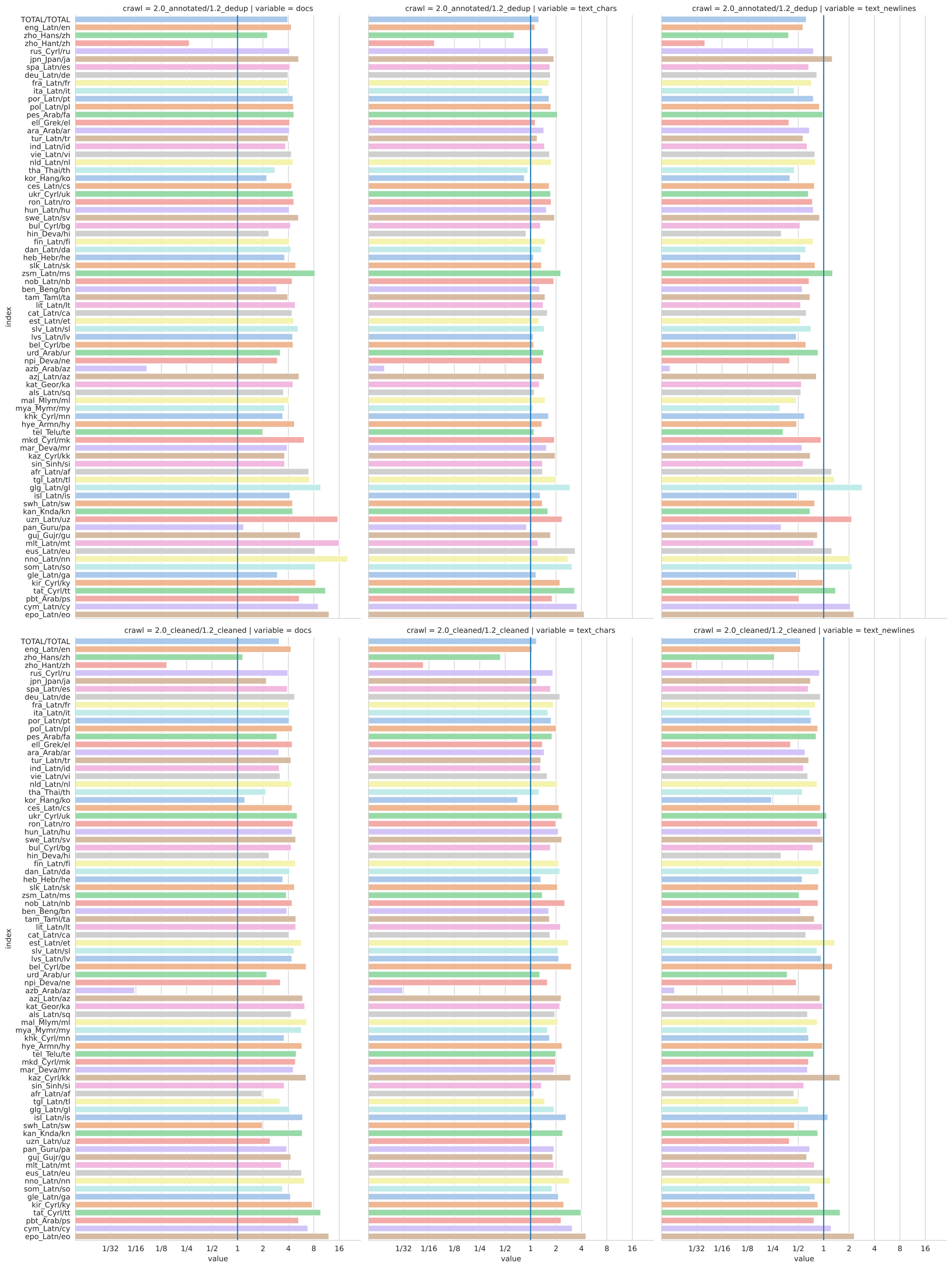
crawl = 2.0_annotated/1.2_dedup | variable = docs
crawl = 2.0_annotated/1.2_dedup | variable = text_chars
crawl = 2.0_annotated/1.2_dedup | variable = text_newlines
crawl = 2.0_cleaned/1.2_cleaned | variable = docs
crawl = 2.0_cleaned/1.2_cleaned | variable = text_chars
crawl = 2.0_cleaned/1.2_cleaned | variable = text_newlines
index
TOTAL/TOTAL
eng_Latn/en
zho_Hans/zh
zho_Hant/zh
rus_Cyrl/ru
jpn_Jpan/ja
spa_Latn/es
deu_Latn/de
fra_Latn/fr
ita_Latn/it
por_Latn/pt
pol_Latn/pl
pes_Arab/fa
ell_Grek/el
ara_Arab/ar
tur_Latn/tr
ind_Latn/id
vie_Latn/vi
nld_Latn/nl
tha_Thai/th
kor_Hang/ko
ces_Latn/cs
ukr_Cyrl/uk
ron_Latn/ro
hun_Latn/hu
swe_Latn/sv
bul_Cyrl/bg
hin_Deva/hi
fin_Latn/fi
dan_Latn/da
heb_Hebr/he
slk_Latn/sk
zsm_Latn/ms
nob_Latn/nb
ben_Beng/bn
tam_Taml/ta
lit_Latn/lt
cat_Latn/ca
est_Latn/et
slv_Latn/sl
lvs_Latn/lv
bel_Cyrl/be
urd_Arab/ur
npi_Deva/ne
azb_Arab/az
azj_Latn/az
kat_Geor/ka
als_Latn/sq
mal_Mlym/ml
mya_Mymr/my
khk_Cyrl/mn
hye_Armn/hy
tel_Telu/te
mkd_Cyrl/mk
mar_Deva/mr
kaz_Cyrl/kk
sin_Sinh/si
afr_Latn/af
tgl_Latn/tl
glg_Latn/gl
isl_Latn/is
swh_Latn/sw
kan_Knda/kn
uzn_Latn/uz
pan_Guru/pa
guj_Gujr/gu
mlt_Latn/mt
eus_Latn/eu
nno_Latn/nn
som_Latn/so
gle_Latn/ga
kir_Cyrl/ky
tat_Cyrl/tt
pbt_Arab/ps
cym_Latn/cy
epo_Latn/eo
1/32
1/16
1/8
1/4
1/2
1
2
4
8
16
value