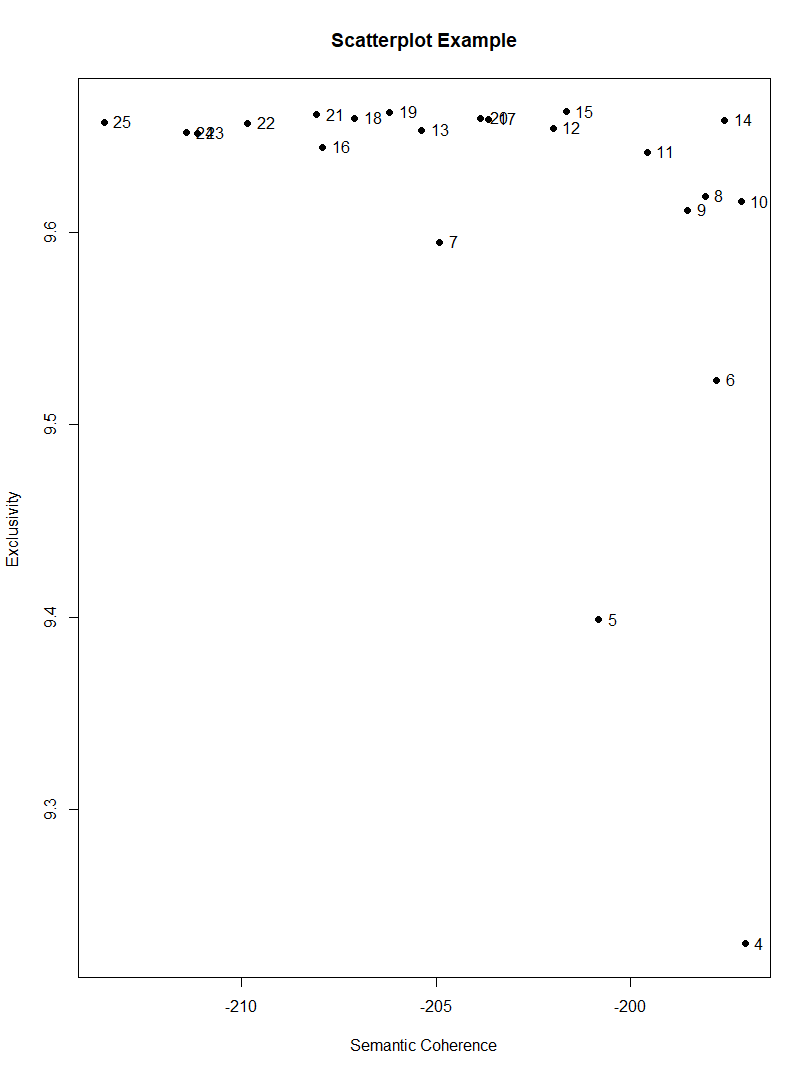
Assignment –TWO

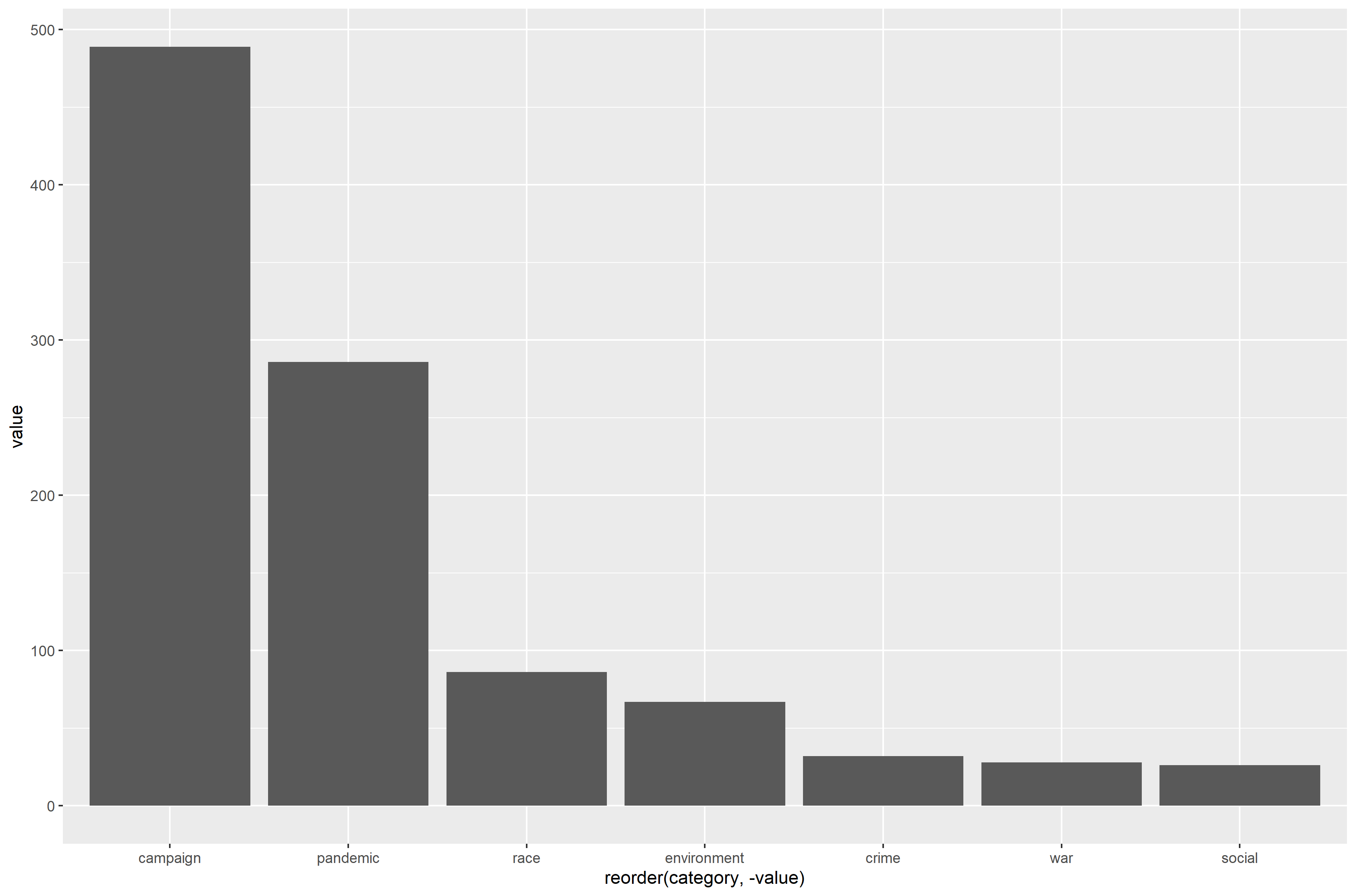
First, we ran an LDA and after estimated the optimal number of K for further STM calculations. As plotted below, K = 14 indicates the high exclusivity while still showing high values on the semantic coherence.



Results **without** topic prevalence covariates

Results **with** topic prevalence covariates

Assignment – THIRD PART



We analyzed the latest 1’500 tweets from Joe Biden. For semi-supervised classification we created a dictionary as stated in the code with the following categories: campaign, pandemic, race, environment, crime, war and social.

In total 990 tweets were assigned to the suggested categories, therefore in 510 tweets none of the words from our dictionary were found.

The tweets cover the time period from September 2020 until April 2021. Taking this into account it makes sense, that campaigning was a very prevalent topic. Obviously, the pandemic is also a key topic in general politics these days, therefore also in the tweets from Joe Biden.

As a critic of our dictionary would be our composition of the social dictionary words. Due to a lack of knowledge of specific key words in this topic, this category may be more prevalent that depicted in our graph.