

Scientific Writing in a Nutshell

Summer 2024

Dr. Schapranow

**Design IT.
Create Knowledge.**

www.hpi.de



Key Facts



Submission no later than Sun Aug 18, 2024 containing:

- Contributions, prototypes, models, source code as a zip archive,
- Brief instruction how to make it working, and
- Your final report as PDF version + source, e.g. Latex/Word, figures, etc.

Final report

- Format: IEEE two-column conference, A4 templates available at <https://www.ieee.org/conferences/publishing/templates.html>
- Length max. 4 (TaC4DH) or 6 (AI4DH) full pages (excl. appendices, references, figures) respectively

Accessing the templates

Microsoft Word

- [US letter](#) (DOC, 30 KB) **Updated Jan 2019**
- [A4](#) (DOC, 30 KB) **Updated Jan 2019**

[LaTeX Template Instructions](#) (PDF, 63 KB) [Be sure to use the template's **conference** mode.]

- [Template](#) (ZIP, 700 KB) **Updated October 2019**
- [LaTeX Bibliography Files](#) (ZIP, 309 KB)

Overleaf

- When working in Overleaf, the template is available at <https://www.overleaf.com/gallery/tagged/ieee-official>

Can Emojis Make Science?

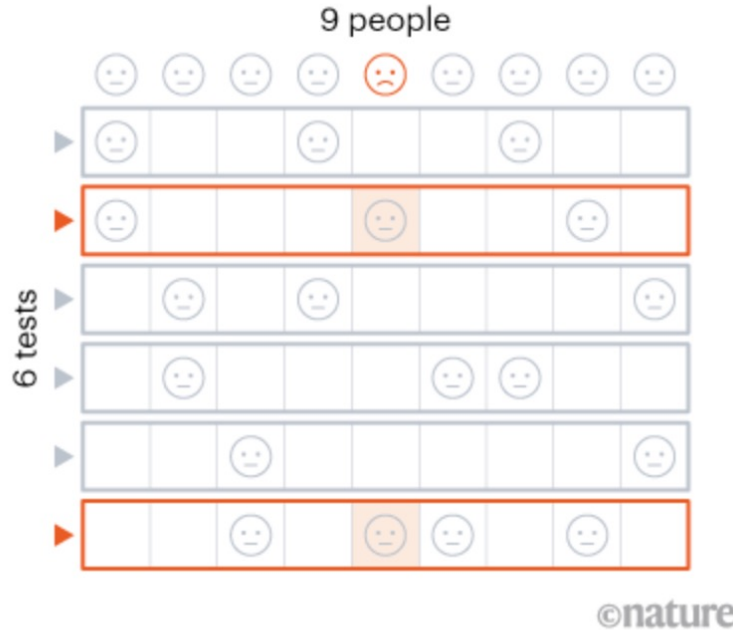


Emojis May Help Science!



Method 4

This method uses only one round of testing. Samples are distributed into a matrix of overlapping groups.



nature



NEWS · 10 JULY 2020

The mathematical strategy that could transform coronavirus testing

Four charts show how pooling samples from many people can save time or resources.

Smriti Mallapaty

Sources: Dorfman, R. *Ann. Math. Statist.* 14, 436-440 (1943); Mutesa, L. *et al.* Preprint at <https://arxiv.org/abs/2004.14934> (2020).; Manoj Gopalkrishnan

Structure

The Fairytale Approach

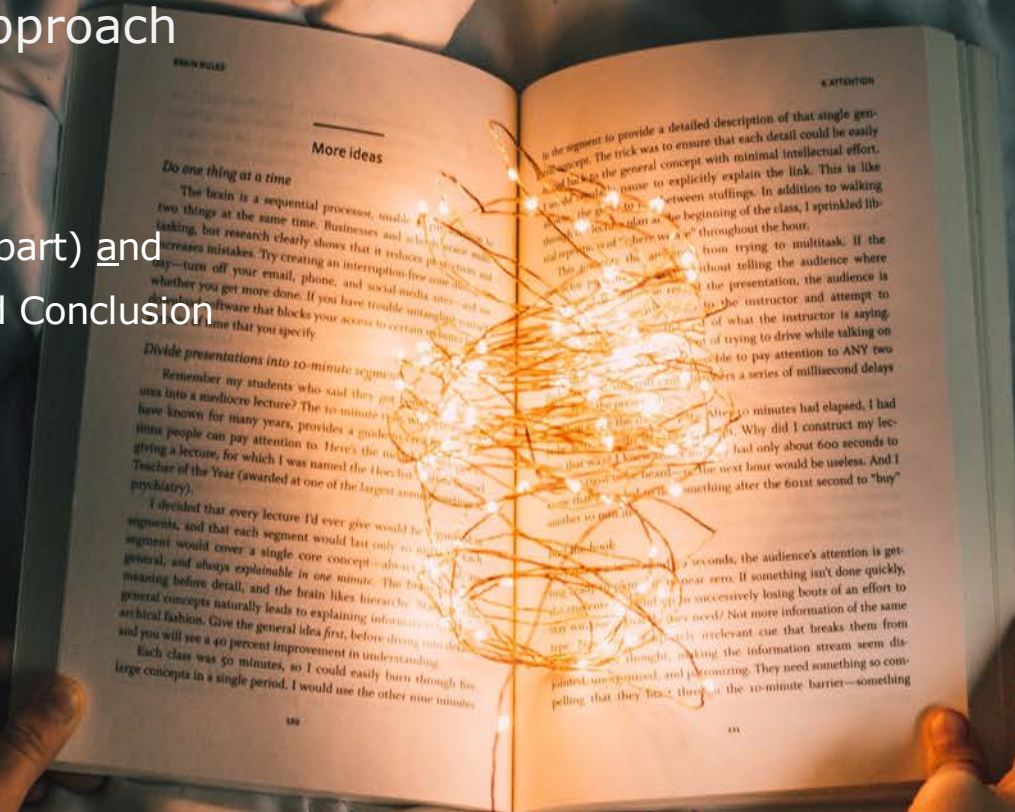
1. Introduction
2. Main part
3. Conclusion



Structure

The IMRAD Approach

1. Introduction
2. Methods
3. Results (Main part) and
4. Discussion and Conclusion



Structure

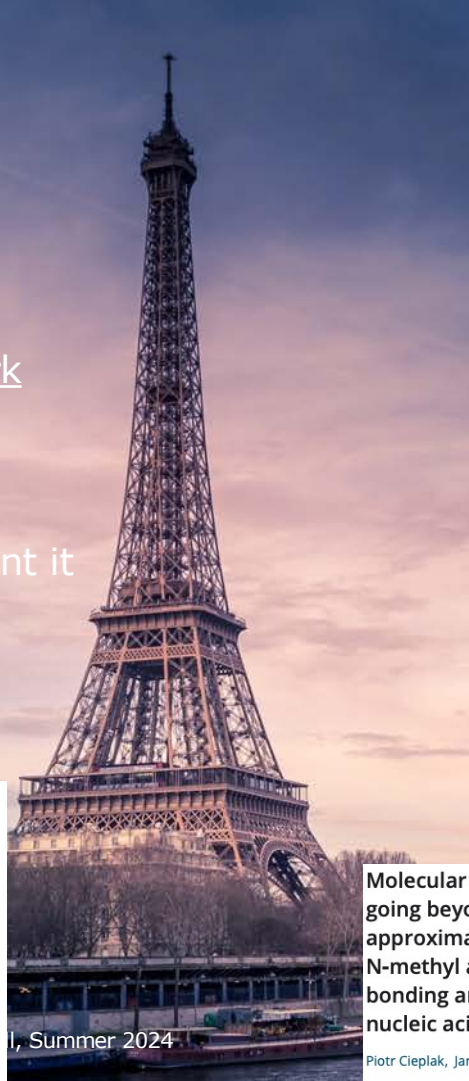
The Engineering Papers Approach

1. Introduction
2. Related Work
3. Methods
4. Contribution(s)
5. Evaluation
6. Discussion
7. Conclusion and Outlook



Title

- Aim: Make reader curious about your work
- Catchy, but precise
- Not too long as you want it to be indexed



The nucleotide sequence of a 3.2 kb segment of mitochondrial maxicircle DNA from *Crithidia fasciculata* containing the gene for cytochrome oxidase subunit III, the N-terminal part of the apocytochrome *b* gene and a possible frameshift gene; further evidence for the use of unusual initiator triplets in trypanosome mitochondria

Paul Sloof ✉, Janny van den Burg, Arthur Voogd, Rob Benne

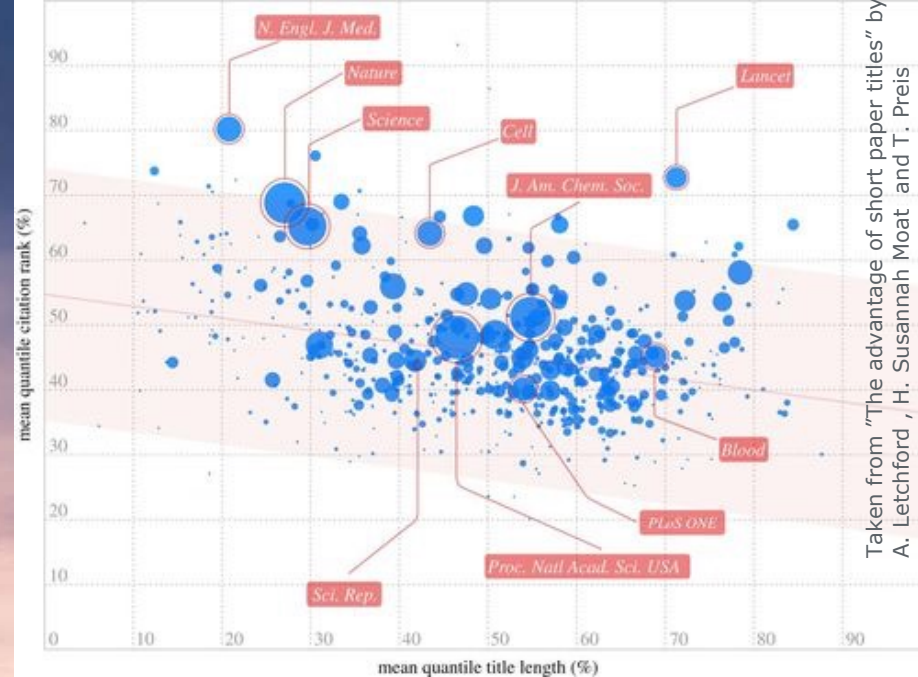
Nucleic Acids Research, Volume 15, Issue 1, 12 January 1987, Pages 51–65,
<https://doi.org/10.1093/nar/15.1.51>

Published: 12 January 1987 [Article history](#) ▼

, Summer 2024

Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases

Piotr Cieplak, James Caldwell, Peter Kollman ✉



Taken from "The advantage of short paper titles" by A. Letchford, H. Susannah Moat and T. Preis



Abstract



- Aim: Make the reader understand the scope of your contribution
- Answer the following four questions:
 1. What is the state of the art?
 2. What are the limitations of the current state?
 3. What is your contribution?
 4. How does your contribution improve the state of the art?
- Review your abstract at the end of your work to check whether it is still aligned

1. Introduction

- Aim: Brief intro into the topic not an academic lecture
- Guide the reader to the topic in brief
- Name some well-known work
- You might want to place an important figure or graph on the first page to make your work look unique
- The last paragraph shares details about the structure of your document, e.g. "The remainder of the work is structured as follows: Sect. 1..."



Towards An Integrated Health Research Process: A Cloud-based Approach

Matthieu-P. Schapranow¹, Matthias Ullacker¹, Murat Sariyar², Sebastian Semler²,
Johannes Fichte³, Dietmar Schielke³, Kismet Ekinci⁴ and Thomas Zahn⁴

¹ Hasso Plattner Institute, Enterprise Platform and Integration Concepts,
August-Bebel-Str. 88, 14482 Potsdam, Germany

² Technology, Methods, and Infrastructure for Networked Medical Research,
Charlottenstr. 42, 10117 Berlin, Germany

³ data experts gmbh, Wödlögler Str. 12, 17033 Neubrandenburg, Germany

⁴ Gesundheitswissenschaftliches Institut Nordost der AOK Nordost, Wilhelmstr. 1, 10963 Berlin, Germany

E-Mail: ¹{matthieu.schapranow|matthias.ullacker}@hpi.de, ²info@tmf-ev.de,

³{johannes.fichte|dietmar.schielke}@data-experts.de, ⁴{kismet.ekinci|thomas.zahn}@nordost.aok.de

Abstract—Today, health research and health care generate a steadily increasing amount of data. Making these available for secondary use cases is essential for efficiency gains in health research, e.g. by reducing time- and cost-intensive acquisition of data.

In this contribution, we introduce our SAHRA software platform enabling reproducible research, e.g. by combining multiple data sources, performing data de-identification, and content filtering. We define an innovative research process combining retrospective and prospective research for the first time.

Thus, authorized users, e.g. clinical researchers, are able to gain access through our system relevant research data and to perform interactive analyses. As a result, existing sensitive health data is securely transformed into de-identified research data, which can be used to improve future health research.

Index Terms—Health Research Process; Data De-Identification; Sensitive Data; Research Data; Software Architecture

I. INTRODUCTION

Currently, healthcare data is acquired in numerous ways, e.g. by doctors during visits, laboratories, and health insurance companies. On the one hand, these data are considered very sensitive and require specific measures and safeguards defined by law to protect the human individual they describe. On the other hand, access to data sets from these different contexts allows gaining new insights, e.g., regarding the course of diseases or additional requirements for an elderly society.

In this contribution, we share first results of our Smart Analysis Health Research Access (SAHRA) cooperation project. Its aim is to provide scientific analysis methods incorporating latest In-Memory Database Technology (IMDB) for analysis of longitudinal health data [1]. These methods support research and the development of innovative solutions and products by providing real-time analysis of longitudinal health data for the first time.

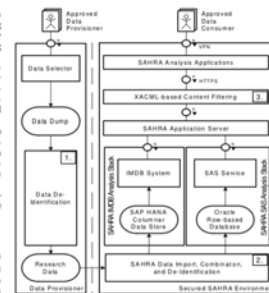


Fig. 1. Software architecture depicting the interaction between data provider and data consumer.

Public and governmental institutes as well as small and mid-sized healthcare enterprises are the target audience of the project. Latest data protection and privacy measures are taken to protect any data on the SAHRA platform in compliance with latest German and international data protection laws.

Our contribution focuses on our SAHRA software platform and especially on how to make sensitive health data available for research. We share real-world experiences of

2. Related Work



- Aim: Show you know other's work and can set your contribution in the context
- Perform literature review, share results, e.g. IEEE, PubMed, ACM, SpringerLink, Google Scholar, Elsevier Scopus, etc.
- Name selected approaches and stress how your work differs, extends, matches, etc.
- You might want to create a structured tabular overview, e.g. selected features, approaches

Table 1: Tools surveyed. Abbreviations: GP-General Purpose, BSP-Biomedical Software Platform. Legend: ●-Fully supported; ◐-Partially supported; ○-Not supported.

Category	Tool	Functional Requirements																				Non-Functional			
		Preparation			DS Selection			Prod. Handling			M. Generation			Eval. & Validation			Interpretation			Presentation			NF		
		F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	NF1	NF2	NF3
GP	Statistical Packages [1, 30, 40]	○	○	○	●	○	○	●	●	●	●	○	○	●	●	○	○	○	○	○	○	○	○	○	○
	ML Toolkits [14, 21, 28, 39, 41]	○	○	○	●	○	○	●	●	●	●	○	○	●	●	○	○	○	○	○	○	○	○	○	○
	Scientific Computing [17, 25, 51]	○	○	○	●	○	○	●	●	●	●	○	○	●	●	○	○	○	○	○	○	○	○	○	○
	ML Suites [16, 22, 23, 31]	○	○	○	●	○	○	●	●	●	●	○	○	●	●	○	○	○	○	○	○	○	○	○	○
BSP	transSMART [49]	◐	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	ATLAS [19]	○	○	○	●	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	MLBCD [36]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	ExplicU [11]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	PIP R package [46]	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
	MORPHER	●	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

3. Methods



- Aim: Outline how you worked methodically to get your findings
- What particular methods have you used, e.g. design thinking, software engineering, machine learning methods, quantitative measurements, etc.
- Why did you select them, e.g. referenced by others, make your work comparable to other's work, etc.

4. Contribution(s)

- Aim: Share most relevant details of your work
- Describe what you achieved, i.e. results and deliverables
- Highlight your findings, what is unique, what are you proud of, what do you want to share with others
- Share detailed insights, prevent just general sentences
- Clearly distinguish b/w your own work and what you have built upon
- Prevent story-telling, e.g. ~~“First, I did this, but it fails. Then, I tried this. At the end...”~~



4a. Benchmarks and Results



- Aim: Compare your work and support reproducibility
- Describe your specific setup for obtaining results, e.g. computing infrastructure, measurement devices, precision of clocks or measurement devices, what was measured, how often, etc.
- What metric(s) did you measure and for what reason
- Which parameters were kept fixed / varied – hint: keep most parameters fixed and change one parameter at a time
- What results did you acquire
- Please do not place interpretations or assessments here → better move them to your evaluation

TABLE 1. Simulation results displaying AUCROC for the different analysis cohorts and patient outcomes. Abbreviations: IHID = Intermittent Hemodialysis, CRRT = Continuous Renal Replacement Therapy, MLP = Multilayer Perceptron, RF = Random Forest, BRL = Bayesian Rule Lists, LR = Logistic Regression and BRR = Bayesian Ridge Regression, LOS ICU = Length of Stay in the ICU.

Outcome	Complete cohort				Acute patients				HD patients				CRRT patients									
	MLP	RF	BRL	LR	BRR	MLP	RF	BRL	LR	BRR	MLP	RF	BRL	LR	BRR	MLP	RF	BRL	LR	BRR		
90-days mortality	0.81	0.84	0.76	0.71	0.79	0.81	0.85	0.79	0.81	0.83	0.82	0.79	0.69	0.79	0.77	0.78	0.72	0.66	0.72	0.76	0.66	0.72
Renal Recovery	0.78	0.81	0.88	0.77	0.88	0.84	0.81	0.68	0.72	0.78	0.82	0.81	0.80	0.78	0.80	0.86	0.71	0.78	0.76	0.86	0.76	0.86
Yieldable Days	0.81	0.79	0.71	0.74	0.80	0.64	0.64	0.68	0.65	0.61	0.81	0.74	0.78	0.71	0.79	0.75	0.64	0.79	0.64	0.79	0.64	0.79
LOS ICU	0.81	0.86	0.82	0.73	0.82	0.78	0.64	0.69	0.61	0.73	0.80	0.82	0.78	0.78	0.80	0.79	0.64	0.79	0.63	0.79	0.63	0.79

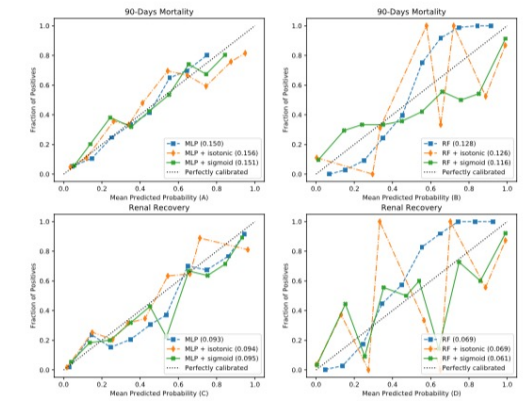
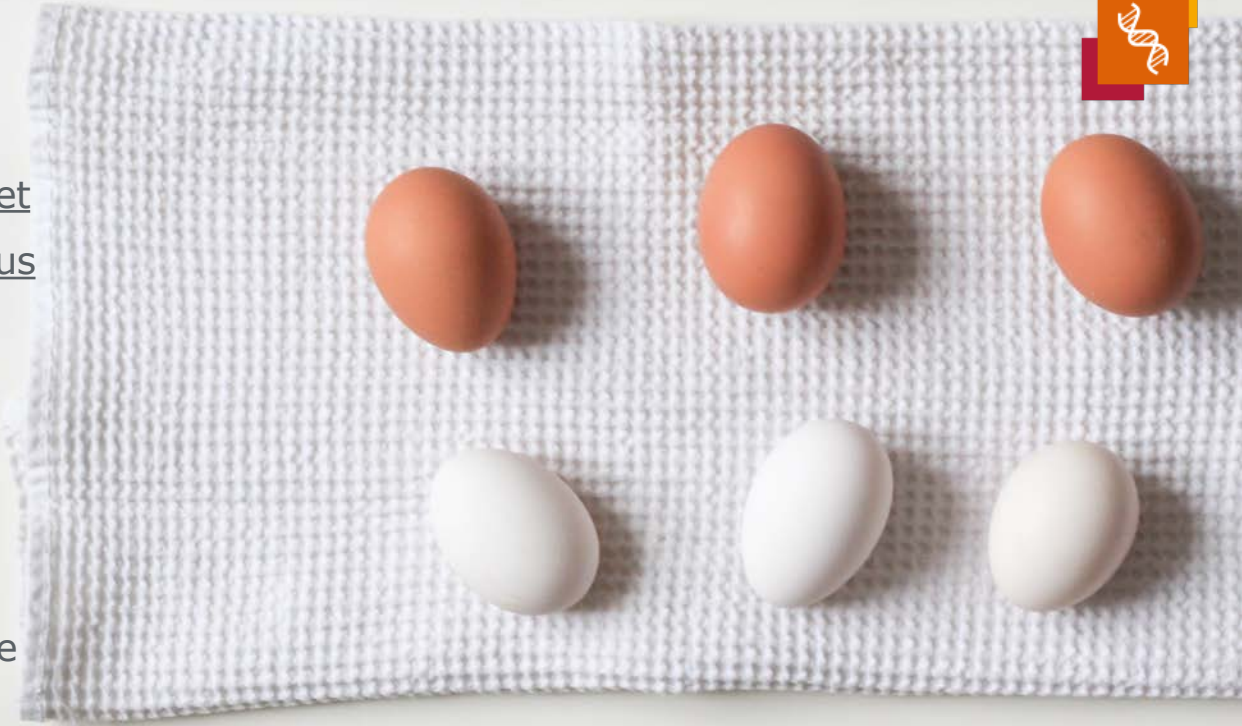


Figure 4. Model calibration depicted for Multilayer Perceptron (MLP) and Random Forest (RF) for the outcomes of 90-days mortality and renal recovery.

5. Evaluation

- Aim: Help the reader to interpret your findings/observations; focus on your observations!
- “I tried to produce identical eggs, but my results differ in size and color.”
- “We measured a mean response time of 110 ms with a standard deviation of 15%, which is 10% faster than comparable approaches.”
- “We would have expected to double the throughput, but stuck at a factor of 1.5x.”



6. Discussion



- Aim: Discuss your findings, what did the evaluation show, how to proceed with it
- Did your results meet your expectations, yes/no why?
- You might want to explain why you meet/lack behind certain assumption
- What could be possible explanations for your observations/measurements
- What next steps/changes need to be conducted to understand the source for your results
- Do you need to change assumptions, data, parameters, etc.

7. Conclusion & Outlook

- Conclusion (Aim: Sum up your findings / main contribution)
 - Help to recall your findings / contributions in a nutshell
 - Stress why and how your contributions extend the current state of the art; why is your work important
- Outlook (Aim: Guide the reader through your planned next steps)
 - Do you plan certain next steps, experiments
 - How do you plan to proceed your research
 - How will you address issues identified during your evaluation



Bibliography



- Recommendation: Use a bibliography management tool to make your life easier
- Always check (even if you use existing bib snippets for publications):
 - Check for consistency, e.g. upper- vs. lowercase of titles and journal names (use curly brackets in bib files to lock format “{MySQL}”)
 - More than three authors? → First author’s name + “et al.” (can be achieved by “Schapranow and others” in bib files)
 - Online references? → Please add URL and “last checked” date – Hint: Check all online references prior to submission at the same date

Style of Writing



- Be precise and formal, e.g. "I or we did X and Y to achieve Z."
- Underline statements by adding appropriate references at the end of the sentence
- Write about facts and not about opinions or your individual challenges – the discussion section is the right place to share your personal opinion
- Consider splitting complex sentences in multiple parts
- Please prevent "already mentioned, as state before, again..."

- Use commas when appropriate, e.g.
 - With conjunctions,
 - With most "w-"pronouns
 - Must not be placed before that

The Perfect Length of the Paper

*THE UNSUCCESSFUL SELF-TREATMENT OF
A CASE OF "WRITER'S BLOCK"¹*

DENNIS UPPER

VETERANS ADMINISTRATION HOSPITAL, BROCKTON, MASSACHUSETTS



"I have studied this manuscript very carefully with lemon juice and X-rays and have not detected a single flaw in either design or writing style. I suggest it be published without revision. Clearly it is the most concise manuscript I have ever seen-yet it contains sufficient detail to allow other investigators to replicate Dr. Upper's failure. In comparison with the other manuscripts I get from you containing all that complicated detail, this one was a pleasure to examine. Surely we can find a place for this paper in the Journal-perhaps on the edge of a blank page."

REFERENCES

¹Portions of this paper were not presented at the 81st Annual American Psychological Association Convention, Montreal, Canada, August 30, 1973. Reprints may be obtained from Dennis Upper, Behavior Therapy Unit, Veterans Administration Hospital, Brockton, Massachusetts 02401.

*Received 25 October 1973.
(Published without revision.)*

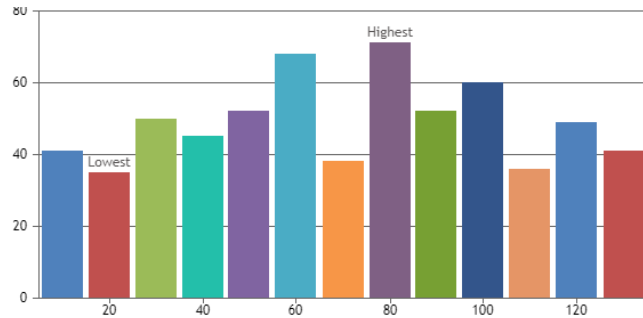
Formal Aspects



- Check spelling and grammar repeatedly
- If a new line ends with a single word, consider rephrasing the sentence
- Place multiple references together at the end of the line to support readability, e.g. “[1,7-10,15]”.
- Check if anything – figures, tables or text – reaches into the paper margin
- Place figures either at top / bottom of a page, prior to referring to them
- Refer to all figures, tables, equations, listings, etc. within the text
- Add an appropriate caption to your figures, tables, equations, listings, etc. to summarize them for the reader

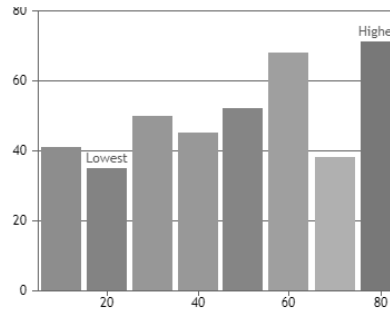
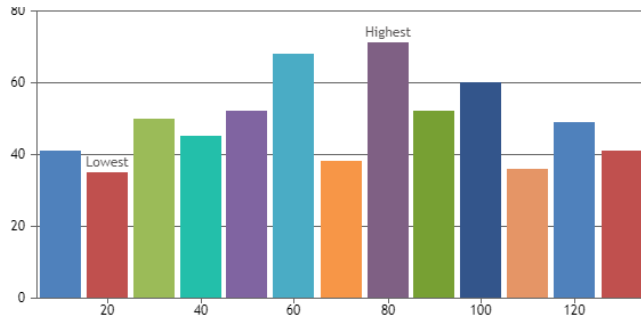


Figures, Charts, Listings, ...

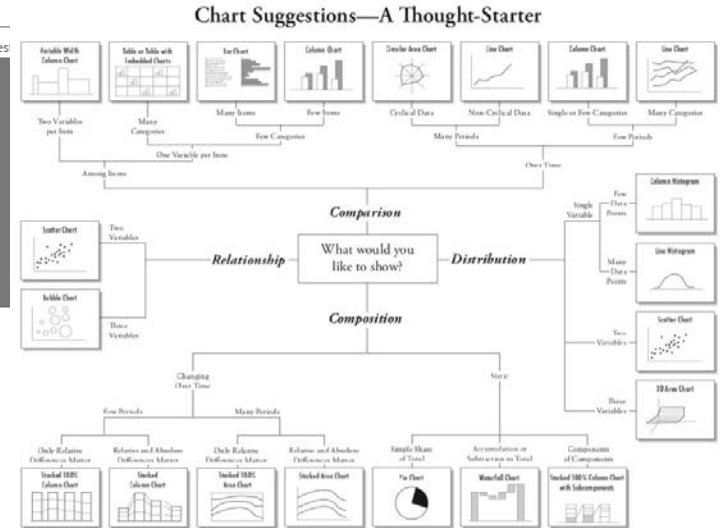


- “A picture is worth a thousand words”
- Creates an long-lasting association in the reader’s mind
- A strong form of communicating findings

Figures, Charts, Listings, ...



- “A picture is worth a thousand words”
- Creates an long-lasting association in the reader’s mind
- A strong form of communicating findings
- Use visual hatchings to make charts readable even in grayscale printing
- Text in figures should be the same size as the remainder of your document
- Resolution should be adequate for printing, i.e. at least 300dpi



LaTeX: Tips and Tricks



- Dos
 - Introduce `\label{fig:abc}` for anything you want to reference, e.g. section, figures, charts, listings, etc.
 - Use `\autoref` or `\cref` instead of `<labelname>~\ref`
 - For references, use a `.bib` file and an appropriate bib editor, e.g. BibDesk, Mendeley
 - Check line breaks in latex source because they have a special meaning, e.g. `indent`

- Don'ts
 - Prevent footnotes, e.g. for URLs; instead add a proper reference
 - Prevent manual formatting, e.g. bold or italics, because it is managed by `.sty/.cls`
 - Manual line breaks (`\\`) or new pages (`\newpage`)

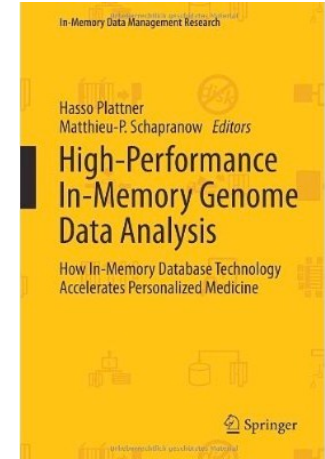
Happy writing!



Dr. Matthieu-P. Schapranow
Group Leader & Scientific Manager Digital Health Innovations

Hasso Plattner Institute
Digital Health Cluster
Rudolf-Breitscheid-Str. 187
14482 Potsdam, Germany

we.analyzegenomes.com
@AnalyzeGenomes



Supported by:



Federal Ministry
of Education
and Research



Federal Ministry
for Economic Affairs
and Energy



European
Commission

Horizon 2020
European Union funding
for Research & Innovation

on the basis of a decision
by the German Bundestag