



The *graph* is always greener on the other side

graphing and visuals tips, and what to avoid

Stephen Ketcham

Summary

- **Focus for this presentation on visuals** (graphs, charts, tables):
 - What's effective, what's not (the visual's purpose)
 - Underlying data in a visual: representative?
 - What misrepresents reality
 - Examples: good and bad
 - What to avoid, what's good and a few visuals for discussion
 - Watch out for these types of visuals!
 - Forecasting uncertainty
 - Combine data in unique ways without jumping to conclusions

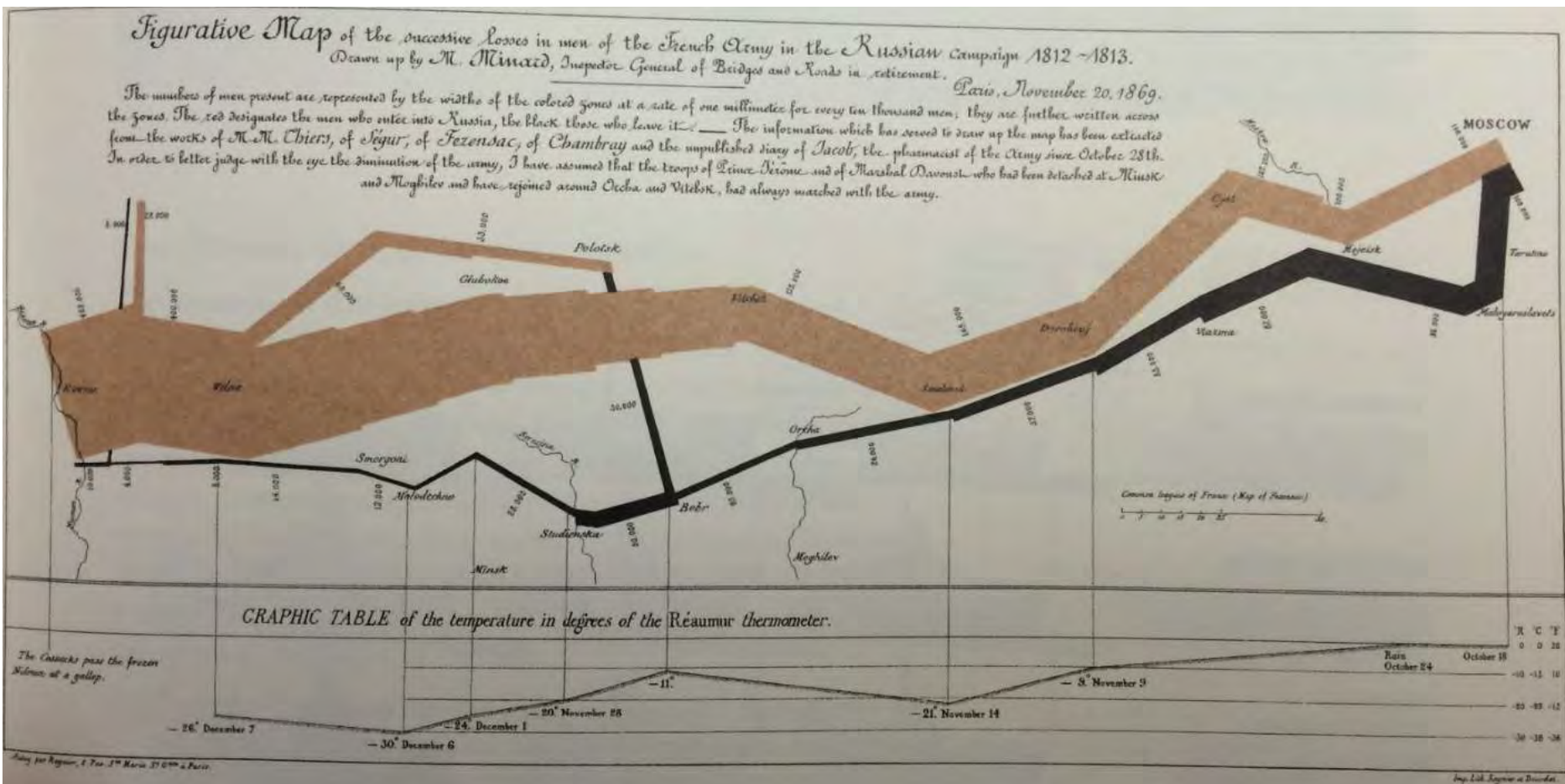
Introduction

- **The Purpose of Visuals:**
 - summarize information efficiently: Can show important drivers of CERs (cost estimating relationships)
 - quantify absolute and relative differences among “related” data (simple correlation illustrated or causation CER assumed? $y = f(x)$)
 - Trends are easily shown: change over time (time series)
- **The Power of Visuals:**
 - Visuals are remembered
 - Like a “killer app” for a mobile device
 - Can convince or mislead

Good Data, Good Graphs?

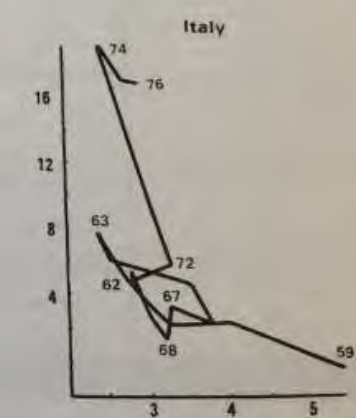
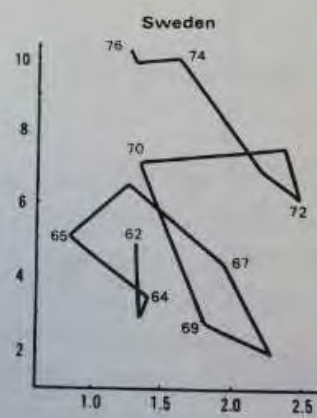
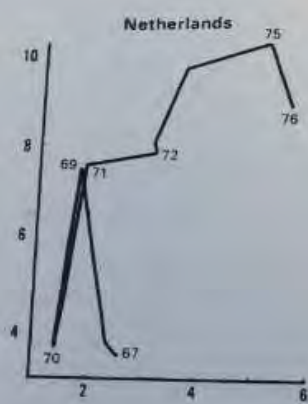
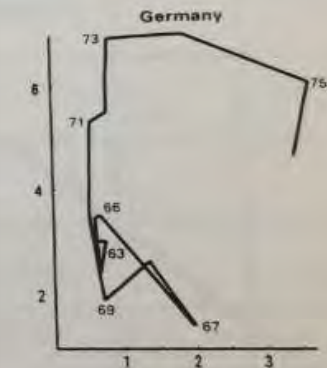
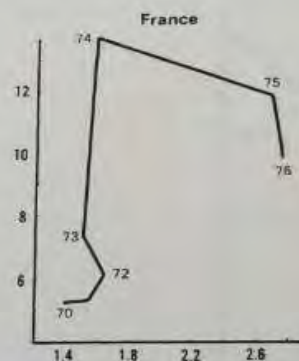
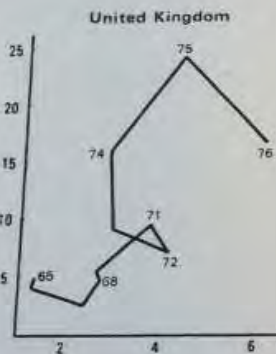
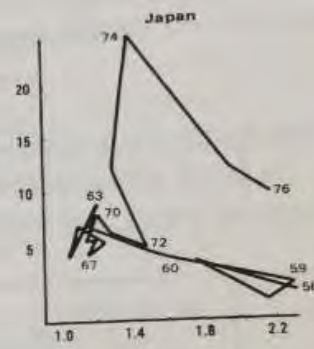
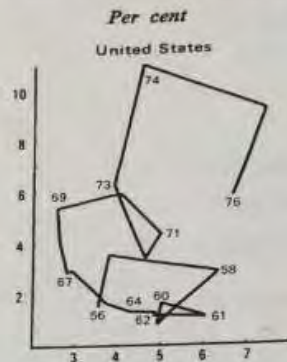
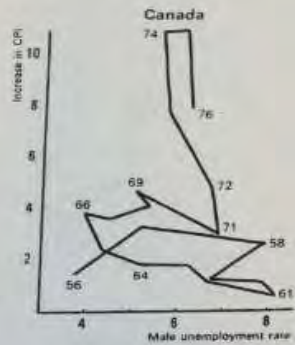
- **Statistical properties of the under-lying data**
 - Is the graph representative??
(sample reflects true population/reality)
 - Bias is avoided:
 - bias from incomplete sample
 - bias from sampling method
 - Has the data been normalized ?
 - Projections/extrapolations are grounded on solid assumptions
 - Uncertainty of the future is treated both:
 - statistically (confidence intervals)
 - honestly (past uncertainty/forecasting error is accounted for and adjustments are attempted to improve the forecast (model performance is tested via in-sample forecasts))

Good Data, Good Graphs?



Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard*, 1845-1869. Translation and drawing by Dawn Finley, Elaine Morse, respectively, 2002. – excerpt from Tufte, *The Visual Display of Quantitative Information*, 41.

The “Philips Curve”



- Proven economic theory from the 1960s breaks down in the 70s with the advent of “stagflation”
 - Breakdown in inverse relationship seen across many developed countries
 - big drivers high inflation – OPEC oil
 - Slow productivity growth, economic weakness, regulation?, price controls?

Paul McCracken, et al., *Towards Full Employment and Price Stability* (Paris, 1977, 106).

What to avoid presenting


- **Graphs of legacy costs illustrating implausible near-term outcomes – why unlikely in the mid-term?**
 - Costs that ignore the potential for solutions via other avenues (cheaper alternative via O&M funding enabling capital outlay deferral?)
 - The lingering positive effects of prior system enhancements/tech refreshes
 - Low cost midterm technical workarounds (configuration management, system consolidation, maintenance procedural changes, system performance acceptance limits changes, another competing system makes the system less critical)
 - Usage of available shadow inventory/field spares, or excess like-system remitted parts

What to avoid presenting

- Graphs that shows highly uncertain (and discounted) out-year effects dominating the outcome
- Unlabeled or poorly labeled graphs and charts
- Graphs that are not spaced correctly and distort
- Composite average relationships
- Cumulative average graphs
- Cherry-picked samples (*un-representative*)
 - Small
 - Favorable to the business case

What to avoid – confusing composites

System Supportability Scoring

1-2	Green	
3-4	Yellow	
5	Red	




Line Replaceable Unit (LRU) 1

Composite Score 4.8 

Line Replaceable Unit (LRU) 2

Composite Score 3.3 

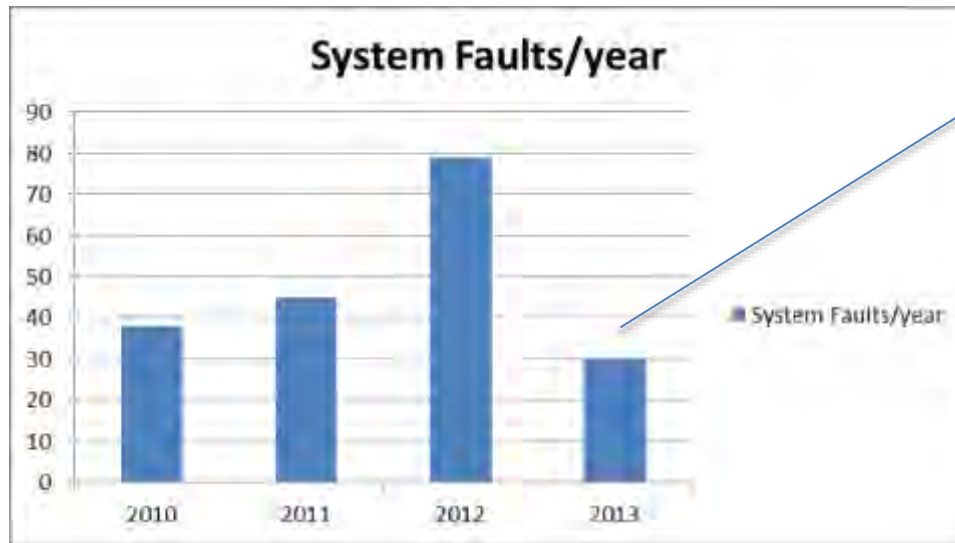
Piece Part Scoring

part 1	
part 2	
part 3	

Piece Part Scoring

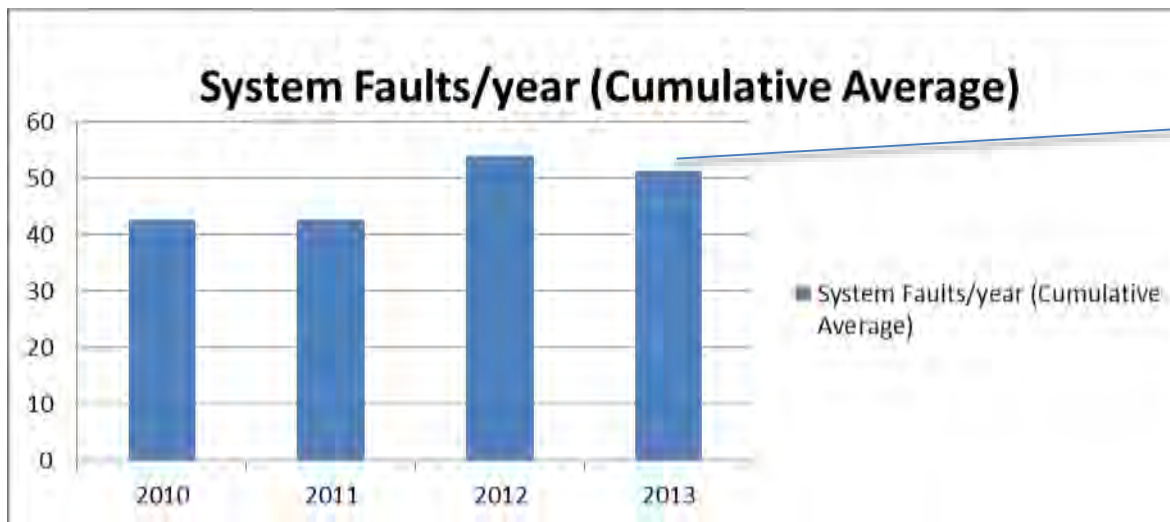
part 1		1
part 2		4
part 3		5

What to avoid – cumulative averages/smoothing



Large marginal decrease shown here clearly

If data is not noisy, then why smooth?



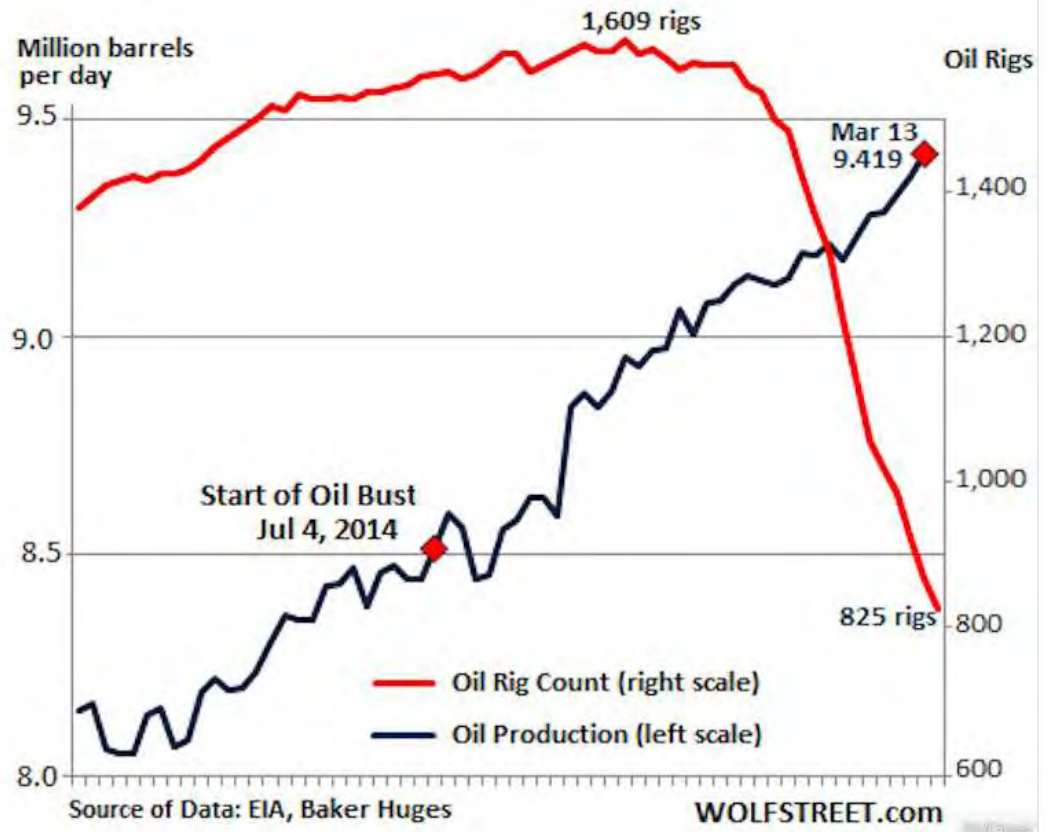
The decrease here is hidden

Avoid presenting – incorrect spacing/skewed

- If the economic relationship between rig count and production is production vs. *lagged* rig count then why not show it that way?
- why only a 15-month history?

The US Fracking Bust

Oil Production Boom, Oil Rig Count Plunge
Weekly, 2014 - 2015 YTD

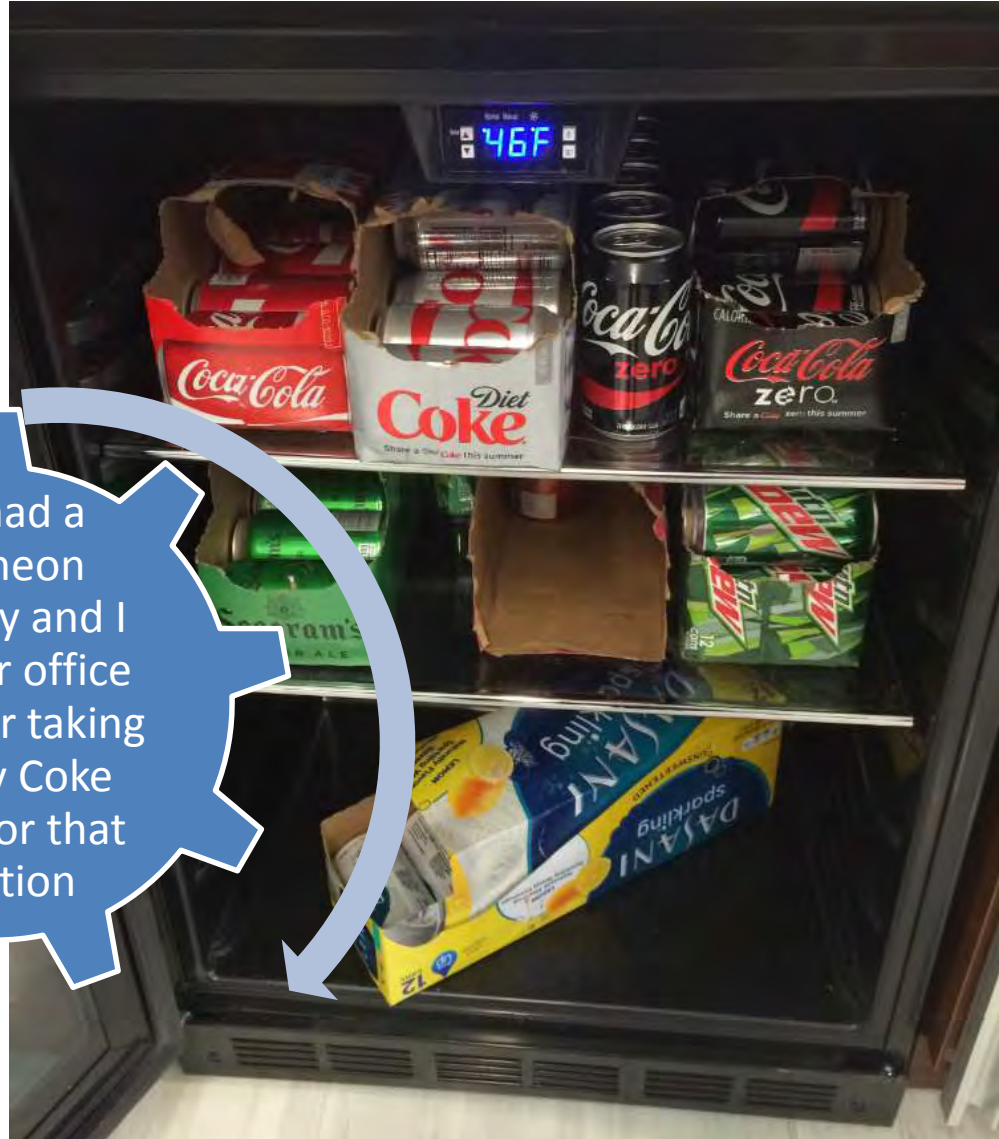
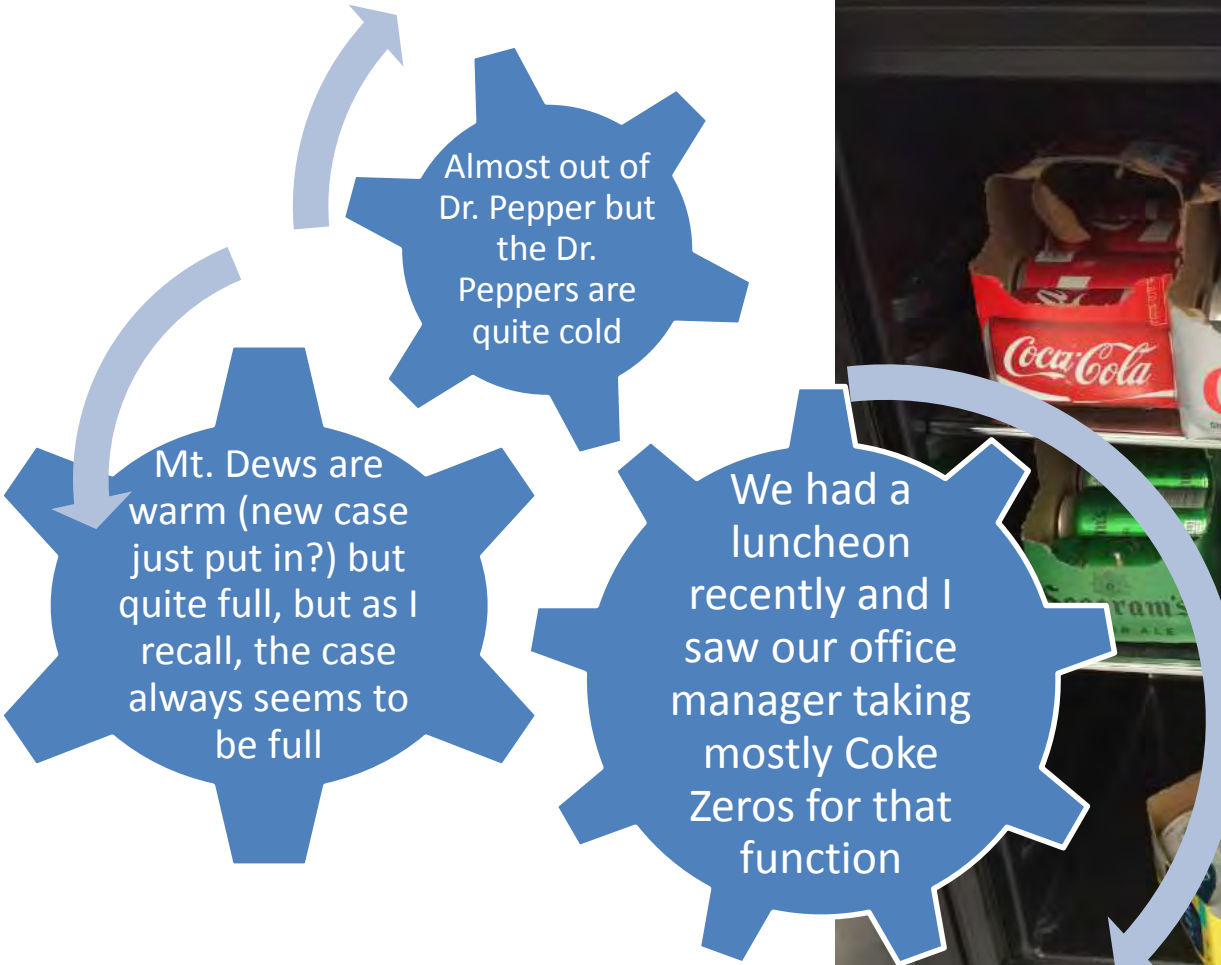


Source: <http://www.businessinsider.com/the-us-oil-bust-is-getting-uglier-2015-3>

Labels!!

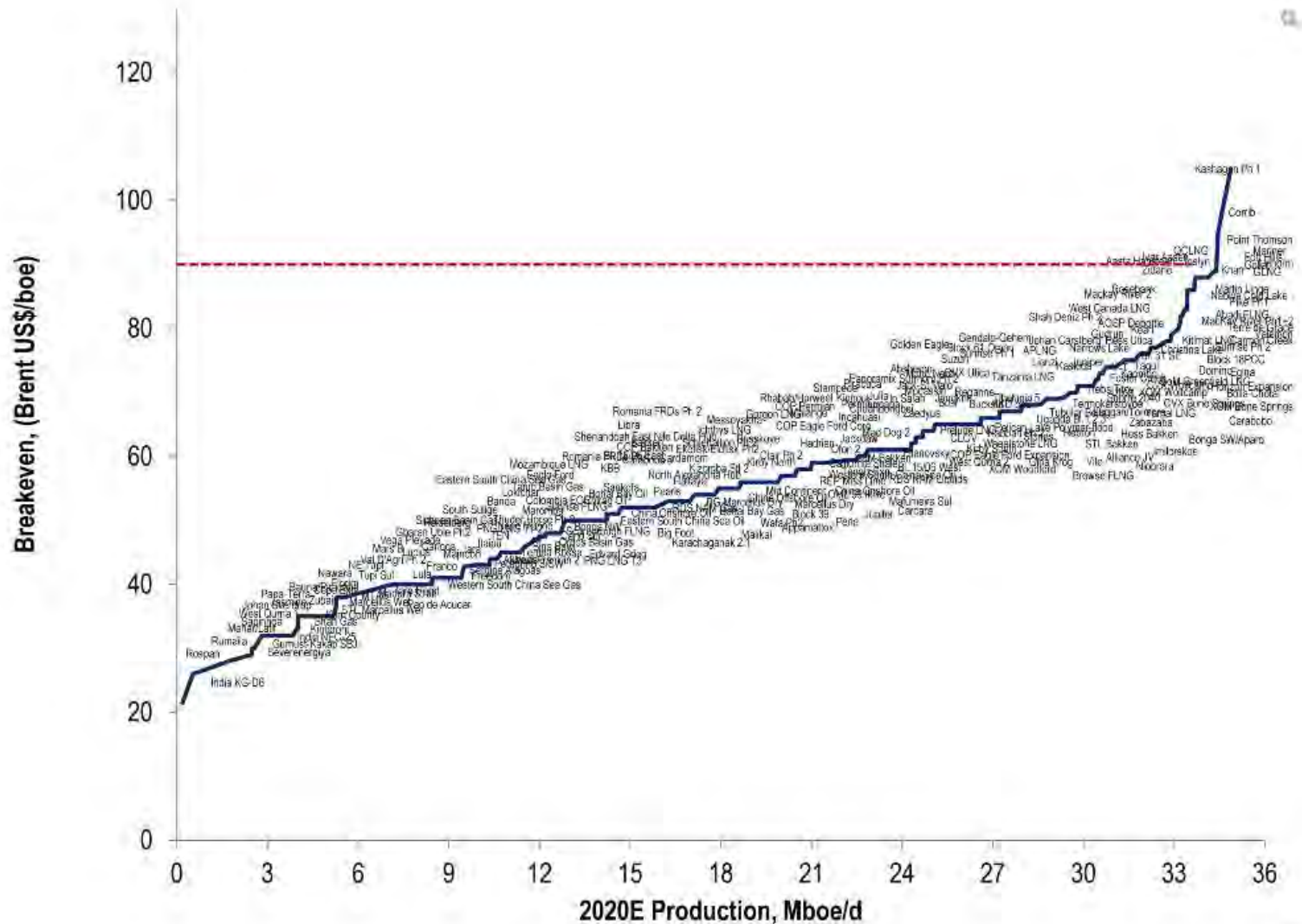
- **Always**
 - title your visuals
 - label each axis
 - Detail sample size, sample period (gives reader idea of how complete/representative your data is)

Examples: Cobec fridge problem - we need more data!



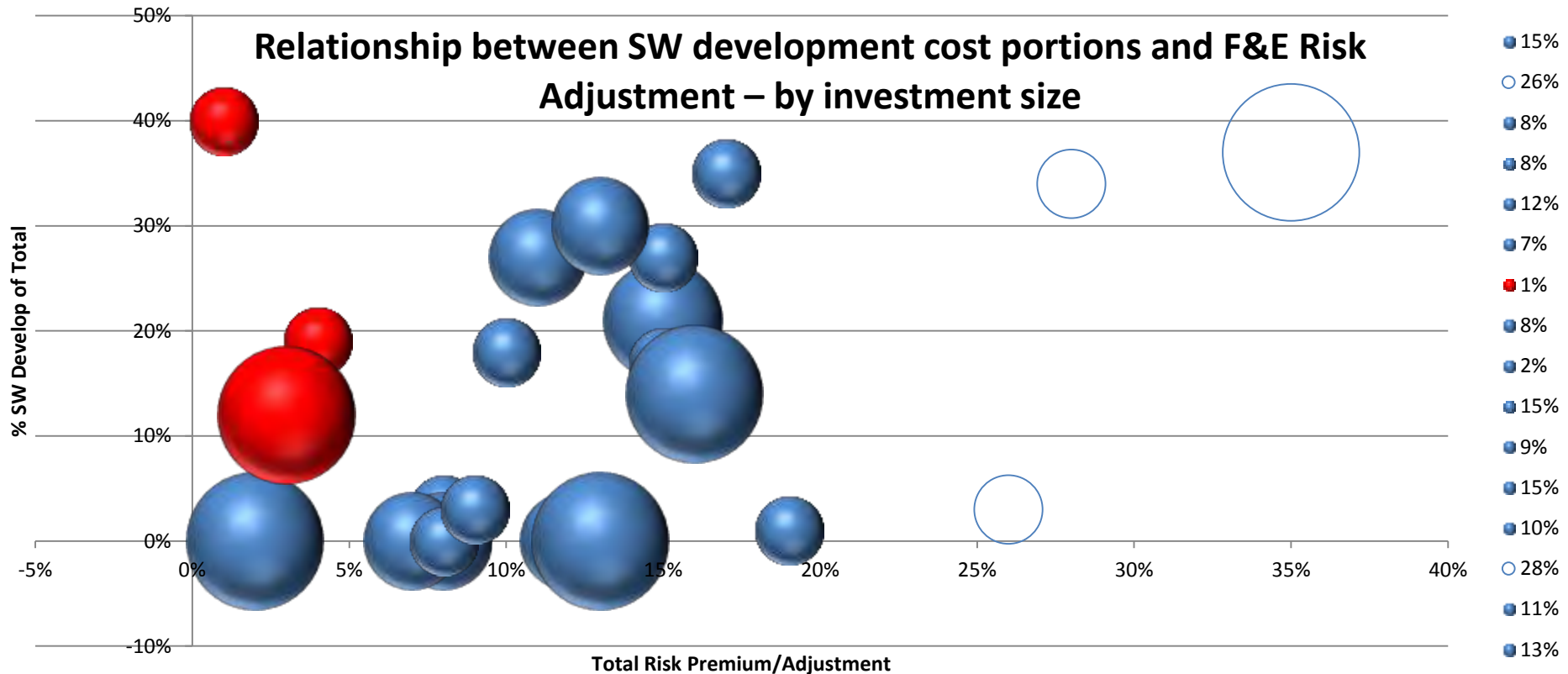
Examples: the Informative

Oil Production breakeven points - global *source Ed Morris Citi Inc. Nov 2014*

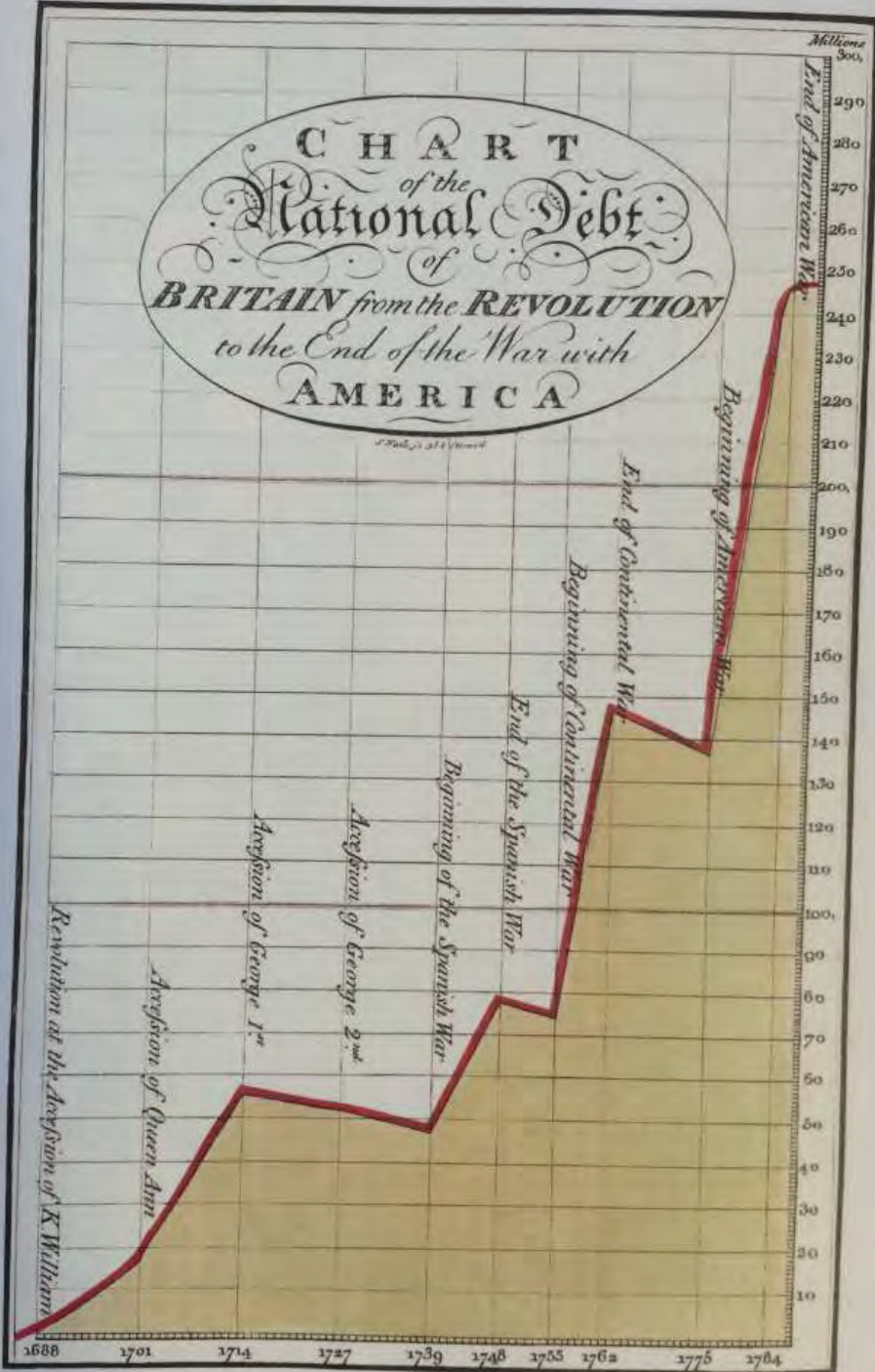


the Informative - choosing how to display your visual

Does SW cost really drive overall risk?



the Informative



British National Debt

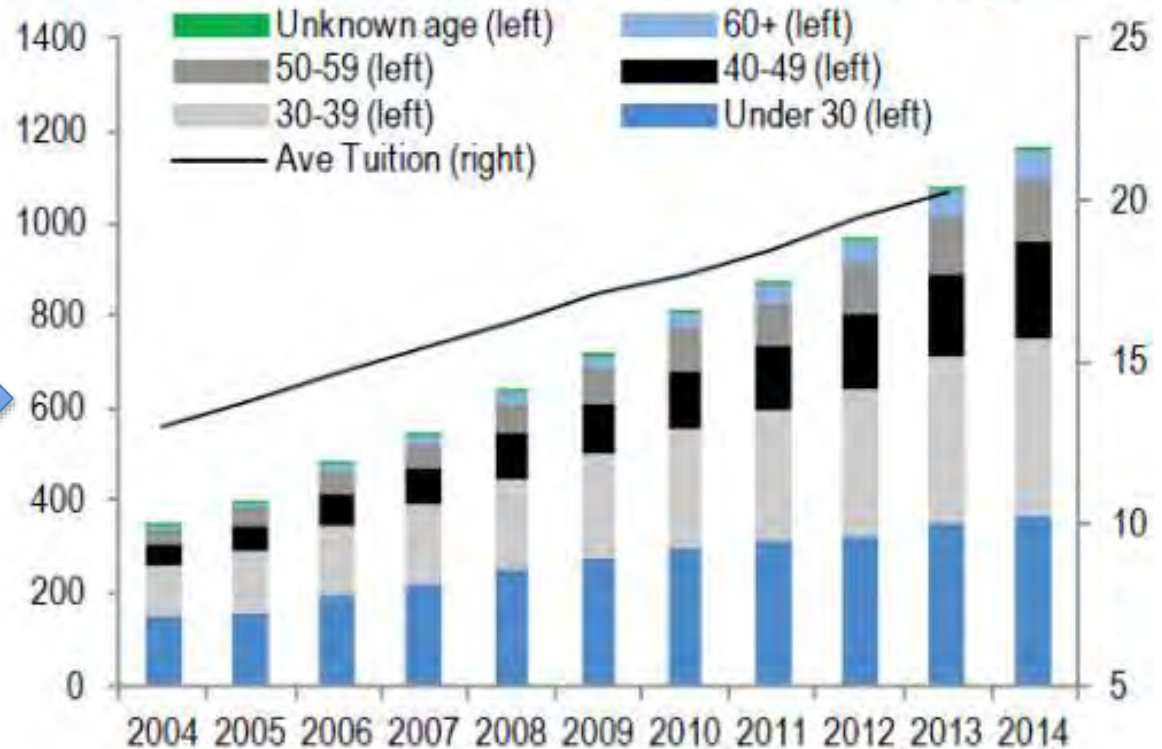
- *Emphasis* on key drivers of debt *but*
- X axis labels on key “debt drivers” dates like wars and the installation of new kings/queens are still in regular [non-skewed] intervals*

*Tufté, Edward. *The Visual Display of Quantitative Information*, 148.
Graph at left: Playfair, William, *The Commercial and Political Atlas*.

Good or Bad graph?

- The title looks to be true based on the data shown but 40-49 year olds' debt is also increasing
- Adjusted for inflation?
- Right vertical axis label? (\$Bs?)
- No real action/drivers in this chart, so it's a bit uninteresting

Exhibit 4: Tuition and student loan debt continued to increase, a drag for young first-time home buyers



Source: J.P. Morgan, Federal Reserve of New York, NCES

The dark blue and light grey bars, which represent borrowers under the age of 40, show the bulk of student loan debt is owed by young people.

Enhancing the Un-informative, correcting the Misleading

- **Correct the cumulative and composite average graphs (incremental)**
 - Why? Better to look *at the margin*, the most current and relevant data unless volatility in the numbers warrants hiding/smoothing out variance in the data
 - Composite averages can also hide the true cost/risk driver
- **Time series:**
 - **Add data** to show a more complete and accurate historical trend
 - If cyclical in nature, do we show beyond 2008 (US economic peak) ? What about only 2010-2014 recovery only? – **normalize for cycles or include a full cycle in your dataset**
 - Is system “learning” being represented accurately (SW, HW rollouts may take 2-3 years to achieve learning effects – are the high learning costs the first 2 years being used as a selling point for the next tech refresh?) – **correct by including the latest cost information**

Enhancing the Un-informative, correcting the Misleading

- **Regression can be graphical regression!**
 - presenting low R^2 , high standard error regressions
 - Ignoring relevant but inconvenient data points
 - Blind extrapolation of weak CERs is dangerous: when the inherent uncertainty around the predicted value(s) is ignored decision makers may not be aware of the risks

Enhancing the Un-informative, correcting the Misleading

- **Prescriptions:**

- Consider relationship causality carefully before getting graphical!
 - **Example:** Apple stock price = $f(\text{size of Iphone})$ *or*
Apple stock price = $f(\text{historical earnings, futures earnings}(\text{branding effects, new products}), \text{economic wealth effects})$
- Does the relationship make sense, what else could be a factor
- Always scatterplot, if the scatterplot doesn't clearly show a relationship don't get fancy with a regression; better to illustrate the complexity (or lack thereof) of the relationship rather than average data out
- Consider regression statistics carefully, and report them, e.g SE
- Research the outliers, **find out why** they are outliers
- Always draw confidence/prediction intervals

Conclusion

- Underlying data in a visual: representative?
- What's effective, what's not (the visual's purpose)
- What misrepresents reality
- Examples: good and bad
- Watch out for these types of visuals!
- Forecasting uncertainty
- Combine data in unique ways without jumping to conclusions
- Make your visual a centerpiece of your business case/briefing, tying everything together