Measurement, Scaling, and Dimensional Analysis
2019 ICPSR Summer Program

Adam M. Enders
Department of Political Science
University of Louisville

June 25, 2019

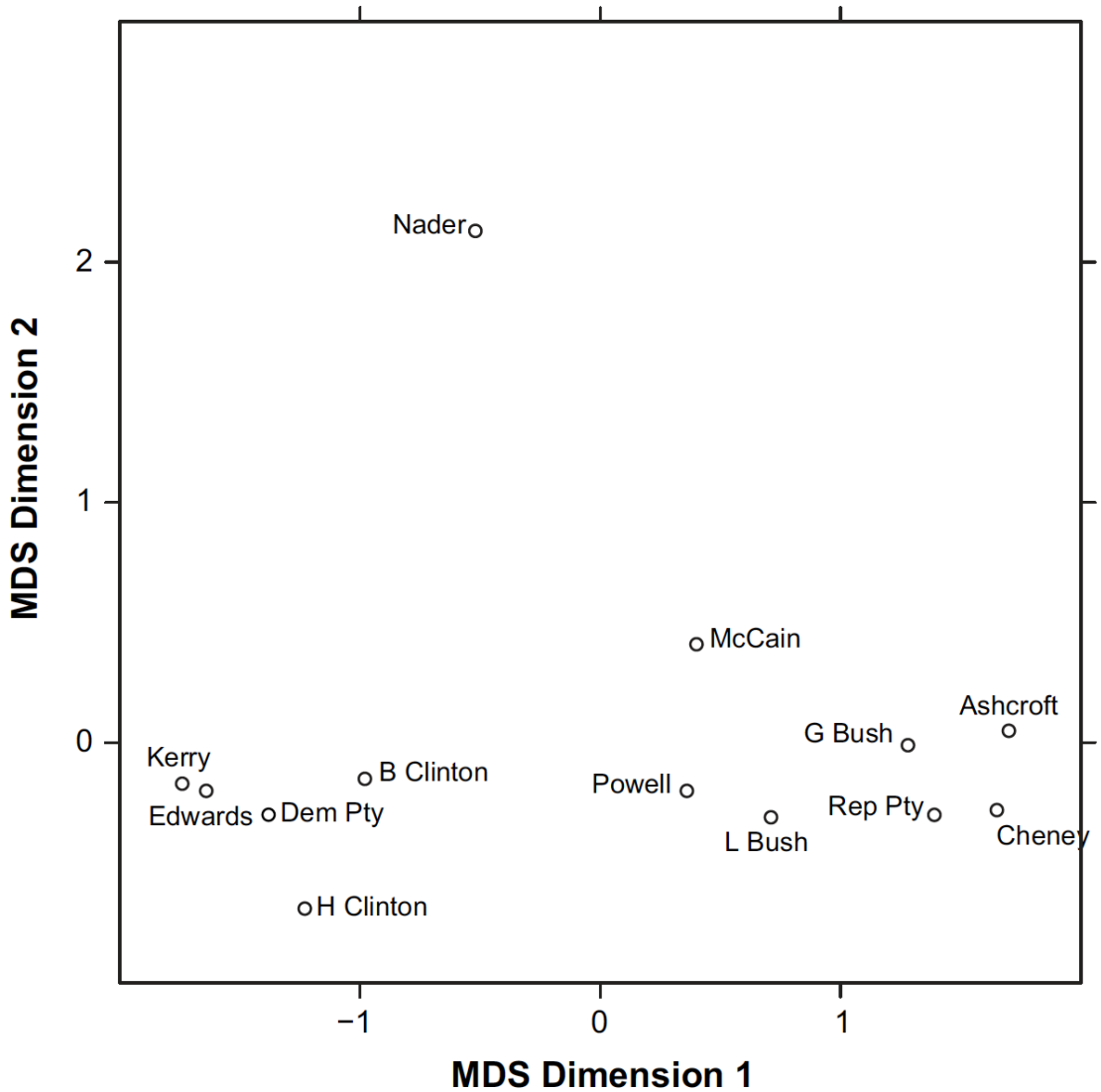
Goal of Course

- **Explore a family of techniques that will help us better measure the things we're interested in**
  - Students will (hopefully) walk away with a better sense of:
    1. What it really means to measure something
    2. How to go about measuring social phenomena in a deliberate, theoretical, and empirically rigorous way
    3. How to lend substantive interpretations to measurements and convince others of those interpretations

What is Scaling?

- People speak about scaling analysis as if its a single method called "scaling"
- There is a commonality to the methods we label "scaling"
  - They are all geometric representations of data
- The models themselves provide information about the substantive processes that produce the data (DGP in some sense)
- Producing geometric structures lies at the heart of everything we'll be doing

Example: 2004 ANES Feeling Thermometers



Why "Scale" Things?

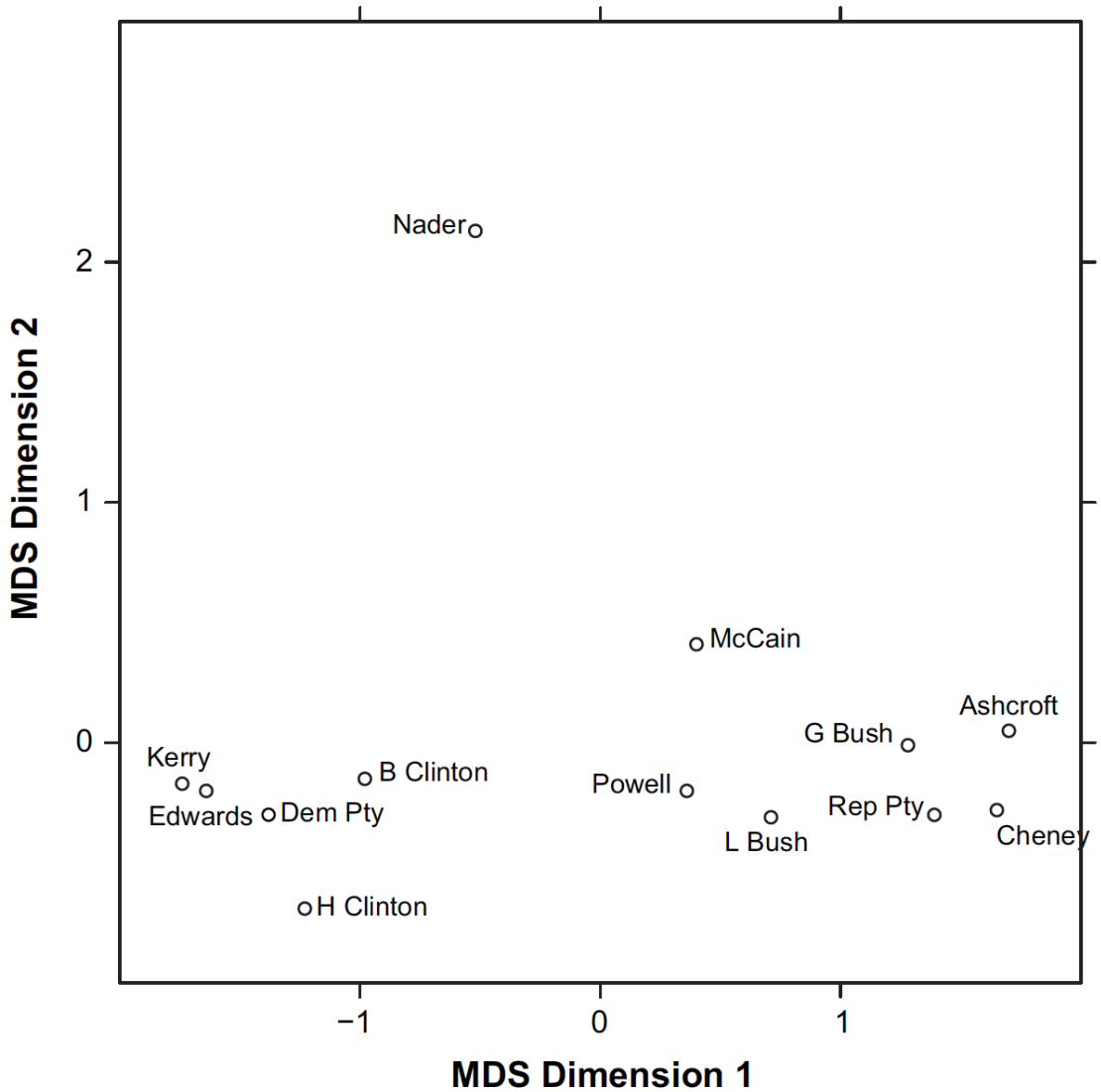
Four main objectives (in no particular order)

1. Data reduction
   - We live in an information age and we're confronted by big data that is hard to examine in raw form
   - 1000s of units on 100s or 1000s of variables
   - Information is useless if we can't comprehend it
   - So, we reduce data down, distill it, from an incomprehensible mass to some digestible that also retains the interesting, important properties of the original data
   - Examples
     - Principal components analysis
     - Factor analysis is used this way, but sort of a misuse
     - Likert scales/additive scales

Example: 2004 ANES Feeling Thermometers

wbush	obama	mccain	biden	palin	hclinton	bclinton	rice	limbaugh
85	25	60	15	90	20	20	95	52
60	0	70	50	70	30	0	50	40
85	30	60	50	70	40	40	100	50
70	30	70	50	85	40	30	70	70
70	0	85	50	60	0	0	70	40
85	15	70	50	70	0	0	100	50
70	30	70	15	100	40	30	100	60
0	100	50	60	15	50	70	50	0
60	30	85	15	100	60	40	50	40
85	0	85	50	85	0	0	70	50
100	0	70	50	85	0	0	100	70
70	15	85	15	85	15	50	85	70
70	85	85	100	70	85	70	100	0
30	40	85	100	70	85	85	70	0
85	0	70	0	85	0	0	100	70
40	40	70	50	85	85	70	70	50
30	60	60	85	60	70	85	30	40
30	100	0	70	0	50	85	70	0
0	70	30	70	15	70	70	30	0
40	70	60	70	50	70	85	60	50
60	15	85	30	70	60	40	70	60
100	40	100	40	100	60	50	100	100
60	40	85	50	60	20	30	60	50
0	100	15	60	0	60	40	60	0
30	70	60	50	30	60	60	30	15
70	70	50	50	50	50	50	50	50
25	85	50	50	40	85	100	60	50
1	70	80	80	40	90	100	50	10
25	85	35	90	30	60	75	15	10
0	70	15	60	30	70	15	0	0

Example: 2004 ANES Feeling Thermometers



Why "Scale" Things?

2. Assess the dimensionality of the information we're analyzing
   - We try to understand the number and nature of the distinct sources of variability in a set of data
   - Think about market researchers trying to figure out what aspects of breakfast cereal people pay attention to when buying cereal
     - Cost, the box, sugar content, organic, where its produced?
     - These are all potential sources of variability in people's behavior when it comes to purchasing cereal
     - Assessing dimensionality is a way of figuring out which of those distinct sources of variability is actually present in people's minds

Why "Scale" Things?

- We use dimensionality analysis to separate the interesting and important sources of variability from the other potential sources of variability (error) that don't matter/aren't interesting/important
  - Error isn't mistakes, just information that isn't useful or interesting
- Deal with the "curse of dimensionality"
  - The useful thing about making geometric representations is that they can be represented visually very nicely
  - The problem: we can't see into more than 3-4 dimensions very easily
  - Analytically we aren't limited by the curse of dimensionality. We can engage $n - 1$ dimensions. But we can't produce graphical representations beyond 3-4.

Why "Scale" Things?

3. Measurement
   - Extremely powerful measurement tools
   - Can extract information from incomprehensible data measured at ordinal and nominal levels and produce comprehensible interval level information
   - Example: Likert scales
     - Likert scales take ordinal information and produces an interval level scale
   - Measurement is, itself, a theory and theories are meant to be tested
   - Scaling methods are ways of testing those theories about our measures

Why "Scale" Things?

4. Statistical graphics
   - All methods we use are amenable to being represented in pictorial form
   - A picture is worth 1000 words – worth 1000 numbers too
   - Visualizing data is extremely useful in communicating our findings to reviewers, editors, colleagues, the public, etc.
     - A lot more powerful that tables of coefficients or equations
   - Science is an inherently social enterprise – visualizing data via statistical graphics makes the socializing, the communication more efficient and easier.

Who Are We?

- Instructor: **Adam Enders**
  - Assistant professor of political science
  - University of Louisville
  - Public opinion and political behavior research
  - Substantive research interests:
    - Conspiratorial thinking
    - Polarization
    - Partisanship and ideology
  - Methodology interests:
    - Latent variable modeling of all sorts (particularly IRT and MDS)
    - In addition to class: SEM, differential item functioning, optimal scaling
- Teaching Assistant: **Tyler Girard**
  - Graduate student, political science
  - University of Western Ontario
  - Interests in international political economy, norms

Who Are You?

1. Name
2. Affiliation (university, department, employer...whatever makes sense)
3. Field of study/work
4. Why are you interested in the course?
   - What methodologies are you most interested in?
   - What does your work look like, what problems are you trying to solve?
   - Is there anything relevant to class that isn't on the syllabus?

Course Material

- Course webpage: `www.adamenders.com/teaching`
  - Click "Scaling and Dimensional Analysis"
  - Password: "ICPSR2019"
- Will post slides, data, code, and homework assignments
- *Please do not share course materials with non-participants*

See course syllabus for more information

Summer Program Details

- Figure out what's useful to you and do that
  - Don't worry about what course(s) you initially registered for
  - Spend the first day or two exploring different courses, but make selections after that
  - Three courses is tough; wouldn't recommend more than two plus supplemental lectures
- Lectures
  - Blalock Lectures: in evening, topical
  - Mathematics for Social Scientists: recommend II (focus on matrix algebra), if any
  - R lectures tonight through July 6 @ 5:30 PM: recommend if no prior experience with R
- Office hours
  - Most faculty (myself included!) are more than happy to discuss research projects
  - Find a way to apply course material to your interests, your data – get a paper out of it!

Questions???