# Exploratory data analysis in environmental health

Stéphane Joost & Mayssam Nehme

# Publishing scientific articles & data

# Publishing scientific articles & data

1. Instructions for semester project description (cf ad hoc PDF on Moodle)
2. Description of the structure of a standard scientific paper
3. Publish an open dataset

4. Manage references
5. Online collaborative writing applications

6. Exercise: Upload your dataset on Zenodo Sandbox

# Exploratory data analysis in environmental health

Stéphane Joost & Mayssam Nehme

# Instructions for the description of the group semester project

# Description of the project

- The different groups will have to submit a description of the research project until **November 8**, 2024.
- This description has to summarize the project idea on two A4 pages max
- It has to contain the formulation of 1 or 2 working hypotheses your research will address
- e.g. “Geneva areas showing high vegetation indices favor physical activity and therefore overweight should be less frequent in the corresponding zones”.

# Structure of the description of the project (1)

1. Introduce a context for the problematics you identified. Shortly explain what will be the object of your research.

   For example, that you are interested in studying the spatial distribution of individuals showing overweight and the potential higher health risk for these persons when confronted to either high land surface temperatures, or exposed to a high noise level, or consuming a lot of sugar sweetened beverages.

# Structure of the description of the project (2)

2. Find 2 (recent) scientific articles to mention in your description which results were obtained by scientists in this domain and which methods they used
3. Describe briefly the different datasets you think you will need to conduct your project

   You do not need to describe them into the details, just mention what type of information you think you will need – also in the case you think you will use more (additional) data than the data already available.

# Structure of the description of the project (3)

4. Also describe the global approach you will use to implement your research
5. Formulate 1 or 2 working hypotheses you will try to verify or reject

   In the successive weeks, you will be allowed to adjust the hypotheses, but you need this starting point to initiate the thinking process

6. Indicate at the end of your A4 pages the full references of the 2 articles you used to get initial information on the topic (see point 2)
7. Indicate to us how you think you will organize the work within the group (globally how will you distribute tasks and share work, according to which arguments/criteria?). Who does what? (precise list)

# Exploratory data analysis in environmental health

Stéphane Joost & Mayssam Nehme

## Structure of a standard scientific paper

# Structure of a standard scientific paper

- Title
- Author, co-authors, affiliation
- Introduction
- Data
- Methods
- Results
- Discussion
- Conclusion
- References
- Contributions

# Title

- As explicit as possible
- Marketing : sell the study you carried out
- Cannot be too long
- Should not sound too scholarly or pedantic like
- Give a correct idea of the content
- Make a reader want to read it
- A good title makes the difference
- There is no point (punctuation) after a title

J Med Pract Manage. Nov-Dec 2014;30(3):166-7.

**Titles are terrific: creating titles that will attract attention**

Neil Baum

PMID: 25807616

**Abstract**

There are millions of blogs posted every day. However, the majority of them don't get read. Often this is because the title is not engaging and does not motivate the reader to read the content of the blog. Your blog content has to be informative and engaging, but first the blog must have a title that makes the reader want to take the time to read the blog. This article will discuss techniques and methods to use to create engaging titles that will get your blog read and, more importantly, move the reader from the blog to your Web site, and then to go online and make an appointment with your practice.

# Authors, co-authors, affiliation

- Authors and co-authors are mentioned with their last name and the abbreviation of their first name
- With several co-authors, the order in which the names are displayed has a meaning
- The first name is the main author, the last name is the senior author
- After the name, authors have to indicate their affiliation (laboratory, institute, and university
- Here “Section of Environmental Engineering (SIE), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland”
- As regards the order of the names for the short and long papers, you will use the alphabetical order

**The Origin of Chemical Elements**

R. A. ALPHER*
*Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland*

AND

H. BETHE
*Cornell University, Ithaca, New York*

AND

G. GAMOW
*The George Washington University, Washington, D. C.*
February 18, 1948

# αβγ paper

PHYSICAL REVIEW VOLUME 73, NUMBER 7 APRIL 1, 1948

## Letters to the Editor

*PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is five weeks prior to the date of issue. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length.*

### The Origin of Chemical Elements

R. A. ALPHER*
*Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland*
AND
H. BETHE
*Cornell University, Ithaca, New York*
AND
G. GAMOW
*The George Washington University, Washington, D. C.*
February 18, 1948

AS pointed out by one of us,[1] various nuclear species must have originated not as the result of an equilibrium corresponding to a certain temperature and density, but rather as a consequence of a continuous building-up process arrested by a rapid expansion and cooling of the primordial matter. According to this picture, we must imagine the early stage of matter as a highly compressed neutron gas (overheated neutral nuclear fluid) which started decaying into protons and electrons when the gas pressure fell down as the result of universal expansion. The radiative capture of the still remaining neutrons by the newly formed protons must have led first to the formation of deuterium nuclei, and the subsequent neutron captures resulted in the building up of heavier and heavier nuclei. It must be remembered that, due to the comparatively short time allowed for this process,[1] the building up of heavier nuclei must have proceeded just above the upper fringe of the stable elements (short-lived Fermi elements), and the present frequency distribution of various atomic species was attained only somewhat later as the result of adjustment of their electric charges by $\beta$-decay.

Thus the observed slope of the abundance curve must not be related to the temperature of the original neutron gas, but rather to the time period permitted by the expansion process. Also, the individual abundances of various nuclear species must depend not so much on their intrinsic stabilities (mass defects) as on the values of their neutron capture cross sections. The equations governing such a building-up process apparently can be written in the form:

$$\frac{dn_i}{dt}=f(t)(\sigma_{i-1}n_{i-1}-\sigma_i n_i) \quad i=1,2,\cdots 238, \qquad (1)$$

where $n_i$ and $\sigma_i$ are the relative numbers and capture cross sections for the nuclei of atomic weight $i$, and where $f(t)$ is a factor characterizing the decrease of the density with time.

We may remark at first that the building-up process was apparently completed when the temperature of the neutron gas was still rather high, since otherwise the observed abundances would have been strongly affected by the resonances in the region of the slow neutrons. According to Hughes,[2] the neutron capture cross sections of various elements (for neutron energies of about 1 Mev) increase exponentially with atomic number halfway up the periodic system, remaining approximately constant for heavier elements.

Using these cross sections, one finds by integrating Eqs. (1) as shown in Fig. 1 that the relative abundances of various nuclear species decrease rapidly for the lighter elements and remain approximately constant for the elements heavier than silver. In order to fit the calculated curve with the observed abundances[3] it is necessary to assume the integral of $\rho_n dt$ during the building-up period is equal to $5\times10^4$ g sec./cm$^3$.

On the other hand, according to the relativistic theory of the expanding universe[4] the density dependence on time is given by $\rho\leqq10^6/t^2$. Since the integral of this expression diverges at $t=0$, it is necessary to assume that the building-up process began at a certain time $t_0$, satisfying the relation:

$$\int_{t_0}^{\infty}(10^6/t^2)dt\leqq5\times10^4, \qquad (2)$$

which gives us $t_0\leqq20$ sec. and $\rho_0\leqq2.5\times10^8$ g sec./cm$^3$. This result may have two meanings: (a) for the higher densities existing prior to that time the temperature of the neutron gas was so high that no aggregation was taking place, (b) the density of the universe never exceeded the value $2.5\times10^8$ g sec./cm$^3$ which can possibly be understood if we

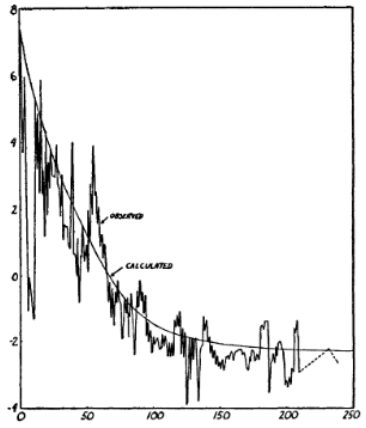
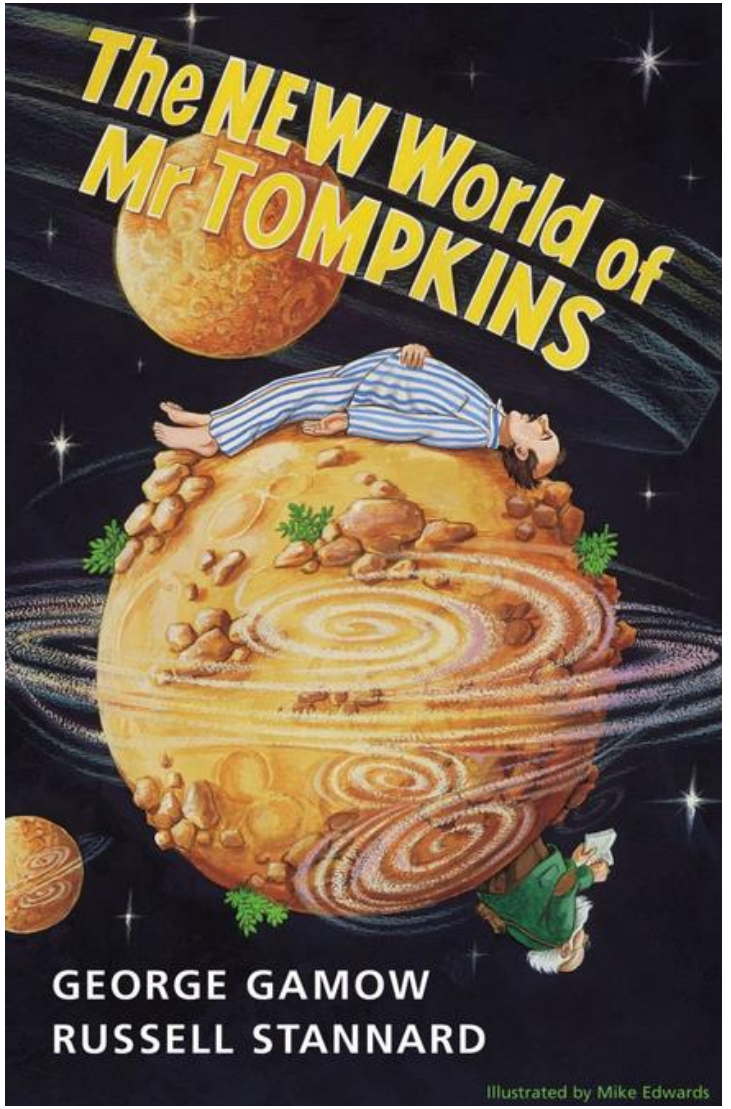



FIG. 1.
Log of relative abundance
Atomic weight






- *Monsieur Tompkins au pays des merveilles : Histoire de c, G et h* [« Mr. Tompkins in wonderland »] (trad. de l'anglais par Geneviève Guéron, ill. John Hookham), Paris, Dunod, 1953, 98 p. (OCLC 716581738)
- *Monsieur Tompkins explore l'atome* (trad. de l'anglais par Geneviève Guéron), Paris, dunod, 1959 (OCLC 878117711)
- *Mr Tompkins* in Paperback (Cambridge University Press, rééd.1993) (ISBN 0-521-44771-2)
- *Le nouveau monde de M.Tompkins* (Ed. Le Pommier, 2002) (ISBN 2-7465-0086-8)

# Introduction

- Present the context of the study
- Global introduction about the thematics
- Expose the problematics
- State of the art of papers having addressed the same issue
- Shortly indicate the approach they used and their main conclusions
- The end of the introduction should present your working hypothesis
- Short explanation about how you will proceed to verify or infirm this hypothesis (approach or methodology)

# Data

- Describe all the data you will use in your analysis
- Mention the references that describe these data
- Everything matters: Spatial resolution of geodata, the date when the data were collected / The type of data, the original projection / The source of information
- E.g. Landsat 8 image recorded on June 26, 2018 / provide a reference paper to Landstat 8 / Explain which Landsat bands you used / Explain how you characterized the cells from the raster data
- The vector data – the hectometric grid - produced by the Swiss Federal Statistical Office, STATPOP 2016
- The spatial resolution of the cells is 100m x 100m.

# Methods

- Explain how you analyzed the data (a good idea to start the method section is a paragraph (~5-6 lines) describing the methodology used (the successive analytical steps)
- List the approaches, methods and software
- Describe how they work
- Software: mention the functions used
- Refer to existing references
- For instance «We used Geoda (Anselin et al. 2006) to implement an exploratory approach…»
- Specific index or statistics: explain how it was built, mention references

# Results

- Only **describe** the results you obtained
- Section strictly descriptive – no interpretation, no discussion
- Do not start with a map, a table of results or a graph, but with descriptive text
- Go directly to the facts, to the results
- E.g. « Among the 56 models tested, the best one, according to the ranking was obtained with the following parameters…”
- Impossible to describe all sets of results: necessary to select the main ones
- Identify the most important results and establish a hierarchy among them, choose the ~3 main ones

# Discussion

- Start from the description you did in the previous section
- If you already selected the main results, you may need to further restrict to the set of most interesting results to interpret and discuss
- Propose an interpretation and possible explanations for these results
- **Use what others said in previous papers** related to your disoveries
- **Link** your results with the output of others, compare your results with results obtained elsewhere
- Highlight what is new and original and was not observed before

# Conclusion

- Summarize your paper and its main outcome
  - Propose a very short reminder about the working hypothesis
  - A very short reminder about the method
  - A very short reminder of the main results obtained and the corresponding interpretation you propose

# References

- Mention the list of all articles, books, reports, web pages, etc. used
- References must be abbreviated in the body of the text
- Example: «Richard Feynman gives several interesting examples that reveal the misdeeds of pseudo-sciences (Feynman 2010)». In the references of your paper : *«Feynman, R.P. (2000) Vous voulez rire, Monsieur Feynman! Odile Jacob. Paris.»*
- Use alphabetical order of the name of the main authors when you list your references
- You will get a DOI when you will upload your dataset to Zenodo Sandbox. Mention it in your bibliography:

*Gaillard, S. (2018) Vegetation data at the hectare level, Zenodo, DOI: 10.1111/1755-0998.12991*

# Contributions

- Describe the respective contribution of each co-author (all members of the group)
- Explain who did what in writing the manuscript, processing the data, creating maps, write the text, create the maps, create the plots, etc.
- Use the initials of each co-author
- Here is an example:

  *«SJ and RG looked for references and drafted the state of the art, SJ, RG and CS wrote the introduction and conclusion, CS performed the calculations and wrote the results sections, SJ and CS elaborated the thematic maps.»*

- This is an important and compulsory section

# Publish data on an open repository

- Why?
  - Personal commitment to Open Science
  - Funders' requirements
  - Institutional policy
  - Journal policy
  - Good practice

https://sandbox.zenodo.org/

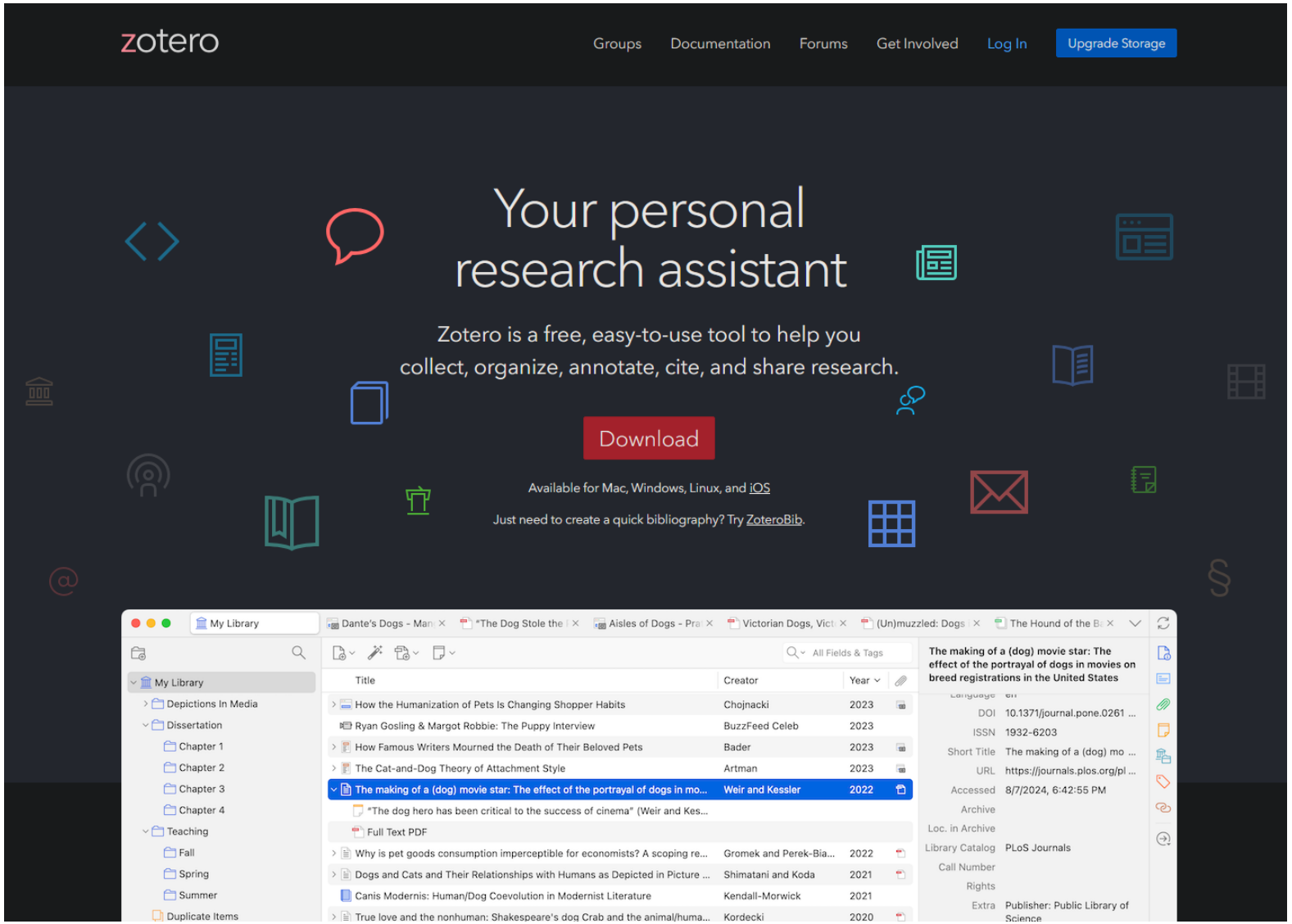
# Publish data

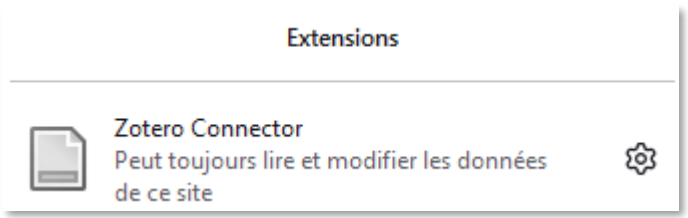
- Analysis results are generally published to advance knowledge in all areas of scientific research.
- But it is also useful to publish datasets for the same reasons.
- And to make these data accessible to the entire scientific community, it's very important to publish them in the “open” domain.
- When an article is published in a scientific journal, it receives a unique identifier called “DOI” for “Digital Object Identifier”.
- This makes it easier to identify publications and to refer to them in other publications.
- Following the same model, “open” datasets published in data repositories also receive a DOI.

# Exercise on data publication

- To complete the exercise of analyzing and publishing research results that you are carrying out this semester, we ask you to upload the open dataset that you developed as part of Exercise 3 (Geneva's inhabited hectares characterized by environmental data on noise, vegetation and air pollution).
- You have to place your vector dataset (.gpkg or .shp) in a .zip file, then upload this file to the Zenodo data repository.
- We will use Zenodo's Sandbox test space for this exercise (https://sandbox.zenodo.org/).
- The aim is to obtain a DOI for your open dataset
- When you will write your article (semester project), you'll be able to refer to your dataset and cite it as a reference
- Tasks are described in exercise 6 on Moodle

# Manage the references



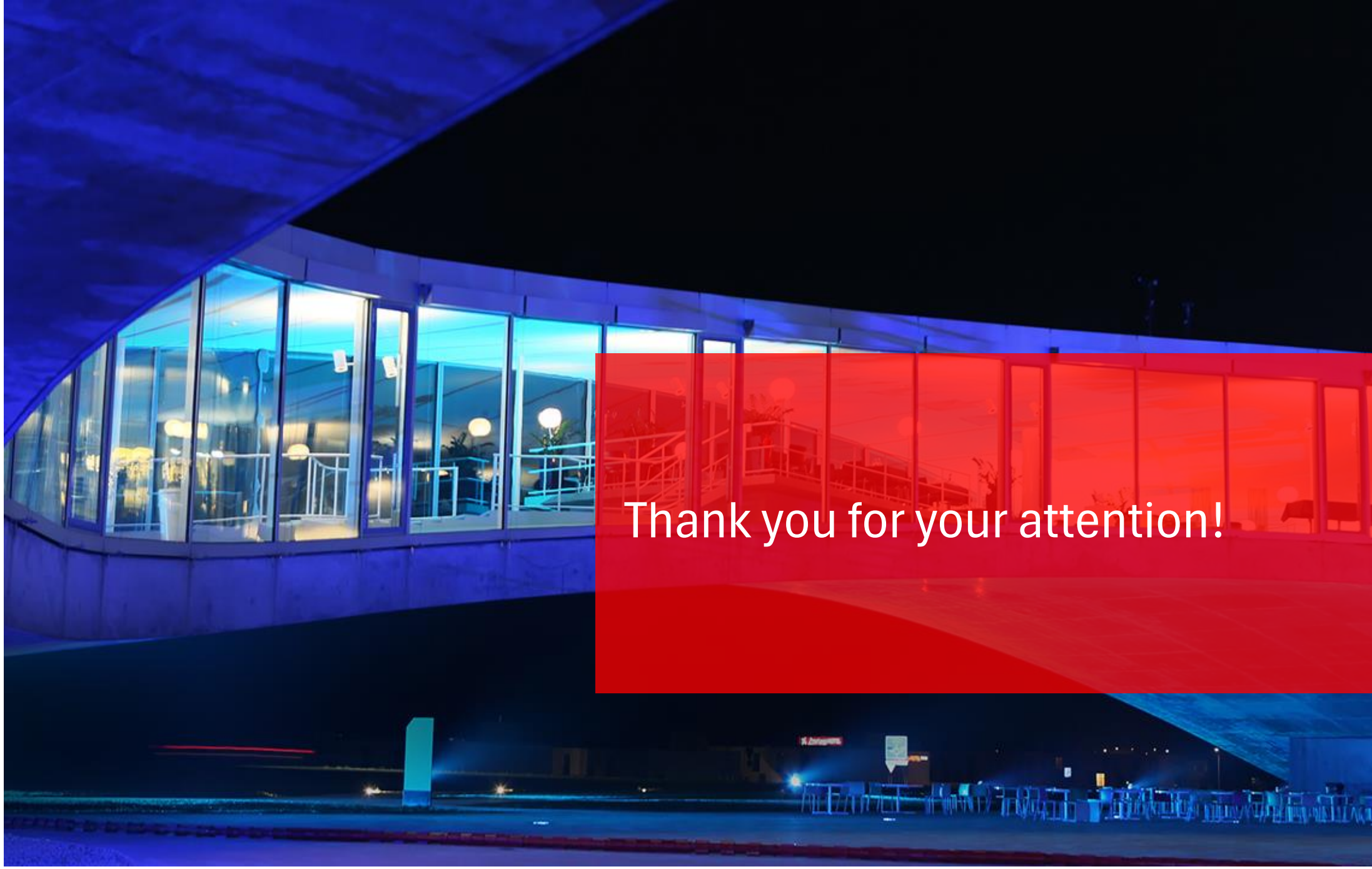



# Collaborative writing

- Overleaf
- Authorea
- Share LaTeX
- Onedrive, sharepoint (ms-word)
- Google Docs with Zotero


# References


- Cargill, M., & O'Connor, P. (2013). Writing Scientific Research Articles: Strategy and Steps. John Wiley & Sons.
- Gastel, B., & Day, R. A. (2016). How to Write and Publish a Scientific Paper, 8th Edition. ABC-CLIO.
- O'connor, M., & Woodford, F. P. (1975). Writing scientific papers in English. An ELSE-Ciba Foundation guide for authors. Writing Scientific Papers in English. An ELSE-Ciba Foundation Guide for Authors.
- SWALES, J., & NAJJAR, H. (1987). The Writing of Research Article Introductions. Written Communication, 4(2), 175–191. doi:10.1177/0741088387004002004
- Way, C. W. V. (2007). Writing a Scientific Paper. Nutrition in Clinical Practice, 22(6), 636–640. doi:10.1177/0115426507022006636


Thank you for your attention!