Has Politics Irrevocably Entered Academic Research?: An Investigation into the last 50 Years of Publications

May 3, 2023
STAT E-109

By Rachel Krasner, Jeff Schiesser,
Mohamad El-Khadra, and Nazaf Showkat

Introduction

There is an unmet need to investigate how academic journals over time have evolved. The broadening of academic research is of utmost importance to study given the considerable changes the last decade has brought. This paper aims to provide a comprehensive study on the analysis of how research interests have evolved in the past half century, specifically in terms of political science. We are interested to see if there is evidence that academic publications of most fields have become more politicized in the last decade than in previous ones. Please note the use of the word "politicized" which does not belong solely to any political party but to the open-ended arena of politics. By examining a curated list of politically driven terms, we use publication records to see how these terms map onto a variety of fields over time. In this analysis, we demonstrate a predictive inspection on the political terms over a given period. We hope to present our findings in a non-partisan, non-judgmental way, and the reader will ultimately determine the strengths and directionality of these findings.

Research Question and Hypothesis

Within the scope of this project, we aim to investigate the growth rate of chosen political terms (Figure 1) in comparison to an expected exponential growth model. From this, we seek to evaluate based on the null hypothesis that the usage of the terms does not exceed baseline growth of 5%. In order to reject the null hypothesis, the data must be statistically significant from the expected growth.

$\mathbf{H_0 <= 5\%}$, for each individual term

$\mathbf{H_A > 5\%}$, for each individual term

Data

In this project, we obtained an API key through S2ORC: The Semantic Scholar Open Research Corpus which contains "81.1M English-language academic papers spanning many academic disciplines."[1] This corpus contains data from 1970 up to the present. We want to see if the rate of growth per topic within a field exceeds the predicted baseline rate of growth for publications in academia overall using a baseline percentage from this reference: "Since 1952, science has grown exponentially without restrictions with an annual growth rate of 5.08% and a doubling time of 14.0 years." (Bornmann, L., Haunschild, R. & Mutz (2021).[2] We round the annual growth rate down to a conservative 5% for comparison in our models.

We used a custom Python based program to pull word frequencies by year as well as by discipline from the raw JSON data provided by S2ORC. This first pass amounts to 2,690,470 papers. The data are then stratified by chosen political terms which we outline below in Figure 1.

The Future of Democracy	**Education**	**Abortion**	**Immigration**	**Foreign Policy**	**Issues around race and ethnicity**
Conservative Progressivism Nationalism Patriotism Political Tribalism	Transgender	Abortion Activism Feminism Sexism Misogyny	Immigration	Terrorism Human rights	Racism Prejudice Civil rights Discrimination Diversity Inclusion Politically correct

Figure 1. Chosen Political Terms based on Pew Research Center 2022 midterm election priority topics.[3]

In addition to the segmentation of the data on the chosen political terms, we also stratified and evaluated the occurrence of these terms across fields or time periods. Specifically, this data was stratified from the time periods of 1970 to 2020. Our aim is to investigate term usage by year.

After further filtering papers to make sure that chosen political topics are referenced in either the title or the abstract of the paper as a way of determining the focus of the research paper, the final database contains 1,784,056 papers on which we test our hypothesis: do the chosen political terms exceed the expected 5% growth rate in the period from 1970 - 2020?

Methods

Prior to finding the appropriate regression model, we conducted exploratory analysis to understand the spread and nature of the data in question. Once the data was imported, we used histograms and bar charts to understand at a basic level what to expect from our data and anticipate any hurdles we could expect from any abnormalities or outliers.

In addition, the correlation matrix is a common statistical tool which allows us to examine correlations for more than two variables at a given time. In this project, we are looking for a correlation between terms for the previous 50 years. We grouped the total data by term followed by frequency counts by year. We applied the correlation matrix on the grouped dataset to examine the correlation p-values and correlation coefficient.

Next, we ran a regression on our dataset using a Poisson exponential growth predictor. The Poisson method is appropriate for count data such as frequencies over a time span. We fitted the model based on our reference code (see Appendix). Furthermore, we created a prediction model to obtain the best fit predicted growth rate for each term's respective year range and count. After defining a predicted exponential growth function, we then modeled a second exponential growth function with an expected 5% growth rate. From there, we were able to plot the two models from the same first data point: predicted growth rate and expected growth rate. We performed this same analysis for each of the 24 terms. We defined exponential growth using the following formula:

$$y = a(1+r)^t$$

where y is the count for a given year, a is the initial count, r is the growth rate, and t is the time in years.

Results

Exploratory Analysis

In our initial exploration of the data, one thing that immediately was clear was that the data was incomplete past the year 2020, which is why our analysis includes data from 1970 until that 2020 cutoff. From there, it is evident that the shape of the histogram by year was the approximate shape of an exponential curve (Figure 2). Once the range of the dates were set, we looked at the data by term to ensure our terms were adequately represented and to be aware of any gaps or lack of representation as those could be potential issues for our model and would need to be accounted for (Figure 3). As is apparent from the chart, there are some terms with limited representation but not so much to be removed from our analysis. As such, we noted them and continued the analysis.

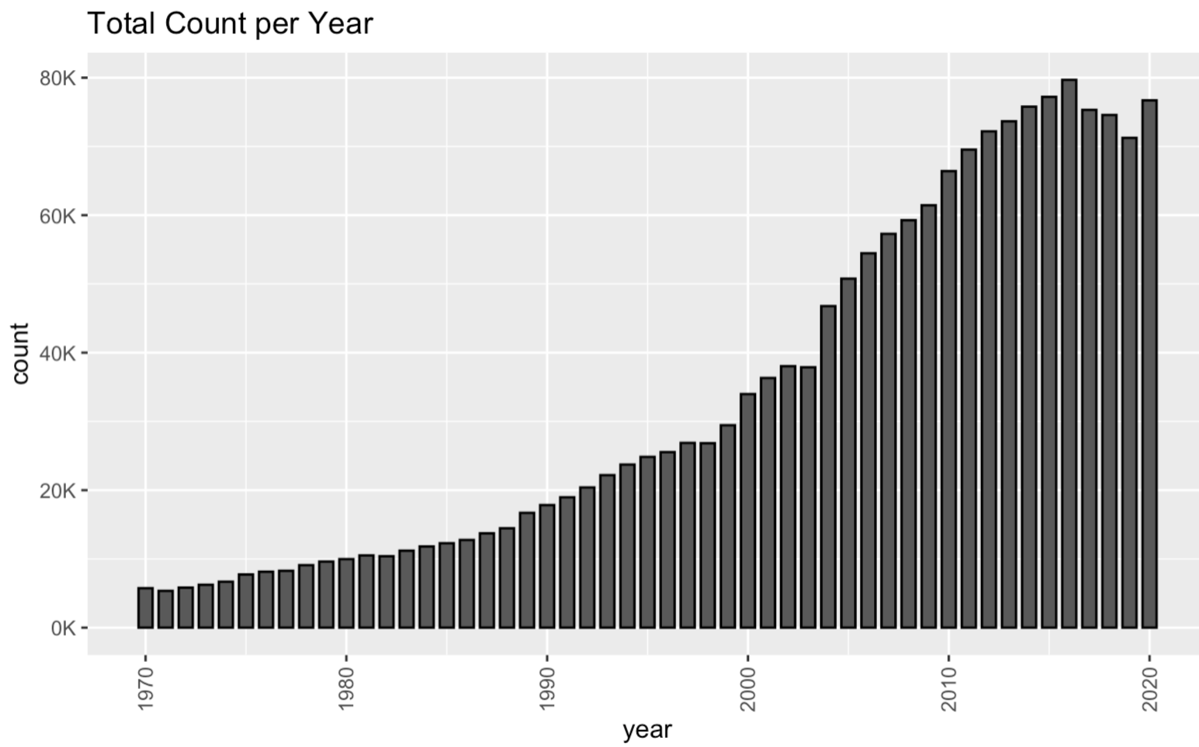


Figure 2. Histogram exploring cumulative count of gathered published articles from 1970 - 2020.

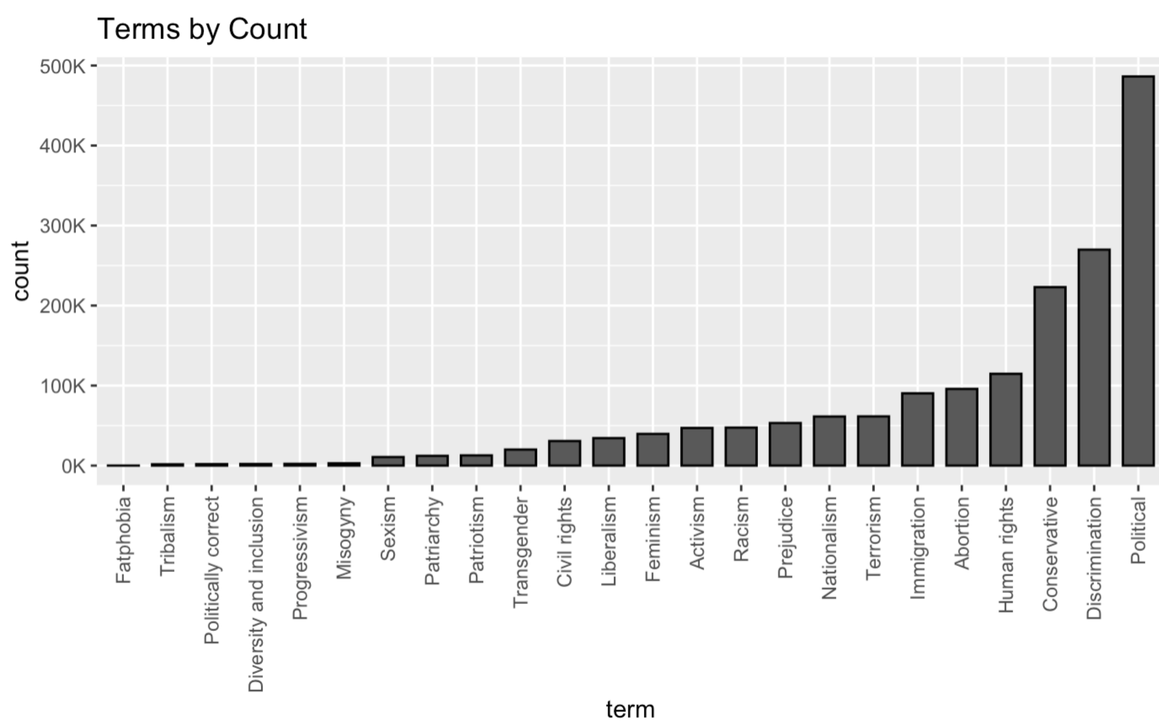


Figure 3. Barplot exploring cumulative count of published articles per individual term.

In continuation of our exploratory analysis, we created a correlation matrix to see if any terms displayed ties to one another through the 50 year time span. What we did not take into account is that a correlation matrix of 24 terms by 24 terms would create an overwhelming number of Pearson's r values to consider. After flattening the matrix into a more comprehensible list (see Appendix), it is easier to judge that almost all of the terms used in this paper are highly correlated with one another. As we are plotting individual terms against an expected growth pattern and not against each other, we determined that these correlations are important to note but would not necessarily interfere with predicting increasingly complex models.

At this point in our analysis, we were comfortable with our understanding of the data, had cleaned up where necessary and noted any potential pitfalls in the terms. From here, we were ready to proceed with fitting an exponential growth curve to our terms.

Exponential Growth Model

The table below (Figure 4) houses the set of predicted exponential growth rates for each of the 24 terms. Likewise, corresponding p-values produced through a poisson.test() method allowed for a comparison test between two Poisson generated predicted rates (ie. predicted growth rate versus 5% expected growth rate). As noted, all terms but 3 are determined to be statistically significant ($p < 0.001$) with predicted growth rates all above the baseline 5% expected value. The three terms which fall below the 5% expected level are Abortion (predicted rate = 4.11%), Conservative (predicted rate = 4.99%), and Political (predicted rate = 3.64%). On the higher end of the predicted rates, the top three predicted growth rates are for Diversity and inclusion (predicted rate = 29.24%), Transgender (predicted rate = 20.5%), and Fatphobia (predicted rate = 14.14%). However, it is worth stating that for these three high growth rate terms, they each appeared sparsely in the publication literature until the 1990s and later, at which point the frequency of these terms dramatically increased in the last twenty years (2000-2020).

Figure 5 displays two of the twenty-four term plots showcasing examples of higher and lower than predicted growth rates versus the expected 5% growth rate. For all twenty-four plots, see the Appendix.

	Terms <chr>	Predicted Growth Rate <chr>	P Values <dbl>
1	Abortion	4.11%	1.0000e+00
2	Activism	10.69%	0.0000e+00
3	Civil rights	6.58%	0.0000e+00
4	Conservative	4.99%	1.0000e+00
5	Discrimination	5.63%	0.0000e+00
6	Diversity and inclusion	29.24%	0.0000e+00
7	Fatphobia	14.14%	2.3127e-10
8	Feminism	7.74%	0.0000e+00
9	Human rights	9%	0.0000e+00
10	Immigration	7.69%	0.0000e+00
11	Liberalism	7.25%	0.0000e+00
12	Misogyny	9.41%	0.0000e+00
13	Nationalism	7.3%	0.0000e+00
14	Patriarchy	8.87%	0.0000e+00
15	Patriotism	8.52%	0.0000e+00
16	Political	3.64%	1.0000e+00
17	Politically correct	5.02%	3.5815e-06
18	Prejudice	7.29%	0.0000e+00
19	Progressivism	5.59%	3.3159e-04
20	Racism	8.76%	0.0000e+00
21	Sexism	8.01%	0.0000e+00
22	Terrorism	9.62%	0.0000e+00
23	Transgender	20.5%	0.0000e+00
24	Tribalism	6.82%	6.0062e-41

Figure 4. Table of predicted growth rates and p-values for each term.

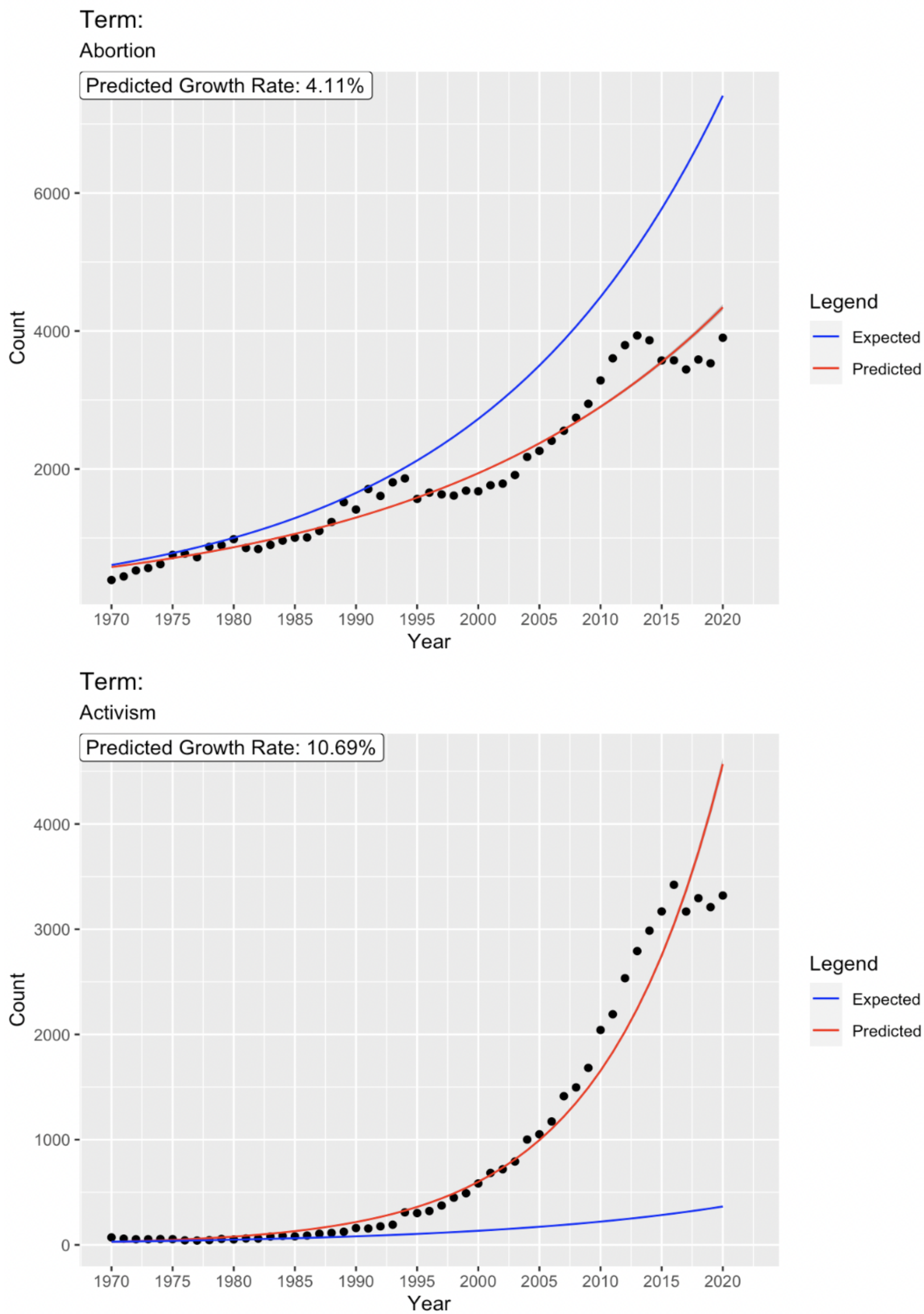


Figure 5. Two of the twenty-four plots are shown here. The top plot for Abortion has a lower 4.11% predicted growth rate (shown in red) than expected while the term Activism has a higher predicted growth rate of 10.69%. The blue line is the 5% expected growth rate for the respective term.

Segmented Exponential Growth Models

After looking at the overall predicted growth trends, we decided to do a more in depth analysis of a handful of terms by decade. We wanted to see if the overall predicted growth held constant through the five decade time span or if rates varied dramatically during different time points. As this is an exponential growth model, the nuances of a particular decade segment may differ significantly from the overall model–and without segmentation, it can be challenging to see those nuances in earlier decades as they are visually dwarfed by the data points closer to the present (assuming those terms are seeing exponential growth).

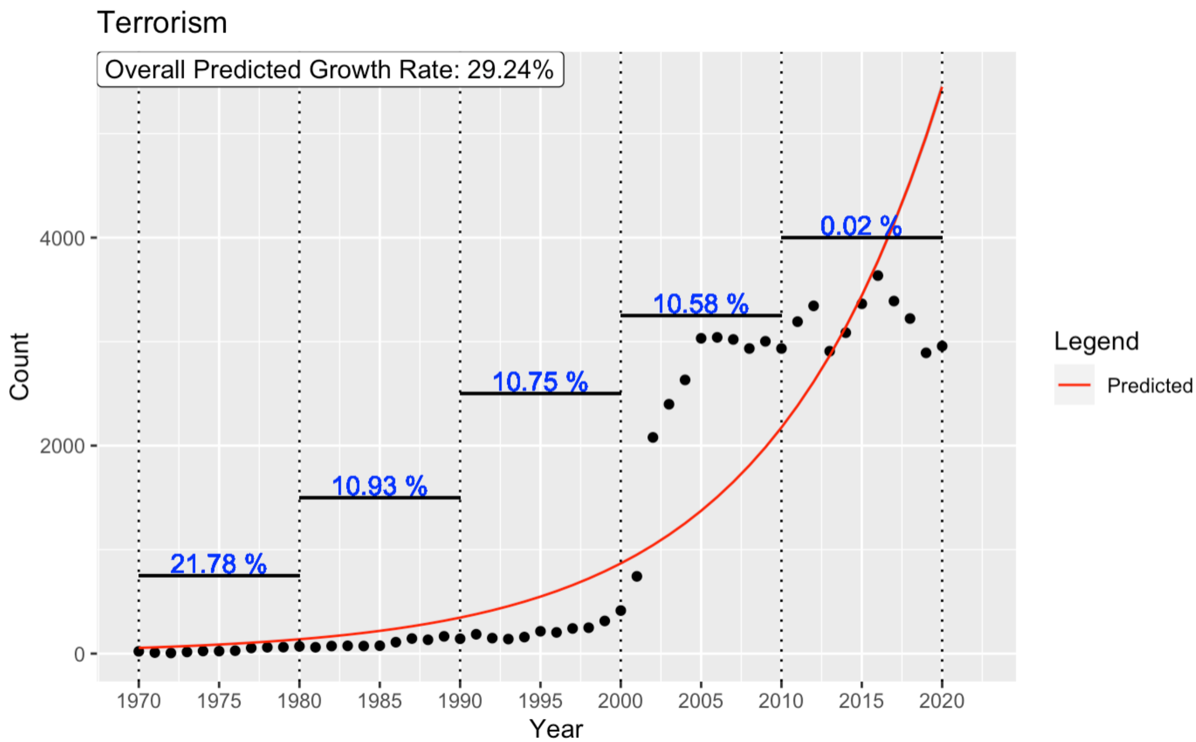


Figure 6. Predicted growth for the term Terrorism divided by decade.

As the term with highest overall growth rate, the segmented graph on Terrorism is a fantastic example of a few things. Firstly, the clear jump that happened after 2001 makes sense given the historical context of events at the time (ie. 9/11 attack). Also, since that jump happened near a decade cutoff, if taken without the full context no individual decade would have shown as high of a growth rate as the full range does. And finally, the growth rate in the 1970s is quite high, despite being graphically hard to see. This further validates the benefit this segmentation can bring.

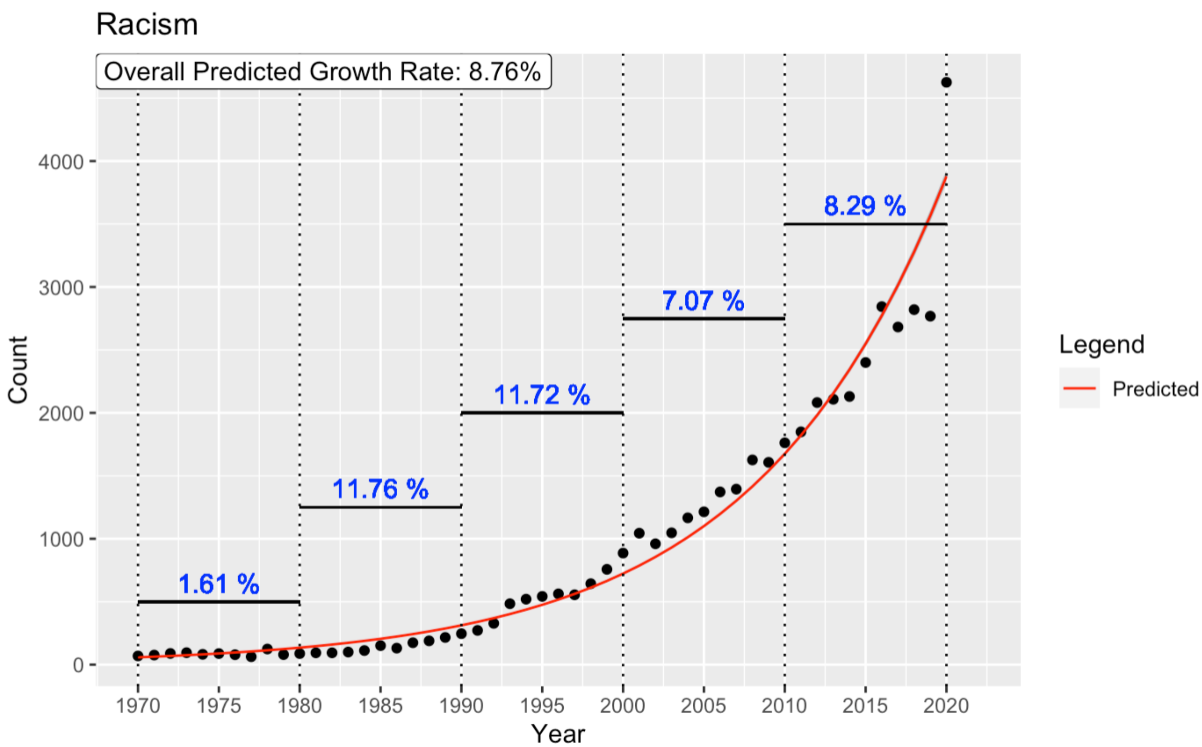


Figure 7. Predicted growth for the term Racism divided by decade.

We also chose to take a closer look at the term Racism. Unlike many of the other terms examined in this set of analyses, Racism displays a generally steady rate of growth since the 1980s. Its overall predicted rate of growth (8.76%) falls within the range of growth from 1980 to 2020 (7.07% - 11.76%). It's possible that if we were to look more closely within each decade, say by year, there would be fluctuations not visible as a result of the smooth curvilinear model.

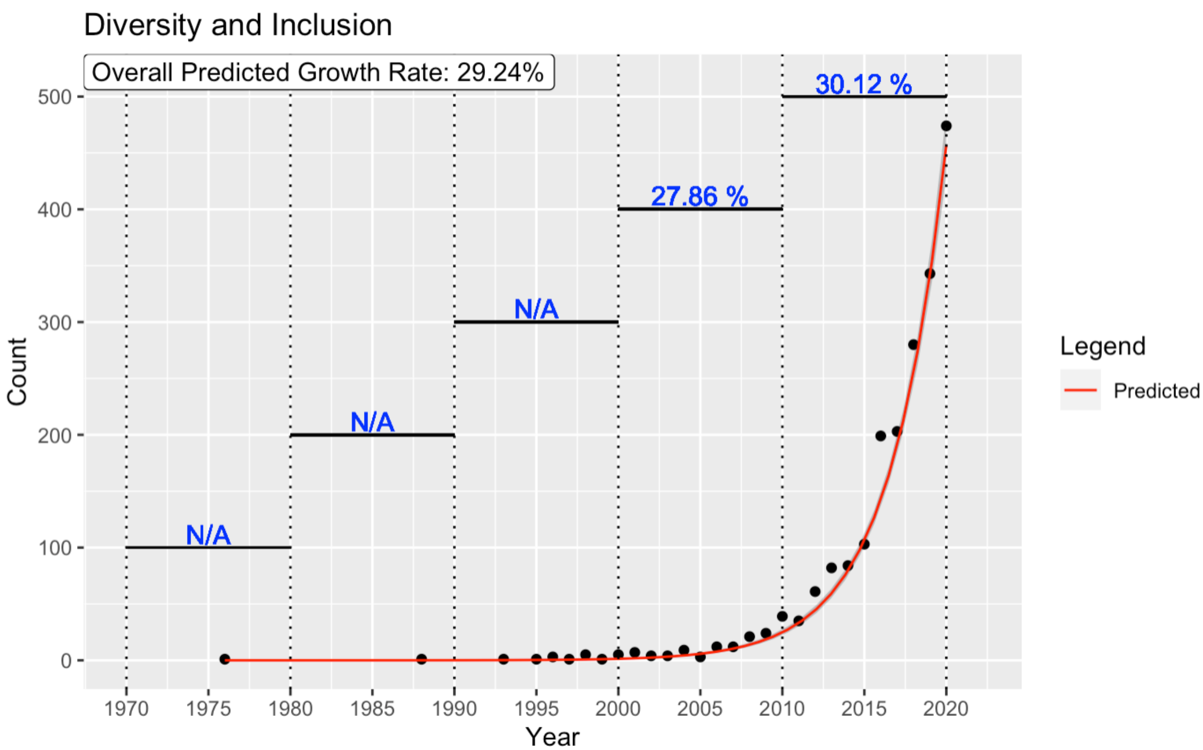


Figure 8. Predicted growth for the term Diversity and inclusion divided by decade.

Diversity and inclusion also depicts an interesting segmented trend not otherwise visible. From 1975 until the mid-1990s, there were not enough data points (total count < 20) to run a viable statistical analysis per decade. However, from 2000 onwards, the number of publications including this term began to rapidly grow with the highest growth of 30.12% from 2010 to 2020.

We looked at two final terms: Liberalism and Conservative. We chose these two terms for their inherent connection to politics. Unlike in the three previous examples, the plots for Liberalism and Conservative (Figures 9 and 10) show an interesting drop off in growth (-1.89% and -0.56%, respectively). Particularly for the term Conservative, even though the overall predicted growth rate was positive (4.99%) and close to the expected 5% baseline rate, the individual segments detail the declining rate per decade from 1970 until 2020. Again, this trend would be obscured if looking only at the total predicted growth rate.

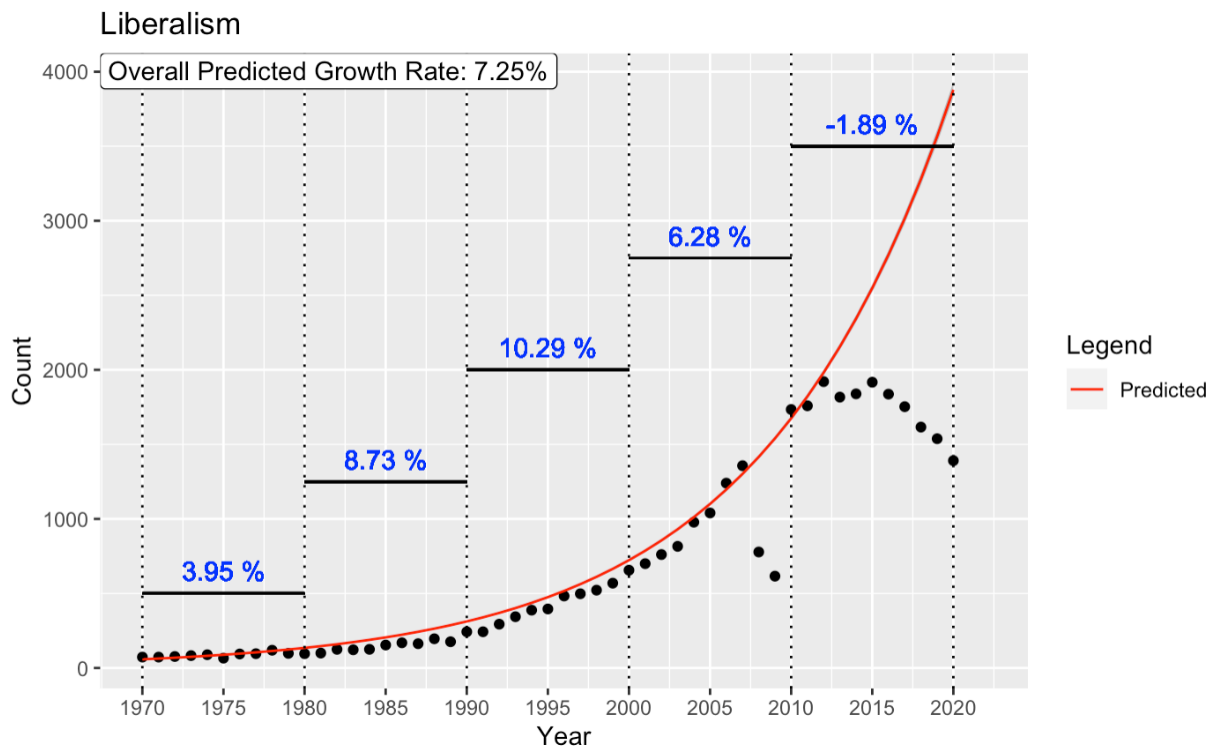


Figure 9. Predicted growth for the terms Liberalism divided by decade.

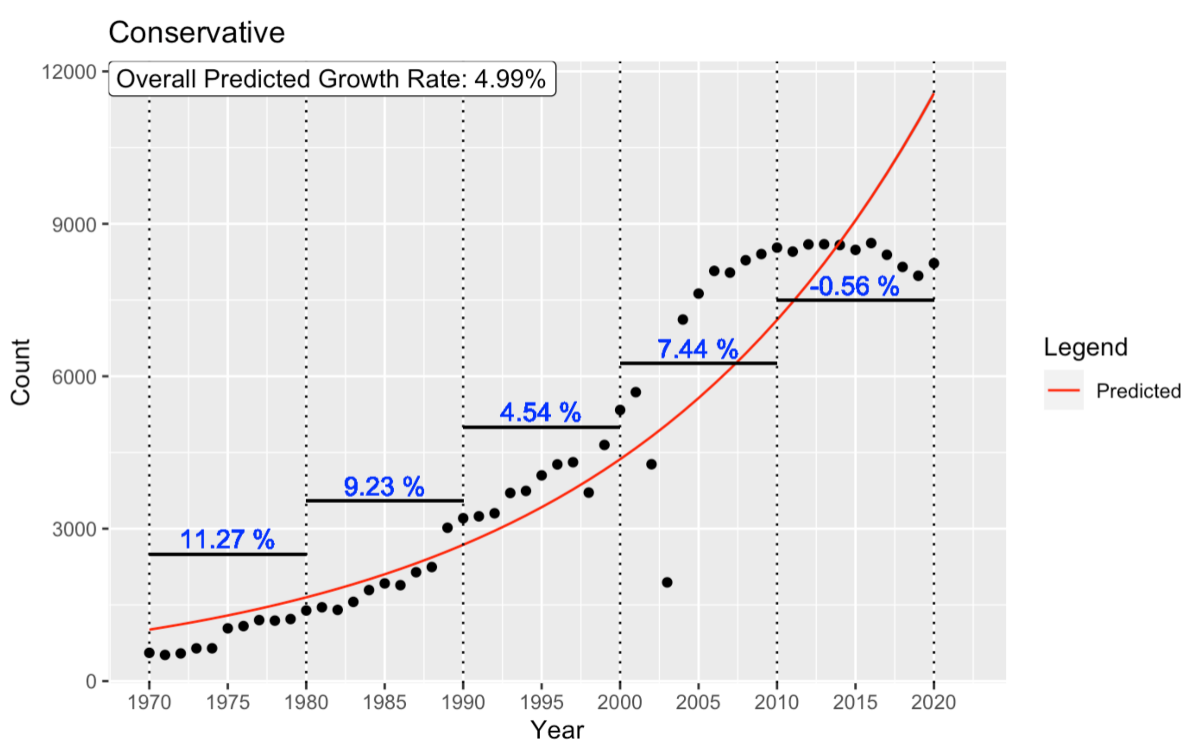


Figure 10. Predicted growth for the terms Conservative divided by decade.

Discussion

Overall, we can confidently say that publications have in fact become more politicized in the last fifty years. Though three of the terms we collected failed to exceed the expected 5% threshold rate, the remaining 21 terms exhibit a range of exponential growth patterns that contribute to the impression that politics has indeed entered academic publications more so than in previous eras. While we hesitate to definitively conclude that each of these highly specific, individualized terms directly demonstrate politicization, the evidence presented here is suggestive.

With further analysis of terms into segments, we can also see how rates of growth varied from decade to decade, and even have a drop off in frequency in some cases. Our assumption is that these trends fluctuate with societal happenings (presidential elections, war, social upheaval and riots, to name a few examples). While social movements can start up quickly, research does require a lengthy amount of time with a lag in the follow-up publication date. As a result, it is difficult to map social changes directly onto publication timelines but word frequencies do offer a proxy for how these two systems are intertwined.

And of course, the evolution of modern technologies (ie. the Internet, computers with increasingly higher processing power, and corresponding statistical methods) dramatically shifted the landscape and potential power for doing research in academia. In many of the graphs, there is a noticeable jump in publications from the mid-1990s onwards after the world wide web became available to the general public. As with societal shifts, technological progress is also closely intertwined with modern research and associated publication rates .

Because of the complicated nature of the data, we encountered several challenges in the development of the statistical analyses necessary to answer our hypothesis. First and foremost, though the data collection process occurred prior to the onset of this project, the quality of the data collected from the published source was poor. Upon closer investigation, it was apparent that the datasource had many “holes” and did not necessarily capture the terms we were interested in. We cleaned this data (as mentioned in the methods section) which still left a large number of data points to use in our analysis (1.78 million).

The next primary challenge was determining the appropriate statistical model to apply to this data. After trial and error, and referring to various resources, we were able to determine that a Poisson model would fit our data best as this analytical approach was specifically created to compute count data over time. Using this predictive model, we were able to relatively easily fit the expected growth model. Likewise, after a bit of investigative work, we found the poisson.test() method to appropriately compare our predicted and expected models. In a similar vein, we applied exponential growth to each of the segmented decade plots though there may be better model fits depending on individual segment behaviors. In some decades, it may be more appropriate to apply linear or logistic regression analyses, but this would require closer examination of the individual data segments as well as understanding if there is enough power and data points to continue to run statistical analyses on each decade.

The final challenge lay in the composition of this paper. Generating models and plots is one thing; finding meaning in these figures is another. Luckily, our model outputs are straightforward to interpret but because of the magnitude of data, sifting through each term's corresponding plot and output consumes the ultimate resource which is time.

Future Directions

This exploration into academic publications is one of the first forays into this realm. As a result, there are many different questions and directions that arise a result of this preliminary set of analyses.

The set of terms used here is very limited. Applying Natural Language Processing (NLP) models on this set of publications would provide greater insight into the context in which each of these terms is used. These models can look into the valence each term and the general direction of the papers as a whole. Additionally, each term has a set of closely related words and synonyms which can be projected into NLP models to further deepen the understanding of the context for each publication. Likewise, expanding the database of publications beyond the S2ORC would also increase the reach and scope of investigating these kinds of research questions.

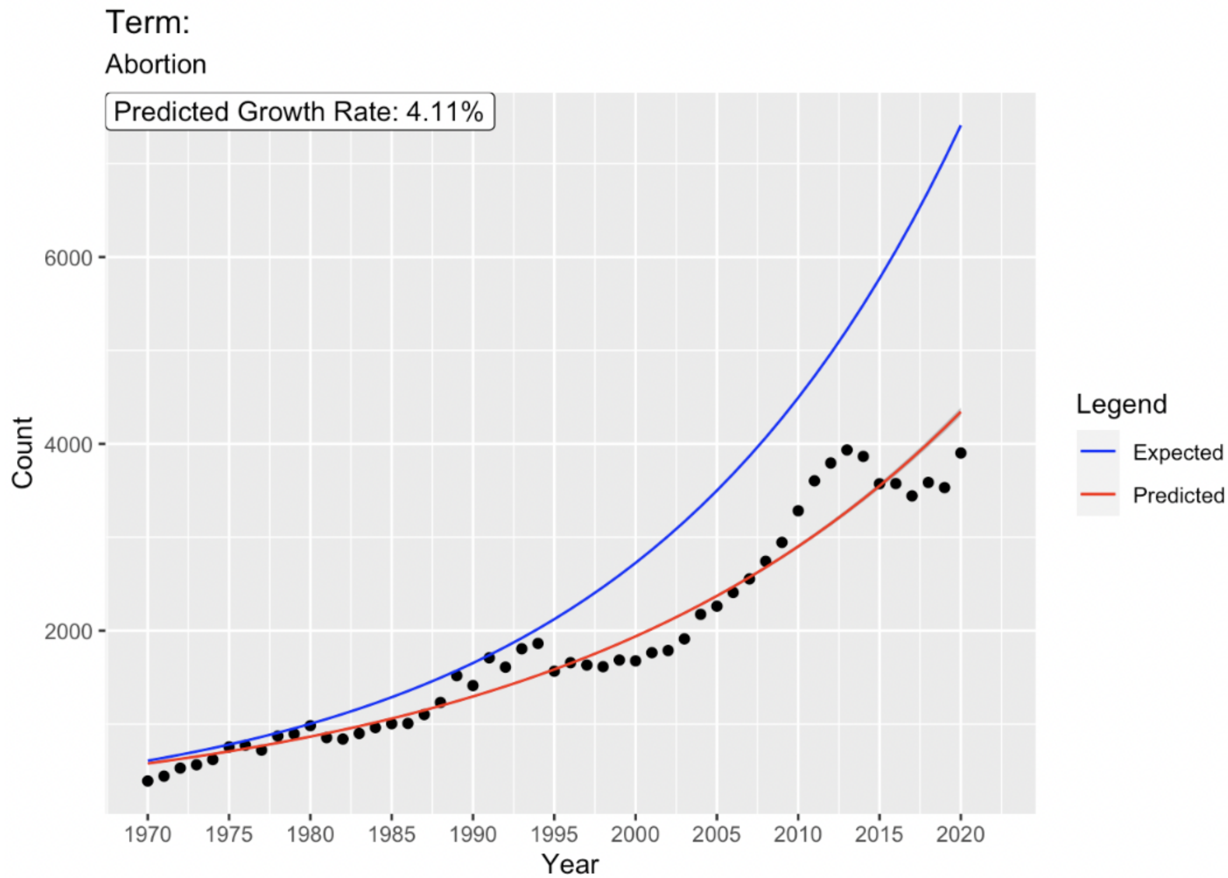
In parallel, the way terms are used are highly dependent on the field of publication. This study takes a broadstroke across academic publications which inherently misses the nuances of how terms come to be and evolve in a given research domain. There is a myriad of potential studies each field could do on their own list of politicized terms to determine if there is increasing outside political influence.

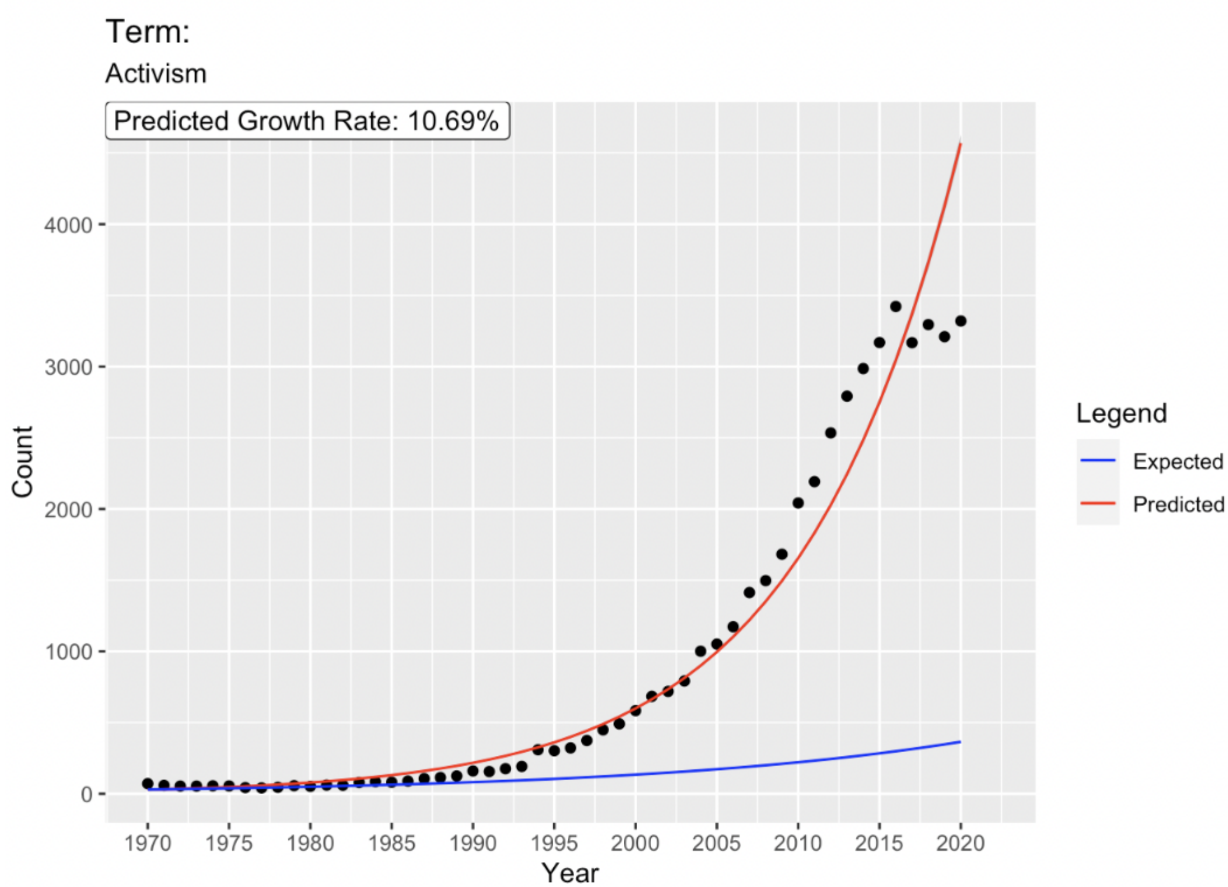
Finally, segmenting the data to look deeply into a particular decade would provide even more clues as to possible interactions between academia and the political sphere. Within each individual year and decade, there are countless events which contribute to the social realm. Each of these events has the potential to influence research trends. Drawing lines between external events and the internal workings of academic research is a rich and yet poorly investigated direction of work.

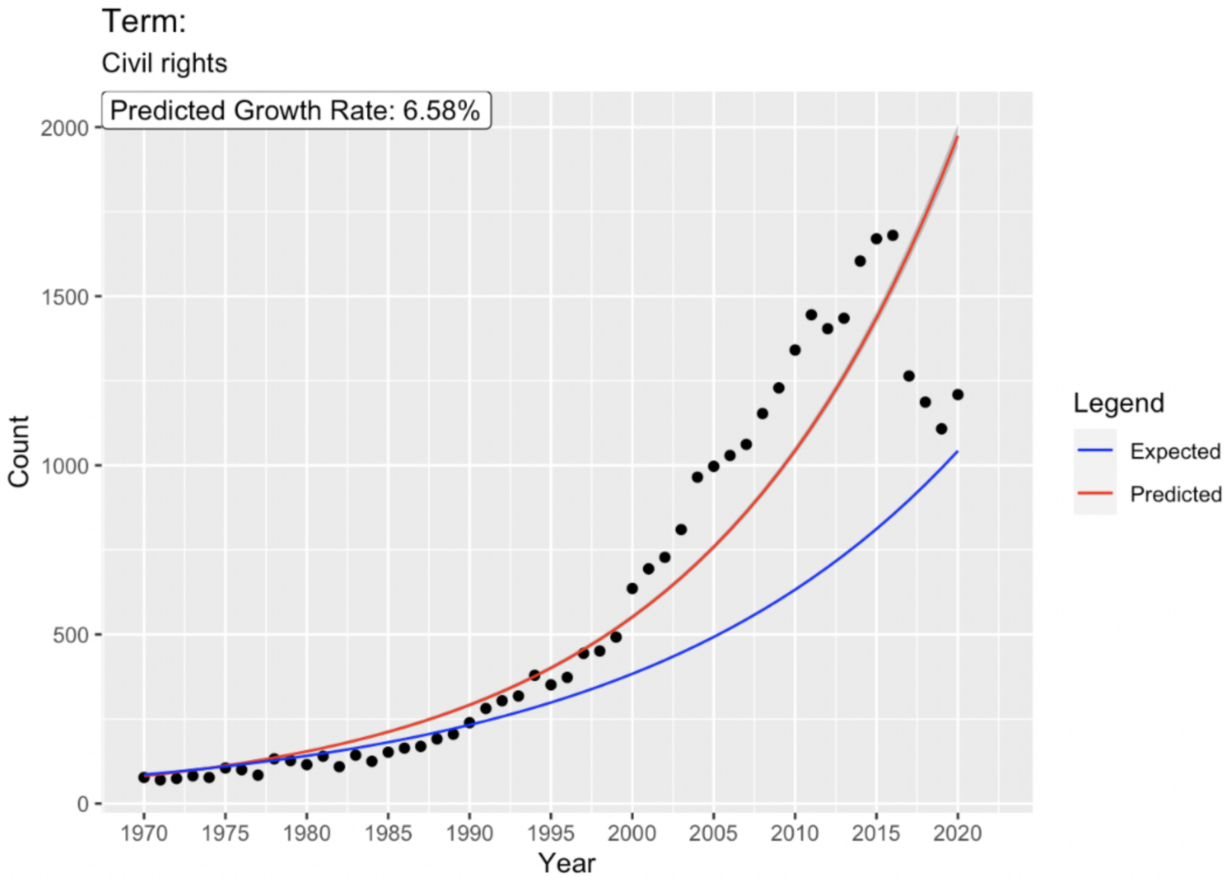
This list of future directions is short but there are countless other directions to take this line of thought. We believe that starting with any of these potential research directions will open up further questions and need for analytical prowess. We hope that this work will inspire others to be introspective of their own research questions, findings, and the connections that link institutions to one another.

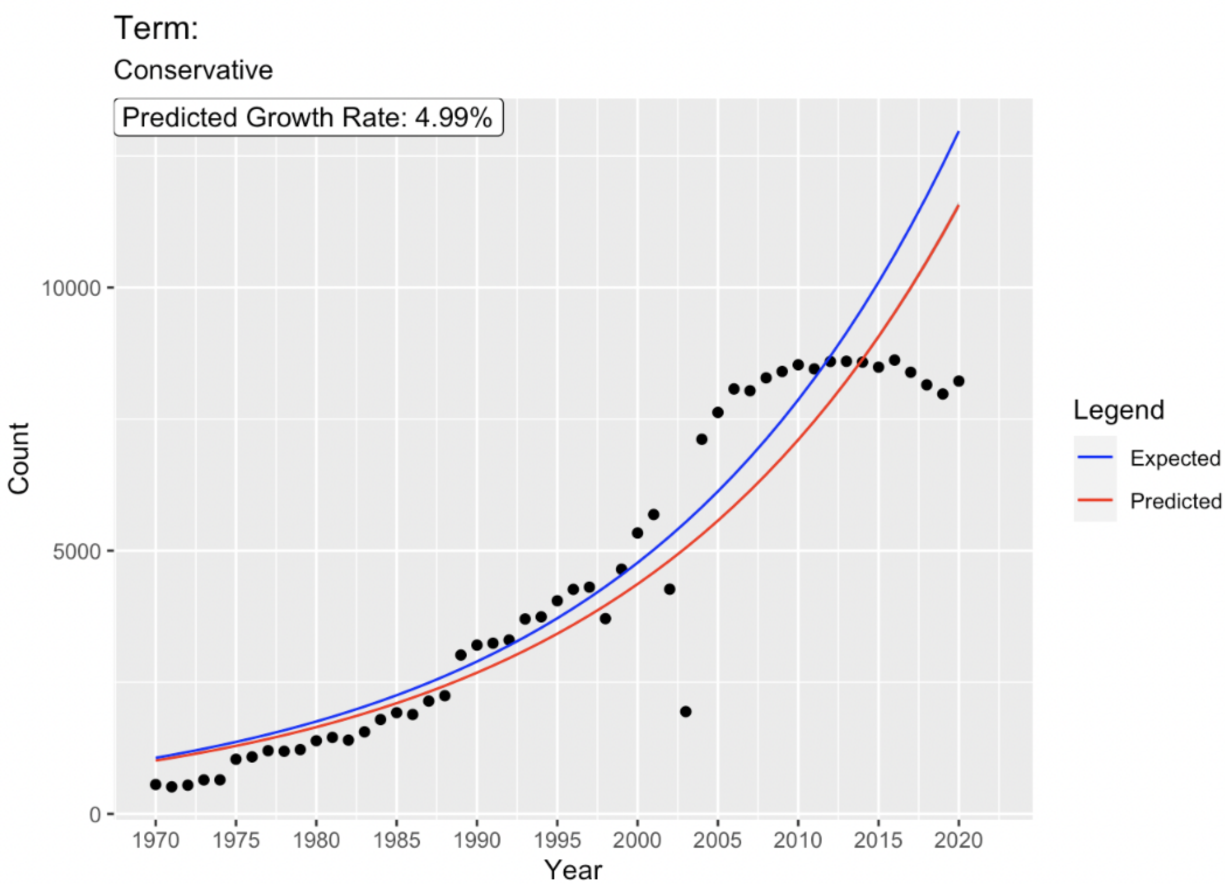
References and Appendix

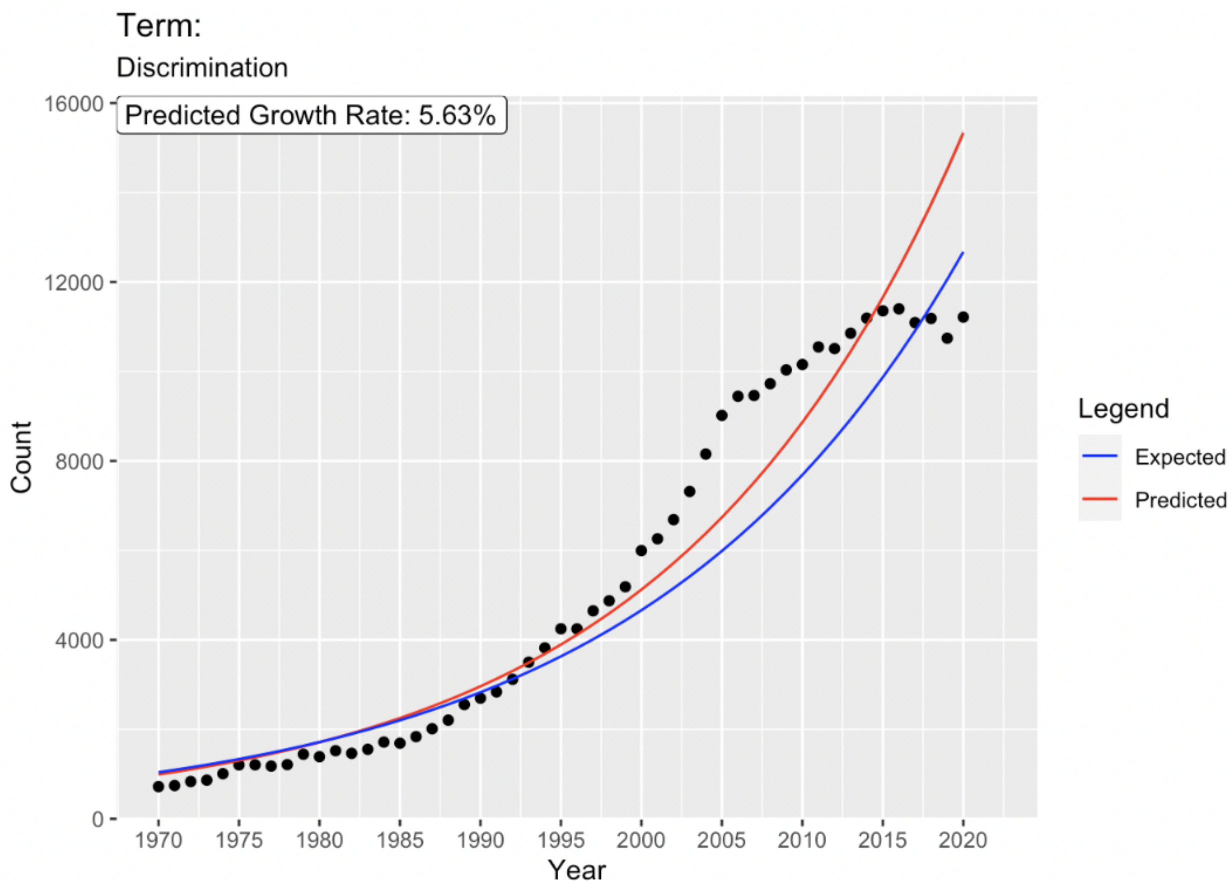
1. Rozado, David. Themes in Academic Literature: Prejudice and Social Justice. *National Association of Scholars.* Summer (2022). DOI: 10.51845/35.2.5
2. Bornmann, L., Haunschild, R. & Mutz, R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanit Soc Sci Commun* **8**, 224 (2021). https://doi.org/10.1057/s41599-021-00903-w
3. https://www.pewresearch.org/fact-tank/2022/11/03/key-facts-about-u-s-voter-priorities-ahead-of-the-2022-midterm-elections/

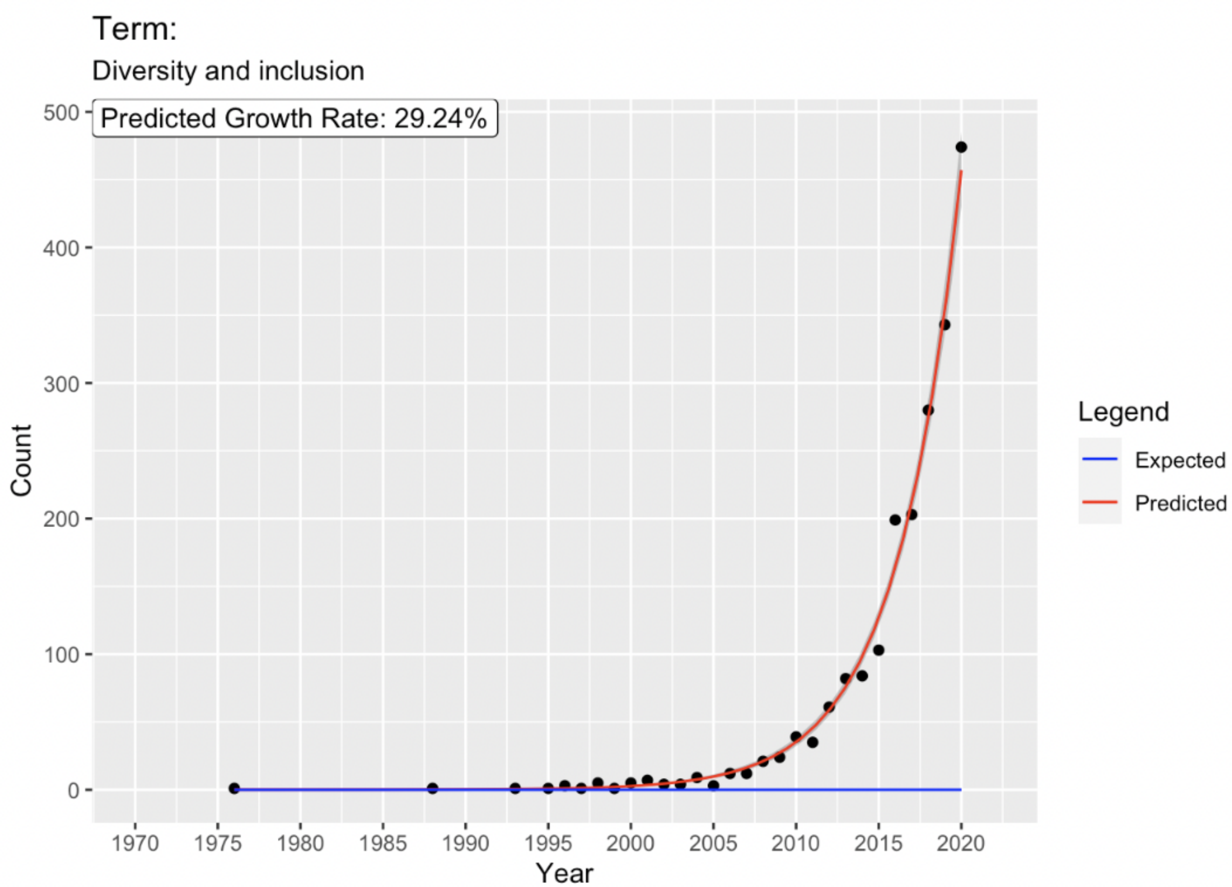

Term:
Abortion
Predicted Growth Rate: 4.11%
Count
6000
4000
2000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

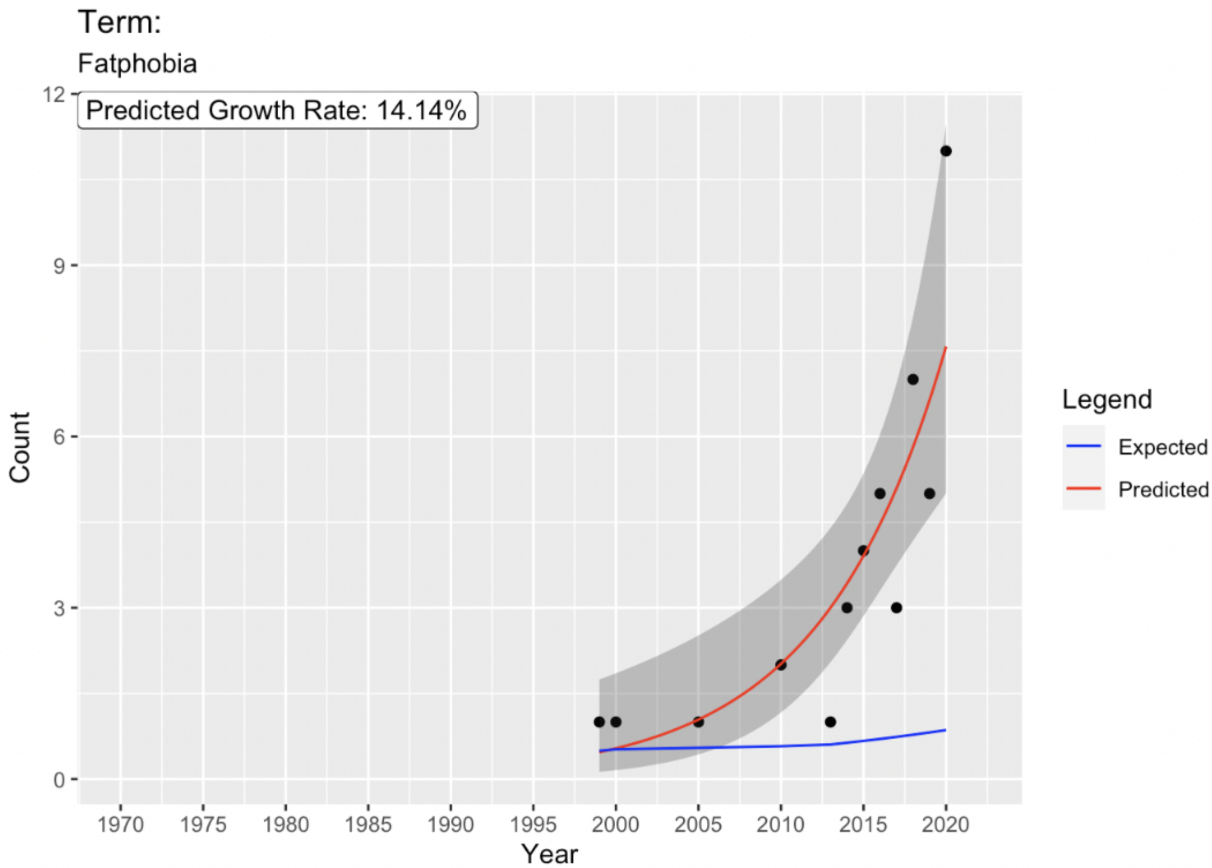

Term:
Activism
Predicted Growth Rate: 10.69%
Count
4000
3000
2000
1000
0
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

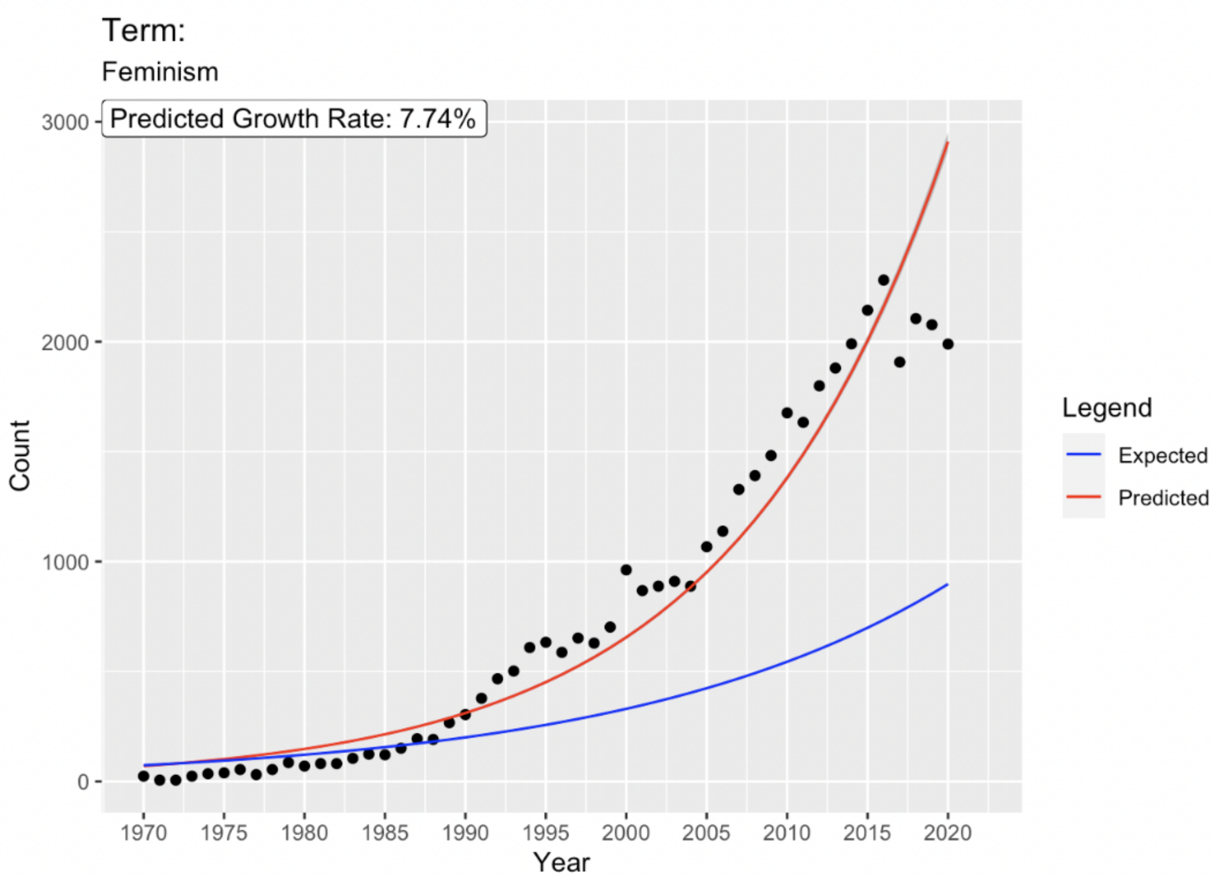

Term:
Civil rights
Predicted Growth Rate: 6.58%
2000
1500
1000
500
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

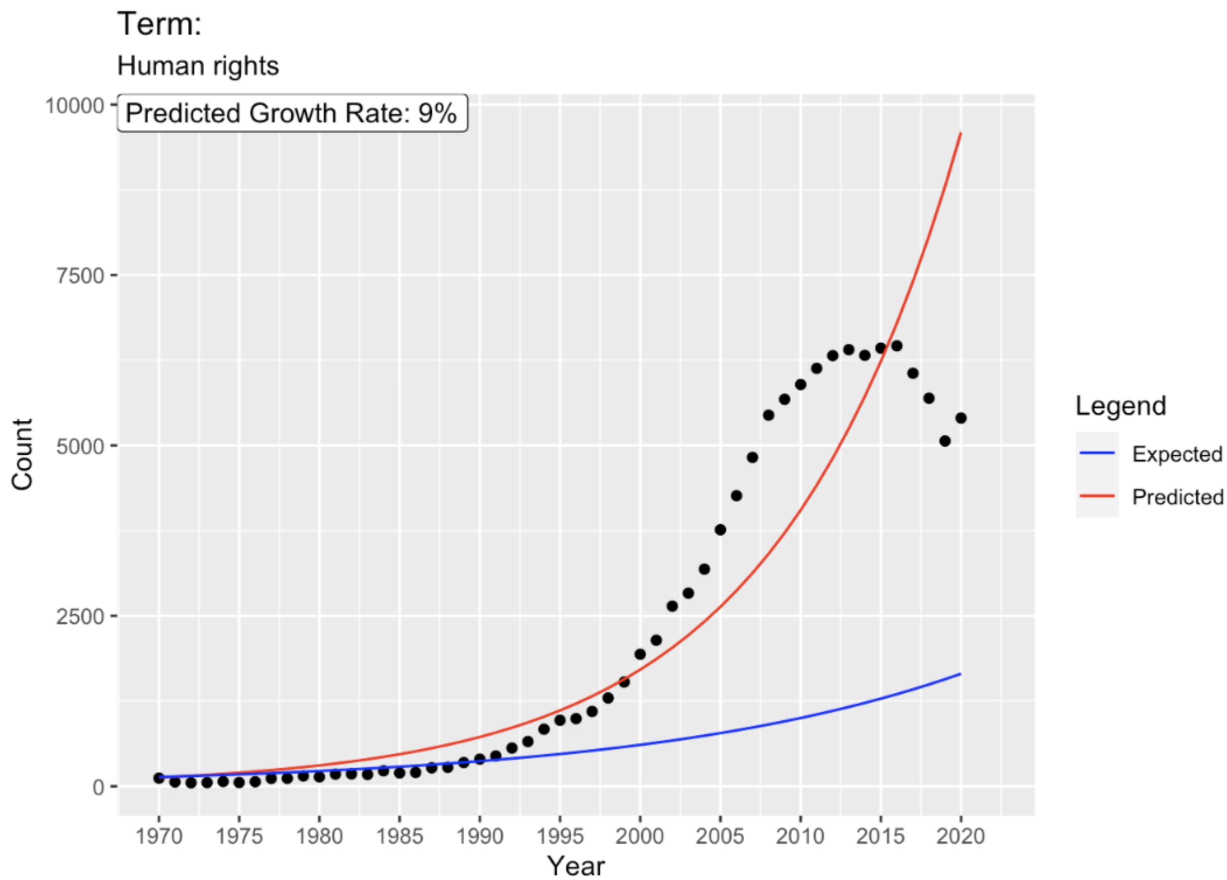

Term:
Conservative
Predicted Growth Rate: 4.99%
10000
5000
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

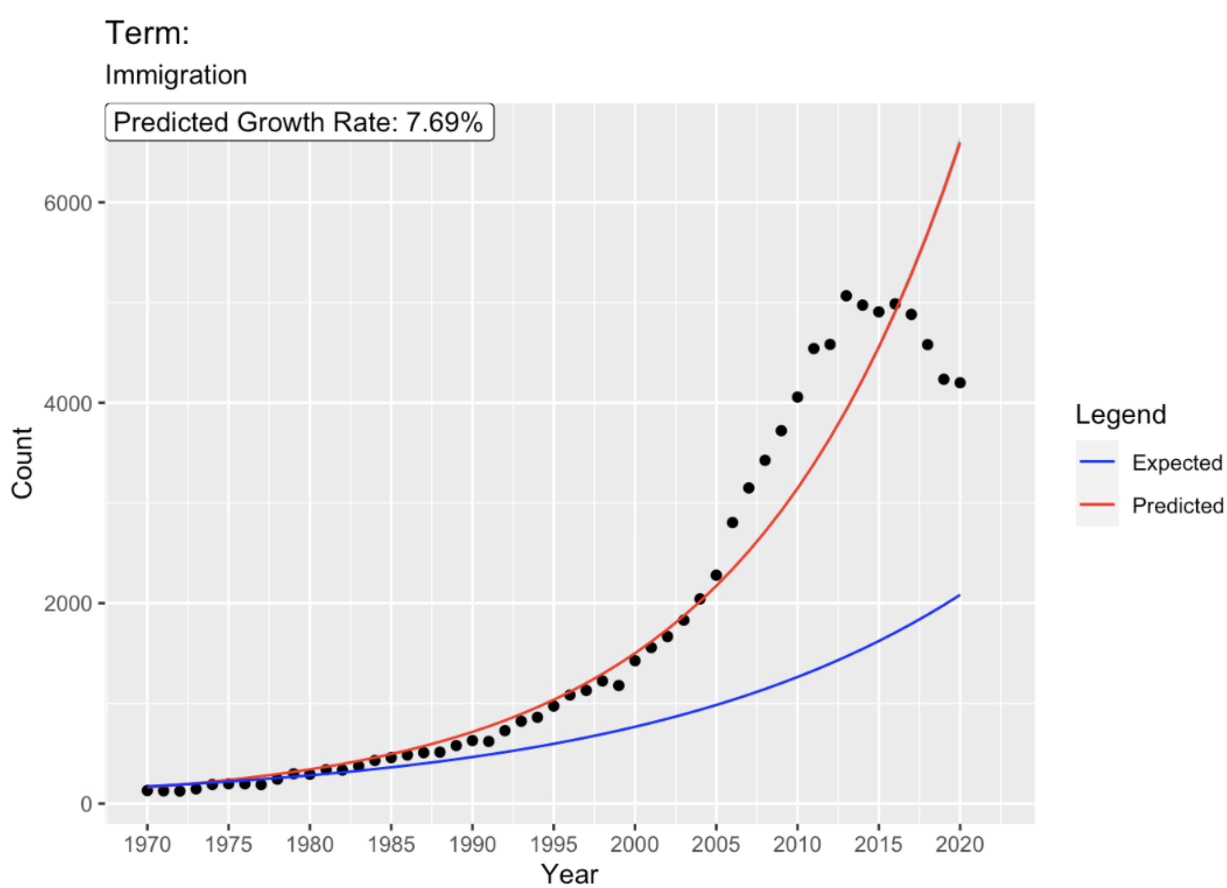

Term:
Discrimination
Predicted Growth Rate: 5.63%
Count
Year
Legend
Expected
Predicted
0
4000
8000
12000
16000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020

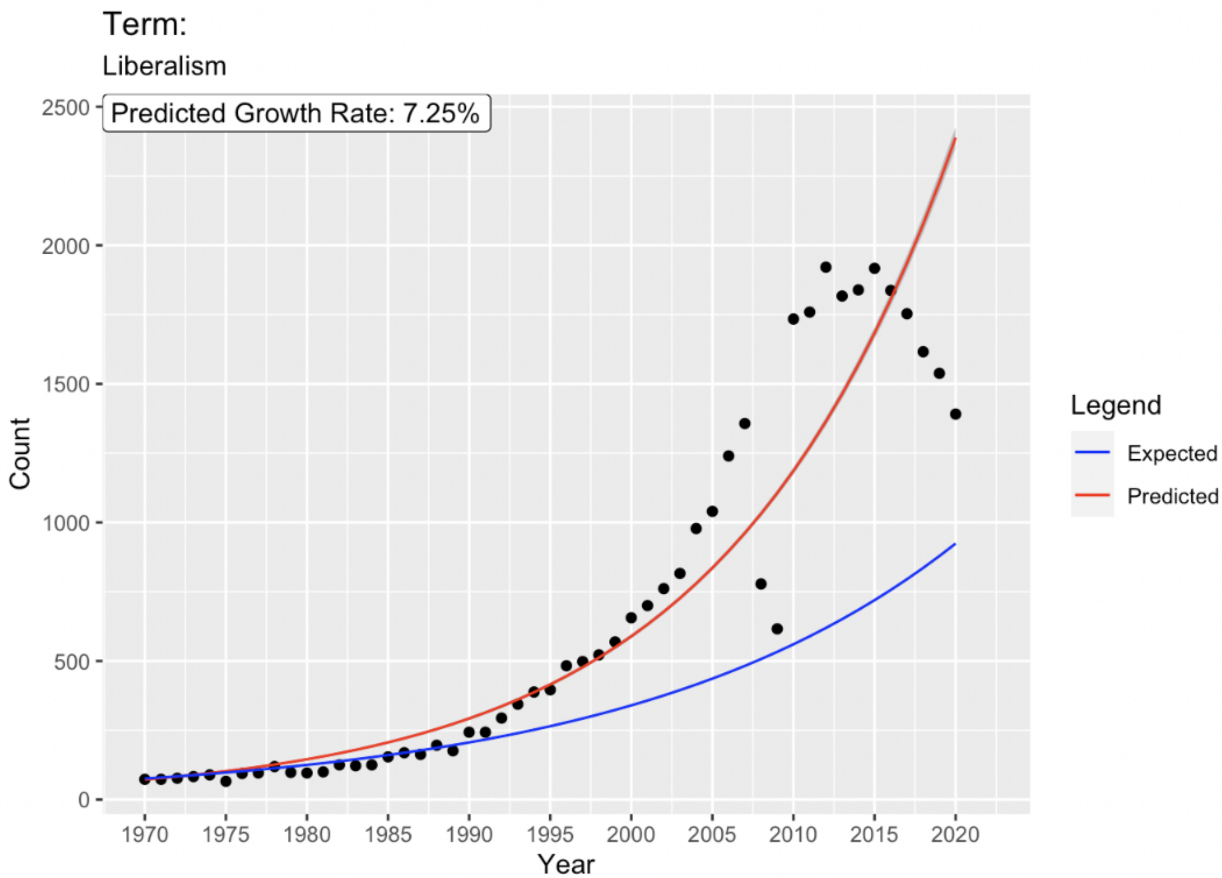

Term:
Diversity and inclusion
Predicted Growth Rate: 29.24%
Count
Year
Legend
Expected
Predicted
0
100
200
300
400
500
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020

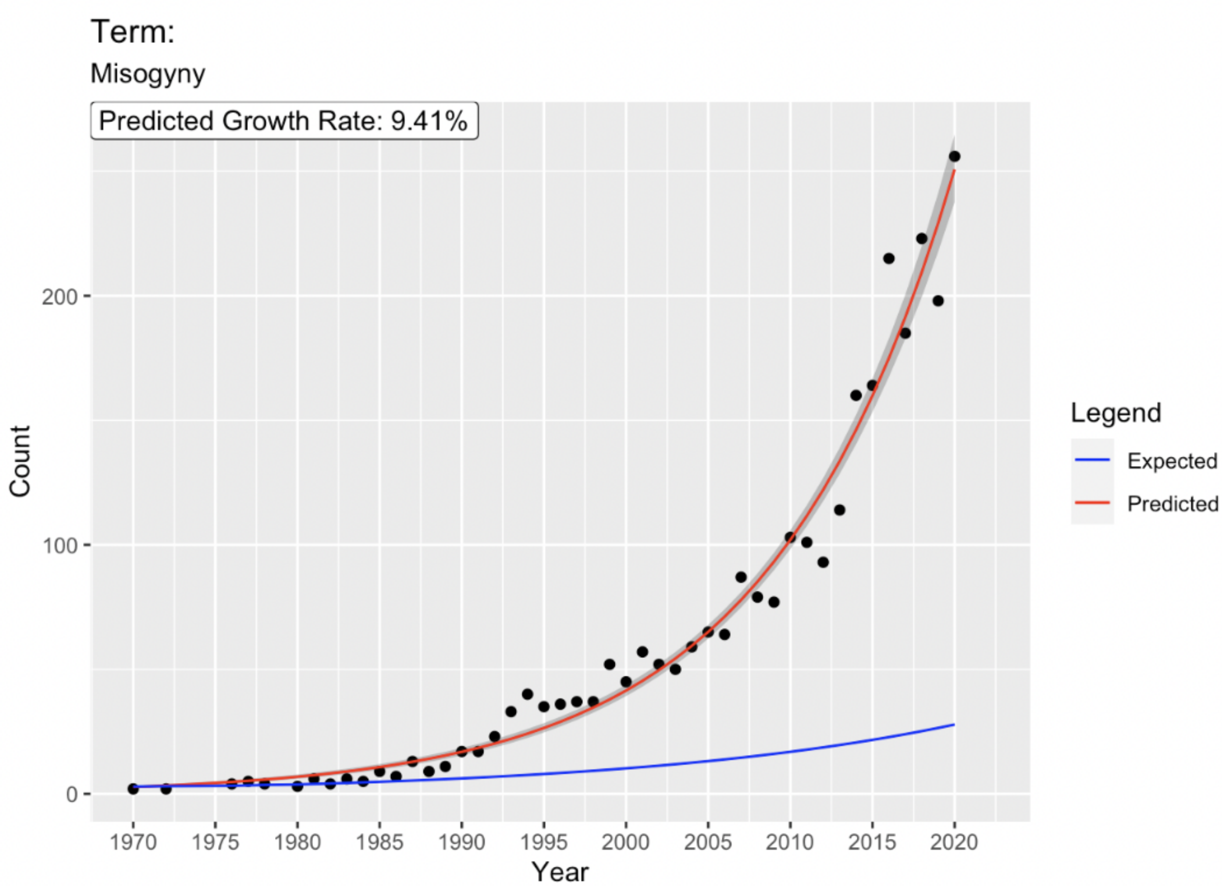

Term:
Fatphobia
Predicted Growth Rate: 14.14%
Count
Year
0
3
6
9
12
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Legend
Expected
Predicted

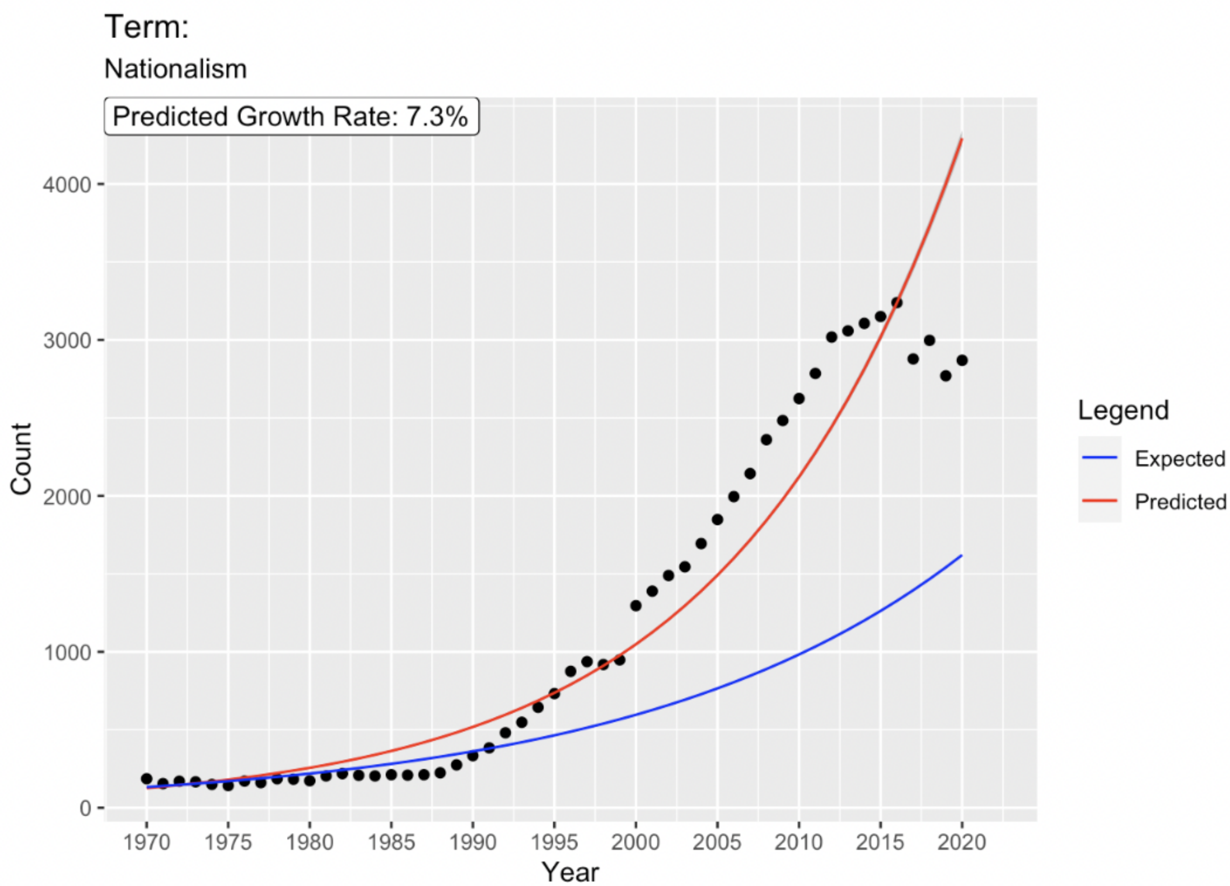

Term:
Feminism
Predicted Growth Rate: 7.74%
Count
Year
0
1000
2000
3000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Legend
Expected
Predicted

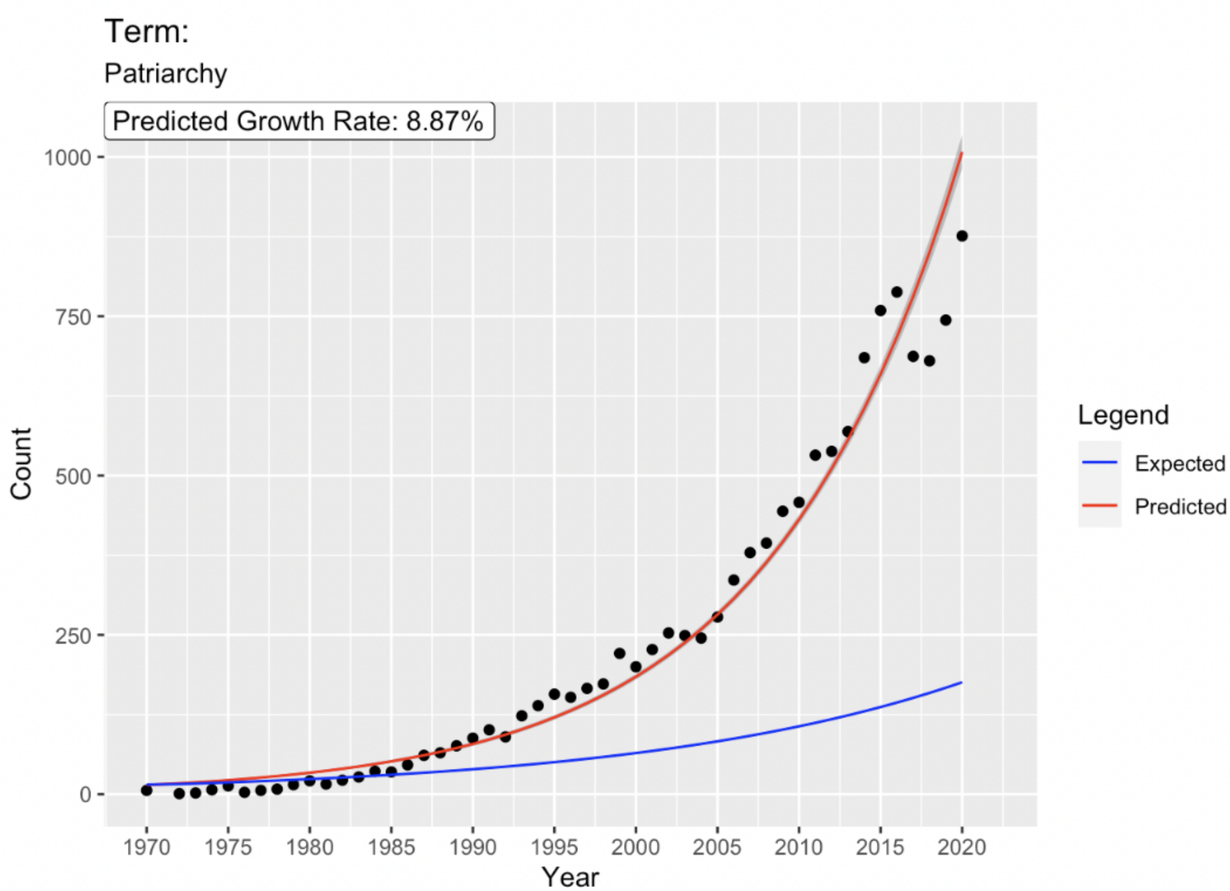

Term:
Human rights
Predicted Growth Rate: 9%
Count
Year
Legend
Expected
Predicted

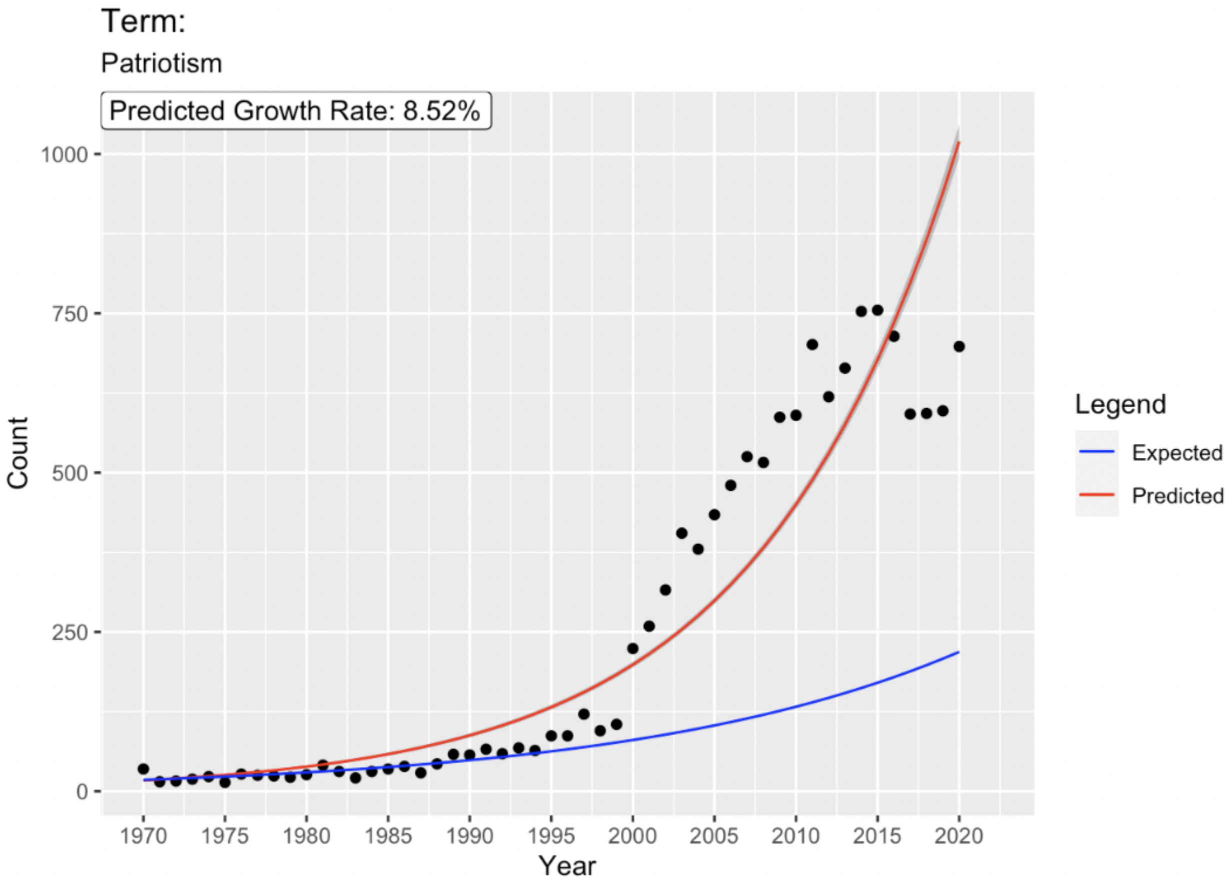

Term:
Immigration
Predicted Growth Rate: 7.69%
Count
Year
Legend
Expected
Predicted

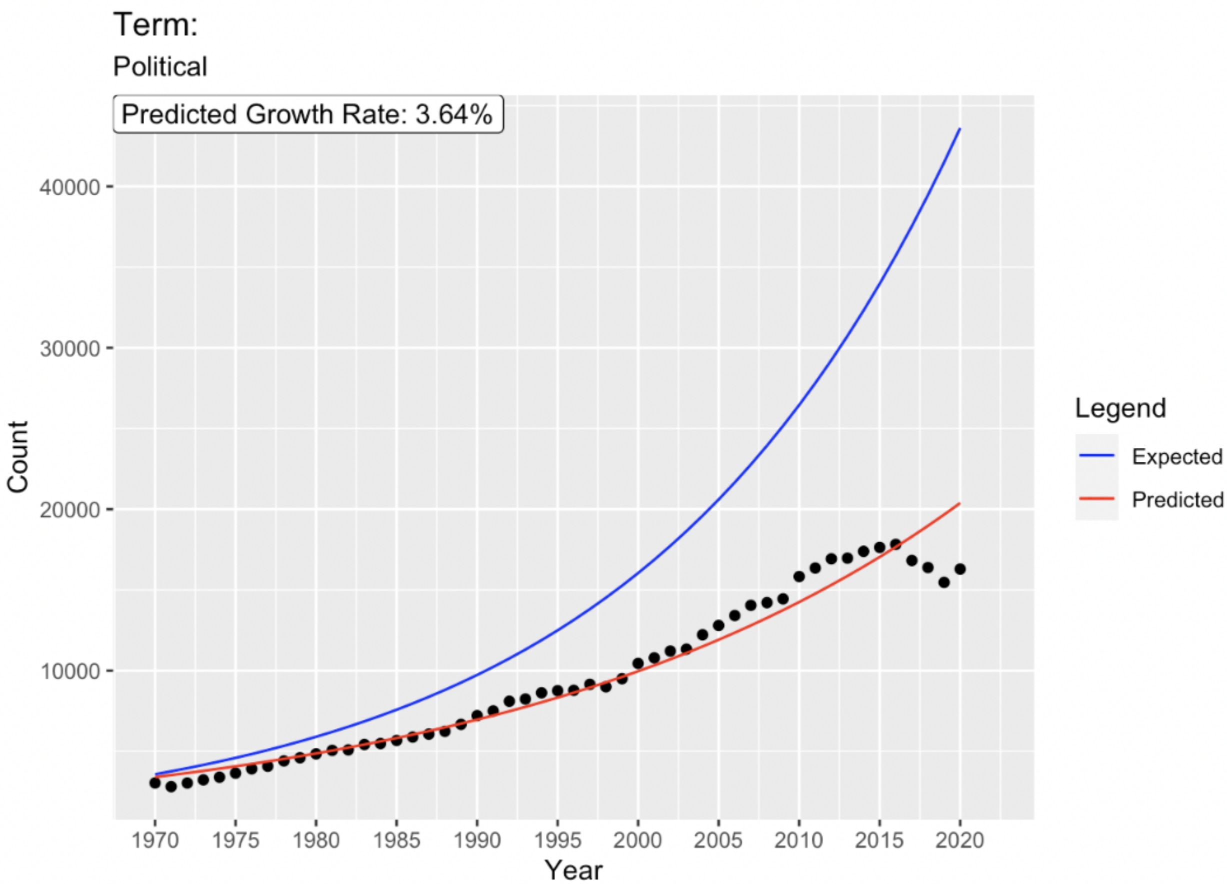

Term:
Liberalism
Predicted Growth Rate: 7.25%
2500
2000
1500
1000
500
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted


Term:
Misogyny
Predicted Growth Rate: 9.41%
200
100
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

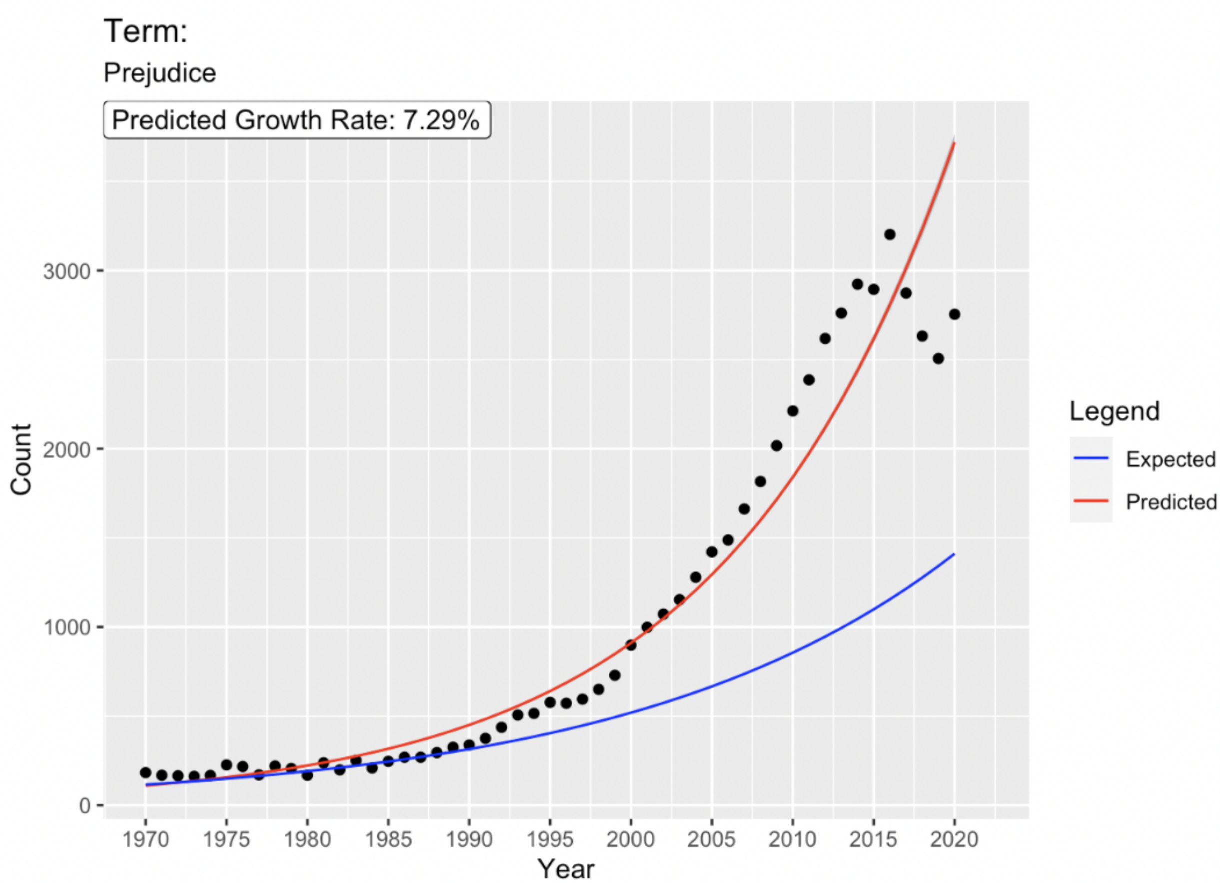

Term:
Nationalism
Predicted Growth Rate: 7.3%
Count
Year
Legend
Expected
Predicted


Term:
Patriarchy
Predicted Growth Rate: 8.87%
Count
Year
Legend
Expected
Predicted

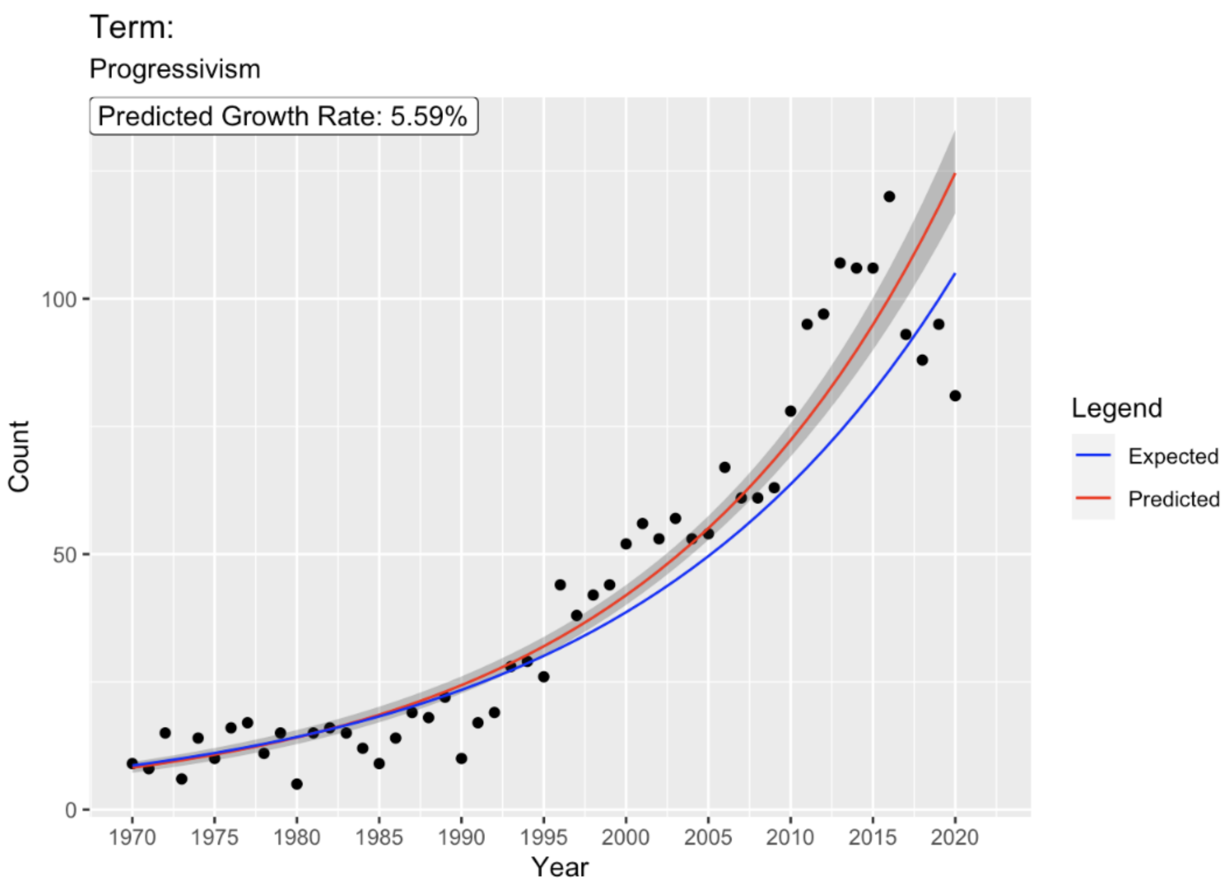

Term:
Patriotism
Predicted Growth Rate: 8.52%
1000
750
500
250
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

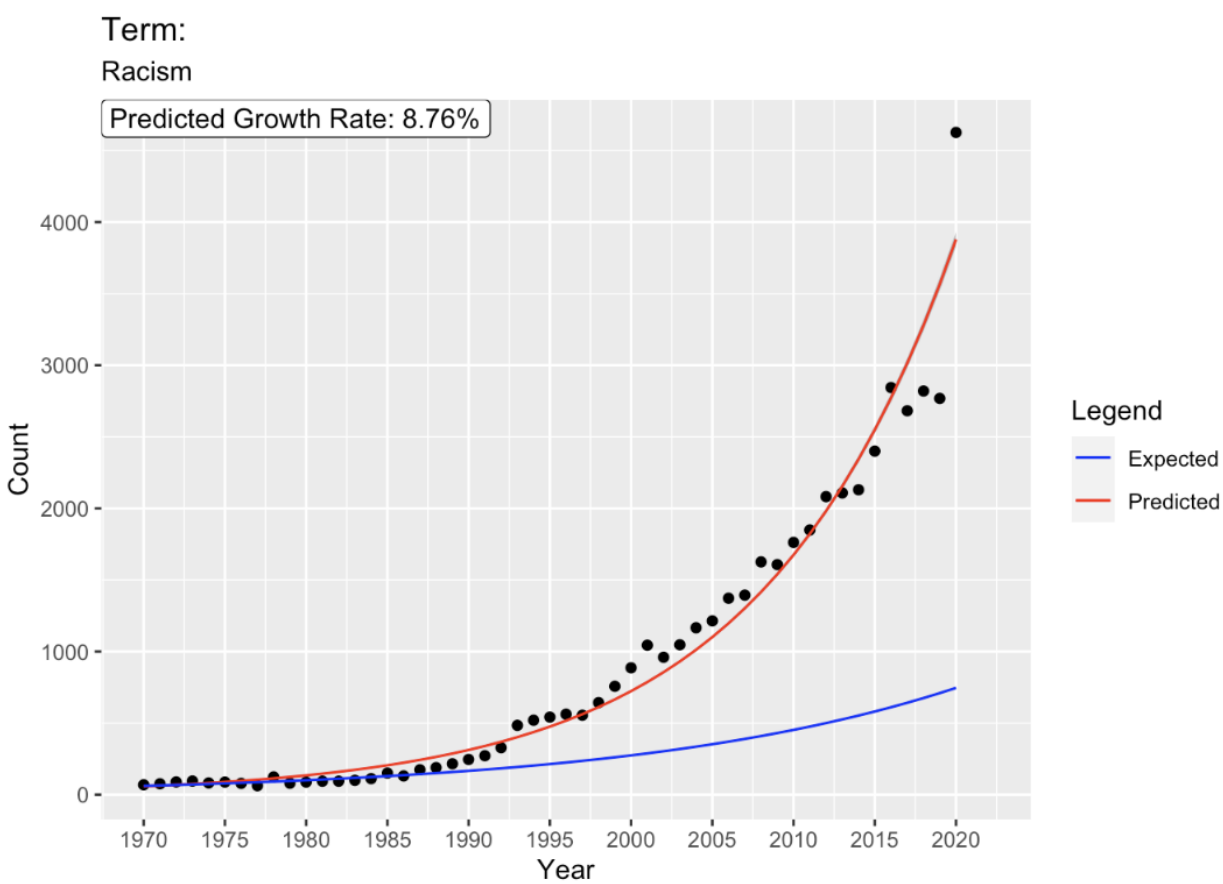

Term:
Political
Predicted Growth Rate: 3.64%
40000
30000
20000
10000
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

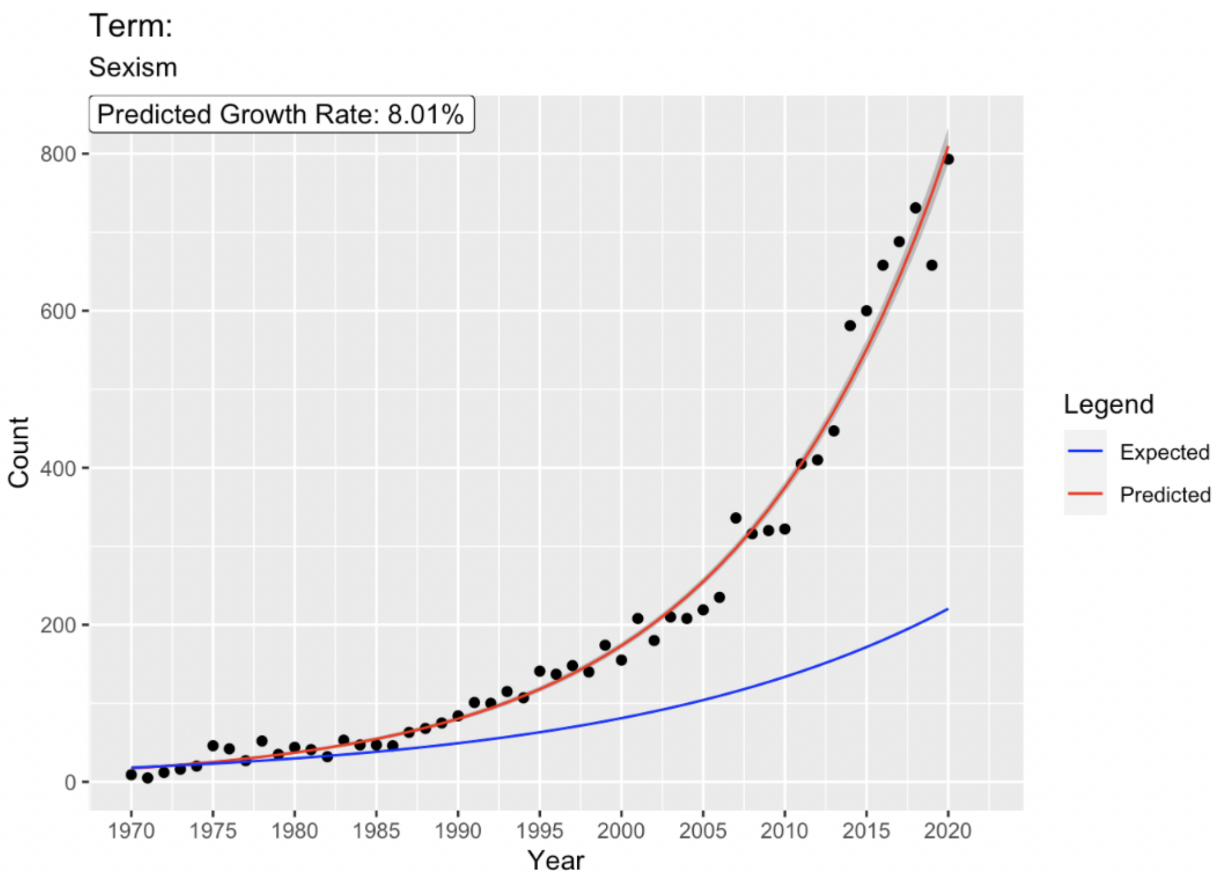
Term:
Politically correct
Predicted Growth Rate: 5.02%
Count
Year
Legend
Expected
Predicted



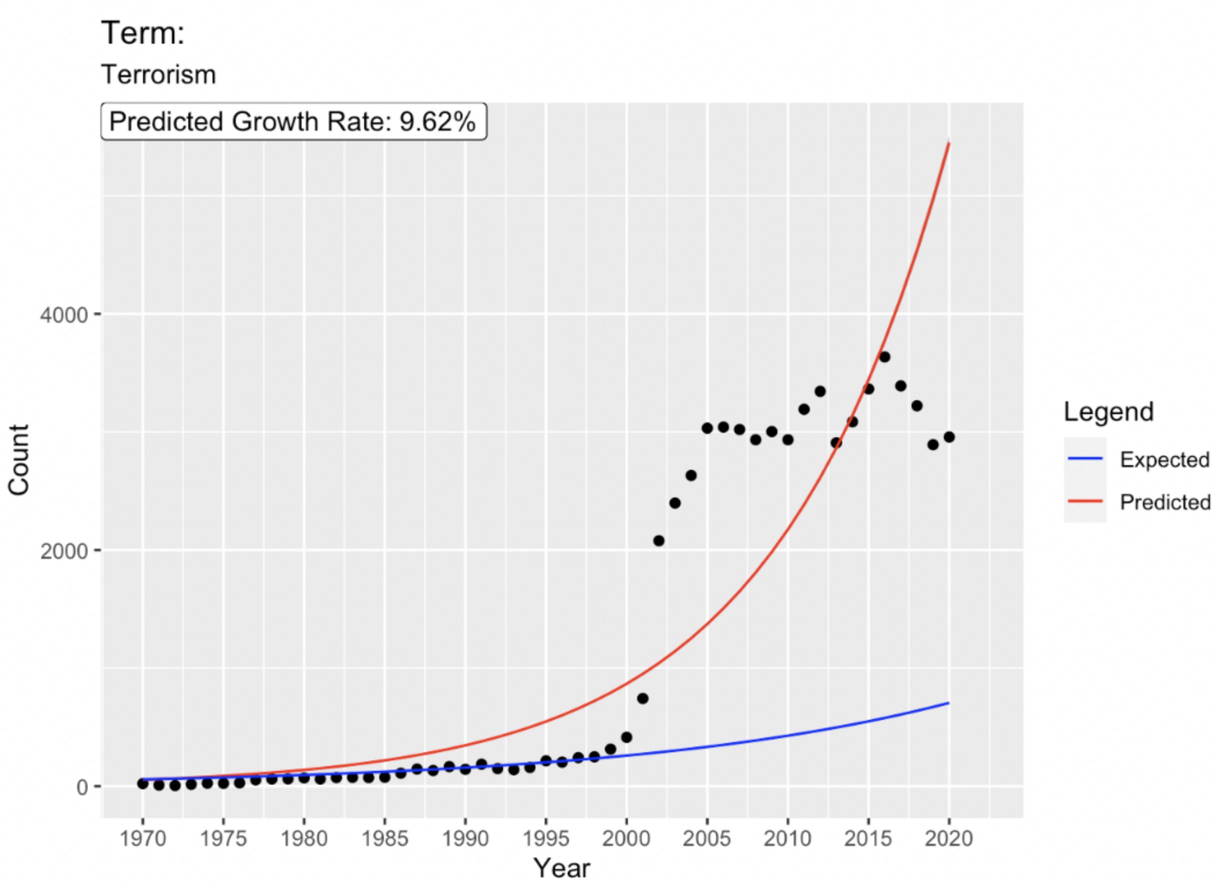
Term:
Prejudice
Predicted Growth Rate: 7.29%
Count
Year
Legend
Expected
Predicted


Term:
Progressivism
Predicted Growth Rate: 5.59%
Count
0
50
100
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

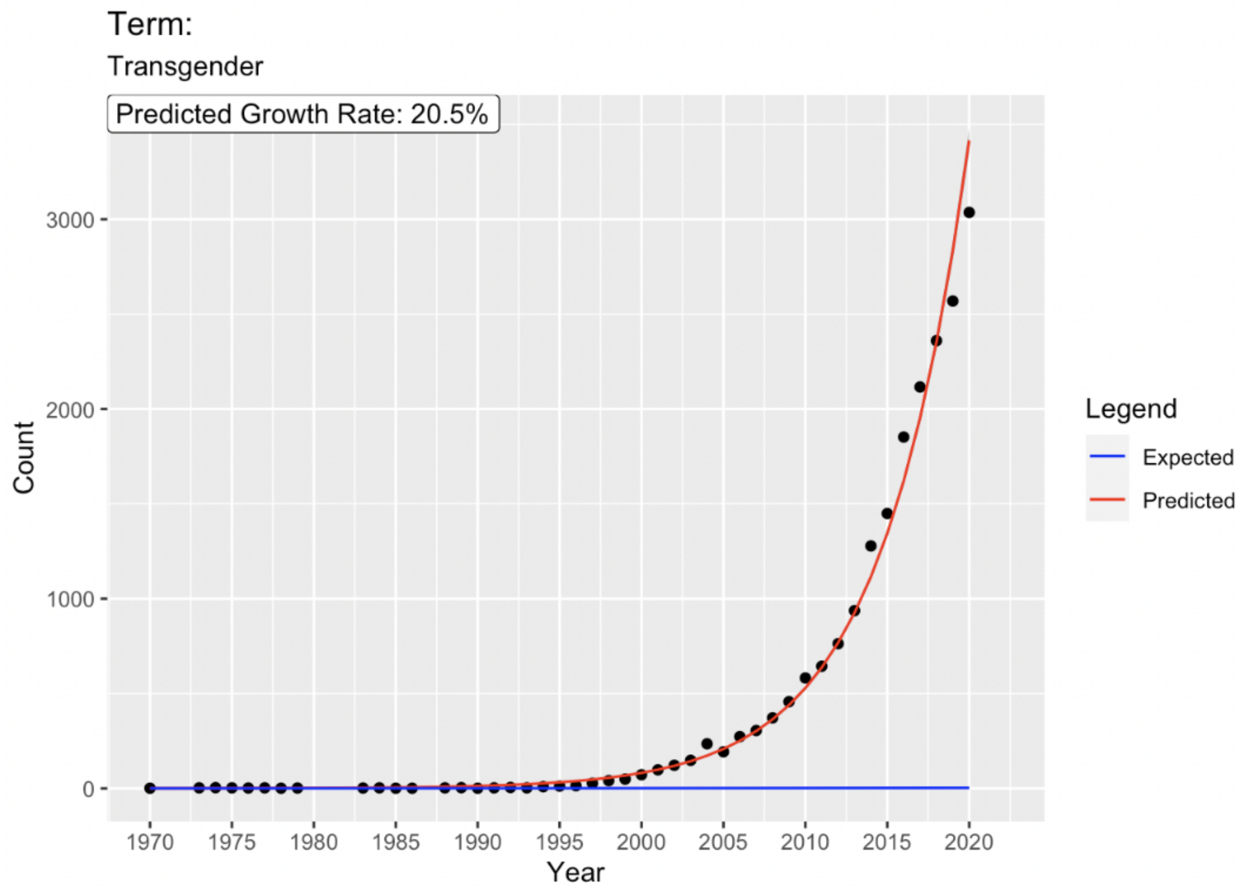

Term:
Racism
Predicted Growth Rate: 8.76%
Count
0
1000
2000
3000
4000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted



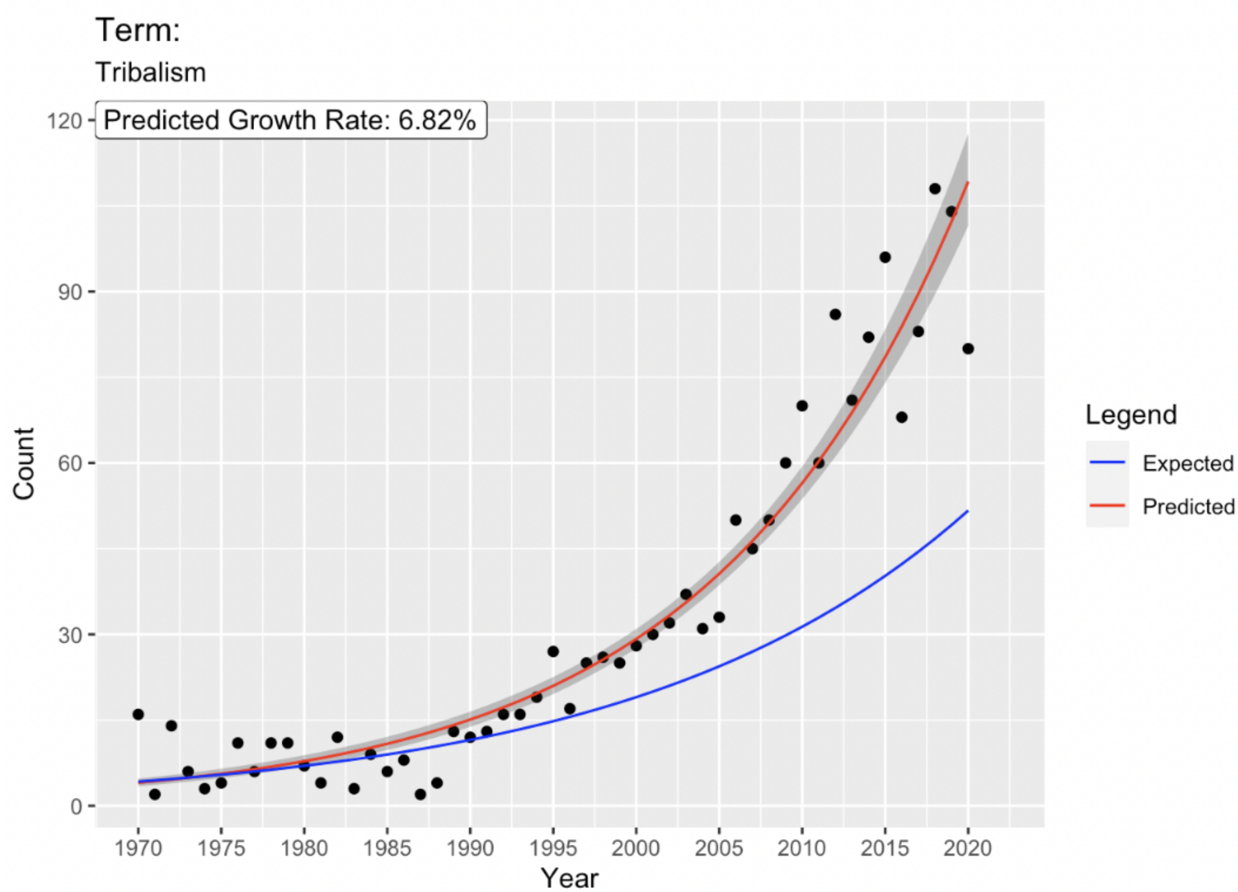
Term:
Sexism
Predicted Growth Rate: 8.01%
Count
0
200
400
600
800
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted



Term:
Terrorism
Predicted Growth Rate: 9.62%
Count
0
2000
4000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted



Term:
Transgender
Predicted Growth Rate: 20.5%
3000
2000
1000
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted



Term:
Tribalism
Predicted Growth Rate: 6.82%
120
90
60
30
0
Count
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Legend
Expected
Predicted

Stats Group Project

2023-04-08

```r
# delete rows with missing data (1100 deleted)
data <- na.omit(data)

# convert to data frame
data <- data.frame(data)

# capitalize first letter in term column
data$term <- capitalize(data$term)

# ensure all terms for each field in the data set are clean and showing as
expected
unique(data$term)
unique(data$field)
unique(data$year)

# 2021 and 2022 have yet to be completely populated, which is evidenced by
the below histogram.
# as such, we are going to set the range of this study to be from 1970 to
2020

# grouping data by year for plotting
years <- data %>%
  group_by(year) %>%
  dplyr::summarise(count = n())

# plotting the histogram
ggplot(years, aes(x = year, y = count)) +
  geom_histogram(binwidth = 1, stat = "identity", width = 0.7, color =
"black") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  # changing units of labels
  scale_y_continuous(labels = function(count) paste0(count/1000, "K")) +
  ggtitle("Total Count per Year")

## Warning in geom_histogram(binwidth = 1, stat = "identity", width = 0.7, :
## Ignoring unknown parameters: `binwidth`, `bins`, and `pad`

# subset data to 1970-2020 (deleting incomplete 2021, 2022)
data <- subset(data, data$year<=2020)
```

```
# we then will ensure terms look acceptable as they'll be the other focus of
our analysis

# grouping data by term for plotting
terms <- data %>%
  group_by(term) %>%
  dplyr::summarise(count = n())

# ordering the new data by count (ascending)
terms$term <- reorder(terms$term, terms$count)

# plot the bar chart
ggplot(terms, aes(x = term, y = count)) +
  geom_bar(stat = "identity", width = 0.7, color = "black") +
  # change x axis labels to be vertical
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  # changing units of labels
  scale_y_continuous(labels = function(count) paste0(count/1000, "K")) +
  ggtitle(("Terms by Count"))

# it does appear that we have some terms without few occurrences
comparitively.
# these 5 terms are the least frequent
terms %>% arrange(count) %>% head(5)

# Despite the lower numbers, we will proceed with full set of terms while
keeping this in mind.

# for the base of our analysis, we will will be looking at Term usage by
Year.
# grouping data by term and year, and summarizing the count of terms by year.
byterm <- data %>%
  group_by(term, year) %>%
  dplyr::summarise(count = n())

## `summarise()` has grouped output by 'term'. You can override using the
## `.groups` argument.

# pivot df so that terms are column name
data_wide <- spread(byterm, term, count)
cor(data_wide[,-1], method="spearman", use="pairwise.complete.obs")
data_cor <- rcorr(as.matrix(data_wide[,-1]))
```

```
# Extract the correlation coefficients
data_cor$r

# Extract p-values
data_cor$P

# ++++++++++++++++++++++++++++
# flattenCorrMatrix
# ++++++++++++++++++++++++++++
# cormat : matrix of the correlation coefficients
# pmat : matrix of the correlation p-values
flattenCorrMatrix <- function(cormat, pmat) {
  ut <- upper.tri(cormat)
  data.frame(
      row = rownames(cormat)[row(cormat)[ut]],
      column = rownames(cormat)[col(cormat)[ut]],
      cor  =(cormat)[ut],
      p = pmat[ut]
      )
}

flattenCorrMatrix(data_cor$r, data_cor$P)

##                        row                 column      cor          p
## 1                      Abortion                  Activism  0.94555844
0.000000e+00
## 2                      Abortion            Civil rights  0.95920370
0.000000e+00
## 3                      Activism            Civil rights  0.92857534
0.000000e+00
## 4                      Abortion            Conservative  0.94982526
0.000000e+00
## 5                      Activism            Conservative  0.87257920
0.000000e+00
## 6                Civil rights              Conservative  0.95195036
0.000000e+00
## 7                      Abortion            Discrimination  0.96385987
0.000000e+00
## 8                      Activism            Discrimination  0.92064883
0.000000e+00
## 9                Civil rights              Discrimination  0.98192836
0.000000e+00
## 10                 Conservative            Discrimination  0.97117136
0.000000e+00
```

```
## 11                   Abortion Diversity and inclusion  0.64876502
1.408160e-04
## 12                   Activism Diversity and inclusion  0.77057947
1.008736e-06
## 13             Civil rights Diversity and inclusion  0.45053162
1.418181e-02
## 14             Conservative Diversity and inclusion  0.45878054
1.230648e-02
## 15             Discrimination Diversity and inclusion  0.56183951
1.515752e-03
## 16                   Abortion                Fatphobia  0.55777460
5.950784e-02
## 17                   Activism                Fatphobia  0.65852253
1.988513e-02
## 18             Civil rights           Fatphobia  0.27591140 3.853664e-01
## 19             Conservative           Fatphobia  0.38256153 2.196943e-01
## 20             Discrimination                Fatphobia  0.53515756
7.298198e-02
## 21  Diversity and inclusion           Fatphobia  0.92546260 1.597316e-05
## 22                   Abortion                Feminism  0.97801320
0.000000e+00
## 23                   Activism                Feminism  0.96836422
0.000000e+00
## 24             Civil rights                  Feminism  0.97235219
0.000000e+00
## 25             Conservative                  Feminism  0.94579226
0.000000e+00
## 26             Discrimination                Feminism  0.97848452
0.000000e+00
## 27  Diversity and inclusion                  Feminism  0.68606121
3.985953e-05
## 28             Fatphobia              Feminism  0.59953110 3.935682e-02
## 29                   Abortion            Human rights  0.95878160
0.000000e+00
## 30                   Activism          Human rights  0.94361490
0.000000e+00
## 31             Civil rights            Human rights  0.98873092
0.000000e+00
## 32             Conservative            Human rights  0.94703142
0.000000e+00
## 33             Discrimination          Human rights  0.97880474
0.000000e+00
## 34  Diversity and inclusion           Human rights  0.50293975
```

```
5.422456e-03
## 35           Fatphobia        Human rights  0.37400788 2.310438e-01
## 36                 Feminism         Human rights  0.97110860
0.000000e+00
## 37                 Abortion         Immigration  0.97597613
0.000000e+00
## 38                 Activism         Immigration  0.97255809
0.000000e+00
## 39           Civil rights           Immigration  0.97917517
0.000000e+00
## 40           Conservative           Immigration  0.93674158
0.000000e+00
## 41           Discrimination         Immigration  0.96773222
0.000000e+00
## 42  Diversity and inclusion         Immigration  0.59107220
7.345891e-04
## 43           Fatphobia              Immigration  0.42271126
1.709905e-01
## 44                 Feminism         Immigration  0.98299791
0.000000e+00
## 45           Human rights             Immigration  0.98987087
0.000000e+00
## 46                 Abortion               Liberalism  0.95221570
0.000000e+00
## 47                 Activism               Liberalism  0.94211183
0.000000e+00
## 48           Civil rights           Liberalism  0.96540789 0.000000e+00
## 49           Conservative           Liberalism  0.91823039 0.000000e+00
## 50           Discrimination               Liberalism  0.95233280
0.000000e+00
## 51  Diversity and inclusion         Liberalism  0.52181188 3.693427e-03
## 52           Fatphobia              Liberalism  0.29120224 3.584571e-01
## 53                 Feminism               Liberalism  0.96182006
0.000000e+00
## 54           Human rights           Liberalism  0.95984196 0.000000e+00
## 55           Immigration            Liberalism  0.96998418 0.000000e+00
## 56                 Abortion               Misogyny  0.89604793
0.000000e+00
## 57                 Activism               Misogyny  0.96238163
0.000000e+00
## 58           Civil rights                 Misogyny  0.84796590
1.036948e-13
## 59           Conservative                 Misogyny  0.81855018
```

```
3.687717e-12
## 60            Discrimination                Misogyny  0.87257908
2.664535e-15
## 61  Diversity and inclusion                 Misogyny  0.90242890
2.238565e-11
## 62            Fatphobia               Misogyny  0.88002125 1.595477e-04
## 63                  Feminism                Misogyny  0.93285230
0.000000e+00
## 64            Human rights                 Misogyny  0.85799382
2.575717e-14
## 65            Immigration             Misogyny  0.89726366 0.000000e+00
## 66            Liberalism              Misogyny  0.86553159 8.437695e-15
## 67                  Abortion          Nationalism  0.96941144
0.000000e+00
## 68                  Activism          Nationalism  0.95836009
0.000000e+00
## 69            Civil rights            Nationalism  0.98923076
0.000000e+00
## 70            Conservative            Nationalism  0.95062455
0.000000e+00
## 71            Discrimination          Nationalism  0.98745219
0.000000e+00
## 72  Diversity and inclusion           Nationalism  0.58758985
8.036761e-04
## 73            Fatphobia               Nationalism  0.49523028
1.016160e-01
## 74                  Feminism          Nationalism  0.98880332
0.000000e+00
## 75            Human rights            Nationalism  0.99266068
0.000000e+00
## 76               Immigration          Nationalism  0.99057294
0.000000e+00
## 77            Liberalism              Nationalism  0.96957606
0.000000e+00
## 78                  Misogyny          Nationalism  0.89440432
0.000000e+00
## 79                  Abortion              Patriarchy  0.96129176
0.000000e+00
## 80                  Activism                Patriarchy  0.98701364
0.000000e+00
## 81            Civil rights            Patriarchy  0.93878823 0.000000e+00
## 82            Conservative            Patriarchy  0.89996815 0.000000e+00
## 83            Discrimination                Patriarchy  0.94390475
```

```
0.000000e+00
## 84  Diversity and inclusion           Patriarchy  0.80002177 1.908001e-07
## 85             Fatphobia              Patriarchy  0.76151702 4.005402e-03
## 86                  Feminism                Patriarchy  0.98201630
0.000000e+00
## 87             Human rights           Patriarchy  0.94318655 0.000000e+00
## 88             Immigration            Patriarchy  0.96773373 0.000000e+00
## 89             Liberalism             Patriarchy  0.93677000 0.000000e+00
## 90                  Misogyny                Patriarchy  0.97218734
0.000000e+00
## 91             Nationalism            Patriarchy  0.96418735 0.000000e+00
## 92                  Abortion                Patriotism  0.95228512
0.000000e+00
## 93                  Activism                Patriotism  0.94743832
0.000000e+00
## 94             Civil rights           Patriotism  0.98672602 0.000000e+00
## 95             Conservative              Patriotism  0.92818996
0.000000e+00
## 96             Discrimination               Patriotism  0.97447735
0.000000e+00
## 97  Diversity and inclusion           Patriotism  0.54952270 2.017184e-03
## 98             Fatphobia              Patriotism  0.49621429 1.008322e-01
## 99                  Feminism                Patriotism  0.96716312
0.000000e+00
## 100            Human rights           Patriotism  0.99077311 0.000000e+00
## 101            Immigration            Patriotism  0.98164999 0.000000e+00
## 102            Liberalism             Patriotism  0.95710959 0.000000e+00
## 103                 Misogyny                Patriotism  0.87289328
2.664535e-15
## 104            Nationalism            Patriotism  0.98804861 0.000000e+00
## 105            Patriarchy             Patriotism  0.95150903 0.000000e+00
## 106                 Abortion                Political  0.97993952
0.000000e+00
## 107                 Activism                Political  0.91841598
0.000000e+00
## 108            Civil rights           Political  0.97919412 0.000000e+00
## 109            Conservative           Political  0.97208015 0.000000e+00
## 110            Discrimination               Political  0.99013502
0.000000e+00
## 111 Diversity and inclusion           Political  0.55255249 1.882141e-03
## 112            Fatphobia              Political  0.45468572 1.375212e-01
## 113                 Feminism               Political  0.98098079
0.000000e+00
```

```
## 114           Human rights           Political  0.96809015 0.000000e+00
## 115           Immigration            Political  0.96843186 0.000000e+00
## 116           Liberalism             Political  0.95751766 0.000000e+00
## 117                 Misogyny               Political  0.87349994
2.442491e-15
## 118           Nationalism            Political  0.97965735 0.000000e+00
## 119           Patriarchy             Political  0.94569531 0.000000e+00
## 120           Patriotism             Political  0.96028753 0.000000e+00
## 121                 Abortion    Politically correct  0.79716844
2.108255e-09
## 122                 Activism    Politically correct  0.69539483
1.262418e-06
## 123           Civil rights      Politically correct  0.87801446
4.507505e-13
## 124           Conservative      Politically correct  0.83593425
6.509548e-11
## 125           Discrimination    Politically correct  0.88676524
1.272316e-13
## 126 Diversity and inclusion     Politically correct  0.06909899
7.268001e-01
## 127           Fatphobia   Politically correct -0.20435444 5.240670e-01
## 128                 Feminism    Politically correct  0.81747056
3.777445e-10
## 129           Human rights     Politically correct  0.85077758
1.343015e-11
## 130           Immigration       Politically correct  0.80563375
1.054774e-09
## 131           Liberalism  Politically correct  0.84221625 3.404343e-11
## 132                 Misogyny    Politically correct  0.60872319
5.012298e-05
## 133           Nationalism       Politically correct  0.85652536
6.958878e-12
## 134           Patriarchy  Politically correct  0.72702475 2.350362e-07
## 135           Patriotism  Politically correct  0.83700693 5.838774e-11
## 136           Political   Politically correct  0.88587741 1.452172e-13
## 137                 Abortion               Prejudice  0.96984516
0.000000e+00
## 138                 Activism               Prejudice  0.98505515
0.000000e+00
## 139           Civil rights           Prejudice  0.97591912 0.000000e+00
## 140           Conservative           Prejudice  0.92278068 0.000000e+00
## 141           Discrimination               Prejudice  0.96306030
0.000000e+00
```

```
## 142 Diversity and inclusion           Prejudice  0.65440420 1.176317e-04
## 143            Fatphobia              Prejudice  0.52212772 8.162588e-02
## 144                  Feminism              Prejudice  0.98599812
0.000000e+00
## 145            Human rights              Prejudice  0.98219214
0.000000e+00
## 146            Immigration            Prejudice  0.99467617 0.000000e+00
## 147            Liberalism             Prejudice  0.96790153 0.000000e+00
## 148                  Misogyny              Prejudice  0.92564113
0.000000e+00
## 149            Nationalism            Prejudice  0.98917971 0.000000e+00
## 150            Patriarchy             Prejudice  0.98126247 0.000000e+00
## 151            Patriotism             Prejudice  0.98079445 0.000000e+00
## 152            Political              Prejudice  0.96168449 0.000000e+00
## 153      Politically correct              Prejudice  0.78953657
3.830429e-09
## 154                  Abortion         Progressivism  0.95453544
0.000000e+00
## 155                  Activism         Progressivism  0.94719071
0.000000e+00
## 156            Civil rights           Progressivism  0.97327115
0.000000e+00
## 157            Conservative           Progressivism  0.91968365
0.000000e+00
## 158            Discrimination         Progressivism  0.96186542
0.000000e+00
## 159 Diversity and inclusion           Progressivism  0.56297652
1.475464e-03
## 160                 Fatphobia         Progressivism  0.32162084
3.080013e-01
## 161                  Feminism         Progressivism  0.97703048
0.000000e+00
## 162            Human rights           Progressivism  0.96174986
0.000000e+00
## 163            Immigration          Progressivism  0.97301693
0.000000e+00
## 164            Liberalism       Progressivism  0.97332089 0.000000e+00
## 165                  Misogyny         Progressivism  0.88265429
4.440892e-16
## 166            Nationalism            Progressivism  0.97990863
0.000000e+00
## 167            Patriarchy       Progressivism  0.95313184 0.000000e+00
## 168            Patriotism       Progressivism  0.95987341 0.000000e+00
```

```
## 169           Political        Progressivism  0.96515620 0.000000e+00
## 170      Politically correct         Progressivism  0.83656249
6.108447e-11
## 171           Prejudice        Progressivism  0.97558262 0.000000e+00
## 172                 Abortion                Racism  0.92349005
0.000000e+00
## 173                 Activism                Racism  0.95756416
0.000000e+00
## 174           Civil rights                  Racism  0.88853388
0.000000e+00
## 175           Conservative                  Racism  0.86541508
2.220446e-16
## 176           Discrimination                Racism  0.91375882
0.000000e+00
## 177 Diversity and inclusion                 Racism  0.88797887
1.322547e-10
## 178           Fatphobia              Racism  0.91809579 2.526712e-05
## 179                 Feminism                  Racism  0.94347346
0.000000e+00
## 180           Human rights                  Racism  0.90258956
0.000000e+00
## 181           Immigration                   Racism  0.92132613
0.000000e+00
## 182           Liberalism                    Racism  0.88748596
0.000000e+00
## 183                 Misogyny                Racism  0.97345091
0.000000e+00
## 184           Nationalism                   Racism  0.92787462
0.000000e+00
## 185           Patriarchy                    Racism  0.97570915
0.000000e+00
## 186           Patriotism                    Racism  0.91676706
0.000000e+00
## 187           Political              Racism  0.90273321 0.000000e+00
## 188      Politically correct                Racism  0.67948251
2.720260e-06
## 189           Prejudice              Racism  0.94301235 0.000000e+00
## 190           Progressivism                 Racism  0.90084841
0.000000e+00
## 191                 Abortion                Sexism  0.92892661
0.000000e+00
## 192                 Activism                Sexism  0.98116027
0.000000e+00
```

```
## 193           Civil rights                     Sexism  0.88872500
0.000000e+00
## 194           Conservative                     Sexism  0.85575422
1.332268e-15
## 195           Discrimination                     Sexism  0.90523763
0.000000e+00
## 196 Diversity and inclusion                    Sexism  0.86558752
1.348925e-09
## 197           Fatphobia                 Sexism  0.80774074 1.483219e-03
## 198                 Feminism                  Sexism  0.95478491
0.000000e+00
## 199           Human rights                     Sexism  0.90154314
0.000000e+00
## 200           Immigration                      Sexism  0.93700003
0.000000e+00
## 201           Liberalism                       Sexism  0.90267769
0.000000e+00
## 202                 Misogyny                  Sexism  0.99091356
0.000000e+00
## 203           Nationalism                      Sexism  0.92844479
0.000000e+00
## 204           Patriarchy                       Sexism  0.98568669
0.000000e+00
## 205           Patriotism                       Sexism  0.91101328
0.000000e+00
## 206           Political                 Sexism  0.90446619 0.000000e+00
## 207      Politically correct                   Sexism  0.64446604
1.261844e-05
## 208           Prejudice                 Sexism  0.95566523 0.000000e+00
## 209           Progressivism                    Sexism  0.91500710
0.000000e+00
## 210                   Racism                  Sexism  0.97439418
0.000000e+00
## 211                 Abortion                  Terrorism  0.90682761
0.000000e+00
## 212                 Activism                  Terrorism  0.90068548
0.000000e+00
## 213           Civil rights                 Terrorism  0.96011662
0.000000e+00
## 214           Conservative              Terrorism  0.90906155 0.000000e+00
## 215           Discrimination                   Terrorism  0.96000773
0.000000e+00
## 216 Diversity and inclusion              Terrorism  0.46393179 1.124401e-02
```

```
## 217          Fatphobia               Terrorism  0.41491081 1.798659e-01
## 218                Feminism                Terrorism  0.92506449
0.000000e+00
## 219          Human rights            Terrorism  0.97115303 0.000000e+00
## 220          Immigration             Terrorism  0.94466767 0.000000e+00
## 221          Liberalism              Terrorism  0.92696426 0.000000e+00
## 222                Misogyny                Terrorism  0.81950541
3.318013e-12
## 223          Nationalism             Terrorism  0.95890613 0.000000e+00
## 224          Patriarchy              Terrorism  0.89914021 0.000000e+00
## 225          Patriotism              Terrorism  0.97458303 0.000000e+00
## 226               Political                Terrorism  0.93269897
0.000000e+00
## 227      Politically correct               Terrorism  0.82412939
2.051206e-10
## 228          Prejudice               Terrorism  0.94229323 0.000000e+00
## 229          Progressivism               Terrorism  0.91531368
0.000000e+00
## 230                Racism            Terrorism  0.87185335 0.000000e+00
## 231                Sexism            Terrorism  0.86207193 4.440892e-16
## 232                Abortion          Transgender  0.78283502
2.103659e-10
## 233                Activism          Transgender  0.90533105
0.000000e+00
## 234          Civil rights            Transgender  0.69616429
1.097801e-07
## 235          Conservative            Transgender  0.66059549
7.891103e-07
## 236          Discrimination          Transgender  0.72811872
1.443048e-08
## 237 Diversity and inclusion          Transgender  0.96405823
0.000000e+00
## 238          Fatphobia               Transgender  0.87538235
1.913136e-04
## 239                Feminism          Transgender  0.81726335
7.453815e-12
## 240          Human rights            Transgender  0.72909626
1.350177e-08
## 241          Immigration             Transgender  0.78934167
1.173814e-10
## 242              Liberalism          Transgender  0.74179170
5.541201e-09
## 243                Misogyny          Transgender  0.95915257
```

```
0.000000e+00
## 244           Nationalism            Transgender  0.76858982
7.058865e-10
## 245           Patriarchy            Transgender  0.89632931 0.000000e+00
## 246           Patriotism             Transgender  0.75315391
2.387842e-09
## 247           Political              Transgender  0.72091544
2.335752e-08
## 248      Politically correct         Transgender  0.39483647
1.716536e-02
## 249           Prejudice              Transgender  0.82663140
2.656764e-12
## 250           Progressivism          Transgender  0.75360546
2.307132e-09
## 251                 Racism           Transgender  0.92098649
0.000000e+00
## 252                 Sexism           Transgender  0.94435190
0.000000e+00
## 253           Terrorism              Transgender  0.69723582
1.029958e-07
## 254                 Abortion               Tribalism  0.93907201
0.000000e+00
## 255                 Activism               Tribalism  0.96786120
0.000000e+00
## 256           Civil rights           Tribalism  0.91673401 0.000000e+00
## 257           Conservative           Tribalism  0.88066354 0.000000e+00
## 258           Discrimination               Tribalism  0.92827119
0.000000e+00
## 259 Diversity and inclusion          Tribalism  0.72073069 1.035023e-05
## 260                Fatphobia               Tribalism  0.59623644
4.073881e-02
## 261                 Feminism               Tribalism  0.96463675
0.000000e+00
## 262           Human rights           Tribalism  0.93629177 0.000000e+00
## 263           Immigration              Tribalism  0.95667314
0.000000e+00
## 264           Liberalism             Tribalism  0.93282924 0.000000e+00
## 265                 Misogyny               Tribalism  0.92176665
0.000000e+00
## 266           Nationalism            Tribalism  0.95443171 0.000000e+00
## 267           Patriarchy             Tribalism  0.96024092 0.000000e+00
## 268           Patriotism             Tribalism  0.93298624 0.000000e+00
## 269           Political              Tribalism  0.92168248 0.000000e+00
```

```
## 270      Politically correct                Tribalism  0.72443557
2.720280e-07
## 271            Prejudice              Tribalism  0.95882857 0.000000e+00
## 272            Progressivism                Tribalism  0.93297667
0.000000e+00
## 273                  Racism              Tribalism  0.92538987 0.000000e+00
## 274                  Sexism              Tribalism  0.95061449 0.000000e+00
## 275            Terrorism              Tribalism  0.89275259 0.000000e+00
## 276               Transgender                Tribalism  0.85892114
4.440892e-14
```

Source to model a time-count exponential curve:
[*https://stats.stackexchange.com/questions/261769/problems-with-plotting-exponential-curve-and-data-in-the-same-plot-when-values-a*](https://stats.stackexchange.com/questions/261769/problems-with-plotting-exponential-curve-and-data-in-the-same-plot-when-values-a)

Reference code for loop:
[*https://stackoverflow.com/questions/64500868/how-do-i-create-multiple-plots-by-looping-through-a-vertical-data-frame-in-r*](https://stackoverflow.com/questions/64500868/how-do-i-create-multiple-plots-by-looping-through-a-vertical-data-frame-in-r)

```
#initializing temporary table to store pvalues
tmp_table = matrix(ncol=3)

# save one plot per term
# looping over each individual term
for (i in unique(byterm$term)) {

  # filtering individual term into temporary vector
  xterm <- byterm %>%filter(term==i)

  # fitting model based on reference code
  mod <- glm(count~year, data = xterm, family = "poisson")

  # length.out measures length of x-axis which is 50 (1970-2020)
  pred.df <- data.frame(year = seq(min(xterm$year), max(xterm$year),
length.out = 51))

  # creating prediction model
  pred <- predict(mod, newdata = pred.df, se.fit = TRUE)

  # fitting exp count
  pred.df$count <- exp(pred$fit)

  # generating lowest count value for ribbon (CI)
  pred.df$countmin <- exp(pred$fit - 2 * pred$se.fit)
```

```
# generating highest count value for ribbon (CI)
pred.df$countmax <- exp(pred$fit + 2 * pred$se.fit)

# calculate the percentage growth rate
pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100

# Set up time variable to get year range for each term
time <- length(xterm$year)
range <- 1:time

# Define exponential growth function
exp_growth <- function(x, a, b) a * exp(b * x)

# Set growth rate
growth_rate <- 0.05  # 5% per year

# Get intercept value from predicted exponential curve
start_count <- pred.df$count[1]
articles <- exp_growth(range, start_count, growth_rate)

# data frame with year range and expected count
expected.df <- data.frame(year = xterm$year, count = articles)

# plotting model
colors <- c("Predicted" = "red", "Expected" = "blue")
plot <- ggplot(xterm,aes(x=year,y=count)) +

# setting x-limits for count
scale_x_continuous(limits=c(1970,2022),breaks=c(1970,
                                                1975,1980,1985,1990,1995,
                                                2000,2005,2010,2015,2020)) +
geom_point() +
geom_ribbon(data = pred.df, aes(ymin = countmin, ymax = countmax), alpha =
0.3) +

# plotting fitted curves
geom_line(data = pred.df, aes(y = count, color="Predicted")) +
geom_line(data = expected.df, aes(y = count, color="Expected")) +
annotate(geom = 'label', label = paste0("Predicted Growth Rate: ",
round(pred_rate, digits = 2), "%"),
            x = -Inf, y = Inf, hjust = 0, vjust = 1) +
labs(x = "Year",
```

```
      y = "Count",
      color = "Legend") +
      scale_color_manual(values = colors) +
  ggtitle("Term:", i)

  # showing plot
  plot(plot)

  #running one-sided poisson test for expected vs predicted growth rates
  ptest <- poisson.test(
  x = c(sum(as.integer(expected.df$count)), sum(as.integer(pred.df$count))),
  T = c(length(expected.df$count), length(pred.df$count)),
  alternative = 'less'
  )

  #saving and formatting pvalue
  pvalue <- as.numeric(ptest[3])
  results = c(i, pred_rate, pvalue)
  #saving term and pvalue to table
  tmp_table<-rbind(tmp_table,results)

}

#formatting stats_df
stats_df <- data.frame(tmp_table)
colnames(stats_df) <- c("Terms", "Predicted Growth Rate", "P Values")
stats_df <- stats_df[-1,]
rownames(stats_df) <- 1:nrow(stats_df)
stats_df$`P Values` <- signif(as.numeric(stats_df$`P Values`), 5)
stats_df$`Predicted Growth Rate` <- round(as.numeric(stats_df$`Predicted
Growth Rate`), 2)
stats_df$`Predicted Growth Rate` <- paste0(stats_df$`Predicted Growth Rate`,
"%")
stats_df

# Terrorism
terrorism <- subset(byterm, byterm$term == "Terrorism")
d1 = data.frame(subset(terrorism, terrorism$year <=1980))
d2 = data.frame(subset(terrorism, terrorism$year >=1980 & terrorism$year
<=1990))
d3 = data.frame(subset(terrorism, terrorism$year >=1990 & terrorism$year
<=2000))
d4 = data.frame(subset(terrorism, terrorism$year >=2000 & terrorism$year
<=2010))
```

```
d5 = data.frame(subset(terrorism, terrorism$year >=2010 & terrorism$year
<=2020))
yr_list = list(d1,d2,d3,d4,d5,terrorism)
yr_tmp = list()
for (i in 1:6){
  dt = yr_list[[i]]
  # fitting model based on reference code
  mod <- glm(count~year, data = dt, family = "poisson")
  # calculate the percentage growth rate
  pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100
  yr_tmp[[i]] <- pred_rate
}
#Convert list to dataframe
yr_df <- data.frame(yr_tmp)
#Add correct column name per decade
names(yr_df) <-
c("1970-1980","1980-1990","1990-2000","2000-2010","2010-2020", "1970-2020")
#Round numbers to 2 digits and add %
yr_df <- round(yr_df, 2)
yr_df[] <- Map(paste, yr_df, "%")
yr_df

##   1970-1980 1980-1990 1990-2000 2000-2010 2010-2020 1970-2020
## 1   21.78 %   10.93 %   10.75 %   10.58 %    0.02 %      9.62 %

#to export to png
# library(gridExtra)
# png("yr_df.png", height = 50*nrow(yr_df), width = 200*ncol(yr_df))
# grid.table(yr_df)
# dev.off()

# Subset term
terrorism <- subset(byterm, byterm$term == 'Terrorism')

# fitting model based on reference code
mod <- glm(count~year, data = terrorism, family = "poisson")

# length.out measures length of x-axis which is 51 (1970-2020)
pred.df2 <- data.frame(year = seq(min(terrorism$year), max(terrorism$year),
length.out = 51))

# creating prediction model
pred <- predict(mod, newdata = pred.df, se.fit = TRUE)
```

```
# fitting exp count
pred.df2$count <- exp(pred$fit)

# generating lowest count value for ribbon (CI)
pred.df2$countmin <- exp(pred$fit - 2 * pred$se.fit)

# generating highest count value for ribbon (CI)
pred.df2$countmax <- exp(pred$fit + 2 * pred$se.fit)

# calculate the percentage growth rate
pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100


colors <- c("Predicted" = "red", "Expected" = "blue")
ggplot(terrorism,aes(x=year,y=count)) +
  scale_x_continuous(limits=c(1970,2022),breaks=c(1970,
                                                    1975,1980,1985,1990,1995,
                                                2000,2005,2010,2015,2020)) +
  geom_point() +
  geom_ribbon(data = pred.df2, aes(ymin = countmin, ymax = countmax), alpha =
0.3) +
  geom_line(data = pred.df2, aes(y = count, color="Predicted")) +
  # Add vertical lines
  geom_vline(xintercept = c(1970,1980,1990,2000,2010,2020),
             linetype="dotted") +
  # Add growth rate per decade
  geom_segment(aes(x = 1970, y = 750, xend = 1980, yend = 750))+
  geom_text(aes(1975,
             y = 750,
             label = "21.78 %"),
             color = 'blue',
             size = 4,
             nudge_y = 120) +
  geom_segment(aes(x = 1980, y = 1500, xend = 1990, yend = 1500))+
  geom_text(aes(1985,
             y = 1500,
             label = "10.93 %"),
             color = 'blue',
             size = 4,
             nudge_y = 120) +
  geom_segment(aes(x = 1990, y = 2500, xend = 2000, yend = 2500))+
  geom_text(aes(1995,
             y = 2500,
```

```
            label = "10.75 %"),
            color = 'blue',
            size = 4,
            nudge_y = 120) +
  geom_segment(aes(x = 2000, y = 3250, xend = 2010, yend = 3250))+
  geom_text(aes(2005,
            y = 3250,
            label = "10.58 %"),
            color = 'blue',
            size = 4,
            nudge_y = 120) +
  geom_segment(aes(x = 2010, y = 4000, xend = 2020, yend = 4000))+
  geom_text(aes(2015,
            y = 4000,
            label = "0.02 %"),
          color = 'blue',
            size = 4,
            nudge_y = 120) +
  annotate(geom = 'label', label = paste0("Overall Predicted Growth Rate: ",
round(pred_rate, digits = 2), "%"),
            x = -Inf, y = Inf, hjust = 0, vjust = 1) +
  labs(x = "Year",
      y = "Count",
      color = "Legend") +
  ggtitle("Terrorism") +
  scale_color_manual(values = colors)

# Diversity and Inclusion
# Not enough values until 2000
diversity <- subset(byterm, byterm$term == "Diversity and inclusion")
d4 = data.frame(subset(diversity, diversity$year >=2000 & diversity$year
<=2010))
d5 = data.frame(subset(diversity, diversity$year >=2010 & diversity$year
<=2020))
yr_list = list(d4,d5,diversity)
yr_tmp = list()
for (i in 1:3){
  dt = yr_list[[i]]
  # fitting model based on reference code
  mod <- glm(count~year, data = dt, family = "poisson")
  # calculate the percentage growth rate
  pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100
  yr_tmp[[i]] <- pred_rate
}
```

```r
#Convert list to dataframe
yr_df <- data.frame(yr_tmp)
#Add correct column name per decade
names(yr_df) <- c("2000-2010","2010-2020","1970-2020")
#Round numbers to 2 digits and add %
yr_df <- round(yr_df, 2)
yr_df[] <- Map(paste, yr_df, "%")
yr_df

##   2000-2010 2010-2020 1970-2020
## 1   27.86 %   30.12 %   29.24 %

# Subset term
diversity <- subset(byterm, byterm$term == 'Diversity and inclusion')

# fitting model based on reference code
mod <- glm(count~year, data = diversity, family = "poisson")

# length.out measures length of x-axis which is 51 (1970-2020)
pred.df2 <- data.frame(year = seq(min(diversity$year), max(diversity$year),
length.out = 51))

# creating prediction model
pred <- predict(mod, newdata = pred.df, se.fit = TRUE)

# fitting exp count
pred.df2$count <- exp(pred$fit)

# generating lowest count value for ribbon (CI)
pred.df2$countmin <- exp(pred$fit - 2 * pred$se.fit)

# generating highest count value for ribbon (CI)
pred.df2$countmax <- exp(pred$fit + 2 * pred$se.fit)

# calculate the percentage growth rate
pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100


colors <- c("Predicted" = "red", "Expected" = "blue")
ggplot(diversity,aes(x=year,y=count)) +
  scale_x_continuous(limits=c(1970,2022),breaks=c(1970,
                                                 1975,1980,1985,1990,1995,
                                                 2000,2005,2010,2015,2020)) +
  geom_point() +
```

```
  geom_ribbon(data = pred.df2, aes(ymin = countmin, ymax = countmax), alpha =
0.3) +
  geom_line(data = pred.df2, aes(y = count, color="Predicted")) +
  # Add vertical lines
  geom_vline(xintercept = c(1970,1980,1990,2000,2010,2020),
             linetype="dotted") +
  geom_segment(aes(x = 1970, y = 100, xend = 1980, yend = 100))+
  # Add growth rate per decade
  geom_text(aes(1975,
             y = 100,
             label = "N/A"),
             color = 'blue',
             size = 4,
             nudge_y = 12) +
  geom_segment(aes(x = 1980, y = 200, xend = 1990, yend = 200))+
  geom_text(aes(1985,
             y = 200,
             label = "N/A"),
             color = 'blue',
             size = 4,
             nudge_y = 12) +
  geom_segment(aes(x = 1990, y = 300, xend = 2000, yend = 300))+
  geom_text(aes(1995,
             y = 300,
             label = "N/A"),
             color = 'blue',
             size = 4,
             nudge_y = 12) +
  geom_segment(aes(x = 2000, y = 400, xend = 2010, yend = 400))+
  geom_text(aes(2005,
             y = 400,
             label = "27.86 %"),
             color = 'blue',
             size = 4,
             nudge_y = 12) +
  geom_segment(aes(x = 2010, y = 500, xend = 2020, yend = 500))+
  geom_text(aes(2015,
             y = 500,
             label = "30.12 %"),
             color = 'blue',
             size = 4,
             nudge_y = 12) +
  annotate(geom = 'label', label = paste0("Overall Predicted Growth Rate: ",
```

```
round(pred_rate, digits = 2), "%"),
            x = -Inf, y = Inf, hjust = 0, vjust = 1) +
  ggtitle("Diversity and Inclusion") +
  labs(x = "Year",
      y = "Count",
      color = "Legend") +
      scale_color_manual(values = colors)

# Racism
racism <- subset(byterm, byterm$term == "Racism")
d1 = data.frame(subset(racism, racism$year <=1980))
d2 = data.frame(subset(racism, racism$year >=1980 & racism$year <=1990))
d3 = data.frame(subset(racism, racism$year >=1990 & racism$year <=2000))
d4 = data.frame(subset(racism, racism$year >=2000 & racism$year <=2010))
d5 = data.frame(subset(racism, racism$year >=2010 & racism$year <=2020))
yr_list = list(d1,d2,d3,d4,d5, racism)
yr_tmp = list()
for (i in 1:6){
  dt = yr_list[[i]]
  # fitting model based on reference code
  mod <- glm(count~year, data = dt, family = "poisson")
  # calculate the percentage growth rate
  pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100
  yr_tmp[[i]] <- pred_rate
}
#Convert list to dataframe
yr_df <- data.frame(yr_tmp)
#Add correct column name per decade
names(yr_df) <-
c("1970-1980","1980-1990","1990-2000","2000-2010","2010-2020","1970-2020")
#Round numbers to 2 digits and add %
yr_df <- round(yr_df, 2)
yr_df[] <- Map(paste, yr_df, "%")
yr_df

##   1970-1980 1980-1990 1990-2000 2000-2010 2010-2020 1970-2020
## 1  1.61 %    11.76 %   11.72 %    7.07 %      8.29 %      8.76 %

# Subset term
racism <- subset(byterm, byterm$term == 'Racism')

# fitting model based on reference code
mod <- glm(count~year, data = racism, family = "poisson")
```

```
# length.out measures length of x-axis which is 51 (1970-2020)
pred.df2 <- data.frame(year = seq(min(racism$year), max(racism$year),
length.out = 51))

# creating prediction model
pred <- predict(mod, newdata = pred.df, se.fit = TRUE)

# fitting exp count
pred.df2$count <- exp(pred$fit)

# generating lowest count value for ribbon (CI)
pred.df2$countmin <- exp(pred$fit - 2 * pred$se.fit)

# generating highest count value for ribbon (CI)
pred.df2$countmax <- exp(pred$fit + 2 * pred$se.fit)

# calculate the percentage growth rate
pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100


colors <- c("Predicted" = "red", "Expected" = "blue")
ggplot(racism,aes(x=year,y=count)) +
  scale_x_continuous(limits=c(1970,2022),breaks=c(1970,
                                                  1975,1980,1985,1990,1995,
                                                  2000,2005,2010,2015,2020)) +
  geom_point() +
  geom_ribbon(data = pred.df2, aes(ymin = countmin, ymax = countmax), alpha =
0.3) +
  geom_line(data = pred.df2, aes(y = count, color="Predicted")) +
  # Add vertical lines
  geom_vline(xintercept = c(1970,1980,1990,2000,2010,2020),
            linetype="dotted") +
  # Add growth rate per decade
  geom_segment(aes(x = 1970, y = 500, xend = 1980, yend = 500))+
  geom_text(aes(1975,
            y = 500,
            label = "1.61 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  geom_segment(aes(x = 1980, y = 1250, xend = 1990, yend = 1250))+
  geom_text(aes(1985,
            y = 1250,
```

```
            label = "11.76 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  geom_segment(aes(x = 1990, y = 2000, xend = 2000, yend = 2000))+
  geom_text(aes(1995,
            y = 2000,
            label = "11.72 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  geom_segment(aes(x = 2000, y = 2750, xend = 2010, yend = 2750))+
  geom_text(aes(2005,
            y = 2750,
            label = "7.07 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  geom_segment(aes(x = 2010, y = 3500, xend = 2020, yend = 3500))+
  geom_text(aes(2015,
            y = 3500,
            label = "8.29 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  annotate(geom = 'label', label = paste0("Overall Predicted Growth Rate: ",
round(pred_rate, digits = 2), "%"),
            x = -Inf, y = Inf, hjust = 0, vjust = 1) +
  ggtitle("Racism") +
  labs(x = "Year",
      y = "Count",
      color = "Legend") +
      scale_color_manual(values = colors)

# Liberalism
liberalism <- subset(byterm, byterm$term == "Liberalism")
d1 = data.frame(subset(liberalism, liberalism$year <=1980))
d2 = data.frame(subset(liberalism, liberalism$year >=1980 & liberalism$year
<=1990))
d3 = data.frame(subset(liberalism, liberalism$year >=1990 & liberalism$year
<=2000))
d4 = data.frame(subset(liberalism, liberalism$year >=2000 & liberalism$year
<=2010))
d5 = data.frame(subset(liberalism, liberalism$year >=2010 & liberalism$year
```

```
<=2020))
yr_list = list(d1,d2,d3,d4,d5,liberalism)
yr_tmp = list()
for (i in 1:6){
  dt = yr_list[[i]]
  # fitting model based on reference code
  mod <- glm(count~year, data = dt, family = "poisson")
  # calculate the percentage growth rate
  pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100
  yr_tmp[[i]] <- pred_rate
}
#Convert list to dataframe
yr_df <- data.frame(yr_tmp)
#Add correct column name per decade
names(yr_df) <-
c("1970-1980","1980-1990","1990-2000","2000-2010","2010-2020","1970-2020")
#Round numbers to 2 digits and add %
yr_df <- round(yr_df, 2)
yr_df[] <- Map(paste, yr_df, "%")
yr_df

##   1970-1980 1980-1990 1990-2000 2000-2010 2010-2020 1970-2020
## 1  3.95 %      8.73 %   10.29 %  6.28 %   -1.89 %  7.25 %

# Subset term
liberalism <- subset(byterm, byterm$term == 'Liberalism')

# fitting model based on reference code
mod <- glm(count~year, data = liberalism, family = "poisson")

# length.out measures length of x-axis which is 51 (1970-2020)
pred.df2 <- data.frame(year = seq(min(liberalism$year), max(liberalism$year),
length.out = 51))

# creating prediction model
pred <- predict(mod, newdata = pred.df, se.fit = TRUE)

# fitting exp count
pred.df2$count <- exp(pred$fit)

# generating lowest count value for ribbon (CI)
pred.df2$countmin <- exp(pred$fit - 2 * pred$se.fit)

# generating highest count value for ribbon (CI)
```

```
pred.df2$countmax <- exp(pred$fit + 2 * pred$se.fit)

# calculate the percentage growth rate
pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100


colors <- c("Predicted" = "red", "Expected" = "blue")
ggplot(liberalism,aes(x=year,y=count)) +
  scale_x_continuous(limits=c(1970,2022),breaks=c(1970,
                                               1975,1980,1985,1990,1995,
                                               2000,2005,2010,2015,2020)) +
  geom_point() +
  geom_ribbon(data = pred.df2, aes(ymin = countmin, ymax = countmax), alpha =
0.3) +
  geom_line(data = pred.df2, aes(y = count, color="Predicted")) +
  # Add vertical lines
  geom_vline(xintercept = c(1970,1980,1990,2000,2010,2020),
          linetype="dotted") +
  # Add growth rate per decade
  geom_segment(aes(x = 1970, y = 500, xend = 1980, yend = 500))+
  geom_text(aes(1975,
          y = 500,
          label = "3.95 %"),
          color = 'blue',
          size = 4,
          nudge_y = 150) +
  geom_segment(aes(x = 1980, y = 1250, xend = 1990, yend = 1250))+
  geom_text(aes(1985,
          y = 1250,
          label = "8.73 %"),
          color = 'blue',
          size = 4,
          nudge_y = 150) +
  geom_segment(aes(x = 1990, y = 2000, xend = 2000, yend = 2000))+
  geom_text(aes(1995,
          y = 2000,
          label = "10.29 %"),
          color = 'blue',
          size = 4,
          nudge_y = 150) +
  geom_segment(aes(x = 2000, y = 2750, xend = 2010, yend = 2750))+
  geom_text(aes(2005,
          y = 2750,
```

```
              label = "6.28 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  geom_segment(aes(x = 2010, y = 3500, xend = 2020, yend = 3500))+
  geom_text(aes(2015,
            y = 3500,
            label = "-1.89 %"),
            color = 'blue',
            size = 4,
            nudge_y = 150) +
  annotate(geom = 'label', label = paste0("Overall Predicted Growth Rate: ",
round(pred_rate, digits = 2), "%"),
            x = -Inf, y = Inf, hjust = 0, vjust = 1) +
  ggtitle("Liberalism") +
  labs(x = "Year",
      y = "Count",
      color = "Legend") +
      scale_color_manual(values = colors)

# Conservative
conservative <- subset(byterm, byterm$term == "Conservative")
d1 = data.frame(subset(conservative, conservative$year <=1980))
d2 = data.frame(subset(conservative, conservative$year >=1980 &
conservative$year <=1990))
d3 = data.frame(subset(conservative, conservative$year >=1990 &
conservative$year <=2000))
d4 = data.frame(subset(conservative, conservative$year >=2000 &
conservative$year <=2010))
d5 = data.frame(subset(conservative, conservative$year >=2010 &
conservative$year <=2020))
yr_list = list(d1,d2,d3,d4,d5, conservative)
yr_tmp = list()
for (i in 1:6){
  dt = yr_list[[i]]
  # fitting model based on reference code
  mod <- glm(count~year, data = dt, family = "poisson")
  # calculate the percentage growth rate
  pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100
  yr_tmp[[i]] <- pred_rate
}
#Convert list to dataframe
yr_df <- data.frame(yr_tmp)
#Add correct column name per decade
```

```
names(yr_df) <-
c("1970-1980","1980-1990","1990-2000","2000-2010","2010-2020","1970-2020")
#Round numbers to 2 digits and add %
yr_df <- round(yr_df, 2)
yr_df[] <- Map(paste, yr_df, "%")
yr_df

##   1970-1980 1980-1990 1990-2000 2000-2010 2010-2020 1970-2020
## 1   11.27 %    9.23 %      4.54 %      7.44 %   -0.56 %  4.99 %

# Subset term
conservative <- subset(byterm, byterm$term == 'Conservative')

# fitting model based on reference code
mod <- glm(count~year, data = conservative, family = "poisson")

# length.out measures length of x-axis which is 51 (1970-2020)
pred.df2 <- data.frame(year = seq(min(conservative$year),
max(conservative$year), length.out = 51))

# creating prediction model
pred <- predict(mod, newdata = pred.df, se.fit = TRUE)

# fitting exp count
pred.df2$count <- exp(pred$fit)

# generating lowest count value for ribbon (CI)
pred.df2$countmin <- exp(pred$fit - 2 * pred$se.fit)

# generating highest count value for ribbon (CI)
pred.df2$countmax <- exp(pred$fit + 2 * pred$se.fit)

# calculate the percentage growth rate
pred_rate <- (exp(summary(mod)$coefficients[2,1]) - 1) * 100


colors <- c("Predicted" = "red", "Expected" = "blue")
ggplot(conservative,aes(x=year,y=count)) +
  scale_x_continuous(limits=c(1970,2022),breaks=c(1970,
                                                   1975,1980,1985,1990,1995,
                                                   2000,2005,2010,2015,2020)) +
  geom_point() +
  geom_ribbon(data = pred.df2, aes(ymin = countmin, ymax = countmax), alpha =
0.3) +
```

```
  geom_line(data = pred.df2, aes(y = count, color="Predicted")) +
  # Add vertical lines
  geom_vline(xintercept = c(1970,1980,1990,2000,2010,2020),
             linetype="dotted") +
  # Add growth rate per decade
  geom_segment(aes(x = 1970, y = 2500, xend = 1980, yend = 2500))+
  geom_text(aes(1975,
             y = 2500,
             label = "11.27 %"),
             color = 'blue',
             size = 4,
             nudge_y = 300) +
  geom_segment(aes(x = 1980, y = 3550, xend = 1990, yend = 3550))+
  geom_text(aes(1985,
             y = 3550,
             label = "9.23 %"),
             color = 'blue',
             size = 4,
             nudge_y = 300) +
  geom_segment(aes(x = 1990, y = 5000, xend = 2000, yend = 5000))+
  geom_text(aes(1995,
             y = 5000,
             label = "4.54 %"),
             color = 'blue',
             size = 4,
             nudge_y = 300) +
  geom_segment(aes(x = 2000, y = 6250, xend = 2010, yend = 6250))+
  geom_text(aes(2005,
             y = 6250,
             label = "7.44 %"),
             color = 'blue',
             size = 4,
        nudge_y = 300) +
  geom_segment(aes(x = 2010, y = 7500, xend = 2020, yend = 7500))+
  geom_text(aes(2015,
             y = 7500,
             label = "-0.56 %"),
             color = 'blue',
             size = 4,
             nudge_y = 300) +
  annotate(geom = 'label', label = paste0("Overall Predicted Growth Rate: ",
round(pred_rate, digits = 2), "%"),
             x = -Inf, y = Inf, hjust = 0, vjust = 1) +
```

```
ggtitle("Conservative") +
labs(x = "Year",
    y = "Count",
    color = "Legend") +
    scale_color_manual(values = colors)
```