**Statistical Analysis with R.**

# Instructor Email

Son Nguyen snguyen4@bryant.edu

Canvas and <https://bryantstats.github.io/math421/>

# Course Description

This course covers the application of R in a wide range of subjects in data analysis. The statistics subjects include descriptive statistics; hypothesis testing; probability distribution; Bayesian statistics; predictive modelling; unsupervised learning. Students also learn about how to write functions in R, Rmarkdown, and various R famous packages such as ggplot2, caret, dplyr.

**Prerequisites**

MATH 350 or AM 332

# Desired Learning Outcomes

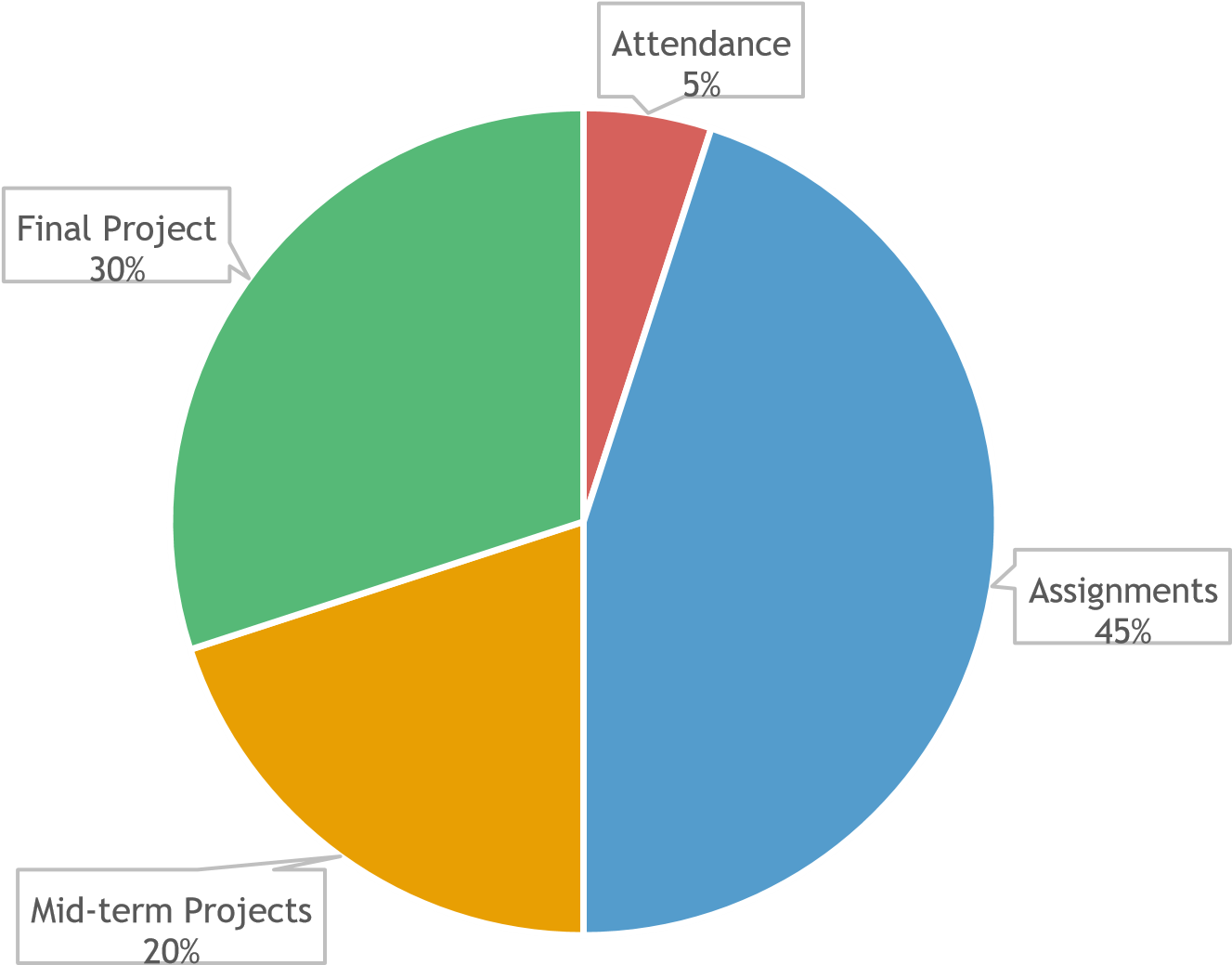
After completing this course, you should be able to

* implement a variety of data exploration techniques (such as plot variables, histograms, calculating statistics) on data using R.
* perform statistical inferences, such as hypothesis testing, confidence interval, ANOVA on data using R.
* build statistical models, such as linear regression, logistic regression, decision tree on data using R.
* implement a variety of techniques in statistical unsupervised learning on data using R.
* produce a data-interactive document and slides presentation to communicate their works using R.

# Office Hours

You can find me at my office (A8 - Suite A) from 11:55AM-12:55PM on MWF. If the office hours do not work, please feel free to schedule another time with me.

**Course Materials**

All the course materials will be provided online. **Grades**

# Attendance

Attendance will be checked in every class meeting. Missing less than four class meetings will guarantee you the full credits for attendance**.**

# Assignments

The assignments include, but not limited to, solving statistical problems, writing reports about a statistical technique, and presenting statistical analysis on a dataset.

# Datacamp

You are required to register an account in Datacamp and to be added to the class. You will be assigned to take Datacamp courses throughout the semester. You are responsible for the materials mentioned in the assigned Datacamp courses.

# Midterm and Final Projects

Projects should be done individually. The logistics and other details of the projects will be updated on the class webpage and Canvas.

# Late Work

Late assignments are penalized 10% for each day late. You can resubmit your work to improve your score, but the late penalty will apply.

# Grading Scale

A 92.45 - 100% C 72.45 - 76.44%

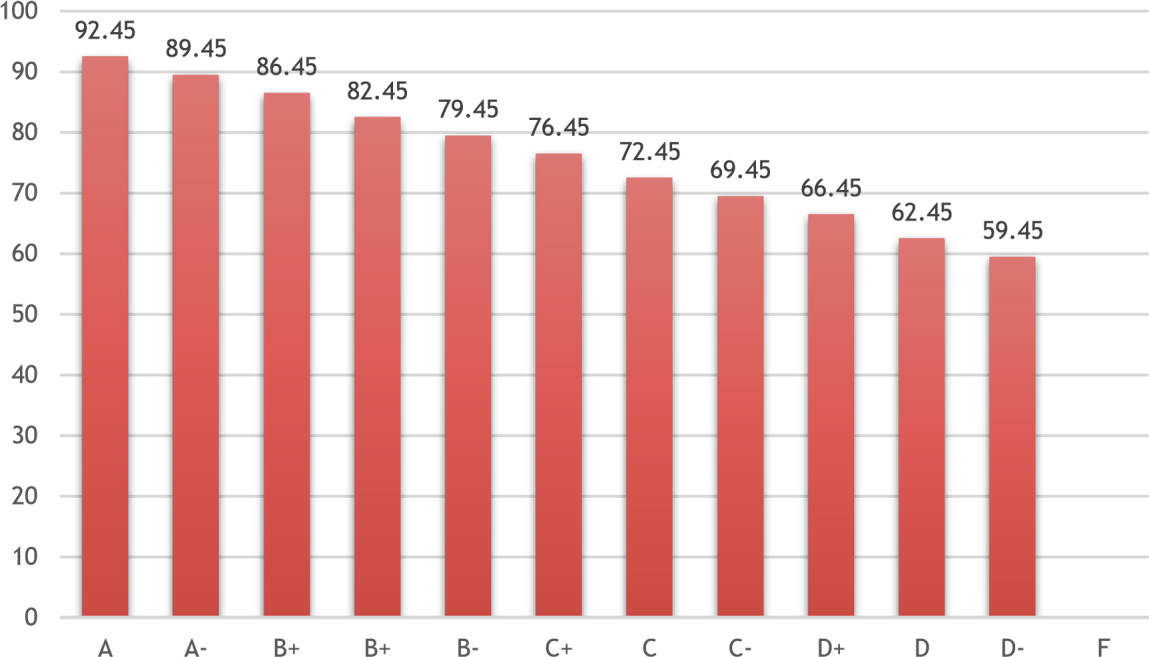
A- 89.45 - 92.44% C- 69.45 - 72.44%

B+ 86.45 - 89.44% D+ 66.45 - 69.44%

B 82.45 - 86.44% D 62.45 - 66.44%

B- 79.45 - 82.44% D- 59.45 - 62.44%

C+ 76.45 - 79.44% F Below 59.44%



# Tentative Topic and Weekly Schedule

|  |  |
| --- | --- |
| Week | Topic |
| 1 | R Basics |
| 2 | Data wrangling with dplyr and tidyverse |
| 3 | Data wrangling with dplyr and tidyverse |
| 4 | Handle Missing Values |
| 5 | Data Visualization with ggplot2 |
| 6 | Data Visualization with ggplot2 |
| 7 | Predictive Modelling |
| 8 | **Midterm Project** |
| 9 | Predictive Modelling |
| 10 | Text Mining |
| 11 | Text Mining: Twitter Data |
| 12 | If Statement, Loop and Functions |
| 13 | Final Project - Writing Shiny apps |
| 14 | Final Project - Writing Shiny apps |
| 15 | Final Project - Writing Shiny apps |

# Academic Misconduct

Cheating will result in an “F” as your final grade and may result in your expulsion from the University. All cheating will be reported to the Chair of the Mathematics Department and Academic Advising.

# Regarding Diversity

In this course, and all your courses at Bryant, and throughout the Bryant learning community, we value and respect diversity. This includes differences in race, ethnicity, nationality, gender, gender identity, sexuality, socioeconomic status, ability, and religion.