

Instructions

The American Bar Association (ABA) provides pro bono (that is, free of charge) legal services across the United States via an online platform available in some states and territories. The platform allows people who qualify (based on state-based low-income status) to post legal questions and receive legal advice from volunteer lawyers. The ABA would like to anticipate the sorts of legal questions that arise so that they can prepare volunteers to address those questions, better know how and when to recruit lawyers with specific expertise, and know how to advise state partners on general trends they are seeing. This information will help the ABA and their volunteer attorneys better adapt to new laws, societal changes, natural disasters, and other events that may cause an increase (and, in some cases, decrease) in the demand for pro bono legal services on particular legal topics.

Your data contains actual conversations between clients and lawyers (redacted to maintain confidentiality) and basic demographic information about the client. These exchanges have been previously classified into categories and sub-categories by the people posting questions; potentially, specific conversations might be sorted into different categories by platform administrators. Your challenge is to provide advice to the ABA concerning any patterns or trends you see in these conversations that would help the ABA advise its state partners, create resources to respond to identified patterns, and devise outreach strategies to better reach potential user populations and volunteers. For example, ABA would like to help its lawyers better “connect” with clients, which means they should use similar language to describe issues as their clients, should understand the cultural, societal and emotional context of the clients’ messages, and should understand how to best listen and engage with clients.

Data notes: The data you are provided is a full history of all complete conversations from August 5, 2016 through January 24, 2022. There might be some data from earlier periods before the program was fully launched. A conversation is marked as “closed” at the request of either the client, the attorney, or an administrator. Conversations are automatically marked as “closed” if there has been no communication for 10 days. Conversations may be closed because they have been completed, or because an administrator has decided it is not appropriate for this web site, or for many other reasons. If a potential client wishes to use

the services, they must provide personal information. This information is used to determine eligibility for participation. Eligibility criteria vary by state (or territory) and depends on income. Not all states may have provided data on eligibility criteria.

In addition: (1) Texts of posts in **questionposts.csv** will have commas in them, watch out when reading; (2) The data include a “state” labeled “US/ABA Federal”, which is used for questions concerning immigration or coming from veterans; and (3) The data do not provide an obvious means for identifying which posts are from the attorney and which from the client.

You are required to use R Markdown for all visualizations and fitting models to the data, **answer the following questions:**

1. Your work should include **at least 3 different visualizations** using using **ggplot2** and making use of **color, faceting, theme among others**. Each visualization must be accompanied by at least one paragraph of text. This text should include an interpretation of your visualization as well as what is interesting about your visualization. A strong visualization will be accompanied by text explaining what patterns or insights it helps us glean from the data. Each visualization should include a caption that fully explains how to understand your visualization (that is, explain all the components, legends, etc.). A good guideline is that someone who has not read your report should be able to look at just a visualization and its caption and fully understand what that visualization is showing.
2. Propose **one research question and fit an appropriate model to answer this research question**. Explain why you proposed this statistical method and carefully interpret your results.

Finally, your final report should be about 4-8 pages (maximum) and must include a one-page concluding/summary describing your main takeaways and what you learned through this challenge.