# 6

Creating Effective Charts

1. Figure 6A is missing a legend; 6B is missing axis titles, axis labels, and units of measurement.

3. Identify the task and types of variables, then state the appropriate type of chart.
   a. Three-way association between one continuous and one nominal predictor (date and type of scenario, respectively), and a continuous outcome (number of people receiving degrees). Multiple-line chart, to show projected number by date (on the *x* axis) in the number of people receiving college degrees (on the *y* axis), with different lines and line styles for low, medium, and high scenarios (identified in the legend). Notes about data sources and assumptions used in each scenario.
   b. Two-way (bivariate) association between transportation mode (nominal) and cost (continuous). Simple bar chart, with one bar for each transportation mode on the *x* axis and cost on the *y* axis.
   c. Composition (univariate) of a nominal variable. Pie chart to illustrate the percentage (or number of cases) from rural, suburban, and urban areas.
   d. Distribution of one categorical variable (educational attainment) within another categorical variable (continent). Stacked bar chart, with bars for U.S. native-born people and each continent of origin, and one slice for each educational attainment level. Each bar totals 100% of that continent's immigrants (on the *y* axis) to illustrate composition while correcting for different numbers of immigrants across continents.
   e. Association between several nominal independent variables (gender, occupation, and region) and a continuous dependent variable (relative odds of being laid off in the past year). High/low/close chart ("high" and "low" show the upper and lower 95% confidence limits), with the independent variables on the *x* axis and the odds ratios on the *y* axis.
   f. Association between a continuous independent variable (percentage body fat) and a continuous dependent variable (systolic blood pressure). Single-line chart with the percentage body fat on the *x* axis and blood pressure on the *y* axis, each labeled with its respective units.
   g. Net effects of an interaction between two categorical independent variables (tercile of student's class rank and mother's educational attainment) and a continuous independent variable (first-year college GPA). Clustered bar chart with one cluster for each category of mother's education on the *x* axis and a different bar color for each tercile of class rank (in the legend). *Y* axis shows predicted mean first-

year college GPA. Notes specifying data source and other variables controlled in the model (or naming a table in which those estimates are shown), identifying the reference categories for class rank and mother's education, and defining symbols used to denote statistical significance.

5. Create a stacked bar chart, after answering the given questions.
    a. Counties arranged on the *x* axis in descending order of total number of unhealthy ozone days
    b. A different color slice for each level of ozone warning, identified in the legend
    c. Number of unhealthy ozone days goes on the *y* axis
    d. Same title as table 6A: "Number of unhealthy ozone days by level of warning for selected counties in Indiana, 1996–1998"

7. Create charts showing the specified patterns from analysis by Fussell and Massey (2004).
    a. Chart to portray the association between age in years and relative odds of first trip to the United States, compared to 15-year-olds.

**Relative Odds of First Trip to the United States, Men,
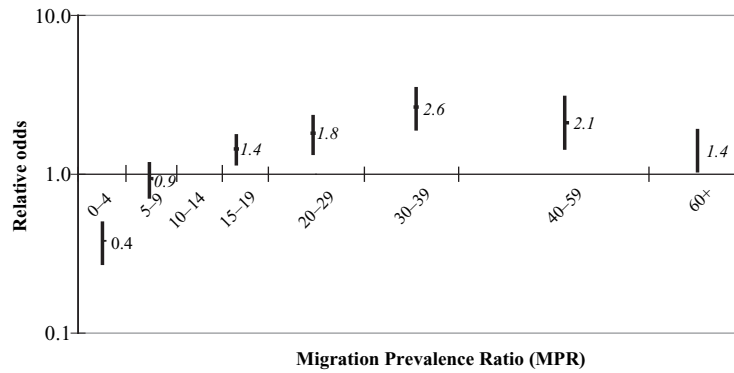1987–1998 Mexican Migration Project**
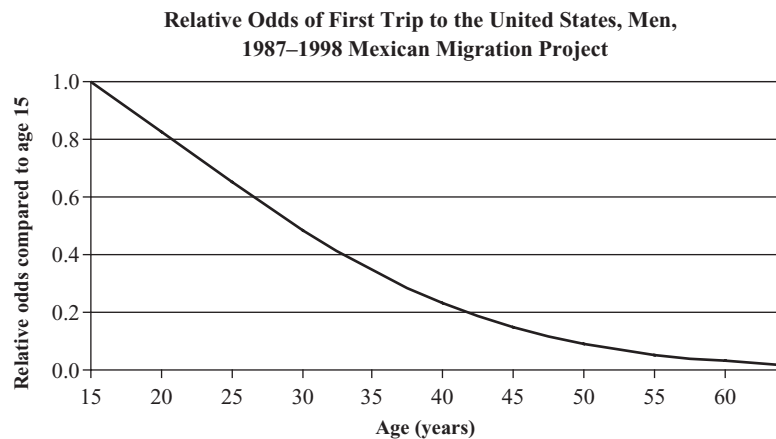
Based on model controlling for marital status, number of children, education, labor force experience, family migrant history, and migration prevalence ratio. Reference category = 15 year olds.

Figure 6C.

b. Chart to portray the association between the migration preva-
lence ratio and relative odds of first trip to the United States, with
95% confidence intervals.

**Relative Odds and 95% Confidence Interval (CI) of First Trip to the United States,
by Migration Prevalence Ratio, Men, 1987–1998, Mexican Migration Project**

Compared to MPR = 10-14. Based on model controlling for age, marital status,
number of children, education, labor force experience, and family migrant history.

Figure 6D.

Comments: A logarithmic scale was used to preserve symmetry in ap-
parent sizes of odds ratios above and below 1.0; see "Charts to Display
Logistic Regression Results" on page 157 of *Writing about Multivariate
Analysis* for an explanation. Spacing of categories on *x* axis is propor-
tional to actual width of the Migration Prevalence Ratio (MPR) cate-
gories: 5-year-wide MPR categories (e.g., 0–4, 15–19) appear half as
wide as 10-year-wide MPR categories (e.g., 30–39), which are half as
wide as the 20-year-wide MPR category (40–59).

9. Create charts to accompany the specified GEE descriptions of results
from Mensch et al. (2003).
a. Chart presenting odds ratios of reporting a sensitive behavior by
mode of interview among boys.

**Odds Ratios of Reporting Ever Had Specified Sensitive Behaviors
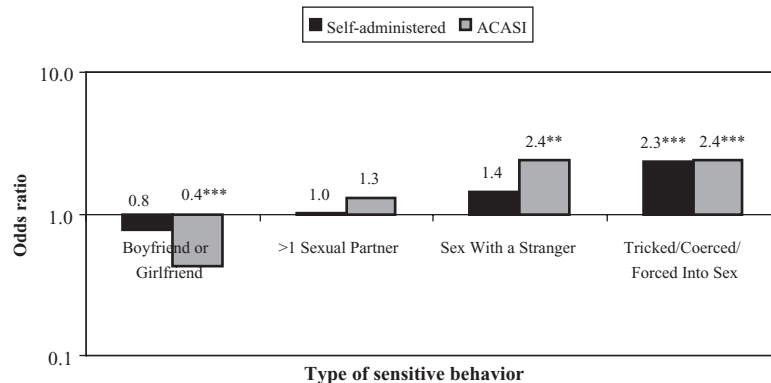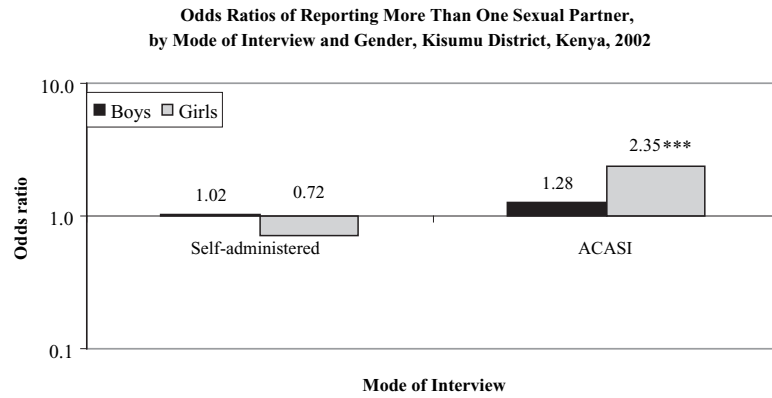by Mode of Interview, Boys, Kisumu District, Kenya, 2002**

Figure 6E.

b. Chart of the association between mode of interview and odds ratios of reporting having had more than one sexual partner by gender.

**Odds Ratios of Reporting More Than One Sexual Partner,
by Mode of Interview and Gender, Kisumu District, Kenya, 2002**

ACASI = audio computer-assisted self-interviewing. Reference category = Interviewer administered.
Compared to children of the same gender.
* p < 0.05; ** p < 0.01; *** p < 0.001.

Figure 6F.

Comments: A logarithmic *y* scale was used on figures 6E and 6F to pre-serve symmetry in apparent sizes of odds ratios above and below 1.0.