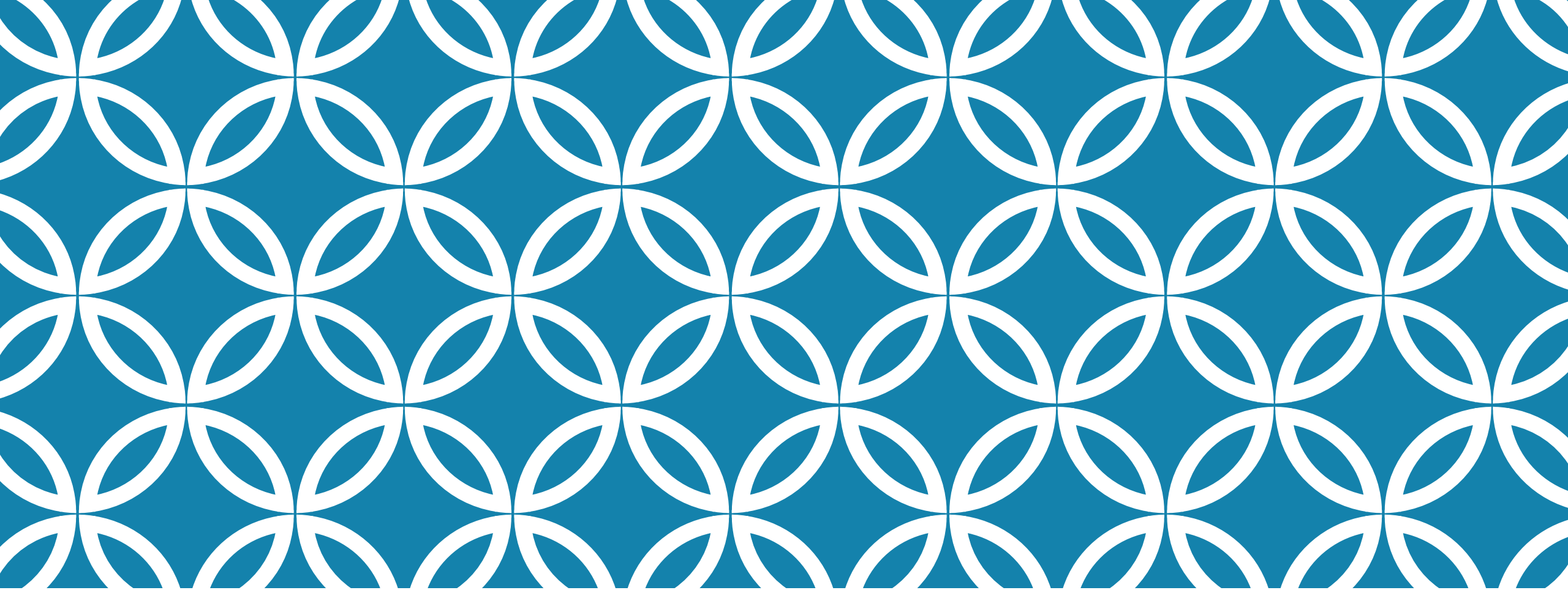


LING 0700 LAB: WEEK 1

TA: Wesley Mark Lincoln

WELCOME!

- Introductions
- Course admin
- Lab practice

INTRODUCTION

- **Wesley** Mark Lincoln
- Singaporean
- Native languages: English, Singlish, Mandarin, Hokkien
- Other languages: French, Spanish, Korean, Thai…

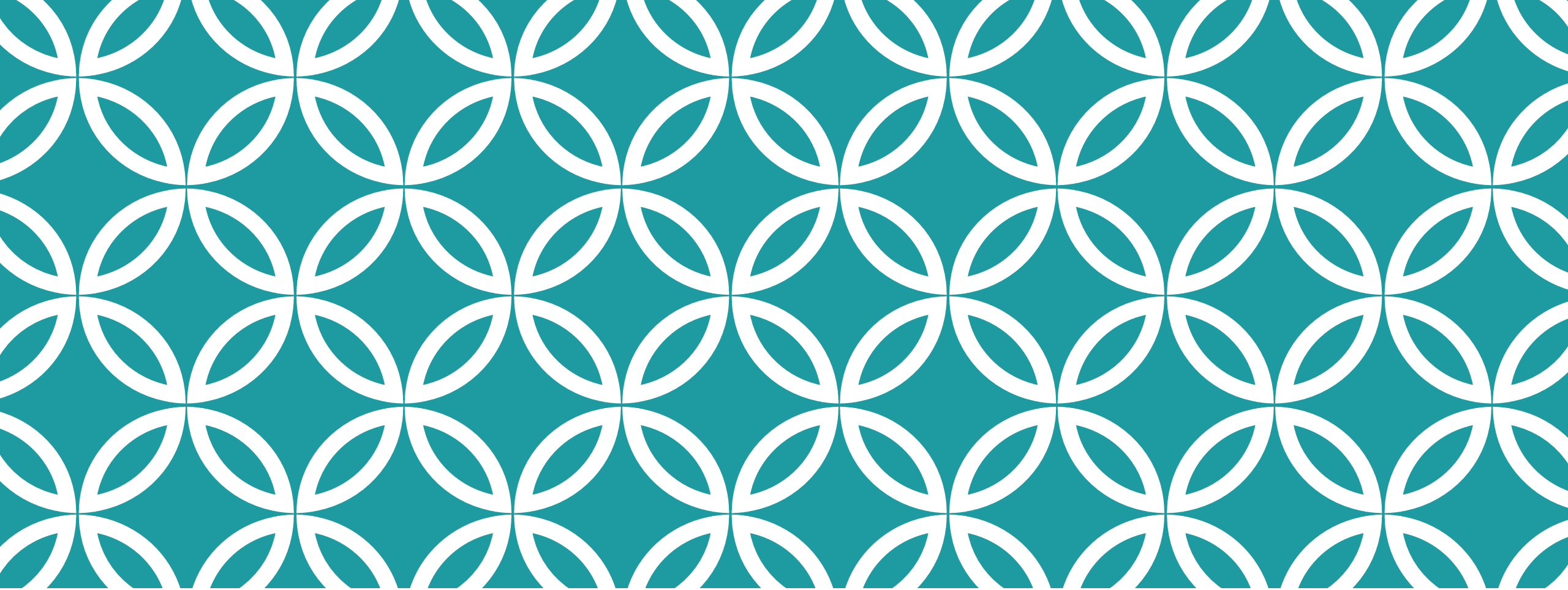
INTRODUCTION

- National University of Singapore (2018–2023)
  - Life Sciences (specialised in Evolutionary Biology)
  - English Language and Linguistics

- Current second-year PhD student in Linguistics
  - Evolutionary linguistics
  - Sociolinguistic variation in sound *and* structure

YOUR TURN! PLEASE TELL US…

- Your name
- Your major(s), minor(s), and year
- Languages you know or like
- Anything else you'd like us to know



COURSE ADMIN

COURSE STRUCTURE

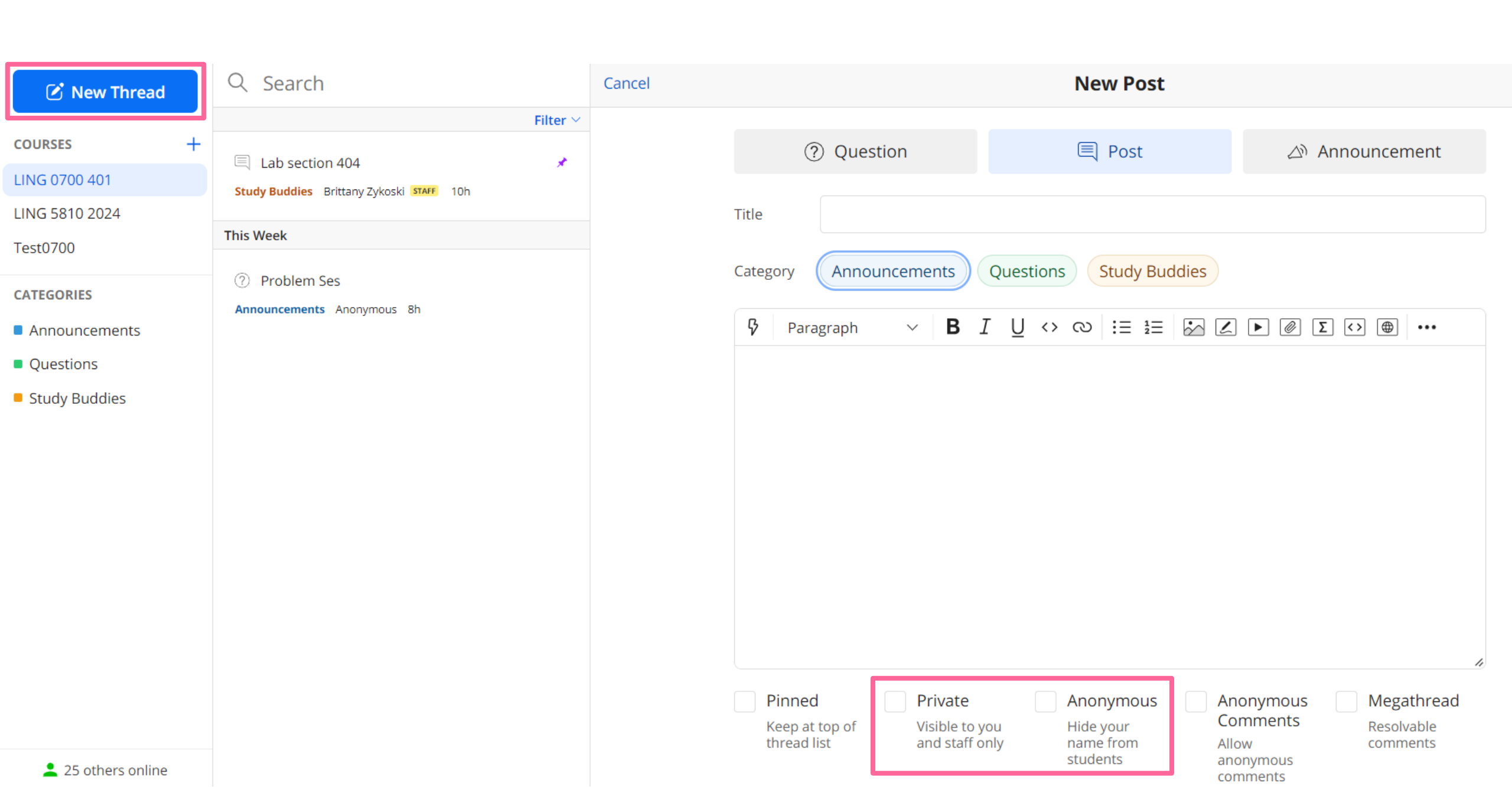
…	M	T	W	R	F	weekend	M	T	W	R	F	weekend	…
	PS up	Lecture		Lecture Labs	Labs		PS due	Lecture		Lecture Labs PS key up	Labs		

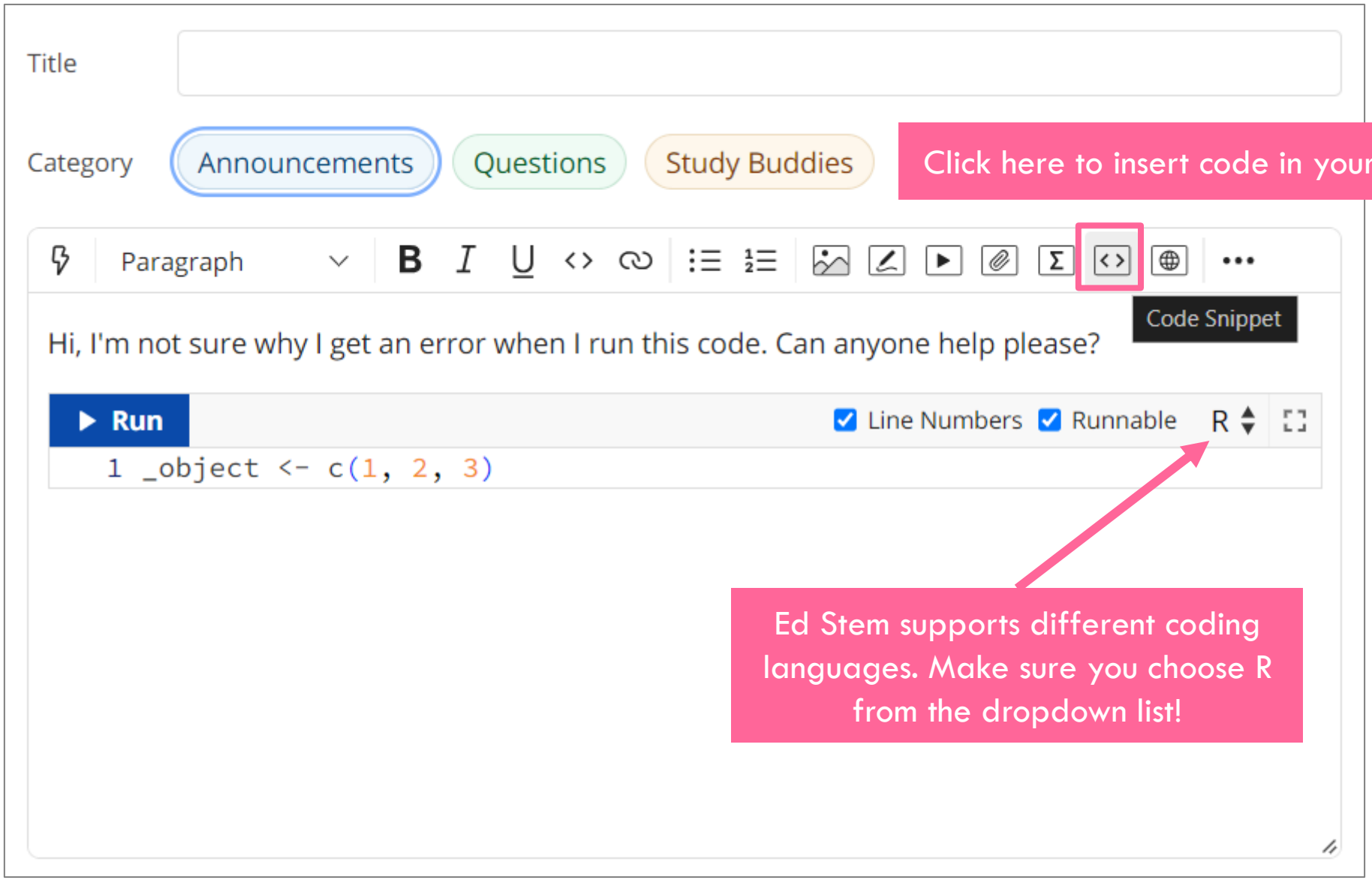
Specific dates are all listed at the end of the syllabus:

https://kathrynschuler.com/datasci/

ASKING CONTENT QUESTIONS

- Ed Discussion is the go-to platform for this
  - Written, centralised format
  - Option to post anonymously
  - Others may have asked the same question(s) you have (and may have already gotten an answer)

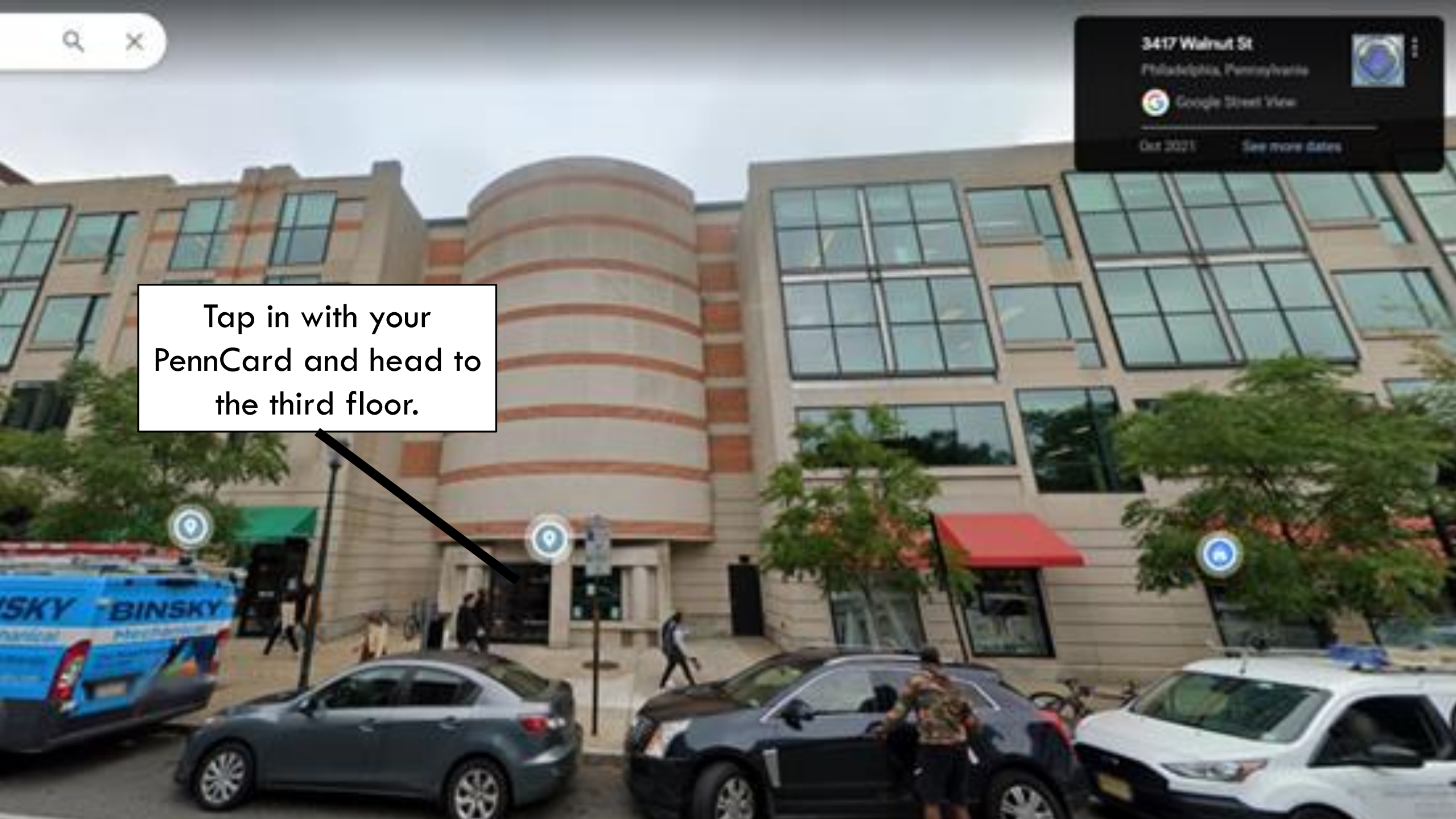

New Thread
Search
Cancel
New Post
COURSES
LING 0700 401
LING 5810 2024
Test0700
CATEGORIES
Announcements
Questions
Study Buddies
25 others online
Filter
Lab section 404
Study Buddies Brittany Zykoski STAFF 10h
This Week
Problem Ses
Announcements Anonymous 8h
Question
Post
Announcement
Title
Category
Announcements
Questions
Study Buddies
Paragraph
Pinned
Keep at top of thread list
Private
Visible to you and staff only
Anonymous
Hide your name from students
Anonymous Comments
Allow anonymous comments
Megathread
Resolvable comments


Title
Category
Announcements
Questions
Study Buddies
Click here to insert code in your post
Paragraph
Code Snippet
Hi, I'm not sure why I get an error when I run this code. Can anyone help please?
Run
Line Numbers
Runnable
R
1 _object <- c(1, 2, 3)
Ed Stem supports different coding languages. Make sure you choose R from the dropdown list!

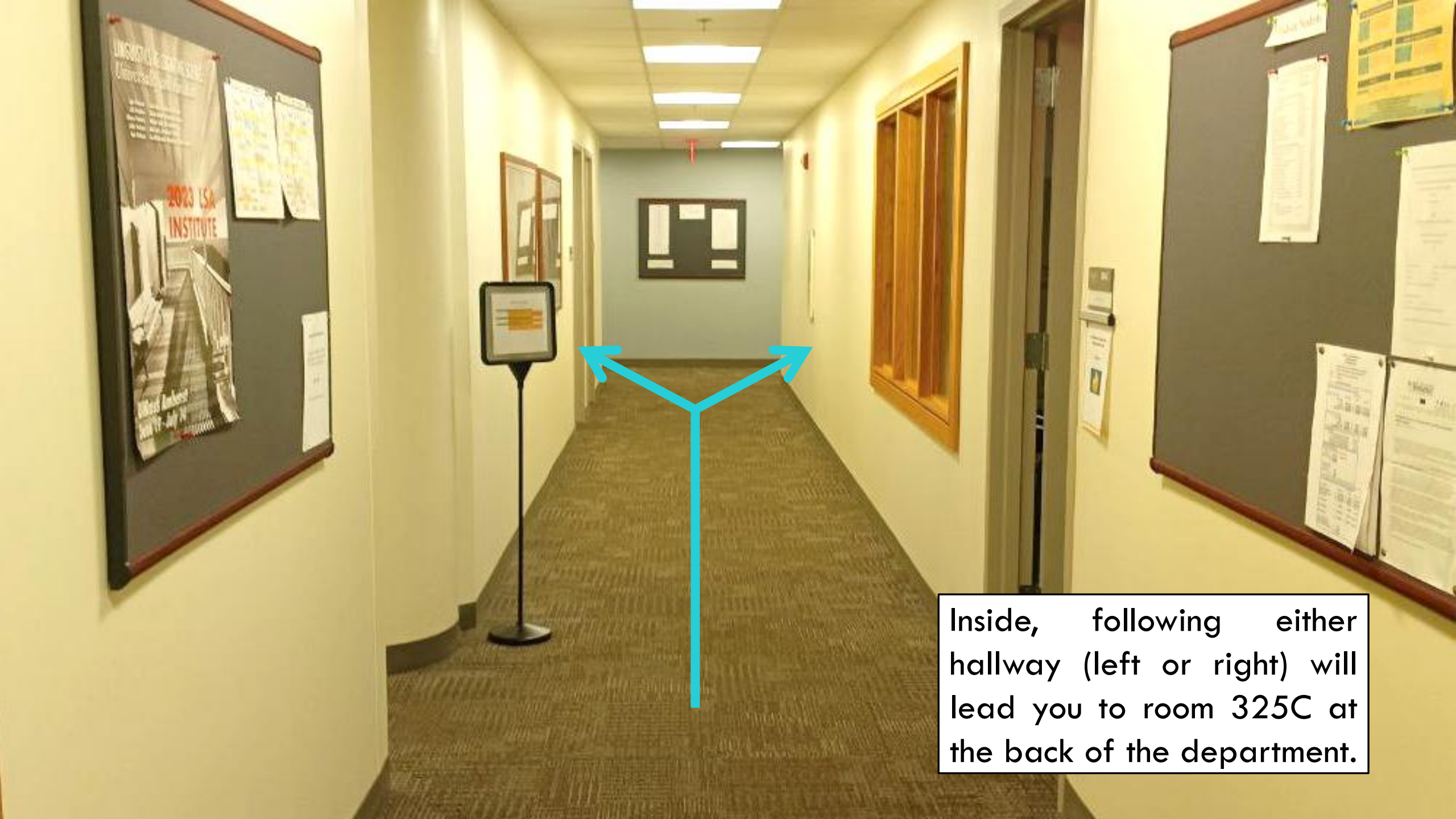
ASKING CONTENT QUESTIONS

Office hours:

- Mondays from 2pm–3pm
- Department of Linguistics, Room 325C
  3401C Walnut Street
  Suite 300C
- Please come prepared with specific questions!


3417 Walnut St
Google Street View
Oct 2021
See more dates
Tap in with your PennCard and head to the third floor.


Enter the Department of Linguistics here.



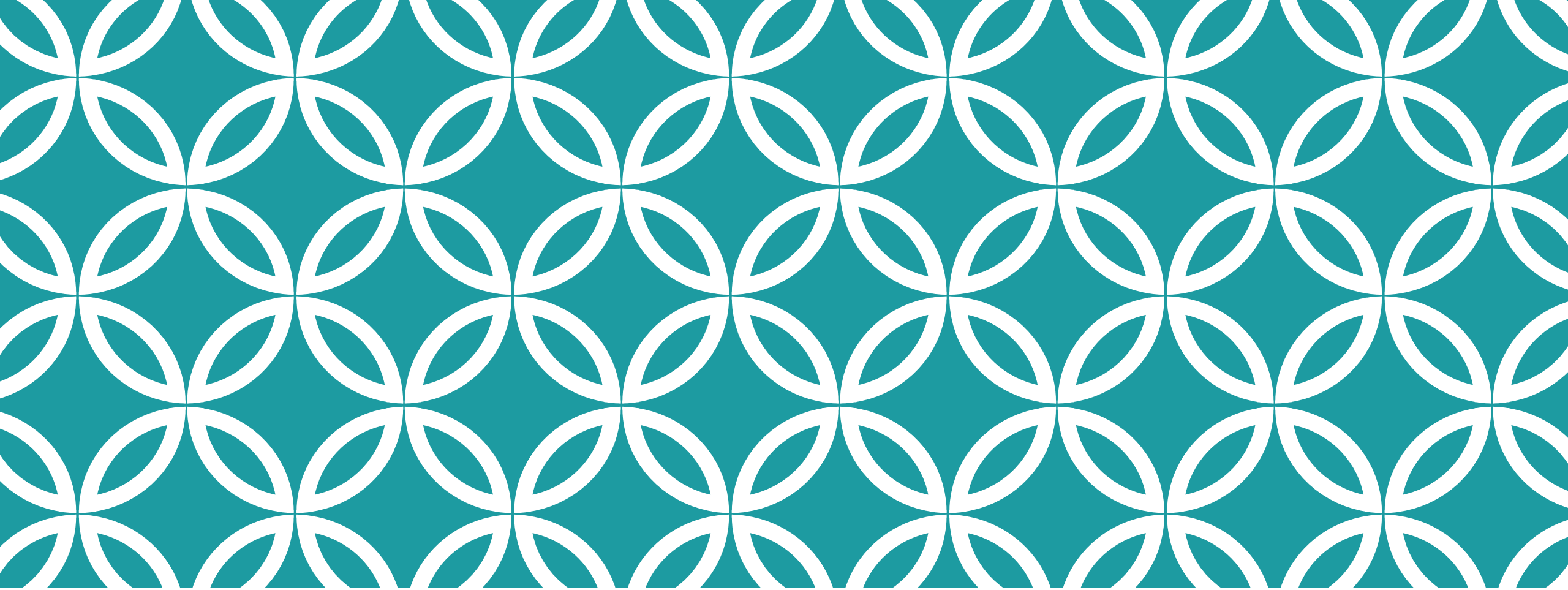
Inside, following either hallway (left or right) will lead you to room 325C at the back of the department.


325C

CONTACTING ME

Email: wlincoln@sas.upenn.edu

- Non-content questions (e.g., admin, illness, absence) you would prefer not to ask publicly on Ed
- Include “LING 0700” in the subject header
- I will do my best to reply within 24 hours, prioritising urgent matters



LAB EXERCISES

GOOGLE COLAB

R is a programming language.

[Python, Julia, MATLAB, C++, JavaScript…]

Google Colab is a development environment where you can run R.

[RStudio, Visual Studio Code, Jupyter Notebook…]

R BASICS: VALID VARIABLE NAMES

`childAge`	OK!
`response_time`	OK!
`1stPlaceWinner`	NO! Cannot start with numeral
`2fast2furious`	NO! Cannot start with numeral
`pi`	NO! Protected name

R BASICS: VALID VARIABLE NAMES

`.15`	NO! Initial <.> cannot be followed by numeral
`.object`	OK!
`260`	NO! Cannot start with numeral
`15n`	NO! Cannot start with numeral
`n15`	OK!
`_hi`	NO! Cannot start with underscore

R BASICS

Which of the following occur in the code block below?

```
# compute the mean of x and y
mean(c(x,y))
```

Comment

- ❌ a message
- a function
- a comment
- an expression

Functions: `mean()`, `c()`

Expressions: `mean(c(x,y))`, `c(x,y)`, `x`, `y`

VECTORS

Suppose we construct a vector with `c(1, "two", 3, 4, 5, 6)` and assign it to `x`. What will the following code block return?

```
typeof(x)
```

character ✓

What is the previous question an example of?

- attribute addition
- explicit coercion
- ✓ implicit coercion
- none of the above

The item "two" is a character.
Since vectors are atomic and can only have one type of data, all other items are **silently** changed to character, i.e.: c("1", "two", "3", "4", "5", "6"). This is termed **implicit coercion**.

VECTORS

Reminder: implicit coercion follows a given hierarchy:

character > double > integer > logical

If a mix of types are fed into a vector, R coerces the type to the highest in the hierarchy. For example, if you have doubles and logicals, R coerces the vector to double. If you have characters, doubles, and integers, R coerces it to character.

VECTORS

What will the following code block return?

```
x <- 1:4
y <- matrix(x, ncol=2, nrow=2)
typeof(y)
```

integer ✅

- The `:` operator creates a range of **integers** – so `x` is of type `integer`
- `y` inherits the data type `integer`
- In other situations, you specify that you want an integer using `L`, e.g.:

```
x <- c(1L, 2L, 3L)
```

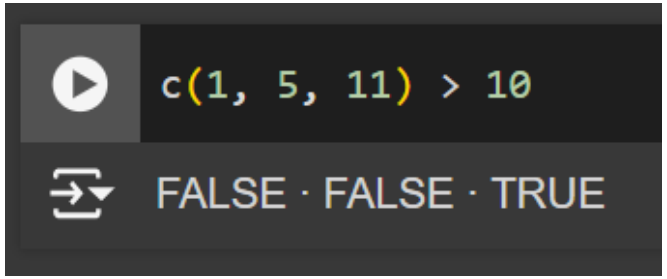
VECTORS

Suppose we run the following code. What will `any(x)` return?

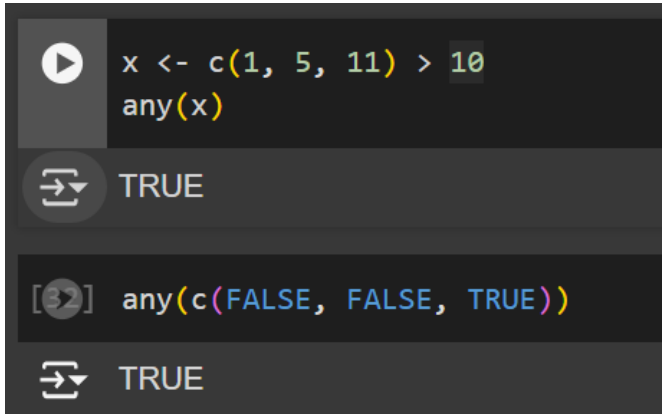
```
x <- c(1, 5, 11) > 10
```

- TRUE

Operations like `>` are vectorised in R, so they apply to each position of a vector. Let's break this code down.



Here, R goes through each position of the vector and checks whether the value is >10. If true, it returns `TRUE` in that position, and if false, it returns `FALSE` in that position.



Therefore, these two code chunks are equivalent. `any()` returns `TRUE` if any position in its input is `TRUE`, and false if no position in in its input is `TRUE`.

SUBSETTING

Suppose we run the following code. What will `m[1, 2]` return?

```
m <- matrix(c(10,20,30,40), nrow=2, ncol=2)
```

```
[6]  c(10, 20, 30, 40)
     10 · 20 · 30 · 40

[7]  m <- matrix(c(10,20,30,40), nrow=2, ncol=2)
     m
     A matrix:
     2 × 2 of
     type dbl
     10  30
     20  40

     m[1,2]
     30
```

Creates a vector of type `double`

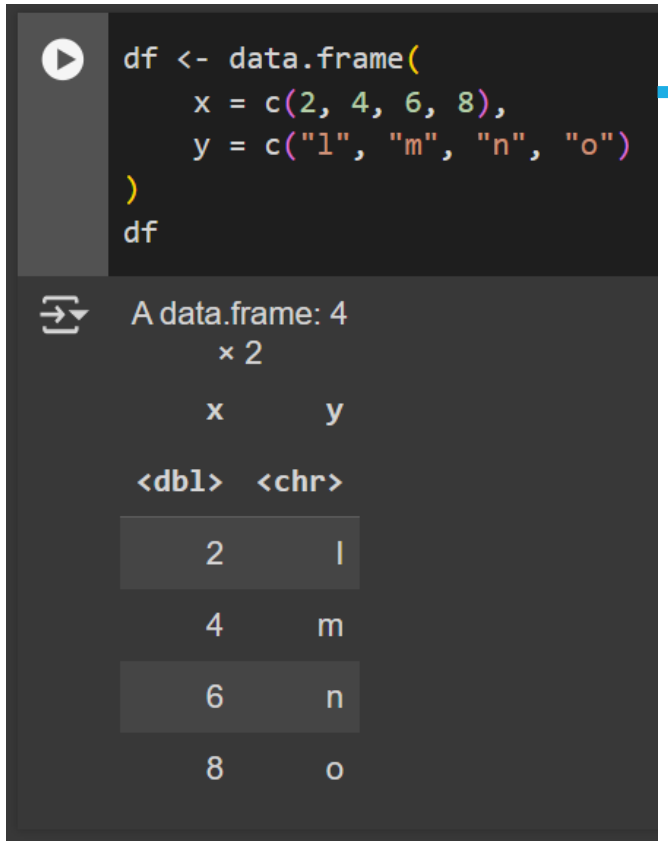
Uses the data in the vector to populate a 2×2 matrix `m`, filling columns from top to bottom, left to right

Returns the subset of data in column 1, row 2 (i.e., `30`)

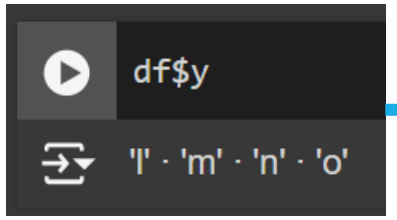
SUBSETTING

Suppose we run the following code. What will `df$y[4]` return?

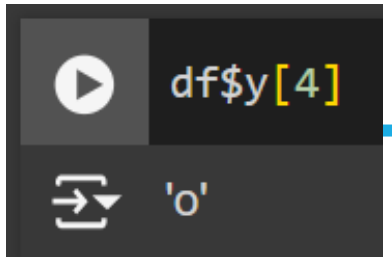
```
df <- data.frame(
    x = c(2, 4, 6, 8),
    y = c("l", "m", "n", "o")
)
```



1. Note the structure of `df`: each vector that we feed into `data.frame()`, i.e. `x` and `y`, becomes a column in `df`.



2. You can use $ and a column name to subset the data.



3. The output of df$y can be further subsetted by choosing only one index, i.e. 4.

SUBSETTING

```
[34] df <- data.frame(
         x = c(2, 4, 6, 8),
         y = c("l", "m", "n", "o")
     )

[43] # We saw that we can isolate (subset) a column using:
     df[2]
     df["y"]

[48] # We can also select a specific cell using df[r,c]
     # where r=row number, c=column number
     df[1,2]

[51] # To get all rows for a given column, use df[,c]
     df[,2]

     # To get all columns for a particular row, use df[r,]
     df[2,]
```