Data Visualization

A product of human design

TECOLOTE
R E S E A R C H

May 14-17, 2019
Prepared for ICEAA 2019

Introduction

Professional background
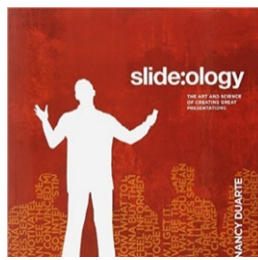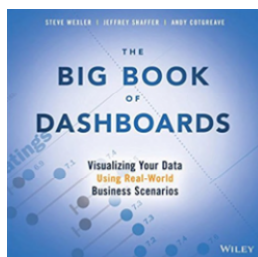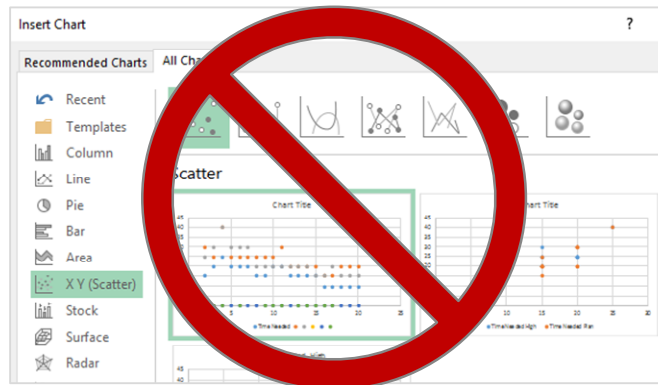

↑ **Last ~5 years**

My data viz background

Theme: a product of human design

Presentation objectives

- Provide a **brief introduction** to data visualization.

- Provide the audience some **actionable guidance**.

- Convince you that **it matters**.

Definition

Presentation of data in a

Graphical manner

Area Graph   Bar Chart   Box & Whisker Plot

Gantt Chart   Heatmap   Histogram

For the purpose of

Exploration

Understanding

Communication

Focus of presentation

But why should you care? →

Reference: http://www.perceptualedge.com/blog/?p=1897

Why should you care?

- It's relevant to your work.

- You are probably not very good at it.

- You've failed to convey… and convince a decision maker on a key issue.

Uninformed… highly subjective… data viz creation process

1. Grab all the data

2. Make a chart (optional)

3. Add everything that could be useful

4. Make it *better*

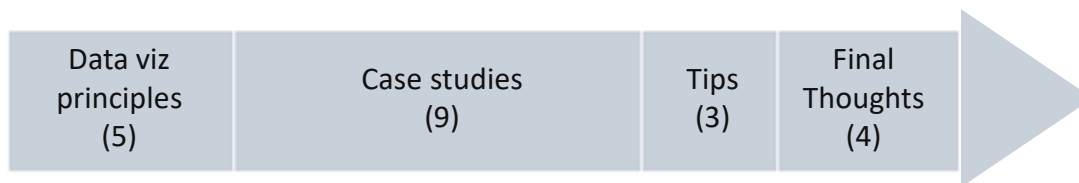5. Put main message at the bottom (optional)
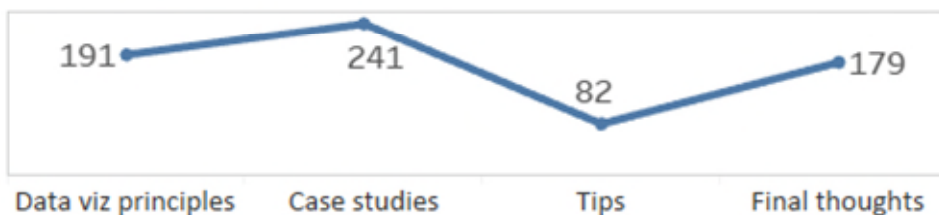
Systematic

Subjective
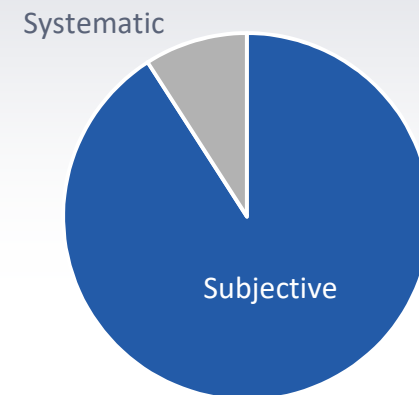
Breakdown of decision process

Overview

Presentation timeline

 # slides per section

| Data viz principles (5) | Case studies (9) | Tips (3) | Final Thoughts (4) |

More detail...

 # words per section

191 • 241 • 82 • 179

Data viz principles — Case studies — Tips — Final thoughts

Useless detail...

Picture = 1000 words

 # equivalent words per section

11K — 16K — 6K — 5K

Data viz principles — Case studies — Tips — Final thoughts
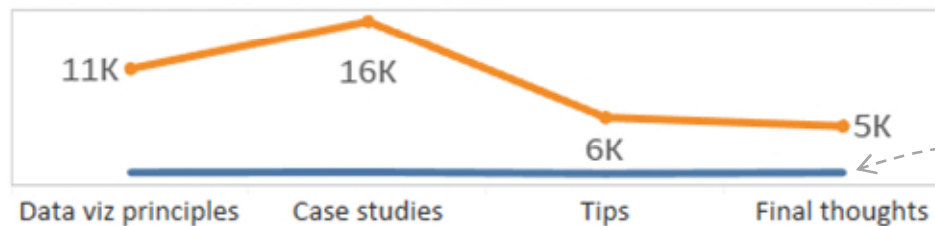
'#words per section' reference line

Data visualization principles
3 concepts to get you started

Pre-attentive processing

Getting you to understand without thinking.

How many 5's are there?

Attentive processing

```
98734979027564790289472862409240603707057 0279072
80320802900730250127023700837408207872027 2007083
24780260270379377570970737797066746209709 4702780
92797970972309723097959275092727979873497 2608027
```
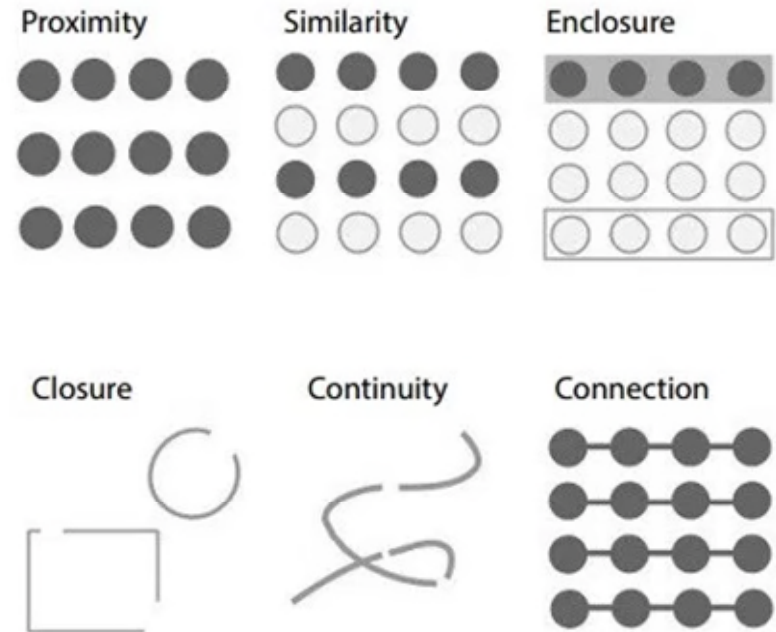
Pre-attentive processing

```
98734979027564790289472862409240603707057 0279072
80320802900730250127023700837408207872027 2007083
24780260270379377570970737797066746209709 4702780
92797970972309723097959275092727979873497 2608027
```

A more comprehensive definition: http://www.infovis-wiki.net/index.php/Preattentive_processing

Gestalt principles

Methods to differentiate your data points/series.

- We visually and psychologically attempt to make **order out of chaos**.

- Gestalt principles describe **how our mind organizes individual visual elements into groups**, to make sense of the entire visual.

- When designing a visual, these principles can be **used to highlight patterns that are important to us, and downplay other patterns**.
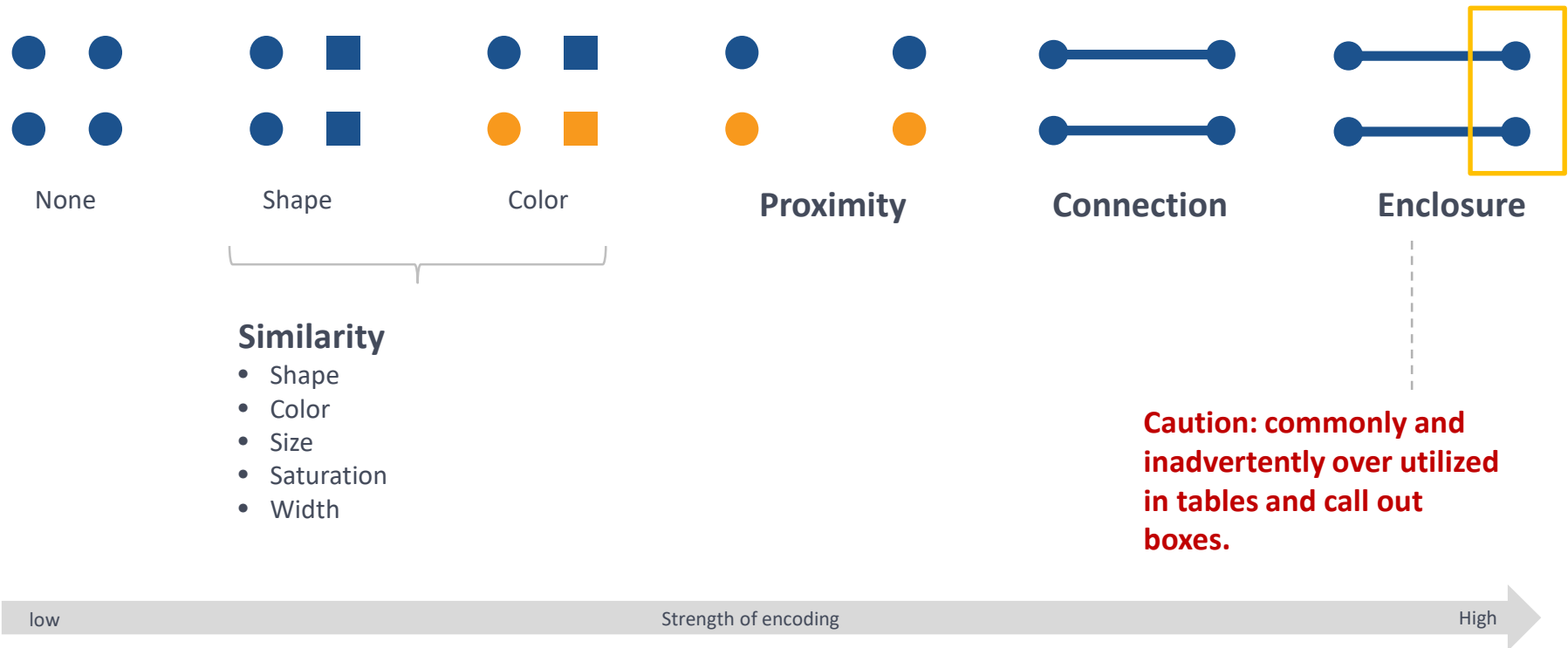
References
- http://www.fusioncharts.com/blog/2014/03/how-to-use-the-gestalt-principles-for-visual-storytelling-podv/
- https://www.smashingmagazine.com/2014/03/design-principles-visual-perception-and-the-principles-of-gestalt/
- https://emeeks.github.io/gestaltdataviz/section1.html
- https://excelcharts.com/data-visualization-excel-users/gestalt-laws
- http://facweb.cs.depaul.edu/sgrais/gestalt_principles.htm

Gestalt principles: varying degrees of strength

None   Shape   Color   **Proximity**   **Connection**   **Enclosure**

Similarity
- Shape
- Color
- Size
- Saturation
- Width

Caution: commonly and inadvertently over utilized in tables and call out boxes.
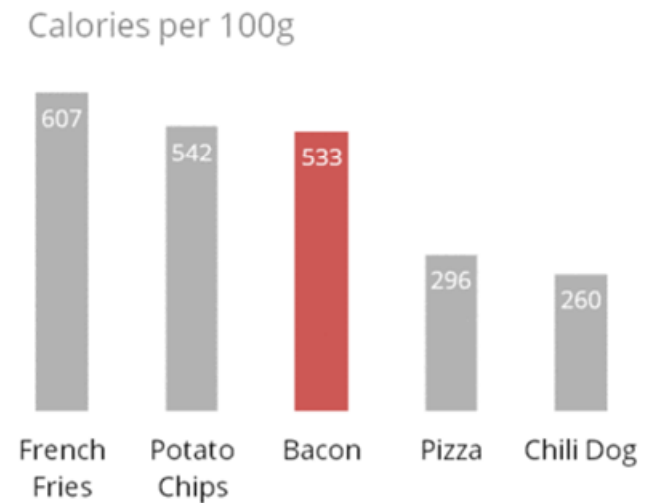
low   Strength of encoding   High
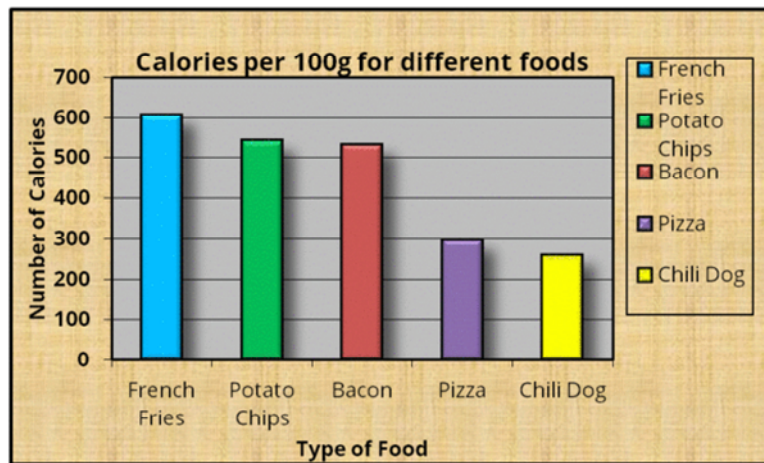
Key points
- The various methods to encode data have different strengths.
- When we over-encode… we create complexity… which leads to slower understanding.

Data ink ratio

Getting rid of chart junk

Key points

- Reducing cognitive overload
- Not about being minimal for the sake of being minimal

$$\text{Data ink} = \frac{\text{Ink used to describe data}}{\text{Ink used to describe everything else}}$$

Chart reference: http://www.darkhorseanalytics.com/blog/data-looks-better-naked
Add'l info: http://www.infovis-wiki.net/index.php/Data-Ink_Ratio

Case Studies

Disclaimer: original case studies performed on actual data.
Examples have been sanitized for ICEAA.

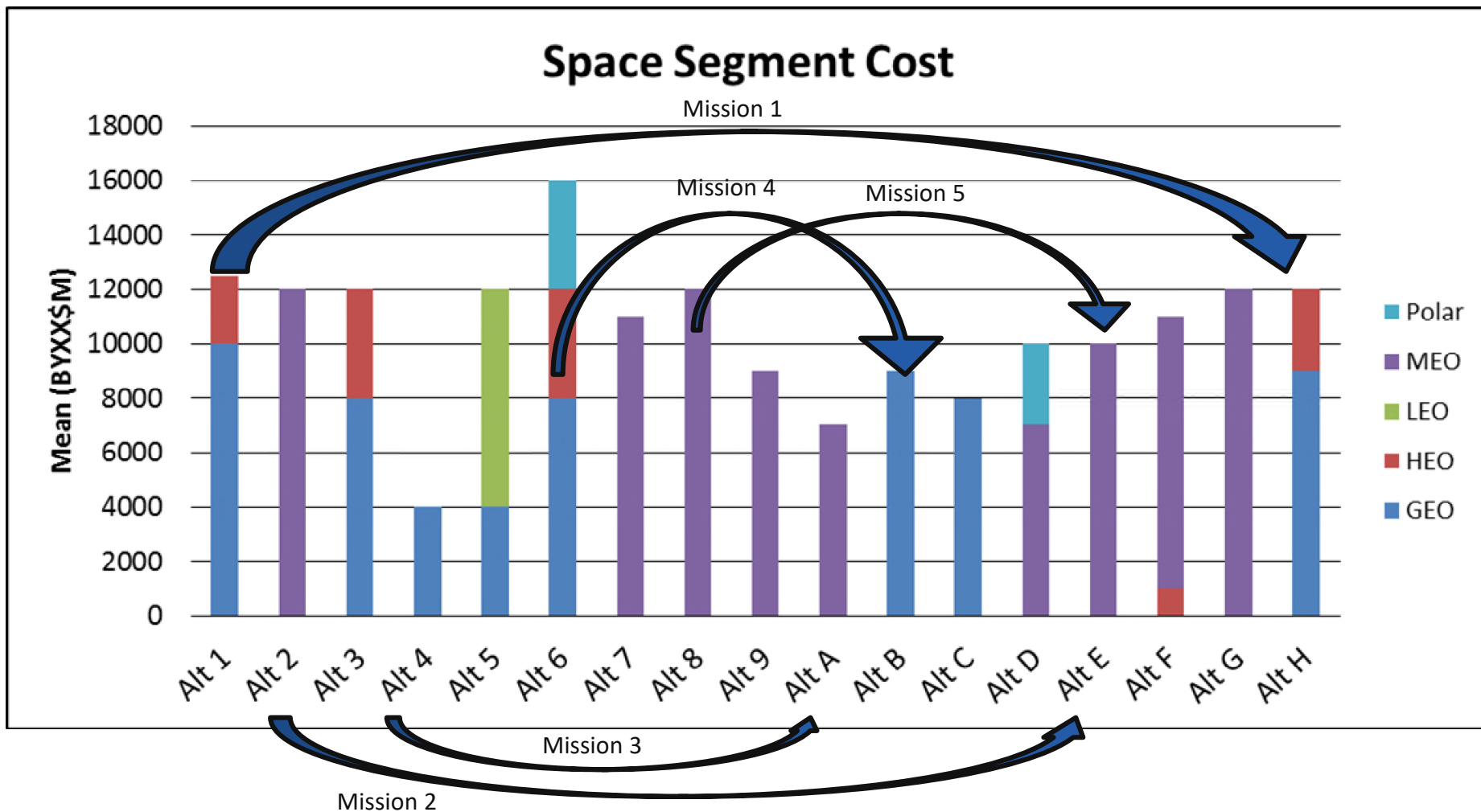# Case Study #1
Analysis of Alternatives

Scenario: 2 cost estimates performed on different combination of satellites.

Lesson learned: Break down complex charts into more simple parts.

Round 1

Round 2

VS

Alternatives 1-9

Alternatives A-H
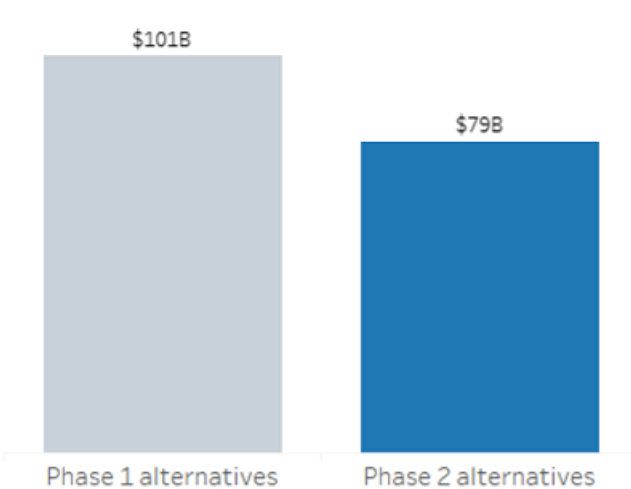
Alternative Re-constructing

Case Study #1: original

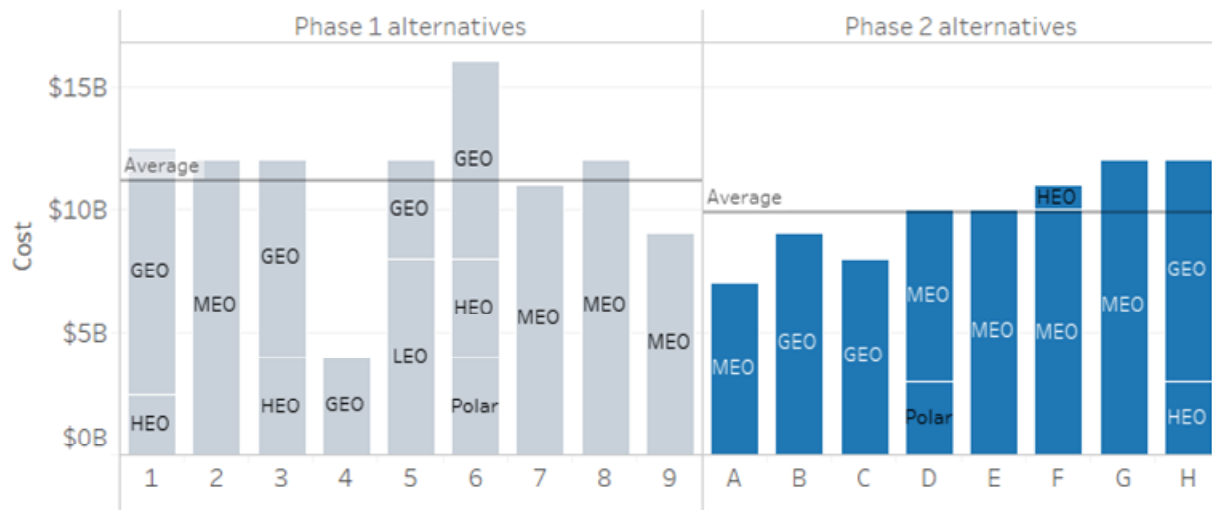**Alts A-H built from lessons learned in first phase of analysis**

Phase 2 (Alt A-H) built from lessons learned in Phase 1 (Alt 1-9) of analysis
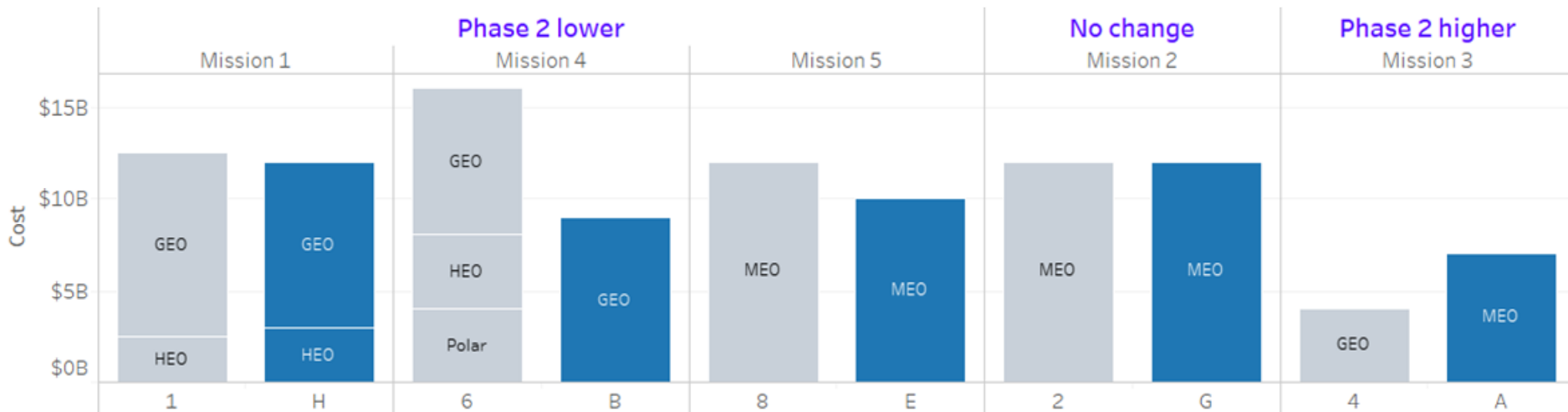## Case Study #1: revised

1) Comparison at the total
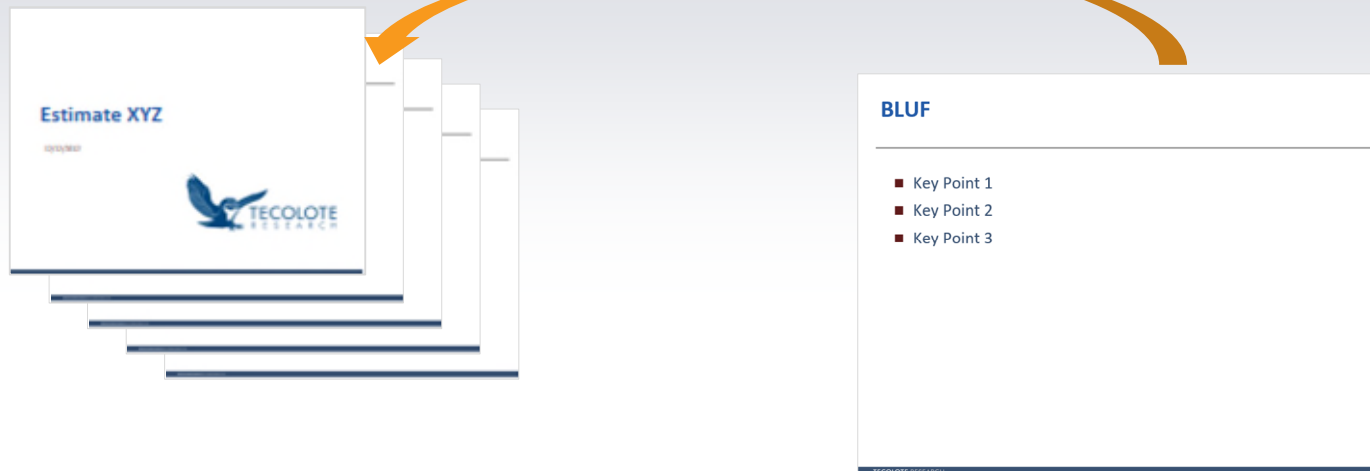
2) Detailed comparison

3) Re-configured missions

Case Study #2
Bottom Line Up Front

Scenario: Presentations to upper management commonly start with a summarized key point slide (a.k.a. the BLUF)

Lesson learned: Visualizing your numbers can have a bigger impact than inserting them in text.

Bottom Line Up Front (BLUF)
Case Study #2: Original

- Single Greatest Estimate (SGE) Total by Appropriation: $5,600M

- Current Draft SGE (w/Fact of Life) changes: $5,600M (-$300M)

- Background

- Fact of Life Changes
  - +$10M: Design changes
  - +$30M: Change Fee
  - +$20M: Other costs
  - -$170M: Change in scope
  - -$190M: Wraps

- Recommended Actions
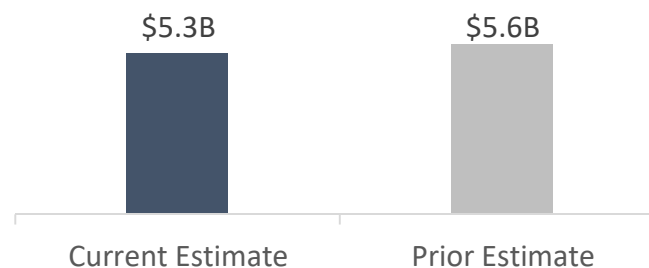 - 2-Letter leadership Draft SGE approval

# Bottom Line Up Front (BLUF)
Case Study #2: Revised

Key points

- Meeting objective: to obtain 2-Letter Draft estimate approval.

- Estimate is $300M less than prior year.

Current vs Prior Estimate

$5.3B

$5.6B

Current Estimate

Prior Estimate

How the $300M delta breaks down

$-170M, Change in scope

$-190M, Wraps

$20M, Other Costs

$10M, Design changes

$30M, Fees

-400   -350   -300   -250   -200   -150   -100   -50   0   50   100

TY $M
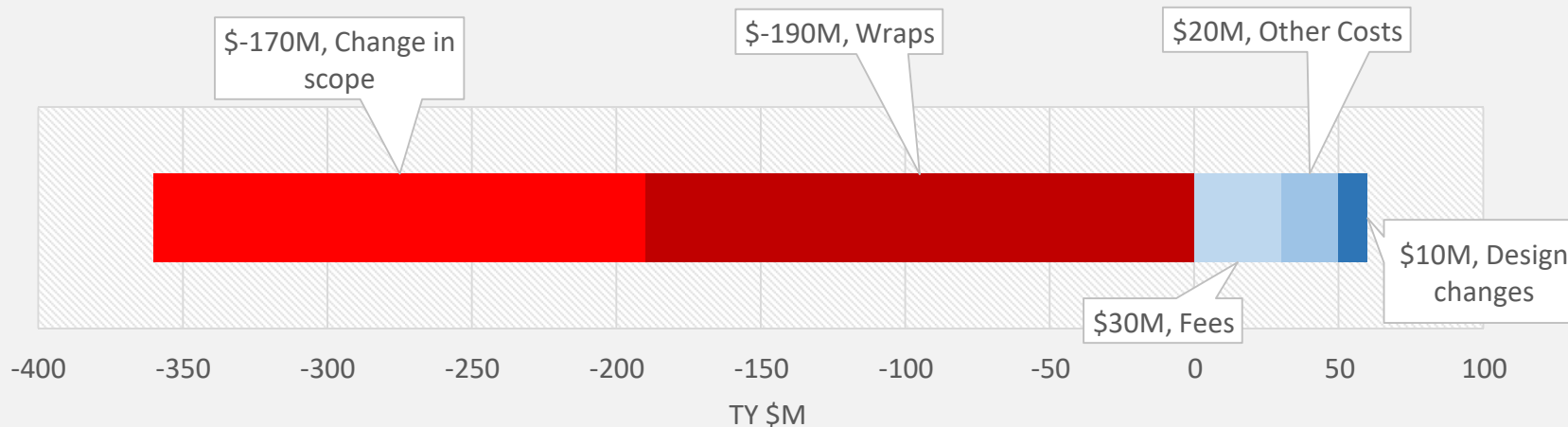
Case Study #3
Cost estimates

**Scenario:** Showing the results of your estimate to leadership.

Lesson learned: Experiment!

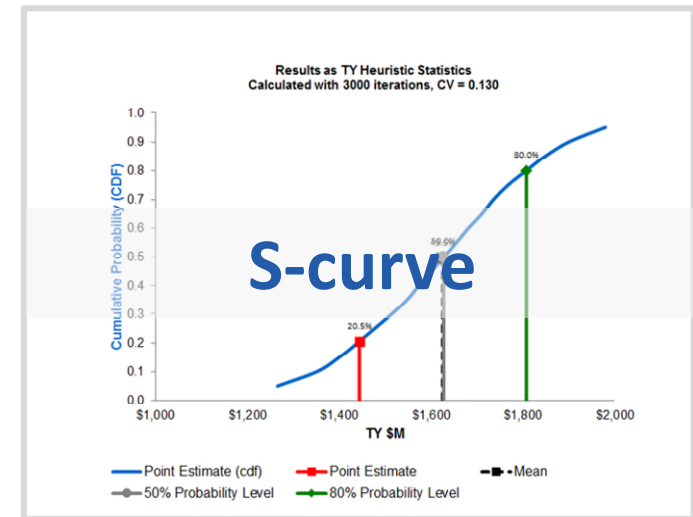# Typical visualizations shown during estimate reviews
Case study #3: original

Table of results

WBS	Estimate	Independent Estimate	Delta	Methodology
Total	123	456	333	
Space Vehicle	123	456	333	
System 1	123	456	333	
Nonrecurring	123	456	333	
Payload	123	456	333	ABC
Bus	123	456	333	DEF
SEITPM	123	456	333	GHI
Recurring	123	456	333	
Payload	123	456	333	ABC
Bus	123	456	333	DEF
SEITPM	123	456	333	GHI
Fee	123	456	333	

System Name

Budget comparisons

Legend: ■ Estimate ■ Budget
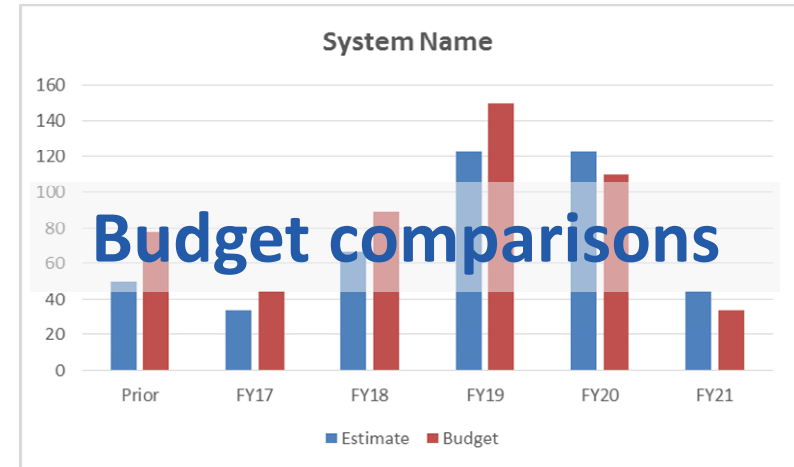
Total	Prior	FY17	FY18	FY19	FY20	FY21	Total
Estimate	123	123	123	123	123	123	738
Budget	456	456	456	456	456	456	2736
Delta	333	333	333	333	333	333	1998

Results as TY Heuristic Statistics
Calculated with 3000 iterations, CV = 0.130

S-curve

Legend: — Point Estimate (cdf), ■ Point Estimate, ■ Mean, ● 50% Probability Level, ● 80% Probability Level

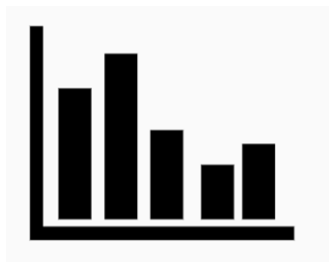# Estimate Dashboard

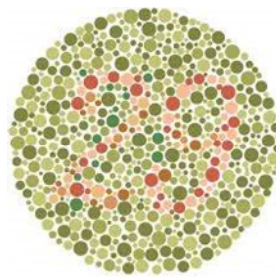## *Case study #3: Revised*
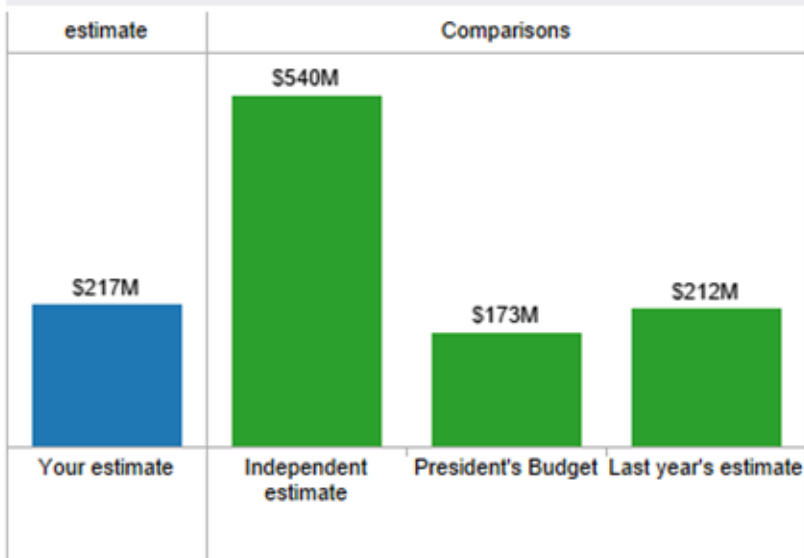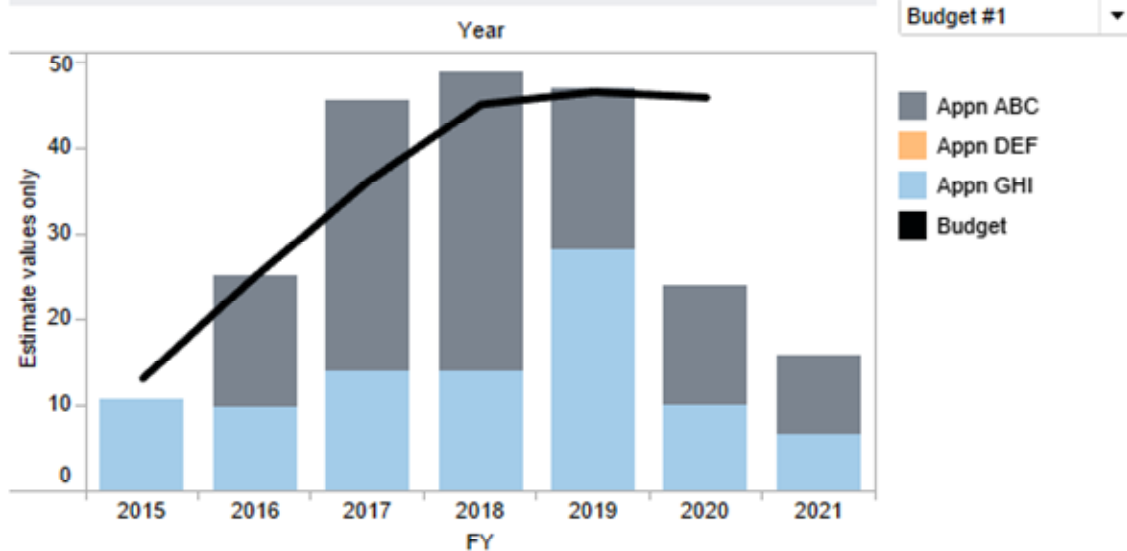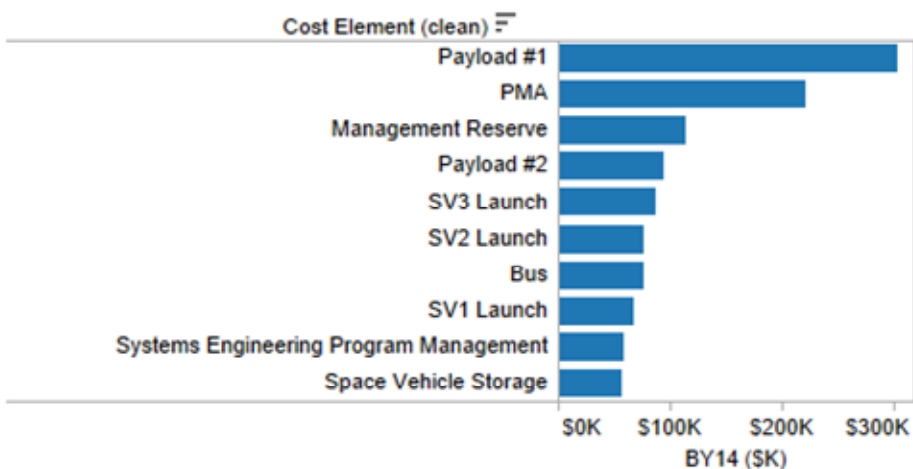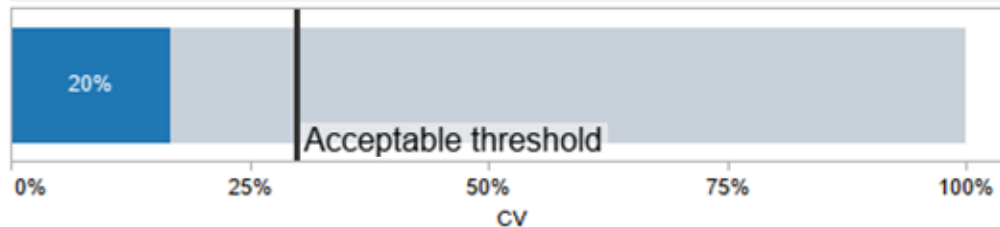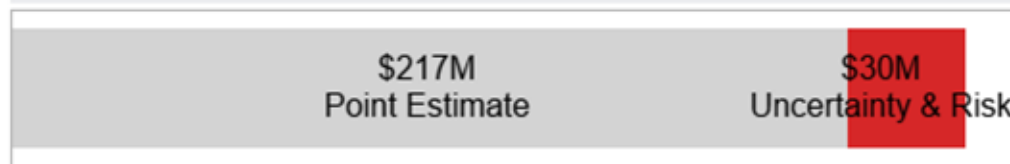
Data visualization tips

Practical tips

Start with your key points

- Key point 1
- Key point 2
- Key point 3
- ...
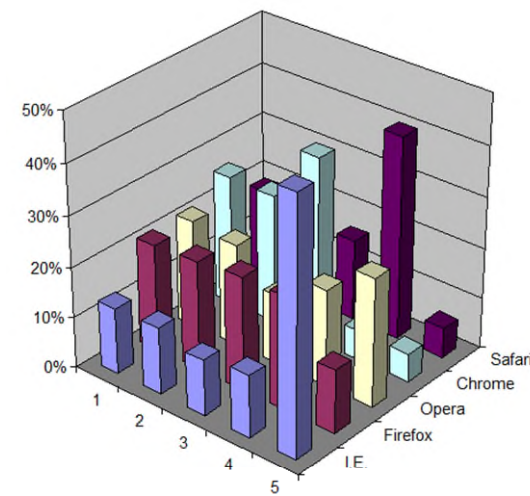
Consider color blindness

STOP BOLDING EVERYTHING

Limit the number of colors

AE (C-E)	Program 21
Coriolis	Program 22
DMSP 5D1 (1-4)	Program 23
DMSP 5D2 (8-1..	Program 24
Geolite	Program 25
GPS (1-8)	Program 26
GPS (9-11)	Program 27
LCROSS	Program 28
Mightysat II	Program 29
OSO	Program 30
P72-2	Program 31
P78-1	Program 32
P78-2	RHESSI
Program 15	S3
Program 16	SMS
Program 17	STPSat 3
Program 18	STPSat 5
Program 19	
Program 20	

Label your data directly if possible

Legend

- blueberries
- rasperries
- grapes

Direct labeling
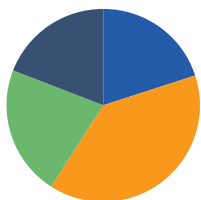
blueberries

raspberries

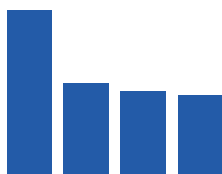grapes
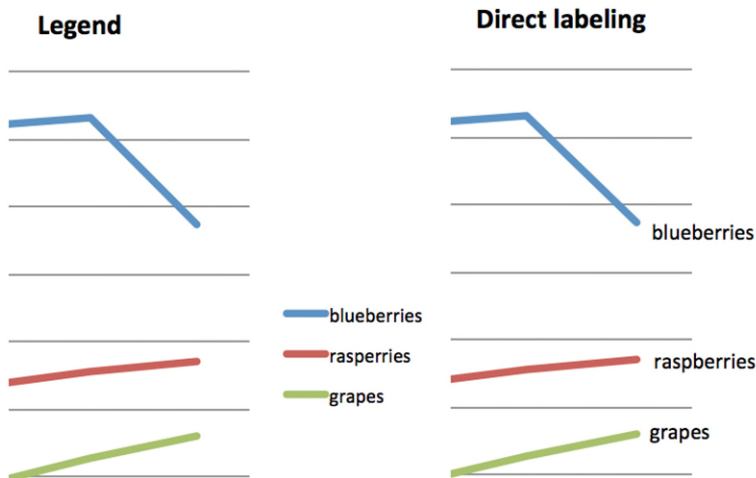
Special effects aren't cool

Practical tips (continued)

Bars, not pies

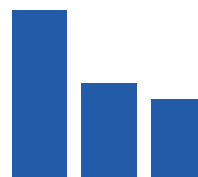No

Yes

Vertical or horizontal?

Ordered or sequential data

Categorical

Use light grey for reference lines

A	B	C
1	2	3
1	2	3
1	2	3

No

A	B	C
1	2	3
1	2	3
1	2	3

Yes

Right align numbers in tables.
Make sure decimal places are identical.
Always add commas to large numbers.

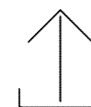| Left align (no)	Center (no)	Right align (yes)	Center (yes)
1234.56	123.45	1,235.56	1
123.4	123.4	123.40	2
23.45	23.45	23.45	3

Final thoughts

Perspective on data viz software

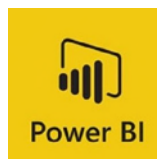**Main elements of data visualization software**

Data integration

Charting capability

Reporting (i.e. sharing)

BUY or DIY?

COTS

Code

No one size fits all... many things to consider

Default behaviors
- Axis colors
- Number formatting
- Color options
- Intelligent labeling

Features
- Interactivity
- Dashboards
- Data integration

Other considerations
- Accessibility to audience (i.e. reporting)
- Cost – organization dependent
- Ease of use – user dependent

Conclusions

- I hope you realize DV can be an effective skillset in your work.

- Start to pay attention to DV… yours and others.

- I have resources to get started (see next slide).

- Reach out if you have questions or want a second opinion.

Resources

Books

- 'Show me the numbers' book (data visualization bible)
- 'Story telling with data' book (learn to tell data stories)
- 'Information dashboard design' book (learn to design dashboards)
- 'The truthful art' book (data viz and communication)
- Leveraging data visualization techniques for Tecolote deliverables (IR&D)

Media

- http://www.visualisingdata.com/ (monthly digest of best data viz)
- http://www.storytellingwithdata.com/ (before/after examples)
- http://annkemery.com/ (before/after examples)
- https://policyviz.com/ (various topics)
- http://www.informationisbeautiful.net/ (examples of amazing design)
- https://excelcharts.com/ (various topics)
- http://flowingdata.com/ (various topics)
- https://eagereyes.org/ (various topics)