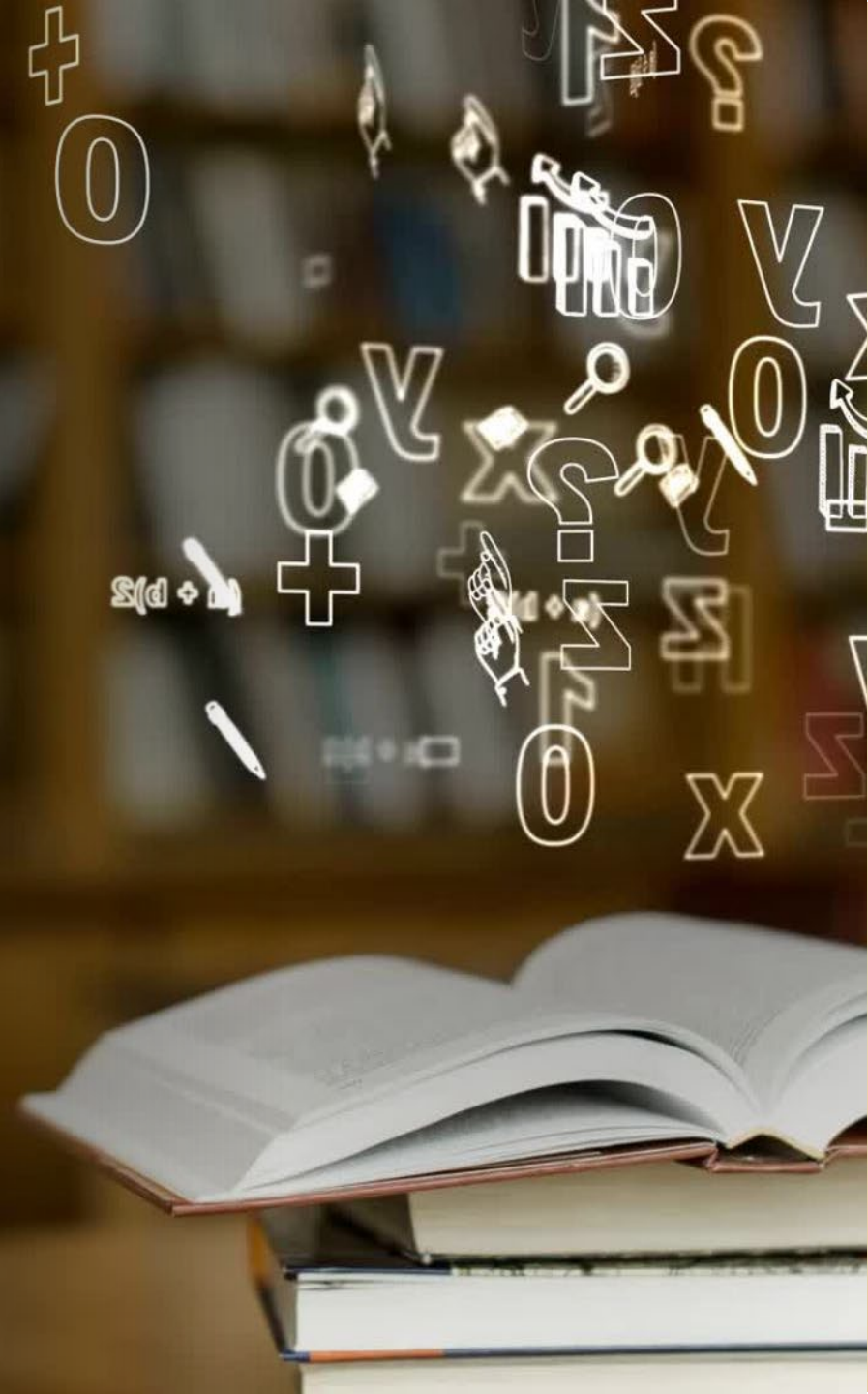




How Libraries Can Leverage AI to Help Researchers Find Joy in Research Again

By Joshua Been and Millicent Weber

Baylor University Libraries



Outline

Before → After → How

- We will use this format to frame all of our examples

AI Chat Interface Examples

- Survey Design & Qualtrics
- Data Prep and Analysis
- Research Data Management documentation

Local AI Workflows

- Extracting and Classifying Primary Documents
- Supporting Evidence Synthesis
- Automating Identification of Human Settlements
- OCR Complex Rows and Columns

Survey Design Challenges

Rate your agreement: The librarians provide helpful research support and timely responses.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

How often do you use the library?

Never

A little

A moderate amount

A lot

A great deal

How satisfied are you with library services?

Unsatisfied

Somewhat satisfied

Satisfied

Very satisfied

How many times have you checked out books from the library this semester?

0–5 times

5–10 times

10–15 times

What library service do you use the most?

Library webchat

Study rooms

Databases

Survey Challenges

- Poor question wording
- Ambiguous/Double-barreled questions
- Overlapping and/or improper response scales
- Not measuring what you think you are

Survey: Before and After

Rate your agreement: The librarians provide helpful research support and timely responses.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

How often do you use the library?

Never

A little

A moderate amount

A lot

A great deal

How many times have you checked out books from the library this semester?

0–5 times

5–10 times

10–15 times

Library staff provide helpful research support.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

Library staff respond to inquiries in a timely manner.

Strongly disagree

Disagree

Neither agree nor disagree

Agree

Strongly agree

Approximately how often do you use library services?

Never

Rarely (a few times per semester)

Occasionally (monthly)

Frequently (weekly)

Very frequently (multiple times per week)

How many times have you checked out books from the library this semester?

0

1–5

6–10

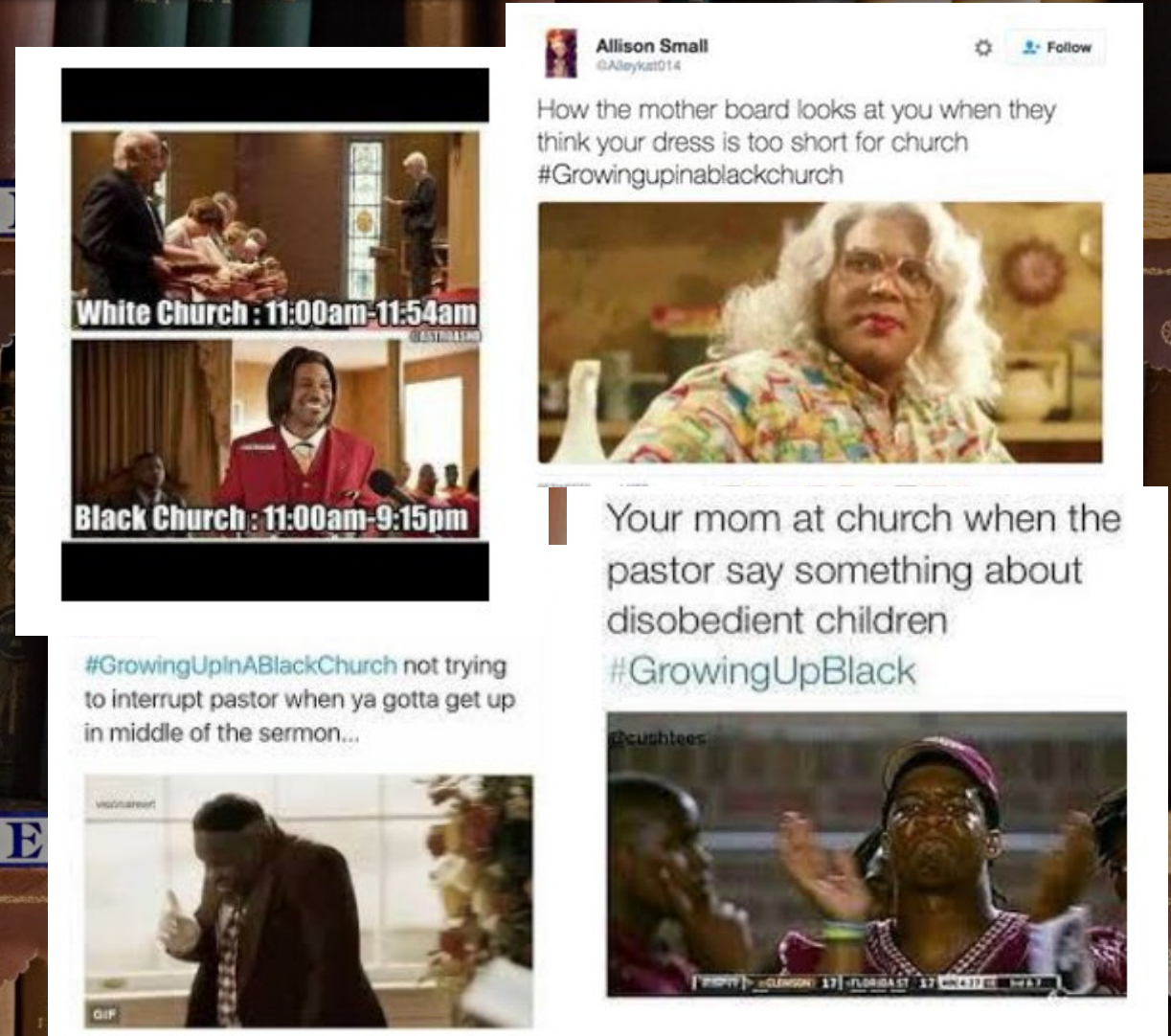
11–15

16 or more

Black Church Memes Research Project - Before

Messy Image Data

- Researchers didn't have the image files; only copy/pasted versions on a PDF
- Need to qualitatively theme 100+ images
- Didn't have standardized method for theming
- Maybe a survey format?



Black Church Memes Research Project - After



Your mom at church when the pastor say something about disobedient children
#GrowingUpBlack



PDF of disorganized meme images

Clean Qualtrics survey for theming and analyzing each image



Allison Small
@Alleykat014

How the mother board looks at you when they think your dress is too short for church
#Growingupinablackchurch



Does this meme reflect Traditional Black Christianity or Digital Black Christianity?

- ☐ Traditional Black Christianity (i.e., the meme expresses traditional Black Church focuses)
- ☐ Digital Black Christianity (i.e., the meme appears designed to attract youth or extend ministry beyond physical services)

Primary theme that matches this meme:

- ☐ Growing Up in the Black Church
- ☐ Black Church Choir / Worship
- ☐ Intergenerational Social Commentary
- ☐ Race, Culture & Social Issues
- ☐ General Faith Inspiration & Encouragement

(If applicable) Secondary theme that matches this meme:

- ☐ Growing Up in the Black Church
- ☐ Black Church Choir / Worship

Survey Design - How

Black Church Meme Project

- AI extracted the images
- Human created Qualtrics survey template
- Template QSF file uploaded into AI
- AI generated the new survey (QSF file) with all images and questions

AI as your Survey Design Assistant

- Can improve your questions and generate new ones
- Can create QSF file from your list of questions
- Other uses: Pilot-test surveys (simulate survey-takers/data)

Create a new project

Survey

Name

Untitled project

Folder

Projects and Programs

How do you want to start your survey?

Create a blank survey project

Create a blank survey project

Import a QSF file

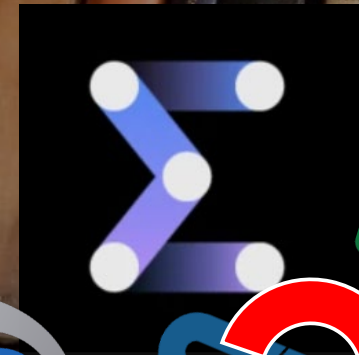
Copy a survey from an existing project

Use a survey from your library

Data Prep and Analysis

The Pain: Tedious, time-consuming Data Cleaning

- Standardizing/transforming columns
- Coding
- Identifying outliers and missing data
- Knowing which tools, procedures, analyses to do
- Human error



Food Pantry Research Project - Before

Messy Data

- Client survey data (paper/pencil)
- Student workers – Excel manual data entry
- Data in wrong format, not coded
- Can't analyze in this format
- PI not data savvy

	A	B	C	D	E	F	G	H	I	J	K
1	Time_usin	Visit_frequ	Reasons_hard_t	Staff_welc	Wait_reas	Comfortab	Food_fres	Variety_av	Choose_o	Volunteers	Cho
2	1-2 years	A few times	It's not easy for	Rarely	Often	Often	Rarely	Sometime	Always	Often	Yes
3	1-6 month	Once a mor	I can't get there	Never	Rarely	Sometime	Never	Never	Never	Always	Yes
4	1-6 month	A few times	The food pantry	Often	Rarely	Often	Often	Never	Often	Often	No
5	First time	First time	I worry about wl	Never	Always	Often	Always	Always	Often	Always	Yes
6	1-2 years	A few times	I have no trouble	Always	Sometime	Sometime	Never	Never	Never	Often	No
7	1-2 years	A few times	It's not easy for	Often	Rarely	Rarely	Never	Rarely	Never	Always	No
8	7-11 mont	A few times	I can't get there	Often	Never	Sometime	Often	Always	Never	Sometime	Yes
9	1-6 month	Once a mor	The food pantry	Sometime	Sometime	Sometime	Always	Sometime	Rarely	Sometime	No
10	2+ years	A few times	I worry about wl	Often	Often	Often	Never	Rarely	Never	Always	No
11	1-6 month	A few times	I have no trouble	Rarely	Never	Rarely	Often	Never	Often	Often	No
12	2+ years	Once a wee	It's not easy for	Always	Rarely	Rarely	Rarely	Sometime	Always	Sometime	No
13	1-2 years	Once a mor	I can't get there	Never	Rarely	Sometime	Rarely	Often	Often	Rarely	No
14	7-11 mont	Once a mor	The food pantry	Sometime	Never	Rarely	Always	Never	Sometime	Sometime	No
15	1-2 years	A few times	I worry about wl	Never	Never	Sometime	Never	Often	Always	Always	Yes
16	1-6 month	Once a wee	I have no trouble	Always	Sometime	Sometime	Often	Often	Rarely	Rarely	Yes
17	1-6 month	A few times	It's not easy for	Rarely	Rarely	Sometime	Rarely	Often	Sometime	Always	Yes
18	7-11 mont	Once a mor	I can't get there	Always	Sometime	Sometime	Often	Rarely	Always	Always	No
19	2+ years	A few times	The food pantry	Always	Sometime	Often	Always	Sometime	Never	Rarely	No
20	1-6 month	Once a mor	I worry about wl	Sometime	Always	Rarely	Rarely	Always	Often	Never	No
21	2 years	A few times	I have no trouble	Always	Sometime	Rarely	Rarely	Rarely	Often	Sometime	No

✖ ✓ fx ▾		=AVERAGE(E2:L2) + AVERAGE(Q5:S5)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
----------	--	----------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

After: Clean, Coded Data

	A	B	C	D	E	F	G	H	I	J	K
1	Time_usin	Visit_freque	Reasons_hard	Staff_welc	Wait_reas	Comfortab	Food_fres	Variety_av	Choose_o	Volunteers	Food_amd
2	4	5	It's not easy fi	2	4	4	2	3	5	4	4
3	2	4	I can't get the	1	2	3	1	1	1	5	1
4	2	4	The food pan	2	2	4	4	1	4	4	2
5	1	1	I worry about	1	5	4	5	5	2	5	4
6	4	3	I have no trou	5	3	3	1	1	1	4	3
7	4	3	It's not easy fi	4	2	2	1	2	1	5	1
8	3	3	I can't get the	4	1	3	4	5	1	3	3
9	2	4	The food pan	3	3	3	5	3	2	3	5
10	5	3	I worry about	4	4	4	1	2	1	5	4
11	2	5	I have no trou	2	1	2	4	1	4	4	3
12	5	6	It's not easy fi	5	2	2	2	3	5	3	3
13	4	4	I can't get the	1	2	3	2	4	4	2	3
14	3	4	The food pan	3	1	2	5	1	3	3	4
15	4	5	I worry about	1	1	3	1	4	5	5	1
16	2	6	I have no trou	5	3	3	4	4	2	2	4
17	2	3	It's not easy fi	2	2	3	2	4	3	5	4
18	3	4	I can't get the	5	3	3	4	2	5	5	4
19	5	5	The food pan	5	3	4	5	3	1	2	1
20	2	4	I								3
21	4	3	I								4

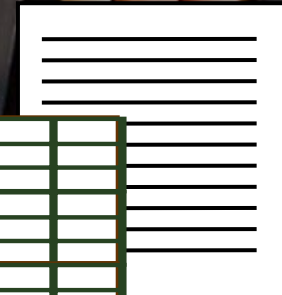
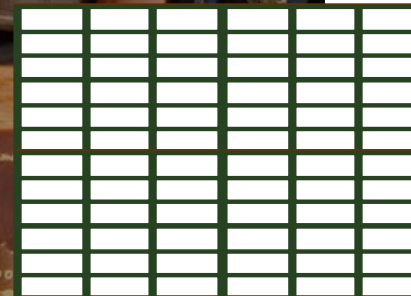
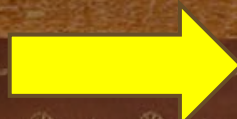
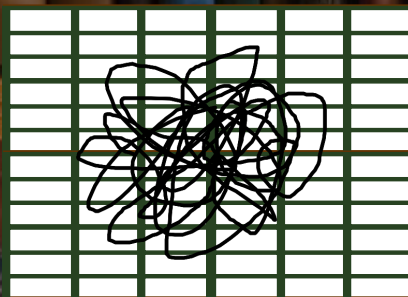
Variables coded numerically, with the codebook (shown to the right) generated in a separate sheet in the same file.

	A	B	C
1	Variable	Code	Label
2	Time_using_pantry	1	First time
3	Time_using_pantry	2	1–6 months
4	Time_using_pantry	3	7–11 months
5	Time_using_pantry	4	1–2 years
6	Time_using_pantry	5	More than 2 years
7	Visit_frequency	1	First time
8	Visit_frequency	2	Once a year
9	Visit_frequency	3	A few times a year
10	Visit_frequency	4	Once a month
11	Visit_frequency	5	A few times a month
12	Visit_frequency	6	Once a week or more
13	Staff_welcome	1	Never
14	Staff_welcome	2	Rarely
15	Staff_welcome	3	Sometimes
16	Staff_welcome	4	Often
17	Staff_welcome	5	Always
18	Food_fresh	1	Never
19	Food_fresh	2	Rarely
20	Food_fresh	3	Sometimes
21	Food_fresh	4	Often
22	Food_fresh	5	Always
23	Food_variety_satisfaction	1	Very dissatisfied
24	Food_variety_satisfaction	2	Somewhat dissatisfied
25	Food_variety_satisfaction	3	Neutral
26	Food_variety_satisfaction	4	Somewhat satisfied
27	Food_variety_satisfaction	5	Very satisfied
28	Food_fresh	1	Never
29	Food_fresh	2	Rarely

Food Pantry Research Project - How

AI helps you find joy in data cleaning

- Uploaded the Excel data file to Copilot.
- Explained to Copilot how we wanted the variables coded.
- Copilot output an Excel file with the data coded.
- Then had Copilot create a Codebook on a new sheet in that Excel file.



Health Research Project - Before

Large, Disorganized Dataset

- Heart rate data, testing effects of taking work breaks on heart rate variability (HRV)
- 500,000+ rows of data
- Incorrect format for analysis
- Data needed large-scale transformations
- Unsure how to analyze

Mixed ANOVA?

rr_interval?

Pvthon or R or SPSS?

Linear Mixed Effects Model?

RMSSD?

Condition x timepoint

	A	B	C	D	E	F
1	participant	condition	time_point	date	time	RR_interval_ms
556911	SG6041	vr	1	10/14/2025	15:45:30	783
556912	SG6041	vr	1	10/14/2025	15:45:30	800
556913	SG6041	vr	1	10/14/2025	15:45:30	792
556914	SG6041	vr	1	10/14/2025	15:45:30	703
556915	SG6041	vr	1	10/14/2025	15:45:30	728
556916	SG6041	vr	1	10/14/2025	15:45:30	703
556917	SG6041	vr	1	10/14/2025	15:45:30	672
556918	SG6041	vr	1	10/14/2025	15:45:30	647
556919	SG6041	vr	1	10/14/2025	15:45:30	703
556920	SG6041	vr	1	10/14/2025	15:45:30	703
556921	SG6041	vr	1	10/14/2025	15:45:30	688
556922	SG6041	vr	1	10/14/2025	15:45:30	672
556923	SG6041	vr	1	10/14/2025	15:45:30	695
556924	SG6041	vr	1	10/14/2025	15:45:30	728
556925	SG6041	vr	1	10/14/2025	15:45:30	720
556926	SG6041	vr	1	10/14/2025	15:45:30	663
556927	SG6041	vr	1	10/14/2025	15:45:30	680
556928	SG6041	vr	1	10/14/2025	15:45:30	672
556929	SG6041	vr	1	10/14/2025	15:45:30	680
556930	SG6041	vr	1	10/14/2025	15:45:30	655
556931	SG6041	vr	1	10/14/2025	15:45:30	647
556932	SG6041	vr	1	10/14/2025	15:45:30	680
556933	SG6041	vr	1	10/14/2025	15:45:30	688
556934	SG6041	vr	1	10/14/2025	15:45:30	688

Health Research Project - After

	A	B	C	D	E	F
1	participant	condition	time_point	date	time	RR_interval_ms
556911	SG6041	vr	1	10/14/2025	15:45:30	783
556912	SG6041	vr	1	10/14/2025	15:45:30	800
556913	SG6041	vr				
556914	SG6041	vr				
556915	SG6041	vr				
556916	SG6041	vr				
556917	SG6041	vr	1	10/14/2025	15:45:30	672
556918	SG6041	vr	1	10/14/2025	15:45:30	647
556919	SG6041	vr	1	10/14/2025	15:45:30	703
556920	SG6041	vr	1	10/14/2025	15:45:30	703
556921	SG6041	vr	1	10/14/2025	15:45:30	688
556922	SG6041	vr	1	10/14/2025	15:45:30	672
556923	SG6041	vr	1	10/14/2025	15:45:30	695
556924	SG6041	vr	1	10/14/2025	15:45:30	728
556925	SG6041	vr	1	10/14/2025	15:45:30	720
556926	SG6041	vr	1	10/14/2025	15:45:30	663
556927	SG6041	vr	1	10/14/2025	15:45:30	680
556928	SG6041	vr	1	10/14/2025	15:45:30	672
556929	SG6041	vr	1	10/14/2025	15:45:30	680
556930	SG6041	vr	1	10/14/2025	15:45:30	655
556931	SG6041	vr	1	10/14/2025	15:45:30	647
556932	SG6041	vr	1	10/14/2025	15:45:30	680
556933	SG6041	vr	1	10/14/2025	15:45:30	688
556934	SG6041	vr	1	10/14/2025	15:45:30	680

~200 rows per participant per timepoint. 4 timepoints. ~800 rows per participant. 250 participants. Over 500,000 rows of data.



	A	B	C	D	E
1	participant	condition	time_point	RMSSD	n_rr
933	SC1335	talk	1	41.78016	331
934	SC1335	talk	2	88.17525	498
935	SC1335	talk	3	274.281	246
936	SC1335	talk	4	142.9488	386
937	SG0304	none	1	100.4221	269
938	SG0304	none	2	75.10183	696
939	SG0304	none	4	80.20232	689
940	SG6041	vr	1	32.923	448
941	SG6041	vr	2	46.2382	821
942	SG6041	vr	3	44.50574	591
943	SG6041	vr	4	118.2824	459
944	SH5906	insta	1	74.80873	362
945	SH5906	insta	2	78.67421	454
946	SH5906	insta	3	88.85753	534
947	SH5906	insta	4	128.993	483
948	SK2515	cross	1	37.23097	360
949	SK2515	cross	2	66.39835	59
950	SK2515	cross	3	53.25963	516
			4	79.85192	544
			1	40.452	410
			2	49.20221	901
			3	101.4253	561
			4	77.60119	440
			1	114.7504	201

Converted heartbeat interval data to proper format for analysis (RMSSD). 1 row per timepoint. Now just 1000 rows of data.

RMSSD_by_participant_timepoint

Health Research Project - How

AI Assists and Guides Data Prep/Analysis

- Didn't upload data to Copilot.
- Described to Copilot what the columns were and what we wanted to calculate.
- Output Python code for Google Colab.
- Guidance on analyses, instructions for SPSS and R

```
1 # defining the RMSSD calculation function - telling python how to
  calculate RMSSD
2 def compute_rmssd(rr_series):
3     """
4     Compute RMSSD from a series of RR intervals (ms).
5     """
6     rr = rr_series.dropna().to_numpy()
7
8     # Need at least 2 RR intervals to calculate RMSSD, so this is
9     # checking to make sure the data has at least 2 RR data points per
10    # timepoint per participant. And then running the calculation below.
11    if len(rr) < 2:
12        return np.nan
13
14    rr_diff = np.diff(rr)
15    rmssd = np.sqrt(np.mean(rr_diff ** 2))
16    return rmssd
```

```
1 # Calculate RMSSD per participant per timepoint - This produces one
  RMSSD value per timepoint per participant, then merges it back into
  the original dataframe as a new column.
2 rmssd_df = (
3     df
4     .groupby(["participant_id", "time_point"], as_index=False)
5     .agg(RMSSD=("RR_interval_ms", compute_rmssd))
6 )
```



Research Data Management

Before: Common challenges in RDM

- Documentation is one of the biggest gaps in research workflows.
- **The issue:** Forgetting to include a README file, and providing poor metadata.

After: How AI enhances data documentation workflows

- Draft README files (dataset description, file structure, variable definitions/coding, etc.)
- Draft structured metadata (title, description/abstract, keywords, etc.)
- AI can tailor output for specific repositories (such as TDR/Dataaverse)

Poor README

HRV Study Dataset

Variables:

- `participant_id`: unique participant identifier
- `condition`: break condition
- `time_point`: measurement point during the session
- `RMSSD`: HRV value
- `n_rr`: number of RR intervals used in calculation

Break Conditions:

- `cross`: crossword puzzle
- `insta`: Instagram scrolling
- `vr`: virtual reality nature scenes
- `talk`: conversation with another student
- `medit`: meditation
- `none`: no break (control)

Good README

README FILE

HRV, Breaks and Personal Virtues at Work

Overview

This dataset contains tabular physiological measurements associated with a study examining the effects of short break activities on stress. The dataset includes heart rate variability (HRV) measurements collected from participants at multiple time points under different break conditions. HRV is represented using the RMSSD metric. The dataset is intended to support analysis of how different types of breaks influence physiological stress responses.

Files and Directory Structure

- `RMSSD_by_participant_timepoint.csv`
 - Tabular dataset containing HRV-related measurements for participants across multiple time points and conditions.
 - Each row represents a single observation for a participant at a specific time point.

Data Description

File Formats

- CSV (Comma-Separated Values)

Variables / Columns

- `participant_id`: An anonymized identifier for each participant.
- `Condition`: A categorical variable indicating the experimental condition or break type. There are 6 break conditions:
 - `cross`: Crossword puzzle
 - `insta`: scrolling Instagram
 - `vr`: VR headset with nature scenes
 - `talk`: talking with another student
 - `medit`: meditation activity
 - `none`: no break (control condition)
- `time_point`: Numeric indicator of measurement timing within an experimental session. Ranges from 1 to 4. Timepoint 1 is baseline, Timepoint 2 is during a stressful activity, Timepoint 3 is during the break condition, Timepoint 4 is during another round of the stressful activity.
- `RMSSD`: Numeric variable representing Root Mean Square of Successive Differences (HRV metric).
- `n_rr`: Numeric variable indicating the number of RR intervals used to calculate RMSSD.

Units of Measurement

- Units for `RMSSD` are milliseconds.
- Units for `n_rr` are a numeric count of how many RR intervals were used to calculate RMSSD for the respective participant at the respective timepoint.

Methodology

The dataset originates from a controlled experimental study involving:

- Participants completing break activities

Excerpt of Metadata Output

=== METADATA SUGGESTIONS ===

Title

Suggested Value:

HRV, Breaks and Personal Virtues at Work Dataset

Confidence: High

Justification: Directly derived from user-provided project title.

Description

Suggested Value:

This dataset contains heart rate variability (HRV) measurements collected from participants in a study examining the effects of short break activities on physiological stress. Data [include](#) repeated measurements of RMSSD across multiple time points and experimental conditions representing different types of breaks.

Confidence: Medium

Justification: Description aligns with provided project context and file structure but extends beyond file-only evidence and should be verified.

Keywords

Suggested Value:

- heart rate variability
- HRV
- RMSSD
- stress
- breaks
- experimental study
- physiology
- repeated measures

Confidence: Medium

Justification: Reflects variables and inferred research focus; some terms (e.g., stress, experimental study) are inferred from provided description.

The prompt instructed Copilot to provide values for a specified list of Dataverse metadata fields.

Subject

Suggested Value:

Medicine, Health and Life Sciences

Confidence: Medium

Justification: HRV and physiological measurements strongly align with biomedical/health sciences, but no explicit disciplinary label present in files.

Kind of Data

Suggested Value:

Observational data; Experimental data; Time-series (repeated measures)

Confidence: Medium

Justification: Structure (participant × time point) supports repeated measures; experimental framing inferred from description rather than file alone.

Notes for Researcher (Optional but Useful for TDR Deposit)

To strengthen deposit quality, consider adding:

- A codebook with full variable definitions and units
- Explicit definitions of time points
- Expanded methodology section
- Data collection protocol
- Instrumentation details
- License (e.g., CC-BY 4.0)
- ORCID-linked author metadata

The prompt also instructed Copilot to give its confidence rating and justification for each metadata value it suggested.



Shifting Focus to: Advanced AI for Applied Research Support

Research 12k+ Simón Bolívar Documents

```
article-05.txt
File Edit View
DOCUMENTO 5 CARTA A SU TIO DON CARLOS PALACIOS BLANCO, DESDE BILBAO, ACERCA DE LA PROVISION DE FONDOS RECIBIDOS EN
ESPANA Y DE LOS ENVIOS DE CACAO PARA RESPONDER DE ELLOS LE ANUNCIA SU PROXIMA MARCHA DE BILBAO* -

Portada del sitio > 1) Periodo (20MAR 1799 AL 23DIC 1824) Correspondencia Personal () > DOCUMENTO 5 CARTA A SU TIO
DON CARLOS PALACIOS BLANCO, DESDE BILBAO, ACERCA ()

DOCUMENTO 5 CARTA A SU TIO DON CARLOS PALACIOS BLANCO, DESDE BILBAO, ACERCA DE LA PROVISION DE FONDOS RECIBIDOS EN
ESPANA Y DE LOS ENVIOS DE CACAO PARA RESPONDER DE ELLOS LE ANUNCIA SU PROXIMA MARCHA DE BILBAO*

Bilbao, 29 de diciembre de 1801

Estimado tio Carlos: Acabo de recibir [una letra] [1] de tres mil ochocientos ochenta y seis pesos fuertes en
vales reales, que mi [tio Pedro ha cons]eguido

[2] de la casa de Munoz y Orea, a pagarla el 18 de Junio de [1802] [3]en dinero efectivo Por lo que suplico a Vmd
me remita, por mi cuenta y riesgo, doscientas fanegas de cacao, dirigiendolas a cualquiera de los puertos de esta
peninsula, teniendo especial cuidado de comisionar un comerciante de credito conocido y hombria de bien; y dandole
la orden de entenderse directamente con el Sr Dn Bernardo Rodriguez de Toro

[4] quien dispondra del dinero o del cacao, como mejor convenga

Mi marcha sera dentro de dos o tres meses

Ln 1, Col 1 2,676 characters 100% Unix (LF) UTF-8
```

Research 12k+ Simón Bolívar Documents



What type of document?

- Letter?
- Proclamation?
- Legal?
- Other?

Parties?

- If a letter: From and To
- Otherwise, Author

Many other extracted information:

- Date
- Subject themes
- Locations
- Many more...

Simón Bolívar Letters

Before: The
2,000-Hour
Marathon



12,000+ historical documents. Manually reading, classifying, and extracting metadata would take a full researcher-year of tedious data entry. The joy of history is buried in clerical work.

After: The
Coffee-Cup
Database

Wake up, pour coffee, and search. All 12,000 documents are fully parsed, tagged by theme, and searchable by date and location. The researcher instantly steps back into creative scholarship.

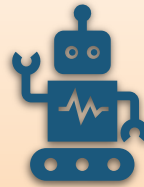
Simón Bolívar Letters: How?



The Brain (Local Open-Weight LLM)

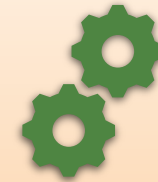
No Cloud, No Cost, Total Privacy, Less Impact on Environment

Model: gpt-oss:120b



The Conveyor Belt (Python Automation)

Automating the Repetitive Work



The Blueprint (Structured JSON)

Forcing the AI to Be Transparent and Consistent

Output Excerpt

	A	B	C	D	E	F	G
1	File	Letter	Sender	Sender Location	Receiver	Receiver Location	Topics
2	article-01.	Yes	Simon Bolivar	Veracruz	Pedro Palacios Bla	Caracas	Personal Relation
3	article-02.	Yes	Simon Bolivar	Madrid	Don Pedro Palacio	Cadiz	Personal Relation
4	article-03.	Yes	Simon Bolivar	Madrid	Don Pedro Palacio	Unknown	Personal Relation
5	article-04.	Yes	Simon Bolivar	Bilbao	Pedro Palacios Bla	Unknown	Personal Relation
6	article-05.	Yes	Simón Bolívar	Bilbao	Don Carlos Palacio	Caracas	Economic Policie
7	article-06.	Yes	Simon Bolivar	Bayona, 13 de ene	Don Francisco Jose	Unknown	Personal Relation
8	article-07.	Yes	Don Simon Bolivar y	Santander	Don Pedro Rodrigu	Villa y Corte de Mad	Personal Relation
9	article-08.	Yes	Don Simon Bolivar	Madrid	Unknown	Unknown	Personal Relation
10	article-09.	Yes	Simon Bolivar	Madrid	Sr Pbro Licenciado	Madrid	Personal Relation
11	article-10.	Yes	Don Simon de Boliva	Madrid	Unknown	Unknown	Personal Relation
12	article-100	Yes	Antonio Muñoz Teba	Caracas	Director de Rentas	Unknown	Military Strategy
13	article-100	Yes	Simón Bolívar	Los Cayos, San Lui	Alejandro Pétion	Haiti	International Dip

Evidence Synthesis



Evidence Synthesis & The Rigor of Extraction

Before: Cognitive Fatigue

- Extracting complex scientific variables (dosages, sample sizes, exact p-values) from dense PDFs is exhausting.
- Human transcription errors are common, and reconciling differences between two manual extractors takes endless meetings.

After: Audited Efficiency

- The researcher reviews a pre-populated spreadsheet where every extracted data point is accompanied by its exact source quote and page number.
- Every extraction contains the cited evidence and the model's rationale for the decision.
- The workload is diminished, and meetings are spent on synthesis, not spreadsheet reconciliation.



Evidence Synthesis: How

Long-Context & Verified Extraction

- **Deep Document Reading:** Utilizing larger, long-context local models optimized for parsing 30+ page scientific papers.
- **Guaranteed Data Types:** Implementing strict schemas to force the LLM to output clean, structured scientific data (no conversational fluff).
- **Absolute Provenance:** Requiring the AI to cite the exact paragraph or table, transforming the researcher's role from "data entry" to "expert editor."

C9	Few studies have explored student perceptions of stress in PA education (Hernandez et al., 2010; Johnson et al., 2020; O'Brien et al., 2012). Additionally, researchers have given little attention to the structure										
	A	B	C	D	E	F	G	H	I	J	K
1	file_name	design_type	design_type_quote	data_emphasis	design_diagram_des	mixed_methods_not	research_questions	research_questions	purpose_aim_stated	purpose_aim_stated	extraction_status
2	#10900 - Cho 2022.pdf	Sequential Mixed Met	to describe the ment	Mixed	N/A	N/A	TRUE	1) Examine the perce	FALSE	N/A	SUCCESS
3	#1101.pdf	Sequential Mixed Met	The 10-min actual sc	Mixed	N/A	N/A	FALSE	N/A	TRUE	The purpose of this s	SUCCESS
4	#11480 - Choi 2018.pdf	Explanatory Sequenti	An Explanatory Seque	Mixed	N/A	N/A	TRUE	Four research questi	TRUE	The purpose of this s	SUCCESS
5	#11826 - Choi 2022 .pdf	Sequential Mixed Met	Sequential Mixed Met	Mixed	N/A	N/A	FALSE	N/A	TRUE	The purpose of this s	SUCCESS
6	#11979 - Chase 2022.pdf	Sequential Mixed Met	Study 1 used qualitat	Mixed	N/A	N/A	FALSE	N/A	TRUE	The current dissertati	SUCCESS
7	#12466 - Chisum 2018.pdf	Sequential Mixed Met	The model by Mumfo	Mixed	N/A	N/A	TRUE	Research Questions :	TRUE	The purpose of this s	SUCCESS
8	#12516 - Chesaniuk 2023 .pdf	Sequential Mixed Met	Sequential Mixed Met	Mixed	N/A	N/A	FALSE	N/A	TRUE	Aim 1, Hypothesis 1: i	SUCCESS
9	#12922 - Cetrone 2023.pdf	Sequential Mixed Met	Few studies have exp	Mixed	N/A	N/A	TRUE	The extent to which fa	TRUE	The primary purpose	SUCCESS
10	#13127 - Chaivisit 2021.pdf	Sequential Mixed Met	a sequential mixed r	Mixed	Figure 3. Mixed Methc	N/A	FALSE	N/A	TRUE	The purpose of this s	SUCCESS
11	#13713 - Ceron 2016.pdf	Sequential Mixed Met	observations took pl	Mixed	N/A	N/A	FALSE	N/A	TRUE	The comparisons of i	SUCCESS
12	#15233 - Chesson 2017.pdf	Sequential Mixed Met	A design thinker refle	Mixed	N/A	N/A	TRUE	N/A	FALSE	N/A	SUCCESS

Geospatial / Ukraine

Before: High Stakes & Crushing Timelines

- **The Scene:** A Ukrainian-born political science faculty member in 2022 needs to analyze human settlement agglomerations using remote sensing data.
- **The Pain:** This would require many, many hours of digging through image pixels to draw boundaries around settlements.

After: Accelerated Insight

- **The Scene:** The researcher has a complete, highly accurate vector map of settlement agglomerations, ready for immediate political science and spatial analysis.
- **The Reality:** Instead of spending months clicking on a screen, the researcher gets the foundational data in a fraction of the time.

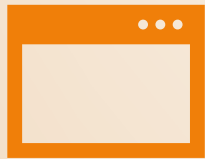
AI-Identified Agglomerations

Workflow:

- Guided the researcher from manual digitization towards a geospatial AI Deep Learning toolset.



Breaking the OCR Barrier



Before: Visual Fragmentation

Classical OCR tools will have problems creating a single table with this layout.



After: AI Spatial Extraction

Researcher provides multiple volumes and the AI reads the page like a human scholar.



How

Swap text-only AI model for a model with vision (qwen3.5:122)

Modern models with vision capabilities possess true spatial reasoning: margins, tables, and column alignments.

Intelligent OCR: Output

D3	Pas-de-Calais - Bethune - Argny				
	A	B	C	D	E
1	Name	Profession	Domicile	Property_Location	Date
2	LORADOUR Mureillac		Vienne - Poitiers - Asnières	Vienne - Poitiers - Asnières	19 novembre 1792
3	LORAGUEL	Valet de Chambre du dernier tyran	Pas-de-Calais - Bethune - Argny	Pas-de-Calais - Bethune - Argny	13 juillet 1792
4	LORAILLE	Capitaine D'Anstrasie	Calvados - Lizieux - Lairenomière	Calvados - Lizieux - Lairenomière	1er août 1792
5	LORAIN	Chanoine	Loir et Cher - Vendôme - Morée	Loir et Cher - Vendôme - Morée	
6	LORAS (frère)		Rhône et Loire - Lyon - Lyon	Rhône et Loire - Lyon - Lyon	10 avril 1793
7	LORDA (femme)	Marquise	Paris - Paris - Paris	Paris - Paris - Paris	29 août 1793

Summing Up

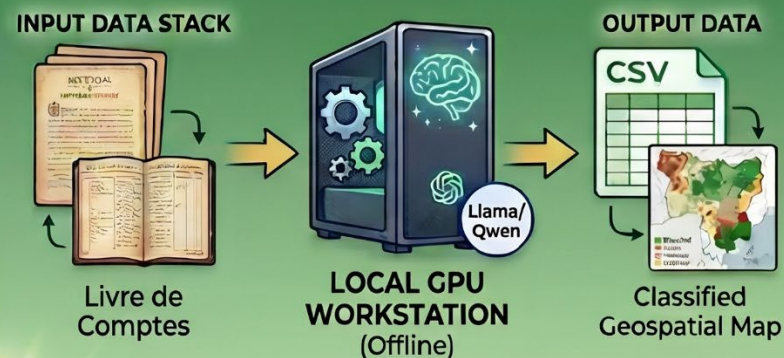
THE LIBRARY AI SPECTRUM: *Matching the Tool to the Task*

Cloud AI Chat Interfaces (The Collaborative Copilot)



Ad-hoc & Interactive
Third-Party Privacy
Create Code for Analysis
Supporting Documentation for Research
Best for Non-Repetitive Tasks

Local Automated Pipelines (The Heavy-Duty Engine)



Mass Scale Automation (e.g., 12k+ Docs),
100% Data Privacy, Zero Recurring Fees
Complex Visual & Spatial Reasoning
Best for Structural & Repetitive Tasks



FINDING JOY IN RESEARCH



Questions?

