**MAT 114 – INTRODUCTION TO DATA SCIENCE** BEGINNING FALL 2020

3 hrs./wk. – 3 cr.

Catalog Description: Introduction to Data Science will provide students with data literacy skills in order to understand techniques in data manipulation, visualization and interpretation. This project based course will allow students to utilize a toolkit of statistical software to perform data science methods. Ethical issues related to data privacy, authenticity and security will be addressed alongside an introduction to artificial intelligence.

Prerequisite: MAT 016, MAT 120, or equivalent

Text: R for Data Science Authors: Garrett Grolemund and Hadley Wickham Free for students: <https://r4ds.had.co.nz/>

Big Data Demystified Author: David Stephenson PhD

Syllabus

|  |  |  |
| --- | --- | --- |
| Suggested Timeline | Text Chapter | Topics |
| Week 1 | 1 –2 | Introduction- The Data Science Model |
| Week 2 | 3.1 – 3.4 | Data Visualization-Grammar of Graphics, problems with data |
| Week 3 | 3.4 – 3.10 | Data Visualization- Facets, objects, transformations |
| Week 4 | 4.1 – 4.4 | Workflow Basics, coding, naming, functions |
| Week 4 | 28.2.1 and 28.3 | Graphics Communication |
|  |  | Quiz 1-ggplot() |
| Week 5/6 | 12.1 – 12.7 | Tidy Data |
| Week 7 | 18 | Pipes |
|  |  | Quiz 2 Data Cleaning |
|  |  | Midterm Exam Swirl Project |
| Week 8/9 | 5.1 to 5.7 | Data Transformation, filters, sub- setting, grouping, missing values |
| Week 9 | 6 | Workflow Scripts |
| Week 10 | 13 | Relational Data- Filtering and joins |
|  |  | Quiz 3 Joins |
| Week 11 | 23.4.1 | Categorical Modeling |
| Week 12 |  | Cluster Analysis |
| Week 13 |  | Quiz 4 Modeling |
| Week 14/15 |  | Final Exam Project Due |
|  |  |  |
|  |  |  |
| Integrated into the  Course | 14-16 | Strings, Factors and Dates and Times |
| Additional Technology |  | Tableau |

Students are expected to adhere to the policies of the County College of Morris. These can be accessed at [www.ccm.edu/academics/academic-policies/](http://www.ccm.edu/academics/academic-policies/).

**Statement of Course LEARNING OUTCOMES**

* Produce and interpret data visualizations, including dashboards, graphs, charts and maps to describe and explore large data sets.
* Produce and interpret numerical summaries to describe and explore large data.
* Investigate and explore relationships between more than two variables, multivariate analysis.
* Solve problems utilizing programming languages for data scientists.
* Apply exploratory data analysis and transformations to identify trends and problems within the data
* Clean and prepare data for analysis.
* Identify problems with messy and missing data.
* Recognize questions and problems that can be investigated using data.
* Communicate findings and outcomes based on data science techniques.
* Explain issues related to data privacy, security and authenticity.
* Explain advances in artificial intelligence.