Using “Big Data” in a Political Science Research Methods Course: A Description and Initial Assessment of a Social Media Analysis Assignment

William R. Wilkerson

SUNY Oneonta

Funding for this project came from SUNY Innovative Information Technology Grant (IITG) Program jointly granted to SUNY Oneonta and the Center for Computational Research at the University of Buffalo.

Note: This is a preliminary draft. Any and all comments and suggestions are appreciated.

Introduction

Teaching research methods to political science undergraduates can be a genuine challenge. Political science undergraduates do not coming enter our classrooms on the first day being particularly interested in the “science” part of political science. Making matters worse, research methods is often a required course. Many among our science majors are stressed about the idea of talking, thinking, and writing about numbers of any kind, let alone using mathematics and statistics to learn about the world around them (Bos and Schneider 2009; Buchler 2009; Bernstein and Allen 2013). As an instructor of research methods, I regularly observe in my class that students are interested in politics, but are not inherently interested in how political science research is conducted. To be successful as instructors we must cultivate their interest. We must help them become motivated to study what we must acknowledge can be a daunting subject. While students may come to our courses on civil rights and liberties, international human rights, terrorism, or fill in almost any subject oriented political science course here, with motivations to learn about a subject they are passionate about or eager to learn about, we must do more in research methods to develop and sustain this motivation.

This paper explores an assignment developed to help political science students learn new skills and expose them to new areas of research and data, in a way that will hopefully motivate them to learn more about political science research methods. This assignment has students gathering and analyzing social media data—big data—using a mix of open source and commercial tools on a topic of their choosing.

Student Motivation

In their excellent summary on the subject of teaching and learning, *How Learning Works*, Ambrose et al. argue that motivation is a key to learning: "[s]tudent's motivation determines, directs, and sustains what they do to learn (2010, 5)." This motivation is critically important as Ambrose et al. note, "[l]earning is not something done *to* students, but rather something students themselves do (3)." One of our central tasks as instructors then is to do what we can to increase motivation.

It turns out political science students are not alone in having low motivation and negative attitudes about courses that use statistics.[1] These attitudes are typical of attitudes of college students toward statistics courses overall (Ball & Pelco, 2006; Mallow et al., 2012; Tremblay et al., 2000), with some college students viewing statistics as the worst course taken in college (Hogg, 1991). Many of the challenges that statistics instructors face are related to a lack of student motivation and preparation for these courses (for a review, see Tiskhovskaya & Lancaster, 2012). Some of the causes of students' lack of motivation and engagement in statistics courses include math and statistics anxiety (Gal & Ginsburg, 1994), a lack of sufficient background knowledge (Batanero et al., 1994; Garfield & Ben-Zvi, 2008), and a failure to see the relevance or value of statistics for their degree or future career (Hood, Creed, & Newmann, 2012). Students' low motivation is likely to hinder their development of statistical skills and knowledge, achievement in these courses (Hood,

[1] Thanks to Dr. Sara Fulmer, a faculty member in Oneonta's Educational Psychology and Reading Department for insight into this literature.

Creed, & Newmann, 2012; Schutz, Drogsz, White, & Distefano, 1998), and pursuit of statistical knowledge and skills after the course (Gal, Ginsburg, & Schau, 1997; Schau, 2003).

How can we increase motivation among our students? Ambrose and colleagues (83-85) have a number of suggestions:

- "Connect the material to students' interests,"
- "Provide authentic, real world tasks,"
- "Show relevant to student's current academic lives,"
- "Demonstrate the relevance of higher level skills to students' future professional lives,"
- "Identify and reward what you value," and
- "Show your passion and enthusiasm for the discipline."

As political scientists, like instructors in other disciplines where the goal is to apply methodological tools to a discipline where the students have shown pre-existing interest, we have a leg up on the traditional statistics class as we have a topic, politics, that students have expressed interest in. Promoting research in this way should help increase student motivation. Research has found that statistics students learn more when they view the content as relevant (Marsick & Watkins, 2001; Marek, Christopher, & Walker, 2004; Navarro, 2013). Furthermore, our courses can help develop skills that can be useful in future political science courses, in graduate schools, and in their careers. These can be keys to student motivation (Brophy, 2008; Hulleman & Harackiewicz, 2009; Newmann et al., 1996). Data gathering, data analysis, critical thinking and report writing are all examples. When students are aware that these skills are important, it can help to motivate them.

Motivation can also potentially come from the data that our students analyze and the tools they use both interms of content and type. Analyzing "big data" with open-sources tools is one potential example of how students can develop skills and learn new tools, thus motivating them to learn more in political science research methods courses.

Big Data: What is it and why use it in class?

Big data is an umbrella term (Press 2014). Big data refers to datasets that are so big that it is impractical to analyze using traditional social science techniques and tools. This data can take many forms that are applicable in a wide-variety of fields. Everything from medical records to phone logs, meteorological data to social media posts can be a source of big data. While controversial, big data analysis has been growing in the social sciences, the natural sciences, and even the humanities. Mayer-Schoenberger and Cukier argue that big data is "A revolution that will transform how we live, work, and think (2013)." While the early euphoria has been tempered by critics, most still acknowledge that analyzing very large datasets that require machine learning are likely to provide insights into our lives in various ways (Ruths and Pfeffer 2014).

Educators in a variety of applied fields including business (Cannon, Godwin & Goldberg 2013), journalism (Bryant & Uzma 2014), and engineering (Xian & Madhaven 2014) recognize that some familiarity with big data will be essential for the workplace in the foreseeable future. Information/library sciences scholars (Eisenberg, Johnson & Berkowitz, 2010) have also advocated engagement with big data writ large in order to attempt to enhance future professional needs, as well as the potential for these technologies to make undergraduates into active learners (Dominic & Pilomenraj 2014). Social science students will also benefit from being conversant in this area.

Political scientists have begun to mine social media to understand mass and elite political behavior (Barbera 2014; Bekafigo & McBride 2013; Clark 2014). The Power Research Internet Project (2014) has found that 74% of adult Internet users, about 64% of adults, use social media. Teen use is even higher at 81%. Social media then—Facebook, Twitter, Reddit, blogs, etc.—is a potential form of big data that is being used by political science students. Many of our students are active on at least one social media platform. With this in mind a group of faculty at SUNY Oneonta set out to bring big data to undergraduate classrooms in the social sciences and philosophy. We hope that exposing students to big data will increase motivation to learn in various ways, while at the same help our students develop new skills that will be of use in their future endeavors.

Course Background

POLS 200, Approaches to Political Science, is a required course for political science majors at SUNY Oneonta, a residential, primarily residential institution of about 5,800 students. Students completing the political science minor and various education content concentrations also take the course, but the vast majority of students are political science majors. Class size ranges from 15-24 and the course is offered three out of every four semesters.

The course is taught in a flexible computer lab—computers can be recessed into the furniture when not in use creating tables. Over time the course has moved from a heavy lecture format with exams and two research assignments, to few lectures, much in-class work with frequent small assignments, including a significant amount of small group work, a few online quizzes, and a series of projects on a topic that each student selects at the beginning of the semester. POLS 200 has the following student learning outcomes:

- Understand key issues and concepts in studying research methods in political science;
- collect and assess data from a variety of sources; find and assess political science literature;
- describe data and explain relationships between variables, present data in charts and tables, and use spreadsheet and statistical software to do this;
- write a research paper based on statistical data;
- explore social media as a source of social science data (added spring 2014).

These learning outcomes, aside from the last, are fairly standard for undergraduate political science research methods courses. Some of the assignments are less typical (Turner and Theis 2009). Along with my departmental colleagues, we have decided to place some emphasis on practical numerical literacy skills—just plain data analysis as

Klass (2012) calls it—along with traditional goals of methods courses in understanding the research process and producing quantitative research. Table 1 shows the major topic assignments that each student must complete.

Table 1—POLS 200 Assignments, Fall 2014

Assignment	Description
Topic statement & research question	Students define a topic for the assignments to come and get practice in defining specific research questions.
Annotated bibliography	Students research the research question and provide a 10 citation bibliography. Source must come from books, articles, the Internet, and include data.
Creating a table & chart	Students find data and present it in a table and chart emphasizing readability and ease of interpretation. A brief write up is required.
Quantitative analysis	A traditional 6–8 pages quantitative analysis using techniques, either contingency tables or regression, appropriate to their data.
Big data/social media	The topic of this paper. Students gather, analyze and present social media data.
Presentation	A 6:40 presentation on their topic explaining what students have learned during the semester.
Topic reflection	A 500 word, end of semester examination of their successes and failures during the semester. What were they proud of? What did they learn? What could they do better?

Topic self selection is intended to increase student motivation (Ambrose 2010). While it does occasionally create difficulties for some assignments—for example finding appropriate datasets for quantitative analysis—students and I work together to do the best we can in exploring their chosen topics. Students use a variety of tools during the semester, Powerpoint, Excel and SPSS for presenting and analyzing data.

No single assignment counts for more than 100 of the 525 point semester total and the topic assignments in sum account for 400 points.

The Big Data/Social Media Assignment

The assignment was used for the first time in the spring 2014 semester and was repeated fall 2014. It is fairly simple in design and the report is modest. Execution, discussed in more detail in the next section, is more complex and requires some in class time as well as outside of class support.[2] The assignment comes late in the semester, during fall 2014 we didn't start it until late November, and has two weeks or less of class time dedicated to it each semester. In fall 2014 a week and a half of class time was used on the assignment. Two full weeks were used spring 2014. Students are allowed to use finals week to complete the assignment, giving students over three weeks to finish the work. Specifically in this assignment students will:

- Learn about big data and machine learning at an introductory level;
- Show their understanding of—introduced in earlier assignments—to gather data, understand research design and present data;
- Be exposed to the idea of data privacy in social media and develop an understanding of their social media presence.

The assignment is described to students as follows:

You will be looking at patterns in social media, specifically Twitter, on your topic. You will write a 750-900 word paper reporting on your findings. For this assignment you will:

1. Gather social media, most likely tweets, using the tool Trackur and analyze them using RapidMiner.
2. Center your paper around a chart or charts that summarizes your results. You may also want to include a table as well, but a graphic is required for each set of data that you discuss in your paper.
3. Discuss Your Process. You need to explain your Trackur search(es). What terms did you search? When were the searches run? How big was the dataset you produced? You also must explain the processing you did in RapidMiner. (Most of you will use a word count process. Note that and any other processes you used.)

[2] A few students did similar assignments in other courses in political science or sociology concurrently or in a previous course.

Beyond that, if you filtered your data to limit terms note the parameters you used. For example: "I limited the analysis to those words that appeared at least 100 times in my dataset and also eliminated #SCOTUS. I did this because . . ."

4. Present Your Analysis. Ask yourself at least two questions: Did the time period matter in my search? What broader themes appear in the word counts I produced?

In Class Work

We spend time in class signing up for both accounts in Trackur—a web-based commercial service for gathering social media data—and for VIDIA—a web-based high performance computing environment SUNY Oneonta developed with SUNY Buffalo's Center for Computational Research (CCR). This was done in a class period where the idea of "big data" was introduced and we examined the tools in a "hands on" way. We also explored social media and privacy as we searched in various social media platforms. At the end of this class we expect that students will develop enough proficiency in Trackur to gather the data they will need for the assignment. After a week or 10 days we come back to the assignment in class. In this second class session we collect the data and move it into the VIDEA platform for analysis. In a third session we continue to work on the assignment to allow stragglers to begin analysis and to help other students to refine their analysis. It is common for some students to need help outside of class to complete this assignment.

In between these class sessions we cover content analysis as a topic. This discussion is explicitly connected to social media data. Students complete a small group project where they review newspaper front pages, develop a coding scheme, code the data and present their results to the class. This is intended to prime students to think about coding data and looking for the broader meaning of text as data.

The Environment: Trackur, Vidia & RapidMiner

Gathering data, especially on a large scale, is difficult. Purchasing data can be expensive and is for the most part out of the reach of faculty for teaching purposes. Similarly, analyzing this data can be difficult as well. Tools for analyzing large-scale text datasets are less developed on standard software platforms commonly used by undergraduates on most college campuses such as SPSS. Even when available there are problems of getting students access outside of class. When students have access through emulation or in other settings, support because of location and different computer configurations can be difficult.

In the summer of 2012, a group of Oneonta faculty and information technology professionals saw the value in getting students access to big data and set out to find tools to make the data gathering and analysis doable for undergraduate social science majors.[3]

Gathering Data: Trackur

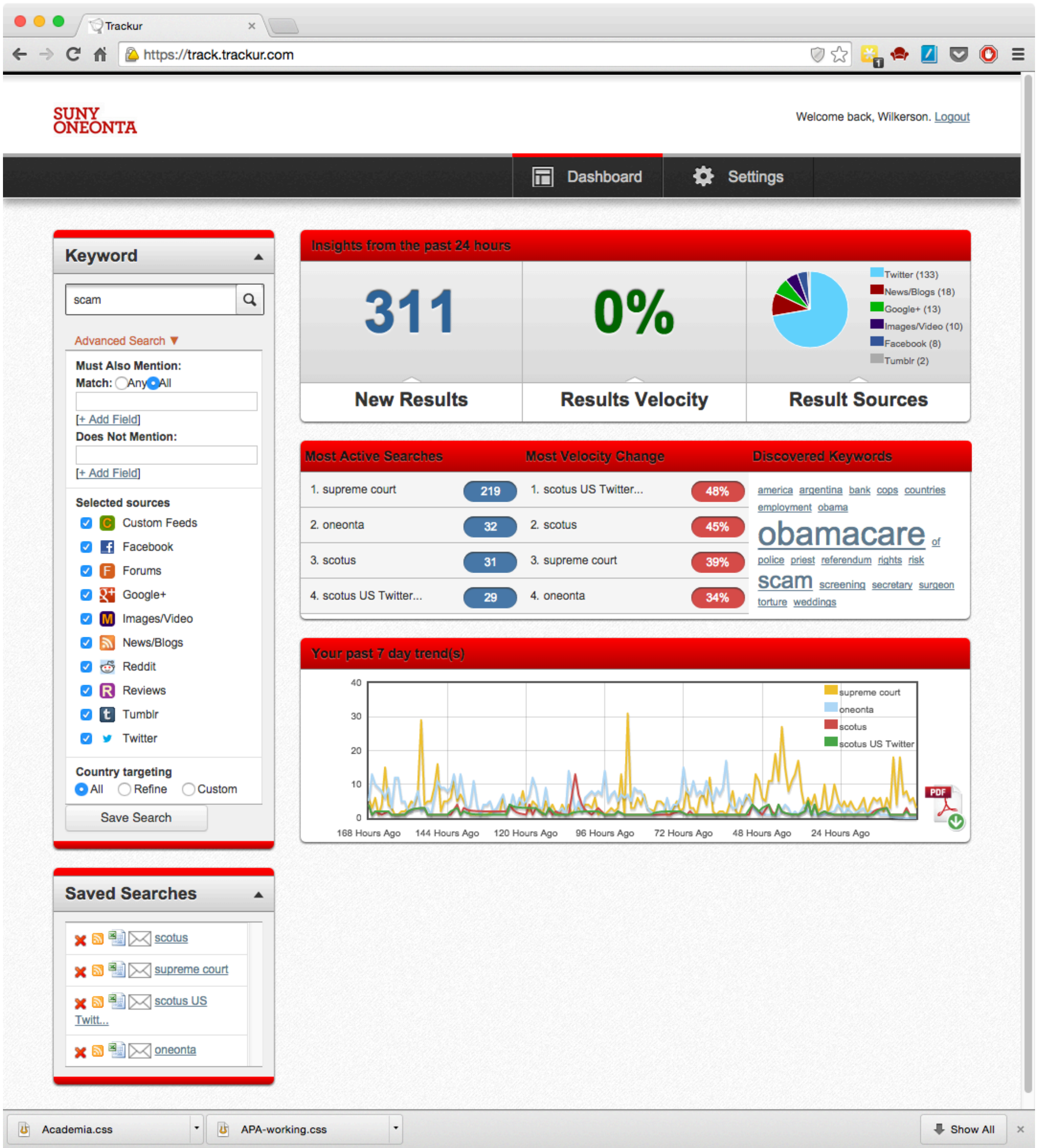
Having tried a number of social media capture tools beginning in 2012 including Discovertext, GNIP, Trackur, ContenSeer, X1 and a homegrown tool, we have settled on Trackur as a data gathering tool since the fall 2013 semester. Trackur is primarily focused on gathering social media data for businesses interested in what people are saying about their brands. Trackur connects to APIs for a variety of social media platforms including Facebook, Reddit, Tumbler, Google+, as well as a wide range of forums and news sites. For simplicity, both in analysis and supporting students, we have focused on Twitter data, tweets, for assignments.

Trackur is notably easy to use. After creating an account, something that Teaching Learning and Technology Center (TLTC) staff help with, students are provided with a clean, simple interface to begin searching in a box that is common on many web pages. (See Figure 1 along the upper left.) Trackur gives users the ability to save searches and control the sources of searches as well as the nationality of searches. Once saved, as shown on the bottom right, a user can broadcast results as an RSS feed or export results as a .csv file delivered by email that can be opened in a spreadsheet for review and editing or read directly by most data analysis packages. The standard Trackur dashboards provide information on the number of posts and the percentage changes in posts with the keyword(s) that give the user some insight as to the frequency of the search terms used. This can be helpful as studs explore social media and finalize their searches.

---

[3] I had earlier had students analyzing social media data in a capstone seminar, The Internet and Politics. I had used twitter data in POLS 200 prior to the current assignment as well.

Figure 1—The Trackur Search Window



Aside from minor headaches, such as emailed files ending up in email junk folders, Trackur has worked well for teaching purposes. Its simple interface and the ease of getting readable datasets have made Trackur popular with students and easy to support.

There are limits to the software, however. Trackur does not provide access to the full twitter "pipe," it's API provides a sample using an unknown methodology. Casual tests using searching in the Twitter app show this to be a minimal issue. Another potential issue is the fact that Trackur only sends out the 3,000 most recent results, although many more may be available in the search window. Depending on the popularity of the topic, gathering data over even a short period requires consistent trips to the site to send another search export. Multiple files must be combined and duplicates must be culled. None of this work is particularly difficult, but it requires a diligence that can be difficult.

Overall, Trackur has worked well. The lone problem that limits use is the funding needed to subscribe to Trackur. At over $4,000 per year it is reasonably priced in comparison to competitors, but is still expensive for academic departments.[4] For this reason, we continue to look for open-source, or other alternatives, for data gathering.

Analyzing Data: Vidia and RapidMiner

Central to our SUNY IIT grant was the development of an environment for analyzing big data, something that is beyond the current capabilities of Oneonta's information technology capabilities. This is a common problem for primarily undergraduate institutions (PUIs).[5] Oneonta has developed a strong working relationship with the University of Buffalo's Center for Computational Research (CCR) to create a collaborative environment, VIDIA, where PUIs can conduct intensive data analysis not otherwise possible. The Center for Computational Research is a leading academic supercomputing facility, maintaining over 8,000 processing cores and 500 TB of storage. It has extensive experience both with the development of virtual organizations and analysis tools for high performance computing users.

VIDIA is based on HUBzero, an open-source, web-based, scientific collaboration platform

---

4 Funding has come from the SUNY IITG grant and more recently from academic departments, the Dean of Social Science, and Information Technology.

5 This paper focuses on teaching uses of VIDIA, but the environment has outstanding potential for supporting research as well. Several Oneonta faculty are using the environment for these purposes.

developed by Purdue University.[6] VIDIA provides several advantages. First, it provides a notably high performance environment for data analysis of all kinds. Our usage so far in classes has been limited to analyzing Twitter results for the big data assignment using RapidMiner—a leading open-source data analytics and data mining package—but the environment is robust. It currently hosts, RapidMiner, R Studio—an integrated development environment for R, the open-source data analysis and programming environment—and PSPP, an open-source alternative to SPSS. It has the potential to do much more. We are discussing adding PYTHON for a user in Physics, for example. Second, it is attached to massive storage managed as part of CCRs cluster. Third, and this is key for support, as a web-based client, VIDIA allows us to have a uniform environment for all users in terms of software and data. Finally, VIDIA also allows for collaboration in various ways and has much of the functionality of course management systems such as Blackboard including document sharing, a gradebook and communication tools.[7]

Unlike Trackur, VIDIA does present some challenges for end-users. The process of inputting data, analyzing it and making changes is foreign to most if not all students. The process of doing analysis and its challenges are as follows:

1. Data, a .csv file, must be exported from Excel and then must be imported into VIDIA using WebDAV. Mapping a drive is not difficult, but it is not something most students are familiar with. RapidMiner's interface is complicated. (See Figure 2.)

2. Once in the environment, it looks foreign. The modular, flow chart metaphor, with a "box" for each step in the process is confusing for many. Clicking on the boxes often presents subroutines or options which must be explained.

3. Students are mostly running word frequency analyses. Some students ran k-means clustering processes as well.

---

[6] CCR has previous experience with HUBzero for research-only environments, but VIDIA is their first experience with developing a teaching environment.

[7] I plan to further integrate VIDIA into POLS 2015 fall 2015.

4. Making changes to processes is not intuitive for many students. Working with the TLTC, we built processes for students to start with, and in demonstrations we went through the process for changing parameters in processes. For example, results can be limited to the most common results. They can “prune” data, exclude common words like “the” using built in dictionaries and identify specific words or phrases such as the keywords searched for, from analysis. Students can “stem” words, a process of merging the results of words with similar roots, spelling and capitalizations. For example, obama, Obama and OBAMA would be combined into a single term. Several stemming algorithms are available.

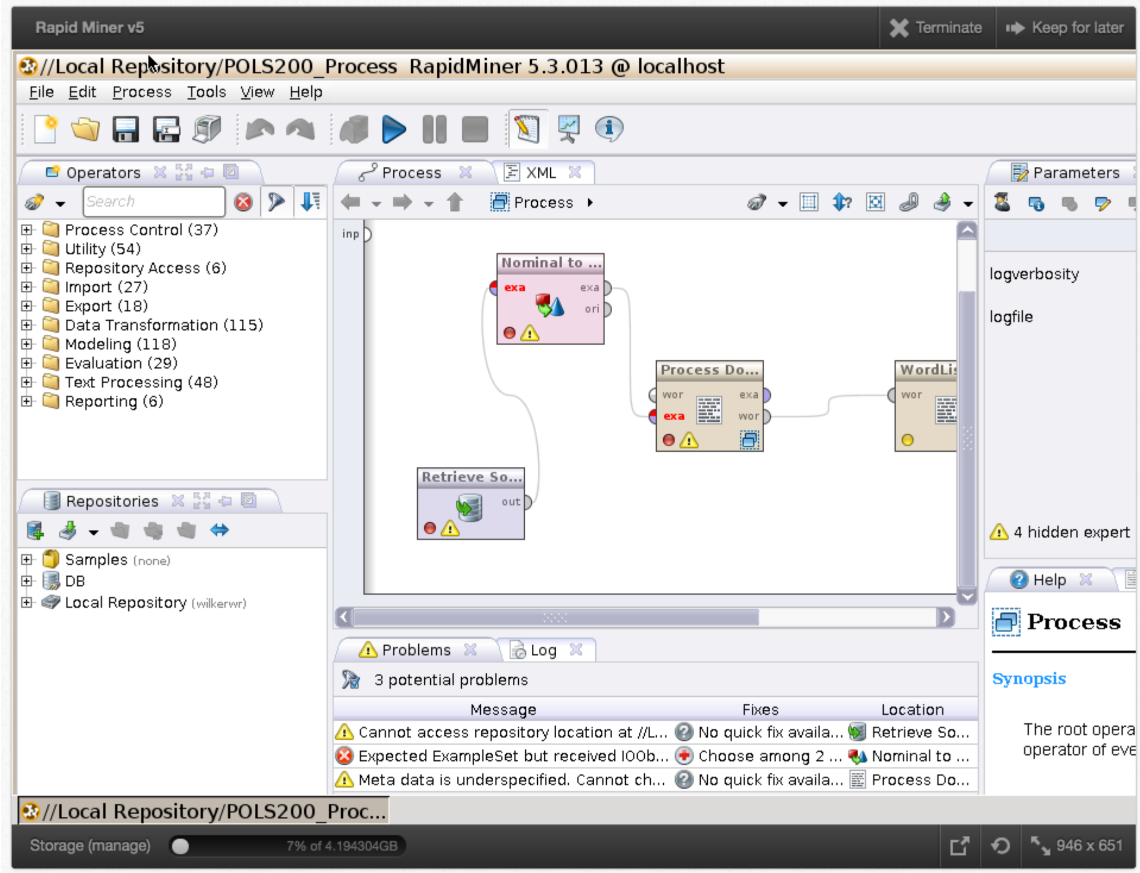
5. Results must be exported. The visualizations of the data in RapidMiner are not user-friendly or easily interpreted. Given my goal to develop data presentation skills, I had students export their data to Excel. Results were in the shared drive they created.

Students are well supported in their work in RapidMiner. The Director of the TLTC ran two work sessions in both classes and both the TLTC director and the instructor were available for extended office hours. In addition, a web site with written instructions and video tutorials was available to students.[8] Despite the complexity, we have been able to walk students successfully through this process. The Director of the TLTC and I have by trial and error developed a hands-on approach that has worked well. Students are not allowed to sit back and observe. We require students to work along with us. This fall, based on a question from a student about the modular processes, we walked student through the details. We both saw light bulbs going on as we did this. Students seemed more adept at making changes to their processes after this. This was used in other courses with similar results.

---

[8] http://www.oneonta.edu/bigdata/docs.asp

Figure 2—Rapid Miner Interface in VIDIA



Data and Methods

To examine the success and impact of the assignment during the spring 2014 and fall 2104 semesters a variety of measures were used. First, students were surveyed at the end of the semester about their knowledge of big data and their attitudes about the assignment as well as demographic information and their use and knowledge of social media. Questions centered on the assignment, the support they received to help complete the assignment and the environment used to gather information and analyze data. The survey was designed by a group of four faculty in political science and sociology who are active in the big data teaching project at Oneonta. Thirty six of 37 (97.3%) of students enrolled in POLS 200 during the two semesters completed this survey. This included each student who completed the assignment.[9] Students were asked to complete the survey before the final exam period. To encourage completion, this was a graded assignment worth three points toward their semester grade with non-completion earned a score of zero.

Second, I reviewed two end of semester assignments to see if students referred to the big data findings and if so, how they referred to them. Students completed an end of semester presentation of what they learned. Students were asked to write an approximately 500 word—self-evaluation of their semester's work. Was the big data assignment used in these end of semester assignments? If so, how was it used?

The students in the course, were traditional college age, the oldest was born in 1991, and almost exclusively sophomores and juniors. Most students were political science majors. The classes were largely male (72.35 of students) and white (88.5%).[10] Students were users of social media, especially Facebook and Twitter. Few were users of other platforms we asked about.[11] (See Table 2.) Two-thirds of students in each class, 24 of 36 total, had at least heard of the term "big data" before we undertook the project. Three students, one in the spring section and two in the fall section were concurrently or had in the previous semester completed a big data assignment.

Table 2—Student Use of Social Media Platforms

	At Least Daily	A Few Times Per Week	LT Once Per Week	Never
Fall 2014				
FacebookUse	12 (80.00 %)	2 (13.33 %)	1 (6.67 %)	0 (0.00 %)
TwitterUse	5 (33.33 %)	4 (26.67 %)	3 (20.00 %)	3 (20.00 %)
BlogUse	1	2	1	11

9 One student in the spring 2015 course stopped attending before the assignment was due.

10 The gender makeup was unusual as political science majors are about 55% men. The racial makeup of the class was close to college percentages.

11 A discussion with the class this fall found other social media, especially YikYak and also Instagram, were also commonly used by the students in the class.

	At Least Daily	A Few Times Per Week	LT Once Per Week	Never
	(6.67 %)	(13.33 %)	(6.67 %)	(73.33 %)
GooglePlusUse	1 (6.67 %)	2 (13.33 %)	1 (6.67 %)	11 (73.33 %)
WordPressUse	0 (0.00 %)	0 (0.00 %)	1 (6.67 %)	14 (93.33 %)
RedditUse	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)	15 (100.00 %)
Spring 2014				
FacebookUse	16 (76.19 %)	4 (19.05 %)	1 (4.76 %)	0 (0.00 %)
TwitterUse	8 (38.10 %)	3 (14.29 %)	0 (0.00 %)	10 (47.62 %)
BlogUse	1 (4.76 %)	1 (4.76 %)	3 (14.29 %)	16 (76.19 %)
WordPressUse	1 (4.76 %)	1 (4.76 %)	3 (14.29 %)	16 (76.19 %)
RedditUse	1 (4.76 %)	1 (4.76 %)	2 (9.52 %)	17 (80.95 %)
GooglePlusUse	1 (5.00 %)	2 (10.00 %)	3 (15.00 %)	14 (70.00 %)

Results: Student Attitudes Toward the Project

The students were generally positive about the assignment with Fall 2014 students being more enthusiastic than the Spring 2014 students. Table 3 shows responses to a series of questions about their satisfaction with the assignment from various perspectives. Two items were aimed at gauging whether students believed whether they had the time and support to complete the big data assignment. These items read "I had sufficient time to complete the Big Data assignment, and "I had sufficient support to complete the Big Data assignment." Students were generally satisfied with the amount of time to complete the assignment along with the support they had for completing the assignment. Two-thirds of spring 2014 students agreed or strongly agreed they had sufficient time to complete their assignment. Ten of 21 students agreed and four strongly agreed. Most of the remaining students—five of seven—were neutral. Students were more positive in fall 2014. Over half (53.3%) strongly agreed while 80% were positive overall. The remaining 20% were "neutral." Similarly, students felt well supported. Eighty-one percent of spring 2014 students at least agreed that they were well supported with one-third strongly agreeing. Only two students (9.5%) disagreed. Fall 2014 students were both more positive and more negative about the support they were given, with 40% strongly agreeing and 33.3% more agreeing. In contrast, 20% disagreed that they were well supported.

Table 3—Student Evaluations of the Assignment

	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
Fall 2014					
Support	6 (40.00 %)	5 (33.33 %)	1 (6.67 %)	2 (13.33 %)	1 (6.67 %)
TimeToComplete	8 (53.33 %)	4 (26.67 %)	3 (20.00 %)	0 (0.00 %)	0 (0.00 %)
LearnExper	8 (53.33 %)	5 (33.33 %)	1 (6.67 %)	0 (0.00 %)	1 (6.67 %)
Repeat	7 (46.67 %)	5 (33.33 %)	1 (6.67 %)	0 (0.00 %)	2 (13.33 %)
Spring 2014					
Support	7 (33.33 %)	10 (47.62 %)	2 (9.52 %)	2 (9.52 %)	0 (0.00 %)
TimeToComplete	4 (19.05 %)	10 (47.62 %)	5 (23.81 %)	2 (9.52 %)	0 (0.00 %)

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
LearnExper	3 (14.29 %)	11 (52.38 %)	4 (19.05 %)	2 (9.52 %)	1 (4.76 %)
Repeat	5 (23.81 %)	4 (19.05 %)	10 (47.62 %)	1 (4.76 %)	1 (4.76 %)

Interestingly, the spring 2014 students felt better supported despite experiencing many more technical problems. The VIDIA system was down more than once and students noted this. For example, "One of the only things I really feel that needs improving on this course, was how ineffective it was to use Rapid Miner. This was the only tool we used that seemed ineffective." And even more bluntly: "Make sure its working and you know how to use it before make you make us do it for a grade." And, "Before putting an assignment that requires a lot of moving parts on the syllabus, just make sure that all the possible drawbacks are dealt with and don't exist." Twelve of 21 students made comments critical of RapidMiner. The responses to the Likert scale may have been an "us against the world" phenomena, seeing us fighting against a frustrating tool. In contrast, students in the fall 2014 section had many fewer comments about RapidMiner. They did ask for more time to complete the assignment though. These comments were less widespread, but are typified by "Just a longer period of learning about Big Data," and, "[m]ore scheduled time to deal with one on one issues."

Each class was also asked two summative questions about the assignment, one direct—a response to the statement "The Big Data assignment was a worthwhile learning experience"—the other less direct—"I would encourage my instructor to use a Big Data assignment the next time he/she teaches this course." In evaluating their learning experience, two-thirds of spring 2014 students at least agreed, with over one-half (52.4%) agreeing and 14.2% strongly agreeing. Only 14% students disagreed. In fall 2014 the response was overwhelming positive with over one-half (53.3%) strongly agreeing and 86.7% of the class at least agreeing with the statement. Only one student in the fall semester disagreed that the assignment was a worthwhile learning experience. Should the assignment be repeated for future classes? The spring 2014 students were notably less positive. Only 9 of 21 students (42.9%) agreed or strongly agreed, and the modal category with 10 responses (47.6%) was "neutral". The fall 2014 group was again decidedly more positive. Twelve of 15 students at least agreed that the assignment should be repeated with the modal category being "strongly agree." Only two students disagreed, although both were in the "strongly disagree" category.

Table 4 shows the results of two additional questions asked only of fall 2014 students to assess their view of the assignment to assess their satisfaction with the assignment and its place in the course as a whole. Students wanted more exposure to research using social medial. In response to the statement "I would like for the course to spend more time on big data," 14 of 15 (93.3%) of students at least agreed with 9 (60%) strongly agreeing. Students were also asked "If given the opportunity, I would prefer an expanded big data assignment over the quantitative analysis assignment." The quantitative analysis assignment is generally considered the hardest assignment of the semester, asking students to analyze data on their topic using either contingency tables or regression as appropriate. Sixty percent of students responded positively to this suggestion—six strongly agreed and three agreed—with the remaining students responding "disagree" (4) or strongly disagree" (1).

Table 4—Do Students Want More Big Data Analysis?

	Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**
ReplaceQA	6 (40.00 %)	3 (20.00 %)	0 (0.00 %)	5 (33.33 %)	1 (6.67 %)
MoreOnBigData	9 (60.00 %)	5 (33.33 %)	0 (0.00 %)	0 (0.00 %)	1 (6.67 %)

In free responses following the final two questions in fall 2014, students noted that they found working with social media data, big data, interesting. One student noted, "I found it interesting and I wish more of the semester was dedicated to analyzing it more, instead of just leaning which variable(s) are which, etc." Some students wrote in more depth. One common theme was that they wanted to explore their data more. This sentiment was capture by quotes from three students.

> "I feel like there is still so much I don't know, and think it would benefit me to know more."
>
> "I found it interesting but I felt a little rushed so I didn't get to look into all angles."
>
> "Big data was the best part of the class for me. I'm not sure grade wise, but it was fun. I had said before, and again, I would want more time to have additional time to breakdown by time." "Having more time to divide data and maybe working with others that focus on different social media outlets and comparing the results between each other and different social medias."

The final quote about breaking down data over time was also reflected in this comment that they would "perhaps look at trends at various points in the semester." Students wanted more time and tools to explore how their topic was reflected in social media.

Sentiments like this are not typically expressed about traditional quantitative analysis assignments, although in the free responses to whether to replace quantitative analysis with big data, two students expressed strong support for the more traditional assignment. Most strongly one student noted, "Yes, I loved the regression analysis and I wish I got more chance(s) to play with it more." A couple of students thought the quantitative analysis was too difficult. More students reemphasized that big data was more interesting, and saw it as future oriented.

Overall, students have been positive about the assignments. Despite technical problems in spring 2014, there was enthusiasm in working with the data. With technical issues out of the way in fall 2014 students were clearly looking for more in this assignment.

Results: Evidence of Big Data in Self Reflection and Class Presentation

Did students make use of their big data assignment in their end of semester summative work? In brief: not much in spring 2014, significantly more in fall 2015. Table 5 breifly summarizes the data.

Table 5—References to Big Data in End of Semester Reflective Work, Spring 2015 & Fall 2014

Assignment	Noted	Not Noted
Spring 2014 Presentation (Note: I have no slides for 7 students.)	4 (28.6%)	10 (71.4%)
Fall 2014 Presentation	10 (66.7%)	5 (33.3%)
Fall 2014 Reflection	12 (80%)	3 (20%)

Students clearly make more use of their work with social media in fall 2014. Most presentations that used the presented a chart of key terms and discussed their results at least briefly. Three students discussed this work across at least three slides of the 20 slide presentation. While the time of the assignment in both semesters was similar the timing of the presentation differed somewhat, as some students gave their presentation on the last day of class, before the big data assignment was due. As noted earlier, technical issues also frustrated students.

In their fall 2014 reflections, students referred to the big data assignment as a challenge, even the biggest challenge of the semester (three students mentioned this), but they also talked about the assignment in terms of skills developed (three), the excitement of the new tools (three), the satisfaction of completing the analysis (three), and two students reflected more broadly on social media data in their lives and its potential for social science. These reflections are all positive indicators that the students are highly motivated to engage in this work with the potential for expansion.

Lessons and Looking Forward

The experience with this social media as big data assignment in POLS 200 shows that it has potential to teach students about data analysis, research design from a new perspective. Students were notably enthusiastic about the assignment during both semester, especially so when in the fall 2014 semester when there were no significant technical glitches. Students were motivated by the data and their topic to work through a complex analysis environment that was foreign to all students. Especially when VIDIA worked well, as in fall 2014, students wanted more opportunities to explore this data in more detailed ways. With high levels of interest, students may be better motivated to use traditional data analysis tools to derive meaning from their data.

Having students investigate social media also has the potential to teach students in a very direct way about data privacy. In an open-ended question about what students learned, several students focused on privacy issues although this was not a major focus of the assignment. Examples of comments were:

> "It is amazing how easy it is to find information."
>
> "Everyone's business is everywhere!"
>
> "Everything you say counts."
>
> "Everything is kept somewhere on a database and nothing is anonymous."

These comments make clear that at least some students have a better grasp of data privacy than before. This is notable as many students have attended sessions on managing their online profiles in high school and college.

Moving forward possibilities for improvement and expansion include:

- Developing a method to gather or make available data that is free. One possibility we have begun to explore is using R, the open-source data analysis and programming environment, along with a related application Shiny,[12] that moves processes to the Web in the form of apps, to gather data. R packages are available to access Twitter and to more broadly scrape data from the Internet. This has potential to get out of the trap of paying commercial providers for access today.

[12] A political science example of a Shiny app can be found at SWIID, the The Standardized World Income Inequality Database.

While these apps may not be appropriate for research purposes, they will work well for teaching processes.

- Make truly big data available to students. Another path is to develop or acquire large datasets that students can access to explore as they wish. There are large datasets available by Stanford University's SNAP Group.
- Build out VIDIA and continue build more support for RapidMiner. This high performance computing environment has great potential for teaching.
- In terms of teaching POLS 200. I will work to begin this assignment earlier in the semester. I will also consider creating an option for students to use big data/social media analysis in the longer data analysis assignment.

If others are interested in exploring the tools in VIDIA, we are happy to set up an account for you and investigate your using it with students and your colleagues.

Bibliography

Ambrose, Susan A., Michael W. Bridges, Michele DiPietro, Marsha C. Lovett, and Marie K. Norman. 2010. *How Learning Works: Seven Research-Based Principles for Smart Teaching.* Jossey-Bass.

Ball, Christopher T and Lynn E Pelco. "Teaching Research Methods to Undergraduate Psychology Students Using an Active Cooperative Learning Approach." *International Journal of Teaching and Learning in Higher Education* 17, no. 2 (2006): 147–54.

Barbera, P. 2014. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* (2014): 1-16.

Batanero, Carmen, Juan D Godino, A Vallecillos, DR Green, and P Holmes. "Errors and Difficulties in Understanding Elementary Statistical Concepts." *International Journal of Mathematical Education in Science and Technology* 25, no. 4 (1994): 527–47.

Bekafigo, M. A. and A. McBride. 2013. "Who Tweets About Politics?: Political Participation of Twitter Users During the 2011gubernatorial Elections." *Social Science Computer Review* 31, no. 5: 625–43.

Bernstein, Jeffrey L. and Brooke Thomas Allen. "Overcoming Methods Anxiety: Qualitative First, Quantitative Next, Frequent Feedback Along the Way." *Journal of Political Science Education* 9, no. 1 (2013): 1–15.

Bos, Angela L. and Monica C. Schneider. 2009. ''Stepping Around the Brick Wall: Overcoming Student Obstacles in Methods Courses.'' *PS: Political Science & Politics* 42(2): 375–383.

Brophy, Jere. "Developing Students' Appreciation for What is Taught in School." *Educational Psychologist* 43, no. 3 (2008): 132–41.

Bryant, Antony and Uzma Raja. "In the Realm of Big Data." *First Monday* 19, no. 2 (2014):

Buchler, Justin. 2009. ''Teaching Quantitative Methodology to the Math Averse.'' *PS: Political Science & Politics* 42(3): 527–530.

Cannon, DM, JH Godwin, and SR Goldberg. "Goldman Sachs and Cloud Surfing." Wiley Online Library (2013):

Clark, Tom S, Jeffrey K Staton, Eugene Agchtein, and Yu Wang. "Revealed Public Opinion on Twitter: The Supreme Court of the United States Same-Sex Marriage Decisions."

Dominic, Maria, Sagayaraj Francis, and Anthony Pilomenraj. "E-Learning in Web 3.0." *International Journal of Modern Education & Computer Science* 6, no. 2 (2014).

Eisenberg, Mike, Doug Johnson, and Bob Berkowitz. "Information, Communications, and Technology (Ict) Skills Curriculum Based on the Big6 Skills Approach to Information Problem-Solving." *Library Media Connection* 28, no. 6 (2010): 24–27.

Gal, Iddo and Lynda Ginsburg. "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework." *Journal of Statistics Education* 2, no. 2 (1994): 1–15.

Garfield, Joan B and Dani Ben-Zvi. "Developing Students' Statistical Reasoning." C*onnecting Research and Teaching Practice*. The Netherlands: Springer (2008):

Hogg, Robert V. "Statistical Education: Improvements Are Badly Needed." *The American Statistician* 45, no. 4 (1991): 342–43.

Hood, Michelle, Peter A Creed, and David L Neumann. "Using the Expectancy Value Model of Motivation to Understand the Relationship Between Student Attitudes and Achievement in Statistics." *Statistics Education Research Journal* 11, no. 2 (2012): 72–85.

Hulleman, Chris S and Judith M Harackiewicz. "Promoting Interest and Performance in High School Science Classes." *Science* 326, no. 5958 (2009): 1410–12.

Jure, Leskovec and Krevl Andrej. 2014. Snap Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/.

Klass, Gary M. *Just Plain Data Analysis: Finding, Presenting, and Interpreting Social Science Data.* Rowman & Littlefield Publishers, 2012.

Mallow, Jeffry, Helge Kastrup, Fred B Bryant, Nelda Hislop, Rachel Shefner, and Maria Udo. "Science Anxiety, Science Attitudes, and Gender: Interviews From a Binational Study." *Journal of Science Education and Technology* 19, no. 4 (2010): 356–69.

Marek, Pam, Andrew N Christopher, and Bonnie J Walker. "Learning By Doing: Research Methods With a Theme." *Teaching of Psychology* 31, no. 2 (2004): 128–31.

Marsick, Victoria J and Karen E Watkins. "Informal and Incidental Learning." *New directions for adult and continuing education* 2001, no. 89 (2001): 25–34.

Mayer-Schönberger, Viktor and Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform how We Live, Work, and Think.* New York: Houghton Mifflin Harcourt.

Navarro, Daniel. "Learning Statistics With R: A Tutorial for Psychology Students and Other Beginners (Version 0.4)."

Newmann, Fred M, Helen M Marks, and Adam Gamoran. "Authentic Pedagogy and Student Performance." *American Journal of Education* (1996): 280–312.

Pew Rsearch Internate Project. 2014. Key Indicators: Three Technology Revolutions. http://www.pewinternet.org/three-technology-revolutions/

Press, Gil. 2014. *Forbes*. "12 Big Data Definitions: What's Yours?" http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/

Ruths, Derek and Jurgen Pfeffer. 2014. Social media for large studies of behavior. *Science* (346: 6213(November 28): 1063-1064.

Schau, C. "Students' attitudes: The "other" important outcome in statistics education." *Proceedings of the Joint Statistical Meetings* (2003): 3673–81.

Schutz, Paul A, Lisa M Drogosz, Vicki E White, and Chris Distefano. "Prior Knowledge, Attitude, and Strategy Use in an Introduction to Statistics Course." L*earning and Individual Differences* 10, no. 4 (1998): 291–308.

Solt, Frederick. 2014. "The Standardized World Income Inequality Database." Working paper. SWIID Version 5.0, October 2014. http://myweb.uiowa.edu/fsolt/swiid/swiid.html

Tremblay, Paul F, RC Gardner, and Gerwin Heipel. 2000. "A Model of the Relationships Among Measures of Affect, Aptitude, and Performance in Introductory Statistics.". *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 32:40.

Turner, Charles C., and Cameron G. Thies. 2009. "What We Mean By Scope and Methods: A Survey of Undergraduate Scope and Methods Courses". *PS: Political Science & Politics* 42:367.

Xian, H and K Madhavan. "Anatomy of Scholarly Collaboration in Engineering Education: A Big-data Bibliometric Analysis." *Journal of Engineering Education* (2014):