

Tutorial on Data Visualization for Graduate Students

Uichin Lee
KAIST

Why Viz **Design Principles** **Charting 101**

Slides largely from CS171 by Prof. Pfister at Harvard University

Visualization

To convey information through graphical representations of data

Visualization Goals

- **Record** information
- **Analyze** data to support reasoning
- **Reveal patterns** to support hypotheses
- **Communicate** ideas to others

Record

Leonardo DaVinci, ca. 1500 Galileo Galilei, 1616

The History of Visual Communication
The Galileo Project, Rice University

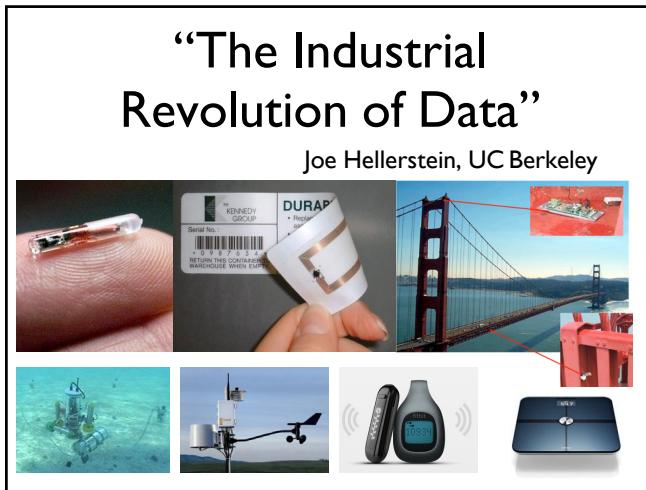
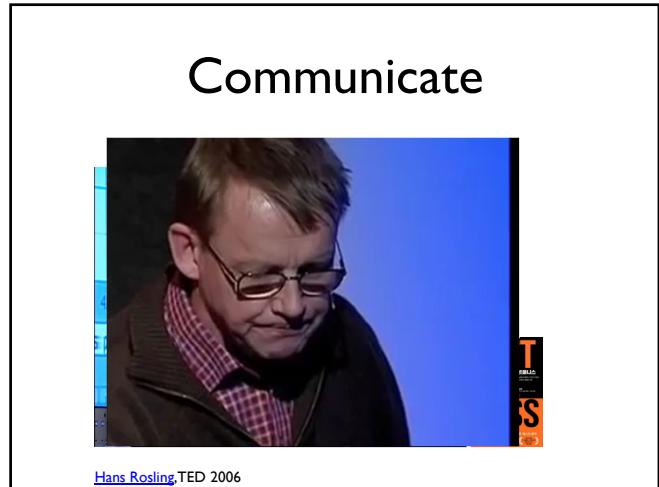
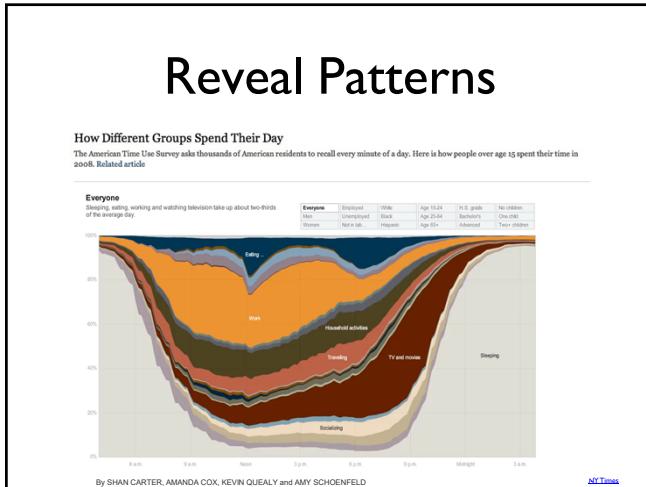
Analyze

W. Playfair, 1786 W. Playfair, 1801

wikipedia.org

Reveal Patterns

John Snow, 1854 E.Tufte, Visual Explanations, 1997



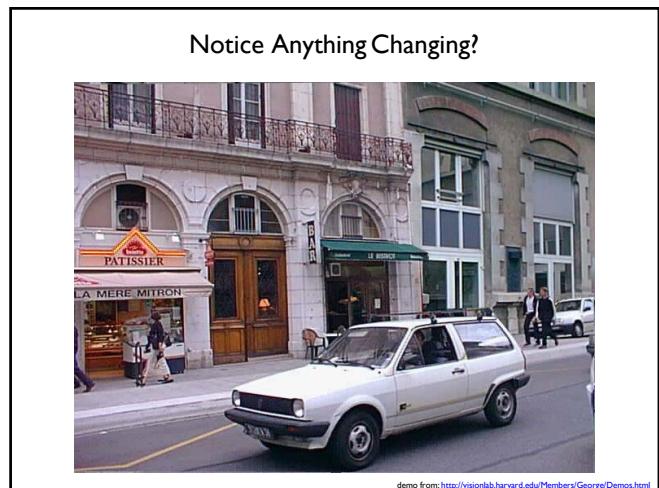
Microsoft Excel: sd_evar_bgs3.dfr

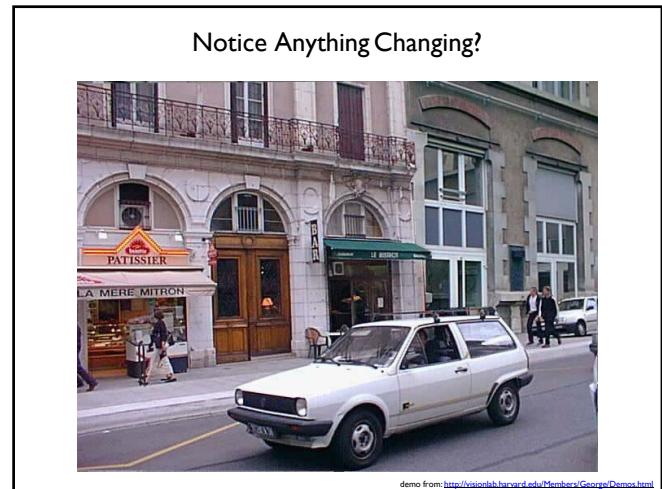
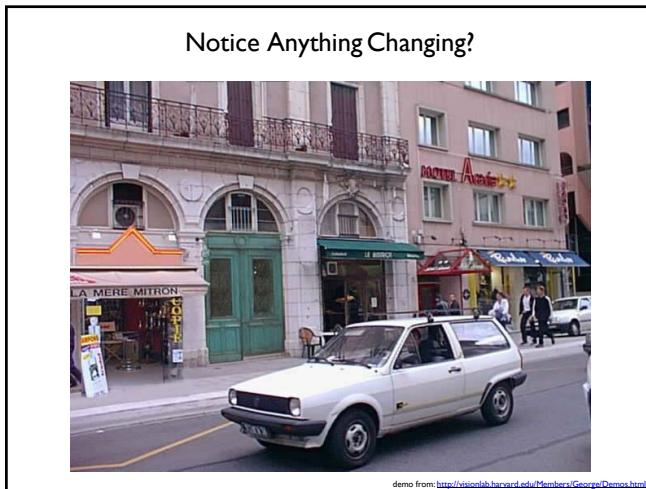
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ObjectID	ST_CNT	STC_CONTRACT	BTRIPS	POP2000	POP2005	POP00_SQMI	POP05_SQMI	WHITE	BLACK	AMERI_ESIAN			
2	25044	06	073	06973	003207	1060710032073	2033	2160	3835.8	4075.5	1244	133	13	21
3	25305	06	073	06973	003207	1060710032073	1855	1803	1987.1	1927.2	1616	12	10	14
4	25306	06	073	06973	003207	1060710032073	424	445	6057.1	6337.1	394	1	1	15
5	25308	06	073	06973	008324	1060730083243	660	688	6600.0	6880.0	590	4	0	41
6	25309	06	073	06973	008324	1060730083244	1908	1994	3028.6	3165.1	1712	15	2	10
7	25310	06	073	06973	008324	1060730083245	524	540	794.9	7942.9	5111	11	0	11
8	25311	06	073	06973	008324	1060730083246	568	613	7100.0	7322.9	5118	7	0	28
9	25312	06	073	06973	008324	1060730083247	812	833	5075.0	5206.3	724	7	2	42
10	25313	06	073	06973	008327	1060730083271	2836	3321	1524.7	1785.5	2414	12	1	25
11	25314	06	073	06973	008327	1060730083272	2179	2614	3961.6	4752.7	1769	8	4	31
12	25315	06	073	06973	008329	1060730083291	2251	2707	831	980.1	1930	11	3	26
13	25316	06	073	06973	008329	1060730083291	4334	4919	4422.4	5019.4				

Matthew Ericson, NY Times

Limited Cognition!

Notice anything changing?
Get ready!





Limited Cognition

Which **gender or income level group** shows different effects of **age** on triglyceride levels?

Income Group	Males		Females	
	Under 65	65 or Over	Under 65	65 or Over
0-\$24,999	250	200	375	550
\$25,000+	430	300	700	500

Slide after Stephen Koslyn, Clear and to the Point

Visual Queries

Triglyceride Level

Males Females

Legend: 0-\$24,999 (blue), \$25,000+ (orange), 0-\$24,999 (green), \$25,000+ (red)

Slide after Stephen Koslyn, Clear and to the Point

“It is things that make us smart.”

Donald Norman

The History of Visual Communication

“It is things that make us smart.”

Donald Norman

Visual Thinking Collection, Dave Grey

Idea Maps, by Jamie Nast

Visualization

- Uses perception to free up cognition
- Serves as an external aid to augment working memory
- Boosts our cognitive abilities

“Visualization is really about external cognition, that is, how resources outside the mind can be used to boost the cognitive capabilities of the mind.”



Stuart Card

Tutorial on Data Visualization for Graduate Students

Why Viz	Design Principles	Charting 101
---------	-------------------	--------------

• Goals: record, analyze, reveal patterns, communicate
 • Viz = External cognition

Slides largely from CS171 by Prof. Pfister at Harvard University

Design

“Most people make the mistake of thinking **Design is what it looks like.**
People think it's this veneer – that the designers are handed this box and told, 'Make it look good!'

That's not what we think design is.
 It's not just what it looks like and feels like.
Design is how it works. — Steve Jobs



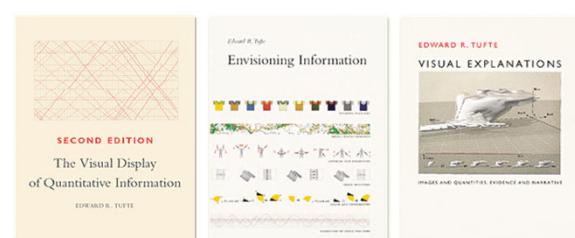
Design

“Design must be **functional**, and functionality must be translated into visual aesthetics, without any reliance on gimmicks that have to be explained.”



F. A.
Porsche

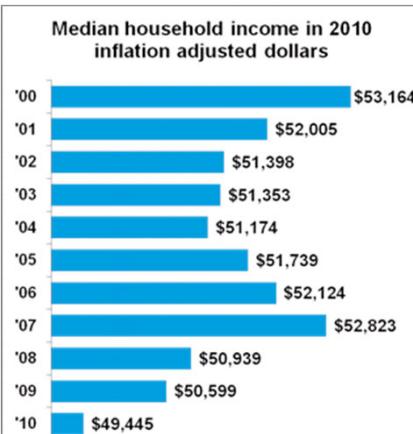
Edward Tufte



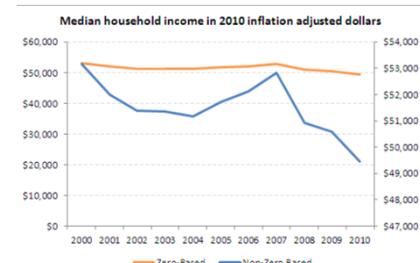
Outline

- Graphical Integrity
- Visualization Design Principles
- Graphic Design Principles

Graphical Integrity

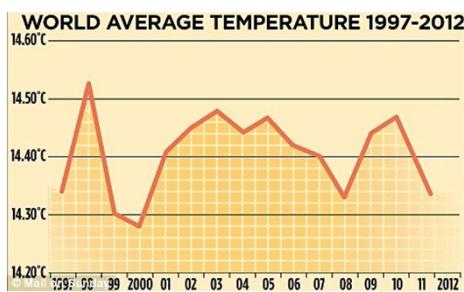


Scale Distortions

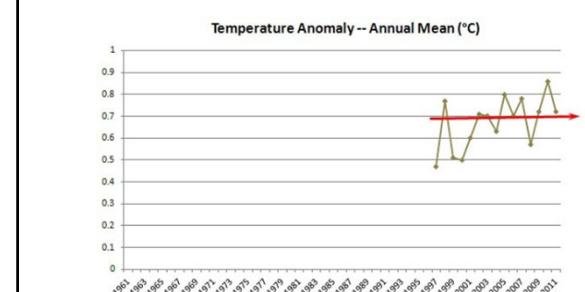


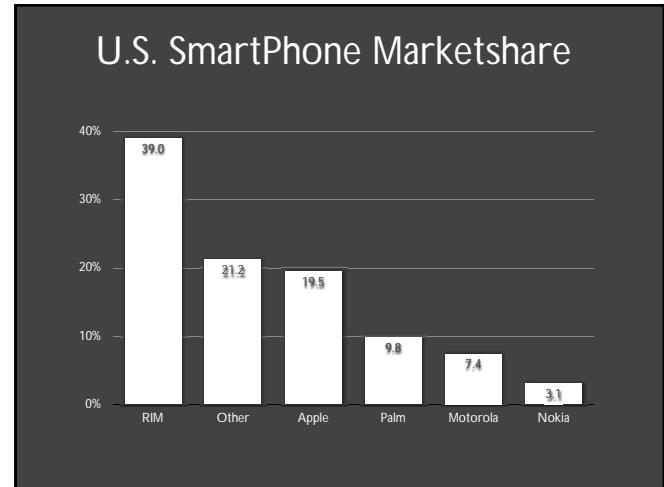
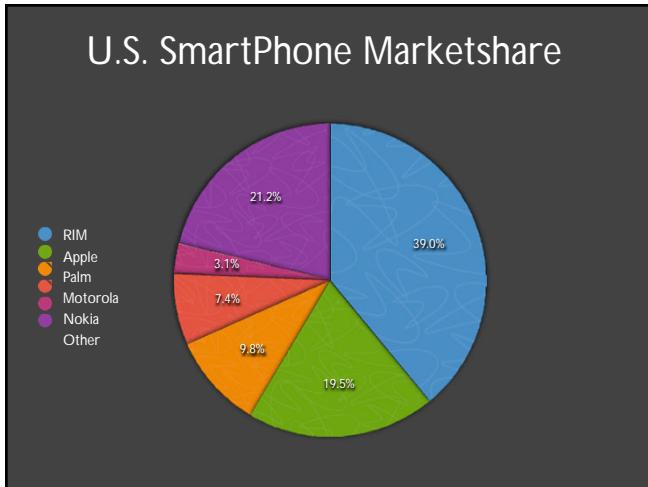
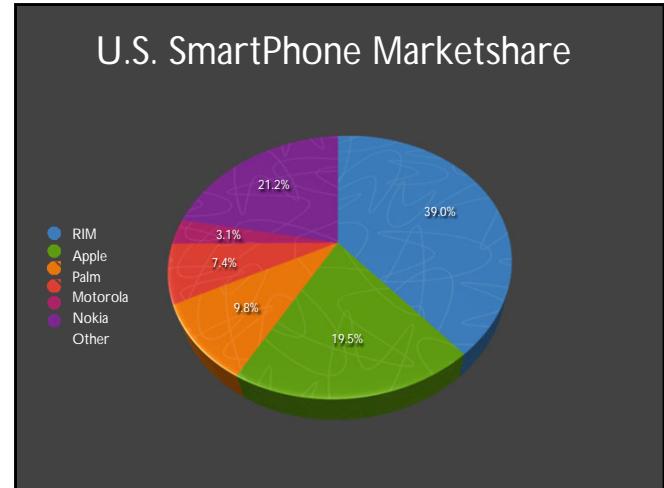
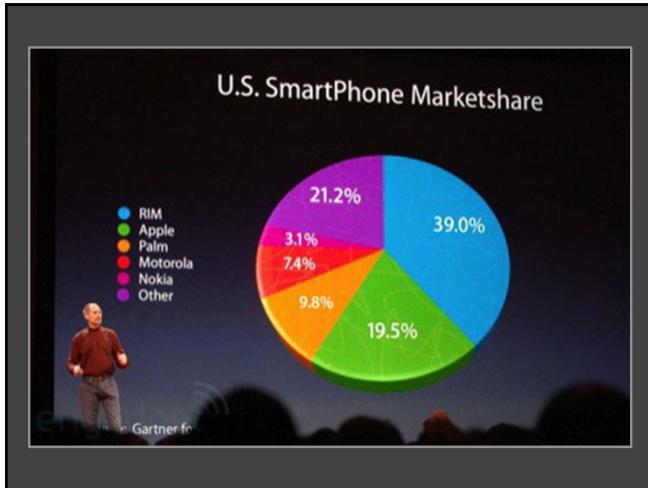
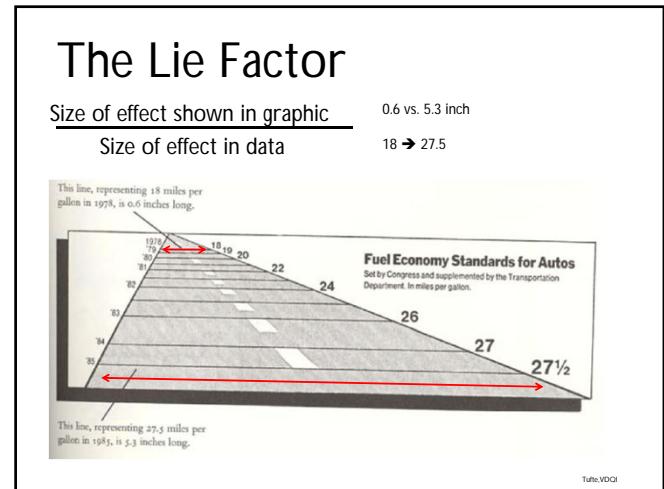
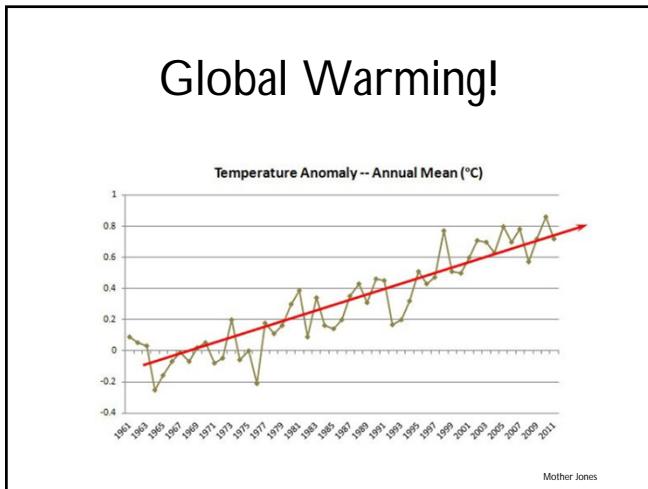
A. Kriebel/VizWiz

Global Warming?



Global Warming?





Tufte's Integrity Principles

- Clear, detailed, and thorough labeling and appropriate scales
- Size of the graphic effect should be directly proportional to the numerical quantities (beware of “lie factor”)
- Show data variation, not design variation

Visualization Design Principles

Maximize Data-Ink Ratio , with reason

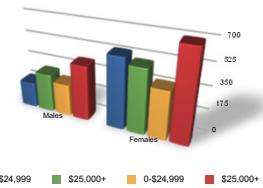
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

= proportion of a graphic's ink devoted to the non-redundant display of data-information
= 1.0 - proportion of a graphic that can be **erased without loss of data-information**

Above all else show the data!

Maximize Data-Ink Ratio , with reason

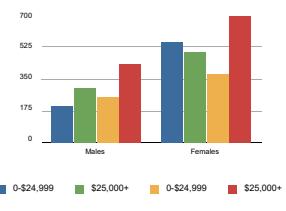
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



How can we maximize data-ink ratio?

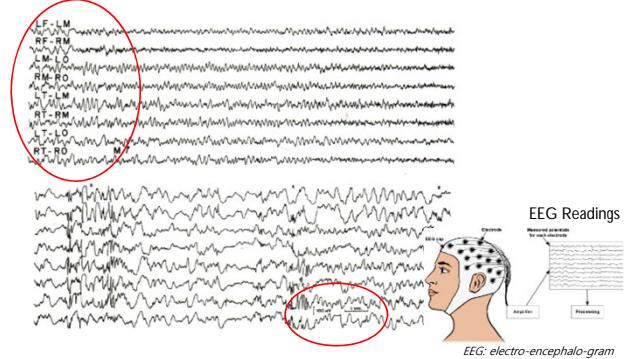
Maximize Data-Ink Ratio , with reason

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



After erasing non-data ink

Maximize Data-Ink Ratio , with reason

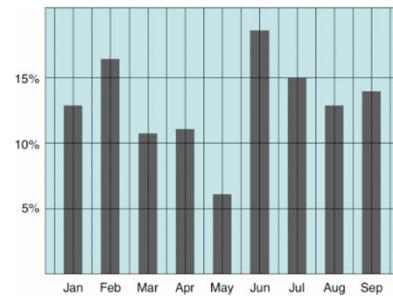


Maximize Data-Ink Ratio , with reason



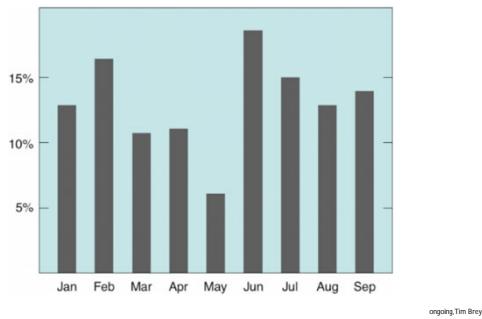
Avoid Chartjunk

Extraneous visual elements that distract from the message

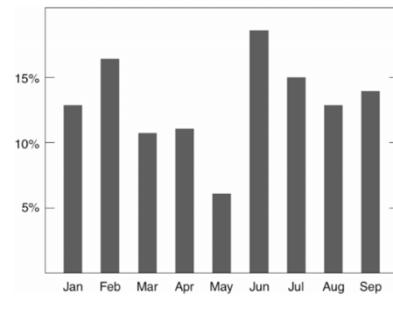


ongoing Tim Brey

Avoid Chartjunk



Avoid Chartjunk

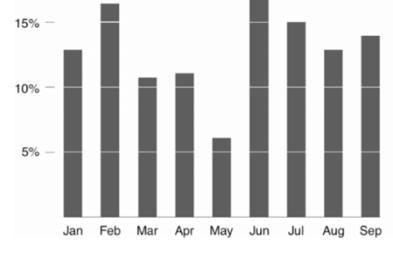


ongoing Tim Brey

Avoid Chartjunk

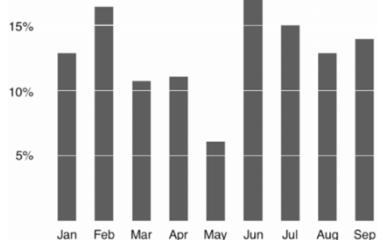


Avoid Chartjunk



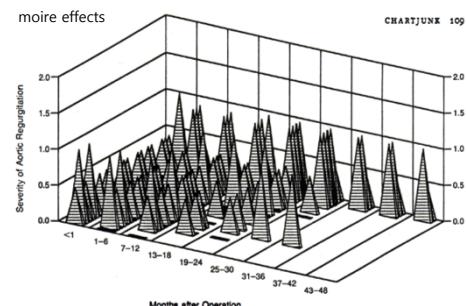
ongoing Tim Brey

Avoid Chartjunk

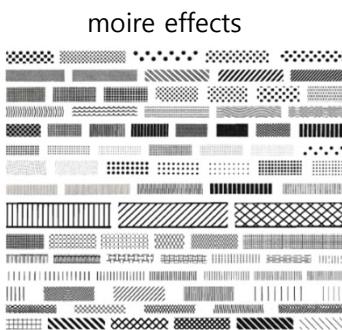


originally Tim Bracy

Avoid Chartjunk

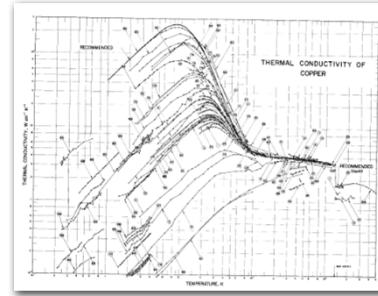


Avoid Chartjunk

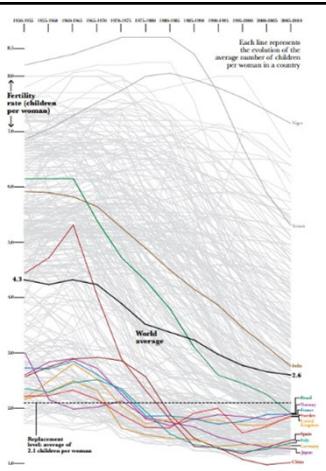


Increase Data Density

$$\text{Data density} = \frac{\text{Number data items}}{\text{Area of data in graphic}}$$

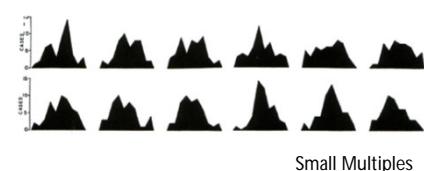


Ho et al., "Thermal Conductivity of the Elements: A Comprehensive Review"
J. Phys. Chem. 1974



Increase Data Density

$$\text{Data density} = \frac{\text{Number data items}}{\text{Area of data in graphic}}$$



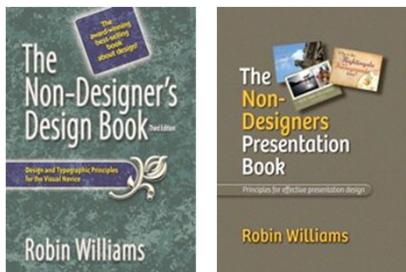
Edmond A. Murphy, "One Cause? Many Causes? The Argument from the Bimodal Distribution," Journal of Chronic Diseases, 17 (1964), 309.

Tufte's Design Principles

- Maximize data-ink ratio
- Avoid chart junk
- Increase data density

Graphic Design Principles

Robin Williams



**Contrast
Repetition
Alignment
Proximity**

Robin Williams

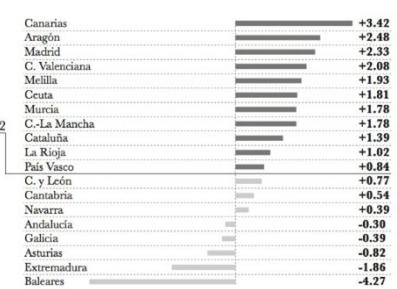
Principle of Contrast

If two items are not exactly the same, then make them different. Really different.

Don't be a wimp.

Unemployment rates by region (in October)

Percentage change compared to previous month



A. Cairo,
The Functional Art

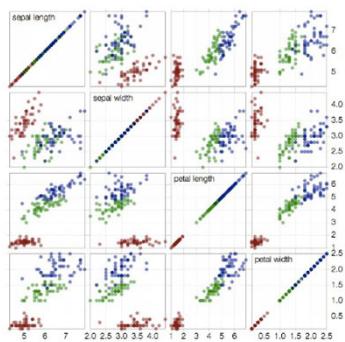
Contrast Repetition Alignment Proximity

Robin Williams

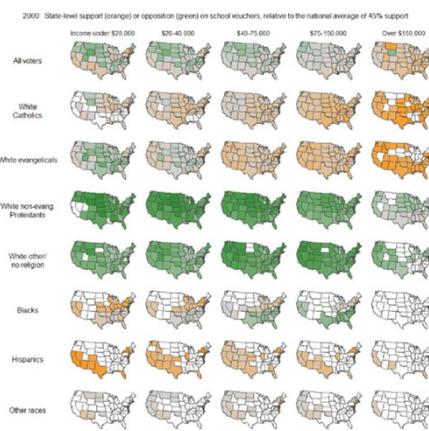
Principle of Repetition

Repeat some aspects of the design throughout the entire piece

Small Multiples



D3, M. Bostock



A. Gelmann

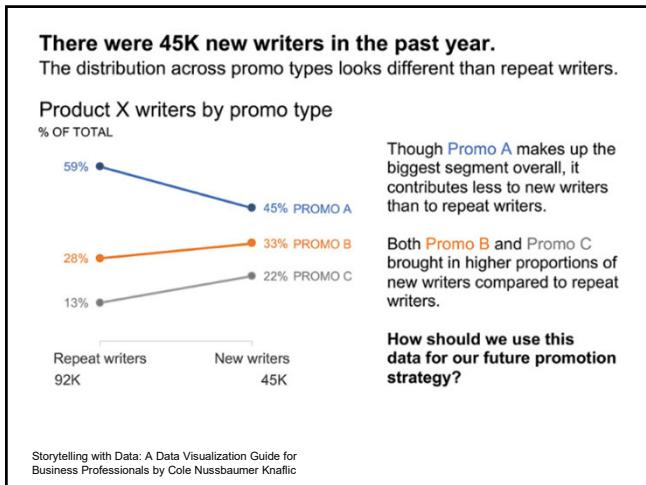
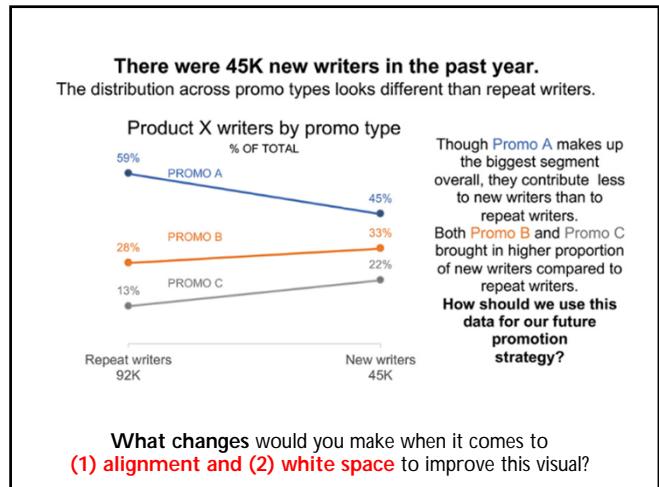
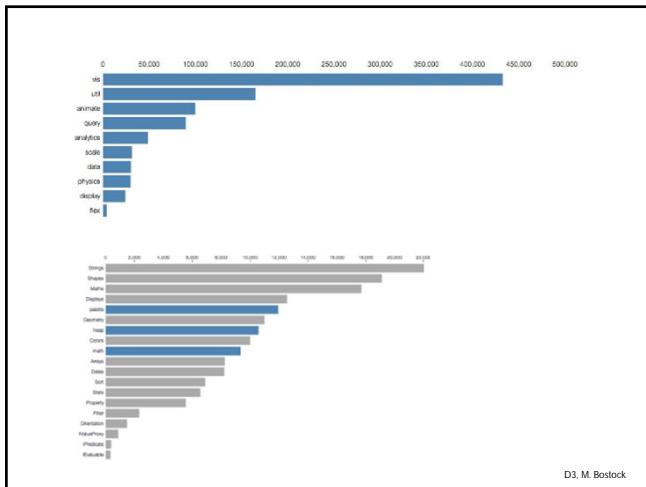
Contrast Repetition Alignment Proximity

Robin Williams

Principle of Alignment

Nothing should be placed on the page arbitrarily

Every item should have a visual connection with something else

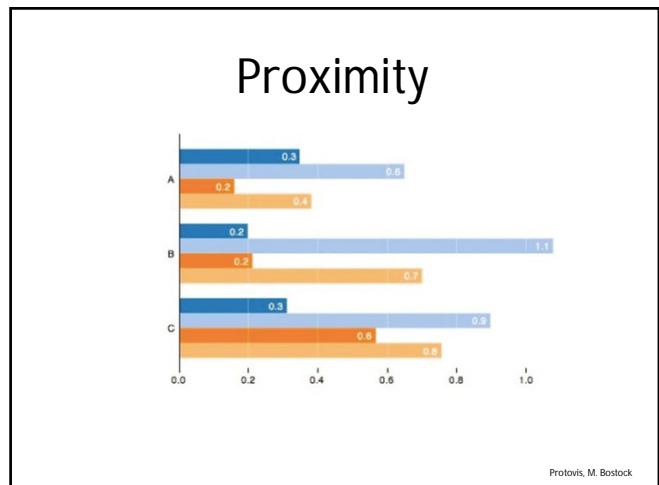


Contrast Repetition Alignment Proximity

Robin Williams

Principle of Proximity

Group related items together . . .
as physical closeness implies a relationship



CRAP + Color Use

Follow "Get it Right in Black and White"

[Get it right in black and white. Stone. 2010.](https://web.archive.org/web/20160225055303/http://www.stonesc.com/wordpress/2010/03/get-it-right-in-black-and-white)

James Clifton Thomas
123 Penny Lane
Portland, OR 97211
(888) 555-1212

PROFILE:
A multi-talented, hard-working young man, easy to get along with, dependable, and joyful.

ACCOMPLISHMENTS:
January 2008-present Web designer and developer, working with a professional team of creatives in Portland.

May 2006-January 2006 Pecker Full of Posies Day Care Center. Changed diapers, taught magic and painting, wiped noses, read books to and danced with babies and toddlers. Also coordinated schedules, hired other teachers, and developed programs for children.

Summer 2006 Updated the best-selling book, *The Non-Designer's Web Book* with my mom (Robin Williams) and John Tollett.

1997-2000 Developed and led a ska band called Lead Veins. Designed the web site and coordinated a national tour.

EDUCATION:
2002-2005 Pacific Northwest College of Art, Portland, Oregon; B.A. in Printmaking
1999-2000 Santa Rosa High School, Santa Rosa, California
1997-1998 Santa Fe High School, Santa Fe, New Mexico
1982-1986 Peppy Creek Daycare Center, Santa Rosa, California

PROFESSIONAL AFFILIATIONS:
Grand National Monotype Club, Executive Secretary, 2000-2002
Jerkin Williams, Member of Portland President, 1999-present
Local Organization of Children of Robin Williams, 1982-present

Hobbies:
Snowboarding, skateboarding, tap dancing, cooking, magic, music (trumpet, drums, vocals, bass guitar), portrait drawing

References available on request.

Tvsfpq wC
Ls{ \$s\$y tvszic

James Clifton Thomas
123 Penny Lane
Portland, OR 97211
(888) 555-1212

PROFILE:
A multi-talented, hard-working young man, easy to get along with, dependable, and joyful.

ACCOMPLISHMENTS:
January 2008-present Web designer and developer, working with a professional team of creatives in Portland.

May 2006-January 2006 Pecker Full of Posies Day Care Center. Changed diapers, taught magic and painting, wiped noses, read books to and danced with babies and toddlers. Also coordinated schedules, hired other teachers, and developed programs for children.

Summer 2006 Updated the best-selling book, *The Non-Designer's Web Book* with my mom (Robin Williams) and John Tollett.

1997-2000 Developed and led a ska band called Lead Veins. Designed the web site and coordinated a national tour.

EDUCATION:
2002-2005 Pacific Northwest College of Art, Portland, Oregon; B.A. in Printmaking
1999-2000 Santa Rosa High School, Santa Rosa, California
1997-1998 Santa Fe High School, Santa Fe, New Mexico
1982-1986 Peppy Creek Daycare Center, Santa Rosa, California

PROFESSIONAL AFFILIATIONS:
Grand National Monotype Club, Executive Secretary, 2000-2002
Jerkin Williams, Member of Portland President, 1999-present
Local Organization of Children of Robin Williams, 1982-present

Hobbies:
Snowboarding, skateboarding, tap dancing, cooking, magic, music (trumpet, drums, vocals, bass guitar), portrait drawing

References available on request.

Tutorial on Data Visualization for Graduate Students

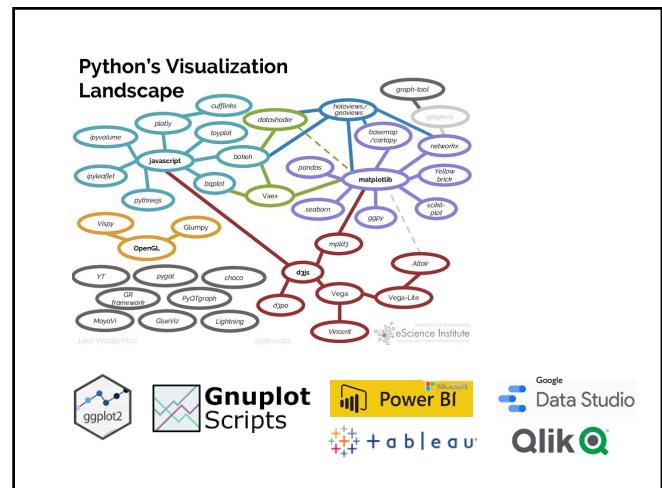
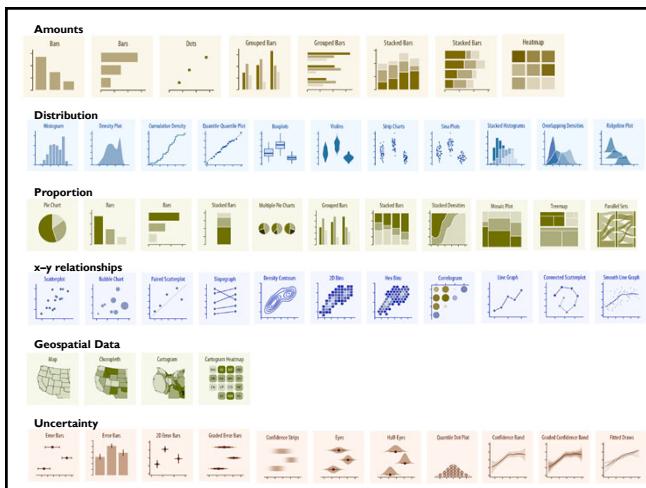
Why Viz

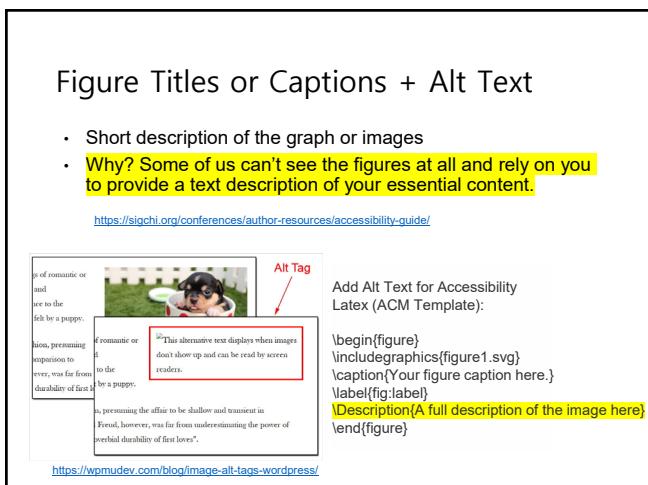
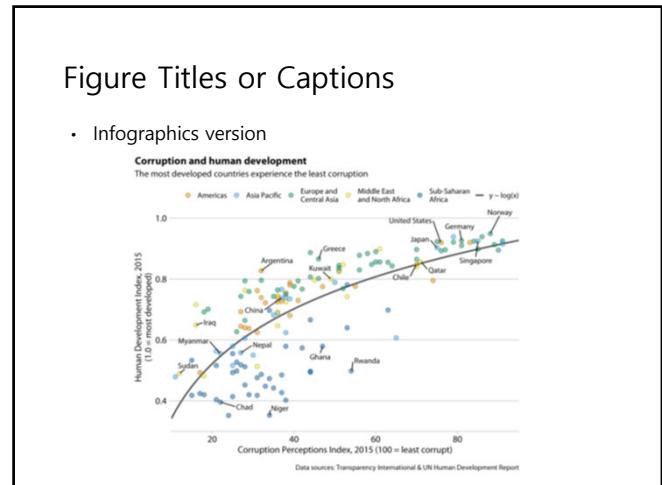
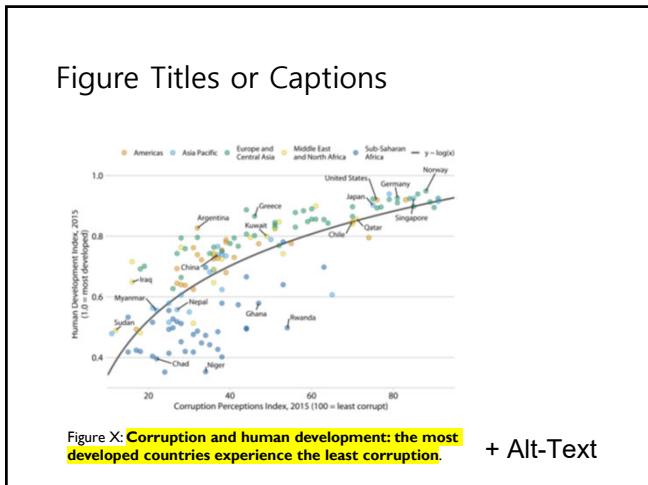
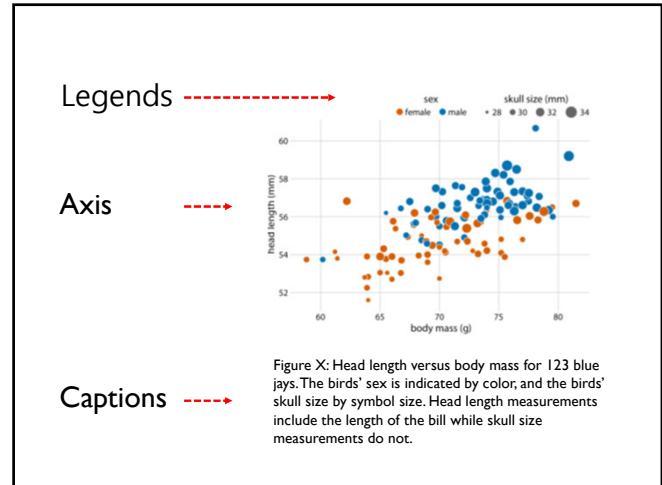
- Goals: record, analyze, reveal patterns, communicate
- Viz = External cognition

Design Principles

- Integrity: avoid scale distortion & lie factor
- Viz design: mix data-link ratios, avoid chartjunk, increase data density
- Design principles: CRAP - contrast, repetition, alignment, proximity (grouping)

Charting 101





Tables

a

Rank	Title	Amount
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

b

Rank	Title	Amount
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

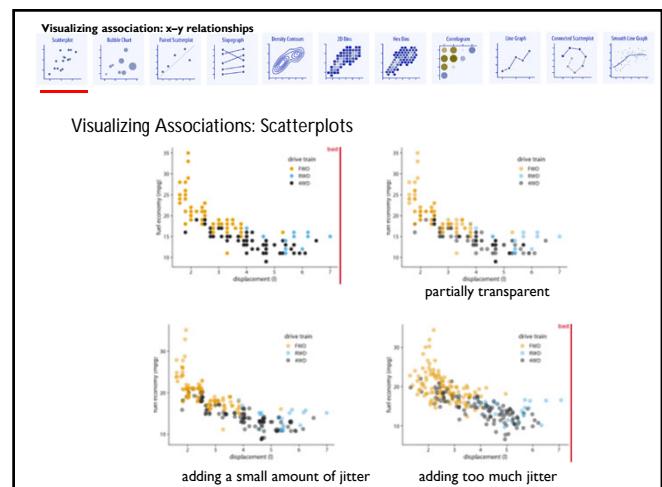
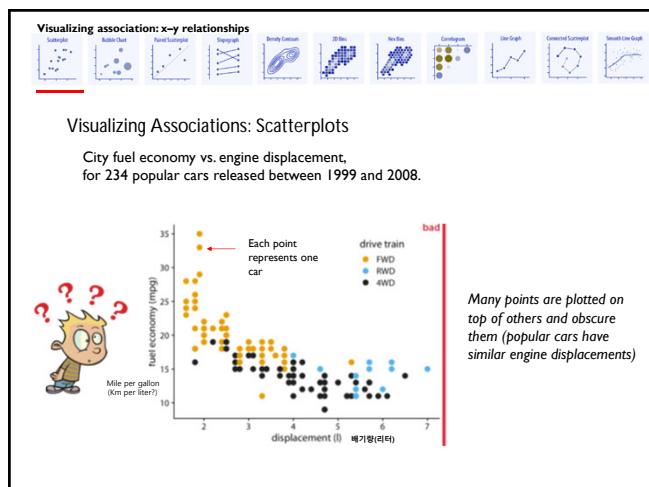
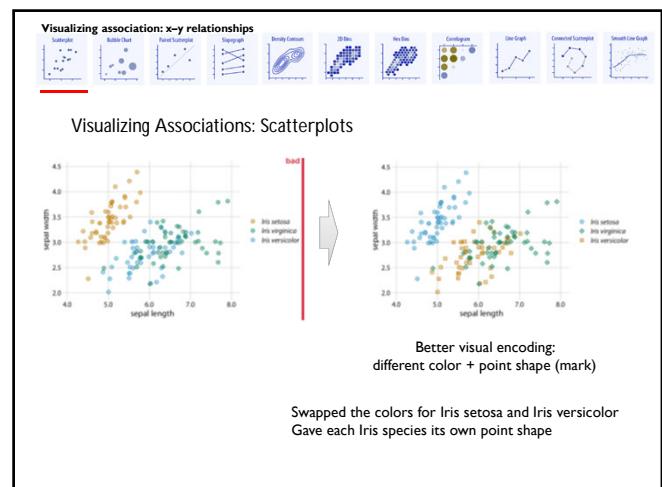
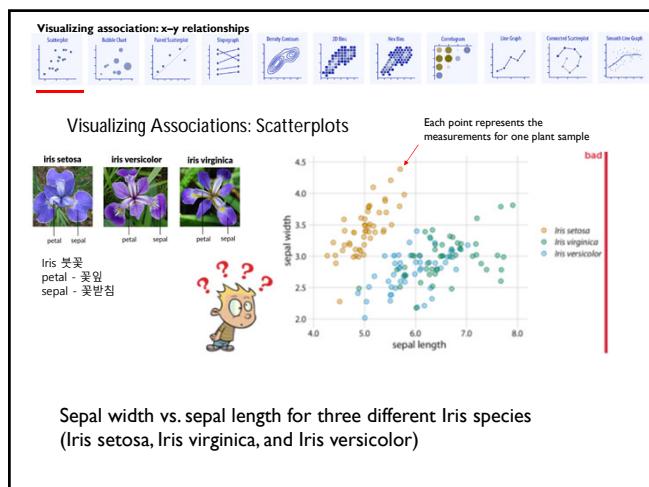
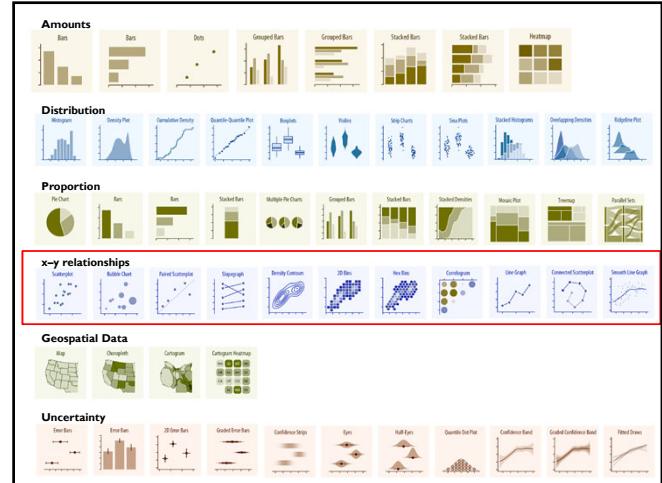
Tables

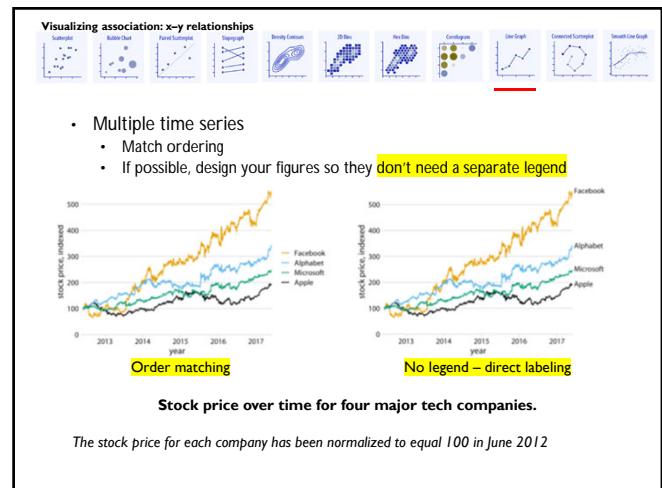
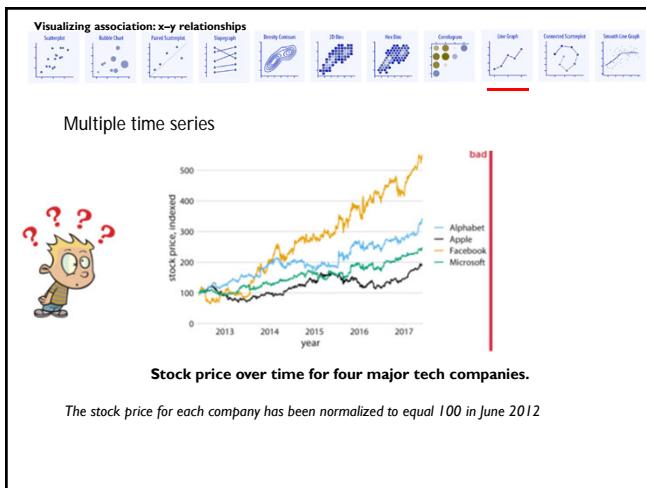
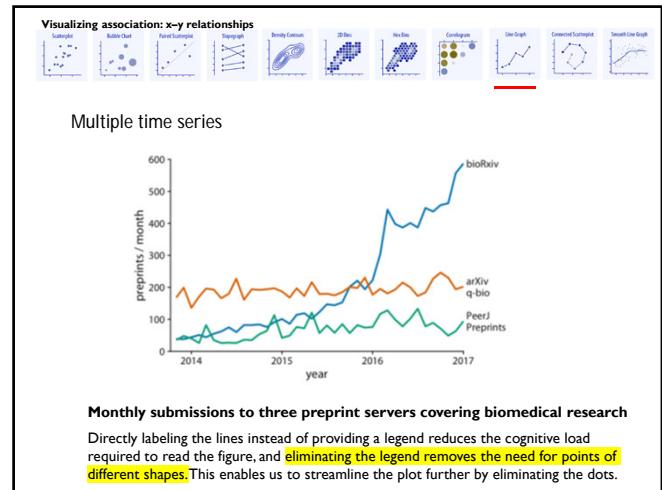
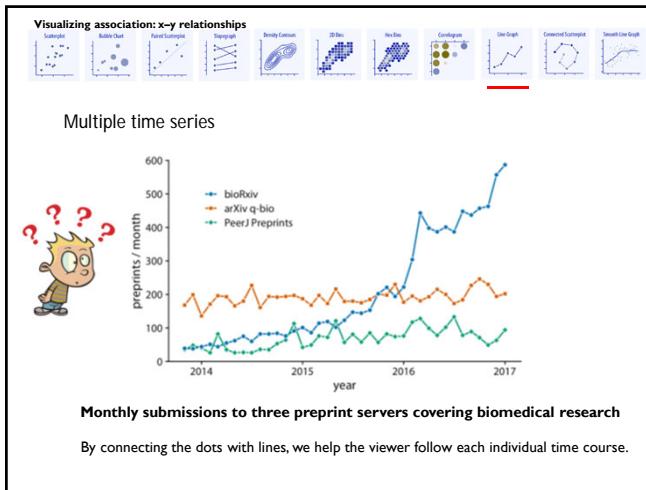
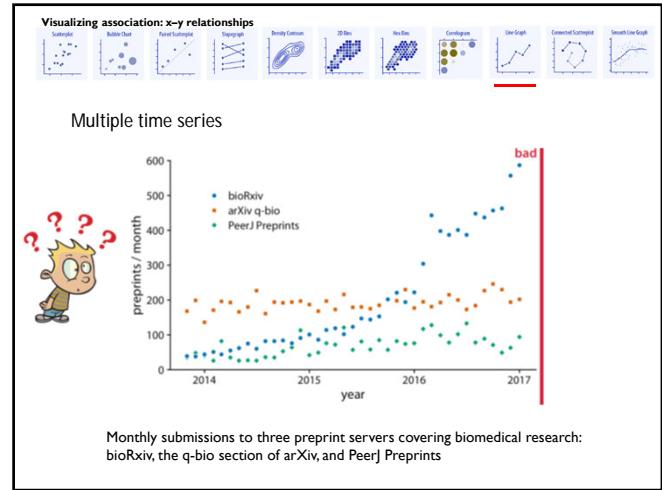
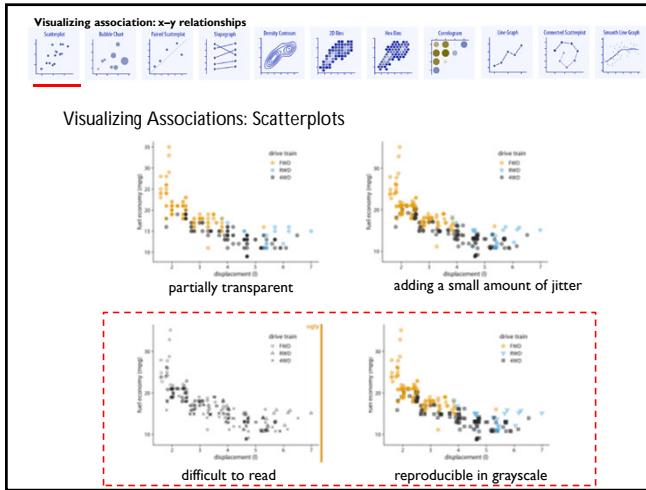
ugly		
Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

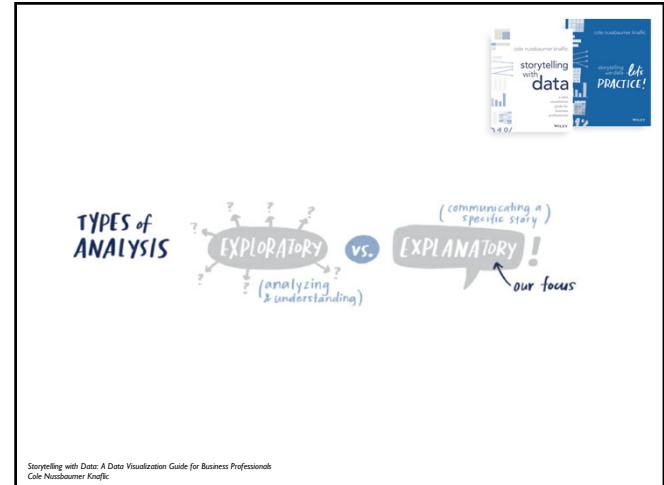
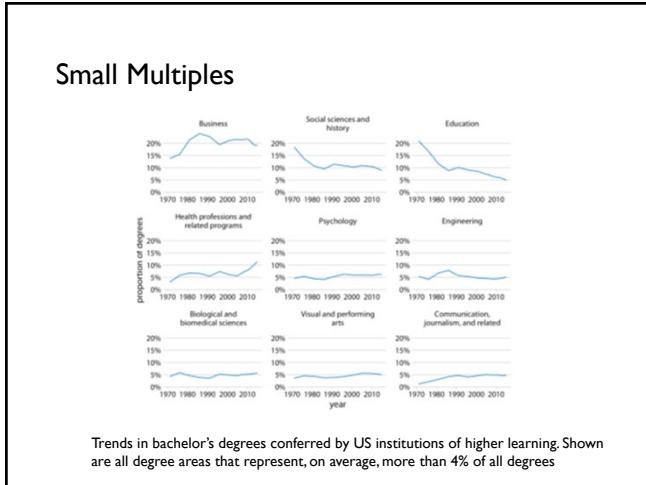
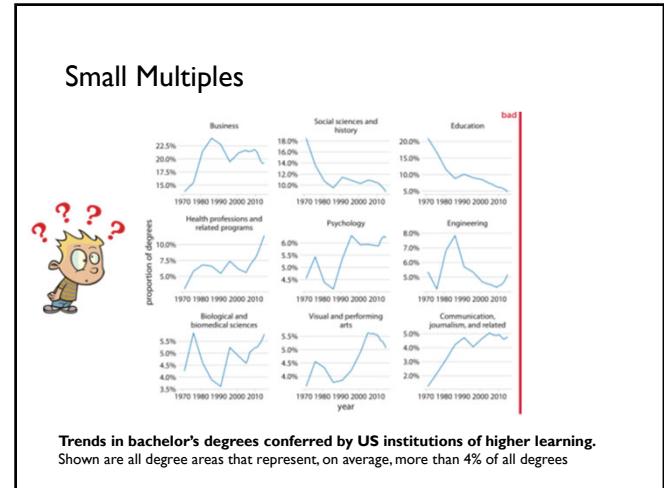
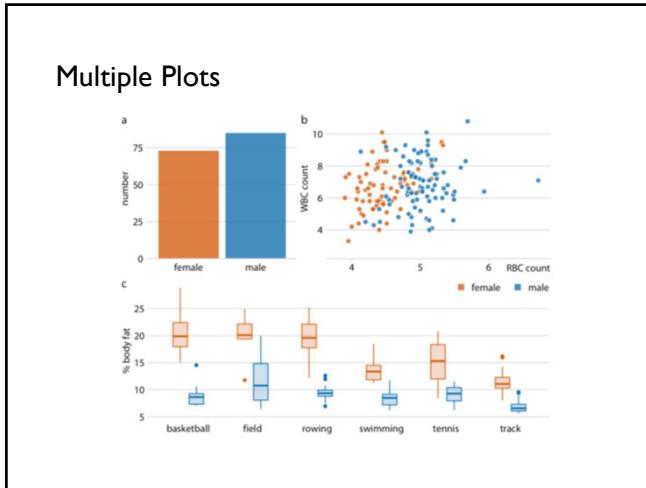
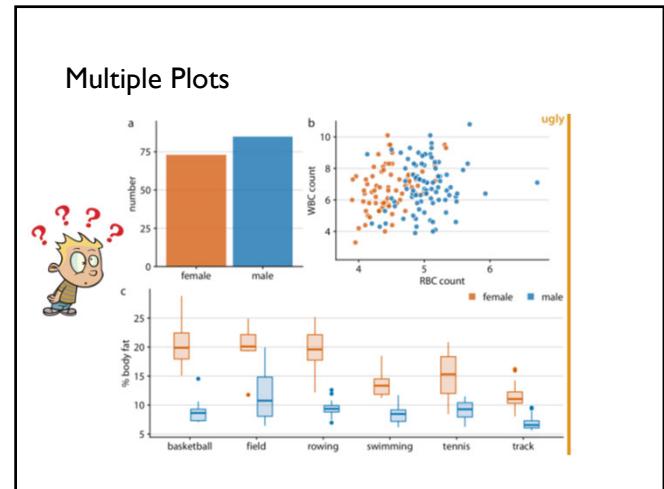
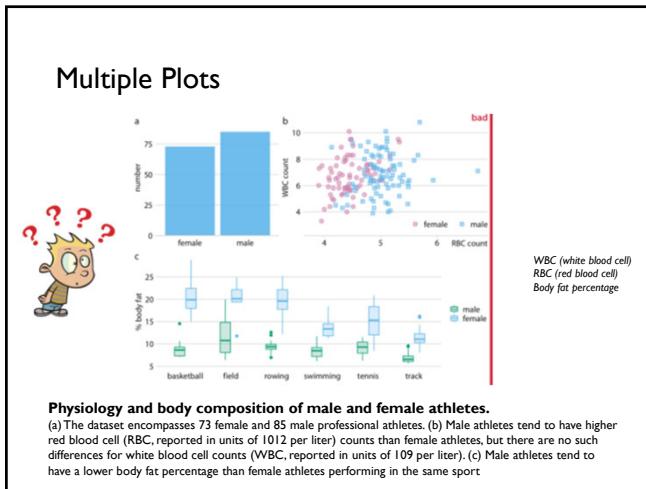
ugly		
Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

ugly		
Rank	Title	Amount
1	Star Wars: The Last Jedi	\$71,565,498
2	Jumanji: Welcome to the Jungle	\$36,169,328
3	Pitch Perfect 3	\$19,928,525
4	The Greatest Showman	\$8,805,843
5	Ferdinand	\$7,316,746

- Do not use vertical lines
- Do not use horizontal lines between data rows (Horizontal lines as a separator between the title row and the first data row or as a frame for the entire table are fine.)
- Text columns should be left aligned
- Number columns should be right aligned and should use the same number of decimal digits throughout
- Columns containing single characters should be centered
- The header fields should be aligned with their data; i.e., the heading for a text column will be left aligned and the heading for a number column will be right aligned







Explanatory Data Viz

"for communicating a specific story"

Understand the context

Choose an effective visual

Eliminate clutter

Focus attention

Tell a story

Make sure that you sketch charts before you use tools!
If possible, sketch multiple alternatives

Storytelling with Data: A Data Visualization Guide for Business Professionals
Cole Nussbaumer Knaflic, 2015

Don't Dump Data into Charts

Survey results: summer learning program on science

PRE: How do you feel about doing science?

Category	Percentage
Bored	5%
Not great	11%
OK	25%
Kind of interested	38%
Excited	19%

POST: How do you feel about doing science?

Category	Percentage
Bored	6%
Not great	12%
OK	14%
Kind of interested	30%
Excited	36%

A large red 'X' is drawn across both charts.

**Telling a story using an assertion statement to articulate your unique point of view and convey what's at stake.
(This applies to any kind of scientific presentation)**

assertion: Pilot program was a success

supporting visual evidence: Bar chart showing student feelings about science before and after the pilot program.

Category	Before (%)	After (%)
Bored	20%	10%
Not great	40%	20%
OK	30%	50%
Kind of interested	10%	30%
Excited	10%	30%

supporting textual summary: BEFORE program, the majority of children (40%) felt just OK about science. AFTER program, more children were Kind of interested (30%) & Excited (38%) about science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic, 2015

Tutorial on Data Visualization for Graduate Students

Uichin Lee, KAIST

Why Viz

- Goals: record, analyze, reveal patterns, communicate
- Viz = External cognition

Design Principles

- Integrity: avoid scale distortion & lie factor
- Viz design: max data-ink ratio, avoid chartjunk, increase data density
- Design principles (CRAP): contrast, repetition, alignment, proximity (grouping)

Charting 101

- Exploratory vs. Explanatory Viz
- Exploratory viz for scientific presentation: 1) understand the context 2) choose an effective visual 3) eliminate clutter 4) focus attention 5) tell a story
- Avoid common pitfalls in charting!