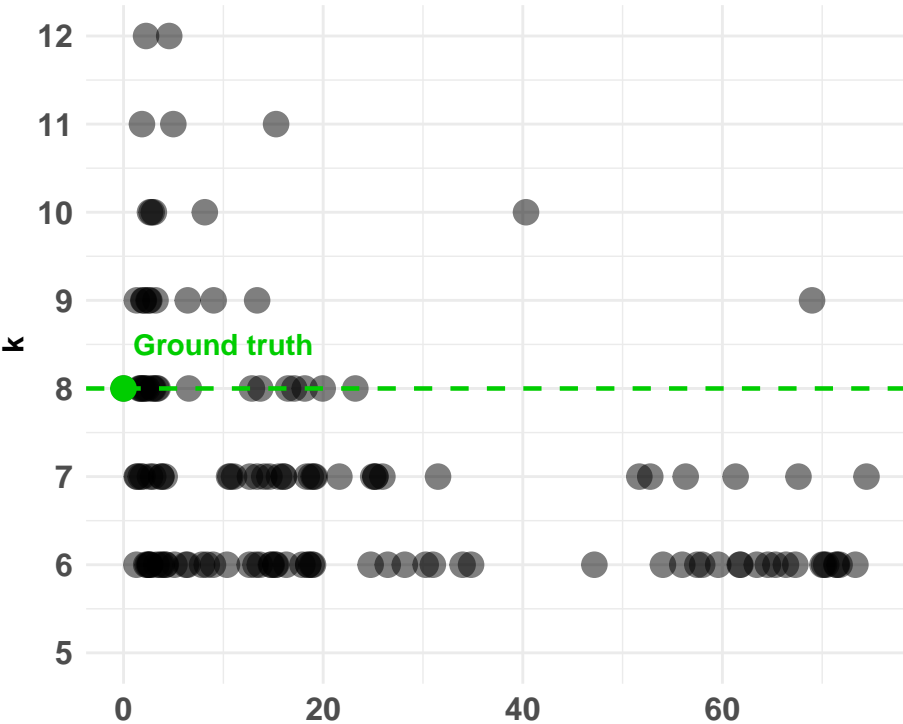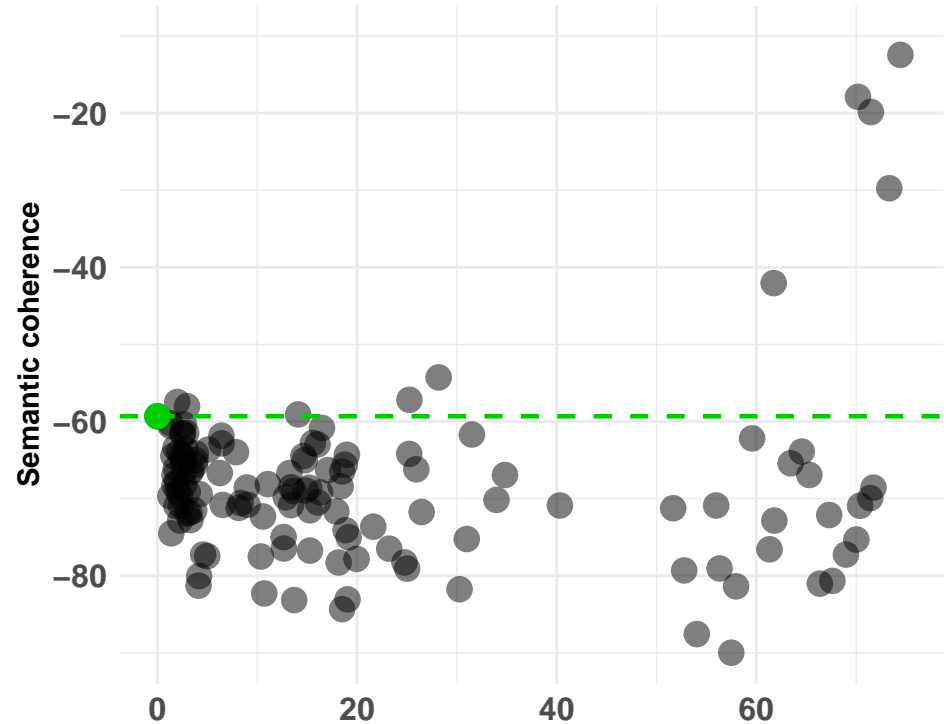Structural topic models of Old Books dataset, built with R package stm. Each point represents a noise/engine version (n=132) of the corpus. X axes cropped.

b) Suggested k values

(k intercepts for held−out likelihood high points.)

b) Semantic coherence (k=8)

Word error (percent)