

The Real Work of the Courts

Overview and Motivation

Sustainable development and a healthy economy crucially depend on a functioning judiciary. Though this is hardly a novel insight, up until a decade ago there was surprisingly little theoretical and empirical research on how the judiciary influences social welfare and economic growth. Ramello and Voight (2012)

The popular media report only the small fraction of cases that have sensational value, or that make big changes in the law. While these outlier cases are individually interesting, they leave most people with a distorted impression of what really happens in legal proceedings in general, and in Florida's courts in particular. Millions have followed famous murder trials or been entertained by the judge challenging the public defender to a fistfight. What is not widely known, however, is that the Florida judiciary is so large and diverse that some degree of aberration is inevitable. Our project will show you why that is the case.

Our goal is to visually develop the size and complexity of the work of the Florida courts, and to suggest some conclusions about how well they fulfill their functions. At the first level of analysis, we inquire into the nature of the caseload is, not just the raw number of cases filed, but what kinds of cases, and whether these quantities vary over time and geography. We ask what may be driving changes in volume and composition of cases, and depict some of the hypothesized relationships. For example, one might expect that the primary driver of the demand for judicial services would be the underlying growth trends in the population. While the data do support this proposition, they also render it unmistakable that there must be other forces at work. Recently, at least, most raw counts of new cases have been declining, and at the same time the state's population has been aging. For certain types of cases, such as criminal prosecutions, one might expect there to be a relationship between these, if older people are less prone to crime. For other types of cases, such as civil lawsuits, one might expect economic factors to influence how many new cases are filed.

Incoming case counts, of course, are only the first half of the inquiry, as the more interesting question is what becomes of those cases once they enter the system. The traditional metric for assessing the performance of a judicial system is the "clearance rate" which is simply the ratio of cases resolved to new cases file during a specific period. The clearance rate became a standard measure because the data is readily collectible, not because it is the best imaginable measure of performance. It is cross-sectional rather than longitudinal, meaning it compares current period resolutions to current period filings. It tells us only whether the overall level is rising or falling, but nothing about how long the resolved cases had been pending: a system with a ten year average time-to-resolution could have the same clearance rate as a system with one year average duration. Nevertheless, since clearance rate is well defined in the literature, there is some comparability of those measures across partitions of a system as well as across systems generally.

Like filing rates, clearance rates are subject to variation. They change over time and are not geographically uniform. Accordingly, we depict case resolutions and clearance rates over time, subject to

partition by geography and case type. In addition, we offer the opportunity to drill down to a relatively fine level of detail regarding resolutions-- whether the case was resolved by jury trial, nonjury trial, pretrial judgment, settlement or dismissal. By making this level of detail discoverable, we hope to visually communicate what actually happens in the Florida courts.

Related Work

There is an abundance of statistical data about court caseloads, but little by way of interpretive visualization. The [Administrative Office of the United States Courts](#) maintains detailed data for cases in federal courts. The [National Center for State Courts](#), through its [Court Statistics Project](#), collects some cross-sectional data intended to be comparable across states, with considerable effort made toward presenting data visually. Most state courts maintain some caseload data; among the largest states [California](#) and [Florida](#) provide websites with some level of interactivity, while [Texas](#) and [New York](#) report workload statistics in static .pdf annual reports. The majority of state court reports provide bar, line, and pie visualization; among the largest states only Florida provides a [text-based user-defined query interface](#). Although it is obvious that incoming court caseloads must be a function of some exogenous factors, there is practically no existing work on the identification, much less quantification, of what those factors are (Ramello and Voight 2012).

Questions

What questions are you trying to answer?

Just how many cases are initiated in the court system annually? What kinds of cases are being filed, and how many of each? Does the volume or composition of cases change over time? Does it change as population grows, or in response to economic events? Does it change geographically, from north to south or between urban and rural areas? Are the courts keeping up with the caseloads? Are they resolving as many cases as they are receiving? How many are resolved through trial vs. settlement? Does the result vary from time to time or place to place?

How did these questions evolve over the course of the project?

Initially, we were looking at the input and output numbers, and the ratio between them, known as the “clearance rate.” We were planning to depict changes in these numbers over time and space, and to relate them to population levels and court funding levels, on the theory that more population would result in more cases. In addition, we planned to inquire whether changes in the form or level of functioning of the judiciary influenced the clearance rates. On graphing the filing and disposition data over time, a pattern was evident, where the filing rates increased dramatically in 2008-09 when the economic downturn started. This led us to obtain economic data to see if it might show a relationship between economic swings and court cases generally.

What new questions did you consider in the course of your analysis?

The recent time path of the filing and disposition data required a reconsideration of the paradigm that initially motivated the choice of this topic. It became clear that simply visualizing the composition of the normal business of the courts would not be fully consistent with the recent data, being driven as it was

by some upheaval during the great recession. The original thesis, that the system is so large that occasional aberrations are inevitable, did not account for large-scale underlying changes in economic activity that could dominate the periodic fluctuations. Our project started to examine the work of the courts during normal times, now it examines the work of the courts during normal times *and* during times it is subjected to an exogenous shock, which fortuitously tells a more interesting story.

Data

The primary data source was the Florida Office of the State Courts Administrator, who upon request provided three tables that contained all the filing and disposition data in its Summary Reporting System going back to 1987. The first table contained annual filing data, classified according to county and case type; the second contained resolution data, similarly classified, and the third contained detailed disposition data, reporting how many cases of each type and location were resolved by jury trial, by pre-trial ruling, by dismissal, etc. The OSCA also provided budgetary data for the judicial budget compared to the overall budget, and for the subcomponents of the judiciary budget. All of this data was in excellent condition and needed no scraping or cleaning, though the files were in highly redundant csv flat format. The one file that did need to be scraped was the OSCA's display of the court structure, in which the parent node is the state, first children are appellate districts, then circuits, then counties. But it was a small job; correctly aligning each of Florida's 67 counties with its parent nodes took less than a half-hour.

The integration of data from other sources into the visualization is what extends this project beyond the existing work. Data regarding population demographics was obtained from census, and disaggregated time series of economic growth rates (gdp) were obtained from Bureau of Economic Analysis. In addition, a number of sources were used to obtain cross-sectional comparative data, including the federal courts data repository, and that of the National Center for State Courts. The NCSC data was not readily obtainable from NCSC, which insisted in exposing its data only through its "data viewer," but with a little detective work the original data set was located at ICPSR. Because that data is at present limited to observations for the year 2010, all of the cross-sectional comparison observations will be standardized on that year. Accordingly we obtained annual reports from the other comparably large states: [California 2010](#), [Texas 2010](#), [Florida 2010](#), [New York 2010](#). These reports are not used as data sources for visualization, but rather for comparative purposes in tooltip text.

We also obtained a longitudinal data sample from the clerk of the circuit court in Hillsborough County Florida, which contains the City of Tampa. The sample consisted of 2600 observations, 100 per year over 26 years, for each case type. Each observation had an identifying index, a filing date, and a closure date, along with the number of "events" as that term is defined in this particular clerk's record keeping system. This was the only data from which we could obtain reliable estimates of how long cases stayed open. To prevent bias arising from missing values for recently opened cases, we partitioned and reported the data by the year of closure, not the year of opening. Thus the data is interpreted as, conditioned upon the case being closed in year x, how many days did it take?

Exploratory Data Analysis

What visualizations did you use to initially look at your data?

Using the data in a customized version of Bostock's d3 showreel allowed us to survey several standard visualization layouts for each data set. This quickly showed that some data are better displayed using overlapping, instead of additive layouts. For example, adding filing and disposition numbers together has no ready interpretation, so they are better depicted in an overlapping fashion.

What insights did you gain? How did these insights inform your design?

The most conspicuous insight derived from the exploratory charting of the data was that the recent history is not representative of the day-to-day business of the court we initially set out to show. Instead, we will need to account for the surge in cases that occurred during the recent economic downturn. The disposition data at first blush suggest the court system was unprepared for the surge but reacted appropriately over time, showing a surge in dispositions about as large as the surge in filings, lagged by about two years.

The different nature of the recent changes compared to the historical ones was part of the motivation for choosing a vertical implementation of the multi-axis view in the discovery page. It places the most recent activity at the top, highlighting its recency and unusualness, and lets the user scroll to more historical, and quieter, times.

Design Evolution

The initial design mockup submitted with the proposal raised concern both with our TF and design studio that there could be an over use of the opening figurehead, and that the collapsing node menu system could create a chaotic and confusing distraction from the main visualization. With these words of caution in mind, we are moving forward in developing these features to the extent that a more complete evaluation can be done. The scale and position of the domain expert is used to sequence the narrative content and plays an important role in priming the expectations of the user. The exploding node based menu system is used to provide an interesting way to move sequentially through a branching domain expert narrative or user directed data control narrative.

Implementation

Our implementation consists of two pages, the landing and the discovery. The landing page shows the project's datasets in a showreel-style animation. While this display does permit the user to isolate a specific data set and view, it is meant primarily to serve as background for the opening narrative and invite the user to the next page. The discovery page contains the interactive visualization itself.

The left side uses a map of the state and a set of radio buttons to permit the user to isolate a specific county, a group of counties (a circuit), or a group of circuits (a district). Across the bottom a timeline and bar graph offer the opportunity to select a single year or group of years. Planned, but not yet visible in the prototype, is an additional display to show the detailed resolution data upon selection of place and time.

Evaluation

What did you learn about the data by using your visualizations?

It is too early to fully answer this question, but not too early so say what we didn't learn. As mentioned above, all the data obtained from public sources was cross-sectional, and did not tell us anything about the length of time an average case stayed open, or the number of events per case that occurred. We obtained a sample of longitudinal data, observing the starting and resolution dates of 2600 cases of each type, 100 per year over 26 years, from one county in Florida. That data also contained event counts per case, and it was hoped that there would be a clear pattern to report from that sample. Unfortunately, the

event count data was very clustered around low values with sporadic large—very large—values without any discernible pattern. So ultimately, the only use we were able to make of that sample data was to draw a time series of the time-to-resolution which did show a clear pattern of decreasing after the system recovered from the glut of mortgage foreclosure cases.

How did you answer your questions? How well does your visualization work, and how could you further improve it?

Our exploratory look at the data revealed that we had more to say than just the day-to-day operation of the courts; we have an instance of the courts responding to a stressor, namely a surge in cases without any warning or resources reserved against such an occurrence. To further improve the project at this point, we are going to incorporate the feedback from the design studio participants, and of course from our TF Samuel, and integrate the narrative of the response of the courts to the unusual demands they faced during the financial crisis.

References

Ramello and Voight 2012. *The economics of efficiency and the judicial system*, International Review of Law and Economics 32:1.