Rotman

INTRO TO DATA VISUALIZATION

Part I - Basic Principles & Good Practices

September 3, 2024 Prepared by Jay Cao / TDMDAL

Website: https://tdmdal.github.io/mma-dv-2024/

Rotman School of Management
UNIVERSITY OF TORONTO

What is Data Visualization (DV)

- Graphical representations of data and information
  - Encode data/info via “visual variables”: position, size, shape, color, brightness, orientation, texture and motion (Bertin, 1967; Mackinlay, 1986)

- Many forms
  - Basic charts and plots (e.g., line chart, bar chart, etc.)
  - Maps (choropleth map, symbol map, etc.)
  - Interactive dashboards
  - Infographics (not our focus)

Why DV

- “A picture is worth a thousand words”
  - Abstract to reduce the complexity and communicate insights
  - Promote comprehension (i.e., get the message across)

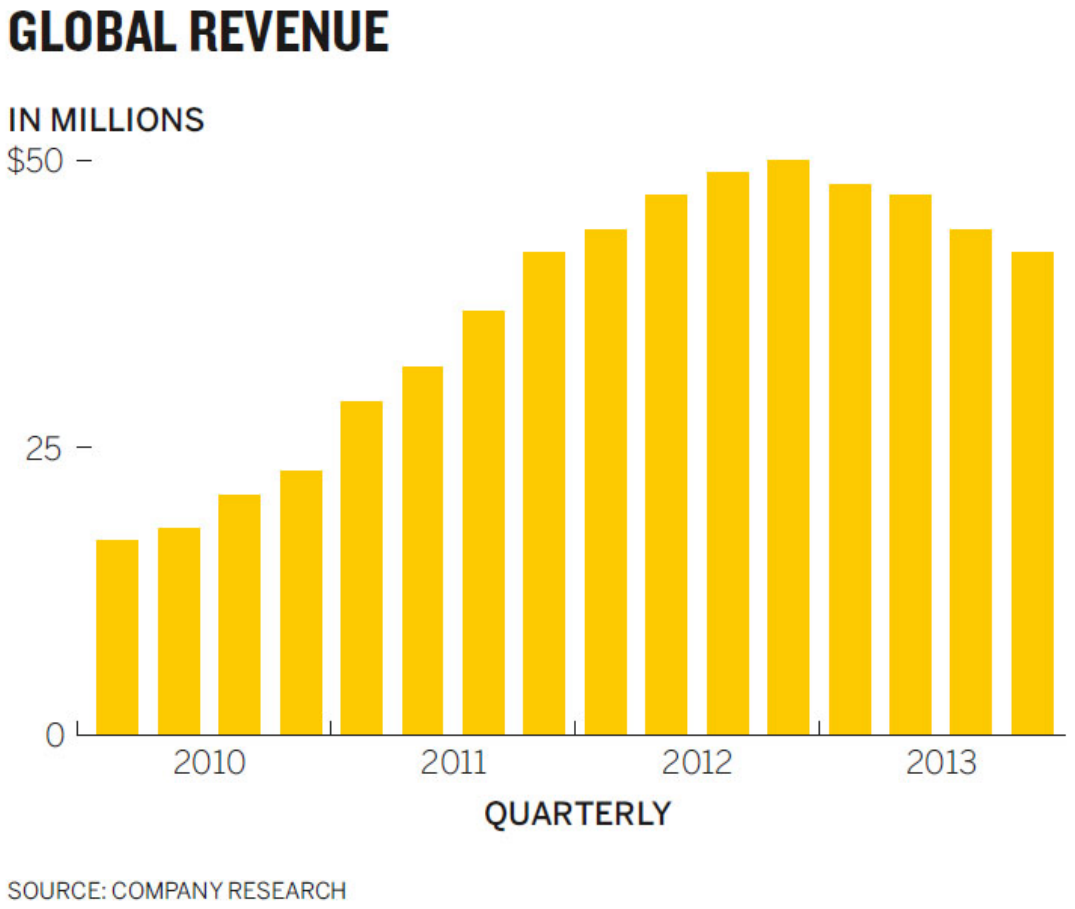
- Facilitate exploratory analysis
  - Identify patterns, trends, and outliners
  - Guide feature engineering, modeling, etc.

Plan for this Workshop

- Establish a way to think about and achieve good visualizations (Part I)
  - i.e., what's a good chart, and a good way/workflow to make one

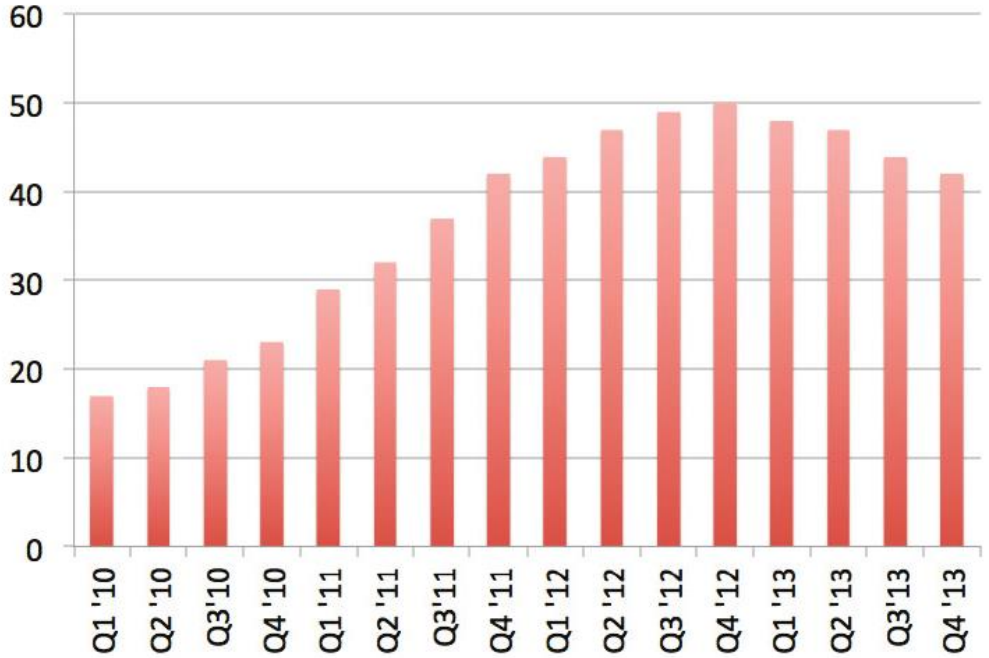
- Learn Python tools to produce good visualizations (Part II, III, & IV)
  - Plot basic default charts and customize them (Matplotlib)
  - Plot maps (Matplotlib & Geopands)
  - Create dashboard (Quarto & Plotly Express)

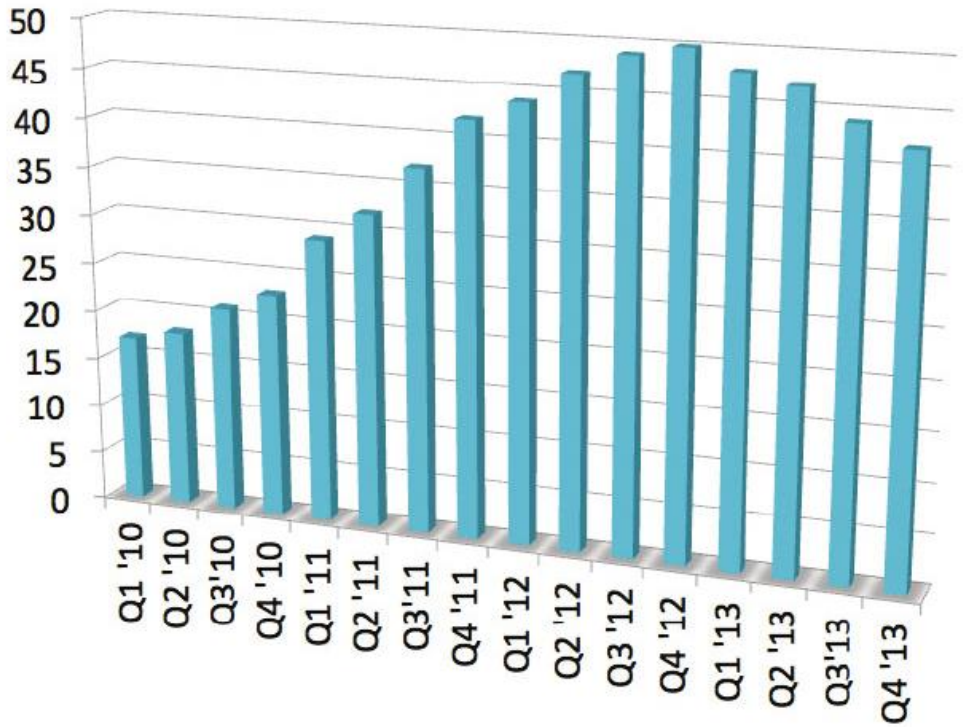
Motivation – Is this a Good Chart?



This example (including figures) is from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

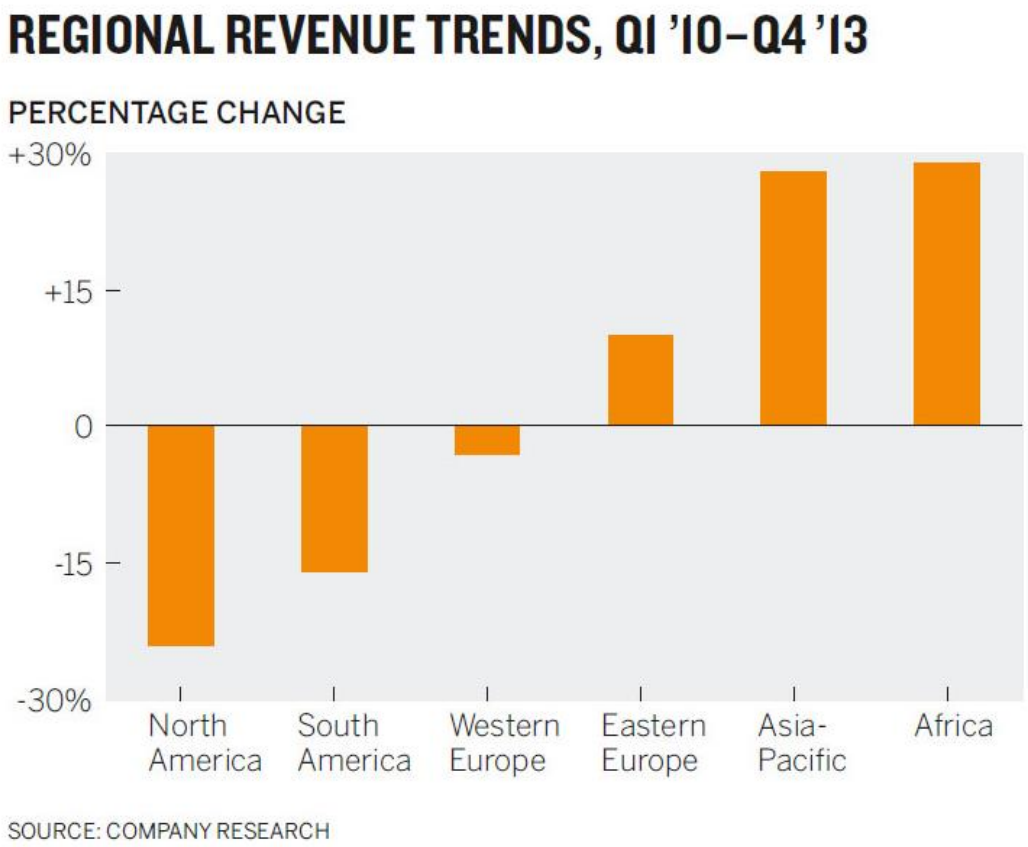
Is this a Good Chart, Visually?

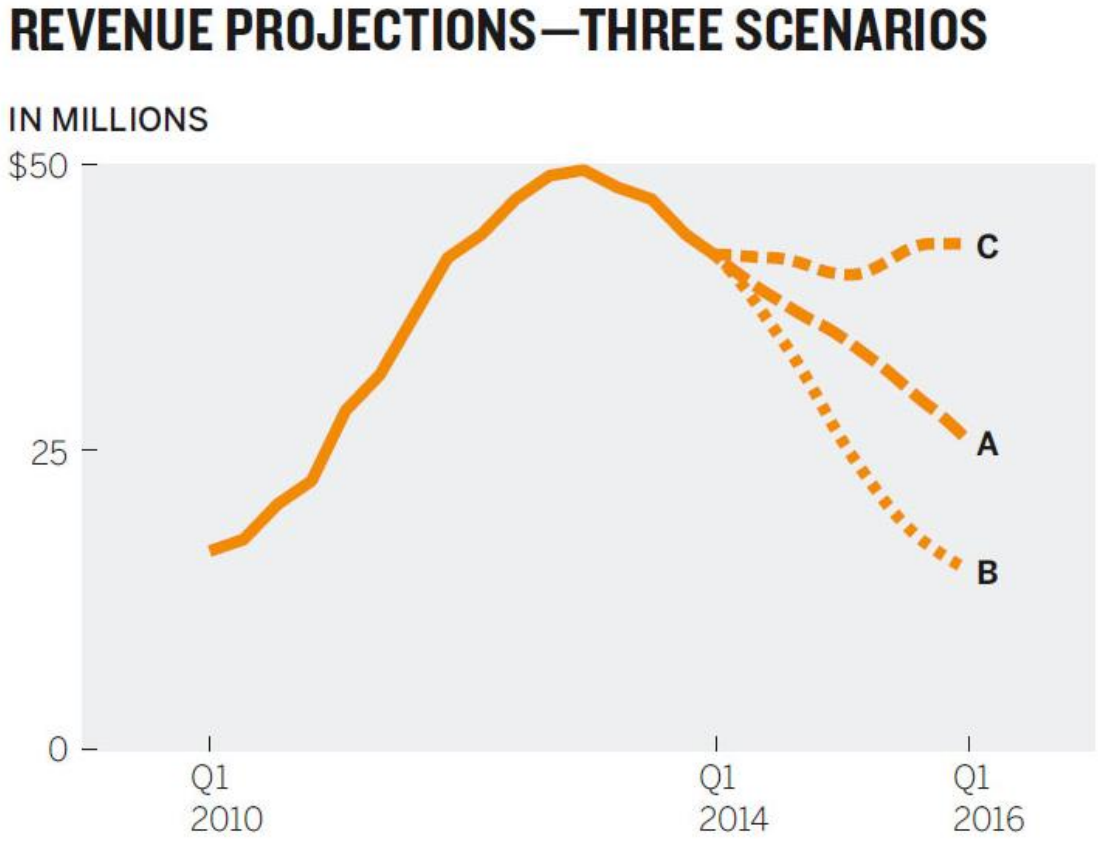




This example (including figures) is from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

Is this a Good Chart, Message-wise?

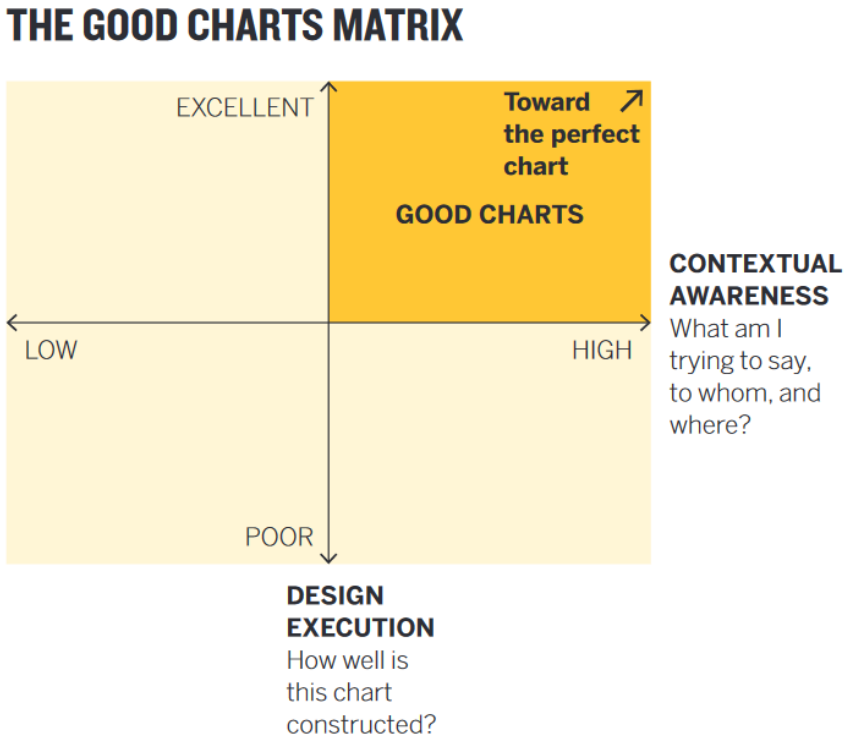




This example (including figures) is from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

Visual Design & Idea Communication

- A good chart has good visual designs
- A good chart communicates ideas effectively
  - i.e., it sends relevant messages to the targeted audience in the most efficient way
  - high in contextual awareness. It answers these questions: what am I trying to say, to whom, why is it important to say it, and how to say it



From page 9 of Good Charts by Scott Berinato

Ref. Intro section from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

Goal for Part I

- Establish a workflow to think about and produce good visualizations
  - Step 1 Plan
  - Step 2 Create
  - Step 3 Refine

- Learn basic principles & good practices to produce good charts

Two Main References

- Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato, Harvard Business Review Press, [2016]

- Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLOS Computational Biology 10(9): e1003833.
  - Paper is also Ch6 of Scientific Visualization: Python & Matplotlib by Rougier
  - Companion code for the paper (here) and chapter 6 of the book (here)


Editorial

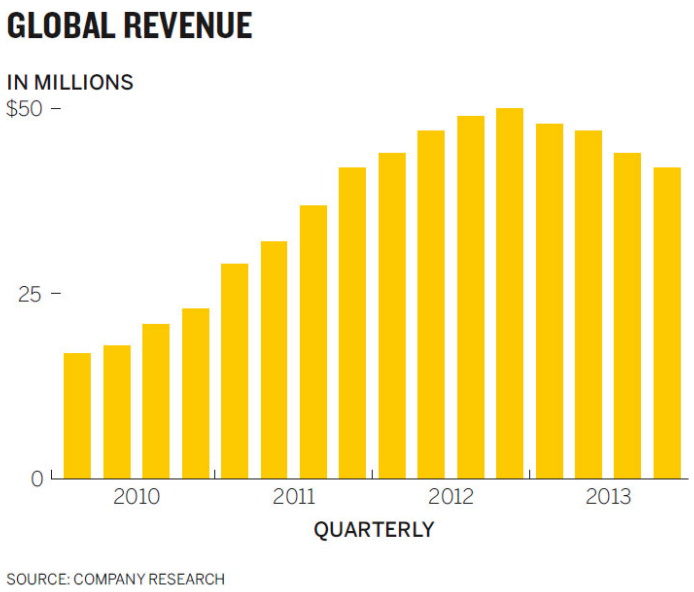
Ten Simple Rules for Better Figures

Step 1: Plan

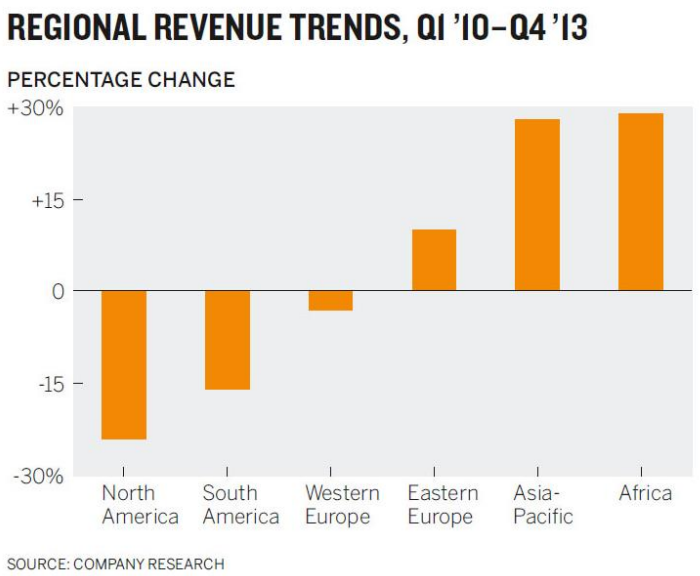
1. Know your audience
2. Identify your message
3. Know your medium
4. Choose a chart type

Ref: The first three bullet points are rule 1 to 3 in Rougier et. al. (2014)

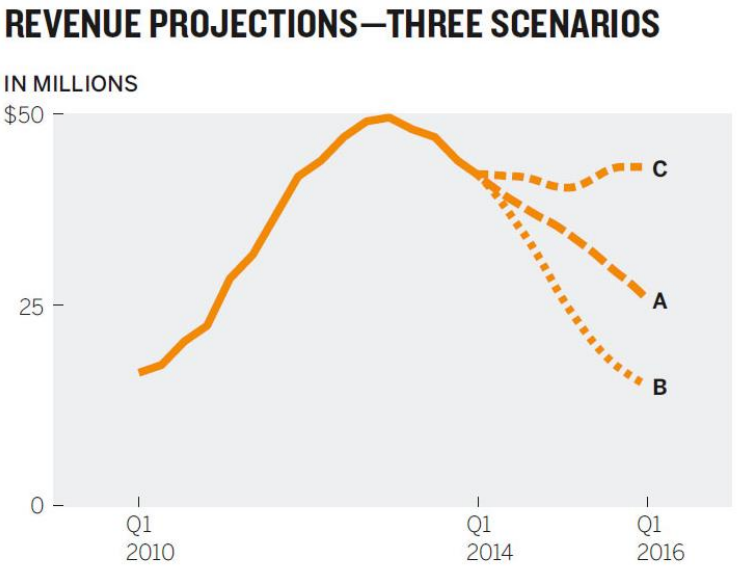
Know Your Audience & Identify Your Message



Audience: Public shareholders or newly hired junior managers who want to understand some basic facts about the company's global revenue
Message: simple fact: global revenue grew from $20M in the first quarter of 2010 to…, Since 2013, however, there was a declining trend



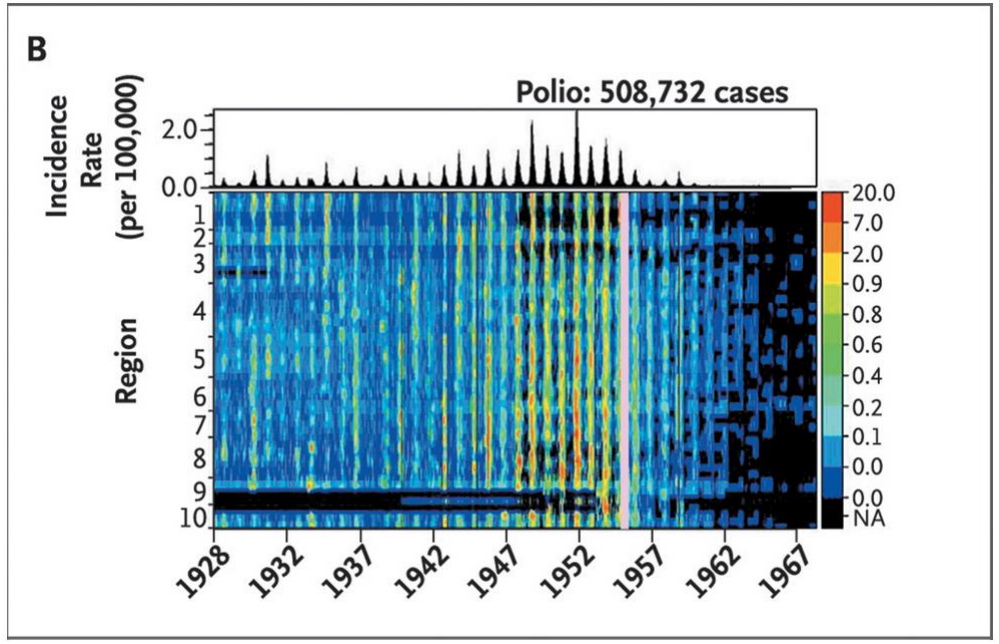
Audience: Sales executives who are looking for markets to invest in to reverse the global revenue declining trend
Message: some insight: negative revenue growth in America but strong growth in Asia Pacific and Africa, and therefore…



Audience: Executives who want to understand future revenue trajectories based on different investment strategies and market conditions.
Message: predictions based on data and modeling: if scenario A…

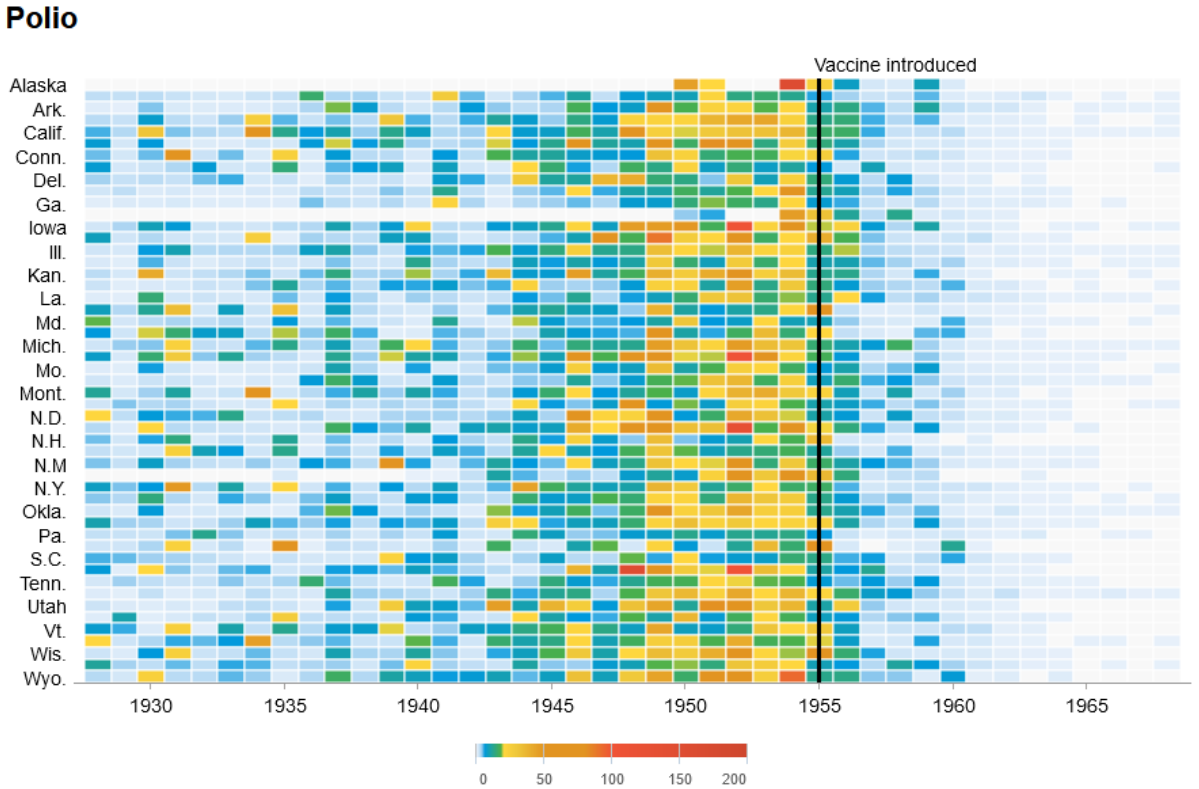
This example (including figures) is from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

Know Your Audience & Know Your Medium



New England Journal of Medicine

Audience: Specialist (doctors, scientists, etc.)
Medium: Scientific journal



"Battling Infectious Diseases in the 20th Century: The Impact of Vaccines" WSJ

Audience: General public
Medium: daily newspaper published in broadsheet format and online

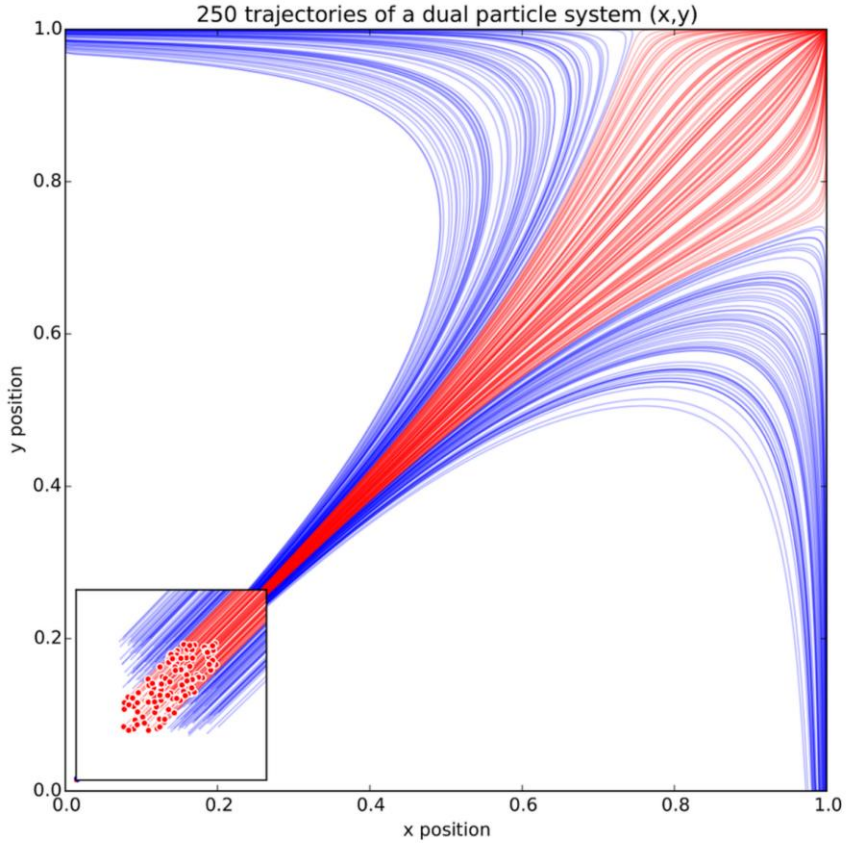
This example (including figures) is from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

Adapt to the Medium – Good Practice

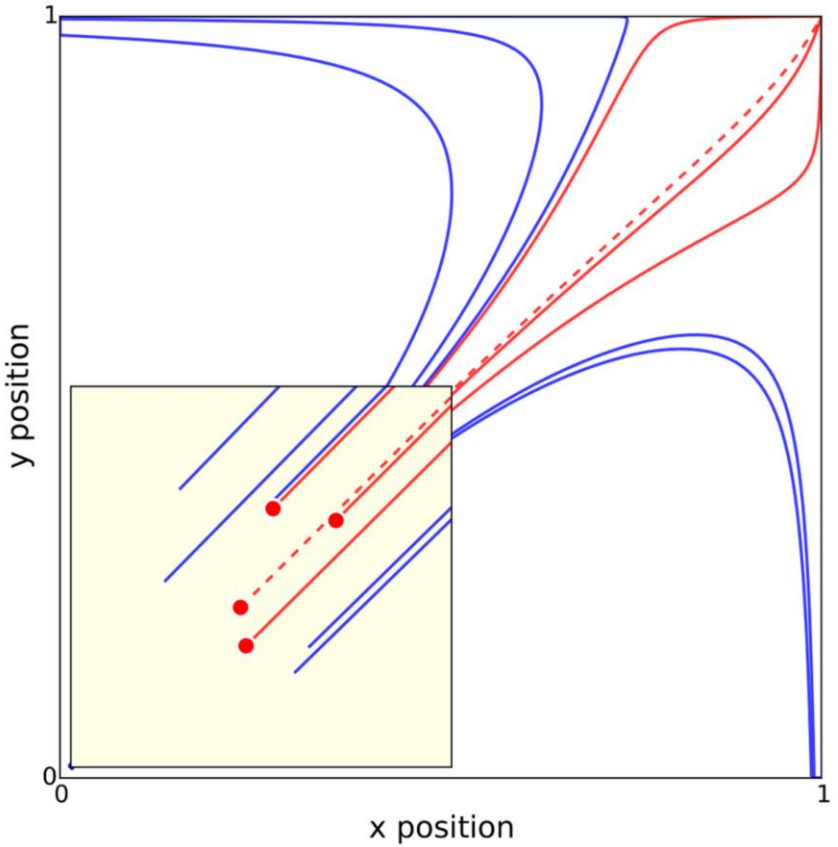
	Projector screen (oral presentation)	Paper (report, journal article)	Computer monitor (blog, web)
Ways of viewing and interacting with a figure	- limited viewing time - viewed from a distance - cannot change view orientation	- can be viewed as long as needed - can be viewed closely - can change view orientation	?
How to adapt?	- keep a figure simple, but with visually salient messages - easily viewable visual elements (thicker lines, bigger fonts, strong color contrast, no vertical texts, etc.) - easily referable visual elements (marker types, etc.)	- more details in the figure - caption with more explanations - if black and white print, then…	?

Ref: Rougier et. al. (2014)

Adapt to the Medium – An Example



For a journal article

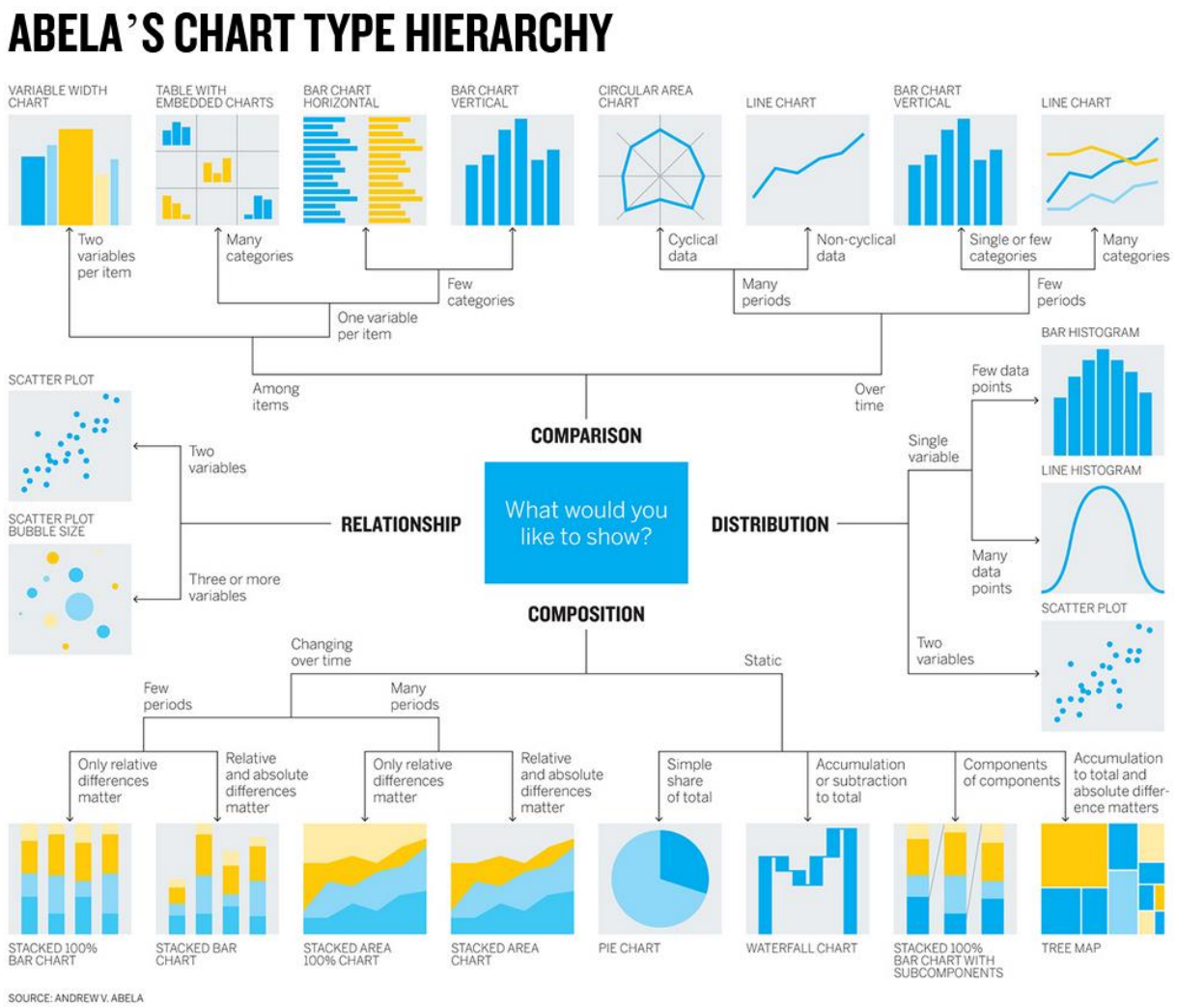


For presentation screen

Ref: Figure 3 in Rougier et. al. (2014)

Choose a Chart Type

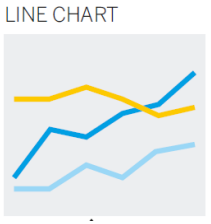
- A chart choosing cheat sheet
  - By no means complete
- What do you plan to show?
  - Comparison
  - Relationship
  - Distribution
  - Composition
  - Maps
  - Network
  - Logic



Note: 1) Original figure by Abela in his book *Advanced Presentations by Design: Creating Communication that Drives Action.*
2) This is a re-plot on page 83 of Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato.
3) Many versions exist online, see for example, here, here and here.

Step 1 Plan – An Example: The Mag 7

- Context
  - You are a data scientist in an investment research team at an online trading and investing firm. Your company provides trading tools, market data, and research reports to help its clients manage their online investments.
  - When preparing the end-of-year client meeting in 2023, your team noticed that a group of 7 tech stocks significantly outperformed the S&P 500. Your team tasked you to produce a visualization of this finding/fact for a presentation slide in the meeting.
  - The group (the Magnificent 7): Nvidia, Meta, Tesla, Amazon, Alphabet, Microsoft, & Apple
- **Audience**
  - company clients who uses the company platform to trade and do investment research
- **Message**
  - Present the fact: the mag 7 tech stocks has grown ~75% in 2023, significantly outpacing S&P500 (~20%) and other companies -- S&P500 companies excluding the mag 7. (~10%)
- **Media**
  - A presentation slide / projector screen / computer screen (for online audience)
- **Choose a chart**
  - Bar chart vs Line chart? After some discussions with your team, you decided to use **Line chart** to display and contrast the returns of mag 7, S&P, and S&P (excluding mag 7) over time in 2023.
- Collect data
  - Three time series: cumulative returns for mag 7, S&P500, and S&P500 excluding mag 7.



Step 2: Create - Get the Right Tool

- Too many tools
  - Spreadsheet: Excel, Google Sheets, Apple Numbers
  - BI analytics software: Tableau, PowerBI, Qlik
  - Online graphic tools: Datawrapper, RAWGraphs
  - Design tools: Adobe Illustrator
  - Programming languages
    - Python: matplotlib, seanborn, plot.ly
    - R: ggplot2, plot.ly
    - Javascript: D3, Observable, Google Chart, Highcharts, plot.ly

- Choose one and master it
  - Python's Matplotlib will be our main tool in this workshop

Ref: "Get the Right Tool" is rule 10 in Rougier et. al. (2014)

Step 2: Create – Prototype / Default Plot

- Any plotting tool has a default style for standard charts
  - It's a good starting point

- **However**, the default plot cannot be the final plot
  - Unless you are exploring the data on your own…
  - It rarely meets the standard for publication or formal presentation

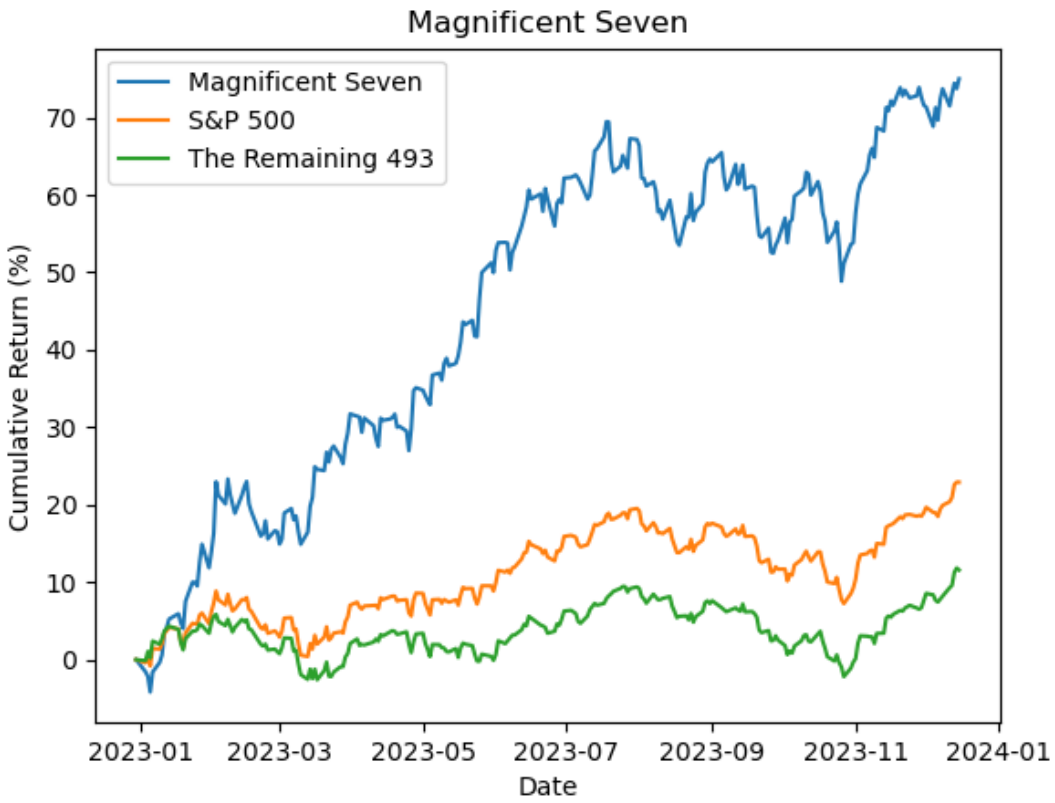
- "Do Not Trust the Default"
  - "… they are good enough for any plot but they are best for none."

Ref: "Do Not Trust the Default" is rule 5 in Rougier et. al. (2014)

Step 2: Create – The Mag 7 Example

- Choose Python's Matplotlib library as the tool for this static line plot

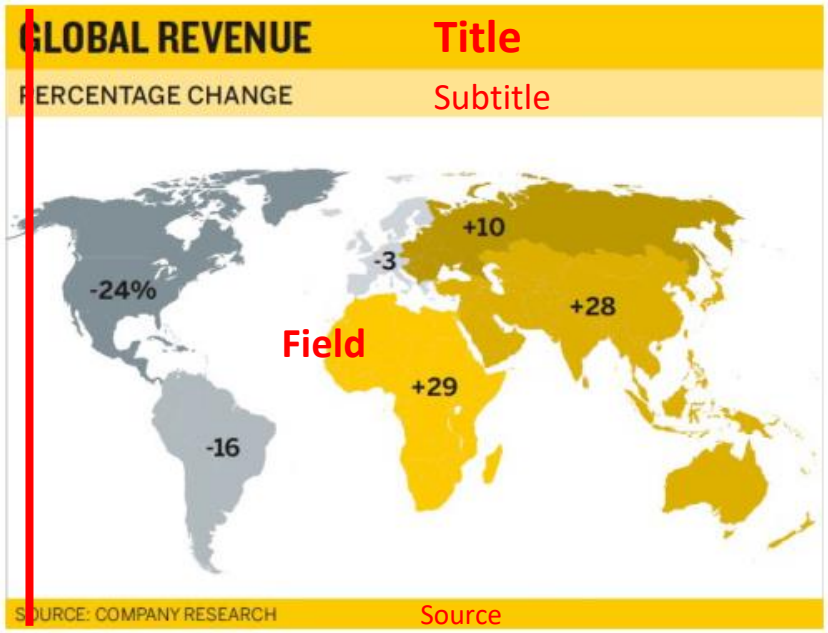
- The default line plot is OK but far from publication-ready

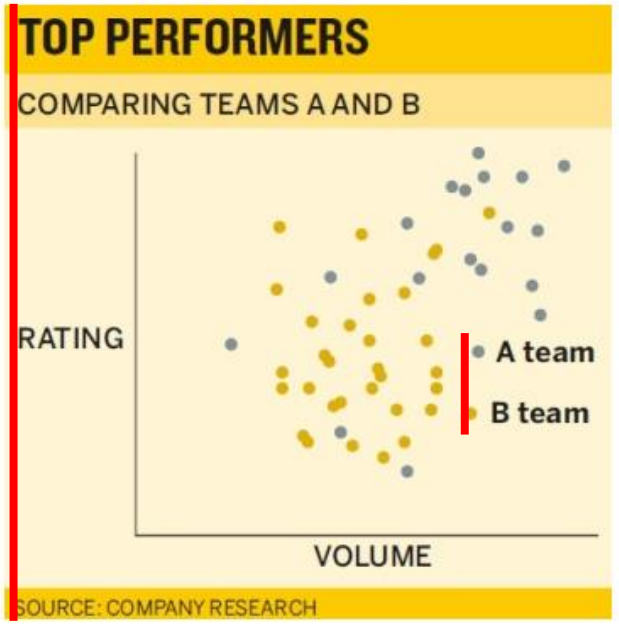


default line plot from Matplotlib

Step 3: Refine to Impress - Consistency

- Consistent layout
- Consistent alignment
- Consistent color scheme



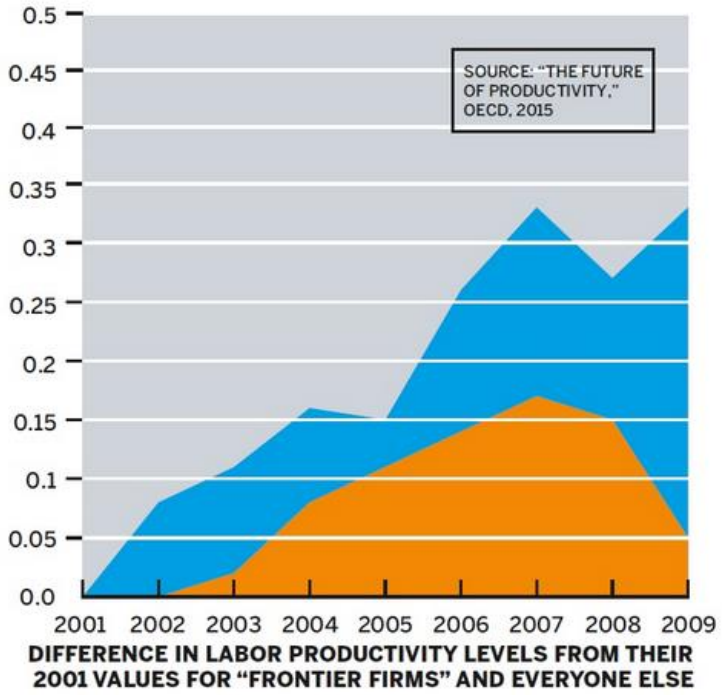


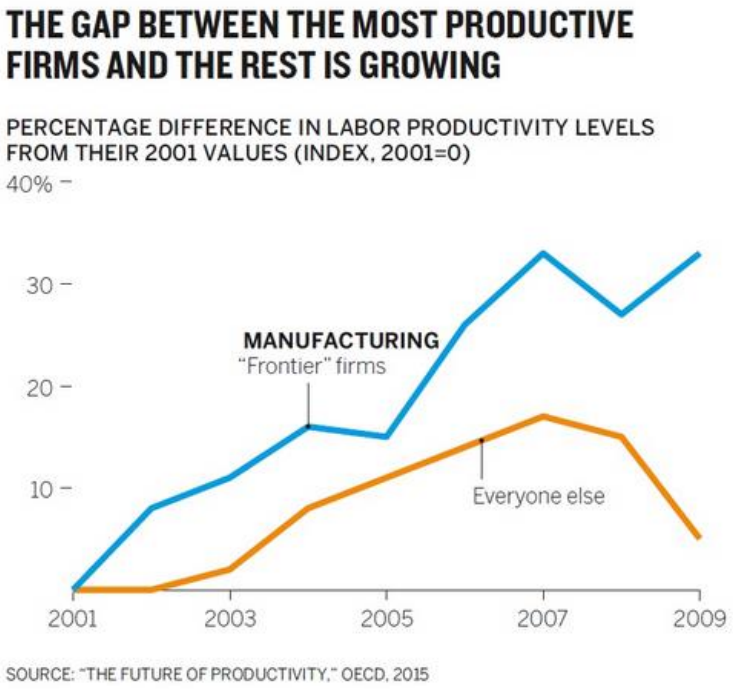
An example: consistent layout and alignment

Ref. Chapter 5 from [Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations] by Scott Berinato

Refine to Impress – Clarity and Simplicity

- Avoid “chartjunk”





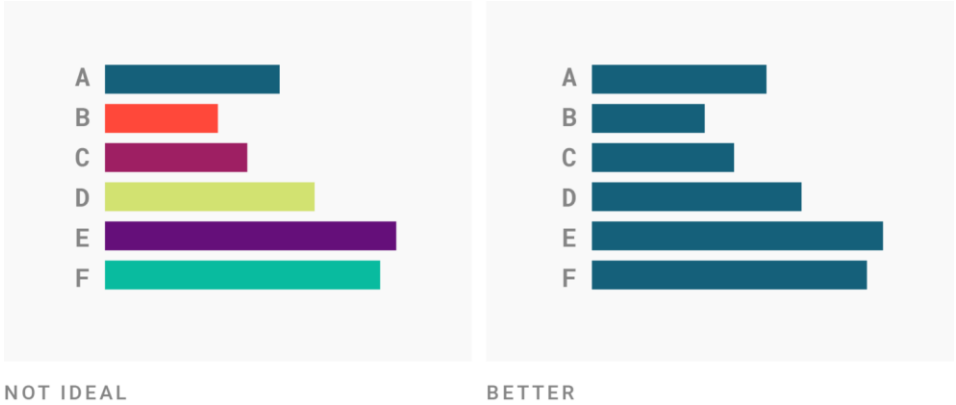
Ref. 1 Chapter 5 from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato
Ref. 2 “Avoid ‘chartjunk’” is rule 8 in Rougier et. al. (2014)

Refine to Impress - Avoid “chartjunk”

Ref: Data Looks Better Naked by Darkhorse Analytics; Versions for Tables, Pies, and Maps.

Refine to Impress – Color

- Excessive/unnecessary use of color is bad



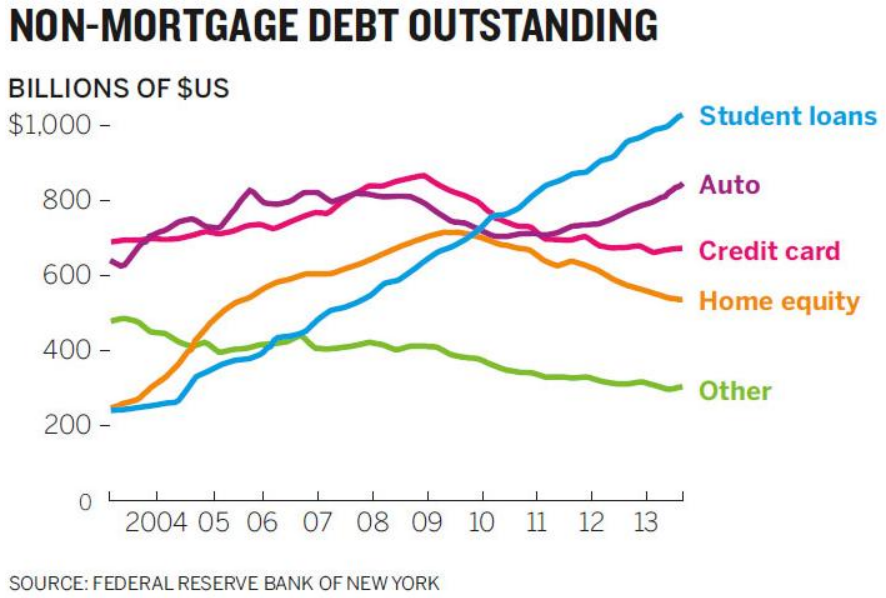
- Use meaningful colors

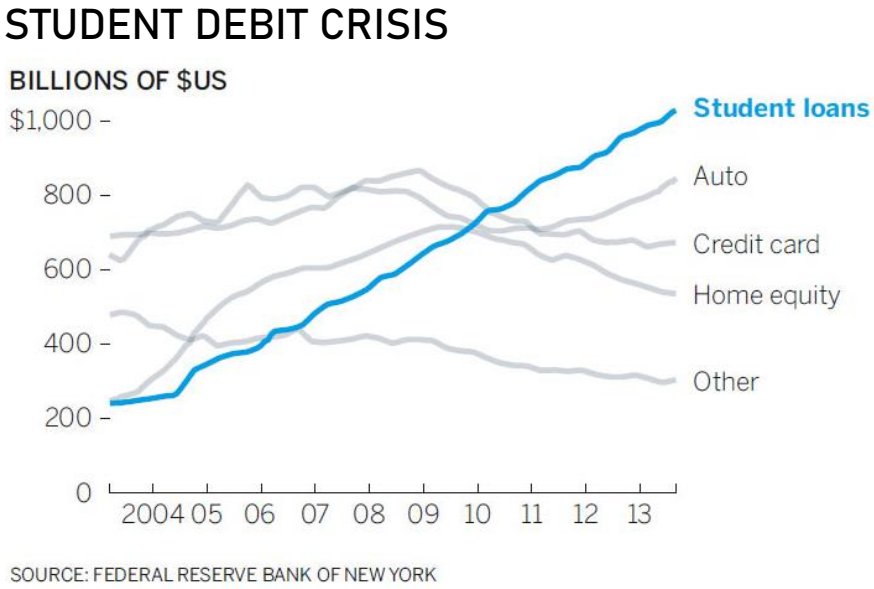
- Attend to color blindness

Ref: 1) https://blog.datawrapper.de/10-ways-to-use-fewer-colors-in-your-data-visualizations/
2) https://clauswilke.com/dataviz/color-pitfalls.html

Refine to Persuade

- Hone the main idea.
- Make it stand out.
- Adjust what's around it.





Ref. Chapter 6 from Good Charts: the HBR Guide to Making Smarter, More Persuasive Data Visualizations by Scott Berinato

Persuade but Don’t Mislead

- One example, “truncated y-axis”

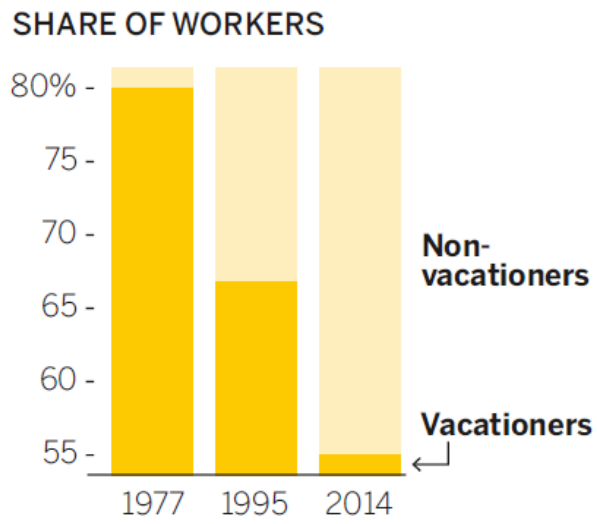
THE DATA

SHARE OF WORKERS

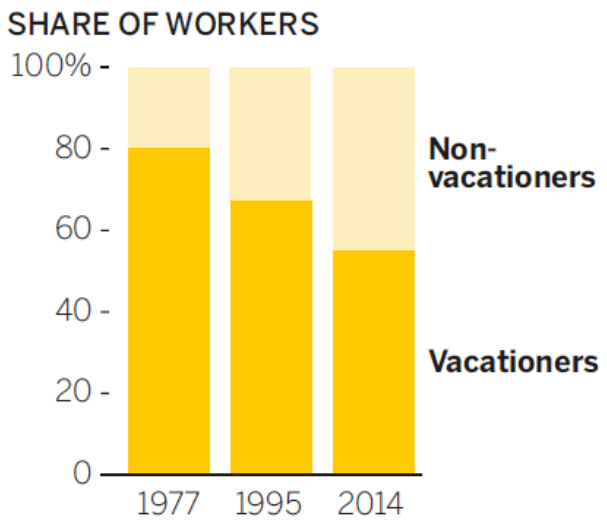
	Vacationers	Non-vacationers
1977	80%	20%
1995	67	33
2014	55	45

SOURCE: BUREAU OF LABOR STATISTICS, VOX

TRUNCATED AXIS



FULL AXIS



From Chapter 7 of Good Charts by Scott Berinato

Note: Many other ways to lie with charts. See, for example, How to Lie with Charts and How Charts Lie (and a talk by the *How Charts Lie* author).

Step 3: Refine – The Mag 7 Example