

JRN 263: Big Data, Small Screens: The Art and Science of Responsive Data Journalism

Carl V. Lewis

Fall 2017 - University of Florida College of Journalism and Communications

| “Data journalism is the new punk. Anyone can do it.”

- **Simon Rogers, *The Guardian*

| * “If it doesn’t work on mobile, it doesn’t work*.”

-Brian Boyer, NPR Visuals Team Leader. * * *

COURSE LECTURER/LAB INSTRUCTOR: CARL V. LEWIS

- Contact Email: cvl2103@caa.columbia.edu (preferred contact method)
- Contact Phone: [912.816.7007](tel:912.816.7007) (if any *urgent* matters arise, you may send a brief text message to this number stating your name and the course).
- Course Homepage: <http://ddjrn1263.github.io>
- Course Location/Time: *Weimer 204, MWF, 1 - 1:50 P.M. Lab: M, 2:00 p.m.-5:00 p.m.*
- Dedicated Office Location/Hours: Weimer 103, MWF, 2 -3 p.m.
- On-Demand Office Hours Reservations: <http://bit.ly/dataj-office-hours>

I. Rationale

Unlike any time before in our lives, we have access to vast amounts of free information—a phenomena scholars refer to as “[big data](#).” With the right tools, we can start to make sense of all this data to see patterns and trends that would otherwise be invisible to us. By transforming numbers into graphical shapes of and interactive web apps, we can allow users to understand the stories those numbers hide. This process of gaining new insights from data and crafting visual stories to convey that data in an accessible and engaging format is part of what we call “[data journalism](#)”—a specialty subset of journalism that reflects the increased role that numerical data is used in the production and distribution of information in the digital era.

Data journalism, which typically uses [visualization](#) as its primary means of communication, has rapidly emerged during the past decade as an in-demand skill-set across nearly all sectors of digital media—helping audiences understand complex social and public policy issues ranging from climate change, to the U.S. Electoral College, to the growing the influence of money in politics. Yet more recently — as the majority of users worldwide now access the web via mobile devices with smaller screen sizes and touch-screen interfaces — the task of creating clear, usable and truthful data visualizations has taken on an added level of complexity. We now find ourselves forced to squeeze large amounts of data into limited screen real-estate, and having to rethink the way we approach the workflow of data journalism on a fundamental level to account for what we

call *responsive design* techniques. Developing theoretical models — and possibly even practical solutions to — this broader challenge of displaying big data on small screens will guide our thought process as we learn the history, theory, guidelines and basic technical workflow of data journalism.* *

II. Overview and Learning Outcomes

This four-credit hour lecture and lab course will approach data journalism from both a scholarly and a technical perspective. For purposes of this course, we define ‘data journalism’ in the broadest sense as both an art and a science, and do not differentiate between disparate yet closely-related sub-branches of data journalism such as ‘*data-driven journalism*,’ ‘*database journalism*,’ and its earlier predecessor, ‘*computer assisted reporting*.’ As such, we will cover the entire data retrieval, analysis and communicative process, including data sourcing, reporting and computational processing methods* *in addition to** data visualization, perceptual accuracy, design principles, front-end development basics, and publishing methods. The course has two high-level learning goals:

To analyze, reflect and apply the historical, theoretical, statistical and ethical principles behind the nascent field of data journalism by engaging with key texts in the fields of information design, interactive visualization, computer-assisted reporting (CAR) and what Phillip Meyer calls “precision journalism.” To impart basic-to-intermediate technical skill-sets needed to tell data-driven stories visually and interactively on all screen sizes and devices via responsive design techniques.

Part *information design*, part basic statistics, part *computer science*, part *data science* and part *shoe-leather reporting*, students should leave this course with a thorough grasp of both the *process* and the *product* of obtaining, processing, analyzing and visualizing massive datasets and spreadsheets to convey an accurate and cohesive visual narrative to a non-professional audience on large, medium and small screen sizes alike.

** Students are not expected to become technical experts in data journalism, much less in responsive data visualization, in this course.** Understanding the *concepts* behind* *the tools we use for data journalism will be much more important than learning any one discrete tool itself. This assumption aligns with my fundamental belief as a data journalism practitioner: that practical knowledge of computational technology can only reach its full potential in practice when undergirded by sound academic reflection on how we create, perceive and interact with data displays for general audiences — either as journalists, as media consumers, or citizens of the world.

That said, the successful student will complete this course having grasped the following more specific technical and analytical concepts on a basic level:

- The use of *visual thinking* and *data storytelling* to create order out of disorder and make sense out of abstract numbers in an era of *information overload*.
- How to locate and retrieve data from the web in the correct format for the task at hand as well as how to request data from government agencies via a FOIA request— tasks that are analagous to “finding sources” in traditional reporting.
- The use of spreadsheets for data analysis, formatting, manipulation, tabulation and exploration — or what Derek Willis *calls* “interviewing” the data. Using pivot tables, column sort, recurring formulas, and simple overview charts for analysis.
- The journalistic ability to find the data *lede* — or main story — amid large, raw datasets often obscured by extraneous data and information overload. Specifically, deduping, finding fuzzy matches, using find and replace
- The statistical ability to “fact-check” what the data appears to indicate with basic statistics; primarily, calculating the level of correlation, standard deviation and other important values such as the *median*, *percent change*, variance, *range*, *z-score* and *p-value*. And why using just the average (mean) to describe a dataset should be avoided.
- The application of *computational thinking*, *numeracy*, *graphicacy* and what Mary Jo Webster calls a ‘data mindset’

- How to analyze and critique visualizations in newspapers, blogs, books, TV, etc., and how to propose alternatives that would improve them — otherwise known as “critique by redesign.”
- A general grasp of [best practices](#) for selecting the appropriate visualization type that conveys the story at hand (a.k.a., choosing the right chart type or, possibly, no chart at all).
- The role of [visual perception](#) of size, space, distance, position and color in giving meaning to data visualizations — and how those visual elements can [distort](#) the truth if used improperly.
- How visualizations and graphical displays can mislead or confuse if overly-embellished or deceptively presented.
- Mapmaking made [easy](#), and common geospatial data formats such as GeoJSON, TopoJSON, KML and SHP. Also, binding geographic data to numerical data via column joins.
- The power of design to enhance cognition and usability (emotional design, human-centered design) or, alternatively, when implemented for the sake of aesthetics alone, abstract and complicate the presentation, causing [information overload](#).
- The workflow of creating a basic modern web app (HTML *for structure*, CSS *for style*, JS *for interactivity*, JSON* for data*) and basic syntax structure for each.
- What an iFrame is (translation:* a ‘window’ into another webpag*e), and how to make iFrames responsive without distorting the data dimensions of a visualization (tip:
- The primary file formats of data storage for spreadsheets ([CSV](#)), geospatial information ([SHP](#), [KML](#), [GeoJSON](#), [TopoJSON](#)) and fast-scaling data for web applications ([APIs: JSON, XML](#)).
- Useful [current tools](#) to speed up and reduce the programming knowledge necessary to create basic-to-intermediate complexity visualizations.
- The value of exploratory visualizations, interactive web apps and immersive media in personalizing the user experience and providing greater relevance, lending an air of objective credibility, weakening confirmation bias and increasing engagement and time spent.
- General strategies to apply responsive design techniques to data visualizations using Twitter Bootstrap and responsive iFrames *when possible* (in some cases, the scale of the dataset may be so large and multivariate that there is currently no acceptable solution to displaying it in a truly responsive fashion, but an **automatic A+, my scholarly admiration, recommendation letters for life, and probably some super important industry award and fellowships if you happen to devise a solution to this vexing problem in the field!**).
- How to achieve the modern-day version of what pioneer information design scholar Edward Tufte calls “[graphical excellence](#).”

These are all meant merely as *predicted* outcomes for taking this course, not definitive requirements. The level of depth we reach will vary depending upon the class’ ability to keep pace and your individual skill-sets coming into the class. What I do fundamentally expect you to gain is an understanding of and, hopefully, a *passion* for data storytelling. In other words, I aim for you to make a sincere effort to explore what makes strong data visualizations both functional and truthful; what makes data journalism both an art and a science; what makes data journalism both emotionally persuasive and quantitatively sound.

While no specific prerequisite courses or prior programming knowledge is necessary to take this course, at least one (1) introductory-level reporting class or equivalent journalistic experience is *recommended*, but not mandatory. While the subject matter may seem complex or daunting at first, I assure you that—with a little enthusiasm and self-belief coupled with sincere effort—you will eventually have an “Aha!” moment and get it. Trust me,* it’s not rocket science*, but I’m confident it can be just as rewarding as rocket science if you come into class open-minded and eager to learn.

Specific data journalism tools and technologies fluctuate rapidly, but key concepts such as numeracy, graphicacy and visual perception do not. Having said that, this course will – for purposes of experience, workflow and possible portfolio-quality end-products – cover the real-world application of current “Web 2.0” responsive data visualization technologies and standards. These conventions include the basic building blocks of

a modern native web app — HTML5, CSS3 and JavaScript — as well as the most commonly used data storage formats in currency today: CSV (comma-separated-value files) and JSON (JavaScript Object Notation). We'll also examine a number of JavaScript *libraries* (basically, collections of goodies) for responsive data journalism such as [d3.js](#) and [Leaflet.js](#).

If any or all of these terms sound scary or foreign to you at the start of the course, **do not fear:** They frightened me initially as well. You will not be expected to leave this course as a JavaScript engineer, nor a data scientist, nor a web developer. But you will be expected to carry out simple, guided in-class assignments that will help you to begin to understand how these various technologies appear in text (or, more precisely, code) and how they interact together to render a web page in the modern browser. **In addition, if you at any point feel like you need extra help understanding *any* concept or technology outside of class, you are encouraged to schedule one-on-one instruction time with me using the [schedule reservation booking page at the beginning of this syllabus](#).** Please allow 24 hours before an upcoming deadline to schedule a meeting to ensure you leave adequate time for us to meet. As your instructor, I will be acting first and foremost in my official capacity as an educator, but also in an unofficial capacity as might a professional editor.

For data journalism *production* purposes, we'll primarily be employing a number of open-source, graphical, drag-and-drop tools that require little-to-no coding knowledge, and instead focus the bulk of our efforts more upon data sourcing, processing, analysis, design and presentation. *That said, we will still be conducting regular exercises and assignments that introduce basic programming concepts and practices*, as graphical user interface tools cannot in all cases do all the heavy lifting for us. The more complex the dataset, the higher the likelihood some coding may be needed (and, in turn, the higher the likelihood you will need to harness the interactive power of JavaScript).

III. Course Materials

Each class will include different assigned readings from mostly free online sources (blog posts, essays, social media, open-source e-books or excerpts from other scholarly sources). Any readings not already listed in course schedule at the end of this syllabus will be posted to the [course website](#) and Blackboard schedule at least three (3) weeks prior the class in which we discuss them.

That said, four (4) books will be required for purchase or loan for the duration of the course:

1. Cairo, Alberto. *The Functional Art: An Introduction to Information Graphics and Visualization*. Berkeley, CA :New Riders, 2013. Print.
2. Norman, Don. *The Design of Everyday Things*. Basic Books, 2013. Print.
3. The New Precision Journalism
4. Learning Responsive Data Visualization

The three titles above will all be available in the Campus Bookstore, all less than \$20.00 each. In addition, students will be expected to download and have on-hand the full-text of two free e-books:

1. *The Data Journalism Handbook: How Journalists Can Use Numbers to Improve the News* from O'Reilly Press (available for free download at datajournalismhandbook.com).
2. Chiasson, Trina and Dyanna Gregory. (2015) *Data+Design: A simple introduction to preparing and visualizing information*. Infoactive. (available for free download at <https://infoactive.co/data-design/>)

You will also want to follow *at least three* of these blogs, as I will likely refer to them in class discussion time, thus benefiting your participation grade if you, too, have been following them:

Evergreen Data: <http://stephanieevergreen.com/category/blog/> *FlowingData:* <http://flowingdata.com/> *
Information is Beautiful: <http://www.informationisbeautiful.net/blog/> * *Junk Charts:* <http://junkcharts.typepad.com/>
PolicyViz: <http://policyviz.com/> *Storytelling with Data:* <http://www.storytellingwithdata.com/> * *Ann's*

Blog: <http://annkemery.com/blog/> * Storybench: <http://storybench.org> * Data-Driven Journalism: <http://datadrivenjournalism.org>
* Subreddit r/dataisbeautiful: * Visualising Data by Andy Kirk:

We will also read other online posts and brief excerpts from a number of other works on the subjects of information design and data journalism, most of which are listed in the course schedule.

All software used will be freeware or open-source licensed for educational use, with the exception of Microsoft Excel and Adobe Creative Suite, both of which are available on lab computers in the event you do not own copies yourself.

V. Course Deliverables and Projects

No exams or quizzes will be administered as a part of this course. Your grade hinges entirely upon the quality of the data journalism you produce, the intellectual rigor of your critiques and reading responses, and the level of participation you demonstrate by offering meaningful class discussion.

Course grades will be determined by weighing the scores for each of the following deliverables:

- **WEEKLY DATA JOURNALISM CRITIQUES (“Monday Makeovers”) - 10% OF FINAL COURSE GRADE:** Offer one (1) thoughtful, well-articulated critique of a popular and/or novel example of data journalism of your choice each week in the form of a roughly three-hundred fifty (350) word blog post published no later than 8 a.m. the morning before each Monday class. The piece of data journalism you critique can be something you see in the news that week, or any memorable digital artifact you can find that blends design, data and interaction. Each Monday, at the start of class, two students will be chosen beforehand based upon the quality of their blog critiques and the novelty of the examples they find to present their critiques and findings briefly to the class for discussion.
- **BLOG POST RESPONSES TO CLASS READINGS - ** 20% OF FINAL COURSE GRADE:**** Reflect critically upon assigned readings for each class, and compose one (1) roughly two-hundred fifty (250) word blog post response to *one or more* of the texts prior to each class on that day’s assigned reading(s). The reading responses must be posted to your class blog no later than 10 a.m. the morning *before* each class. *A high-scoring blog post will do more than summarize or paraphrase the author’s argument.* It will, instead, seek to confront the text(s) at hand critically and with logically sound evidence. It will either (a) form an opinion in favor of or in opposition to the author’s claims, backed by concrete examples, (b) evaluate the strengths or weaknesses of the author’s logic, or (c) compare/contrast the author’s stated views with that of another scholar or practitioner’s work. **Do not attempt to craft your responses without having first read the text(s). Your attempts will not go unnoticed.** Discussing off-topic subjects, employing vague generalizations, failing to cite specific evidence from the text, or simply restating the author’s introductory argument, among other ploys, are all red flags of posts written without having read the assigned text. Such ruses meant to shirk assigned readings can be easily detected, and will be pointed out to the student – especially if occurring more than once – in private email feedback. While no formal numeric scores will be assigned to individual blog posts, weak or shallow responses will result in a commensurate deduction in the ‘reading responses’ component of your final grade (which, again, accounts for 20% of your course grade). In addition, missing more than five (5) reading responses and/or composing more than seven (7) responses that reflect a clear failure to have read the text will automatically render your *reading response* grade as zero (0), and could also lead to subsequent deductions in the *participation* component of your course grade as well (which also accounts for another 20 percent). This means, in theory, failing to read assigned texts and, thus, failing to compose sufficiently critical reading responses could, on its own, trigger an automatic D or F in the course. Feedback on reading responses will take place via occasional comments on your blog posts or, if necessary due to consistently poor performance, private email correspondence. So, take the readings seriously—even if they may seem unrelated at the time to the more practical components of the course, I assure you the readings and the lab work will tie in if you allow yourself to absorb the texts enough to recognize their practical applications. *In the highly*

unlikely event the majority of the class at any point reflects a continued pattern of failure to meet the minimum bar for critically-analyzed blog responses, the instructor reserves the right to begin instituting reading quizzes at the beginning of each class **in addition** to composing the responses already required. I neither expect nor want this to happen, as rote memorization of facts or concepts is not the goal of this course, and quizzes can cause undue anxiety that may distract students from absorbing the meaning of the readings. But, I feel compelled to add this cautionary disclaimer to the syllabus should it come to that point at any time.

- **IN-LAB, GUIDED TUTORIAL WORK - 10% OF FINAL COURSE GRADE:** Follow along with in-class/lab programming, data analysis, statistical and visualization tutorials and submit finished work from each tutorial via email no later than the start of the next class meeting. Tutorials will be paced in a manner that you should reasonably be able to complete them in-class as they are given. They will also as a failsafe be made available as screencasts uploaded to the class website for later review. But, if you still encounter difficulty completing the tutorial work before the next class, I encourage you to schedule time to meet with me outside of class for further assistance. I am more than happy to help.
- **‘DEADLINE’ DATA JOURNALISM PROJECTS - 30% OF FINAL COURSE GRADE:** Produce a total of three (3) individual ‘deadline’ data journalism/data visualization projects that might be reasonably produced in one to two (1-2) days of full-time newsroom work on a topic of current interest. Projects may take the form of a simple interactive chart, map, dashboard, data-driven game/animation or other visualization, and should include any necessary explanatory text or reporting to compliment the visualization. The projects will be evaluated on technique, design, clarity, truthfulness, the integrity of the underlying data and the story being told (*Note: Each project is not necessarily weighed equally here –greater weight will be assigned to higher scoring projects; thus, if you score poorly on your first project and then exceedingly well on your second and third, you could still in theory receive a high score overall in this category*).
- **PARTICIPATION AND DISCUSSION - 20% OF FINAL COURSE GRADE:** Participate meaningfully in class discussions on peer critiques and assigned readings and — to a lesser extent — during lab tutorials. This does not mean participation for the sake of participation. This means offering insightful, unique, critical analysis of the text and/or subject at hand that either contributes and furthers discussion or raises otherwise new insights. You will be given formal written feedback from me regarding your performance in this aspect of the course three times during the semester: once at week six (6), again at week twelve (12), then once again two (2) weeks prior to the end of term. Because data journalism is almost always a collaborative effort that requires a ‘data mindset,’ your ability to speak fluently the language of data, the web, and visual communications in a group setting is of paramount importance.
- **FINAL GROUP PROJECT - 20% OF FINAL COURSE GRADE:** By the end of the course, you should be able to contribute meaningfully to a final group project in teams of five (5) that assumes one or more of the following formats, all in a journalistic context:
 - * An exploratory or searchable data-driven news application that allows the user to personalize his or her experience with the narrative. Think [Dollars for Docs](#) by ProPublica or [Mapping America: Every City, Every Block](#) by NYT Graphics.
 - * An interactive chart or data visualization that displays more than one (1) variable either in single or multiple views, preferably responsively. Because this format alone usually requires less effort and often benefits from further context, it should include along with it *at least one* or more of the following additional components: a conventional text-based narrative, multimedia assets, or animated illustrations. You may here also wish to cite anecdotal evidence collected through original reporting, including remarks from experts who specialize in the topic of your project.
 - * A visualization that allows for dynamic data exploration among multiple hierarchies of information and responds appropriately to a range of screen sizes . Accompanying the visual(s) may or may not — depending on the project scope — be a textual. image-based and/or video narrative describing the effects and providing human voices to complement the story conveyed by the data.

This project will be presented in front of a panel of well-known data and visual journalists one (1) week prior

to its final due date (Dec. 14) so that judges may offer constructive feedback on ways you might improve your project before it is officially submitted to me via email during the week of finals no later than Dec. 17 at 10 a.m. Upon each group member's written consent, I intend to publish the final projects on a dedicated server and domain under the team's bylines, as well as to make them available as open-source repositories on GitHub, for purposes of preservation and showcasing your work.

Assessment BREAKDOWN

Reading Response Evaluation Criteria:

- *Clear Thesis* – One-third (1/3).
- *Thoughtful Thesis* – One-third (1/3)
- *Illustrative Examples/Textual Evidence* – One-third (1/3)

Data Journalism Critiques Evaluation Criteria:

- *Timeliness and/or novelty of chosen visualization* – One-third (1/3).
- *Critical analysis of strengths and/or weaknesses of visualization* – One-third (1/3)
- *Suggestions for how visualization might be improved or, alternatively, ways in which it might have distorted the true visual meaning were it not well-thought out* – One-third (1/3)

Deadline Data Journalism Projects Evaluation Criteria:

- *Effective use of data in conveying relevant information related to narrative* – One-third (1/3).
- *Use of appropriate visual display of data (i.e. picking the right chart/medium, using color appropriately,* – One-third (1/3)
- *Clarity and truthfulness of visuals* - One-third (1/3)

Final Group Project Assessment Criteria:

- *What is the story, why is it important, what data backs up conclusions? Is this publication-ready?* - One-third (1/3)
- *Ability to improve project after feedback from judges* – One-third (1/3).
- *Use of multimodal communicative media, if appropriate* - One-third (1/3)

Redo Policy

Given that data visualization is an inherently visual specialization often best critiqued by redesign, you will each get one chance to redesign or “redo” any single one (1) of the three (3) deadline data journalism projects if you are unhappy with your initial score. This effectively allows you to use the feedback given to you from me to improve the visualization design and/or journalistic strength of your original project and submit it again for re-evaluation within ten (10) days of the date the project score and feedback was returned to you. You also will receive one (1) free pass to use should you fail to submit a critique/reading before deadline. With this pass, as long as you complete the post within one week of the due date, you will not be penalized for the late submission.

Final course grades will be determined using the standard academic format rounding to the nearest integer, with a grade of C or C+ roughly representing the average grade:

Grade Percentage A 93-100 A- 90-92 B+ 87-89 B 83-86 C+ 77-79 C 73-76 C- 70-72 D+ 67-69 D 65-66 E/F Below 65

*As the course components above suggest, this course aims to blend theory/discussion equally with practice/production of actual data journalism. Producing stellar works of data journalism but failing to post reading responses and critiques – or to participate meaningfully in class discussion – will **not** be sufficient to succeed in this course. Likewise, diligent blog posts and participation without a sincere attempt to put theory into practice with actual data journalism projects will also **not** meet criteria for course success.*

VI. Attendance Policy and Other Miscellaneous Stipulations

More than three (3) unexcused absences from lectures or lab time will automatically render your participation grade (which accounts for 20% of your total grade) as zero (0). More than eight (8) unexcused absences will automatically trigger failure of the class.

Given that a large component of the class is structured around your blog post responses and critiques, I will offer feedback regularly in the form of posting comments to your blog entries. It is recommended but in no way mandated that you use [WordPress](#), [Tumblr](#) or [Ghost](#) to host your blog; whatever platform or site architecture you choose, ensure I and your classmates can access it and leave meaningful comments. If a blog post is deemed to lack sufficient evidence of having read the text, and thus not deserving of serious critical commentary, I will instead email you with feedback privately. If you receive a comment on the post from me, consider that as effectively the equivalent of an ‘A’ for that day’s response. If you receive neither a comment nor negative private feedback, that means you successfully showed evidence of having read the text that day and should not have cause for concern. I generally will attempt to leave a constructive comment on one of your posts at least once every two (2) weeks if you are successfully carrying them out.

As previously mentioned, in-lab tutorial work should be submitted via email, and all blog post responses and critiques will be submitted by posting them directly to your blog (I will begin checking for posts at 8 a.m., so altering the published time on posts won’t work if late—but nice try!). Any work submitted past the stated deadline will at most receive a fifty (50) score and may receive as low as zero (0) at the discretion of the instructor.

Because this course will focus heavily on ensuring all or mostly all produced data visualizations adapt to a variety of different screen sizes, the use of mobile phones in class will be permitted as a prototyping tool during lab times to test visualizations (having said that, please respect the time of your peers and do not use your phone to text, Snapchat or carry out any other unrelated social communication during lab times or lectures ad nauseum; *doing so overtly will result in an automatic deduction of twenty points from your participation score*).

You are welcome to bring your own laptop to class or to use the lab computers. If bringing your own machine, please either ensure it is running the Mac OS X operating system, or that you can follow along using the operating system you have without falling behind. We will be using MacOSX Sierra for purposes of in-class tutorials, along with Sublime Text (free; sublimetext.com), iTerm2 (free; iterm2.com), Cyberduck (free; Cyberduck.io), Python V 2.7 (<http://> and Tableau Public (free; <http://public.tableau.com>). You’ll also want to create a Google account in the incredibly rare event you don’t already have one, as we’ll be using [Google Sheets](#) for most spreadsheet-related data analysis. *For a more thorough list of possible tools and technologies you may wish to use in your individual and group work, I have previously put together a collection of curated tools for students of data journalism and data visualization technologies to use at <http://dataviz.tools>. Or, you may always ask me for a solution during lab time or office hours if unable to find help elsewhere.*

VII. Academic Integrity

Each student in this course is expected to abide by the UF Code of Academic Integrity. Any work submitted by a student in this course for academic credit will be the student’s own work.

You are encouraged to study together and to discuss information and concepts covered in lecture and the sections with other students. You can give “consulting” help to or receive “consulting” help from such students. However, this permissible cooperation should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an e-mail, an e-mail attachment file, a flash drive or a hard copy. Specifically in regard to lab tutorial work, it is important that you show the process of your work, as the entire class will ultimately be creating the same visualization.

Should copying occur, both the student who copied work from another student and the student who gave material to be copied will both automatically receive a zero (0) for the assignment. Penalty for violation of this Code can also be extended to include failure of the course and University disciplinary action.

VII. Diversity

The UF College of Journalism and Communications values diversity, in the broadest sense of the word – gender, age, race, ethnicity, nationality, income, religion, education, geographic, physical and mental ability or disability, sexual orientation. We recognize that understanding and incorporating diversity in the curriculum enables us to prepare our students for careers as professional communicators in a global society. As communicators, we understand that journalism, advertising and other forms of strategic communication must reflect society in order to be effective and reliable. We fail as journalists if we are not accurate in our written, spoken and visual reports; including diverse voices and perspectives improves our accuracy and truthfulness. In advertising, we cannot succeed if we do not understand the value of or know how to create advertising that reflects a diverse society and, thus, appeals to broader audiences.

Moreover, given that this course involves elements of traditionally ‘STEM’ disciplines such as Computer Science and Data Analytics, I pledge, as your instructor, to make the learning process as inclusive of students of all backgrounds as possible. Unnecessary abstractions, overly-complicated terminology, or ‘state-of-the-art’ code syntax that’s hard for a non-developer to read, among other needless complexities, all help construct artificial barriers to entry that discourage the beginner from getting started in technology and contribute further to an already insider’s culture (think Silicon Valley). This, in my belief, is one of the single largest reasons for the current lack of diversity in the tech community.

VII. Course Schedule and Format

This course is inherently interdisciplinary, combining elements of design, statistics, computer science, human-computer interaction and communication studies. But it is still, at its core, a *journalism* course in the sense that its key aim is to help students learn how to distill complex information of public or societal importance into a cogent and engaging narrative for users. In that regard, then, the course schedule or reading list may be subject to change in response to a major news event, the publication of a relevant piece of scholarship that contributes to the flow of class discussion, or the release of a new technology that expands the possibilities of individual or group projects. I also reserve the right to adjust the course schedule as necessary depending on the pace at which the class grasps key concepts.

The schedule will be embedded as a auto-updating Google Document into the course website as well as on Blackboard, and will be modified as we proceed to reflect any changes or recalibration of the curriculum pace that may occur. In the event any changes do occur, I will notify the class immediately via email.

Disclaimer on Fair Use of Copyrighted Materials:

Any assigned readings from copyrighted works not listed for purchase in Section IV of this syllabus will take the form of short excerpts of no more than one to two chapters which I will provide for you on Blackboard in PDF form at least three (3) weeks prior to the date of the class. The reproduction and distribution of these excerpts conforms as closely as possible to current “Fair Use” educational [guidelines](#) set forth by the U.S. Copyright Office. As such, it is incumbent upon you, the student, not to redistribute these materials outside of the classroom in digital or print form for any non-class related purposes during or after course completion. This restriction does not apply to non-copyrighted resources or readings directly, openly available and linked to on the web in the schedule below.

Date Reading(s) Lecture Lab M, Aug. 28, 2017 Your syllabus! No readings/blog posts due for first class. What is 'data journalism' and why should I care? Defining "The Many Words for Visualization" (available [here](#)). Visualization vs. infographic vs. illustration. Introduction to 'makeover Monday' critiques. **HELLO, WORLD!**: Data Journalism with Skittles (group activity); Data Visualization 101 (screencast on course website); Starting your course blog; Connecting via FTP (file transfer protocol) to your UF server space. Configuring our computers.

W, Aug. 30, 2017 1.) S. Slovic and P. Slovic. Numbers and Nerves: Information, emotion and meaning in a world of data. 2016. Introduction, p. 1-21. Chapter 12, p. 157-159. AND 2.) *The Data Journalism Handbook*, Ch. 1-2. Available: <http://datajournalismhandbook.org> AND Sunne, Samantha. (2016) *Diving into Data Journalism: Strategies for getting started or going deeper*. <https://www.americanpressinstitute.org/publications/reports/strategy-studies/data-journalism/> How might we apply Slovic's "balanced or multidimensional approach to 'data'" to journalism and communicative media? What would that look like, given what you've read from *The Data Journalism Handbook*? We'll look at some classic data journalism examples, discuss their pros and cons and identify what exactly makes them successful. n/a F, Sept. 1, 2017 1.) Howard, Alexander. (2015) "The Art and Science of Data-Driven Journalism." Tow Center for Digital Journalism at Columbia University. P. 1-19. <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>. AND 2.) *Boyd, Danah and Kate Crawford. (2011). "Six Provocations for Big Data."* <http://ssrn.com/abstract=1926431> *Types of data journalism; the public's need for data literacy through storytelling and visuals; how "data-driven journalism" is the future; the role of computation in journalism.* n/a M, Sept. 4, 2017 1.)** *Manovich, Lev. The Language of New Media. 2001. Chapter 5, p. 213-236. (Available for download here: http://cvlassets.s3.amazonaws.com/Manovich-Lev_The_Language_of_the_New_Media.pdf).* AND 2.) *Shazna, L. "Visual Literacy in the Age of Data."* Blog post, Source by Open News. URL: <https://source.opennews.org/en-US/learning/visual-literacy-age-data> AND 3.) *Diakopoulos, Nick. (2013). Storytelling with Data: What Are the Impacts on the Audience? Blog post. Source: http://www.nickdiakopoulos.com/2013/04/12/storytelling-with-data-what-are-the-impacts-on-the-audience/ What is, in Manovich's opinion, the relationship between database and narrative? We'll look at the non-linear structure employed in most pieces of data journalism, discuss how visual learning patterns differ from narrative learning patterns and Manovich's call for "infosthetics."* **DATA BASICS : Using Data as a Journalistic "Source"** (group activity); *The Spreadsheet as Modern Day Reporter's Notebook; basic spreadsheet use; rows, columns, ranges, formulas.* W, Sept. 6, 2017 1.) *Norman, Don. (2001) The Design of Everyday Things. Chapters 1-3.* AND ** 2.)** *Walter, A. (2001) Designing for Emotion. A Book Apart. Print. Chapter 3. (Handout)* AND 3.) *Kosara, R. (2008). "What is Visualization? A Definition."* Blog post. <https://eagereyes.org/criticism/definition-of-visualization> *The role of emotion in usability; the effects of aesthetics on cognition and memory; affordances, signifiers and feedback; emotional design, design thinking and human-centered design. Question: Why do you think Norman's book was originally titled "The Psychology of Everyday Things"?* n/a F, Sept. 8, 2017 1. *Nyhan, B. et. al. (2012). Opening the Political Mind: The roles of information deficits and identity threat in the prevalence of misperceptions. (Available for download: http://www.dartmouth.edu/~nyhan/opening-political-mind.pdf.* AND 2. *Tufte, Edward. Data Analysis for Politics and Policy. Chapter 1. (Handout) Data stories may seem to audiences as 'more credible' and less partisan. How might visual displays of data create misperceptions? How might they enhance credibility and lead to a higher rate of correct perceptions? What are the ethical responsibilities unique to data journalism?* n/a M, Sept. 11, 2017 1.) *Cairo, A.* The Functional Art. Chapters 1 and 2.* AND 2.) "Technology and History: Kranzberg's Laws." (1986). Technology and Culture, Vol. 27, No 3. P. 544-560. <https://www.jstor.org/stable/3105385> AND 3.) *Bell, Emily. 2012. "Journalism by Numbers."* Columbia Journalism Review. http://www.cjr.org/cover_story/journalism_by_nu 4.) *Segal, E. and Jeffrey Heer. " http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf*

Not every story is just going to fall into your lap. But how do you get hold of the data that no-one wants to give you? A look at making FOIA requests. Also, a brief overview of Florida's "Sunshine Laws" that require open government data be accessible (slides: <http://www.slideshare.net/visualization-in-the-newsroom>)

OBTAINING DATA: Use of open-government portals to download datasets; State of Florida and U.S. Open Data Laws; requesting data from a government agency; scraping data with programming languages such as Python; collecting your own datasets through aggregation or experimentation; basic data file types to look for; tools for scraping data from websites and PDFs ([Tabula](#), [import.io](#)). W, Sept. 13, 2017 1.) Meyer, P. (1991). *The new precision journalism*. Bloomington: Indiana University Press. Ch. 1-2. AND 2.) Cohen, Sarah. *Numbers in the Newsroom: Using Math and Statistics in the News*. Investigative Reporters and Editors, 2014. Ch. 1-3. http://cvlassets.s3.amazonaws.com/Numbers_in_the_Newsroom_Second_Edition.pdf AND 3.) *Data Journalism Handbook*, Ch. 5. Basic statistics for journalists using Excel/Sheets. Percent change; rates; standard deviation; probability; Biggest Math Mistakes for Journalists. n/a F, Sept. 15, 2017 1.) Numbers in the Newsroom, Ch. 4-9. AND 2.) Stray, Jonathan. *The Curious Journalist's Guide to Data*. * *Columbia Journalism Review*, 2016. http://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php AND 3.) Coddington, Mark. (2014). Clarifying Journalism's Quantitative Turn. *University of Texas at Austin*. <http://dx.doi.org/10.1080/21670811.2014.976400> How might you "cook" raw data? Cleaning datasets; refining data integrity; narrowing datasets to relevant variables only. "Like all journalism, data journalism requires editing." n/a M, Sept. 18, 2017 Quartz's "Guide to Bad Data." Available: <https://github.com/Quartz/bad-data-guide/> 2.) Interviewing Data for News Stories: <http://dwillis.github.io/interviewing-data/> Interviewing your data —(a) Meeting the Data: backgrounding the source/getting to know the data (b) Asking Questions — Finding the flaws, counting sums, interestingness and (c) Drawing Conclusions. **ANALYZING YOUR DATA:** "Interviewing" Your Data With Google Sheets (exercise); Causation and correlation; statistical significance, p value, z value and standard deviation. Normal vs. skewed distribution. Role of outliers. Pivot tables. Tutorial: Clustering Campaign Finance Data with OpenRefine. W, Sept. 20, 2017 1.)**_ Lupton, Ellen.* Thinking With Type, 2010. *p. 97-101. AND **_ 2.) McGhee, J. (2010) Journalism in the Age of Data. *Documentary*, Stanford University. Available: <http://datajournalism.stanford.edu/> AND 3.) Norman, Don. *The Design of Everyday Things*. Ch. 4-5. n/a F, Sept. 22 1.) Chiasson, Trina and Dyanna Gregory. (2015) *Data+Design: A simple introduction to preparing and visualizing information*. Infoactive. Available for free download: <http://infoactive.co/data-design> AND 2.) Wing, Jeanette M. (2006) "Computational thinking." *Communications of the ACM*. 49 (3). <https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf> TK n/a LABOR DAY - NO CLASS n/a n/a n/a W, Sept. 26, 2017 1.) Tufte, E. (1990). *The Visual Display of Quantitative Information*. AND 2.) Illinsky and Steele, Designing Data Visualizations, chapter 4 3.) Norman, Don. (2005) *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books. Print. Ch. 1. Charts, charts and more charts; choosing the right chart type for your data; need to go beyond static Excel charts for big datasets; avoiding "chartjunk." F, Sept. 28, 2017 1.)** Wong, D. M. (2013). *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*. United States: WW Norton & Co. Ch. 2-5. AND 2.) Cairo, Alberto. *The Truthful Art*. Ch. 1. How to lie with charts; baselines, axes, why pie charts are almost never a good idea; importance of providing source. n/a M, Oct. 1, 2017 1.) Mollerup, Per. *Data Design: Visualising quantities, locations, connections*. Bloomsbury, 2015. Print. Ch. 2-4. AND 2.) "Data Visualization Checklist" by Evergreen and Emery: http://stephanieevergreen.com/wp-content/uploads/2014/05/DataVizChecklist_May2014.pdf Primer to HTML and CSS; the iFrame as a "window" into another webpage; fluid grid system basics; **VISUALIZE YOUR DATA:** Why make basic charts? There are large numbers of tools available to make simple embedable charts that help you tell your story. This session will show how to use a selection of open-source tools such

as Google Charts, [HighCharts](#), [DataWrapper](#), [Chart Tool](#) and Quartz's ChartBuilder. Also, a walkthrough of eight chart types from [Plotly](#). W, Oct. 3, 2017 The Functional Art: Ch. 3-7. How Data Journalism is Different (American Press Institute): <https://www.americanpressinstitute.org/publications/reports/strategy-studies/how-data-journalism-is-different/> n/a F, Oct. 5, 2017 DUE: First 'Deadline' Data Journalism Project by 10 a.m. n/a n/a COLUMBUS DAY - NO CLASS n/a n/a n/a W, Oct. 10, 2017 The Data Journalism Handbook, Ch. 4-8 n/a F, Oct. 12, 2017 "What Makes a Visualization Memorable": <http://www.storybench.org/understanding-what-makes-a-visualization-memorable/> n/a M, Oct. 15, 2017 Mollerup, Per.* Data Design: Visualising quantities, locations, connections. *Bloomsbury, 2015. Print. Ch. 1.* AND Wolfe, Jeremy. "Visual Search." http://www.scholarpedia.org/article/Visual_search AND Cooper, Alan and DESIGNING FOR THE EYE AND THE MIND: Color, hierarchy, typography, white space, grid systems, W, Oct. 17, 2017 1.) Ayres, Paul and John Sweller (2010) "The Split-Attention Principle in Multimedia Learning." AND. "The Architecture of a Data Visualization": <https://medium.com/accurat-studio/the-architecture-of-a-data-visualization-470b807799b4#.4269conf4> AND 3.) Boardman, Richard. "Bubble trees: The visualization of hierarchical information structures." <http://cvlassets.s3.amazonaws.com/Boardman%20Unknown%20-%20Bubble%20trees%20The%20visualization%20of%20hierarchical%20information%20structures-annotated.pdf> F, Oct. 19, 2017 Tufte, E. (1990) *The Visual Display of Quantitative Data.* M, Oct. 15, 2017 1. Hinderman, Bill. Responsive Data Visualization for the Web.* AND 2.) Lewis C. (2012) "Making the case for hover interactions in maps." Blog post. <http://carlvlewis.net/making-the-case-for-hover-interactions-in-maps/> AND 3.) Ericson, Matthew. "When Maps Shouldn't Be Maps." <http://www.ericson.net/content/2011/10/when-maps-shouldnt-be-maps/> MAPPING MADE EASY: Using CARTO to create interactive clustered bubble map of incidents of mass shootings from [massshooting-tracker.com](#). Building a choropleth map with Leaflet.js; cartograms; recognizing the various forms geospatial data takes. W, Oct. 17, 2017 F, Oct. 19, 2017 M, Oct. 20, 2017 Bubbletree. MAKING DATA INTERACTIVE: JavaScript Magic Show (d3.js, d3plus, DataWrapper, Excel2D3). Value of interaction in personalizing and bringing deeper engagement to data stories. W, Oct. 22, 2017 "Responsive Web Design": <http://alistapart.com/article/responsive-web-design> People hear statistics, but they feel stories M, Oct, 25, 2017 No assigned reading today but a blog post still expected. In lieu of a reading response, compose a blog post on the simultaneous rise of data visualization literacy (graphicacy) and growing mobile information consumption. You may cite sources on the divergence of these two trends, if you can find them. If not, form your own draft hypothesis MAKING IT RESPONSIVE: WTF is an SVG? W, Oct. 27, 2017 Wu, Ashley. "Why Mobile Data Visualization Shouldn't Hurt." Blog post. <https://source.opennews.org/en-US/articles/mobile-data-visualization-shouldnt-hurt/> AND MobileV.is: <http://mobilev.is/> F, Oct. 29, 2017 M, Nov. 1, 2017 1. Roam, D. *The Back of the Napkin* FROM DATA TO BROWSER: Wireframing and Paper Prototyping; responsive grid systems ([Skeleton](#), [Bootstrap](#)), How to Choose the Right Chart; Picking Colors; CSS Deep Dive. W, Nov. 3, 2017 DUE: Second 'Deadline' Data Journalism Project F, Nov. 5, 2017 M, Nov. 7, 2017 FROM BROWSER TO WEB APP: The basics of HTML and CSS and how they interact to create the visual look of a webpage; iFraming in our visualization projects into a Skeleton template; a gentle introduction to JavaScript for interaction. W, Nov. 9, 2017 F, Nov. 11, 2017 M, Nov. 12, 2017 FROM VISUALIZATION TO DASHBOARD: Using [Tableau](#) to create a Florida state budget dashboard; using [DataSeed](#) to do similar task, then reproducing the same visualization in [Quadri-gram](#). Brief look at more advanced D3.js-based libraries (collection of goodies) including Crossfilter, Crosslet and C3.js. W, Nov. 14, 2017 F, Nov. 16, 2017 DUE: Third 'Deadline' Data Journalism Project M, Nov, 18, 2017 DISCOVERING THE OPEN-SOURCE LANDSCAPE: Basics of GitHub; creating repositories for your first and second class projects; finding tools to help tell final projects on GitHub. Specifically, we'll use [Cloudstitch](#) to pull data from a Google Sheet to create a zoomable treemap of spending data; then we'll create a force network visualization of our friend circles using [Onodo](#). NO CLASS - THANKSGIVING BREAK Bored over break? Optional reading list here: <!--link to come--> n/a n/a NO CLASS - THANKSGIVING BREAK Bored over break? Optional reading list here: <!--link to come--> n/a n/a M, Nov. 29, 2017 PACKAGING YOUR DATA JOURNALISM INVESTIGATION: Deep dive into Bootstrap's responsive CSS grid system; adaptive vs. responsive design; combining text, multimedia and interactive JS visualizations into multimodal narrative; W, Nov. 31, 2017 F, Dec. 2, 2017 DUE: Presenting group projects This

class lecture time will be reserved for each group to deliver a 5-minute presentation of its final project, followed by feedback from judges. M, Dec. 12, 2017 http://www.r2d3.us/visual-intro-to-machine-learning-part-1/?utm_content=buffera0264&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer Machine Learning/Predictive Analytics 'EDITING' DATA JOURNALISM: W, Dec. 14, 2017 F, Dec. 9, 2017 DUE: Final revised group projects

Pro-tip on readings: Most classes will include two (2) or more readings tangentially related; typically, at least one of these readings will be more theoretical in nature, while at least one more will adopt a more practical or technical tone. The secret to writing successful reading responses and excelling in class discussion is for you to draw connections between the two. * * *

ABOUT THE INSTRUCTOR: (*Endnote: Add this syllabus to your phone's home screen at http://bit.ly/uf_bigdata*)

-30-(?)