A Visualization Course for Journalism Students

Stefan Jänicke   stjaenicke@imada.sdu.dk

IMADA, Southern University of Denmark, Odense, Denmark

Abstract— Visualization as a means to communicate information relevant to society in a visual form has become more and more important in the recent years. As the prior medium has changed from print to online media, writers of online articles and bloggers regularly serve their readers with interactive visual interfaces that give intuitive insights into the discussed topics. The change of how information is brought to the public and the large scale of digital information available have also changed teaching contents of journalists who learn aspects of data processing and representation. This year, 21 journalism students participated in my interdisciplinary visualization course. This paper reflects on theoretical course contents, conducted student projects and the feedback of journalism students on how a visualization course could be modified to perfectly fit the needs of becoming data journalists.

1 INTRODUCTION

This year, my course on information visualization for the digital humanities (equivalent to five ECTS credits) was offered for Masters students of journalism for the first time. The course typically involves Masters students from different subjects such as computer science, digital humanities, and humanities, but this time, 21 out of 23 participants were journalism students. Nevertheless, the course was suited for the demands of becoming data journalists as it includes a balanced training of theoretical and practical knowledge focusing on developing intuitive visualizations not only for domain experts but also for casual users. In this paper, I share my experiences in teaching theoretical and practical contents of the course. Accompanied with the feedback given by five participating journalism students, I aim to guide other teachers to establish related courses on the intersection of visualization and journalism.

My course included a theoretical training discussed in Section 2 in the form of a lecture giving a broad overview of general visualization aspects, visual means to enhance detailed views of text in particular and data in general, typical overview visualizations providing visual data summaries, and the Information Seeking Mantra [8] as a bridge between overview and detail. The theoretical course was complemented with a practical training discussed in Section 3, in which the students carried out little visualization projects themselves. On the basis of data sets being of interest to the public, the students processed the data according to visualization ideas and developed Web-based applications used for interactive visual data exploration. The main focus was that participating journalism students learned how to computationally process related data sets, to map occurring data features to visual variables, and to develop interactive visualizations that intuitively present complex topics to casual users.

2 VISUALIZATION LECTURE

Originally, my course was designed for students related to digital humanities (computer science, humanities, digital humanities). Looking at the skills of the journalism students who joined this year, they are comparable to humanities students with basic programming skills. The main purpose for their participation is that analyzing large data sources, e.g., the Offshore Leaks Database[1], becomes more and more important due to their relevance to society, and the visual presentation of the gained insights to the public plays a crucial role when publishing the journalistic work. The digital humanities is related in that scenario as visualizations designed in that domain require to be easily comprehensible—journalists need to design visualizations easy to be understood by casual users. In addition, related works exist where
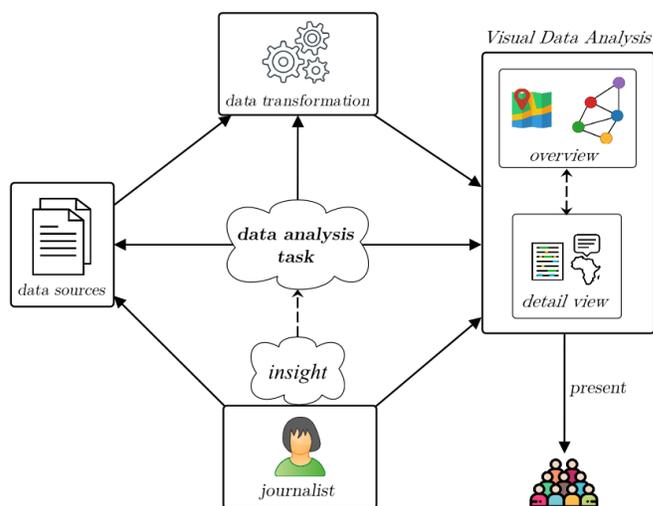

Fig. 1. Visual data analysis process for journalists adapted from [5].

journalistic tasks are performed on historical data sources by digital humanities scholars. An example is given by Kaufman [6] who quantitatively analyzed the Henry A. Kissinger Papers and presented her outcomes with graph and tag cloud visualizations.

Journalistic work can also be described, similarly to the visual text analysis process for digital humanities scholars [5], by the visual data analysis process shown in Figure 1. Journalistic projects start with a central idea—the *data analysis task*—directed towards a specific data source that is of interest to the public. Data transformation aspects and the developed visualizations strongly relate to this task and the underlying data. Overview and detail visualizations have two functions in this context. On the one hand, they aid at forming an argument for the journalist when analyzing the data, on the other hand, they serve to make the written argument in articles easier understandable by the readers. Referring to this visual data analysis process, I taught the following related aspects:

- **Visualization Design:** First, I gave an overview of general guidelines of visualization design based on Munzner's book [7]. Acquiring the capability to break down project ideas into the relevant questions *WHAT* (data shall be visualized), *WHY* (a visualization is necessary, in other words, what are the user tasks to be supported), and *HOW* (data can be visualized to support the intended user tasks) was the main focus in this regard. The great opportunity was that students learned the same terminology

---
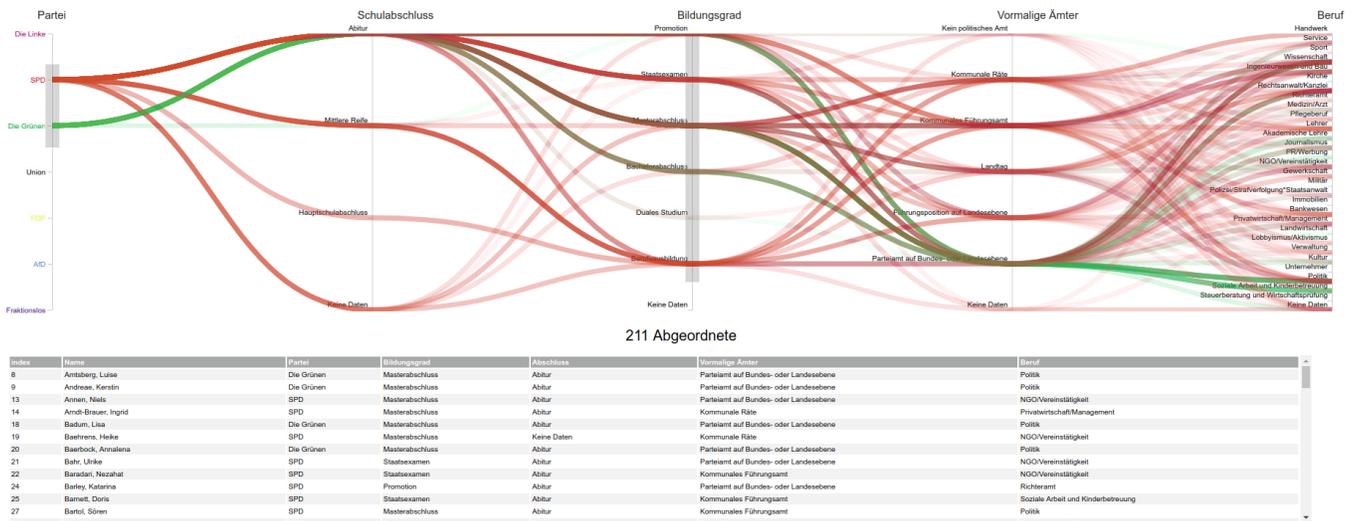
[1] https://offshoreleaks.icij.org/

Fig. 2. A Perfect Setup project: Analyzing the biographies of 709 politicians of the German Bundestag.

for visual design independent on their academic backgrounds in journalism or digital humanities. In addition, learning data and task abstraction strategies denoted an appropriate contrast to the many concrete projects discussed in the course.

- **Overviews:** Throughout the course, I presented diverse overview techniques—frequently used to visually communicate insights made by humanities scholars to casual users, thus, also being relevant for journalists aiming to communicate their insights to the public—to the students: heat maps, tag clouds, maps, timelines and graphs [5]. On the one hand, theoretical and technical details were outlined, on the other hand, practical examples from the related works underpin the value of overview techniques for (digital humanities) scholars, journalists and casual users alike.

- **Information Seeking Mantra:** On the bridge between detailed views and overviews, I discussed the main differences among both techniques using Shneiderman's mantra "Overview first, zoom and filter, details-on-demand" [8]. This included the discussion of interaction techniques to switch between both views. Many related visualizations (designed for related tasks in digital humanities) that have been designed implement the information seeking mantra as casual users are provided with overview visualizations, and means of interaction lead them to interesting patterns within the underlying data source.

- **Detail views:** Writing texts is a fundamental task for journalists. In one lecture, I described visual means how the reading of a text can be enhanced. This included the use of different techniques to highlight information important for the reader, and means how to annotate texts with supportive information—including overview visualizations capable of supporting a journalist's argument made in the article.

As two digital humanities students joined the course as well, the domain was briefly introduced, and related terminology like *close reading* for detailed views or *distant reading* for overview visualizations were used. The course further included primarily the discussion of visual designs made to support (digital) humanities scholars in analyzing texts. Those were extended with some related visualizations from journalistic articles. All in all, the course was modified to serve the needs of journalism and digital humanities students both facing similar data analysis tasks.

3 STUDENT PROJECTS

The practical training in the form of student projects was announced during the first lecture. I asked students to think about project ideas on visualizing data relevant to society having in mind that visualizations should be ready to be offered for readers of online articles. The theoretical part of the course covered half of the semester, and students brought forth their ideas throughout the whole time frame. The longer the theoretical course took, the more precise students were able to formulate project ideas using the learned technical terms. Most of the students were only rudimentary trained in computer science, thus, the technical realization of the projects was a crucial step for some of the teams. Partially, I helped them how to apply existing tools and visualizations to their data. Also, students took charge of services provided by the university for computer science students of early semesters. Nevertheless, few groups were also able to implement sophisticated visualizations with less technical support. In the following, the conducted projects are briefly discussed. The project results were presented during the last semester week. All groups reflected the carried out steps and taken design decisions using the learned process models and visualization terminology. A successful project participation was mandatory for students to be approved for the oral course exam.

Biografies of the German Bundestag Politicians A group of students got inspired by the New York Times visualization [1] reporting on the career paths of the politicians in the U.S. congress. They argued that it would be highly interesting for the public to see the biographies of German Bundestag politicians, in other words, to see the career steps to become a Bundestag politician. As opposed to the NYT visualization, the students had an easier understandable visualization design in mind. As of June 2019, the students collected the biographies from the Bundestag homepage[2] focusing on different aspects of political careers: political party, school degree, education degree, former political position and former practised profession. In order to extract those information for 709 politicians, they implemented rules to parse the crawled biographical texts accordingly. In order to browse the gained data, they adapted a parallel coordinates visualization variant that can be used to interactively search for career patterns. Figure 2 compares members of the SPD and Die Grünen Bundestag factions. One can find out that all members of Die Grünen had a position on federal or state level before joining the Bundestag. Instead, many SPD politicians having a position in the commune were elected for the Bundestag.

Social Network of German Hip Hop Musicians The prior goal of this project was developing a database of German hip hop musicians, and, accompanied with a list of relations, a corresponding social network visualization. Based on a list offered by Wikipedia[3], individual musicians and basic biographical information could be extracted. While the given list of bands was used to define close relations between two musicians as they worked together, a loose relation was defined when two musicians were having a similar age and were living in the same city. The students also planned to define close relations when two musicians recorded a song together, but the project time was too short for a manual solution and the programming skills of the participating students were not sufficient to perform this task automatically. However, the resultant database included 352 hip hop musicians with a multitude of relations recorded. Gephi [3] was used for visualizing different subsets of the graph as shown in Figure 3.
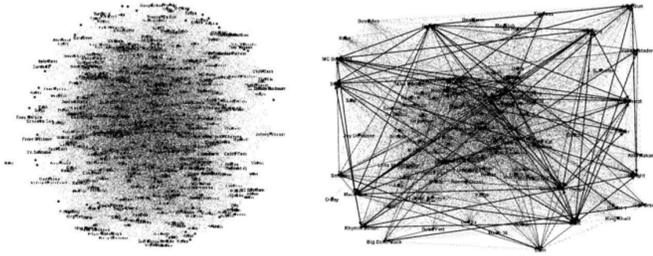

Fig. 3. Social network visualization of German hip hop musicians visualized with Gephi.

German Michelin Restaurants A group of four students focused on the history of German Michelin restaurants. They were able to generate an aestehtic visual output in the form of an interactive map accompanied with a time slider that allows to explore changes throughout the years (see Figure 4). The star size encodes the number of Michelin stars (1, 2, or 3) and a star is highlighted in orange once the number of stars change. When transforming historical data that was provided in lists of Michelin restaurants per year, the students had to encode uncertainties as some restaurants were already closed leading to geocoding errors. As the main focus of the project was comparing the coverage of Michelin restaurants in Eastern and Western Germany, in such cases, cities were chosen instead of exact positions keeping global relationships stable. To support the comparison task, numbers of Michelin restaurants could be compared by state using a horizontal bar chart. The students reflected that the visualization visually underpins the fact that only very few Michelin restaurants are present in the Eastern part of Germany, which is not only related to the division of Germany before 1990, but also related to the large distance of Eastern German states to France being the origin of the Guide Michelin.

Comparative Analysis of Fake News Websites A group of students quantitatively analyzed articles from German websites that are known for publishing fake news articles focusing on the question: *What are the topics central to the reporting of the three most popular fake news websites Philosophia Perennis, Compact Magazin and Journalisten Watch?* After crawling respective articles, RStudio[4] was used for data processing purposes like tokenization and word frequency counting. Next to German stopwords, the students blacklisted words that systematically occured in each article of a provider. For visualization purposes, students used TagPies [4] to support comparatively analyzing frequently used words (see Figure 5). The students used the visualization and summarized their results as follows. All the three analyzed realms use the terms *children* (Kinder), *violence* Gewalt, Facebook, *fear* (Angst) and EU (European Union) in similar frequencies. However, some tags placed in the outer bounds of the TagPie sectors

³   https://de.wikipedia.org/wiki/Liste_von_Hip-Hop-Musikern_Deutschlands

⁴ https://www.rstudio.com/
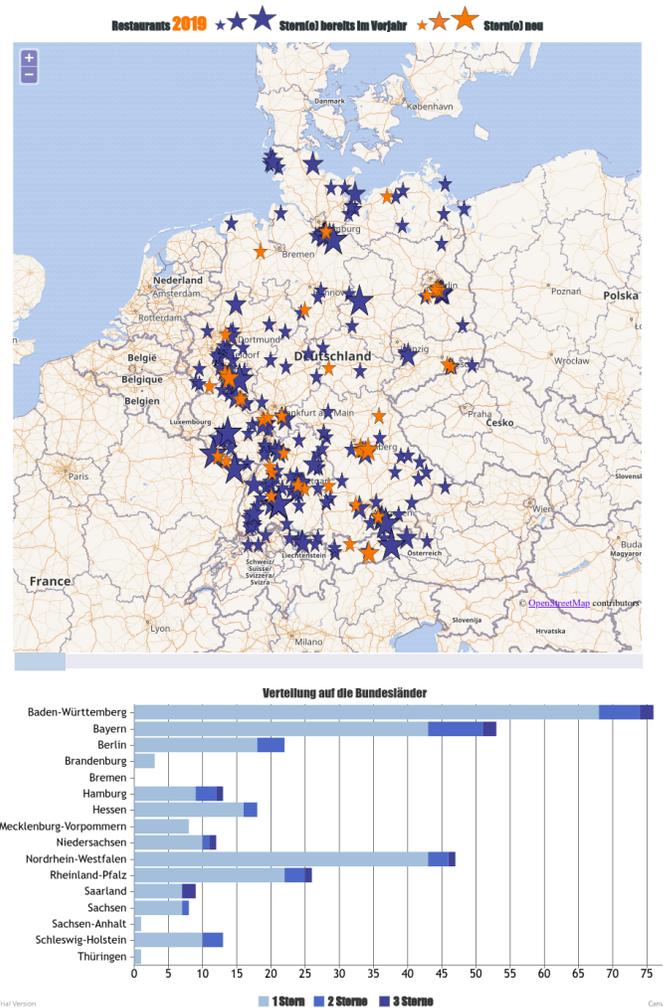
Fig. 4. Geospatial-temporal analysis of German Michelin restaurants.

indicate different objectives. While words like *church* (Kirche), *islamization* (Islamisierung) and *culture* (Kultur) outline the religious focus of Philosophia Perennis, Compact Magazin rather thematizes political issues indicated by words like China, Putin and NATO. Journalisten Watch rather discusses broad topics ranging from migration to media and German politics.

Education Spending in German States The German Federal Statistical Office is responsible for collecting and providing objective information on politics, government, economy, administration and citizens. Available data sets are often subject to online articles due to being of interest to diverse social classes, and charts and maps are offered to the readers to easier comprehend the discussed topic. One student group looked at how much money German states spent for educating pupils—data available from 2011 to 2016. They implemented a choropleth map with uMap[5], the result is shown in Figure 6. The students decided to color the state polygons according to the spending amounts in 2016, but they choose to minimal color differences to make national spending differences salient. When clicking on a state, a line chart informs on the respective spending trend. The original inquiry—comparing the spendings of different states—is not easy to figure out as states need to be individually clicked and memorized for comparison. The discussion of the project result included potential improvements to meet the intended purpose of the visualization.

⁵ https://umap.openstreetmap.fr/de/

Fig. 5. Analyzing the themes of German Fake News Websites.

Fig. 6. Education spending in German states focusing on Saxony.

Music Preferences in Eastern and Western Germany   Comparing the music features favored in different German cities was focused in this project. More precisely, students wanted to visually communicate the differences of music heard in Eastern and Western Germany. Therefore, the students selected five cities from Eastern and Western Germany each and analyzed the songs from the corresponding Spotify playlists[6] for each of those cities according to the music features *happy, dancable, explicit* and *energetic*. The mean value from a city's playlist for each of those features is mapped to the range $[0, 1]$ that is used to scale an emoji icon that reflects the corresponding music feature best. The exact value is shown on mouseover. The resulting maps are shown in Figure 7. As the students extracted the playlists manually, they restricted their observation to few German cities. They admitted that an appropriate answer to the given comparative analysis interest is hard to be found: the number of chosen cities was subjective, the equal amount of representatives from East and West did not stand for the population ratio, and the number of songs in the playlists differed. However, the maps revealed some interesting information on city level. Music heard in Frankfurt is typically dancable and explicit, but less happy and energetic. In Saxony, people listen to energetic music that is less happy, dancable and explicit.

4 FEEDBACK

After completing the course, I asked journalism participants to fill out an open question questionnaire on positive and negative aspects of the course that was primarily designed for students related to digital humanities. Below, I summarize the collective feedback of six students who all joined different projects.

Which course contents do you consider to be particularly relevant for your future work as a journalist and why?   There was a consensus among the students on the value of the course for them as becoming data journalists. They liked the broad overview of visualization techniques, for what kind of data they can be applied, and how they can support data analysis tasks. Of great importance was the discussion of advantages and disadvantages, and rules to be considered when designing visualizations at the hand of related examples. Lastly, the discussion of interdisciplinary aspects of digital humanities scholarship was seen relevant as journalists who typically cooperate with designers and developers face similar communication problems.

[6] https://spotifymaps.github.io/musicalcities/

Which course contents do you consider to be less relevant?
Students criticized that most of the discussed related visualizations were too scientific and too descriptive, and they wished to include more examples from journalism. Further, some topics of the course were too deeply discussed (e.g., close reading, or temporal data visualizations based on space-time cube operations [2]). This might be due to the fact that some visualizations were seen as less relevant for journalism activities.

What aspects considered to be relevant in the future were missing?   Five students missed practical hands-on sessions throughout the theoretical part of the course. On the one hand, they wished to apply existing visualization tools to data sets—a common work practice of journalists—, on the other hand, the programming of sample snippets was seen valuable before beginning the student projects. In addition, more details on the technical realization of related projects was desired. Finally, one student wished a broader overview of data processing techniques next to Named Entity Recognition (NER) discussed in the course. However, this would require an extension of the course to more than 5 ECTS credits in the future.

To what extent are you satisfied with your project results, and do you see room for improvement?   Overall, students were satisfied as they would never have been able to design related visualizations without the knowledge gained in the course. One representative reply in this regard was *"Based on my level of knowledge at the beginning of the semester, I am very satisfied with the result, as our finished visualization actually fulfills all the functions it was supposed to fulfill."* At least for one project, the visualization of biographies of German Bundestag politicians, a publication is planned.

What conceptual limits did you reach during the project?   All projects faced data acquisition problems. On the one hand, this relates to limited data processing skills, on the other hand, to the limited time frame of around five weeks given to work on the projects. While in one project, numbers needed to be manually extracted from old documents for which OCR failed, other projects suffered from the long response time of data deliveries—seemingly all valuable lessons learned for future projects.

How do you assess the relevance of the steps taken to develop the visualization for your future work as a journalist?   *"The development of visualization has brought me a lot further, because I can now imagine the implementation of such projects much more concretely."*   All students described the value of the project work in a similar way. Most of the participating students joined such projects for the first time, and the course taught them strategies how
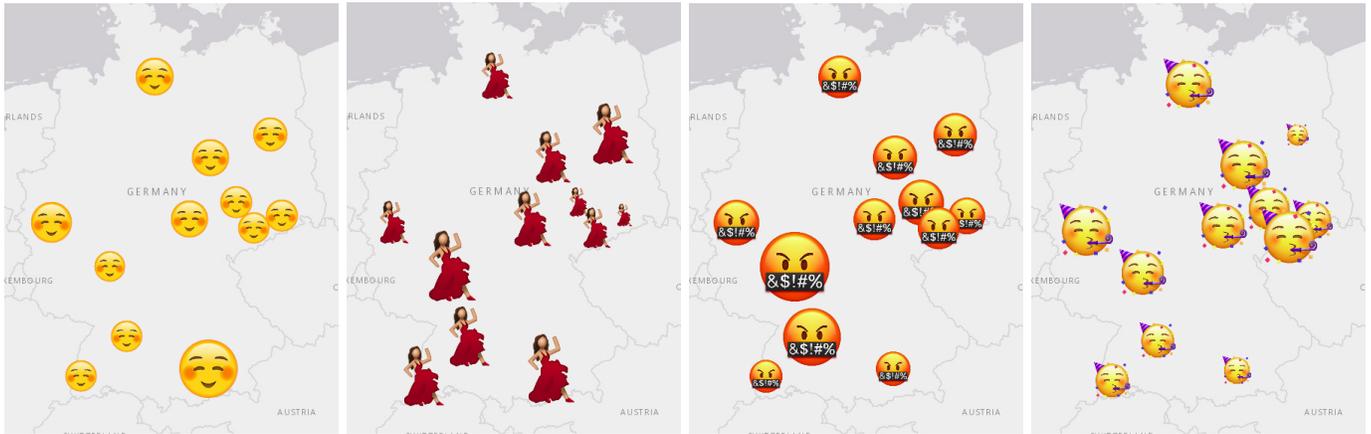
Fig. 7. Favored music styles in 10 German cities.

to approach them. The students valued the high relevance of the conducted projects for their future, also they saw how the undertaken steps could be applied to support other data analysis tasks. Lastly, the exchange with other groups facing similar data processing or visualization tasks was important for them.

Finally, one student remarked that a conceptual model should be proposed for journalists as opposed to (digital) humanities scholars [5]. Figure 1 aims to serve this request.

5 Conclusion

With my experiences in teaching visualization design principles to journalism students who aim to offer visual interfaces to casual users to underline arguments discussed in their articles, I hope to serve others with valuable information when designing a course for a comparable group of students. In general, the course played a significant role considering the future work of the students as data journalists. Theoretical lecture contents should be accompanied with practical hands-on sessions in order to easen the project start for students.

The results illustrated in the figures are the visual designs as presented by the groups on the last course day. Even though most visualizations leave room for improvements due to some inappropriate design decisions, I, personally, was impressed by the quality of project outcomes considering the fact that many students had neither computer science nor visualization skills before joining the course.

One of the key factors was that students were allowed to choose their own project topics. The longer the course took, the more they were able to generate ideas on their own—being interesting from a technical data analysis as well as a societal perspective.

Acknowledgments

References

[1] New York Times: How Every Member Got to Congress, 2019. https://www.nytimes.com/interactive/2019/01/26/opinion/sunday/paths-to-congress.html (Retrieved 2019-07-08).

[2] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A Review of Temporal Data Visualizations Based on Space-Time Cube Operations. In R. Borgo, R. Maciejewski, and I. Viola, editors, *EuroVis - STARs*. The Eurographics Association, 2014.

[3] M. Bastian, S. Heymann, M. Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.

[4] S. Jänicke, J. Blumenstein, M. Rücker, D. Zeckzer, and G. Scheuermann. TagPies: Comparative Visualization of Textual Data. In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: IVAPP,*, pages 40–51. INSTICC, SciTePress, 2018.

[5] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. Visual Text Analysis in Digital Humanities. 36:226–250, 2017.

[6] M. Kaufman. 'Everything on Paper Will Be Used Against Me': Quantifying Kissinger. In *Proceedings of the Digital Humanities 2015*, 2015.

[7] T. Munzner. *Visualization Analysis and Design*. CRC press, 2014.

[8] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Visual Languages, Proceedings*, pages 336–343, 1996.