The ‘Work and Corona’ Dashboard

Fabian Kalleitner^{*1}, Lukas Schlögl² und Licia Bobzien³

Abstract

Open-science efforts often remain limited to the scientific community. Attempts to increase the audience, focus largely on communicating selected research results. The Work-and-Corona (WoCo) project aims to exemplify how quantitative social science projects can make their data accessible, thus connecting open science with public communication of scientific research. The *Work and Corona* (WoCo) project provides flexible, free and easy to use access to public opinion data in the context of the COVID-19 pandemic in Austria. To that end, we programmed a data visualisation dashboard (<https://woco.univie.ac.at/dashboard/>) that displays aggregated survey data from the *Austrian Corona Panel Project*, an academically designed public opinion survey. This paper describes the dataset used, the process of data aggregation and processing and a set of technical details on the programming of the dashboard. The paper further describes the workflow of the project enabling other researchers to provide similar forms of public data access. We provide full access to our code.

Keywords: COVID-19, social sciences, open data, dashboard, survey

* correspondent Author - fabian.kalleitner@univie.ac.at

¹ University of Vienna, Department of Economic Sociology, Vienna, Austria

² University of Vienna, Department of Political Science, Vienna, Austria

³ Hertie School, Berlin, Germany

Introduction

Over the last decade, the open science movement contributed immensely to the advancement of reliable, transparent and verifiable research. Currently, more researchers than ever publish their scientific papers via openly accessible archives and research funders increasingly demand research to be published open access (McKiernan et al. 2016). Further, more and more journals also require researchers to provide access to the datasets and the code used in their papers (e.g. policy of Public Opinion Quarterly, Journal of Politics, American Economic Association). The COVID-19 pandemic has accelerated this trend in scientific publishing standards (see e.g. Daly et al. 2020). It has further pushed co-operative and collaborative research, the bundling of resources, and has helped speed up the process from research design to publication.

Despite these advances, the audience of the open science movement still remains limited to the scientific community. Attempts to increase the audience focus largely on communicating selected research results to the broader public via scientific blogs, policy papers or contributions to media. Although the pandemic led to an increase in the visibility of such formats, these developments have not yet been accompanied by an increase in easily accessible scientific data. While many official and private actors provide information on COVID-19 infections and vaccinations via public dashboards, the same ‘data revolution’ has not yet materialised to the same extent in other areas of data collection.

Despite early innovators like Hans Rosling (a public health expert and father of the visualisation tool ‘Gapminder’) and the constant rise in data journalism, the scientific community has made only limited efforts in providing easy to use data access points. A famous exception is ourworldindata.org (2021), a data visualisation website curated by a team surrounding Max Roser (University of Oxford). With an easy-to-use dashboard and transparent data sources, this tool enables a broad audience to access and interact with aggregated data. The site provides researchers with country-level macro data on a range of social, economic and public health topics such as Covid-19 vaccination rates (Mathieu et al. 2021). *OurWorldInData* itself builds on a revolution in data sharing policies by national statistical offices, which also resulted in open-access digital databases such as the World Bank’s *World Development Indicators* or the *OECD.Stat* database.

OurWorldInData highlights the potential in publicly accessible easy to use data visualizations. However, to date, its focus is mostly on so-called objective “hard facts”, provided by statistical agencies, such as income, work hours, COVID-infections and vaccination numbers or aggregated

measures of these entities such as GDP, or Gini coefficients and other measures (there are some exemptions; see e.g. article on trust). Research studying the COVID-19 pandemic, though, shows that subjectively experienced “soft factors”, such as attitudes, perceptions, beliefs and norms are important to understanding the evolution of the pandemic (Bavel et al. 2020). Not only what people *do* in the pandemic but also what they *think* is important to understanding how individuals and societies deal with the pandemic.

The *Work and Corona* project aims to showcase how academically designed, quantitative social research projects can make their data accessible, thus connecting the open science movement with public communication of scientific research. We do that by creating an easy-to-access data dashboard, flanked by metadata, a tutorial and other information. WoCo takes inspiration from initiative such as the Cosmo data explorer project at University of Erfurt which also aims at providing easy access to data on attitudes during the COVID-19 pandemic (COSMO 2021). Similar to Cosmo, WoCo goes beyond existing tools for online analyses in the social sciences that remain difficult to use and remain limited to calculating summary statistics (see e.g. the data analysis tool of the European Social Survey ESS 2021). The remaining paper explains the specific goals of the WoCo project, describes the data used in the dashboard, the data cleaning and recoding process, and the main features of the dashboard and its statistical background.

Project and Data

The goal of the Work and Corona (WoCo) project is to provide flexible, easy to use public access to scientific public perception data in the context of the COVID-19 pandemic. To do that, we programmed a data dashboard that utilizes aggregated survey results and made this dashboard available within a web environment containing further low-threshold information on the use and interpretation of data.

We draw on a social-science dataset provided by the Austria Corona Panel Project (ACPP) of the University of Vienna. The ACPP team collected panel data from a sample of 1,500 respondents per wave. Sampling includes Austrian inhabitants between the ages of 14-75 via an online access panel by the survey company Marketagent. The sample is quota sampled by region (Bundesland), gender, age-group, and level of education. Further details on the dataset, weighting, and data accuracy are available at Kittel et al. (2021).

The weekly field periods of ACPP started at the end of March 2020 shortly after the first Covid “lockdown” in Austria. After ten weekly survey waves the panel switched to bi-weekly and then to monthly survey intervals (with a gap July-October 2022). At the time of writing this article (December 2021), the dataset includes 27 waves and answers provided by a total of 3,158 unique respondents. The dataset is the largest of its kind in Austria providing public opinion data during the COVID-19 pandemic. The project team provided quick early access to the dataset and published edited datasets for scientific and open use (Kittel et al. 2020) along several Blog-posts detailing selected results (ACPP 2021). With these measures, the ACPP already set high standards of transparency and data sharing. The project team also published a set of scientific papers on various topics (e.g. Bobzien and Kalleitner 2021; Eberl, Huber, and Greussing 2021; Kittel, Kalleitner, and Schiestl 2021; Kritzing et al. 2021; Paul, Eberl, and Partheymüller 2021).

The first Version (V1.0) of the WoCo dashboard published in December 2021 features 100 selected variables of the ACPP-dataset covering a broad thematic range and a long time span. The variables were selected using the following criteria: i) *temporal depth of coverage*: Questions had to have been included in most of the survey waves as the dashboard is intended to allow the user to observe changes over time rather than simple descriptive cross sections. Thus, we selected mostly variables from the ACPP ‘core’ questionnaire or of question modules that were asked repeatedly. ii) *coverage of the sample*: Questions should have been asked to the full sample of respondents rather than to subgroups. To ease interpretation for users, we selected variables that refer to all respondents in the sample. This is especially important since rapid changes of subsamples (e.g. the subgroup of those not vaccinated at point t is likely to be different from the subgroup of those not vaccinated at point $t+1$) could affect response distributions. We make two exceptions to this criterion and include questions asked only to those with children and to those employed/or of working age. Both subsamples change slowly which reduces potential inference issues due to varying sample compositions. iii) *computability and aggregation*: Lastly, the questions should use standard Likert-type answer scales to enable calculation of wave or subgroup averages assuming these scales are interval scaled.

Survey data can contain non-substantive responses such as refusals to answer the question (“no answer” or simply skipped questions) or no opinion answers like “don’t know”. We decided to apply pairwise deletion of these non-substantive responses. We do that to provide constant sample sizes per variable in all functions of the dashboard. Further, we provide links to the survey questionnaire including the full range of answer categories in the variable selection. In addition,

we display the non-substantive response categories (“don’t know” and “no answer”) in a separate information window when we calculate (wave or subgroup) averages. In all visualizations we provide information on the effective sample sizes, excluding those providing non-substantive responses. This also accounts for non-substantive responses in the categorical variables used to calculate subgroup averages (this concerns mostly the income variable).

Dashboard features and methods

The COVID-19 dashboard features four different forms of data visualizations: category shares within each wave [area chart]; averages per wave; averages per wave by subcategory [line charts]; and cross-sectional bar charts of distribution answer categories within one selected wave [bar chart].

Category shares

To calculate the shares of answers within each answer category and survey wave, we calculate the weighted shares of effective answers (excluding “don’t know” and refusals). We use the demographic weights provided in the ACPP dataset (for details see Kittel, Kritzing, et al. 2021). The chart uses respondents’ median start date of the survey to mark the field period of the survey wave (x-axis). This enables us to accurately account for the changing survey intervals over the course of the study. We use a directional colour scheme (red-yellow-blue) to display the categories as all of our variables use directed Likert scales. Users can select what variable(s) they wish to display and can change the direction of the colour scheme.

Averages

This dashboard function displays the weighted average answers within survey waves. Users can select multiple variables and can compare the trajectories of survey means by wave. We automatically recode every scale so that it starts at 0 and increases in value with regard to the name of the variable (e.g. threat perception: high threat > low threat; life satisfaction: very satisfied > not at all satisfied). We provide several additional features and designed the graph to reduce a user’s likelihood to overestimate small between-wave or between-variable differences. First, we enable users to display 95% confidence intervals in the form of whiskers to get a first impression of the minimal statistical variation within the data. Second, we fix the y-axis to start at 0 by default to reduce the likelihood for users to misinterpret the y-axis. Third, we display information on the answer scales of the variables below the graph. These explanations also include a warning message if users select two variables with different answer scales. The graph also allows users to display the

7-day incidence rate of COVID-19 cases in Austria. To calculate these, we use data from the federal Austrian open data portal and the official population of Austria 2020 (Open Data Austria 2021; Statistik Austria 2021). The dashboard also allows users to display important events during the pandemic. We selected the same events that are displayed in the COVID-19 cases dashboard of Austria's public tv broadcaster ORF (ORF.at 2021).

Averages by category

To enable users to get a glimpse into the heterogeneous impact and development of the COVID-19 pandemic, we allow users to differentiate averages between subgroups. Due to the limited sample size per wave, we restrict the number of subgroups. We use categorization by gender (male, female); age (<30, 30-59, >=60); highest level of education (primary, vocational education [Lehre], secondary, tertiary); inhabitants of municipality (<10.000, 10.000-1 Million, >1 Million); disposable household income (low: 1st-3rd decile, middle: 4th-7th decile, high: 8th-10th decile); and pandemic labor-market status (employed as usual, working from home, in furlough [Kurzarbeit], unemployed, not working). Again, users can display 95% confidence intervals and sample sizes are corrected for missing values within selected variables and categorizations.

Cross-section

This dashboard function displays the weighted answer shares of a single variable within the user-selected wave via a bar graph. Users can switch the direction of the colour scheme. Furthermore, users have the option to display an animation that runs through the available wave categories - in essence, a short video showing the changes of the frequency of answer categories in a selected variable wave by wave.

Open code and data use

We utilize an R-Shiny-framework to program the WoCo dashboard (see software section below for detailed references of the software and packages used in the project). Thus, every user has their own instance of R that displays the data. We use a commercial server provided by shinyapps.io to run the R instances. For reasons of data security and to avoid any chance of deidentification of respondents, we decided to not upload any micro data onto external servers. Instead, we do the data processing offline, aggregate the data, and upload only aggregated datasets to a publicly available and thus transparent github-page (<https://github.com/fabiankal/woco>). This also reduces the server processing power needed and decreases the size of the dataset and, thus, the

loading times. The Dashboard itself is imported into a website, hosted by the University of Vienna, via an iframe that scales between different user devices (desktop, mobile).

Data processing is done in three steps. First, the dataset provided @AUSSDA is recoded in a long-format to better fit the R ggplot framework (using the R package panelr). Second, we subset the dataset and manually code direction of the Likert scales to enable the automatic calculation of averages. Third, we apply demographic weights (srvyr), aggregate the data and upload these datasets on github. The code is available on github as well and mostly sticks to the tidy framework.

The WoCo dashboard itself uses a flexdashboard layout. The graphs are either plotted with ggplot2 and then made interactive with ggplotly or fully programmed within the plotly framework. We wrote the code with performance, scalability and adaptability in mind. We intend to update the dashboard when future survey waves of the ACPP become available and increase the range of variables. Furthermore, other researchers can use the code to create aggregated datasets and dashboards of nearly every public opinion data with only a few adaptations of the code.

WoCo actively promotes the use of the data and graphs by others. The visualizations generated by the Work and Corona Dashboard can be shared if appropriate credit is given (Lizenz [CC BY 3.0 AT](#)).

Literature

ACPP. 2021. ‘Corona-Blog’. Retrieved 15 December 2021 (<https://viecer.univie.ac.at/corona-blog/>).

Bavel, Jay J. Van, Katherine Baicker, Paulo S. Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J. Crockett, Alia J. Crum, Karen M. Douglas, James N. Druckman, John Drury, Oeindrila Dube, Naomi Ellemers, Eli J. Finkel, James H. Fowler, Michele Gelfand, Shihui Han, S. Alexander Haslam, Jolanda Jetten, Shinobu Kitayama, Dean Mobbs, Lucy E. Napper, Dominic J. Packer, Gordon Pennycook, Ellen Peters, Richard E. Petty, David G. Rand, Stephen D. Reicher, Simone Schnall, Azim Shariff, Linda J. Skitka, Sandra Susan Smith, Cass R. Sunstein, Nassim Tabri, Joshua A. Tucker, Sander van der Linden, Paul van Lange, Kim A. Weeden, Michael J. A. Wohl, Jamil Zaki, Sean R. Zion, and Robb Willer. 2020. ‘Using Social and Behavioural Science to Support COVID-19 Pandemic Response’. *Nature Human Behaviour* 4(5):460–71. doi: 10.1038/s41562-020-0884-z.

Bobzien, Licia, and Fabian Kalleitner. 2021. ‘Attitudes towards European Financial Solidarity during the Covid-19 Pandemic: Evidence from a Net-Contributor Country’. *European Societies* 23(sup1):S791–804. doi: 10.1080/14616696.2020.1836669.

- COSMO. 2021. 'COVID-19 Snapshot Monitoring (COSMO) - Wwww.Corona-Monitor.De'. Retrieved 15 December 2021 (<https://projekte.uni-erfurt.de/cosmo2020/web/explorer/>).
- Daly, M., Bernhard Ebbinghaus, Lukas Lehner, M. Naczyk, and T. Vlandes. 2020. 'Oxford Supertracker: The Global Directory for COVID Policy Trackers and Surveys, Department of Social Policy and Intervention.' *Oxford Supertracker*. Retrieved 15 December 2021 (<http://supertracker.spi.ox.ac.uk/>).
- Eberl, Jakob-Moritz, Robert A. Huber, and Esther Greussing. 2021. 'From Populism to the "Plandemic": Why Populists Believe in COVID-19 Conspiracies'. *Journal of Elections, Public Opinion and Parties* 31(sup1):272–84. doi: 10.1080/17457289.2021.1924730.
- ESS. 2021. 'Nesstar WebView'. Retrieved 15 December 2021 (<http://nesstar.ess.nsd.uib.no/webview/>).
- Kittel, Bernhard, Fabian Kalleitner, and David W. Schiestl. 2021. 'Peers for the Fearless: Social Norms Facilitate Preventive Behaviour When Individuals Perceive Low COVID-19 Health Risks'. *PLOS ONE* 16(12):e0260171. doi: 10.1371/journal.pone.0260171.
- Kittel, Bernhard, Sylvia Kritzing, Hajo Boomgaarden, Barbara Prainsack, Jakob-Moritz Eberl, Fabian Kalleitner, Noëlle S. Lebernegg, Julia Partheymüller, Carolina Plescia, David W. Schiestl, and Lukas Schlogl. 2020. 'Austrian Corona Panel Project (SUF Edition)'.
- Kittel, Bernhard, Sylvia Kritzing, Hajo Boomgaarden, Barbara Prainsack, Jakob-Moritz Eberl, Fabian Kalleitner, Noëlle S. Lebernegg, Julia Partheymüller, Carolina Plescia, David W. Schiestl, and Lukas Schlogl. 2021. 'The Austrian Corona Panel Project: Monitoring Individual and Societal Dynamics amidst the COVID-19 Crisis'. *European Political Science* 20(2):318–44. doi: 10.1057/s41304-020-00294-7.
- Kritzing, Sylvia, Martial Foucault, Romain Lachat, Julia Partheymüller, Carolina Plescia, and Sylvain Brouard. 2021. "Rally Round the Flag": The COVID-19 Crisis and Trust in the National Government'. *West European Politics* 44(5–6):1205–31. doi: 10.1080/01402382.2021.1925017.
- Mathieu, Edouard, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. 2021. 'A Global Database of COVID-19 Vaccinations'. *Nature Human Behaviour* 5(7):947–53. doi: 10.1038/s41562-021-01122-8.
- McKiernan, Erin C., Philip E. Bourne, C. Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, Brian A. Nosek, Karthik Ram, Courtney K. Soderberg, Jeffrey R. Spies, Kaitlin Thaney, Andrew Updegrove, Kara H. Woo, and Tal Yarkoni. 2016. 'How Open Science Helps Researchers Succeed' edited by P. Rodgers. *ELife* 5:e16800. doi: 10.7554/eLife.16800.
- Open Data Austria. 2021. 'COVID-19: Zeitliche Darstellung von Daten zu Covid19-Fällen je Bundesland - data.gv.at'. Retrieved 15 December 2021 (<https://www.data.gv.at/katalog/dataset/ef8e980b-9644-45d8-b0e9-c6aaf0eff0c0>).
- ORF.at. 2021. 'Lage in Österreich: Coronavirus in Österreich: Täglich aktuelle Daten, Zahlen, Karten und Grafiken'. Retrieved 15 December 2021 (<https://orf.at/corona/daten/>).
- Ourworldindata. 2021. 'Ourworldindata.Org'. *Our World in Data*.

Paul, Katharina T., Jakob-Moritz Eberl, and Julia Partheymüller. 2021. 'Policy-Relevant Attitudes Toward COVID-19 Vaccination: Associations With Demography, Health Risk, and Social and Political Factors'. *Frontiers in Public Health* 9:921. doi: 10.3389/fpubh.2021.671896.

Statistik Austria. 2021. 'Bevölkerung Zu Jahres-/Quartalsanfang'. Retrieved 15 December 2021 (https://www.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bevoelkerung/bevoelkerungsstand_und_veraenderung/bevoelkerung_zu_jahres-_quartalsanfang/index.html).

Software

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Wickham et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

Greg Freedman Ellis and Ben Schneider. 2021. srvyr: 'dplyr'-Like Syntax for Summary Statistics of Survey Data. R package version 1.1.0. <https://CRAN.R-project.org/package=srvyr>

Long JA. 2020. panelr: Regression Models and Utilities for Repeated Measures and Panel Data. R package version 0.7.3, <https://cran.r-project.org/package=panelr>

C. Sievert. 2020. Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida.

Richard Iannone, JJ Allaire and Barbara Borges. 2020. flexdashboard: R Markdown Format for Flexible Dashboards. R package version 0.5.2. <https://CRAN.R-project.org/package=flexdashboard>

Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert and Barbara Borges. 2021. shiny: Web Application Framework for R. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>