

Racial Profiling: Focus on the Person or the Act?

Dominique Haughton

Bentley University, Waltham, USA and Toulouse School of Economics, France

Phong Nguyen

General Statistics Office, Hanoi, Vietnam and University of South Australia, Adelaide, Australia

Lynn Senne

Bentley University, Waltham, USA

In response to allegations of discrimination by police officers, the Massachusetts legislature passed an act in 2000 mandating that the Registry of Motor Vehicles record data on the race, gender, and search status of drivers who were issued a written warning or a ticket. Several previous statistical studies examined this data using regression techniques. This previous body of work, which focuses almost exclusively on demographics, concludes that the disposition of citations—tickets versus warnings—reveal racial and gender bias.

Unlike the earlier work, this paper analyzes the Registry data using Classification and Regression Trees (CART) to determine the extent, if any, of racial and gender profiling. In addition to demographic information, our models consider variables like the situation (e.g., time of day and location), background (e.g. make and model as well as registration of the car), and behavior (exceeding the speed limit and driving at a high rate of speed) as part of the CART analysis. We find that, if the seriousness of the speeding violation—high speed and significant speed over the limit—are considered, the effects of demographic factors such as race and gender disappear. Furthermore, since driving patterns are correlated with these same demographic factors, the demographic factors themselves behave as proxies for serious traffic offenses. This case is accessible to readers with an intermediate-level knowledge of statistics..

Keywords: *Decision Trees, Classification and Regression Trees (CART), Racial Profiling*

Introduction

Law enforcement agencies routinely use profiling as a forensic tool to describe the characteristics of someone who has perpetrated a crime. Such profiles can be physical (e.g., the suspect is a white male, 6' 2" tall, with light brown hair) or psychological (e.g., the crime pattern suggests that the criminal is strongly prejudiced against a

particular religious group). These profiles help guide the decisions and actions of police officers. In the 1980's, for example, the Drug Enforcement Agency developed a profile of drug couriers to help law enforcement officials stop the flow of drugs north from Florida. Following 9/11, the Department of Homeland Security established

guidelines to spot potential hijackers. Such profiling, however, is controversial.

The term “racial profiling” implies police bias. It suggests that the color of a person’s skin may be the primary criterion that leads a patrol officer to take action rather than the officer’s suspicion of a specific individual. This argument rests on the supposition that in some cases either an officer’s prejudice or institutional bias against individuals of a specific minority group may lead to discriminatory behavior (Engel, 2002, p. 21).

Some minority groups are concerned that officers primarily base their decisions to stop, cite and search motorists based on a racial profile. In short, “driving while black” may be the only reason for some officers to pull minority drivers over. Besides being inherently unfair, tickets based on prejudice penalize drivers in two ways: they must pay both a fine and higher insurance premiums. Similar concerns have been raised about gender profiling.

In response to allegations of racial as well as gender discrimination by police officers, the Massachusetts legislature passed Chapter 228 of the Acts of 2000. One of the provisions of this Act mandates that the Registry of Motor Vehicles record data on the race, gender, and search status of drivers who were issued a written warning or a ticket. Because of funding limitations, the Registry computerized these data for just two months from April 1, 2001 and May 31, 2001.

Researchers, using regression techniques to identify predictors for the disposition of citations, have generally concluded that whether a motorist receives a ticket or a warning depends on race, gender, and age. Unlike earlier statistical studies, however, this paper analyzes the Registry data using Classification and Regression Trees (CART) to determine the extent, if any, of racial and gender profiling. In addition to demographics, our research considers variables like the situation (e.g., time of day and location), background (e.g. make and model as well as registration of the car) and behavior (exceeding the speed limit and driving at a high rate of speed) as part of the CART analysis.

Literature

The news media has reported over the years on disparities in traffic stop patterns based on race and gender (Dedman et al 2003 and 2004). Other articles have reported on location (Carroll et al). Many of these articles in the popular press provide anecdotal information.

In 2004, Northeastern University’s Institute on Race and Justice issued two reports—the “Massachusetts Racial and Gender Profiling Study” and the “Massachusetts Racial and Gender Profiling Technical Report”—that review and analyze the data collected by the Registry of Motor Vehicles. Both reports describe population characteristics such as race, gender, and age for each Massachusetts town where the data was collected. The “Profiling Study” asks four questions that are commonly raised about traffic citations:

1. Are non-white drivers who are residents in a community cited more often than their representation in the residential population would predict?
2. Are non-white drivers overall cited more often than their representation in the population of people driving on the roadways would predict?
3. Once stopped are non-white drivers more likely to receive a citation than white drivers?
4. Once stopped are non-white drivers more likely to be subject to a search than white drivers? (Farrell et al, p. 2)

To answer these questions, the companion “Technical Report” provides detailed demographic analysis of:

- Motorists cited for traffic violations in Massachusetts compared to the residential census demographics (pp. 12-69)
- Motorists cited for traffic violations in Massachusetts compared to the driving population estimate (pp. 70-105)
- Motorists searched once cited for traffic violations in Massachusetts (pp. 106-188)
- Motorists cited [sic] compared to motorists warned. (pp. 189-248).

Both reports focus primarily on bivariate descriptive statistics such as the percentage of the white and minority population based on a census benchmark. The reports then compare the percentage for motorists of each race who were stopped to determine if it differed from the percentage one might expect based on the population. The two reports also compare the percentage of citations to the estimated driving population. If the percentage of either differs substantially from the percentage of the comparison population, the authors argue that bias exists.

Although the authors concentrate mainly on demographics, they briefly discuss the driving behavior of motorists that apparently prompted police officers to act. According to the authors, “While all drivers may be more likely to be cited for egregious violations of the law,

differential behavior patterns do not appear to explain away racial differences in citation and warning rates” (Study, Farrell et al, p. 29): they cite the following data to support their point (ibid).

Although the authors consider the difference in ticketing based on race to be statistically significant, they do not include the values of their test statistics such as the p-value. Nonetheless, the authors conclude that “there is no indication from the statewide data that differential violation rates explain away racial differences in dispositions” (Study, Farrell et al, p. 33). In summary, they conclude that their findings support claims of racial profiling.

Table 1. Race versus Speeding Patterns

Speed Relative to Limit	Percentage of Drivers Ticketed		Statistically Significant Difference
	White	Non-White	
≤ 15 MPH	73.6%	82.6%	9 percentage points
> 15 MPH	81.6%	86.8%	5.2 percentage points

The Boston Globe also published a report that analyzes patterns in traffic citations. The report, posted on the newspaper’s Web site in 2003, examines data from every police department in Massachusetts to determine whether race, gender or age played a role in determining which motorists received tickets or warnings. Although this report is no longer available at the Boston Globe Web site, the report and the accompanying data set can be found with this paper on the CSBIGS site. The report also includes analysis by an independent statistician, Prof. Elaine I. Allen of Babson College. Allen employs both descriptive statics as well as bivariate and multivariate models to examine factors that influence the type of citation issued. In addition to calculating the odds ratio, Allen uses logistic regression models to test the effect of race on ticketing while controlling for police, gender, speed above the posted limit, and age. She presents her results in a table that lists statistical significance, the odds ratio and the confidence interval for one of the several logistic regressions she ran. Allen concludes:

There are strong racial differences in whether you are ticketed or just warned. Minorities are significantly more likely to be ticketed even when we take into account the effects of other variables that may influence whether you are ticketed such as age, gender, amount over the speed limit you were traveling, type of police that stopped you and whether this occurred in your own neighborhood or whether you are from outside Massachusetts. There is variability by minority group, with Latinos the most

likely to receive a ticket vs. a warning, followed by blacks and Asians (Technical Report, p.113).

Allen also observes that, among all police departments, state troopers are most likely to give tickets rather than warnings. However, race, gender and age do not appear to affect the disposition of the citation. Although the percentages of minorities, men, and younger drivers who receive tickets from state troopers are slightly higher than the percentages for whites, women and older drivers, the differences are not statistically significant (Technical Report, p.114).

Another report, “How to Correctly Collect and Analyze Racial Profiling Data: Your Reputation Depends on It!” examines problems in data collection and analysis. Published by the Department of Justice, the report observes that many “studies have recognized some of the difficulties in obtaining proper comparison groups and in conducting appropriate statistical tests of the effect of race on stops and searches.” (McMahon et al, p.31). The authors note that some studies “refer to the comparison group issue as the ‘denominator’ problem, as if the primary analytical issue in this research was one of long division or finding just the right measure for the denominator” (McMahon et al, p. 32). The Northeastern study faces this kind of denominator problem when it compares the motorists cited for traffic violations to the residential census demographics or driving population estimates. The Department of Justice report recommends that future studies move beyond bivariate analysis using comparison groups and instead use multivariate methodologies. In addition, it suggests that researchers place more emphasis on carefully designing studies based on thoughtful consideration of analytical issues.

In “Racial Profiling? A Multivariate Analysis of Police Traffic Stop Data” Michael R. Smith and Matthew Petrocelli examine traffic stop data collected in Richmond, Virginia. They find that the citation pattern is more complicated than it appears to be on the surface. Using logistic regression to analyze three factors—Driver (age, race, gender), Officer (age, race, gender, and years of service), and Event (time, reason, location and disposition of stop and the type of search if one is conducted)—they conclude that “Minorities were disproportionately stopped compared with their percentage in the driving-eligible population” (Smith, p. 4). However, they also find that “minority drivers were 50% more likely than white drivers to receive a warning rather than be subjected to a legal sanction whereas whites were more likely to be ticketed or arrested” (Smith, p. 19).

In the U.S. General Accounting Office’s 2006 publication “Racial Profiling: Limited Data Available on Motorist Stops” (2006), the authors suggests that more research needs to be conducted to determine which factors other than race may have occasion the traffic stop (p. 110). They report that they find no “conclusive empirical data from a social science standpoint to determine the extent to which racial profiling may occur” (p. 109). In particular, the authors note that even some “well-designed studies made no distinction between the seriousness of different traffic violations” (p. 110). They also observe that there is little comparative research on traffic violations committed by different racial groups, including possible differences in the type or seriousness of traffic violations. In addition, none of the studies provided information on which traffic violations, if any, were more likely to prompt a stop (p. 110).

In “Theory and Racial Profiling: Shortcomings and Future Directions in Research” Engel et al. argue that “much of the literature on racial profiling is misleading, fails to include crucial explanatory variables, and provides a limited understanding of the phenomenon”(Engel et al, p. 13). They observe that, while purely descriptive research may be interesting and useful, it is not scientific research... [because] the underlying theory guiding racial profiling research is implicit. It is implied that officers make decisions on the basis of citizens’ race (p. 8).

The authors argue that “theories that are not explicitly stated often lead to ‘sloppy’ investigations,” (p. 8), and they call for a more scientific structured approach to data collection and analysis.

Ticket or Warning? Does Profiling Occur on Massachusetts Roads?

The Data

This paper examines the database compiled by the Massachusetts Registry of Motor Vehicles and released under the state’s public records law at the request of the Boston Globe. It contains data on all traffic tickets and warnings written in the State of Massachusetts from April 1, 2001, through May 31, 2001.

The data collected provides information about the offense, the police officer, the driver, and the driver’s vehicle (see Table 2).

Simple Bivariate Analysis

Based on a simple bivariate comparison of race and the type of citation (ticket or warning), it appears that minorities are ticketed more frequently than whites.

Table 2. Data Collected by the Massachusetts Registry of Motor Vehicles

Offense	Driver
Citation type (e.g., warning, ticket, arrest)	Driver license number and state
Location (town or neighborhood)	License class, commercial license
Citation number	Race
Date, time	Sex
Offense (law chapter/section)	Year of birth
Offense description (e.g., speeding, failure to yield)	Home Zip code
Amount of fine	Operator or owner of vehicle
MPH for speeding citations	Vehicle
MPH of speed zone	Plate number and state
Non-inventory vehicle search	Commercial vehicle
Accident	
Court, supplemental (indicates court-issued tickets)	Hazardous-materials vehicle
Reversed (tickets successfully challenged in court)	Make
Police Officer	Model
Police agency	Year
Officer ID (masked)	Color

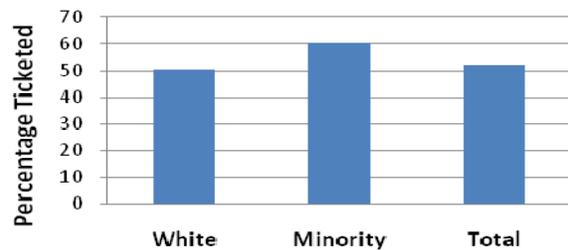


Figure 1. Percentage of ticketed drivers by race

When racial groups are broken down by gender, ticketing also seems to point to a discriminatory pattern (Figure 2).

Within racial groups, white women are ticketed less frequently than white men and minority women less than minority men. Within gender, minority women are ticketed more frequently than white women and minority men more than white men. In fact, minority men receive the highest number of tickets.

In summary, these simple comparisons appear to indicate bias in the disposition of citations.

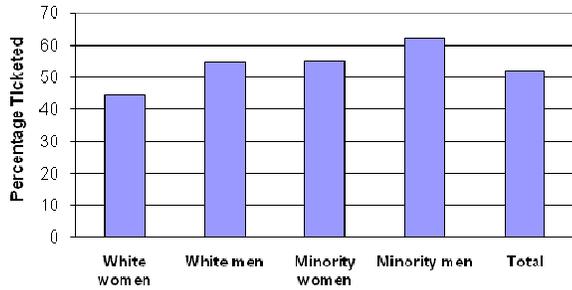


Figure 2. Percentage of ticketed drivers by gender and minority status

Analysis Using CART (Classification and Regression Trees)

Given that the simple percentage charts appear to indicate racial and gender bias, we decided to use Classification and Regression Trees (CART) to identify the main determinants of the likelihood of a driver receiving a ticket.

The CART methodology was originally introduced by Breiman et al. (1984) and has since been widely applied to a very extensive range of areas, from medical to business. A simple internet search reveals more than 14,000 citations of the Breiman et al. book (1984). CART is part of a class of techniques sometimes referred to as “Decision Trees” (not to be confused with the decision trees where one computes conditional probabilities and summarizes them into a tree graph). Decision trees are predictive methods, in that a target variable is singled out and the hope is to build a model with suitable predictors that explains or predicts the target variable well. In that respect, decision trees join the rank of more traditional predictive modeling techniques such a regression models (linear or logistic, for example).

Another main direction within the class of decision trees methods is that referred to as CHAID (Chi-square Automatic Interaction Detector). The original CHAID algorithm, introduced by Kass (1983) and implemented and further developed by Magidson (1993), relies on a collection of chi-square tests of independence between the target variable and each predictor. Because of the ease of interpretation of the trees produced by CART or CHAID (relative to that of interpreting regression models), and because of the ability of both techniques to identify main predictors and interactions, the methods have become extremely popular in areas such as market research and predictive analytics in business. For a comparison of the properties of CART and CHAID, see for example Haughton and Oulabi (1993).

The original CART algorithm proposed by Breiman et al. (1984) is fully implemented in the CART tool from Salford Systems (www.salford-systems.com), and the CHAID algorithm is fully implemented in the Answer Tree tool of SPSS (SPSS Answer Tree 2010). Both algorithms are available in some form in most data mining packages (see for example Deichmann et al. 2003 for a review a data mining packages with decision tree examples). A version of CART exists in the form of an the R package RPART.

CART uses two different types of analysis: classification and regression. The CART algorithm involves binary recursive partitioning to identify predictor variables, i.e., to identify those variables that affect the characteristic we want to predict. The algorithm first creates a list of trees from the smallest tree with only one node to the largest tree with as many nodes as observations, and then prunes the tree to select the one that predicts the dependent variable best on an independent test sample.

A CART tree examines all possible variables as follows. For each node, the CART analysis splits the data into two sub-groups with as little variability as possible in the target variable. To achieve this, at each step, CART selects from among all possible candidate predictors those that split in such a way as to minimize the variability of the target variable within the two new nodes. CART measures the success of the split by the Gini coefficient for a categorical target variable (classification tree) and the within sum of squares for a continuous target variable (a regression tree). Figure 3 shows an example of a node on a tree.

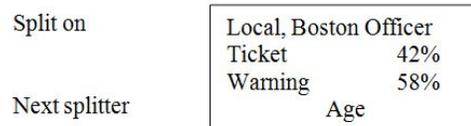


Figure 3. How to Read a CART tree

The first line of a CART node contains the splitter, the break point for dividing the data into two groups. The last line contains the next splitter for the classification.

In a demographic data set, for example, the first node in the classification tree might be gender, a binary variable. The CART analysis then partitions each of the two child nodes (male and female) into two more child nodes. For this second set of child nodes, the next classification might be either white or minority, another binary variable. If a predictor involves more than two categories, CART merges some of the categories in such a way as to obtain two child nodes that are as homogeneous as possible in the target variable. If X is any continuous independent

variable and C is any value taken by X, the CART algorithm uses binary splits such as $X \leq C$ or $X > C$. For example, CART might divide the data into two groups such as ≤ 40 years old or > 40 . On the other hand, if X is a categorical variable coded 1 through 6, then the sample might be split so that $X=2,6$ versus $X=1,3,4,5$. The classification is recursive because the binary partitioning process can be repeated multiple times until CART has determined that no further split can improve the homogeneity of nodes or until some user-set limit has been reached.

Because CART makes no assumptions about the population under investigation, it offers classification advantages over other techniques like logistic regression:

...it is inherently non-parametric. In other words no assumptions are made regarding the underlying distribution of values of the predictor variables. Thus CART can handle numerical data that are highly skewed or multi-modal as well as categorical predictors with either ordinal or non-ordinal structure (Lewis, p. 5).

The CART methodology can also help identify interactions and complex relationships between the target variables and predictors. In Figure 4, we can see that the effect of age on the likelihood of a ticket depends on whether the officer is a local officer or a State Trooper; this exemplifies an interaction between the two predictors Agency and Age. In Figure 5, the fact that MPH over the limit appears in two consecutive splits (at 16.5 and then 9.5 MPH over the limit) would indicate the presence of a complex relationship between the likelihood of a ticket and the predictor MPH over the limit.

In our CART analysis we use the Gini index to measure the homogeneity of classes in the terminal nodes. This index “reaches zero when only one class is present at a node. [In other words]...the Gini index is equal to zero if all cases in a node belong to the same class” (statsoft.com). In our case, the Gini index would be zero if a node contained only warnings or only tickets.

The CART tree in Figure 4 explores all predictors in the data set except speed information. Each block shows the percentage of tickets versus warnings.

The classification and regression tree shows that race, sex and age affect whether a motorist receives a ticket or a warning. Older members of minority groups are about equally likely to receive a ticket instead of a warning from a local police officer. However, whites in that age group

have a smaller (than minorities) probability (36%) of receiving a ticket from a local police officer. For younger drivers it seems that gender is a stronger predictor than race.

The agency issuing the ticket is also a good predictor of the type of citation. Although state troopers issued fewer citations than other agencies, they wrote the highest percentage of tickets. In fact, they gave motorists tickets in more than three quarters of their traffic stops. A quick review of the underlying data reveals that more than half of these tickets were issued to motorists driving 74.5 MPH or faster. This preliminary analysis suggests that absolute speed appears to be important.

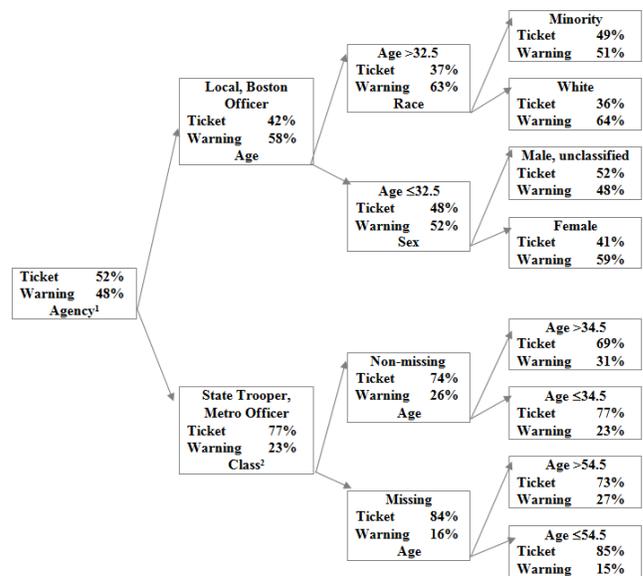


Figure 4. Exploratory CART tree with race, gender and age effects

Notes: ¹Agency: Law enforcement agency (Boston, Police, Local police outside Boston, Metro Police Lower Basin Department, State Trooper); ²Class Driver’s License Type (non-missing: type of license such as passenger vehicle noted on the ticket, missing: type of license not noted)

In our second CART analysis we include not only demographic information but data on the event as well. In other words, we explicitly examine all the data collected by the Registry of Motor Vehicles. Race, gender and age are still included but, in addition, the analysis also looks at other factors like time of day, state plate registration, location, type of infraction, driver’s speed and posted speed limit. Since all the tickets and warnings were issued for speeding, the type of infraction was not relevant. The CART analysis in Figure 5 indicates that both high speed and excessive speed, i.e., speeds above the posted limit are good predictors of citations.

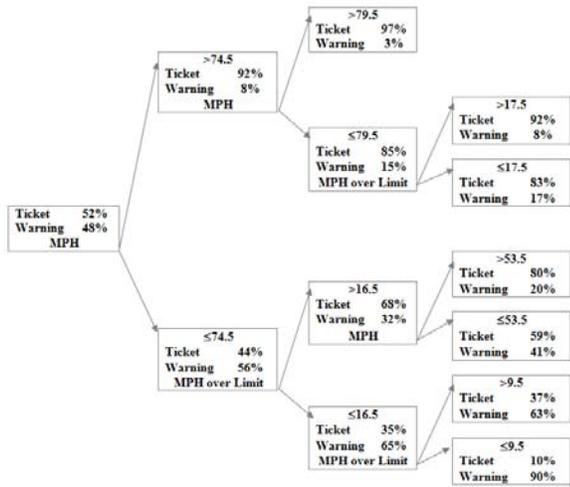


Figure 5. Exploratory CART tree including demographic and speed factors

Although race, gender and age are included in the analysis, these factors no longer appear on the tree. Since the two speed variables are the only factors that appear in the CART tree in Figure 5, they merit further analysis. High speeds almost always led police officers to issues more tickets than warnings, and speeds above the posted limit also led to a higher percentage of tickets. Table 3 shows how the percentage of tickets issued declines as both the actual speed as well as the speed over the limit falls.

Table 3. Fewer tickets for less serious offenses

Speed (MPH)	MPH above Limit	% Tickets
>79.5	n/a	97%
≤79.5	>17.5	92%
≤74.5 but >53.5	>16.5	80%
≤74.5 but ≤53.5	>16.5	59%
≤74.5	≤9.5	10%

For example, at speeds above 79.5 MPH and above, motorists are likely to receive tickets 97% of the time. This pattern is consistent with the fact that state troopers, who more frequently patrol highways with higher speed limits than city streets, issued more tickets than local officers. At speeds <79.5 MPH, the speed above the limit becomes a determining factor. In fact, if a driver is traveling not only ≤16.5 MPH but also ≤9.5 MPH over the limit, police issued tickets 10% of the time.

To summarize, with two measures of excessive speed in the CART analysis, the second decision tree indicates that the seriousness of the violation takes precedence over demographics. In this analysis all the demographic effects as predictors of the disposition of the citation have disappeared at five levels down (three levels are displayed in Figure 5). If both speed factors are included in the

analysis, demographic factors such as race, gender and age are no longer predictors of who will receive a ticket or who will receive a warning.

Given the disappearance of demographic factors altogether, we looked more closely at the relationship between the speed variables and the demographics. Specifically, in order to see if the speed variables are masking the effect of the demographics, we constructed several boxplots. Each plot displays the speeds at which drivers were traveling when they were stopped.

For those individuals pulled over by police officers for speeding violations the plots in Figures 6 and 7 show a consistent pattern in average driving habits across race.

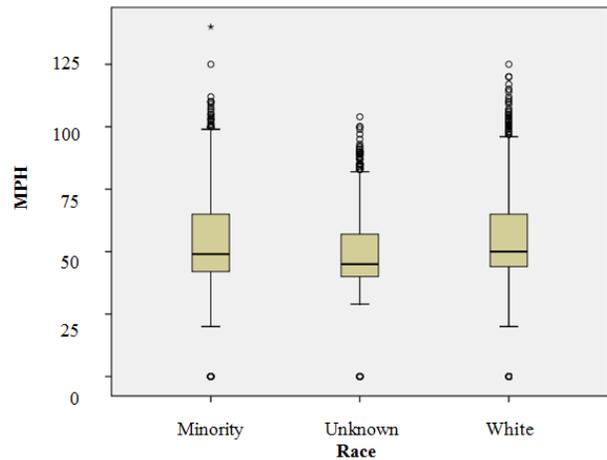


Figure 6. Similar speed pattern for minority and white drivers stopped by police officers

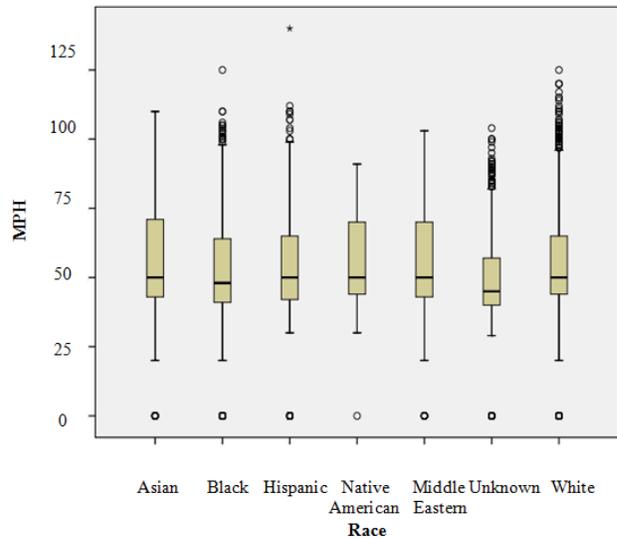


Figure 7. Speed patterns by race for motorists stopped by police officers: only mildly different

The medians within each of the two charts are similar across the classes. The high speeds within the upper quartiles are broadly similar for whites and minorities (Figure 6) although differences do appear when different minority groups are plotted separately (Figure 7).

We also looked at driving patterns by gender within age (Figure 8). Median speeds, which fall slightly for older drivers, are similar across the classes. The spread of the upper quartiles also decline with age. However, for all age groups above 25 the top speeds for men are consistently higher than those for women; top speeds for men also decline with age.

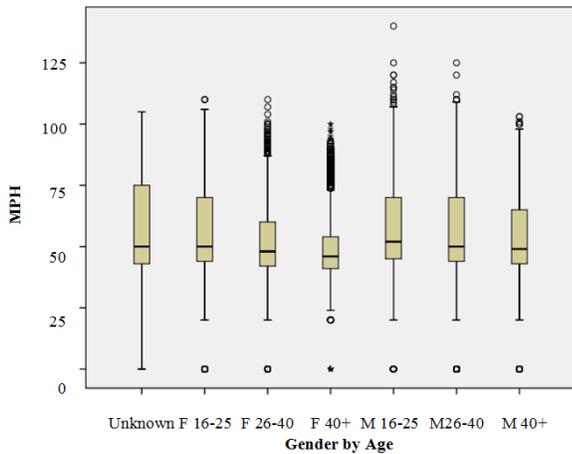


Figure 8. Different speed patterns for men and women by age patterns by race for motorists stopped by police officers: only mildly different

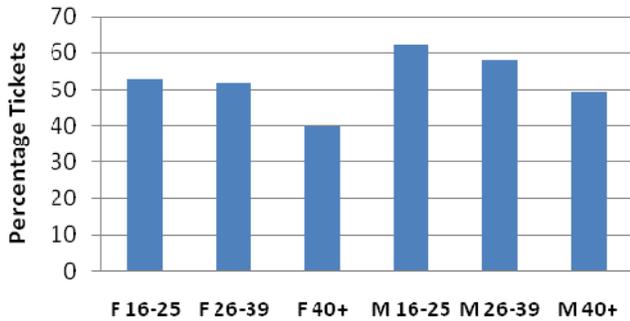


Figure 9. Ticketing follows the speed pattern but not exactly

Although the number of tickets issued declines with age, speed also declines slightly—a fact that may explain why older drivers receive fewer tickets. However, for all age groups, men in the upper quartile were typically driving faster than women when they were ticketed—a fact that may explain why women receive fewer tickets than men.

On the whole the box plots reveal that different demographic groups do indeed exhibit different driving habits. Consequently, demographic data can act as a proxies for behavior when we attempt to explain who gets ticketed and who gets warned. Models built on demographic data can and do find differences in the disposition of citations that may appear to indicate bias when in fact it is conceivable that the disposition is based on significant differences in the underlying driving patterns.

Conclusions

Numerous studies have concentrated on speed but only speed above the posted limit. However, as other researchers have pointed out, this focus on one measure of speeding does not take into account the seriousness of the offense. If both the driver's actual speed and the speed above the limit are considered, the data do not show a pattern of discrimination. Rather the seriousness of the violation takes precedence over demographic factors. Furthermore, since driving patterns are correlated with demographic factors, the demographic factors themselves can act as proxies for the level of seriousness of traffic offenses. In short, if two measures of speed are included—absolute and excessive—the effects of race, gender and age disappear in our analyses.

In addition, any statistical model must take into account a police officer's mindset. Enforcing traffic laws demands flexibility and adaptability. Streets and highways are part of a dynamic environment that requires police officers to make situational decisions. While statistical models can examine patterns in the citation data, to build a valid model of the citation process researchers need to look at the complex human decision-making process that leads to a ticket instead of a warning. If officers issue tickets versus warnings based on the assessment of how dangerously an individual is driving rather than demographics, then race, age, and gender will not be good predictors of the disposition of the citation. In short, if important situational variables are not included in a statistical model, the results of the model may be misleading.

Finally, it is of interest to note that if a model is misspecified, for example if a non-linear relationship between a target and a predictor fails to be included in the model, it can happen that other predictors enter the model significantly, so to speak to attempt to correct this misspecification. When the correct specification is included in the model, such predictors can become insignificant (see for example Haughton and Haughton 1997).

REFERENCES

- Boston Globe 2003. Boston Globe Analysis of Traffic Tickets and Warnings. *Boston Globe*. Accessed July 27, 2003.
- Breiman, L., Friedman J. H., Olshen R. A. & Stone C. J. 1984. *Classification and Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software.
- Carroll, M. M. 2008, February 24. Warning: When It Comes to Speeding Tickets, It Matters Where You Get Stopped. *Boston Globe*.
- Cooper, S. 2006. A Closer Look at Racial Profiling. In S. J. Miller, *Racial Profiling: Issues, Data and Analyzes*. (pp. 25-29). New York: Nova Science Publications.
- Dedman, B. 2003, July 20. Race, Sex, and Age Drive Ticketing. *Boston Globe*, p. A1ff.
- Dedman, B. 2003, March. Romney Backs Tracking Racial Profiling. *Boston Globe*, p. 3.
- Dedman, B. 2004, May 4. Speed Trap: Who Gets a Ticket, Who Gets a Break? *Boston Globe*.
- Deichmann, Haughton D., Eshghi A., Sayek S., Teebagy N. & Topi H. 2003. A review of software packages for data mining. *American Statistician*, 57 (4), 290-309.
- Del Carmen, A. 2008. *Racial Profiling in America*. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Engel, R. S. 2002. Theory and Racial Profiling: Shortcomings and Future Directions in Research. *Justice Quarterly*, 19 (2), 249-274.
- Farrell, A. J. 2004. *Massachusetts Racial and Gender Profiling Study*. Boston: Northeastern University Institute on Race and Justice.
- Haughton, D. & Haughton, J. 1997. Explaining child nutrition in Vietnam. *Economic Development and Cultural Change*, 45, 541-556.
- Haughton, D. & Oulabi, S. 1993. Direct marketing modeling with CART and CHAID. *Journal of Direct Marketing*, 7(3), 16-26.
- Hill, T. A. 2007. Classification and Regression Trees (C&RT). In *STATISTICS Methods and Applications* (pp. 1-13). Tulsa, OK: StatSoft.
- Holbert, S. L. 2006. Making Sense of the Data. *Law Enforcement Technology*, 33 (7), 40-45.
- Holbert, S. L. 2004. *The Color of Guilt and Innocence*. San Ramon: Page Marque Press.
- Kass, G. 1980. An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, 29, 119-127.
- Levenson, M. 2007, September 9. Study of Traffic Stops is Derailed: Police Lagging on Racial Data Collection. *Boston Globe*.
- Lewis, R. J. 2000. An Introduction to Classification and Regression Tree (CART) Analysis. 2000 *Annual Meeting of the Society for Academic Emergency Medicine* (pp. 1-14). San Francisco: Society for Academic Emergency Medicine.
- Magidson, J. 1993. The use of the new ordinal algorithm in CHAID to target profitable segments. *Journal of Database Marketing*, 1, 29-48.
- McMahon, J. J. 2002. *How to Correctly Collect and Analyze Racial Profiling Data: Your Reputation Depends on It*. Washington: U.S. Dept. of Justice, Office of Community Oriented Policing Services.
- Nelson, D. E. 2009. *Racial Profiling (Opposing Viewpoints Series)*. New Haven, MI: Gale Cengage Learning.
- Northeastern University Institute on Race and Justice 2004. *Massachusetts Racial and Gender Profiling Technical Report*. Boston: Northeastern University Institute on Race and Justice.
- Ramirez, D. J. 2006. A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned. In S. J. Miller, *Racial Profiling: Issues, Data and Analyzes* (pp. 57-108). New York: Nova Science Publications.
- Rubin, J. 2008, April 30. LAPD Says Race Not a Factor. *Los Angeles Times*, p. B1f.
- Smith, M. M. 2001. Racial Profiling? A Multivariate Analysis of Police Traffic Stop Data. *Police Quarterly*, 4 (1), 4-27.
- U.S. General Accounting Office. 2006. Racial Profiling: Limited Data Available on Motorist Stops. In S. J. Muffler (Ed.), *Racial Profiling: Issues, Data and Analyzes* (pp. 109-153). New York: Nova Science Publishers.
- SPSS Answer Tree.
<http://www.spss.com/software/statistics/answertree/>, accessed March 2010.
- Therneau, T., Atkinson, B. & Ripley, B. 2010. RPART, <http://cran.r-project.org/web/packages/rpart/rpart.pdf>, accessed March 2010.

Correspondence: dhaughton@bentley.edu