## Methods Used for Analysis of Responses to the 2020 Survey about the SAA’s Principle of Archaeological Ethics

Ben Marwick (corresponding author, bmarwick@uw.edu)

Kaylee Pruski

Raelee Hampton

Eloise Potter

Liying Wang

The survey used several response types, often with multiple types used for a single question. We used a variety of analysis and visualization methods depending on the response type. The R code and data for the analysis reported here are available in a research compendium [(Marwick et al., 2018)](https://www.zotero.org/google-docs/?6BzVLJ) at <http://doi.org/10.17605/OSF.IO/643C8>. To comply with the EU’s General Data Protection Regulation (GDPR) we do not include any responses to the demographic questions in our compendium because these are ‘personal data’. Personal data is information relating to an identified or identifiable person, for example someone who can be identified, directly or indirectly, in particular by reference to a data value.

Single option response

Only one option was allowed, e.g. yes/no. We visualise these responses with a simple ordered bar plot that shows the frequencies of each response.

Multiple responses

One or more options could be selected, e.g. “I am satisfied with the format of the current Principles” / “An aspirational code” / “A living document” / etc. We used an UpSet plot for basic visualisation of these responses. This is an alternative to a Venn Diagram that plots the intersections of a set as a matrix [(Lex et al., 2014)](https://www.zotero.org/google-docs/?WVCsSt).

Five-level Likert item response

Where respondents select one item to indicate their level of agreement or disagreement on a symmetrical agree-disagree scale, e.g. Strongly disagree / Disagree / Neutral / Agree / Strongly agree. We used a diverging stacked proportional bar plot to visualise responses to questions with this Likert-type scale. Although this type of response is widely used in several disciplines, there is disagreement about the most suitable statistical methods for hypothesis testing.

To summarise variation in responses across the demographic categories, we computed “mean disagreement” values for each group by re-scaling the Likert type responses from 1 (strongly agree) to 5 (strongly disagree) and computing the mean within each group. To test for non-random differences between groups, we used the Kruskal-Wallis rank sum test. The test results are summarised on the plots accompanying each survey question, with solid data points indicating a non-random difference, and hollow data points indicating no difference in mean disagreement between the groups. Because Likert type responses may violate assumptions for comparing means or ranks, we also show all the individual data points on the plots to aid in interpretation. To further guide interpretation of the plots, the data point sizes are scaled to the number of respondents in each group.

Free text

Where the respondent can enter any text. e.g. a single word, sentence or paragraphs), or leave the field blank. The amount of text submitted varied greatly between questions and respondents (Figure 1). We take a computational social science approach that combines social science qualitative methods with computer and data science methods (Lazer et al. 2009; Salganik 2019).

We used two sets of methods for analysing the free text: traditional qualitative data analysis (QDA) and more novel exploratory data analysis (EDA) using computational methods.

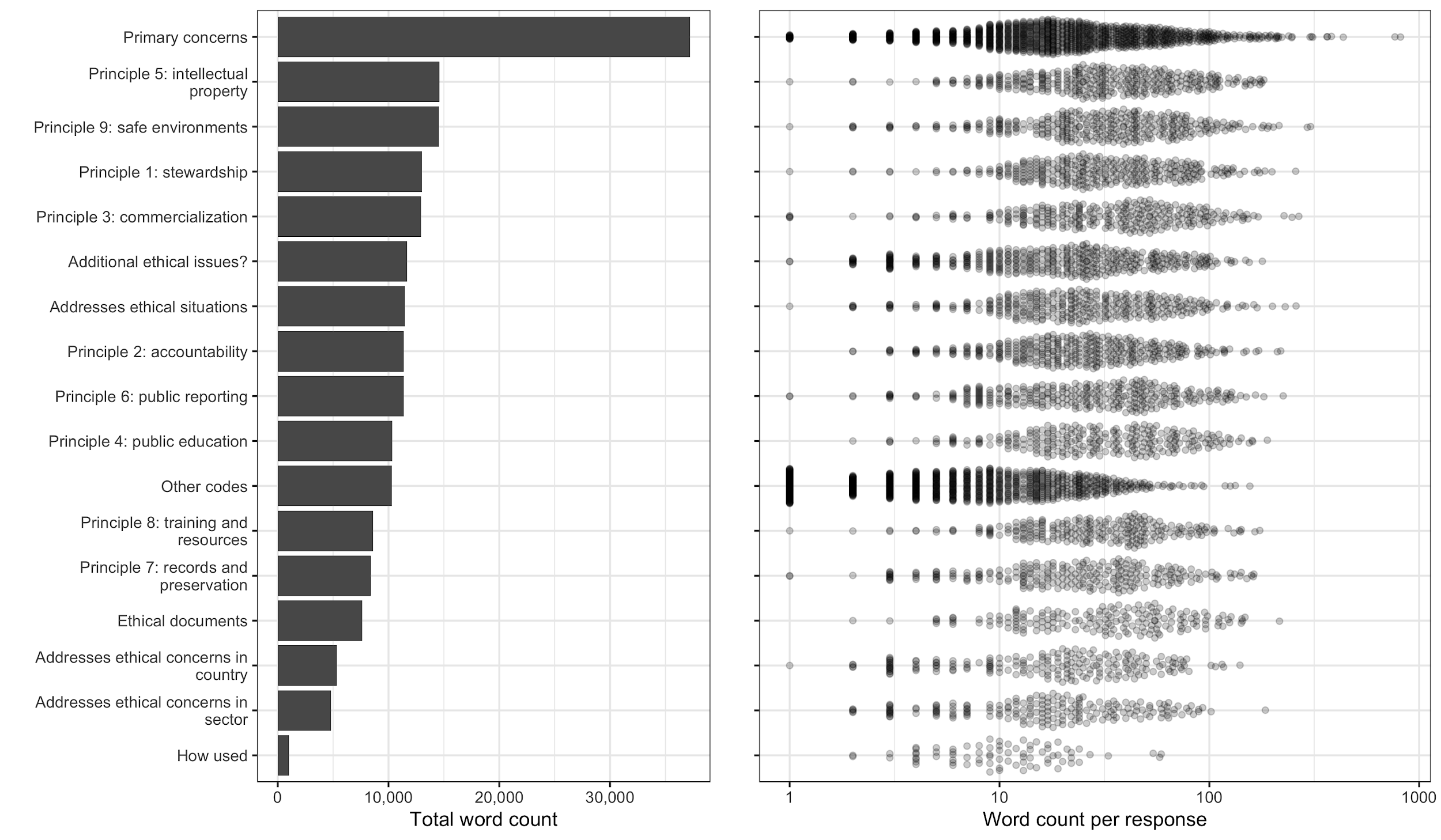


Figure 1. Word count for questions with free text responses.

Qualitative Data Analysis

Surveys that include open‐ended, free‐text responses are often analyzed using qualitative methods (Baumer et al. 2013; Rader, Wash, and Brooks 2012). Analytic methods from grounded theory [(Charmaz, 2006)](https://www.zotero.org/google-docs/?A2DNeY) can generate rich, thick descriptions (Geertz 1973). Our QDA used a grounded-theory approach, and followed typical three steps for text analysis [(Bernard & Ryan, 1998)](https://www.zotero.org/google-docs/?lLc9M4). First we read over all the free text responses to immerse ourselves in the content to infer key recurring themes. We also examined the results of the EDA to inform our thematic analysis. Second, we organised these themes into a small set of tags or codes to classify the key concepts among the responses (Table 1). Third, we studied all the free text responses and applied tags to identify our key themes among the responses. We then analysed the tagged text to identify the relative importance of themes, relationships between themes, and identify examples to summarise the themes. We report on the results of our tagging by organising the data into *Commonalities* (what are the most common tags across many responses?), *Contrasts* (Contrasts: what tags never or rarely appear together; what are examples where one tag has different types of responses?), *Comparisons* (what tags frequently appear together?), *Relationships* (Relationships: do some tags seem to have a cause and effect relationship?), and *Sentiment* (are there positive and negative reactions among the tagged text?).

Table 1. Tags and their descriptions used for the QDA

|  |  |
| --- | --- |
| **Tag** | **Description** |
| Indigenous | also First Nations, descendant communities, rights, involvement, communities, traditional knowledge, consultation, patrimony, decolonising |
| Repatriation | also returning artefacts to Indigenous communities |
| Human remains | also bioarchaeology, human bones |
| Museums | also anything about museums |
| Looting | also illegal antiquities dealing, private collections, hoarding, damage to sites |
| Avocational | also treasure hunting, metal detecting |
| Archiving | also long term preservation of finds, curation, controlling access to data and objects |
| Public engagement | engaging with, communicating with, training members of the public |
| Laws | also state and federal legislation, compliance with laws and policy |
| Relevance | also things like ‘not relevant to me/my sector’, outdated, needs to be updated, |
| Enforcement | also not enforced, need for a high standard, more ‘teeth’, need to be backed up, lack of consequences, needs to be stronger |
| IP and copyright | also intellectual property, data sovereignty, authorship |
| Plagiarism | also lying, dishonesty, stealing work of others |
| Field methods | also sampling |
| Conflicts of interest | profit over quality, protection of the archaeological record, competition |
| Training students | also teaching, experience, quality of education, acquiring professional skills, job preparation |
| Mentoring | also professional relationships between advisors and students |
| Power dynamics | also power imbalances, exploitation of students and junior workers, abuse of power |
| Diversity | also people of diverse backgrounds, also improving access to training |
| Research | also PhD thesis, dissertations, journal articles, books |
| Open science | also open data and open access, information sharing, digital data, transparency, and reproducibility |
| Harassment | also bigotry, abuse, misogyny |
| Sexual harassment | also 2019 SAA, David Yesner, #MeToo |
| Racism | also Black, Latinx communities, exclusion, under-representation |
| Discrimination | age, gender, ethnicity, career stage/type (CRM vs academia) |
| Physical safety | also workplace safety, field work or lab work |
| Reporting | also whistleblowing, making complaints |
| Money | also pay, salary, income, financial concerns, budgeting, job security |
| Collegiality | also collegial and supportive environments, behavior toward other professionals, colleagues |
| Environment | also climate change, fossil fuels, plastics, |

For the question about how the Principles were used, we developed a separate set of tags:

Table 2. Tags used for the question about how the Principles were used.

|  |  |
| --- | --- |
| **Tag** | **Description** |
| Research | to compare and critique with other org or with SAA’s actions |
| Teaching | to show students or prepare for a class |
| Derive | to develop/adopt a code for another organisation |
| Assessment | to guide own behaviour or other people’s behaviour |
| Implement | to guide employee or members of an organisation (without adapting it, just using as-is) |

Exploratory Data Analysis

To complement our QDA, we employed a suite of computational text analysis methods to summarise the free text responses and automatically extract topics. The computational methods are useful for the descriptive analysis of unstructured text data sets, and can provide robust confirmation of qualitative data analysis results. Compared to the grounded-theory approach of QDA, our computational methods are ‘theory-free’, and the results can be independently reproduced by any researcher, regardless of the biases and motivations they bring to their study.

To prepare the text for EDA methods we used the quanteda package for R [(Benoit et al., 2018)](https://www.zotero.org/google-docs/?S3nxlh) to remove stop-words (e.g. ‘the’, ‘and’, etc.) and convert the text into a document-feature matrix (a table where each row is a document and each column is a word, and the cell values are word counts per document). After these preparatory steps we used two techniques for analysis and visualisation.

First, are network plots that show how frequently-used words co-occur with each other. This shows the most frequent words, and also adds additional information in the form of a visual connection between words that often co-occur.

Second are topic models that show computational models of topics present in the responses. As they usually do not make use of human‐curated notions of the topics in a document but instead derive them purely from the data, topic models are often classified under unsupervised machine learning methods. A topic model indicates which topics are most abundant in the responses. A topic model is a statistical analysis of text that assumes there are an arbitrary number (k) topics present in all the responses (we have to choose this number in advance, we use methods to estimate it). It assumes that every unique word in the text has a certain probability that it will appear in each topic. Computing these probability distributions is an iterative process, so each time we run the code we get slightly different combinations of words in the topics. A ‘topic’ then, is a distribution of all the unique words in the text, and one topic is different from another topic by the probability distributions of the unique words. One word can have a high probability of appearing in multiple topics, that’s why we see some duplication of words in different topics. The highest probability words for each topic are shown in the bar plot. Some of our topics are a bit ‘noisy’, and not highly exclusive, distinct, or easily interpretable, this might be improved with some further tuning. This probabilistic approach is an advantage over the methods used for frequency-based methods (such as the word cloud), which are simply based on counts on the unique words, and so only allows a word to have a single meaning.

Topic modelling of survey results has been used in political science (Roberts, Stewart, and Tingley 2016), but as far as we know this is the first application of topic modelling to the results of a survey of archaeologists. Our use of topic models is motivated by Baumer, Mimno, Guha, Quan, and Gay [(Baumer et al., 2017)](https://www.zotero.org/google-docs/?T0BrYA) who found that topic modelling is surprisingly similar to grounded theory-based QDA, demonstrating how similar results could be obtained by applying these two methods to the same research problem.

We computed structural topic models (STM) for some questions with free text responses using the stm package for R [(Roberts, Stewart, Tingley, & others, 2014)](https://www.zotero.org/google-docs/?F3G2V7). Structural topic models use the conventional machine learning algorithm, Latent Dirichlet Allocation, for discovering topics automatically, and also incorporate metadata, such as demographic variables [(Roberts, Stewart, Tingley, Lucas, et al., 2014)](https://www.zotero.org/google-docs/?11vi0z). This enables us to make credible inferences about the effect of covariates such as age and ethnicity on the content of open-ended responses. We summarise these covariate relationships with plots that show estimates of a regression where the free text responses are the units, the outcome is the proportion of each response about a topic in an STM model, and the covariates are demographic metadata associated with each response. This allows us to compute a conditional expectation of topic prevalence given the demographic characteristics of the respondents, and incorporate estimation uncertainty in the visualisation. If the error bar for a data point on these visualizations does not include zero, we consider that it indicates a non-random relationship.

## References

[Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., & Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, *68*(6), 1397–1410. https://doi.org/10.1002/asi.23786](https://www.zotero.org/google-docs/?4FgupJ)

[Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774.](https://www.zotero.org/google-docs/?4FgupJ)

[Bernard, H. R., & Ryan, G. (1998). Text Analysis: Qualitative and Quantitative Methods. In H. R. Bernard (Ed.), *Handbook of Research Methods in Cultural Anthropology* (pp. 595–646). AltaMira Press.](https://www.zotero.org/google-docs/?4FgupJ)

[Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.](https://www.zotero.org/google-docs/?4FgupJ)

[Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248](https://www.zotero.org/google-docs/?4FgupJ)

[Marwick, B., Boettiger, C., & Mullen, L. (2018). Packaging data analytical work reproducibly using r (and friends). *The American Statistician*, *72*(1), 80–88. https://doi.org/10.1080/00031305.2017.1375986](https://www.zotero.org/google-docs/?4FgupJ)

[Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082.](https://www.zotero.org/google-docs/?4FgupJ)

[Roberts, M. E., Stewart, B. M., Tingley, D., & others. (2014). stm: R package for structural topic models. *Journal of Statistical Software*, *10*(2), 1–40.](https://www.zotero.org/google-docs/?4FgupJ)