- Best practices summary
  - **Best practice #1:** Differentiate data with color or shading rather than with shape.
  - **Best practice #2:** The rule of thumb is that for >4 series, you should not rely on color alone for differentiating series. Use direct labeling or some other technique instead.
  - **Best practice #3:** Use primarily neutral or natural colors for the main colors when presenting data. I think this avoids visual fatigue from trying to figure out which brightly colored thing I'm supposed to pay attention to.
  - **Best practice #4:** Don't use a color scheme that has an implied meaning, unless that meaning matches the data.
  - **Best practice #5:** Even for encoding data, use shades of neutral colors rather than Excel defaults or making up a color scheme. It's hard to make up a color scheme that is color blind safe but not too bright.
    - http://colorbrewer2.org is a good resource for getting color blind safe schemes with a variety of colors if you need one
  - **Best practice #6:** Remove unnecessary grid and axis lines from graphs. Use gray for remaining grid lines.
  - **Best practice #7:** Remove grid lines from tables, and use gray for headings.
  - **Best practice #8:** No one should notice the font you use. When I doubt, use Helvetica — while some think it is over-used, it is universally respected by typography designers and is very readable. Be careful with non-standard fonts for presentations.
  - **Best practice #9:**
    - For slides presented in:
      - Ambient light: light background, dark text
      - Dark room: dark background, light text
        - Avoid non-natural/bright colors for the background
- Edward Tufte
  - He's probably the most famous contemporary data presentation writer/presenter
  - He's written some great books
  - There is a lot of good practical advice sprinkled through his books, but he is sometimes opaque about the reasoning behind his opinions.

- Thinking about how we think
  - I've read some other books on data visualizations and presentation, and I really found it helpful to think about how our brains perceive visual information, and how this can be used to optimize the way I present data.
  - What I'm going to do in this presentation is talk briefly about how our brains work in a grossly oversimplified way that is probably inaccurate, but seems to me to be very helpful for improving how data are presented.
  - I'll then go through some practical application of this for presenting data.
- Avoiding tiger attacks
  - Our eyes are drawn to
    - Moving objects
    - Uncommon shapes
    - Bright colors
  - I assume this was an adaptation to help our ancestors avoid getting eaten by tigers.
  - This was one of the most helpful things I read about presenting data because there are a ton of implications from this for presenting data, in terms of
    - How to highlight important information, and
    - How to avoid distracting or overwhelming your audience
  - Pretty much everything else in this presentation references back to this slide.
- Shapes vs. colors
  - Good data presentation often requires using visual techniques to distinguish elements of a graph or table.
    - We know that moving objects are annoying in presentations and impossible on paper
    - So we're left with shapes and colors. Which is most effective?
  - However, we're not as good at picking out shapes as we are at picking out differences in color.
  - *The Functional Art* has a great example of this that I've recreated here. I think it makes the previous point self-evident.
  - One practical application of this is with line graphs: it's much easier to tell colored lines apart than lines versus different shaped markers.

- **Best practice #1:** Differentiate data with color or shading rather than with shape.
- However, there is a limit to this. We've all seen a graph with like 50 series in the legend, which is no good. The viewer has to keep referring back to the legend.
  - This has to do with limits in our working memory.
  - **Best practice #2:** The rule of thumb is that for >4 series, you should not rely on color alone for differentiating series. Use direct labeling or some other technique instead.

- Choosing colors
  - I mentioned before that our eyes are drawn to bright colors, or in other words, colors that we don't expect to see in nature.
  - Aside: before I started started learning about this, I would have probably made this slide like this (with much brighter colors).
  - But if you think about it, it makes sense: our eyes are drawn to the potentially poisonous tree frog or, say, flowers, but not the color of the sidewalk.
  - Tufte says that we should use bright, non-natural colors sparingly. Bright colors should be reserved for highlighting important data.
  - If you use bright colors for everything, it's hard for the audience to know what to pay attention to. Bright colors have an implied meaning: pay attention to me!
  - Here's an example of a weather graph based on a graph in one of Tufte's books.
  - Here's the same graph but with a bright color rather than a neutral color.
  - I find the 2nd graph much more fatiguing to look at.
  - **Best practice #3:** Use primarily neutral or natural colors for the main colors when presenting data. I think this avoids visual fatigue from trying to figure out which brightly colored thing I'm supposed to pay attention to.
    - But your graphics won't be boring because you can still use bright colors to point out important data — especially in slides. (So if you don't have any bright colors, you don't have any important data.)

- Choosing colors: when summarizing data
  - There's a sub-class of the color scheme problem where the colors you use encode a summary of your data.
  - This already came up with the Seinfeld viewers graph. The color wasn't helpful in that graph, but in some cases it is. For example, a bar chart with just 3 series.

- **Best practice #4:** Don't use a color scheme that has an implied meaning, unless that meaning matches the data. In Example 1, the colors imply to me that Site 1 is bad and Site 3 is good. This has nothing to do with the data though.
- Ok, so how do you pick a good color scheme for this bar graph?
- Example 2: I picked some colors that looked ok together to me. They are bright but not too bright.
  - The problem is that they are not color blind safe.
    - Why should you care? There's a good chance someone looking at your data will be colorblind.
- Example 3: Color blind safe colors are too bright
- Example 4: Excel default colors are ugly, but they are color blind safe.
- Example 5: Tufte uses gray in a lot of the graphs in his book. This lets you differentiate the series but is kind of boring.
- Example 6: Shades of blue are "natural" colors, and for 3 series they should be different enough to allow you to differentiate on any screen. Four series might be pushing it.
  - They are color blind safe too.
  - And they work well with highlighting important data with a bright color.
- **Best practice #5:** Even for encoding data, use shades of neutral colors rather than Excel defaults or making up a color scheme. It's hard to make up a color scheme that is color blind safe but not too bright.
  - http://colorbrewer2.org is a good resource for getting color blind safe schemes with a variety of colors if you need one

- Eliminating visual noise
  - For the final graph in the last set of graphs, you might have noticed I did something different with the borders.
  - Here's the default in my graphing software (lots of borders)
  - And again, here's what I did
  - This goes back to the lion in the grass. Lots of extra dark lines make it harder to see the data.
  - Here's another series of graphs that shows the whole spectrum from "aggressive grid" to Tufte's anti-grid

- **Best practice #6:** Remove unnecessary grid and axis lines from graphs. Use gray for remaining grid lines.
- This is another thing that I was doing wrong before I started reading about this.
- The same goes for tables. I've experimented with this and I actually like no grid lines for some tables. I also sometimes like using gray for headings to direct attention to the data.
- This is another thing I was doing wrong before I started looking into this: I always did my tables with black horizontal lines, which now I think is wrong.
- **Best practice #7:** Remove grid lines from tables, and use gray for headings.

▼ More on tables

- I also realized I was doing some other things wrong with tables.
- ▼ I think the following are **good guidelines for tables**:
  - Words should be left-aligned
  - Numbers should be right-aligned and rounded so that the decimals align
  - Heading alignment should match the column (left for words, right for numbers)
  - ▼ You can do some fancy stuff in Excel to get *p* values to align properly, but you need:
    - A fixed-width font (like Courier New)
    - This formula: `=TEXT(ROUND(A1, IF(A1<0.01, 3, 2)),"0.00?")`

▼ Nagging questions

- ▼ There are two things I tried to find definitive answers on and could not
  - ▼ Fonts
    - Serif vs. sans serif? I couldn't find anything particularly scientific about this. I think as long as the font is readable and not obnoxious, either is fine.
    - What is readable? Anything that looks normal and isn't Comic Sans.
    - What is obnoxious? Depends on who you ask.
    - ▼ Note that if you use a non-standard font, you will need to either present from your own computer or from a PDF. If you bring a PPT file to a computer without your fancy font, your presentation will look bad.
      - I always present from a PDF for this reason and to avoid issues with compatibility between different version of PowerPoint.

- **Best practice #8:** No one should notice the font you use. When I doubt, use Helvetica — while some think it is over-used, it is universally respected by typography designers and is very readable. Be careful with non-standard fonts for presentations.

- Background colors for slides
  - Conventional wisdom for PowerPoint is to use either
    - White background, black text
    - Blue background, yellow text
  - This maximizes contrast
  - I think this was more of a problem in the past when projectors weren't as powerful as they are now
  - My problem with the blue/yellow color combination is that the blue is too bright. Now that I know that my audience's eyes will be drawn to bright colors, I want to use that as a tool in my slides, and I can't if my background is the brightest color on the screen.
  - Of course, this all depends on the room and the projector. This is why you should test your slides in the same room with the same projector you're going to present with.
    - You should also set your color scheme in your slide masters, so if you need to change it you can do it without having to adjust every single slide.
  - **Best practice #9:**
    - Ambient light: light background, dark text
    - Dark room: dark background, light text
      - Avoid non-natural/bright colors for the background

- bit.ly/datatips
  - I'm working on a website on data presentation best practices, so these slides and my notes, as well as a bunch of other stuff are up on there.
  - If you want to look at the slides again or read more, this is a good place to start.