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!)
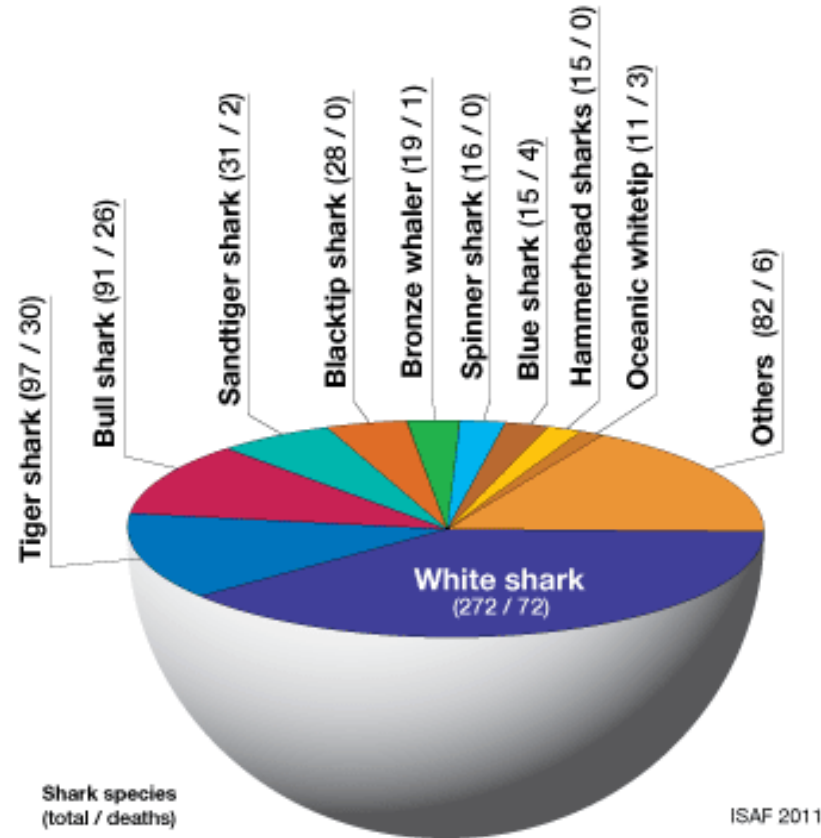
What does this mean?

- 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.

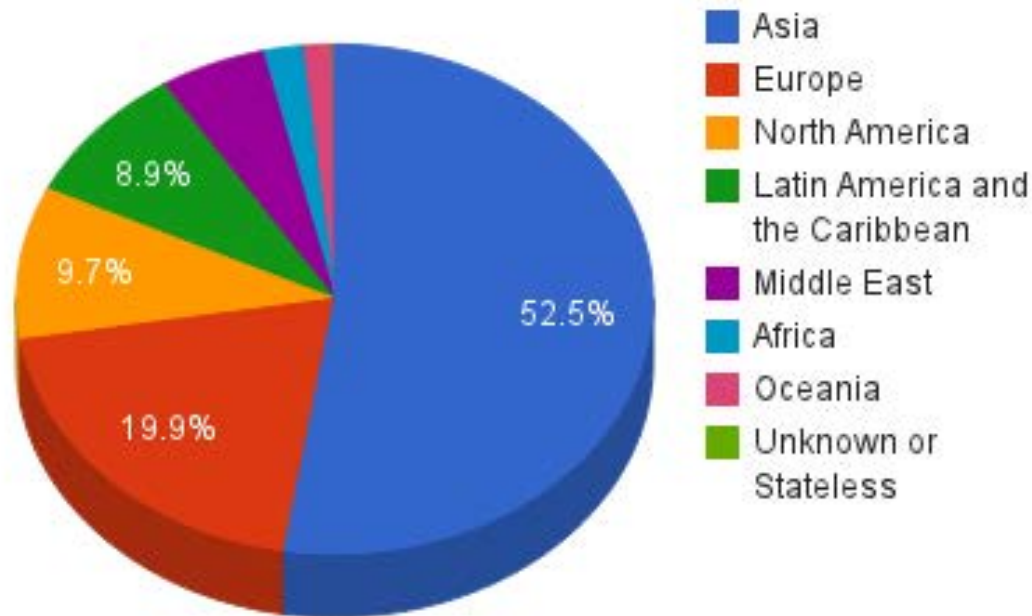- Think for yourself: ultimately subjective!

Bad Visualizations?

Crashes by OS Version Normalized (12/1 - 12/15)

Android Other : 0.54 %
Android 2.2.1 : 1.01 %
Android 3.2.1 : 0.20 %
Android 2.3.7 : 1.04 %
Android 2.3.6 : 0.37 %
Android 2.3.5 : 0.97 %
Android 2.3.4 : 3.65 %
Android 2.3.3 : 3.86 %
Android 2.3.2 : 0.10 %
Android 2.3.1 : 0.05 %
Android 4.0.1 : 1.04 %
Android 2.2 : 0.77 %
Android 3.1 : 0.09 %
Android 2.2.3 : 0.05 %
Android 3.2 : 0.05 %
Android 2.2.2 : 1.30 %
Android 1.6 : 0.40 %
iOS 4.2.8 : 0.29 %
iOS 4.3.1 : 1.38 %
iOS 4.3.3 : 10.66 %
iOS 4.3.2 : 1.36 %
iOS 4.3.5 : 8.96 %
iOS 4.3.4 : 0.52 %

iOS 3.1.3 : 0.00 %
iOS 4.2.10 : 12.64 %
iOS 3.2 : 0.00 %
iOS 5.0 : 4.80 %
iOS 5.1 : 0.02 %
iOS 4.0.2 : 0.23 %
iOS 4.1 : 8.24 %
iOS 4.0 : 0.25 %
iOS 4.3 : 0.96 %
iOS 4.2 : 0.03 %
iOS 4.2.1 : 4.85 %
iOS 4.2.7 : 0.03 %
iOS 4.2.6 : 0.25 %
iOS 4.0.1 : 0.25 %
iOS 3.2.2 : 0.00 %
iOS 5.0.1 : 28.64 %

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

Bad Visualizations?

Source: International Shark Attack File report

Bad Visualization?

MIT International Graduate Students

Legend:
- Asia — 52.5%
- Europe — 19.9%
- North America — 9.7%
- Latin America and the Caribbean — 8.9%
- Middle East
- Africa
- Oceania
- Unknown or Stateless

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

Better Visualization?
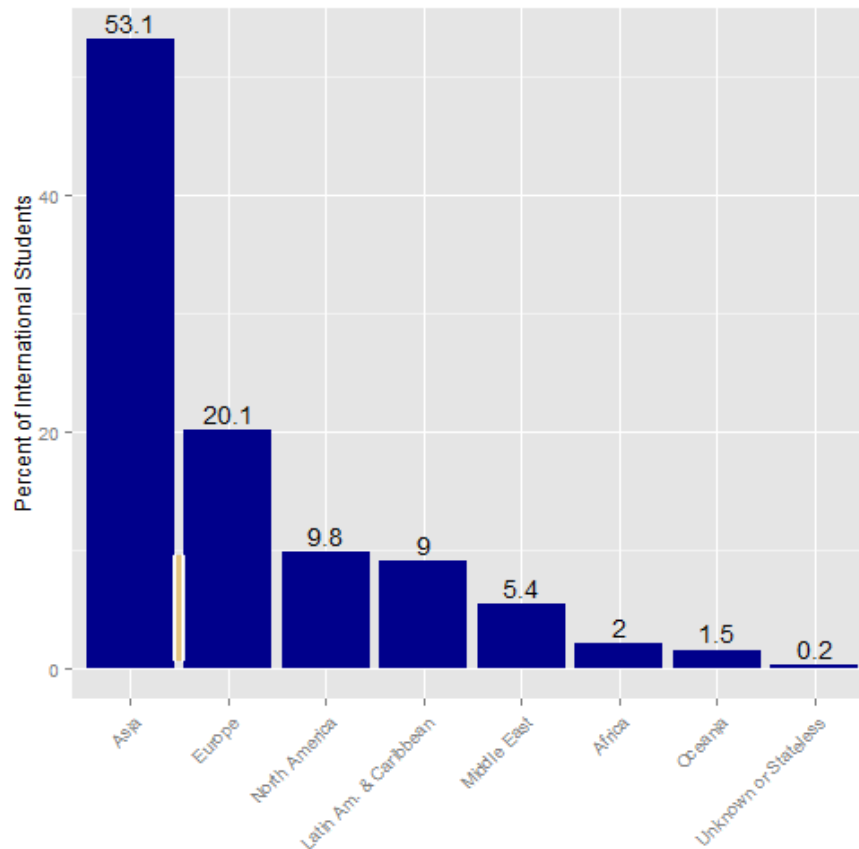
- All data is visible!
- Don't lose small regions.
- Can easily compare relative sizes
- Something to consider is that, for some people and applications, being not as "visually exciting" is a negative.

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.
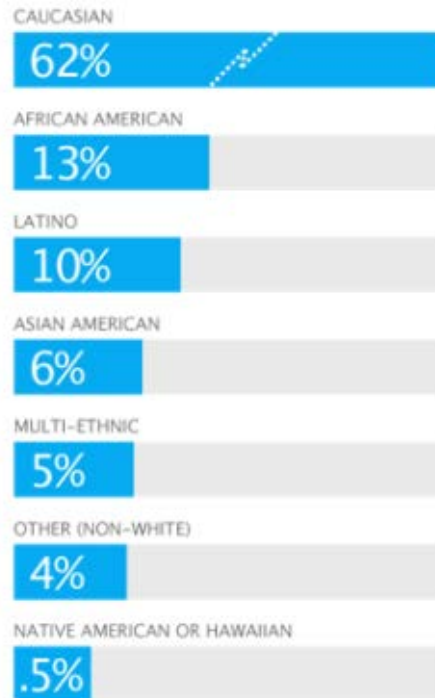
Bad Scales

Source: BBC

Bad Scales
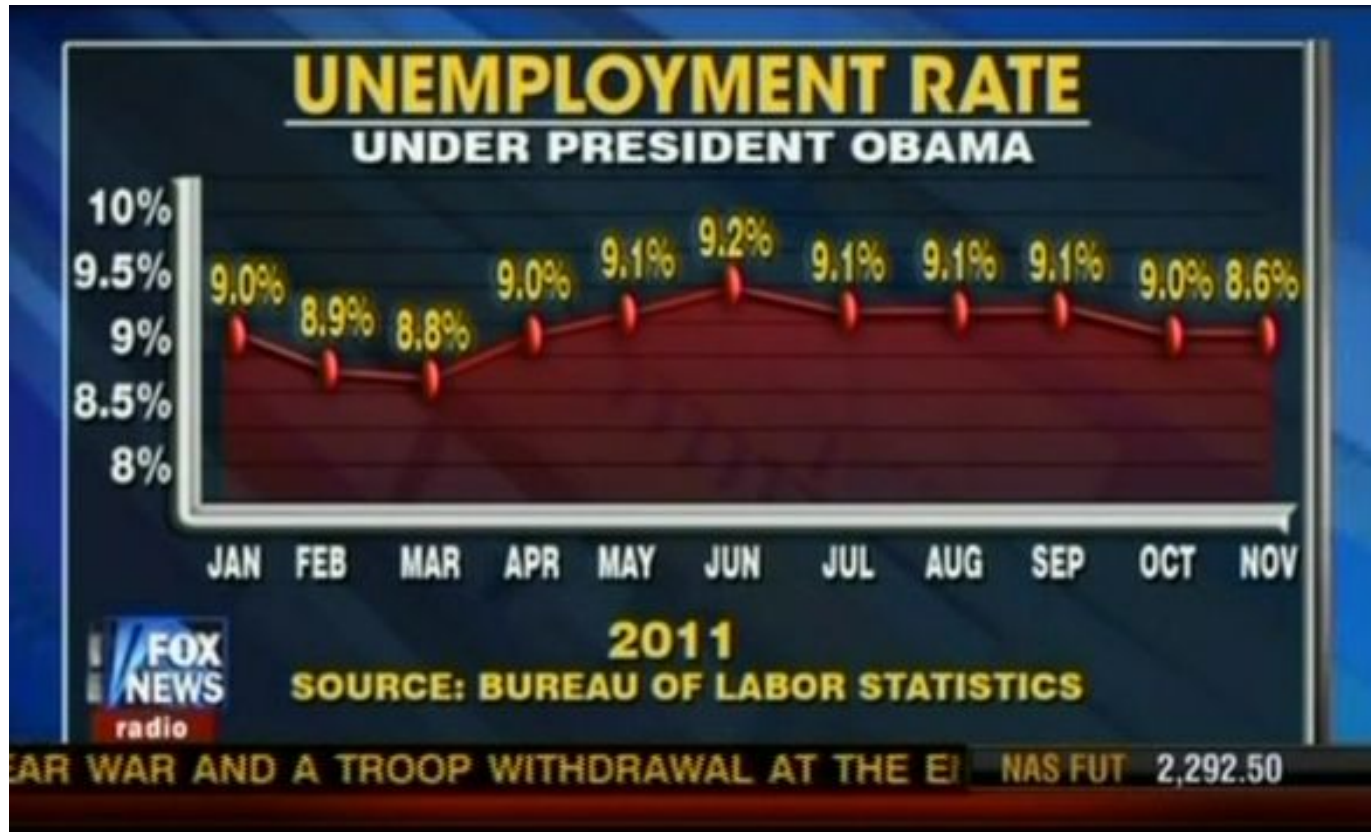
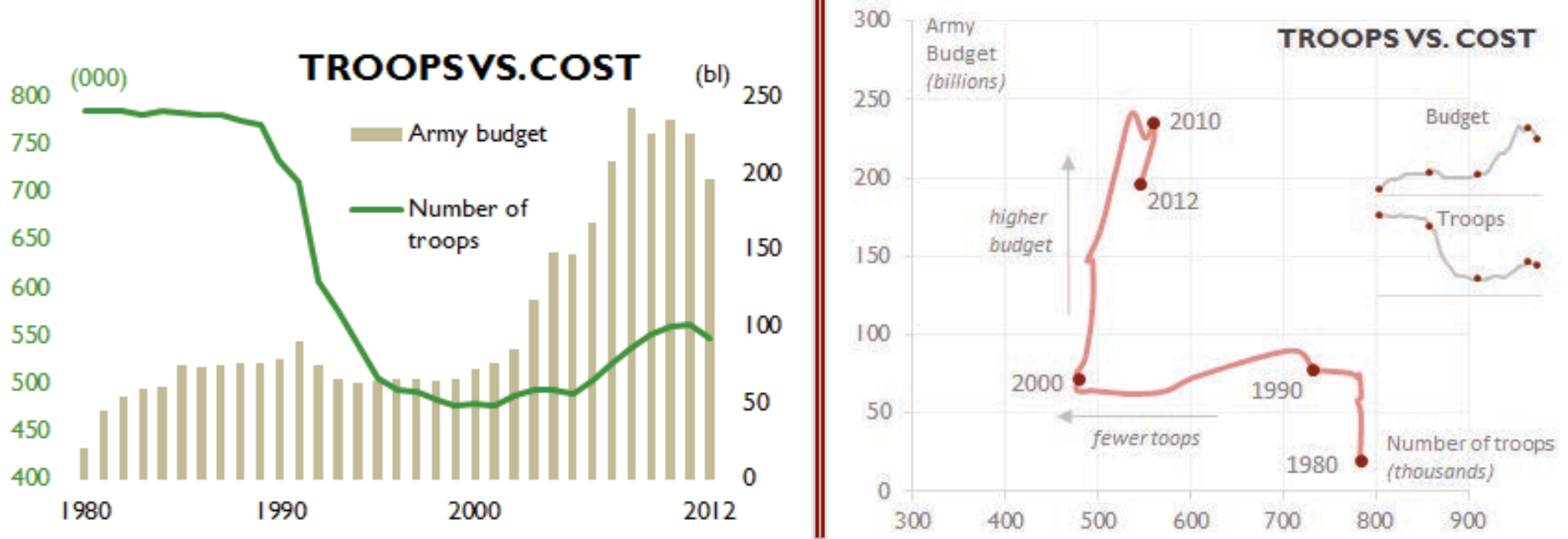Source: Fox News

Bad Scales

Diversity for 2012 Corps Members

Total people of color: 38%

CAUCASIAN
62%

AFRICAN AMERICAN
13%

LATINO
10%

ASIAN AMERICAN
6%

MULTI-ETHNIC
5%

OTHER (NON-WHITE)
4%

NATIVE AMERICAN OR HAWAIIAN
.5%

- "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

http://www.teachforamerica.org/why-teach-for-america/the-corps/who-we-look-for/the-importance-of-diversity
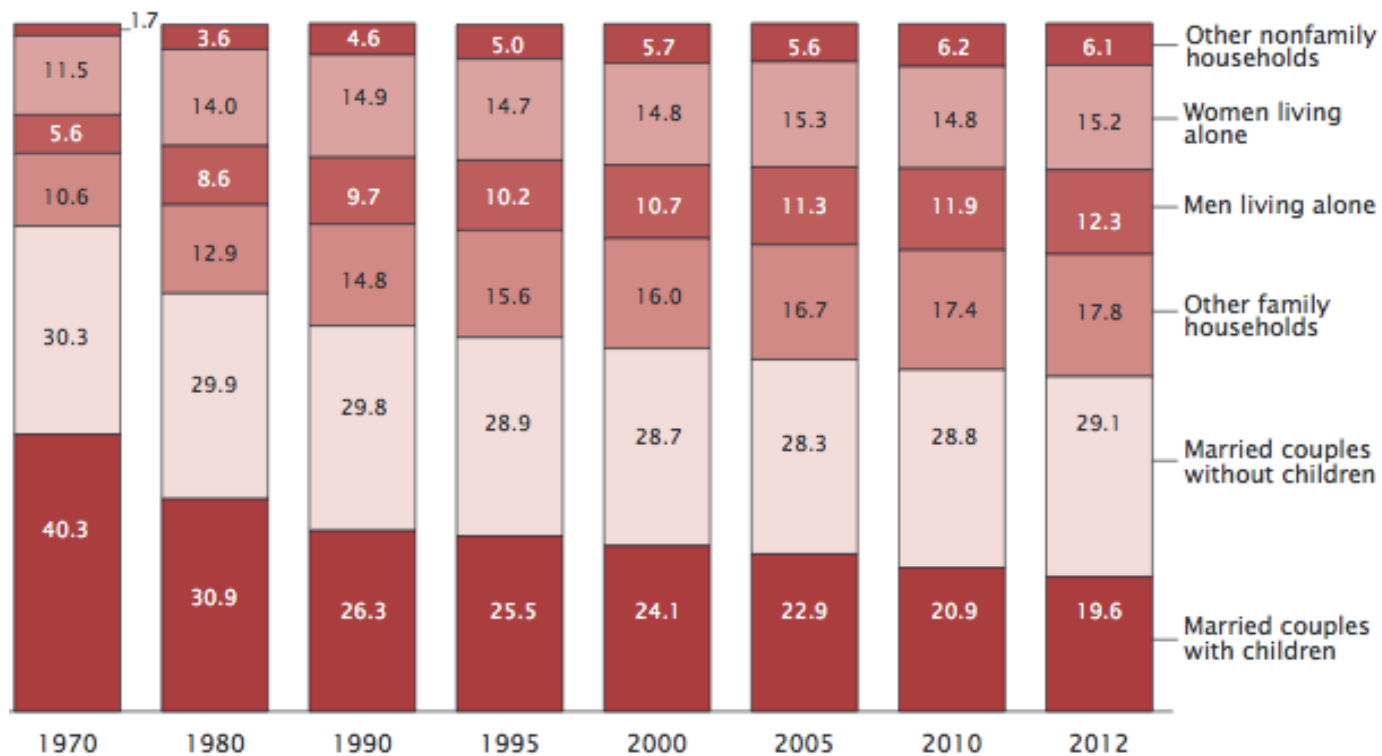
Two Rights Make A Wrong

Source: http://www.excelcharts.com/blog/redraw-troops-vs-cost-time-magazine/

- 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)

	1970	1980	1990	1995	2000	2005	2010	2012	
Other nonfamily households	1.7	3.6	4.6	5.0	5.7	5.6	6.2	6.1	
Women living alone	11.5	14.0	14.9	14.7	14.8	15.3	14.8	15.2	
Men living alone	5.6 / 10.6	8.6	9.7	10.2	10.7	11.3	11.9	12.3	
Other family households	10.6	12.9	14.8	15.6	16.0	16.7	17.4	17.8	
Married couples without children	30.3	29.9	29.8	28.9	28.7	28.3	28.8	29.1	
Married couples with children	40.3	30.9	26.3	25.5	24.1	22.9	20.9	19.6	

Source: U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, selected years, 1970 to 2012.

Family Matters

Households by Type, 1970 to 2012: CPS
(In percent)
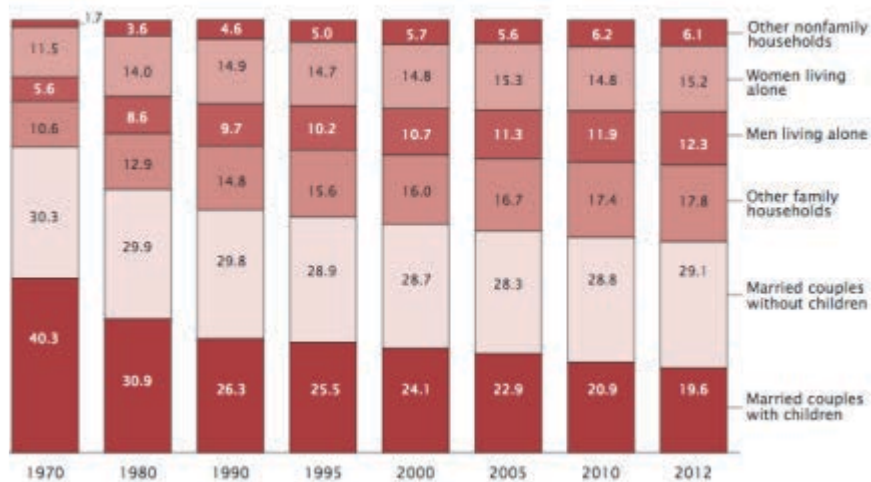
Source: U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, selected years, 1970 to 2012.

- If we are interested in shares within a year, its good.
- If we want to see rates of change, it is pretty much unusable!

- If we want to compare year-to-year, its possible though imperfect.
- Numbers are relative – absolute numbers may reveal, e.g. married couples without children is constant across years.

15.071 Analytics Edge
Spring 2017