bias_analysis

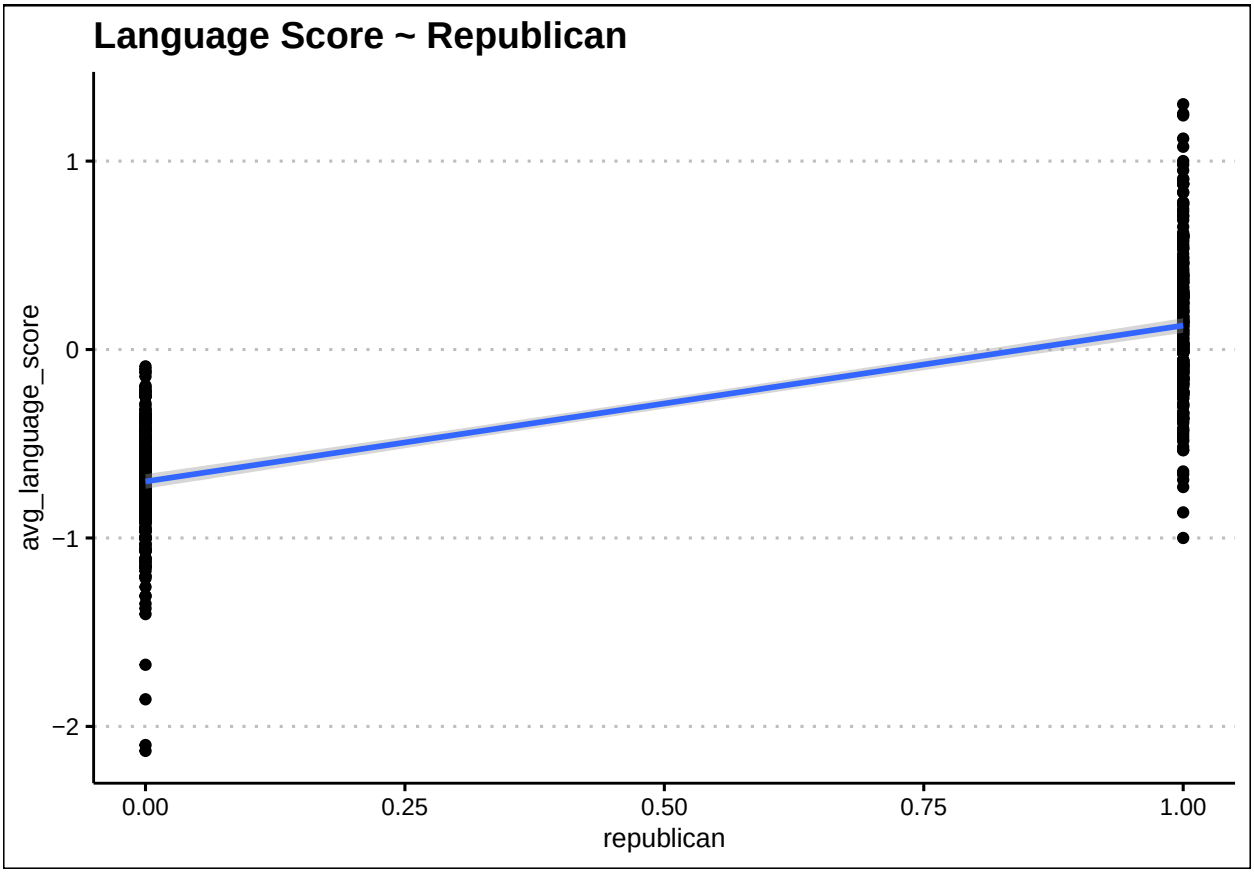
Lindsey Greenhill

12/8/2021

```
# I am now going to give each tweet a score

# dictionary from word dictionary analysis

content_dict <- dictionary(list(republican = c("biden",
                                               "border",
                                               "democrat",
                                               "spend",
                                               "illeg",
                                               "china",
                                               "inflat",
                                               "pelosi",
                                               "trillion",
                                               "afghanistan",
                                               "polici",
                                               "southern",
                                               "communist",
                                           "bidenbordercrisi",
                                           "radic",
                                           "mandat",
                                           "joe",
                                           "dem",
                                           "taxpay",
                                           "socialist"),
                                democrat = c("black",
                                             "work",
                                             "payment",
                                             "violenc",
                                             "vote",
                                             "act",
                                             "childtaxcredit",
                                             "pandem",
                                             "democraci",
                                             "communiti",
                                             "famili",
                                             "child",
                                             "help",
                                             "buildbackbett",
                                             "pass",
                                             "care",
                                             "invest",
                                             "climat",
```

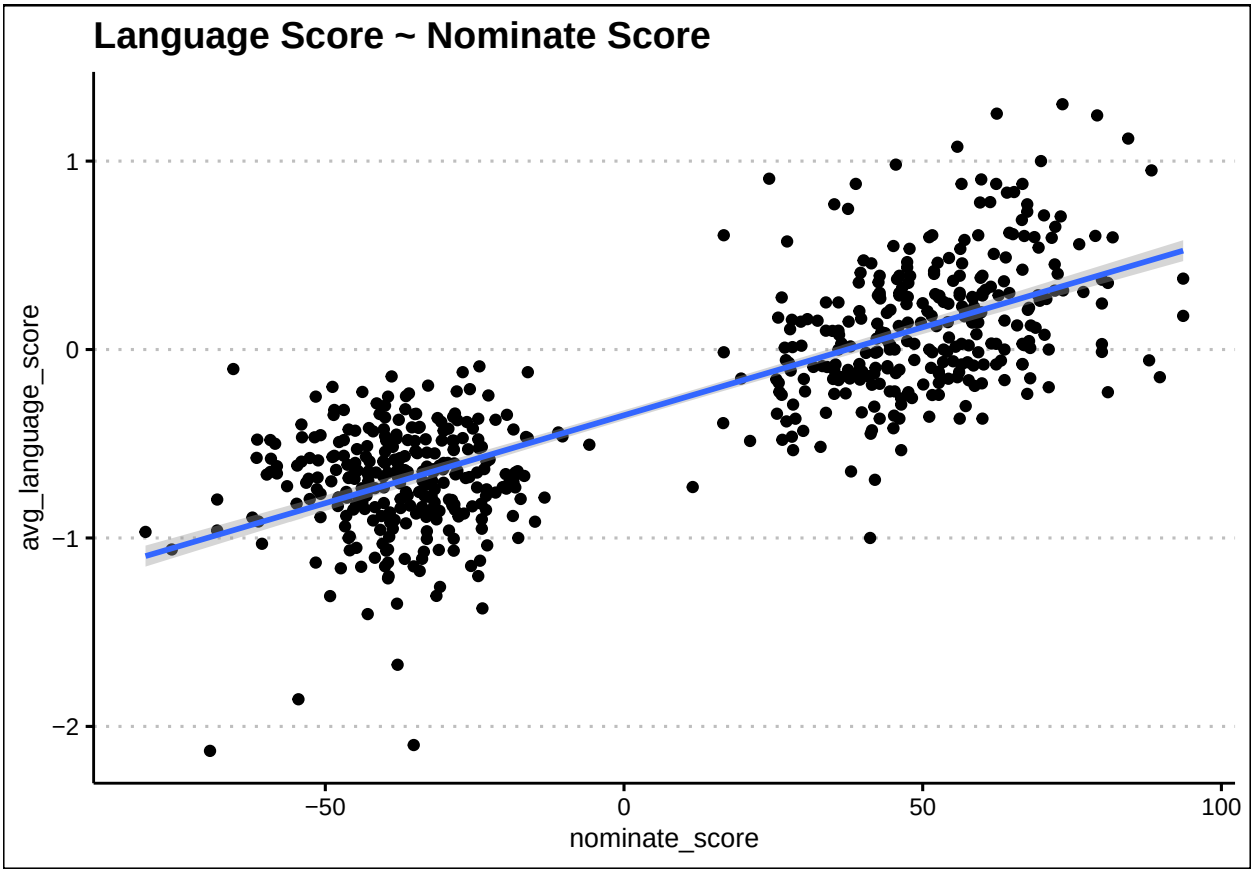
```
                                        "americanrescueplan",
                                        "health")))
```

```
# creating text corpus

text_corpus <- corpus(joined_data, text_field = "text")

# creating dfm for cocntent analysis

content_toks <- tokens(text_corpus,
               remove_punct = TRUE,
               remove_symbols = TRUE,
               remove_numbers = TRUE,
               remove_url = TRUE) %>%
  tokens_tolower() %>%
  tokens_remove(pattern=stopwords("en")) %>%
  tokens_select(min_nchar = 3)

content_dfm <- dfm(content_toks, groups = c("name", "date", "party"))

# selecting words in the dictionaries

content_categories <- dfm_lookup(content_dfm, dictionary = content_dict)

# turning dfm into dataframe

content_df <- convert(content_categories, to = "data.frame")

# changing the quanteda object into a data frame

content_df_cleaned <- content_df %>%
   mutate(party = substr(doc_id, start = str_length(doc_id), stop = str_length(doc_id)),
          name = substr(doc_id, start = 1, stop = str_length(doc_id) - 13))
```




```
## Warning: Removed 7 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 7 rows containing missing values (geom_point).
```



% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, Dec 08, 2021 - 15:42:26

Table 1:

	Dependent variable:	
	avg_language_score	
	(1)	(2)
republican	0.828***	
	(0.028)	
nominate_score		0.009***
		(0.0003)
Constant	−0.700***	−0.349***
	(0.020)	(0.013)
Observations	589	582
R^2	0.600	0.645
Adjusted R^2	0.599	0.644
Residual Std. Error	0.339 (df = 587)	0.320 (df = 580)
F Statistic	879.783*** (df = 1; 587)	1,053.709*** (df = 1; 580)
Note:		*p<0.1; **p<0.05; ***p<0.01