicates the story of those sunspots. Other great examples really important examples follow in this epic as well and

A 1701: Edmond Halley; contour maps showing curves of equal value (an isogonic map, lines of equal magnetic declination for the world, possibly the first contour map of a data-based variable)

B 1765: Joseph Priestley; life spans of 2,000 famous people, 1200 B.C. to 1750 A.D., quantitative comparison by means of bars

C 1782: Charles de Fourcroy; use of geometric, proportional figures (squares) to compare demographic quantities by superposition, an early "tableau graphique"

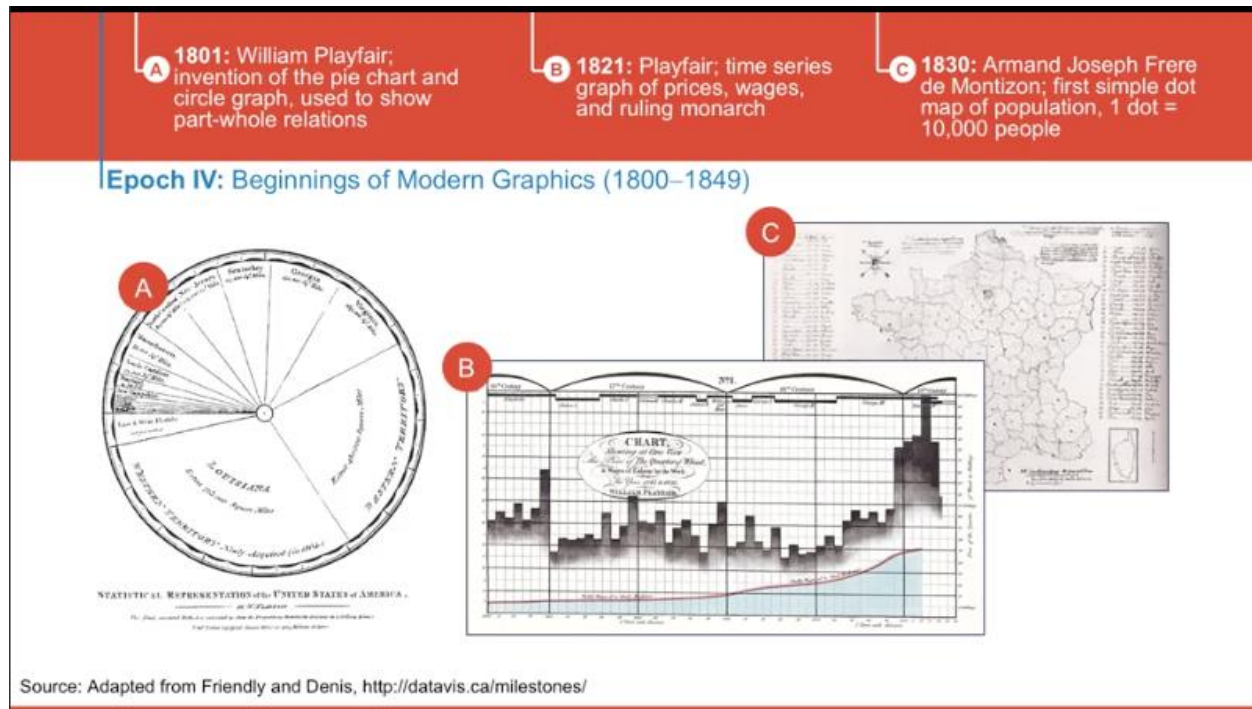
Epoch III: New Graphic Forms (1700s)



Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>



then we move into a third epic of where artists who are familiar with communicating with data began to explore new visual forms, creative forms of communicating data using what today we know to be pre-attentive attributes to get insights in very quickly to their audience including color and shading and others.



Then in the fourth epic, in this what we would call the beginnings of modern graphics, we began to see real true data visualization take root and we are introduced by William Playfair to the pie chart. Something that is still a very important part of any data visualization today, the bar chart, the line chart these visual forms start to take root in this epic that will stay with us now and well into the future.


A 1855: John Snow; use of a dot map to display epidemiological data leads to discovery of the source of a cholera epidemic

B 1857: Florence Nightengale; polar area charts, known as "coxcombs," used in a campaign to improve sanitary conditions in the army

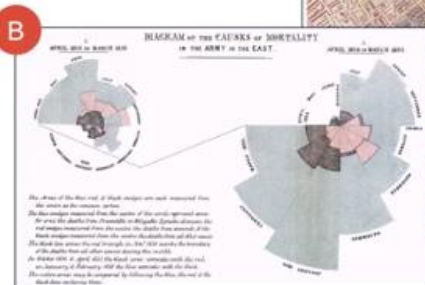
C 1889: Charles Booth; street maps of London showing poverty and wealth by color, transforming existing methods of social survey toward the end of the 19th century

Epoch V: The Golden Age of Data Analytics (1850–1899)


A



B



C

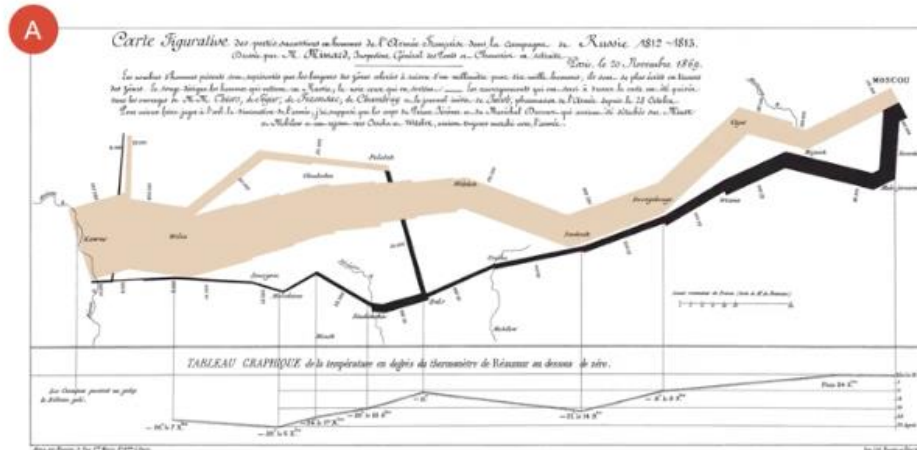


Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

Out of that epic, we move into what we would call the golden age of data analytics and in this phase, the first half or the second half rather of the 19th century, we're actually seeing artists collect data and then use visualization to best communicate that data. Some great examples include John Snow's epidemiological study of cholera in London, in the Broad Street well, understanding through his visual that actually an understanding that changed the way people perceived this disease not as airborne but rather waterborne. Florence Nightingale provides a great example here using visuals to really drive home the need for greater sanitary conditions among the British army and Charles Booth's example of using colour to demonstrate poverty in London grabbing the real emotional sense of that issue.

- A 1869:** Charles Joseph Minard; flow map of Napoleon's March on Moscow frequently called the best graphic ever produced. The illustration depicts Napoleon's army departing the Polish-Russian border bound for Moscow. A thick band illustrates the size of his army at specific geographic points during its advance and retreat. It displays six types of data in two dimensions: the number of Napoleon's troops, the distance traveled, temperature, latitude and longitude, direction of travel, and location relative to specific dates.

Epoch V: The Golden Age of Data Analytics (1850–1899)



Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

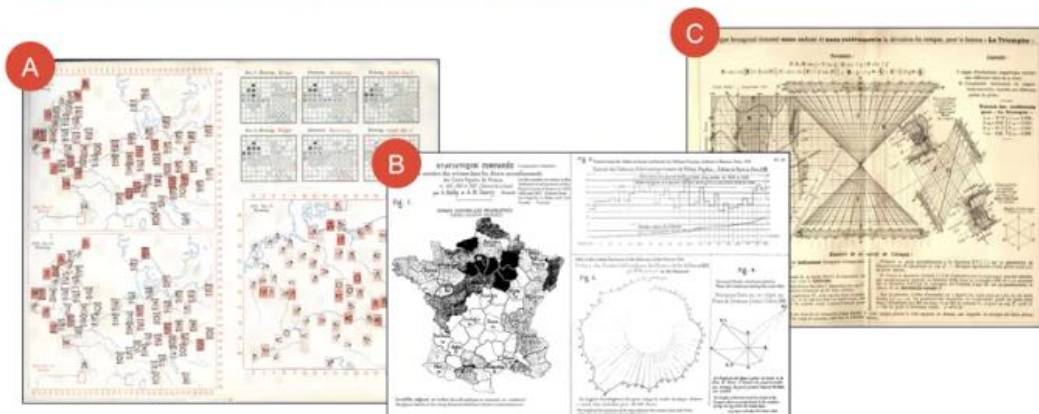
Also in this epic, we're blessed with probably the greatest data visualization of all time. So, much packed into this visual, this visual telling the story of Napoleon's march to Moscow and then return back. The number of soldier lives lost collected in this data, there is temperature, there is distance, there are so many things going on in this beautiful visual and it's serves still today as probably the greatest visual ever created that communicates insight and communicates a story graphically.

- A 1861:** Francis Galton; the modern weather map, a chart showing area of similar air pressure and barometric changes by means of glyphs displayed on a map

- B 1885:** Émile Levasseur; review of all available statistical graphics presented to the Statistical Society of London, classified as figures, maps, and solids

- C 1885:** Charles Lallemand; combination of many variables into multifunction nomograms using 3D, juxtaposition of maps, parallel coordinate and grids

Epoch V: The Golden Age of Data Analytics (1850–1899)



Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>



We do though, in this epic, began to get a little over our skis. Just because we can create graphics, we seem to be in love with creating graphics and so there are several examples of things become a little too complicated overwrought or graphics being brought on top of graphics that do start to obscure much of the meaning that the artists intend.

A 1914: Willard Cope Brinton;
pictograms representing a series of numbers by icons (combining concepts of the bar graph and pictogram of varying sizes)

B 1924: Otto Neurath;
introduction of the ISOTYPE standard (International System of Typographic Picture Education)

C 1932: N.J. Washburn;
spate of articles on experimental tests of statistical graphical forms

Epoch VI: The Modern Dark Age of Data Analytics (1900–1949)

A

Fig. 20. Proportion of College Graduates in Different Professions in 1896-1900 and 1906-1910

Charts of this kind with icons represented in different sizes are usually so drawn that the data are represented by the height of the icons. (Such charts are misleading because, instead of the greatest data, the greatest area is usually that of the icons. Considering the years 1896-1900, the greatest number has about two and one-half times the height of the icons representing public service. The smallest icon represents medicine. Looking at the icons of 1906-1910, it is seen that if the same drawing is represented public service. This kind of graphic work has little real value.)

B

Births and Deaths in Germany in a Year

1911-14
1915-18
1919-22
1923-26

1 child for 250,000 births a year
1 coffin for 250,000 deaths a year

C

Figure 22

Figure 23

Figure 24

Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

Then there's the backlash and we move in the sixth epic into what we will call the modern dark ages of data analytics arise from the academic quarters in revolt for where data visualization has grown to say, we need more accuracy, we need more controls, we need a standardized language, limiting creativity for sake of accuracy and we are introduced to what we know today as a scourge of clip art. These individual icons all representing certain data that are standardized and communicated and frankly threaten to knock the path of data visualization off of its creative tracks. But we do see a few redeeming introductions during this period as well.

- A** **1933:** Henry Beck, an engineering draughtsman, designed a map like an electrical circuit board, using only vertical, horizontal and 45-degree angled lines. He located stations according to available space. The resulting map was geographically inaccurate but easier to use to determine how to get from point A to point B. Beck's idea was soon copied by most subway (and bus) companies around the world.

Epoch VI: The Modern Dark Age of Data Analytics (1900–1949)



Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

One is the famous Henry Beck map of the London underground which beautifully and simplistically describes a very complex system but even more importantly it does not do it with perfect accuracy. It's a revolt against his academic backlash to say that images can be important and can serve a purpose without specificity and without exacting accuracy.

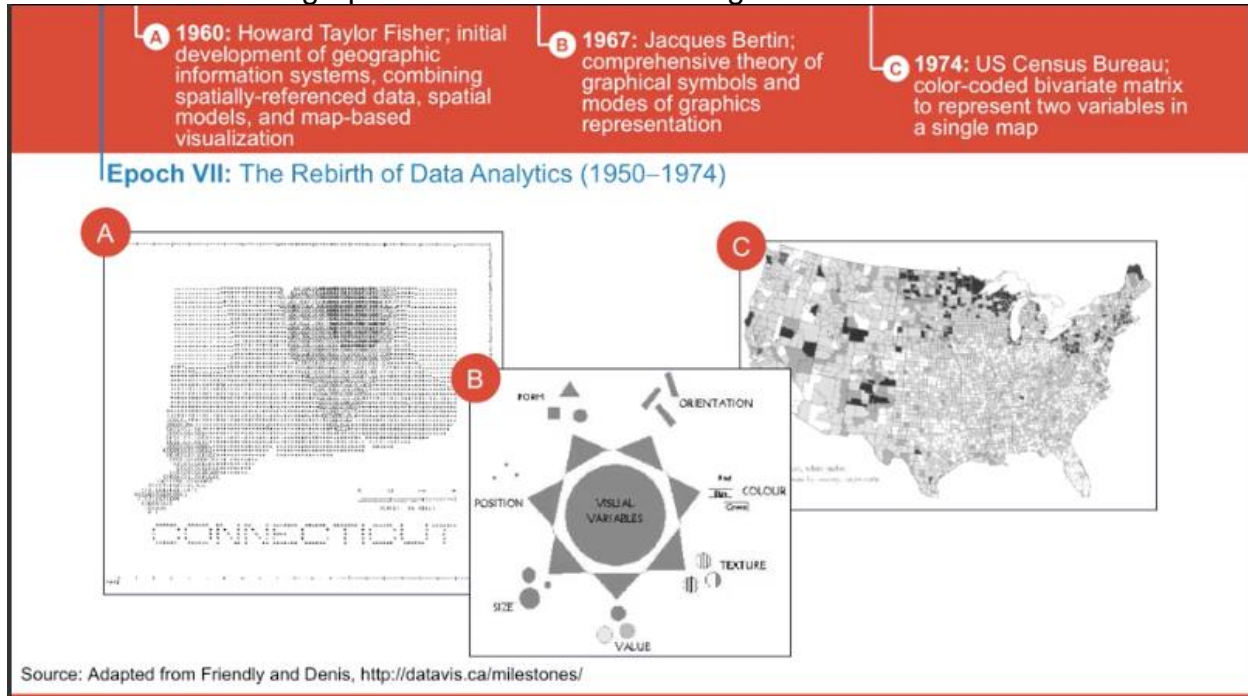
- A** **1944:** Harvard's Mark I, the first digital computer put in service. Known as the 'IBM Automatic Sequence Controlled Calculator' (ASCC), it was 50 feet long and weighed about 5 tons. The machine could perform all four arithmetic operations; moreover, it had special built-in programs to handle logarithms and trigonometric functions. It was slow, requiring 3 to 5 seconds for a multiplication, but it was fully automatic and could complete long computations without human intervention.

Epoch VI: The Modern Dark Age of Data Analytics (1900–1949)

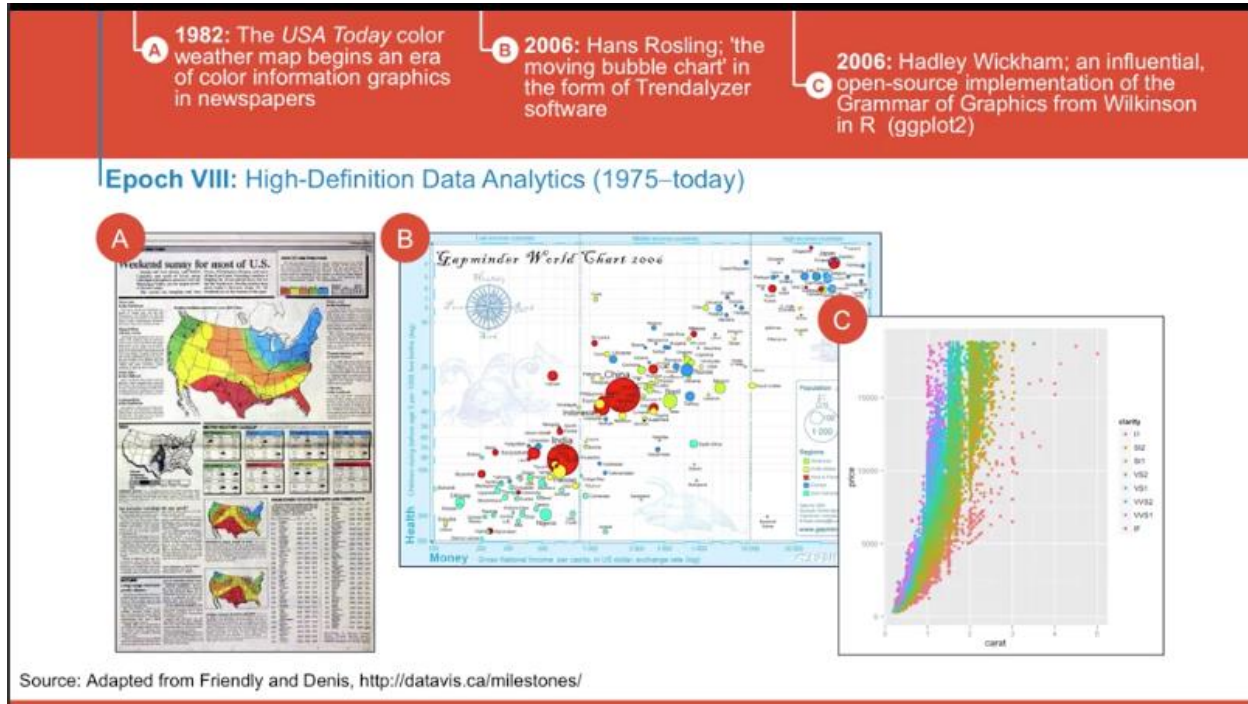


Source: Adapted from Friendly and Denis, <http://datavis.ca/milestones/>

Then we're also in this epic introduced to the computer, the Mark 1 at Harvard, the first computer ever created, the size of a living room which actually ironically held less capability than what we hold in our hand today with a mobile phone. But all the same, this starts to introduce new abilities for data visualization and data communication artists to create new graphics and we take advantage of that



as we move into the seventh epic, the rebirth of data analytics. Here computers began to play a bigger role. We can bring in more data, we can do it more sophisticatedly, we can add different elements that we didn't have before and as we hone these skills,



we move into the final epic, the epic that we still sit in today what we would call high-definition data analytics where visuals are now almost always created by computers or some application that takes in vast amounts of information, leverages all the things that we've learned in our past about data visualization, and the communication with data to create visuals that are effective, efficient, and tremendously powerful. So, those are the eight epics, the way we would trace the history of data visualization.

Studying the History of Dataviz Can Inspire and Inform



Eight distinct epochs trace the long, rich history of data visualization.

Dataviz has evolved from hand-painted maps into sophisticated computer-generated visuals.

Over time, the practice of data visualization has confronted issues of data scarcity, human politics, data misuse, and the scourge of clip art.

Dataviz has emerged as a discipline that offers the ultimate blend of art and science.



Understanding this is important because those eight distinct epics trace a very rich and a very long history that humans have had with communicating with data. It also shows how we've evolved from simple hand painted notes and maps into this place where we are today with tremendous power to generate maps. In fact, now the tools have actually even surpassed our creativity which the tools were always so limiting for humans as they sought to communicate with data. Now we can do just about anything we can think of. We have confronted along this long rich history things like data scarcity, the issue of human politics, data misuse, and of course as I mentioned the scourge of clip art to arrive at what really is today in my estimation the most pure, the most beautiful blend of art and science that we will find this idea of communicating through visuals and using data to create visualizations.

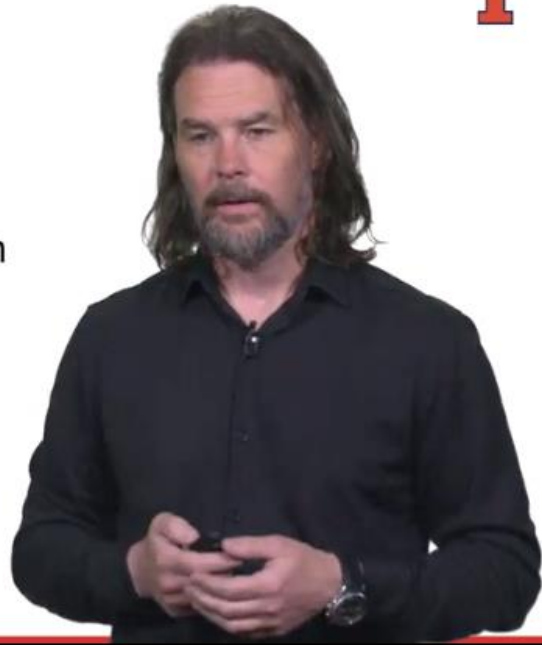


Lecture 1-2: Understanding Today's Dataviz Tools

Understanding Today's Dataviz Tools



Dataviz tools represent an evolution in analytics that can best be described along two important dimensions.



Okay. Understanding Today's Dataviz Tools, let's take a look at this rapidly changing, very dynamic market of data visualization tools. It really helps us here to grab hold of a couple elements of this market, so that we can make sense of it.

Dataviz Tools Can Be Defined by Two Dimensions



Ease of use

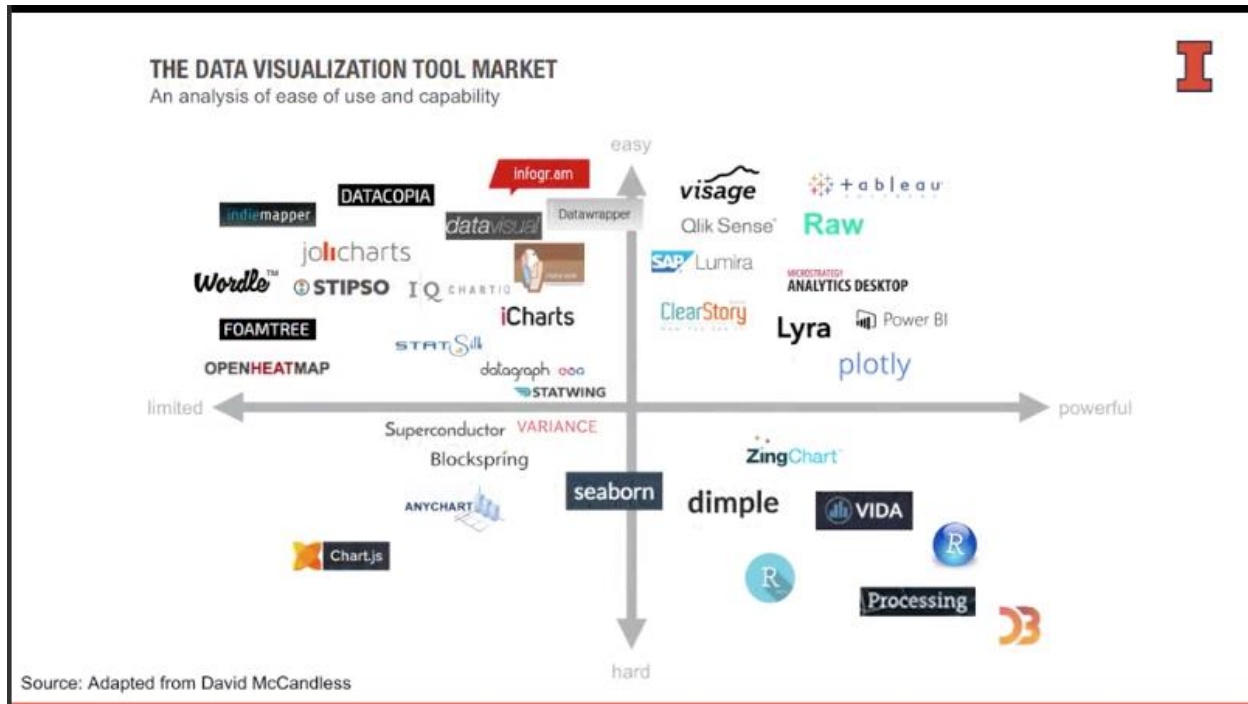
How quickly can one reasonably expect to learn the tool? How intuitive is it? What special skills does it require?

Capability

What is the breadth of the tool's functionality? How flexible is it? How much data can the tool effectively manage?



The two important dimensions that I use when I think about the data visualization tool market is evaluating them in terms of their ease of use. How easy is it for me to learn this tool? How quickly can I access, the things that I need to do with it? How efficiently can I use it, and the capability? Is it a tool that has very broad capability, it does a whole bunch of stuff, or is it something that is more focused, and I'll just use when I need a specific task done?



When you use those two dimensions, the data visualization tool market as it is lays out well across a two-by-two, and you find there's different quadrants of like tools. Right. So, let's take a look at each of these quadrants. When we look at those tools, that are really limited in their capability but yet very easy to use. What you'll find are a number of tools that can do one-off jobs or specific tasks very well for you and very rapidly. Ones that I love particularly in this quadrant is Wordle. When I need a word cloud created, when I've got a bunch of text that I want to be analyzed, I go to Wordle.net. It works perfectly, it doesn't do much more, it doesn't frankly do anything more than create word clouds, but it does it with such ease and with such expertise, that I use that tool anytime I need that specific task handled. The tools that you have up here, those are the things you should keep in your toolbox, things that have one-off capability that will serve you whenever you need them. Right. On the other side of that ledger, in the upper right quadrant, we have the tools that are both powerful and easy to use. Now, these tools, it makes sense to settle a one or a couple to become really the foundation of most of your data visualization work. Make this your go-to tool, it will do just about everything that you need and it's easy to use. So, the graphics that you produce will be done efficiently. Tableau is a perfect tool up in this quadrant, easy to use, well known, many people in the data visualization community accessing and using this and contributing to this tool. It really does become table stakes for most data visualization artists. Right. So, use the tool that's up in this quadrant as that thing that you go to, your go-to tool that you're



using 75% the time. Right. In the lower right quadrant, we find these tools that are tremendously powerful, but a little more awkward, a little more difficult to learn. The power that they have comes at the price of being easy to use. That doesn't mean that you should turn away from these. In fact, these tools can be very very important for you in your communication journey, as you are analyzing data, as you are finding visuals and patterns inside of the information you collected. One of the tools that I love the most, is R. It is an open source tool. It is free. There are thousands of people contributing to the R community, with help files and new capabilities and packages and scripts for this tool. If you had to learn just one tool in this lower right, my money is on R. There are several other tools, that can serve many purposes but I have never found one that I feel works as effectively as R. Now, I'm going to leave out one quadrant, and that is the lower left. Any tool that is really hard to use and doesn't have great capability, I would flush from your toolbox immediately. There isn't any reason to spend time with a tool like that.



Keep a number of data visualization tools in your “toolbox”

- Evaluate dataviz tools in their ability to offer efficiency: the ease with which you can access its capabilities
- Immediately discard hard-to-learn tools that do not offer exceptional capabilities
- Anchor on an easy-to-use tool with broad capabilities
- Utilize other tools when they are the most efficient tool for the job
- Constantly search for new dataviz tools in this dynamic, ever-changing market

So, that's the data visualization market at a high level. The specific tools that I've called out are probably frankly less important than us putting a mental framework around that market, and the way that I recommend we think about it is evaluating them on the ease of use and the access they provide to capability, either broad or more focused. Anything again that is hard to learn but doesn't have exceptional capability for you, shrug that tool side because you can quickly replace it in this growing dynamic market. The tool that you should use most frequently, the tool that's going to become the anchor for your data visualization, will be a tool you find in the upper right, easily use, great capability. The tools in the upper left of that quadrant, lower-right quadrant, you're going to want to use sporadically, just to fit your needs and to serve specific tasks that you have. Now, the trick for us in this market is to remain ever vigilant. There are new tools being added



every single day. This market is shifting and changing so rapidly as our experience with data does as well, that it's almost impossible to stay on top of it. In fact, this snapshot is really there for more illustrative purposes than a direct guidance on tools you use. So, as a data communication artist, challenge yourself to say must in this market. Be aware of the tools that are coming in. Have those tools that you know and use and love, but always be open to the new ones.



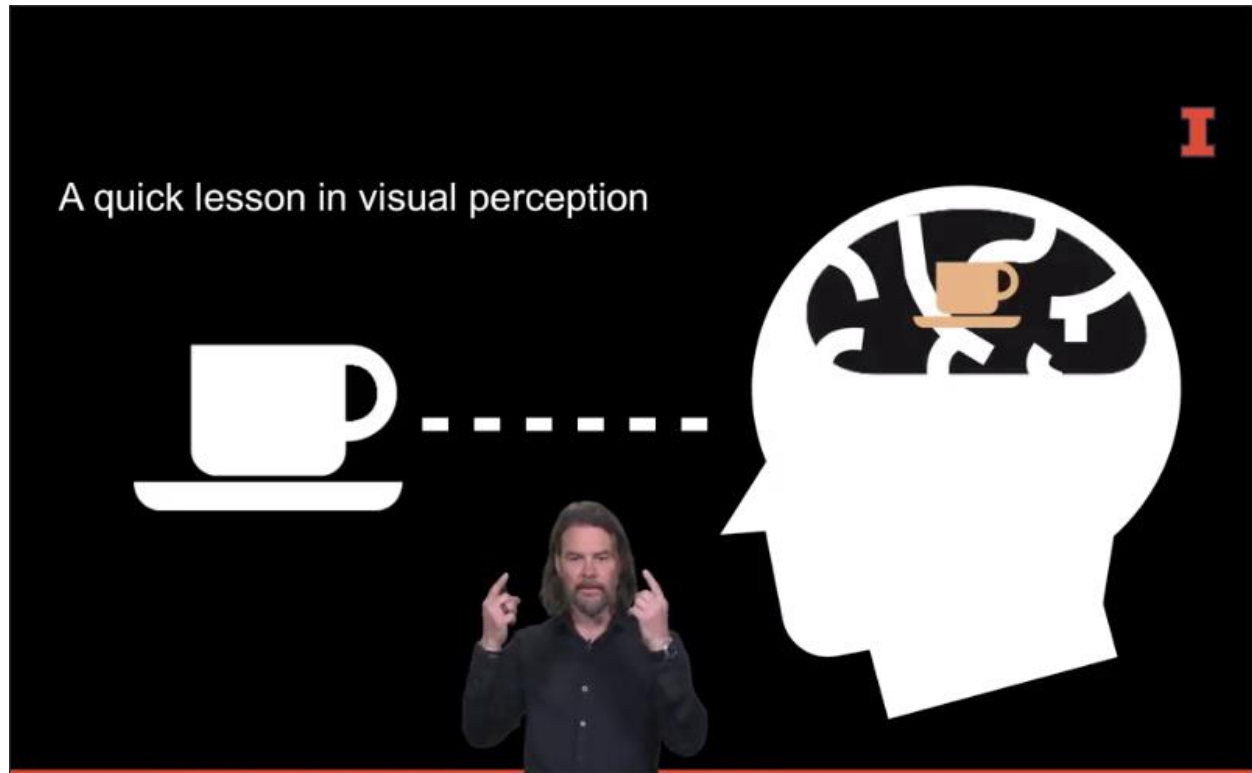
Lecture 1-3: Making Connections With Visuals

Making Connections with Visuals

Visualization is rooted in brain science and is one of the most efficient ways to communicate ideas, insights, and concepts.



Okay, Making Connections with Visuals. Let's take a bit of a step back, and let's actually understand how humans or audience process visuals, right? By understanding that, and understanding what is going on inside of their minds as they are seeing the visuals that we're putting forward helps us better position our messages, use the things that we will call pre-attentive attributes to connect with our audience more quickly. Having that understanding just makes us better at this art.



So, where I want to start is this idea of visual perception. So, our eyes are fantastic tools. But really, when we are seeing things, we're perceiving them in our mind. It's not our eye working, it's our mind working. So, when we are creating images that we are seeking to use as a communication visual to our audience, we should think in terms of how that visual will register in their minds. The mind is a very confusing and complicated place.



Source: Simons & Levin (1998)

 **CHICAGO HARRIS**
PUBLIC POLICY | THE UNIVERSITY OF CHICAGO



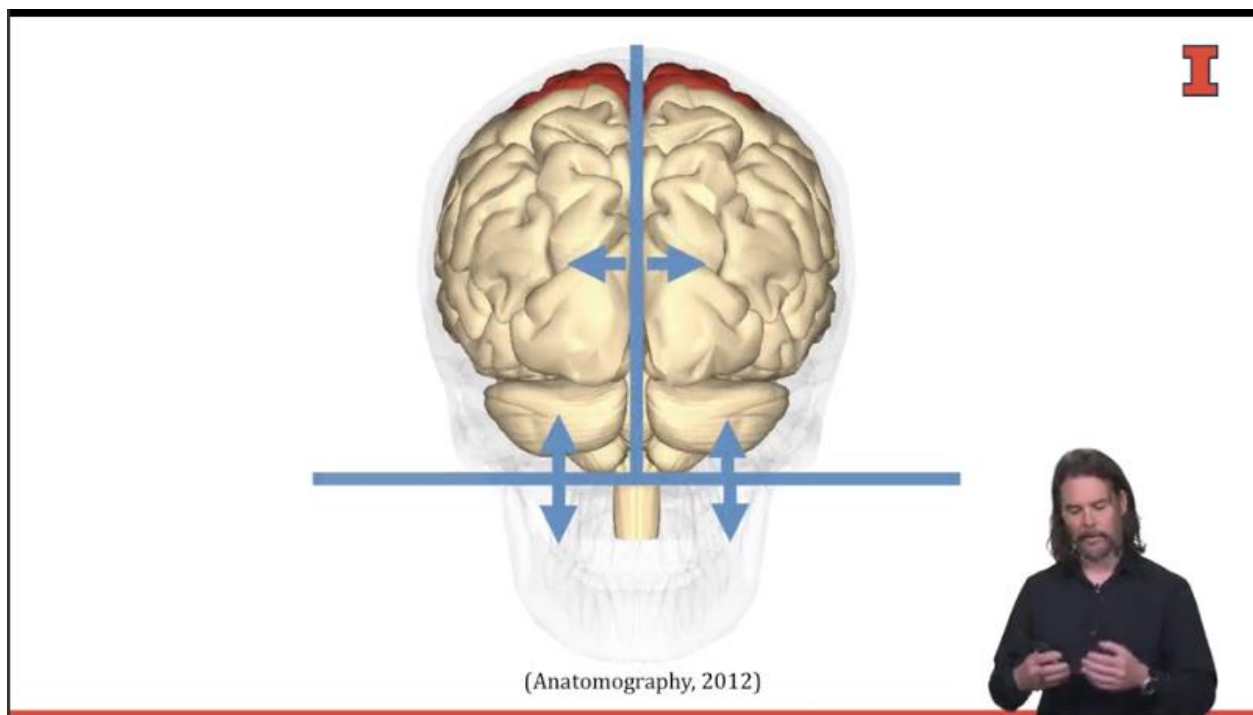
A couple of studies that illustrate this well, the famous Door study by Simons and Levin, where two researchers are experimenting, and one researcher goes and finds a subject in the park, gives them a map, asks them directions to a place, and the subject then is engaged in the map. You are, as a human, we are using our full frontal cortex when you are interpreting maps. This is the most complicated part of our brain. While this is going on and this subject is working to make sense of the map and where they are today and communicating all this out to the researcher, a group of men come through with a door and separate the subject from the researcher. The first researcher then jumps behind the door, and the second researcher pops out. Almost every time, the subject doesn't even notice that now they are talking to another person. They just continue on with the directions on where to go. This is because that prefrontal cortex, when it is fired, it can overload us, and we can miss the simplest of things, but that's not the only issue that we will be up against as we are trying to communicate with data.



A study by Moran Cerf is one that punctuates just how challenging this really can be. Moran had a researcher sit with the subjects, show the subject two pictures of people of the opposite sex. He said, "Which person do you find more attractive?" The subject then chose one, and then the researcher hands them the picture of the person they had chosen and said, "Why? Explain to me why you chose that person." Now, in this case, you are pulling on the lower, more reactive, the fight or flight part of the brain that makes that quick decision on who you find more attractive, not really even understanding why you've done it. It is all instinctual and emotional. But, when you are asked then to communicate and explain why you've done that, now this prefrontal cortex, again, is being fired and working hard to put words to something that has come from a part of our



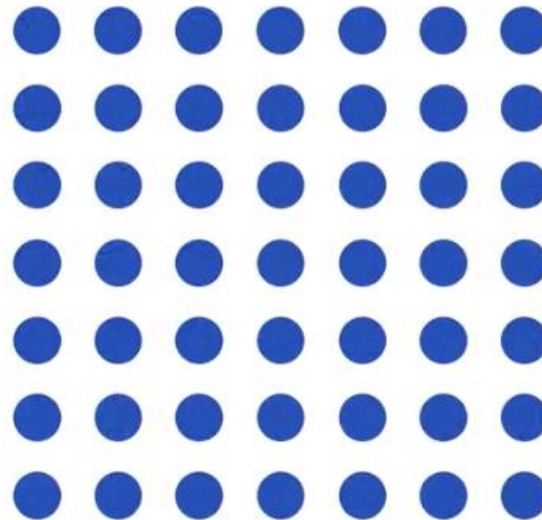
brain which has no capacity for language. It's a very difficult task for the subject, but that was actually not the purpose of this experiment at all. In fact, the researcher was not a researcher at all, but rather a magician. Every fourth card or so through a sleight of hand, they would hand the subject the card of the person they had not chosen. Still, the subject nearly every time continued on and came up with this overall explanation for why they had chosen the person that they actually hadn't chosen in the first place. If we, as an audience, don't understand the flow of a story or understand all the data that is being interpreted or brought to us, we will make up things to fill the gaps. This is the danger that we're up against as we are communicating with data, either confusing our subjects, our audience and getting them to think too much which could, in some of the extreme cases, lead them to make their own narratives up for the story that we are trying to own and tell.



The brain, as I said, is a complicated thing. We understand how the left lobe and the right lobe works, and we think about that quite a bit. But, we don't frequently enough think of the lower brain and the frontal brain, and how we, as data communication artists, need to work our way around that prefrontal cortex, leaving it silent, and calm, and hitting our audience with messages and visuals that strike at the lower brain, the very emotional part of their cognitive understanding.



Our brains are
programmed to
detect differences.

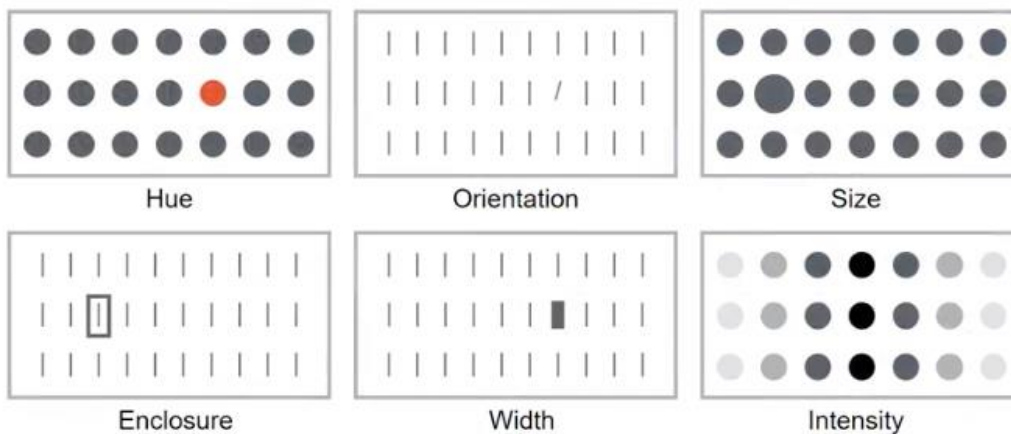


Source: Adapted from Werzowa, Johnson, Lanahan and Shillum

We, as humans, we are designed to pick up differences, differences in shapes, certainly differences in colors.



There are a host of preattentive attributes available

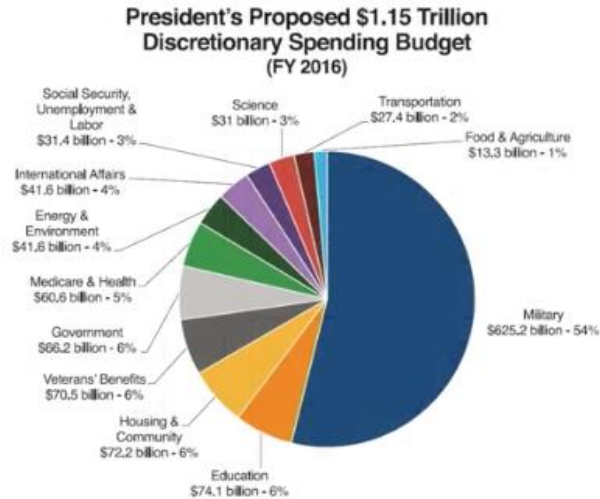


Source: Adapted from Isabel Meirelles, "Design For Information"

As we construct data visualizations, we have a whole host of different, what we would call pre-attentive attributes that we can use in our visuals to communicate points, to draw attention, to really get at that lower part of the brain. That seems very simple for us to use those visuals, those techniques.



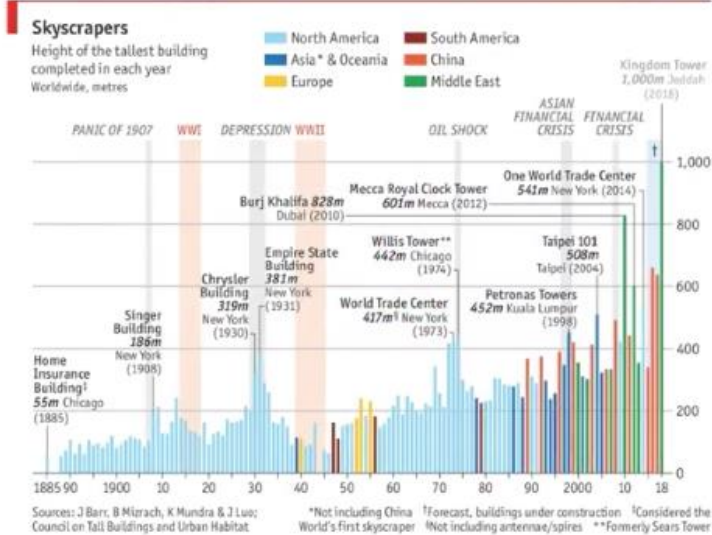
This seems so obvious.
But then why do we still
make graphics like this?



Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations"

But if that's the case, why do we still create visuals like this as Berinato pointed out in his work?

...Or this?



Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations"

Or this?



Or, for God's sake,
like this?



Or, for God's sake, this? If you're creating data visualizations like this, you're probably beyond help. But I would argue that the reason we put visualizations together like this with our understanding of shading, and enclosure and size, and all those things, is because we don't understand the fore rules. We don't understand the guidelines that need to go into creating a data visualization and communicating with data. The purpose of this course is to help you do that.

Use the Workings of Visual Perception to Your Advantage

Design dataviz with the message you want communicated in mind.

Devote yourself to keeping your audience's advanced brain (i.e., prefrontal cortex) as still and quiet as possible.

Build pre-attentive attributes into your visualization to efficiently connect your message to your audience.

Keep in mind that successful dataviz results from more than the use of color, shading, and other pre-attentive attributes.

So, understanding how the brain works will help us create better data visualization.



When we're doing this, we need to begin with the design in our message in mind. We need to keep that front and center. Whatever we're trying to communicate design around that message, we need to really think through the way our audience will interpret the things that we're putting in front of them, and the story that we're trying to communicate, and the visuals we use. Again, our objective there is to keep that prefrontal cortex as still and quiet as possible. We actually don't want our audience thinking much. We want to make a message that is so easy for them to understand and interpret, that they don't need to spend cycles and fire up their prefrontal cortex to fill in blanks of their own or divert attention away from the message that we have. Pre-attentive attributes can certainly help, and we'll learn those certainly in the course as we talk further. But the most important thing for us to understand is that, creating great data visualization and being an effective communicator with data is not a sense of just using color well, or shading, or any of these attributes. They are simply a piece of what is a very large and complicated approach to pulling messages from data, and then communicating them effectively in ways that land squarely with an audience. The purpose of this course will help us understand that full process, right? That full communication journey.

Lecture 1-4: Evaluating The Effectiveness Of Dataviz

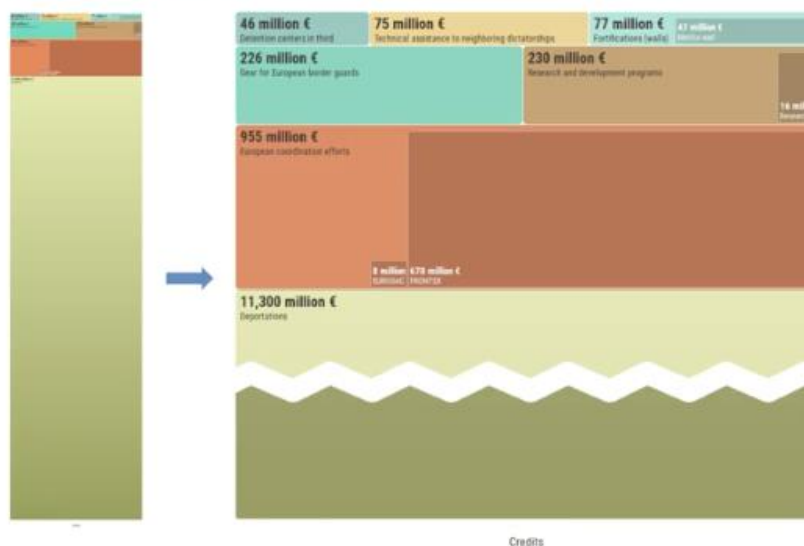
Evaluating the Effectiveness of Dataviz

Creating effective, impactful, and good-looking dataviz requires a successful blend of several key elements.



All right. Evaluating the Effectiveness of Dataviz. Let's take a look at what makes for good data visualization. We're going to do this by looking at some very good visuals, and then actually exploring a definition in a framework that we can use to help us answer that question like: What is good? Here are some of my favorite.

Billion Dollar-o-Gram



Source: Open Budgets EU's "Billion Dollar-o-Gram" generator



This is a beautiful visualization that David McCandless produced, that he calls the



Billion Dollar-o-Gram. So David basically, motivated by the fact that the term billion dollars gets tossed around a lot. He wanted to see what is the relative impact or size of various expenditures that are in that billion dollar range. He scraped data from a number of sources, put them into this graphic, and sized each of the spend according to their relative size, and so it revealed some things that are very interesting. We'll talk more about this visual later in the class.



Another one that I love was this exploration of the impact of the measles vaccination. So, what you're looking at is the occurrences of measles in every single state in the United States from about 1928 all the way into the 2000's, where the darker the color indicates the more occurrence of that disease. You can clearly see what happens in the mid-1960's when that measles vaccination is introduced, how the number of cases, occurrences of that disease falls dramatically, and this visualization does a wonderful job of doing that.



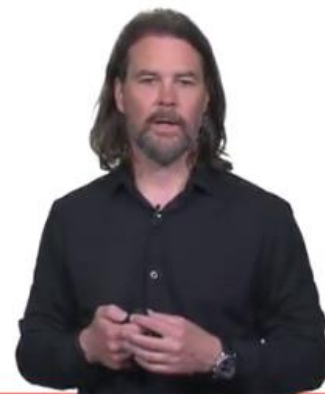
Not every visual needs to be computer-generated either. We're in the midst of this rage that's called photoviz, where visualization is done with real-world objects, and that can be done very effectively. This is one visual that I love and I think communicates very well. You're looking at an ice cube that represents the 1.83 million square miles of the polar icecap in 2008 and four years later the impact of the warming environment on the icecap. I think a beautiful visual, that does this well and really adheres to all the rules that we will talk about in just a moment. So, what does make these visuals so good? What are the things, the elements that they communicate so well, and that others which aren't nearly as effective miss out on?

McCandless Offers a Thorough Definition of Good Data Stories



Well, that idea has been band-aid around quite a bit, and I think the framework that I feel best communicates those ideas of what makes for good data visualization is really captured here. There are four elements, four elements that when all four come together, you have the ability to communicate effectively, you have created a successful data visualization.

McCandless Offers a Thorough Definition of Good Data Stories



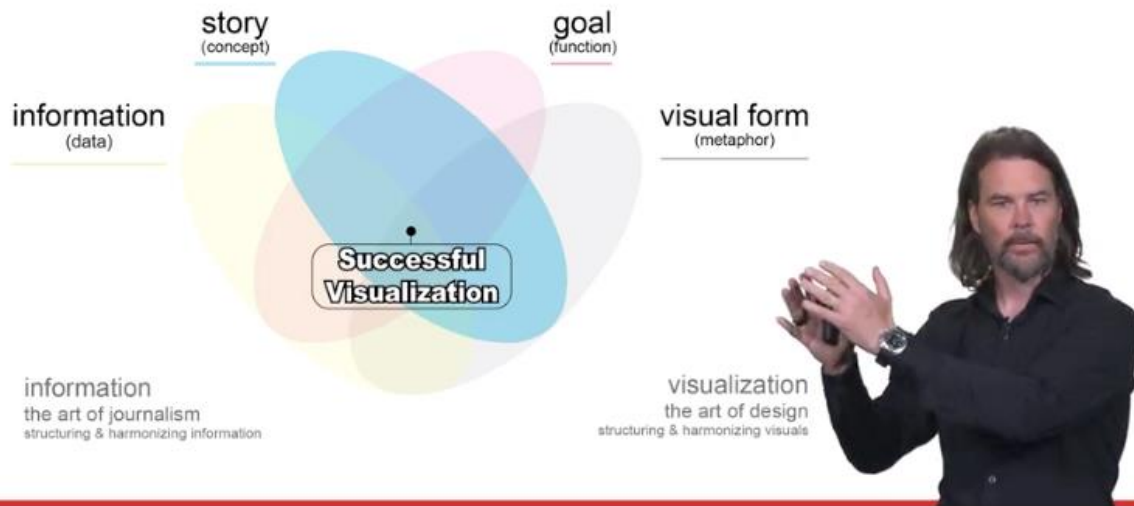


Those four elements are: Information. This is data. This is the data that we're collecting, data that is accurate and deep and robust. The better that source of data is, the better our data visualization will be. If we don't use data as a basis here, we are no longer creating a data visualization, we're creating something completely different. But data is a clear important element of communicating with data.



The next is a goal. This is the functional purpose of the data we've collected. This is an objective that we are working toward, that we'll focus our story, focus the data we collect, and really give us a pinpoint purpose. Without that goal, our data becomes somewhat meaningless and rudderless.

McCandless Offers a Thorough Definition of Good Data Stories



Even if our story is great and we've got great data, if there's nothing that we're trying to obtain or achieve with it, then we don't really have a successful data visualization any longer, we've got something completely different. Knitting these two together is the story. The story is what takes data and moves that toward a purpose. Directs toward that objective. It is the connective tissue between data and an objective, and this is where we are using a narrative to guide our audience. Without that story, we are just throwing facts. Even if we have a solid objective, we're throwing nothing more than figures at an audience, where we have very little opportunity to connect. But even having those three things are not enough.

McCandless Offers a Thorough Definition of Good Data Stories



This is where the idea of visual form comes in. The data we've used, the objective, goal that we have, the story that we've used to knit those things together, those have to take a visual form to effectively communicate with our audience. Otherwise, you are left with a report of text and bullet points that runs the risk of boring our audience, distracting our audience. Visuals become, as we've seen, the most powerful way for us to communicate with an audience. In this case, we are using this visual to achieve a goal that is built out of data and stitch together with a narrative and a story.

McCandless's Framework Brings Quality to All Dataviz Elements



Provides a clear answer to the elusive question: "What makes a data visualization good?"

Goes beyond simply considering the image to evaluate the entire creative process.

Compels the author to pay attention to each individual component of a data visualization.

Ensures all the elements come together to form a single, harmonious dataviz.




So, this framework really does bring all of those things together. It's important for a number of reasons. One is because it does give us this clear answer and understanding to the question; what makes a good data visualization or what makes a visualization good? Which can be somewhat elusive. I think this framework gives us a good handle on that answer. It goes beyond just evaluating the pretty picture. It goes into all the elements of the creative process, which is important because it forces you as an author to then make sure that each element is solid and good. Because without each one, the entire visualization and communication falls apart. Finally, this framework communicates well the idea that these things all form one harmonious message and they all come together as pieces that fit into your data visualization. So, for that reason, I think this framework that has been explored is the right way for us to think about what makes for good data visualization and how we will evaluate the process of communicating with data going forward in this class.

In This Module

Module 1: Pictures You See with Your Brain

Key Concepts

- Studying the history of dataviz
- Understanding today's dataviz tools
- Making connections with visuals
- Evaluating the effectiveness of dataviz



So, that is Module One. In this module, we talked about a number of things. We traced the history of data visualization to give us that long view into how human's relationship with data and communication. We looked at the data visualization tools of today and came up with a good framework on how we can evaluate those tools. We talked about the way our audience will interpret the visuals that we put in front of them, and so we now have an understanding of the brain workings and can use those to our advantage as we seek to communicate with data. Then, finally in this last lesson, we saw a framework that will allow us to evaluate the effectiveness of dataviz, effectively giving us that definition of what is a good data visualization, what are the things that we need to think about in terms of elements when we are putting together a dataviz and seeking to communicate with data.