data-ppf.github.io mar 31, 2020

lecture 10 of 14: data science, 1962-2017

chris wiggins + matt jones, Columbia

student reactions

```
...
16 should
...
13 power
21 ethical/ethics
11 assumptions
1 STEM
```

student reactions: names of things

I found it interesting to read about how controversial the branding of "data science" can be. I suppose that I have never really thought to question the field, and I have largely been too oblivious to have noticed the significance of its emergence. It just seemed to have just shown up, and now it's a hot, in-demand career field. However, after reading this week's texts, I am starting to wonder if its name is misleading—is it really a science?

After completing this week's reading I felt more confused about what data science is than ever before.

themes

- tensions

themes

- tensions
  - between academia + industry

themes

- tensions
  - between academia + industry
    - computers vs math

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)
  - 2000s: .ai, .vc

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)
  - 2000s: .ai, .vc
- sources: experts, evangelists, and expositors

themes

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)
  - 2000s: .ai, .vc
- sources: experts, evangelists, and expositors
- inconvenient truths this week: truth is negotiated

historical & social context

- industrial data powers

historical & social context

- industrial data powers
- academic mathematical statistics

historical & social context

- industrial data powers
- academic mathematical statistics
  - recall neyman/fisher 1955-1956

historical & social context

- industrial data powers
- academic mathematical statistics
  - recall neyman/fisher 1955-1956
- rise of data outside silos

historical & social context

- industrial data powers
- academic mathematical statistics
  - recall neyman/fisher 1955-1956
- rise of data outside silos
  - EPA, ETS, etc. . . ..

historical & social context

- industrial data powers
- academic mathematical statistics
  - recall neyman/fisher 1955-1956
- rise of data outside silos
  - EPA, ETS, etc. . . ..
- "data science" 1997, 2001, 2010, ...

historical & social context

- industrial data powers
- academic mathematical statistics
  - recall neyman/fisher 1955-1956
- rise of data outside silos
  - EPA, ETS, etc. . . ..
- "data science" 1997, 2001, 2010, ...
  - (itself echo of FoDA '62, GLS '93, two cultures '01. . . .)

contemporary context/modern day relevance

- data science everywhere as Donoho says

contemporary context/modern day relevance

- data science everywhere as Donoho says
  - in industry, as job title

contemporary context/modern day relevance

- data science everywhere as Donoho says
  - in industry, as job title
  - in academia: institute $\rightarrow$ MS program $\rightarrow$ major $\rightarrow$ PhD program. . . .

what are the new capabilities this week?

- ML as “technology”

what are the new capabilities this week?

- ML as "technology"
- making sense of data *absent* of type I, type II, or even models

power + truth

- in industry: resources

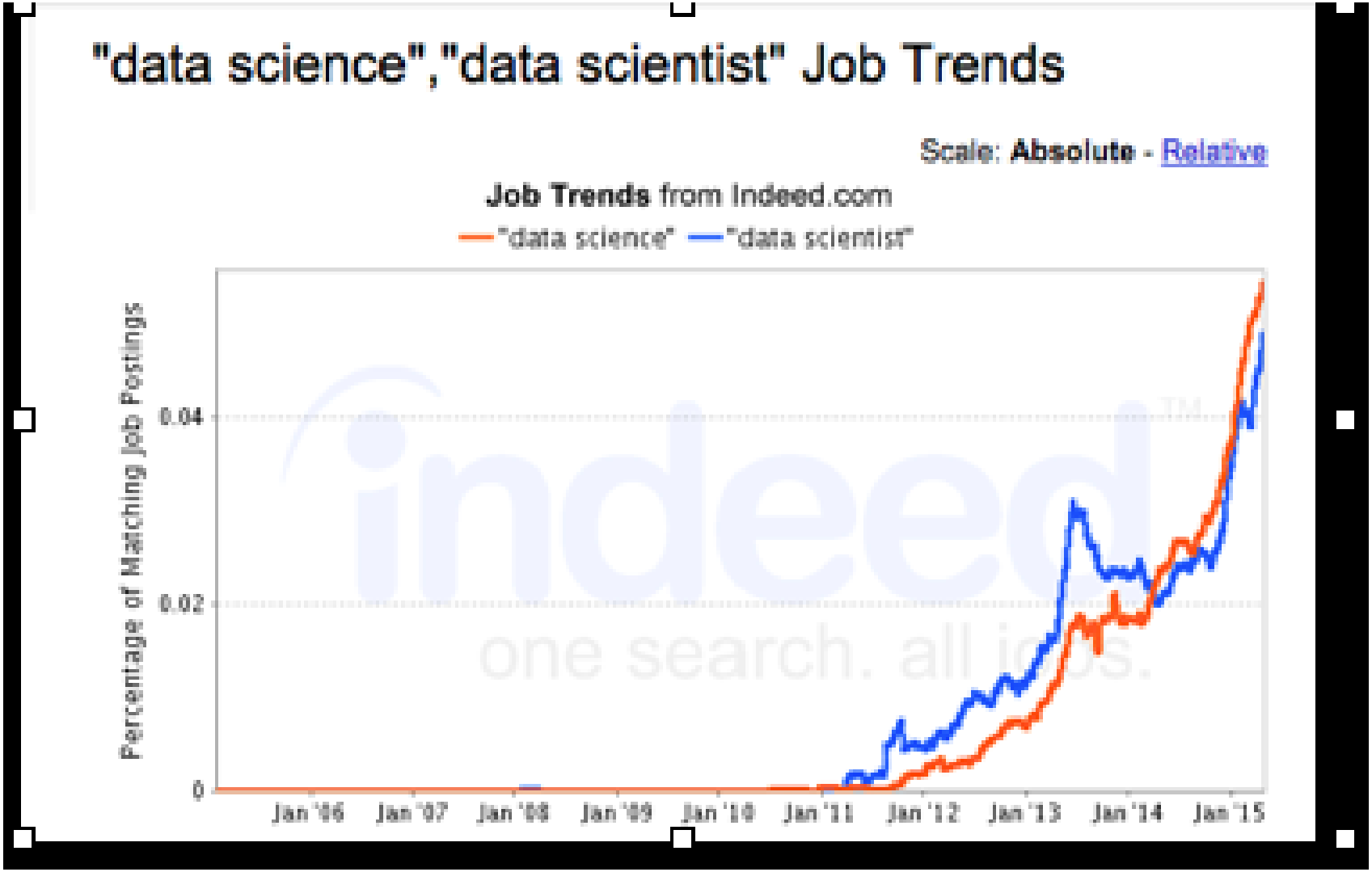


Figure 1: jobs

power + truth

- in industry: resources

2. *Physical, Mathematical, and Engineering Sciences.* For this report (a) *physical sciences* are those sciences concerned primarily with the understanding of the natural phenomena associated with nonliving things; (b) *mathematical sciences* are those sciences which employ logical reasoning with the aid of symbols and which are concerned with the development of methods of operations employing such symbols, including mathematics, pure and applied; astronomy, theoretical mechanics, statistics, logistic research, and computer research exclusive of engineering; (c) *engineering sciences* are those sciences which are concerned with studies directed toward making specific scientific principles usable in engineering practice.

Figure 2: From NSF 1952

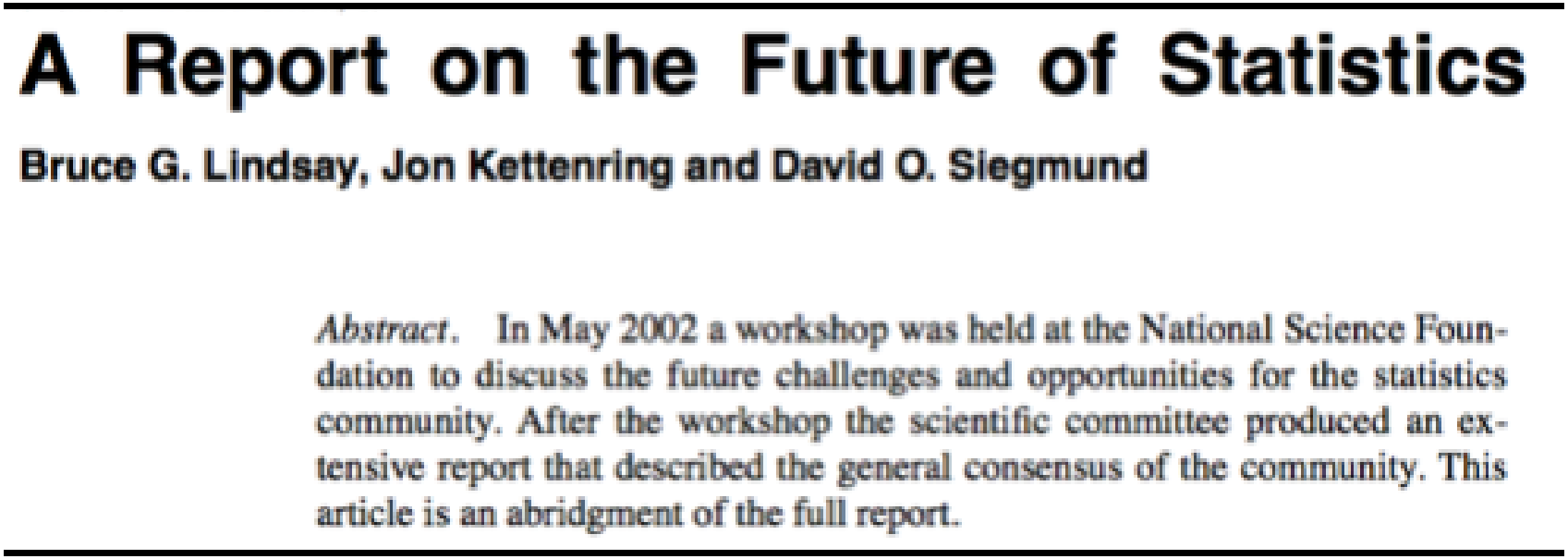
power + truth

- in industry: resources
- in academia: resources

2. *Physical, Mathematical, and Engineering Sciences.* For this report (a) *physical sciences* are those sciences concerned primarily with the understanding of the natural phenomena associated with nonliving things; (b) *mathematical sciences* are those sciences which employ logical reasoning with the aid of symbols and which are concerned with the development of methods of operations employing such symbols, including mathematics, pure and applied; astronomy, theoretical mechanics, statistics, logistic research, and computer research exclusive of engineering; (c) *engineering sciences* are those sciences which are concerned with studies directed toward making specific scientific principles usable in engineering practice.

Figure 2: From NSF 1952

NSF on mathematical statistics



A Report on the Future of Statistics

Bruce G. Lindsay, Jon Kettenring and David O. Siegmund

Abstract. In May 2002 a workshop was held at the National Science Foundation to discuss the future challenges and opportunities for the statistics community. After the workshop the scientific committee produced an extensive report that described the general consensus of the community. This article is an abridgment of the full report.

Figure 3: From NSF 2004

NSF on mathematical statistics

than outward. At a time when statistics is beginning to recover from its "overmathematization" in the post World War II years and engage in significant applications in many areas, the report is a step into the past and not into the future.

Figure 4: Breiman on NSF 2004

Data Science

How get here? What drive process?

- ideas

Data Science

How get here? What drive process?

- ideas
- tech

Data Science

How get here? What drive process?

- ideas
- tech
- power

Data Science

How get here? What drive process?

- ideas
- tech
- power
- cash

Ecologies of instrumental prediction of high-dimension, messy data

- USA

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR
- Japan

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR
- Japan
- France

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR
- Japan
- France
- Netherlands . . .

Not the high cold war fields we thought were crucial

- Operations research

Not the high cold war fields we thought were crucial

- Operations research
- Game Theory

Not the high cold war fields we thought were crucial

- Operations research
- Game Theory
- Formal AI

Not the high cold war fields we thought were crucial

- Operations research
- Game Theory
- Formal AI
- Decision Theory—Mathematical Statistics

Low road of instrumental computational stats

- Vast archives of data (cryptological then commercial in first instance)

Plus

Low road of instrumental computational stats

- Vast archives of data (cryptological then commercial in first instance)

Plus

- Highly instrumental statistical approach

War time data work

Thanks to "war problems," "it was natural to regard statistics as something that had the purpose of being used on data—maybe not directly, but at most at some remove. Now, I can't believe that other people who had practical experience failed to have this view, but they certainly—I would say—failed to advertise it."

- John Tukey interview, 85.

World War 2: statistical cryptography + info processing

“Pattern Recognition”

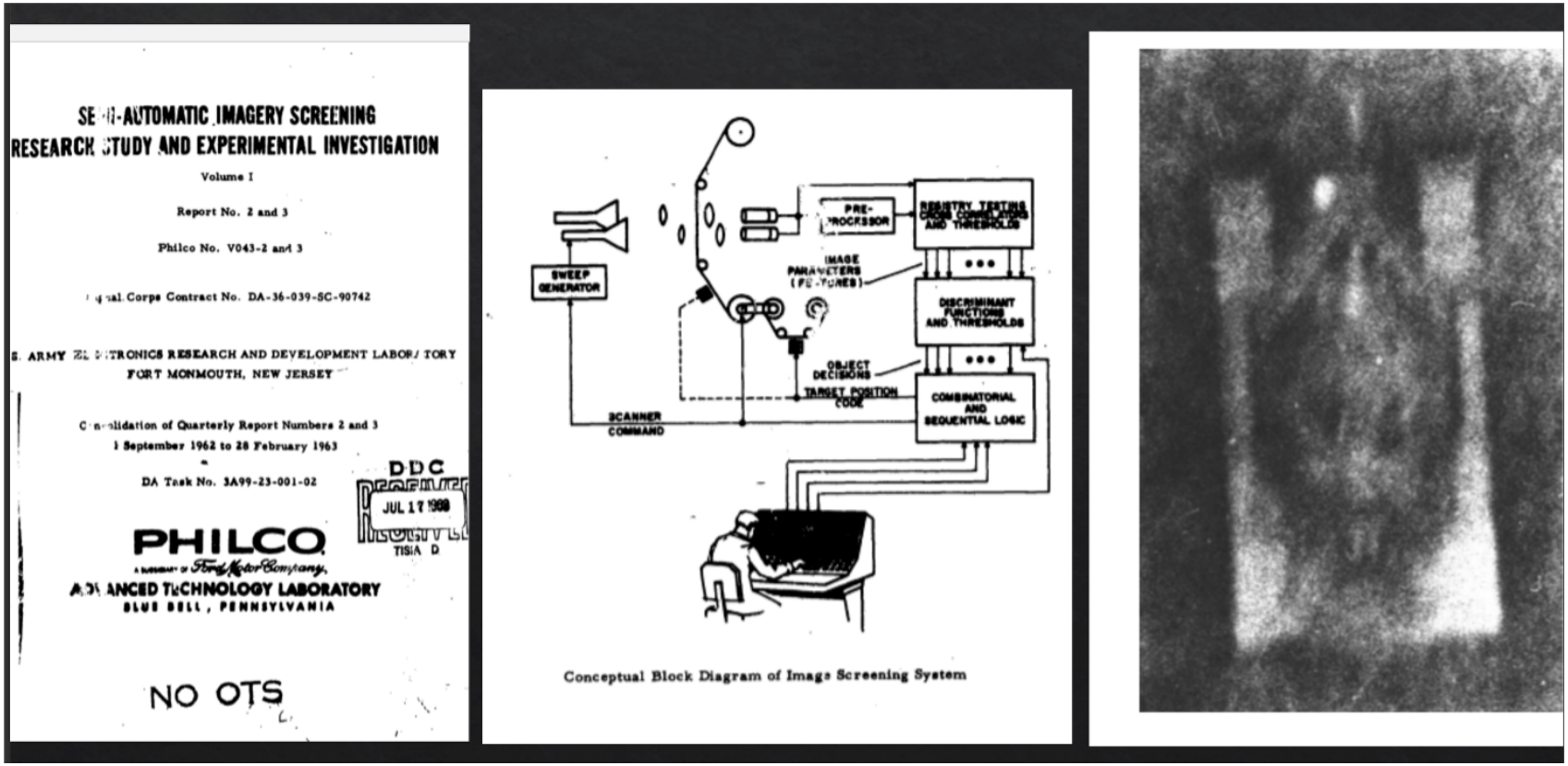
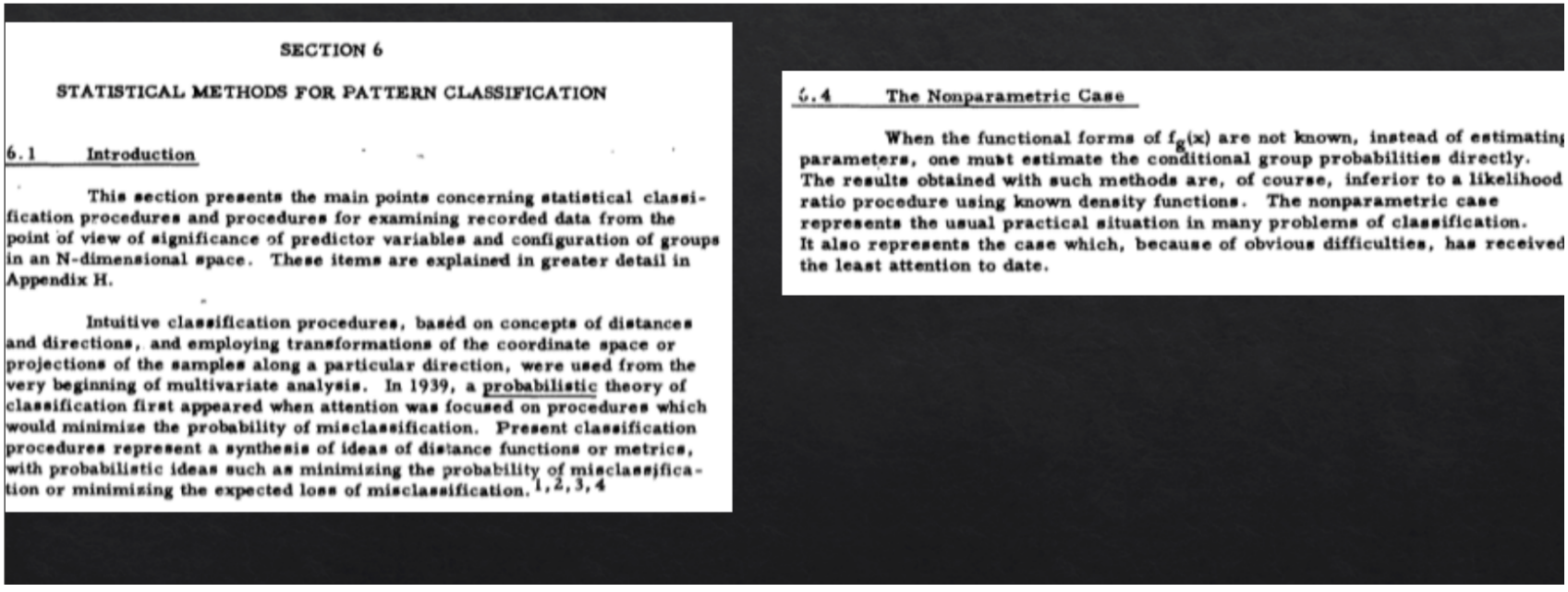


Figure 5: pattern recognition

“Pattern Recognition”



SECTION 6

STATISTICAL METHODS FOR PATTERN CLASSIFICATION

6.1 Introduction

This section presents the main points concerning statistical classification procedures and procedures for examining recorded data from the point of view of significance of predictor variables and configuration of groups in an N-dimensional space. These items are explained in greater detail in Appendix H.

Intuitive classification procedures, based on concepts of distances and directions, and employing transformations of the coordinate space or projections of the samples along a particular direction, were used from the very beginning of multivariate analysis. In 1939, a probabilistic theory of classification first appeared when attention was focused on procedures which would minimize the probability of misclassification. Present classification procedures represent a synthesis of ideas of distance functions or metrics, with probabilistic ideas such as minimizing the probability of misclassification or minimizing the expected loss of misclassification.[1,2,3,4]

6.4 The Nonparametric Case

When the functional forms of $f_g(x)$ are not known, instead of estimating parameters, one must estimate the conditional group probabilities directly. The results obtained with such methods are, of course, inferior to a likelihood ratio procedure using known density functions. The nonparametric case represents the usual practical situation in many problems of classification. It also represents the case which, because of obvious difficulties, has received the least attention to date.

Figure 6: pattern recognition

Cash/Cachet

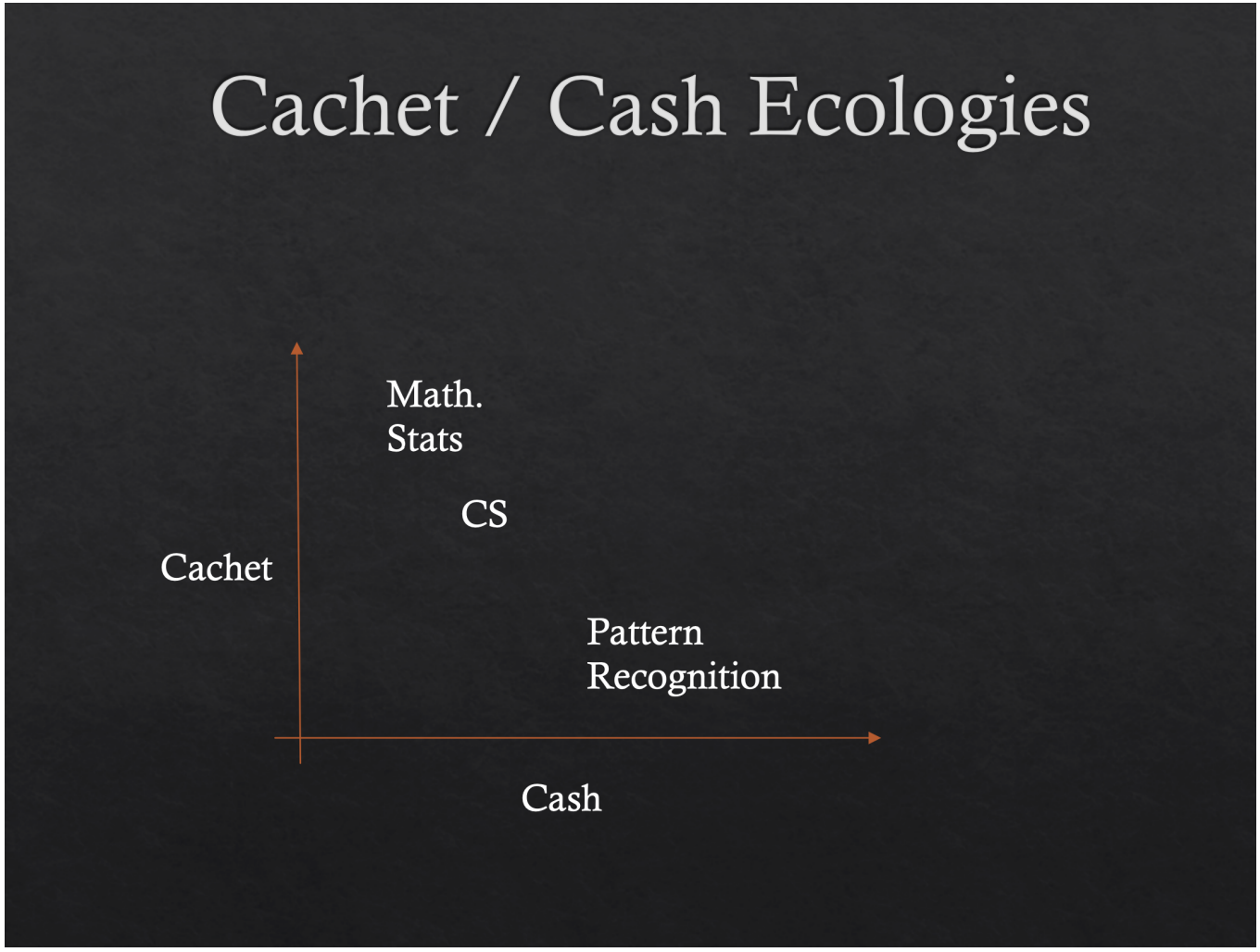


Figure 7: cash-cachet

Ecologies of instrumental prediction of high-dimension, messy data

- USA

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR
- Japan

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR
- Japan
- France

Ecologies of instrumental prediction of high-dimension, messy data

- USA
- USSR
- Japan
- France
- Netherlands . . . *Lots* of stories yet tell

Renegade (heretical?) stats

- Leo Breiman on academic mathematical statistics: "Alice in Wonderland. That is, I knew what was going on out in industry and government in terms of uses of statistics, but what was going on in academic research seemed light years away. It was proceeding as though it were some branch of abstract mathematics." (in Olsen, 196)
- Epistemic virtue: Predictive accuracy » causal model

Example: Bell Labs

Reminder: Bell<->Bletchley

Shannon: mathematical theory of crypto/communications

[24] (P7)

E-1781-B (4-44)

COVER SHEET FOR TECHNICAL MEMORANDA

RESEARCH DEPARTMENT

SUBJECT: A Mathematical Theory of Cryptography - Case 20878 (4)

ROUTING:
1 – HWB-HF-Case Files
2 – CASE FILES
3 – J. W. McRae
4 – L. Espenschied
5 – H. S. Black
6 – F. B. Llewellyn
7 – H. Nyquist
8 – B. M. Oliver
9 – R. K. Potter
10 – C. B. H. Feldman
11 - R. C. Mathes
12 - R. V. L. Hartley
13 - J. R. Pierce
14 - H. W. Bode
15 - R. L. Dietzold
16 - L. A. MacCall
17 - W. A. Shewhart
18 - S. A. Schelkunoff
19 - C. E. Shannon
20 - Dept. 1000 Files

MM– 45-110-92
DATE September 1, 1945
AUTHOR C. E. Shannon
INDEX NO. P 0.4

SECRET

~~ABSTRACT~~

DOWNGRADED AT 3 YEAR INTERVALS
DECLASSIFIED AFTER 12 YEARS
DOD DIR 5200.10

ABSTRACT

A mathematical theory of secrecy systems is developed. Three main problems are considered. (1) A logical formulation of the problem and a study of the mathematical structure of secrecy systems. (2) The problem of "theoretical secrecy," i.e., can a system be solved given unlimited time and how much material must be intercepted to obtain a unique solution to cryptograms. A secrecy measure called the "equivocation" is defined and its properties developed. (3) The problem of "practical secrecy." How can systems be made difficult to solve, even though a solution is theoretically possible.

CONFIDENTIAL

Shannon: mathematical theory of communications

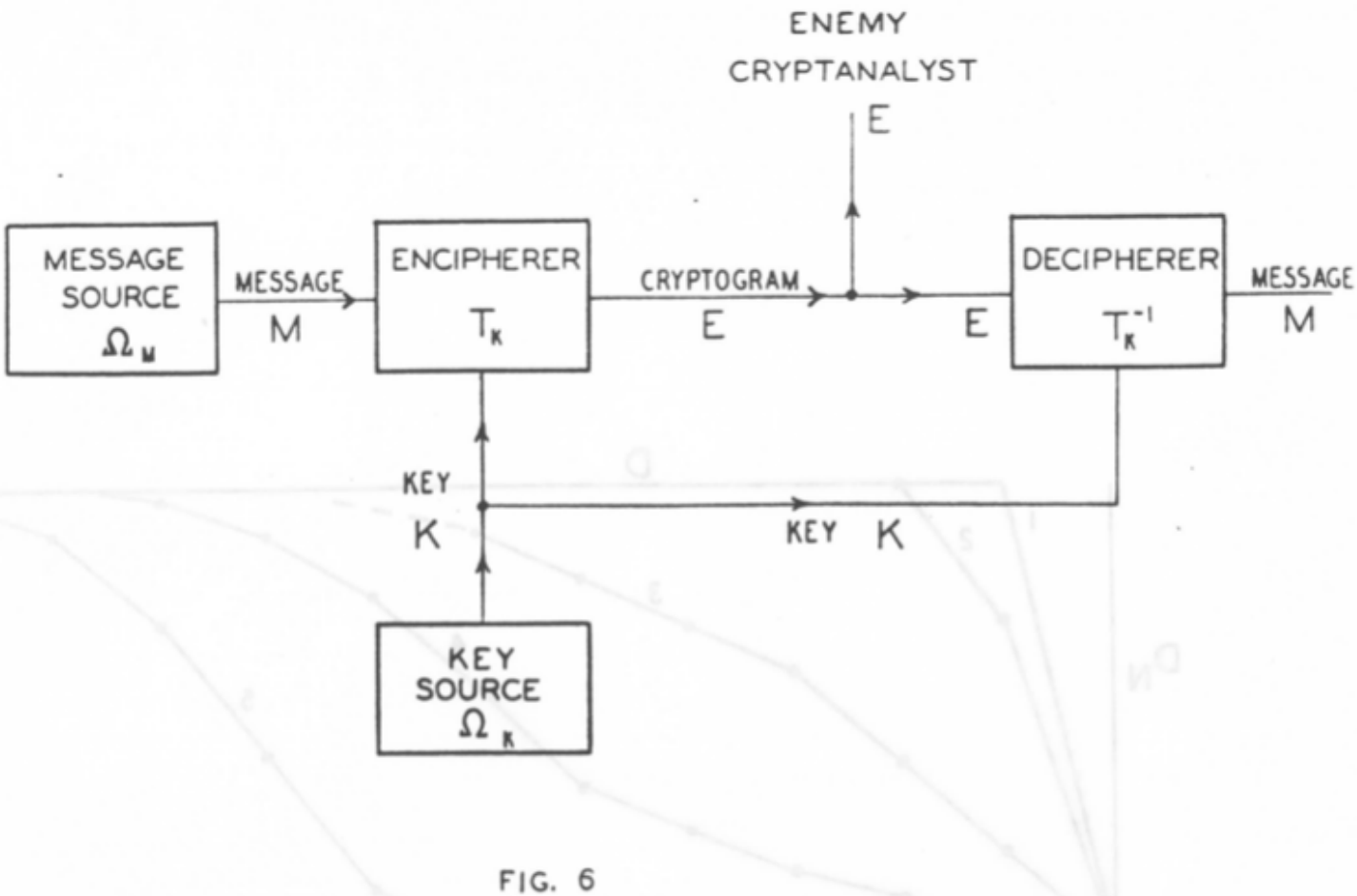


Figure 9: Shannon crypto diagram

Shannon: conversion to civilian

The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

A Mathematical Theory of Communication

By C. E. SHANNON

Introduction

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual

Shannon: from interception to communications

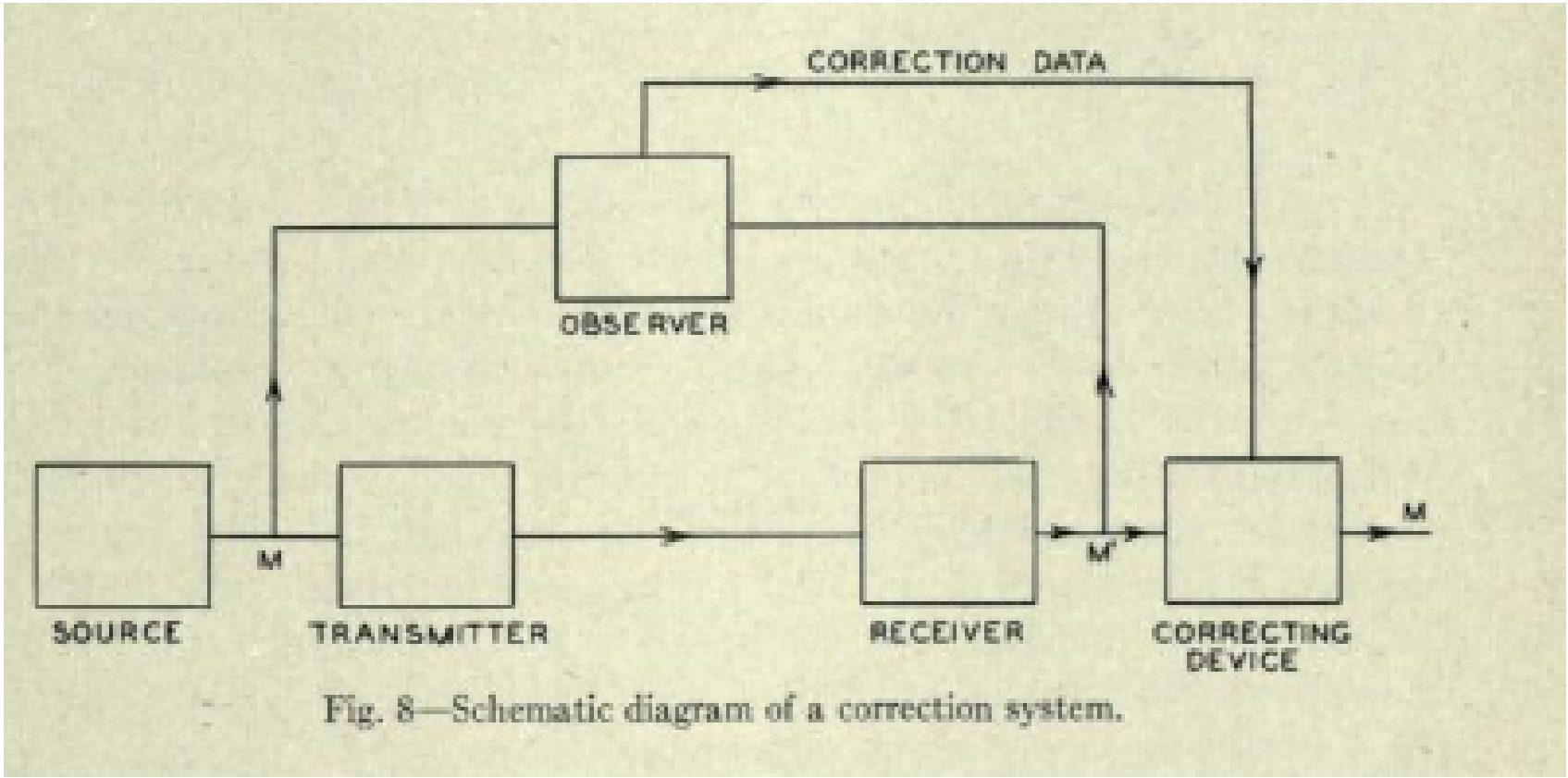


Fig. 8—Schematic diagram of a correction system.

Figure 11: Shannon communications diagram

Shannon->Tukey->Chambers-> Donoho and Yu

- Communications problems as incubator of technologies

it was quite clear that if you look at some of the work of Shannon in connection with cryptography, that must have stemmed from contacts that we had. We had very close contacts with the Bell Laboratories. They were very, let's say, willing to work along with us.

Shannon->Tukey->Chambers-> Donoho and Yu

- Communications problems as incubator of technologies
- Key example of US Cold War R&D

  it was quite clear that if you look at some of the work of Shannon in connection with cryptography, that must have stemmed from contacts that we had. We had very close contacts with the Bell Laboratories. They were very, let's say, willing to work along with us.

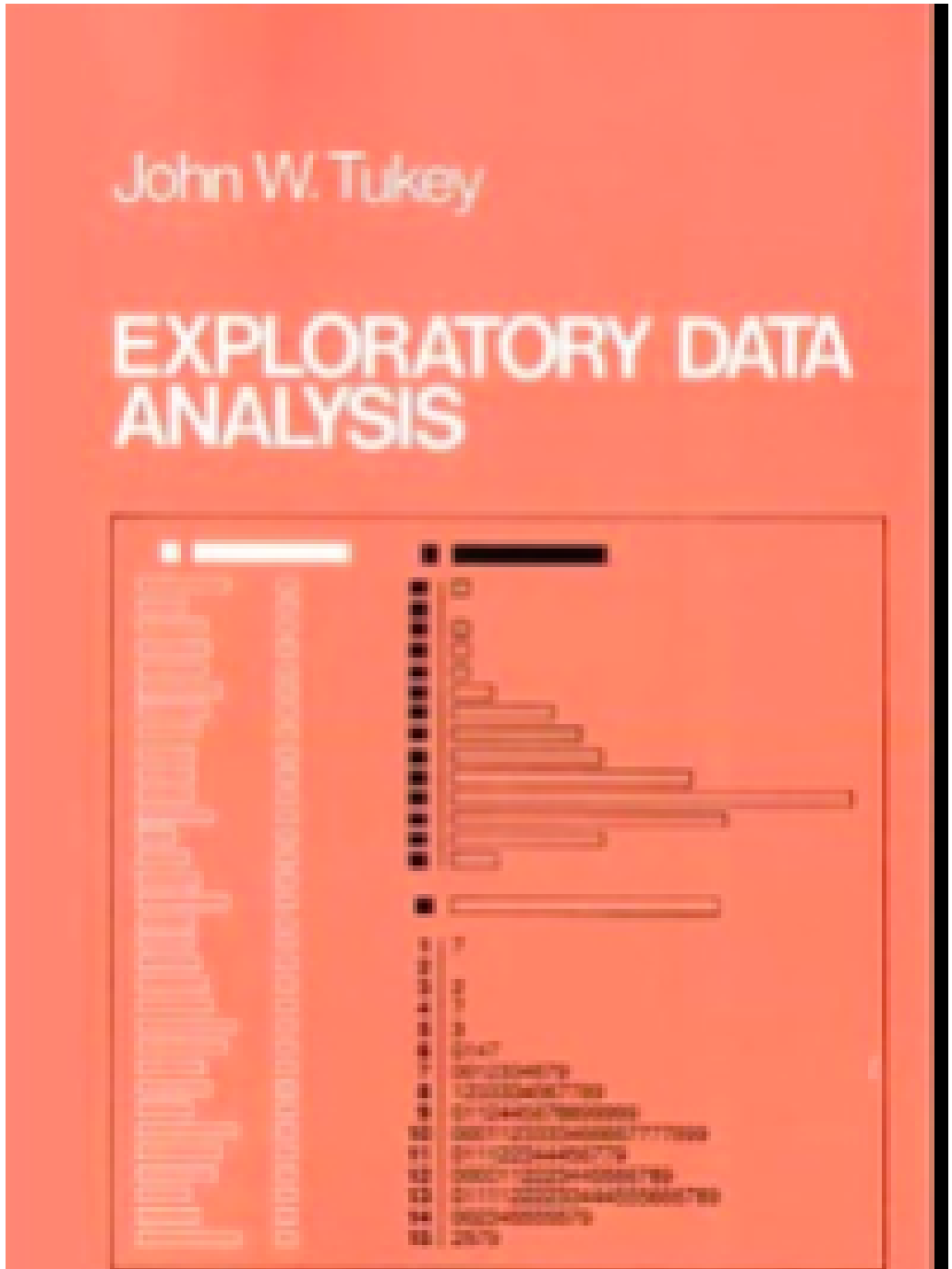
Shannon->Tukey->Chambers-> Donoho and Yu

- Communications problems as incubator of technologies
- Key example of US Cold War R&D

  it was quite clear that if you look at some of the work of Shannon in connection with cryptography, that must have stemmed from contacts that we had. We had very close contacts with the Bell Laboratories. They were very, let's say, willing to work along with us.

- S. Kullback 1982 NSA oral history interview, declassified 2015

Tukey: recall



Tukey: bio

- mathematician

Tukey: bio

- mathematician
- turned statistician

Tukey: bio

- mathematician
- turned statistician
- split career

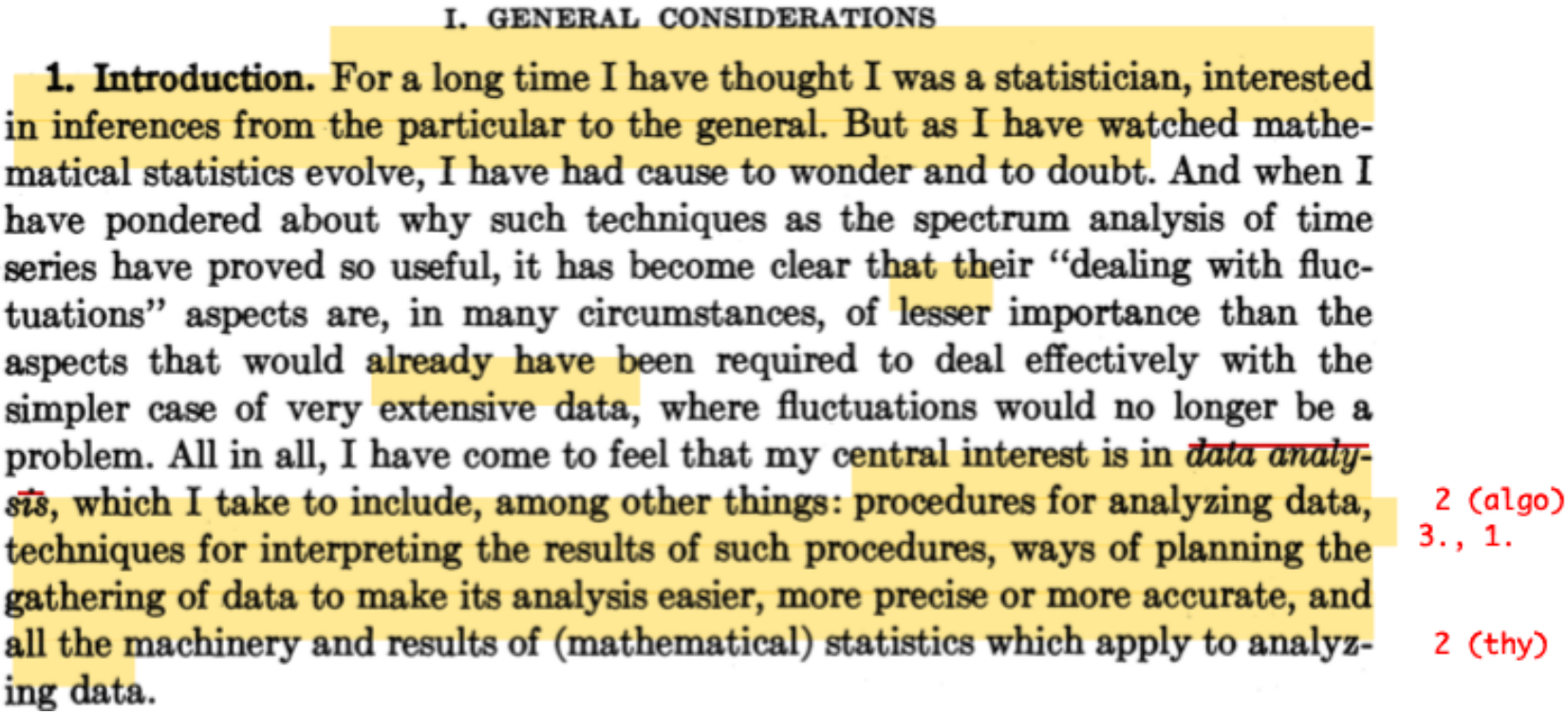
Tukey: bio

- mathematician
- turned statistician
- split career
  - consulting

Tukey: bio

- mathematician
- turned statistician
- split career
  - consulting
  - intelligence work

Tukey FoDA 62 opening quote on identity



I. GENERAL CONSIDERATIONS

1. Introduction. For a long time I have thought I was a statistician, interested in inferences from the particular to the general. But as I have watched mathematical statistics evolve, I have had cause to wonder and to doubt. And when I have pondered about why such techniques as the spectrum analysis of time series have proved so useful, it has become clear that their "dealing with fluctuations" aspects are, in many circumstances, of lesser importance than the aspects that would already have been required to deal effectively with the simpler case of very extensive data, where fluctuations would no longer be a problem. All in all, I have come to feel that my central interest is in *data analysis*, which I take to include, among other things: procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate, and all the machinery and results of (mathematical) statistics which apply to analyzing data.

2 (algo)
3., 1.
2 (thy)

Figure 13: opening

Tukey's call for data analysis

> *Data analysis, and the parts of statistics which adhere to it, must then take on the characteristics of a science rather than those of mathematics, specifically: (1) Data analysis must seek for scope and usefulness rather than security. (2) Data analysis must be willing to err moderately often in order that inadequate evidence shall more often suggest the right answer. (3) Data analysis must use mathematical argument and mathematical results as bases for judgment rather than as bases for proofs or stamps of validity.*

- John W. Tukey, "The Future of Data Analysis," Annals of Mathematical Statistics, no. 1 (Mar 1962): 1–67, https://doi.org/10.1214/aoms/1177704711. 6.

Tukey

- feet in academia but free from it in many respects

Tukey's FoDA

- opens with impostor line

Tukey's FoDA

- opens with impostor line
- attack on mathematization

Donoho Reading: 50 years since Tukey

- Traces history, not only JWT+LB

Donoho Reading: 50 years since Tukey

- Traces history, not only JWT+LB
- GLS'93

Donoho Reading: 50 years since Tukey

- Traces history, not only JWT+LB
- GLS'93
- Cleveland'01

Donoho Reading: 50 years since Tukey

- Traces history, not only JWT+LB
- GLS'93
- Cleveland'01
- interest: baptizing DS as Stats; cakeism

Cleveland's "Data Science: An action plan for..statistics" (2001)

Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics

William S. Cleveland
Statistics Research, Bell Labs
wsc@bell-labs.com

Abstract

An action plan to enlarge the technical areas of statistics focuses on the data analyst. The plan sets out six technical areas of work for a university department and advocates a specific allocation of resources devoted to research in each area and to courses in each area. The value of technical work is judged by the extent to which it benefits the data analyst, either directly or indirectly. The plan is also applicable to government research labs and corporate research organizations.

Figure 14: 2001

Cleveland's "Data Science: An action plan for..statistics" (2001)

The focus of the plan is the practicing data analyst.

One outcome of the plan is that computer science joins mathematics as an area of competency for the field of data science. This enlarges the intellectual foundations. It implies partnerships with computer scientists just as there are now partnerships with mathematicians.

The primary agents for change should be university departments themselves. But it is reasonable for departments to look both to university administrators and to funding agencies for resources to assist in bringing about the change.

Chambers Greater and Lesser Statistics 1993

Greater or Lesser Statistics: A Choice for Future Research

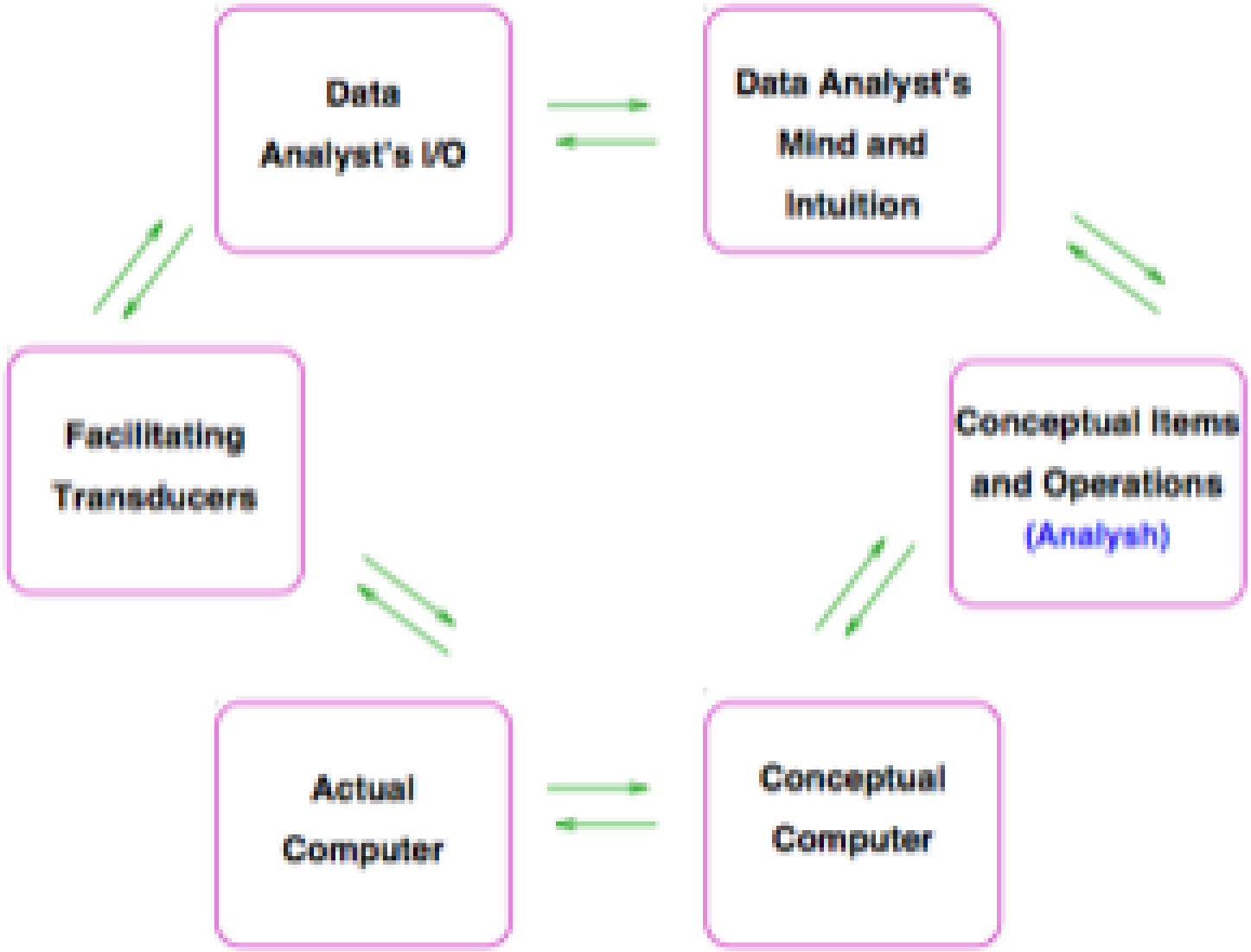
John M. Chambers
AT&T Bell Laboratories, Murray Hill, New Jersey

Abstract

The statistics profession faces a choice in its future research between continuing concentration on traditional topics, based largely on data analysis supported by mathematical statistics, and a broader viewpoint, based on an inclusive concept of learning from data. The latter course presents severe challenges as well as exciting opportunities. The former risks seeing statistics become increasingly marginal in activities to which it can make important contributions.

Figure 15: GLS

Chambers, Bell lineage



Chambers Greater and Lesser Statistics 1993

Greater statistics: learning from data

Three broad categories characterize work in greater statistics:

- preparing data, including planning, collection, organization, and validation

Chambers Greater and Lesser Statistics 1993

Greater statistics: learning from data

Three broad categories characterize work in greater statistics:

- preparing data, including planning, collection, organization, and validation
- analyzing data, by models or other summaries

Chambers Greater and Lesser Statistics 1993

Greater statistics: learning from data

Three broad categories characterize work in greater statistics:

- preparing data, including planning, collection, organization, and validation
- analyzing data, by models or other summaries
- presenting data in written, graphical or other form

Chambers GLS 93 on exhaust

Many mundane . . . activities generate large quantities of potentially valuable data. Examples . . . include retail sales, billing, and inventory management. The data were not generated for the purpose of learning; however, the potential for learning is great. . .

The data usually pass through a computer system nowadays, but aside from the enormous quantity the data are typically thrown away fairly quickly. The computational challenge of collecting and organizing such data is huge. A more clearly statistical challenge is that the data may represent only a portion of the conceptually relevant data; if so, the sample is often biased in crucial ways.

had Statistics fallen behind?

> *Statistics as a discipline exists to develop tools for analyzing data. As such, statistics is an engineering discipline and methodology.*

Yet:

> *Statistics has primarily focused on squeezing the maximum amount of information out of limited data. This paradigm is rapidly diminishing in importance and statistics education finds itself out of step with reality.*

- David Madigan and Werner Stuetzle, "[A Report on the Future of Statistics]: Comment," *Statistical Science* 19, no. 3 (2004): 408.

2014 Statistics re-branding as Data Science

Data Science represents an inevitable (re)-merging of computational and statistical thinking in the big data era. We have to own data science, because domain problems don't differentiate computation from statistics or vice versa, and data science is the new accepted term to deal with a modern data problem in its entirety. Gains for the statistics community are many, and include attracting talent and resources, and securing jobs for our majors, MAs and PhDs.

- Yu, Let us Own Data Science

Neff et al.: Critical Data Studies

Neff: bio

- CC'93!

Neff: bio

- CC'93!
- PhD ethnography

Neff: abstract, 4 critiques

We summarize four critiques that are commonly made in critical data studies:

1. data are inherently interpretive,

Neff: abstract, 4 critiques

We summarize four critiques that are commonly made in critical data studies:

1. data are inherently interpretive,
2. data are inextricable from context,

Neff: abstract, 4 critiques

We summarize four critiques that are commonly made in critical data studies:

1. data are inherently interpretive,
2. data are inextricable from context,
3. data are mediated through the sociomaterial arrangements that produce them, and

Neff: abstract, 4 critiques

We summarize four critiques that are commonly made in critical data studies:

1. data are inherently interpretive,
2. data are inextricable from context,
3. data are mediated through the sociomaterial arrangements that produce them, and
4. data serve as a medium for the negotiation and communication of values.

Neff: abstract, 4 propositions

1. communication is central to the data science endeavor;

Neff: abstract, 4 propositions

1. communication is central to the data science endeavor;
2. making sense of data is a collective process;

Neff: abstract, 4 propositions

1. communication is central to the data science endeavor;
2. making sense of data is a collective process;
3. data are starting, not end points, and

Neff: abstract, 4 propositions

1. communication is central to the data science endeavor;
2. making sense of data is a collective process;
3. data are starting, not end points, and
4. data are sets of stories.

Neff: abstract, 2 calls to action

1. bringing social scientific and humanistic expertise into data science practice

"more ethical" 3x in abstract

Neff: abstract, 2 calls to action

1. bringing social scientific and humanistic expertise into data science practice
- will advance data science and

"more ethical" 3x in abstract

Neff: abstract, 2 calls to action

1. bringing social scientific and humanistic expertise into data science practice

- will advance data science and
- will advance critical data studies.

"more ethical" 3x in abstract

Neff: abstract, 2 calls to action

1. bringing social scientific and humanistic expertise into data science practice

- will advance data science and
- will advance critical data studies.

2. leverage the insights from critical data studies to build new kinds of organizational arrangements

"more ethical" 3x in abstract

Neff: abstract, 2 calls to action

1. bringing social scientific and humanistic expertise into data science practice

- will advance data science and
- will advance critical data studies.

2. leverage the insights from critical data studies to build new kinds of organizational arrangements

- will help advance a more ethical data science.

"more ethical" 3x in abstract

Neff: messages

- critiques benefit from understanding process

A key insight of critical data studies is that interpretation is cooked into the very structures of data and that the work of claiming something as data or a dataset becomes a "rhetorical" move.

Data, as a word, although ends up sounding more authoritative than perhaps those who produce it ever intended

Neff: messages

- critiques benefit from understanding process
  - e.g., not reflecting reflexive data scientists

A key insight of critical data studies is that interpretation is cooked into the very structures of data and that the work of claiming something as data or a dataset becomes a "rhetorical" move.

Data, as a word, although ends up sounding more authoritative than perhaps those who produce it ever intended

Neff: messages

- critiques benefit from understanding process
  - e.g., not reflecting reflexive data scientists
- data scientists benefit from critical data studies

  A key insight of critical data studies is that interpretation is cooked into the very structures of data and that the work of claiming something as data or a dataset becomes a "rhetorical" move.

  Data, as a word, although ends up sounding more authoritative than perhaps those who produce it ever intended

Neff: observations

- subjective design choices: acknowledge rather than eliminate

Neff: observations

- subjective design choices: acknowledge rather than eliminate
- problem choice $>$ problem solution (cf. JWT)

Neff: observations

- subjective design choices: acknowledge rather than eliminate
- problem choice $>$ problem solution (cf. JWT)
- communication is everything

Neff: observations

- subjective design choices: acknowledge rather than eliminate
- problem choice $>$ problem solution (cf. JWT)
- communication is everything
- prominence of ethics

power and principles

how did this capability rearrange power? who can now do what, from what, to whom?

role of rights, harms, justice?

reminder: themes for today

- tensions

reminder: themes for today

- tensions
  - between academia + industry

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)
  - 2000s: .ai, .vc

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)
  - 2000s: .ai, .vc
- sources: experts, evangelists, and expositors

reminder: themes for today

- tensions
  - between academia + industry
    - computers vs math
    - pure vs applied
    - jobs vs. academic power
  - tensions within academia
    - stats vs CS/ML/AI
  - stakes: funding, reputation
- truth, people, and practice
  - constant theme of this course: data as rhetorical claim
- who empowered this field?
  - 1950s/1960s: .mil, .com (esp @ Bell)
  - 2000s: .ai, .vc
- sources: experts, evangelists, and expositors
- inconvenient truths this week: truth is negotiated

up next

- ML= AI2.0

up next

- ML= AI2.0
- ethics and impact

Appendix

- 2020-01-21 : 1 of 14 intro to course

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0
- 2020-04-14 : 12 of 14 ethics

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0
- 2020-04-14 : 12 of 14 ethics
- 2020-04-21 : 13 of 14 present problems: attention economy+VC=dumpsterfire

Appendix

- 2020-01-21 : 1 of 14 intro to course
- 2020-01-28 : 2 of 14 setting the stakes
- 2020-02-04 : 3 of 14 risk and social physics
- 2020-02-11 : 4 of 14 statecraft and quantitative racism
- 2020-02-18 : 5 of 14 intelligence, causality, and policy
- 2020-02-25 : 6 of 14 data gets real: mathematical baptism
- 2020-03-03 : 7 of 14 WWII, dawn of digital computation
- 2020-03-10 : 8 of 14 birth and death of AI
- 2020-03-24 : 9 of 14 big data, old school (1958-1980)
- 2020-03-31 : 10 of 14 data science, 1962-2017
- 2020-04-07 : 11 of 14 AI2.0
- 2020-04-14 : 12 of 14 ethics
- 2020-04-21 : 13 of 14 present problems: attention economy+VC=dumpsterfire
- 2020-04-28 : 14 of 14 future solutions