



# STOR 320 Programming

|

Lecture 12

Yao Li

Department of Statistics and Operations Research

UNC Chapel Hill



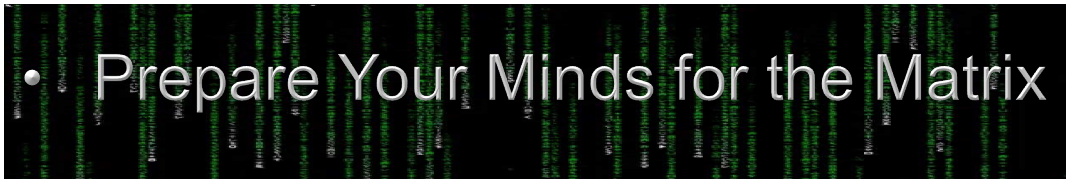
# Introduction

- Reading
  - Chapters 19-21 in R4DS
  - Chapters 14-18 in RP4DS
  - Chapter 7 in AoRP
  - Chapter 4 in FCSPR
- Programming Steps
  - Understand the Problem
  - Inputs and Outputs
  - Create Code
  - Test the Code (Simple Case)
  - Generalize the Code
  - Test Problematic Cases
  - Edit Code to Handle Issues
  - Consider Efficiency



# Setup for Lecture

- Open Tutorial 9
- Packages Required:
  - Tidyverse
  - Ecdat
- Knit Document As You Go
- Read Introduction
- Prepare Your Minds for the Matrix





# Part 1: If Else

- General Construction:

- “If”

```
if (CONDITION) {  
    ACTION  
}
```

- “If-Else”

```
if (CONDITION) {  
    ACTION 1  
} else {  
    ACTION 2  
}
```

- ifelse()

```
ifelse(CONDITION,ACTION1,ACTION2)
```



# Part 1: If Else

- Run Chunk 1
  - Check if Larger than 0
  - If True, Take Log
  - Result When  $x = 3$ ?
  - Result When  $x = -3$ ?
- Run Chunk 2
  - Notice the Difference
  - If-Else to Handle Errors
- Run Chunk 3
  - Situation Not Considered
  - Replace *BLANK* to Lead to Potential Problem



# Part 1: If Else

- Run Chunk 4
  - Replace BLANK with Different Options and Check
  - How Would You Explain this Code to Your Granny?
- Run Chunk 5
  - What is the Difference Between  $y_1$  and  $y_2$ ?
  - Always Look for a Vectorized Solution for Efficiency
- Run Chunk 6
  - Nested `ifelse()` Statements
  - How Would You Explain this to your Mother?



# Part 2: Loops

- General Construction

- “for” Loop

```
for (INDEX in VECTOR) {  
    ACTION FOR EACH INDEX  
}
```

- “while” Loop

```
while (CONDITION) {  
    ACTION UNTIL CONDITION = FALSE  
}
```

- Nested “for” Loops

```
for (INDEX1 in VECTOR1) {  
    for (INDEX2 in VECTOR2) {  
        ACTION  
    }  
}
```



# Part 2: Loops

- Mental Process
  - I Want to Do \_\_\_\_\_  
for Every \_\_\_\_\_  
until \_\_\_\_\_
  - What Type of Object Do You Want Returned?
  - Initiate a Starting Point Based on the Desired Output
  - Try R Code on Single Instance
  - Create the Loop





# Part 2: Loops

- Geometric Series

$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}, \text{ for } |r| < 1$$

- Run Chunk 1
  - What  $a$  did you choose?
  - What  $r$  did you choose?
  - What is the theoretical limit?
  - What pattern exists?
- Run Chunk 2
  - Choose  $a$  and  $r$  that work?
  - Choose  $a$  and  $r$  that don't work?
  - Modify: `if(k>100) break`



# Part 2: Loops

- Geometric Series (Cont.)

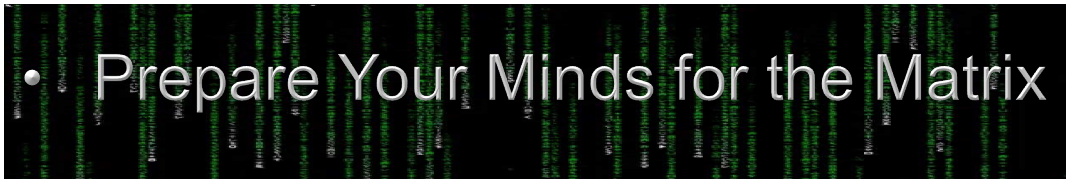
$$\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}, \text{ for } |r| < 1$$

- Run Chunk 3
  - Suppose We Want to Save at Every Step
  - Why? Picture to Examine the Path of the Summation
  - Choose Small  $K < 15$
  - Choose Large  $K > 50$
  - What do You Observe?
  - How Would You Explain This Code to Your Stranged Brother?



# Setup for Lecture

- Open Tutorial 10
- Packages Required:
  - Tidyverse
  - Ecdat
- Knit Document As You Go
- Read Introduction
- Prepare Your Minds for the Matrix





# Part 1: Loops

- Correlation Matrix
  - Definition: Matrix Which Shows the Correlation Between Every Pair of Numeric Variables
  - Used to Understand Strength of Linear Relationships Between Numeric Variables
  - Helpful in Measuring Collinearity
- Run Chunk 1
  - Inspect the Variables in Cigar
  - Inspect the Correlation Matrix
  - Which Variable(s) is Inappropriate for a Correlation Analysis? Why?



# Part 1: Loops

- Run Chunk 2
  - Run First Half – Loops through Every Combination of Columns and Computes Correlation
  - Examine Second Half – Loops Through Every Combination of Columns Excluding the First Column
  - Fill in Blanks with Appropriate Indices so Second Loop Works
  - Run Second Half
- Run Chunk 3
  - Inspect the Variables in H1
  - Uncomment to Print Correlation Matrix
  - What is the Problem?



# Part 1: Loops

- Run Chunk 4
  - Observe the Difference Between the Printed Tibbles
  - What is the Difference?
  - How Would You Explain the First Loop to a Toddler?
  - What is `cat()` doing?
  - How Would You Explain the Second Loop to an Infant?
  - Remember: There Are an Infinite Number of Ways to Do the Same Thing.



# Part 2: SRS

- Important For Simulation Studies
- Known Distributions

Distribution	Density/pmf	cdf	Quantiles	Random Numbers
Normal	<code>dnorm()</code>	<code>pnorm()</code>	<code>qnorm()</code>	<code>rnorm()</code>
Chi square	<code>dchisq()</code>	<code>pchisq()</code>	<code>qchisq()</code>	<code>rchisq()</code>
Binomial	<code>dbinom()</code>	<code>pbinom()</code>	<code>qbinom()</code>	<code>rbinom()</code>

- “d” -> Useful for Plotting Density Curve for Continuous Variables or Probability Mass Function for Discrete Variables
- “p” -> Finds the Probability Less Than Or Equal to a Given Number
- “q” -> Finds Cutoff Points
- “r” -> Generates a Random Sample from the Distribution



# Part 2: SRS

- For SRS, Use “r”
- Run Chunk 1
  - Scenario for  $x_1$ : You Ask BLANK Number of Students Their Grades where Grades Follow a Normal Distribution with Mean=82 and SD=2
  - Scenario for  $x_2$ : You Ask BLANK Number of Students to Roll a Fair Die 10 Times and Tell You the Number of 6's that Appeared.





# Part 2: SRS

- Sampling From Finite Set of Possible Outcomes
- Run Chunk 2
  - Scenario: Flip  $k$  Coins
    - $P(\text{Heads}) = \text{BLANK}$
    - $P(\text{Tails}) = 1 - \text{BLANK}$
  - How would You Explain What the Figure is Showing to a Politician?