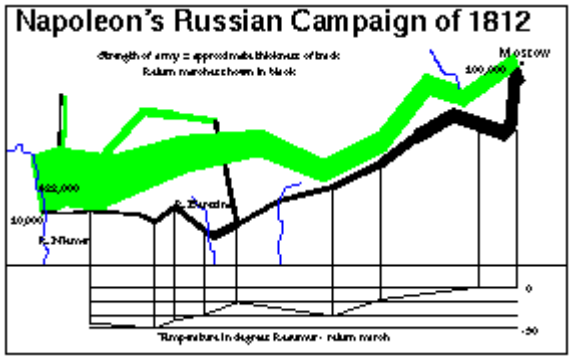
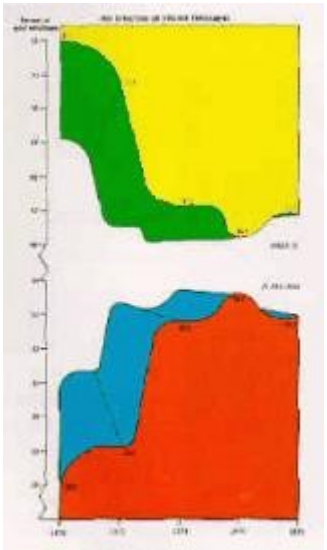
The Visual Display of Quantitative Information



vs.



Excellence!

Excellence in statistical graphics consists of complex ideas communicated with:

- clarity,
- precision,
- efficiency.

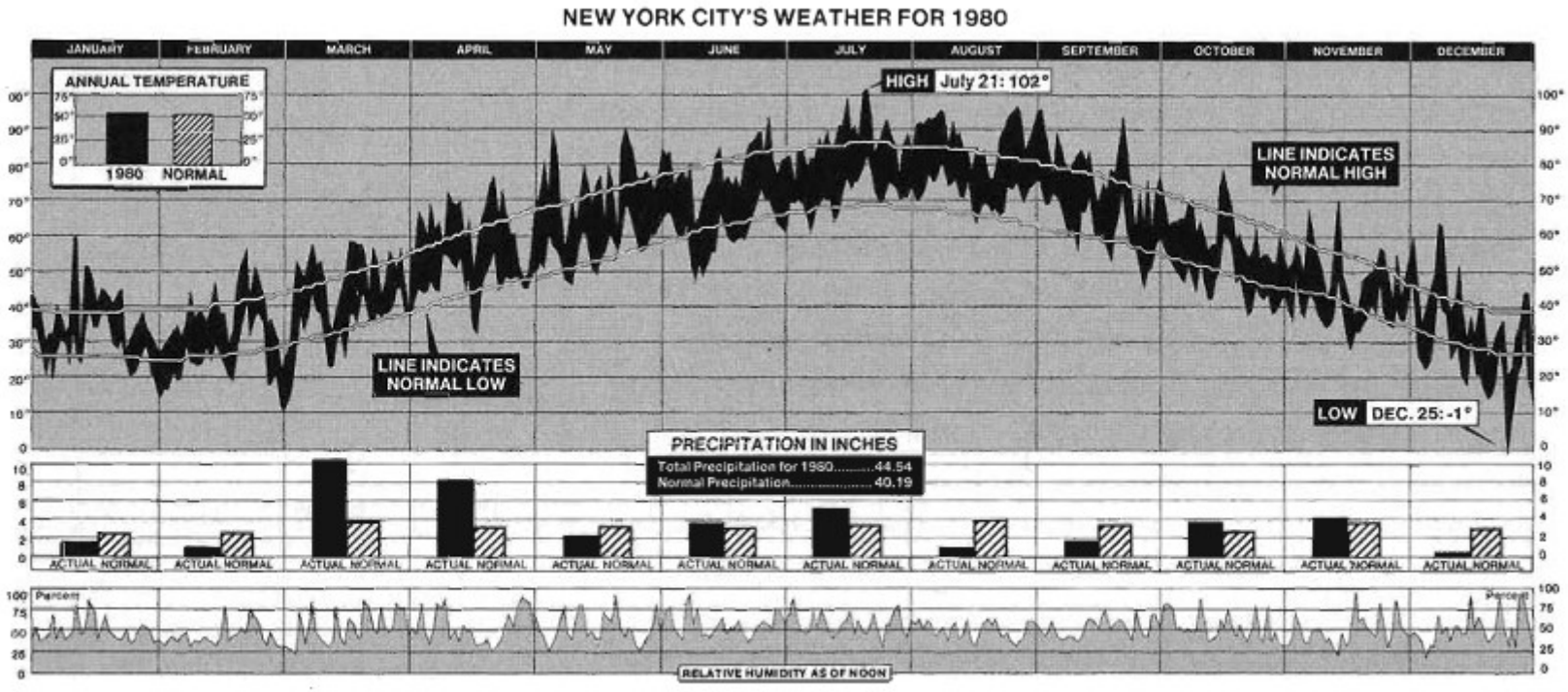
Graphical displays should:

- Show the data
- Make the viewer think about the substance (not the methods/design/technology)
- Avoid distorting the data
- Present many numbers in a small space
- Make large datasets coherent
- Encourage the eye to compare different pieces of data
- Reveal data structure at different levels
- Serve a clear purpose: description, exploration, tabulation, decoration

GOOD graphical displays:

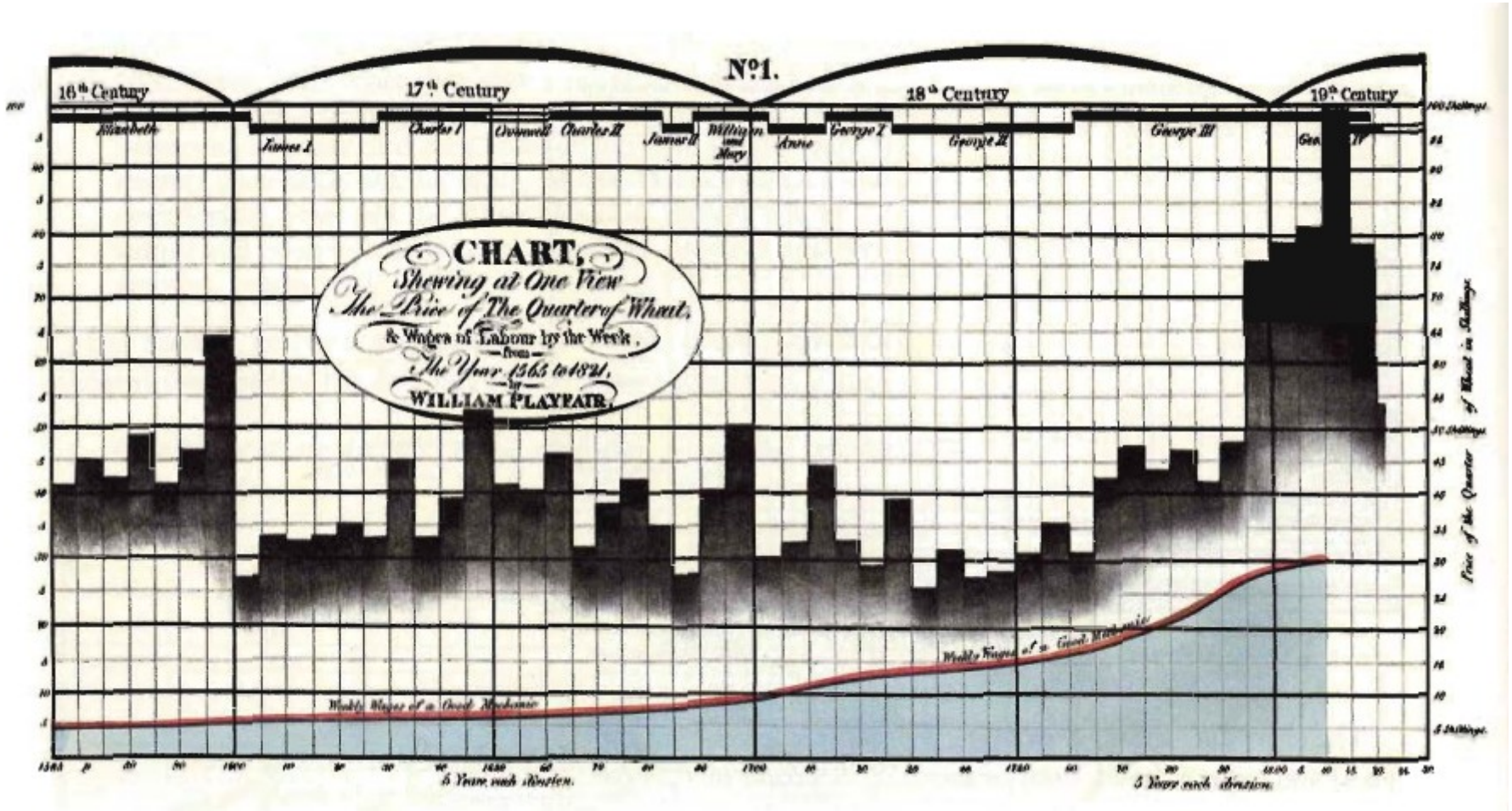
Deaths from cholera in central London, 1854

New York City weather

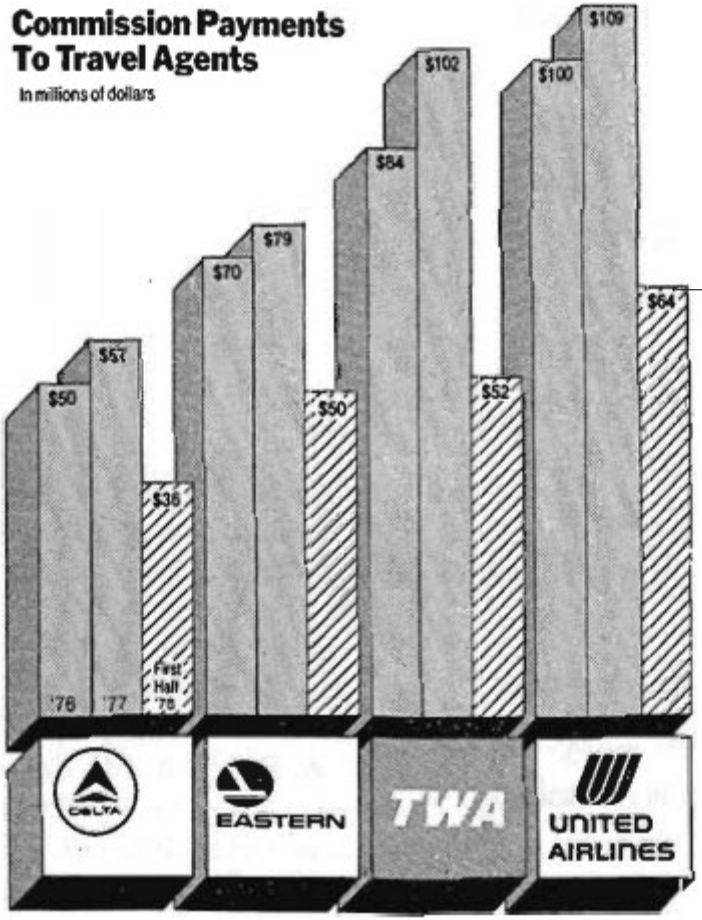


New York Times, January 11, 1981, p. 32.

Wheat prices and wages in the UK



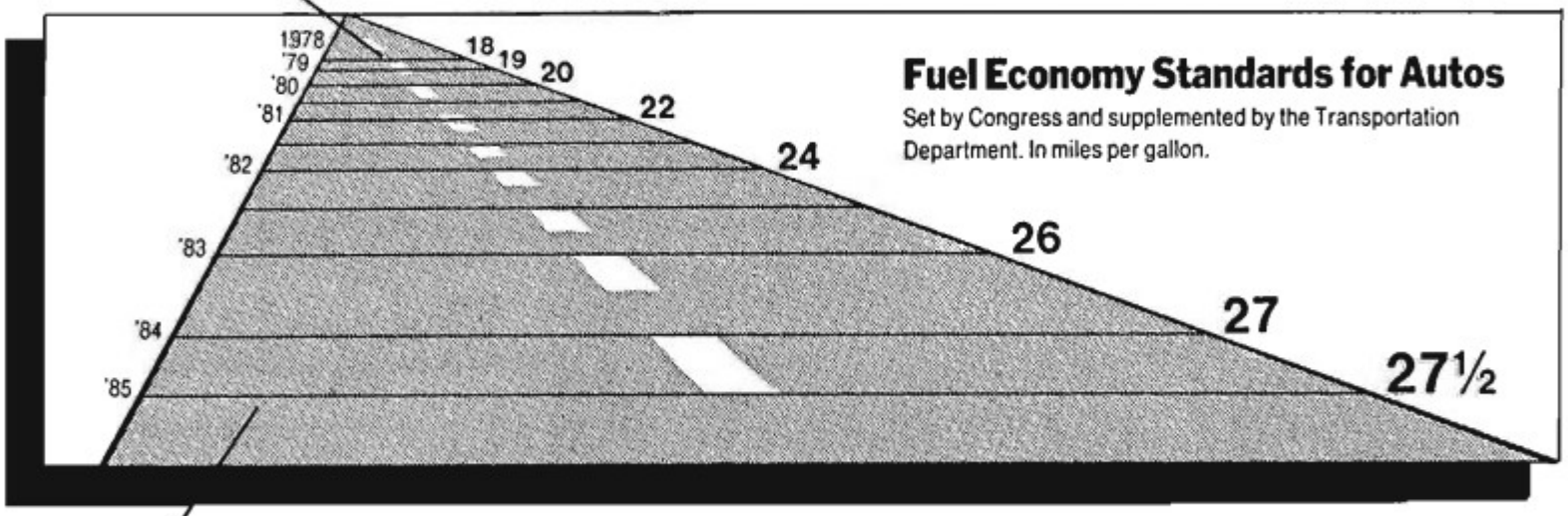
BAD graphical displays:



3D = no extra information

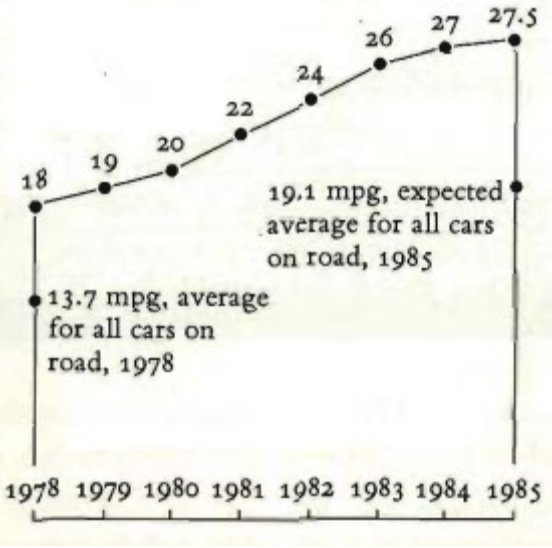
Latest results from only 6 months

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

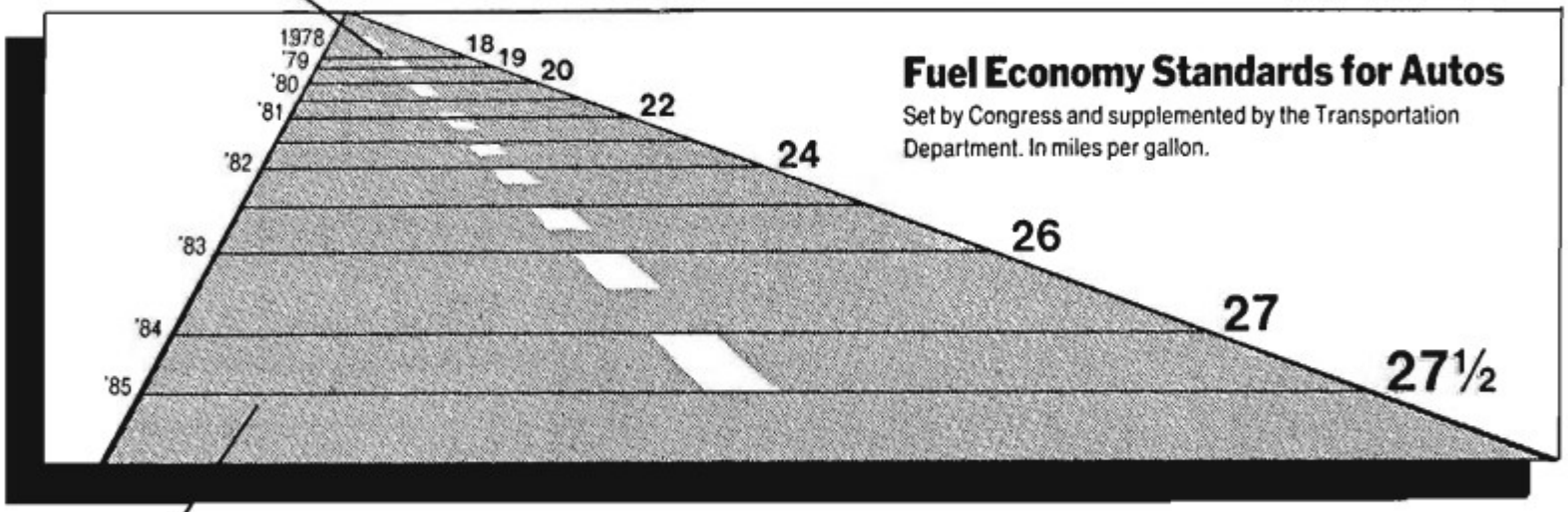


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985

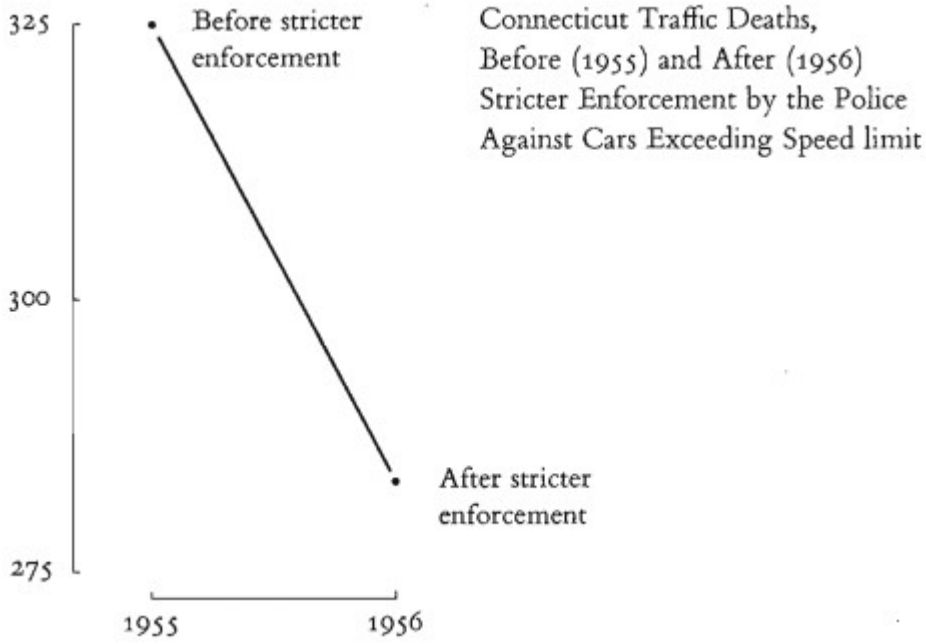


This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

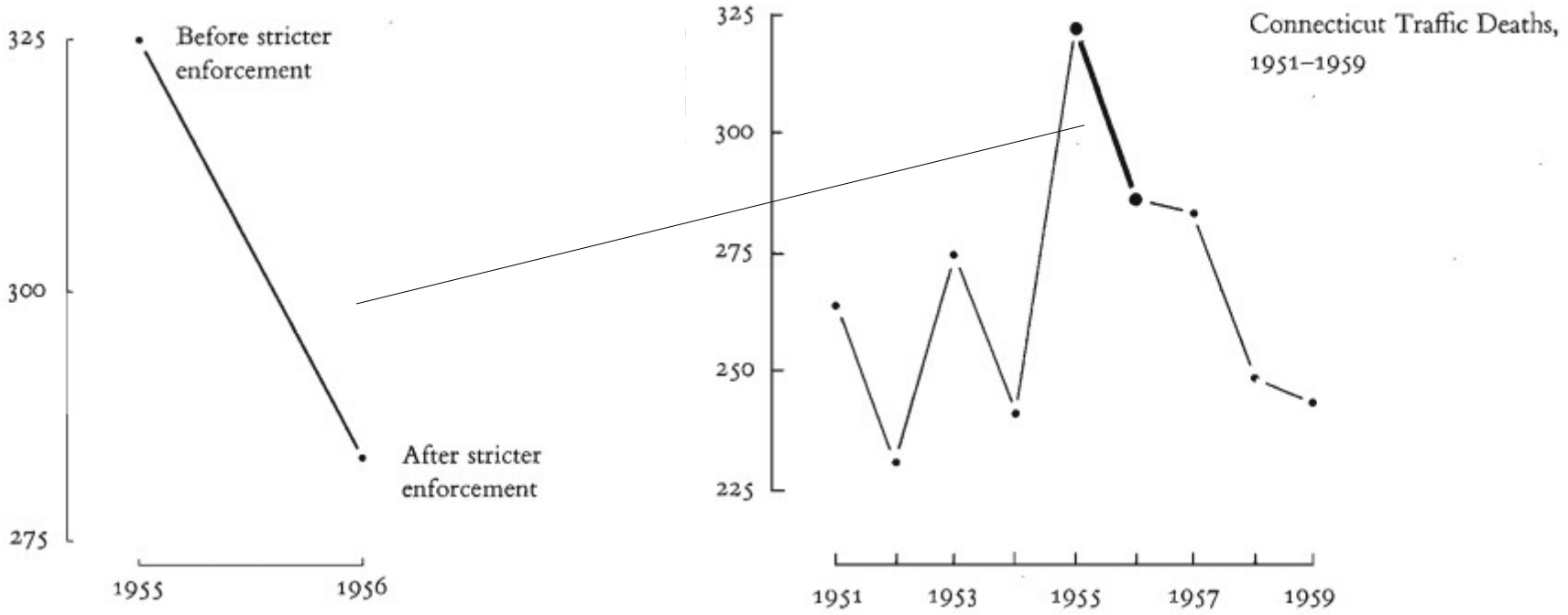


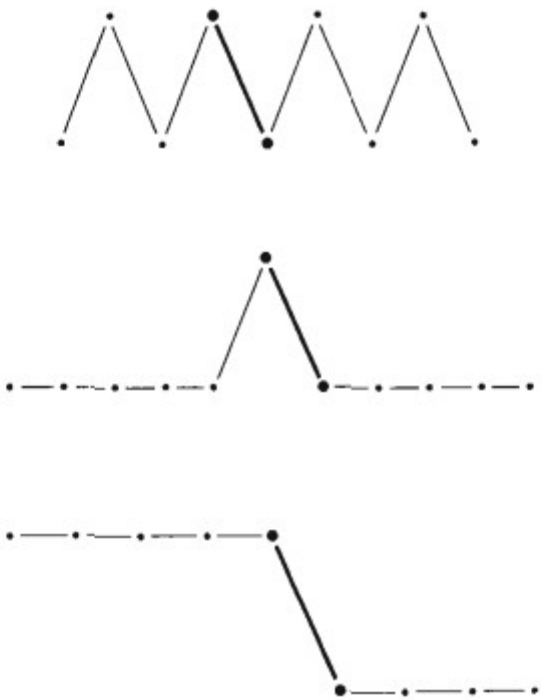
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Traffic deaths

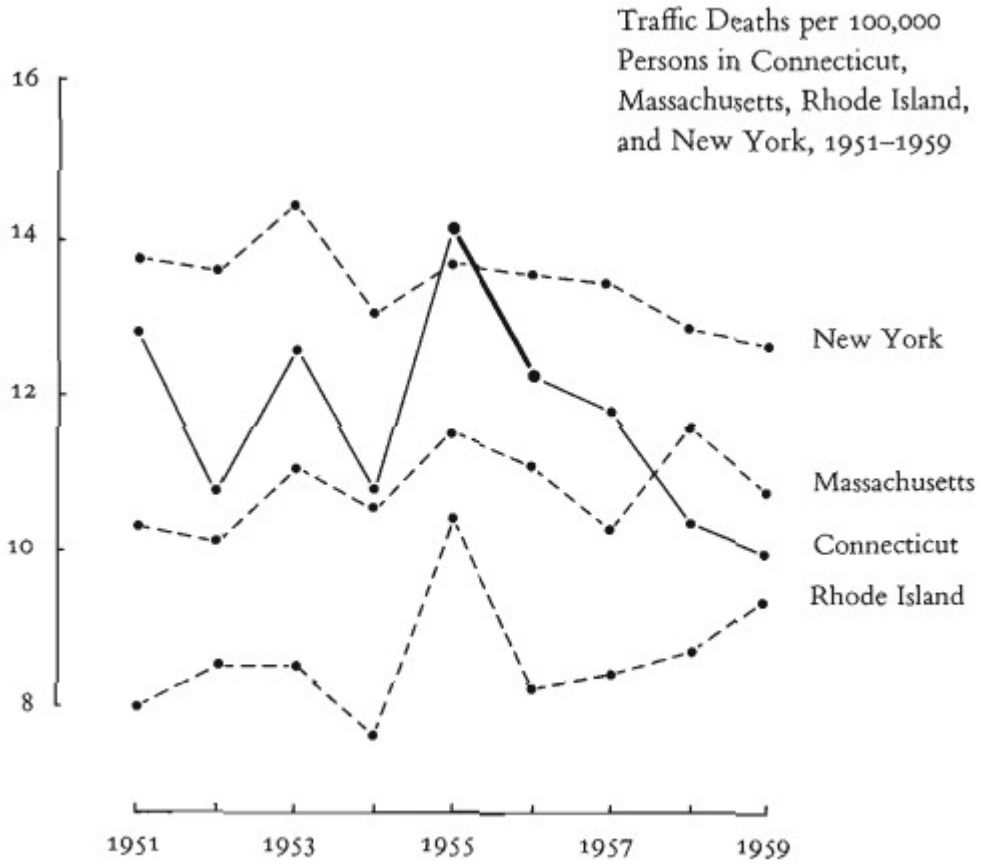


Traffic deaths





Traffic deaths



Theory of Data Graphics

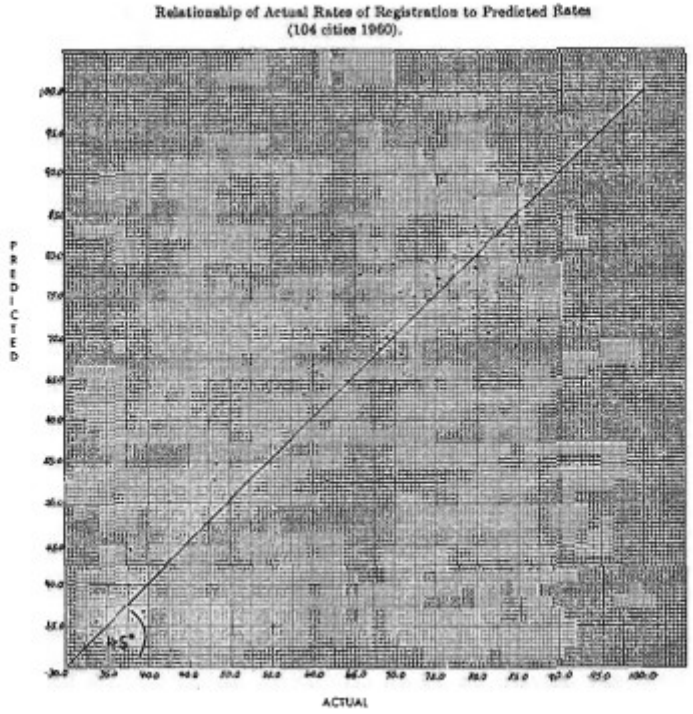
Data:Ink

$$\text{data: ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print graphic}}$$

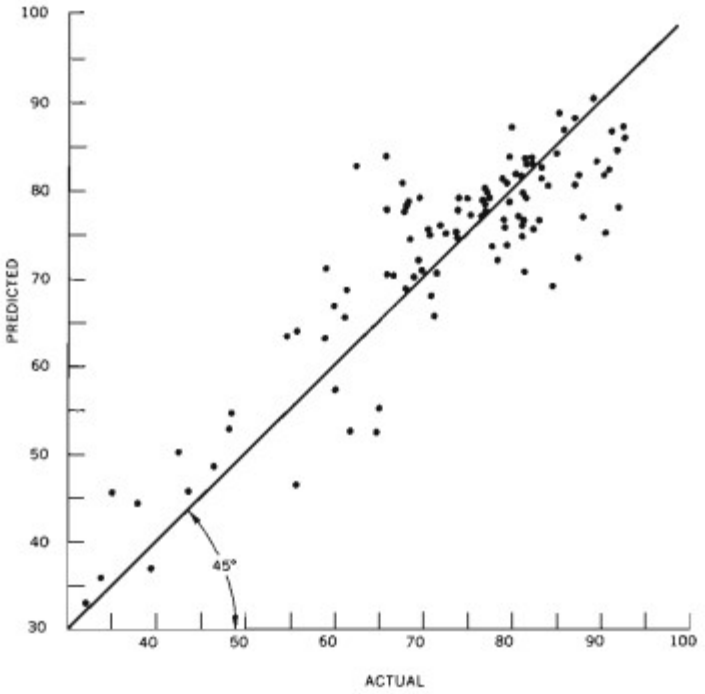
= proportion of graphic's ink devoted to the non-redundant display of data-information

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

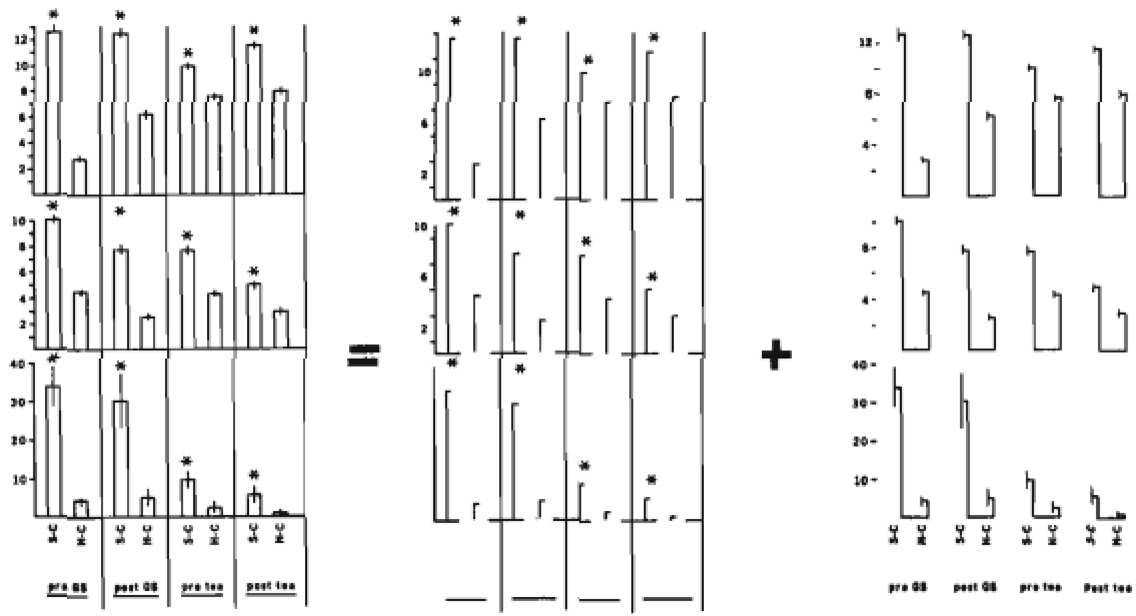
v. low!



data:ink 0.7

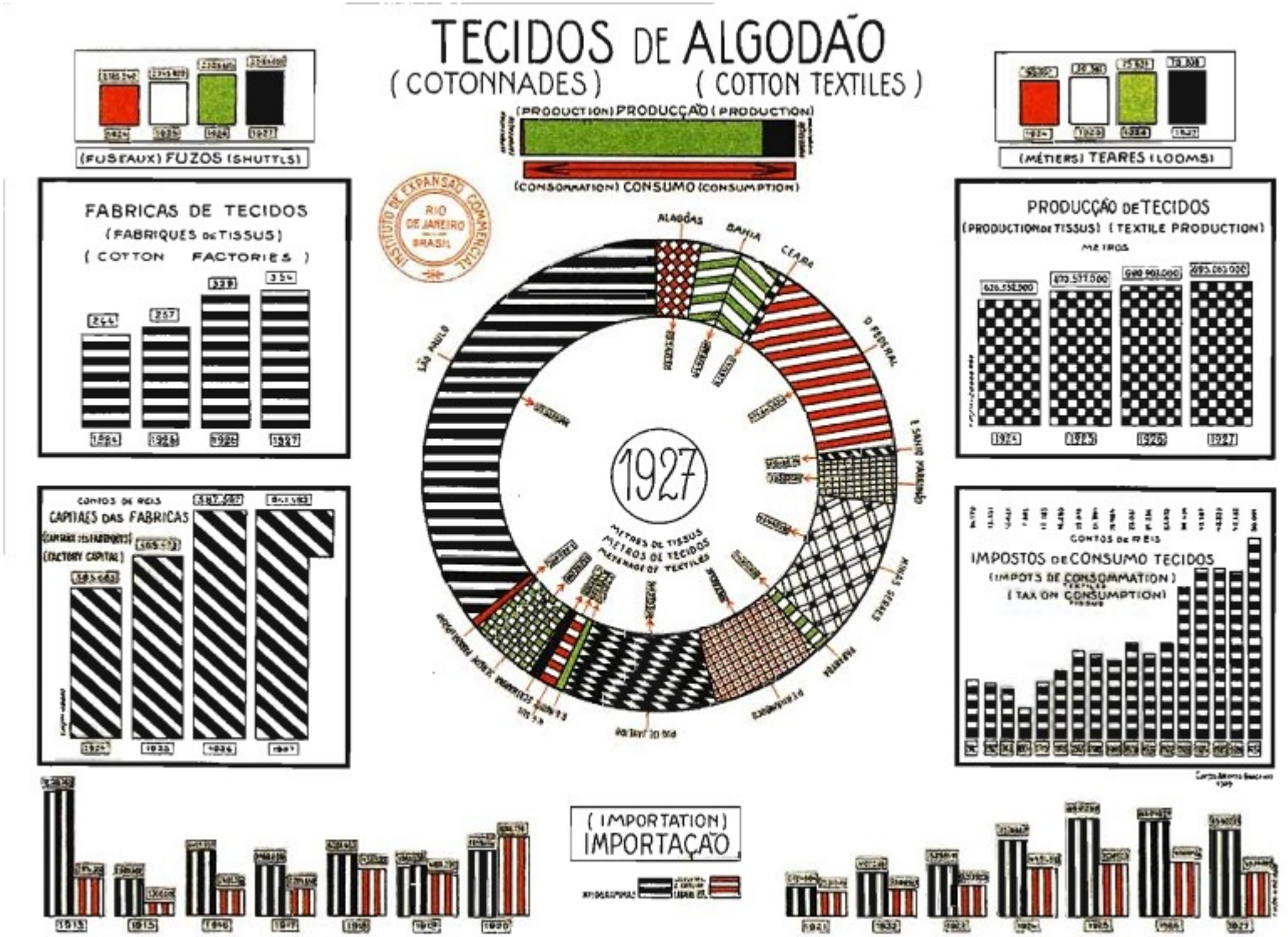


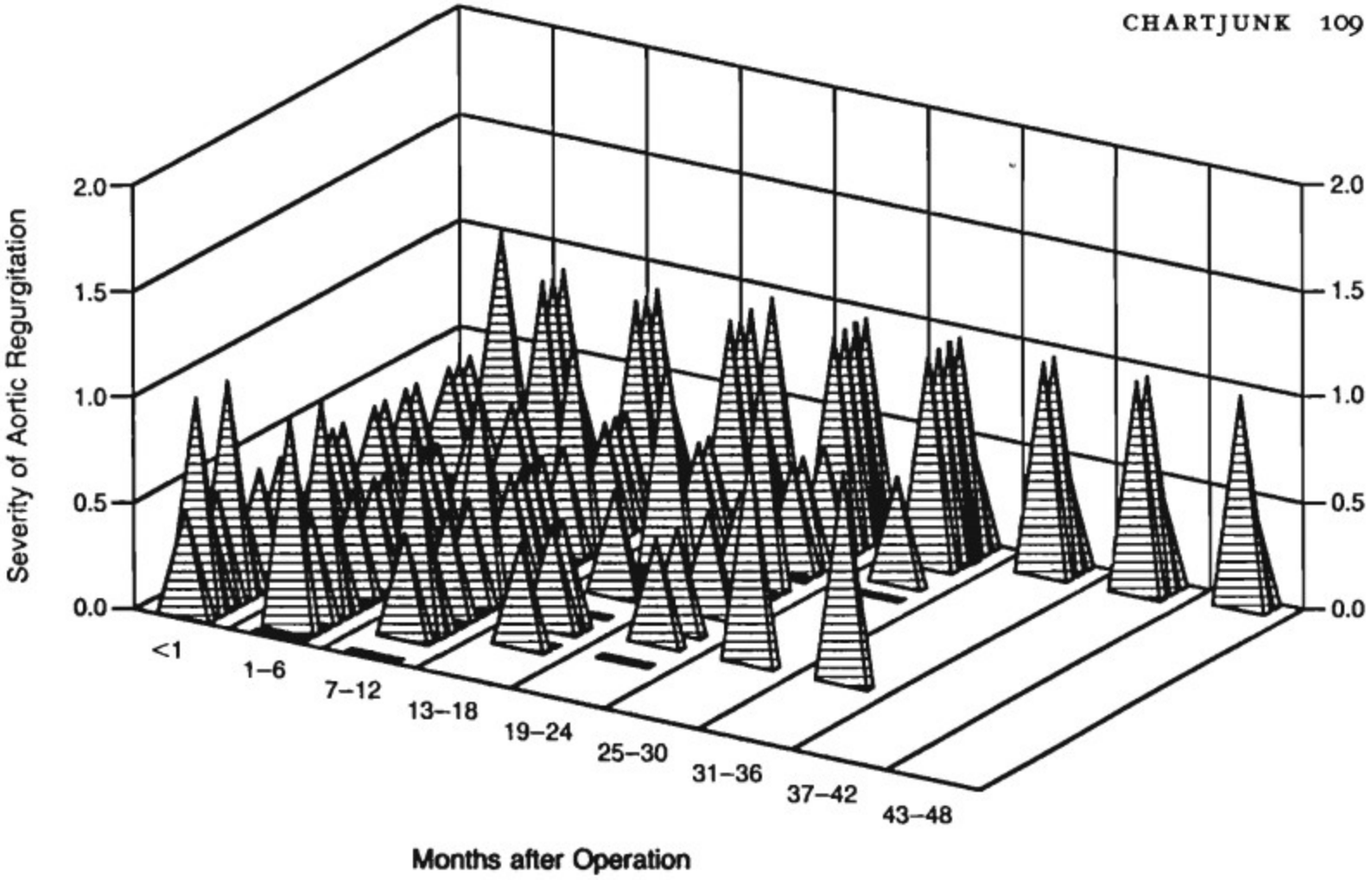
Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

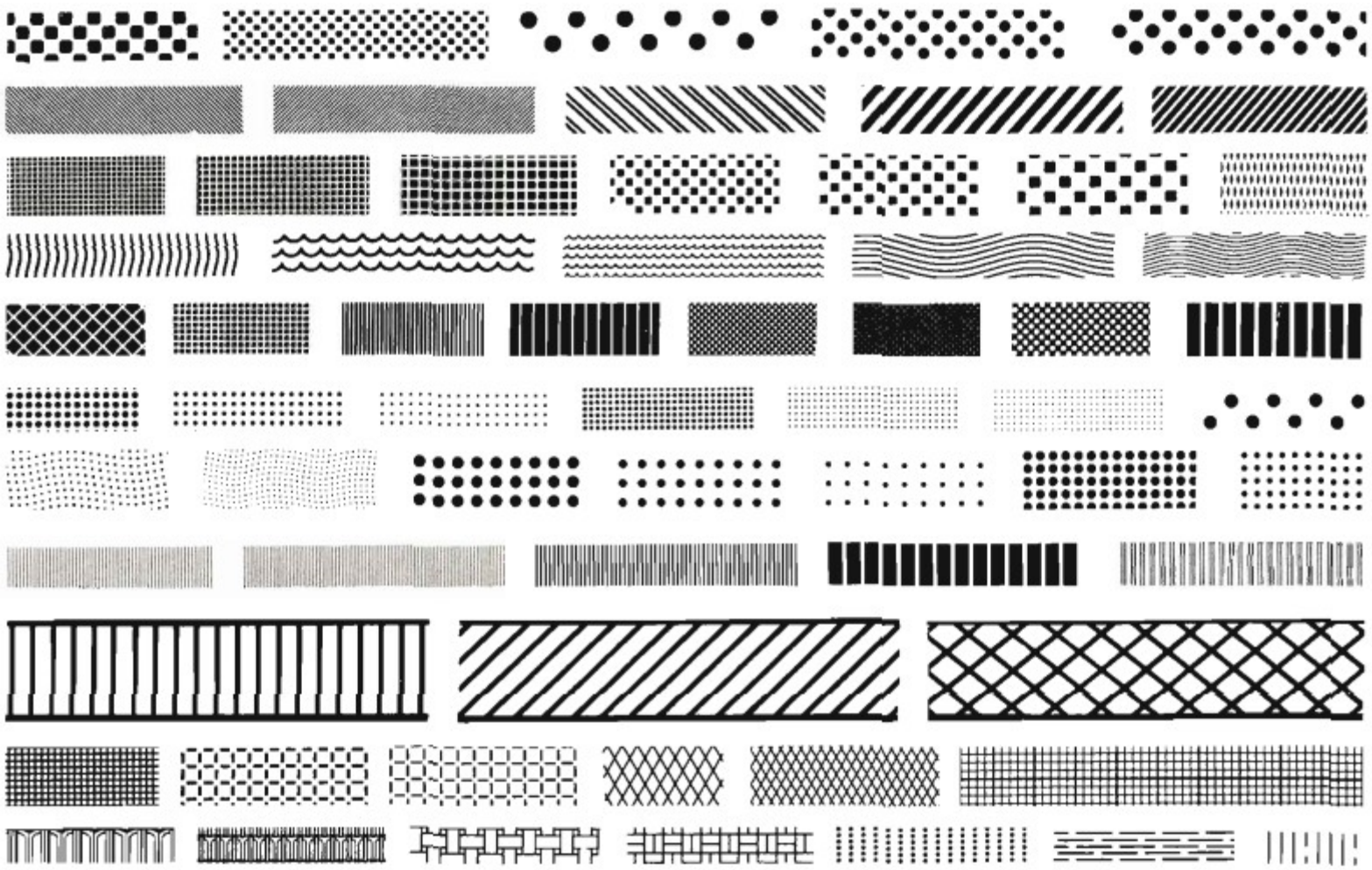


S-C
N-C
pre GS
post GS
pre tea
post tea

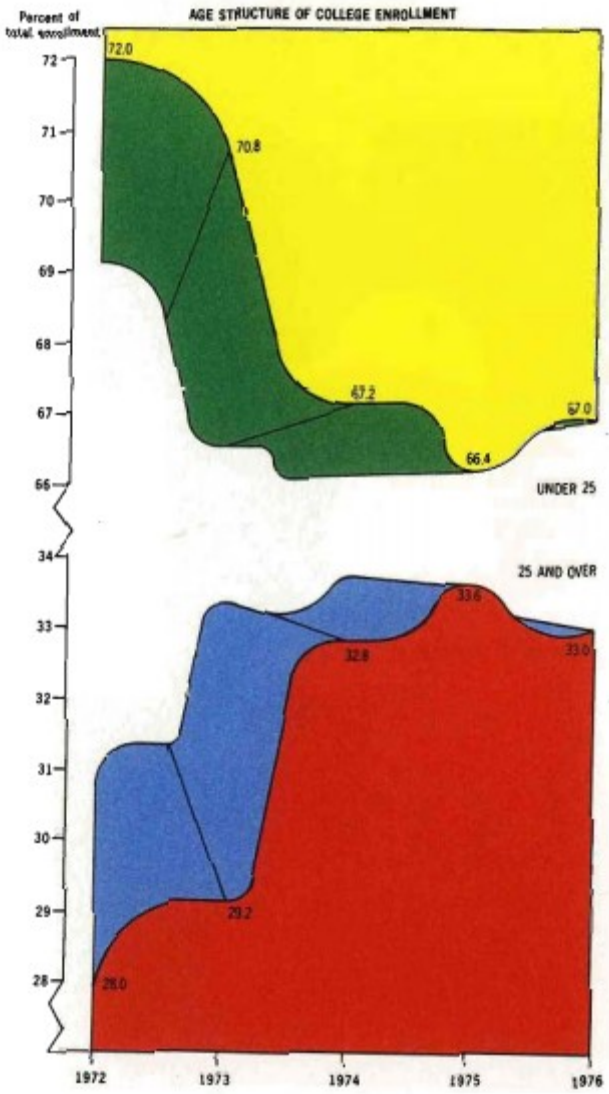
Chartjunk


TECIDOS DE ALGODÃO
(COTONNADES)
(COTTON TEXTILES)
(PRODUCTION) PRODUCÇÃO (PRODUCTION)
(CONSOMMATION) CONSUMO (CONSUMPTION)
(FUSEAUX) FUZOS (SHUTTLES)
(MÉTIERS) TEARES (LOOMS)
FABRICAS DE TECIDOS
(FABRIQUES DE TISSUS)
(COTTON FACTORIES)
PRODUCÇÃO DE TECIDOS
(PRODUCTION DE TISSUS) (TEXTILE PRODUCTION)
METROS
CONTOS DE RÉIS
CAPITAES DAS FABRICAS
IMPOSTOS DE CONSUMO TECIDOS
(IMPOTS DE CONSOMMATION)
(TAX ON CONSUMPTION)
1927
METROS DE TISSUS
METROS DE TECIDOS
METRES OF TEXTILES
(IMPORTATION)
IMPORTAÇÃO


Severity of Aortic Regurgitation
2.0
1.5
1.0
0.5
0.0
<1
1–6
7–12
13–18
19–24
25–30
31–36
37–42
43–48
Months after Operation



Tufte (1983, p.118) says, *"This may well be the worst graphic ever to find its way into print."*



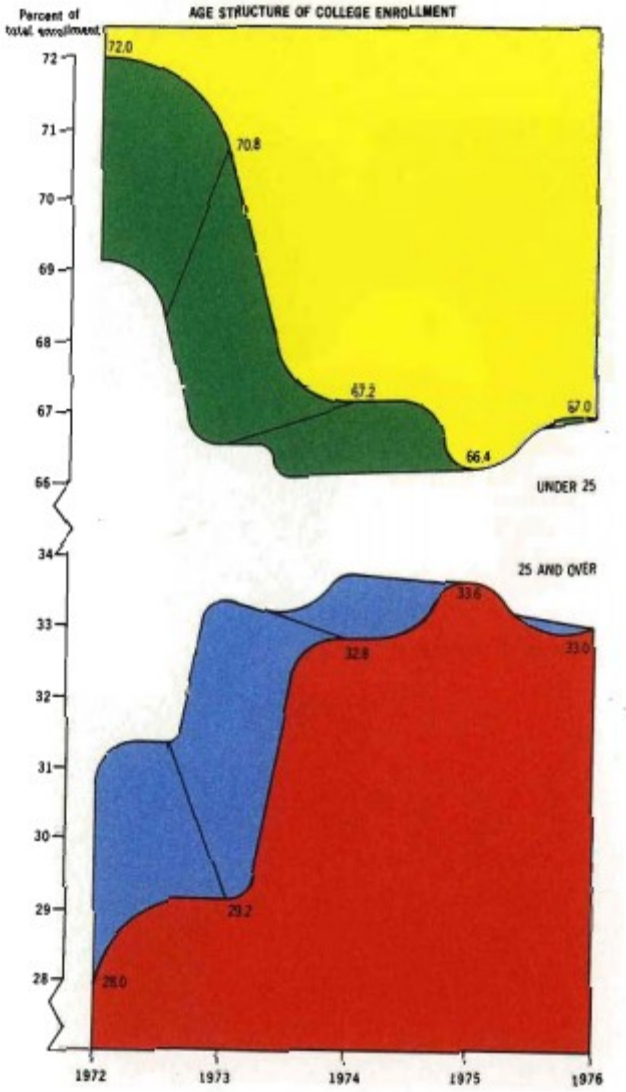
Substance? NO!

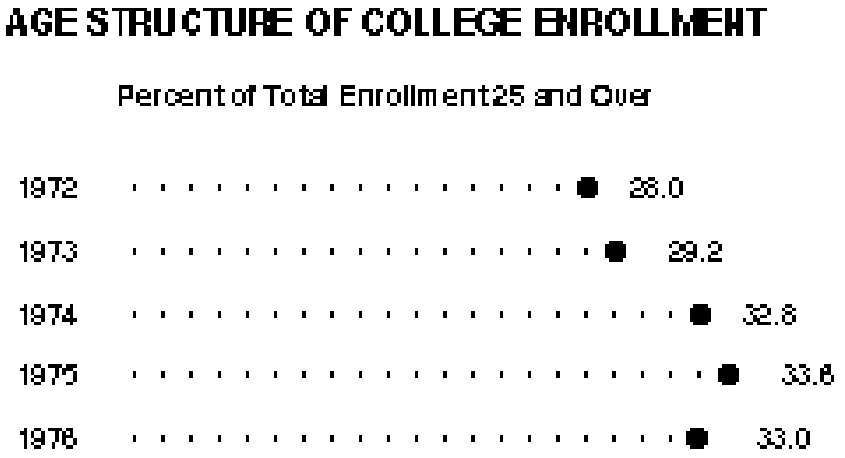
ONLY 5 NUMBERS!!

Colours.. YES
3D effects... YES
Disguised redundancy...YES:

mirror imaging
curved lines

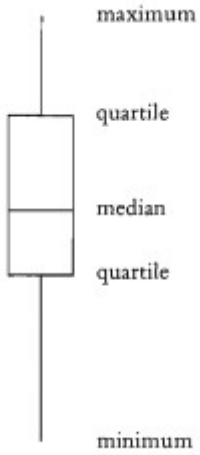
Tufte (1983, p.118) says, *"This may well be the worst graphic ever to find its way into print."*



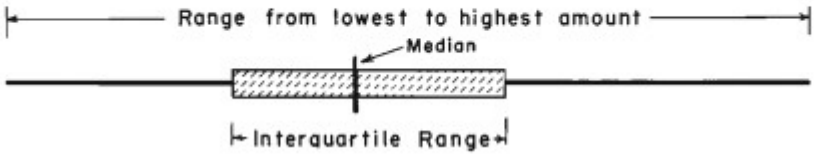


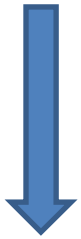
Data-ink maximization

and John Tukey's "box plot"



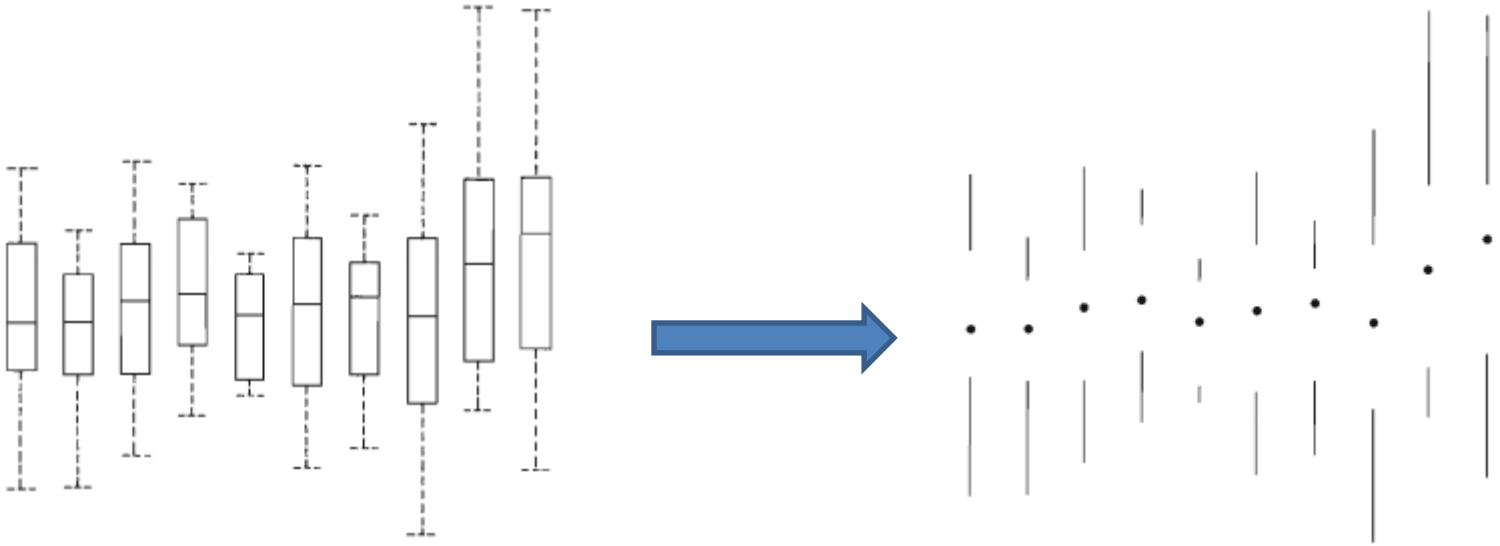
Mary Eleanor Spear's "range bar"

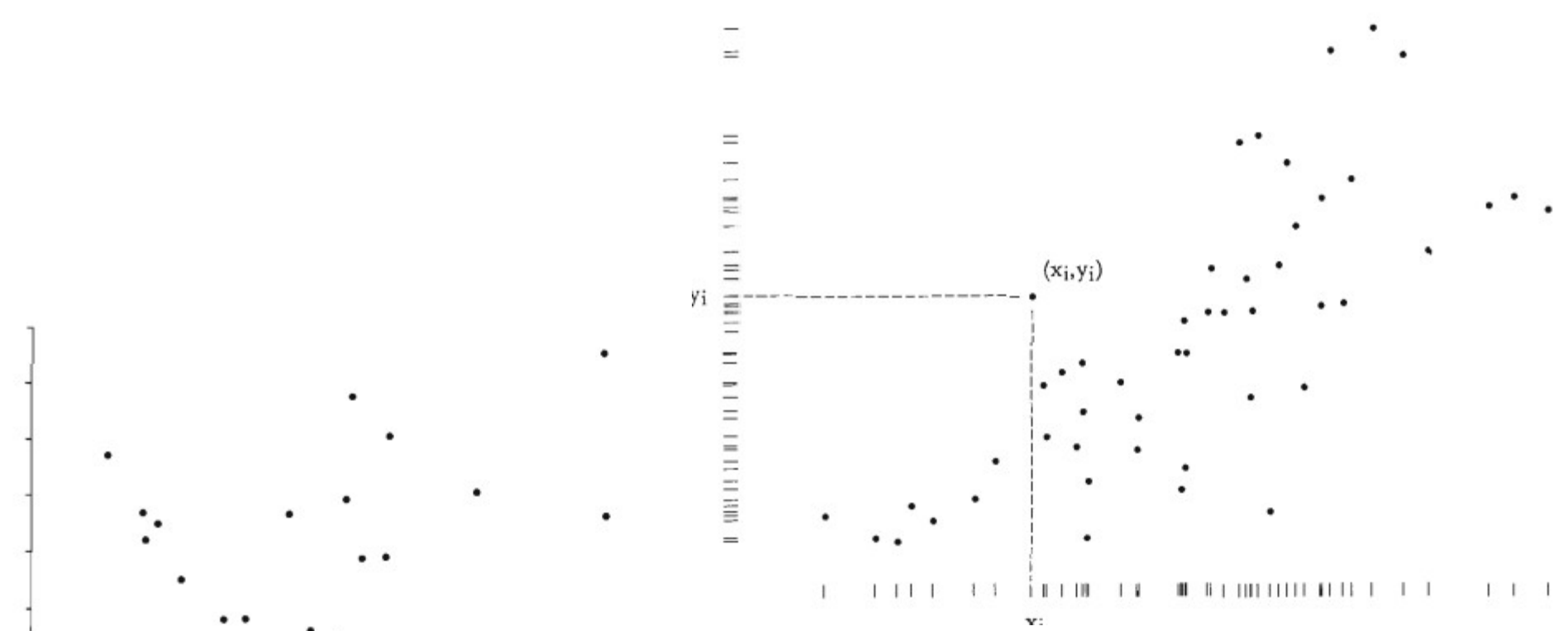




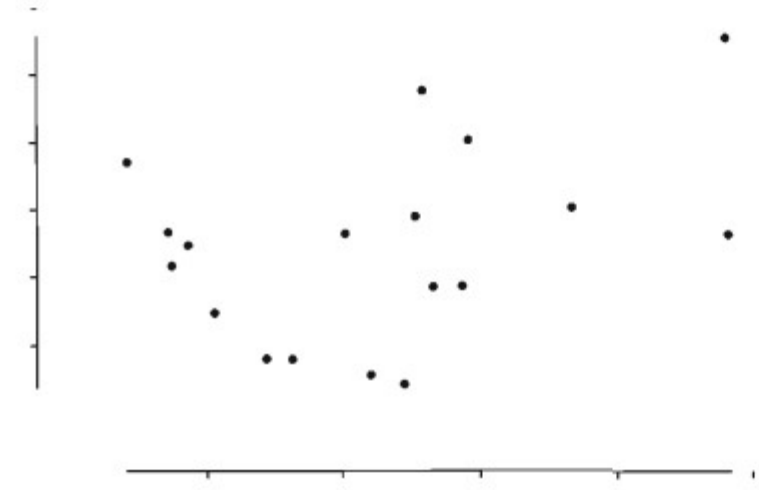
can be mostly erased without loss of information:

_____ . _____

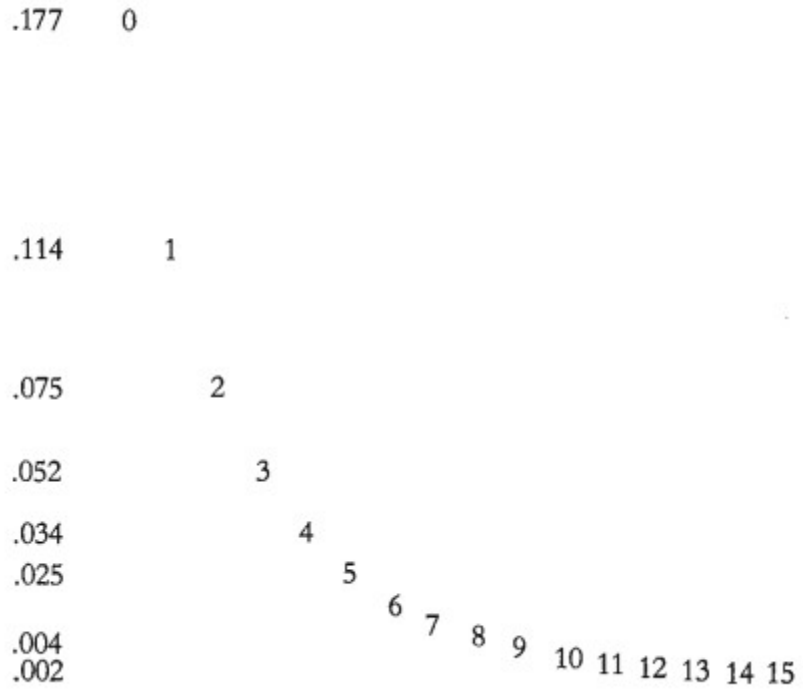


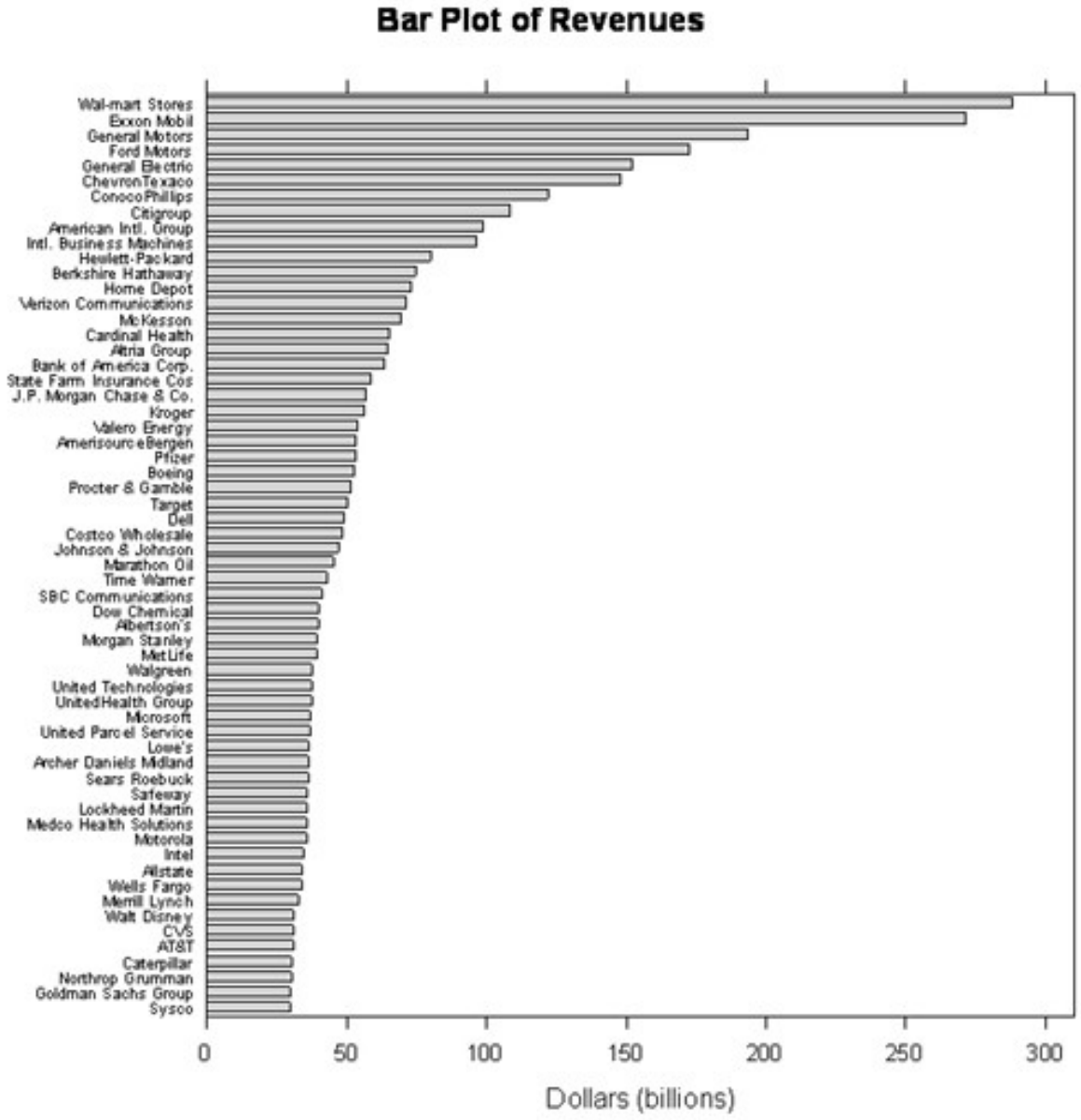


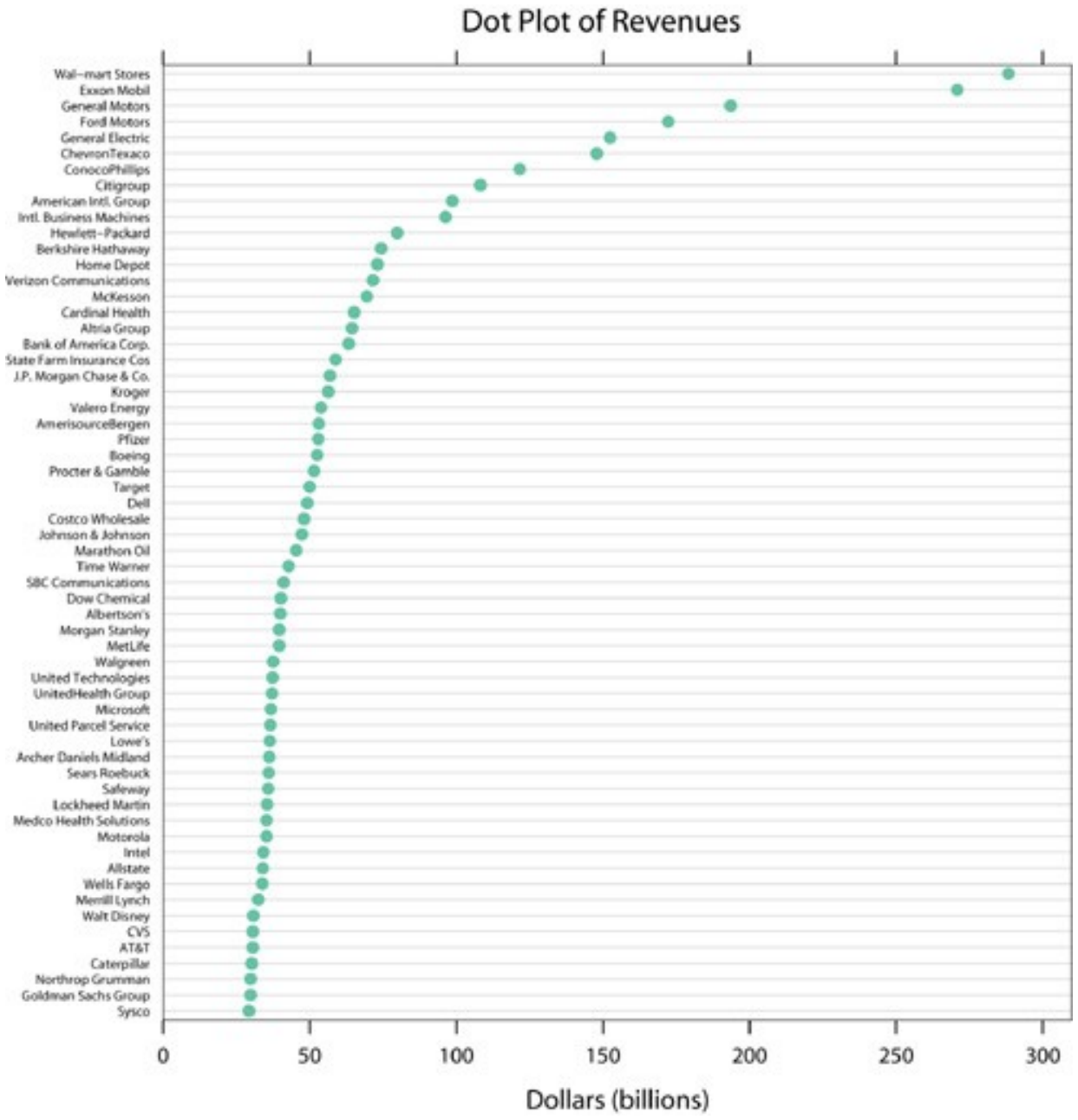
Conventional Scatterplot

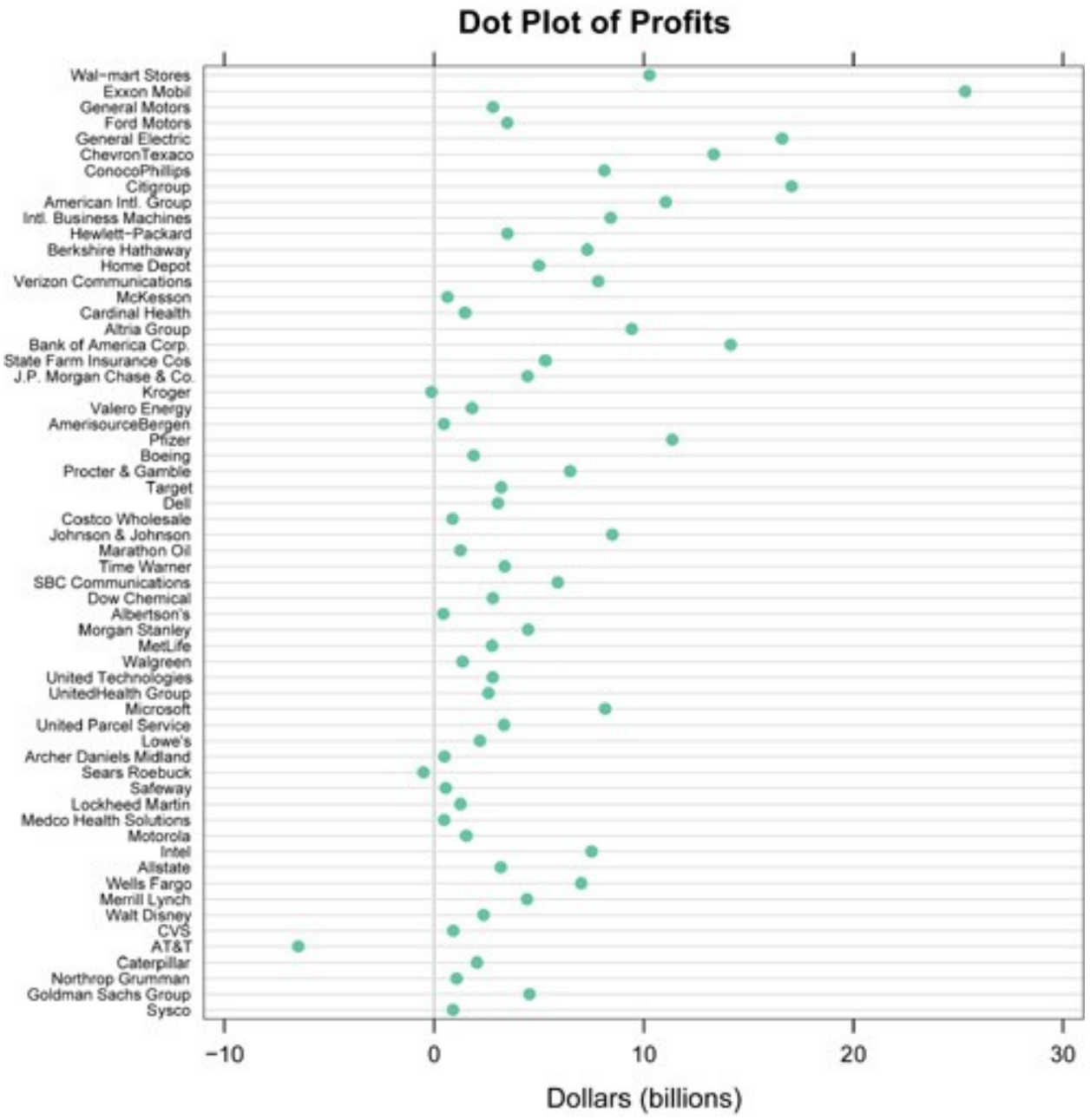


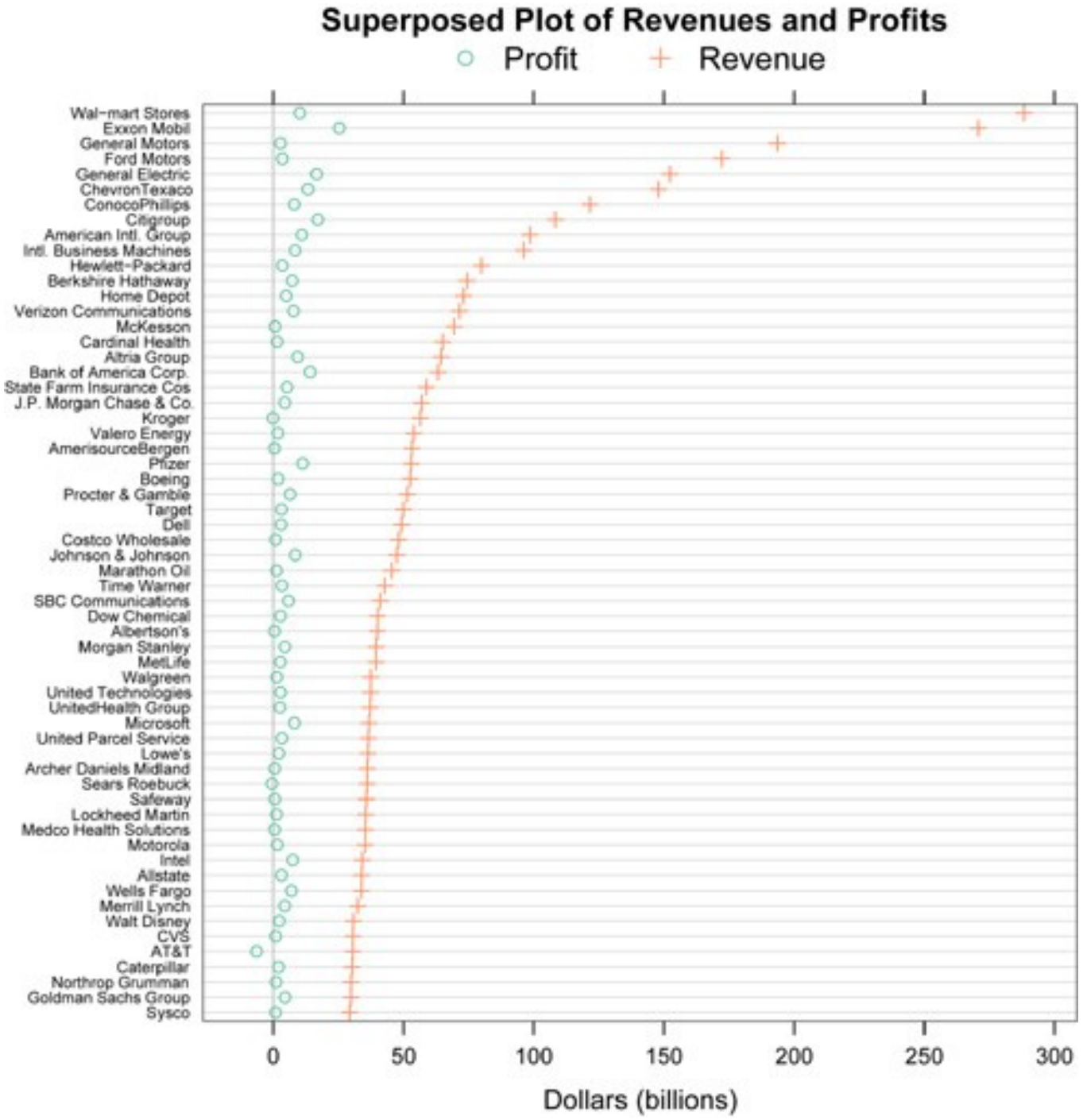
Range-Frame

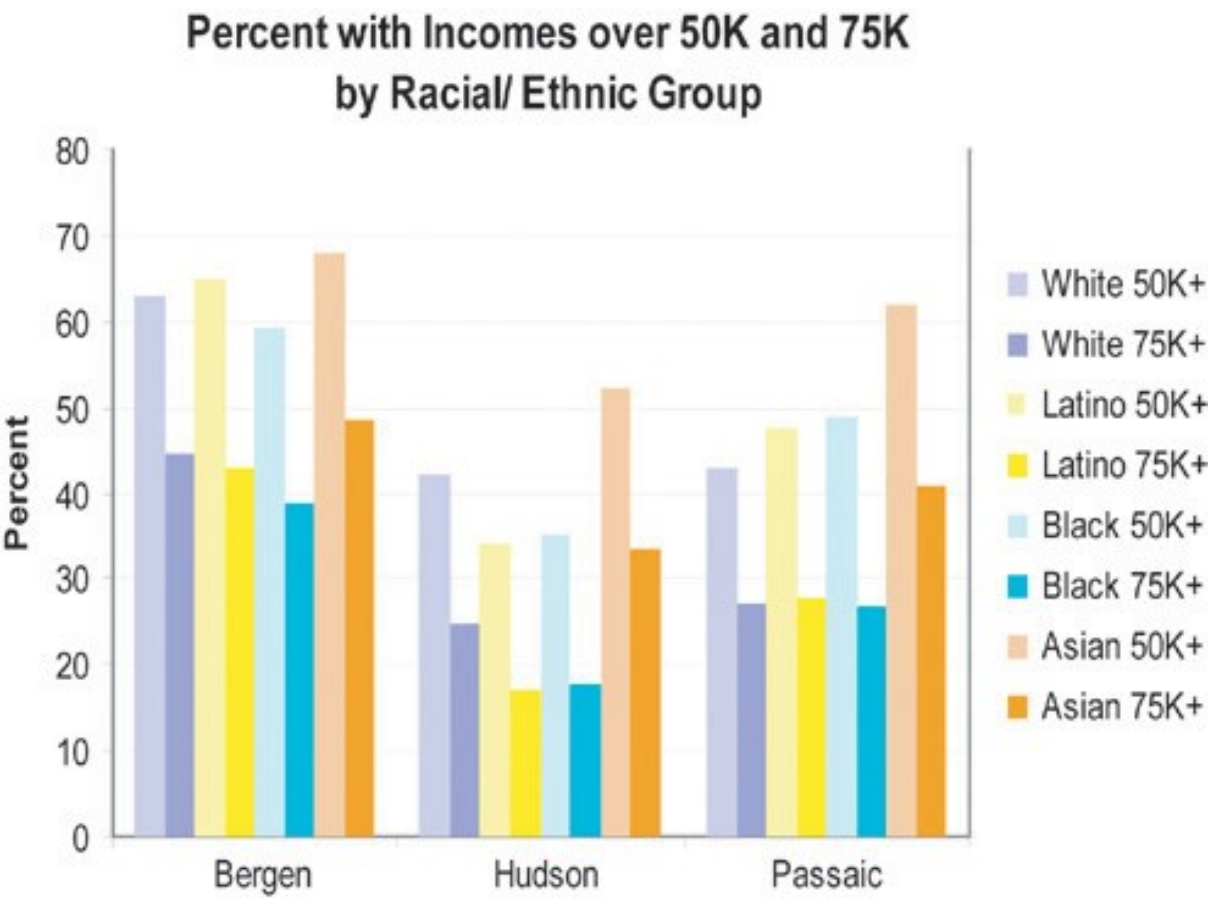

.177 0
.114 1
.075 2
.052 3
.034 4
.025 5
6
7
8
9
10 11 12 13 14 15
.004
.002

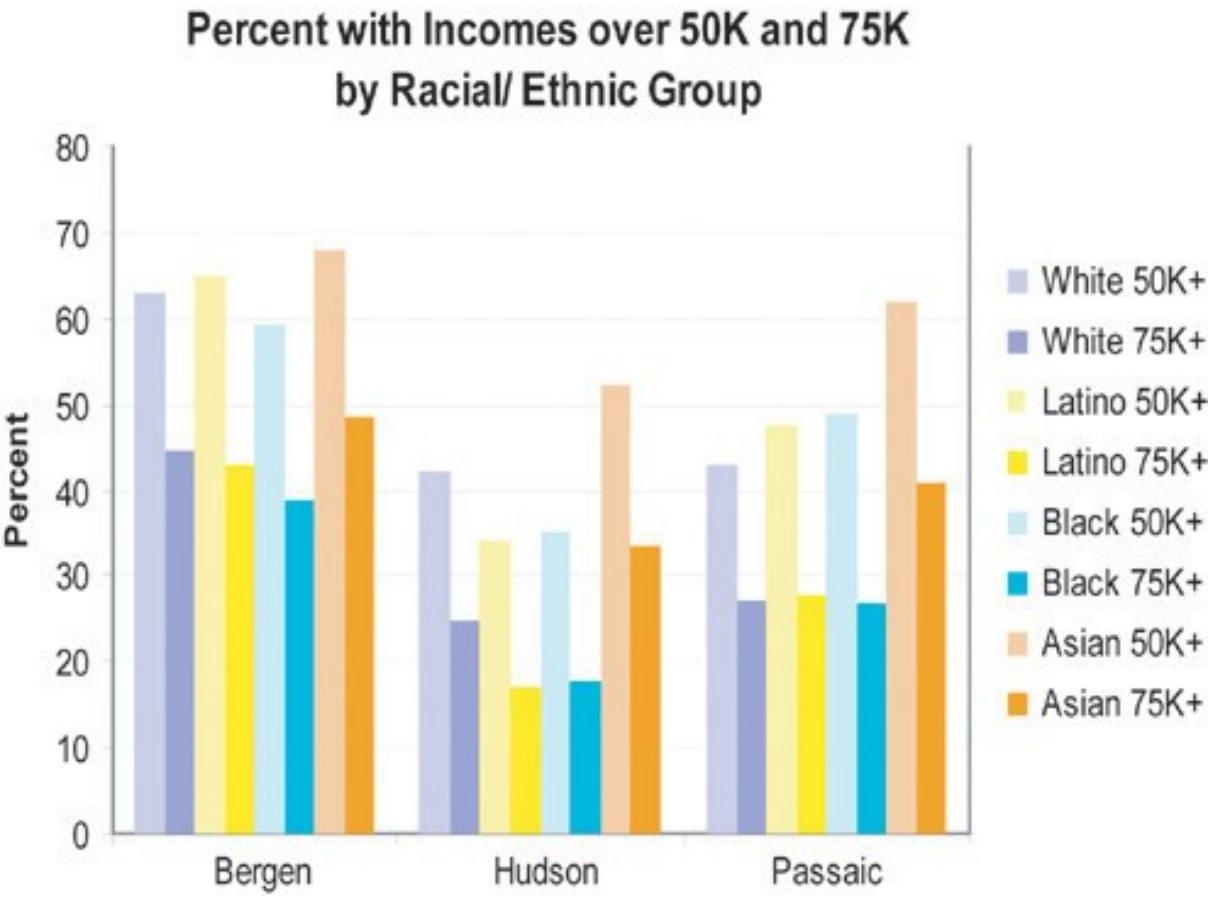

Bar Plot of Revenues
Wal-mart Stores
Exxon Mobil
General Motors
Ford Motors
General Electric
ChevronTexaco
ConocoPhillips
Citigroup
American Intl. Group
Intl. Business Machines
Hewlett-Packard
Berkshire Hathaway
Home Depot
Verizon Communications
McKesson
Cardinal Health
Altria Group
Bank of America Corp.
State Farm Insurance Cos
J.P. Morgan Chase & Co.
Kroger
Valero Energy
AmerisourceBergen
Pfizer
Boeing
Procter & Gamble
Target
Dell
Costco Wholesale
Johnson & Johnson
Marathon Oil
Time Warner
SBC Communications
Dow Chemical
Albertson's
Morgan Stanley
MetLife
Walgreen
United Technologies
UnitedHealth Group
Microsoft
United Parcel Service
Lowe's
Archer Daniels Midland
Sears Roebuck
Safeway
Lockheed Martin
Medco Health Solutions
Motorola
Intel
Allstate
Wells Fargo
Merrill Lynch
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
0
50
100
150
200
250
300
Dollars (billions)

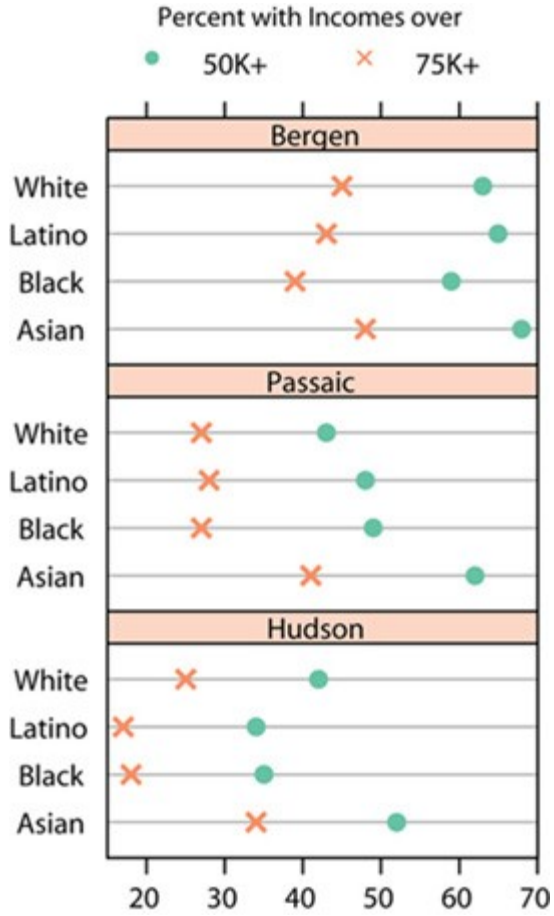

Dot Plot of Revenues
Wal–mart Stores
Exxon Mobil
General Motors
Ford Motors
General Electric
ChevronTexaco
ConocoPhillips
Citigroup
American Intl. Group
Intl. Business Machines
Hewlett–Packard
Berkshire Hathaway
Home Depot
Verizon Communications
McKesson
Cardinal Health
Altria Group
Bank of America Corp.
State Farm Insurance Cos
J.P. Morgan Chase & Co.
Kroger
Valero Energy
AmerisourceBergen
Pfizer
Boeing
Procter & Gamble
Target
Dell
Costco Wholesale
Johnson & Johnson
Marathon Oil
Time Warner
SBC Communications
Dow Chemical
Albertson's
Morgan Stanley
MetLife
Walgreen
United Technologies
UnitedHealth Group
Microsoft
United Parcel Service
Lowe's
Archer Daniels Midland
Sears Roebuck
Safeway
Lockheed Martin
Medco Health Solutions
Motorola
Intel
Allstate
Wells Fargo
Merrill Lynch
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
0
50
100
150
200
250
300
Dollars (billions)

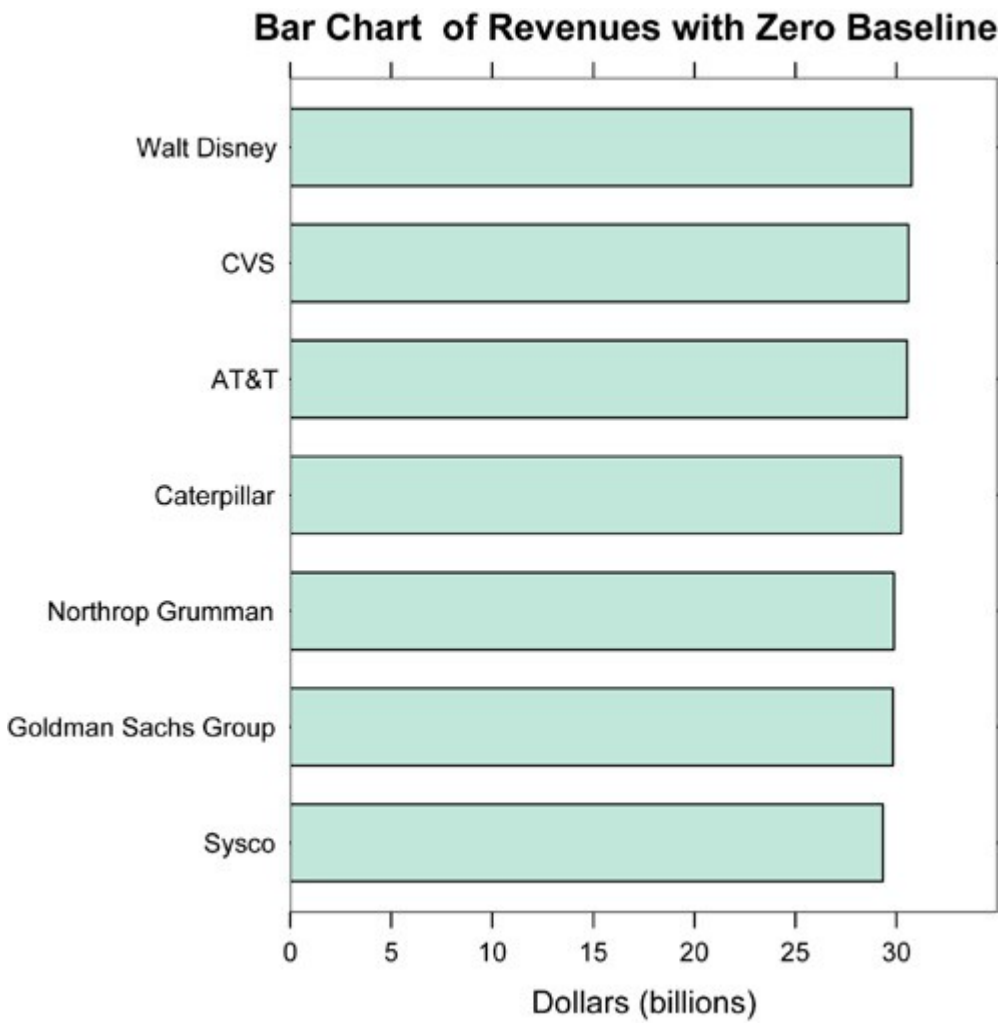

Dot Plot of Profits
Wal−mart Stores
Exxon Mobil
General Motors
Ford Motors
General Electric
ChevronTexaco
ConocoPhillips
Citigroup
American Intl. Group
Intl. Business Machines
Hewlett−Packard
Berkshire Hathaway
Home Depot
Verizon Communications
McKesson
Cardinal Health
Altria Group
Bank of America Corp.
State Farm Insurance Cos
J.P. Morgan Chase & Co.
Kroger
Valero Energy
AmerisourceBergen
Pfizer
Boeing
Procter & Gamble
Target
Dell
Costco Wholesale
Johnson & Johnson
Marathon Oil
Time Warner
SBC Communications
Dow Chemical
Albertson's
Morgan Stanley
MetLife
Walgreen
United Technologies
UnitedHealth Group
Microsoft
United Parcel Service
Lowe's
Archer Daniels Midland
Sears Roebuck
Safeway
Lockheed Martin
Medco Health Solutions
Motorola
Intel
Allstate
Wells Fargo
Merrill Lynch
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
−10
0
10
20
30
Dollars (billions)

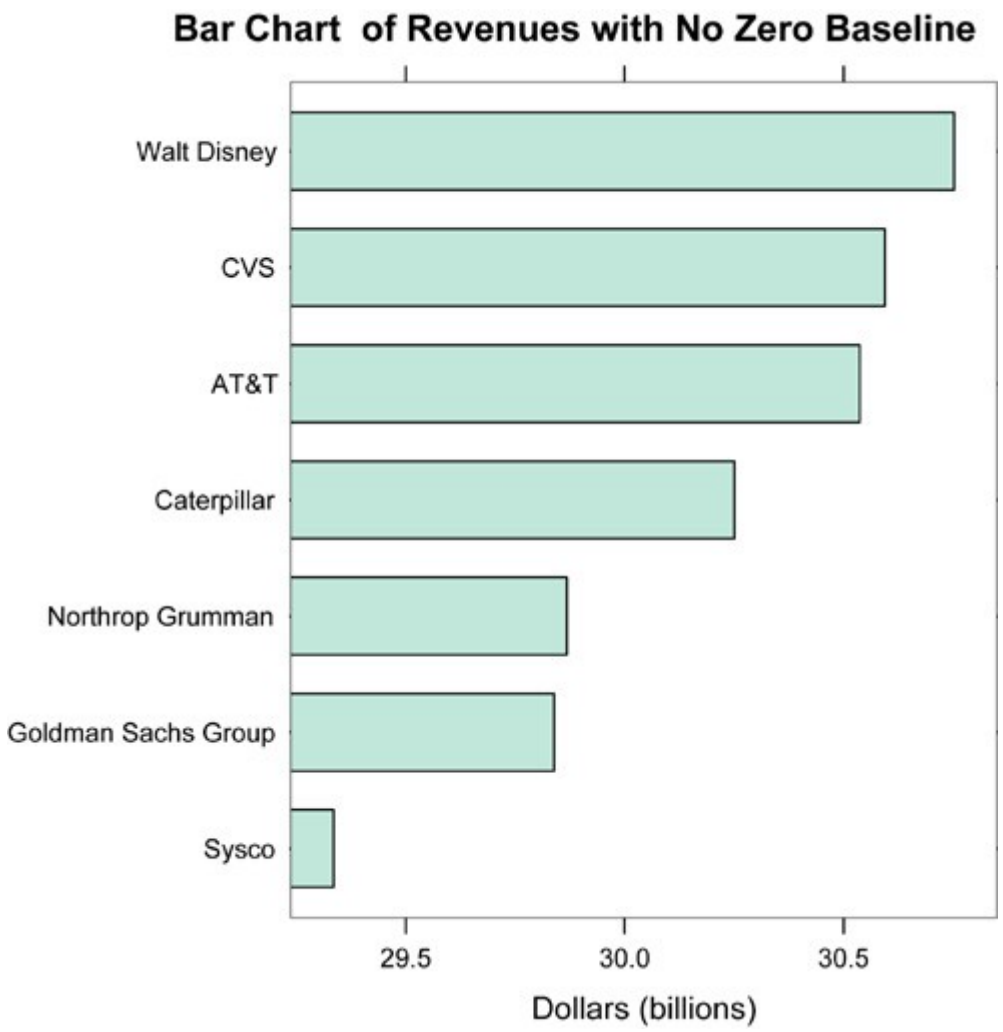

Superposed Plot of Revenues and Profits
Profit
Revenue
Wal−mart Stores
Exxon Mobil
General Motors
Ford Motors
General Electric
ChevronTexaco
ConocoPhillips
Citigroup
American Intl. Group
Intl. Business Machines
Hewlett−Packard
Berkshire Hathaway
Home Depot
Verizon Communications
McKesson
Cardinal Health
Altria Group
Bank of America Corp.
State Farm Insurance Cos
J.P. Morgan Chase & Co.
Kroger
Valero Energy
AmerisourceBergen
Pfizer
Boeing
Procter & Gamble
Target
Dell
Costco Wholesale
Johnson & Johnson
Marathon Oil
Time Warner
SBC Communications
Dow Chemical
Albertson's
Morgan Stanley
MetLife
Walgreen
United Technologies
UnitedHealth Group
Microsoft
United Parcel Service
Lowe's
Archer Daniels Midland
Sears Roebuck
Safeway
Lockheed Martin
Medco Health Solutions
Motorola
Intel
Allstate
Wells Fargo
Merrill Lynch
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
0
50
100
150
200
250
300
Dollars (billions)

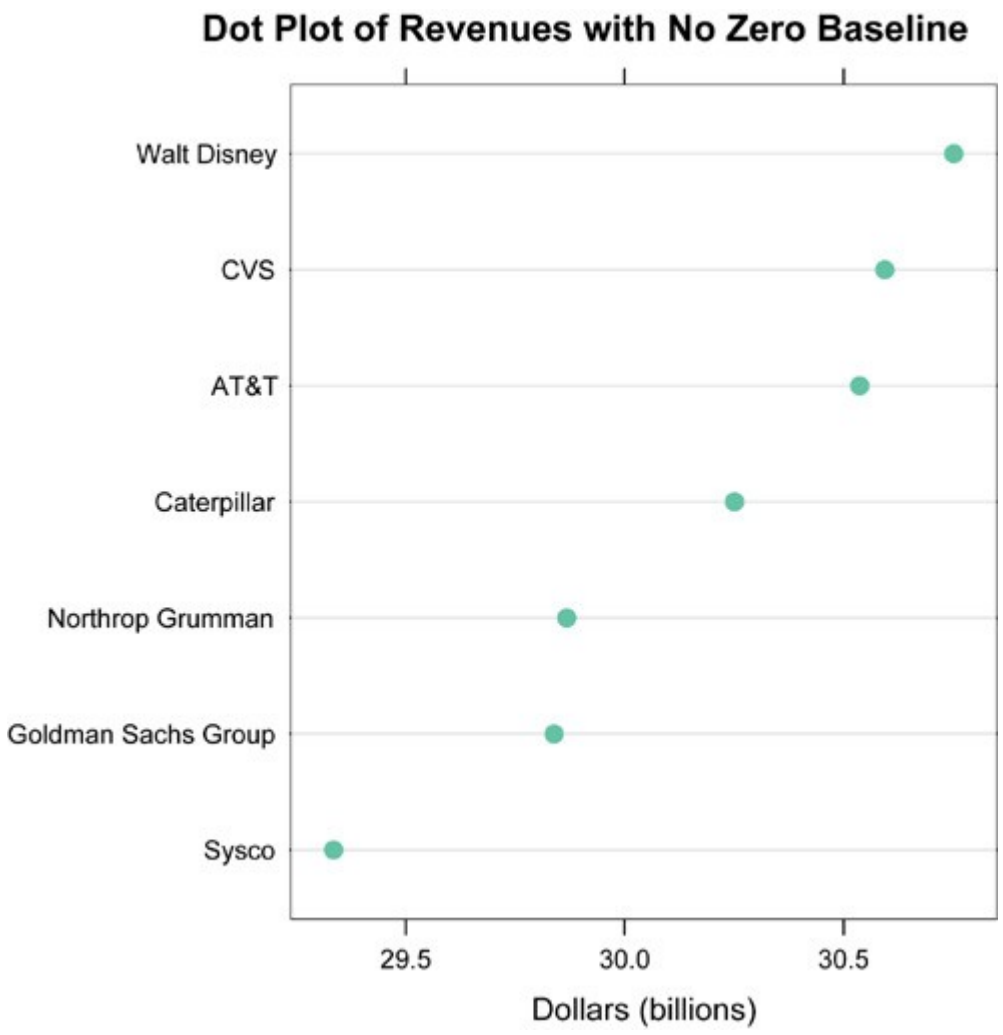

Percent with Incomes over 50K and 75K
by Racial/ Ethnic Group
Percent
80
70
60
50
40
30
20
10
0
Bergen
Hudson
Passaic
White 50K+
White 75K+
Latino 50K+
Latino 75K+
Black 50K+
Black 75K+
Asian 50K+
Asian 75K+


Percent with Incomes over 50K and 75K
by Racial/ Ethnic Group
Percent
80
70
60
50
40
30
20
10
0
Bergen
Hudson
Passaic
White 50K+
White 75K+
Latino 50K+
Latino 75K+
Black 50K+
Black 75K+
Asian 50K+
Asian 75K+


Percent with Incomes over
50K+
75K+
Bergen
White
Latino
Black
Asian
Passaic
White
Latino
Black
Asian
Hudson
White
Latino
Black
Asian
20
30
40
50
60
70


Bar Chart of Revenues with Zero Baseline
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
0
5
10
15
20
25
30
Dollars (billions)


Bar Chart of Revenues with No Zero Baseline
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
29.5
30.0
30.5
Dollars (billions)


Dot Plot of Revenues with No Zero Baseline
Walt Disney
CVS
AT&T
Caterpillar
Northrop Grumman
Goldman Sachs Group
Sysco
29.5
30.0
30.5
Dollars (billions)

Apprehension:

Ability to correctly perceive relations among variables.

Does the graph maximize apprehension of the relations among variables?

Clarity:

Ability to visually distinguish all the elements of a graph.

Are the most important elements or relations visually most prominent?

Consistency:

Ability to interpret a graph based on similarity to previous graphs.

Are the elements, symbol shapes and colors consistent with their use in previous graphs

Efficiency:

Ability to portray a possibly complex relation in as simple a way as possible.

Are the elements of the graph economically used?

Is the graph easy to interpret?

Necessity:

The need for the graph, and the graphical elements.

Is the graph a more useful way to represent the data than alternatives (table, text)?

Are all the graph elements necessary to convey the relations?

Truthfulness:

Ability to determine the true value represented by any graphical element by its magnitude relative to the implicit or explicit scale.

Are the graph elements accurately positioned and scaled?