

A Quantitative Approach to Developing Cartographic Design Guidelines

Aileen R. Buckley

Esri, Redlands, California, USA
abuckley@esri.com

Keywords: cartography, design, text mining and analysis

Introduction

The purpose of this research was to develop a concise set of widely accepted guidelines for basic cartographic design. A set of map design and compilation guidelines would be useful for many purposes. For example, teachers could use the guidelines to help students learn about fundamental cartographic principles. Self-learners could use the guidelines to teach themselves about map making. Organizations and agencies could use the guidelines to help their employees, especially those not familiar with cartography or primarily involved in map making, to create cartographically defensible maps. In fact, the U.S. Census Bureau and Statistics New Zealand (2014) are examples of agencies that have developed such a set of guidelines.

While cartographic textbooks generally include many guidelines, this source is often excessive in content and non-prescriptive in style. Other sources, such as book chapters, articles, or blog posts, are often focused on isolated aspects or elements of map making or design, such as symbols (Akella 2008), map projections (Basaraner and Cetinkaya 2019), or labels (Liao et al. 2019), for example. These sources might instead target a specific map style, such as topographic maps (Raposo and Brewer 2014), climate change maps (McKendry and Machlis 2009), or flood risk maps (Fuchs, Spachinger, and Dorner 2009). Consequently, there are few examples of a set of universally accepted guidelines that provide a foundation for general decision making in map design and compilation.

Compilation of the guidelines should be based on a number of objectives. So that the guidelines cover the wide scope of map making and design, both map compilation and page layout should be considered. So that the guidelines are widely accepted, input from many sources should be gathered. So that the guidelines are authoritative, input should come from expert sources. So that the guidelines are objective and impartial, analytical methods should be used to find structure and meaning in the gathered input.

This paper details an approach used to gather input from many expert sources about important considerations for map design and compilation, perform a quantitative analysis of the input provided, and organize the analysis results in a manner that reflects the practicalities of map making and design.

Methods

Through an informal email from the author in 2018, over 200 cartographers worldwide from academia, government, and the private sector were asked to share ten things they think of as signs that a map was made by someone who is a novice (figure 1). These

signs can be thought of as typical cartographic blunders made by people who are still learning or perhaps have failed to learn the best or correct ways to make maps. The respondents were not asked to rank the items in the list, nor were they required to provide exactly ten items in the list—in fact, some offered more than ten items, and some offered fewer.



Figure 1. Locations of the origins of the email respondents.

These signs can help map makers to know what pitfalls to avoid when making maps, they can help educators to know what to teach when instructing others how to make maps, and they can help map readers to know what to look for to determine the veracity of a map. They can, in a sense, reveal secrets about the ability of the map maker. And they can be used to compile a set of guidelines for basic cartographic design.

Responses

A total of 216 emails were sent out, and 114 responses were received, of which six responses could not be used. For example, one response declined participation, one simply contained a link to a website, and another focused on map use rather than design. The 108 email responses that could be used were compiled into a single large and structured set of texts, called a “text corpus”, in an Excel spreadsheet.

Text Mining and Analysis

An analytical text mining approach with specialized software and computational methods was used to discover useful structures and meaning in the corpus. Data mining on text data, such as documents, emails, or webpages, involves conversion of the corpus into document vectors such that each unique word, called a “token”, is considered an attribute (Kotu and Deshpande 2014). Standard data mining tasks, such as classification and clustering, can then be applied.

Text mining and analysis is an area of research that uses techniques from well-established scientific fields such as “data mining, natural language processing, case-based reasoning, statistics, machine learning, information retrieval and knowledge management” (Verma et al. 2014). It is being used in such diverse application areas as

financial services; retail and customer products; telecommunications and media; and manufacturing, construction, and electronics (Mierswa 2014).

RapidMiner, a world-leading open-source framework for information mining, was used as the data mining and analysis engine to load, preprocess, visualize, and analyze the data (i.e., text corpus). RapidMiner provides a GUI to design and execute analytical workflows called “processes” which consist of multiple “operators” and require a number of “parameters” (figure 1). Each operator performs a single task within the workflow, and the output of each operator provides the input for the next one. Parameters further specify how the corpus or analysis will be restricted or directed. RapidMiner also provides learning schemes, models, and algorithms from extensions. Two specialized text mining extensions—Aylien and Rosette—were used in this research.

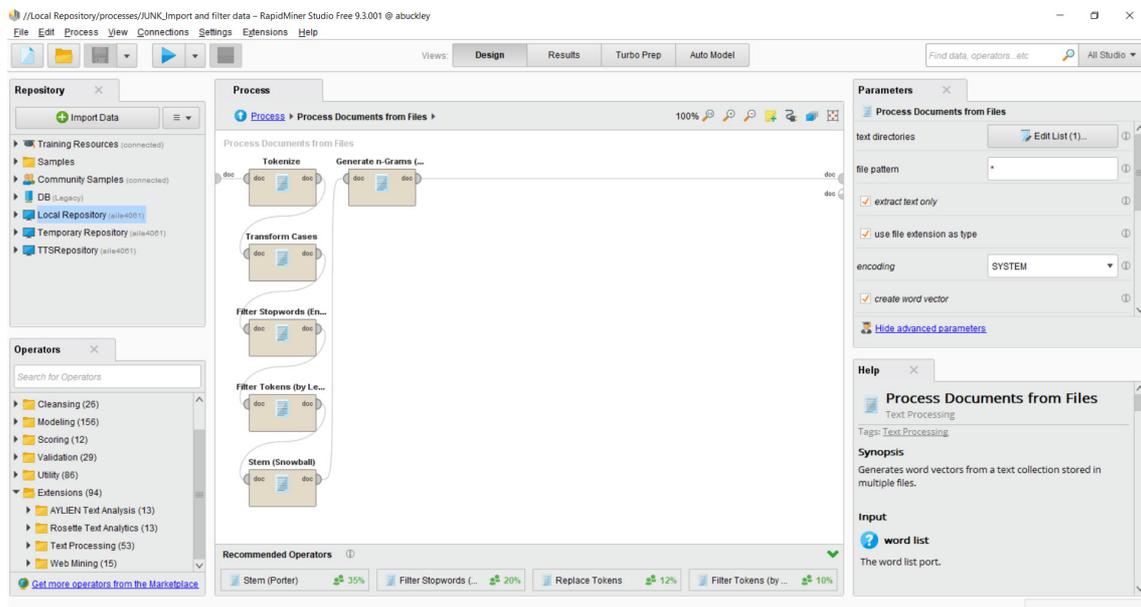


Figure 1. The RapidMiner GUI showing the operators use for text preprocessing.

After loading the text, the first step in the analysis was to preprocess the data. This involved use of the operators and parameters shown in table 1 and in the GUI of figure 1.

Tokenize	An operator that breaks a stream of text into words, phrases, symbols, or other meaningful elements to produce the token].
Transform Cases	An operator used to transform all text to a single case so that upper case and lower case words are recognized and analyzed similarly.
Filter Stopwords	An operator that eliminates very common words, such as, determiners, prepositions conjunctions, and pronouns, (e.g., “the”, “to”, “and”, “him”) which do not add to the analysis.
Filter Tokens	An operator used to filter tokens by such things as length, content, or language. In this research, the tokens were filtered by length (i.e., the number of characters they contained) so that the minimum length of a word was two characters.
Stem	An operator that reduces tokens to the root or “stem” of the word by removing suffixes.

Generate n-Grams	An operator used to find phrases. A 2-gram would be two tokens in a row, a 3 gram would be 3 tokens in a row, and so on. In this research, phrases up to and including three tokens were generated.
Prune	A parameter that removes from the analysis any tokens that occurred less than a specified number of times in the corpus. In this research, any tokens that occurred two or fewer times were removed.

Table 1. RapidMiner analysis processes and parameters used to preprocess the text corpus.

The preprocessed data was then analysed to find the most common tokens, combinations of tokens, and associated tokens. This involved a number of processes in RapidMiner, the results of which are described below.

Results

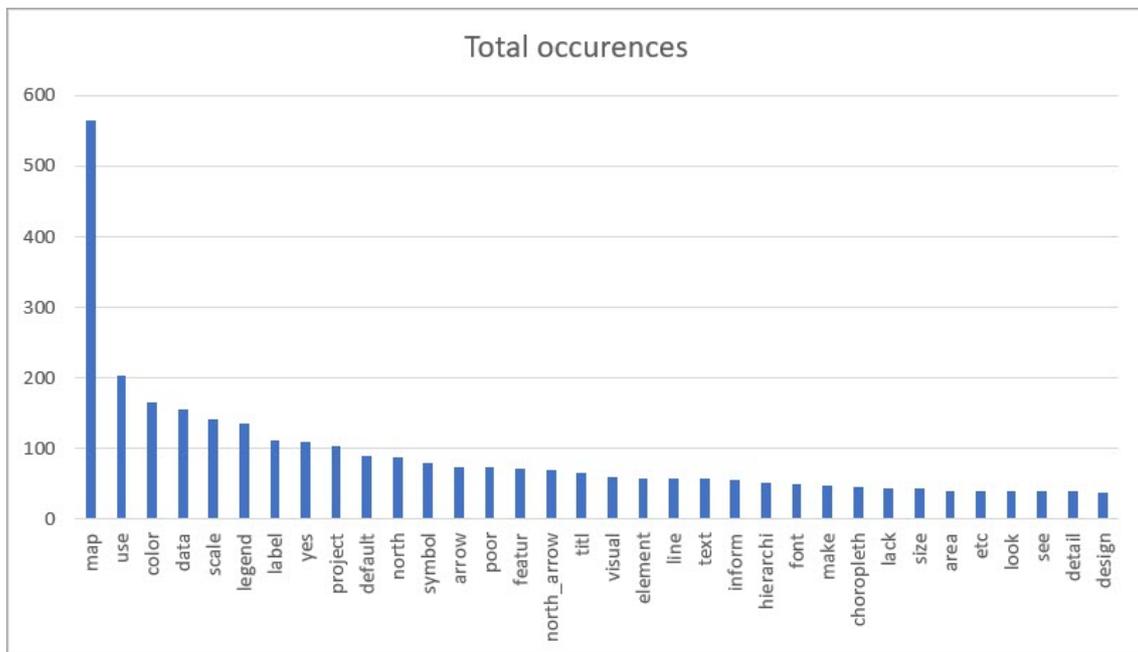
The results of the Word Count process in RapidMiner identified the tokens in the corpus that occurred with the most frequency (table 2). The list of the 35 most common tokens was examined, and 22 words were selected for further analysis since they were recognized by the author as keywords related to map design and compilation.

Word	Total occurrences	Document occurrences	% of responses
map	564	100	93%
use	203	76	70%
color	165	72	67%
data	156	68	63%
scale	141	73	68%
legend	136	67	62%
label	112	56	52%
yes	110	108	100%
project	103	63	58%
default	89	47	44%
north	88	59	55%
symbol	79	42	39%
arrow	73	58	54%
poor	73	37	34%
featur	71	37	34%
north_arrow	70	55	51%
titl	66	49	45%
visual	60	44	41%
element	58	31	29%
line	58	34	31%
text	58	32	30%
inform	56	31	29%
hierarchi	51	40	37%
font	49	32	30%
make	47	31	29%

choropleth	46	29	27%
lack	44	30	28%
size	44	32	30%
area	40	26	24%
etc	40	30	28%
look	40	30	28%
see	40	24	22%
detail	39	28	26%
design	38	27	25%

Table 2: Results of the Word Count process, with the key tokens selected for further analysis highlighted in red.

The Word Count process results can be examined as the total number of times a word occurred in the corpus (occurrences) or the percentage of responses in which they occurred, as shown in figure 2.



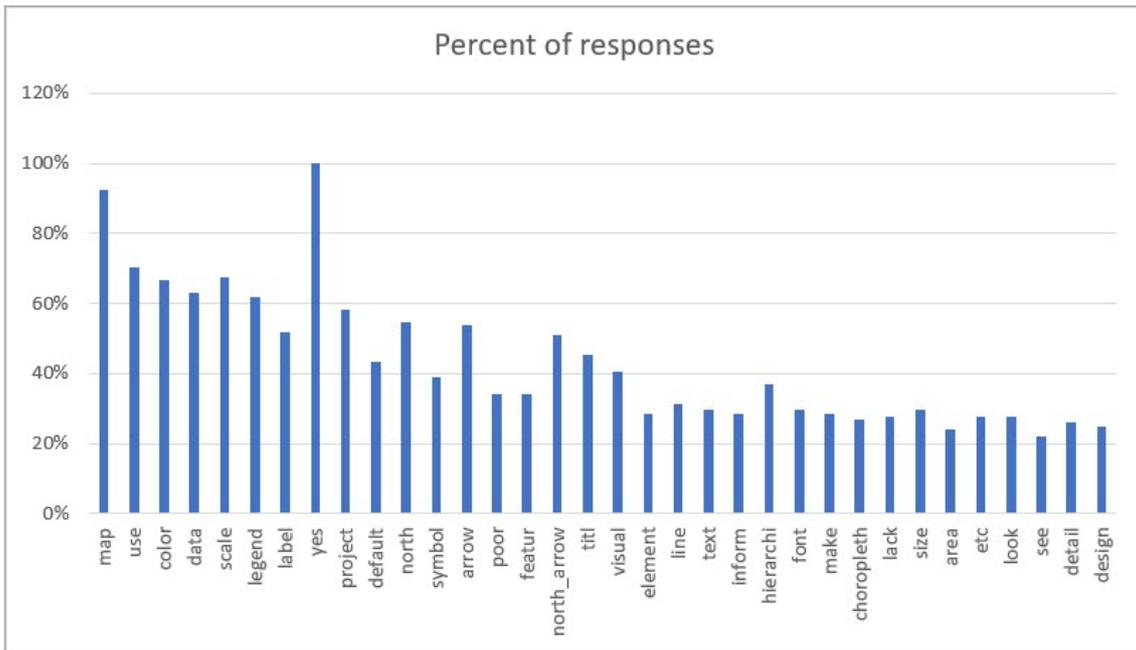


Figure 2. Results of the word county analysis process.

Removal of the most common token, “map”, revealed more variation in the occurrences of the remaining tokens (figure 3).

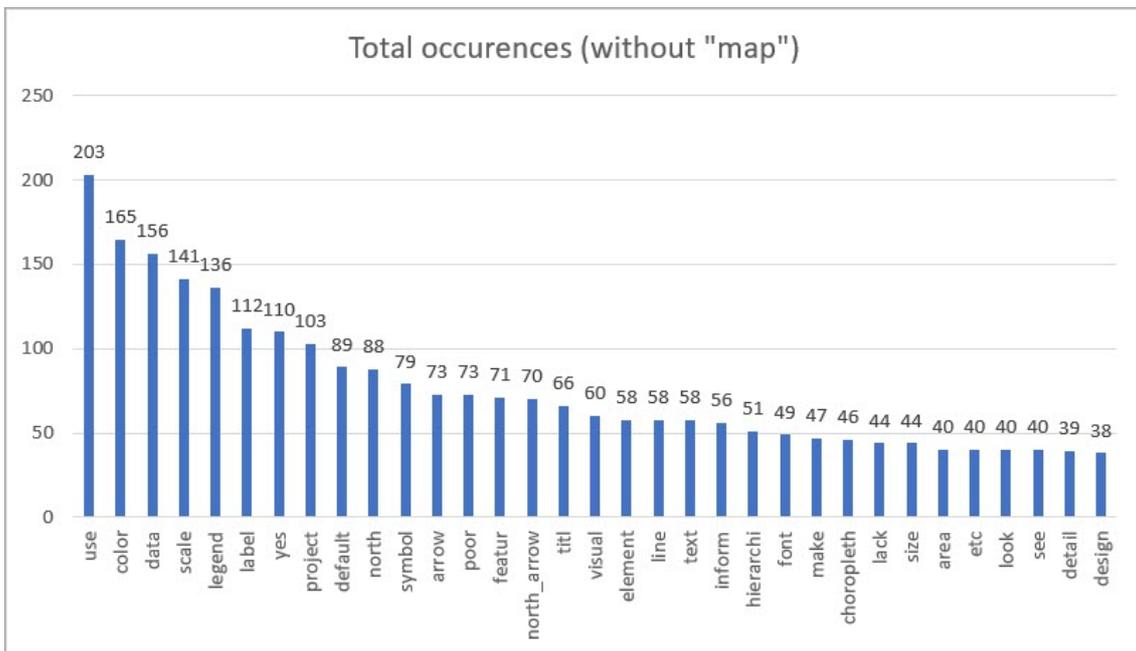


Figure 3. Variation in the occurrences of the most common words when “map” was removed.

To determine if some tokens were often used in conjunction with other tokens, a set of association rules were generated. For example, “map” was often used with “use” in the corpus, as in the upper left box of figure 4. While these analytical results offered some insights, most of the token associations simply confirmed multi-term keywords that the author had suspected; therefore, no further association analysis was performed.

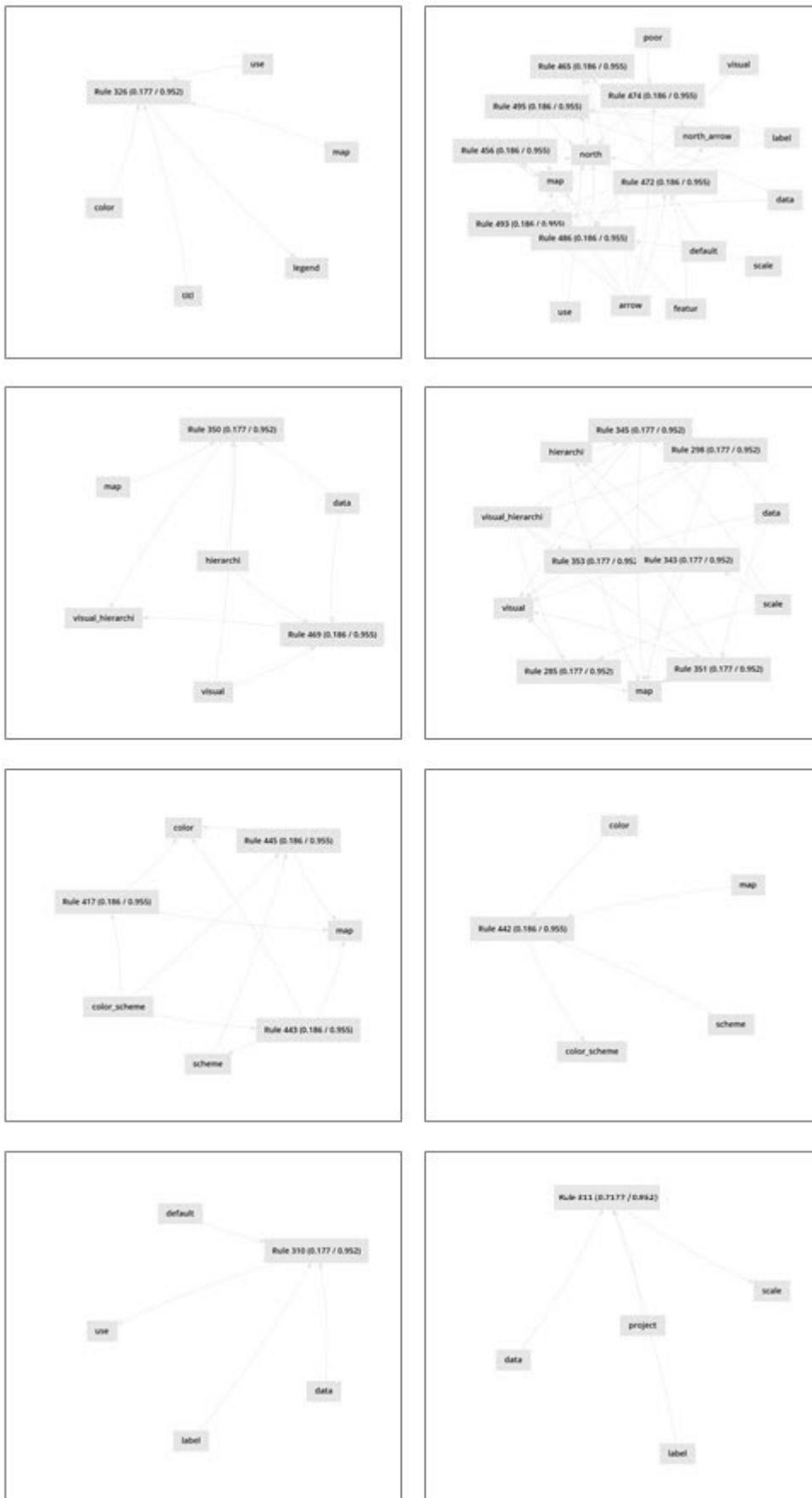


Figure 4. Examples of the results of the Association Rules process in RapidMiner.

Beyond the Text Analysis

After performing text mining and analysis with RapidMiner, the responses were further analyzed relative to the 22 key tokens. Any response with the occurrence of a select key token was copied into a Word document, and the resulting content was grouped based on commonalities. For example, all responses with the key token “north_arrow” were grouped into four categories depending on whether they focused on:

1. North arrows should not always be used on maps
2. The inappropriate use of north arrows
3. North arrows missing on maps
4. The design of north arrows.

Obviously, the groups varied by token. For example, the projection categories included:

1. Projection information in the map marginalia
2. Use of a default projection
3. Use of an inappropriate projection
4. Use of a geographic/unprojected/Plate Carré projection
5. Use of the Web Mercator projection
6. Incorrect projection parameters (e.g., central meridian)

While the key tokens in some responses, such as those above, were fairly isolated from other tokens, others, such as “use”, “scale”, and “design”, were often used in responses that included one or a number of other tokens. For these, responses that contained more than one key token were copied into the documents for each of the key tokens. The relation to other key tokens was also noted in each document.

The result of this part of the analysis was a set of documents, each focused on a key token, which identified the most common guidelines relating to each key token, as well as the association of any key token to other key tokens.

Discussion and Conclusions

Future work will involve publication of the findings as a concise, illustrated primer on cartographic design that describes the most basic elements of this subject and serves as a first resource for novices. While the primer would be targeted toward guiding novices, it is understood those who are uninitiated in a subject sometimes produce products that have more impact than traditional or “correct” products, even when or sometimes because they violate traditional practice.

The primer should describe a set of suggested practices that the readers may find useful to follow as they learn or continue to learn about map making. It would be beneficial to novices, and especially more advanced readers, to add notes about, examples of, and references to more information about deviating from these practices. In addition, due to the brevity of the publication format, it would be useful to provide links and citations to other works to help readers learn more about related cartographic theory and practice.

Additionally, it is recognized that the information used in this research was voluntarily provided in response to a request that essentially solicited input with a negative tone

(thus, no sentiment analysis was performed to quantify the affective state of the responses). The original request was for signs of novice cartography, i.e., things that are wrong with maps made by novices. An essential requirement for production of the primer will be to transform those statements of “what not to do” into guidelines for good practice so that the suggestions are not received as dogmatic and inflexible with no options for creativity and flexibility.

Acknowledgements

Thanks to the many individuals worldwide who responded to my initial email. Without their expertise and their willingness to share their thoughts with me, this work would not have been possible.

References

- Akella, M. K. 2008. First Responders and Crisis Map Symbols: Clarifying Communication. *Cartography and Geographic Information Science* 36 (1), 19–28. doi: [10.1559/152304009787340179](https://doi.org/10.1559/152304009787340179)
- Basaraner, M. and Cetinkaya, S. 2008. Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. *International Journal of Geographical Information Science* 31 (10), 1952–1977. doi: [10.1080/13658816.2017.1346257](https://doi.org/10.1080/13658816.2017.1346257)
- Fuchs, S., Spachinger, K. Dorner, W., Rochman, J., and Serrhini, K. 2009. Evaluating cartographic design in flood risk mapping. *Environmental Hazards* 8 (1), 52–70. doi: [10.3763/ehaz.2009.0007](https://doi.org/10.3763/ehaz.2009.0007)
- Kotu, V. and Deshpande, B. 2014. *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Waltham, MA: Elsevier.
- Liao, H., Wang, X., Dong, W., and Meng, L. 2019. Measuring the influence of map label density on perceived complexity: a user study using eye tracking. *Cartography and Geographic Information Science*, 46 (3), 210–227, doi: [10.1080/15230406.2018.1434016](https://doi.org/10.1080/15230406.2018.1434016)
- McKendry, J. E. and Machlis, G. E. Cartographic design and the quality of climate change maps. *Climatic Change* 95, 219–230. doi: [10.1007/s10584-008-9519-5](https://doi.org/10.1007/s10584-008-9519-5)
- Mierswa, Ingo. 2014. “What This Book is About and What It is Not”, in *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. Boca Raton, FL: CRC Press.
- Raposo, P and Brewer, C. A. 2012. Landscape Preference and Map Readability in Design Evaluation of Topographic Maps with an Orthoimage Background. *The Cartographic Journal* 51 (1), 25–37. doi: [10.1179/1743277412Y.0000000027](https://doi.org/10.1179/1743277412Y.0000000027)
- Statistics New Zealand. 2014. Design principles for maps using New Zealand’s statistical data. Available from <http://archive.stats.govt.nz/~media/Statistics/about-us/statisphere/Files/further-resources-info/design-principles-maps-nz-statistics.pdf>.
- Vermas, Tanu, Renu, and Deepti Gaur. 2014. Tokenization and Filtering Process in RapidMiner. *International Journal of Applied Information Systems (IJ AIS)*, 7 (2), 16–18. doi: [10.5120/ijais14-451139](https://doi.org/10.5120/ijais14-451139).

- A journal article, A. B., Junk, C. D., & Morejunk, E. F. (2019). Title of journal article goes here. *Cartography and Geographic Information Science*, 46, 195. doi: 10.1080/15230406.2019.1574554
- B'Online sources are confusing S. O. (2010). Search for answers at apastyle.org and include issue numbers after volume numbers when there is no doi. *Journal of Articles Without Digital Object Identifiers*, 127 (3), 816-826.
- C magazine article, B. E. (2019, July). Note the last names on this page: Each source type has to be formatted in a different way. [Special issue]. *Prose Magazine*, 126 (5), 96-134.
- D book reference, S. M., Orman, T. P., & Carey, R. (1967). Google scholar's "cite" feature is usually accurate and time-saving. New York, NY: Pearson.
- O'encyclopedia, S. E. (1993). Words. In *The new encyclopedia Britannica* (vol. 38, pp. 745-758). Chicago, IL: Penguin.
- P chapter, P. R., & In an edited volume, J. C. (2001). Scientific research papers provide evidence of frustration with giant style manuals. In P. Z. Wildlife conservation, R. Dawkins, & J. H. Dennett (Eds.), *Research papers are hard work but boy are they good for you* (pp. 123-256). New York, NY: Simon & Schuster.
- Qosenberg, Morris. (1994, September 11). This is how you cite an online news article that has an author. *The Washington Post*. Retrieved from http://www.washingtonpost.com/dir/subdir/2014/05/11/a-d9-11e3_story.html