Visualising the results of the life-events surveys on satisfaction with public administration in Germany

Daniel Kühnhenrich (daniel.kuehnhenrich@destatis.de)
Federal Statistical Office (Destatis)

Sylvana Walprecht (sylvana.walprecht@destatis.de)
Federal Statistical Office (Destatis)

ABSTRACT

Making data understandable via visualisation is an objective of many types of publications. While R offers a vast amount of packages for the task, new R coders may find themselves overwhelmed by the possibilities. This article presents three easy applications. As part of the Federal Government's Better Regulation programme, the Federal Statistical Office of Germany (Destatis) conducts surveys on the satisfaction of citizens and companies with public authorities every two years since 2015. In 2019, Destatis modernised data visualisation in its print publications. A Sankey diagram informs about the usage of digital devices, a word cloud displays reasons of dissatisfaction and a network diagram highlights essential and secondary agencies. They were generated with the R packages networkD3, wordcloud2, and igraph. During the 9th International Conference on the Use of R in Official Statistics (uRos2021), the authors presented these diagrams as part of the "Scientific Session Dissemination and visualization".

Keywords: *public administration, better regulation, Sankey diagram, word cloud, network diagram, life-events surveys, satisfaction*

JEL classification: *H (Public Economics) + Y (Miscellaneous Categories)*

Introduction

Since 2015, the Federal Statistical Office of Germany (Destatis) has conducted life-events surveys every two years as part of the Federal Government's Better Regulation programme. They examine the satisfaction of citizens and companies with public authorities they had contact with over the last two years. In 2019, for example, a random sample of 6,016 citizens and 2,697 companies was interviewed to this end. Their interactions with public authorities are grouped into life events such as childbirth or starting a business. Besides satisfaction, the questionnaire also deals with digitalisation, complexity of procedures and comprehensibility of forms and applications. In 2019, with the objective to modernise the visualisation of the results, Destatis introduced several new types of graphics as

part of its print publications. On 25 January 2021 during the 9th International Conference on the Use of R in Official Statistics (uRos2021), the authors presented these graphics as part of the "Scientific Session Dissemination and visualization" (Universitatea Ecologica Bucuresti, 2021; Walprecht and Kühnhenrich, 2021). This article describes their application in more detail and offers new R coders a place to start in the vast array of available packages. Meanwhile, Destatis (2021) has published the results of the latest surveys.

Methodology

Three types of graphics were introduced into the publications: a Sankey diagram (Kennedy and Sankey, 1898) on the usage of digital devices, a static word cloud (also called tag cloud) (Khusro et al., 2021) for reasons of dissatisfaction and a network diagram (Fleischer and Hirsch, 2001) to highlight essential and secondary agencies in a life event (Destatis, 2019). All three were generated with R, whereas in the surveys before graphics were solely created with SAS and Microsoft Excel.

Sankey diagram

A major focus of the surveys is the digitalisation of public administration. Two questions deal with the citizens' usage of digital devices such as laptops, smartphones or tablets. The first one asked which digital devices respondents employed predominately for private online activities e.g. communication, information, entertainment and shopping, while the other one inquired the medium used for interactions with public authorities. The Sankey diagram in figure 1 illustrates the results in the 2019 survey and compares the usage during these two kinds of activities. The diagram highlights that with 94% the large majority of citizens uses digital devices such as PCs, laptops, smartphones and tablets. However, most of these people filled out paper forms when interacting with public administration. Almost no one employed smartphones and tablets to this end (Destatis, 2019). The Sankey diagram is an appropriate tool to emphasis this discrepancy and to depict the situation for each of the devices in detail.

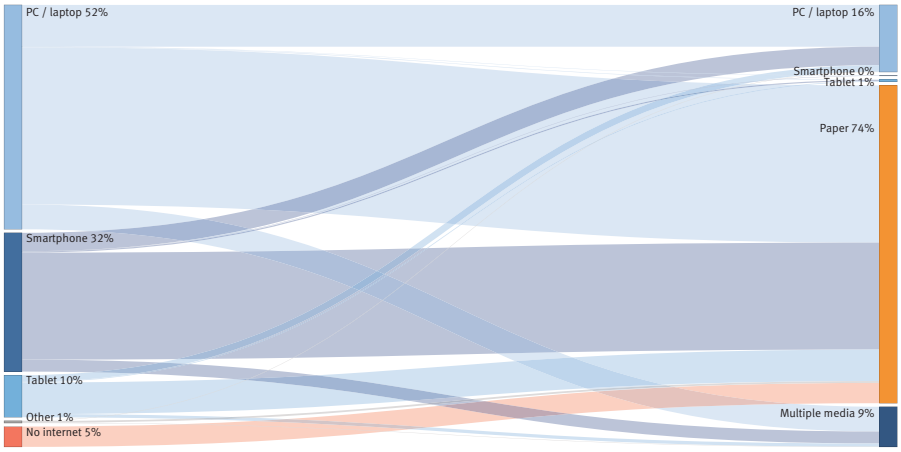

Figure 1: Usage of media for ordinary private activities (left side) and for communicating with public authorities (right side)
Adapted from: Destatis (2019, p. 24).

The **networkD3** package (Allaire et al., 2017) was employed creating the Sankey diagram. It requires two data frames as input:

1. Nodes: it contains the names of the nodes e.g. "PC / laptop" for the left side and "Paper" on the right side.
2. Links: this includes the source numbers of nodes on the left side, target numbers of the nodes on the right side and the respective frequency value e.g. 25% for the link between smartphone users for private online activities and filling out paper forms when in contact with public administration. Furthermore, colour groups were defined.

The `sankeyNetwork` function creates the diagram. It employs the two input data frames i.e. arguments `Links` and `Nodes`, the variables containing the specification of the nodes' names (`Source`, `Target`), the respective frequency values (`Value`) and IDs (`NodeID`). Furthermore, design elements such as the font (`fontFamily`), its size (`fontSize`) and the colour of the nodes and links (`colourScale`) are defined. The arguments `LinkGroup` and `NodeGroup` are necessary to connect the respective nodes with the colour of the data frame in `colourScale`. Finally, the argument `iterations` enables to change the order of the nodes, but is not used in this case.

```
sankeyNetwork(Links = links,
              Nodes = nodes,
              Source = "source",
              Target = "target",
              Value = "value",
              NodeID = "name",
              fontSize= 32,
              fontFamily = "MetaNormalLF-Roman",
              nodeWidth = 50,
              iterations = 0,
              colourScale = my_color,
              LinkGroup = "group",
              NodeGroup = "group")
```

The function produces a JavaScript graphic in an HTML file. Since Destatis published the results as a printed booklet, the HTML file was converted into a vector graphic and slightly adapted prior to publication.

Word cloud

In an open question, respondents who indicated that they had not been satisfied with the respective public authority were asked for the causes of their dissatisfaction. Destatis categorised and coded the 2,895 answer texts manually. The word cloud in figure 2 displays the most important reasons for dissatisfaction with larger terms indicating a higher frequency. It shows that citizens repeatedly criticised the long processing time of their applications. Furthermore, complicated procedures, waiting time at the agency, rejections and unfriendly or seemingly incompetent staff were mentioned often (Destatis, 2019).

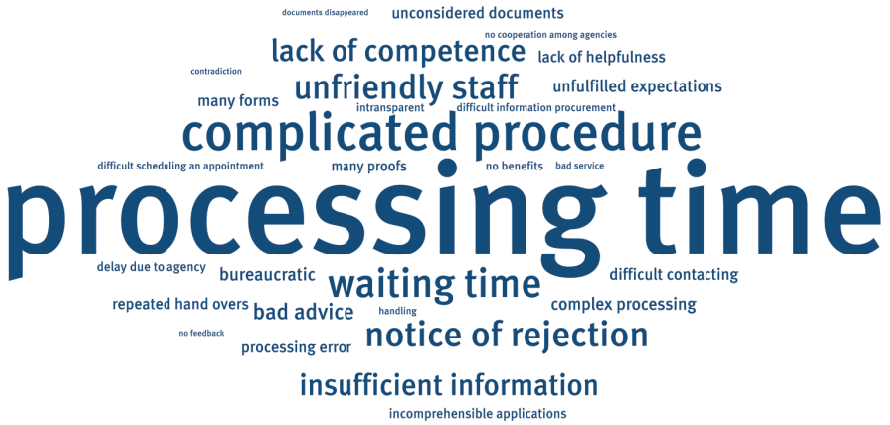

Figure 2: Stated reasons in case of dissatisfaction with public authorities
Adapted from: Destatis (2019, p. 18).

The **wordcloud2** package and its eponymous function were employed to generate the word cloud (Lang and tin Chien, 2018). It requires a data frame containing the terms that are to be displayed and their frequency. A vector was used to define the colours of the words, in this case different shades of blue, while the background colour is white. Since no rotation was needed, the argument `rotateRatio` was set to 0. Furthermore, the font and size of the diagram were specified according to the design guidelines and dimensions required for the print publication.

```
wordcloud2(complains,
           color = colour,
           backgroundColor = "white",
           rotateRatio = 0
           fontFamily = "MetaNormalLF-Roman",
           size = 5, gridSize = 100, widgetsize = c(9600, 5385))
```

Then, the resulting list was exported into an HTML file with the packages **webshot** (Chang, 2019) and **htmlwidgets** (Vaidyanathan et al., 2021). Finally, it was converted into a high-definition PNG file for printing.

Network diagram

At the beginning of the survey, respondents indicated which individual public authorities they had interacted with during the previous two years. With this information, a frequency distribution of contacts was calculated and displayed for the respective life event using a network diagram as shown in figure 3 for the life event "birth of a child". The graphic illustrates that combinations between the registry, parental allowance and family benefits offices, and the statutory health insurance occurred in the majority of cases. Hence, these four authorities play a key role during this life event. New parents interact with the three other agencies – statutory pension insurance, job centre and youth welfare office – less frequently, since they are only important under certain circumstance, e.g. the youth welfare office is responsible for the acknowledgement of paternity of unmarried fathers.

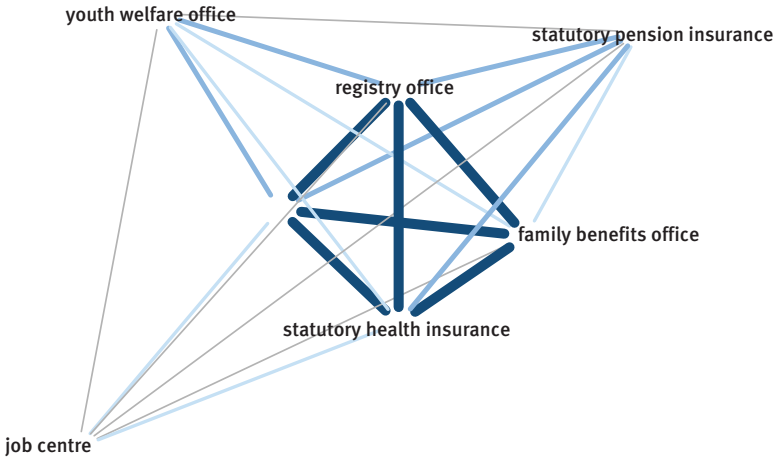

Figure 3: Frequency of combinations of contacts with public authorities in the life event "birth of a child"
Adapted from: Destatis (2019, p. 83).

This diagram was generated using the **igraph** package (Csardi and Nepusz, 2006). It requires a data frame containing two variables describing every possible combination – in the case of 7 public authorities 21 combinations ($= \binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{7!}{(7-2)!2!)}$). A third variable includes the frequency of these contacts occurring together.

The `graph.edgelist` function creates a graph with the public authorities as vertexes i.e. nodes and uses no arrows at the end of the edges i.e. simple lines (`directed = FALSE`) for their connections. The command `E(g)$weight` refers to the vector with the frequencies of each combination. Based on that the thickness and colour of each line are allocated in the two new vectors `E(g)$edge.width` and `E(g)$edge.color`. Finally, the `plot` function creates the network diagram by specifying all needed arguments. Vertexes i.e. nodes are set invisible in this example. Only their label is printed. Since the network diagram is part of the printed booklet, a PDF file was generated.

```
network_mat <- as.matrix(data);
g <- graph.edgelist(network_mat [,1:2],directed = FALSE);
E(g)$weight = as.numeric(network_mat [,3]);
E(g)[weight <= 10 & weight > 0]$edge.color <- "#b3b3b3";
E(g)[weight <= 10 & weight > 0]$edge.width <- 1;

. . .

plot(g, vertex.shape = "none",
        vertex.frame.color = "#ffffff",
        vertex.label.color = "black",
        edge.width = E(g)$edge.width,
        edge.color = E(g)$edge.color,
        vertex.label.family = "mtnolfro"
        asp = 0
);
```

Conclusion

This article described three easy uses to illustrate survey results about the satisfaction with public authorities in Germany. Employing modern and innovative diagrams helps to explain complex relationships between variables or entities. In 2021, Destatis moved away from print to solely digital publications of the life-events survey results (Destatis, 2021). Due to the technical specifications of the website, all graphics were programmed in Highcharts (Highsoft, 2022) – a JavaScript library. Nevertheless, these first applications in R were decisive, since the current interactive implementations were built on those experiences. Furthermore, in R data management, data analysis and graphical illustration stem from one source which is a major advantage especially in explorative data analysis. Finally, the R packages presented allow users to create similar graphics for other purposes and to employ further specifications.

Acknowledgements

The authors wish to thank Carsten Hornbach and Susanne Michalik (both Destatis) for their valuable remarks. Furthermore, the authors wish to thank the uRos2021 organising committee for the invitation to the conference.

Bibliography

J. Allaire, C. Gandrud, K. Russell, and C. Yetman. *networkD3: D3 JavaScript Network Graphs from R*, 2017. URL https://CRAN.R-project.org/package=networkD3. R package version 0.4. [p2]

W. Chang. *webshot: Take Screenshots of Web Pages*, 2019. URL https://CRAN.R-project.org/package=webshot. R package version 0.5.2. [p3]

G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL https://igraph.org. [p4]

Destatis. *Zufriedenheit der Bürgerinnen und Bürger mit behördlichen Dienstleistungen: Ausgewählte Ergebnisse der Lebenslagenbefragung 2019*, 2019. URL https://www.amtlich-einfach.de/SharedDocs/Downloads/Ergebnisse_Buerger_2019.pdf?__blob=publicationFile&v=2. Accessed on 5 May 2022. [p1, 2, 3, 4]

Destatis. *Zufriedenheit der Bürgerinnen und Bürger mit behördlichen Dienstleistungen: Überblick und weitere themenbezogene Detailauswertungen*, 2021. URL https://www.amtlich-einfach.de/DE/Ergebnisse/Buerger2021/Ueberblick_node.html. Accessed on 5 May 2022. [p1, 4]

R. Fleischer and C. Hirsch. Graph drawing and its applications. In M. Kaufmann and D. Wagner, editors, *Drawing Graphs: Methods and Models*, pages 1–22. Springer, Berlin, Heidelberg, 2001. [p1]

Highsoft. *Highcharts Documentation*, 2022. URL https://www.highcharts.com/docs/index. Accessed of 5 May 2022. [p5]

A. B. W. Kennedy and H. R. Sankey. The thermal efficiency of steam engines. report of the committe appointed to the council upon the subject of the definition of a standard or standards of thermal efficiency for steam engines: with an introductory note. *Minutes of the Proceedings of the Institution of Civil Engineers*, 134(1898):278–312, 1898. [p1]

S. Khusro, F. Jabeen, and A. Khan. Tag clouds: Past, present and future. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, 91(2):369–381, 2021. [p1]

D. Lang and G. tin Chien. *wordcloud2: Create Word Cloud by 'htmlwidget'*, 2018. URL https://CRAN.R-project.org/package=wordcloud2. R package version 0.2.1. [p3]

Universitatea Ecologica Bucuresti. *uRos2021 - Scientific Session Dissemination and visualization (Day 2 - session 2)*, 2021. URL https://youtu.be/kjUtApjK6XE?t=2100. Accessed on 5 May 2022. [p1]

R. Vaidyanathan, Y. Xie, J. Allaire, J. Cheng, C. Sievert, and K. Russell. *htmlwidgets: HTML Widgets for R*, 2021. URL https://CRAN.R-project.org/package=htmlwidgets. R package version 1.5.4. [p3]

S. Walprecht and D. Kühnhenrich. *Modern graphics for presenting the results of life-events surveys on satisfaction with public administration*, 2021. URL https://r-project.ro/conference2021/book-of-abstracts-uros2021.pdf#page=21. Accessed on 5 May 2022. [p1]