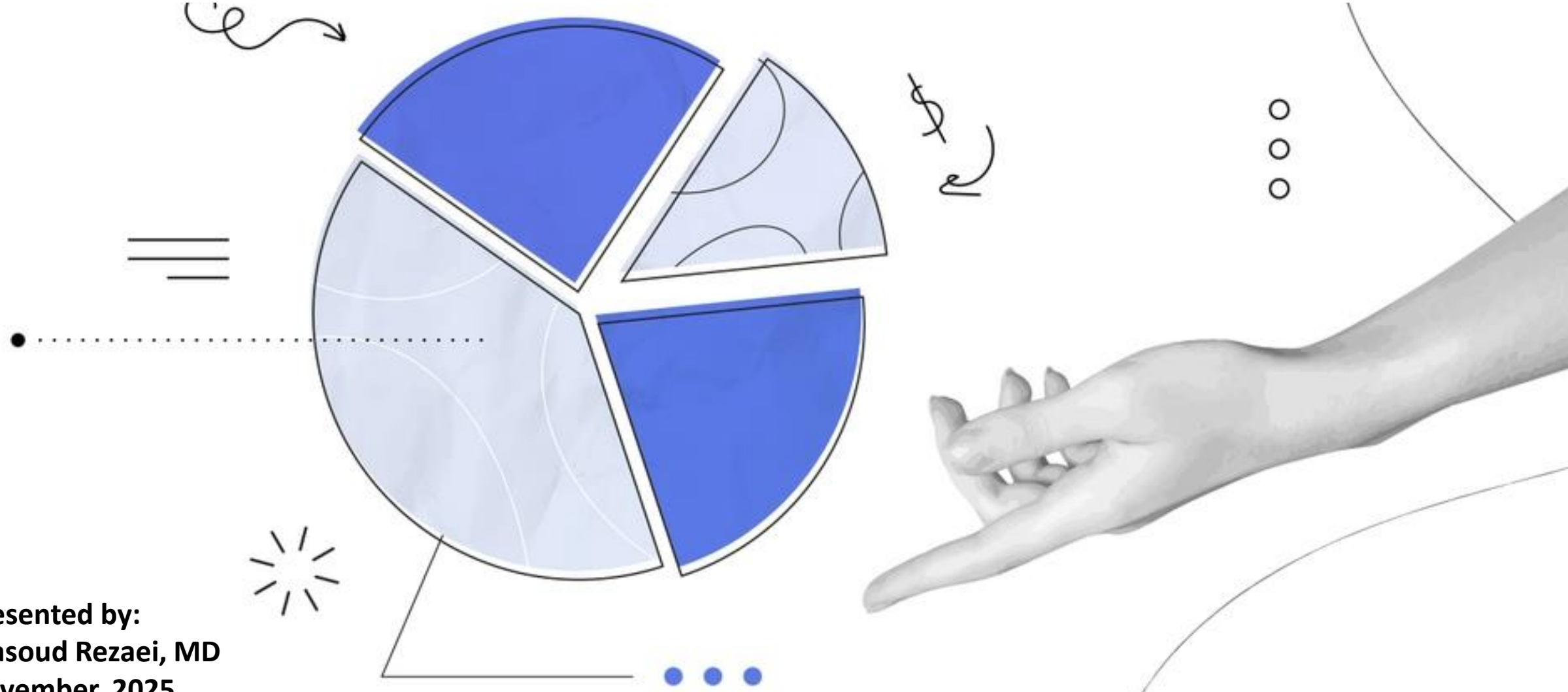
Effective Science Presentation Design



Presented by:
Masoud Rezaei, MD
November, 2025

Table of contents

Introduction

- Identifying the problem.
- Impact of research presentation design on career.
- A change in paradigm: introducing “vibe coding”.

Effective Scientific Figures and Tables

DOI: 10.1177/09760016251324320 ISSN: 0976-0016

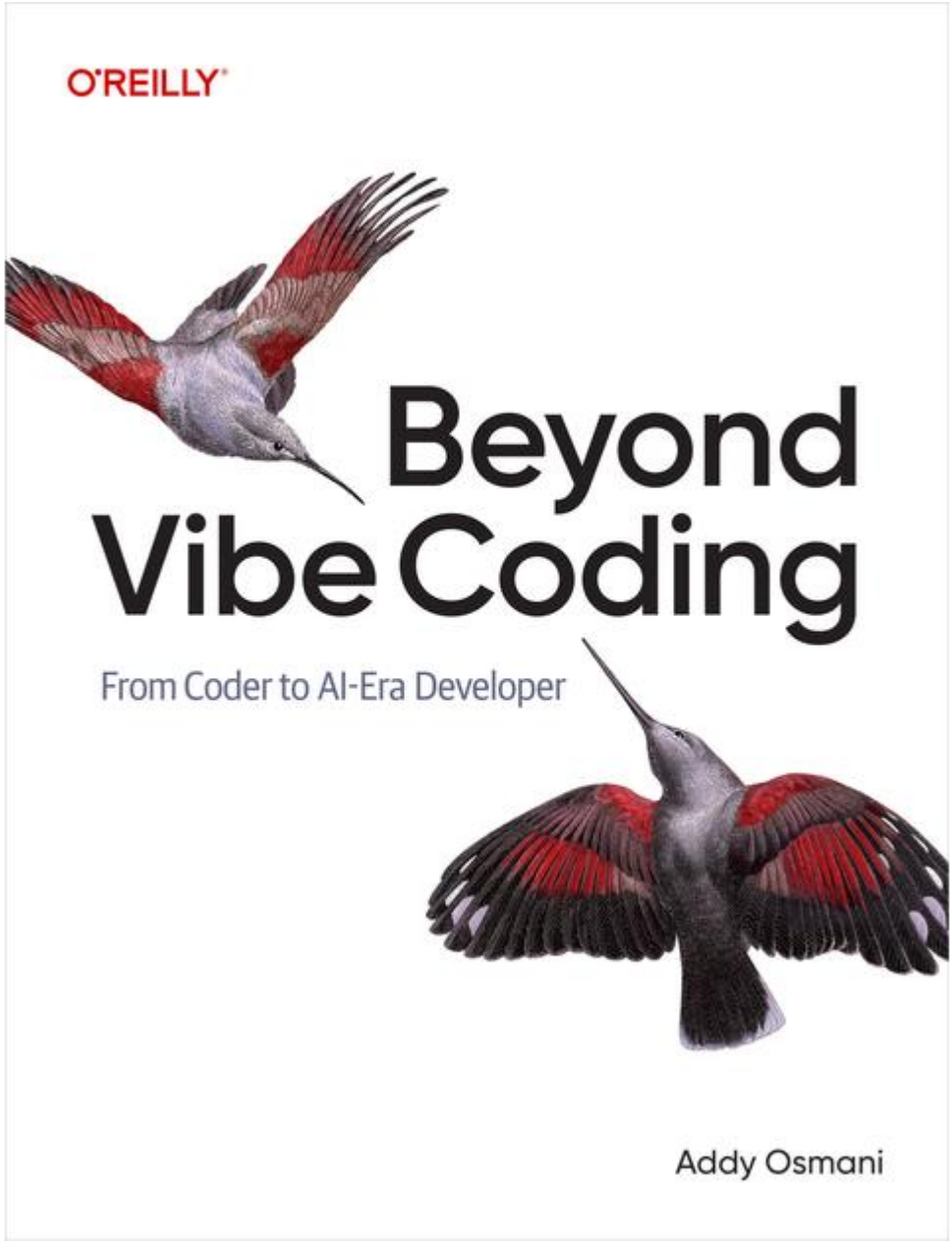
The Impact of Figures and Schematics on Reviewer Ratings and Manuscript Acceptance: A Cross-sectional Survey

Pranav Ajmera, Kapil Shirodkar, Raju Vaishya, Vijay Kumar Jain, Karthikeyan P. Iyengar, Rajesh Botchu

- The survey revealed that the majority of respondents accorded paramount importance to the inclusion of figures in manuscripts (91.7%), with original images (85.7%) and schematics (60.5%) being preferred.
- While the presence of schematics alongside original images was perceived to enhance manuscript acceptance (57.1%), the number of figures did not necessarily correlate with reviewer favourability (25.6% against).

Introduction of vibe coding

Using AI to generate functional code from natural language prompts, especially for those with limited programming experience.



Scientific Design Principles

Design is a plan for arranging elements in such a way as to accomplish a particular purpose.

Charles Eames
Designer and architect

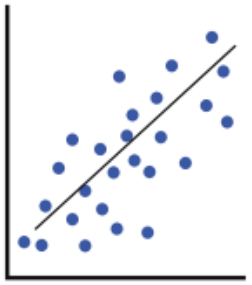
Designing Science Presentations

- Good scientific content does not speak for itself.
- good studies or ideas are rejected by scientific journals, granting agencies, or audiences.

**Great design will not sell an inferior product,
but it will enable a great product
to achieve its maximum potential.**

Thomas J. Watson Jr.
Design-focused former CEO of IBM

The elements of a science presentation



Story + Visual information + Delivery

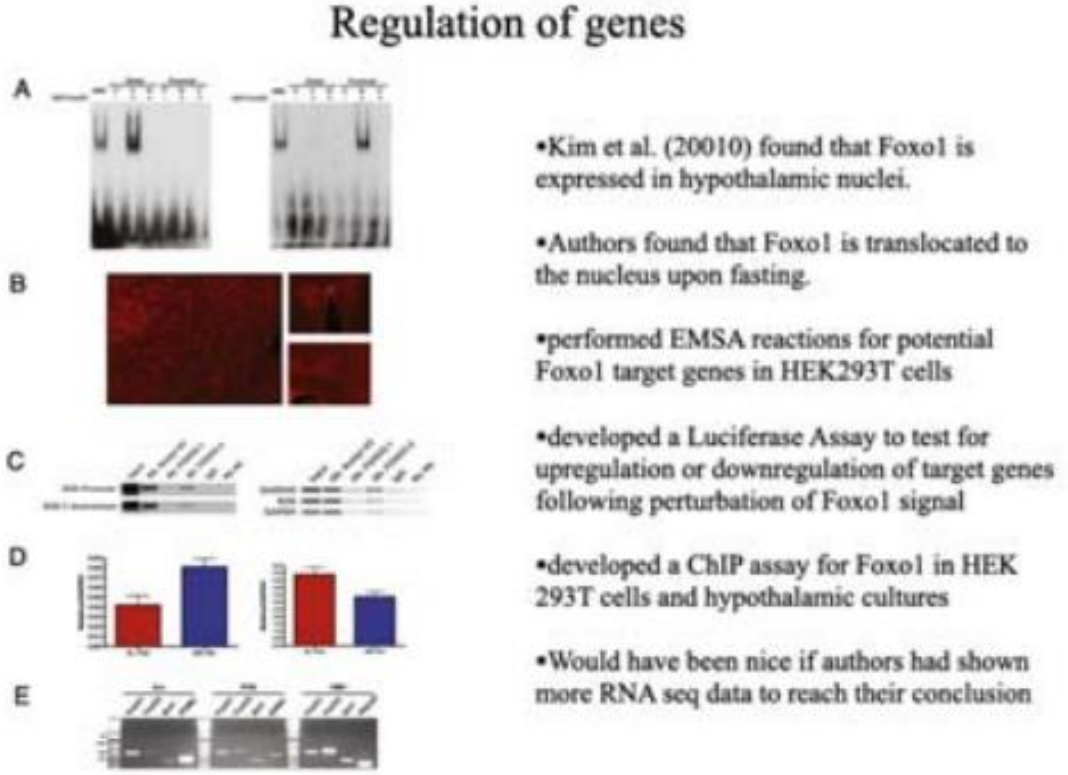
All of your ideas, experiments, results, and conclusions, arranged into a logical and compelling story. The structure and flow of your presentation.

All of the visual aids you use to communicate information. In science presentations, visual information typically consists of text, tables, graphs, photographs, and diagrams.

The narrative presentation of your visual information. In a paper, your narrative is written on the page. In a slide and poster presentation, you deliver your narrative orally and with your body language.

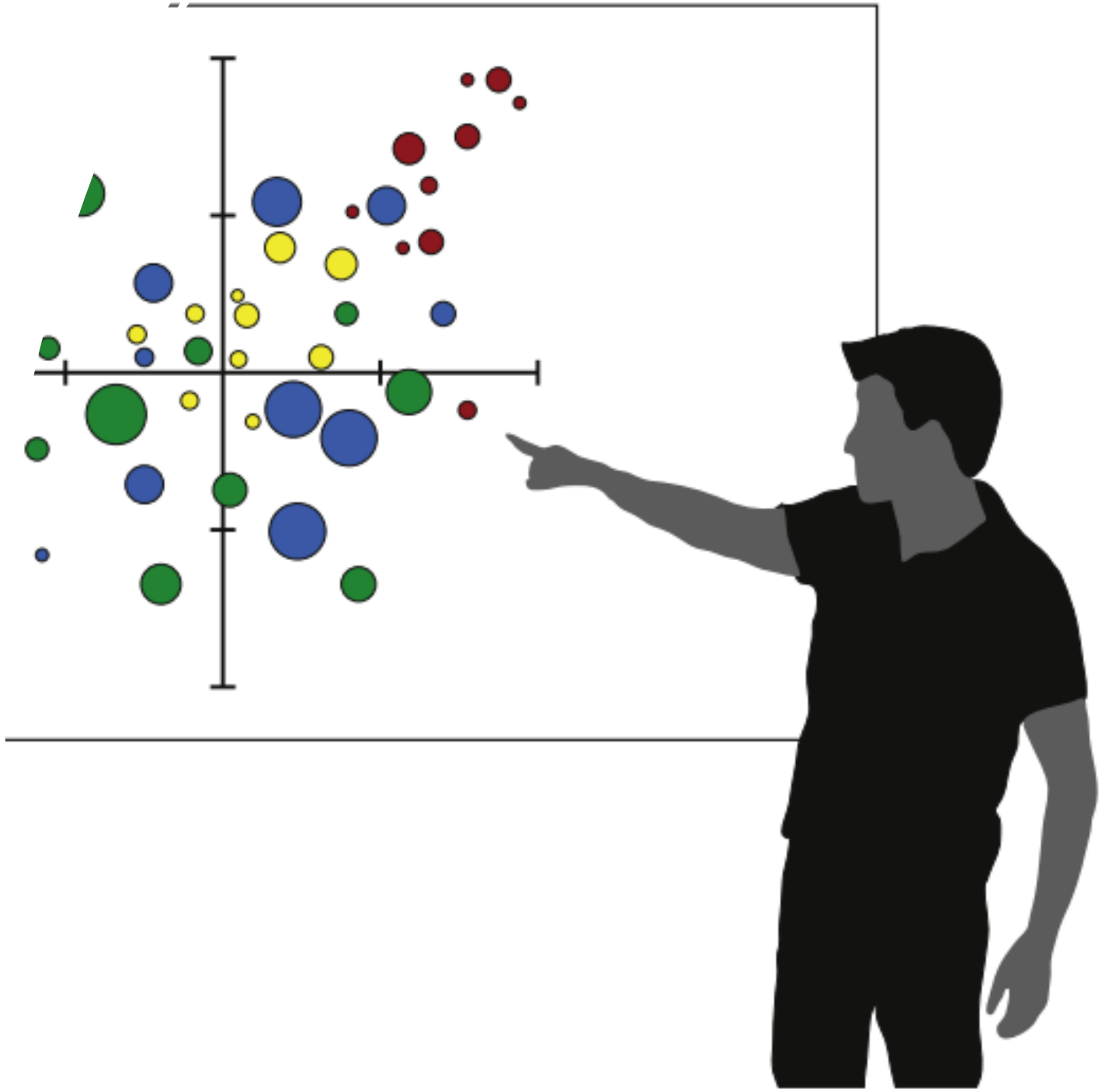
A bad example

- excellent scientific content, but …
- the slide title does not convey a conclusion or take-home point.
- too many figures.
- a lot of text.

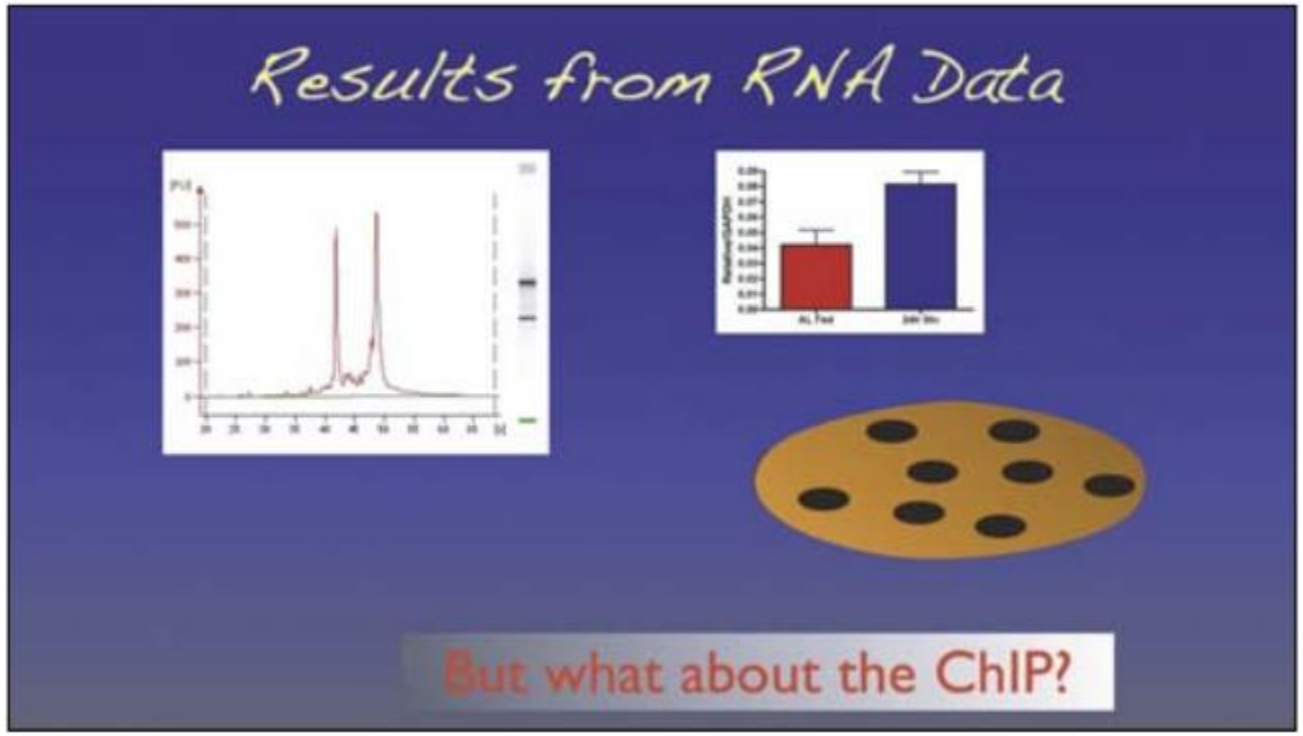


What is design?

- Design is ultimately about determining what impact you want to have on an audience and then establishing the best way to achieve that objective.
- One of the key tenets of good design is striving for simplicity.



Design is not decoration



This slide is poorly designed, not because it is ugly (and many would consider it ugly!), but because it hasn't been designed to optimally communicate information with an audience. Instead of emphasizing the data and a larger conclusion, this slide emphasizes stylized font choices, distracting color patterns, and unnecessary clip art. This slide was *decorated* more than it was designed to communicate meaningful information.

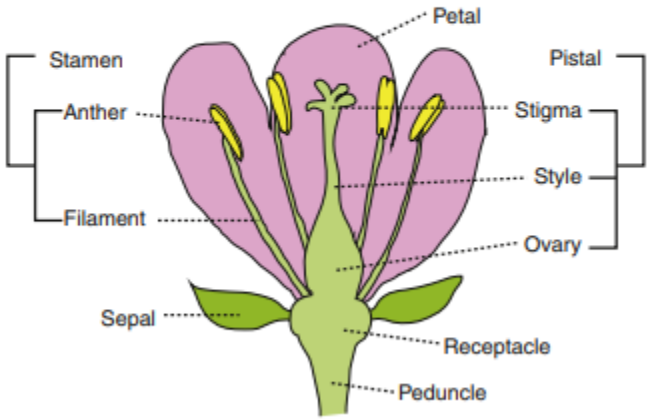
Design is not about what something looks like.
Design is about how it works.

Steve Jobs
Design-focused former CEO of Apple

What simplicity means…

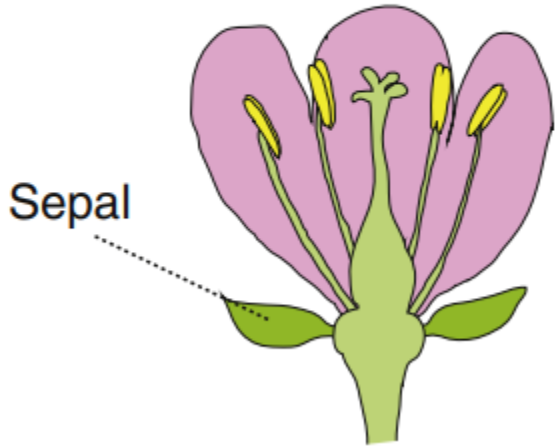
Before

The sepals support the flower petals



After

The sepals support the flower petals



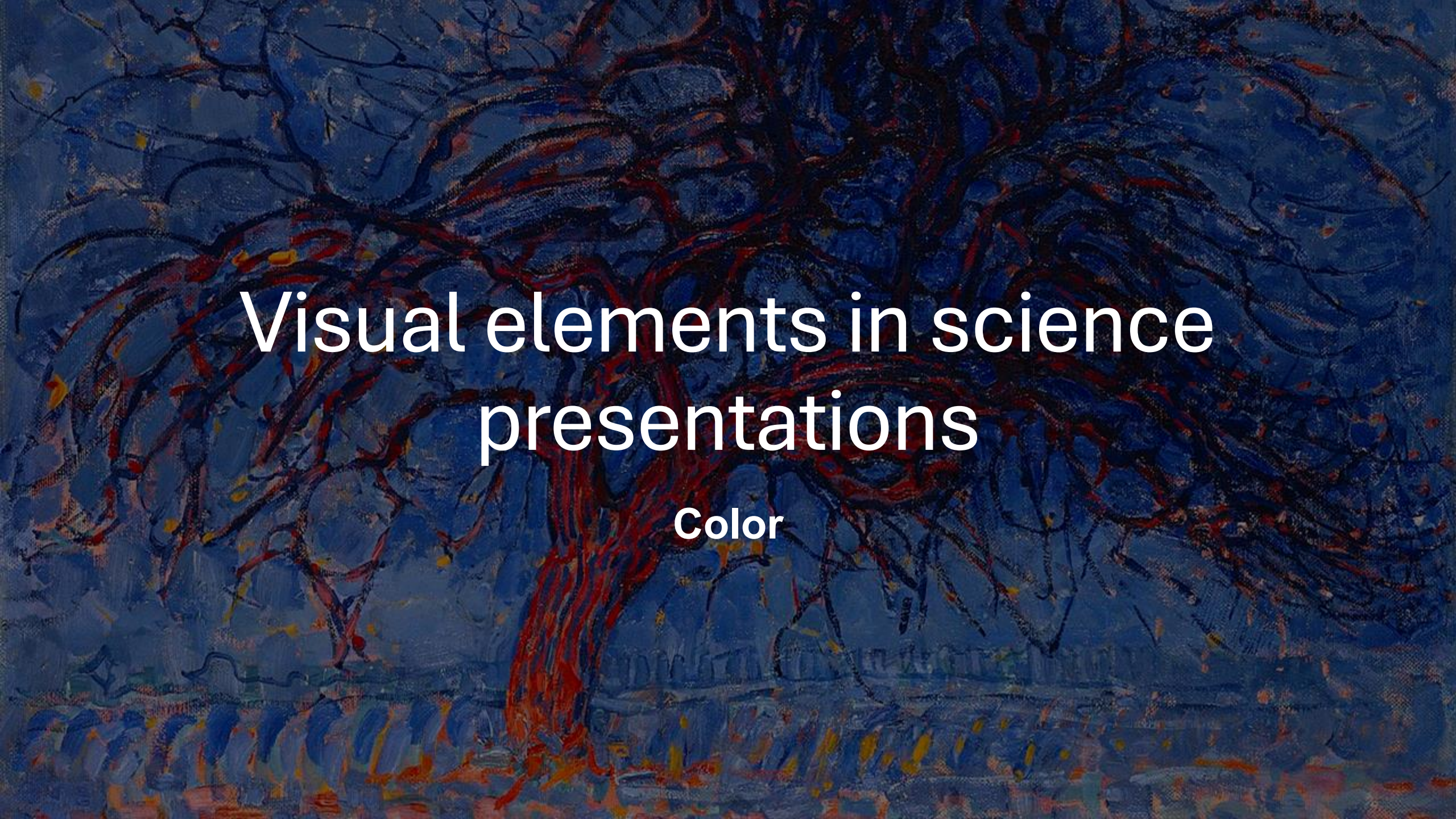
Increasing the simplicity of a presentation isn't the same as "dumming it down." It is about taking away unnecessary distractions or superfluous details so that the audience can focus on what is most important.

Design is ultimately about the audience

- Who are they?

Appreciate good presentation design from other scientists

- Keep a file of research articles that are well written or that contain well-designed figures, regardless of their content.
- In your institution, learn who gives the best talks and always try to attend their presentations.

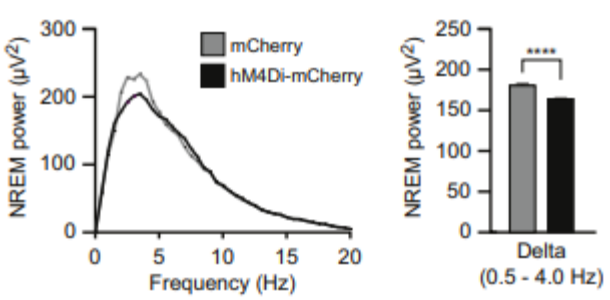


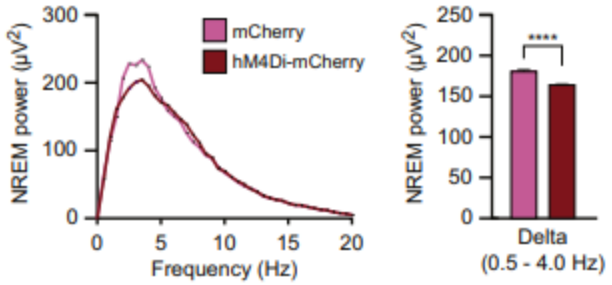
Visual elements in science presentations

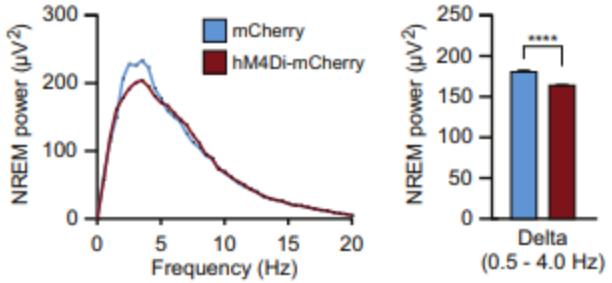
Color

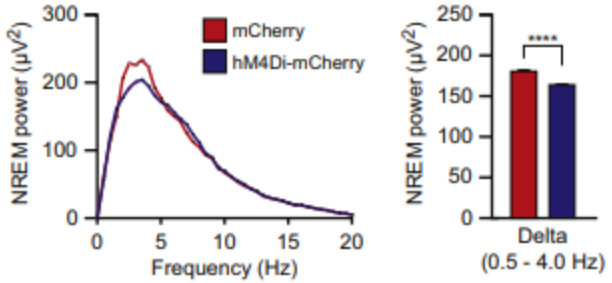
Color functions

- to highlight information.
- to enhance a message.
- to convey an atmosphere or emotion.







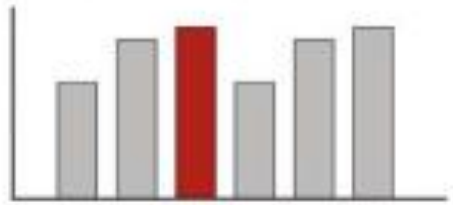


Don't let your colors distract from your message....

Instead, use color to communicate

Don't let your colors overwhelm your data

Instead, use color to emphasize your data



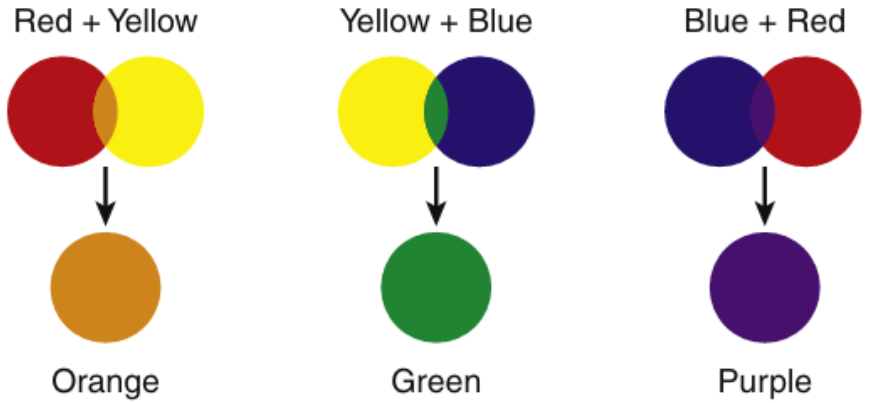
Color is distracting when it is used to decorate

Color is helpful when it is used to design

Terminology of Color

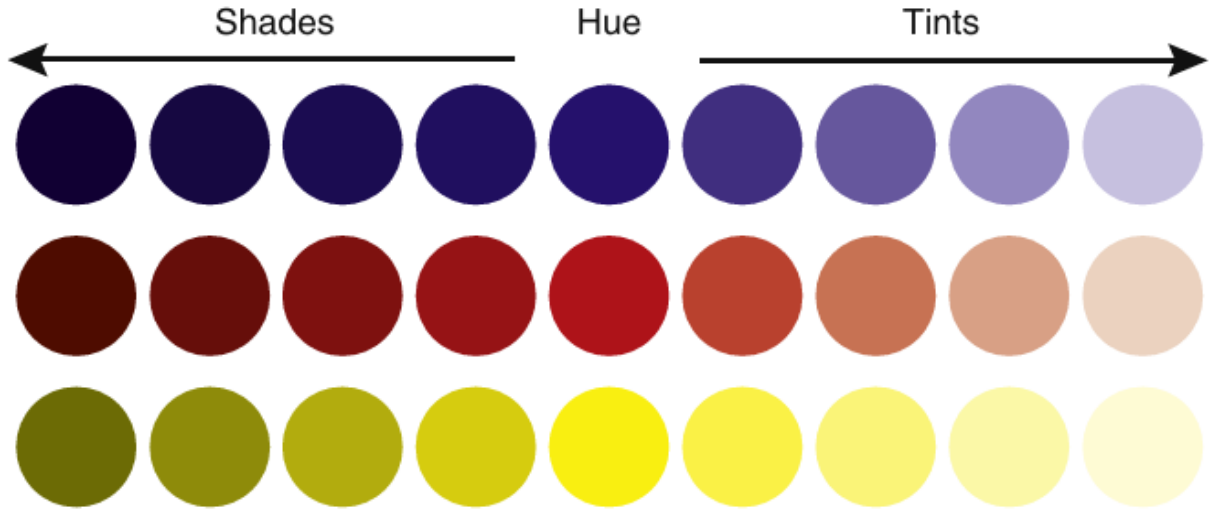
- Hue: a color in its purest, most fundamental form.

- primary colors: red, yellow, and blue.
- Secondary colors: results from the mixing of two of the primary colors.



Terminology of Color

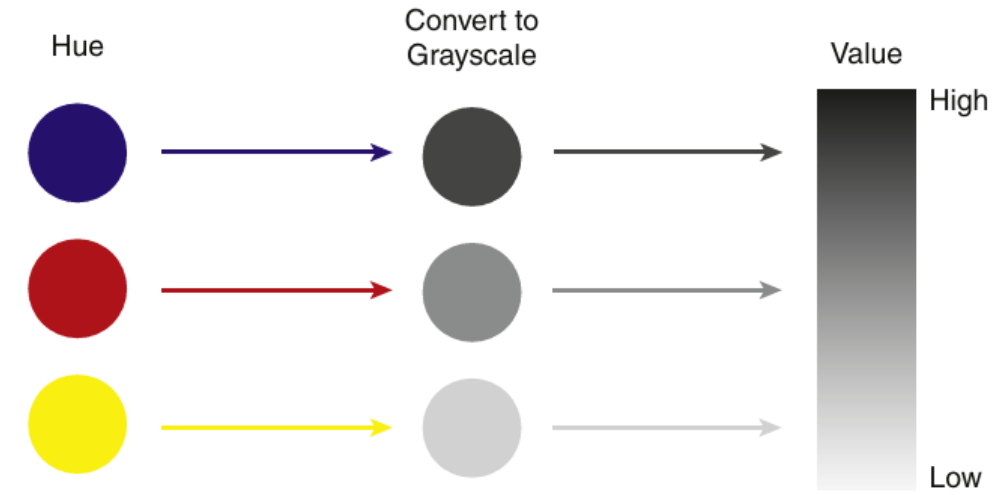
- **Intermediate colors**: result from mixing a primary and secondary color, or multiple secondary colors.
- **Shade** is the amount of black added to a hue.
- **Tint** is the amount of white added to a hue.

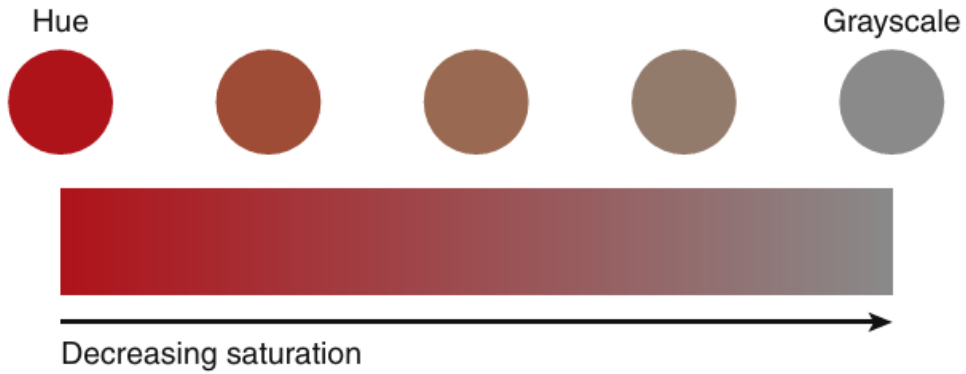


Terminology of Color

- **Value (or intensity):** refers to the inherent lightness or darkness of a color.

- **Saturation:** refers to the degree of hue in a color.





How computers specify color

RGB system

Red = RGB (255,0,0)

Blue = RGB (0,0,255)

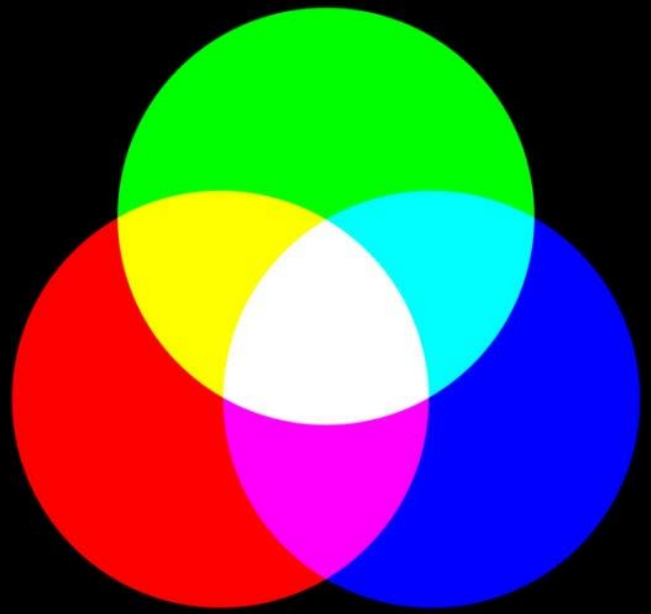
How printers specify color

- Computers specify color in different formats.
- **CMYK system** stands for cyan (C), magenta (M), yellow (Y), and black (K) inks.

Red = CMYK (0,100,100,0)

Blue = CMYK (100,100,0,0)

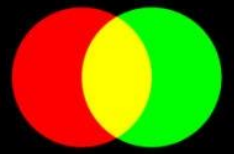
FOR SCREENS



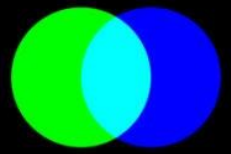
RGB

ADDITIVE COLOR SYNTHESIS

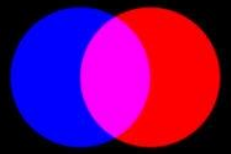
RED • GREEN • BLUE



RED + GREEN

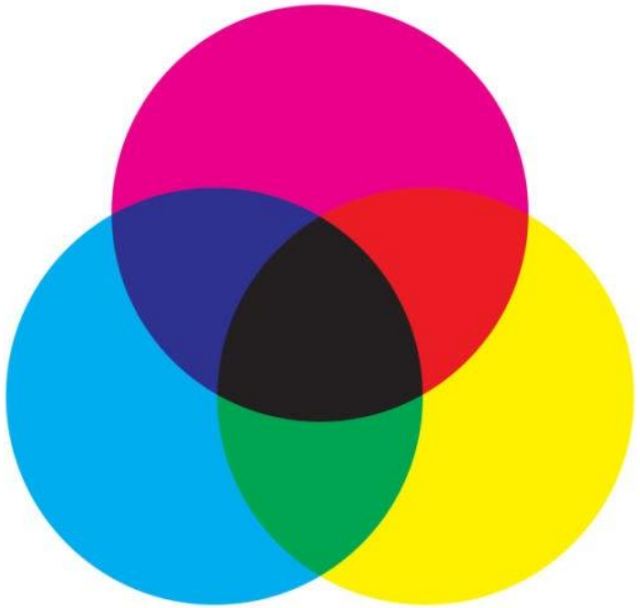


GREEN + BLUE



BLUE + RED

FOR PRINT



CMYK

SUBTRACTIVE COLOR SYNTHESIS

CYAN • MAGENTA • YELLOW • BLACK

CYAN + MAGENTA

MAGENTA + YELLOW

YELLOW + CYAN

How computers specify color

- **Hexvalue** also use combinations of red, green, and blue, but utilize a six-digit number instead of the RGB specification system. Colors are specified in the format "#RRGGBB".
- The degree of each color ranges from #00 to #FF.

Red = ##(FF0000)

Blue = ##(0000FF)

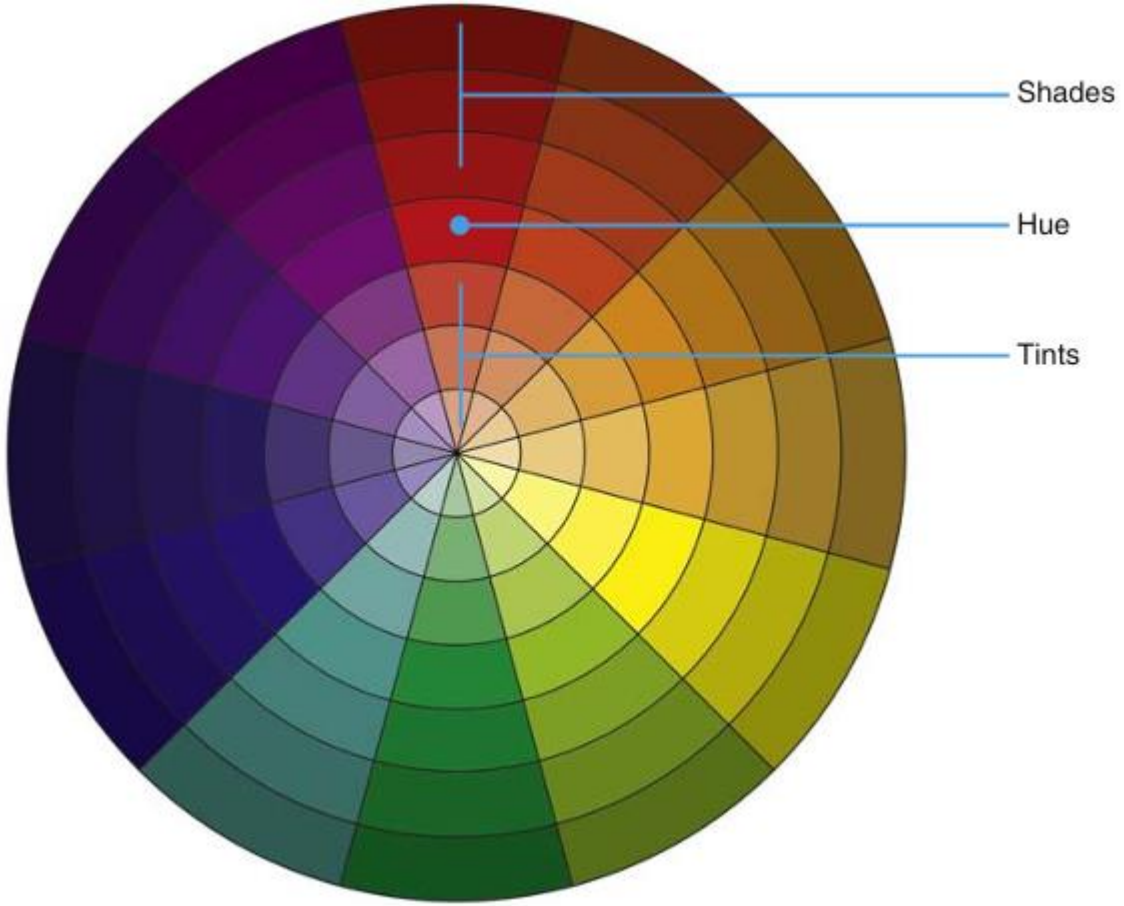
- 16-base Hex digits: 0 1 2 3 4 5 6 7 8 9 A B C D E F.

Color Wheel

- A color wheel can be helpful in selecting colors

Microsoft Color Picker

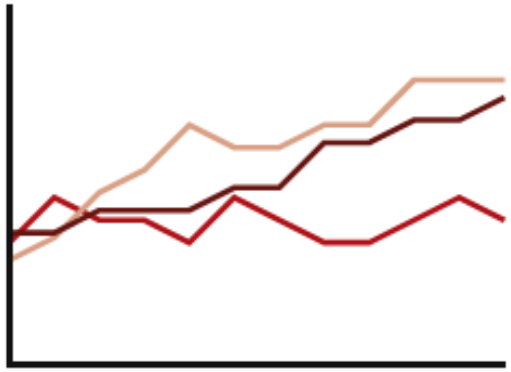
Apple Color Picker

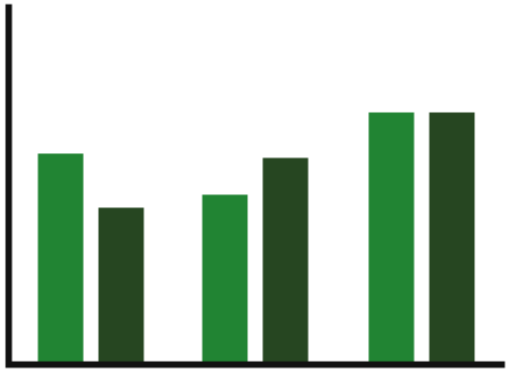


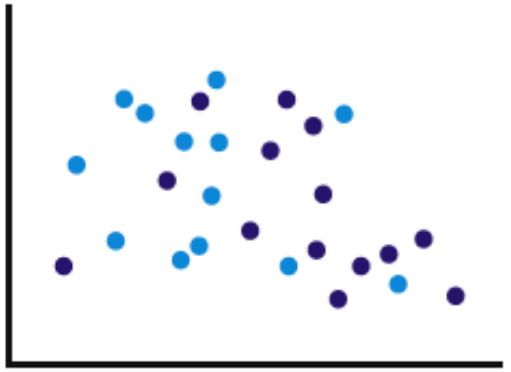
Color choosing strategies

Monochromatic

Only one hue in various shades or tints. The advantage to this strategy is that it creates a consistent, unified look. Even though different datasets may be categorized differently (e.g., different bars on a bar graph), they seem representative of a larger, uniform category.



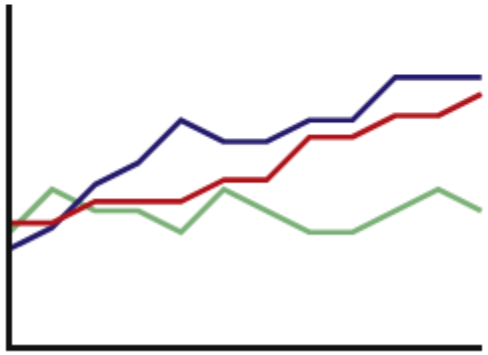


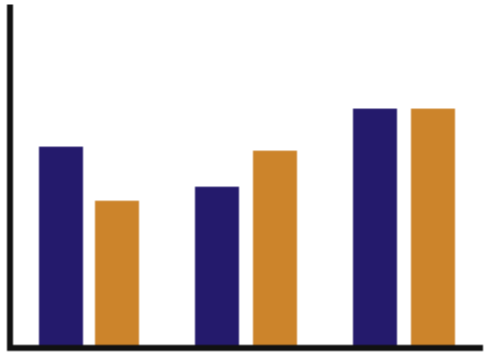


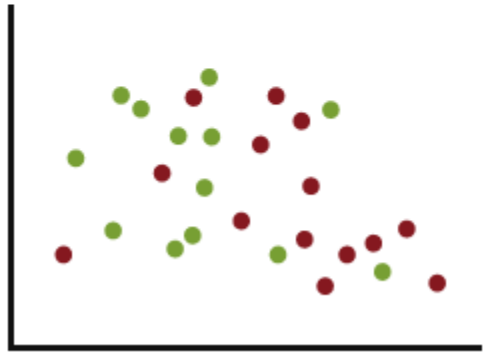
Color choosing strategies

Complementary

Two or three hues on opposite sides of a color wheel. This strategy enhances the difference between categories of data and makes them seem more like opposites.



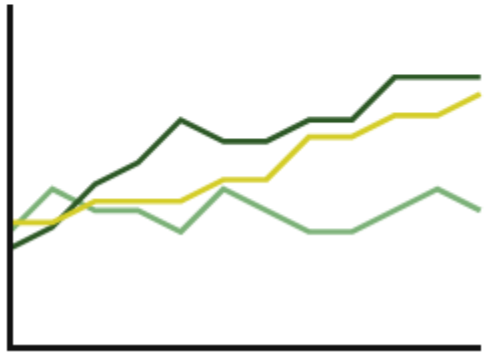


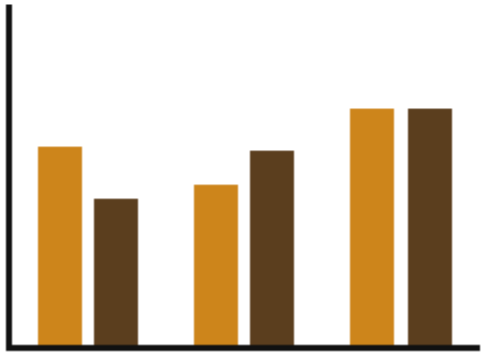


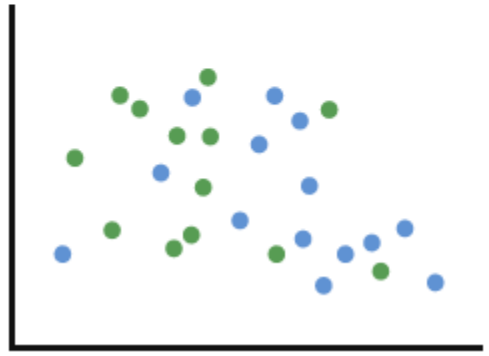
Color choosing strategies

Analogous

Two or three hues that are relatively adjacent on the color wheel. This strategy combines elements of both the monochromatic and complementary strategies, using multiple colors while also achieving a consistent, harmonious look.



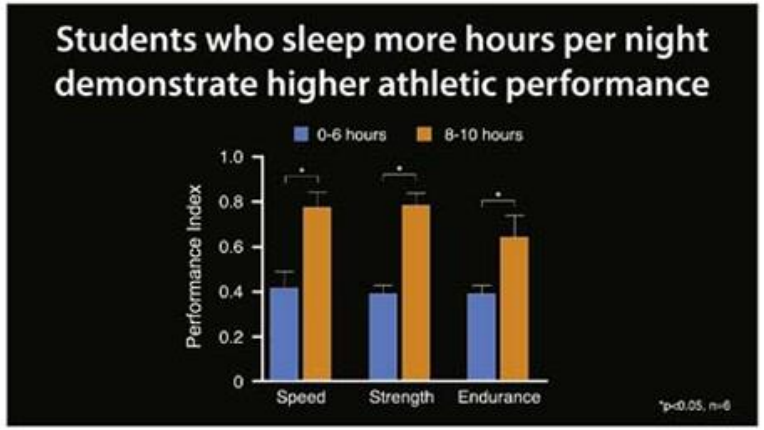




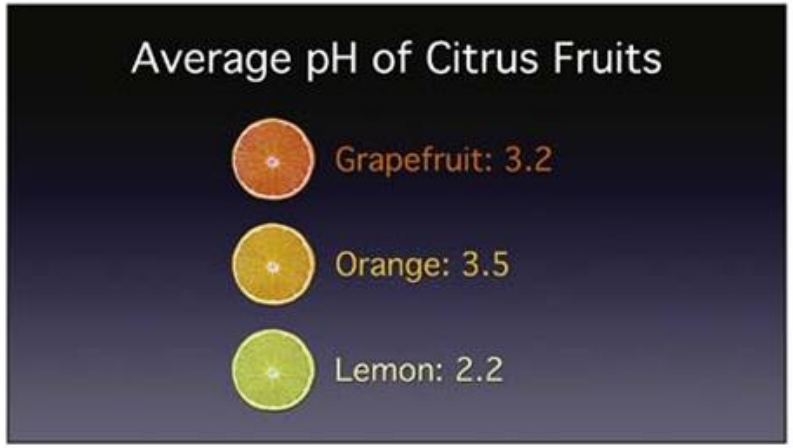
Examples

Monochromatic

Complementary

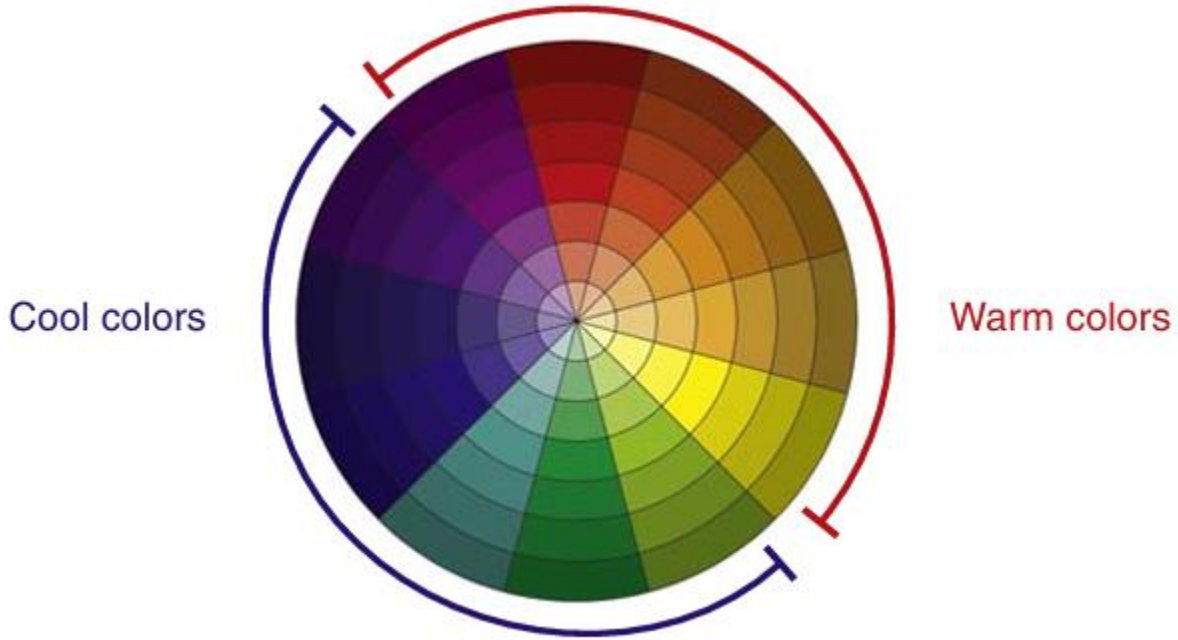


Analogous



Color Temperature

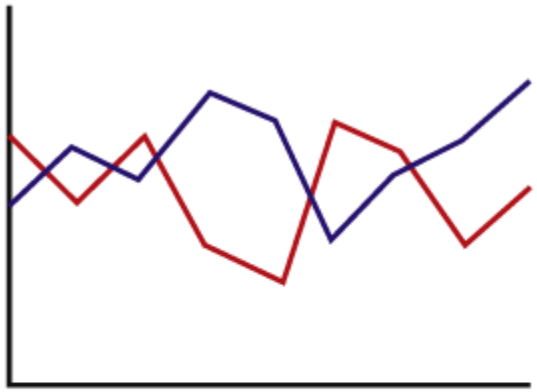
- Colors on opposite sides of the color wheel are often described as “warm” or “cool.”
- Warm colors are associated with energy, vitality, excitement, and fun.
- Cool colors are associated with peace, serenity, and nature.



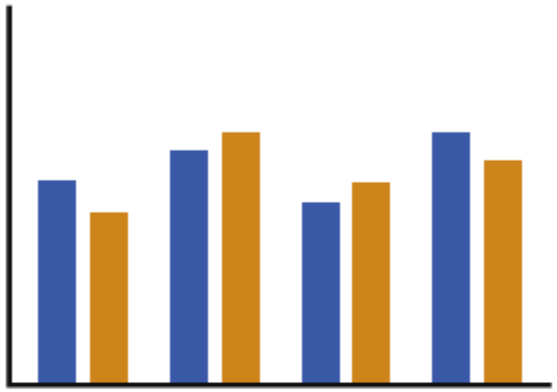
Color Temperature

- Hint: Humans are hard-wired to attend to warm colors more than cool colors.
- we perceive warm colors as being in the foreground and cool colors as being in the background.
- Therefore, choose warm colors to highlight the data that you really want to emphasize.

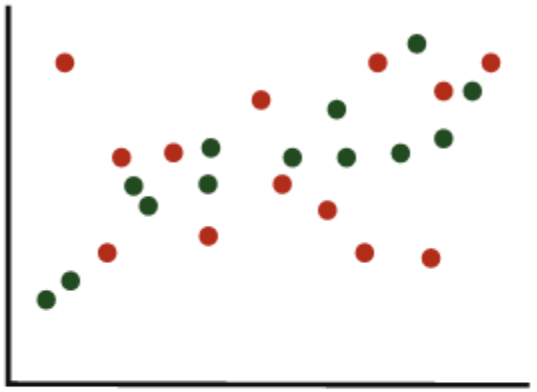
Examples



The red line appears in the foreground even though it is technically placed behind the blue line



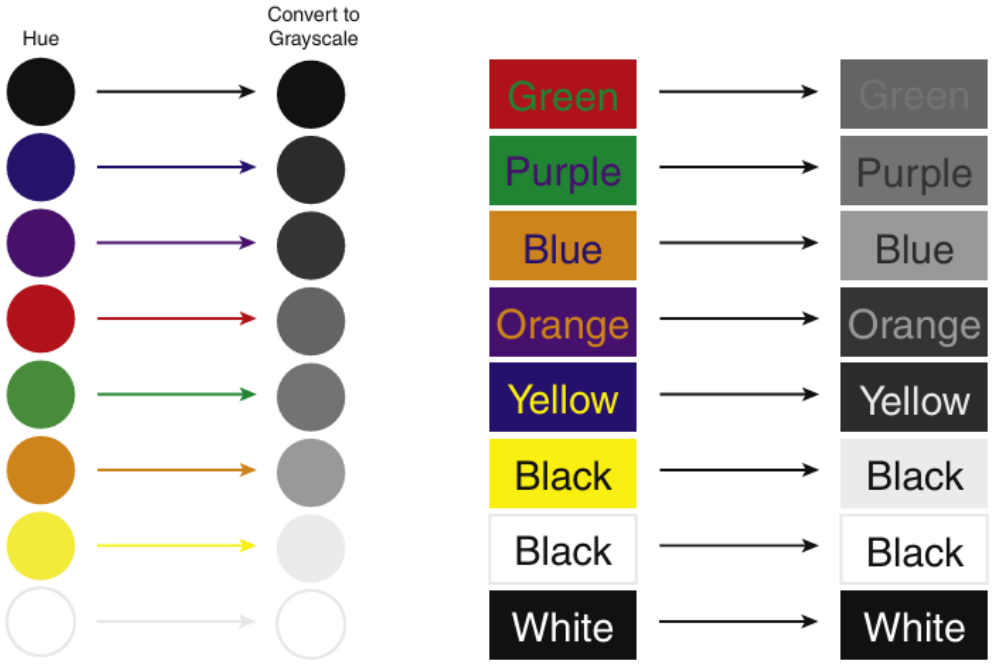
The orange bars attract attention much more than the blue bars



The red data points stand out even though the green data points show more of a trend

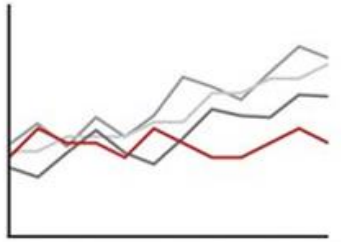
Choosing Proper Back/Foreground Colors

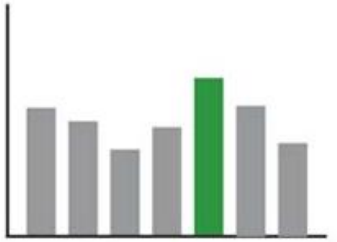
- The best foreground and background combinations are black on a white background or white on a black background (optimal contrast).
- If using color, a warm color foreground usually stands out on a darker cool color background.



Use color to highlight salient information

- For color to be effective in attracting an audience's attention, it should ideally be used in isolation.





	X	Y	Z
A	15.4	12.3	11.1
B	14.8	15.8	19.9
C	10.4	10.6	14.7
D	10.9	41.2	14.1
E	14.2	16.3	12.1

ATTTGACGATGAGCGCTAGCATGGACCGAT
TAAACTGCTACTCGCGATCGTACCTGGCTA

Emotional associations of different colors

- As a further reading for those who are interested.

White: purity, simplicity, innocence, clean, spacious, milk, cotton, clouds

Red: love, hate, passion, hot, stop signal, blood, berries, heart

Yellow: light, cheerful, sunny, optimistic, summer, dry, wheatfield, cornfield

Green: natural, environment, healthy, go signal, grass, vegetables, trees

Orange: autumn, fruity, fun, sporty, pumpkin, Halloween, caution sign, oranges

Blue: peaceful, natural, tranquil, calm, positive, melancholy, cold, sky, air, water, ocean, ice

Brown: rustic, earthy, woody, cozy, dirt, wilderness, cabin, outdoors

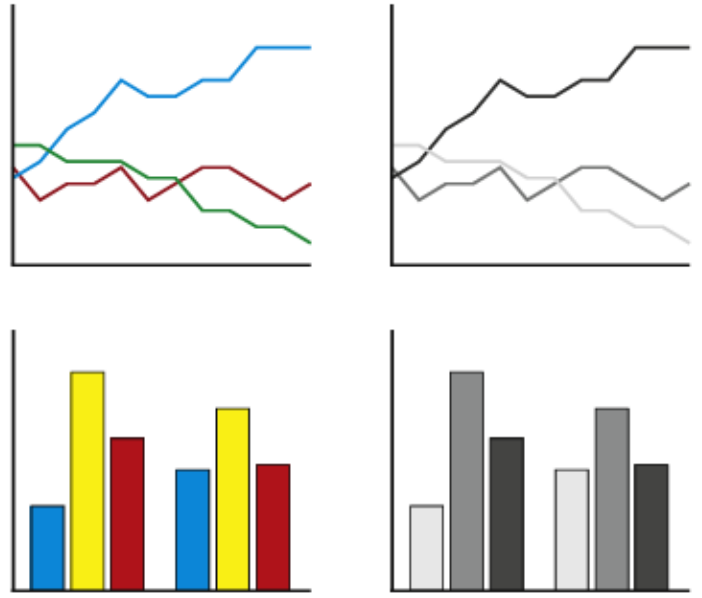
Purple: exotic, creative, sweet, artistic, flowers, candy

Pink: soft, delicate, young, sweet, feminine, flowers, baby, candy

Black: powerful, formal, corporate, classy, night, suit, briefcase, judge

Black and white are colors, too

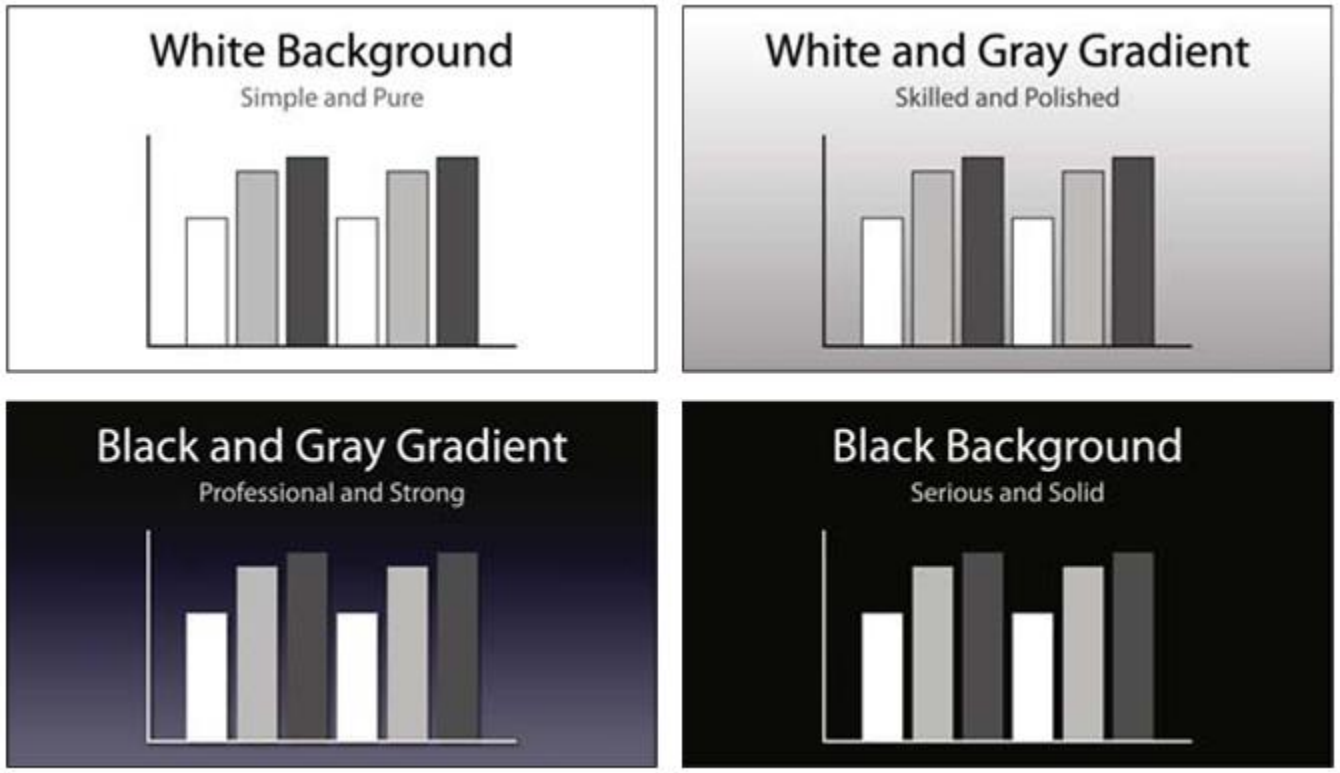
- Some journals charge high fees for color figures that could easily be made in black and white.
- Well-chosen **shades of gray** can be used in place of multiple colors.



Is the color in these figures really necessary?

Replacing the color with shades of gray reduces printing costs and improves the contrast between data sets.

Examples


White Background
Simple and Pure
White and Gray Gradient
Skilled and Polished
Black and Gray Gradient
Professional and Strong
Black Background
Serious and Solid

Typography

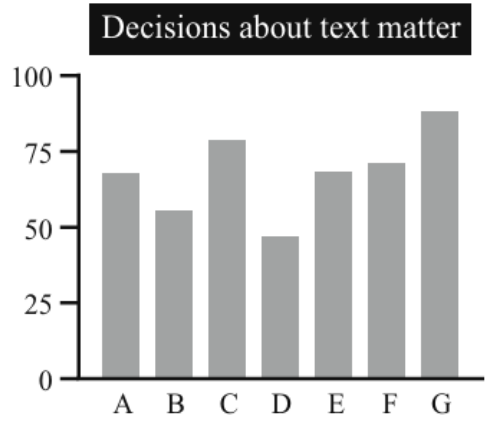
Decisions about text matter.

Before

DECISIONS ABOUT TEXT MATTER!

Decisions about text matter

- They affect the legibility, tone, and professionalism of your presentation

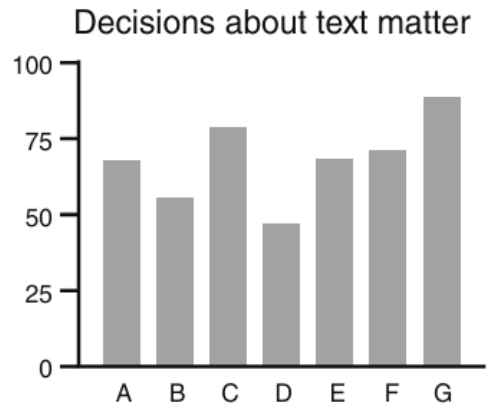


After

Decisions about text *matter!*

Decisions about text matter

- Affect legibility
- Affect tone
- Affect professionalism



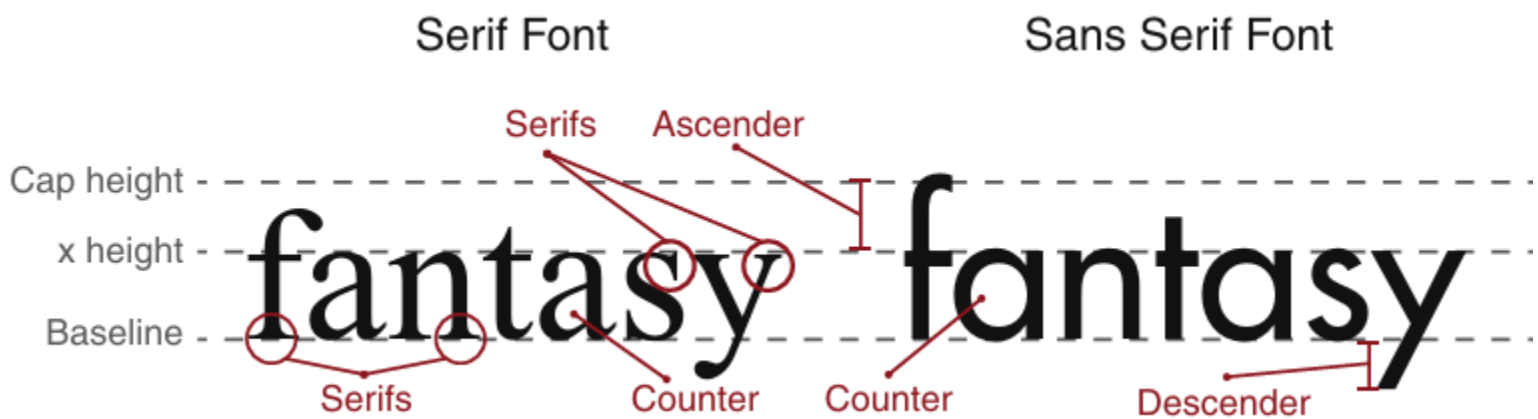
Characteristics of a font

- There are two families of fonts: serif vs sans serif.
- **Serif fonts** are good for smaller character sizes (10-14 pts) in multiple lines of type.
- **Sans serif** fonts are usually perceived as most simple and pure, and best for slide and poster presentations.

Font	Personality
Garamond	classic, refined
Georgia	elegant, mature
Times New Roman	professional, traditional

Font	Personality
Calibri	formal, neutral
Century Gothic	grand, optimistic
Helvetica	simple, pure, contemporary

Example


Serif Font
Sans Serif Font
Serifs
Ascender
Cap height
x height
Baseline
fantasy
fantasy
Serifs
Counter
Counter
Descender

less standard/routine sans serif fonts

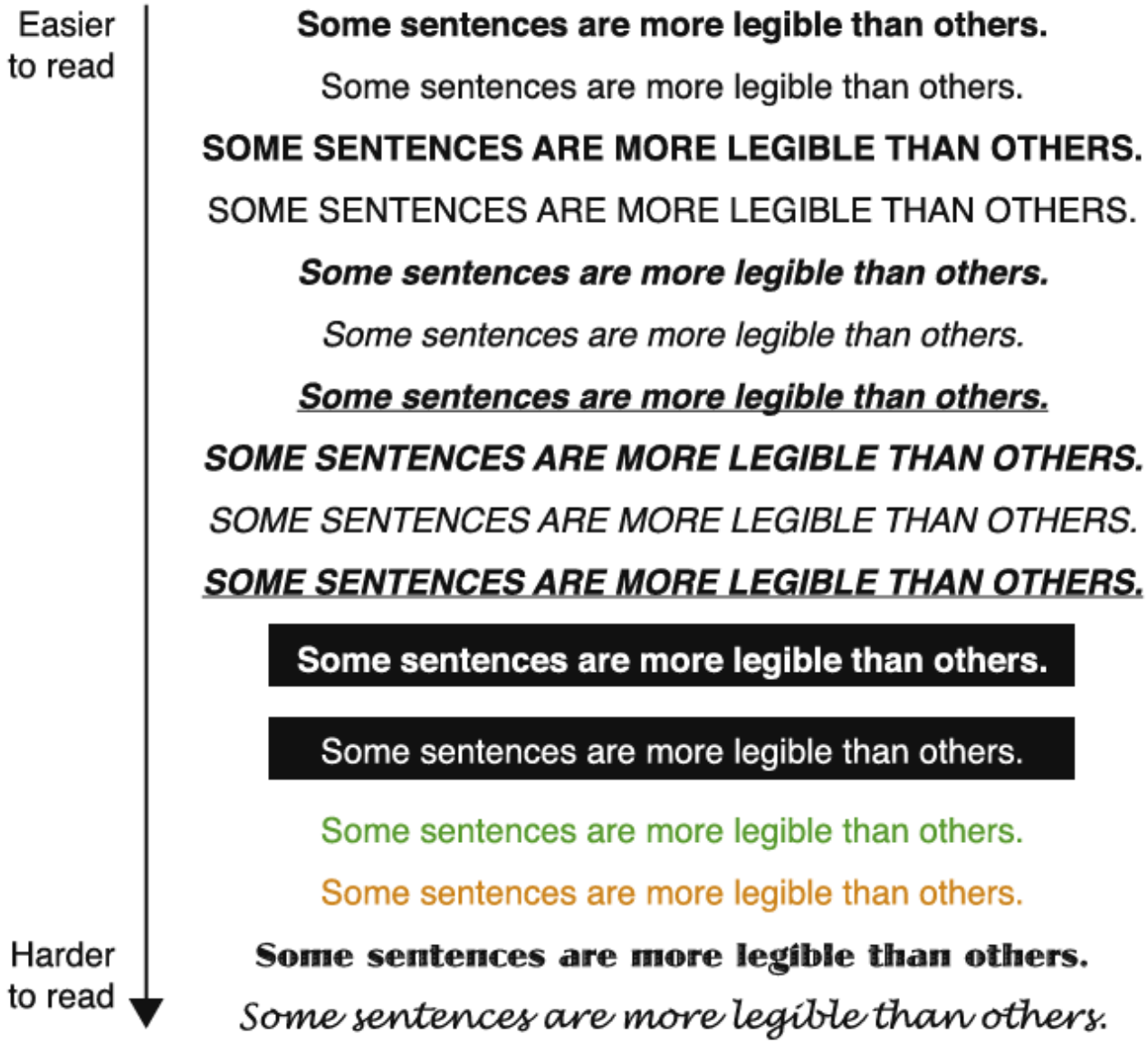
Font	Personality
Futura	fun, utopia
Gill Sans	warm, friendly
Myriad Pro	jovial, friendly, casual
Chalkboard	informal, fun
Comic Sans	silly, fun
Marker Felt	**informal, creative**

Non-proportional (“monospaced”) fonts

- they are great for writing letters in a sequence, such as sequences of DNA, amino acids, or computer code.

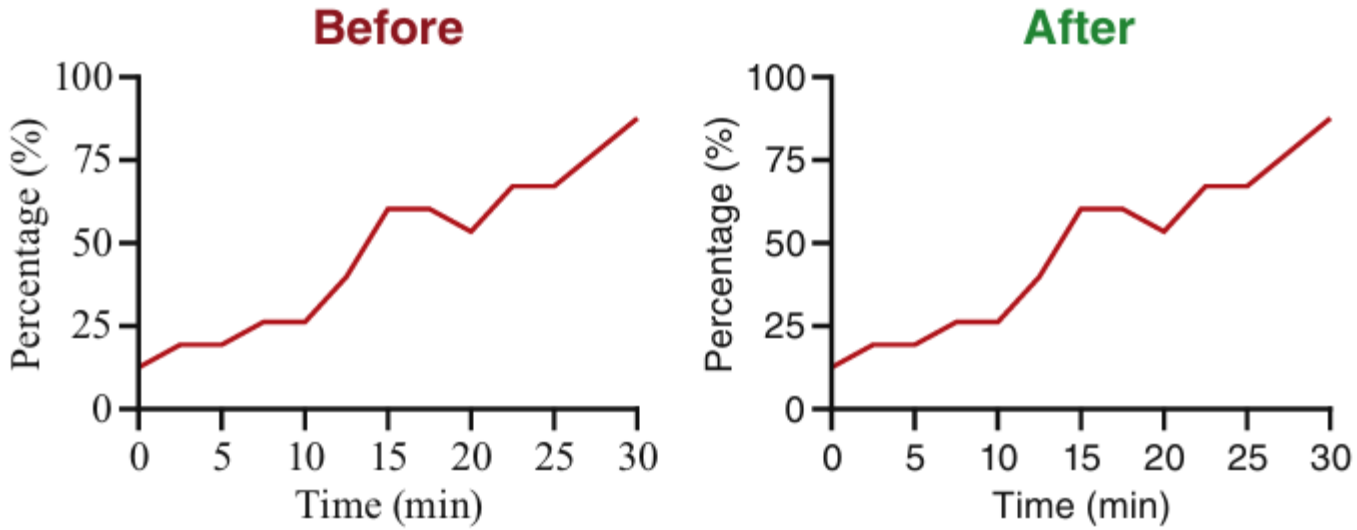
Font	Personality
Courier	retro, nerdy
Letter Gothic	simple, elementary
Lucida Sans Typewriter	informal, quirky

Font legibility



Font considerations for numbers

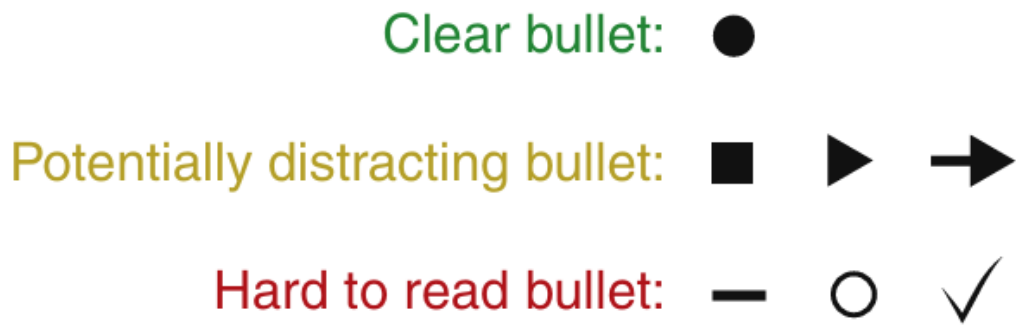
- In figures, numbers in sans serif fonts are always the most legible. When in doubt, use **Helvetica**. It's always a good choice.



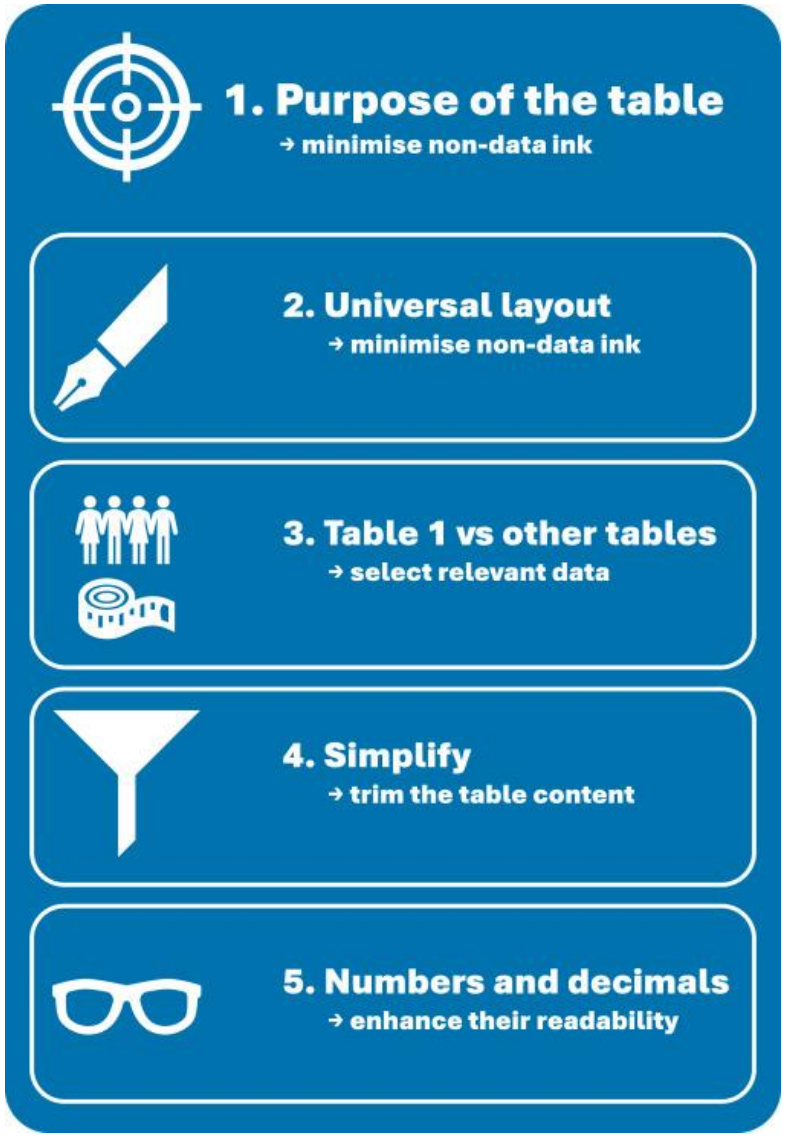
Making the best use of bullets

Making the best use of bullets

- Try not to include more than four to six items.
- Use consistent verb tense.
- end with punctuation.



Tables


1. Purpose of the table
→ minimise non-data ink
2. Universal layout
→ minimise non-data ink
3. Table 1 vs other tables
→ select relevant data
4. Simplify
→ trim the table content
5. Numbers and decimals
→ enhance their readability

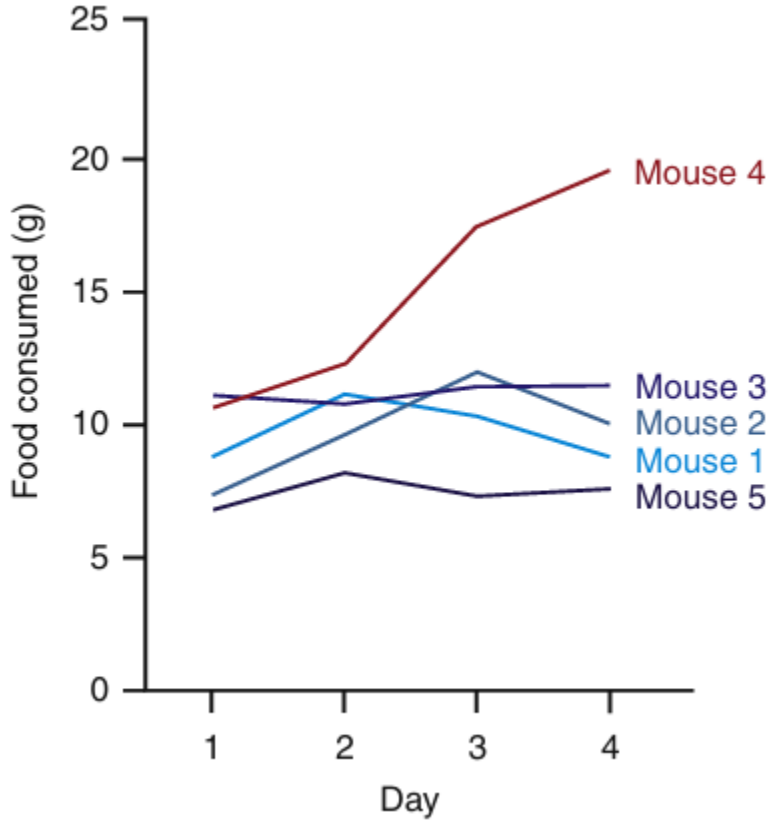
When to use a table instead of text or a figure?

- Use a table instead of text when it is easier to access single values of information from much larger datasets.
- In contract, a graph is much better than a table for visually conveying the differences, patterns, trends, or interactions between values.

Table vs figure

Table 2. Food consumed (g) in five individuals following injection of compound JSB3341.

Mouse	Day 1	Day 2	Day 3	Day 4
1	8.2	10.9	10.2	8.0
2	7.4	9.6	11.9	10.0
3	11.3	10.6	11.5	11.5
4	10.5	12.4	16.5	19.2
5	6.9	7.8	7.2	7.5



The table on the left contains the same information as the graph below, however, the relationship between the data is more apparent and accessible in the graph.

Anatomy of a table

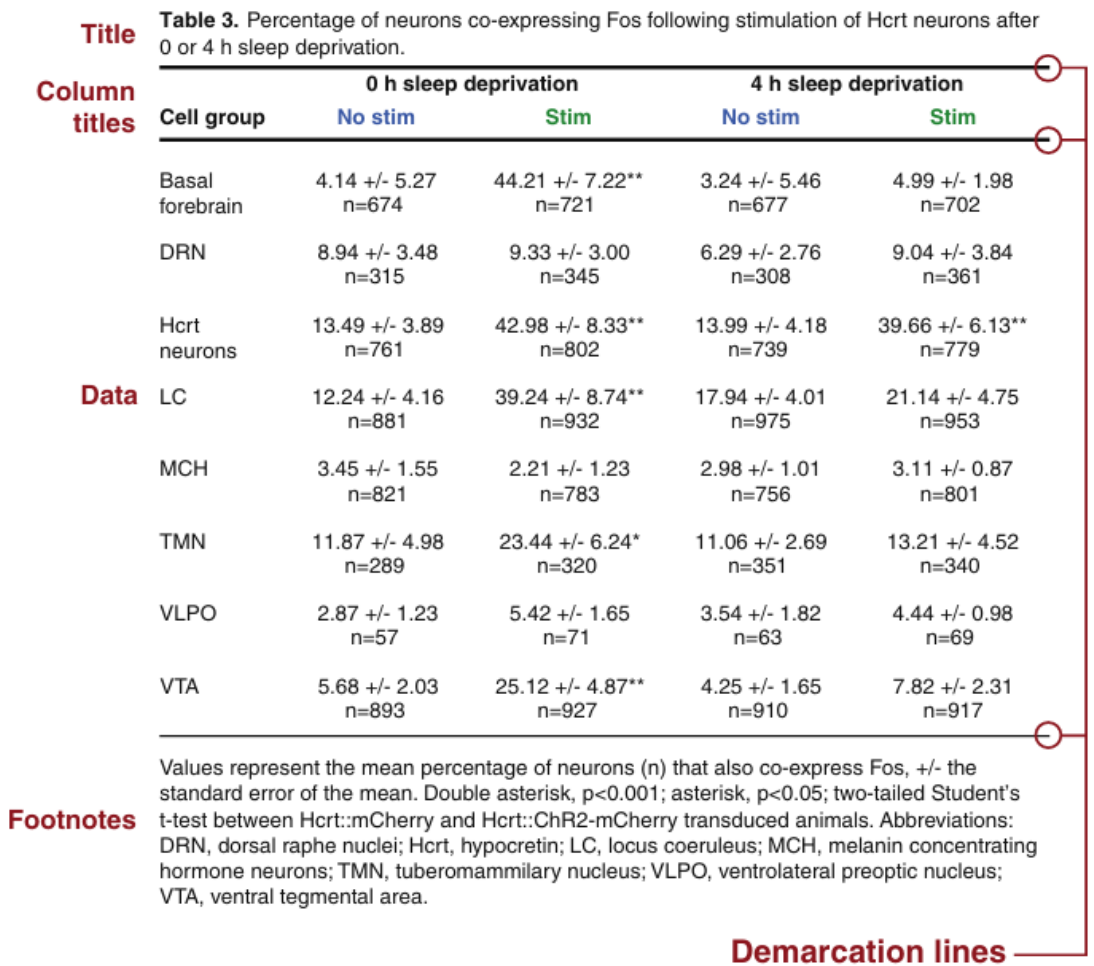


Table 3. Percentage of neurons co-expressing Fos following stimulation of Hcrt neurons after 0 or 4 h sleep deprivation.

	0 h sleep deprivation		4 h sleep deprivation	
Cell group	**No stim**	**Stim**	**No stim**	**Stim**
Basal forebrain	4.14 +/- 5.27 n=674	44.21 +/- 7.22** n=721	3.24 +/- 5.46 n=677	4.99 +/- 1.98 n=702
DRN	8.94 +/- 3.48 n=315	9.33 +/- 3.00 n=345	6.29 +/- 2.76 n=308	9.04 +/- 3.84 n=361
Hcrt neurons	13.49 +/- 3.89 n=761	42.98 +/- 8.33** n=802	13.99 +/- 4.18 n=739	39.66 +/- 6.13** n=779
LC	12.24 +/- 4.16 n=881	39.24 +/- 8.74** n=932	17.94 +/- 4.01 n=975	21.14 +/- 4.75 n=953
MCH	3.45 +/- 1.55 n=821	2.21 +/- 1.23 n=783	2.98 +/- 1.01 n=756	3.11 +/- 0.87 n=801
TMN	11.87 +/- 4.98 n=289	23.44 +/- 6.24* n=320	11.06 +/- 2.69 n=351	13.21 +/- 4.52 n=340
VLPO	2.87 +/- 1.23 n=57	5.42 +/- 1.65 n=71	3.54 +/- 1.82 n=63	4.44 +/- 0.98 n=69
VTA	5.68 +/- 2.03 n=893	25.12 +/- 4.87** n=927	4.25 +/- 1.65 n=910	7.82 +/- 2.31 n=917

Values represent the mean percentage of neurons (n) that also co-express Fos, +/- the standard error of the mean. Double asterisk, $p<0.001$; asterisk, $p<0.05$; two-tailed Student's t-test between Hcrt::mCherry and Hcrt::ChR2-mCherry transduced animals. Abbreviations: DRN, dorsal raphe nuclei; Hcrt, hypocretin; LC, locus coeruleus; MCH, melanin concentrating hormone neurons; TMN, tuberomammilary nucleus; VLPO, ventrolateral preoptic nucleus; VTA, ventral tegmental area.

Anatomy of a table

- A table should be completely comprehensible on its own.
- Information in a well-designed table should be easy to understand and access, even if the table is removed from the rest of a presentation.
- The **title** of your table must adequately and completely describe the contents of the table.
- Usually, the title is a sentence fragment (lacking a verb).

Table titles

Before Injection coordinates
After Injection coordinates of all brain regions targeted with AAV

Before Planetary probes
After Probes that have landed on other planets

Before Lizards studied
After Sex, weight, and snout-vent length of *Sceloporus malachiticus* individuals

Anatomy of a table

- **Column titles**
  - must be short and specific.
  - To make these titles shorter, it is usually okay to employ abbreviations that normally you wouldn't use (e.g., "temp" for temperature).
- **Demarcation lines**
  - These lines should be placed above and below column headings and at the very bottom of the table.
- **Footnotes**
  - describe or clarify information from the table in more detail.
  - always placed immediately beneath the table.

Logically formatting a table

- Categories of comparative data should be presented vertically in columns, not horizontally in rows.

Table 4. Area, length, and maximum depth of the three largest African lakes.

Lake	Malawi	Tanganyika	Victoria
Area (km^2)	30,044	32,893	69,485
Length (km)	579	676	322
Depth (m)	706	1470	84

Table 4. Area, length, and maximum depth of the three largest African lakes.

Lake	Area (km^2)	Length (km)	Depth (m)
Malawi	30,044	579	706
Tanganyika	32,893	676	1470
Victoria	69,485	322	84

Both tables present the same data, but the table on the right is organized more logically and is easier to read. The eye tends to read down columns more naturally than read across rows, so comparative statistics should be arranged vertically.

Logically formatting a table

- Use a hierarchical organization to emphasize the categories you think are most important.

Table 5. Number of men and women selected by NASA to be astronauts by year of selection.

	Men			Women		
	1980	1990	2000	1980	1990	2000
Mission specialist	9	12	7	2	4	3
Pilot	8	6	7	0	1	0
Total	17	18	14	2	5	3

Table 5. Number of men and women selected by NASA to be astronauts by year of selection.

	1980		1990		2000	
	Men	Women	Men	Women	Men	Women
Mission specialist	9	2	12	4	7	3
Pilot	8	0	6	1	7	0
Total	17	2	18	5	14	3

The table on the left emphasizes the comparison between men and women. The table on the right emphasizes the comparison between years of selection.

Logically formatting a table

- Horizontal entries of data should not be listed randomly. Order information in alphabetical or numerical order depending on which data you want to emphasize.

Table 6. Diameter and mass of planets in relation to the Earth.

Planet	Diameter	Mass
Mercury	0.38	0.06
Venus	0.95	0.82
Earth	1.00	1.00
Mars	0.53	0.11
Jupiter	11.21	317.80
Saturn	9.45	95.20
Uranus	4.01	14.60
Neptune	3.88	17.20

Table 6. Diameter and mass of planets in relation to the Earth.

Planet	Diameter	Mass
Earth	1.00	1.00
Jupiter	11.21	317.80
Mars	0.53	0.11
Mercury	0.38	0.06
Neptune	3.88	17.20
Saturn	9.45	95.20
Uranus	4.01	14.60
Venus	0.95	0.82

Table 6. Diameter and mass of planets in relation to the Earth.

Planet	Diameter	Mass
Jupiter	11.21	317.80
Saturn	9.45	95.20
Uranus	4.01	14.60
Neptune	3.88	17.20
Earth	1.00	1.00
Venus	0.95	0.82
Mars	0.53	0.11
Mercury	0.38	0.06

The table on the left lists the planets in order from the sun. The center table lists the planets in alphabetical order. The table on the right lists the planets in descending order of diameter. Any order is acceptable as long as it follows an understandable hierarchy and emphasizes the main point that you want to convey.

Logically formatting a table

- Text and number alignment

Don't:

Align the major items on the lefthand side flush right or center

Align text entries flush right

Align whole numbers center or flush left

Align numbers with decimals or +/– symbols in the center or flush left or right

Table 7. Average mass and length of ten of the heaviest mammals.

Animal	Environment	Avg. mass (kg)	Avg. length (m)
Asian elephant	Terrestrial	4,150	6.8
Blue whale	Aquatic	110,000	25.5
Fin whale	Aquatic	57,000	20.6
Giraffe	Terrestrial	1,015	5.1
Gray whale	Aquatic	19,500	13.5
Hippopotamus	Terrestrial	1,800	4.0
Humpback whale	Aquatic	29,000	13.5
Sperm whale	Aquatic	31,250	13.3
Walrus	Terrestrial	944	2.8
White rhinocerus	Terrestrial	2,100	4.4

Table 7. Average mass and length of ten of the heaviest mammals.

Animal	Environment	Avg. mass (kg)	Avg. length (m)
Asian elephant	Terrestrial	4,150	6.8
Blue whale	Aquatic	110,000	25.5
Fin whale	Aquatic	57,000	20.6
Giraffe	Terrestrial	1,015	5.1
Gray whale	Aquatic	19,500	13.5
Hippopotamus	Terrestrial	1,800	4.0
Humpback whale	Aquatic	29,000	13.5
Sperm whale	Aquatic	31,250	13.3
Walrus	Terrestrial	944	2.8
White rhinocerus	Terrestrial	2,100	4.4

Do:

Align the major items on the lefthand side of a table flush left

Align text entries in the center or flush left

Align whole numbers flush right

Align numbers with decimals or +/– symbols centered on the decimal point or +/–

Choosing to add gridlines on tables

- In written and poster presentations, it is usually best to avoid using gridlines to separate rows and columns of information.

Name	Data	Data	Data	Data
Item A	2.3	1.9	8.7	9.0
Item B	4.0	7.2	9.1	5.5
Item C	0.4	0.8	5.2	0.6
Item D	8.0	9.4	1.0	4.2
Item E	6.3	3.5	8.0	6.0

Name	Data	Data	Data	Data
Item A	2.3	1.9	8.7	9.0
Item B	4.0	7.2	9.1	5.5
Item C	0.4	0.8	5.2	0.6
Item D	8.0	9.4	1.0	4.2
Item E	6.3	3.5	8.0	6.0

Choosing to add gridlines on tables

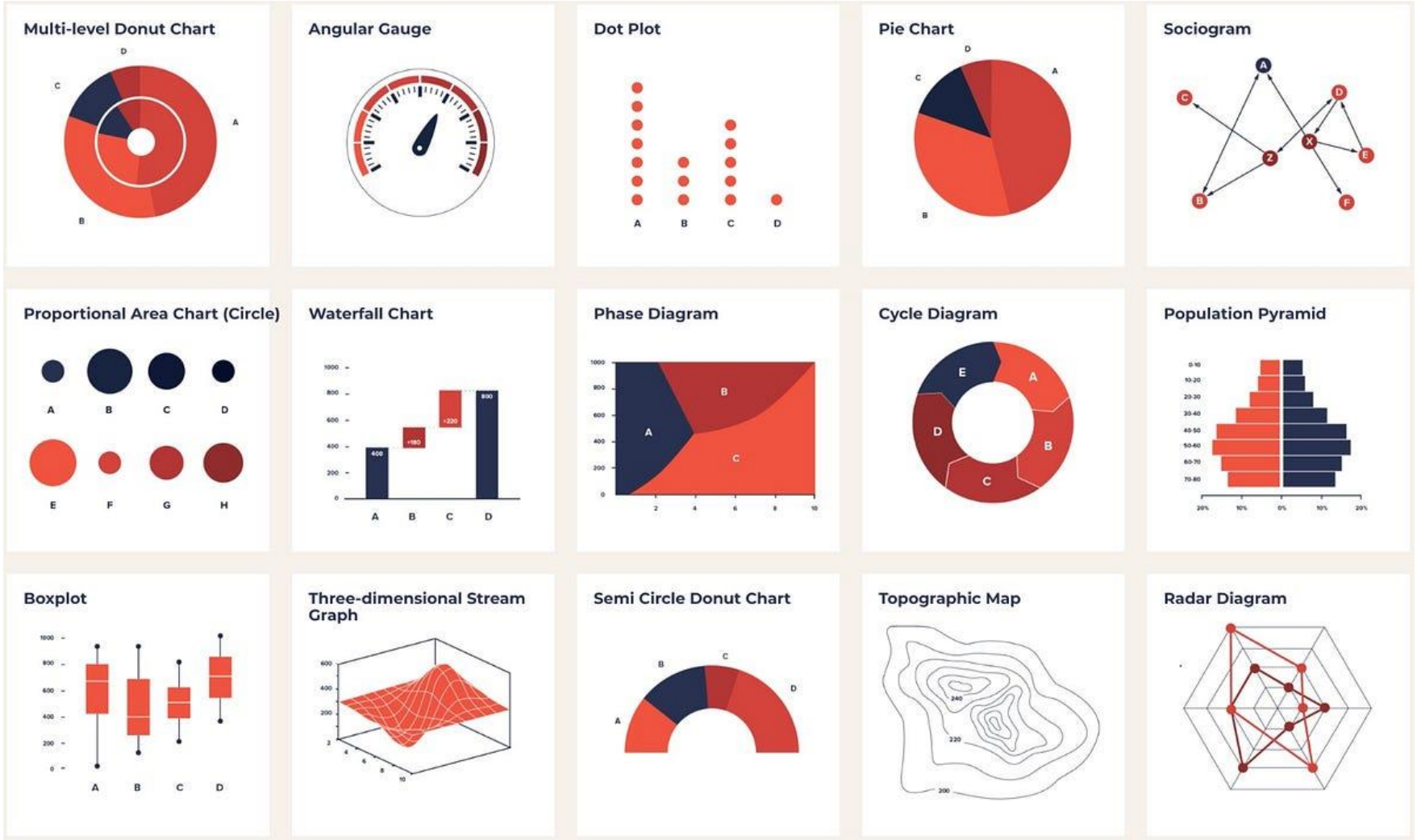
- in a large table, subtle gridlines can help guide the reader in asea of numbers.
- If it is helpful, add light horizontal gridlines every three to five rows. Alternatively, add subtle gray shading every other row.

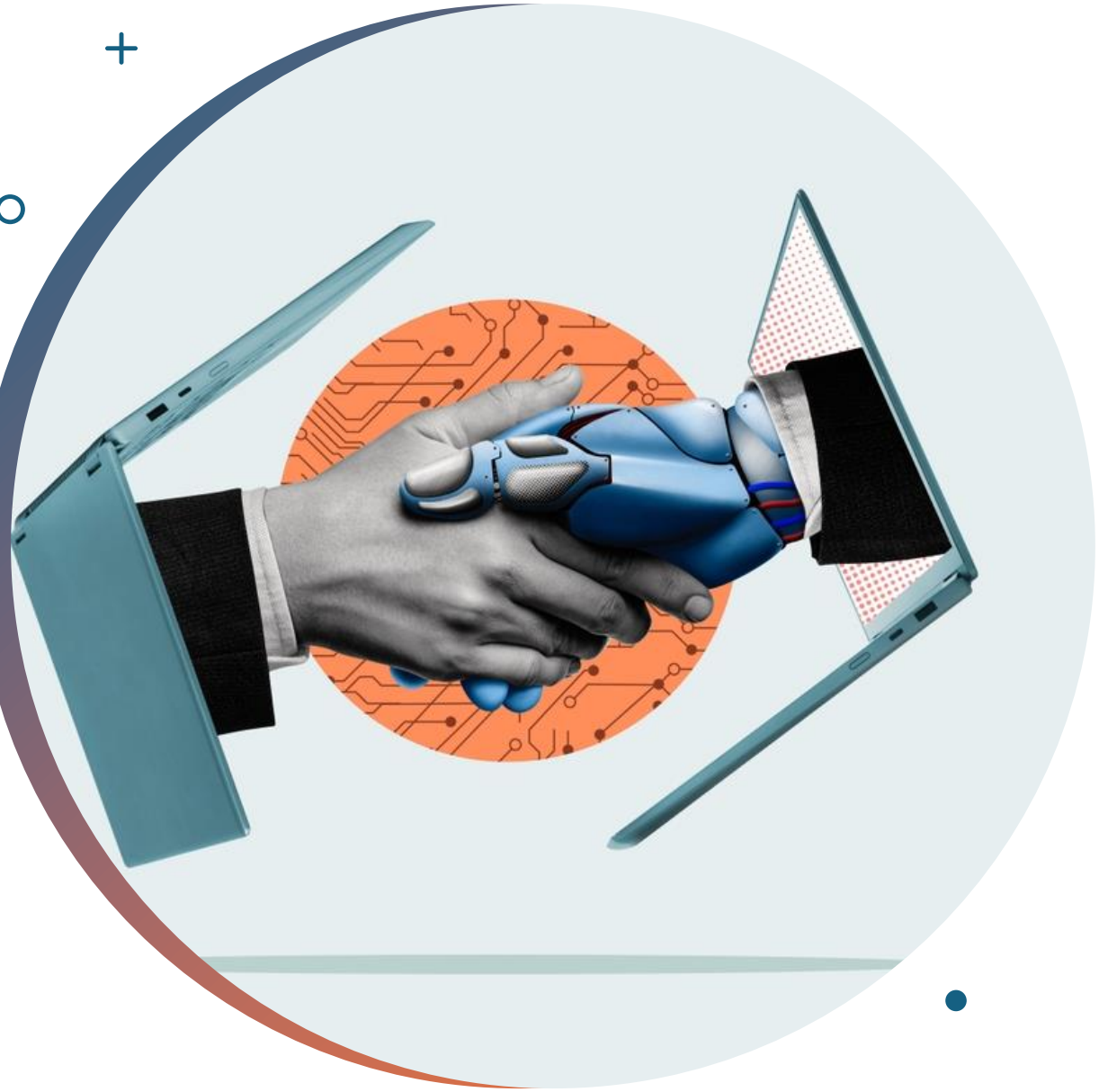
Name	Data	Data	Data
Item A	2.3	1.9	8.7
Item B	4.0	7.2	9.1
Item C	0.4	0.8	5.2
Item D	8.0	9.4	1.0
Item E	6.3	3.5	8.0
Item F	0.5	1.7	3.8
Item G	7.0	1.4	9.2
Item H	1.6	0.3	8.1
Item I	4.7	9.2	3.5
Item J	9.1	4.8	3.2
Item K	7.1	4.2	3.3
Item L	8.0	2.8	4.7

Name	Data	Data	Data
Item A	2.3	1.9	8.7
Item B	4.0	7.2	9.1
Item C	0.4	0.8	5.2
Item D	8.0	9.4	1.0
Item E	6.3	3.5	8.0
Item F	0.5	1.7	3.8
Item G	7.0	1.4	9.2
Item H	1.6	0.3	8.1
Item I	4.7	9.2	3.5
Item J	9.1	4.8	3.2
Item K	7.1	4.2	3.3
Item L	8.0	2.8	4.7

Name	Data	Data	Data
Item A	2.3	1.9	8.7
Item B	4.0	7.2	9.1
Item C	0.4	0.8	5.2
Item D	8.0	9.4	1.0
Item E	6.3	3.5	8.0
Item F	0.5	1.7	3.8
Item G	7.0	1.4	9.2
Item H	1.6	0.3	8.1
Item I	4.7	9.2	3.5
Item J	9.1	4.8	3.2
Item K	7.1	4.2	3.3
Item L	8.0	2.8	4.7

Exploratory and analytical data analysis





Hands-on examples

- Design a standard table.
- Create a standard scatter plot.
- Perform between group comparison and regressions.

Suggested statistical analysis platforms

Statisty

RAWGraphs

1. ggquickeda: publication ready tables

- Requires R (version > 4.x), R-tools, and R Studio
- Run these commands to install the shiny app:
    1. install.packages("ggquickeda")
    2. library(ggquickeda)
    3. run_ggquickeda()

C:/Program Files/R/R-4.4.1/library/ggquickeda/shinyapp - Shiny

http://127.0.0.1:7237 | Open in Browser | Publish

Inputs | Graph Options | How To

Click Browse to choose csv file to upload with

(•) comma (,) ( ) or semicolon (;) separators, or use sample data

Browse... No file selected

[x] Character Variables as Factors?

[ ] Numeric Variables -99 as Missing?

Please select a data set

[ ] Plot a matrix of all Y variables

Please select a data set

Categorize/Cut | Merge factor levels

Recode/Reorder Categories

Combine Two/Three Variables | Filters

One Row by ID(s) | Division/Inverse/Rounding

Reorder Facets or axis Levels

Please select a data set
Please select a data set
Please select a data set
Please select a data set
Please select a data set
Please select a data set
Please select a data set
Please select a data set

X/Y Plot | Export Plots | Experimental Plotly | Descriptive Stats | Data | Plot Code

Please select at least one x or at least one y.

Please select a data set
Please select a data set
Please select a data set
Please select a data set

Points, Lines | Color/Group/Split/Size/Fill Mappings | Boxplots/Violins | Quantile Regression

2. Statisty.app

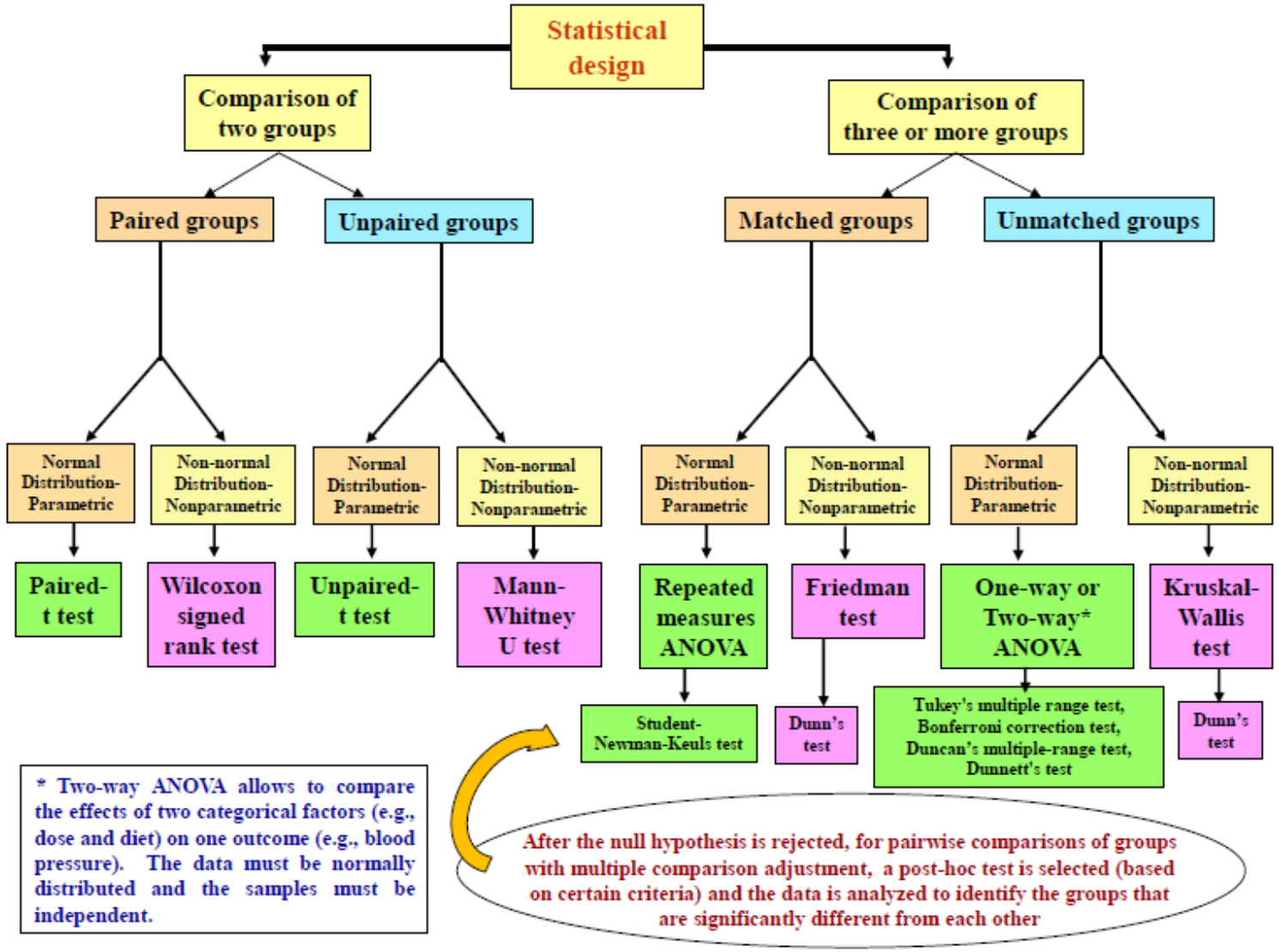
Statisty

Install

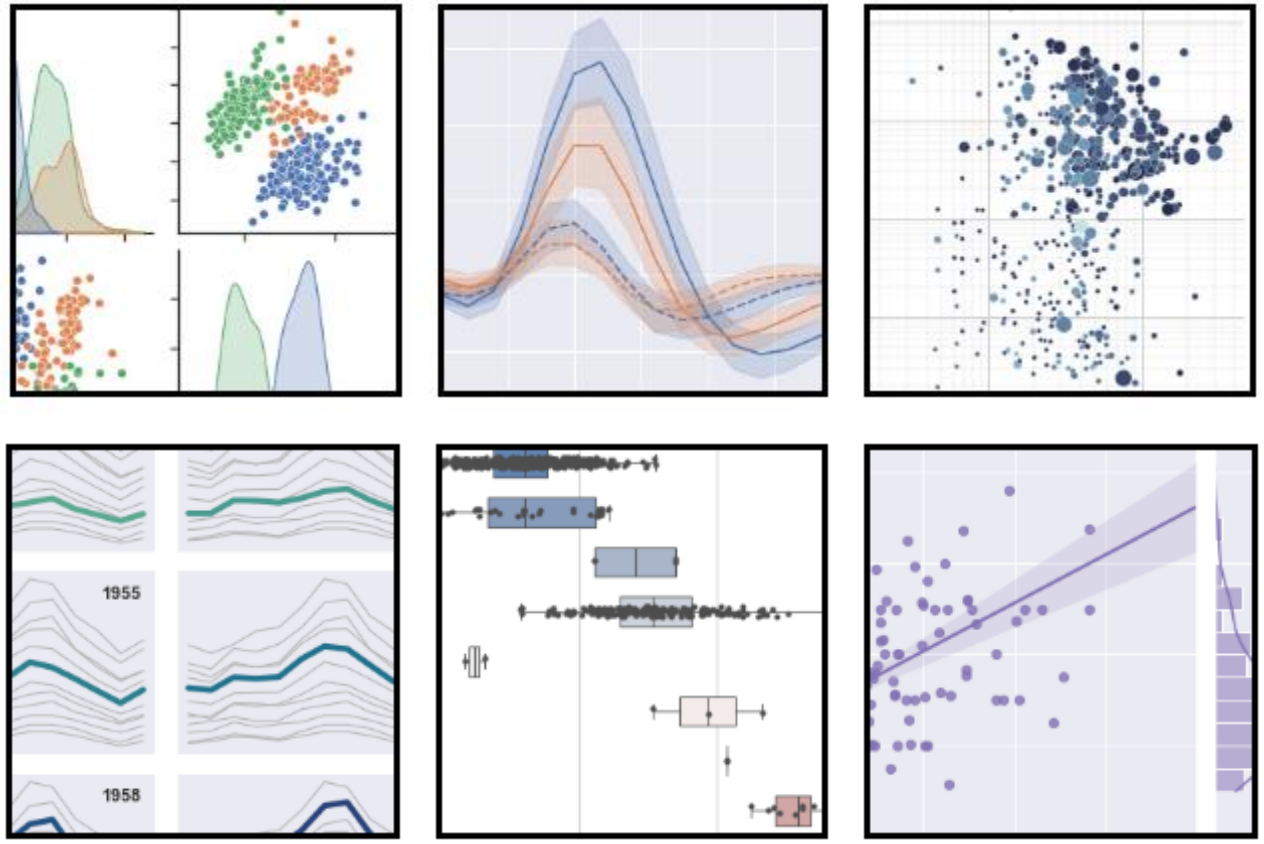
Clear Table Export / Import Settings

metrisch	metrisch	nominal	metrisch	metrisch	nominal	nominal	nominal
Admission Code	**Age (Years)**	**Age group**	**Age (Months)**	**Weight (kg)**	**Cardiac Disease**	**Metabolic Disease**	**Immune De**
3450725	0.75	< 1 year	9	9	no	no	no
3450781	4.92	? 1 year	59	20	no	no	no
3460729	0.92	< 1 year	11	9	no	no	no
3452096	4	? 1 year	48	15	no	no	no
3447132	1.42	? 1 year	17	12	no	no	no
3422441	13.75	? 1 year	165	55	no	no	no
3398428	0.83	< 1 year	10	7	no	no	no
3393003	1.17	? 1 year	14	9	no	no	no
3395204	1.08	? 1 year	13	6	yes	no	no
3380830	4.75	? 1 year	57	16	no	no	no
3376290	3.67	? 1 year	44	15	no	no	no
3370560	0.17	< 1 year	2	4	no	no	no

Choosing the right test



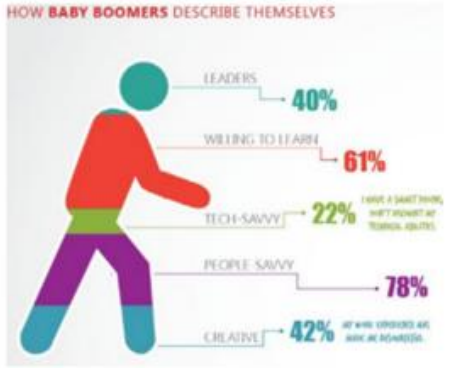
statistical data visualization



Why this topic matters?

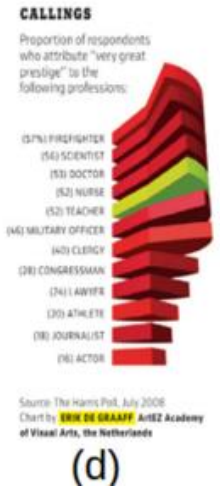
- Misinformation due to shape violation:
  - (a) using 3D pie chart for comparing proportion
  - (b) using wrong visualization chart type
  - (c) pie chart should be used
  - (d) bar chart should be used

(a)



(b)

(c)

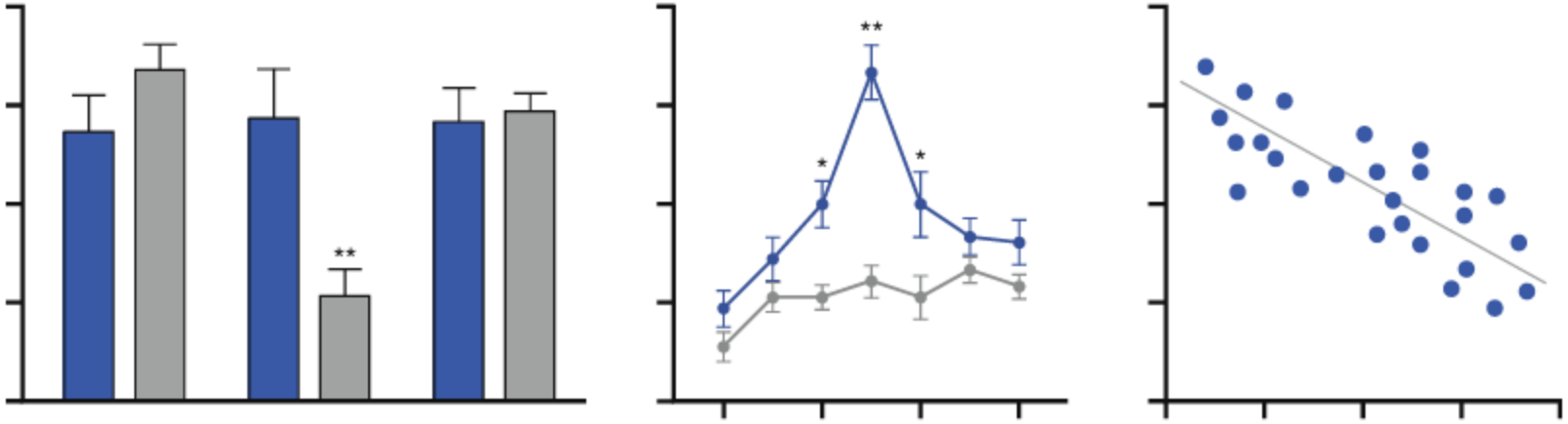


(d)

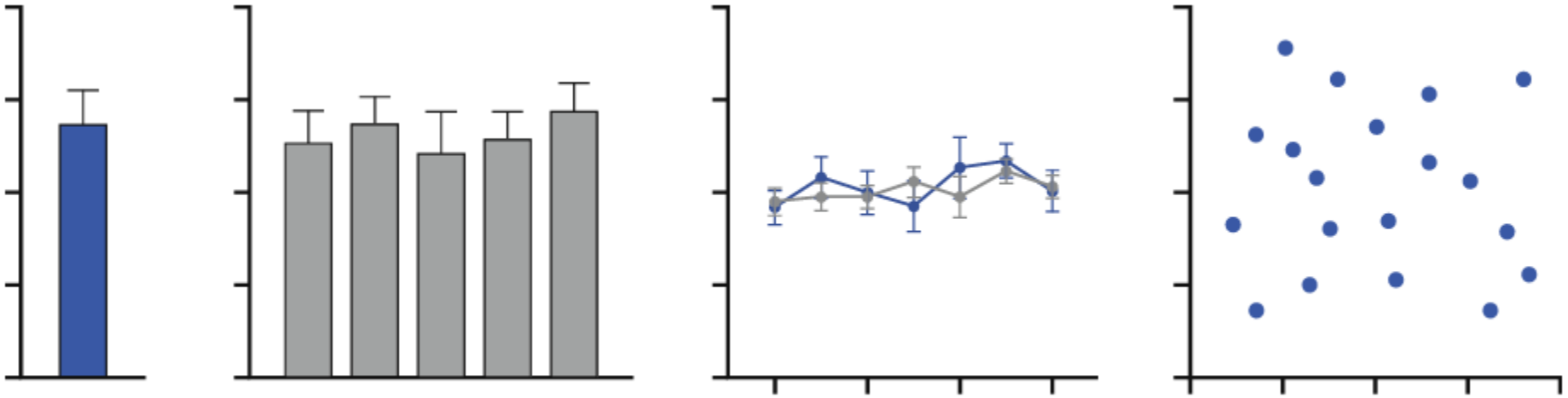
Nguyen VT, Jung K, Gupta V. Examining data visualization pitfalls in scientific publications. Visual Computing for Industry, Biomedicine, and Art. 2021 Oct 29;4(1):27.

When to use a graph

- Use a graph to visually communicate patterns, trends, or relationships among data.



Bad examples



Anatomy of a graph

- Audiences naturally focus on figures more than written text or oral narration.

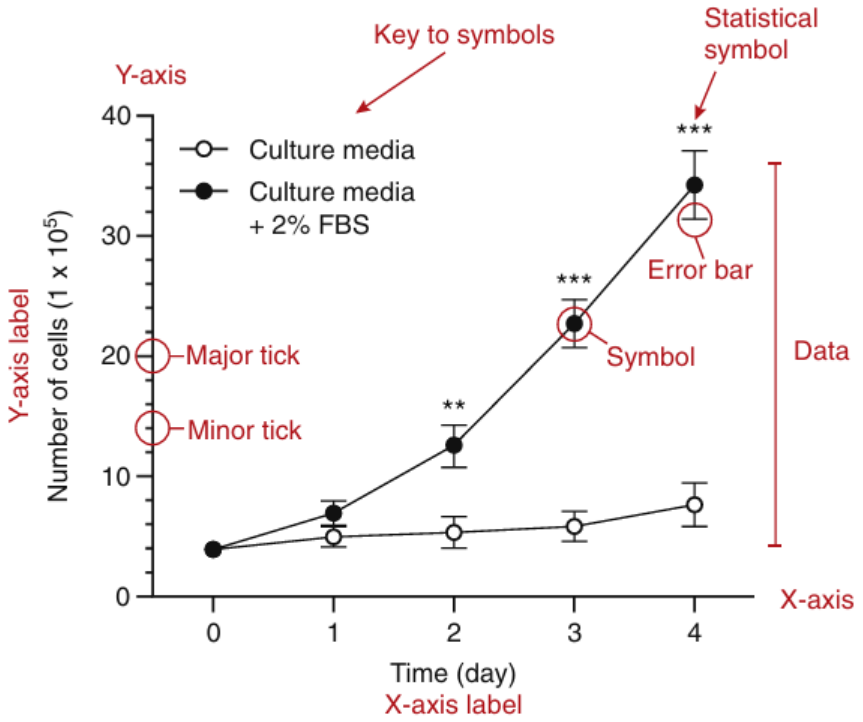
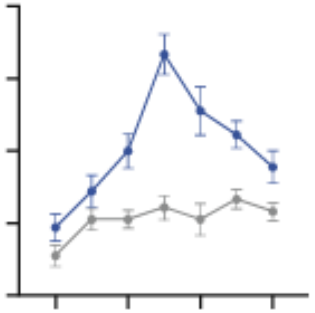


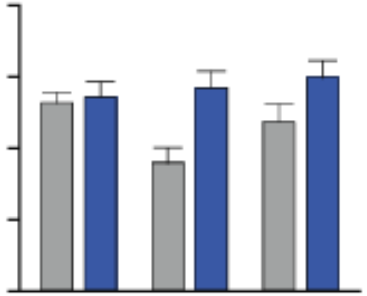
Figure legend

Figure 1. Addition of 2% fetal bovine serum (FBS) to culture medium increases the growth of DK cells. Data represent the mean +/- standard deviation; n=6 trials; **p<0.05, ***p<0.0001, repeated measures ANOVA followed with Bonferroni post hoc tests.

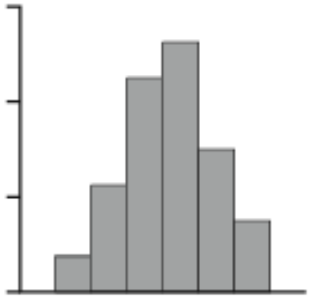
Five fundamental categories of graphs



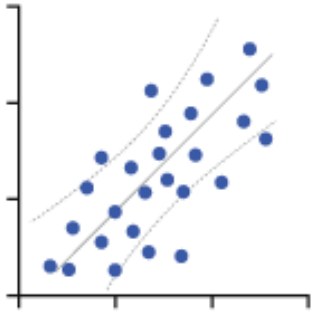
Line graph
Visualizes a trend of continuous data, often over time



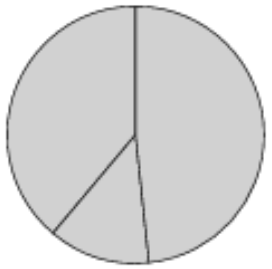
Bar graph
Compares discrete quantities of non-continuous data



Histogram
Reports the distribution of data and the frequency with which they occur

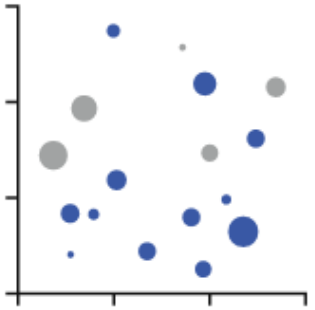


Scatterplot
Displays the relationship between two continuous variables

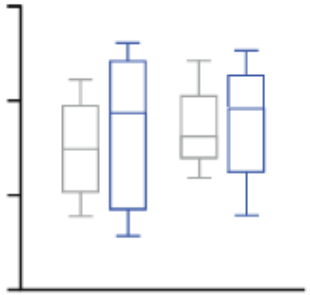


Pie chart
Shows the proportional values that make up a whole

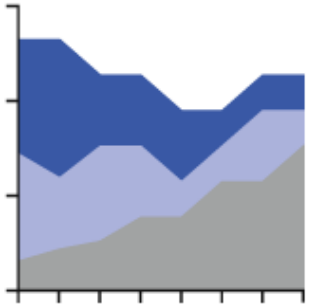
Specialized graphs



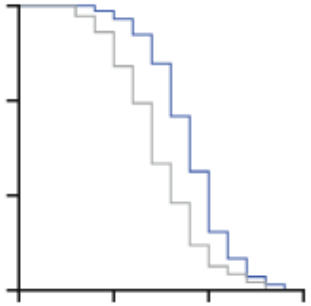
Bubble scatterplot
Displays the relationship between two continuous variables and a third variable represented by color and/or size



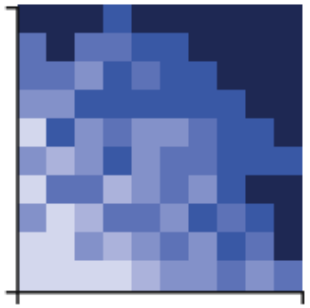
Box and whisker graph
Compares the maximum, upper quartile, median, lower quartile, and minimum values of one or more datasets



Area chart
Represents the cumulative totals of multiple, continuous data series, often over time



Decay/growth graph
Reports the cumulative decay or growth of one or more populations over time

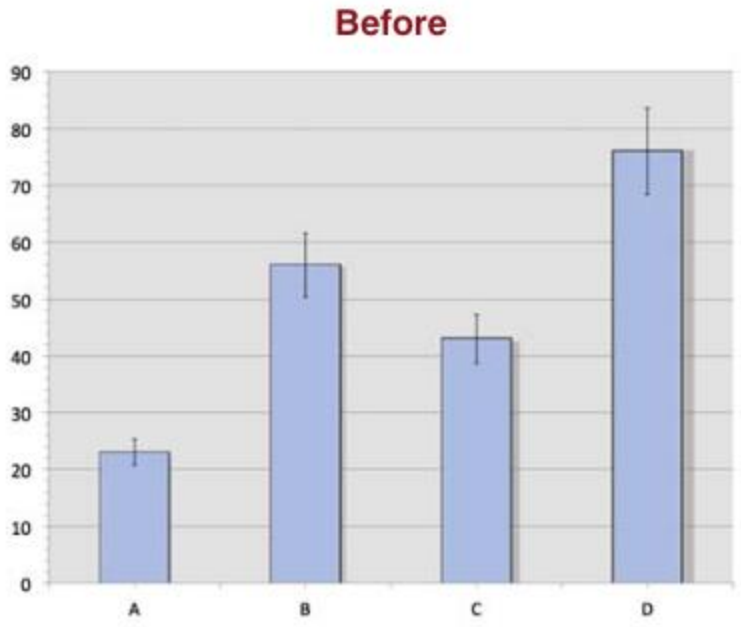


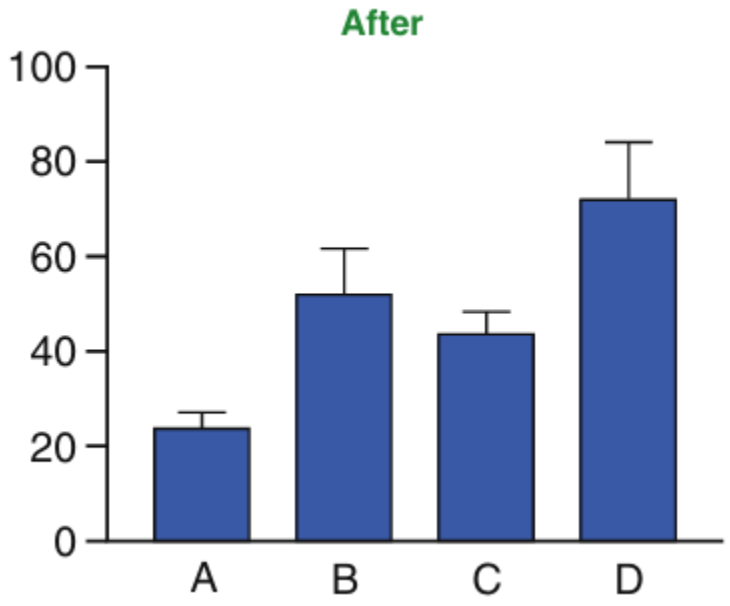
Heat map
Displays the relationship between two variables and the intensity of a third variable represented by a gradient of color

Radar chart
Compares values between two or more data series across multiple variables

General design considerations for graphs

- Never trust a computer to do all of the design work for you.



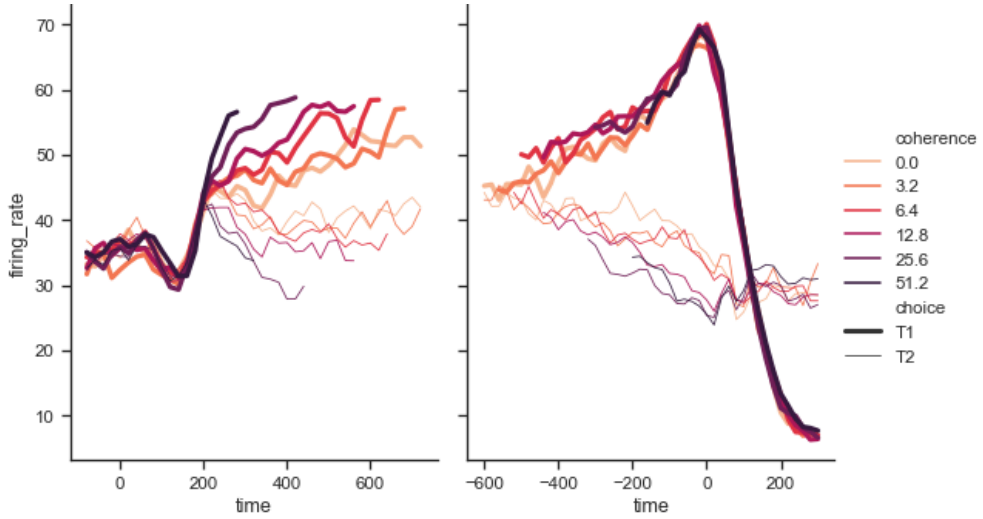


Anatomy of a graph

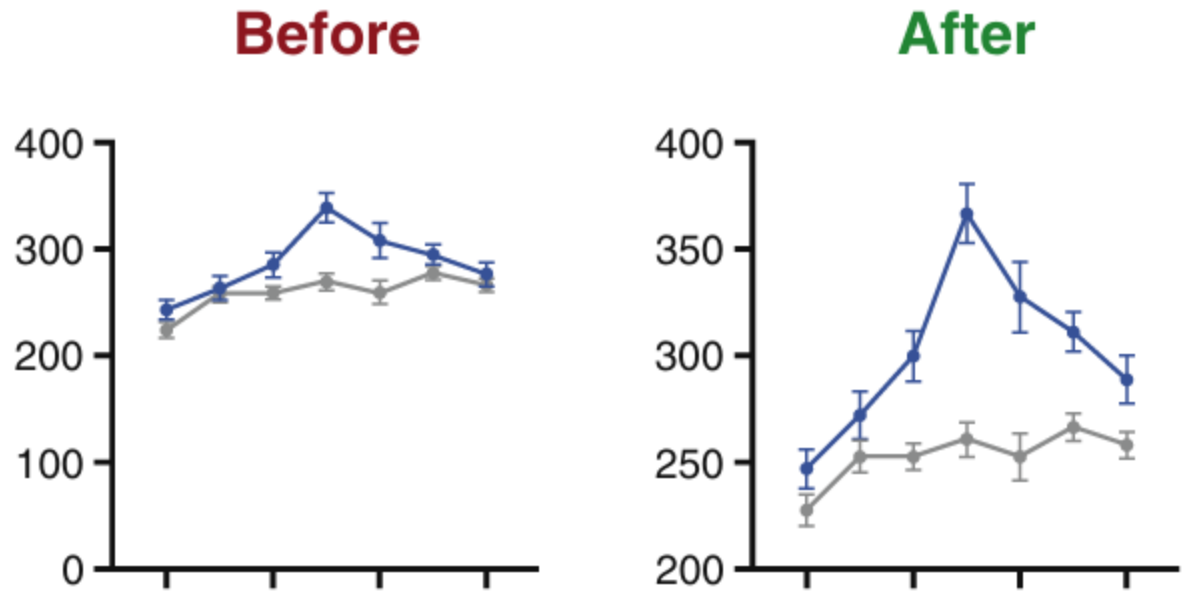
- **Background color**
  - For written presentations, this will almost always be white.
- **Color of datasets**
  - try to keep those colors consistent across your entire presentation.
- **Gridlines**
  - only use gridlines if you need them. They are often unnecessary.
- **Pleasing increments of scale**
  - Use increments that people naturally use when counting, such as multiples of 2, 5, 10, 20, 25, 100.
- **Font**
  - Helvetica or Arial are always good choices.

Designing line graphs

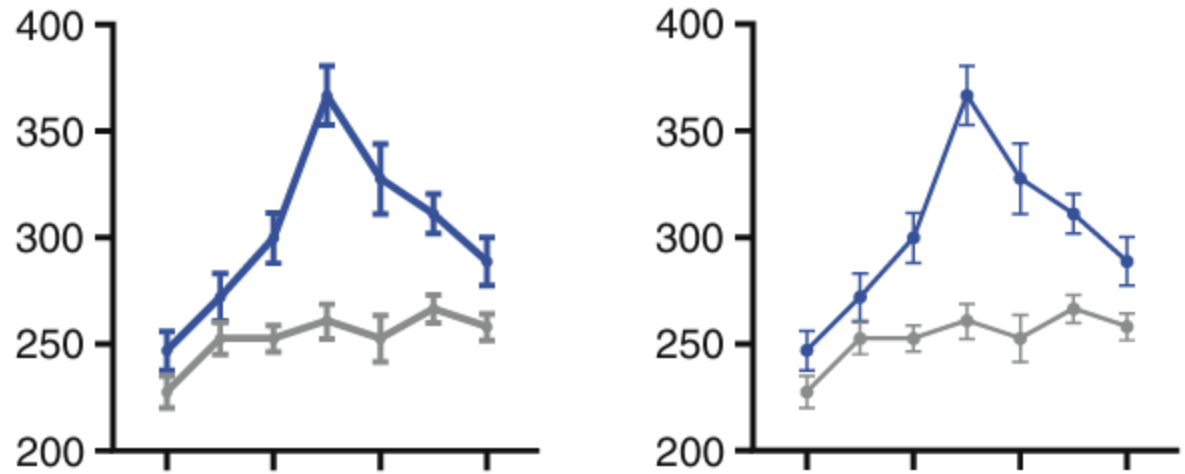
- Line graphs are used to display continuous data series and show trends over time.
- Often the most interesting conclusions in these graphs are about how the data change overtime, or how different categories of data compare with each other.



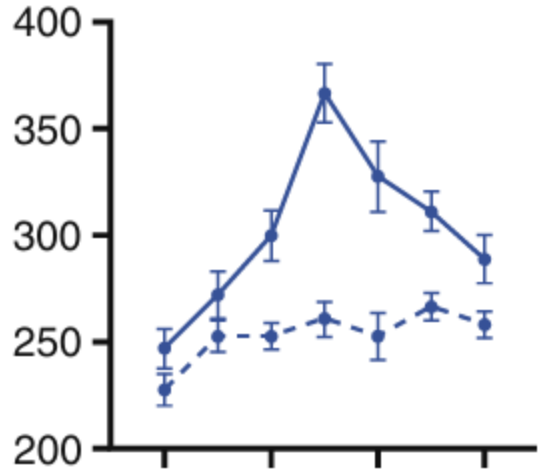
To take full advantage of the size of your graph, the range in height of your data should ideally take up about three-fourths the height of the Y-axis. It is okay to not start the Y-axis at zero as long as your origin is clearly labeled.

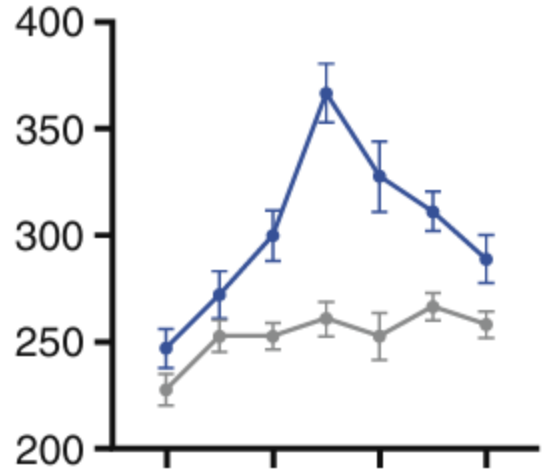


To help your audience visualize discrete values of data, choose line weights that are about half as big as the data symbols themselves. Choose even lower line widths for the error bars.

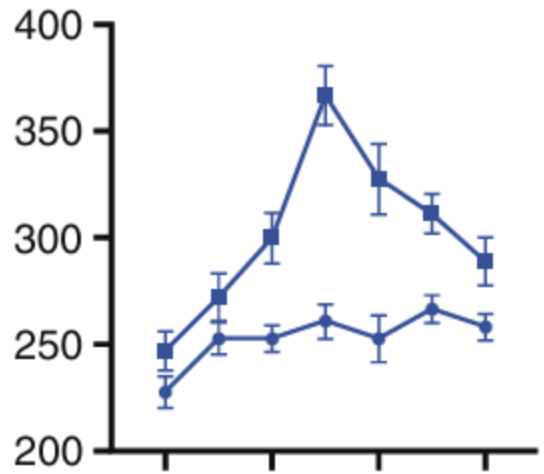


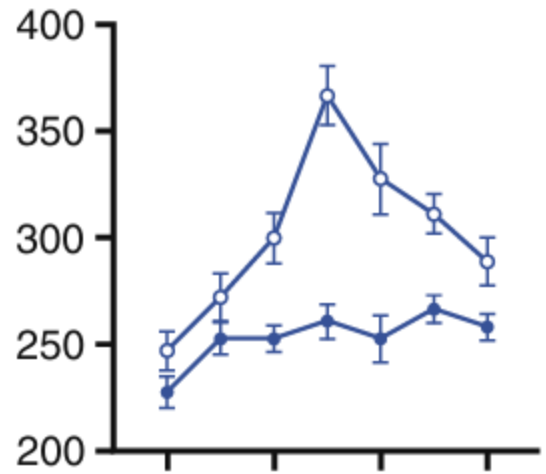
Dashed lines are hard to read. To increase the contrast between two different categories of data, use different colors or shades.





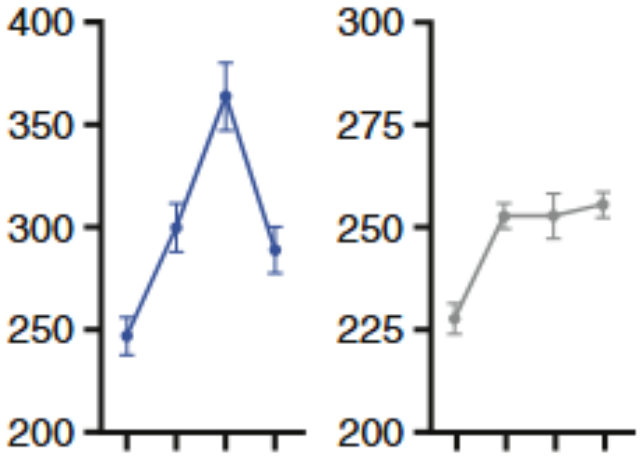
Alternatively, vary the symbols used to represent data. It is usually easiest to distinguish between closed and open symbols (● versus ○) rather than different shapes (● versus ■).



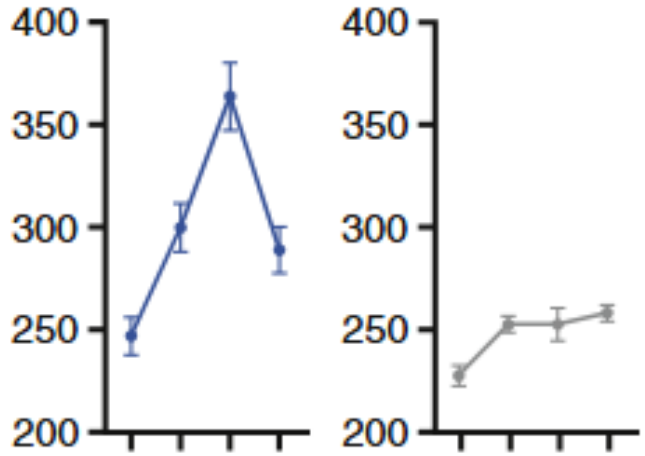


When presenting two graphs next to each other that represent the same experiment, make sure that your axes are similar in scale. Otherwise, you risk misrepresenting your data.

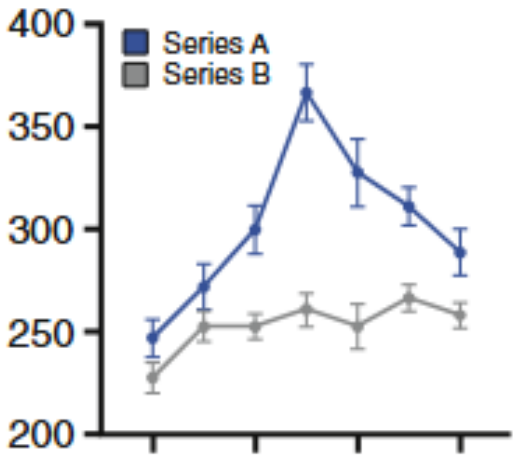
Before

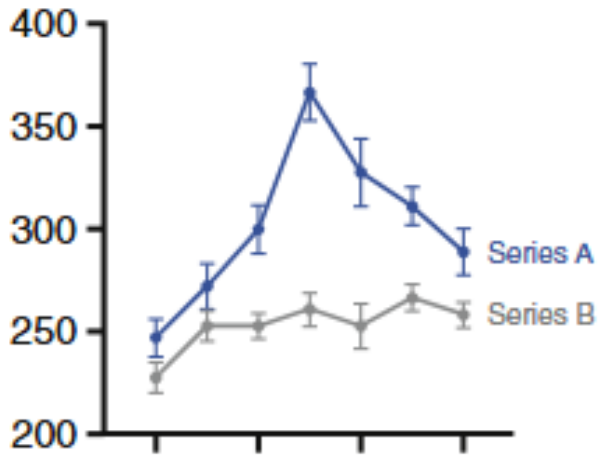


After

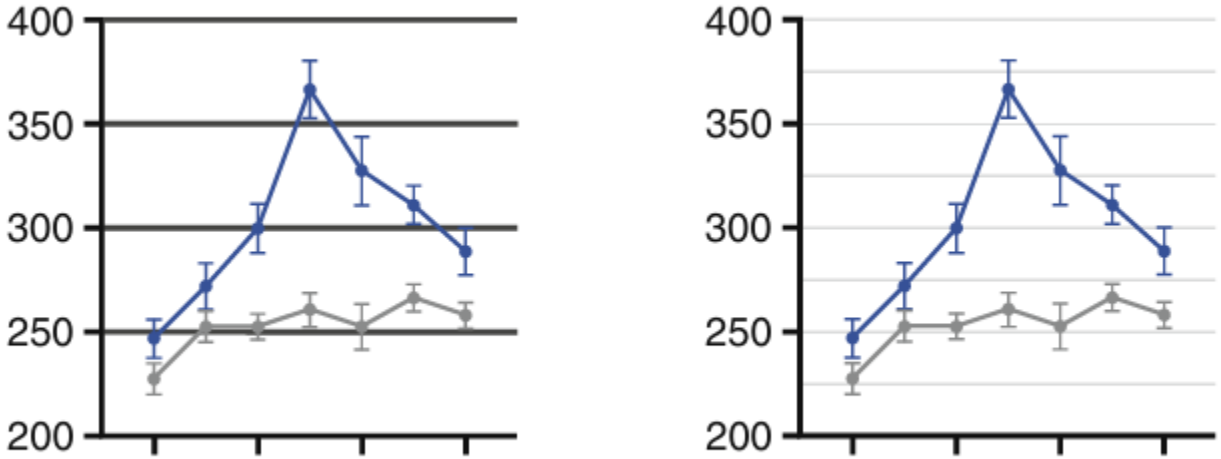


To make it easier for your audience to match data labels to specific lines, try to place your labels next to the lines themselves rather than a far-off spot.

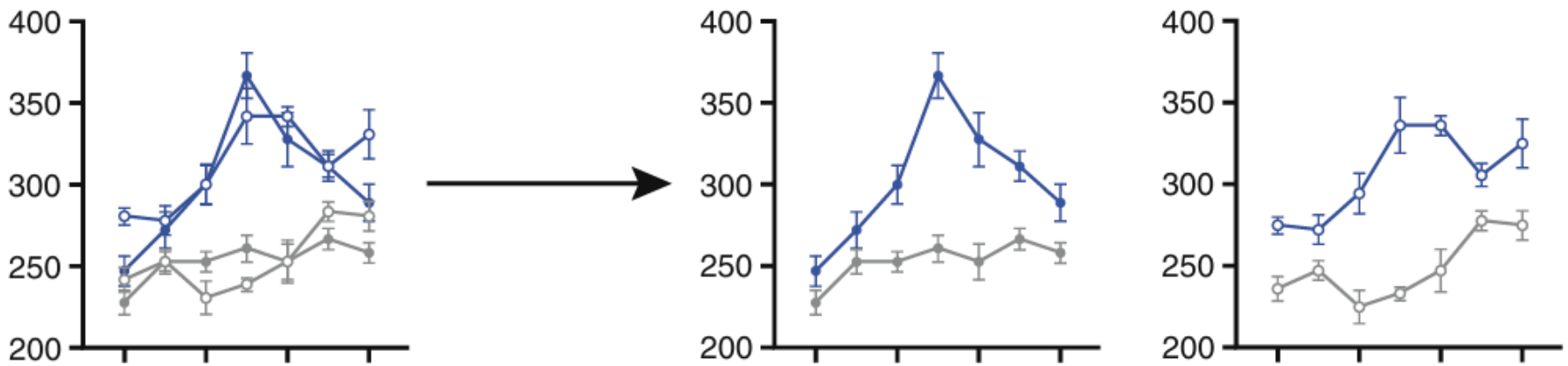




Don’t overwhelm your chart with an overwhelming gridline. If you feel a gridline is helpful, place subtle, light gridlines in the background.

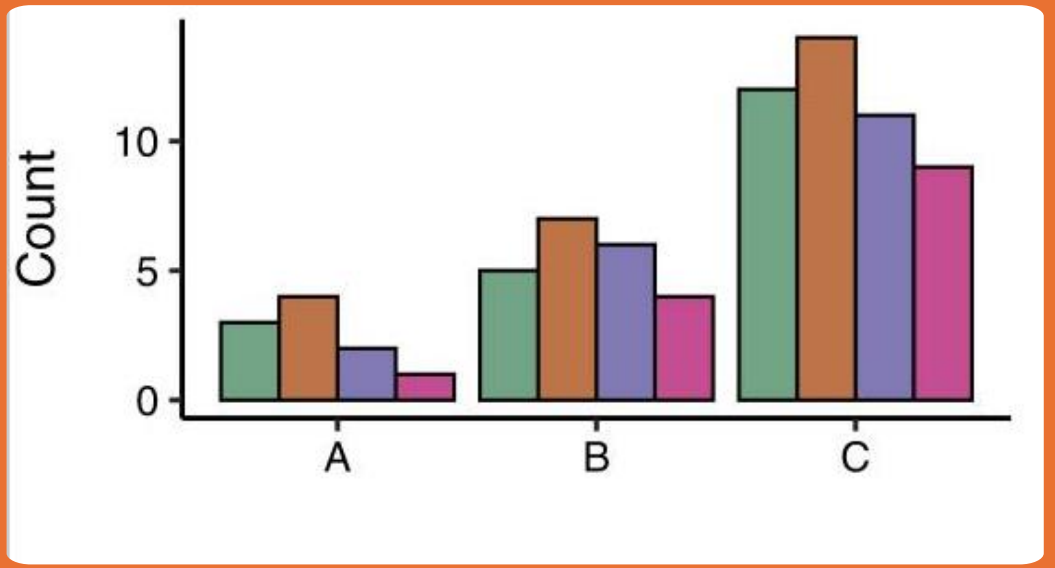


To optimize clarity in a single graph, keep the maximum number of lines to about three or four. If you need more (or if your lines look crowded), separate your graphs into multiple panels.



Designing bar graphs

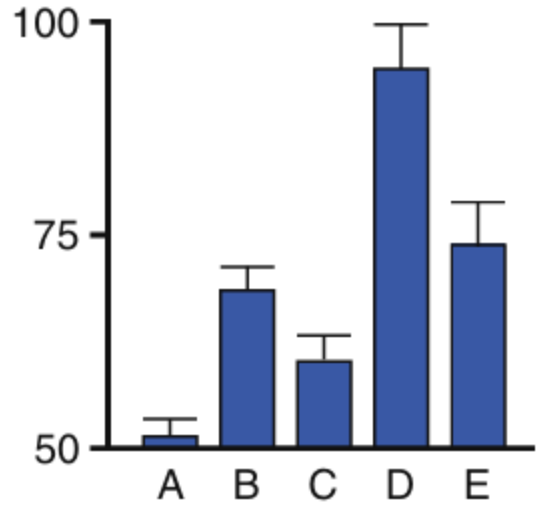
Bar graphs are used to compare discrete quantities of non-continuous data.

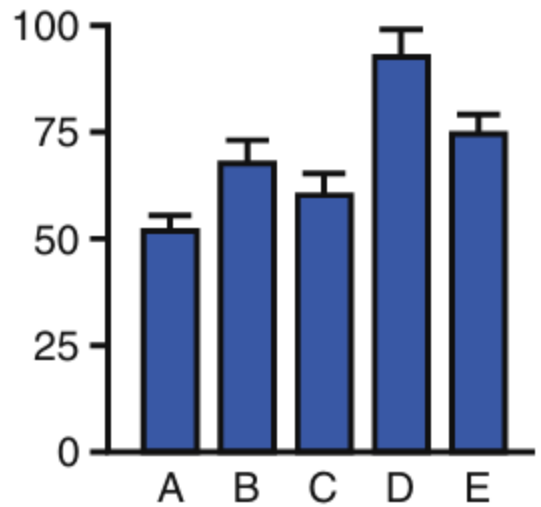


In most cases it is best to start the Y-axis of a bar chart at zero. Because the height of a bar represents a discrete value, charts that don't start at zero can be misleading. The exception is for data in which zero does not specify a value of "nothing," but exists on a relative continuum, for example, when measuring temperature.

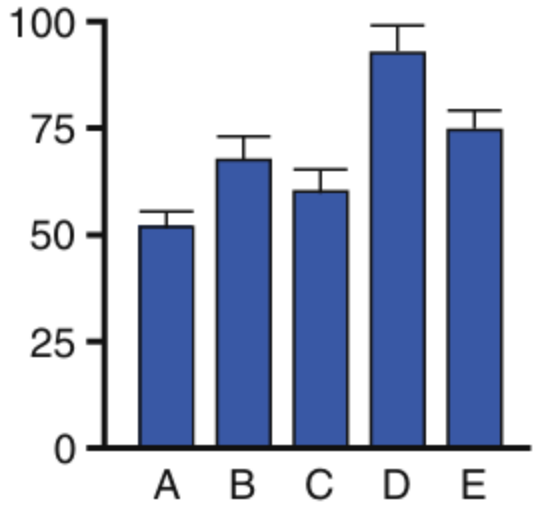
Don't let the thickness of the lines outlining the bars overwhelm the bars themselves. The most visible, prominent bars are contained within lines that are very fine but visible.

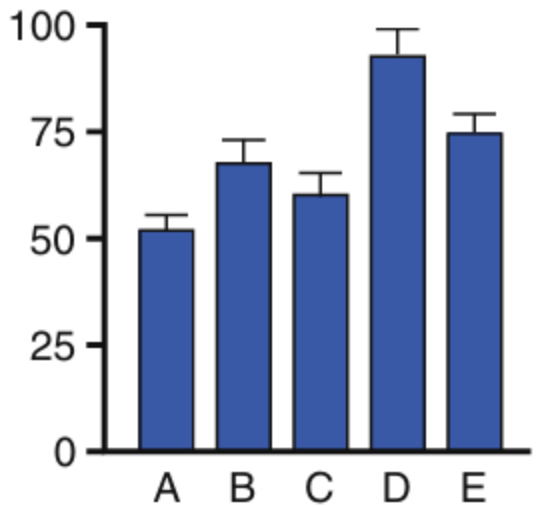
Before



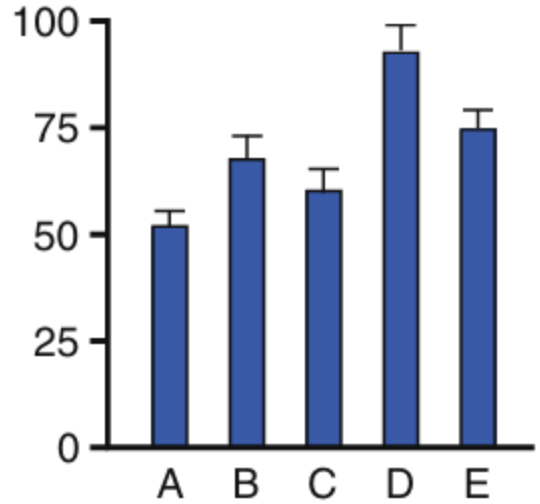


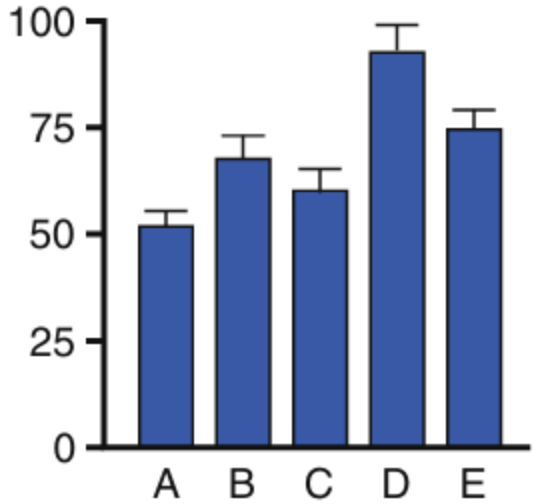
After



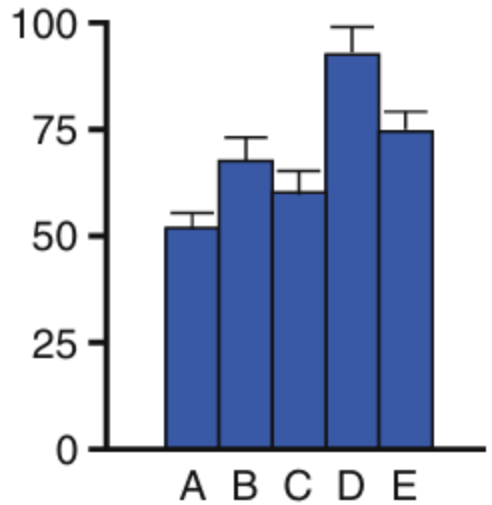


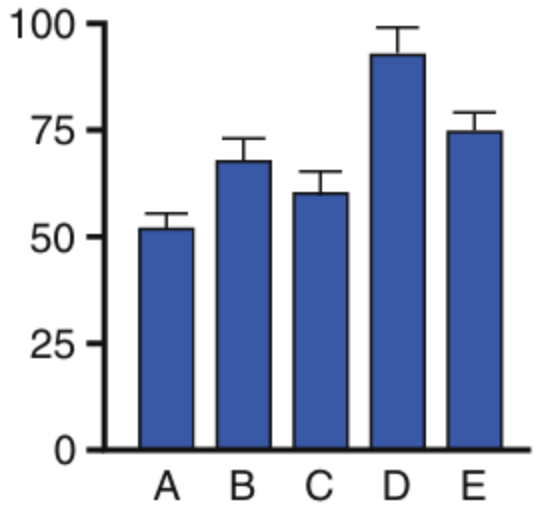
To increase visiblity of the individual bars, avoid bar widths that are too thin or thick. If the bars are too spread out, it can be more difficult to visually relate them to each other. The ideal spacing between bars is usually about one-third their width.



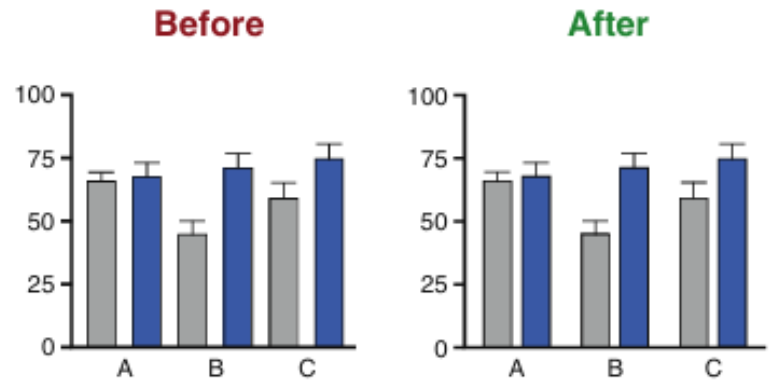


To help readers visualize different categories of data, don't place individual bars directly next to each other.

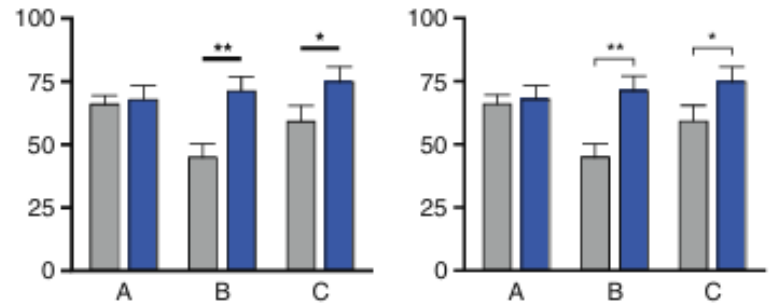




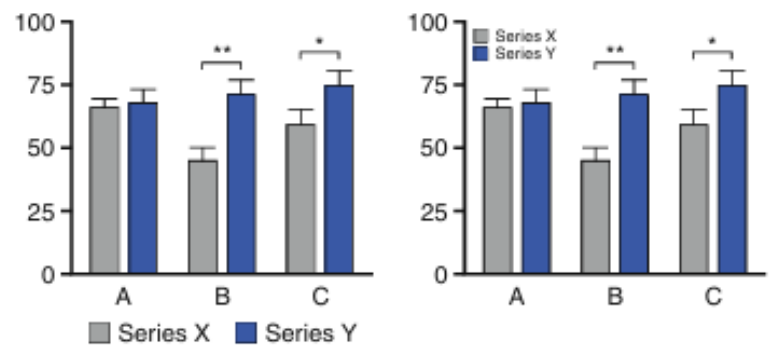
In bar charts with two or more categories of data, place a larger space between the different variables on the X-axis than between the different bars. Be sure to keep the order of the bars consistent across the X-axis.



Don't let statistical information overwhelm a graph. Show statistical differences between two bars using a thin line with subtle overhangs to aid the viewer's eye.

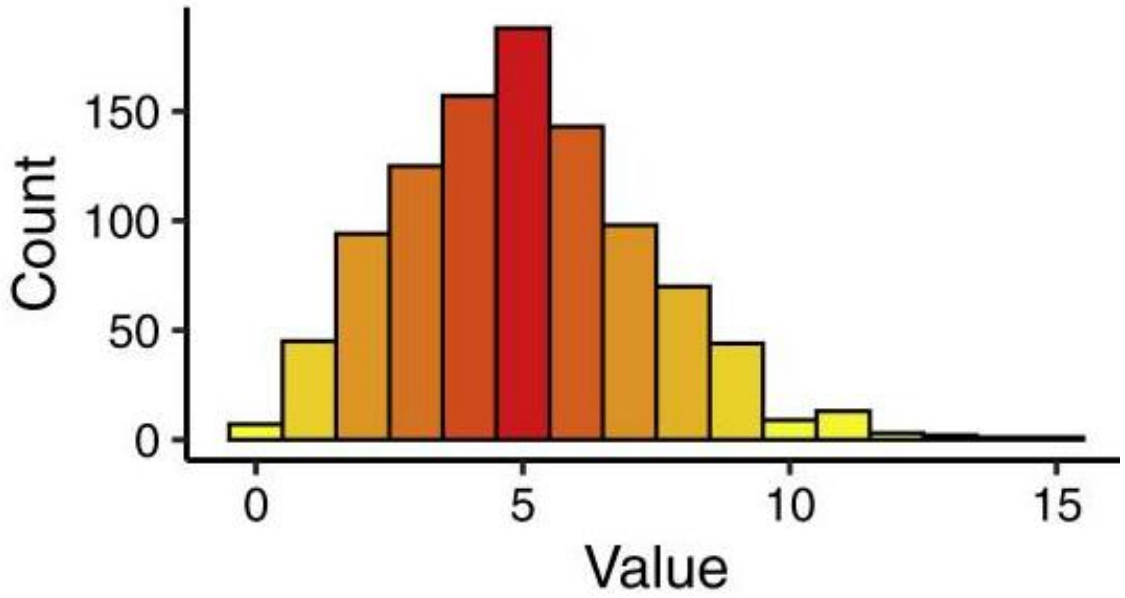


Place keys to data categories either to the right or on the top of a bar graph. If possible, the best way to save space is to place the key within the area of the bar graph itself.



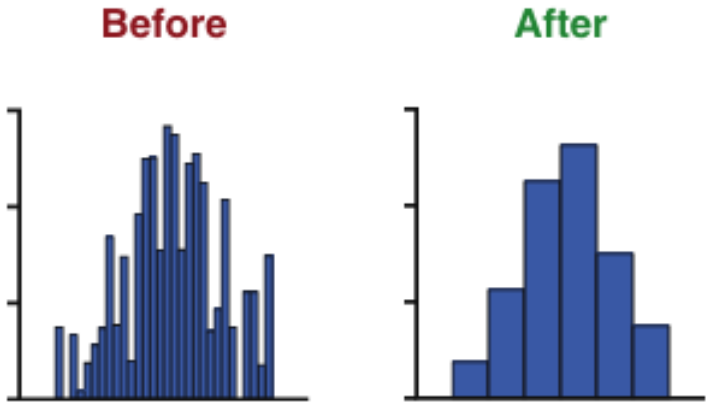
Designing histograms

A histogram shows the distribution of data and the relative frequency with which they occur.

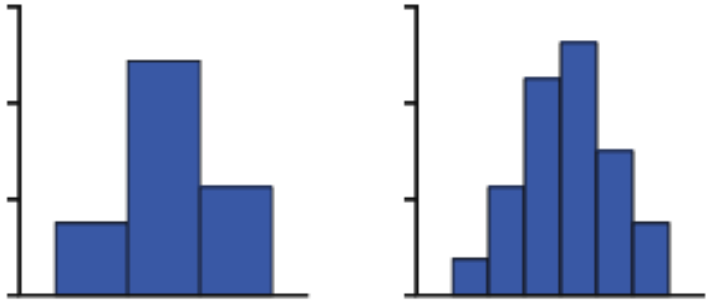


Anatomy of histogram

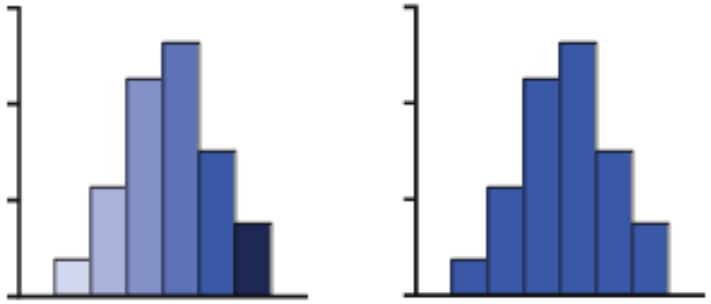
Group data into separate “bins” to increase the clarity of the overal trend and reduce the effects of outliers.



Depending on your data, it is usually best to have at least five bins. It is harder for readers to make conclusions about a dataset in a histogram with four or fewer bins.

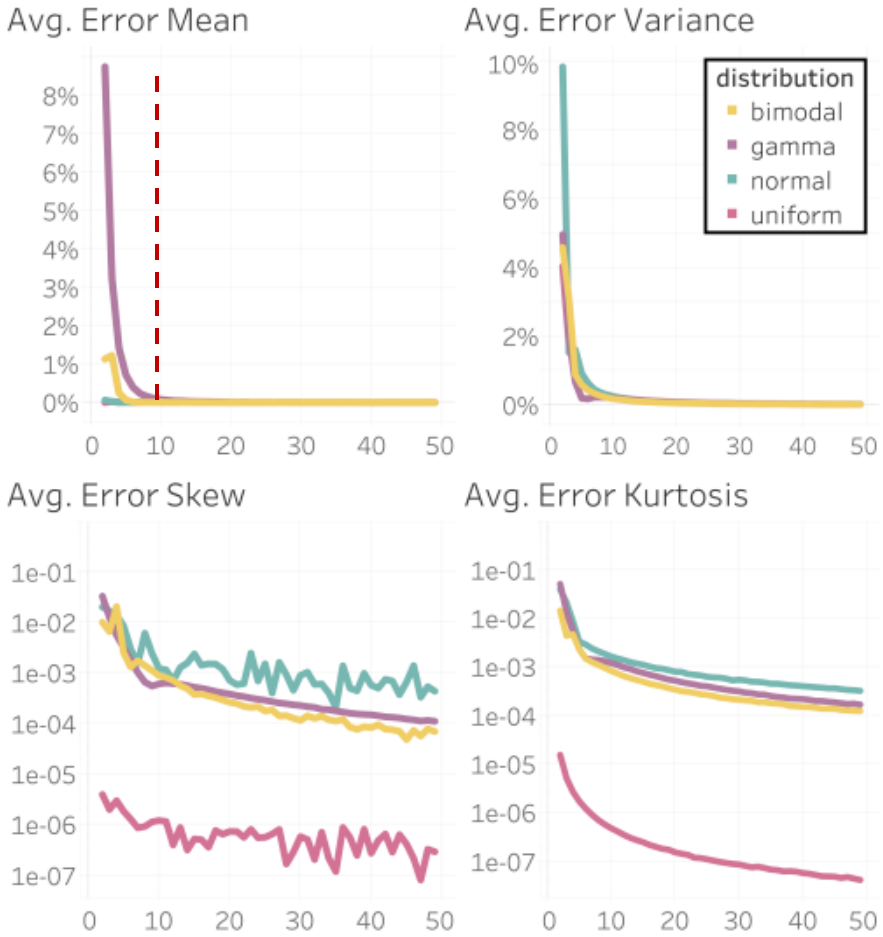


Don’t unnecessarily assign different colors to different bins. If the histogram represents a single dataset, use the same color for all bins throughout.



How many bins should be put in a regular histogram?

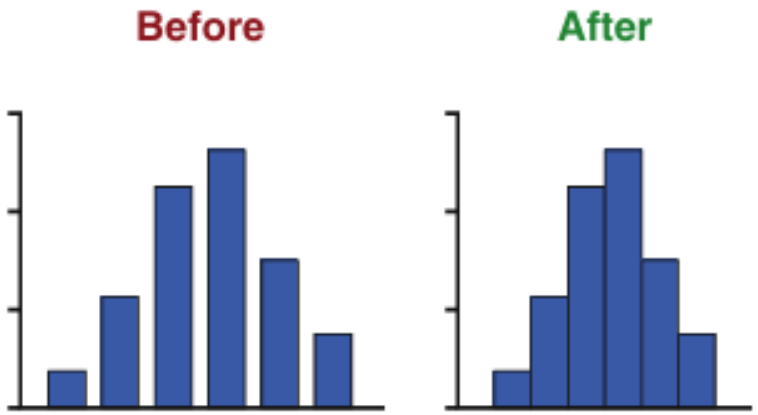
- For larger sample sizes, a larger number of bins increases the recognition of the correct distribution. However, beyond 20 bins, the detection rate does not increase significantly anymore.



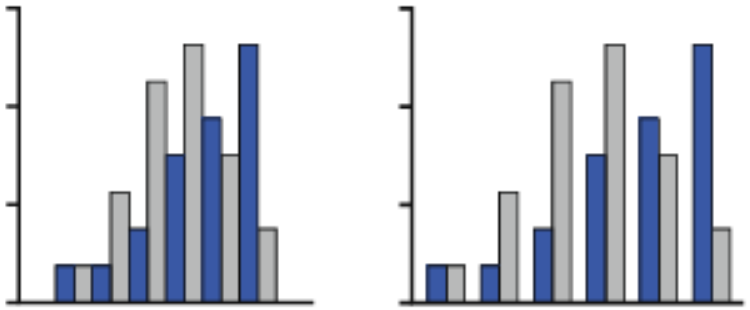
Sahann R, Müller T, Schmidt J. Histogram binning revisited with a focus on human perception. In2021 IEEE Visualization Conference (VIS) 2021 Oct 24 (pp. 66-70). IEEE.

Anatomy of histogram

Don't use spacing between bars on a histogram as you would on a bar chart. For frequency data, it is usually easier to perceive data as part of a unified dataset when the bars are placed directly adjacent to each other.

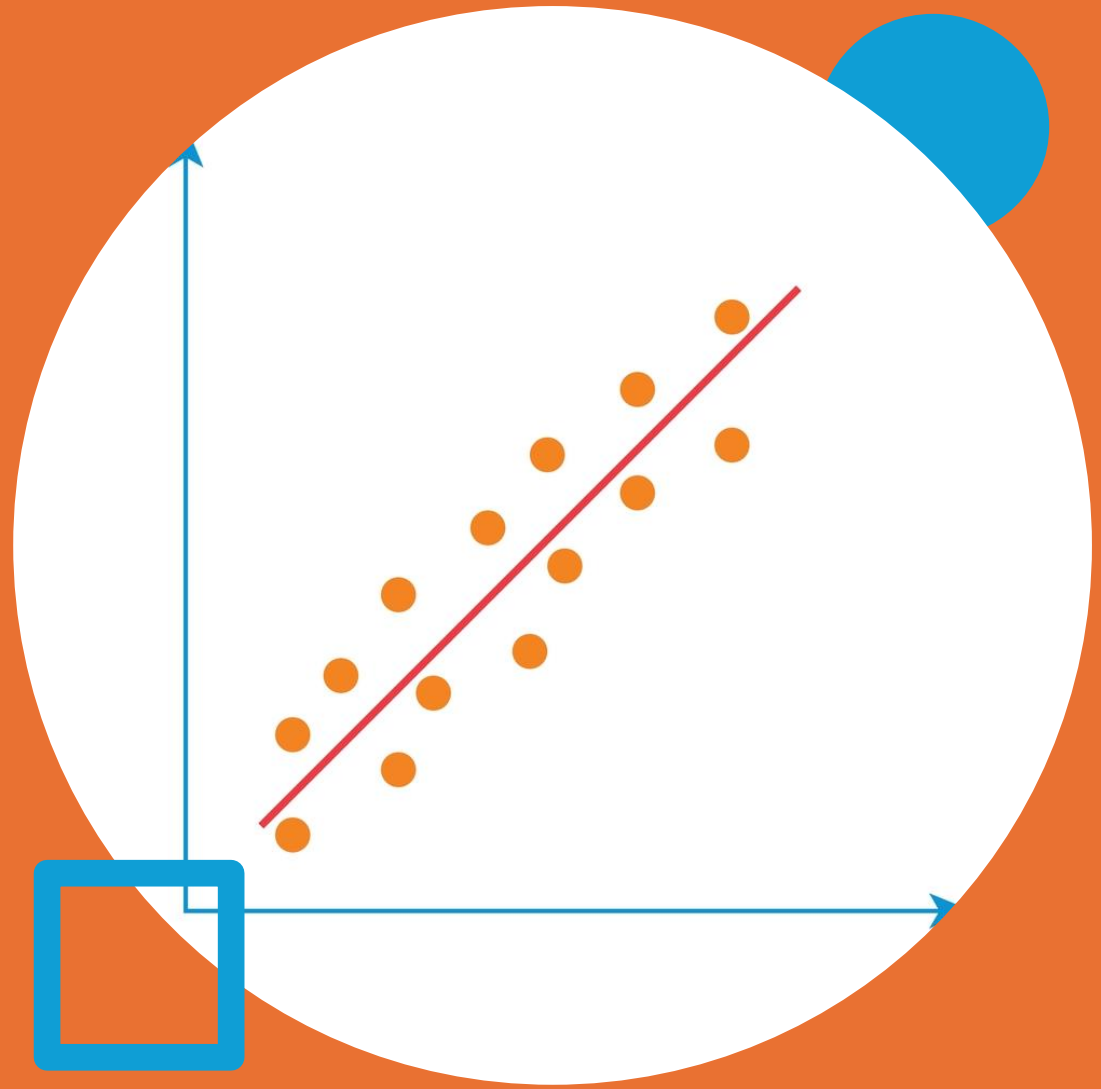


In contrast, use spacing between bars if there are two or more datasets plotted on the same histogram so that it is easier for readers to identify the bin value of each pair of bars...

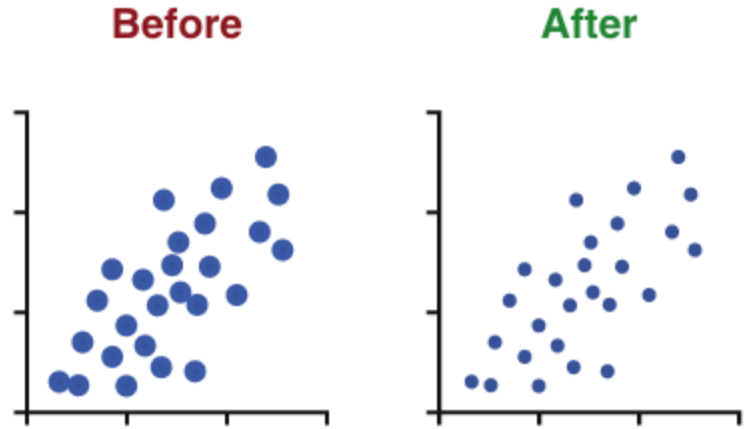


Designing scatterplots

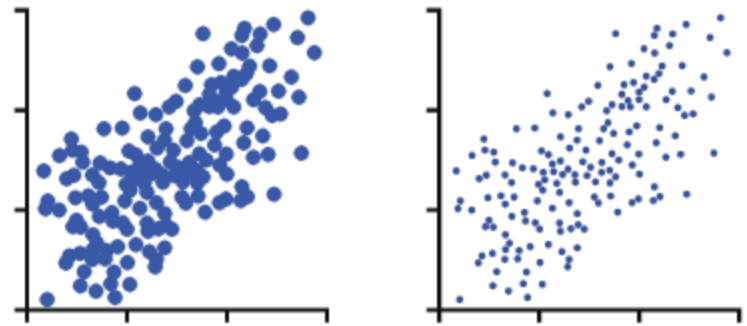
Scatterplots are used to show the relationship between two continuous variables.



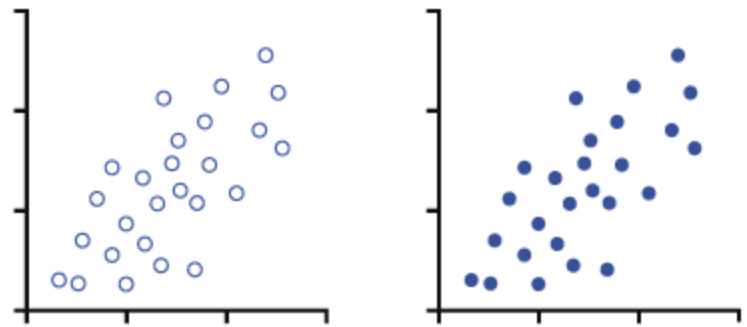
The size of the dots on a scatterplot depends on the quantity of your data. In general, a good starting point is to use dots that are 2-3 times larger than the line thickness of the axes.



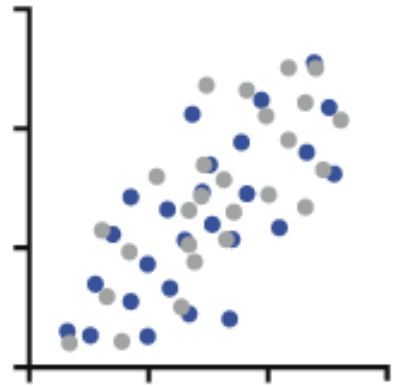
If you have a relatively large amount of data and it is hard to differentiate between separate dots, shrink the size of the dots by 1 pt. until most of the datapoints are distinct.

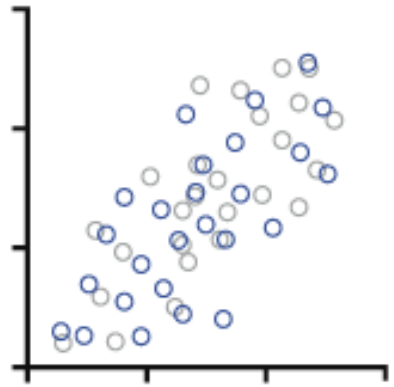


Closed (filled) dots are easier to perceive than open dots....

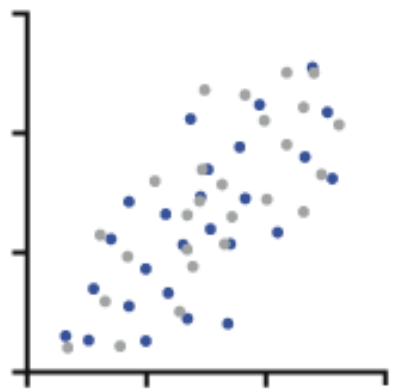


...however, when plotting two categories of data on the same scatterplot, open dots make it easier to view both datasets at once.





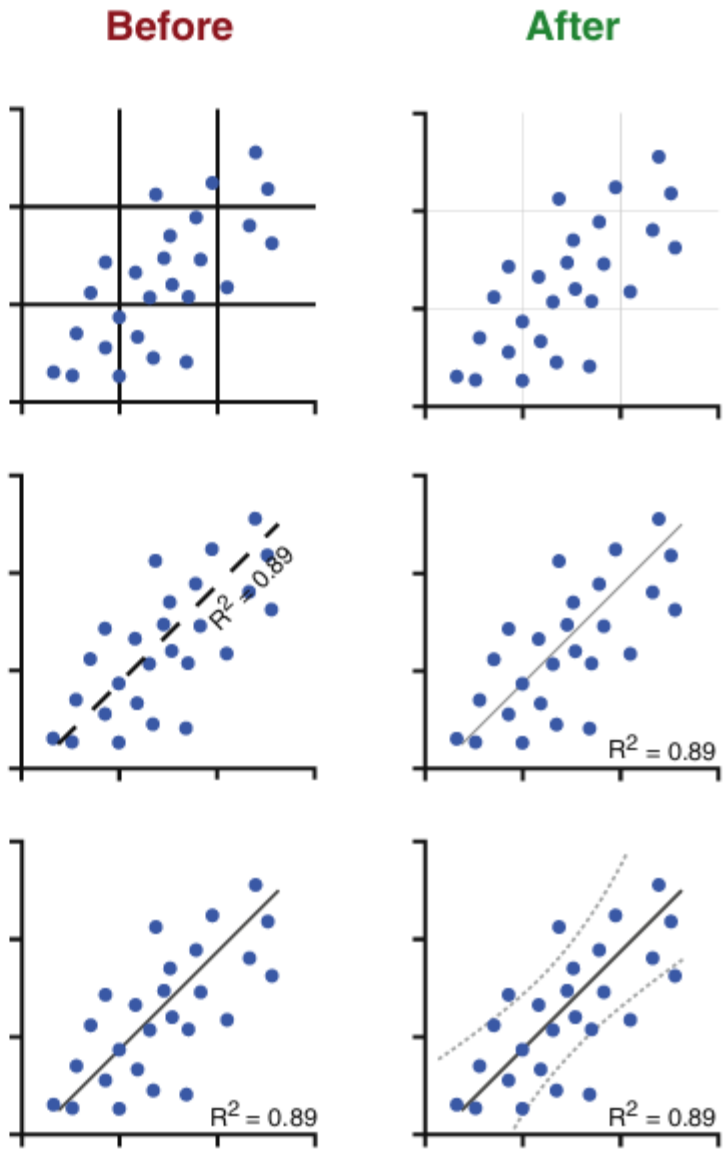
Alternatively, you might try decreasing the size of the dots.



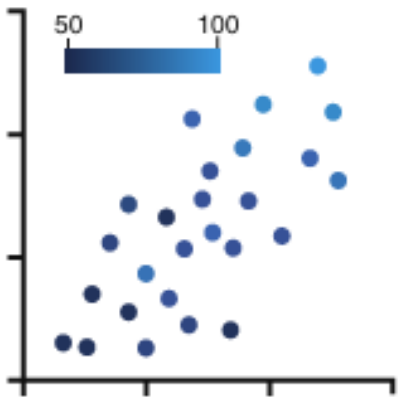
When you want to help your audience determine specific values of the individual datapoints, be cognizant about not overwhelming your scatterplot with distracting gridlines. Instead, place subtle gridlines in the background.

Use a line of best fit to represent a statistical statement about the relationship between the variables. Make sure the line stands out from the individual dots but does not overwhelm the scatterplot. Place statistical values somewhere on the chart where they do not clutter the data.

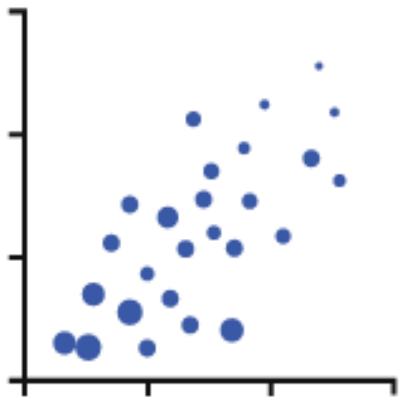
Consider using subtle lines to show other statistical parameters, such as 95% confidence intervals.



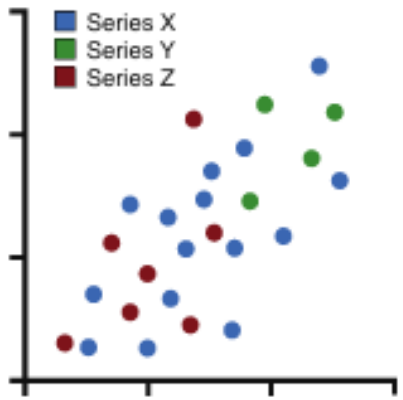
Scatterplots show the relationship between two continuous variables, but you can visualize a third variable by altering the color of the dots, or, in a bubble scatterplot, by altering the size of the dots.



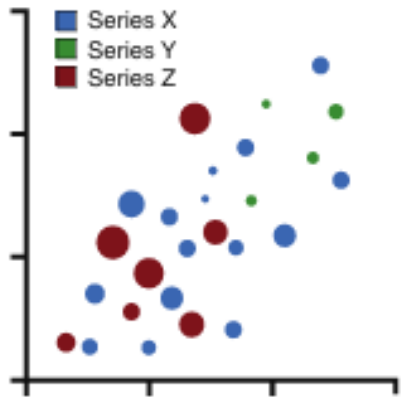
Depict a third, continuous variable via the color of the dots.



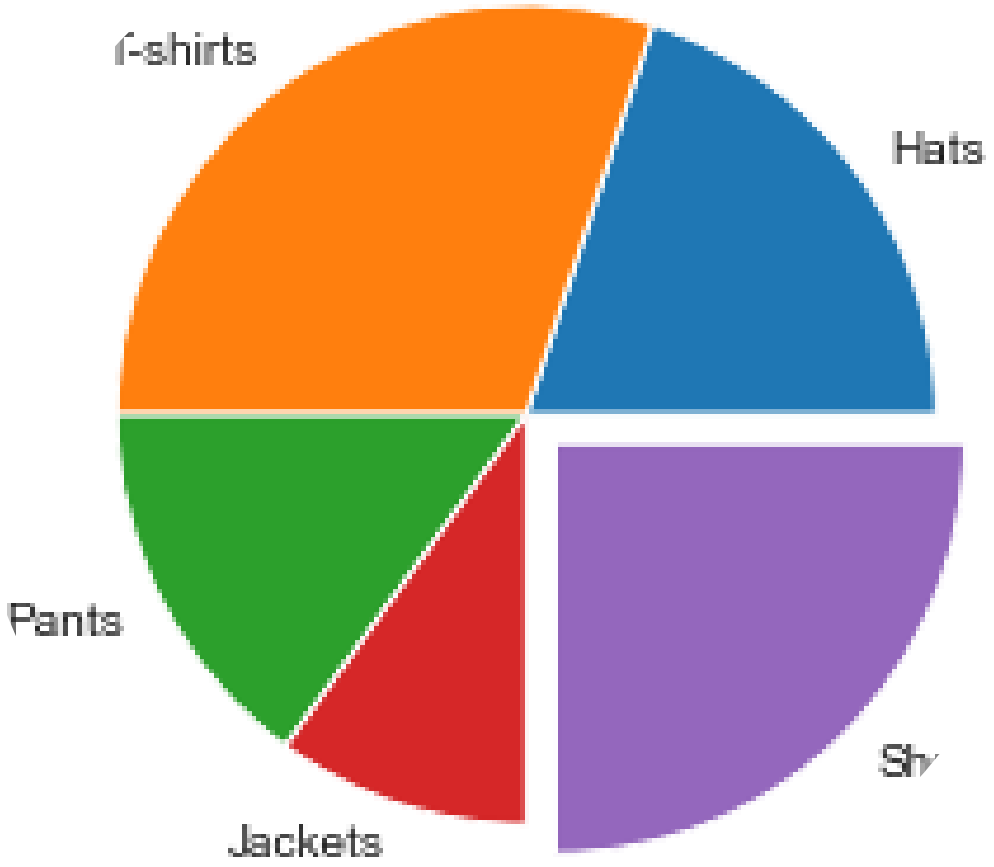
Depict a third, continuous variable via the area of the dots.



Depict different datasets via discrete colors of dots.



Depict different datasets via discrete colors of dots and a third, continuous variable via the area of the dots.



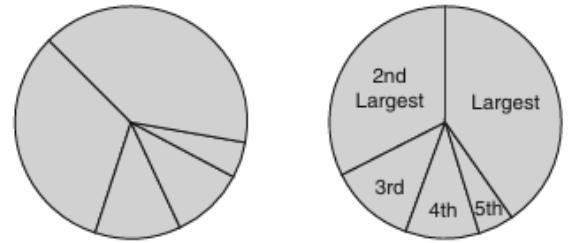
Designing pie charts

Use a pie chart to show relative proportions of a whole.

Before

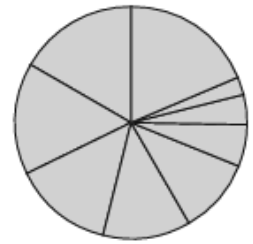
After

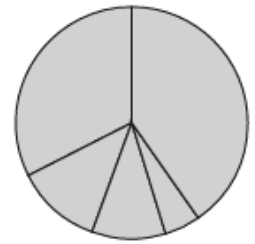
Try to start the largest slice of a pie chart at the 12:00 position. Usually it is easiest for audiences to perceive the relative proportions of the slices if the largest slice runs to the right of the 12:00 position, the second largest slice runs to the left of the 12:00 position, and the rest of the slices descend in value counter-clockwise.



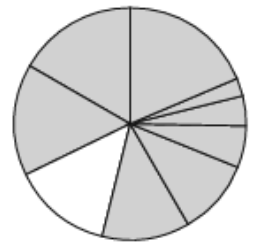
Assigning dark shades, bright hues, or multiple colors for different slices in a pie chart can distract your audience from the data within the chart. If it is necessary to apply different shades or colors (for example, when keeping different categories of data consistent from figure to figure), choose monochromatic colors, as described in Ch. 3, and try not to use distracting hues.

It is difficult for people to perceive differences between the sizes of slices if there are more than five or six slices. Otherwise, the visual distinction between different slices becomes meaningless.

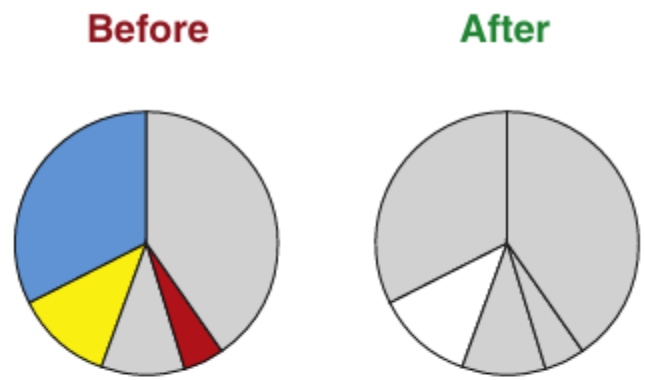




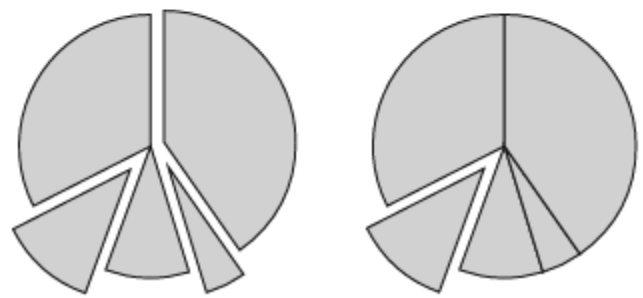
An exception occurs if you wish to highlight a single slice relative to all of the other slices in the pie.



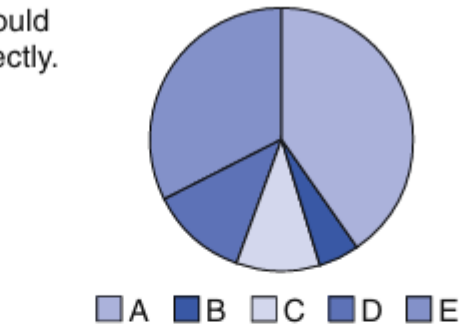
Try not to highlight more than one or two slices or to use highlighting colors that overwhelm the rest of the pie chart.

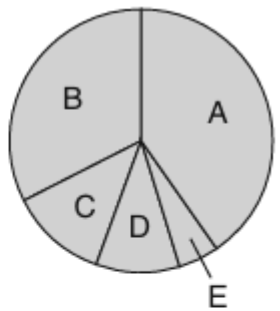


An alternative to highlighting a single slice with color is to use an “exploded pie” to emphasize a single slice. Don’t use this method to emphasize more than one slice at a time.



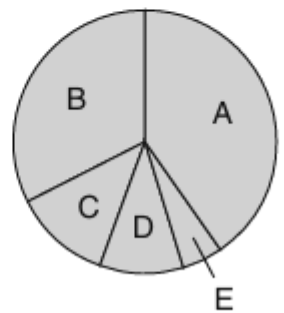
Instead of using a separate key (as you would for a line or bar graph), label the slices directly. If your label won't fit on the pie, place it immediately adjacent to the slice.

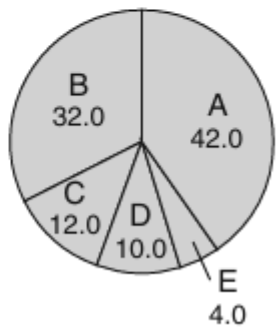




If you want to label each slice with the exact value or percentage of the whole, do so directly on the slice and not in a separate legend.

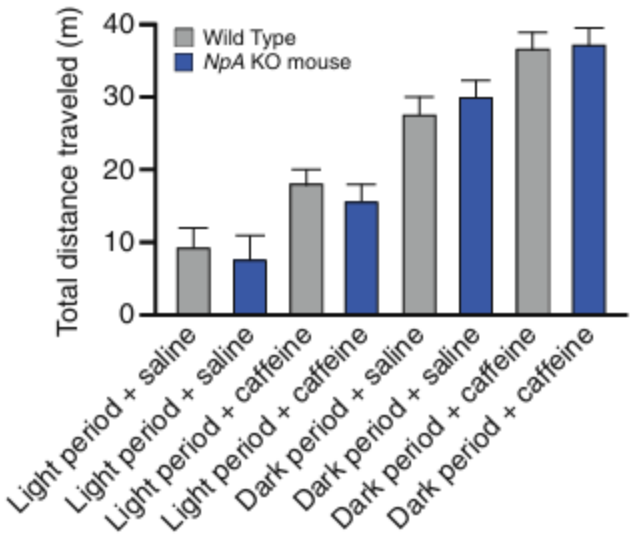
A: 42.0
B: 32.0
C: 12.0
D: 10.0
E: 4.0



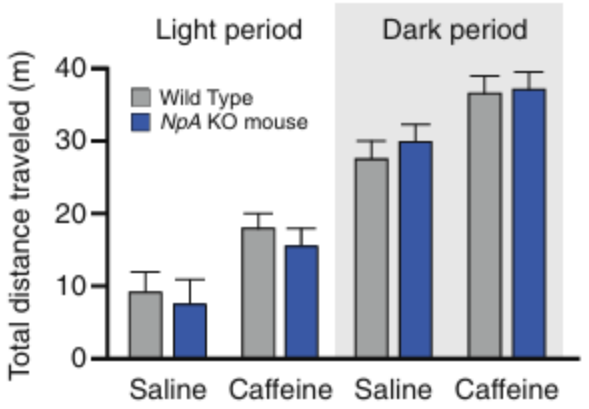


Help your audience visualize what is most important

Before

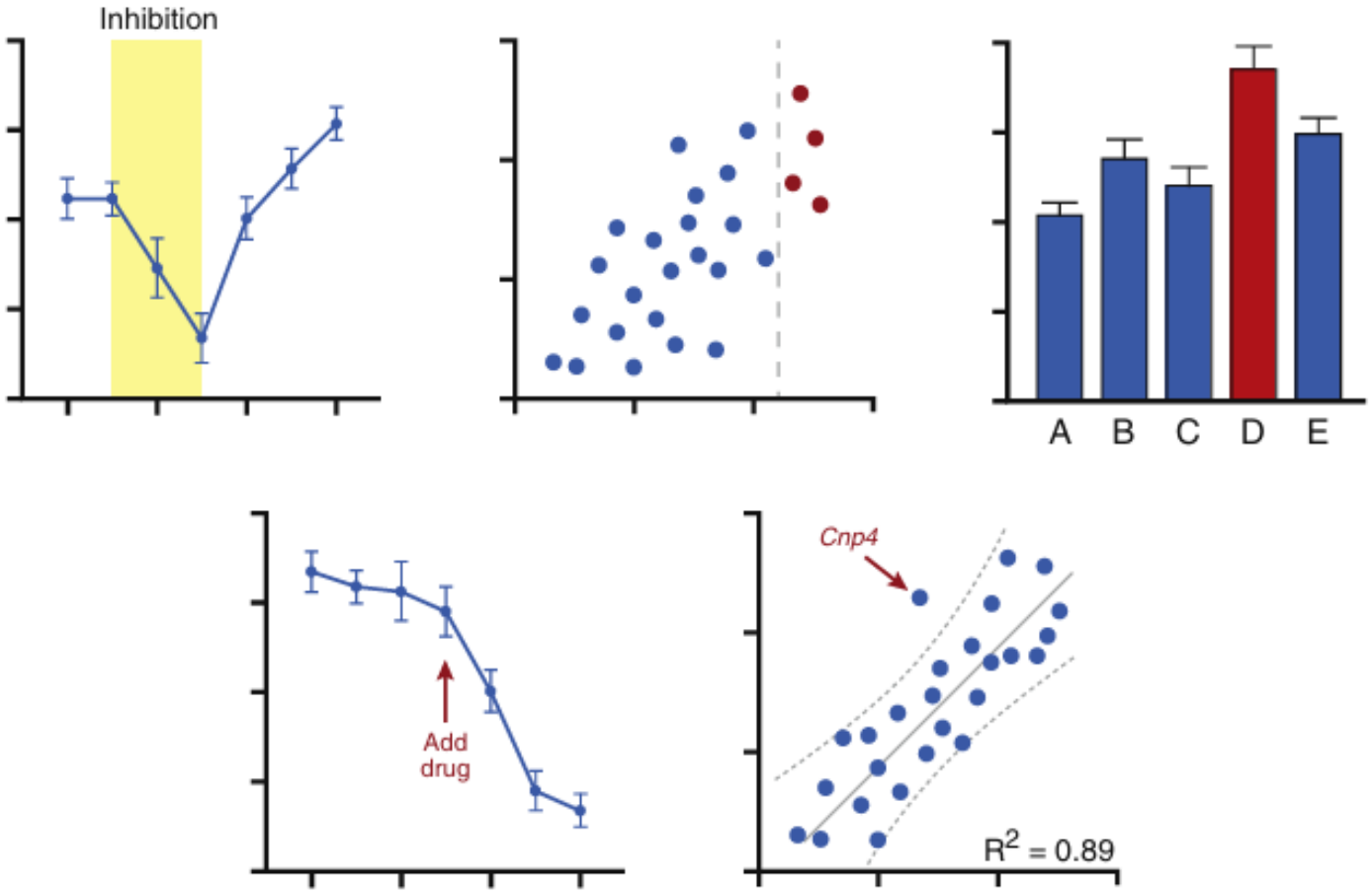


After



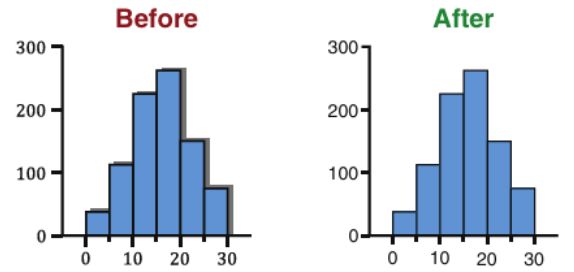
Help your audience visualize what is most important

When appropriate, highlight data points that are particularly meaningful to you.

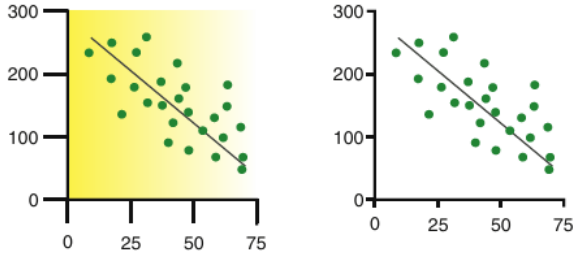


Reduce clutter, decoration, and distractions

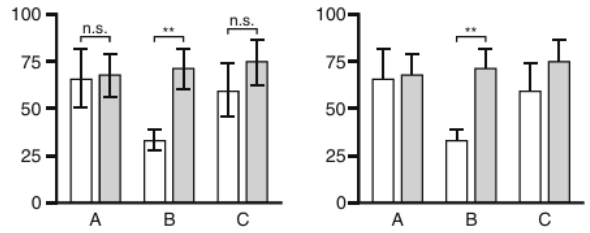
The chart on the right is more clear because the serif font has been changed to sans serif and the shadow behind the bars has been removed.



In the chart on the left, the gradient isn't necessary. Additionally, the length of the tick marks is too long relative to the size of the chart.



Error bars on bar charts only need to be placed above the bar. If you want to emphasize that there is no significant difference between two sets of data, "n.s." can be useful. However, in many cases, it is only necessary to emphasize when there is a significant difference.



Many presenters think that 3D graphs look more professional and exciting than 2D. In reality, 3D charts contain awkward corners, shadows, and viewing angles that obscure the representation of data and cause difficulties in interpretation.

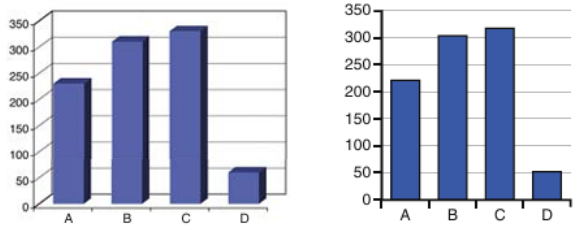


Figure legends

- Information to include in a figure legend includes:
  - A specific title
  - Definitions as to what various symbols and shapes represent, including datapoints and scalebars.
  - Definitions of abbreviations
  - *n* values, including the number of subjects, trials, sessions, etc.
  - Definitions of statistical significance and reporting of statistical tests used

A good example

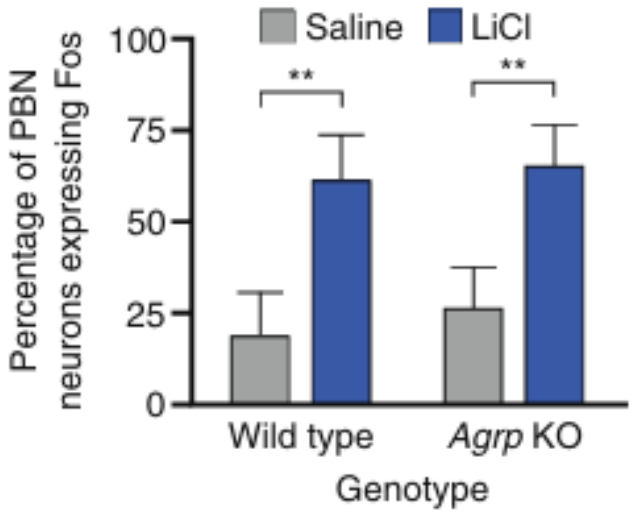


Figure 2. Intraperitoneal administration of LiCl causes an increase in the percentage of PBN neurons expressing Fos in both wild type and *Agrp* KO mice. n=8 animals per condition; **p<0.001, two-way ANOVA between genotype and treatment followed by Tukey posthoc test. AgRP, agouti-related protein; PBN, parabrachial nucleus.

The best graph titles not only describe experiments, they communicate the meaning of the results.

Before	Mass of rats over time
After	Rats on a high fat diet increase mass over time compared with rats fed normal chow
Before	Annual income versus socioeconomic background
After	Annual income positively correlates with socioeconomic background
Before	Effects of microstimulation of FEF on visual perception
After	Microstimulation of the FEF increases performance on a visual perception task
Before	Flies lacking *Smu2*
After	*Smu2*-null flies show no preference for sucrose versus quinine
Before	Average annual rainfall in Seattle during the 1990s
After	Average annual rainfall in Seattle remained stable during the 1990s
Before	Three-year prognosis after treatment
After	Increase in survival rate 3 years after treatment

Example prompt for a graph

Prompt

Write a python code to read my file data.csv, and make a modern, scatter plot, using the following setting:

x axis: Prolactin in logarithmic scale

y axis: TSH

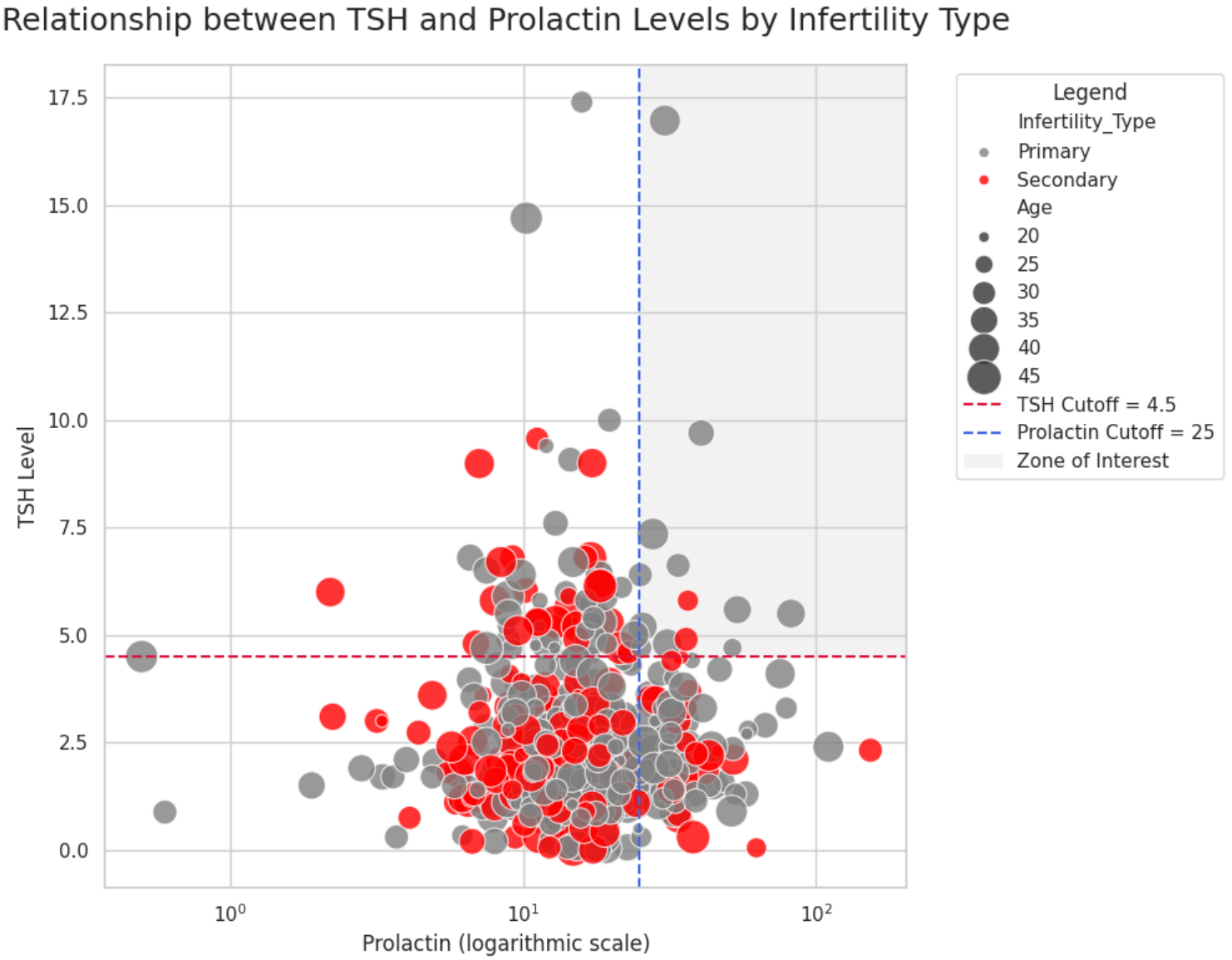
size: use Age as the size of the data points

Gridlines: use TSH cut off value of 4.5, and 25 for prolactin.

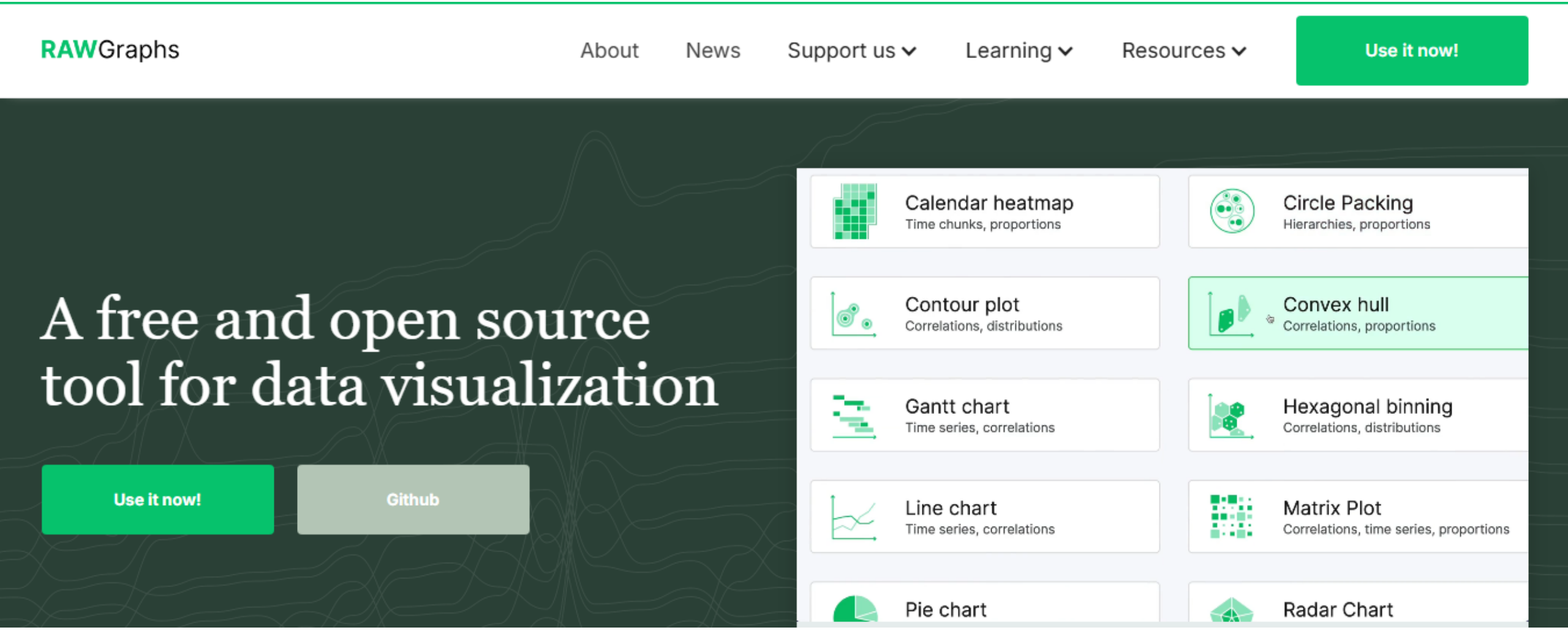
Zone of interest: Use a 10% shade of grey for the background color part that TSH > 4.5, and prolactin > 25. use red color for those data points with Infertility_Type=Secondary in this region.

- **Address**
  - https://colab.research.google.com/

Result


Relationship between TSH and Prolactin Levels by Infertility Type
TSH Level
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
10^0
10^1
10^2
Prolactin (logarithmic scale)
Legend
Infertility_Type
Primary
Secondary
Age
20
25
30
35
40
45
TSH Cutoff = 4.5
Prolactin Cutoff = 25
Zone of Interest

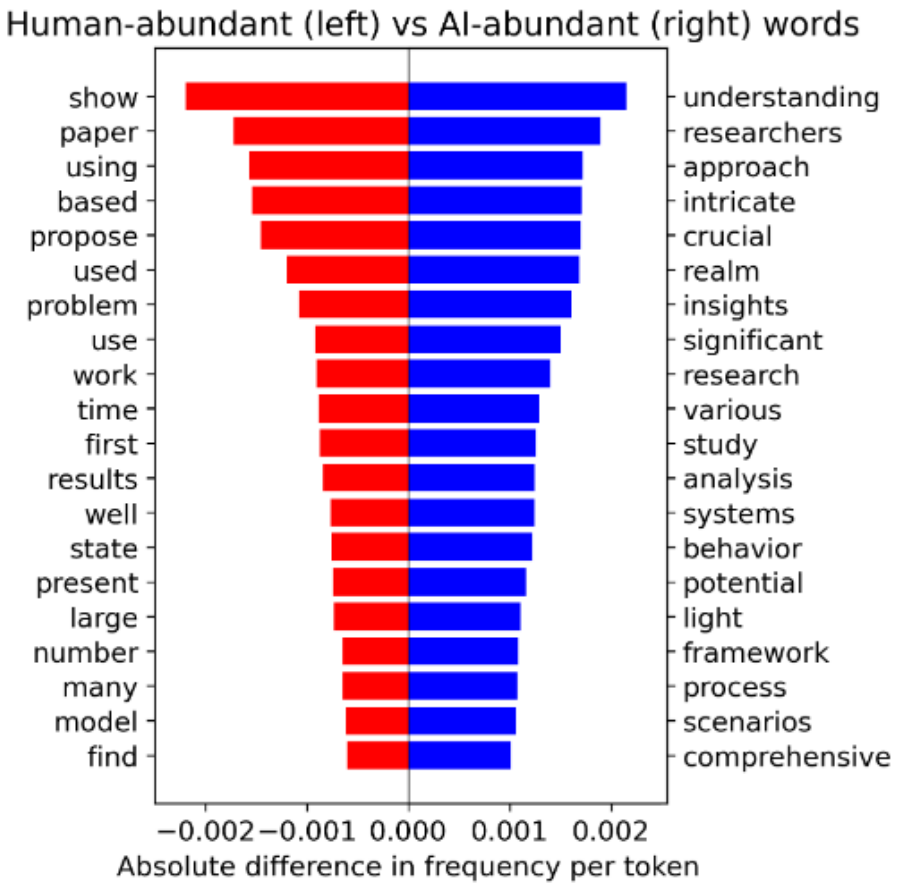
3. rawgraphs.io



Effective results writing

Try to avoid colloquialism and slang

Wordy	Concise
A total of …	*(Omit)*
All of …	All …
At the present time …	At present …
At this point in time …	At present …
Based on the fact that …	Because …
Both of …	Both …
During the course of …	During … or In …
Figure 1 shows that …	… (Figure 1)
For the purpose of …	For … or To …
Has been shown to be …	Is …
In light of the fact that …	Because …
In order to …	To …
In a small amount of …	Rarely … or Occasionally …
In previous years …	Previously …
In may be that …	Perhaps …
It should be mentioned …	*(Omit)*
It is interesting to note that	Of interest is …
More often than not …	Usually …
… quite unique	… unique
Really …	*(Omit)*
… small in size	… small
The reason is because …	The reason is …
… very …	*(Omit)*
… would seem to suggest …	… suggest …



Astarita S, Kruk S, Reerink J, Gómez P. Delving into the utilisation of chatgpt in scientific publications in astronomy. arXiv preprint arXiv:2406.17324. 2024 Jun 25.

Singular versus plural

- Probably the most misused plural word is “data”.
- The word “data” should always be used as plural.

Incorrect	Correct
The data shows	The data show
The data suggests	The data suggest
The data supports …	The data support …
The data is not …	The data are not …

Considerations

- **Active vs passive**: use the passive voice in scientific writing.
- **Verb tense**: In general, it is best to use the past tense to describe actions and experiments and the present tense to describe knowledge accepted as valid conclusions.

Verb tense

- **Abstract**
  - Use the present tense to introduce relevant background information.
  - Use the past tense to summarize what you did.
- **Introduction**
  - mostly consist of information in the present tense.
- **Methods and Results**
  - should be in the past tense.
- **Discussion**
  - a mixture of past and present.
- **Conclusion**
  - should be in the present tense

Commonly misused or incorrect words

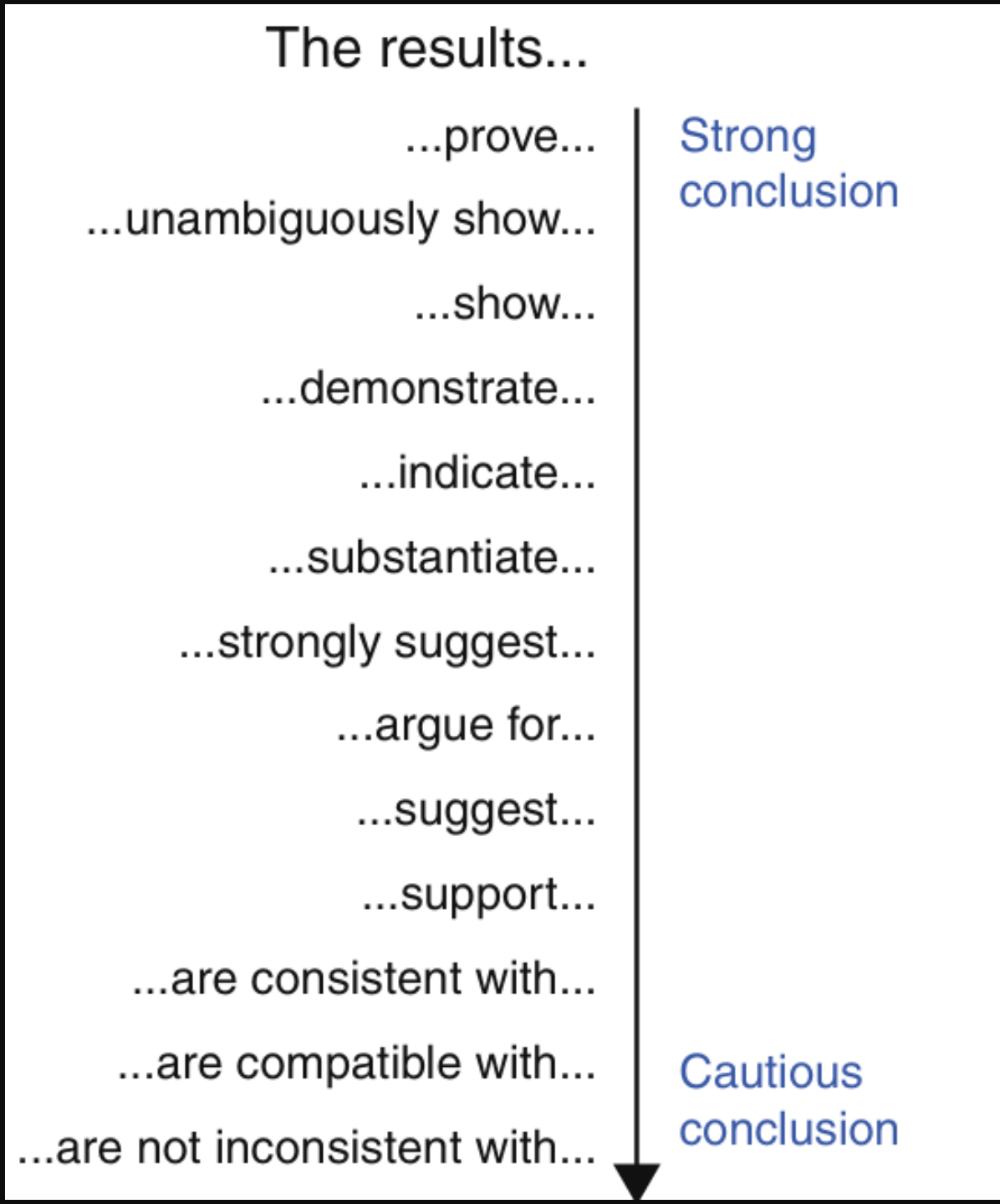
Adaption	*Adaption* is not a word. The correct word is *adaptation*.
As	Don't use *as* in place of the word *because*. *As* means *in the same way that …* or *at the moment that …* For example: "English as a second language," or "The egg cooked as the water boiled."
Comprise	The word *comprise* means *to contain*, as in "the solar system comprises the sun and planets. It does not mean *to constitute* or *to make up*.
Correlate to	Items/concepts might be related *to* each other, but they are always correlated *with* one another.
Could of	*Could of* is grammatically incorrect; the correct usage is *could have*.
Dilemma	*Dilemma* doesn't simply mean *a difficult problem* or a *quandary*. Instead, it means a decision between two equally good or bad choices.
Due to	The word *due* implies a debt or deadline. Therefore, don't use *due* when you mean to say *because of*.
Different than	Two items are different *from* each other, not different *than* each other.
Experience(d)	*Experience* implies sensation. Therefore, only use this word in reference to living, sensing creatures/objects, as opposed to statements like "the Earth experienced a cooling period," or "The forest experienced high winds."

Cont'd

Irregardless	Irregardless is not a word. Use *regardless*.
Literally	*Literally* means that *something is completely and totally true in a literal sense.* Because this word can be used for dramatic effect, sometimes people use it when they actually should use the word's exact opposite, *figuratively*, which means that something is not true in a literal sense.
Peruse	A common misconception is that *peruse* means *to browse or skim lightly*. It actually means the opposite: *to read with great care and attention to detail.*
Significant	In science presentations, only use *significant* in reference to statistical significance. When not describing statistics, consider using the words *substantial*, *notable*, or *remarkable*.
Since	Don't use *since* in place of the word *because*. Since connotes time, as in "I haven't seen you since this morning."
Thing	The word *thing* is okay to use when referring to an object, but many people use it to represent an abstract concept: "The thing I don't like about the paper …"; "The thing I like about your proposal …"; "Here's the thing …" Try your best to be more specific. Use thing if you feel you have to, but more specific words force you to communicate better with a reader.
This	Don't use *this* in isolation. The problem with *this* is that it isn't descriptive enough. This just won't do. Instead, be specific: "This concept …," "This question …," "This phenomenon …," etc.
Where	*Where* should always be used in reference to a location. Don't use *where* instead of *in which*, or *for which*. For example: "This is a protocol *in which* we need to pay careful attention to detail.
While	*While* should always be used in reference to time. Don't use *while* instead of *and, but, although,* or *whereas*. For example: "The inner planets are mostly rocky, whereas the outer planets are mostly gaseous."

The burden of proof

Consider the possibility that future studies may present contradictory findings that challenge your conclusions.



Guidelines for writing about numbers

- Write out single-digit whole numbers as words and use numerals for numbers greater than nine.
- Be consistent when writings words versus numerals for a category of information.
- Never start a sentence with a numeral.

We used eight mice in this study.
We used 26 mice in this study.

This study consisted of 6 girls and 14 boys.

Guidelines for writing about numbers

- Use numerals to report statistics, provide quantitative data using units of measurement, and when describing dates, times, pages, figures, and tables.
- Always spell out simple fractions.
- Hyphenate all compound numbers from twenty-one through ninety-nine.

Only 5% of neurons transduced with GFP showed co-expression with Fos (Figure 1). All stimulation episodes lasted 5 h (between 13:00 and 18:00).

One-half of the mice were fed chow and one-half were fed sucrose pellets.

Sixty-three adult males took part in the study.

Description of tables and figures

- Example prompt:

I am working on the relationship between age, sun exposure duration, season of birth, and vitamin D level in patients with multiple sclerosis (MS). Look at the attached table, act as a researcher and try to describe the findings and patterns. Also conform the following the rules:

Rules

1. The word “data” should always be used as plural.
2. avoid colloquialism and slang.
3. use past tense and passive voice.
4. Consider the possibility that future studies may present contradictory findings that challenge your conclusions; thus, be cautious about the burden of proof you use.
5. Write out single-digit whole numbers as words and use numerals for numbers greater than nine.
6. Be consistent when writings words versus numerals for a category of information.
7. Never start a sentence with a numeral.
8. Use numerals to report statistics, provide quantitative data using units of measurement, and when describing dates, times, pages, figures, and tables.

Hint & examples

1. The results section should be written such that each 1-2 paragraphs exists as a "mini-paper," complete with its own rationale, statement of methods, results, and conclusion.
2. Example:
    1. To determine if Tav1 interacts with Farr3 and Farr6, the full-length versions of these proteins in HEK293T cells overexpressed and co-immunoprecipitation experiments performed using an antibody to Tav1. it was found that Tav1 interacted with both Farr3 and Farr6, indicating that Tav1 interacts with either receptor ex vivo.

Table

Age, sun exposure duration, season of birth, and vitamin D level in three groups of participants

Demographic characteristics	MS group	Siblings	Controls	P
Age (mean ± SD)	30.3 ± 7.5	31.2 ± 8.7	31.1 ± 8.2	0.8
Education level (year) (mean ± SD)	13.8 ± 2.6	13.4 ± 3.1	14.5 ± 3.1	0.2
Daily sun exposure duration (min)	52.5 ± 41.4	60.0 ± 48.4	59.5 ± 36.5	0.6
Serum vitamin D level (nmol/l) (mean)	9.7 ± 7.9	9.4 ± 9.9	7.5 ± 11.6	0.5
Season of birth				0.3
Spring	19	11	11	
Summer	15	15	15	
Autumn	6	8	11	
Winter	5	11	8	
Median	6.5	6	4.2	
Vitamin D status				0.2
Deficient	39	38	42	
Insufficient	6	6	1	
Sufficient	0	1	2	

Main references