

From Spreadsheets to Scripts: Modernizing IR/IE Workflows with R

NCAIR 2026

Logan King Haywood Community College

A little about Haywood CC



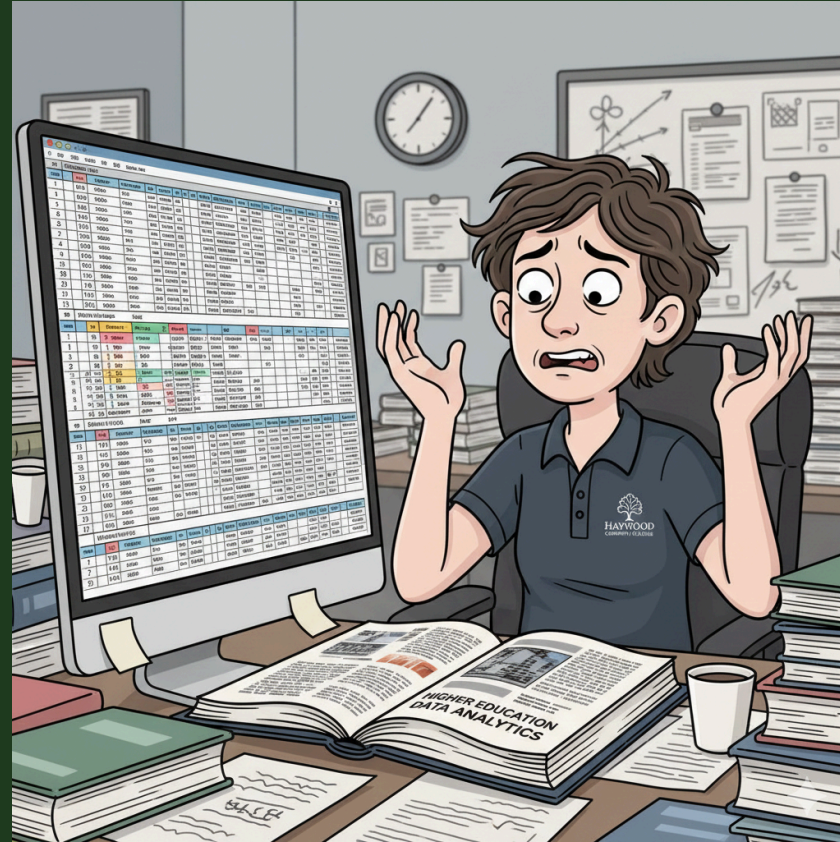
- ~1300 Students in Curriculum
- ~2700 Students in Continuing Education (25-26)
- 30+ Academic Programs
- Over 50% of Courses offered online

Mission: We serve the educational needs and economic growth of our community, by promoting lifelong student learning and success.

- 2027 Aspen Prize Nominee
- Unique programs in Natural Resources and Professional Crafts
- Have a Lumberjack Team!



Is This You?



Source: Google Gemini

Session Overview

What will you be able to do after today:

1. Evaluate the benefits and trade-offs of transitioning from Excel to R
2. Identify 2–3 IR/IE tasks currently done in Excel that can be migrated to R
3. Formulate next steps for beginning your own transition

A quick note on what this session is NOT:

```

1 age = int(input("what is your age? "))
2 time = int(input("what is the time of the movie? "))
3
4 #Store in variable whether or not the user is above the age of 13
5 if(age > 13):
6     adult = "true"
7 else:
8     adult = "false"
9
10 #If an adult, follow this line
11 if(adult == "true"):
12     if(time >= 2000):
13         #Post 10 PM, $8 Ticket + $4 Matinee
14         notAllowed = "false"
15         price = 12
16     elif(time <= 1600):
17         #Before 6 PM, $8 Ticket + $5 Matinee
18         notAllowed = "false"
19         price = 13
20     else:
21         #Standard ticket price, $8
22         notAllowed = "false"
23         price = 8
24
25 #If not an adult, follow this line
26 if(adult == "false"):
27     if(time >= 2000):
28         #If the user is underage and trying to join Midnight Madness, set
29         #notAllowed variable to true
30         notAllowed = "true"
31     else:
32         notAllowed = "false"
33         price = 6
34
35 print()
36
37 #If the user is not allowed in Midnight Madness, print message.
38 #Else, print ticket price
39 if(notAllowed == "true"):
40     print("Midnight Madness is not available for children under 14 years of age!")
41 else:
42     print("Your ticket price is $", price)
43

```



Source: Visit Haywood

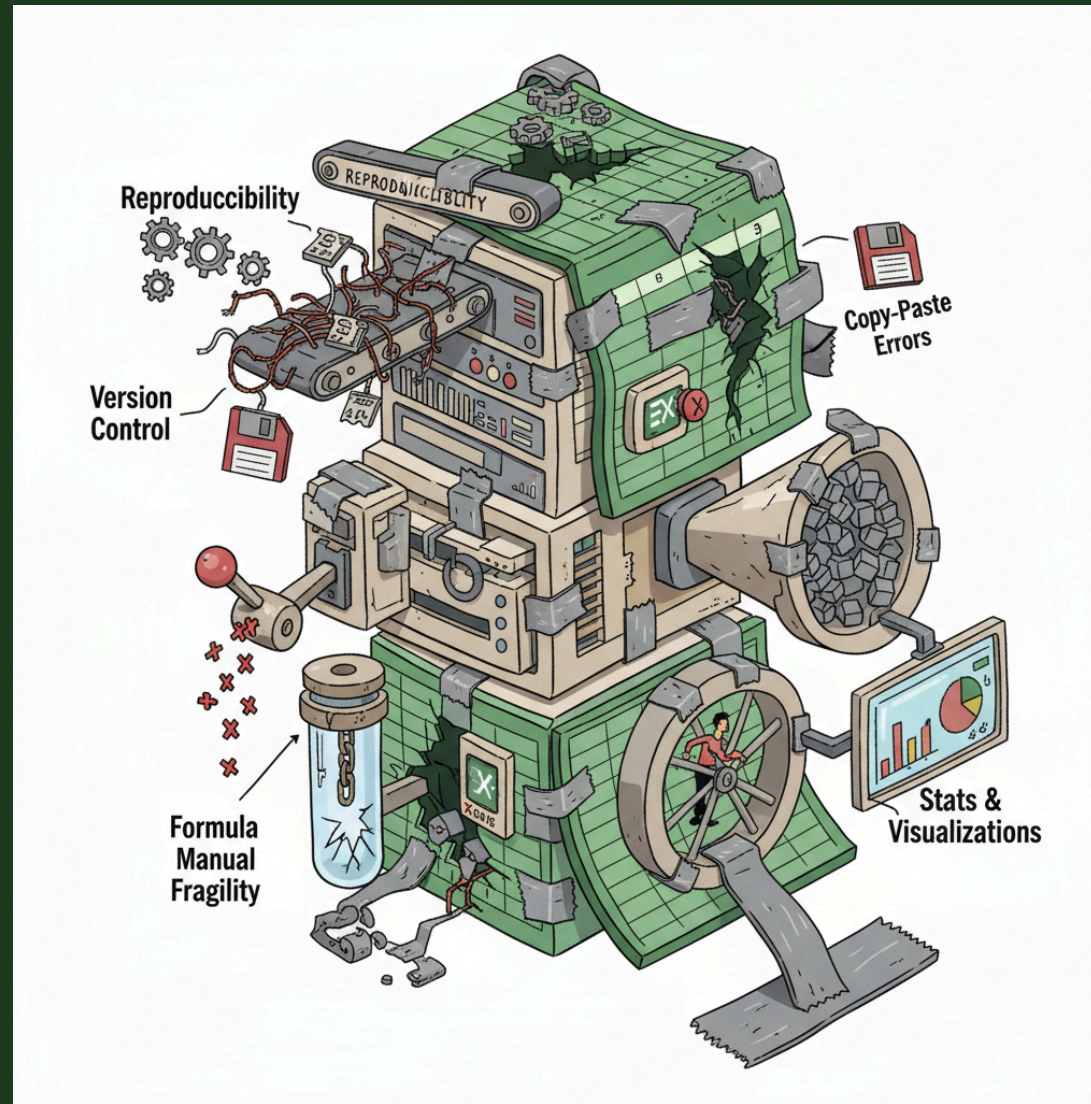
Source: Reddit

The Reality of IR/IE Workflows in Excel

Most IR/IE offices live in Excel, and for good reason. It's accessible, flexible, and everyone knows it.

- **Reproducibility:** Can you recreate last year's report from scratch? Can someone else?
- **Version control:** Which file is the real one? What changed between versions?
- **Copy-paste errors:** Manual data movement introduces mistakes that are hard to catch.
- **Formula fragility:** One deleted row can silently break a calculation downstream.
- **Scalability limits:** Large datasets slow Excel to a crawl or crash it entirely.
- **Tedious manual updates:** Each new reporting cycle means repeating the same steps by hand.
- **Stats & Visualizations:** Excel's built-in stats tools are limited and not designed for complex analyses or visualizations.

We don't use Excel because it's the best tool for everything. We use it because it's the tool we have.



Source: Google Gemini



From Spreadsheets to Scripts

Transitioning to R can address each of these pain points directly.

Excel Challenge	R Solution
Reproducibility	Scripts document every step. Run it again and again, get the same report.
Version control	Works seamlessly with Git and Github; track every change.
Copy-paste errors	Data flows through code, not manual transfers.
Formula fragility	Logic is explicit, testable, and auditable.
Scalability	Handles millions of rows without breaking a sweat.
Tedious manual updates	Automate the repetitive stuff. Spend time on analysis.
Stats & Visualizations	Powerful libraries for complex analyses and stunning visuals.

The core toolkit:

Programming
Language



Integrated
Development
Environment (IDE)



Reporting &
Visualization



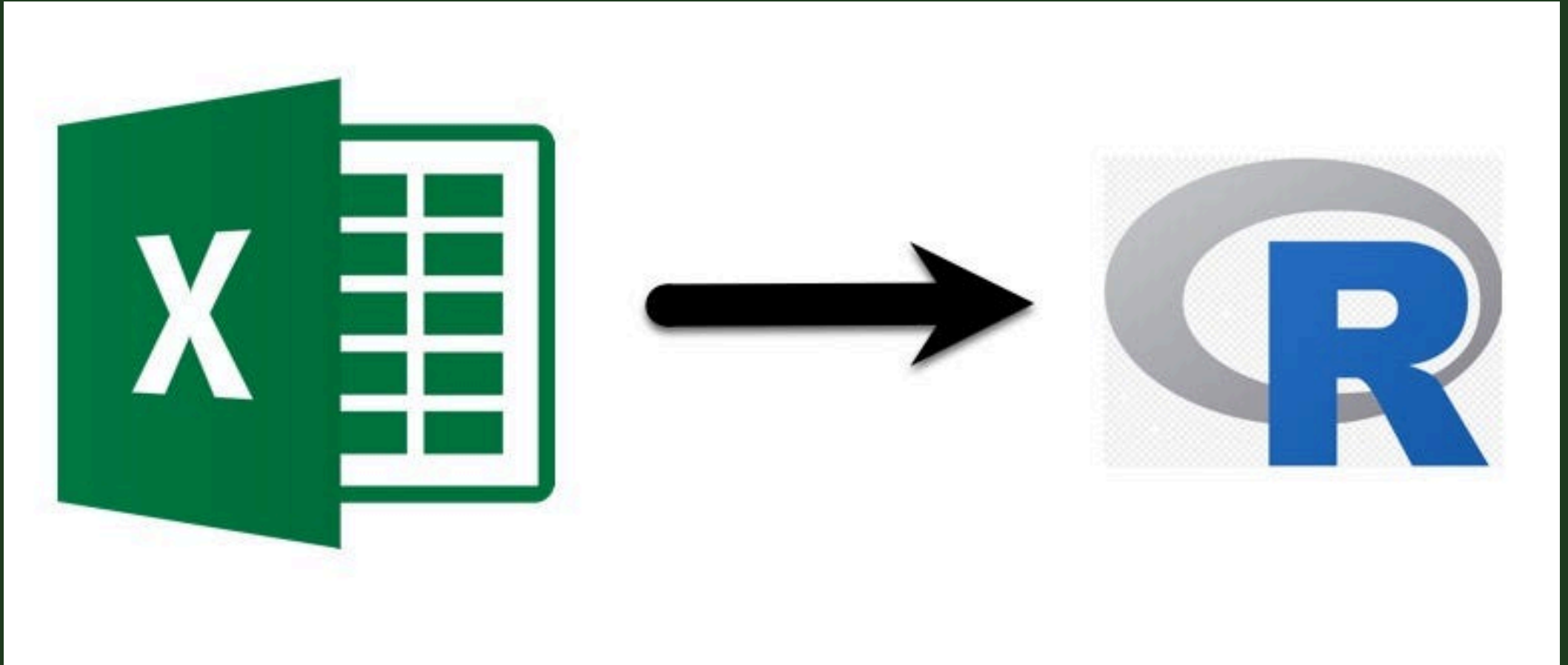
Interactive Apps and
Dashboards



Source: Posit & R-Project



Live Examples: Excel → R



Source: Daran Johnson Blog

Example 1: IPEDS Comparisons

The Excel workflow:

- Pull peer data manually from IPEDS Data Center
- Manually manipulate data to build 1 table
- Build pivot tables for each metric by hand
- Recreate charts every reporting cycle from scratch
- Static output: one Excel file nobody else can easily update

IPEDS_Peers_2024_FINAL_v2_USE_THIS.xlsx

The result: Hours of work to answer one question:

“How do we compare?”

```
1 library(readxl); library(dplyr); library(gt)
2
3 df <- read_excel("IPEDS_NC_CC_Enrollment_Fake.xlsx", sheet = "I
4
5 df |>
6   mutate(
7     pct_ft      = ENRFT / ENRTOT,
8     pct_women   = UGENRW / UGENRTOT,
9     pct_black   = UGENR_BLACK / UGENRTOT,
10    pct_deg_seek = DEG_SEEKING / UGENRTOT
11  ) |>
12  select(INSTNM, ENRTOT, pct_ft, pct_women, pct_black, pct_deg_
13  arrange(desc(ENRTOT)) |>
14  gt() |>
15  fmt_integer(columns = ENRTOT) |>
16  fmt_percent(columns = c(pct_ft, pct_women, pct_black, pct_deg
```

One script. Swap the file next year. Done.

Example 1: Potential Output

IPEDS Peer Enrollment Comparison — NC Community Colleges

Fall 2023 · Source: IPEDS EF Survey · Haywood CC highlighted

Institution	Total Enroll.	% Full-Time	% Women	% Black	% Deg-Seeking
Wake Tech CC	24,180	38.2%	54.8%	28.3%	72.4%
Central Piedmont CC	18,450	41.1%	53.4%	31.2%	69.8%
Guilford Tech CC	12,870	35.5%	51.9%	29.8%	71.2%
Forsyth Tech CC	9,340	37.8%	52.7%	27.1%	73.1%
Johnston CC	5,620	42.1%	56.1%	18.4%	68.3%
Haywood CC	1,310	31.2%	57.2%	4.2%	76.1%
Isothermal CC	1,080	29.8%	54.3%	3.8%	74.4%
Mayland CC	890	27.5%	53.1%	2.9%	71.8%

Generated by R · readxl + dplyr + gt · Haywood CC Office of Institutional Research

Example 2: SLO Assessment

The Excel workflow:

- Collect rubric scores across faculty via emailed spreadsheets
- Manually consolidate into a master workbook
- Calculate means by hand or with fragile cell references
- Re-paste into Word or PowerPoint for the assessment report
- Repeat every semester with a new set of files

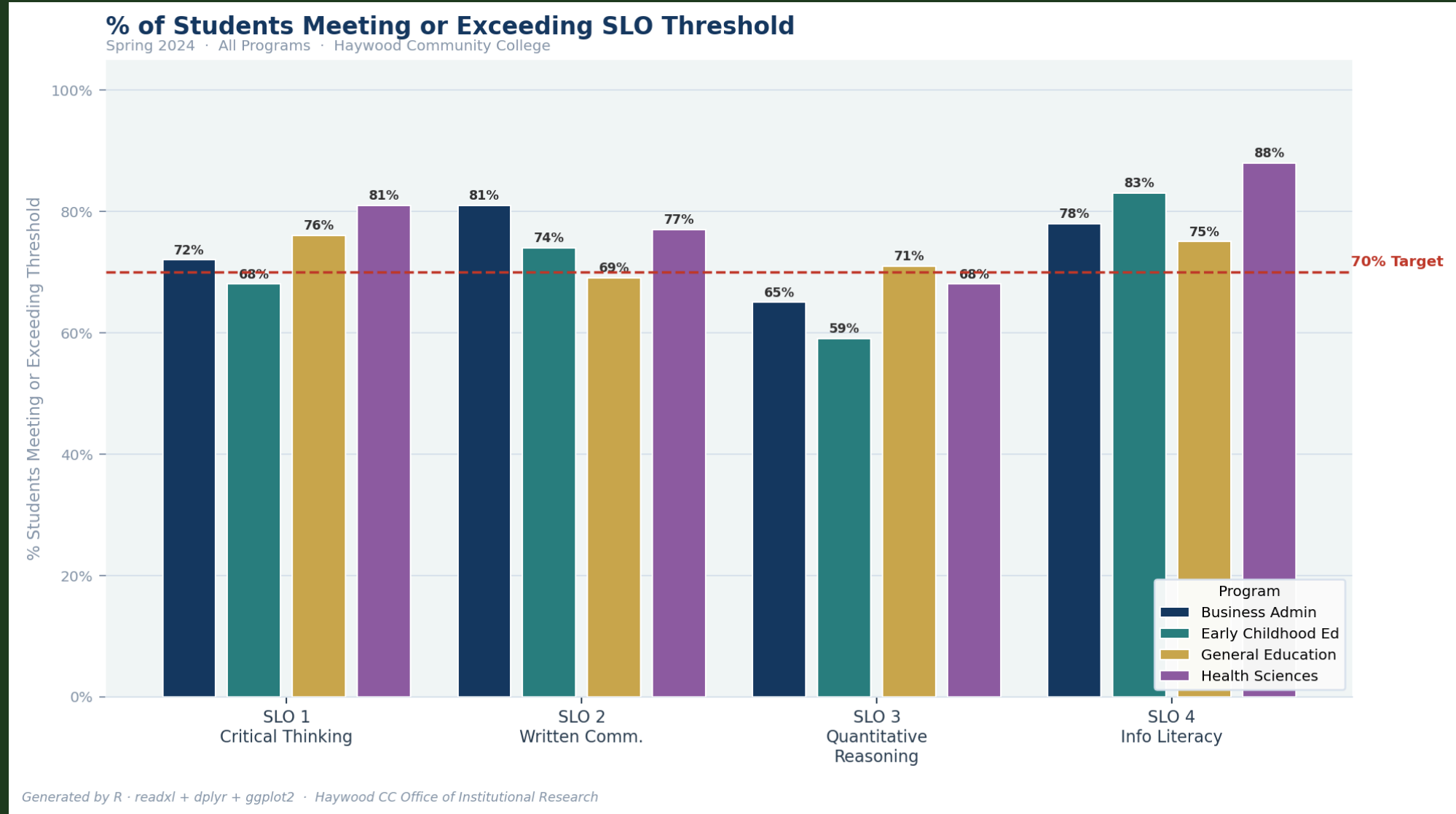
SLO_Spring2024_CONSOLIDATED_v3_Logans_version_FINAL.xlsx

The result: Hours lost to data wrangling before any actual analysis begins.

```
1 library(readxl); library(dplyr); library(ggplot2)
2
3 df <- read_excel("SLO_Rubric_Scores.xlsx")
4
5 df |>
6   group_by(program, outcome) |>
7   summarise(
8     n          = n(),
9     mean_score = mean(score, na.rm = TRUE),
10    pct_meets  = mean(score >= 3, na.rm = TRUE)
11  ) |>
12  ggplot(aes(x = outcome, y = pct_meets, fill = program)) +
13  geom_col(position = "dodge") +
14  scale_y_continuous(labels = scales::percent) +
15  labs(
16    title = "% of Students Meeting or Exceeding SLO Threshold",
17    x = "Student Learning Outcome",
18    y = NULL, fill = "Program"
19  ) +
20  theme_minimal()
```

Add a new semester's file. Rerun. Updated report, no copy-paste required.

Example 2: Potential Output



Example 3: Statistical Analysis

The Excel workflow:

- Use Data Analysis ToolPak, if it's even installed
- Results paste as static, unformatted output into the sheet
- Re-run manually every time data changes
- No easy path to regression, SEM, or forests
- Visualizations require separate power chart stepss

retention_analysis_2024_stats_attempt2.xlsx

The result: Analysis limited by what Excel's toolbox can do, not by your research questions.

```

1 library(readxl); library(tidymodels); library(probably)
2
3 df <- read_excel("Student_Retention_Fake.xlsx") |>
4   mutate(retained = factor(RETAINED, levels = c("No", "Yes")))
5
6 # Define model spec and recipe
7 log_spec <- logistic_reg() |> set_engine("glm")
8
9 rec <- recipe(retained ~ FIRST_GEN + PELL_ELIG + ENROLL_INTENSI
10              data = df) |>
11   step_dummy(all_nominal_predictors())
12
13 # Fit via workflow
14 wf <- workflow() |>
15   add_recipe(rec) |>
16   add_model(log_spec) |>
17   fit(data = df)
18
19 # Tidy coefficients with odds ratios
20 wf |>
21   tidy(exponentiate = TRUE, conf.int = TRUE) |>
22   filter(term != "(Intercept)") |>
23   select(term, estimate, conf.low, conf.high, p.value)

```

Logistic regression, in eleven lines. Swap the outcome variable and rerun for any question.



Benefits *and* Tradeoffs

It's not a magic wand, but it is a better tool for a lot of what we do.

Benefits

-  **Reproducibility:** Scripts are documentation; anyone can rerun your work
-  **Automation:** Repetitive tasks run in seconds, not hours
-  **Scalability:** Millions of rows, no sweat
-  **Version control:** Git tracks every change, with notes
-  **Professional outputs:** Quarto produces polished reports and presentations
-  **Free and open source:** No licensing costs

Tradeoffs & Challenges

-  **Learning curve:** R takes time to learn, but there are plenty of resources
-  **Migration investment:** Rebuilding existing workflows takes real effort up front
-  **Organizational adoption:** IT, leadership, and colleagues all need buy-in
-  **Not everything should migrate:** A quick one-off calculation? Excel is fine
-  **Support structures:** You may be the only R user in your office (for now)

First Steps for Your Own Transition

How do you know what to migrate? Look for tasks that are repetitive, data-heavy, error-prone, or frequently updated.

How to approach the migration:

1. Start with *one* task, not your most complex report
2. Rebuild it in R while keeping the Excel version as a check
3. Validate your R output against the known Excel result
4. Once confident, retire the Excel version

How to build organizational support:

- Show your work: share a polished Quarto report with your supervisor
- Involve IT early: especially around data access and file storage
- Find an internal champion: one curious colleague can make a difference
- Be patient with yourself and others: adoption takes time

Resources, Tools, and Community

Tools

- [R Project](#)
- [RStudio](#)
- [Positron](#)
- [Quarto](#)
- [Shiny](#)

Learning

- [R for Data Science](#) — Free online book; the best place to start
- [Quarto Documentation](#)

Community

- [Posit Communities](#)
- [ThIRsdays](#) — IR-focused R tutorials by David Eubanks, Furman University

AI & Coding Assistants

- GitHub Copilot, Claude, and ChatGPT can help you learn R syntax, debug errors, and understand unfamiliar code
- Use them as a tutor, not a shortcut

Q&A



Thank You!

Connect with me!



Logan King



rking3387@haywood.edu

