

OSCAR: A Semantic-based Data Binning Approach

Vidya Setlur*

Michael Correll†

Sarah Battersby‡

Tableau Research

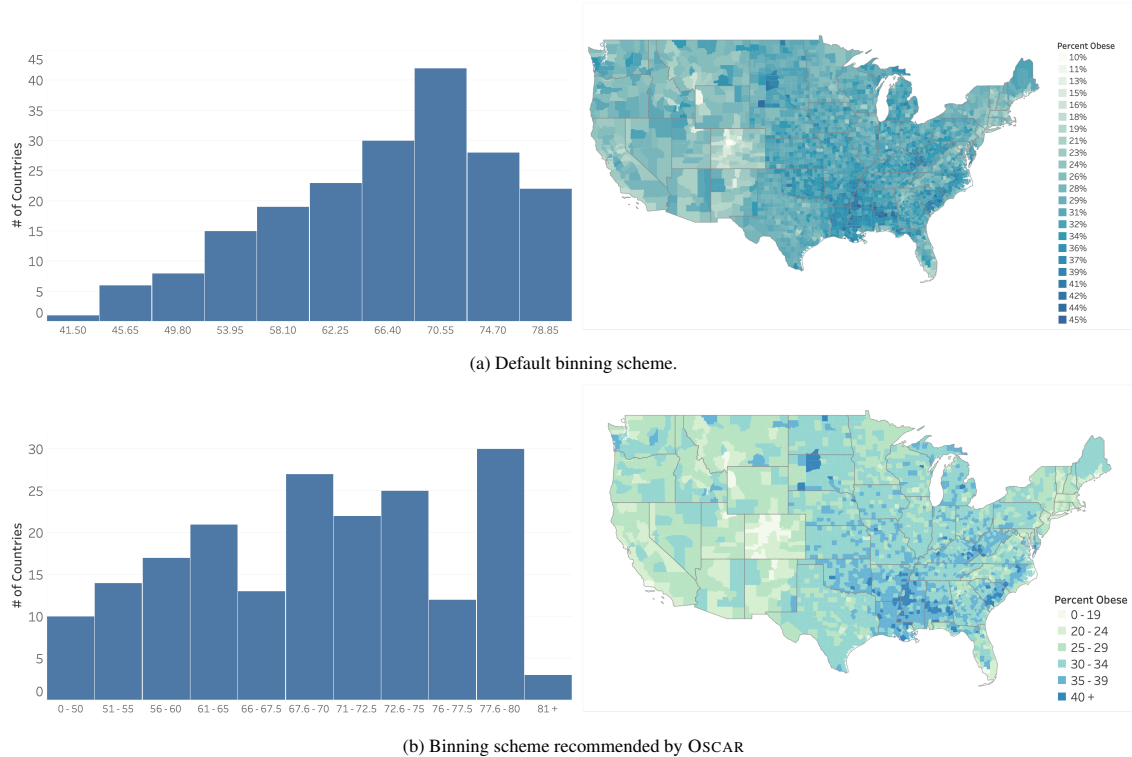


Figure 1: Visualizations showing comparisons of bins for data on per-country *life expectancy* (left) and per-U.S. county *obesity* rates (right). The top-row bins are computed based on statistical properties, while the bottom-row bins are computed by OSCAR. Semantic bins have benefits for legibility, reducing the number of bins (i.e., the visual complexity of the map or histogram), and taking advantage of non-uniformity to either highlight areas of interest or compress long tails of the distribution into single bins.

ABSTRACT

Binning is applied to categorize data values or to see distributions of data. Existing binning algorithms often rely on *statistical* properties of data. However, there are *semantic* considerations for selecting appropriate binning schemes. Surveys, for instance, gather respondent data for demographic-related questions such as age, salary, number of employees, etc., that are bucketed into defined semantic categories. In this paper, we leverage common semantic categories from survey data and Tableau Public visualizations to identify a set of semantic binning categories. We employ these semantic binning categories in OSCAR: a method for automatically selecting bins based on the inferred semantic type of the field. We conducted a crowdsourced study with 120 participants to better understand user preferences for bins generated by OSCAR vs. binning provided in Tableau. We find that maps and histograms using binned values generated by OSCAR are preferred by users as compared to binning schemes based purely on the statistical properties of the data.

Keywords: Data-driven semantics, binning, constraints, geospatial.

Index Terms: Human-centered computing—Visualization—

*vsetlur@tableau.com

†mcorrell@tableau.com

‡sbattersby@tableau.com

1 INTRODUCTION

Binning of quantitative attribute data is a prerequisite step for many aspects of data visualization—the bins can be used to reduce continuous data down to more manageable categories, preserve data privacy through aggregation, generate histograms, or create breaks for ordinal color scales. Different binning schemes prioritize different goals: for instance, the selection of the number of histogram bins may be chosen to minimize error compared to an unknown but estimated population distribution [10], or the bins used for a color palette set in order to maximize both intra-bin coherence and inter-bin difference [15]. Yet, these schemes based on statistical properties can ignore what might be the most crucial property of a binning scheme meant for use in visualization: the *legibility and semantic coherence of the bins*. The ultimate consumers of these binning schemes are human readers of charts and maps, and so a human-centered binning scheme ought to leverage not just the *statistics* of the quantitative field in question but also *data semantics* as well as the *legibility and interpretability* of the resulting bins.

In this paper, we present OSCAR¹, a human-centered binning technique that leverages data semantics and legibility constraints to suggest bins for quantitative data for use in histograms, maps, and

¹ The name OSCAR is inspired by the beloved Sesame Street character who embraces his life living out of a trash *bin* [30].

other charts. OSCAR leverages information from Tableau Public [35] to suggest common bin sizes for a particular field based on the field’s name (or those of fields with semantically similar names), then allows the user to refine the resulting bins. The resulting bins provide a number of useful features for legibility (Fig. 1), including a focus on particular values of interest for specific use cases, a respect for the grain of the data, and the use of non-uniform bins to condense long tails or outliers into single bins. OSCAR addresses attribute “binning” or classification; it does not perform any spatial binning.

We conduct a crowdsourced evaluation of the bins generated from OSCAR and find a general preference for “semantic” bins over bins created via software defaults such as Tableau’s [34]. Beyond our binning scheme, findings also suggest generalizable principles for “human-centered” binning, such as user preferences for “nice” bin boundaries (e.g., whole numbers or numbers rounded to multiples of 5 or 10) and bin boundaries with appropriate granularities that do not collapse information about critical semantic values (e.g., having finer-grained bins for higher age values for *life expectancy*). In addition to specific recommendations for creating human-legible bins, preliminary evidence indicates that OSCAR provides useful bin semantics for creating charts that better match the needs of users.

2 RELATED WORK

2.1 Histogram Binning

Selecting the correct number of bins in a histogram is often portrayed as a tradeoff—too few bins and the distinct shape of a distribution is lost; too many, and the resulting noisy histogram makes shape information difficult to recover. A common assumption is also that the quantitative data in a histogram are *samples* from an unknown *population* distribution: the choice of the number of bins is characterized as an estimation problem, and common histogram binning schemes (such as Sturges’ rules [29], the Freedman-Diaconis rule [10], and Scott’s rule [28]) are based on minimizing error under certain assumptions about the population distribution (e.g., for Sturges’ rule, that it is a unimodal Gaussian) and under certain definitions of error (e.g., for Scott’s rule, Mean Integrated Squared Error).

However, histograms are generated for diverse audiences and for diverse purposes, and these rules may fall short for different human-scale tasks. For instance, Correll et al. [8] find that common rules may generate too few bins for people to reliably identify data quality issues like missing data in distributions (rather than merely their shape). Conversely, Sahann et al. [26] find that relatively few bins are sufficient for viewers to reliably distinguish between different population distributions and suggest there are diminishing returns for creating more bins. Lastly, Gopal Lolla et al. [13] find that important shape information can be lost within traditionally assigned bins and suggest an error function that incorporates shape information when selecting bin boundaries for histograms. We point to these works to suggest the potential benefits of a *human-centered* binning schema that is mindful of how people read (or, just as importantly, fail to read [6, 17, 19]) histograms and how they are employed for a variety of analytical goals beyond estimating the shape of a distribution.

2.2 Cartographic binning

With maps, binning provides an opportunity to explore patterns across spatial distributions. The color encoding on the map enforces visual grouping of regions based on bin category. While continuous or un-classed maps are valuable for maintaining absolute numeric data relationships [36], it is more common to use discrete bins to emphasize the similarity between locations. These bins should be meaningful for the dataset *and* the question(s) being explored; they may be driven by data distribution (e.g., standard deviation), be more arbitrary and unrelated to data distribution (e.g., equal interval), or tied to specific visual / data values of relevance to the cartographer or map reader (e.g., diverging bins arranged around the income requirement for a Federal assistance program) [33]. Some of the most

commonly provided cartographic binning methods [7] are equal interval, Jenk’s optimal or natural breaks [15], mean/standard deviation, quantiles, and “pretty breaks” (rounding to nice-looking numbers), though there are generally also options for manual bin range selection providing more opportunity to tailor the view based on user goals or understanding of data distribution. Less commonly implemented for commercial usage, but still academically interesting to consider are automated methods such as those relying on genetic algorithms [4], or proximity-based binning schemes [23] to encourage more spatially compact, homogeneous regionalization on the map. The use of color also places a constraint on the number of bins: MacEachren [20] has noted that while more detailed maps may be interpretable, the information retained from map reading tends to be reduced to roughly three ordinal categories (high, medium, low).

2.3 Incorporating semantics into visualization design

Other techniques have augmented visualization designs through the use of automatic lookups meant to resonate with the semantic backgrounds of viewers. Kim et al. [18] create personalized analogies based on location data to help users better interpret distances (e.g., explaining concepts in terms of the distance from San Francisco to San Jose might be easier to conceptualize for a Bay Area resident than for the distance from London to Reading). Most similar to our work is that of Setlur & Stone [31], who employ a linguistic approach for automatically generating semantically resonant color palettes for categorical data. Their algorithm uses Google Image data to find associations between words and colors (e.g., for a bar chart of vegetable produce sales, assigning green and orange colors to the data values ‘broccoli’ and ‘carrot’, respectively). Inspired by this work, we demonstrate a technique for applying external domain knowledge and semantics of common binning patterns and categories to help inform reasonable defaults for histograms and breaks for classes in data. For data attributes that do not have semantic associations, we apply best practice binning techniques to provide default bins.

3 OSCAR BINNING PROCESS

The algorithm employs a two-step process for creating more useful bins for numerical data attributes: (1) Semantic bins for fields that have semantic lookups and (2) default bins that create human-legible bin breaks based on the underlying statistical properties in the data.

3.1 Generate Semantic Bin Lookup

To generate a lookup of semantic categories and their corresponding bins, we employ a data-driven approach of mining both a public corpus of survey questionnaires [38] and published Tableau Public [35] visualizations containing binned fields. Leveraging a corpus of prevalent semantic categories provides recognizable and familiar bin breaks that are often used in data analysis. The core idea of our technique is employing Latent Dirichlet Allocation (LDA) [5], a popular topic modeling technique to extract topics from a given corpus as proxies for semantic bin concepts.

Build the LDA model for bin concepts. We build an LDA model from the binned field names from Tableau dashboards along with common demographic information commonly found in surveys that have binned numerical responses such as age, salary, population, etc. The LDA model is trained using MALLET [22] for 1000 iterations with the top 100 binned field names. We then apply the LDA model on the survey corpus to get probabilities for bin concepts in each survey to generate a lookup of strings and their associated bin sizes.

Create a list of related concepts. For each bin concept, we have a label name and a set of related concepts such as synonyms. Enriching the bin concept with related concepts increases the probability of a match with the LDA model. These seed lists are created using a thesaurus service [25] and Wordnet synsets [9]. For example, the

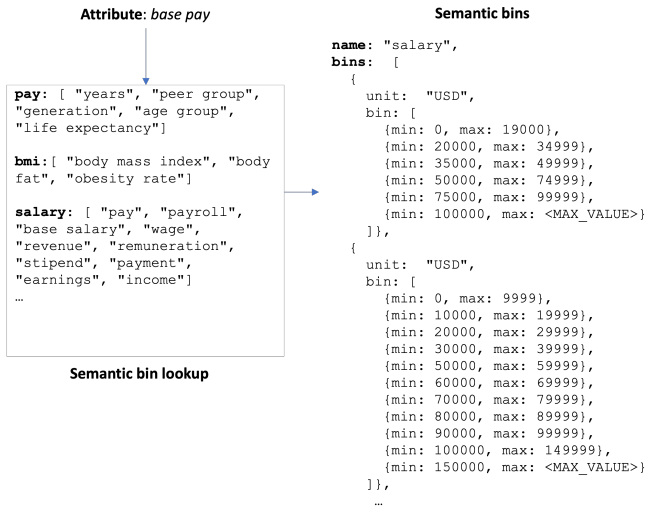


Figure 2: Semantic bin lookup for the attribute *base pay* in OSCAR.

bin concept ‘salary’ has a seed list: [‘pay’, ‘payroll’, ‘base salary’, ‘wage’, ‘remuneration’, ‘stipend’, ‘earnings’, ‘income’].

Align bin concept and bin breaks. The final step to creating the semantic bin lookup is an alignment between the bin concept c , along with its related concept list (together we denote as $R(c)$), and question topic t in the surveys with every topic being mapped to at most one concept. We align each t to c with the maximum score, $S(c, t)$ that measures the summed probabilities of c and $R(c)$ in t : $S(c, t) = \sum_{w \in R(c)} p(w|t)$. We remove all alignments where $S(c, t) < a_{threshold}$. In practice, we found that $a_{threshold} = .06$ provides reasonable results for precision and relevance.

3.2 OSCAR binning algorithm

3.2.1 Compute Semantic Bins

Given an attribute, OSCAR first checks if there is a match with the attribute and the semantic bin lookup, as described in Section 3.1. We employ fuzzy matching and lemmatization to match attribute strings to bin concepts in the table [21]. If there is a match and there is a semantic bin option, these bins are applied to the attribute to generate the corresponding visualization (e.g., a histogram or a map shown in Figure 1b). If multiple options of semantic bins are available, the semantic bin option that is closest to the data bounds of the attribute is selected. An overview of this process is illustrated in Figure 2.

3.2.2 Compute Default Bins

In the absence of semantic bins, OSCAR attempts to use smart defaults to select human-legible bins (Figure 3). After choosing a binning based on the statistical properties of the distribution (e.g., Sturges rule [29]), OSCAR performs the following optimizations:

Constrain the number of bins to a maximum number. For use in a color ramp, the designer might wish to limit the number of bins in order to maximize the discriminability of colors or reduce the complexity of the legend (see Figure 1a, right for an example of a perhaps overly complex color legend). For use in a histogram, the designer might wish to make sure that there are not so many bins that features in the distribution or the labels of bins are illegible (see Figure 3a). While these maximums are to some extent dependent on contingent properties such as display resolution or color ramp choice, we note that the default maximum number of bins in VegaLite [27] is 20, and the maximum number of bins for stepped color ramps in ColorBrewer [14] is 12 (and even then, only with a subset of ramps).

Round the bin extents to match the precision of the data. Many existing rules for selecting histogram bins can produce arbitrary precision floating point bin boundaries. These boundaries can be misleading if they promise or suggest precision beyond the precision of the data. For instance, integer data should not have decimal bins (see Figure 3b), and data expressed in terms of round millions of dollars should not have bin widths of tens of thousands of dollars.

Round the bin extents to convenient values of 5 or 10. We draw inspiration here from VegaLite’s [27] “nice” operator, which rounds bin or scale extents to an appropriate power of 10 given the precision of the value (e.g., $9 \rightarrow 10$, $0.9 \rightarrow 1.0$, etc.).

Shift bins such that 0 does not occur within a bin. Again drawing inspiration from VegaLite [27]’s “anchor” operator, we believe there is a semantic distinction between positive and negative numbers for a wide variety of quantitative fields. By shifting bin boundaries, we can ensure that 0 falls between bins.

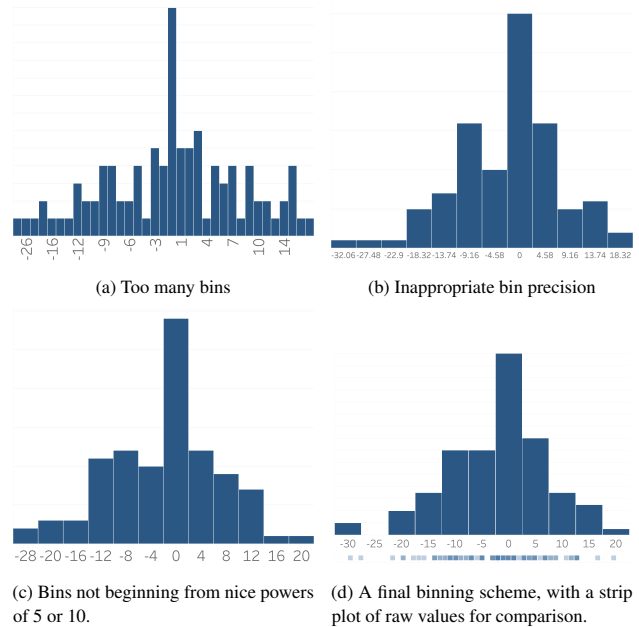


Figure 3: Default binning (Fig. 3d) heuristics applied on a dataset of 100 integers drawn from a normal distribution. The bins are 1) limited in number (especially for use in color ramps, where the ability to distinguish colors is limited, compared to Fig. 3a), 2) consistent with the data grain (e.g., if the data are integers, then the bins should not be decimal numbers as in Fig. 3b), and 3) rounded to “nice” values (say, powers of 5 or 10, rather than widths of 4 in Fig. 3c).

4 EVALUATION

In this study, we explored the high-level research question: What are participants’ preferences when viewing charts with and without semantic binning? We hypothesized the following:

- **H1:** Participants will prefer semantic bins over default bins to better reflect the categories for the corresponding data attribute.
- **H2:** Participants will not prefer coarse semantic bins as it may be harder to discern distribution patterns.
- **H3:** Participants will not prefer non-integer bin starts in the default bins as these bins are harder to interpret.

4.1 Participants

We conducted a power analysis to determine the number of participants required to find an overall difference in preference rankings.

With a medium effect size of ~ 0.49 , our analysis suggests that a target sample of 112 would yield 95% power to detect an overall difference between preference rankings at an alpha level of 0.05.

We recruited 125 participants from Amazon Mechanical Turk. To qualify for the study, participants were required to be located in the USA and have a 95% acceptance rate on previous tasks. We compensated participants at a rate of \$2.00 for a six-minute study. After excluding participants who failed attention checks (e.g., failing to select a specific answer in a multiple-choice question) or entered nonsensical responses, we ended up with 120 participants, with 64 that identified as women ($M_{age} = 34.05$, $SD_{age} = 10.52$), 52 as men ($M_{age} = 36.65$, $SD_{age} = 11.88$), and four chose to not disclose.

The participants completed a subjective graph literacy report [12] and reported an average value of 3.85 out of 6 ($SD = 0.82$, 1 = low self-reported literacy, 6 = high self-reported literacy), suggesting that most participants were comfortable with visualizations but did not identify as visualization experts. Only 7 people reported that they create visualizations often for work or as a hobby, and 21 people reported rarely interacting with visualizations in their daily lives.

4.2 Stimulus and Design

We created two sets of ranking stimuli - one set using OSCAR, each differing in the number of bins (5 – 14 bin categories) and the other based on Tableau binning functionality [1]. For generating these bins, we used five datasets: CDC obesity health data [2], World Indicators [11], US Census commute data [37], country Gini coefficient [3], and Titanic passenger data [16]. Six sets of histograms and four sets of maps were generated, resulting in a total of ten preference tasks. We denote the OSCAR binned histograms and maps as $Histo_{oscar1}$, $Histo_{oscar2}$, Map_{oscar1} , and Map_{oscar2} , with $oscar1$ having less number of bins than $oscar2$; the default binned histogram and map as $Histo_{default}$ and $Map_{default}$ respectively.

4.3 Procedure

Participants were provided a link and a brief introduction to our survey. Participants viewed all the ten preference tasks in random order. For each task, participants were shown the default and semantic binning variants of chart images and were asked, “Which chart is more useful for showing the distribution of the data.” They were provided a free-text response field to add feedback about their ranking choices. At the end of the survey, participants reported demographic information and completed the self-report visual literacy test.

4.4 Analysis

We conducted a Friedman Rank Sum test using the PMCMRplus R package [24] to compare preference rankings for the visualizations, with post-hoc pair-wise comparisons via Conover’s test with Bonferroni’s correction to determine the ranking differences.

4.5 Results

There is a significant difference in user preferences semantically binned charts overall, supporting **H1**. For the histogram charts, we observed $Histo_{oscar1}$ over $Histo_{default}$ (Friedman $\chi^2 = 34.62$, $p < 0.001$), $Histo_{oscar2}$ over $Histo_{default}$ (Friedman $\chi^2 = 20.59$, $p < 0.001$), and $Histo_{oscar2}$ over $Histo_{oscar1}$ (Friedman $\chi^2 = 19.36$, $p < 0.001$). For maps we observed Map_{oscar1} over $Map_{default}$ (Friedman $\chi^2 = 54.25$, $p < 0.001$), Map_{oscar2} over $Map_{default}$ (Friedman $\chi^2 = 68.32$, $p < 0.001$), and Map_{oscar2} over Map_{oscar1} (Friedman $\chi^2 = 69.33$, $p < 0.001$).

Observations confirm that users prefer finer-grained semantic bins over coarser-grained ones (**H2**). Also, low preference rankings for default binned charts containing non-integer bin starts support **H3**.

Feedback from the participants also reflected these observations. P27 stated, “The obesity numbers seemed to be a bit arbitrary (referring to the default bins) and wanted the images that showed

something I’m familiar with.” Other comments on bin granularity included, “I don’t like the histogram clumping up everything. Prefer to see it all spread out [P76]” and “I don’t see that much color variation with the smaller categories. Liked seeing more differences [P07].” There was also feedback against non-integer bin breaks – “What do those decimals even mean? I can’t wrap my head around them [P71]” and “It’s easy on the eyes to see the whole numbers; they are less scary compared to all those floating numbers [P100].”

5 DISCUSSION AND FUTURE WORK

By relying on prior bins constructed by users of Tableau Public that are likely to have semantic importance, as well as by affording defaults that are oriented around the human legibility and interpretability of bins, OSCAR was able to produce binning schemes that were often preferred over binning schemes that are only sensitive to the distribution. We view these initial successes as evidence that the automatic integration of semantic information from usage data can improve visualization design and that the careful setting of defaults can influence the success or failure of a visualization design. However, we see several areas for further improvement and study, both in terms of the design and evaluation OSCAR as well as for human-centered visualization design, especially for mass audiences.

Provide user affordances for bin refinement and repair. Our semantic binning schemes could result in poor outcomes for a variety of reasons, such as poorly chosen defaults or concept drift as data semantics change over time (e.g., an early map of COVID case data would need to be re-binned as the scale of the pandemic changed). To counteract these potential failures or shifts, we envision OSCAR to take more of a mixed-initiative role in dealing with underspecification, allowing users to “repair” [32] instances of under-specification, or directly compare different binning schemes (as in Figure 1), or to tailor their binning schema to particular contexts (for instance, putting different constraints on bins to be used for generating color ramps, as opposed to bins used to generate histograms).

Evaluate quality of semantic bins for various analytical tasks.

The semantic bin lookup method employs a corpus-driven approach that identifies bin breaks prevalent for various commonly occurring concepts such as salary, population, and age. While the LDA model provides an effective approach for identifying probable semantic categories with corresponding bins, further investigation needs to explore the quality of these bins based on user intent and the analytical task at hand. Our evaluation is preliminary and focused on user preference. It remains future work to investigate if there could be a preference-performance gap for certain tasks (such as identifying missing data or modes) [8], or determine if there are quantifiable benefits to semantic bins beyond their legibility.

6 CONCLUSION

This paper presents a technique, OSCAR that automatically selects bins based on the inferred semantic type of the data attribute. Using a combination of data-driven semantic lookup information obtained from public survey corpora and Tableau dashboards containing binned fields, OSCAR provides semantic bins and smart defaults to generate human-legible bins. We conducted a crowdsourced user preference study with 120 participants to better understand user preferences for bins generated by OSCAR vs. default binning provided in Tableau. We find that maps and histograms using binned values generated by OSCAR are preferred by users as compared to binning schemes based purely on the statistical properties of the data. These preliminary results indicate that OSCAR provides useful bin semantics that could be incorporated into visual analysis workflows to create more semantically meaningful charts.

REFERENCES

- [1] Tableau Help: Create bins from a continuous measure. https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculations_bins.html.
- [2] Cdc nutrition, physical activity, and obesity: Data, trends and maps, 2022. <https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html>.
- [3] Gini coefficient by country 2022, 2022. <https://worldpopulationreview.com/country-rankings/gini-coefficient-by-country>.
- [4] M. P. Armstrong, N. Xiao, and D. A. Bennett. Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Annals of the Association of American Geographers*, 93(3):595–623, 2003.
- [5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022, mar 2003.
- [6] L. Boels, A. Bakker, W. Van Dooren, and P. Drijvers. Conceptual difficulties when interpreting histograms: A review. *Educational Research Review*, 28:100291, 2019. doi: 10.1016/j.edurev.2019.100291
- [7] C. A. Brewer and L. Pickle. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers*, 92(4):662–681, 2002.
- [8] M. Correll, M. Li, G. Kindlmann, and C. Scheidegger. Looks good to me: Visualizations as sanity checks. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):830–839, 2019. doi: 10.1109/TVCG.2018.2864907
- [9] C. Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.
- [10] D. Freedman and P. Diaconis. On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4):453–476, 1981. doi: 10.1007/BF01025868
- [11] Gapminder. World development indicators, 2022. CC-BY Dataset: <https://gapminder.org/data/>.
- [12] R. Garcia-Retamero and E. T. Cokely. Communicating health risks with visual aids. *Current Directions in Psychological Science*, 22(5):392–399, 2013.
- [13] S. V. Gopal Lolla and L. L. Hoberock. On selecting the number of bins for a histogram. In *DMIN 2011: proceedings of the 2011 international conference on data mining (Las Vegas NV, July 18-21, 2011)*, pp. 344–350, 2011.
- [14] M. Harrower and C. A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003. doi: 10.1179/000870403235002042
- [15] G. Jenks. Optimal data classification for choropleth maps. occasional paper no. 2. *University of Kansas, Department of Geography*, 1977.
- [16] Kaggle. Titanic dataset, 2022. <https://www.kaggle.com/c/titanic/data>.
- [17] J. J. Kaplan, J. G. Gabrosek, P. Curtiss, and C. Malone. Investigating student understanding of histograms. *Journal of Statistics Education*, 22(2):null, 2014. doi: 10.1080/10691898.2014.11889701
- [18] Y.-S. Kim, J. Hullman, and M. Agrawala. Generating personalized spatial analogies for distances and areas. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 38–48. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858440
- [19] S. Lem, P. Onghena, L. Verschaffel, and W. V. Dooren. On the misinterpretation of histograms and box plots. *Educational Psychology*, 33(2):155–174, 2013. doi: 10.1080/01443410.2012.674006
- [20] A. M. MacEachren. The role of complexity and symbolization method in thematic map effectiveness. *Annals of the Association of American Geographers*, 72(4):495–513, 1982.
- [21] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, USA, 2008.
- [22] A. K. McCallum. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>, 2002.
- [23] M. S. Monmonier. Maximum-difference barriers: An alternative numerical regionalization method. *Geographical analysis*, 5(3):245–261, 1973.
- [24] T. Pohlert and M. T. Pohlert. Package ‘pncmrplus’. *R Foundation for Statistical Computing, Vienna, Austria*, 2018.
- [25] Programmable Web. Thesaurus API, 2022.
- [26] R. Sahann, T. Müller, and J. Schmidt. Histogram binning revisited with a focus on human perception. In *2021 IEEE Visualization Conference (VIS)*, pp. 66–70, 2021. doi: 10.1109/VIS49827.2021.9623301
- [27] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2017. doi: 10.1109/TVCG.2016.2599030
- [28] D. W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 12 1979. doi: 10.1093/biomet/66.3.605
- [29] D. W. Scott. Sturges’ rule. *WIREs Computational Statistics*, 1(3):303–306, 2009. doi: 10.1002/wics.35
- [30] Sesame Street. Oscar the Grouch™, 2022. <https://www.sesamestreet.org/>.
- [31] V. Setlur and M. C. Stone. A linguistic approach to categorical color assignment for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):698–707, 2016. doi: 10.1109/TVCG.2015.2467471
- [32] V. Setlur, M. Tory, and A. Djalali. Inferencing underspecified natural language utterances in visual analysis. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, IUI '19, p. 40–51. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3301275.3302270
- [33] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard. *Thematic cartography and geovisualization*. CRC Press, 2014.
- [34] Tableau Software. Tableau Calculations, 2022. https://help.tableau.com/current/pro/desktop/en-us/calculations_bins.htm.
- [35] Tableau Software. Tableau Public, 2022. <https://public.tableau.com/en-gb/s/>.
- [36] W. R. Tobler. Choropleth maps without class intervals? *Geographical analysis*, 5(3):262–265, 1973.
- [37] U.S. Census Bureau. 2019 American Community Survey 5-year estimates, 2019.
- [38] D. Zavala-Rojas, D. Sorato, L. Hareide, and K. Hofland. The Multilingual Corpus of Survey Questionnaires: a tool for refining survey translation. *Meta: Journal des traducteurs*.