Which of these two graphs do you prefer and why?

Mentimeter: Which graphic do you prefer and why?

African Countries by GDP

South Africa
Egypt
Nigeria
Algeria
Morocco
Angola
Libya
Tunisia
Kenya
Ethiopia
Ghana
Cameroon

https://visual.ly/community/Infographics/economy/african-countries-gdp
Learning outcomes

Graphics is communication

Strengths and challenges of human perception

Three principles of effective communication:

- Have a clear purpose
- Show the data clearly
- Make the message obvious

Apply knowledge in two mini-makeovers
The goal of graphics is **communication**

“A graph is **more effective** than another if its quantitative information can be **decoded** more **quickly** or more **easily** by most observers.”

Nancy Robbins, *Creating More Effective Graphs*
The communication process

Sender → Encoding → Channel → Decoding → Receiver

Feedback
Graphics can speed up knowledge acquisition by:

A. Playing to strengths of human brain function:
   I. Pre-attentive visual processing
   II. Pattern recognition

B. Accounting for challenges of human perception
   I. Perceive relative differences (not absolutes)
   II. Assign meaning depending on context (e.g., red and blue in US vs UK)
   III. Variation in ability (e.g., color vision deficiency)
A. Perception is sometimes **SLOW** …

How many times does ‘5’ appear above?
… and sometimes FAST!

How many times does ‘5’ appear above?
SYSTEM 1
Intuition & instinct

95%

Unconscious
Fast
Associative
Automatic pilot

SYSTEM 2
Rational thinking

5%

Takes effort
Slow
Logical
Lazy
Indecisive

Source: Daniel Kahneman
A.i. Take advantage of pre-attentive visual processing

Stephen Few's Information Dashboard Design
A.i. Take advantage of pre-attentive visual processing

- **Very precise quantitative perception**
  - Length
  - Width
  - Size
  - Intensity
  - Blur

- **Not very precise quantitative perception**
  - Orientation
  - Shape
  - Enclosure
  - Added marks

- **No quantitative perception**
  - 2D position

Stephen Few's Information Dashboard Design
A.ii. Humans seek patterns

SIMILARITY
Objects that share similar attributes (e.g., color or shape) are perceived as a group.

PROXIMITY
Objects that are close together are perceived as a group.
A.ii. Humans seek patterns

**ENCLOSURE**
Objects that appear to have a **boundary around them** (e.g., formed by a line or area of common color) are perceived as a group.

**CONNECTION**
Objects that are **connected** (e.g., by a line) are perceived as a group.

https://emeeks.github.io/gestaltdataviz/section1.html
A.ii. Humans seek patterns

**CLOSURE**
Open structures are perceived as closed, complete, and regular whenever there is a way that they can be reasonably interpreted as such.

**CONTINUITY**
Objects that are aligned together or appear to be a continuation of one another are perceived as a group.
B.i. We perceive *relative difference*

Simultaneous contrast can make the *same* colors look *different*
B.i. We perceive **relative difference**

Simultaneous contrast can make **different** colors look the **same**

http://www.engr.colostate.edu/ECE666/Handouts/WritingPapers/UsingColorEffectively.pdf
B.ii. Understand that **meaning** depends on **context**

https://www.informationisbeautiful.net/visualizations/colours-in-cultures/
B.iii. Account for variation in perception

Red–green color blindness affects up to 8% of males and 0.5% of females of Northern European descent.

THREE PRINCIPLES OF EFFECTIVE COMMUNICATION

1. Have a clear purpose
2. Show the data clearly
3. Make the message obvious
THREE PRINCIPLES OF EFFECTIVE COMMUNICATION

1. Have a clear purpose
2. Show the data clearly
3. Make the message obvious
1. Have a clear purpose

a) Understand the question you are trying to answer

b) Identify the quantitative evidence to answer that question

c) Know your audience and focus the design to support their needs
1a. Understand the question you are trying to answer

- Data exploration,
- Convey information,
- Deliver a message,
- Convince an audience,
- Support a decision?
- Any / all of the above .. ?
1b. Identify quantitative evidence to answer that question

Collect new data

Find existing data

...
1c. Know your audience; focus design to support them

The audience is not you!

https://www.momtastic.com/
1c. Know your audience; focus design to support them

A. Public

B. Academic

African Countries by GDP

- South Africa: $285.4 b
- Egypt: $188.4 b
- Nigeria: $173 b
- Algeria: $140.6 b
- Morocco: $91.4 b
- Angola: $75.5 b
- Libya: $62.3 b
- Kenya: $39.6 b
- Ethiopia: $28.5 b
- Ghana: $26.2 b
- Cameroon: $22.2 b

GDP Calculation:
Gross Domestic Product (GDP) is the market value of all final goods and services produced within a country in a given period (2000 - 2005).

https://visual.ly/community/Infographics/economy/african-countries-gdp
Makeover #1
Mentimeter: Worst or best graphic ever … ? and why?

(1) Worst graph ever

(10) Most amazing graph ever
[GOOD] Communication to wide audience

**Title:** Clear, but no message

**Message:** More red recently = hotter

**Color:** Harder to make direct quantitative comparisons

**Good use of** diverging color scheme

**X axis:**
- Clear annual scale
Compare scientific journal graphic with public-facing graphic

- Shows data on 2 axes
- Shows uncertainty
- Describes data sources
- Longer times series, maybe harder to understand pace of change?
THREE PRINCIPLES OF EFFECTIVE COMMUNICATION

1. Have a clear purpose
2. Show the data clearly
3. Make the message obvious
2. Show the data clearly

a) Choose the appropriate graph type to display your data
b) Avoid misrepresentation (use appropriate scales)
c) Maximize data to ink ratio (reduce distraction, less is more)
Encoding numeric vs categorical data

**NUMERIC data**
- Points
- Lines
- Bars
- Color

**CATEGORICAL data**
- Shape/symbol
- Line type
- Color

Very precise quantitative perception
- Length
- 2D position

Not very precise quantitative perception
- Width
- Size

No quantitative perception
- Orientation
- Shape
- Intensity
- Enclosure
- Blur
- Added marks

*Stephen Few's Information Dashboard Design*
2a. Choose appropriate graph type to display data

TOOL TIP!
https://www.data-to-viz.com/
2a. Choose appropriate graph type to display data

- Pie
- Vertical bar
- Line
- Scatter

 TOOL TIP!
2a.i. Pie charts

Good if you have only few (2, 3) groups

Parts sum to whole (100%)

Start at 12 noon

Order parts

Stacked bar plots may be better than pies

2a.ii. Bar charts ...

Rank bars by the same attribute
Baseline = 0
Wide bars (>2x white space)
Same color
Horizontal text ...

… Horizontal bar charts

Plot negative bars to left

Bar plots with many bars ... dot plots

No need to start baseline at 0

Life expectancies of countries in the Americas, for the year 2007. Source: https://clauswilke.com/dataviz/
2a.iii. Line charts

Use sensible y-axis: Line covers 2/3 y-axis range

Label lines directly

Only shade area below if baseline = 0

Use color and width for different lines

>4 lines = small multiples

Monthly submissions to the preprint server bioRxiv, shown as a line graph without dots. Omitting the dots emphasizes the overall temporal trend while de-emphasizing individual observations at specific time points. It is particularly useful when the time points are spaced very densely. Source: https://clauswilke.com/dataviz/
Use small multiples

Align panels

Identical x- and y- axis range

Trends in Bachelor’s degrees conferred by U.S. institutions of higher learning. Shown are all degree areas that represent, on average, more than 4% of all degrees. Source: https://clauswilke.com/dataviz/
2a.iv. Scatterplots

Use sensible axes, usually low to high

Large points, identified

Clear axis labels, with units
2b. Avoid misrepresentation: start bars at 0
2b. Avoid misrepresentation: don’t distort data

- Genome size (Gbp) ~ Area
- Genome size (Gbp) ~ Length

- Wheat
- Human
- Arabid.
2b. Avoid misrepresentation: don’t go against convention

Here, y-axis goes **DOWN** ..
2c. Maximize data:ink, within reason

Before

After

Remove to improve (the data-ink ratio)
2c. Maximize data:ink (reduce distraction, less is more)

Data Looks Better Naked Series

Remove to improve the data:ink ratio

Remove to improve the data tables edition

Remove to improve the pie chart edition

BAR CHARTS

DATA TABLES

PIE CHARTS

CHOROPLETH MAPS

EXAMPLES!

https://www.darkhorseanalytics.com/portfolio-data-looks-better-naked
2. Display the data clearly

Try and show as much raw data as possible
Makeover #2
Mentimeter: Worst or best graphic ever … ? and why? What improvements would you make?

(1) Worst graph ever

(10) Most amazing graph ever
[BAD!] Main problem: more color than information

**Message:** What is the message?

**Data:**
- Time series data should usually be lines and not bars
- 3D bar graphs are hard to interpret and see all the data

**Y axis:**
- Too many 0s … 4.5M

**X axis:**
- Vertical text is hard to read
- Why use the hyphen?

**Image:**
- Too many bananas!

**Color:**
- Years are color-coded; should be country
Remake original …

Ecuador exports a lot of bananas!
THREE PRINCIPLES OF EFFECTIVE COMMUNICATION

1. Have a clear purpose
2. Show the data clearly
3. Make the message obvious
3. Make the message obvious

a) Minimize mental arithmetic

b) Use proximity and alignment to aid in comparisons

c) Use colors and annotations to highlight important details
3a. Add meaningful information to tell the whole story

Reference lines,
Benchmark effects,
Inferences,
Variation, etc.

Confidence intervals widen with smaller sample size. Chocolate bars from Canada and Switzerland have comparable mean ratings and comparable standard deviations (indicated with simple black error bars). However, over three times as many Canadian bars were rated as Swiss bars, and therefore the confidence intervals (indicated with error bars of different colors and thickness drawn on top of one another) are substantially wider for the mean of the Swiss ratings than for the mean of the Canadian ratings.

https://clauswilke.com/dataviz/
3a. Minimize mental arithmetic: Aid comparisons
3a. Minimize mental arithmetic: Label data directly
3a. Minimize mental arithmetic: Use effective redundancy

https://clauswilke.com/dataviz/
3a. Minimize mental arithmetic: Use clearly different colors, symbols, etc.
3a. Minimize mental arithmetic: Avoid 3D

3D distorts the data

The same 3D pie chart shown from four different angles.

Rotating a pie into the third dimension makes pie slices in the front appear larger than they really are and pie slices in the back appear smaller.

https://clauswilke.com/dataviz/
3a. Minimize mental arithmetic: Use sensible axes

Tick marks at natural increments (e.g., 0.2, 0.25, 1, 2, 5, 10, 25, 50, 100, … )

Keep to same decimal place

Maybe include hundred, millions, etc in axis title

Use scales that people understand …
3a. Minimize mental arithmetic: Use larger axis labels
3b. Use proximity and alignment to aid in comparisons

Align small multiples

Same axes and limits on all panels

Same line color and type (and symbol, points, etc) across panels/figures

https://clauswilke.com/dataviz/
3c. Use colors to highlight important details

“avoiding catastrophe becomes the first principle in bringing color to information:

Above all, do no harm.”

Do not encode too much or irrelevant information
THE USE OF COLOR IN DATA VISUALIZATION

SEQUENTIAL
color is ordered from low to high

DIVERGING
two sequential colors with a neutral midpoint

CATEGORICAL
contrasting colors for individual comparison

HIGHLIGHT
color used to highlight something

ALERT
color used to get reader’s attention

https://clauswilke.com/dataviz/
Do account for **color-vision deficiencies**

Approximately **8% of males** and **0.5% of females** suffer from some sort of color-vision deficiency.

A red–green contrast becomes indistinguishable under red–green cvd (deuteranomaly or protanomaly):

- **Original**
  - Red
  - Orange
  - Yellow
  - Green
  - Brown
  - Gray
- **Deuteranomaly**
  - Brown
  - Gray
- **Protanomaly**
  - Gray
- **Tritanomaly**

https://clauswilke.com/dataviz/color-pitfalls.html#not-designing-for-color-vision-deficiency

Please, just don’t …
THREE PRINCIPLES OF EFFECTIVE COMMUNICATION

1. Have a clear purpose
2. Show the data clearly
3. Make the message obvious
R CODE: The evolution of a ggplot

The Evolution of a ggplot

CODE TIP!

https://www.cedricscherer.com/2019/05/17/the-evolution-of-a-ggplot-ep.-1/
Resources
Examples of good charts

https://www.cedricscherer.com/2021/05/09/contributions-30daychartchallenge-2021/
Books and guides


Data Visualization 101 ebook: https://visage.co/content/data-visualization-101/

Wickham & Grolemund. R for Data Science. Chapter 3: Data Visualization https://r4ds.had.co.nz/data-visualisation.html
Tools for making graphs: R

https://swirlstats.com/
https://r-graphics.org/
https://stat545.com/graphics-overview.html
https://www.cedricscherer.com/2019/08/05/a-ggplot2-tutorial-for-beautiful-plotting-in-r/
https://www.cedricscherer.com/2019/05/17/the-evolution-of-a-ggplot-ep.-1/
Tools for making graphs: Others

Most software has horrible defaults. Datawrapper is actually ok!

Data Wrapper: https://www.datawrapper.de/

Tableau: https://www.tableau.com/academic/students

Excel: https://stephanieevergreen.com/how-to/

Canva: https://www.canva.com/
Color

Comprehensive list of color palettes in r: https://github.com/EmilHvitfeldt/r-color-palettes

Palettes: https://colorbrewer2.org/

Scales: https://clauswilke.com/dataviz/color-basics.html

Pitfalls: https://clauswilke.com/dataviz/color-pitfalls.html
Cheatsheets

Design Principles
https://stephanieevergreen.com/data-visualization-checklist/

ggplot
How do you get actionable insights?

1. **Research**
   - Identify problem, need, & data

2. **Edit**
   - Understand, clean, & shape your data

3. **Plot**
   - Apply design principles to choose chart, add layers, text, refine

4. **Review**
   - Check for errors, areas of confusion, opportunities for improvement
Research

Who is your audience?
What do they need?
What is your question?
Is it timely? Is it relevant?
Can you answer it?
Is there data? How good is it?
Can you get it? How quickly? At what cost?
Recently, you may have noticed that global warming has been getting some attention in the news. Global warming refers to the idea that the world's average temperature has been increasing over the past 150 years, may be increasing more in the future. *Age - 3 Categories Crosstabulation*

<table>
<thead>
<tr>
<th>Age - 3 Categories</th>
<th>18-35 years</th>
<th>35-54 years</th>
<th>55+ years</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>94.4%</td>
<td>14.6%</td>
<td>15.7%</td>
<td>14.1%</td>
</tr>
<tr>
<td>Don't know</td>
<td>14.8%</td>
<td>12.6%</td>
<td>13.7%</td>
<td>13.6%</td>
</tr>
<tr>
<td>Yes</td>
<td>75.5%</td>
<td>73.0%</td>
<td>70.6%</td>
<td>72.3%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Recently, you may have noticed that global warming has been getting some attention in the news. Global warming refers to the idea that the world's average temperature has been increasing over the past 150 years, may be increasing more in the future. *Education - 4 Categories Crosstabulation*

<table>
<thead>
<tr>
<th>Education - 4 Categories</th>
<th>Less than high school</th>
<th>High school</th>
<th>Some college</th>
<th>Bachelor's degree or higher</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>19.0%</td>
<td>13.6%</td>
<td>16.1%</td>
<td>12.3%</td>
<td>14.1%</td>
</tr>
<tr>
<td>Don't know</td>
<td>23.4%</td>
<td>19.7%</td>
<td>14.6%</td>
<td>7.9%</td>
<td>13.6%</td>
</tr>
<tr>
<td>Yes</td>
<td>57.7%</td>
<td>66.7%</td>
<td>69.3%</td>
<td>70.8%</td>
<td>72.3%</td>
</tr>
</tbody>
</table>
Plot

NORTHERN HEMISPHERE

Departures in temperature (°C)
from the 1961–1990 average

Year

1000 1200 1400 1600 1800 2000

Data from thermometers (red) and from tree rings, corals, ice cores and historical records (blue).
Global temperatures have increased by over 1°C
How do you get actionable insights?

Identify problem, need, & data

Research

Understand, clean, & shape your data

Edit

Apply design principles to choose chart, add layers, text, refine

Plot

Check for errors, areas of confusion, opportunities for improvement

Review