

How AI Can Help Bring Your Research into the First League

Laura Johanesová, 10.6.2026

Disclaimer

This presentation was created with AI - I deliberately didn't change it and used it exactly as AI gave it to me, 0 edits. The AI **was provided with the text** and it was prompted to create a presentation. I specified the style, the colour palette, how many slides I want, what logos to use. The prompt was quite extensive.

Yet, we don't even have to be graphical experts to see that the presentation looks quite **bad**. There is misalignment, some elements are clearly missing, some bullet points are not spaced out properly, AI has also changed some text and we can recognise the signature "AI writing".

Because we are able to say what is a good presentation and what is a bad presentation, we wouldn't just take this presentation and consider it a job well done, straight away. We would **critically look at it**, think about what the AI gave us and use our own skill and knowledge to decide what is okay and what has to be done/written in a different way.

But **if we know nothing** or very little about a certain topic and we decide to use AI to help us, we need to **be aware that AI makes mistakes all the time**. And the advice or guidance it gives us is very often not the best one. And if we cannot say what is true and what isn't, it's **best to double check the AI's outputs** or at least **use 1-2 other AI tools** to give you another critical opinion.

How AI Can Help Bring Your Research into the First League

Laura Johanesová
CodeBridge College

For university teachers and researchers

AI can help us work with data faster, more clearly and more professionally.

RESEARCH
JUDGEMENT

reproducible

professional

clear

draft

The promise of the webinar

A practical introduction to AI as a smart expert assistant for data work.

- process quantitative and qualitative data more quickly
- create graphs and visualisations at a stronger standard
- speed up your own research without losing judgement
- transfer data and AI literacy into teaching

**The goal is not automation
for its own sake.
The goal is clearer research
thinking.**

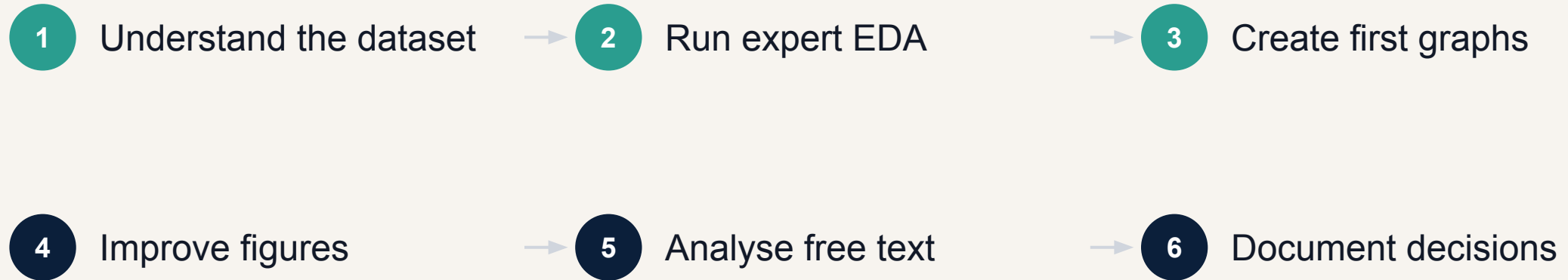


AI as assistant

Researcher as owner

What you will see today

One coherent path: from a messy dataset to a defensible research output.



**Main practical example today:
Exploratory Data Analysis.**

EDA is a perfect visual entry point into serious data work.

You do not need to become a data scientist

You can borrow the most useful data science tools and use them responsibly.

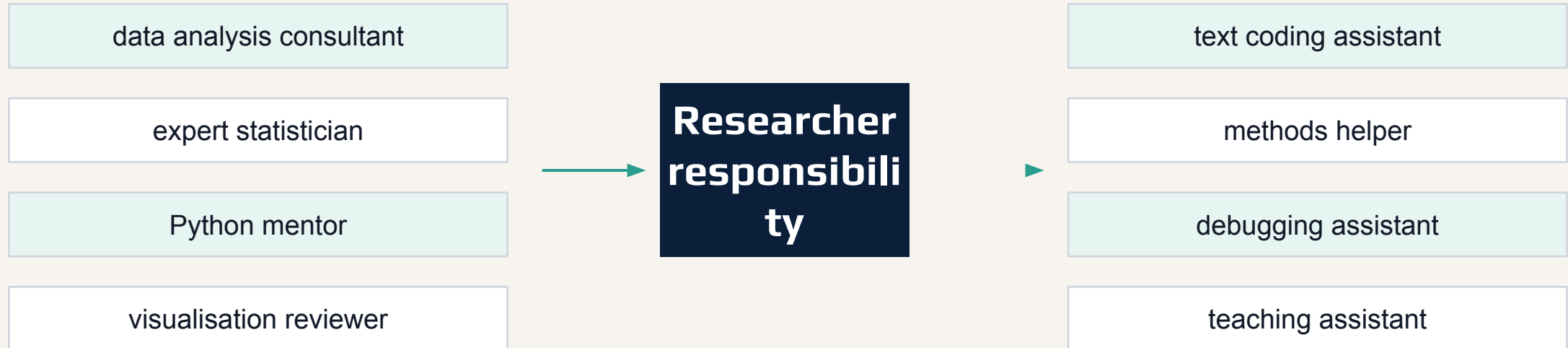
- look at a dataset before making claims
- ask good questions with respect to the research goal
- check data quality and avoid obvious traps
- create clear graphs and use AI safely
- avoid overclaiming when the evidence is only descriptive

BORROW TOOLS

**KEEP
RESPONSIBILITY**

AI as a research assistant

AI is most useful when each role has a clear purpose.



AI helps with speed, structure and clarity. The researcher stays responsible.

Where AI fits in the research workflow

Use AI to make your thinking clearer at each stage, not to replace the thinking.



- inspect a dataset and find useful variables
- suggest data quality checks and first visualisations
- explain statistical options and assumptions
- write, improve and debug code
- draft a methods paragraph or teaching exercise

Exploratory Data Analysis is the detective phase

Before we calculate, we investigate.

- look carefully at what the data actually contain
- find problems before formal analysis
- understand variables and units of analysis
- create first tables and graphs
- decide what can and cannot be analysed



EDA turns “I have a file” into “I understand what this file can support.”

Two questions before any graph

These two questions prevent most confused analyses.

1. What does one row represent?

- a person, patient, student, school or publication
- one country, one answer or one event
- the unit of analysis decides what counts as evidence

2. What is the main goal?

- every dataset can produce hundreds of graphs
- we want useful graphs connected to a question
- the question guides which variables to inspect

If we misunderstand the row, we misunderstand the whole analysis.

Data comes in three forms

Most real research data are mixed, so each column needs a different inspection strategy.

Structured data

survey data
clinical records
publication databases
learning analytics

Unstructured data

interviews
student essays
PDF documents
emails and notes

Mixed data

numbers
categories
dates
free text
metadata

**Numbers need distributions. Categories need counts. Text needs examples and coding.
Dates need time patterns.**

Expert EDA workflow

A compact workflow that works with almost any dataset.

- | | |
|-------------------------------|---------------------------------|
| 1 Load data | 1 Check duplicates |
| 2 Check rows and columns | 2 Inspect key variables |
| 3 Define the unit of analysis | 3 Create first graphs |
| 4 Inspect column names | 4 Write limitations |
| 5 Check data types | 5 Prepare context for AI |
| 6 Check missing values | 6 Decide the next analysis step |

Professional EDA is not “just playing with data”. It is structured investigation.

Start with a data inventory

A data inventory tells us what we have before we interpret anything.

column name

data type

missing values

unique values

example values

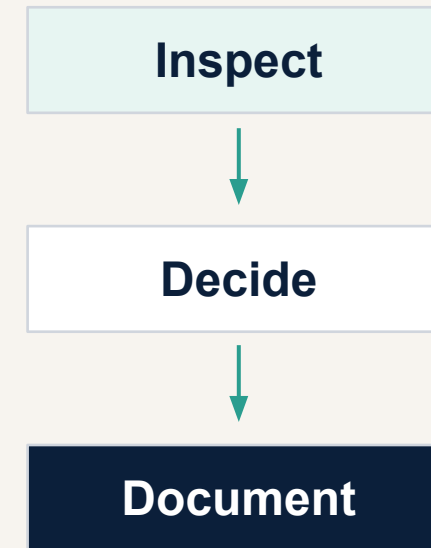
research relevance

This is the first professional output of EDA. A good inventory saves hours of confusion later.

Data quality decisions: missingness and duplicates

Problems are not always errors. First inspect, then document the decision.

- Which variables are missing, how much is missing and for which groups?
- Does missingness change the interpretation of a result?
- Are there identical rows, repeated IDs or repeated answers?
- Are duplicates harmless or do they inflate results?



Inspect variables according to their type

Numbers, categories, text and dates do not need the same checks.

Numeric

min / max
mean / median
quartiles
outliers
distribution shape

Categorical

counts
percentages
rare categories
labels
group sizes

Text

missing values
text length
language
examples
possible themes

Dates

response period
peaks
gaps
seasonality
time patterns

Always ask: are the values plausible?

First graphs should answer one question

Start simple. Remove graphs that do not serve the research question.

Missingness

Which variables have gaps?

Distribution

How is the numeric variable shaped?

Bar chart

Which category is most common?

Boxplot by group

How do groups differ?

Scatter plot

Do two numeric variables move together?

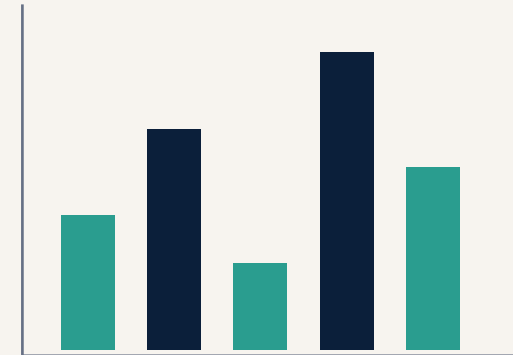
Line chart

What changes over time?

Heatmap

Where are grouped percentages concentrated?

One graph = one question



From first graph to publication-quality figure

A strong figure is easy to read, honest and reproducible.

clear title

axis labels and units

readable font size

source note

simple legend

sensible scale

sample size visible

saved file and caption

Good figures can be reproduced. That is what makes them professional.

AI can help with statistics, cautiously

Use AI to understand options and assumptions, not to outsource judgement.

- which test fits a descriptive question
- which assumptions need checking
- how to interpret p values and confidence intervals
- how to compare groups and describe uncertainty
- how to avoid causal overclaiming

Example prompt:

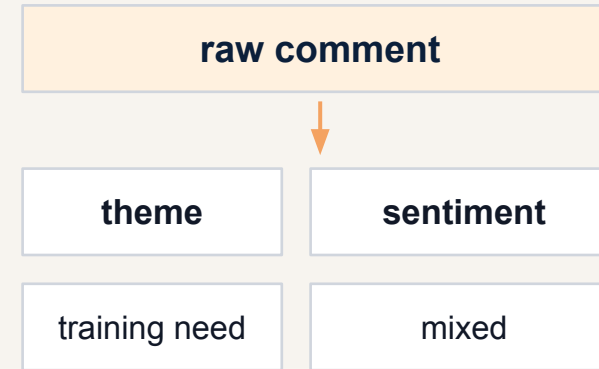
I have one categorical outcome and one categorical group variable. Which descriptive statistics and simple tests should I consider? Explain the assumptions in simple language.

**AI should explain.
The researcher decides.**

Free text inside a table

Open comments, essays and notes become useful when we create structure from them.

- inspect missing values, text length, language and duplicates
- read short examples and very long examples
- look for common words and possible themes
- define the purpose before coding
- draft a first codebook with AI, then verify it



From text to structure

Information extraction finds a specific piece of information and saves it as a variable.

Example text:

“I use ChatGPT for literature search, lesson planning and Python code.”



uses_ai_for_literature_search = yes

uses_ai_for_teaching = yes

uses_ai_for_coding = yes

mentioned_tools = ChatGPT

- people, organisations, countries and dates
- tools, methods, outcomes and recommendations
- barriers, topics, claims and action points

Codebook-driven qualitative workflow

AI assisted coding is useful only when the categories are defined, checked and revised.

1. Keep raw text safe

2. Read a small sample

3. Define the goal

4. Draft categories

5. Create codebook

6. Code sample

7. Check ambiguities

8. Revise

9. Code more text

10. Report limitations

A codebook defines: code name, definition, include when, exclude when, example phrase, ambiguity and human verification.

Working with unstructured documents

Documents need structure first; tables make them easier to inspect.

- PDF reports, transcripts, policy documents and articles
- meeting notes, emails and open-ended assignments
- AI can help turn documents into inspectable tables

document ID	title
date	source
section	paragraph
topic	summary
extracted facts	codes
quotes	

Give AI good context

Do not paste the whole dataset. Give a compact, anonymised summary.

dataset source

research question

row meaning

column names

data types

missing value summary

value counts

small anonymised sample

known limitations

Better context produces better advice and protects sensitive data.

AI agents need clear boundaries

Structured instructions make AI work safer and more useful.

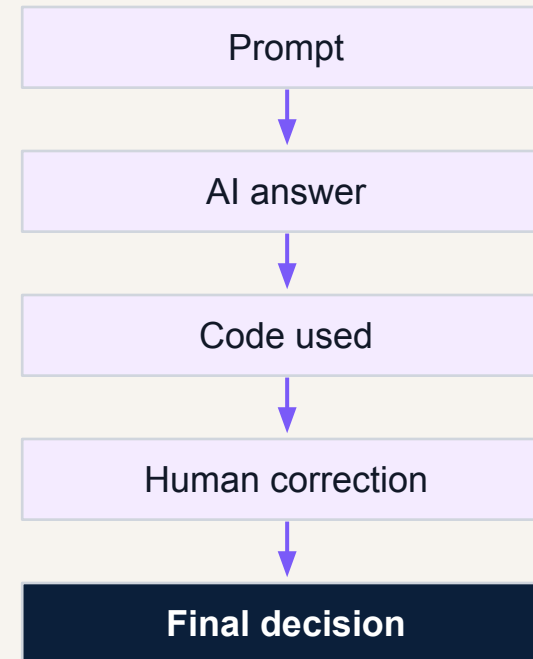
Task	suggest first graphs	Forbidden actions	no final findings
Input	column summary + research question	Expected output	table of graph ideas
Allowed actions	propose graph ideas	Verification	researcher checks variables and runs code

Good AI work is structured work.

Verification, reproducibility and audit trail

AI outputs are drafts. Serious analysis needs evidence of what happened.

- code uses real column names
- percentages are calculated correctly
- filters and row removal are visible
- interpretation is supported by data
- sample size and missing values are clear
- limitations and human decisions are saved



Transfer the workflow into teaching

Students learn data literacy and AI literacy together.

- give students a real dataset and a research question
- ask for a data inventory, missingness table and two first graphs
- include one AI prompt, one AI response and one verification note
- finish with one limitation paragraph

Assess: clarity, data quality thinking, correct graphs, careful interpretation, transparent AI use and reproducibility.

Final takeaway

**AI helps with the work.
The researcher keeps
responsibility.**

speed

structure

coding

visualisation

statistics

documentation

teaching

The deeper skill is research judgement.

Be curious

Be careful

Check the data

Check the code

Check the graph

Check the claim

Document the process