```python
nt = pd.read_csv('/Users/indre/Downloads/archive/nyt-articles-2020.csv')
```
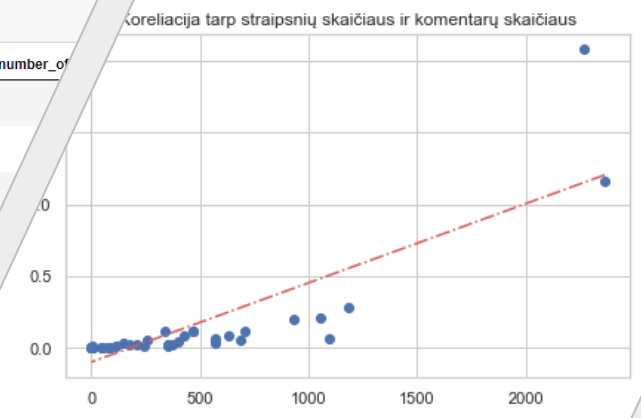
```python
nt = nt.rename(columns={'n_comments':'number_of_comments'})
```

```python
nt['pub_date'] = pd.to_datetime(nt['pub_date']).dt.date
```

```python
nt_new = nt[[ 'section', 'headline', 'abstract', 'keywords', 'pub_date', 'number_of_comments']]
nt_new
```

	section	headline	abstract	keywords	pub_date	number_of
0	Opinion	Protect Veterans From Fraud	Congress could do much more to protect America...	['Veterans', 'For-Profit Schools', 'Financial ...	2020-01-01	
1	Crosswords & Games	'It's Green and Slimy'	Christina Iverson and Jeff Chen ring in the Ne...	['Crossword Puzzles']	2020-01-01	
2	Science	Meteor Showers in 2020 That Will Light Up Nigh...	All year long, Earth passes through streams of...	['Meteors and Meteorites', 'Space and Astronom...	2020-01-01	
3	Science	Sync your calendar with the solar system	Never miss an eclipse, a meteor shower, a rock...	['Space and Astronomy', 'Moon', 'Eclipses', 'S...	2020-01-01	
4	Science	Rocket Launches, Trips to Mars and More 2020 S...	A year full of highs and lows in space just en...	['Space and Astronomy', 'Private Spaceflight', ...	2020-0	
...	...	...	...	...		
16782	Opinion	What It Takes to Heal From Covid-19	Survivors can get better, but they need help.	['Chronic Condition (Health)', 'Coronavirus (2...		
16783	Sports	Padres Jolt M.L.B. With Bold Moves to Set Up W...	While many teams continued to assess the finan...	['San Diego Padres', 'Major League Baseball', ...		
16784	Business Day	Their Finances Ravaged, Customers Fear Banks W...	Banks have the power to decide whether to let ...	['Banking and Financial Institution', 'Coron...		
16785	Food	Should Wine Be Among Your Health Resolutions?	The new category of 'clean wines' is an effort...	['Wines', 'Grapes', 'Die... Nutrition'...		
16786	Technology	Microsoft Says Russian Hackers	The hackers gained more access	['Microsoft Corp', 'US...		

```python
from matplotlib_venn import venn2
plt.figure(figsize=(12,9))
set1 = set(frequent_words_bysections['word_us'])
set2 = set(frequent_words_bysections['word_business'])

plt.title('U.S. ir The Business sekcijų keywords palyginimas')

venn = venn2([set1, set2])
venn.get_label_by_id('100').set_text('\n'.join(map(str,set1-set2)))
venn.get_label_by_id('110').set_text('\n'.join(map(str,set1&set2)))
venn.get_label_by_id('010').set_text('\n'.join(map(str,set2-set1)))
# venn.get_label is quoted from https://stackoverflow.com/questions/55717203/plot-actual-set-item
```

prediction, 'r', linestyle='-.')
layout()
('Koreliacija tarp straipsnių skaičiaus ir komentarų skaič

.0, 'Koreliacija tarp straipsnių skaičiaus ir komentarų skaič

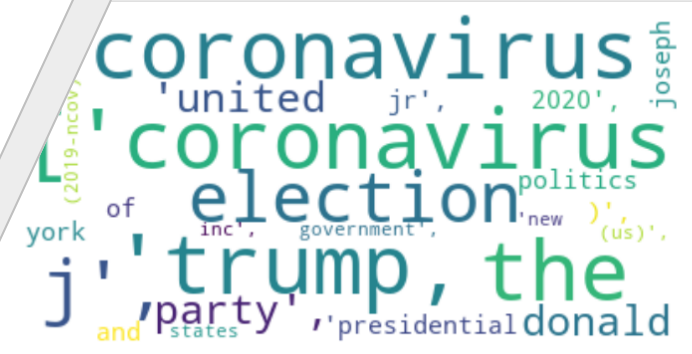Koreliacija tarp straipsnių skaičiaus ir komentarų skaičiaus

3.0.1 Išvada – koreliacija silpna, ant strapsniu section arba keywords :)

dCloud, STOPWORDS
as plt
sections['word_all']
.index.map(str)
().generate_from_frequencies(counts)
(12,9)
oud(max_font_size=50, max_words=100, background_color="white").generate_from_freque
et(STOPWORDS)
date(["corona", "trump", "2020",])
WordCloud(stopwords=stopwords, background_color="white").generate_from_frequencies(coun
ordcloud, interpolation='bilinear')
ff")