Bibliothek der Universität Zürich | ZB Zentralbibliothek Zürich

Universität Zürich[UZH]

Create data documentation for your research data

Coffee Lectures HS 2024

Dr. Melanie Röthlisberger

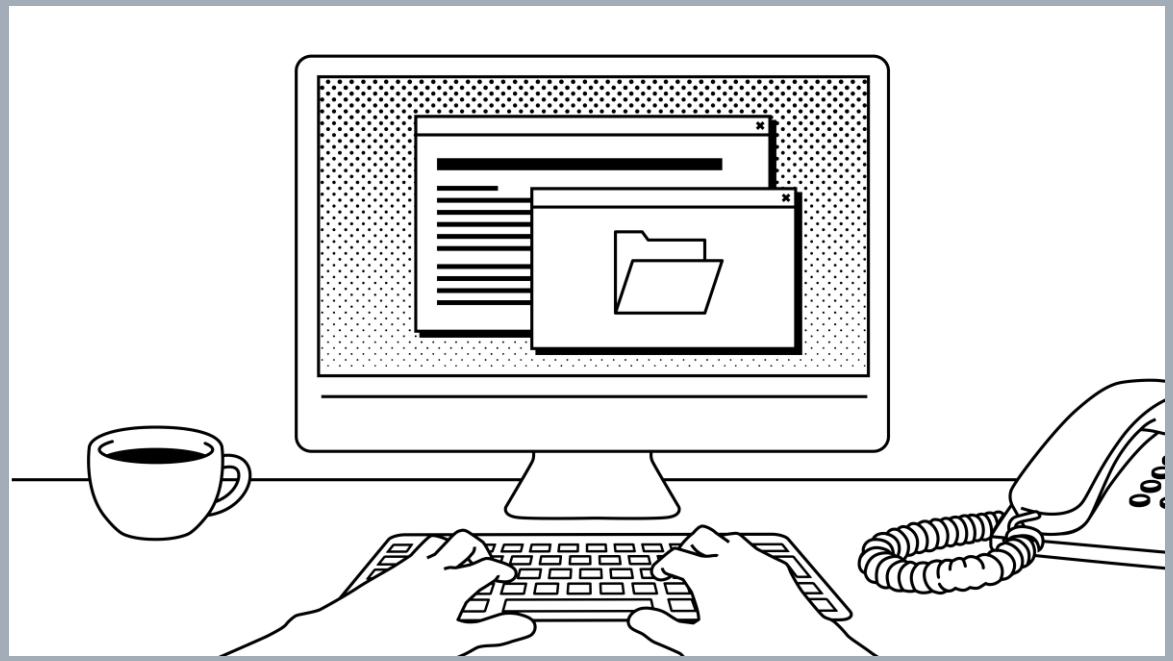
More information:

https://www.ub.uzh.ch/en/wissenschaftlich-arbeiten/mit-daten-arbeiten/daten-dokumentieren.html

Bibliothek der Universität Zürich | ZB Zentralbibliothek Zürich

Universität Zürich[UZH]

I use a dataset for my research paper. How should I describe this dataset in my paper?



Source:
Zentrale Informatik, SIVIC/MELS. (2021). Illustrations of the Game "Open Up Your Research". Zenodo. https://doi.org/10.5281/zenodo.5707726, CC-BY-NC-SA 4.0 International

Why is data documentation important?

When is data documentation important?

When you want to REUSE existing data because only with good data documentation...

… can you verify the quality and usefulness of a dataset.

… can you understand the dataset.

When you collect your OWN data because only with good data documentation ...

… can others understand your data (and how you processed it).

… can you ensure the quality of your data.

… can you reuse the data in the future or make it potentially reusable for others.

Icons by Freepik and Eucalyp at flaticon.com.

What is good data documentation?

Who has created the data?

What is in the data?

When was the data created / collected?

Where was the data created / collected?

How was the data created and processed?

Why was the data created?

Icons by Freepik at www.flaticon.com.

Example of data documentation in a repository

Effects of short photoperiod and carbohydrate consumption on sleep, liver steatosis, and the gut microbiome in diurnal grass rats

Shankar, Anusha, Cornell University

Deal, Cole, Colorado State University, https://orcid.org/0000-0003-1677-2729

McCahon, Shelby, University of Alaska Fairbanks

Callegari, Kyle, University of Alaska Fairbanks

Seitz, Taylor, University of Alaska Fairbanks

Yan, Lily, Michigan State University

Drown, Devin, University of Alaska Fairbanks

Williams, Cory, Colorado State University System

nushiamme@gmail.com, coledeal@colostate.edu, smccahon@uidaho.edu, tjseitz@alaska.edu, yanl@msu.edu, dmdrown@alaska.edu, cory.williams@colostate.edu

Published Dec 18, 2023 on Dryad. https://doi.org/10.5061/dryad.n2z34tn3g

Cite this dataset

Shankar, Anusha et al. (2023). Effects of short photoperiod and carbohydrate consumption on sleep, liver steatosis, and the gut microbiome in diurnal grass rats [Dataset]. Dryad. https://doi.org/10.5061/dryad.n2z34tn3g

Abstract

Seasonal affective disorder (SAD) is a recurrent depression triggered by exposure to short photoperiods, with a subset of patients reporting hypersomnia, increased appetite, and carbohydrate craving. Dysfunction of the microbiota–gut–brain axis is frequently associated with depressive disorders, but its role in SAD is unknown. Nile grass rats (Arvicanthis niloticus) are potentially useful for exploring the pathophysiology of SAD, as they are diurnal and have been found to exhibit anhedonia and affective-like behavior in response to short photoperiods. Further, given grass rats have been found to spontaneously develop metabolic syndrome, they may be particularly susceptible to environmental triggers of metabolic dysbiosis. We conducted a 2x2 factorial design experiment to test the effects of short photoperiod (4h:20h Light:Dark (LD) vs. neutral 12:12 LD), access to a high concentration (8%) sucrose solution, and the interaction between the two, on activity, sleep, liver steatosis, and the gut microbiome of grass rats. We found that animals on short photoperiods showed disrupted activity and sleep patterns but maintained robust diel rhythms and similar subjective day lengths as controls in neutral photoperiods. We found no evidence that photoperiod influenced sucrose consumption. By the end of the experiment, some grass rats were overweight and exhibited signs of non-alcoholic fatty liver disease (NAFLD) with micro- and macro-steatosis. However, neither photoperiod nor access to sucrose solution significantly affected the degree of liver steatosis. The gut microbiome of grass rats varied substantially among individuals, but most variation was attributable to parental effects and the microbiome was unaffected by photoperiod or access to sucrose. Our study indicates short photoperiod leads to disrupted activity and sleep in grass rats but does not impact sucrose consumption or exacerbate metabolic dysbiosis and NAFLD.

README: SAD rats: Effects of short photoperiod and carbohydrate consumption on sleep, liver steatosis, and the gut microbiome in diurnal grass rats

README: SAD rats: Effects of short photoperiod and carbohydrate consumption on sleep, liver steatosis, and the gut microbiome in diurnal grass rats

General Info

Dataset contains 4 zipped files, corresponding to data used for analysis. AnimalData.zip contains data files (in .csv format) to analyze data corresponding to animal weights, fat pad data, liver steatosis data, and phase 1 and phase 2 animal metadata. ConsumptionData.zip contains all data (in .csv format) necessary for analyze sucrose/water consumption stats. microBiome.zip contains data files (in .csv, .gz, and .txt format) necessary to analyze fecal/intestine microBiome data. SleepData.zip contains all data to analyze sleep data obtained from the piezo system.

We used a 2x2 experimental design, with animals subjected to either a short (4h Light: 20h Dark [4:20 LD]) or neutral (12:12 LD) photoperiod and provided with either access to 8% sucrose or only water. Room temperatures were always maintained at 22°C. Grass rats were given ad libitum food (Prolab RMH 2000 5P06 rodent chow from PMINutrition International, St. Louis, Missouri, USA) and ad libitum water access from a water bottle. Every three days, slip cages and bottles were changed, and animals were weighed. Prior to the experiment, animals were co-housed with littermates on a neutral photoperiod (lights on: 05:00-17:00). Animals were then moved into individual cages and provided two weeks to acclimate to their experimental rooms prior to altering photoperiod. After acclimation, we kept 23 rats (11 males, 12 females) on this neutral photoperiod, and shifted 22 rats (10 males, 12 females) to the short photoperiod treatment (lights on: 09:00-13:00). Four weeks later, half of the animals were given a second bottle with sucrose solution whereas the other half were given a second bottle of water. At the end of the experiment, animals were euthanized (using CO2), weighed, fat pads were extracted and weighed, liver samples were taken, and fecal pellets were extracted from the large intestine. Liver samples were fixed in 10% buffered formalin, trimmed, and then stained with a hematoxylin & eosin (H&E) stain. Fecal samples from the large intestine were immediately frozen on dry ice and then stored at -80°C for further analyses. Fecal samples were also collected directly from animal tubs (Trial 2 only) for microbiome analyses immediately prior to photoperiod treatment (0 weeks), after 4 weeks of photoperiod treatment, and at the end of the experiment after 7 weeks of photoperiod treatment and 3 weeks of sucrose treatment.

Description of the data and file structure

All data filenames are referenced in the appropriate code (upload on github: https://github.com/ckdeal/Grass_Rats)

Data files:

AnimalData:

Data in this folder are used to calculate animals weight data, fat pad data, liver steatosis data. Folder has metadata files to correlate animals to other experimental information.

Animal_weights_forMassChange.csv
Chamber_Animal_IDs_Sex_bothPhases.csv
Fat_pads.csv
Merged_Liver.csv
Meta_litter_parents.csv

Source: Screenshots of a dataset, to be found at https://doi.org/10.5061/dryad.n2z34tn3g, CC0 1.0 Universal.

Example of data documentation in a seminar paper / thesis paper

Corpora & Data

The International Corpus of English (ICE) corpus comprises written and spoken texts (and various genres) from various English-speaking countries. For this study, data was collected from nine ICE corpora. The Corpus of Global web-based English (GloWbE) is a web-based corpus that includes texts from websites (news articles, blogs, forums) across the world.

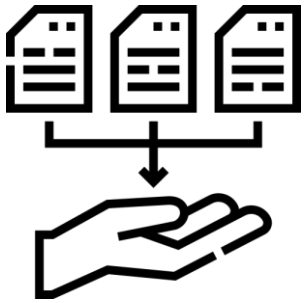
- **Verb List**: A list of 86 alternating dative verbs was compiled from previous literature.
- **Corpus Extraction**: A python script used this verb list to select all constructions with a verb followed by two noun phrases or pronouns, with an optional 'to' between them.
- **Filtering False Positives**: Observations that looked like dative constructions but did not contain a verb, recipient, and theme were excluded.
- **Defining Variable Context**: The dataset was restricted to dative observations that occurred in a context where the alternating variant would be grammatically acceptable and semantically equivalent.
- **Annotation**: Each dative variant was annotated for language-external and language-internal factors, such as register, genre, file number, speaker number, text number, mode, and corpus.

This process resulted in a dataset of 13,171 dative tokens to be analyzed. The dataset was then used to explore variation in the dative alternation across different English varieties.

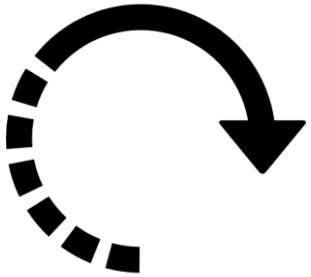
Statistical Tools: Describes the statistical toolkit used for analysis, including random forest analysis, mixed-effects logistic regression, and dialectometric techniques. For each tool (e.g. library) used, it mentions the version used.

Methodological Challenges: Highlights potential shortcomings (data bias, annotation subjectivity, generalizability) and advantages (quantitative approach, scope of dataset) of the statistical methods applied in the study.

How to create data documentation



Start during data collection



Fairly complete

Keep notebook

Include readme

Follow standards

Word-template by Cornell university

https://www.ub.uzh.ch/en/wissenschaftlich-arbeiten/mit-daten-arbeiten/daten-dokumentieren.html

CESSDA data management expert guide

https://dmeg.cessda.eu/Data-Management-Expert-Guide/2.-Organise-Document

Data documentation initiative

https://ddialliance.org/

Datacite Metadata Generator

https://dhvlab.gwi.uni-muenchen.de/datacite-generator/

Exercise

Let’s have a look at an example together: For the dataset found under

https://doi.org/10.5255/UKDA-SN-851835

Can you answer all questions below?

- **Who** created the dataset?
- **What** is in the dataset?
- **When** was the dataset created?
- **Where** was the dataset created / the data collected?
- **How** were the data created and processed?
- **Why** was the dataset created?

This work is based on: Regula Zwicky. 2024. Teaching Tool «Data management for your Studies»: Data documentation. Universität Zürich; two more slides were added regarding the «how» of data documentation.