

## DIVERSITY-RELATED PROJECTS IN AN INTRODUCTORY STATISTICS COURSE

Kirsten Doehler and Laura Taylor  
Department of Mathematics and Statistics  
Elon University, Elon, NC 27244, USA  
[kdoehler@elon.edu](mailto:kdoehler@elon.edu)

*This paper discusses a diversity-related introductory statistics project where students are provided data on individuals who receive Supplemental Nutrition Assistance Program (SNAP) benefits, also known as food stamps. The multi-part project included two written reports and a presentation. Students analyzed the SNAP data to explore multiple demographic and inferential questions. After analyzing the data and investigating articles about food insecurity, each group developed a meal plan based on the budget of SNAP recipients to help foster empathy. Students presented their analyses and meal planning portions of their project. The focus of this paper is to provide details about the project. Additionally, the authors discuss various goals of the project, which include helping some students understand challenges faced by individuals from a low socioeconomic status.*

### INTRODUCTION

Projects can be a beneficial experience for statistics students, as they provide opportunities for students to be actively engaged in learning. Evidences for using projects are widespread in statistics education literature (Carnell, 2008; Spence, Bailey, & Sharp, 2017), though some advocate for having students collect data (Hogg, 1991). Collaborative assignments and projects are endorsed as a high-impact practice as well (Association of American Colleges & Universities, 2017). Since statisticians naturally work in collaborative environments, experience working on group projects builds teamwork skills and mimics real-world experiences. Numerous recommendations from the Guidelines for Assessment and Instruction in Statistics Education (GAISE) are supported through this project, including integrating real data with a context and purpose, fostering active learning, and using technology to analyze data (GAISE, 2016). The authors have previous experience using projects in different levels of statistics courses. However, the focus of the multi-part project discussed in this paper is to help students gain both statistical skills while simultaneously thinking critically about issues of diversity and social justice.

### PROJECT HISTORY

Initially we integrated diversity themes by obtaining diversity-related data to be used in group projects. Numerous data sets for this earlier version of the project assignment were obtained from the [diversitydata.org](http://diversitydata.org) website (The Heller School for Social Policy and Management, 2017). Most of the data for the first version of the diversity-themed projects included a variable for race/ethnicity and another variable such as the percent of children living in low-income neighborhoods. Data were available for many metro areas throughout the United States. However, this data frequently contained only one numeric variable and one or more categorical variables. It was challenging to find data that could be analyzed using a wide variety of statistical methods that allowed students to develop their own research questions.

Delaney and Scheller's 2015 article "Who gets food stamps? White people, mostly" and the "food stamp" social media challenge, famously attempted by Gwyneth Paltrow (Bever, 2015), prompted our interest in having students explore characteristics of individuals who receive Supplemental Nutrition Assistance Program (SNAP) benefits. The data referenced in the article is found by clicking on "SNAP Quality Control Data" from the United States Department of Agriculture Food and Nutrition Service website (2017, "Research"). Unlike earlier diversity-related data that we had considered, the SNAP Quality Control data bases lend themselves to a variety of analysis methods covered in the course since they contain a variety of quantitative and categorical variables. They also incorporate race and socioeconomic status variables. Currently, separate databases are available for 1996 through 2016.

**PROJECT DATA**

The project utilizes SNAP data from 2013, which contains data on 792 variables from 49,569 households (108,142 individuals). The data are raw in the sense that categorical variables are coded and require referencing the technical documentation (also available on the “SNAP Quality Control Data” link) to determine the values (Filion, Eslami, Bencio, & Schechter, 2014). The authors modified the raw data by reorganizing the structure so that individuals are represented by a row in the data (as opposed to all members of a household being on the same row of the data). From this data set, two subsetted data sets were created, “SNAP\_sample” and “HOH\_sample,” as discussed below for use in the two written parts of the project. As is typical in real-world data, there are missing values.

*The “SNAP\_sample” Data*

For the first project report, students are provided with information on the race, age, and sex of a random sample of 500 individuals from the Quality Control Database of SNAP beneficiaries. The data contains individuals who were SNAP beneficiaries in 2013 that ranged from head of household, dependents, and other adults. The first several rows of the data are shown in Figure 1; this image also appears on the project instructions. There are 22 race categories, and students use the technical documentation to collapse these 22 categories into six broader categories (Filion, Eslami, Bencio, & Schechter, 2014).

|   | A    | B   | C   |
|---|------|-----|-----|
| 1 | race | age | sex |
| 2 | 2    | 32  | 2   |
| 3 | 7    | 23  | 2   |
| 4 | 7    | 41  | 2   |
| 5 | ?    | 52  | 1   |

Figure 1: The first three observations in the “SNAP\_sample” data set

*The “HOH\_sample” Data*

The data for the second project report contains information on a random sample of 500 head of households that received SNAP benefits as opposed to all recipients, as seen in Figure 2. This data includes the same variables as those present in the “SNAP\_sample” data along with variables for years of education, employment status, and a status variable related to whether the amount of SNAP benefits received was correct, too high, or too low. The choice to have a second data set that only includes head of households was made since the original data set contained all beneficiaries including children. Variables such as years of education and employment status are also reported for children in the original data, and without some additional expertise in data manipulation, analysis of these variables could underestimate education and employment status.

|   | A      | B    | C     | D      | E   | F   |
|---|--------|------|-------|--------|-----|-----|
| 1 | STATUS | race | YRSED | EMPSTB | age | sex |
| 2 | 2      | 2    | 12    | 4      | 55  | 2   |
| 3 | 1      | 7    | 12    | 1      | 51  | 2   |
| 4 | 1      | 7    | 14    | 1      | 39  | 2   |
| 5 | ?      | ?    | 12    | 1      | 22  | ?   |

Figure 2. The first three observations in the “HOH\_sample” data set

**PROJECT OUTLINE**

The diversity-themed project has three components: two written reports and a presentation. Students work in the same group for all parts of the project. The project discussed here involves these components:

- Analysis and written report on the race of SNAP recipients using the “SNAP\_sample” data;
- Analysis and written report of data on a student-selected question related to the head of households for SNAP recipient families based on the “HOH\_sample” data; and
- Presentation on the analysis of “HOH\_sample” data and meal plan for the student teams developed with the budget of SNAP recipients.

For the written portions of the project, students are expected to communicate details of their analysis and results in a manner that can be understood by a non-statistical audience. Students are instructed to write a report to executives at a SNAP benefits office so that these employees can reference the report in a press release about the characteristics of SNAP beneficiaries. Goals for the project are listed in Table 1.

Table 1: Project Learning Goals on Part 1 (P1), Part 2 (P2), and Presentation (Pres)

| Students will...   | P1 | P2 | Pres |
|--|----|----|------|
| ... recognize and evaluate any preconceived notions on the race of SNAP recipients and connect the trends in the data to larger population trends. | ✓  |    |      |
| ... formulate their own question about real data that can be answered using methods learned in the course.   |    | ✓  |      |
| ... gain experience working with real, large, raw data sets.   | ✓  | ✓  |      |
| ... select an appropriate analysis method for quantitative or categorical data.  | ✓  | ✓  |      |
| ... perform statistical analysis using statistical software.   | ✓  | ✓  |      |
| ... communicate statistical results in writing to a general audience.  | ✓  | ✓  |      |
| ... familiarize themselves with the basic provisions of SNAP benefits and develop a meal plan that meets budgetary restrictions.                   |    |    | ✓    |
| ... communicate statistical results orally to a general audience.  |    |    | ✓    |

### Part 1

In Part 1, students use statistical software to condense the 22 race categories into the following six categories (Black or African American, Hispanic or Latino, Multi-racial, Other races, Unknown, White). Students report a confidence interval for the proportion of SNAP recipients who identify as each of these six races. Students do not use the variables age and sex in this part of the project, but those variables could be used in Part 2. Students do not present this portion of the project. However, instructors facilitate class discussion to highlight the similarities between the proportions in the SNAP population (based on their analysis) and the proportions of each race in the general population.

### Part 2

In Part 2, students develop a statistical question that can be answered with either of the provided data sets to investigate further demographic characteristics of the SNAP beneficiaries. Students must complete an appropriate statistical analysis based on the question they formulate. Possible analysis methods based on the provided data include: t-tests, ANOVA, and Chi-Square tests.

### Presentation

Students present their work from Part 2 of the project. The presentation also has two additional requirements. First, students research a diversity issue and discuss relevant information from the literature. Second, students determine the amount of SNAP benefits their project group (called a statistics family) would receive and develop a one-week meal plan. The amount of SNAP benefits is based on both the number of people and net monthly income of their “statistics family.” This amount is found on the United States Department of Agriculture Food and Nutrition Service website (2016, “Supplemental Nutrition Assistant Program (SNAP)”). The meal plan portion of the presentation also consists of a grocery list with prices. This helps to verify that the meal plan developed meets financial limits and raises awareness about the costs of basic necessities such as groceries.

## CONCLUSION

There were several pitfalls that students fell into during this project. For example, some students struggled to interpret the confidence intervals for Part 1 correctly. Students would interpret their intervals as the proportion of each race that received SNAP benefits as opposed to the proportion of SNAP beneficiaries who identified as each race. Additionally, some students had issues identifying the types of analyses that should be performed to answer the question that they formulated. Other students developed a question after they had performed a successful or interesting analysis. For the presentation, a few groups missed the intent of developing the meal plan, which was to understand the challenges faced by SNAP recipients to feed their family balanced and nutritious meals on a limited budget. For those students, they took pride in coming in below budget, which trivialized the challenges of families facing food insecurity. To help minimize this effect, instructors had students respond to a survey about how nutritious students felt their meal plan was and whether they would be happy eating those meals for one week, two weeks, etc. Finally, students and instructors sometimes struggled with understanding what could be purchased using SNAP benefits and who was eligible to receive those benefits.

During project presentations, it is recommended that instructors facilitate class discussion around issues such as access to education and impediments to equality. It may also be pertinent to address the comparison between the amount of SNAP benefits some families are receiving each week and the current average cost of college tuition per week. Overall, this project has been a gateway for opening the door to discussions about inequality using empirical evidence and fostering empathy in a statistics classroom.

## REFERENCES

- Association of American Colleges & Universities. (2017). *High-Impact Practices*. Retrieved from <https://www.aacu.org/resources/high-impact-practices>
- Bever, L. (2015, April 17). A hungry Gwyneth Paltrow fails the food-stamp challenge four days in. *The Washington Post*. Retrieved from [https://www.washingtonpost.com/news/morning-mix/wp/2015/04/17/a-hungry-gwyneth-paltrow-fails-the-food-stamp-challenge-four-days-in/?utm\\_term=.b7c1f597d26f](https://www.washingtonpost.com/news/morning-mix/wp/2015/04/17/a-hungry-gwyneth-paltrow-fails-the-food-stamp-challenge-four-days-in/?utm_term=.b7c1f597d26f)
- Carnell, L.J. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education*, 16(1).
- Delaney, A. & Scheller, A. (2015, February 28). Who gets food stamps? White people, mostly. *Huffington Post*. Retrieved from [https://www.huffingtonpost.com/entry/food-stamp-demographics\\_n\\_6771938.html](https://www.huffingtonpost.com/entry/food-stamp-demographics_n_6771938.html)
- Filion, K., Eslami, E., Bencio, K., & Schechter, B. (2014). *Technical Documentation for the Fiscal Year 2013 Supplemental Nutrition Assistance Program Quality Control Database and the QC Minimodel*. Retrieved from <https://host76.mathematica-mpr.com/fns/Download.aspx?>
- GAISE College Report ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report*. <http://www.amstat.org/education/gaise>
- Hogg, R.V. (1991). Statistical education: Improvements are badly needed. *The American Statistician*, 45, 342-343.
- Spence, D. J., Bailey, B., & Sharp, J. L. (2017). The impact of student-directed projects in introductory statistics. *Statistics Education Research Journal*, 17(1).
- The Heller School for Social Policy and Management. (2017). *Data for diverse and equitable metropolitan areas*. Retrieved from <http://diversitydata.org/>
- United States Department of Agriculture Food and Nutrition Service. (2017). *Research*. Retrieved from <https://www.fns.usda.gov/ops/research-and-analysis>
- United States Department of Agriculture – Food and Nutrition Service. (2016). *Supplemental Nutrition Assistant Program (SNAP)*. Retrieved from <https://www.fns.usda.gov/snap/how-much-could-i-receive>