The Good, the Bad, and the Ugly
Visualization Recitation

15.071x – The Analytics Edge
Great Power, Great Responsibility

• There are many ways to visualize the same data.

• You have just seen how to make quite attractive visualizations with ggplot2, which has good default settings, but judgement is still required, e.g. do I vary the size, or do I vary the color?

• Excel, etc. can also be used to make perfectly acceptable visualizations – or terrible ones.
What is the difference?

• **Good visualizations…**

  Clearly and accurately convey the key messages in the data

• **Bad visualizations…**

  Obfuscate the data
  (either through ignorance, or malice!)
What does this mean?

- Visualizations can be used by an analyst for their own consumption, to gain insights.

- Visualizations can also be used to provide information to a decision maker, and/or to convince someone.

- Bad visualizations hide patterns that could give insight, or mislead decision makers.
Today

• We will look at some examples of visualizations taken from a variety of sources.

• We’ll discuss what is good and bad about them

• We will switch in to R to build better versions ourselves.

• Think for yourself: ultimately subjective!
Bad Visualizations?

Source: http://www.forbes.com/sites/tomiogerontomio/2012/02/02/does-ios-crash-more-than-android-a-data-dive/
Bad Visualizations?

Source: International Shark Attack File report
Bad Visualization?

- Not all points can be labeled, so data is lost.
- Colors are meaningless, are close enough to be confusing, but are still needed to make it at all readable.
- 3D adds nothing, visible volume is larger than true share.
Better Visualization?

- All data is visible!
- Don’t lose small regions.
- Can easily compare relative sizes
- Something to consider is that, for some people and applications, being not as “visually exciting” is a negative.
On a World Map?

- Possible with this data, but still a bit tedious to create because we need to determine which countries lie in which region.
- Shading all countries in region the same color is misleading – countries in, e.g. Latin America, will send students at different rates.
- We have access to per country data – we will plot it on a world map and see if it is effective.
Bad Scales

Source: BBC
Bad Scales

Source: Fox News
Bad Scales

- “Caucasian” bar is truncated – would be as wide as this slide!
- Every bar has its own scale – compare “Native American” to “African American”.
- Only thing useful is the numbers.
- Minor: mixed precision, unclear rounding applied

Two Rights Make A Wrong


- Different units suggest (non-existent) crossover in 1995
- Transformation shows true moments of change
Family Matters

Households by Type, 1970 to 2012: CPS
(In percent)

Family Matters

- If we are interested in shares within a year, it’s good.
- If we want to see rates of change, it is pretty much unusable!
- If we want to compare year-to-year, it’s possible though imperfect.
- Numbers are relative – absolute numbers may reveal, e.g., married couples without children is constant across years.