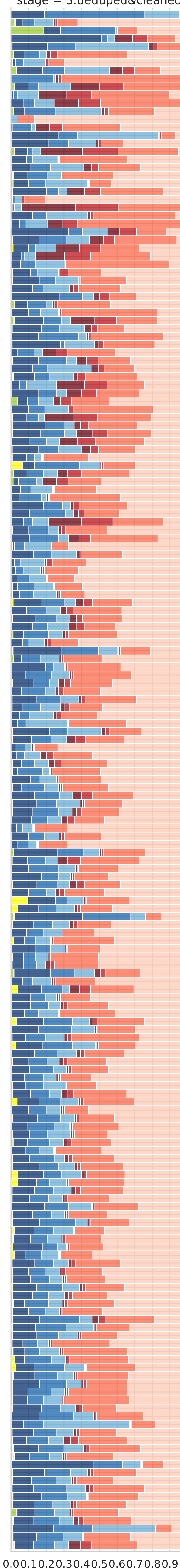
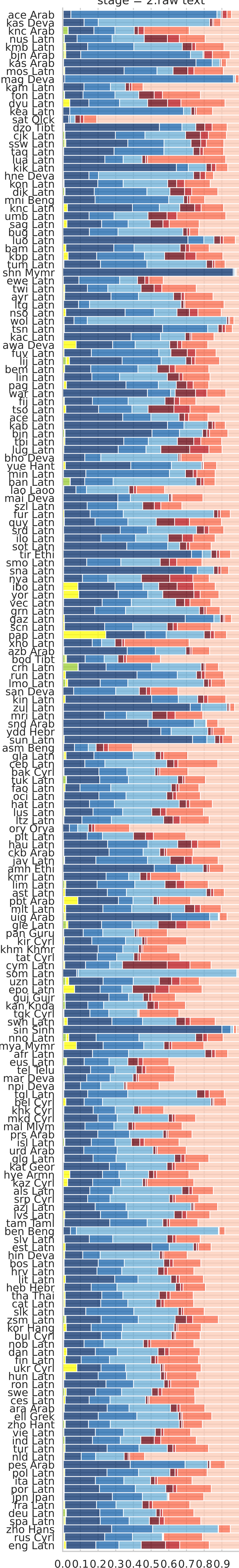


stage = 2:raw text

stage = 3:deduped&cleaned

language



group

cc_r

cc_n

cc_m

cc_o

ia_n

ia_m

ia_o

ia_archivebot

ia_survey