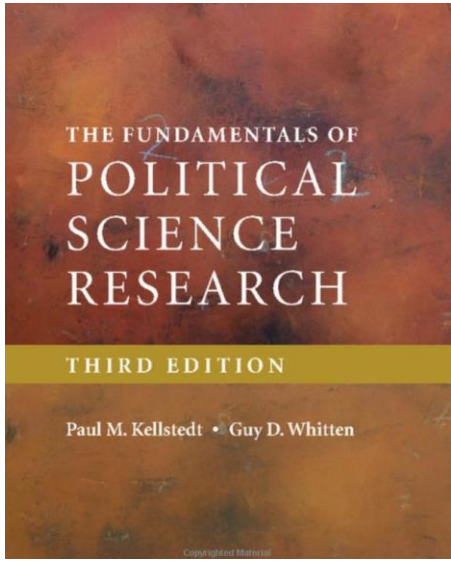
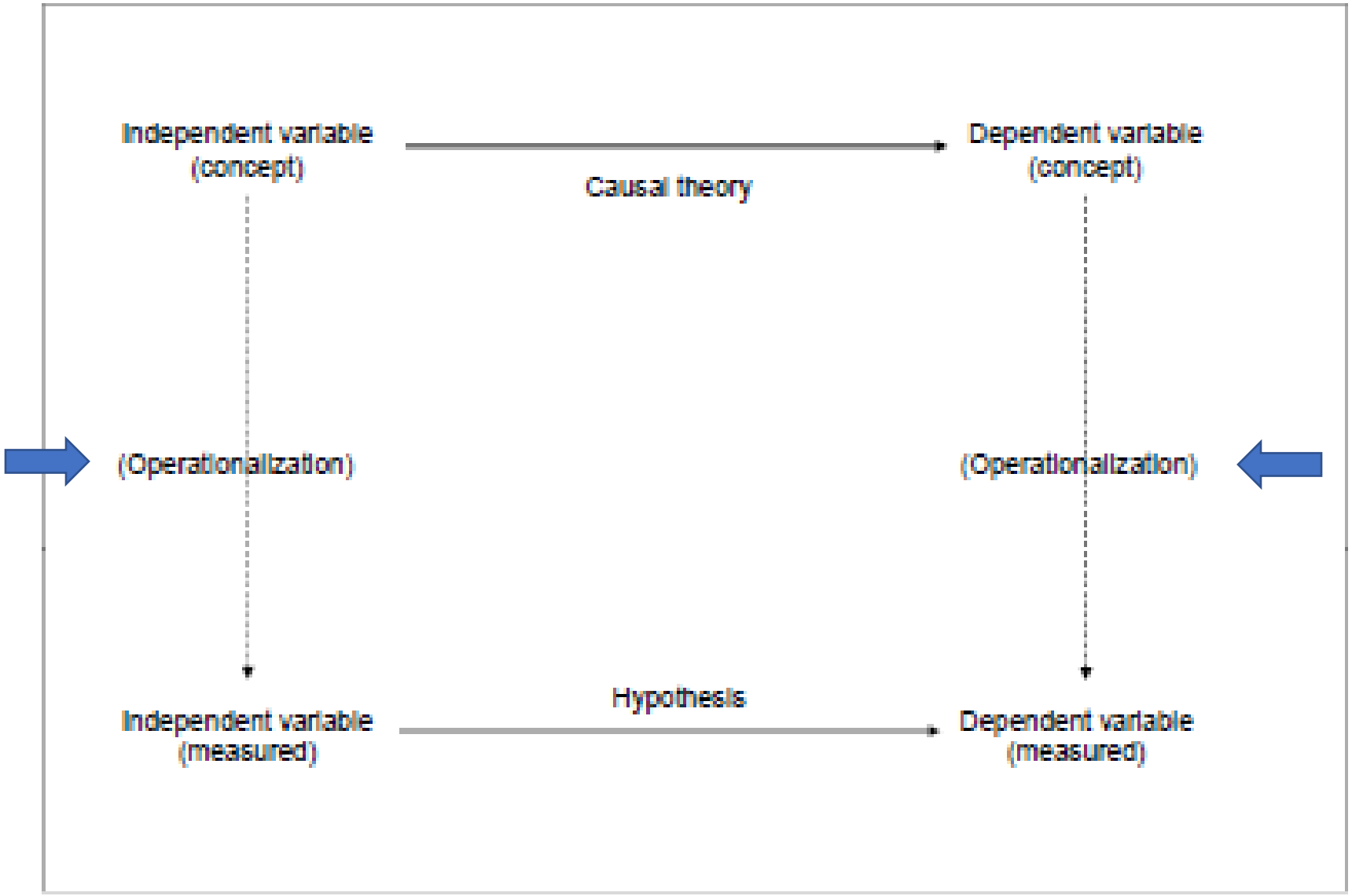


The Fundamentals of Political Science Research, 3rd edition

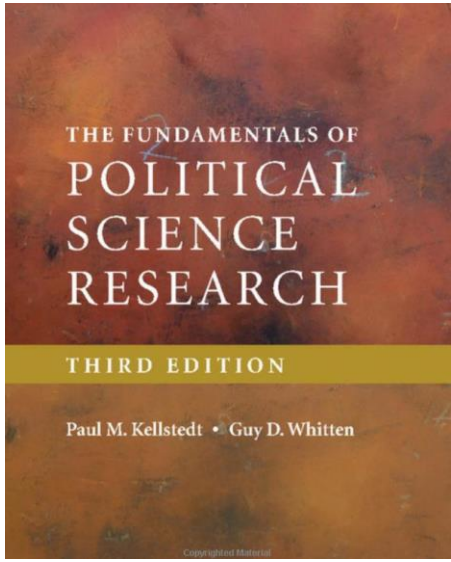
Chapter 5: Measuring Concepts of Interest

How do we evaluate our theories?

- We need to make ***empirical observations***. In other words, we need to compare our abstract theoretical ideas with reality. (Remember, “empirical” just means “based on observations.” They might be quantitative or in-depth qualitative.)
- We need to be as confident as possible that our concepts in our theory correspond as closely as possible to our empirical observations.
- This is called the **problem of “measurement**.”

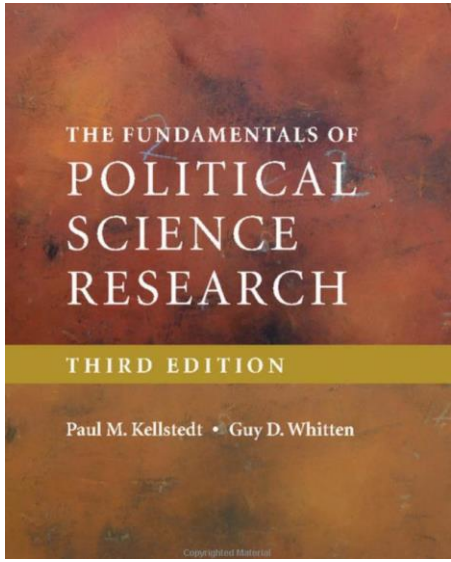

Independent variable
(concept)
Causal theory
Dependent variable
(concept)
(Operationalization)
(Operationalization)
Hypothesis
Independent variable
(measured)
Dependent variable
(measured)

Do you see the disconnect?



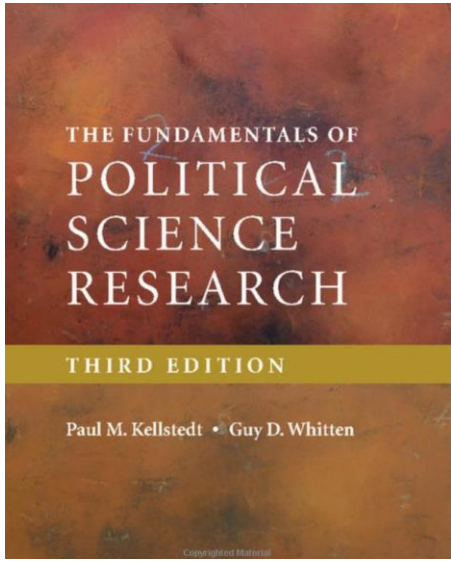
- The relationship that we care about most is one we cannot ***directly observe***. We therefore have to rely on *potentially imperfect measures* of the concepts we care about
- Three issues of measurement
  - Clarity
  - Reliability
  - Validity

Conceptual clarity



- What is the exact nature of the concept we're trying to measure?
- Example: How should a survey question measure “income”?
  - 1 “What is your income?” How would you measure it?
- Clearer the concepts, easier to measure

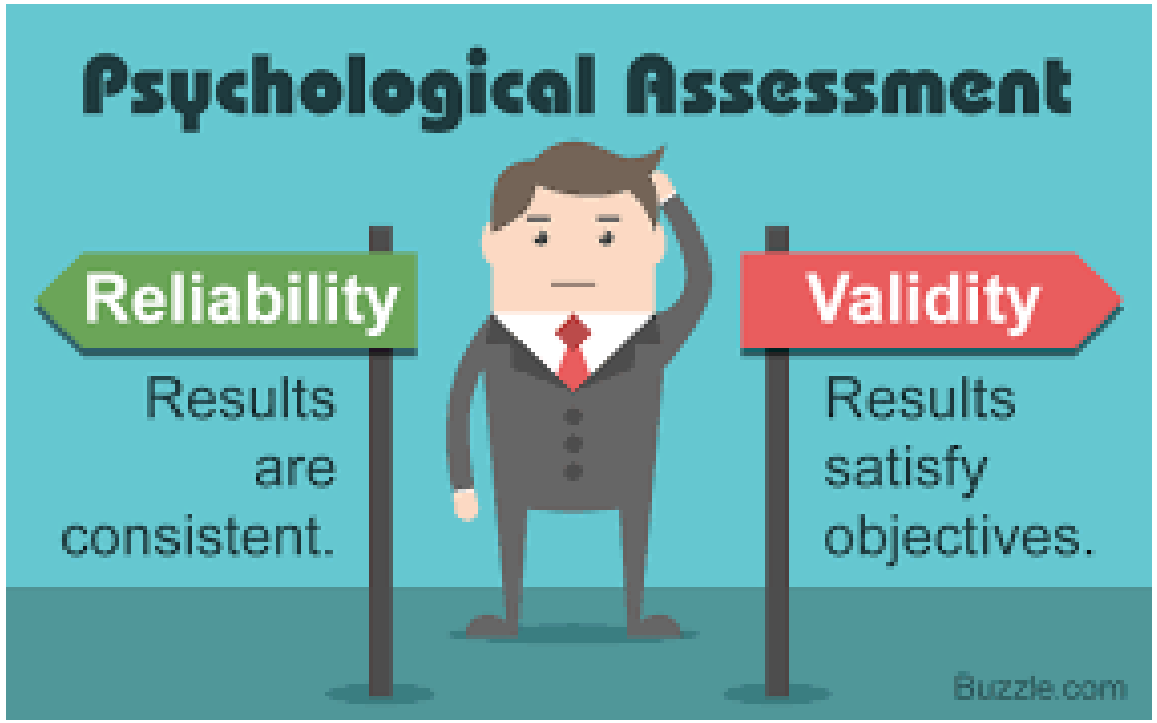
Conceptual clarity



- What is the exact nature of the concept we're trying to measure?
- Example: How should a survey question measure “income”?
  - 1 “What is your income?”
  - 2 “What is the total amount of income earned in the most recently completed tax year by you and any other adults in your household, including all sources of income?”
- Clearer the concepts, easier to measure

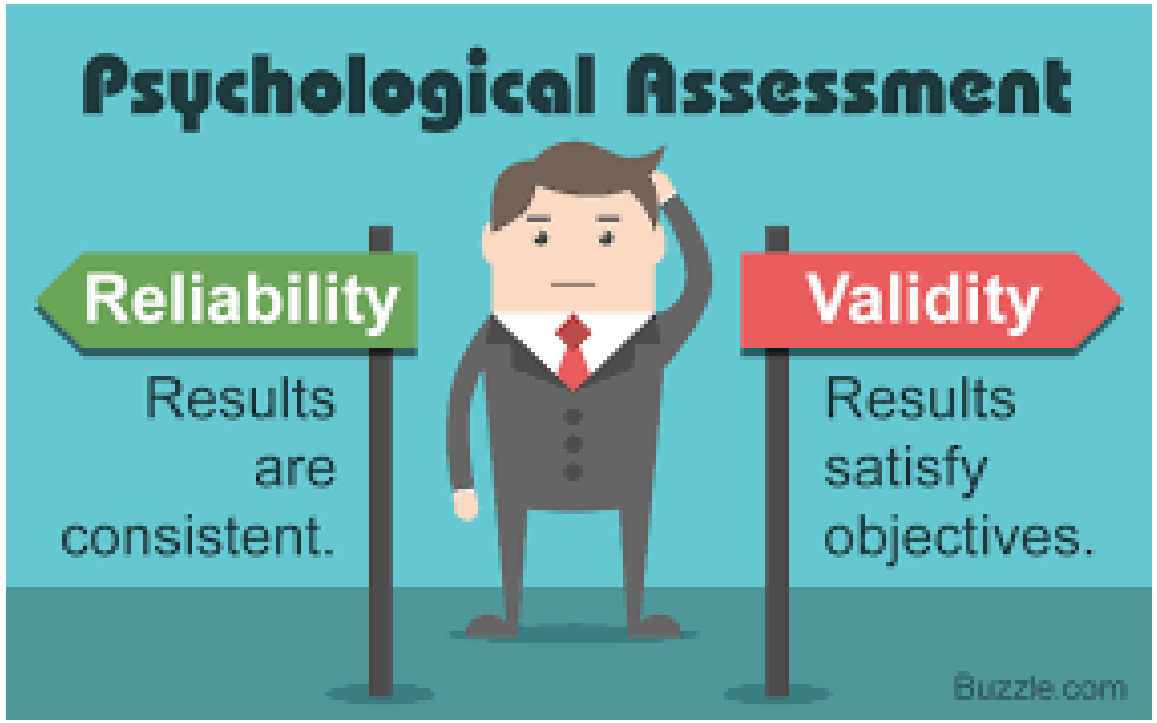
Reliability

- An operational measure is *repeatable* or *consistent*; that is, applying the *same measurement* rules to the same case or *observation* will **reproduce identical results**.



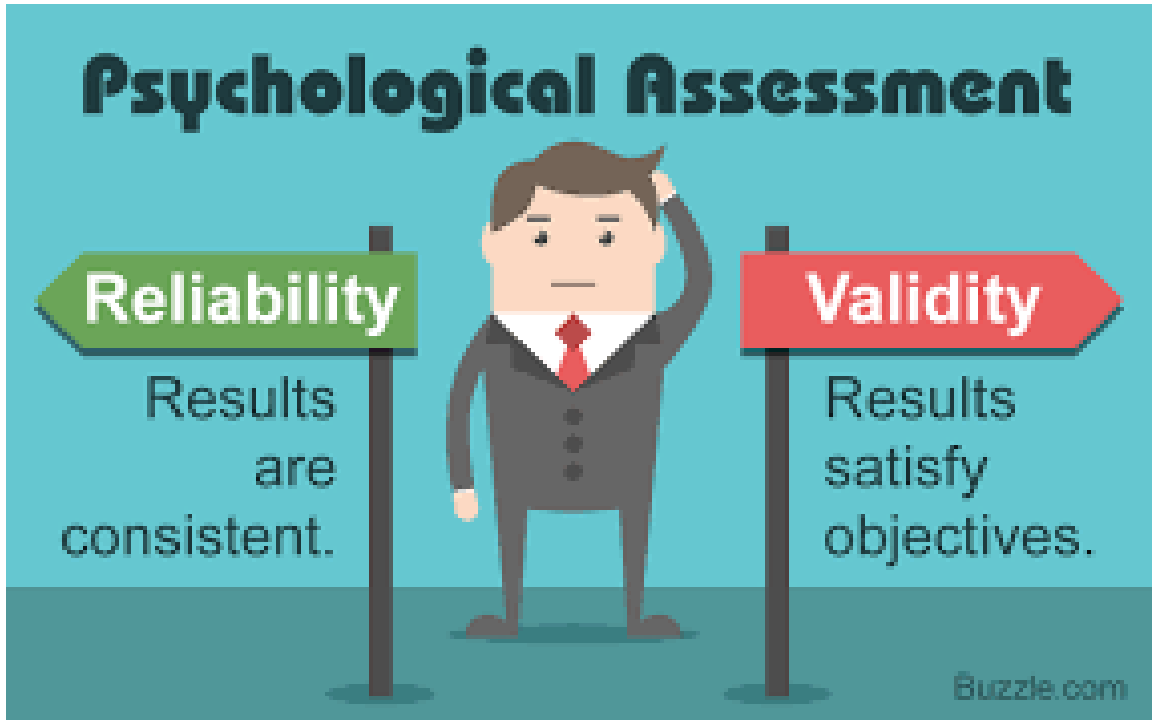
Validity

- You measure what you mean:
  - A **valid** measure accurately represents the concept that it is supposed to measure.
- There are three types of measurement validity that scientists are concerned about:
  - 1 Face validity: on its face, it makes sense
  - 2 Content validity: what are the **essential elements** of this concept
  - 3 Construct validity: if this measure is **highly correlated with other commonly used measure**
  - E.g., measuring civil war?



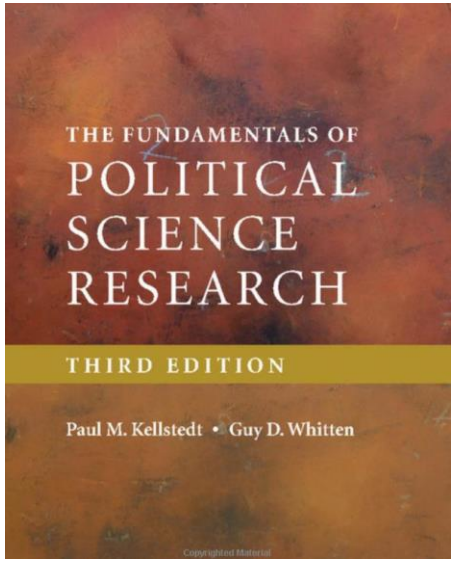
Validity

- You measure what you mean:
  - A **valid** measure accurately represents the concept that it is supposed to measure.
- There are three types of measurement validity that scientists are concerned about:
  - 1 Face validity: on its face, it makes sense
  - 2 Content validity: what are the **essential elements** of this concept
  - 3 Construct validity: if this measure is **highly correlated with other commonly used measure**
  - E.g., civil war (100 battle death)

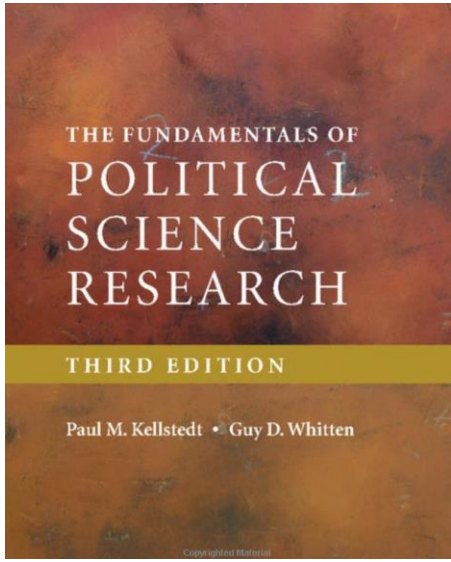


Controversy 1: Measuring democracy

- How do we measure democracy? Thoughts?



Controversy 1: Measuring democracy

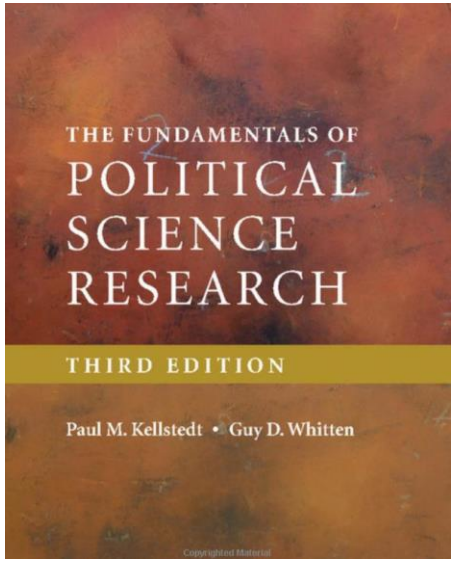


- How do we measure democracy? Thoughts?
- At the conceptual level, what does it mean to say that Country A is "more democratic" than Country B?
- Two core attributes to a democracy: "contestation" and "participation."
- Measuring democracy: the Polity IV measure.
  - It measures democracy with annual scores ranging from -10 (strongly autocratic) to +10 (strongly democratic) for every country on earth from 1800 – 2004.

Criticisms of Polity IV

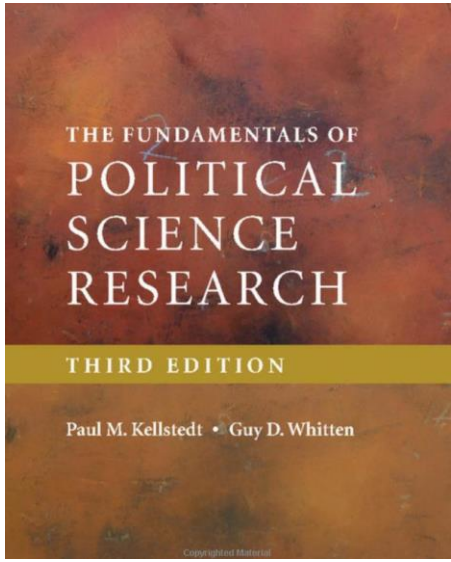
- But the Polity measure includes only "**contestation**" and lacks information about a country's level of "***participation***."
- If suffrage is expanding, our measures of democracy ought to incorporate that reality. But Polity IV measure does not fully encompass what it means, conceptually, to be more or less democratic. (This is content validity.)

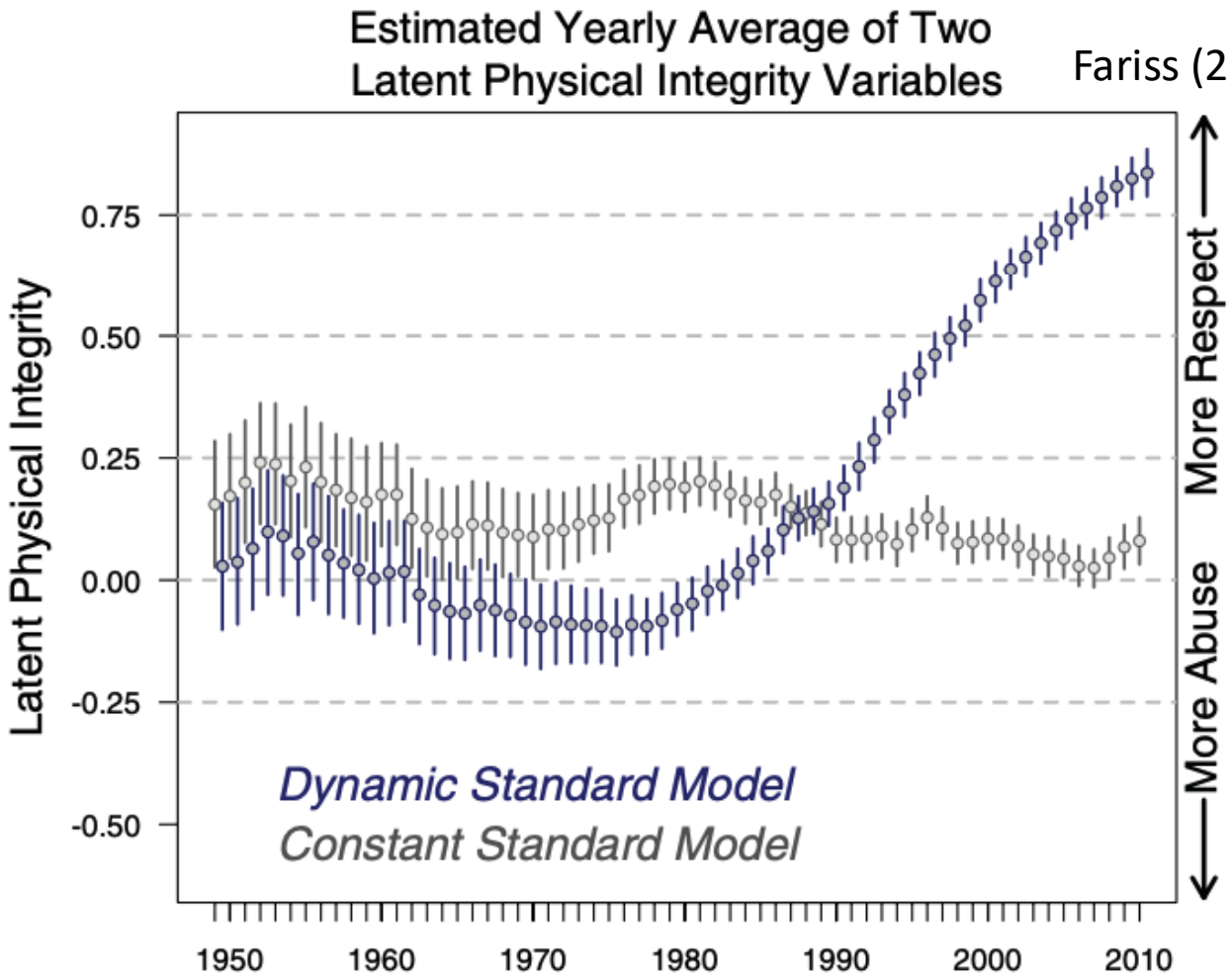
Controversy 2: Measuring human rights violations



- The data show that human rights practices have not improved over the past 35 years
- By looking into how we measure HR standards, we find that we miss a systematic change in the way monitors, like Amnesty International and the U.S. State Department, encounter and interpret information about abuses.

Changing Standards of Human Rights





Fariss (2014)

Measuring security communities in IR?

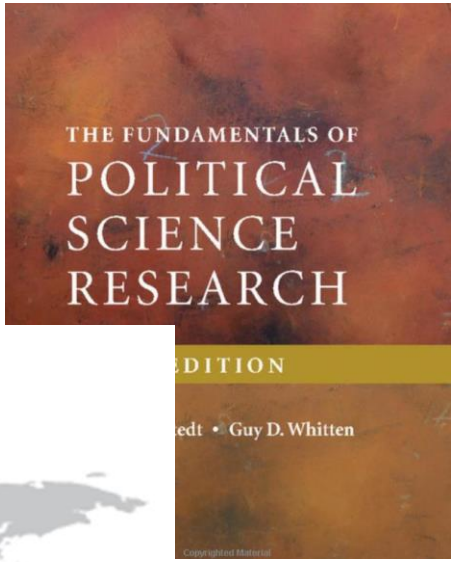
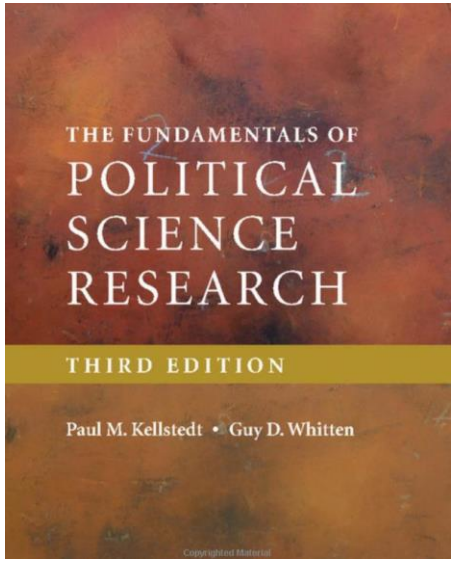


Figure 2. Visualizations of joint-production security communities in 1999 at different resolution levels. *A*, High resolution (seven communities); *B*, medium resolution (three communities); *C*, low resolution (two communities).

So what should we do?

Match your concept and data (measurement) as closely as possible

Use available data wisely and critically



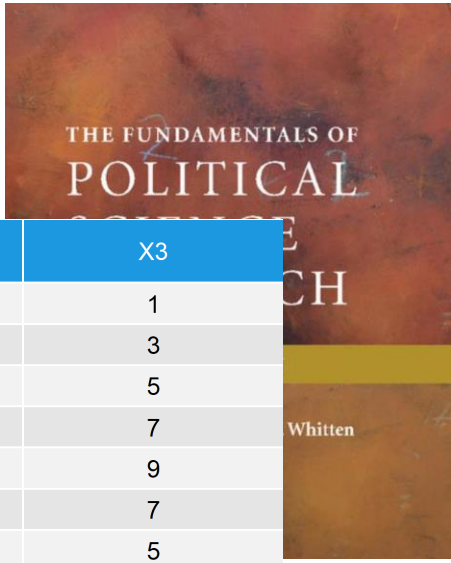
The Fundamentals of Political Science Research, 3rd edition

Chapter 6: Getting to Know Your Data

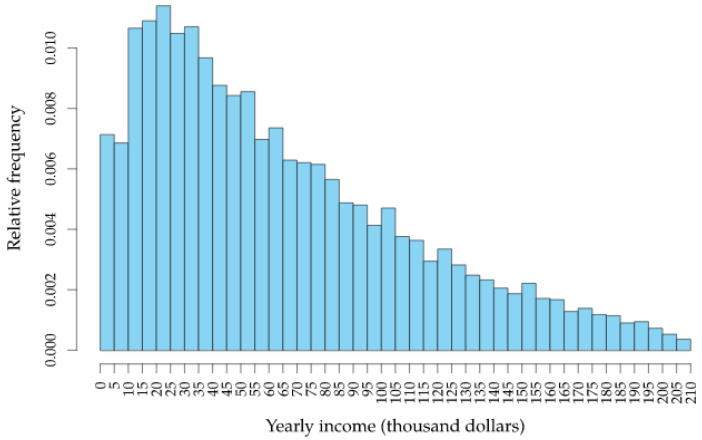
(describe your data)

What do data “look like”?

- After you load a dataset, what should you do first? Thoughts?

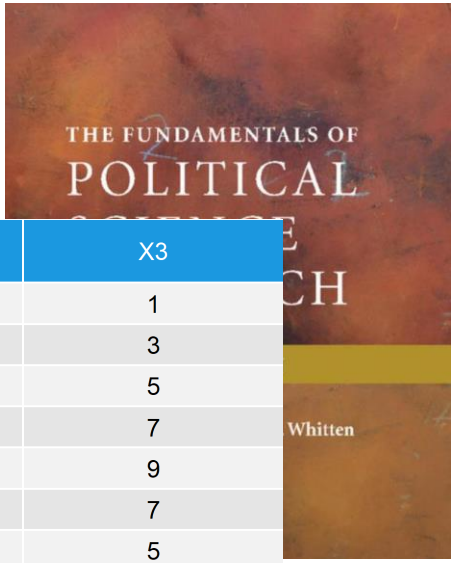


X1	X2	X3
9	NA	1
6	5	3
NA	2	5
9	1	7
2	5	9
5	8	7
NA	0	5

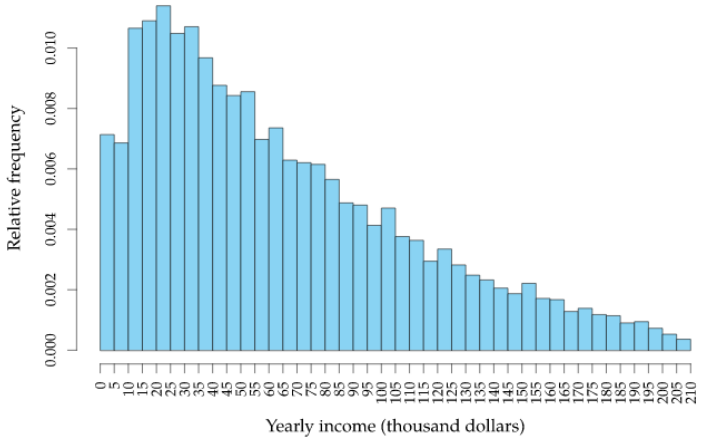


What do data “look like”?

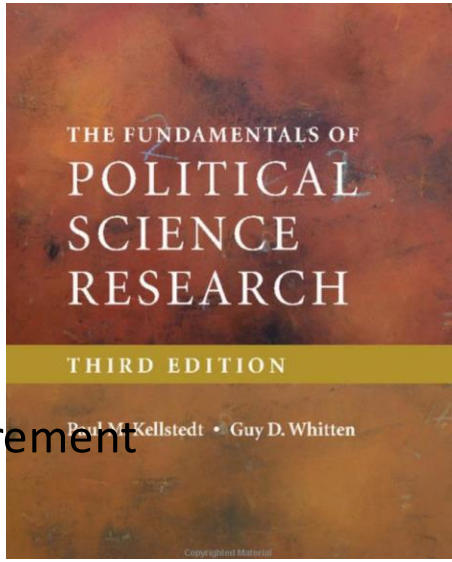
- After you load a dataset, what should you do first? Thoughts?
- Get a good idea of the *types of values* that the individual variables take on before moving on to test for causal connections between two or more variables.
  - Beware of NAs
- What do “typical” values for a variable look like?
- How tightly clustered (or widely dispersed) are these values?
  - This is the preliminary step that will help you to avoid errors



X1	X2	X3
9	NA	1
6	5	3
NA	2	5
9	1	7
2	5	9
5	8	7
NA	0	5

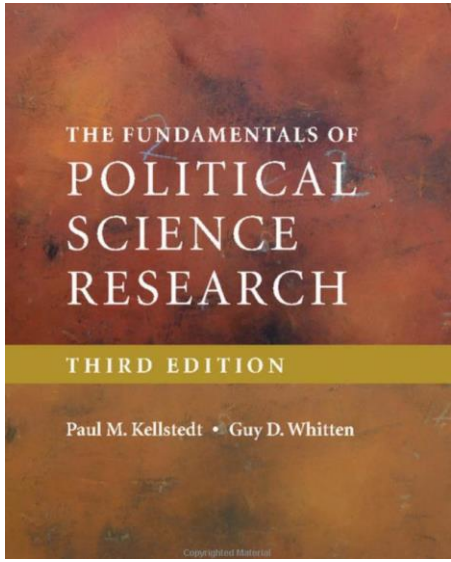


What is a variable's measurement metric?



- One major way to distinguish among variables is **measurement metric**. A variable's measurement metric is the ***type of values*** that the variable takes on
- Binary variable:
  - Gender: Male or not
  - A war outcome: Win or lose
- Categorical variables:
  - Religion: Catholic, Muslim, nonreglious → transforming them into binary
- Ordinal variables:
  - Level of repression: no repression, non-lethal repression, lethal repression
- Continuous variables:
  - Age
  - Percentage of gender inequality: 0 to 100 %

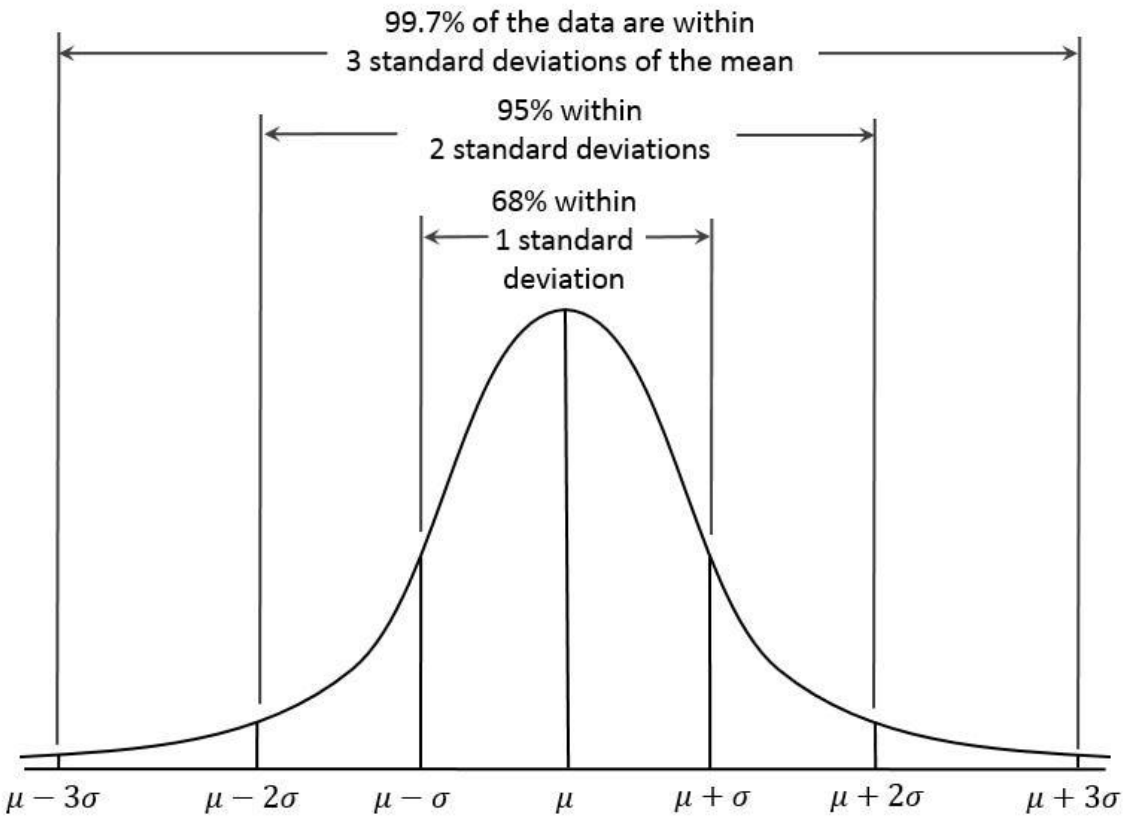
Variable statistics



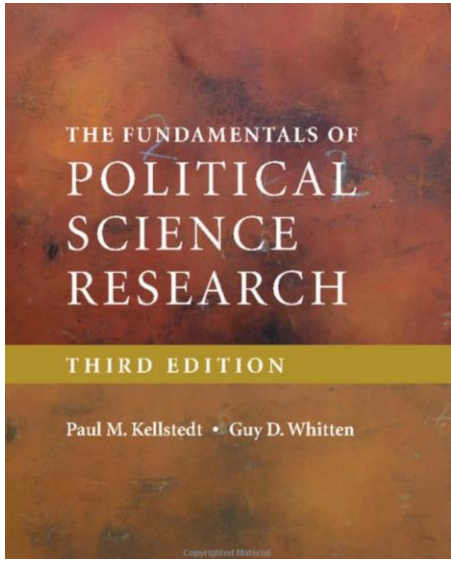
- There are two types of descriptive statistics that are most relevant in the social sciences:
  - 1. Measures of **central tendency** tell us about typical values for a particular variable
  - 2. Measures of **variation** or (dispersion) tell us the distribution (or spread, or range) of values that it takes across the cases for which we measure it.
- Importantly, for different variable types just discussed, some descriptive statistics are appropriate (and some aren't).

Measures of central tendency and variation

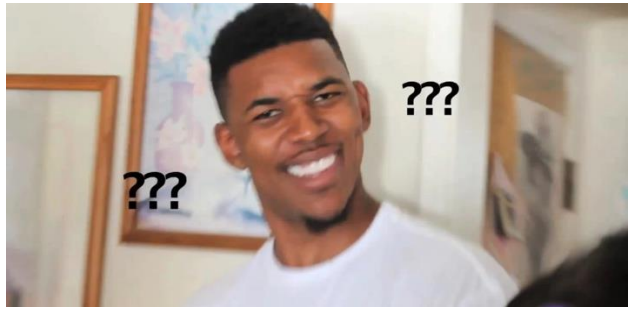
- Central tendency: Mode, median, mean
  - When should you use the median rather than the mean?
- Dispersion: Standard deviation



Describing categorical variables



- The only measure of central tendency that is appropriate for a categorical variable is the **mode**, which we define as the most frequently occurring value.
- We can't compute the mean or average
  - E.g., Male =1, Female =0 → mean = 3.5?



- The simplest thing to do is create a table, or a graph.

Show you categorical data: table

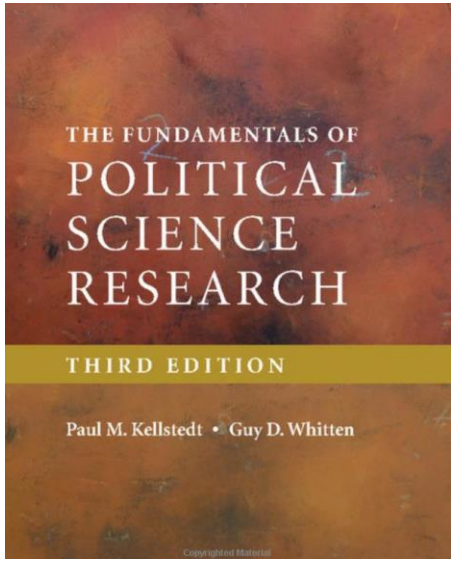
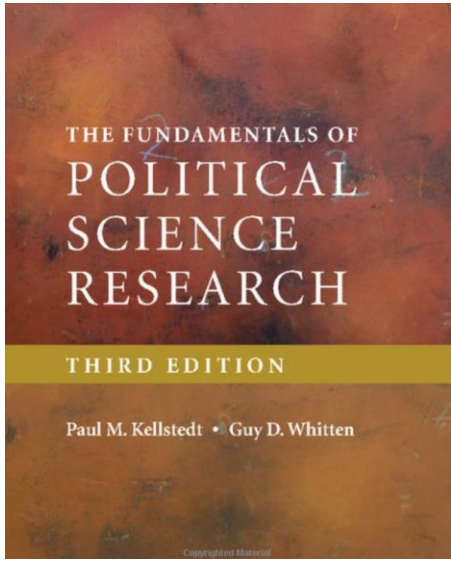


Table: Frequency table for religious identification in the 2004 ANES

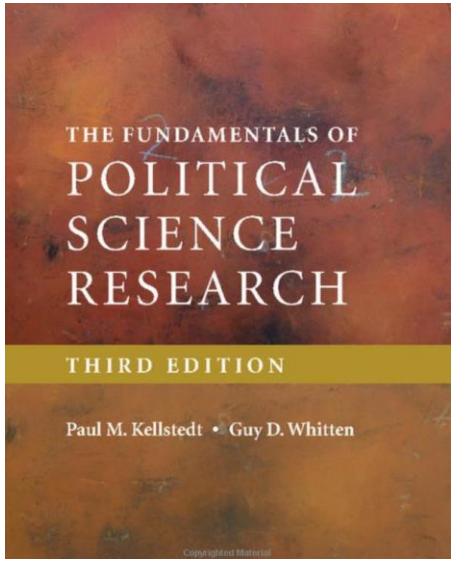
Category	Number of Cases	Percent
Protestant	672	56.14
Catholic	292	24.39
Jewish	35	2.92
Other	17	1.42
None	181	15.12

Central tendency: The median



- One measure of central tendency for continuous variables is the median value of the variable. The median value is the value of the case that sits at the ***exact center of our cases*** when we rank them from the smallest to the largest observed values.
- When we have ***an even number of cases***, we *average the value of the two center-most ranked cases* to obtain the median value. This is also known as the value of the variable at the 50 percent rank.

Central tendency: The mean

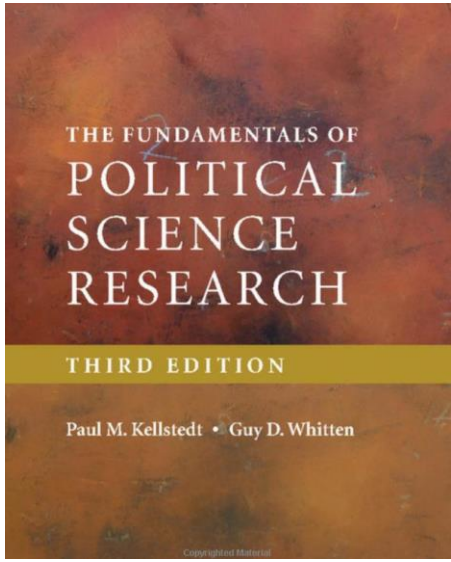


- Another measure of central tendency is the mean value or "average" value for the variable. For a variable Y, the mean value is depicted and calculated as:

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

where Y , known as "Y-bar," indicates the mean of Y , which is equal to the sum of all values of Y across individual cases of Y , Y_i , divided by the *total number of cases "n."*

Variation: The standard deviation

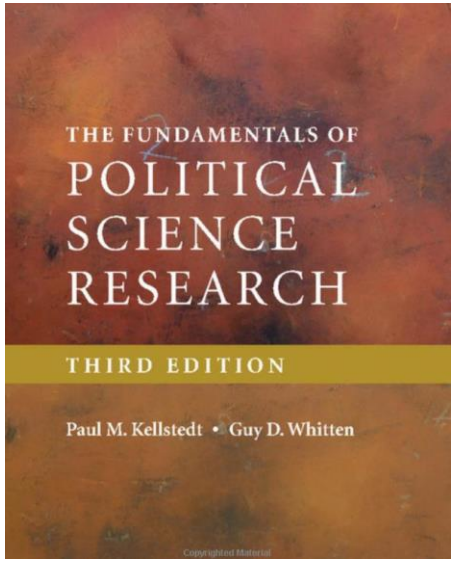


- The most intuitive measure of variance is the **standard deviation**:

$$sd(Y) = sd_Y = s_Y = \sqrt{var(Y)} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$$

Why do we want to use the sum of squared deviations?
Thoughts?

Variation: The standard deviation



- The most intuitive measure of variance is the **standard deviation**:

$$sd(Y) = sd_Y = s_Y = \sqrt{var(Y)} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$$

Roughly speaking, this is the average difference between values of Y (Y_i) and the mean of Y ( ). At first glance, this may not be apparent. But, the important thing to understand about this formula is that the purpose of squaring each difference from the mean and then taking the square root of the resulting sum of squared deviations ***is to keep the negative and positive deviations from canceling each other out.***

Show you categorical data: table

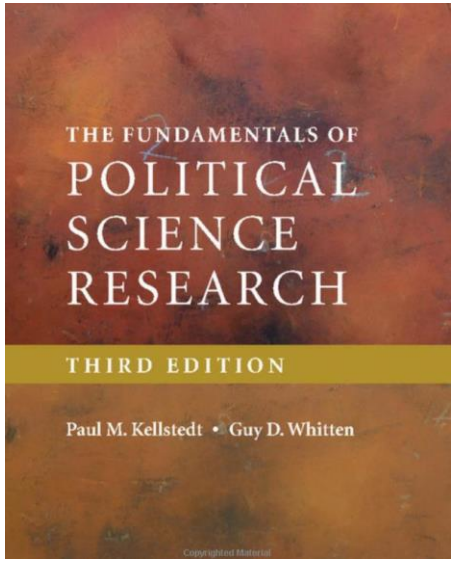
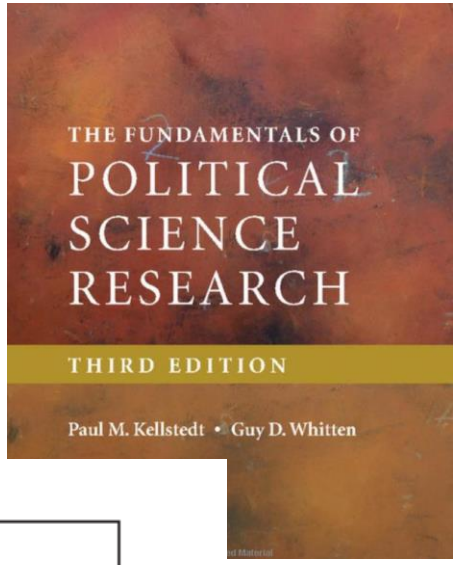


Table: Frequency table for religious identification in the 2004 ANES

Category	Number of Cases	Percent
Protestant	672	56.14
Catholic	292	24.39
Jewish	35	2.92
Other	17	1.42
None	181	15.12

Show your continuous data: histogram



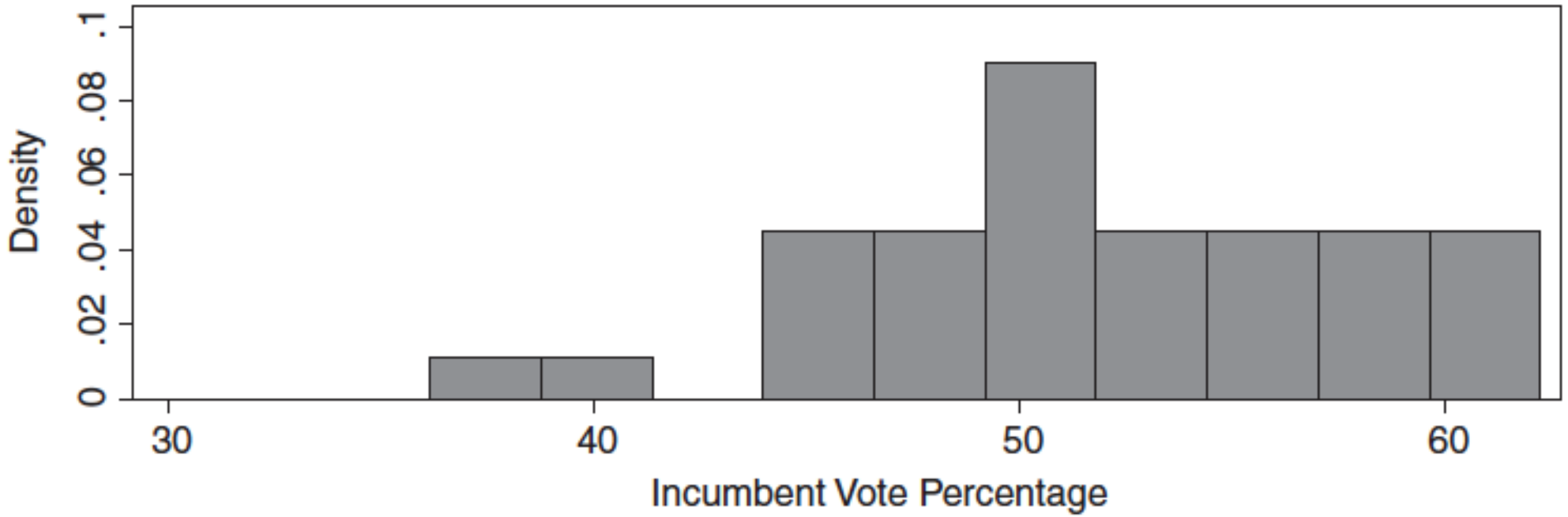


Figure 5.9. Histograms of incumbent-party presidential vote percentage, 1876–2008, depicted with 2 and then 10 blocks.

Show your cross-sectional data: map