CODING AND COLLABORATION:
DATA ANALYTICS IN THE LAW SCHOOL
CLASSROOM

Charlotte S. Alexander* & Nicole G. Iannarone†

N. Iannarone:

Thank you for joining us today. I am Nicole Iannarone, and I am excited to present with Professor Charlotte Alexander. I am with Drexel University Kline School of Law; Professor Alexander is from the Robinson College of Business at Georgia State University. In Spring 2019, we worked together in a co-taught Legal Analytics class,¹ a course that arose from the work Charlotte began when she founded the Legal Analytics Lab at Georgia State University.² We will be discussing how we taught data analytics in the law school classroom through cross-disciplinary collaboration between students pursuing law and master’s in data science and analytics degrees.

C. Alexander:

Several years ago, at Georgia State, I founded and continue to run our Legal Analytics Lab. I am a lawyer by training. The idea is that those of us who have legal expertise can pair with colleagues who have expertise

---

* Connie D. and Ken McDaniels Women Lead Associate Professor of Law and Analytics, J. Mack Robinson College of Business and the College of Law at Georgia State University, Founder and Director, Legal Analytics Lab.
† Assistant Professor of Law, Drexel University Thomas R. Kline School of Law.

We are grateful to Dr. Mohammadjavad Feizollahi, Khalifeh al Jadda, and Professor Anne Tucker for their expertise and collaboration in teaching the course, as well as the law and analytics students for their curiosity and diligence in learning to use text analytics to study legal documents. Thank you to the organizers and attendees at the 2021 Emory Conference on the Teaching of Transactional Law & Skills for the opportunity to present this work and for their feedback.

¹ The course description for this iteration of Legal Analytics is as follows: “This course introduces students to the emerging field of legal analytics, which employs computational and statistical modeling, analysis, and visualization of legal data to accomplish both descriptive and predictive goals. For analytics students, the course introduces the U.S. legal system and legal reasoning, legal materials, and the problems and questions present in the law. For law students, the course offers an introduction to basic coding, as well as to the theory and applications of text mining, natural language processing, machine learning and other methods for managing and analyzing unstructured data such as that found in legal documents.”
² See LEGAL ANALYTICS LAB, https://robinson.gsu.edu/academic-departments/insight/innovation-labs/legal-analytics-lab/ (describing Legal Analytics Lab and projects).
in data science, computer science, or software engineering. If we start to
view legal questions as data questions, at least in part, then we can extract
really interesting insight from masses of legal text that previously would
require many hours of associate or law student power to analyze. The lab
is a larger effort that we have underway at Georgia State. The class that we
are going to describe, Legal Analytics, is the classroom portion of what we
are doing at the intersection of data science and law.

N. Iannarone:

Collaboration is at the heart of the intersection of data science and
law in the Legal Analytics course. Cross-disciplinary collaboration is a key
characteristic of legal practice. It is also a key characteristic of data
analytics. That is why we brought together students at law and business
schools in the course. As lawyers, we recognize that competent
representation requires engaging with technology. For example, the Model
Rules of Professional Conduct note that understanding technology is an
essential element of the ethical duty of competence. Though not all
jurisdictions have explicitly adopted the Model Rules’ requirement of
technological competence, it is impossible to practice law without
technology, making technological knowledge at the very least an implicit
element of the duty of competence.

In addition, lawyers have to collaborate. We work with other
lawyers, clients, and experts to deliver legal services. So do data scientists.
Data scientists training to become experts will work with clients and need
to learn their role in their client’s industry. Both sets of experts, lawyers,
and data scientists must translate their expertise to their respective
audiences, who may be novices in those particular areas. With those ideas
in mind, we developed a course framework for legal analytics so that both
law and data scientists could gain information from each other and learn
together.

C. Alexander:

3 MODEL RULES OF PRO. CONDUCT r. 1.1 cmt. 8 (AM. BAR ASS’N. 2020) (stating that
attorneys must maintain technical competence).
4 See Nicole Iannarone, What Every Attorney Should Know About Technology in Practice, 26
PIABA B.J. 59, 61–63 (2019) (describing ethical duty of technological competence and
duty owed in jurisdictions that have not adopted ABA Model Rule 1.1 cmt. 8).
5 See, e.g., MODEL RULES OF PRO. CONDUCT r. 1.4(b) (“A lawyer shall explain a matter to
the extent reasonably necessary to permit the client to make informed decisions regarding
the representation.”).
The enrollment in the course that we are describing today comes from law students and from the Masters in Data Science and Analytics program, which is housed in the business school. The course has been offered for about three years, using different projects each time it is taught. Nicole and I will discuss the first iteration of the course, which had a more transactional focus. However, subsequent iterations of the course used different projects, exposing students to other areas of focus. For example, later runs of the class have analyzed speeches from the Congressional Record (in collaboration with a political science colleague) and financial disclosures made by franchisors to potential franchisees (in collaboration with a professor from the business school’s hospitality department). The overall idea is to bring together students from different programmatic and disciplinary backgrounds. We expose them to real-world problems at the intersection of data science and law. We place them on interdisciplinary teams. In order for the course to function well, there is a substantial investment of faculty time at the front end to guide the students’ progression through the course.

N. Iannarone:

When designing the first iteration of Legal Analytics, we had to start by choosing a set of legal documents - a corpus of information — for students to analyze. Finding the right corpus is somewhat of a Goldilocks problem: the corpus should neither be too big nor too small — but somehow just right. Charlotte and I had begun collaborating on our own legal analytics research, analyzing securities arbitration awards in the Financial Industry Regulatory Authority (FINRA) dispute resolution forum,\(^6\) where each of us brought different areas of expertise to the project. As we began that project, we realized the corpus of arbitration awards we were studying were consistently presented, with each award containing the same segments and information.\(^7\) Because the awards were so consistent, and the data set sufficiently contained, we agreed that the FINRA awards would be ideal for the data analysis aspects of the course. At the same time, looking at resolution of securities disputes would also allow us to introduce a unique aspect of the U.S. legal system with our data analytics students without overwhelming them.

\(^6\) [FINRA Dispute Resolution Services](https://www.finra.org/arbitration-mediation) (“FINRA operates the largest securities dispute resolution forum in the United States”).

\(^7\) See, e.g., [FINRA Rule 12904(e)](https://www.finra.org/sites/default/files/documents/rulebooks/2019/Arbitration-and-Securities-Litigation-Rulebook-2019.pdf) (detailing the information required to be reported in each award issued in the FINRA forum).
While we needed a consistent set of documents to constitute the corpus, those documents could not be so consistent that they were identical in all ways. The corpus needed to present some messy situations to provide students with data problems they would need to resolve. Accordingly, while we had award documents with consistent segments, we chose this corpus because there were some differences in how information was reported in each award segment. The awards were sufficiently consistent so that they would allow law students to quickly understand the substantive legal area of the law and focus the bulk of their time on learning how to code, use data analytics tools, resolve data issues, and apply statistical methods. On the flip side, the data set would permit the data analytics students to learn about the legal system in a manageable way over one semester while being challenged with the analytics problems presented by the data. Using a data set with which Charlotte and I were already familiar permitted us to focus on the students’ skill and substantive knowledge development, because we each had very strong understanding of how our respective fields of expertise – securities arbitration and legal analytics – intersected with the corpus of FINRA award documents.

C. Alexander:

The course logistics have remained consistent over three iterations of the course. The students are broken up into teams. The teams themselves are composed of some law students and some analytics students. Over the arc of the semester, students are introduced to the substantive area of law and the data set, and they then identify a particular research project that they are going to tackle, given the raw materials that we provide to them. In this case, the corpus or the body of texts that we worked with were the FINRA arbitration awards. After introducing them to arbitration, the FINRA forum, and data analytics basics, we begin the semester with the teams identifying a research question that they want to answer. In each subsequent week, we provide them with instruction in a different analytics method to tackle their research questions, using text as our data source.

---


9 See id. (Illustrating results of collaboration between data analytics expert and securities arbitration expert in analyzing FINRA awards data).
After understanding the substance of the legal area, the first conceptual set of methods that the students have to understand is how to think of text as data. Consider a student who is interested in the outcomes of these FINRA arbitration proceedings, for example. How do you go turn hundreds of pages of texts where the outcomes are buried into a nice, clean Excel spreadsheet that identifies a single outcome for each arbitration? We begin the semester with a conceptual introduction to this process of thinking of text as data.

We progressed through the semester, each week providing the entire class with specific instruction with modeled code. Students have their hands on keyboards and are learning how to run code. They are troubleshooting and reinforcing their learning in teams. They are taking the code or methods they have learned each week and are going back to their team and applying those methodologies to their particular research question that they are pursuing over the arc of the semester.

The semester culminates with each team giving a presentation and submitting a written report, discussing the answer to their research question that they posed at the beginning of the semester. If any of you have been involved with projects like this, you know that it is very rare that you have the same research question at the end as you did at the beginning. There is a lot of iteration in refining the work and refining the students’ focus along the way, all while fitting their interests to the methods that they have available.

While the teams continued to work on their projects, outside presenters participated in class sessions. We had data scientists including innovation and technology professionals from Alston & Bird\(^{10}\) and a guest from Ernst & Young’s\(^{11}\) legal department who use analytics to advise clients on legal questions and problems. Experts from the FINRA forum, including the director of FINRA Dispute Resolution’s Southeastern Regional Office, joined class to share expertise on the securities arbitration forum and answer questions about the award documents students were analyzing. These guests met another goal of this course: to expose students to potential career paths and applying the skills and methodologies that they were learning in class.

Towards the end of the semester, some class time was set aside so teams could obtain professor feedback on the status of their projects, or


if they had questions on either the data or legal sides of their work. Prior
to breaking into their teams for focused collaboration, we led focused
discussions with the entire class to surface issues and questions common
to multiple teams so the class could strategize about how to resolve those
issues. Students then broke into their teams for dedicated collaboration
time. During team breakouts, Nicole and I rotated from team to team to
calculate that they were on track with their research questions, answer questions,
and to provide them with feedback on the progression of their projects.

These elements – weekly methods instruction, team application,
and outside presenters – all worked together towards the teams’ final
research reports and presentations to the class. To give you a sense of the
kind of work the students were able to accomplish, we will show you some
excerpts from one of our student teams and give you a flavor of the high-
quality products they were able to generate.

This is an excerpt from one of our student teams:

**Original Research Question:**

Does the aggregate number of arbitrations a public
arbitrator has heard (our independent variable) impact the
amount of compensatory damages awarded to a claimant
in proportion to the compensatory damages requested
(our dependent variable)? If so, are we able to develop a
model that will generate an accurate prediction of the
proportion of the amount requested that will be awarded
based on the number of previous cases a public arbitrator
has heard?

**Hypothesis:**

When a public arbitrator first begins hearing cases the
arbitrator will sympathize with the customer claimant
more and thus make an award that is closer to the amount
requested. We hypothesize, however, that as the number
of cases an arbitrator has heard increases, the arbitrator
will award increasingly lower amounts in comparison to
the amount requested.
This team’s original research question was regarding the number of arbitrations that a public arbitrator had heard. Specifically, what is the association between the number of arbitrations and the amount of compensatory damages awarded to the claimant? The team had an interesting hypothesis about arbitrator behavior. They were curious as to the extent which arbitrators would sympathize with claimants. However, they hypothesized that as the number of cases an arbitrator heard increases, the arbitrator will award a lower amount of compensatory damages as he or she becomes more experienced.

If you unpack this question and hypothesis, think about the corpus with which the students were working. We handed them thousands of arbitration awards and they had to turn that raw text into data that they could then use to answer these questions. The team was required to revise their research question as they delved into the data to try and get a better handle on how to quantify claimant success in the forum. As a result, they simplified their question to “Does an arbitrator’s experience impact a claimant’s outcome?” Text analysis with real-world text is messy and hard, which is the meta lesson we wanted the students to encounter through this real-world iterative process.

The students were able to generate an interactive dashboard in which you could plug in data to see the results of their findings. The data could be parsed by specific regions or viewed as nationwide results. Students could focus on panel composition. They were able ultimately to compute a percent of relief awarded after identifying the amount awarded and the amount requested. They were also able to break down their results by arbitrator. They thought about how they could present this information in a way that would be useful to practitioners who appear in the forum and choose arbitrators through a striking and ranking process.12

N. Iannarone:

As we looked at the students’ findings, we helped guide them towards how the results they discovered would be useful to their clients and other audiences who would be interested in their analysis. So, for example, because we were in Atlanta at the time that this class was run, one group of students focused on the Atlanta FINRA hearing location. They attempted to answer how claimants fare in security arbitration cases

12 See FINRA RULE 12402(d) (describing striking and ranking process for one arbitrator cases); FINRA RULE 12403(c) (describing striking and ranking process for three arbitrator cases).
in the Atlanta hearing location. In particular, they addressed the question of what percentage claimants recover of their initial damages request, comparing the Atlanta hearing location to the rest of the nation. As shown in the below chart, they discovered that claimants recovered 43.71% of the compensatory damages overall, but when the results were analyzed by hearing location, claimants in Atlanta did not fare as well, recovering only 39.4% of the amount requested.

Figure 1: Percentage of Compensatory Relief Awarded Claimants Nationwide and in Atlanta Hearing Location

This team then dug deeper, identifying the most frequently appearing arbitrators, both nationwide and in the Atlanta hearing location, to ascertain how claimants fare before the most popular arbitrators. This deeper dive provided insightful information: in securities arbitration cases involving the most frequently appearing Atlanta arbitrator, claimants recover 49.14% of the compensatory damages they requested. However, when they analyzed the results of the most frequently appearing arbitrator nationwide, the students discovered that claimants recovered a slightly lower percentage of the compensatory damages they sought, 41.55%.
These results illustrate that having command of what the data show and the information they can reveal can guide lawyers’ decisions in practice.

The student teams’ results helped us focus the class on real-world application of their knowledge. In particular, it provided the students experience thinking about how they would present their results to leaders in the industry who could potentially be their future clients. When we think about real-world application, we think about regulators, lawyers practicing in the forum, and arbitrators, those who examine the industry and are interested in what is actually happening in the forum. What type of claims are customers bringing against their stockbrokers? How are those claims being resolved? By looking at the data, are we seeing anything strange happening in various hearing locations?

Accordingly, we invited relevant outside audiences to attend the students’ final presentations. The number of outside guests was astonishing, as nearly everyone we invited was extremely interested in seeing the results of the students’ work. Thirty invited guests attended and asked questions of each presenting group. Guests included professors from the law and business schools. The Investor Advocate, Ombudsman, and Chief Economist from the U.S. Securities & Exchange Commission’s Office of the Investor Advocate participated either in person or via video conference. From FINRA, the Executive Vice President and Director of Dispute Resolution Services, and directors of the Southeastern and Midwest Regional Dispute Resolution Services Offices participated. In addition, several Atlanta lawyers at major firms who represented both claimants and respondents in the FINRA Dispute Resolution Services forum, practiced in securities law, or were interested in practical applications of data analytics in law attended.\textsuperscript{13} Finally, several experts whose work focuses on high-level innovation and technology in their particular practices and firms also joined the final presentations.\textsuperscript{14} Presenting before such a knowledgeable group gave students an opportunity to hear what their clients and partners are going to want moving forward. We were also able to share with students the extremely

\textsuperscript{13} We are grateful to the following lawyers who attended the students' presentations and provided feedback to students: Leo Kogan, Eversheds Sutherland; Scott Wandstradt, Arnall Golden Gregory; Brian Mink, Troutman Pepper; Tod Sawicki, Alston & Bird; G. Wayne Hillis, Parker Hudson Rainier & Dobbs; Frank Lichtmas, Lichtmas Law; and Mike Bishop, Smiley Bishop & Porter.

\textsuperscript{14} We are also grateful to the experts who use analytics in their daily work for their support of the students' learning: Chris Haley, Troutman Sanders; Ed Sohn, Thomson Reuters; Taimur Ghaznavi, Ernst & Young; and Dawn Edwards Martin, Counsel on Call.
positive feedback we received from professors, regulators, experts, and practitioners on the students’ work.

We started today by defining the students’ work in the first iteration of the class as transaction. You might be thinking, “What do securities arbitration awards have to do with transactional law?” The student teams’ results relate directly to policy decisions and regulator questions. At the onset, we mentioned that Charlotte and I are studying FINRA award data set in our own collaborative research. That is because we believe the findings that can be unlocked through analytics tools can influence positive change in the near mandatory FINRA arbitration regime. There are numerous questions about arbitration that can be answered by analyzing FINRA awards, and we hope that those answers will help guide policymaking in the arena. Thus, in addition to exposing students to how they can use data analytics tools in their own representation of clients, we explored with students how these tools can assist in improving the overall legal system.

Charlotte has been very innovative in bringing data analytics into the law school curriculum. Her subsequent iterations of the course using the framework we described have been really interesting. I would like her to share about how she has run the class with different types of data sets.

C. Alexander:

Nicole and I collaborated on the first offering of the Legal Analytics class, which we described today. In subsequent iterations of the class, I have worked with other faculty members and colleagues who have

---

15 See, generally, Alexander & Iannarone, supra note 8, at 1741–54 (identifying data reporting barriers to measuring percentage of damages claimants recover); id. at 1754–58 (recommending award reporting enhancements to more readily study consumer experiences in FINRA forum); Nicole G. Iannarone, Structural Barriers to Inclusion in Securities Arbitrator Pool, 46 WASH. L. REV. 1389 (2021) (analyzing FINRA arbitration awards to determine if new entrants who diversified FINRA’s arbitrator roster are being selected to oversee smaller claims proceedings, identifying potential barriers to new entrants’ inclusion in the FINRA forum, and suggesting transparency and policy interventions to improve inclusion in the forum). The authors are currently collectively researching the impact that the gender of parties, party representatives, and arbitrators has on arbitration outcomes in the FINRA arbitration forum. Professor Iannarone is currently researching the Covid-19 shift to virtual hearings in the FINRA forum and how claimants fared in the forum in virtual versus live hearings from March 2020 through August 2021.
different research interests.\textsuperscript{16} As a result, we have used different sets of documents as raw materials. We have run a version of this class looking at franchise disclosure documents ("FDD").\textsuperscript{17} These are the required disclosures of financial and other material information that franchise companies have to provide to potential franchisees. We were able to answer several interesting research questions about companies, what they disclose, and how they choose to disclose it with those FDDs.

We also ran a version of the class looking at Congressional Record texts.\textsuperscript{18} This was accomplished with a political science colleague of mine, and the students were quite interested in how Congress discussed COVID relief. One of our teams focused specifically on COVID relief. The Congressional Record class was Spring 2021, and it was conducted entirely virtually.

To finish, I will take a step back and reflect on the benefits and drawbacks of this kind of class. There can be true magic when it all works. We have students working across disciplines, really stretching themselves to learn new skill sets. They are also working on their soft skills, communication, teamwork, problem solving, critical thinking, and presentation. Generally, that has been the feedback we have received from the students. It has been hard and a different kind of work for them, but the students find it to be quite valuable.

Students may not be able to develop a clear answer to their research question, given their resources and time in a single semester. We hope that what we are doing is teaching them the potential of using data analytics tools to answer legal questions, or to get a start in answering legal questions. A downside is that this class requires a substantial investment of time and work from professors. Also, as we all know, teams can break down. There is a lot of soft skill work on the part of professors in shepherding the students through the semester, though we believe that the time investment is worth it in terms of the students’ learning outcomes.

\textsuperscript{16} Professor Susan Smelker co-taught the class focused on Congressional Record text in Spring 2021 and Professors Smelker and Ben Lawrence co-taught the class focused on FDDs in Spring 2020.
\textsuperscript{18} Examples of Congressional Record text may be found at https://www.congress.gov/congressional-record.
We will be happy to talk more in our Q&A time. We would love to hear your thoughts on what we have presented.

C. Alexander: It looks like there is a question in the chat: “Could you please say more on the time-consuming nature of using data analytics for folks not trained in the field?” Yes. If you have a research question that you are trying to answer using data, the perfect data set is something that comes to you in a spreadsheet with all of the rows and columns filled in, without a question about what each row and column represents. Then you are off to the races because you have nice, clean data as your input. Then you may have to make choices about what model you are going to use or what associations you are going to explore. In our example, we started with thousands of pages of PDFs. The body of documents that Nicole described were a set of FINRA arbitration awards.

We had to go from text in a PDF form to a nice, clean spreadsheet that represented different aspects of what the text said numerically. In the example we gave, the team was interested in a number of arbitrations that an arbitrator had heard and the association with between that and compensatory damages received. We had to walk the teams through writing the computer code that would dive into those documents, pull out those key pieces of information, turn it all into zeros, ones, or other numbers, put it into a spreadsheet, and then make the choices about what model to use to find and explore that association. It is messy, it is a real-world problem, and all of it starts with text. I think that is the key innovation here: we can analyze text-like data if we know how to think about it, and we can teach students how to think about it.
N. Iannarone: Building on what Charlotte said: the other messy aspects that came with the students research projects is once you convert the text to data, there are still numerous questions that need to be explored. We learned very early on that one of the questions was, “how do we even define if an outcome is a success?” It became such a sticky question that Charlotte and I wrote an entire law review article on what it means to win, lose, or to obtain favorable results in a FINRA proceeding.19 Even once we had a workable data set for the students, there were additional translation questions that the law and data students resolved together so they could then define what their research questions would be able to do, with certain limitations.

That aspect of the course presented substantial work for the instructors. The weekly meetings with the teams, which Charlotte and I rotated through to assist with analytics and substantive forum questions, respectively, were demanding. Teams had excellent questions outside of class as well. We had to think about when questions arose, how much we intervened and how much we turned the questions back to students and said, “Well, how might we interpret this? What would it mean for your research if you handled the question in this way as opposed to another way?” There was a great deal of interaction from beginning to end as students cleaned the data, developed their data set, and then answered their hypotheses. Despite the amount of work that the students and we put into the course, we feel that the course was extremely successful.

Thank you for your questions, and we appreciate the opportunity to discuss the course with you today.

19 See generally Alexander & Iannarone, supra note 8.
Speaker 1: Thank you. It was very interesting to hear about your incredibly complex and innovative course. Thank you for sharing that with us.