

# DSC 205: Introduction to Advanced Data Science

Dr. Marcus Birkenkrahe

---

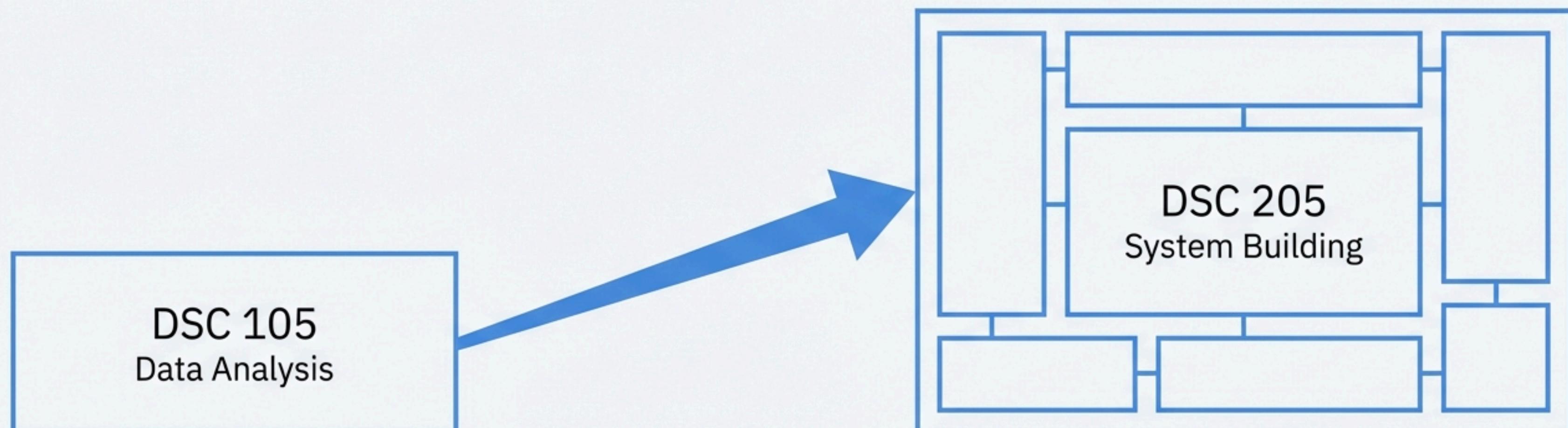
Spring 2026

## Course Logistics

- **Meetings:** Monday, Wednesday, Friday | 14:00–14:50
- **Location:** Derby Science Center, Room 209

# The Journey Continues: Beyond Foundational R

This course builds directly upon the foundation established in DSC 105. We transition from fundamental data analysis to the discipline of professional statistical programming. The focus shifts from simply using R to mastering its functional and object-oriented capabilities to build robust, efficient, and scalable data science solutions.



Functional Programming

Object-Oriented Principles

Statistical Computing Infrastructure

# From Apprentice to Architect

## The Analogy

“If DSC 105 was about learning to read and write sentences in the language of data...”

“...then DSC 205 is about becoming a software architect. You will learn to build the library (the infrastructure), automate the printing press (the Shell), and ensure the delivery trucks are running at maximum speed (C++ optimization).”

## Core Principles

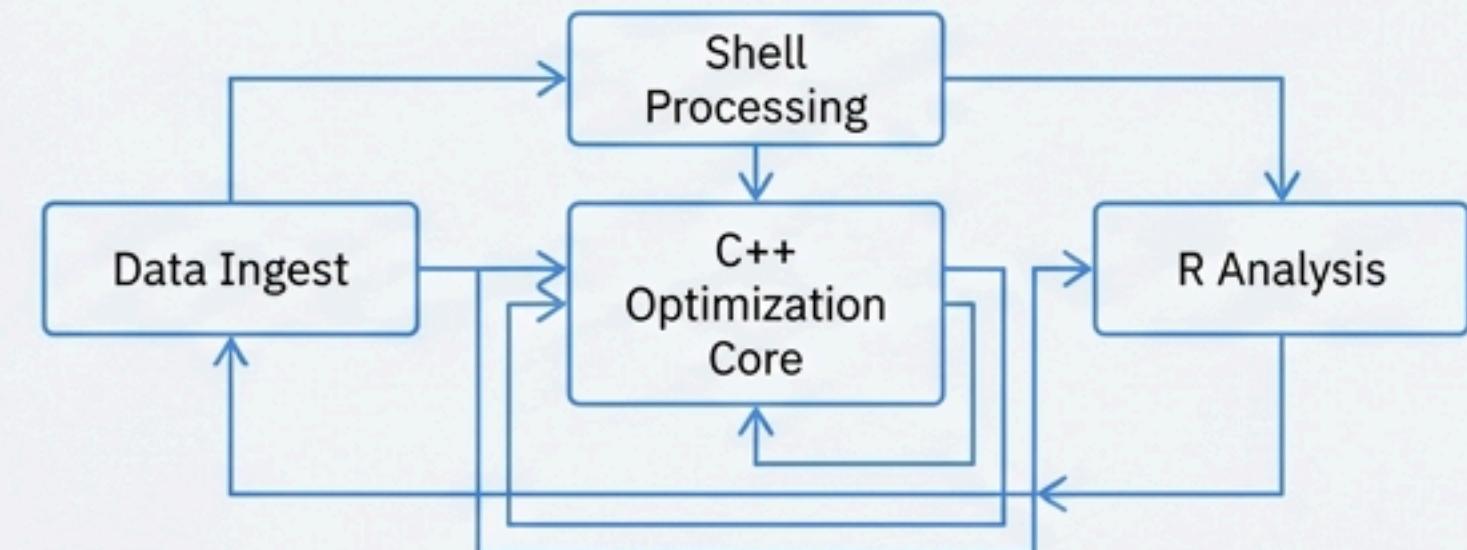
⟳ **Reproducibility:** Ensuring your work is a clear, reliable blueprint that can be understood and replicated by others.

⚡ **Optimization:** Engineering your solutions for maximum performance and efficiency.

## The Sentence

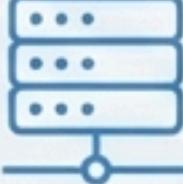
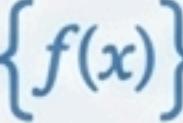
```
data <- read.csv("data.csv")
summary(data)
```

## The System



# Your Learning Objectives: A Blueprint for Mastery

Upon completion of DSC 205, you will be able to:

-  Master the infrastructure for advanced statistical computing, from data import to transformation.
-  Develop advanced data visualizations as part of rigorous exploratory data analysis.
-  Understand and implement basic predictive modeling strategies.
-  Process and pipeline data directly from the command-line (UNIX).
-  Design and write efficient, well-documented, and robust functions in R.
-  Achieve significant performance gains through data science **optimization** with C++.

# The Four Pillars of the Curriculum

## 1.

### Advanced R Programming

Moving beyond basic scripts to control flow, iteration, and sophisticated functional patterns.

```
results <-  
lapply(data,  
summary_function)
```

## 2.

### Data Processing in the Shell

Using powerful UNIX tools to download, clean, transform, and pipe data outside of the R environment.

```
cat data.log | grep  
"ERROR" | awk '{print  
$5}'
```

## 3.

### Writing Robust Functions

Designing reusable, well-documented, and validated functions—the building blocks of any serious project.

```
calculate_risk <-  
function(age,  
pressure) {  
...  
}
```

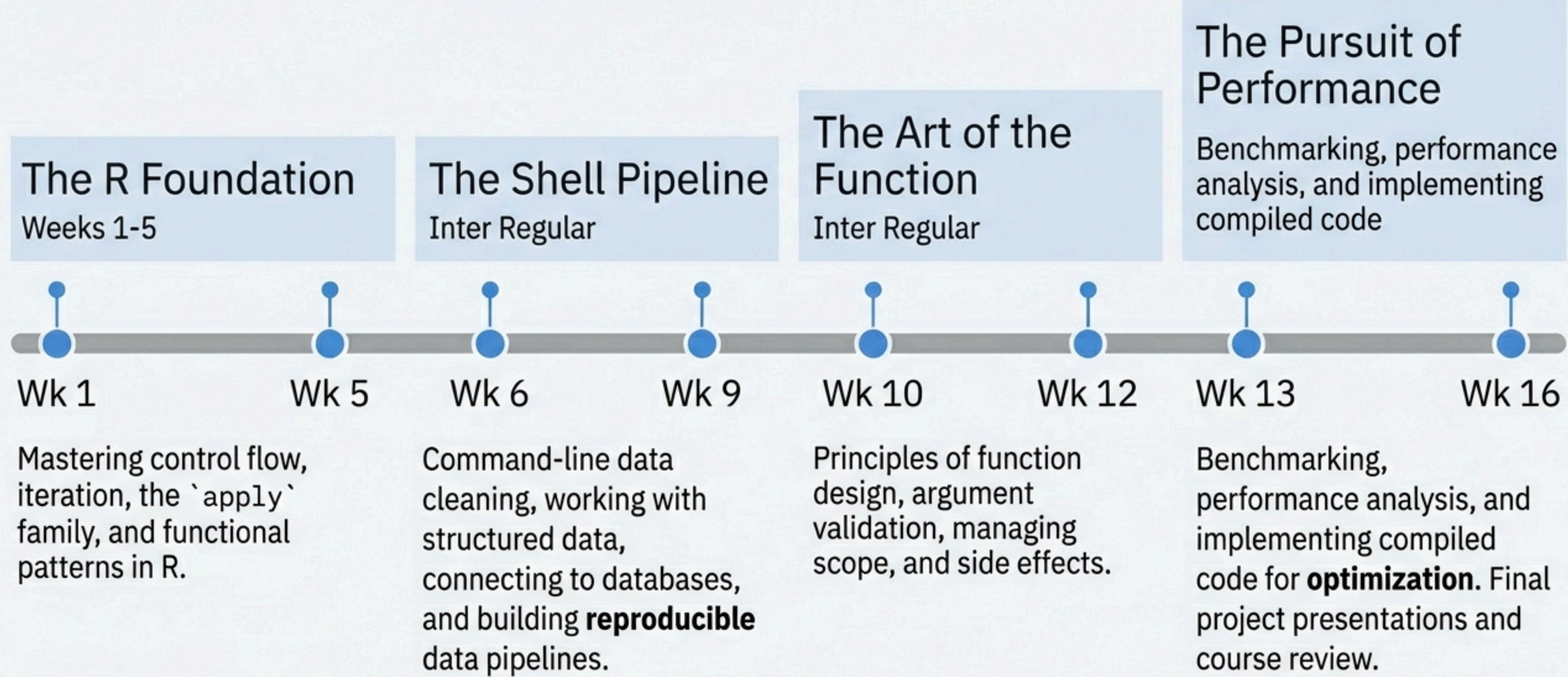
## 4.

### Optimization with C++ (Rcpp)

Identifying performance bottlenecks and accelerating R code by interfacing with high-performance compiled extensions.

```
// [[Rcpp::export]]  
NumericVector  
fast_sum(NumericVect  
or x) { ...  
}
```

# The 16-Week Roadmap to Mastery



# Your Advanced Technical Toolset

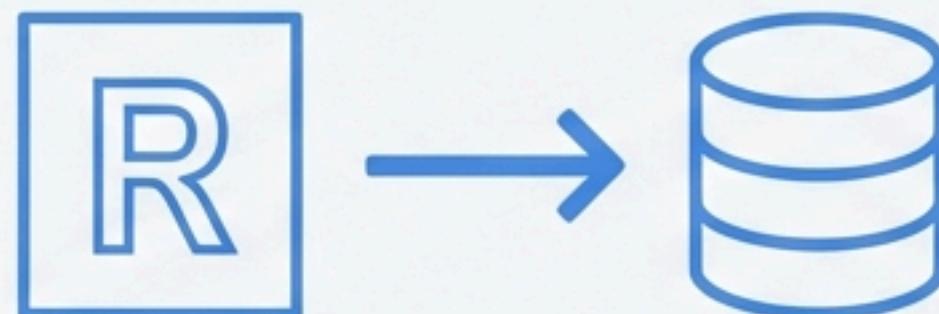
## UNIX Command-Line

Learn to leverage a suite of powerful tools (`grep`, `awk`, `sed`, etc.) for fast and efficient data manipulation and building automated pipelines.

```
$ cat data.csv | grep "keyword" | wc -l  
42
```

## RSQLite

Integrate and query databases directly from your R environment, mastering the fundamentals of structured data storage.



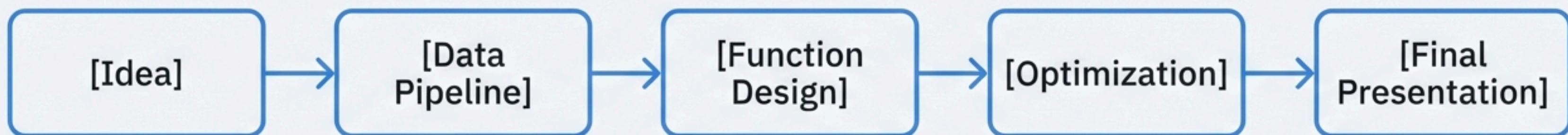
## Rcpp

Bridge the gap between R and C++ to accelerate computationally intensive tasks, turning R into a high-performance computing environment.

```
#include <Rcpp.h>  
using namespace Rcpp;  
  
// [[Rcpp::export]]  
double meanC(NumericVector x) {  
    int n = x.size();  
    double total = 0;  
    for(int i = 0; i < n; ++i) {  
        total += x[i];  
    }  
    return total / n;  
}
```

# The Term Project: Your Architectural Showcase

This is a cumulative project, undertaken individually or in a team, that serves as the capstone for the course.



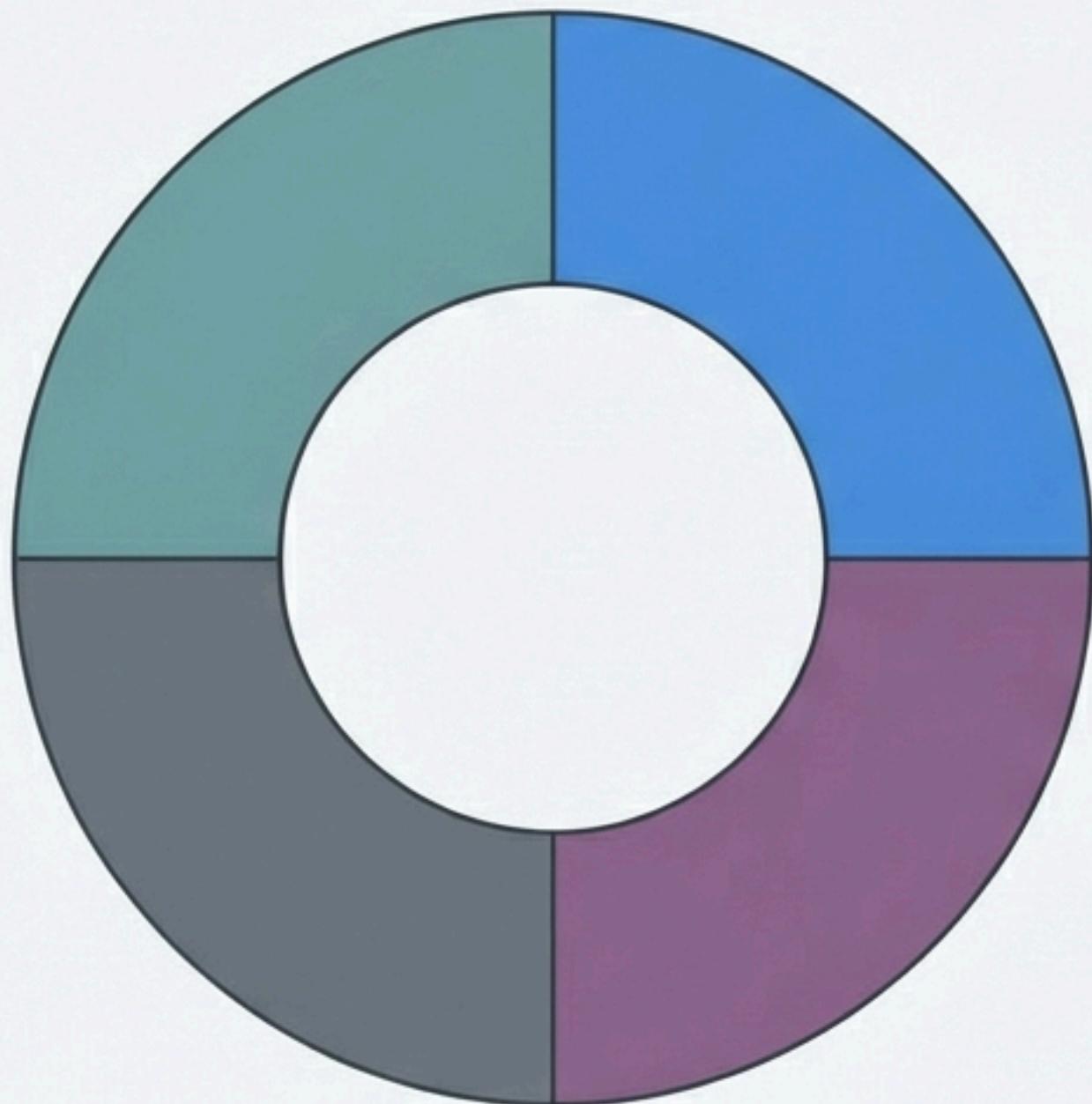
## Core Focus

The project is evaluated on three key dimensions of professional data science work:

-  **Programming Discipline:** Clean, documented, and well-structured code.
-  **Reproducibility:** A complete, automated workflow that can be run and verified by others.
-  **Performance:** Justified use of **optimization** techniques to create an efficient solution.

**Process:** Your progress will be evaluated via monthly project sprint reviews.

# Grading & Assessment: Measuring Your Progress



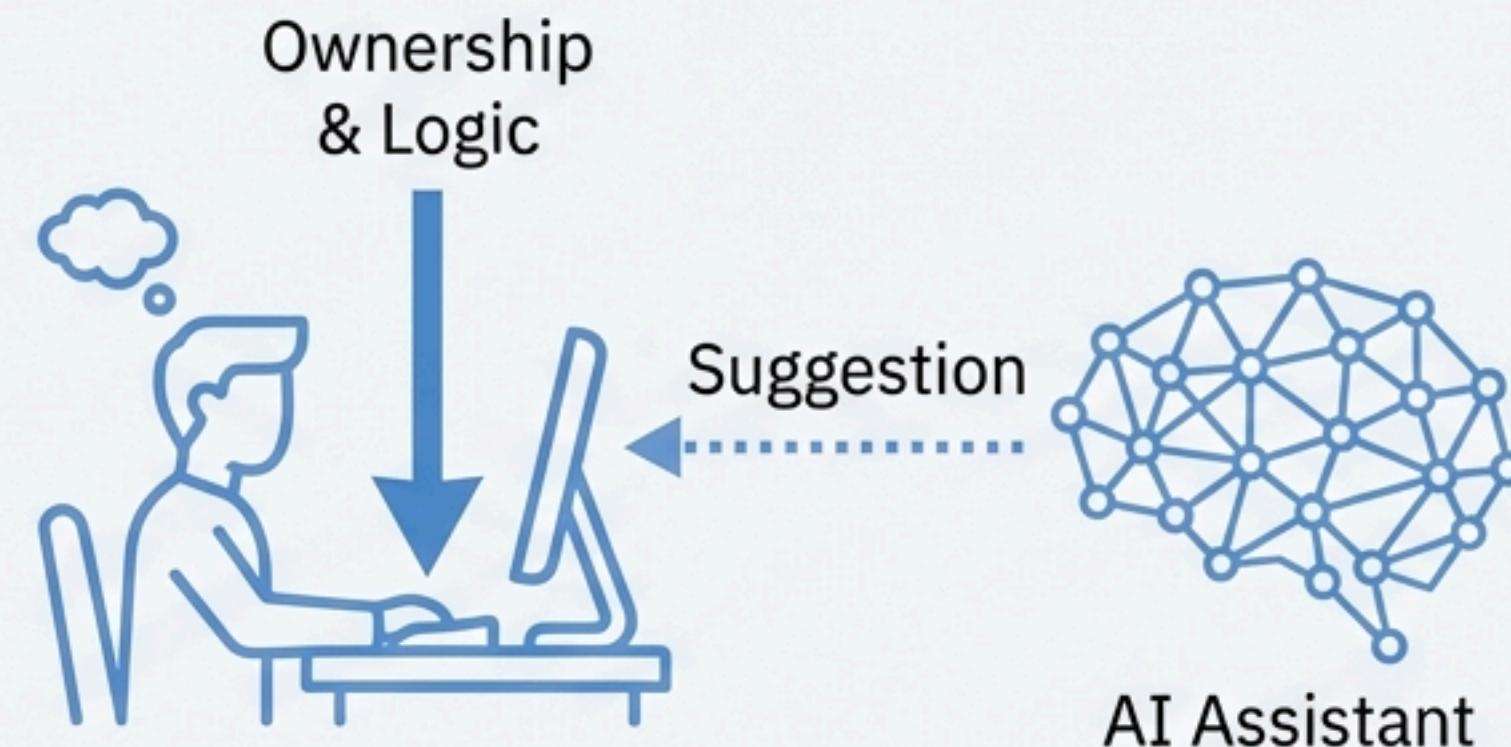
- 25%: Weekly Programming Assignments
- 25%: Weekly Multiple Choice Tests
- 25%: Monthly Project Sprint Reviews
- 25%: Optional Final (Oral) Exam

## Key Details

- **Passing Grade:** A minimum of 60% of all available points is required to pass.
- **Optional Final Exam:** The final is an oral exam based on a random selection of weekly test questions. It is only required if you wish to improve your grade. If the exam score is lower than your current average, it will be ignored.

# AI-Assisted Programming: A Tool, Not a Crutch

**Permitted Use:**  
You may use AI tools for idea generation, code drafting, and exploring alternative approaches.



**Prohibited Use:** Use of AI that obscures authorship, bypasses conceptual understanding, or violates the Lyon College Honor Code is not permitted.

**“Students are responsible for the correctness, interpretation, and justification of all submitted work and be able to explain and defend their analyses and code.”**

**Guiding Principle:** Use AI to augment your intellect, not to replace it.

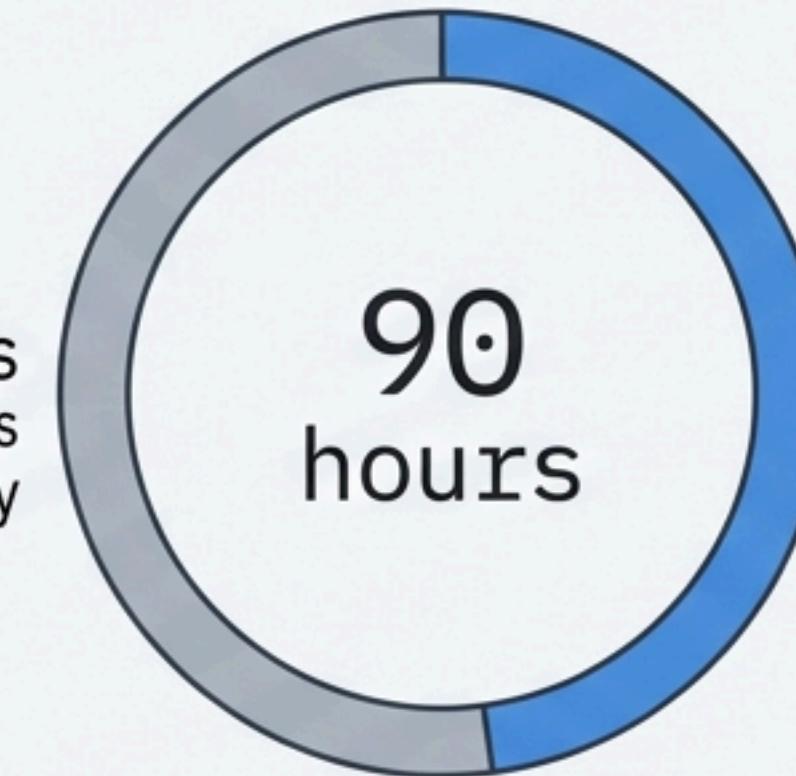
# Expectations for Success

## Prerequisites

- Successful completion of DSC 105 ('Introduction to Data Science').
- *Or* equivalent knowledge demonstrated by completing chapters 1-15 of 'fasteR: Fast Lane to Learning R!' or Part I of Davies' 'The Book of R'.

## Estimated Term Workload

~42 hours  
Assignments & Study



48 hours  
In-class time

This represents a weekly workload of approximately 5-6 hours outside of class, with peaks around project sprint reviews.

This course rewards discipline, curiosity, and a commitment to building robust, high-performance systems.