

6. Creating Effective Charts

PROBLEM SET

1. List what is missing from the charts in figures 6A and 6B.

**Age distribution of the elderly population
United States, 2000**

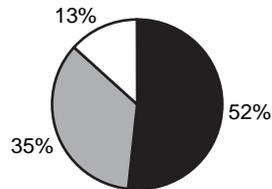


Figure 6A.

Median sales price of new single-family homes, by region, United States, 1980–2000

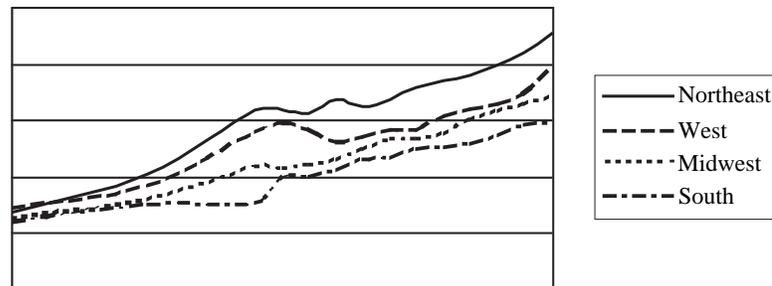


Figure 6B.

2. Answer the following questions for figures 6.4 and 6.5 in *Writing about Multivariate Analysis, 2nd Edition*.
 - a. Who is described by the data?
 - b. To what date or dates do the data pertain?
 - c. Where were the data collected?
 - d. What criteria were used to organize the values of the variables on chart axes? (Hint: Consider type of variable.)
 - e. What are the units of measurement? Are they the same for all numbers shown in the chart?
 - f. Are there footnotes to the chart? If so, why? If not, are any needed?

3. For each of the following topics, identify the type of task (e.g., univariate distribution, bivariate association, or relationship among three variables), and types of variables to be presented (nominal, ordinal, interval, or ratio), then state which type of chart would be most suitable, using the guidelines in table 6.1 on pp. 140–41 of *Writing about Multivariate Analysis, 2nd Edition*.
 - a. Projected number of people receiving college degrees by region of the country from 2010 to 2025 under three different scenarios about rates of college attendance and completion
 - b. Average commuting costs per month, by mode of transportation (bicycle, bus, car, train, walk, other); one number per type of transportation
 - c. Number of cases in a study sample from rural, suburban, and urban areas
 - d. Educational attainment distribution (<HS, =HS, >HS) for native-born US residents and immigrants from other North American countries, Africa, Asia, Australia and New Zealand, Europe, and Latin America in the year 2000
 - e. Estimated odds ratios and 95% confidence intervals for gender, major occupation category (blue collar, white collar, service, other), and region (four major census regions) from a logistic regression of being laid off in the past year
 - f. Overall effect of a quadratic specification of percentage body fat in an OLS model of systolic blood pressure (millimeters of mercury [mm Hg])
 - g. Overall effects of an interaction between tercile of a student's own high school class rank and their mother's educational attainment (<HS, =HS, >HS) on the student's first-year college grade point average (GPA). Results are based on an OLS regression controlling for gender, race, and family income, using data from the high school classes of 1995 through 2000. The top tercile of each variable in the interaction is the reference category.
4. Use the data in table 5.5 (p. 89 of *Writing about Multivariate Analysis, 2nd Edition*) to create a chart comparing the racial composition of the NHANES III study sample to that of all US births. Include a complete title, labels, legend, and notes.
5. Draft one or more charts to present the findings shown in table 6A.
 - a. Use the criteria in table 6.1 on pp. 140–41 of *Writing about Multivariate Analysis, 2nd Edition* to determine which type of chart matches the number and types of variables.
 - b. Indicate which variables would go on the axes and which would go in the legend. Hint: Consider whether panels are needed, and if so, which portions of the table go into each panel.

TABLE 6A. Means and standard deviations of psychiatric symptoms by gender and pubertal timing, African American children, 1997 Family and Community Health Study

Psychiatric symptoms	Early maturers (E)	On-time maturers (O)	Late maturers (L)	F-statistic	p-value	Post-hoc comparisons ^a
Girls						
Attention deficit hyperactivity disorder	(N = 88) 7.08 (5.62)	(N = 286) 5.66 (5.42)	(N 111) 4.78 (4.97)	4.46	0.01	E>O; E>L
Conduct disorder	3.64 (2.11)	3.62 (1.93)	3.49 (2.18)	1.82	0.16	
Generalized anxiety disorder	4.20 (2.59)	3.64 (2.73)	3.37 (2.66)	2.34	0.10	E>O
Major depression	7.19 (4.51)	6.16 (4.83)	5.04 (4.85)	4.96	0.01	E>L; O>L
Oppositional defiance disorder	3.37 (3.16)	2.68 (2.85)	2.07 (2.44)	5.06	0.01	E>O; E>L
Social anxiety	4.62 (2.71)	4.40 (2.78)	4.18 (2.88)	0.60	0.55	
Boys						
Attention deficit hyperactivity disorder	(N = 77) 7.41 (5.53)	(N = 256) 6.24 (5.38)	(N = 78) 4.72 (4.85)	4.81	0.01	E>O; E>L
Conduct disorder	2.39 (2.91)	2.08 (2.95)	1.88 (2.56)	0.61	0.55	
Generalized anxiety disorder	4.27 (2.79)	3.43 (2.76)	2.76 (2.47)	5.81	<0.05	E>O; E>L
Major depression	7.53 (4.62)	5.84 (4.64)	4.97 (4.72)	5.98	0.00	E>O; E>L
Oppositional defiance disorder	3.57 (3.03)	2.92 (3.01)	2.39 (2.86)	2.93	<0.05	E>O
Social anxiety	4.55 (2.47)	3.59 (2.69)	3.28 (2.44)	5.17	0.01	E>O; E>L

Adapted from Ge et al., "Pubertal Maturation and African American Children's Internalizing and Externalizing Symptoms," *Journal of Youth and Adolescence* 35, no. 4 (2006):528-537, table IV.

Notes: Symptoms based on Diagnostic Interview Schedule for Children, Version 4 (DISC-IV). I intentionally changed the order in which the symptoms are listed in the table from that used by the authors. All numbers are unchanged from the article. Post-hoc comparisons are based on $p < 0.05$. "E" = "early maturer," "O" = "on-time maturer," "L" = "late maturer."

- c. Include complete titles, axis labels, and footnotes to define terms and indicate statistical significance.
 - d. In table 6A, the types of psychiatric symptoms are arranged in alphabetical order. What principle(s) would you use to reorganize the order of those symptoms to improve the coordination of the chart with the associated prose? Explain why you chose those criteria, with reference to the guidelines in chapter 6.
6. Use the criteria in chapter 3 to assess the findings in table 6A in terms of
- a. The statistically significant findings
 - b. Substantively meaningful findings
 - c. The additional information you would need to evaluate causality of the associations
7. Create a stacked bar chart to present the data shown in table 6B, allowing the bar height to vary to show total number of ozone days. To help you plan your chart, answer the following questions, then draw an approximate stacked bar chart, allowing the level to vary by county.
- a. Which variable goes on the x axis, and what principle would you use to organize its values?
 - b. Which variable goes in the slices (and legend)?
 - c. Which variable goes on the y axis, and in what units is it measured?
 - d. What is the title for the chart?

TABLE 6B. Number of unhealthy ozone days by level of warning for selected counties in Indiana, 1996–1998

	Level of warning ^a		
	Unhealthy for sensitive groups	Unhealthy	Very unhealthy
Allen	25	0	0
Clark	29	3	1
Elkhart	15	0	0
Floyd	27	6	0
Hamilton	31	3	0
Hancock	28	2	0
Lake	29	2	0
La Porte	26	6	1
Madison	27	3	0
Marion	32	3	0
Porter	25	3	0
Posey	14	1	0
St. Joseph	21	1	0
Vanderburgh	32	2	0
Vigo	25	1	0
Warrick	40	3	0

^aUnhealthy for sensitive groups = 0.085–0.104 parts per million (ppm); Unhealthy = 0.105–0.124 ppm; Very unhealthy = 0.125–0.374 ppm.

Source: American Lung Association.

8. Revise your chart from the previous question to illustrate the relative importance (share) of different levels of ozone warning in each county.
 - a. What aspects of each chart remain the same as in the previous question? What aspects change?
 - b. What are the advantages and disadvantages of the two versions of the chart with reference to this topic and data?

9. Fussell and Massey (2004) used data from the Mexican Migration Project to study relationships among demographic factors, human capital, social capital in the family and community, and migration from Mexico to the United States (table 6C). Use that information to create charts showing the following patterns. Hint: Use a spreadsheet, following the guidelines in appendix D of *Writing about Multivariate Analysis, 2nd Edition*.
 - a. The association between age in years and relative odds of first trip to the United States, compared to 15-year-olds. Allow age to vary from 15 to 64 years.
 - b. The association between migration prevalence ratio and relative odds of first trip to the United States, with 95% confidence intervals.

TABLE 6C. Estimated log-odds of first trip to the United States, men, 1987–1998 Mexican Migration Project

	Log-odds	Standard error
<i>Demographic background</i>		
Age (years)	−0.003	0.02
Age-squared	−0.001	0.0002
Ever married	−0.09	0.06
Number of minor children in household	0.01	0.01
<i>Human capital</i>		
Years of education	−0.04	0.006
Months of labor-force experience	−0.002	0.0007
<i>Social capital in the family</i>		
Parent a prior US migrant	0.51	0.05
Siblings prior US migrants	0.36	0.02
<i>Social capital in the community</i>		
Migration prevalence ratio ^a		
0–4	−0.99	0.15
5–9	−0.09	0.12
(10–14)		
15–19	0.35	0.10
20–29	0.57	0.13
30–39	0.95	0.15
40–59	0.74	0.19
60 or more	0.34	0.15
Intercept	−3.31	0.26
−2 log likelihood	23,369.2	
df	26	

Source: Adapted from Elizabeth Fussell and Douglas S. Massey, “The Limits to Cumulative Causation: International Migration from Mexican Urban Areas,” *Demography* 41, no. 1 (2004): 151–71, table 2. <http://muse.jhu.edu/journals/demography/v041/41.1fussell.pdf>.

Note: Model also includes controls for occupational sector, internal migratory experience, community characteristics, and Mexican economic and US policy context.

^aThe migration prevalence ratio = (the number of people aged 15+ years who had ever been to the US/the number of people aged 15+ years) × 100.

10. In a study of sexual behavior among youths in Kenya, Mensch and colleagues (2003) evaluated whether audio computer-assisted self-interviewing (ACASI) produces more valid reporting of sexual activity and related sensitive behaviors than face-to-face interviews or self-administered written interviews. Their results are reported in table 6D. Use that information to create charts

- a. to accompany a “Generalization, example, exception” (GEE) description of whether reporting a sensitive behavior differs by mode of interview among boys;
- b. to accompany a GEE description of whether the association between mode of interview and reporting having had more than one sexual partner differs by gender.

TABLE 6D. Odds ratios from logistic regressions of reporting sensitive behaviors, by mode of interview and gender, Kisumu District, Kenya, 2002

Behavior	Boys	Girls
Ever had a boyfriend or girlfriend		
Interviewer-administered	1.00	1.00
Self-administered	0.78	0.82
ACASI ^a	0.43***	0.69*
Ever had more than one sexual partner		
Interviewer-administered	1.00	1.00
Self-administered	1.02	0.72
ACASI ^a	1.28	2.35***
Ever had sex with a stranger		
Interviewer-administered	1.00	1.00
Self-administered	1.43	1.24
ACASI ^a	2.42**	4.25***
Ever tricked/coerced/forced into sex		
Interviewer-administered	1.00	1.00
Self-administered	2.33***	1.89**
ACASI ^a	2.40***	3.35***

Source: Adapted from Barbara S. Mensch, Paul C. Hewett, and Annabel S. Erulkar, “The Reporting of Sensitive Behavior by Adolescents: A Methodological Experiment in Kenya,” *Demography* 40, no. 2 (2003): 247–68, table 2. <http://muse.jhu.edu/journals/demography/v040/40.2mensch.pdf>.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

^aACASI = audio computer-assisted self-interviewing.