

Discovery

Gary King
Institute for Quantitative Social Science
Harvard University

covering joint work with
Justin Grimmer (Harvard) and Eleanor Powell (Harvard ↔ Yale)

(talk @ Institute for Qualitative and Multimethod Research, Syracuse University, 5/26/09)

- Gary King and Eleanor Neff Powell. 2008. “How Not to Lie Without Statistics”
- Justin Grimmer and Gary King. 2009. “Quantitative Discovery of Qualitative Information: A General Purpose Document Clustering Methodology”

<http://GKing.harvard.edu>

Is Qualitative or Quantitative Research Better?

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

If sufficient information is quantified, statistics wins

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative

If sufficient information is quantified, statistics wins

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little
- 6 crude variables beat 83 law professors (Martin et al., 2004)

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little
- 6 crude variables beat 83 law professors (Martin et al., 2004)
- Simple quantitative models forecast elections better than expert pundits (Gelman and King, 1993; Campbell, 2005)

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little
- 6 crude variables beat 83 law professors (Martin et al., 2004)
- Simple quantitative models forecast elections better than expert pundits (Gelman and King, 1993; Campbell, 2005)
- 284 experts forecast the political future with “less skill than simple extrapolation” (Tetlock, 2005)

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little
- 6 crude variables beat 83 law professors (Martin et al., 2004)
- Simple quantitative models forecast elections better than expert pundits (Gelman and King, 1993; Campbell, 2005)
- 284 experts forecast the political future with “less skill than simple extrapolation” (Tetlock, 2005)
- statistical model out-performs physicians in determining cause-specific mortality rates (King and Lu, 2008)

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little
- 6 crude variables beat 83 law professors (Martin et al., 2004)
- Simple quantitative models forecast elections better than expert pundits (Gelman and King, 1993; Campbell, 2005)
- 284 experts forecast the political future with “less skill than simple extrapolation” (Tetlock, 2005)
- statistical model out-performs physicians in determining cause-specific mortality rates (King and Lu, 2008)
- Hundreds of head-to-head contests, mostly with the same conclusion (Meehl, 1954; Grove, 2005)

Is Qualitative or Quantitative Research Better?

This question has an answer (hard to know from the qualitative methods literature!)

If insufficient information is quantified, qualitative judgment wins

- The vast majority of human decisions are qualitative
- If not, we wouldn't have survived as a species

If sufficient information is quantified, statistics wins

- The information required: shockingly little
- 6 crude variables beat 83 law professors (Martin et al., 2004)
- Simple quantitative models forecast elections better than expert pundits (Gelman and King, 1993; Campbell, 2005)
- 284 experts forecast the political future with “less skill than simple extrapolation” (Tetlock, 2005)
- statistical model out-performs physicians in determining cause-specific mortality rates (King and Lu, 2008)
- Hundreds of head-to-head contests, mostly with the same conclusion (Meehl, 1954; Grove, 2005)
- The **march of quantification** across fields of academia, professions, commerce, sports, etc. (Moneyball, SuperCrunchers, Numerati) ☰ ↻ 🔍

3 Steps

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

- **Quantitative**: focus on measurement and verification \rightsquigarrow assumes & underplays conceptualization

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

- **Quantitative**: focus on measurement and verification \rightsquigarrow assumes & underplays conceptualization
- **Qualitative**: focus on (& iterate between) conceptualization and measurement \rightsquigarrow ignores or underplays verification

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

- **Quantitative**: focus on measurement and verification \rightsquigarrow assumes & underplays conceptualization
- **Qualitative**: focus on (& iterate between) conceptualization and measurement \rightsquigarrow ignores or underplays verification

Proposed Solutions

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

- **Quantitative**: focus on measurement and verification \rightsquigarrow assumes & underplays conceptualization
- **Qualitative**: focus on (& iterate between) conceptualization and measurement \rightsquigarrow ignores or underplays verification

Proposed Solutions

- Do both qualitative & quantitative analysis, separately

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

- **Quantitative**: focus on measurement and verification \rightsquigarrow assumes & underplays conceptualization
- **Qualitative**: focus on (& iterate between) conceptualization and measurement \rightsquigarrow ignores or underplays verification

Proposed Solutions

- Do both qualitative & quantitative analysis, separately
- Atheoretical unsupervised learning algorithms

3 Steps

- 1 **Conceptualization** (e.g., a categorization scheme)
- 2 **Measurement** (e.g., classifying objects into categories)
- 3 **Verification** (e.g., testing a hypothesis that could be wrong)

Styles & Approaches (the cause of many misunderstandings)

- **Quantitative**: focus on measurement and verification \rightsquigarrow assumes & underplays conceptualization
- **Qualitative**: focus on (& iterate between) conceptualization and measurement \rightsquigarrow ignores or underplays verification

Proposed Solutions

- Do both qualitative & quantitative analysis, separately
- Atheoretical unsupervised learning algorithms
- \rightsquigarrow Integrated quant/qual: **Computer-assisted qualitative analysis**

The Problem: Discovery from Unstructured Text

- Examples: scholarly literature, news stories, medical information, blog posts, comments, product reviews, emails, social media updates, audio-to-text summaries, speeches, press releases, legal decisions, etc.

The Problem: Discovery from Unstructured Text

- Examples: scholarly literature, news stories, medical information, blog posts, comments, product reviews, emails, social media updates, audio-to-text summaries, speeches, press releases, legal decisions, etc.
- 10 minutes of worldwide email = 1 LOC equivalent

The Problem: Discovery from Unstructured Text

- Examples: scholarly literature, news stories, medical information, blog posts, comments, product reviews, emails, social media updates, audio-to-text summaries, speeches, press releases, legal decisions, etc.
- 10 minutes of worldwide email = 1 LOC equivalent
- An essential part of discovery is **classification**: “one of the most central and generic of all our conceptual exercises. . . . the foundation not only for conceptualization, language, and speech, but also for mathematics, statistics, and data analysis. . . . Without classification, there could be no advanced conceptualization, reasoning, language, data analysis or, for that matter, social science research.” (Bailey, 1994).

The Problem: Discovery from Unstructured Text

- Examples: scholarly literature, news stories, medical information, blog posts, comments, product reviews, emails, social media updates, audio-to-text summaries, speeches, press releases, legal decisions, etc.
- 10 minutes of worldwide email = 1 LOC equivalent
- An essential part of discovery is **classification**: “one of the most central and generic of all our conceptual exercises. . . . the foundation not only for conceptualization, language, and speech, but also for mathematics, statistics, and data analysis. . . . Without classification, there could be no advanced conceptualization, reasoning, language, data analysis or, for that matter, social science research.” (Bailey, 1994).
- We focus on **cluster analysis**: discovery through (1) classification and (2) simultaneously inventing a classification scheme

The Problem: Discovery from Unstructured Text

- Examples: scholarly literature, news stories, medical information, blog posts, comments, product reviews, emails, social media updates, audio-to-text summaries, speeches, press releases, legal decisions, etc.
- 10 minutes of worldwide email = 1 LOC equivalent
- An essential part of discovery is **classification**: “one of the most central and generic of all our conceptual exercises. . . the foundation not only for conceptualization, language, and speech, but also for mathematics, statistics, and data analysis. . . Without classification, there could be no advanced conceptualization, reasoning, language, data analysis or, for that matter, social science research.” (Bailey, 1994).
- We focus on **cluster analysis**: discovery through (1) classification and (2) simultaneously inventing a classification scheme
- (We analyze text; our methods apply more generally)

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)
- $Bell(5) = 52$

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)
- $Bell(5) = 52$
- $Bell(100) \approx$

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)
- $Bell(5) = 52$
- $Bell(100) \approx 10^{28} \times$ Number of elementary particles in the universe

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)
- $Bell(5) = 52$
- $Bell(100) \approx 10^{28} \times$ Number of elementary particles in the universe
- Now imagine choosing the *optimal* classification scheme by hand

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)
- $Bell(5) = 52$
- $Bell(100) \approx 10^{28} \times$ Number of elementary particles in the universe
- Now imagine choosing the *optimal* classification scheme by hand
- **Qualitative-only approaches are hopeless**

Why Johnny Can't Classify (Optimally)

- $Bell(n)$ = number of ways of partitioning n objects
- $Bell(2) = 2$ (AB, A B)
- $Bell(3) = 5$ (ABC, AB C, A BC, AC B, A B C)
- $Bell(5) = 52$
- $Bell(100) \approx 10^{28} \times$ Number of elementary particles in the universe
- Now imagine choosing the *optimal* classification scheme by hand
- **Qualitative-only approaches are hopeless**
- That we think of all this as astonishing ... is astonishing

Why HAL Can't Classify Either

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps,...

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps, . . .
 - **Well-defined** statistical, data analytic, or machine learning foundations

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps, . . .
 - **Well-defined** statistical, data analytic, or machine learning foundations
 - How to add substantive knowledge:

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps, . . .
 - **Well-defined** statistical, data analytic, or machine learning foundations
 - How to add substantive knowledge: With few exceptions, **who knows?!**

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps, . . .
 - **Well-defined** statistical, data analytic, or machine learning foundations
 - How to add substantive knowledge: With few exceptions, **who knows?!**
 - The literature: **little guidance on when methods apply**

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps, . . .
 - **Well-defined** statistical, data analytic, or machine learning foundations
 - How to add substantive knowledge: With few exceptions, **who knows?!**
 - The literature: **little guidance on when methods apply**
 - **Deriving such guidance**: difficult or impossible

Why HAL Can't Classify Either

- The Goal — an optimal application-independent cluster analysis method — is mathematically impossible:
 - **No free lunch theorem**: every possible clustering method performs equally well on average over all possible substantive applications
- Existing methods:
 - **Many choices**: model-based, subspace, spectral, grid-based, graph-based, fuzzy k -modes, affinity propagation, self-organizing maps, . . .
 - **Well-defined** statistical, data analytic, or machine learning foundations
 - How to add substantive knowledge: With few exceptions, **who knows?!**
 - The literature: **little guidance on when methods apply**
 - **Deriving such guidance: difficult or impossible**
 - (Perhaps true by definition in unsupervised learning: If we knew the DGP, we wouldn't be at the discovery stage.)

If Ex Ante doesn't work, try Ex Post

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)
 - For discovery (our goal): No problem

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)
 - For discovery (our goal): No problem
 - For estimation & confirmation: more difficult or biased

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)
 - For discovery (our goal): No problem
 - For estimation & confirmation: more difficult or biased
- Complicated concepts are easier to define ex post:

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)
 - For discovery (our goal): No problem
 - For estimation & confirmation: more difficult or biased
- Complicated concepts are easier to define ex post:
 - “I know it when I see it” (Justice Stewart's definition of obscenity)

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)
 - For discovery (our goal): No problem
 - For estimation & confirmation: more difficult or biased
- Complicated concepts are easier to define ex post:
 - “I know it when I see it” (Justice Stewart's definition of obscenity)
 - Anchoring Vignettes (on defining concepts by example)

If Ex Ante doesn't work, try Ex Post

- **Methods and substance must be connected** (no free lunch theorem)
- The usual approach fails: can't do it by understanding the model
- We do it **ex post** (by qualitative choice)
 - For discovery (our goal): No problem
 - For estimation & confirmation: more difficult or biased
- Complicated concepts are easier to define ex post:
 - “I know it when I see it” (Justice Stewart's definition of obscenity)
 - Anchoring Vignettes (on defining concepts by example)
- **But how to choose from an enormous list of clusterings?**

Our Idea: Meaning Through Geography

Our Idea: Meaning Through Geography

wide at **SuperPages.com**

195

Car

C

Cartage New England Inc
36 Adams Ln Ipswich 01908..... 978 356-9960

Cartagena Lydia
19 Levitt Bus 02118..... 617 323-7639

Cartagena Avelth
9 Pleasant Row 02119..... 617 442-0780
B Hot 02136..... 617 361-5253

Carte Nicholas
12500 N. Green Can 02139..... 617 576-1061

Carte Nicholas
18 Appleton Station 02114..... 617 695-6996

Cartegosa O & Millard Bus 02118..... 617 338-8219

Carton Tho J Sr & Claire
1 Poydell Rd Mt 02181..... 617 698-6163

Thomas & Kuthbert
50 Thompson Ln 02186..... 617 696-6919

Carter A Bus 02111..... 617 327-2257

Carter A Industry..... 617 442-5230

Carter A 31 Bethune Wy Roxbury 02119..... 617 442-1219

Carter A 208 Johnson Av Cambridge 02139..... 617 492-4174

A M 250 Worcester Av 02115..... 617 266-7153

Adams 381 Centre St Mt 02181..... 617 698-9074

Alice 108 Elmwood St 02115..... 617 425-0193

Alice 45 Market Cambridge 02139..... 617 945-2711

Andrew F 27 West Av Som 02143..... 617 625-7623

Carter Angus MD
118 Beacon Dr 02144..... 617 739-1022

Carter Athens
777 Newbury Boston 02114..... 617 536-6329

B E 68 Chatham Av Mat 02124..... 617 296-6911

Carter Barbara L MD
15th New England Medical Center Box 02111

Carter Becky Jo 02114..... 617 636-0051

Carter Becky Jo 02114..... 617 523-4368

Carter Bernard J
121 State C Bus 02118..... 617 567-3430

Bitshah 20 Minkley Dr 02134..... 617 298-8713

Blake 26 W Irving Bus 02118..... 617 367-9911

Blair 30 Park Pl Bus 02116..... 617 423-0020

Carter B Burgess Consultants Inc
71 East C Bus 02141..... 617 225-0200

Carter C 2000 Concord Av Mt 02136..... 617 782-2118

Carter C 2000 Concord Av Mt 02136..... 617 569-1545

C 359 Harvard Can 02138..... 617 491-4822

C 639 Brook Hb Mt 02136..... 617 296-6392

C B 88 41 Burroughs Ave 02118..... 617 524-2524

Carter F 51 Woburn Bus 02116..... 617 327-1105

Faye & Ricky
107 Columbia Av Bus 02118..... 617 437-7331

Francis S 184 Temple W Bus 02135..... 617 323-6963

Franklin & Anne
251 Mt Auburn Can 02138..... 617 354-6798

Fred 45 Newmarket Avm 02139..... 617 524-3078

Fred W Hinchey Rd Mt 02136..... 617 698-1343

G & B R Harvard Bus 02134..... 617 434-9906

G T 27 Franklin Av Som 02145..... 617 623-7121

Gayle 25 Promenade Dr 02134..... 617 825-0322

Geo S 115 Woburn Hb Mt 02118..... 617 522-2215

George 125 Hudson Bus 02114..... 617 367-9548

Carter Halliday Associates
187 S Dwyer Bus 02111..... 617 456-1689

Carter Harry E
38 Bayway Hb W Bus 02135..... 617 325-5465

Carter Hide Co Inc
145 Summer Bus 02115..... 617 542-7907

Carter Hilary 41 Harvey Can 02138..... 617 874-2750

Horace
942 Walnut Av Roxbury 02119..... 617 442-5307

Howard Jr 28 Ntra Dns Mt 02119..... 617 445-5552

J Can..... 617 354-2680

J 15 Chatham Bus 02146..... 617 232-7990

J 518 Harvard Bn 02146..... 617 730-9483

J 775 Ws Pkwy West Roxbury 02132..... 617 323-5574

Carter J Jacques MD
1 Brookline Pl Bus 02146..... 617 735-8787

Carter J M
1415 Columbia St Bus 02117..... 617 464-1040

Carter J M Ornamental Ironworks
1000 North St Som 02145..... 617 436-5353

Carter J Veal Co
50 Newmarket St Bus 02118..... 617 442-1775

Carter James
1573 Cambridge St Can 02138..... 617 492-1214

James 102 Fisher Av Roxbury 02139..... 617 739-2193

James 27 Gold Star Rd Cambridge 02140..... 617 876-8841

Jan 1 14 Rockwood Rd Mt 02135..... 617 363-0773

Janis 114 Adams Rd Newton 02460..... 617 964-9435

Jeffrey 41 Warren Av 02118..... 617 987-2163

John 15 Mountford Rd 02119..... 617 426-5999

John 227 Summer Bus 02118..... 617 423-4334

John 45 Westford Rd Som 02149..... 617 252-1235

June O 329 A Summit Av Bn 02136..... 617 734-6199

K 38 Browning Av Dorchester 02124..... 617 265-9456

K 77 Forest Dorchester 02122..... 617 282-1593

Carter Nella E
323 Manchester Av Bus 02115..... 617 267-6483

Nicholas S F
115 Northrop Av Mt 02186..... 617 498-5307

Nick 21 Fairford Bus 02116..... 617 267-5222

Nick & Debbie
106 Harvard Rd Newton 02459..... 617 527-0480

Nicole..... 617 698-0713

Norman G
38 Chickadee Dr 02132..... 617 822-1203

P 46 Cranwood Pl Bus 02135..... 617 427-4754

P E 501 E South St Bus 02127..... 617 266-4213

P L 48 Hollings Bus 02121..... 617 427-0170

P R 10 Bonnet Can 02138..... 617 983-9692

Paul & Constance
114 Anson Av W Bus 02138..... 617 325-2036

Paul E 200 E South St Bus 02127..... 617 268-4546

Paul M 27 Union Pl 02135..... 617 787-2115

Carter Pier Driving Inc LP Beaver 13
Vermont 02192..... Middleboro Twp 781 235-8488

Carter Prudence
40 Franklin Wadsworth 02127..... 617 393-3782

Prudence
49 Francis Wadsworth 02127..... 617 926-7063

Reginald
386 Braintree Dorchester 02122..... 617 541-2843

Renée & Andrew
10 Walnut Bus 02108..... 617 720-3765

Carter Rice David
Buckley Center Publishing 511 Main Wilmington 01887
161 Free Del T & Thru..... 800 638-1671

Carter R 100 Cambridge Prof 613 Main Wilmington
161 Free Del T & Thru..... 500 619-7447

Carter R 100 Cambridge Prof 613 Main Wilmington
161 Free Del T & Thru..... 800 648-7447

Headquarters 613 Main Wilmington 01887
161 Free Del T & Thru..... 978 988-7447

Ingalls Center 143 Main Wilmington 01887
161 Free Del T & Thru..... 800 638-1673

Carter Richard
1879 Concord Av Brighton 02117..... 617 987-0836

Richard A 97 Mt Vernon Bus 02136..... 617 566-7293

Carter Richard A MD
176 Concord Av Bus 02116..... 617 267-0710

Richard S 3 Worcester St Bus 02132..... 617 268-0448

Robert L 175 Schinola Av Can 02116..... 617 864-1535

Roger 120 South St Bus 02115..... 617 424-6148

Roy M Concord Av Can 02116..... 617 491-6115

Royle 18 Sepia Way 02128..... 617 241-6114

C

Our Idea: Meaning Through Geography

wide at SuperPages.com

195

Car

C

17 566-1282 Carriage New England Inc
36 Main St Ipswich 01938... 978 356-9960

17 447-4101 Cartegama Lydia
19 Jordan Ave 02111... 617 323-7639

800 257-9981 Cartegama Avelth
F Pleasant Hill 02119... 617 442-0780
B Hvt 02119... 617 361-5253

17 566-1282 Jessica 50 Otisbur Ct 02129... 617 241-0152

17 364-5188 Lucille 134 Harvard Cam 02139... 617 491-5621
M 95 Sweet Rd 02139... 617 323-9713

361-0380 Melvin 91 Green Can 02139... 617 576-1061

Carre Nicholas
18 Apollon Street 02114... 617 695-6996

17 628-8248 Cartegama O 4 Willow St 02118... 617 338-8219

Carton Tho & S Clave
1 Poydras Rd MO 02188... 617 698-6163

17 445-5116 Thomas S Kuthbier
50 Thompson Ln 02186... 617 696-6919

17 822-2982 Carter A 80 02111... 617 327-2257

17 627-5712 A Industry... 617 442-5230

17 569-2698 A 31 Bethune Wy Roxbury 02119... 617 442-1219

17 667-5190 A M 250 Main St 02115... 617 492-4174

17 569-1417 Adams 381 Corbett St 02118... 617 698-9074

17 569-1417 Alice 108 Elmwood St 02115... 617 425-0193

17 338-9110 Andrew F 22 Vine St 02143... 617 625-7623

17 825-9158 Carter A 181
1181 Beacon Dr 02146... 617 739-1022

17 296-1593 Carter Athens
777 Newbury Boston 02116... 617 536-6329

17 670-2078 B E 68 Chatham St MO 02136... 617 296-6911

17 623-7001 Carter Barbara L MD
150 New England Medical Center Box 02111

17 296-4275 Carter Becky S 02114... 617 636-0051

17 542-1521 Bernard J
711 State St 02118... 617 567-3430

17 364-5232 Bishop 25 Minkley Dr 02114... 617 298-8713

17 541-5649 Blake 26 W Irving St 02118... 617 367-9911

17 739-2662 Carter Broussard Co
26 Park Pl Box 02116... 617 423-0201

17 679-8030 Carter & Burgess Consultants Inc
71 East St 02114... 617 225-0200

17 541-3948 Carter C 2800 Commonwealth Ave 02116... 617 782-2118

17 436-1513 Carter D 100 Forest Hill 02118... 617 569-1545

17 569-4119 C 359 Harvard Cam 02138... 617 491-8222

17 436-1513 C 40 Brook Hill 02118... 617 296-6392

800 569-8782 C B 41 Burroughs Ave 02118... 617 326-9238

17 327-1105 Carter F 24 Woburn St 02116... 617 327-1105

Faye & Ricky
107 Columbia Av Box 02116... 617 437-7331

Francis S 114 Temple St Box 02116... 617 323-6963

Franklin & Anne
251 Mt Auburn Cam 02138... 617 354-6798

Fred 47 Waverfield Ave 02138... 617 524-3078

Fred W Hinchey Rd 02138... 617 698-1343

G & R Harvard 02124... 617 436-9906

G T 27 Franklin Av Son 02116... 617 623-7121

Gayle 25 Promenade Dr 02114... 617 825-0322

Geo S 115 Woburn Hill Ave 02116... 617 522-2215

George 125 Nantua Box 02114... 617 367-9548

Carter Halliday Associate
187 S Deyou St 02111... 617 456-1689

Carter Harry E
18 Bayway Rd W 02118... 617 325-5465

Carter Hide Co Inc
140 Summer St 02111... 617 542-7867

Carter Henry 41 Newry Cam 02116... 617 876-2750

Horace
342 Walnut Av Roxbury 02119... 617 442-5307

Howard Jr 28 Nebra Dns 02119... 617 445-5552

J Cam
415 Cambridge St 02146... 617 354-2680

J 15 Chatham St 02146... 617 232-7990

J 518 Harvard Bldg 02146... 617 730-9483

J 775 Wey Plow West Roxbury 02132... 617 323-5574

Carter J JAQUES MD
1 Brookline Pl Box 02146... 617 735-8787

Carter J M
1410 Columbia St S Box 02117... 617 464-1040

Carter J M Ornamental Ironworks
1000 Washington St 02116... 617 343-5353

Carter J Veal Co
50 Newmarket St 02116... 617 442-1775

Carter James
1573 Cambridge St Cam 02138... 617 492-1214

James 102 Fisher Av Roxbury 02138... 617 739-2193

James
27 Gold Star Rd Cambridge 02140... 617 876-8841

Jan L 140 Reservoir Rd Mt 02118... 617 363-0773

John 114 Adams Rd Newton 02459... 617 964-9435

Jeffrey 41 Warren Av Box 02111... 617 426-5994

John 15 Mountford St 02119... 617 987-2163

John 127 Summer St 02118... 617 423-4334

John 45 Mountford Rd St 02119... 617 252-1235

June O 329 A Summit Av Box 02136... 617 734-6199

K 38 Browning Av Dorchester 02124... 617 265-9456

K 77 Forest Dr Dorchester 02122... 617 282-1972

Carter Nella E
323 Manchester Av Box 02115... 617 267-6483

Nicholas S F
115 Rumpheys Av MO 02186... 617 698-5307

Nick 21 Fairfield Box 02116... 617 267-5222

Nick & Debbie
104 Harvard Hill Newton 02459... 617 527-0480

Nicole
38 Chikobutov Dr 02135... 617 822-1203

P 46 Cranston Pl 02119... 617 427-4754

P E 101 E South St Box 02127... 617 266-4213

P L 44 Hurlingham Box 02111... 617 427-9170

P R 10 Fenwick Ave 02138... 617 983-5692

Paul & Constance
114 Anson Av W Box 02138... 617 325-2036

Paul G 201 E South St Box 02127... 617 268-4546

Paul M 27 Union St 02135... 617 787-2115

Frankington 02132... 617 266-4888

Carter Prudence
40 Franklin Waltham 02127... 617 393-3782

Prudence
40 Franklin Waltham 02127... 617 986-7063

Reginald
386 Braintree Dorchester 02122... 617 541-2843

Renee & Andrew
30 Walnut St 02108... 617 720-3765

Carter Rice David
Bosley Centre Publishing 513 Main Wilmington 01887
140 Free Dal T & Thon... 800 638-1671

Richard 1000 Industrial Park 613 Main Wilmington
140 Free Dal T & Thon... 800 619-7447

Carl See Printing 613 Main Wilmington
140 Free Dal T & Thon... 800 648-7447

Headquarters 613 Main Wilmington 01887
Carl... 978 968-7447

Ingalls Center 143 Main Wilmington 01887
140 Free Dal T & Thon... 800 638-1673

Carter Richard T
1879 Commonwealth Av Brighton 02117... 617 987-0836

Carl See Printing 613 Main Wilmington 01887... 617 566-7293

Carter Richard A MD
176 Commonwealth Av Box 02116... 617 267-0710

Richard K
13 Worcester St Box 02137... 617 268-0448

Robert L 175 Scholastic Av Cam 02116... 617 564-1535

Roger 100 Concord St 02115... 617 424-6148

Roy 41 Concord Av Cam 02116... 617 491-6115

Royle 18 Sepia Way 02129... 617 241-8418

C



Our Idea: Meaning Through Geography

wide at SuperPages.com

195

Car

C

17 566-1282	Carriage New England Inc 36 Main St Ipswich 01938... 978 356-9960	237 327-1105	Carter F 30 Hickory Hill 02115... 617 327-1105	323 267-6483	Carter Nella E 323 Manchester Ave Box 02115... 617 267-6483
17 447-4101	Cartagena Lydia 19 Jordan Ave 02115... 617 323-7639	617 437-7331	Faye & Ricky 107 Columbia Ave Box 02115... 617 437-7331	617 498-5307	Nicholas S F 115 Northgate Ave 02186... 617 498-5307
800 257-9981	Cartagena Avelth 9 Pleasant Hill 02119... 617 442-0780 R Hot 02119... 617 361-5253	617 524-3078	Franklin & Anne 721 Mt Auburn Court 02130... 617 354-4798	617 267-5222	Nick 21 Fairfield Box 02116... 617 267-5222
17 566-1282	Jacques 50 Decatur Cir 02129... 617 241-0152	617 436-6906	Fred 45 Waverford Ave 02134... 617 436-6906	617 527-0480	Nick & Debbie 104 Harriet Rd Newton 02459... 617 527-0480
17 364-5188	Lucille 134 Harvard Cam 02139... 617 491-5621 M 95 Stone Hill 02115... 617 323-9713	617 825-0322	Fred 50 Hinchey Rd Mt 02136... 617 698-1343	617 498-0713	Nicole 156 Harriet Rd Newton 02459... 617 498-0713
361-0380	Carte Nicholas 115 South Hill 02115... 617 576-1061	617 522-2215	G 27 Franklin Ave Son 02146... 617 623-7121	617 822-1203	Shurman 38 Chickadee Dr 02135... 617 822-1203
17 628-8248	Carte Nicholas 18 Apollon Boston 02114... 617 695-6996	617 327-1105	Gayle 25 Promenade Dr 02134... 617 327-1105	617 427-0170	Thurman P L 48 Hattings Box 02114... 617 427-0170
17 445-5116	Cartes O 4 Willow Hill 02118... 617 338-8219	617 456-1689	Geo S 115 Wynn Hill Rd Apt 02134... 617 367-9548	617 983-5692	Paul & Constance 114 Anson Ave Wm 02138... 617 325-2036
17 627-2982	Carton Thos J Sr & Claire 1 Poydian Rd Mt 02186... 617 698-6163	617 325-5465	George 125 Nashua Box 02114... 617 367-9548	617 268-4546	Paul M 201 W South St Wm 02137... 617 268-4546
17 627-5712	Thomas & Kathleen 50 Thompson Ln Mt 02186... 617 696-6919	617 542-7067	Carter Harry E 182 Spring Hill W Wm 02137... 617 325-5465	617 787-2115	Prudence 40 Franklin Wadsworth 02179... 617 393-3782
17 569-2698	Carter A Box 02111... 617 327-2257	617 876-2750	Carter Hide Co Inc 145 Summer Box 02110... 617 542-7067	617 926-7063	Prudence 40 Franklin Wadsworth 02179... 617 926-7063
17 667-5190	A 31 Bethune Wy Roxbury 02119... 617 442-1219	617 354-2680	Carter Henry 41 Newry Cam 02186... 617 876-2750	617 541-2843	Reginald 40 Francis Wadsworth 02179... 617 541-2843
17 569-1417	A M 250 Woodside Ave Box 02115... 617 266-7153	617 232-7990	Horace 342 Millard Ave Roxbury 02119... 617 442-5307	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 338-9110	Adams 381 Courts St Mt 02186... 617 698-9074	617 323-5474	Howard Jr 20 Nebra Drive Mt 02119... 617 445-5552	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 825-9158	Alice 108 Elmwood Box 02215... 617 425-0193	617 739-9483	J 25 Chatham Box 02146... 617 232-7990	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 296-1593	Alice 45 Market Cambridge 02139... 617 945-2711	617 323-5474	J 518 Harvard Box 02146... 617 739-9483	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 670-2078	Andrew F 22 Visual Art Soc 02143... 617 625-7623	617 739-1022	J 775 Wey Plow Way Roxbury 02132... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 623-9001	Carte Nicholas 115 South Hill 02115... 617 576-1061	617 464-1040	Carter J 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 296-4275	Carter Avelth 9 Pleasant Hill 02119... 617 442-0780	617 464-1040	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 542-1521	Carter Becky Jo 02114... 617 634-0051	617 442-1775	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 364-5232	Carter Bernard J 711 State St E 02115... 617 567-3430	617 492-1214	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 541-5649	Carter Bishah N 02134... 617 298-8713	617 998-8435	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 739-2662	Carter Blake 26 W Irving Box 02118... 617 367-9911	617 926-5999	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 679-0030	Carter 200 Commonwealth Ave 02139... 617 782-2118	617 987-2163	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 436-1513	Carter 359 Harvard Cam 02139... 617 569-1545	617 422-4334	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
17 569-4119	Carter 359 Harvard Cam 02139... 617 569-1545	617 282-1235	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
800 569-8782	Carter 359 Harvard Cam 02139... 617 569-1545	617 285-9456	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843
	Carter 359 Harvard Cam 02139... 617 569-1545	617 282-1235	Carter J M 181 Rowley Box 02115... 617 323-5474	617 720-3765	Renee & Andrew 306 Braintree Dorchester 02122... 617 541-2843



↪ We reduce a (conceptual) geography of clusterings

A New Strategy

A New Strategy

- 1 Code text as numbers (in one or more of several ways)

A New Strategy

- ① **Code text as numbers** (in one *or more* of several ways)
- ② **Apply all existing clustering methods** (that have been used by at least one person other than the author) to the data — each representing different substantive assumptions (<15 mins)

A New Strategy

- 1 **Code text as numbers** (in one *or more* of several ways)
- 2 **Apply all existing clustering methods** (that have been used by at least one person other than the author) to the data — each representing different substantive assumptions (<15 mins)
- 3 Develop an **application-independent distance** metric between clusterings

A New Strategy

- ① **Code text as numbers** (in one *or more* of several ways)
- ② **Apply all existing clustering methods** (that have been used by at least one person other than the author) to the data — each representing different substantive assumptions (<15 mins)
- ③ Develop an **application-independent distance** metric between clusterings
- ④ Create a **metric space of clusterings**, and a 2D projection

A New Strategy

- 1 **Code text as numbers** (in one *or more* of several ways)
- 2 **Apply all existing clustering methods** (that have been used by at least one person other than the author) to the data — each representing different substantive assumptions (<15 mins)
- 3 Develop an **application-independent distance** metric between clusterings
- 4 Create a **metric space of clusterings**, and a 2D projection
- 5 Introduce the **local cluster ensemble** to summarize any point, including points with no existing clustering

A New Strategy

- 1 **Code text as numbers** (in one or more of several ways)
- 2 **Apply all existing clustering methods** (that have been used by at least one person other than the author) to the data — each representing different substantive assumptions (<15 mins)
- 3 Develop an **application-independent distance** metric between clusterings
- 4 Create a **metric space of clusterings**, and a 2D projection
- 5 Introduce the **local cluster ensemble** to summarize any point, including points with no existing clustering
- 6 Propose a new **animated visualization**: use the local cluster ensemble to explore the space of clusterings (smoothly morphing from one into others)

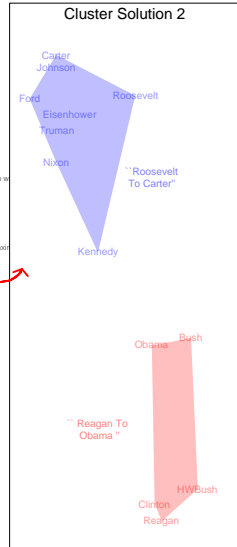
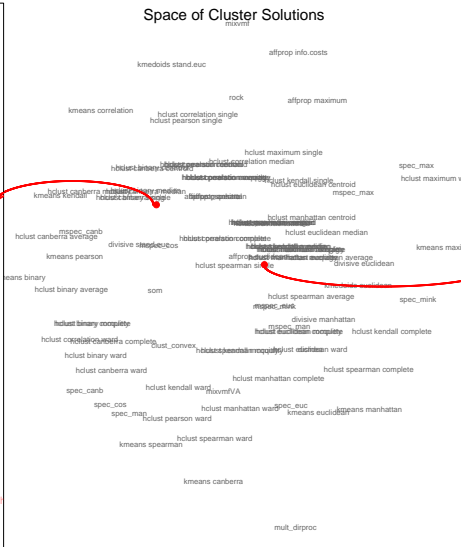
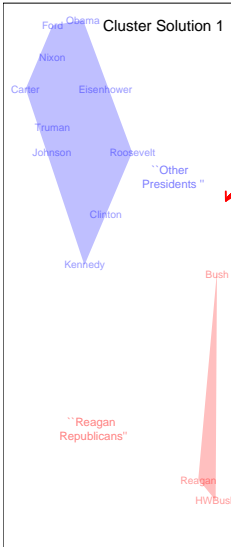
A New Strategy

- 1 **Code text as numbers** (in one or more of several ways)
- 2 **Apply all existing clustering methods** (that have been used by at least one person other than the author) to the data — each representing different substantive assumptions (<15 mins)
- 3 Develop an **application-independent distance** metric between clusterings
- 4 Create a **metric space of clusterings**, and a 2D projection
- 5 Introduce the **local cluster ensemble** to summarize any point, including points with no existing clustering
- 6 Propose a new **animated visualization**: use the local cluster ensemble to explore the space of clusterings (smoothly morphing from one into others)

↪ **meaning revealed through a *geography* of clusterings**

Many Thousands of Clusterings, Sorted & Organized

You choose one (or more), based on insight, discovery, useful information, . . .



Application-Independent Distance Metric: Axioms

Application-Independent Distance Metric: Axioms

- 1 Clusterings with more **pairwise document agreements** are closer (we prove: pairwise agreements encompass triples, quadruples, etc.)

Application-Independent Distance Metric: Axioms

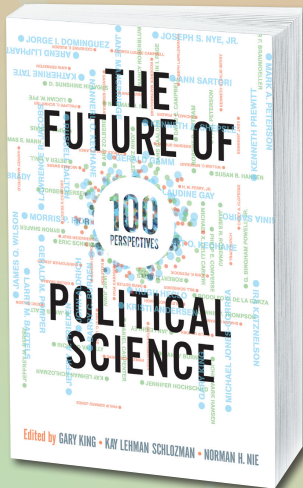
- 1 Clusterings with more **pairwise document agreements** are closer (we prove: pairwise agreements encompass triples, quadruples, etc.)
- 2 **Invariance**: Distance is invariant to the number of documents (for any fixed number of clusters)

Application-Independent Distance Metric: Axioms

- 1 Clusterings with more **pairwise document agreements** are closer (we prove: pairwise agreements encompass triples, quadruples, etc.)
- 2 **Invariance**: Distance is invariant to the number of documents (for any fixed number of clusters)
- 3 **Scale**: the maximum distance is set to $\log(\text{num clusters})$

Application-Independent Distance Metric: Axioms

- 1 Clusterings with more **pairwise document agreements** are closer (we prove: pairwise agreements encompass triples, quadruples, etc.)
 - 2 **Invariance**: Distance is invariant to the number of documents (for any fixed number of clusters)
 - 3 **Scale**: the maximum distance is set to $\log(\text{num clusters})$
- ↪ **Only one measure satisfies all three** (the “variation of information”)



Available March 2009: 304pp
Pb: 978-0-415-99701-0: **\$24.95**
www.routledge.com/politics

THE FUTURE OF POLITICAL SCIENCE

100 Perspectives

Edited by Gary King, Harvard University, Kay Lehman Schlozman, Boston College
and Norman H. Nie, Stanford University

“The list of authors in *The Future of Political Science* is a ‘who’s who’ of political science. As I was reading it, I came to think of it as a platter of tasty hors d’oeuvres. It hooked me thoroughly.”

—Peter Kingstone, University of Connecticut

“In this one-of-a-kind collection, an eclectic set of contributors offer short but forceful forecasts about the future of the discipline. The resulting assortment is captivating, consistently thought-provoking, often intriguing, and sure to spur discussion and debate.”

—Wendy K. Tam Cho, University of Illinois at Urbana-Champaign

“King, Schlozman, and Nie have created a visionary and stimulating volume. The organization of the essays strikes me as nothing less than brilliant. . . It is truly a joy to read.”

—Lawrence C. Dodd, Manning J. Dauer Eminent Scholar in Political Science,
University of Florida

 **Routledge**
Taylor & Francis Group
an **informa** business

Evaluators' Rate Machine Choices Better Than Their Own

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Pairs from	Overall Mean	Evaluator 1	Evaluator 2
Random Selection	1.38	1.16	1.60

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Pairs from	Overall Mean	Evaluator 1	Evaluator 2
Random Selection	1.38	1.16	1.60

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Pairs from	Overall Mean	Evaluator 1	Evaluator 2
Random Selection	1.38	1.16	1.60
Hand-Coded Clusters	1.58	1.48	1.68

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Pairs from	Overall Mean	Evaluator 1	Evaluator 2
Random Selection	1.38	1.16	1.60
Hand-Coded Clusters	1.58	1.48	1.68
Hand-Coding	2.06	1.88	2.24

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Pairs from	Overall Mean	Evaluator 1	Evaluator 2
Random Selection	1.38	1.16	1.60
Hand-Coded Clusters	1.58	1.48	1.68
Hand-Coding	2.06	1.88	2.24
Machine	2.24	2.08	2.40

Evaluators' Rate Machine Choices Better Than Their Own

- Scale: (1) unrelated, (2) loosely related, or (3) closely related
- Table reports: mean(scale)

Pairs from	Overall Mean	Evaluator 1	Evaluator 2
Random Selection	1.38	1.16	1.60
Hand-Coded Clusters	1.58	1.48	1.68
Hand-Coding	2.06	1.88	2.24
Machine	2.24	2.08	2.40

p.s. The hand-coders did the evaluation!

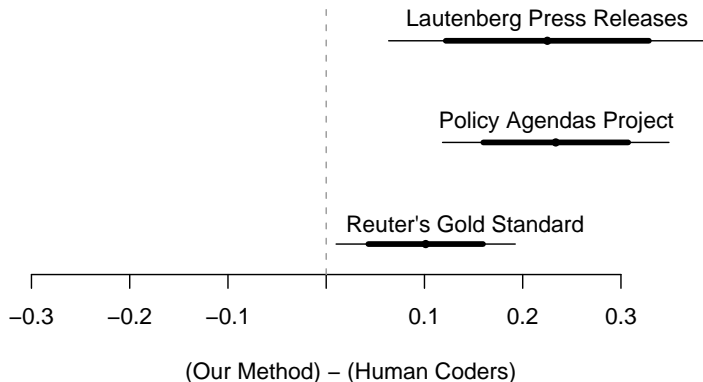
Cluster Quality Experiments

Cluster Quality Experiments

Scale: $\text{mean}(\text{within clusters}) - \text{mean}(\text{between clusters})$

Cluster Quality Experiments

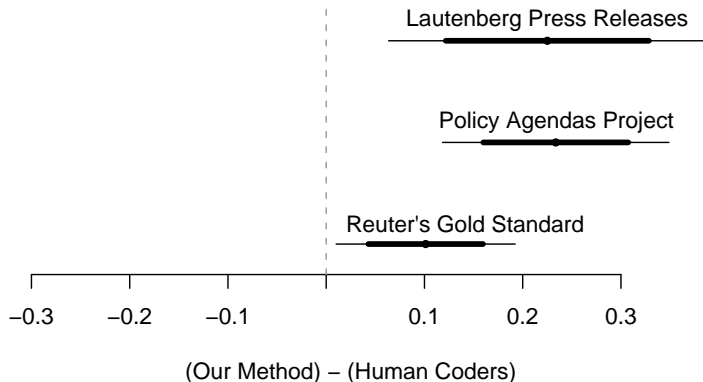
Scale: $\text{mean}(\text{within clusters}) - \text{mean}(\text{between clusters})$



Lautenberg: 200 Senate Press Releases (appropriations, economy, education, tax, veterans, ...)

Cluster Quality Experiments

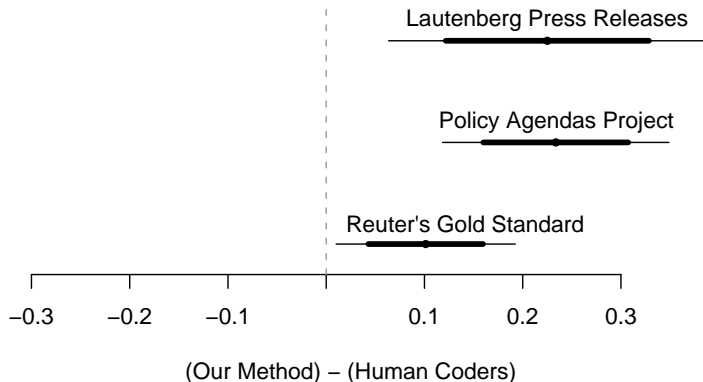
Scale: $\text{mean}(\text{within clusters}) - \text{mean}(\text{between clusters})$



Policy Agendas: 213 quasi-sentences from Bush's State of the Union (agriculture, banking & commerce, civil rights/liberties, defense, ...)

Cluster Quality Experiments

Scale: $\text{mean}(\text{within clusters}) - \text{mean}(\text{between clusters})$



Reuter's: financial news (trade, earnings, copper, gold, coffee, ...); "gold standard" for supervised learning studies

What do Members of Congress Do?

What do Members of Congress Do?

Substantive example of a finding, using our approach

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising
 - 2 Credit Claiming

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising
 - 2 Credit Claiming
 - 3 Position Taking

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising
 - 2 Credit Claiming
 - 3 Position Taking
- We find one more: **Partisan Taunting**

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising
 - 2 Credit Claiming
 - 3 Position Taking
- We find one more: **Partisan Taunting**
 - “Senator Lautenberg Blasts Republicans as ‘Chicken Hawks’ ”
[Government Oversight]

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - ① Advertising
 - ② Credit Claiming
 - ③ Position Taking
- We find one more: **Partisan Taunting**
 - “Senator Lautenberg Blasts Republicans as ‘Chicken Hawks’ ”
[Government Oversight]
 - “The scopes trial took place in 1925. Sadly, President Bush’s veto today shows that we haven’t progressed much since then.” [Healthcare]

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising
 - 2 Credit Claiming
 - 3 Position Taking
- We find one more: **Partisan Taunting**
 - “Senator Lautenberg Blasts Republicans as ‘Chicken Hawks’ ” [Government Oversight]
 - “The scopes trial took place in 1925. Sadly, President Bush’s veto today shows that we haven’t progressed much since then.” [Healthcare]
 - “John Kerry had enough conviction to sign up for the military during wartime, unlike the Vice President, who had a deep conviction to avoid military service” [Government Oversight]

What do Members of Congress Do?

Substantive example of a finding, using our approach

- David Mayhew's (1974) famous typology
 - 1 Advertising
 - 2 Credit Claiming
 - 3 Position Taking
- We find one more: **Partisan Taunting**
 - "Senator Lautenberg Blasts Republicans as 'Chicken Hawks' " [Government Oversight]
 - "The scopes trial took place in 1925. Sadly, President Bush's veto today shows that we haven't progressed much since then." [Healthcare]
 - "John Kerry had enough conviction to sign up for the military during wartime, unlike the Vice President, who had a deep conviction to avoid military service" [Government Oversight]
 - ↪ **Is this what it means to be a member of a political party?**

More Informative Discoveries

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work
- For each: created 2 clusterings from each of 3 methods, including ours

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work
- For each: created 2 clusterings from each of 3 methods, including ours
- Created info packet on each clustering (for each cluster: exemplar document, automated content summary)

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work
- For each: created 2 clusterings from each of 3 methods, including ours
- Created info packet on each clustering (for each cluster: exemplar document, automated content summary)
- Asked for $\binom{6}{2}=15$ pairwise comparisons

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work
- For each: created 2 clusterings from each of 3 methods, including ours
- Created info packet on each clustering (for each cluster: exemplar document, automated content summary)
- Asked for $\binom{6}{2}=15$ pairwise comparisons
- Both cases a Condorcet winner:

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work
- For each: created 2 clusterings from each of 3 methods, including ours
- Created info packet on each clustering (for each cluster: exemplar document, automated content summary)
- Asked for $\binom{6}{2}=15$ pairwise comparisons
- Both cases a Condorcet winner:

“Immigration”:

Our Method 1 → vMF 1 → vMF 2 → Our Method 2 → K-Means 1 → K-Means 2

More Informative Discoveries

- Found 2 scholars analyzing lots of textual data for their work
- For each: created 2 clusterings from each of 3 methods, including ours
- Created info packet on each clustering (for each cluster: exemplar document, automated content summary)
- Asked for $\binom{6}{2}=15$ pairwise comparisons
- Both cases a Condorcet winner:

“Immigration”:

Our Method 1 → vMF 1 → vMF 2 → Our Method 2 → K-Means 1 → K-Means 2

“Genetic testing”:

Our Method 1 → {Our Method 2, K-Means 1, K-means 2} → Dir Proc. 1 → Dir Proc. 2

Intended contributions

- An encompassing cluster analytic approach for discovery

Intended contributions

- An encompassing cluster analytic approach for discovery
- A new approach to evaluating results in unsupervised learning

Intended contributions

- An encompassing cluster analytic approach for discovery
- A new approach to evaluating results in unsupervised learning
- Especially useful for the ongoing spectacular increase in the production and availability of unstructured text

Intended contributions

- An encompassing cluster analytic approach for discovery
- A new approach to evaluating results in unsupervised learning
- Especially useful for the ongoing spectacular increase in the production and availability of unstructured text
- **Multiple approaches: Integrated, not separate**

For more information:

<http://GKing.Harvard.edu>