Introduction to R Workshop

Amrom Obstfeld
Association for Pathology Informatics Summit
May 6, 2019
Course Introduction
Goals and Objectives

• Advocate for the use of R as a means of improving reproducibility in clinical data analysis
• Demonstrate how R is used to perform analyses of laboratory operational data
• Establish a basis of understanding in the 'tidy' approach to data analysis within the framework of R
<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00 am - 8:00 am</td>
<td>BREAKFAST - BALLROOM LOBBY - 2ND FLOOR</td>
</tr>
<tr>
<td>8:00 am - 8:10 am</td>
<td>Instructor Introductions</td>
</tr>
<tr>
<td>8:10 am - 9:50 am</td>
<td>Introduction to R and RStudio for Reproducible Reporting</td>
</tr>
<tr>
<td>9:50 am - 10:10 am</td>
<td>REFRESHMENT BREAK - BALLROOM LOBBY - 2ND FLOOR</td>
</tr>
<tr>
<td>10:10 am - 11:50 am</td>
<td>Data Wrangling</td>
</tr>
<tr>
<td>12:00 pm - 1:00 pm</td>
<td>LUNCH - BALLROOM LOBBY - 2ND FLOOR</td>
</tr>
<tr>
<td>1:00 pm - 2:50 pm</td>
<td>Data Understanding</td>
</tr>
<tr>
<td>2:50 pm - 3:10 pm</td>
<td>REFRESHMENT BREAK - BALLROOM LOBBY - 2ND FLOOR</td>
</tr>
<tr>
<td>3:10 pm - 5:00 pm</td>
<td>Exploratory Data Analysis</td>
</tr>
</tbody>
</table>
Lectures

Your Turn

Introduce yourself to your neighbors
- Who are you?
- Where are you from?
- What do you do with data?
- Have you ever used R?

```
# R Notebook

This is an R Notebook. R Notebooks are written in R Markdown. An R Notebook is like an electronic lab notebook, but for data analysis. You can use R Notebooks to take notes, write code, and you can run that code and see the results in the same document.

To take notes, simply edit the text in this document. For example, edit the following line to replace XXX with your name:
```

My name is XXX, and I'm editing an R Notebook!
```

In an R Markdown document, code goes into "code chunks". Each code chunk starts with three back-ticks (```) and the letter "r" in curly brackets. It ends with a line that only has three backticks (```). The RStudio editor makes the background color of code chunks grey. This way it's easy to see where all the code chunks are. You can run the code in a code chunk by clicking the green triangle in the upper right corner of the code chunk. The results will appear beneath the chunk. Try it!
```

```r
plot(cars)
```

Good job!
```

You can open a new R Notebook by going to **File > New File > R Notebook**.
```

```r
3:00
```
Course Materials

Introduction to R
Association for Pathology Informatics Summit 2019

Course Textbook
Slides with course notes

Cheat sheet

Resources for Future Learning

Course textbook: Data Transformation with dplyr:

- Workshop on using dplyr for data manipulation

- Resources:
  - Course materials
  - Slides with course notes
  - Cheatsheets
  - Resources

Massive Open Online Courses (MOOCs) online courses

- Introduction to Data Science by Harvard
  - The definitive text on data science written by the leading R developers

- Data Science for Business by Columbia University
  - A course to teach the business of data science

- Coursera and edX courses
  - A variety of courses on R programming and data science

Resources

- Cheatsheets

- Workshop on using dplyr for data manipulation

- Resources:
  - Course materials
  - Slides with course notes
  - Cheatsheets
  - Resources
https://github.com/amromeo/api_r2019
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
Stephan Kadauke

Assistant Professor of Clinical Pathology and Laboratory Medicine

University of Pennsylvania Perelman School of Medicine

Assistant Director of the Cell and Gene Therapy Laboratory

Children's Hospital of Philadelphia
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 Laboratory Medicine and Pathology, University of Minnesota

Director of Laboratory Medicine and Pathology Informatics, Department of Laboratory Medicine and Pathology, University of Minnesota

Laboratory Medicine and Pathology Chief Medical Informatics Officer, Fairview Health Services
Amrom Obstfeld

Assistant Professor of Clinical Pathology and Laboratory Medicine

University of Pennsylvania Perelman School of Medicine

Director of Hematology and Coagulation Laboratories

Children's Hospital of Philadelphia
Who are you?
What is your job title/role?

21 responses

- Medical student
- Resident or clinical fellow
- Graduate/undergraduate student
- Postdoctoral researcher
- Administrative director
- Informatics or laboratory staff
- Pathologist (non-trainee)
- Other faculty (non-pathology)
Have you taken any previous courses in computer programming (in any language)?

21 responses

61.9% Yes
38.1% No
If you have programming experience, please indicate your level of proficiency (or check "No experience")

21 responses

- 47.6% Advanced (e.g. fluent coder, reading/writing complex code is a fundamental component of your work)
- 28.6% Intermediate (e.g. can write and execute your own code for simple analysis)
- 19% Basic (e.g. can load and execute code written by others)
- 19% No experience
Have you ever used (or tried programming in) R in the past?

21 responses

- Yes: 33.3%
- No: 66.7%
I believe that the benefits of making a data analysis reproducible justify a significant upfront investment.

21 responses
Your Turn

Introduce yourself to your neighbors

Who are you?
Where are you from?
What do you do with data?
Have you ever used R?
Final Tips

- **Remember that computers are not actually that smart**: R will do exactly what you tell it to do, even if that’s not what you want! Your code can’t have errors because R can’t figure out what you mean.

- **Take the “copy, paste, and tweak” approach**: It is often much easier to taking existing code that you know works and modify it to suit your ends, rather than trying to write new code from scratch. Exercise ctrl-C/ctrl-V!
Final Tips

- **The best way to learn to code is by doing**: Find a project, make a goal, and push yourself to use R!
- **Practice is key!**