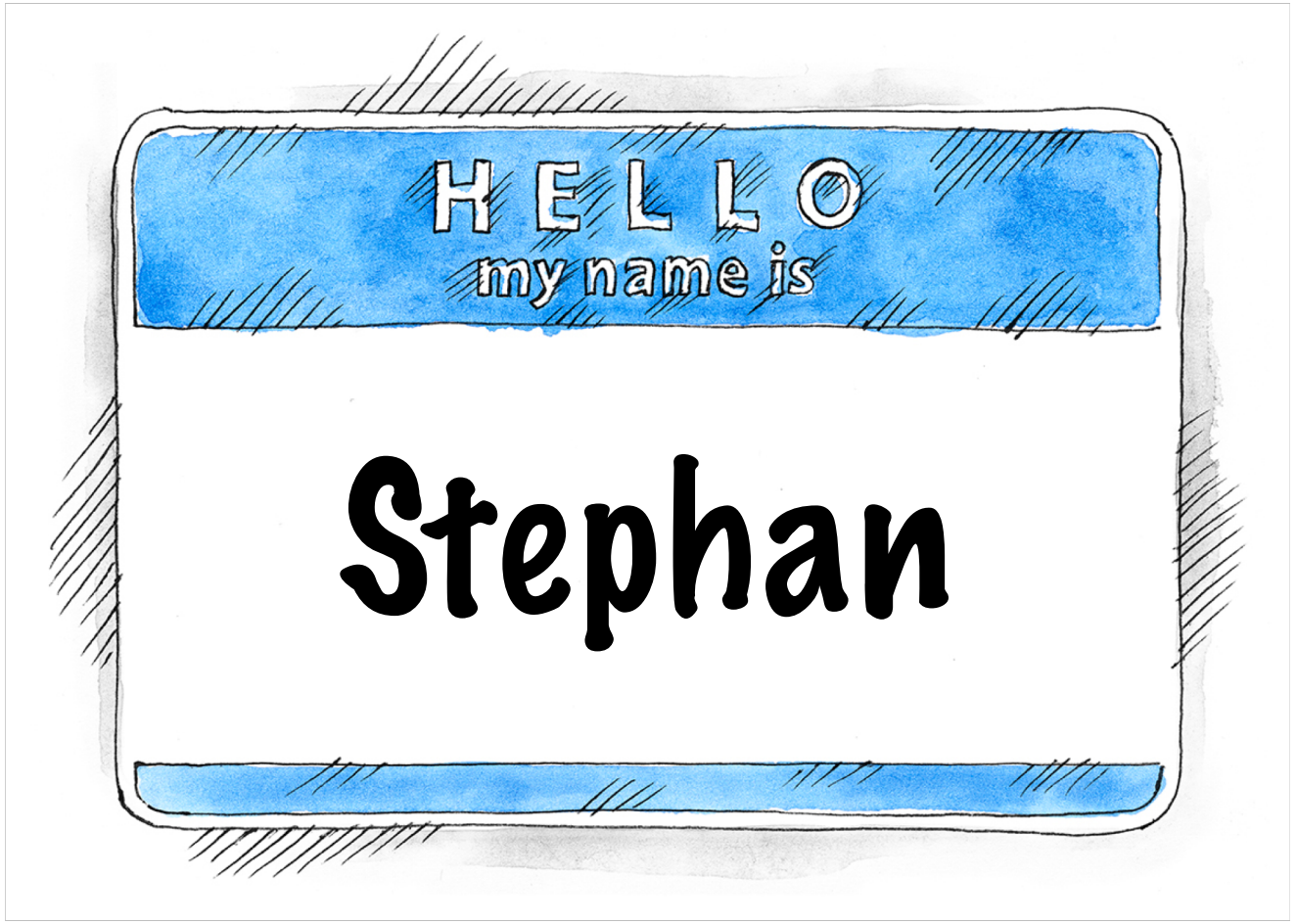
# R at PENN/CHOP:
# Improving Collaboration Between Clinicians and Analysts

Stephan Kadauke, MD, PhD

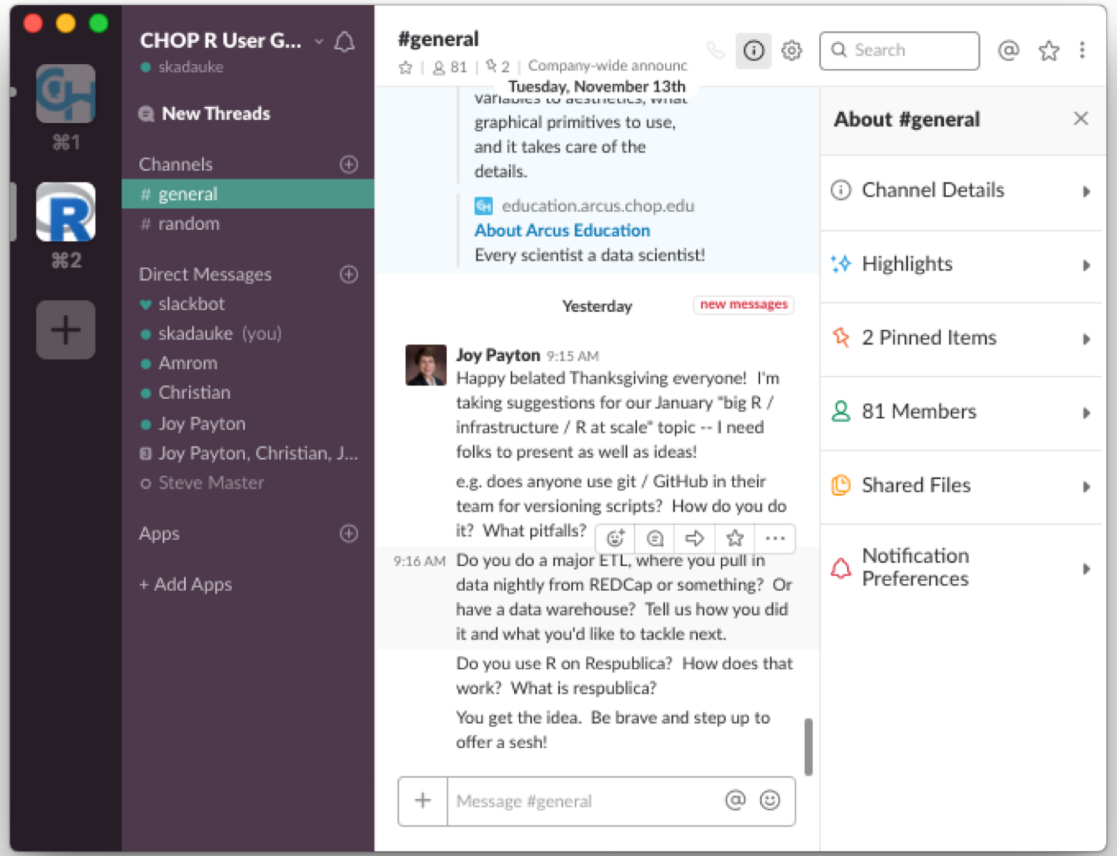
PhillyR Meetup

February 27, 2019




made in Germany


chopr

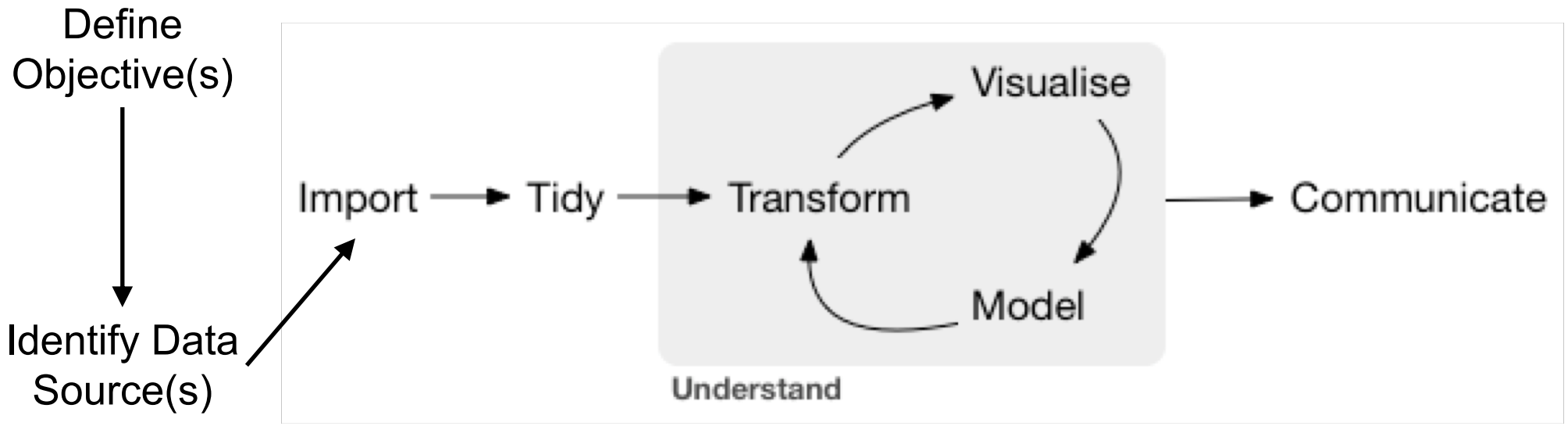

CHOP R User G...
skadauke
New Threads
Channels
# general
# random
Direct Messages
slackbot
skadauke (you)
Amrom
Christian
Joy Payton
Joy Payton, Christian, J...
Steve Master
Apps
+ Add Apps
#general
81 | 2 | Company-wide announc
Search
Tuesday, November 13th
graphical primitives to use,
and it takes 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
- Discuss barriers to effective clinician-analyst collaboration
- Describe a course in Reproducible Clinical Data Analysis for clinicians aimed at improving clinician-analyst collaboration

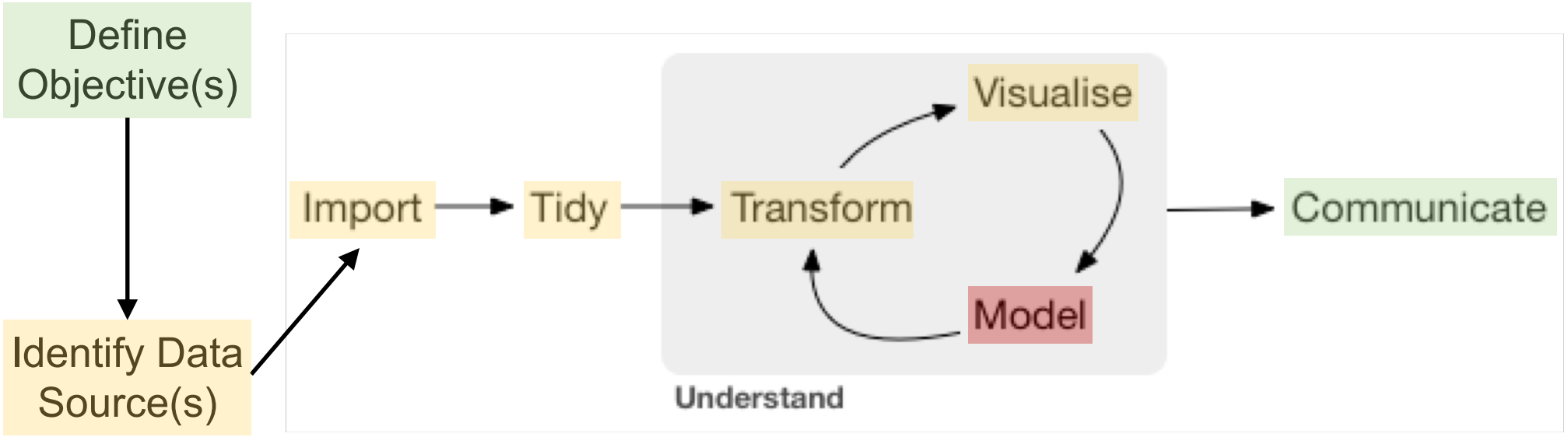
# Analytics Maturation


Adapted from:
Mark Wu (Tableau)

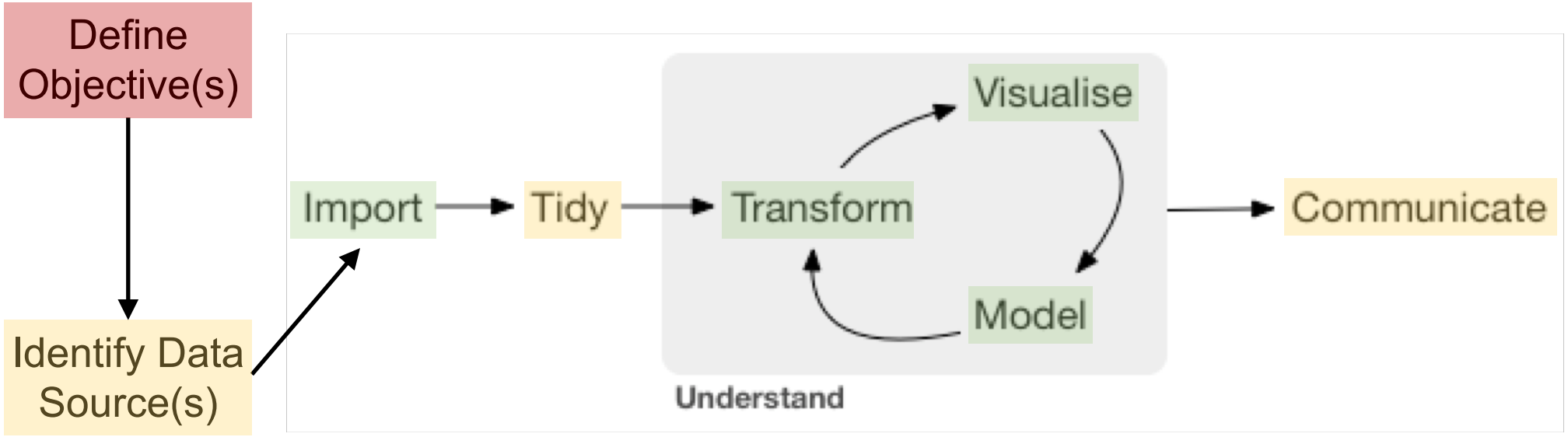



Adapted from: Wickham, Grolemund.
R for Data Science (2017)

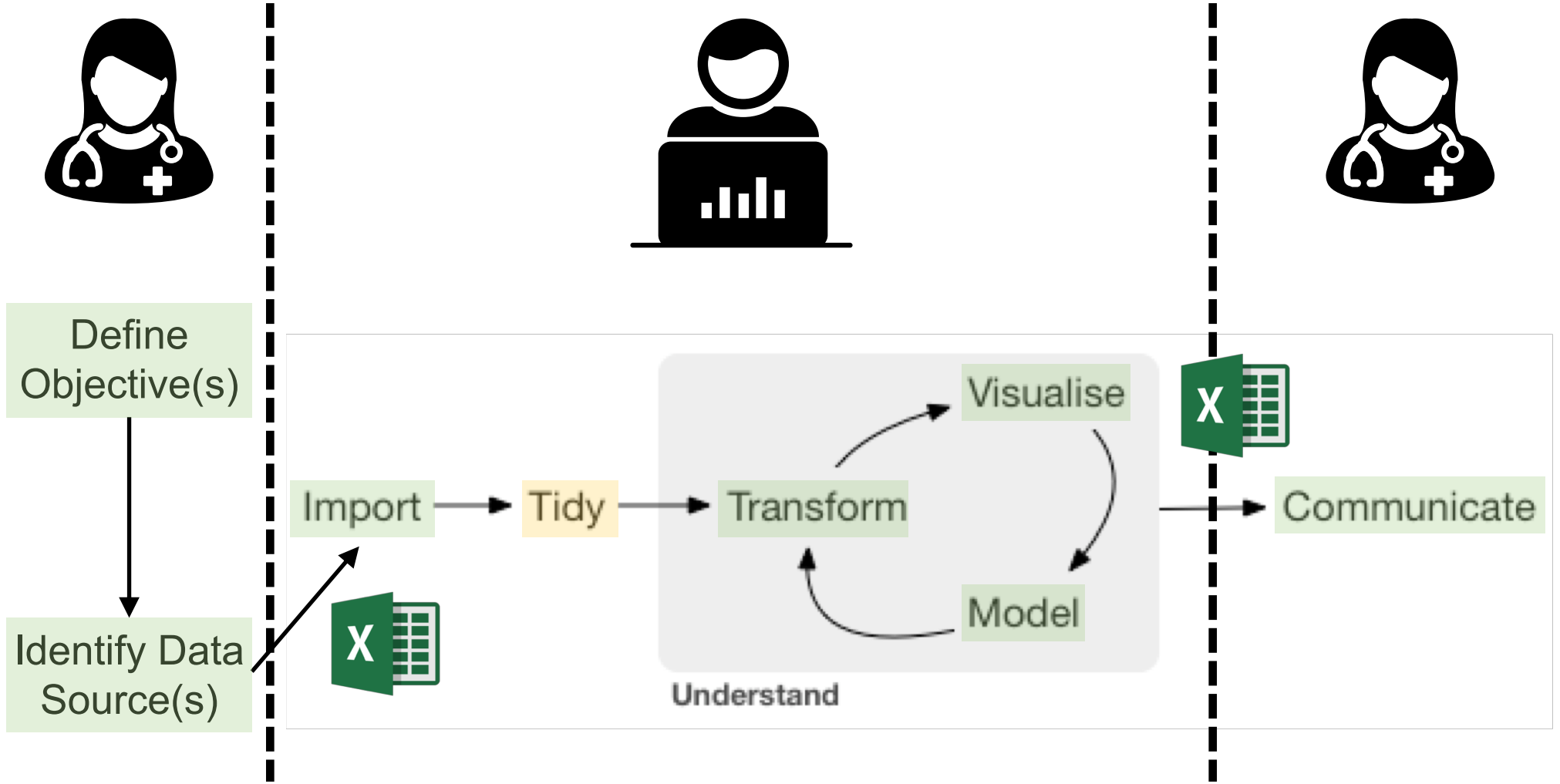
# Clinician-driven Analysis



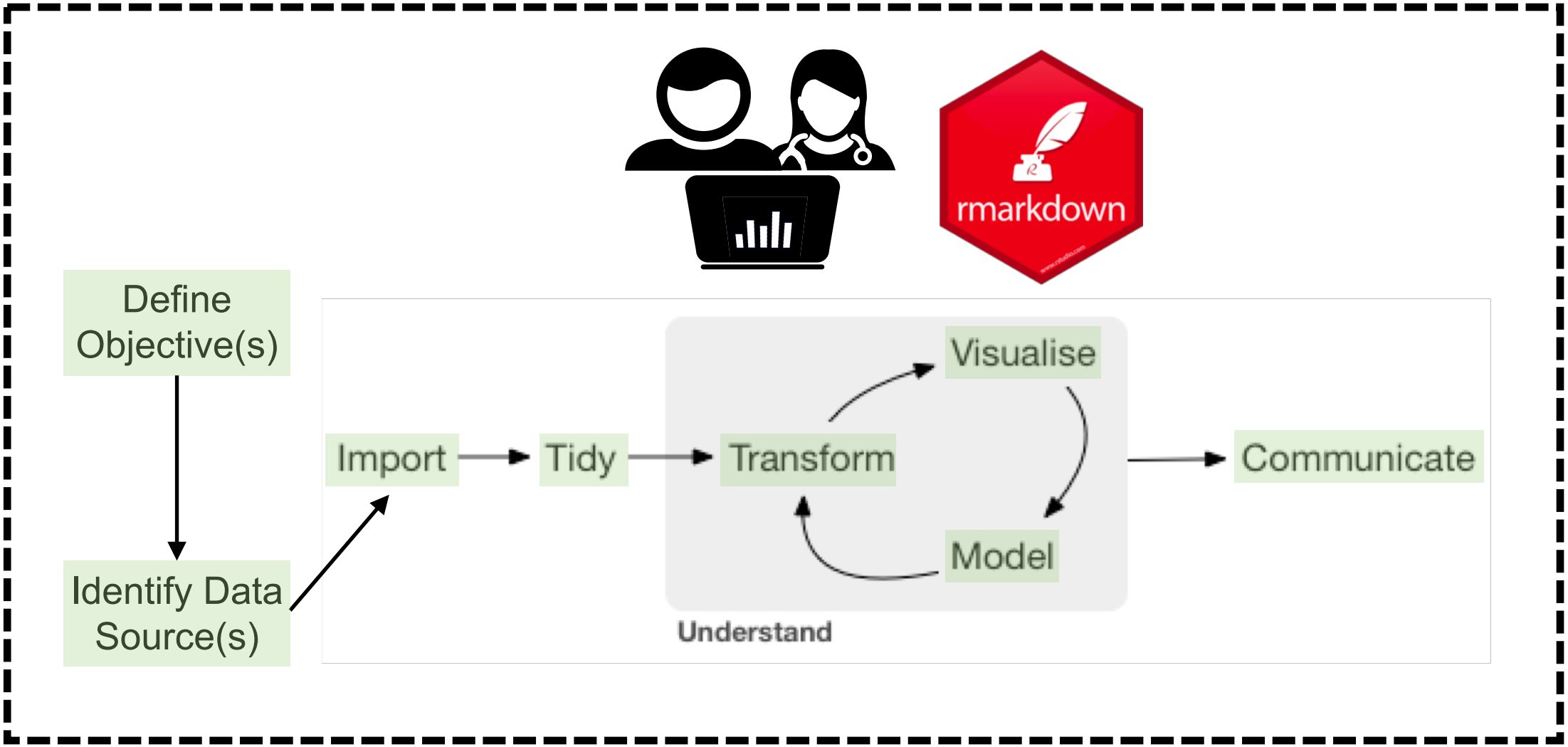
# Analyst-driven Analysis



# Collaboration: Waterfall model



# Collaboration: Agile model

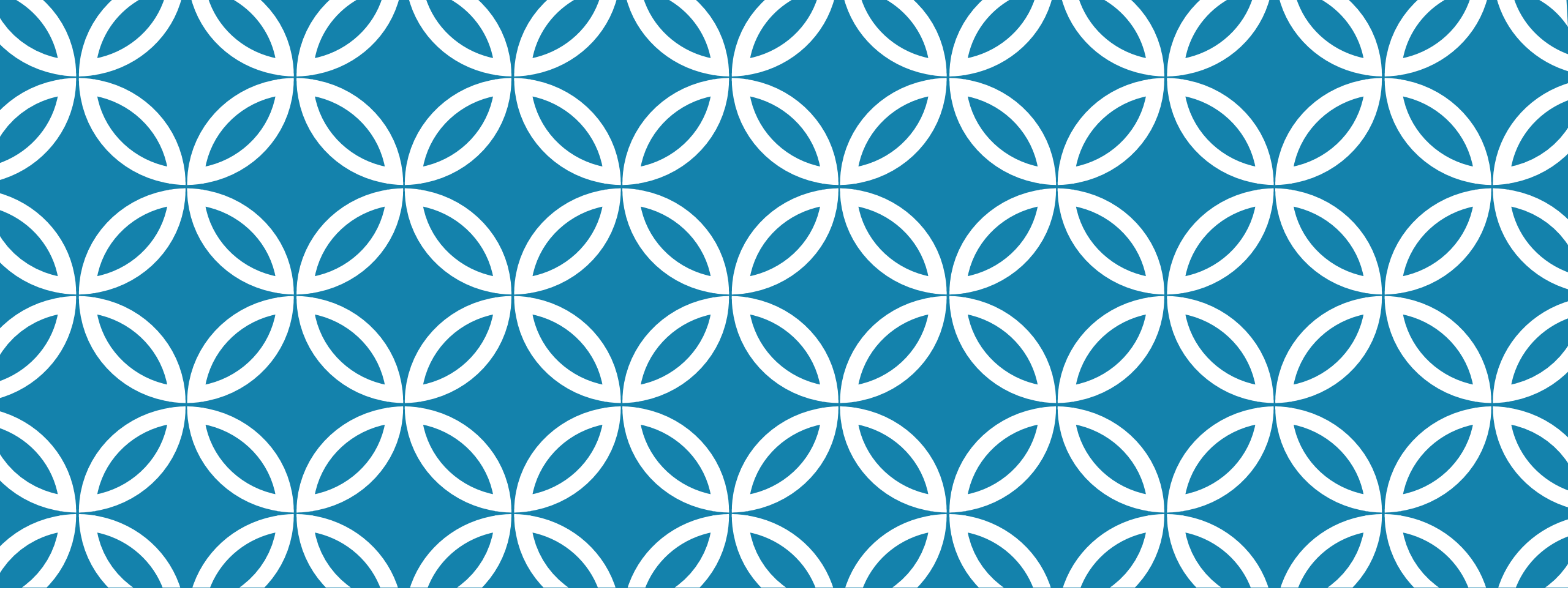


# Point-and-Click Is Not Reproducible

- 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
  - Stats is not a pre-med requirement! And we didn’t really care for physics either.
- 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 other, more senior clinicians
  - … and more senior clinicians don’t 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

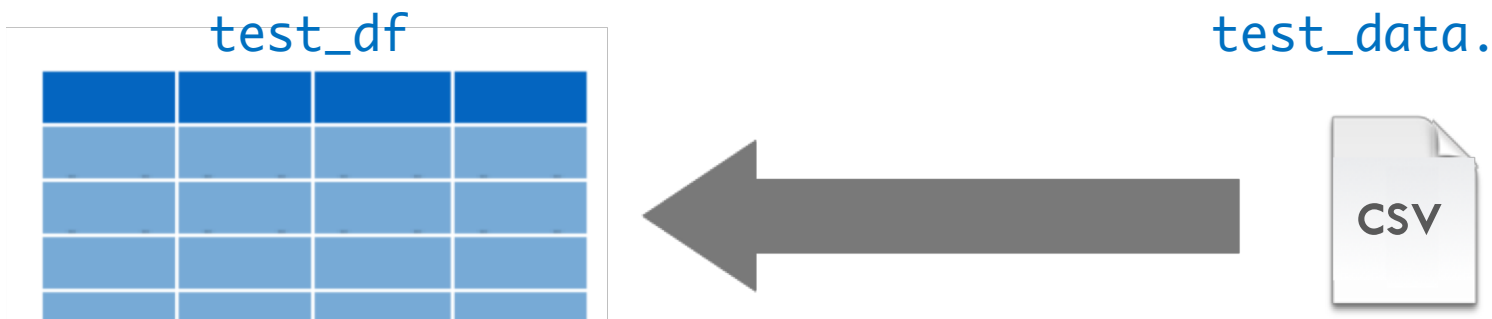
1. Define “Reproducibility” and Explain its Importance
2. Learn How to Use R/RStudio to Import Data from Files and Databases; Transform Data; and Visualize Data
3. Create a Reproducible Report Addressing a Clinical Question

# read_csv()

data frame to read data into

name of CSV file

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

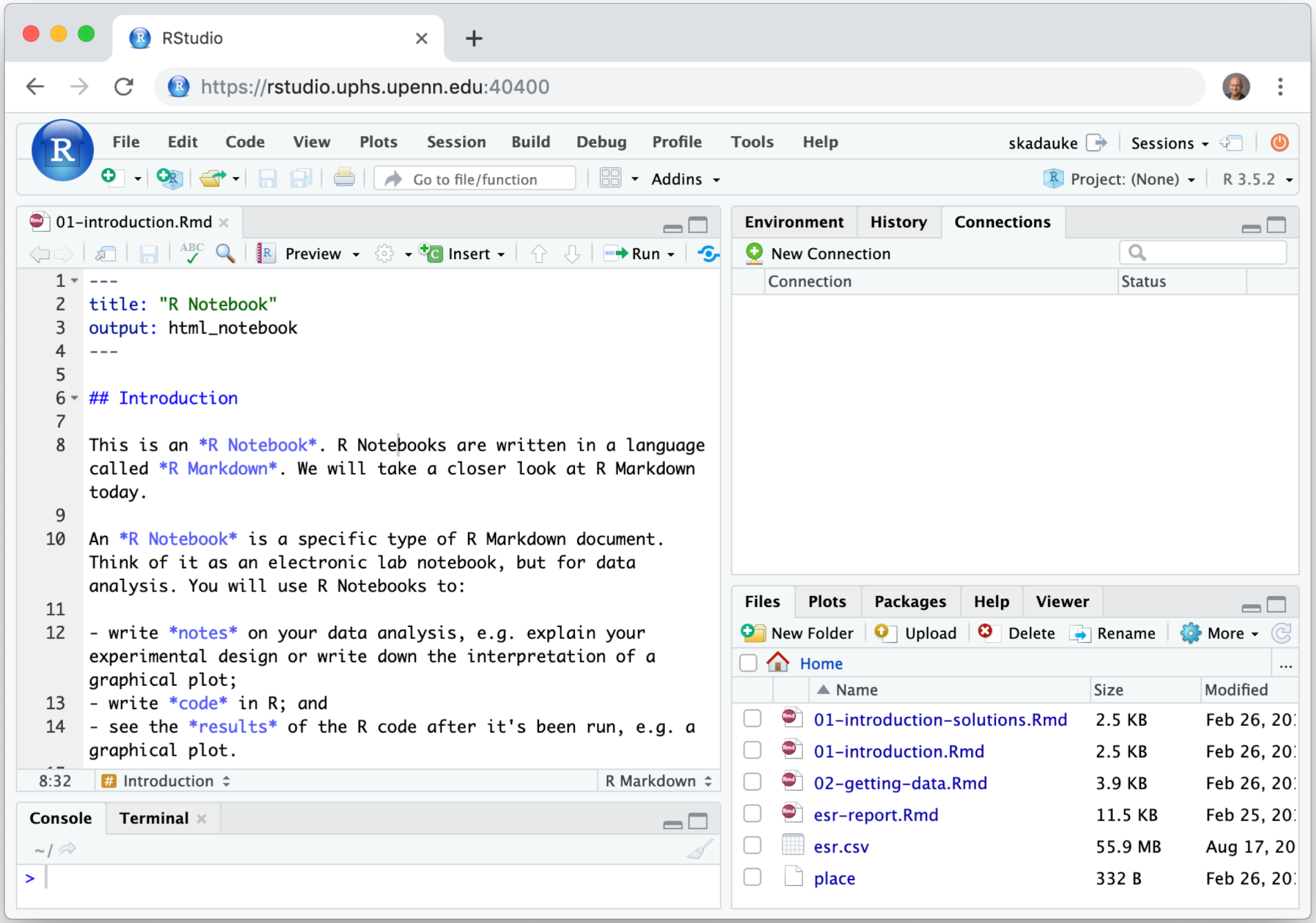



# Your Turn #2

Log onto the RStudio Server by pointing your browser to:

https://example.com

Open `02-getting-data.Rmd`. Work through the section "Your Turn #2".


RStudio
https://rstudio.uphs.upenn.edu:40400
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

# Summary

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

- Agile collaboration requires a reproducible workflow, which in turn requires coding. This is a problem because clinicians generally don’t code.

- A course in reproducible clinical data analysis tailored to clinicians could lower the barrier to collaboration with analysts. This course was recently introduced at Penn and we’re looking to open up a version to the public in the near future.

# Thank You!