Dictionary Analysis

Riccardo Ruta

5/2022

Contents

RICORSO ALLA RETORICA POPULISTA.

At the level of political parties, which ones make most use of populist rhetoric?

At the level of individual politicians, which ones make most use of populist rhetoric?

- Rooduijn & Pauwels: Rooduijn, M., and T. Pauwels. 2011. "Measuring Populism: Comparing Two Methods of Content Analysis." West European Politics 34 (6): 1272–1283.

- Decadri & Boussalis: Decadri, S., & Boussalis, C. (2020). Populism, party membership, and language complexity in the Italian chamber of deputies. Journal of Elections, Public Opinion and Parties, 30(4), 484-503.
- Grundl: Gründl J. Populist ideas on social media: A dictionary-based measurement of populist communication. New Media & Society. December 2020.
- Decadri & Boussalis + Grundl: this is simply a more extended version of the D&B dictionary, which also contains some terms taken from Grundl.

1) First step, import the words and create the dictionary

```
# import dictionaries file
dict <-  read_excel("data/populism_dictionaries.xlsx")
variable.names(dict)
```

```
## [1] "Rooduijn_Pauwels_Italian"
## [2] "Grundl_Italian_adapted"
## [3] "Decadri_Boussalis"
## [4] "Decadri_Boussalis_Grundl_People"
## [5] "Decadri_Boussalis_Grundl_Common Will"
## [6] "Decadri_Boussalis_Grundl_Elite"
```

```
# create the dictionary
Rooduijn_Pauwels_Italian <-
  dictionary(list(populism =
                    (dict$Rooduijn_Pauwels_Italian
                     [!is.na(dict$Rooduijn_Pauwels_Italian)])))

Grundl_Italian_adapted <-
  dictionary(list(populism =
                    dict$Grundl_Italian_adapted
                  [!is.na(dict$Grundl_Italian_adapted)]))

Decadri_Boussalis <-
  dictionary(list(populism =
                    dict$Decadri_Boussalis
                  [!is.na(dict$Decadri_Boussalis)]))

Decadri_Boussalis_Grundl <-
  dictionary(list(people =
                    dict$Decadri_Boussalis_Grundl_People
                  [!is.na(dict$Decadri_Boussalis_Grundl_People)],
                  common_will =
                    dict$`Decadri_Boussalis_Grundl_Common Will`
                  [!is.na(dict$`Decadri_Boussalis_Grundl_Common Will`)],
                  elite =
                    dict$Decadri_Boussalis_Grundl_Elite
                  [!is.na(dict$Decadri_Boussalis_Grundl_Elite)]))
```

2) Import the DFM prepared in previous steps and apply dictionaries

I) Decadri_Boussalis_Grundl

Level of sparsity

daily: 12.08%

weekly: 0.55%

monthly: 0%

```
# Daily Dictionary analysis with Decadri_Boussalis_Grundl on the whole dataset
dfm_dict1  <- dfm_lookup(dfm_weight, dictionary = Decadri_Boussalis_Grundl)
# Group by date
dfm_by_date1 <- dfm_group(dfm_dict1, groups= date)
dfm_by_date1
```

```
## Document-feature matrix of: 839 documents, 3 features (12.08% sparse) and 3 docvars.
##             features
## docs              people common_will      elite
##   2020-01-01 0.04761905  0          0.06666667
##   2020-01-02 0.42636838  0.05882353 0.39453879
##   2020-01-03 0.57511141  0.25000000 0.53910534
##   2020-01-04 1.35115485  0          0.28968254
##   2020-01-05 1.94643434  0.08333333 0.23852814
##   2020-01-06 0.75485528  0          1.16776316
## [ reached max_ndoc ... 833 more documents ]
```

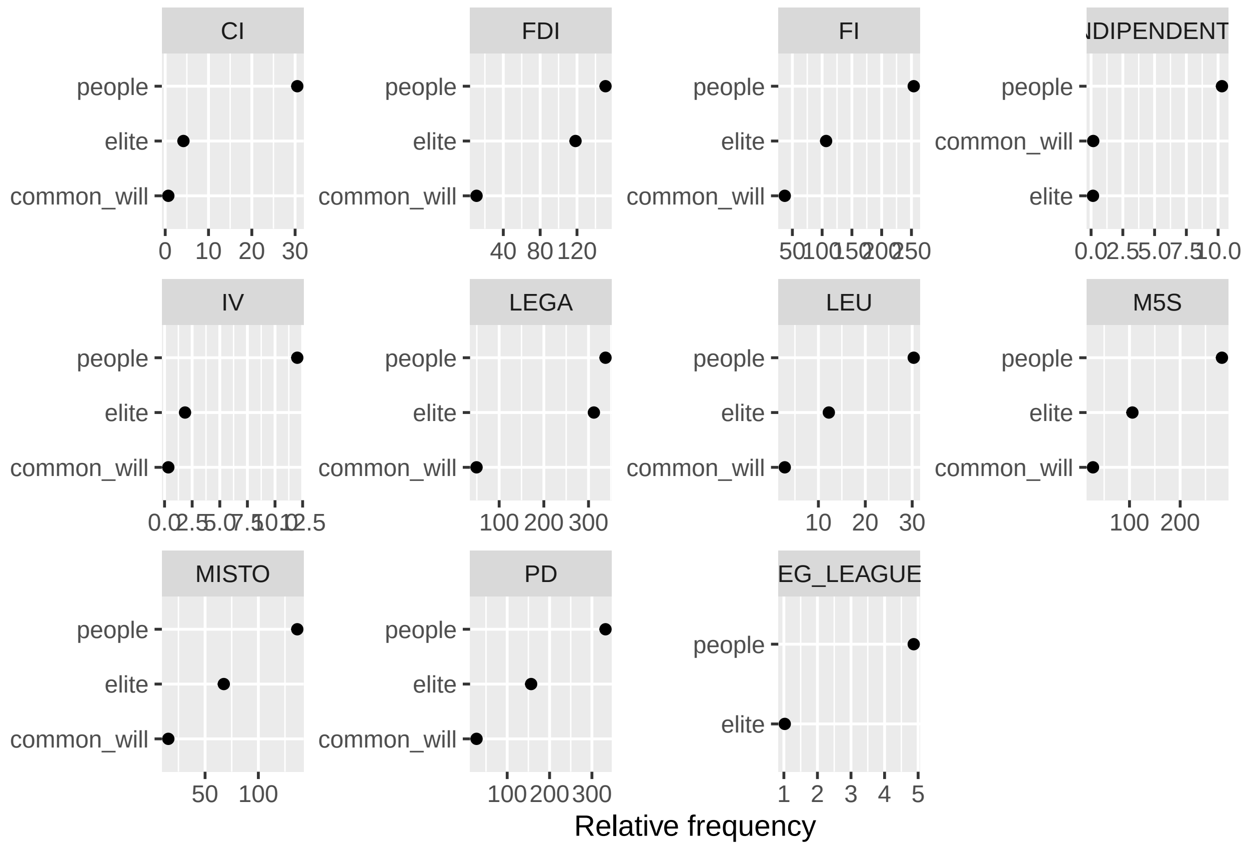
```
# Group by week
dfm_by_week1 <- dfm_group(dfm_dict1, groups= week)
dfm_by_week1
```

```
## Document-feature matrix of: 121 documents, 3 features (0.55% sparse) and 1 docvar.
##     features
## docs    people common_will     elite
##    1  4.346688   0.3921569  1.528521
##    2 10.917519   2.1764234  5.389965
##    3 16.217653   1.9467875  7.994374
##    4 17.996818   1.3673553  8.399551
##    5 17.661895   0.9670610  6.979555
##    6 11.257576   0.7009804 10.688212
## [ reached max_ndoc ... 115 more documents ]
```

```
# Group by month
dfm_by_month1 <- dfm_group(dfm_dict1, groups= month)
dfm_by_month1
```

```
## Document-feature matrix of: 28 documents, 3 features (0.00% sparse) and 1 docvar.
##     features
```

```
## docs   people common_will    elite
##    1 63.08421    6.720617 30.07071
##    2 51.95882    3.532448 36.73581
##    3 59.69107    3.065409 26.04866
##    4 51.97619    1.975715 38.92381
##    5 49.45054    1.012790 35.60162
##    6 43.57187    1.785750 39.61004
## [ reached max_ndoc ... 22 more documents ]
```



Looking at the populist rhetoric for each party divided into the 3 components people-centrism, anti-elitism and common-will, we note that the most frequent components is People-centrism.

II) Rooduijn_Pauwels_Italian

Level of sparsity

daily: 0.60%

weekly: 0.%

monthly: 0%

```
# Daily Dictionary analysis with Rooduijn_Pauwels_Italian on the whole dataset
dfm_dict2  <- dfm_lookup(dfm_weight, dictionary = Rooduijn_Pauwels_Italian)
# Group by date
```

```r
dfm_by_date2 <- dfm_group(dfm_dict2, groups= date)
dfm_by_date2
```

```
## Document-feature matrix of: 839 documents, 1 feature (0.60% sparse) and 3 docvars.
##             features
## docs           populism
##   2020-01-01 0.06666667
##   2020-01-02 0.34691975
##   2020-01-03 0.53910534
##   2020-01-04 0.28968254
##   2020-01-05 0.23852814
##   2020-01-06 1.16776316
## [ reached max_ndoc ... 833 more documents ]
```

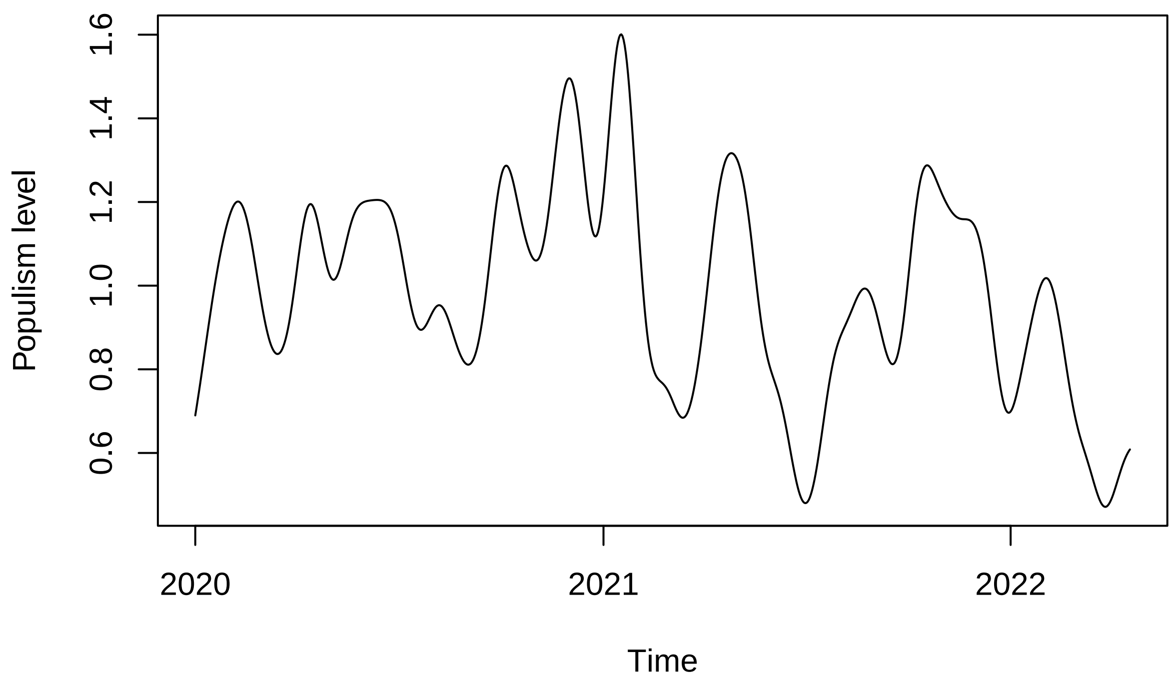
```r
# Group by week
dfm_by_week2 <- dfm_group(dfm_dict2, groups= week)
dfm_by_week2
```

```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs  populism
##    1  1.480902
##    2  5.140665
##    3  7.493528
##    4  7.872910
##    5  6.339158
##    6 10.353899
## [ reached max_ndoc ... 115 more documents ]
```

```r
# Group by month
dfm_by_month2 <- dfm_group(dfm_dict2, groups= month)
dfm_by_month2
```

```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1 28.10591
##    2 34.76596
##    3 24.91863
##    4 37.43421
##    5 32.79228
##    6 37.74417
## [ reached max_ndoc ... 22 more documents ]
```

General level of populism in time



Most populist party

```r
# Most populist party
dfm_dict2_tstat_party <- textstat_frequency(dfm_dict2, groups = party_id)
kable(dfm_dict2_tstat_party %>% slice_max(frequency, n = 20))
```

	feature	frequency	rank	docfreq	group
6	populism	303.9474786	1	1919	LEGA
10	populism	149.7512641	1	1671	PD
2	populism	113.7388243	1	1124	FDI
3	populism	98.6906136	1	941	FI
8	populism	87.6625041	1	1119	M5S
9	populism	60.9720255	1	669	MISTO
7	populism	11.7023384	1	175	LEU
1	populism	3.7116701	1	45	CI
5	populism	1.8540424	1	26	IV
11	populism	1.0264294	1	11	REG_LEAGUES
4	populism	0.0833333	1	1	INDIPENDENTE

Most populist politician

```
dict2_tstat_nome <- textstat_frequency(dfm_dict2, groups = nome)

kable(dict2_tstat_nome %>% slice_max(frequency, n = 20))
```

	feature	frequency	rank	docfreq	group
194	populism	42.115152	1	146	FERRERO Roberta
472	populism	15.910436	1	160	SGARBI Vittorio
341	populism	14.112659	1	77	MORANI Alessia
24	populism	13.999694	1	52	BALDELLI Simone
179	populism	13.821584	1	48	FAGGI Antonella
271	populism	13.095709	1	149	LANNUTTI Elio
217	populism	12.884799	1	39	FREGOLENT Sonia
450	populism	12.806346	1	64	RUSPANDINI Massimo
326	populism	12.518396	1	192	MELONI Giorgia
427	populism	12.257891	1	40	RIVOLTA Erica
106	populism	10.788399	1	68	CECCHETTI Fabrizio
283	populism	10.783981	1	108	LOLLOBRIGIDA Francesco
260	populism	10.778644	1	76	IEZZI Igor Giancarlo
230	populism	10.648954	1	155	GARNERO SANTANCHE' Daniela
303	populism	10.133849	1	78	MALAN Lucio
447	populism	9.885108	1	29	RUFA Gianfranco
455	populism	9.561830	1	93	SALVINI Matteo
360	populism	9.110910	1	105	NOBILI Luciano
35	populism	8.689617	1	57	BAZZARO Alex
501	populism	8.495460	1	32	TONELLI Gianni

III) Grundl_Italian_adapted

Level of sparsity

daily: 0.24%

weekly: 0.0%

monthly: 0%

```
# Daily Dictionary analysis with Grundl_Italian_adapted on the whole dataset
dfm_dict3  <- dfm_lookup(dfm_weight, dictionary = Grundl_Italian_adapted)
# Group by date
dfm_by_date3<- dfm_group(dfm_dict3, groups= date)
dfm_by_date3
```

```
## Document-feature matrix of: 839 documents, 1 feature (0.24% sparse) and 3 docvars.
##             features
## docs           populism
##   2020-01-01 0
##   2020-01-02 0.16894258
##   2020-01-03 0.34545455
##   2020-01-04 0.08333333
```

```
##   2020-01-05 1.49644922
##   2020-01-06 1.64472860
## [ reached max_ndoc ... 833 more documents ]
```

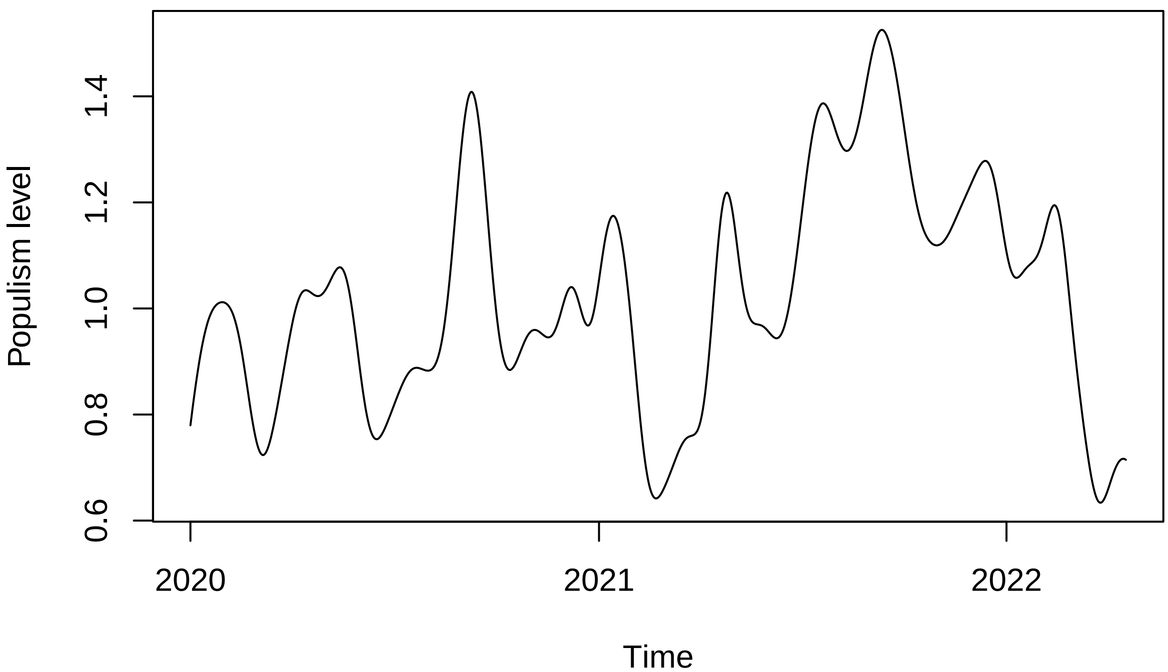
```r
# Group by week
dfm_by_week3 <- dfm_group(dfm_dict3, groups= week)
dfm_by_week3
```

```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1 2.094180
##    2 7.732182
##    3 7.349727
##    4 6.090898
##    5 7.887386
##    6 7.021892
## [ reached max_ndoc ... 115 more documents ]
```

```r
# Group by month
dfm_by_month3 <- dfm_group(dfm_dict3, groups= month)
dfm_by_month3
```

```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1 30.09665
##    2 26.23980
##    3 22.99661
##    4 32.36833
##    5 33.50214
##    6 21.44168
## [ reached max_ndoc ... 22 more documents ]
```

General level of populism in time



Most populist party

```r
# Most populist party
dict_3_tstat_party <- textstat_frequency(dfm_dict3, groups = party_id)
kable(dict_3_tstat_party %>% slice_max(frequency, n = 20))
```

	feature	frequency	rank	docfreq	group
6	populism	225.678708	1	2075	LEGA
10	populism	153.269683	1	2017	PD
8	populism	133.053746	1	1724	M5S
3	populism	131.838292	1	1524	FI
2	populism	99.425177	1	1087	FDI
9	populism	86.092041	1	997	MISTO
7	populism	15.213765	1	231	LEU
1	populism	10.602522	1	157	CI
5	populism	2.559005	1	40	IV
4	populism	1.983671	1	31	INDIPENDENTE
11	populism	1.505044	1	22	REG_LEAGUES

Most populist politician

```
dict_3_tstat_nome <- textstat_frequency(dfm_dict3, groups = nome)

kable(dict_3_tstat_nome %>% slice_max(frequency, n = 20))
```

	feature	frequency	rank	docfreq	group
287	populism	23.033031	1	240	LANNUTTI Elio
210	populism	19.501980	1	110	FERRERO Roberta
562	populism	19.042283	1	131	VITO Elio
275	populism	16.483870	1	120	IEZZI Igor Giancarlo
494	populism	15.974269	1	184	SGARBI Vittorio
341	populism	11.063928	1	159	MELONI Giorgia
15	populism	10.731212	1	120	ANZALDI Michele
298	populism	10.659433	1	98	LOLLOBRIGIDA Francesco
74	populism	10.645964	1	97	BORGHI Claudio
476	populism	9.238862	1	122	SALVINI Matteo
248	populism	9.004085	1	139	GARNERO SANTANCHE' Daniela
96	populism	8.438949	1	103	CANGINI Andrea
546	populism	8.339166	1	106	URSO Adolfo
224	populism	8.162373	1	101	FONTANA Lorenzo
472	populism	7.850014	1	68	RUSPANDINI Massimo
44	populism	7.832168	1	120	BERGESIO Giorgio Maria
165	populism	7.565932	1	92	DE MARTINI Guido
141	populism	7.036558	1	43	CROSETTO Guido
446	populism	7.000320	1	47	RIVOLTA Erica
359	populism	6.861311	1	73	MORELLI Alessandro

IV) Decadri_Boussalis

Level of sparsity

daily: 0%

weekly: 0.0%

monthly: 0%

```
# Daily Dictionary analysis with Decadri_Boussalis on the whole dataset
dfm_dict4  <- dfm_lookup(dfm_weight, dictionary = Decadri_Boussalis)
# Group by date
dfm_by_date4<- dfm_group(dfm_dict4, groups= date)
dfm_by_date4
```

```
## Document-feature matrix of: 839 documents, 1 feature (0.00% sparse) and 3 docvars.
##             features
## docs          populism
##   2020-01-01 0.1142857
##   2020-01-02 0.8566215
##   2020-01-03 1.2364390
##   2020-01-04 1.6408374
```

```
##   2020-01-05 2.2375941
##   2020-01-06 1.9814420
## [ reached max_ndoc ... 833 more documents ]
```

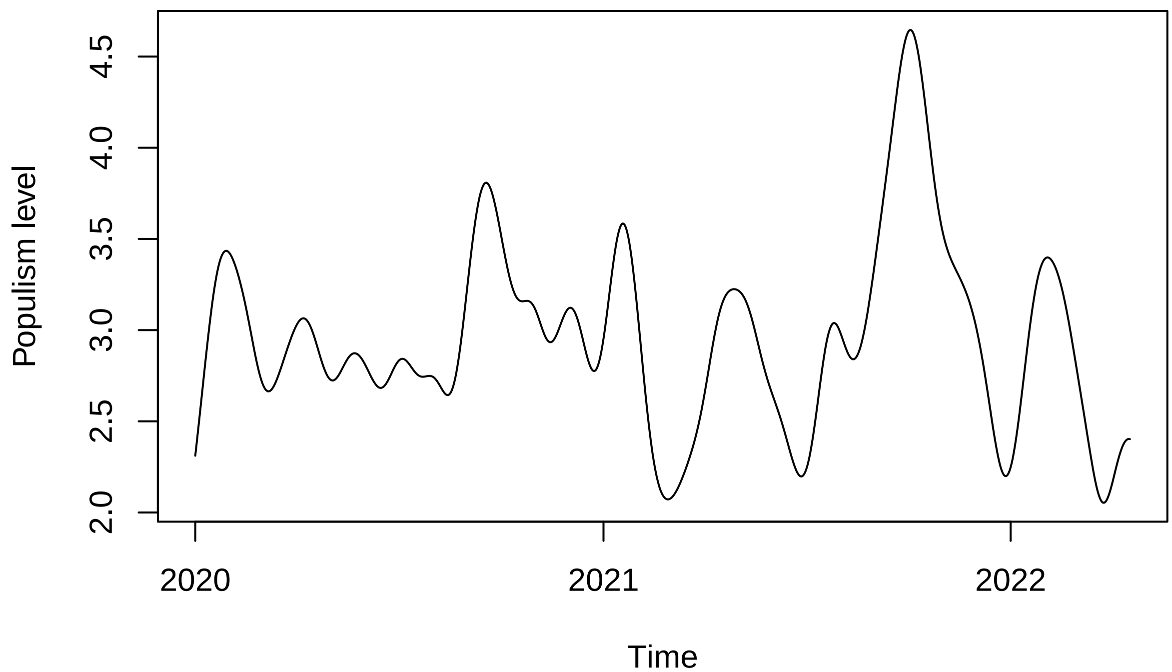
```r
# Group by week
dfm_by_week4 <- dfm_group(dfm_dict4, groups= week)
dfm_by_week4
```

```
## Document-feature matrix of: 121 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs  populism
##    1  6.085778
##    2 16.762233
##    3 24.352100
##    4 26.321730
##    5 24.812520
##    6 22.284239
## [ reached max_ndoc ... 115 more documents ]
```

```r
# Group by month
dfm_by_month4 <- dfm_group(dfm_dict4, groups= month)
dfm_by_month4
```

```
## Document-feature matrix of: 28 documents, 1 feature (0.00% sparse) and 1 docvar.
##     features
## docs populism
##    1 93.79618
##    2 88.79620
##    3 85.17899
##    4 90.99191
##    5 83.84470
##    6 82.69573
## [ reached max_ndoc ... 22 more documents ]
```

General level of populism in time



Most populist party

```r
# Most populist party
dict_4_tstat_party <- textstat_frequency(dfm_dict4, groups = party_id)
kable(dict_4_tstat_party %>% slice_max(frequency, n = 20))
```

	feature	frequency	rank	docfreq	group
6	populism	651.348390	1	5672	LEGA
10	populism	493.532735	1	6417	PD
8	populism	376.966170	1	5178	M5S
3	populism	376.609606	1	4532	FI
2	populism	270.814483	1	2960	FDI
9	populism	202.466904	1	2463	MISTO
7	populism	44.919508	1	659	LEU
1	populism	35.105322	1	506	CI
5	populism	14.132863	1	197	IV
4	populism	10.615825	1	153	INDIPENDENTE
11	populism	6.122696	1	93	REG_LEAGUES

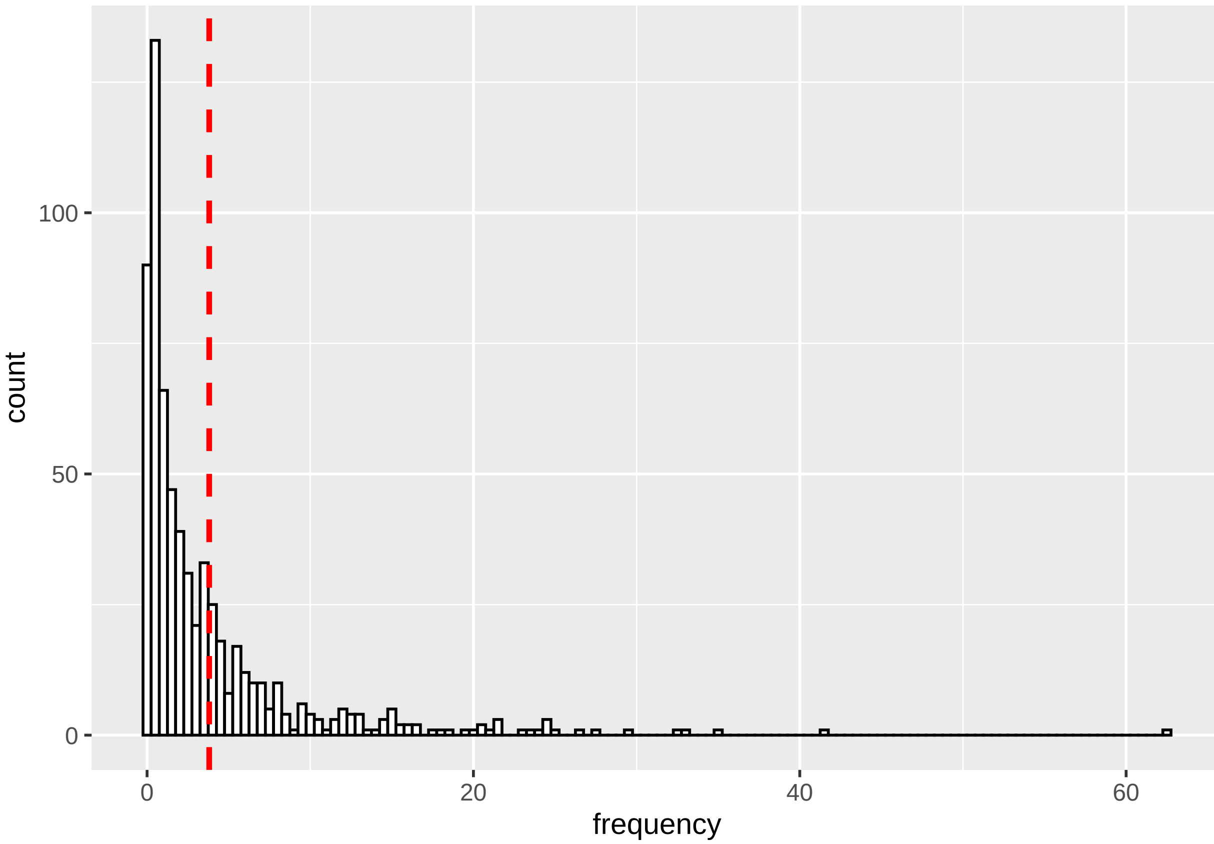
Most populist politician

```
dict_4_tstat_nome <- textstat_frequency(dfm_dict4, groups = nome)

kable(dict_4_tstat_nome %>% slice_max(frequency, n = 20))
```

	feature	frequency	rank	docfreq	group
236	populism	62.66405	1	282	FERRERO Roberta
560	populism	41.70723	1	443	SGARBI Vittorio
329	populism	34.85565	1	397	LANNUTTI Elio
391	populism	33.15912	1	496	MELONI Giorgia
344	populism	32.36912	1	358	LOLLOBRIGIDA Francesco
540	populism	29.61242	1	368	SALVINI Matteo
27	populism	27.44810	1	135	BALDELLI Simone
280	populism	26.74696	1	372	GARNERO SANTANCHE' Daniela
530	populism	24.85093	1	184	ROTONDI Gianfranco
68	populism	24.50676	1	252	BONACCINI Stefano
220	populism	24.35617	1	122	FAGGI Antonella
317	populism	24.31241	1	207	IEZZI Igor Giancarlo
128	populism	23.82148	1	195	CECCHETTI Fabrizio
585	populism	23.63509	1	327	TAJANI Antonio
80	populism	22.82617	1	240	BORGHI Claudio
161	populism	21.54784	1	158	CROSETTO Guido
39	populism	21.35229	1	202	BAZZARO Alex
47	populism	21.29380	1	318	BERGESIO Giorgio Maria
535	populism	20.92822	1	140	RUSPANDINI Massimo
365	populism	20.38171	1	185	MALAN Lucio

```
# TEST
ggplot(dict_4_tstat_nome, aes(x=frequency)) +
  geom_histogram(binwidth=.5, colour="black", fill="white") +
    geom_vline(aes(xintercept=mean(frequency, na.rm=T)),   # Ignore NA values for mean
               color="red", linetype="dashed", size=1)
```


100
50
0
count
0
20
40
60
frequency