Visualizing Crash Data by Age Group, Time, and Year

Richard M. Heiberger

Abstract

Does the tendency of teenage drivers to be involved in automobile accidents increase dramatically with the number of passengers in the car? Believing this to be so, legislators in several states require that junior license holders be prohibited from driving with more than one passenger. Is the GES 2000–2014 data consistent with earlier 2001 Insurance Institute for Highway Safety results? Are the restrictions on junior license holders still justified?

Introduction

Figure 1 is an interaction plot of the 2001 Insurance Institute for Highway Safety data showing that teenagers have different changes in crash rates (trips per crash) as the number of passengers increases.

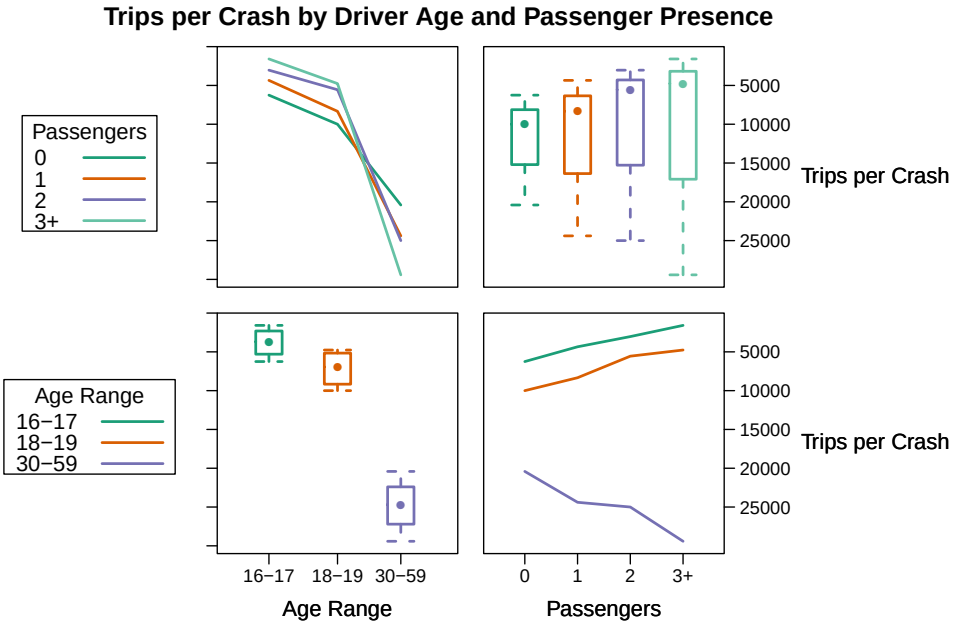

Figure 1: Interaction plot of reciprocal crash rate data. HH2 Figure 14.18 from Heiberger and Holland (2015). Based on a 2001 report from the Insurance Institute for Highway Safety (Williams, 2001) The message is best seen in the bottom row. The boxplots in the lower-left panel show a difference in the small number of trips per crash for teens compared to the larger number of trips per crash for the adults. The lines in the lower-right panel show that adults increase the number of trips per crash with more passengers, meaning that they drive more safely with more passengers. But teens decrease the number of trips per crash with more passengers, meaning that they drive less safely with more passengers.

*rmh@emple.edu Temple University, Department of Statistical Science, Fox School of Business

Display of DOT NASS data

Can the DOT National Automotive Sampling System data be used to answer a question about rates? The DOT data is a survey containing only weighted numerator information about crashes. It is missing the denominator "Number of Trips".

A different type of measure is needed. Figure 2 shows a mosaic plot of conditional percentage, the percent of crashes that have 3 or more occupants in the car. Each panel shows the proportion of crashes with the stated number of occupants in the specified age group and for the specified time of day. Wee see that at all times of day, the teens involved in crashes have more occupants than the adults.

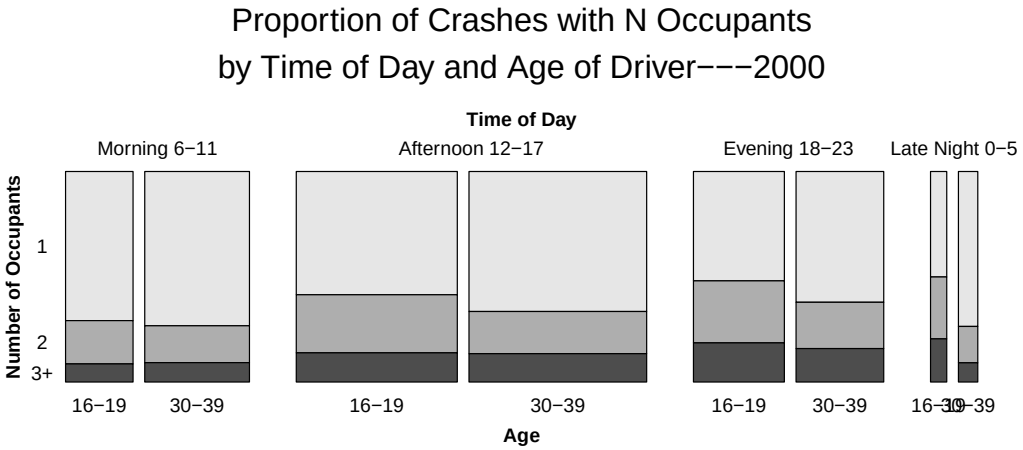

Figure 2: Mosaic plot comparing teenage drivers to 30-year old drivers by time of day. (6-hour ranges beginning at 0600 (6AM)) and by number of occupants in the car. Each partitioned vertical bar shows the proportions of one, two, three-or-more passengers conditional on the specified age group and time of day. The width of each bar is proportional to the weighted number of crashes by that age group in that time period. In each time of day the teens have higher proportions of multiple occupants than the older drivers. In the evening (18-23) and early morning (0-5, think of it as late night), the percentage of multiple occupants are higher than daytime for both age groups, and much higher for teens than older drivers. Conditional on being in a crash, the teens are more likely to have more passengers than the older drivers.

Does the pattern observed for 2000 in Figure 2 hold for other years? Figure 3 shows the Afternoon (12:00–17:59) plots of conditional proportions for all ages (by decades) for the years 2000–2014. There is no visibly significant difference over years. In all years there are decreases in multiple occupants as the groups ages increase.

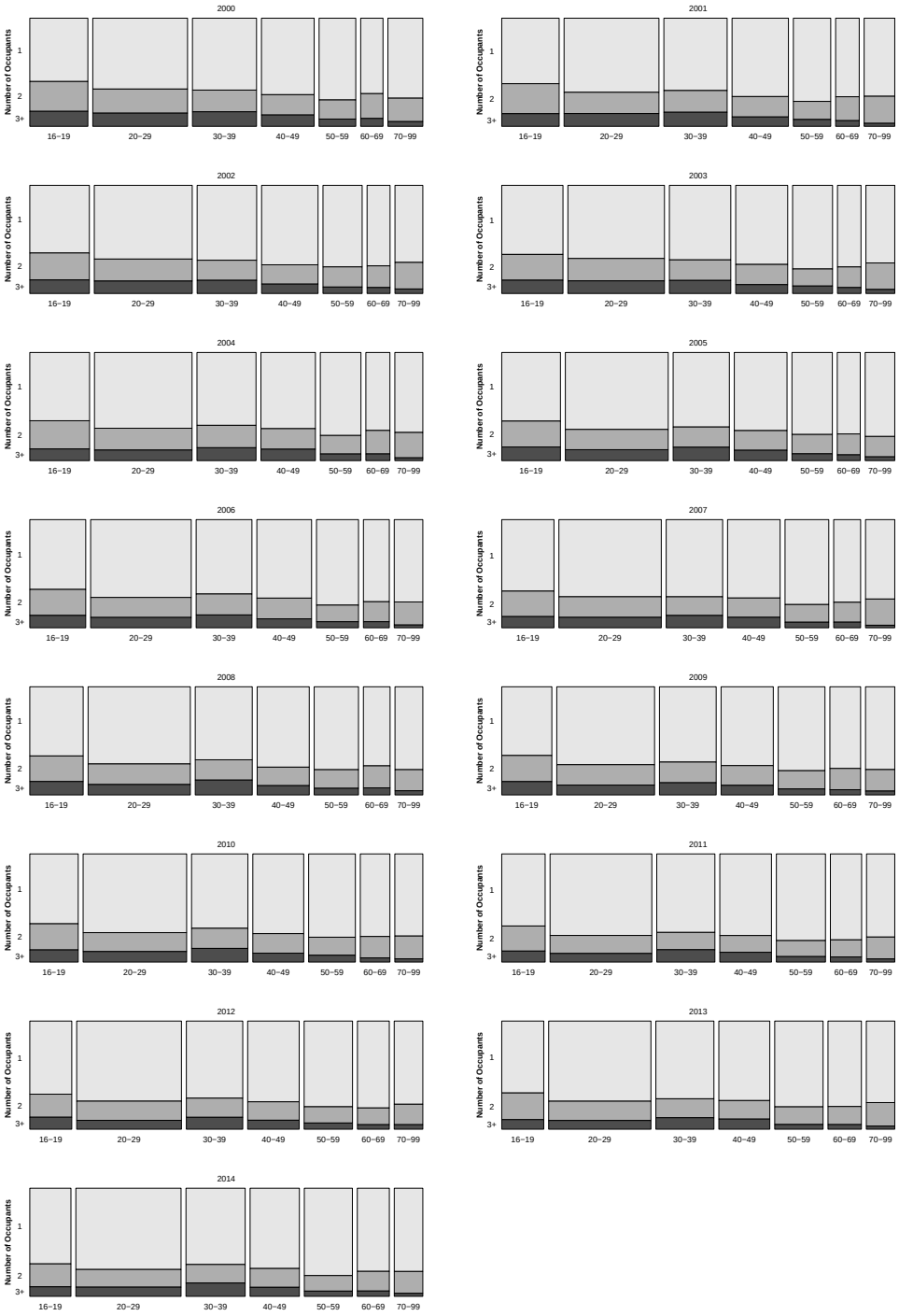

Figure 3: Afternoon conditional proportions for all ages (grouped by decades) for the years 2000–2014. There is no visibly significant difference over years. In all years there are decreases in multiple occupants as the groups ages increase.

Conclusion

For the original question: "Should teenage drivers be prohibited from driving with more than one passenger?" We cannot calculate the crash rates from this DOT data because the total number of trips is not available. What we can see is that the available data is consistent over the time period 2000–2014. It is therefore reasonable to believe that the unobserved variable on "Number of Trips" is also consistent over time, and therefore that the crash rates are consistent over time.

If the rates reported by Willams justified the restrictions on teenage driving in 2001, then it is likely that the more recent rates would also justify the restrictions.

Note: All graphs were drawn with R (R Core Team, 2016). Figure 1 uses the `HH::interaction2wt` function. Figures 2 and 3 use the `vcd::mosaic` function.

References

Richard M. Heiberger (2015). *HH: Statistical Analysis and Data Display: Heiberger and Holland.* R package version 3.1-31, `http://CRAN.R-project.org/package=HH`.

Richard M. Heiberger and Burt Holland (2015). *Statistical Analysis and Data Display: An Intermediate Course with Examples in R.* Springer-Verlag, New York, second edition. `http://www.springer.com/978-1-4939-2121-8`.

David Meyer, Achim Zeileis, and Kurt Hornik (2015). *vcd: Visualizing Categorical Data.* R package version 1.4-1, `https://CRAN.R-project.org/package=vcd`.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org/`.

US Department of Transportation General Estimates System (GES) (2015). *National Automotive Sampling System (NASS).* `ftp://ftp.nhtsa.dot.gov/GES/`.

A. F. Williams (2001). Teenage passengers in motor vehicle crashes: A summary of current research. Technical report, Insurance Institute for Highway Safety, Arlington, VA.