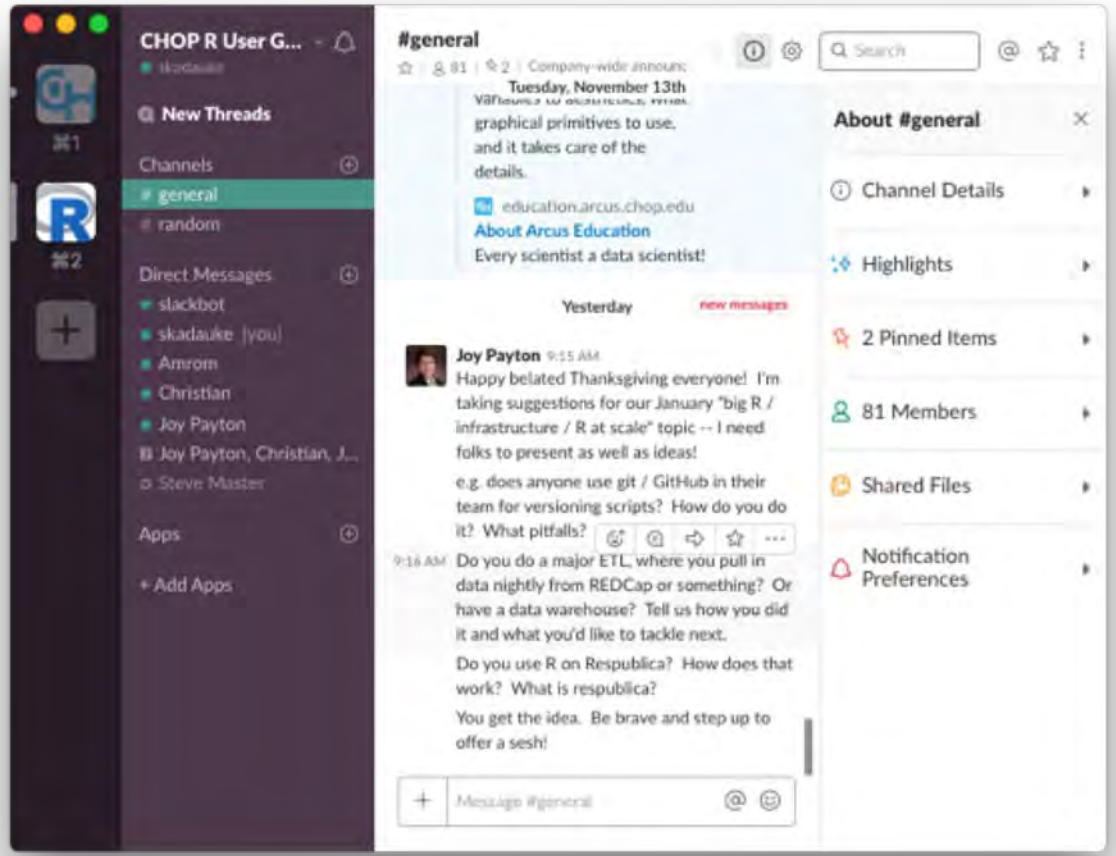
Mind the Gap:
Improving Collaboration Between Clinicians and Analysts

Stephan Kadauke
Christian Minich

Advancing Analytics for Children's Hospitals 2019
June 6, 2019


chopr

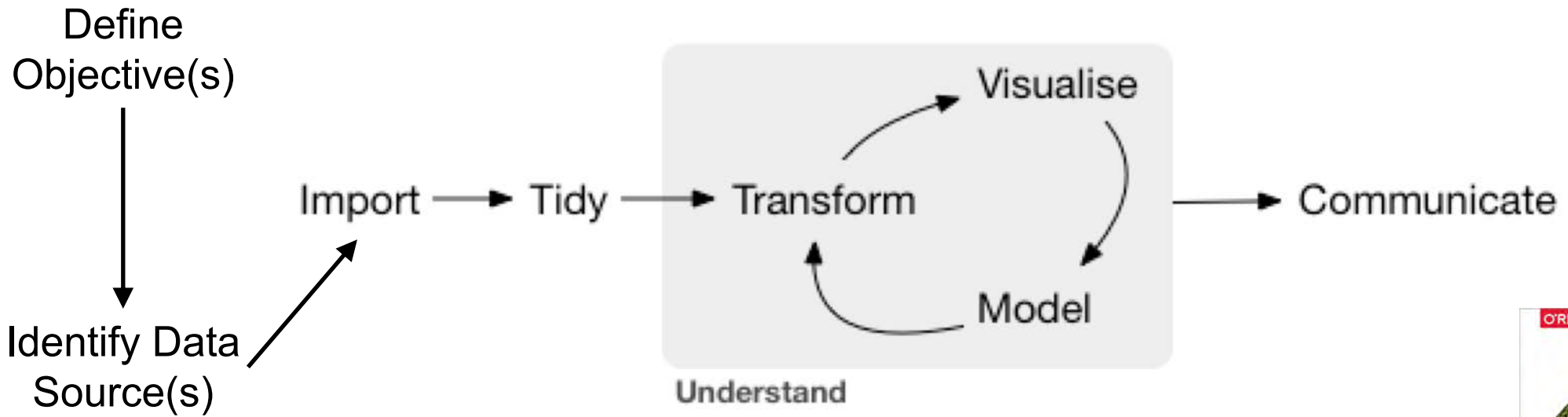

CHOP R User G...
New Threads
Channels
general
random
Direct Messages
slackbot
skadauke (you)
Amrom
Christian
Joy Payton
Joy Payton, Christian, J...
Steve Master
Apps
+ Add Apps
#general
Search
Tuesday, November 13th
graphical primitives to use,
and it takes care of the
details.
education.arcus.chop.edu
About Arcus Education
Every scientist a data scientist!
Yesterday
new messages
Joy Payton 9:15 AM
Happy belated Thanksgiving everyone! I'm taking suggestions for our January "big R / infrastructure / R at scale" topic -- I need folks to present as well as ideas!
e.g. does anyone use git / GitHub in their team for versioning scripts? How do you do it? What pitfalls?
9:16 AM Do you do a major ETL, where you pull in data nightly from REDCap or something? Or have a data warehouse? Tell us how you did it and what you'd like to tackle next.
Do you use R on Respublica? How does that work? What is respublica?
You get the idea. Be brave and step up to offer a sesh!
Message #general
About #general
Channel Details
Highlights
2 Pinned Items
81 Members
Shared Files
Notification Preferences

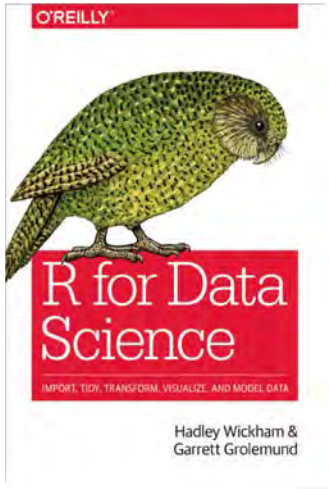
Objectives

- Propose a collaborative model for clinical data analytics projects
- Describe a course in Reproducible Clinical Data Analysis for clinicians aimed at improving clinician-analyst collaboration
- Outline a general approach for analysts (and other technical users) to become conversant in the clinical subject area domain in which they work

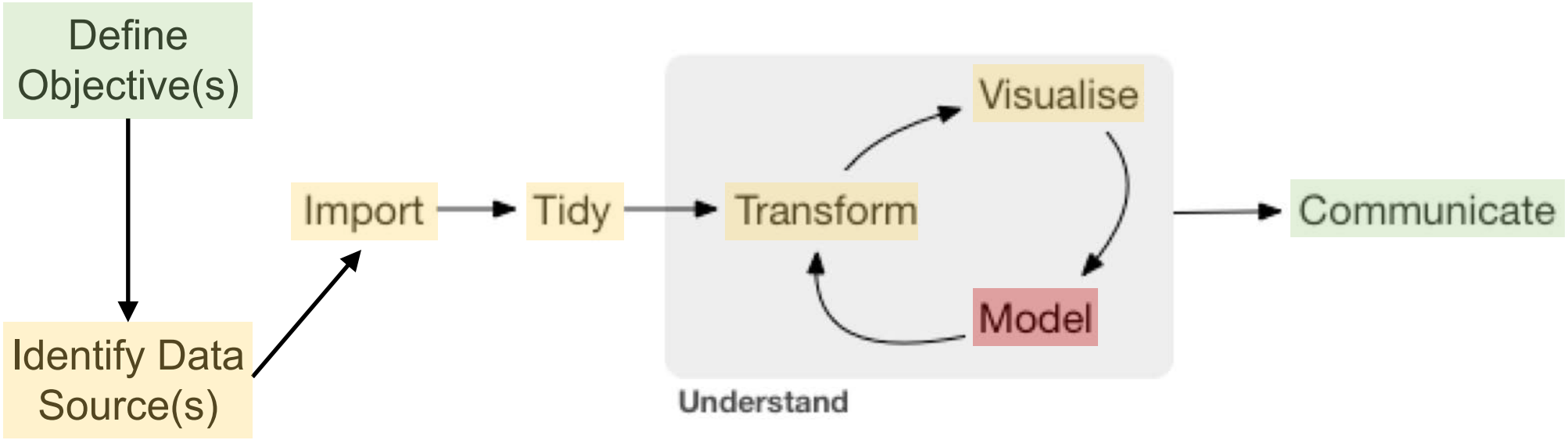
Analytics Maturation

Adapted from:
Mark Wu (Tableau)

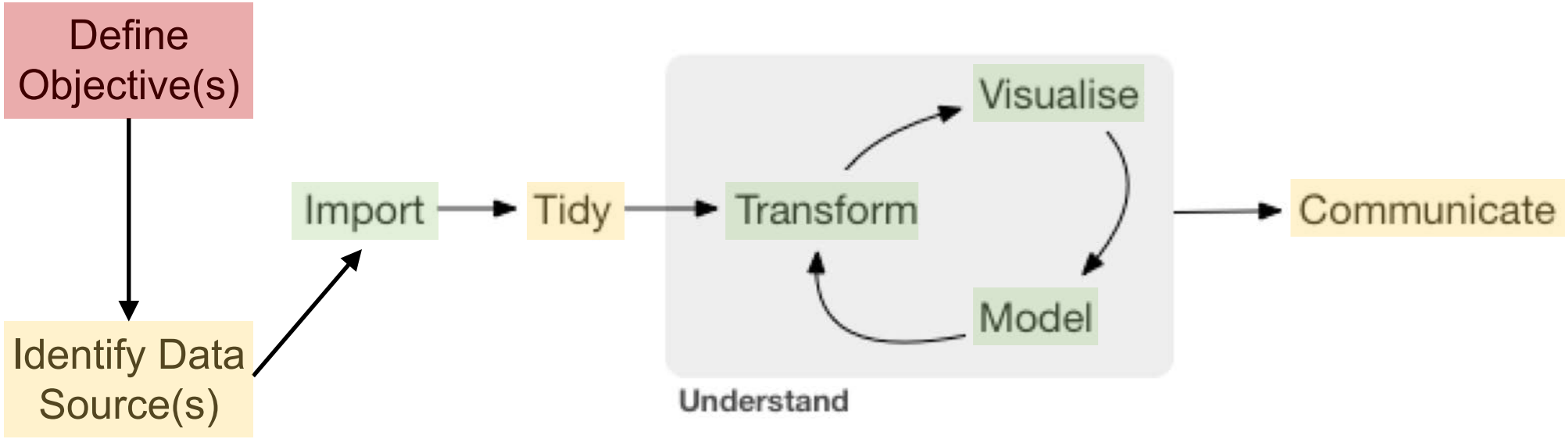

Define
Objective(s)
Identify Data
Source(s)
Import
Tidy
Transform
Visualise
Model
Understand
Communicate


O'REILLY
R for Data
Science
IMPORT, TIDY, TRANSFORM, VISUALIZE, AND MODEL DATA
Hadley Wickham &
Garrett Grolemund

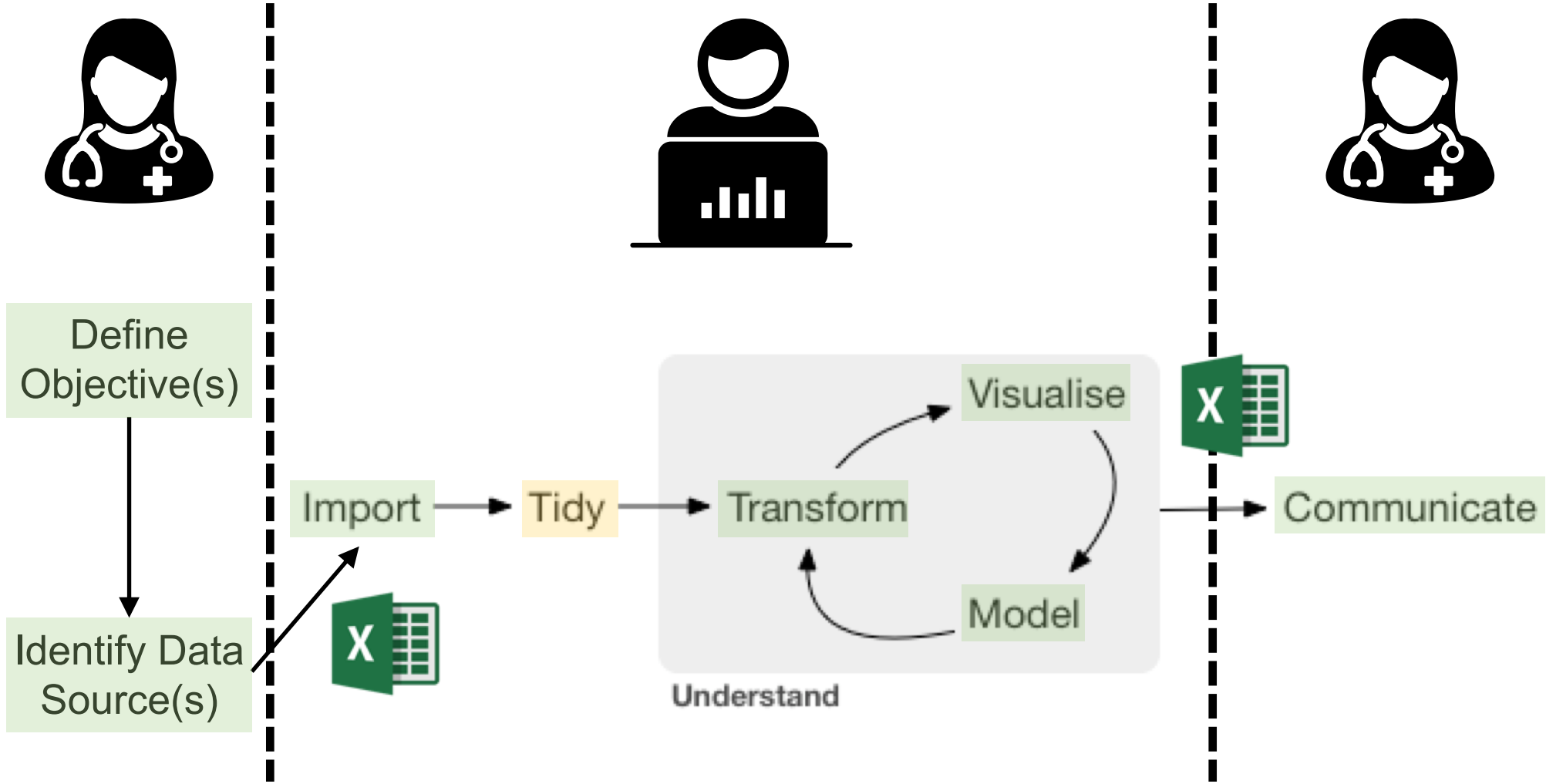
Purely Clinician-driven Analysis



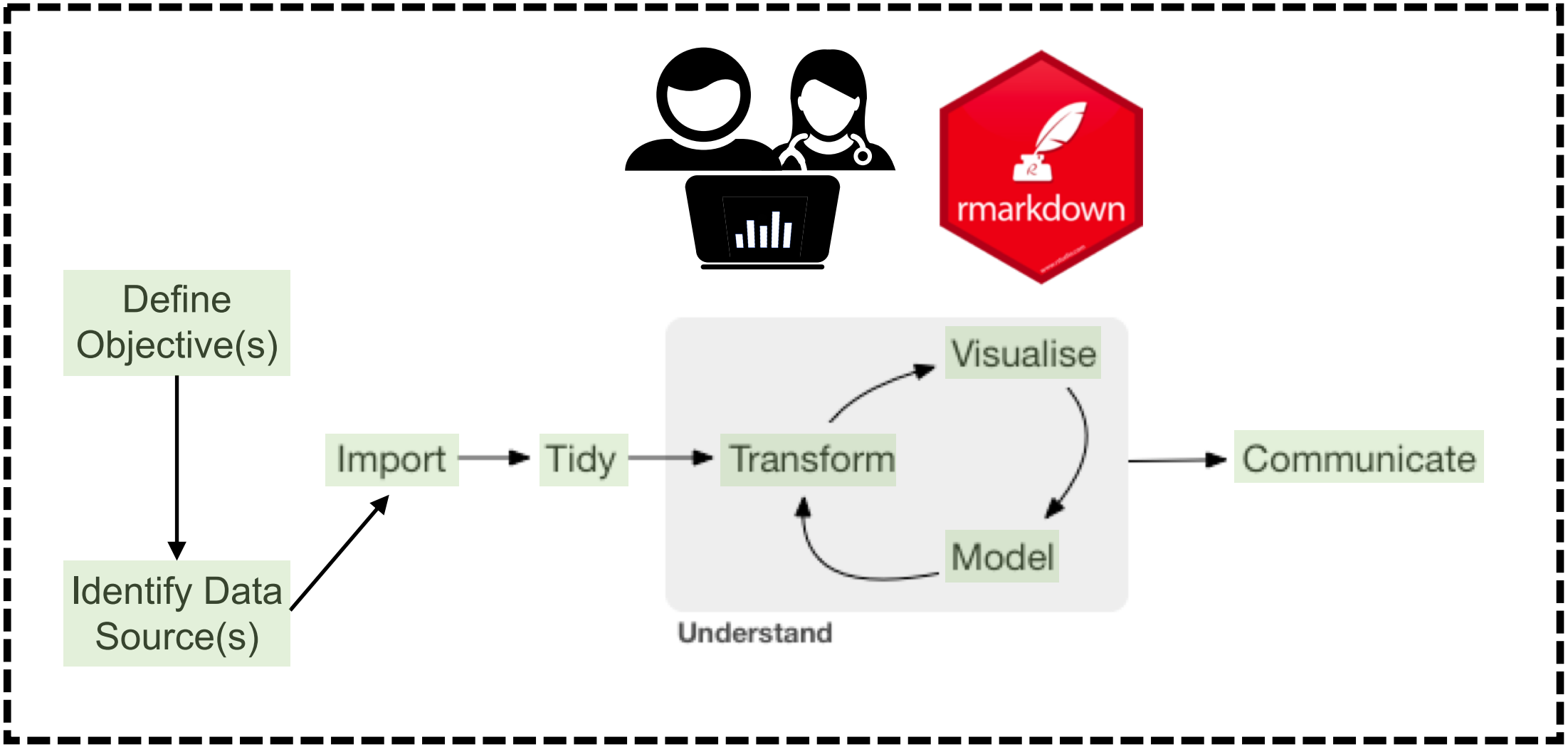
Purely Analyst-driven Analysis



Collaboration: Waterfall model



Collaboration: Agile model

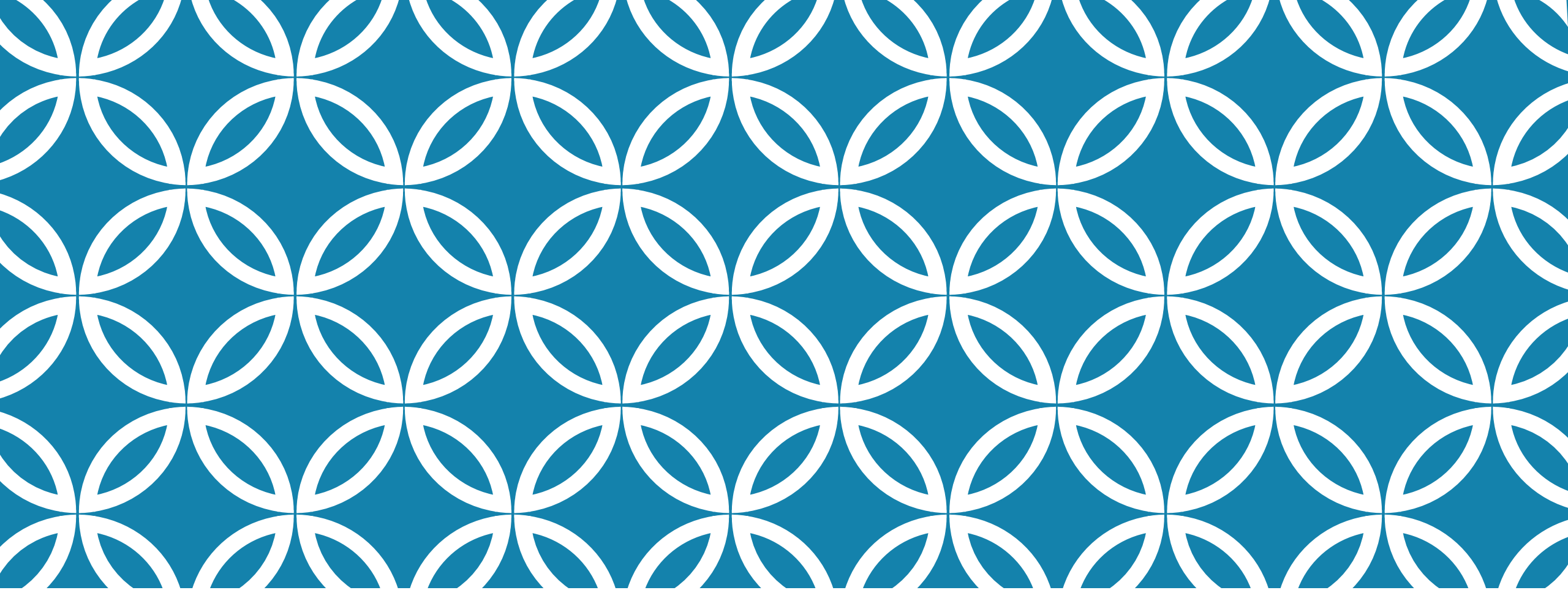


Point-and-Click Is Not Reproducible

- Excel does not record user actions
- Manual documentation of user actions is error-prone
- Manual analyses cannot be repeated on new data sets

Clinicians Don’t Code

- We have limited quantitative/statistics skills
  - Statistics is not a pre-med requirement 🤯
- We have serious time constraints
  - If you ask us to do something hard or time consuming, you need to tell us why it’s important for our patients.
- We learn most things from more senior clinicians
  - … and more senior clinicians generally don’t know how to code either.



Reproducible Clinical Data Analysis with R and RStudio

Goals

1. Appreciate Reproducibility in Data Analysis
2. Learn a Practical Way to Analyze Clinical Data Reproducibly

Objectives

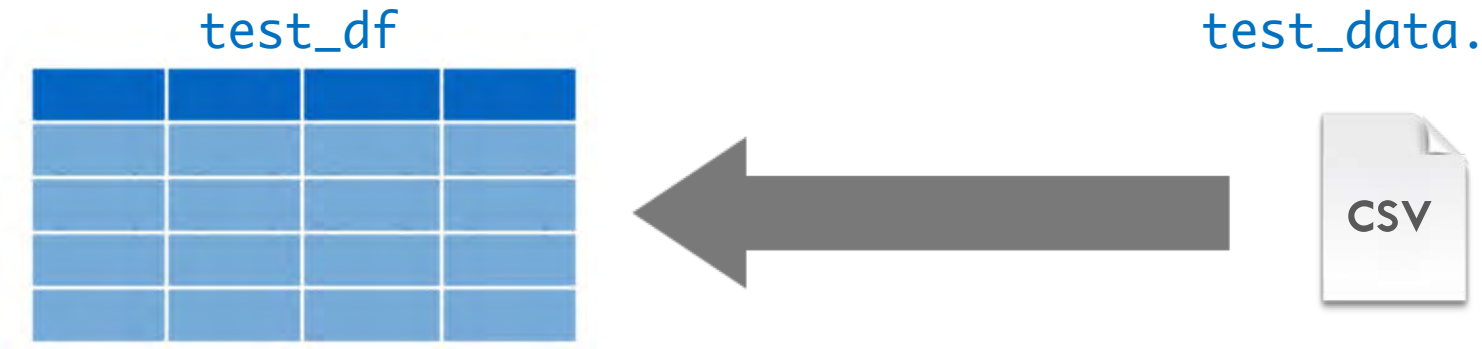
1. Define “Reproducibility” and Explain its Importance
2. Learn How to Use R/RStudio to Import Data from Files and Databases; Transform Data; and Visualize Data
3. Create a Reproducible Report Addressing a Clinical Question

read_csv()

data frame to read data into

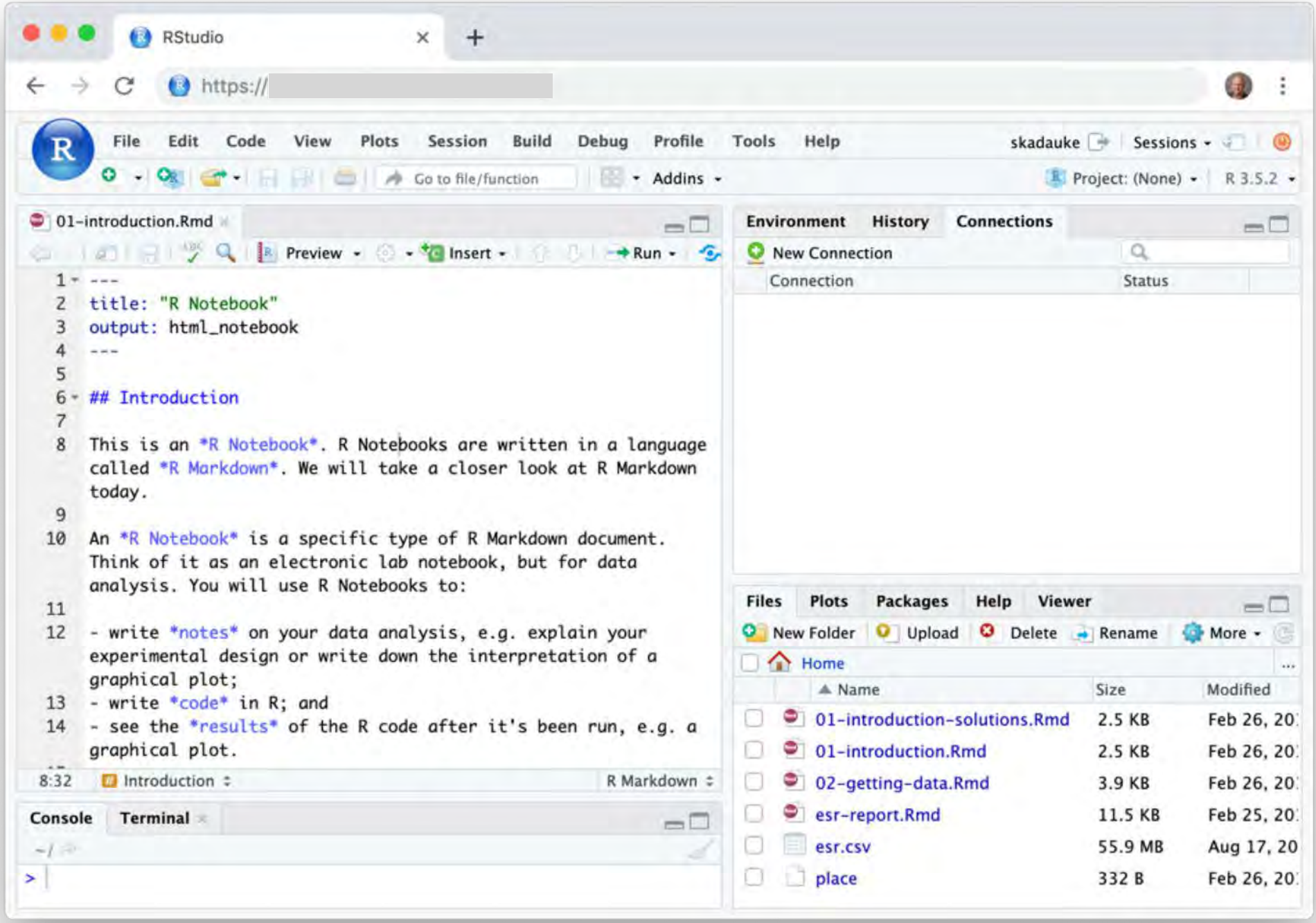
name of CSV file

```
test_df <- read_csv("test_data.csv")
```



Your Turn

Open `01-introduction.Rmd`. Work through the section "Your Turn #1".


RStudio
https://
File Edit Code View Plots Session Build Debug Profile Tools Help
skadauke
Sessions
Go to file/function
Addins
Project: (None)
R 3.5.2
01-introduction.Rmd
Preview
Insert
Run
1 ---
2 title: "R Notebook"
3 output: html_notebook
4 ---
5
6 ## Introduction
7
8 This is an *R Notebook*. R Notebooks are written in a language called *R Markdown*. We will take a closer look at R Markdown today.
9
10 An *R Notebook* is a specific type of R Markdown document. Think of it as an electronic lab notebook, but for data analysis. You will use R Notebooks to:
11
12 - write *notes* on your data analysis, e.g. explain your experimental design or write down the interpretation of a graphical plot;
13 - write *code* in R; and
14 - see the *results* of the R code after it's been run, e.g. a graphical plot.
8:32
Introduction
R Markdown
Console
Terminal
~/
>
Environment
History
Connections
New Connection
Connection
Status
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
01-introduction-solutions.Rmd 2.5 KB Feb 26, 20
01-introduction.Rmd 2.5 KB Feb 26, 20
02-getting-data.Rmd 3.9 KB Feb 26, 20
esr-report.Rmd 11.5 KB Feb 25, 20
esr.csv 55.9 MB Aug 17, 20
place 332 B Feb 26, 20

Introduction

Getting Data

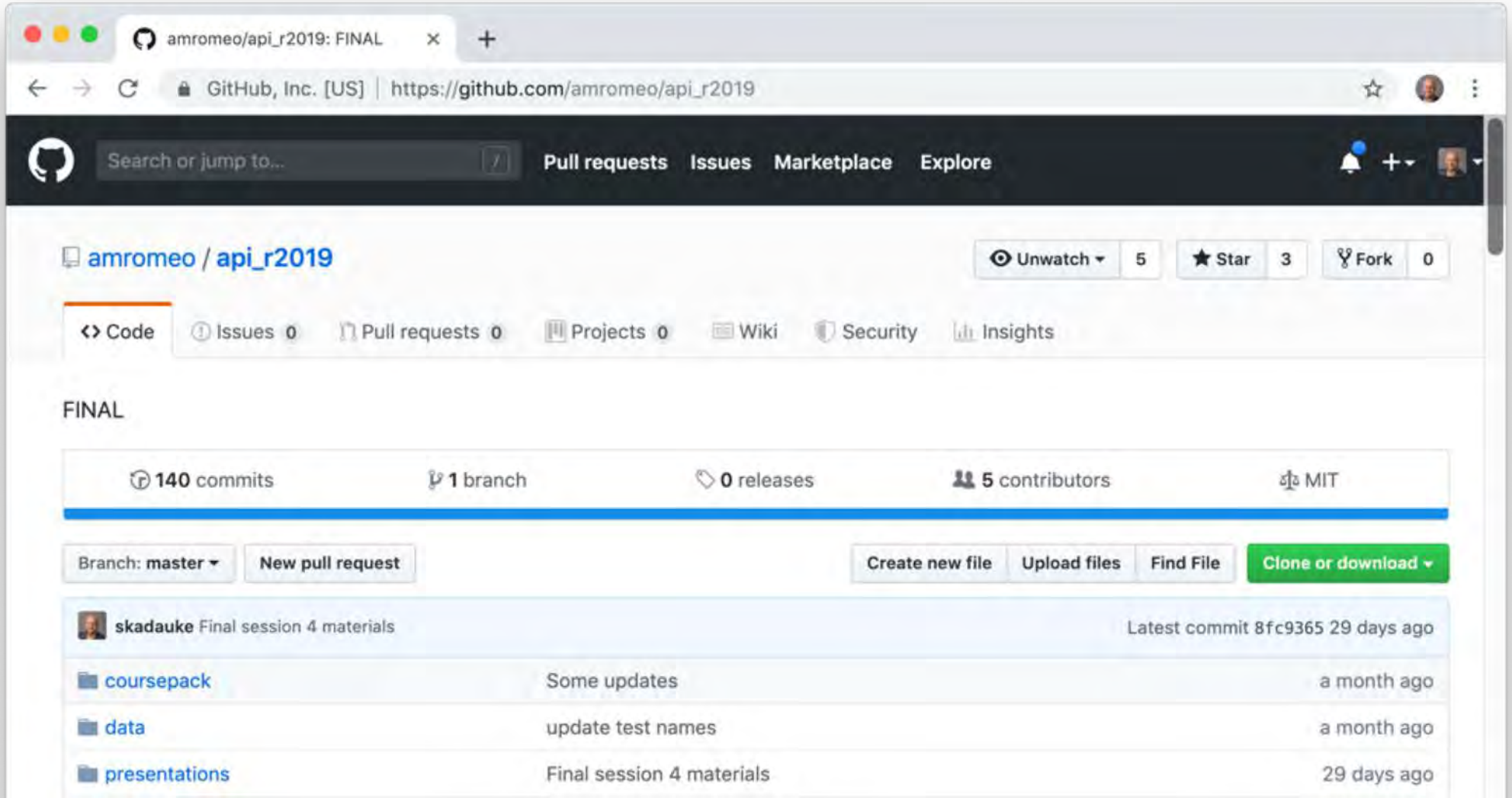
Exploring Data

Reproducible Reports

“Hackathons”

Course Project Presentation

https://github.com/amromeo/api_r2019



The Room Where it Happens: Genchi Genbutsu

- **Every** analyst or engineer should “see for themselves”
- Education is great; experience is **necessary**
- Being in the room makes you a participant

Getting there: the analyst perspective

- Make it a team norm
- Make it mandatory?
- Need leadership buy-in
- Best with your own guide

Tensions with a cross-functional model

- More Meetings
- Less immediate productivity
- Can be uncomfortable

Summary

- To achieve a “culture of analytics,” subject matter experts and technical users must closely collaborate, ideally by working together through all stages of the analysis.

- A course in reproducible clinical data analysis tailored to clinicians could lower the barrier to productive collaboration with technical users.

- Technical users should have an immersion experience in a relevant clinical domain.

Thank You!