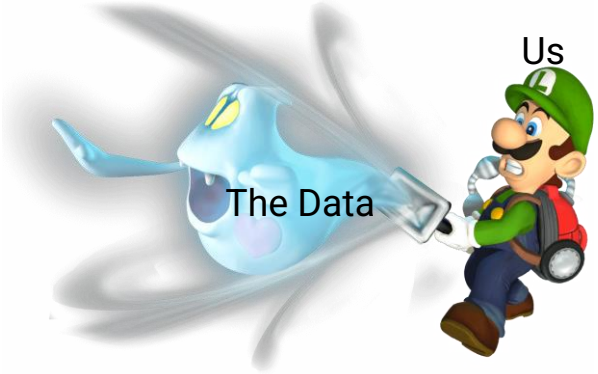
A Statistical Analysis of ABA Pro-Bono Civil Legal Consultations

The Paranormal Distributions
Eli Card (xe/xem)
Noah Cline (he/they)
Carissa French (she/her/hers)
Brendan Pinkerton (they/he)

Base Analysis and Data Cleaning



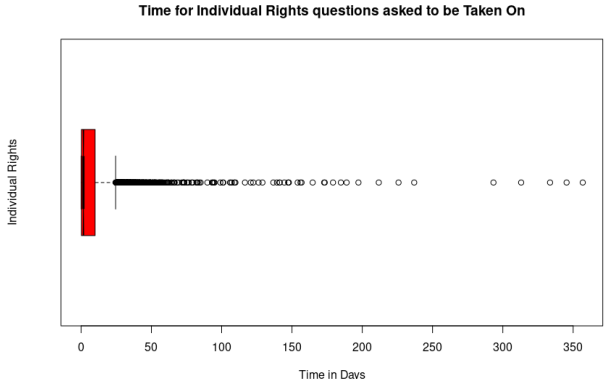
Interesting Observations

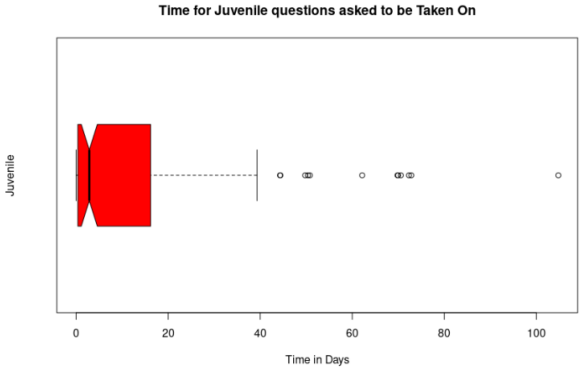
- Missing **states**
  - Alaska was not in statesites.csv, but had many questions
  - Idaho was there, but only had 1 question, whose post did not appear to be recorded
- HTML
  - One Lawyer sent an HTML file as a response
- Zip Codes
  - Had some zip codes with only 4 digits
- Questions that took longer than 10 days to get a **response**
  - We found that the time **varied by category**
  - Time data contained numerous outliers skewing towards abnormally long response times, up to 375+ days.

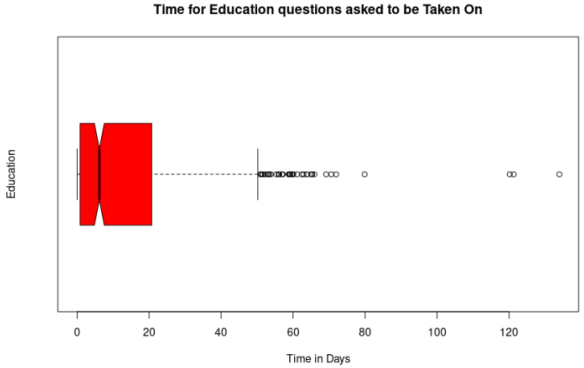
Data Cleaning

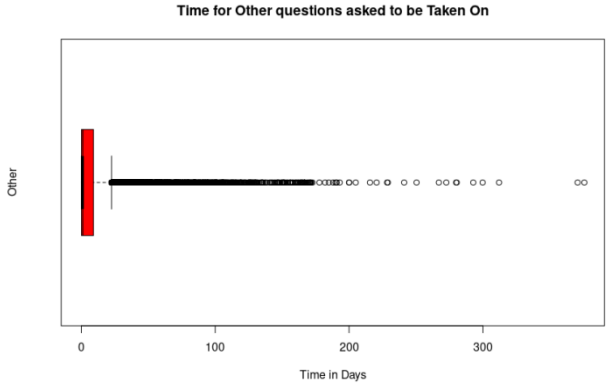
- Work around “ and , in questionposts.csv
- We determined that subcategories are state specific, and thus largely useless.

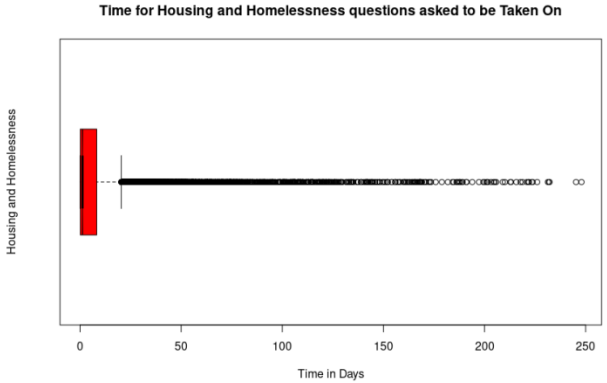
Category Box and Whisker Plots



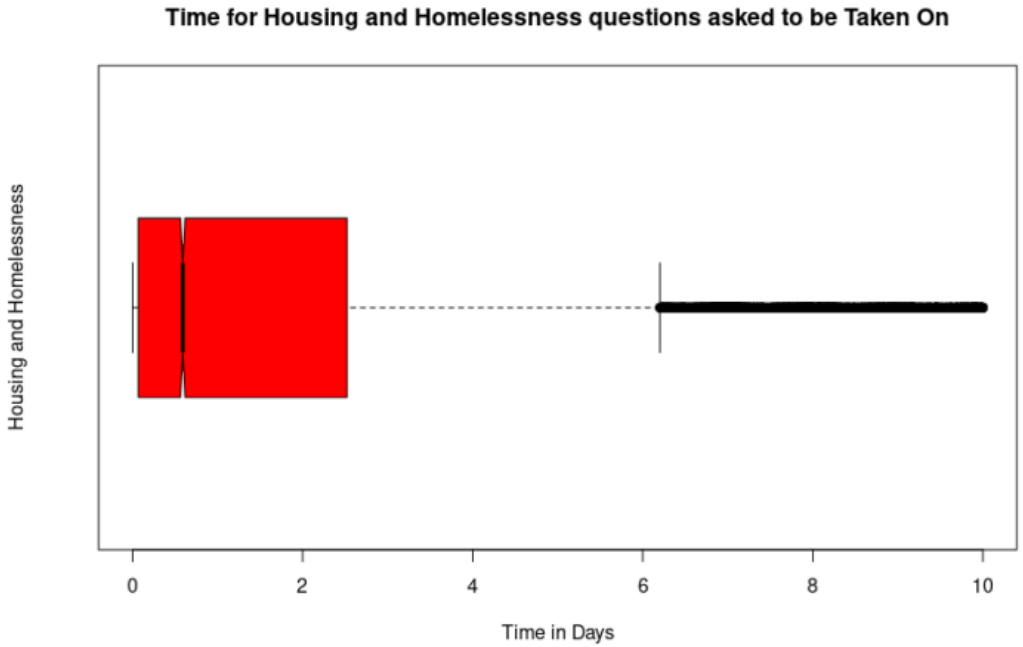








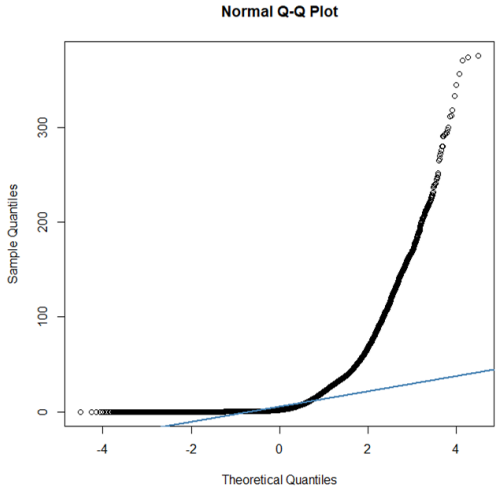
Category Box and Whisker Plots
(cleaning outliers longer than 10+ days)



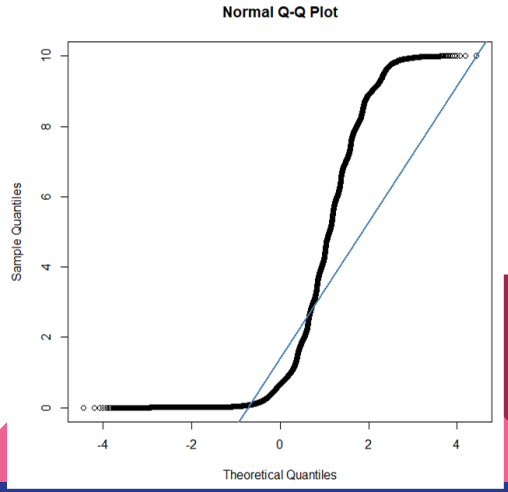
Check For Normality

- Points should form a straight line
- These plots are for the difference in time between questions being asked and being taken on
- CONCLUSION: data is not normal and tends toward extreme values

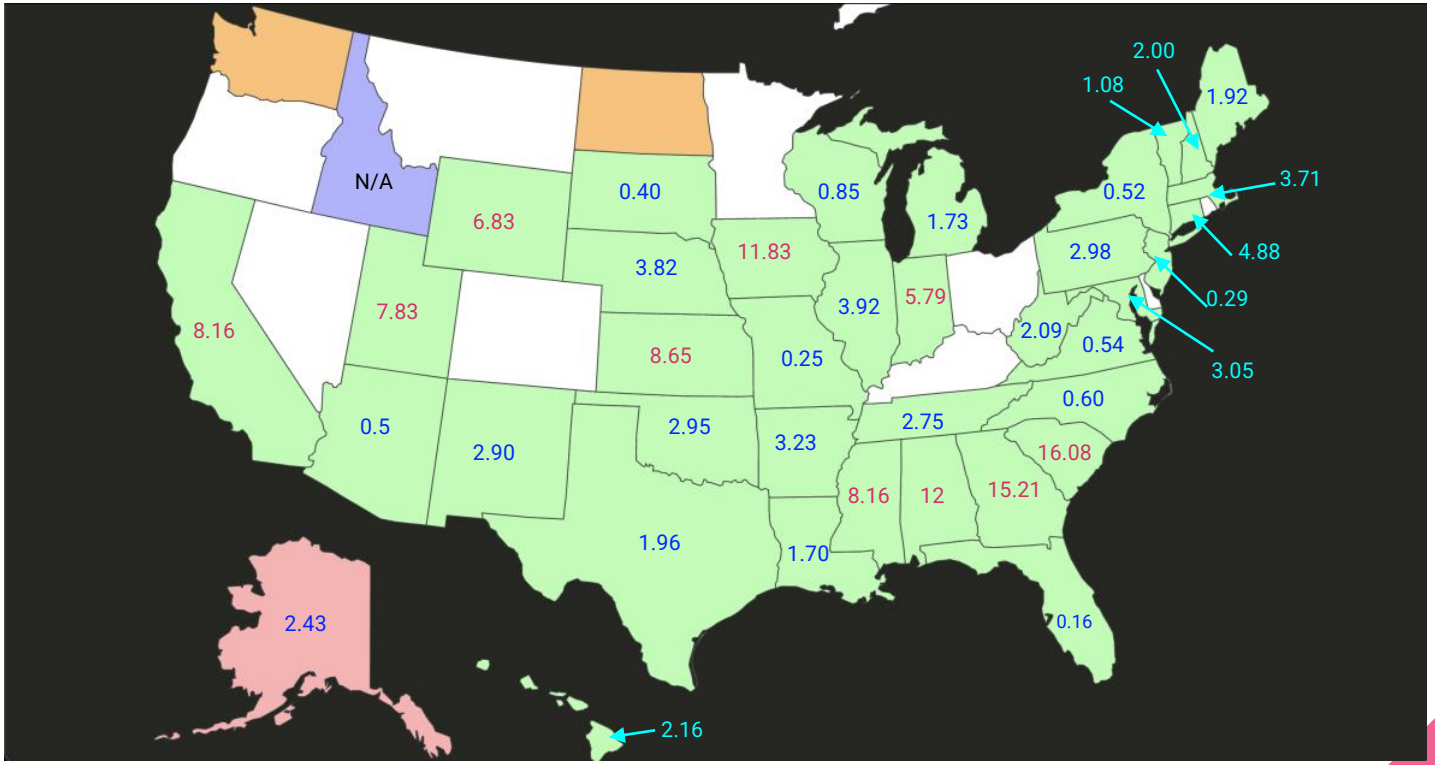
(all responses)



(responses <= 10 days)



State Discrepancies



Numbers given are median response times in days:
Blue -> less than 5 days Red -> greater than 5 days

State Key

Green: In State Sites Questions Posted

Blue: In State Sites One Question Posted

No Question

Gold: In State Sites No Questions Posted

Red: Not In State Sites Has Questions Posted

Blank: Not In State Sites No Questions Posted

Proportionally Most Asked & Taken Category By State:

Most Asked Category	Most Taken Category
CA, NY, US - Other	NY, US - Other
MA - Housing and Homelessness	MA, VT - Housing and Homelessness
All Others - Family and Children	All Others - Family and Children

Large Differences (>5%) in percentages:

State_Abbr	Category	Percentage_Difference	Percentage_Taken	Percentage_Asked
PA	Housing and Homelessness	9.3795093795094	26.9841270	17.6046176
CA	Other	9.1635620909834	24.3992606	33.5628227
AZ	Other	8.4156996223863	17.4871418	25.9028414
VT	Housing and Homelessness	7.0345313601128	34.8484848	27.8139535
MS	Other	6.8874233351737	13.9653415	20.8527648
OK	Family and Children	6.7035689880494	51.2763596	44.5727906
PA	Other	6.4935064935065	19.0476190	25.5411255
AZ	Housing and Homelessness	5.7712187956906	10.7274063	16.4986251
KS	Consumer Financial Questions	5.6623186962727	15.7024793	10.0401606
ME	Family and Children	5.44329305632431	60.2715877	54.8282947
MI	Family and Children	5.4269916035224	39.4736842	34.0466926
GA	Family and Children	5.2873319394115	44.1725692	38.8852373
CA	Housing and Homelessness	5.0739530607245	27.4491682	22.3752151

Correlation Testing:

Pearson Correlation Test:

- We tested the correlation between the percentage of the total asked and taken questions in each state, and found that with the exception of US data (which took only Other questions), for all states there is a significant correlation between these for all states.

  (maximum p-value for any state: 1.332e-05 [CA])

- Doing likewise for categories rather than states, we find even more significant correlations, with a maximum p-value for any category being Income Maintenance at p = 2.647e-11.

Conclusion & Further Steps

On Time Data:

- Large Outliers Indicate Systemic Issues

On Categorical Data:

- Numerous state's attorneys took up disproportionately more questions in some categories (e.g. PA and Housing and Homelessness questions) or did not take up disproportionately more questions in some categories (CA and Other questions).
- The ABA could use this information to find or train attorneys in such states to be more suited to answer questions they underperform in to improve overall responses and attempt to rectify these biases and get all categories of questions taken up as are asked.