

# ABA DATA ANALYSIS

The Pirates (we use R!!!)

Jacob Beaird & Moe Toyoda

Southeast Missouri State University

# Data Overview

- Our analysis utilizes "questions.csv", "questionposts.csv", "clients.csv", and "attorneys.csv"
- Laws and demographics differ from state to state, so we focused our analysis on data from Missouri alone
- Our analysis is designed to be applicable to other states as well

# Objectives and Plan

## Objective I: Optimize client/attorney pairs

Plan: Determine the "typical" client to ask questions in each category and attorneys who are most skilled in each category. Pair these people together.

## Objective II: Prepare attorneys for conversations in each category

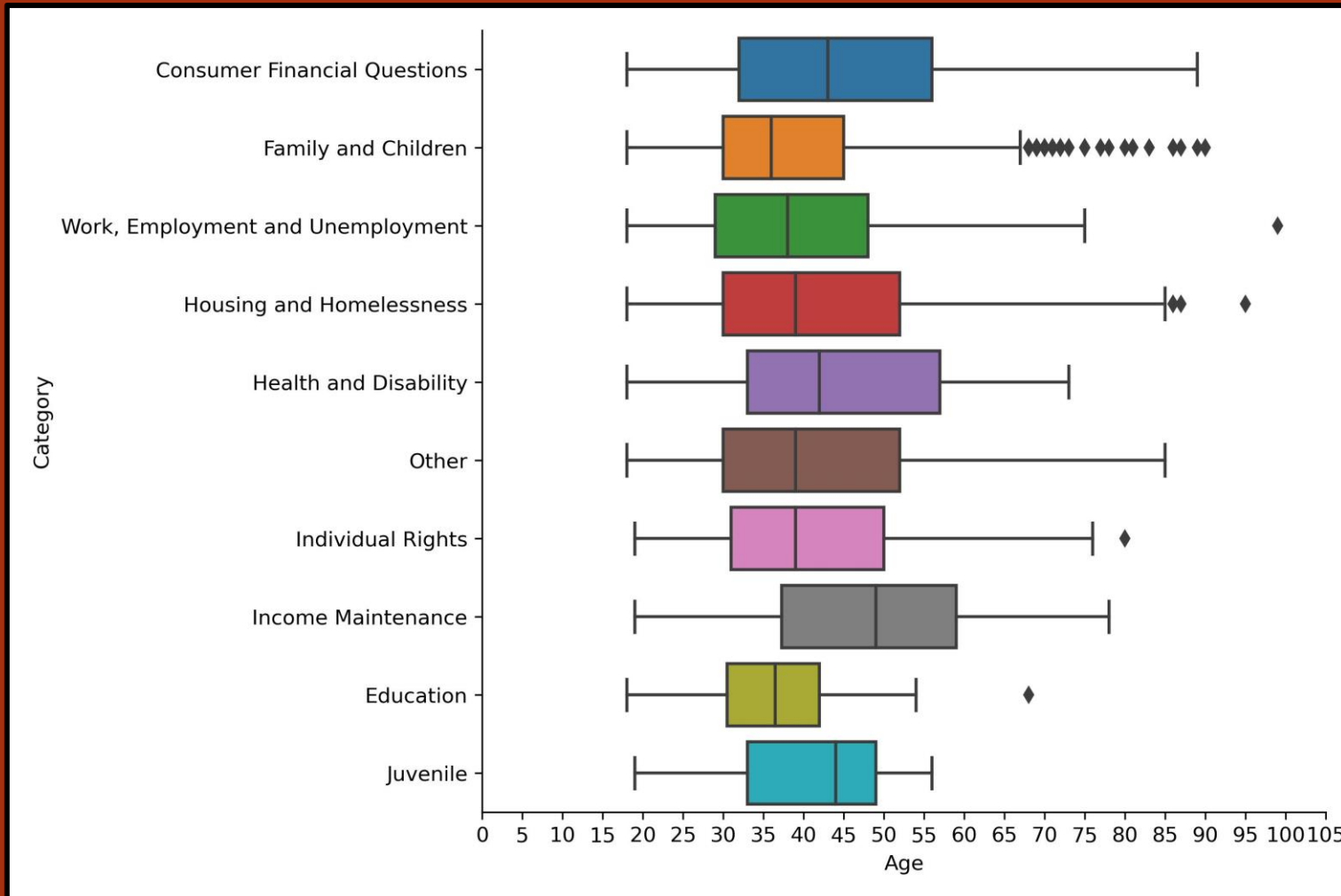
Plan: Determine most frequent topics of conversation and present this information in a digestible manner.

# Client Demographic Analysis - Numeric

Category	Age		NumberInHousehold		AnnualIncome		AllowedIncome	
	mean	median	mean	median	mean	median	mean	median
<b>Consumer Financial Questions</b>	44.2	43.0	2.5	2.0	24987.9	24000.0	42748.8	32470.0
<b>Education</b>	36.0	36.5	3.3	3.5	26916.8	22608.0	57545.0	60790.0
<b>Family and Children</b>	38.4	36.0	3.0	3.0	20540.4	18000.0	50493.0	51350.0
<b>Health and Disability</b>	44.0	42.0	2.4	2.0	21777.0	18000.0	40411.6	32470.0
<b>Housing and Homelessness</b>	41.8	39.0	2.6	2.0	21046.8	19000.0	44484.5	32470.0
<b>Income Maintenance</b>	47.5	49.0	2.1	2.0	14764.9	10280.0	33490.5	32470.0
<b>Individual Rights</b>	40.8	39.0	2.7	2.0	17619.5	14000.0	45664.1	32470.0
<b>Juvenile</b>	40.1	44.0	3.5	3.0	30492.7	20000.0	60234.7	51350.0
<b>Other</b>	41.5	39.0	2.6	2.0	19717.4	15000.0	43732.4	32470.0
<b>Work, Employment and Unemployment</b>	39.9	38.0	2.7	2.0	23304.9	20000.0	45197.6	32470.0

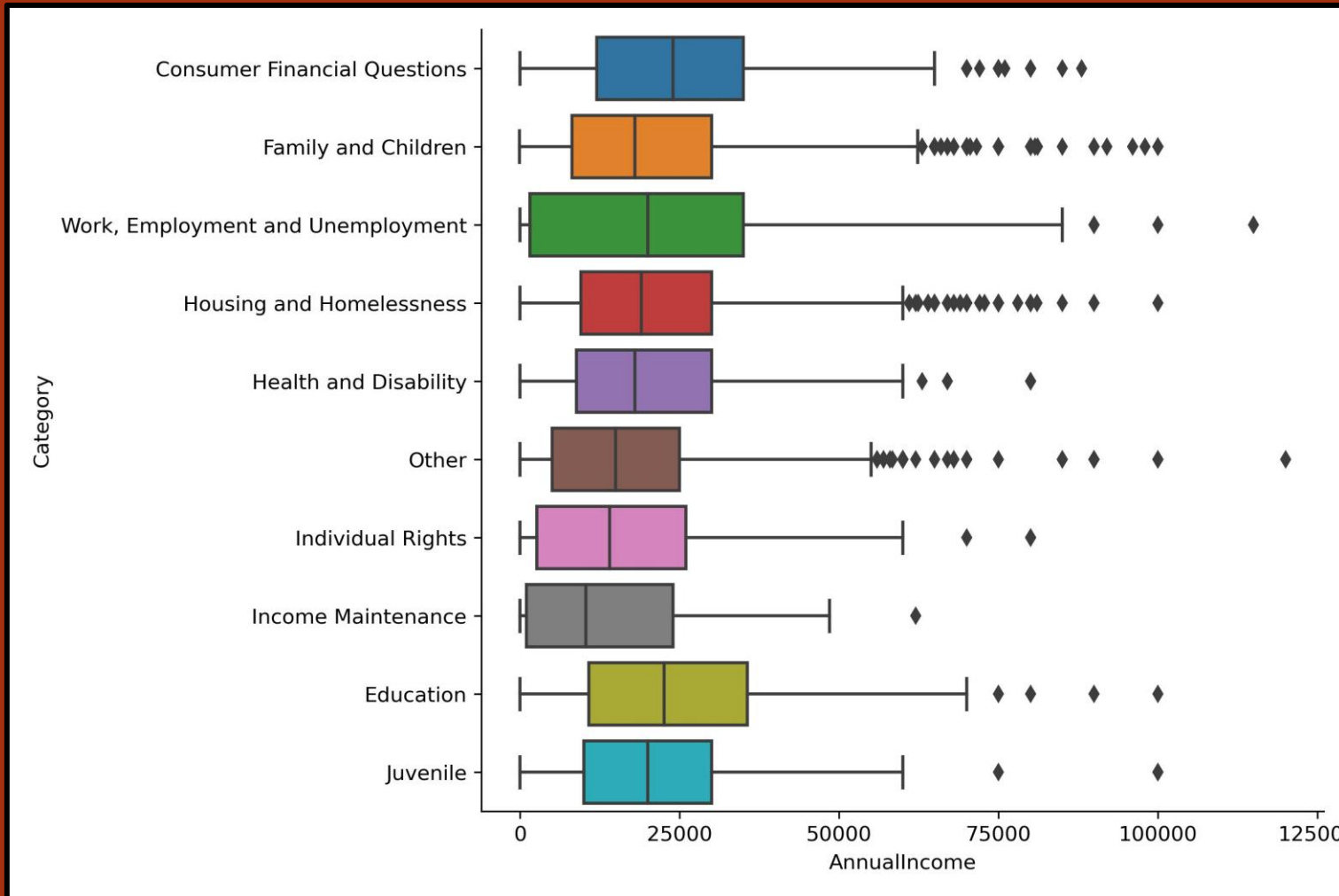
Let's explore each demographic factor more deeply with some boxplots!

# Age



- "Consumer Financial Questions" and "Income Maintenance" tend to be oldest
- "Family and Children" and "Education" tend to be youngest

# AnnualIncome



- "Consumer Financial Questions" and "Education" tend to have highest income
- "Work, Employment, and Unemployment" has largest spread of incomes

# Client Demographic Analysis - Categorical

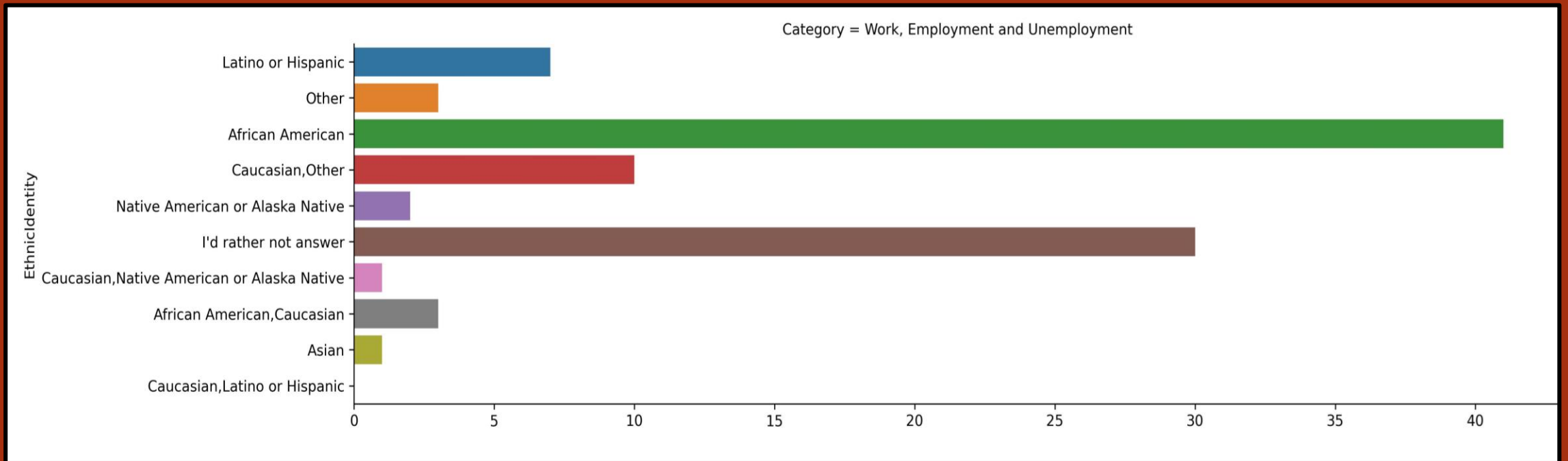
At first glance, this result does not appear interesting at all!

Let's remove dominant categories and analyze the remaining distributions.

	EthnicIdentity	Gender	MaritalStatus
Category	Mode		
Consumer Financial Questions	Caucasian	Female	Single
Education	Caucasian	Female	Single
Family and Children	Caucasian	Female	Single
Health and Disability	Caucasian	Female	Single
Housing and Homelessness	Caucasian	Female	Single
Income Maintenance	Caucasian	Female	Single
Individual Rights	Caucasian	Female	Single
Juvenile	Caucasian	Female	[Married / remarried, Single]
Other	Caucasian	Female	Single
Work, Employment and Unemployment	Caucasian	Female	Single

# EthnicIdentity

Remove category "Caucasian", which dominates each question category by frequency



- Many clients choose not to identify their ethnicity



# Let's Build a Model!

Based on these descriptive statistics, we select seven demographic features to predict question category:

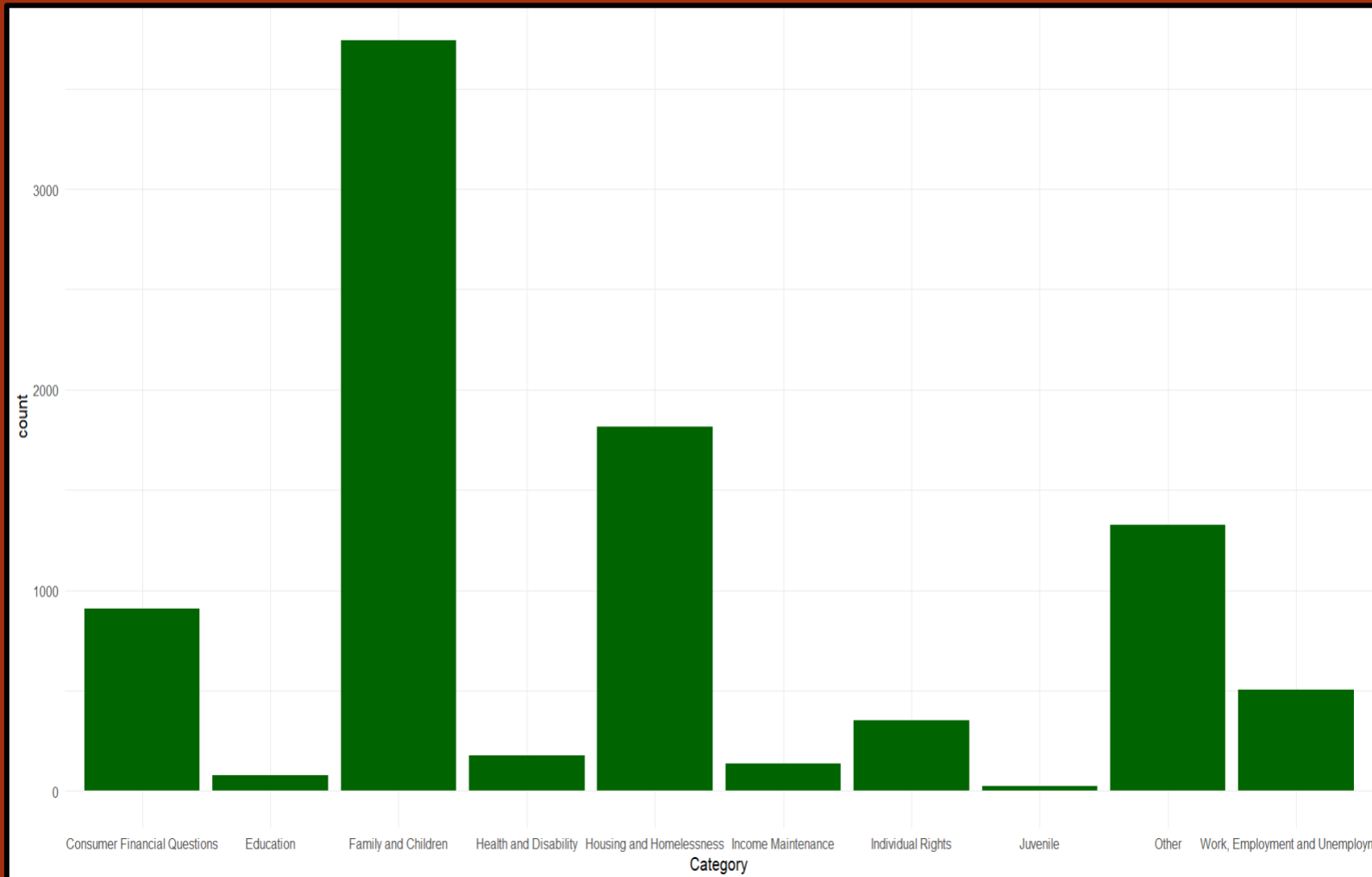
NumberInHousehold, Age, AnnualIncome, AllowedIncome, EthnicIdentity, MaritalStatus, Gender

---

The presence of many categorical variables leads us to select the model:

**Random Forest Classifier**

# Structure



- Two-layer model
- First layer: Classify "Family and Children" or not
- Second Layer: Classify into remaining categories
- Only first layer was completed

# Under the Hood

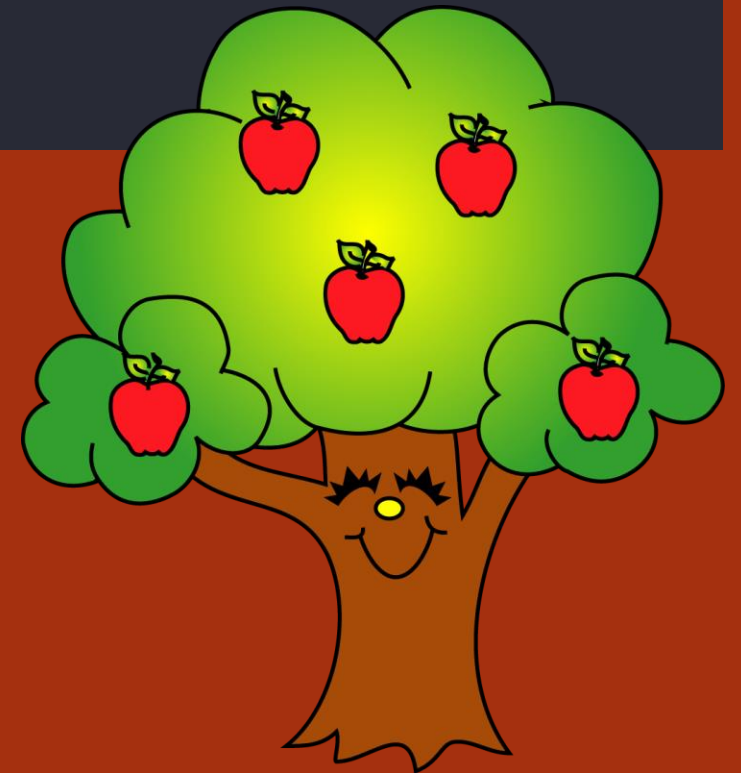
```
rf.mo1<-rfsrc(FnC~EthnicIdentity+Age+MaritalStatus+Gender+  
NumberInHousehold+AnnualIncome+AllowedIncome,  
data=mo.train)
```

```
predict(rf.mo1, mo.test)
```

Confusion matrix:

	predicted		
observed	0	1	class.error
0	1074	392	0.2674
1	508	527	0.4908

Misclassification error: 0.3598561



# Use Cases for the ABA

- Anticipate user questions before they are asked
- Determine how many specialized attorneys are needed per category
- Predict future trends in the demand for legal services

# ATTORNEY AND CLIENTS MATCHING

Table: Top 10 attorneys who are most experienced in Financial Questions

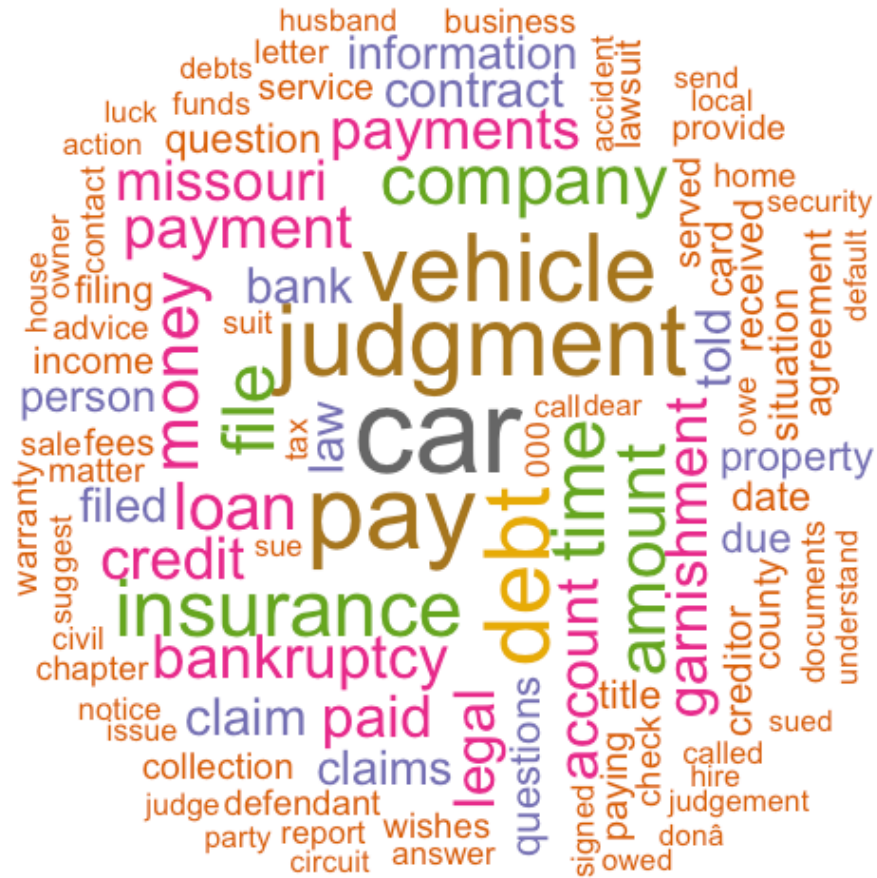
AttorneyUno	Category	n	County
05DD8A38-5C07-4796-8ED5-699E05424614	Consumer Financial Questions	285	St. Louis
735D4273-3EAC-4C2A-AF1C-F1BCFAFFEED8	Consumer Financial Questions	193	St. Louis
B89840FE-ECFE-4565-B7E1-9A2D8B6B170C	Consumer Financial Questions	108	St. Louis
390837A4-C502-402F-8E79-4CCF17E4E2BB	Consumer Financial Questions	48	St. Louis
62919210-9501-47E0-8CF3-D0B7527919C4	Consumer Financial Questions	39	St. Louis
8C3DE4A8-576C-455D-B3B2-9D6599DB5AC3	Consumer Financial Questions	33	St. Louis
03BA978D-3CA1-4D54-A125-828D3396C645	Consumer Financial Questions	28	St. Louis
3712EA2C-F291-481A-B045-D9E171D40563	Consumer Financial Questions	28	Jackson
7BEE59A5-23AF-40FB-B654-3751107A997B	Consumer Financial Questions	12	Clinton
E33AC9AF-D16F-49FC-83E2-A769F51E5F5D	Consumer Financial Questions	11	Perry



# WHY WORD CLOUD?

- Provide information to ABA so that they can;
  - anticipate the sorts of issues that arise
  - prepare volunteers to address those questions
  - know how to advise state partners on general trends they're seeing

# FINANCE



- In finance-related questions, clients were often seeking services for vehicle-payment-related problems
- Bankruptcy is another big topic questioned often in finance



# CONCLUSION

- Our focus was to provide insight into the data from both clients' and attorneys' perspectives
  - Match any given client with an attorney best suited to answer questions
  - Help to prepare attorneys by providing what types of questions are in demand as well as the trending topic within each category
  - Increase the accuracy and efficiency of their services in Missouri

## MOVING FORWARD.....

- Add layer(s) to Random Forest Classifier (improving model accuracy)
- performing an identical analysis of the selected state's data will help the ABA increase the accuracy and efficiency of their services

Thanks!