

# The art of data investigations

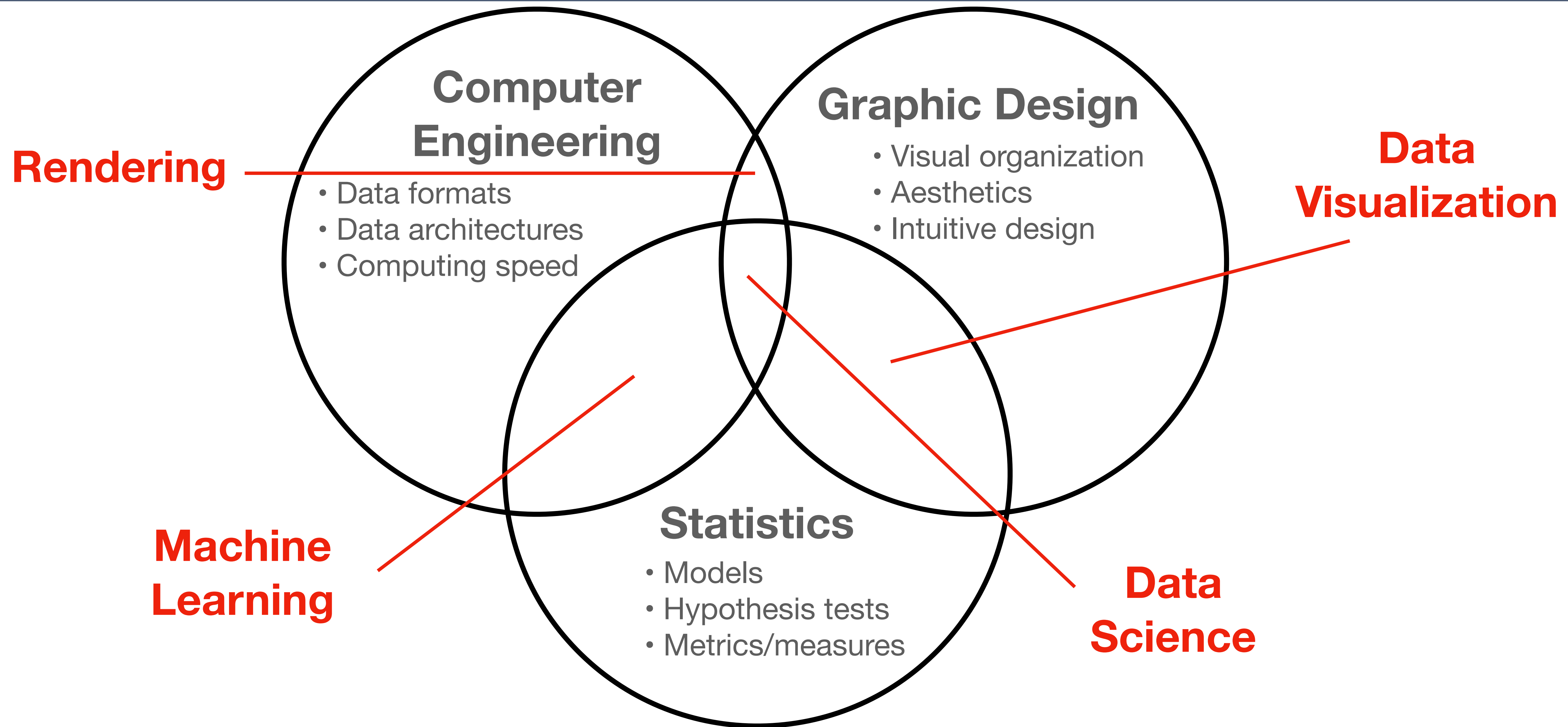
# What is data science?

**Engineering:**  
To build

**Art & Design:**  
To communicate

**Science:**  
To understand

# What is data science?



# What is data science?

**Data science** is an **inter-disciplinary** field that uses scientific methods, processes, algorithms and systems to extract **knowledge** and insights from many structural and **unstructured data**.<sup>[1][2]</sup> - Wikipedia



What is the story that the data is trying to tell you?

**Data science is fundamentally the process of investigation**



# Principles of data investigations

1. Clearly identify what you are asking.
2. Determine the parameters & constraints of your query.
3. Know the context of your evidence.
4. Be systematic in your search.
5. Disturb the data as little as possible.
6. See the story from as many perspectives as possible.
7. Strive for simplicity over control.
8. Communicate your story effectively.

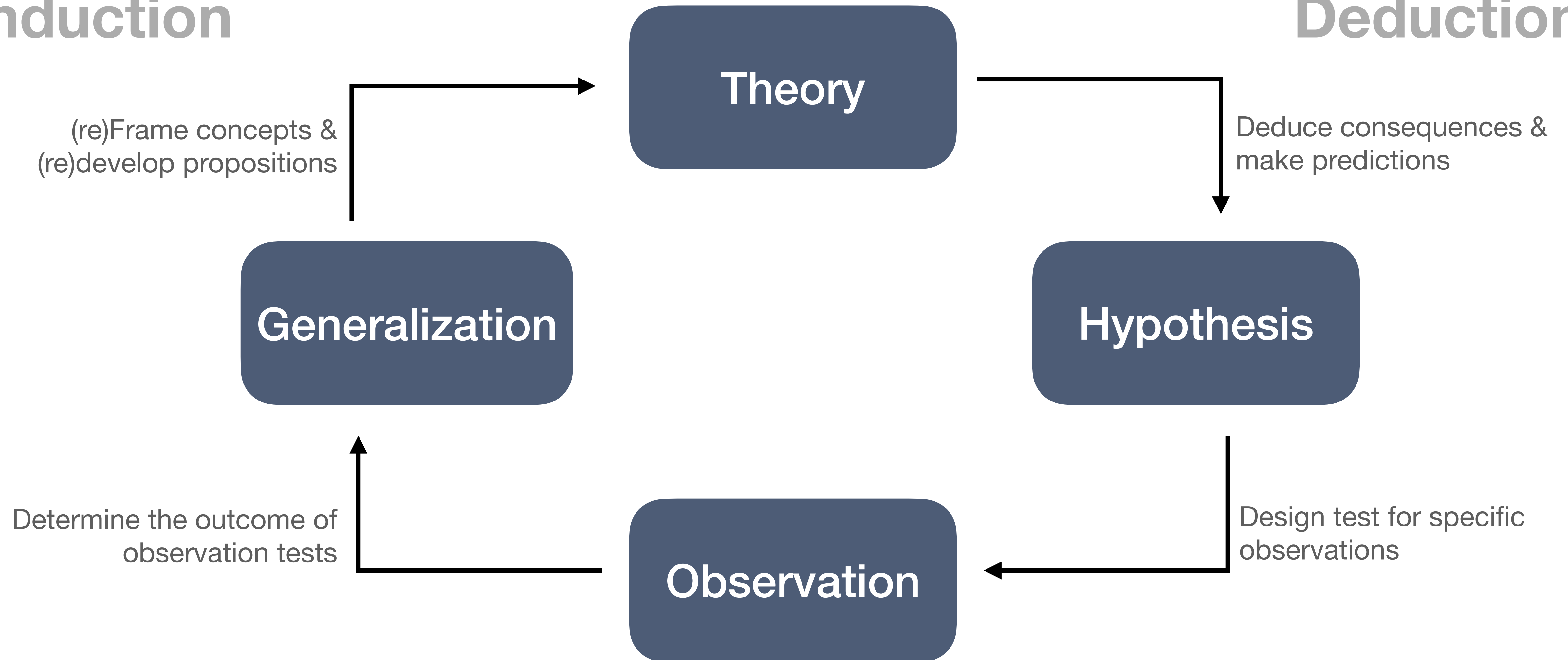


“The Adventure of the Abbey Grange”  
- Sidney Paget

# Walter Wallace's logic “wheel”

Induction

Deduction

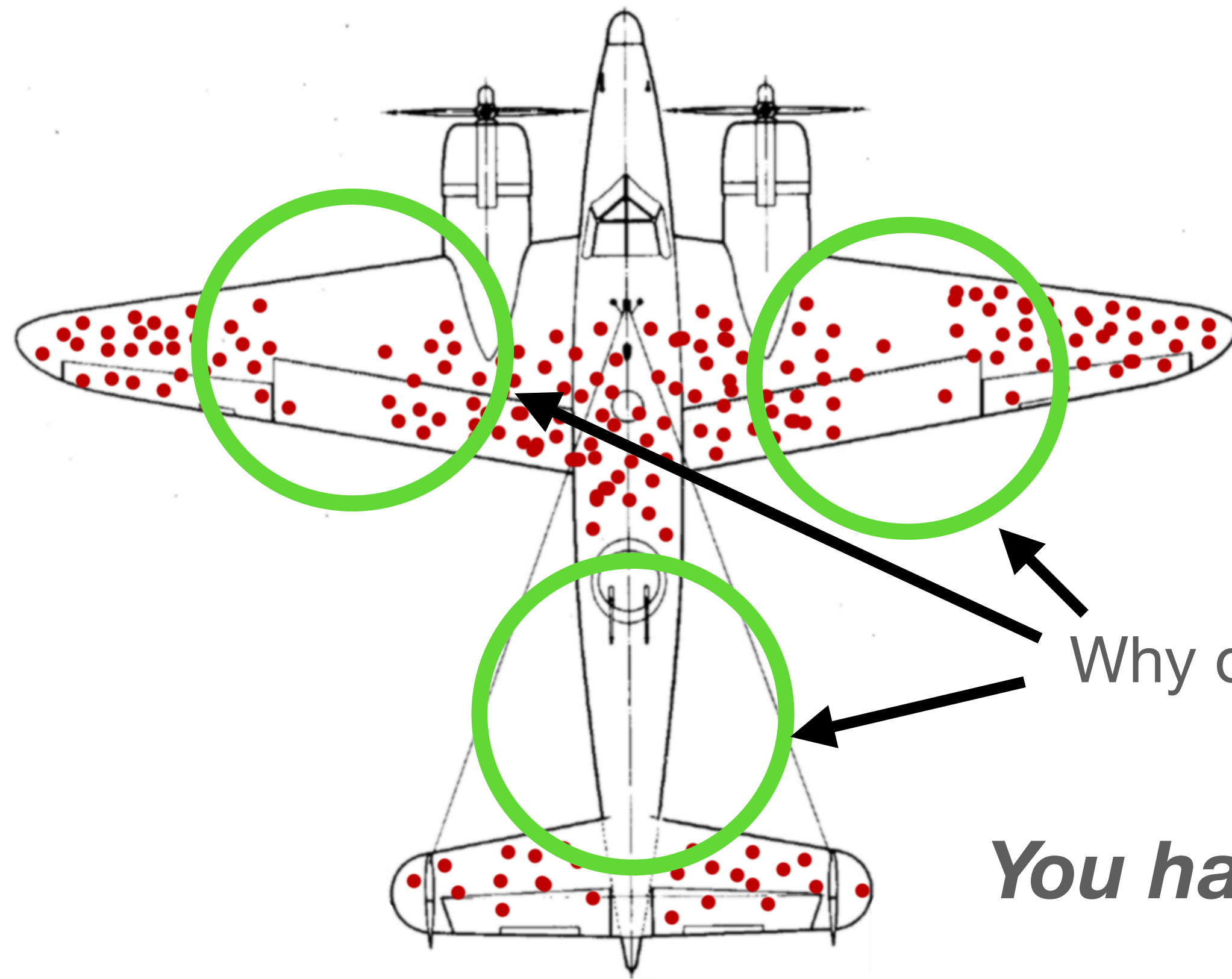


Wallace, W. L. (1971). *The Logic of Science in Sociology*.



# Seeing past the data

## Abraham Wald's airplane dilemma



- Context: World War II
- Question: Where to reinforce planes with armor?
- Data: Maps of where planes that returned from battle were shot.

Why do we not see bullet holes here?

***You have to understand the nature of your data before you can infer anything from it.***



**This class**

# Goal of the class

Make you all more effective data detectives by showing you how to extract intuition and knowledge from your data.

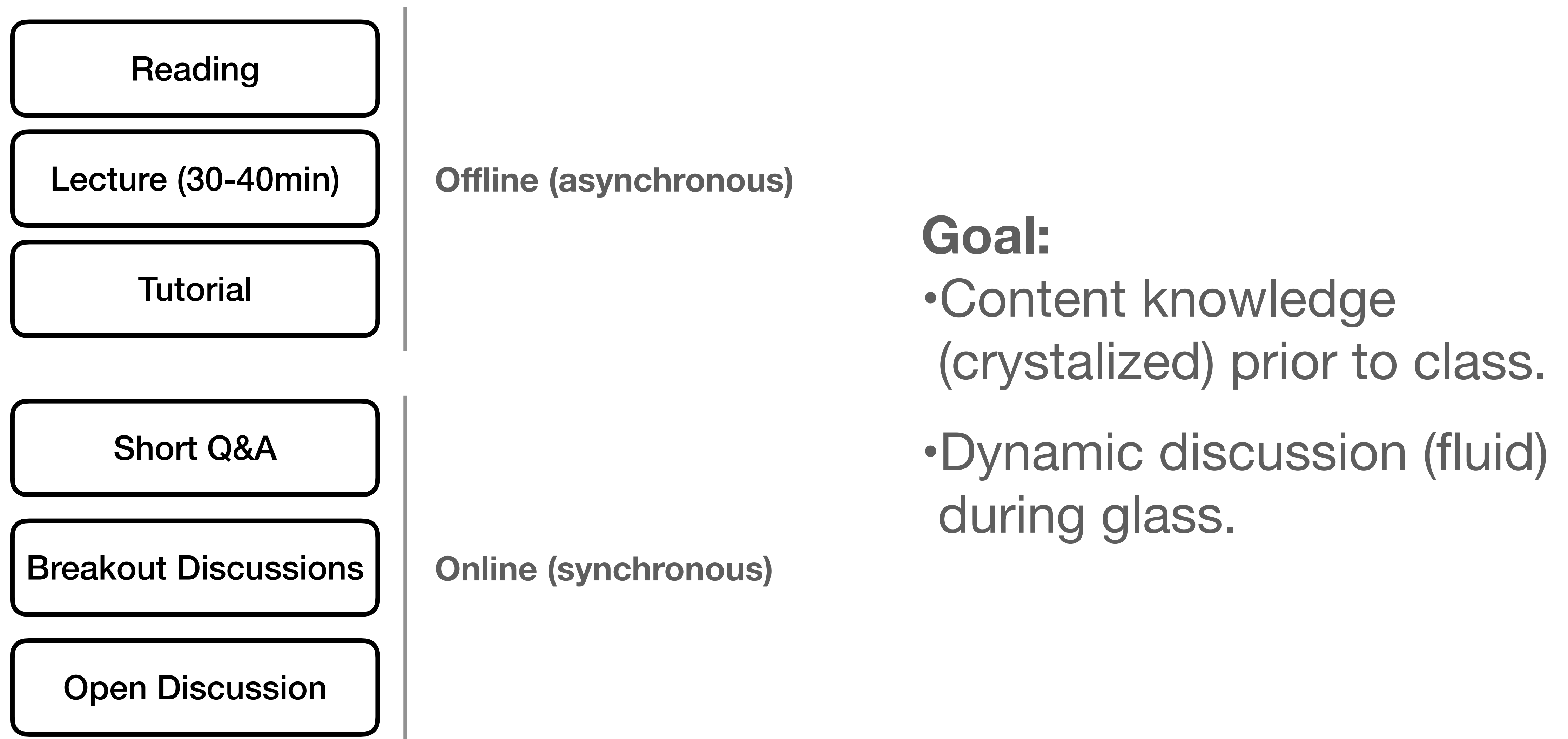
# Learning objectives for this class

1. Understand basic principles of statistical theory, measurement, and experimental design;
2. Be able to clean and organize data efficiently;
3. Be well versed the execution and interpretation of data analysis;
4. Use information resources to find appropriate data science tools;
5. Communicate statistical results effectively in multiple modalities;
6. Be a critical consumer of data science techniques and their application in empirical research.

# Prior knowledge

1. Introductory level understanding of probability theory and statistics (CMU 36-309, 86-309, or equivalent)
2. Basic familiarity with R or similar functional data analysis languages.

# Class structure



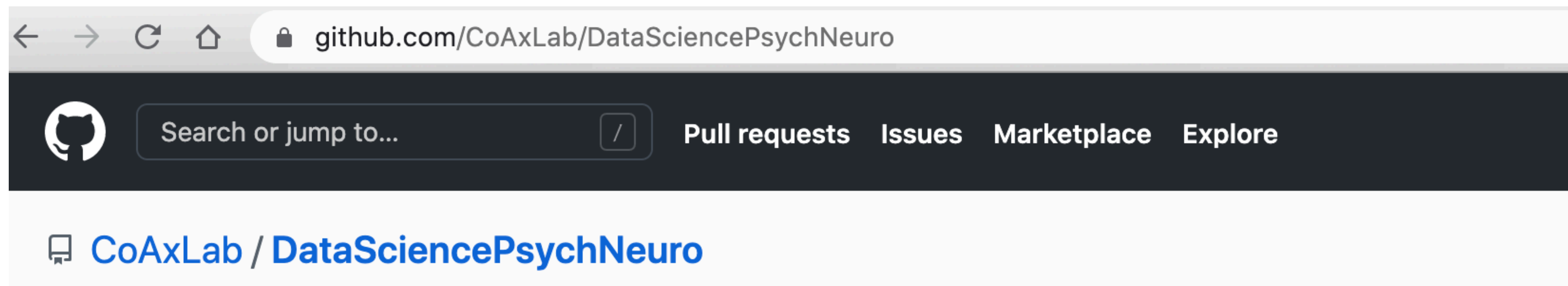


# Resources


## 1. Texts:

- Textbook: James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R (Vol. 6). New York: springer. (<http://www-bcf.usc.edu/~gareth/ISL/>).
- Supplemental book: Hadley Wickham & Garrett Grolemund (2016). R for Data Science. O'Reilly (<https://r4ds.had.co.nz/>).
- Auxiliary readings will be posted on Canvas/Github for class sections covering material not in the main textbook.

## 2. Github Repository: <https://github.com/CoAxLab/DataSciencePsychNeuro>



# Resources

 [CoAxLab](#) / [DataSciencePsychNeuro](#)

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

---

## Home

Cognitive Axon Laboratory edited this page on Mar 30 · 41 revisions

---

This Wiki will serve as a useful resource for tools associated with class lectures and labs. Keep checking as it is updated.

## Required tools for the course

---

[R \(version 3.4.3\)](#)  
[Jupyter Notebooks](#)  
[A Github Account](#)

## Useful support tools

---

[RStudio](#): A better interactive environment than the R console  
[GitKraken](#): A GUI interface for Github.  
[Jupyter Themes](#): Spruce up your notebook style. Pure aesthetics.  
[Latex Math Symbols Cheatsheet](#)

## Useful tutorials

---

**Getting R working with Jupyter Notebooks**