SC4024/CZ4124
Data Visualization

Assistant Professor WANG Yong

yong-wang@ntu.edu.sg

CCDS, Nanyang Technological University

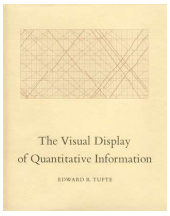
1

Chapter 8.1
Design Graph to Enlighten: Principles and Best Practices

2

Outline

- Visualization Principles by Prof. Edward Tufte
  - Graphical Integrity
  - Data-ink Ratio Maximization
- Other Best Practices



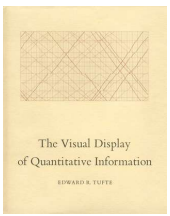
Prof. Edward Tufte from Yale University

3

3

Outline

- Visualization Principles by Prof. Edward Tufte
  - **Graphical Integrity**
  - Data-ink Ratio Maximization
- Other Best Practices



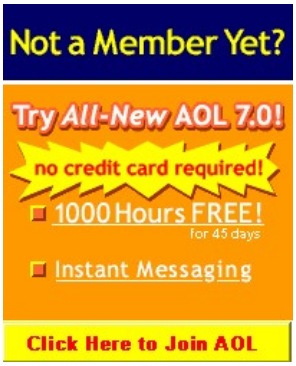
Prof. Edward Tufte from Yale University

4

4

Examples of Misleading Visualizations

- Big discount or benefits?

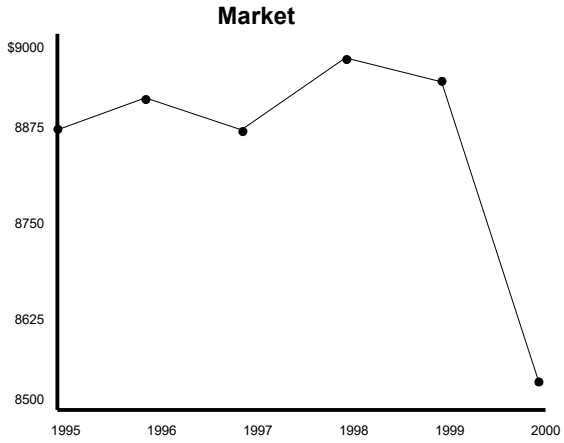


5

5

Examples of Misleading Visualizations

- Is it a market crash?

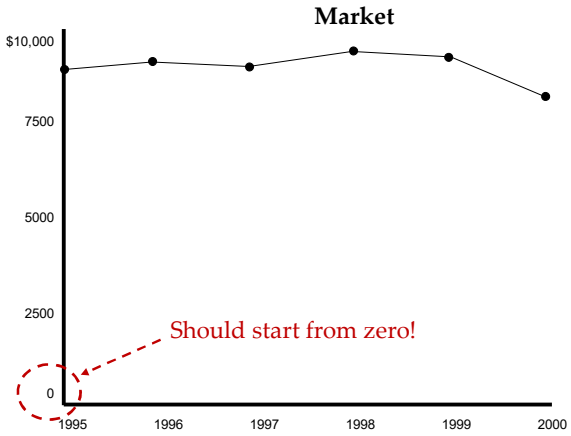


6

6

Graphical Integrity

- **Principle: Graphics must not quote data out of context**
  - Showing the entire scale



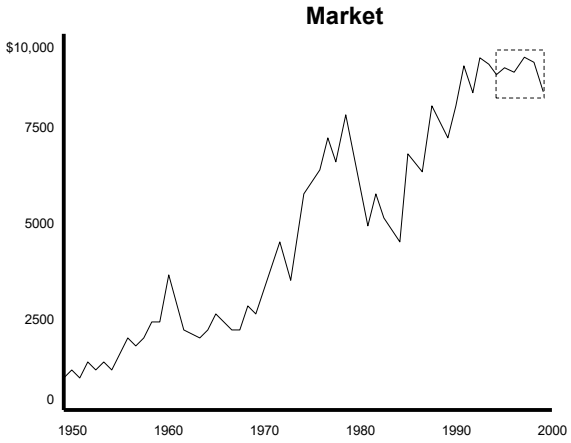
If data visualization does not start at zero, the viewer can be tricked into seeing false information.

7

7

Graphical Integrity

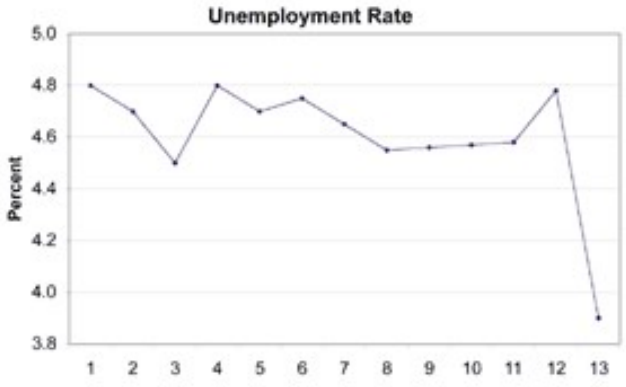
- **Principle: Graphics must not quote data out of context**
  - Showing the entire scale



8

8

Examples of Misleading Visualizations

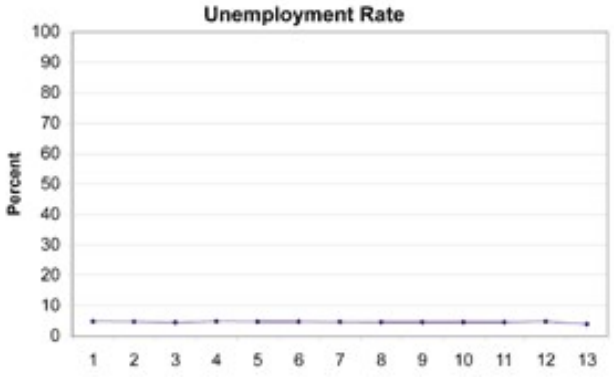


Unemployment Rate: A misleading visualization

9

9

Examples of Misleading Visualizations



Unemployment Rate: A visualization telling the truth

10

10

Graphical Integrity

- **Principle: Visual attribute value should be directly proportional to data attribute value**

  - Visual attributes: the visual encoding channels used to encode the underlying data, e.g., bar height, color density, circle radius, etc.

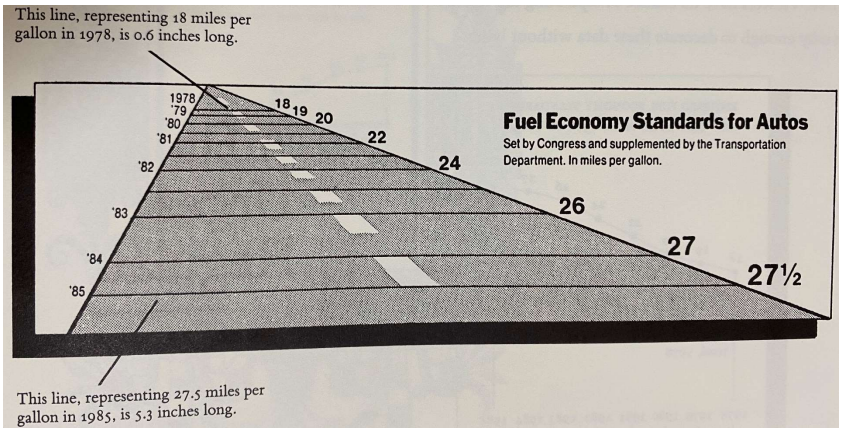
- Lie factor = $\frac{size\ of\ effect\ shown\ in\ graphic}{sizes\ of\ effect\ in\ data}$
- Truth = 1.0

11

11

Fuel Economy Standards for Autos

- Lie factor = ?



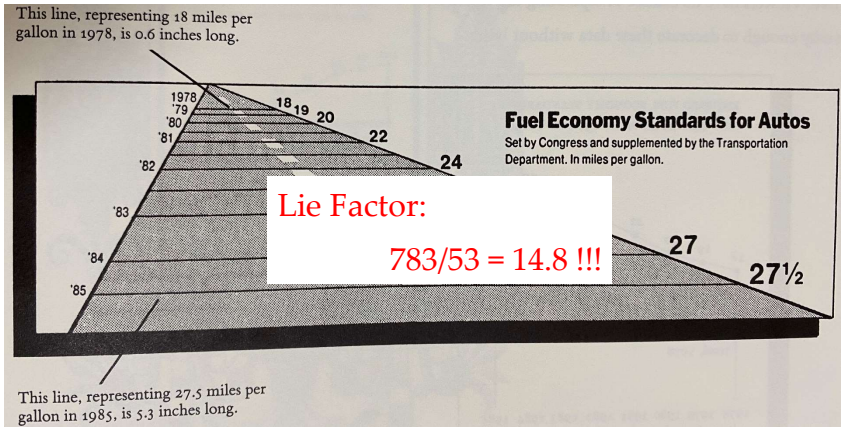
Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

12

12

Fuel Economy Standards for Autos

- Size of effect in data:
  (27.5 – 18.0)/18.0 = 53%
- Size of effect shown in graphic:
  (5.3 – 0.6)/0.6 = 783%



Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

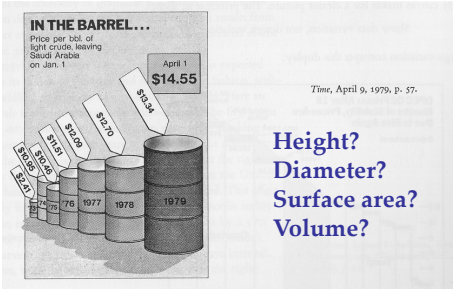
13

13

Graphical Integrity

- **Principle: The number of information-carrying dimensions (i.e., visual encoding channels) should not exceed the number of dimensions in the data.**

For example, avoid using two (or three) varying dimensions to show one-dimensional data



The visualization uses the height and width simultaneously of the bottle to represent the crude oil price, which is problematic:

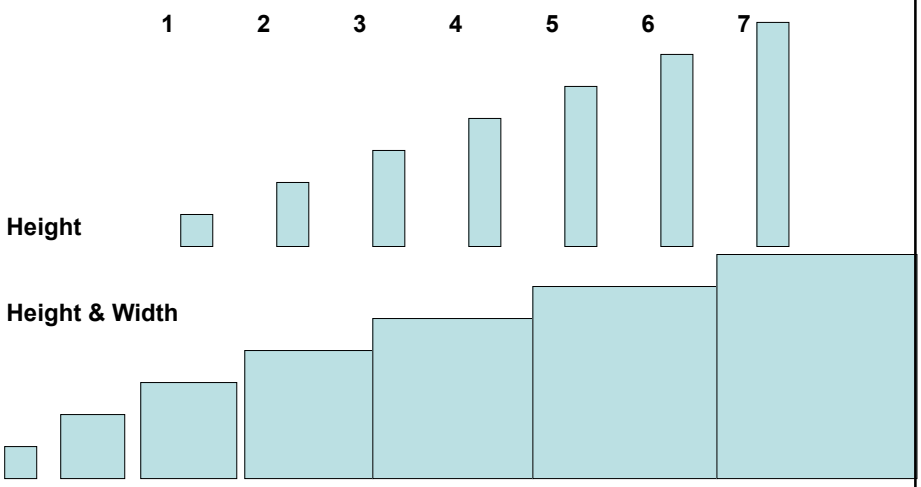
- Ambiguity in perception: Without an explicit indication, it is difficult for viewers to know which visual channel is used to encode the data.
- Space inefficiency: it results in the usage of a larger space.

Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

14

14

Height & Width Encoding



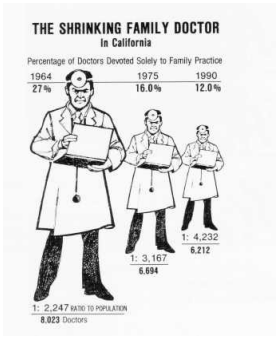
15

15

Graphical Integrity

- **Principle: The number of information-carrying dimensions (i.e., visual encoding channels) should not exceed the number of dimensions in the data.**

For example, avoid using two (or three) varying dimensions to show one-dimensional data



Questions:

- By reading the visualization itself, are you able to accurately identify which visual channel is used to encode the underlying data?
- How should we fix the issue?

Use only one visual encoding channel (e.g., only width or height)!

Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

16

16

Rankings of Visual Encoding Effectiveness

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Value)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area	Texture	Containment
Volume	Connection	Density (Value)
Density (Value)	Containment	Color Sat
Color Sat Color Hue Texture	Length	Shape
Connection	Angle	Length
Containment	Slope	Angle
Shape	Area (Size)	Slope
	Volume	Area
	Shape	Volume

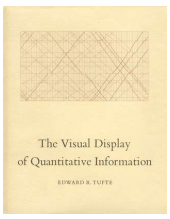
Mackinlay, Jock. "Automating the design of graphical presentations of relational information." *Acm Transactions On Graphics (Tog)* 5.2 (1986): 110-141.

17

17

Outline

- Visualization Principles by Prof. Edward Tufte
  - Graphical Integrity
  - Data-ink Ratio Maximization



Prof. Edward Tufte from Yale University

- Other Best Practices

18

18

Data-ink Ratio Maximization

- Data-ink is the non-erasable ink used for the presentation of data
- Reduce the non-data-ink (a.k.a., chartjunk)
  - Remove unnecessary non-data-ink
  - Emphasize data-ink
- Enhance the data-ink

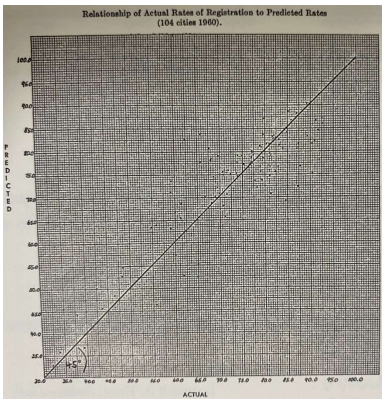
$$\frac{\text{Data Ink}}{\text{Total Ink}} = \text{Data Ink Ratio}$$

19

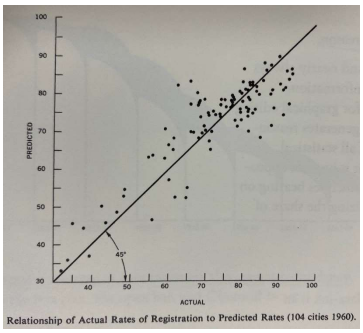
19

Data-ink Ratio Maximization

- For the following two visualizations that shows exactly the same data, which one is more clear for you to understand the underlying data?



Low data-ink ratio



High data-ink ratio

Tufte, Edward R. *The visual display of quantitative information*. Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

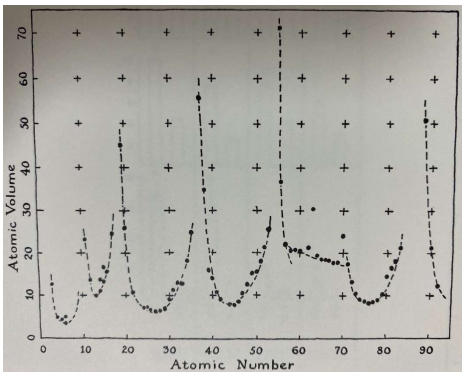
20

20

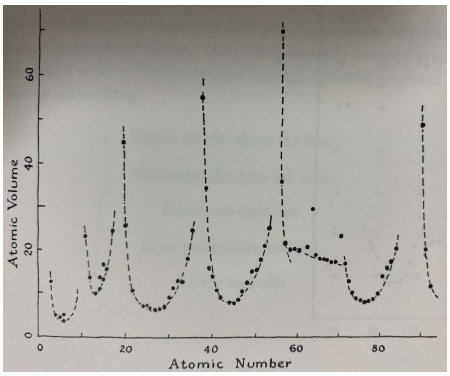
Data-ink Ratio Maximization

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- For the following two visualizations that shows exactly the same data, which one is more clear for you to understand the underlying data?



Low data-ink ratio



High data-ink ratio

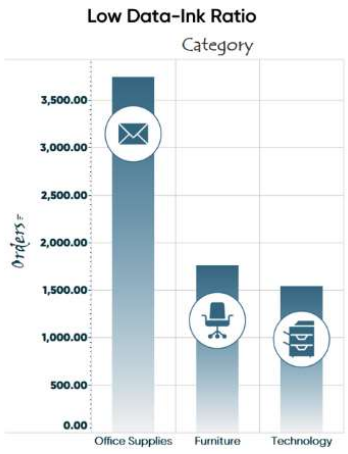
Tufte, Edward R. *The visual display of quantitative information.* Vol. 2. No. 9. Cheshire, CT: Graphics press, 1983.

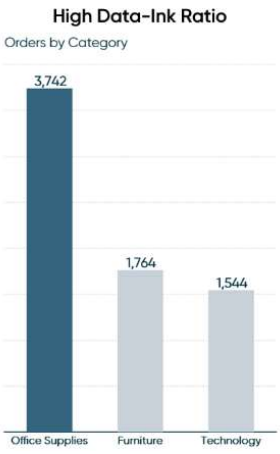
21

21

Data-ink Ratio Maximization

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE





It is not necessary to put the icons on the bar chart, as it does not bring more information.

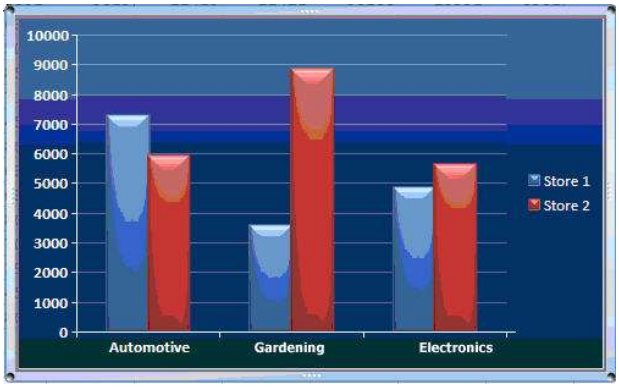
22

22

Data-ink Ratio Maximization: Avoid Chartjunk

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Avoid using unnecessary color shading for the bar



- The color shading here is useless and can actually be misleading for viewers!
- The background color is also meaningless.

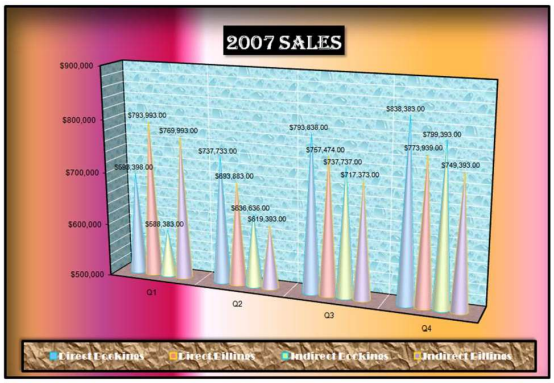
23

23

Data-ink Ratio Maximization: Avoid Chartjunk

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Avoid colorful or wallpaper background



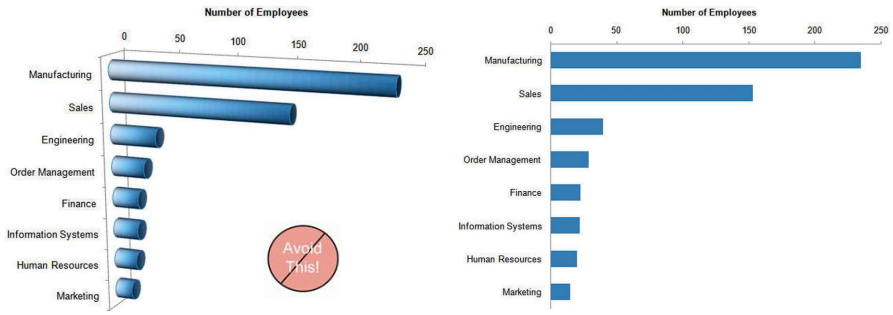
- The colorful background is also meaningless.

24

24

Data-ink Ratio Maximization: Avoid Chartjunk

- Avoid using 3D effects in graphics



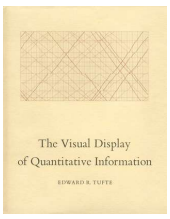
- The 3D effect makes it harder for viewers to perceive the underlying data accurately.

25

25

Outline

- Visualization Principles by Prof. Edward Tufte
  - Graphical Integrity
  - Data-ink Ratio Maximization



Prof. Edward Tufte from Yale University

- Other Best Practices

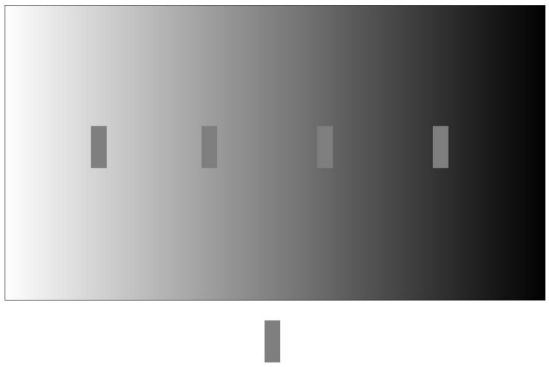
26

26

Practical Guides for Using Color in Charts

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Rule 1: If you want different objects of the same color in a graph to look the same, make sure that the background- the color that surrounds them – is consistent.



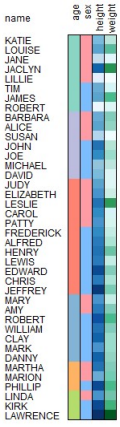
27

27

Practical Guides for Using Color in Charts

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Rule 2: If you want objects in a graph to be easily seen, use a background colour that constracts sufficiently with the object



	name	age	sex	height	weight	mean weight by age
1	KATIE	12	F	59	95	99.000
2	LOUISE	12	F	61	123	99.000
3	JANE	12	F	55	74	99.000
4	JACLYN	12	F	66	145	99.000
5	LILLIE	12	F	52	64	99.000
6	TIM	12	M	60	84	99.000
7	JAMES	12	M	61	128	99.000
8	ROBERT	12	M	51	79	99.000
9	BARBARA	13	F	60	112	94.714
10	ALICE	13	F	61	107	94.714
11	SUSAN	13	F	56	67	94.714
12	JOHN	13	M	65	98	94.714
13	JOE	13	M	63	105	94.714
14	MICHAEL	13	M	58	95	94.714
15	DAVID	13	M	59	79	94.714
16	JUDY	14	F	61	81	100.833
17	ELIZABETH	14	F	62	91	100.833
18	LESLIE	14	F	65	142	100.833
19	CAROL	14	F	63	84	100.833
20	PATTY	14	F	62	85	100.833
21	FREDERICK	14	M	63	93	100.833
22	ALFRED	14	M	64	99	100.833
23	HENRY	14	M	65	119	100.833
24	LEWIS	14	M	64	92	100.833
25	EDWARD	14	M	69	112	100.833
26	CHRIS	14	M	64	99	100.833
27	JEFFREY	14	M	69	113	100.833
28	MARY	15	F	62	92	108.286
29	AMY	15	F	64	112	108.286
30	ROBERT	15	M	67	128	108.286
31	WILLIAM	15	M	65	111	108.286

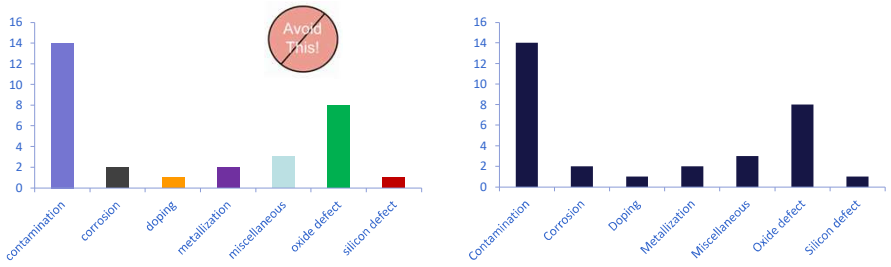
28

28

Practical Guides for Using Color in Charts

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Rule 3: Use different colors only when they correspond to different meanings in the data



It is not necessary to use different colors for different bars.

29

29

Practical Guides for Using Color in Charts

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Rule 4: Use soft, natural colours to display most information and bright and/or dark colours to highlight information that requires greater attention

The grids and labels are in soft gray, while the line charts are in bright yellow and blue.

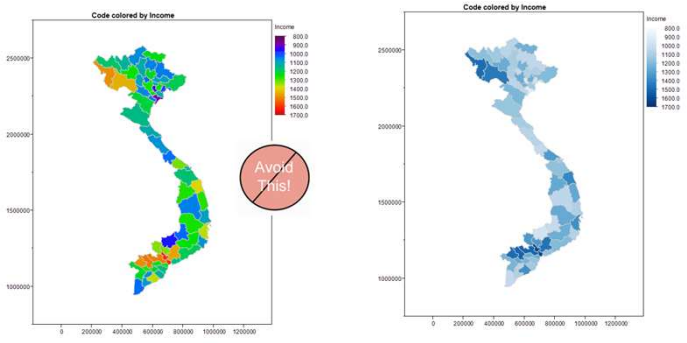
30

30

Practical Guides for Using Color in Charts

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Rule 5: When using colour to encode a sequential range of quantitative values, stick with a single hue and vary intensity from pale colours for low values to increasingly darker and brighter colours for high values, i.e., use **sequential colormap**



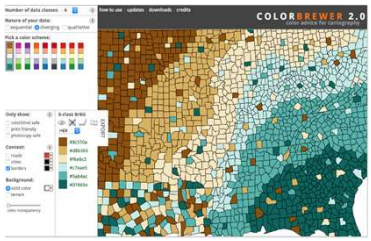
31

31

Colormap Design

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Colormap: a mapping between data values and colors
- Two types of colormaps
  - Categorical (aka., qualitative)
  - Ordered (aka., quantitative): sequential + diverging
- A very useful tool: Colorbrewer2 http://colorbrewer2.org/

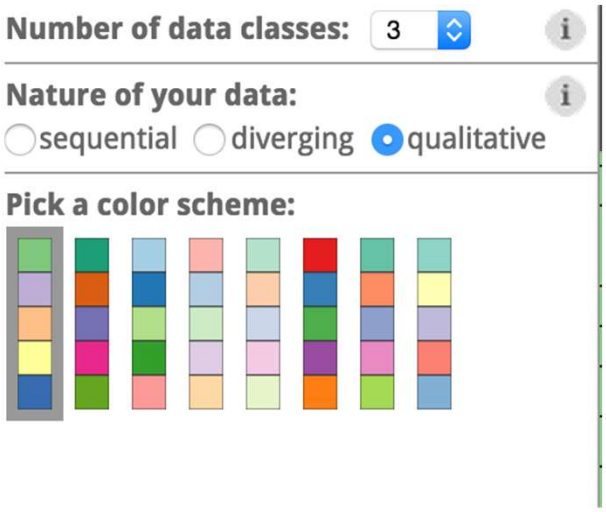


32

32

Categorical Colormaps

- Use color to encode categories and groupings



33

33

Rules for Categorical Colormaps

- The maximum number of colors which can be used is usually between 6 to 12
- The 12 recommended colors

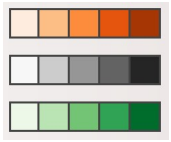
1. Red 2. Green 3. Yellow
4. Blue 5. Black 6. White
7. Pink 8. Cyan 9. Gray
10 Orange 11. Brown 12. Purple

34

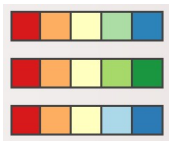
34

Ordered Colormaps

- **Sequential Colormap**
  - Ramp in luminance, possibly also hue.
  - Typically higher values map to darker colors.



- **Diverging Colormap**
  - Useful when data has a meaningful "midpoint".
  - Use neutral color (e.g., gray) for midpoint.
  - Use two saturated colors for endpoints.



- **Limit number of steps in color to 3-9**

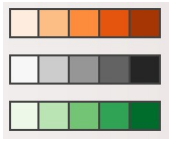
  Why?

35

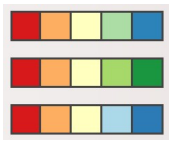
35

Ordered Colormaps

- **Sequential Colormap**
  - Ramp in luminance, possibly also hue.
  - Typically higher values map to darker colors.



- **Diverging Colormap**
  - Useful when data has a meaningful "midpoint".
  - Use neutral color (e.g., gray) for midpoint.
  - Use two saturated colors for endpoints.



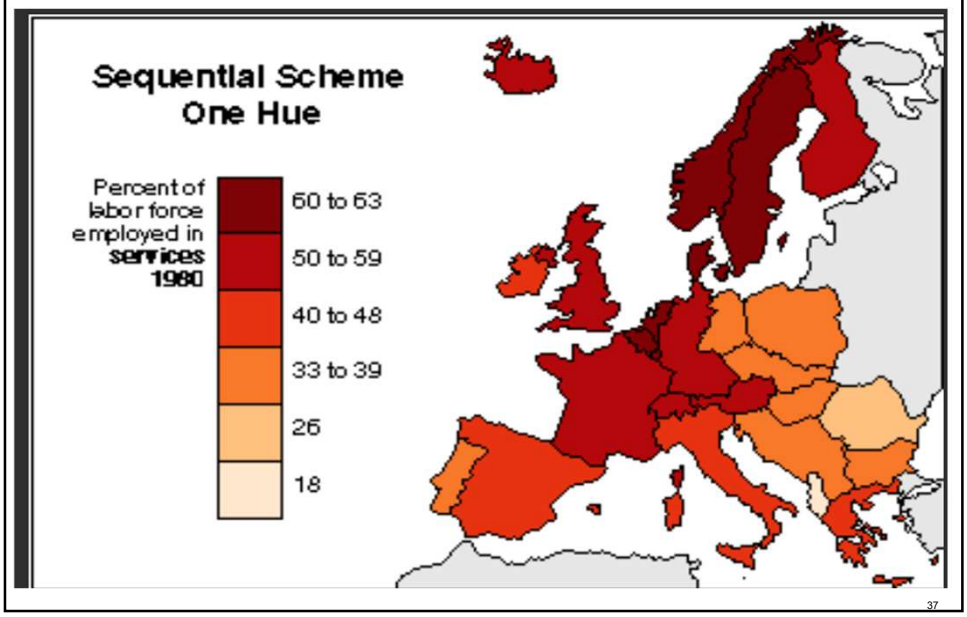
- **Limit number of steps in color to 3-9**

  Avoid simultaneous contrast; Need to hold mappings in memory

36

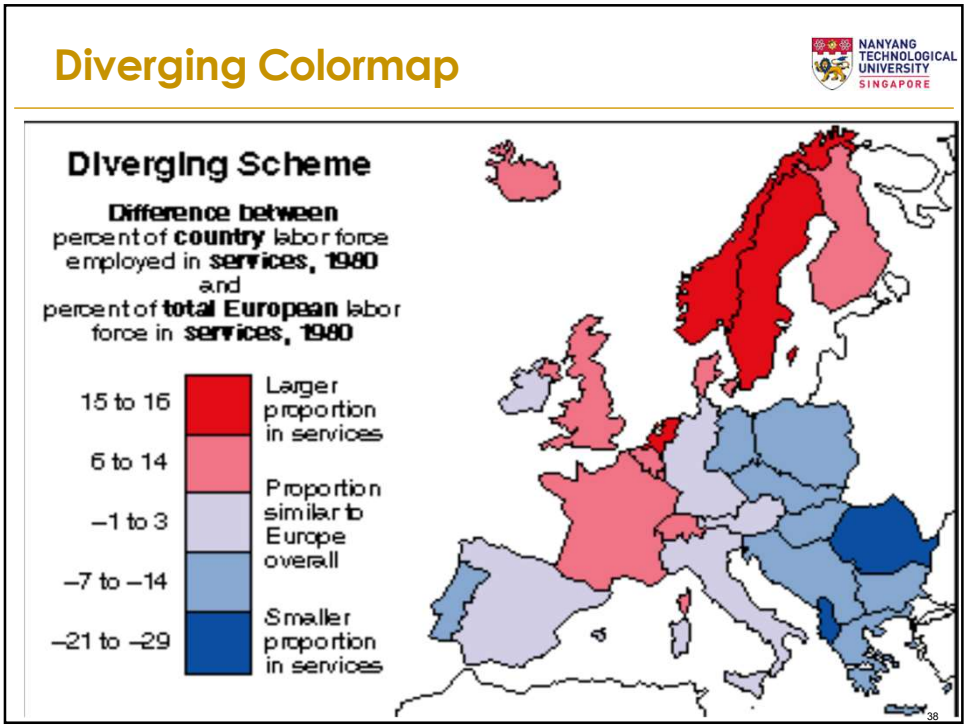
36

Sequential Colormap



37

Diverging Colormap



38

Practical Guides for Using Color in Charts

- Rule 6: Respect the **color blind and** be colorblind friendly.

39

39

Practical Guides for Using Color in Charts

- Rule 7: Pay attention to the meaning of different colors
  Example: Logo color

40

40

Practical Guides for Using Color in Charts

- Rule 7: Pay attention to the meaning of different colors

Example: Logo color

- Red = Adventure. Energy. (Coca Cola)
- Pink = Compassion. Nurturing. (Breast Cancer Research Foundation; Barbie)
- Orange = Cheerful. Warm. Optimistic. (Amazon)
- Yellow = Happy. Bright. Sunny. (McDonalds)
- Green = Health and Healing. Environment. Organic. (Whole Foods Market; Starbucks)
- Blue = Honesty. Trust. Dependability. (Facebook; Pfizer)
- …

41

41

Practical Guides for Using Color in Charts

- Rule 7: Pay attention to the meaning of different colors

Use colors that viewers are familiar with

- "red" for hot, dangers, cautions, & warnings

>> watch out for culture conflicts (e.g., red in Asian cultures)

>> watch out for differences of conventions: the meaning of red and green in different stock markets (China vs. USA)

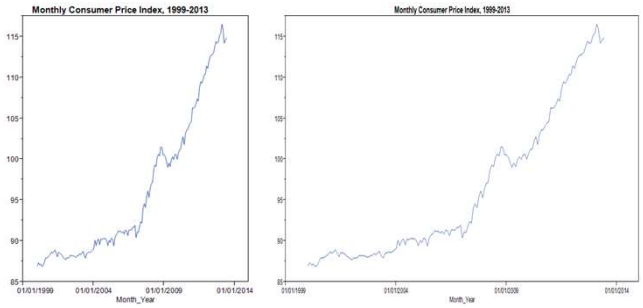
42

42

What is the Best Aspect-ration of a Two-variable Visualization?

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- Should not manipulate the aspect ratio to intentionally exaggerate or downplay the rate of change.
- The general consensus seems to be that an aspect ratio that keeps the line segments around 45 degrees going up (or -45 degrees going down) is the best for detecting trends.



Heer, Jeffrey, and Maneesh Agrawala. "Multi-scale banking to 45 degrees." *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006): 701-708.

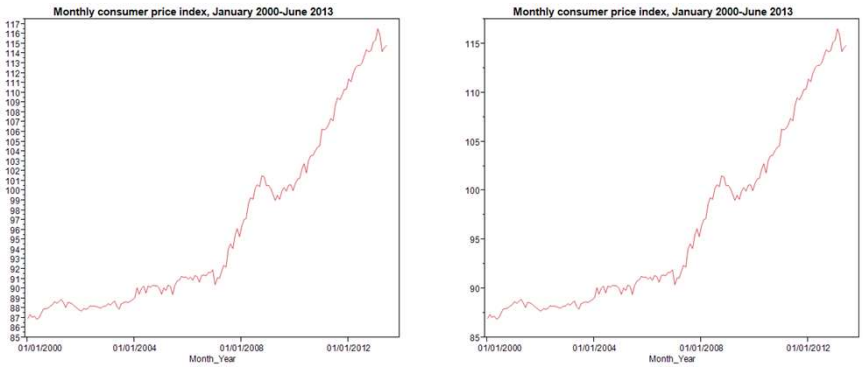
43

43

How many tick marks should you use?

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- There is no exact number that works best in all circumstances, and the size of the graph is a factor that must considered: the longer the scale line, the more tick marks it should contain.

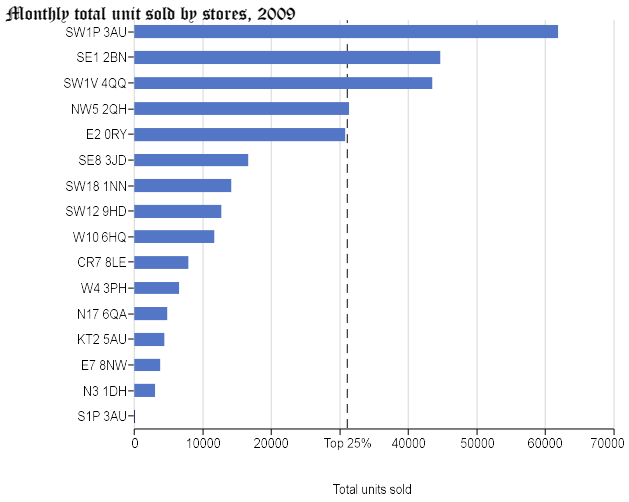


44

44

Graph Typography

- Avoid using artistic fonts

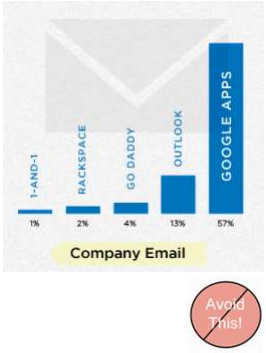


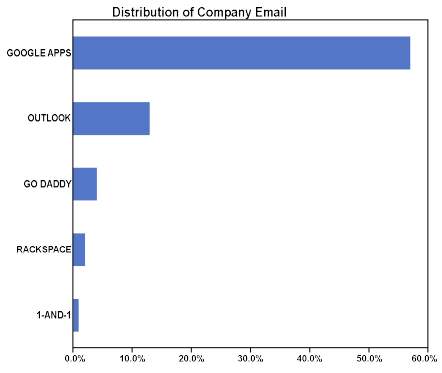
45

45

Graph Labeling

- Orientation of label should be reader friendly



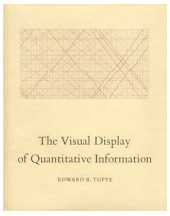


46

46

Recap

- Visualization Principles by Prof. Edward Tufte
  - Graphical Integrity
  - Data-ink Ratio Maximization



Prof. Edward Tufte from Yale University

- Other Best Practices

47

47

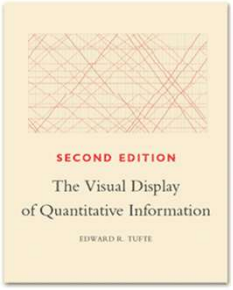
Questions?

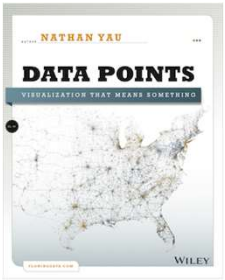
Thank You

48

Reference

- Tufte, Edward (2nd Edition) The Visual Display of Quantitative Information, Graphics Press LLC, Connecticut, USA
- Yau, Nathan (2013) Data Points: Visualization that means something, John Wiley & Sons, Inc.



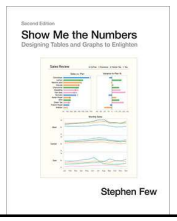


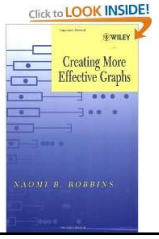
49

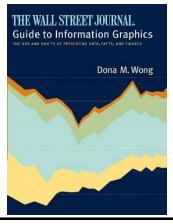
49

Reference

- Few, Stephen (2012) (2nd edition) **Show Me the Numbers: Designing Tables and Graphs to Enlighten**, Analytics Press, Oakland, USA
- Robbins, Naomi B. (2005) **Creating More Effective Graphs**, John Wiley & Sons, New Jersey, USA
- Wong, Dona M. (2010) **The Wall Street Journal Guide to Information Graphics**, W. W. Norton & Company, Inc. New York







50

50

Information Graphics and Data Visualization blogs

- Junk Charts (http://junkcharts.typepad.com/junk_charts/)
- Perceptual Edge (http://www.perceptualedge.com/blog/)
- Statistical Graphics and more (http://www.theusrus.de/blog/)
- EagerEyes (http://eagereyes.org/)
- Visualizing Economics (http://visualizingeconomics.com/)
- The Global Sociology Blog (http://globalsociology.com/)

51

51