1. Intro
   1. Intro statistics (background info, summary stats (head, shape, etc.)
   2. Leads into data cleaning 🡪 our initial word cloud 🡪 bar plots 🡪 table
   3. Background research
2. Into Analysis
   1. Prompt 4 goes in intro analysis that predates our hypoth
   2. We analyze and “find” that religious texts originating from the same religion tend to have similar themes 🡪 quantify using background research and word count / tf idf SIMILILARITY
   3. From this we hypothesize that since these religions come from distinct geographical regions, we would expect that religious texts coming from different regions would be different from each other 🡪 quantify using background research and word count / tf idf differences 🡪 noun similarities are based on the fact that they are all religious text
   4. Geographical infographic and Heatmaps /histogram from barry etc.
   5. Hypothesis test and “find” that we’re right duh
3. Old Testament Question (using tf-idf/word count similarities and differences, barry’s random forest, and background research)
4. Buddhism and Taoism 🡪 from our hypoth we should find that these are different 🡪BUT we can include research that shows that Buddhism has extended beyond india and say that based on our hypothesis findings that even though this is the case, we would expect Chinese Buddhism to be more similar to Taoism 🡪 we can justify this using background research comparison of Chinese Buddhism and Taoism 🡪 written year or chronological info important