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

1

Speaker Conflicts

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

2

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

3

4

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

5

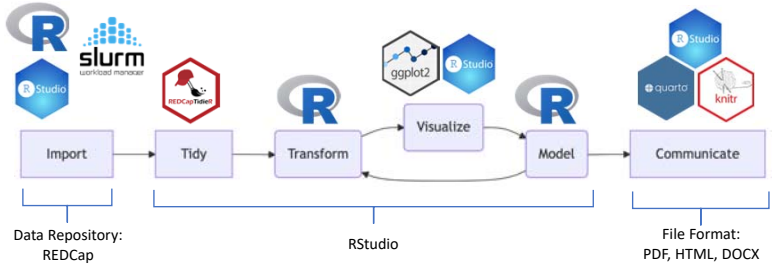
Problem-based Learning Design

- 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

6

Data Science Workflow

- Iterative cycle of transforming, visualizing and modeling

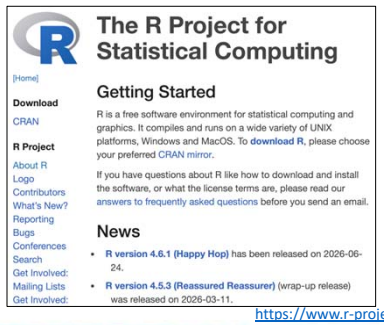


https://r4ds.hadley.nz/

7

Getting Started

- Obtain R from **C**omprehensive **R** **A**rchive **N**etwork (CRAN)
  - https://cran.r-project.org/
- Install Base R
- Open-source software



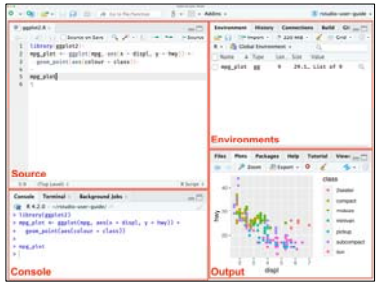
https://www.r-project.org

8

Getting Started

- Obtain and install an **I**ntegrated **D**evelopment **E**nvironment (IDE)
- **V**isual **S**tudio (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/

Base R

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

```r
# Get help of a function
help("name")

# Download and install a package, REDCapTidieR from CRAN
install.packages("REDCapTidieR")
# Load the package REDCapTidieR
library(REDCapTidieR)

# Load a build-in dataset
data(Indometh)

# Assign variable
x <- "name"
# Data frame
df <- data.frame(x = 1:3, y = c('a', 'b', 'c'))

# Read and write data
df <- read.csv('filename.csv')
write.csv(df, 'filename.csv')
```


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

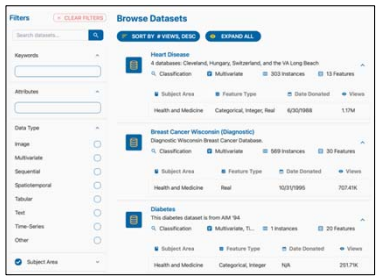
- Rdatasets archive
  - 3648 datasets

	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	Remifentanil	Pharmacokinetics of Remifentanil	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

Practice Datasets

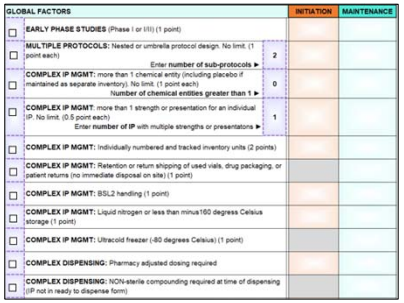
- UC Irvine Machine Learning Repository
  - 689 datasets
  - 113 datasets in health and medicine
- Python import available
  - 57 in health and medicine
- Comma Separate Values (CSV)



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

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

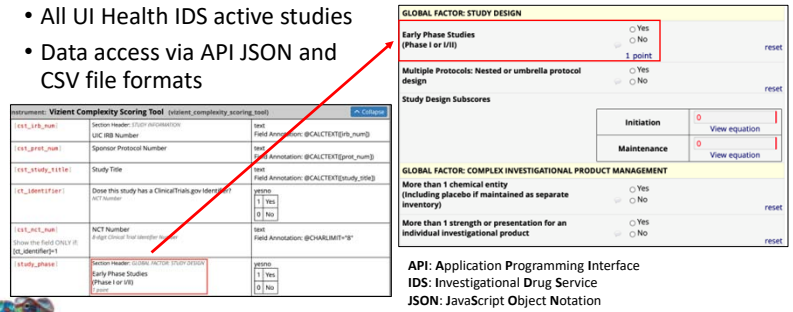


AJHP 2023;80:1557-1563 https://doi.org/10.1093/ajhp/zxad138

13

Practice Dataset – REDCap Complexity Scores

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



API: **A**pplication **P**rogramming **I**nterface
IDS: **I**nvestigational **D**rug **S**ervice
JSON: **J**ava**S**cript **O**bject **N**otation

14

API: Application Programming Interface

- Enable communication between computers or computer programs

PHP | Perl | Python | Ruby | Java | R | C# | cURL

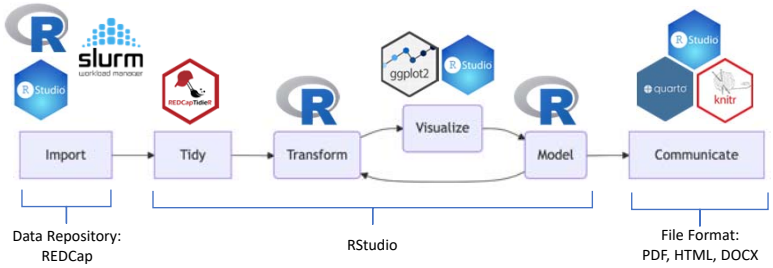
Displayed in the box below is the code you would use to execute this API request in the selected programming language.

```
#!/usr/bin/env Rscript
token <- "[redacted]"
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)
```


15

Data Science Workflow

- Iterative cycle of transforming, visualizing and modeling



https://r4ds.hadley.nz/

16

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.uic.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

17

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

18

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

19

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

2026 ICHP Annual Meeting

20

Student Experience

Learning to code without prior programming background

21

Starting point

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

22

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

23

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

24

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)
```


25

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.

26

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)
```


27

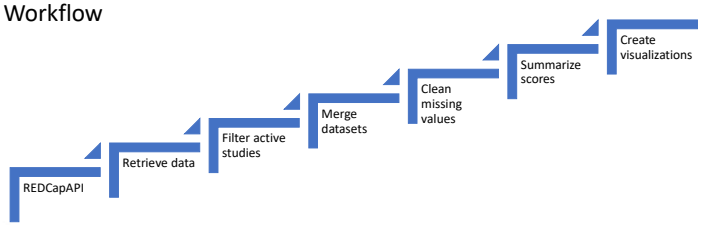
Sixth Challenge – Debugging my code

- 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.

28

Putting Everything Together

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



29

Result

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

30

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

31

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.

32

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

33

Assessment 5

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

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

34

Weekly Complexity Scores

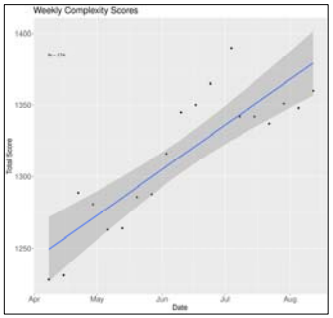
- Slurm Workload Manager
- Automate R script execution

```
#!/bin/bash

# ============================================
# SLURM JOB CONFIGURATION
# ============================================
# These #SBATCH lines tell SLURM what resources you need
# They MUST come before any other commands (except comments)

#SBATCH --job-name=cst_repeat          # Name shown in squeue
#SBATCH --time=00:05:00                # Maximum runtime (HH:
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --partition=batch_gpu
#SBATCH --gpus-per-node=1
#SBATCH --account=pmpr_dchan_chi
#SBATCH --mem=4G
#SBATCH --output=logs/job_%j.out
#SBATCH --error=logs/job_%j.err
```

```
"weekly")
# Run every X days (good for week
# You can use ANY number of days
# ↓ CHANGE TO YOUR DESIRED NUMBE
INTERVAL="7 days"
# Common examples: "7 days" (week
;;
```

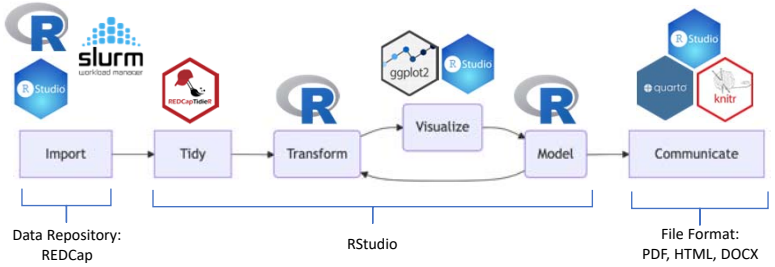


SLURM: **S**imple **L**inux **U**tility for **R**esource **M**anagement

35

Data Science Workflow

- Iterative cycle of transforming, visualizing and modeling



https://r4ds.hadley.nz/

36

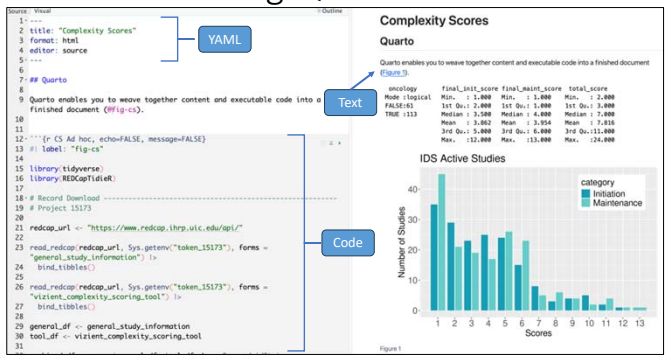
Communicate using Quarto

An open source scientific and technical publishing system

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

37

Communicate using Quarto



38

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

39

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

40

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

41

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

42