Creating a healthier world.

Tabular and Graphical Presentation of Data

Epidemiology Methods II

S. Das Pamnani, MBBS, MPH

Chighaf Bakour MD, MPH

General Principles

- What is your study question? What points would you like to convey through tables and graphs?
- Know your data. What type of variables do you have? (nominal, ordinal, discrete, continuous)
- Look at distributions of all study variables
- Create as many tables and graphs as you like for your own use (background)

Typical Sequence of Analysis

1. Data Cleaning
2. Descriptive Analyses
3. Main analysis for exposure → outcome
4. Secondary analysis
5. Creating Tables and Graphs with results
6. Preparation of oral presentation or conference poster.
7. Preparation of final tables and graphs for publication (usually 2-6 for a journal article).
8. Write the final manuscript

Why Tables and Graphs?

- Summarize relevant data and results in a concise and easy to understand manner
- Save readers time and energy
- Reduce word count
- If poorly constructed, can lead to incorrect interpretation or misinterpretation of results
- Readers who go beyond the abstract of a paper are likely to examine the graphs and tables next.

TABULAR PRESENTATION OF DATA

When to Use Tables

- Written documents (reports, journal articles) typically present most results in tabular form.
- Research Posters for conferences.
- More concise format than graphs.
- In oral presentations, only VERY simple tables should be presented.

Components of a Good Table

- Title
- Ordering and Organization of the Results
- Column Headings
- Row Headings
- Data in the Cells
- Footnotes (internal use)
- Footnotes (external use)

Titles of Tables & Graphs

- Must be comprehensive and "stand alone".
- Clear and concise. Includes "main message"
- Describes person, place and time — what, where, and when — of the data in the table.
- Precede the title with a table number.

The Problem with Titles

- Bad Title: "Table 1. Descriptive characteristics"
- Bad Title: "Figure 1. Flowchart of study selection"
- Bad Title: "Descriptive characteristics of cases and controls"
- Bad Title: "Percent of Smokers in the Study"
- Good Title: "Percent of Study Participants Who Smoked at Baseline, FHS, 1949-1951"

More on Titles

- What comes after the "of" should be your denominator.
- Use precise language.
- Titles should be 1-2 lines long.
- Extra information about the study population, exclusions, caveats, etc. should be included in a footnote.

Ordering and Organization of Results

- What is the "main message" of your table?
- Organize the rows and columns to emphasize your main message.
- Sort data by the most meaningful variable
- Try to maintain consistent organization across different tables in the same report.

Sorting

8a. Youth Television Watching

	Percent of 9-year-olds who watch more than 5 hours of television per weekday
Canada	14.9
Denmark	6.0
Finland	6.1
France	5.5
Germany	4.4
Ireland	11.8
Italy	9.2
Netherlands	12.6
Spain	17.5
Sweden	4.7
United States	21.5

8b. Youth Television Watching

	Percent of 9-year-olds who watch more than 5 hours of televison per weekday
United States	21.5
Spain	17.5
Canada	14.9
Netherlands	12.6
Ireland	11.8
Italy	9.2
Finland	6.1
Denmark	6.0
France	5.5
Sweden	4.7
Germany	4.4

Source: Uri Bronfenburger, et. al. The State of Americans (New York: The Free Press, 1996); qtd. In William Bennett, The Index of Leading Cultural Indicators (New York: Broadway Books, 1999), p. 230

Presenting Meaningful Data

1a: Murders* in Ten Largest US Cities, 1998		1b: Murder Rates* in Ten Largest US Cities, 1998	
Chicago	703	Detroit	43.0
New York	633	Chicago	25.6
Detroit	430	Philadelphia	23.3
Los Angeles	426	Dallas	23.1
Philadelphia	338	Phoenix	15.1
Houston	254	Houston	14.1
Dallas	252	Los Angeles	11.8
Phoenix	185	New York	8.6
San Antonio	89	San Antonio	8.1
San Diego	42	San Diego	3.5
*Murder and non-negligent manslaughter		*Murder and non-negligent manslaughter per 100,000 population	

source: Bureau of Justice Statistics: http://www.ojp.usdoj.gov/bjs/data/cities92.wk1

Column and Row Headings

- Be clear, spell things out – if you use acronyms then footnote them in the table.
- Always include units for numerical data
- Use consistent headings across different tables.
- Use consistent categories across tables.
- Show totals for rows and columns, where appropriate. If you show percentages (%), also give their total (always 100).

Data in the Cells

- In most cases, your numbers should include only 1-2 decimal places – avoid to many decimal places
- E.g. RR = 1.1 vs. RR = 1.13987
- Include confidence intervals and/or P-values where applicable.
- Include SD, SE, or CI for means

Footnotes (Personal Use)

- Always include the name of the SAS program that produced the data in the table – even if you just used “proc print”.
- This creates a “trail of bread crumbs”.
- Slight adjustments to your analysis might result in an entire table having to be re-done.
- Add other notes to yourself as needed.

Footnotes (external use)

- Details about variable definitions, coding, abbreviations.
- Details about exclusions and inclusions.
- Indications of statistical significance.
- Source of data, if not original.
- Covariates included in analysis
- Missing or unknown data (within the table or in footnote)

Oral Presentations

- Only include important results
- One report table might need to be broken down into as many as 8-10 slides.
- Don’t paste huge tables onto slides and then say “sorry you can’t read this”!!
- Use large fonts and clear formatting

Table 1. Description of Cases with Congenital Heart Defects by Maternal and Infant Characteristics and Vital Status, Texas Birth Defects Registry, 1999-2007

	NH*-White		NH-Black		Hispanic		Total	
	Cases	Deceased	Cases	Deceased	Cases	Deceased	Cases	Deceased
Characteristics	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Maternal Age								
< 20 years	966 (9.0)	107 (11.7)	443 (14.6)	74 (18.6)	2,552 (15.7)	246 (16.9)	3,961 (13.2)	427 (15.4)
20-29 years	5,403 (50.4)	448 (49.1)	1,683 (55.4)	208 (52.3)	8,559 (52.7)	716 (49.2)	15,645 (52.1)	1,372 (49.6)
30-39 years	3,917 (36.5)	308 (33.7)	815 (26.8)	105 (26.4)	4,565 (28.1)	413 (28.4)	9,297 (31.0)	826 (29.9)
40+ years	440 (4.1)	50 (5.5)	96 (3.2)	11 (2.8)	576 (3.5)	81 (5.6)	11,12 (3.7)	142 (5.1)
Maternal Education								
< High school	1,435 (13.4)	149 (16.3)	573 (18.9)	79 (19.8)	8,033 (49.4)	744 (51.1)	10,041 (33.5)	972 (35.1)
High school	3,041 (28.4)	280 (30.7)	1,101 (36.3)	161 (40.5)	4,675 (28.8)	404 (27.7)	8,817 (29.4)	845 (30.5)
> High school	6,097 (56.8)	465 (50.9)	1,286 (42.3)	137 (34.4)	3,257 (20.0)	264 (18.1)	10,640 (35.4)	866 (31.3)
Missing	153 (1.4)	19 (2.1)	77 (2.5)	21 (5.3)	287 (1.8)	44 (3.0)	517 (1.7)	84 (3.0)
Infant Sex								
Males	5,467 (51.0)	467 (51.2)	1,517 (50.0)	201 (50.5)	8,086 (49.8)	761 (52.3)	15,070 (50.2)	1,429 (51.6)
Females	5,254 (49.0)	444 (48.6)	1,518 (50.0)	195 (49.0)	8,156 (50.2)	692 (47.5)	14,928 (49.7)	1,331 (48.1)
Missing	5 (0.0)	2 (0.2)	2 (0.1)	2(0.5)	10 (0.1)	3 (0.2)	17 (0.1)	7 (0.3)
Birth Weight & Gestational Age								
< 37 weeks, < 1500 grams	666 (6.2)	147 (16.1)	436 (14.4)	94 (23.6)	1,087 (6.7)	230 (15.8)	2,189 (7.3)	471 (17.0)

Common Errors

- Tables too large → Difficult to follow
- Tables that are too simple: All information included in text (unnecessary table)
- Inadequate definition of symbols/abbreviations
- Tables not cited in text
- Numbers do not add up
- Data in table do not agree with text

Examples of Good Tables

- Table 1 in any manuscript usually presents the characteristics of study participants. Use percentage to show the distribution of selected variables (table shell on next slide)
- You can compare the characteristics of study groups (cases/controls, exposed/unexposed).
- P-values for difference in distribution of selected variables between the two groups can also be presented in table 1.

Table Shell 4.9b Selected Characteristics of Case and Control Participants, SAFE Study—Miami, 1987–1989

		CASES Number	CASES (Percent)	CONTROLS Number	CONTROLS (Percent)
Age	65–74	___	( )	___	( )
	75–84	___	( )	___	( )
	≥85	___	( )	___	( )
Sex	Male	___	( )	___	( )
	Female	___	( )	___	( )
Race	White	___	( )	___	( )
	Black	___	( )	___	( )
	Other	___	( )	___	( )
	Unknown	___	( )	___	( )
Ethnicity	Hispanic	___	( )	___	( )
	Non-Hispanic	___	( )	___	( )
	Unknown	___	( )	___	( )
Hours/day spent on feet					
	≤1	___	( )	___	( )
	2–4	___	( )	___	( )
	5–7	___	( )	___	( )
	≥8	___	( )	___	( )
Smoking status					
	Never smoked	___	( )	___	( )
	Former smoker	___	( )	___	( )
	Current smoker	___	( )	___	( )
	Unknown	___	( )	___	( )
Alcohol use (drinks / week)					
	None	___	( )	___	( )
	<1	___	( )	___	( )
	1–3	___	( )	___	( )

Presentation of Main Results

Table Shell 4.9c Relationship Between Physical Activity (Vigorous and Mild) and Fracture, SAFE Study—Miami, 1987–1989

		CASES No.	CASES (Percent)	CONTROLS No.	CONTROLS (Percent)	Odds Ratio (95% Confidence) Interval)
Vigorous Activity	Yes	___	( )	___	( )	___(___-___)
	No	___	( )	___	( )	
Mild Activity	Yes	___	( )	___	( )	___(___-___)
	No	___	( )	___	( )	

Adapted from: Stevens, JA, Powell KE, Smith SM, Wingo PA, Sattin RW. Physical activity, functional limitations, and the risk of fall-related fractures in community-dwelling elderly. Annals of Epidemiology 1997;7:54-61.

The Good, the Bad, and the Disaster!

- Critically evaluate the tables in the following slides
- Overall impression of each table
- Suggested improvements
- Look at title, column and row headings, footnotes, and data within the cells.

TABLE 1 Michigan Hand Outcomes Questionnaire Patient-Reported Pain Score for Rheumatoid Arthritis

Age (y)	Score (Out of 100)
23	18
28	3
32	9
35	3
36	34
37	16
44	36
45	46
49	41
49	39
52	45
52	76
53	47
55	64
57	61
58	52
58	94
65	83
69	63
69	46
69	67
72	75
78	54
81	63

- What is the conclusion?
- This looks like a whole dataset.
- Readers will need to do all the work.

TABLE 1 Michigan Hand Outcomes Questionnaire Patient-Reported Pain Score for Rheumatoid Arthritis

Age (y)	Score (Out of 100)
23	18
28	3
32	9
35	3
36	34
37	16
44	36
45	46
49	41
49	39
52	45
52	76
53	47
55	64
57	61
58	52
58	94
65	83
69	63
69	46
69	67
72	75
78	54
81	63

BAD

Table 2

TABLE 2 Patient-Reported Pain Scores

Age Group	N	Mean Score
20–39	6	13.83333
40–59	11	54.63636
60–79	6	64.66667
≥80		63.00000

Table 3

Needs Improvement

TABLE 2 Patient-Reported Pain Scores

Age Group	N	Mean Score
20–39	6	13.83333
40–59	11	54.63636
60–79	6	64.66667
≥80		63.00000

Better than the previous one – it is interpretable
*Too many digits, lacks units or references.
* A small, simple table that could easily be explained in text.

Table 4

Table 4.1b Reported Cases of Primary and Secondary Syphilis by Age—United States, 2002

Age Group (years)	CASES Number	Percent
≤14	21	0.3
15–19	351	5.1
20–24	842	12.3
25–29	895	13.0
30–34	1,097	16.0
35–39	1,367	19.9
40–44	1,023	14.9
45–54	982	14.3
≥55	284	4.1
Total	6,862	100.0*

* Actual total of percentages for this table is 99.9% and does not add to 100.0% due to rounding error.

Data Source: Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2002. Atlanta: U.S. Department of Health and Human Services; 2003.

Table 4

Table 4.1b Reported Cases of Primary and Secondary Syphilis by Age—United States, 2002

Age Group (years)	CASES Number	Percent
≤14	21	0.3
15–19	351	5.1
20–24	842	12.3
25–29	895	13.0
30–34	1,097	16.0
35–39	1,367	19.9
40–44	1,023	14.9
45–54	982	14.3
≥55	284	4.1
Total	6,862	100.0*

* Actual total of percentages for this table is 99.9% and does not add to 100.0% due to rounding error.

Data Source: Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2002. Atlanta: U.S. Department of Health and Human Services; 2003.

Good

Good title, clear headings, good organization.

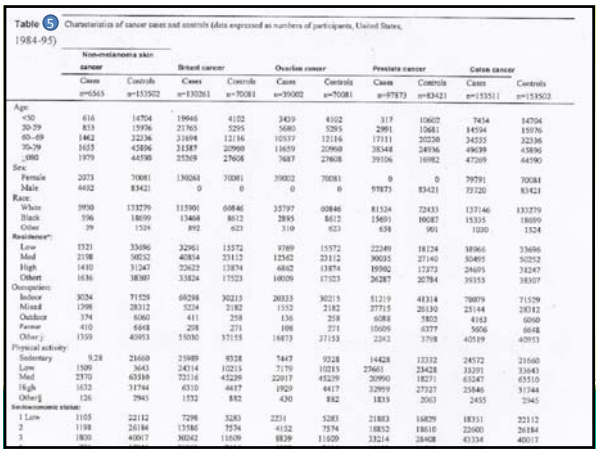


Table 5 Characteristics of cancer cases and controls (data expressed as numbers of participants, United States, 1984-95)

	Non-melanoma skin cancer		Breast cancer		Ovarian cancer		Prostate cancer		Colon cancer	
	Cases n=6565	Controls n=153502	Cases n=130261	Controls n=70081	Cases n=39002	Controls n=70081	Cases n=97873	Controls n=83421	Cases n=153511	Controls n=153502
Age:										
<30	616	14704	19946	4102	3439	4102	317	10602	7454	14704
30-39	853	15976	21765	5295	5680	5295	2991	10681	14594	15976
40–69	1462	32336	31694	12116	10537	12116	17111	20220	34555	32336
70-79	1653	45896	31587	20969	13659	20969	38348	24936	49039	45896
≥80	1979	44590	25269	27608	7687	27608	39106	16982	47269	44590
Sex:										
Female	2073	70081	130261	70081	39002	70081	0	0	79791	70081
Male	4492	83421	0	0	0	0	97873	83421	73720	83421
Race:										
White	5930	133279	115901	60846	35797	60846	81324	72433	137146	133279
Black	596	18699	13468	8612	2895	8612	15691	10087	15335	18699
Other	39	1524	892	623	310	623	658	901	1030	1524
Residence*:										
Low	1321	33696	32961	13572	9789	15572	22249	18124	38966	33696
Med	2198	50252	40854	23112	12562	23112	30035	27140	50495	50252
High	1410	31247	22622	13874	6862	13874	19502	17273	24695	31247
Other	1636	38307	33824	17523	10009	17523	26287	20784	39353	38307
Occupation:										
Indoor	3024	71529	69298	30215	20333	30215	51219	41314	78879	71529
Mixed	1398	28312	5224	2182	1552	2182	27715	26130	25144	28312
Outdoor	374	6060	411	258	136	258	6088	5802	4163	6060
Farmer	410	6648	208	271	108	271	10606	6377	5606	6648
Other‡	1359	40953	55030	37155	16873	37155	2242	3798	40519	40953
Physical activity:										
Sedentary	928	21660	25989	9328	7447	9328	14428	12332	24572	21660
Low	1509	3643	24314	10215	7179	10215	27661	23428	33293	33643
Med	2370	63510	72216	45239	22017	45239	20950	18271	63247	63510
High	1632	31744	6310	4417	1929	4417	32959	27327	25846	31744
Other§	126	2945	1532	882	430	882	1833	2063	2455	2945
Socioeconomic status:										
1 Low	1105	22112	7298	3283	2231	5283	21883	16829	18351	22112
2	1198	26184	13586	7574	4152	7574	18852	13610	22600	26184
3	1800	40017	30242	11609	8829	11609	33214	28408	43334	40017

Table 5

Bad

- Too large, difficult to interpret
- Would have been much easier to read if they provided percentages
- Better if broken into two or more tables (problem: Journals allow a limited number of tables)

GRAPHICAL PRESENTATION OF DATA

Graphs

- Show trends and patterns in the data.
- Paint an interesting picture and make a visual impact.
- Reveal relations between variables in the data.
- Used to present data that is too numerous or complicated to describe adequately in the text

Guidelines for Preparing Graphs

- Be sure that all figures are cited/numbered within the text
- Limit the number of illustrations to include only those that provide essential information
- Number each figure in the order in which they are referred to in the text.
- Note that figures and tables should be numbered separately.

Preparing Graphs

- Make sure the graph is clear and readable. Font size should be large enough to see
- Maximize the data to formatting ratio!!!
- Graph should be fully understandable even outside of the context of the paper.
- Be HONEST!

Elements of a Good Graph

- Appropriate type of graph for the data.
- A clear concise title
- A clear, descriptive legend for each graph
- Scale appropriate for the data
- Minimal number of lines on a graph
- Abbreviations and symbols defined

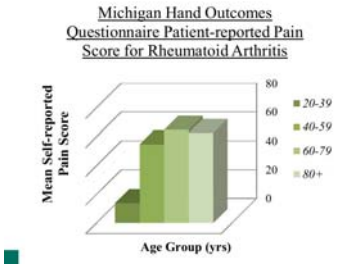
Common Errors

- Information in the text is duplicated in graphs, or information in graphs is duplicated in tables.
- The wrong type of graph is chosen to represent the data.
- The graph is not plotted to scale. Data is not labeled, is inconsistent, or exaggerated.
- Misuse of pseudo three-dimensional graphs.
- Design elements interfere with clarity of graph or figure (Too much formatting)

The Good, the Bad, and the Misleading

- Please examine the graphs in the following slides.
- Is the type of graph appropriate for the data?
- Is the message presented clearly?
- Title? Labels? Scaling?
- Any suggested improvements?

Graph 1



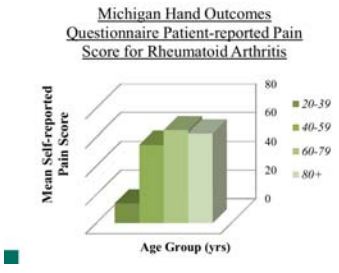
Graph 1

Title too large for the figure. Year?

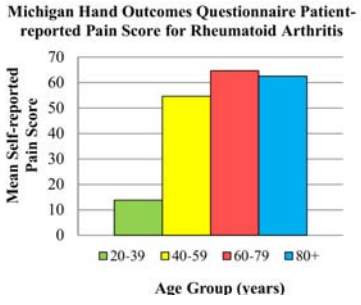
Why 3-D?
Which line corresponds to the measurement (front or back)?

Monochromatic color scheme may be hard to distinguish.

Too much use of underlining, bold typeface, italics…

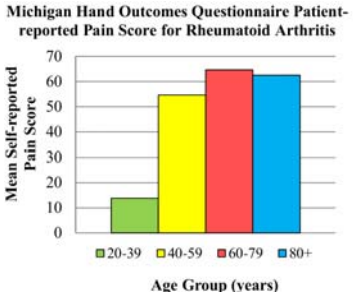


Graph 2

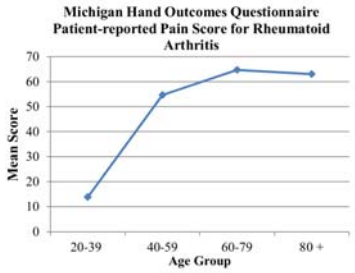


Graph 2

Same data as in Graph 1, but in 2-D. Better Representation of the data.
•Values are not distorted by the skewed perspective.
• Category labels are more space-efficient.
•The graph, not its title, occupies the most space.
•Colors can be distin-guished, even by a color-blind reader

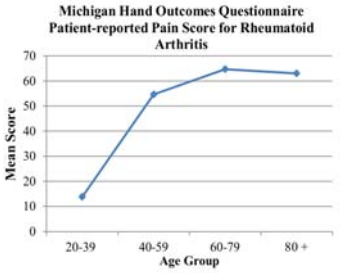


Graph 3

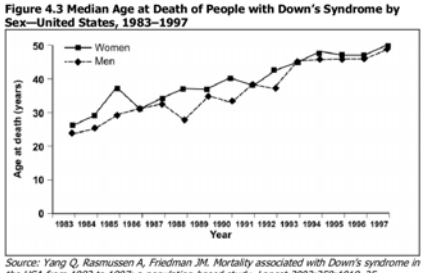


Graph 3

- Same data as in #1 & 2
- WRONG TYPE of graph!
- Discrete points connected in a line graph.
- The connecting segments suggest that there are values between age groups that fall on the lines,
- when in fact the author cannot know this!



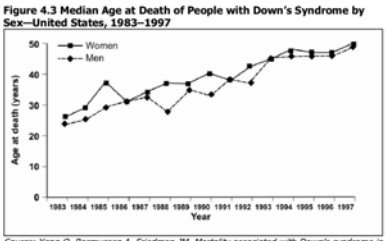
Graph 4



Source: Yang Q, Rasmussen A, Friedman JM. Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. Lancet 2002;359:1019–25.

Graph 4

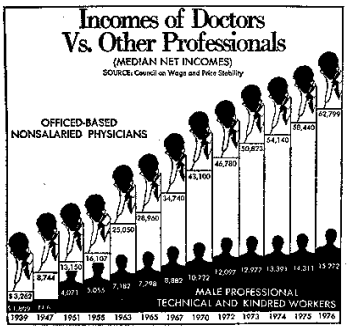
- Title ✓
- Legend ✓
- Type ✓
- Footnote ✓
- Scale ✓
- GOOD!



Source: Yang Q, Rasmussen A, Friedman JM. Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. Lancet 2002;359:1019–25.

Misleading Graphs

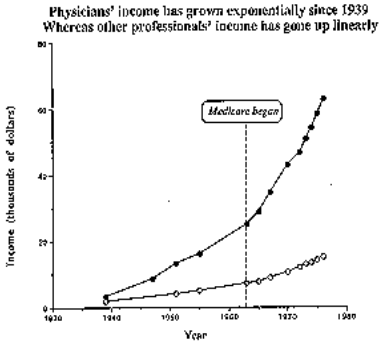
- Scale changed partway through the axis, creating a false impression
- The images distract the attention away from the irregular scale!
- Conveys the impression that doctors incomes increased about linearly, with some slowing down in the later years



A More Accurate Version

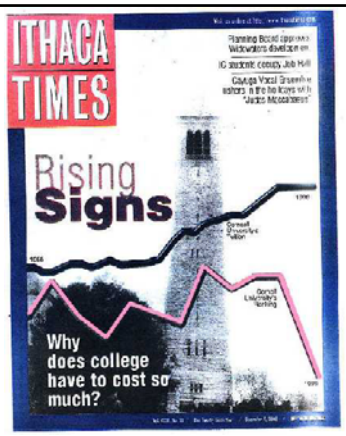
A re-drawing of the graph, with year values spaced appropriately, gives an entirely different message

Doctors' income now appears to have increased exponentially, while the incomes of other professionals seems to have increased linearly.



This image, from the cover of the *Ithaca Times (Dec. 7, 2000)* may well be **the most misleading graph ever published.**

The Top line is labeled "Cornell's Tuition" and the bottom one "Cornell's Ranking".

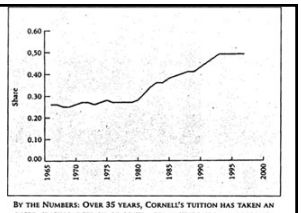


The individual graphs as presented inside.

Top graph: Cornell's tuition (1965-2000).

Bottom: Cornell's Ranking (1988-1999)

What's wrong?



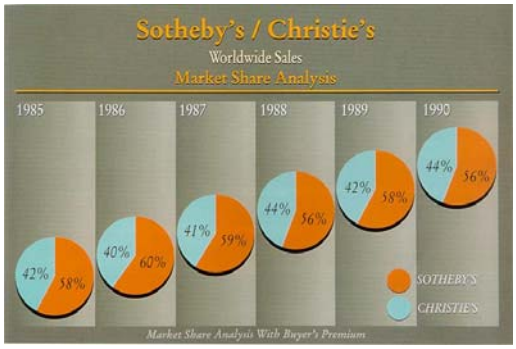
By the Numbers: Over 35 years, Cornell's tuition has taken an increasingly larger share of its median student family income.

Pecking Order: Over 12 years, Cornell's ranking in *US News & World Report* has risen and fallen erratically.

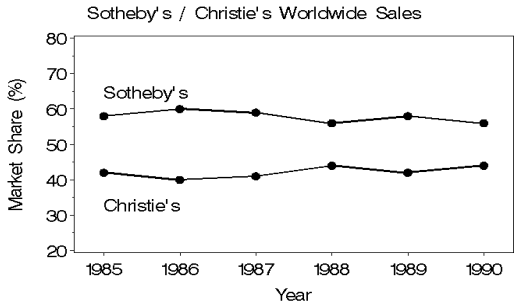
What is Wrong?

1. The ranking graph covers an 11 year period, the tuition graph 35 years, yet they are shown simultaneously (the same apparent width) on the same horizontal "scale".
2. The ranking graph is placed under the tuition graph creating the impression that cost exceeds quality.
3. The differing time units are cleverly disguised by printing them rotated 90°.
4. And here is the masterstroke: the sharp "drop" in the ranking graph over the past few years actually represents the fact that **Cornell's rank has IMPROVED from 15th TO 6th ...**

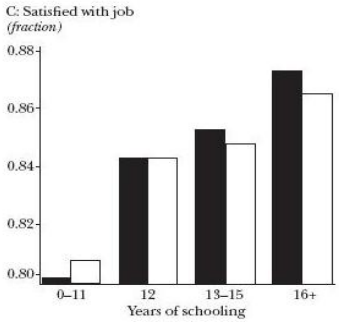
What Does this Graph Say?



A Much Better Graph

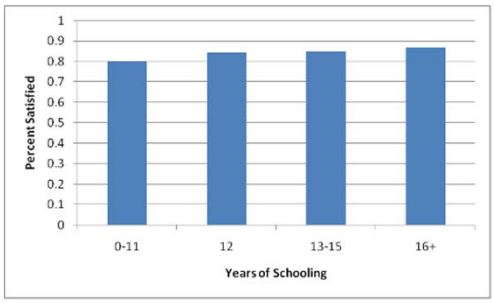


One more…



http://scienceblogs.com/principles/2011/07/10/great-moments-in-deceptive-gra/

How does it look now?



Thank you

University of South Florida College of Public Health
our practice is our passion.