

Teaching Pharmacy Students Coding Through REDCap and Automated Data Tracking

Speakers:

David Chan, PharmD, PhD, BCOP, CCRP

Coordinator, Investigational Drug Service
Clinical Assistant Professor, Retzky College of Pharmacy
University of Illinois at Chicago

Yuri Song

PharmD Candidate (2028)
UIC Retzky College of Pharmacy



2026 ICHP Annual Meeting

Speaker Conflicts

- The speakers have no conflicts of interest with ineligible companies to disclose



Learning Objectives

- Evaluate problem-based learning as a method for teaching data science
- Demonstrate the use of REDCap and other data source to support code learning
- Discuss the R language in the context of data science workflow
- Identify an example of a data science project amendable to automation
- Examine data science workflow: data acquisition, data wrangling, data visualization and communication



2026 ICHP Annual Meeting



"Artwork by Allison Horst". Licensed for open use [CC-BY-4.0](https://creativecommons.org/licenses/by/4.0/). <https://allisonhorst.com/allison-horst>

Coding for Pharmacy Students

- No or little background in computer programming
- Special Projects in Pharmacy Practice
 - Independent study
- Introduce basic data science skills
 - Coding, Data import, data wrangling (cleaning, preparation), and communication
- Provides a framework for student-centered learning
 - Focus on problem solving
 - Self-directed, self-motivated, active learning
- Reproducible, reusable and shareable R code



2026 ICHP Annual Meeting

Problem-based Learning Design

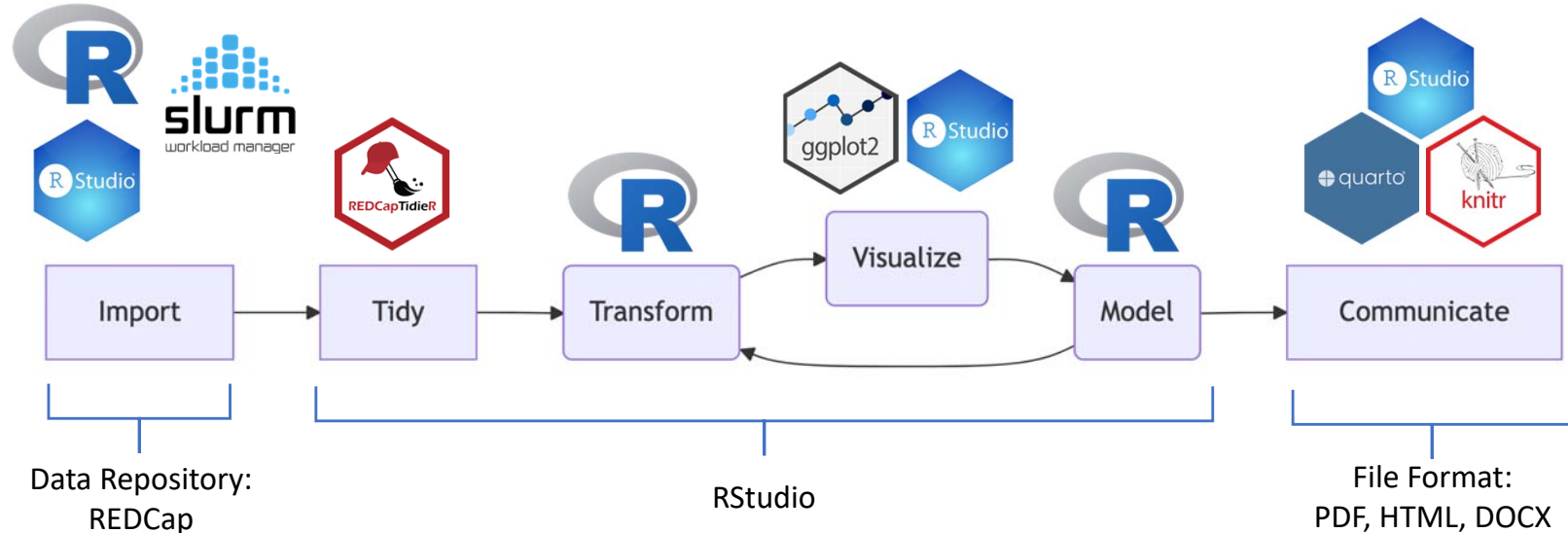
- Identify an existing REDCap project for basic data science knowledge acquisition
 - Complexity scores
 - Temperature logs
- Guided discovery and mentoring
 - Starter R script, REDCap videos: <https://projectredcap.org/resources/videos/>
 - Reading list
 - Technical feedbacks
- Reproducible code to assess reproducibility, quality and credibility of an analysis



2026 ICHP Annual Meeting

Data Science Workflow

- Iterative cycle of transforming, visualizing and modeling



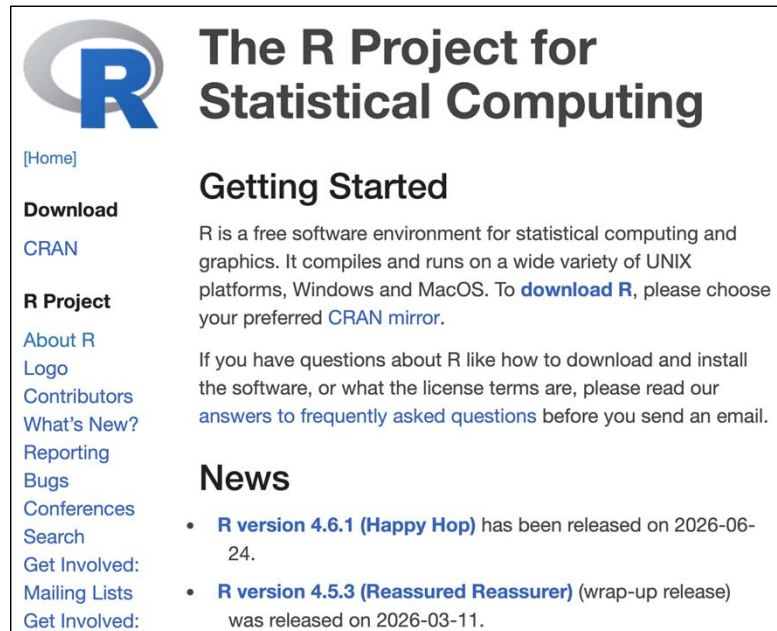
<https://r4ds.hadley.nz/>



2026 ICHP Annual Meeting

Getting Started

- Obtain R from **Comprehensive R Archive Network (CRAN)**
 - <https://cran.r-project.org/>
- Install Base R
- Open-source software



The R Project for Statistical Computing

[Home]

Download

[CRAN](#)

R Project

[About R](#)
[Logo](#)
[Contributors](#)
[What's New?](#)
[Reporting](#)
[Bugs](#)
[Conferences](#)
[Search](#)
[Get Involved:](#)
[Mailing Lists](#)
[Get Involved:](#)

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 4.6.1 (Happy Hop)** has been released on 2026-06-24.
- **R version 4.5.3 (Reassured Reassurer)** (wrap-up release) was released on 2026-03-11.

<https://www.r-project.org>

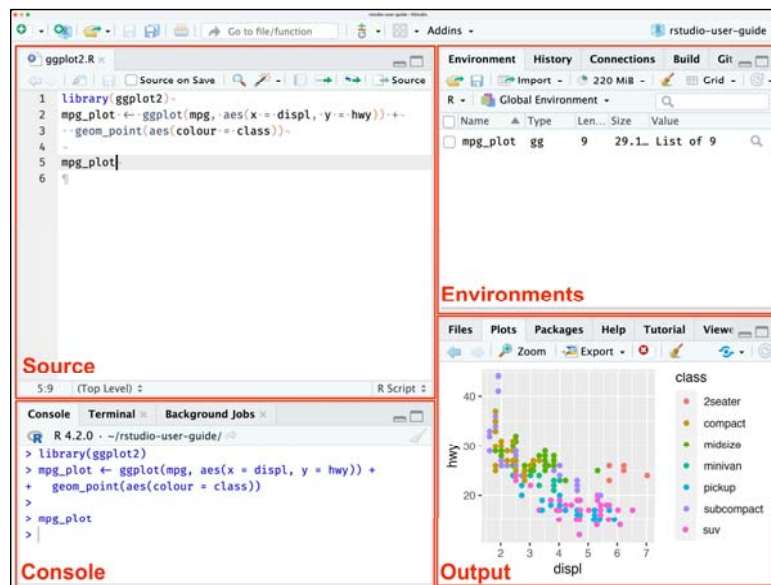


2026 ICHP Annual Meeting

Getting Started

- Obtain and install an Integrated Development Environment (IDE)
- Visual Studio (VS) Code®
 - <https://code.visualstudio.com>
- Positron®
 - <https://positron.posit.co>
- RStudio® Desktop
 - <https://posit.co/products/open-source/rstudio>

RStudio Primary Panes



<https://docs.posit.co/ide/user/ide/get-started/>



2026 ICHP Annual Meeting

Base R

- Getting help
- Using packages
- Assigning variables
- Reading and writing data

```
1 # Get help of a function
2 help("name")
3
4 # Download and install a package, REDCapTidieR from
  CRAN
5 install.packages("REDCapTidieR")
6 # Load the package REDCapTidieR
7 library(REDCapTidieR)
8
9 # Load a build-in dataset
10 data(Indometh)
11
12 # Assign variable
13 x <- "name"
14 # Data frame
15 df <- data.frame(x = 1:3, y = c('a', 'b', 'c'))
16
17 # Read and write data
18 df <- read.csv('filename.csv')
19 write.csv(df, 'filename.csv')
```



2026 ICHP Annual Meeting

Practice Datasets from Base R/R Packages

- R built-in datasets
 - 89 datasets

Indometh	Pharmacokinetics of Indomethacin
InsectSprays	Effectiveness of Insect Sprays
JohnsonJohnson	Quarterly Earnings per Johnson & Joh
LakeHuron	Level of Lake Huron 1875–1972
LifeCycleSavings	Intercountry Life–Cycle Savings Data
Loblolly	Growth of Loblolly Pine Trees
Nile	Flow of the River Nile
Orange	Growth of Orange Trees
OrchardSprays	Potency of Orchard Sprays
PlantGrowth	Results from an Experiment on Plant
Puromycin	Reaction Velocity of an Enzymatic Re
Theoph	Pharmacokinetics of Theophylline

- Rdatasets archive
 - 3648 datasets

	Package	Item	Title	CSV	Doc	Rows	Cols
1035	datasets	Indometh	Pharmacokinetics of Indomethacin	CSV	Doc	66	3
1087	datasets	Theoph	Pharmacokinetics of Theophylline	CSV	Doc	132	5
2117	medicaldata	indometh	Cohort Study of the Pharmacokinetics of Intravenous Indomethacin	CSV	Doc	66	3
2126	medicaldata	theoph	Cohort Study of the Pharmacokinetics of Oral Theophylline	CSV	Doc	132	5
2307	nlme	Cefamandole	Pharmacokinetics of Cefamandole	CSV	Doc	84	3
2336	nlme	Remifentanyl	Pharmacokinetics of Remifentanyl	CSV	Doc	2107	12
2339	nlme	Tetracycline1	Pharmacokinetics of tetracycline	CSV	Doc	40	4
2340	nlme	Tetracycline2	Pharmacokinetics of tetracycline	CSV	Doc	40	4

<https://vincentarelbundock.github.io/Rdatasets/index.html>



2026 ICHP Annual Meeting

Practice Datasets

- UC Irvine Machine Learning Repository
 - 689 datasets
 - 113 datasets in health and medicine
- Python import available
 - 57 in health and medicine
- Comma Separate Values (CSV)

Filters ✕ CLEAR FILTERS

Search datasets...

Browse Datasets 🔍 SORT BY # VIEWS, DESC 🔍 EXPAND ALL

Heart Disease
4 databases: Cleveland, Hungary, Switzerland, and the VA Long Beach
🔍 Classification 📊 Multivariate 📄 303 Instances 📄 13 Features

Subject Area: Health and Medicine Feature Type: Categorical, Integer, Real Date Donated: 6/30/1988 Views: 1.17M

Breast Cancer Wisconsin (Diagnostic)
Diagnostic Wisconsin Breast Cancer Database.
🔍 Classification 📊 Multivariate 📄 569 Instances 📄 30 Features

Subject Area: Health and Medicine Feature Type: Real Date Donated: 10/31/1995 Views: 707.41K

Diabetes
This diabetes dataset is from AIM '94
🔍 Classification 📊 Multivariate, TI... 📄 1 Instances 📄 20 Features

Subject Area: Health and Medicine Feature Type: Categorical, Integer Date Donated: N/A Views: 251.71K

<https://archive.ics.uci.edu>

<https://archive-beta.ics.uci.edu> (beta site)



2026 ICHP Annual Meeting

Vizient® Complexity Scoring Tool

- For study initiation and maintenance
 - Categories: low, medium, or high complexity
 - Derived from 322 responses
- Objectively measure the complexity of clinical trials
- Spreadsheet-based data collection

GLOBAL FACTORS		INITIATION	MAINTENANCE
☐	EARLY PHASE STUDIES (Phase I or I/II) (1 point)		
☐	MULTIPLE PROTOCOLS: Nested or umbrella protocol design. No limit. (1 point each) Enter number of sub-protocols ▶	2	
☐	COMPLEX IP MGMT: more than 1 chemical entity (including placebo if maintained as separate inventory). No limit. (1 point each) Number of chemical entities greater than 1 ▶	0	
☐	COMPLEX IP MGMT: more than 1 strength or presentation for an individual IP. No limit. (0.5 point each) Enter number of IP with multiple strengths or presentations ▶	1	
☐	COMPLEX IP MGMT: Individually numbered and tracked inventory units (2 points)		
☐	COMPLEX IP MGMT: Retention or return shipping of used vials, drug packaging, or patient returns (no immediate disposal on site) (1 point)		
☐	COMPLEX IP MGMT: BSL2 handling (1 point)		
☐	COMPLEX IP MGMT: Liquid nitrogen or less than minus160 degrees Celsius storage (1 point)		
☐	COMPLEX IP MGMT: Ultracold freezer (-80 degrees Celsius) (1 point)		
☐	COMPLEX DISPENSING: Pharmacy adjusted dosing required		
☐	COMPLEX DISPENSING: NON-sterile compounding required at time of dispensing (IP not in ready to dispense form)		

AJHP 2023;80:1557-1563 <https://doi.org/10.1093/ajhp/zxad138>



2026 ICHP Annual Meeting

Practice Dataset – REDCap Complexity Scores

- All UI Health IDS active studies
- Data access via API JSON and CSV file formats

Instrument: **Vizient Complexity Scoring Tool** (vizient_complexity_scoring_tool) [Collapse](#)

[cst_irb_num]	Section Header: STUDY INFORMATION UIC IRB Number	text Field Annotation: @CALCTEXT([irb_num])
[cst_prot_num]	Sponsor Protocol Number	text Field Annotation: @CALCTEXT([prot_num])
[cst_study_title]	Study Title	text Field Annotation: @CALCTEXT([study_title])
[ct_identifier]	Dose this study has a ClinicalTrials.gov Identifier? <i>NCT Number</i>	yesno 1 Yes 0 No
[cst_nct_num]	NCT Number <i>8-digit Clinical Trial Identifier Number</i> Show the field ONLY if: [ct_identifier]=1	text Field Annotation: @CHARLIMIT="8"
[study_phase]	Section Header: GLOBAL FACTOR: STUDY DESIGN Early Phase Studies (Phase I or I/II) <i>1 point</i>	yesno 1 Yes 0 No

GLOBAL FACTOR: STUDY DESIGN

Early Phase Studies
(Phase I or I/II) ☐ Yes ☐ No *1 point* [reset](#)

Multiple Protocols: Nested or umbrella protocol design ☐ Yes ☐ No [reset](#)

Study Design Subscores

Initiation	0 View equation
Maintenance	0 View equation

GLOBAL FACTOR: COMPLEX INVESTIGATIONAL PRODUCT MANAGEMENT

More than 1 chemical entity
(Including placebo if maintained as separate inventory) ☐ Yes ☐ No [reset](#)

More than 1 strength or presentation for an individual investigational product ☐ Yes ☐ No [reset](#)

API: Application Programming Interface
IDS: Investigational Drug Service
JSON: JavaScript Object Notation



2026 ICHP Annual Meeting

API: Application Programming Interface



- Enable communication between computers or computer programs

The screenshot shows a web interface with a row of buttons for different programming languages: PHP, Perl, Python, Ruby, Java, R, C#, and cURL. The 'R' button is highlighted with a red box. Below the buttons, a text box contains the following R script:

```
#!/usr/bin/env Rscript
token <- "████████████████████████████████████████"
url <- "https://www.redcap.ihrp.uic.edu/api/"
formData <- list("token"=token,
  content='project',
  format='json',
  returnFormat='json'
)
response <- httr::POST(url, body = formData, encode = "form")
result <- httr::content(response)
print(result)
```

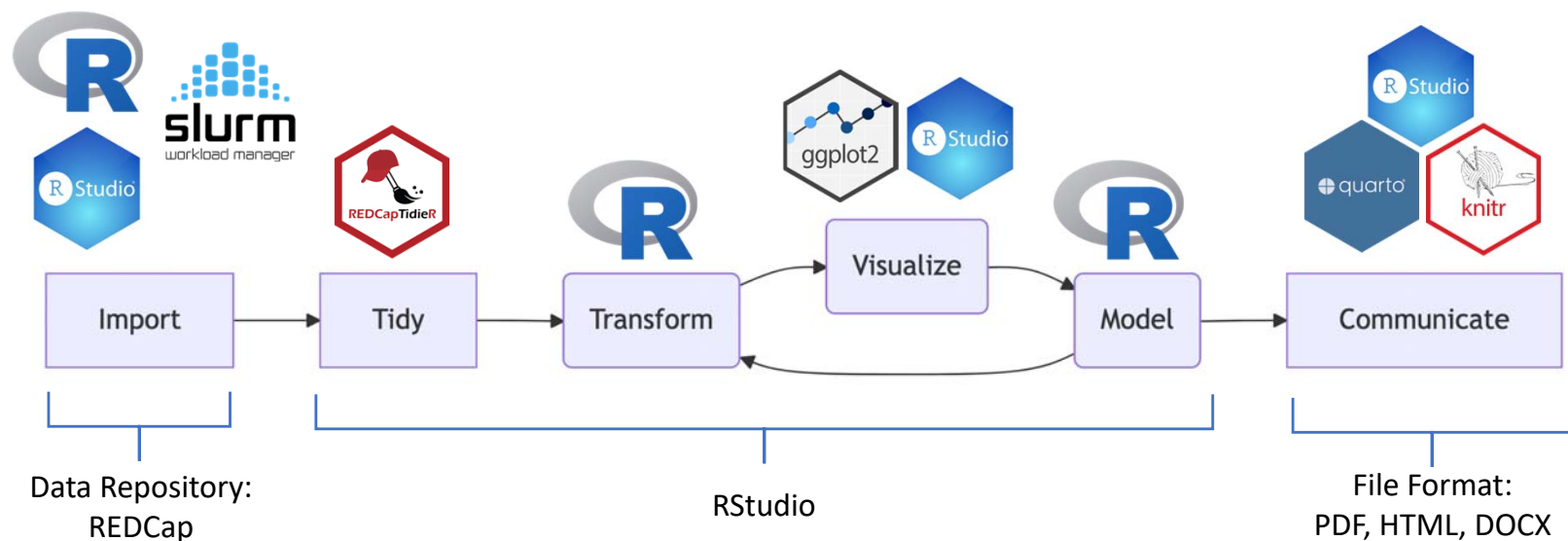
The line `response <- httr::POST(url, body = formData, encode = "form")` is highlighted with a red box.



2026 ICHP Annual Meeting

Data Science Workflow

- Iterative cycle of transforming, visualizing and modeling



<https://r4ds.hadley.nz/>



2026 ICHP Annual Meeting

Data Import, Tidy and Transform

- R package: REDCapTidieR

- Functions

- `read_redcap()`
- `bind_tibbles()`

- Arguments

- uniform resource locator (URL) of the REDCap server
- token
- forms (optional)

- Pipe `|>`

```
read_redcap(
  "https://www.redcap.ihrp.ui
c.edu/api/",
  Sys.getenv("token_15173"),
  forms =
  "general_study_information"
) |>
bind_tibbles()
```

- Query the REDCap API
- Retrieve, transform and bind data



2026 ICHP Annual Meeting

Assessment 1

Identify the components of problem-based learning as a method for teaching data science:

- A. Mentor guided
- B. Promote active learning
- C. Solving “real-world” problems
- D. All the above



2026 ICHP Annual Meeting

Assessment 2

Where can "real-world" datasets be obtained directly to support code learning?

- A. REDCap projects
- B. Base R and/or R Packages
- C. UC Irvine Machines Learning Repository
- D. All the above



2026 ICHP Annual Meeting

Assessment 3

Which software environment support(s) R language for data science projects?

- A. RStudio IDE
- B. VS Code
- C. Both A and B
- D. None of the above



Student Experience

Learning to code without prior programming background



2026 ICHP Annual Meeting

Starting point

- No prior coding experience
- First real research programming opportunity
- Goal: solve a real IDS workflow problem



2026 ICHP Annual Meeting

First Challenge – Understanding the Data

- Q: How can I identify active IDS studies?
- A: I need to understand how the data was organized in REDCap.

REDCap Form	Purpose
General Study Information	Study status (Active/Closed)
Complexity Scoring Tool	Initial and Maintenance complexity scores



2026 ICHP Annual Meeting

Second Challenge – Finding the Right Records

- I expected to analyze the Complexity Scoring dataset directly
- Problem
 - Complexity Scoring: 165 records
 - Active studies: 152 records
- Solution: Filter active studies using General Study Information



2026 ICHP Annual Meeting

Third Challenge – Combining Two Datasets

- Problem: Active studies and complexity scores lived in two separate tables, linked only by Record ID
- I needed to combine them automatically not by comparing IDs by eye
- Result: active study records could be linked to available complexity scores using Record ID

```
open_ids <- general_df |>
  filter(tolower(ids_status)
    == "open") |>
  distinct(record_id) |>
  drop_na(record_id)
merged_active <- open_ids
  |>
  inner_join(comp_scores,
    by = "record_id") |>
  drop_na(record_id)
```



Fourth Challenge – Results mismatched

- Expected: 152 active studies
 - Any score (initial or maintenance): ~120
 - Both scores: ~101
- Finding: Some studies simply hadn't reached the maintenance-scoring stage yet.



2026 ICHP Annual Meeting

Fifth Challenge – Handling zero scores

- Problem: **score of 0** broke downstream math and implied "zero complexity"

→ Fix: **recode 0 to 1**

if_else() = decision-making

Score = 0

→ Yes: Replace with 1

→ No: Keep original score

```
adj_init_score <-  
if_else(final_init_score ==  
0, 1, final_init_score)
```

```
adj_maint_score <-  
if_else(final_maint_score  
== 0, 1, final_maint_score)
```



2026 ICHP Annual Meeting

Sixth Challenge – Debugging my code

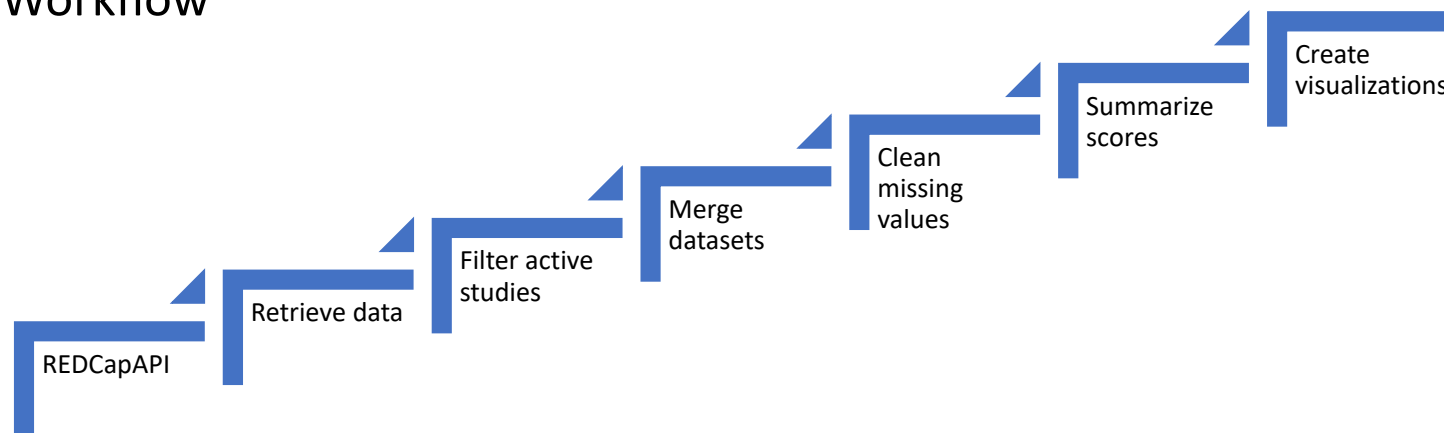
- Problem: `filter()` stopped working
- Cause: The REDCap data had not been imported because those lines were commented out
- Solution:
 - Reload the data
 - Clear the Environment
 - Re-run the script
- Lesson: Check the Environment panel first, before assuming the code itself is broken.



2026 ICHP Annual Meeting

Putting Everything Together

- After solving these challenges, I combined everything into one automated workflow.
- Workflow



2026 ICHP Annual Meeting

Result

- Automatically generated a reproducible complexity summary for active IDS studies
- Same report regenerates every week without manual work required



2026 ICHP Annual Meeting

Assessment 4

Which best describes a project suited for automation?

- A. One-time literature review
- B. Recurring report from a live database
- C. One-time analysis for a presentation
- D. Manual audit performed once



2026 ICHP Annual Meeting

What I learned

- This project taught me that data science isn't about knowing every function upfront.
- It's about learning to ask the right question, debug patiently, and verify before trusting your data.
- I learned to understand a database before writing any code.
- I learned to validate data instead of trusting it blindly.
- I learned to build a reproducible workflow that can be rerun automatically with updated data.



2026 ICHP Annual Meeting

Why It Matters

- This same pattern — pull, clean, verify, report — applies anywhere data needs to be tracked automatically over time
- Example use cases
 - IDS operations
 - Quality improvement
 - Medication-use evaluation
 - Clinical research
 - Automated reporting



2026 ICHP Annual Meeting

Assessment 5

Which step prepares raw data for analysis by cleaning and organizing records?

- A. Acquisition
- B. Wrangling
- C. Visualization
- D. Communication



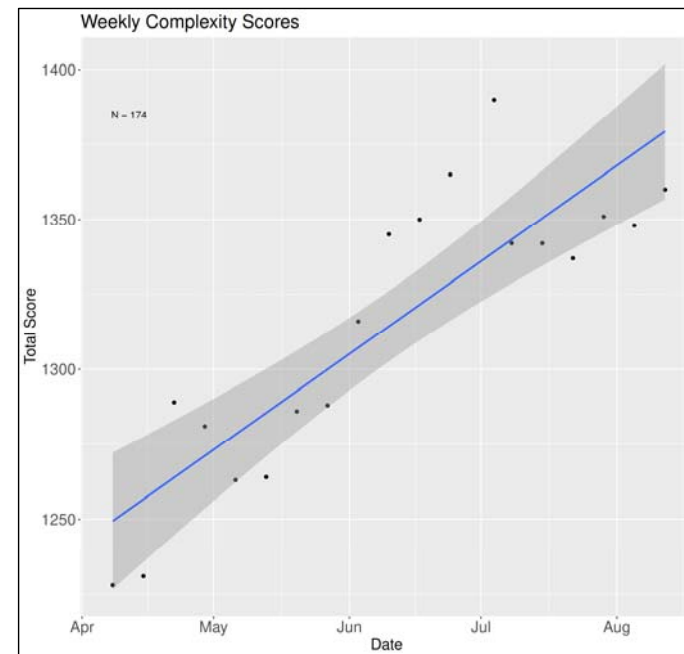
Weekly Complexity Scores

- Slurm Workload Manager
- Automate R script execution

```

1  #!/bin/bash
2
3  # =====
4  # SLURM JOB CONFIGURATION
5  # =====
6  # These #SBATCH lines tell SLURM what resources you need
7  # They MUST come before any other commands (except comments)
8
9  #SBATCH --job-name=cst_repeat           # Name shown in queue
10 #SBATCH --time=00:05:00                # Maximum runtime (HH:MM:SS)
11 #SBATCH --nodes=1
12 #SBATCH --ntasks=1
13 #SBATCH --partition=batch_gpu
14 #SBATCH --gpus-per-node=1
15 #SBATCH --account=pmpr_dchan_chi
16 #SBATCH --mem=4G
17 #SBATCH --output=logs/job_%j.out
18 #SBATCH --error=logs/job_%j.err
19
20 # =====
21 # WEEKLY COMPLEXITY SCORES
22 # =====
23 # Run every X days (good for weekly)
24 # You can use ANY number of days
25 # ↓ CHANGE TO YOUR DESIRED NUMBER OF DAYS
26 INTERVAL="7 days"
27 # Common examples: "7 days" (weekly), "14 days" (bi-weekly), "30 days" (monthly)
28 # ;;;

```



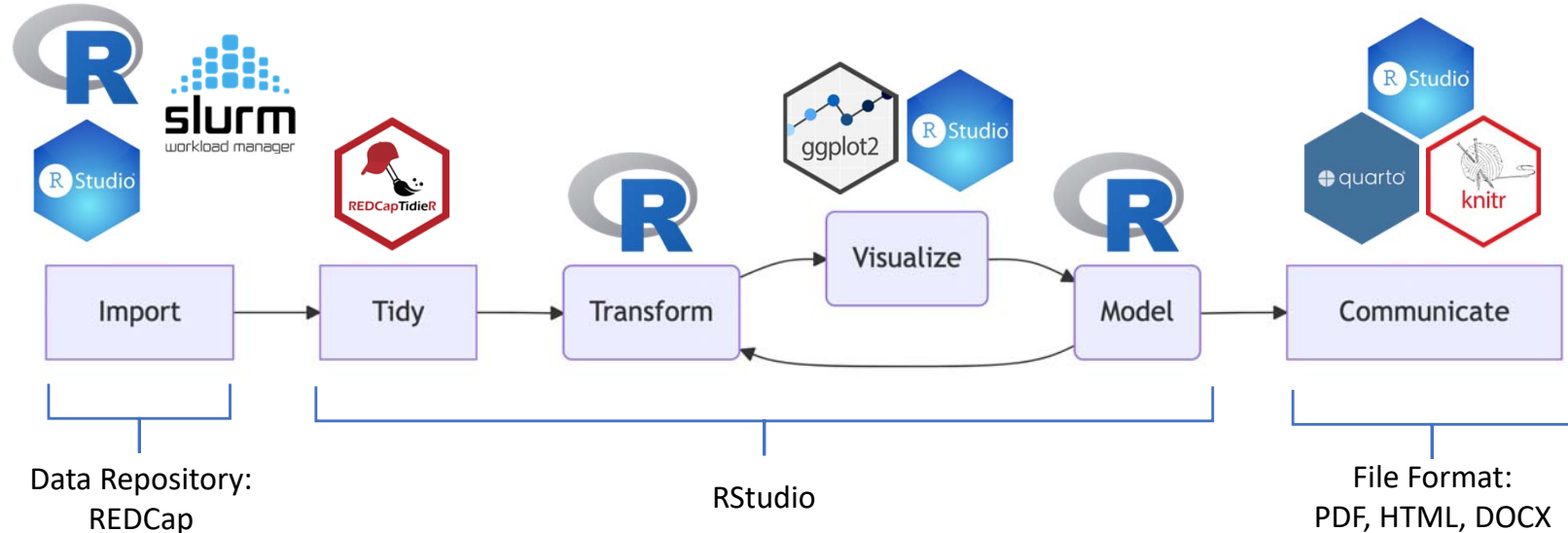
SLURM: Simple Linux Utility for Resource Management



2026 ICHP Annual Meeting

Data Science Workflow

- Iterative cycle of transforming, visualizing and modeling



2026 ICHP Annual Meeting

Communicate using Quarto



An open source scientific and technical publishing system



2026 ICHP Annual Meeting

<https://quarto.org/docs/get-started/hello/rstudio.html>

Communicate using Quarto

Source

```

1 ---
2 title: "Complexity Scores"
3 format: html
4 editor: source
5 ---
6
7 ## Quarto
8
9 Quarto enables you to weave together content and executable code into a
10 finished document (@fig-cs).
11
12 ```{r CS Ad hoc, echo=FALSE, message=FALSE}
13 #| label: "fig-cs"
14
15 library(tidyverse)
16 library(REDCapTidieR)
17
18 # Record Download -----
19 # Project 15173
20
21 redcap_url <- "https://www.redcap.ihrp.uic.edu/api/"
22
23 read_redcap(redcap_url, Sys.getenv("token_15173"), forms =
24   "general_study_information") |>
25   bind_tibbles()
26
27 read_redcap(redcap_url, Sys.getenv("token_15173"), forms =
28   "vizient_complexity_scoring_tool") |>
29   bind_tibbles()
30
31 general_df <- general_study_information
32 tool_df <- vizient_complexity_scoring_tool
33

```

YAML

Text

Code

Complexity Scores

Quarto

Quarto enables you to weave together content and executable code into a finished document (Figure 1).

oncology	final_init_score	final_maint_score	total_score
Mode: logical	Min.: 1.000	Min.: 1.000	Min.: 2.000
FALSE: 61	1st Qu.: 2.000	1st Qu.: 1.000	1st Qu.: 3.000
TRUE: 113	Median: 3.500	Median: 4.000	Median: 7.000
	Mean: 3.862	Mean: 3.954	Mean: 7.816
	3rd Qu.: 5.000	3rd Qu.: 6.000	3rd Qu.: 11.000
	Max.: 12.000	Max.: 13.000	Max.: 24.000

IDS Active Studies

Figure 1



2026 ICHP Annual Meeting

Summary

- R is relatively easy to learn, and R packages provide add-on functions that are sharable and reusable
- Learning R is an iterative process by solving challenges
- Coding with the R imparts essential computing skills that are necessary for working with data



2026 ICHP Annual Meeting

Questions?

Acknowledgements:

IDS Pharmacists

Rebeca Gansari, PharmD, BCPS

Brian Leav, PharmD

Students/Trainees

Kara Brown, PharmD

Laila Hammad, PharmD

Neera Kadiyala, PharmD

Jessica Mamalio, PharmD

Tom Szwajnos, PharmD



2026 ICHP Annual Meeting

References

1. Allaire J, Dervieux C (2025). R Interface to 'Quarto' Markdown Publishing System. R package version 1.5.1, <https://doi.org/10.32614/CRAN.package.quarto>.
2. Bryan J. STAT 545. <https://stat545.com/>.
3. Data Science in a Box. <https://datasciencebox.org/>.
4. Hanna R, Kadauke, S, Porter, E (2026). REDCapTidieR: Extract 'REDCap' Databases into Tidy 'Tibble's. R package version 1.2.5.9000, <https://chop-cgtinformatics.github.io/REDCapTidieR/>.
5. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010.
6. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208. doi:10.1016/j.jbi.2019.103208.
7. Higgins PDR (2025). Reproducible Medical Research with R. https://bookdown.org/pdr_higgins/rmrwr/.
8. Horst A. Artwork <https://allisonhorst.com/allison-horst>
9. Long J. (2019) R Cookbook 2nd edition. <https://rc2e.com>.
10. Martinez-Plumed F, Hernández-Orallo J. Training Data Scientists Trough Project-Based Learning. IEEE-RITA. 2023;18(3):295-304. doi: 10.1109/RITA.2023.3302954.
11. Posit team (2026). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. <https://posit.co/>.
12. R core Team (2025). R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>.
13. Song K, Yu M, McLuckie R, et al. Development of complexity categories for an investigational drug services complexity scoring tool to assess pharmacy effort in clinical trial initiation and maintenance. *Am J Health Syst Pharm.* 2023;80(21):1557-1563. doi:10.1093/ajhp/zxad138.
14. Wickham H, Çetinkaya-Rundel M, Grolemund, G (2023). R for Data Science 2nd edition. <https://r4ds.hadley.nz/>.
15. Xie Y (2025). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.50, <https://yihui.org/knitr/>.



2026 ICHP Annual Meeting

Teaching Pharmacy Students Coding Through REDCap and Automated Data Tracking

Speakers:

David Chan, PharmD, PhD, BCOP, CCRP

Coordinator, Investigational Drug Service
Clinical Assistant Professor, Retzky College of Pharmacy
University of Illinois at Chicago

Yuri Song

PharmD Candidate (2028)
UIC Retzky College of Pharmacy



2026 ICHP Annual Meeting