Infusion of Data Science Tools and Knowledge into Introductory Statistics

Implementation Manual

Mathematics & Statistics Department

6/21/2022

DRAFT

To Be Updated Based on Feedback from Professional Development Sessions

Funded by NSF HBCU-UP Targeted Infusion Project 2106945.

Contents

Chapter 1
Overview of the Project

1.1. Project Goals

Infusing Data-Centered Pedagogy and Data-Analytical Skills into Introductory Statistics is an innovative instructional reconceptualization and redesign project aiming to transform the teaching of introductory statistics (intro stats) at North Carolina A&T State University (NCA&T) through targeted infusions of data science (DS) knowledge and big data analytics tools in the high-stakes intro stats course to enhance the statistical and data-analytical skills of and promote DS literacy among underrepresented minority (URM) students. The project seeks to achieve three main goals: (1) Enhance students' statistical knowledge and data-analytical skills gained from the intro stats course; (2) Create a pipeline for the new DS programs offered at A&T; and (3) Build a faculty cadre capable of and committed to teaching intro stats using a data-centered pedagogy to promote data literacy among undergraduate students.

This project is a faculty-driven, theory-informed and evidence-generating project that takes a develop-incorporate-assess approach to redesign the general education intro stats course at NCA&T resulting in a novel and replicable course design that does not only ensure development of statistical and data-analytical skills at early stages of college education for URM students but also prepares and entices STEM and non-STEM students to pursue DS education and careers.

The intro stats course at NCA&T serves more than 600 students annually (46% STEM majors, 82% African Americans, and 69% females). The redesign of the course includes the structure, content, and pedagogy. The course redesign adds a virtual statistical computing lab to the course in a way that carefully coordinates the lab sessions with the class sessions. The class and lab activities are designed to train students to think structurally with data, become data-savvy, and expose them, early and frequently, to the elements of the DS workflow. The redesigned course incorporates components directed to boosting students' knowledge of DS educational and career opportunities.

Learn more about this project and access project materials and products through the project's website *here*.

1.2. Funding Source

The National Science Foundation (NSF) Historically Black Colleges and Universities Undergraduate Program (HBCU-UP), through Targeted Infusion Projects, supports the development, implementation, and study of evidence-based innovative models and approaches for improving the preparation and success of HBCU undergraduate students so that they may pursue STEM graduate programs and/or careers.

This project is supported by a grant from the NSF HBCU-UP Target Infusion Projects Program under Grant No. HRD 2106945.

1.3. Project Team

The project team is comprised of experienced statisticians and mathematicians who are both effective educators and well-established researchers. The following is a description of the coordination of project activities among the team members.

Dr. Sayed Mostafa (PI and project coordinator) is an assistant professor of Statistics, course coordinator of MATH224 and director of the Statistical Computing Lab. He is responsible for overall project management. He leads the team efforts for developing materials for the new course design, teaches the new course with the virtual statistical computing lab supplement during the implementation, evaluation, and generalization stages of the new course design, leads the internal evaluations and data analyses, organizes the training workshops and lunch & learn for the faculty development, and leads dissemination efforts by presenting at conferences and seminar talks at peer institutions.

Dr. Guoqing Tang (Co-PI), professor and chair of the Mathematics & Statistics Department, manages the engagement of faculty and graduate research assistants in the project and handles the scheduling of course sections included in the study.

Dr. Seongtae Kim, Co-Principal Investigator, is an Associate Professor of Statistics in the College of Science and Technology at NCA&T, where he teaches both introductory and upper-level statistics and data science courses and serves as the Director of the Statistical Consulting and Analytics Center. He will participate in developing material for the new course design, will implement the new course design in his course sections, will participate in collection of internal evaluation data, will play a key role in disseminating the project results and products by giving conference and/or seminar talks and writing journal publications, and will present at the faculty development workshops and lunch & learn.

Dr. Tamer Elbayoumi, Co-Principal Investigator, is an Assistant Professor of Statistics in the College of Science and Technology at NCA&T. He will participate in creating and revising course materials, implementing the new design, and collecting internal evaluation data. He will also contribute to disseminating the project results and products by giving conference and/or seminar talks and writing journal publications. Finally, he will be responsible for updating the project page on the Open Science Framework (OSF).

Dr. Mingxiang Chen, Co-Principal Investigator, is a Professor of Mathematics in the College of Science and Technology at NCA&T. She has taught introductory statistics for about ten years. She will help with creating and revising course materials, will implement the new course design in her course sections, will help with evaluation data collection, and will facilitate generalizing the new course design to all course sections.

Dr. Monique Matelski (External Evaluator), Director of Research & Evaluation, at Cobblestone Applied Research & Evaluation, Inc, is responsible for project evaluation and assessment.

Chapter 2

Infusion of Data Science Tools and Knowledge into Introductory Statistics

In this chapter, we describe the process of infusing data science tools and knowledge within the intro stats course to enhance students' data-analytical skills and awareness and aspirations of data science.

2.1. Revised course content:

From:

Table 1. Content of the intro stats course under current course design.

1. Introduction (basic concepts)
- Descriptive vs inferential statistics
- Types of data (quantitative vs qualitative)
- Sample vs population
- Sampling methods (simple random sampling and systematic sampling)

2. Descriptive statistics
- Describing data graphically (manually construct various types of graphs)
- Numerical summaries (manually compute central tendency and variability measures, and standardized scores)

3. Introduction to probability
- Basic probability terminologies (sample spaces, events, complementary events, and unions and intersections of events)
- Additive rule and disjoint events
- Multiplicative rule and independence
- Conditional probability

4. Probability distributions
- Use formulas to manually compute expectation and variance of a given discrete probability distribution
- Use binomial formula to compute probabilities about binary variables
- Use normal table to compute probabilities and percentiles for normal random variables

5. Sampling distribution of sample mean
- Central limit theorem
- Use normal table to compute probabilities about the sample mean

6. Confidence intervals
- Use formula, calculator and normal table to compute confidence interval for the population mean/proportion

7. Hypothesis testing
- Perform 5 systematic steps and use calculator and normal table to compute p-value and reject/retain the null hypothesis about the population mean/proportion

To:

Table 2. Content of the intro stats course under the new course design.

1. Introduction to elements of data analysis
- Data analysis workflow (research question, data acquisition, cleaning, wrangling, visualization, modeling, and interpretation)

2. Data collection/acquisition
- Target population vs sample
- Sampling variation and generalization
- Sampling and resampling
- Data from designed experiments

3. Univariate descriptive statistics
- Graphics (bar charts, dot plots, histograms, boxplots, and density plots)
- Numerical summaries (five-number summary, mean, standard deviation, and standardized scores) and detect outliers

4. Bivariate relations
- Simple linear regression

5. Probability, chance models and sampling distributions
- Basic probability rules, conditional probability and independence
- Binomial and normal probability models
- Sampling distribution of sample mean/proportion with simulations

6. Inference for one population mean/proportion
- Construction and interpretation of confidence intervals
- Classical t-tests and resampling tests for one mean/proportion
- How large is the evidence (effect size)?
- Statistical versus practical significance

7. Inference for two population means/proportions
- Construction and interpretation of confidence intervals for difference bet. two means/proportions

• Scatterplots, correlation and causation • Contingency tables for categorical variables • Faceted plots for displaying relations across different levels of categorical variables	• Classical t-tests and permutation tests for two groups • Using plots to check assumptions **8. Multivariate relations** • Multiple linear regression & analysis of variance

2.2. Textbook:

The free ***OpenIntro Statistics*** (https://www.openintro.org/book/os/) textbook is adopted for the course. This book covers the main topics in a modern introductory statistics course.

2.3. MyOpenMath Online Homework System

The free online homework system MyOpenMath (https://www.myopenmath.com/info/lti.php) is adopted for the course. Neither instructors nor students need to pay for this homework system. Further, students only use the homework link in the learning management system to access and complete homework assignments without needing to create an account on MyOpenMath.

The following is a description of the how to integrate MyOpenMath assignments within Blackboard.

Blackboard Integration - Manual Process

To manually create a link to MyOpenMath in Blackboard:

- Under the Build Content dropdown, select Web Link
- Give the link a Name
- For URL, enter https://www.myopenmath.com/bltilaunch.php
- Click the checkbox for "This link is to a Tool Provider"
- If you have a global college-wide key and secret, you will not see or need to fill out the following two entries.
    - For Key, enter LTIkey_####_1, where #### is your Course ID
    - For Secret, enter the secret shown in your MyOpenMath Course Settings
- If you are setting up a link to an individual assessment, set "Enabled Evaluation" to Yes, and enter the points possible. Whole course links to not return grades, so there is no need to enabled evaluation for those.
- Click Submit
- If the link you just created shows "(invalid link)" next to the name, they you'll need to ask your Blackboard administrator to approve LTI connections to MyOpenMath.
- Click on the link you just created.
- If this is your first time using this key and secret, MyOpenMath will ask you, the teacher, to sign into your MyOpenMath account. This is necessary to establish a connection between your LMS account and your MyOpenMath account. You will not need to do this step again, and students will not be asked to sign in and will not need a MyOpenMath account.

- If this is your first link from this course, MyOpenMath will ask you to select the MyOpenMath course you want to connect your LMS course with.
- Select the assignment you want to link with, or select "Whole Course Placement" if you want a single link to your entire MyOpenMath course, then click Make Placement.

Be aware assessment links created manually in this way will lose their association with MyOpenMath assessments when the course is copied in Blackboard, so using the Export/Import process is recommended if you are creating individual assessment links.

For other Learning Management Systems (LMS) such as Canvas,
See: https://www.myopenmath.com/help.php?section=lti

2.4. Virtual Statistical Computing Lab

The course redesign complements the class sessions with a virtual statistical computing lab and carefully coordinate the activities of class and lab sessions. The virtual lab sessions use **RStudio Cloud**, which (i) provides students with free and effortless access to computing in R/RStudio, (ii) reduces the faculty and students' effort to deal with device-specific issues with the R/RStudio software, and (iii) removes the logistic restrictions associated with physical computer labs without adding substantial cost (currently it costs only $15/month per instructor allowing 300 hours usage and unlimited number of student participants). These lab sessions are held using Zoom. This makes the weekly contact time of the course 4 hours (3-hour class session plus 1-hour virtual lab session). During class sessions, R is used to demonstrate concepts, introduce new commands, and run data analyses. During virtual lab sessions – staffed by an instructor and a graduate teaching assistant (GTA)– students are guided to further explore concepts via simulations, to practice using R commands introduced in class, and to analyze real datasets and make data-driven decisions. This design brings data analysis to the front and center of the course structure, which is aligned with the principles of the data-centered pedagogy. The repeated practice in making data-informed decisions with feedback from instructor and peers helps students develop critical thinking skills, which are necessary for the big data era. Further, during virtual lab sessions, students work in groups on major-related data analysis projects that aim to help students understand statistics as a problem-solving and decision-making process rather than a collection of unrelated methods and formulas. These projects help promote collaborative learning and teamwork spirit among students.

2.4.1. R Shiny Reading Assignments

Interactive R shiny tutorials are assigned as weekly reading assignments for students to complete prior to the computing lab session. These tutorials are used to help students prepare for the weekly lab sessions. The tutorials used are inspired by and largely derived from the AppliedStatsInteractive of Adam Gilbert.

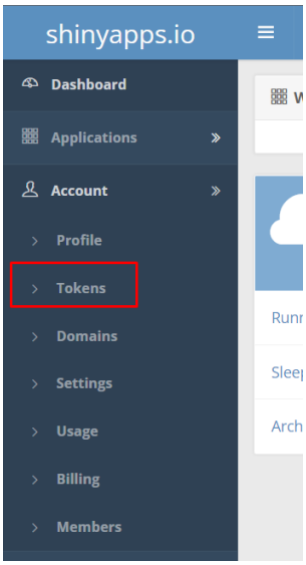
In the following, we describe how to set-up these tutorials and publish them on the R Shiny server for students to work on.

2.4.1.1. How to connect RStudio to IntroToStatNCAT shiny account

Here we will go through a step-by-step process of connecting the IntroToStatNCAT shiny account to the RStudio IDE on your local computer to be able to publish the reading assignments.

Step 1: Go to https://www.shinyapps.io/ and log in using the IntroToStatNCAT account through google.

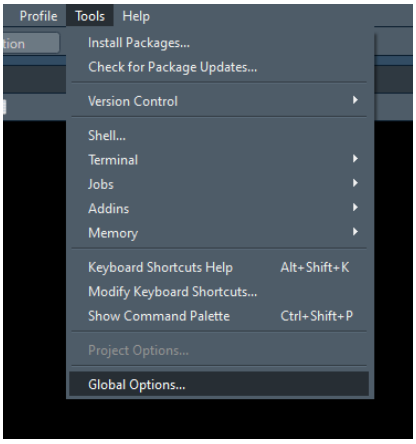
Step 2: Locate Account on the left panel and choose Tokens



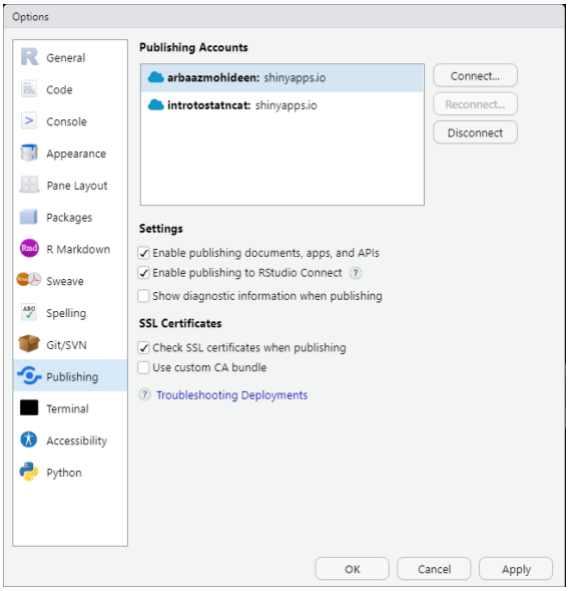
Step 3: Click on Show and then click on “Copy to Clipboard”

Step 4: Open RStudio on your local computer

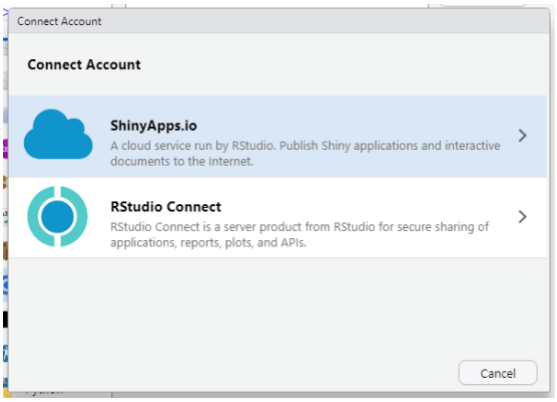
Step 5: Go to Tools -> Global Options in the Top panel



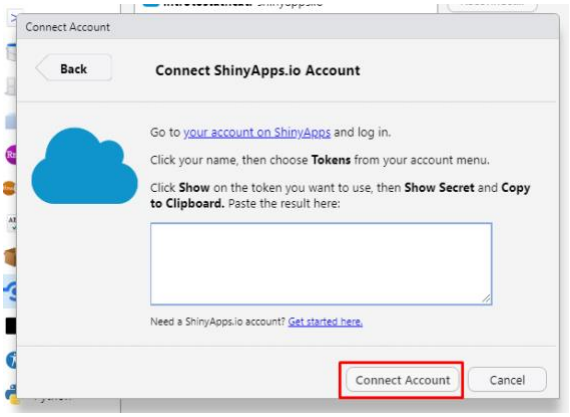
Step 6: Select the category “Publishing”



Step 7: Press connect and select ShinyApps.io



Step 8: Paste the secret that we copied in step 3 into the space provided



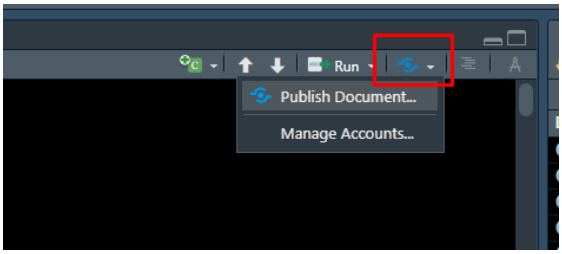
Step 9: Click Connect Account. Now you have connected the ShinyApps account to your local computer.

2.4.1.2. How to publish for the first time

Here we will show instructions on how to publish a reading assignment to the shiny app servers for the first time. That is, this assignment has not been published already.

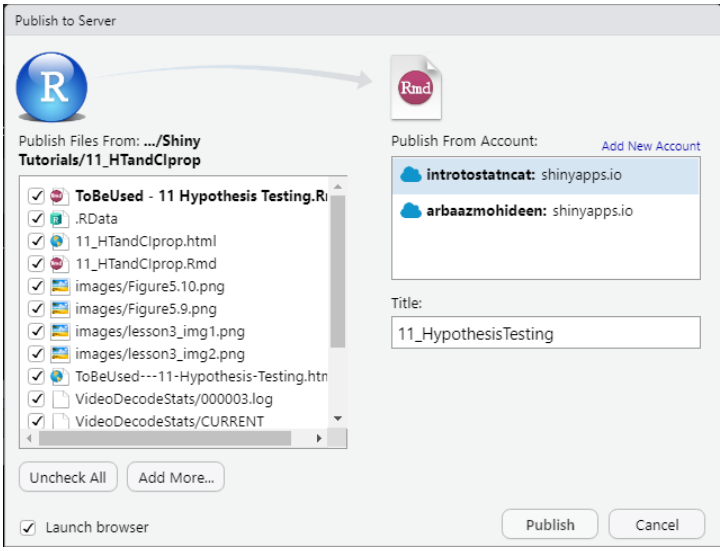
Step 1: Open the RMD for the reading assignment that needs to be published.

Step 2: Locate the publish icon on the top right of the RMD pane in RStudio.



Step 3: Now click on the drop-down menu and select Publish Document.

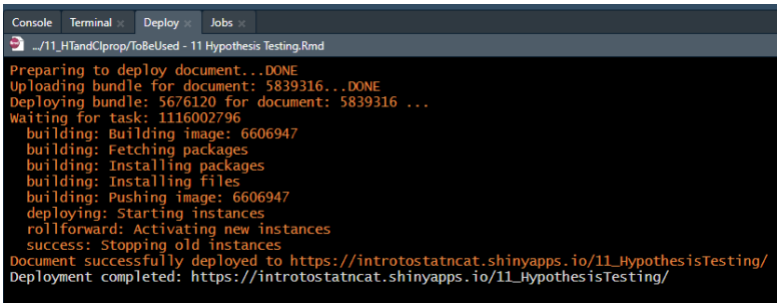
Step 4: For instruction purposes, we will make use of the Hypothesis Testing reading assignment. Once you complete step 3. Then you will see this,



On the left, it is all the files in the folder where the main RMD file is located. On the right, you will find all the accounts you have connected to your RStudio. In our case, we have a personal account and our target account.

Step 5: Make sure to select “introtostatncat” and then press Publish. P.S. Make sure the title has no white spaces between the characters. Title is the extension of the link for the reading assignment once it is published.

Step 6: Once you press Publish, it will take a while to deploy the documents. You should see this message in the RStudio Console pane when its published.



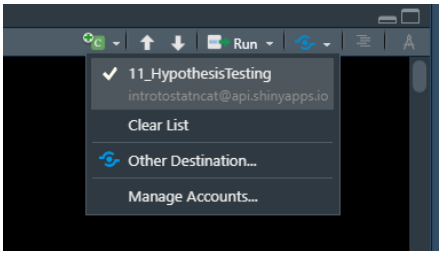
Step 7: Wait till the server has been deployed. Once the server deployment is complete then it will automatically open a new tab in your default browser. This indicates the completion of the reading assignment being published.

2.4.1.3. How to update an already published app

Here we will show how to update an already published app such as the Hypothesis Testing reading assignment that was published in 2.4.1.2. Let us say that we encounter a typo or a mistake in one of the plots. Now we will go through the steps to fix it in the shiny app.

Step 1: Fix the mistake in the main RMD file.

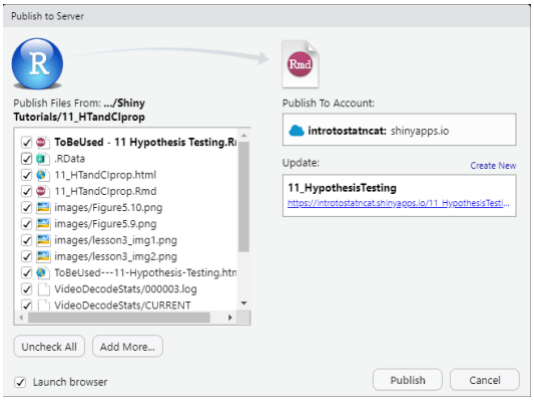
Step 2: Just like before, we will locate the publish icon on the top right of the RMD pane.



But this time it is different because this RMD was already used to publish a shiny app earlier. So, this will show up on the list with the exact title name we used the first time we published it.

Step 3: Click on the option with the check mark (first highlighted option).

Step 4: This time you will not get an option for which account you want to publish this app to as we already did that step in the previous section. Now you have “Update” instead of title as this information was already provided.

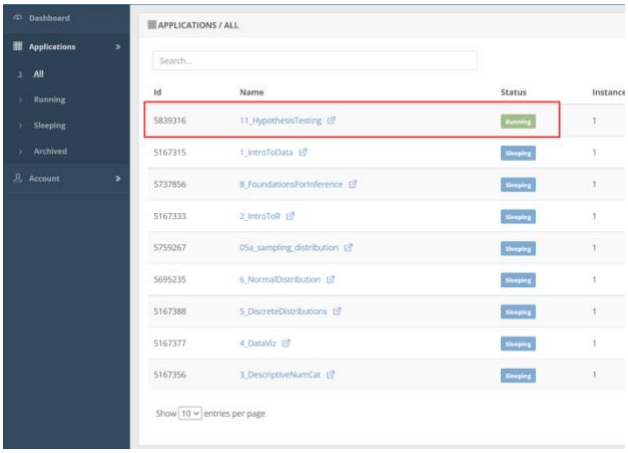


Step 5: Click on publish and wait till the server has been deployed. Just like the previous section. Once the app opens in a new tab, you will see that the changes made to the main RMD file is reflected in the app.

2.4.1.4. Adjusting ShinyApps (3X-Large) settings (Only for basic plan)

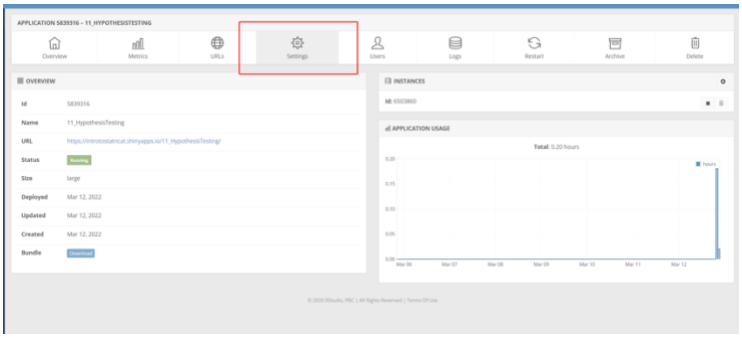
This step is particularly important as this decides the memory allocation for a given shiny app. In simpler terms, it dictates how many students can access the app at once. If one too many students access the app at once, then nobody will be able to open the app itself, stopping them from finishing their assignment.

Step 1: Once the app has been published. Go to https://www.shinyapps.io/ and locate Application in the left pane and click on All. This will show all the apps that has been published under this account.

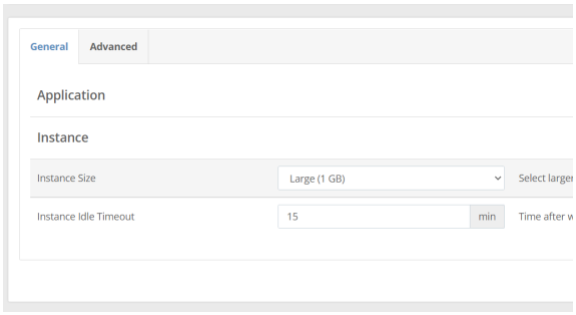


Step 2: Click on the recently uploaded. In our case, we will press on “11_HypothesisTesting”.

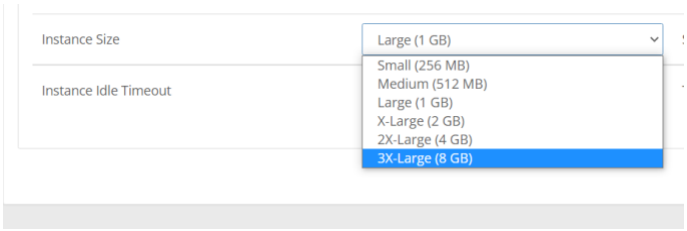
Step 3: Click Settings



Step 4: Locate Instance Size.

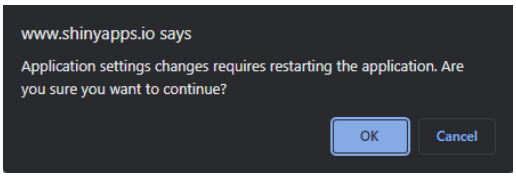


Step 5: By default, it will be set to “Large (1 GB)”. We will change this to “3X-Large (8 GB)” by clicking on the drop-down button.



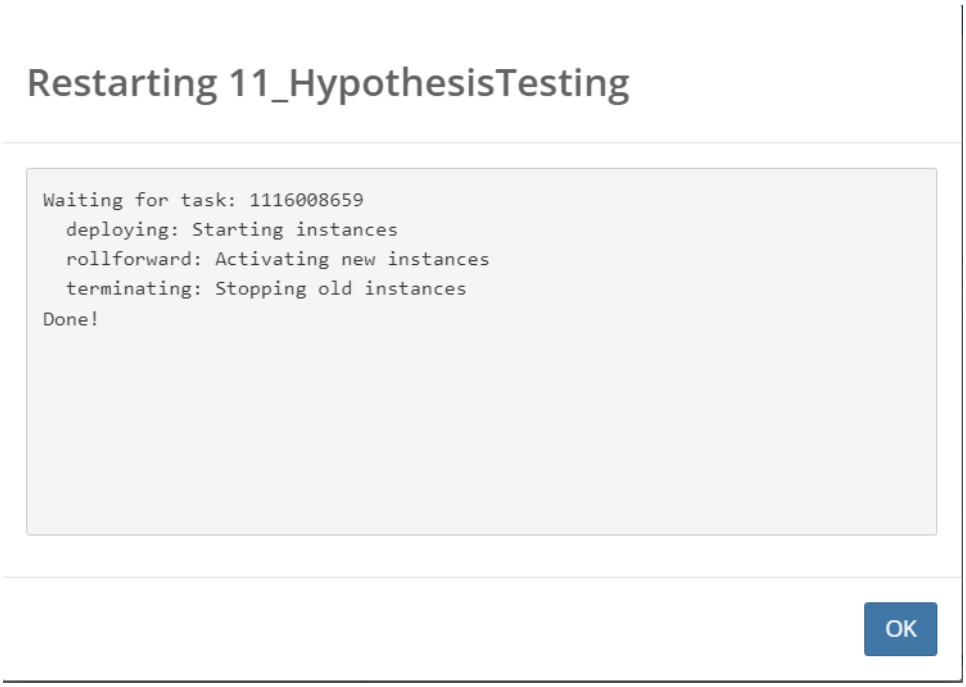
Step 6: Click on Save Settings.

Step 7: Now you will be given this pop-up:



Step 8: Click on OK to proceed. This will lead to the server being restarted.

Step 9: Once you get this message



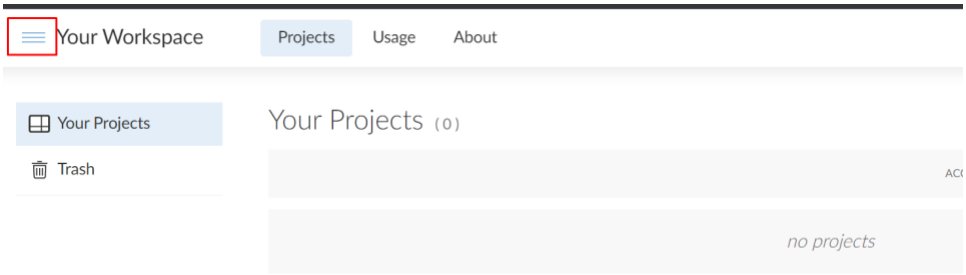
Go ahead and press OK. Now you are good to use the application link for students reading assignments.
P.S. You can acquire the link to the app from the overview tab. From there you can copy the URL.

2.4.2. RStudio Computing Labs

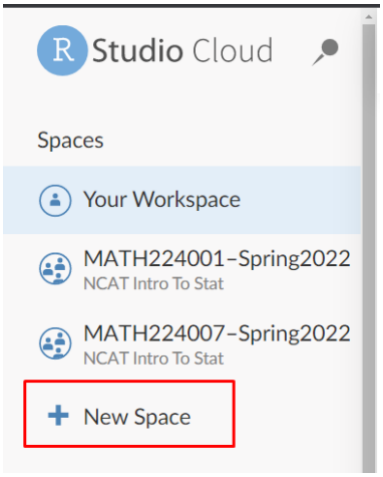
2.4.2.1. Setting up workspace (First Time)

Step 1: Go to https://rstudio.cloud/ and log in with the IntroToStatNCAT account using google.

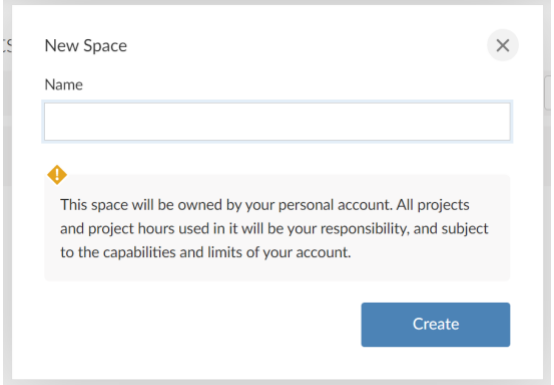
Step 2: Now press on the three blue lines on the top left part of the window.



Step 3: Now press “New Space”



Step 4: Now you will be prompted to enter a name for the workspace



Apply the format: Math224Section-SemesterYear. You can see examples of the name in Step 3.

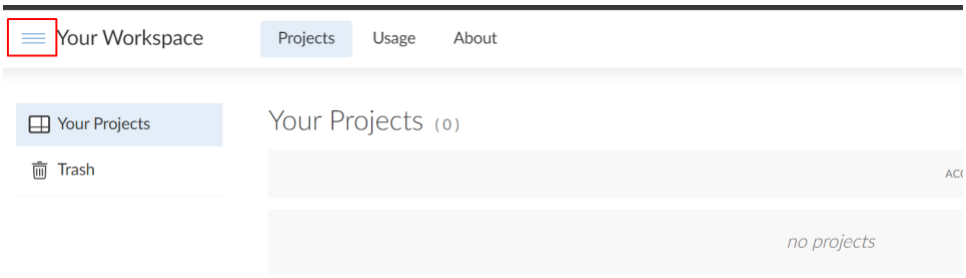
Now the workspace has been created.

2.4.2.2. Setting up workspace (Future)

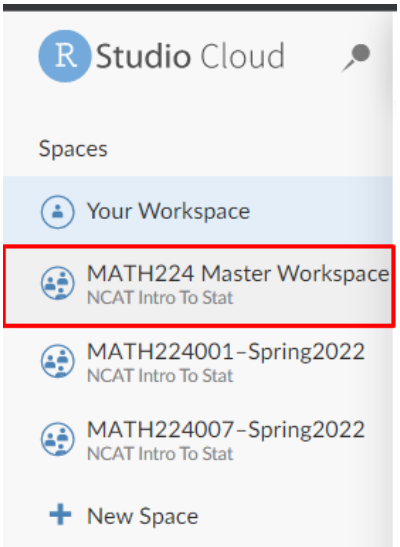
We already have workspace set up from the prior semesters. So, we will perform the following steps to copy the pre-existing workspace with members and projects. The good thing about copying an already existing workspace is that it will copy all the projects, but it will all the members (students and TAs) and provide us with a clean workspace.

Step 1: Go to https://rstudio.cloud/ and log in with the IntroToStatNCAT account using google.

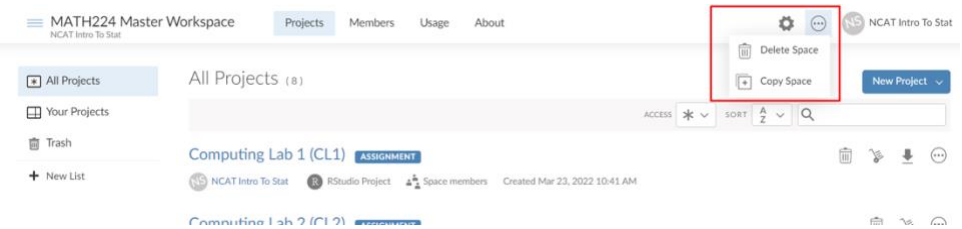
Step 2: Now press on the three blue lines on the top left part of the window.



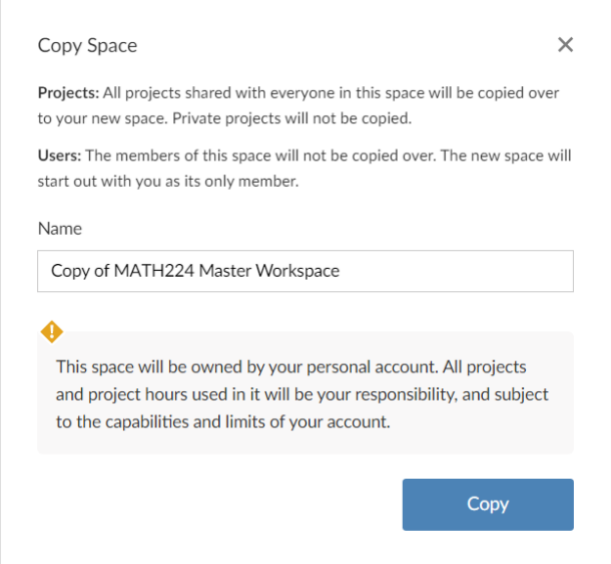
Step 3: Go MATH224 Master Workspace



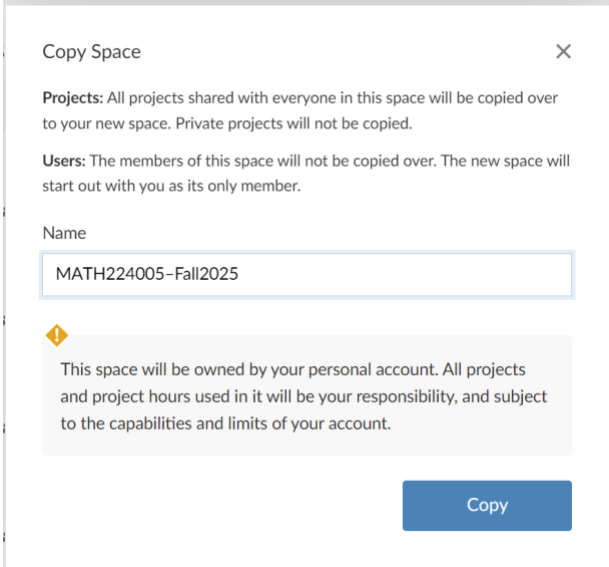
Step 4: Click on the circle with three dots inside towards the top right of the window.



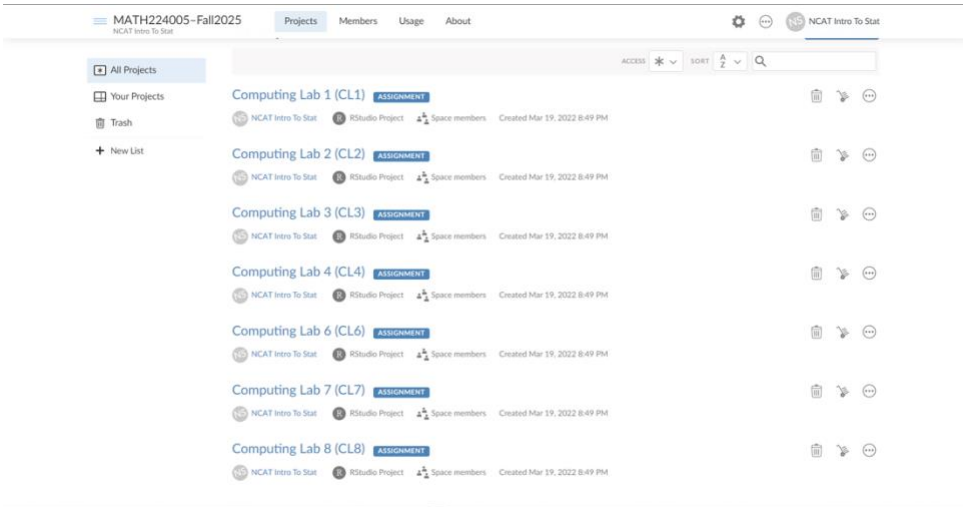
Step 5: Press Copy Space. Now you will be prompted to add a name for the new workspace.



Step 6: Enter the name of the format: Math224Section-SemesterYear. So, for Section 5 for Fall 2025 would be:



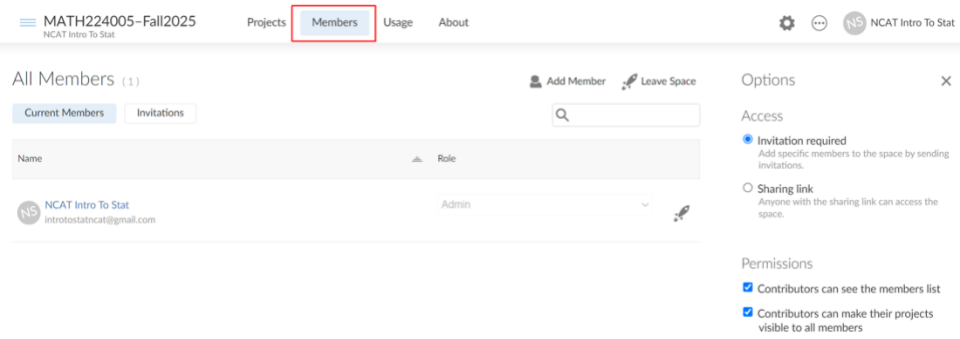
As you can see below, it copied all the main public projects so far and removed the previous members' individual projects.



2.4.2.3. Adding Other Instructors/TAs to Workspace

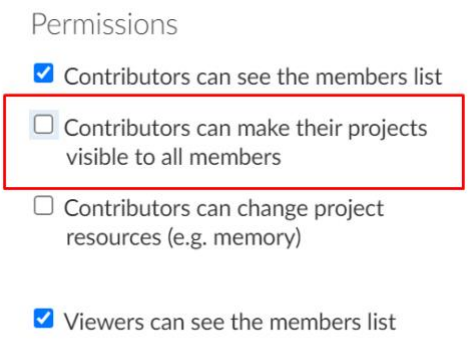
Step 1: Enter the workspace

Step 2: Locate the “Members” tab and click

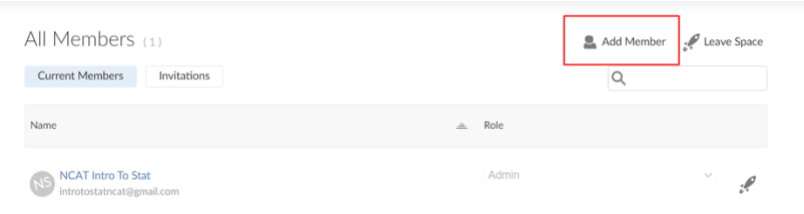


As you can see, the main account will be the only member of the workspace.

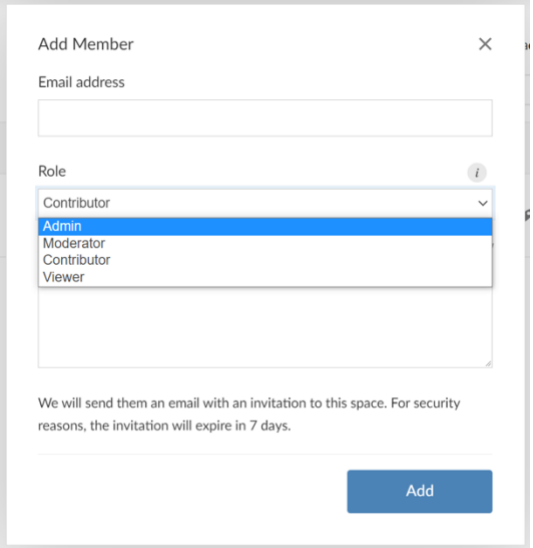
Step 3: Uncheck the 2nd option under Permissions on the right



Step 4: Click on “Add member”



Step 5: Enter the Instructors email and set the Role to Admin



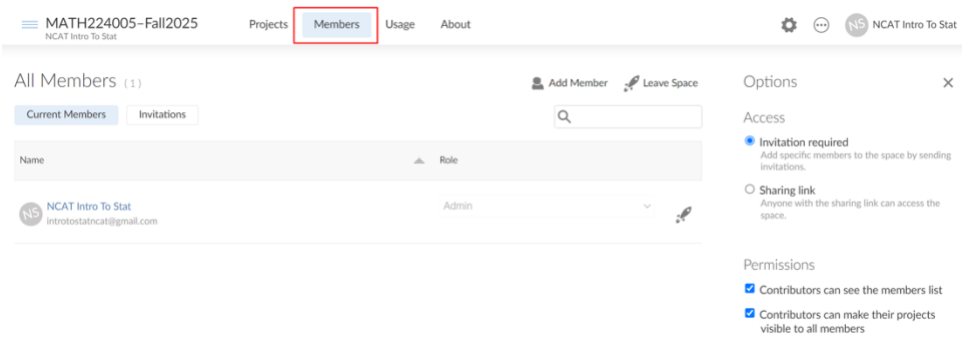
Step 6: Press Add. This will send an invitation to the provided email through RStudio Cloud.

Step 7: Enter the TA email and set the Role to Moderator. Now press Add.

2.4.2.4. Inviting students to join workspace

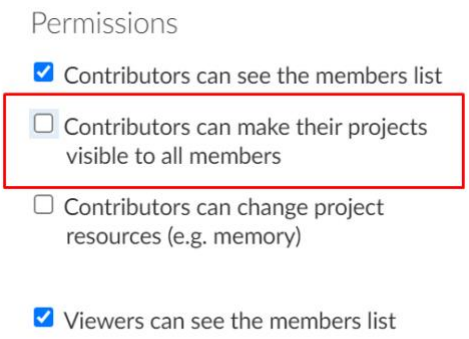
Step 1: Enter the workspace

Step 2: Locate the “Members” tab and click



As you can see, the main account will be the only member of the workspace.

Step 3: Uncheck the 2nd option under Permissions on the right



Step 4: Click on “Add member”

Step 5: Add the first students email and set the Role to Contributor. Now press “Add”

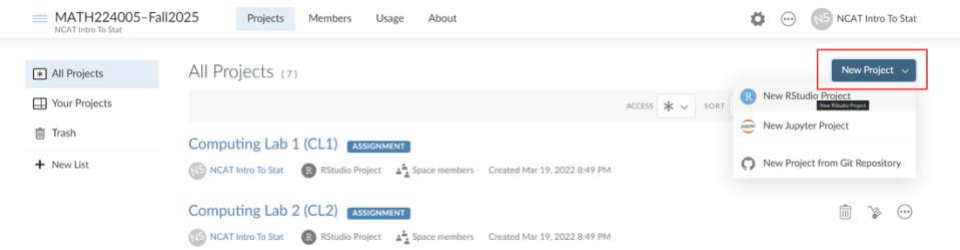
Step 6: Repeat Step 5 for all the students in each section.

2.4.2.5. Creating Lab projects

We would already have all the lab projects set up when we copy the workspace. But if there is a need for a new lab project then follow the steps:

Step 1: Enter the workspace.

Step 2: Click on “New Project” button on the right. This will provide you with a drop-down menu. Then select “New RStudio Project”



Step 3: While the project is deployed, go ahead, and change the name of the project. You can find it on the top of the window.

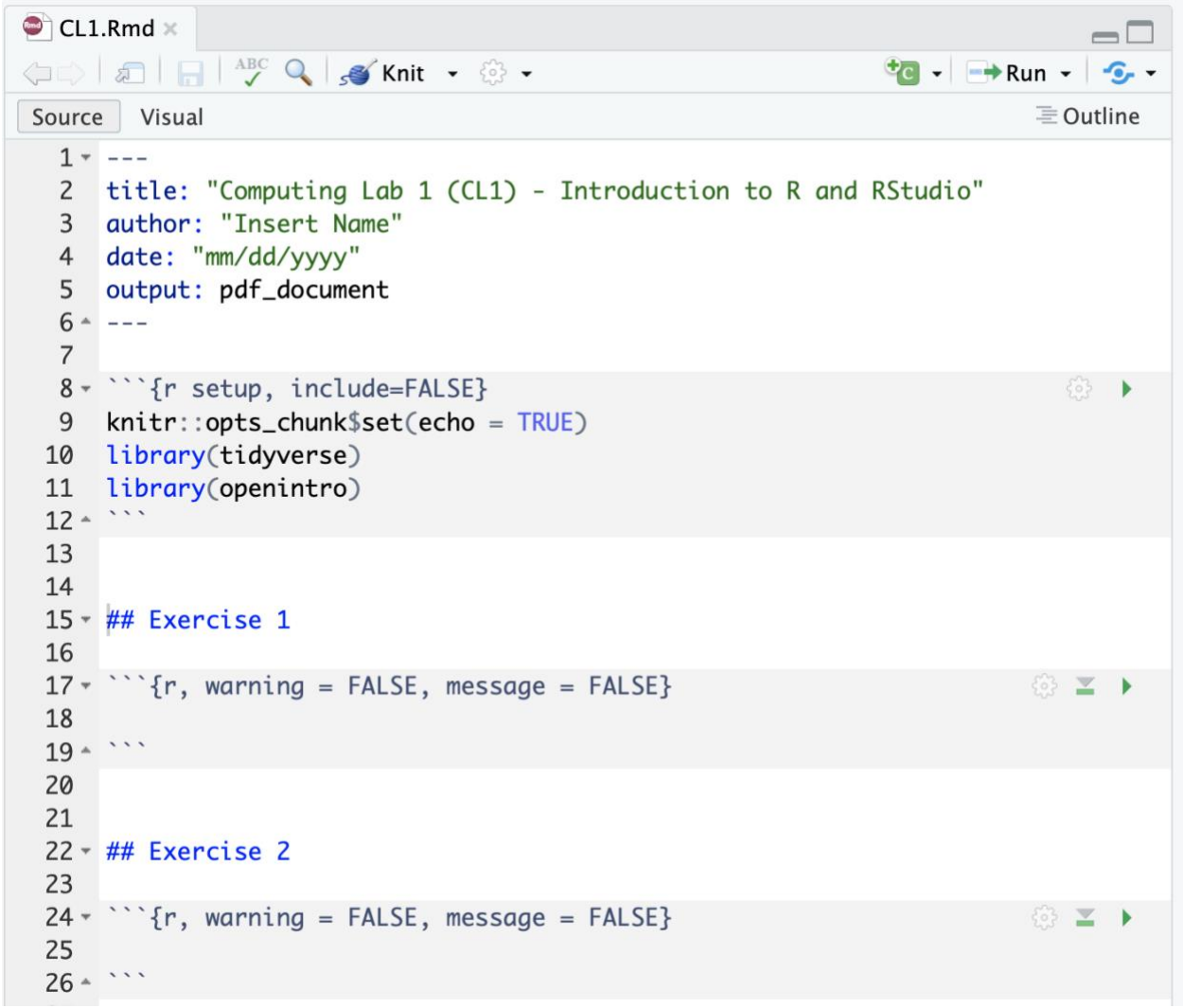
Step 4: Make a new RMD file and save it. Make sure to name the file with CL# (# - Lab number).

Step 5: Add the template for the Lab (Imitate the previously made labs with the Exercise heading and code chunk placements)

Step 6: Knit the template. This way, the empty lab report is found in the file's environment.

2.4.2.6. R Markdown Lab Templates

As described in previous section, an R markdown template for the lab report is created under each computing lab (project) in the RStudio cloud workspace. When students access the computing lab for the day, they find the markdown template which they will use to prepare their lab reports. A sample lab report template is shown below. These templates can be copied from the master RStudio workspace.

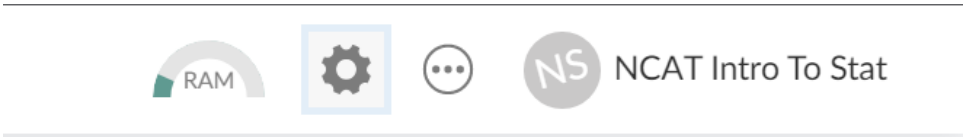


2.4.2.7. Open projects to students

Step 1: Enter the workspace.

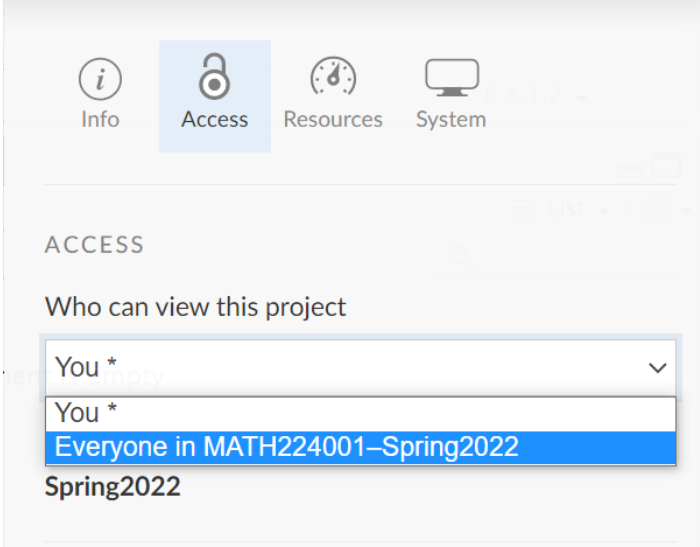
Step 2: Click on the lab that needs to be opened for students. This project (Lab) will originally be set to private.

Step 3: Click on Settings towards the top right of the window.

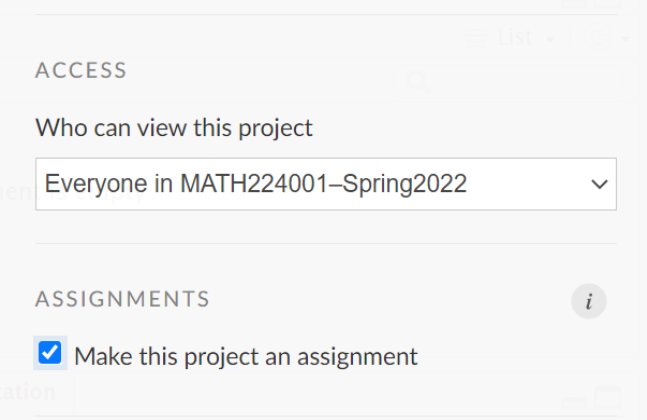


Step 4: Click on Access

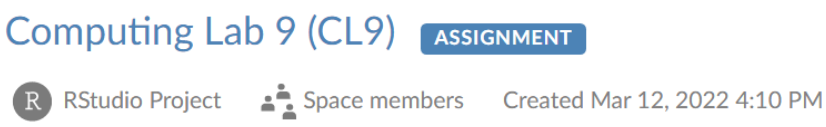
Step 5: Click on the drop-down menu and select “Everyone in the workspace”



Step 6: Check the option to turn the project into an assignment



Step 7: Now the students are ready to start the assignment.



2.5. Data Analysis Project

2.5.1. Creating Groups by Major

Step 1: Extract the file with student's names and majors from the beginning of the semester intro stat survey data

Step 2: Open the excel file with student information. This file will contain information such as First Name, Last Name, User ID, Email and Major.

Step 3: Fill in the missing Major cells using Discussion Board #1 entries.

Step 4: Fix the Major names as this information is pulled from a survey because students can enter the same major in multiple ways.

Step 5: Sort the Major column alphabetically from A to Z

Step 6: Separate each major with an empty row

Step 7: Now divide each subsection into 3 students or at least 2.

Step 8: If the subsection has students from the same major then good but if they are not then we want to match them up with majors that are related to each other. For example, Computer Science, Mathematics, and Information Technology are closely related. Mathematics, Physics and Engineering are related. Animal Science, Biology and Chemistry are related.

Step 9: We will assign a group number in the order the groups are divided now. It should look something like this.

Major	Group
AET - Applied Engineering Technologies	1
AET - Applied Engineering Technologies	1
engineering	1
Animal Science	2
Animal science	2
Animal science	2
Biology	3
Biology	3
Biology	3
Biology	4
Biology	4
Biology	4
Computer Science	5
Computer Science	5
Computer systems technology	5
Information of Technology	6
Information Technology	6
informational Tech	6
Kinesiology	7
Kinesiology	7
Physics	7
Liberal studies	8
Liberal studies	8
Math Education	8
Biology (stem ec student)	9
STEM EARLY COLLEGE	9
STEM EARLY COLLEGE	9

2.5.2. Assigning Data Sets by Major

Step 1: Open the groups excel file from the previous section and open the datasets excel file. The datasets excel file contains information such as dataset name, number of rows and columns, potential main variable, and major to which the dataset relates to the most.

Step 2: Now we will match the majors from both files and assign a dataset that will work the best for the students in that group.

2.6. Blog Posts about Data Science Educational and Career Opportunities

A series of blog posts are designed and embedded in the LMS (Blackboard) to provide students with information on the current trends in the data science job market and data scientists salaries via. Some of these informational posts/videos are also be discussed during class and lab sessions. Additionally, these blog posts are continuously updated to provide students with information about the data science educational opportunities at and outside A&T. These blog posts are hosted on the course landing (start here) page which students see once they login to the course in the LMS.

In Spring 2022, we made the following four blog posts about Data Science educational opportunities and current trends in the Data Science job market:

1) Why become a Data Scientist?;
2) How to become a Data Scientist at NCA&T?;
3) Undergraduate Certificate in DSA at NCA&T; and
4) Data Science & Analytics Research Experience for Undergraduates at NCA&T.

2.7. Discussion Board Assignments

A series of discussion board assignments that ask students to watch a video or read an article about the power of Statistics and Data Science, data science in action in real applications, and the use of Statistics in the student's own major field of study. Students write their own posts and comment on at least one post from their peers.

The descriptions of 3 discussion board assignments used in Spring 2022 are as follows:

1) The Joy of Stats:

Due: Friday/Sunday 11:59 PM EST

Instructions:
Create a new thread with your name when you start your own posting. For responses, you can just use the reply button. Each posting should be on different dates. Make sure you follow the discussion rubric to get proper credit for the posting.
You are required to post your own response and critique one other response.

- Please make your initial post by **Friday at 11:59 PM EST**, and then
- your 1 **critique response to another student's posting by Sunday at 11:59 PM EST**.

Topic:
For this week's discussion, you need to watch this 4-minute presentation (https://youtu.be/jbkSRLYSojo?list=PLkIselvEzpM5nFf4nu--q2OGccA_0alYO) by Hans Rosling where he uses data animation in real space to tell a story based on 200 years of data for 200 countries.

After watching the video, make your discussion post reflecting on the following:

- What is one peice of information you learned about the world from video?
- How does the presentation show the power of statistics and data visualization for telling data-driven stories?

2) The NBA Data Scientist:

Due: Friday/Sunday 11:59 PM EST

Instructions:
Create a new thread with your name when you start your own posting. For responses, you can just use the reply button. Each posting should be on different dates. Make sure you follow the discussion rubric to get proper credit for the posting.
You are required to post your own response and critique one other response.

- Please make your initial post by **Friday at 11:59 PM EST**, and then
- your 1 **critique response to another student's posting by Sunday at 11:59 PM EST**.

Topic:
For this week's discussion, you need to watch this 8-minute video (https://youtu.be/MpLHMKTolVw?list=PLkIselvEzpM5nFf4nu--q2OGccA_0alYO) about how statistical analysis and data science are used in helping NBA coaches devise smarter strategies to win.
After watching the video, make your discussion post reflecting on the following:

- How do statistical analysis and data science help NBA teams win games?

3) Statistics in Your Field

Due: Friday/Sunday 11:59 PM EST

Instructions:
Create a new thread with your name when you start your own posting. For responses, you can just use the reply button. Each posting should be on different dates. Make sure you follow the discussion rubric to get proper credit for the posting.
You are required to post your own response and critique one other response.

- Please make your initial post by **Friday at 11:59 PM EST**, and then
- your 1 **critique response to another student's posting by Sunday at 11:59 PM EST**.

Topic:
For this week's discussion, you need to find a research article/study or news report from your field of study (major) that uses statistical analysis. Create a post discussing the following:

- How statistics was used in the article/study/report?
- What statical methods do you usually hear about in your field/major?

2.8. Example Syllabus and Course Schedule

See Appendix.

2.9. Course Shell Design

The course shell was redesigned where all material for each week were organized together within a learning module. Fifteen learning modules were created to organize the material for the 15 weeks of semester. The following figures display the organization of the course shell.

Learning Units

Build Content | Assessments | Tools | Partner Content

Course Introduction and Orientation

Click on the link "Course Introduction and Orientation" above to access the content within the folder

Unit 1 – Introduction to data

Click on the link "Unit 1" above to access the content within the folder

Unit 2 – Summarizing Numerical Data

Click on the link "Unit 2" above to access the content within the folder

Unit 3 – Summarizing Categorical Data

Click on the link " Unit 3" above to access the content within the folder

Unit 4 - Probability I

Click on the link " Unit 4" above to access the content within the folder

Unit 5 - Probability II

Click on the link "Unit 5" above to access the content within the folder

Unit 6 - Distributions of Random Variables

Click on the link "Unit 6" above to access the content within the folder

Unit 7 - Distributions of Random Variables (Continued)

Click on the link "Unit 7" above to access the content within the folder

Unit 8 - Foundations for Inference I

Click on the link "Unit 8" above to access the content within the folder

Unit 9 - Foundations for Inference I (Continued)

Click on the link "Unit 9" above to access the content within the folder

Unit 10 - Foundations for Inference II

Click on the link "Unit 10" above to access the content within the folder

Unit 11 - Inference for Numerical Data I

Click on the link "Unit 11" above to access the content within the folder

Unit 12 - Inference for Numerical Data II

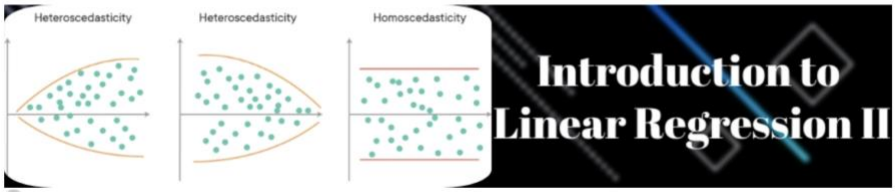
Click on the link "Unit 12" above to access the content within the folder

Unit 13 - Introduction to Linear Regression I

Click on the link "Unit 13" above to access the content within the folder

Unit 14 - Introduction to Linear Regression II

Click on the link "Unit 14" above to access the content within the folder



Unit 15 - Multiple linear regression

Click on the link "Unit 15" above to access the content within the folder

Chapter 3

Survey and Assessments

3.1. DS Awareness, Aspirations and Readiness Survey

3.2. Intro Stats Diagnostic Test

3.3. Focus Groups

Chapter 4

Professional Development Activities

4.1. Internal Faculty and Graduate Students Workshops

4.2. Workshops/Talks at Peer HBCUs

Chapter 5
Frequently Asked Questions

Appendix: Sample Course Syllabus and Schedule

Spring 2022 Course Syllabus

College of Science and Technology
Department of Mathematics and Statistics

NOTE: Students are responsible for reading, understanding and following the syllabus.

Undergraduate Course Information

Course Name: Introduction to Probability and Statistics

Course Number/Section: MATH224.007

Credit Hours: 3.00

Days and Times: MWRF 12:00-12:50pm

Class Location: MWF → Marteena Hall 312 and R → Virtual Computing Lab (Zoom*)

*You can attend the virtual lab on Zoom from Room 201 in Marteena Hall or any quite place.

Instructor Contact Information

Instructor: Dr. Sayed Mostafa

Office Location: 3005 Smith Hall

Email Address: sabdelmegeed@ncat.edu

Office Phone: 336-2853104

Communication

Students can expect to receive an answer to all communications (e.g., email, phone, course messages) within 48 hours, i.e., within two business days.

If you do not hear from me in 48 hours, please forward the email to me again.

Graduate teaching assistant: Arbaaz Mohideen (amohidee@aggies.ncat.edu)

TA Office Hours: TBA

Student Hours (Virtual via Zoom Link in Blackboard)

These are times students may visit the professor without an appointment to request the assistance they need.

10:00 AM ☒ / PM ☐ – 11:00 AM ☒ / PM ☐

Monday ☒ Tuesday ☐ Wednesday ☐ Thursday ☐ Friday ☒

Course Prerequisites

List the course Prerequisites here.

Minimum C Grade in MATH104 or MATH110 or MATH111 or MATH131

Course Description

This is a general introductory statistics course covering graphical and numerical summaries of data, basic rules of probability, introduction to discrete and continuous random variables, sampling variability, and the fundamentals of estimation and hypothesis testing for means and proportions.

General Education Student Learning Objectives/Outcomes (SLO)

By the end of this course, students will be able to do the following:

- Summarize data using graphical techniques such as histograms, barcharts, stemplots, boxplots, and scatterplots.
- Compute summary statistics for measures of central tendency (e.g., means, medians) and measures of spread and variability (e.g., standard deviation, interquartile range).
- Use the laws of probability to compute the probability of events (and their unions intersections and complements).
- Use discrete and continuous random variables, such as the binomial and normal distributions to estimate probabilities.
- Use the normal distribution and the Central Limit Theorem to approximate probabilities for sample proportions and means
- Estimate population means and proportions by confidence intervals.
- Perform and interpret tests of significance on population parameters.

Required Textbooks and Materials

Any course-level subscriptions and tools linked in Blackboard Learn learning management system (LMS) should be listed here. The Blackboard LMS must have links to their student data privacy statement.

Required Texts: OpenIntro Statistics (4th Ed) by Christopher D. Barr, David M. Diez, and Mine Çetinkaya-Rundel. Free pdf available here: https://www.openintro.org/book/os/

Required Materials:

- You need to create free RStudio Cloud account here: https://rstudio.cloud
- Access to Blackboard. Course syllabus, lecture notes, quizzes, homework assignments, and scores will be accessible in Blackboard.
- **All tests will be taken through Blackboard using Respondus Lockdown browser**.
- **A computer with webcam and microphone.**
- **You need to have Respondus Lockdown browser installed to take tests in the course**. For more details and downloading instructions, follow this link: http://www.respondus.com/lockdown/download.php?id=922833142
- **NCAT email account**: Students use their same email login credentials to access the course in Blackboard, which is the platform for this online course. Course syllabus, course materials and updates are posted on Blackboard.
- Students are expected to **actively participate in Discussion Board** to discuss the course contents and to post and/or answer questions.
- You are allowed to use any calculator for the course. Basic calculator that takes square root should be sufficient.

Grading Policy

Course Grade Scale *[Undergraduate level courses]*

92-100	A	80-81.99	B-	66-67.99	D+
90-91.99	A-	78-79.99	C+	60-65.99	D
88-89.99	B+	70-77.99	C	00-59.99	F
82-87.99	B	68-69.99	C-		

Grading Allocation

Course grades are based on a weighted grading scale of 100%. The breakdown for the course is as follows:

Work	Percentage	Notes
Homework	15%	Weekly HW assignments completed in Blackboard
Reading Assignments	10%	Weekly reading assignments in the form of interactive R Shiny tutorials with embedded knowledge check questions
Lab Reports	10%	Weekly lab attendance and submission of lab report due by the end of the day of each lab session
Data Analysis Project	10%	Work on a semester-long data analysis project in a group of 3 students.
Participation	5%	Discussion board assignments, in-class participations, pre- and post-diagnostic tests and the intro stats survey
Midterm Exams	30%	Best 2 scores out of 3 midterm exams **(each worth 15%)**
Final Exam	20%	**Comprehensive**

Homework (HW)

- These assessments will be completed in MyOpenMath by clicking the assignment link in Blackboard. You do **NOT** need to create an accounton MyOpenMath. Just access the weekly HW assignment from the unit folder under "Learning Units" in Blackboard.
- HW is due **by 11:59pm on Sundays**.
- You can try up to **2 versions** of each HW question and you have **3 attempts** on each version. If you get the question or part of it wrong, you can try the same question two more times. If you still want to improve, you can try a different version of the question.
- **Late submission penalty**: If you submit your homework late but by 11:59pm on the day before the next major exam opens, you will receive 80% of points earned.
- Your lowest HW score will be dropped.

Reading Assignments

- You will need to complete a weekly reading assignment in the form of interactive R Shiny tutorials with embedded knowledge check questions.
- The reading assignments will prepare you for the computing lab session on Thursdays and therefore are due before the lab session (**by 12:00noon on Thursday**).
- To complete the reading assignment for the week, click the reading assignment link in the week's unit folder under "Learning Units" in Blackboard.
- Each reading assignment is worth 20 points.
- You have unlimited time before the due date to complete the reading assignment.
- Follow instructions under the reading assignment to submit your work in Blackboard.
- Your lowest reading assignment score will be dropped.

Computing Lap Reports

- You will attend a weekly virtual lab on Thursday via Zoom (Starting in Week 2).
- During the lab session, you will follow the instructor's demo to complete some coding tasks on the lab template and prepare and submit your lab report for the session.
- Lab reports must be submitted in Blackboard **by 11:59pm on Thursday** (the day of the lab session).

Data Analysis project

- You will work on a semester-long data analysis project in a group of **3** students.
- Your group will be assigned a dataset to analyze throughout the semester with support and guidance from instructor and TAs.
- Your group will prepare and submit two progress reports and a final report.
- More details about the project will be given in class.

Participation

- Complete discussion board assignments by making and responding to posts in the Blackboard "Discussion Board".
- You will also receive participation points for completing the pre- and post- intro stats survey which will be announced in class and email invitations will be emailed to each student.
- You will also receive participation points for completing the pre- and post-diagnostics tests in Blackboard during First Week and Last Week of semester, respectively.
- Other participation points will be assigned for in-class participations.

Exams

- There will be **three midterm exams** (see course schedule) and a **comprehensive final exam**.
- Exams are closed book, closed notes. A formula sheet will be provided for each exam.
- **All exams will take place online through Blackboard Respondus Lockdown Browser.**
- **For exam dates, see course schedule.**

Course Policies

Use Of Blackboard as The Learning Management System

Blackboard is the primary online instructional and course communications platform. Students can access the course syllabus, assignments, grades, and learner support resources. Students are encouraged to protect their login credentials, complete a Blackboard orientation and log in daily to course.

[Other Course Polices]

Describe all policies specific to your course, e.g. Make-Up Exams, Late Work, Extra Credit, Special Assignments, Homework, Quizzes, Exams, Academic Dishonesty, Attendance/Participation, or Classroom Citizenship, etc. Utilize any headings as needed. Here you could include any common policies that you feel require special mention.

Make-Up Exams Only university-recognized excuses (e.g. Immediate Family Loss, Illness, etc.) are accepted. The student should notify the instructor at least 72 hours prior to missing quiz or exam and official documentation should be presented, e.g., Doctor's note. The student missing a quiz or exam with valid documented excuse is responsible to contact the instructor to reschedule for making-up missed quiz or exam.

Extra Credit Any extra credit opportunities will be announced in class and/or via Blackboard.

Late Work Late work shall not be accepted except due to a university-excused absence on due date.

Special Assignments There shall be no special assignments.

Class Schedule [or Course Plan]

Week	Date	Module Title	Sections	Assignments Due
Week 1	01/10-01/16	Ch1: Introduction to Data	1.1 Case study: using stents to prevent strokes 1.2 Data basics 1.3 Sampling principles and strategies 1.4 Experiments	**Sunday, 01/16 -** RA1, HW1, DB1
Week 2	01/17-01/23	**Monday,** 01/17	**Martin Luther King Day (no class)**	**Wednesday, 01/19 – Pre-Intro Stats Survey** **Wednesday, 01/19 – Pre-diagnostic Test** **Thursday, 01/20 -** RA2, CL1 **Friday, 01/21 -** DB2 **Sunday, 01/23 -** HW2, DB2
		Ch2: Summarizing Numerical Data	2.1 Examining Numerical Data Lab 1: Intro to R	
Week 3	01/24-01/30	Ch2: Summarizing Categorical Data	2.2 Considering Categorical Data 2.3 Case Study: Malaria Vaccine Lab 2: Exploratory Data Analysis Part 1	**Thursday, 01/27 -** RA3, CL2 **Sunday, 01/30 -** HW3
Week 4	01/31-02/06	**Exam 1**		Monday, 01/31
		Ch3: Probability I	3.1 Defining Probability 3.2 Conditional Probability Lab 3: Exploratory Data Analysis Part 2	**Thursday, 02/03 -** RA4, CL3 **Sunday, 02/06 -** HW4
Week 5	02/07-02/13	Ch3: Probability II	3.3 Sampling from a small population 3.4 Random Variables Lab 4: Introducing DAP Part I	**Thursday, 02/10 -** RA5, CL4 **Sunday, 02/13 -** HW5
Week 6	02/14-02/20	Ch4: Distributions of Random Variables	4.3 Binomial distribution 3.5 Continues distributions 4.1 Normal distribution Lab 5: Probability	**Thursday, 02/17-** RA6, CL5 **Friday, 02/18 –** DB3 **Sunday, 02/20 -** HW6, DB3
Week 7	02/21-02/27	Ch4: Distributions of Random Variables	4.1 Normal distribution (Cont'd) Lab 6: Normal Distribution	**Thursday, 02/24 –** RA7, CL6 **Sunday, 02/27 -** HW7
		Exam 2		Friday, 02/25
Week 8	02/28-03/06	Ch5: Foundations for Inference I	5.1 Point Estimates and Sampling Variability Lab 7: Sampling distributions	**Thursday, 03/03 –** RA8, CL7 **Sunday, 03/06 – Submit DAP Part I**
		Friday 03/04	**Wellness Day (no class)**	
Week 9	03/07-03/13	**Spring Break**		
Week 10	03/14-03/20	Ch5: Foundations for Inference II	5.2 Confidence Intervals for a proportion 5.3 Hypothesis Testing for a Proportion Lab 8: Confidence Intervals for proportions	**Thursday, 03/17 -** CL8 **Sunday, 03/20 –** HW8
Week 11	03/21-03/27	**Tuesday,** 03/22	**College of S. &T. Wellness Day**	**Thursday, 03/24 -** RA9, CL9
		Ch5: Foundations for Inference III	5.3 Hypothesis Testing for a Proportion Lab 9: Inference for categorical data	
Week 12	03/28-04/03	**Tuesday,** 03/29	**University Mental Health Day (no class)**	**Thursday, 03/31 -** RA10, CL10 **Sunday, 04/03 –** HW9
		Ch6: Inference for Categorical Data	6.1 Inference for a Single Proportion 6.2 Difference of Two Proportions Lab 10: Introducing DAP Part II	
Week 13	04/04-04/10	Ch7: Inference for Numerical Data I	7.1 One-sample Means with the t-distribution 7.2 Paired Data Lab 11: Inference for Numerical data	**Thursday, 04/07 -** RA11, CL11 **Sunday, 04/10 -** HW10
Week 14	04/11-04/17	**Friday,** 04/15	**University Holiday (no class)**	
		Ch7: Inference for Numerical Data II	7.3 Difference of Two Means Review for Exam 3	
		Exam 3		**Wednesday, 04/13 – Thursday, 04/14**
		Monday, 04/18	**University Mental Health Day (no class)**	

Week 15	04/18-04/24	Ch8: Introduction to linear regression I	8.1 Fitting a line, residuals, and correlation 8.2 Lease squares regression Lab 12: Introducing DAP Part III	**Thursday, 04/21 -** CL12
Week 16	04/25-05/01	Ch8: Introduction to linear regression II	8.4 Inference for linear regression 9.1 Introduction to multiple regression Lab 13: Correlation and Linear Regression	**Thursday, 04/28 -** RA12, CL13 **Friday, 04/29 – Post-Intro Stats Survey** **Friday, 04/29 – Post-diagnostic Test** **Sunday, 05/01 -** HW11
Week 17	05/02-05/06	Ch9: Multiple and logistic regression I	9.2 Model Selection 9.3 Checking model conditions using graphs Lab 14: Multiple linear regression	**Thursday, 05/05 -** RA13, CL14 **Sunday, 05/08 -** HW12
	05/09-05/13	**Final Exam**		**Opens on Tuesday, 05/10 at 12:05am and closes on Wednesday, 05/11 at 11:59pm**

** These descriptions and timelines are subject to change at the discretion of the instructor.*

Please refer to the Common Policies file for all other University policies. The Common Policies should also be provided to all students or available in the course Blackboard shell.