

How to Truly "Excel" at Data Analysis and Visualization: An Introduction to the R Programming Language

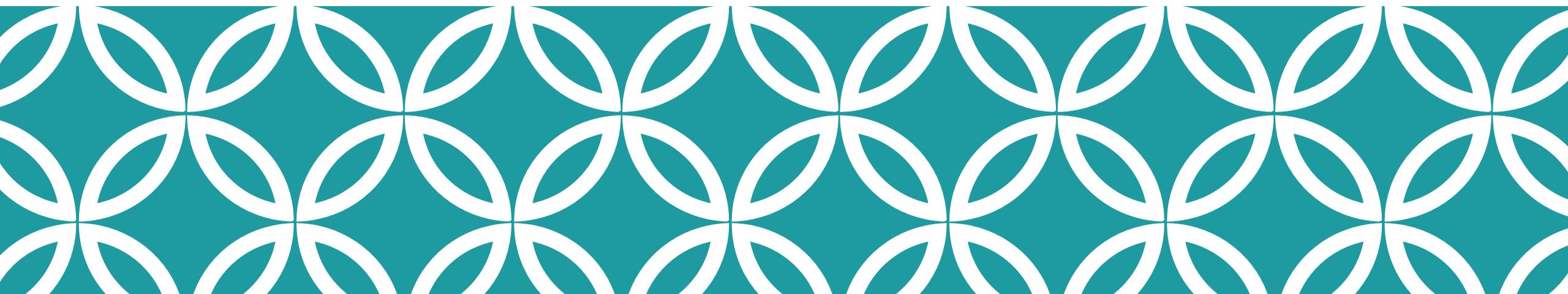
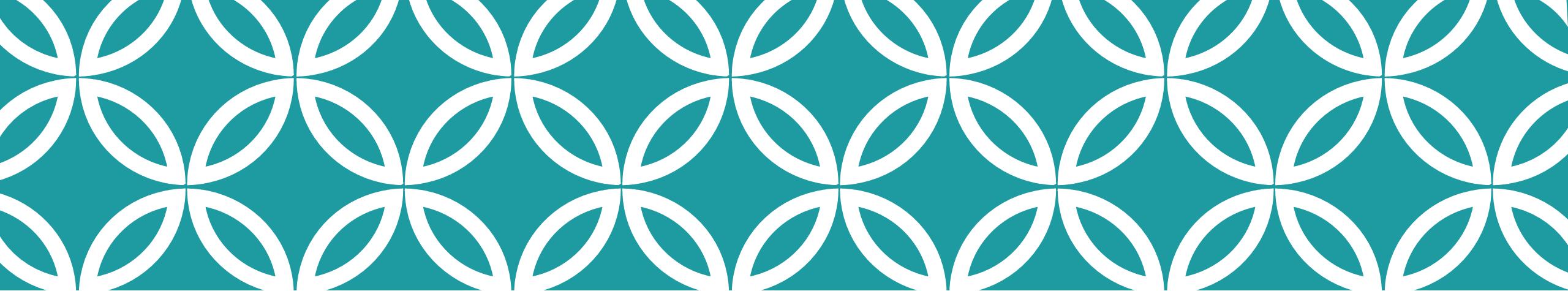
**Daniel Herman
Patrick Mathias
Joseph Rudolf
December 13, 2020**

Course Introduction

Course Goals

1. Gain familiarity with the R programming language and RStudio software
2. Learn some fundamentals of coding using Tidyverse tools on a clinical data set
3. Develop workflows for reproducible data analysis using RMarkdown

December 13 2020	Session	Instructor
8:30 am - 8:50 am	Introduction to the Course	Daniel Herman Patrick Mathias Joseph Rudolf
9:00 am - 9:45 am	Introduction to R and RStudio	Joseph Rudolf
10:00 am - 10:45 am	Reproducible Reporting	Patrick Mathias
11:00 am - 11:00 am	Data Visualization	Joseph Rudolf
12:30 pm - 1:15 pm	Data Slicing	Patrick Mathias
1:30 pm - 2:15 pm	Data Transformation	Daniel Herman
2:30 pm - 3:15 pm	Statistical Analysis	Daniel Herman



Who are we?

Daniel Herman

Assistant Professor of Pathology and
Laboratory Medicine

University of Pennsylvania Perelman School
of Medicine

Director, Endocrinology Laboratory

Hospital of the University of Pennsylvania



Patrick Mathias

Assistant Professor, Department of
Laboratory Medicine

University of Washington School of
Medicine

Associate Medical Director, Laboratory
Medicine Informatics



Joseph Rudolf

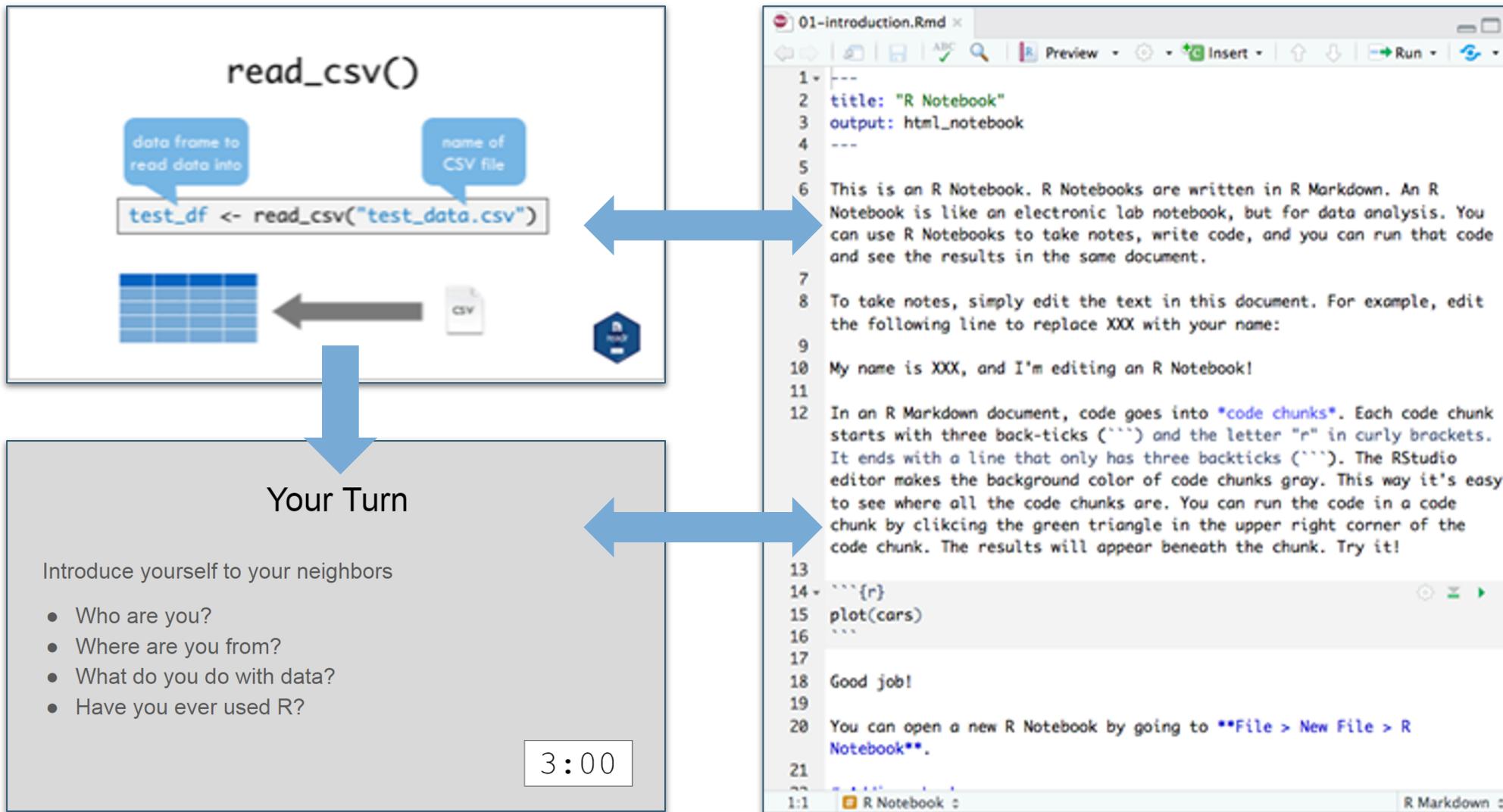
Assistant Professor, Department of Pathology,
University of Utah Medical School

Medical Director, Automated Core Laboratory, ARUP
Laboratories



Workshop Workflow

Sessions

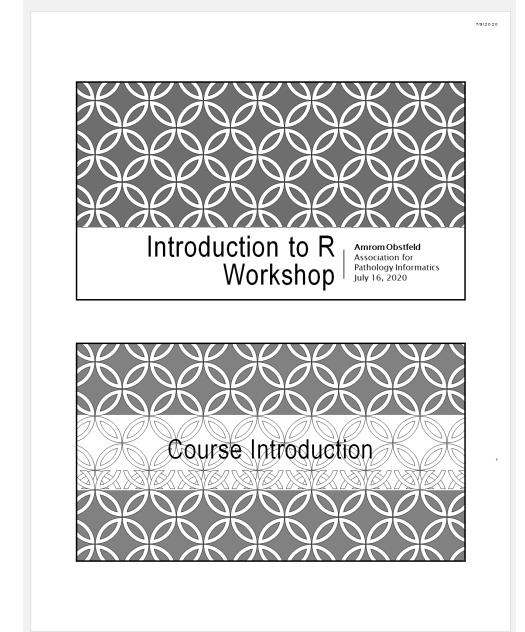


Configuring Your Setup



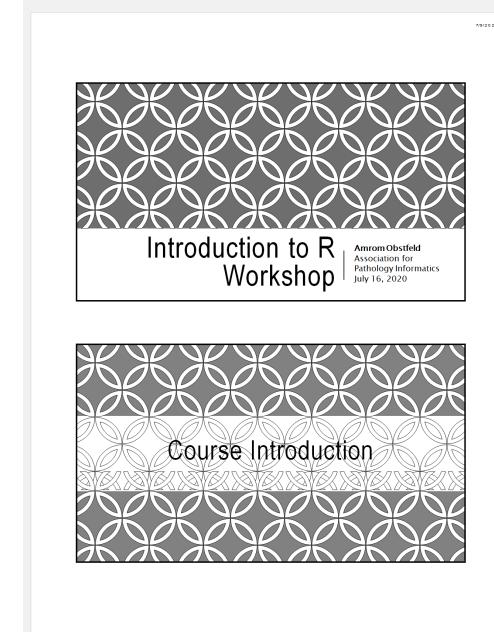
A screenshot of the RStudio interface. The top menu bar includes File, Edit, Code, View, Project, Workspace, Plots, Tools, and Help. The left sidebar shows a file tree with "diamondPricing.R" and "formatPlot.R" selected. The workspace pane displays the "diamonds" dataset with 53940 observations and 10 variables. The console pane shows R code for generating a scatter plot. The main plot area is titled "Diamond Pricing" and shows a scatter plot of Price vs. Carat, with points colored by Clarity. The legend indicates clarity levels: I1 (red), SI2 (orange), SI1 (yellow), VS1 (green), VS2 (light green), VSI (cyan), VVS2 (blue), VVS1 (light blue), and IF (pink).

```
library(ggplot2)
source("plots/formatPlot.R")
view(diamonds)
summary(diamonds)
avesize <- round(mean(diamonds$carat), 4)
clarity <- levels(diamonds$clarity)
p <- qplot(carat, price,
           data=diamonds, color=clarity,
           xlab="Carat", ylab="Price",
           main="Diamond Pricing")
format.plot(p, size=24)
```



Workshop Coursebook

- Print out of all slides
- Appendix
 - Cheat Sheets
 - Useful resources



Using Zoom



- Participants muted
- Chat window
- Non-verbal feedback

A video call interface featuring a scenic mountain landscape as the background. The video feed is muted by a participant named Joseph Rudolf.

Participants (1)

JR Joseph Rudolf (Host, me) X End

Controls:

- yes (green checkmark)
- no (red X)
- go slower (left arrow)
- go faster (right arrow)
- more (dots)
- clear all (grey square)

Buttons:

- Invite
- Mute All
- More ▼

Chat:

To: Everyone

Type message here...

Bottom Controls:

- Unmute
- Stop Video
- Security
- Participants 1
- Chat
- Share Screen ▲
- Reactions
- More
- End

Getting Help

- During presentation – Enter your question in the chat, an instructor will respond to the group or direct message you as appropriate.
- Between sessions – Feel free to unmute yourself and converse with instructors.



Who are you?

Your Turn

Where are you today?

Type your name and location in the chat.

Pre-Course Survey Results

Time for the Course!!