

**1. Course number and name**

CptS 475: Data Science

**2. Credits and contact hours**

3 credits, 3 lecture hours

**3. Instructor's or course coordinator's name**

Assefaw Gebremedhin

**4. Textbook, title, author, and year**

The course does *not* require or use a textbook. Instead select chapters from a variety of books (listed below) and other resources are used as starting points and are supplemented by instructor developed lecture notes and reading material. The lecture notes and reading material are made available on the course's website. Below is a list of (the more frequently) referenced books:

G. James, D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. Springer. ISBN 978-1461471370. (Freely available online)

C. O'Neil and R. Schutt. *Doing Data Science, Straight Talk from The Frontline*. O'Reilly, 2014. ISBN 978-1-449-35865-5.

H. Wickham and G. Grolemund. *R for Data Science*. O'Reilly, 2017. ISBN 978-1-491-91039-9. (Freely available online)

J. Leskovek, A. Rajaraman and J. Ullman. 2014. *Mining of Massive Datasets*. v 2.1, Cambridge University Press. (Freely available online)

N. Yao. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Wiley Publications. ISBN-13: 978-0470944882.

I. Goodfellow, Y. Bengio, and A. Courville. 2016. *Deep Learning*. MIT Press. ISBN 9780262035613. (Freely available online)

**5. Specific course information**

a. *Catalog description*: The data science process, data wrangling, exploratory data analysis, linear regression, classification, clustering, principal components analysis, recommender systems, data visualization, data and ethics, and effective communication.

b. *Prerequisites or corequisites*: Familiarity with algorithm design and analysis, basic linear algebra, basic probability and statistics.

**6. Specific goals for the course**

By the end of the course, students will be able to

- Describe what Data Science is and the skill sets needed (1a).
- Describe the Data Science Process (1a, 1c, 1e).
- Use R to carry out statistical modeling and analysis (2a, 2g, 6a).
- Carry out exploratory data analysis (1e, 6a, 6d).
- Use effective data wrangling approaches to manipulate data (6a).
- Apply machine learning algorithms for predictive modeling (2a, 2b, 6a, 6c).
- Apply effective resampling methods to assess model performance (2e, 6b).

- Apply learning and mining methods to discover patterns, trends and anomalies in data (6a, 6b, 6c, 6d).
- Create effective visualization of data to communicate or persuade (3c, 3d, 3e).
- Reason around ethical and privacy issues in data science conduct, and apply ethical practices (4a, 4f).
- Work effectively in teams on data science projects (5b, 5d, 5g).
- Apply knowledge gained in the course to carry out a project and write a technical report and produce a presentation (1a-e, 2a-c, 2e, 2f, 3a-e, 5b, 5d, 5g, 6a-d, 7a-g).

## **7. Brief list of topics to be covered**

- What is Data Science?
- Introduction to R
- Exploratory Data Analysis
- The Data Science Process
- Data Wrangling
- Linear Regression
- Classification
- Resampling Methods
- Unsupervised Learning
- Data Visualization
- Tree-based Methods
- Time-series Data Mining
- Intro to Deep Learning
- Data Science and Ethical Issue