

Белорусский государственный университет



Проректор по учебной работе и
образовательным инновациям
О.И. Чуприс
2019 г.

Регистрационный № УД- 7207/уч.

МЕТОДЫ ВИЗУАЛИЗАЦИИ В АНАЛИЗЕ БОЛЬШИХ ДАННЫХ СРЕДСТВАМИ ЯЗЫКА R

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-31 80 09 Прикладная математика и информатика

профилизация

Компьютерный анализ данных

2019 г.

Учебная программа составлена на ОСВО 1-31 80 09-2019 и учебного плана G31-073/уч. от 11.04.2019 г.

СОСТАВИТЕЛЬ:

В.В. Мушко, доцент кафедры технологий программирования факультета прикладной математики и информатики Белорусского государственного университета, кандидат физико-математических наук.

РЕЦЕНЗЕНТ:

Б.А. Залесский, заведующий лабораторией обработки и распознавания изображений Объединенного института проблем информатики Национальной академии наук Беларуси, доктор физико-математических наук.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики
(протокол № 15 от 30 апреля 2019 г.);

Научно-методическим Советом БГУ
(протокол № 5 от 28 июня 2019 года).

Заведующий кафедрой
теории вероятностей и математической статистики

 А.Ю. Харин

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины – ознакомление студентов II ступени (магистрантов) с основами визуализации данных, а также основными возможностями пакета ggplot2, расширяющего базовые графические возможности системы статистических вычислений языка R.

Задачи учебной дисциплины:

1. Изучение основных подходов и методов графического анализа данных и способов визуализации данных;
2. Формирование практических умений и навыков работы с пакетом ggplot2 в рамках R и RStudio.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием (магистра).

Учебная дисциплина относится к компоненту учреждения высшего образования и входит в модуль «Специальные методы анализа в прикладных задачах».

Основой для изучения дисциплины «Методы визуализации в анализе больших данных средствами языка R» является дисциплина I ступени высшего образования «Теория вероятностей и математическая статистика». Дисциплина «Методы визуализации в анализе больших данных средствами языка R» способствует успешному изучению дисциплин II ступени высшего образования «Многомерный статистический анализ» модуля «Методы и программные средства прикладной математики», «Методы статистического анализа сложных данных» модуля «Анализ сложных данных» и «Визуальная аналитика данных с временной структурой» модуля «Визуализация и интеллектуальный анализ», прохождению практики и написанию магистерских работ.

Требования к компетенциям

Освоение учебной дисциплины «Методы визуализации в анализе больших данных средствами языка R» должно обеспечить формирование следующих универсальных и специализированных компетенций:

универсальные компетенции:

УК-5. Обладать способностью в минимальные сроки изучать и профессионально эксплуатировать программные системы, модули и библиотеки.

специализированные компетенции:

СК-1. Обладать навыками использования основных методов визуализации больших данных средствами языка R.

В результате освоения учебной дисциплины студент должен:

знать:

- основные целевые функции процесса визуализации данных;
- основные подходы к визуализации данных;
- основные виды графиков, области их применения и типичные ограничения для каждого графического инструмента;
- особенности анализа и визуализации конкретных типов данных;
- типовые способы визуально эффективного представления результатов исследования
- потенциальные ошибки, возможные при визуализации данных;
- основные графические возможности пакета R ggplot2;

уметь:

- использовать различные методы визуализации данных для подготовки отчетов о результатах анализа;
- выбирать адекватные, оптимальные инструменты визуализации данных для эффективной поддержки принятия решений;
- эффективно использовать инструментальный программный обеспечения наук о данных – пакет ggplot2 среды R – для генерации визуальных представлений данных;

владеть:

- теоретическими знаниями базовых концепций и типовых практических инструментов, необходимых для анализа и визуализации больших данных;
- техническими навыками выбора адекватных инструментов эффективного графического анализа данных различных типов;
- умениями пользоваться инструментальной базой на практике.

Структура учебной дисциплины

Дисциплина изучается в первом семестре. Всего на изучение учебной дисциплины «Методы визуализации в анализе больших данных средствами языка R» отведено:

– для очной формы получения высшего образования – 126 часов, в том числе 40 аудиторных часов, из них: лекции – 20 часов, лабораторные занятия – 20 часов.

Трудоемкость учебной дисциплины составляет 3 зачетные единицы.

Форма текущей аттестации по учебной дисциплине – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Введение

Тема 1.1. Описание пакета ggplot2. Установка ggplot2 и начало работы. Грамматика графических элементов. Данные, используемые в примерах. Дополнительные ресурсы для изучения ggplot2. Расширения, созданные на основе ggplot2.

Тема 1.2. Функция qplot(). Аргументы функции qplot(). Построение диаграмм рассеяния с помощью qplot(). Другие примеры использования qplot(): линии тренда, одномерные диаграммы рассеяния, диаграммы размахов, гистограммы, графики плотностей распределения вероятностей, полигоны частот, столбиковые диаграммы. Категоризованные графики. Два способа организации панелей.

Тема 1.3. Построение графиков слой за слоем. Аргументы функции ggplot(). Требования к данным. Присваивание эстетических атрибутов. Группирование данных. Геометрические объекты, реализованные в ggplot2. Статистические преобразования. Общие аргументы geom- и stat- функций. Добавление слоев при помощи функций семейства stat.

Тема 1.4. Подготовка графиков к публикации. Стили. Создание составных рисунков. Экспорт графиков из среды R.

Раздел 2. Основные типы статистических графиков

Тема 2.1. Визуализация одномерных распределений вероятностей. Точечные диаграммы Уилкинсона. Столбиковые диаграммы. Гистограммы. Полигоны частот. Графики плотностей распределения вероятностей. Функции распределения. Квантильные графики.

Тема 2.2. Визуализация 2D- и 3D-распределений. Контурные плотности вероятности. Изолинии. Сотовые диаграммы.

Тема 2.3. Визуализация сводной статистической информации о количественных переменных. Диаграммы диапазонов. Диаграммы размахов. Скрипичные диаграммы.

Тема 2.4. Визуализация зависимостей. Диаграммы рассеяния. Линии тренда. Линии квантильной регрессии.

Тема 2.5. Визуализация временных рядов. Тепловые карты. Географические карты. Визуализация наблюдаемых временных рядов. Тепловые карты. Географические карты.

Тема 2.6. Другие геометрические объекты. «График-щетка». Горизонтальные и вертикальные линии. Прямоугольные области. Отрезки. Ломаные линии. Многоугольники. Площадь под кривой. Текстовые аннотации.

Раздел 3. Шкалы и системы координат

Тема 3.1. Шкалы и их основные типы. Шкалы и их основные типы. Аргументы, общие для всех scale-функций.

Тема 3.2. Шкалы положения. Шкалы положения для количественных переменных. Шкалы положения для дат и времени. Шкалы положения для качественных переменных.

Тема 3.3. Цветовые шкалы. Цветовые шкалы для количественных переменных. Цветовые шкалы для качественных переменных.

Тема 3.4. Пользовательские шкалы для качественных переменных. Тожественные шкалы. Пользовательские шкалы для качественных переменных. Тожественные шкалы.

Тема 3.5. Системы координат. Декартова система и ее разновидности. Полярная система. Картографические проекции.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования с применением дистанционных образовательных технологий

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов		Форма контроля знаний
		Лекции	Лабораторные занятия	
1	Введение	4	2	
1.1	Описание пакета ggplot2	1		собеседования
1.2	Функция qplot()	1		собеседования
1.3	Построение графиков слой за слоем	1		собеседования
1.4	Подготовка графиков к публикации	1	2	рефераты
2	Основные типы статистических графиков	12	12	
2.1	Визуализация одномерных распределений вероятностей	2	2	письменные отчеты
2.2	Визуализация 2D- и 3D-распределений	2	2	письменные отчеты
2.3	Визуализация сводной статистической информации о количественных переменных	2	2	письменные отчеты
2.4	Визуализация зависимостей	2	2	письменные отчеты
2.5	Визуализация временных рядов. Тепловые карты. Географические карты	2	2	письменные отчеты
2.6	Другие геометрические объекты	2	2	письменные отчеты
3	Шкалы и системы координат	4	6	
3.1	Шкалы и их основные типы	1	1	контрольная работа № 1
3.2	Шкалы положения	0.5	1	отчеты
3.3	Цветовые шкалы	0.5	1	отчеты
3.4	Пользовательские шкалы для качественных переменных. Тожественные шкалы	1	2	контрольная работа № 2, отчеты
3.5	Системы координат	1	1	Отчеты
ИТОГО		20	20	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Wickham H., Grolemund G. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. — O'Reilly Media, 2017. — 522 p.
2. Chang W. R Graphics Cookbook: Practical Recipes for Visualizing Data, 2nd Edition. — O'Reilly Media, 2017. — 643 p.
3. Wickham H. ggplot2: Elegant Graphics for Data Analysis, 3rd Edition. — Springer International Publishing, 2019. — 260 p.
4. Мастицкий С. Э. Визуализация данных с помощью ggplot2. — М.: ДМК Пресс, 2016. — 222 с.

Перечень дополнительной литературы

1. Kassambara A. ggplot2: Guide to Create Beautiful Graphics in R. Volume 1, 2nd Edition.— CreateSpace Independent Publishing Platform, 2016. — 238 p.
2. Teutonico D. ggplot2 Essentials. — Packt Publishing, 2015. — 234 p.
3. Rahlf T. Data Visualisation with R. 100 Examples. — Springer International Publishing, 2017. — 385 p.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Для диагностики компетенций используются следующие формы:

1. *Устная форма*: собеседования;
2. *Письменная форма*: рефераты, письменные отчеты по аудиторным практическим упражнениям, контрольные работы;
3. *Устно-письменная форма*: отчеты по домашним практическим упражнениям с их устной защитой.

Формой текущей аттестации по дисциплине «Методы визуализации в анализе больших данных средствами языка R» учебным планом предусмотрен зачет.

При формировании итоговой оценки используется рейтинговая оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний и текущей аттестации в рейтинговую оценку:

- Формирование оценки за текущую успеваемость:
- отчеты – 40 %;

- подготовка реферата – 30 %;
- контрольные работы – 30 %.

Рейтинговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости и зачетной оценки с учетом их весовых коэффициентов. Оценка по текущей успеваемости составляет 30 %, зачетная оценка – 70 %.

Примерная тематика лабораторных занятий

Занятие № 1. Расширения, созданные на основе ggplot2.

Занятие № 2. Визуализация одномерных распределений вероятностей.

Занятие № 3. Визуализация 2D- и 3D-распределений.

Занятие № 4. Визуализация сводной статистической информации о количественных переменных.

Занятие № 5. Визуализация зависимостей.

Занятие № 6. Визуализация временных рядов. Тепловые карты. Географические карты.

Занятие № 7. Другие геометрические объекты.

Занятие № 8. Отчеты по домашним практическим упражнениям с их устной защитой.

Занятие № 9. Отчеты по домашним практическим упражнениям с их устной защитой.

Занятие № 10. Отчеты по домашним практическим упражнениям с их устной защитой.

Рекомендуемая тематика контрольных работ:

Контрольная работа № 1. Географические карты.

Контрольная работа № 2. Цветовые шкалы.

Текущий контроль знаний проводится в соответствии с учебно-методической картой дисциплины.

Описание инновационных подходов и методов к преподаванию учебной дисциплины (эвристический, проективный, практико-ориентированный)

При организации образовательного процесса используется практико-ориентированный подход, который предполагает:

- освоение содержания образования через решения практических задач;
- приобретение навыков эффективного выполнения разных видов профессиональной деятельности;
- ориентацию на генерирование идей, реализацию групповых студенческих проектов, развитие предпринимательской культуры;

- использованию процедур, способов оценивания, фиксирующих сформированность профессиональных компетенций.

При организации образовательного процесса используется метод группового обучения, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую функционирование разных типов малых групп, работающих как над общими, так и специфическими учебными заданиями.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине следует использовать современные информационные ресурсы: разместить на образовательном портале комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля, тематика рефератов и др., список рекомендуемой литературы, информационных ресурсов и др.).

Темы реферативных работ

1. Расширения, созданные на основе ggplot2.

Примерный перечень вопросов к зачету

1. Грамматика графических элементов.
2. Стили. Создание составных рисунков.
3. Экспорт графиков из среды R.
4. Визуализация одномерных распределений вероятностей.
5. Визуализация 2D- и 3D-распределений.
6. Визуализация сводной статистической информации о количественных переменных.
7. Визуализация зависимостей.
8. Визуализация временных рядов.
9. Тепловые карты.
10. Географические карты.
11. Другие геометрические объекты.
12. Шкалы и их основные типы.

13. Шкалы положения.
14. Цветовые шкалы.
15. Пользовательские шкалы для качественных переменных.
16. Тожественные шкалы.
17. Системы координат.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Многомерный статистический анализ	Кафедра теории вероятностей и математической статистики	нет	Оставить содержание учебной дисциплины без изменения, протокол № 15 от 30 апреля 2019 г.
Методы статистического анализа сложных данных	Кафедра теории вероятностей и математической статистики	нет	Оставить содержание учебной дисциплины без изменения, протокол № 15 от 30 апреля 2019 г.
Визуальная аналитика данных с временной структурой	Кафедра теории вероятностей и математической статистики	нет	Оставить содержание учебной дисциплины без изменения, протокол № 15 от 30 апреля 2019 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 201_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
