

## シラバス

授業科目名	年度	学期	開講曜日・時限	学部・研究科など	担当教員	配当年次	単位数
AI・データサイエンスツールⅢ	2022	前期	他	学部間共通科目	酒折 文武、武石 智香子、中村 周史	1~4年次 配当	2

### 授業形式

すべての授業回についてオンデマンド型授業を行います。

### 履修条件・関連科目等

履修者数の上限は200名程度です。

### 授業で使用する言語

日本語

### 授業で使用する言語（その他の言語名）

### 授業の概要

AI・データサイエンス全学プログラムの応用基礎レベルに位置づけられ、スキルを身につけるツール科目の中でも、今後の応用力につながるツール科目である。BI（ビジネス・インテリジェンス）ツールツールやデータ分析言語を用いて、社会的な実課題に対してデータを活用することができる基本的なスキルを学んでいく。

前半は Exploratory というツールを扱う。大規模なデータを扱って可視化やデータ分析を行うデータサイエンスでは、R や Python などでのプログラミングが必須だと思われがちである。しかしデータを扱う現場では Tableau や Microsoft Power BI などのBI（ビジネス・インテリジェンス）ツールを用いて、簡単な操作で効率的に可視化や分析を行い、迅速な意思決定につなげている。この科目で扱う Exploratory は「探索的データ分析ツール」と銘打っているが高機能な BI ツールとも考えられ、学生は無料で使用できることも大きな特徴である。この科目の前半では、それぞれ分野でデータ分析や可視化の活用法を体得したい学生を対象とし、Exploratory を用いてデータの加工、可視化、分析を行い、他人に伝えるための技法について学ぶ。

後半は R というツール（言語）を扱う。R はデータ加工、可視化、分析などを行うための、データサイエンスでは欠かせないプログラミング言語であり、オープンソースで誰でも無料で新たなパッケージを開発・配布することが可能であるという特徴をもつ。同様の用途で用いられる Python 言語と比べて汎用性には劣るが、統計的な分析や可視化ならば R のほうが簡潔に行うことが可能である。さらに、tidyverse と呼ばれるパッケージ群を活用することで、データの加工を自由自在に行うことができる。この科目前半で扱った Exploratory も内部ではこの R が動いている。この科目の後半では、R 言語に触れたことのない学生を対象にして、tidyverse を用いたデータ加工や可視化を含んだ、データ分析の基礎的な技術を学ぶ。

授業は実習を含むオンデマンド型である。実習部分については、動画あるいは静的なウェブページの内容を参考にして、自身の PC で同じことを実行しながら進め、適宜演習問題にチャレンジしながら理解を深める。わからないところは manaba のスレッドなどを活用して講師やTAによるサポートが受けられるようにする。

### 科目目的

データサイエンスを各専門分野やビジネスで活用していくには、ビッグデータをラングリングし、視覚化・分析するツールが不可欠である。本科目の目的は、プログラムの基礎知識を要することなくこれらを行うことができる BI（ビジネス・インテリジェンス）ツールと、データサイエンスにおける基礎言語の一つである R の基礎、そしてデータ加工、可視化、データ分析の基本的な考え方を理解することである。

### 到達目標

本科目の到達目標は、数万件程度までの比較的小規模のデータについて、データ加工、可視化、基本的なデータ分析を自身で行えるようになることである。

### 授業計画と内容

#### ■ 前半：BIツール

##### 第1回 AIとBI、データサイエンスとツール

ガイダンス、AIとBI、データサイエンスとAI、データサイエンスと様々なツール、Exploratory と R、Exploratory のインストール、Exploratory 機能体験

##### 第2回 データの読み込みとサマリー

構造化データと非構造化データ、Exploratory とデータ、ファイルからのデータの読み込みの方法、統計学でのデータ、Exploratory における変数のタイプ、データフレームとサマリー、テーブルとソート、ハイライト・相関機能の活用

##### 第3回 グラフや表による可視化と分析

データ分析のアプローチとサイクル、可視化目的に応じたグラフ表現、クロス集計とピボットテーブル

##### 第4回 データラングリング

データ加工・クレンジングの重要性、新規変数の作成、層別集計とフィルタ、ロング型とワイド型の変換、ピンとブランチ、複数のデータフレームの結合とマージ

#### 第5回 データ・アナリティクス～統計と機械学習

データ分析と統計学・機械学習、線形回帰モデルによる予測、k-平均法によるクラスタリング、Exploratory で実行できる他のデータアナリティクス、その先

#### 第6回 ダッシュボードやノートによるレポート作成

ダッシュボードとその意義、Exploratory におけるダッシュボードの作成、ダッシュボードへの動的コンテンツの追加、Exploratory におけるノート・スライドの作成、Markdown 記法と数式、ダッシュボードやノート・スライドのエクスポート

#### 第7回 Exploratory のさらなる活用

ウェブスクレイピング技術、Twitterとの連動とテキストデータの分析、コロナ地図、住所データの扱い

#### ■ 後半：R

#### 第8回 RとRStudio

R と RStudioの概要、Exploratory に加えて R 言語を学ぶ意味、R・RStudio のインストール、コマンドの実行とスクリプト・四則演算・様々な関数、オブジェクトと代入、ベクトル、行列、スクリプトの保存

#### 第9回 データの読み込みと操作

プロジェクトと作業ディレクトリの設定、read.csv によるデータの読み込みと data.frame、基本的なデータハンドリング、パッケージのインストールと読み込み、read\_csv によるデータの読み込みと tibble、Excelファイルなど他アプリファイルの読み込み、パイプ

#### 第10回 グラフ作成とカスタマイズ

基本機能によるグラフ作成・グラフの他の機能・グラフの保存、ggplot2 によるグラフ作成の基本、ggplot の補足とグラフの保存、plotly による動的なグラフ、グラフィックの文法

#### 第11回 データラングリングの詳細

フィルタ filter、行の並べ替え arrange・列名の変換 rename、列の取り出し select・並べ替え relocate、新しい変数の作成 (mutate・transmute)、集計 summarise とグループ化 group\_by、pivot\_wider と pivot\_longer によるデータの変換、データタイプの変換、データフレームの結合とマージ

#### 第12回 Rでのデータ・アナリティクス

R での重回帰分析、ロジスティック回帰分析と識別、caret パッケージによる機械学習プログラミング、k-NN 法による予測、ランダムフォレストによる予測、勾配ブースティング木による予測

#### 第13回 Rmarkdown によるレポート作成

R Notebook と R Markdown、R Notebook によるノート作成、R Markdown によるドキュメント作成、R Notebook・R Markdown の記法の詳細、R Markdown によるスライドの作成、R Markdown の発展的課題

#### 第14回 Shinyでウェブアプリ構築

推薦システムと協調フィルタリング、アプリ制作の準備、Shiny App による協調フィルタリングの実装

### 授業時間外の学修の内容

授業終了後の課題提出

### 授業時間外の学修の内容（その他の内容等）

### 授業時間外の学修に必要な時間数/週

- ・毎週 1 回の授業が半期（前期または後期）または通年で完結するもの。1 週間あたり 4 時間の学修を基本とします。
- ・毎週 2 回の授業が半期（前期または後期）で完結するもの。1 週間あたり 8 時間の学修を基本とします。

### 成績評価の方法・基準

種別	割合 (%)	評価基準
レポート	70	<ul style="list-style-type: none"><li>●前半 (Exploratory) レポート：35% 与えられたテーマに合致したデータを収集し、Exploratory を用いたデータ加工、可視化、分析を適切に行って考察をまとめることができるかを問う。</li><li>●後半 (R) レポート：35% 与えられたデータについて、R言語を用いて効率的にかつ適切にデータ加工、分析を行い、考察ができるかどうかを問う。</li></ul>
その他	30	<ul style="list-style-type: none"><li>●自動採点式の確認問題による基本操作、基本的な考えかたの確認 Exploratory や R を適切に用いることができるかを問う。</li></ul>

### 成績評価の方法・基準（備考）

---

**課題や試験のフィードバック方法**

---

授業時間に限らず、manabaでフィードバックを行う

**課題や試験のフィードバック方法（その他の内容等）**

---

**アクティブ・ラーニングの実施内容**

---

実習、フィールドワーク

**アクティブ・ラーニングの実施内容（その他の内容等）**

---

**授業におけるICTの活用方法**

---

その他

**授業におけるICTの活用方法（その他の内容等）**

---

各自のPCでExploratoryおよびRの実習を行う。

**実務経験のある教員による授業**

---

はい

**【実務経験有の場合】実務経験の内容**

---

（武石）シンクタンク勤務経験有り。

**【実務経験有の場合】実務経験に関連する授業内容**

---

（武石）時代は大きく変わり、ツールも進化したが、実務時代に行っていた現場のニーズを起点として最新ツールの活用に挑戦できるための基礎力を養う授業を行う。

**テキスト・参考文献等**

---

特定のテキストは使用しない。毎回資料を配布する。

**オフィスアワー**

---

manaba を通じて連絡する。

**その他特記事項**

---

動画配信型オンライン授業を受講できる環境と、実習を行うことのできるPCが必要です。

**参考URL**

---

Exploratory: <https://exploratory.io>  
R: <https://www.r-project.org>  
RStudio: <https://rstudio.com>

**コメント1**

---

**コメント2**

---

**コメント3**

---

**コメント4**

---