

APPLIED DATA SCIENCE (DSCI)

DSCI 330. Cognition and Computation. 3 Units.

An introduction to (1) theories of the relationship between cognition and computation; (2) computational models of human cognition (e.g. models of decision-making or concept creation); and (3) computational tools for the study of human cognition. All three dimensions involve data science: theories are tested against archives of brain imaging data; models are derived from and tested against datasets of e.g., financial decisions (markets), legal rulings and findings (juries, judges, courts), legislative actions, and healthcare decisions; computational tools aggregate data and operate upon it analytically, for search, recognition, tagging, machine learning, statistical description, and hypothesis testing. Offered as COGS 330, COGS 430, DSCI 330 and DSCI 430.

DSCI 332. Spatial Statistics for Near Surface, Surface, and Subsurface Modeling. 3 Units.

This course is on spatial modeling of near surface, surface, and subsurface data, also known as geostatistical modeling. Spatial modeling has its origins in predictive modeling of minerals in subsurface formations, from which many examples are used in this class. Students will learn the basics of spatial models in order to understand how they are built from various data types and how their uncertainties are assessed and risk reduced. Students will be expected to learn the rudimentary navigation of R Studio, execute pre-written publically available R code (provided), and make simple modifications. Graduate students will be expected to learn the above and develop a 10 week modeling project focused on the use of spatial modeling methods with R using data relevant to their specific discipline or interest. These projects will include preparing datasets to be executed in R code scripts. Resulting scripts will be placed in a git repository for use by other students as open source resources along with documentation demonstrating the reproducible spatial modeling science and analyses for these problems. Geostatistical (spatial) mapping is applicable across many disciplines. Examples of graduate projects from previous classes include subsurface modeling (geology), earthquake mapping (geophysics/civil engineering), soil stability modeling (civil engineering), aquifer characterization (hydrology), and pollution/contaminant mapping (environmental studies/medicine). Offered as DSCI 332 and DSCI 432.

DSCI 351. Exploratory Data Science. 3 Units.

In this course, we will learn data science and analysis approaches to identify statistically significance relationships and better model and predict the behavior of these systems. We will assemble and explore real-world datasets, perform clustering and pair plot analyses to investigate correlations, and logistic regression will be employed to develop associated predictive models. Results will be interpreted, visualized and discussed. We will introduce basic elements of statistical analysis using R Project open source software for exploratory data analysis and model development. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and munging functions, and a rich selection of statistical packages, used for data analytics, model development and prediction. This will include an introduction to R data types, reading and writing data, looping, plotting and regular expressions, so that one can start performing variable transformations for linear fitting and developing structural equation models, while exploring for statistically significant relationships. The M section of DSCI 351 is for students focusing on Materials Data Science. Offered as DSCI 351, DSCI 351M and DSCI 451. Prereq: (ENGR 130 or ENGR 131 or CSDS 132 or ECSE 132 or DSCI 134) and (STAT 312R or STAT 201R or SYBB 310 or PQHS 431).

DSCI 351M. Exploratory Data Science. 3 Units.

In this course, we will learn data science and analysis approaches to identify statistically significance relationships and better model and predict the behavior of these systems. We will assemble and explore real-world datasets, perform clustering and pair plot analyses to investigate correlations, and logistic regression will be employed to develop associated predictive models. Results will be interpreted, visualized and discussed. We will introduce basic elements of statistical analysis using R Project open source software for exploratory data analysis and model development. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and munging functions, and a rich selection of statistical packages, used for data analytics, model development and prediction. This will include an introduction to R data types, reading and writing data, looping, plotting and regular expressions, so that one can start performing variable transformations for linear fitting and developing structural equation models, while exploring for statistically significant relationships. The M section of DSCI 351 is for students focusing on Materials Data Science. Offered as DSCI 351, DSCI 351M and DSCI 451. Prereq: (ENGR 130 or ENGR 131 or CSDS 132 or ECSE 132 or DSCI 134) and (STAT 312R or STAT 201R or SYBB 310 or PQHS 431).

DSCI 352. Applied Data Science Research. 3 Units.

This is a project based data science research class, in which project teams identify a research project under the guidance of a domain expert professor. The research is structured as a data analysis project including the 6 steps of developing a reproducible data science project, including 1: Define the ADS question, 2: Identify, locate, and/or generate the data 3: Exploratory data analysis 4: Statistical modeling and prediction 5: Synthesizing the results in the domain context 6: Creation of reproducible research, Including code, datasets, documentation and reports. During the course special topic lectures will include Ethics, Privacy, Openness, Security, Ethics. Value. The M section of DSCI 352 is for students focusing on Materials Data Science. Offered as DSCI 352, DSCI 352M and DSCI 452. Prereq: (CSDS 133 or DSCI 134 or ENGR 130 or ENGR 131 or CSDS 132 or ECSE 132) and (STAT 312R or STAT 201R or SYBB 310 or PQHS 431 or OPRE 207) and (DSCI 351 or (SYBB 311A and SYBB 311B and SYBB 311C and SYBB 311D) or SYBB 321 or MKMR 201).

DSCI 352M. Applied Data Science Research. 3 Units.

This is a project based data science research class, in which project teams identify a research project under the guidance of a domain expert professor. The research is structured as a data analysis project including the 6 steps of developing a reproducible data science project, including 1: Define the ADS question, 2: Identify, locate, and/or generate the data 3: Exploratory data analysis 4: Statistical modeling and prediction 5: Synthesizing the results in the domain context 6: Creation of reproducible research, including code, datasets, documentation and reports. During the course special topic lectures will include Ethics, Privacy, Openness, Security, Ethics. Value. The M section of DSCI 352 is for students focusing on Materials Data Science. Offered as DSCI 352, DSCI 352M and DSCI 452. Prereq: (CSDS 133 or DSCI 134 or ENGR 130 or ENGR 131 or CSDS 132 or ECSE 132) and (STAT 312R or STAT 201R or SYBB 310 or PQHS 431 or OPRE 207) and (DSCI 351 or (SYBB 311A and SYBB 311B and SYBB 311C and SYBB 311D) or SYBB 321 or MKMR 201).

DSCI 353. Data Science: Statistical Learning, Modeling and Prediction. 3 Units.

In this course, we will use an open data science tool chain to develop reproducible data analyses useful for inference, modeling and prediction of the behavior of complex systems. In addition to the standard data cleaning, assembly and exploratory data analysis steps essential to all data analyses, we will identify statistically significant relationships from datasets derived from population samples, and infer the reliability of these findings. We will use regression methods to model a number of both real-world and lab-based systems producing predictive models applicable in comparable populations. We will assemble and explore real-world datasets, use pair-wise plots to explore correlations, perform clustering, self-similarity, and logistic regression develop both fixed-effect and mixed-effect predictive models. We will introduce machine-learning approaches for classification and tree-based methods. Results will be interpreted, visualized and discussed. We will introduce the basic elements of data science and analytics using R Project open source software. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and assembly functions, and a rich selection of statistical packages, used for data analytics, model development, prediction, inference and clustering. With this background, it becomes possible to start performing variable transformations for linear regression fitting and developing structural equation models, fixed-effects and mixed-effects models along with other statistical learning techniques, while exploring for statistically significant relationships. The class will be structured to have a balance of theory and practice. We'll split class into Foundation and Practicum a) Foundation: lectures, presentations, discussion b) Practicum: coding, demonstrations and hands-on data science work. The M section of DSCI 353 is for students focusing on Materials Data Science. Offered as DSCI 353, DSCI 353M and DSCI 453.

DSCI 353M. Data Science: Statistical Learning, Modeling and Prediction. 3 Units.

In this course, we will use an open data science tool chain to develop reproducible data analyses useful for inference, modeling and prediction of the behavior of complex systems. In addition to the standard data cleaning, assembly and exploratory data analysis steps essential to all data analyses, we will identify statistically significant relationships from datasets derived from population samples, and infer the reliability of these findings. We will use regression methods to model a number of both real-world and lab-based systems producing predictive models applicable in comparable populations. We will assemble and explore real-world datasets, use pair-wise plots to explore correlations, perform clustering, self-similarity, and logistic regression develop both fixed-effect and mixed-effect predictive models. We will introduce machine-learning approaches for classification and tree-based methods. Results will be interpreted, visualized and discussed. We will introduce the basic elements of data science and analytics using R Project open source software. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and assembly functions, and a rich selection of statistical packages, used for data analytics, model development, prediction, inference and clustering. With this background, it becomes possible to start performing variable transformations for linear regression fitting and developing structural equation models, fixed-effects and mixed-effects models along with other statistical learning techniques, while exploring for statistically significant relationships. The class will be structured to have a balance of theory and practice. We'll split class into Foundation and Practicum a) Foundation: lectures, presentations, discussion b) Practicum: coding, demonstrations and hands-on data science work. The M section of DSCI 353 is for students focusing on Materials Data Science. Offered as DSCI 353, DSCI 353M and DSCI 453.

DSCI 354. Data Visualization and Analytics. 3 Units.

Data Visualization and Analytics students will learn data visualization and analytics techniques focused on different types of data such as time-series, spectral, or image data science problems. This class will focus on increasing analysis of complex data sets through visualization by enhancing exploratory data analysis and data cleaning. This class will focus on creating effective data visualizations to communicate data analytics results to different audiences. Different datasets will be provided to develop different types of visualizations and analytics. Types of data visualizations include in interactive plots (e.g., bar graphs change over time), applications that allow users to adjust the visualizations based on their decisions (e.g., shiny applications), interactive maps, 3-D plots of data, etc. Discussing how an audience understands information and brings in data as well as the ethics of making data visualizations will be discussed. The class will also include ways to increase modeling and analysis with effective visualizations for credible, data-driven decision making. This will include a git repository for other students to use these codes as open source resources and the preparation of reproducible data science analyses for different types of problems. Offered as DSCI 354, DSCI 354M, and DSCI 454. Prereq: (DSCI 351 or DSCI 351M) and (DSCI 353 or DSCI 353M).

DSCI 354M. Data Visualization and Analytics. 3 Units.

Data Visualization and Analytics students will learn data visualization and analytics techniques focused on different types of data such as time-series, spectral, or image data science problems. This class will focus on increasing analysis of complex data sets through visualization by enhancing exploratory data analysis and data cleaning. This class will focus on creating effective data visualizations to communicate data analytics results to different audiences. Different datasets will be provided to develop different types of visualizations and analytics. Types of data visualizations include in interactive plots (e.g., bar graphs change over time), applications that allow users to adjust the visualizations based on their decisions (e.g., shiny applications), interactive maps, 3-D plots of data, etc. Discussing how an audience understands information and brings in data as well as the ethics of making data visualizations will be discussed. The class will also include ways to increase modeling and analysis with effective visualizations for credible, data-driven decision making. This will include a git repository for other students to use these codes as open source resources and the preparation of reproducible data science analyses for different types of problems. Offered as DSCI 354, DSCI 354M, and DSCI 454. Prereq: (DSCI 351 or DSCI 351M) and (DSCI 353 or DSCI 353M).

DSCI 430. Cognition and Computation. 3 Units.

An introduction to (1) theories of the relationship between cognition and computation; (2) computational models of human cognition (e.g. models of decision-making or concept creation); and (3) computational tools for the study of human cognition. All three dimensions involve data science: theories are tested against archives of brain imaging data; models are derived from and tested against datasets of e.g., financial decisions (markets), legal rulings and findings (juries, judges, courts), legislative actions, and healthcare decisions; computational tools aggregate data and operate upon it analytically, for search, recognition, tagging, machine learning, statistical description, and hypothesis testing. Offered as COGS 330, COGS 430, DSCI 330 and DSCI 430.

DSCI 432. Spatial Statistics for Near Surface, Surface, and Subsurface Modeling. 3 Units.

This course is on spatial modeling of near surface, surface, and subsurface data, also known as geostatistical modeling. Spatial modeling has its origins in predictive modeling of minerals in subsurface formations, from which many examples are used in this class. Students will learn the basics of spatial models in order to understand how they are built from various data types and how their uncertainties are assessed and risk reduced. Students will be expected to learn the rudimentary navigation of R Studio, execute pre-written publically available R code (provided), and make simple modifications. Graduate students will be expected to learn the above and develop a 10 week modeling project focused on the use of spatial modeling methods with R using data relevant to their specific discipline or interest. These projects will include preparing datasets to be executed in R code scripts. Resulting scripts will be placed in a git repository for use by other students as open source resources along with documentation demonstrating the reproducible spatial modeling science and analyses for these problems. Geostatistical (spatial) mapping is applicable across many disciplines. Examples of graduate projects from previous classes include subsurface modeling (geology), earthquake mapping (geophysics/civil engineering), soil stability modeling (civil engineering), aquifer characterization (hydrology), and pollution/contaminant mapping (environmental studies/medicine). Offered as DSCI 332 and DSCI 432.

DSCI 451. Exploratory Data Science. 3 Units.

In this course, we will learn data science and analysis approaches to identify statistically significance relationships and better model and predict the behavior of these systems. We will assemble and explore real-world datasets, perform clustering and pair plot analyses to investigate correlations, and logistic regression will be employed to develop associated predictive models. Results will be interpreted, visualized and discussed. We will introduce basic elements of statistical analysis using R Project open source software for exploratory data analysis and model development. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and munging functions, and a rich selection of statistical packages, used for data analytics, model development and prediction. This will include an introduction to R data types, reading and writing data, looping, plotting and regular expressions, so that one can start performing variable transformations for linear fitting and developing structural equation models, while exploring for statistically significant relationships. The M section of DSCI 351 is for students focusing on Materials Data Science. Offered as DSCI 351, DSCI 351M and DSCI 451.

DSCI 452. Applied Data Science Research. 3 Units.

This is a project based data science research class, in which project teams identify a research project under the guidance of a domain expert professor. The research is structured as a data analysis project including the 6 steps of developing a reproducible data science project, including 1: Define the ADS question, 2: Identify, locate, and/or generate the data 3: Exploratory data analysis 4: Statistical modeling and prediction 5: Synthesizing the results in the domain context 6: Creation of reproducible research, Including code, datasets, documentation and reports. During the course special topic lectures will include Ethics, Privacy, Openness, Security, Ethics. Value. The M section of DSCI 352 is for students focusing on Materials Data Science. Offered as DSCI 352, DSCI 352M and DSCI 452.

DSCI 453. Data Science: Statistical Learning, Modeling and Prediction. 3 Units.

In this course, we will use an open data science tool chain to develop reproducible data analyses useful for inference, modeling and prediction of the behavior of complex systems. In addition to the standard data cleaning, assembly and exploratory data analysis steps essential to all data analyses, we will identify statistically significant relationships from datasets derived from population samples, and infer the reliability of these findings. We will use regression methods to model a number of both real-world and lab-based systems producing predictive models applicable in comparable populations. We will assemble and explore real-world datasets, use pair-wise plots to explore correlations, perform clustering, self-similarity, and logistic regression develop both fixed-effect and mixed-effect predictive models. We will introduce machine-learning approaches for classification and tree-based methods. Results will be interpreted, visualized and discussed. We will introduce the basic elements of data science and analytics using R Project open source software. R is an open-source software project with broad abilities to access machine-readable open-data resources, data cleaning and assembly functions, and a rich selection of statistical packages, used for data analytics, model development, prediction, inference and clustering. With this background, it becomes possible to start performing variable transformations for linear regression fitting and developing structural equation models, fixed-effects and mixed-effects models along with other statistical learning techniques, while exploring for statistically significant relationships. The class will be structured to have a balance of theory and practice. We'll split class into Foundation and Practicum a) Foundation: lectures, presentations, discussion b) Practicum: coding, demonstrations and hands-on data science work. The M section of DSCI 353 is for students focusing on Materials Data Science. Offered as DSCI 353, DSCI 353M and DSCI 453.

DSCI 454. Data Visualization and Analytics. 3 Units.

Data Visualization and Analytics students will learn data visualization and analytics techniques focused on different types of data such as time-series, spectral, or image data science problems. This class will focus on increasing analysis of complex data sets through visualization by enhancing exploratory data analysis and data cleaning. This class will focus on creating effective data visualizations to communicate data analytics results to different audiences. Different datasets will be provided to develop different types of visualizations and analytics. Types of data visualizations include in interactive plots (e.g., bar graphs change over time), applications that allow users to adjust the visualizations based on their decisions (e.g., shiny applications), interactive maps, 3-D plots of data, etc. Discussing how an audience understands information and brings in data as well as the ethics of making data visualizations will be discussed. The class will also include ways to increase modeling and analysis with effective visualizations for credible, data-driven decision making. This will include a git repository for other students to use these codes as open source resources and the preparation of reproducible data science analyses for different types of problems. Offered as DSCI 354, DSCI 354M, and DSCI 454. Prereq: DSCI 451 and DSCI 453.