



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Wizualizacja danych w R [S1DSwB1>WDwR]

Przedmiot

Kierunek studiów

Data Science w biznesie

Rok/Semestr

3/6

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

pierwszego stopnia

Język oferowanego przedmiotu

polski

Forma studiów

stacjonarne

Wymagalność

obligatoryjny

Liczba godzin

Wykład

30

Laboratorium

0

Inne

0

Ćwiczenia

30

Projekty/seminaria

0

Liczba punktów ECTS

4,00

Koordynatorzy

dr hab. inż. Yevhen Revtiuk

evhen.revtiuk@put.poznan.pl

Wykładowcy

Wymagania wstępne

1. Podstawowa znajomość matematyki i statystyki: Zrozumienie podstawowych pojęć statystycznych, takich jak średnia, mediana, odchylenie standardowe, rozkłady prawdopodobieństwa, korelacja i regresja, będzie pomocne w analizie danych. 2. Znajomość obsługi komputera: Umiejętność obsługi komputera, w tym korzystania z systemu operacyjnego, instalowania oprogramowania oraz podstawowych narzędzi biurowych. 3. Podstawowa znajomość języka R: Znajomość podstaw języka R, w tym pracy z danymi, podstawowych funkcji oraz pakietów (np. dplyr, ggplot2, tidyr). Kurs zakłada, że studenci posiadają umiejętności takie jak importowanie danych, manipulowanie danymi, podstawowa analiza oraz używanie funkcji języka R do wizualizacji prostych wykresów. 4. Zrozumienie podstawowych pojęć związanych z danymi: Umiejętność pracy z różnymi typami danych, takimi jak dane liczbowe, kategoryczne, czasowe, oraz rozumienie podstawowych pojęć związanych z jakością danych (np. brakujące wartości, outliery).

Cel przedmiotu

Celem przedmiotu „Wizualizacja danych w R” jest zapoznanie studentów z podstawowymi metodami wizualizacji danych przy użyciu języka R. Studenci nauczą się tworzyć różnorodne wykresy i wizualizacje, które ułatwią interpretację danych oraz wspomogą proces podejmowania decyzji. Kurs ma na celu rozwój umiejętności analizy danych poprzez wizualizację, poznanie narzędzi i pakietów w R, takich jak ggplot2 i plotly, oraz wdrożenie najlepszych praktyk w zakresie prezentacji wyników analizy. Po zakończeniu kursu studenci będą potrafili samodzielnie tworzyć przejrzyste i efektywne wykresy, dostosowane do charakterystyki danych oraz potrzeb odbiorcy.

Przedmiotowe efekty uczenia się

Wiedza:

1. Charakteryzuje techniki eksploracji i wizualizacji danych w R, w tym pakiety ggplot2, plotly, lattice oraz base R graphics [DSB1_W02].
2. Opisuje metody przygotowania danych do wizualizacji, w tym czyszczenie, transformację i agregację danych [DSB1_W03].
3. Wyjaśnia zasady percepcji wizualnej i teorię kolorów w kontekście projektowania skutecznych wizualizacji danych [DSB1_W07].

Umiejętności:

1. Dobiera odpowiednie metody wizualizacji danych w R, dostosowując je do rodzaju danych i celów analizy [DSB1_U02].
2. Analizuje i wizualizuje dane pochodzące z różnych źródeł, wykorzystując różnorodne typy wykresów, takie jak wykresy punktowe, słupkowe, histogramy, pudełkowe oraz mapy cieplne [DSB1_U04].
3. Projektuje i wdraża interaktywne wizualizacje danych, wykorzystując zaawansowane techniki facetingu, wizualizacji wielowymiarowej oraz geospacialnej [DSB1_U08].
4. Wykorzystuje narzędzia uczenia maszynowego do analizy zależności i predykcji trendów na podstawie wizualizacji danych [DSB1_U09].

Kompetencje społeczne:

1. Krytycznie analizuje własne umiejętności w zakresie wizualizacji danych, dążąc do ich doskonalenia i dostosowania do aktualnych standardów [DSB1_K01].
2. Angażuje się w inicjatywy związane z analizą i wizualizacją danych, promując dobre praktyki wizualizacji w analizie biznesowej i nauce [DSB1_K03].

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

1. Ocena projektów końcowych (50%)

• Opis: Studenci przygotowują projekt końcowy, w którym tworzą kompleksową wizualizację danych na podstawie rzeczywistego zbioru danych. Projekt musi obejmować analizę danych, dobór odpowiednich narzędzi wizualizacyjnych oraz prezentację wyników.

• Kryteria oceny:

- Jakość i poprawność wizualizacji (czy wykresy są odpowiednie do rodzaju danych, czy są czytelne i estetyczne).
- Umiejętność zastosowania technik wizualizacji w kontekście analizy danych.
- Oryginalność podejścia, kreatywność w wyborze narzędzi i metod.
- Przejrzystość raportu, uzasadnienie doboru wykresów, klarowność komunikacji wyników.

2. Ćwiczenia praktyczne i zadania domowe (50%)

• Opis: Regularne zadania domowe oraz ćwiczenia praktyczne podczas zajęć, polegające na tworzeniu wizualizacji danych przy użyciu R. Zadania obejmują zarówno proste wykresy, jak i bardziej zaawansowane analizy z wieloma zmiennymi.

• Kryteria oceny:

- Dokładność wykonania zadań i poprawność stworzonych wykresów.
- Zastosowanie odpowiednich technik wizualizacji w zależności od danych.
- Terminowość i samodzielność wykonania zadań.
- Umiejętność dostosowania wykresów do różnych grup odbiorców (np. wykresy do prezentacji vs. wykresy do publikacji).

Treści programowe

1. Przypomnienie podstawowych informacji z zakresu statystyki będącej podstawą analiz i wizualizacji w pakiecie R.
 - tworzenie różnych statystyk celem ich późniejszej wizualizacji.
2. Wprowadzenie do R i narzędzi do wizualizacji
 - Podstawowe zasady programowania w R
 - Instalacja i konfiguracja R oraz RStudio
 - Omówienie pakietów: ggplot2, plotly, lattice, base R graphics
3. Podstawy wizualizacji danych w ggplot2
 - Koncepcja "Grammar of Graphics"
 - Struktura wykresów w ggplot2 (aes(), geom_, theme_)
 - Tworzenie wykresów punktowych, słupkowych, liniowych
4. Przetwarzanie i przygotowanie danych do wizualizacji
 - Czyszczenie i transformacja danych (pakiet dplyr, tidyr)
 - Łączenie danych i filtrowanie
 - Formatowanie danych dla wykresów
5. Zaawansowane wykresy i techniki wizualizacji
 - Histogramy, wykresy pudełkowe, mapy cieplne
 - Wykresy wielowymiarowe i faceting
 - Wizualizacja macierzy korelacji
 - Wizualizacja danych na mapach za pomocą leaflet
6. Dobre praktyki wizualizacji danych
 - Percepcja wzrokowa i teoria kolorów
 - Jak unikać błędów i manipulacji w wizualizacji
7. Analiza przypadków i projekty studenckie
 - Wizualizacja danych rzeczywistych (np. dane finansowe, pogodowe, społeczne)
 - Tworzenie raportów i prezentacji danych
 - Przegląd narzędzi wspierających analizę danych (Markdown, R Notebooks)

Tematyka zajęć

1. Omówienie możliwości wykorzystania wizualizacji w pakiecie R w przypadku data science na przykładzie problemów z zakresu statystyki, zarządzania i marketingu - część realizowana na wykładach.
2. Wprowadzenie do wizualizacji danych i języka R
 - Znaczenie wizualizacji danych w analizie i prezentacji wyników.
 - Przegląd dostępnych narzędzi do wizualizacji danych w R.
 - Instalacja i konfiguracja pakietów R (np. ggplot2, plotly).
 - Podstawowe operacje na danych w R: importowanie, przetwarzanie, czyszczenie danych.
3. Podstawowe typy wykresów
 - Wykresy punktowe (scatter plots), histogramy, wykresy słupkowe.
 - Różnice między wykresami i kiedy je stosować.
 - Tworzenie podstawowych wykresów w ggplot2.
4. Wykresy liniowe i wykresy czasowe
 - Tworzenie wykresów liniowych do analizy danych szeregów czasowych.
 - Wykorzystanie pakietu ggplot2 do rysowania wykresów czasowych.
 - Manipulacja osiami i etykietami na wykresach.
5. Wizualizacja danych kategoriycznych
 - Wykresy słupkowe (bar plots) i wykresy kołowe.
 - Porównanie różnych rodzajów wykresów kategoriycznych i ich zastosowanie.
 - Praca z danymi w postaci tabel i wykresy grupowane.
6. Wizualizacja zależności między zmiennymi
 - Wykresy pudełkowe (box plots) i wykresy rozrzutu (scatter plots).
 - Wykorzystanie wykresów do analizy zależności między zmiennymi.
 - Wykresy z grupowaniem danych (np. wykresy z kolorami reprezentującymi zmienną grupującą).
7. Praca z mapami i danymi geograficznymi
 - Wizualizacja danych geograficznych w R: mapy, wykresy geospatial.
 - Wizualizacja danych przestrzennych: lokalizacje, granice, analiza geograficzna.
8. Estetyka i dostosowanie wykresów
 - Personalizacja wykresów: kolory, etykiety, tytuły, legendy.
 - Dostosowanie wykresów pod kątem odbiorcy: wykresy do publikacji, wykresy do prezentacji.
 - Omówienie zasad dobrego projektowania wykresów (np. zasada Tuftego).

9. Przykłady analizy danych i prezentacji wyników

- Analiza przypadków: jak wybrać odpowiedni typ wykresu w zależności od danych.
- Tworzenie raportów z wynikami analizy danych.
- Prezentacja wyników wizualnych: jak efektywnie komunikować dane za pomocą wykresów.

10. Podsumowanie kursu i projekt końcowy

- Omówienie projektów końcowych: samodzielne tworzenie wizualizacji danych.
- Podsumowanie zdobytej wiedzy i umiejętności.
- Prezentacja projektów przez studentów, analiza wyników.

Metody dydaktyczne

1. Wykłady interaktywne:

- Krótkie wykłady wprowadzające w tematykę wizualizacji danych w R, z omówieniem kluczowych pojęć, narzędzi oraz zasad dobrego projektowania wykresów.
- Podczas wykładów mogą być prezentowane przykłady zastosowań wizualizacji w różnych dziedzinach (np. analiza ekonomiczna, nauki społeczne, statystyka, zarządzanie) oraz dyskusja na temat najlepszych praktyk.
- Interaktywność wykładów polega na angażowaniu studentów w zadawanie pytań oraz omawianie przykładów na żywo.

2. Ćwiczenia praktyczne (laboratorium):

- Zajęcia oparte na rozwiązywaniu problemów, gdzie studenci w praktyce wykorzystują język R oraz pakiety do tworzenia wykresów i analizowania danych.
- Studenci otrzymują dane do analizy i na ich podstawie tworzą odpowiednie wizualizacje, rozwiązując konkretne zadania.
- Ćwiczenia prowadzone w formie warsztatów umożliwiają indywidualną pracę, podczas której prowadzący może udzielać wsparcia w razie trudności

3. Konsultacje indywidualne

- Studenci będą mogli skorzystać z konsultacji indywidualnych z prowadzącym zajęcia w celu omówienia trudności związanych z materiałem lub zadaniami. Konsultacje będą dotyczyły zarówno teorii, jak i praktycznych aspektów programowania w R.

Literatura

Podstawowa:

1. Hadley Wickham, Garrett Grolemund - R for Data Science (2017)
2. Kieran Healy - Data Visualization: A Practical Introduction (2018)
3. Claus O. Wilke - Fundamentals of Data Visualization (2019)

Uzupełniająca:

1. Hadley Wickham - ggplot2: Elegant Graphics for Data Analysis (2016),
2. Nathan Yau - Data Points: Visualization That Means Something (2013)
3. Fernando Miguez - Data Visualization with R (2022)
4. David McCandless - Information is Beautiful (2009)

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	100	4,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	60	2,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	40	1,50