Inf2 – Foundations of Data Science
S2 Week 8: Writing workshop

THE UNIVERSITY *of* EDINBURGH
informatics

FOUNDATIONS
OF
DATA
SCIENCE

Announcements

- Week 8's Lecture
  "Software Engineering for Data Science"
  Emma Pead, Senior Data Scientist, head of the Data Intelligence Hub for the Biomedical AI CDT
- Project presentation sign-ups will be released next week
  - Presentations will be in weeks 9 and 10
  - During regular workshop slots (both in 5.04 and 5.01)
  - Each presentation is 5 min (3 of you talking and 2 for Q&A
- Written update – feedback next week

Aims

- Make you feel better & more confident about writing
- Give you some tips on how to write
- Give you some advice on things to avoid
- Suggest exercises you can do to improve your writing

Overview

- About writing
- What makes writing effective?
- How do we write?
- Storytelling
- Levels of writing
- Referencing
- Resources

About writing

Why do we write?

- Science and industry: **to communicate ideas** to **other people**
  - Not to impress with the beauty of our writing...
  - … Or intimidate the reader with our superior knowledge
- **To help us think**
- Other reasons to write:
  - To entertain
  - To persuade
  - To instruct

For whom do we write?

- You are writing for **other people**, not yourself
- It's **your job** to help them understand
- Try to keep them in mind when writing
  - what they know – e.g. general methods
  - what they don't know – e.g. how you have applied those methods
  - They can get distracted and lose concentration.
  - Are you making it easy for them to follow your writing?
  - How are you going to gain and hold their interest?

Writing is hard

- "Writing a book is a horrible, exhausting struggle, like a long bout of some painful illness" - George Orwell in *Why I write*

- But writing can be learned

- And the best way of learning is by doing it – a lot

What makes writing effective?

Exercise

Identify features of effective and ineffective writing in submissions from a previous year in the "**Assessment -> Examples of Previous Coursework -> Project** folder in **Learn**.

- fds-project-option-2-individual
- fds-project-option-3-individual

Features of effective scientific writing

- Accurate
- Evidenced
- Informative

- Logically organised
- Clear
- Concise
- Motivating

- At right level for audience
- Coherent

Features of ineffective scientific writing

- Missing figure numbers
- No background or motivation
- Missing explanation

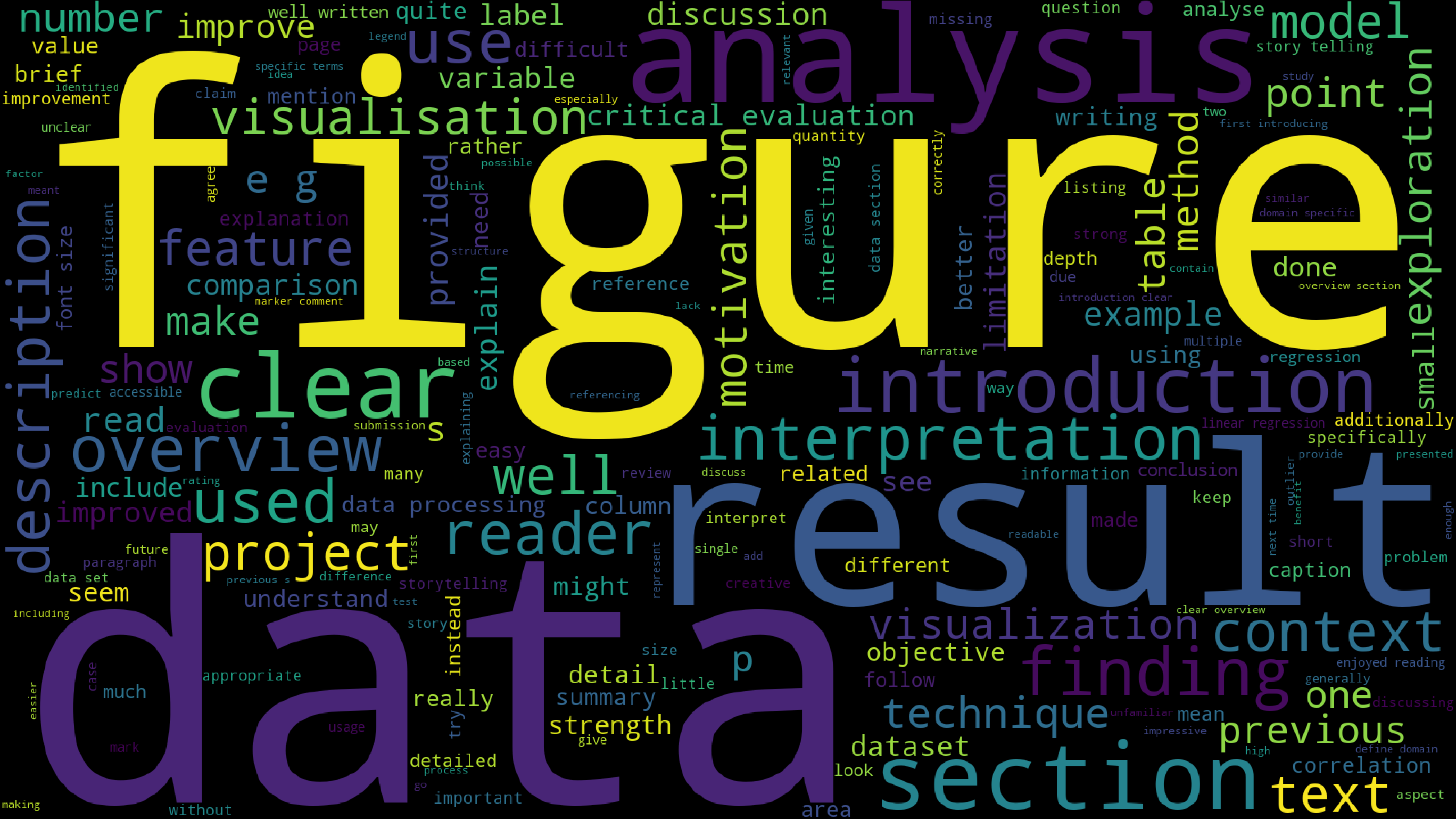
figure
data
result
analysis
section
introduction
interpretation
clear
visualisation
context
finding
reader
overview
use
used
project
description
feature
visualization
technique
model
well
point
exploration
motivation
text
previous
number
improve
discussion
critical evaluation
show
read
make
comparison
table
method
one
dataset
example
small
e g
variable
label
quite
well written
value
brief
improvement
difficult
limitation
explain
provided
include
improved
s
p
see
related
done
strength
detail
summary
objective
correlation
caption
column
data processing
understand
might
seem
mention
explanation
need
interesting
writing
rather
better
listing
depth
reference
information
conclusion
additionally
specifically
using
regression
different
really
detailed
instead
much
appropriate
follow
mean
look
area
important
without
storytelling
story telling
analyse
question
missing
page
specific terms
legend
idea
identified
claim
unclear
factor
meant
agree
significant
font size
marker comment
explanation
structure
possible
think
quantity
correctly
especially
two
first introducing
study
relevant
similar
domain specific
strong
due
contain
overview section
introduction clear
lack
narrative
multiple
time
based
predict
accessible
evaluation
submission
referencing
way
linear regression
easy
many
rating
review
discuss
interpret
made
keep
provide
presented
outlier
benefit
next time
enough
may
future
paragraph
first
data set
previous s
difference
readable
represent
single
add
creative
short
problem
test
including
story
size
little
try
case
easier
usage
mark
give
process
go
making
clear overview
enjoyed reading
generally
discussing
unfamiliar
impressive
high
define domain
aspect
data section
given

How do we write?

The writing process

- Some lucky people can write a document from the beginning to the end in one go
- Most of us can't
- We need to split writing into multiple activities:
  - Creating a structure
  - The first draft
  - Re-writing, editing and revising
- Good to leave plenty of time between a draft and the edit

Creating a structure

- The structure in the project template is similar to much scientific writing:
  - Introduction: What's the area of the data? Why is it interesting to investigate? What have others done? What questions am I asking?
  - Data [Methods]: What is the data? How am I processing it?
  - Exploration and analysis [Results]: What can we learn from the data? What are the answers to the questions?
  - Discussion and conclusions: What did I find? What were the strengths and limitations of my own work? How does what I find compare with other knowledge? How could this work be improved?

The first draft

- Write quickly, trying not to go back and fix things
- Don't worry about details – get down the main "building blocks"
- After the writing:
  - Put in references (in BibTeX)
  - spell check
  - grammar check (Langtool.org is good free alternative to Gammarly)
- Ideally! – wait a day or two

Revising and editing

- **Read** through your draft, marking up problems as you go
  - E.g. by printing out & using a pen
  - Using a tablet
  - Comments on a PDF
  - The *change of context* is helpful
- Work through your comments, ticking them off as you go
- Repeat!

Final check

- Spell check
- Grammar check
- Is every figure referred to in the text?
  - "Figure 1 shows...."
- Are all references present?
- Read your document out loud!
  - Painful, but it helps you spot errors and gaps in the logic

Storytelling

The words and figures should tell a story

- There needs to a logical progression from one point to another
- The pictures (figures) tell much of the story, but the reader needs some interpretive text to be helped to see what to note in the figure
- Reading over by a fellow group member can help!

Levels of writing

Your document has multiple levels of structure

- Section structure (given)
- Within sections, paragraphs
- Within paragraphs sentences
- Organisation needs to be **structured** and **coherent**

Paragraphs

- Structure: one main idea/topic per paragraph
- Usually one core/topic sentence, usually the opening sentence, which tells the reader what the paragraph is about
- Paragraphs needn't all be the same length; some topics require more space
  - But try to avoid very long paragraphs
- One paragraph should follow from another

Paragraphs that are logically coherent

- "I have decided to create another dataset with the distances from munros to the top largest cities in Scotland, such as Edinburgh, Glasgow and Perth. To check this I have conducted PCA on the data …"
  - In what sense does the 2nd sentence follow from the first?
  - How does doing PCA **check** the distances?

Sentences

- Sentences should:
  - Express a small idea
  - Be clear and concise
  - Can be as long as they need to be – but beware of very long sentences

Tips for writing concise sentences

- Put the action in the verb
  - X is an indication that Y → X indicates that Y
- Use active voice:
  - my model is described in section 3 → I describe my model in section 3 or section 3 describes my model.
- Eliminate wordy phrases and weak verbs:
  - Our results provide evidence that → Our results suggest that

Tips for writing clear sentences: "This"

- "Interestingly, clusters did coincide with the established popularity ratings, even though the separation between them is not very clear. The reason behind **this** is other possible munro characteristics or possible details of the relief, which I am not focusing on in this report."
  - What does **this** mean?
- Every time "this" is used, the reader needs to work out what it connects to
- Adding a noun can helps avoid ambiguity, e.g. "this data"

Tips for writing clear sentences: comparisons

- "Hu et al [7] found that fewer commuting trips were made on the Chicago bikeshare system during the pandemic."
  - Than when?
- "Hu et al [7] found that fewer commuting trips were made on the Chicago bikeshare system during the pandemic *than in the pre-pandemic year*."

Tips for writing clear sentences: remove unneeded words and add information

- I have decided to create another dataset with the distances from munros to the top largest cities in Scotland, such as Edinburgh, Glasgow and Perth.

  →

- I created a dataset with the distances from **each** Munro to the **n** largest cities in Scotland, including Edinburgh, Glasgow and Perth.

Referencing

References

- All not generally-known statements should have a citation
  - No citation needed for standard methods such as linear regression, PCA
  - Citations needed for people who've done previous work, datasets
  - Citation not needed in every sentence, as long as context is clear.
- Parenthetical or with author name? Either is fine...
  - Hu et al. [7] reported that the proportion of bikeshare commuting trips made in Chicago was substantially lower during the pandemic.
  - The proportion of bikeshare commuting trips made in Chicago was substantially lower during the pandemic [7].

References section

- Each reference should provide enough information to find work
- Journal article:
  - Author, Year, Title, Journal Title, Journal Volume, Page number/range, Digital Object Identifier (URL)
  - Sterratt, D (1995), "Waves in a tank", *Journal of Interesting Ideas* **94**: 67-89
- News article:
  - Author name, Date, Title, Publication, URL, Date retrieved
- BibTeX has many types to cover different types of work
- Less formal references (e.g. blogs, wikipedia pages, etc.) are best referenced in a footnote rather than in the bibliography

Resources

Resources and exercises (links on course web pages)

- Google Technical Writing One - Exercises on lots of aspects of writing, particularly at the sentence and paragraph level
- Sharon Goldwater's short guide to writing clear and concise sentence
- Tips on writing clear, concise sentences from the University of Wisconsin Madison Writing Center - a much larger set of tip
- The University of Manchester Phrasebank - A great resource for useful ways of phrasing things appropriately in scientific writing

What about AI?

AI Policy

- For this course, we are allowing some use of GenAI...
- …but you should be careful about it
- If you use GenAI, we strongly suggest that you try writing your report without the use of GenAI first – part of the benefit of writing is that it helps you develop your thinking, and this coursework provides a valuable opportunity to learn how to do this.
- UoE's official guidance on using GenAI: https://information-services.ed.ac.uk/computing/comms-and-collab/elm/generative-ai-guidance-for-students/using-generative

AI can be harmful for learning

"Not surprisingly, use of ChatGPT was likely to develop tendencies for procrastination and memory loss and dampen the students' academic performance." - Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students.

Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance

First published: 10 December 2024

Yizhou Fan, Luzhen Tang, Huixiao Le, Kejie Shen, Shufang Tan, Yueying Zhao, Yuan Shen, Xinyu Li, Dragan Gašević

The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, Nicholas Wilson

CHI 2025 | April 2025

AI can have consequences

AI • ACADEMIC RESEARCH

NeurIPS, one of the world's top academic AI conferences, accepted research papers with 100+ AI-hallucinated citations, new report claims

By **Sharon Goldman**
AI Reporter

January 21, 2026, 9:00 AM ET

Add us on

Lawyer caught using AI-generated false citations in court case penalised in Australian first

Victorian solicitor stripped of his ability to practise as a principal lawyer after acknowledging he did not verify case list

AI - Overall

- AI can still be useful and good!
- But you need to be responsible in how you use it!