Project 2 Proposal

Load Packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
```

Load Course Data

```
course_data <- read_csv(here::here("data/course_catalog.csv"))
```

```
## Rows: 2703 Columns: 13

## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr  (9): Subject, Catalog, Descr, Section, Pat, Mode, Descr 1, Career, Term...
## dbl  (2): Unique Class Identifier, Cap Enrl
## time (2): Mtg Start, Mtg End

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
course_data <- course_data %>%
  rename(location = `Descr 1`,
         class_identifier = `Unique Class Identifier`,
         catalog_number = Catalog,
         enroll_cap =`Cap Enrl`,
         days = `Pat`,
         mtg_start = `Mtg Start`,
         mtg_end = `Mtg End`,
         term = `Term Descr`) %>%
  filter(location != "NA",
         Descr != "NA",
         !grepl('ML', location),
         !grepl('406 Oregon St 0114', location),
         !grepl('See Instructor/Department', location),
         !grepl('Thesis',  Descr),
         Descr != 'FIRST-YEAR SEMINAR (TOP)')
```

```r
course_data <- course_data %>%
  mutate(location = case_when(
    grepl('Classroom Building', location ) ~ 'Classroom Buiding',
    grepl('Allen', location) ~ 'Allen',
    grepl('Art Building', location) ~ 'Art Building',
    grepl('Bell Tower', location) ~ 'Bell Tower',
    grepl('Biddle', location) ~ 'Biddle',
    grepl('Biological Sciences', location) ~ 'Biological Sciences',
    grepl('Bivins', location) ~ 'Bivins',
    grepl('Branson Hall', location) ~ 'Branson Hall',
    grepl('Bridges House', location) ~ 'Bridges House',
    grepl('Brodie', location) ~ 'Brodie',
    grepl('Bryan Center', location) ~ 'Bryan Center',
    grepl('Chesterfield', location) ~ 'Chesterfield',
    grepl('Crowell', location) ~ 'Crowell',
    grepl('Divinity', location) ~ 'Divinity School',
    grepl('Duke Chapel', location) ~ 'Duke Chapel',
    grepl('East Duke', location) ~ 'East Duke',
    grepl('FITZPATRICK', location) ~ 'Fitzpatrick',
    grepl('Fitzpatrick', location) ~ 'Fitzpatrick',
    grepl('Franklin Center', location) ~ 'Franklin Center',
    grepl('French Science', location) ~ 'French Science',
    grepl('Friedl Bldg', location) ~ 'Friedl',
    grepl('Fuqua', location) ~ 'Fuqua',
    grepl('Grainger Hal', location) ~ 'Grainger Hall',
    grepl('Gray', location) ~ 'Gray',
    grepl('Gross Hall', location) ~ 'Gross Hall',
    grepl('Hudson Hall', location) ~ 'Hudson Hall',
    grepl('Languages', location) ~ 'Languages',
    grepl('LSRC', location) ~ 'LSRC',
    grepl('Nanaline', location) ~ 'Nanaline',
    grepl('Nasher', location) ~ 'Nasher',
    grepl('Old Chemistry', location) ~ 'Old Chemistry',
    grepl('Page', location) ~ 'Page',
    grepl('Perkins', location) ~ 'Perkins',
    grepl('Physics', location) ~ 'Physics',
    grepl('Reuben-Cooke', location) ~ 'Reuben-Cooke',
    grepl('Rubenstein Hall', location) ~ 'Sanford',
    grepl('Rubenstein Arts', location) ~ 'Rubenstein Arts Center',
    grepl('Sanford', location) ~ 'Sanford',
    grepl('Smith Warehouse', location) ~ 'Smith Warehouse',
    grepl('Social Sciences', location) ~ 'Social Sciences',
    grepl('Teer', location) ~ 'Teer',
    grepl('The Ark', location) ~ 'The Ark',
    grepl('Trent', location) ~ 'Trent Hall',
    grepl('West Duke', location) ~ 'West Duke',
    grepl('White', location) ~ 'White Lecture Hall',
    grepl('Wilkinson', location) ~ 'Wilkinson',
    grepl('Wilson Center', location) ~ 'Wilson Center',
    TRUE ~ location))

locations <- course_data %>%
  distinct(location)
```

```
glimpse(course_data)
```

```
## Rows: 2,408
## Columns: 13
## $ class_identifier <dbl> 1.7901e+14, 1.7901e+14, 1.7901e+14, 1.7901e+14, 1.790~
## $ Subject          <chr> "AMES", "LIT", "CINE", "VMS", "AAAS", "AAAS", "AAAS",~
## $ catalog_number   <chr> "161", "213", "255", "232", "190S", "190S", "102", "1~
## $ Descr            <chr> "JAPANESE CINEMA", "JAPANESE CINEMA", "JAPANESE CINEM~
## $ Section          <chr> "1", "1", "1", "1", "4", "1", "1", "1", "1", "1", "1"~
## $ enroll_cap       <dbl> 40, 40, 40, 40, 18, 13, 28, 28, 32, 32, 32, 32, 32, 1~
## $ days             <chr> "MW", "MW", "MW", "MW", "MW", "TH", "TTH", "TTH", "T"~
## $ mtg_start        <time> 15:30:00, 15:30:00, 15:30:00, 15:30:00, 19:00:00, 10~
## $ mtg_end          <time> 16:45:00, 16:45:00, 16:45:00, 16:45:00, 20:15:00, 12~
## $ Mode             <chr> "In Person", "In Person", "In Person", "In Person", "~
## $ location         <chr> "Classroom Buiding", "Classroom Buiding", "Classroom ~
## $ Career           <chr> "UGRD", "UGRD", "UGRD", "UGRD", "UGRD", "UGRD", "UGRD~
## $ term             <chr> "2022 Spring Term", "2022 Spring Term", "2022 Spring ~
```

High Level Overview

Create an R Shiny app that allows students to build their academic schedule and provides additional insights to their schedule through data visualization.

Project Description/Goals/Motivation

Dukehub is an academic portal used by students, faculty, and advisors to view courses, make tuition payments, and view transcripts. Each semester students spend countless crafting their schedules and back-up schedules. Students can find classes using a "simple search" by term and subject area. Students can also find classes using an "advanced class search" inputting the course attributes, meeting times, instructor name, location, or the number of units. Duke launched DukeHub 2.0 in 2020 to improve user experience and add more features. DukeHub 2.0 is very useful, but our team wishes it would provide more information about our courseload. Our team's goal is to create an improved DukeHub using a shiny app which allows students to select courses for their schedule and provides additional information about their schedule based on their selection. Specifically, we hope to create three visualzations based on student class selection:

- Course schedule in a calendar (classes colored based on student selection). If there is overlapping class times, an error will occur (visually).

- Geospatial visualzation showing the distance traveled during a specific day or week.

- Visualization showing the population density in each building on campus during a given time.

To complete our project, we will be using a course catalog data set that we requested from the Duke University Registrar's Office. This data set contains 2,408 observations of 13 variables. Each observation in the data set represents a course which is offere to undergraduates during the Spring 2022 term. There are 13 variables in the dataset: `class_identifier`, `Subject`, `catalog_number`, `Descr`, `Section`, `enroll_cap`, `days`, `mtg_start`, `mtg_end`, `Mode`, `location`, `career`, `term`.

Additonally, we are creating a distance dataframe which describes the distance between buildings.

Weekly Plan of Attack

- Week 1 of project (week of Mon, Oct 18): Kate, Yihan, and Kartik contributed 3 potential ideas each and we chose one that would be the most interested. We decided to build a R Shiny app similar to Dukehub.

- Week 2 of project (week of Mon, Oct 25):

- Kate: project proposal motivation and goal.
- Yihan: project proposal (introduce brief description of each dataset including the reason why you chose the particular dataset, its dataset, its dimensions and any other relevant metadata)
- Kartik: weekly plans

- Week 3 of project (week of Mon, Nov 1):

  - Kate: Finish proposal; finish creating computer science data
  - Yihan: Finish creation of repository;finish creating dataset
  - Kartik: Finish proposal;finish creating dataset

- Week 4 of project (week of Mon, Nov 8): Conduct peer review on project proposals, and optionally, submit in an updated version of your proposal.

  - Kate: work on peer review and update proposal; work on UI
  - Yihan: work on peer review and update proposal; work on UI
  - Kartik: work on peer review and update proposal; work on UI

- Week 5 of project (week of Mon, Nov 15): Continue working on your project.

- Week 6 of project (week of Mon, Nov 22): Continue working on your project.

- Week 7 of project (week of Mon, Nov 29): Conduct another round of peer review.

- Week 8 of Project(Due Date): Present our project

Organization of Project Repository:

- Data folder contains 5 datasets:
  - course_catalog.csv: All classes offered to undergraduates at Duke Unversity in the Spring 2022 term.
  - building_dist.csv: Distances between each building group
  - building_group.csv: grouping each building into regional groups
  - README.md for data folder
- Proposal Folder :
  - proposal.rmd
  - README.md
- DukeHub3.0
  - ui.R
  - server.R