RESEARCH • METHODS • CRITICAL AI LITERACY

AI Augmented Workflows

Three ways to investigate one Census research question

David Gustavsen • Jon Wheeler
College of University Libraries & Learning Sciences | University of New Mexico

AI Use Disclosure

How AI contributed to this workshop—and where responsibility remains.

AI-assisted development

ChatGPT and Duck.ai supported planning, drafting, code generation, data analysis, and slide design.

Human review and accountability

The presenters reviewed and revised AI-generated material and remain responsible for its accuracy and interpretation.

Instructional—not validated research

Analyses are demonstrations of research workflows and have not been independently validated.

Start with the room

Audience participation while attendees arrive

Three Mentimeter questions

What does “AI” mean to you?

How often do you use AI?

How confident do you feel using AI?

Or use code **4592 6445** at menti.com

What do the responses tell us?

AI technologies change rapidly. We might answer differently tomorrow!

Definitions

Where do definitions overlap or diverge? Are people thinking about tools, techniques, or social systems?

Experience

Do frequent users report more confidence? Are there notable exceptions?

Expectations

What forms of AI assistance would feel useful—or risky—in research?

About the presenters

David Gustavsen

Assistant Professor
College of University Libraries & Learning Sciences
University of New Mexico

AI requires critical engagement with its capabilities and limitations. We should avoid both hype and reflexive dismissal.

Jon Wheeler

Associate Professor
College of University Libraries & Learning Sciences
University of New Mexico

AI literacy should emphasize quality of use and social benefits—not only competency.

What participants should leave with

Objectives

Cogitate upon critical AI literacy for computational workflows.

Demonstrate AI-assisted strategies for finding and analyzing Census data.

Identify practical and methodological limitations.

Takeaways

AI outputs require mediation.

Efficiency gains depend on task characteristics.

AI can augment—not automatically validate—computational analysis.

Our motivating research question

How does educational attainment compare with broadband access across households in New Mexico?

RESEARCH LANGUAGE

Available data only support comparison. The descriptive Census analyses shown here primarily support associations—not causal effects.

One question, three workflows

A controlled comparison of who does the work and who checks it

01

Human-led

AI supports bounded tasks

02

Hybrid

Human + AI share tasks

03

AI-directed

Human QA/QC checkpoints

Not a ladder of “good to better.” Every scenario relies on human expertise, in different ways.

The dataset sets the boundaries

There is a Census product that already addresses the motivating question, table B28006. We're building upon it.

B28006 measures

Educational attainment × presence of a computer × type of internet subscription.

Universe: household population age 25 and over.

It does not directly measure

Causal effects of education.

Broadband quality, affordability, reliability or infrastructure.

Individual households as the unit in this table.

Use precise wording: The universe and unit of observation must be clearly defined.

SCENARIO 1

Human-directed workflow

AI helps with discovery, documentation and communication—not the core analysis.

LEAD: JW

Step 1 · Search the Census site

Work directly in a published, pre-aggregated Census product

1	**Open**	data.census.gov
2	**Search**	“new mexico educational attainment internet access”
3	**Inspect**	2024 ACS 5-year table B28006
4	**Verify**	Geography, universe, counts, margins of error, and definitions

Source: data.census.gov | Search and interface labels may change.

Step 2 · Read before calculating

Counts can answer some questions, data may allow more detailed comparisons.

What the table gives us

Three education groups.

Computer and internet subscription categories.

Estimates and margins of error.

Considerations

What exactly counts as broadband?

Do differences exceed uncertainty?

Which comparisons are most useful or robust?

Step 3 · Use AI for discovery

An actual preparation prompt — results will differ.

> “I am using census data to compare educational attainment and broadband access across households in New Mexico. I found table B28006 at data.census.gov, but it only shows counts and margins of error. What other products are available from the US Census?”

Demonstrated tool: duck.ai (Fast / GPT-6 Luna during preparation). Model availability can change.

Step 4 · Audit the answer

AI offers leads, not verified research design

Plausible suggestions

Calculate group percentages from counts.

Explore S2801 and DP02, or other tables referenced in the AI output.

Consider county-level comparison.

Verification tasks

Confirm the table IDs exist.

Check each universe, year and geography.

Evaluate precision and recall.

Analytical boundary

Separate within-table comparisons from correlations across topics.

Avoid claiming household-level or causal findings.

Scenario 1 · What efficiencies are gained? At what cost?

AI reduces discovery friction—not the burden of establishing valid comparisons.

Value

Vocabulary, table IDs, alternative approaches, and documented caveats.

Risk

Plausible but mistaken tables, overconfident claims.

SCENARIO 2

Hybrid human–AI workflow

Humans define, inspect and validate; AI supports targeted insertion points.

LEAD: JW

Move from one table to 33 counties

Same question; new unit of comparison

County-level question

Across New Mexico counties, is the share of adults with a BA+ associated with the share living in a household with broadband?

Not the same as...

Are individuals (or households) with higher levels of educational attainment more likely to have broadband access?

Each point in a county-level correlation represents a county, not a person or household.

The hybrid workflow: deliberate AI insertion points

HUMAN	Define variables + verify Census definitions
AI	Suggest tables and outline the process
HUMAN	Download / inspect the source CSV
HUMAN	Calculate NM + Bernalillo manually
AI	Generate repeatable analysis for counties
HUMAN	**Audit results, uncertainty and interpretation**

Step 1 · Download county data

A reproducible manual baseline

01	**Find B28006**	2024 ACS 5-year Detailed Tables
02	**Set geography**	New Mexico → counties; include all 33
03	**Download**	Export CSV for readability or ZIP for machine-ready data
04	**Preserve**	Keep original source, notes, release year and margins of error

Step 2 · Work one example by hand

New Mexico state + Bernalillo County as a reference check

Broadband % among age 25+ = 100 × (Educational_Grouping::Computer_Access::Broadband_Access) / Educational_Grouping

Check the numerator

Broadband-subscription counts across all three educational groupings.

Check the denominator

Total household population 25+ (not number of households) in each educational grouping.

Step 3 · Ask AI to scale the analysis (buyer beware!)

NOT USED: This prompt was suggested by ChatGPT and does not generate the comparison of interest (breakdown by educational attainment is not included).

"Using the exported 2024 ACS 5-year B28006 county table, calculate for each NM county: (1) BA+ share of the population 25+, and (2) the share of that population living in households with broadband. Use the documented Census cell codes. Show the formula for each variable, retain county GEOIDs, check against our Bernalillo result, and generate a scatterplot and Pearson correlation. Do not interpret this as a household-level relationship."

Step 3 · Ask AI to scale the analysis

USED: Downloaded CSV data from the Census had to be converted to TAB delimited format for pasting into duck.ai

"I have US Census table B28006, educational attainment and internet access, for all counties in New Mexico. The data are from the 2024 ACS 5 year estimate. I would like to compare rates of broadband access within each county based on educational attainment. I know I can't upload a CSV file but I can paste the data into the next prompt window."

Manual calculation to establish ground truth

Label (Grouping)	New Mexico!!Estimate	New Mexico!!Percentages_c	Bernalillo County, New Mexico!!Estimate	Bernalillo County, New Mexico!!Percentages_c
Total:	1,433,486		470,317	
Less than high school graduate or equivalency:	167,969		43,213	
Has a computer:	143,159	0.852294173329602	38,867	0.899428412746164
With dial-up Internet subscription alone	253	0.00150623031630837	170	0.000393400134218869
With a broadband Internet subscription	121,769	0.724949246587168	34,639	**0.8015**87485247495
Without an Internet subscription	21,137	0.125838696426126	4,211	0.0974475273644505
No computer	24,810	0.147705826670398	4,346	0.100571587253836
High school graduate (includes equivalency) , some college or associate's degree :	816,442		246,201	
Has a computer:	768,825	0.941677424728272	236,571	0.960885617848831
With dial-up Internet subscription alone	1,345	0.00164739197640494	437	0.00177497248183395
With a broadband Internet subscription	701,534	0.859257608011347	220,933	**0.8973**6841036389
Without an Internet subscription	65,946	0.0807724247405205	15,201	0.0617422350031072
No computer	47,617	0.0583225752717278	9,630	0.0391143821511692
Bachelor's degree or higher:	449,075		180,903	
Has a computer:	442,597	0.985574792629294	179,019	0.989585579011957
With dial-up Internet subscription alone	748	0.00166564605021433	267	0.00147592908906983
With a broadband Internet subscription	425,721	0.947995323720982	173,390	**0.9584**6945600681
Without an Internet subscription	16,128	0.0359138228580972	5,362	0.0296401939160766
No computer	6,478	0.0144252073707065	1,884	0.0104144209880433

Broadband subscription by householder education (AI output)

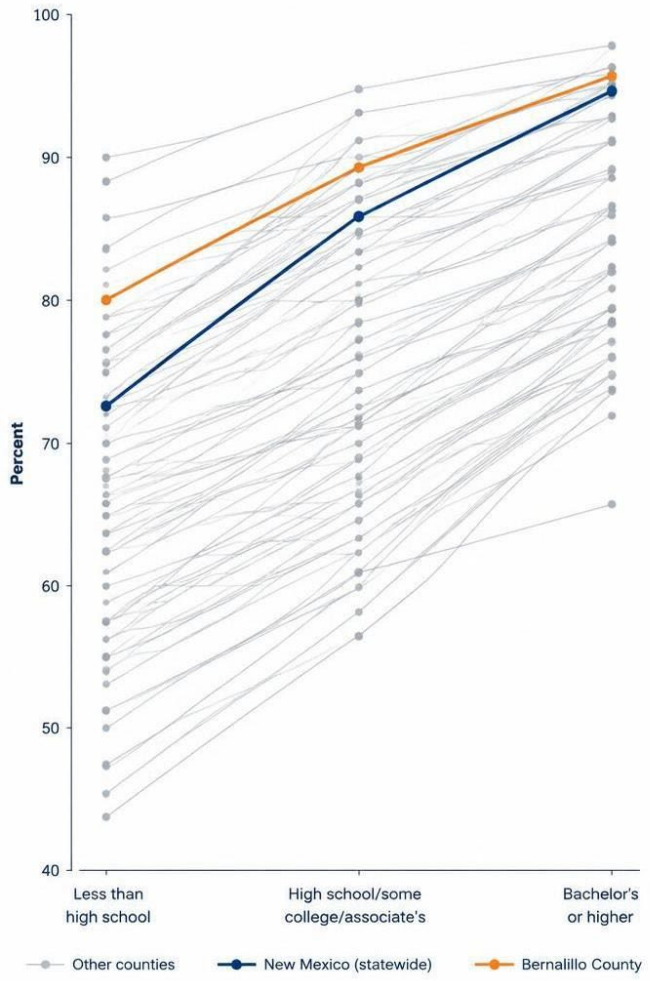
New Mexico counties · ACS 2024 5-year estimates · rates (%)

County	Less than high school	High school / some college / associate's	Bachelor's or higher	Gap (pp)
Bernalillo	80.2%	89.7%	95.8%	+15.6
Catron	80.3%	69.2%	95.1%	+14.8
Chaves	69.6%	85.8%	91.7%	+22.1
Cibola	72.7%	79.7%	90.3%	+17.6
Colfax	64.3%	79.4%	93.6%	+29.3
Curry	79.7%	88.9%	96.1%	+16.4
De Baca	54.5%	87.1%	75.2%	+20.7
Doña Ana	77.8%	91.4%	96.4%	+18.6
Eddy	72.0%	90.4%	93.9%	+21.9
Grant	79.3%	88.2%	92.2%	+12.9

Gap = bachelor's or higher rate minus less-than-high-school rate; percentage points.

Broadband subscription by householder education, New Mexico counties (ACS 2024 5-year)

County broadband subscription rates



Broadband subscription rates by county (sorted by gap)

County	Less than high school	High school/some college/associate's	Bachelor's or higher
Hidalgo	51.7	81.7	99.3
Catron	80.3	69.2	95.1
Lincoln	74.7	84.5	94.8
Sierra	88.0	85.6	91.8
Los Alamos	76.9	94.1	94.2
Doña Ana	77.8	91.4	96.4
Curry	79.7	88.9	96.1
Roosevelt	71.6	89.0	97.6
Sandoval	67.4	89.7	95.8
Bernalillo	80.2	89.7	95.8
Santa Fe	73.7	87.7	96.1
San Juan	58.7	72.8	87.2
Taos	69.3	82.3	95.4
Cibola	72.7	79.7	90.3
Valencia	72.9	85.4	92.6
Eddy	72.0	90.4	93.9
Grant	79.3	88.2	92.2
Socorro	43.6	66.8	95.9
Quay	48.8	76.3	89.1
Rio Arriba	54.4	77.9	87.0
Torrance	75.6	86.5	88.7
Chaves	69.6	85.8	91.7
Lea	79.5	89.0	92.0
Otero	75.6	88.7	93.8
San Miguel	52.5	80.9	92.2
Mora	83.6	78.7	68.7
Union	55.6	86.9	95.7
Colfax	64.3	79.4	93.6
McKinley	41.9	83.2	89.6
Luna	76.1	82.5	89.0
Harding	60.0	81.8	62.7
Guadalupe	70.2	82.6	83.5
Cibola	88.8	82.6	93.5
Guadalupe	70.2	82.6	83.6

Low (≈40%) — High (≈100%)

Rates are estimates; small-county values may be imprecise.

Step 4 · Audit the county correlation

A plot is a useful sanity check but requires validation.

Data checks

All 33 counties?

Correct table and year?

Correct denominator?

No suppressed or missing values?

Statistical checks

Influential counties?

Linear vs nonlinear pattern?

Uncertainty in x and y?

Interpretive checks

Geographic correlation ≠ individual association.

Correlation ≠ causal impact.

Subscription ≠ quality or affordability.

Scenario 2 · What changes?

AI can scale repetitive work after the researcher establishes a trustworthy example.

Potential improvement

Reusability, clear steps, repeatable code, more time for interpretation.

What remains human work

Validation, conceptual decisions, uncertainty and ethical claims.

SCENARIO 3

AI-directed workflow

AI plans and carries out work, Humans audit

DG

Scenario 3 · Working design

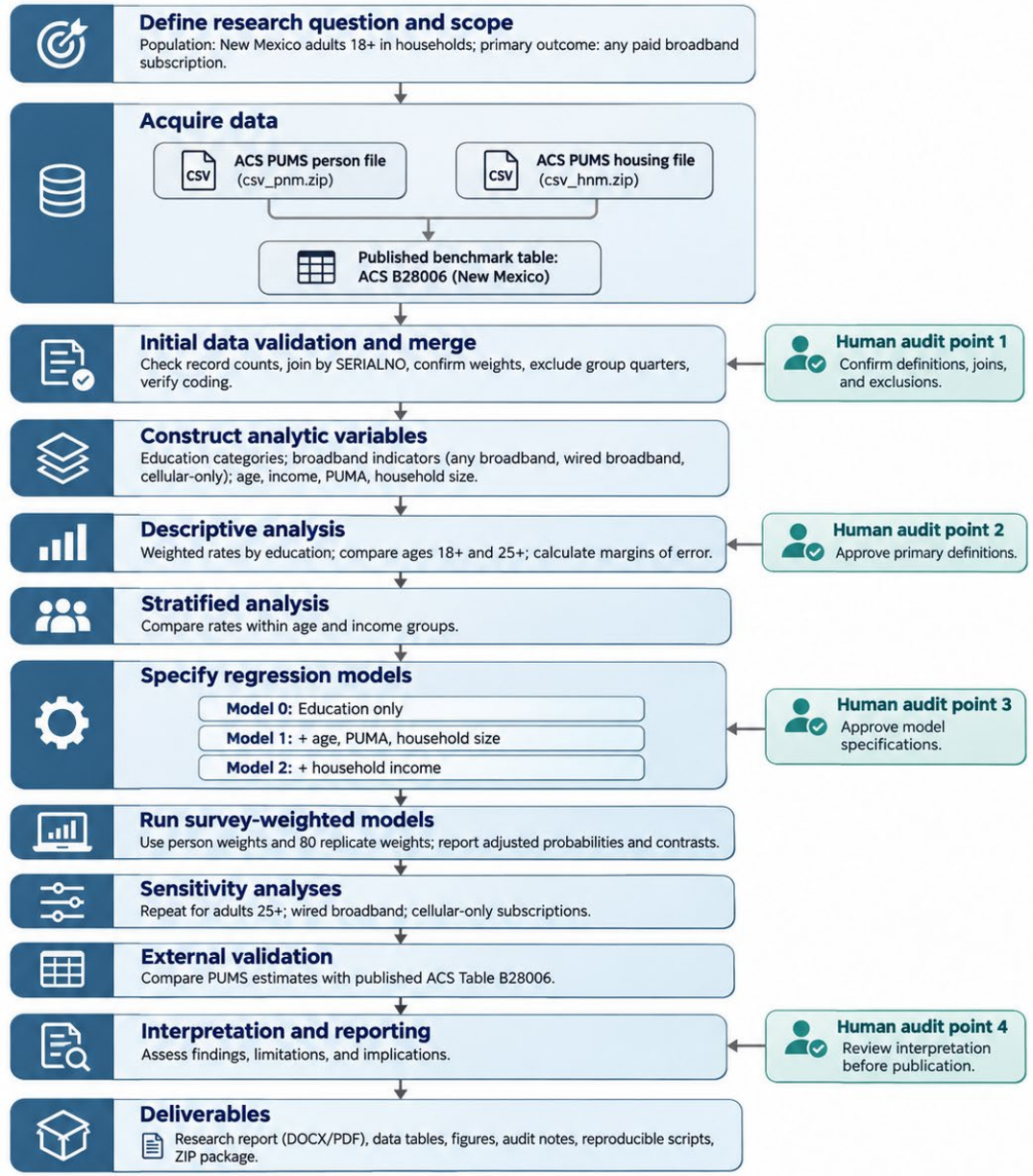
INPUT	Research question + data source + constraints
AI WORK	Plan → retrieve → clean → analyze → visualize
HUMAN CHECK	**Audit provenance, choices, reproducibility and claims**
REVISE	Correct errors and document unresolved uncertainty

Tuesday 9:01 AM

I'd like to study the relationship between education and broadband access in New Mexico using data from the American community survey. Create a plan to find this data and analyze it. If you need me to download or provide any data, let me know and provide instructions. create sensible human audit points.

Workflow: Analyzing the Relationship Between Education and Broadband Access in New Mexico

ACS 2020–2024 Five-Year Data



Key principle: *Use transparent human review at major decision points.*

Tuesday 9:01 AM

scripts.

7. Immediate next step

The best next step is to **establish the published B28006 benchmark, then build the 18+ analysis from the two New Mexico PUMS files**.

For the PUMS stage, upload `csv_pnm.zip` available, the next work should be validatin immediately fitting statistical models.

This sequence gives us a clear point to verif conclusions.

I recommend proceeding with adults **18+** as the primary population and **any paid broadband subscription, including cellular-only plans**, as the primary outcome. We would retain ages 25+ and wired broadband as secondary comparisons.

Do you approve those two definitions?

After that decision, the next steps are to validate the PUMS estimates against published Census Table B28006, examine the relationship in more detail, and review the variables proposed for regression before fitting models.

- nm_acs2024_5y_broadband_descriptive_preview.csv
  Spreadsheet
- nm_acs2024_5y_broadband_audit_notes.md
  Document
- build_nm_broadband.py
  Code

agree with those definitions

RESEARCH REPORT | 10 NOVEMBER 2026

Educational Attainment and Household Broadband Subscriptions in New Mexico

An analysis of the 2020–2024 American Community Survey 5-Year Public Use Microdata Sample

Executive summary

This study examines how educational attainment is associated with living in a household that has a paid broadband internet subscription in New Mexico. The analysis covers 74,370 sampled adults aged 18 or older living in households, representing approximately 1.62 million adults after survey weighting. The outcome includes cellular data plans, cable/fiber/DSL, satellite, and other non-dial-up subscriptions; it is not a measure of connection quality or individual internet use. [1, 2]

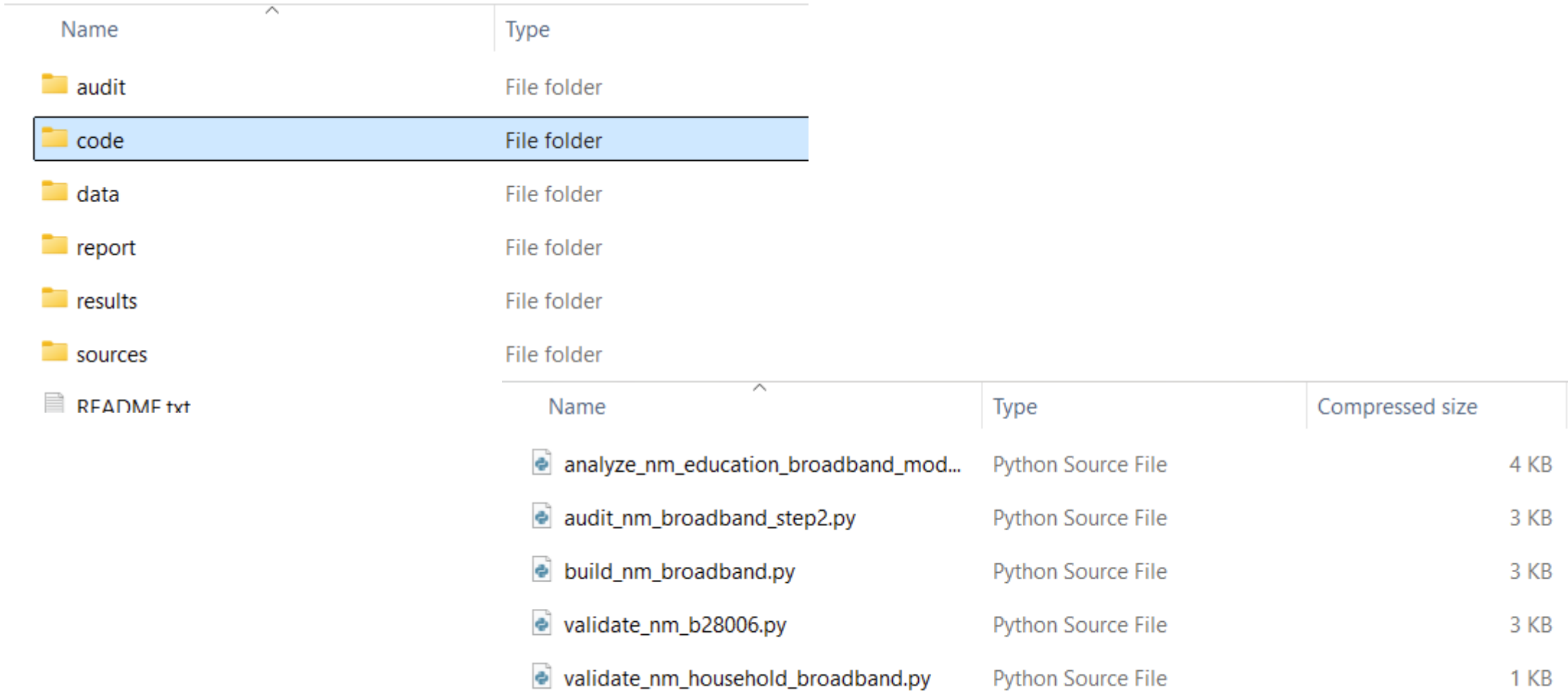
Broadband subscription prevalence increases across all six educational attainment groups, from 75.9% among adults without a high school credential to 95.7% among those with a graduate or professional degree. The unadjusted difference is 19.8 percentage points (90% confidence interval [CI], 18.5–21.1). After adjusting for age, Public Use Microdata Area (PUMA), household size, and household income, the modeled difference is 14.3 points (90% CI, 13.1–15.6).

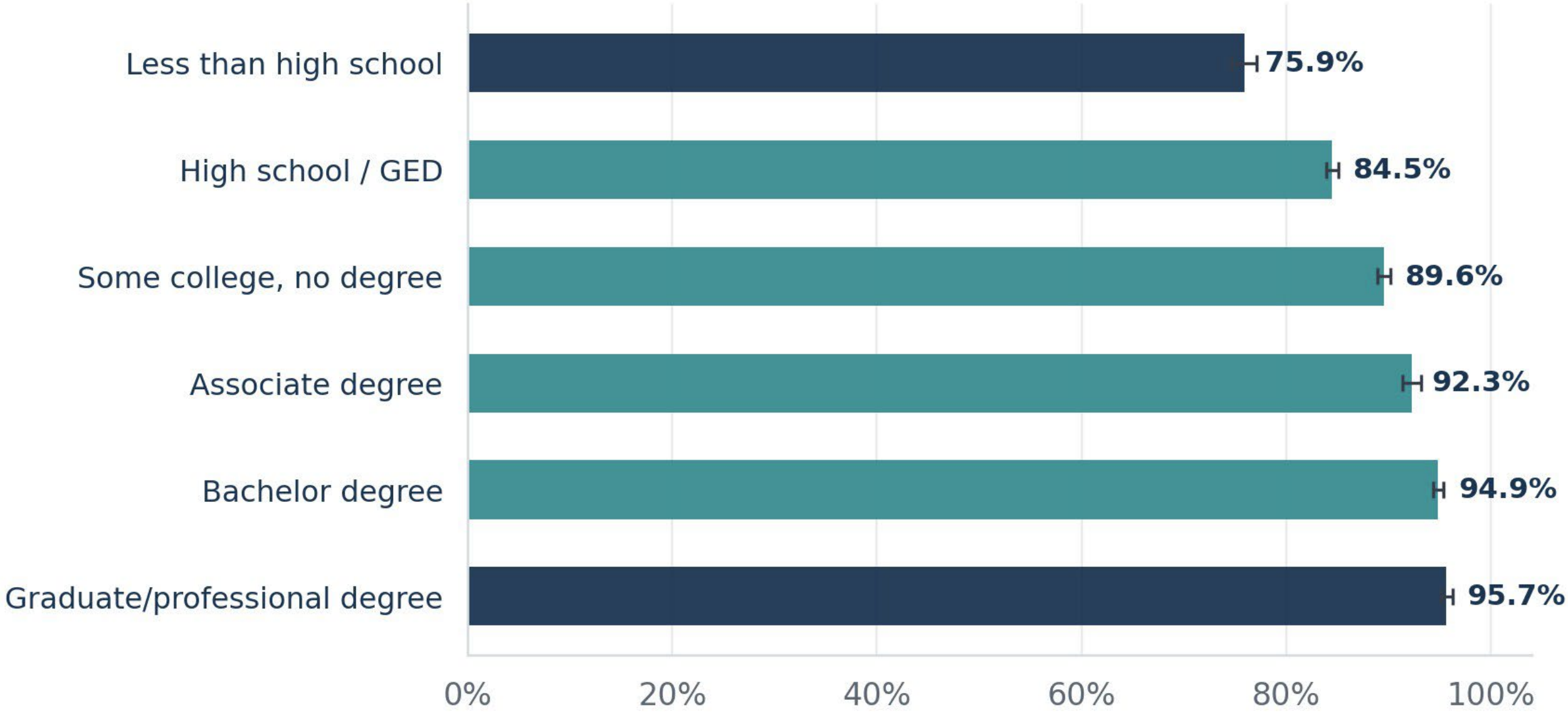
Education-related differences are greater for cable/fiber/DSL service than for any broadband subscription. Cellular-only plans, in contrast, are more common among adults in lower education categories. The findings are consistent with an education-related digital divide in subscription type as well as overall subscription status, but do not establish causation.

RESULTS AT A GLANCE

Unadjusted gap	Income-adjusted gap	Cable/fiber/DSL gap
+19.8 percentage points	+14.3 percentage points	+22.6 percentage points
Graduate vs. less than HS	Graduate vs. less than HS	Graduate vs. less than HS

All gaps compare graduate/professional degrees with less than high school. Confidence intervals use ACS 90% conventions. Differences represent associations in household subscription status, not effects of educational attainment.


Name
Type
audit
File folder
code
File folder
data
File folder
report
File folder
results
File folder
sources
File folder
README.txt
Name
Type
Compressed size
analyze_nm_education_broadband_mod...
Python Source File
4 KB
audit_nm_broadband_step2.py
Python Source File
3 KB
build_nm_broadband.py
Python Source File
3 KB
validate_nm_b28006.py
Python Source File
3 KB
validate_nm_household_broadband.py
Python Source File
1 KB


Less than high school
75.9%
High school / GED
84.5%
Some college, no degree
89.6%
Associate degree
92.3%
Bachelor degree
94.9%
Graduate/professional degree
95.7%
0%
20%
40%
60%
80%
100%

Education	Sample n	Weighted adults	Broadband	90% MOE (pp)
Less than high school	8,704	200,159	75.9%	±1.19
High school / GED	19,867	440,586	84.5%	±0.59
Some college, no degree	16,754	370,592	89.6%	±0.60
Associate degree	6,268	140,973	92.3%	±0.88
Bachelor degree	12,219	260,014	94.9%	±0.50
Graduate/professional degree	10,558	204,971	95.7%	±0.56

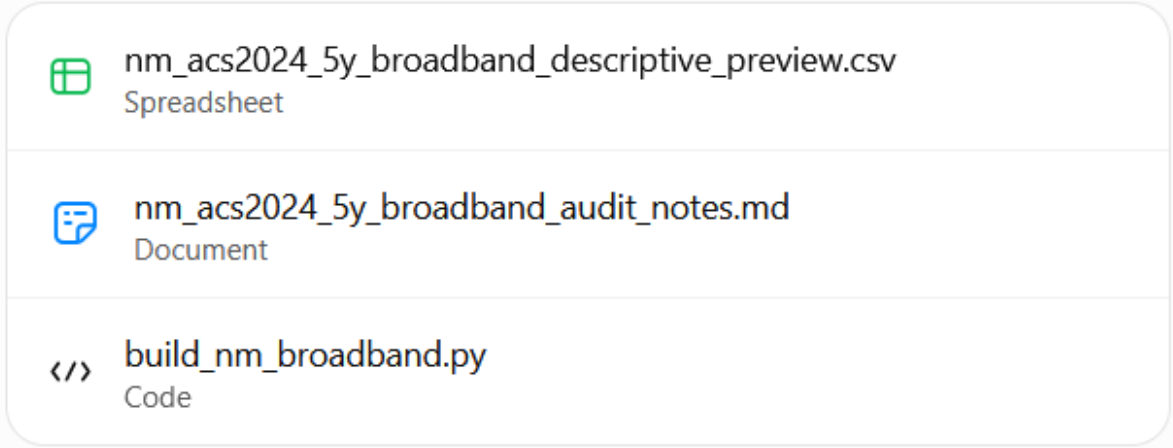
Table 1. Descriptive education-specific subscription rates, New Mexico adults aged 18+. MOE = margin of error; percentage points (pp).

statistic	jw	jw/ai	dg/ai
less than HS/GED with broadband pct	63	63	75.9
HS diploma or GED with broadband pct	76.3	76.3	84.5
some college/associate's with bbnd pct	87.4	87.4	89.6
bachelor's degree with broadband pct	92.8	92.8	94.9
grad degree with broadband pct	94.3	94.3	95.7

I recommend proceeding with adults **18+** as the primary population and **any paid broadband subscription, including cellular-only plans**, as the primary outcome. We would retain ages 25+ and wired broadband as secondary comparisons.

Do you approve those two definitions?

After that decision, the next steps are to validate the PUMS estimates against published Census Table B28006, examine the relationship in more detail, and review the variables proposed for regression before fitting models.



agree with those definitions

statistic	jw	jw/ai	DG/AI 1	DG/AI Revise
less than HS/GED with broad	63	63	75.9	67.4
HS diploma or GED with broa	76.3	76.3	84.5	79.6
some college/associate's with	87.4	87.4	89.6	88.2
bachelor's degree with broac	92.8	92.8	94.9	93.8
grad degree with broadband	94.3	94.3	95.7	95.2

Compare all three workflows

DIMENSION	HUMAN-LED	HYBRID	AI-DIRECTED
Researcher role	Analyst	Collaborator + supervisor	Evaluator + accountable owner
AI role	Information assistant	Task-focused contributor	Primary executor
Typical value	Discovery efficiency	Scale + consistency	Speed and breadth
Primary risk	Unverified suggestions	Hidden process choices	Opaque chains of decisions
Human oversight	Continuous	Planned AI insertion points	Human QA/QC checkpoints

The three takeaways

01	**Mediate outputs**	Verify definitions, data, code and claims.
02	**Match effort to complexity**	AI may speed up coding while adding validation work.
03	**Augment deliberately**	Choose assistance or autonomy based on task and consequence.

A reusable researcher checklist

Before

Question, data authority, permitted use, unit of analysis

During

Prompt constraints, provenance, code review, spot-checks

After

Uncertainty, reproducibility, interpretation, disclosure

Thank you!

Questions, conversation, and networking

David Gustavsen · Jon Wheeler

College of University Libraries & Learning Sciences
University of New Mexico