The Real Business of the Courts

Overview

> 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 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. Using data counting every Florida court case commenced or resolved from 1986-2013, over 80 million cases, our project shows why that is true.

Implementation

This product visually develops the real business of the Florida courts. It portrays the size and complexity of the work of the Florida judiciary, the types of cases, their geographic variety, and their results.

The opening page, called landing, is a little more interactive than most cover pages, containing an animated series of views of each of our data tables, with the opportunity to stop the animation to look at any one of the views
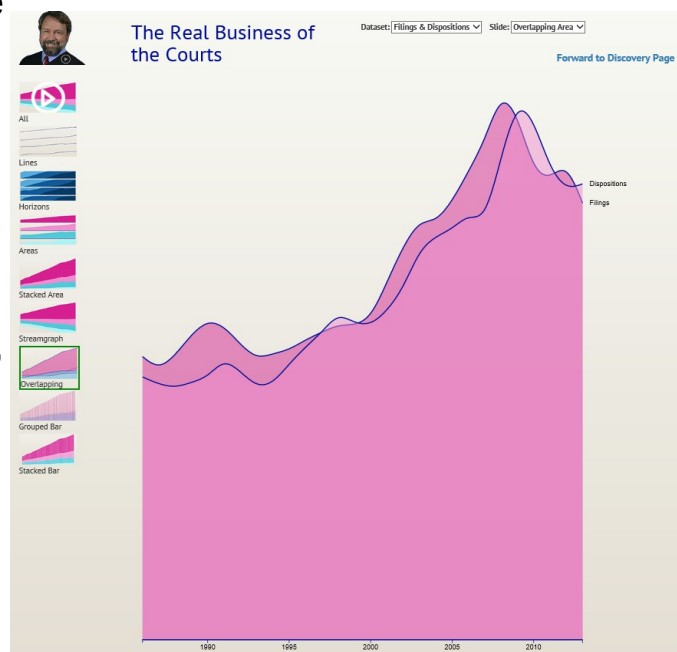

Illustration 1: Discovery Page

individually.

The main interactive page, called discovery, features a big timeline that shows the number of cases per year over the period 1986-2013. It can be updated to show cases for a particular region using the map, or by the type of case, using the bar chart (called "DOC") below the map. The next bar chart down shows the population of the selected region, by age ranges. Below the big timeline is a sankey diagram that shows the manner in which the selected cases were resolved.

The timeline itself acts to select a single year or span of years, which updates all the other elements to reflect data for the selected period.

Demographic and economic data are represented by population, by age, and per capita personal income, as a proxy for economic activity. These can be toggled on and off by the selectors above the timeline. That location also contains a data display control, permitting selection of zero-based, range-based, or indexed display. For indexed display each value is expressed as a percentage of its 1986 value.

In assessing court performance, the analysis starts with examination of the nature of the caseload, in units denominated "filings." A filing is the event that opens a legal case. These are depicted as heights on the timeline and sankey diagram, and as width on the case type bar graph.

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. Accordingly, we also display data for "dispositions" which are the events that close cases. Cases can be disposed by default, by pretrial rulings, by trial, or by some other process. The discovery page permits the user to identify the gross number of resolutions,
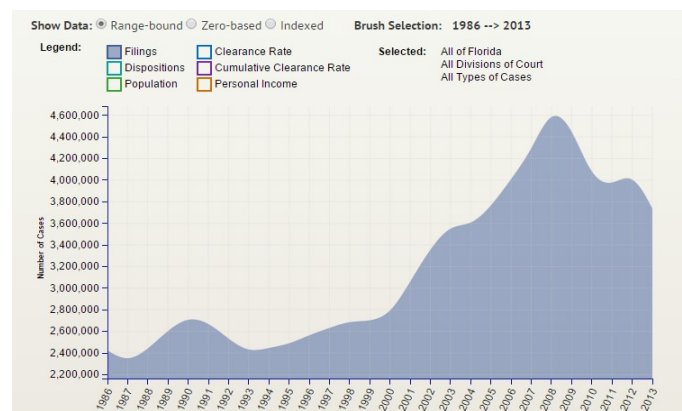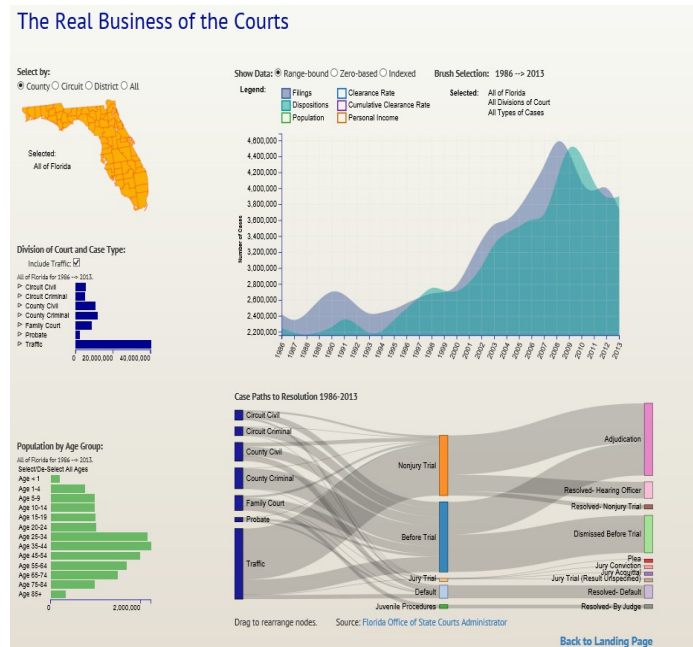
and to break them down into a fine level of detail

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 filed during a specific period. If the clearance rate is less than 100%, a backlog of pending cases is building; if it is more, the court is working down its inventory of pending cases. The clearance rate is depicted as a line that can be superimposed on the timeline by checking the associated box; the dashed line represents 100%. The clearance rate tends to fluctuate; in the overall view it is hard to tell whether a backlog is building or diminishing.
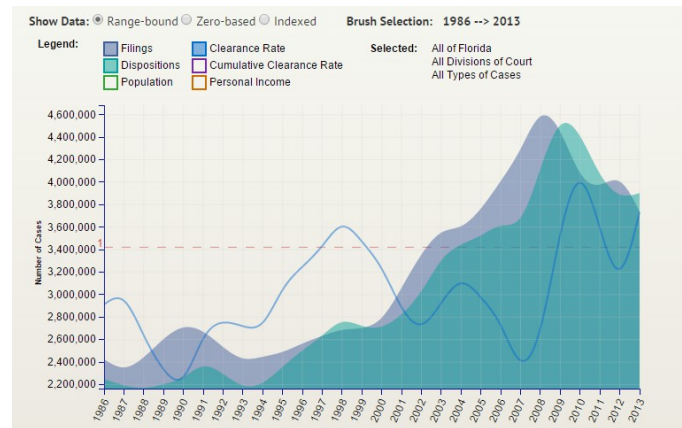
The clearance rate became a standard measure because the data is readily collectible, not because it is the best imaginable measure of performance. It is cross-sectional rather than longitudinal, meaning it compares current period resolutions to current period filings. It tells us only whether the overall 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.
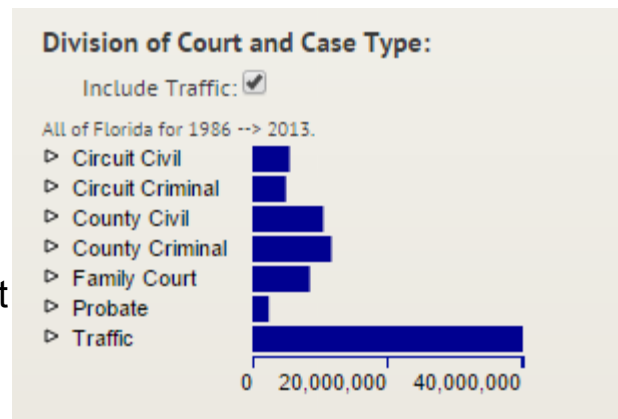
We introduce an extension of the clearance rate, the cumulative clearance rate, which provides a more visually meaningful measure of output compared to input over time. It depicts whether the court's backlog is building over time.

In considering whether the filings, dispositions, and clearance rates are driven by population, economic factors, or something else, it is crucial to observe that not all cases are the same. A traffic ticket and a capital murder trial are both counted as one case. Accordingly, it is necessary to be able to look at
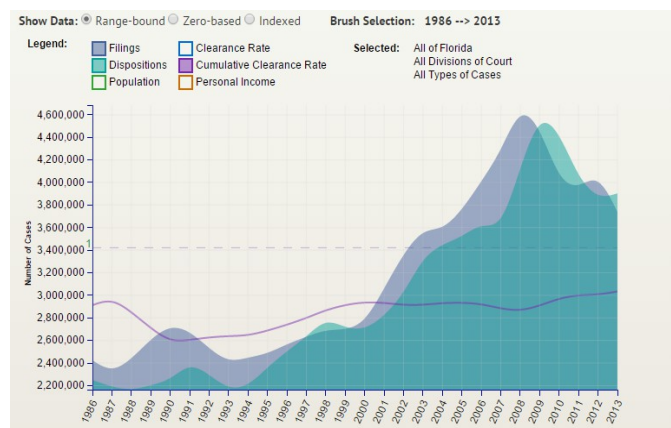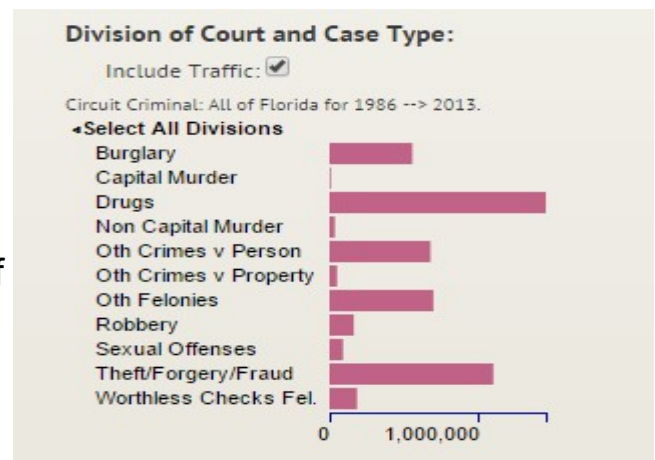
case types separately, which is accomplished by the bar graph below the map.

The Division Of Court ("DOC") element displays the number of cases by division, as filtered by the geographic selection from the map and the time-span selections from the timeline. It also provides a second layer of selection itself, as each division of court has several components of its own, called case types. Selecting one division of court by clicking its bar or its text causes the timeline and sankey to display data for only those divisions. It also causes the sub-menu of case types for that division to be exposed.
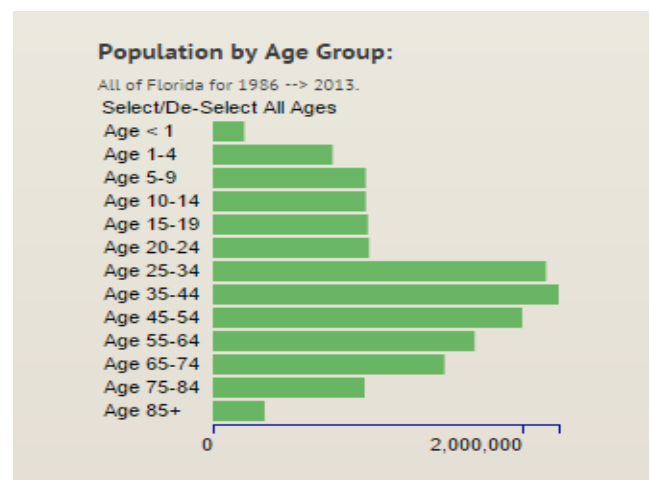
Division of Court and Case Type:

Include Traffic: ✔

All of Florida for 1986 --> 2013.
▷ Circuit Civil
▷ Circuit Criminal
▷ County Civil
▷ County Criminal
▷ Family Court
▷ Probate
▷ Traffic

0   20,000,000   40,000,000

For example, clicking "Circuit Criminal" causes the timeline to show filings, dispositions, and (if selected) clearance rates for felony cases only, and causes the case types histogram to show the sub-menu containing the various types of felony cases. One can then select one of the case types, such as Robbery, which will cause the timeline and sankey to display data only for Robbery cases.

Division of Court and Case Type:

Include Traffic: ✔

Circuit Criminal: All of Florida for 1986 --> 2013.
◂Select All Divisions
Burglary
Capital Murder
Drugs
Non Capital Murder
Oth Crimes v Person
Oth Crimes v Property
Oth Felonies
Robbery
Sexual Offenses
Theft/Forgery/Fraud
Worthless Checks Fel.

0   1,000,000

Clicking on a second or subsequent case type will additively select that case type also: with Robbery selected one can also select Drugs which results in the timeline and sankey displaying the sum of the data for the Robbery and Drugs case types.

The histogram on the bottom displays and subsets population demographics by age. It exists for the purpose of exploring the extent to which population factors may be correlated with filings, dispositions, and clearance rates. The display is responsive to geographic selection, case type selection, and time span selection. Clicking one bar of the chart selects that bar alone; clicking a second or subsequent bar additively

Population by Age Group:

All of Florida for 1986 --> 2013.
Select/De-Select All Ages
Age < 1
Age 1-4
Age 5-9
Age 10-14
Age 15-19
Age 20-24
Age 25-34
Age 35-44
Age 45-54
Age 55-64
Age 65-74
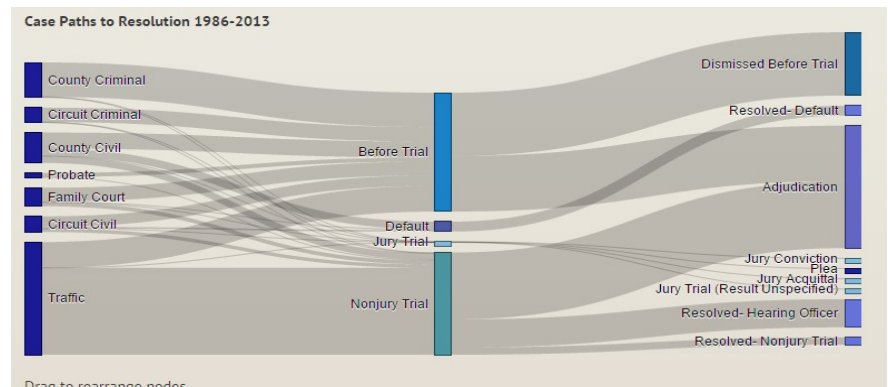Age 75-84
Age 85+

0   2,000,000

selects that age band also. The population line is displayed on the timeline, with an option to toggle it on or off; it has no effect on the sankey.

The map is used to select a geographic subdivision or arbitrary set of subdivisions, and it displays the selection. The smallest unit available is the county; Florida has 67. These can be grouped as the user may desire, simply by clicking on an unselected county to add it to the set. Clicking on a selected county will deselect it. For convenience, the user may group the counties all at once by circuit, which is a set of counties, or by appellate district, which is a set of circuits. Changing the map selection updates the timeline to show data only for the selected region, updates the DOC histogram to show the cases by division or by case type, similarly updates the age demographics histogram, and updates the sankey to reflect cases only for that region (and selected type and selected years).
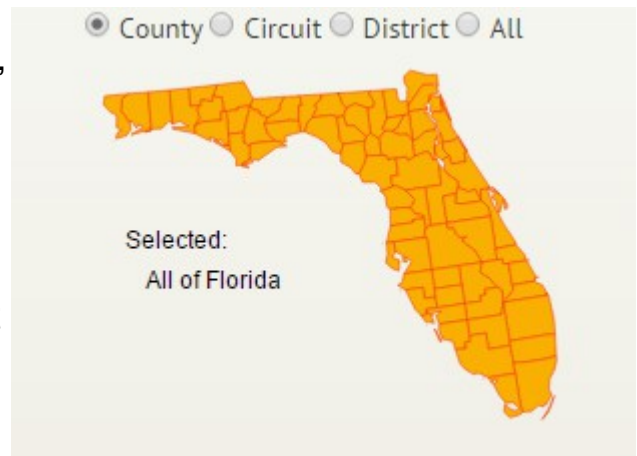
The final element is the sankey diagram. It is the only major design element in the discovery page that does not function as a selector. It reflects the selections made on the map, the DOC histogram, and the years selected from the brush function of the timeline, and shows the number of cases of each type, the path by which they were resolved, and their ultimate resolution. Hovering over a node will pop up a tooltip that displays the number of cases the node represents; hovering over a link will display the source, destination, and size of that link.

The sankey is interactive in one respect: a node, such as the node for jury trial, can be dragged into a new position to allow the user to isolate its links better.

Motivation

The discovery page is intended to ienable inquiry into the size and composition of the caseload; not just the raw number of cases filed, but what kinds of cases,

and whether these quantities vary over time and geography. Some important questions are what may be driving changes in volume and composition of cases, and whether the data support 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. These questions and more are whatwe hope to develop, and to enable each user to explore.

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 funding 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 as a function of population growth 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, still appears true, but it 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 1986. 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. Most of this data was in excellent condition and needed very little 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.

During testing, two kinds of problems in the detail disposition data were discovered. Some of the numeric values had been coerced to strings with quote marks, resulting in sums reporting NaN. Each quoted number was replaced with its numeric value. Second, there was a discontinuity in the data, as the reporting standards of probate cases changed in 1995. This was addressed by merging the pre-1995 data items with the post-1995 items in the same column. The following dendogram depicts the detailed disposition data structure:
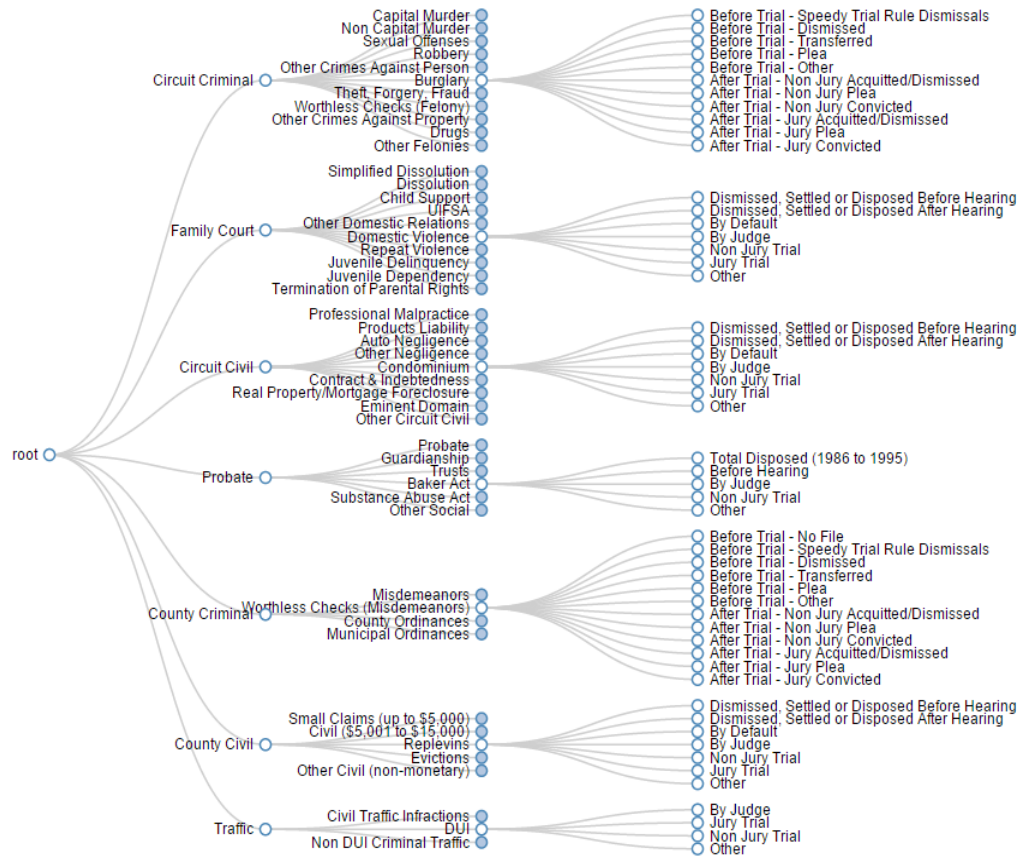
Illustration 2: Dendogram Showing Structure of Detailed Disposition Data

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. The gdp data, however, was not available by county, so the per capita personal income measure is instead used as a proxy for economic activity.

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 would 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.

As it happened we chose not to use the cross-state comparative data. It simply did not fit in well with our narrative.

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 2,600 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?

Even though the sample size was large, the data yielded very inconsistent results, so we report the sample data only in the landing page which makes no effort to be analytical.

Evolution of the Design

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. Ultimately, we concluded that this view would best be used on the cover page, in a rotating display of different depictions of various raw data sets, just to show what d3 can do with our data. But the process led to our choice of the overlapping area layout as the primary design element of the discovery page.

At the time, we had a different design paradigm in mind for the cover page. The original concept was to use warm colors to evoke Florida and make the domain expert
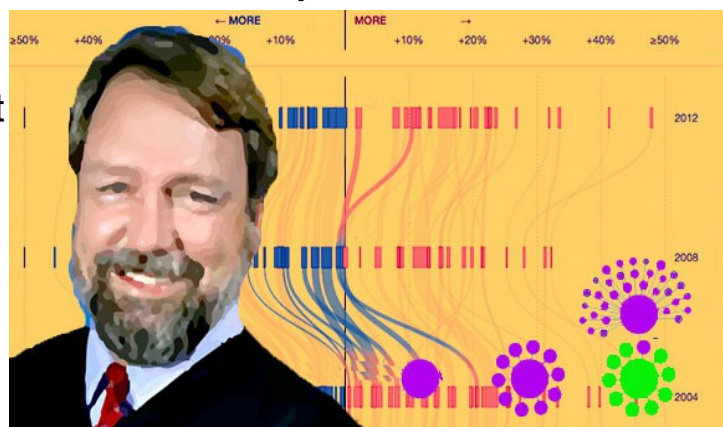
the major graphic element, with menu options arrayed in a collapsible force-directed layout. It was a visually rich opening page but after peer feedback and meetings with our TF we determined to go in a different direction.

We had in mind a version of the sankey diagram, vertically oriented as in the teaser on the course website. The basic idea remained, but the decision to make the sankey horizontal, and below the timeline, led to the layout that ultimately resulted, after a few intermediate efforts to settle on one.
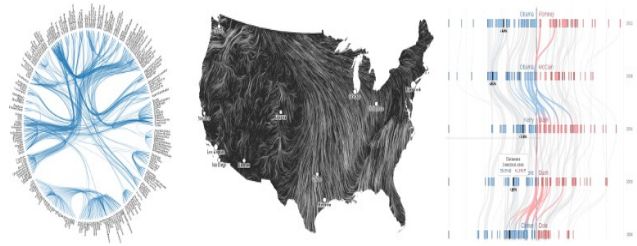
For the display of the case paths to resolution, we considered several other layouts, and went so far as to build working models of each one.
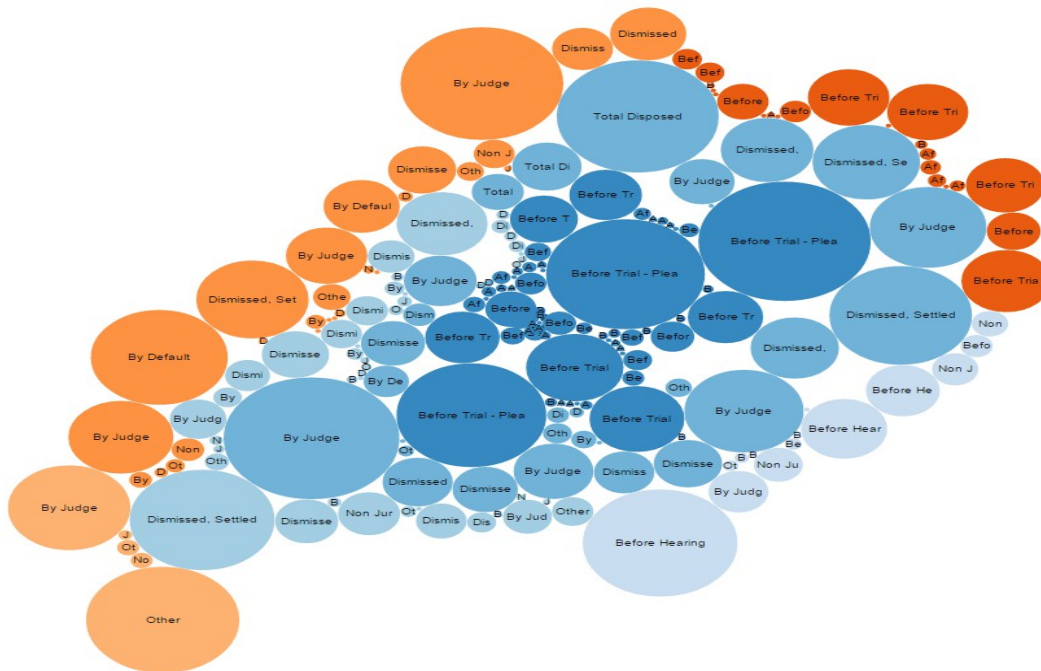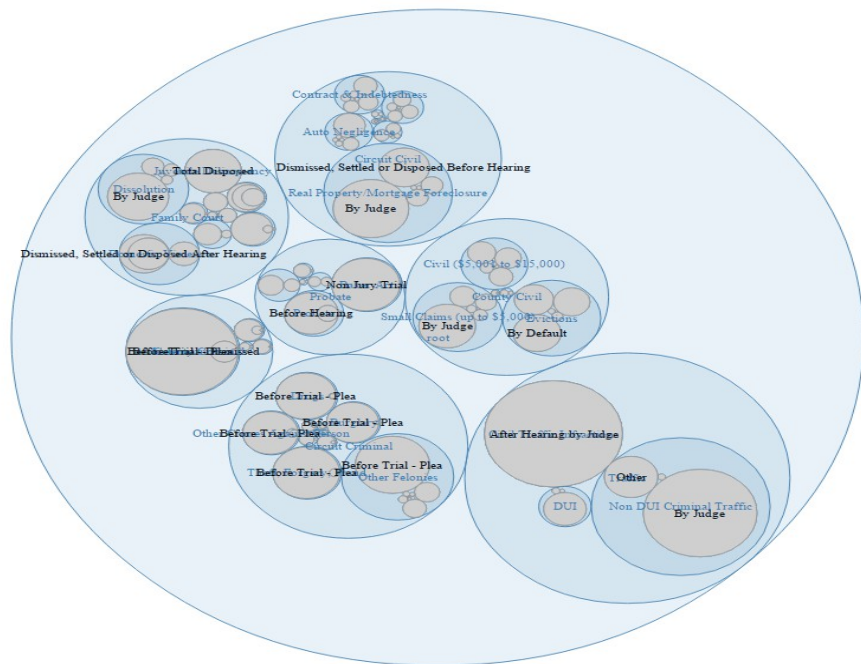

Illustration 3: Bubble Chart

Disposition Type Detail
circle packing
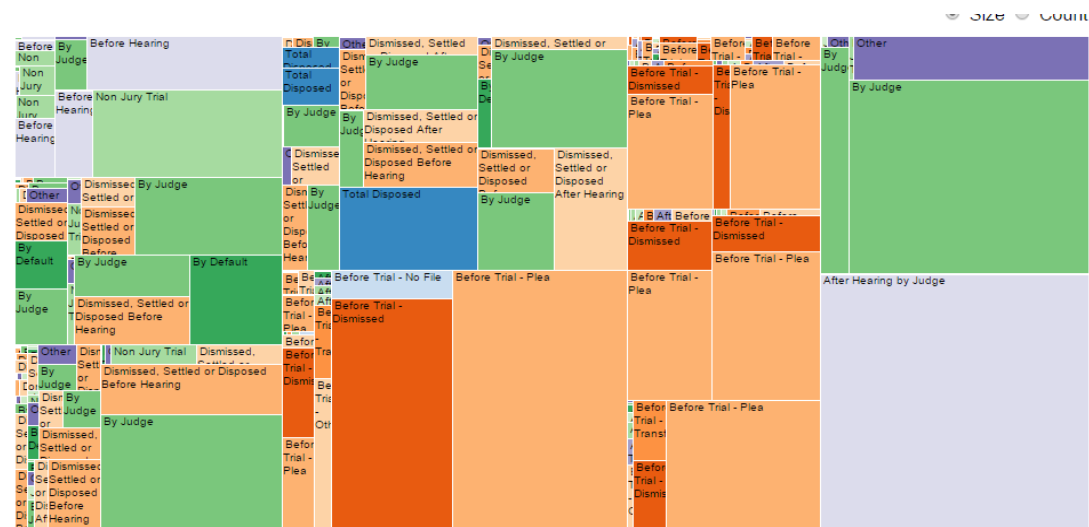*Illustration 4: Circle Packing Layout*

Illustration 5: Treemap

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 he 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.

Evaluation

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

The most striking conclusion form visualizing the data is that the recent economic downturn had a detectable effect on most types of cases, not just mortgage foreclosures and collection actions. Divorce cases plummeted while child support cases rose. Property crimes and robbery increased, but murders fell. Ultimately it seems that the court system is still reeling from the effects of that major economic event, which has dominated the effects of population growth.

There was also a lot of finding out what we did not 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 2,600 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.

At this point, the visualization works well to give a highly detailed inspection of the number and composition of cases filed and disposed, selectable by region, case type, and year. Some more explanatory tooltips, or perhaps other potentially comparative data lines, could be potential improvements.

References

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