# VISUALIZATION

**Graphical Excellence**

# Overview

- Graphical Excellence
- Graphical Integrity
- Design Aesthetics

# Overview

- Graphical Excellence
- Graphical Integrity
- Design Aesthetics

# Features of good Graphical Displays - Tufte

- They should show the data.
- They should encourage the user to think about the substance of the data.
- They should make large datasets coherent to the reader.
- They should reveal the data at several levels of detail.
- They should avoid distorting what the data have to say.
- They should serve a reasonably clear purpose: description, exploration, tabulation or decoration
- They should be closely integrated with the statistical and verbal descriptions of a data set

# Principals of Graphical Excellence - Tufte

I. Graphical excellence is the well-designed presentation of interesting data—a matter of substance, of statistics, and of design.

II. Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency.

III. Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.

IV. Graphical excellence is nearly always multivariate.

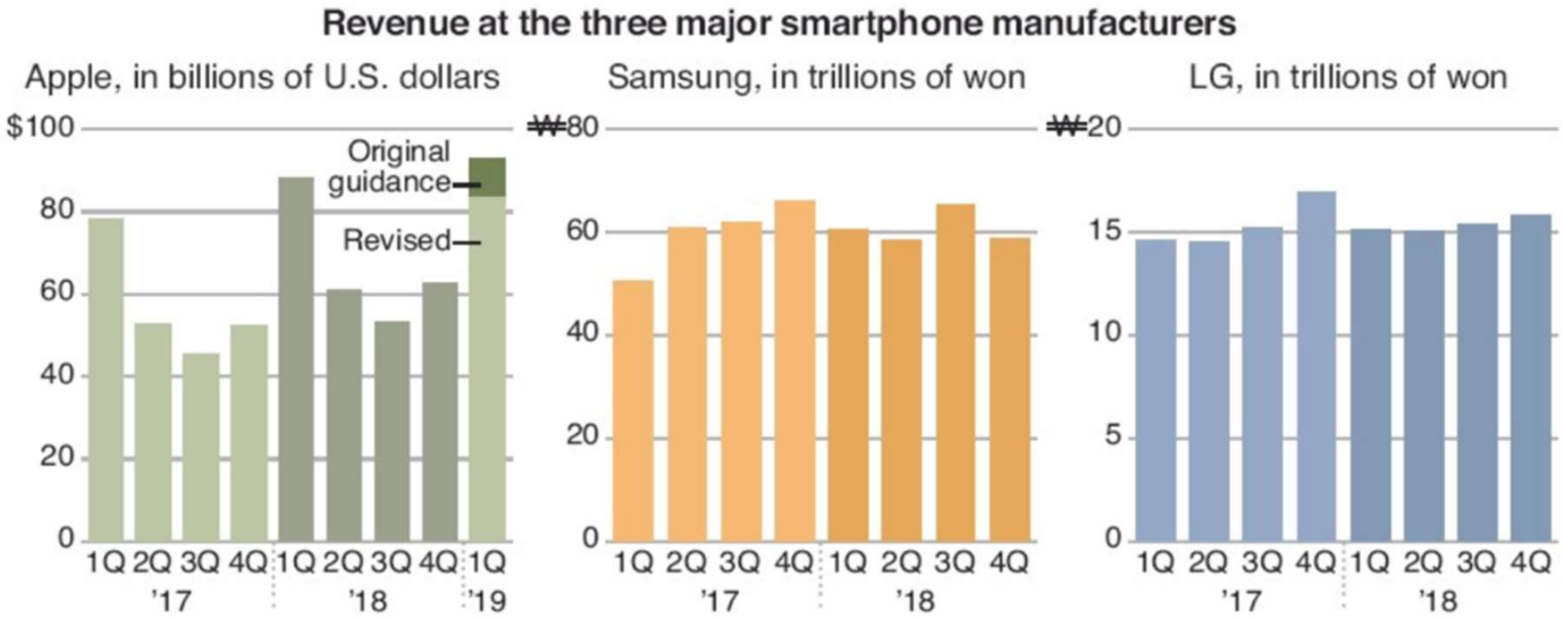
V. And graphical excellence requires telling the truth.

# Overview

- Graphical Excellence
- Graphical Integrity
- Design Aesthetics

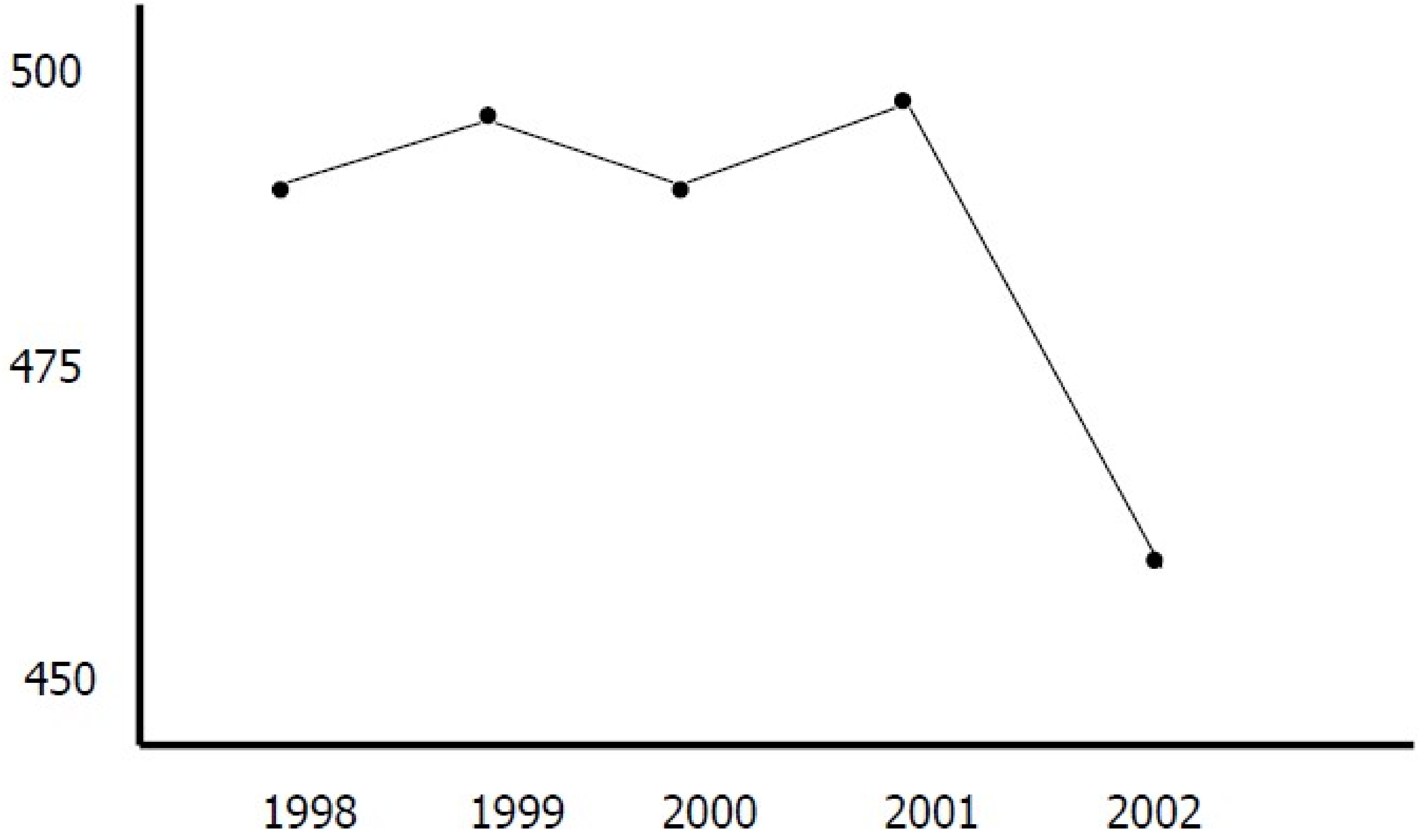
# What is wrong?

Standardised currency e.g., USD, not used

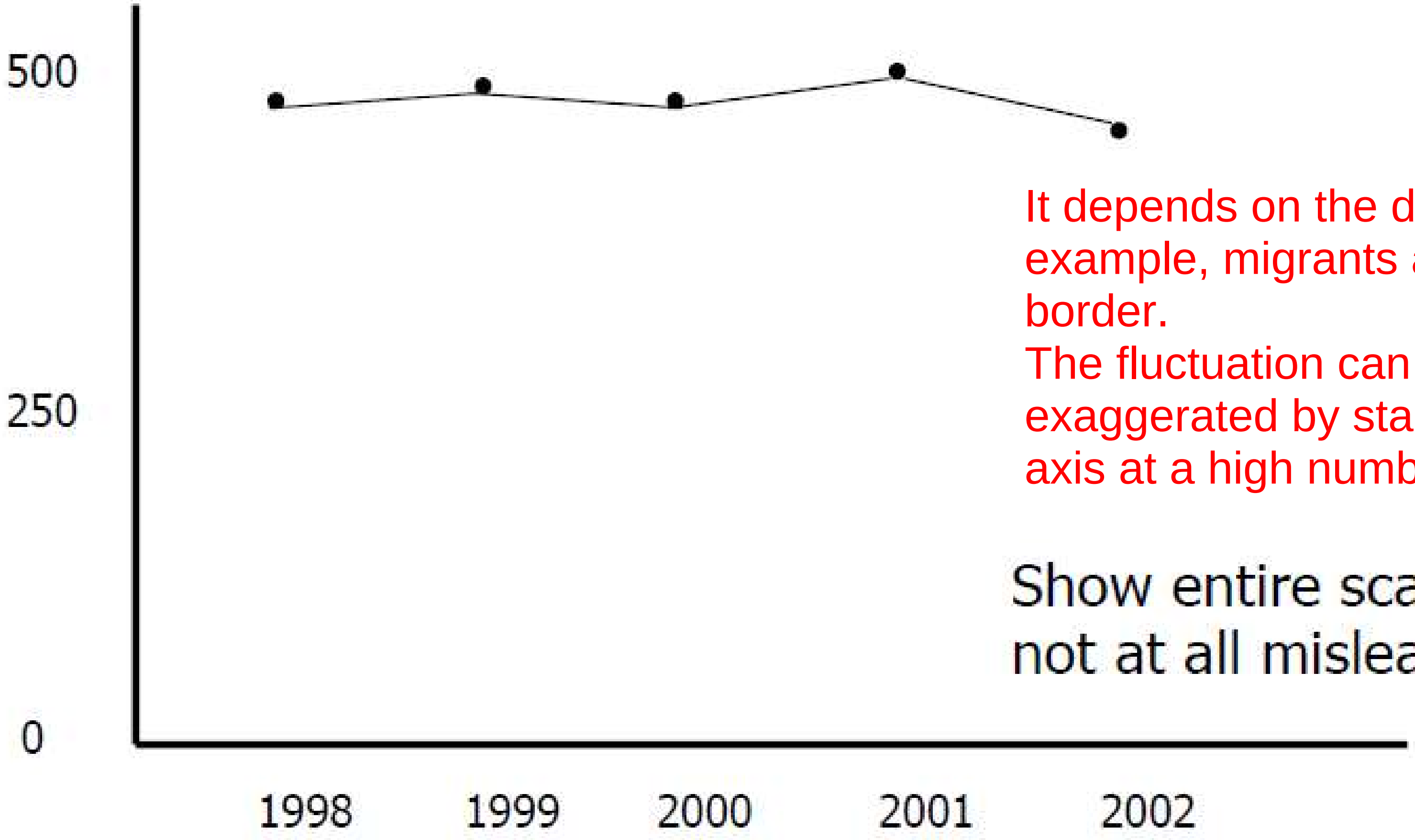


# Graphical Integrity

No units, no labels, no title.



# Scale



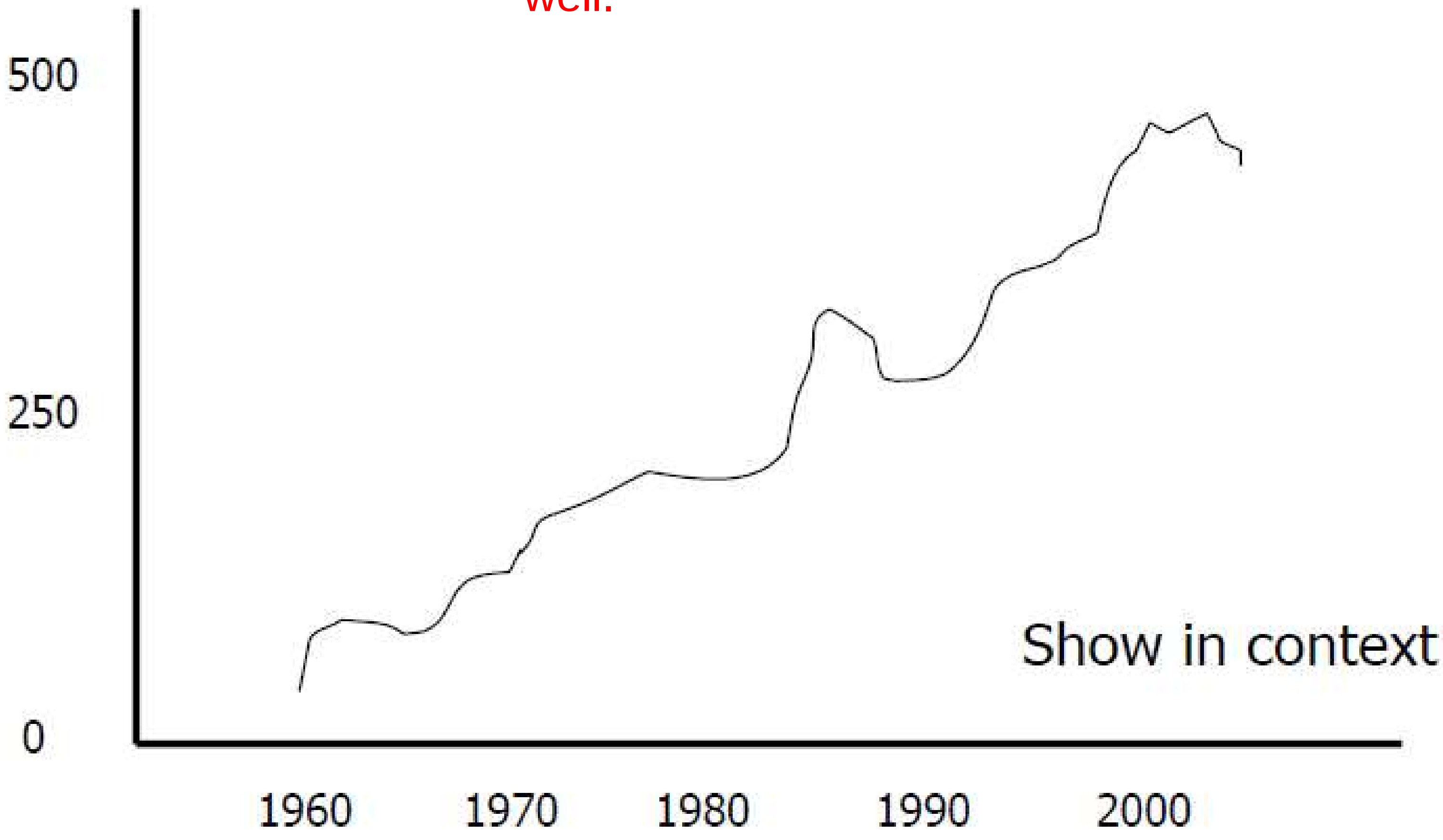
It depends on the data. For example, migrants at the US border.

The fluctuation can be exaggerated by starting the y axis at a high number.

Show entire scale – not at all misleading

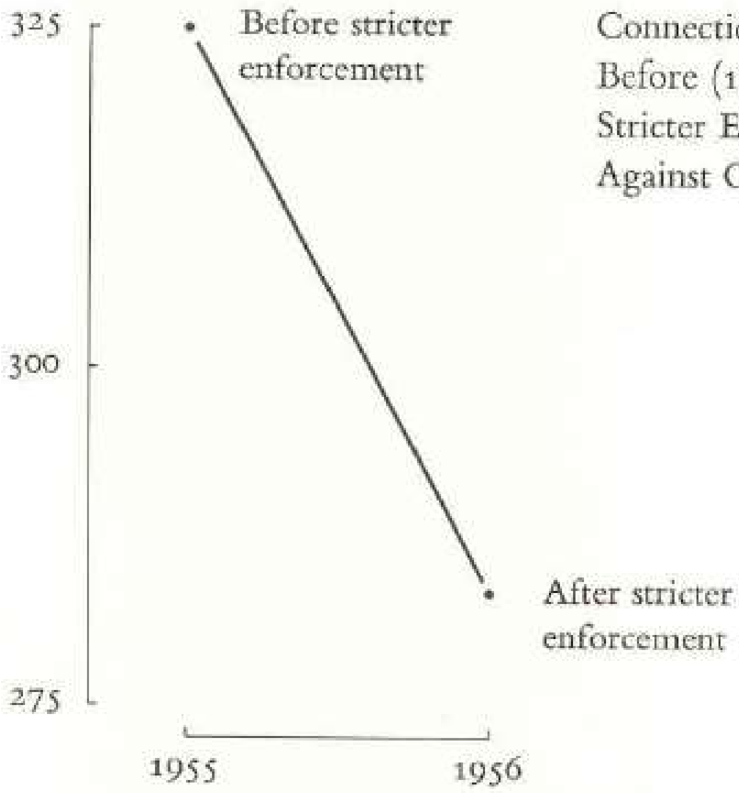
# Context

Example: When showing fluctuations, one should show the trend in prior years as well.



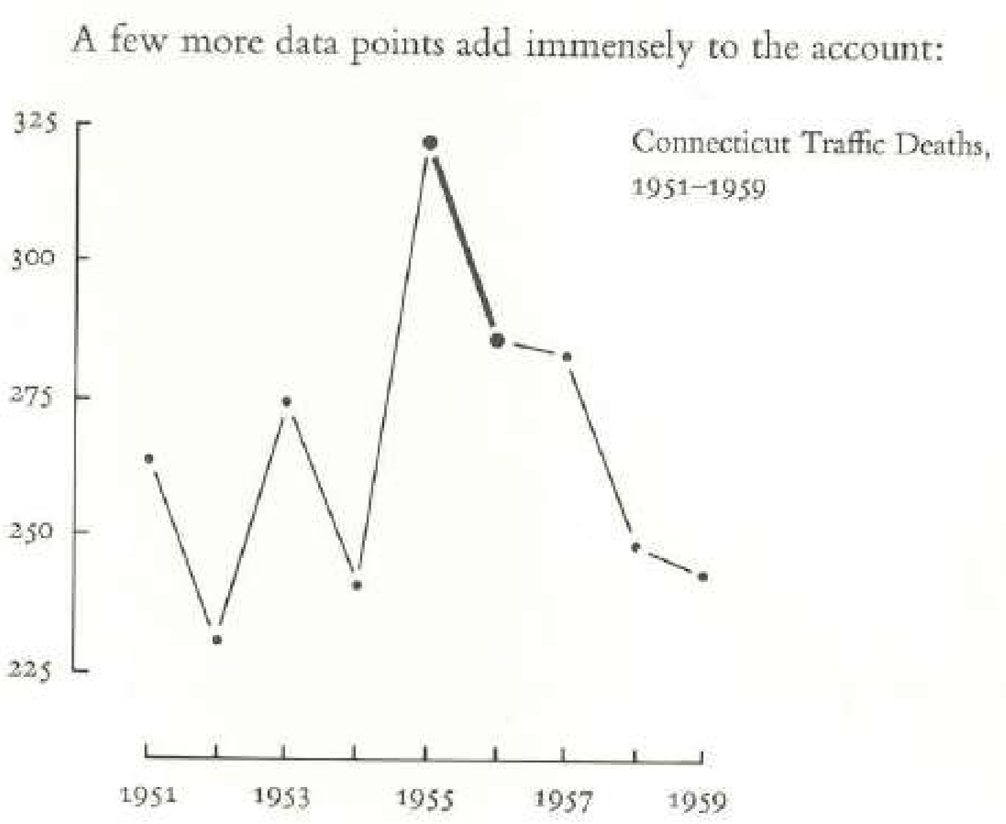
# Chart Integrity

- Where’s baseline?
- What’s scale?
- What’s context?



Connecticut Traffic Deaths,
Before (1955) and After (1956)
Stricter Enforcement by the Police
Against Cars Exceeding Speed limit

Great work, Connecticut
You get a prize !



A few more data points add immensely to the account:

Oops, not so fast!!

Show the context!

Tufte Vol 1, p. 74

# Tufte's Six Principles of Graphical Integrity

I. The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities measured.

II. Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.

III. Show data variation, not design variation.

IV. In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.

V. The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

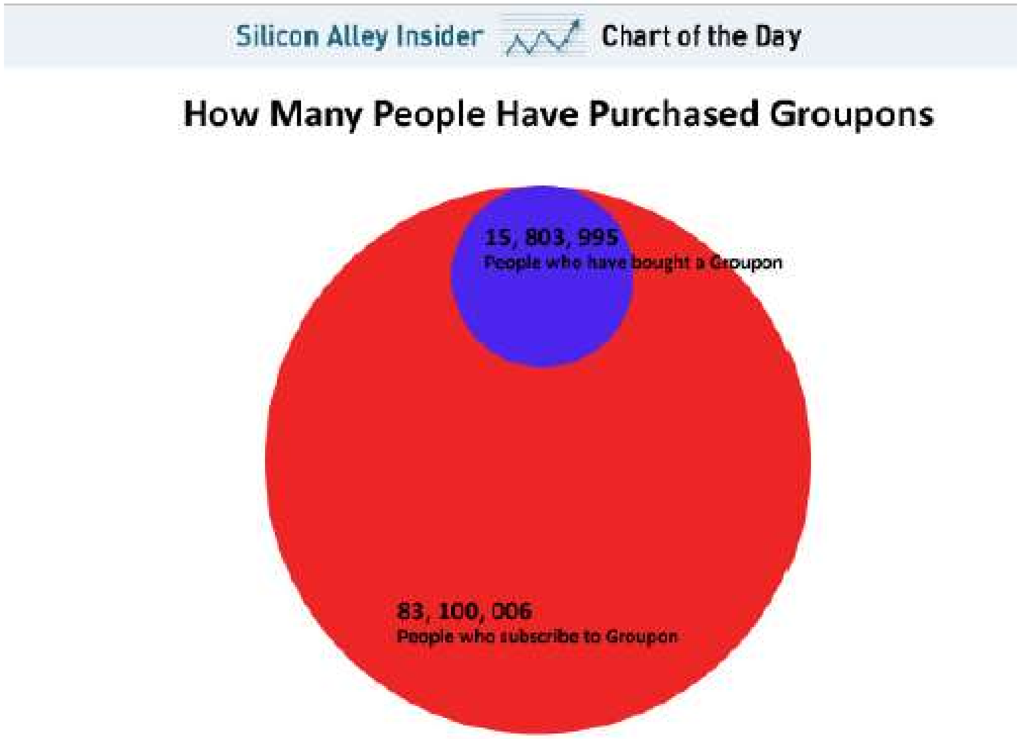
VI. Graphics must not quote data out of context

# Measuring Misrepresentation

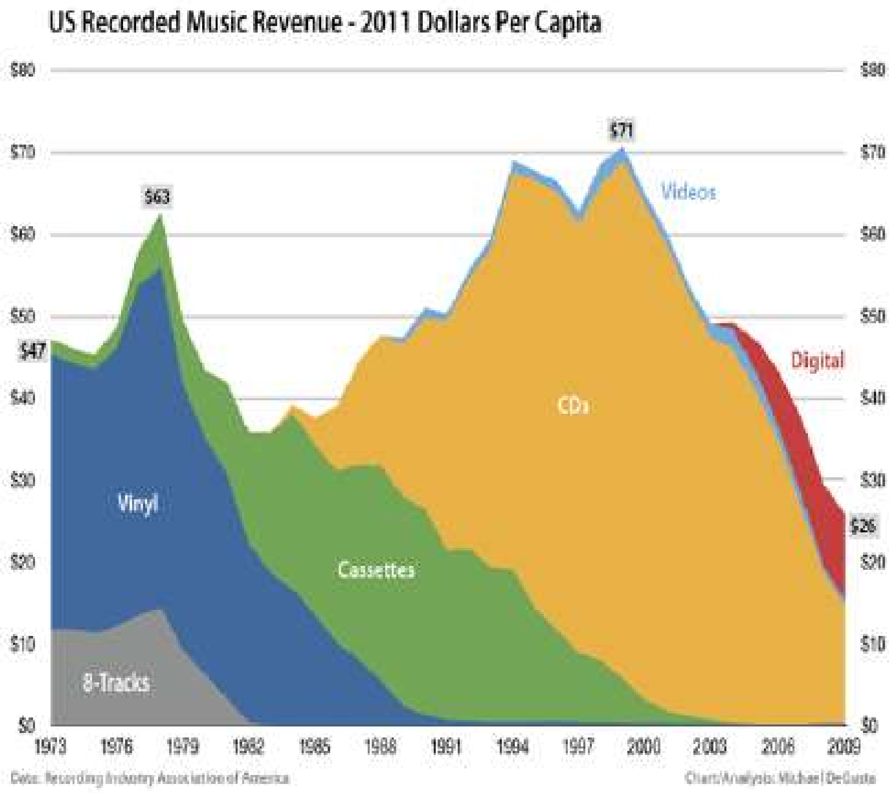
- Visual attribute value should be directly proportional to data attribute value
- Lie factor = (Size of effect shown in graphic)/(Size of effect in data)

# Lie Factor

- Graphical area ratio ~9.7
- Data ratio ~5.3
- Lie factor 9.7/5.3 = 1.8

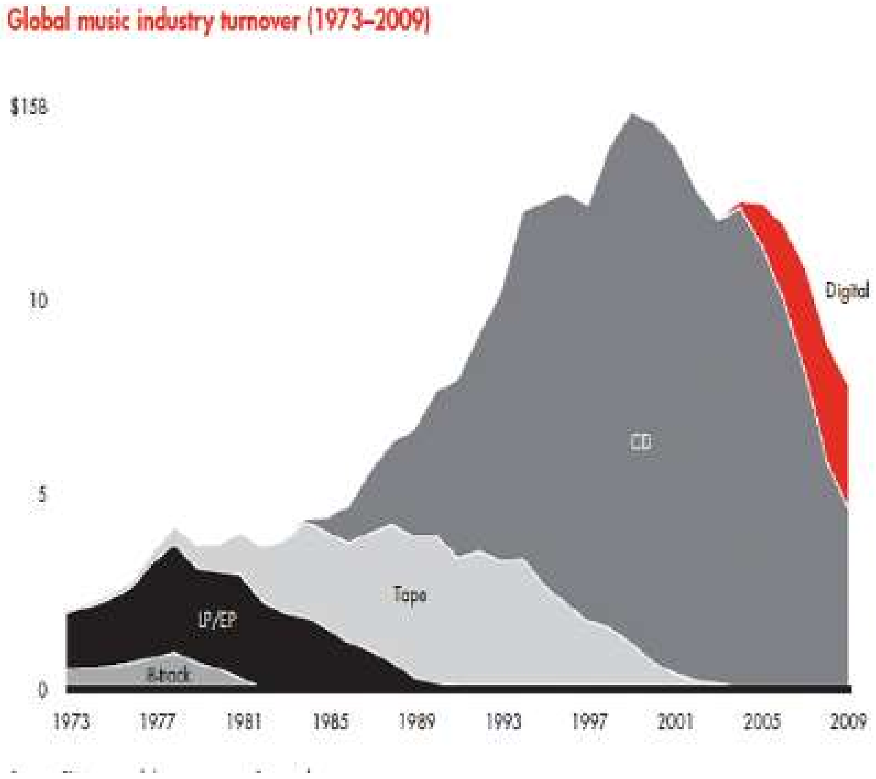


# Financial data should be adjusted for inflation


US Recorded Music Revenue - 2011 Dollars Per Capita

Data: Recording Industry Association of America

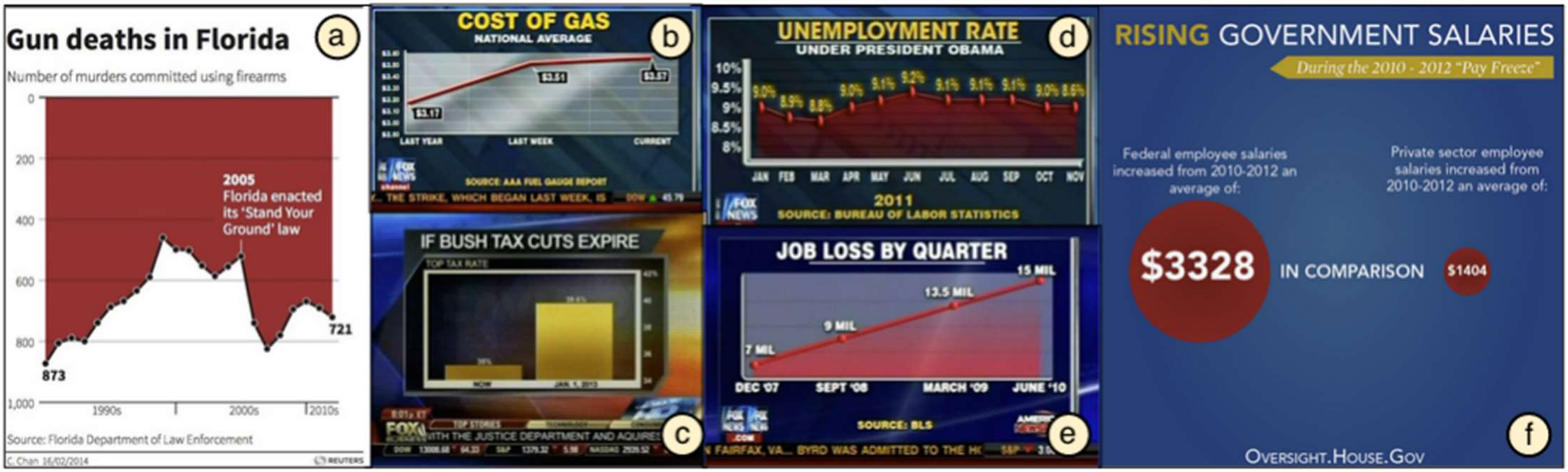
Chart/Analysis: Michael DeGusta


Global music industry turnover (1973–2009)

# Deceptive Visualizations

- Visualizations can make messages more accessible, comprehensible and persuasive
- Visual representations can also be easily misused and misunderstood – even by their creators
- Definition: “a graphical depiction of information, designed with or without an intent to deceive, that may create a belief about the message and/or its components, which varies from the actual message“
  - A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini, "How Deceptive are Deceptive Visualizations?", CHI '15.

# Real-world Examples



- Some of the real-world data visualization examples which might lead to misinterpretation of message, hence to deception
  - Manipulation of axis orientation/scale [a,c]
  - Use of disproportionate sizes [f]
  - Incorrect representation [d]
  - Non-linear scales [b,e]

# Types of Deceptions

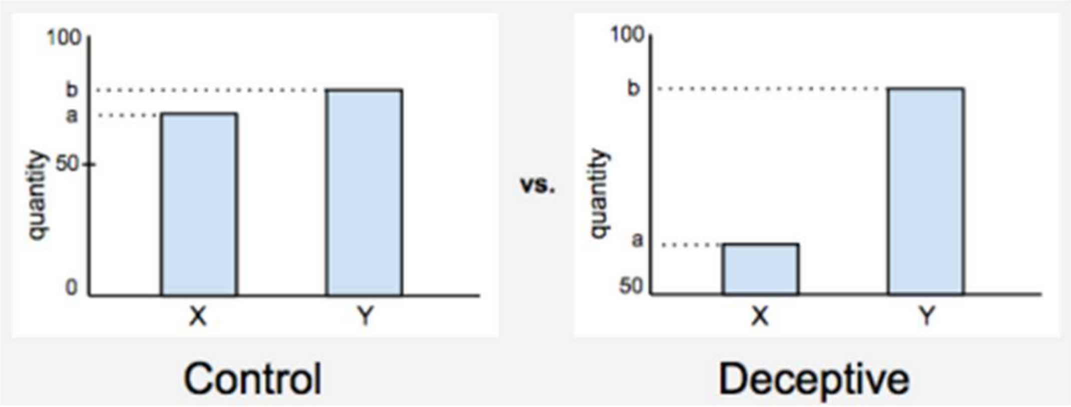
- Deception may occur at two levels
  - Chart level: The user reads the chart incorrectly, and/or processes an incorrect estimate of the data presented
  - Message level: Users interpret the message incorrectly
    - Message Exaggeration/Understatement:
      - This kind of deception happens when the fact is not distorted, however, but the extent of the presented fact is tweaked, i.e., the fact is exaggerated.
      - For example, if a chart compares two quantities - A and B, where A is bigger than B, but the users are presented with the fact that A is bigger than B, but the extent is exaggerated.
    - Message Reversal:
      - This type of deception happens when a visualization encourages users to interpret the fact in the message incorrectly.
      - For example, if a chart compares two quantities - A and B, where A is bigger than B, the users perceive the message as A is smaller than B.

# Distortion Techniques

- Some of the visualization techniques that are widely used for deception:
  - Truncated Axis
  - Area as quantity
  - Aspect ratio
  - Inverted axis

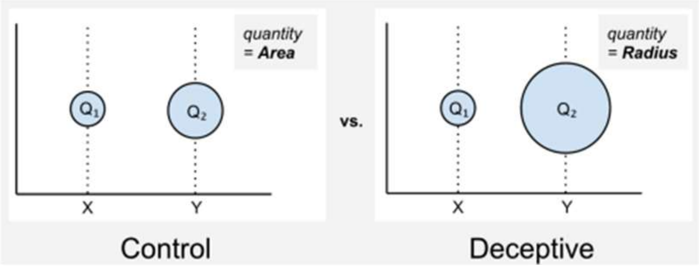
# Truncated Axis

- In the truncated axis visual distortion, one or more of the axes of a chart are altered by changing the minimum and maximum values presented on the scale
- Leads to exaggeration or understatement of the quantities presented



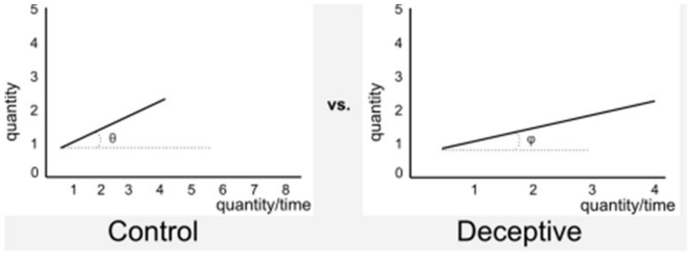
# Area as Quantity

- Encoding quantitative data with size has faced serious criticisms in the visualization community, and is a process that requires careful mapping of data with graphics.
- Although no guidelines are available about how to map the actual data with graphical area, it is believed that a one-to-one mapping between the data and the graphical area is least prone to distortion
- Leads to exaggeration or understatement of the quantities presented



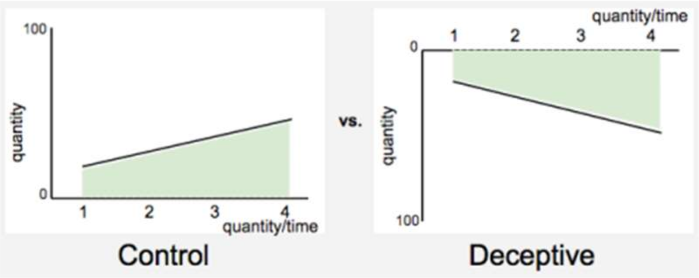
# Aspect Ratio

- This type of distortion primarily affects line-charts as it directly impacts the rate of increase or decrease of one quantity over another.
- May impact other visualizations such as bar-charts also.
- This type of distortion also leads to message exaggeration/understatement.



# Inverted Axis

- Human beings relate directions with trends, such as: upwards - increase, downwards - decrease, right - front/progress, left - back/receding
- This directional interpretation makes inverted axis one the most common distortion techniques
- Leads to reversal of the message, and makes the users susceptible to drawing false conclusions

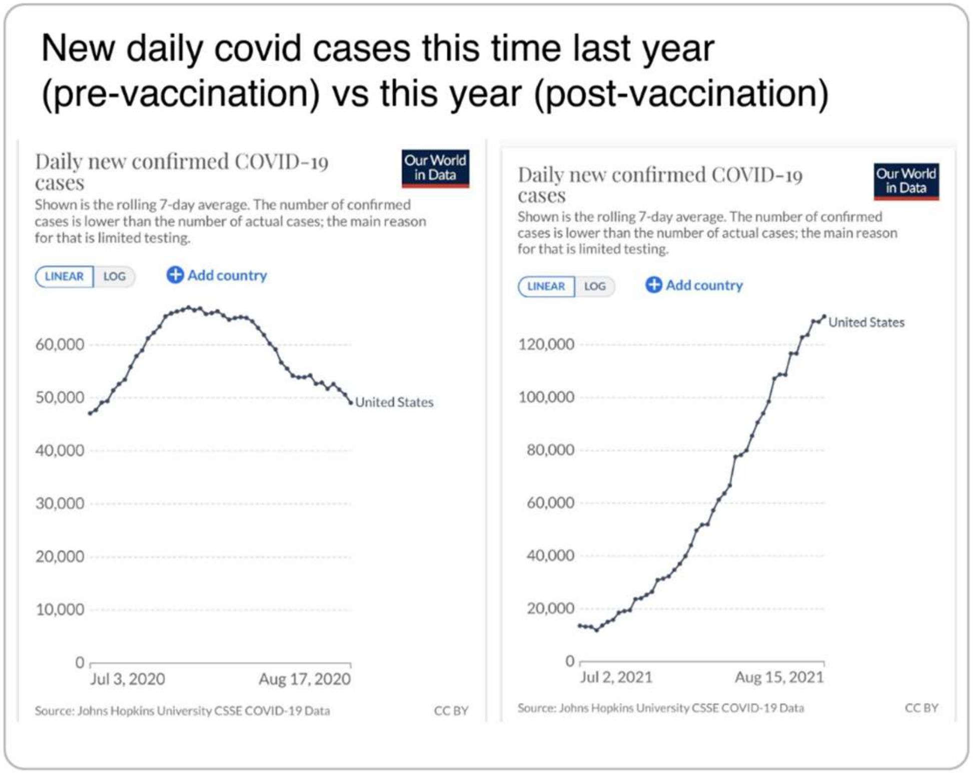


# Interpreting Visualizations Incorrectly

- 6 common types of reasoning errors:
  - Cherry-picking
  - Setting an arbitrary threshold
  - Incorrectly inferring causality
  - Issues caused by data validity
  - Failing to account for statistical nuance
  - Misrepresenting scientific results.

# Cherry Picking

- Visualization posts employ cherry-picking when the main conclusion is consistent with the incomplete evidence presented but likely would not be generalizable with more representative evidence.
- COVID case curves for the US for the periods of early July through mid August for the years 2020 (before the vaccination campaign) and 2021 (during the vaccination campaign).



New daily covid cases this time last year (pre-vaccination) vs this year (post-vaccination)

Daily new confirmed COVID-19 cases

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Source: Johns Hopkins University CSSE COVID-19 Data

Daily new confirmed COVID-19 cases

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

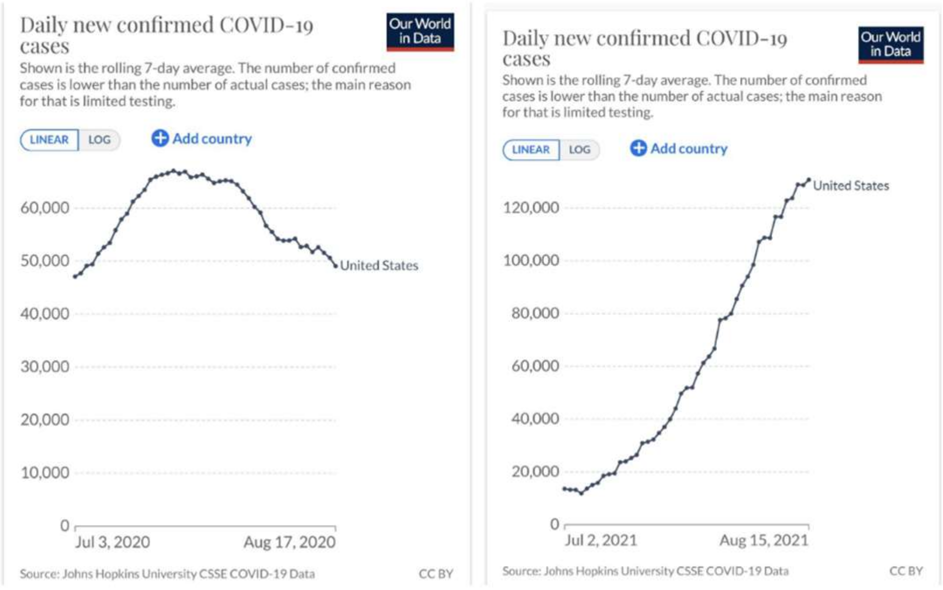
Source: Johns Hopkins University CSSE COVID-19 Data

- Because the number of cases in August 2021 is higher than in August 2020, has the vaccination campaign failed?

# Cherry Picking

- The implication is based on a single county
- Carefully selects the time frame that most effectively supports the argument, omitting a large drop in cases in Spring 2021.

New daily covid cases this time last year (pre-vaccination) vs this year (post-vaccination)



Daily new confirmed COVID-19 cases

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

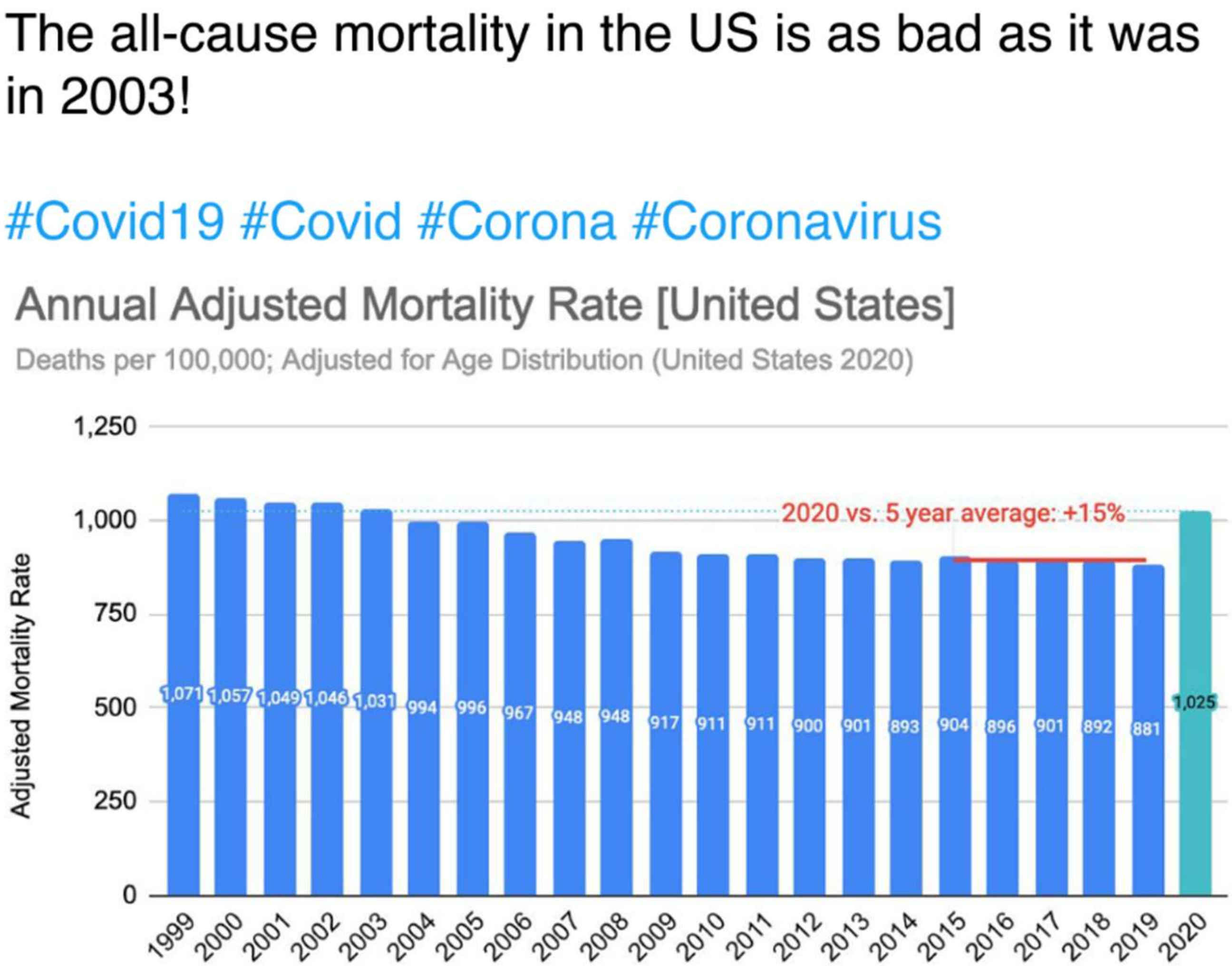
Daily new confirmed COVID-19 cases

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

# Setting an arbitrary threshold



The all-cause mortality in the US is as bad as it was in 2003!

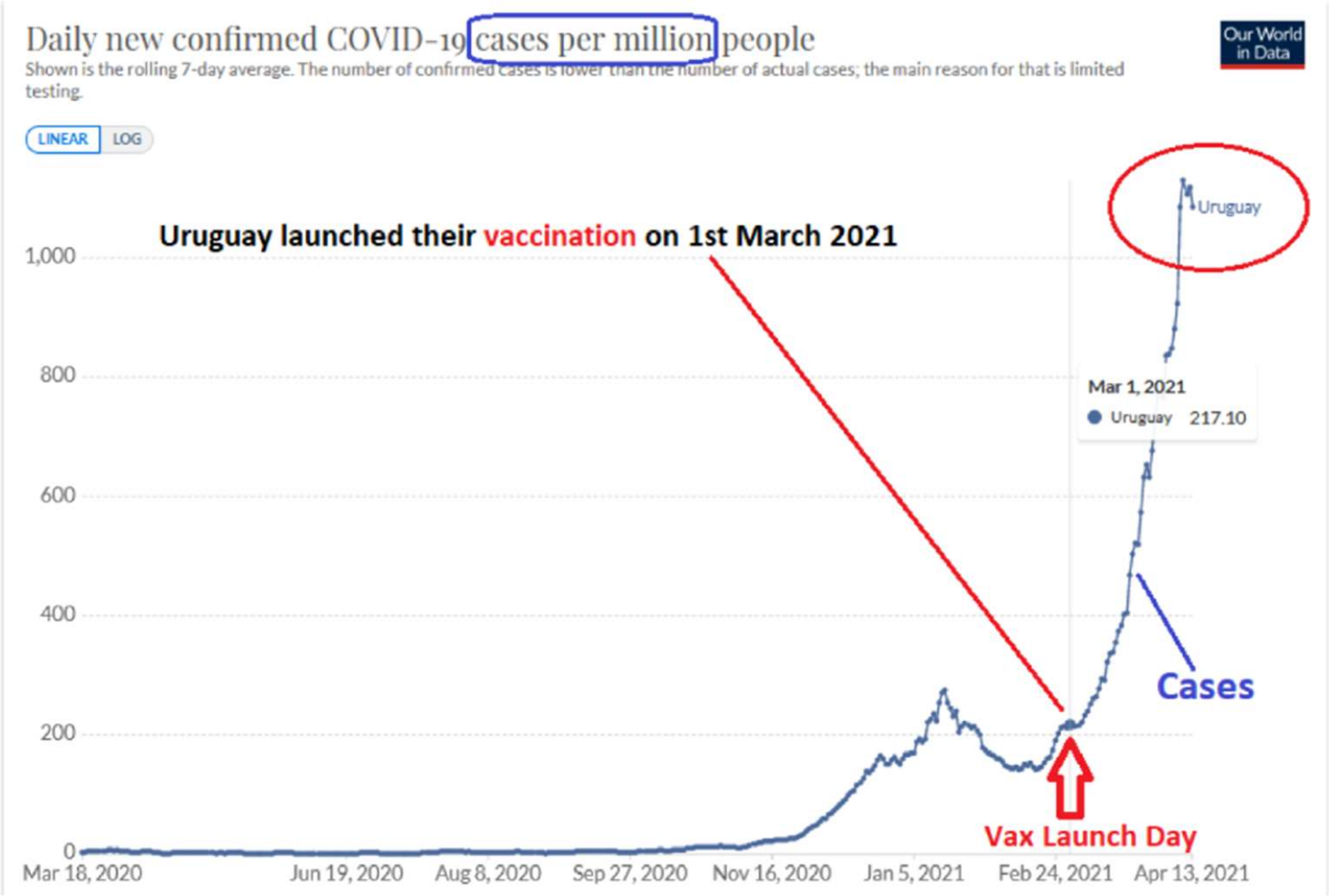
#Covid19 #Covid #Corona #Coronavirus

Annual Adjusted Mortality Rate [United States]

Deaths per 100,000; Adjusted for Age Distribution (United States 2020)

- An author shares a chart of the annual mortality rate in the US and argues that an increase in deaths of “only” 15% is not significant enough.

# Incorrectly inferring causality



Daily new confirmed COVID-19 cases per million people

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Uruguay launched their vaccination on 1st March 2021

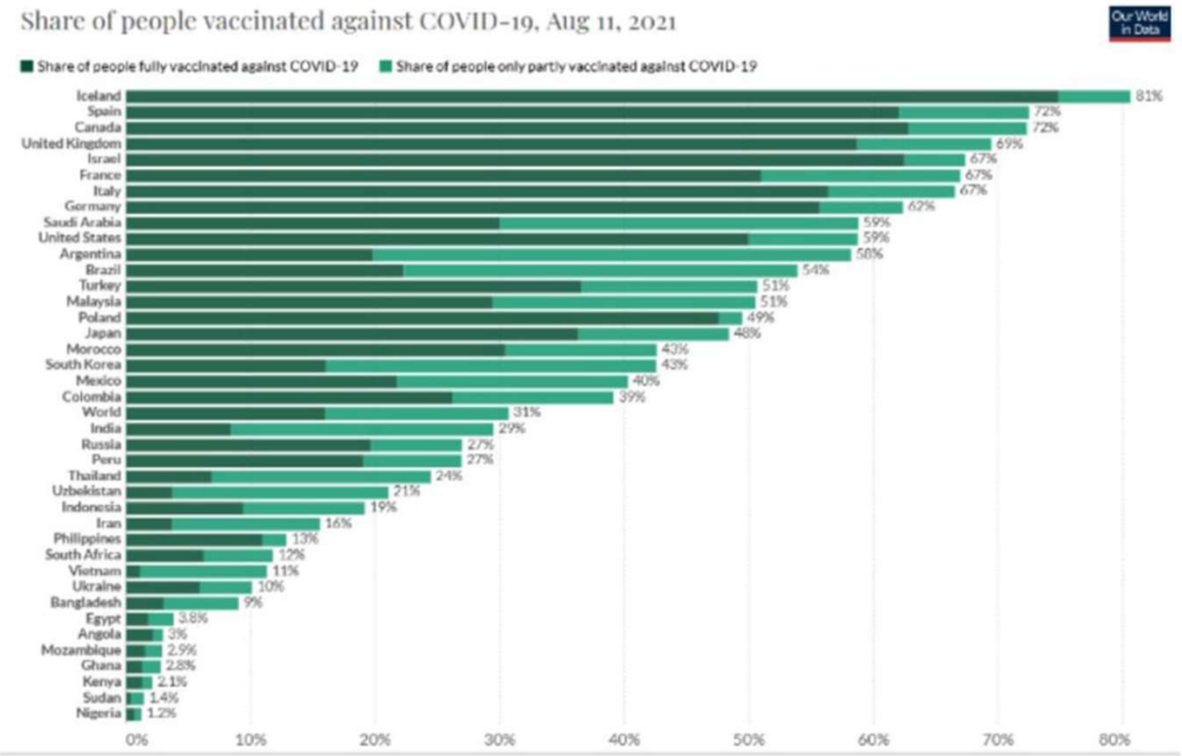
- Suggests that the sharp increase in COVID-19 cases in Uruguay—a prominent feature of the chart—was caused by the vaccination campaign and that vaccines are harmful.

# Issues With data validity

Most vaccinated: Iceland 81%

Least vaccinated: Nigeria 1.2%

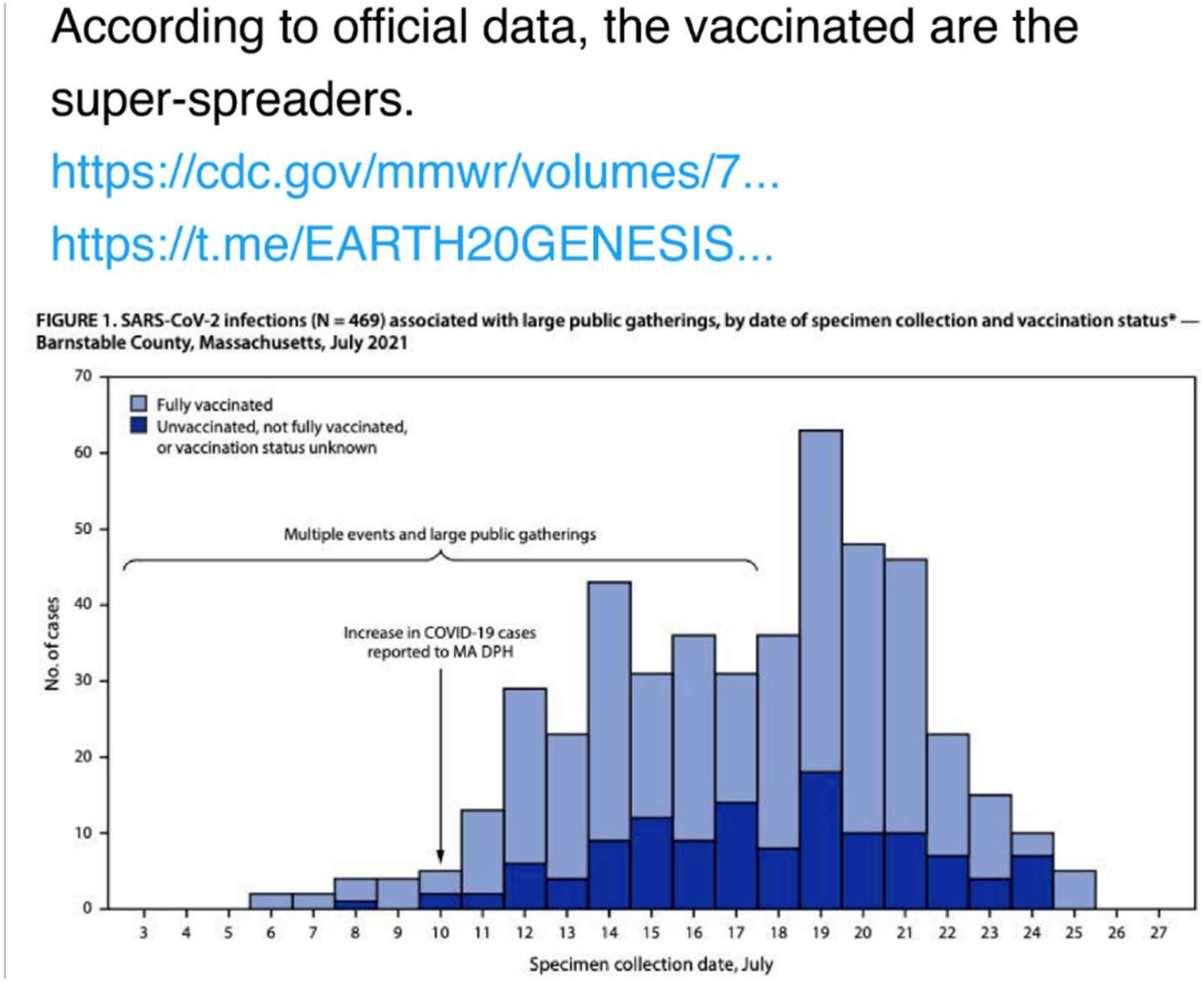
Iceland has 119 times more Covid cases



Share of people vaccinated against COVID-19, Aug 11, 2021

- Visualization suggests Iceland is more vaccinated than Nigeria, it is experiencing more COVID-19 cases, implying that vaccines are not effective.
- Fails to account, however, for the fact that Iceland had a much higher testing rate—roughly 200 times as high at the time of posting—making it unreasonable to compare the two countries.

# Failure to account for statistical nuance



According to official data, the vaccinated are the super-spreaders.

https://cdc.gov/mmwr/volumes/7...

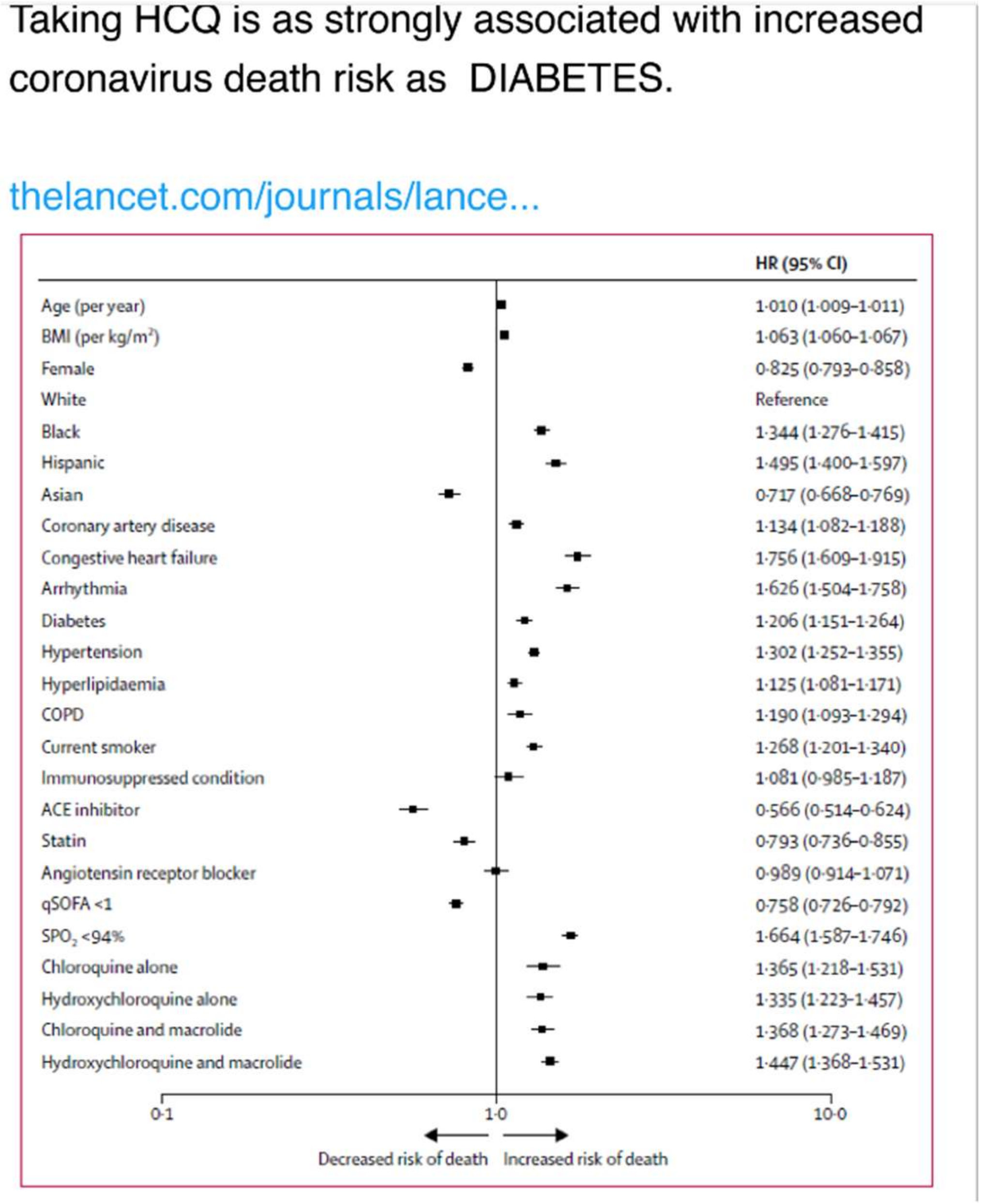
https://t.me/EARTH20GENESIS...

FIGURE 1. SARS-CoV-2 infections (N = 469) associated with large public gatherings, by date of specimen collection and vaccination status* — Barnstable County, Massachusetts, July 2021

▨ Fully vaccinated
■ Unvaccinated, not fully vaccinated, or vaccination status unknown
Multiple events and large public gatherings
Increase in COVID-19 cases reported to MA DPH
No. of cases
0
10
20
30
40
50
60
70
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
Specimen collection date, July

- Suggests since there are more vaccinated cases, the vaccinated are “super-spreaders.”
- Fails to account for the high proportion of vaccinated in the general population—likely as high as 95% at the time of data collection.

# Misrepresentation of scientific results



Taking HCQ is as strongly associated with increased coronavirus death risk as DIABETES.

thelancet.com/journals/lance...

| | HR (95% CI) |
|---|---|
| Age (per year) | 1·010 (1·009–1·011) |
| BMI (per kg/m²) | 1·063 (1·060–1·067) |
| Female | 0·825 (0·793–0·858) |
| White | Reference |
| Black | 1·344 (1·276–1·415) |
| Hispanic | 1·495 (1·400–1·597) |
| Asian | 0·717 (0·668–0·769) |
| Coronary artery disease | 1·134 (1·082–1·188) |
| Congestive heart failure | 1·756 (1·609–1·915) |
| Arrhythmia | 1·626 (1·504–1·758) |
| Diabetes | 1·206 (1·151–1·264) |
| Hypertension | 1·302 (1·252–1·355) |
| Hyperlipidaemia | 1·125 (1·081–1·171) |
| COPD | 1·190 (1·093–1·294) |
| Current smoker | 1·268 (1·201–1·340) |
| Immunosuppressed condition | 1·081 (0·985–1·187) |
| ACE inhibitor | 0·566 (0·514–0·624) |
| Statin | 0·793 (0·736–0·855) |
| Angiotensin receptor blocker | 0·989 (0·914–1·071) |
| qSOFA <1 | 0·758 (0·726–0·792) |
| SPO$_2$ <94% | 1·664 (1·587–1·746) |
| Chloroquine alone | 1·365 (1·218–1·531) |
| Hydroxychloroquine alone | 1·335 (1·223–1·457) |
| Chloroquine and macrolide | 1·368 (1·273–1·469) |
| Hydroxychloroquine and macrolide | 1·447 (1·368–1·531) |

- Sometimes users accept any scientific findings that align with their prior beliefs at face value —such as studies on the efficacy of certain types of medication that have not yet been peer-reviewed, reproduced, or otherwise scrutinized
- For instance, this tweet argues against the use of hydroxychloroquine noting that it leads to an increased risk of mortality.
- The figure comes from a study that has since been retracted due to concerns about the veracity of the data.

# Overview

- Graphical Excellence
- Graphical Integrity
- Design Aesthetics

# Design Aesthetics

- Set of principles to help guide designers
  I. Maximize data-ink ratio
  II. Maximize data density
  III. Content is king
  IV. Avoid “Chart Junk”
  V. Utilize multifunctioning graphical elements
  VI. Use small multiples
  VII. Be careful with color

# Maximize data-ink ratio

- Data ink ratio = (Data ink)/(Total ink used in graphic)

  = the ratio of elements in a visual representation conveying information to the total elements in the visualization

  = 1 – (proportion of graphic data that can be erased without loss of information)

- Data ink ratio can be used to make sure that all the elements displayed in a visualization are relevant to the information being displayed and that no element in the chart be redundant.

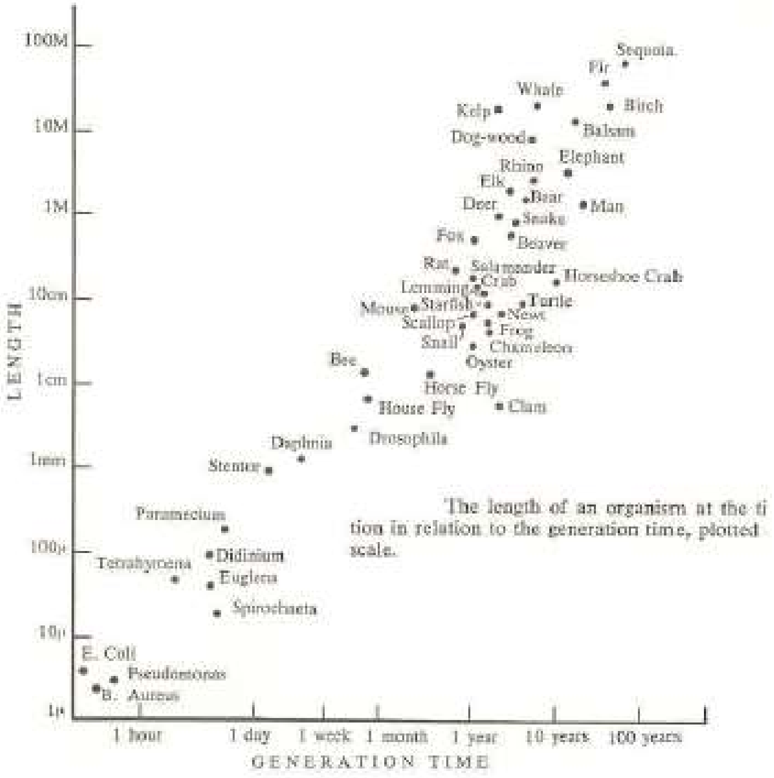
# Reasons for low data-ink ratio

- The use of 3-D effects and shadow effects in a data visualization doesn’t add any extra information. Hence, it decreases the data-ink ratio.
- The use of background images can also be unnecessary in the visualizations and may decrease the data-ink ratio.
- Unnecessary borders and grid lines don’t convey any information to the user. Most of the time, grids and borders are redundant and decrease the data-ink ratio.
- Adding redundant legends, bold labels, and other decorative elements also reduces the data-ink ratio.

# Tufte's 5 Laws of Data Ink

1. **Show the data**: Keep in mind that we need to show the data to the viewer. Hence, we should show all the relevant data in the chart. The data should be the number one priority.
2. **Maximize the data-ink ratio**: While presenting the data, we should focus on maximizing the data-ink ratio.
3. **Erase non-data ink**: To increase the data-ink ratio, we should erase all the elements of the visualization that don't contribute any information like 3-D effects, grids, annotations, colors, and borders.
4. **Erase redundant data ink**: We should also delete the elements that show redundant data from the visualization. The elements that often fall in this category are legends, labels, and information unrelated to the visualization.
5. **Revise and edit**: While creating any visualization, we should critically evaluate it and make sure that we have proper elements in the visualization and that there is no redundancy.

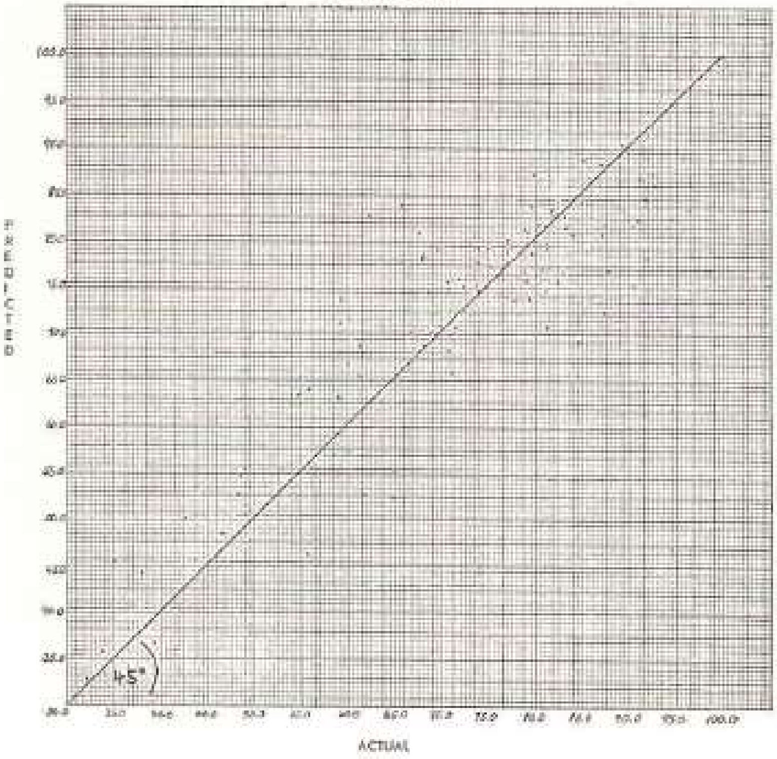
# Data-Ink Ratio



The length of an organism at the ti
tion in relation to the generation time, plotted
scale.

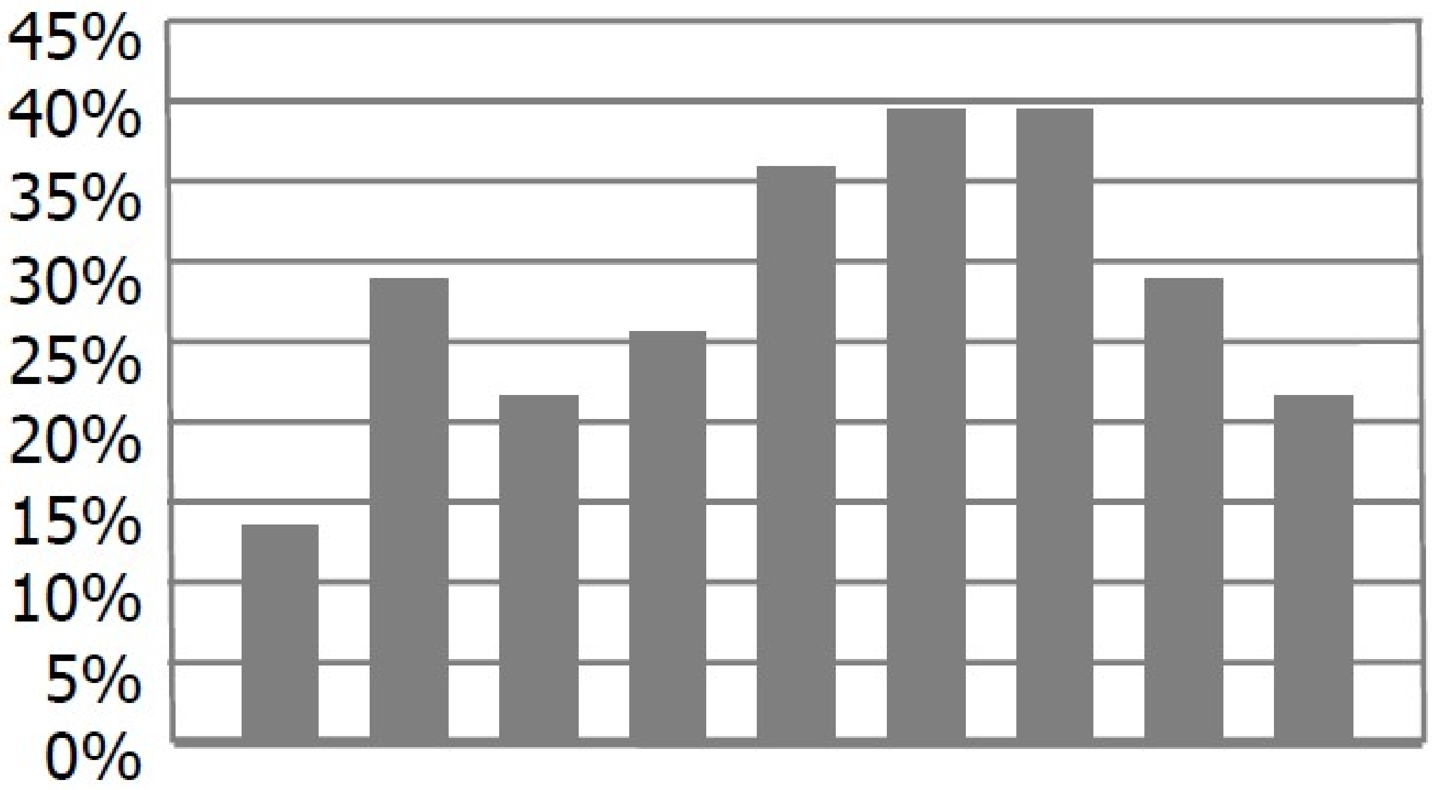
Good

Relationship of Actual Rates of Registration to Predicted Rates
(104 cities 1900).

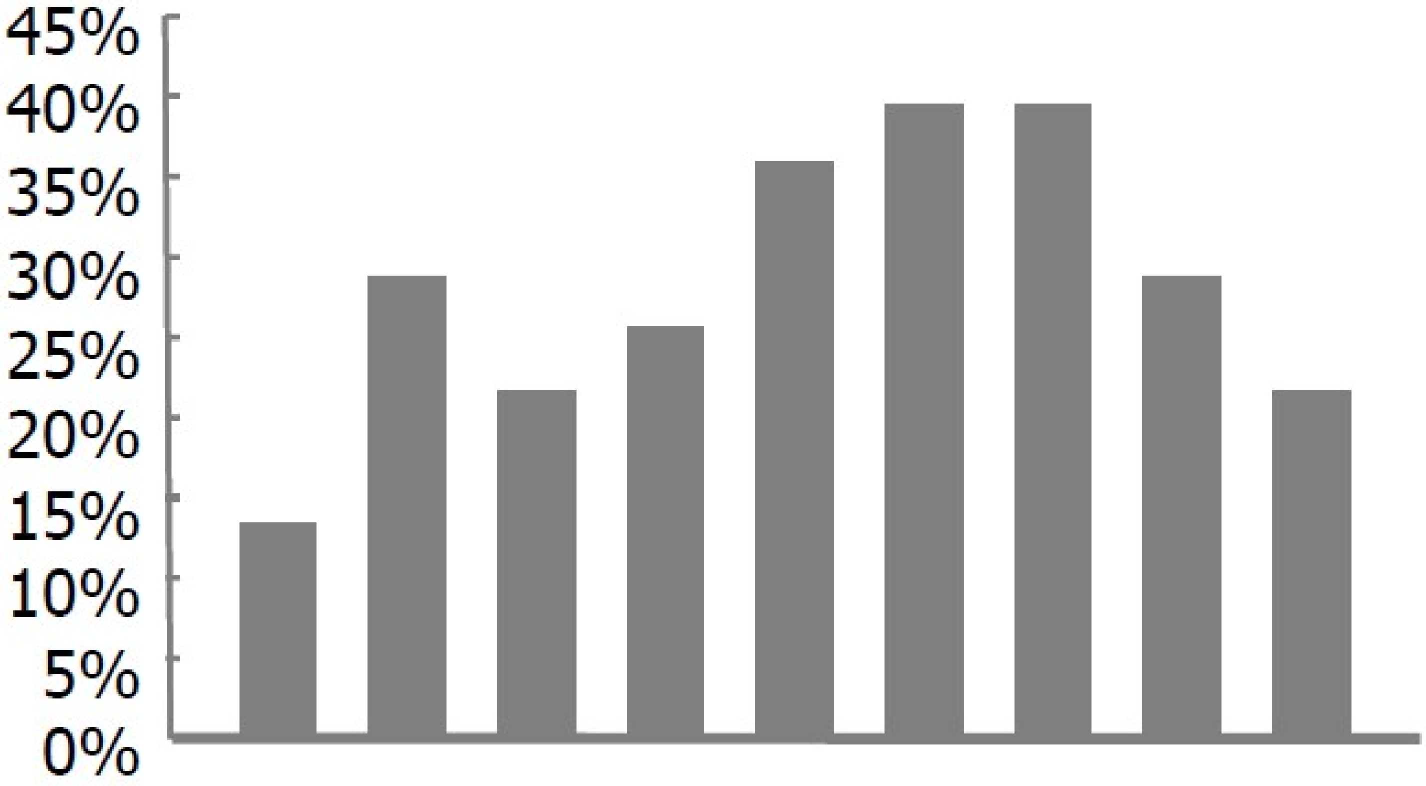


Bad

# Bar Chart - Bad

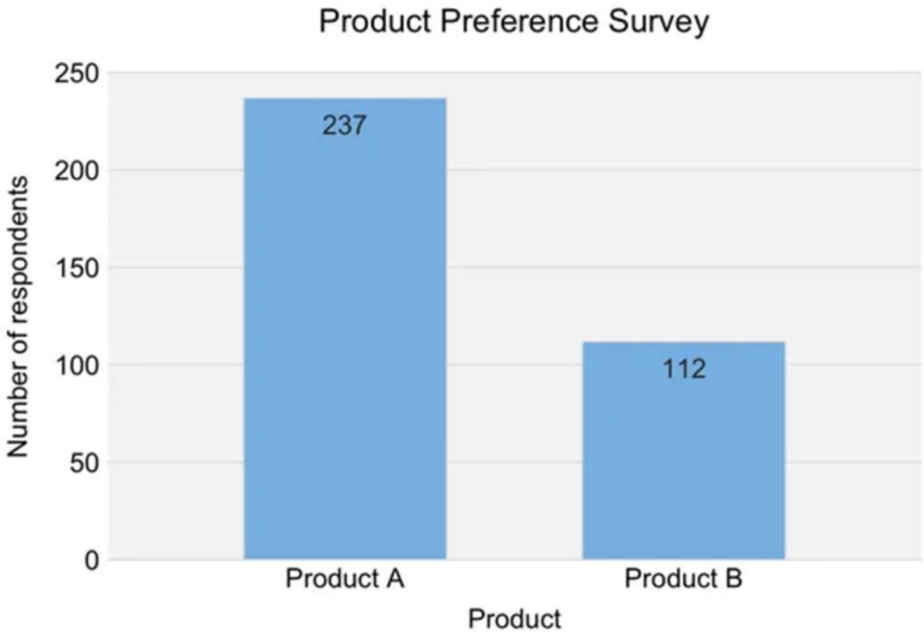


# Bar Chart - Good

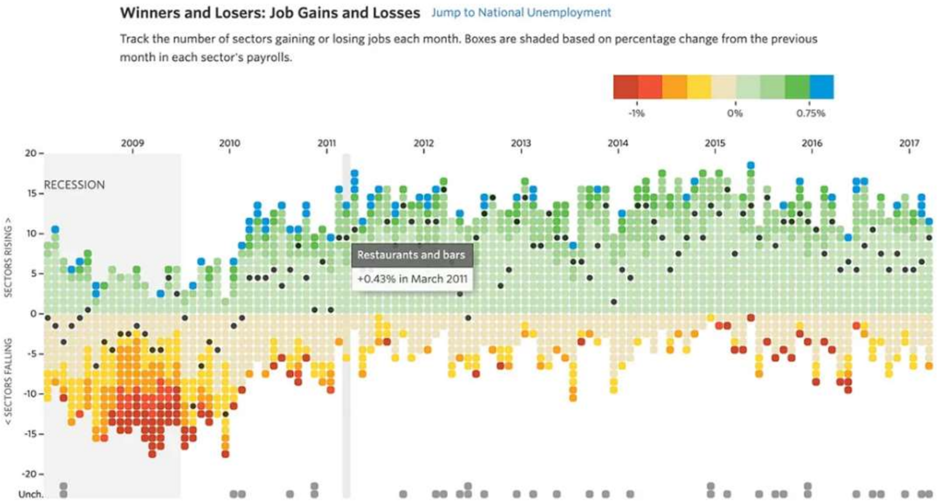


# Data Density

- Data Density = (Number of entries in data matrix)/(Area of data graphic)



Bad Data Density


Winners and Losers: Job Gains and Losses

Track the number of sectors gaining or losing jobs each month. Boxes are shaded based on percentage change from the previous month in each sector's payrolls.

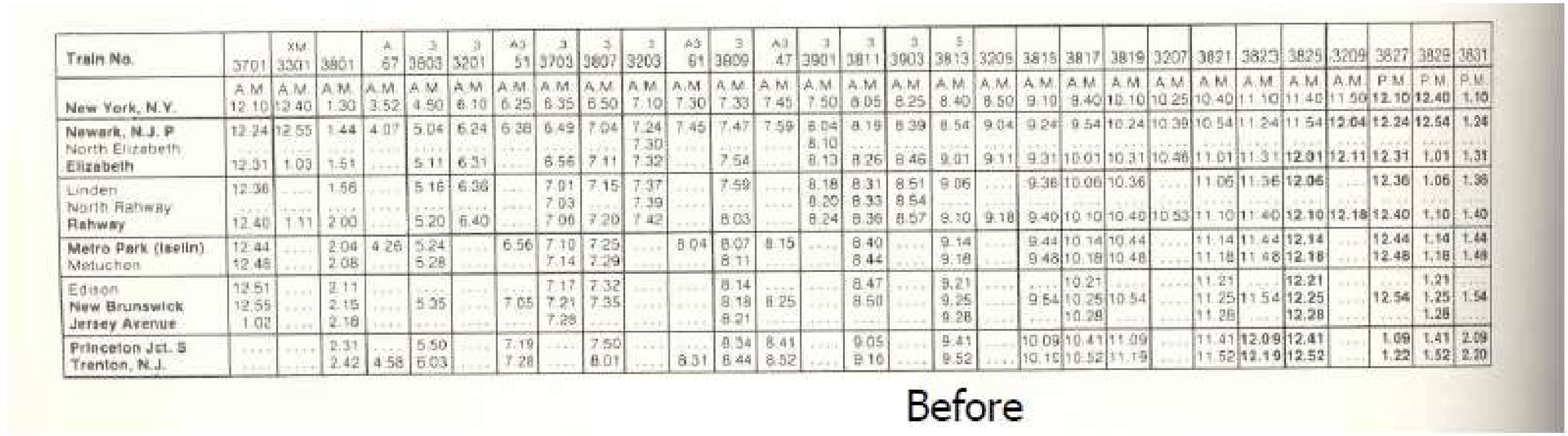
Good Data Density

# Maximize Data Density

- “Data-rich designs give a context and credibility to statistical evidence. Low-information designs are suspect: what is left out, what is hidden, why are we shown so little? High-density graphics help us to compare parts of the data by displaying much information within the view of the eye: we look at one page at a time and the more on the page, the more effective and comparative our eye can be. The principle, then, is: Maximize data density and the size of the data matrix, within reason.”
- Edward Tufte

# Content is King

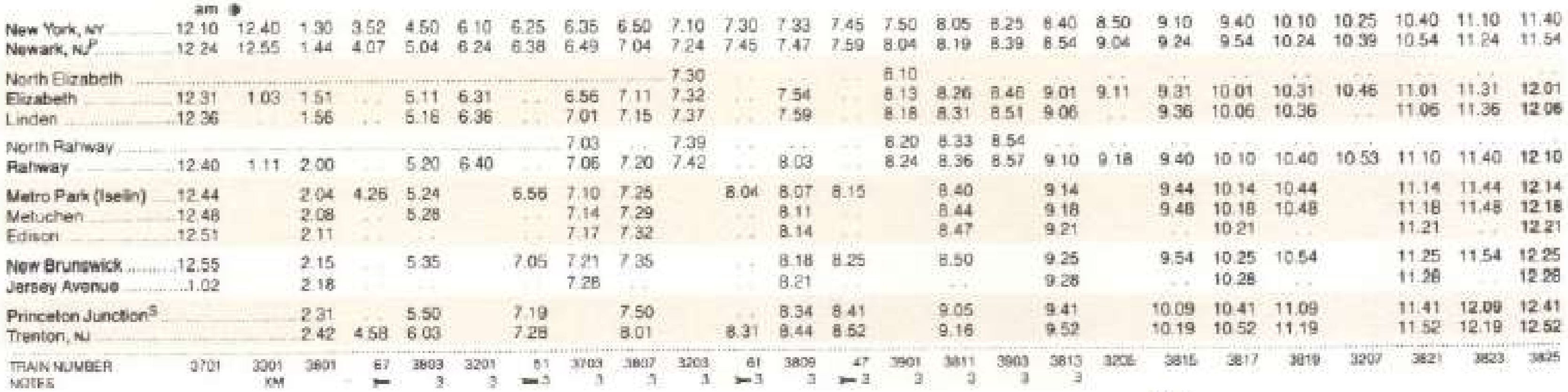
- Avoid separate legends and keys -- Just have that information in the graphic
- Make grids, labeling, etc., very faint so that they recede into background
- Integrate text, chart, graphic, map into a coherent narrative



| Train No. | 3701 | XM<br>3301 | 3801 | A<br>67 | 3<br>3803 | 3<br>3201 | A3<br>51 | 3<br>3703 | 3<br>3807 | 3<br>3203 | A3<br>61 | 3<br>3809 | A3<br>47 | 3<br>3901 | 3<br>3811 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. |
| **New York, N.Y.** | 12.10 | 12.40 | 1.30 | 3.52 | 4.50 | 6.10 | 6.25 | 6.35 | 6.50 | 7.10 | 7.30 | 7.33 | 7.45 | 7.50 | 8.05 |
| **Newark, N.J. P** | 12.24 | 12.55 | 1.44 | 4.07 | 5.04 | 6.24 | 6.38 | 6.49 | 7.04 | 7.24 | 7.45 | 7.47 | 7.59 | 8.04 | 8.19 |
| North Elizabeth | .... | .... | .... | .... | .... | .... | .... | .... | .... | 7.30 | .... | .... | .... | 8.10 | .... |
| **Elizabeth** | 12.31 | 1.03 | 1.51 | .... | 5.11 | 6.31 | .... | 6.56 | 7.11 | 7.32 | .... | 7.54 | .... | 8.13 | 8.26 |
| Linden | 12.36 | .... | 1.56 | .... | 5.16 | 6.36 | .... | 7.01 | 7.15 | 7.37 | .... | 7.59 | .... | 8.18 | 8.31 |
| North Rahway | .... | .... | .... | .... | .... | .... | .... | 7.03 | .... | 7.39 | .... | .... | .... | 8.20 | 8.33 |
| **Rahway** | 12.40 | 1.11 | 2.00 | .... | 5.20 | 6.40 | .... | 7.06 | 7.20 | 7.42 | .... | 8.03 | .... | 8.24 | 8.36 |
| **Metro Park (Iselin)** | 12.44 | .... | 2.04 | 4.26 | 5.24 | .... | 6.56 | 7.10 | 7.25 | .... | 8.04 | 8.07 | 8.15 | .... | 8.40 |
| Metuchen | 12.48 | .... | 2.08 | .... | 5.28 | .... | .... | 7.14 | 7.29 | .... | .... | 8.11 | .... | .... | 8.44 |
| Edison | 12.51 | .... | 2.11 | .... | .... | .... | .... | 7.17 | 7.32 | .... | .... | 8.14 | .... | .... | 8.47 |
| **New Brunswick** | 12.55 | .... | 2.15 | .... | 5.35 | .... | 7.05 | 7.21 | 7.35 | .... | .... | 8.18 | 8.25 | .... | 8.50 |
| **Jersey Avenue** | 1.02 | .... | 2.18 | .... | .... | .... | .... | 7.28 | .... | .... | .... | 8.21 | .... | .... | .... |
| **Princeton Jct. S** | .... | .... | 2.31 | .... | 5.50 | .... | 7.19 | .... | 7.50 | .... | .... | 8.34 | 8.41 | .... | 9.05 |
| **Trenton, N.J.** | .... | .... | 2.42 | 4.58 | 6.03 | .... | 7.28 | .... | 8.01 | .... | 8.31 | 8.44 | 8.52 | .... | 9.16 |

| Train No. | 3<br>3903 | 3<br>3813 | 3205 | 3815 | 3817 | 3819 | 3207 | 3821 | 3823 | 3825 | 3209 | 3827 | 3829 | 3831 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | A.M. | P.M. | P.M. | P.M. |
| **New York, N.Y.** | 8.25 | 8.40 | 8.50 | 9.10 | 9.40 | 10.10 | 10.25 | 10.40 | 11.10 | 11.40 | 11.50 | 12.10 | 12.40 | 1.10 |
| **Newark, N.J. P** | 8.39 | 8.54 | 9.04 | 9.24 | 9.54 | 10.24 | 10.39 | 10.54 | 11.24 | 11.54 | 12.04 | 12.24 | 12.54 | 1.24 |
| North Elizabeth | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... |
| **Elizabeth** | 8.46 | 9.01 | 9.11 | 9.31 | 10.01 | 10.31 | 10.46 | 11.01 | 11.31 | 12.01 | 12.11 | 12.31 | 1.01 | 1.31 |
| Linden | 8.51 | 9.06 | .... | 9.36 | 10.06 | 10.36 | .... | 11.06 | 11.36 | 12.06 | .... | 12.36 | 1.06 | 1.36 |
| North Rahway | 8.54 | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... | .... |
| **Rahway** | 8.57 | 9.10 | 9.18 | 9.40 | 10.10 | 10.40 | 10.53 | 11.10 | 11.40 | 12.10 | 12.18 | 12.40 | 1.10 | 1.40 |
| **Metro Park (Iselin)** | .... | 9.14 | .... | 9.44 | 10.14 | 10.44 | .... | 11.14 | 11.44 | 12.14 | .... | 12.44 | 1.14 | 1.44 |
| Metuchen | .... | 9.18 | .... | 9.48 | 10.18 | 10.48 | .... | 11.18 | 11.48 | 12.18 | .... | 12.48 | 1.18 | 1.48 |
| Edison | .... | 9.21 | .... | .... | 10.21 | .... | .... | 11.21 | .... | 12.21 | .... | .... | 1.21 | .... |
| **New Brunswick** | .... | 9.25 | .... | 9.54 | 10.25 | 10.54 | .... | 11.25 | 11.54 | 12.25 | .... | 12.54 | 1.25 | 1.54 |
| **Jersey Avenue** | .... | 9.28 | .... | .... | 10.28 | .... | .... | 11.28 | .... | 12.28 | .... | .... | 1.28 | .... |
| **Princeton Jct. S** | .... | 9.41 | .... | 10.09 | 10.41 | 11.09 | .... | 11.41 | 12.09 | 12.41 | .... | 1.09 | 1.41 | 2.09 |
| **Trenton, N.J.** | .... | 9.52 | .... | 10.19 | 10.52 | 11.19 | .... | 11.52 | 12.19 | 12.52 | .... | 1.22 | 1.52 | 2.20 |

Before



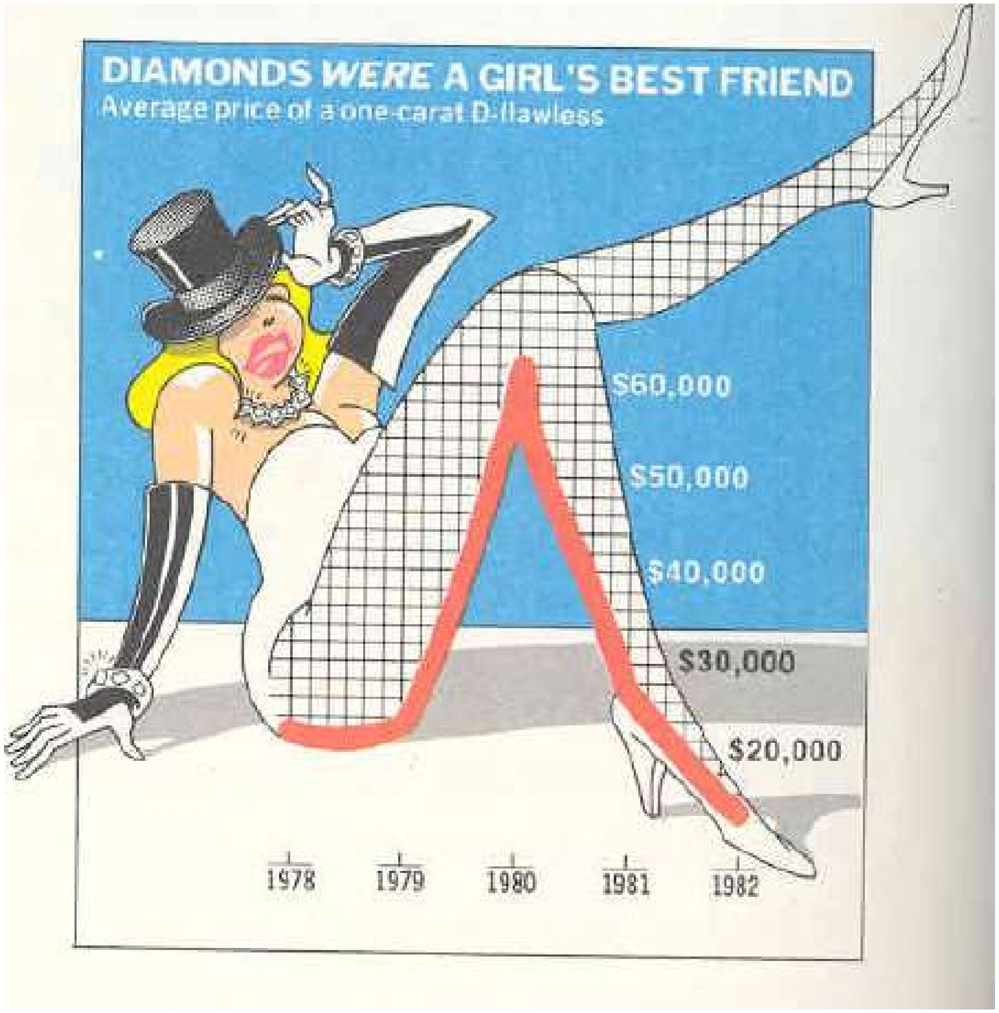
| | am | ● | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **New York, NY** | 12.10 | 12.40 | 1.30 | 3.52 | 4.50 | 6.10 | 6.25 | 6.35 | 6.50 | 7.10 | 7.30 | 7.33 | 7.45 |
| **Newark, NJ P** | 12.24 | 12.55 | 1.44 | 4.07 | 5.04 | 6.24 | 6.38 | 6.49 | 7.04 | 7.24 | 7.45 | 7.47 | 7.59 |
| North Elizabeth | | | | | | | | | | 7.30 | | | . . |
| **Elizabeth** | 12.31 | 1.03 | 1.51 | . . | 5.11 | 6.31 | . . | 6.56 | 7.11 | 7.32 | . . | 7.54 | . . |
| Linden | 12.36 | . . | 1.56 | . . | 5.16 | 6.36 | . . | 7.01 | 7.15 | 7.37 | . . | 7.59 | . . |
| North Rahway | | | | | | | | 7.03 | . . | 7.39 | . . | . . | . . |
| **Rahway** | 12.40 | 1.11 | 2.00 | . . | 5.20 | 6.40 | . . | 7.06 | 7.20 | 7.42 | . . | 8.03 | . . |
| **Metro Park (Iselin)** | 12.44 | | 2.04 | 4.26 | 5.24 | | 6.56 | 7.10 | 7.25 | | 8.04 | 8.07 | 8.15 |
| Metuchen | 12.48 | | 2.08 | . . | 5.28 | | . . | 7.14 | 7.29 | | . . | 8.11 | . . |
| Edison | 12.51 | | 2.11 | . . | . . | | . . | 7.17 | 7.32 | | . . | 8.14 | . . |
| **New Brunswick** | 12.55 | | 2.15 | . . | 5.35 | | 7.05 | 7.21 | 7.35 | | . . | 8.18 | 8.25 |
| **Jersey Avenue** | 1.02 | | 2.18 | . . | . . | | . . | 7.28 | . . | | . . | 8.21 | . . |
| **Princeton Junction S** | | | 2.31 | . . | 5.50 | | 7.19 | | 7.50 | | . . | 8.34 | 8.41 |
| **Trenton, NJ** | | | 2.42 | 4.58 | 6.03 | | 7.28 | | 8.01 | | 8.31 | 8.44 | 8.52 |
| TRAIN NUMBER | 3701 | 3301 | 3801 | 67 | 3803 | 3201 | 51 | 3703 | 3807 | 3203 | 61 | 3809 | 47 |
| NOTES | | XM | | ▬ | 3 | 3 | ▬3 | 3 | 3 | 3 | ▬3 | 3 | ▬3 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **New York, NY** | 7.50 | 8.05 | 8.25 | 8.40 | 8.50 | 9.10 | 9.40 | 10.10 | 10.25 | 10.40 | 11.10 | 11.40 |
| **Newark, NJ P** | 8.04 | 8.19 | 8.39 | 8.54 | 9.04 | 9.24 | 9.54 | 10.24 | 10.39 | 10.54 | 11.24 | 11.54 |
| North Elizabeth | 8.10 | . . | | . . | . . | . . | . . | . . | . . | . . | . . | . . |
| **Elizabeth** | 8.13 | 8.26 | 8.46 | 9.01 | 9.11 | 9.31 | 10.01 | 10.31 | 10.46 | 11.01 | 11.31 | 12.01 |
| Linden | 8.18 | 8.31 | 8.51 | 9.06 | . . | 9.36 | 10.06 | 10.36 | . . | 11.06 | 11.36 | 12.06 |
| North Rahway | 8.20 | 8.33 | 8.54 | . . | . . | . . | . . | . . | . . | . . | . . | . . |
| **Rahway** | 8.24 | 8.36 | 8.57 | 9.10 | 9.18 | 9.40 | 10.10 | 10.40 | 10.53 | 11.10 | 11.40 | 12.10 |
| **Metro Park (Iselin)** | | 8.40 | | 9.14 | | 9.44 | 10.14 | 10.44 | | 11.14 | 11.44 | 12.14 |
| Metuchen | | 8.44 | | 9.18 | | 9.48 | 10.18 | 10.48 | | 11.18 | 11.48 | 12.18 |
| Edison | | 8.47 | | 9.21 | | . . | 10.21 | . . | | 11.21 | . . | 12.21 |
| **New Brunswick** | | 8.50 | | 9.25 | | 9.54 | 10.25 | 10.54 | | 11.25 | 11.54 | 12.25 |
| **Jersey Avenue** | | . . | | 9.28 | | . . | 10.28 | . . | | 11.28 | . . | 12.28 |
| **Princeton Junction S** | | 9.05 | | 9.41 | | 10.09 | 10.41 | 11.09 | | 11.41 | 12.09 | 12.41 |
| **Trenton, NJ** | | 9.16 | | 9.52 | | 10.19 | 10.52 | 11.19 | | 11.52 | 12.19 | 12.52 |
| TRAIN NUMBER | 3901 | 3811 | 3903 | 3813 | 3205 | 3815 | 3817 | 3819 | 3207 | 3821 | 3823 | 3825 |
| NOTES | 3 | 3 | 3 | 3 | | | | | | | | |

After

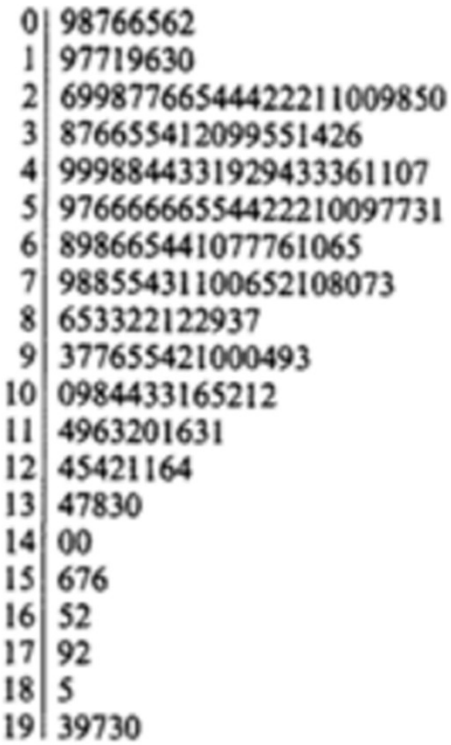
# Avoid chart junk

- Extraneous visual elements that detract from message



# Utilize Multifunctioning Graphical Elements

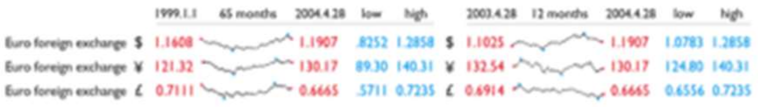
- Graphical elements that have many functions
- Example: Stem and leaf plot
  - Constructs distribution of the variables with the numbers themselves



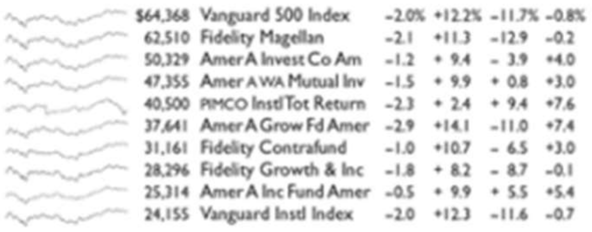
```
 0| 98766562
 1| 97719630
 2| 69987766544422211009850
 3| 876655412099551426
 4| 999884433192943336110 7
 5| 97666666554422210097731
 6| 898665441077761065
 7| 988554311006521080 73
 8| 6533221 22937
 9| 377655421000493
10| 0984433165212
11| 4963201631
12| 45421164
13| 47830
14| 00
15| 676
16| 52
17| 92
18| 5
19| 39730
```

# Use small multiples

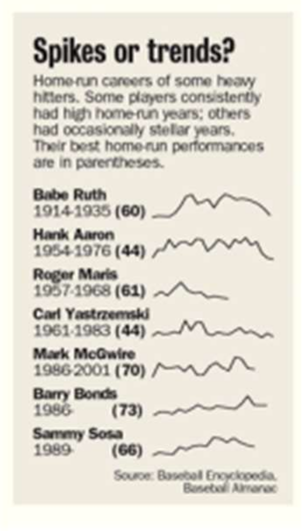
- Repeat visually similar graphical elements nearby rather than spreading far apart
- Example: Sparklines



| | | 1999.1.1 | 65 months | 2004.4.28 | low | high | | 2003.4.28 | 12 months | 2004.4.28 | low | high |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Euro foreign exchange | $ | 1.1608 | | 1.1907 | .8252 | 1.2858 | $ | 1.1025 | | 1.1907 | 1.0783 | 1.2858 |
| Euro foreign exchange | ¥ | 121.32 | | 130.17 | 89.30 | 140.31 | ¥ | 132.54 | | 130.17 | 124.80 | 140.31 |
| Euro foreign exchange | £ | 0.7111 | | 0.6665 | .5711 | 0.7235 | £ | 0.6914 | | 0.6665 | 0.6556 | 0.7235 |



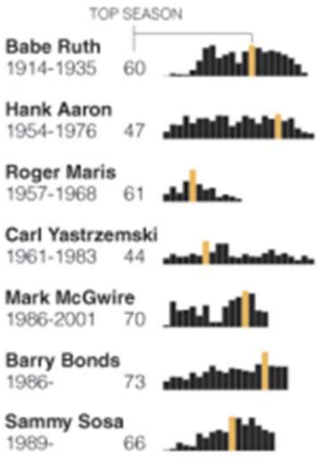
| | | | | | | |
|---|---|---|---|---|---|---|
| | $64,368 | Vanguard 500 Index | -2.0% | +12.2% | -11.7% | -0.8% |
| | 62,510 | Fidelity Magellan | -2.1 | +11.3 | -12.9 | -0.2 |
| | 50,329 | Amer A Invest Co Am | -1.2 | + 9.4 | - 3.9 | +4.0 |
| | 47,355 | Amer A WA Mutual Inv | -1.5 | + 9.9 | + 0.8 | +3.0 |
| | 40,500 | PIMCO Instl Tot Return | -2.3 | + 2.4 | + 9.4 | +7.6 |
| | 37,641 | Amer A Grow Fd Amer | -2.9 | +14.1 | -11.0 | +7.4 |
| | 31,161 | Fidelity Contrafund | -1.0 | +10.7 | - 6.5 | +3.0 |
| | 28,296 | Fidelity Growth & Inc | -1.8 | + 8.2 | - 8.7 | -0.1 |
| | 25,314 | Amer A Inc Fund Amer | -0.5 | + 9.9 | + 5.5 | +5.4 |
| | 24,155 | Vanguard Instl Index | -2.0 | +12.3 | -11.6 | -0.7 |

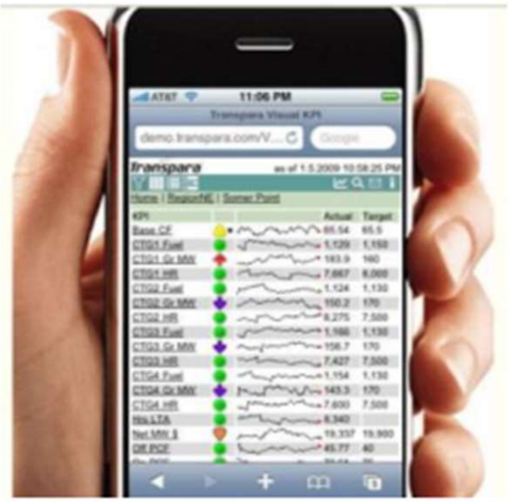


Spikes or trends?

Home-run careers of some heavy hitters. Some players consistently had high home-run years; others had occasionally stellar years. Their best home-run performances are in parentheses.

Source: Baseball Encyclopedia, Baseball Almanac





| KPI | | | Actual | Target |
|---|---|---|---|---|
| Base CF | ● | | 65.54 | 65.5 |
| CTG1 Fuel | ● | | 1,129 | 1,150 |
| CTG1 Gr MW | ● | | 183.9 | 160 |
| CTG1 HR | ● | | 7,867 | 8,000 |
| CTG2 Fuel | ● | | 1,124 | 1,130 |
| CTG2 Gr MW | ◆ | | 150.2 | 170 |
| CTG2 HR | ● | | 8,275 | 7,500 |
| CTG3 Fuel | ● | | 1,166 | 1,130 |
| CTG3 Gr MW | ◆ | | 156.7 | 170 |
| CTG3 HR | ● | | 7,427 | 7,500 |
| CTG4 Fuel | ● | | 1,154 | 1,130 |
| CTG4 Gr MW | ◆ | | 143.3 | 170 |
| CTG4 HR | ● | | 7,600 | 7,500 |
| Hrs LTA | ● | | 8,340 | |
| Net MW $ | ● | | 19,337 | 19,900 |
| Off PCF | ● | | 45.77 | 40 |

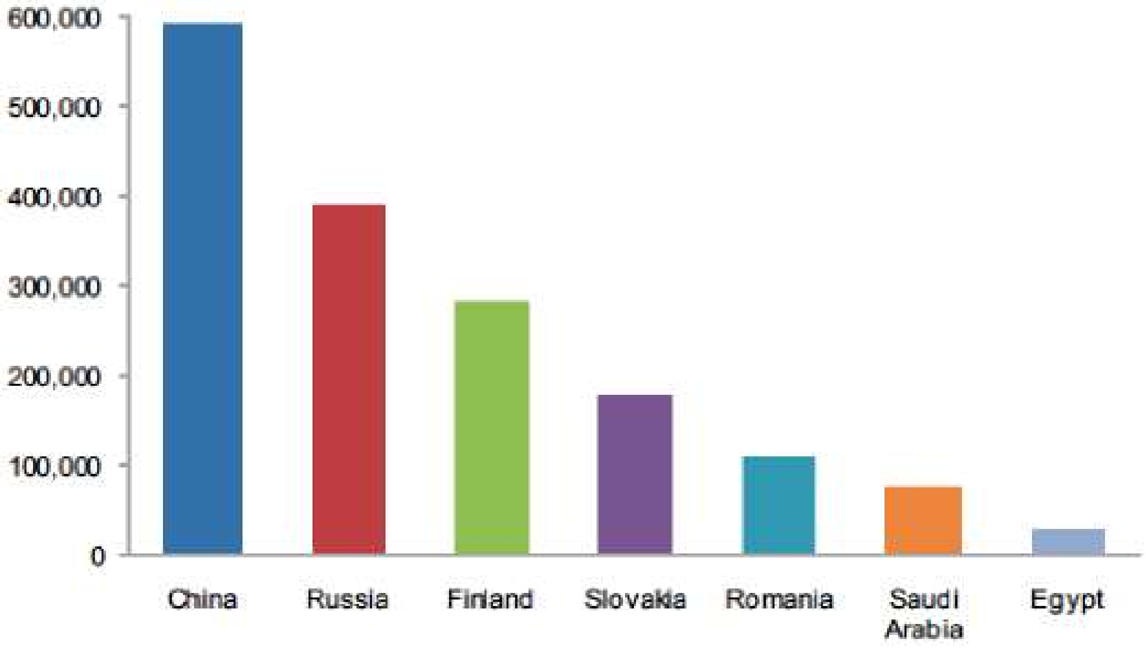
# Use Color to Communicate, not to Decorate!

- To differentiate information marks (red states, blue states)
- To show continuous variation of a value (as in temperature range)
- To call attention/highlight

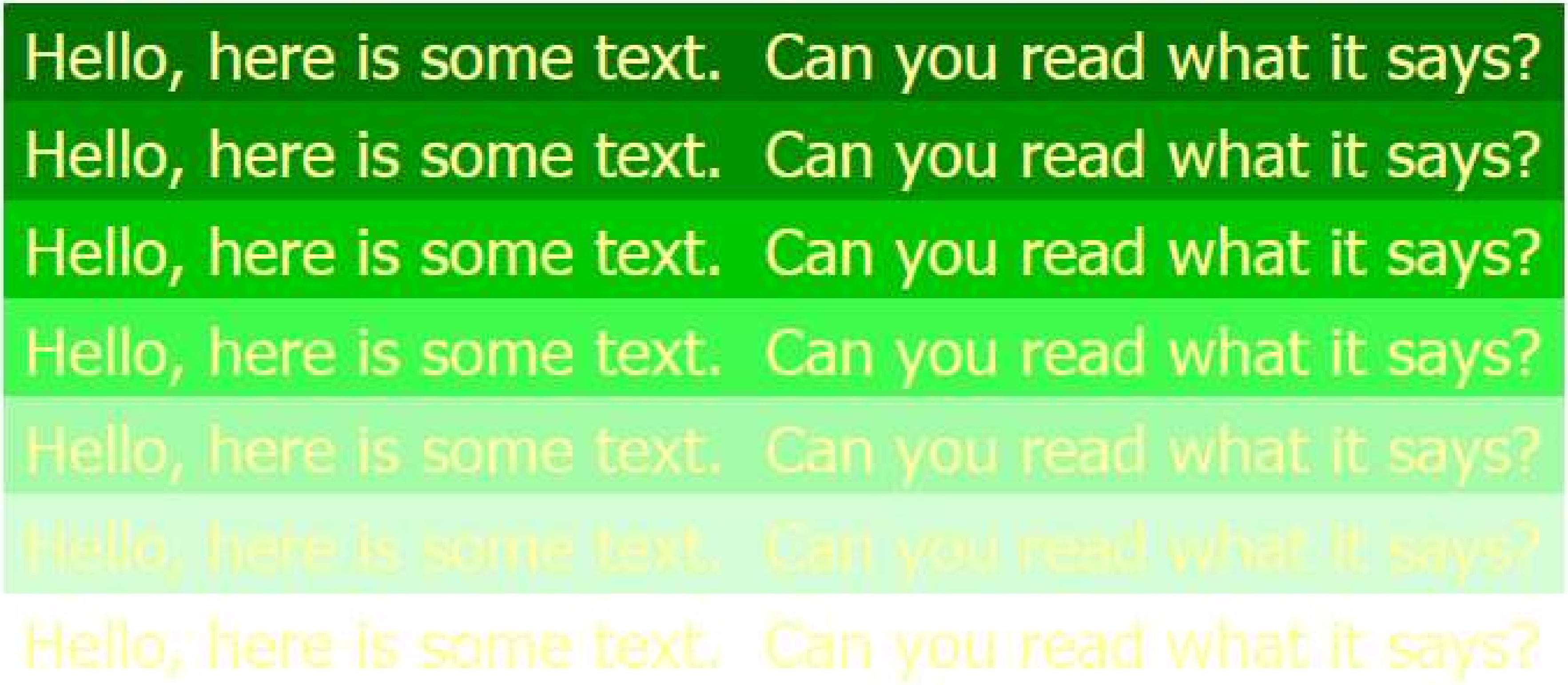
- “The often scant benefits derived from coloring data indicate that even putting a good color in a good place is a complex matter. Indeed, so difficult and subtle that avoiding catastrophe becomes the first principle in bringing color to information: Above all, do no harm.”
    - *Tufte*

# Color should be used to communicate

Of what use is color in this bar chart?



# Color Contrast for Visibility



Hello, here is some text. Can you read what it says?

Hello, here is some text. Can you read what it says?

Hello, here is some text. Can you read what it says?

Hello, here is some text. Can you read what it says?

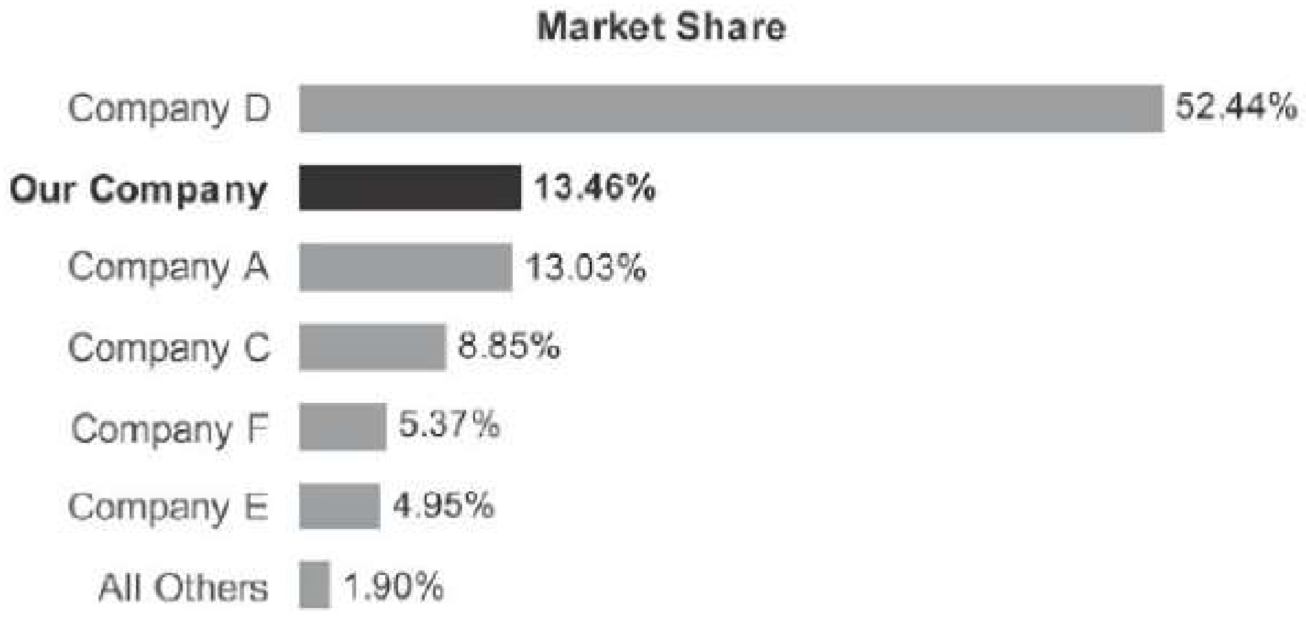
Hello, here is some text. Can you read what it says?

Hello, here is some text. Can you read what it says?

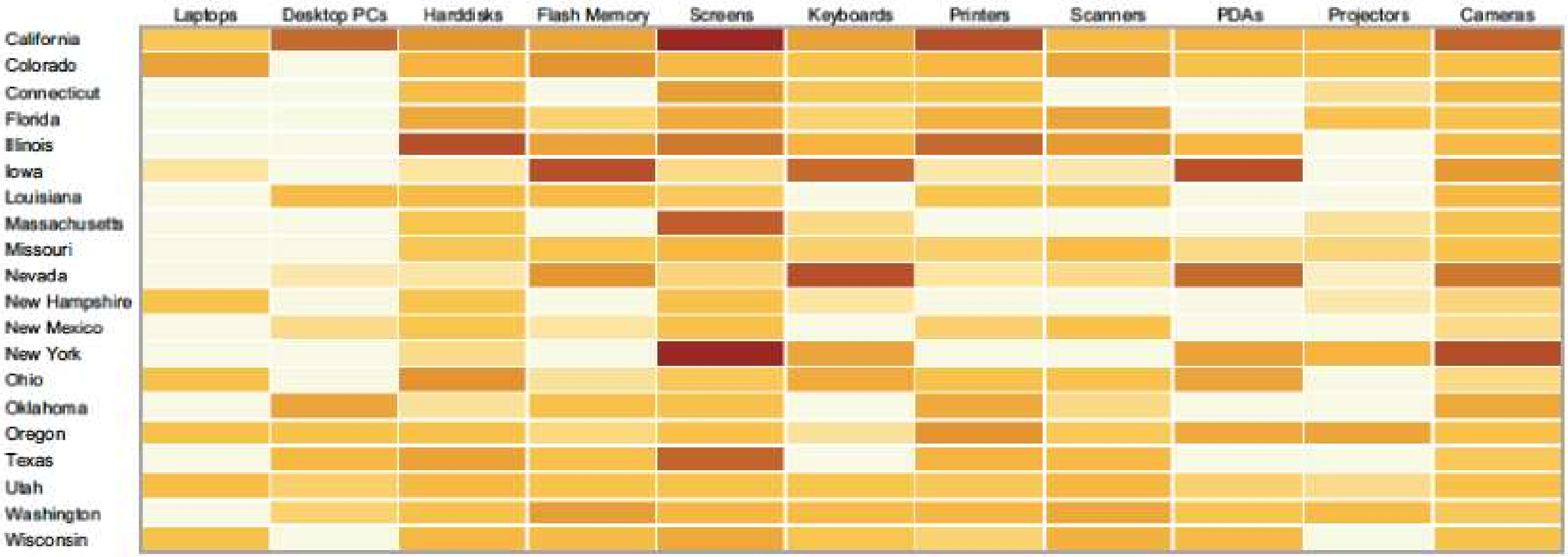
Hello, here is some text. Can you read what it says?

# Use Soft Colors

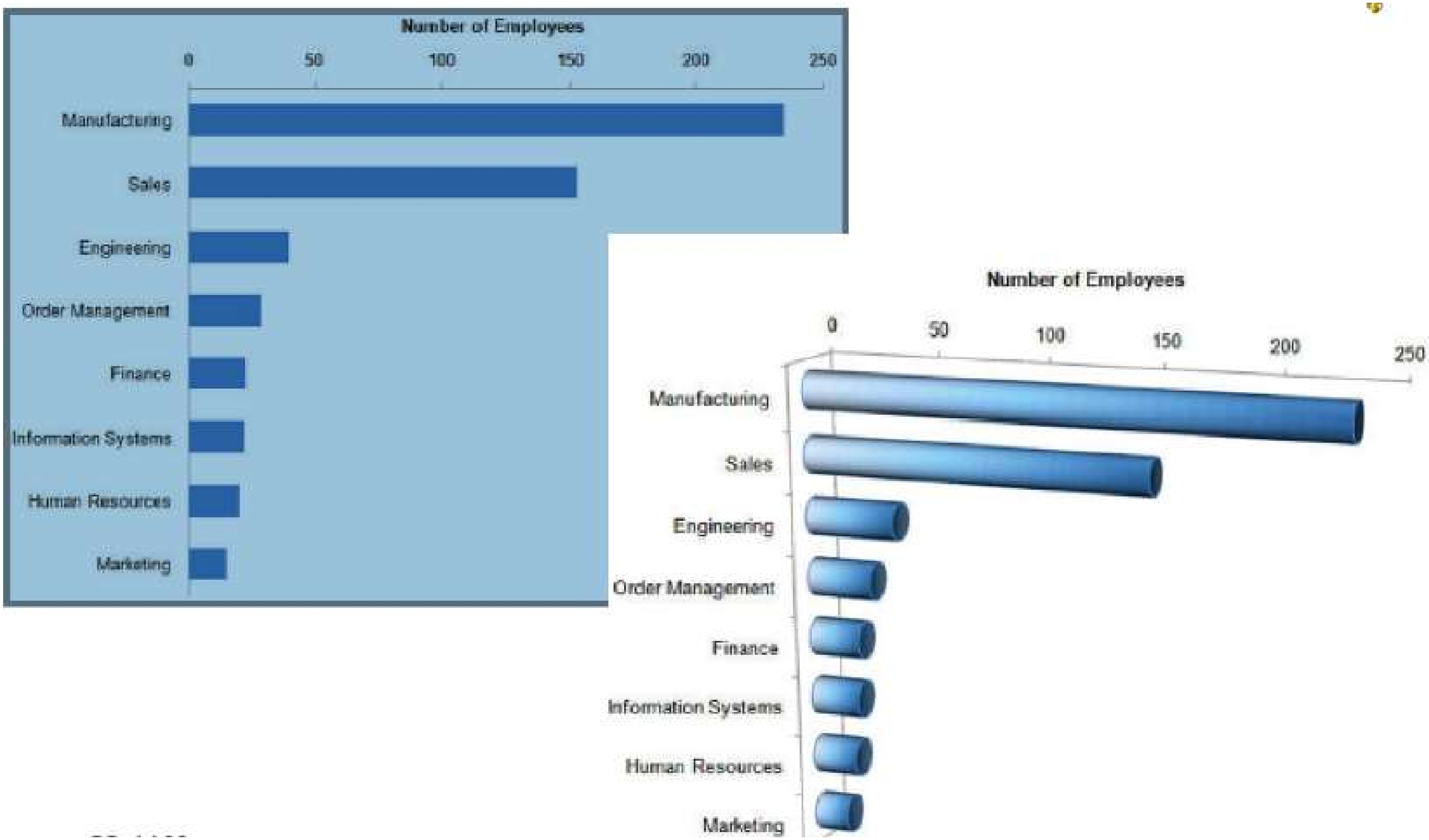
- Use mostly soft colors, with strong contrasting colors for emphasis
- Different shades of gray are colors too!



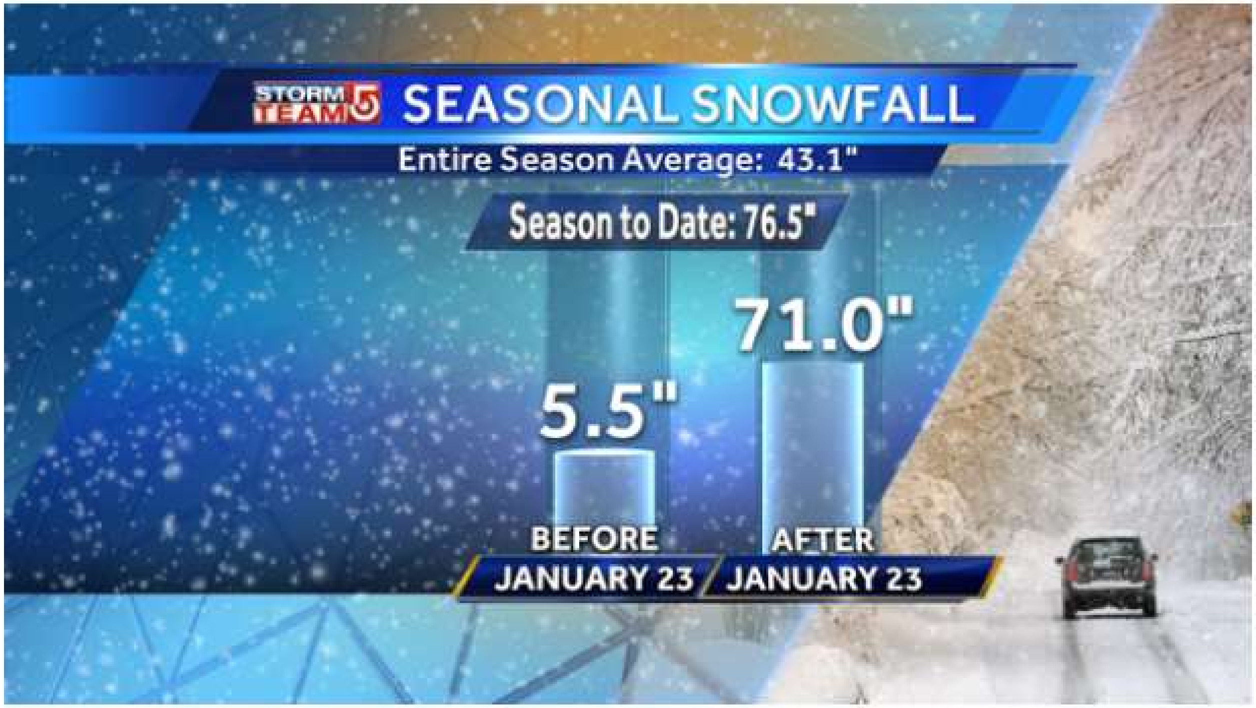
# Ordered Hues are meaningful



# Do not use fancy visual effects



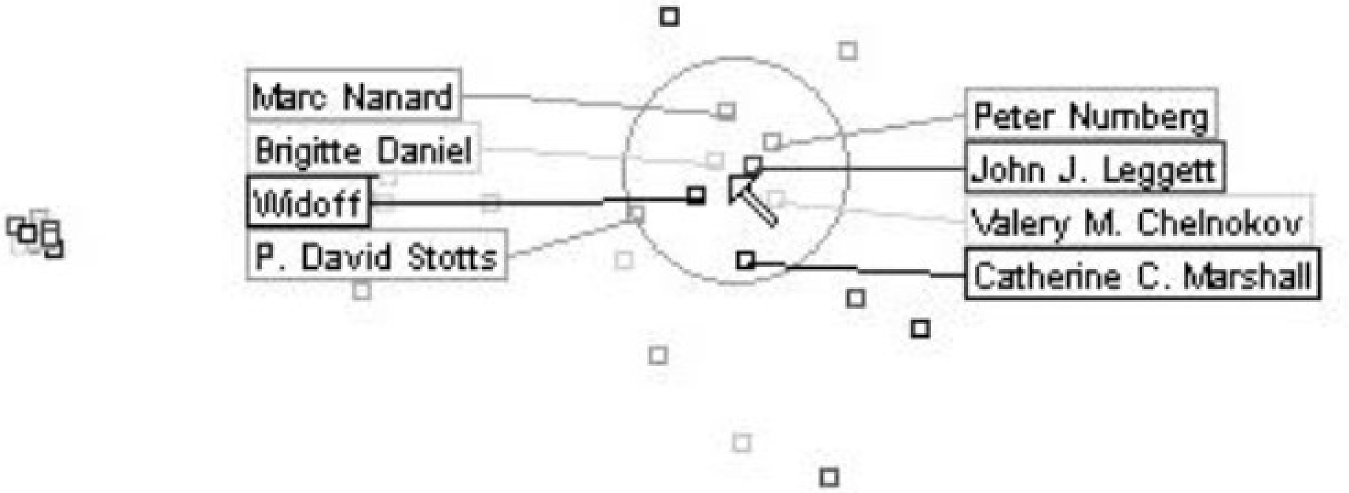
# Good bar chart?



Rule: Use channel proportional to data!

# Labeling

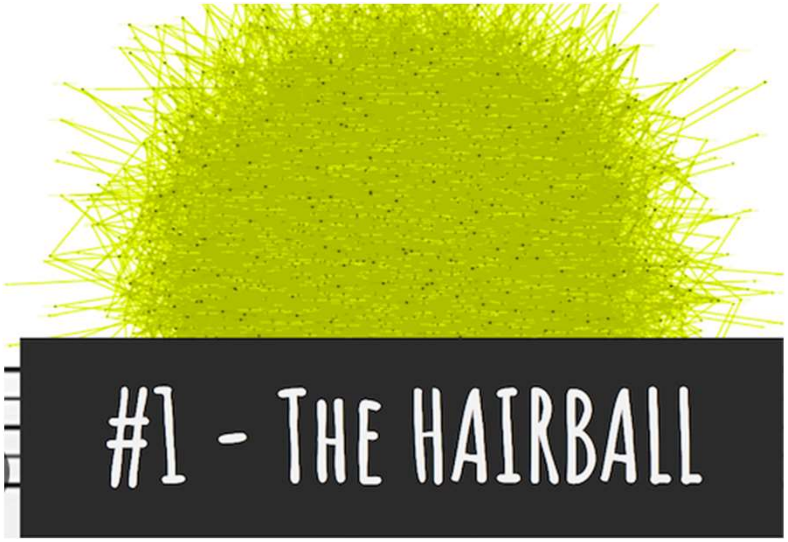
- Labeling is difficult to do when so many entities exist
- Exentric Labeling: Does not appear until user hovers over data points

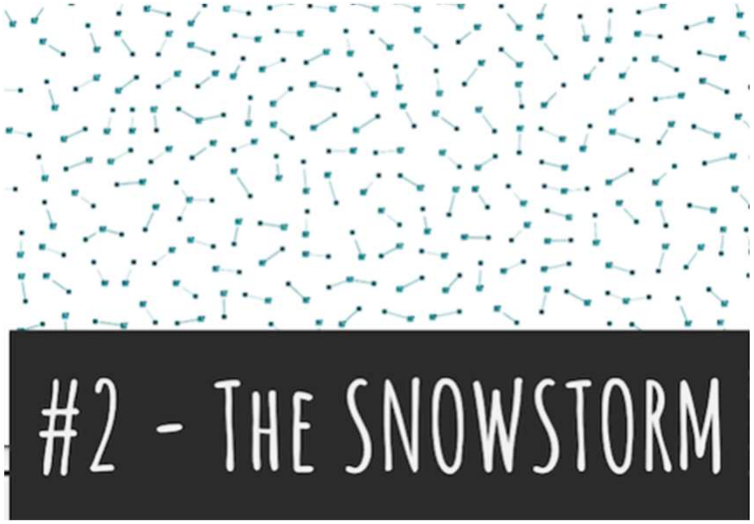


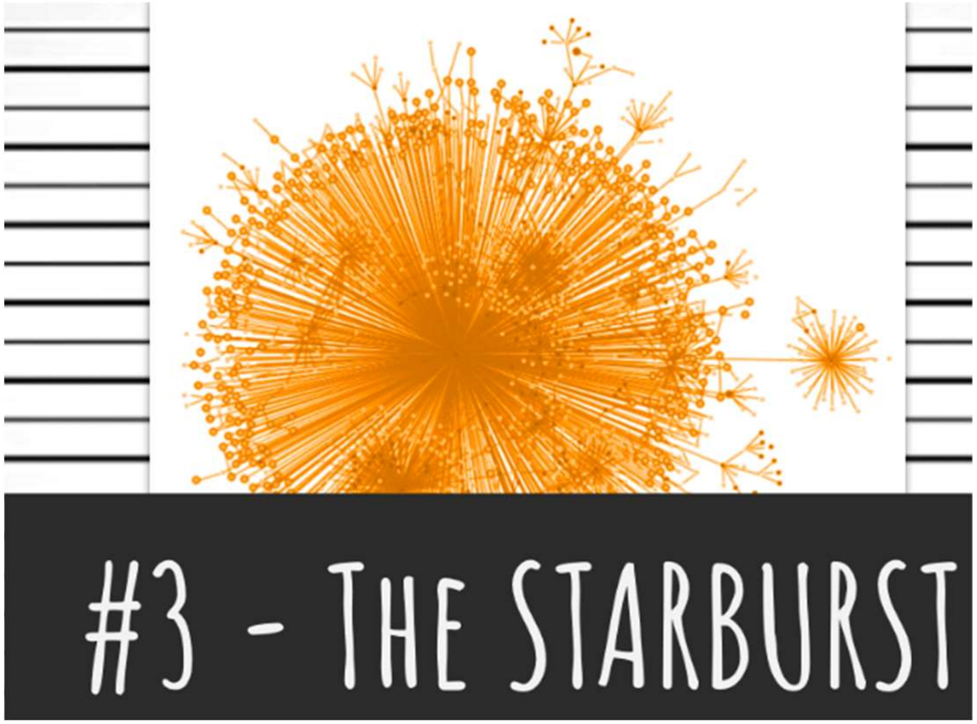
# Avoid  Bad Visualizations - Graph

1. **The hairball** – showing connections that are so dense, they can’t be usefully visualized.
2. **The snowstorm** – packed with many small, separate components where nothing stands out.
3. **The starburst** – where almost every connection is between a single central node and every other node.

# Avoid  Bad Visualizations - Graph







# Books on Design Principles

- Tufte, The Visual Display of Quantitative Information (Best 100 books of 20th century – Amazon)
- Tufte, Envisioning Information
- Tufte, Visual Explanations
- Few, Now you see it