

A Service Learning Project in an Introductory Course: visualizing relationships in 2-1-1 data

Carolyn Cuff and Sam Hockenberry '19

Westminster College, New Wilmington, Pennsylvania 16172

Description of Course

Audience: Liberal arts majors with a steady increase in journalism and broadcast communication majors each year. Meets quantitative reasoning requirement. Separate statistics courses exist for science, math, social science, and business majors. Typically taken by first and second year students

Content: Descriptive statistics including two and more variables
Normal probability models, introduction to binomial models
Collecting data with emphasis on surveys and sampling
Confidence intervals and hypothesis testing for proportions

Content Emphasis: Statistics used in mass media, observational studies and categorical variables

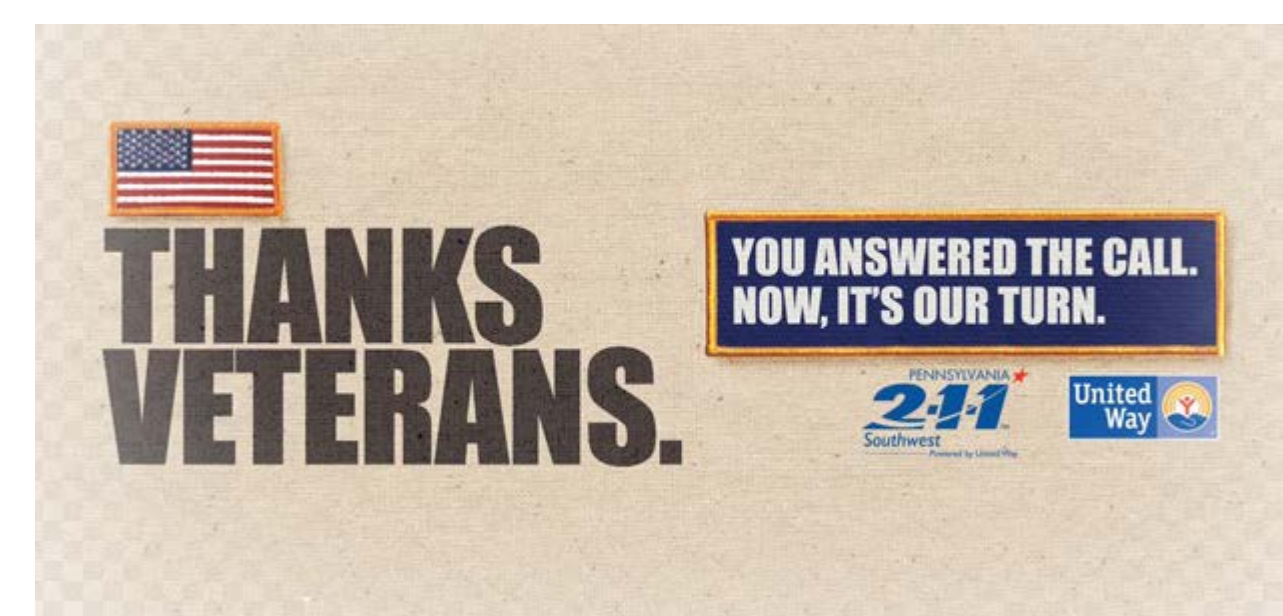
Text: Gould and Ryan, *Essential Statistics: Exploring the world through data*. Use MyStatLab for homework, assigned daily

Software: StatCrunch with class group, D2L

Class format: 3 times a week, 1 hour each meeting, students bring laptops when told, interactive lecture, group problem solving

Evaluation: MyStatLab (unlimited attempts) 15%
Project interim activities 20%
Research Day poster presentation 5%
Final presentation 10%
Exams (equally weighted no final) 50%

Successful Project Data Spring, 2017



PA 2-1-1 Southwest is part of the national 2-1-1 call centers initiative that seeks to provide an easy-to-remember telephone number and web resource for finding health and human services both for everyday needs and in crisis situations. In partnership with United Way, this agency maintains a large database of clients and resource providers. Understanding the needs of their clients will aid in both ascertaining that resources providers are sufficient and obtaining funding for the services. One group of particular interest at this point is called "Veterans" but also includes active service individuals or their immediate families. Local Veteran organizations, PA 2-1-1, and the United Way of Allegheny County (Pittsburgh) are interested in basic demographics of the callers and potential relationships between the variables.

Below is a cut of the cleaned data. The data set consists of over 11,000 cases and 30 variables. Note, most of the variables are categorical; very appropriate for this level course.

Actively Serving	Veteran	Month	ContactDateTime	Primary Reason For Calling	Referral Name	Type Of Contact	Single Parent	CallType Detail	County	Gender	How Heard	Intake Line	Family Member Working	First Call	Marital Status	Age	Number Of Adults In Household	Number Of Children	Taxonomy Category	Source Of Referral	zip	city
No	Yes	6	06/17/2015 13:38:24.44324 EDT	Utilities	Duquesne Light Utility	Phone	NA	Referral	ALLEGHENY	Female	Agency - Duquesne Light	After Hours	No	Yes	Single	54	2	NA	Electric Service Payment Assistance	Agency/Organization	15221	Pittsburgh
No	Yes	1	01/18/2015 02:48:13.25686 EST	Consumer Services	Supportive Services for Veteran Families	Phone	NA	Information (Phone # Requests)	BEAVER	Male	NA	After Hours	NA	Yes	Unknown	NA	NA	NA	Property Tax Exemption Information	Media/TV/Radio	15001	Aliquippa
No	Yes	2	02/24/2015 13:56:47.40186 EST	Taxes	IRS VITA	Phone	NA	Tax	MERCER	Male	Library	Public Form - Taxes	NA	Yes	Unknown	NA	NA	NA	Tax Preparation Agency/Organization	Agency/Organization	16154	Transfer
No	Yes	2	02/09/2015 18:44:16.46037 EST	NA	NA	NA	NA	Tax	ALLEGHENY	Male	NA	After Hours	NA	No	Unknown	NA	NA	NA	NA	NA	15215	Pittsburgh

GAISE Recommendations in Service Learning Project

Many **GAISE 2016** recommendations are directly addressed in the successful project. The numbers refer to the numbered recommendation in that document.

2. Students should be able to recognize questions for which the investigative process in statistics would be useful and should be able to answer questions using the investigative process.

Analysis: Approximately 50% of students had substantial difficulty. Questions were posed along the lines of **why** do women call for help with mental health issues more often than men, **how** can we address needs of clients, **what** is the connection between the county and the reason for calling.

Remediation: Substantial time, both in class and with individual students, was spent discussing the role good analysis of data could play helping 2-1-1 directors answer their own questions. Students were grouped according to interests shown my type of questions asked (e. g. mental health, family services, demographics). See also GAISE 3 analysis.

3. Students should be able to produce graphical displays and numerical summaries and interpret what graphs do and do not reveal.

Analysis: Basic graphical analysis of demographics helped students formulate better questions. For example, comparing the proportion of callers who are women with the proportion who are men is valuable information but does not answer why do women call more often. The process of "write a question, use data to answer that question, what questions are raised" became natural because students **owned** the data.

GAISE #2 and #3 tie together better when students examine and reexamine questions and graphs.

4. Students should recognize and be able to explain the central role of variability in the field of statistics.

Analysis: Unexpected positive results. Very rich in class discussions of why data vary. For example, the number of calls vary by month for utility help, by age and month for tax help. Very rich discussions of importance in reporting variation and not just a measure of central tendency. Very rich discussions about considering other related variables to minimize overall variance. Students began to investigate potential variables that might explain the variation seen. Analysis of outliers became important. For example, the spike in calls related to food in month 8 was not "just" ignored.

6. Students should gain experience with how statistical models, including multivariable models, are used.

Analysis: A large dataset allows partitioning the data in meaningful ways. Partitioning leads to easy discussions of "another" variable. Since data was presented as "we are interested in how the counties might differ," students naturally considered data by county.

What I learned: consider how you present the data, consider augmenting textbook data set with another variable if only one or two variables, don't pare down the textbook dataset to make it easier for students.

Problem encountered: StatCrunch is limited in data visualizations or difficult to use compared to R. Three variables easy - adding another layer difficult.

9. Students should demonstrate an awareness of ethical issues associated with sound statistical practice.

Analysis: Presentation to directors forced students to consider the conclusions that are justifiable. Real people, real problems met real considerations of ethical issues with consequences.

Large Data Sets in Intro Course

Advantages:

Understand value of data summary – MUST summarize data
Create many different models with same data – understand modeling **process** rather than understand **some** model
Working with more than one or two variables is natural
Partition dataset to give parts to students (early in the course or if weak students)
Real data ☺

Disadvantages:

Messy – will need cleaning which takes a lot of time and these students cannot do
Overwhelming to students – need to find way to introduce pieces

Conclude: Advantages can outweigh Disadvantages

Lessons from Failures

Don't try to collect data. Time required >>> Value to students;
Previous project: campus wide batteries recycling

Get data PRIOR to semester – Format may not be usable.
Previous semester same data set but not usable, Data Science students cleaned for about 100 person hours, still not usable.

Every project will take 2 trials. First trial is learning experience for instructor and students. Second trial is learning experience for students

Project Interim Activity Timing

Week	Points	Goal	Assignment
1	25 individual	Take project seriously	Summarize PA 2-1-1 after in-class hour long presentation by Director of agency
2	9 partner of choice	Identify categorical variable, become familiar with data set, use StatCrunch	Determine the number of levels and give examples of the levels for three categorical variables
6	25 partner of choice	Pose questions that can be answered with statistical analysis	Write three questions that can be answered using the data set
7	25 individual	Present statistical findings	Write a brief abstract for the poster. Include information on 211, information on statistical questions
8	29 group assigned by similar questions	Create graphs illustrating relationships between 2 or more variables	First draft graphs. Formulate 6 questions that consider relationship between two or more variables. Extensive feedback was given.
9	12 same group	Analyze graphs	First draft analysis. Having created better graphs, analyze the graphs noting and commenting on relationships
10	10 same group	Present findings in coherent way in a poster	First draft poster. Given a list of elements – checked to see if there. First time many students had created a poster
12	Group	Discuss at individual level	Poster presentation to Peers during Research Day at College Graded by Business Statistics Students
14	Class		Oral presentation to Agency

Finding projects

Reach out to a program or agency for a project:

Build a relationship with some non-profit agency. Connect with key partners or stakeholders through your own civic engagement. Use those relationships to find others. Think - friend of a friend

Attend presentations offered by the agency to get an overview of the work then you can quickly identify ways in which the class could assist. Watch your local news agency for announcements of such or call the local newspaper staff

Within your (smaller) college some typical needs

Registrar – rich source of data typically not analyzed. Some examples
Change of Majors - from what to what to, when, why (trends)
Internships (especially not for credit) - where, what major/semester

Use of Library, use of public computer labs – when, where, why
Use of other facilities such as recreation areas, parking areas

Additional information

Available at
<http://www.westminster.edu/staff/ccuff/uscots2017>

- This poster
- Brief annotated bibliography of service learning
- Syllabus of the course; csv file of data; original data before cleaning
- Student posters of 2-1-1 project
- Not as successful project work