

# Information Visualization Crash Course

(AKA Information Visualization 101)

Dr. Chad Stolper  
Now at Google  
Georgia Tech CS PhD Alum



What is Infovis?

Why is it Important?

Human Perception

Chart Basics

(If Time, Some Color Theory)

The Shneiderman Mantra

Where to Learn More

What is Information Visualization?

# Information Visualization

“The use of **computer**-supported, **interactive**, **visual** representations of abstract data to **amplify cognition.**”

Card, Mackinlay, and Shneiderman 1999

# Communication

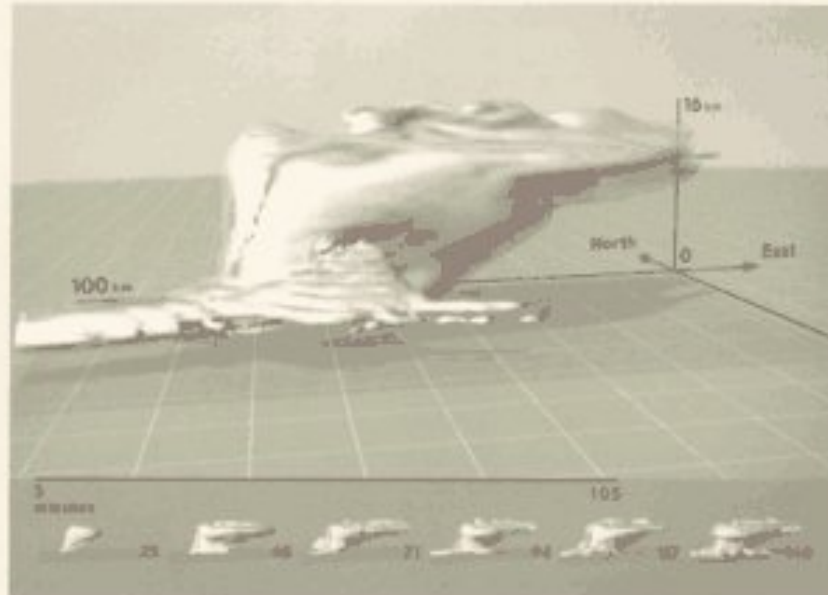
## Exploratory Data Analysis (EDA)

# Communication

(gone wrong)

EDWARD R. TUFTE

# VISUAL EXPLANATIONS



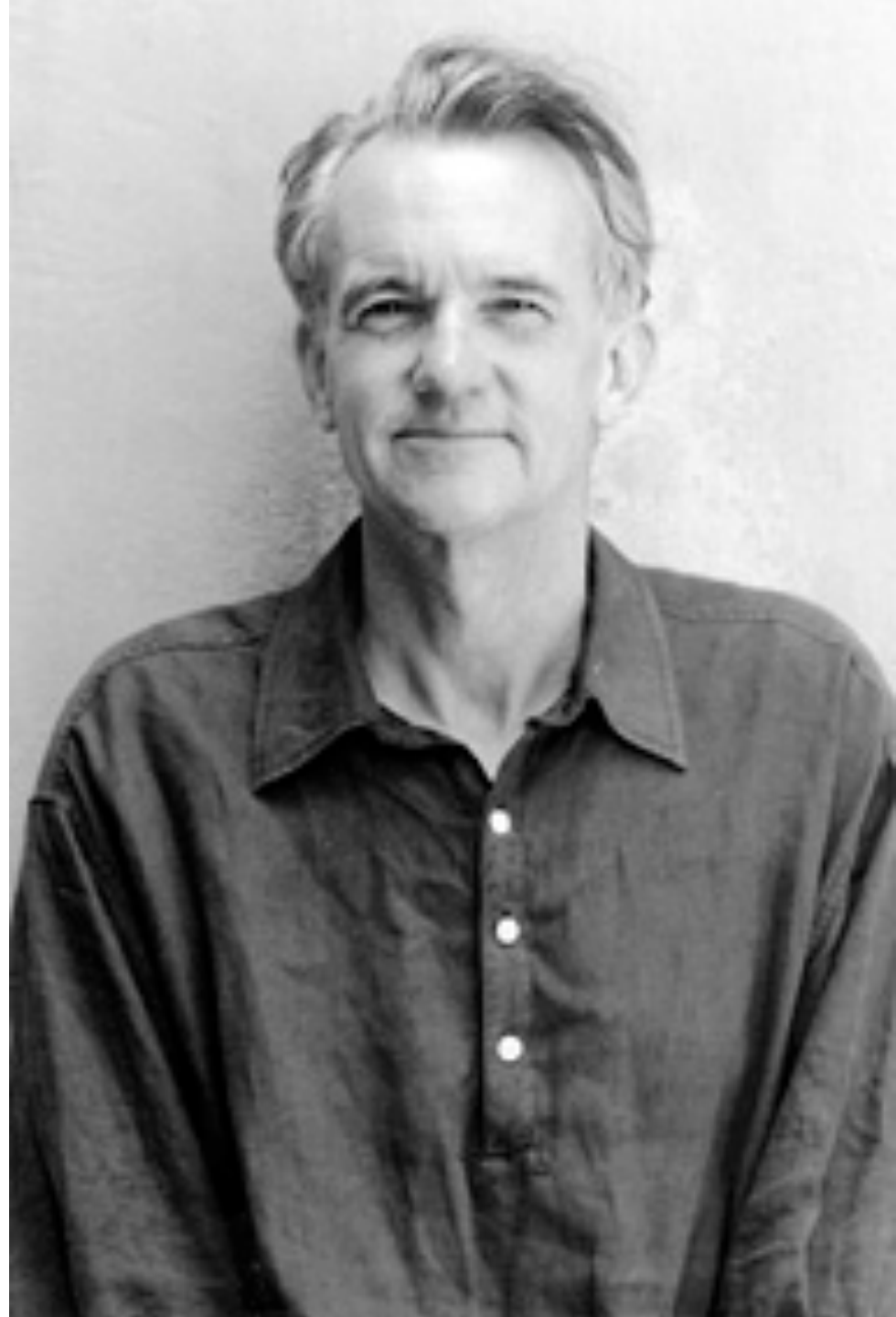
IMAGES AND QUANTITIES, EVIDENCE AND NARRATIVE

# Edward Tufte

An American statistician and professor emeritus of political science, statistics, and computer science at Yale University.

He is noted for his writings on information design and as a **pioneer in the field of data visualization.**

-Wikipedia

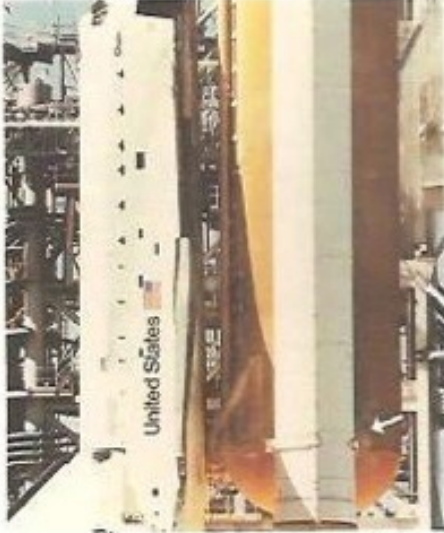




# Space Shuttle Challenger

January 28, 1986

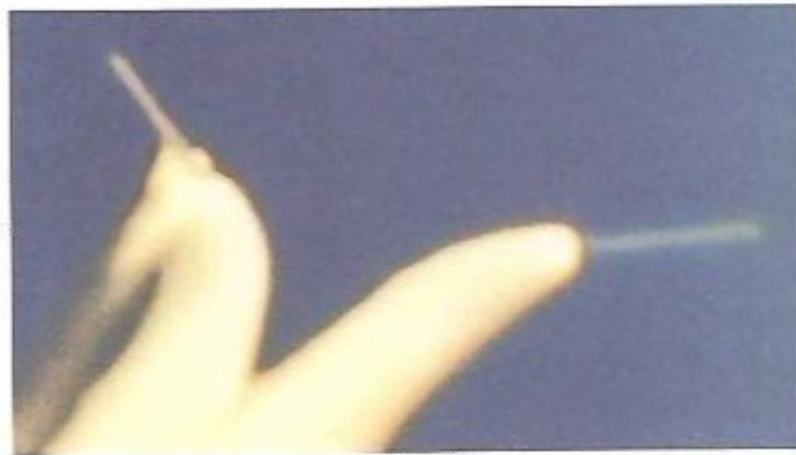
Morning Temperature: 31°F



Less than 1 second after ignition, a puff of smoke appeared at the aft joint of the right booster, indicating that the O-rings burned through and failed to seal. At this point, all was lost.



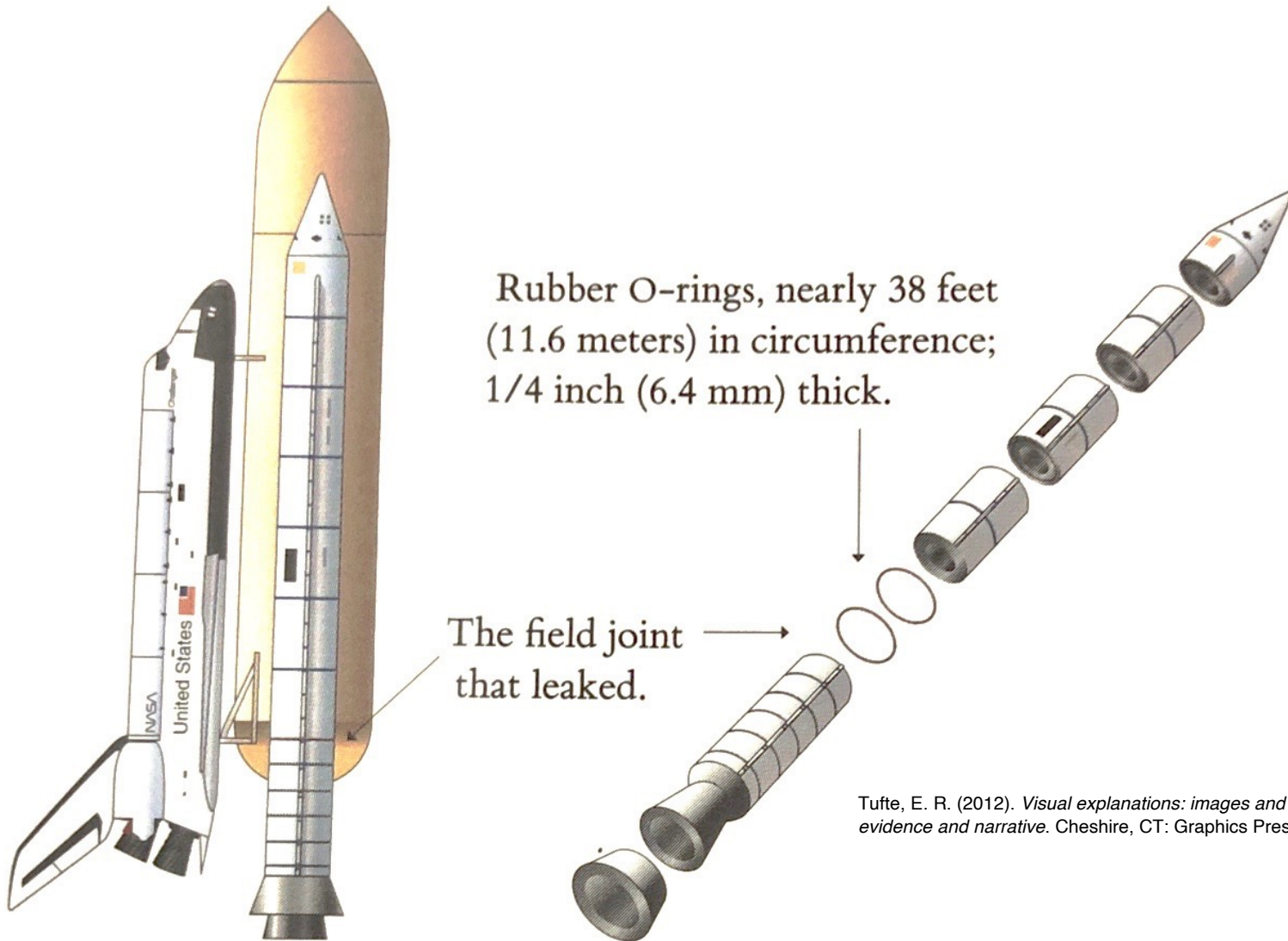
On the launch pad, the leak lasted only about 2 seconds and then apparently was plugged by pyroclastic material and insulation as the shuttle rose, flying through rather strong cross-winds. Then 58.788 seconds after ignition, when the Challenger was 6 miles up, a flicker of flame emerged from the leaky joint. Within 10 seconds, the flame grew and engulfed the fuel tank (containing liquid hydrogen and liquid oxygen). That tank ruptured and exploded, destroying the shuttle.



As the shuttle exploded and broke up at approximately 73 seconds after launch, the two booster rockets crisscrossed and continued flying wildly. The right booster, identifiable by its failure plume, is now to the left of its non-defective counterpart.



The flight crew of Challenger 51-L. Front row, left to right: Ellison S. Onizuka, pilot; Francis R. (Dick) Scobee, commander; Ronald E. McNair. Back row: Ellison S. Onizuka, S. Christa McAuliffe, Gregory B. Burch, Judith A. Resnik.



Rubber O-rings, nearly 38 feet (11.6 meters) in circumference; 1/4 inch (6.4 mm) thick.

The field joint that leaked.

Tufte, E. R. (2012). *Visual explanations: images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press.

# Most Watched Science Experiment

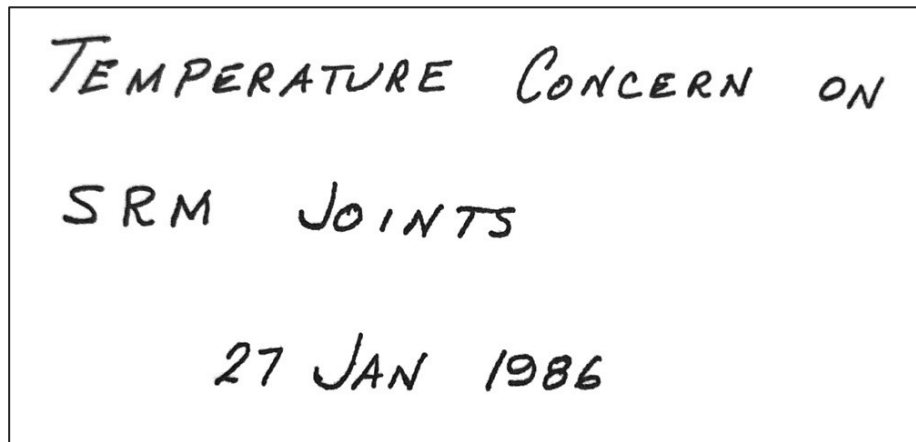


Richard Feynman, Physics Nobel laureate explained how rubber became rigid in cold temperate

YouTube video:  
<https://youtu.be/6Rwcbsn19c0>

How did this happen?

Engineers at **Morton Thiokol**, the rocket maker, presented on the day before and recommended **not to launch**.



TEMPERATURE CONCERN ON  
SRM JOINTS  
27 JAN 1986


## CONCLUSIONS :

- TEMPERATURE OF O-RING IS NOT ONLY PARAMETER CONTROLLING BLOW-BY  
SRM 15 WITH BLOW-BY HAD AN O-RING TEMP AT 53°F  
SRM 22 WITH BLOW-BY HAD AN O-RING TEMP AT 75°F  
FOUR DEVELOPMENT MOTORS WITH NO BLOW-BY WERE TESTED AT O-RING TEMP OF 47° TO 52°F  
DEVELOPMENT MOTORS HAD PUTTY PACKING WHICH RESULTED IN BETTER PERFORMANCE
- AT ABOUT 50°F BLOW-BY COULD BE EXPERIENCED IN CASE JOINTS
- TEMP FOR SRM 25 ON 1-26-86 LAUNCH WILL BE 29°F 9 AM  
38°F 2 PM
- HAVE NO DATA THAT WOULD INDICATE SRM 25 IS DIFFERENT THAN SRM 15 OTHER THAN TEMP

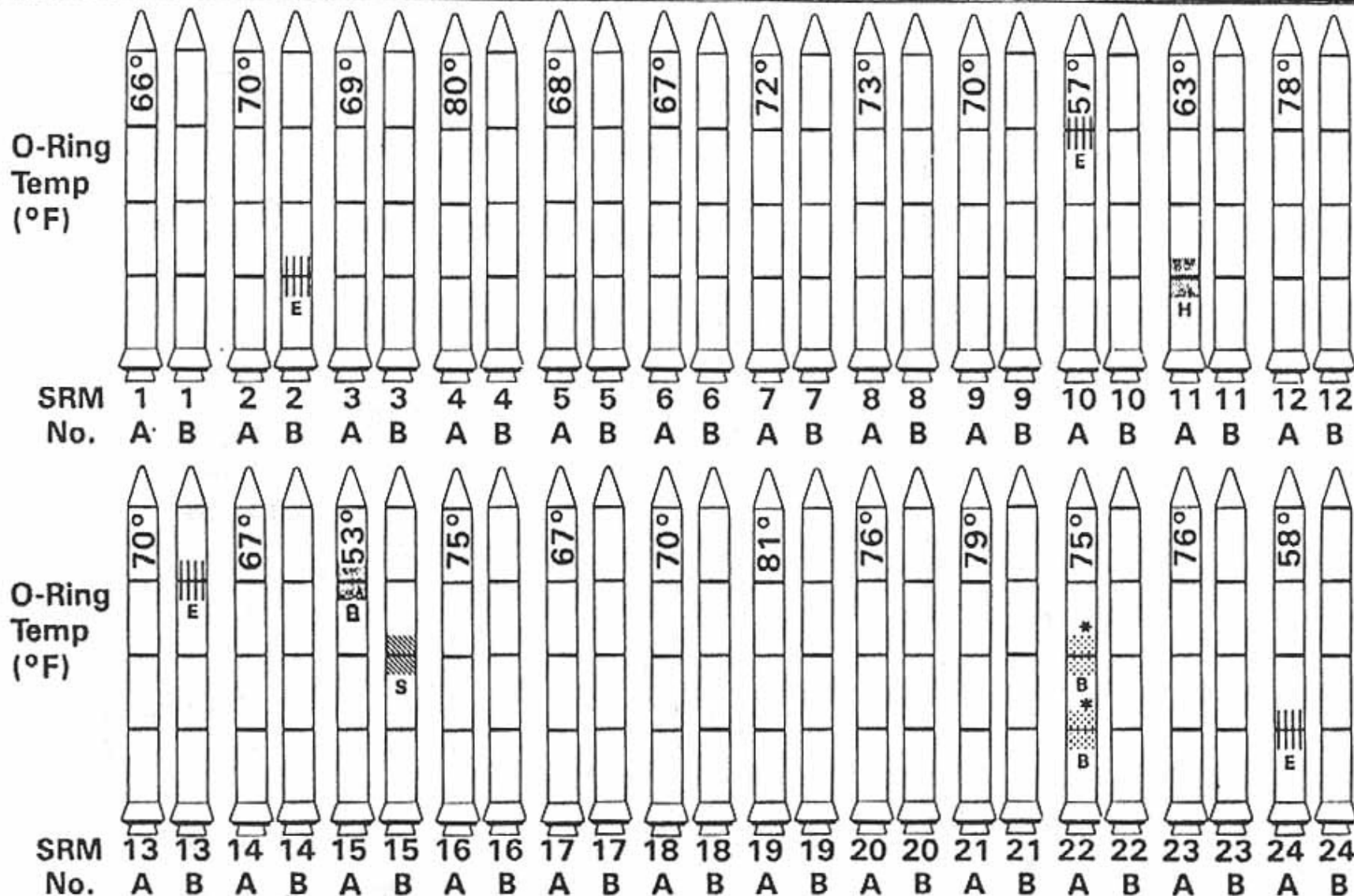
## RECOMMENDATIONS :

- O-RING TEMP MUST BE  $\geq 53$  °F AT LAUNCH  
DEVELOPMENT MOTORS AT 47° TO 52°F WITH PUTTY PACKING HAD NO BLOW-BY  
SRM 15 (THE BEST SIMULATION) WORKED AT 53 °F
- PROJECT AMBIENT CONDITIONS (TEMP & WIND) TO DETERMINE LAUNCH TIME

## RECOMMENDATIONS :

- 
- O-RING TEMP MUST BE  $\geq 53$  °F AT LAUNCH  
DEVELOPMENT MOTORS AT 47° TO 52°F WITH PUTTY PACKING HAD NO BLOW-BY  
SRM 15 (THE BEST SIMULATION) WORKED AT 53 °F
  - PROJECT AMBIENT CONDITIONS (TEMP & WIND) TO DETERMINE LAUNCH TIME

# History of O-Ring Damage in Field Joints (Cont)



MORTON THIOKOL, INC.  
Wasatch Operations

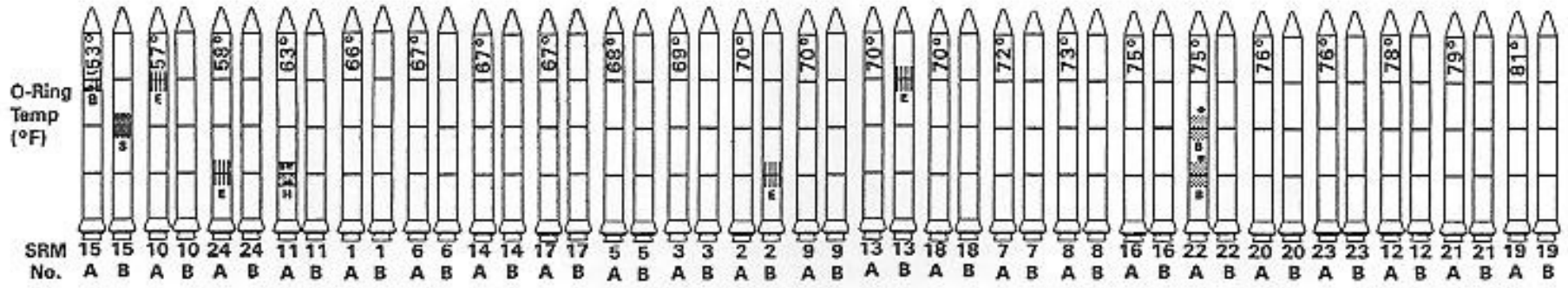
\* No Erosion

86486-1E

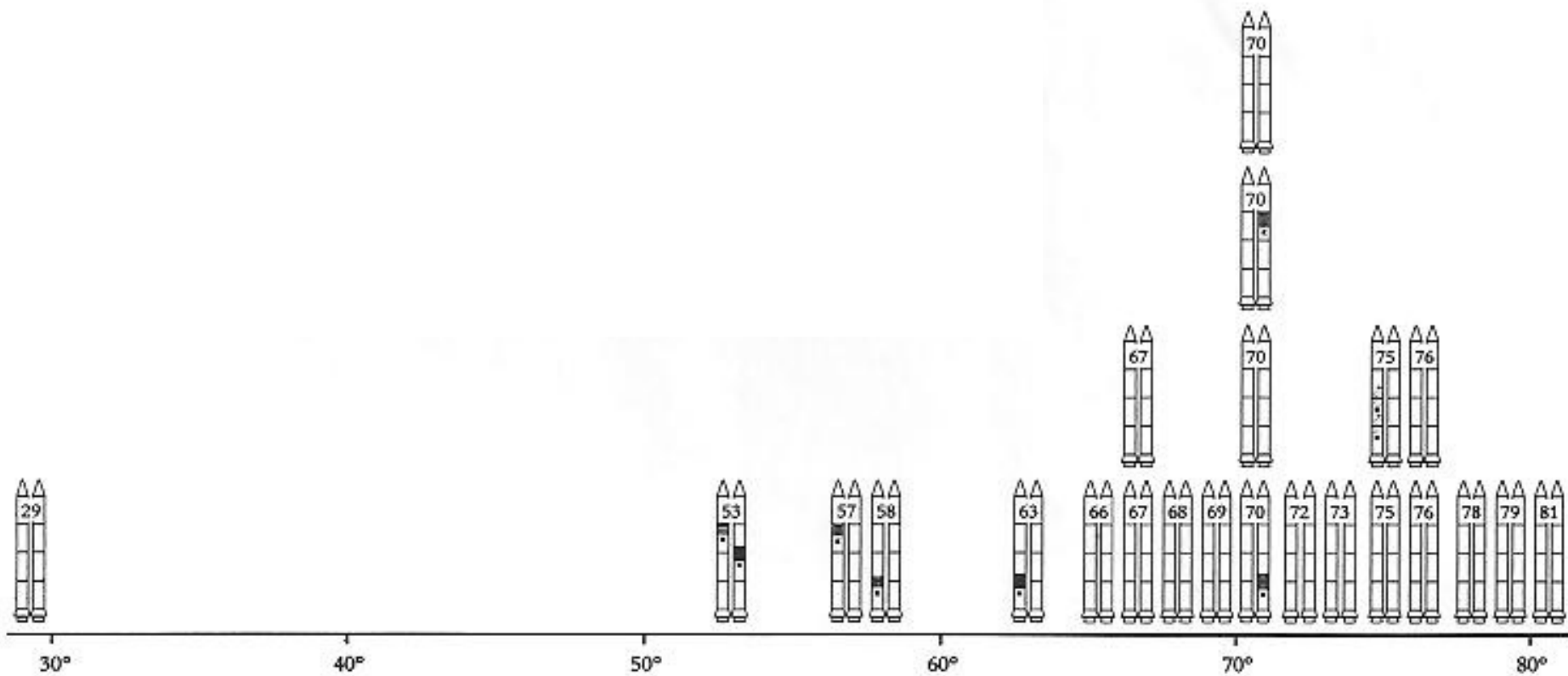
INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

[Ref. 2/26-2 2 of 3]



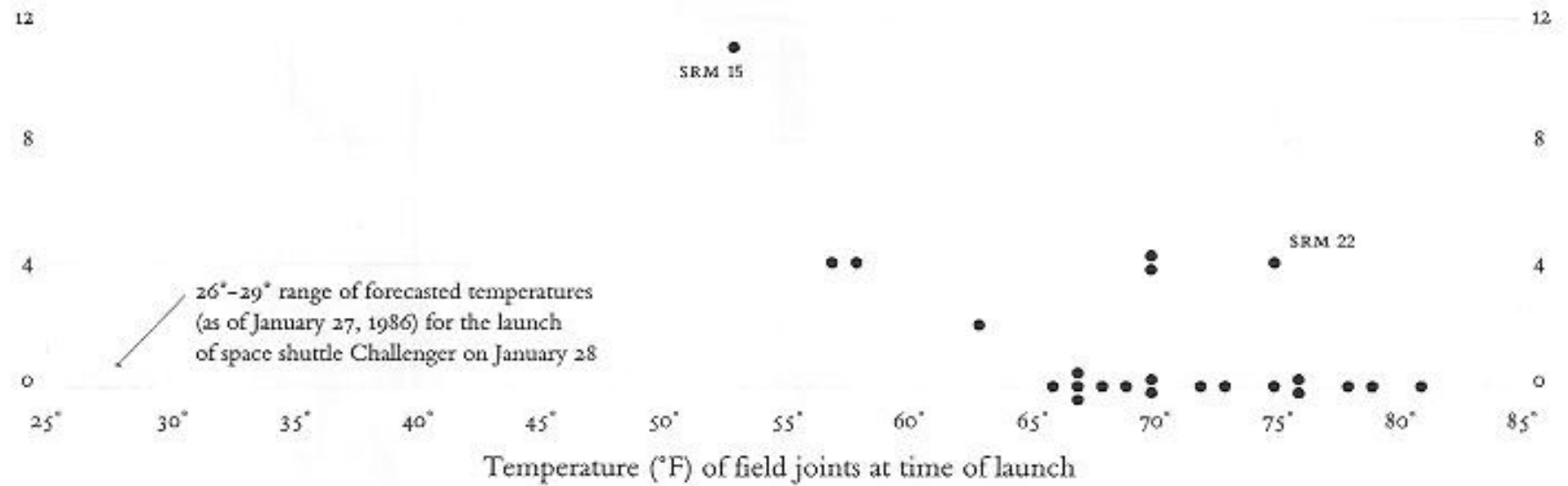


\* No Erosion



Flight	Date	Temperature °F	Erosion incidents	Blow-by incidents	Damage index	Comments
51-C	01.24.85	53°	3	2	11	Most erosion any flight; blow-by; back-up rings heated.
41-B	02.03.84	57°	1		4	Deep, extensive erosion.
61-C	01.12.86	58°	1		4	O-ring erosion on launch two weeks before Challenger.
41-C	04.06.84	63°	1		2	O-rings showed signs of heating, but no damage.
1	04.12.81	66°			0	Coollest (66°) launch without O-ring problems.
6	04.04.83	67°			0	
51-A	11.08.84	67°			0	
51-D	04.12.85	67°			0	
5	11.11.82	68°			0	
3	03.22.82	69°			0	
2	11.12.81	70°	1		4	Extent of erosion not fully known.
9	11.28.83	70°			0	
41-D	08.30.84	70°	1		4	
51-G	06.17.85	70°			0	
7	06.18.83	72°			0	
8	08.30.83	73°			0	
51-B	04.29.85	75°			0	
61-A	10.30.85	75°		2	4	No erosion. Soot found behind two primary O-rings.
51-I	08.27.85	76°			0	
61-B	11.26.85	76°			0	
41-G	10.05.84	78°			0	
51-J	10.03.85	79°			0	
4	06.27.82	80°			?	O-ring condition unknown; rocket casing lost at sea.
51-F	07.29.85	81°			0	

O-ring damage  
index, each launch



So, communication is  
**extremely important.**

Visualization can help with that –  
**communicate ideas and insights.**



Hans Rosling:

# The best stats you've ever seen

TED2006 · 19:50 · Filmed Feb 2006

Subtitles available in 48 languages

[https://www.ted.com/talks/hans\\_rosling\\_the\\_best\\_stats\\_you\\_ve\\_ever\\_seen](https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen)

Visualization can also help with  
**Exploratory Data Analysis (EDA)**

But why do you need to  
explore data at all???

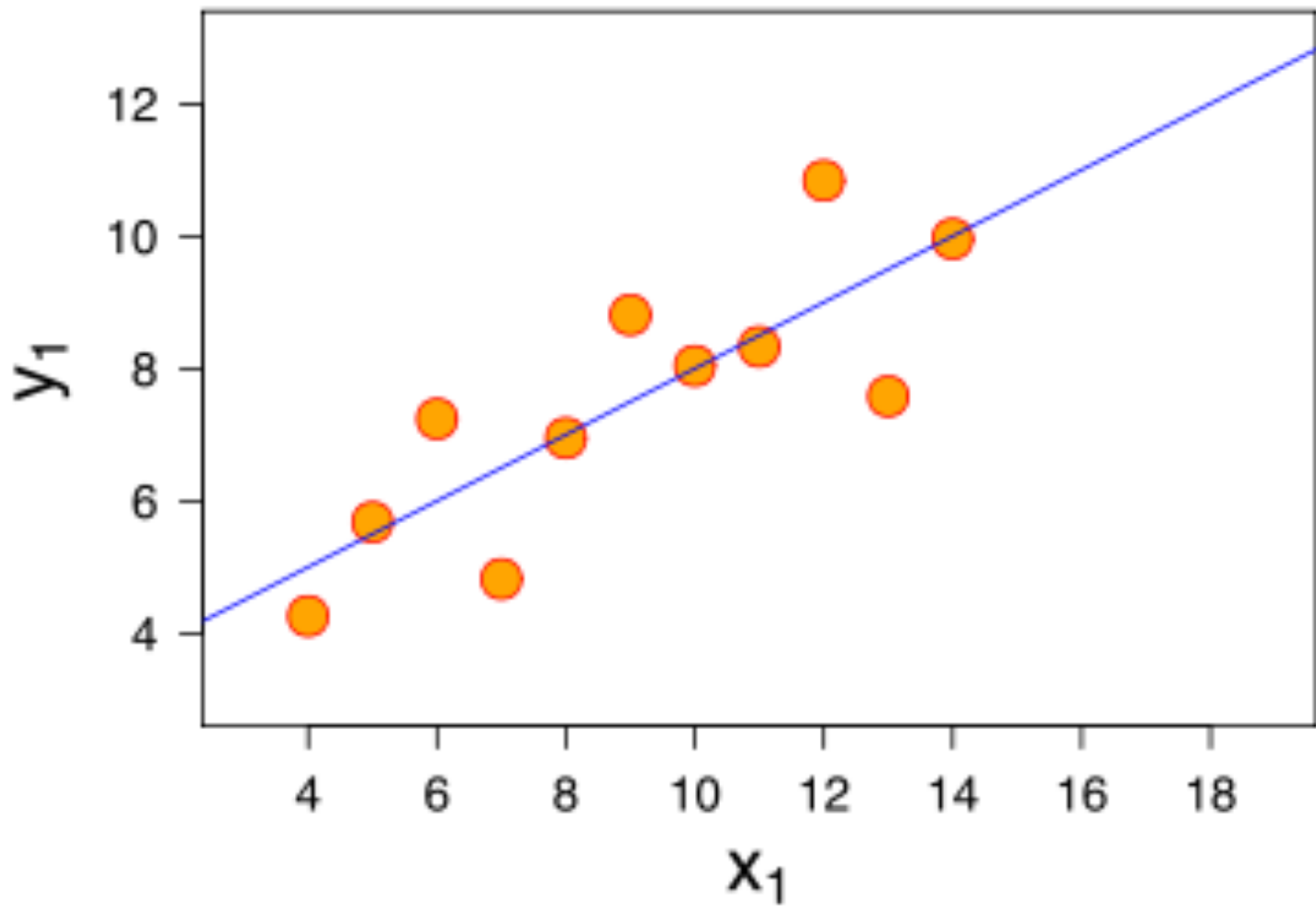
“There are three kinds of lies:  
lies, damned lies, and statistics.”

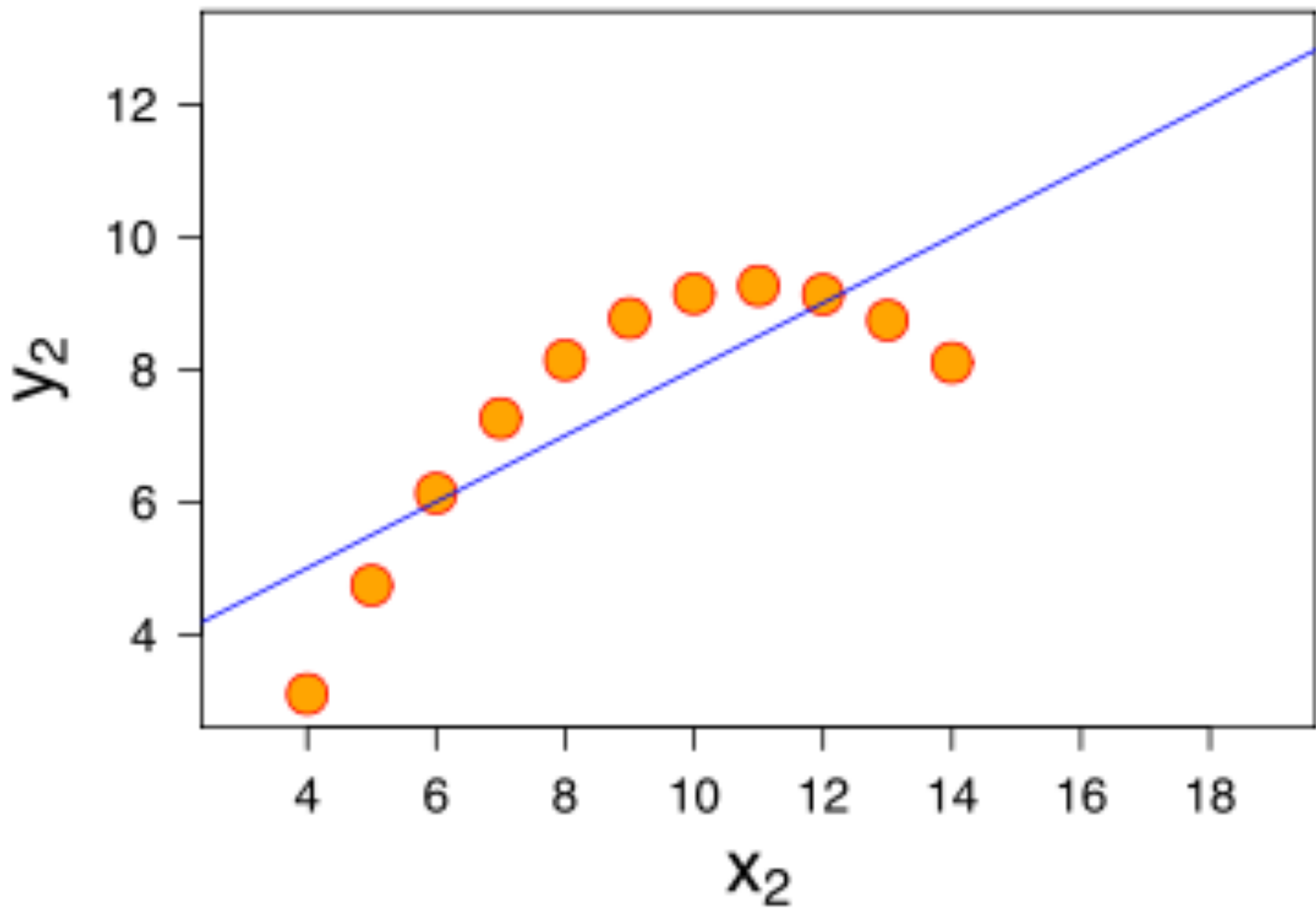


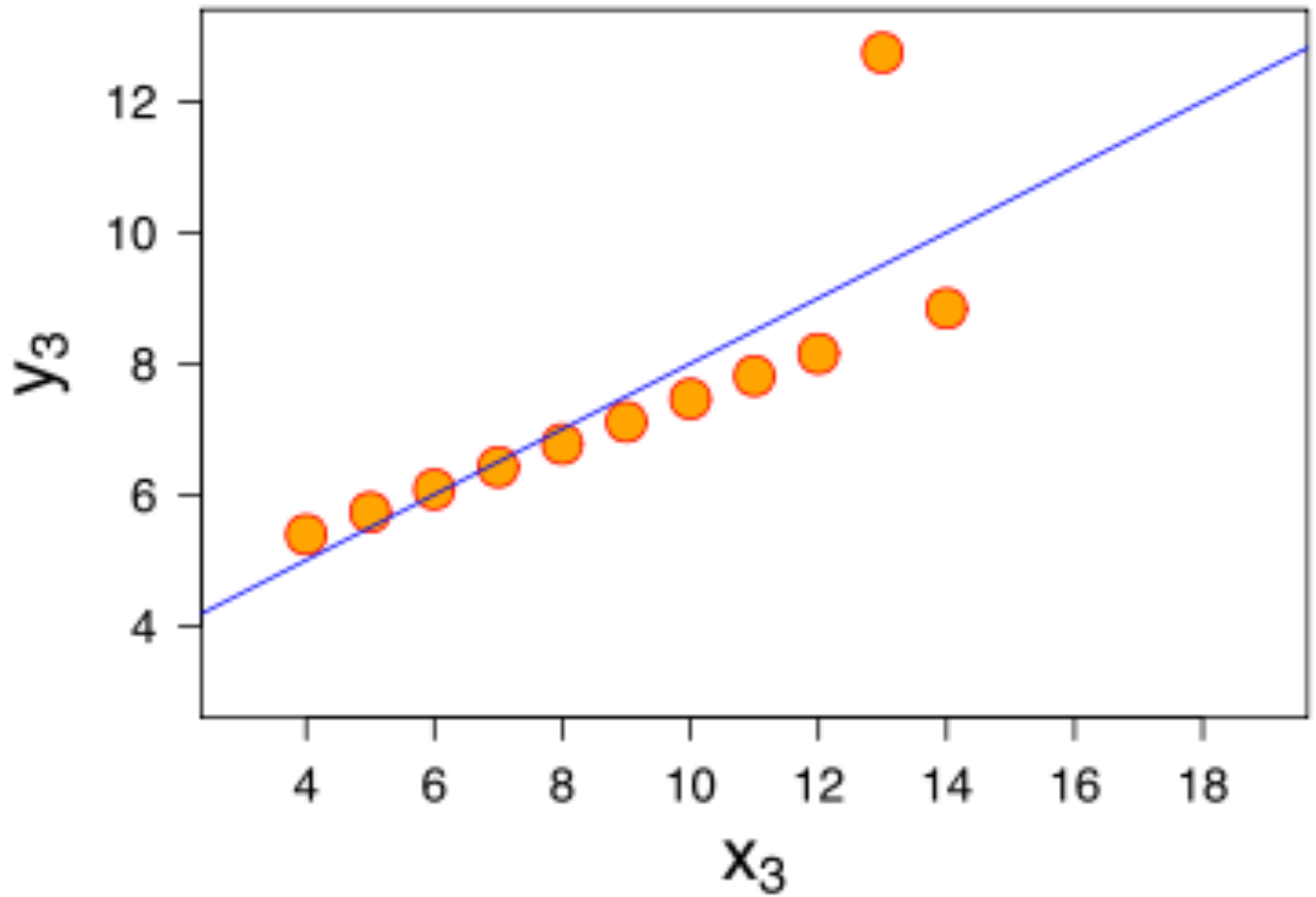
# Mystery Data Set

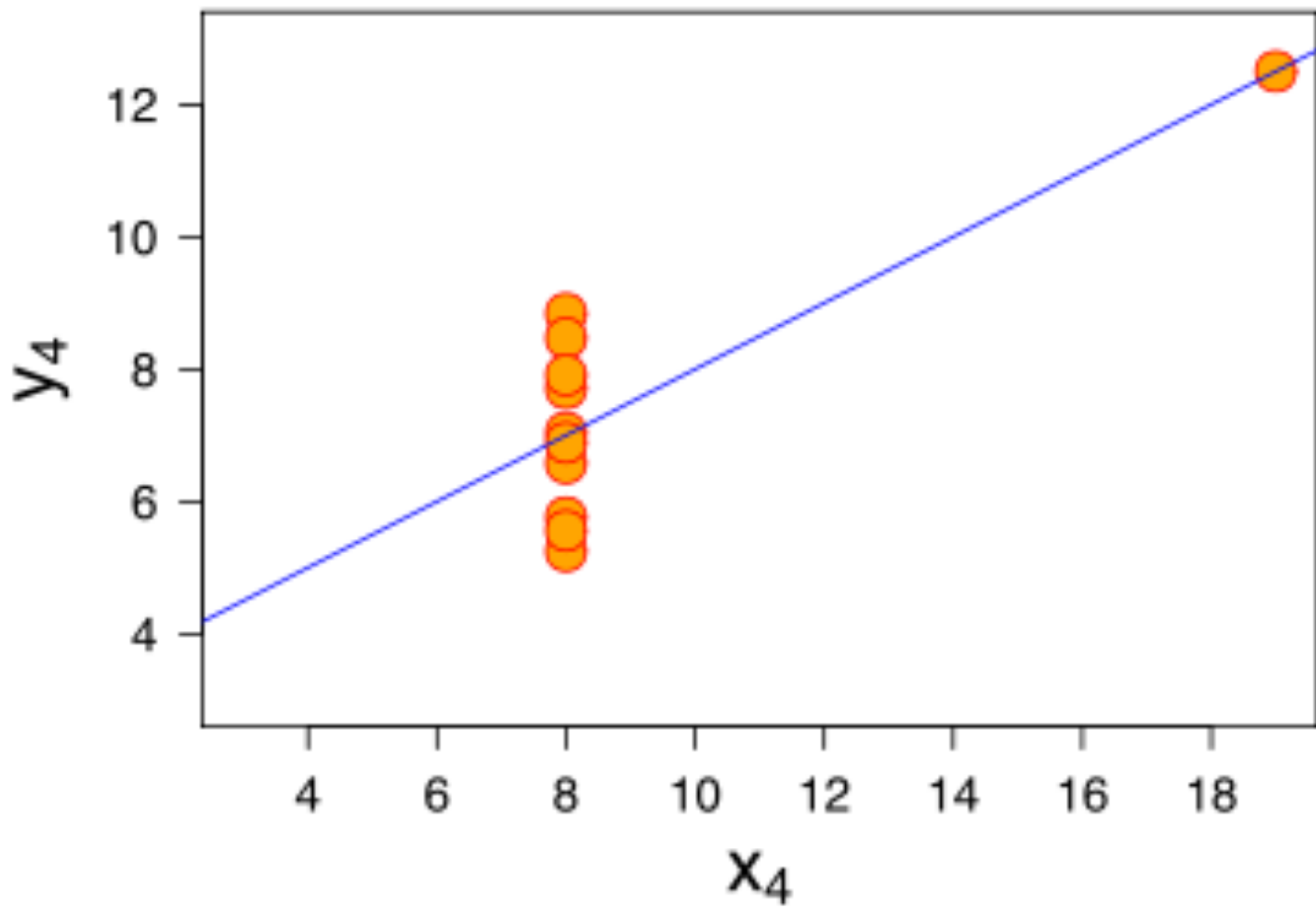
# Mystery Data Set

Property	Value
mean( x )	9
variance ( x )	11
mean( y )	7.5
variance ( y )	4.122
correlation ( x,y )	0.816
Linear Regression Line	$y = 3 + 0.5x$

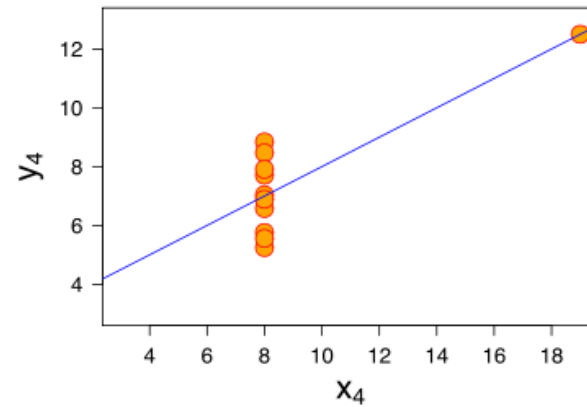
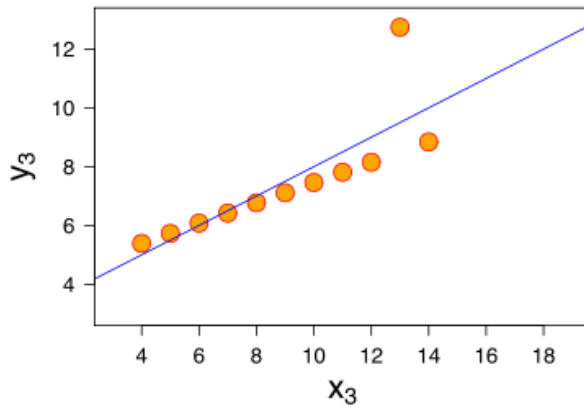
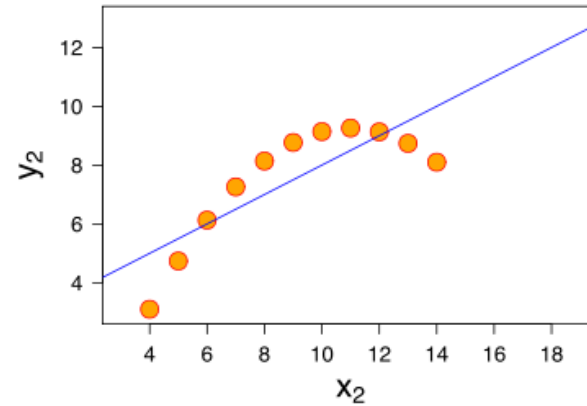
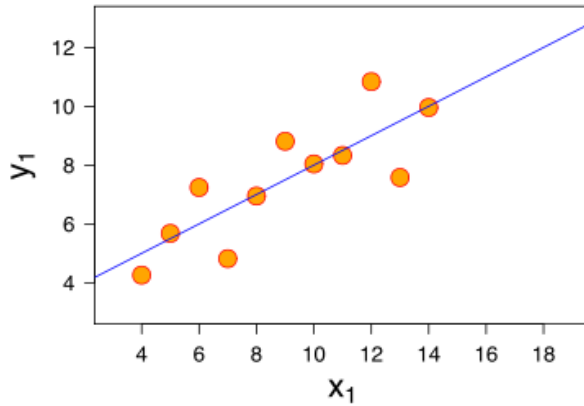






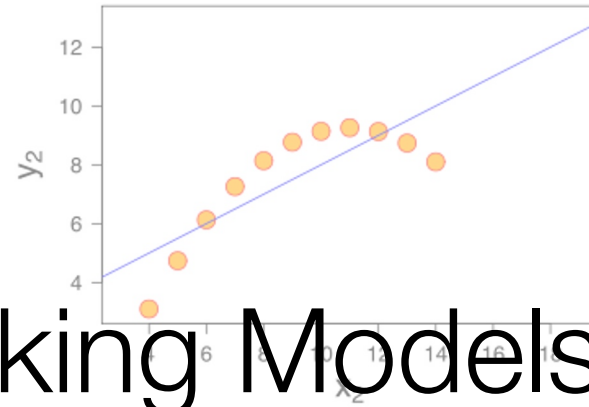
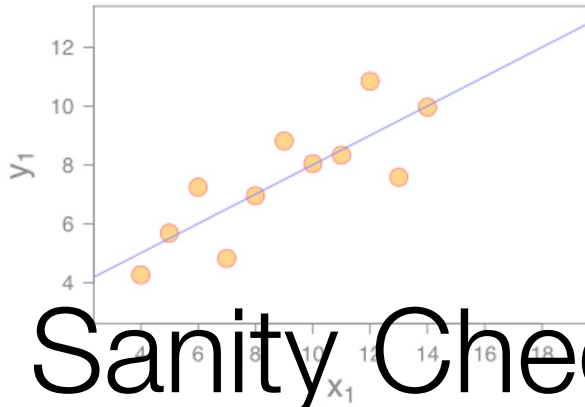


# Anscombe's Quartet

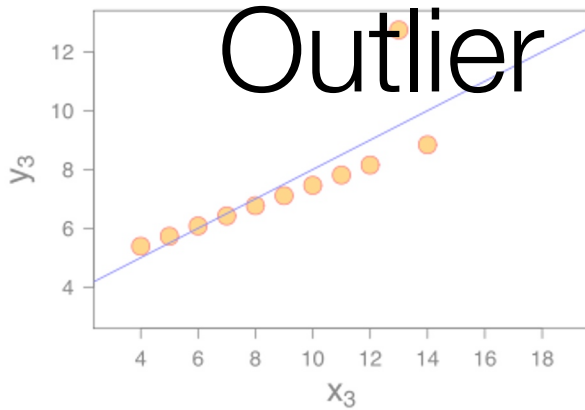


[https://en.wikipedia.org/wiki/Anscombe%27s\\_quartet](https://en.wikipedia.org/wiki/Anscombe%27s_quartet)

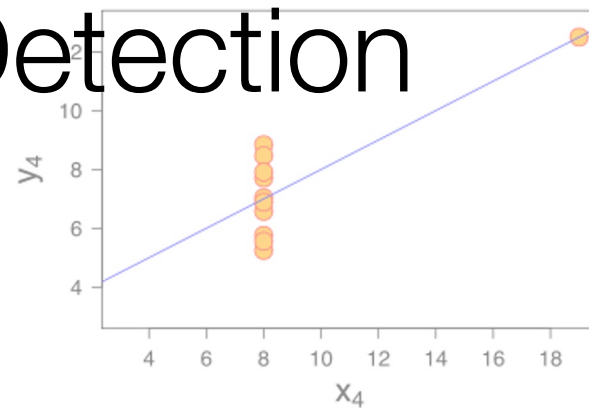
# Anscombe's Quartet



Sanity Checking Models



Outlier Detection





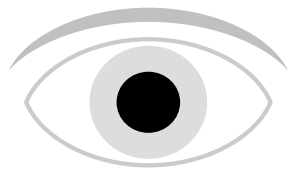
Data visualization leverages  
human perception

Name the five senses.

Sense	Bandwidth (bits/sec)
Sight	10,000,000
Touch	1,000,000
Hearing	100,000
Smell	100,000
Taste	1,000

# A (Simple) Model of Human Visual Perception

# A (Simple) Model of Human Perception



## **Stage 1**

→ Parallel detection of  
basic features into  
an iconic store



## **Stage 2**

Serial processing of  
object identification and  
spatial layout

# Stage 1: Pre-Attentive Processing

Rapid

Parallel

Automatic

(Fleeting = lasting for a short time)

# Stage 2: Serial Processing

Relatively Slow  
(Incorporates Memory)  
Manual

# Stage 1: Pre-Attentive Processing

The eye moves every 200ms  
(so this processing occurs every  
200ms-250ms)



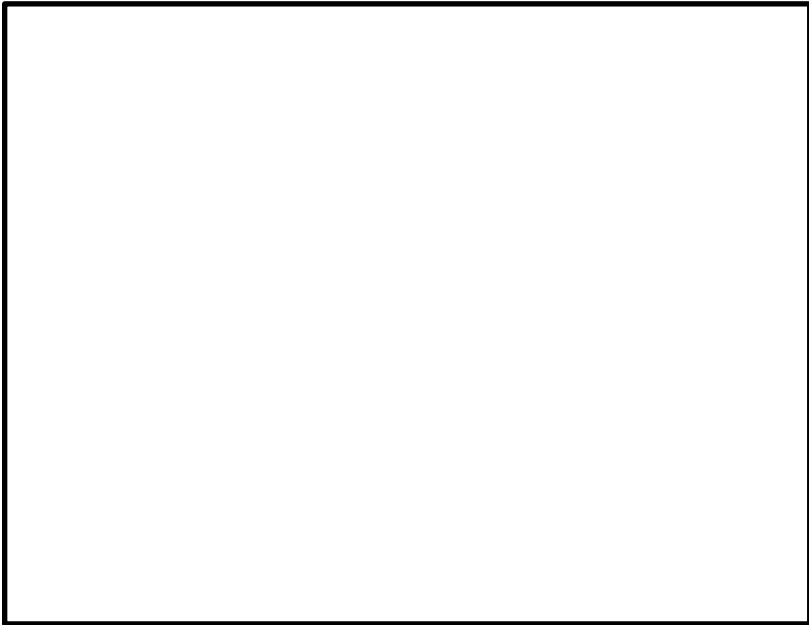
# Example

1281768756138976546984506985604982826762  
9809858458224509856458945098450980943585  
9091030209905959595772564675050678904567  
8845789809821677654876364908560912949686

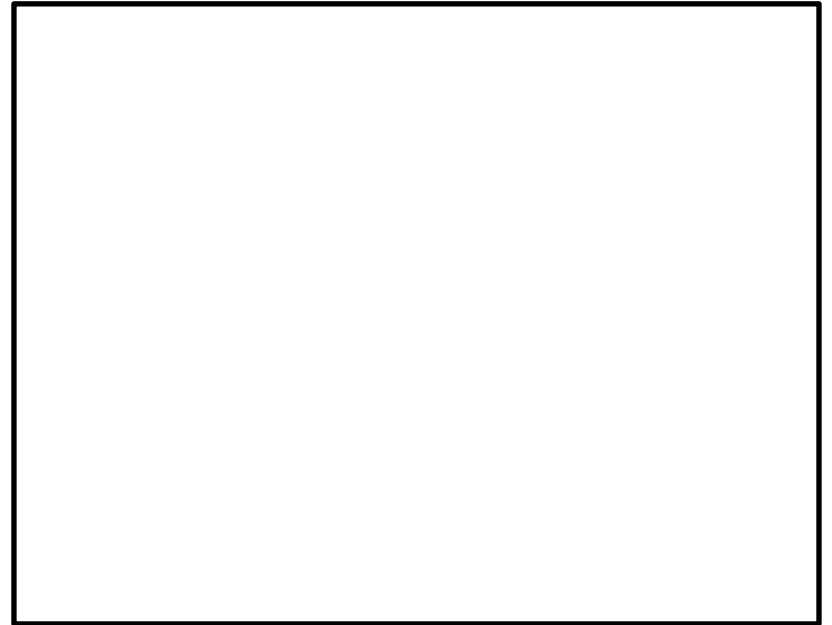
# Example

12817687561**3**8976546984506985604982826762  
980985845822450985645894509845098094**3**585  
90910**3**0209905959595772564675050678904567  
8845789809821677654876**3**64908560912949686

A few more examples from  
Prof. Chris Healy at NC State



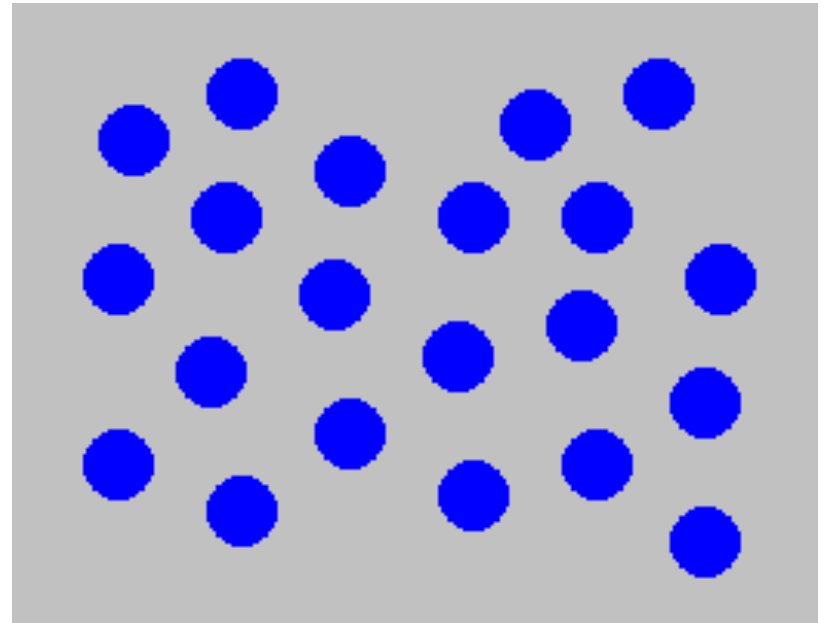
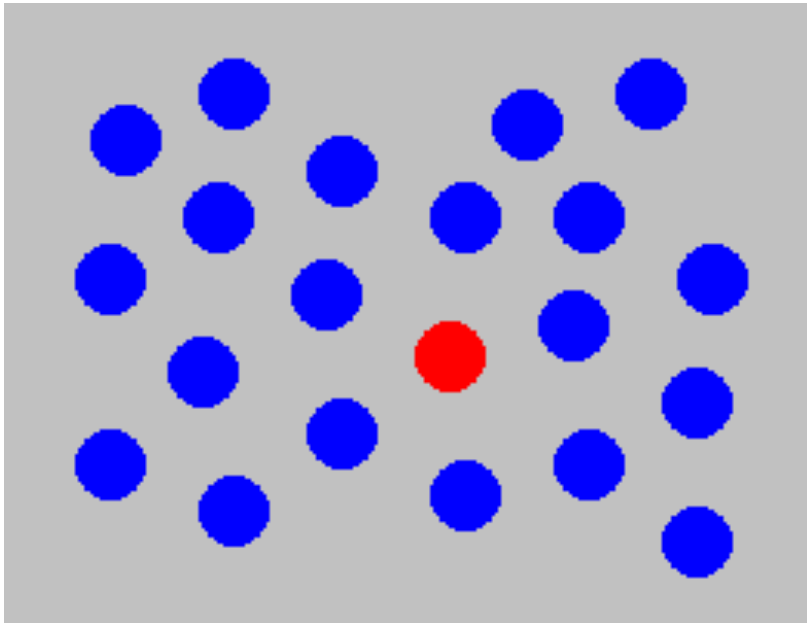
Left Side



Right Side

Raise your hand if a **RED DOT**  
is present...

(On the left or on the right?)

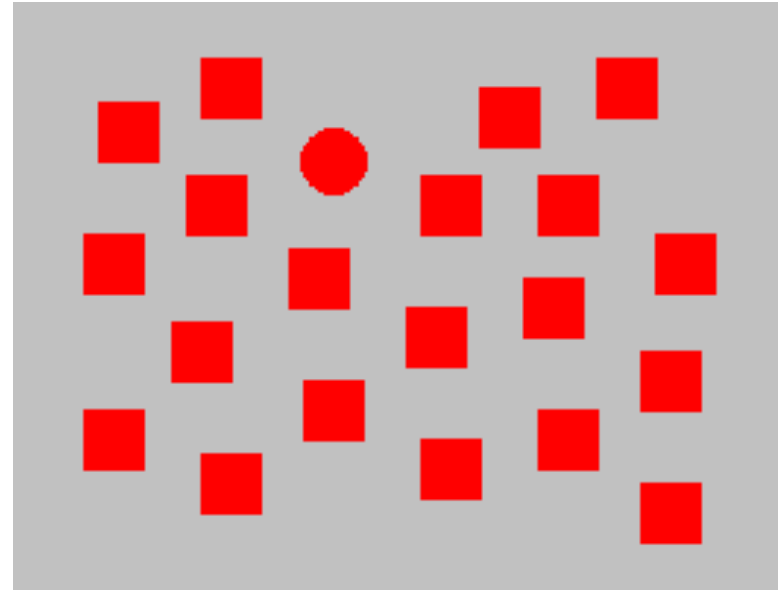
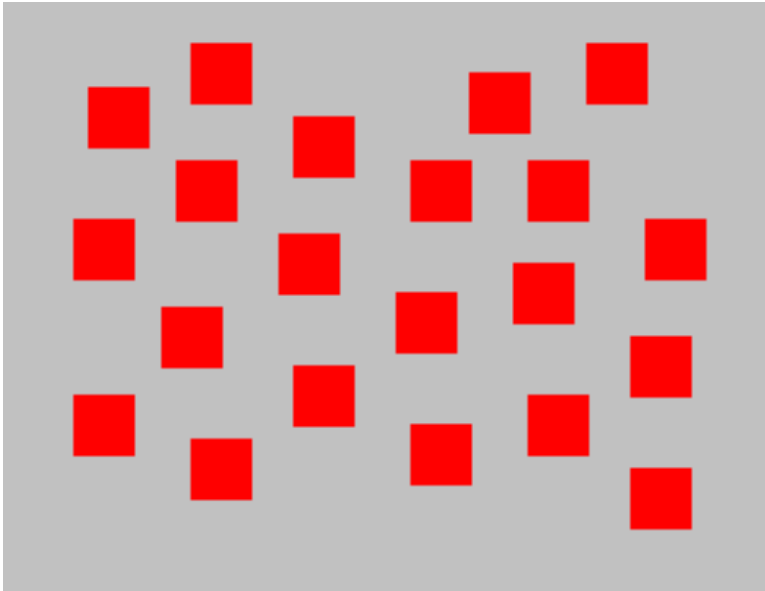




Color (hue) is pre-attentively processed.



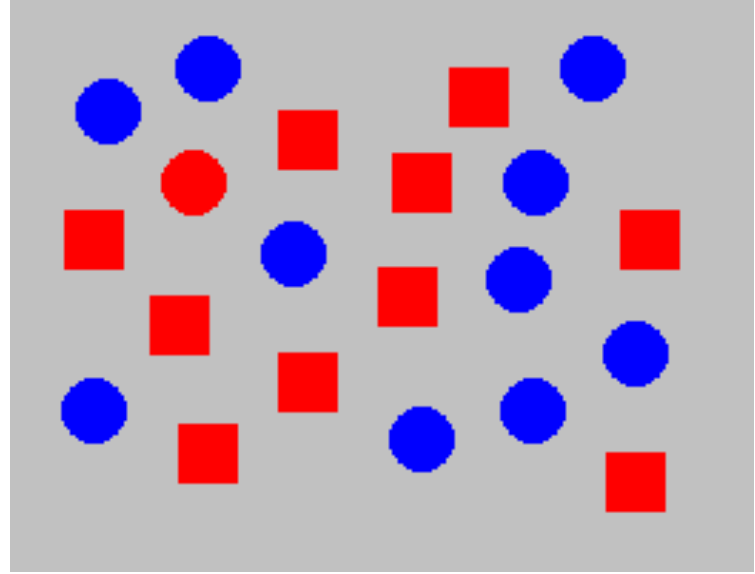
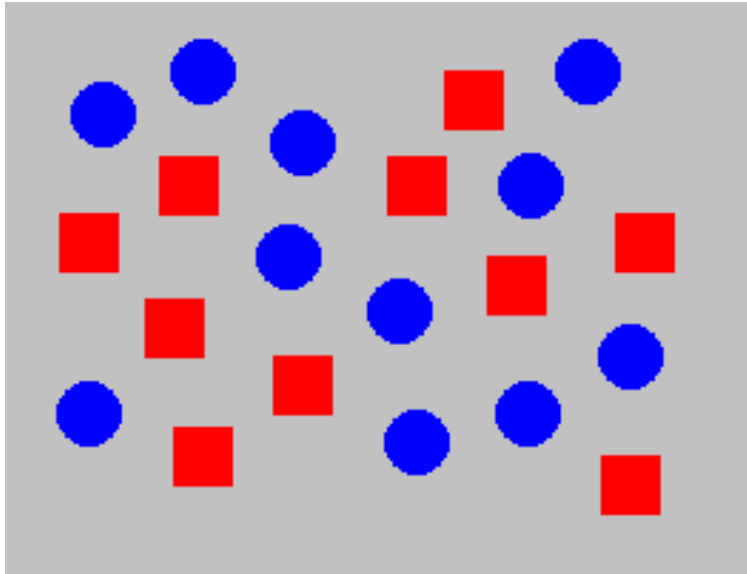
Raise your hand if a RED DOT  
is present...





Shape is pre-attentively  
processed.

Determine if a RED DOT is  
present...



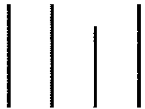

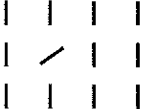
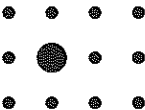
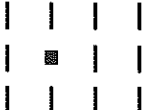
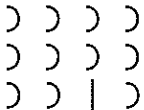
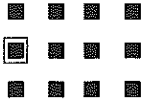
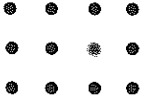
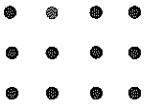
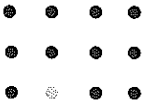





Hue and shape together are  
NOT pre-attentively processed.



# Pre-Attentive Processing

- length
- width
- size
- curvature
- number
- terminators
- intersection
- closure
- hue
- lightness
- flicker
- direction of motion
- binocular lustre
- stereoscopic depth
- 3-D depth cues
- lighting direction

Group	Attribute		
Form	<p data-bbox="575 139 658 165">Length</p> 	<p data-bbox="973 139 1056 165">Width</p> 	<p data-bbox="1340 139 1470 165">Orientation</p> 
	<p data-bbox="591 358 643 384">Size</p> 	<p data-bbox="973 358 1056 384">Shape</p> 	<p data-bbox="1340 358 1470 384">Curvature</p> 
	<p data-bbox="562 572 672 598">Enclosure</p> 	<p data-bbox="983 572 1047 598">Blur</p> 	
Color	<p data-bbox="591 789 643 815">Hue</p> 		<p data-bbox="954 789 1064 815">Intensity</p> 
Spatial Position	<p data-bbox="542 1003 691 1029">2-D Position</p> 		<p data-bbox="915 1003 1103 1029">Spatial Grouping</p> 
Motion	<p data-bbox="562 1218 672 1243">Direction</p> 		

Stephen Few  
 “Now You See It”  
 pg. 39

Pre-Attentive → Cognitive

# Gestalt Psychology

Berlin, Early 1900s

# Gestalt Psychology

Goal was to understand  
pattern perception

Gestalt (German) = “seeing the whole picture all at once”  
instead of a collection of parts

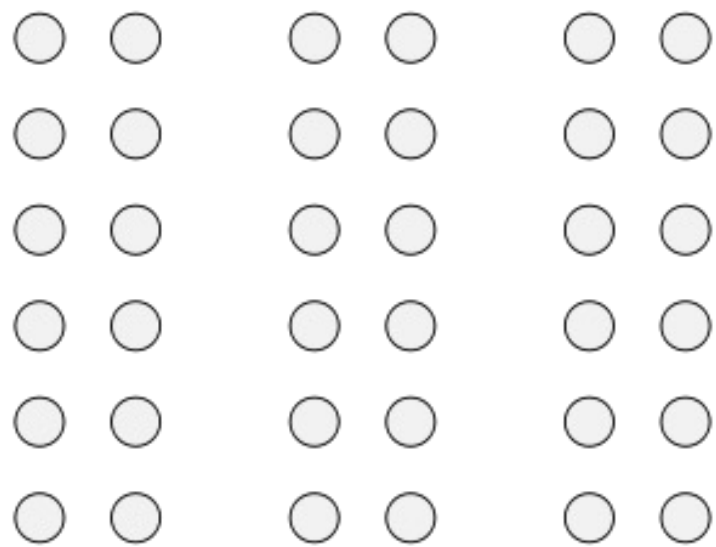
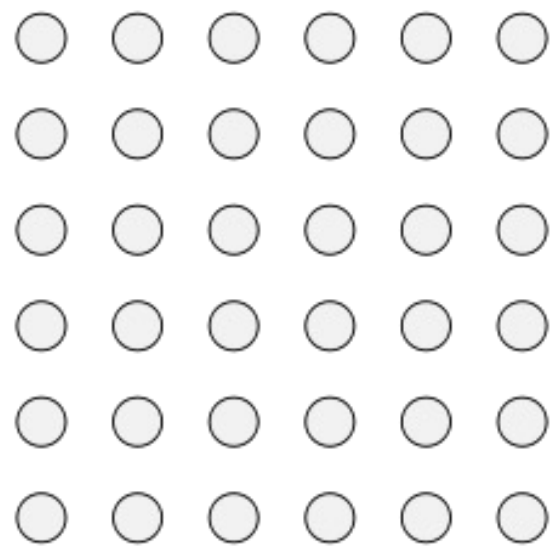
Identified 8 “Laws of Grouping”

<http://study.com/academy/lesson/gestalt-psychology-definition-principles-quiz.html>

# Gestalt Psychology

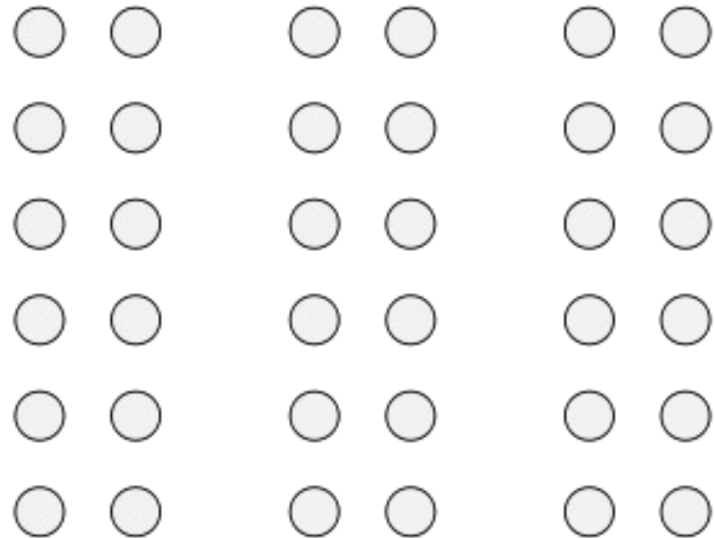
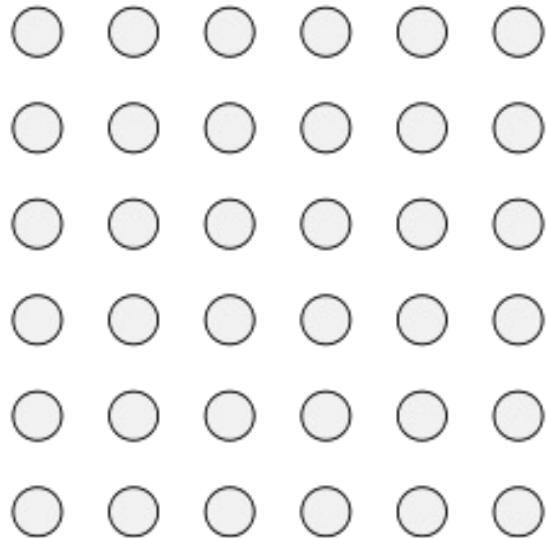
1. Proximity
2. Similarity
3. Closure
4. Symmetry
5. Common Fate
6. Continuity
7. Good Gestalt
8. Past Experience

How many groups are there?

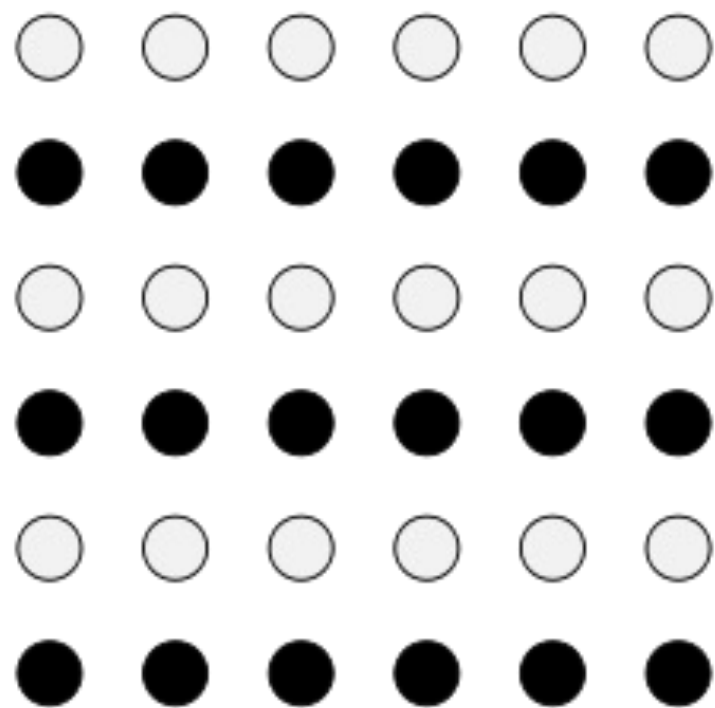




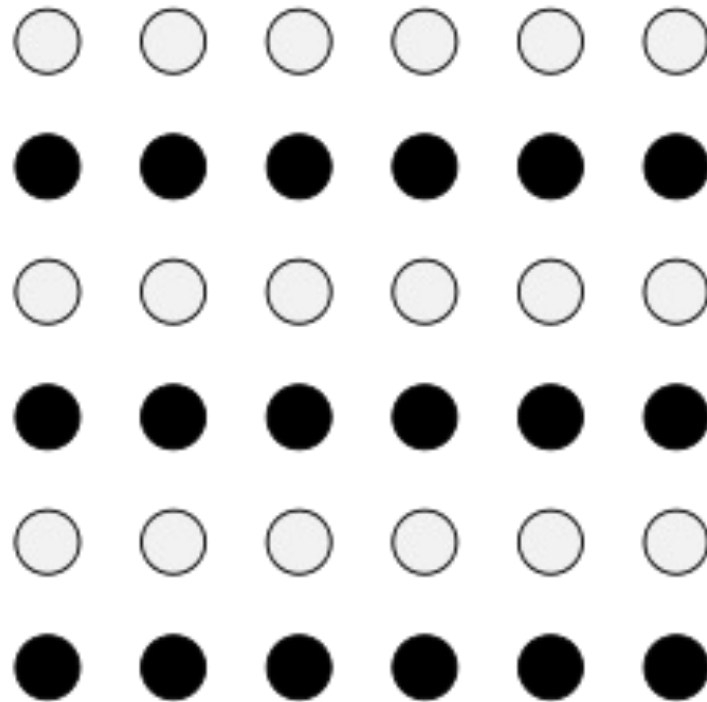
# Proximity



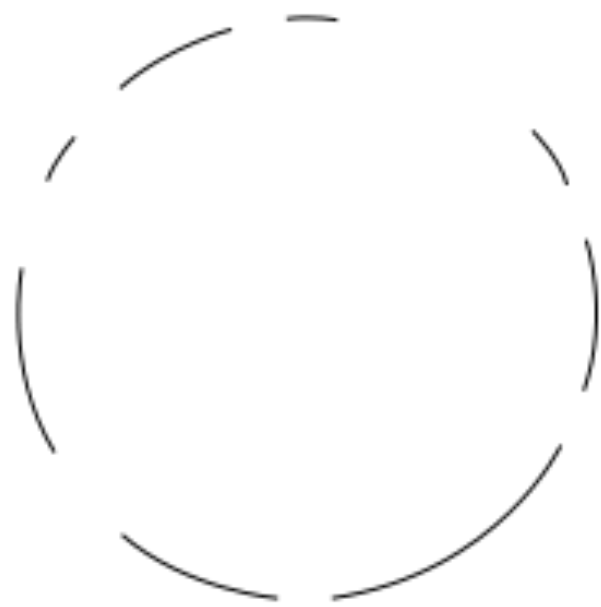
How many groups are there?



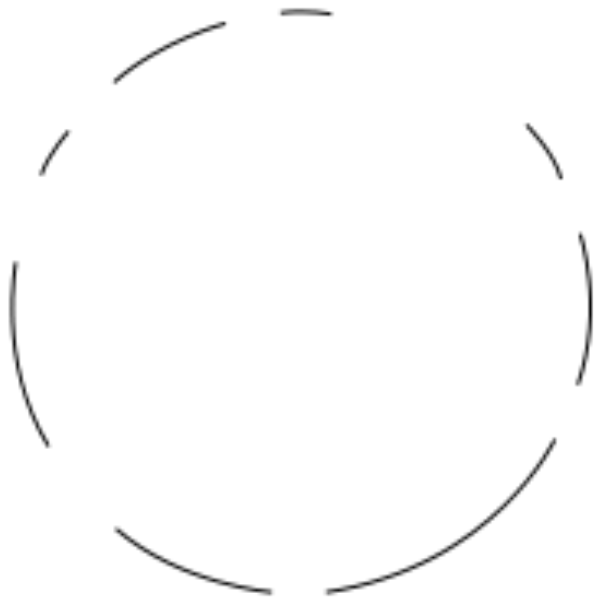
# Similarity



How many shapes are there?



# Closure



How many items are there?



( )

{ }

[ ]

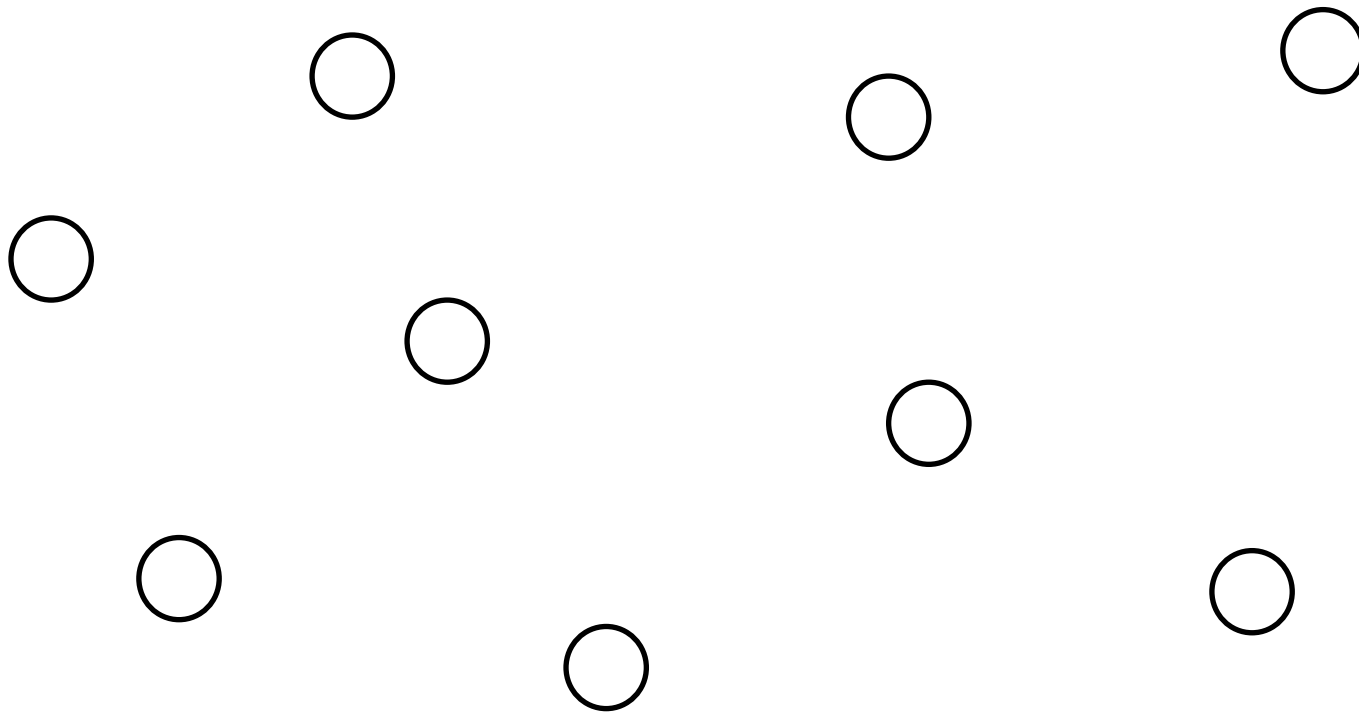
# Symmetry

( )

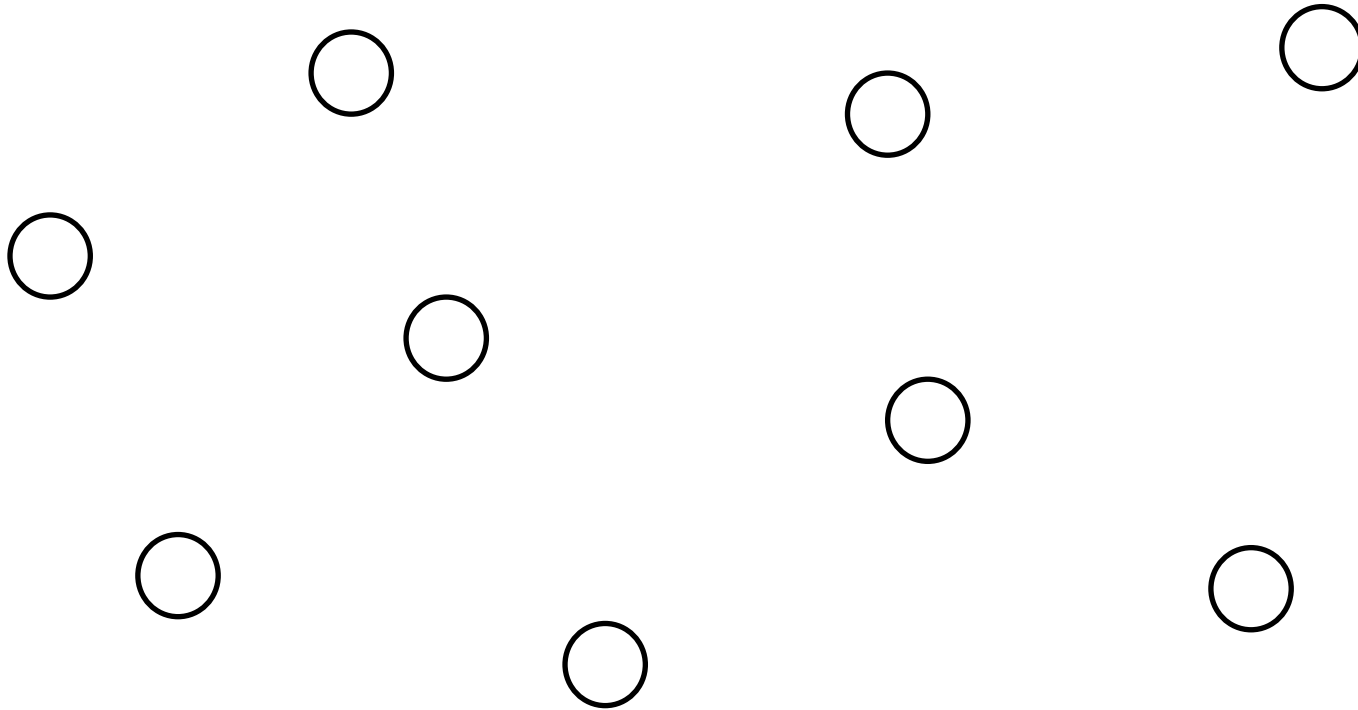
{ }

[ ]

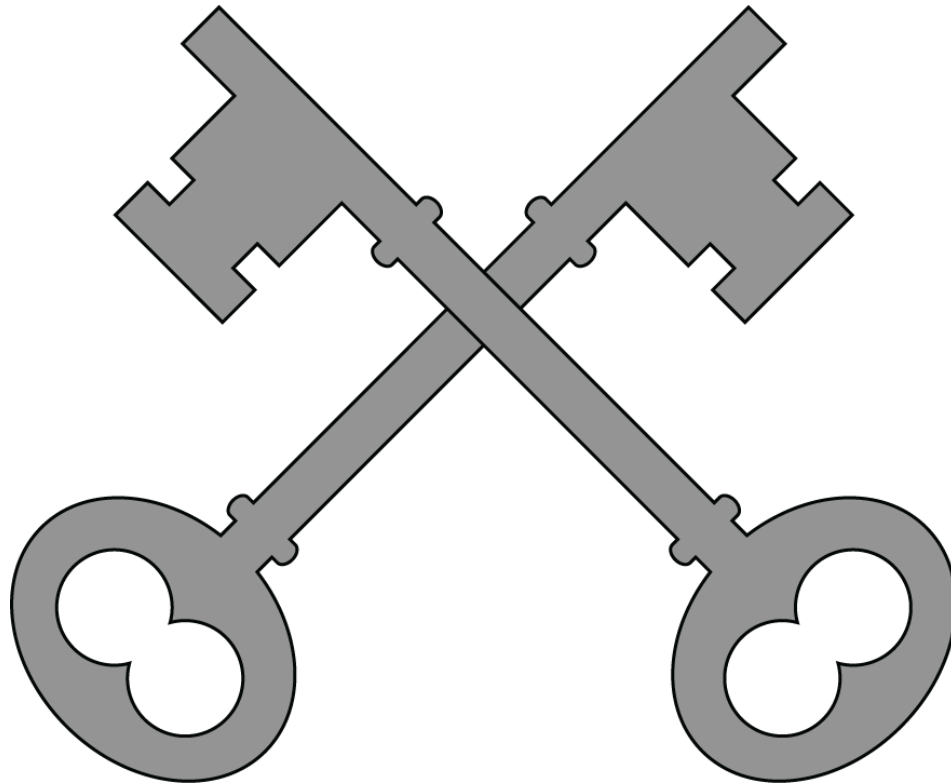
How many sets are there?



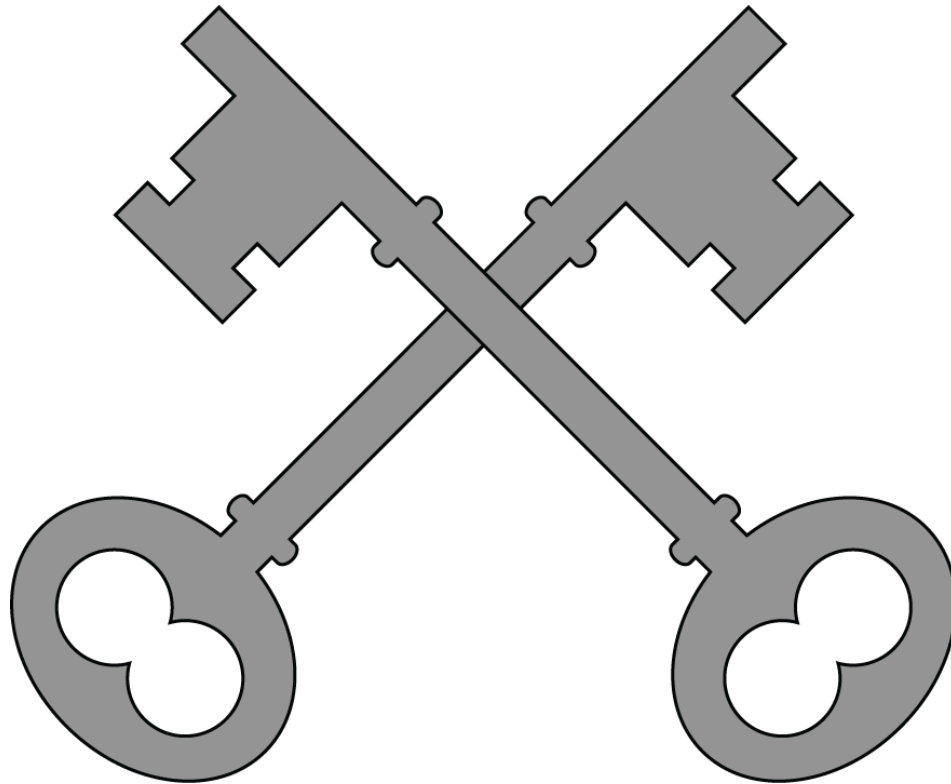
# Common Fate



How many objects are there?

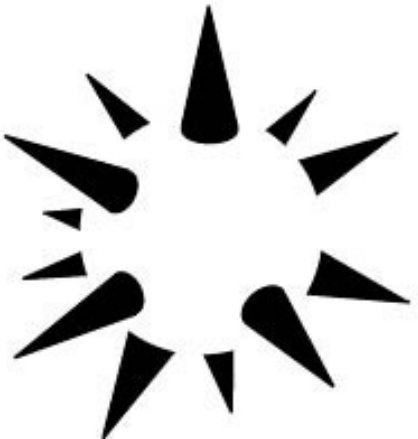
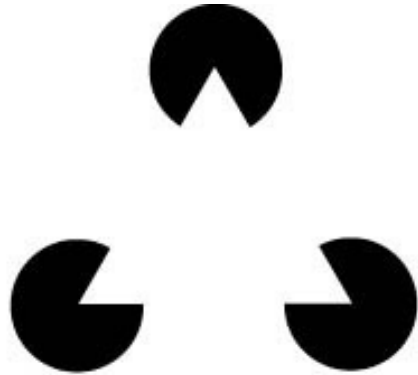


# Continuity

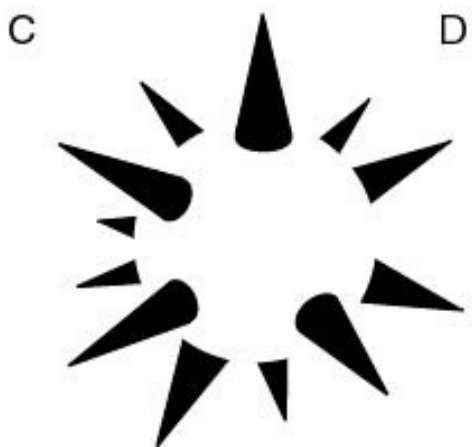




How many objects are there?



# Good Gestalt



What is this word?

**FLIGHT**

Past Experience

**FLIGHT**

# Pre-Attentive Processing

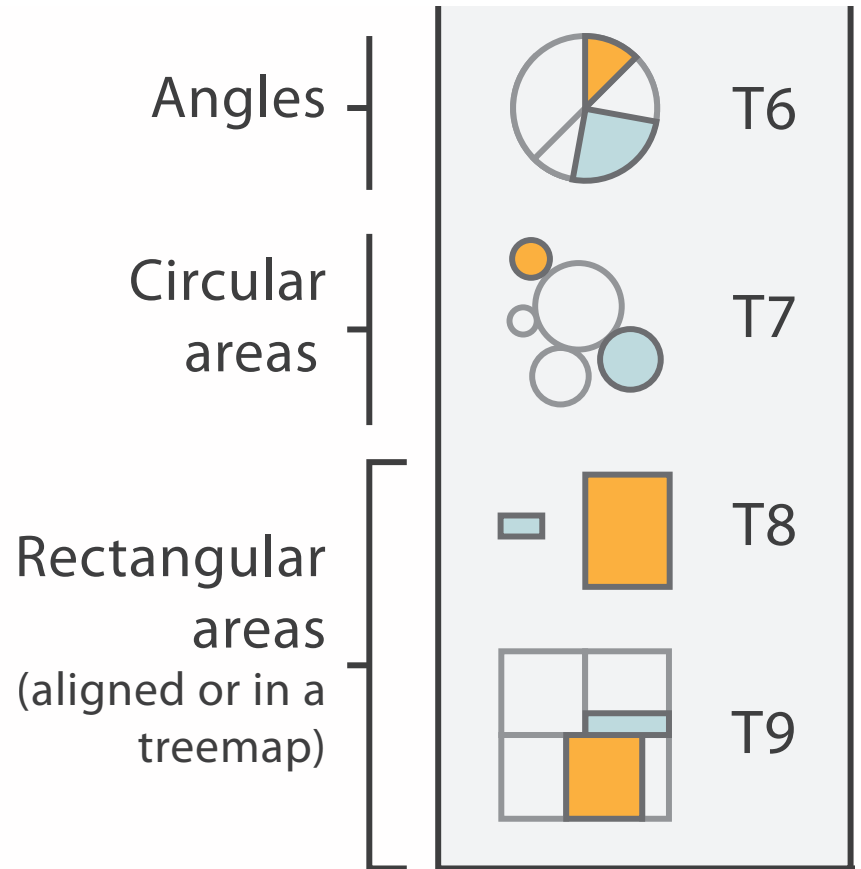
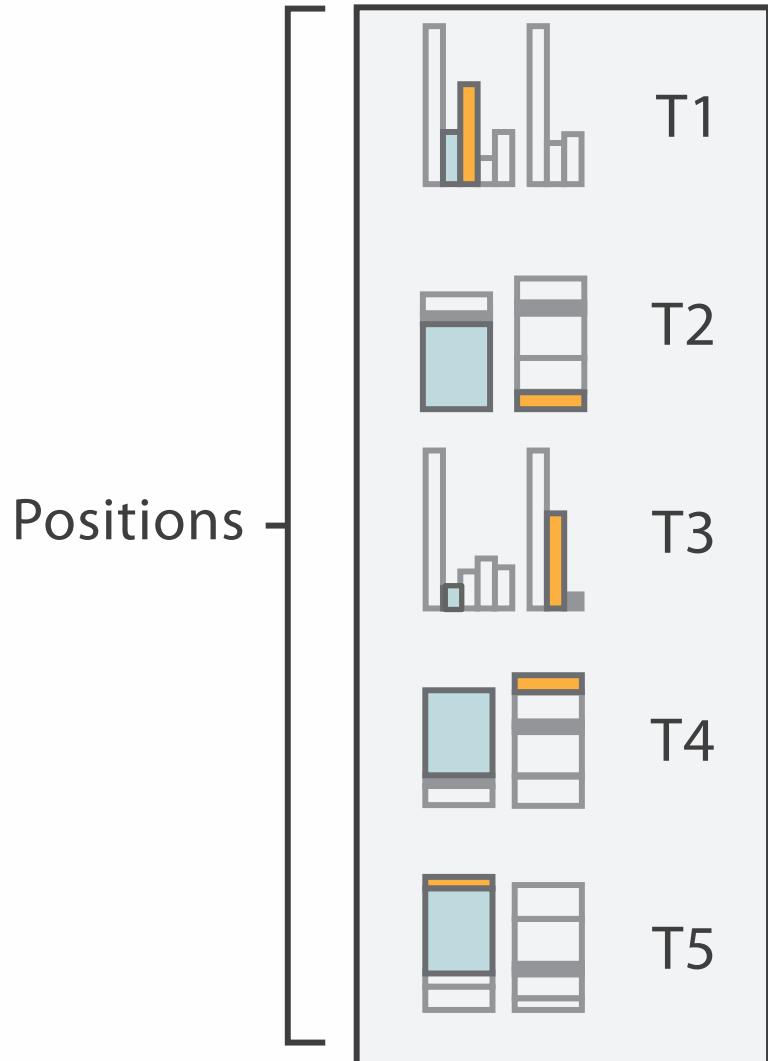
## Gestalt Laws

Detect Quickly

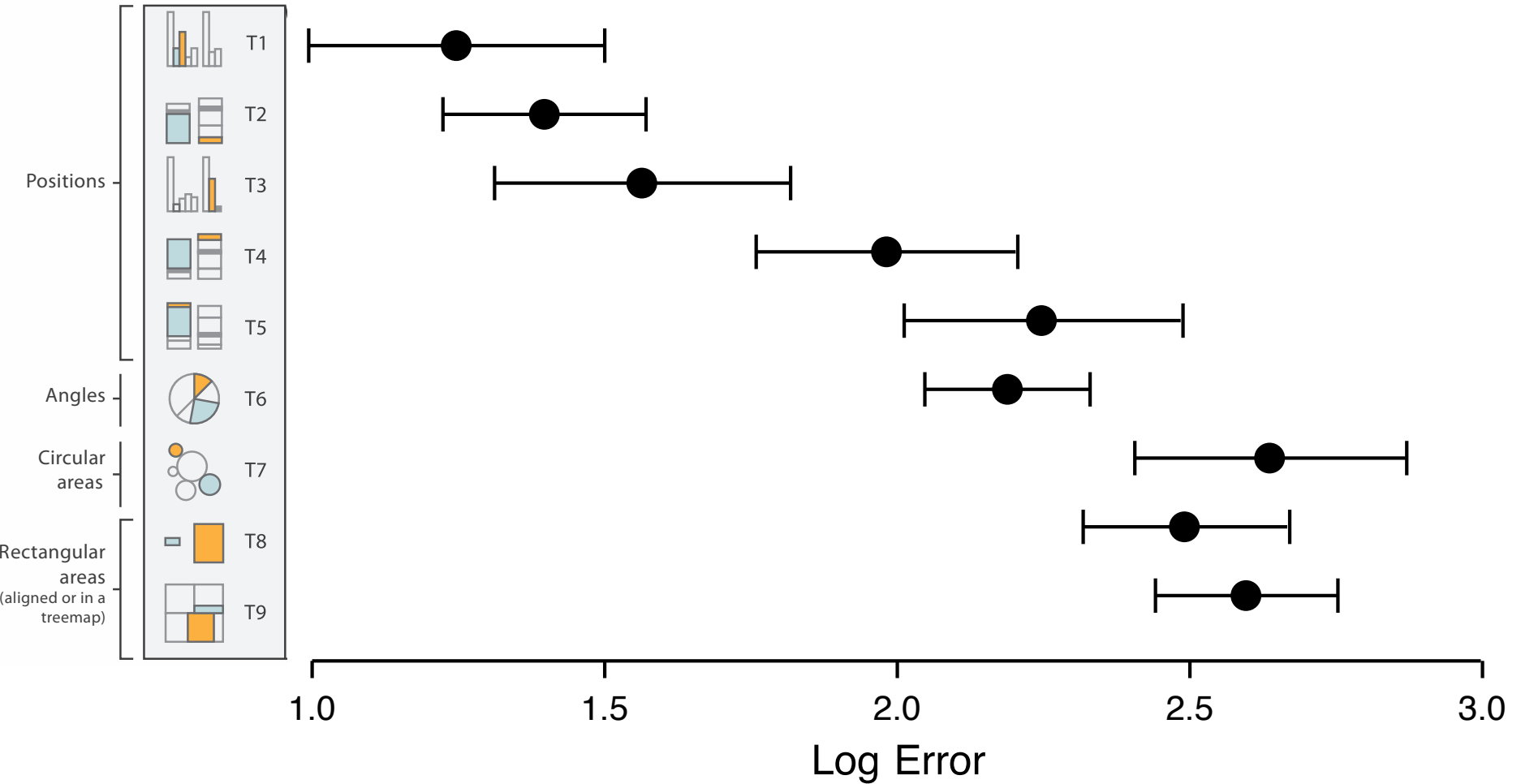


Detect quickly **does NOT mean**  
detect accurately

Ideally you want both.



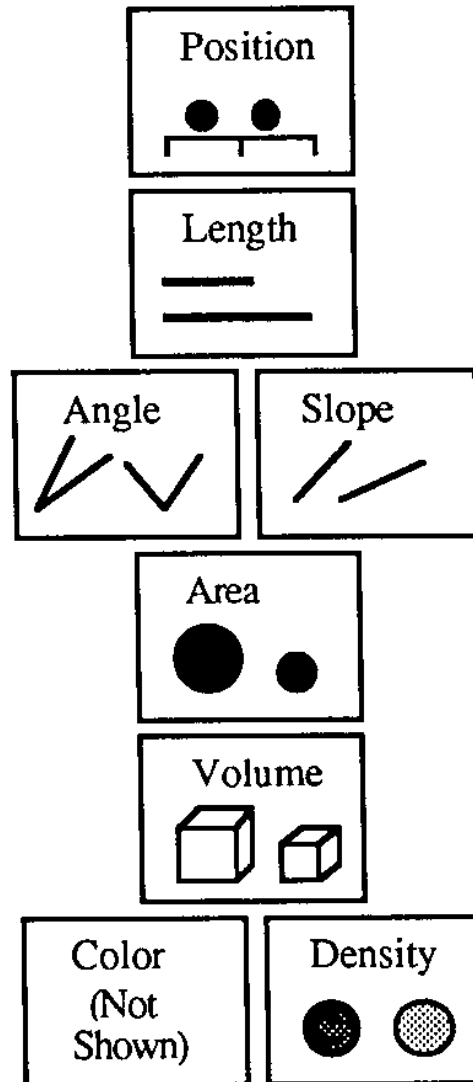
# Crowdsourced Results




More accurate




Less accurate

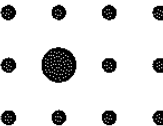



Precision of Quantitative Perception	Attribute	Example	Description
--------------------------------------	-----------	---------	-------------

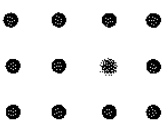
Very precise	Length		Longer = greater
--------------	--------	--	------------------

	2-D Position		Higher or farther to the right = greater
--	--------------	--	--

Not very precise	Width		Wider = greater
------------------	-------	--	-----------------

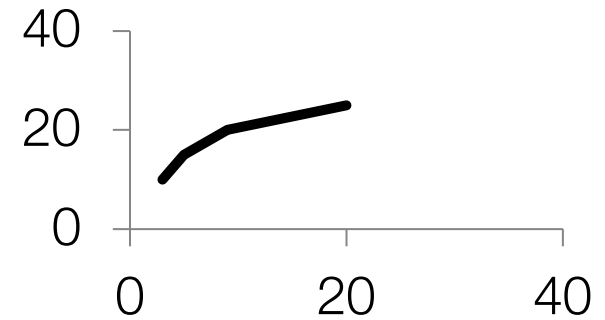
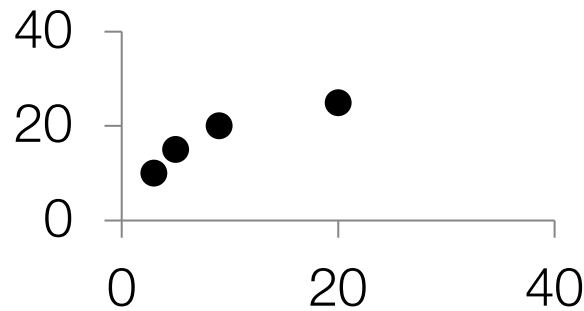
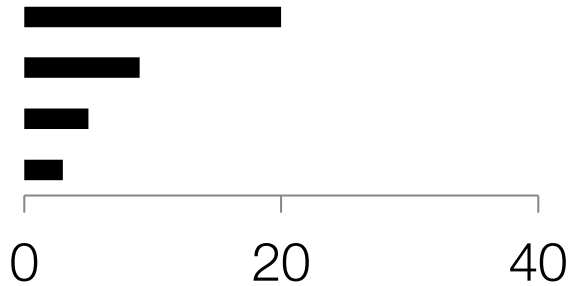
	Size		Bigger = greater
--	------	--	------------------

	Intensity		Darker = greater
--	-----------	---	------------------

	Blur		Clearer = greater
--	------	--	-------------------

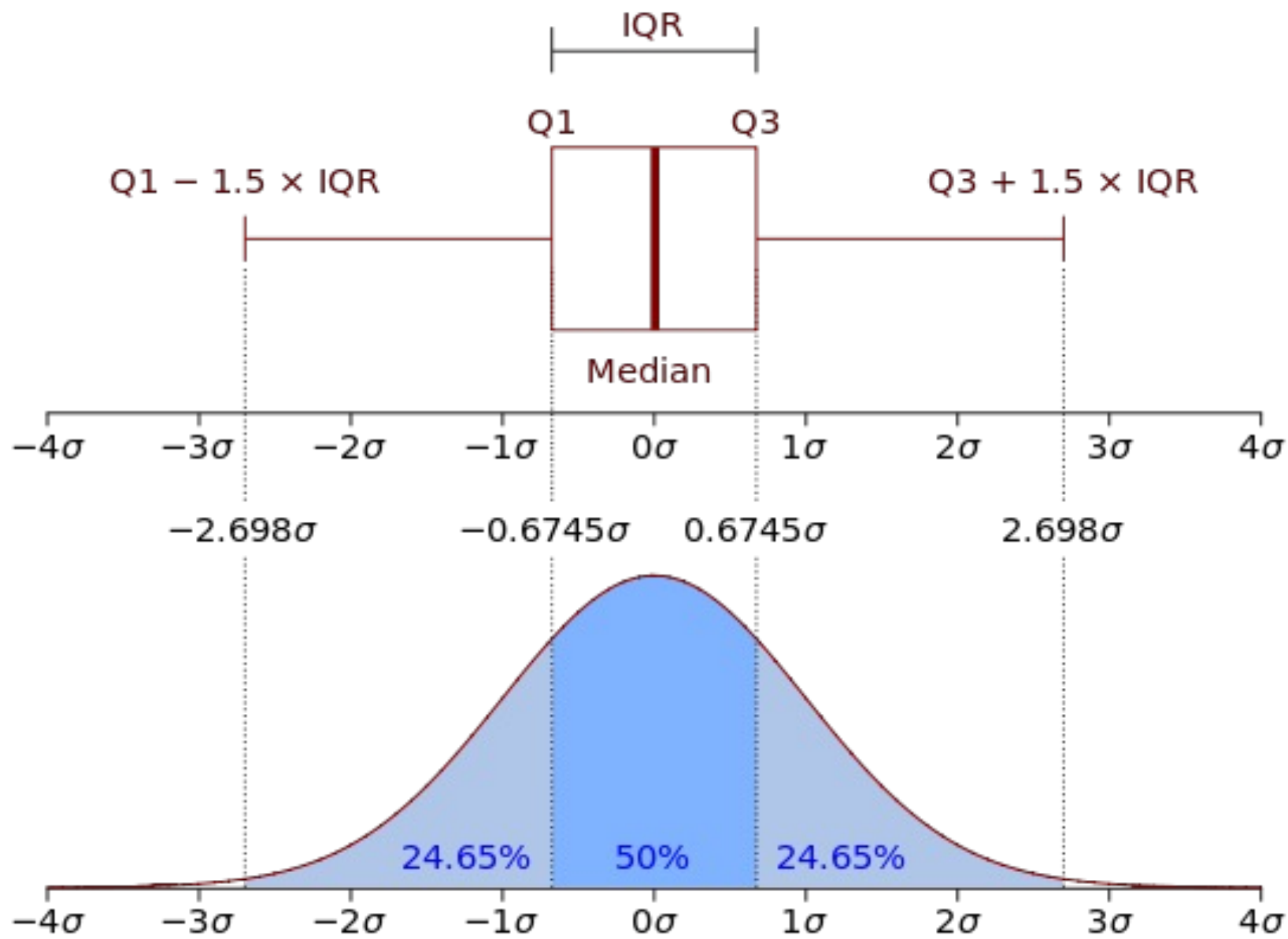
What does this tell us?

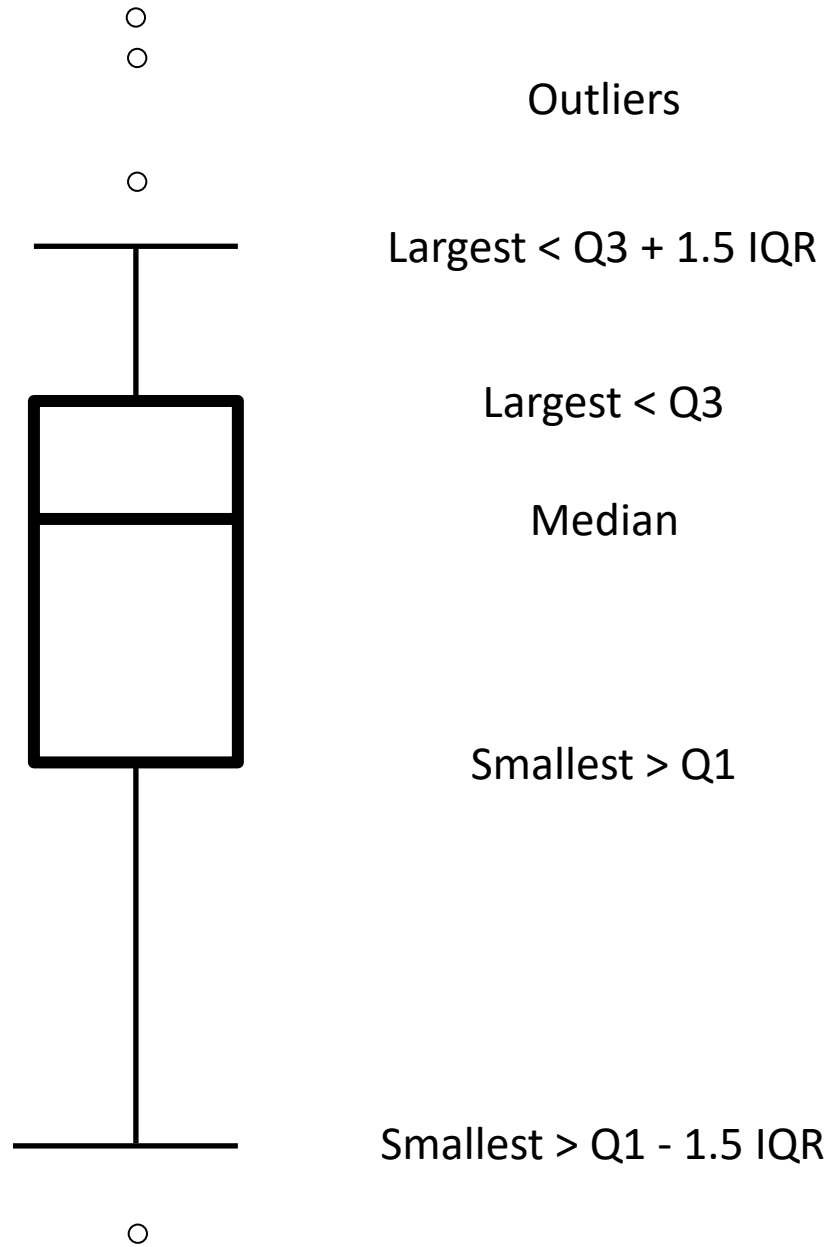
Bar charts, scatterplots, and line charts are *really* effective for quantitative data



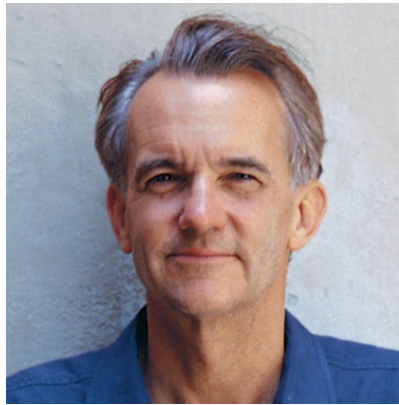
(and for statistical distributions)  
Tukey Box Plots







# Tufte's Chart Principles



Edward Tufte

# Tufte's Chart Principles

**DO NOT LIE!**

# Tufte's Chart Principles

**DO NOT LIE!**

Maximize Data-Ink Ratio

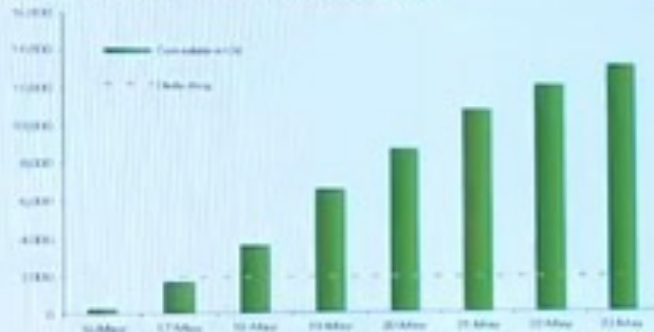
Minimize Chart Junk

## Subsea Oil Collection



- Avg circa 2,000 bbl per day
- Total of 13,500 bbls collected

Cumulative Oil Collected, bbls

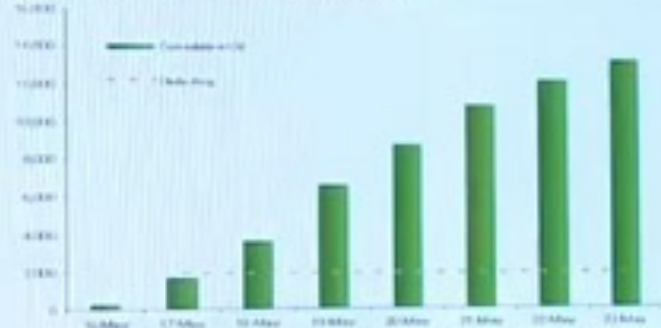


Flow Injection Tube Tool (FITT)

SAMSUNG

“Cumulative”

Cumulative Oil Collected, bbls

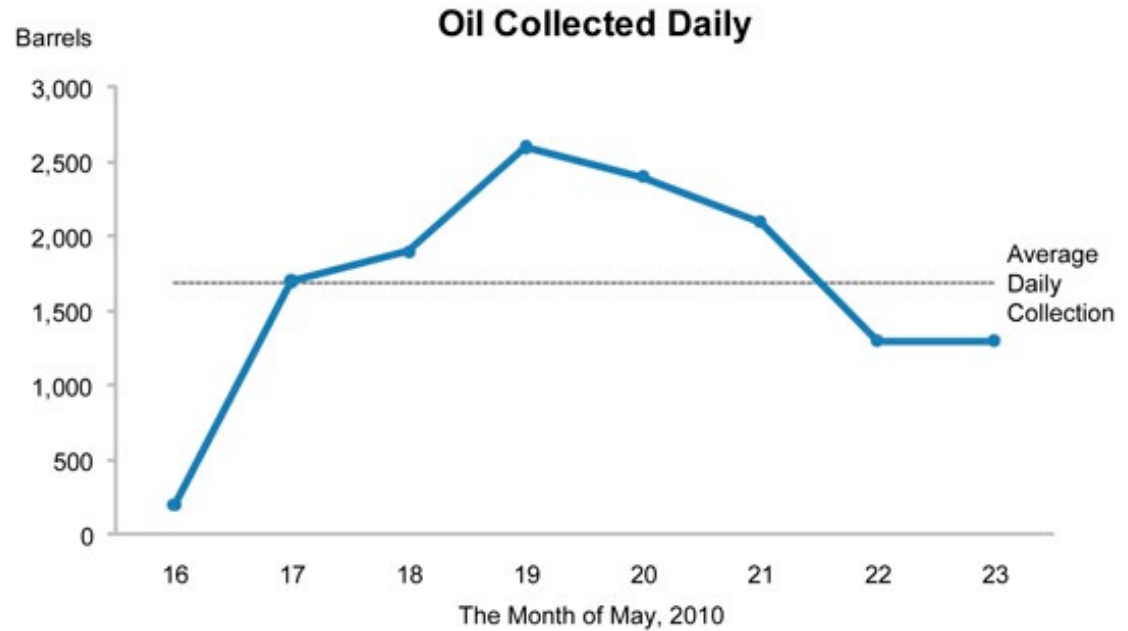
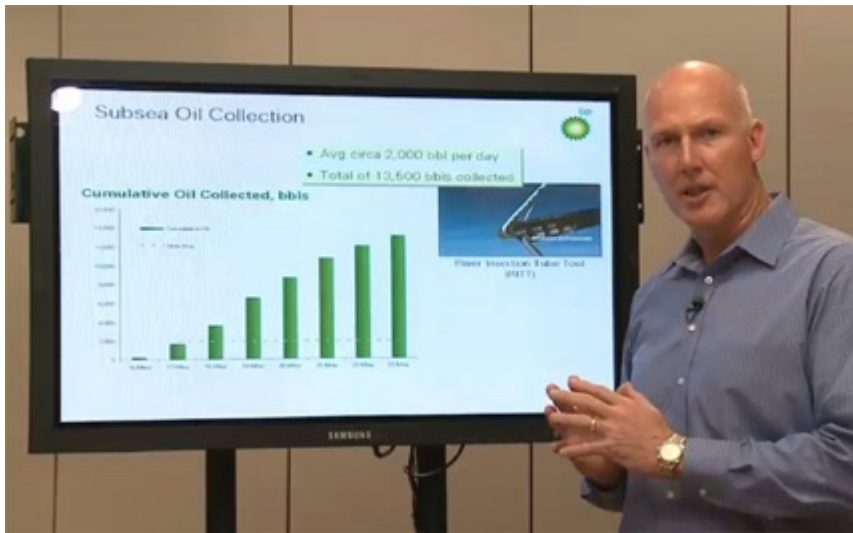


- Avg circa 2,000 bbl per day
- Total of 13,500 bbls collected

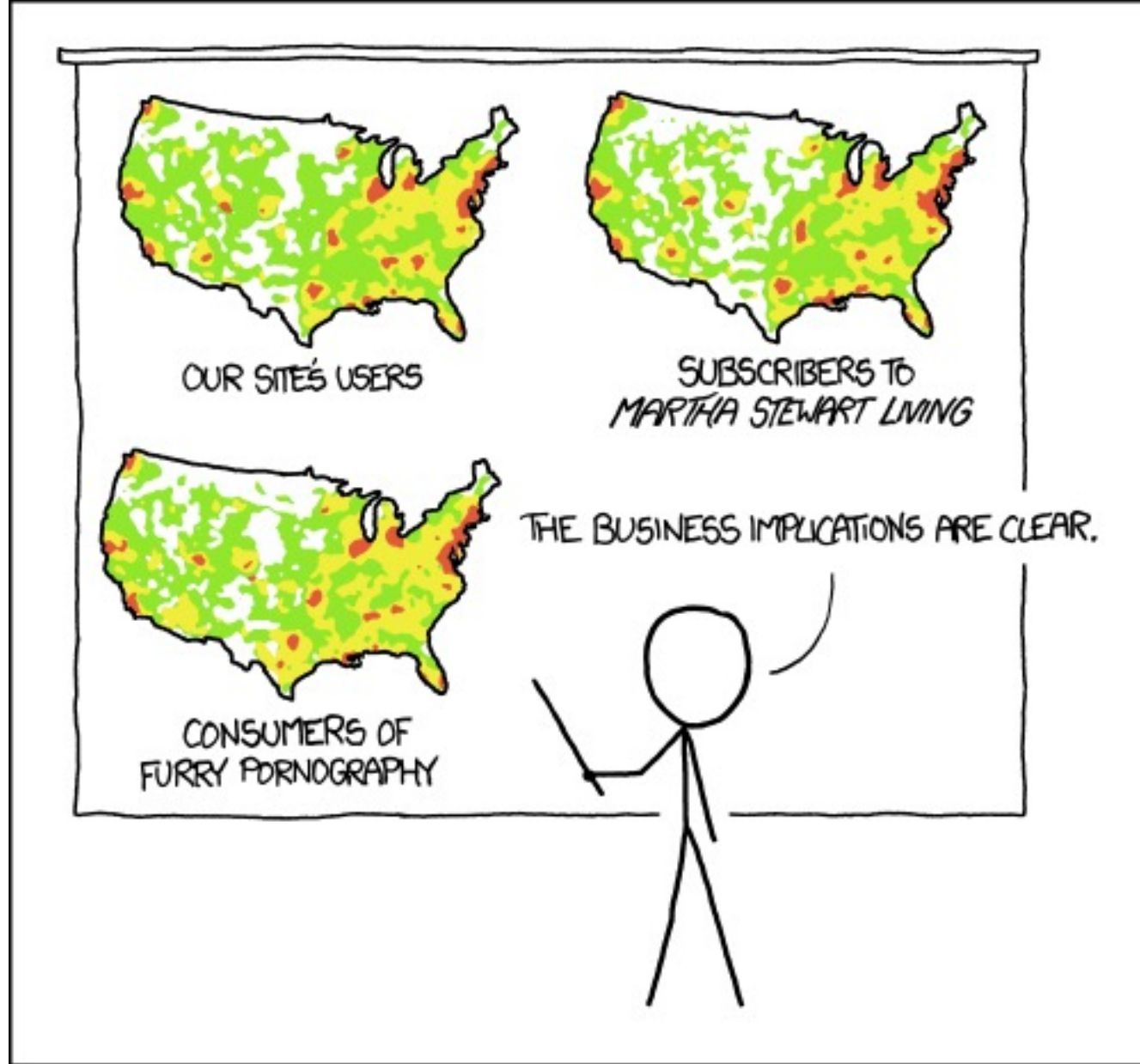


Flow Injection Tube Tool (FITT)









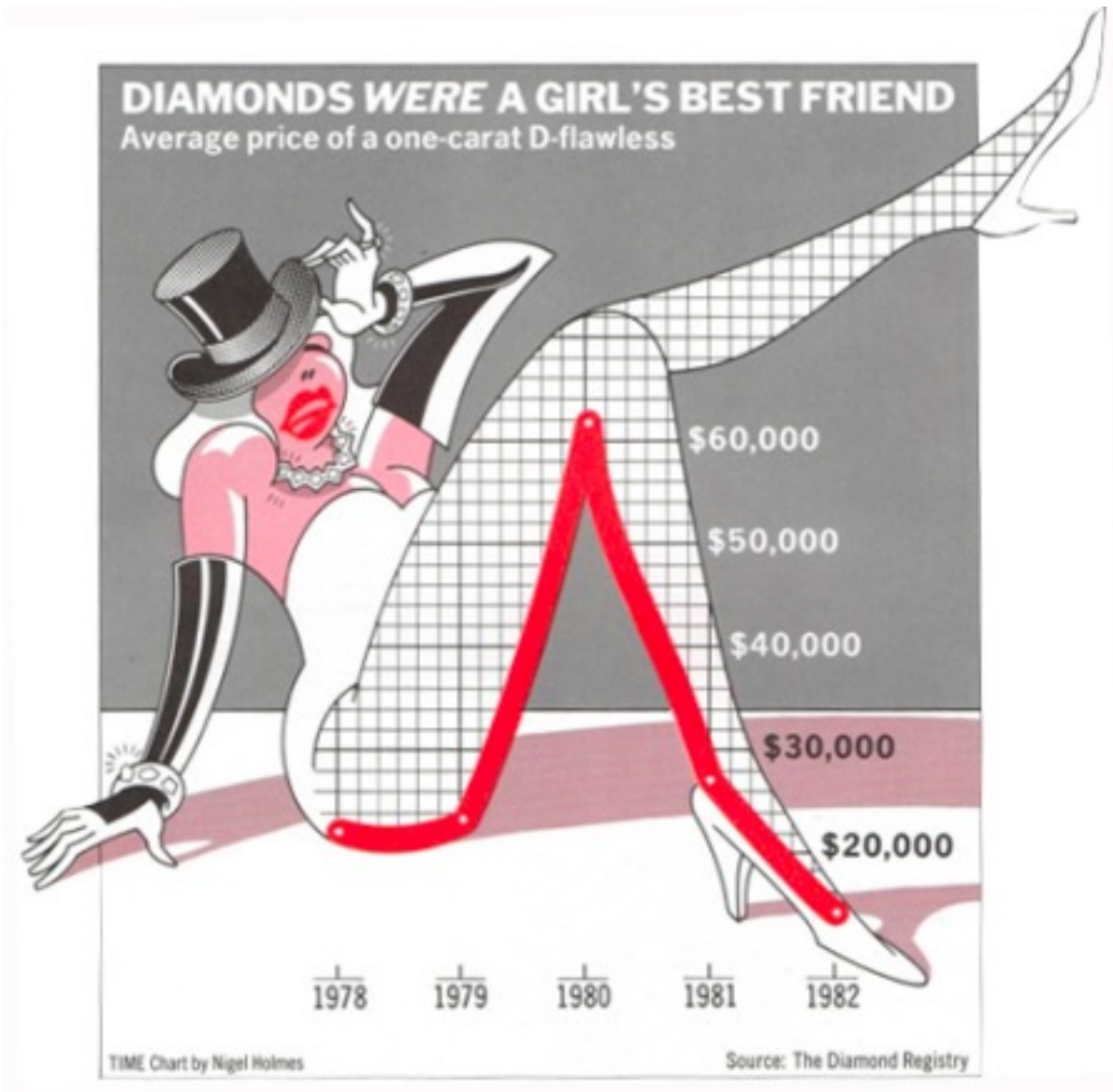
PET PEEVE #208:  
GEOGRAPHIC PROFILE MAPS WHICH ARE  
BASICALLY JUST POPULATION MAPS

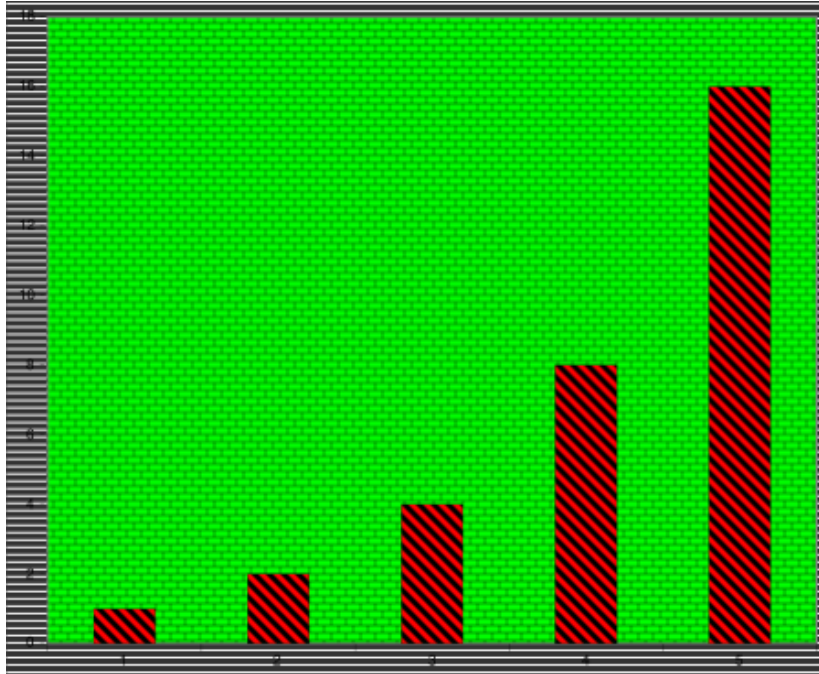
# Tufte's Chart Principles

DO NOT LIE!

Maximize Data-Ink Ratio

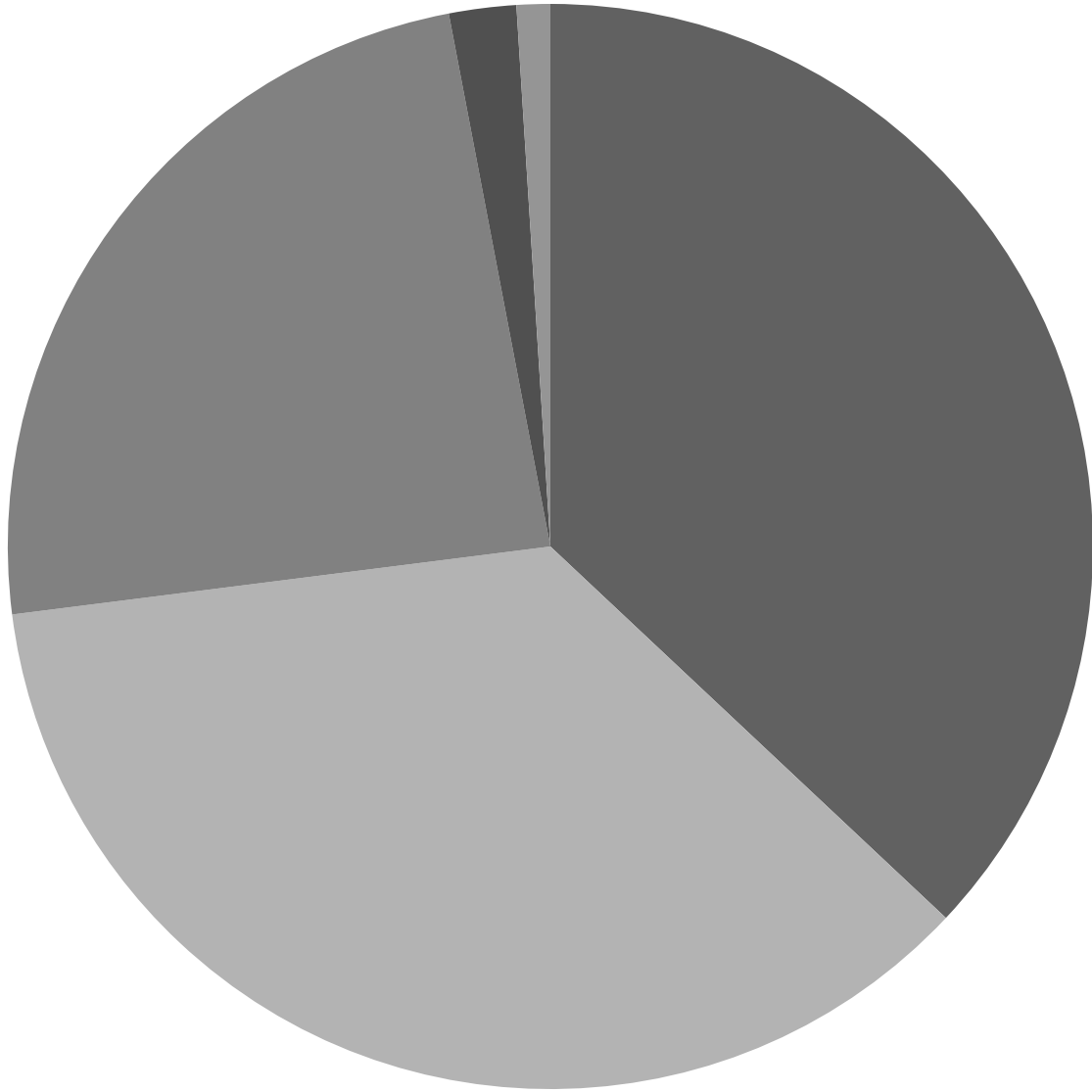
Minimize Chart Junk

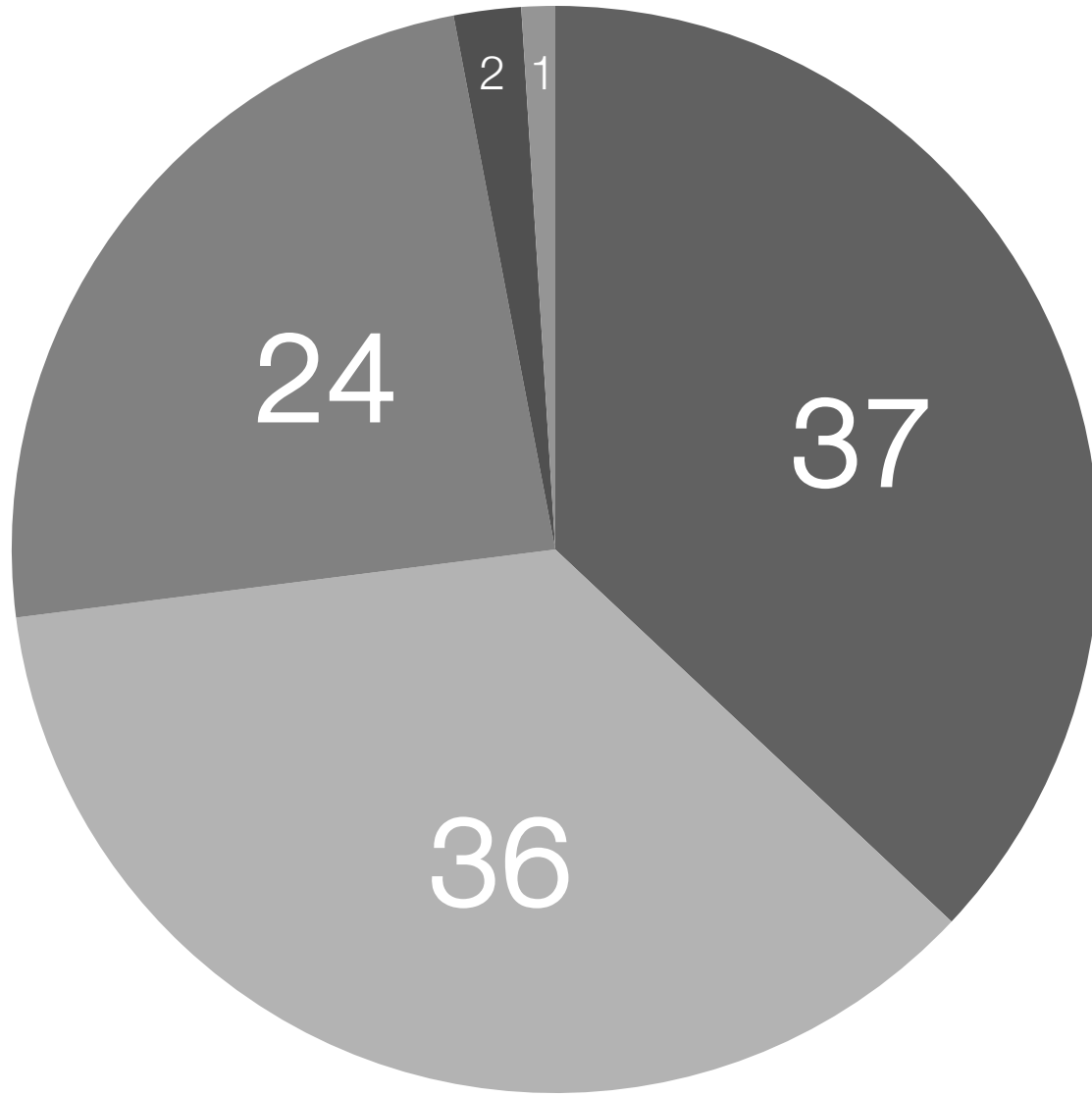




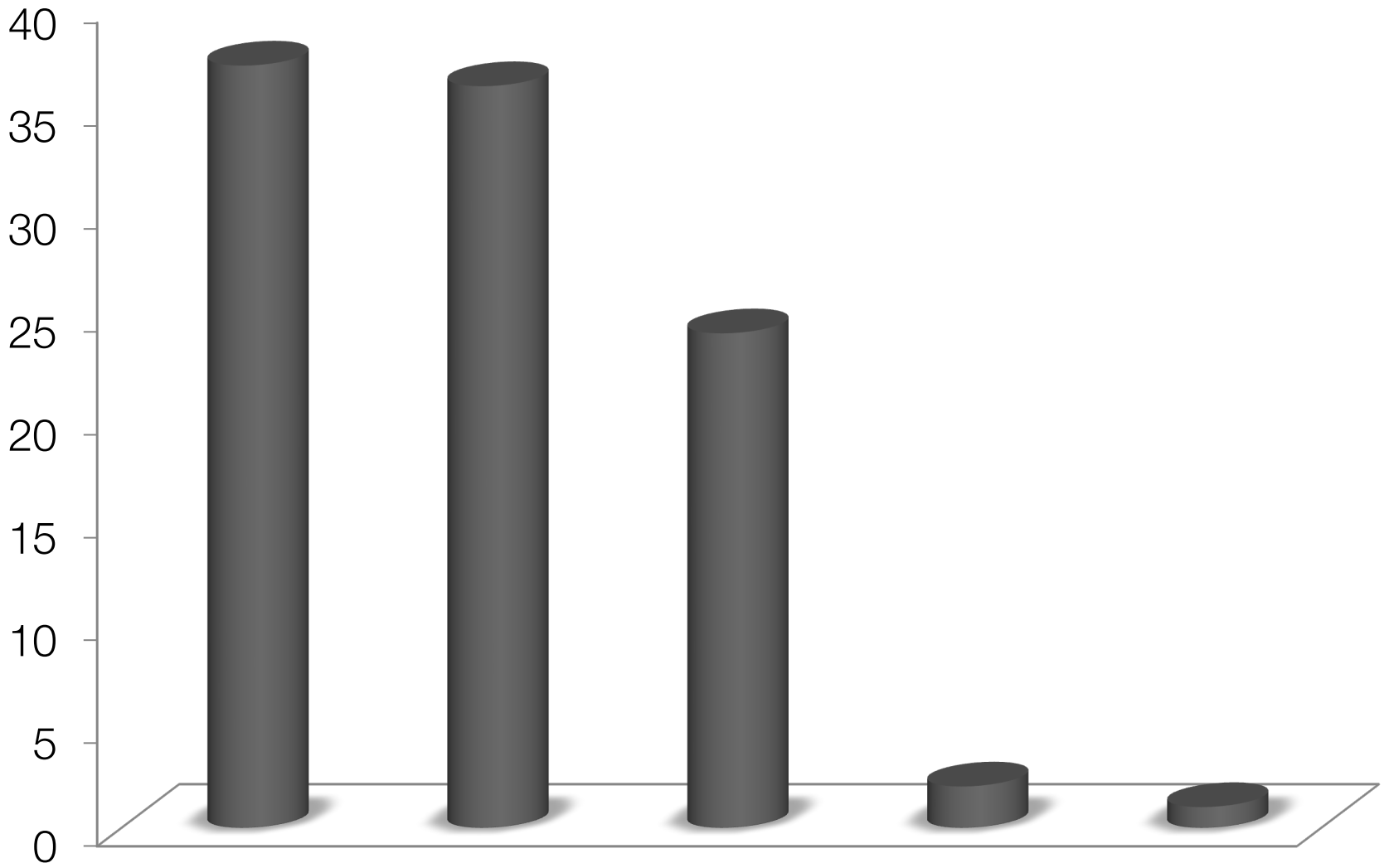
Please...

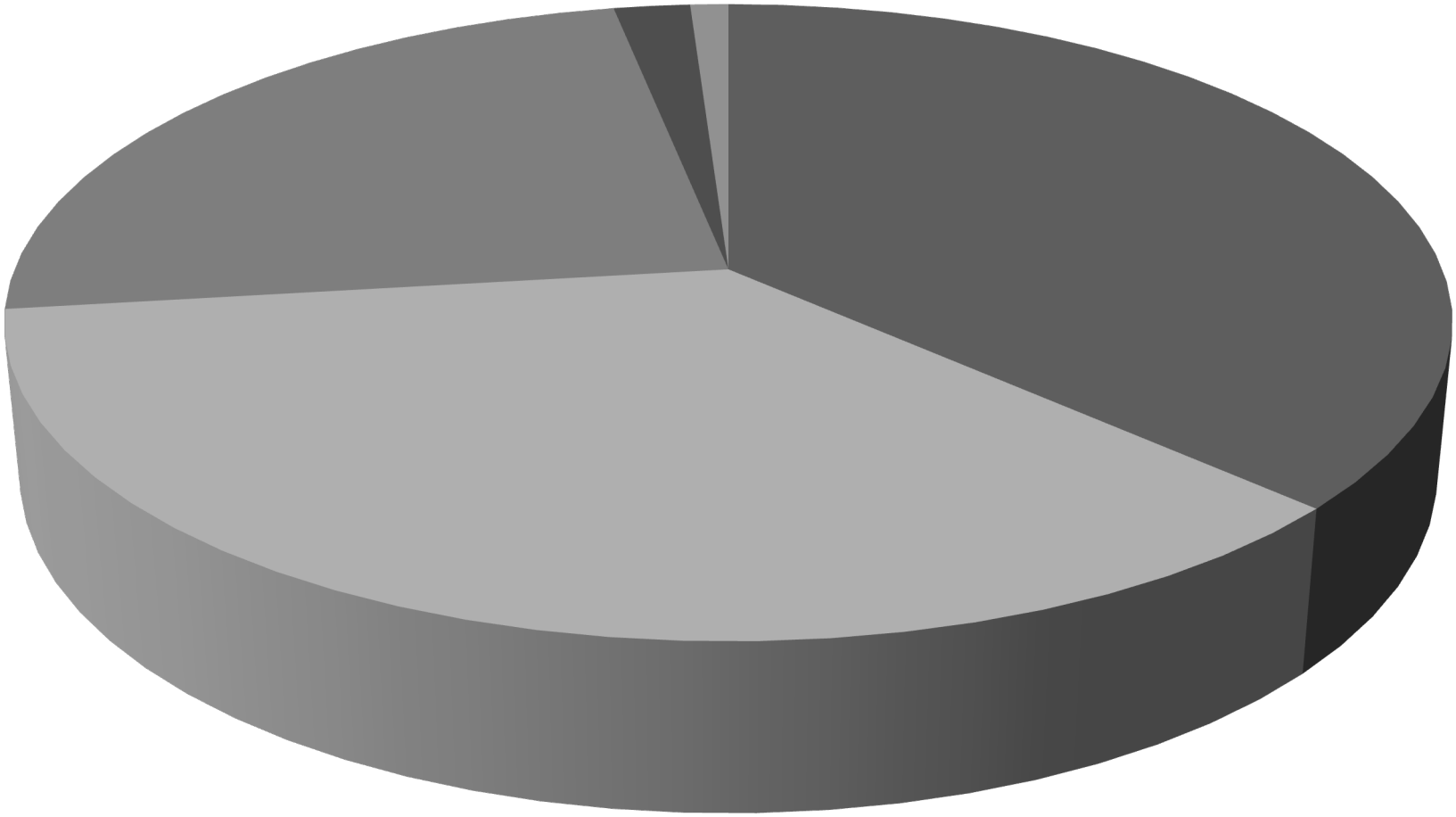
No pie charts.  
No 2.5D charts.





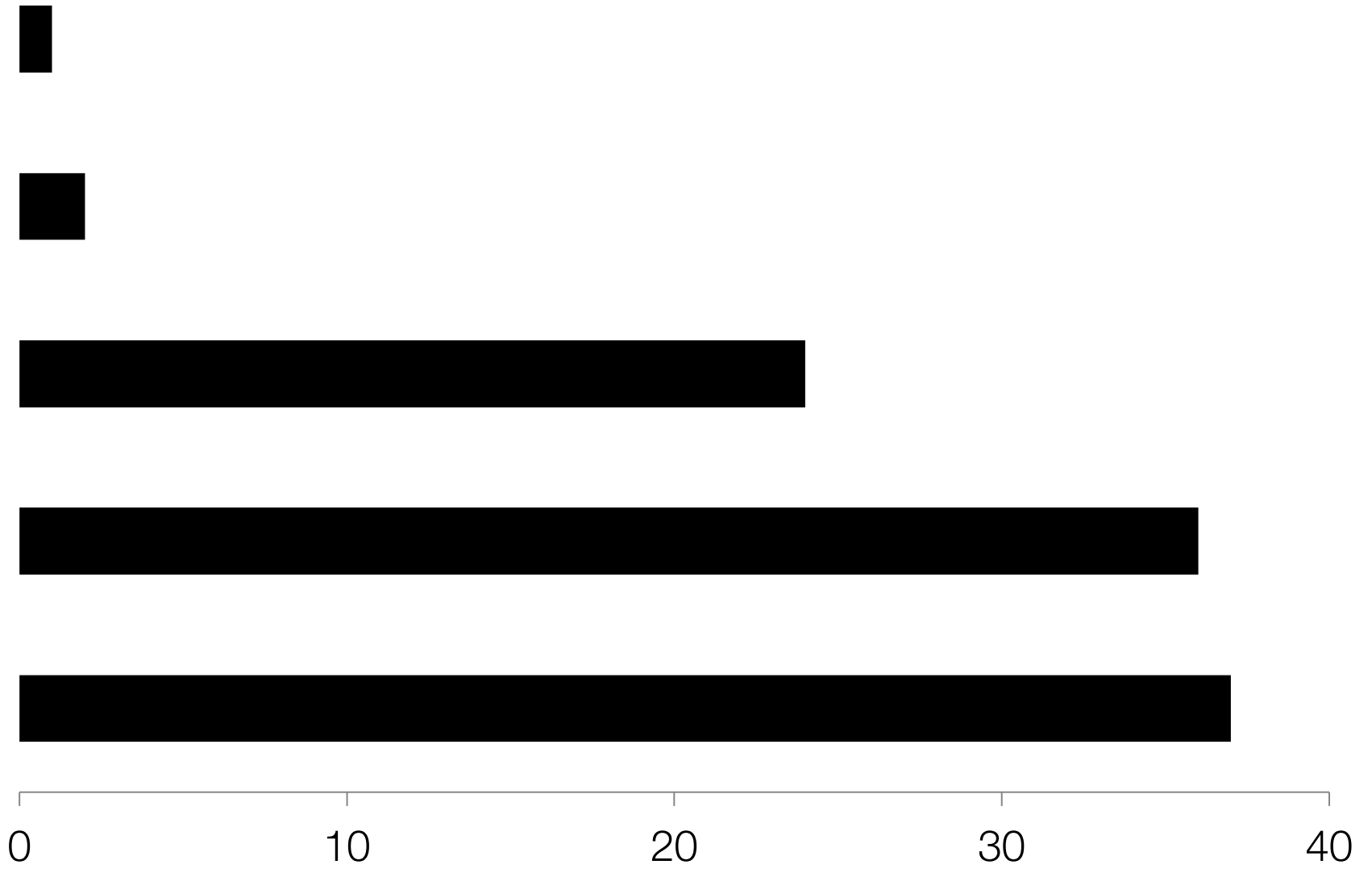




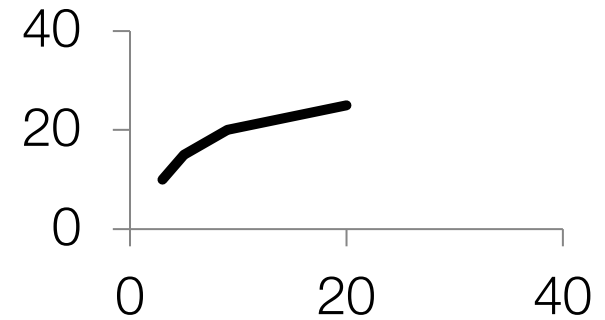
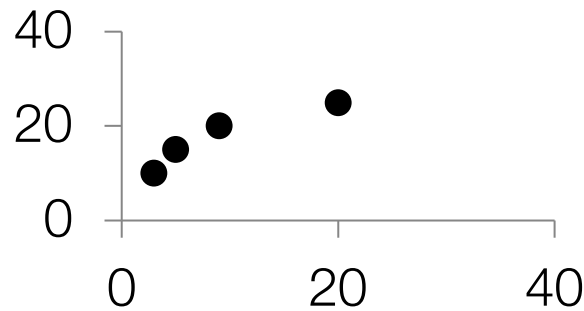
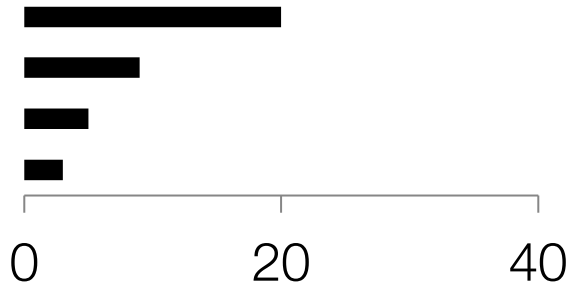




PLEASE DON'T  
EVER DO THIS!

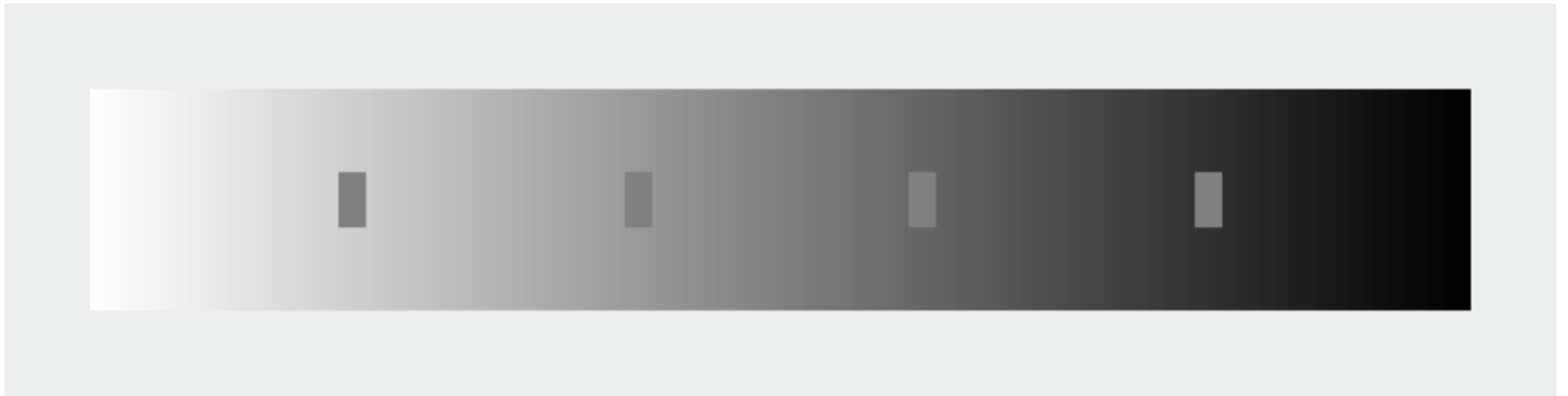


Bar charts, scatterplots, and line charts are *really effective* for quantitative data



Anyone else bored  
by my color choices?

In fact, grayscale can be risky...



In fact, grayscale can be risky...





# Color is Powerful

# Color

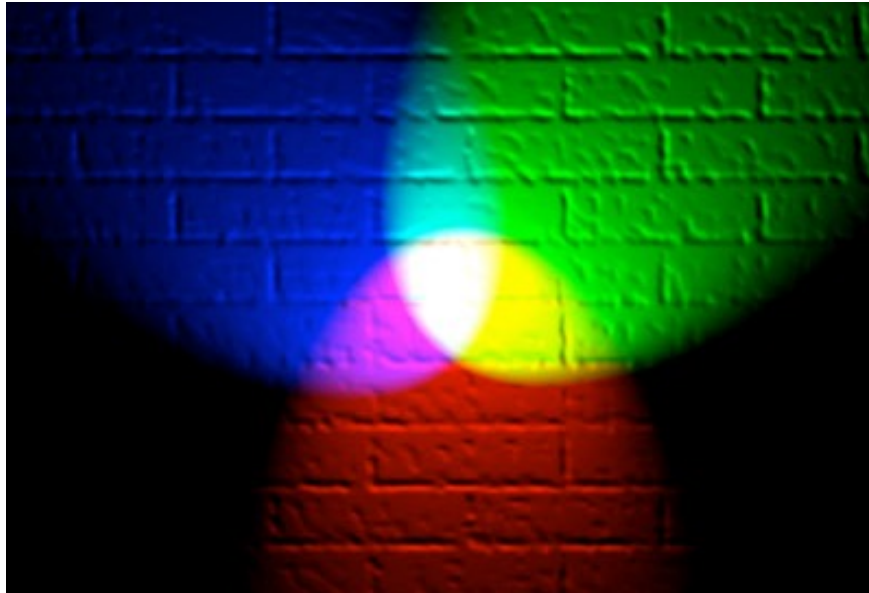
Call attention to information

Increase appeal

Increase memorability

Another dimension to work with

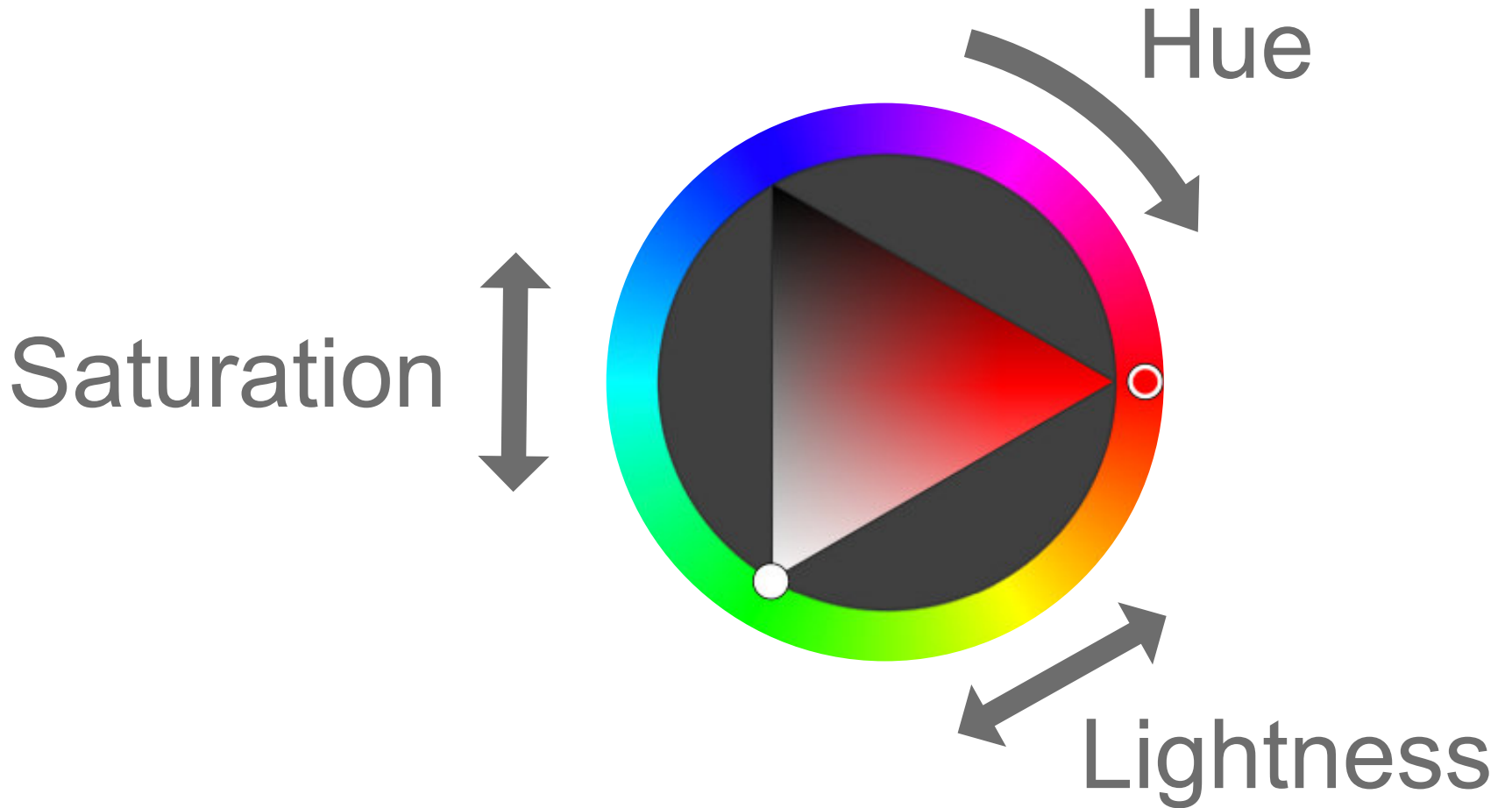
# Have you heard of RGB?



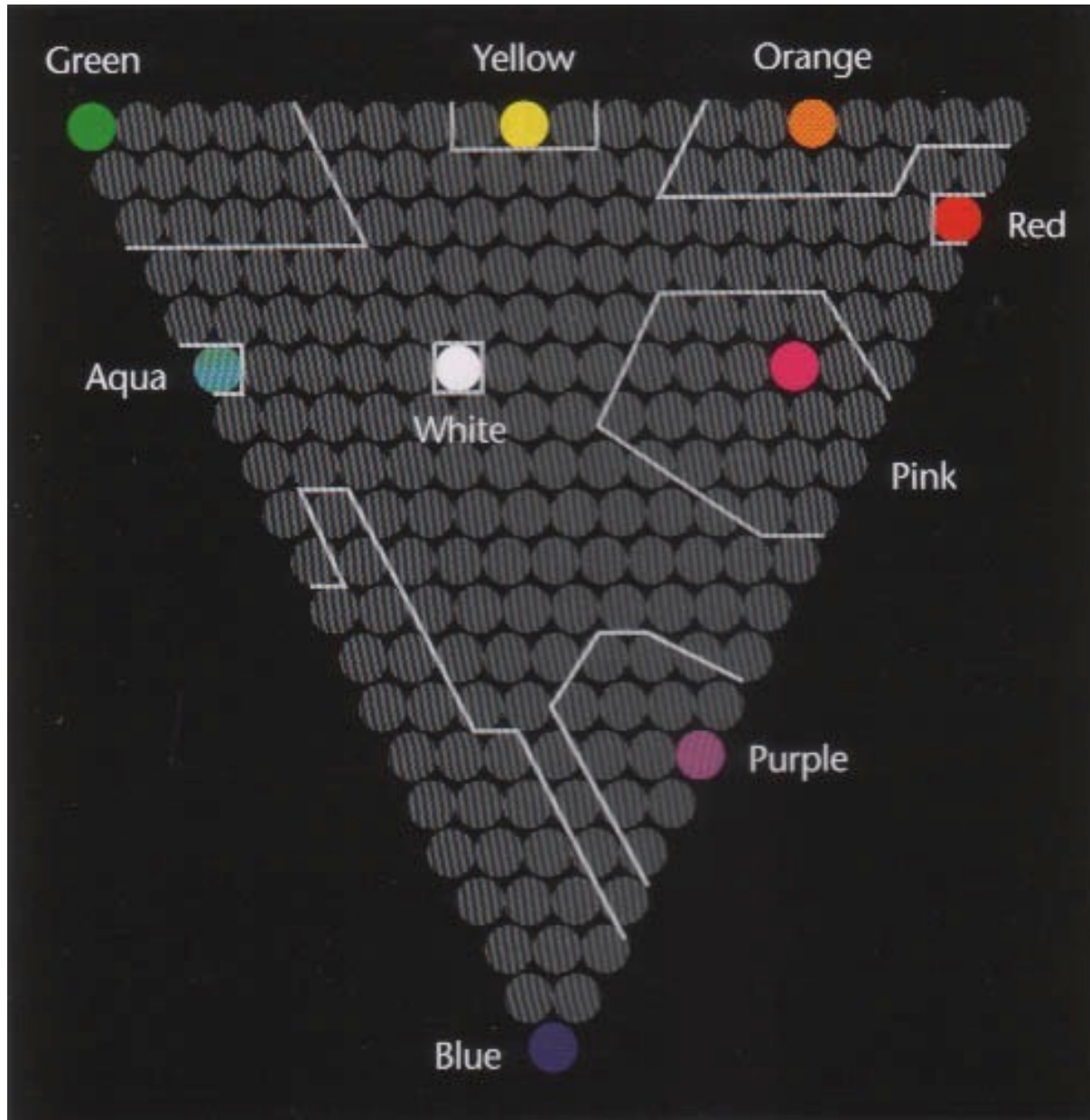
Additive color model: colors create by mixing  
**red, green, blue** light

We see in RGB,  
but we don't interpret in RGB...

# HSV Color Model



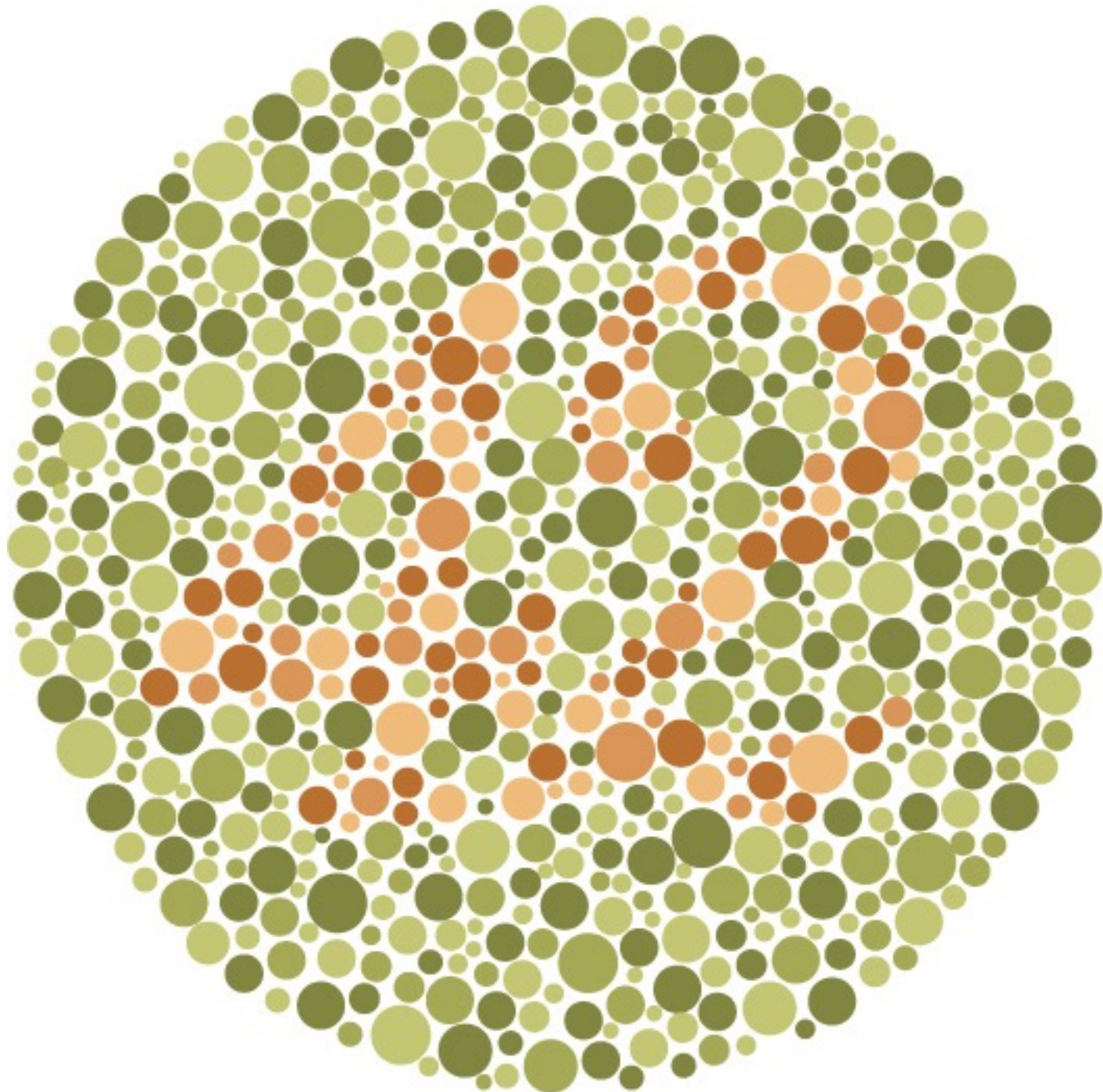
# Hue



Post & Greene, 1986

# Hue and Colorblindness

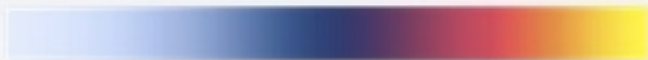
10% of males and 1% of females  
are Red-Green Colorblind







Surface temperature (°C)



0

45

Sep 30, 2014

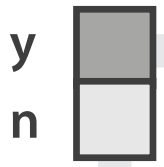
NOAA's Latest High Resolution Weather Model is Released

# Color and Quantitative Data

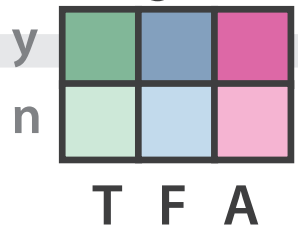
Can you order these (low  $\rightarrow$  hi)?



**Binary**



*Categorical*

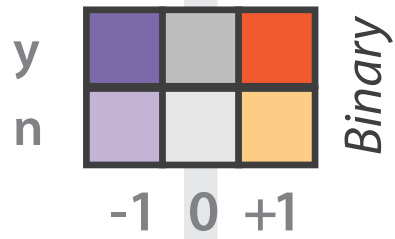


*Binary*



**Categorical**

*Diverging*



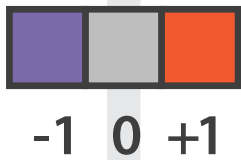
*Binary*

*Categorical*

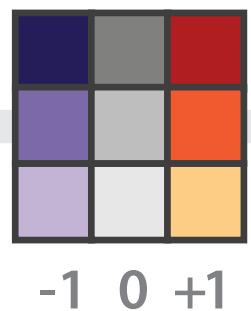


*Sequential*

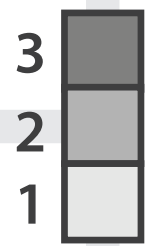
**Diverging**



*Diverging*



*Sequential*



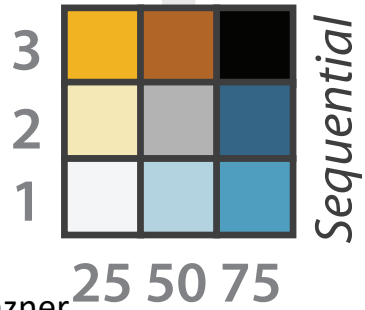
**Sequential**

*Diverging*



*Diverging*

*Sequential*



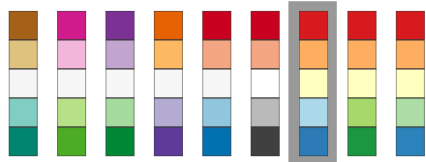
*Sequential*

# Color Brewer for Picking Color Scales

Number of data classes:

Nature of your data:  
 sequential  diverging  qualitative

Pick a color scheme:



Only show:  
 colorblind safe  
 print friendly  
 photocopy safe

Context:

roads  
 cities  
 borders

Background:

solid color  
 terrain

7-class RdYlBu

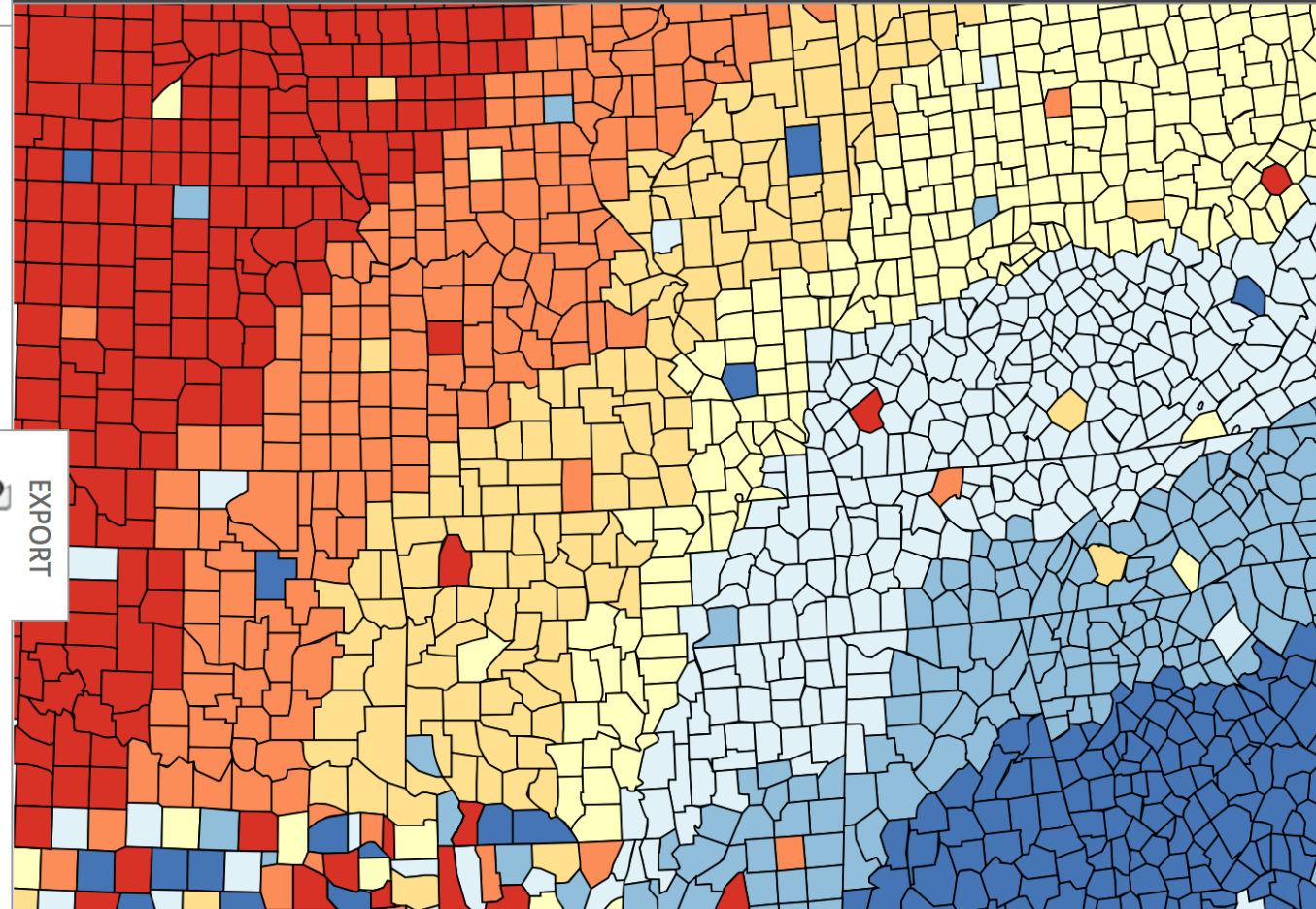


HEX

	#d73027
	#fc8d59
	#fee090
	#ffffbf
	#e0f3f8
	#91b1db
	#4575b4

[how to use](#) [updates](#) [downloads](#) [credits](#)

**COLORBREWER**  
color advice for cartog



# Chroma.js Color Palette Helper

This [chroma.js](#)-powered tool is here to help us [mastering multi-hued, multi-stops color scales](#).

1 What kind of palette do you want to create?

Palette type:  sequential  diverging

Number of colors:

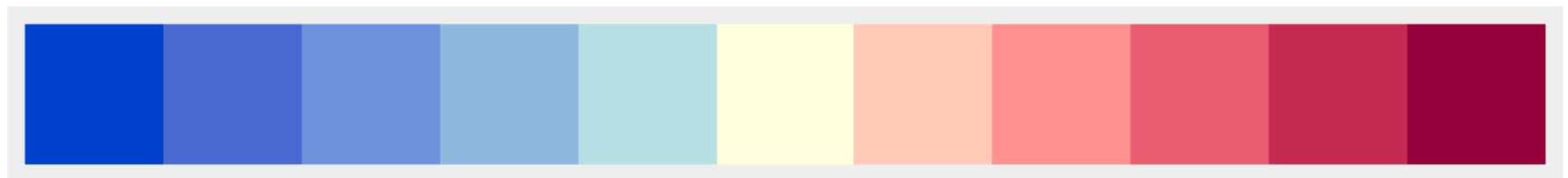
2 Select and arrange input colors

3 Check and configure the resulting palette

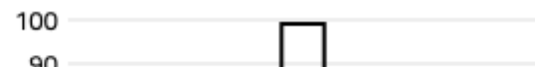
correct lightness  bezier interpolation

✓ This palette is colorblind-safe.

simulate:  normal  deut.  prot.  trit.



lightness



saturation



hue



Overview  
Zoom+Filter  
Details on Demand

Shneiderman Mantra  
(Information-Seeking Mantra)

<https://www.mat.ucsb.edu/g.legrady/academic/courses/11w259/schneiderman.pdf>





<http://visual.ly/every-single-death-game-thrones-series>



# NameGrapher

Explore the historical popularity of United States baby names

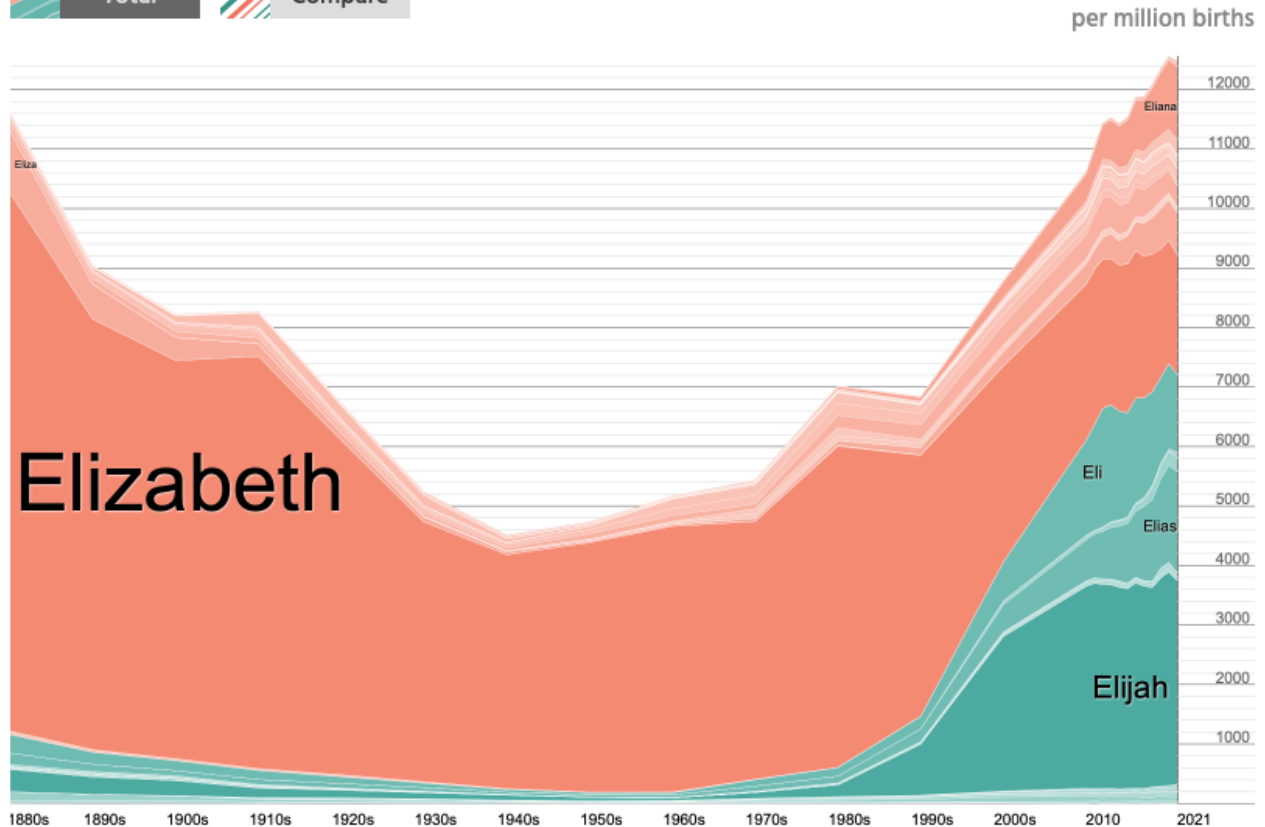
Start typing in the lefthand text box and the graph will update. Click the option buttons for different types of results.  
*Tips:* Be sure to check out both the "total" and "compare" views, and try multiple search terms separated by commas.

Starts with Ends with Contains Exact Match Girls Boys Both

Total Compare

- Eli M
- Elizabeth F
- Elijah M
- Elias M
- Eliana F
- Eliza F
- Elise F
- Elisa F
- Elisabeth F
- Elianna F
- Elian M
- Elisha M
- Elinor F
- Elissa F
- Eliseo M

And 31 more...



# NameVoyager: Explore baby names and name trends letter by letter

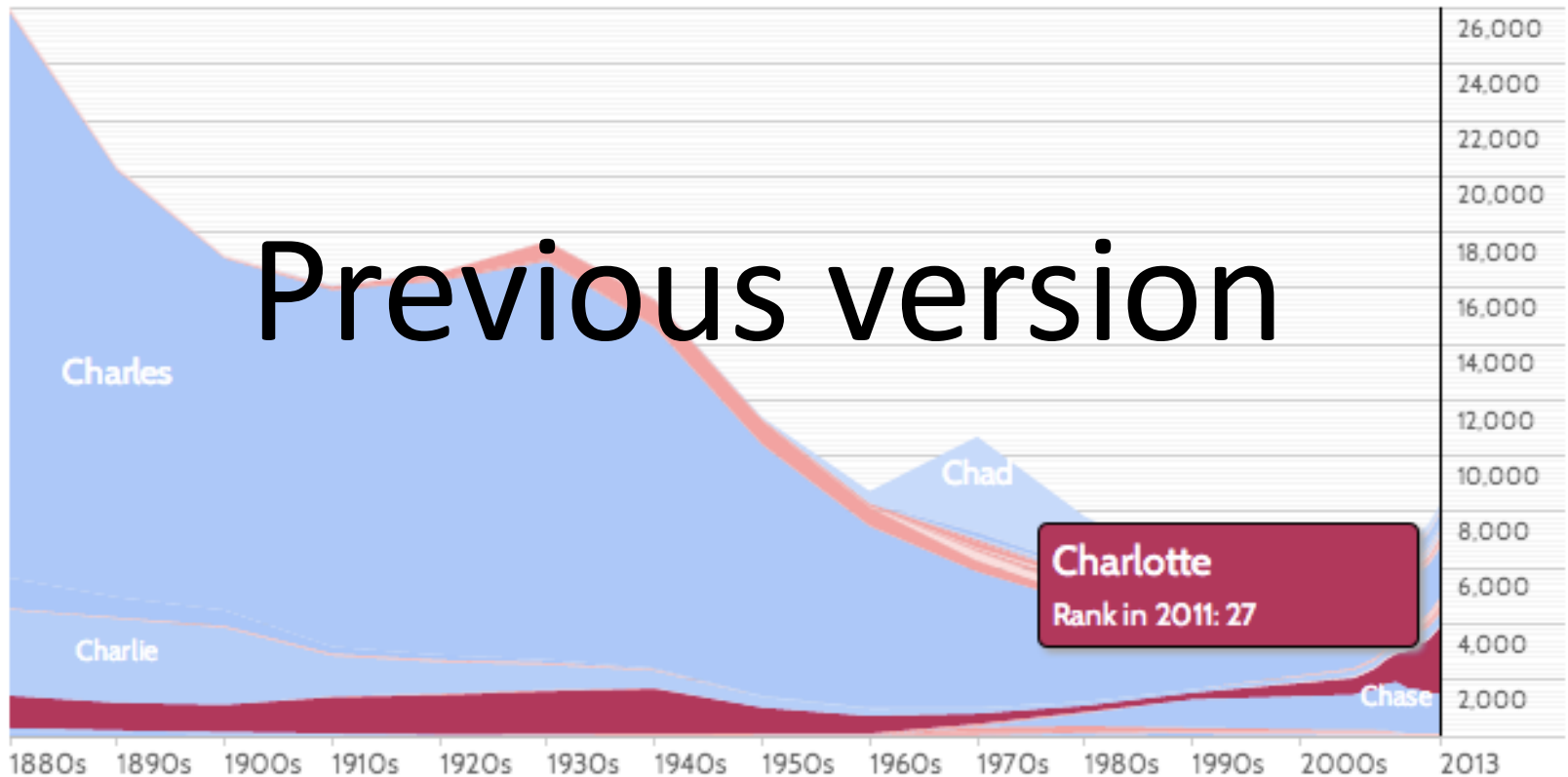
Looking for the perfect baby name? [Sign up for free](#) to receive access to our expert tools!

Baby Name >   Both  Boys  Girls

Current rank: boys	1000	500	100	25	1
girls	1000	500	100	25	1

Names starting with 'CHA' per million babies

per million births



Previous version

Click a name graph to view that name. Double-click to read more about it.

[enlarge](#)

Where to learn more?

# Our Courses

We offer classes in Information Visualization and Visual Analytics for undergraduates as well as graduates in the Spring and Fall semesters.

## Undergraduate

**CS 4460 : Introduction to Information Visualization**

Fall

Spring

**CX 4242 : Data and Visual Analytics**

Fall

Spring

## Graduate

**CS 6730 : Data Visualization: Principles and Applications**

Fall

**CS 7450 : Information Visualization**

Spring

**CS 7451 : Human-Centered Data Analysis**

Spring

**CSE 6242 : Data and Visual Analytics**

Fall

Spring

**CP 8883 : Spatial Network Analysis**

Spring

<https://vis.gatech.edu/courses/>

# How to Make Good Charts

- Edward Tufte's One-Day Workshop
  - <http://www.edwardtufte.com/tufte/courses>
- Edward Tufte, *Visual Display of Quantitative Information*
  - [http://www.edwardtufte.com/tufte/books\\_vdqi](http://www.edwardtufte.com/tufte/books_vdqi)
- Stephen Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*
  - [http://www.amazon.com/Show-Me-Numbers-Designing-Enlighten/dp/0970601972/ref=la\\_B001H6IQ5M\\_1\\_2?s=books&ie=UTF8&qid=1385050724&sr=1-2](http://www.amazon.com/Show-Me-Numbers-Designing-Enlighten/dp/0970601972/ref=la_B001H6IQ5M_1_2?s=books&ie=UTF8&qid=1385050724&sr=1-2)

# Visualization Theory “Books”

- Tamara Munzner VIS Tutorial and Book
  - <http://www.cs.ubc.ca/~tmm/talks.html>
  - <http://www.cs.ubc.ca/~tmm/vadbook/>
- Colin Ware, *Information Visualization: Perception for Design*
  - <http://www.amazon.com/Information-Visualization-Perception-Interactive-Technologies/dp/1558605118>
- Stephen Few, *Now You See It*
  - [http://www.amazon.com/Now-You-See-Visualization-Quantitative/dp/0970601980/ref=pd\\_bxgy\\_b\\_img\\_z](http://www.amazon.com/Now-You-See-Visualization-Quantitative/dp/0970601980/ref=pd_bxgy_b_img_z)
- Edward Tufte, *Envisioning Information*
  - [http://www.edwardtufte.com/tufte/books\\_ei](http://www.edwardtufte.com/tufte/books_ei)
- Edward Tufte, *Visual Explanations*
  - [http://www.edwardtufte.com/tufte/books\\_visex](http://www.edwardtufte.com/tufte/books_visex)
- Edward Tufte, *Beautiful Evidence*
  - [http://www.edwardtufte.com/tufte/books\\_be](http://www.edwardtufte.com/tufte/books_be)
- Tamara Munzner, *Visualization Analysis & Design*
  - <http://www.amazon.com/Visualization-Analysis-Design-AK-Peters/dp/1466508914>

# Perception and Color Websites

- Chris Healy, NC State
  - <http://www.csc.ncsu.edu/faculty/healey/PP/index.html>
- Color Brewer
  - <http://colorbrewer2.org/>
- Maureen C. Stone (Color Links, Blog, Workshops)
  - <http://www.stonesc.com/color/index.htm>
- Subtleties of Color by Robert Simmon of NASA
  - <http://blog.visual.ly/subtleties-of-color/>

# Visualization Blogs

- Flowing Data by Nathan Yau
  - <http://flowingdata.com/>
- Information Aesthetics by Andrew Vande Moere
  - <http://infosthetics.com/>
- Information is Beautiful by David McCandless
  - <http://www.informationisbeautiful.net/>
- Visual.ly Blog
  - <http://blog.visual.ly/>
- Indexed Comic by Jessica Hagy
  - <http://thisisindexed.com/>



# Infographics

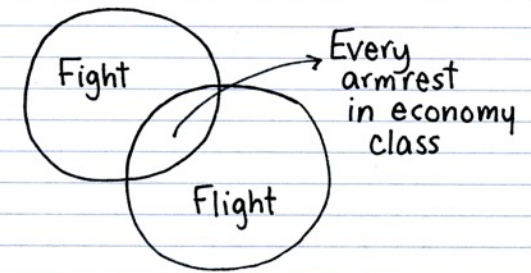
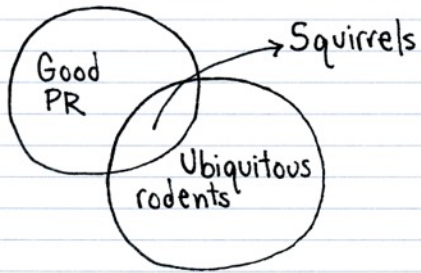
[Visual.ly/view](#)

([wtfviz.net](#))

Thanks!

Chad Stolper

chadstolper@gatech.edu



# Questions?

 Chad Stolper

chadstolper@gatech.edu