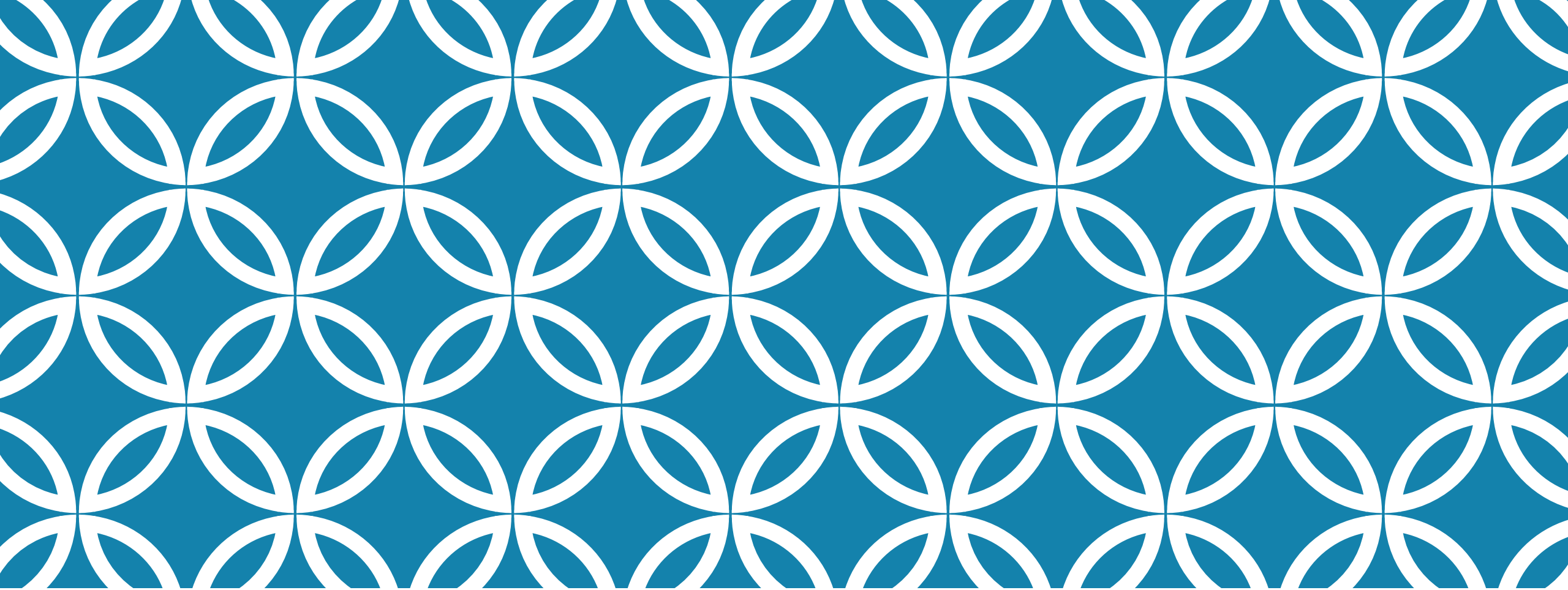


# I DON'T GO LOOKING FOR DATA … DATA USUALLY FINDS ME

Mason Garrison

Friday, March 5, 2021

# PREAMBLE

Most interesting aspect of my work (or at least to me)

- However, it is also the least documented
  - Primarily lives in footnotes, personal statements, and appendices

Audience Mission (should you choose to accept it)

- Where can this project/tale/talk live?

Process outlined today is descriptive, not prescriptive

# OUTLINE

Preamble

Finding Data

- Discuss two major approaches to data
  - Question Driven (Confirmatory)
  - Data Driven (Exploratory)

Recovering Data

Wrangling Data

Advantages | Disadvantages ?

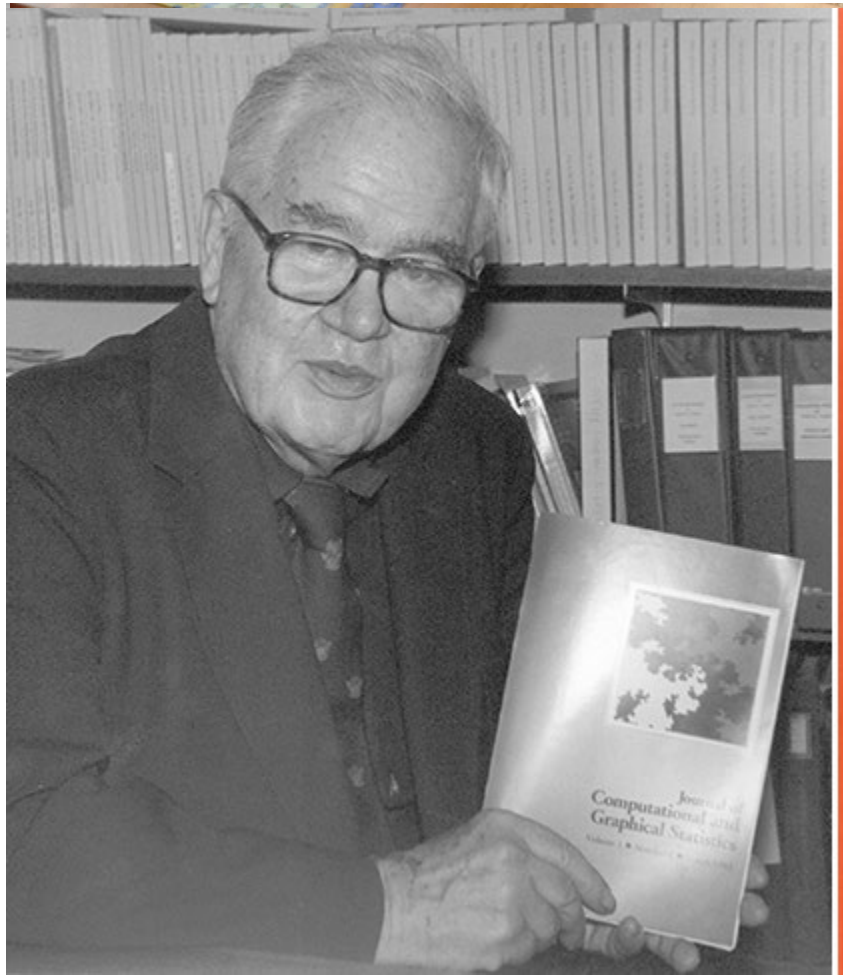
# TWO MAJOR APPROACHES

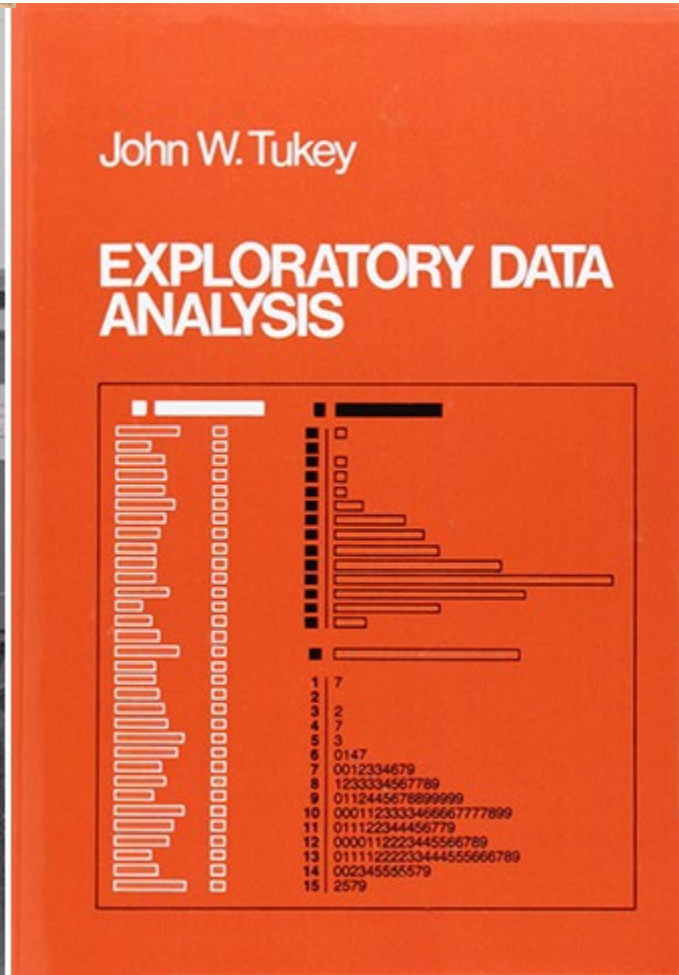
## Exploratory

- Descriptive
- Data-Driven
- Data Scientist

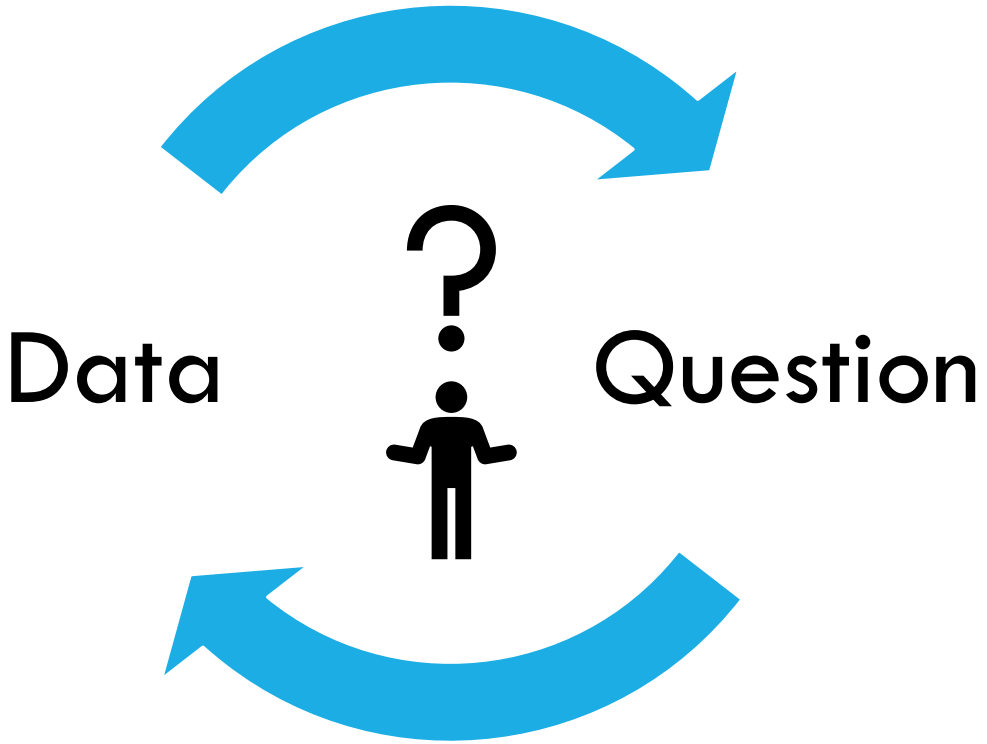
## Confirmatory

- Hypothesis Testing
- Question-Driven
- Research Scientist



# WHERE DO YOU START?



# CONFIRMATORY APPROACH TO ARCHIVAL DATA

**Question-Driven**

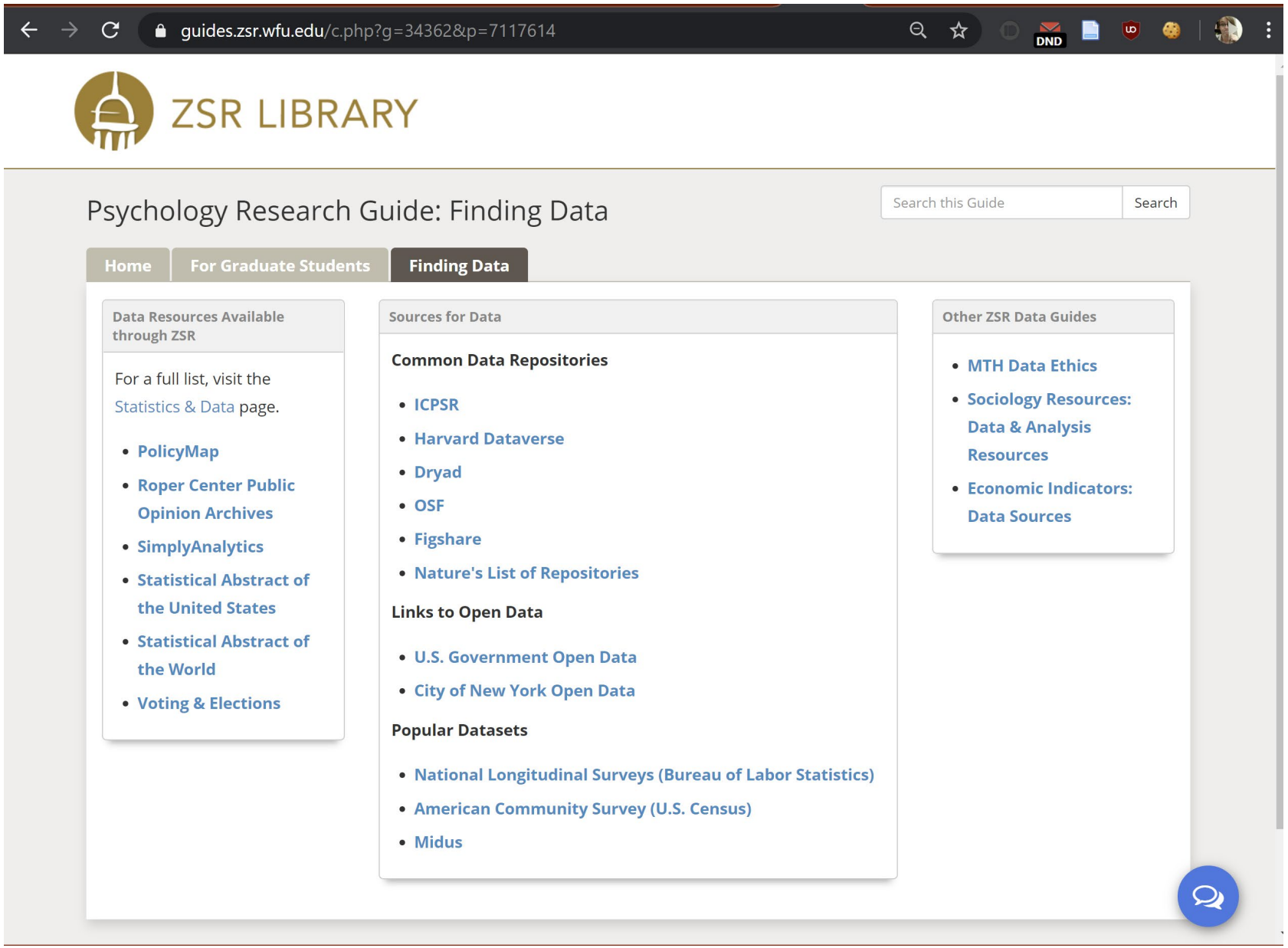
Types of Questions

- Theory
  - E.g., Do smart girls delay sex?
- Measure
  - E.g., Is Coding Speed from the ASVAB a decent proxy for conscientiousness?
- Subjects
  - E.g., Where can I find Twins Raised Apart?
- Model
  - E.g., How do I illustrate my dual mediated survival model?
- Replication
  - E.g., Can I replicate my finding in another sample?
- External
  - E.g., Can I address reviewer two's concern about reliability of difference scores?

# CONFIRMATORY APPROACH TO ARCHIVAL DATA

These questions narrow your search…

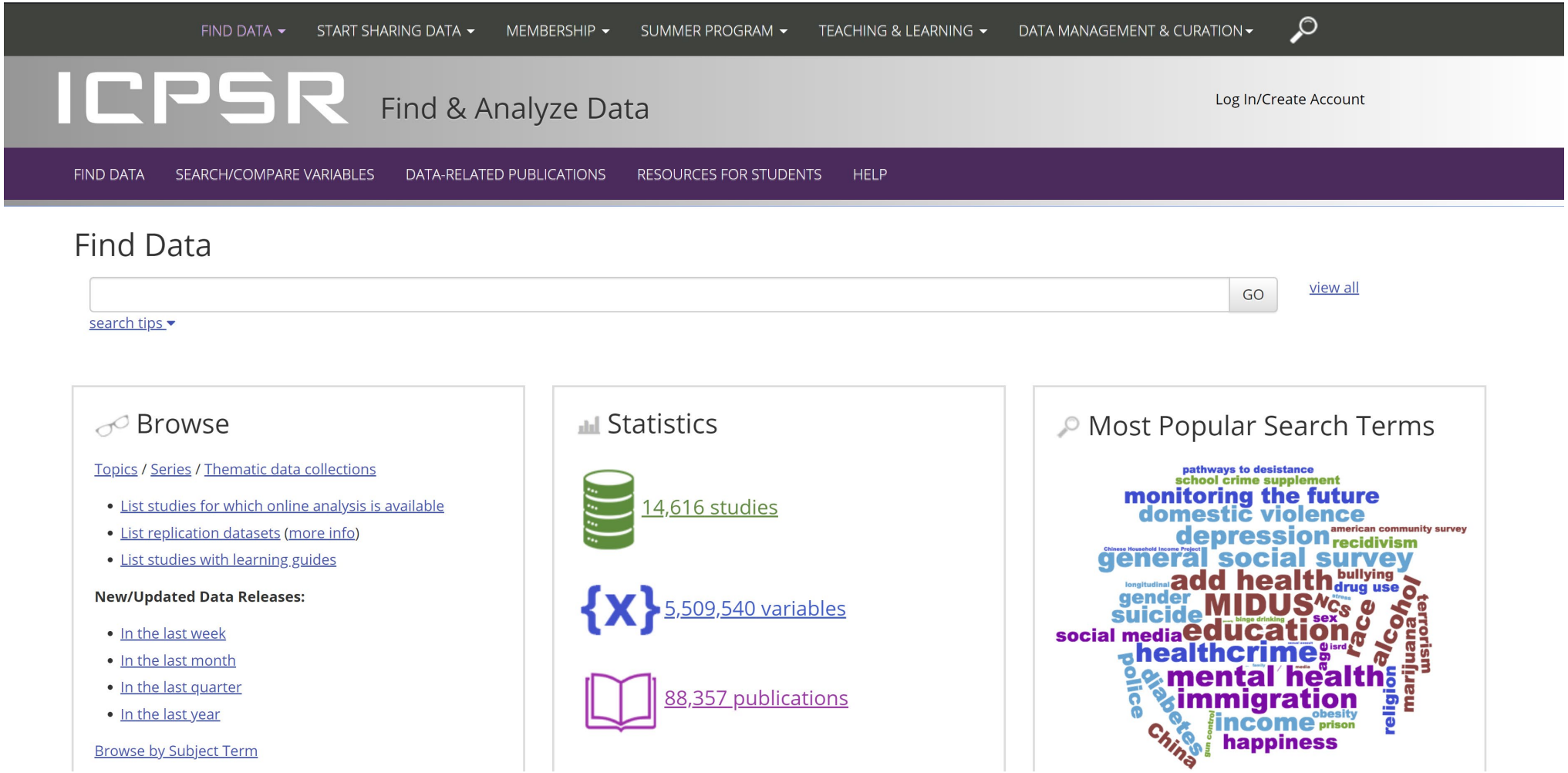
- Otherwise the scope of data is overwhelming
  - The wonderful Kathy Shields
    - helped me add a section to the library website to get you started
    - guides.zsr.wfu.edu/psychology



# PLACES TO START LOOKING

My favorite three:

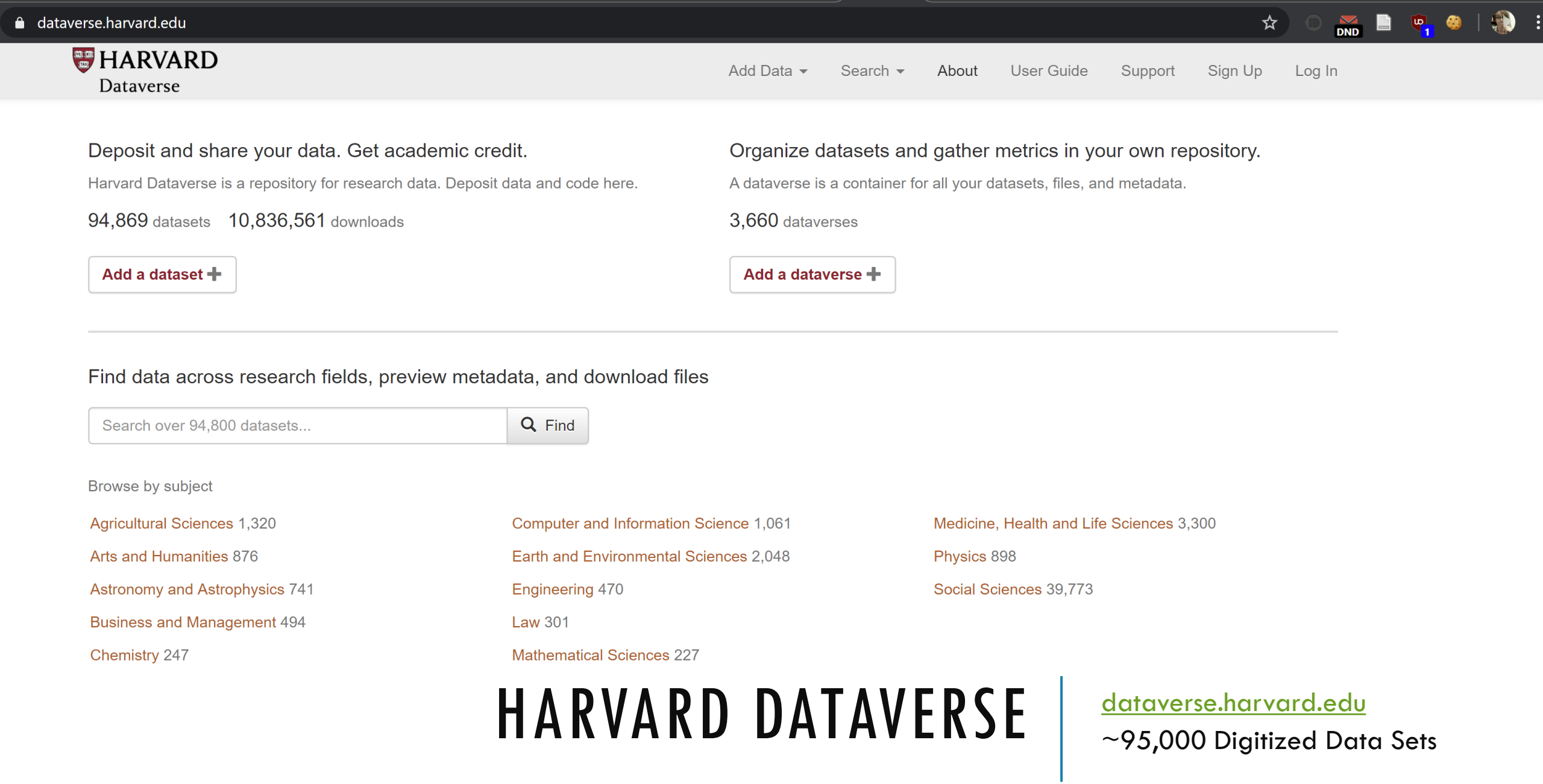
- University of Michigan's ICPSR
- Harvard's Dataverse
- Data.gov




# ICPSR: INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

icpsr.umich.edu

~15,000 Digitized Data sets




# HARVARD DATAVERSE

dataverse.harvard.edu

~95,000 Digitized Data Sets

# EXPLORATORY APPROACH TO ARCHIVAL DATA

## Data-Driven

- Already have a data set in mind
- My earliest experiences fall into this category
  - (and most interesting)
- How does data find you?
  - Referral
    - E.g., A speaker tells you about an intergenerational data set partially run by the BLS
  - Reading
    - E.g., A(n)? Historian using aspects of a marriage study
  - Rumor
    - E.g., A control group is mentioned in the original write-up of the Terman study (1921ish)
  - Serendipity
    - E.g., You fly to SPSP, talk to the person sitting next to you about a study you were in...

# PLACES TO START LOOKING

Sometimes, your specific dataset is already:

- digitized,
- de-identified, and
- clearly located.

Sometimes, only parts of it are…

Other times, it gets interesting…

# DATA RETRIEVAL RANGES WILDLY

## Can be as easy as:

- downloading a data file

## Can be an unwieldly as:

- Appling for access to the "digitized" dataset on the Dataverse for an econometrics class
- Waiting for a panel of two researchers to convene for approval
- Learning that one researcher (E. Lowell Kelly) on the panel died in 1986 and the other (James Conley) is nowhere to be found.
- Teaming up with a 2nd-year assistant professor (Josh Jackson) to find Conley
- Tracking down James Connolly, who legally changed his name in 1992 (or so) from James Conley
- Convincing Jim to share the data
- Determining where the data are
- Retrieving boxes from the basement of:
  - Henry A. Murray Research Archive (Part of Harvard's Dataverse)
  - Jim Connolly's office
  - Bentley Library (Part of Michigan's ICPSR)

The more hurdles between you and the data
The less likely it is that you'll be scooped.

# SILVER-LINING FOR DIFFICULT DATA RETRIEVAL

# INTERMISSION

# BIT OF CONTEXT: MARRIAGE STUDY

Between 1935 and 1938, 300 engaged couples participated in a longitudinal study on personality and newly-formed marriages

(KCLS; Kelly & Conley, 1987; Jackson, Connolly, Garrison et al 2015).

The original aim of the study was to understand what characteristics influenced marital compatibility.

However, subsequent follow-ups broadened the scope of the study.

# BIT MORE CONTEXT

Participants were primarily from Connecticut with upper- and middle-class backgrounds in their early to mid 20s.

Acquaintance–ratings of personality were obtained from close friends – most raters come from the participant's wedding party (Kelly, 1977).

The study has multiple follow-ups in the late 1940s, 1954/55, 1979-1981, and 2012-2014.

# DATA WRANGLING

| Less Interesting | More Interesting |
| --- | --- |
| Digitized Data | Undigitized |
| Systematic Data Structure | Unsystematic Data Structure |
| Computer Readable Storage Format<br>▪ csv<br>▪ tab-delimated, etc | Non-Standard Format<br>▪ Magnetic Tapes<br>▪ Punchcards |
| Measures in current use | Measures no longer in use |
| No Item-Level Data | Yes, Item-Level Data |
| Clear Documentation | No Documentation |

# DIGITIZING DATA DEPENDS ON THE FORMAT

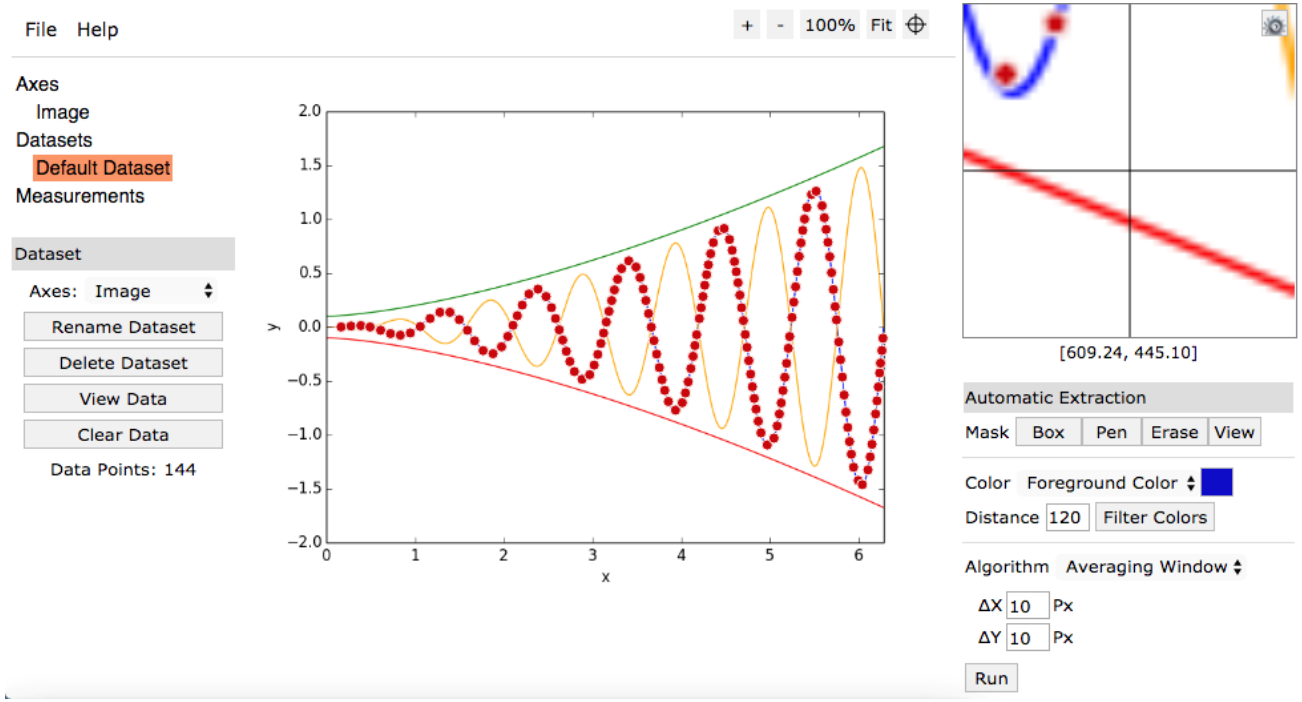
## Formats

- Magnetic Tapes
- Punchcards
- Tally Sheets
- Handwritten Forms
- Typed Forms
- Figures
- Photographs

## Some Digitization is Easily Automated

- OCR Software (optical character recognition)
  - can convert some text

- WebPlotDigitizer
  - can extract data from figures
    - automeris.io/WebPlotDigitizer



# WHY IS THE MEASURE NO LONGER IN USE

## Less Interesting

Newer Version Available

Cheaper Alternative Available

Shorter Version Available

Researcher interest has waned

## More Interesting

Unreliable

Invalid

Theory has changed

# ITEM-LEVEL DATA

If you can get the item-level responses, DO IT!

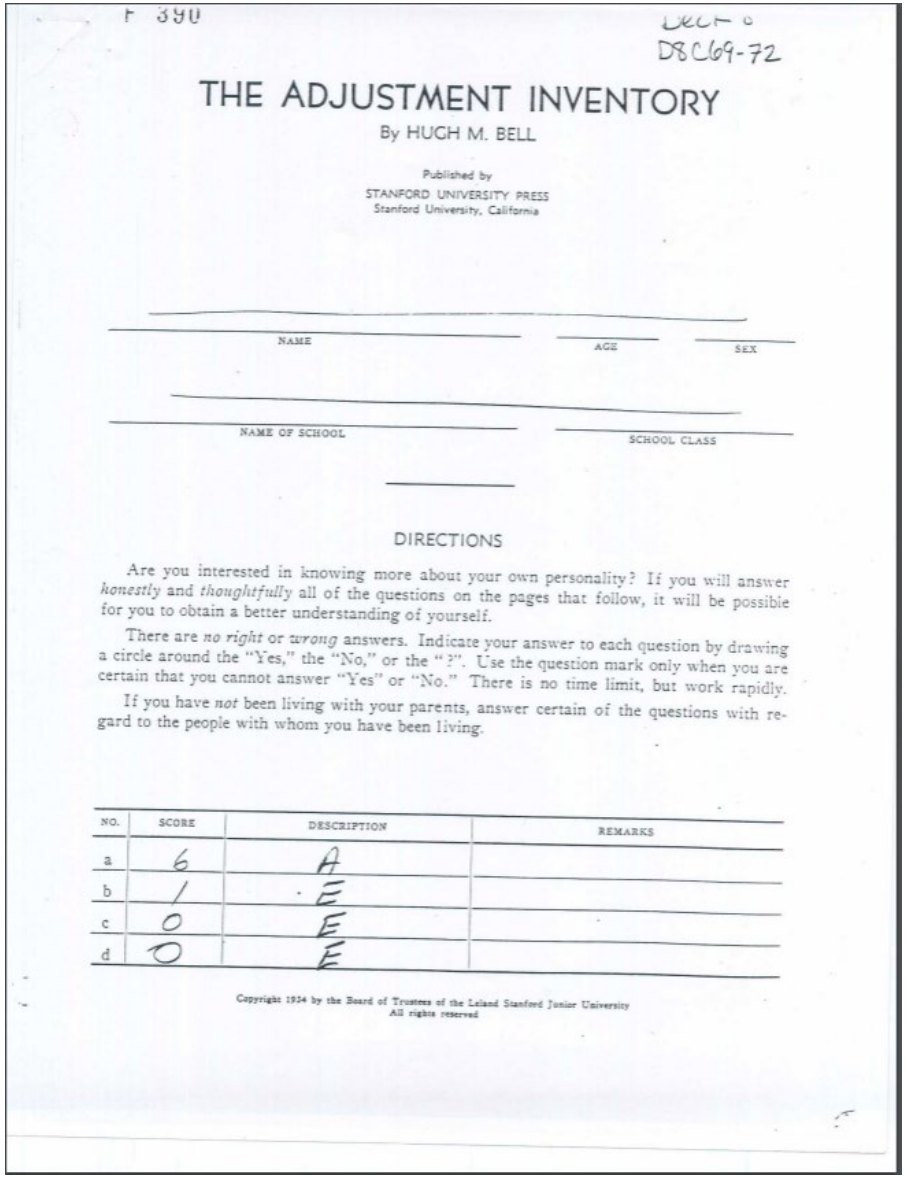
You can rescore an old scale, into something meaningful.

More important in older measures

Example:

- Bell Adjustment Inventory (1934) Contains Health Data
- Kelly Personality Rating Scale (1940) captures the Big Five

# BELL ADJUSTMENT INVENTORY



F 390

D8C69-72

## THE ADJUSTMENT INVENTORY

By HUGH M. BELL

Published by
STANFORD UNIVERSITY PRESS
Stanford University, California

NAME AGE SEX

NAME OF SCHOOL SCHOOL CLASS

### DIRECTIONS

Are you interested in knowing more about your own personality? If you will answer *honestly* and *thoughtfully* all of the questions on the pages that follow, it will be possible for you to obtain a better understanding of yourself.

There are *no right* or *wrong* answers. Indicate your answer to each question by drawing a circle around the "Yes," the "No," or the "?". Use the question mark only when you are certain that you cannot answer "Yes" or "No." There is no time limit, but work rapidly.

If you have *not* been living with your parents, answer certain of the questions with regard to the people with whom you have been living.

| NO. | SCORE | DESCRIPTION | REMARKS |
|---|---|---|---|
| a | 6 | A | |
| b | 1 | E | |
| c | 0 | E | |
| d | 0 | E | |

# THE ADJUSTMENT INVENTORY

By HUGH M. BELL

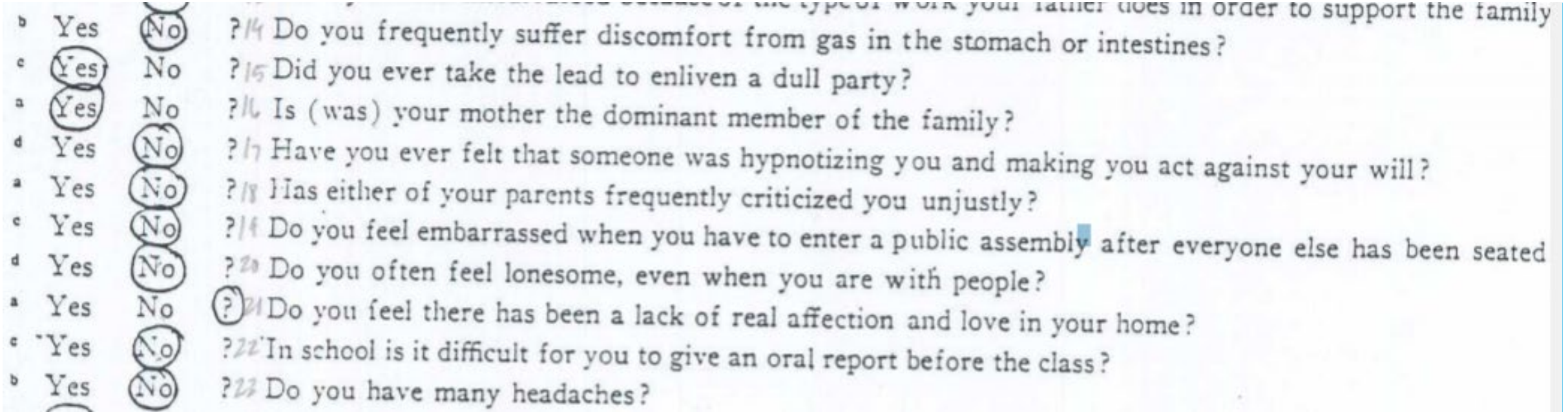
Published by
STANFORD UNIVERSITY PRESS
Stanford University, California

## DIRECTIONS

Are you interested in knowing more about your own personality? If you will answer *honestly* and *thoughtfully* all of the questions on the pages that follow, it will be possible for you to obtain a better understanding of yourself.

There are *no right* or *wrong* answers. Indicate your answer to each question by drawing a circle around the "Yes," the "No," or the "?". Use the question mark only when you are certain that you cannot answer "Yes" or "No." There is no time limit, but work rapidly.

If you have *not* been living with your parents, answer certain of the questions with regard to the people with whom you have been living.




# WAY MORE INFORMATION THAN A SUM!!!

# KELLY PERSONALITY RATING SCALE

Graphical Scale with 36 items

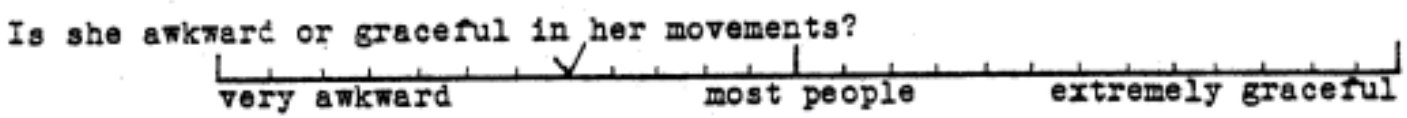





TABLE 1

TRAIT IN TERMS OF QUESTION[1] ASKED RATER

| No. in Scale | | Coef. of Reliability |
|---|---|---|
| 2. | How intelligent is she? | .86 |
| 3. | How does she meet new social situations? | .84 |
| 4. | How social is she? | .79 |
| 5. | Is she physically attractive? | .78 |
| 6. | Is she nervous? | .77 |
| 7. | Is she popular with other people? | .77 |
| 8. | What is her attitude about religion? | .72 |
| 1. | Is she physically energetic? | .72 |
| 9. | How does she meet her appointments? | .71 |
| 11. | How courteous is she? | .69 |
| 12. | How jealous is she? | .69 |
| 13. | What sort of a voice has she? | .69 |
| 14. | How coöperative is she? | .66 |
| 15. | How cultured is she? | .65 |
| 16. | Is she awkward? | .65 |
| 17. | How well does she stick to a task? | .65 |
| 18. | How is she with regard to money? | .64 |
| 19. | How honest and fair is she? | .63 |
| 31. | How vain is she? | .63 |
| 20. | How does she dress? | .61 |
| 21. | Is she patient? | .60 |
| 22. | Does she possess common sense? | .60 |
| 23. | Are her interests wide or narrow? | .60 |
| 24. | What is her usual disposition? | .59 |
| 25. | How conventional is she? | .57 |
| 26. | Is she a good sport? | .56 |
| 27. | Is she generally quiet or boisterous? | .56 |
| 28. | How sincere is she? | .51 |
| 29. | What kind of a temper does she have? | .50 |
| 30. | Is she boring or entertaining? | .50 |
| 36. | How selfish is she? | .49 |
| 10. | How much initiative does she have? | .49 |
| 32. | Is she tactful? | .49 |
| 33. | Does she have definite ideas? | .48 |
| 35. | Can you count on her to do a thing? | .34 |
| 34. | How good is her sense of humor? | .31 |

[1]The questions as here listed are as they appear in the *F* form. Identical questions using "he," "his," and "him" are used in the *M* form.

# KPRS

## Graphical Scale with 36 items

PERSONALITY RATING SCALE

Instructions: This is a device to determine personality traits: to find out what sort of person one is in the opinion of his friends and acquaintances. Although all of one's friends do not have the same opinion with regard to his various traits, it has been found that when the same person is rated by a number of different people, the average of the ratings constitutes a good measure of his personality.

_____, an acquaintance of yours is desirous of obtaining such a measure of her personality and wishes you, along with several others, to give your best judgment as to what sort of person she is by rating her on the traits listed below. Since you may not have had experience with this sort of rating, the following example is given:

Think of this person for a moment. Is she awkward or graceful? Very few people are either very awkward or very graceful in all situations, so instead of asking you to check one word or the other you are given the opportunity of indicating the exact point on the scale where you think she belongs, thus: if you think she is considerably more awkward than most people you know, but not the most awkward, you would perhaps check her thus:

Is she awkward or graceful in her movements?

very awkward — most people — extremely graceful

It will be appreciated if, on the following pages, you would rate this person in this manner on each of the traits listed. In each case, ask yourself what kind of person she is and then place a check mark on the line at the point which you think best describes her. Think of the person always in terms of the average person who would be rated at the exact center of the scale. The ends of the scale should be thought of as indicating the most extreme individuals whom you know, with all others falling at various points between. Be careful not to judge a person high on all traits simply because she is high on some. Often people are very high on some traits and very low on others.

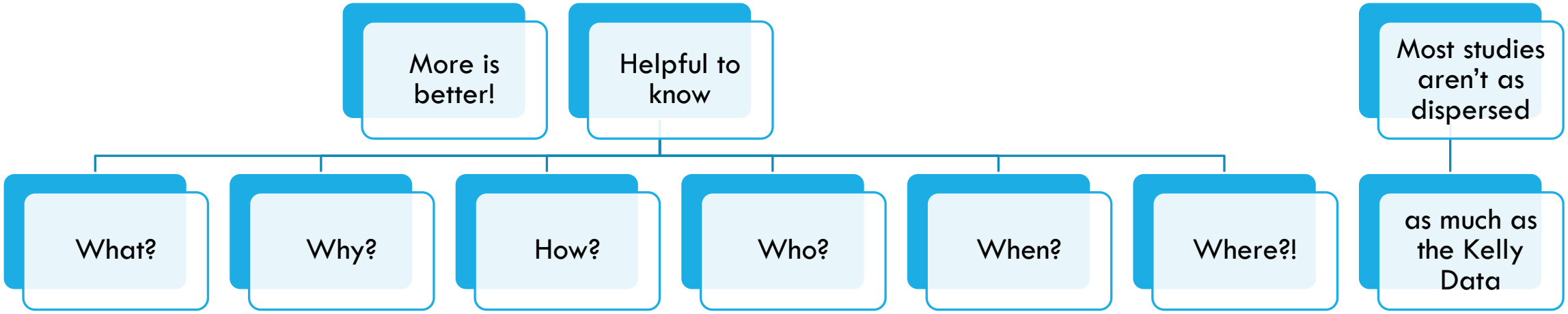


TABLE 1

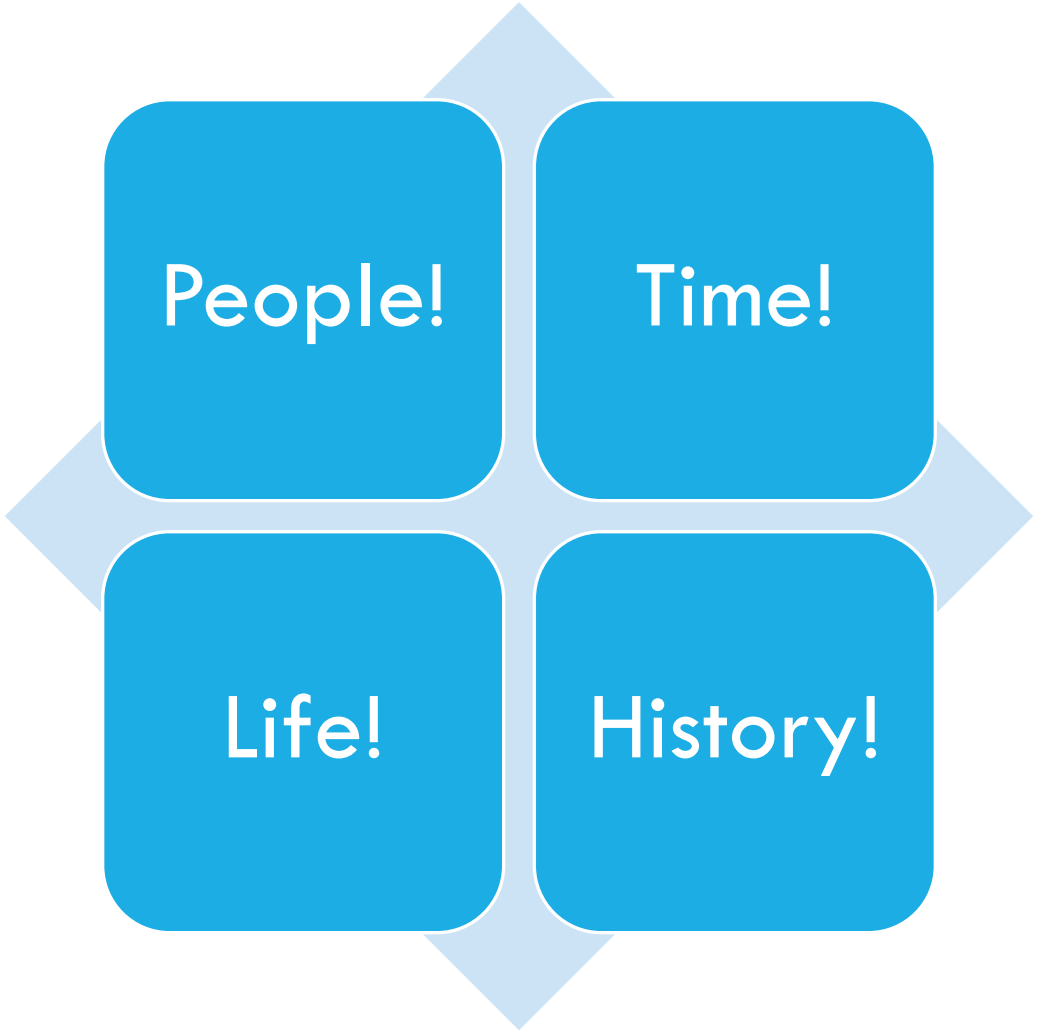
TRAIT IN TERMS OF QUESTION[1] ASKED RATER

| No. in Scale | | Coef. of Reliability |
|---|---|---|
| 2. | How intelligent is she? | .86 |
| 3. | How does she meet new social situations? | .84 |
| 4. | How social is she? | .79 |
| 5. | Is she physically attractive? | .78 |
| 6. | Is she nervous? | .77 |
| 7. | Is she popular with other people? | .77 |
| 8. | What is her attitude about religion? | .72 |
| 1. | Is she physically energetic? | .72 |
| 9. | How does she meet her appointments? | .71 |
| 11. | How courteous is she? | .69 |
| 12. | How jealous is she? | .69 |
| 13. | What sort of a voice has she? | .69 |
| 14. | How coöperative is she? | .66 |
| 15. | How cultured is she? | .65 |
| 16. | Is she awkward? | .65 |
| 17. | How well does she stick to a task? | .65 |
| 18. | How is she with regard to money? | .64 |
| 19. | How honest and fair is she? | .63 |
| 31. | How vain is she? | .63 |
| 20. | How does she dress? | .61 |
| 21. | Is she patient? | .60 |
| 22. | Does she possess common sense? | .60 |
| 23. | Are her interests wide or narrow? | .60 |
| 24. | What is her usual disposition? | .59 |
| 25. | How conventional is she? | .57 |
| 26. | Is she a good sport? | .56 |
| 27. | Is she generally quiet or boisterous? | .56 |
| 28. | How sincere is she? | .51 |
| 29. | What kind of a temper does she have? | .50 |
| 30. | Is she boring or entertaining? | .50 |
| 36. | How selfish is she? | .49 |
| 10. | How much initiative does she have? | .49 |
| 32. | Is she tactful? | .49 |
| 33. | Does she have definite ideas? | .48 |
| 35. | Can you count on her to do a thing? | .34 |
| 34. | How good is her sense of humor? | .31 |

[1]The questions as here listed are as they appear in the *F* form. Identical questions using "he," "his," and "him" are used in the *M* form.




# DOCUMENTATION

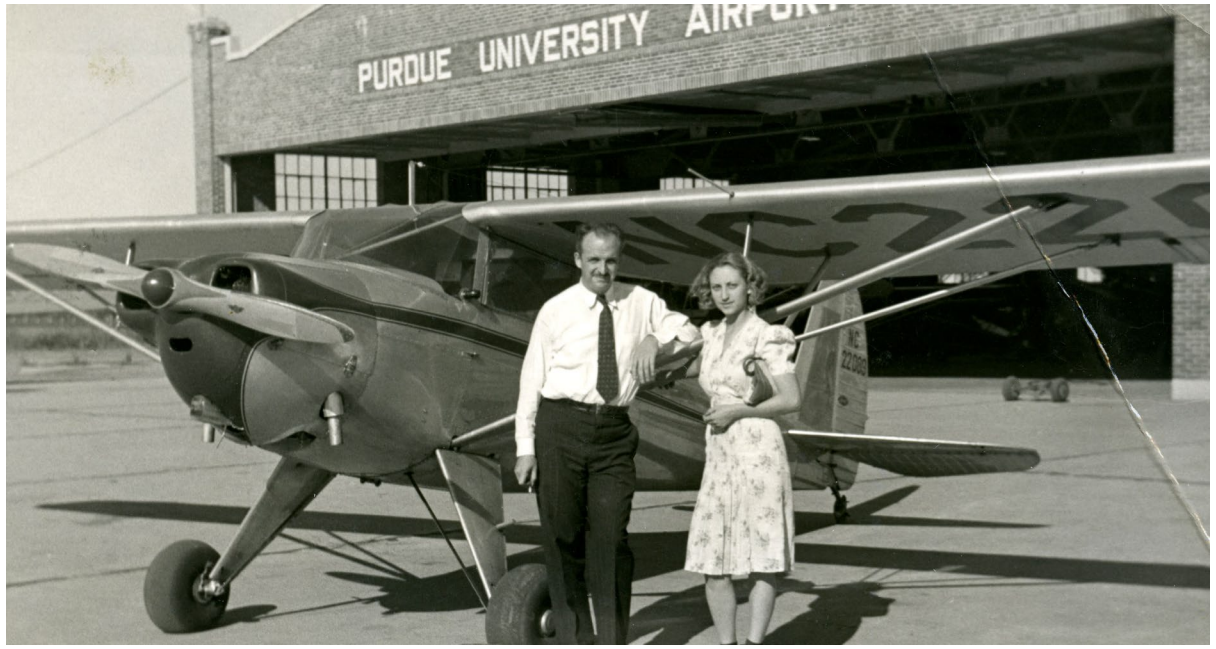


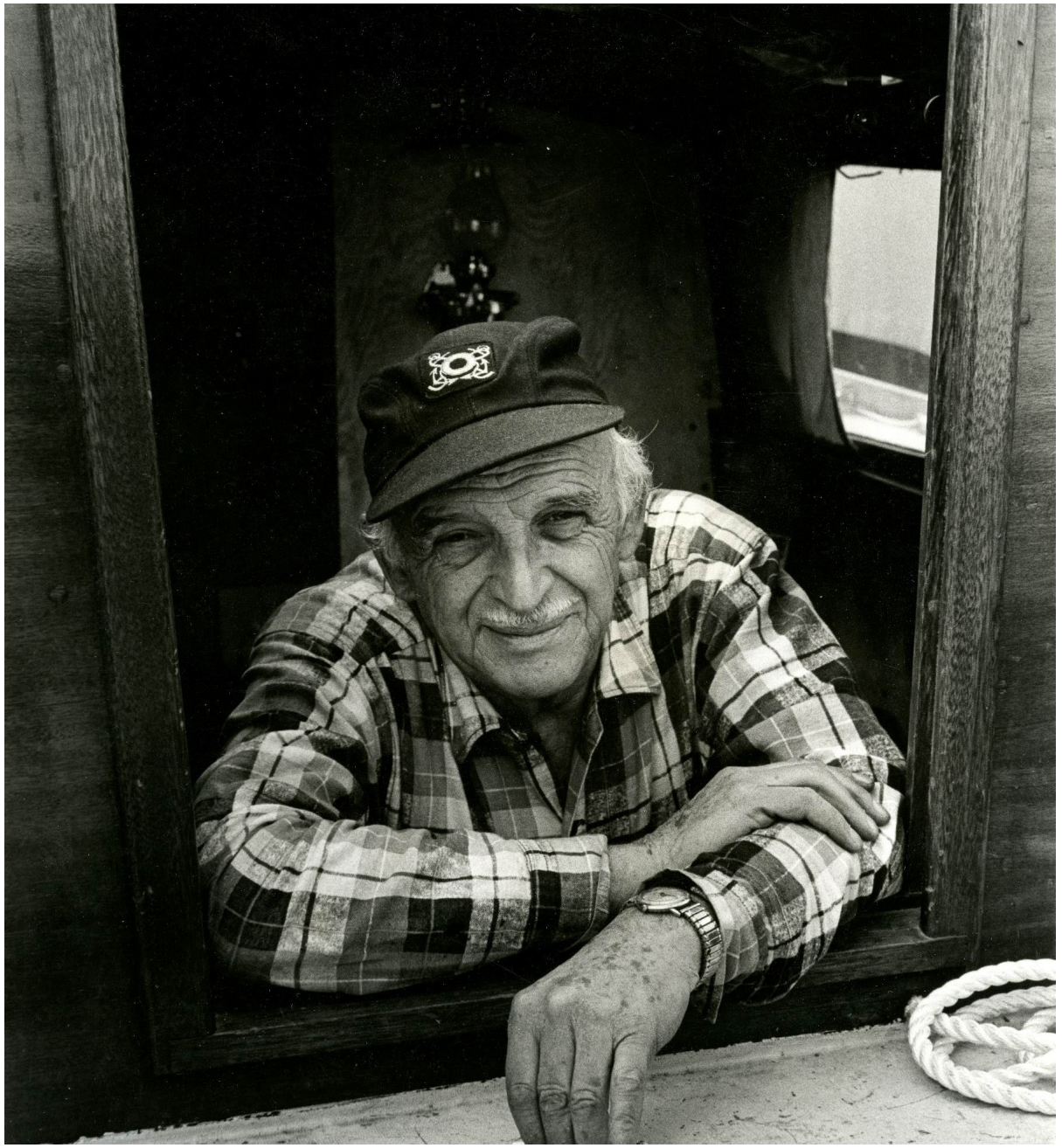

# WHY IS IT SO SPREAD OUT?

| | |
|---|---|
| Lewis Terman | Stanford |
| Clarence Darrow | University of Hawaii |
| Nazis | University of Berlin -> Austria & Switzerland |
| Great Depression | University of Connecticut -> Purdue |
| More Nazis | Airforce |
| No More Nazis | Purdue |
| Robert Marquis | University of Michigan |
| John F. Kennedy | Peace Corp |

# WHY IS IT SO SPREAD OUT?

There's more, but I'll spare you


PURDUE UNIVERSITY



| Advantages | Disadvantages |
| --- | --- |
| • Cool Data<br>• Longer Timespan<br>• Bigger Sample | • Old<br>• Messy<br>• Difficult<br>• Generalizability? |

# ADVANTAGES | DISADVANTAGES ?

# FINAL THOUGHTS

and that's why you always

- ~~leave a note~~
- document your data!!!

# ALTERNATIVE TITLES

Let’s talk about data

Stop! Data time!

It's Raining Data

Data Are a Girl's Best Friend

Datasets Keep Fallin' on My Head

Data Data Data I Don't Love You You Don't Love Me Aha Aha Aha

Just Data

Do you want to find a dataset?