on the charge of simple possession of a small amount of marijuana, for which enforcement proce-dures allowed police discretion. An officer could release an arrestee with a summons ("Form 9") to appear in court, or take the person to a police station for questioning ("Form 10") or booking ("Form 11.1") or order the person held in jail for a bail hearing ("Show cause").

The statistical issue was whether the data on these arrests showed evidence of differential treat-ment in relation to skin color, particularly in the treatment of blacks vs. whites, controlling, of course, for other factors. Statistical tests on these data (χ^2 tests, loglinear models, logistic regres-sion) showed overwhelming evidence of differential treatment of blacks and whites. However, tables of these results do not reveal the nature of this association.

Figure 1.7 is an example of a graph designed for *analysis*—a mosaic display (Chapter 5) show-ing the frequencies of those arrested on this charge by skin color and release type. The size of each rectangle shows the frequency and these are shaded in relation to the asociation between skin color and release—blue for positive associations (more than expected under independence) to red for negative associations.
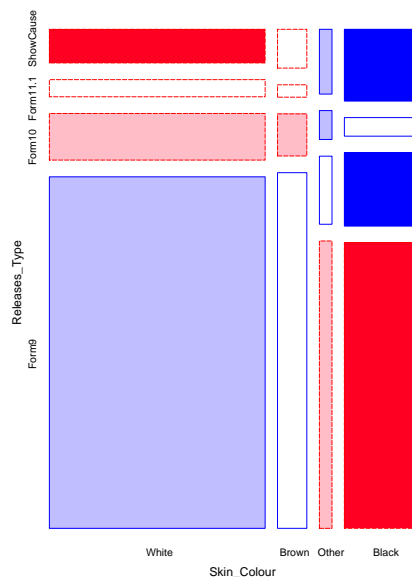

Figure 1.7: Mosaic display showing the relationship between skin color and release type for those arrested on a charge of simple possession of marijuana in Toronto, 1996-2002.

{fig:arrests0-mosaic}

Once you know how to read such graphs, the pattern is clear: blacks were indeed more likely to be held for more severe treatment, whites were more likely to be released with a summons. But this is hardly a graph that would be clear to a general audience, and would require a good deal of explanation.

In contrast, Figure 1.8 shows a redesign of this as a *presentation graphic* prepared by the *Star* and published on December 11, 2002 in conjunction with a meeting between the newspaper and the Toronto Police Services Board to consider the issue of racial profiling. The police vehemently denied that racial profiling was taking place. The revision makes the point immediately obvious and compelling in the following ways:

• It announces the conclusion in the figure title: "Same charge, different treatment"

- The text box at the top provides the context for this conclusion
- Skin colors "Brown" and "Other", which were of low frequency were removed, and the release categories "Form 10" and "Form 11.1" were combined as "released at station."
- The graphic is still a mosaic display, however, it now shows explicitly the number of charges laid against whites and blacks and the percentage of each treatment.
- The labels for Whites and Blacks were enhanced by indicating what a reader should see for each.
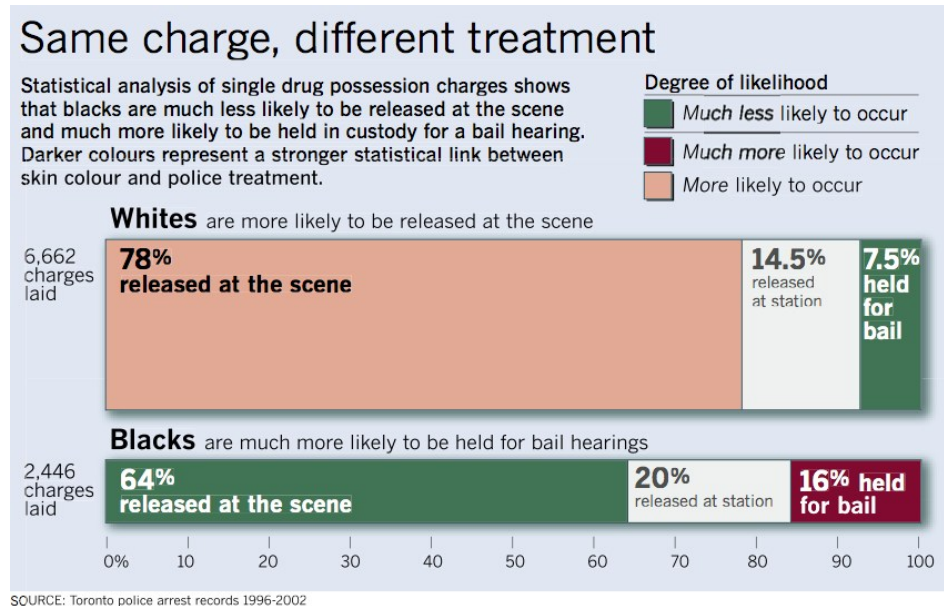- The legend for color is titled non-technically as "degree of likelihood."

Figure 1.8: Redesign of Figure 1.7 as a presentation graphic. *Source*: Graphics department, *The Toronto Star*, December 11, 2002. Used by permission.

{fig:arrests0-star}

Clear communication is not achieved without effort. The revised graph required several iterations and emails between the graphic designer and the statistical consultant (the first author of this book) in the few hours available before the newspaper went to press. The main question was, "what are we trying to show here?" Starting with the original Figure 1.7 mosaic, we asked "what can we remove?" and "what can we add?" to make the message clearer.

△

1.4.2   Categorical data require different graphical methods

{sec:intro-catdata}

We mentioned earlier, and will see in greater detail in Chapter 7 and Chapter 9, that statistical models for discrete response data and for frequency data are close analogs of the linear regression and ANOVA models used for quantitative data. These analogies suggest that the graphical methods commonly used for quantitative data may be adapted directly to categorical data.

Happily, it turns out that many of the analysis graphs and diagnostic displays (e.g., effect plots, influence plots, added variable and partial residual plots, etc.) that have become common adjuncts in the analysis of quantitative data have been extended to generalized linear models including logistic regression (Section 7.5) and loglinear models (Section 11.6)

Unhappily, the familiar techniques for displaying raw data are often disappointing when applied