The COVID-19 pandemic has stressed judiciaries around the world. These class projects explore potential collaborations with country counterparts to facilitate e-justice amid covid. Broad applications of NLP or causal inference can intersect fruitfully with recent developments in optimization, statistical learning, and machine learning through the DE JURE program whose aim is to revolutionize how legitimacy and equality in justice systems are measured, understood, and enhanced–with a sightline to access and equity–to meet sustainable development goals.

The projects outlined below are inspired from the class syllabus and ongoing counterpart interests, but discussions can deviate from the proposed topics and projects can be pursued in other countries where data infrastructure permits. To accomodate the breadth of student interests, the list is organized between (1) e-justice and development, (2) disparities, and (3) grammar of law. Some datasets are described at the end.

1 E-Justice and Development[1]

1.1 Peru

- Personalized case-based teaching: In this project we focus on how to improve judicial training via technological innovations. The Judicial Academy of Peru is in the transition from theory to case-based teaching. This transition from theory to case-based teaching is a moment that characterizes legal education throughout the past century in many countries in the world, not only in the legal practice, but also in other disciplines such as medicine. In this context, DE JURE proposes leveraging Artificial Intelligence (AI) to bring this transition a step forward. The goal is to use AI to develop a method of personalized case-based teaching using the history of a judge's past decisions. In specific, in a promotion program for judges, this tool will leverage the history of the judge's own written decisions to evaluate how such judge would decide on a similar case in comparison to a curricular example or to other similar decisions by peer judges. The example of personalized teaching based on the history of decisions is developed in other settings. For instance, the decision data can reveal interesting and varying behavior among portfolio managers. Some may be more loss-averse than others, holding on to underperforming investments longer than they should. Others may be overconfident, possibly taking on too large a position in each investment. The analysis identifies these behaviors and—like a coach—provides personalized feedback that highlights behavioral changes over time, suggesting how to improve decisions. This opportunity to use AI to improve the training and performance of judges is of interest to many countries.

- Social-emotional learning: Despite the importance of high-stake judicial decisions on litigants' well-being and economic development, there is little evidence on how to improve judicial performance. In

[1]Contact: Manuel Maqueda (mramosmaqueda@worldbank.org) and Daniel Chen (dchen9@worldbank.org)

this project, we use a randomized controlled trial (RCT) to evaluate the social-emotional exercises on the performance of judges and prosecutors in Peru's Judicial Academy. We test the impact of our interventions on both soft and hard skills that aim to improve not only educational attainment and social-emotional skills, but also contribute to fairer and more efficient judicial decision-making. These writing exercises (advice-giving) involved self-distancing or self-reflection, which theories of psychological motivation suggest can increase wisdom, intrinsic motivation, empathy, and a sense of social identification with the occupation. This advice was randomly assigned for others to read and grade. The Machine Learninng (ML) opportunity is to identify what advice improves decision-making and which social-emotional learning exercises impact decisions and disparities.

- Conciliation: Conciliation and mediation are a large part of the legal system in many developing countries. ALEGRA Free Legal Aid (Asistencia Legal Gratuita) is interested to improve the performance of conciliators. DE JURE has developed an open-source app to show conciliators' performance, explore the data for patterns in different ways, download reports, and also request advice, suggest advice, and log records. The Peru team will design three sets of randomized control trials: who should ask and receive advice, who should send advice, and impact of self-tracking and showing features that predict performance or improvement.

1.2 Chile

- Personalized nudges: Previous studies have shown that behavioral nudges can be a cost-effective tool to influence changes in people's actions. Civil servants, like judges, in many settings have fixed salaries and life tenure. Motivation and absenteeism can be problems. This project tests whether nudging legal actors through informing them on their performance in absolute and relative terms can improve productivity. Moreover, we test if there is any difference if the information about performance is given in contrast and relation to the self's past performance or if the information is relative to others' performance. The ML problem is to estimate what is the ideal sequence of personalized nudges showing worst performing statistics and most improved statistics to motivate state actors in low resource settings.

- WhatsApp access to courts: The COVID-19 pandemic has stressed the Chilean Judiciary's (PJUD) need to develop new means of connecting with citizens. DE JURE's technical counterparts have developed the Conecta digital platform, which allows citizens to access PJUD's services through a video call, chat or WhatsApp. Conecta's goal is to facilitate access to PJUD's services to the general population, as well as to generate greater confidence in the justice system among citizens. At the same time, the platform also aims to optimize resources that can reduce the workload of PJUD officials. Human call receivers are given text to copy and paste, which raises the potential for chatbots to support humans especially at night and on weekends. This AI opportunity is to

deploy a chatbot and measures its impacts on trust and access to justice. (Note: the relationship on this project is more rudimentary, but the conceptual framework can apply elsewhere.)

1.3 Kenya

- Actionable Information and Accountability: A major defect of judiciaries in developing countries is their slowness. In Kenya, the average civil case takes over 4 years to resolve. In India, there are 33 million cases pending, and some cases have taken 40 years to conclude. A national RCT displays recommendations for how judges can improve a deficient statistic. As part of the first digitized daily court records, DE JURE grafted a nation-wide experiment, developing an algorithm that identifies the major greatest sources of court delays. One treatment arm provided actionable information–these sources of court delays. The second treatment arm provided information and accountability to committees who could potentially deliberate about this information. Information and accountability increased court efficiency. The Kenya team will design ideal algorithm to RCT when it becomes monthly after COVID-19.

- Mediation Assignment: Simpler alternative dispute resolution mechanisms may enhance speed of justice. The Kenyan judiciary has launched a program allowing the judiciary to assign cases to mediators for settlement out of court. They want to assign mediators to cases based on peformance–settlement rates, case duration, attention, and (non) reopening of disputes. The Kenya team will design two RCTs: a multi-arm bandit framework with capacity constraints for optimal specialization and a second one that accounts for learning.

- Virtual Mediation: Virtual courtrooms promise to facilitate justice amid social distancing needs. A recent study of virtual reality courtrooms finds evidence of racial in-group bias. The AI problem is to personalize debiasing of mediators depending on their susceptibility to extraneous factors. The Kenya team will have a second project to determine which cases should be done virtually.

1.4 Brazil

- Detecting and Prosecuting Slave Labor: How may AI and ML methods aide human decision-making? The usual narrative is backlash to AI. A recent study found that in Kentucky, when judges were given decision-support, it ended up increasing disparities - not because the algorithm was biased - in fact the algorithm would have resulted in lower disparities. But the judges selectively paid attention to the algorithm, which resulted in greater disparities. Consider an incremental approach leveraging recent theoretical insights from social preference economics. The core insight is that judges are moral decision-makers, you're right or wrong, good or bad, and to understand what motivates these decision-makers, one might turn to self-image motives - "I think I'm a good person - a good judge" - a topic of active research in recent years. Each stage leverages motives of self-image,

self-improvement, self-understanding, and ego. In stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms). Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies and reminding the human of her prior choices in similar situations. Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns. Then, in stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts. The Brazil team will work with the Brazilian Federal Prosecutor office to design the RCTs that carefully estimate the impacts of the incremental AI framework. (Note: this data ecosystem is more rudimentary, but the conceptual framework can apply to elsewhere.)

1.5 Pakistan

- Legal Uber App: Some courts are highly congested whereas others have slack. This project applies the sharing economy principles of low transaction costs, coordination and information networks and better resource allocation and utilization. The app displays a dashboard interface that allows the supply side user, for instance a judge in this scenario, to view case availability. Cases on the app are prioritized by two factors, namely, the amount of time for which the case has been idle and the region in which it exists. Judges on the app can volunteer into an incentive scheme that allows them to earn points as they complete the cases. These points can then be spent on various rewards like access to interns, working from home allowance, flexible scheduling, and public recognition. The AI/ML problem is to solve the optimal dynamic point system and estimate its impact on court and litigant outcomes. (Note: this data ecosystem is more rudimentary, but the conceptual framework can apply to e.g., India.)

- Training Empathy to Elite Civil Servants: Randomizing different schools of thought on cultivating prosociality suggests that training the utilitarian value of empathy elevates pro-sociality among high-stakes decision makers in Pakistan. One month after training, treated civil servants display 0.4-0.6 sigma greater altruism; two to four months after, orphanage visits and blood donations double. Field and lab results suggest improved theory of mind in strategic dilemmas: blood donations only increased when treated individuals were told their exact blood type was needed, and training improved cooperation, coordination, guessing the decisions of others, and honesty. Treated civil servants also chose empathy books in a book lottery, scored higher grades on a soft-skills course, and increased language of social cohesion in social media. The Pakistan team will use ML to investigate the broader impacts on language use in social media (Twitter and Facebook), impacts on who they choose to follow, and impacts on the followers' language use.

1.6 India

- Legal Search: Fair and effective justice systems require "google translate" to past precedents and to communicate to judges. This project has two aims. First, to investigate the impact of legal search engines on case resolution (verdicts and duration) and ideas (diversity of citations and memes). Second, to use AI/ML to assess best practices in search to specific arguments that minimize communication frictions to judges, and speed justice.In 2011, Kanoon made free and public the corpus of Indian high court opinions. Preliminary analysis by DEJURE finds that Kanoon enabled the courts to cite themselves and increased the diversity of citations. The broader scientific interest is the impact of Google on Wikipedia is hard to know, since Wikipedia didn't exist prior to Google; here, we can study common law, to see the polarization or democratization of justice.

- Law Platform for Automated Impact Analysis of Judicial Decisions: Legal scholars and judges have long made arguments about laws and regulations and justified their arguments with theories about the effects of these legal rules. This situation resembles the field of medicine a century ago: prior to the advent of clinical trials, there were only theories without rigorous causal evidence. A growing body of empirical research now demonstrates that causal inference is possible in when cases are randomly assigned to judges. Randomizing cases to judges predicted to be harsh or lenient generates the inference on the long-run causal impacts of the length of sentences. This project has four parts: First, automatically identify the nearest previous cases when a case appears; second, fast-decision classification of the prior cases' directionalities; third, use document embeddings for low-dimensional representation of legal dicta and reasoning; fourth, use judge embeddings based on the history of their writings and citations to predict their verdicts on cases. This can be used to support judges in estimating the potential impacts of their rulings on downstream economic outcomes.

1.7 Bangladesh

- Apps for Gender-Based Violence: In many settings, lack of trust or access to justice stymie recognition of wrongs and rule of law. The Legal Uber App addresses cases actually filed, but what about cases that never file? In Bangladesh, the government has deployed an app called JOY for gender-based violence. An emergency button sends the victim's GPS, picture, and audio recording. The Bangladesh team develops builds off open-source software beiwe.org that collects passive features for additional social-emotional data. Part of the ML challenge is to minimize the amount of phone data needed to capture data predictive of abuse. The second challenge is to automate nudges to minimize abuse.

1.8 Estonia

- Quadratic Voting for Survey Research: Services provided by local governments vary substantially and are non-harmonized. There is little information available that would enable citizens and policymakers to assess and deliberate how these services should be further improved. Quadatic Voting for Survey Research facilitates measurement of citizen preferences and therefore deliberation. The antecedent challenge is whether a new survey methodology actually elicits data that is better predictive of behavior, for instance, in predicting voter turnout. An RCT randomized exposure to QVSR or Likert. One ML project is to assess whether survey responses in which the respondents were exposed to either QVSR or Likert on several questions are more predictive of their actual turnout.A second is to evaluate if a particular survey method is better in predicting the probability of voting using covariates and the survey responses. The third task is to identify the survey questions (and maybe covariates) that may have causal influence on the prediction probability. The live survey method can be improved, for instance, a tinder-interface for policy preferences, or the ability to propose new policy questions for other survey participants. (Note: the framework will be piloted on existing US data.)

2 Disparities[2]

- Emotions: Rule of law relies on judges enforcing the law in an impartial and predictable manner. This project assess indifference of judges to defendants revealed through their susceptibility to behavioral factors. Can judicial irritation decrease by intentionally preventing random fluctuations in the number of violent charges or the types of people (race, gender, number of past convictions) that a judge sees by spreading cases evenly throughout the week/month? Are we able to detect increased irritation/aggressiveness on weeks/days when a judge sees many violent cases in their non-criminal opinions (by maybe looking at their use of hyperbolic language regarding right/wrong)? If we are able to extract "racial slant" from language, are "racially slanted" judges easier/harsher on people of the same skin tone when they see more people of that skin tone on a single day? We can do many CATE estimations on the "racial slant" dimension. Or, more generally, we can try to understand the emotional states of judges over time via opinions and see if that corresponds to higher than average/lower than average sentences during that same time period. Are periods of higher emotional intensity in writing following periods of many violent crimes?

- Sentencing Under Fairness Constraints: This project uses offline contextual bandit to find optimal sentencing with "fairness constraints" to minimize recidivism. Then applying policy to test set to see CATE on different groups (race, age, etc.) for when policy was enacted or when policy was ignored. "Treatments" include different incarceration lengths, restitution amounts, probation lengths, etc.

[2]Contact: Ryan Carroll (rcarroll1@worldbank.org) and Daniel Chen (dchen9@worldbank.org)

- Early Predictability: Rule of law relies on judges being attentive to legally relevant features. This project develops an indicator of snap judgements. What does it mean for cases to be "too" predictable? The team will use features prior to case opening vs. features inclusive of case hearings. This index can be compared with measures of implicit attitudes of judges (textual or audio). How can this information be used to debias judicial proceedings?

- Impact of Economics on Justice: This project has two dimensions. First, assess the causal impact of financial payments to judges on how they vote, how they write, how they cite? How would one detect corruption in judicial corpora? Second, judges who attended law and economics conferences show greater racial and gender sentencing disparities. Can ML shed light on the reason for these disparities?

- Supreme Court Scorecasting: U.S. Supreme Court Prediction (or generally, predicting appeals court decisions using features of lower-court decisions) but also using oral arguments (similar data exists for U.S. Circuit Courts). How much time is needed ("nowcasting") and interpreting features in isolation (recordings, transcriptions, case and judge history). Does the model fit with other data (audio of other speakers' ideologies, or recordings from police stops)? Modeling social influence via phonetic accomodation in oral arguments. One project involves applying the Bayesian Echo Chamber model to the audio data (not just textual data) in oral arguments. Nothing is known about whether judges' questions or how lawyers' respond to such questions are predictive of outcomes.

- Legal Schools of Thought: A perennial question is how much law clerks affect judicial opinions. Several prominent studies suggest traces of the writing styles of clerks are reflected in the final opinions. This project examines legal schools of thought using the law school training of clerks. Several episodes of unraveling in the judicial clerkship market (where judges had to make hires with less information about clerks) allow studying the efficiency and equity consequences of the current clerkship market system.

- Death Penalty: Construct topographical map of WW1 loyalty using universe of courts martial and see if it explains the deterrent or anti-deterrent effect of the death penalty across different military crimes. Analyze physical descriptions of soldiers (youth, complexion, looks, occupation), especially outcome heterogeneity, e.g., do young soldiers respond more to executions? Model geographic spillover / news of executions using geo-location of military units.

3 Grammar of Law[3]

- Precedent: Economic development relies on judges enforcing the law in an impartial and predictable

[3]Contact: Sandeep Bhupatiraju (sbhupatiraju@worldbank.org), Henrik Sigstad (henrik.sigstad@econ.uio.no), and Daniel Chen (dchen9@worldbank.org)

manner. We know, however, very little about which features of a legal text tend to "bind" judges into deciding according to "the law" as opposed to according to their own personal ideologies. We have data on 2.2 million district court opinions and 300,000 circuit court opinions setting binding precedents. By using an event study design, one can estimate the causal effect of a precedent on future judicial decisions. One can then use CATE estimation to characterize the textual features contained in the most influential precedents. For instance, are precedents relying on vague or subjective language less effective in constraining future judges than precedents that rely on objective/bright-line language?

- Motivated Reasoning: How do district court judges appointed by Democrat Presidents differ from judges appointed by Republican Presidents? CATE estimation, together with the fact that cases are randomly assigned to judges, offers a way to answer this important open question: What characterizes the cases where Democratic (or Republican) appointed judges tend to favor the plaintiff? How different are judges appointed by Democrat and Republican Presidents in their decisions? How has this changed over time? We provide access to 2.2 million district court opinions with judge identifiers.

- Transmission: We can also develop causal effects of visiting judges from other circuits. What kind of language and legal ideas are "contagious", and which type of judges are more influential? The project could use ML/NLP in various stages: (1) Predicting the identity of a judge by their language and use of precedent. (2) Characterizing which type of language and legal ideas are more contagious. (3) Using CATE estimation to characterize the most influential judges. For short-term effects, we can also exploit the repeated random assignment to judicial teams.

- Polarization: Repeated random assignment to judicial teams also identifies the causal effects of being in minority, majority, or uniformity positions. Mechanisms of social cohesion can be measured, including assimilation, integration, confirmity, dis-assimilation, radicalization, egotism, persuasion, other-ing, sectism, etc.

- Causal Memes: Other possibilities include using writings from judges prior to their appointment (in JSTOR academic articles or district courts), the influence of lawyers' briefs, and the development of "points of law" in previously expert-collated datasets. The random assignment to panels can also develop estimates of causal memes from specific judges and the tendency to use particular phrases, citations, or concepts.

- Motivated Reasoning II: The *Fact-Value Distinction* is widely considered a source of conflict between science and ethics—the distinction between what can be known to be true and the personal preferences of individuals. An important step for automated moral reasoning is the ability to make this distinction. Court opinions have been previously annotated to distinguish between facts and

legal reasoning of a case. We want to train a model to classify text as fact or reasoning. This will be useful for asking whether a decision follows reasoning, or instead judges use reasoning as an ex post rationalization of a subconscious decision (motivated reasoning).[4] Do differences in fact descriptions explain final decisions? Do judges distort fact descriptions? Can the model be used to ask if legal areas be classified as objective or subjective? Are citation patterns reflective of these differences?

- Genealogy of Ideology Project: a) predicts the memetic *phrases and citations* that are likely to be passed along the network in forward citation, but do not otherwise appear in a distant case in the citation graph; b) detects propagation (peer effects) when the underlying network is partially observed (for example, in oral arguments); c) identifies influential ideas and thought leaders; d) predicts how the judge votes on the next panel, using only the history of who the judges previously sat with on panels and how the judges' votes aligned with the panelists. We use random assignment of judges to make causal inferences. Then one can test population biology theories like whether greater volatility (be they environmental or economic) in a jurisdiction leads to more legal innovation, more generation of memes, and more citations. Study cultural evolution in legal documents. Predict citations (number, or positive/negative treatment) in the circuit case based on the district court case. Measure innovation in laws (courts and/or statutes). Compute Bayesian Surprise in legal documents.

- Precedent II: Predict forward citations from the backward citations, citation networks can be the channels of propagation of ideas and backward citations can be analysed to predict the forward citations. For this, ML models can be trained to identify the case level variables (i.e. judge data, backward citations, time taken, length of text, etc.) in the citation network, which can effectively predict the forward citations of the cases.

- Precedent III: Generate a page-rank of judicial influence via the citations and generation of agreement or dissent (on 3-judge panels). Identify parts of opinions that spark citations, for identifying sections that lead to polarized citations, and for thinking of citation recommendation to judges.

- Judicial Mean Fields: Can we score judges on statutory interpretation, like textualism or originalism, critical legal studies, etc.? This model could then be used to simulate counterfactuals. For example, how would the decision in a case change by switching out the authoring judge? How would the style of language change for a different circuit? This will give a topography of ideology in the U.S. judiciary.

- Grammar of Law: Universal grammar is the theory of an innate component of the language faculty, independent of sensory experience. *The Grammar of Law Project*: a) exploits parallel multilin-

[4] For instance, are judges that are predicted to decide in favor of a plaintiff more likely to adopt pro-plaintiff reasoning from new precedents? If yes, this would indicate that the reasoning is used as ex post rationalization.

gual legal databases to identify equivalent legal phrases (to identify the 'molecules' of law); b) uses these molecules to automate the detection of legal inconsistencies; c) uses the molecules, and sentiment/treatment, to automate encoding of moral views. One potential approach is a convolutional neural network to recover feature mappings that are predictive of decisions, and feature mappings that have an impact on higher-court decisions. Is there an optimal deviation in legal consistency? How much legal innovation engenders subsequent (positive) citations? Is there a trade-off between innovation and reversal likelihood? Does the legal text or the structure of its citations predict subsequent treatment in terms of importance (citations), controversy (dissent), and mistake (reversal)?

- Legal Knowledge: Using a new benchmark dataset focused on legal summaries or holdings, the work will be primarily focused on building baseline models to evaluate the task difficulty. Evaluation will be done using BLEU (bilingual evaluation understudy) and other similar metrics leveraging state-of-the-art Natural Language Generation (NLG) techniques to generate line by line predictions of case summaries. *Legal Knowledge Representation:* Adapt or deploy state of the art knowledge representation embeddings and benchmark against expert labeled legal outlines. Leverage data on text and citations. Usage scenarios would explain what the rule means, answer questions about cases, automated chatbots.

- Originalism: Assist judges and legal scholars in determining the "original meaning" of a concept. Auxiliary work intends to increase access to justice by offering users the possibility of highlighting a part of an argument in a case and then showing all the related cases. Use case—someone can highlight a part of the court opinion and trace its original meaning or see the other cases that use similar arguments. Identify Circuit splits (two cases in different Circuits that have reached opposing conclusions on the law for the same set of facts). Automated discovery of the emergence of new legal issues.

- Demography of Ideas: Considers life tenure of judges and the relationship between aging, health, and output. Aging judges appear to become more lenient in criminal and asylum courts and use simpler language. Predict early death or dementia, or simply retirement, using judicial corpora. Does health covary with judicial predictability or unpredictability?

- Networks: Making a citation network graph and parsing it using AI algorithms like BFS/IDS/DFS to find shortest/optimum paths to the root cases, and doing the document vectorisation approach. More specifically, consider a node representing a case. By seeing this graph, we can conceptualize that the shorter the path, the more related are the cases. Validate using document vectors. Explore further to find related judgements to a landmark judgement in a specific type/domain of cases by traversing the graph. Do the same for judges as nodes.

4 U.S. Datasets

- Digitized universe of U.S. Courts of Appeals cases from 1880 to 2013 (roughly 380,000 cases), identities of randomly assigned judges sitting on the panels (who is authoring the opinions, writing dissents, or writing concurrences), hand-labeled judges' biographies, hand-labeled legal topics, citation network among the cases, seating network among the judges, 2 billion N-grams of up to length eight, and original text (for use with neural nets). Linked to publicly available Supreme Court datasets, U.S. District Court docket datasets, geocoded judge seats, (some) biographies of judicial clerks, 5% random sample that was hand-labeled for hundreds of features including vote ideology, presence of oral arguments, and administrative data from Administrative Office of the U.S. Courts (date of key milestones, e.g., oral arguments, when was the last brief filed, etc.). To study equal treatment before the law and equality based on recognition of difference and to study the transmission of legal thought. We also have the digitized universe of U.S. State Supreme Court cases from 1947–1994 (roughly 400,000 cases), identities of judges sitting on the panels, hand-labeled biographies, a citation network, and original text.

- Judges are randomly assigned in the U.S. Courts of Appeals. Data on 25 polarized legal areas have already been collected and hand-coded: sexual harassment, eminent domain, free speech, abortion, church-state separation, affirmative action, gay rights, disability rights, campaign finance, capital punishment, criminal appeals, desegregation, sex discrimination, punitive damages, federalism, National Labor Review Board, environmental protection, National Environmental Policy Act, Federal Communications Commission, Title VII, First Amendment, Eleventh Amendment, standing, contracts, and corporate veil piercing. To study the channels through which legal regulations have their effects.

- Digitized World War I British archival datasets, including universe of deserters (including names and often their birthplace) reported in military diaries, police gazettes, and handwritten military trials, commuted and executed capital sentences (which historians believe was random), geocoded casualties, maps, officer lists, and order of battle. To study the role of legitimacy in legal compliance and differential effects of the death penalty.

- Digitized universe of administrative data on 1 million refugee asylum cases and 15 million hearing sessions and their time of day across 50 courthouses and 20 years (with randomly assigned judges), hand-labeled biographical data of judges, and dozens of features on the case and the defendant. We know when the asylum case was assigned, whether the hearing was an individual hearing or whether multiple individuals were scheduled in the same session, how many cases were scheduled for sessions during a day for that judge, whether this was an in person hearing or by audio or video, whether it was a written or oral order, whether there are other related applications for relief filed

by the individual and the judge's ruling on each, ethnicity of the applicant, the reason for the case and the judge. Data linkages have been made to daily weather and local sporting events.

- Digitized universe of 1 million criminal sentencing decisions across 94 U.S. District Courts from 1992–2009 (with randomly assigned judges), hand-labeled biographical data of judges, and an 83 page codebook on features of the case and the defendant. Data comes from U.S. Sentencing Commission. Linkages to judge identity were obtained (not publicly available) and hand-labeled biographical data of judges incorporated. Data linkages have been made to daily weather and local sporting events.

- Digitized universe of individuals in a federal prosecutor's office over a decade with many stages of random assignment. New Orleans is the largest city and metropolitan area in the state of Louisiana. The Orleans Parish District Attorney's Office and its prosecuting attorneys are responsible for enforcing state criminal laws and local ordinances to protect and serve the citizens of New Orleans and surrounding areas. The current dataset is from 1988 to 1999 and provides detailed information on approximately 430,000 charges and 280,000 cases (involving 145,000 defendants) filed or adjudicated during this timeframe. The data collected also contains detailed information regarding each individual offender, such as social security number and the corresponding prosecutor and judge. The dataset is rare: vertical linkages from the time of arrest, including those sent home without a trial, otherwise do not exist. There is a 594 page codebook.

- Digitized speech patterns in U.S. Supreme Court oral arguments since 1955–longitudinal data on speech intonation (linguistic turns) are rare. Linked to hand-labeled oral advocates' biographies, (some) lawyers' faces, clipped identical introductory sentences, ratings of their traits, and publicly available U.S. Supreme Court databases containing dozens of additional features and preceding U.S. Circuit Court data. Actual analysis of speech patterns is statistically challenging, since speech is modified dynamically. A common measure for variation in speech patterns considers resonances of vowel sounds. In order to properly measure these, the starting locations of all distinct vowel sounds have been manually flagged. A machine then measured vowel resonances and assigned to each vowel sound a multidimensional continuous quantity. Therefore, the size of the oral data set is much larger than the size of the underlying text. Text is traditionally treated with discrete models. Speech measurements (for example, resonances) by contrast are continuous. Recordings of U.S. Circuit Court oral arguments for a limited time period.