Quantitative Text Analysis

CEU Political Science

Petro Tolochko

Contents of the Course

Contents of the Course

- Text Representation
- Pre-processing
- Regular Expressions and Classification with dictionaries
- Machine Learning and Classification
- Topic modeling/unsupervised clustering
- Multilingual text analysis
- Word Vectors
- Large Language Models and Transformers • Scraping and Using APIs
- Validation
- Ethics and Data Security
- Critical reflection on the methods

Contents of the Course

- Text Representation
- Pre-processing
- Regular Expressions and Classification with dictionaries
- Machine Learning and Classification
- Topic modeling/unsupervised clustering
- Multilingual text analysis
- Word Vectors
- Large Language Models and Transformers • Scraping and Using APIs
- Validation
- Ethics and Data Security
- Critical reflection on the methods

Subject to Change

Course Structure

- Theoretical Input
- Practical Sessions
- Attendance / Participation
- Final Paper

Introduction

Introduction

- Post-Doc, Computational Communication Science Lab, University of Vienna
- Research focus: Text complexity, social networks
- petro.tolochko@univie.ac.at

Introduction

- Name?
- Background?
- Experience with Text Analysis?
- Programming?
- What are the expectations and wishes for the course?

Objectives

- Getting to know methods of automated text analysis
- Practical Challenges
- Critical Reflection
- Utility for your own projects

Course assessment

- Participation + Engagement + Take-home Assignment

Course assessment

- Participation + Engagement + Take-home Assignment

Personal Feedback

- Informal feedback on your projects
- Research question, Data, Methods, Current struggles

Course repository

- https://github.com/PeterTolochko/CEU_text_analysis

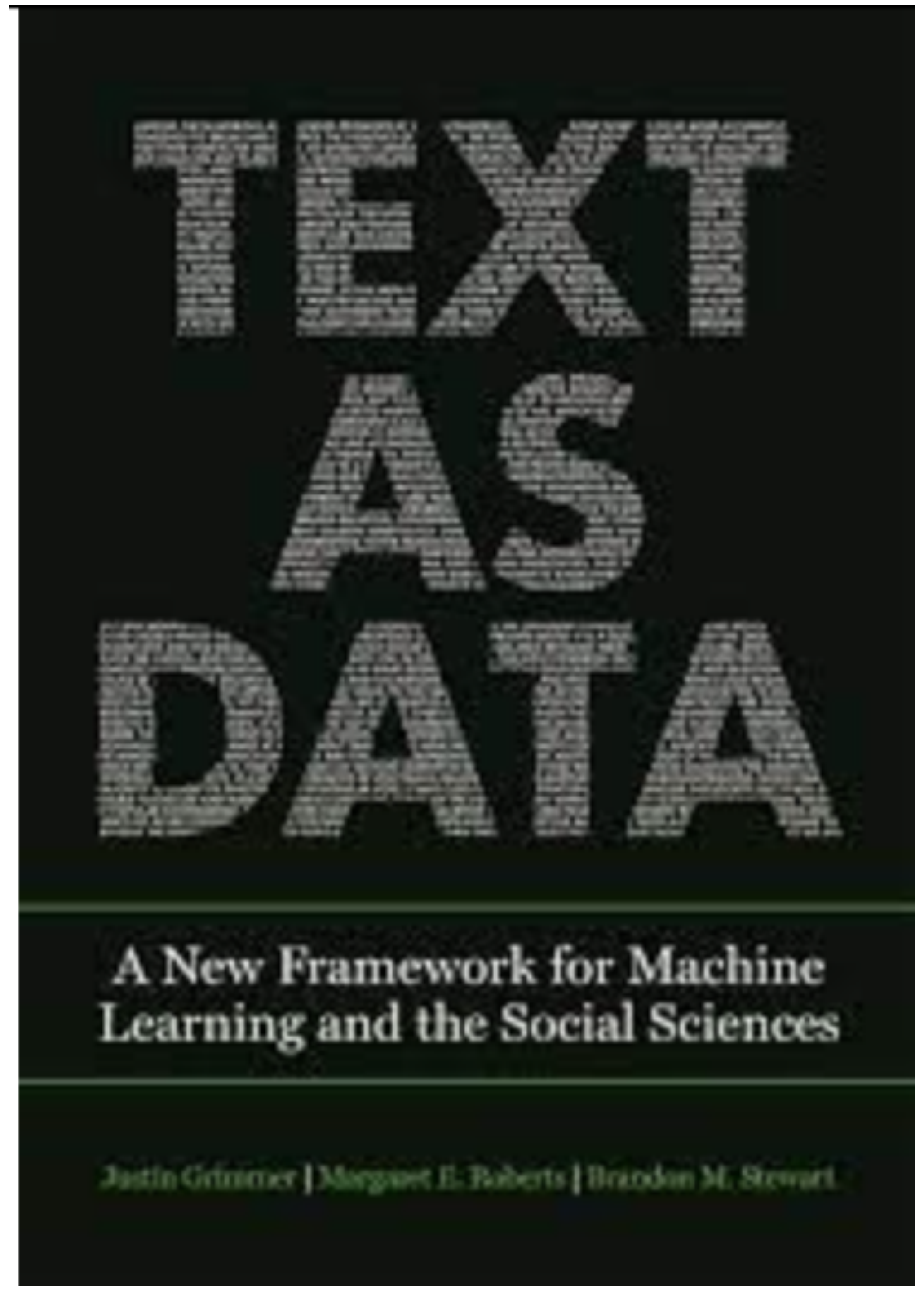
Today

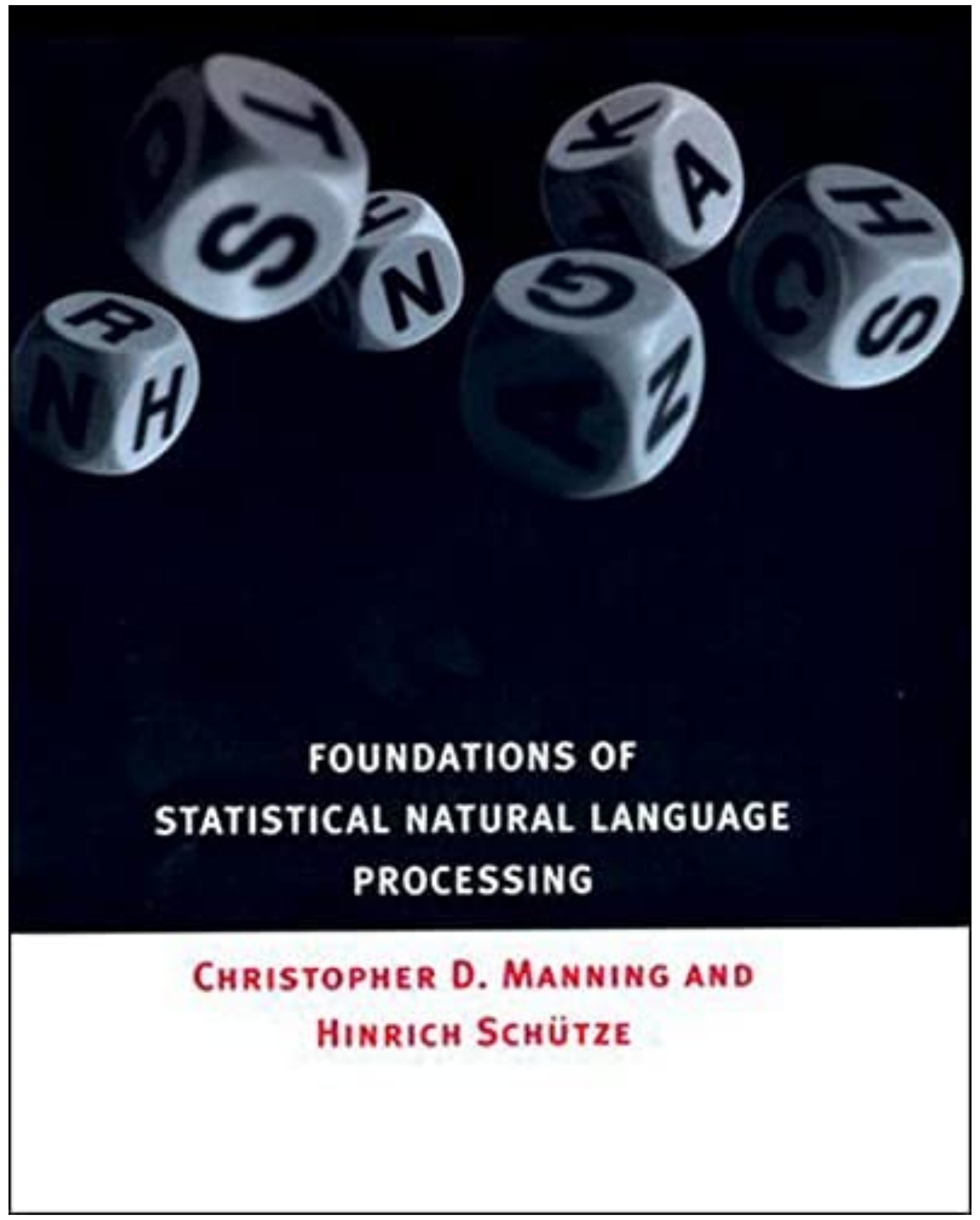
- Text representation
- R / GitHub refresher

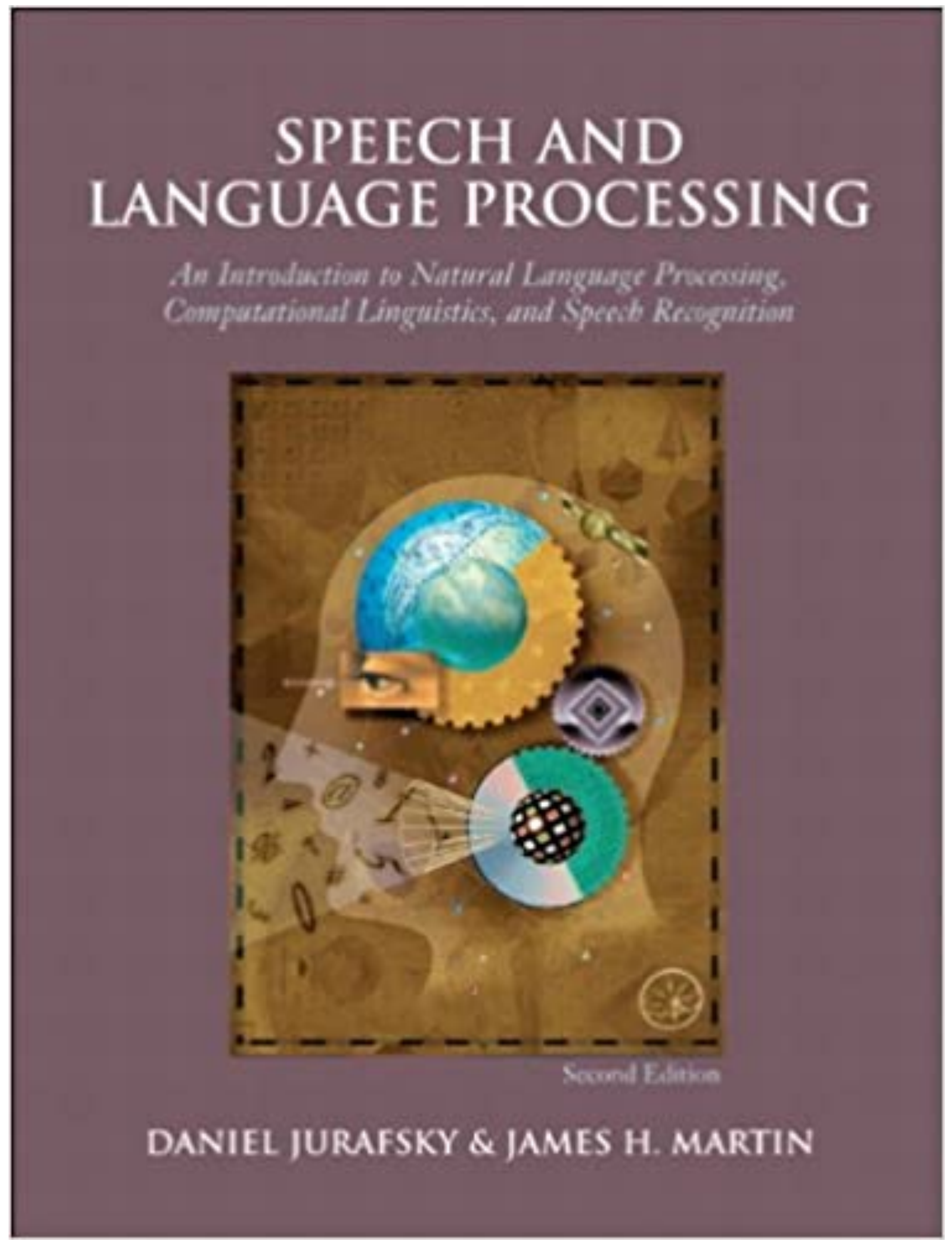
Questions?

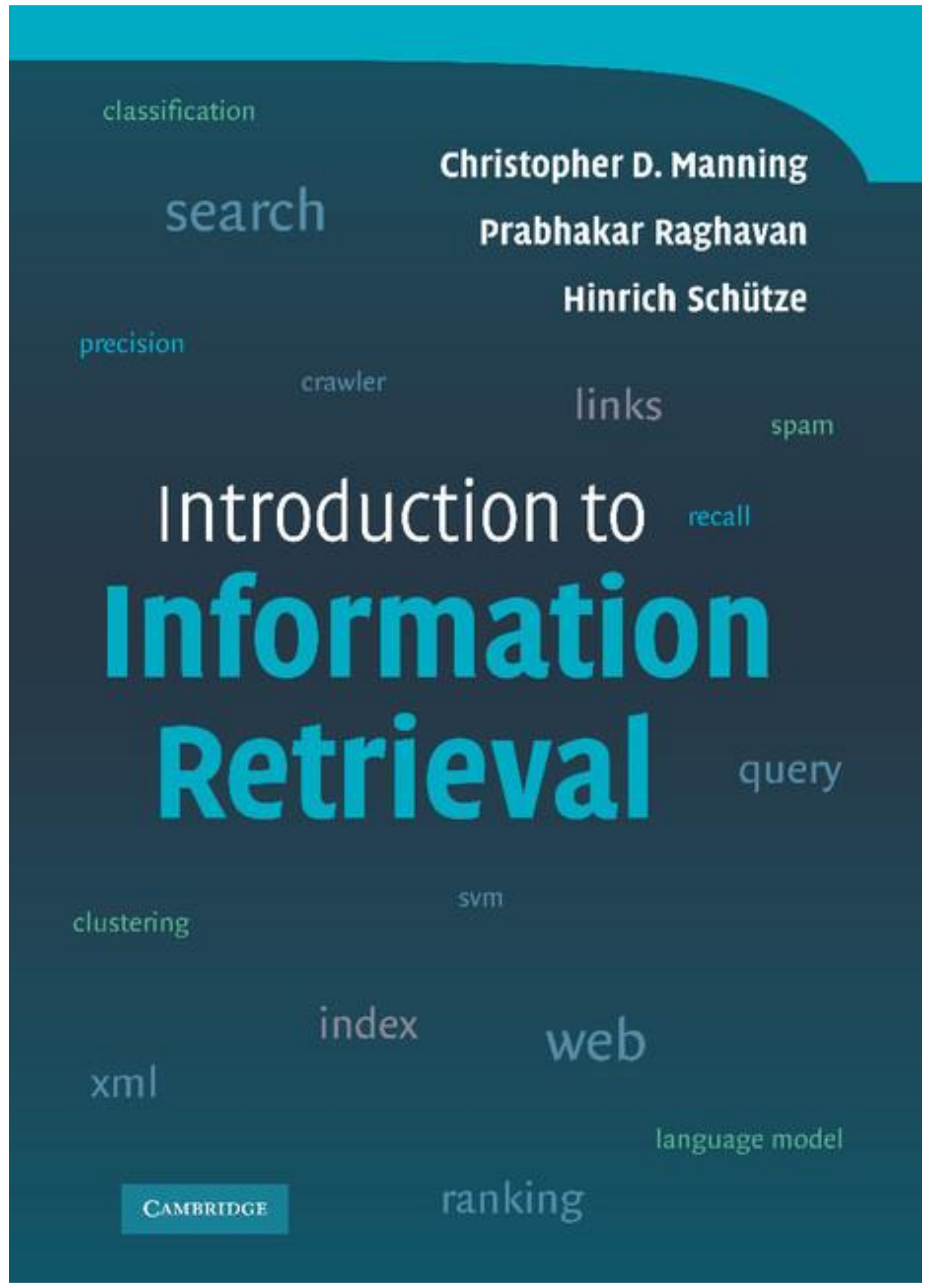
Resources

Resources

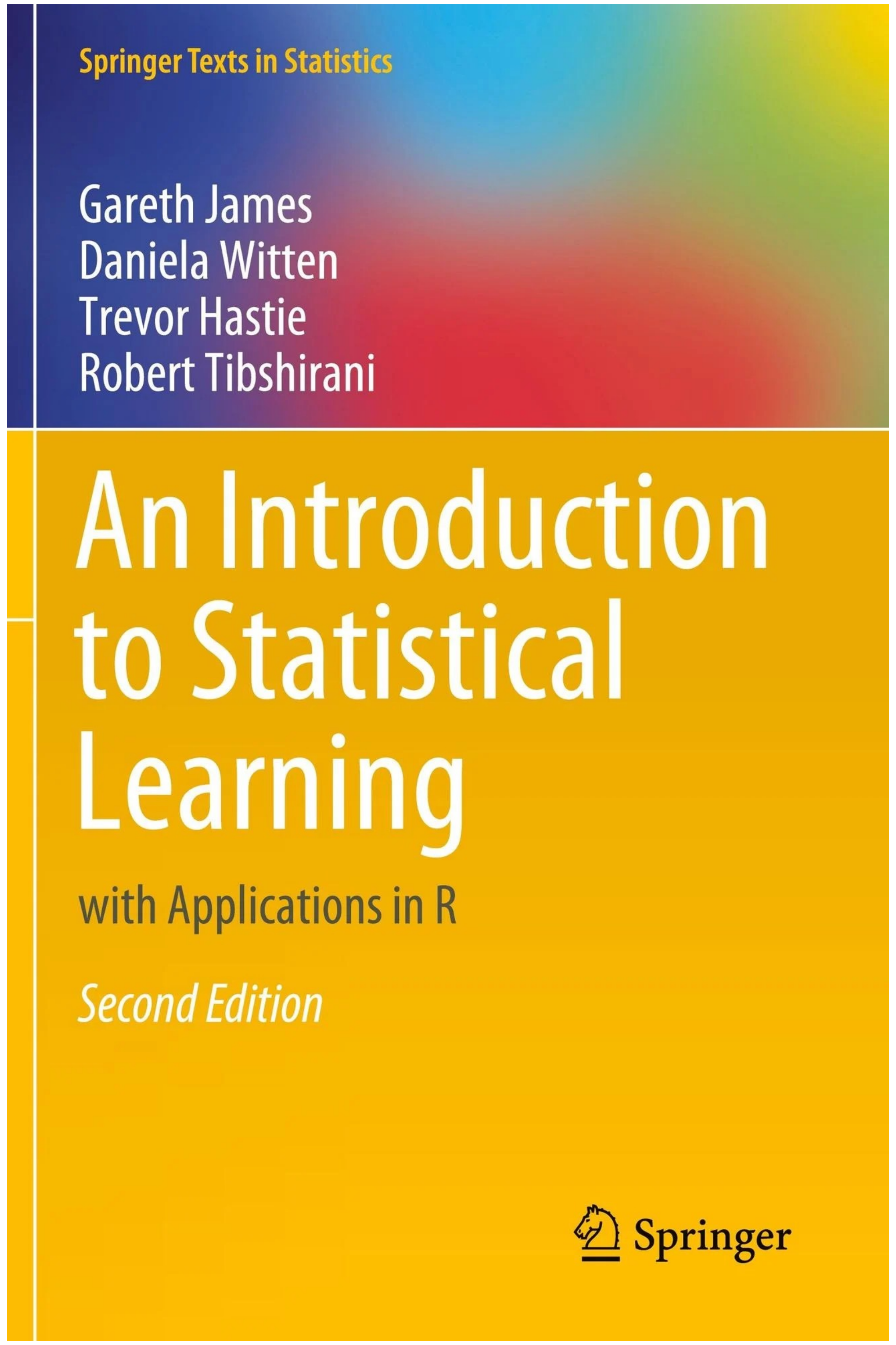

TEXT AS DATA
A New Framework for Machine Learning and the Social Sciences
Justin Grimmer | Margaret E. Roberts | Brandon M. Stewart

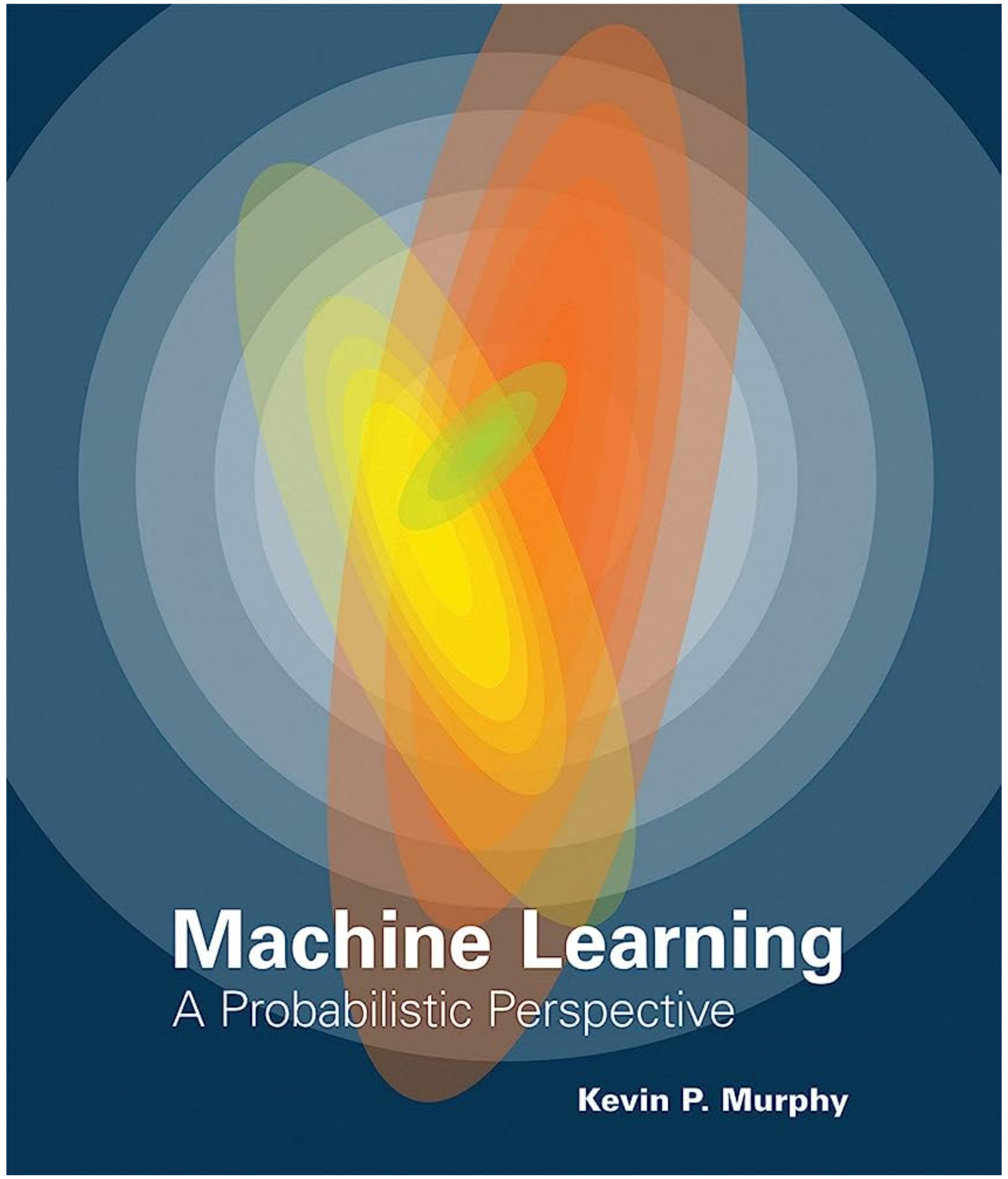

FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING
CHRISTOPHER D. MANNING AND HINRICH SCHÜTZE


SPEECH AND LANGUAGE PROCESSING
An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition
Second Edition
DANIEL JURAFSKY & JAMES H. MARTIN


Christopher D. Manning
Prabhakar Raghavan
Hinrich Schütze
Introduction to Information Retrieval
classification
search
precision
crawler
links
spam
recall
query
clustering
svm
index
web
xml
language model
ranking
CAMBRIDGE

Resources


Springer Texts in Statistics
Gareth James
Daniela Witten
Trevor Hastie
Robert Tibshirani
An Introduction to Statistical Learning
with Applications in R
Second Edition
Springer


Machine Learning
A Probabilistic Perspective
Kevin P. Murphy

Resources

- *Advanced Data Analysis from an Elementary Point of View*
    - Cosma Rohilla Shalizi
    - ***Free online***

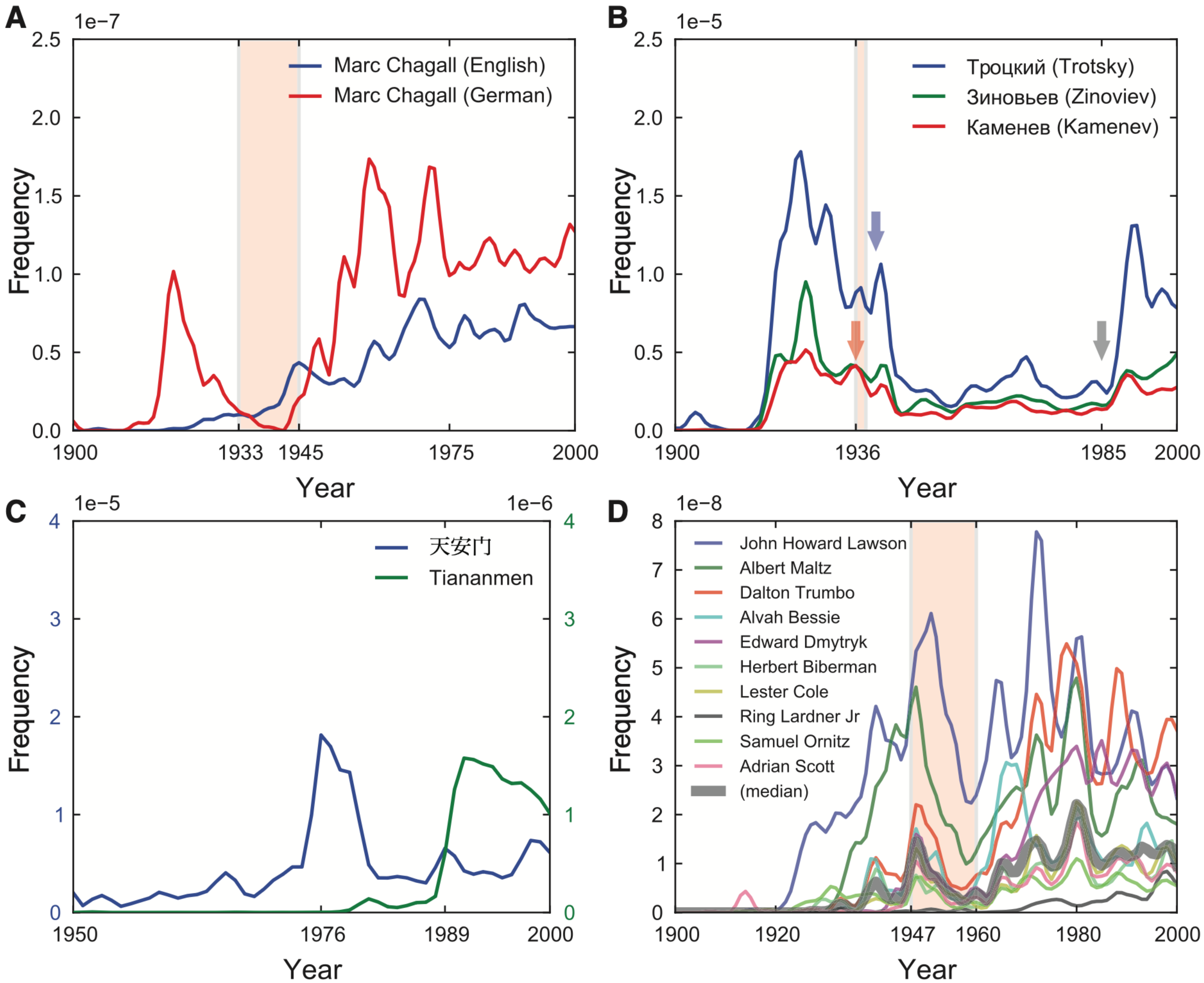
Why analyse texts?

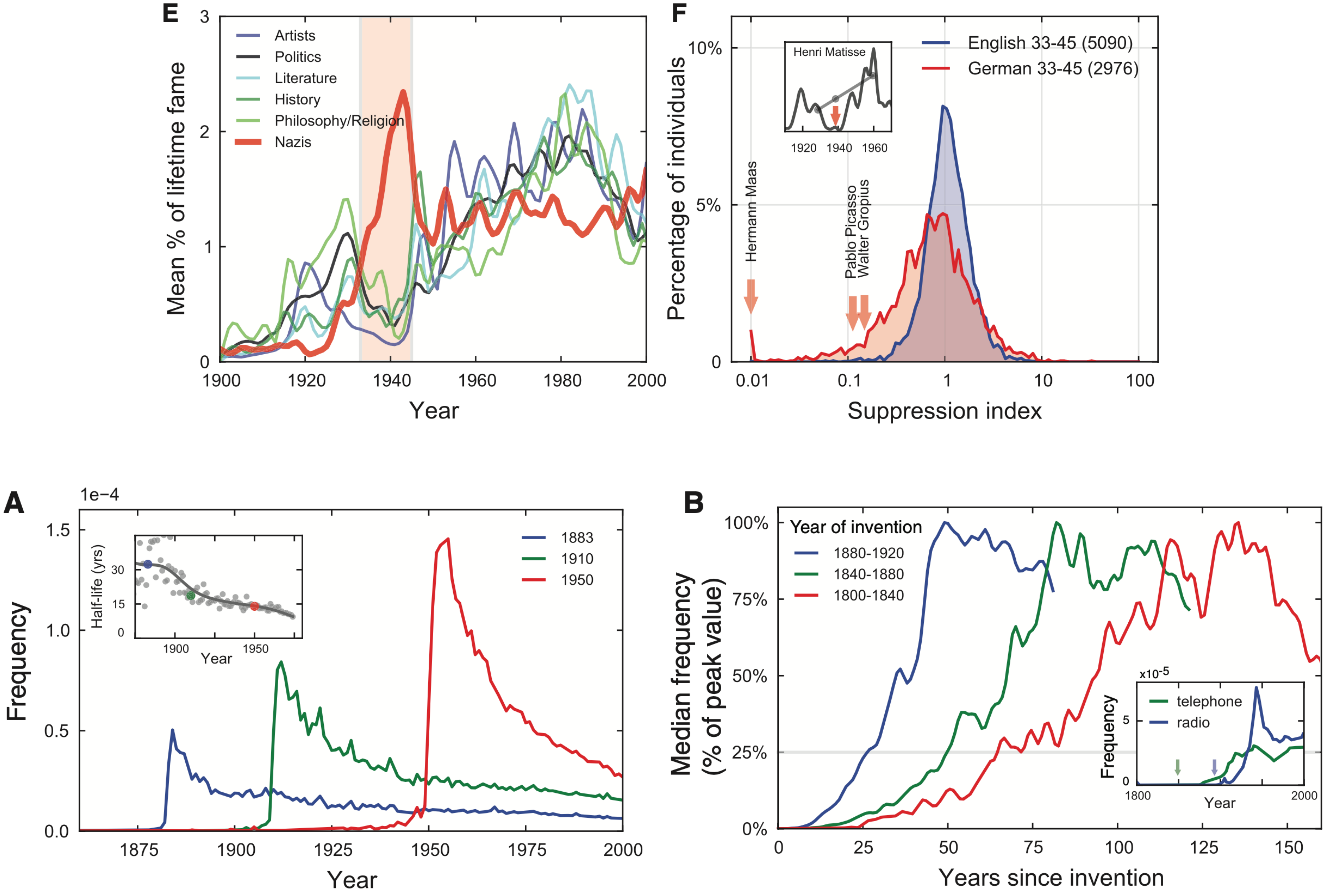
- Vast quantities of data available
- Texts carry meaning
- Texts capture social behaviour

Quantitative Analysis of Culture Using Millions of Digitized Books

Article *in* Science · January 2011

DOI: 10.1126/science.1199644 · Source: PubMed


A
1e−7
Marc Chagall (English)
Marc Chagall (German)
Frequency
Year
1900
1933
1945
1975
2000
B
1e−5
Троцкий (Trotsky)
Зиновьев (Zinoviev)
Каменев (Kamenev)
Frequency
Year
1900
1936
1985
2000
C
1e−5
1e−6
天安门
Tiananmen
Frequency
Year
1950
1976
1989
2000
D
1e−8
John Howard Lawson
Albert Maltz
Dalton Trumbo
Alvah Bessie
Edward Dmytryk
Herbert Biberman
Lester Cole
Ring Lardner Jr
Samuel Ornitz
Adrian Scott
(median)
Frequency
Year
1900
1920
1947
1960
1980
2000



E
Mean % of lifetime fame
Artists
Politics
Literature
History
Philosophy/Religion
Nazis
0
1
2
3
1900
1920
1940
1960
1980
2000
Year
F
Percentage of individuals
Henri Matisse
1920
1940
1960
English 33-45 (5090)
German 33-45 (2976)
Hermann Maas
Pablo Picasso
Walter Gropius
0
5%
10%
0.01
0.1
1
10
100
Suppression index
A
1e−4
Frequency
Half-life (yrs)
0
15
30
1900
1950
Year
1883
1910
1950
0.0
0.5
1.0
1.5
1875
1900
1925
1950
1975
2000
Year
B
Median frequency
(% of peak value)
Year of invention
1880-1920
1840-1880
1800-1840
x10-5
Frequency
telephone
radio
0
5
1800
2000
Year
0%
25%
50%
75%
100%
0
25
50
75
100
125
150
Years since invention

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Predicting and Interpolating State-Level Polls Using Twitter Textual Data

Nicholas Beauchamp Northeastern University

Abstract: *Spatially or temporally dense polling remains both difficult and expensive using existing survey methods. In response, there have been increasing efforts to approximate various survey measures using social media, but most of these approaches remain methodologically flawed. To remedy these flaws, this article combines 1,200 state-level polls during the 2012 presidential campaign with over 100 million state-located political tweets; models the polls as a function of the Twitter text using a new linear regularization feature-selection method; and shows via out-of-sample testing that when properly modeled, the Twitter-based measures track and to some degree predict opinion polls, and can be extended to unpolled states and potentially substate regions and subday timescales. An examination of the most predictive textual features reveals the topics and events associated with opinion shifts, sheds light on more general theories of partisan difference in attention and information processing, and may be of use for real-time campaign strategy.*

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Policy Diffusion: The Issue-Definition Stage

Fabrizio Gilardi University of Zurich
Charles R. Shipan University of Michigan
Bruno Wüest Forschungsstelle sotomo

Abstract: *We put forward a new approach to studying issue definition within the context of policy diffusion. Most studies of policy diffusion—which is the process by which policymaking in one government affects policymaking in other governments—have focused on policy adoptions. We shift the focus to an important but neglected aspect of this process: the issue-definition stage. We use topic models to estimate how policies are framed during this stage and how these frames are predicted by prior policy adoptions. Focusing on smoking restriction in U.S. states, our analysis draws upon an original data set of over 52,000 paragraphs from newspapers covering 49 states between 1996 and 2013. We find that frames regarding the policy's concrete implications are predicted by prior adoptions in other states, whereas frames regarding its normative justifications are not. Our approach and findings open the way for a new perspective to studying policy diffusion in many different areas.*

Verification Materials: The data and materials required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at https://doi.org/10.7910/DVN/QEMNP1.

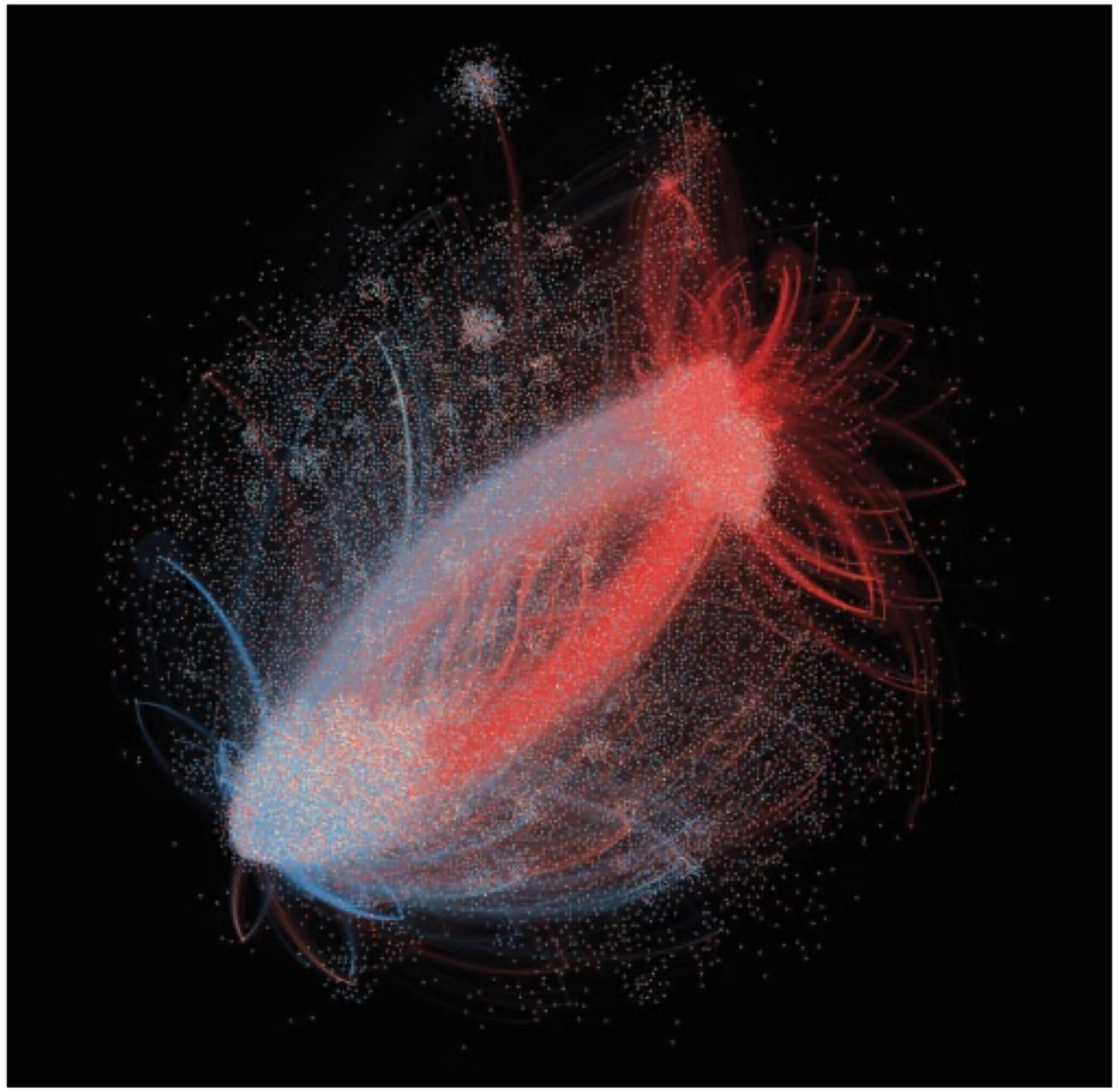
General Article

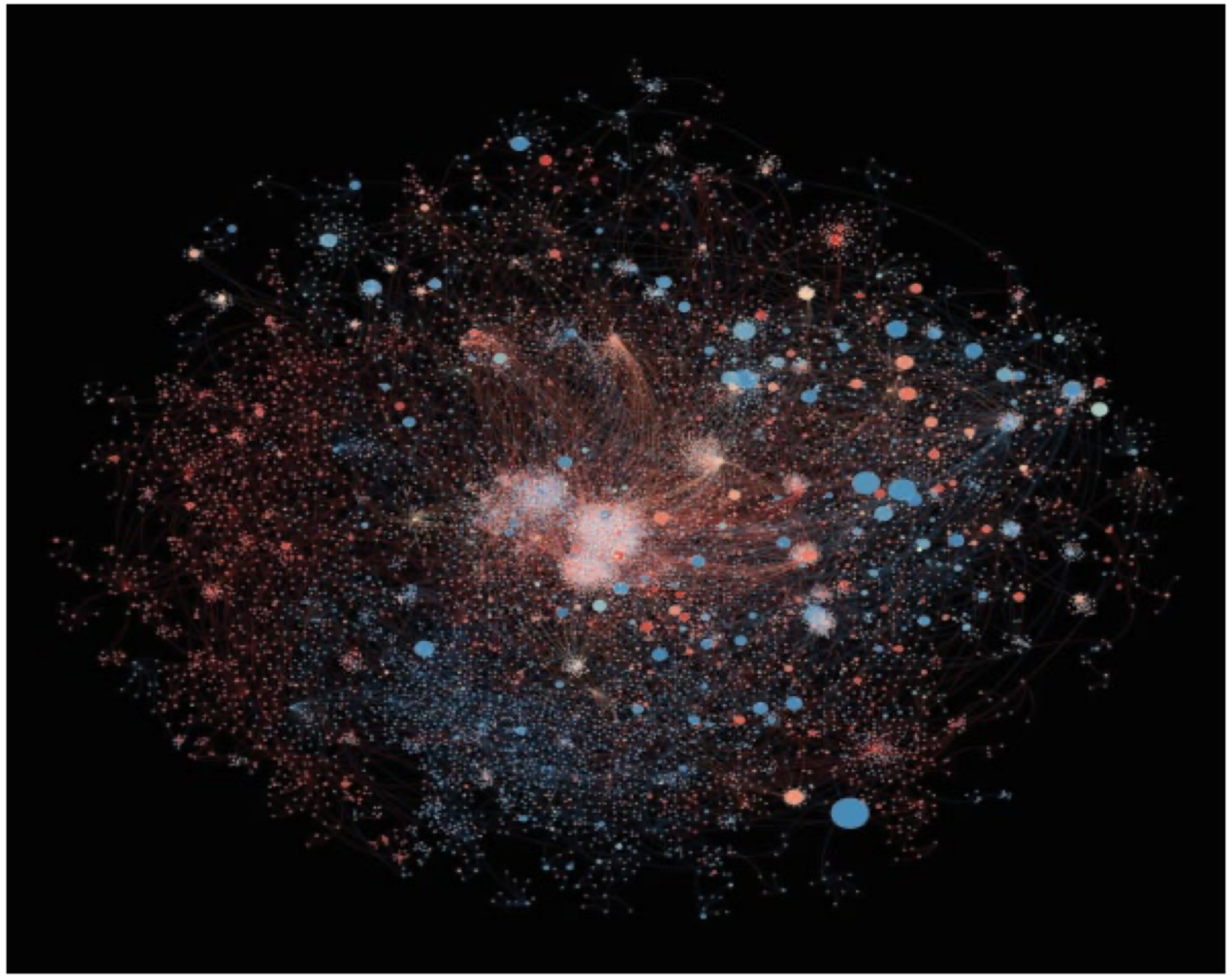
Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?

Psychological Science
2015, Vol. 26(10) 1531–1542

DOI: 10.1177/0956797615594620
pss.sagepub.com

Pablo Barberá[1], John T. Jost[1,2,3], Jonathan Nagler[3], Joshua A. Tucker[3], and Richard Bonneau[4]
[1]Center for Data Science, [2]Department of Psychology, [3]Department of Politics, and [4]Center for Genomics and Systems Biology, New York University





Text Representation

What is text?

- Data
- Unstructured
- (Highly) Multidimensional

    - In general, difficult to work with (if you’re not human)

From Text to Structure

- We need to “structure” the text before we perform analyses
- Different ways to ***represent*** text so that computers “understand”
- Different ways to model text so that both (we and computer) “understand”
    - Different research questions
    - Different ways to think about what text is

Document-Term Matrix

$$X = \begin{bmatrix} 1 & 1 & \ldots & 0 \\ 2 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \\ 2 & 0 & \ldots & 3 \end{bmatrix}$$

X = N * K matrix
N = number of documents
K = number of terms/features

Example

- Corpus (Collection of texts)
- Document 1: “John loves ice-cream”
- Document 2: “John loves oranges”
- Document 3: “Marry hates ice-cream”

- N? K?

Docs	icecream	john	loves	oranges	hates	marry
1	1	1	1	0	0	0
2	0	1	1	1	0	0
3	1	0	0	0	1	1

(Terms)

N = 3
K = 6

X = 3 * 6 matrix

Types & Tokens

- Types: unique words in a text
- Tokens: all words in the text

- Types and tokens in our example corpus?

Types & Tokens

- Types: unique words in a text
- Tokens: all words in the text

- Types and tokens in our example corpus?
- 6 types (unique words) / 9 tokens (total length of the corpus)

Bag-of-Words representation

- Representation of text as a bag of words
- Collection of words
- Order is irrelevant
- Each text is represented as a ***count*** of words contained in ti

Docs	icecream	john	loves	oranges	hates	marry
1	1	1	1	0	0	0
2	0	1	1	1	0	0
3	1	0	0	0	1	1

Terms is printed above the term columns.

Multinomial Model of Language

- You can think of a text as a draw from a Multinomial distribution

Binomial Distribution

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Binomial Distribution

$$\binom{n}{k} p^k (1-p)^{n-k}$$

n = number of events
k = number of successes

Multinomial Distribution

$$\frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

n = text length
k = size of vocabulary
p = probability of a word

Example Texts

- Text_1 = "banana banana banana banana chocolate"
- Text_2 = "chocolate chocolate chocolate banana fudge"
- Text_3 = "banana banana"
- Text_4 = “ice-cream ice-cream fudge ice-cream"
- Text_5 = "fudge fudge fudge”
- Text_6 = "ice-cream ice-cream fudge fudge"

Example Texts

- Text_1 = "banana banana banana banana chocolate"
- Text_2 = "chocolate chocolate chocolate banana fudge"
- Text_3 = "banana banana"
- Text_4 = "icecream icecream fudge ice-cream"
- Text_5 = "fudge fudge fudge"
- Text_6 = "ice-cream ice-cream fudge fudge"

John

Example Texts

- Text_1 = "banana banana banana banana chocolate"
- Text_2 = "chocolate chocolate chocolate banana fudge"
- Text_3 = "banana banana"

John

- Text_4 = "icecream icecream fudge ice-cream"
- Text_5 = "fudge fudge fudge"
- Text_6 = "ice-cream ice-cream fudge fudge"

Mary

Document Term Matrix

banana	chocolate	fudge	icecream
4	1	0	0
2	0	0	0
1	3	1	0
0	0	1	3
0	0	3	0
0	0	2	2

Document Term Matrix

banana	chocolate	fudge	icecream
7	4	1	0

John Rates

banana	chocolate	fudge	icecream
1	3	7	5

Mary Rates

Document Term Matrix

John Language Model

banana	chocolate	fudge	icecream
7	4	1	0

John Rates

banana	chocolate	fudge	icecream
1	3	7	5

Mary Rates

Mary Language Model

New Texts

- new_text_1 = "ice-cream fudge fudge"
- new_text_2 = "chocolate chocolate banana banana”

New Texts

- new_text_1 = "ice-cream fudge fudge"
- new_text_2 = "chocolate chocolate banana banana”
- What’s the probability that they have been generated by John or Marry?

New Texts

banana	chocolate	fudge	icecream
0	0	2	1
2	2	0	0

Probability Spoken by John

- new_text_1 = "ice-cream fudge fudge"

$$Pr(b = 0, ch = 0, f = 2, i = 1) =$$
$$\frac{3!}{0!0!2!1!} 0.583^0 \times 0.333^0 \times 0.08^2 \times 0^1 = 0$$

Probability Spoken by John

- new_text_2 = "chocolate chocolate banana banana"

$$Pr(b = 2, ch = 2, f = 0, i = 0) =$$
$$\frac{4!}{2!2!0!0!} 0.583^2 \times 0.333^2 \times 0.08^0 \times 0^0 = 0.23$$

Probability Spoken by Marry

- new_text_1 = "ice-cream fudge fudge"
- P = 0.18

- new_text_2 = "chocolate chocolate banana banana"
- P = 0.0008

Vector Space Representation

- Representation of texts as vectors in a multidimensional space

Multidimensional?

- Position on a map:
    - X = Longitude Y = Latitude
- Position in a real world:
    - X = Longitude Y = Latitude Z = Height
- Point in time and space
    - X = Longitude Y = Latitude Z = Height T = Time

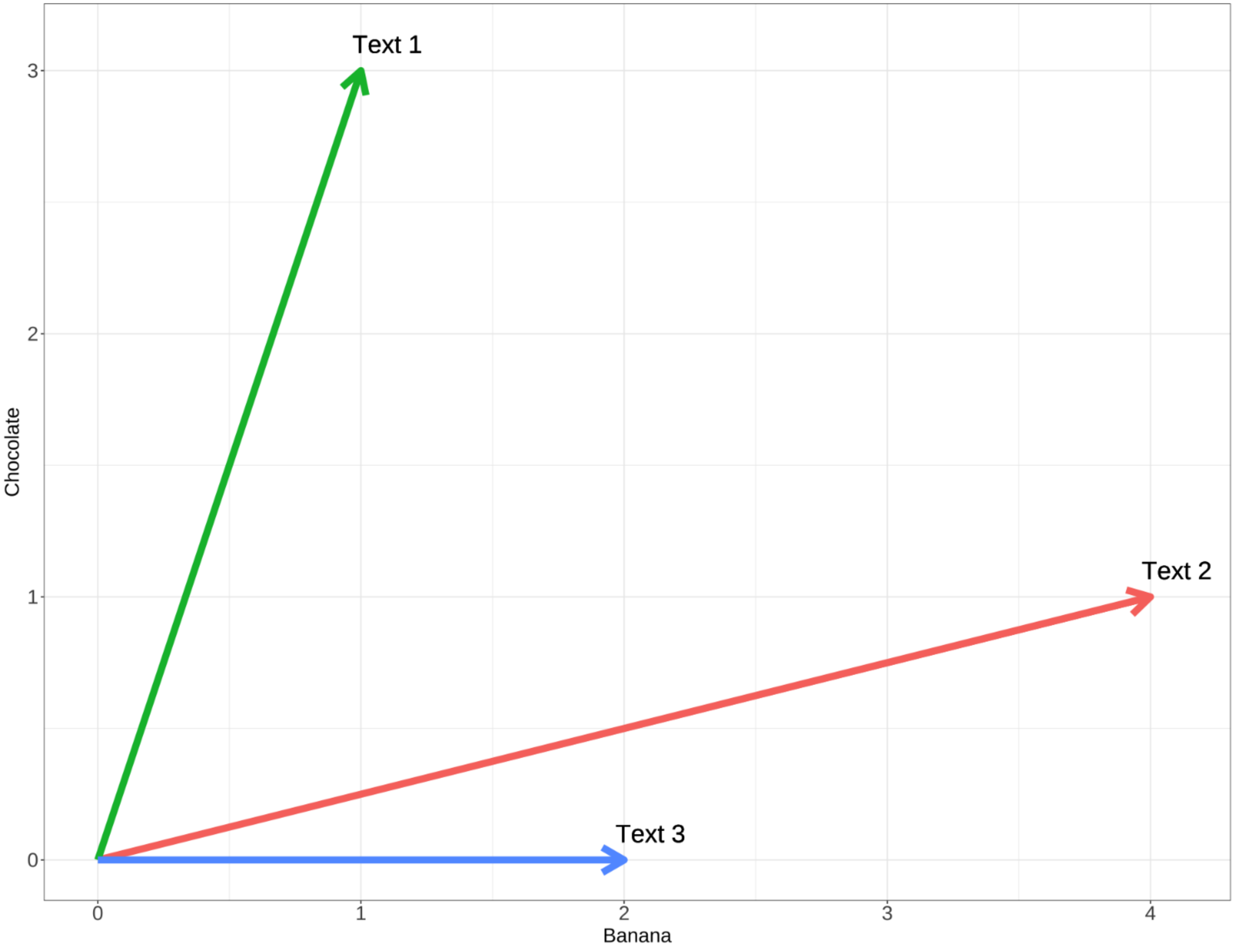
Multidimensional?

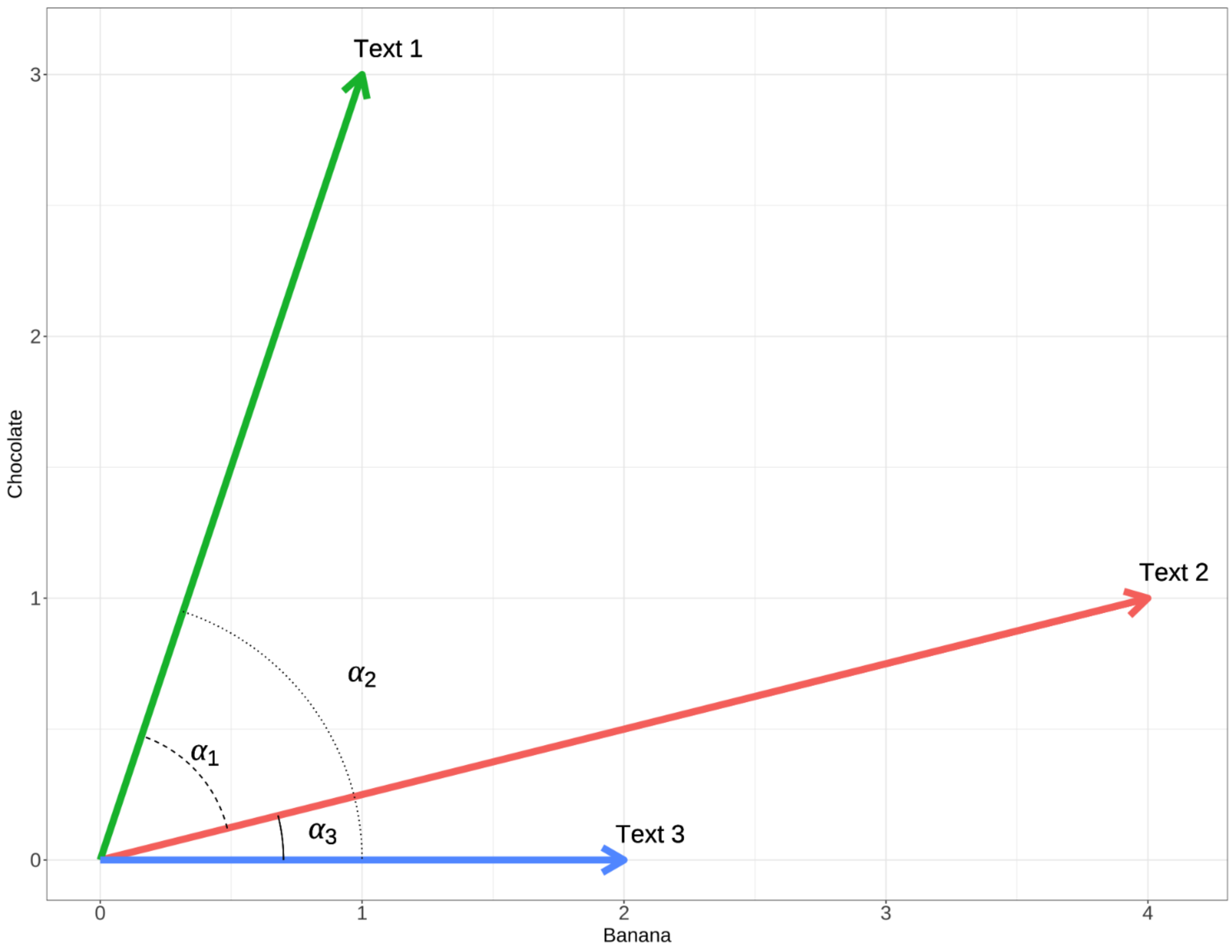
- Number of dimensions is the number of data points needed to describe an object in space
    - ***Coordinates*** in the context of ***geographical position***
    - ***Words*** in the context of position of text within a ***linguistic space***

Two-dimensional space

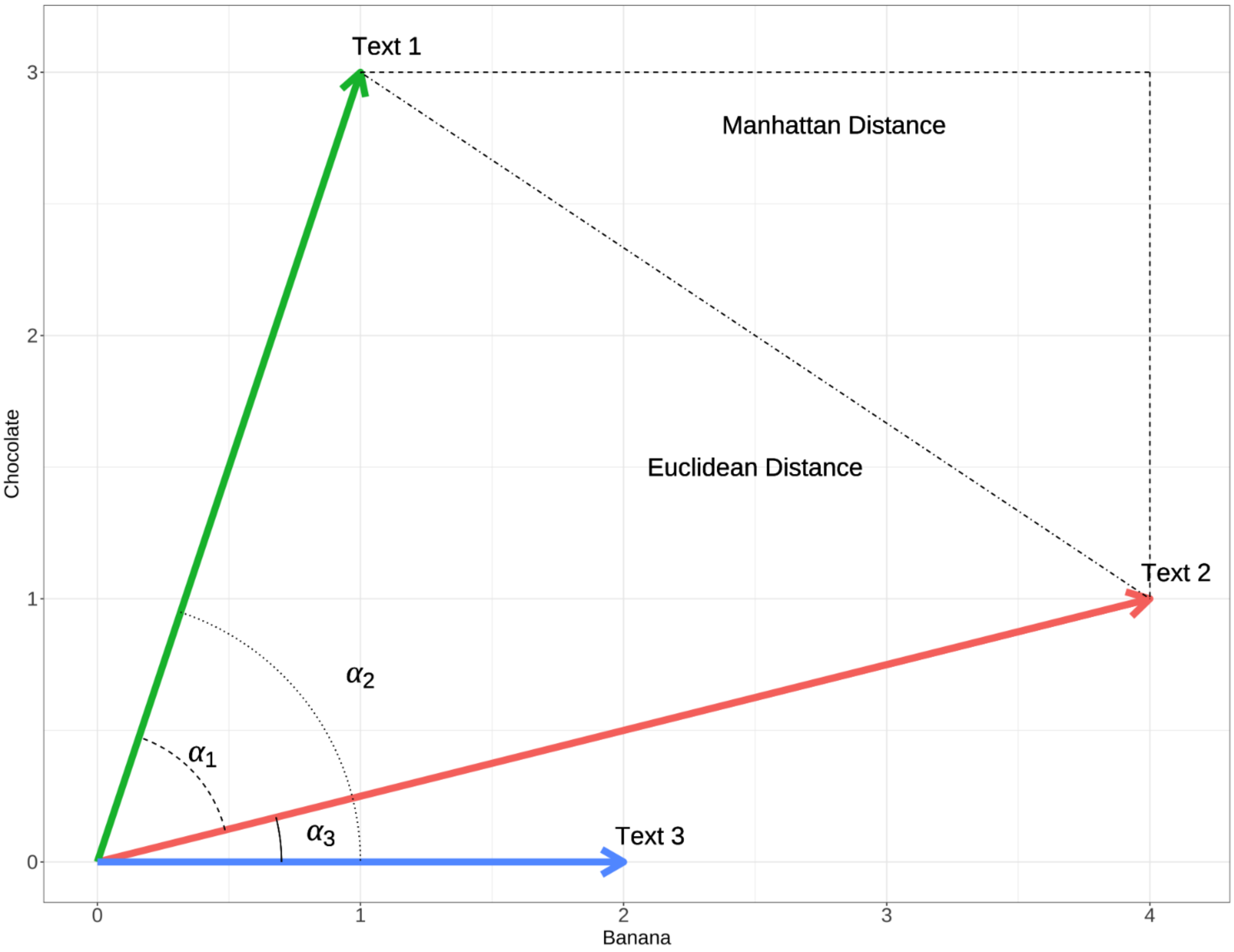
- vocabulary = {“banana”, “chocolate”}

- Text 1 = “chocolate, chocolate, chocolate, banana”
- Text 2 = “banana, banana, banana, banana, chocolate”
- Text 3 = “banana, banana”


Text 1
Text 2
Text 3
Chocolate
Banana
0
1
2
3
4



Text 1
Text 2
Text 3
α_1
α_2
α_3
Chocolate
Banana
0
1
2
3
4



Text 1
Manhattan Distance
Euclidean Distance
Text 2
α_2
α_1
α_3
Text 3
Chocolate
Banana
0
1
2
3
4

Multidimensional?

- Estimated ~1,022,000 words in English language
- ~ 1,022,000-dimensional space


[adult swim]

- Same math, though

Distributional Hypothesis / Word Embeddings

- The new thing kids do…

Distributional Hypothesis / Word Embeddings

- The new thing kids do… at least since ‘57…

Distributional Hypothesis / Word Embeddings

- In most cases, the meaning of a word is its use (Wittgenstein, 1953)
- Distributional hypothesis: difference of meaning correlates with difference of distribution (Harris, 1954)
- A word is characterized by the company it keeps (Firth, 1957)
- Words which are similar in meaning occur in similar contexts (Rubenstein & Goodenough, 1965)
- Words with similar meanings will occur with similar neighbors if enough text material is available (Schütze & Pedersen, 1995)

Distributional Hypothesis / Word Embeddings

- Each word is embedded in an n-dimensional vector space (typically, 300 dimensions)

- Neural network that "learns" the context of a word by looking at the words around it

- Word2Vec (Mikolov, 2013) – one of the most famous methodologies

Distributional Hypothesis / Word Embeddings

- Each word is embedded in an n-dimensional vector space (typically, 300 dimensions)
- Neural network that “learns” the context of a word by looking at the words around it
- Word2Vec (Mikolov, 2013) – one of the most famous methodologies

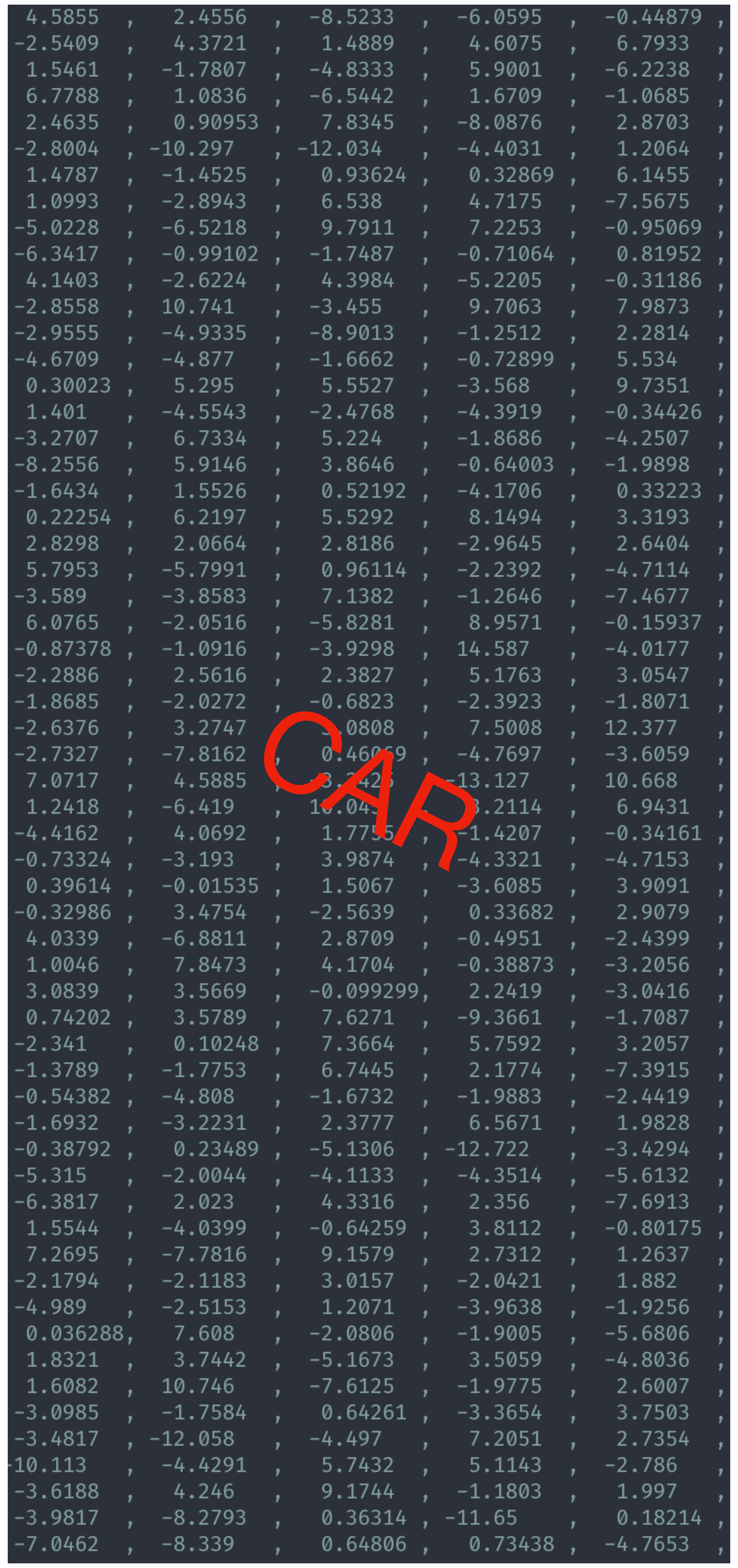
```
[ 0.20778 , -2.4151  ,  0.36605 ,  2.0139  , -0.23752 , -3.1952  ,
 -0.2952  ,  1.2272  , -3.4129  , -0.54969 ,  0.32634 , -1.0813  ,
  0.55626 ,  1.5195  ,  0.97797 , -3.1816  , -0.37207 , -0.86093 ,
  2.1509  , -4.0845  ,  0.035405,  3.5702  , -0.79413 , -1.7025  ,
 -1.6371  , -3.198   , -1.9387  ,  0.91166 ,  0.85409 ,  1.8039  ,
 -1.103   , -2.5274  ,  1.6365  , -0.82082 ,  1.0278  , -1.705   ,
  1.5511  , -0.95633 , -1.4702  , -1.865   , -0.19324 , -0.49123 ,
  2.2361  ,  2.2119  ,  3.6654  ,  1.7943  , -0.20601 ,  1.5483  ,
 -1.3964  , -0.50819 ,  2.1288  , -2.332   ,  1.3539  , -2.1917  ,
  1.8923  ,  0.28472 ,  0.54285 ,  1.2309  ,  0.26027 ,  1.9542  ,
  1.1739  , -0.40348 ,  3.2028  ,  0.75381 , -2.7179  , -1.3587  ,
 -1.1965  , -2.0923  ,  2.2855  , -0.3058  , -0.63174 ,  0.70083 ,
  0.16899 ,  1.2325  ,  0.97006 , -0.23356 , -2.094   , -1.737   ,
  3.6075  , -1.511   , -0.9135  ,  0.53878 ,  0.49268 ,  0.44751 ,
  0.6315  ,  1.4963  ,  4.1725  ,  2.1961  , -1.2409  ,  0.4214  ,
  2.9678  ,  1.841   ,  3.0133  , -4.4652  ,  0.96521 , -0.29787 ,
  4.3386  , -1.2527  , -1.7734  , -3.5637  , -0.20035 , -3.3013  ,
  0.99951 , -0.92888 , -0.94594 ,  1.5124  , -3.9385  ,  2.7935  ,
 -3.1042  ,  3.3382  ,  0.54513 , -0.37663 ,  2.5151  ,  0.51468 ,
 -0.88907 ,  1.011   ,  3.4705  , -3.6037  ,  1.3702  ,  2.3468  ,
  1.6674  ,  1.3904  , -2.8112  ,  2.237   , -1.0344  , -0.57164 ,
  1.0641  , -1.6919  ,  1.958   , -0.78305 ,  0.14741 ,  0.51083 ,
  1.8278  , -0.69638 ,  0.90548 ,  0.62282 , -1.8315  , -2.8587  ,
  0.48424 , -2.0527  , -0.53808 , -2.3472  ,  1.0354  , -1.8257  ,
 -0.3892  , -0.24943 ,  0.8651  , -1.5195  ,  1.2166  , -2.698   ,
 -0.96698 ,  2.2175  , -0.16089 , -0.49677 , -0.19646 ,  1.3284  ,
  4.0824  ,  1.3919  ,  0.80669 , -1.0316  , -0.28056 , -1.8632  ,
  0.47716 , -0.53628 ,  1.3853  , -2.1755  , -0.2354  ,  2.4933  ,
 -0.87255 ,  1.4493  , -0.10778 , -0.44159 ,  1.3462  ,  4.4211  ,
 -1.8385  ,  0.3985  ,  0.47637 , -0.60074 ,  3.3583  , -0.15006 ,
 -0.40495 ,  2.7225  , -1.6297  ,  0.86797 , -4.1445  , -2.7793  ,
  1.1535  , -0.011691,  0.9792  , -1.0141  ,  0.80134 ,  0.43642 ,
  1.4337  ,  2.8927  ,  0.82871 , -1.1827  , -1.3838  ,  2.3903  ,
 -0.89323 ,  1.1461  , -1.7435  ,  0.8654  , -0.27075 , -0.78698 ,
  1.5631  , -0.5923  ,  0.098082, -0.26682 ,  1.6282  , -0.77495 ,
  3.2552  ,  1.7964  , -1.4314  ,  1.2336  ,  2.3102  , -1.6328  ,
  2.8366  , -0.71384 ,  0.43967 ,  1.5627  ,  3.079   , -0.922   ,
 -0.43981 , -0.7659  ,  1.9362  , -2.2479  ,  1.041   ,  0.63206 ,
  1.5855  ,  3.4097  , -2.9204  , -1.4751  , -0.59534 , -1.688   ,
 -4.1362  ,  2.745   , -2.8515  ,  3.6509  , -0.66993 , -2.8794  ,
  2.0733  ,  1.1779  , -2.0307  ,  2.595   , -0.12246 ,  1.5844  ,
  1.1855  ,  0.022385, -2.2916  , -2.2684  , -2.7537  ,  0.34981 ,
 -4.6243  , -0.96521 , -1.1435  , -2.8894  , -0.12619 ,  2.9577  ,
 -1.7227  ,  0.24757 ,  1.2149  ,  3.5349  , -0.95802 ,  0.080346,
 -1.6553  , -0.6734  ,  2.2918  , -1.8229  , -1.1336  ,  1.8884  ,
  2.4789  , -0.66061 ,  2.0529  , -0.76687 ,  0.32362 , -2.2579  ,
  0.91278 ,  0.36231 ,  0.61562 , -0.15396 , -0.42917 , -0.89848 ,
  0.17298 , -0.76978 , -2.0222  , -1.7127  , -1.5632  ,  0.56631 ,
 -1.354   ,  2.6261  ,  1.9156  , -1.5651  ,  1.8315  , -1.4257  ,
 -1.6861  , -0.51953 ,  1.7635  , -0.50722 ,  1.388   , -1.1012  ],
```

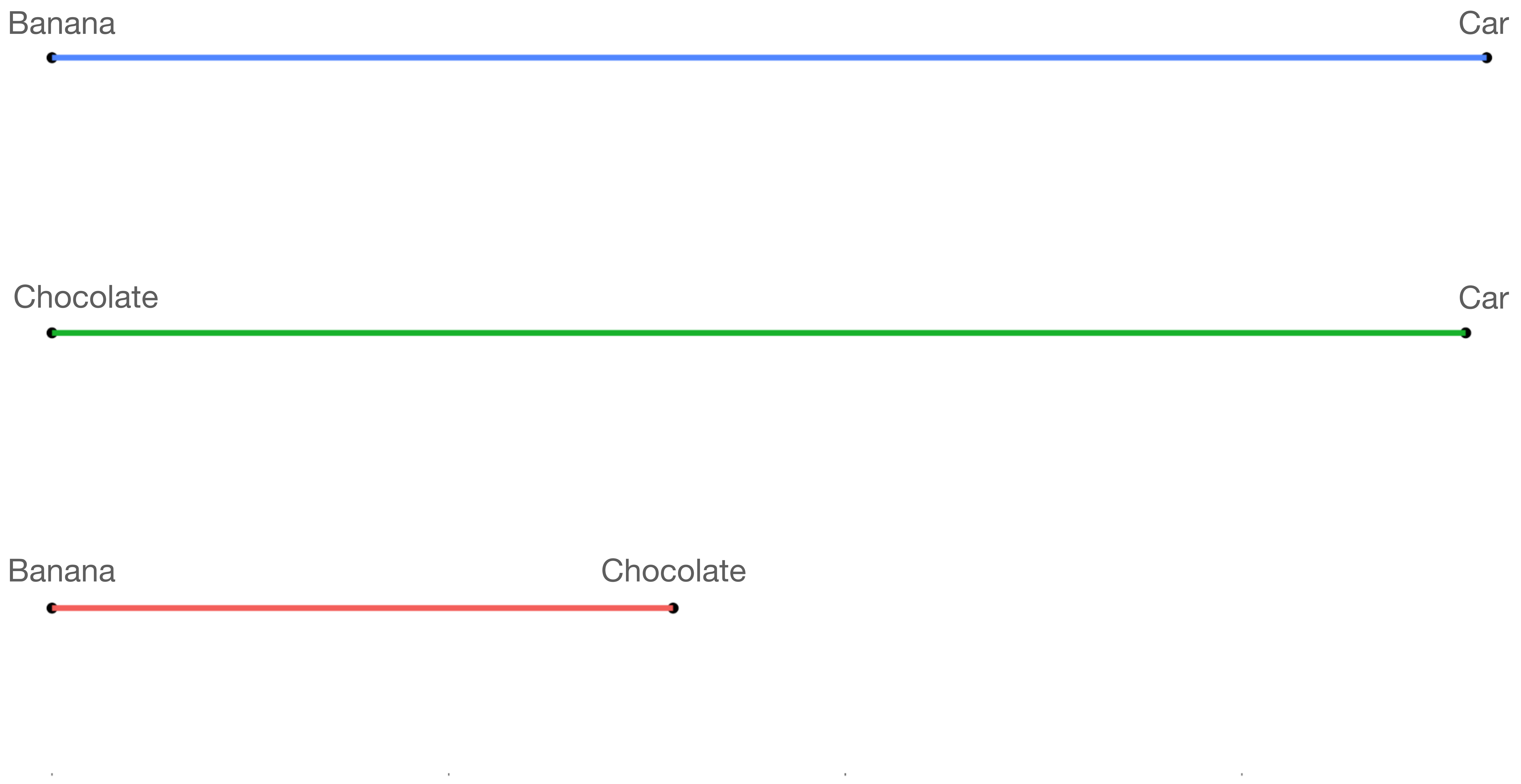
```
[-7.5251e-01, -3.3480e+00, -2.9293e+00,  3.6773e+00,  6.7698e-01,
 -4.6221e+00,  1.7471e+00,  2.9072e+00, -1.1218e+00,  1.9050e+00,
  6.1861e+00, -1.5307e+00, -2.1315e-01, -4.9000e-01,  3.1568e+00,
 -3.2417e+00,  9.3068e-02, -1.6506e+00,  1.8947e+00, -3.6223e+00,
 -1.4505e+00,  2.8421e+00, -1.6908e+00, -4.7524e-01, -5.5192e+00,
 -1.5492e+00, -3.2481e+00,  4.3969e+00, -9.3570e-01,  1.1385e+00,
  2.3373e+00,  2.3882e+00, -1.5618e+00, -7.5315e-01, -5.9527e-01,
 -1.9020e+00,  1.0459e+00,  2.2420e-01, -9.4108e-01,  4.8844e-01,
 -2.2083e+00,  2.5857e+00,  2.2875e+00,  6.3069e-01,  3.4058e+00,
  2.0758e+00, -1.0663e+00, -8.4464e-01, -5.2534e-02, -7.9447e-01,
  3.0140e+00, -8.9454e-01,  2.1576e+00, -3.0407e+00,  1.3439e+00,
 -2.1920e-01, -6.2846e-02,  1.1478e+00,  2.8001e+00,  2.6444e+00,
  2.6262e+00,  2.2010e-02,  1.4596e+00, -1.1558e+00,  1.8789e-01,
  9.4600e-01, -2.9744e+00, -2.2531e+00,  7.7054e-01, -5.4315e-01,
 -2.2618e+00,  2.2210e+00, -1.2964e+00,  1.0105e+00,  5.8169e-01,
  3.5617e-01, -2.4568e-01, -2.0808e+00,  3.5410e+00, -5.2889e-01,
 -2.8393e-01,  4.8163e-01,  1.7635e+00,  7.4050e-01,  6.7875e-01,
 -2.2662e-01,  4.8440e+00,  8.9114e-01, -2.5486e+00, -6.9544e-01,
  6.2643e-01,  2.9279e-01,  3.1008e+00, -3.2464e+00,  1.5747e+00,
 -1.1993e+00,  3.0120e+00, -1.0923e+00, -1.1773e+00, -2.2735e+00,
 -1.2936e+00, -1.3023e+00,  1.0400e+00, -9.1724e-01,  1.0221e+00,
  8.9763e-01, -3.4229e+00,  2.7322e+00, -2.2374e-01,  4.8981e-01,
  5.6333e-01, -1.3467e+00,  9.6163e-01, -7.6177e-01, -3.1454e-01,
 -1.3337e+00,  2.9608e+00, -3.7193e+00, -1.1941e+00, -1.0349e+00,
 -2.5313e+00,  1.7521e+00, -1.5778e+00,  5.4771e-02,  4.6839e-01,
 -1.9399e+00, -1.3847e-01, -3.9830e+00,  4.9884e+00, -4.8193e-01,
 -1.1701e+00,  6.2994e-01,  2.9822e+00, -2.4728e-01, -2.6371e+00,
  1.0801e+00, -2.2238e+00, -5.2203e+00,  1.5099e+00, -1.8284e+00,
 -3.9196e-01,  1.7773e+00, -6.8698e-01, -1.0951e+00, -1.5319e+00,
  2.8311e+00,  2.9736e-01, -1.5198e+00,  1.3076e+00,  5.9841e-01,
 -5.2798e-01, -5.5499e-01, -3.4522e+00, -1.3115e+00, -1.0095e+00,
  3.3329e+00,  6.5440e+00, -1.3999e+00,  2.3499e+00, -2.4218e+00,
  6.9150e-01, -1.4240e+00,  3.3080e-01, -1.2254e+00,  3.7678e+00,
 -1.6502e+00, -1.6829e+00,  2.3409e-01,  8.3192e-01,  2.0174e+00,
 -2.6225e+00, -3.7696e-01, -2.1272e-01,  3.4416e+00, -3.6619e+00,
 -2.1298e+00,  9.7029e-01, -5.1133e-02,  8.2768e-01, -1.2364e+00,
  5.9028e-02, -5.1805e-01, -1.0821e+00, -1.7695e+00, -2.9489e-02,
 -2.5980e+00, -4.9045e-02, -6.9158e-01,  9.1347e-01, -4.6027e-01,
 -1.9500e+00,  2.0457e+00, -1.7526e+00,  2.7582e+00,  3.6836e-02,
 -2.3929e+00, -1.3635e+00,  2.1516e+00,  1.1975e+00, -1.2935e+00,
  1.4003e+00, -1.5616e+00, -7.1990e-01, -1.2839e-01,  1.5071e+00,
 -2.8197e+00, -5.9906e-01, -3.1609e-01,  4.8745e+00,  1.7453e+00,
  4.0927e+00, -5.4239e-01, -6.3825e-04,  3.3456e+00,  1.4135e+00,
  2.0070e-01,  1.8593e+00,  1.0568e+00, -2.4357e+00,  2.3165e+00,
  5.5872e-01, -1.6893e+00, -2.2931e+00,  1.6865e+00, -2.2543e+00,
 -1.6019e+00,  6.9584e-01,  2.5392e+00, -3.3192e-01,  3.2114e+00,
 -2.1623e+00, -9.7765e-01, -8.8937e-01, -5.1731e-01, -2.1990e+00,
  4.1397e+00,  4.2648e-01,  4.6854e+00,  1.0355e+00,  1.4013e+00,
 -1.0843e-01, -5.9694e-01, -4.0420e-01,  4.2305e+00, -5.2332e-01,
  4.1959e+00, -2.2805e-02, -6.3223e-01, -6.5072e-01, -1.7390e+00,
 -3.0675e+00,  3.9072e+00, -2.6396e+00, -2.4627e+00, -3.1164e-01,
 -2.5056e+00, -1.6382e+00,  3.2290e+00, -2.6652e+00, -7.3372e-01,
  9.6020e-01,  4.3881e+00, -1.2500e+00,  1.2498e+00,  1.9080e-01,
  1.9253e+00,  1.8284e+00, -2.3579e+00, -3.3646e+00,  6.8795e-01,
  1.2263e+00, -9.3136e-01,  5.5192e-01,  1.1171e+00, -2.8175e+00,
 -2.6307e+00,  1.4002e-01,  3.1652e-01, -5.7089e-01, -1.2883e+00,
  9.8610e-01, -1.0684e+00, -7.9920e-02, -2.6351e+00, -1.4276e+00,
 -5.3942e-01, -1.3570e+00, -6.0974e-01, -2.2030e+00,  2.0585e+00,
 -4.2257e-01, -1.5917e+00, -1.1557e+00, -2.8138e+00, -2.9554e+00,
```

```
  4.5855  ,   2.4556  ,  -8.5233  ,  -6.0595  ,  -0.44879 ,
 -2.5409  ,   4.3721  ,   1.4889  ,   4.6075  ,   6.7933  ,
  1.5461  ,  -1.7807  ,  -4.8333  ,   5.9001  ,  -6.2238  ,
  6.7788  ,   1.0836  ,  -6.5442  ,   1.6709  ,  -1.0685  ,
  2.4635  ,   0.90953 ,   7.8345  ,  -8.0876  ,   2.8703  ,
 -2.8004  , -10.297   , -12.034   ,  -4.4031  ,   1.2064  ,
  1.4787  ,  -1.4525  ,   0.93624 ,   0.32869 ,   6.1455  ,
  1.0993  ,  -2.8943  ,   6.538   ,   4.7175  ,  -7.5675  ,
 -5.0228  ,  -6.5218  ,   9.7911  ,   7.2253  ,  -0.95069 ,
 -6.3417  ,  -0.99102 ,  -1.7487  ,  -0.71064 ,   0.81952 ,
  4.1403  ,  -2.6224  ,   4.3984  ,  -5.2205  ,  -0.31186 ,
 -2.8558  ,  10.741   ,  -3.455   ,   9.7063  ,   7.9873  ,
 -2.9555  ,  -4.9335  ,  -8.9013  ,  -1.2512  ,   2.2814  ,
 -4.6709  ,  -4.877   ,  -1.6662  ,  -0.72899 ,   5.534   ,
  0.30023 ,   5.295   ,   5.5527  ,  -3.568   ,   9.7351  ,
  1.401   ,  -4.5543  ,  -2.4768  ,  -4.3919  ,  -0.34426 ,
 -3.2707  ,   6.7334  ,   5.224   ,  -1.8686  ,  -4.2507  ,
 -8.2556  ,   5.9146  ,   3.8646  ,  -0.64003 ,  -1.9898  ,
 -1.6434  ,   1.5526  ,   0.52192 ,  -4.1706  ,   0.33223 ,
  0.22254 ,   6.2197  ,   5.5292  ,   8.1494  ,   3.3193  ,
  2.8298  ,   2.0664  ,   2.8186  ,  -2.9645  ,   2.6404  ,
  5.7953  ,  -5.7991  ,   0.96114 ,  -2.2392  ,  -4.7114  ,
 -3.589   ,  -3.8583  ,   7.1382  ,  -1.2646  ,  -7.4677  ,
  6.0765  ,  -2.0516  ,  -5.8281  ,   8.9571  ,  -0.15937 ,
 -0.87378 ,  -1.0916  ,  -3.9298  ,  14.587   ,  -4.0177  ,
 -2.2886  ,   2.5616  ,   2.3827  ,   5.1763  ,   3.0547  ,
 -1.8685  ,  -2.0272  ,  -0.6823  ,  -2.3923  ,  -1.8071  ,
 -2.6376  ,   3.2747  ,   3.0808  ,   7.5008  ,  12.377   ,
 -2.7327  ,  -7.8162  ,   0.46069 ,  -4.7697  ,  -3.6059  ,
  7.0717  ,   4.5885  ,  -3.2425  , -13.127   ,  10.668   ,
  1.2418  ,  -6.419   ,  10.045   ,   3.2114  ,   6.9431  ,
 -4.4162  ,   4.0692  ,   1.7755  ,  -1.4207  ,  -0.34161 ,
 -0.73324 ,  -3.193   ,   3.9874  ,  -4.3321  ,  -4.7153  ,
  0.39614 ,  -0.01535 ,   1.5067  ,  -3.6085  ,   3.9091  ,
 -0.32986 ,   3.4754  ,  -2.5639  ,   0.33682 ,   2.9079  ,
  4.0339  ,  -6.8811  ,   2.8709  ,  -0.4951  ,  -2.4399  ,
  1.0046  ,   7.8473  ,   4.1704  ,  -0.38873 ,  -3.2056  ,
  3.0839  ,   3.5669  ,  -0.099299,   2.2419  ,  -3.0416  ,
  0.74202 ,   3.5789  ,   7.6271  ,  -9.3661  ,  -1.7087  ,
 -2.341   ,   0.10248 ,   7.3664  ,   5.7592  ,   3.2057  ,
 -1.3789  ,  -1.7753  ,   6.7445  ,   2.1774  ,  -7.3915  ,
 -0.54382 ,  -4.808   ,  -1.6732  ,  -1.9883  ,  -2.4419  ,
 -1.6932  ,  -3.2231  ,   2.3777  ,   6.5671  ,   1.9828  ,
 -0.38792 ,   0.23489 ,  -5.1306  , -12.722   ,  -3.4294  ,
 -5.315   ,  -2.0044  ,  -4.1133  ,  -4.3514  ,  -5.6132  ,
 -6.3817  ,   2.023   ,   4.3316  ,   2.356   ,  -7.6913  ,
  1.5544  ,  -4.0399  ,  -0.64259 ,   3.8112  ,  -0.80175 ,
  7.2695  ,  -7.7816  ,   9.1579  ,   2.7312  ,   1.2637  ,
 -2.1794  ,  -2.1183  ,   3.0157  ,  -2.0421  ,   1.882   ,
 -4.989   ,  -2.5153  ,   1.2071  ,  -3.9638  ,  -1.9256  ,
  0.036288,   7.608   ,  -2.0806  ,  -1.9005  ,  -5.6806  ,
  1.8321  ,   3.7442  ,  -5.1673  ,   3.5059  ,  -4.8036  ,
  1.6082  ,  10.746   ,  -7.6125  ,  -1.9775  ,   2.6007  ,
 -3.0985  ,  -1.7584  ,   0.64261 ,  -3.3654  ,   3.7503  ,
 -3.4817  , -12.058   ,  -4.497   ,   7.2051  ,   2.7354  ,
-10.113   ,  -4.4291  ,   5.7432  ,   5.1143  ,  -2.786   ,
 -3.6188  ,   4.246   ,   9.1744  ,  -1.1803  ,   1.997   ,
 -3.9817  ,  -8.2793  ,   0.36314 , -11.65    ,   0.18214 ,
 -7.0462  ,  -8.339   ,   0.64806 ,   0.73438 ,  -4.7653  ,
```


BANANA


CHOCOLATE


CAR


Banana
Car
Chocolate
Car
Banana
Chocolate

R / GitHub refresher