**990 Mission Stemmer, Version 1**

(Paxton, Velasco, and Ressler 2019)

***Data File:***

A list of stems designed to aid pre-processing of Form 990 and Form 990EZ Mission Statements (Part III) and Primary Exempt Purposes (Part I). Saved in .csv format as **Mission\_Stemmer\_PaxtonVelascoRessler\_v1.csv**.

***Citation:***

When using the stemmer, ***please cite the correct*** ***version*** to aid replication:

Paxton, Pamela, Kristopher Velasco, and Robert Ressler. 2019. Mission Stemmer v.1. [Computer file]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]

***Coding Description:***

e-filed 990 and 990EZ mission statements (Part III) and primary exempt purposes (Part I) released by the IRS through Amazon Web Services contain inflections and derivations of words that could be stemmed to aid analysis. The 990 Mission Stemmer identifies and suggests stems for common words used in nonprofit mission statements.

Common stemming algorithms such as the Porter algorithm are ‘strong’ and reduce words in ways that sometimes conflate meanings in the nonprofit realm. For example, many nonprofits may refer to fundraising as *advancement* which should not be stemmed to “advanc” along with *advance*. In the third sector we may wish to distinguish *associates* from *associations* from *associated* nonprofits. Or *equal* from *equality*. *Managing* vs. *managers* vs. *management* isanother example where our stemmer would differ from Porter which stems all to “manag.” On the other hand, Porter would stem *christ* and *christian* separately while our stemmer collapses them under the stem “christ.” Even simple suffix stripping can conflate words that those working with nonprofits may wish to separate – such as *operate* and *operated*. Or *house* and *housing*.

Each research project is unique. But this stemmer can form a basis on which a researcher can build, depending on their research question. Or, it can be used in conjunction with another stemmer, such as the Porter algorithm, if a researcher agrees with these suggested overrides.

To create the stemmer, all words of frequency 100 or higher in mission statements from 2011-2015 (3703 words) were inspected by a team of three individuals. Any suggested stems were approved by at least one additional coder. Our goal was to reduce inflected or derived words to a root stem but also to maintain differences that are relevant to the third sector.

The file reproduces these 3703 words suggests 1668 stems.

Note that these corrections apply to mission statements where words are lowercase and properly spelled.

***List of Variables:***

word original word

stem suggested stem

***Suggestions:***

We welcome suggestions for changes or additional stems. Please use the form available at www.pamelapaxton.com/990-mission-stemmer-suggestion-form to contribute your suggestions for version 2 of the stemmer.

***Additional Questions:***

For further questions or concerns regarding the use of this data, contact:

Dr. Pamela Paxton

Department of Sociology and School of Public Affairs

University of Texas at Austin

305 E 23rd St, A1700

Austin, TX 78712-1086

Phone: (512) 232-6323

email: ppaxton@prc.utexas.edu