6. *Creating Effective Charts*

SOLUTIONS

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 ordinal 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 separate bars for US native-born people and each continent of origin, and one slice for each educational attainment level (in the legend). 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, the odds ratios on the *y* axis, and a reference line at $y = 1.0$.
   f. Association between a continuous independent variable (percentage body fat) and a continuous dependent variable (systolic blood pressure). Single-line chart calculated from the regression coefficients and input values of percentage body fat, with the percentage body fat on the *x* axis and blood pressure on the *y* axis, each labeled with its respective units.

g. Overall effects of an interaction between two ordinal 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). The *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. Charts to portray the relationships shown in table 6A.

a. Clustered bar chart with two panels to display the association among type of disorder (six nominal independent variables, one for each of six types), pubertal timing (ordinal categorical independent variable), mean number of symptoms (continuous dependent variable), and gender (nominal independent variable).
b. The type of disorder goes on the *x* axis variables, pubertal timing in the legend, and mean number of symptoms on the *y* axis, with one panel for boys and one for girls.
c. Figures 6C1 and 6C2.
d. In figures 6C1 and 6C2, the types of psychiatric disorders are arranged into one set of clusters for the three internalizing disorders

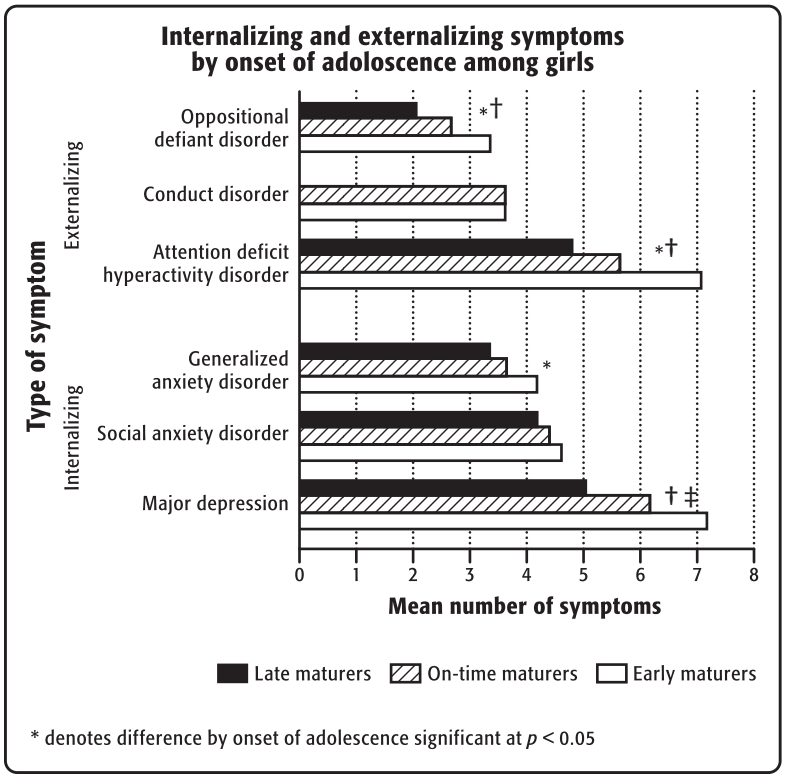


Figure 6C1.

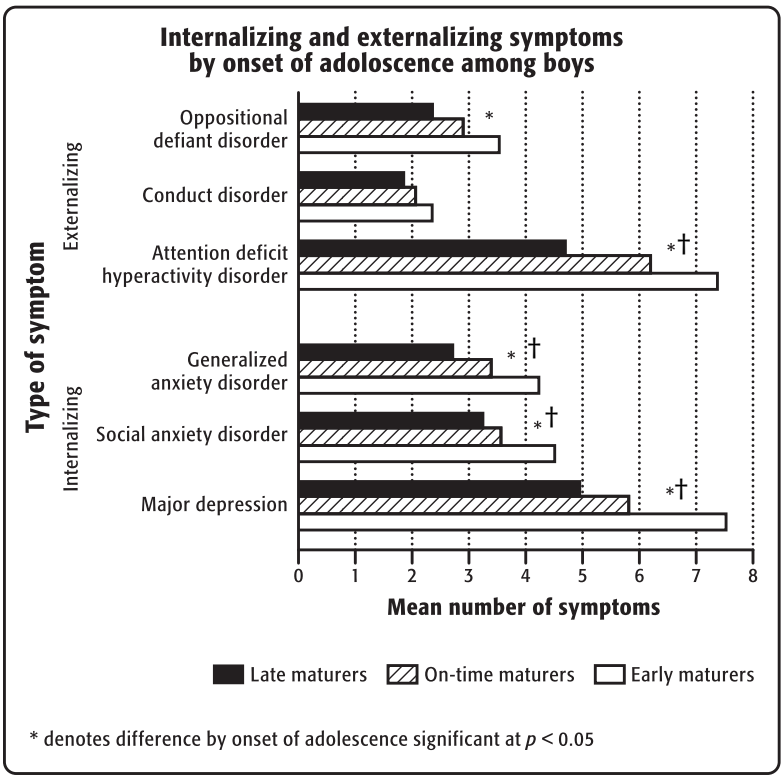


Figure 6C1. and C.2. Mean number of psychiatric symptoms by type of disorder, timing of pubertal maturation, and gender among African American children, 1997 Family and Community Health Study.
Panel 1: Girls
Panel 2: Boys
Source: Ge et al. 2006. "Pubertal Maturation and African American Children's Internalizing and Externalizing Symptoms." *Journal of Youth and Adolescence*. 35(4):528–537. Table IV.
* denotes early maturer > on-time maturer ; † denotes early maturer > late maturer; ‡ denotes on-time maturer > late mature at p < 0.05 based on post-hoc tests.

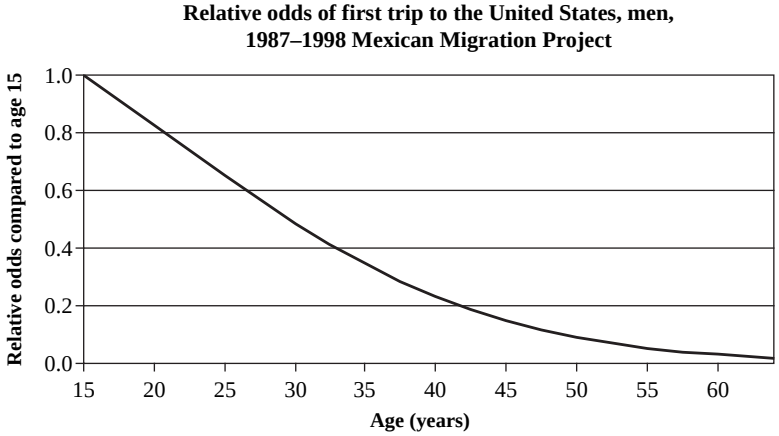
and another set of clusters for the three externalizing disorders, following those conceptual groupings which are mentioned in the title to the article (see footnote to table 6A). The categories of pubertal timing are retained in ordinal sequence, fitting the conceptual meaning of that variable. There is one panel for each gender with one cluster for each type of symptom because that is consistent with the statistical tests, which test whether the mean number of symptoms differ across pubertal timing groups within each gender. Separately within the sets of internalizing and externalizing disorders, the conditions are arranged in descending order of mean number of symptoms.

7. Stacked bar charts, based on the given answers.

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

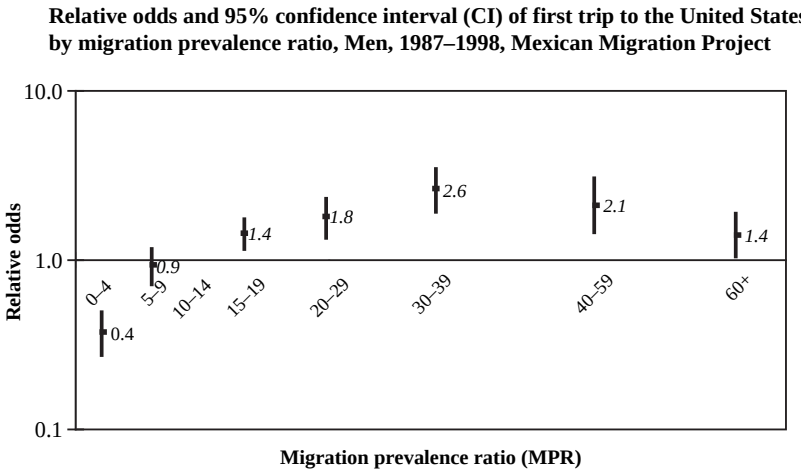
9. Create charts showing the specified patterns from analysis by Fussell and Massey (2004).



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 6D.

a. Figure 6D portrays the association between age in years and relative odds of first trip to the United States, compared to 15-year-olds.



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 6E.

b. Figure 6E portrays the association between the migration prevalence ratio and relative odds of first trip to the United States, with 95% confidence intervals.

Comments: A logarithmic scale was used to preserve symmetry in apparent sizes of odds ratios above and below 1.0; see "Charts to Display Logistic Regression Results" on pp. 147–49 of *Writing about Multivariate Analysis, 2nd Edition* for an explanation. Spacing of categories on the *x* axis is proportional to actual width of the Migration Prevalence Ratio (MPR) categories: 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).