CMPT 880/890
Writing labs

Outline

- Presenting quantitative data in visual form
- Tables, charts, maps, graphs, and diagrams
- Information visualization

Why present data?

- Data is evidence
- “Let the data speak for itself”

0.304864	0.23958	0.823592	0.220172	0.166601	0.21703	0.379094	0.888709	0.604231	0.289701	0.69145
0.71293	0.247149	0.020057	0.975823	0.385206	0.920183	0.234845	0.423193	0.129751	0.106599	0.655337
0.567528	0.319104	0.938442	0.412772	0.395283	0.224406	0.550002	0.898919	0.246467	0.414323	0.549748
0.291932	0.717761	0.681091	0.863136	0.015325	0.205211	0.708156	0.536779	0.718398	0.497219	0.280734
0.280663	0.440627	0.416695	0.840344	0.838419	0.177335	0.233469	0.264562	0.612207	0.691207	0.900643
0.286143	0.068406	0.575928	0.887799	0.80249	0.569998	0.240159	0.890158	0.777556	0.388025	0.442095
0.521428	0.447712	0.191131	0.174906	0.590936	0.905836	0.141818	0.852352	0.738399	0.250746	0.082781
0.589554	0.575669	0.647259	0.203487	0.247181	0.608416	0.06918	0.558221	0.455767	0.185315	0.770355
0.978779	0.309747	0.312057	0.867348	0.225182	0.992644	0.356866	0.177884	0.38365	0.354664	0.014401
0.075315	0.969136	0.517158	0.342862	0.89708	0.797646	0.988338	0.915916	0.72729	0.823894	0.521123
0.196744	0.386795	0.124285	0.47188	0.791731	0.221197	0.889446	0.257013	0.332555	0.968394	0.869113
0.682876	0.808793	0.810331	0.811081	0.884151	0.168341	0.395406	0.824098	0.330086	0.116181	0.742156
0.600207	0.645369	0.786829	0.875453	0.298435	0.036772	0.198626	0.696289	0.250192	0.715313	0.862473
0.894277	0.32921	0.72527	0.80072	0.570146	0.757558	0.169824	0.638069	0.406614	0.556612	0.199865
0.231218	0.900314	0.713952	0.822915	0.512035	0.988635	0.420775	0.837481	0.205448	0.940655	0.853948
0.758824	0.624205	0.138129	0.019501	0.965186	0.472978	0.611144	0.965701	0.00493	0.240562	0.617086
0.201367	0.73526	0.638996	0.557445	0.608896	0.516824	0.938281	0.291482	0.351785	0.682883	0.859275
0.008101	0.001418	0.806907	0.023895	0.922967	0.780037	0.532817	0.897387	0.752964	0.245308	0.868724
0.316687	0.085128	0.558296	0.629962	0.518876	0.206519	0.477239	0.7339	0.659047	0.973449	0.415621
0.837432	0.248366	0.093685	0.304459	0.103778	0.422256	0.284179	0.259666	0.000185	0.190551	0.760345
0.832536	0.259659	0.396741	0.088027	0.254271	0.761775	0.230421	0.973154	0.10831	0.836571	0.418838
0.977804	0.040827	0.996275	0.948693	0.208362	0.444456	0.480483	0.862432	0.657024	0.035052	0.711469
0.772297	0.324586	0.240894	0.960926	0.790059	0.961345	0.01602	0.545091	0.802129	0.09003	0.671506
0.760326	0.231381	0.209754	0.214668	0.19684	0.922351	0.570911	0.220543	0.505307	0.250886	0.593244
0.349302	0.670905	0.091617	0.602575	0.290617	0.013253	0.618324	0.941295	0.936434	0.242322	0.339894
0.880004	0.339471	0.673688	0.083408	0.987561	0.714778	0.013714	0.302457	0.944677	0.28802	0.677895
0.856214	0.304482	0.092897	0.761928	0.757804	0.498051	0.619115	0.298062	0.065046	0.730203	0.04564
0.756057	0.999927	0.540508	0.026911	0.816253	0.109978	0.501679	0.061164	0.860826	0.283988	0.116396
0.188162	0.426021	0.974475	0.803214	0.633659	0.66571	0.899155	0.105477	0.541041	0.465141	0.931354
0.443124	0.223574	0.75241	0.600703	0.075607	0.916727	0.49712	0.543559	0.763193	0.586542	0.682322

Why present data?

- Data is evidence
- “Let the data speak for ~~itself~~ your argument”
- Remember:
  - you are making a claim
  - how will your evidence aid your argument?
  - what are you trying to show?
- What kind of visual presentation will convey the necessary information to your readers?

Displaying quantitative information

- Table or chart?
- *Tables* are useful when:
  - it is important to see individual values
  - it is important to compare individual values
  - precision is important
  - there is only a small amount of data
- *Charts* are useful when:
  - the message can be seen in the shape of the data
  - e.g.: trends, relationships, exceptions

Guiding principles

- All of the data you present in a chart or table should be relevant to your argument
- The data should stand out clearly without distraction
- Should you make your charts and tables fancy?
  - No.

Designing tables

- Do you need a table?
  - Data can be presented inline in your text
    - “In 1996, on average, men earned $32,144 a year, and women $23,710, a difference of $8,434.”
    - This takes up much less space than a table!
  - Use a table only when the numbers make the paragraph confusing or clumsy

Designing tables

- What conclusion(s) do you want the reader to draw?
  - Too much information will obscure the comparisons
- What rows and columns?
- What order for the rows and columns?
- How many rows and columns?
  - heuristics: 7x7 maximum; 100 items maximum

Designing tables

- Formatting guidelines:
  - Rows and columns should be clearly labeled
  - White space instead of lines between columns
  - Lines between rows for grouping
  - Use appropriate precision
    - Fewer significant digits = easier to read
  - Do not repeat symbols such as % or $ within the table
    - Include them in the row or column titles.
  - Should you make the table fancy?
    - No.

0.304864	0.23958	0.823592	0.220172
0.71293	0.247149	0.020057	0.975823
0.567528	0.319104	0.938442	0.412772
0.291932	0.717761	0.681091	0.863136
0.280663	0.440627	0.416695	0.840344

	June	July	August	Average
EEE	0.97	0.56	0.79	0.77
HP	0.61	0.76	0.52	0.63
Apple	0.70	0.39	0.29	0.46
Lenovo	0.28	0.61	0.18	0.36
Dell	0.70	0.01	0.33	0.35

	June	July	August	Average
EEE	0.97	0.56	0.79	0.77
HP	0.61	0.76	0.52	0.63
Apple	0.70	0.39	0.29	0.46
Lenovo	0.28	0.61	0.18	0.36
Dell	0.70	0.01	0.33	0.35

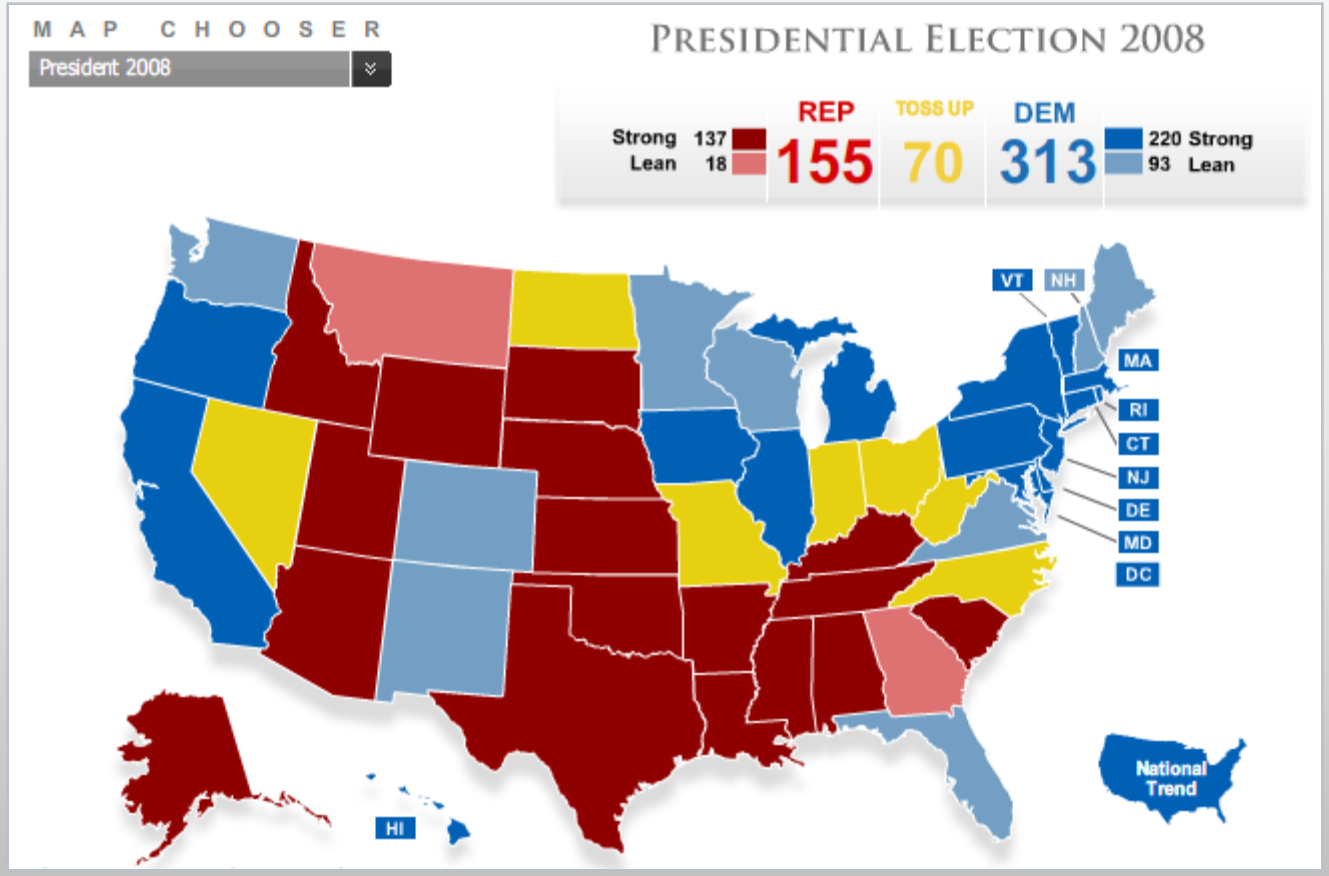
	Area	Fresh	Marginal	Brackish	Saline	All types
Western Australia	*2 622 000*	578	1 240	652	261	2 740
Queensland	*1 174 800*	1 760	683	255	144	2 840
New South Wales	*595 900*	881	564	431	304	2 180
South Australia	*486 100*	102	647	375	86	1 210
Northern Territory	*236 700*	994	3 380	43	10	4 420
Victoria	*103 700*	469	294	691	30	862
Australia	***5 226 440***	**4 831**	**6 877**	**1 833**	**835**	**14 376**

9/26/08

State	Pollster	Date	Obama	McCain	Barr	Nader
FL	ARG	9/23-25	47%	46%	-	-
FL	Rass	9/24	47%	48%	-	-
FL	SV	9/20-22	45%	48%	-	-
FL	Zogby	9/9-12	41%	51%	4%	1%
PA	SV	9/20-22	47%	46%	-	-
PA	Rass	9/24	49%	45%	-	-
PA	Zogby	9/15-19	44%	46%	3%	-
OH	Rass	9/23	46%	47%	-	-
OH	Zogby	9/9-12	44%	48%	5%	-
MI	DFP	9/22-24	51%	38%	1%	-
MI	SV	9/22-24	48%	45%	-	-
NC	Zogby	9/9-12	47%	45%	2%	1%
VA	Rass	9/25	50%	45%	-	-
VA	Zogby	9/9-12	43%	48%	3%	-
MO	SUSA	9/23-24	46%	48%	-	-
CO	ARG	9/23-25	45%	48%	-	-
NM	Zogby	9/9-12	45%	43%	2%	-
WV	Rass	9/24	42%	50%	-	-
NH	SV	9/22-24	46%	45%	-	-
NH	Zogby	9/9-12	43%	46%	4%	2%

Leaning towards	State	Lead	Electoral Votes	Sum of votes
Obama	Vermont	34%	3	137
	Hawaii	30%	4	
	Rhode Island	21%	4	
	Illinois	15%	21	
	Massachusetts	15%	12	
	Connecticut	15%	7	
	California	14%	55	
	Maryland	14%	10	
	Maine	14%	4	
	Minnesota	12%	10	
	Oregon	10%	7	
Obama	Delaware	9%	3	70
	New York	8%	31	
	New Jersey	7%	15	
	New Hampshire	6%	4	
	Michigan	5%	17	
Undecided	Wisconsin	3%	10	130
	Iowa	3%	7	
	Pennsylvania	2%	21	
	Washington	2%	11	
	Colorado	1%	9	
	Ohio	1%	20	
	Indiana	2%	11	
	New Mexico	2%	5	
	Nevada	3%	5	
	North Carolina	4%	15	
	Virginia	4%	13	
	South Dakota	4%	3	
McCain	Missouri	5%	11	77
	West Virginia	5%	5	
	Florida	8%	27	
	Texas	9%	34	
McCain	Arizona	10%	10	121
	Arkansas	10%	6	
	Montana	11%	3	
	South Carolina	13%	8	
	Mississippi	13%	6	
	North Dakota	14%	3	
	Alabama	15%	9	
	Louisiana	17%	9	
	Kentucky	18%	8	
	Georgia	18%	15	
	Nebraska	18%	5	
	Wyoming	19%	3	
	Kansas	23%	6	
	Tennessee	24%	11	
	Oklahoma	33%	7	
	Utah	39%	5	
	Idaho	39%	4	
	Alaska	39%	3	

Lead indicates the percentage lead of winner over opponent in last published opinion poll
All opinion poll data taken on 9/14 from: http://www.usaelectionpolls.com/
Analysis by http://www.jenslapinski.com

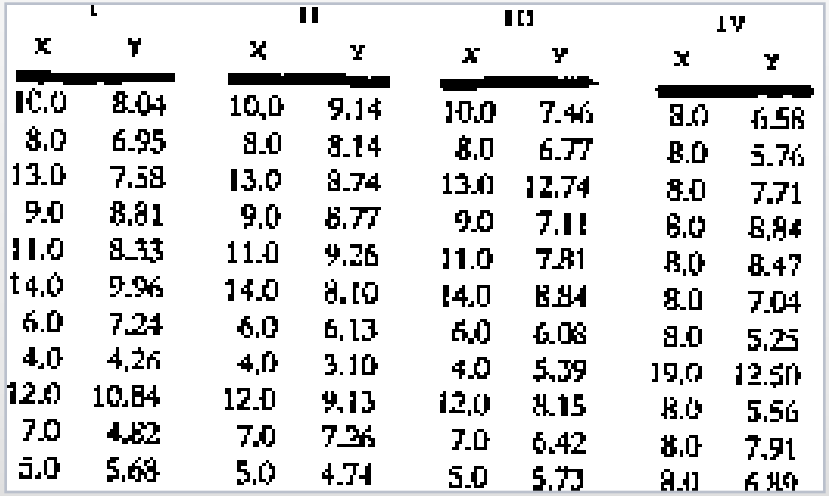

MAP CHOOSER
President 2008
PRESIDENTIAL ELECTION 2008
REP
TOSS UP
DEM
Strong 137
Lean 18
155
70
313
220 Strong
93 Lean
VT
NH
MA
RI
CT
NJ
DE
MD
DC
HI
National Trend

From tables to charts

Charts

- Less precise than tables, but greater impact
- Use a chart when the message can be seen in the shape of the data
  - trends, relationships, exceptions

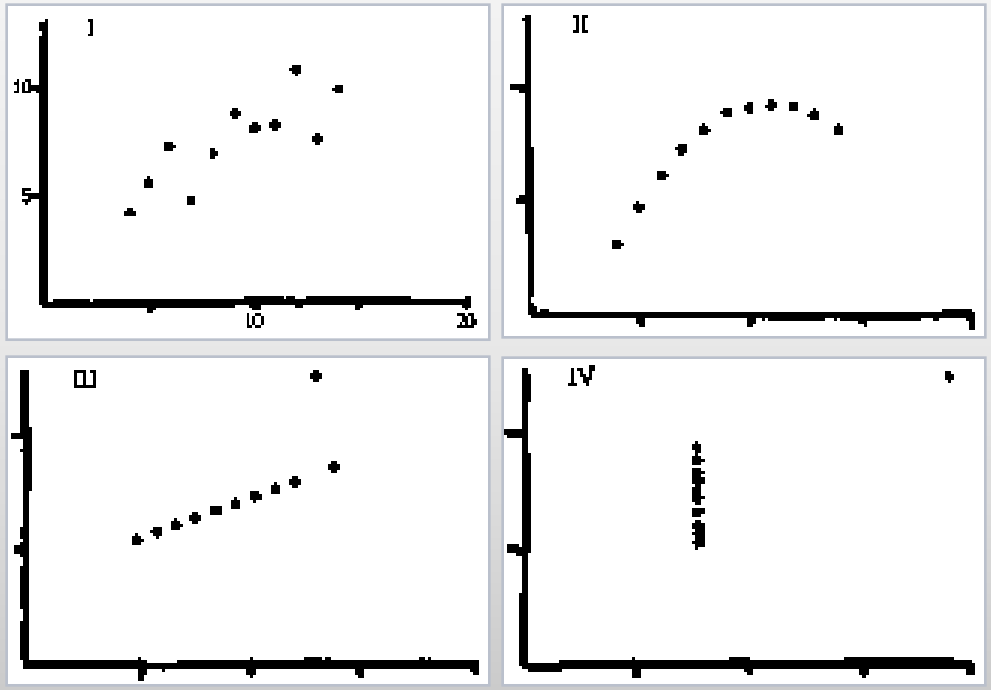
Tables may not tell the story best



I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

N:	11
mean Xs :	9.0
mean Ys :	7.5
standard error of slope estimate:	0.1
sum of squares:	110.0
regression sum of squares:	27.5
residual sum of squares of Y:	13.8
correlation coefficient:	0.8
r squared:	0.7
regression line:	Y=3+0.5X

Visual representation



Do I deserve a tax break?

Taxcomp.xls

	A	C
1	Market value (*$1000)	Improvement
2	140.0	31,120.
3	147.0	29,980.
4	151.0	38,120.
5	152.0	34,360.
6	155.0	40,710.
7	170.0	21,620.
8	172.0	42,100.
9	178.0	41,070.
10	180.0	34,210.
11	180.0	44,090.
12	182.0	55,960.
13	185.0	45,170.
14	185.0	46,820.
16	193.4	50,200.
17	194.5	71,860.
18	197.0	48,460.
19	203.0	40,720.
20	205.0	56,600.
21	213.0	42,780.
22	221.0	58,770.
23	225.0	58,960.
24	245.0	48,910.
25	248.0	62,620.
26	278.0	58,580.
27	302.5	72,200.
28	308.0	67,320.

Do I deserve a tax break?

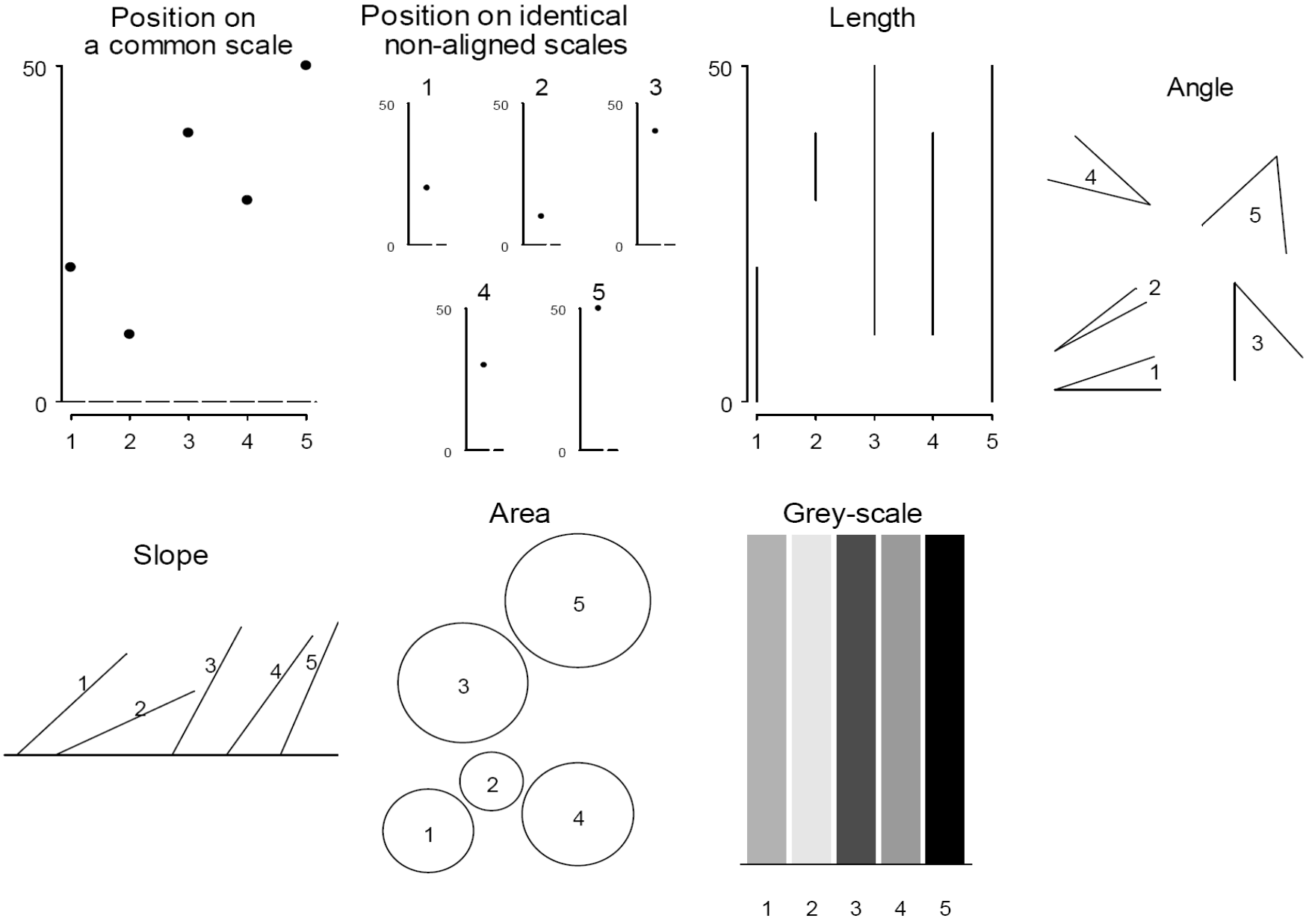
Types of charts

- Hundreds of types!
- Most of the time, use the simple types:
  - bar chart
  - line chart
  - histogram
  - scatterplot
  - pie chart

Types of attribute data

- Nominal data
  - Category data that can only be compared for equality
  - e.g. apple, orange
- Ordinal data
  - Ordering and ranking based on < and >
  - e.g. restaurant ratings
- Interval data
  - Ordering and arithmetic possible but no natural zero
  - e.g. temperatures, dates
- Ratio data
  - Ordered, natural zero
  - e.g. height, weight, age, length

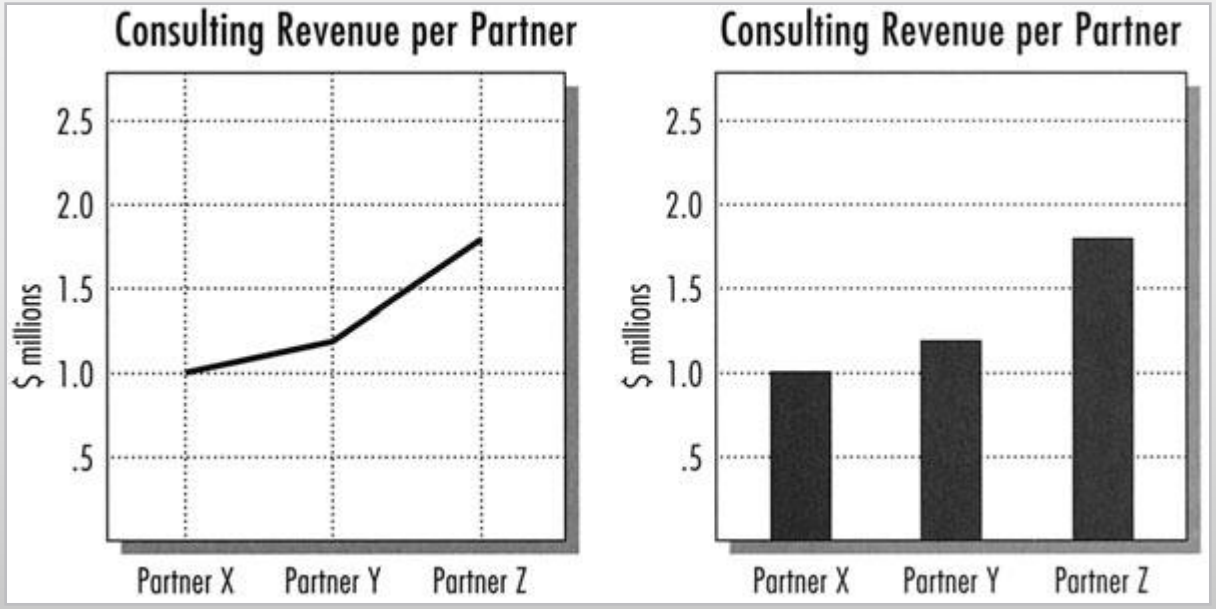
Ease of visual comparison


Position on a common scale
50
0
1 2 3 4 5
Position on identical non-aligned scales
1
2
3
4
5
50
0
Length
50
0
1 2 3 4 5
Angle
4
5
2
3
1
Slope
1
2
3
4
5
Area
5
3
2
1
4
Grey-scale
1 2 3 4 5

Bar charts

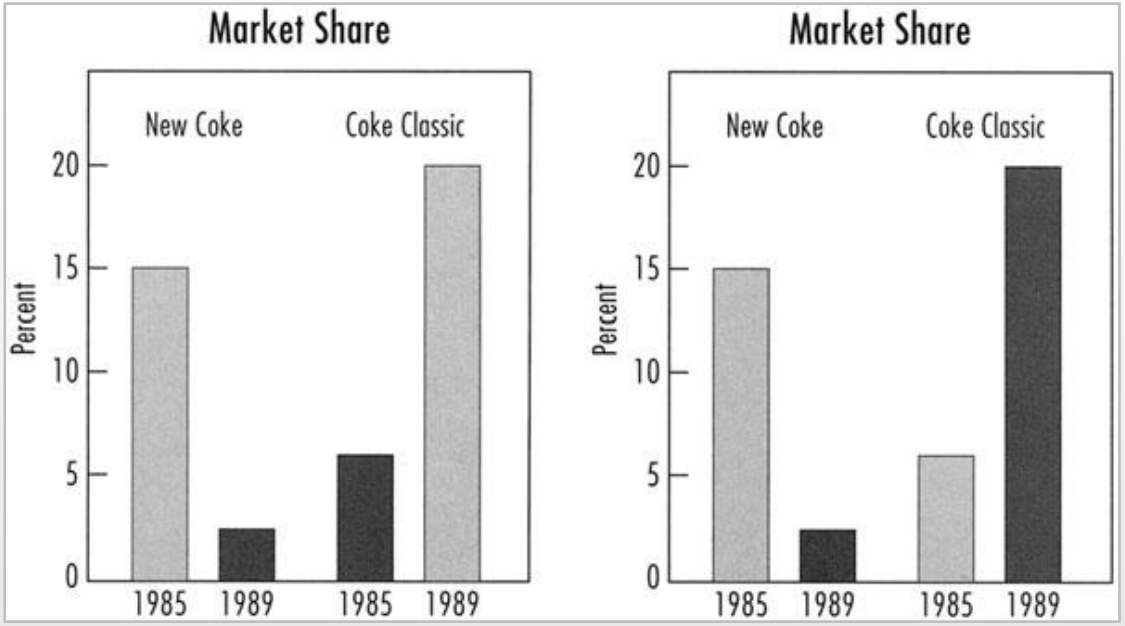
- Use a bar graph to illustrate differences between specific point values
- Don’t use a bar graph to illustrate trends
- Use a bar graph for categorical data
- Arrange categories (the X axis) to reinforce your story

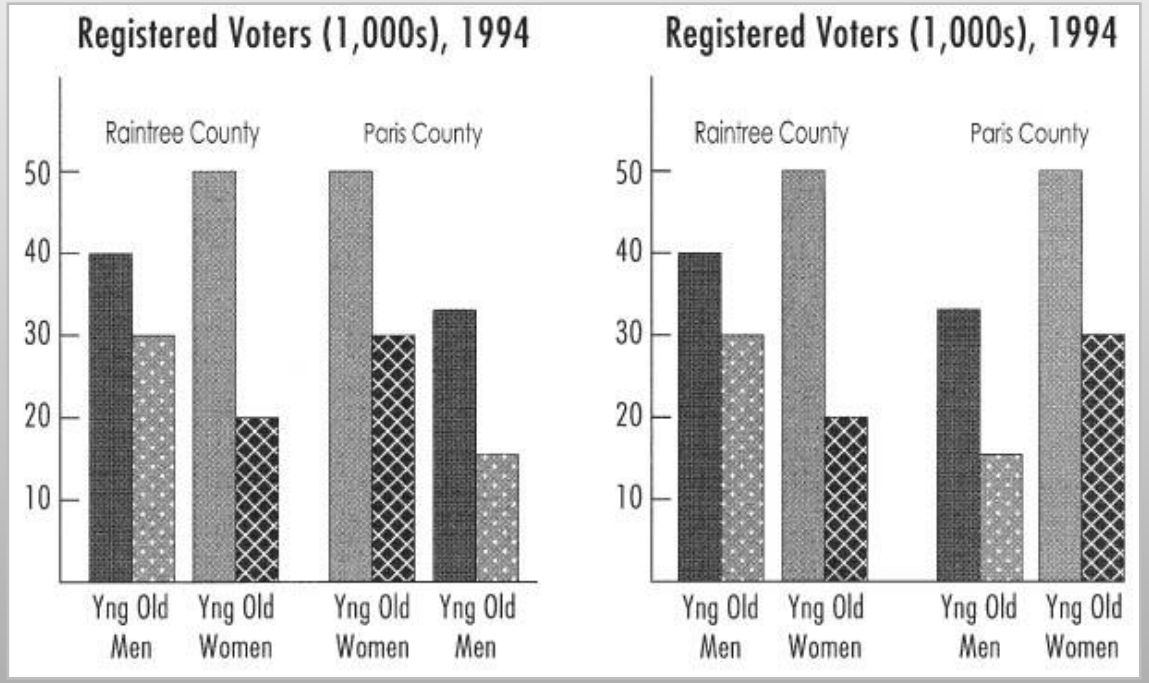
Bar charts


Consulting Revenue per Partner
$ millions
2.5
2.0
1.5
1.0
.5
Partner X
Partner Y
Partner Z
Consulting Revenue per Partner
$ millions
2.5
2.0
1.5
1.0
.5
Partner X
Partner Y
Partner Z

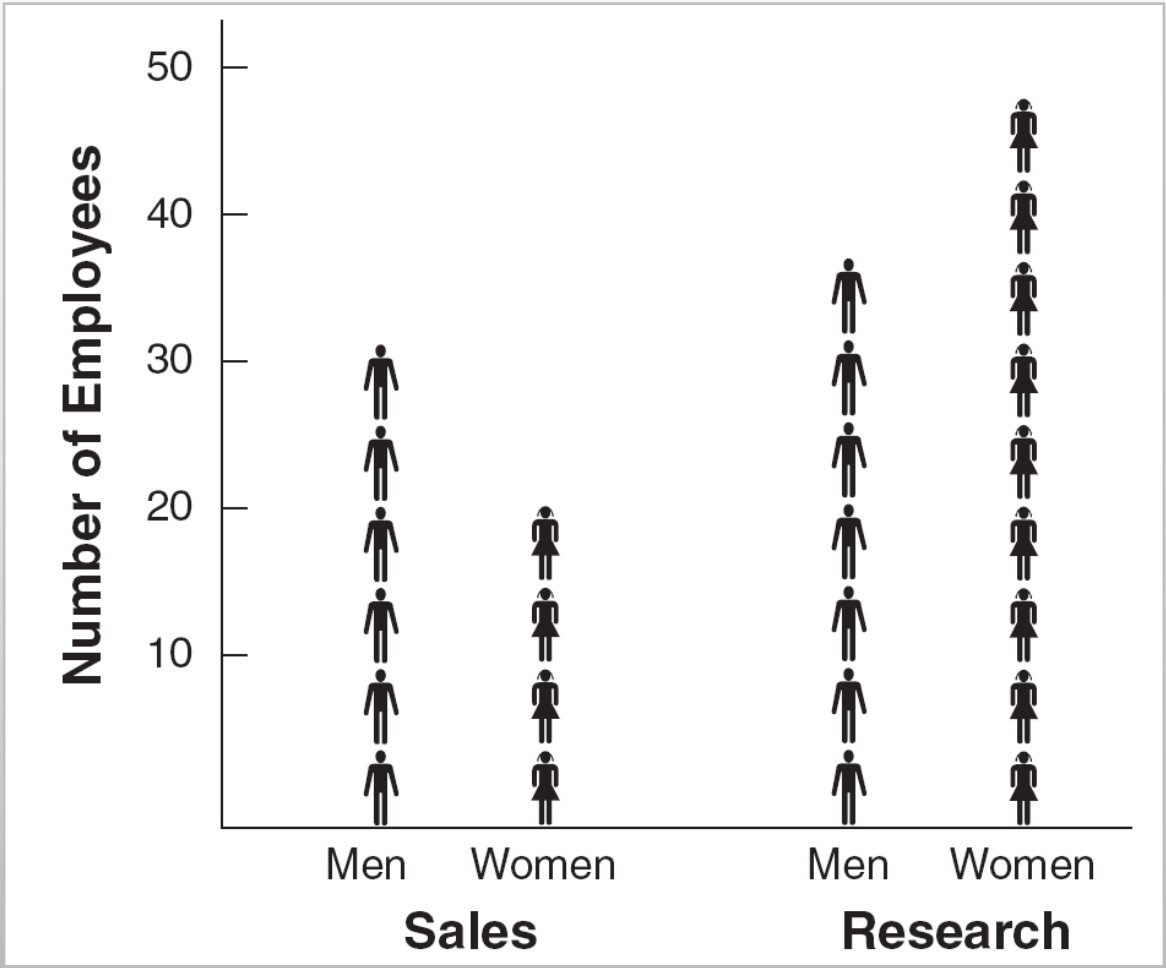
Bar charts

- Use a horizontal bar graph if the labels are too long to fit under a vertical bar graph
- Arrange and mark corresponding bars in the same way


Market Share
New Coke
Coke Classic
Percent
20
15
10
5
0
1985
1989
1985
1989
Market Share
New Coke
Coke Classic
Percent
20
15
10
5
0
1985
1989
1985
1989

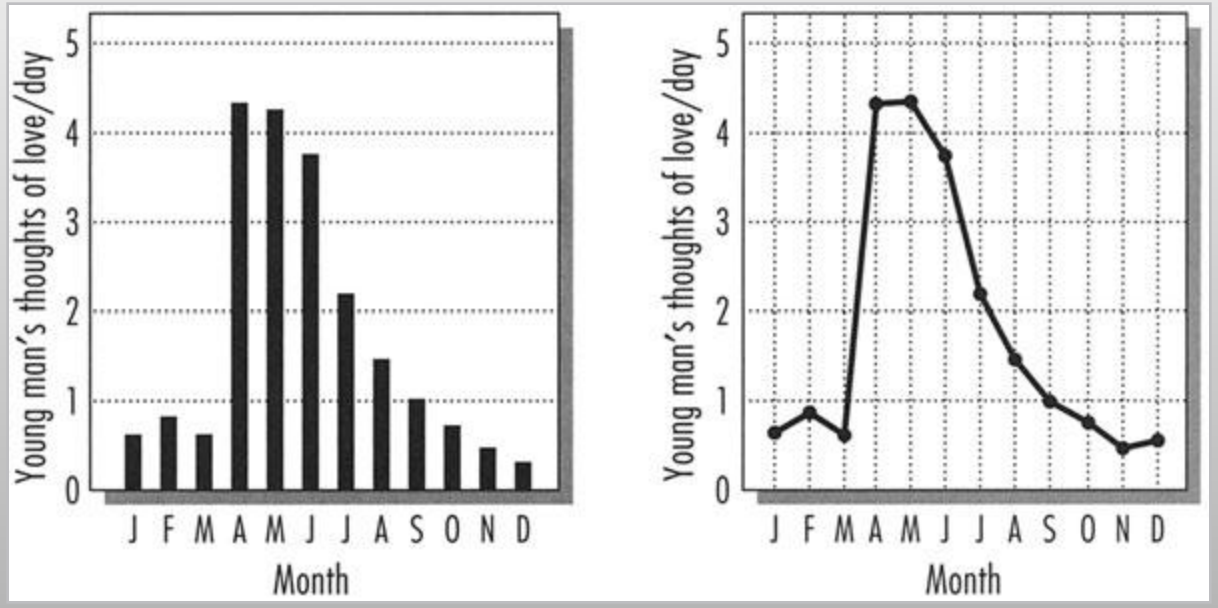

Registered Voters (1,000s), 1994
Raintree County
Paris County
50
40
30
20
10
Yng Old Men
Yng Old Women
Yng Old Women
Yng Old Men
Registered Voters (1,000s), 1994
Raintree County
Paris County
50
40
30
20
10
Yng Old Men
Yng Old Women
Yng Old Men
Yng Old Women

Bar charts: isotype



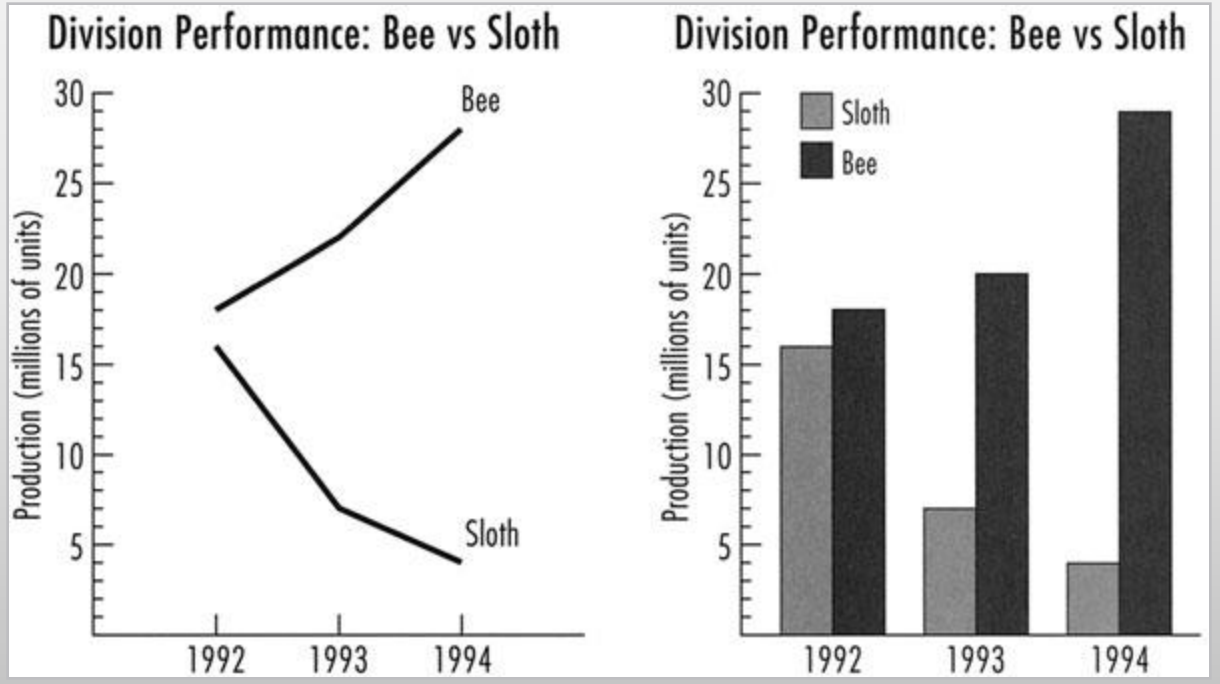
Line charts

- Use a line graph if the X axis has a continuous scale
  - e.g., time, distance, temperature
  - lines allow interpolation
- Use a line graph to display trends



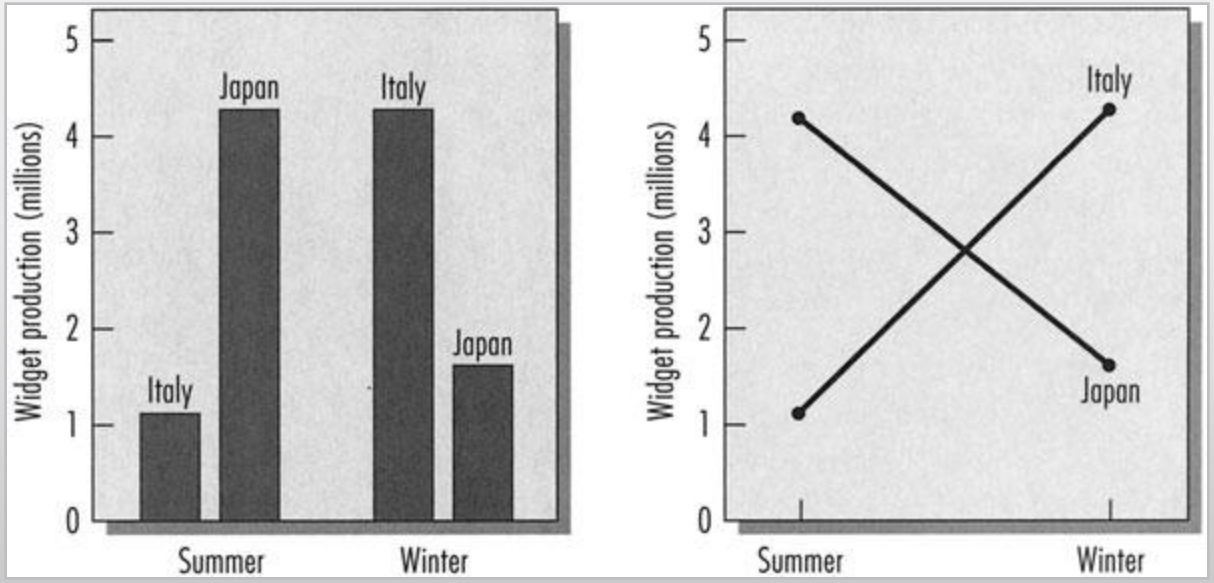
Line charts

- Don’t use a line graph to show point values



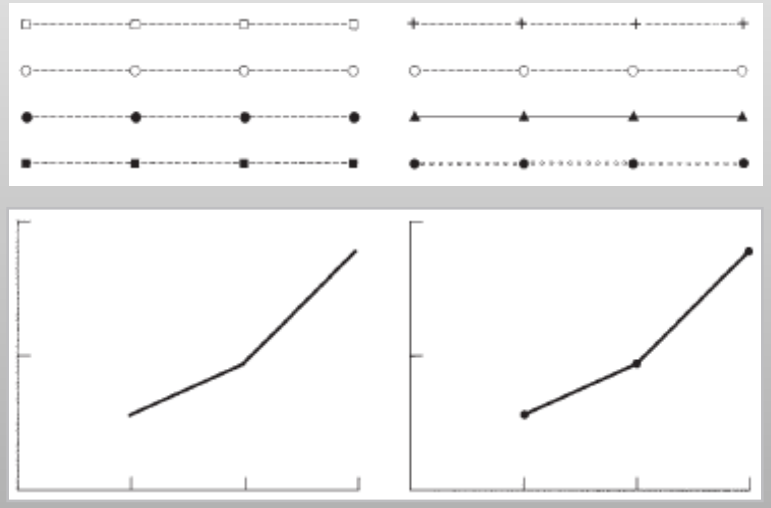
Line charts

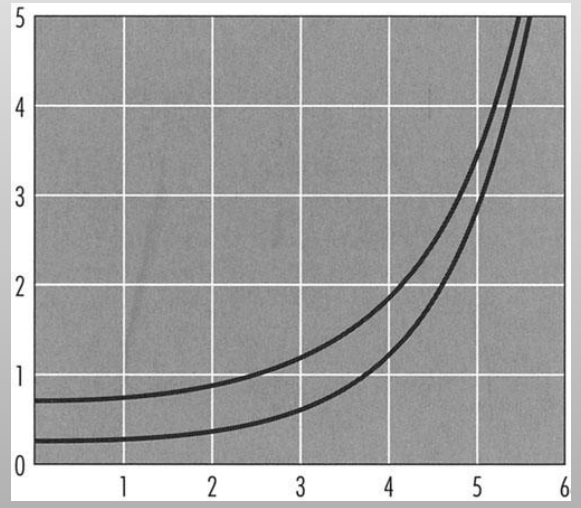
- Use a line graph to display interactions

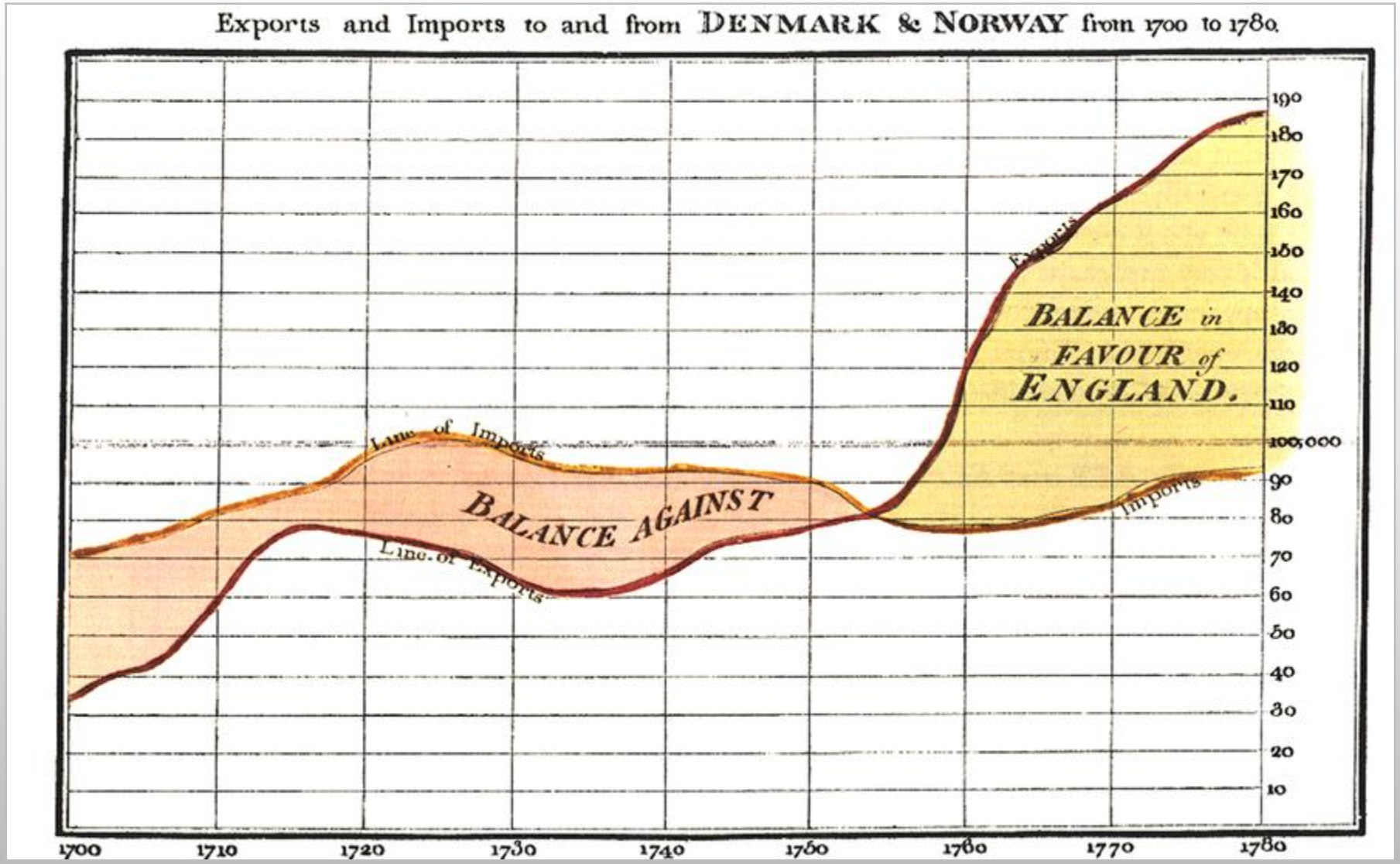


Line charts

- Formatting guidelines:
  - X and Y axes must be clearly identifiable and appropriately labeled
  - Ensure that points and lines are discriminable
  - If lines connect points, make sure the points are visible
  - Use a grid when precise values are important

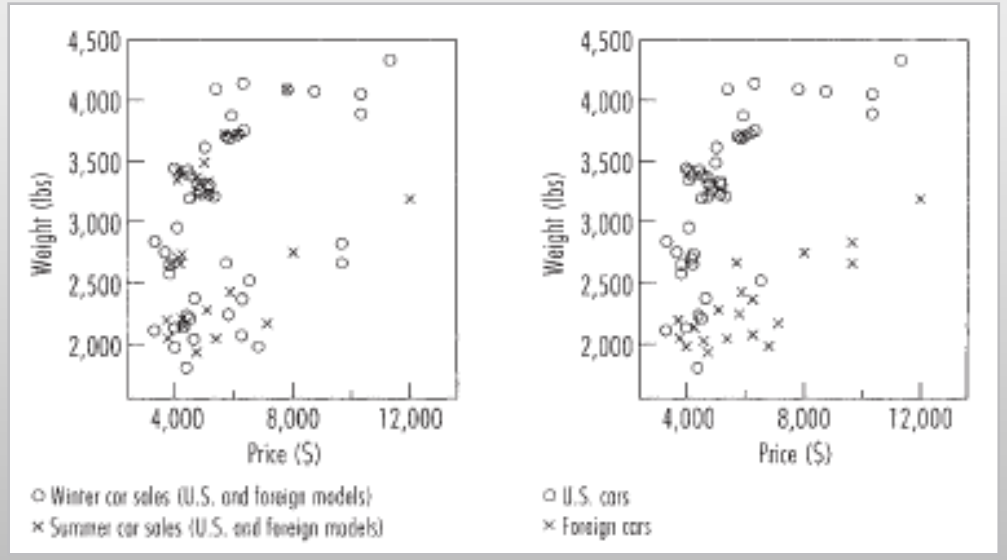





Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.
BALANCE in FAVOUR of ENGLAND.
Exports
Line of Imports
BALANCE AGAINST
Line of Exports
Imports
190
180
170
160
150
140
130
120
110
100,000
90
80
70
60
50
40
30
20
10
1700
1710
1720
1730
1740
1750
1760
1770
1780

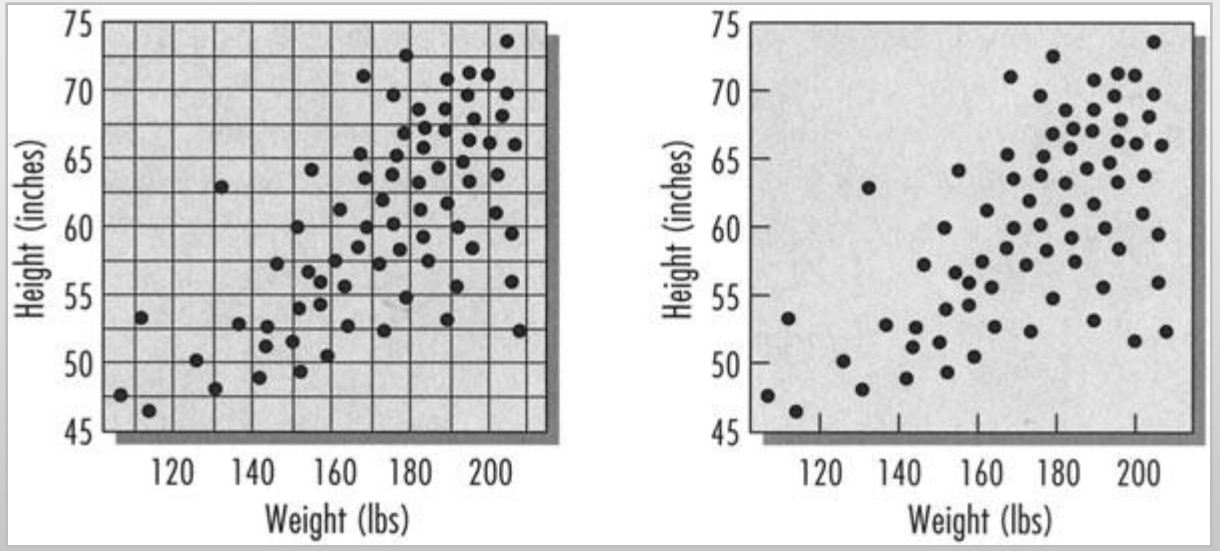
Scatterplots

- Use a scatterplot to convey an overall impression of the relationship between two variables
  - More than two becomes confusing



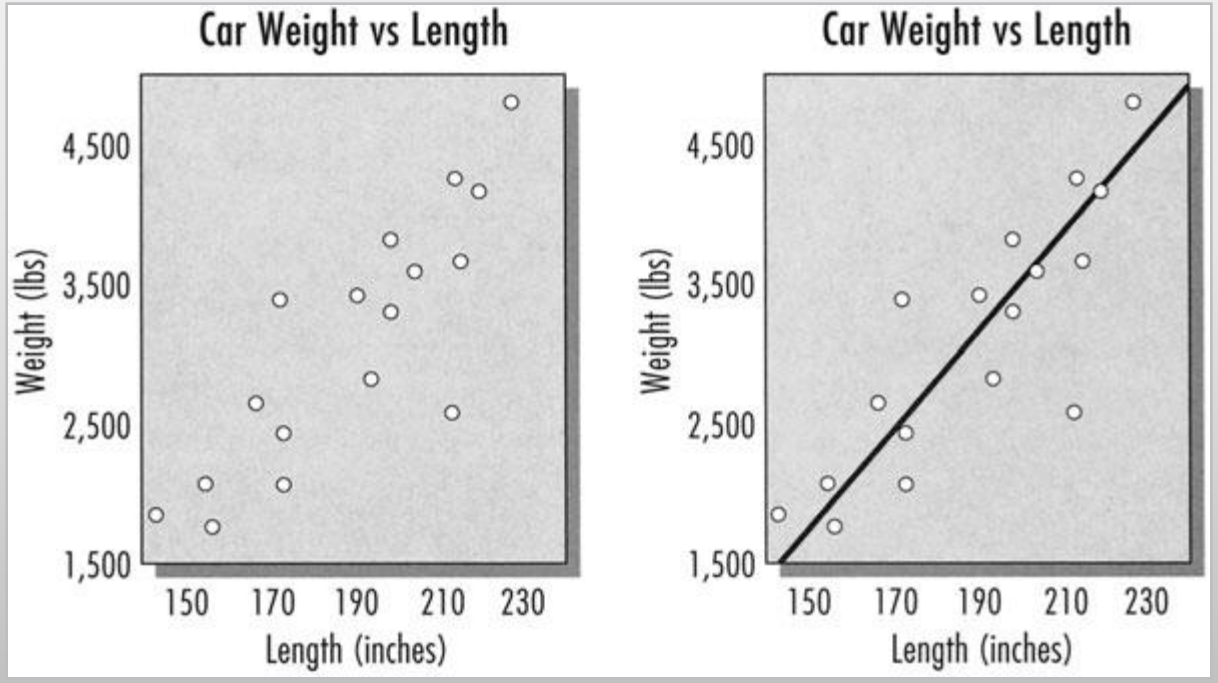
Scatterplots

- Scatterplots are not good for showing precise values
- Grids not usually used



Scatterplots

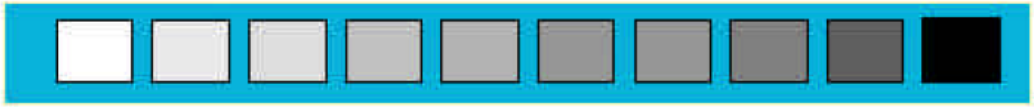
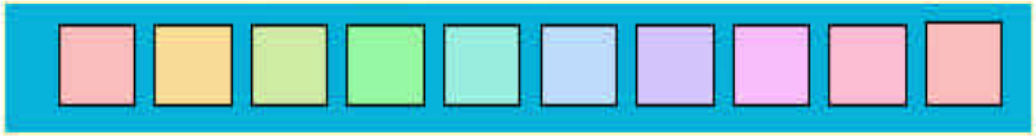
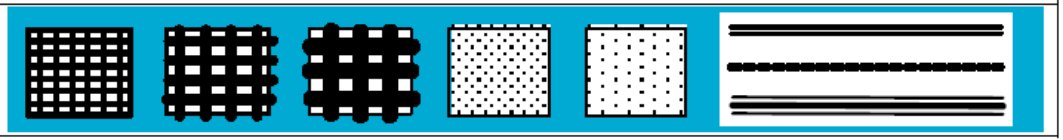
- Fit a line through a scatterplot to show how closely two variables are related



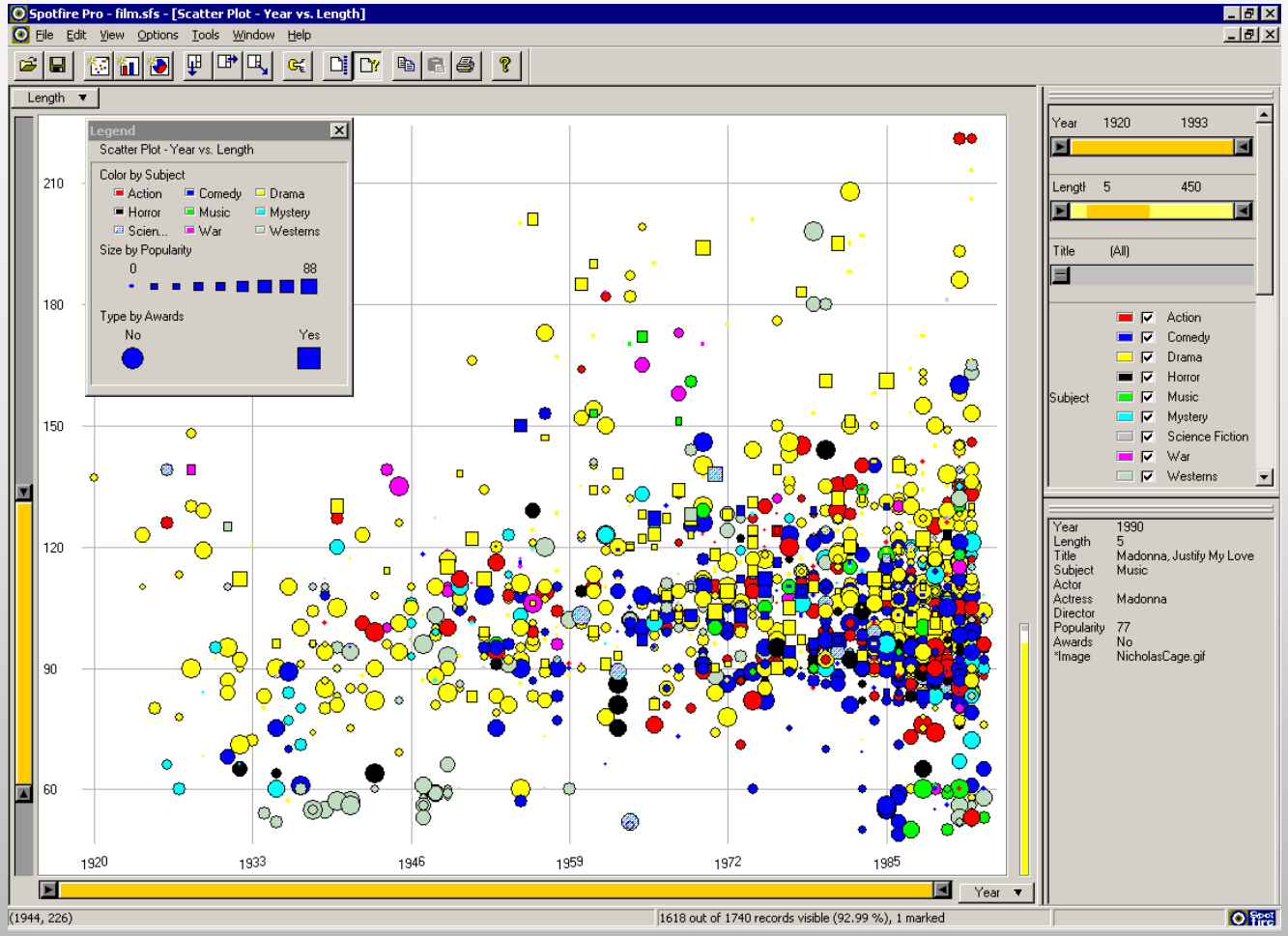
Scatterplots

- What if you want to show more attributes?
- Change the appearance of an individual mark
- Visual variables

Visual variables

Bertin's Original Visual Variables	
Position changes in the x, y location	
Size change in length, area or repetition	
Shape infinite number of shapes	
Value changes from light to dark	
Colour changes in hue at a given value	
Orientation changes in alignment	
Texture variation in 'grain'	

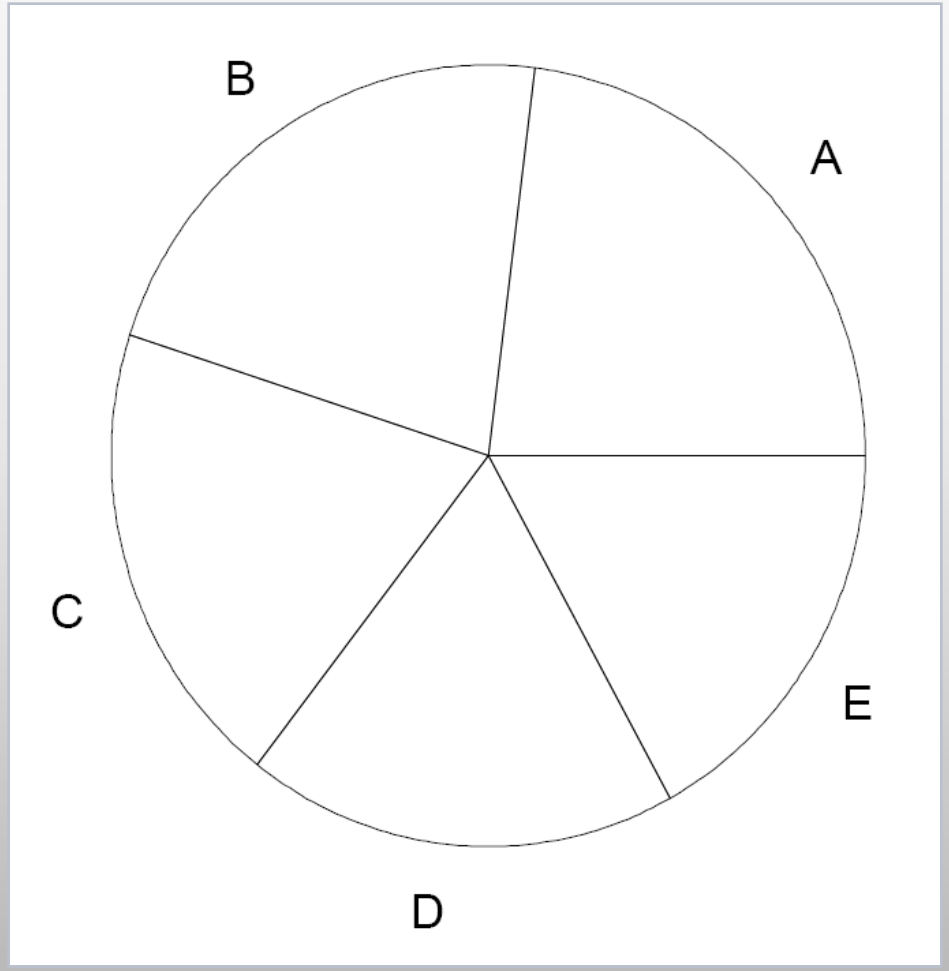
Multi-dimensional scatterplot



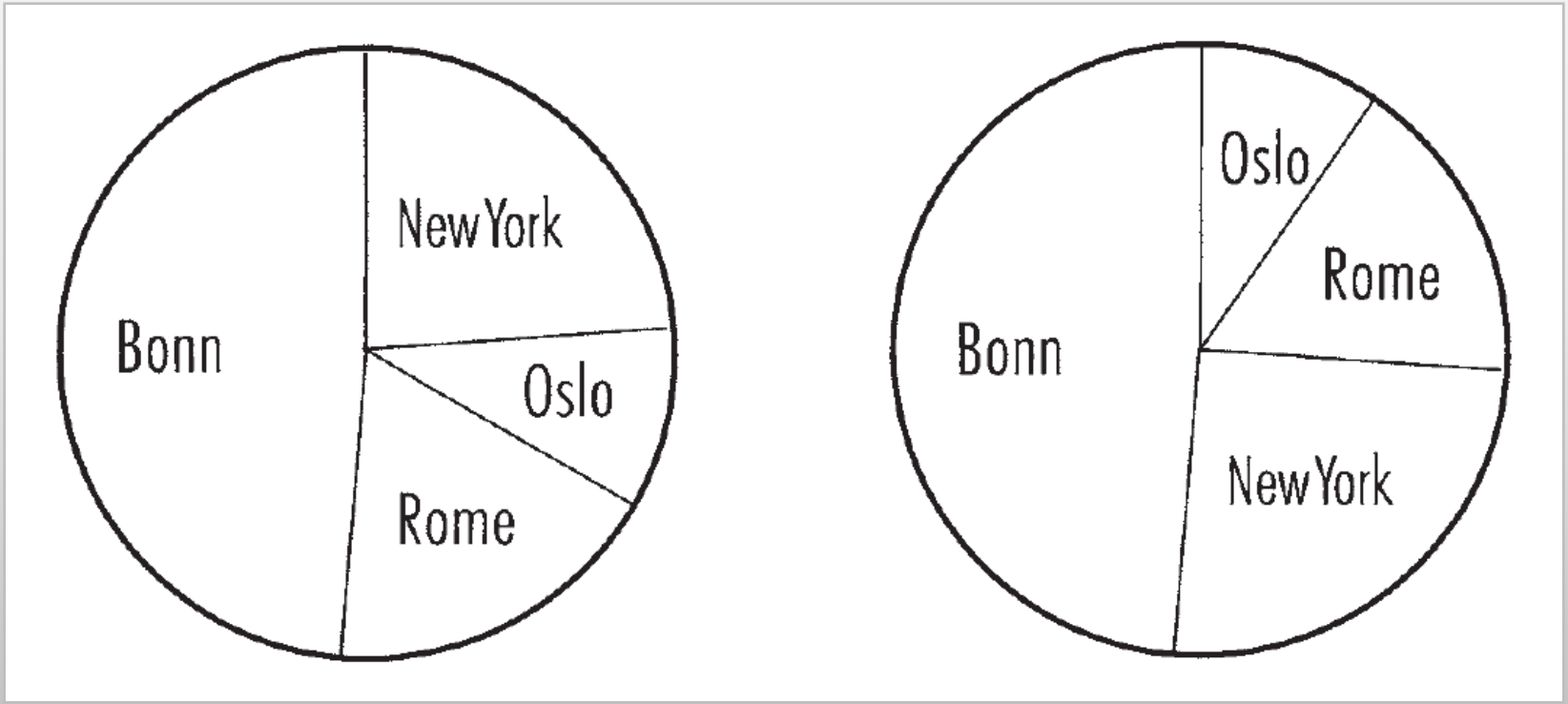
Pie charts

- Use a pie to convey approximate relative amounts
- Label wedges if precise values are important
- Arrange wedges in order of size
  - unless the data demands a different organization

Which is the largest?



Clockwise ordering

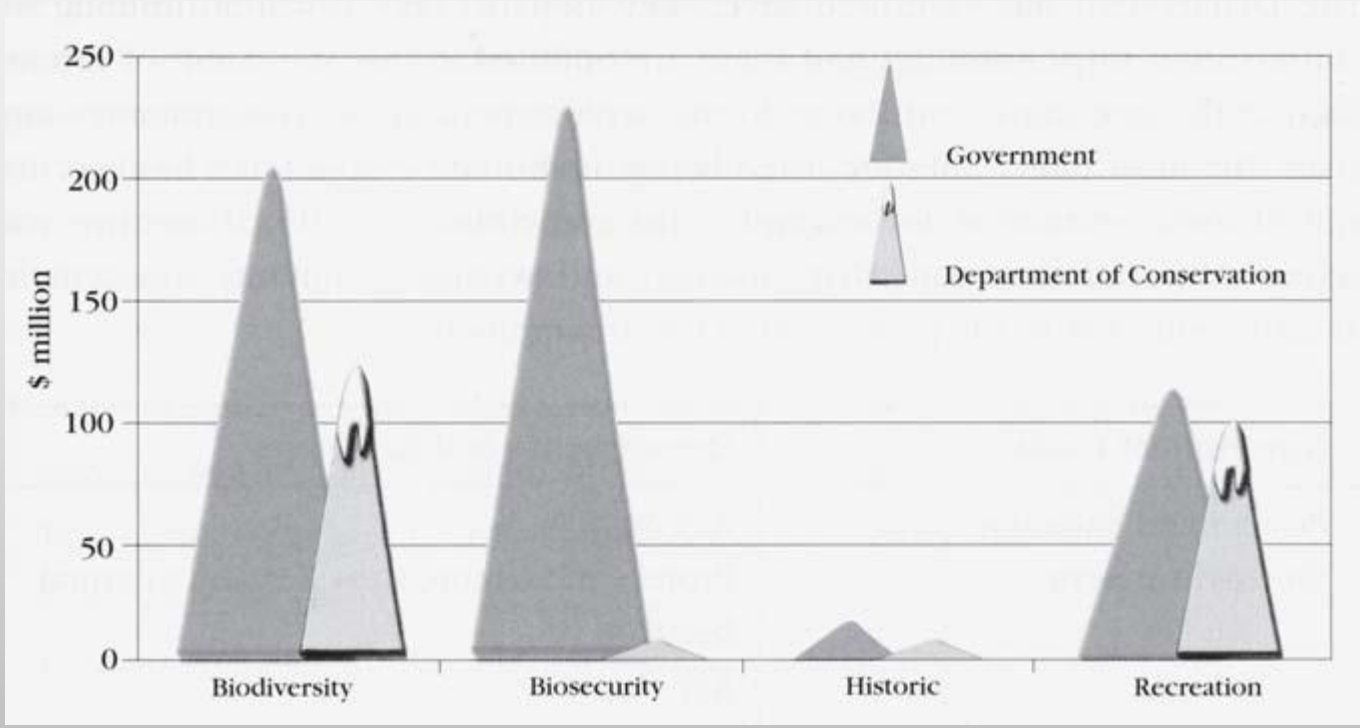


Lie factors

- Are you misleading the reader?

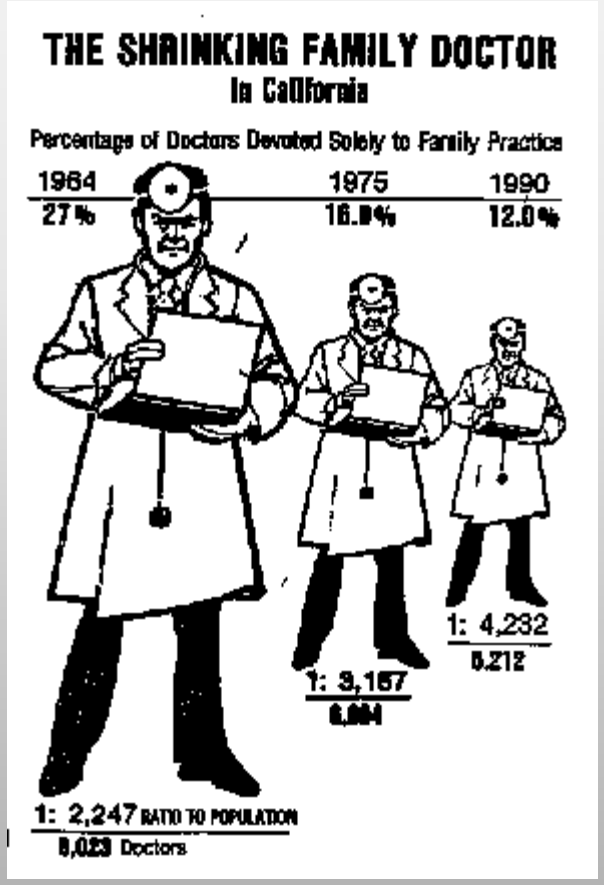
Lie factors

- Height or area? (or volume?)



Lie factors

- Height or area? (or volume?)



IMMIGRATION
1911 to 1921

ONTARIO
652,042

QUEBEC
338,985

ALBERTA
286,691

BRITISH COLUMBIA AND YUKON
262,257

SASKATCHEWAN
253,728

MANITOBA
220,663

MARITIME PROVINCES 169,554

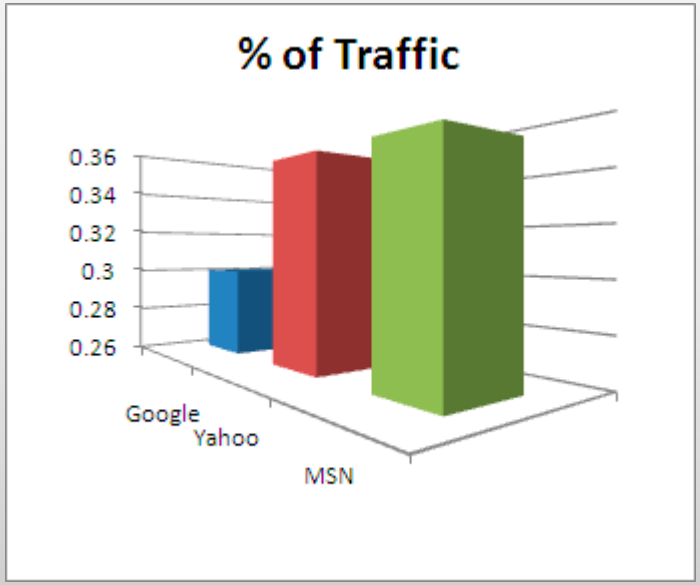
ONTARIO, the great English speaking province of Canada, owes its success to the United Empire Loyalists, those intrepid pioneers who settled along the shores of the Bay of Quinte, and the Niagara River. They were Ontario's first settlers and they built up settlements which today are thriving communities and their offspring are still carrying on.

That Ontario offers farmers and farm laborers greater opportunities than can be found in any other province of Canada, and is recognized by the incoming settler, is proved by immigration returns. During the period 1911 to 1921, 652,000 immigrants settled in Ontario, a number approximately one hundred per cent. (100 %) greater than that of any other province and representing thirty-seven per cent. (37%) of the total number of immigrants entering Canada during that period.

Ontario is the ideal province for the home seeker; its farm lands possess a marvelously rich and fertile soil, are well timbered and well watered; they provide lumber for building the home, farm buildings and fences; game of all kinds frequent the forests while the numerous rivers and lakes abound in fish of every description

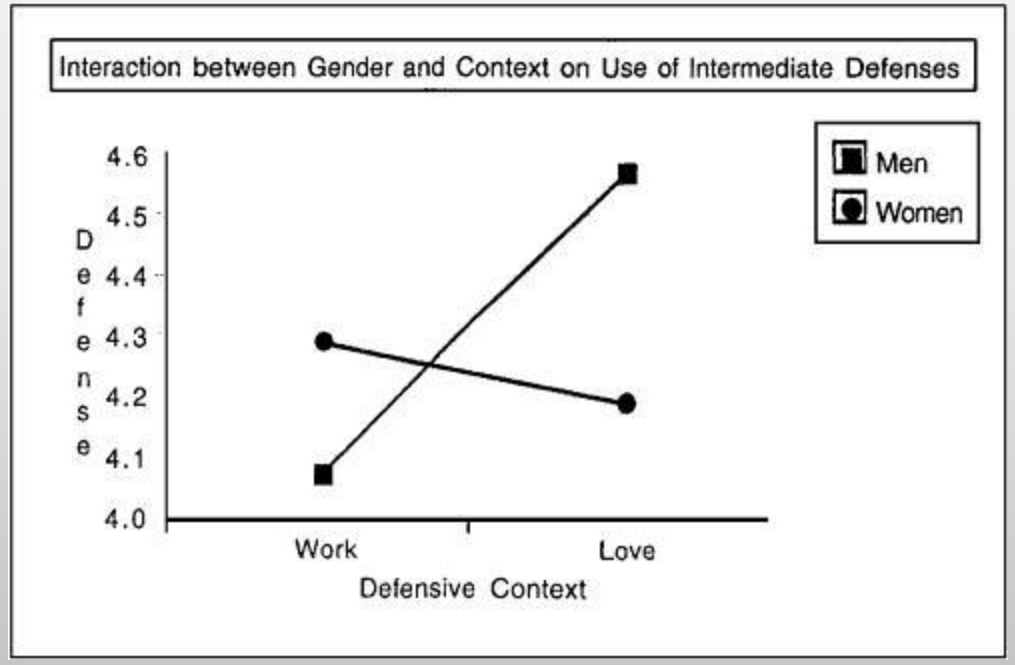
Lie factors

- Scale on Y axis



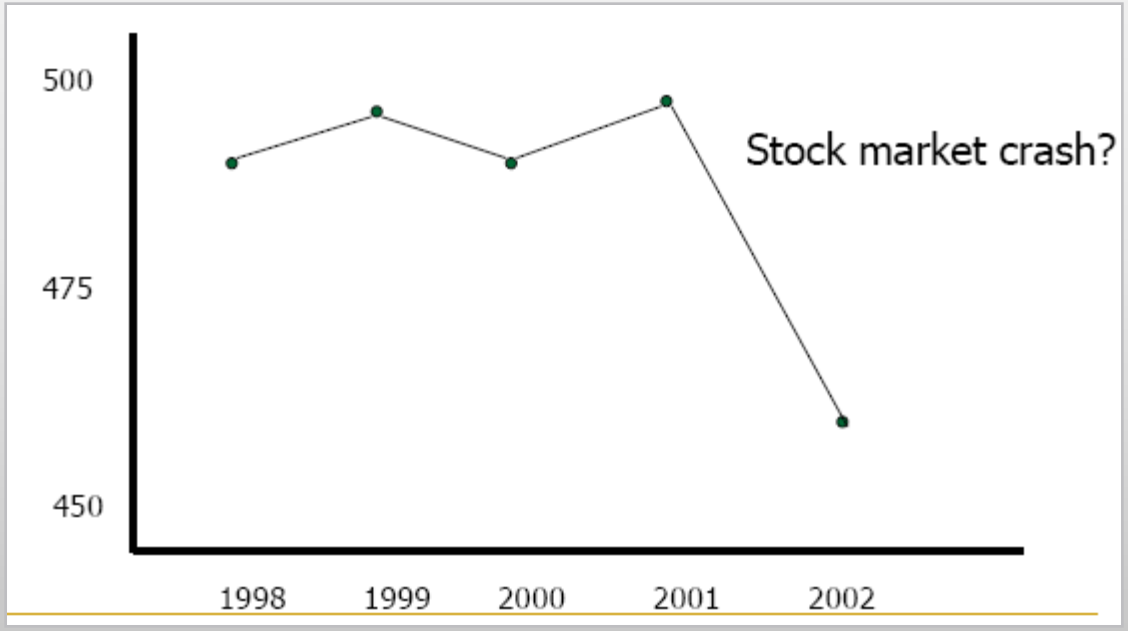
Lie factors

- Scale on Y axis



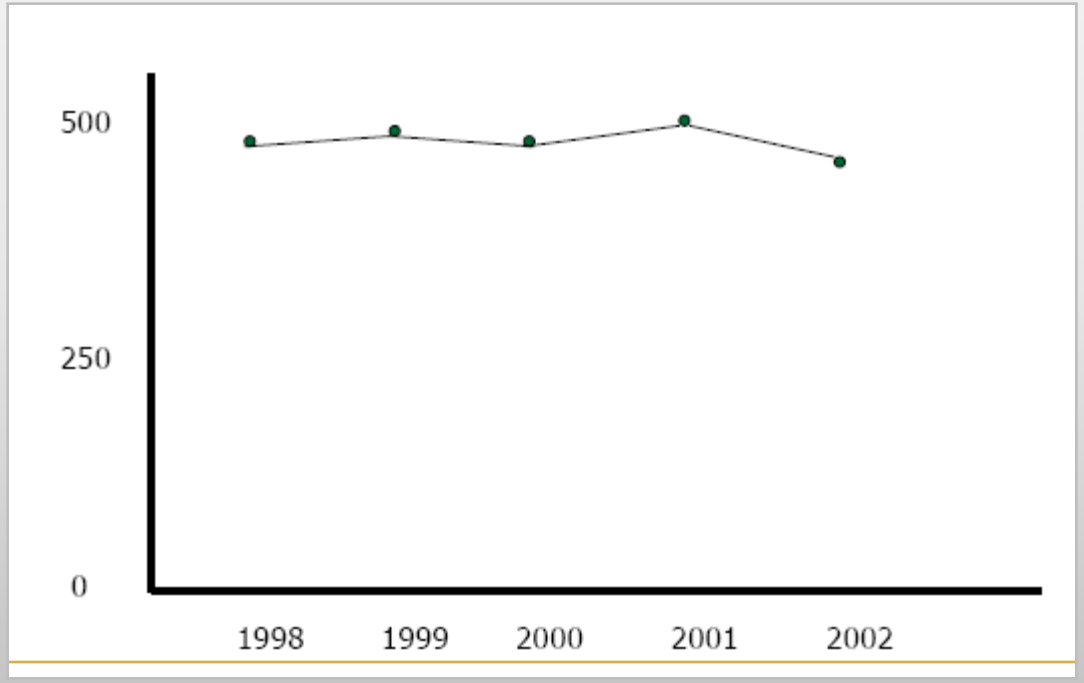
Lie factors

- Scale on Y axis



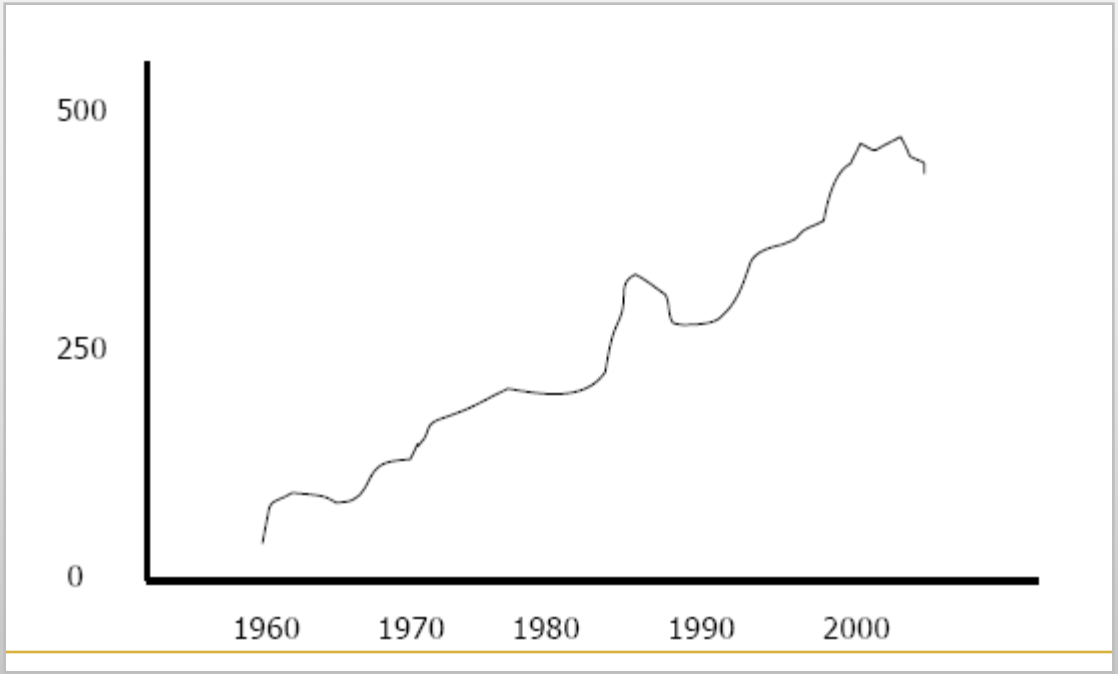
Lie factors

- Scale on Y axis



Lie factors

- Show appropriate context



Lie factors

- Switched X and Y axes

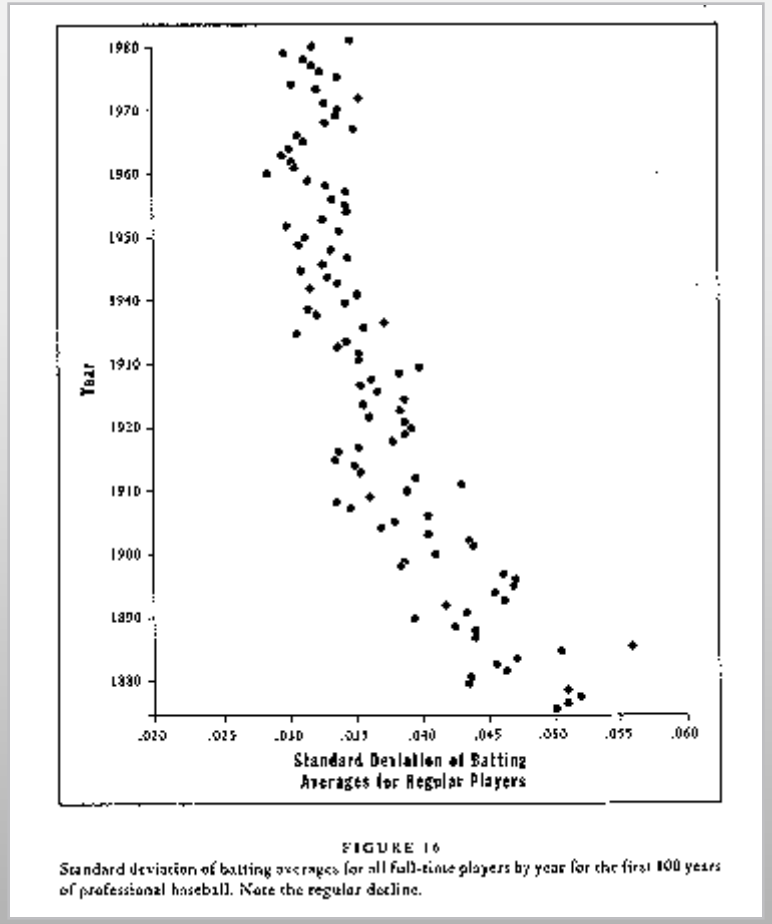
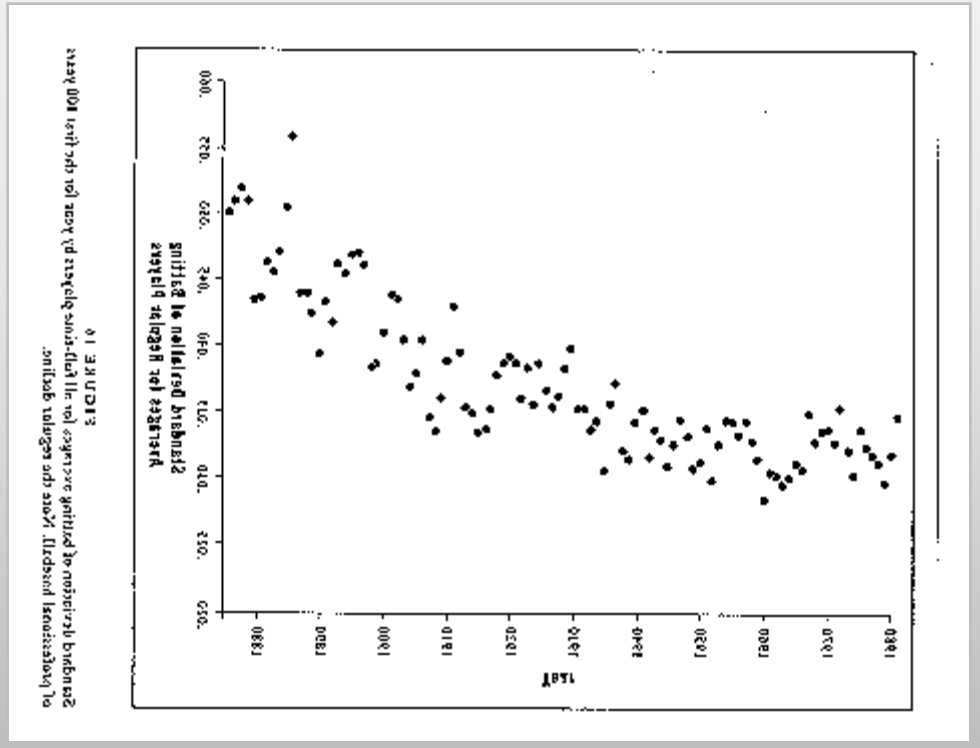


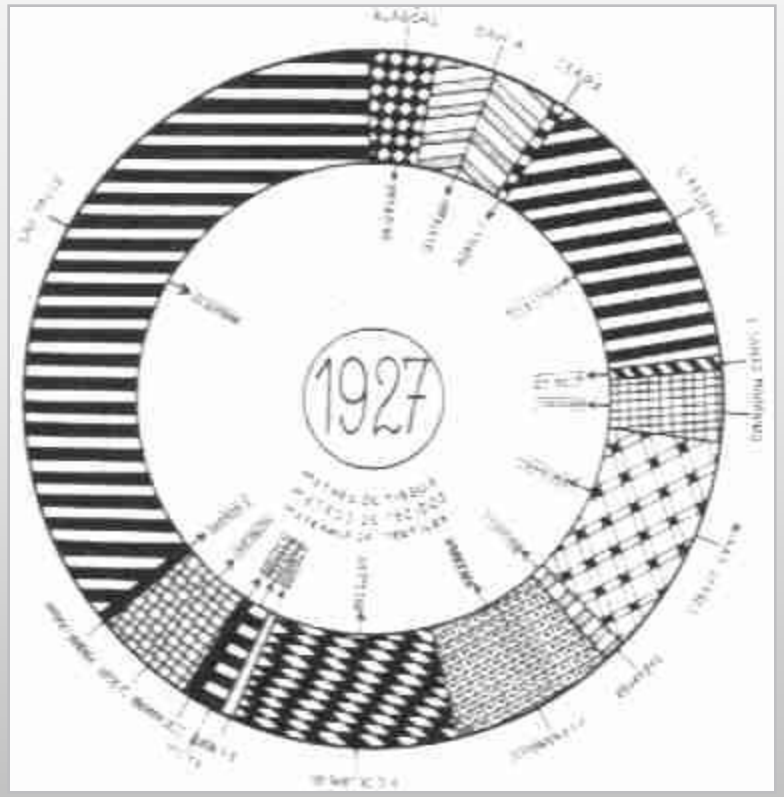
FIGURE 16
Standard deviation of batting averages for all full-time players by year for the first 100 years of professional baseball. Note the regular decline.



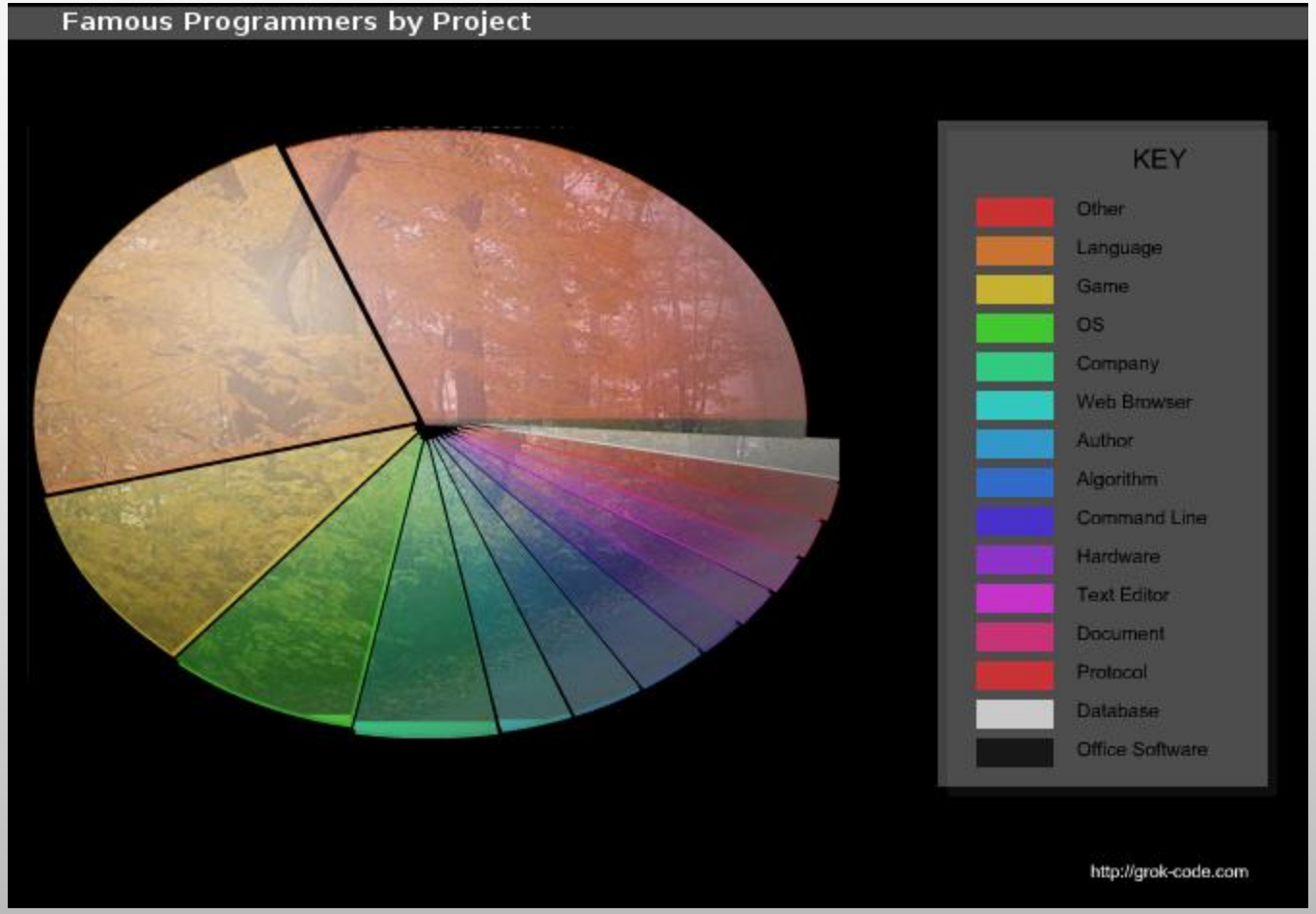
Chartjunk

- Remember that your goal is to convey meaning to your reader in support of your argument
- Chartjunk is any extra ink that does not do this

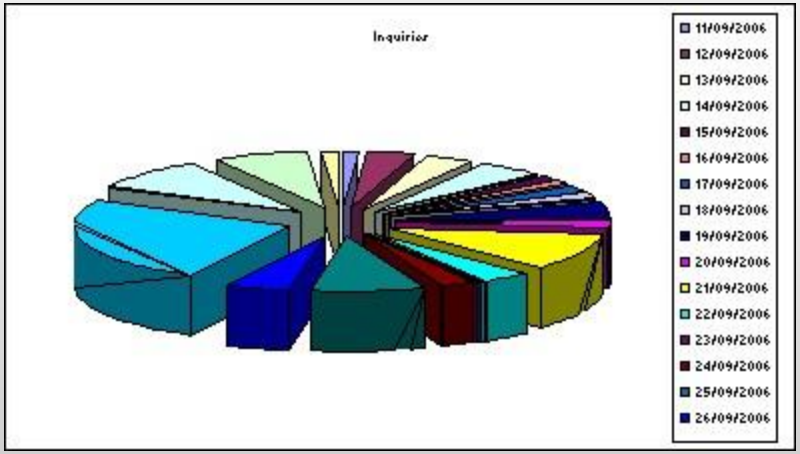
Chartjunk



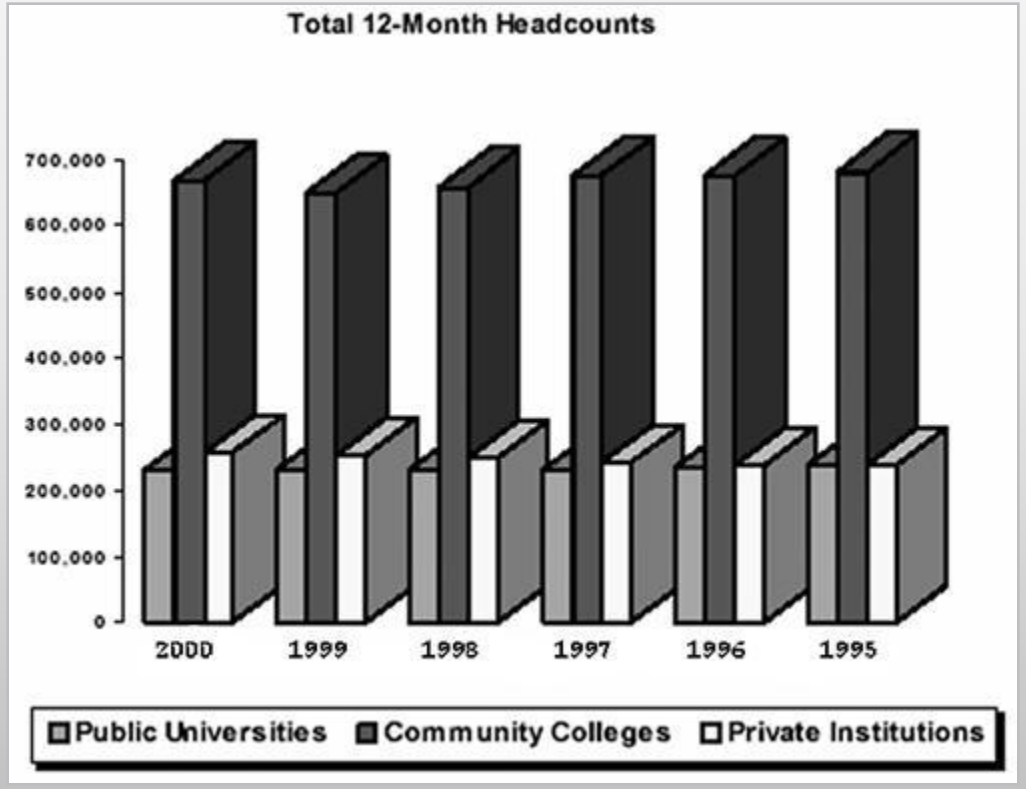
Chartjunk


Famous Programmers by Project
KEY
Other
Language
Game
OS
Company
Web Browser
Author
Algorithm
Command Line
Hardware
Text Editor
Document
Protocol
Database
Office Software
http://grok-code.com

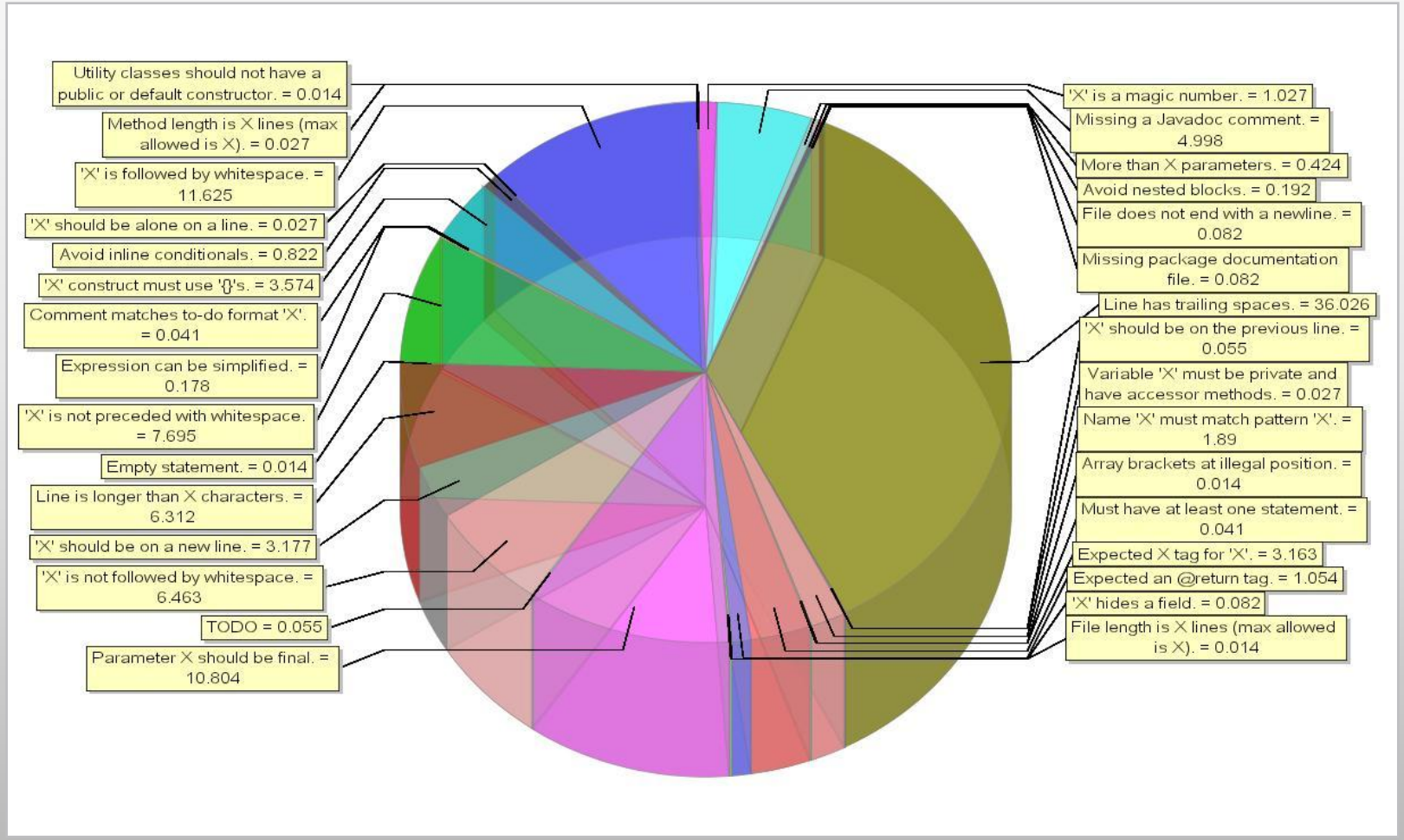
3D effects: chartjunk



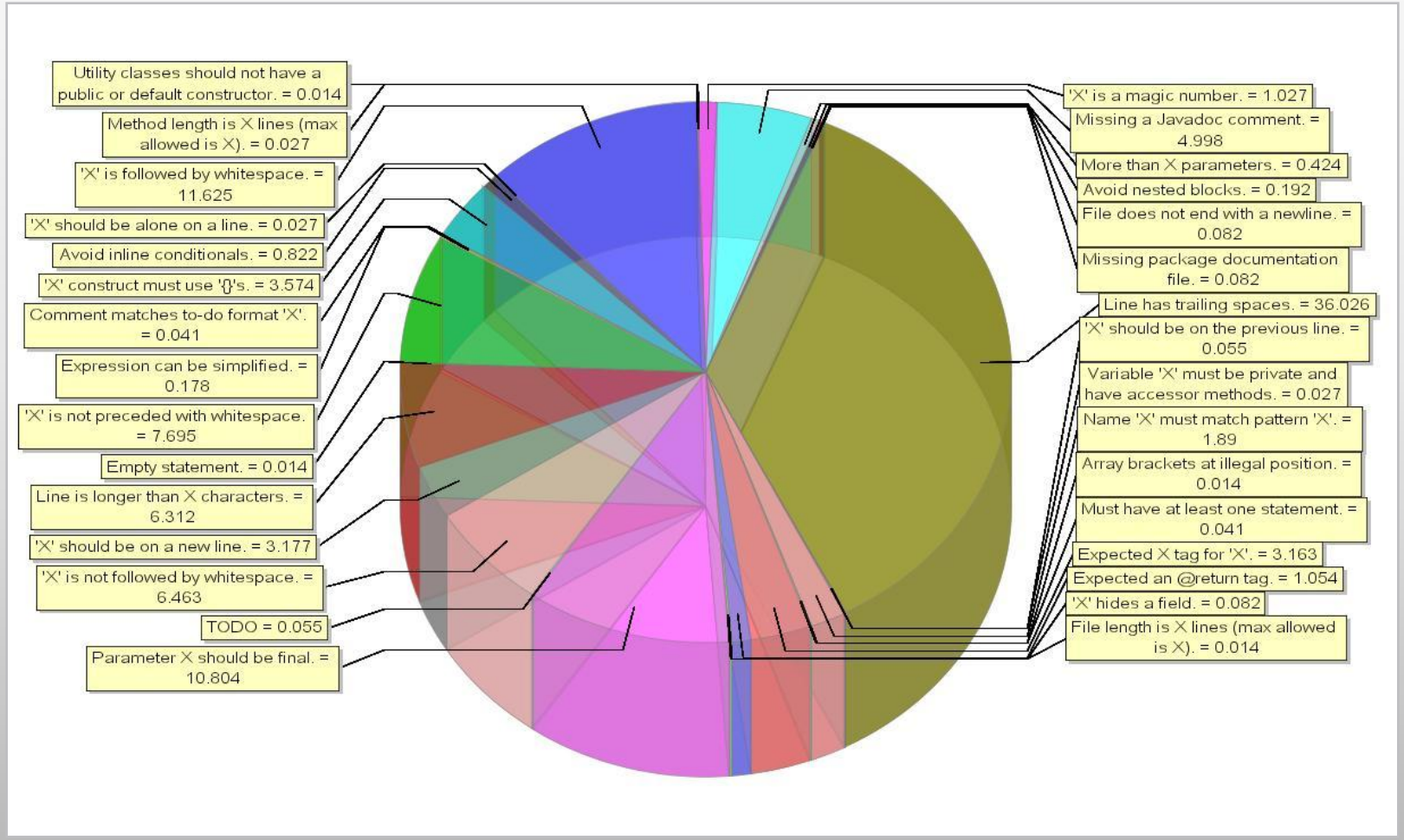
3D effects: chartjunk



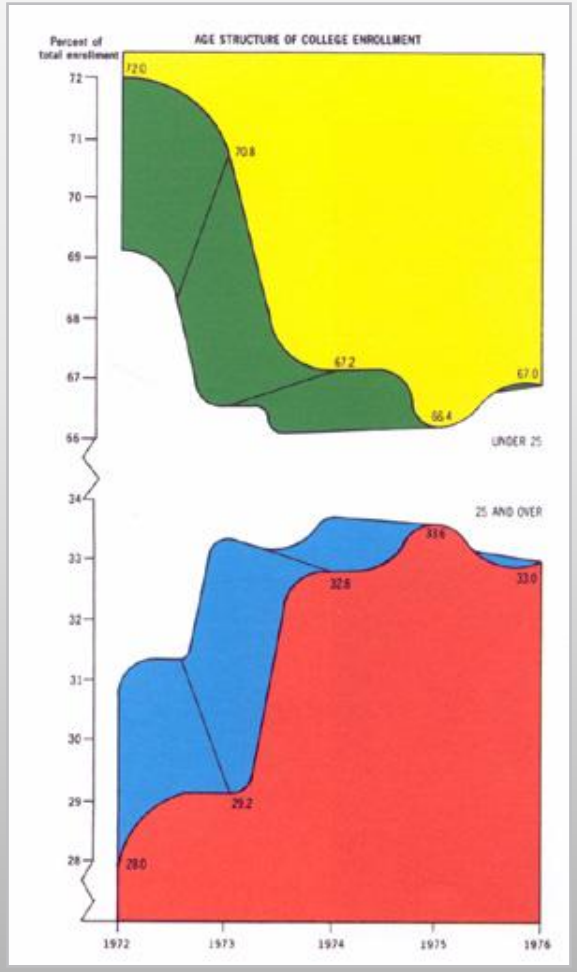
3D effects: chartjunk



3D effects: chartjunk

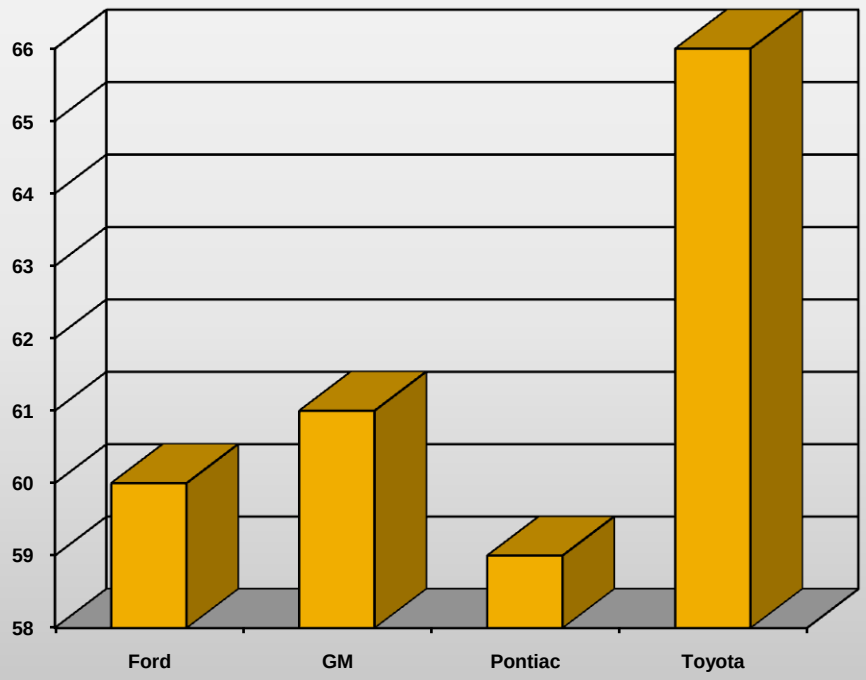


“The worst graphic ever made”

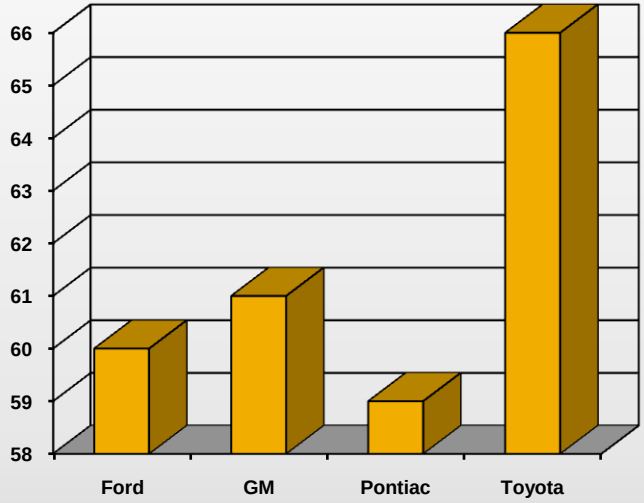


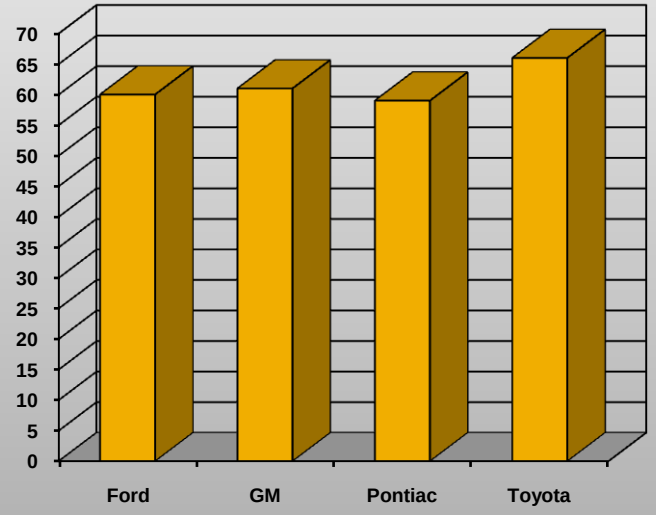
Analyse the following charts

Maintenance cost per year

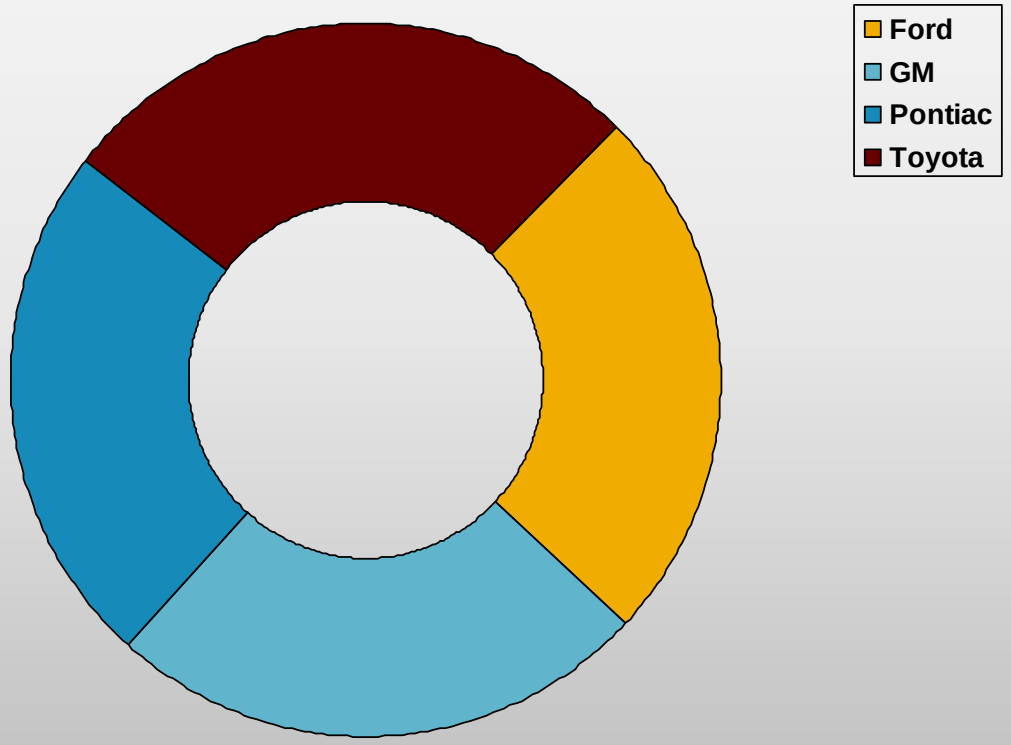


Maintenance cost per year

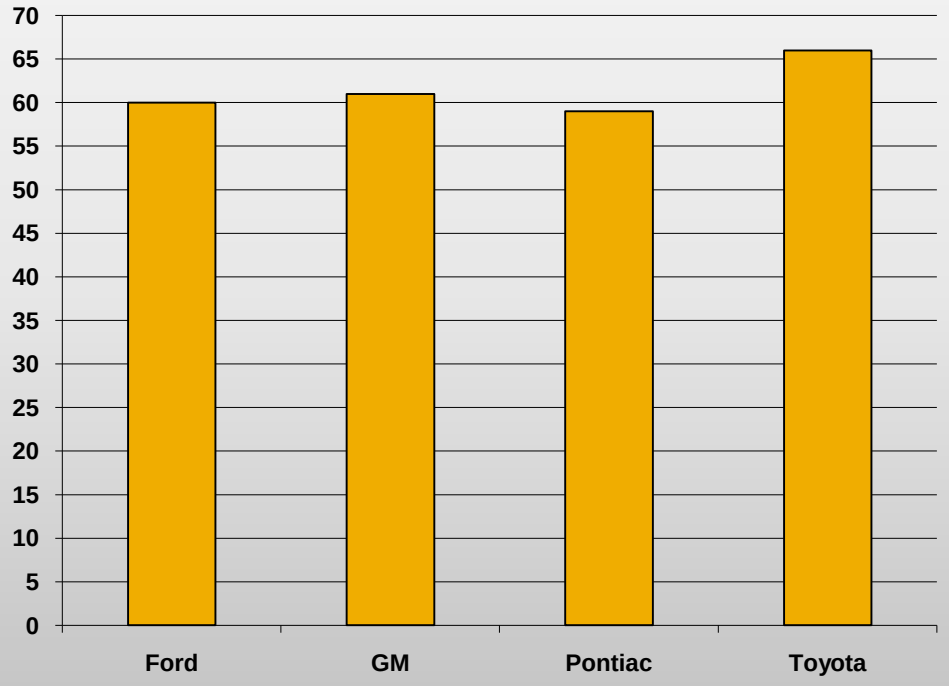




Maintenance cost per year



Maintenance cost per year



Don’t be lazy

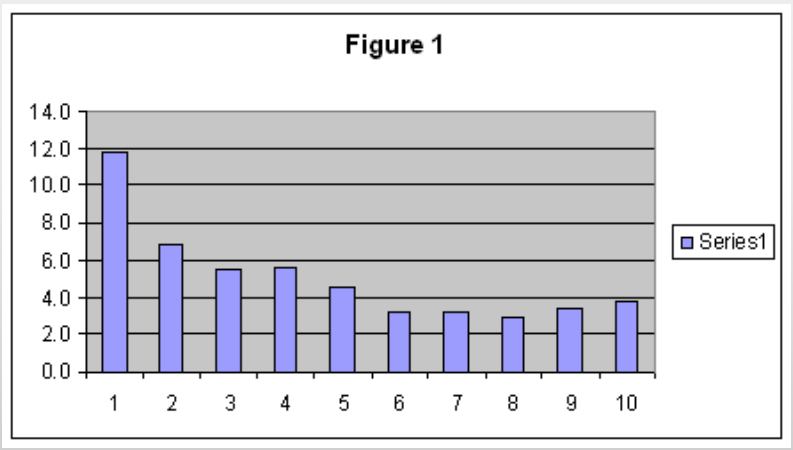


Figure 1
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
1
2
3
4
5
6
7
8
9
10
Series1

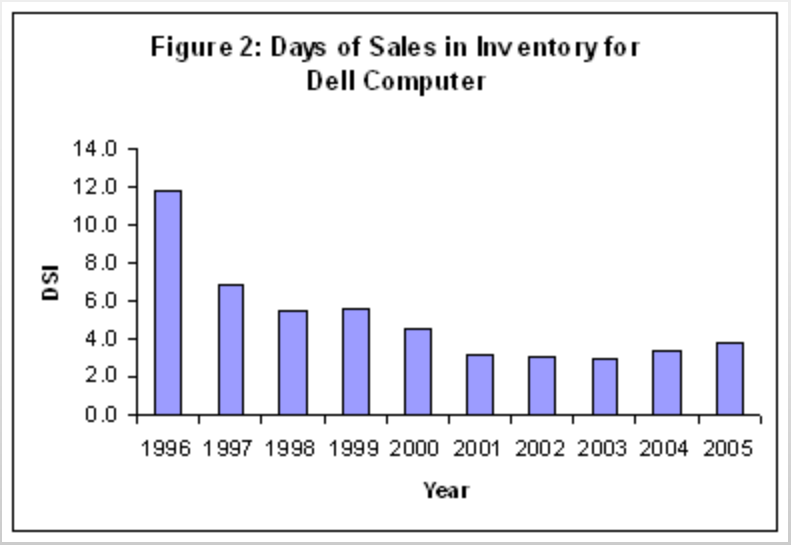


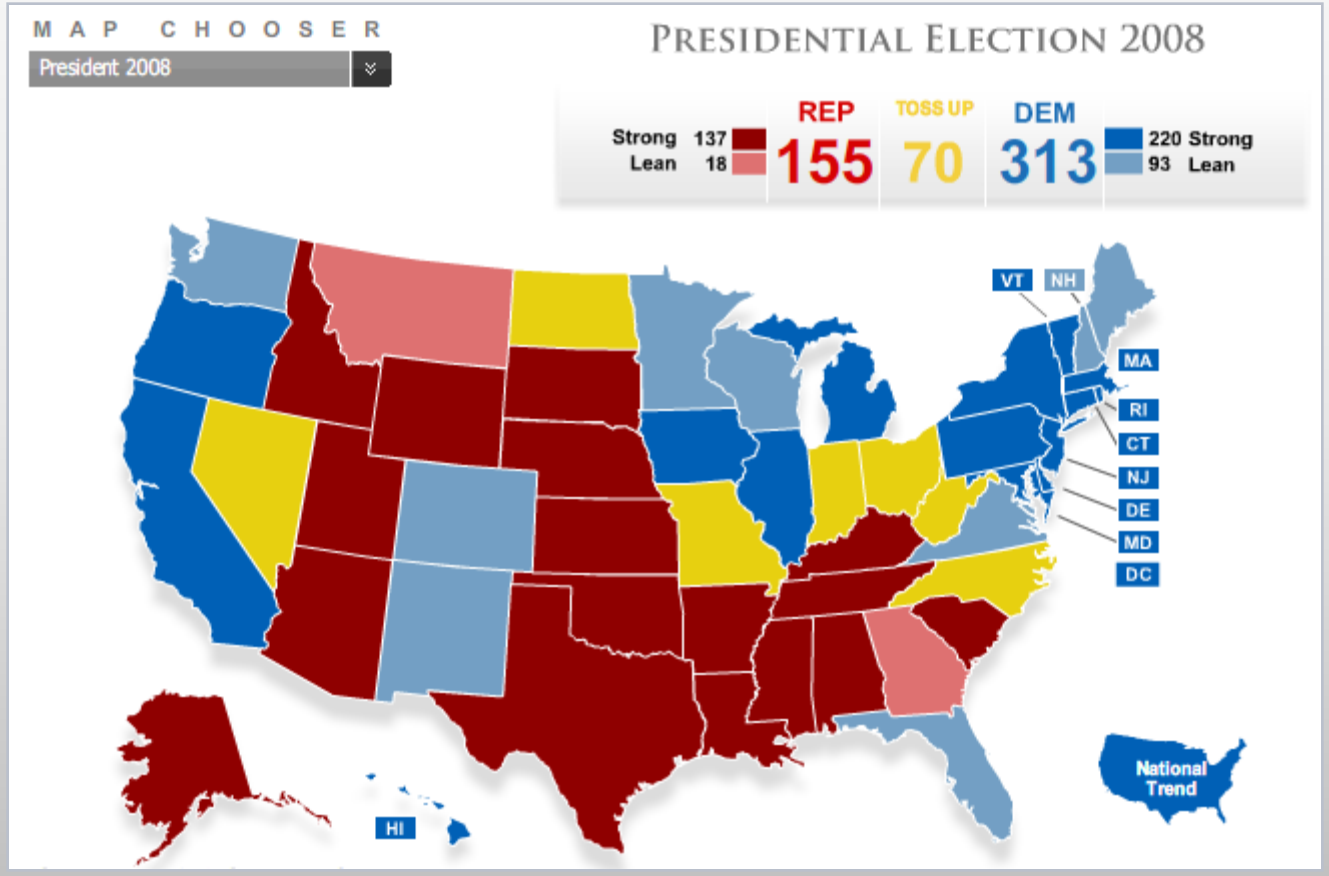
Figure 2: Days of Sales in Inventory for Dell Computer
DSI
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
Year

Maps

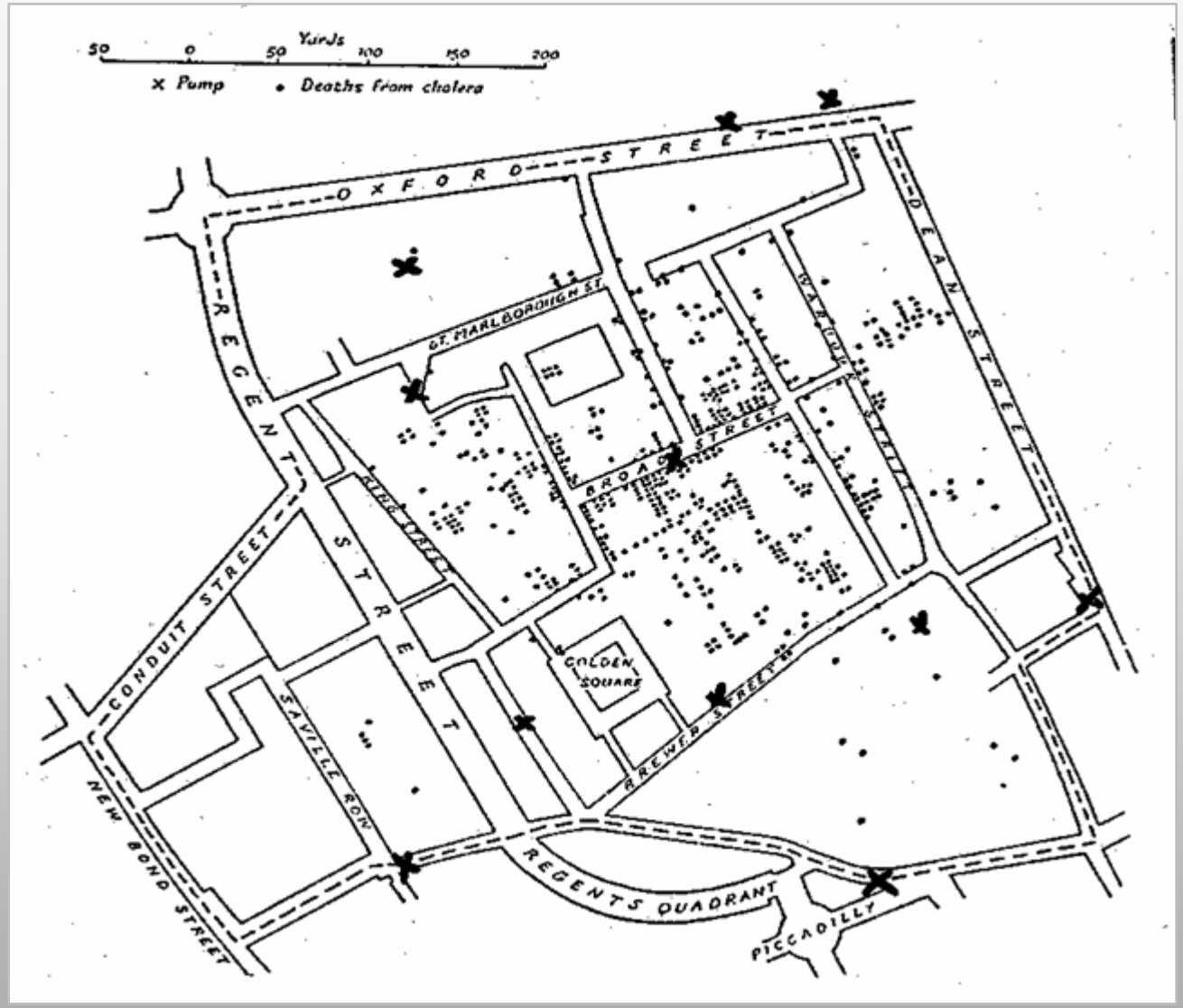
Maps

- Use a map when your data has location information
- X and Y are now used for location

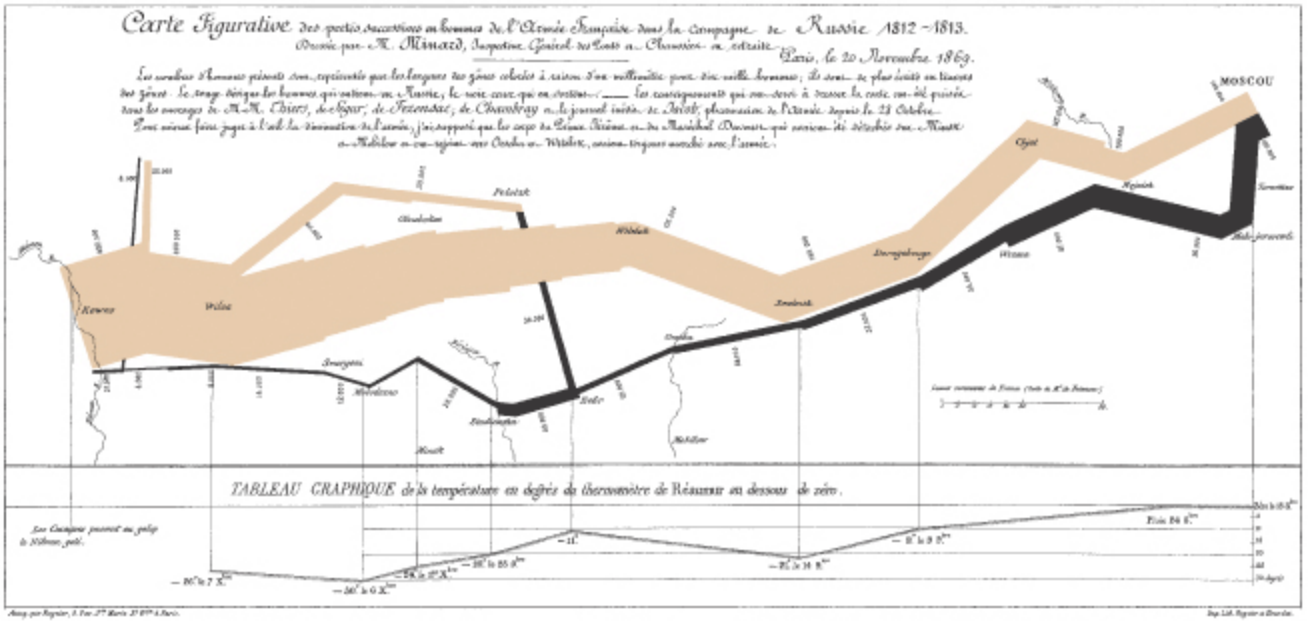
Election forecast



Cholera epidemic 1854



Napoleon's march on Moscow



Napoleon's March to Moscow The War of 1812

Charles Joseph Minard

This classic of Charles Joseph Minard (1781–1870), the French engineer, shows the terrible fate of Napoleon's army in Russia. Described by E. J. Marey as seeming to defy the pen of the historian by its brutal eloquence, this combination of data map and time-series, drawn in 1869, portrays the devastating losses suffered in Napoleon's Russian campaign of 1812. Beginning at the left on the Polish-Russian border near the Neman River, the thick band shows the size of the army (422,000 men) as it invaded Russia in June 1812. The width of the band indicates the size of the army at each place on the map. In September, the army reached Moscow, which was by then sacked and deserted, with 100,000 men. The path of Napoleon's retreat from Moscow is depicted by the darker, lower band, which is linked to a temperature scale and dates at the bottom of the chart. It was a bitterly cold winter, and many froze on the march out of Russia. As the graphic shows, the crossing of the Berezina River was a disaster, and the army finally straggled back into Poland with only 10,000 men remaining. Also shown are the movements of auxiliary troops, as they sought to protect the rear and the flank of the advancing army. Minard's graphic tells a rich, coherent story with its multivariate data, far more enlightening than just a single number bouncing along over time. Six variables are plotted: the size of the army, its location on a two-dimensional surface, direction of the army's movement, and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn.

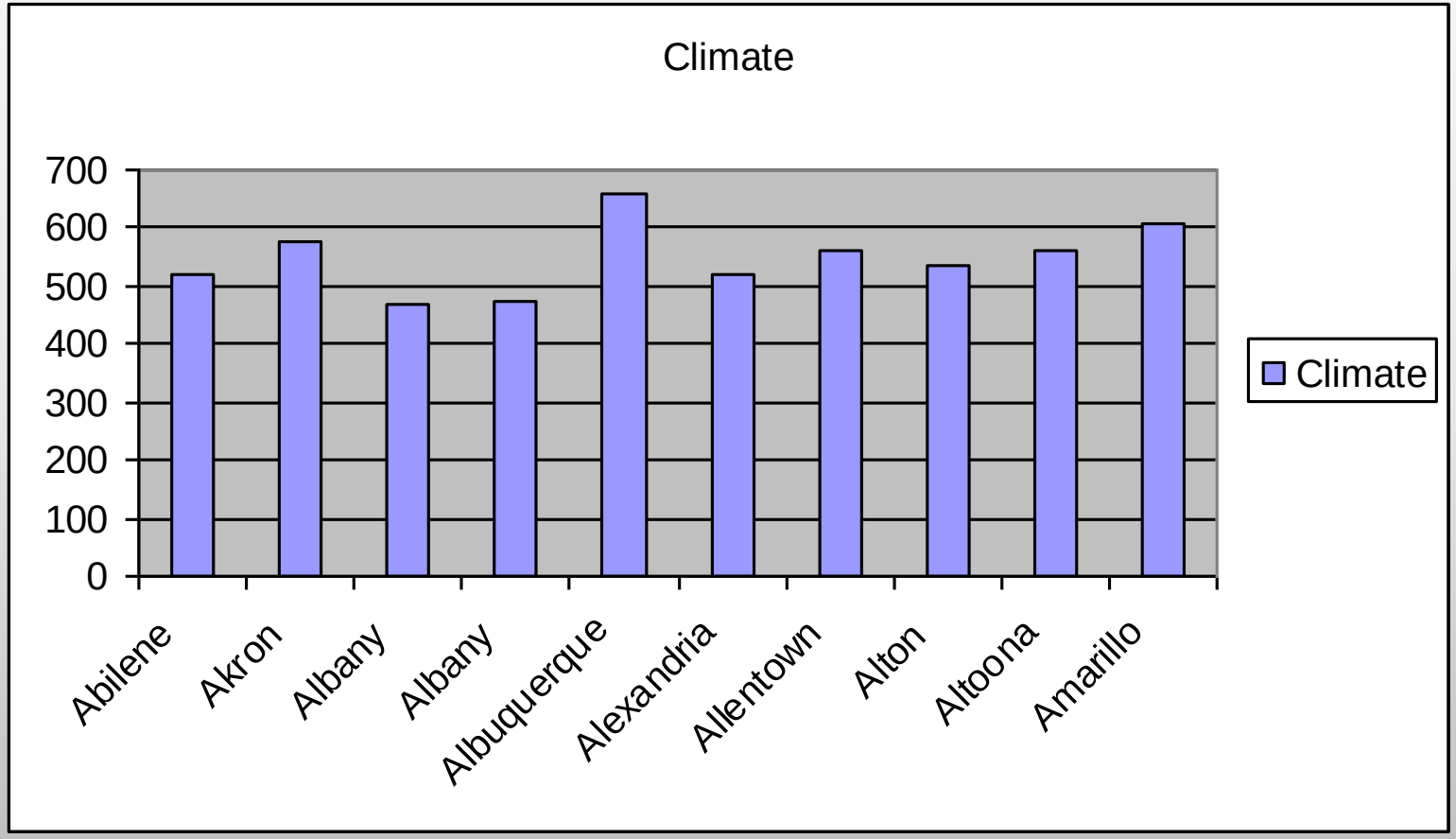
Edward R. Tufte, *The Visual Display of Quantitative Information* Graphics Press Box 430 Cheshire, Connecticut 06410

Exercise

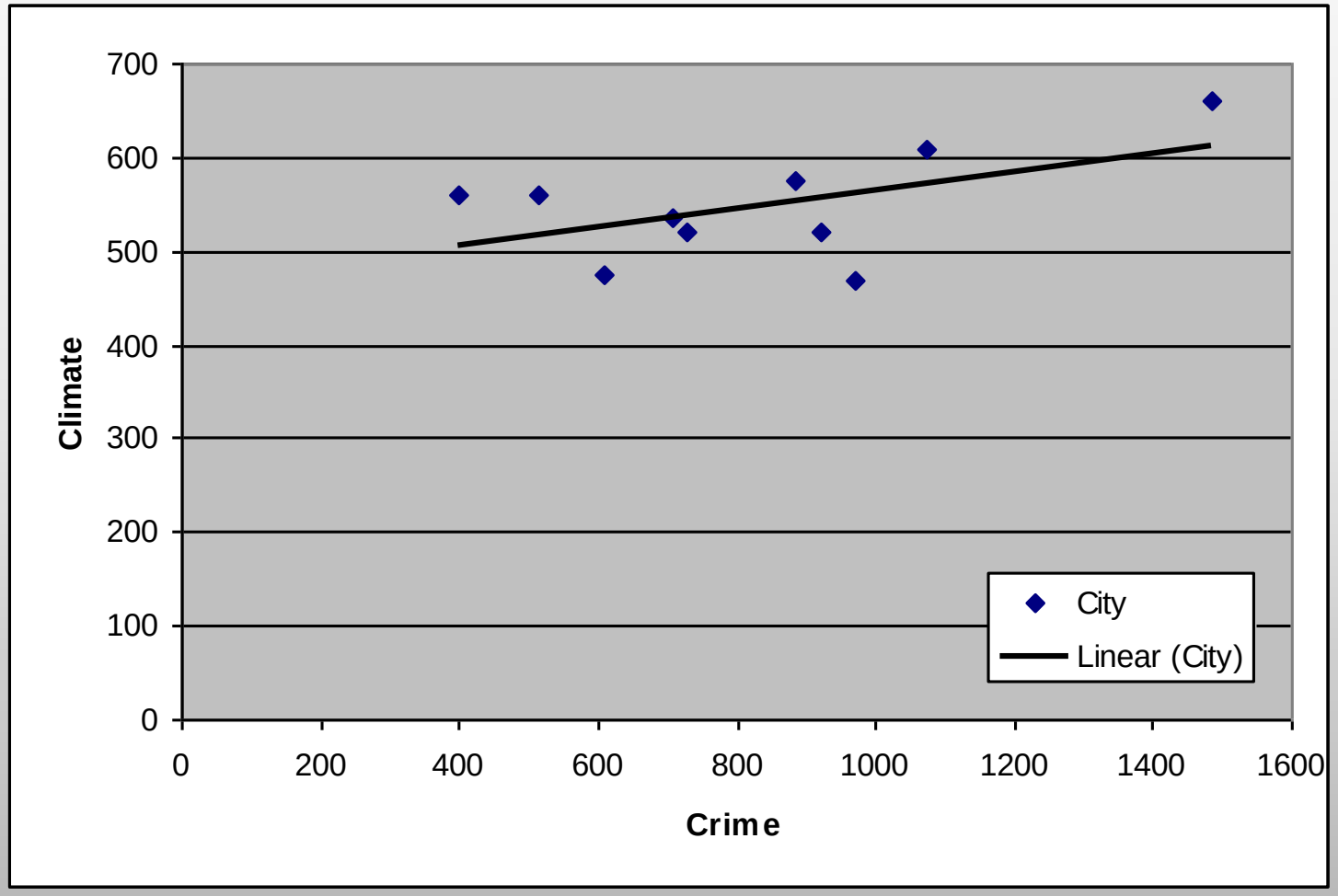
Where to live in the U.S.?

City	Climate	Housing	HlthCare	Crime	Transp	Educ	Arts	Recreat	Econ	Long	Lat	Pop
Abilene	521	6200	237	923	4031	2757	996	1405	7633	-99.689	32.559	110932
Akron	575	8138	1656	886	4883	2438	5564	2632	4350	-81.518	41.085	660328
Albany	468	7339	618	970	2531	2560	237	859	5250	-84.158	31.575	112402
Albany	476	7908	1431	610	6883	3399	4655	1617	5864	-73.7983	42.7327	835880
Albuquerque	659	8393	1853	1483	6558	3026	4496	2612	5727	-106.65	35.083	419700
Alexandria	520	5819	640	727	2444	2972	334	1018	5254	-92.453	31.302	135282
Allentown	559	8288	621	514	2881	3144	2333	1117	5097	-75.4405	40.6155	635481
Alton	537	6487	965	706	4975	2945	1487	1280	5795	-90.1615	38.794	268229
Altoona	561	6191	432	399	4246	2778	256	1210	4230	-78.395	40.515	136621
Amarillo	609	6546	669	1073	4902	2852	1235	1109	6241	-101.849	35.383	173699

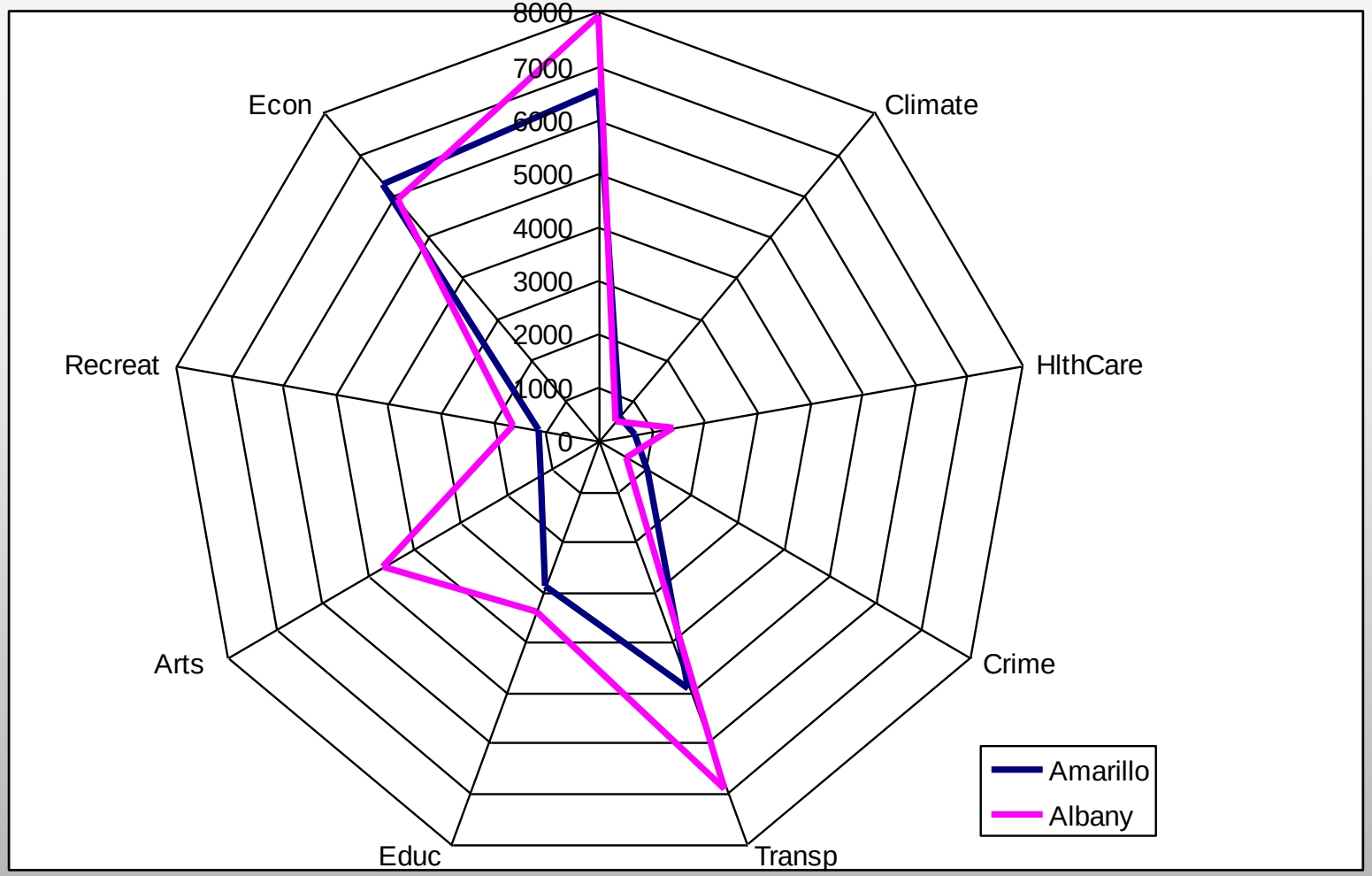
Climate


Climate
700
600
500
400
300
200
100
0
Abilene
Akron
Albany
Albany
Albuquerque
Alexandria
Allentown
Alton
Altoona
Amarillo
Climate

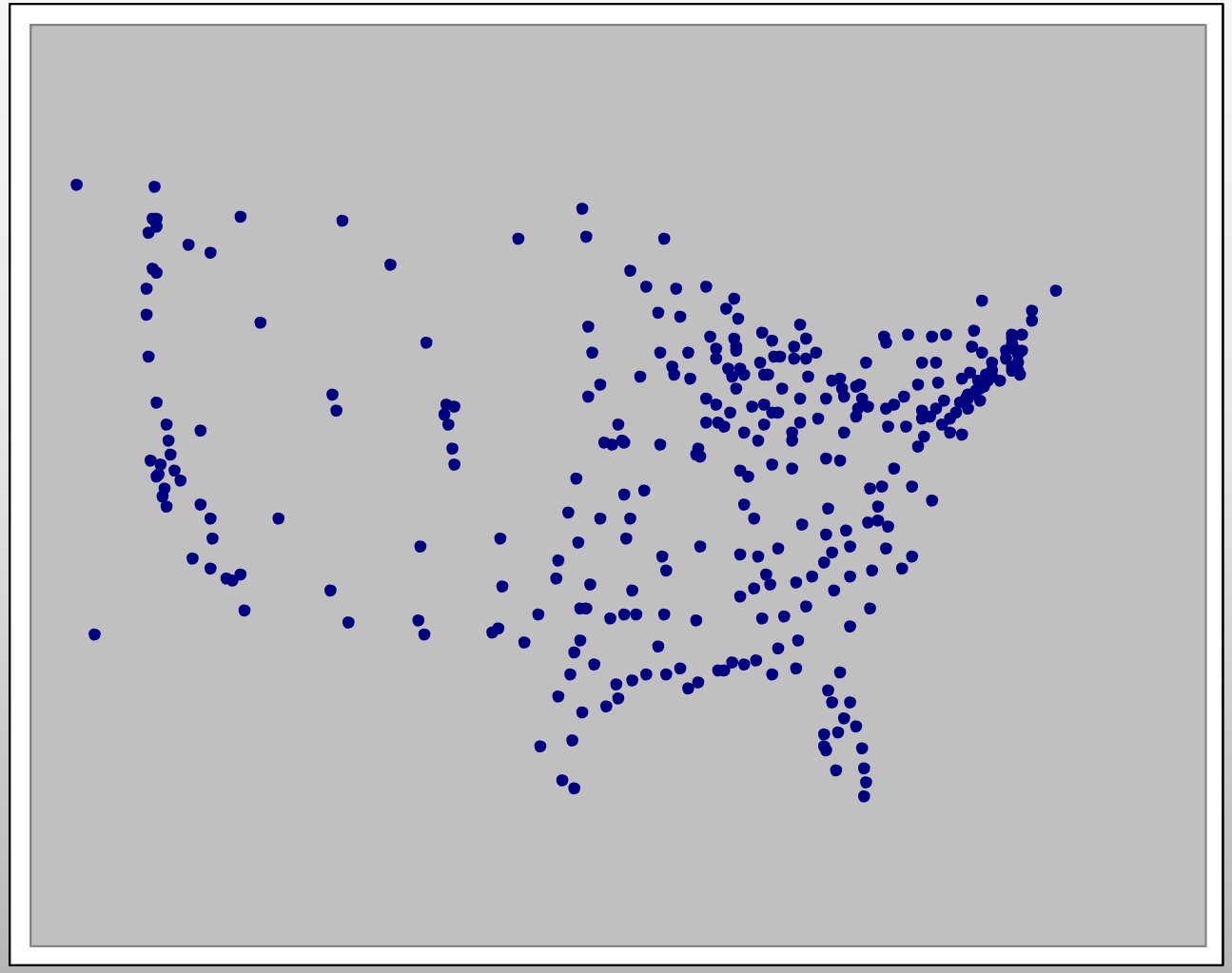
Crime vs. climate



Multiple attributes – radar plot



Map



Going further

- If these basic types of charts don't fit your need, there are many resources to help you find out more
- Edward Tufte, *The Visual Display of Quantitative Information*
- Stephen Kosslyn, *Elements of graph design*
- Robert Spence, *Information Visualization*
- Tools: www.infovis-wiki.net/

Going further