|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Statistics & Analytics* | **17-MAR-2021** | **5-MAY-2021** |  |  | **Wednesday** | **7:30 pm - 10:00 pm ET** |  |

**MVIS 5301.01 *Statistics & Analytics* [3.0 credits]**

**MAR 17 – MAY 5, 2021**

**Faculty: Mr. John T. Miller**

**Live Class Day/Time: Wednesdays 7:30 pm - 10:00 pm ET** (3/17, 3/24, 3/31, 4/7, 4/14, 4/21, 4/28, 5/5)

**FACULTY COMMUNICATION**

* **MICA Email:** [*jmiller04@mica.edu*](mailto:jmiller04@mica.edu)
* **Phone: 240-755-7016**
* **Office Hours: By Appointment**

The best way to reach me is by email at *jmiller04@gmail.com*. My policy is to respond to student emails within 24 hours, if not sooner. The next best way is to txt message me at **240-755-7016**. If necessary, you can call directly, and please leave a detailed voicemail message. Unfortunately, I often have my cellphone turned off between the hours of 6 am and 2 pm due to work requirements so I appreciate your patience.

***I will respond to all communications within 24 hours of receipt.***

I am available for additional 1:1 sessions via Zoom or Google Hangouts. The students are expected to take the initiative in contacting me for any additional sessions.

**COURSE DESCRIPTION**

This course introduces students to research methods for gathering and analyzing information and data that are central to visualization. Topics will include data mining, data aggregation, grouping methods, data cleaning, and evaluation. Students will explore tools and applications that help establish and/or understand statistical relationships with data and information and how to begin transforming research into visualizations.

R (<https://www.r-project.org/>) is a free, open source language and environment for statistical computing and graphics. R has an incredibly rich set of capabilities, and this course will focus on loading data, performing basic statistical operations, and producing graphical outputs. By using the functionality provided by the tool students will understand the entire pipeline of data processing and the connections between data processing concepts and their visual representations. Students wishing to use something other than R for homework or presentations must obtain permission of the instructor.

**STUDENT LEARNING OBJECTIVES**

After completion of the course the student will be able to:

1. Download, prepare, and verify a data set

2. Calculate and describe statistical summaries of the data

3. Examine relationships between the data elements

4. Produce a summary report explaining the contents of a data set

**REQUIRED MATERIALS**

Below are required textbooks, software, or other resources necessary to instruction and learning in this course.

* “Learning Statistics with R (≥V0.5)”, Dan Navarro; Available as a pdf and in print form.  
  <https://learningstatisticswithr.com/>
* “R for Data Science”, Hadley Wickham & Garrett Grolemund; <http://r4ds.had.co.nz/>
* “The R Project for Statistical Computing (V4.0.2)” <http://www.r-project.org/index.html>
* “R Studio”, (RStudio Desktop V1.3) <http://www.rstudio.com/>

These are available free on Windows, Mac, and Linux. The course examples will be done using the Windows based versions of these tools

Students will also need to be able to open and do simple edits with .XLSX files and use a text editor (e.g. <http://notepad-plus-plus.org/> , <http://www.sublimetext.com/> or <http://www.jedit.org/> (NOT MSWord).

**Optional / Other Useful Resources**

* “R Graphics Cookbook”, Winston Chang; <http://www.cookbook-r.com/>
* “R Cookbook”, Paul Teetor; <http://shop.oreilly.com/product/0636920174851.do>
* “Statistics (the easy easy) with R”, Nicole Radziwill, <http://www.radziwill.cc/> (I love this book)
* “MBA Fundamentals: Statistics”, Paul Thurman.
* “Understanding Basic Statistics”, Charles and Corrinne Brase. (It was used in previous versions of this class; I leave slides referencing it up on our Canvas course site. It explains the fundamentals very well)
* [Google](https://www.google.com/)
* Stack Overflow. (A fantastic resource for debugging.)
* StatQuest with Josh Starmer, <https://www.youtube.com/user/joshstarmer> (Corny and effective.)
* Subreddit, Data is Beautiful, <https://www.reddit.com/r/dataisbeautiful/>
* #MakeoverMonday is a weekly data visualization challenge with hundreds of passionate data people. Use the hashtag on YouTube and Twitter to see participants’ entries or check out <https://www.makeovermonday.co.uk/data/> It can be a lot of Tableau but the principles are the same
* R-Bloggers, excellent for all skill levels <https://www.r-bloggers.com/>

**LIBRARY GUIDE AND OTHER ADDITIONAL RESOURCES**

Students can find additional resources for this course and the program at [**http://libguides.mica.edu/infovis**](http://libguides.mica.edu/infovis)**.** There is a section on “Find Data” which you may find especially useful for the final project.

**COURSE EXPECTATIONS**

The live (synchronous) class takes place every **WEDNESDAY** from **7:30 pm - 10:00 pm ET**.

* + - Students are expected to attend all meetings of each class. Unexcused absences from as few as two classes will result in a failing grade.
    - Students must notify faculty prior to being absent or as soon as possible after an absence.
    - Students will be responsible for all activities and assignments during their absence, including watching the live lecture recording.

The course consists of readings, student-driven discussion and presentations, online activities and discussion forums, and hands-on assignments.

* Readings, resources, and assignments will be available on SUNDAY of each course week.
* Students are responsible for completing on time to ensure their ability to participate appropriately and progress successfully in the program.

Students are expected to observe the key rules of “netiquette” as detailed in the **MICA Policies + Resources** area. Courtesy drives all course policies. Courtesy means being available for every class, arriving for live sessions on time, handing in well-prepared assignments on time, and participating in class discussions in a cooperative, collegial spirit.

During the lectures, it is common to have an open chat window for all participants. Please *do not send me any private chat message during class*, as it can distract me from the lecture. A general chat question or asking a question at a pause is always appreciated. We learn via your questions. Please do send private emails anytime, and I will respond ASAP.

**Assignments**

Part of your grade is how well you adhere to deadlines. Unless otherwise noted, assignments will be due by the following **TUESDAY by 8:00 p.m. EST**. Late assignments will receive the grade of zero points, as I will be reviewing the solutions in class.

Because of the nature of this online course, please be advised that “I had a computer problem” or “I was not connected” —and so on—are not valid excuses for not having an assignment on deadline, so plan ahead. Assignments will only be accepted using the Moodle learning environment; no assignments will be accepted via email.

If you do have any issues, problems, barriers, etc. *please contact me ASAP*. Given the remote / on-line nature of this program, it is imperative that we all communicate via postings, chats, and email.

Course assignments should be completed using the software programs cited in the assignment description.

* + All assignments should be uploaded using the following naming convention:

FirstLast\_AssignmentTitle-#\_YYYY-MM-DD

ex.

JohnMiller\_ ILoveMVIZ-1\_2020-11-02.pdf

JohnMiller\_ ILoveMVIZ-1\_2020-11-02.zip

* + Assignments completed using word processing software (Microsoft Word, Google Drive document, etc.) should have 1 inch margins on all sides and the document should be double-spaced (see the MLA model.) The assignment heading will be in the upper left corner of the page and should include student’s full name, assignment name, and due date.
* Documents should be converted to PDF
  + Assignments completed using the statistical s/w and associated results should be placed in a single .ZIP file using the above naming convention.
* All documents and program should be placed inside a folder using the following naming conventions and directory structure:
  + FirstLast\_AssignmentTitle-#\_YYYY-MM-DD
  + JohnMiller\_ILoveMVIZ-1\_2020-11-02 *(folder)*
    - myRFile.R *(file)*
    - myMarkDownFile.RMD (*file*)
    - dataset.csv *(file)*
    - datset.xlsx *(file)*
    - write-up.pdf *(file)*

The instructor should be able to unzip the file into a single top-level folder and run the programs within.

**Final Project**

During Week 8, during a final presentation of 8 minutes, participants will demonstrate their ability to answer a business question and provide a clear recommendation using data in R (ask me for an exception if you prefer another coding language). There should be a clear demonstration of competence in the class objectives, including an exploratory data analysis component and engaging graphics.

**Student Participation**

Regular, informed class participation is a weighted portion of the Overall Course Grade (*see below*). The minimum expectation is that students attend all live sessions and participate meaningfully in the sessions through the chat feature or by using a microphone headset as specified in the program’*s Technical Requirements* document.

Courses that have discussion forums, group sessions, a wiki, a glossary, or any other assignment that the faculty has created to inform and benefit the entire cohort will be graded both as the individual assignment and the student’s meaningful contribution will be included as a portion of the course participation grade.

By meaningful participation/contribution, the student is expected to add to or build upon, and ultimately further inform, the class discussion through research, experience, and/or access to resources that are appropriate to the topic of discussion and beneficial to the entire cohort.

**OVERALL COURSE GRADING**

* Class attendance, preparation, and participation (10%)
* Short assignments, discussion forum posts, and quizzes (10%)
* Four weekly projects (50%, or 12.5% each)
* Final Project (30%)

The final project and weekly projects are weighted more so that as you gain more experience and confidence in R, you have an opportunity do well in the course.

* Students must receive a B or better to pass a course and to maintain a B or better in all coursework to continue in the program.
* Assignment grades will be completed and posted in the gradebook approximately 1 week after the assignment due date.
* Final course grades will be available via the MICA Portal Unofficial Transcript no later than 2 weeks after the course end date.

**MICA’s ACADEMIC POLICY**

**Academic Policy statements are** published online in MICA’s Academic Bulletin: http://www.mica.edu/Programs\_of\_Study/Academic\_Bulletin.html, including a definition of plagiarism, ADA Compliance, and guidelines for students with extended illness or cause for legitimate absences.

**CREDITS**

The instructor is greatly appreciative of Mr. Rob Rolleston, MICA faculty, in providing a framework for this course that best benefits our students. Much of this content represent his ideas.

**COURSE SCHEDULE**

**Week-0 (NA)**

* Subject: Set-Up Tools & Getting Head Start
* Assignment: Due Date **Mar-17, e.g. first day of class**
  + Install R and R-Studio, see Pre-Class Assignment on class website
  + Accept invitation to join class at <https://www.datacamp.com>
    - You should have access for 6-months, starting MAR-2020
  + *Optional –Other students have found these useful. There is a lot of material, for many different topics.* 
    - Feel free to playing around. Some good videos for this course are:
      * Introduction to the Tidyverse
      * Data Manipulation with dplyr
      * Introduction to Data Visualization with ggplot2
    - Especially if you have not done any programming, I encourage you to start working thru these courses:
      * Introduction to R
      * Reporting with R Markdown
      * There is no “Intro to RStudio IDE.” However other instructors find this useful:

<https://www.youtube.com/channel/UCctc3RvC6n5dJXP_NR7wfjQ/videos>  
Look for “Introduction to R and RStudio” by Hefin Rhys

**Week-1 (03/17)**

* Subjects:
  1. Introductions
  2. Getting Started
* Some Suggested Readings:
  + "Learning Statistics with R"
    - section 2.2: "Scales of Measurement"
    - chapter 3: "Getting Started with R"
    - chapter 4: "Additional R Concepts"
    - chapter 8.1: “Basic Programming 8.1 Scripts”
  + “R for Data Science” ( <https://r4ds.had.co.nz/> )
    - Welcome-Introduction (<https://r4ds.had.co.nz/introduction.html> )
    - Explore-Workflow: Basics (<https://r4ds.had.co.nz/introduction.html>)
    - Explore-Workflow: Scripts (<https://r4ds.had.co.nz/workflow-scripts.html> )
    - Communicate-R Markdown (<https://r4ds.had.co.nz/r-markdown.html> )
* Assignment#1: Due Date: MAR-23 8:00PM EST

<https://r4ds.had.co.nz/workflow-basics.html#exercises-7>

(.R file with answers commented ### out, answer all of 4.4 questions)

<https://r4ds.had.co.nz/workflow-scripts.html#exercises-14>

(in that same .R file list one tip and one common mistake)

**Week-2 (03/24)**

* Subjects:
  1. Getting & Organizing Data
  2. Make a Plot
* Some Suggested Readings:
  + "Learning Statistics with R"
    - Chapter 6: “Drawing Graphs” … only if you want to learn about “Base Graphics”
  + “R for Data Science” ( <https://r4ds.had.co.nz/> )
    - Explore-Data Visualization (<https://r4ds.had.co.nz/data-visualisation.html> )
    - Tidy Data (<https://r4ds.had.co.nz/tidy-data.html> )
    - Data Import (<https://r4ds.had.co.nz/data-import.html> )
* Assignment#2: Due Date: MAR-30 8:00PM EST

(taken from “R for Data Science” by Wickham)

“Basic Visualization”

<https://r4ds.had.co.nz/data-visualisation.html#exercises-3>

Exercises 3.6 (all except for 3.6.6)

<https://r4ds.had.co.nz/tidy-data.html#case-study>

Exercises 12.6.1 (all)

You need to read through the case study to work the questions. You can do it!

**Week-3 (03/31)**

* Subjects:
  1. Categorical Data
  2. Reducing Data
* Some Suggested Readings:
  + “Learning Statistics with R”
    - Chapter 7: “Pragmatic Matters”
    - Chapter 12: “Categorical data analysis”
  + “R for Data Science” ( <https://r4ds.had.co.nz/> )
    - Wrangle-Introduction (<https://r4ds.had.co.nz/wrangle-intro.html> )
    - Wrangle-Factors (<https://r4ds.had.co.nz/factors.html> )
    - Wrangle-Relations data (<https://r4ds.had.co.nz/relational-data.html> )
* Assignment #3: Due Date: APR-06 8:00PM EST

(taken from “R for Data Science” by Wickham)

“Manipulating Data”

<https://r4ds.had.co.nz/factors.html>

Exercises 15.3.1 (all)

<https://r4ds.had.co.nz/relational-data.html>

Exercises 13.2.1 (all)

* Final Project Assignment:
  1. Data Choice Due APR-14 8:00PM EST (WK5)
  2. Exploratory Summary Due APR-27 8:00PM EST (WK7)
  3. Presentation Due MAY-03 8:00PM EST (WK8)
  4. Critiques Due MAY-07 Friday 8:00PM EST (WK8)

**Week-4 (04/07)**

* Subjects:
  1. Descriptive Statistics
  2. Numbers & Distributions
* Some Suggested Readings:
  + "Learning Statistics with R"
    - Chapter 5: "Descriptive Statistics" (up through section 5.6)
    - Chapter 9: "Introduction to Probability”
  + “R for Data Science” ( <https://r4ds.had.co.nz/> )

*Re-read previously assigned sections, good time to review*

Assignment#4: Due Date: APR-13 8:00PM EST

(taken from “R for Data Science” by Wickham)

“Descriptive Statistics & Distributions”

<https://r4ds.had.co.nz/exploratory-data-analysis.html#exercises-15>

Exercises 7.3.4 (1-3 only)

* **Project-I:** Data Choice Due APR-14 8:00PM

**Week-5 (04/14)**

* Subjects:
  1. Scatterplots & Lines
  2. Errors and Confidence
* Some Suggested Readings:
  + "Learning Statistics with R"
    - Chapter 5 "Descriptive Statistics" (sect. 5.7-end)
    - Chapter 15: "Linear Regression", through section 15.4
    - Chapter 10: "Estimating population parameters from a sample"
* Assignment#5: Due Date: APR-20 8:00PM EST

(taken from “Learning Statistics with R” by Navarro)

* “Linear Regression”

Conduct a regression on a dataset of your choice using the Lm() function, p. 460 Provide code (with # comments as needed) and output.

* **Project-II:** Data Exploratory Summary Due APR-27 8:00PM EST

**Week-6 (04/21)**

* Subjects:
  1. Tests of Differences
  2. Hypothesis Testing
* Some Suggested Readings:
  + “Learning Statistics with R”
    - Chapter 11: “Hypothesis testing”
    - Chapter 13: “Comparing two means”
  + “R for Data Science” ( <https://r4ds.had.co.nz/> )
* Assignment #6: Due Date: APR-27 8:00PM EST

(taken from “Learning Statistics with R” by Navarro)

* + Hypothesis Testing & Confidence Intervals

1. Write a possible null and alternate hypothesis related to question about your data or career field. Provide a rough outline of how you would set up a test in R.

2. Define p-values and confidence intervals.

**Week-7 (04/28)**

* Subjects:
  1. Maps
  2. Interactive Rmarkdown & Dashboards
* Some Suggested Readings:
  + “R for Data Science” ( <https://r4ds.had.co.nz/> )
    - Communicate-R Markdown formats

(<https://r4ds.had.co.nz/r-markdown-formats.html> )

* Assignment #7: Review what you have learned. Congratulations, you have come a long way.
  + No Assignment to turn in.
* **Project-III:** Presentations Due MAY-03 8:00PM EST

**Week-8 (05/05)**

* Subject: Student Presentations
* **Project-IV**: Critique other presentations and class in general
  + Due Date: MAY-07 10:00pm EST **–NOTE: DUE Friday after last class—**