

# Designing Effective Visualizations

Yasmin AlNoamany

L&S 88-2

University of California, Berkeley

# Announcements

- Extra credit – 4 points each
  - [Find and critique existing visualization](#)
  - [Getting familiar with D3](#)
- Submitting your projects
  - Presentations on Nov. 28
  - Do you have something after the class?
  - Put all related files (e.g., code, data, presentation, report, all project milestones, and any other related files) to the project in a shareable place such as the university Box, Dropbox, bDrive, or Github (1 extra point).
  - One member of the group emails me a link to the shared folder that contains all the related files to the project.

# Resources

1. [Effectively Communicating Numbers](#), S. Few
  2. [Table and Graph Design for Enlightening Communication](#), S. Few
  3. [Data Visualization Principles](#): Lessons from Tufte
- 
- Optional:
    - [The Seven Stages of Visualizing Data](#)
    - In Defense of Simplicity, A Data Visualization Journey – [blog](#)
    - [6 charts you will see in hell](#)
    - [A Tour through the Visualization Zoo](#), Heers

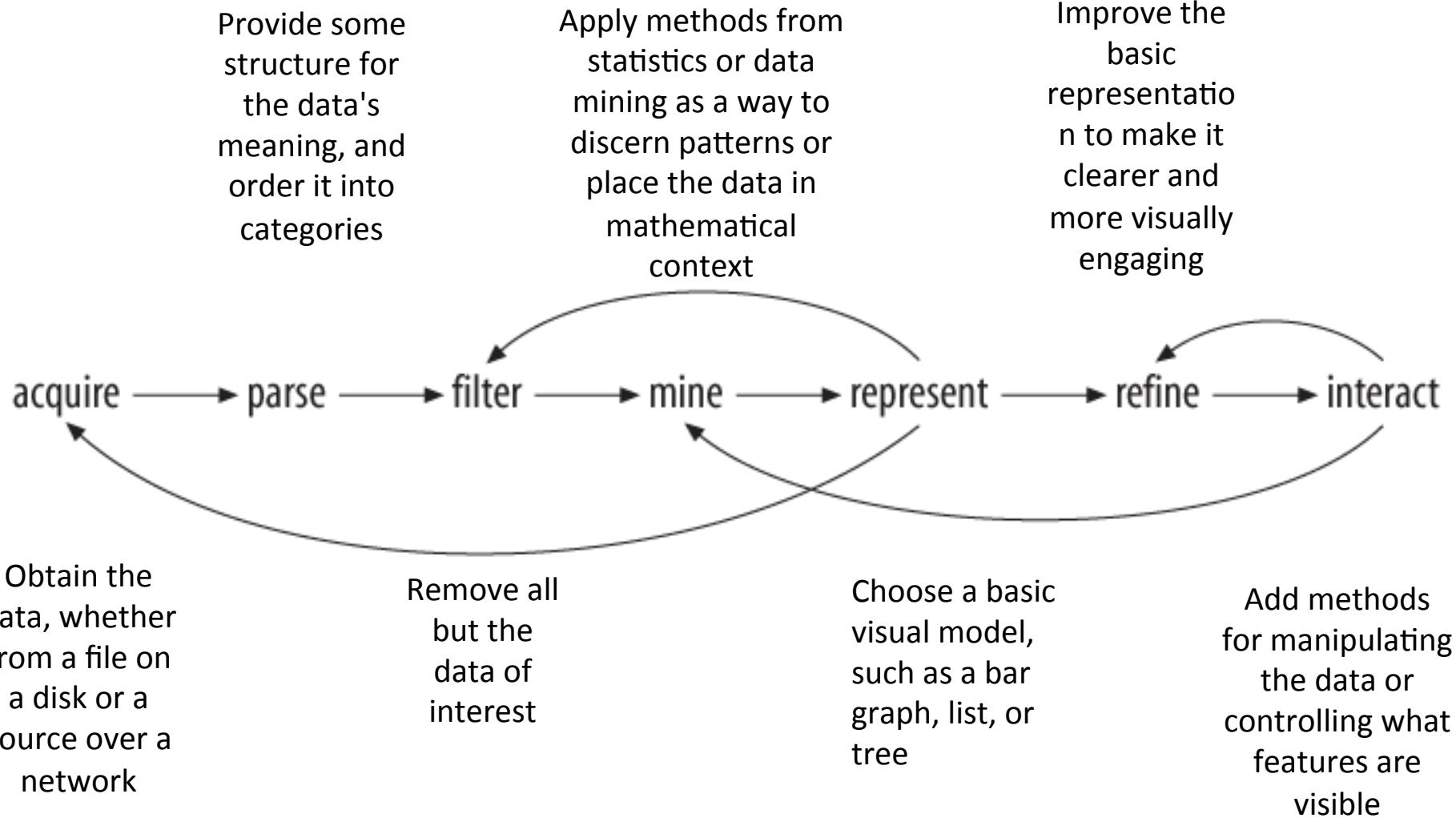
# Outlines

- Data Visualization Process
- How to generate effective visualizations
  - Choosing the right graph
  - Data visualization best practices
- What you should NOT do
- Poor graphs can be transformed

# Objectives

- Identify basic steps of data visualization process
- Distinguish among different types of datasets
- Distinguish among different types of relationships among data values
- Identify multiple data visualization principles
- Determine how to use colors wisely
- Know what not to do in designing effective visualizations
- Turn bad visualizations into good visualizations

# Visualization process 7 stages



# How to generate effective visualizations

# Designing a graph

- Must keep the task in mind
  - why do you need a graph?
  - what questions are being answered?
  - what data is needed to answer those questions?
  - who is the audience?

# Questions to assess the quality of visualization

Welcome to **Visual Trumpery**. Thanks so much for being here. I foresee that, depending on your political leanings, your expectations for this talk are diametrically opposed.

If you are a liberal, you're likely prepared to enjoy a tirade against right-wing politicians. If you are a conservative, I bet you came here ready to get outraged by an unhinged leftist ranter.

I am indeed a ranter, and I love a good tirade, but I'm also going to disappoint you. The very title of this talk is an example of its content. It's intended to prime you and trick you into believing what is not so.

The English word "**trumpery**" means worthless nonsense, something that is showy and deceitful at the same time. Trumpery can occur in text, verbally, or visually. This non-partisan talk focuses on the visual, examining misleading charts, graphs, and data maps designed by individuals and organizations from all over the political spectrum.

I'll use these examples to equip you with a solid understanding of "**graphicacy**," the word I use to refer to visual literacy. I believe a literate, numerate, and graphicate citizenry is the best antidote for a world where trumpery runs rampant.

Join me in the fight against it.

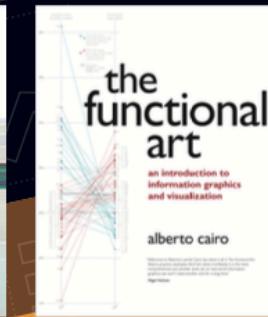
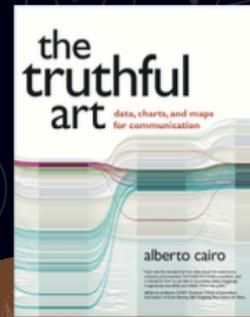
**SUMMARY OF THE TALK**

Here are the four questions I'll mention during the talk to assess the quality of any data visualization:

- 1. Is the graphic based on reliable sources and data?**  
Any document or news story that uses data and graphics ought to clearly identify its sources and link to them. You, the reader, must be able to check whether they are trustworthy, and whether writers and designers handled the data properly.
- 2. Does the graphic include enough information to be truthful?**  
Visualizations should never *simplify* information. They ought to *clarify* it. You'll soon understand the difference, thanks to the examples I'll share with you.
- 3. Is the data correctly represented?**  
Data visualization is based on visual encoding. Numbers are mapped onto spatial properties of objects, like their height, length, size, color, etc. When seeing a chart, graph, or map, always ask yourself: Are the properties representing the data proportional to the data itself?
- 4. Did the journalist or designer take uncertainty into account?**  
Data is hardly ever precise or certain, regardless of what visualizations and news stories often suggest. A good understanding of uncertainty and elementary probability can help you decide what to think or how to act based on what you read.

**ADDITIONAL READINGS**

It's impossible to teach all about graphicacy in just 90 minutes. That's why I created a repository of readings: <http://goo.gl/JwmUcW>  
Finally, if you're interested in my books, I recommend that you first read *The Truthful Art*. A good part of it is related to the Visual Trumpery talk:



# Initial steps for generating good visualizations

- Choosing the right graph
  - Identify your data
  - Determine the best means to encode the values
- Design your visualization to tell a story clearly

# Types of data

- Univariate data
- Bivariate data
- Trivariate data
- Hypervariate data

# Univariate data

## What Goes Where

In univariate representations, we often think of the data *case* as being shown along one dimension, and the *value* in another

	Case <sub>1</sub>	Case <sub>2</sub>	Case <sub>3</sub>	...
Variable <sub>1</sub>	Value <sub>11</sub>	Value <sub>21</sub>	Value <sub>31</sub>	
Variable <sub>2</sub>	Value <sub>12</sub>	Value <sub>22</sub>	Value <sub>32</sub>	
Variable <sub>3</sub>	Value <sub>13</sub>	Value <sub>23</sub>	Value <sub>33</sub>	
...				

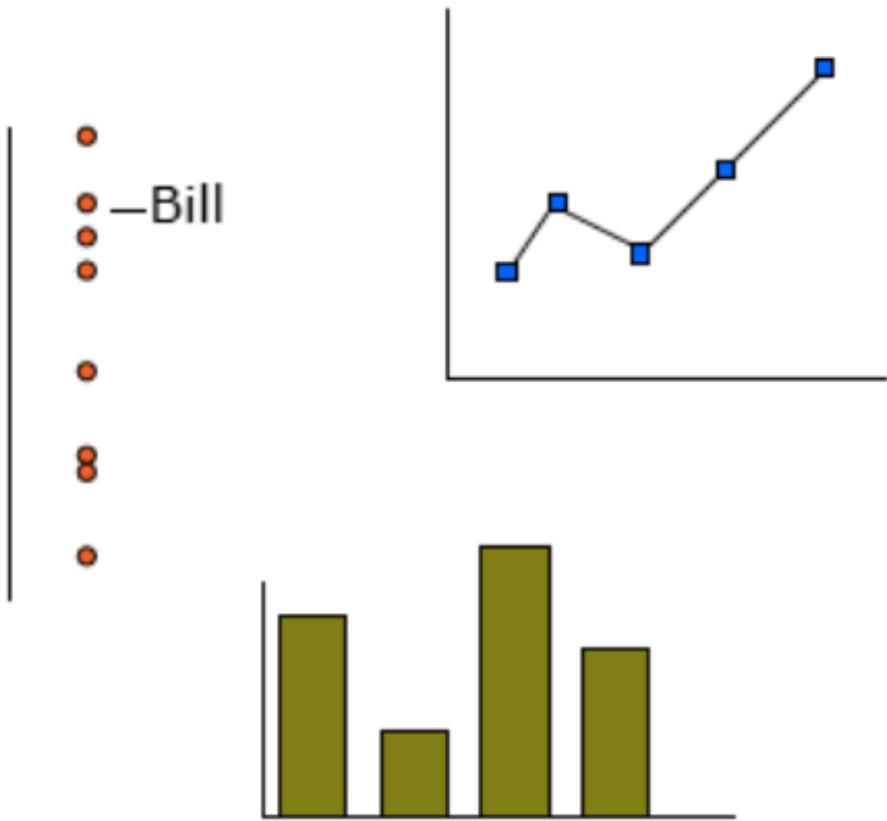
Line graph



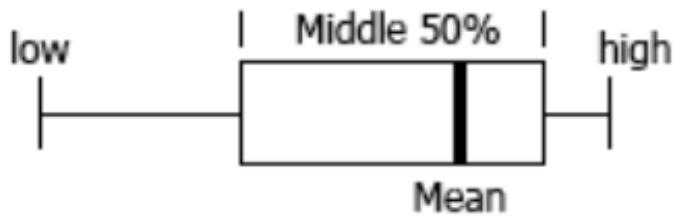
Y-axis is quantitative variable

See changes over consecutive values

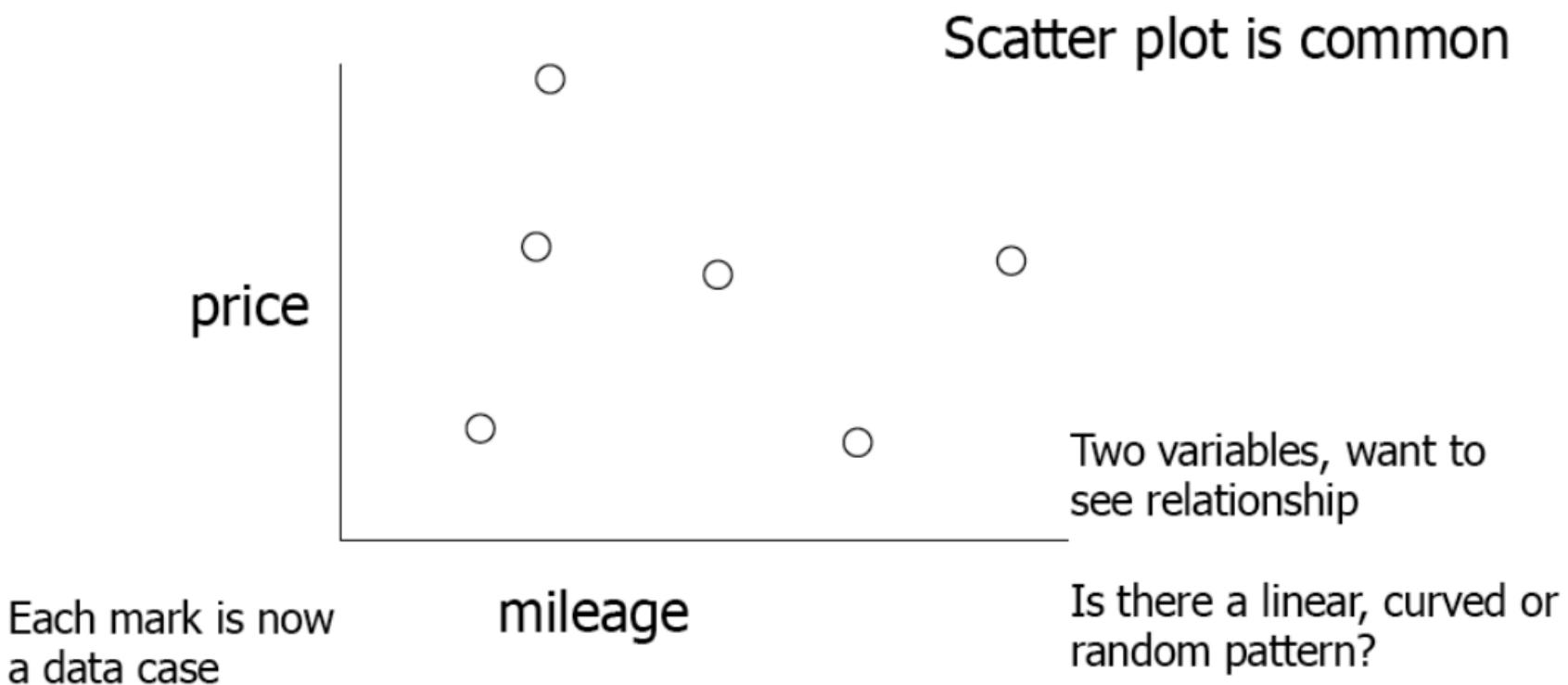
# Univariate data



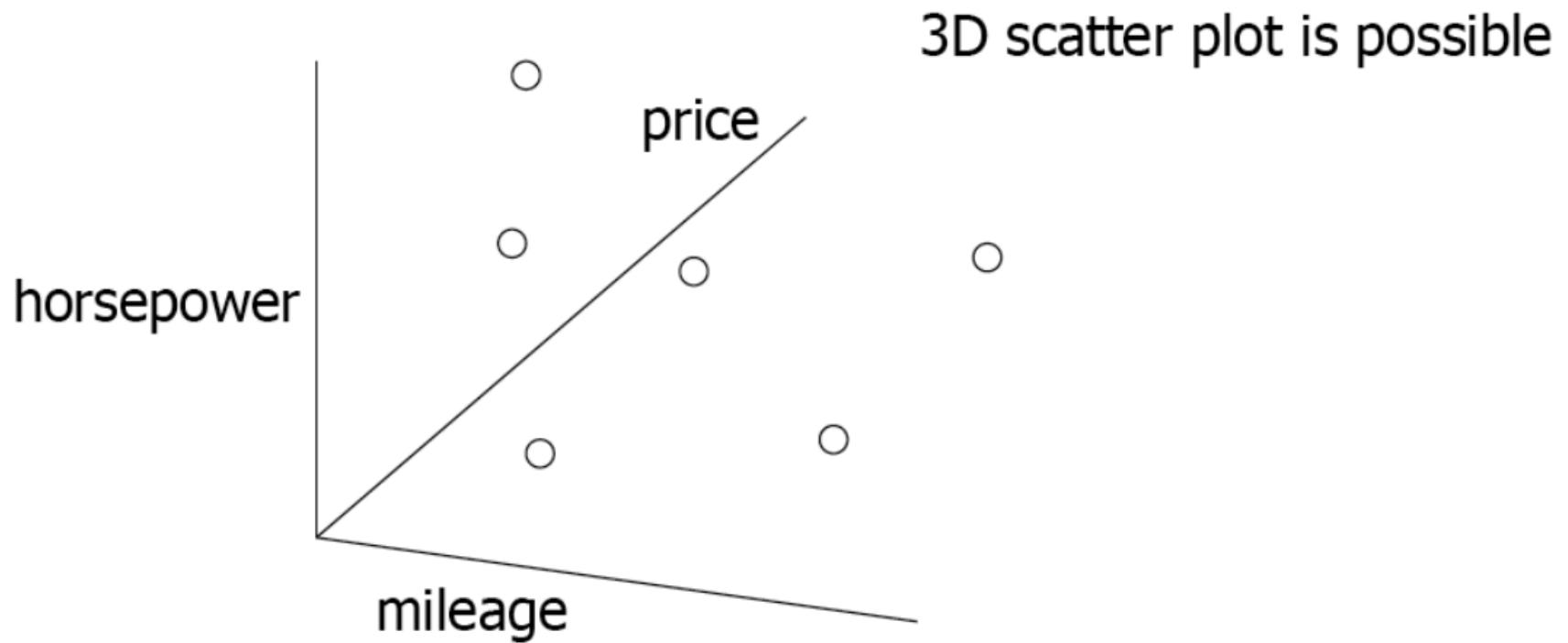
Tukey box plot



# Bivariate data



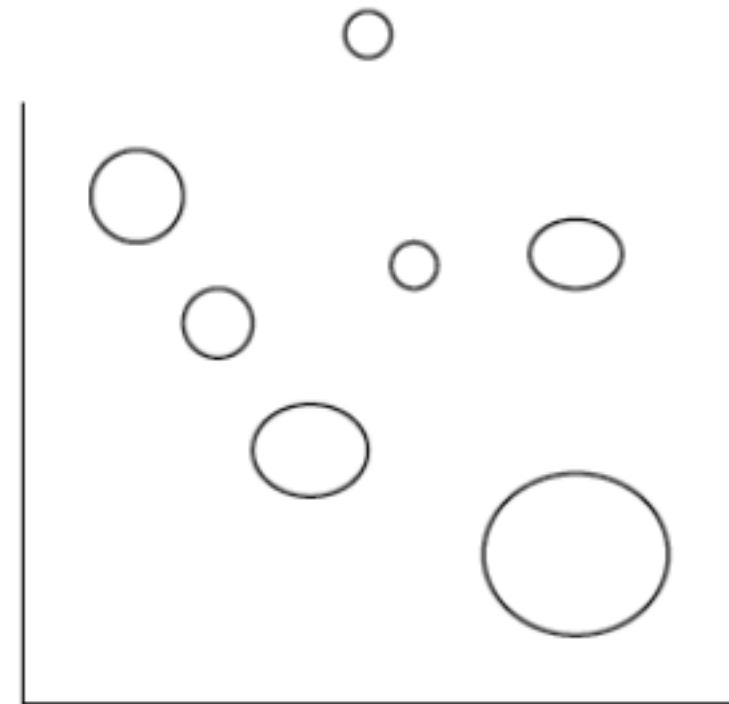
# Trivariate data



# Trivariate data

## Alternative Representation

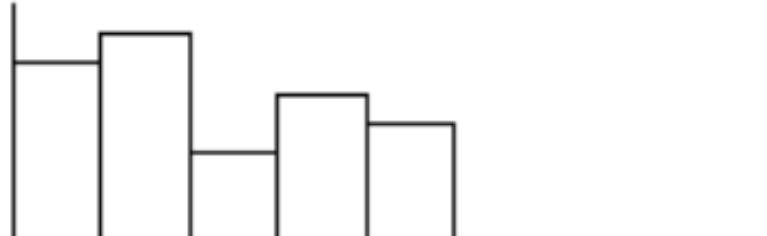
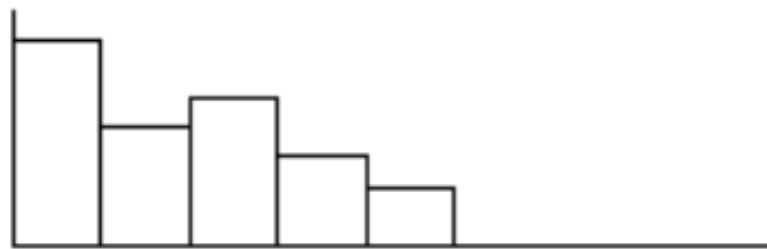
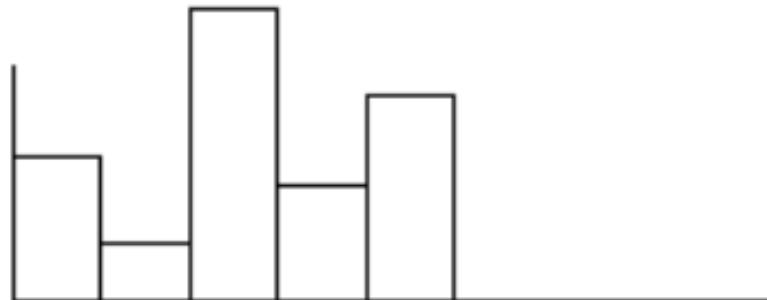
Still use 2D but have  
mark property (size,  
color, etc.) represent  
third variable



# Trivariate data

## Alternative Representation

- Represent each variable in its own explicit way



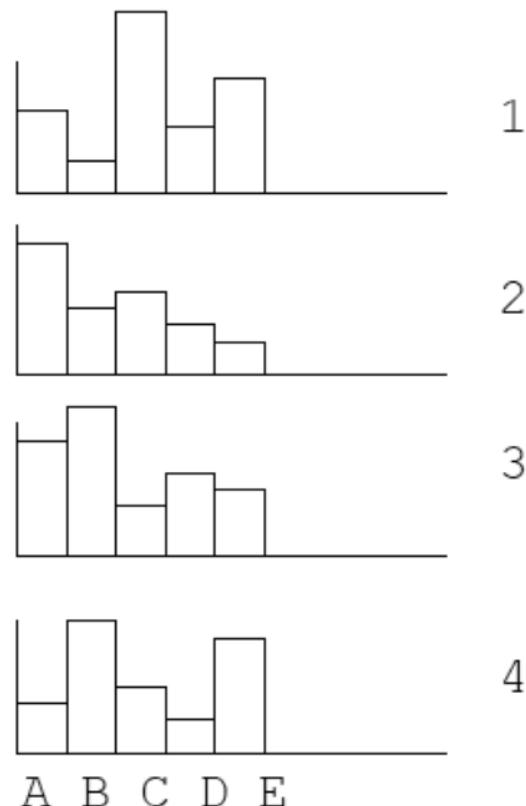
# Hypervariate Data

- What about data sets with more than 3 variables?
  - Often the interesting, challenging ones
  - line graphs, bar graphs, scatter plots
  - small multiples

# Hypervariate Data

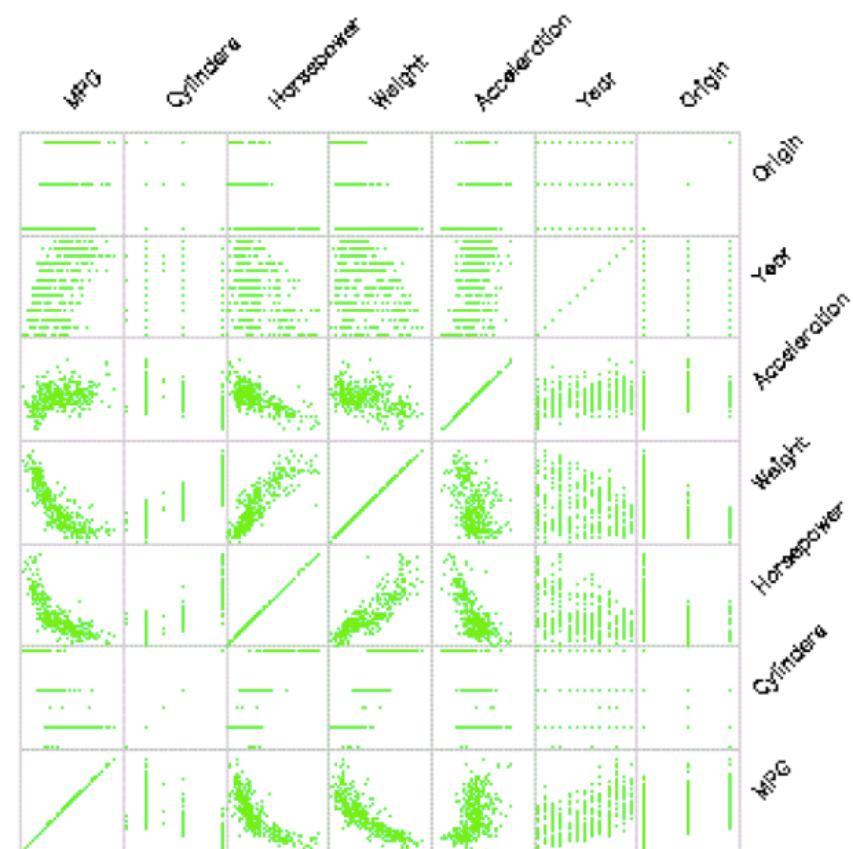
- Multiple Views
  - Give each variable its own display

	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5

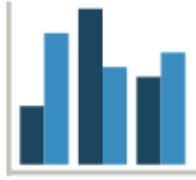


# Hypervariate Data Scatterplot Matrix

Represent each  
possible pair of  
variables in their  
own 2D scatterplot



# Determine the best means to encode the values



## COMPARE ITEMS

Compare values across different categories.



## STATIC COMPOSITION

Part-to-whole relationship.



## CORRELATION

Relationship between two or more variables.



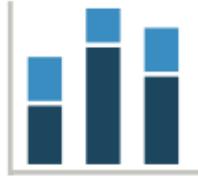
## LOCATION

Show where things happen.



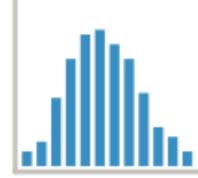
## TIME COMPARISON

Compare changes over a period of time.



## NON-STATIC COMPOSITION

Part-to-whole relationship over time.



## DISTRIBUTION

Frequency of values in a data set.



## KPIs

Key Performance Indicators.

# Seven relationships between values

Nominal comparison

Time-series

Ranking

Part-to-whole

Deviation

Frequency distribution

Correlation

# Seven relationships between values

Nominal comparison

Time-series

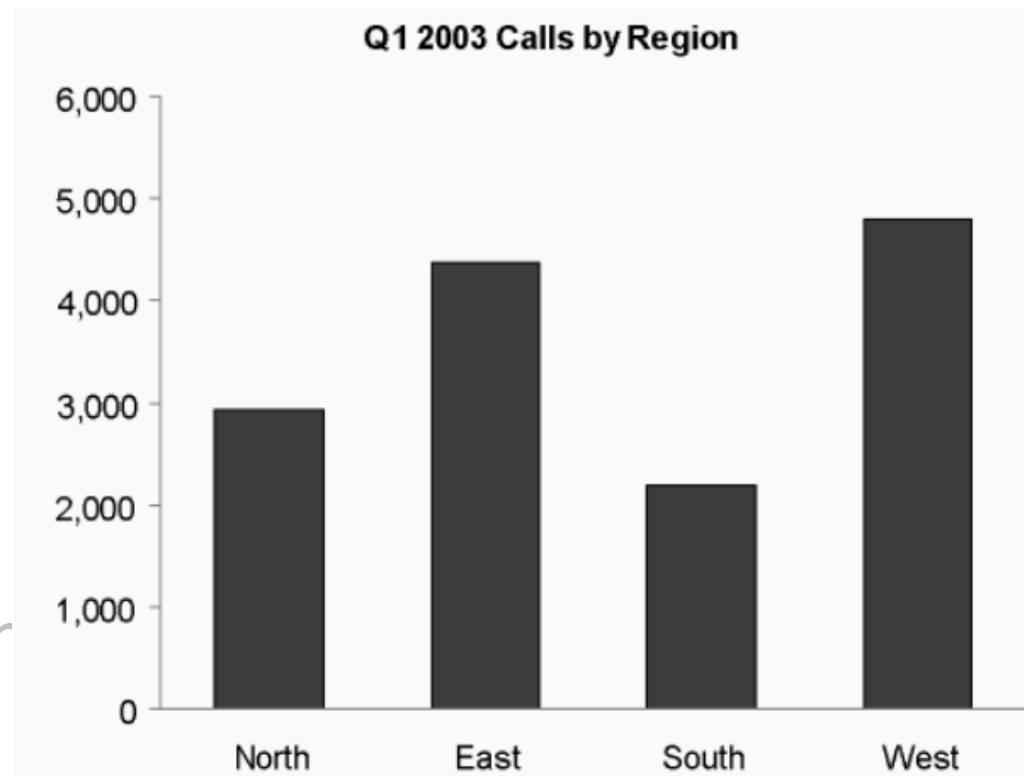
Ranking

Part-to-whole

Deviation

Frequency distribution

Correlation



# Seven relationships between values

Nominal comparison

Time-series

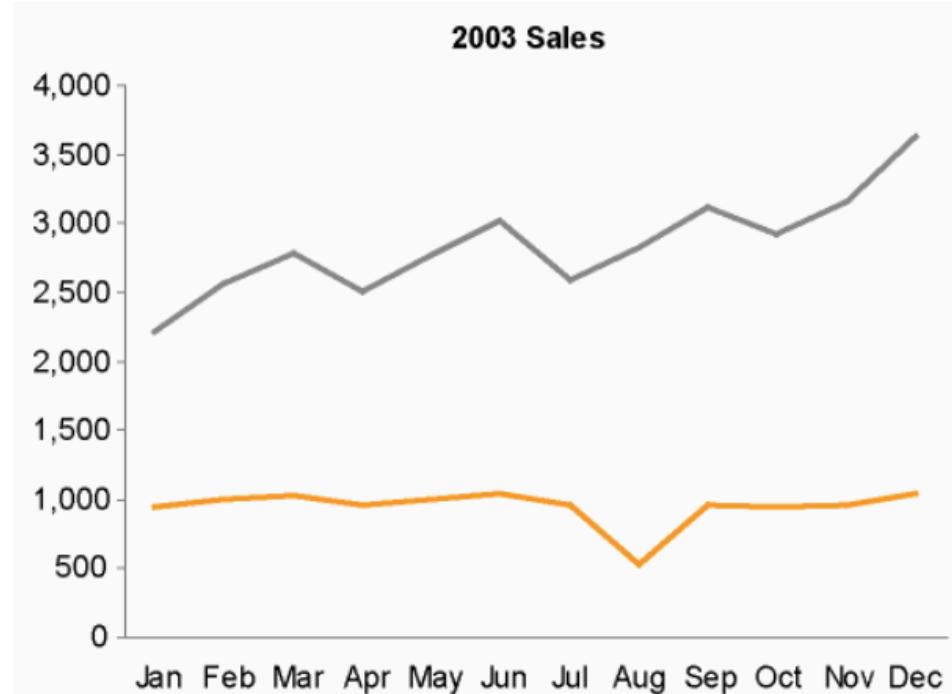
Ranking

Part-to-whole

Deviation

Frequency distribution

Correlation



# Seven relationships between values

Nominal comparison

Time-series

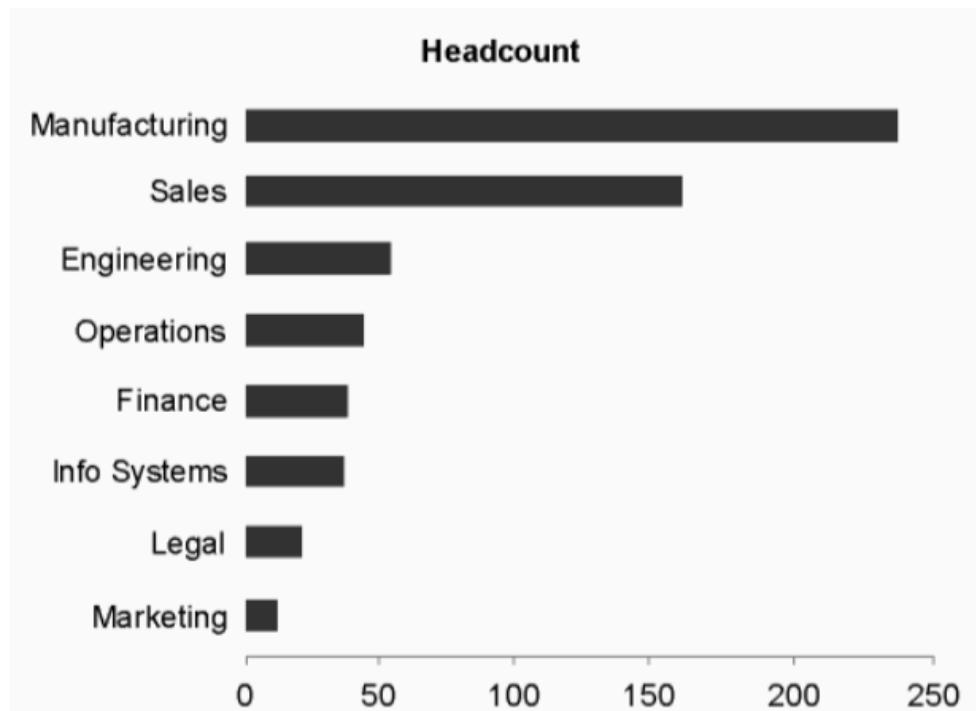
Ranking

Part-to-whole

Deviation

Frequency distribution

Correlation



# Seven relationships between values

Nominal comparison

Time-series

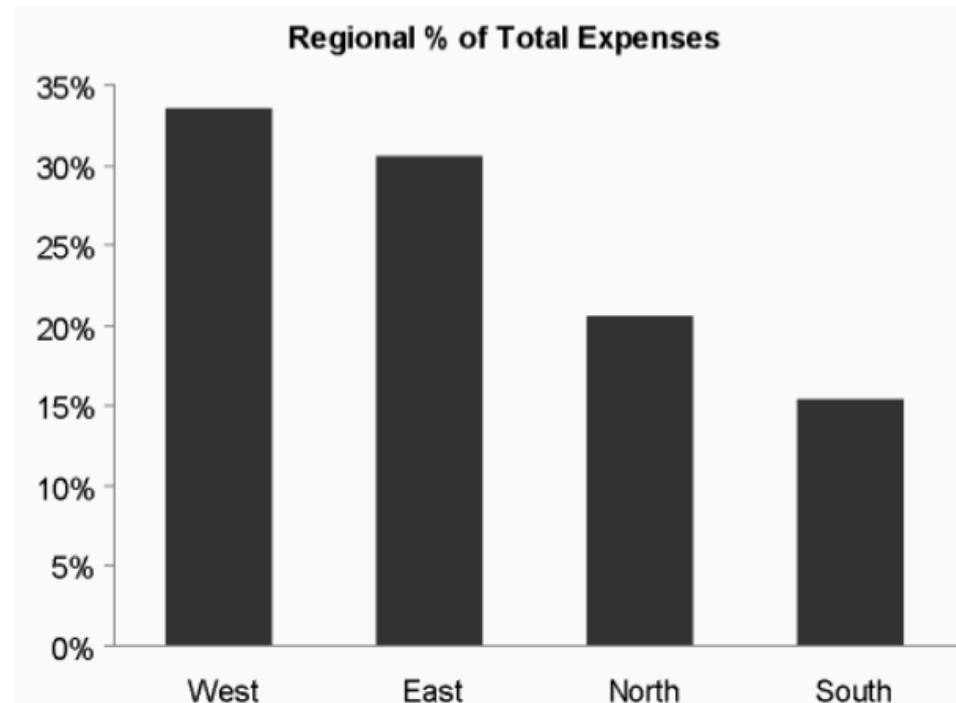
Ranking

Part-to-whole

Deviation

Frequency distribution

Correlation



# Seven relationships between values

Nominal comparison

Time-series

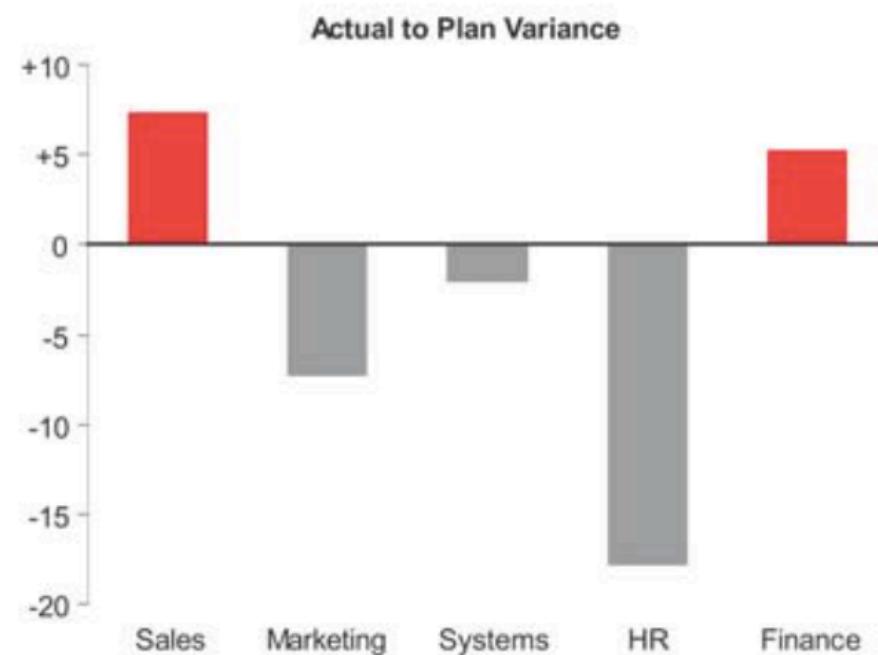
Ranking

Part-to-whole

**Deviation**

Frequency distribution

Correlation



# Seven relationships between values

Nominal comparison

Time-series

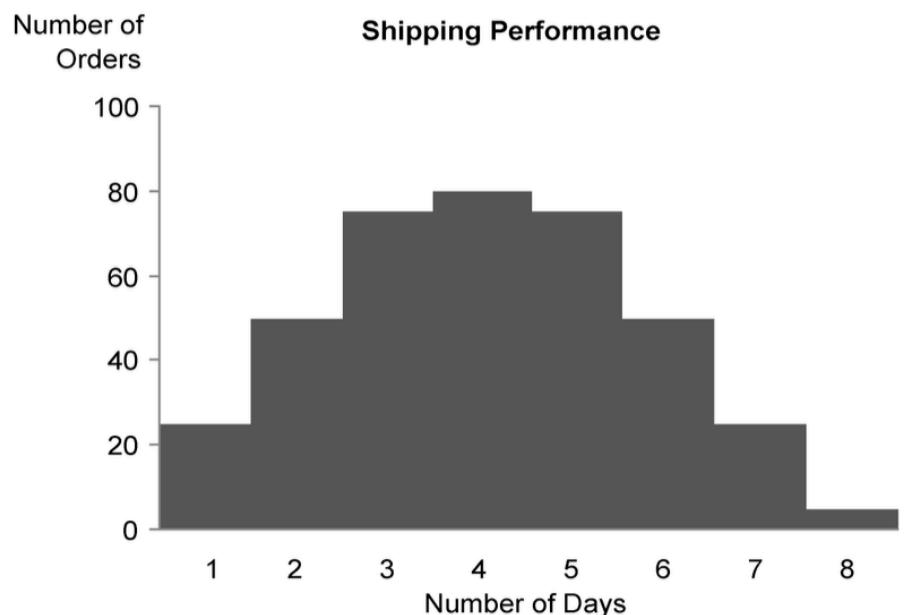
Ranking

Part-to-whole

Deviation

**Frequency distribution**

Correlation



# Seven relationships between values

Nominal comparison

Time-series

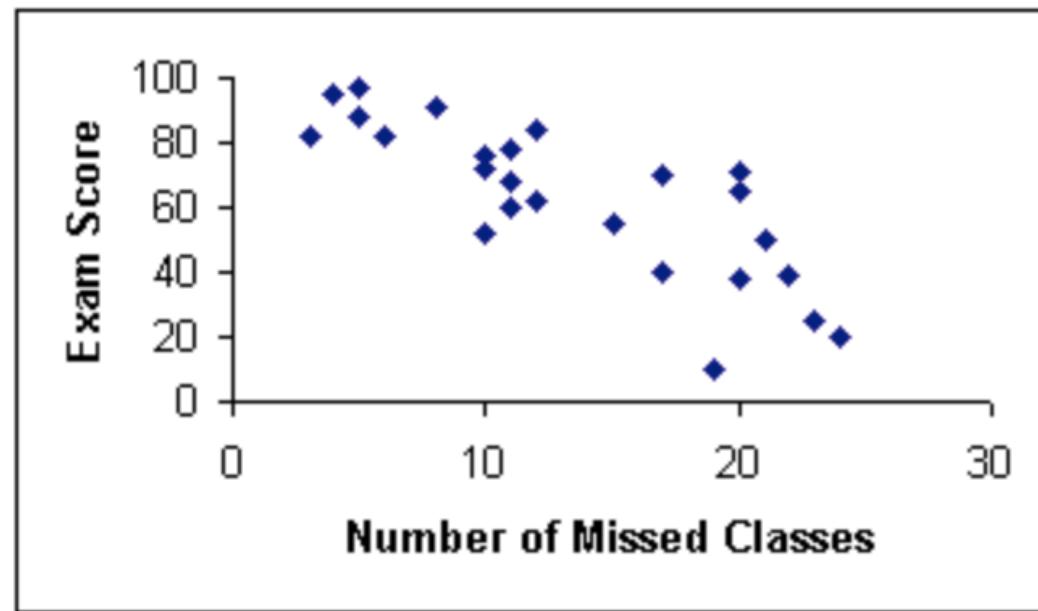
Ranking

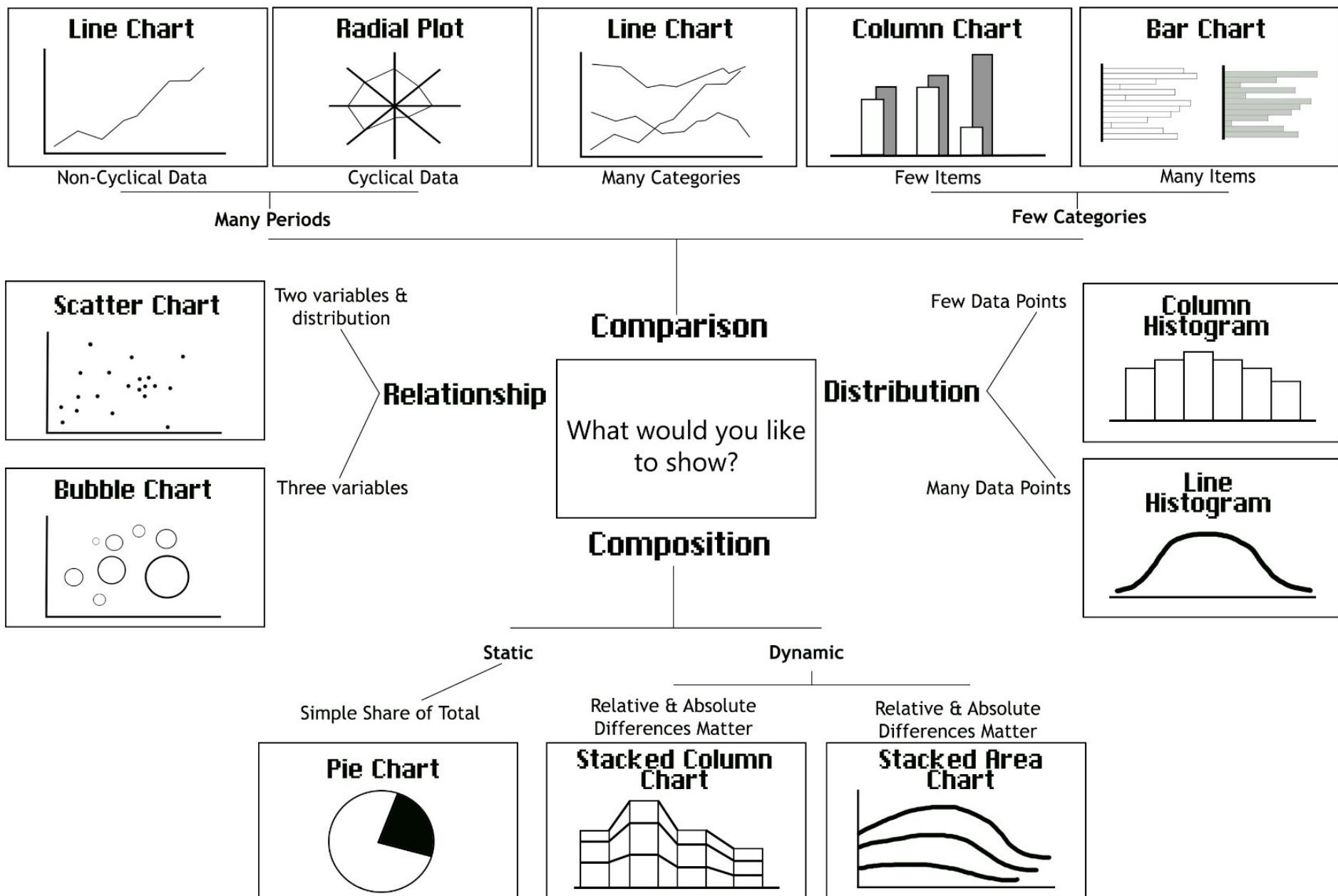
Part-to-whole

Deviation

Frequency distribution

Correlation

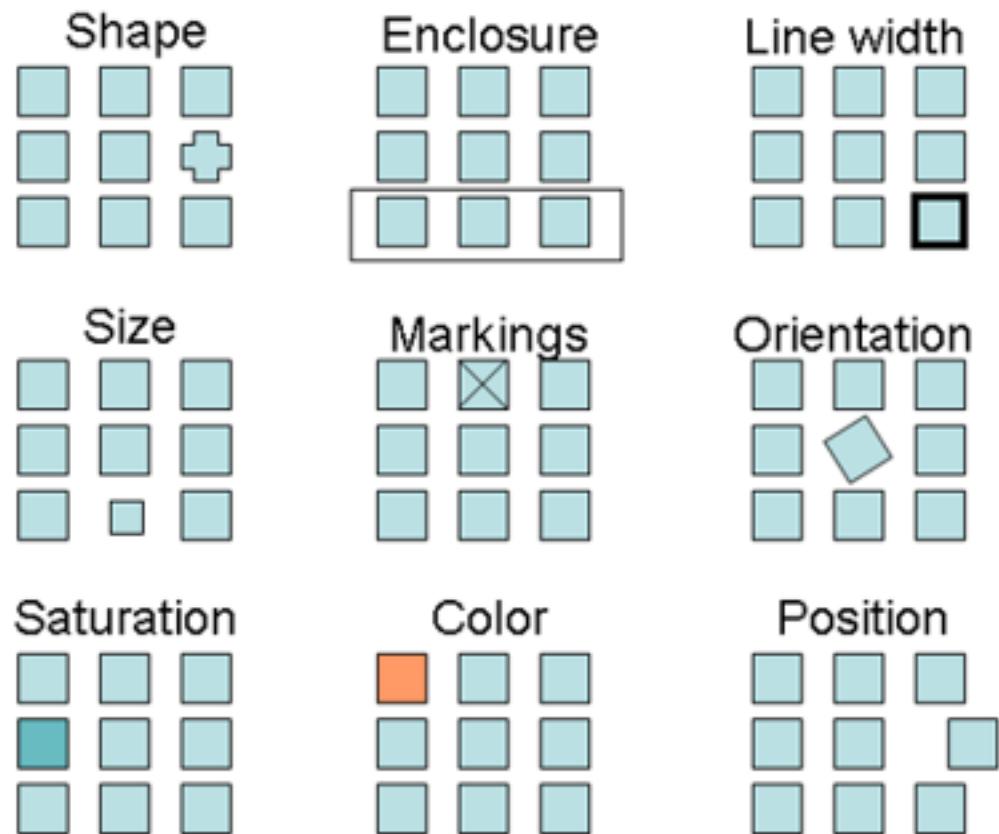




# Data Visualization Best Practices

# Use visual variables for pre-attentive processing

1. Shape
2. Size
3. Saturation
4. Enclosure
5. Marking
6. Color
7. Line Width
8. Orientation
9. Position



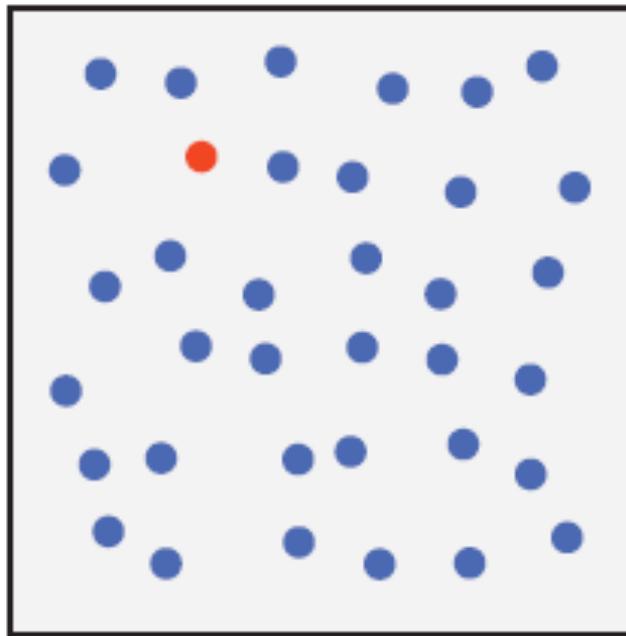
# How many 3s?

128176875613897654698450698560498282676  
298098584582245098564589450984509809435  
8590910302099059595772564675050678904  
567884578980982167765487636490856091294  
9686

# How many 3s?

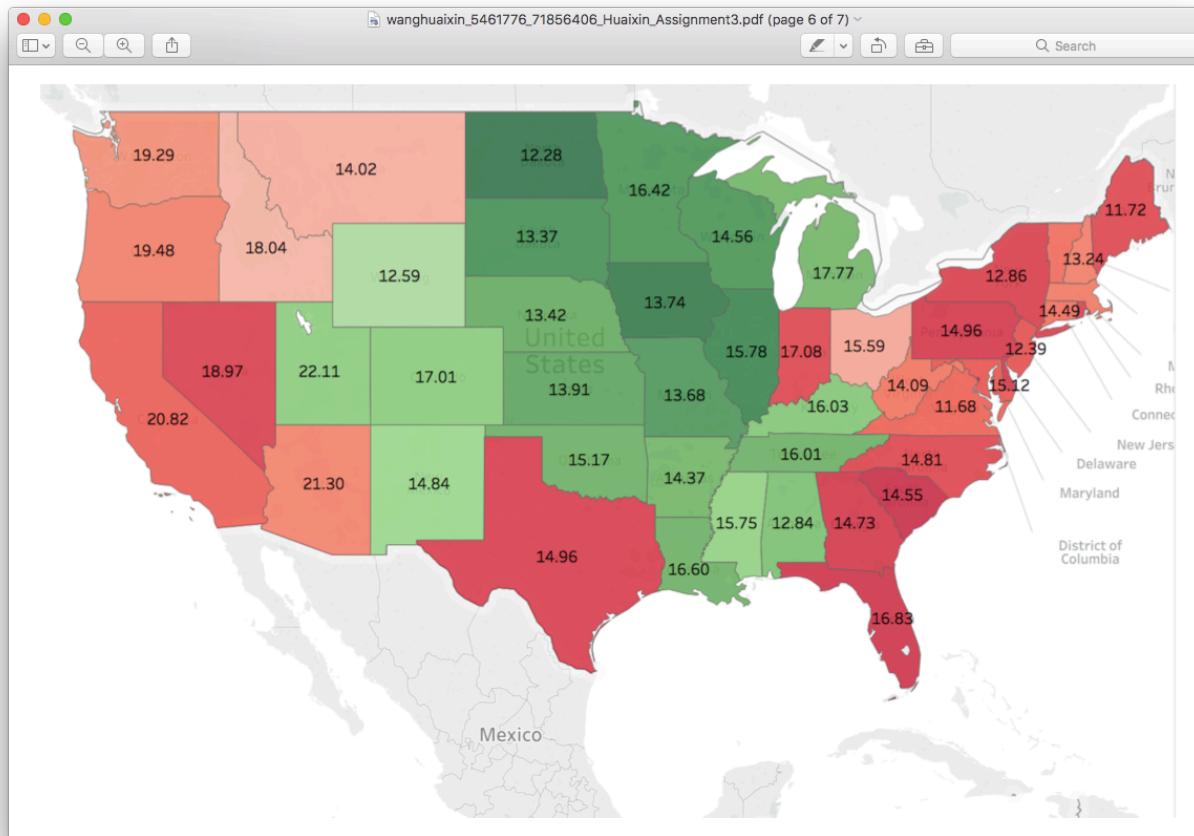
12817687561**3**897654698450698560498282676  
2980985845822450985645894509845098094**35**  
8590910**3**02099059595772564675050678904  
5678845789809821677654876**3**6490856091294  
9686

# Can you find the red dot?



Source: <https://www.csc2.ncsu.edu/faculty/healey/PP/index.html>

# Do you remember Assignment 3?



# Colors can give your charts more meaning



CATEGORICAL

Every category is labeled with a different color.



SEQUENTIAL

Single color represents a metric ordered low to high.

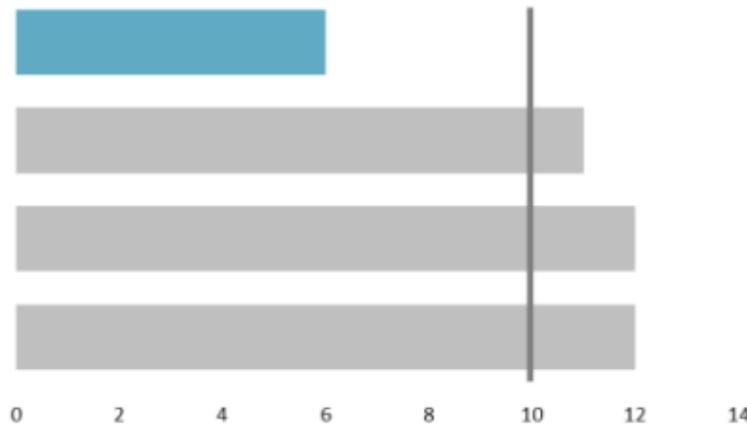


DIVERGING

Two color schemes with critical midpoint value.

- Use colors to highlight different categories or metrics
- Design a chart with a color blind friendly palette

# Use color meaningfully and with restraint



Add emphasis

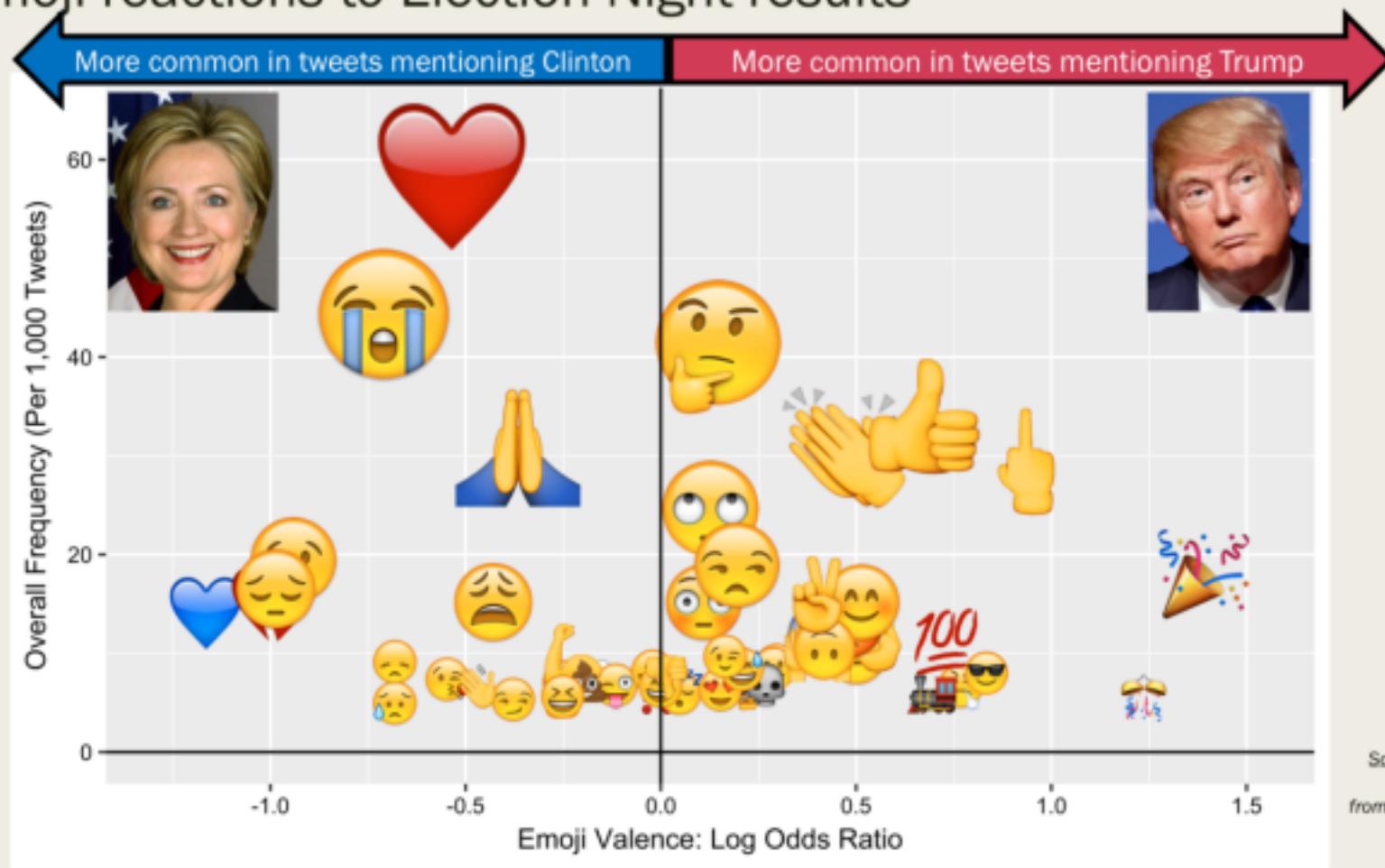


Create the confusing effect of Skittles on a page

# Good use of color

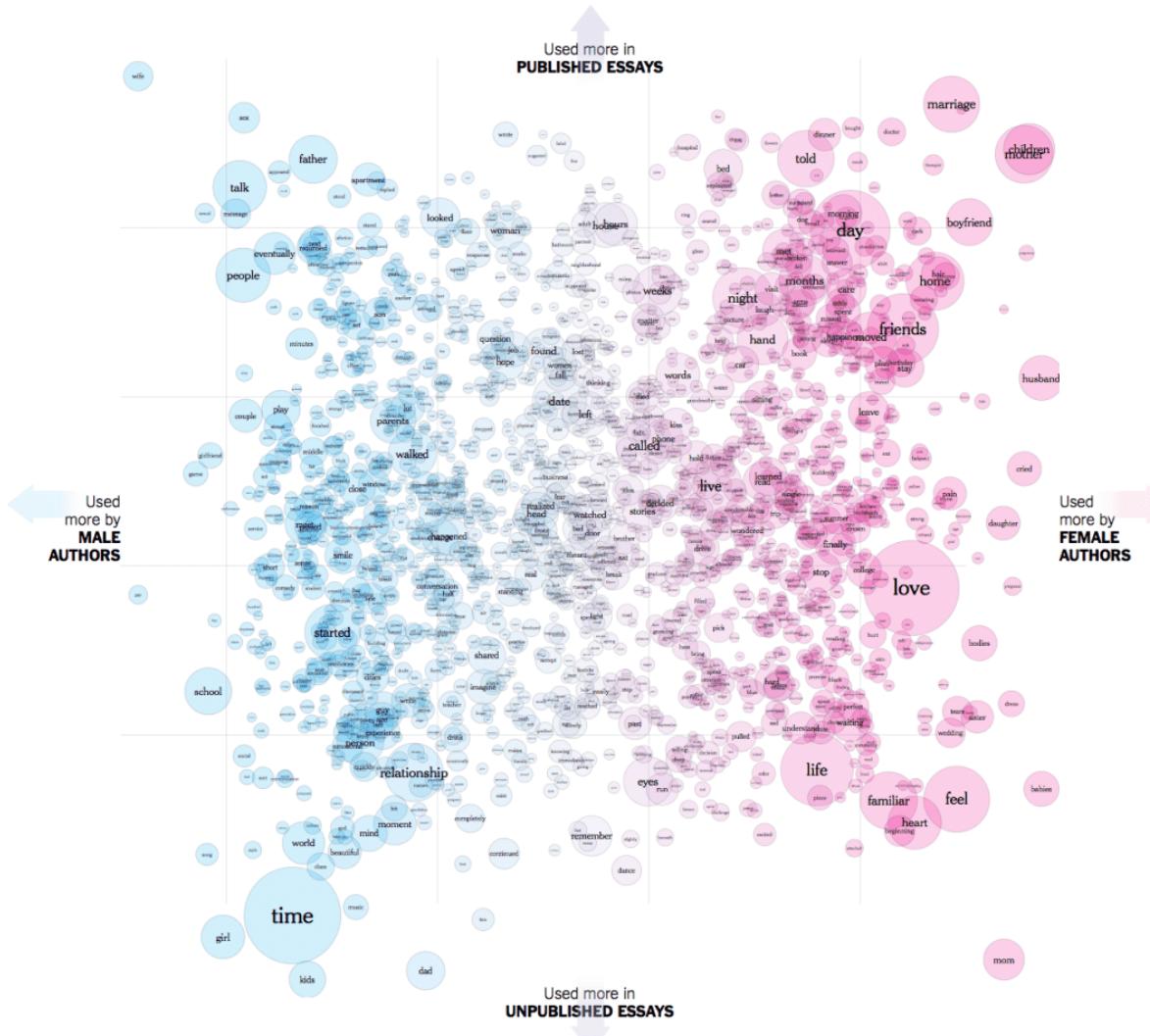
## The top emojis of Election Day 2016

## Emoji reactions to Election Night results

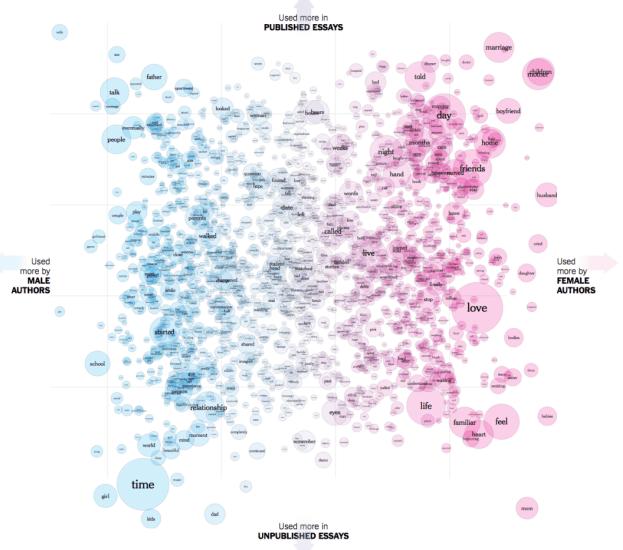


Source: Analysis of sampled  
Election Night tweets,  
from Nov. 9, 2016 12:00am –  
3:30am EST.  
(PRISMOJI)

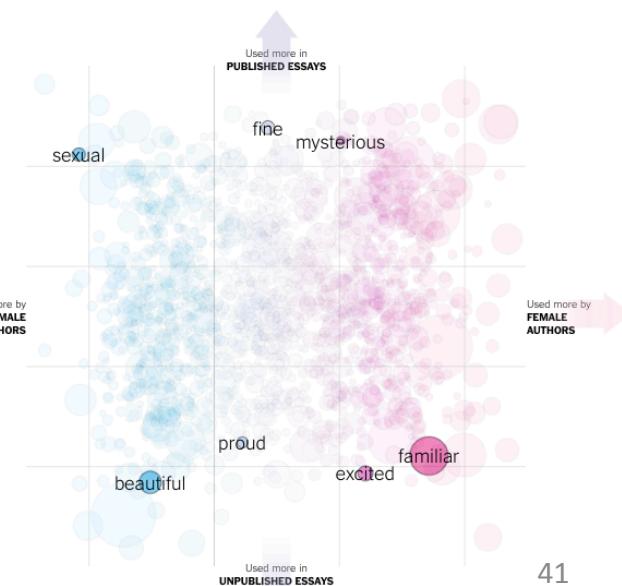
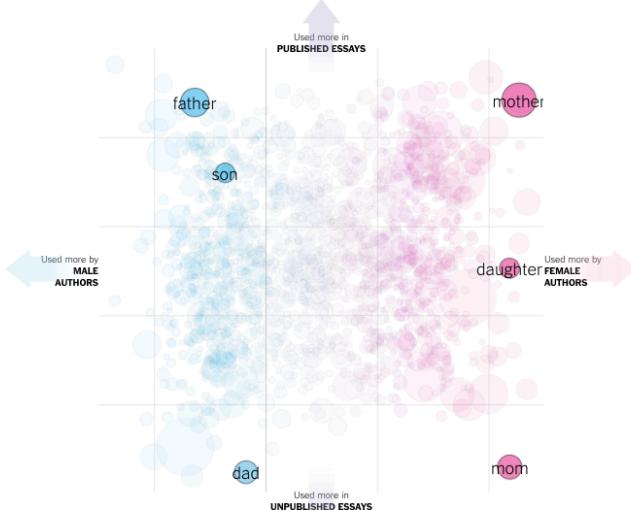
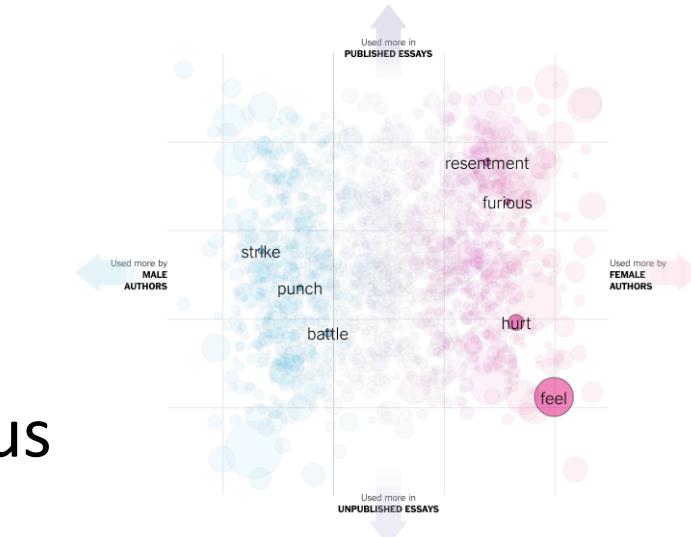
# Words men and women use when they write about love



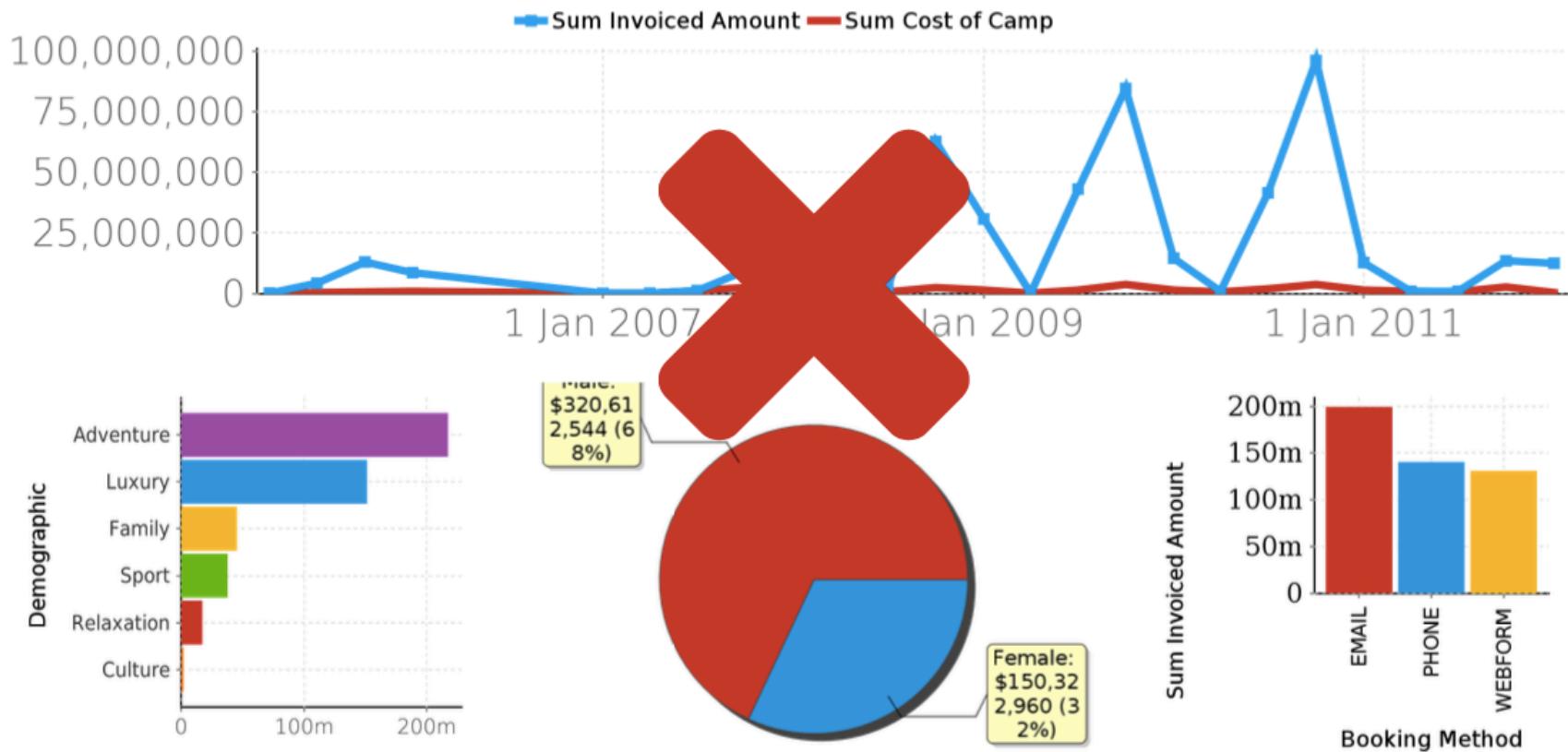
# Always be consistent



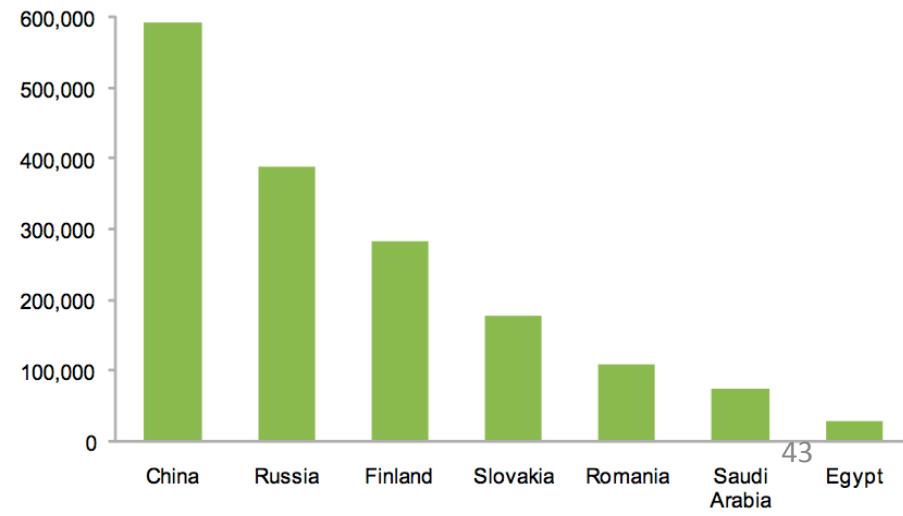
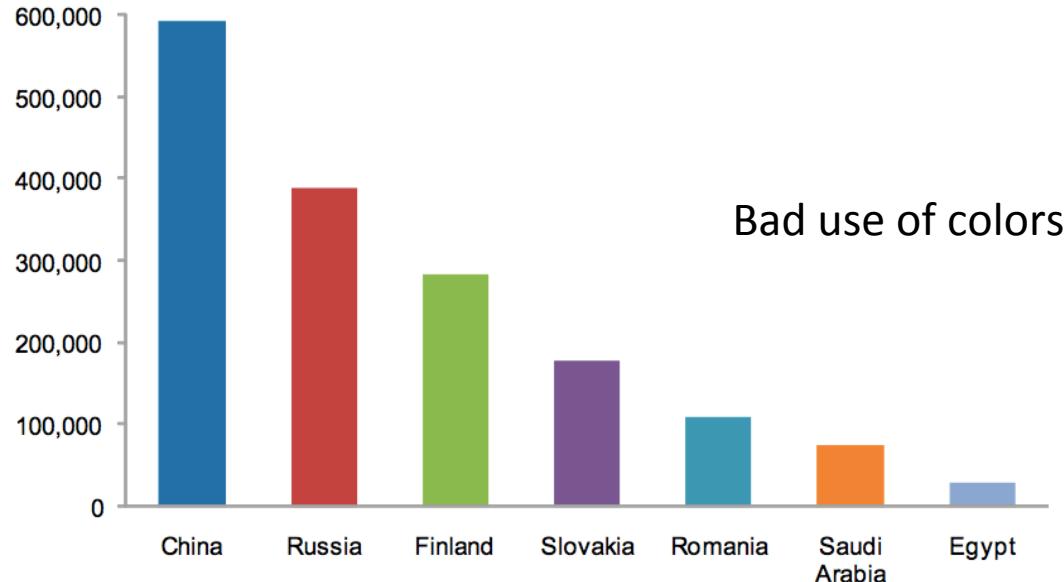
# men versus women



# Always be consistent



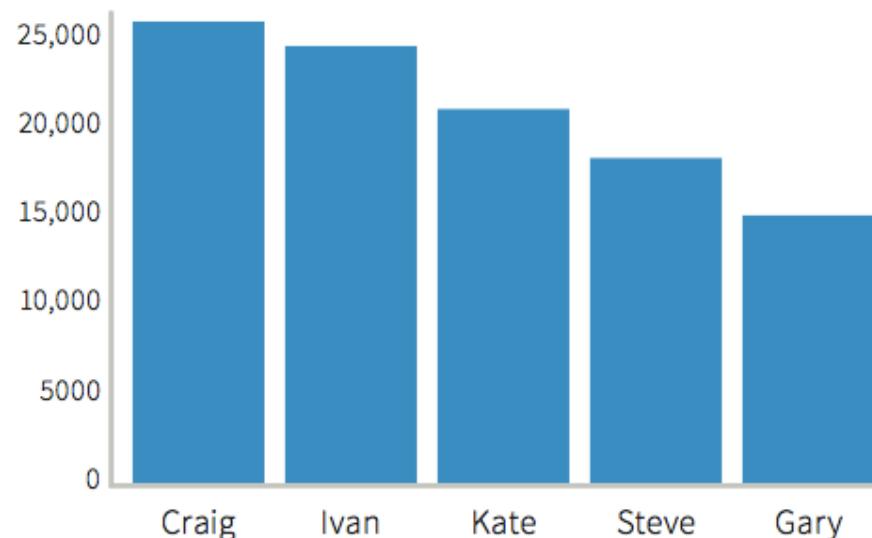
# Use color meaningfully and with restraint



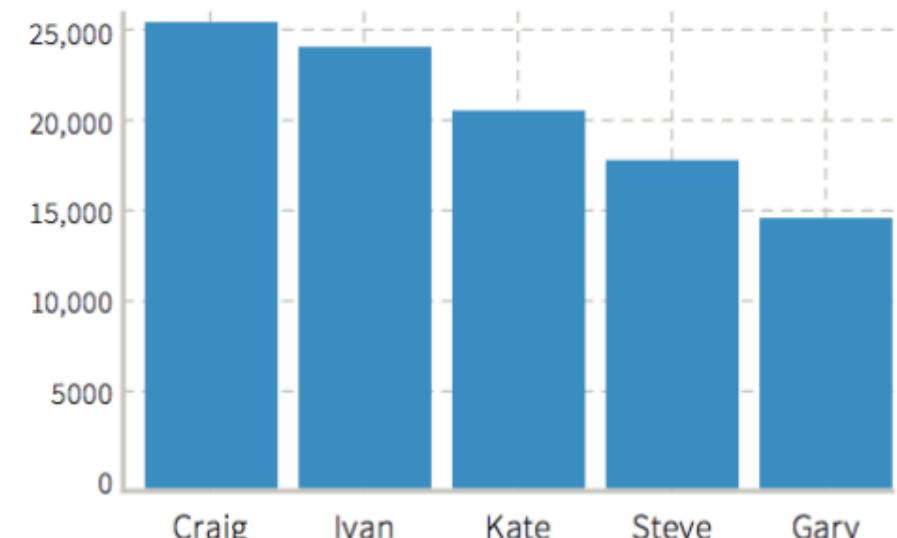
# Add context

## Grid lines communicate key thresholds

NO GRID LINES



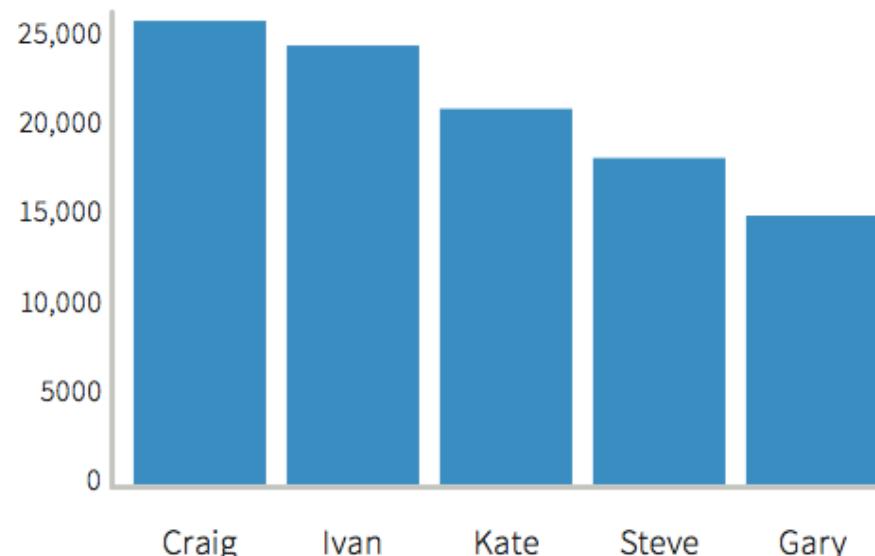
GRID LINES



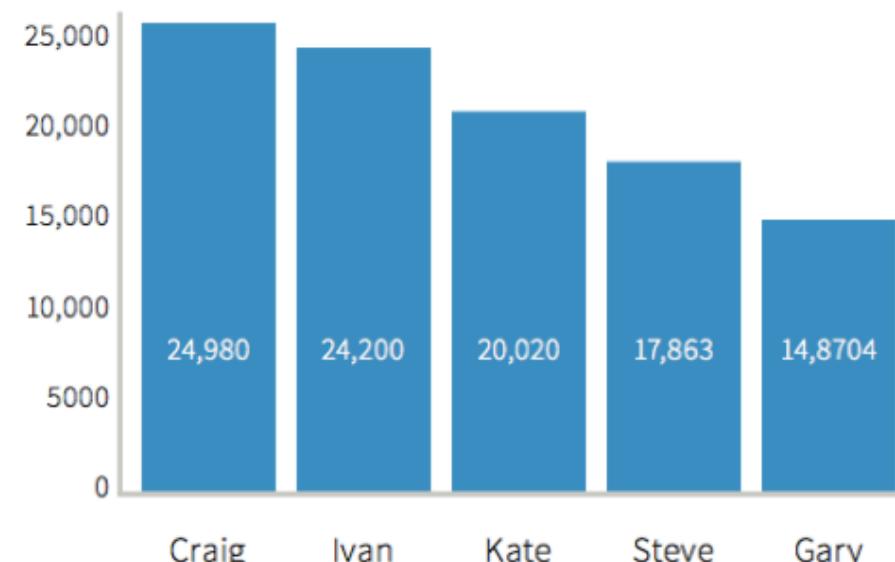
# Add context

## Labels reveal exact values

NO LABELS



LABELS

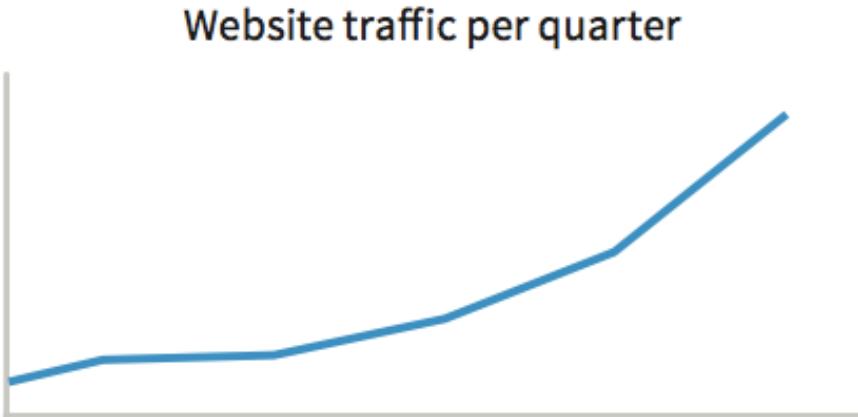


# Add context

## Chart titles frame the story in your data

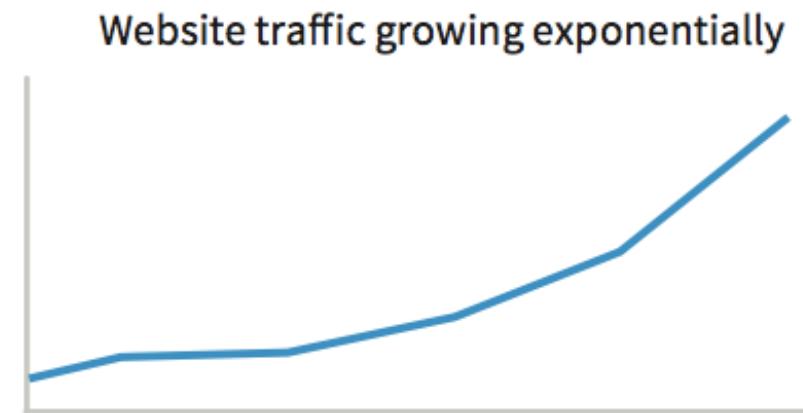
### DESCRIBE THE QUERY

For monitoring data without bias.

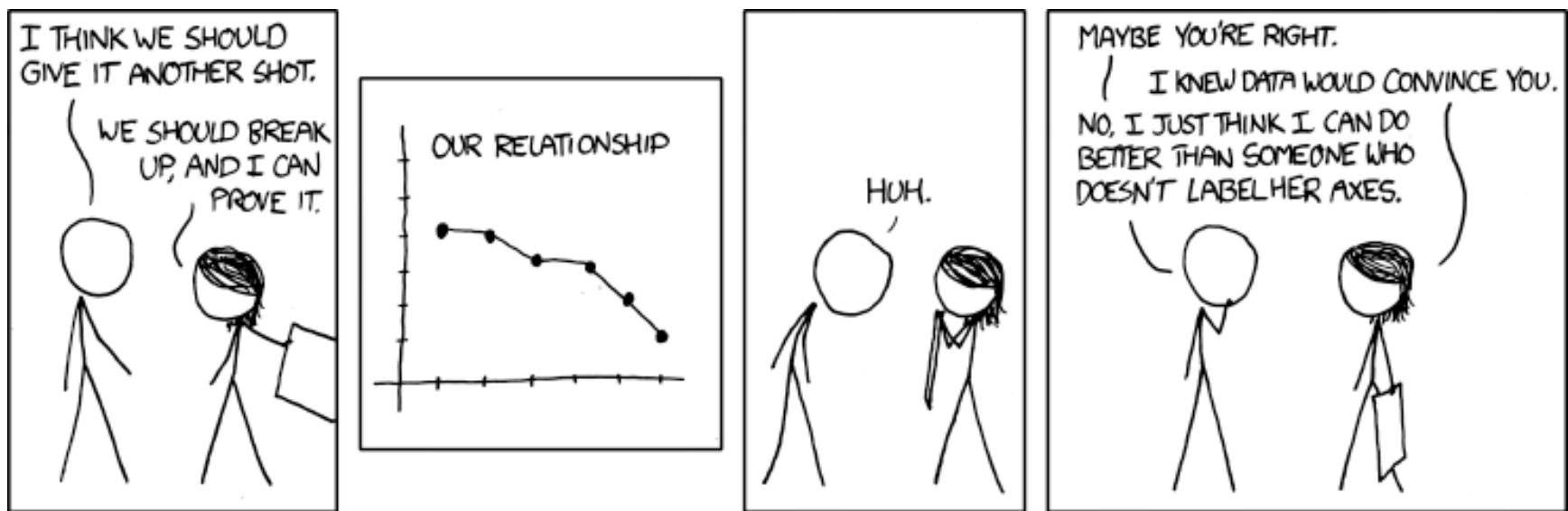


### EXPLAIN THE INSIGHT

Use data to tell a story.

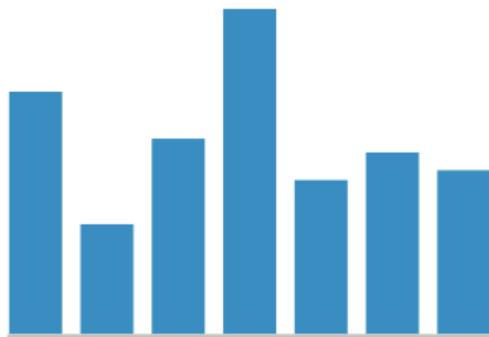


# Axis Labels



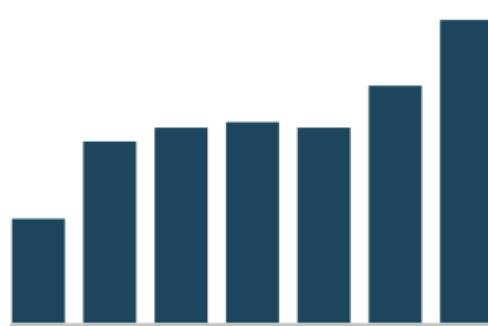
<https://xkcd.com/833/>

# Sort your data to make finding insights easier



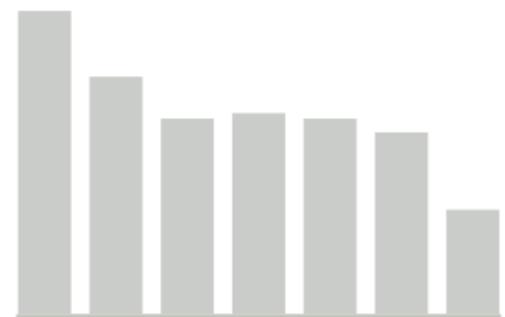
ALPHABETICAL

Quickly find a category.



ASCENDING

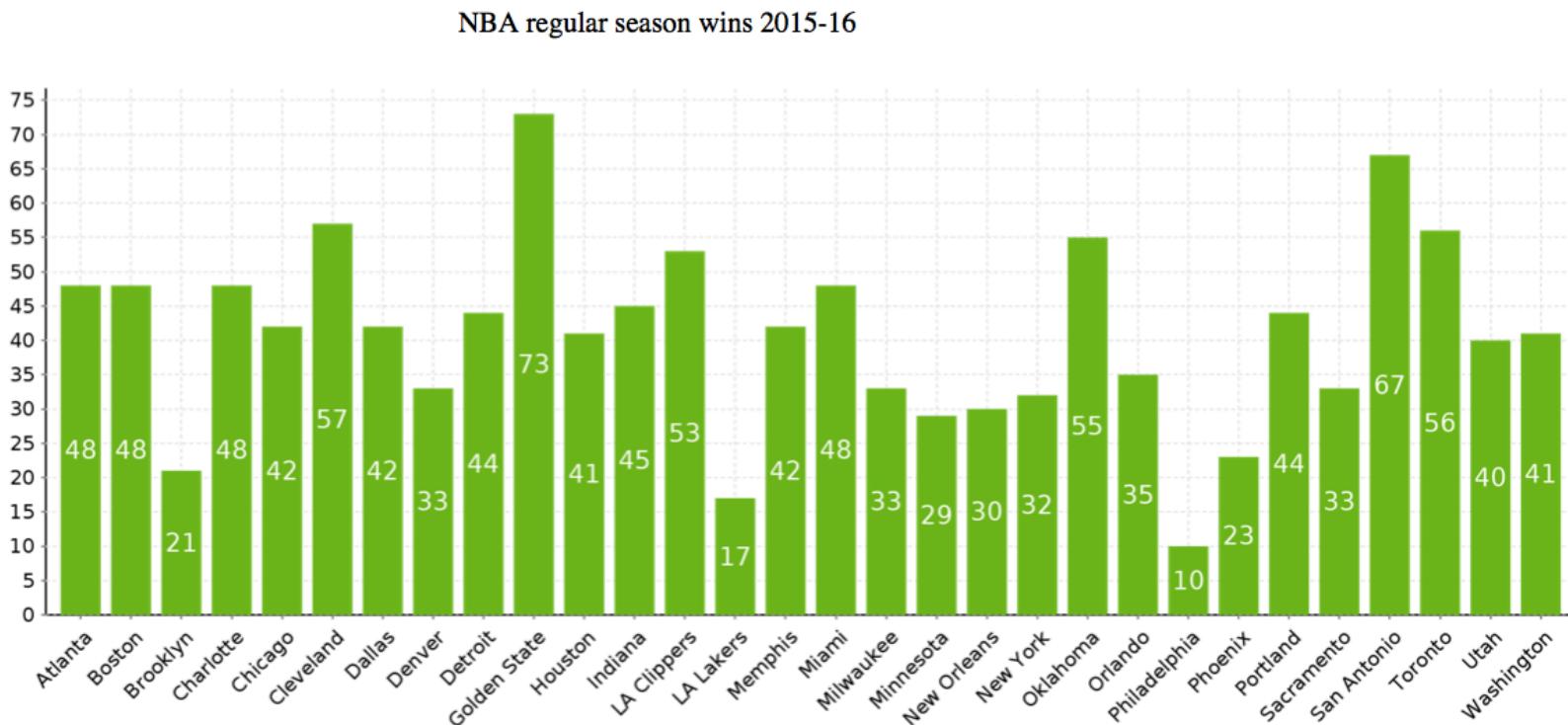
Tell a story in order.



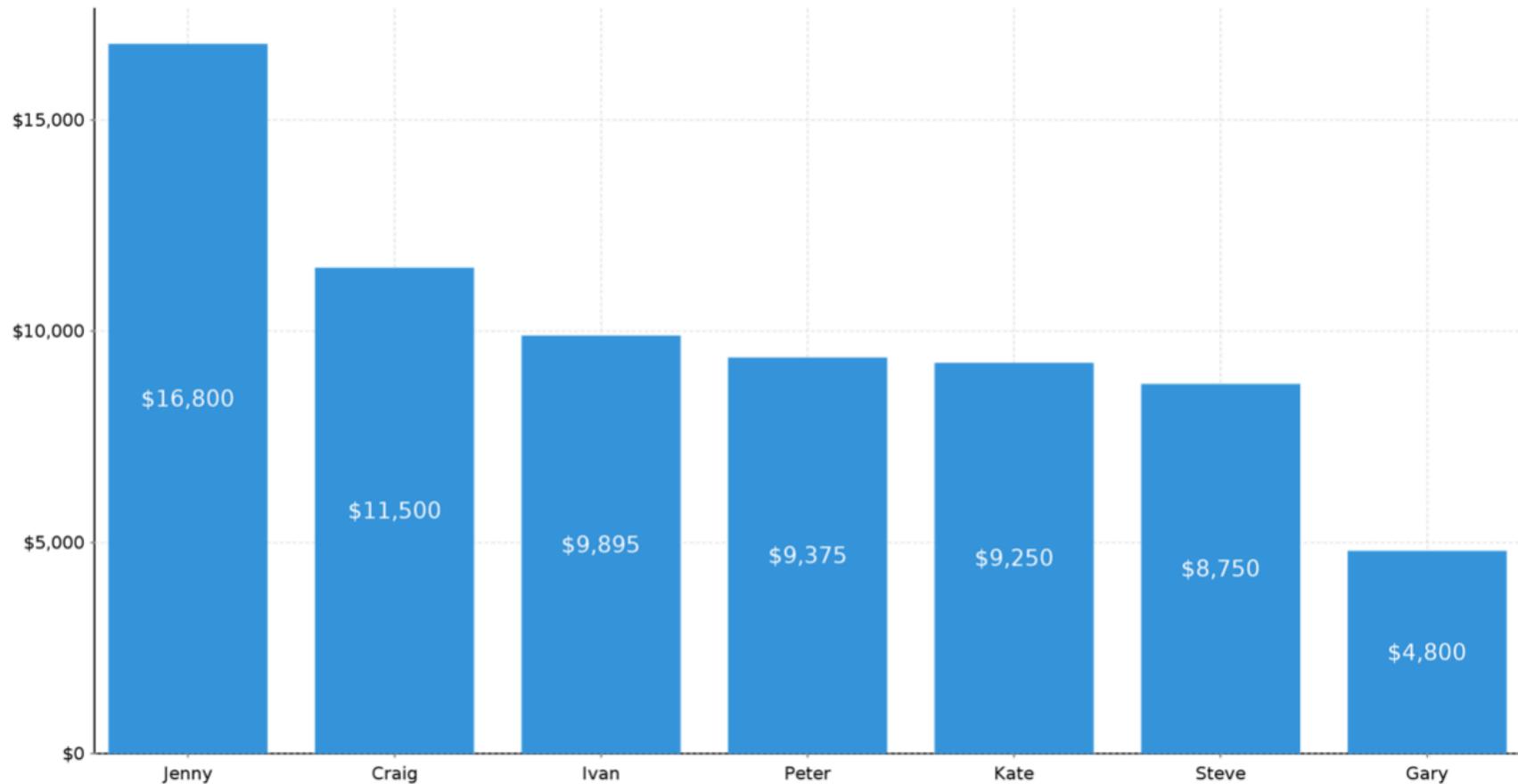
DESCENDING

Compare categories.

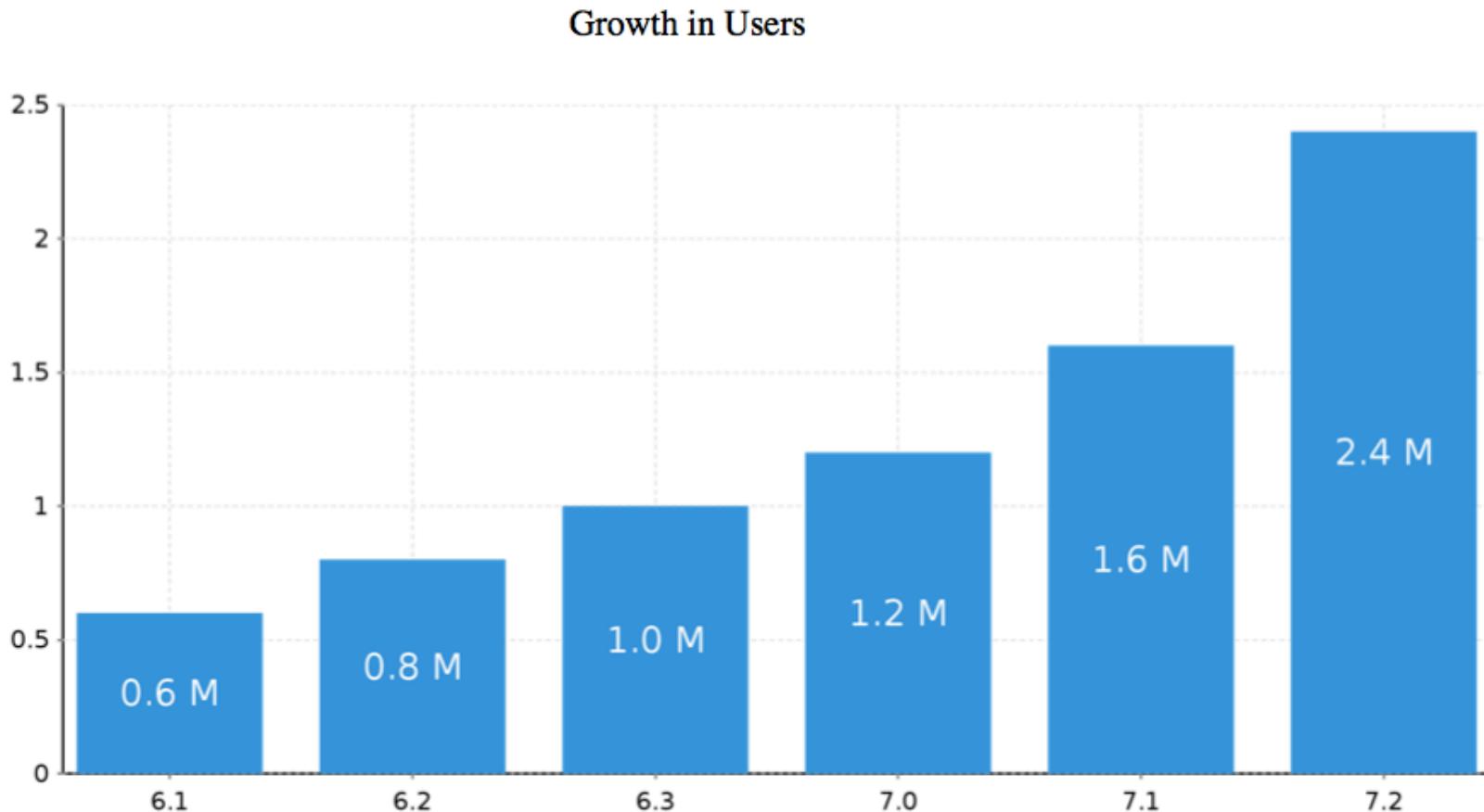
# Sorting Alphabetical helps people find what they are looking for



# Sort by descending - helps compare largest to smallest

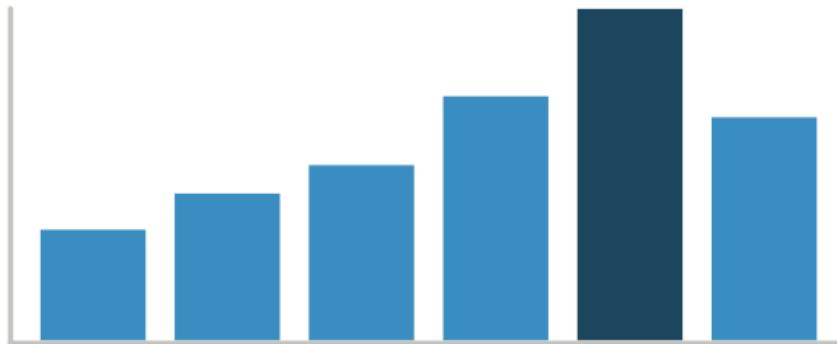


# Sort by ascending - tells a story of growth



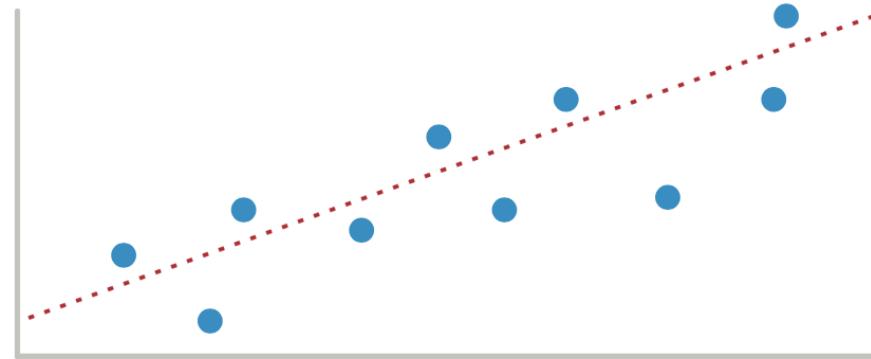
# Highlight what's important

- Draw people's eye to what's important.



## CONDITIONAL FORMATTING

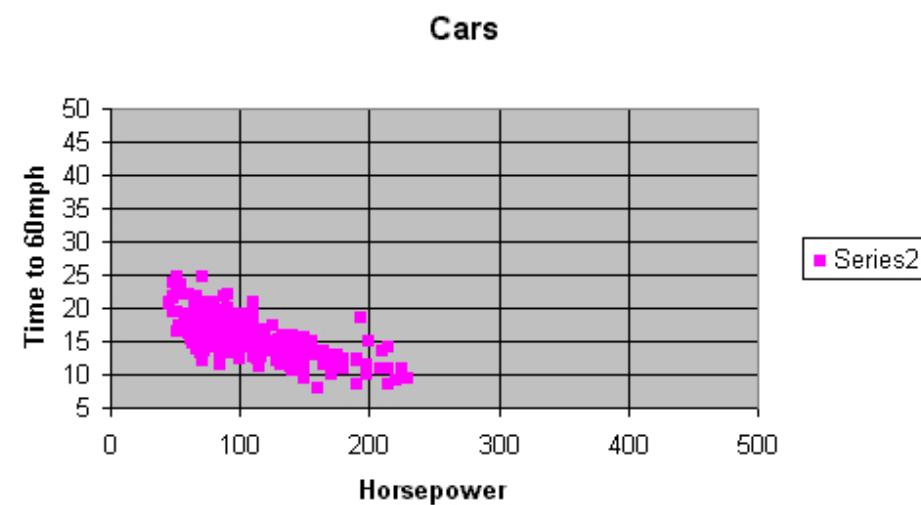
Colors data that is above or below defined thresholds.



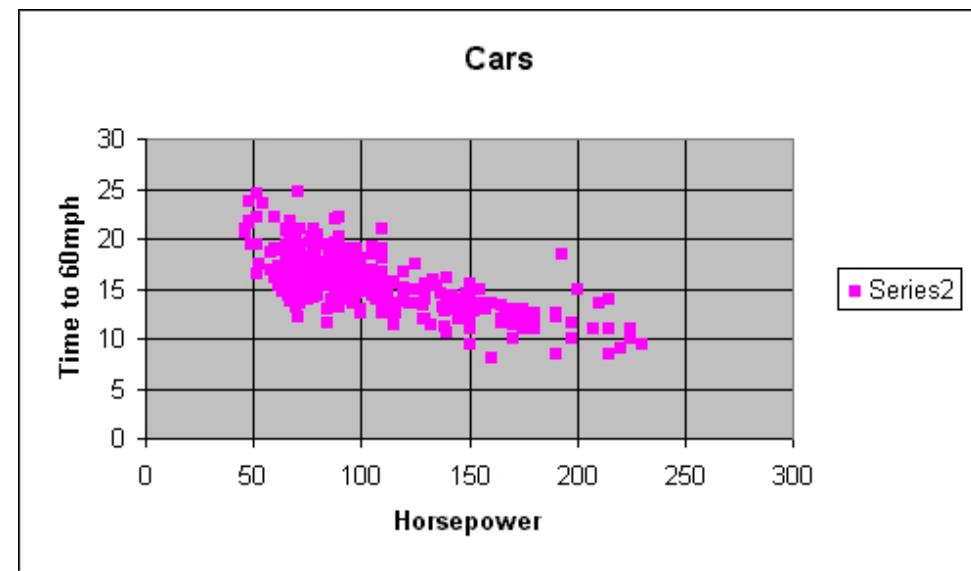
## HIGHLIGHT TRENDS

Uncover patterns in your data.

# Give your chart reasonable data range



illogical



logical

# Limit numbers

## Passengers

**Growth 2.32%**

**Actual 34,485,509**

**Benchmark 33,704,843**



## Passengers

**Growth 2.32%**

**Actual 34.49M**

**Benchmark 33.70M**



# Limit numbers

## Passengers

Growth 2.32%

Actual **34,485,509**

Benchmark 33,704,843

## Passengers

Growth 2.32%

Actual **34.49M**

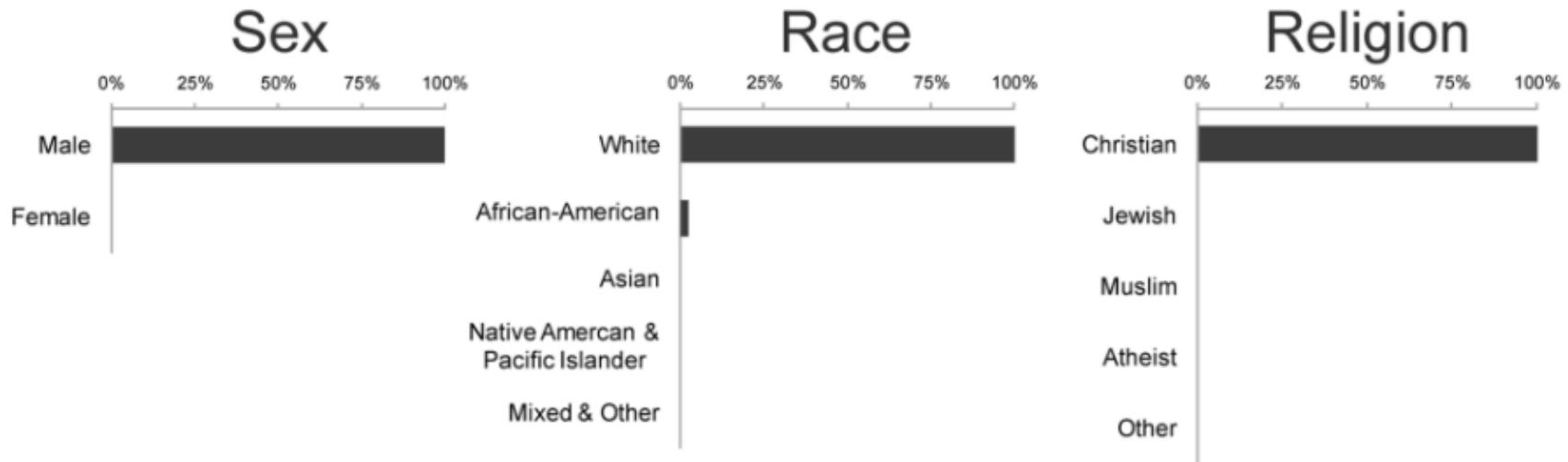
Benchmark 33.70M

# Less is More

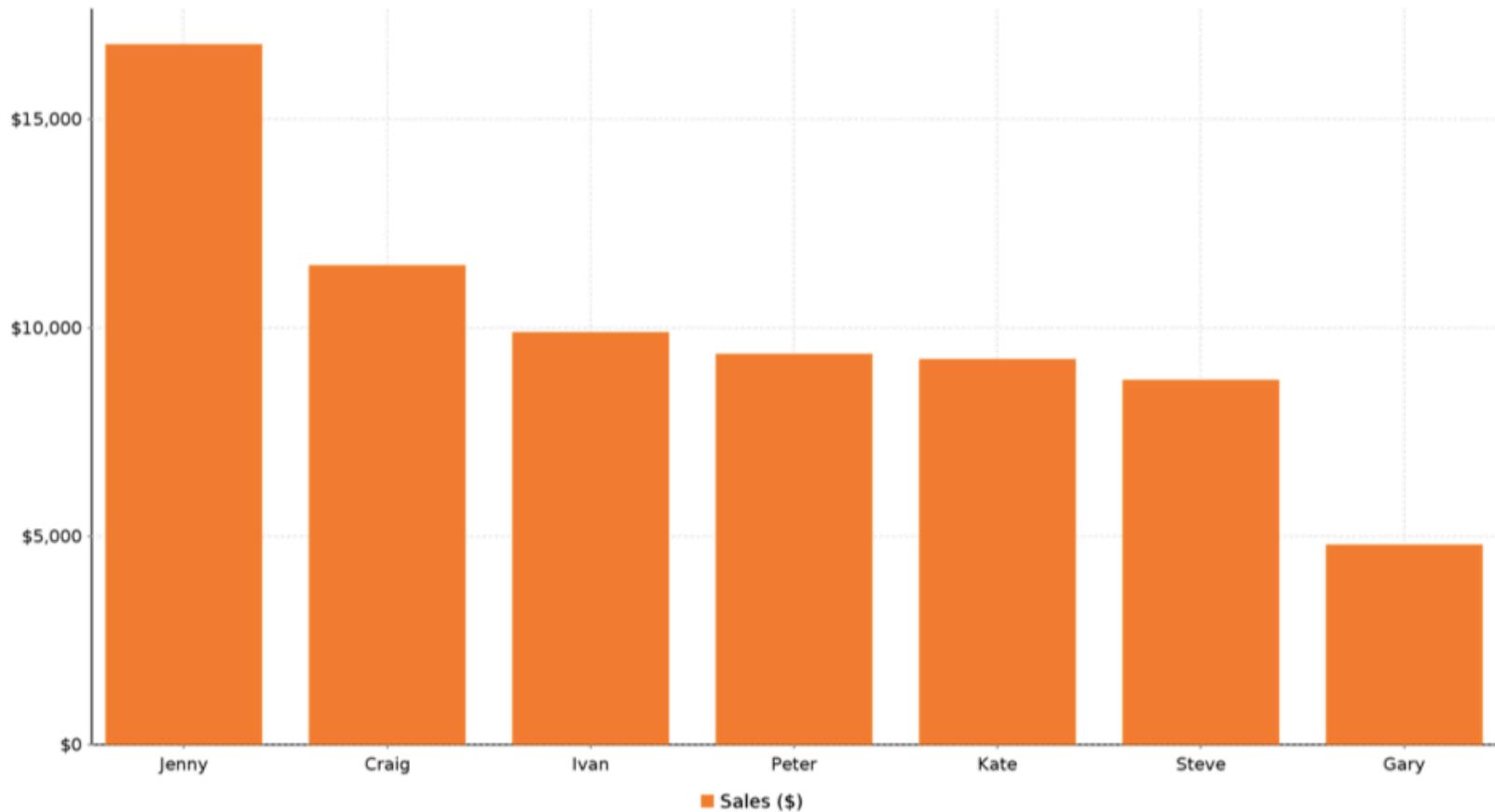
- Focus on what is most important
- You need your dashboard (or report) to be
  - Clear
  - Concise
  - not overly stuffed with charts and objects

# Simple is powerful!

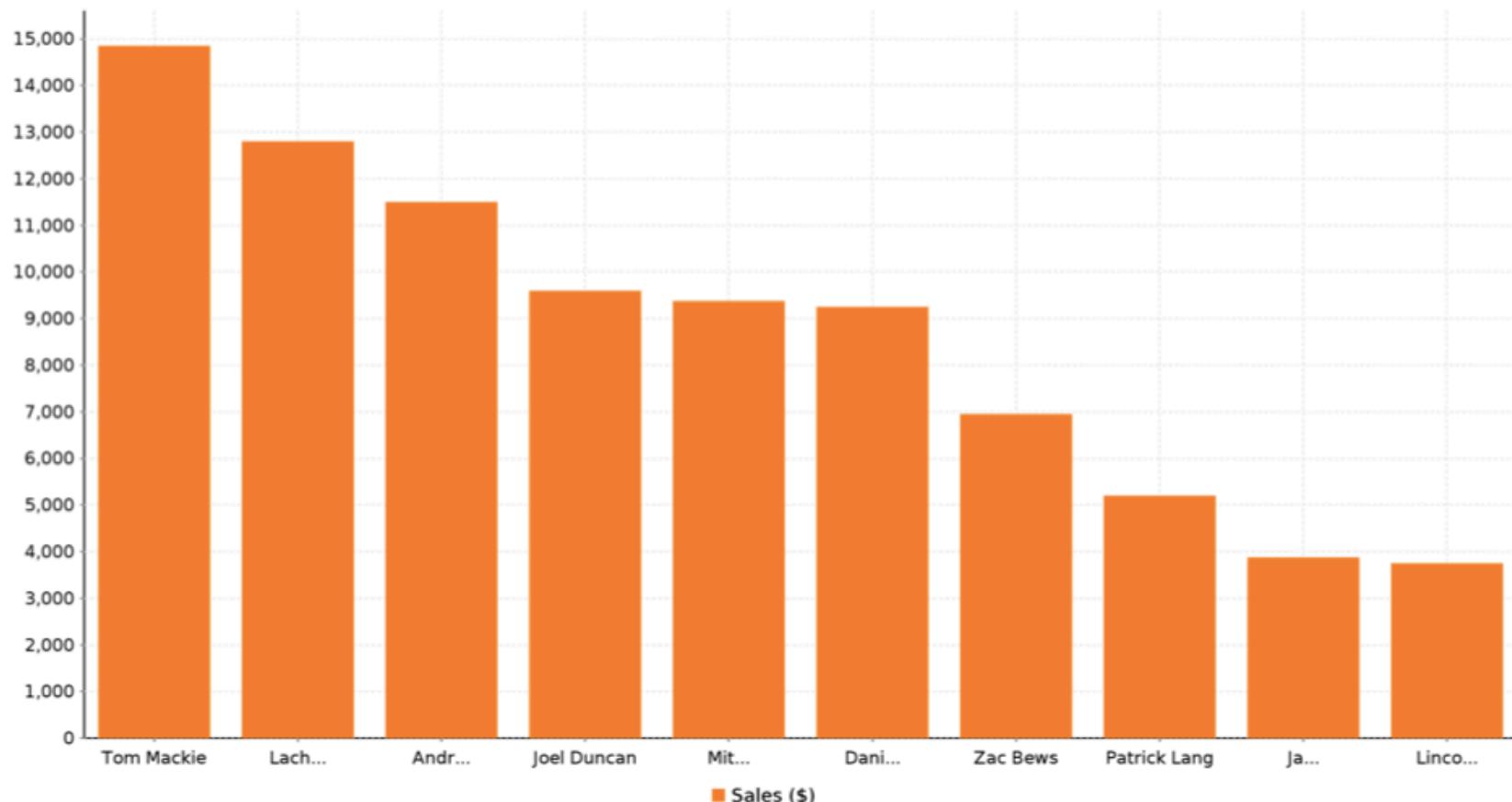
## Statistics about U.S. Presidents



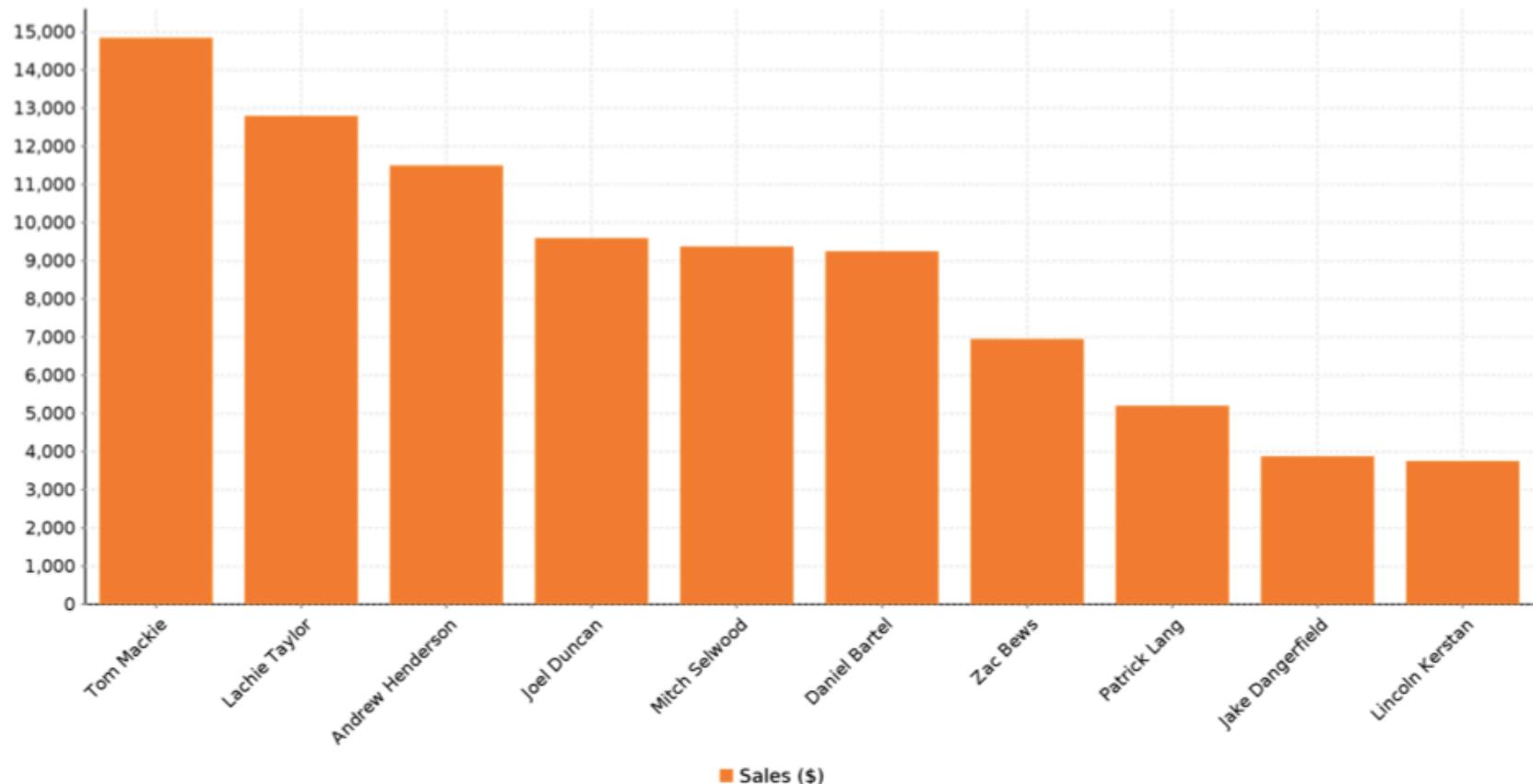
# Column charts are simple and easy to understand...



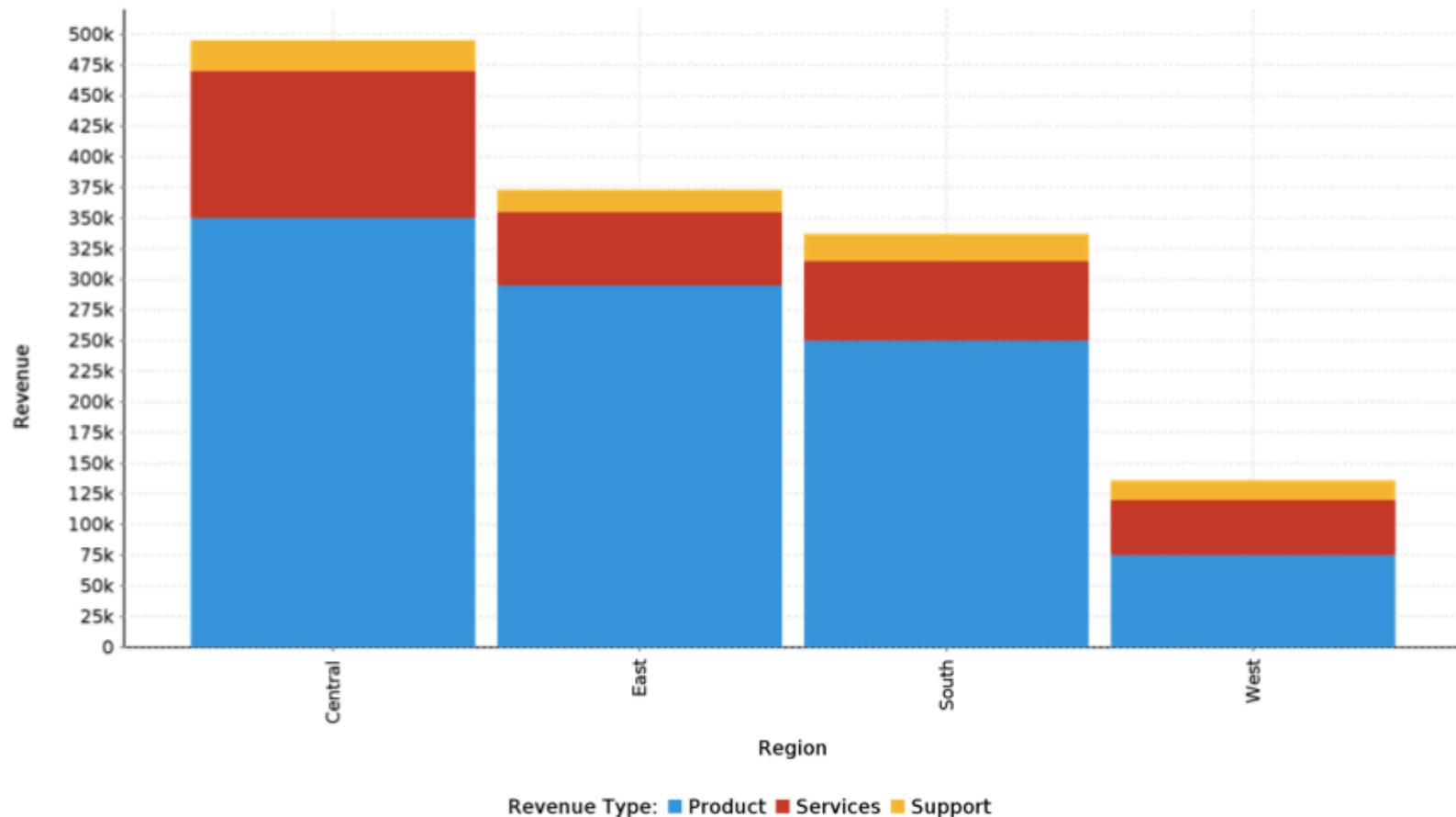
...but become difficult to compare  
with too many categories.



# One solution is to rotate the horizontal font...

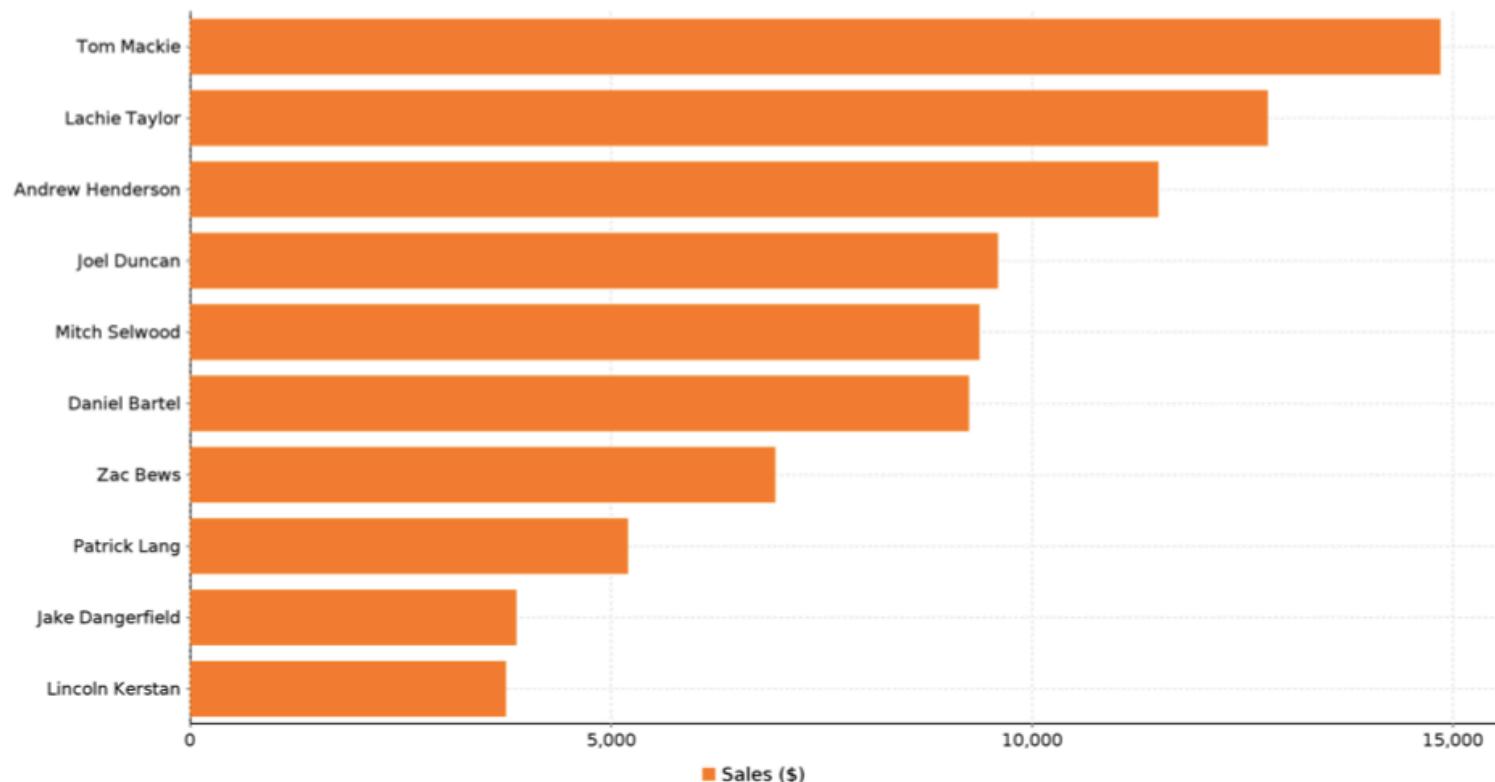


# Try to avoid vertical labels as possible

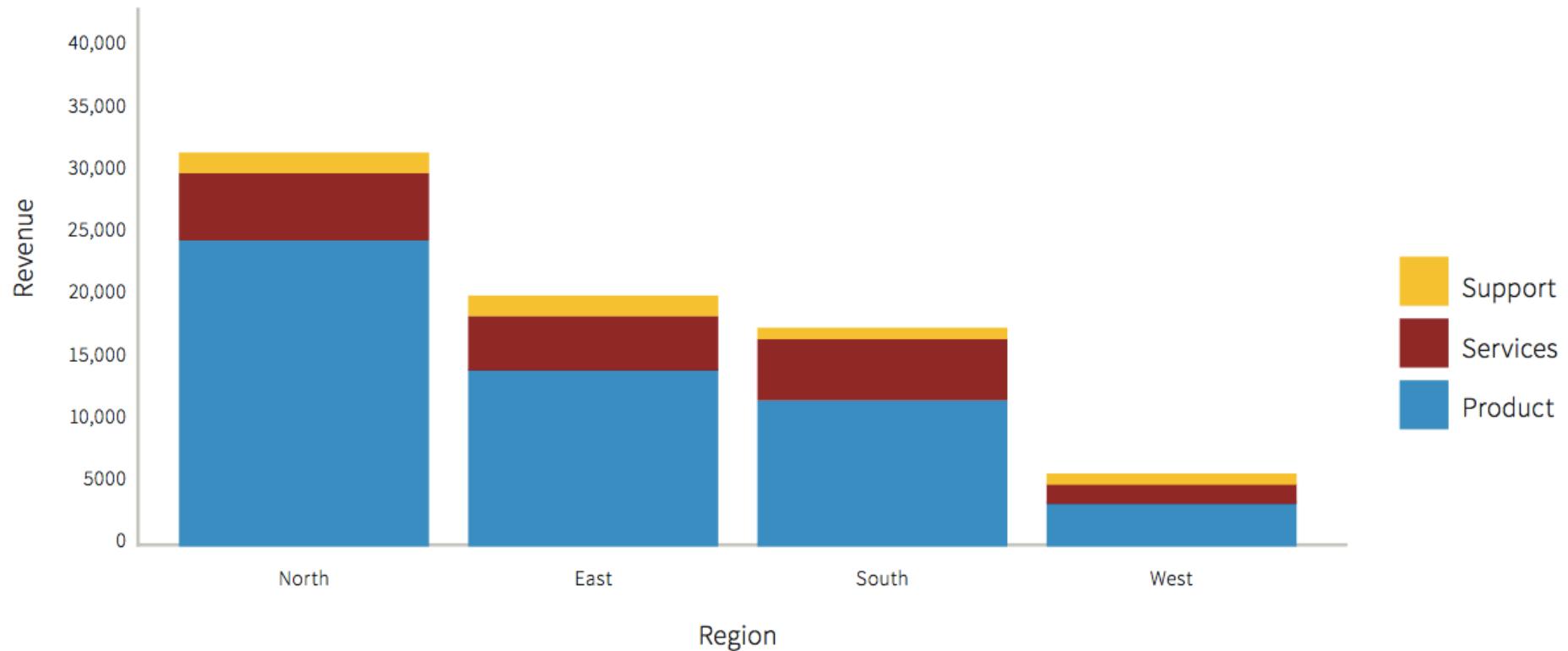


# ...Bar chart best if more than 10 categories

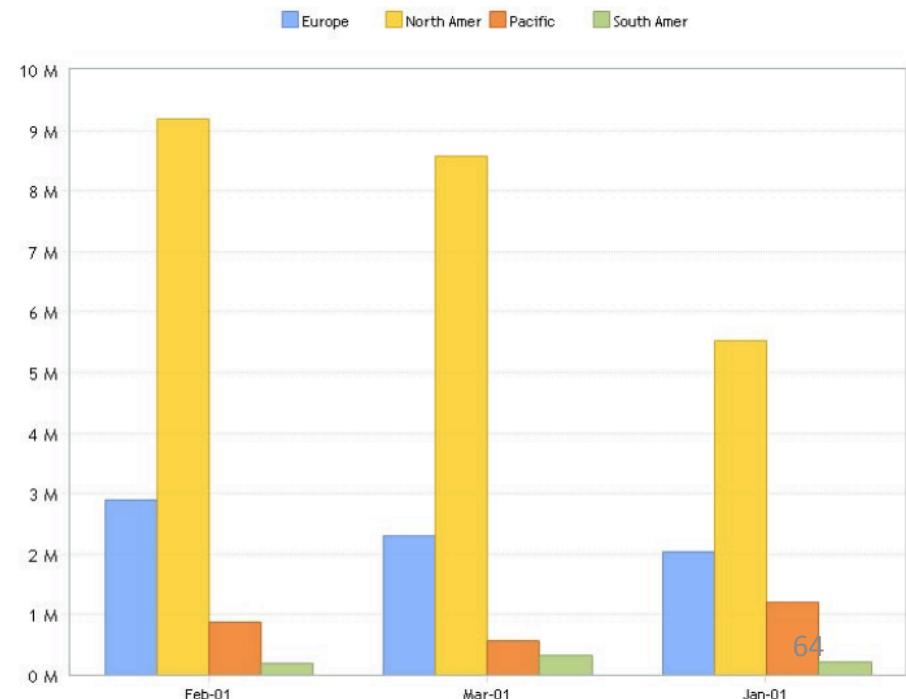
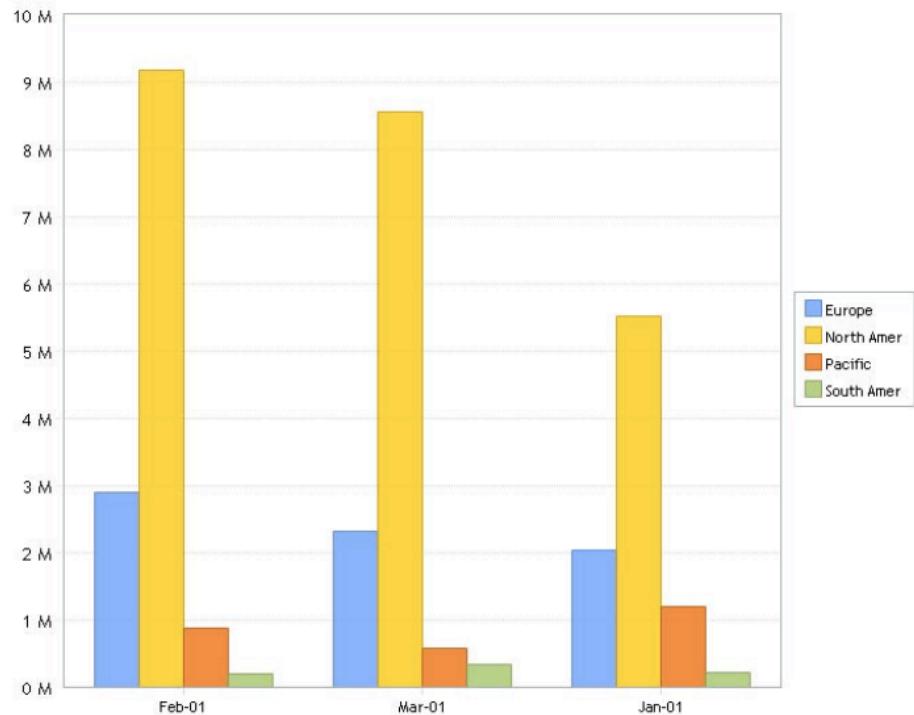
- Horizontal can be good if long text labels or many items



# Use legends as your key



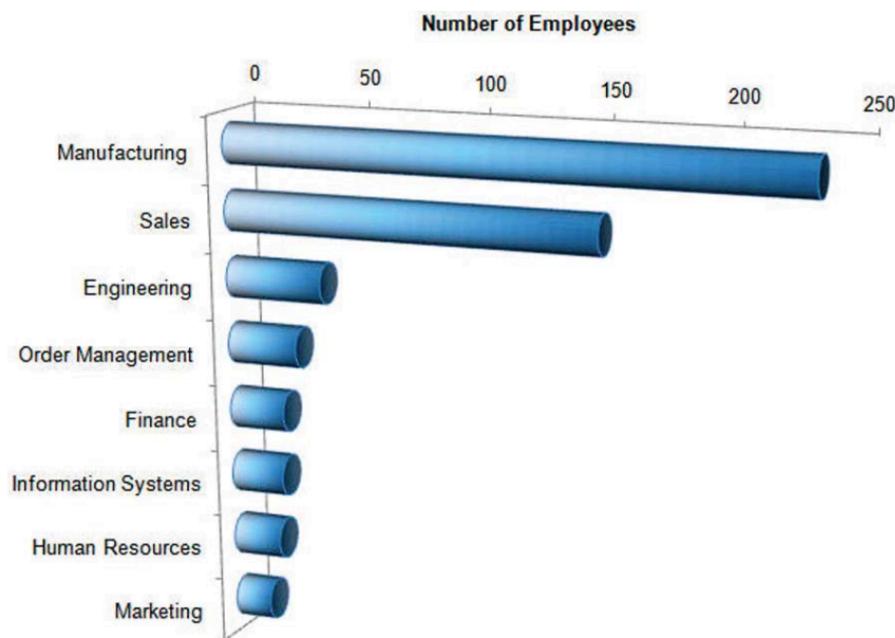
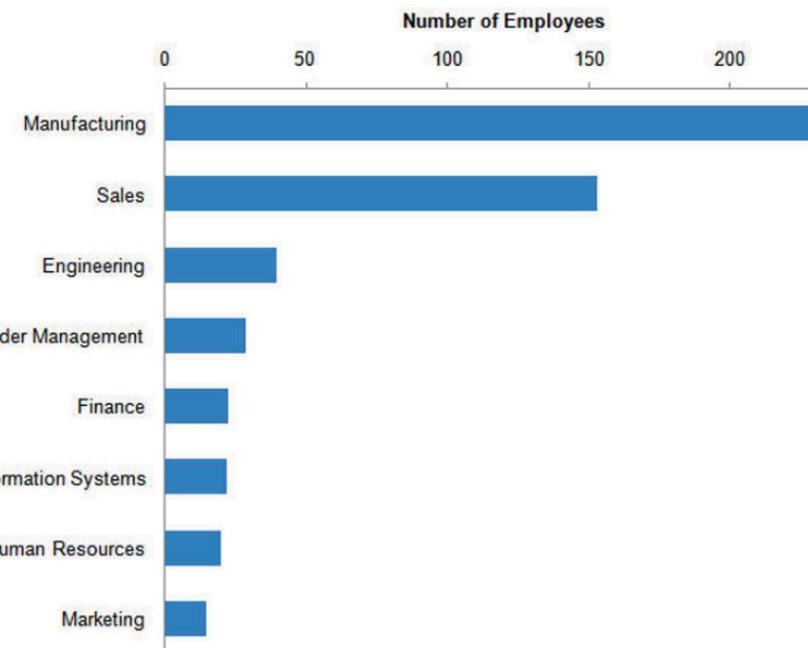
# Legend placement



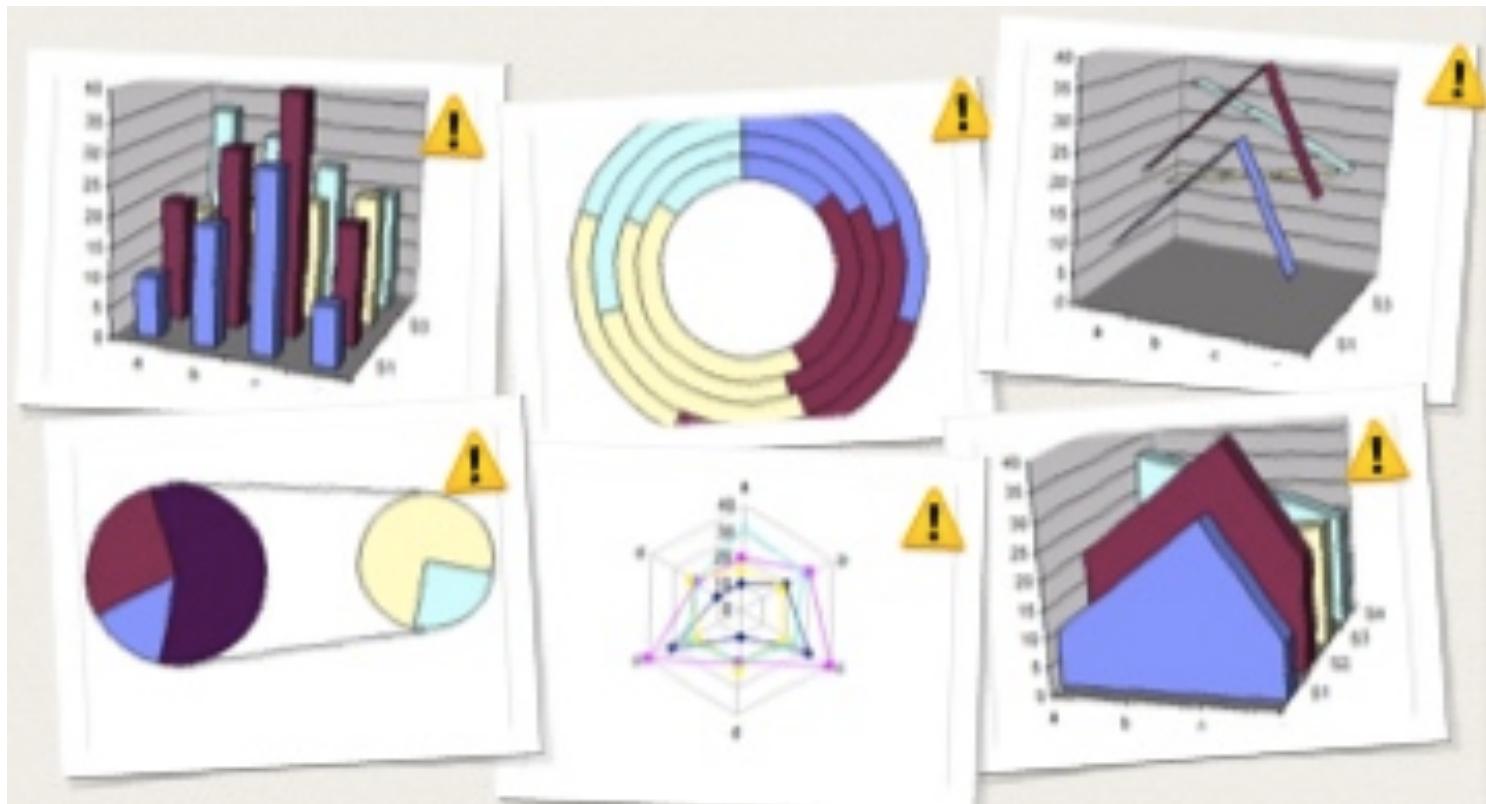
Few, Effectively Communicating Numbers

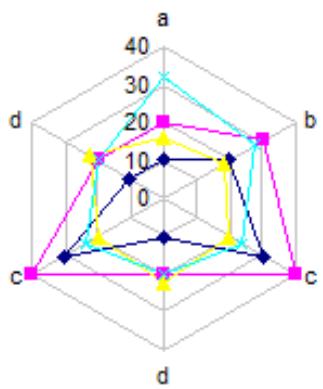
# What not to do!

# Which of the following two graphs is easier to read?

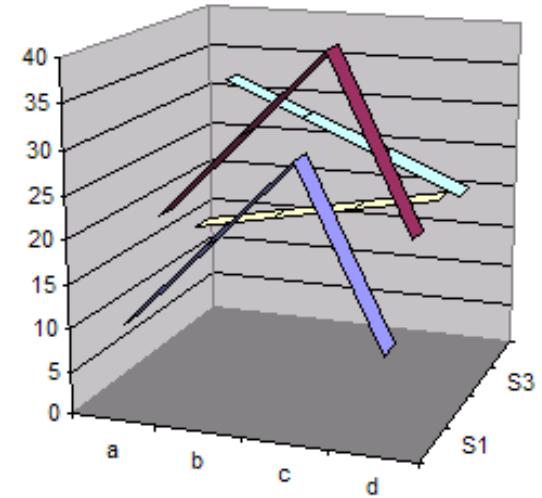


# 6 charts you will see in hell

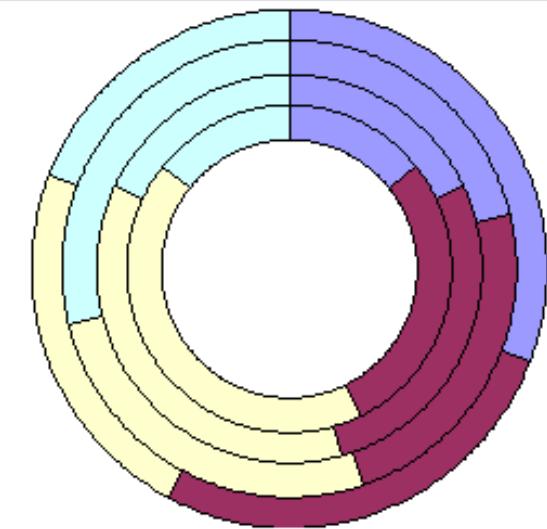




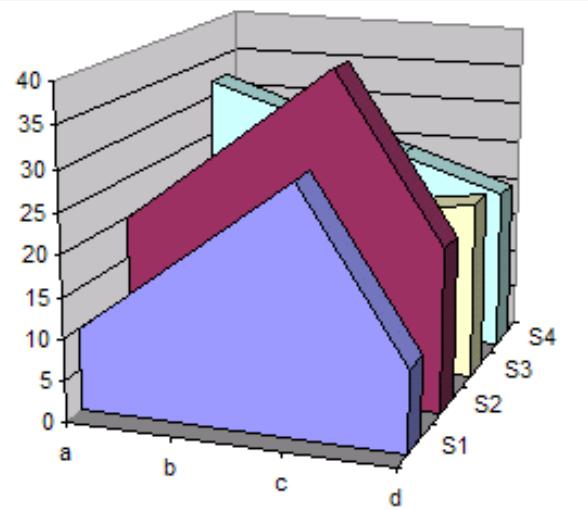
Leave the radar charts for Spidermen



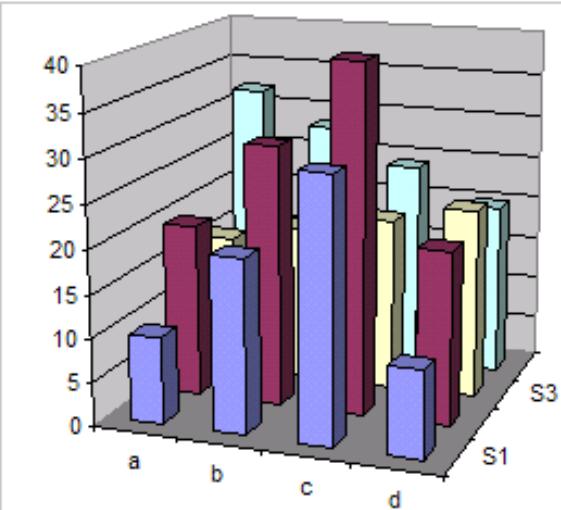
Don't add dimensions to your lines



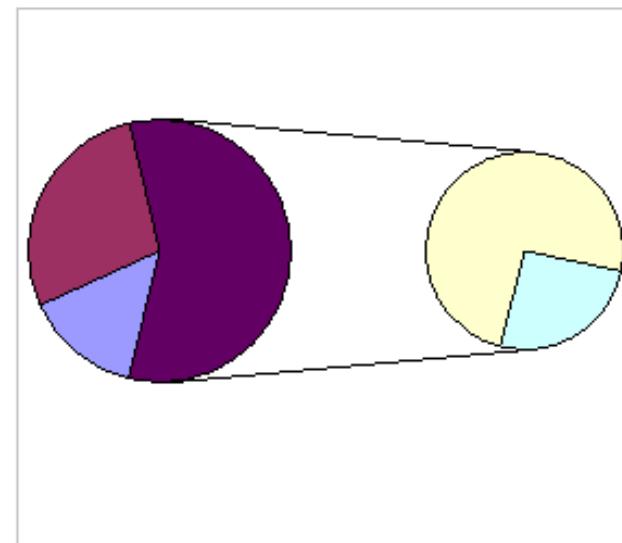
Don't show, just eat your donuts



Save the unstacked area charts till we have x-ray vision

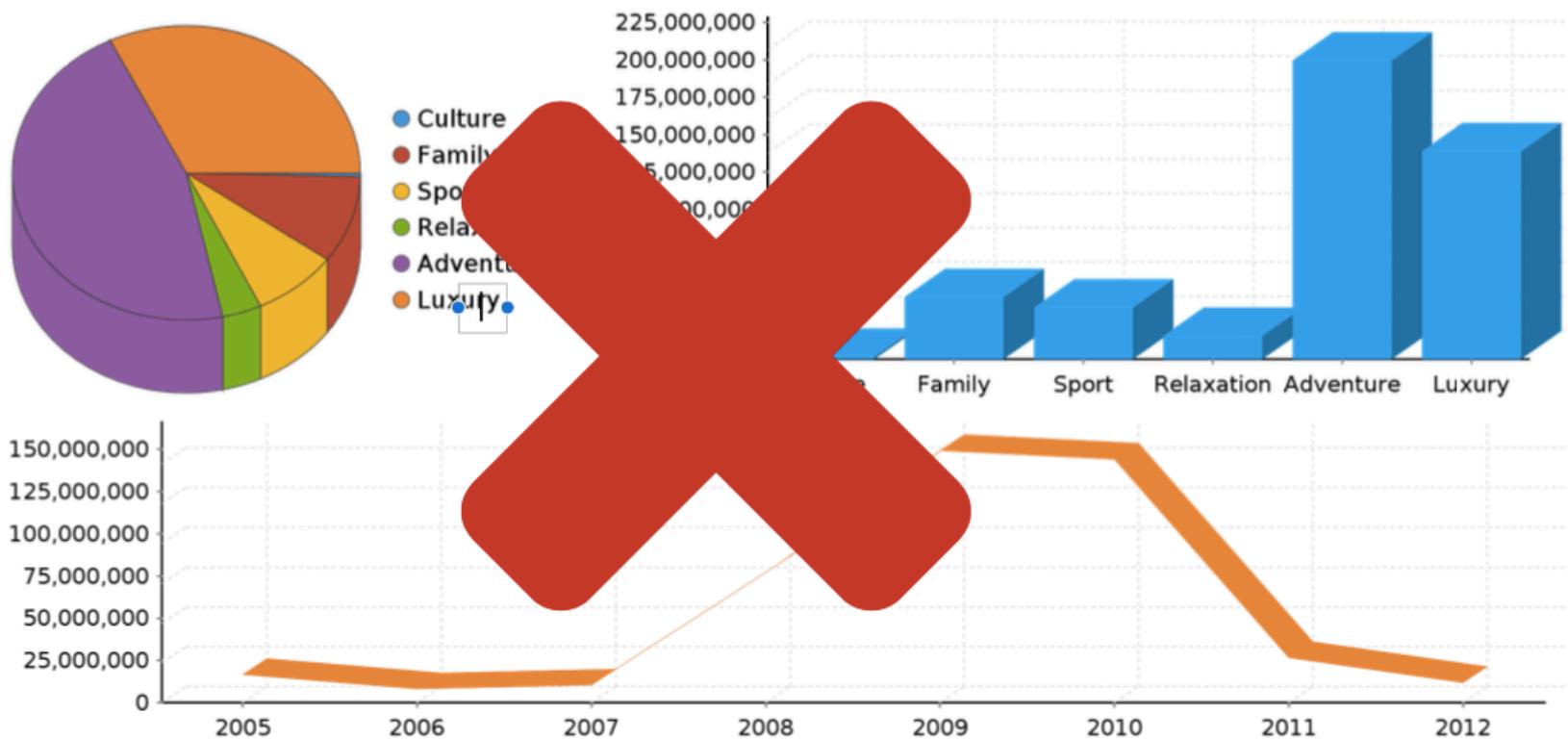


Don't make your charts look like downtown



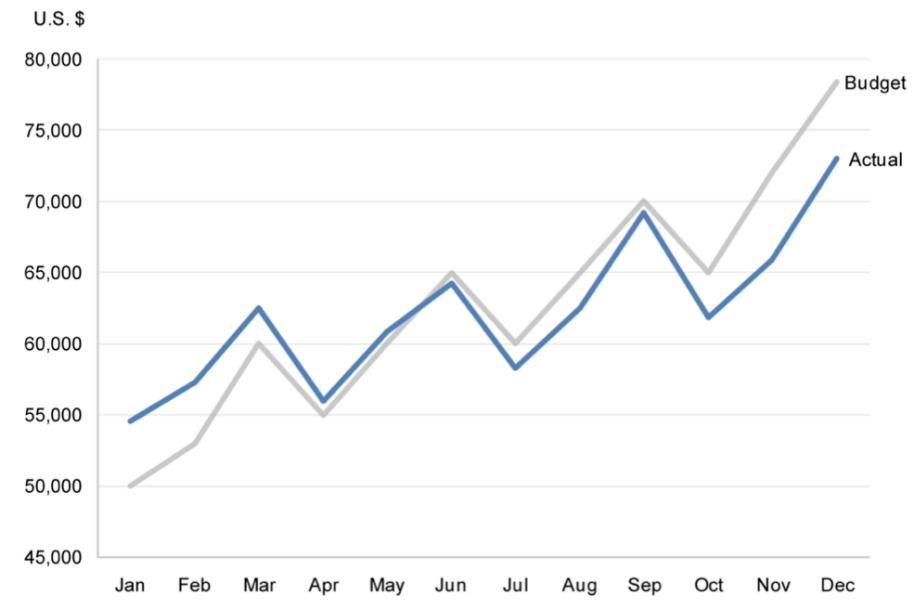
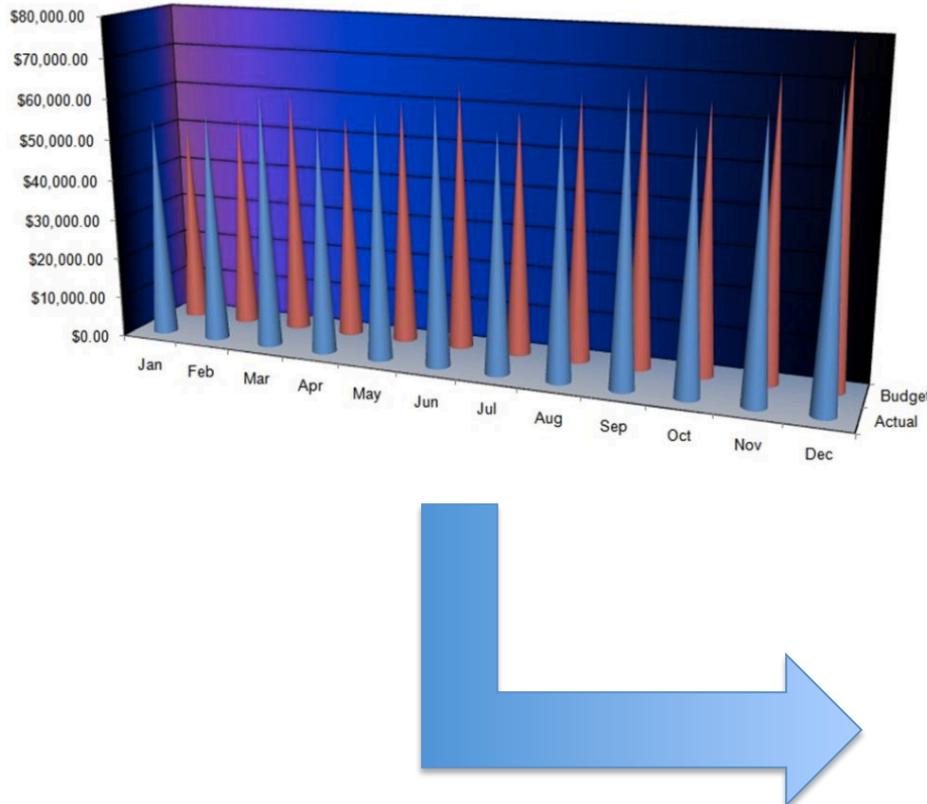
If one Pie is bad, two of them is worst

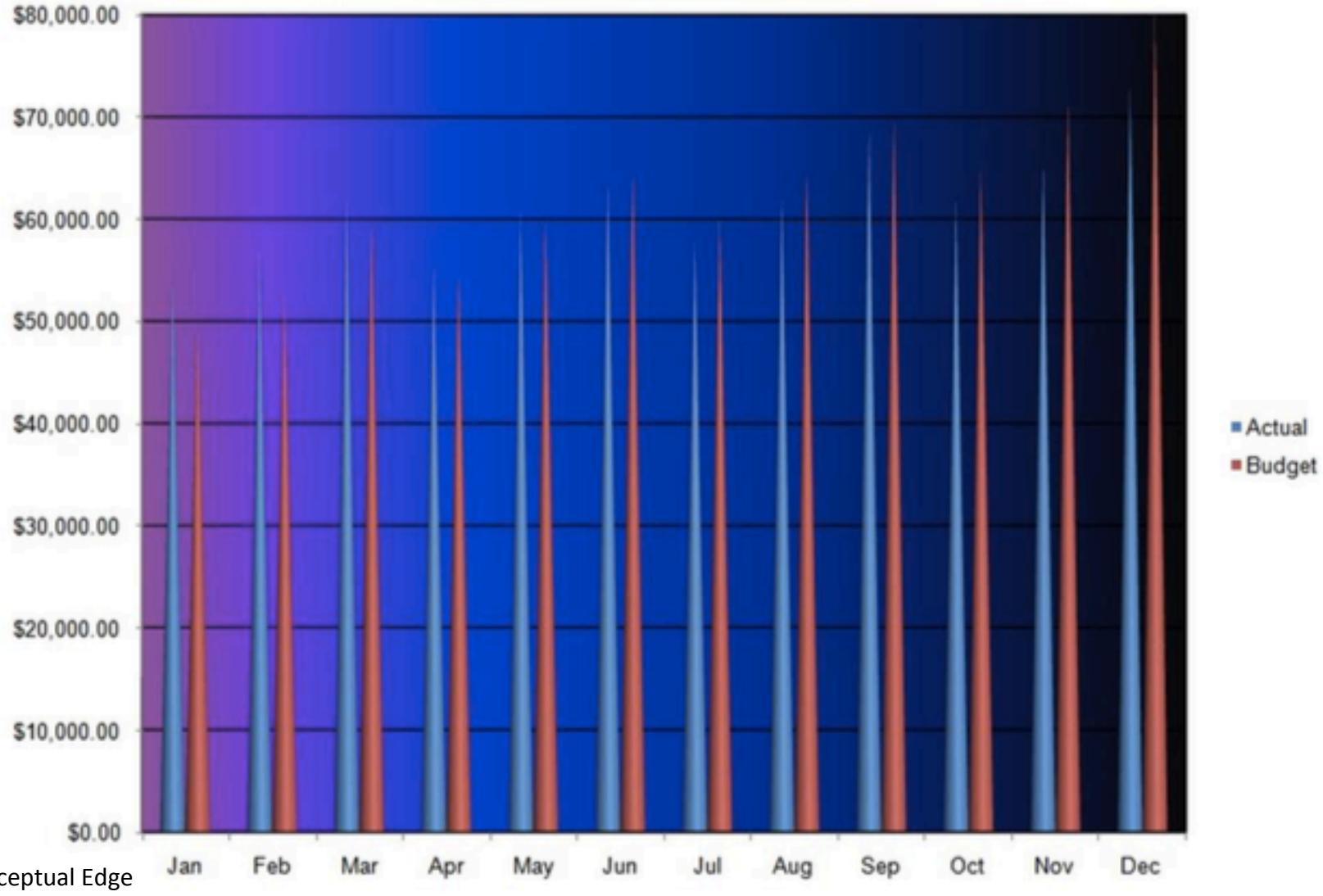
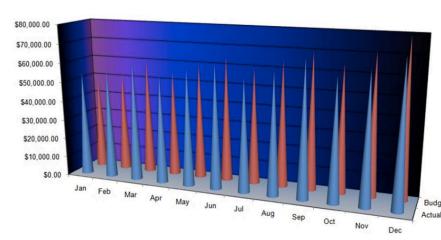
# 3D is not best practice

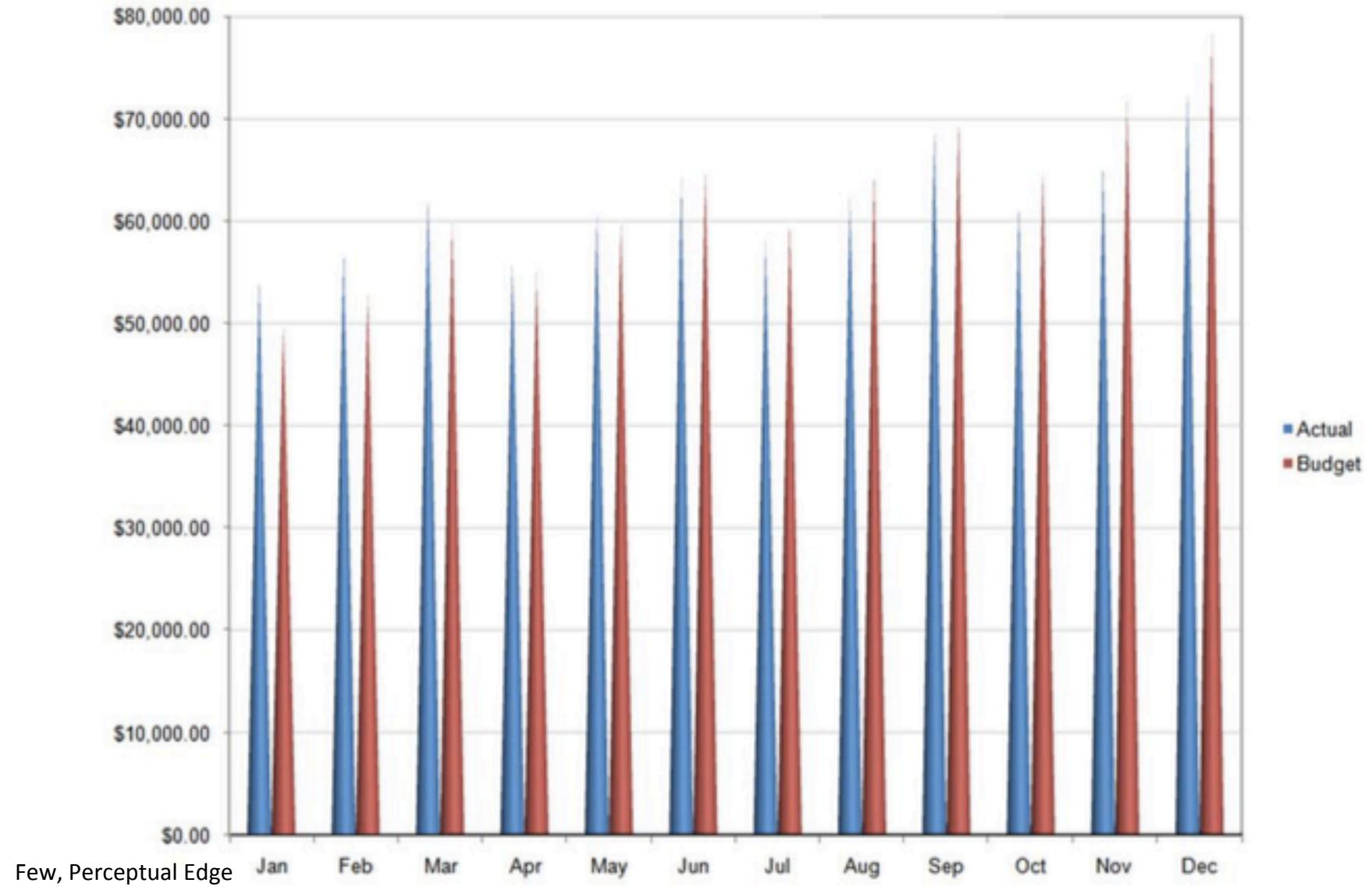


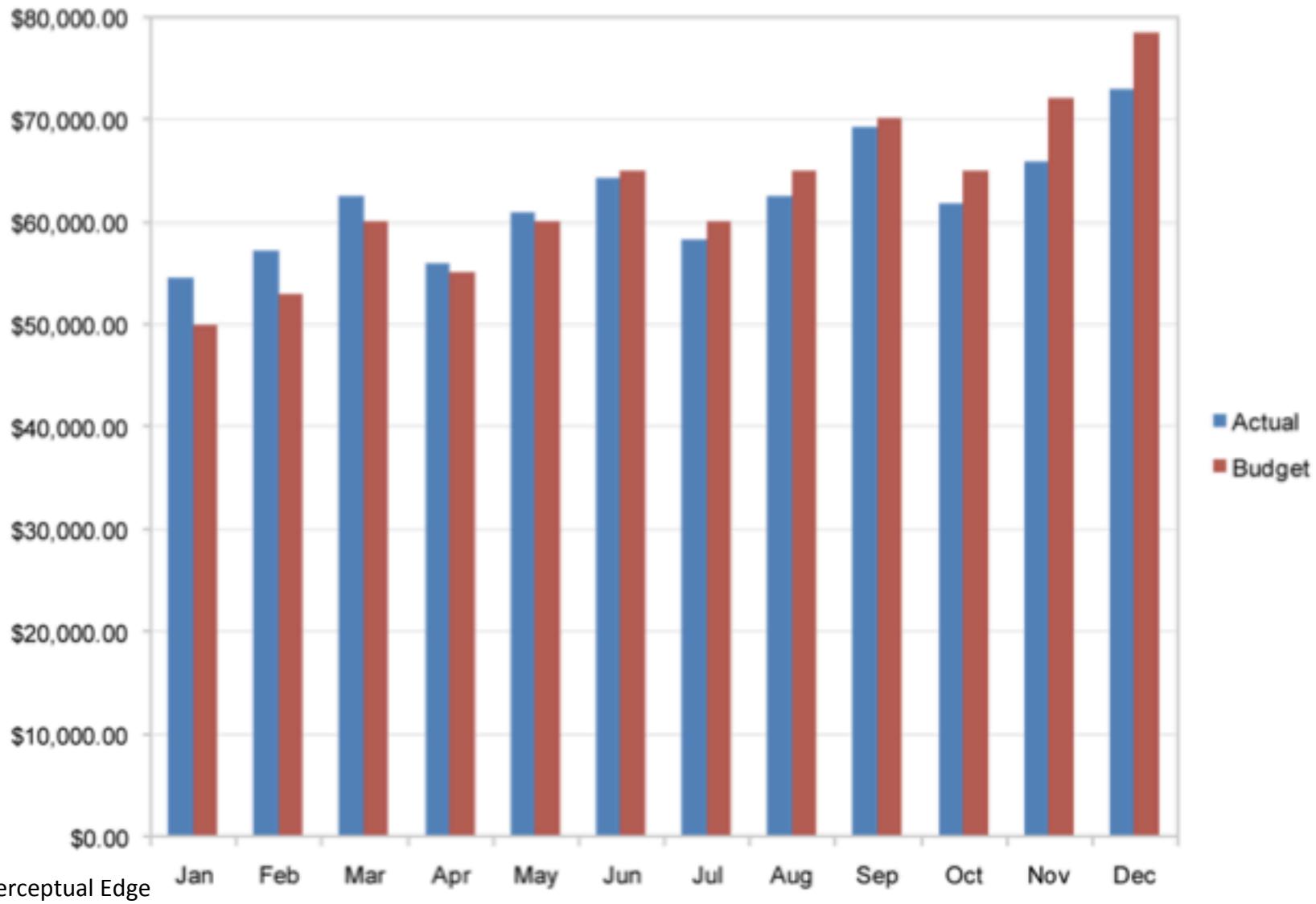
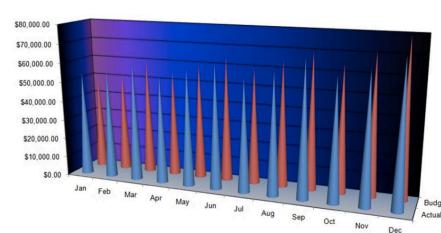
Poor graphs can be transformed

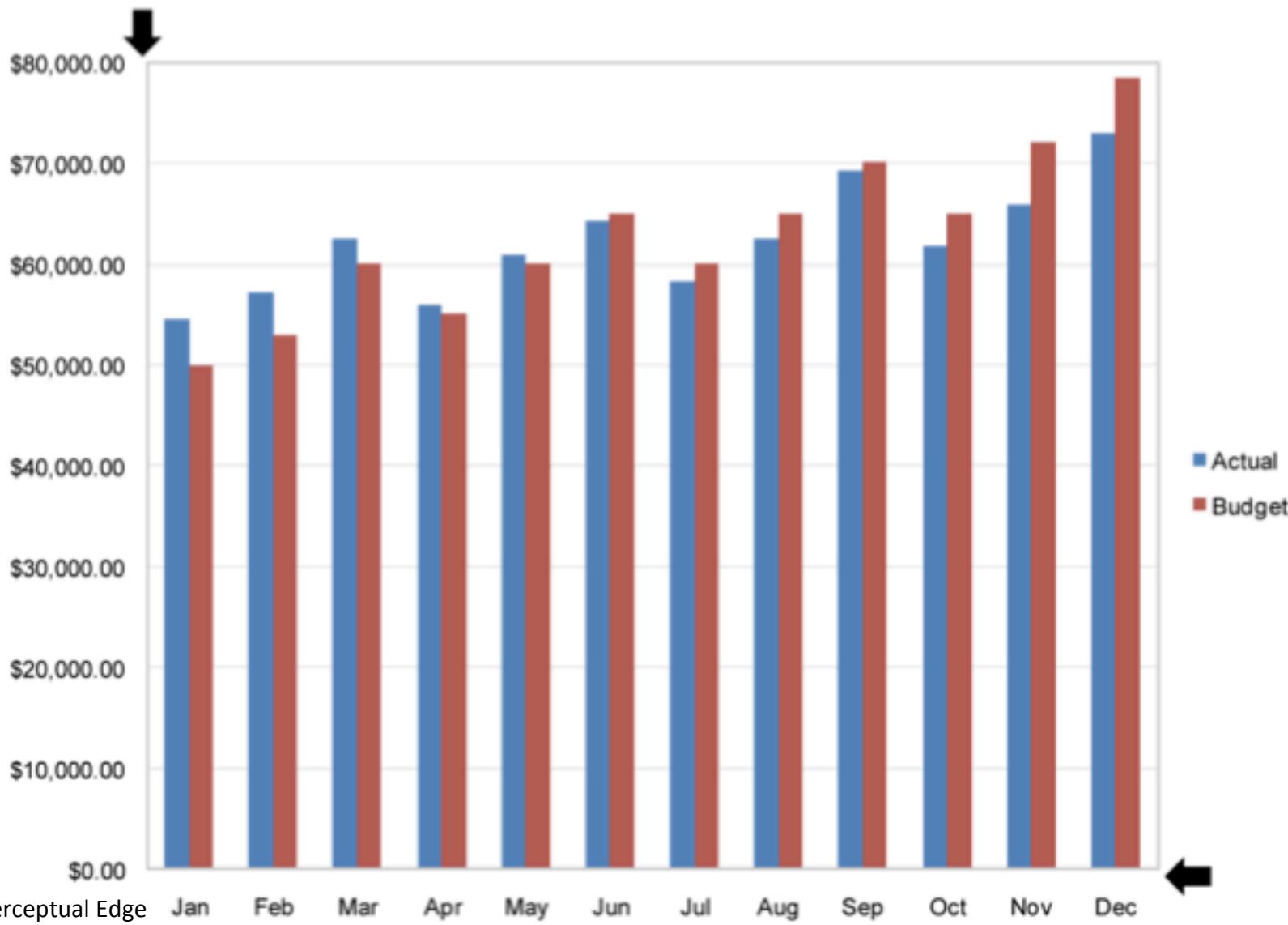
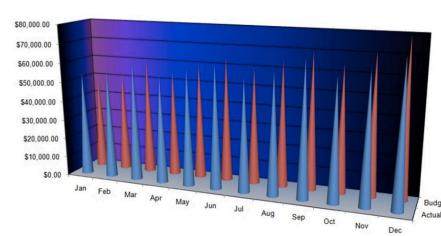
# Try to compare actual expenses to the budget across time

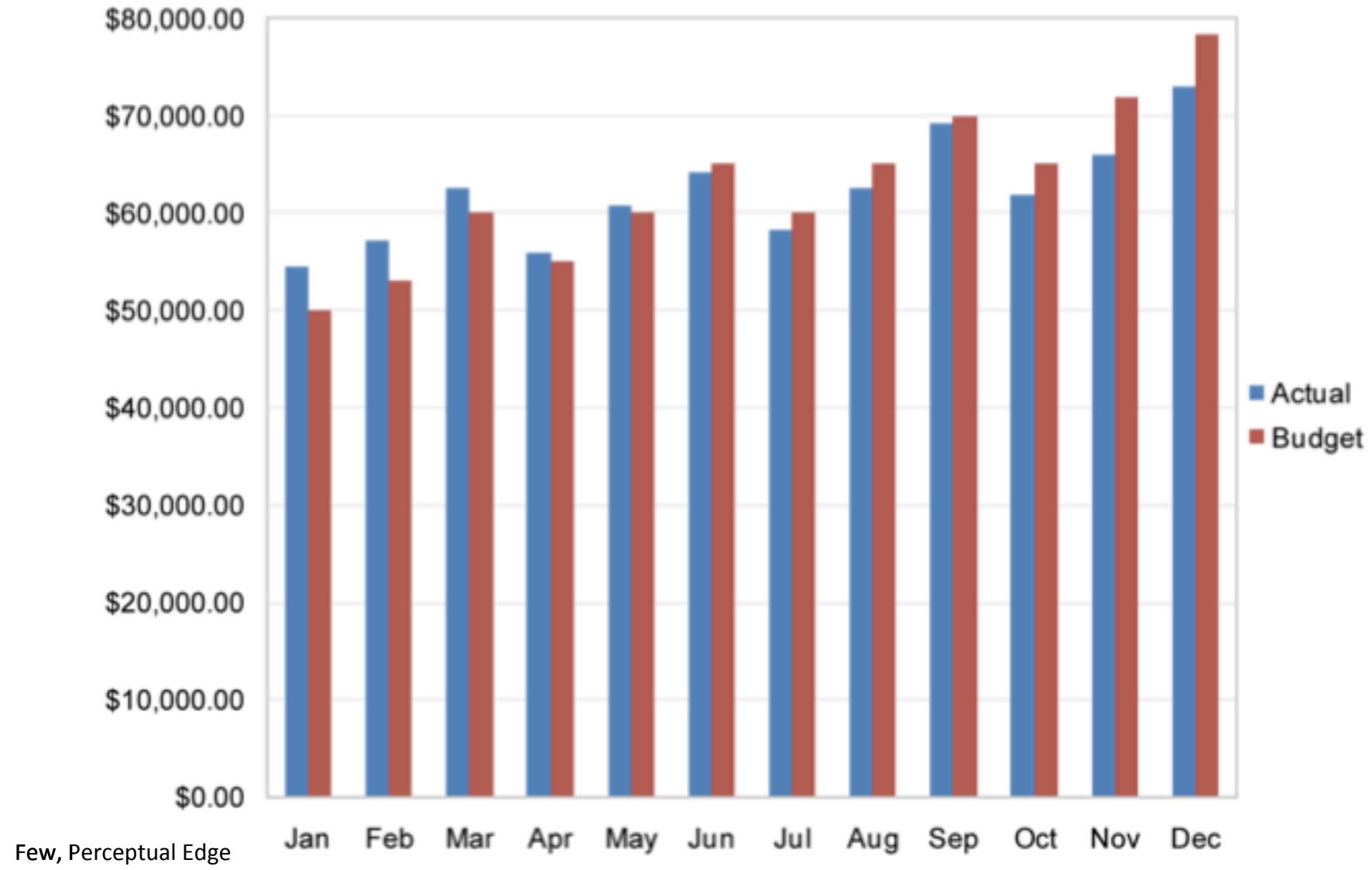
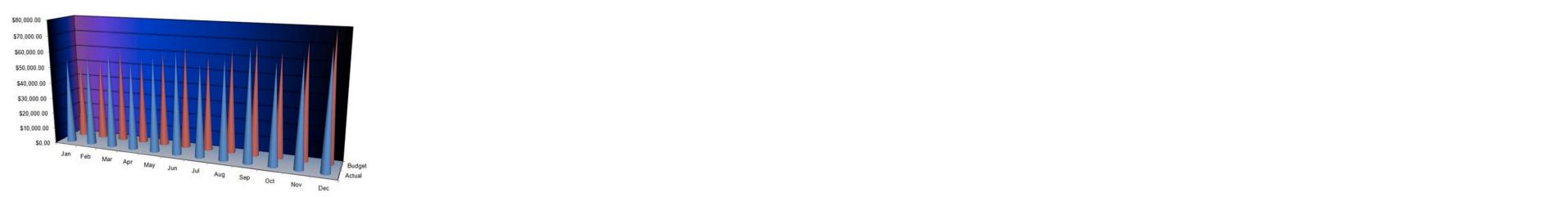


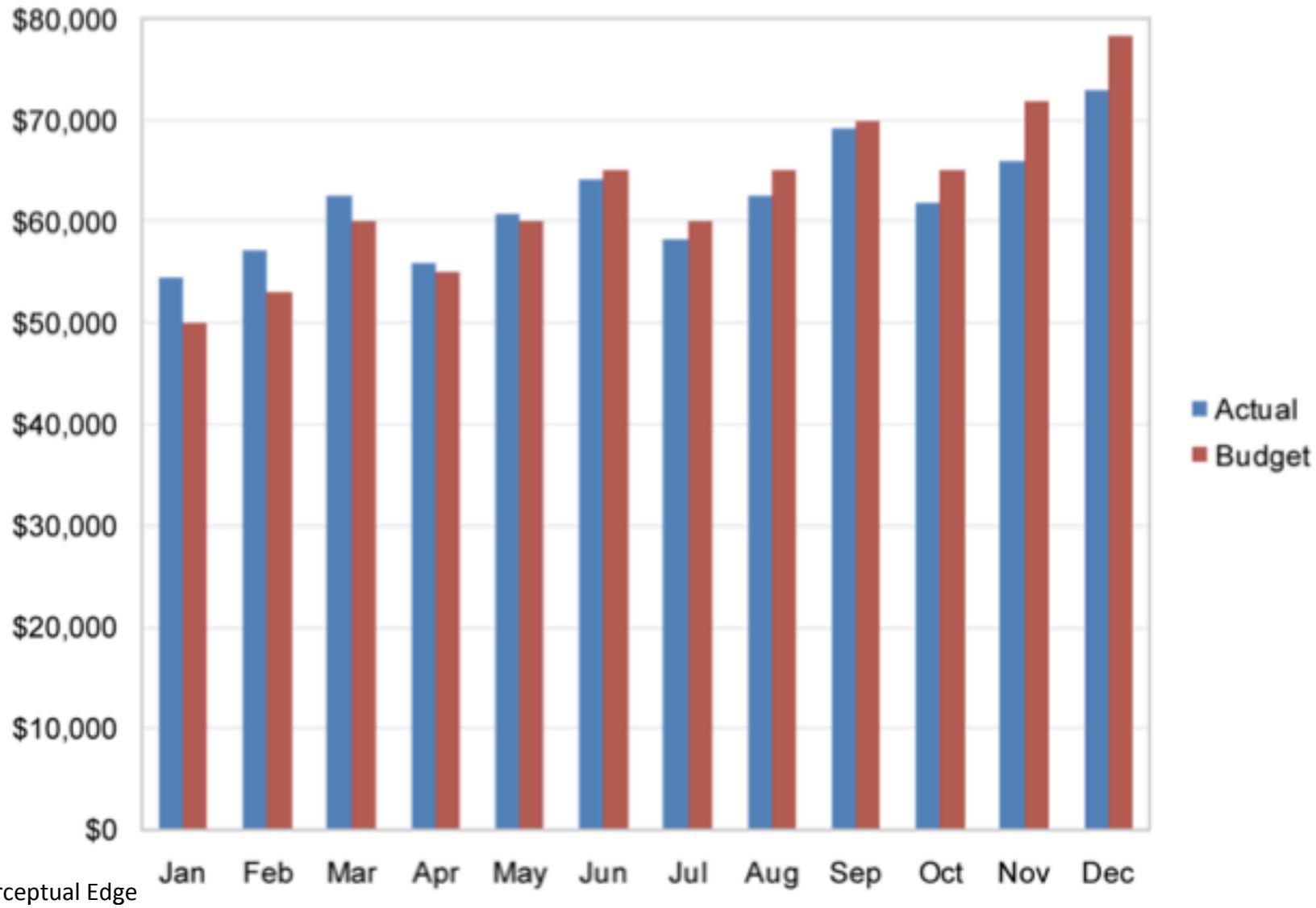
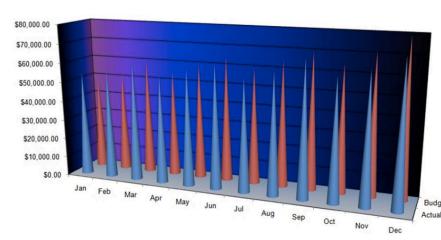


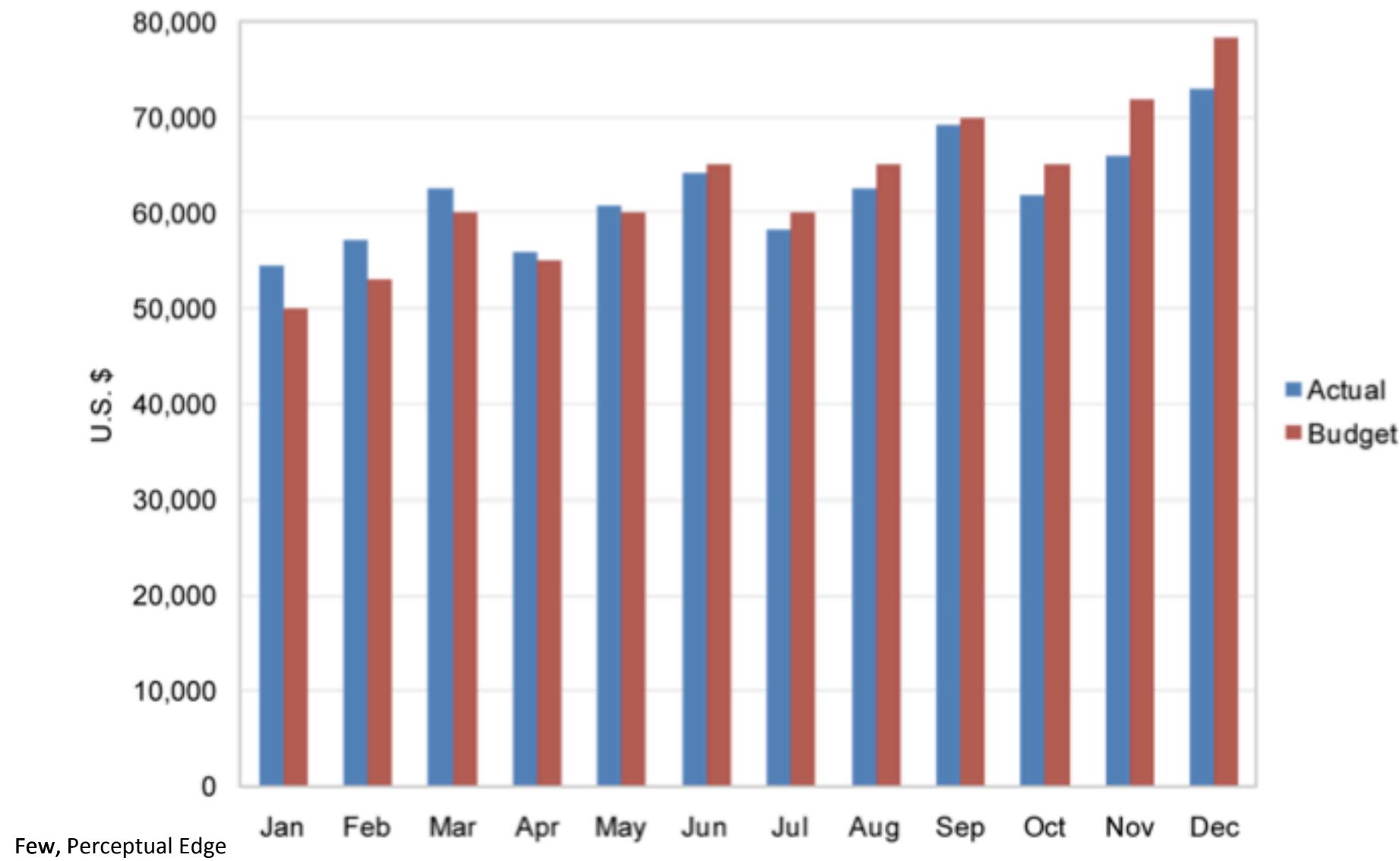
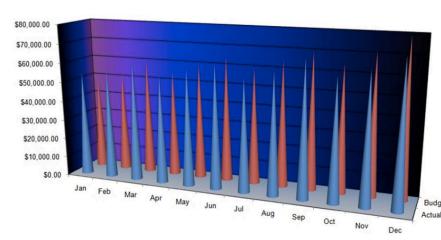


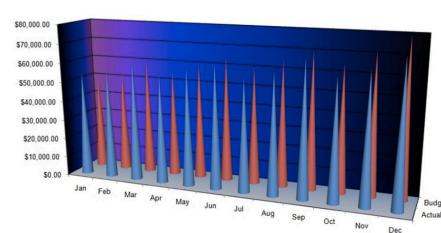




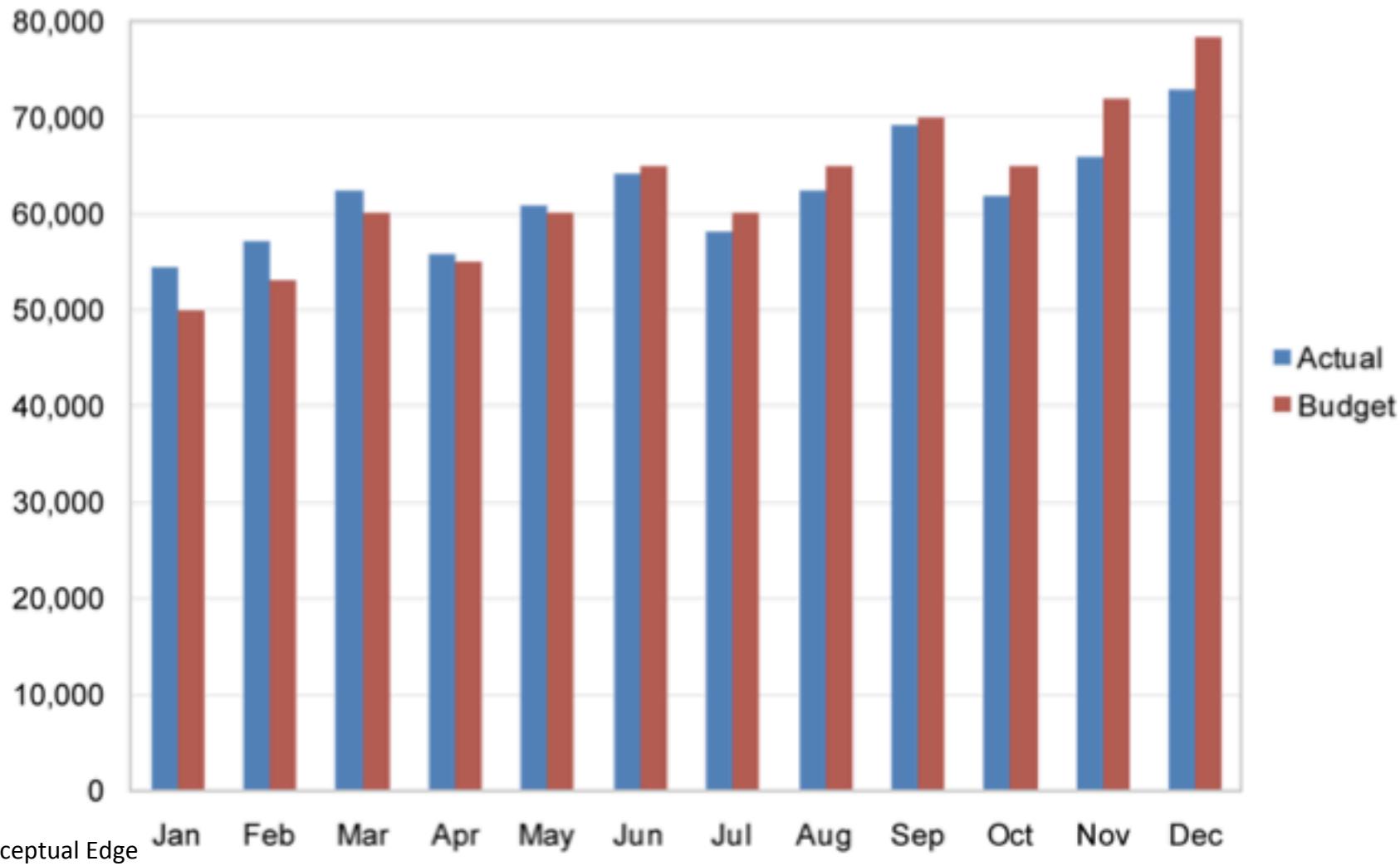


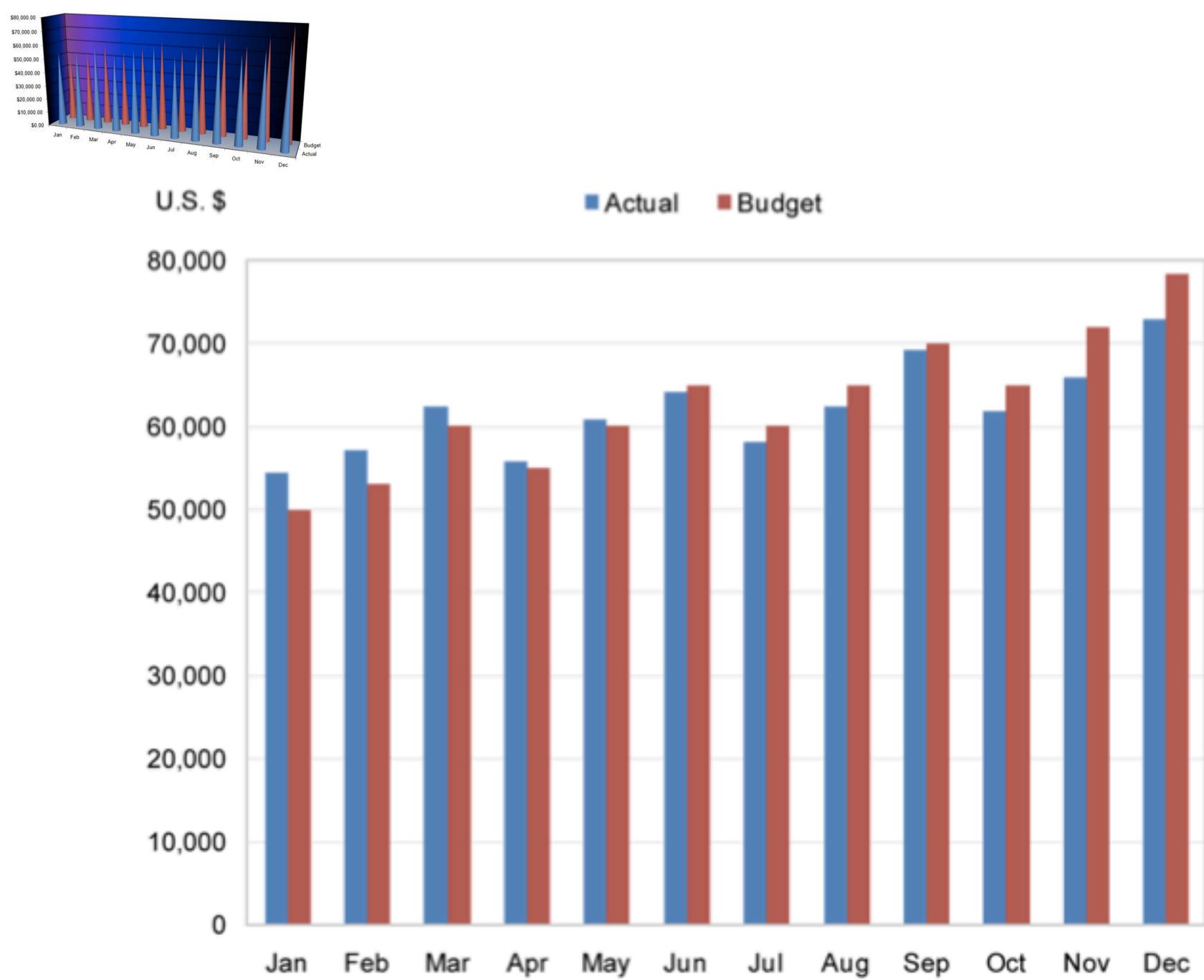


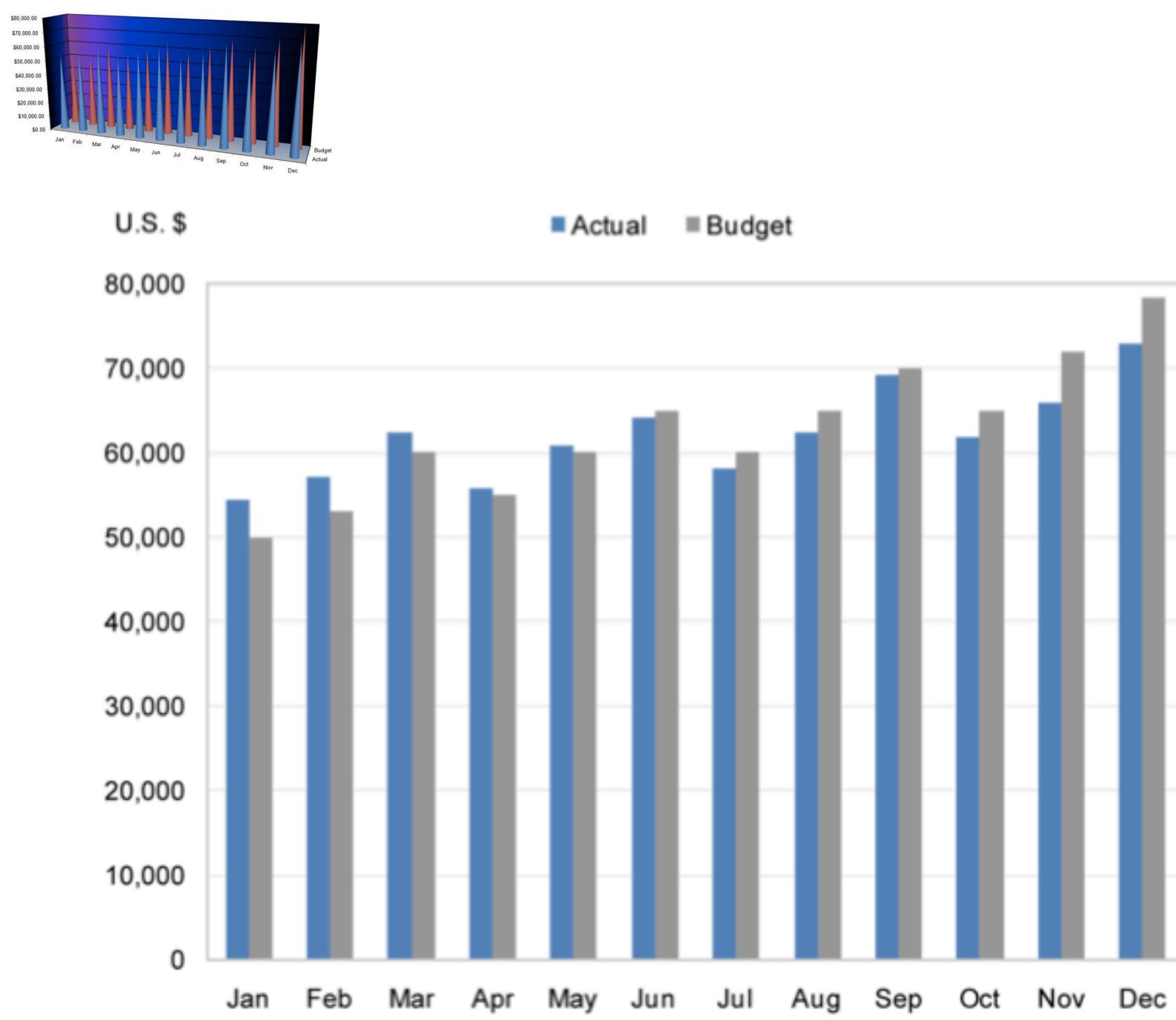


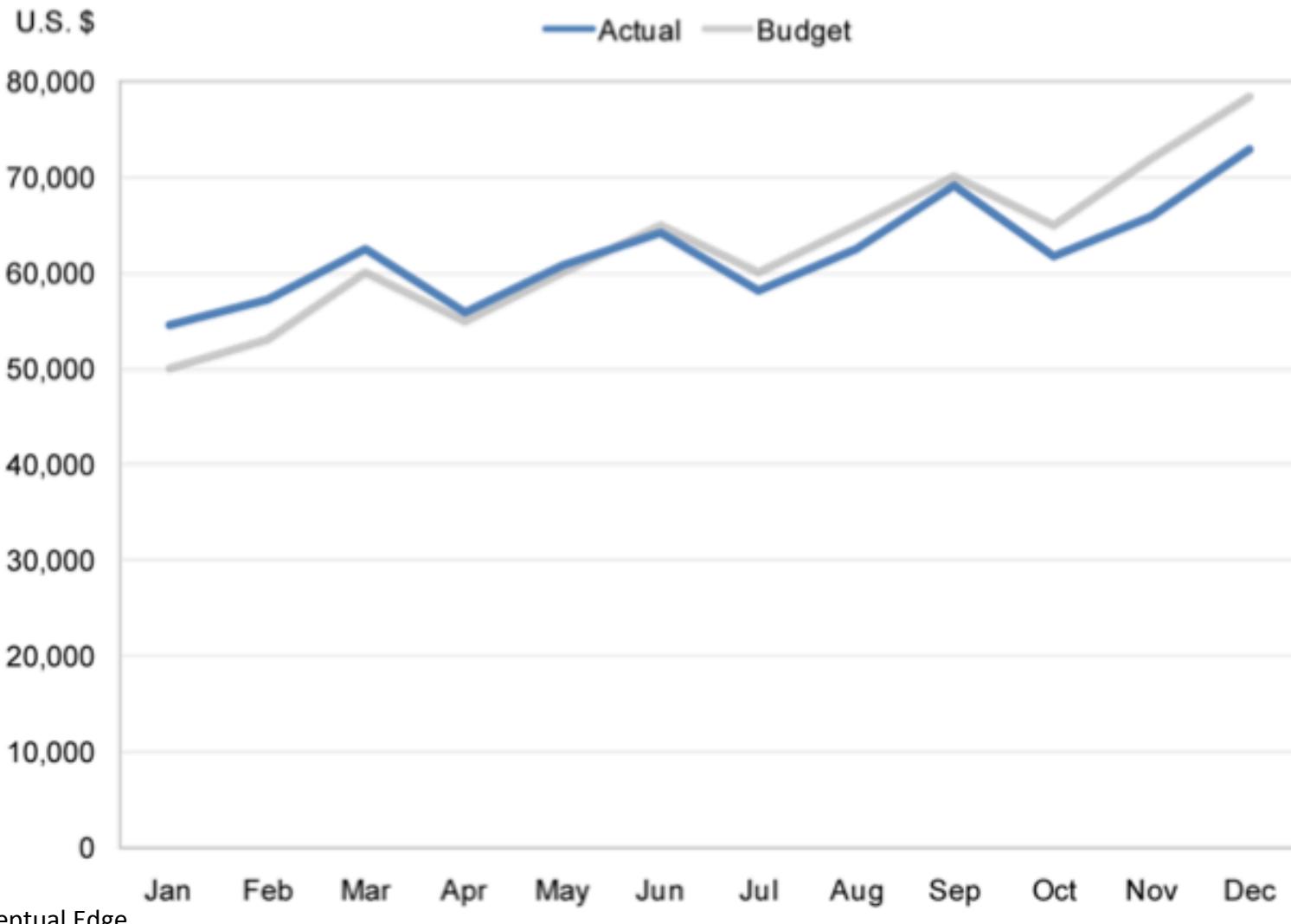
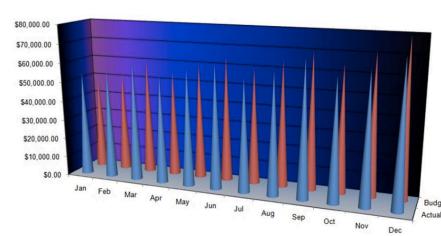


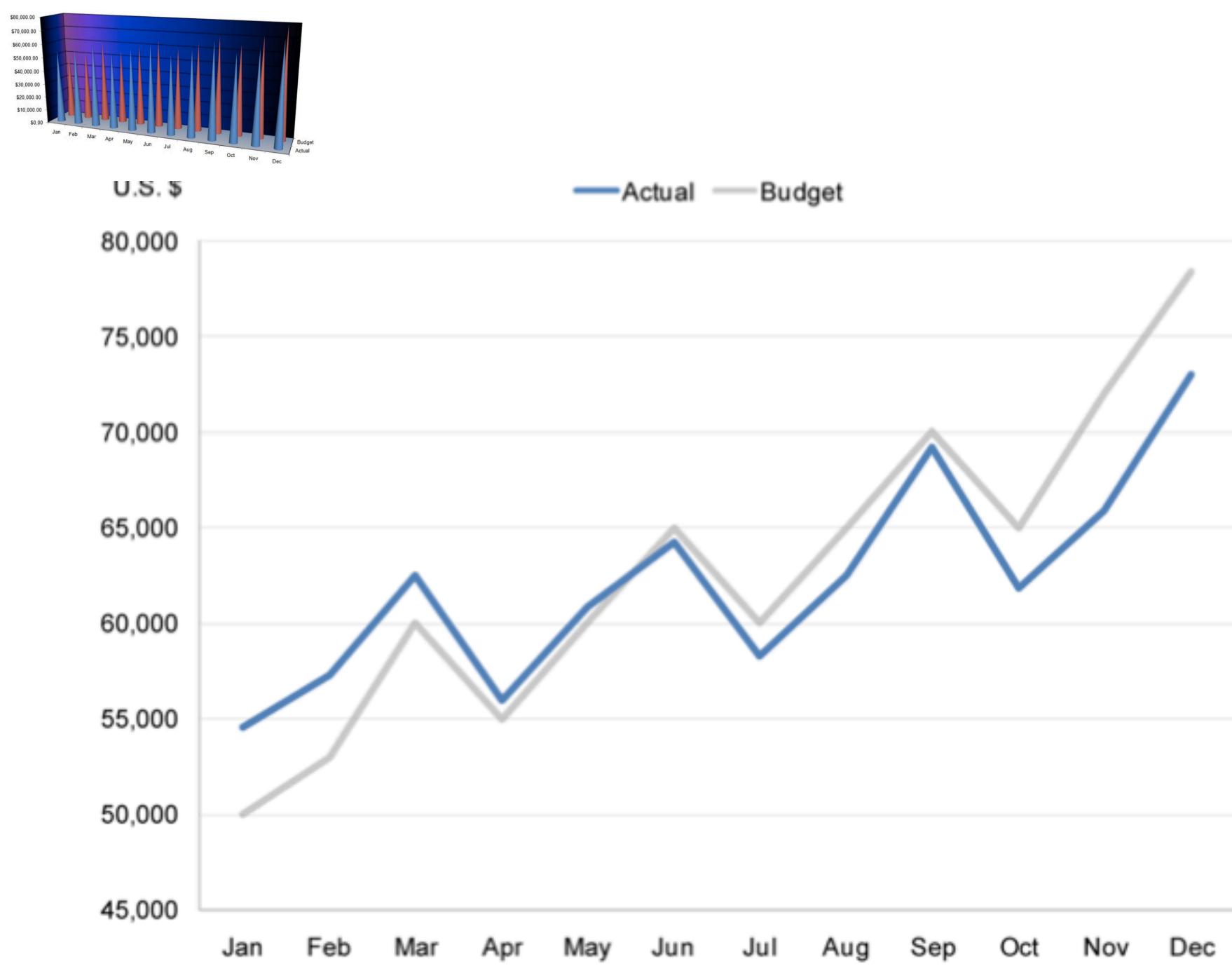
U.S. \$



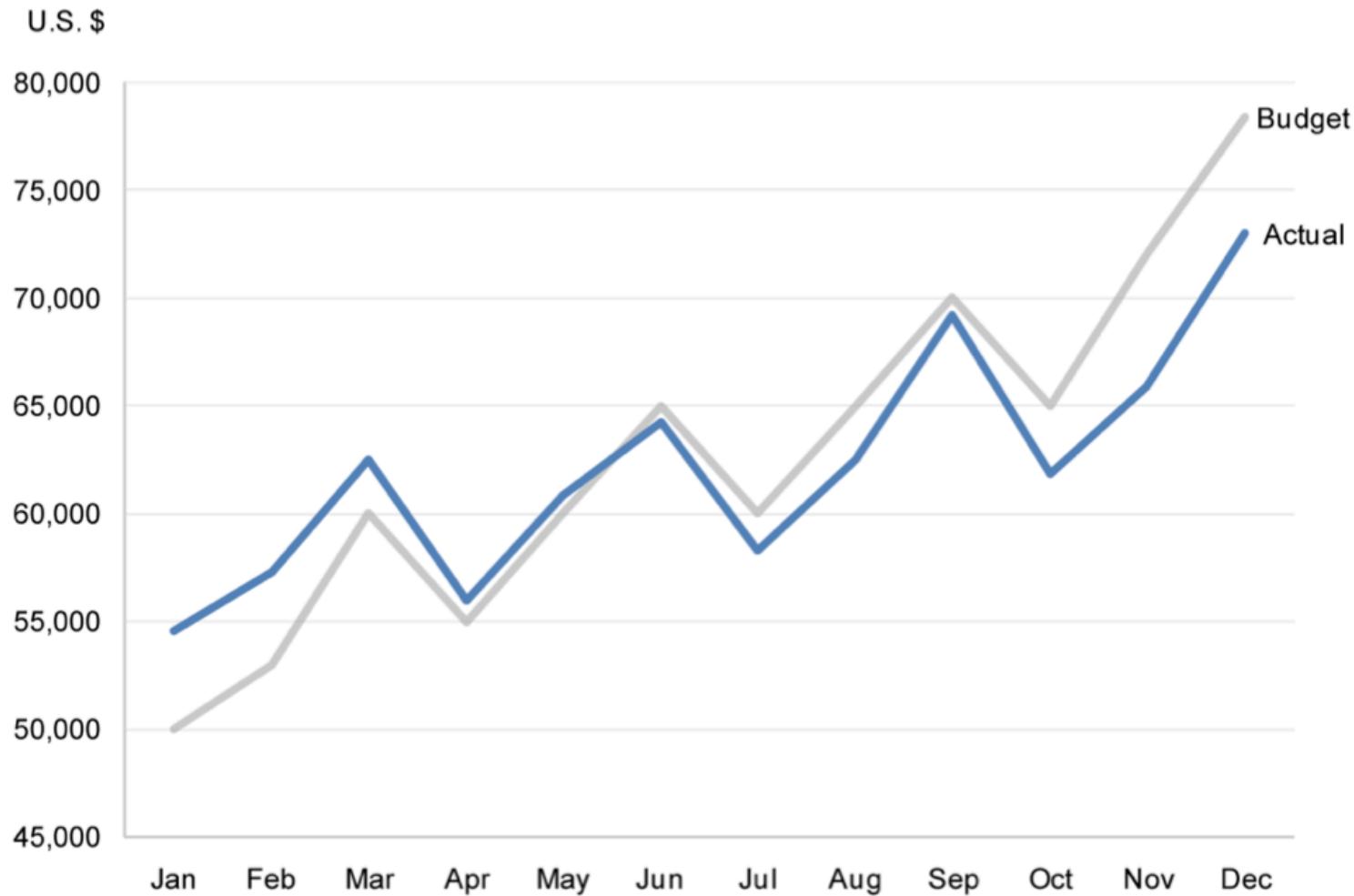




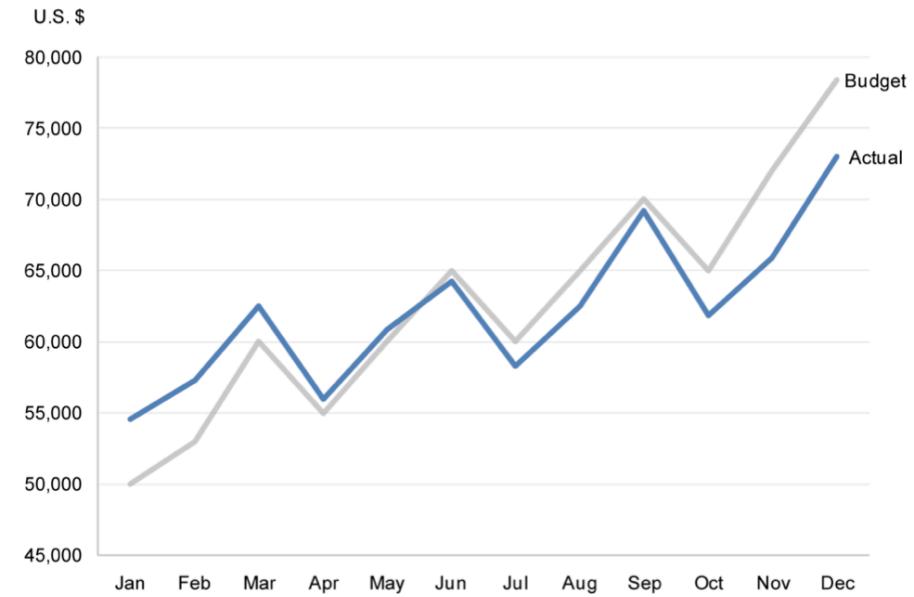
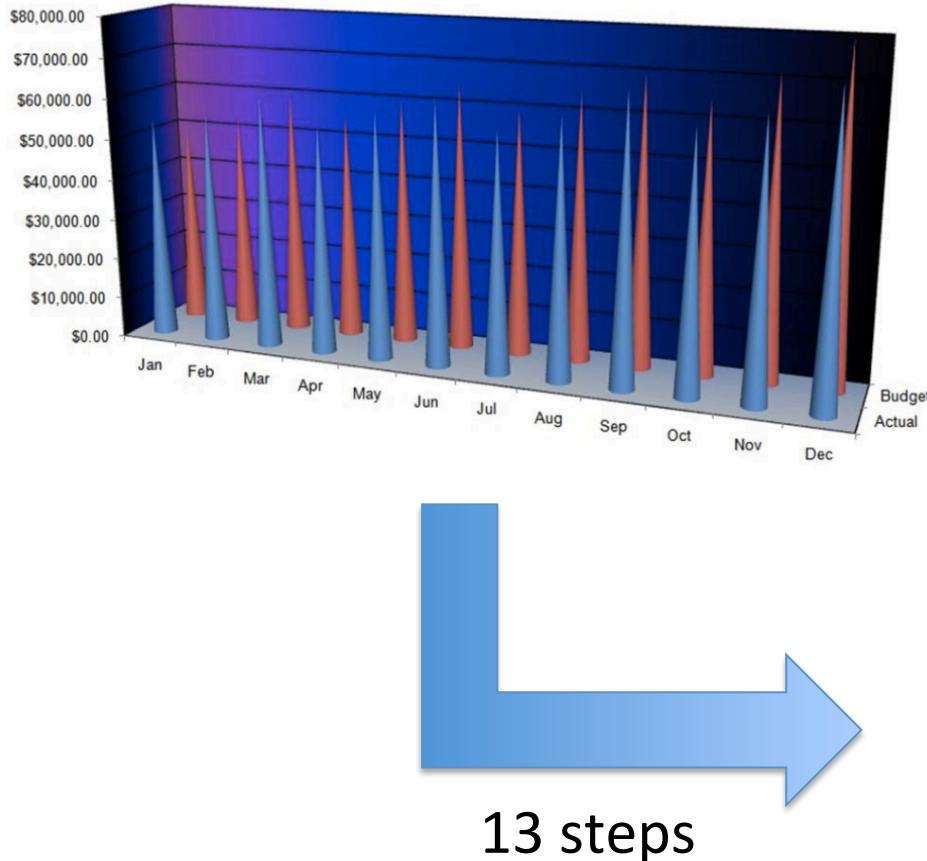




# Is it easier to compare them now?



# Try to compare actual expenses to the budget across time



# Now it is your turn...

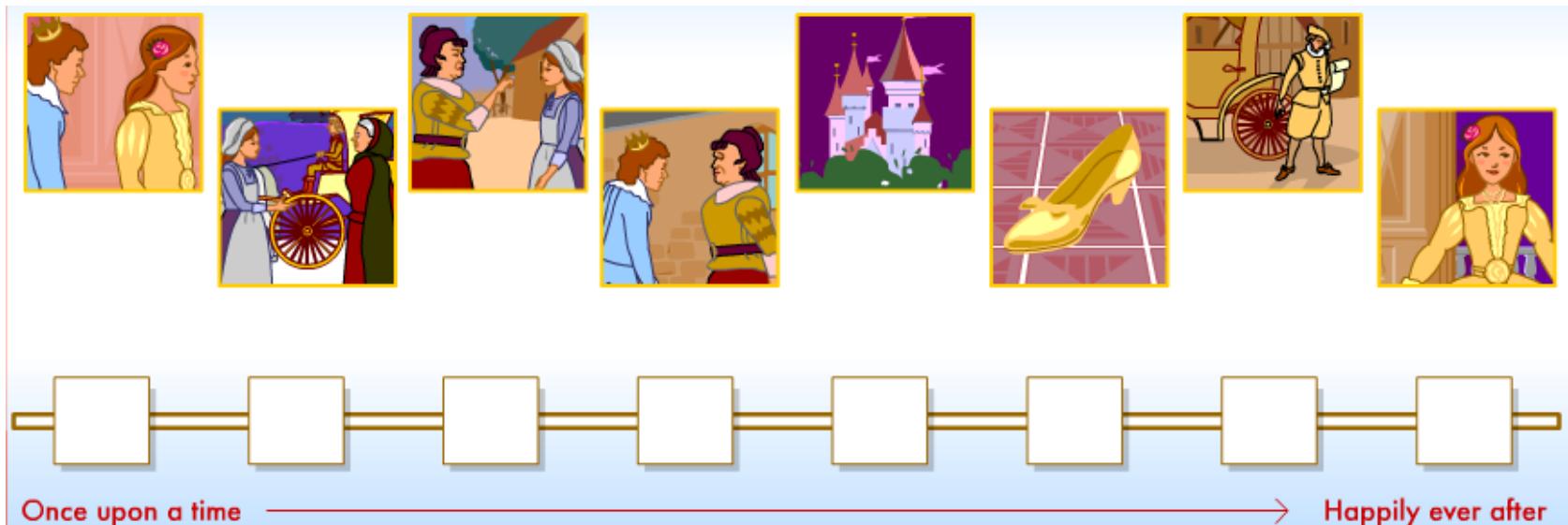
- Draw a chart that reflect your actual learning curve from this lecture versus your expected learning curve ☺

# In order to communicate your information effectively

- Determine your message and identify your data
- Determine if a table, graph, or combination of both is needed to communicate your message
- Determine the best means to encode the values
- Determine where to display each variable
- Determine if particular data should be featured, and if so, how

# Next

## Once upon a time



# Extra slides

# Graphs should

- Show the data
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- Be closely integrated with statistical and verbal descriptions of a data set

# Few's selection and design process

- Should particular data be featured, and if so, how?
  - use bright or dark colors (when using soft colors for everything else)
  - bars - place borders around only those bars that should be highlighted
  - lines - make lines to be highlighted thicker
  - points - make featured points larger or include fill color in them alone

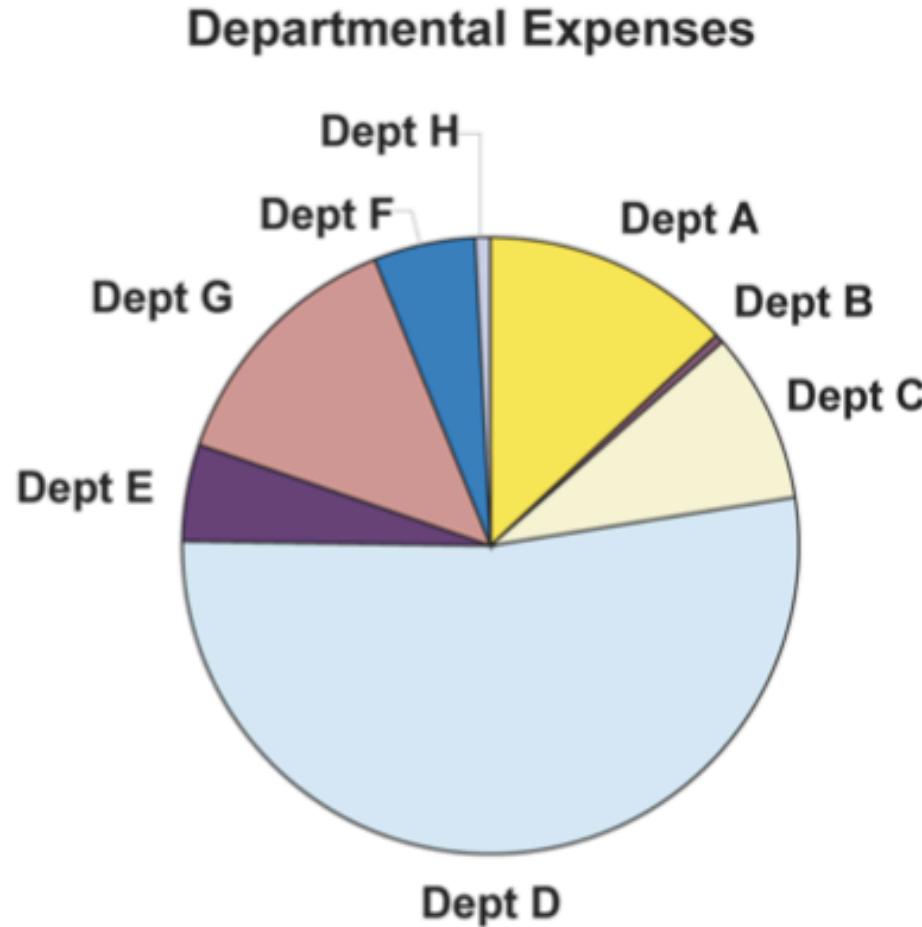
# Few's selection and design process

- Range of the quantitative scale?
  - bars - start from 0
  - lines, points - adjust scale so that it extends a little below the lowest data value and a little above the highest
  - bars - if you want to narrow the scale, switch from bars to points
- If a legend is required, where to place it?
  - the more directly you label the data, the better
  - bar - arrange labels to match the arrangement of bars
- Tick marks required for axis?
  - only necessary on quantitative scales
  - only major tick marks are necessary (5-10 usually enough)

# Few's selection and design process

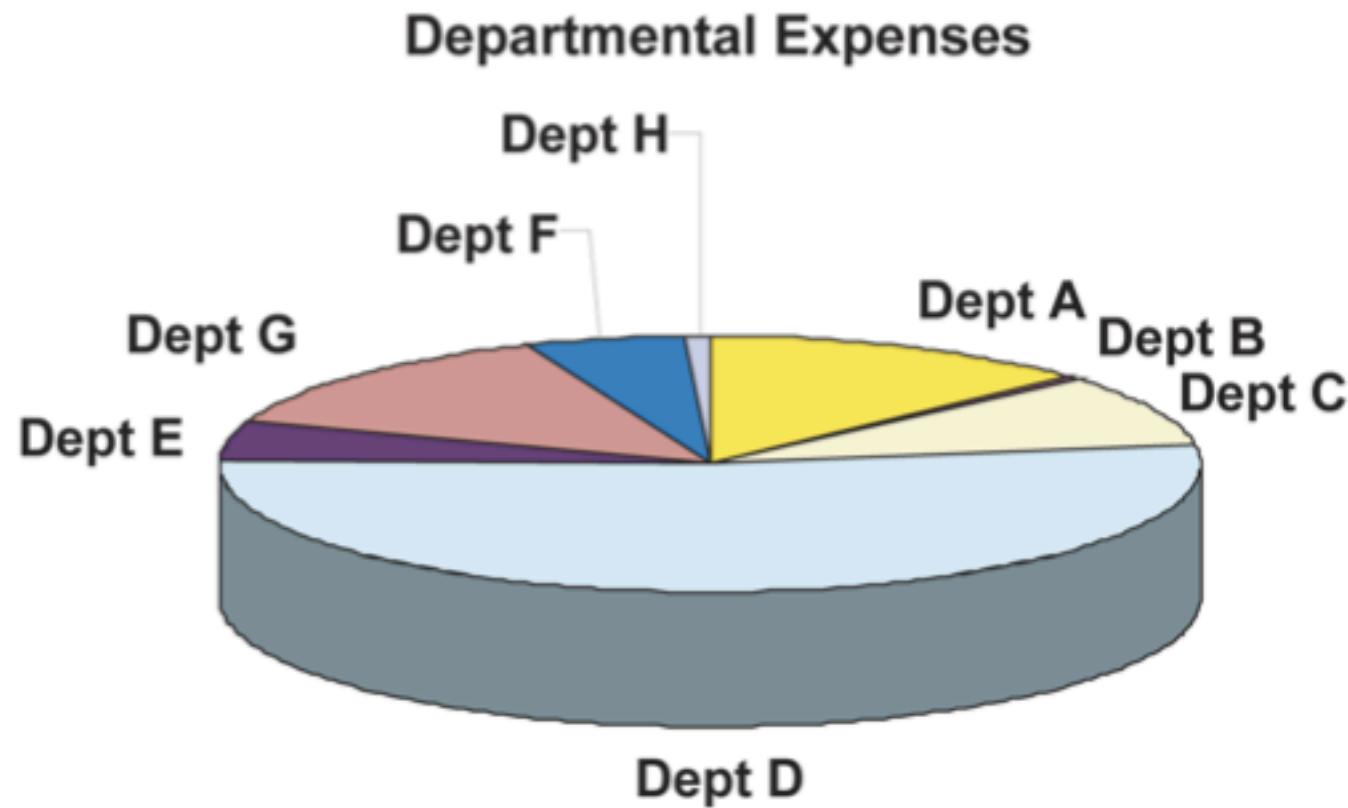
- Best location for the quantitative scale?
  - avoid placing scale on right side unless necessary
  - if quantitative scale ranges between positive and negative, place axis at 0 (but don't let labels interfere with data)
- Grid lines required?
  - values cannot be interpreted with the necessary degree of accuracy
  - subset of points in multiple related scatter plots must be compared
  - if used, make them barely visible
- What descriptive text is needed?
  - title
  - axes titles

# Is the message clear?

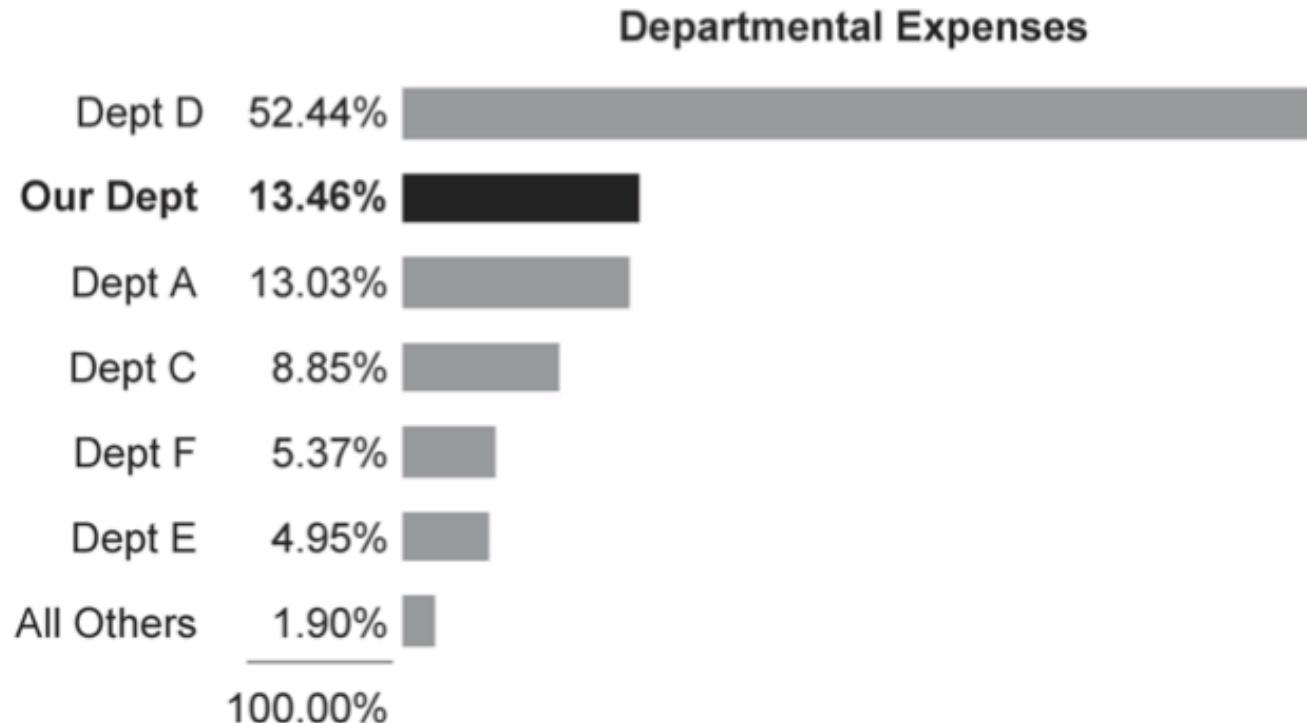


how Department G is doing regarding expenses compared to the other departments

# Does the addition of 3D improve this pie chart?



# This chart convey the message simply and elegantly



ICA12