# ENG1131: Writing through Big Data

by Aaron Beveridge  
 [Course Design](https://www.uc.edu/journals/composition-studies/submissions/course-designs.html) article for Composition Studies

## Course Description

Students in Writing through Big Data write about and research social media trends. In the context of this course, writing is defined under its broader multimodal conception that includes text, image, audio, video and any other digital artifact intended to make, reproduce, or remix meaning. Trends are popular and sometimes viral topcs that circulate among social networks such as Facebook, Twitter, Youtube, Vine, and Instagram. Students for this course will investigate trends occurring within Twitter, and will visualize trend data in order to understand the way writing moves, circulates, and changes over time. All of the work in this course will build toward the final project where students present their semester-long research. Weekly workshops will teach students how to collect writing data and how to visualize this data in order to form an argument about their chosen trend. All of the readings for this course deal with data literacy, with the forming of arguments through data collection and visualization, and with the design and preparation of research presentations.

## Institutional Context

The University of Florida is a large federal land-grant institution housing 16 academic colleges and more than 150 institutes of research. In 2014, UF had a total enrollment of 50,536 students, and UF is among the 10 largest universities in the United States. However, unlike many other top universities, the English department is distinctly separate from the University Writing Program (UWP) at UF—each maintain their autonomy as far as administration, institutional goals, curriculum development, and student instruction are concerned. The primary point of connection between these two entities is facilitating English department graduate students as instructors for many of the UWP course offerings. This allows the English department to fund 100% of its graduate students, while providing the university with the requisite staff for first-year writing, writing across the curriculum, and the Writing Studio (UF’s peer writing/resource lab).

The benefit to such autonomy between the English department and the UWP is the diversity of objectives for courses taught within the English department. Since it is the role of the UWP to prepare students for literate writing practice within the university, writing courses within the English department are able to expand the available genres, mediums, and modalities to encourage a wide range of possiiblities for course development. Writing through Big Data is based on the Writing through Media (WtM) course template envisioned by Gregory Ulmer. As Ulmer explains, WtM was originally “an extension of ‘writing about literature’” that was expanded to include “entertainment and popular culture media (cinema, television, music, video games, pop literature, comics, magazines, zines, and the like).” However, what sets the WtM template apart from the “writing about literature” course is the shift in *methodology* as well as course content. Ulmer explains that this “shift is approached through its rhetorical implications, with the students as makers (and not just consumers) of new media affects.”

UF has a series of introductory writing courses that students must complete called 1101 and 1102. These courses follow traditional composition instruction guidelines: > Composition courses equip students with the skills necessary to complete successfully the reading > and writing requirements of their disciplines. In addition to fulfilling a portion of the Gordon > Rule communication requirement, composition courses offer instruction in methods of writing, > conventions of standard written English, reading and comprehension skills, and ways of making > expository and argumentative prose accessible to readers in varied situations. These courses are > writing-intensive and require extensive practice, and each writer receives feedback for revision. Even though 1131 is not subject to the same guidelines, like 1101 and 1102, 1131 must adhere to the “Gordon Rule” for the amount of text written in each course. Students in 1131 must write at least 6000 words by the end of the semester, and this rule can be intimidating for non-traditional courses that attempt to engage in visual rhetoric and new media. For example, I had to find a way to calculate how infographics and oral/slideware presentations would count toward the required total, and fortunately UF allows instructors to provide equivalent justifications for the required word totals. Students produce two oral/slideware presentations for a total of 20 minutes in my course, and I reverse-engineered word totals based on the average reading time for presenting papers. On average it takes 2 minutes to read 1 page of double-spaced text, and thus 20 minutes of presentation is equivalent to 10 pages of double-spaced text. Since a single double-spaced page contains 250 words, my students’ presentation totals were arguably equivalent to approximately 2500 words.

Once the curriculum guidelines were met, the development of Writing through Big Data fit well with other projects already underway at UF. The English department—in partnership with university libraries and UF’s Research Computing Department—was awarded a Level II Startup Grant from the National Endowment of the Humanities[1](#fn1). This grant funds the continued development of an open source research software called MassMine[2](#fn2)—a social media mining and archiving application that simplifies the process of collecting and analyzing social network data. As the co-creator of MassMine, along with Nicholas Van Horn, my responsibilities include testing MassMine’s core application, the programming of additional data processing and analysis tools, and the development of training materials and pedagogical applications for the humanities. Sidney Dobrin, Chair of the English department at UF and author of *Postcomposition*, is the project director for the grant. Because of the support of Dobrin (and the previous Chair, Kenneth Kidd), the funding provided by the NEH[3](#fn3), and the successful testing and implementation of the software at UF, there has been immense support for the development of a writing studies course focused on social media data mining, analysis, and visualization.

In addition to the support provided by the department, the University of Florida has forwarded multiple initatives focused on interdisciplinary data literacy. Since 2011, Florida’s Research Computing division has been expanding its services and it provides high performance super-computer resources to students and faculty across all disciplines. Research Computing provided free server space and computing resources to all of the students in WtBD. The accessibility of super computer clusters for teaching and research is becoming far more common in universities across the United States, and WtBD was the first humanities course at UF to apply these tools to the study of networked writing. Furthermore, UF has also created an interdisciplinary Informatics Institute to support data-intensive research, and they are one of 16 programs at UF that recieved funding in 2013 from a $15 million pre-eminence package provided to UF by Florida’s legislature[4](#fn4). In addition to the broad interdisciplinary support for courses such as WtBD, UF added a data literacy course to its core requirements for all students as part of its “Grand Challenges Core” intitative.

While WtBD does not immediately seek to fulfill this core requirement, it is within this broader environment of support for Big Data methodologies that WtBD intersects writing studies with data science paradigms. XXXSummarize and conclude.XXX

## Theoretical Rationale

There are two key reciprocal components that form the theoretical rationale for Writing through Big Data. First and foremost, WtBD is a writing studies course—where the object of study is *networked writing*. Networked writing has taken many various forms since the invention of the Internet, and many new forms will no doubt continue to emerge. However, in recent years, social networks have become an important site for the study of networked writing. Because of the immense growth in worldwide users and the influence that this form of networked writing has on culture, politics, journalism, and entertainment, social networks are a valuable resource for studying contemporary intersections of writing and culture. While many aspects of social networks are investigated during WtBD, students focus their research primarily on social network *trends*. Trends provide a macroscopic view of large-scale patterns occurring in the topics and artifacts of networked writing, but trends also provide a way of understanding the networks themselves. As Sidney Dobrin explains regarding networked writing in *Postcomposition*, “writing-as-system interrelates with networks rather than suggest writing is itself an identifiable network” (181). According to Dobrin’s understanding of “writing-as-system,” writing is structured by networks, but writing also restructures the system as it saturates and permeates networks: “Network emerges from writing and depends on its saturation, its fluctuation, and its mass. Without being saturated by and within writing, the network would neither emerge nor evolve, nor would the connections between the nodes and knots serve the network to any degree of circulation” (184). Often with #hashtags serving a nodal function, trends arise from user writing—from the systems and topics already available within a network—but writing also moves and changes those systems by allowing new connections and ideas to emerge.

In addition to networked writing as the object of study for the course, the second key component of Writing through Big Data is data literacy. In “Rhetorical Numbers: A Case for Quantitative Writing in the Composition Classroom,” Joanna Wolfe explains that “there is a paradox in that on one hand our culture tends to represent statistical evidence as a type of ‘fact’ and therefore immune to the arts of rhetoric, but on the other hand we are deeply aware and suspicious of the ability of statistics to be ‘cooked,’ ‘massaged,’ ‘spun,’ or otherwise minipulated.” In other words, Wolfe argues, “Treating numbers as inherently truthful or inherently deceptive is equally naive” (453).

In Writing through Big Data, students spend roughly the first 1/3 of the course determining which trend they want to research during the rest of the semester. While trends are complex in how they emerge and how they may be observed within various social networks (more on this later), trends are simply popular topics or digital artifacts (such as memes or viral videos) that are written about or shared by a large number of users within a network. Trends may arise from particular events and then transform into larger issues or concepts that users continue to discuss, and often these broader concepts become categorical tags that are used to create connections among other newer topics or events (associative connections made through the use of #hashtags). For example, in the case of #blacklivesmatter, motivated by the killing of Michael Brown by police officer Darren Wilson, the protests in the town of Furguson transformed into a much larger trend focused on police violence and institutional racism. The tag #blacklivesmatter was then later used to identify other similar acts of police violence, to motivate protests against institutional racism, and to redirect a broader conversation about racism in contemporary society. Certainly, for a trend like #blacklivesmatter, the amount of attention it recieves within a single network like Twitter (or the extent to which it appears to be “trending”) should not be used to reductively explain its broader social impact and relevance. As students are choosing the trends they will research, they must confront the difference between a trend within a single network like Twitter, and the broader complex media ecology in which networked writing occurs. Just because something is not trending within a particular network does not mean that it is not important or that it does not have value. And this awareness—a critical engagement with how trends are defined and how they are currently used to justify something’s importance, cutural value, or the amount of ongoing attention it ought to receive—this provides the most immediate rationale for Writing through Big Data.

Once students choose a trend, they are provided with a basic research question that frames their work for the course: what is the exigence of your chosen trend? Why are people paying attention? Why does it matter? What may have caused it to gain such a following? Why does it appear to have such momentum? In addition to pulling trend data from Twitter with MassMine, this means that students must conduct outside research and look for news stories, blogs, or other sources that may help contextualize their trend. Students are also encouraged to look at other social networks like Facebook or Reddit to see if their topic is trending on those networks as well. This provides the opportunity to discuss algorithmic affects on how trends function. Obviously, algorithmic engagement cannot be too in-depth within an undergraduate writing course,

Because of the algorithmic control over trends, and the way many topics may be suppressed as a result, students are encouraged to choose topics that may *not* be trending at a high level, and therefore transform the course’s research question into a persuasive project: why *should* people pay attention to your chosen trend? While a course of 15 to 30 students will have little affect on what is trending on a network like Twitter, this reminds students that WtBD is not merely an observational course focused on conducting “objective” research. XXXBurke’s parlorXXX

While many students choose #hashtags that are associated with an event, movement, or broader concept, they also have the option of researching semantic #hashtags that are part of the common internet lexicon. Talk about #salty, and the article the CGB recommended.

Trends are often associated with tagging—or #hashtags—the system that allows users to categorize their

Students often begin with #hashtag trends—the tagging system used by many social networks that allows users to associate their texts, images, and videos with other other

At it’s core, Writing through Big Data takes a “writing about writing” approach to the advanced writing classroom. However, instead of writing about traditional forms of codex prose and essays, students write about *networked writing*—writing that occurs on Twitter’s social network. When students in WtBD “write about writing”—this assumes the broader definition of writing that includes any form of digital inscription or meaning-making.

Students in WtBD spend the first weeks of the semester observing trends on Twitter. The goal is for them to

While rhetoric and writing instructors have worked to reduce the digital divide through multimodal pedagogies and courses focused on new media design, much work remains to confront issues surrounding data literacy and the rhetorics of data visualization. Therefore, the broadest rationale for *Writing through Big Data* is a course designed to meet these challenges by teaching students to produce their own data visualizations in order to understand information design as rhetorical practice.

As explained earlier, however, students work toward such an understanding not only through observing or reading about data visualizations, but primarily through producing data visualizations of their own.

The object of study for such research *is* writing–networked writing that occurs within social media. Get writing about writing quote hereXXX.

XXXExplain the “flipped classroom,” and how this affects the grading of this course.XXX Review recent trends in rhetoric and writing studies that suggest that data literacy needs to be more of a focus.

Using natural language processing in English is not a new concept. Explain the brief history from Ramsay’s *Reading Machines*. However, applying these technologies to writing studies–more specifically, networked writing occuring on social networks–remains a new area of exploration for writing studies research. ent

The benefit of such a methodology for study and research is that it directly leads to the the coureses writing objectives–or, rather, it leads directly leads to the medium through which students will “write” in the course–data visualization.

One of the student’s tasks is to create an infographic that explains the “exigence” of their trend. In fact, for the first 1/2 of the course, this is the reserach question provided to them. Of course, they are encouraged to modify this question or change it entirely once they have dug deeper into their trend.

XXXFor one of their discussion posts, students are asked to find a data visualization that does not allow them to investigate the underlying data and/or the methods of its creation. By so doing, they are introduced to a new awareness of this rhetorical artifact, and how prevalent they have become in contempory media.

## Critical Reflection

## Syllabus

#### Writing through Big Data

Course Instructor: Aaron Beveridge  
 Course Number: ENG1131-1788  
 Regular Class Time: MWF – Period 5  
 Workshop Class Time: W – Period E1-E3  
 Office Hours: W – Period 6  
 Office: TUR 4309  
 Room: CSE E211A

#### Course Description

Students in Writing through Big Data will write about and research social media trends. Trends are popular and sometimes viral topics that circulate among social networks such as Facebook, Twitter, Youtube, Vine, Instagram, etc. For this semester, we will consider trends occurring within Twitter, and we will visualize data on these trends in order to better understand the way writing moves and circulates. All of the work in this course will build toward the final project where students present the data they have collected and visualized. Weekly workshops will teach students how to collect writing data and how to visualize it to support an argument about their given trend. All of the readings either deal with data collection and visualization or with the design and preparation of presentations for student research.

#### Course Rationale

Writing through Big Data is a variation of the Writing through Media series of courses offered by UF’s English Department. As the English Department website explains, the goal of the 1131 course “is to introduce students to the transition underway between literacy and post-literacy (electracy) in contemporary culture.” The students in writing through media courses are intended to be “makers (and not just consumers) of new media.” Writing through Big Data stays true to 1131’s core concepts. Students in WtBD will accomplish two primary goals: (1) study contemporary writing as it occurs within social media, and (2) work to understand data visualization as a form of rhetorical practice by producing and reflecting on data visualizations of writing. In this class we will visualize data resulting from research conducted on social media trends. Trends identify large patterns of writing occurring in social media, and this class will research and visualize current trends as chosen by students.

Writing within this course is understood broadly as any available mode of communication: text, image, audio, video, or any available technical or artistic means of making meaning. In the era of Big Data, writing is undoubtedly chaotic. The flow of information is so continuous and widespread that making sense of writing requires using the tools and techniques of data science. Google CEO Eric Schmidt has said that we now create as much information every 48 hours as we did from the beginning of time until 2003—and he said this 5 years ago. In order for students to understand how to effectively read and write within contemporary, complex forms of mediation, they need to know how to quickly aggregate and reconstruct large amounts of information. In this class students will collect data on writing as it occurs within social networks (specifically, Twitter), and then create visualizations of that writing data to arrive at a more complex understanding of contemporary writing and rhetorical practice.

**Required Texts:**

* Nathan Yau’s *Visualize This*
* Nathan Yau’s *Data Points*
* Garr Reynolds’ *Presentation Zen*

**Required Software/Technology:**  
 *All software listed below is free and open source*

* Laptop (If you do not own a laptop, I will help you secure one for the semester)
* R — <http://cran.r-project.org/>
* R Studio — <http://www.rstudio.com/>
* Inkscape — <https://inkscape.org/en/>
* MassMine

#### Course Assignments:

**Attendance**  
 This class is built around workshop and classroom participation, and therefore attendance is mandatory. Students are allowed to miss 2 of the regular 50 minute classes, and 1 of the 3-hour Wednesday night classes without negatively affecting the final course grade. Upon receiving a 3rd regular absence or a 2nd Wednesday night absence the overall grade in the course will drop by 5 points, and it will continue to drop by an additional 5 points for each subsequent absence. For example, 5 additional absences beyond the 3 allowed absences (8 in total) will drop a student’s total class grade by 25 points.

**Workshop–20 points**  
 The workshop grade is based on completion of activities and research during Wednesday night workshops. If students are absent for a Wednesday night workshop, then activities must be completed and turned in by the Friday class meeting time in order to receive credit. During the workshops, the first 1.5 hours of class will be spent completing the work in the workbook. After workbook work is completed, students are expected to spend the rest of the time working toward their individual research for their final project.

* Workshop criteria:
  + Finish projects, textbook assignments, or class work within the time provided
  + Works consistently
  + Prepared for class—having read assignments ahead of time as directed
  + Provides peer with thoughtful feedback or assistance depending on assignment

**Mini presentation–10 points**  
 The mini presentation allows students to present an abstract of their final project to their peers for feedback. The mini presentation allows students to test their chosen presentation method, as well as practice presenting for the first time. The mini presentation will be part of their final presentation. Students will not present writing data at this time, but they will be expected to have a well-developed research idea to present to their peers.

* Presentation criteria:
  + No less than 5 minutes, no more than 6
  + Presentation appears practiced
  + Presenter does not read from slides, but has prepared notes to assist with presentation
  + A well-developed research idea is presented

**Mid course written exam–20 points**  
 The mid semester written exam will be open-book and open-Internet. The point of the mid term exam is to provide the opportunity for course reflection and synthesis. The exam will be responding to course readings, lectures, and workshop materials. Students will be provided with a general subject ahead of time, but the writing prompt will not be provided until the written exam. Criteria for this assignment will be provided along with the writing prompt at the time of the exam.

* Exam criteria:
  + Minimum of 750 words
  + Displays an understanding of course readings and lectures
  + Adequately responds to the exam questions

**Blog Writing and Responding–20 points**  
 Students will produce 10 discussion posts of 200 words, and will additionally respond to one of their peer’s blog posts for every posted assignment. When a peer responds to an initial blog post, an additional response is required from the writer of the blog post. The 200 word initial post counts as 50% of the credit given to blog posts, and substantive responses to peer posts count as 50% of the credit. There is no minimum word count for peer responses, but responses must be substantive. Simple affirmation like “good job” or “nice post” will not count as a response. A substantive blog response should encourage further conversation, raise a question with the post, try to clarify an idea, or productively add to the topic. The weekly discussion is a core component of this course and will be graded on a weekly basis. The responses and conversations that follow initial posts ought to remain respectful and courteous.

* Blog post criteria
  + Blog posts are due by class time
  + 200 words minimum
  + Relevant to course trend research
* Response criteria:
  + Substantive
  + Timely– initial responses due by Wednesday morning class time
  + Secondary responses by Friday morning class time

\*If blog posts do not meet the basic criteria, students will receive an email directing them on how to improve the quality of their posts. If quality does not improve, no credit will be given for subsequent posts following the email.

\*\*Writing tip: Do not draft your blog posts within the browser window provided by the blog software. Draft your post within your favorite word processor or text editor and then copy/paste your text into the browser. Google docs is also acceptable for drafting posts since it saves frequently and will keep you from losing work.

**Final presentation–30 points**  
 The final presentation motivates the work for the entire semester. The textbook selections and course readings, the workshop, the blog posts–all are designed to move students toward the completion of research for the final presentation. The research presented will investigate a particular trend occurring within Twitter. A trend is defined as either (1) a top ten trending topic, word, or concept from any of the available trend locations as defined by Twitter, or (2) a #hashtag trend of social or political importance that a student finds worthy of further research.

An example research project may be a trend similar to the #ferguson trend that responds to the shooting of unarmed black teenager Michael Brown by a white police officer. By collecting a large sample of tweets associated with the #ferguson trend, students are able to parse their tweets according to other words or #hashtags that appear in the tweets. This data provides the basis for investigating other words or trends associated with #ferguson. Students then pull information on the associated trends to build a more complex representation of how the initial #ferguson trend is associated to other important trends. Students may also map #ferguson and its associated trends to see where and when these trends appear/iterate most frequently since trend data from Twitter includes location and time information. Finally, students compare this data to other “outside” data in order to make a final claim. For example, students could map national data on police shootings of black Americans and compare this to their own data. Students could collect outside data from a handful of various perspectives—allowing them the opportunity to make broader claims about the importance or relevance of their trend.

Final projects do not need to be limited to the types of analyses and visualizations described above, but they need to substantially explore the trend chosen by the student and investigate its relevance, exigence, and impact. The class will collaboratively determine the criteria for the final assignment following the “Final project introduction” lecture on the Friday of week 6. The components of the final assignment to be chosen by the class will be the following:

* Components of final assignment:
  + Presentation software (PowerPoint, Prezi, Google Presentation, etc.)
  + Delivery method (How will this research be re-delivered after presentation? YouTube, Facebook, Twitter, class website. What is our broader audience?)
  + Research (How do we determine a well-investigated research question?)
  + Data (How many data visualizations should be required? What types of data should be presented? How much space in the presentation should be given to data description and analysis?)

#### Total Points for Semester: 100

#### Course Calendar

**Week 1**  
 1.7 Course Introduction and Syllabus Overview  
 E Visualize This chapter 1-2, Software Installation and Intro to Workshop  
 1.9 “What is writing studies?” lecture – Prezi

**Week 2**  
 1.12 Data Points chapter 1 “Understanding Data”  
 1.14 Revising Prose chapters 1-3  
 E Workshop–MassMine/Intro to Research Computing  
 1.16 Blog post due

**Week 3**  
 1.19 No Class  
 1.21 Visualize This chapter 3 “Choosing Tools to Visualize Data” Visualize This chapter 4 “Visualizing Patterns Over Time”  
 E Workshop  
 1.23 Data Points chapter 2 “Visualization: The Medium”, Blog post due

**Week 4**  
 1.26 Data Points chapter 3 “Representing Data”  
 1.28 Visualize This chapter 5 “Visualizing Proportions”  
 E Workshop  
 1.30 Blog post due

**Week 5**  
 2.2 “What is visual rhetoric?” lecture – Prezi, Blog post due  
 2.4 Visualize This chapter 6 “Visualizing Relationships”  
 E Workshop  
 2.6 Data Points chapter 4 “Exploring Data Visually”

**Week 6**  
 2.9 “Final project introduction” lecture – Prezi  
 2.11 Visualize This chapter 7 “Spotting Differences”  
 E Workshop  
 2.13 Blog post due

**Week 7**  
 2.16 Mandatory individual conferences, Blog post due  
 2.18 Mandatory individual conferences  
 E Workshop  
 2.20 Mandatory individual conferences

**Week 8**  
 2.23 Data Points chapter 5 “Visualizing with Clarity”  
 2.25 Review for mid course written exam  
 E Mid course written exam  
 2.27 Workshop, Blog post due

**Week 9**  
 Spring Break

**Week 10**  
 3.9 Data Points chapter 6 “Designing for an Audience”  
 3.11 Visualize This chapter 8 “Visualizing Spatial Relationships” Visualize This chapter 9 “Designing with a Purpose”  
 E Workshop  
 3.13 Blog post due

**Week 11**  
 3.16 Data Points chapter 7 “Where to Go from Here”  
 3.18 No class, Blog post due  
 E No Class, read Presentation Zen  
 3.20 No Class, read Presentation Zen

**Week 12**  
 3.23 Presentation Zen discussion, Blog post due  
 3.25 Workshop  
 E Mini presentations, peer response  
 3.27 Workshop

**Week 13**  
 3.30 Workshop  
 4.1 Workshop  
 E Workshop  
 4.3 Workshop

**Week 14**  
 4.6 Individual conferences  
 4.8 Individual conferences  
 E Workshop  
 4.10 Individual conferences

**Week 15**  
 4.13 Workshop  
 4.15 Workshop  
 E Workshop  
 4.17 Workshop

**Week 16**  
 4.20 Final presentations  
 4.22 Final presentations  
 E Final presentations and course wrap-up

## Works Cited

* <http://ufdc.ufl.edu/AA00025642/00001>[↩](#fnref1)
* <http://www.massmine.org>[↩](#fnref2)
* At the time the class was approved for teaching, the grant application was submitted but not yet funded. However, the department supported the course based on its fulfillment of curriculum guidelines and MassMine’s successful testing and development on Research Computing’s system.[↩](#fnref3)
* <http://www.ocala.com/article/20131009/ARTICLES/131009647>[↩](#fnref4)