Tom,

Thanks, this is very helpful.

1. I want to put two more things on the table – items 1 and 3 from my email of 5/1/2019 5:49pm to LN. In brief, they involve (item 1) missing cases and (item 3) missing Shepard’s treatments. I will forward you an email showing those items and a bit of followup from me on them.
2. ByungKoo and Tom, can you review item 1 from the 5/1/2019 email and see if you can add any detail that might be helpful to LN (and to us)?
3. Tom and ByungKoo, can you put together the tables and graphs that illustrate the key things we are missing? I think there are four key things in addition to the two items noted above:
4. Items 2 and 3 in your email below (missing/duplicate panels)
5. Item 4 (cases with two judges listed, even though there were three)
6. Item 8 (definitions of legal topics)
7. Item 9 (missing legal topics)

Note that a lot of these items could be related to item 1 from the 5/1/2019 email.

1. Perhaps should include, though less important:
2. Obsolete treatment letters. Tom, can you remind me where things stand on that? I believe I asked you about a couple of those letters, but I have forgotten where we left this.
3. Judge IDs are for authoring (opinion, concurring, and dissenting) judges. We have their names, but as Tom has noted, they should be able to give us the IDs pretty easily, which saves us the trouble of identifying which Judge X is the author when there are two judge Xs on a given circuit.
4. Things I’m not sure are worth pursuing right now – what do others think?
5. Didn’t my research cover all the forty cases missing an identifier (item 1 below)?
6. As to  item 5 below (judges missing a name): if none of them appear in panels in the data, then as of now this doesn’t matter, right? (I know that it could become relevant with new data.) Or am I missing something?

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>   
**Sent:** Sunday, June 9, 2019 12:56 PM  
**To:** Stuart M. Benjamin <[benjamin@law.duke.edu](mailto:benjamin@law.duke.edu)>  
**Cc:** Kevin Quinn <[kmq@umich.edu](mailto:kmq@umich.edu)>; Byung Koo Kim <[kimbk@umich.edu](mailto:kimbk@umich.edu)>  
**Subject:** Re: Missing panel plot follow-up

Hi Stuart,

I think that we have a good characterization of the problems uncovered so far.  As I see it, the issues that we have been discussing are:

* Approximately forty cases are missing a court identifier (listed in the 2019-03-12 validation excerpt document).  Some of these have been identified by the team using LN on-line and/or Westlaw.
* Over 200,000 cases are missing a judge panel.  Some of these are duplicated in the data, the duplicates having a panel.
* Some cases with panel are duplicated in the data.
* Certain cases with two panels in the supplied LN data have three judges listed in LN on-line or Westlaw (Bungkoo has identified several).
* Approximately twenty judges are missing a name (judge IDs are listed in the 2019-03-12 validation excerpt document).  However, none of these appear in panels for cases in the data (note that completion of missing panels may change this).
* Judge IDs are needed for authoring (opinion, concurring, and dissenting) judges.
* Citations exist with treatment letters that are classified as obsolete.  Should these be omitted?
* Definitions for legal topics are needed.
* Over 600,000 cases have no legal topics.  Is this OK?

It seems that we could provide the above list, plus anything I have overlooked, along with the tables and graphs that we have generated in exploring the data, and ask for a plan to remedy the issues.  It would also be fair to point out that additional issues might be uncovered as we improve the data and that their continued assistance in resolving them is appreciated.  A restatement of the benefit to LN in supporting our research might help to keep them motivated.

Tom

**From:** Stuart M. Benjamin  
**Sent:** Friday, June 7, 2019 12:07 PM  
**To:** Tom Balmat  
**Cc:** Kevin Quinn; Byung Koo Kim  
**Subject:** RE: Missing panel plot follow-up

I am unsure about how to proceed. We want to go back to LN to alert them to any missing data. Before we go back to LN, we need to be sure what is actually missing. Are we at that point? Do we need to do more to determine exactly what is missing?

Stuart M. Benjamin

Douglas B. Maggs Professor of Law

Duke Law School

210 Science Drive

Box 90360

Durham, NC 27708

[benjamin@law.duke.edu](mailto:benjamin@law.duke.edu)

(919) 613-7275 (phone)

(919) 613-7231 (fax)

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>   
**Sent:** Wednesday, June 5, 2019 5:21 PM  
**To:** Stuart M. Benjamin <[benjamin@law.duke.edu](mailto:benjamin@law.duke.edu)>  
**Cc:** Kevin Quinn <[kmq@umich.edu](mailto:kmq@umich.edu)>; Byung Koo Kim <[kimbk@umich.edu](mailto:kimbk@umich.edu)>  
**Subject:** Missing panel plot follow-up

Hi Stuart,

The following plot is related to the one in my previous e-mail, but this one plots cases with missing panel as a proportion of total cases by court and year.  The similarity of total missing panel and duplicated with panel lines for court IDs 1, 4, 8, and 10 is seen here as in the previous plot.  For all courts, the proportion of cases with no panel, but duplicated with panel is near zero prior to approximately 1990.

Tom

