DATA VISUALIZATION

DATA STEWARDS TRAINING #3

(YIPEE!)

November 14, 2017

Activity time!

What do you hope to get out of today's training?

MY GOALS FOR THE TRAINING

- Understand the basic principles of data viz (so you can use any tool)
- Develop a shared language and understanding for what makes a good data visualization
- Some tips for presenting
- Have fun!
- Note: Practice makes better ☺

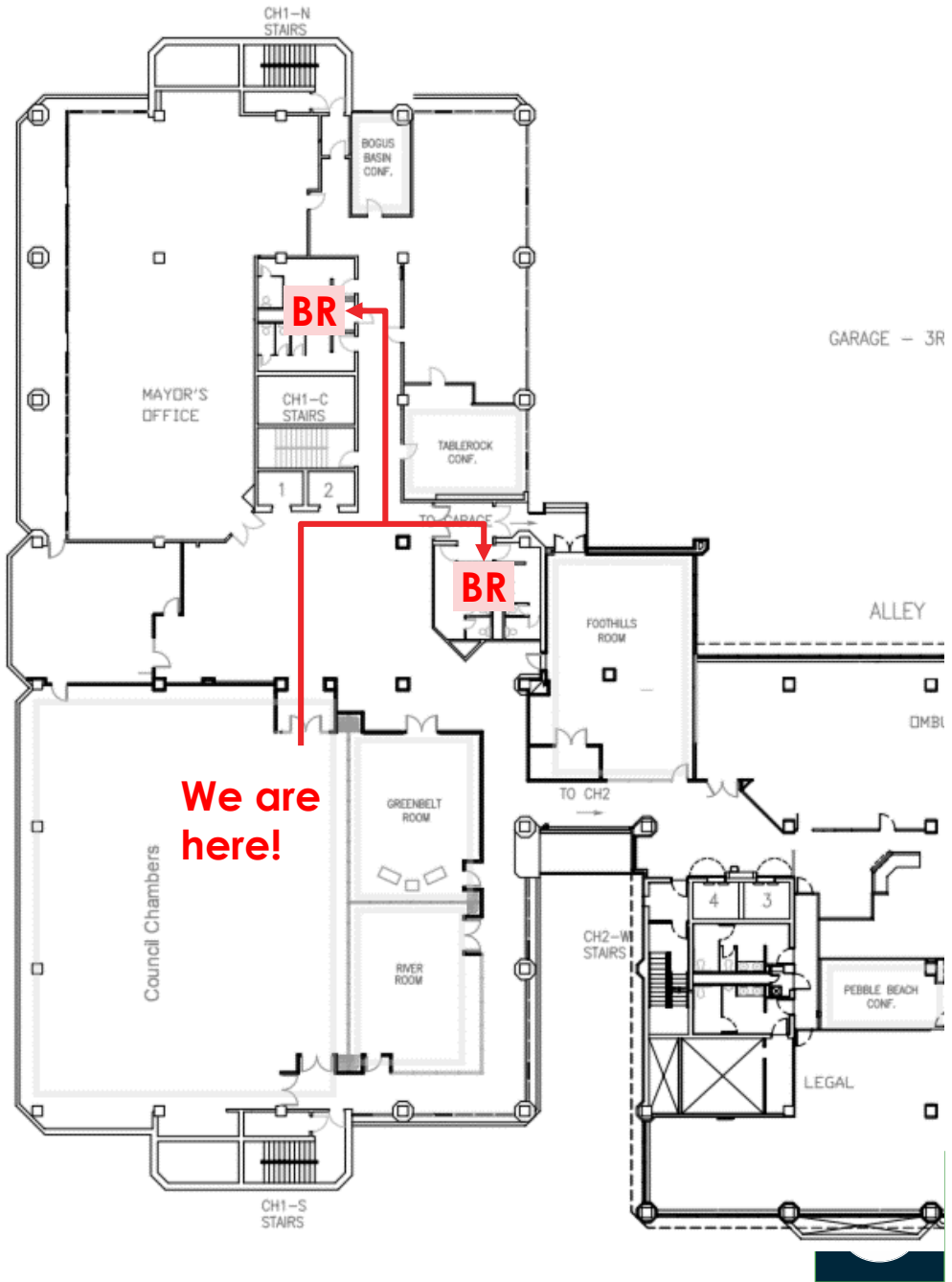
PLAN FOR THE DAY

- Intro and Context Setting
  - Why data viz?
  - What is good data viz?

- Principles of:
  - Data Visualization
  - Data Presentation

HOUSEKEEPING

- Feel free to stand and/or grab a snack
- Feel free to ask questions at any time (but I might not answer)
- Breaks are good!
- So are activities!
- Bathroom locations



CONTEXT SETTING

WHY DATA VIZ? | WHAT IS GOOD DATA VIZ?

WHY DATA VIZ?

It compresses information.

Name	Height	Past crime?	Age
Benicio	6'2	Y	48
Kevin P.	5'5	N	58
Stephen	5'10	Y	49
Gabriel	5'10	Y	65
Kevin S.	5'10	Y	56
Ryan	6'2	N	39
Ben	6'4	N	43
Matt	5'10	Y	45
George	5'11	N	54

VS

JOHNS HOPKINS UNIVERSITY | GovEx

Visual provided by GovEx

WHY DATA VIZ?

Bathroom Directions:

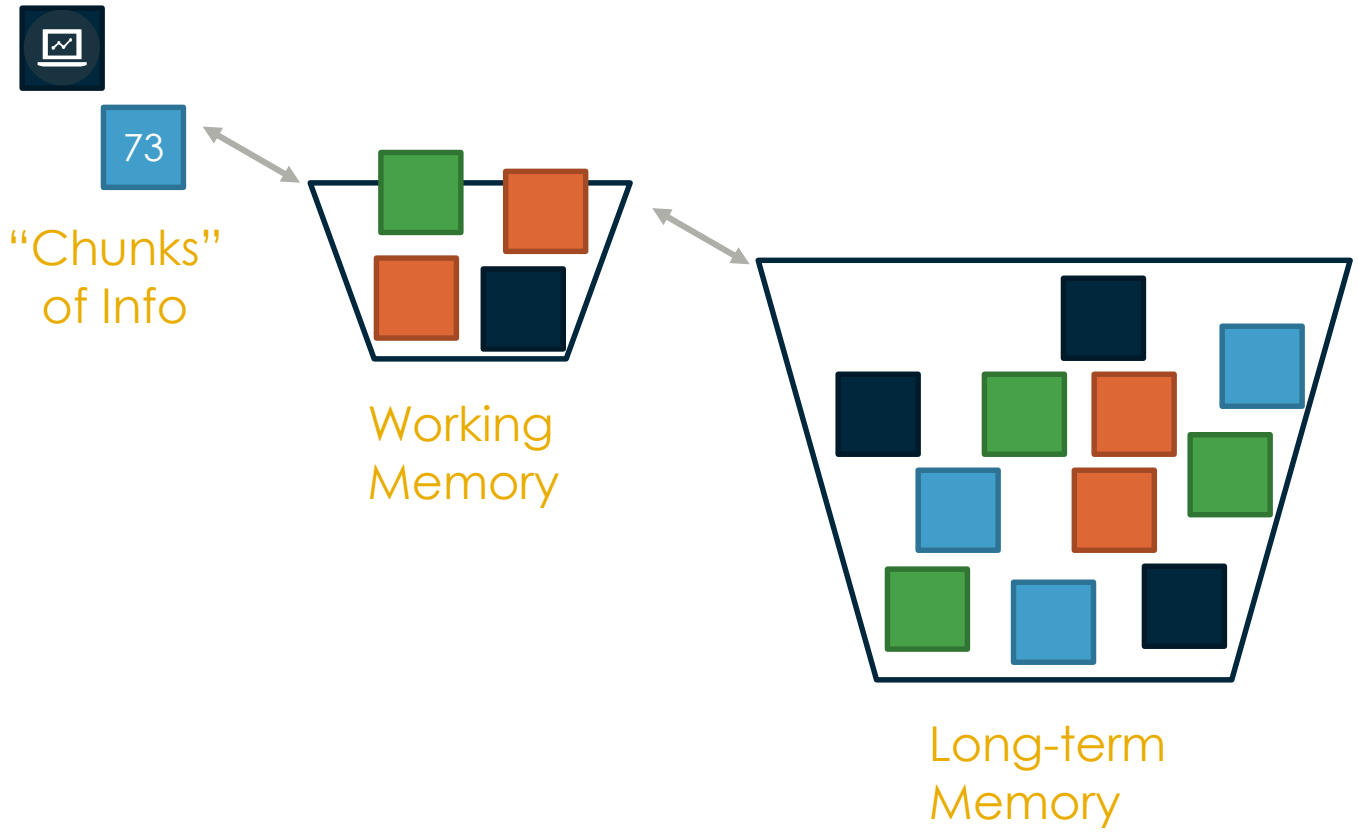
There are two nearby bathrooms.

1. Exit the main doors of Council Chambers and walk down the hallway to the right of the elevators. Bathrooms are on the left.

2. Exit the main doors of Council Chambers, and go through the first double doors on the far side of the lobby. The bathrooms are on the right before the next set of double doors.

VS

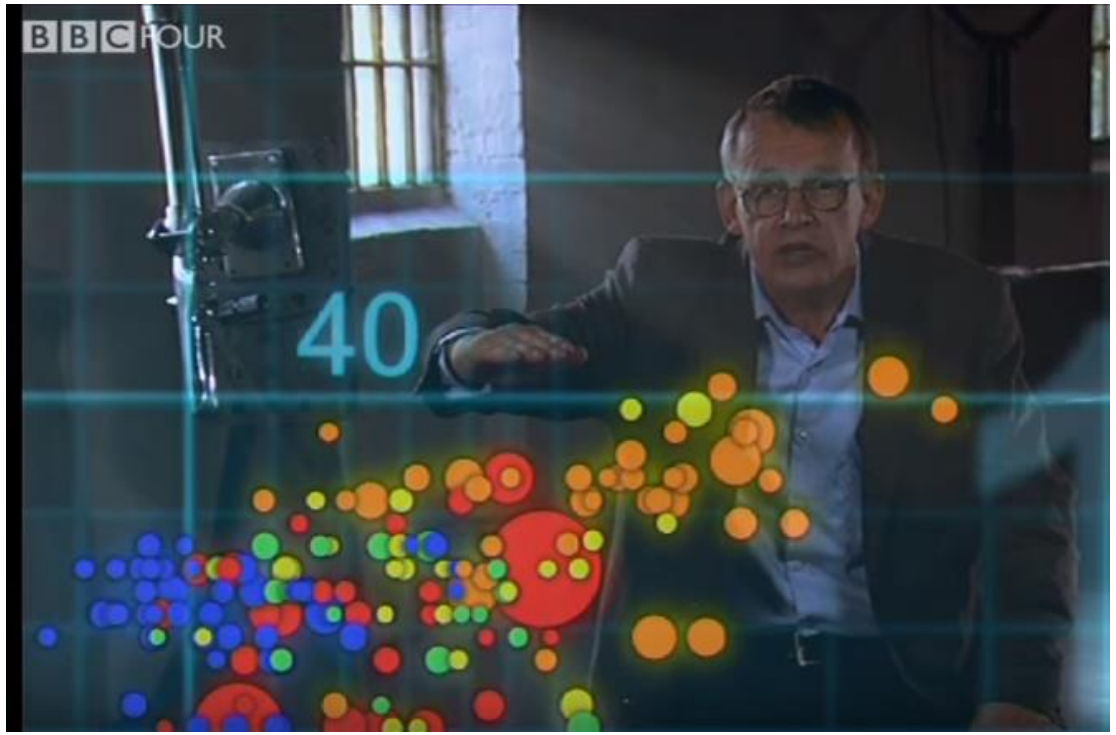
WHY DATA VIZ?



Our working memory is limited.

And we are pattern-seekers by nature.

HANS ROSLING IS FUN TO WATCH



What does Hans do well?

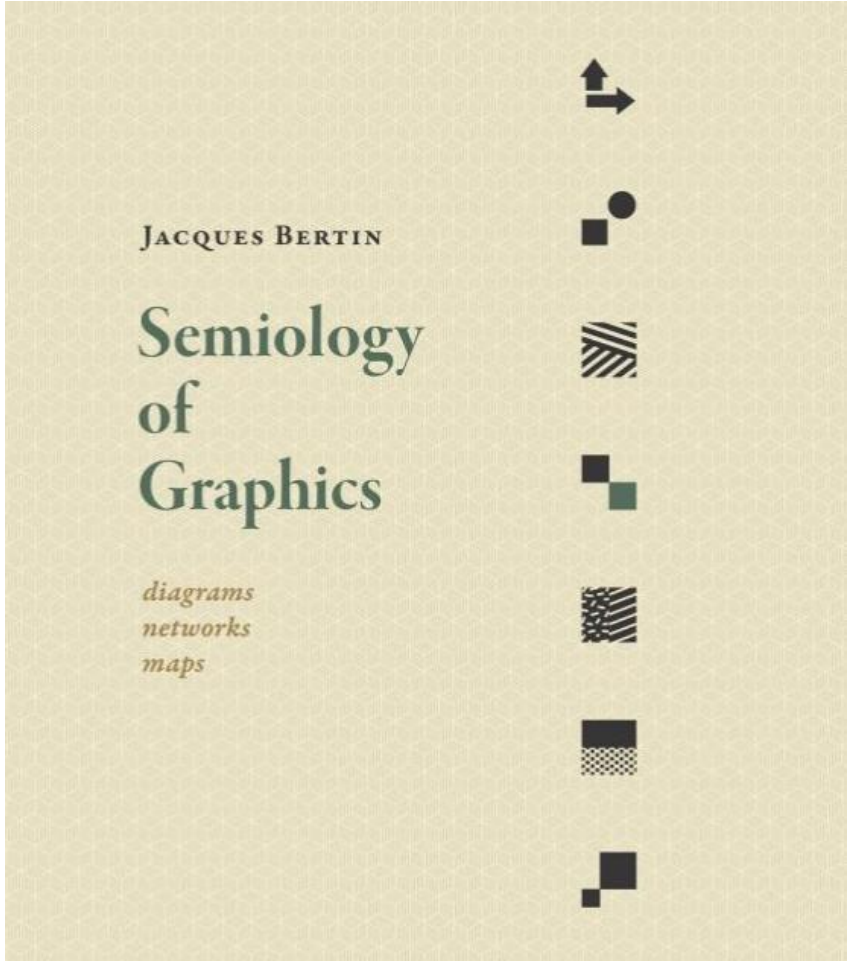
What does he do poorly?

WHAT IS GOOD DATA VIZ?

Principle of Expressiveness – show what you need to, no more, no less

Principle of Effectiveness – use the most efficient method available to visualize your information

Semiologie Graphique
By Jacques Bertin (1967)



WHAT IS GOOD DATA VIZ?

Excellence means…

"…complex ideas communicated with clarity, precision and efficiency."

"…greatest number of ideas in the shortest time with the least ink in the smallest space."

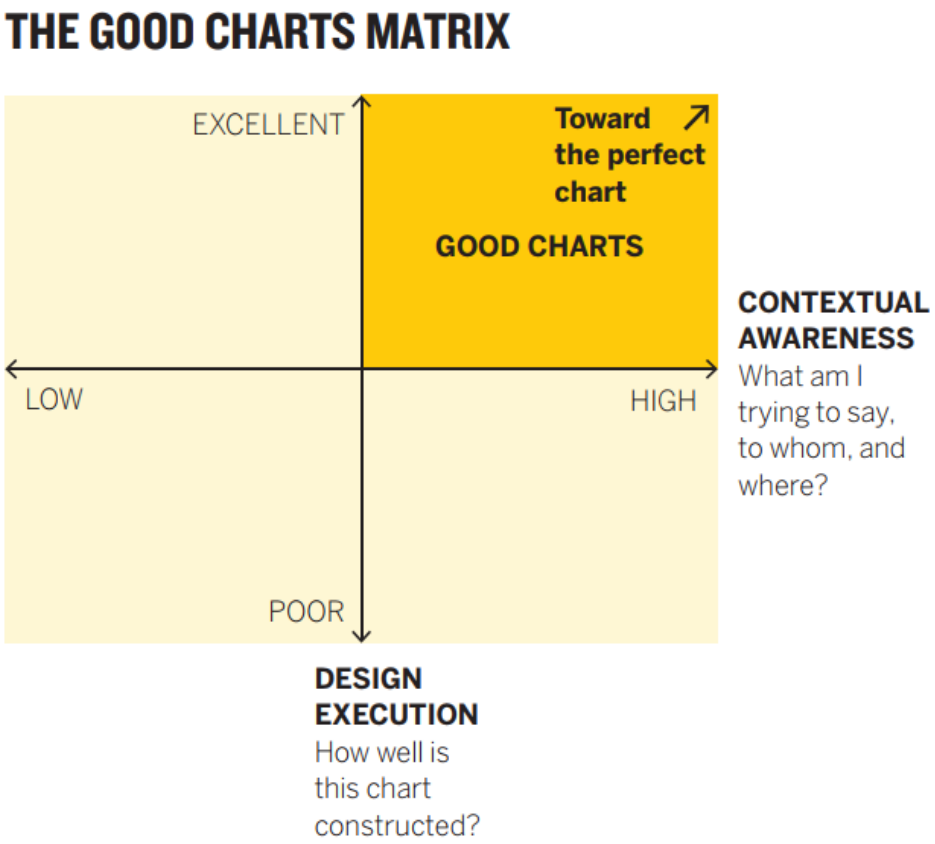
"…requires telling the truth about data."

The Visual Display of Quantitative Information
By Edward Tufte (1983)

WHAT IS GOOD DATA VIZ?

"A perfectly relevant visualization that breaks a few presentation rules is far more valuable – it's *better* – than a perfectly executed, beautiful chart that contains the wrong message, or fails to engage its audience."

Good Charts
by Scott Berinato
*quote and chart



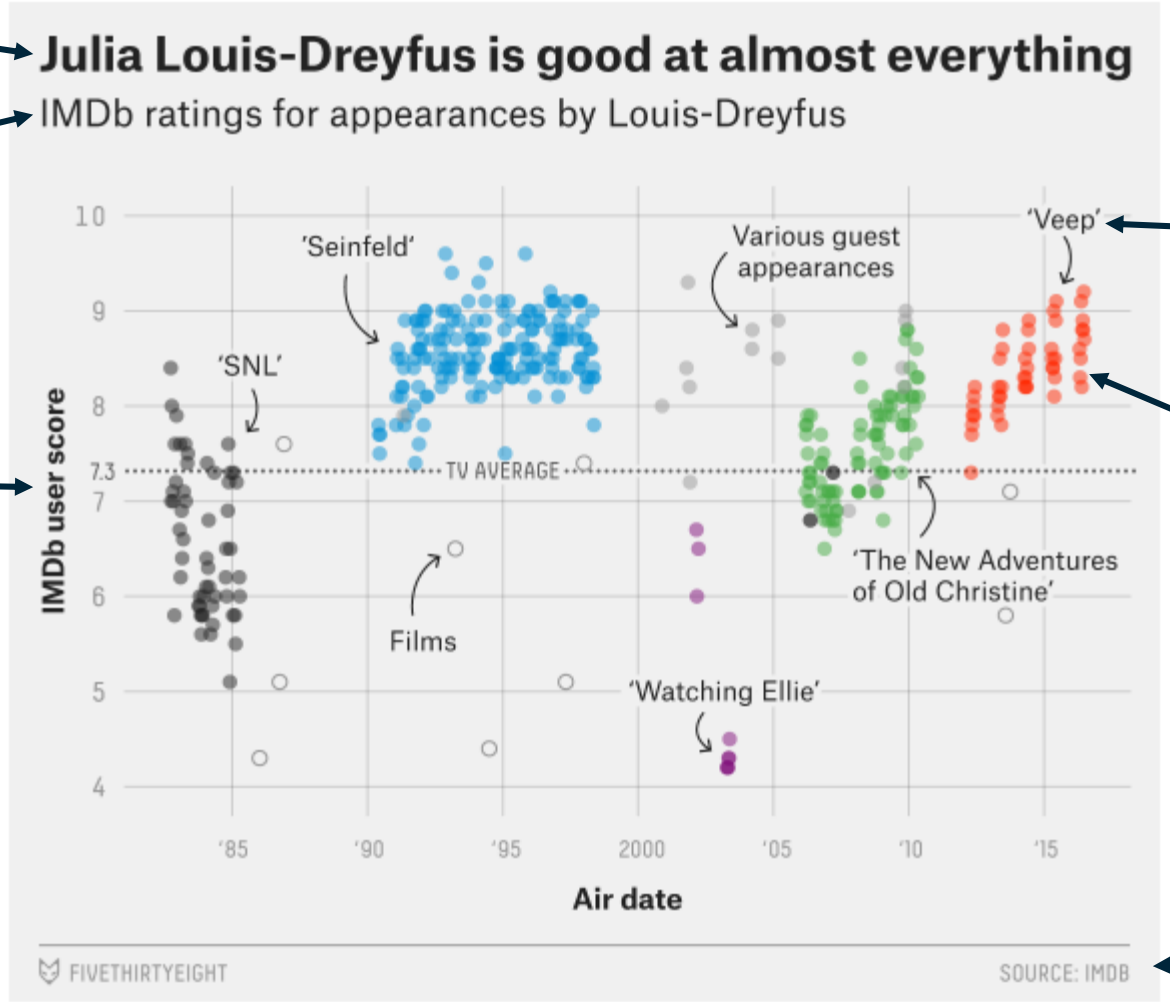
DATA VIZ PRINCIPLES

ANATOMY OF A DATA VIZ

Title

Subtitle

Axes Labels



Direct labels

Use of color

Source

Not shown here:

- Legend
- Data Labels
  (i.e. values)

DATA VIZ PRINCIPLES

#1: **KNOW YOUR MESSAGE**

PRINCIPLE #1:

KNOW YOUR MESSAGE

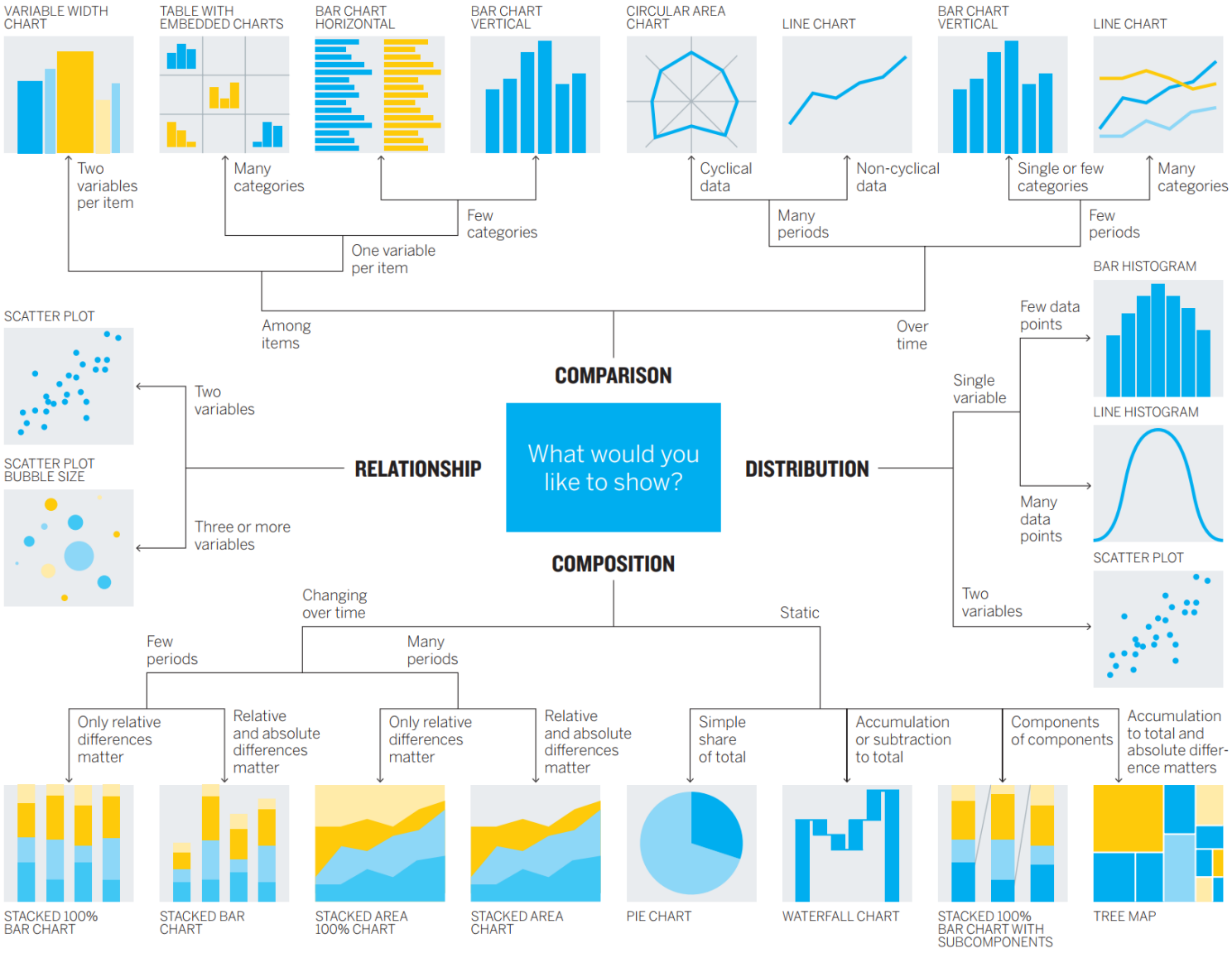
I want to convince _____ that _____.

I want to show that _____.

DATA VIZ PRINCIPLES

#2: **USE THE RIGHT CHART**

ABELA'S CHART TYPE HIERARCHY



SOURCE: ANDREW V. ABELA

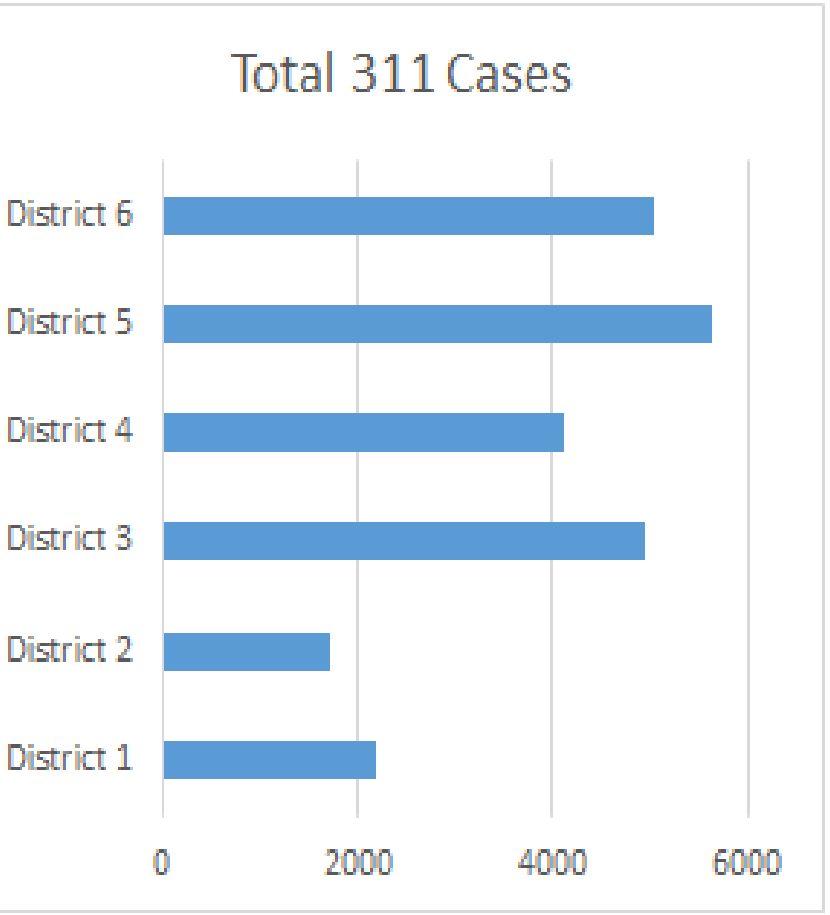
PRINCIPLE #2:

USE THE RIGHT CHART

Thanks to GovEx for the following slides!

Bar and Column Chart

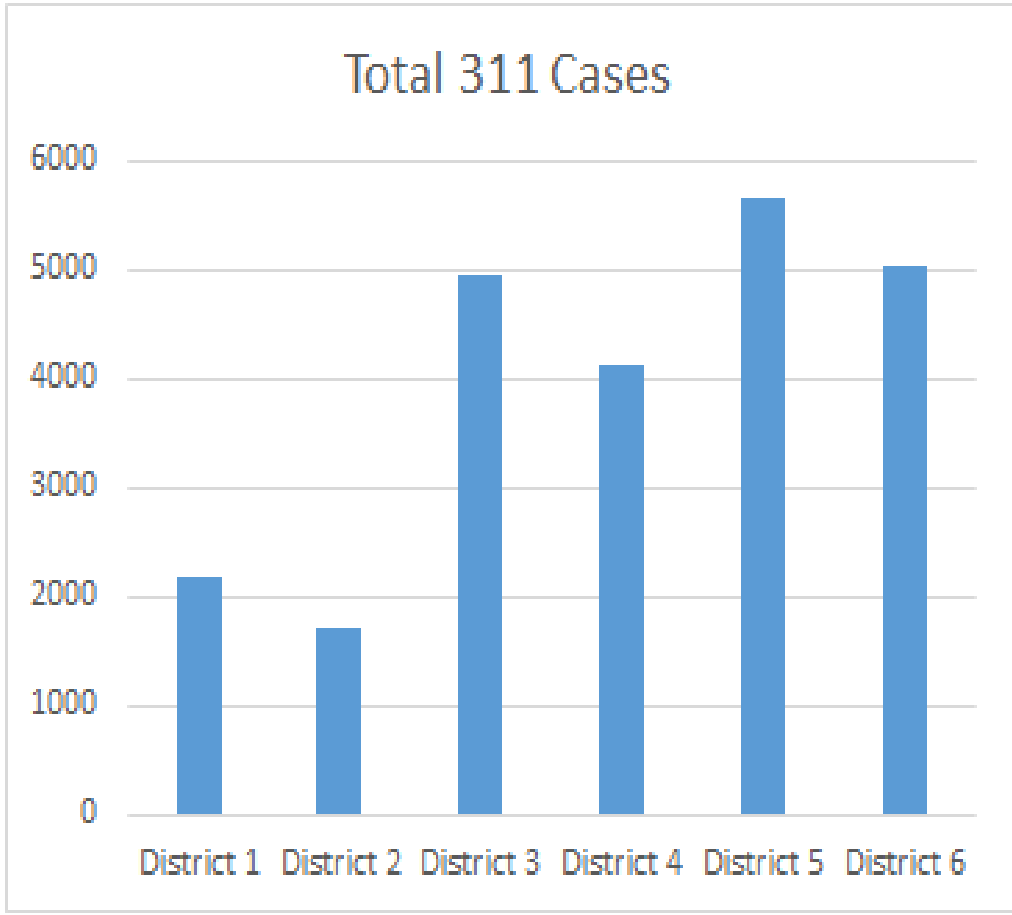
Reveal
- Describe findings
- Identify solutions



Note: Displays data using bars of same width for categories of observations.

Suitable for: Comparing numeric data across categories and revealing high and low points of information at a glance.

Example: Total 311 cases by districts



Highlight Table

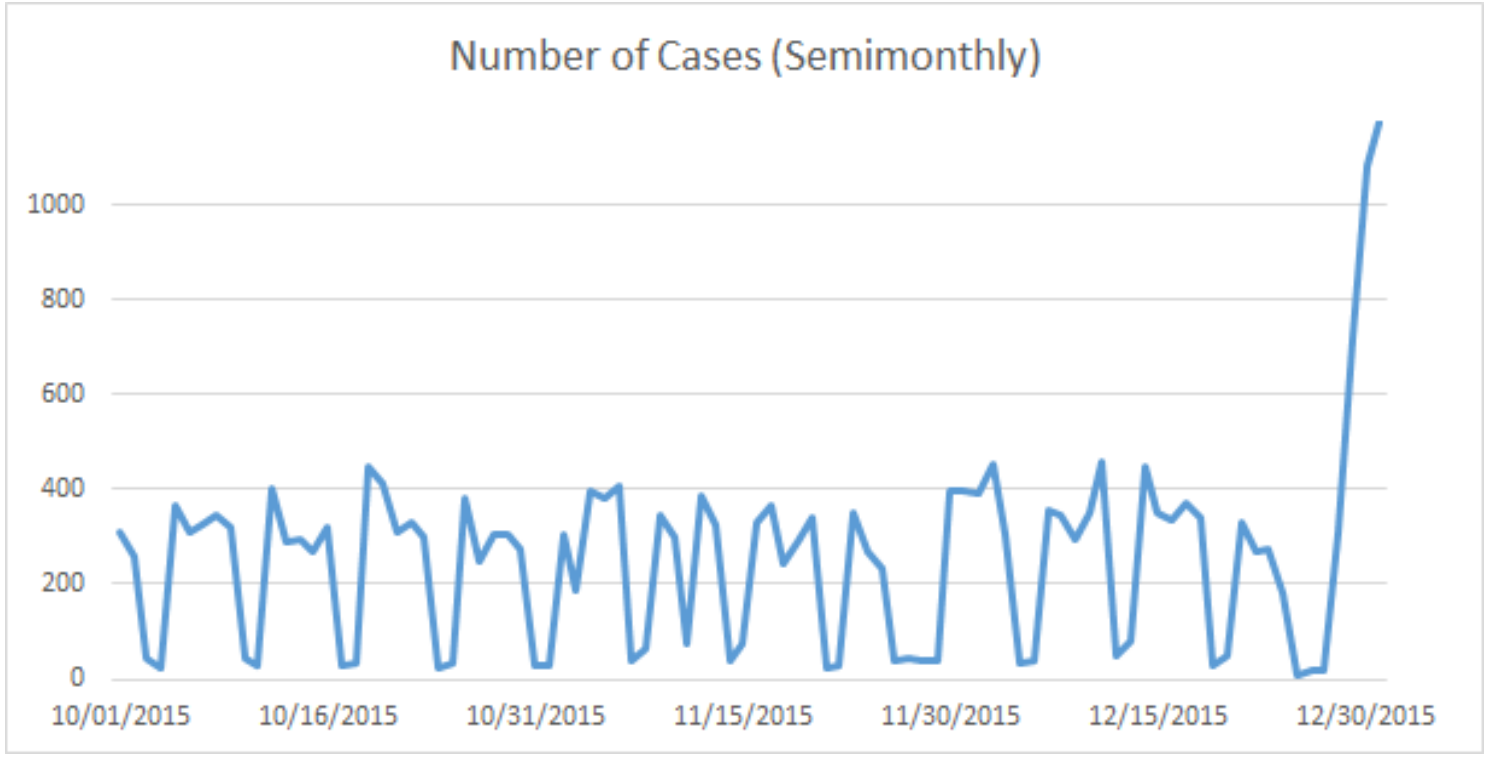
Department	# of Cases
Health	752
NHS	6,847
Parks and Rec	1,116
Public Works	10,274
Water Services	3,585
All other departments	1,086

Note: A table of data highlighted with different colors to provide additional details to unique variables.

Suitable for: Visualizing data at a glance to identify extreme variables and other information of interest.

Example: Number of cases by department.

Line Chart



Reveal
- Describe findings
- Identify solutions

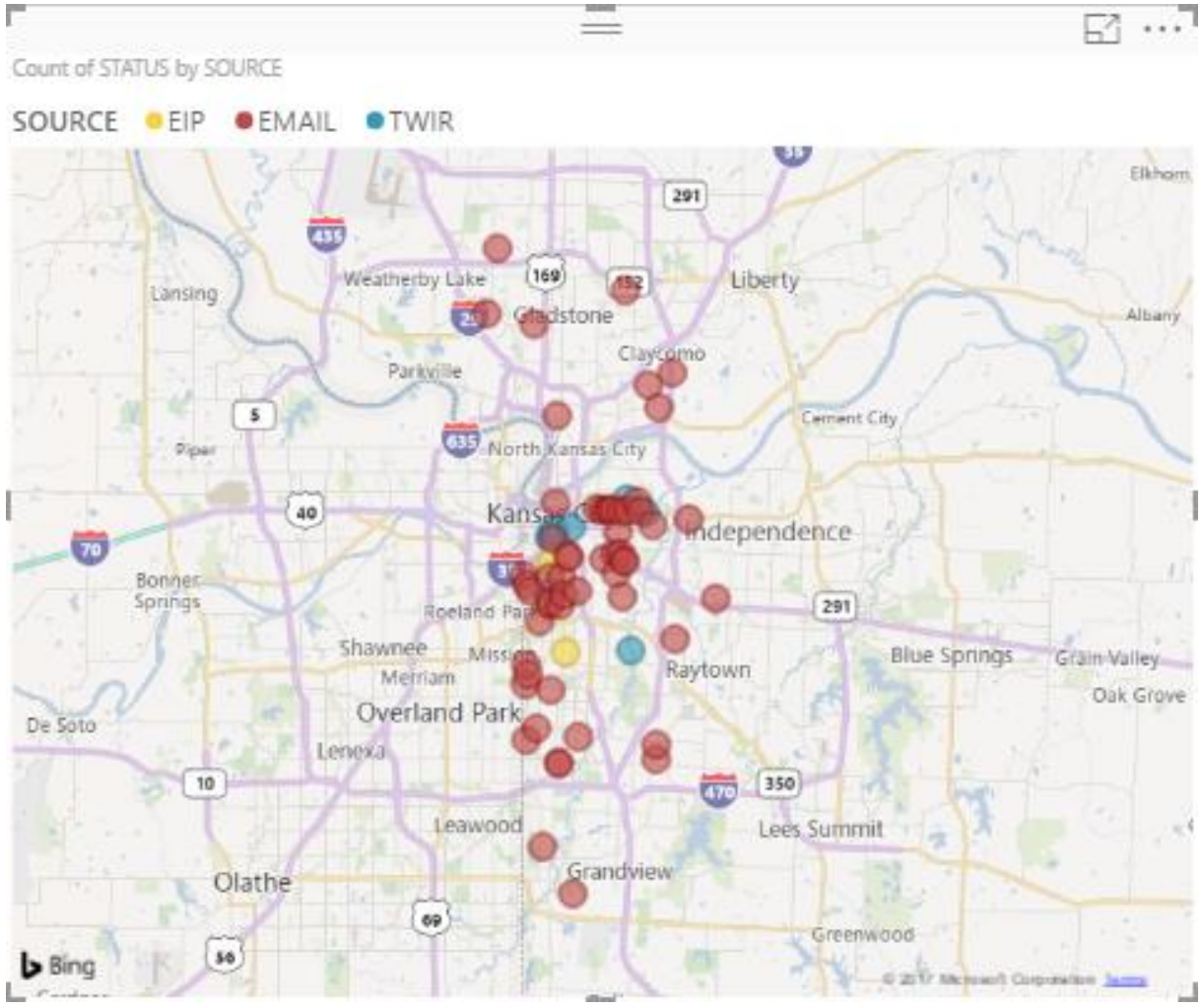
Note: Created by connecting a series of data points with a straight line.

Suitable for: Visualizing trends in data over a period of time.

Example: Showing trend of Number of cases over time.

Bubble Chart

Reveal
- Describe findings
- Identify solutions

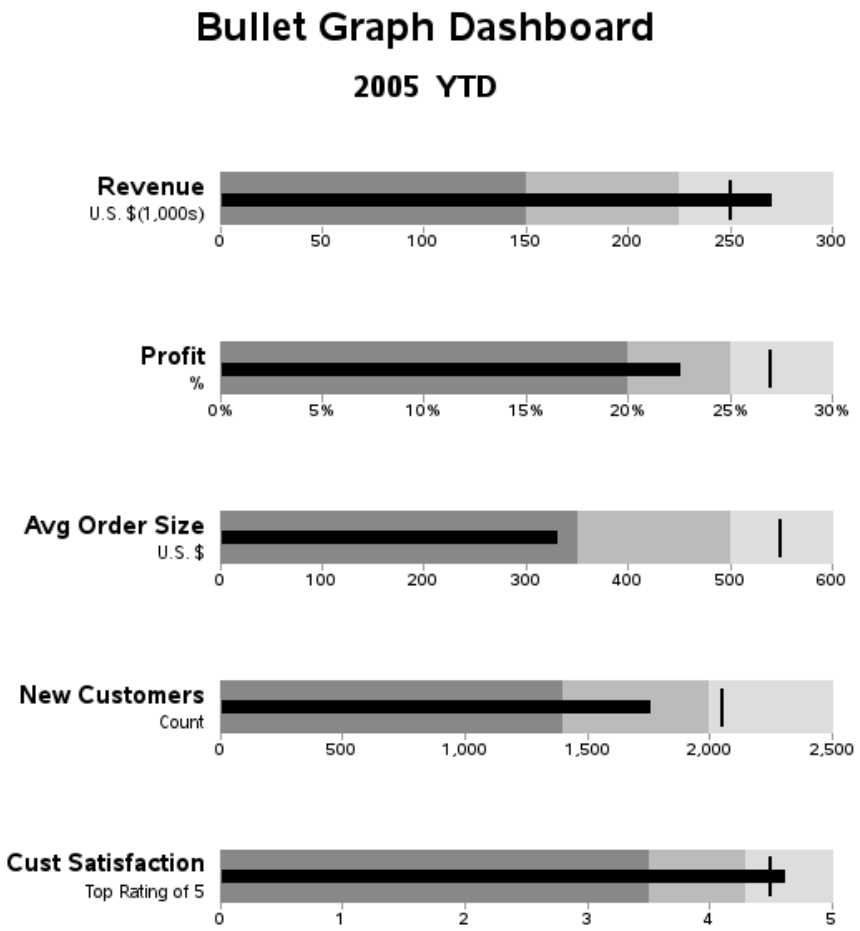


Note: A variation of a scatter plot that displays three dimensions of data.

Suitable for: Comparing relationships between data objects and showing the concentration of data along two axes.

Example: Source of 311 call by location

Bullet Chart



Reveal

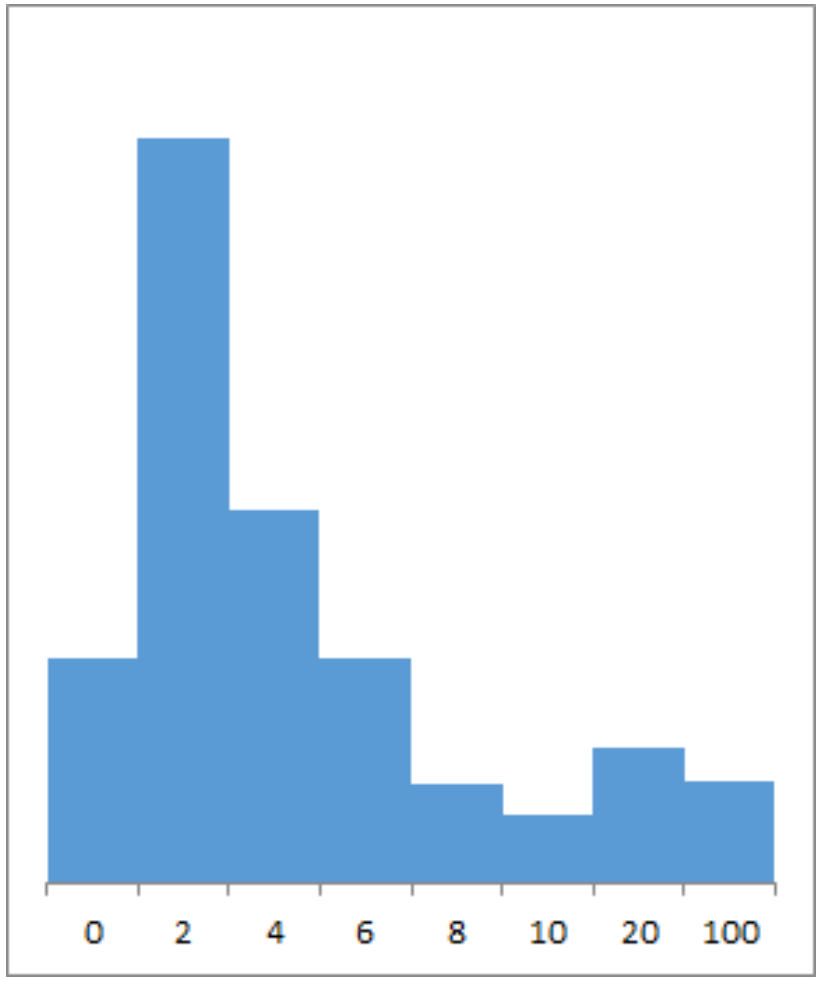
- Describe findings
- Identify solutions

Note: A variation of the bar chart designed to gauge the performance of a system.

Suitable for: Tracking progress and evaluating performance of a metric against a goal.

Example: Actual vs. target/budget

Histogram Chart



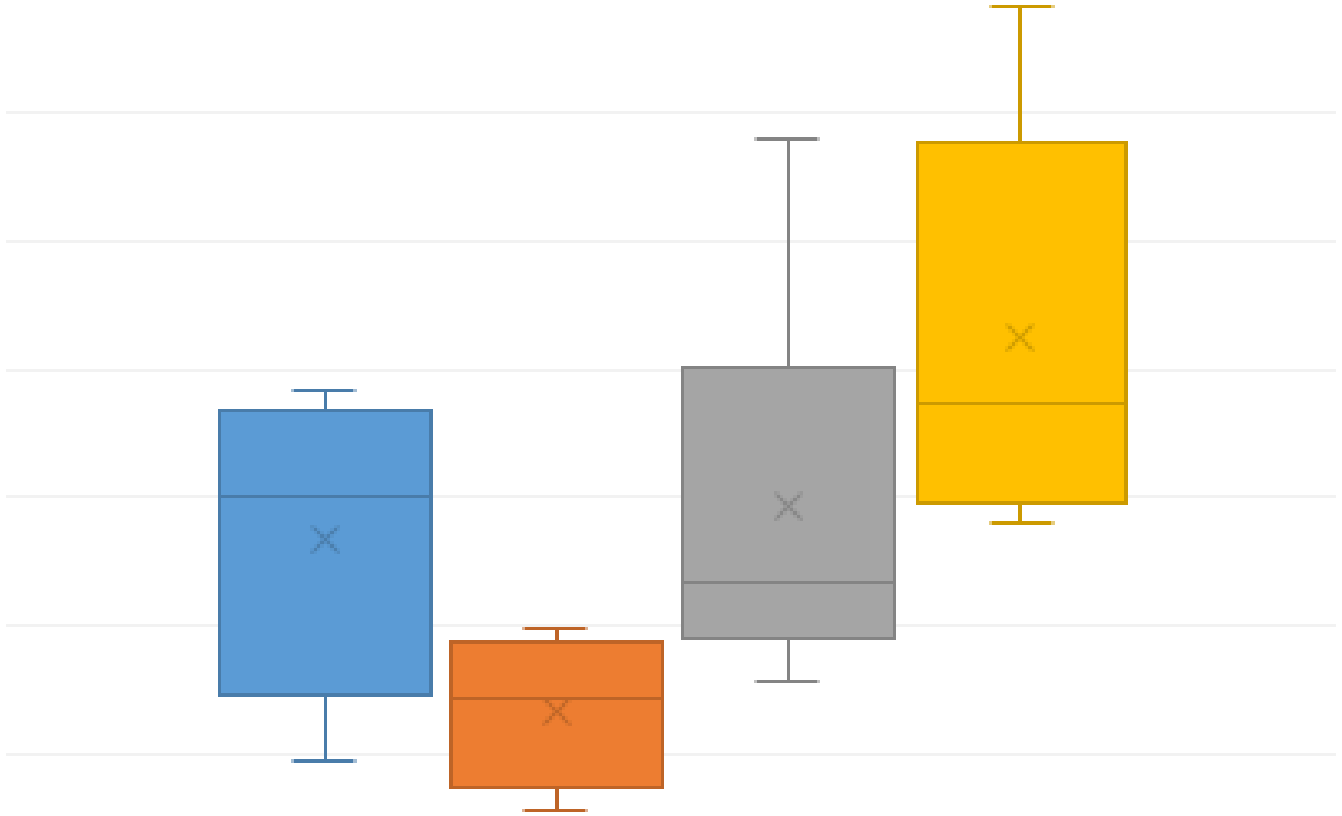
Reveal
- Describe findings
- Identify solutions

Note: Displays information using rectangles of equal width to show the frequency of data in consecutive numerical intervals of equal class size.

Suitable for: Understanding how data is distributed across different groups.

Example: Frequency of Days To Close.

Box and Whisker Plot



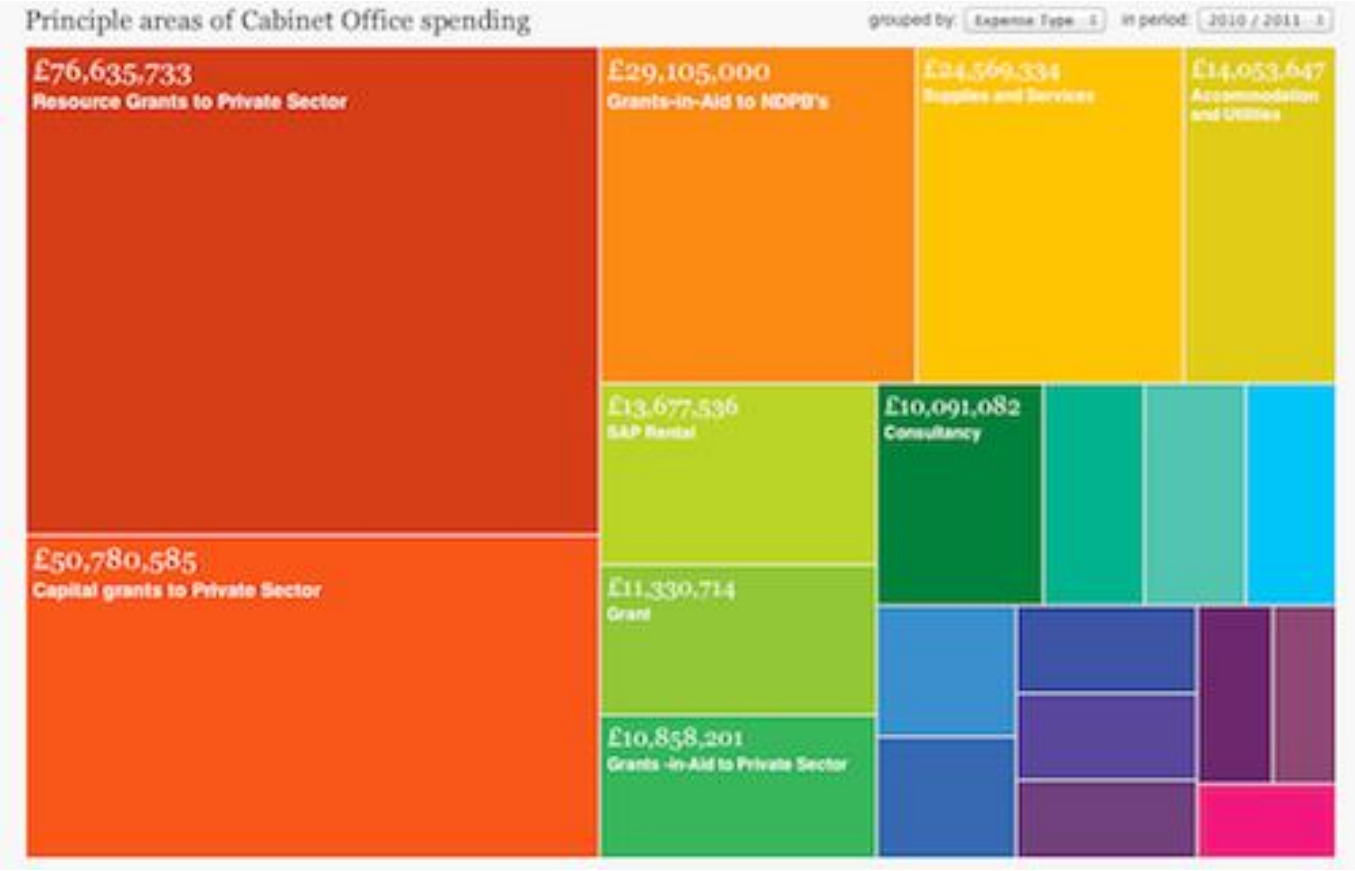
Reveal
- Describe findings
- Identify solutions

Note: Rectangular figure with whiskers displayed on a cartesian coordinate showing the distribution of the data in terms of: minimum, maximum, median, and quartiles.

Suitable for: Understanding the distribution of data sets to see skewness, outliers, and concentrations.

Example: Average Number of Days To Close in different departments

Tree Map



Reveal

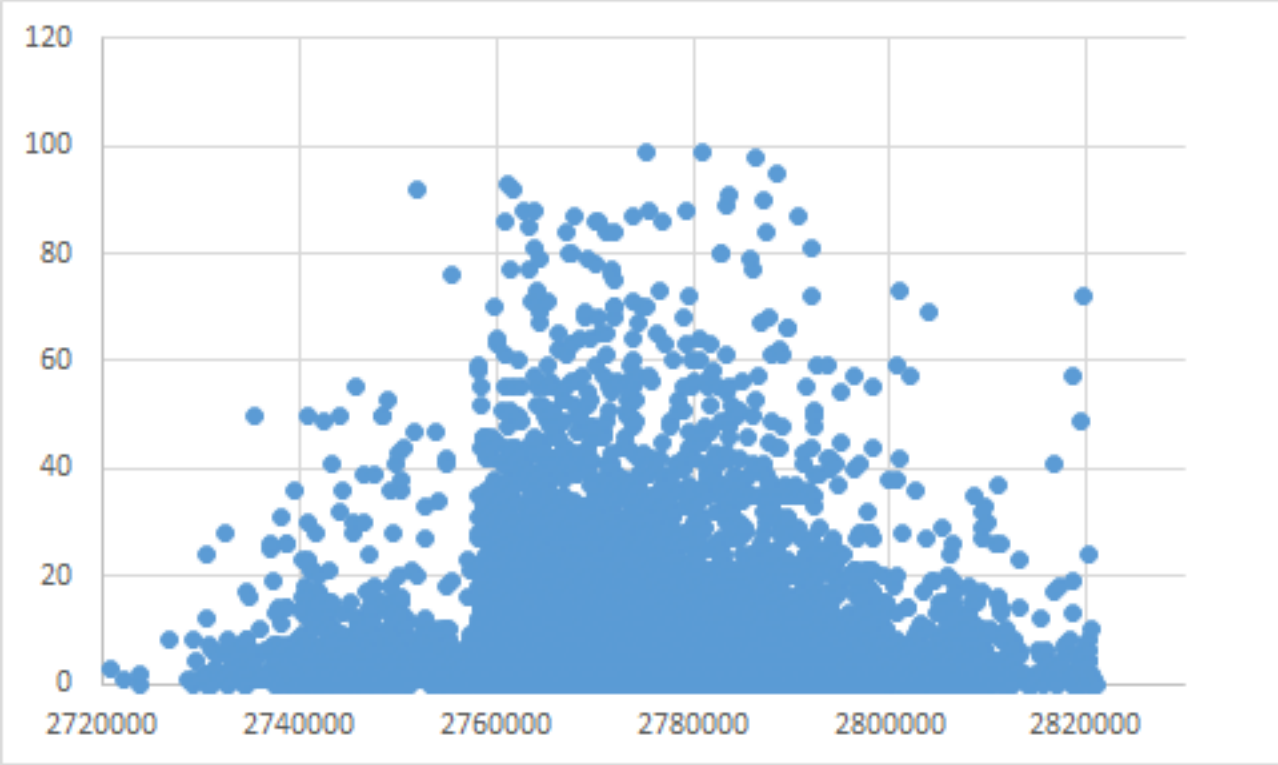
- Describe findings
- Identify solutions

Note: Rectangular chart split up into subrectangles that are sized and ordered in quantitative magnitude.

Suitable for: Visualizing data at a glance and showing hierarchical data as a proportion of a whole

Example: Comparing fiscal budgets between years or departments.

Scatter Plot



Reveal
- Describe findings
- Identify solutions

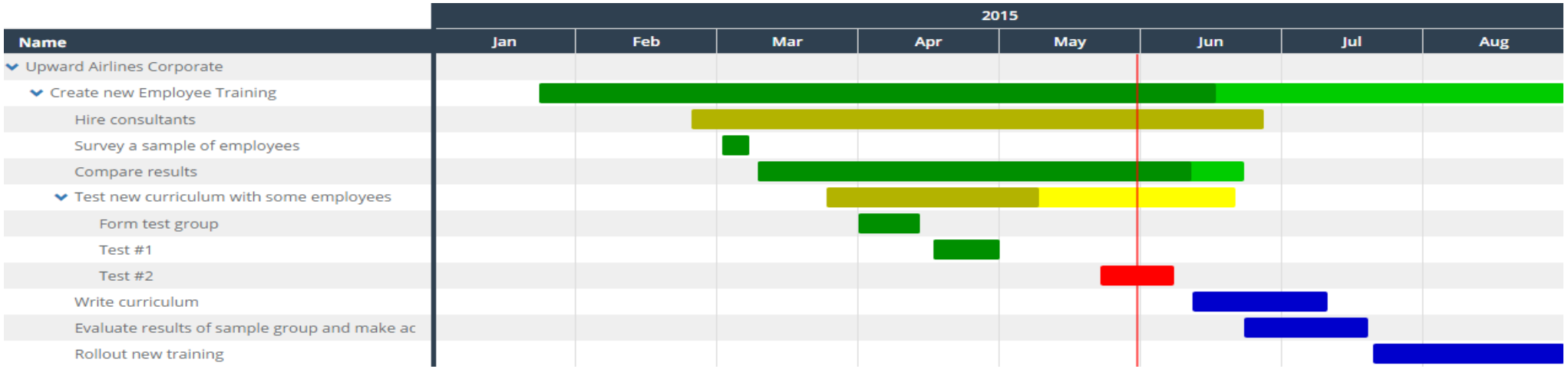
Note: A set of individual dots displayed in a Cartesian plane where each dot denotes an observation for a set of data.

Suitable for: Understanding the relationship between different pieces of information, trends, concentrations, and dispersions.

Example: Relationship between Case ID and Days To Close (no necessary relationship).

Gantt Chart

Reveal
- Describe findings
- Identify solutions



Note: Made up of a series of horizontal lines displaying the amount of work or production completed and uncompleted at different time periods.

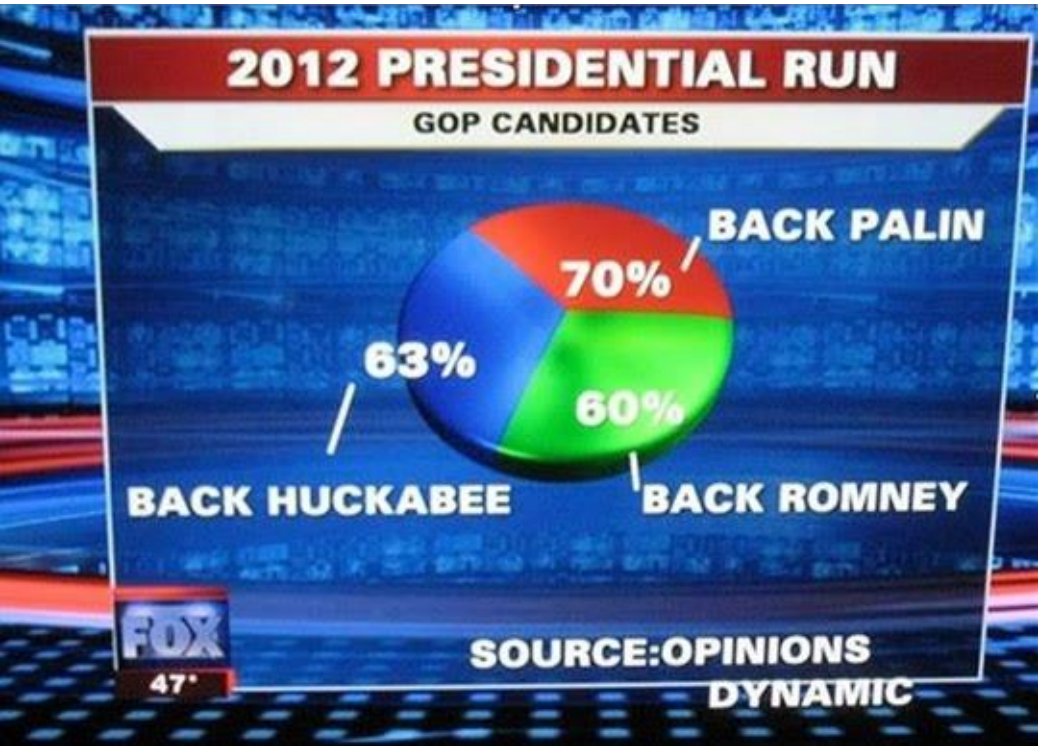
Suitable for: Tracking progress of a project schedule.

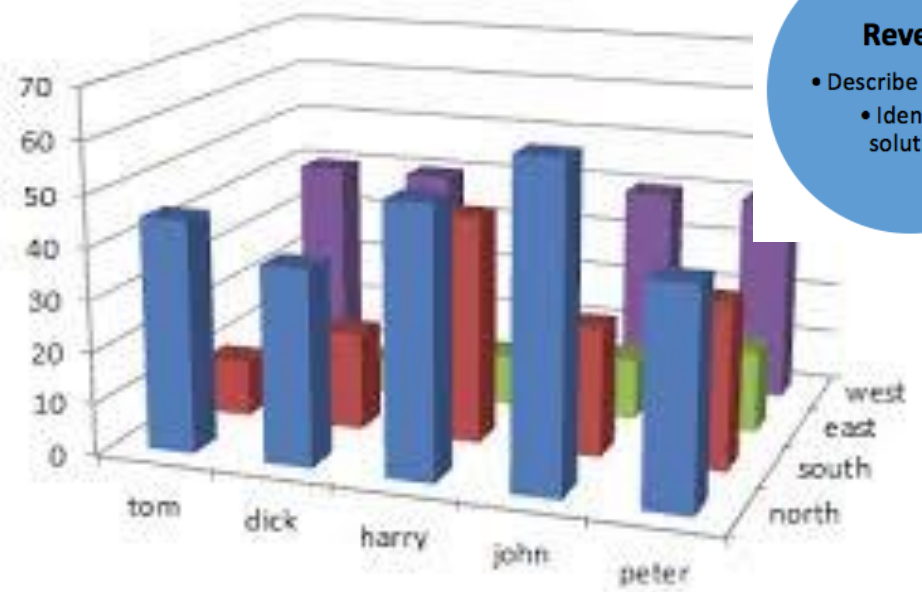
Example: Illustrate city goals, deliverables, owners, and timelines.

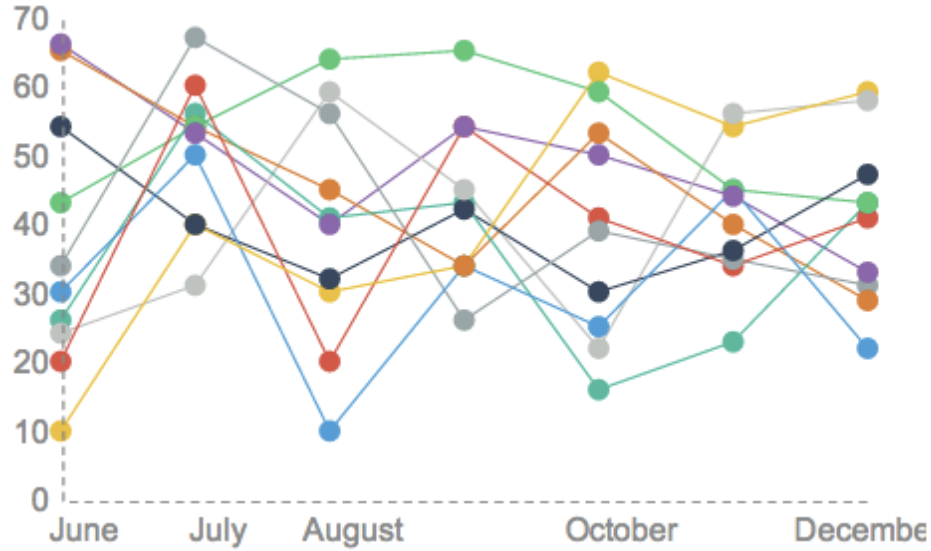
Common Charts for Common Questions

Common Question	Recommended Chart
1. What chart is suitable for analyzing relationships between variables in datasets?	Scatter, Bubble, and Line
2. Which charts are best for comparing values?	Column, Bar, Line, Scatter Plot, Bullet, and Area
3. Which visualization type is right for looking at distribution of data?	Scatter Plot, Line, Bar, and Column
4. What chart is suitable for analyzing trends in datasets?	Line, Dual Axis, and Column
5. Which charts are suitable to show composition?	Pie, Stacked Bar, Stacked Column, Area, and Waterfall

Bad chart examples







Reveal

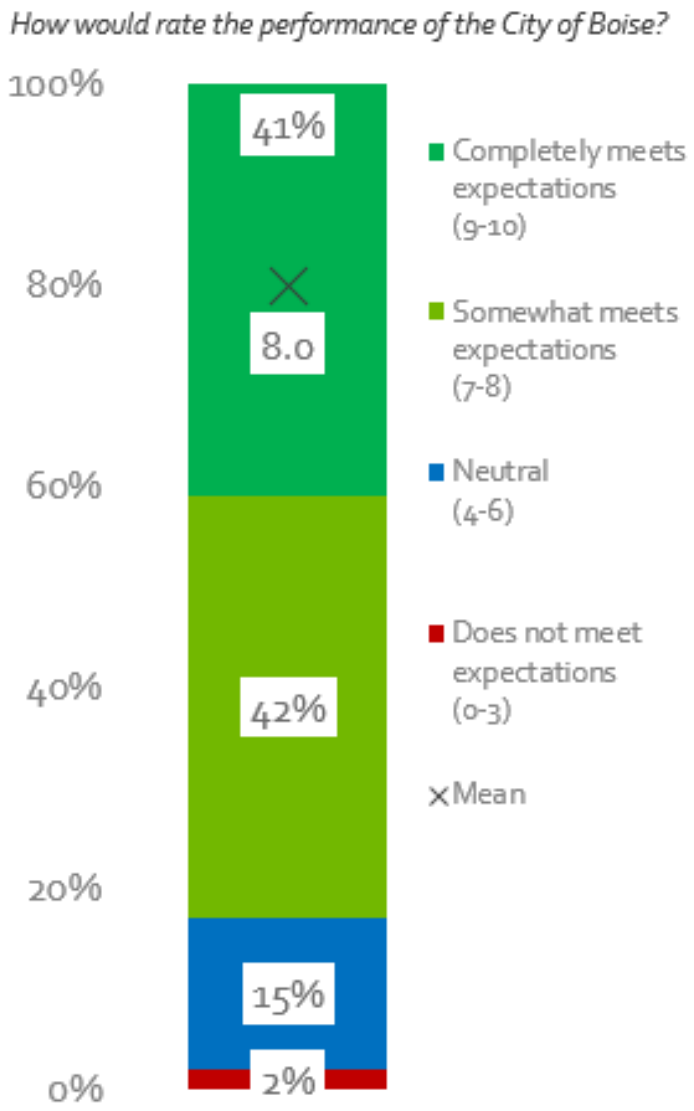
- Describe findings
- Identify solutions

DATA VIZ PRINCIPLES

#3: **COMPARED TO WHAT?**

PRINCIPLE #3:

COMPARED TO WHAT?



83% of residents say the performance of the City of Boise meets expectations.

Is that good?

PRINCIPLE #3:

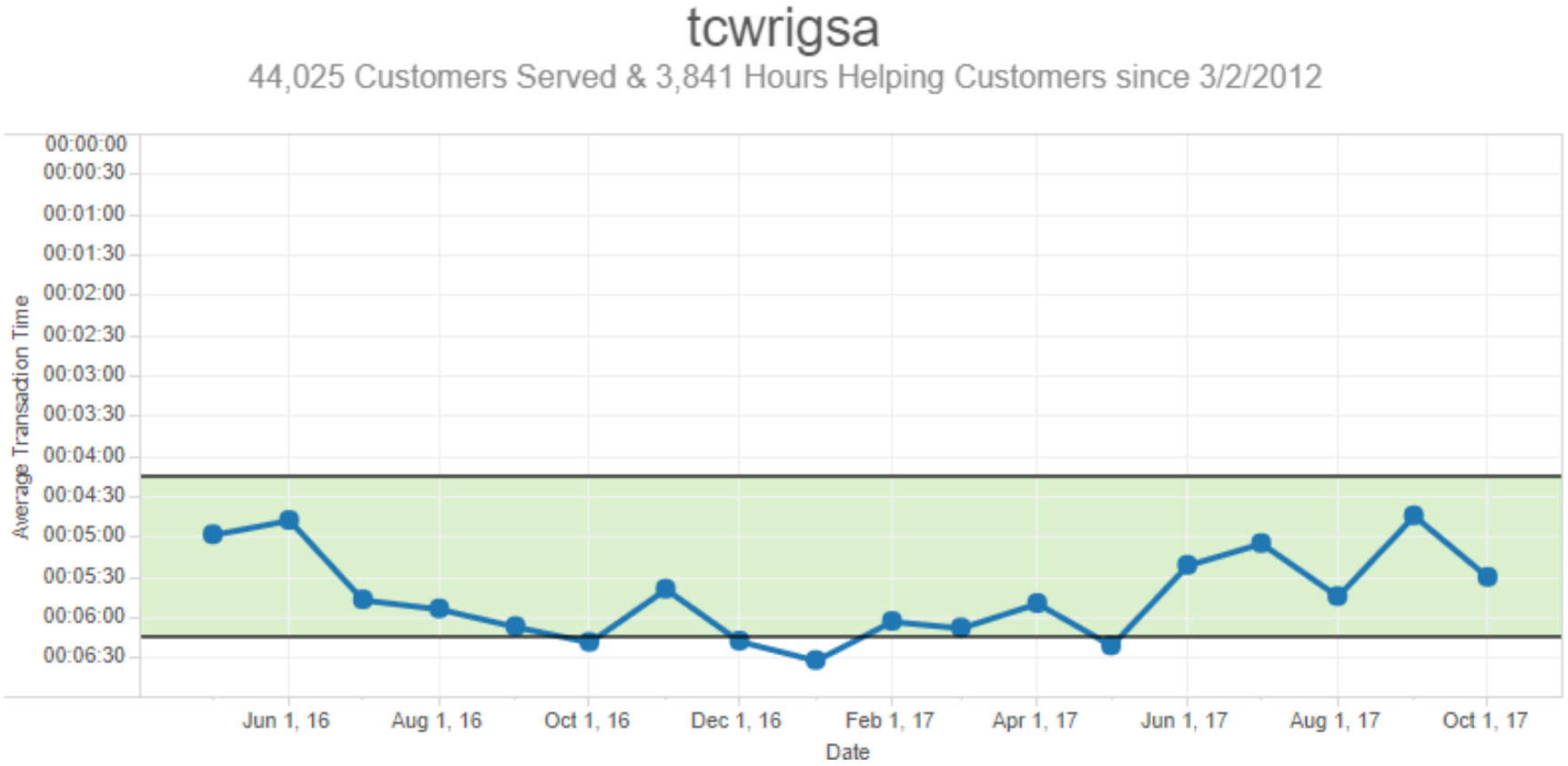
COMPARED TO WHAT?

What can I compare my measure to?

- Target
- Part to whole
- Historical performance
- Peers (people/organizations/places)
- Industry standards
- A second measure

PRINCIPLE #3:

COMPARED TO WHAT?



What comparison is depicted?

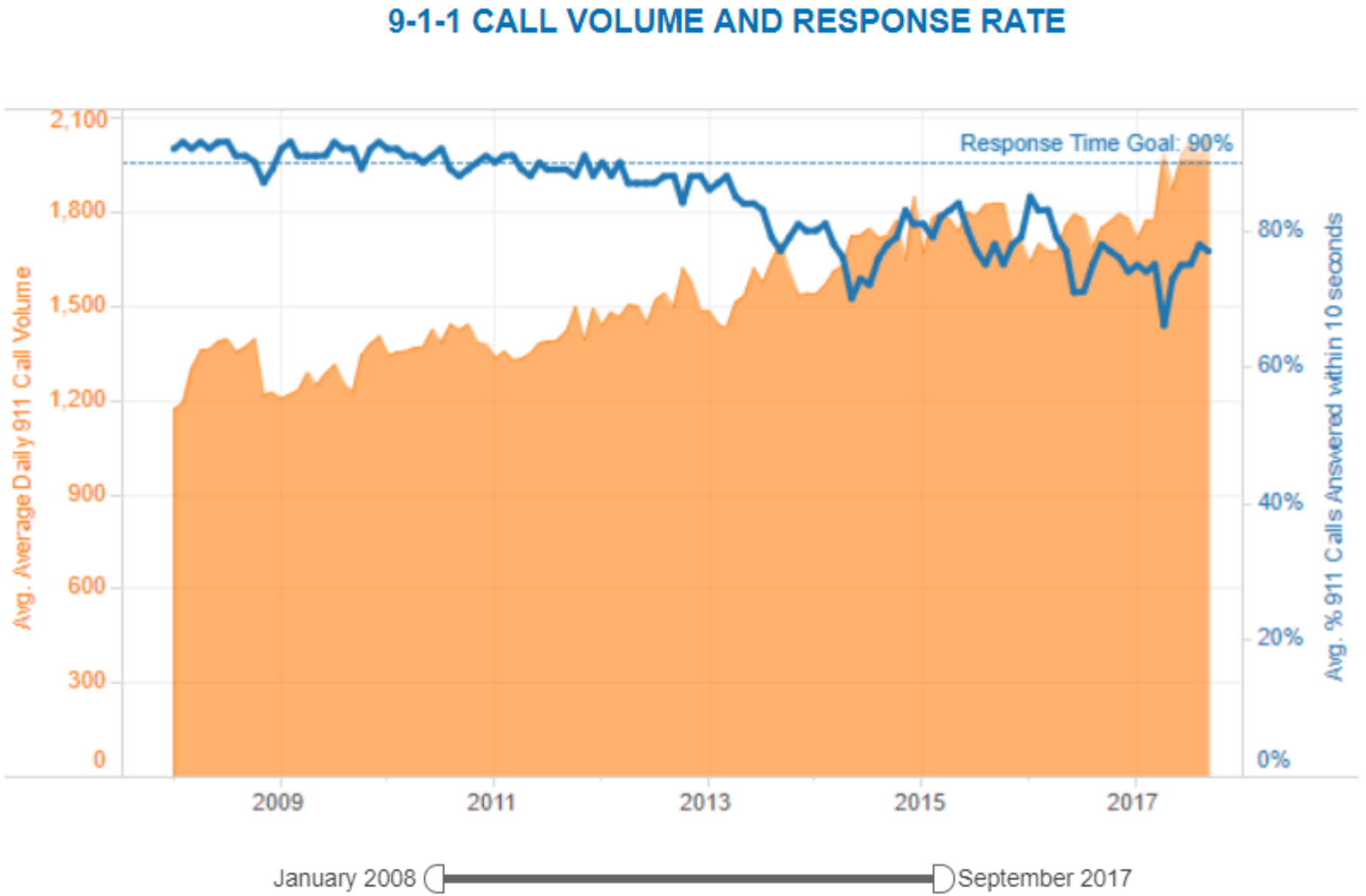
- **Target**
- Part to whole
- **Historical performance**
- Peers
- Industry standards
- A second measure

PRINCIPLE #3:

COMPARED TO WHAT?

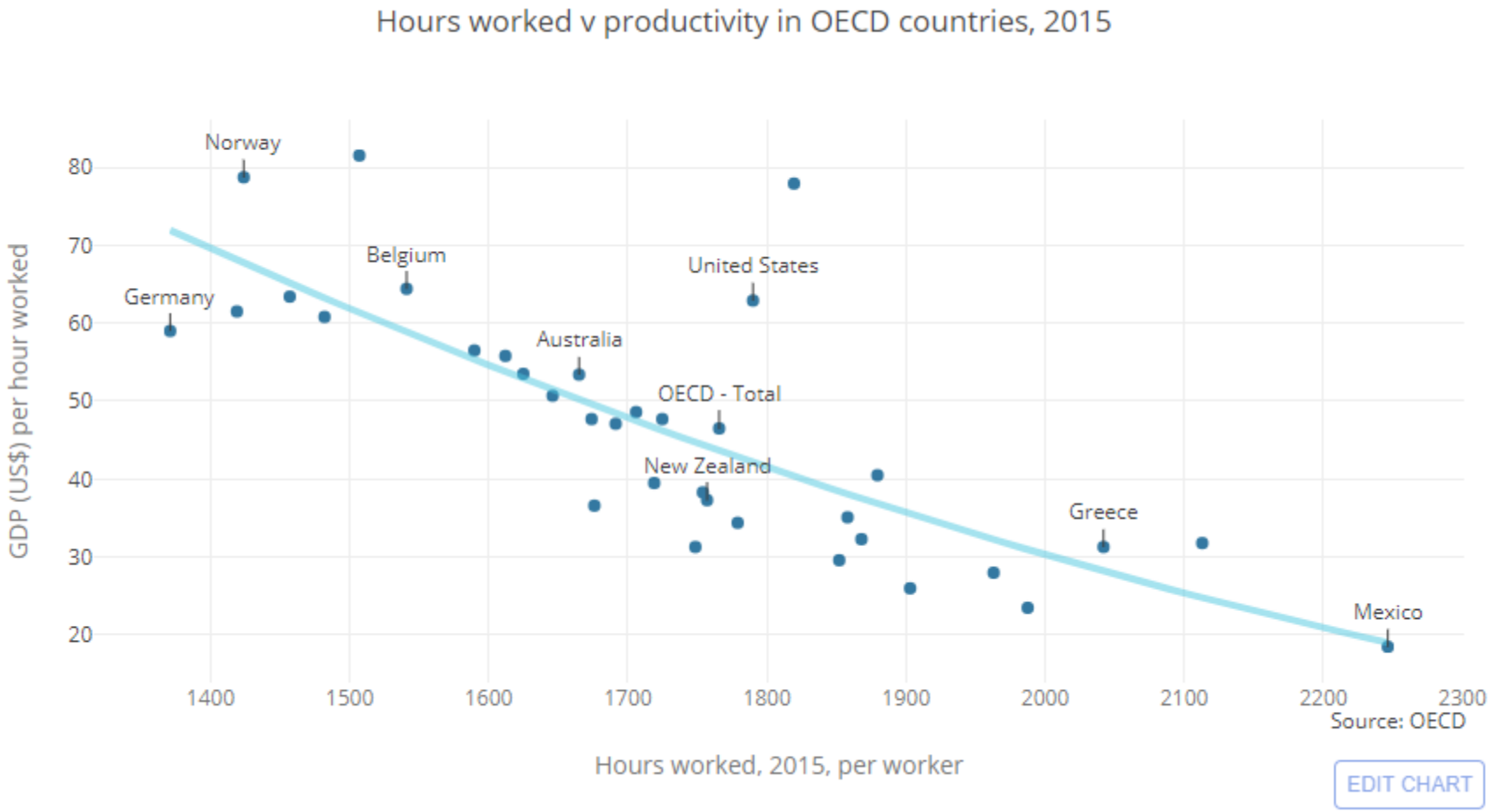
What comparison is depicted?

- **Target**
- Part to whole
- **Historical performance**
- Peers
- Industry standards
- **A second measure**



PRINCIPLE #3:

COMPARED TO WHAT?



What comparison is depicted?

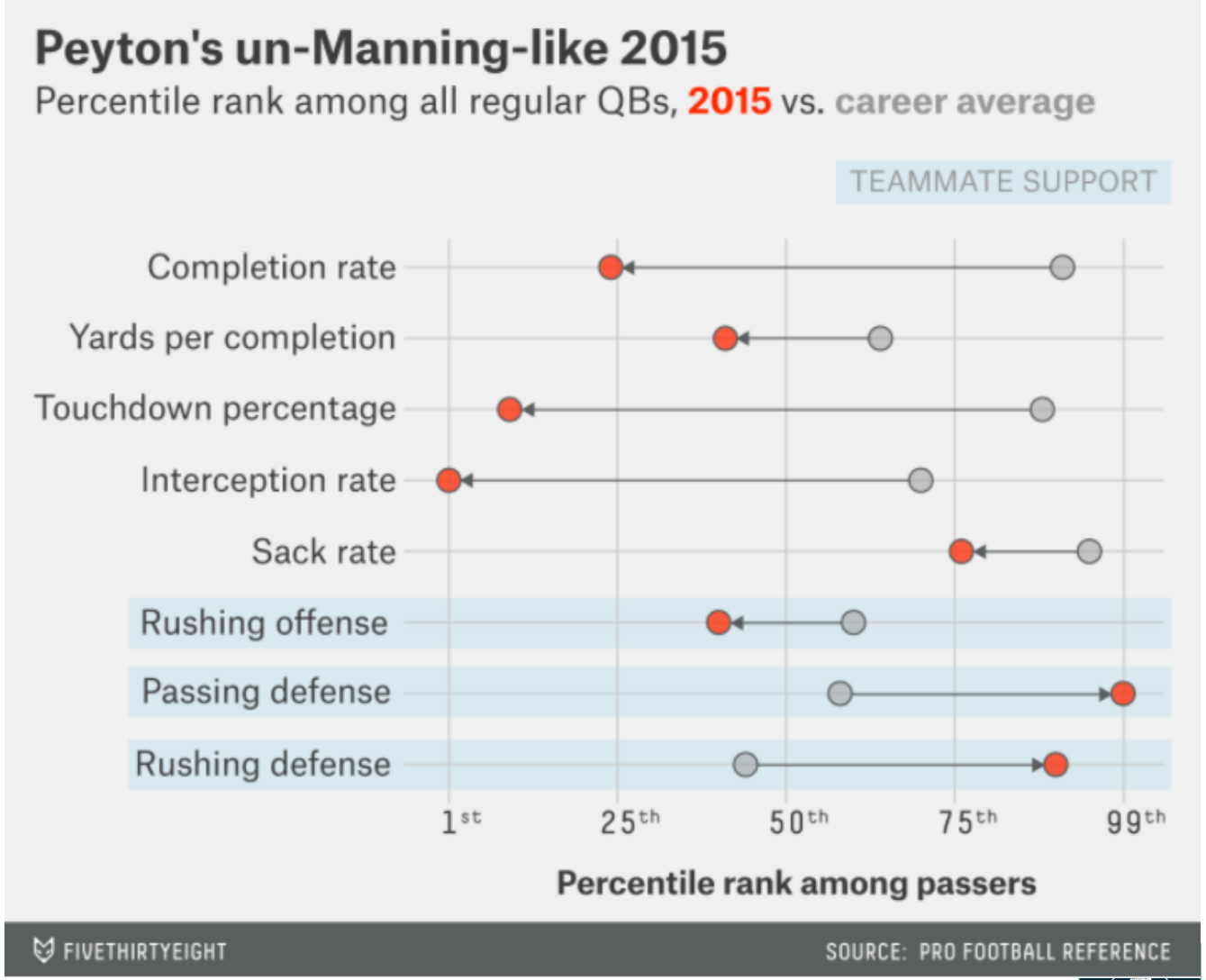
- Target
- Part to whole
- Historical performance
- **Peers**
- Industry standards
- A second measure

PRINCIPLE #3:

COMPARED TO WHAT?

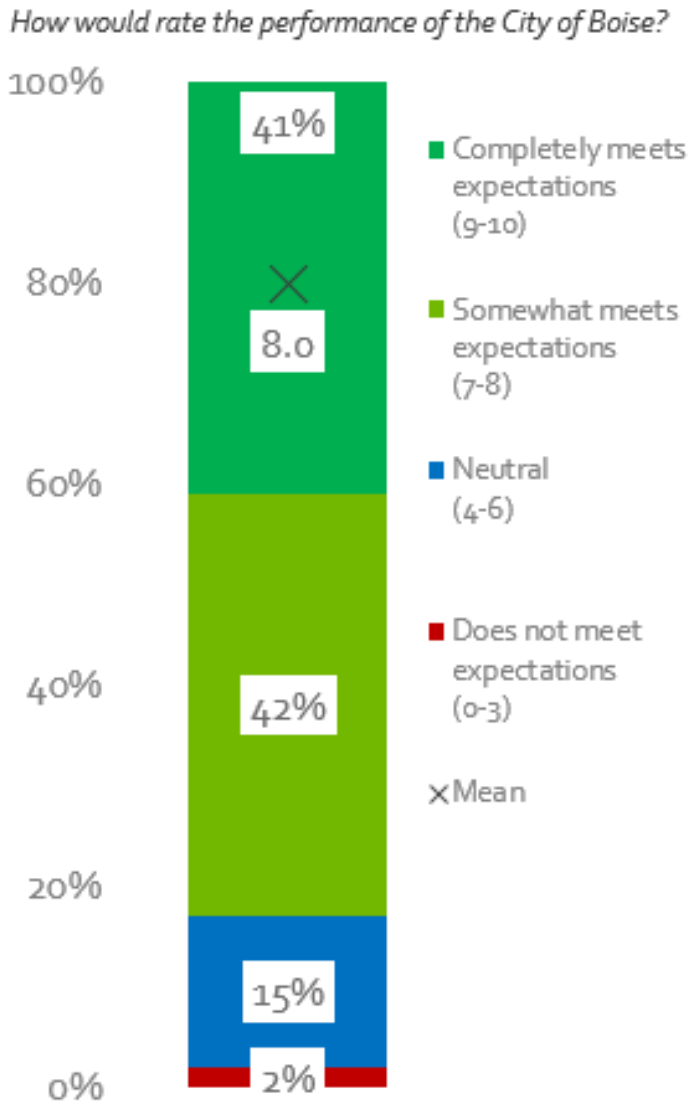
What comparison is depicted?

- Target
- **Part to whole**
- **Historical performance**
- Peers
- Industry standards
- A second measure



PRINCIPLE #3:

COMPARED TO WHAT?

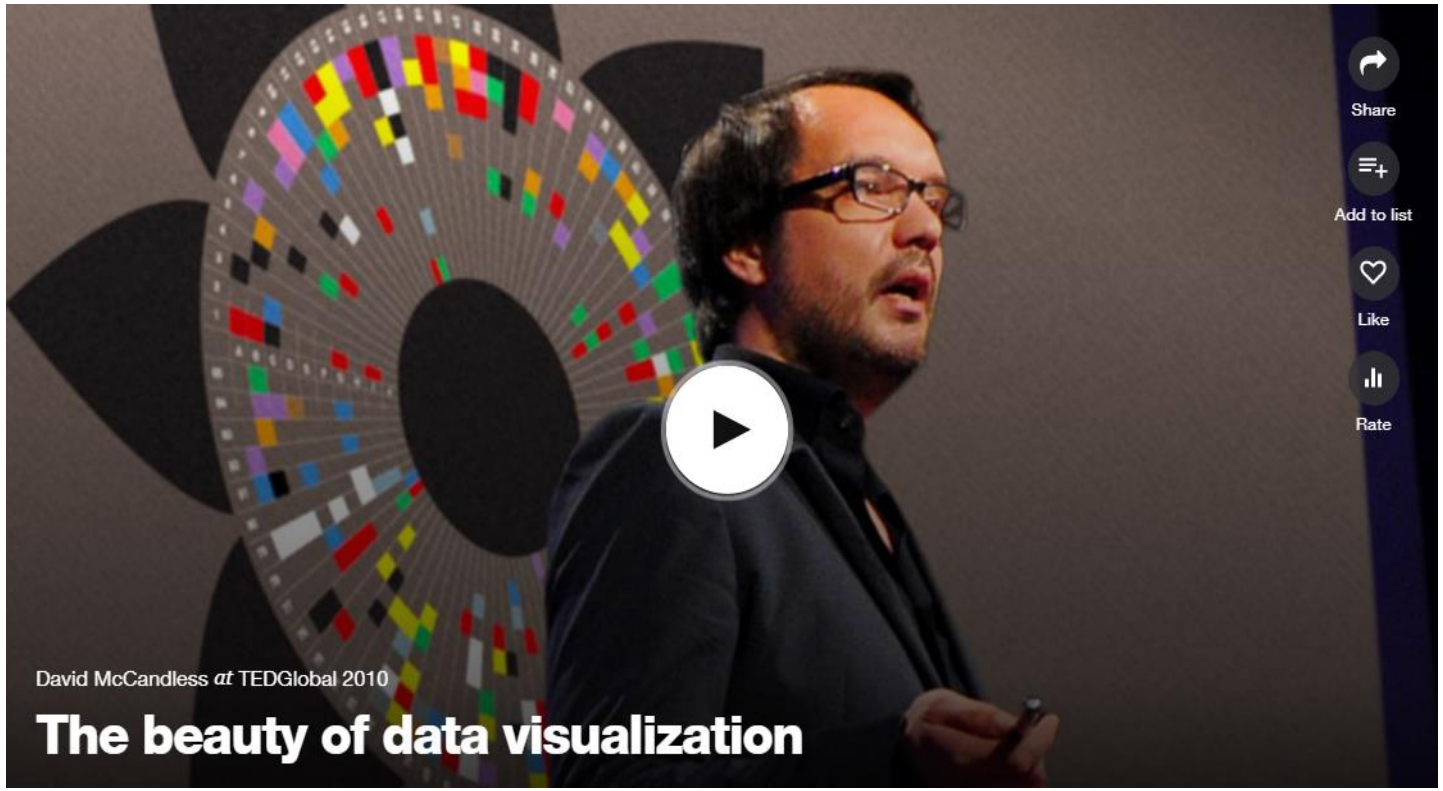


What comparison could/should we use?

- Target
- Part to whole
- Historical performance
- Peers
- Industry standards
- A second measure

PRINCIPLE #3:

COMPARED TO WHAT?



Kyle's takeaways:

1. Use rates whenever you can!
   - Per person
   - Per $
   - Per mile travelled
2. Make sure comparisons are apples to apples.

DATA VIZ PRINCIPLES

#4: **DON'T MISLEAD**

PRINCIPLE #4:

DON'T MISLEAD

The relative size of objects should reflect the data.

PRINCIPLE #4:

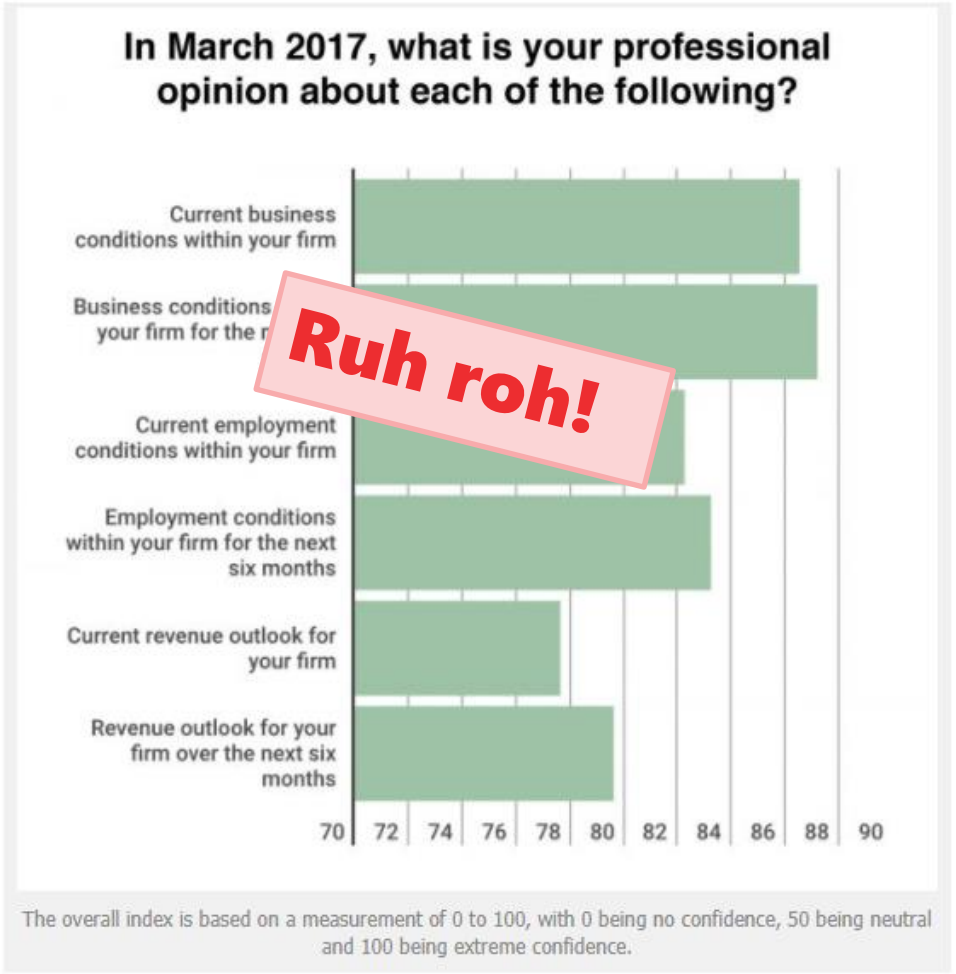
DON'T MISLEAD

Don't truncate axes…usually.

IDAHOBUSINESSREVIEW.com

Idaho Outlook Indicator – Q1 2017

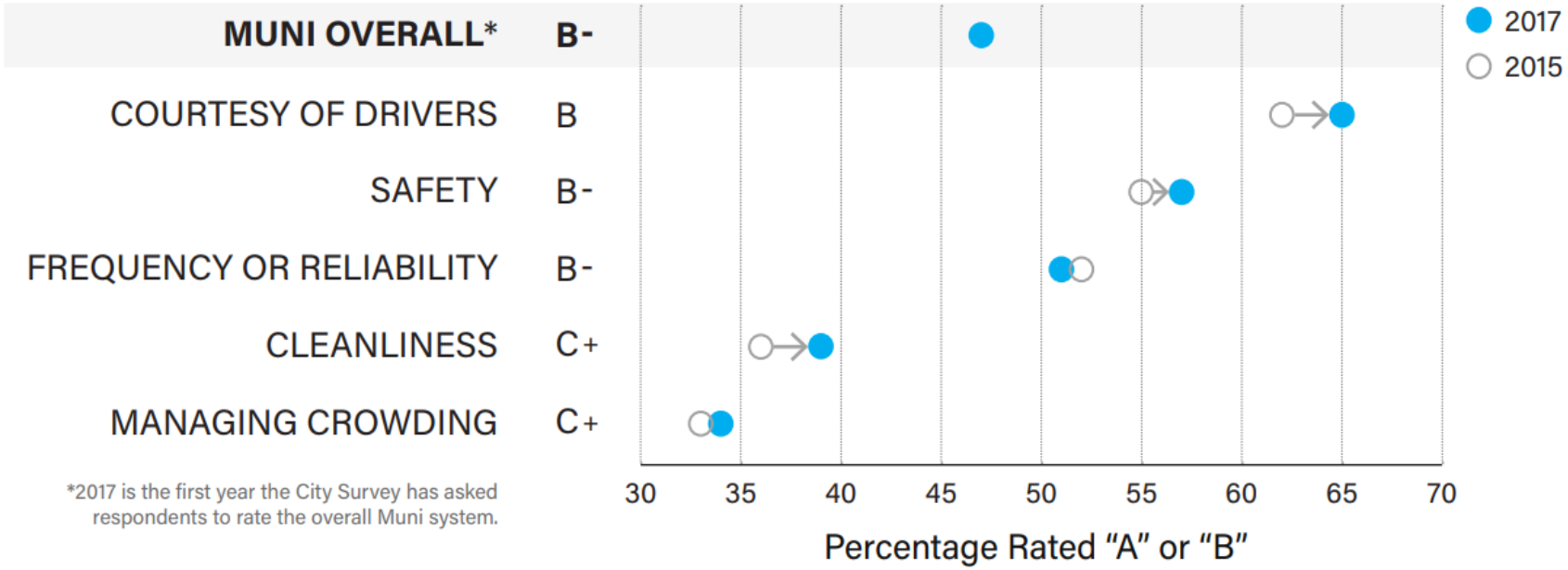
By: Erika Sather-Smith May 5, 2017 Comments Offon Idaho Outlook Indicator – Q1 2017



PRINCIPLE #4:

DON'T MISLEAD

SF City Survey – Transportation Ratings



*2017 is the first year the City Survey has asked respondents to rate the overall Muni system.

DATA VIZ PRINCIPLES

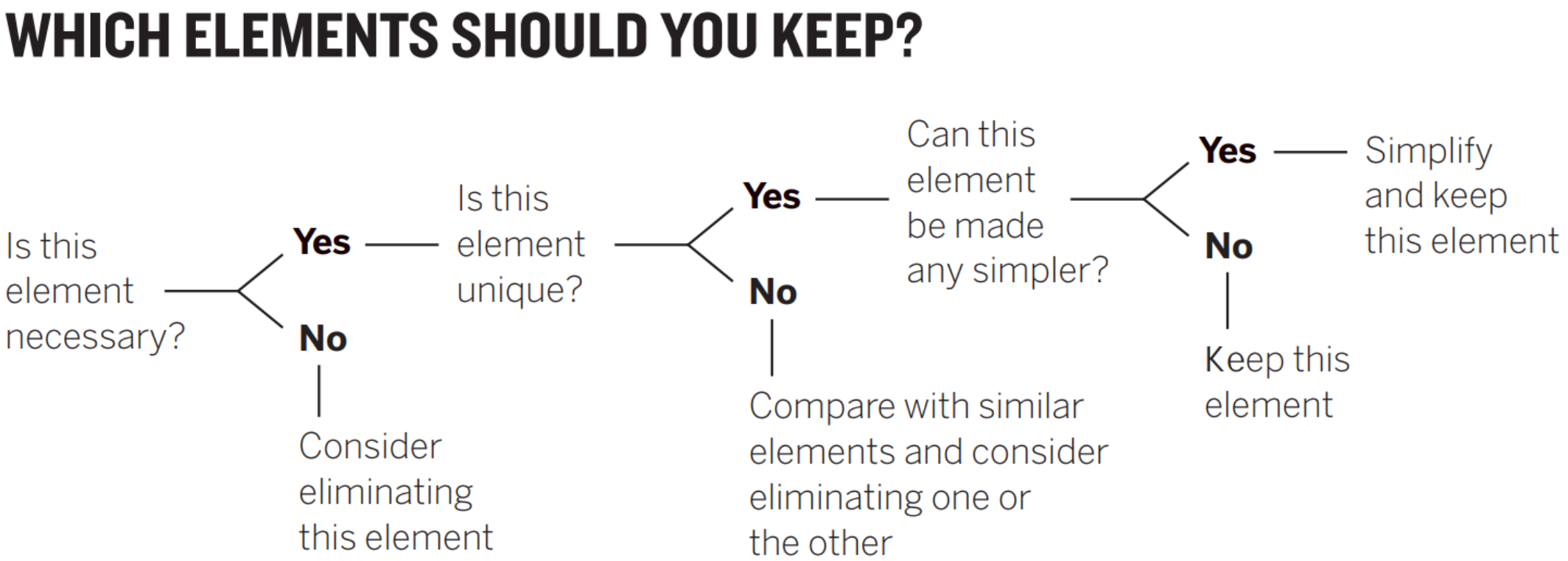
#5: **EFFICIENCY!**

PRINCIPLE #5:

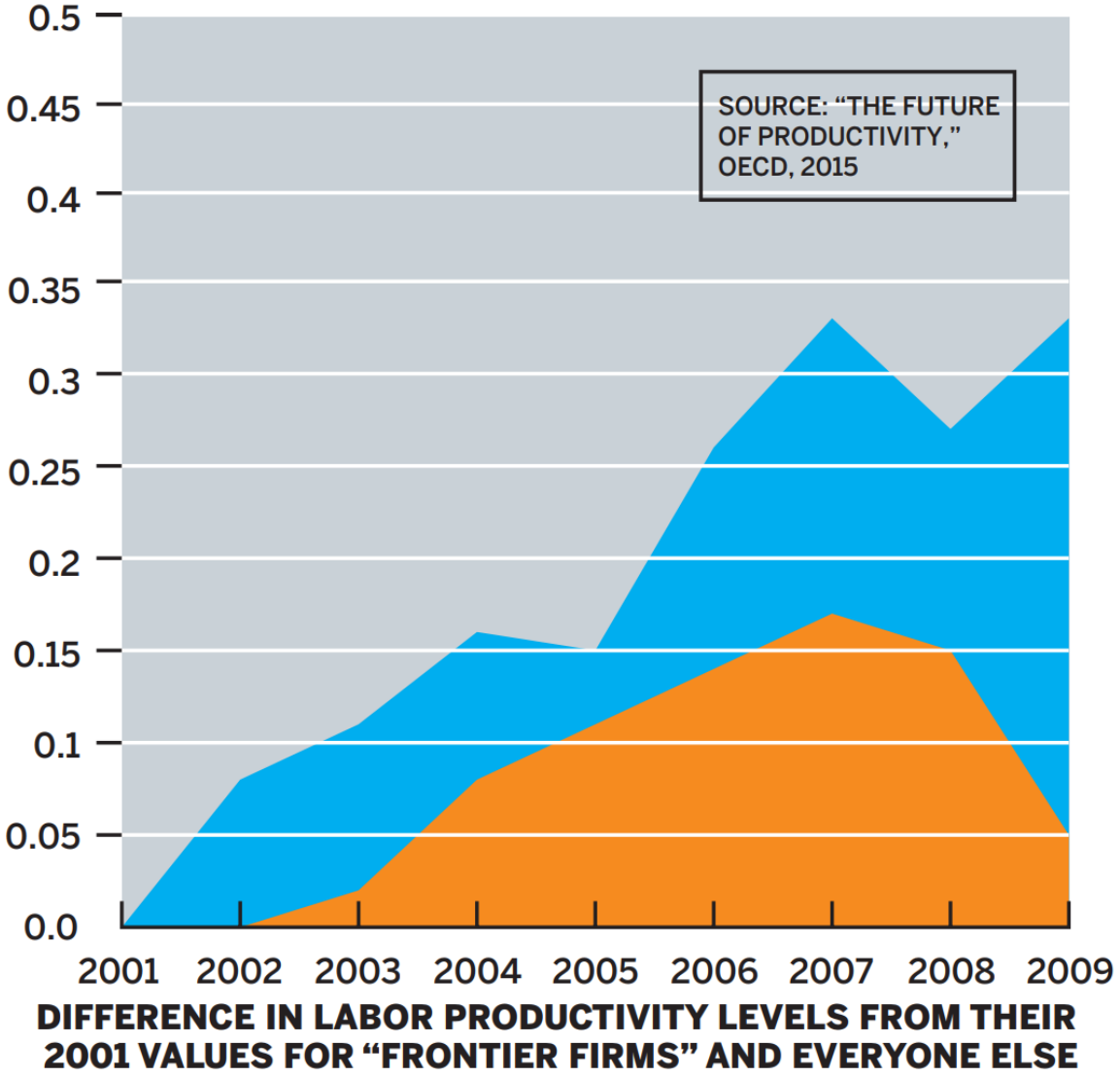
EFFICIENCY!

$$\text{Data Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink}}$$

EFFICIENCY!



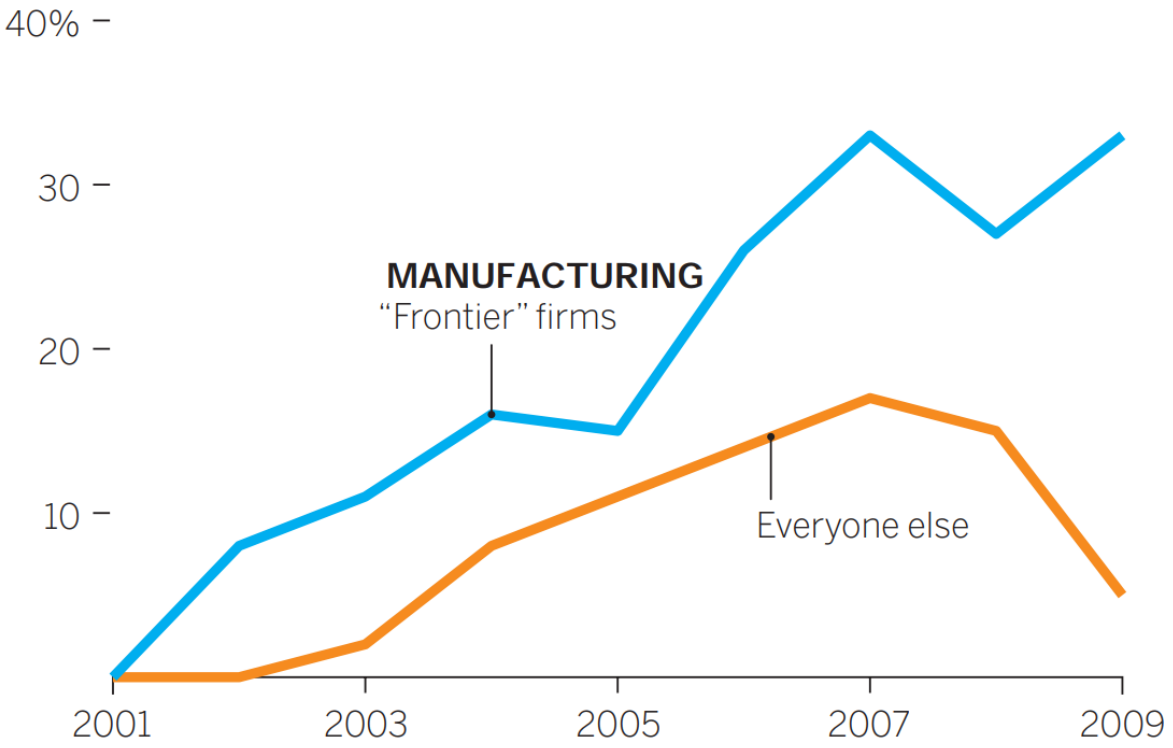
From: *Good Charts* by Scott Berinato



DIFFERENCE IN LABOR PRODUCTIVITY LEVELS FROM THEIR 2001 VALUES FOR "FRONTIER FIRMS" AND EVERYONE ELSE

THE GAP BETWEEN THE MOST PRODUCTIVE FIRMS AND THE REST IS GROWING

PERCENTAGE DIFFERENCE IN LABOR PRODUCTIVITY LEVELS FROM THEIR 2001 VALUES (INDEX, 2001=0)



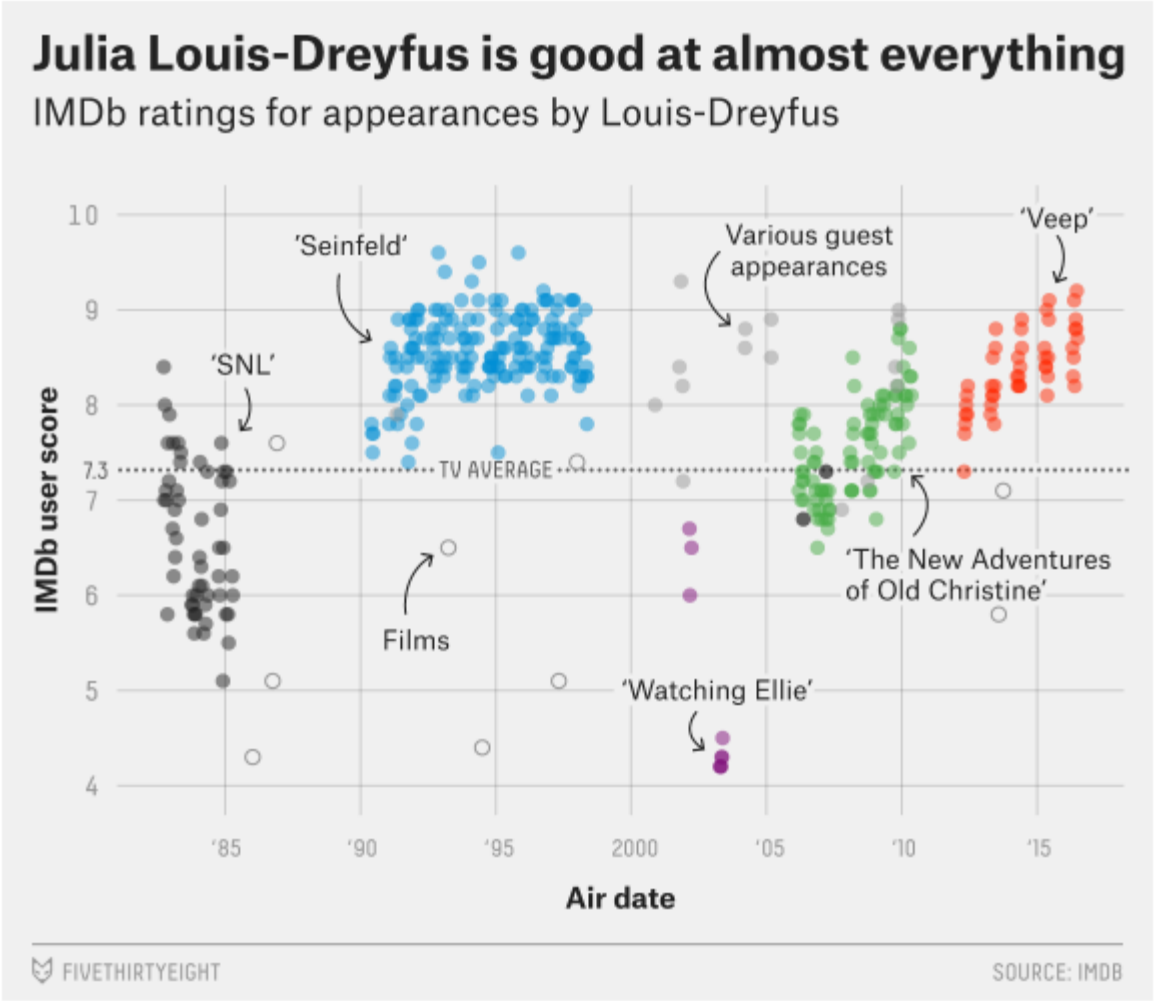
SOURCE: "THE FUTURE OF PRODUCTIVITY," OECD, 2015

DATA VIZ PRINCIPLES

#6: **USE CHART ELEMENTS FOR IMPACT**

PRINCIPLE #6:

USE CHART ELEMENTS FOR IMPACT



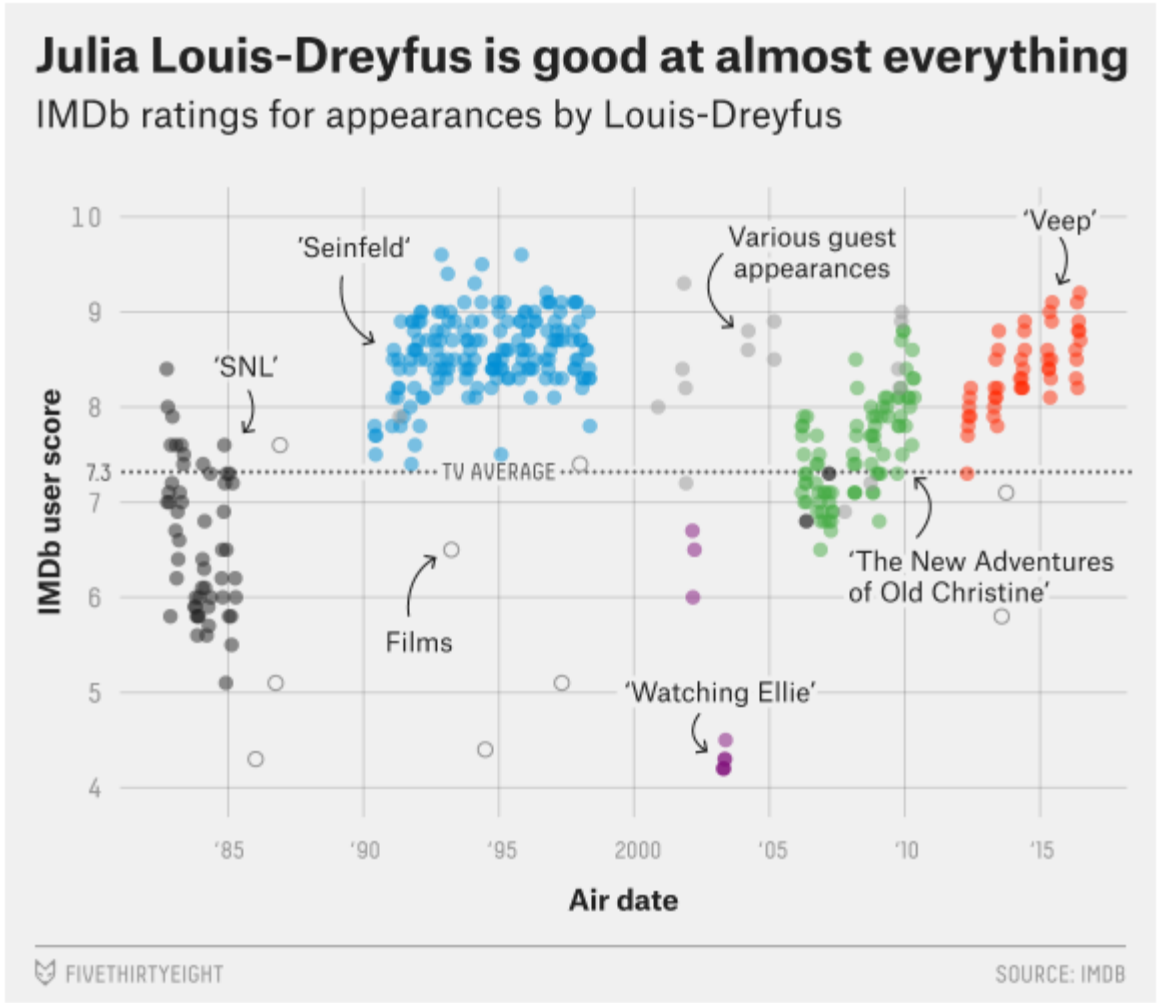
We see what jumps out at us then try to make meaning out of it.

(steep slopes, clusters, outliers, bright colors)

Draw audience attention to your key point.

PRINCIPLE #6:

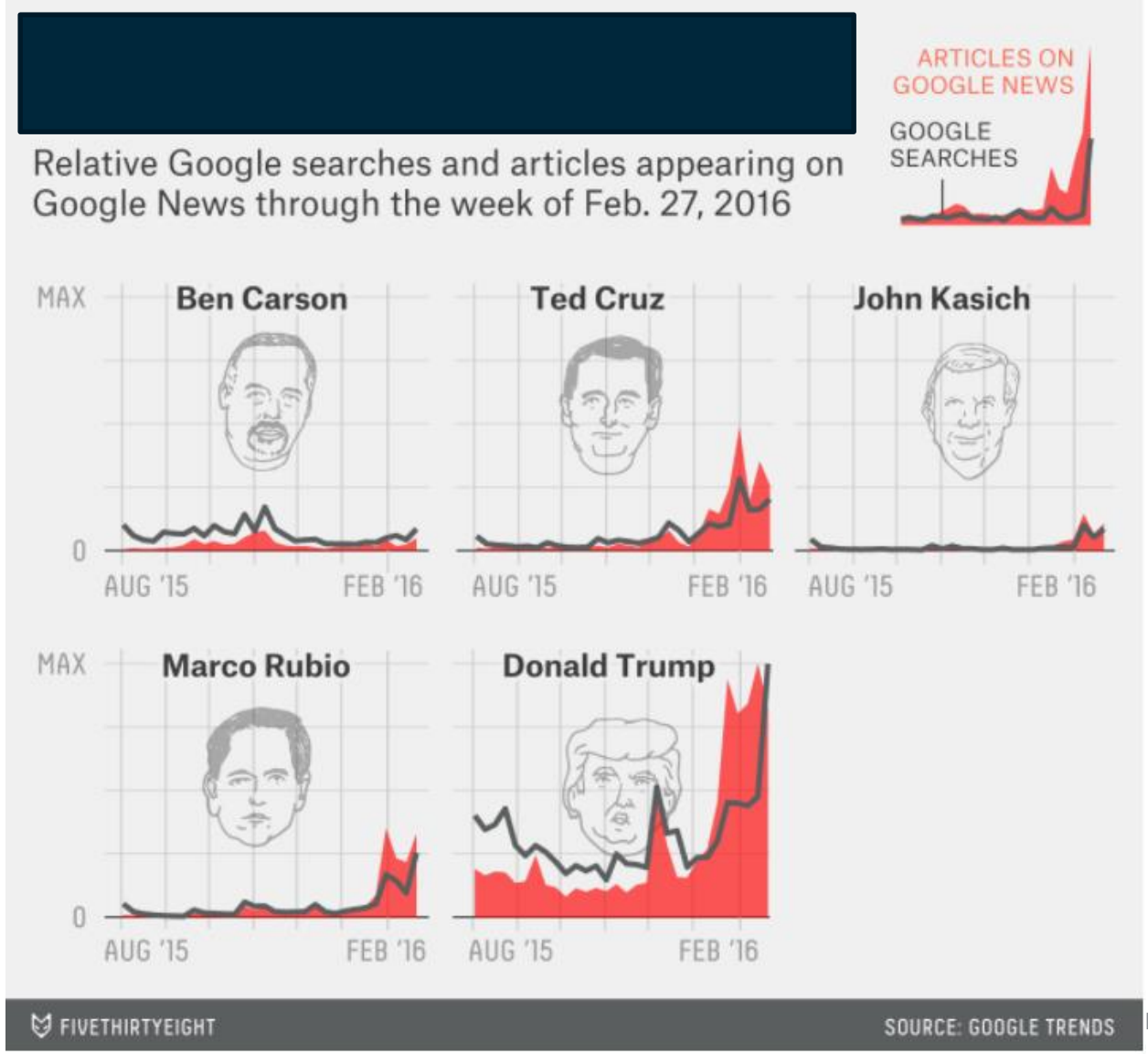
USE CHART ELEMENTS FOR IMPACT



A declarative title tells your audience the key point of the visual.

PRINCIPLE #6:

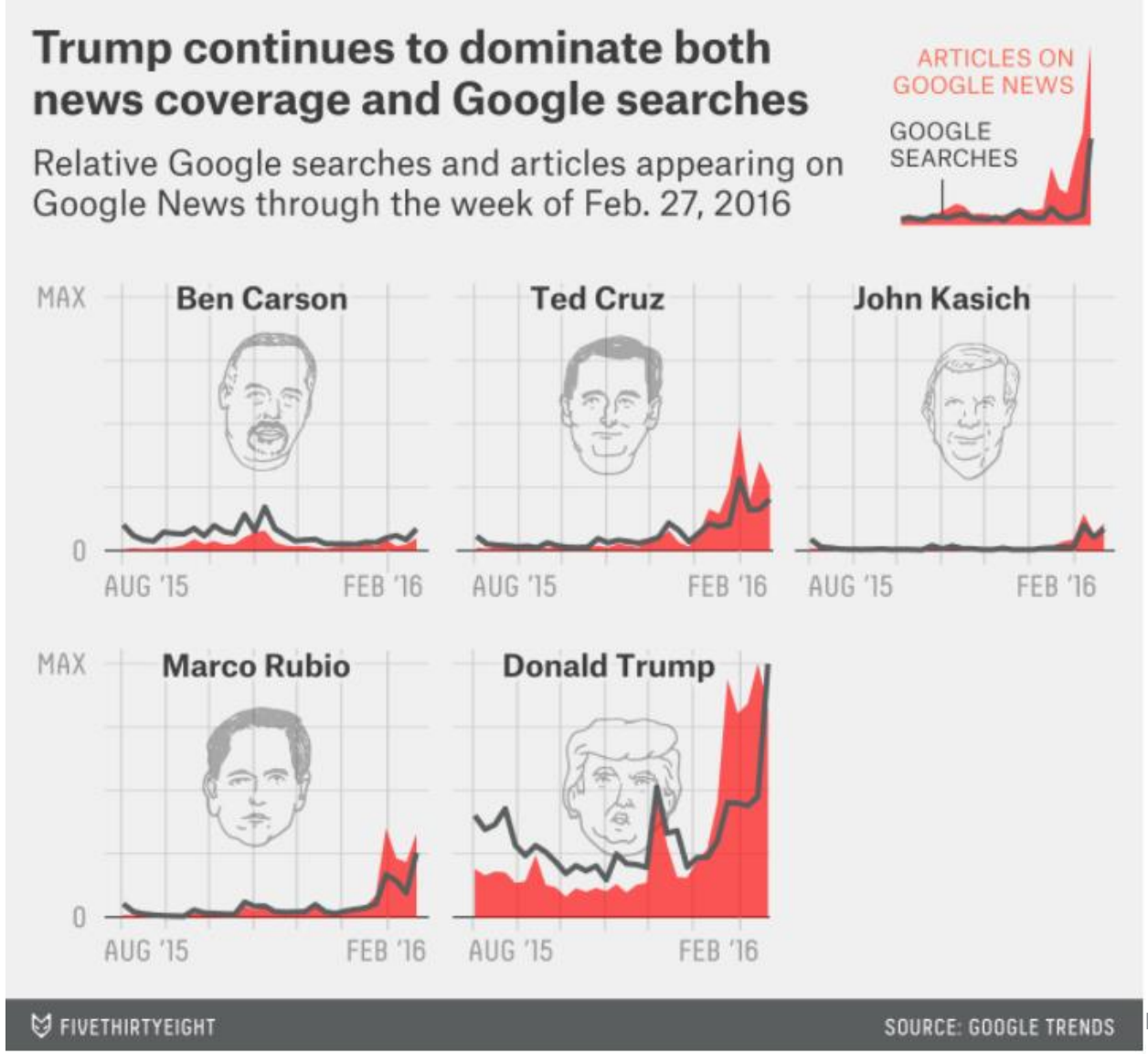
USE ELEMENTS FOR IMPACT



Come up with a declarative title!

PRINCIPLE #6:

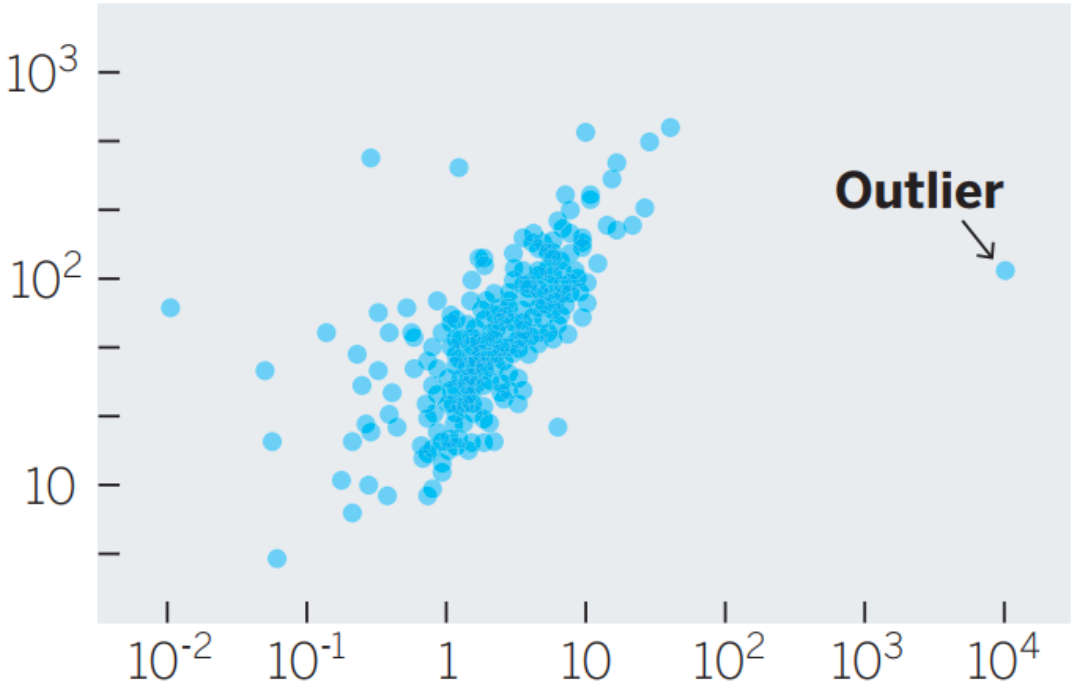
USE ELEMENTS FOR IMPACT

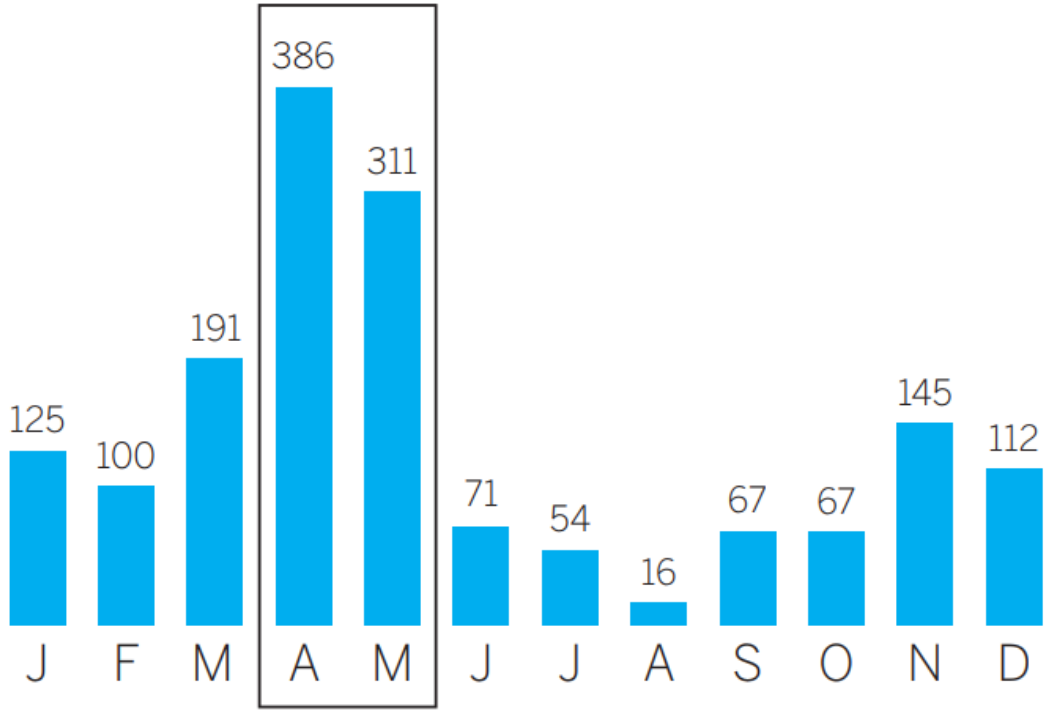


Surprise!

PRINCIPLE #6:

USE ELEMENTS FOR IMPACT



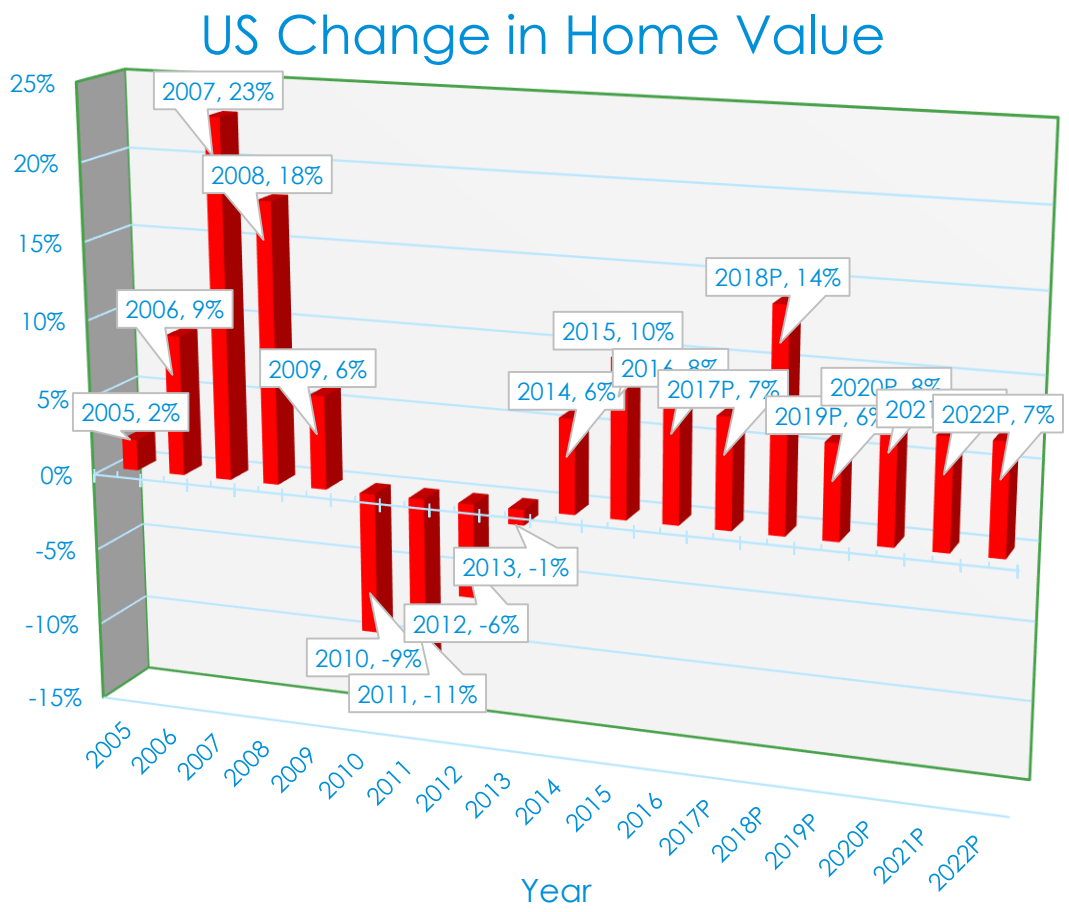


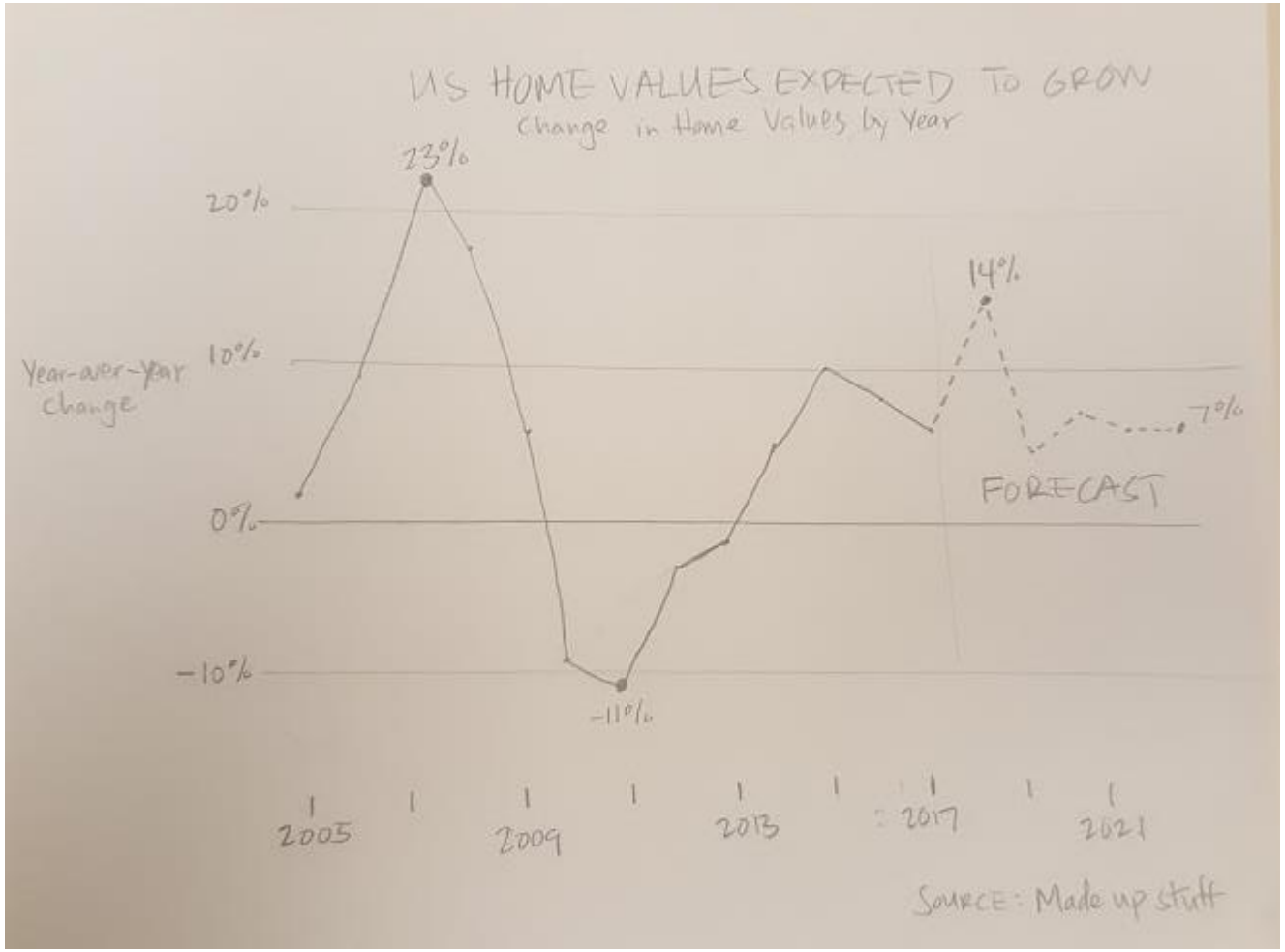
Activity time!

Chart Fixer Upper

PRINCIPLE #5:

EFFICIENCY!





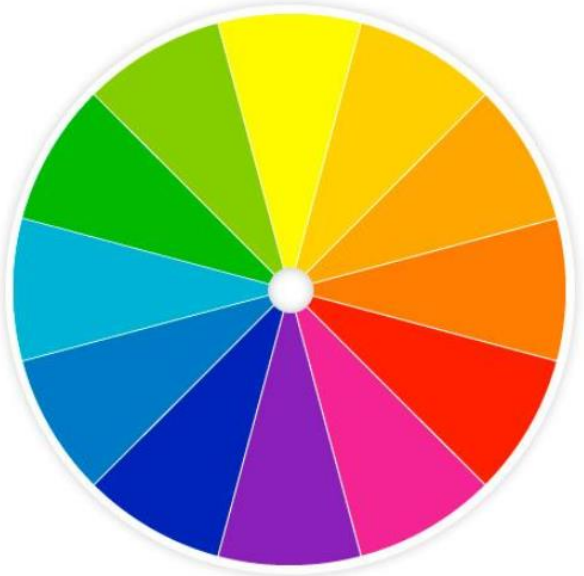
PRINCIPLE #6:

USE ELEMENTS FOR IMPACT

Use color to:

Draw attention to your key point(s).

Make it easy for your audience.

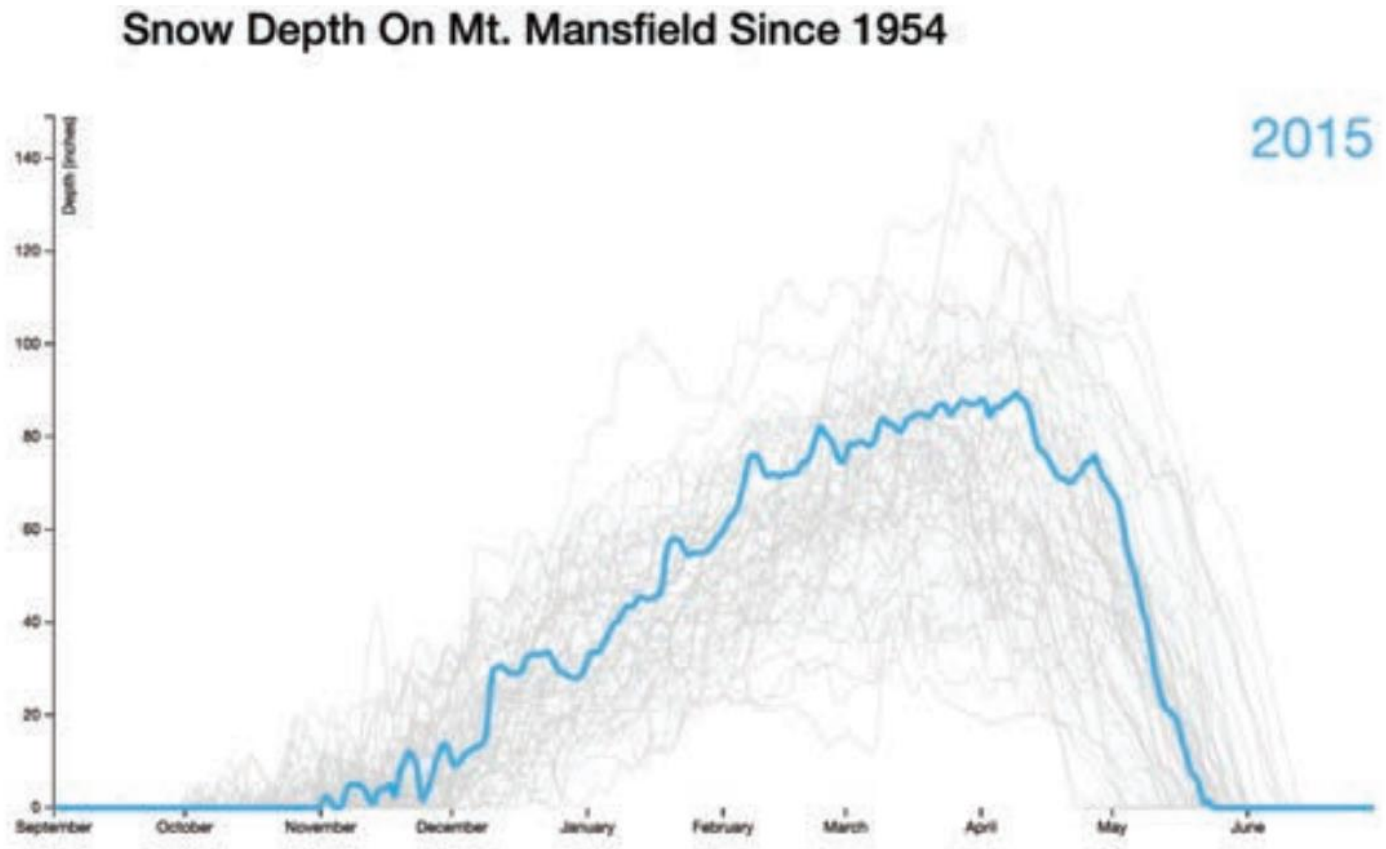


PRINCIPLE #6:

USE ELEMENTS FOR IMPACT

Isolation Effect:

We remember what is most different.

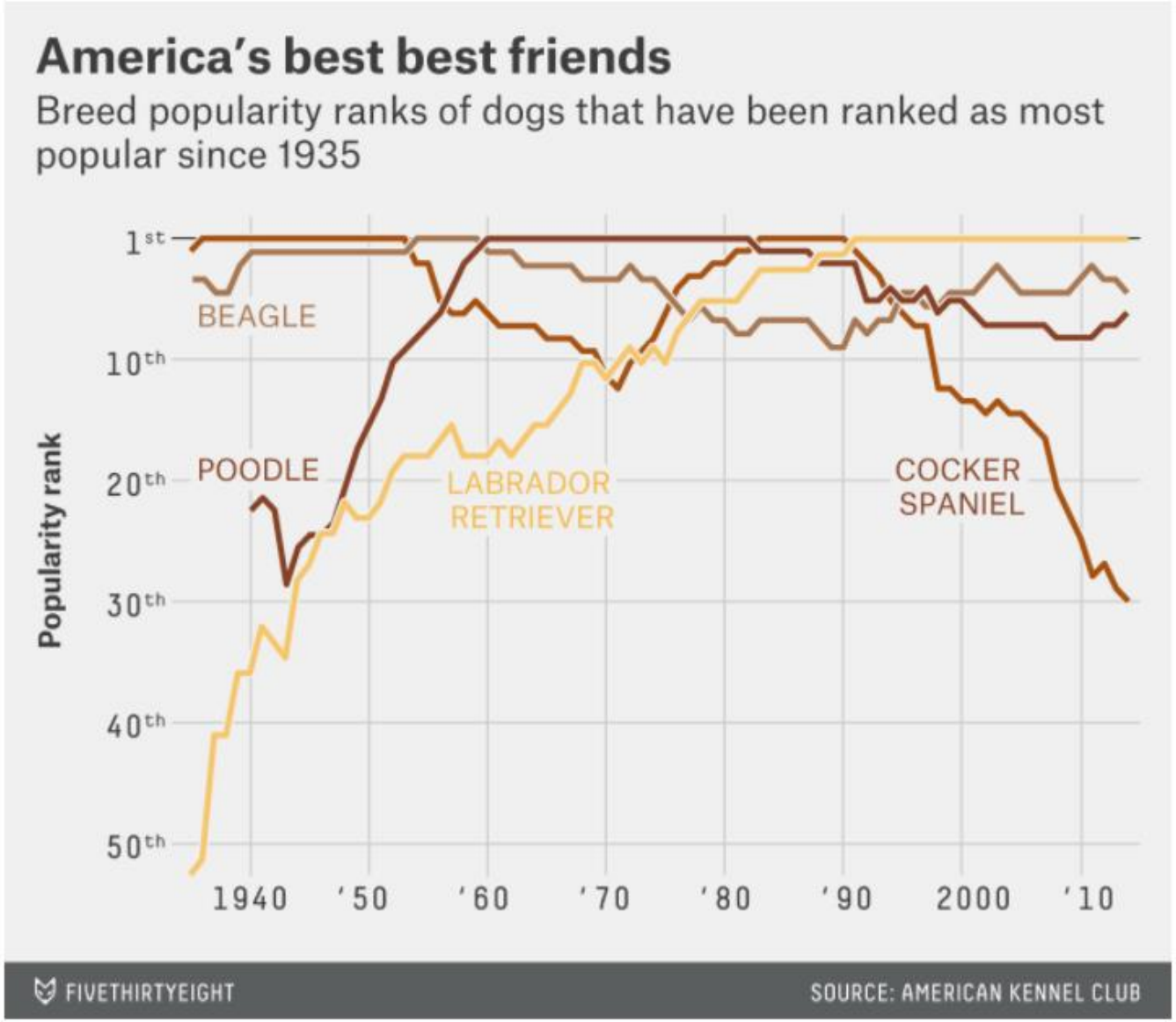


PRINCIPLE #6:

USE ELEMENTS FOR IMPACT

Isolation Effect:

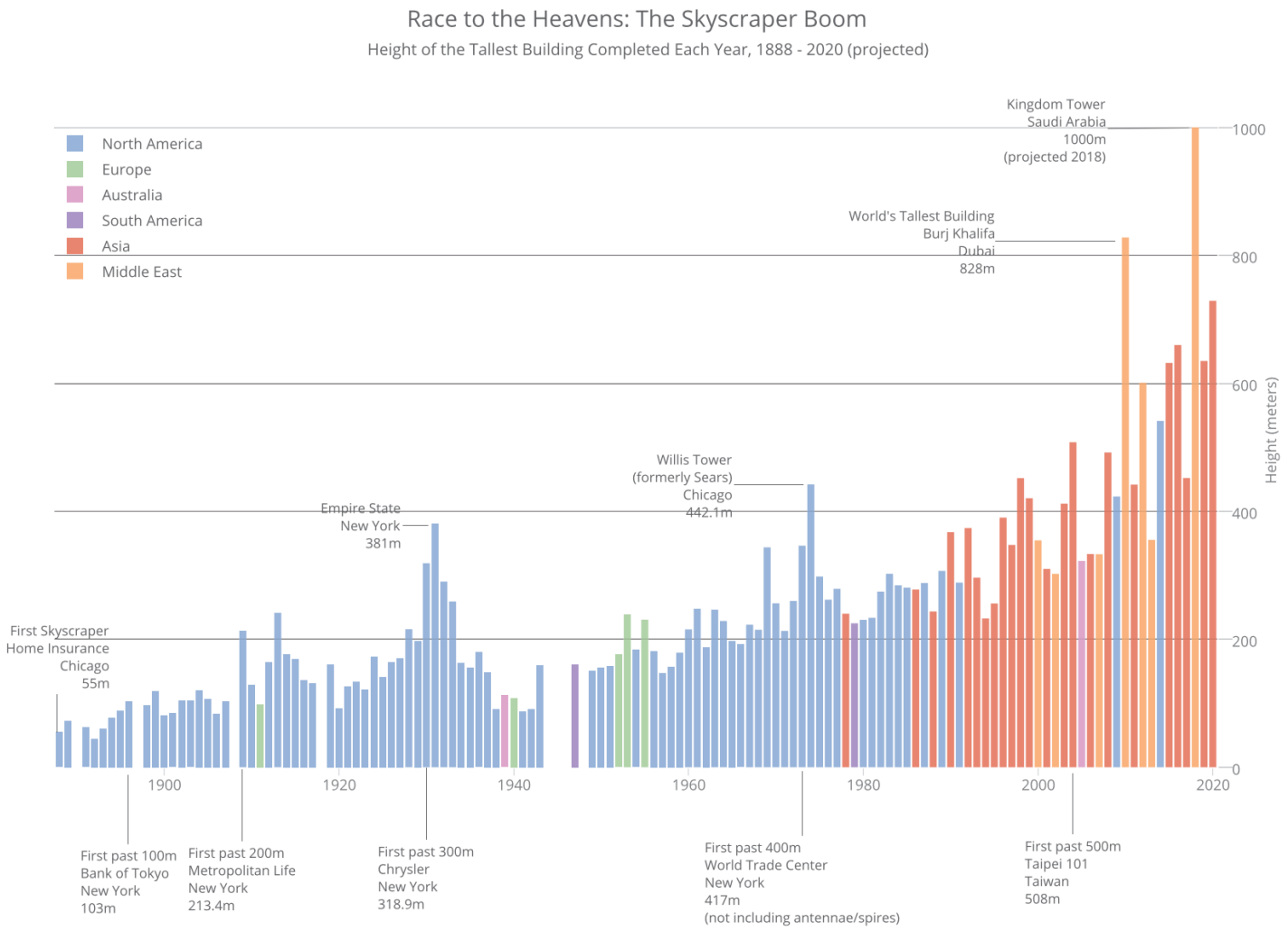
We remember what is most different.



PRINCIPLE #6:

USE ELEMENTS FOR IMPACT

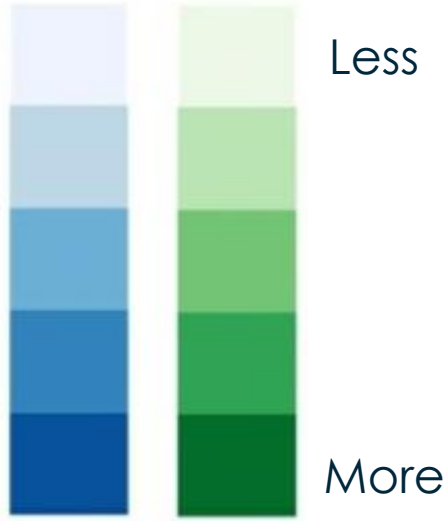
Use colors to group like items.



PRINCIPLE #6:

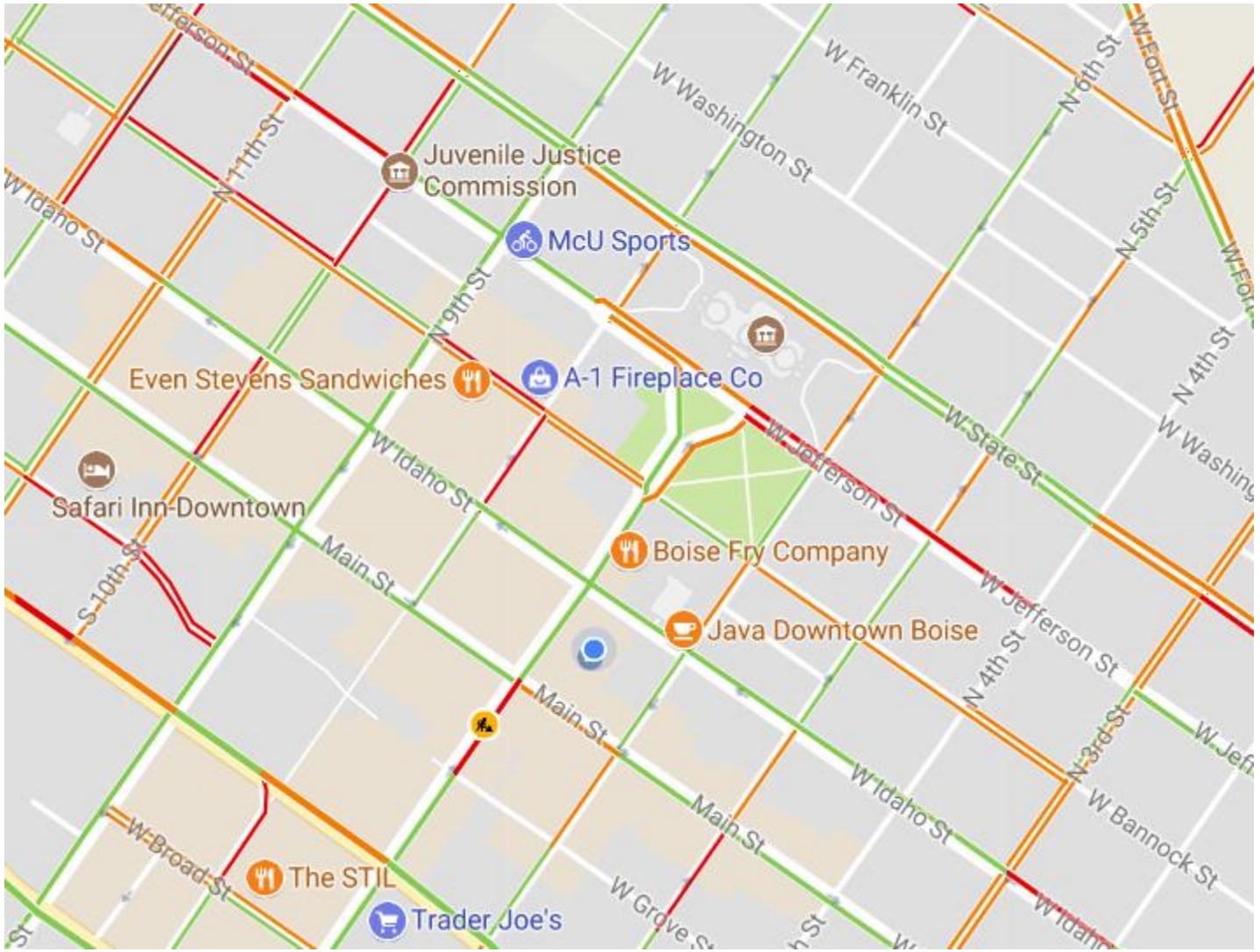
USE ELEMENTS FOR IMPACT

Use existing color conventions.



PRINCIPLE #6:

USE ELEMENTS FOR IMPACT



USE ELEMENTS FOR IMPACT

TRANSPORTATION SCORECARD

Meeting Target | Needs Improvement | Not Meeting Target | No Target

▲ Increase or ▼ Decrease since Prior Reporting Period

TRANSIT PERFORMANCE

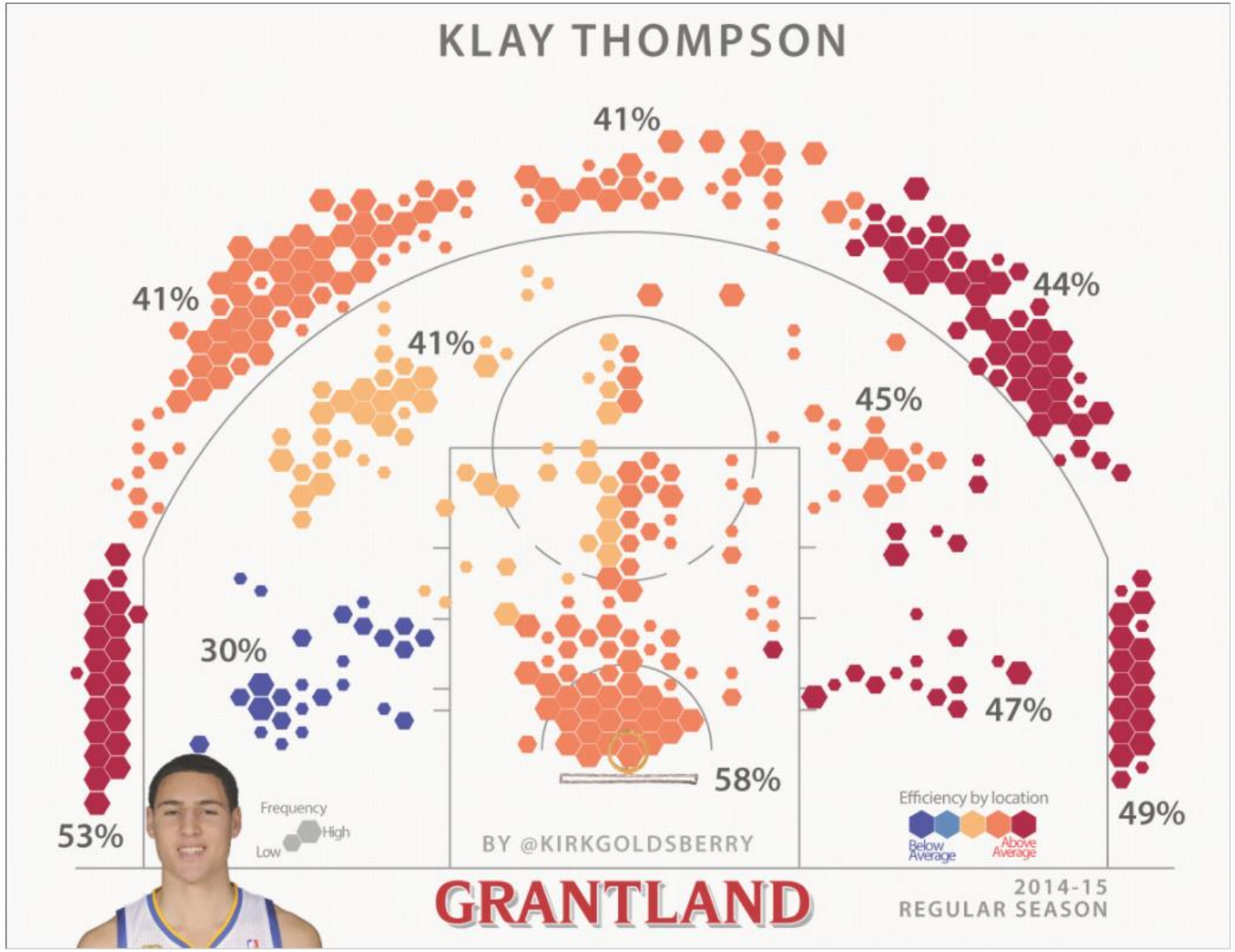
Transit Trips with Bunching or Gaps Between Vehicles
Target: 10.6% combined for bunching and gaps
In July 2017
25.9% ▲

Ridership
Target: 236,995,149 passengers carried (annual)
Average weekday boardings in August 2017
725,080 ▲

Percentage of Scheduled Service Hours Delivered

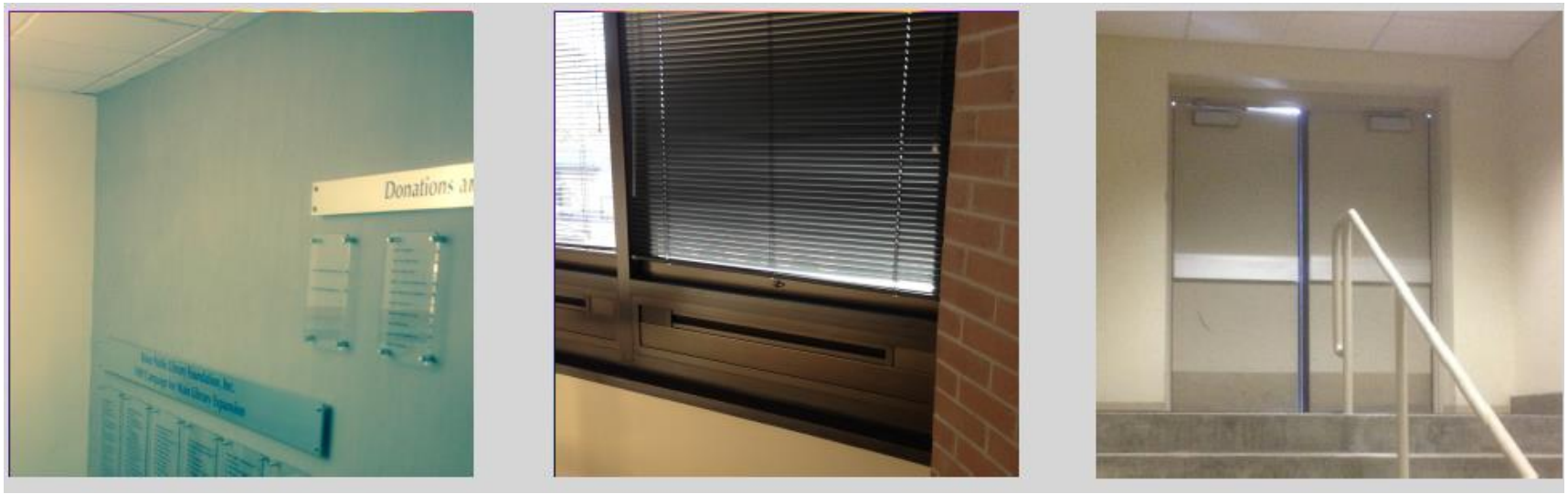
PRINCIPLE #6:

USE ELEMENTS FOR IMPACT



PRINCIPLE #6:

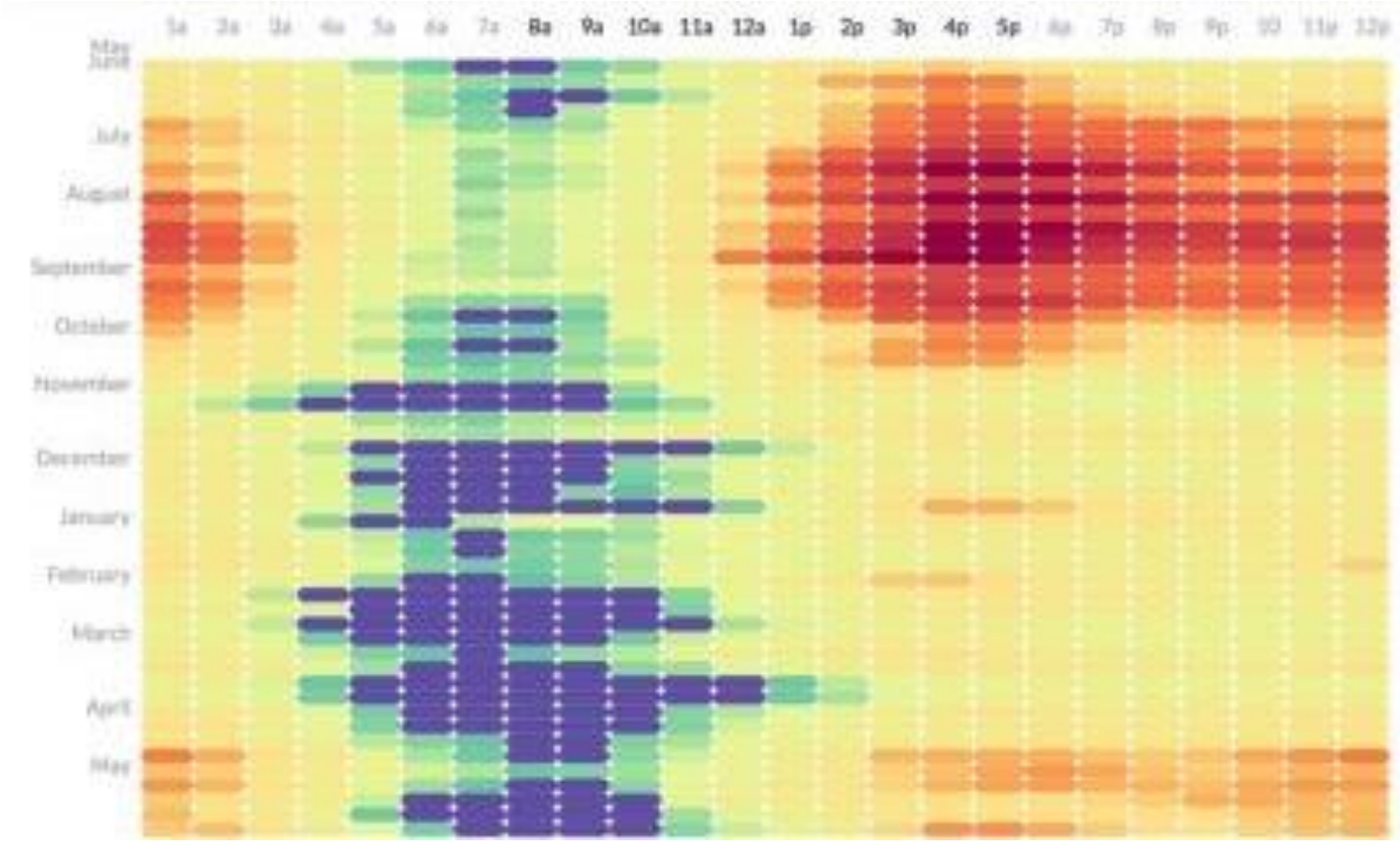
USE ELEMENTS FOR IMPACT



PRINCIPLE #6:

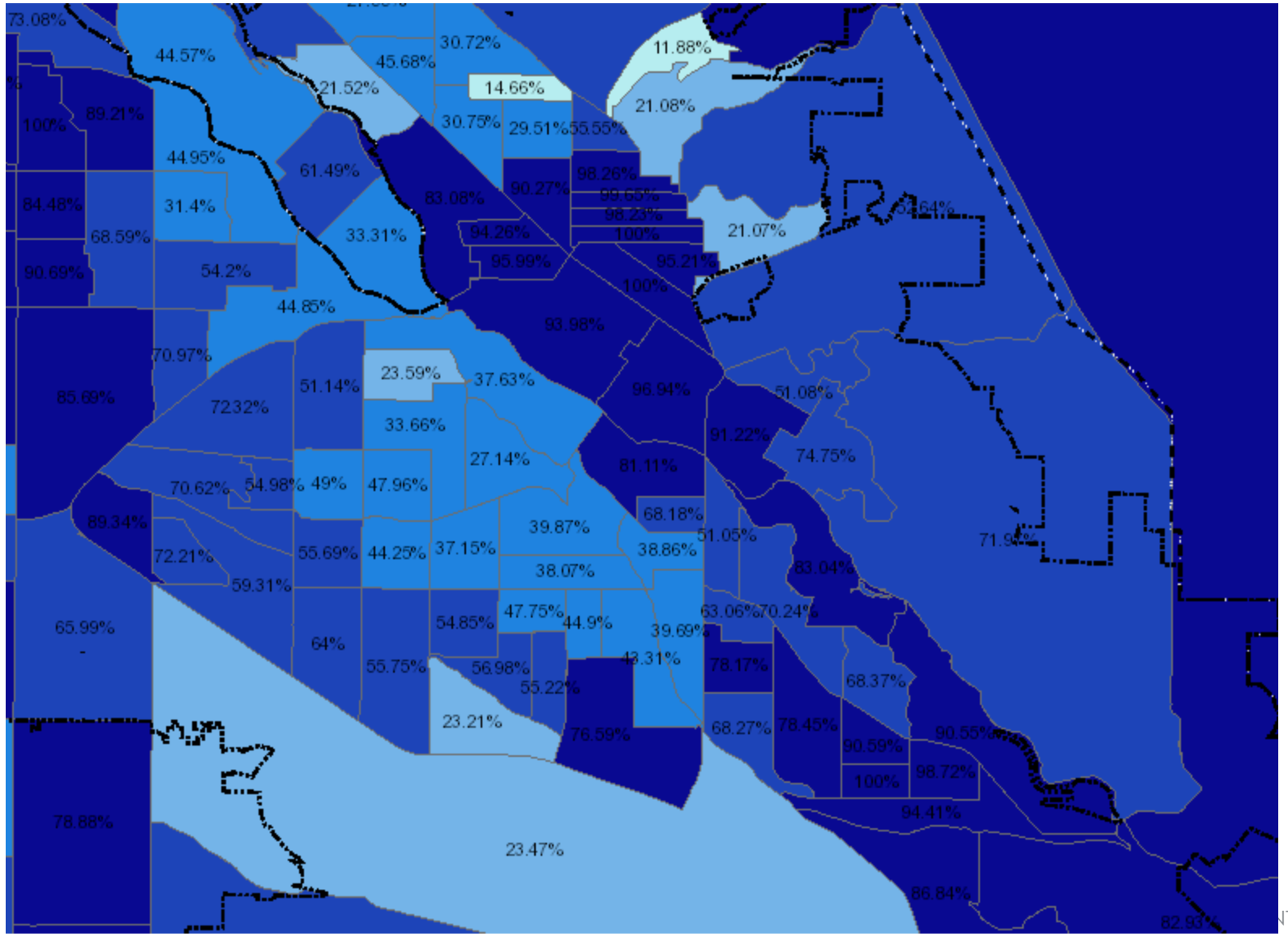
USE ELEMENTS FOR IMPACT

City Hall Energy Use



PRINCIPLE #6:

USE ELEMENTS FOR IMPACT

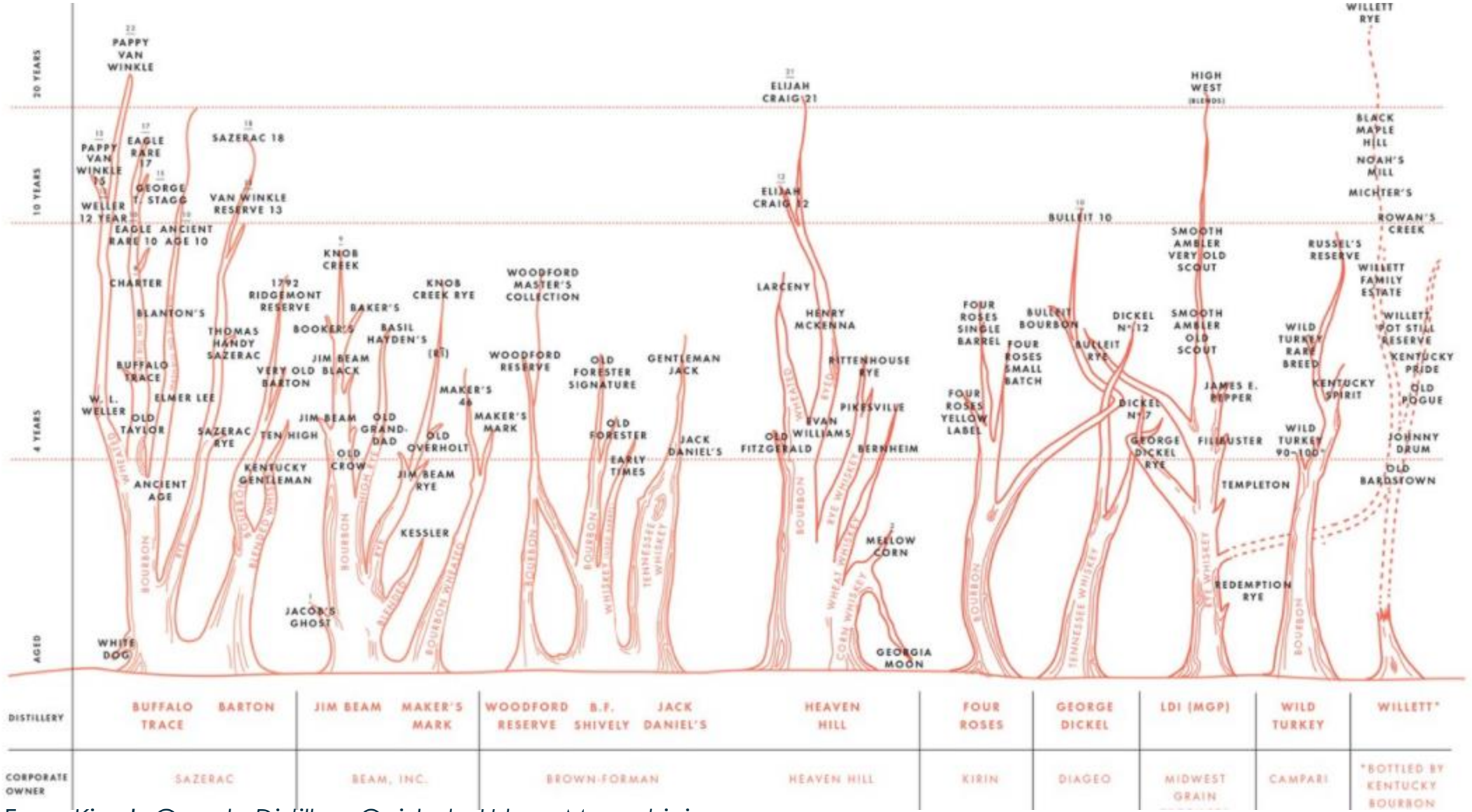


DATA VIZ PRINCIPLES

#7: **DOESN'T HURT TO BE VISUALLY APPEALING**

PRINCIPLE #7:

DOESN'T HURT TO BE VISUALLY APPEALING



From *King's County Distillery Guide to Urban Moonshining*

WHAT IS YOUR FAVORITE PRINCIPLE?

Visualizing Data

1. Know your message
2. Use the right chart
3. Compared to what?
4. Don't mislead
5. Efficiency!
6. Use chart elements for impact
7. It doesn't hurt to be visually appealing (in data viz and life)

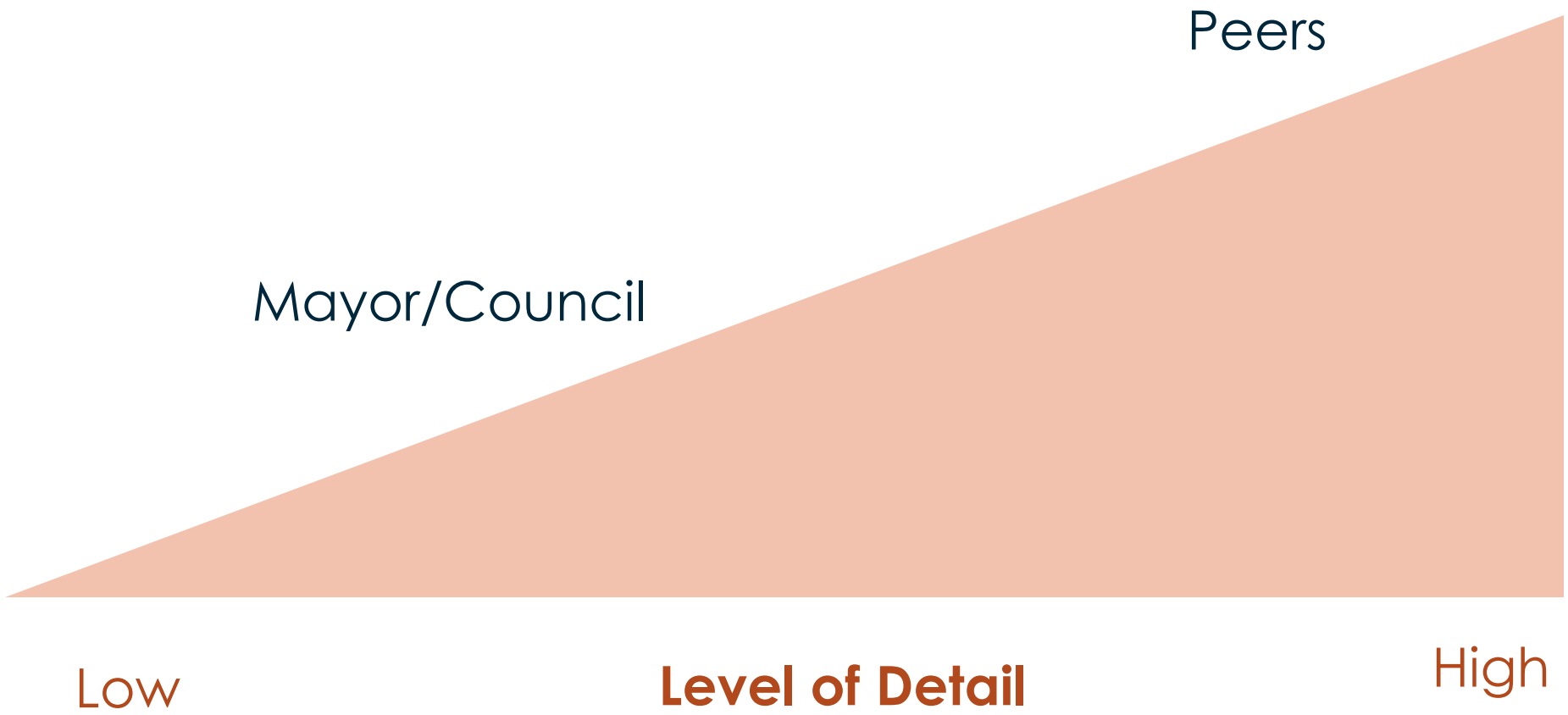
DATA PREZ PRINCIPLES

DATA PREZ PRINCIPLES

#1: **KNOW YOUR AUDIENCE**

PRINCIPLE #1:

KNOW YOUR AUDIENCE



DATA PREZ PRINCIPLES

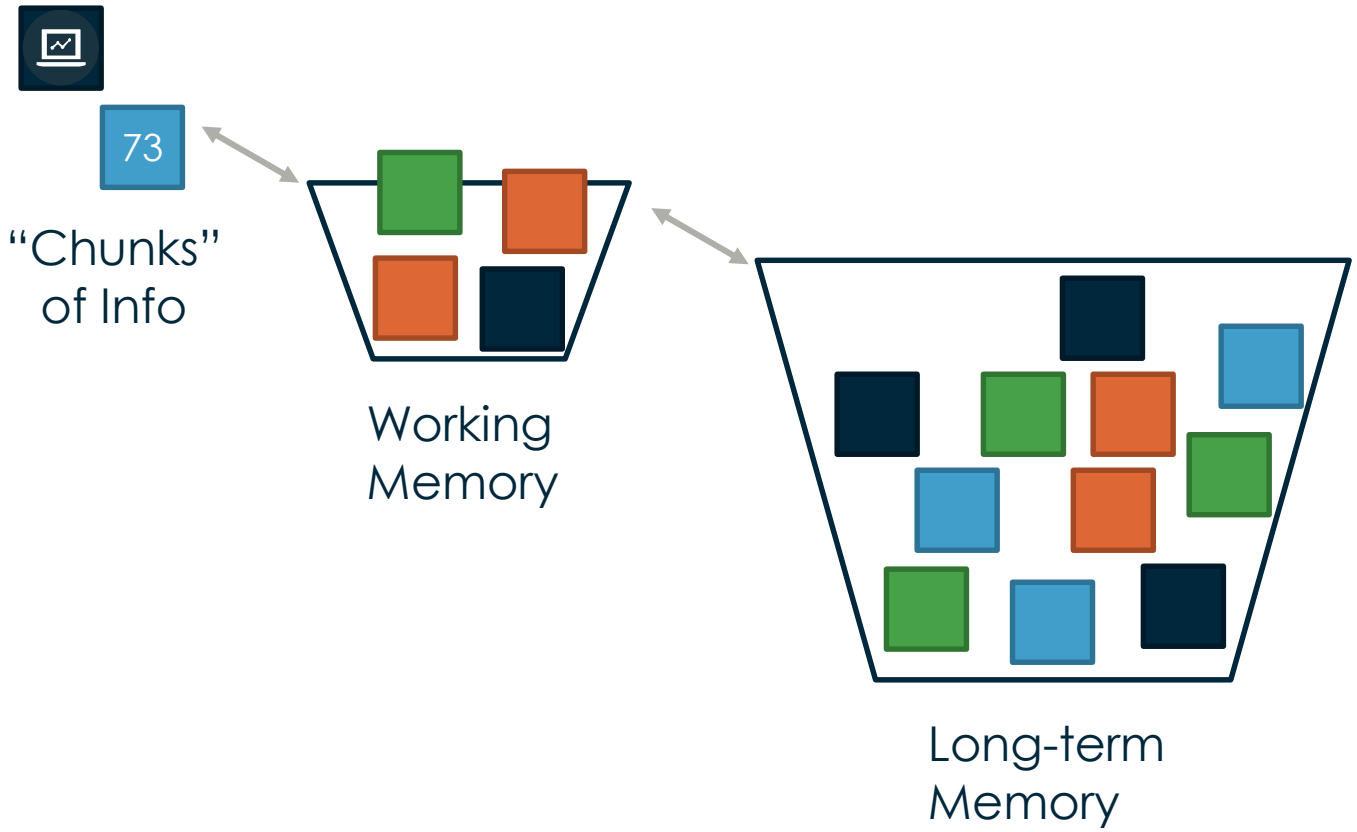
#2: **EFFICIENCY!**

PRINCIPLE #2:

EFFICIENCY!

Be concise.

Know everything, but only present key takeaways.



DATA PREZ PRINCIPLES

#3: **SO WHAT?**

PRINCIPLE #3:

SO WHAT?

Context

Key Findings

Recommendations

PRINCIPLE #3:

SO WHAT?

Resource prioritization

Further study

Program improvements

Recommendations

Maintain status quo

Resource deployment

Change to law, policy, practice.

PRINCIPLE #3:

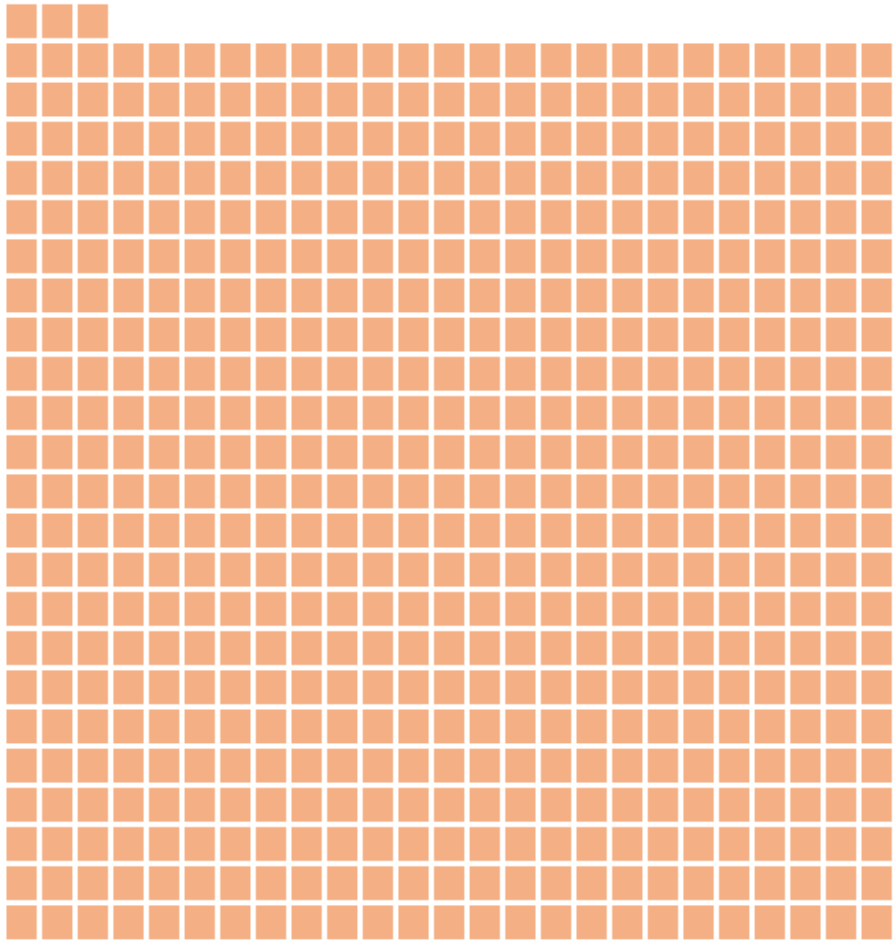
SO WHAT?

Example:

John Snow and the 1854 London Cholera outbreak

PRINCIPLE #3:

SO WHAT?

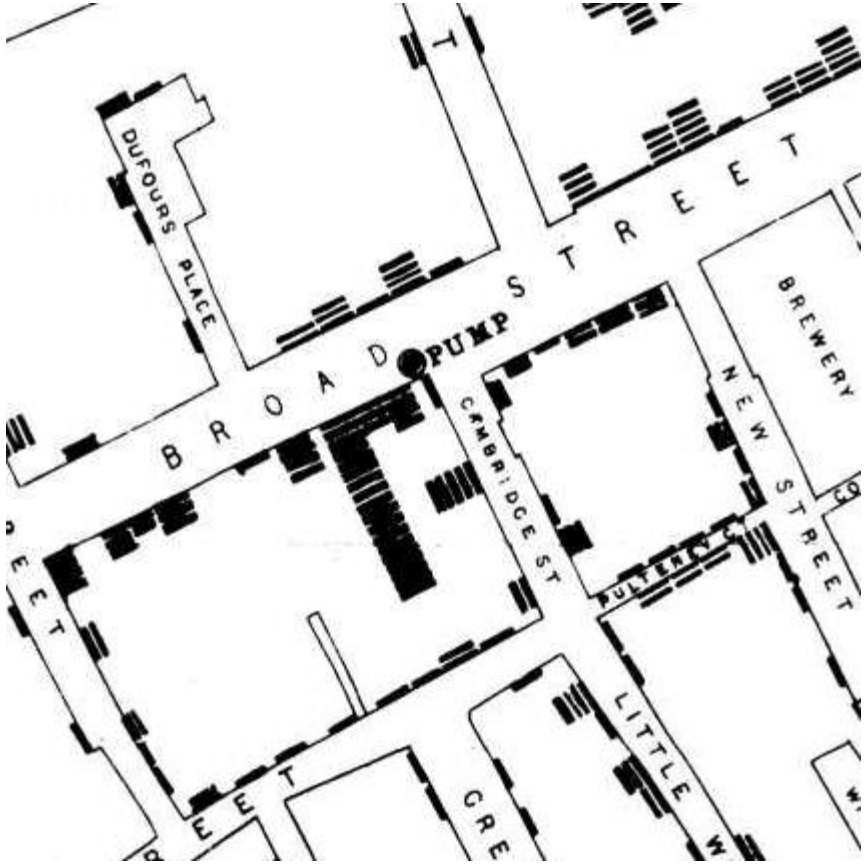


= 1 life

578 residents have died of Cholera in the past 3 weeks.

PRINCIPLE #3:

SO WHAT?



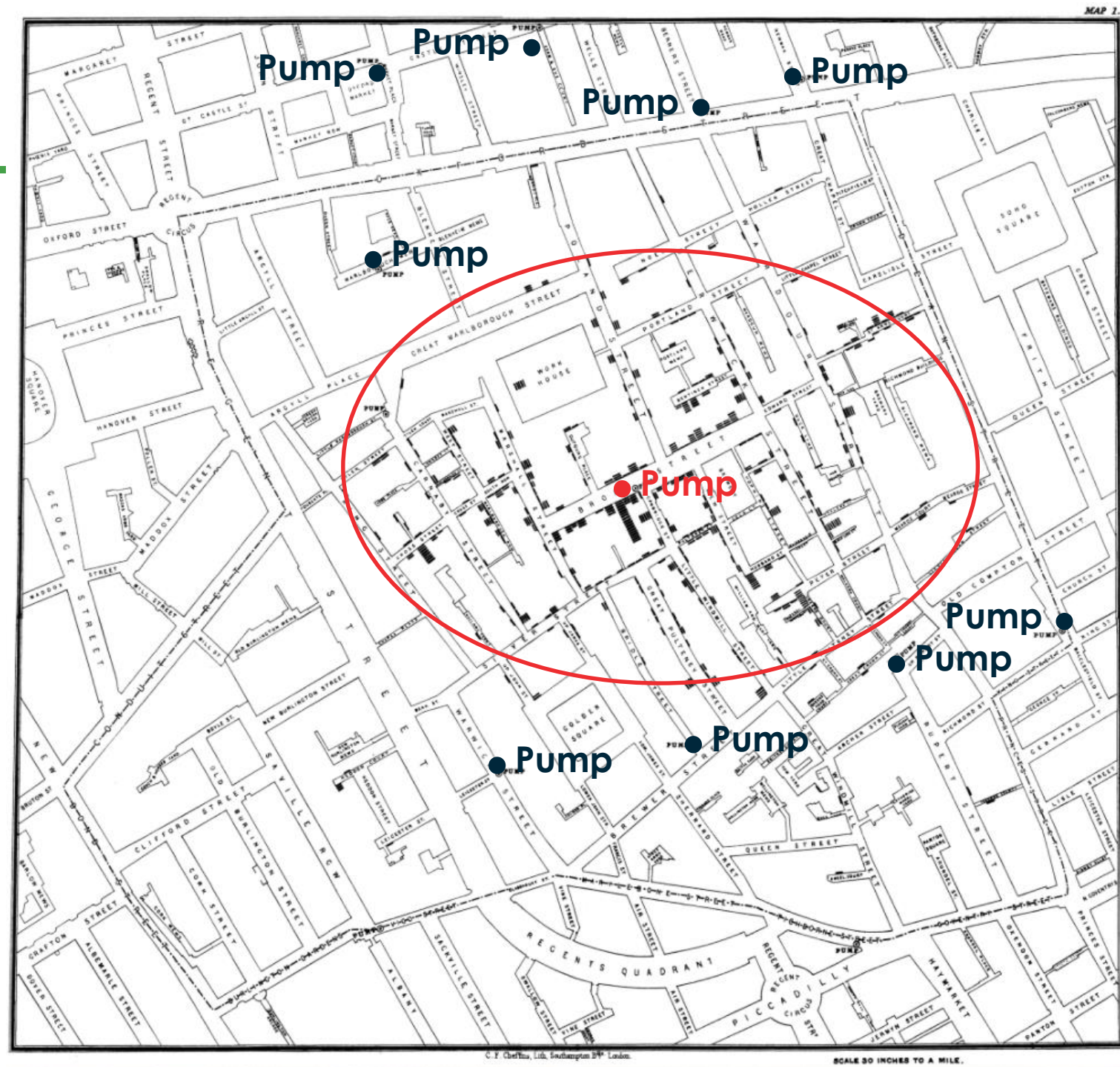
PRINCIPLE #3:

SO WHAT?

Key findings:

99% are within 1/8 mile of one water pump.

Unidentified bacteria found at the pump.



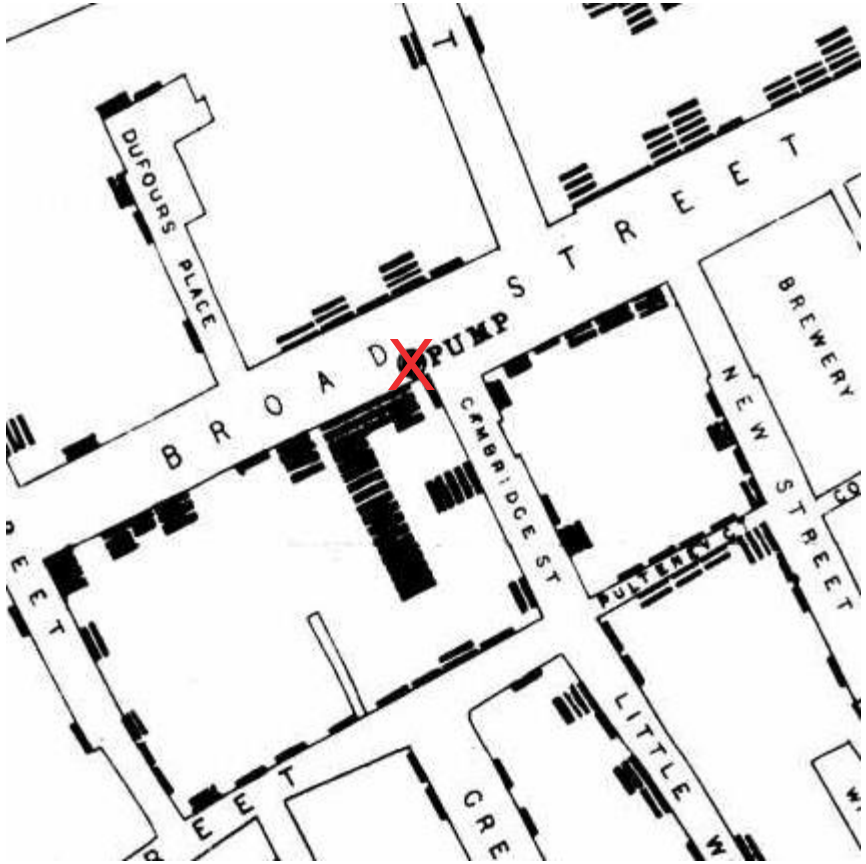
PRINCIPLE #3:

SO WHAT?

Recommendations:

Remove pump handle immediately.

Further study of bacteria needed to prevent future outbreaks.



Activity time!

Practice with presenting

QUICK TIPS ON PRESENTING

- Don't read the slides (don't judge me!)
- Keep it visual and limit text
- Show viz, count to 5, start speaking
- Allow for audience interaction

Activity time!

Hans Redux

KEY PRINCIPLES

Visualizing Data

1. Know your message
2. Use the right chart
3. Compared to what?
4. Don't mislead
5. Efficiency!
6. Use chart elements for impact
7. It doesn't hurt to be visually appealing (in data viz and life)

Presenting Data

1. Know your audience
2. Efficiency! (again)
3. So What?

NEXT TIME

Dashboarding and Power BI!!!

Friday, 8:30am-12:30pm

RESOURCES

Inspiration

- Data Viz Done Right
- NY Times - The Upshot
- The Economist tweets
- FiveThirtyEight
- Information is Beautiful

Use the Right Chart

- Infogram

I have books! You can borrow them!

- Good Charts
- Information Dashboard Design
- Knowledge is Beautiful
- Edward Tufte books (4 of them)

Activity time!

Plus Delta

Appendices

OUTLINE

- Intro/Context – 35 min — 1:00 - 1:35pm
  - Activity – 10 min
  - Slides – 25 min
- Principles of data viz – 100 min — 1:35 - 3:15pm
  - Slides – 80 min
  - Break – 10 min
  - Activity – 10 min
- Presenting data – 30 min — 3:15 – 3:45pm
  - Slides – 20 min
  - Break – 10 min
- Final activities – 75 min — 3:45 – 5:00pm
  - Practice Presenting – 45 min
  - Hans Redux – 20 min
  - Plus/delta – 10 min

TUFTE'S FAVORITE VIZ

"Losses of the French Army in the Russian Campaign"
By Charles Joseph Minard