

# 05 parting remarks

**teaching  
data  
science**

# workshop materials

 [rstudio.com/teach-ds-jsm19](https://rstudio.com/teach-ds-jsm19)



Hello #dsbox

Overview

Philosophy

Topics

Tech stack

Community

Course content

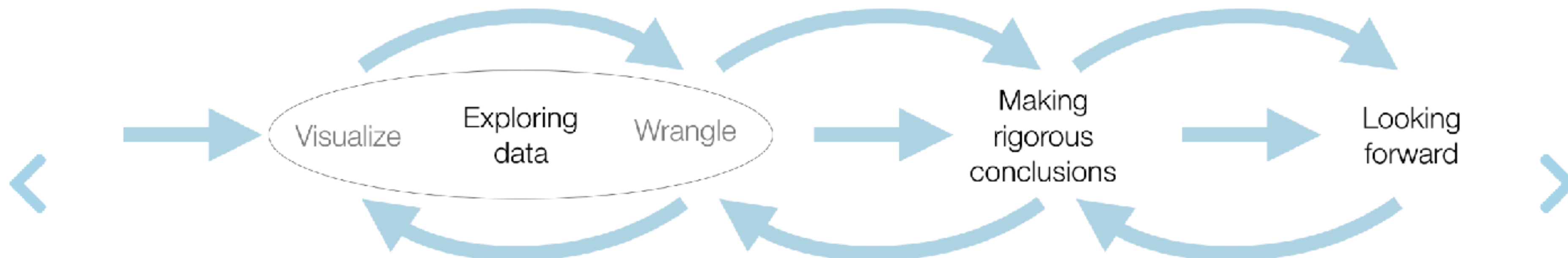
Infrastructure

Pedagogy



## Topics

The course content is organized in three units:



**Unit 1 - Exploring data:** This unit focuses on data visualization and data wrangling. Specifically we cover fundamentals of data and data visualization, confounding variables, and Simpson’s paradox as well as the concept of tidy data, data import, data cleaning, and data curation. We end the unit with web scraping and introduce the idea of iteration in preparation for the next unit. Also in this unit students are introduced to the toolkit: R, RStudio, R Markdown, Git, GitHub, etc.

**Unit 2 - Making rigorous conclusions:** In this part we introduce modeling and statistical inference for making data based conclusions. We discuss building, interpreting, and selecting models, visualizing interaction effects, and prediction and model validity. Statistical inference is introduced from a simulation based perspective, and the Central Limit Theorem is discussed very briefly to lay the foundation for future coursework in statistics.

**Unit 3 - Looking forward:** In the last unit we present a series of modules such as interactive reporting and visualization with Shiny, text analysis, and Bayesian inference. These are independent modules that instructors can choose to include in their introductory data science curriculum depending on how much time they have left in the semester.