Codebook for "Congressional Scandals (1979-2018)."

This dataset compiles information about scandals associated with sitting members of the United States House of Representatives and the United States Senate between 1979 and 2018. The dataset contains information about the member and the scandal in which they were involved.

These data were collected by Michael G. Miller of Barnard College (mgmiller@barnard.edu) and Brian Hamel of UCLA. If you publish with these data, please cite as:

Miller, Michael G., and Brian Hamel. 2021. "Congressional Scandals (1979-2018)." Harvard Dataverse, Version 2. Available at https://doi.org/10.7910/DVN/BID3QM.

Version 2 adds bioguide IDs and cleans up Members' names.

Citation information is also available on the Harvard Dataverse page.

We thank Caroline Clifford and Courtney Couillard for assistance with data collection.

Data collection procedures:

Information about scandals was compiled from a range of sources. The most important of these is:

Basinger, Scott J., Lara M. Brown, Douglas B. Harris, and Girish Gulati. 2014. "Counting and Classifying Congressional Scandals." In *Scandal: An Interdisciplinary Approach to the Consequeces, Outcomes, and Significance of Political Scandals.* Alison Dagnes and Mark Sachleben, eds. pg. 3-28.

Scholars who use these data should cite Basinger et al. as well.

Basinger et al. compiles scandals from the 93rd to 111th Congress. We retain information from these data from the 96th Congress onward. We obtained information on scandals occurring between the 112th and 115th Congress (inclusive) by following similar collection methods described in Basinger et al. We searched a range of sources, including *Congressional Quarterly, The Congressional Quarterly Almanac,* Ethics Committee reports, and newspaper archives, adding all events that might be considered a scandal under the four part classification scheme used in Basinger et al. Each scandal is defined as being a sex, financial, political, or other type of scandal. See Basinger et al. for a description of this coding scheme. We retained this system for the scandals we added.

We also added two fields for all scandals in the data frame, no matter when they occurred. The first is the scandal "break date," which is the first instance we could find of the scandal being reported in a national or local news source. The second is a measure of scandal salience, which is the number of times the scandal was mentioned in the *New York Times* in the thirty days following the break date.

If you discover errors or omissions in these data, please contact Michael G. Miller at mgmiller@barnard.edu.

Note: The unit of analysis in these data is the scandal-politician. As such, members involved in more than one scandal may appear multiple times in the data frame. Similarly, scandals involving more than one member (such as the House Banking Scandal) will appear multiple times in the data frame.

Variables:

bioguide: String containing the Member ID from the Biographical Directory of the United States Congress.

name: String containing the name of the politician involved in this scandal.

house: Binary variable indicating whether the politician is a member of the U.S House of Representatives (1) or the U.S. Senate (0)

congress: Nominal numeric variable indicating the Congress in which the scandal occurred.

state: String containing the postal abbreviation of the state represented by this politician in Congress.

cd: Nominal numeric variable indicating the district represented by this politician in Congress. For Senators, this value is missing.

party: The party of which this politician is a member.
D=Democratic
R=Republican

breakdate: The first date the scandal was reported in media, in M/D/Y format.

salience: Numeric variable indicating the number of times the scandal is mentioned in the *New York Times* in the 30-day period following the break date.
Min=0
Max=200

court_outcome: String containing basic information about the scandal's treatment in the civil and/or criminal justice system.

political_outcome: String containing basic information about how the scandal affected the Member politically.

type: String classifying the scandal within the 4-party typology described in Basinger et al.: Financial, Sex, Political, or Other.

notes: String containing other contextual information and/or details that may be helpful.