ENHANCING THE UNDERSTANDING OF STATISTICS THROUGH STUDENT-CONDUCTED DATA GATHERING

Mary W. Gray
American University, Washington DC, USA
mgray@american.edu

The role of real data in helping to develop statistical thinking in students has been generally recognized, but most often existing data is analyzed. A series of cooperative projects among graduate and undergraduate students demonstrated that the collection of data can also play an important role in establishing enthusiasm as well as learning among a heterogeneous group of students. Topics selected, including an examination of the impact of voter ID laws in the state of Virginia, highlight the engagement of students in the study of evidence-based policy

INTRODUCTION

For many years, the importance of undergraduate students learning to work with real data has been understood (Cobb 1991). Incorporating consulting activities (Smucker and Bailer 2015) and using case study techniques (Valaitis and Gray 2006) have also become widespread. With the current emphasis on evidence-based public policy, data relevant to policy decisions are increasingly included in statistics learning. While the resulting improvement in statistical thinking is cited (Hardin *et al* 2015), less has been said about substantive increases in civic education and understanding of the implications of evidence-based decision making for the students and the broader public. A series of cooperative efforts of graduate (design) and undergraduate (execution) students in exit polls and other forms of data gathering on such issues of public policy as voting, education, employment, environment and health has expanded not-only the statistical knowledge and skills of students, but has contributed to a broader understanding of the role of statistics in legislative, executive or judicial decision-making.

METHOD

At American University undergraduate students in all disciplines – liberal arts and science, business, communication, public affairs, and international service – are required to take at least one quantitative course, generally in elementary statistics. This makes for classes composed of a heterogeneous group of students in both background and interests. As is often the case with required courses not closely related to their interests, students may express little enthusiasm for, or engagement in, the learning of statistics or the importance of doing so either in their own fields of study or as participants in civil society.

At the graduate and advanced undergraduate level, courses in the Department of Mathematics and Statistics provide training in a variety of techniques, both through an overview of statistical methods and in specialized courses, in the case at hand in survey sampling, to students in a variety of disciplines in addition to statistics, such as public health, psychology, economics, and international relations. There is a strong consensus with the view that dealing with real data is important in learning statistical thinking and analysis (ASA 2014), and even more so with student-generated data. Although often the undergraduate students design their own projects involving the collection and analysis of data, in some cases, in

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE),
July 2016, Berlin, Germany. iase-web.org/Conference_Proceedings.php

particular the example discussed here in detail, the students in the survey sampling class are responsible for design and complex analysis and the undergraduates for collection and the more elementary parts of the analysis.

The major effort addressed here evolved over a series of studies of the impact of so-called Voter ID laws, involving students at several other universities as well as at American University. Whether stringent voter ID laws should govern access to voting has become a political issue; while some see it as preventing voter fraud, others see it as an attempt to make it difficult for segments of the electorate to exercise their right to vote, echoing a past history of discriminatory laws and practices. Thirty-four of fifty states have some sort of Voter ID laws, including the state of Virginia, where currently acceptable forms of ID are

> Valid Virginia Driver's License or ID Card
> Valid United States Passport
> Other government-issued photo identification cards (must be issued by US Government, the Commonwealth of Virginia, or a political subdivision of the Commonwealth) e.g., Permit to carry a concealed weapon
> Tribal enrollment or other tribal ID issued by one of 11 tribes recognized by the Commonwealth of Virginia
> Valid college, university or school student photo identification card (must be from an institution located in Virginia)
> Employee identification card containing a photograph of the voter and issued by an employer of the voter in the ordinary course of the employer's business

In the fall of 2013 there were elections in the state of Virginia involving the state-wide offices of governor, lieutenant governor, and attorney general. The candidate for attorney general won by a few hundred of the over two million votes cast, making clear that even a few votes can make a difference in the outcome. At the election the following year involving national offices a survey (n = 670) was conducted at a sample of precincts stratified by race and income based on US census figures in the city of Alexandria and the counties of Arlington, Fairfax, and the portion of Loudon county adjacent to the other region as well as in the city of Washington and in suburban Maryland, where voters' names are required to be on the rolls at the precincts but there is no voter ID law as such (see Figures 1, 2).

Survey Universe Washington DC Metro Area

Northern Virginia **Maryland**

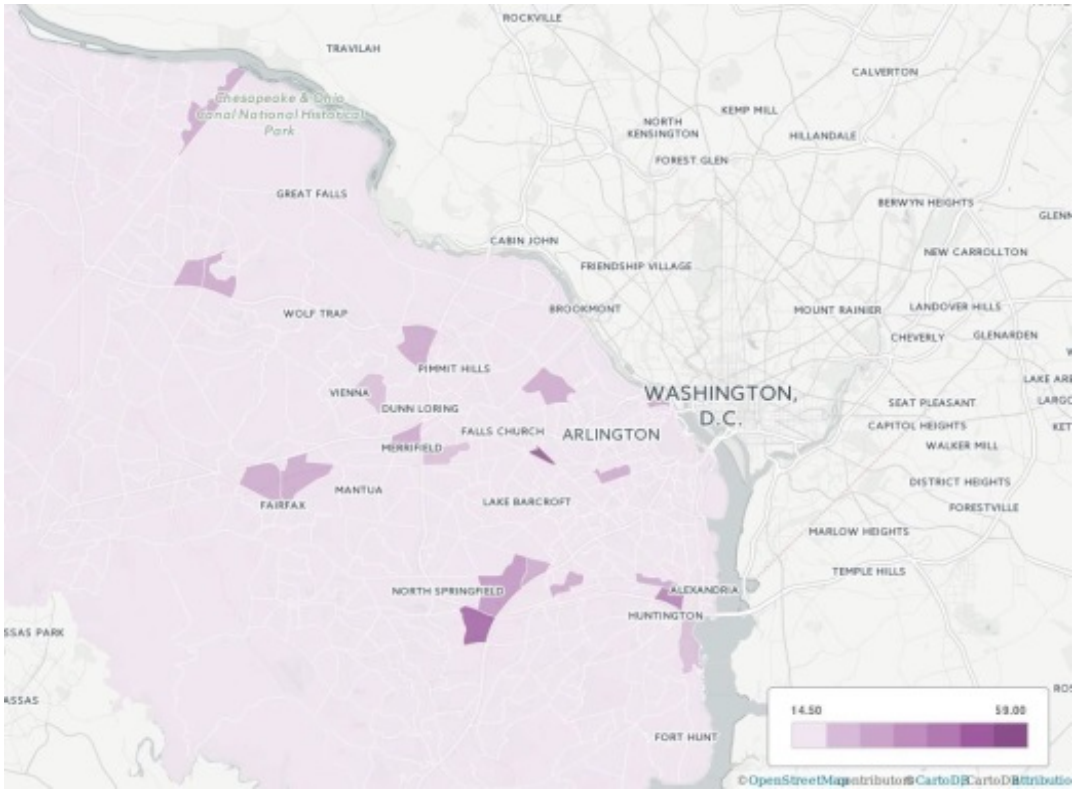


Figure 1. Tracts surveyed

Darkness of the purple indicates the size of the sample

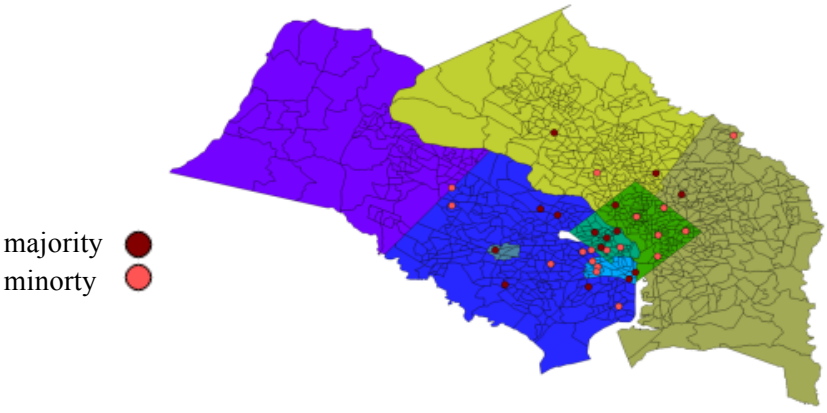
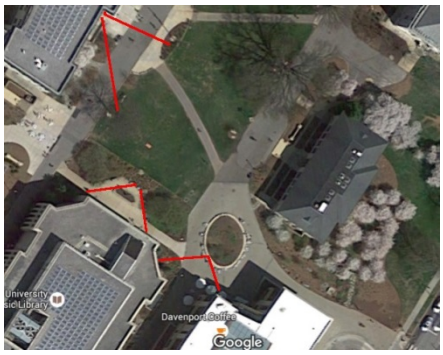


Figure 2: Precints sampled by race

Other projects reflected the diverse interests of students. For example, one survey compared the prices of basic items such as toothpaste in branches of the same pharmacy chain in the upper middle income sections of Washington DC with those of the same size and brand items in the lower income

areas. Another project involved the testing of the announced rationale for various school closing, namely that the selection was based on enrollments and not race, as challengers asserted. Sports statistics are a major interest among students, but the course project involved applying to lacrosse the sort of analysis seen more commonly in baseball or basketball. Another individual project involved using body exposure to predict temperature (Figure 3). Using the resources of Washington's National Zoo, a group of environmental science students studied the behavior of lemurs, while others interested in the environment studied water quality.

Data from study on campus of American University



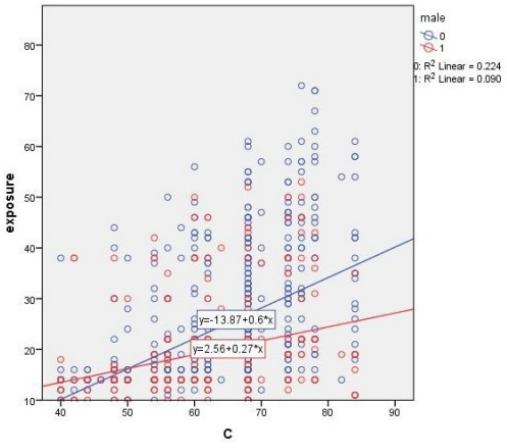


Figure 3. Relation between body exposure and temperature

Data from study at Washington National Zoo

Correlations

Control Variables			1=Red Ruffed, 2=Ringtail	1=IN, 2=LA, 3=LT, 4=FO, 5=SI, 6=OT
1=Keeper in exhibit, 2=OOV, 3=Other & Time & Date	1=Red Ruffed, 2=Ringtail	Correlation	1.000	.047
		Significance (2-tailed)	.	.388
		df	0	335
	1=IN, 2=LA, 3=LT, 4=FO, 5=SI, 6=OT	Correlation	.047	1.000
		Significance (2-tailed)	.388	.
		df	335	0

Figure 4. Behavioral differences between captive red ruffed and ring-tailed lemurs

Drainage Areas and Water Quality in Washington D.C

1 = Clear, 2 = Cloudy, 3 = Opaque * 1 = Fairidge, 2 = Fairfax, 3 = Stedwick Crosstabulation

Count

		1 = Fairidge, 2 = Fairfax, 3 = Stedwick			
		1.00	2.00	3.00	Total
1 = Clear, 2 = Cloudy, 3 = Opaque	1.00	8	12	6	26
	2.00	6	18	9	33
	3.00	20	145	70	235
Total		34	175	85	294

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.941[a]	4	.012
Likelihood Ratio	10.417	4	.034
Linear-by-Linear Association	5.865	1	.015
N of Valid Cases	294		

a. 2 cells (22.2%) have expected count less than 5. The minimum expected count is 3.01.

Figure 5. Site comparison

RESULTS

We present in some detail the results of the voter ID project. 1.8% percent of the voters, all in low income minority precincts in Virginia (minority precincts were those where the African American, Latino/Latina, and Asian populations together constituted a majority) were not allowed to vote because they lacked acceptable Voter ID; no one was turned away in Washington or Maryland. Students at Southern Methodist University in Dallas, Texas, and Swarthmore College in suburban Philadelphia, Pennsylvania, conducted similar surveys in their urban areas with similar results.

Of course the question arises of whether many voters did not show up to vote at the precincts because they were uncertain as to whether they had acceptable forms of identification. Thus in the fall of 2015 a survey was conducted to determine whether Virginians of voting age knew of the requirements of the voter ID law and whether they had an acceptable form of ID. Survey locations came from a list of groceries and pharmacies stratified by race and income. Because there was speculation that students and seniors might be particularly adversely affected by the law, separate samples were drawn at two retirement communities (one majority, one minority) and at two college campuses. In addition, a community in the Shenandoah area to the southwest of Washington was surveyed for comparison purposes. Interviewers asked two questions:

Are you aware of the Virginia voter ID law?
Do you have one of the specified forms of voter ID?
(with the acceptable forms of ID listed)

In the core districts of Northern Virginia, it was found that 57.2% of those interviewed were unaware of the law, but 80.2% currently had an acceptable form of ID (n = 572). In the Shenandoah site, about 10% more were aware of the law and about 5% more had IDs.

There were no substantial differences among the students or those in the retirement communities in comparison with the general results. The gender, race and age of individual interviewees were recorded; although there were no substantial gender differences, older voters were more likely to be aware of the law and to have IDs. The striking result was that the majority voters were substantially more likely than minority voters to know of the law (60.2% to 28.8%) and to have acceptable IDs (95% to 55.5%).

The school closures were indeed strongly related to enrollment data, the prices of items in low income districts were significantly higher than in more affluent areas of the city, while not as reliable as a thermometer, exposure is related to temperature, and behavior of lemurs in captivity was connected to that in natural habitats in somewhat unexpected ways. Unfortunately, there are insufficient data from classes where the projects were not used to judge whether the number of students who continue to study and use statistics in their research increases significantly. Here are a few of the diverse results.

DISCUSSION

Whereas the issue of voter ID has been the subject of considerable political discussion in the US, there has been no systematic study of the possible impact of such laws. Speculation has rather been based on gross demographic information about voter turnout, the composition from existing data of the groups known to be holding or not holding some form of ID, and the difficulties in obtaining acceptable IDs due to distance to places where they are issued, lack of public transportation or personal vehicle, and similar issues. It was a revelation to the largely upper middle class American University undergraduates to learn that not everyone has some form of ID, accustomed as they are to travel requirements. Partisan

organizations and non-partisan groups such as the League of Women Voters obviously find data concerning voter IDs to be important in a variety of "Get out the Vote" contexts; such evidence ought to influence legislators dedicated to evidence-based policy decisions as well. In 2014 similar results were obtained by student surveys in the Dallas TX and Philadelphia PA metropolitan areas; however, commercial and governmental surveys have not adequately address the impact of voter ID laws or the more general voter suppression efforts in areas of the United States.

Research before and after the actual projects engaged the students' interest and understanding of societal problems as well as the use of statistics, especially with respect to voter and voter suppression issues. Also, the underlying cause of the school closure data was the 2009 recession that forced many middle class parents to enroll their students in public rather than private schools.

However, for statistical education purposes the extent to which these projects engage students who usually tend to be less than enthusiastic is the important result. Although so far the long term results are not known, we do know that students from the basic statistics courses have subsequently become engaged in civic society through taking a role in lobbying and voter registration activities and have in some cases changed their majors or at least taken more statistics than the initial course in which they engaged in this or other activities described below. The graduate students who designed the surveys and help train the interviewers have gone on to secure jobs in government or commercial firms or to use surveys in their own research, but as most would do so in any case, one cannot isolate the effects of the projects

FURTHER WORK

Through contacts with other disciplines throughout the university and with non-profit or other interests in the city we have organized similar projects on topics of interest when recognizing an opportunity to engage students in the analysis and often the collection of real data. In particular, one project involved identifying whether phone numbers that were business phone numbers were answered by real people. This is an important issue in states with laws forbidding random calling of phone numbers that apply to any phone numbers answered by real people, including business numbers, as contrasted with being fax or other non-voice lines, answered by recorded announcements.

A useful project that gained attention locally concerned the closure of a number of public elementary schools, described as being due to low enrollment. Allegations were made that the closures were based at least in part on race as the schools that were closed had heavy minority enrollment. However, data collected by students confirmed that the closed schools had low enrollments, in some cases less than 50% of capacity, whereas the areas in which new schools were opened or existing schools enlarged were over enrolled. Residential segregation and gentrification could be identified as the factors in the pattern of low enrollments; the growth in enrollment in some areas followed upon the 2008 economic turndown that caused many middle-income families to move their children from private to public schools. The students not only learned statistical thinking, but how evidence-based policy could be formulated and substantiated.

One food cost survey demonstrated that the same basic items were more expensive in low-income parts of the area than in high-income districts, even in chain stores with locations in both, and led to some price adjustments. However, the attempts to contrast costs of "healthy" food versus "not-very-healthy" food were frustrated by the absence of many items in the first category (such as fresh fruit and vegetables) in low income areas.

In light of a major election scheduled for this fall in the United States, further investigation of voter suppression such as cuts in voting days or hours in certain districts and location and maintenance of sites where IDs may be obtained is planned, with the intention of gathering participation from other locales in the belief that early on it is important for students to learn not only statistical thinking and techniques but commitment to civic society.

REFERENCES

American Statistical Association (ASA) (2014). *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science,* Alexandria, VA. *https://www.amstat.org/asa/education/Curriculum-Guidelines-for-Undergraduate-Programs-in-Statistical-Science.aspx*.

Cobb, G. (1991). Teaching Statistics: More Data, Less Lecturing, *Amstat News,* 182, 4.

Hardin, J., Hoerl, R., Horton, N., Nolan, D., Baumer, B., Hall-Holt, O., Murrell, P., Peng, R., Roback, P., Temple Lang, D., and Ward, M. D. (2015). Data Science in Statistics Curricula: Preparing Students to 'Think with Data,' *The American Statistician,* 69, 343-353.

Smucker, B. J., and Bailer, J. (2015). Beyond Normal: Preparing Undergraduates for the Work Force in a Statistical Consulting Capstone, *The American Statistician,* 69, 300-306.

Valaitis, E., and Gray, M.W. (2006). Using business-style cases in an introductory statistics course. *Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS7, July, 2006), Salvador, Brazil.* Voorburg, The Netherlands: International Statistical Institute. http://iase-web.org/documents/papers/icots7/C131.pdf.

The author would like to thank the students of American University who participated in these projects and Meagan Snow, who provided GIS assistance.