**Data Storytelling**

* Who is your audience?
* What questions are you trying to answer?
* Why should your audience care?
* What are your major insights?
* What impact do you want to have by sharing your finding?

**Storytelling through Exploratory Data Analysis**

[This wonderful Harvard lecture](https://youtu.be/IVyN9kAJ0ak) explores how data storytelling and exploratory data analysis are intrinsically linked. Start from 1:97 and, as you watch the video, pay close attention to the five principles of effective visualizations (33:00) and the flowchart you can use to pick the best method for different situations (43:00). Please do the design exercise with the attending students at (51:35).

[Github:](https://github.com/cs109/2015/blob/master/Lectures/03-EDA.pdf)

* Sometimes you have to look at your data for a long time
* Look for the “that’s funny/weird….” things you notice. There may be a new insight to discover there, e.g. something misclassified
* John Tukey (the “father” of EDA): “The greatest value of a picture is when it forces us to notice what we never expected to see”

Visualization: To convey information through graphical representations of data

**Communicate (Explanatory)**

* + Present data and ideas
  + Explain and inform
  + Provide evidence and support
  + Influence and persuade

**Analyze (Exploratory)**

* + Explore the data
  + Assess a situation
  + Determine how to proceed
  + Decide what to do

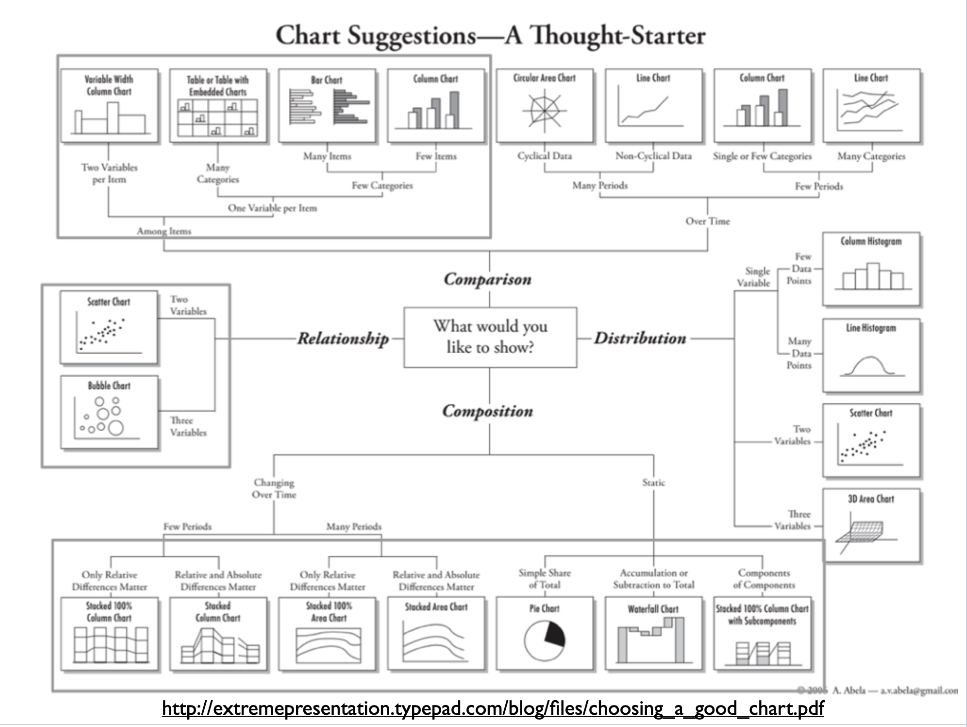
Effective Visualizations:

1. Have graphical integrity
2. Keep it simple
3. Use the right display
4. Use color strategically, very carefully
5. Tell a story with data

Ineffective Visualizations: viz.wtf

Maximize Data-Ink Ratio: Data Ink/Total Ink used in graphic

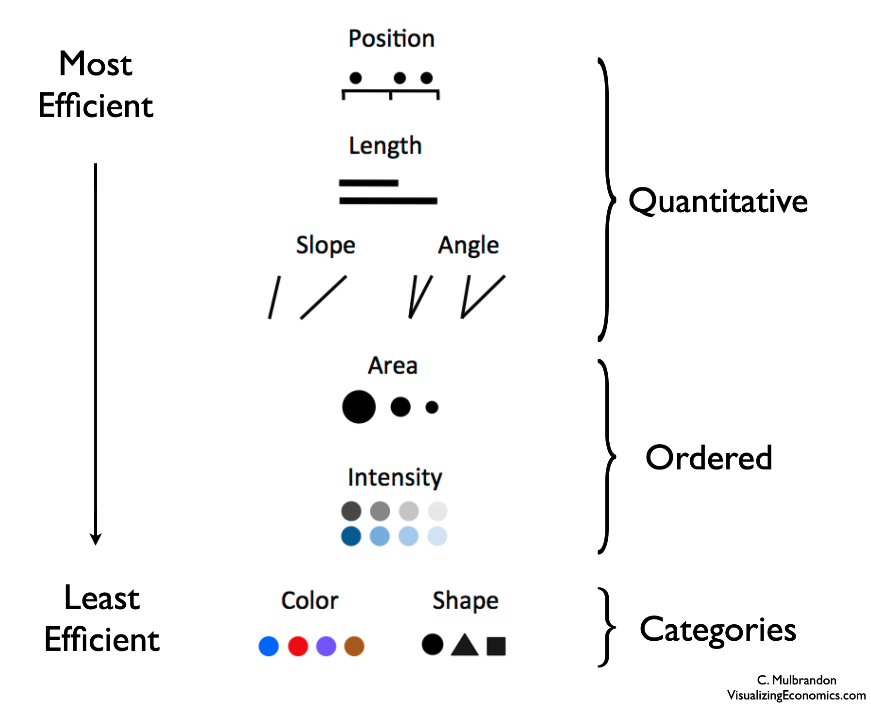
3D is generally bad: 1. Wastes non-data pixels, 2. Distorts the data

Avoid Chartjunk: anything you can remove from your chart that will not change the data display, avoid visual elements that distract form the message

\*Difference bar chart

\*Slope Graph

Most Effective Visualization Features for Quantitative Interpretations by Humans:



Use Color Strategically: W

* Whenever possible, don’t use color
* 5 +- 2 is the max humans can distinguish without too much cognitive effort.
* Don’t use more than 5-8 colors at once
* Can vary luminance and saturation for ordinal data
* DON’T use rainbow: red and yellow visually dominate perception
* DO use greyscale, and then can use color to highlight what you want to be noticed
* Use ColorBrewer tool to select appropriate color scheme. These can all be used in seaborn: https://colorbrewer2.org/#type=sequential&scheme=BuGn&n=3

[Show me the data -- becoming an expert in yourself: Talithia Williams at TEDxClaremontColleges](https://www.youtube.com/watch?v=TDCYJ3_gx2w)

* Tells a story about being 42 weeks pregnant, getting a check, being told her baby was under stress and she needed to be induced
* As a statistician, she basically grills her doctor for the data about this choice, and when she finally gets down to the hard data, she is unconvinced by the data
* She had been collecting data about her cycle for 6 years, lol, which made her confident to be ready to leave.
* Temp, blood pressure, weight
* When you collect your own health data you can take ownership over your health,

Storytelling and Effective Communication

[Harvard Lecture](https://youtu.be/c3ztpGLx_BM), [slides](https://github.com/cs109/2015/blob/master/Lectures/06-StoryTelling.pdf)

This Harvard lecture gives fresh examples of the importance of framing when telling a story with data. As you watch this lecture, be sure to take note of the **key considerations** to make before building a story:

* What questions are you answering?
* Who is your audience?
* why should they care about the information you're sharing?
* What are your major insights and surprises?
* What change do you want to bring about?

Also note the **key steps** to crafting any story itself:

* Introduce interesting characters
* Put them in a predicament
* Resolve the predicament
* Leave room for sequels

The lecturer also gives an incredibly useful piece of advice — useful regardless of the type of audience you have: make the audience aware there’s something important they didn’t know they didn’t know.

**2 fundamental questions to answer in your presentations**:

* 1. What is the goal?
     + Explain future data?
     + Explain or understand a phenomenon?
     + Test a hypothesis?
     + Compare two groups?
     + Dimension reduction?
     + Build a good recommendation system?
     + Decide on a course of action or a policy?
  2. Who Cares?

**IMAC:**

**I:** Inferential Goal (Scientific question of interest) **M:** model (all models are wrong, some are useful)

**A:** algorithms **C:** conclusions and checking

The **C** is crucial: what did we learn? Was the model useful, and how well does it fit? How do we know whether the method is working? Do we need to iterate and improve the model? What are the limitations and future directions?

**Some Key Principles:**

* Remember the Golden Rule (do unto others…) - Know your audience -Tell a story
* Choose and use notation carefully - Read great writers
* Create good sense of direction (with the help of *signposts*), with clear flow of logic

Simple, but memorable examples and stories

**Tell a story!:** any story has a beginning, a middle and an end

* Introduce interesting characters -Put them in a predicament -Resolve the predicament
* But leave room for sequels! (Limitations and future work)

Stories are the most powerful delivery tool for information, more powerful and enduring than any other art form: think stories around a campfire, stories as a way of connection and transmitting information

**Know your audience!:**

* What do they know? - What motivates them? - What experiences do you share?/What are common goals? - What insights can you give them? What tools and “magical gifts”?
  + A presentation should be for everyone, like an advertisement for your paper geared toward technical experts

**Don’t Make Your Audience Think!**

* Your audience does not want to spend cognitive effort on things you know and can just show them
* Lead them through the major steps of your story
* Point out interesting facts and insights using captions and annotations
* **Don’t “bury the lead”**: make the main message as explicit as possible, a well-placed effective visualization that communicates the main message is a good idea for this
* **Make it explicit what the question is you’re answering**

**DO** make your audience aware there is something they didn’t know they didn’t know: use surprise to grab the audience’s attention, curiosity happens when we feel a gap in our knowledge, “you might think you know this, but here’s a new angle on it”

**Framing:** Tell the audience: “Here is the right way to think about the problem I was trying to solve”

* Catch the audience’s attention and frame the story using captions and annotations
* If done well, your insights will seem obvious given this framing. And that’s a good thing!

1. Basic types of framing:
2. **Exploratory: Neutral**, Audience: board meeting, lecture, team meeting
3. **Explanatory: Opinionated**, Audience: Ted Talk, maybe a team meeting

**Successful Data Stories:**

* Target the audience
* Engage and are memorable
* Answer concise questions
* Are carefully designed
* Move us to want to change the world

**Further Reading:**

* The Functional Art, Alberto Cairo
* Presentationzen, Garr Reynolds
* Made to Stick, Chip and Dan Heath
* Resonate, Nancy Duarte

**How to Avoid Death by PowerPoint**

[David JP Phillips, TEDxStockholmSalon](https://youtu.be/Iwpi1Lm6dFo)

1. design principles:
2. One message per slide
   1. Make it simple for human beings, control what your audience will be focusing on
3. Working Memory
   1. Redundancy effect: if you have text on your slides and you’re speaking at the same time, people will remember *nothing*
   2. Use the notes space for your sentences, use the slide for your presentation material: short, sweet bits of text and an image. That is what enhances your message
4. Size
   1. Every time you open your eyes for the rest of your life, you will focus on 4 things: moving objects, signaling colors, contrast-rich objects, big objects.
   2. Powerpoint is defaulted to have big headlines and smaller content—this is backward! Use small headlines to avoid people focusing on that.
   3. Make the focus of the slide the biggest thing on it
5. Contrast:
   1. Controls your focus, use contrast to show audience exactly what to focus on
   2. Default white backgrounds make the slide the most contrast-rich thing in the room; **use dark/black background to pull focus toward you**, the speaker, and keep the slides in their proper place: a visual aid to support the main ideas coming from you, the speaker.
   3. You are the presentation, the slides are the visual aid
6. Objects
   1. How many objects are on the slide?
   2. Don’t put page numbers, lol
   3. The cognitive process of counting objects takes 500% longer than seeing/subitizing
   4. The magic number is 6 objects on a slide, if you have more than that be aware that your audience has to invest 500% more processing power to understand your slide…and they won’t
   5. Want audience to go…ahhhh…in a sigh of relaxation when they see your slides, not ugh….
   6. Number of slides is not the problem, the number of objects on each slide is the problem