

МИНОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:
Директор
А. В. Замятин
«14» *ноября* 20*23* г.



Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине
(Оценочные средства по дисциплине)

Представление знаний и визуализация данных

по направлению подготовки

02.04.02 Фундаментальная информатика и информационные технологии

Направленность (профиль) подготовки:

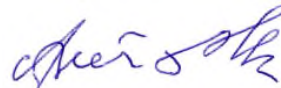
Моделирование систем искусственного интеллекта

ОС составил:
канд. техн. наук, доцент кафедры
теоретических основ информатики



О.В. Марухина

Рецензент:
канд. техн. наук, доцент кафедры
теоретических основ информатики



С.В. Аксёнов

Оценочные средства одобрены на заседании учебно-методической комиссии
института прикладной математики и компьютерных наук (УМК ИПМКН)

Протокол от «08» июня 2023 г. № 02

Председатель УМК ИПМКН,
д-р техн. наук, профессор



С.П. Сущенко

Оценочные средства (ОС) являются элементом системы оценивания сформированности компетенций у обучающихся в целом или на определенном этапе ее формирования.

ОС разрабатываются в соответствии с рабочей программой (РП) дисциплины и включает в себя набор оценочных материалов для проведения текущего контроля успеваемости и промежуточной аттестации по дисциплине «Представление знаний и визуализация данных»

1. Компетенции и результаты обучения, формируемые в результате освоения дисциплины «Представление знаний и визуализация данных»

Компетенция	Индикатор компетенции	Код и наименование результатов обучения (планируемые результаты обучения, характеризующие этапы формирования компетенций)	Критерии оценивания результатов обучения			
			Отлично (зачтено)	Хорошо (зачтено)	Удовлетворительно (зачтено)	Неудовлетворительно (незачтено)
ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	ИОПК-4.1 Анализирует задачи профессиональной деятельности средствами информационных технологий.	ОР-4.1.1. Уметь анализировать задачи профессиональной деятельности средствами информационных технологий.	Демонстрация высокого уровня умения анализировать задачи профессиональной деятельности средствами информационных технологий.	В целом успешные, но содержащие отдельные пробелы знания в умения решать актуальные задачи профессиональной деятельности средствами информационных технологий.	Фрагментарное, неполное знание без грубых ошибок использования современных компьютерных технологий для решения задач профессиональной деятельности средствами информационных технологий.	Не имеет четкого представления об изучаемом материале, допускает грубые ошибки
ПК-4. Способен управлять получением, хранением, передачей, обработкой больших данных.	ИПК-4.1. Осуществляет мониторинг и оценку производительности обработки больших данных.	В результате освоения дисциплины студент будет Уметь: проводить анализ результатов научного исследования с целью грамотного выбора методов и средств визуализации	Демонстрация высокого уровня умения выбирать для реальных систем математические методы решения прикладных задач; математически корректно применять методы	В целом успешные, но содержащие отдельные пробелы знания в умения выбирать для реальных систем математические методы решения прикладных задач;	Фрагментарное, неполное знание без грубых ошибок использования современных компьютерных технологий для	Не имеет четкого представления об изучаемом материале, допускает грубые ошибки при построении

	ИПК-4.2. Использует методы и инструменты получения, хранения, передачи, обработки больших данных.	полученных данных и знаний (ОР-4.1.1); интерпретировать результаты визуализации, грамотно их описывать (ОР-4.1.2). Владеть: навыками использования библиотек современных скриптовых языков (Python, R) для визуализации данных и знаний (ОР-4.2.1); навыками решения прикладных профессиональных задач с использованием методов и средств визуализации данных и знаний (ОР-4.2.2).	исследования предлагаемых моделей; получать основные характеристики исследуемых моделей; выполнять интерпретацию математических результатов; способен совершенствовать и реализовывать новые математические методы решения задач с использованием вероятностных методов логистики.	получать основные характеристики исследуемых моделей; выполнять интерпретацию математических результатов; в умении совершенствовать и реализовывать новые математические методы решения задач с использованием вероятностных методов логистики.	совершенствования и реализации новых математических методов решения прикладных задач с использованием вероятностных методов логистики.	математических моделей и проведении их анализа при решении задач с использованием вероятностных методов логистики.
УК-7. Способен понимать фундаментальные принципы работы современных систем искусственного интеллекта, разрабатывать правила и стандарты взаимодействия человека и искусственного интеллекта и использовать их в социальной и профессиональной деятельности	ИУК-7.1. Применяет современные методы и инструменты для представления результатов научно-исследовательской деятельности.	ОР-7.1.1. Знает современные методы и инструменты для представления результатов научно-исследовательской деятельности.	Знает современные методы и инструменты для представления результатов научно-исследовательской деятельности.	Знает современные методы и инструменты для представления результатов научно-исследовательской деятельности, но допускает незначительные ошибки.	Знает современные методы и инструменты для представления результатов научно-исследовательской деятельности, но допускает много ошибок.	Не знает современные методы и инструменты для представления результатов научно-исследовательской деятельности.

2. Этапы формирования компетенций и виды оценочных средств

№	Этапы формирования компетенций (разделы дисциплины)	Код и наименование результатов обучения	Вид оценочного средства (тесты, задания, кейсы, вопросы и др.)
1.	Раздел 1. Визуализация многомерных данных. Задачи визуализации. Способы визуализации. Цель визуализации, задачи визуализации многомерных данных, классификация по визуализируемым объектам. Выбор правильного типа визуализации. Первичный анализ данных с использованием методов визуализации. Библиотеки Python (или R) для решений задач визуализации. Сравнение полученных визуальных образов. Выполнение лабораторной работы № 1.	ОР-4.1.1, ОР-4.1.2, ОР-4.2.1, ОР-4.2.2, ОР-4.2.2	Опрос на занятиях, подготовка к лабораторным занятиям, публичная защита лабораторной работы № 1.
2.	Раздел 2. Python-библиотеки для визуализации данных в Data Science . Возможности библиотек Matplotlib, Seaborn, Missingno, Altair, Plotly, Vokeh, Pygal, Networkx. Примеры использования перечисленных библиотек. Визуализация многомерных данных с использованием диаграмм Эндрюса. Выполнение лабораторной работы № 2.	ОР-4.1.1, ОР-4.1.2, ОР-4.2.1, ОР-4.2.2, ОР-4.2.2	Опрос на занятиях, подготовка к лабораторным занятиям, публичная защита лабораторной работы № 2.
3	Раздел 3. Визуализация данных средствами дашбордов. Что из себя представляет дашборд, его свойства, отличие от отчета. Модули дашборда. Виды дашбордов. Основные инструменты: Google Sheets, Яндекс.Метрика, Google Analytics, Google Data Studio, Qlik, Power Bi, Owox Bi. Выполнение лабораторной работы № 3. Выбор темы индивидуального проекта.	ОР-4.1.1, ОР-4.1.2, ОР-4.2.1, ОР-4.2.2, ОР-4.2.3	Опрос на занятиях, подготовка к лабораторным занятиям, публичная защита лабораторной работы № 3.
4	Раздел 4. Методы визуализации для решения прикладных задач. Методы визуализации для задач классификации, кластеризации. Визуализация решения задачи временных рядов, демонстрация примеров. Выполнение лабораторной работы № 4. Демонстрация выполнения индивидуальных проектов.	ОР-4.1.1, ОР-4.1.2, ОР-4.2.1, ОР-4.2.2, ОР-4.2.3, ОР-7.1.1.	Опрос на занятиях, подготовка к лабораторным занятиям, публичная защита лабораторной работы № 4.

5	Промежуточная аттестация (по результатам выполнения лабораторных работ (min 70%) и презентации индивидуального проекта -2-3 мин/чел.)	ОР-4.1.1, ОР-4.1.2, ОР-4.2.1, ОР-4.2.2, ОР-4.2.3	Публичное представление и защита результатов индивидуального проекта.
---	---	--	---

3. Типовые контрольные задания или иные материалы, необходимые для оценки образовательных результатов обучения

3.1. Типовые задания для проведения текущего контроля успеваемости по дисциплине «Представление знаний и визуализация данных».

Лабораторная работа № 1

Исходные данные: изучите открытые данные по выборам депутатов Государственной Думы Федерального Собрания Российской Федерации седьмого созыва: http://www.vybory.izbirkom.ru/region/region/izbirkom?action=show&root=1&tvd=100100067795854&vrn=100100067795849®ion=0&global=1&sub_region=0&prver=0&pronetvd=0&vid=100100067795854&type=233. Выберите для дальнейшего анализа данные по одному из округов.

Задание: 1) продумайте и подберите тип визуализации для ответа на следующие вопросы:

- А) какие партии являются лидерами (аутсайдерами) по количеству голосов;
- Б) какие избирательные участки являются лидерами (аутсайдерами) по количеству пришедших на выборы;
- В) есть какие-то закономерности (связи) между количеством голосов за лидирующие партии и конкретными избирательными участками;
- Г) можно ли определить какие-то другие связи?
- Д) сделайте выводы.

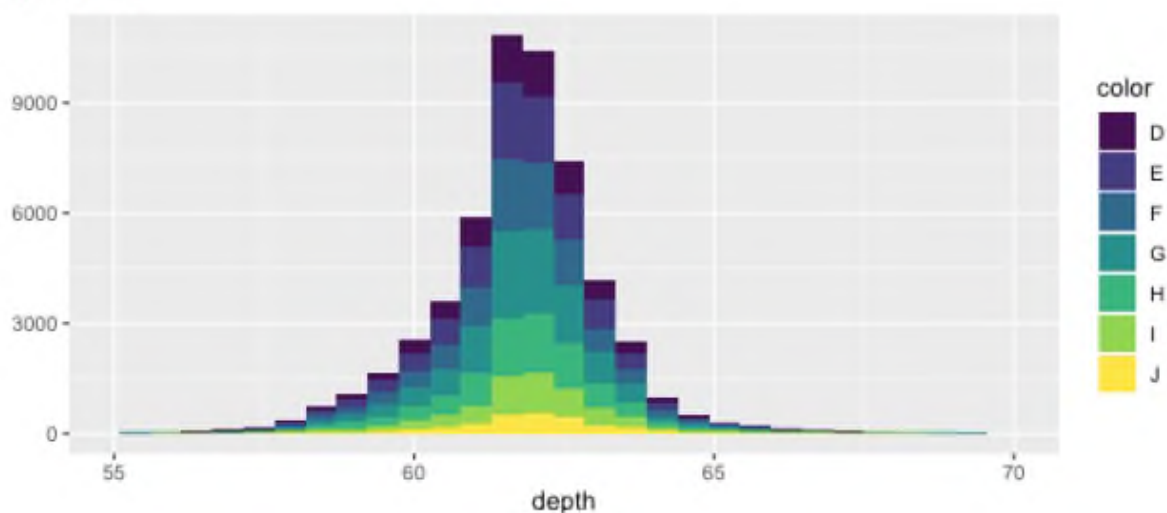
Лабораторная работа № 2

Цель работы: познакомиться с основными элементами графической грамматики основных библиотек визуализации.

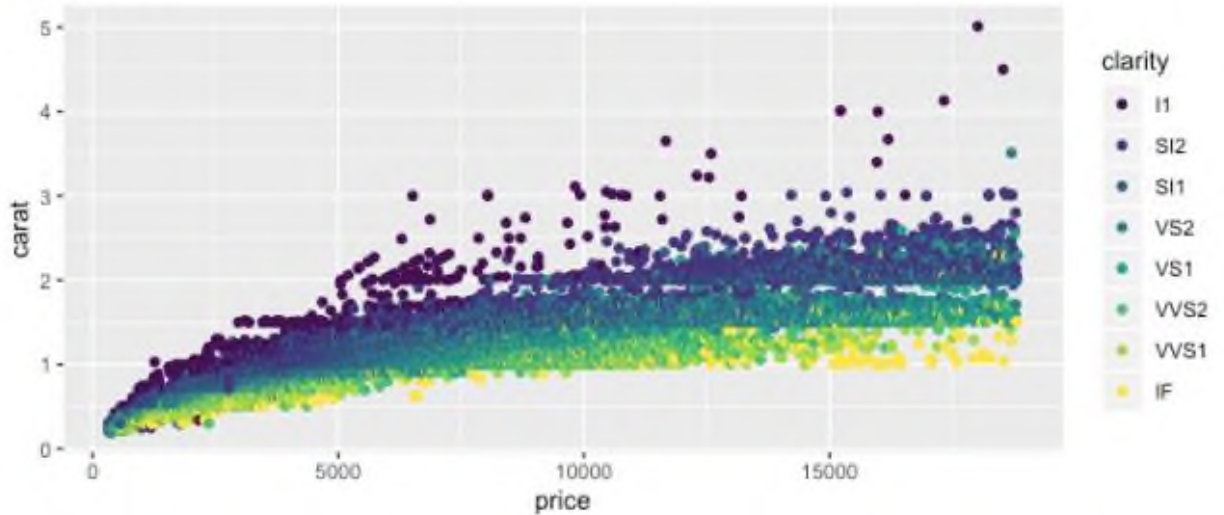
Задания:

Задание 1. С помощью любой из изученных библиотек напишите программный код, строящий следующие графики по набору данных diamonds.

Вариант – 1

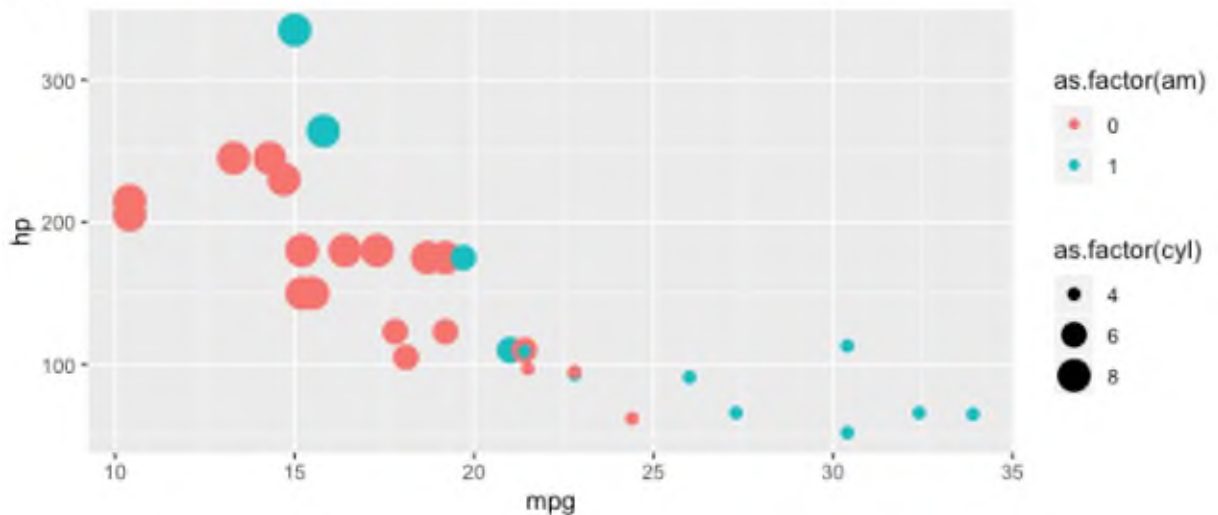


Вариант – 4

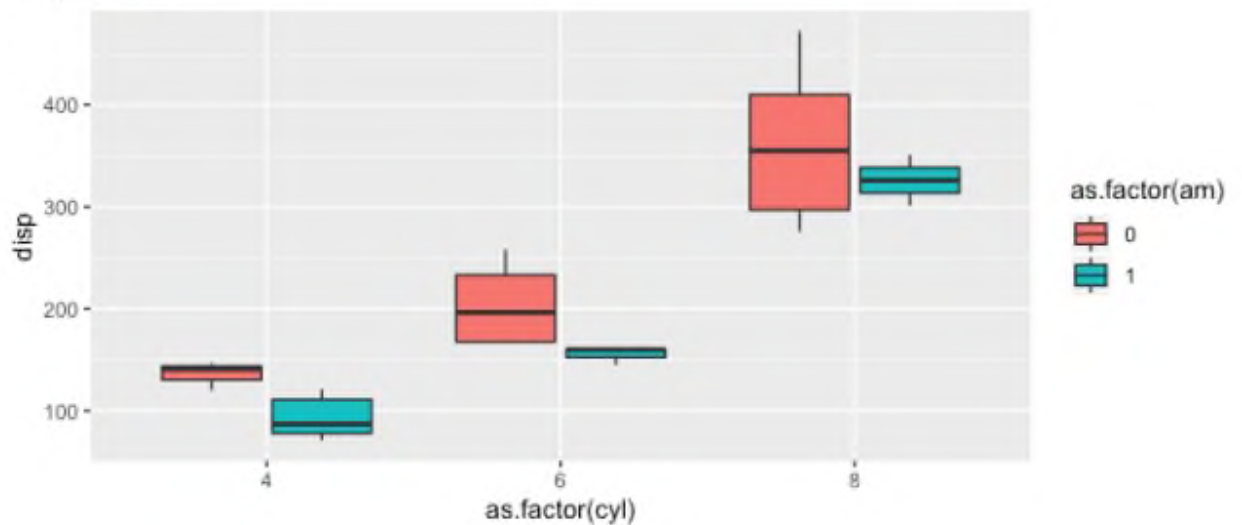


Задание 2. Напишите программный код, строящий следующие графики по набору данных mtcars.

Вариант – 1



Вариант – 4



Задание 3. Постройте два произвольных графика, показав умение использовать различные настройки для набора данных по заданному варианту. Дайте описание набору данных и то, что показывает изображенный график.

№	Набор данных	№	Набор данных	№	Набор данных	№	Набор данных
1	CO2	8	Puromycin	15	iris	22	swiss
2	ChickWeight	9	Seatbelts	16	longley	23	trees
3	DNase	10	Theoph	17	mtcars	24	economics
4	LifeCycleSavings	11	ToothGrowth	18	quakes	25	faithfuld
5	Loblolly	12	USArrests	19	rock	26	midwest
6	Orange	13	freeny	20	stack.x	27	mpg
7	OrchardSprays	14	infert	21	stackloss	28	txhousing

Лабораторная работа № 3

1. Выберите датасет, согласно своего варианта (<https://habr.com/ru/post/452392/>).
Можно выбрать наиболее интересный для себя.
2. Выполните первичный анализ данных, **используя только методы визуализации**:
 - a. Выбросы
 - b. Пропущенные значения (с анализом)
 - c. Коллинерные признаки.
3. Выполните визуализацию зависимости целевой переменной/переменных от информативных признаков. Проанализируйте.
4. Выполните различные виды визуализации, согласно примерам из лекций. Можно добавить свои.
5. Сделайте общие выводы.

ВАРИАНТЫ:

1. [Данные смертей и сражений из игры престолов](#) — этот набор данных объединяет три источника данных, каждый из которых основан на информации из серии книг.
2. [Глобальная база данных терроризма](#) — Более 180 000 террористических атак по всему миру, 1970-2017.
3. [Биткойн, исторические данные](#) — данные биткойнов с интервалом в 1 минуту с избранных бирж, январь 2012 г. — март 2019 г.
4. [FIFA 19 полный набор данных игроков](#) — 18k + FIFA 19 игроков, ~ 90 атрибутов, извлеченных из последней базы данных FIFA.
5. [Статистика видео YouTube](#) — ежедневная статистика трендовых видео на YouTube.
6. [Обзор показателей самоубийств с 1985 по 2016 год](#) — Сравнение социально-экономической информации с показателями самоубийств по годам и странам.
7. [Huge Stock Market Dataset](#) — исторические дневные цены и объемы всех американских акций и ETF.
8. [Индикаторы мирового развития](#) — показатели развития стран со всего мира.
9. [Kaggle Machine Learning & Data Science Survey 2017](#) — Большое представление о состоянии науки о данных и машинного обучения.
10. [Данные о насилии и оружии](#) — полный отчет о более чем 260 тыс. американских инцидентов с применением оружия в 2013-2018 гг.

Лабораторная работа № 4

1. Задание рассчитано на выполнение в группе – 2 человека (по желанию – можно индивидуально).
2. Изучить один из инструментов создания дашбордов, используя документацию и тестовые примеры:
 - Вариант 1. Google Sheets (Excel)
 - Вариант 2. Яндекс.Метрика
 - Вариант 3. Google Analytics
 - Вариант 4. Google Data Studio
 - Вариант 5. Qlik
 - Вариант 6. Tableau
 - Вариант 7. Power Bi
 - Вариант 8. Owox Bi
3. Построить свой дашборд, используя открытые данные любого выбранного вами сайта.
4. В отчет должно входить описание используемой системы, аргументация выбора средств визуализации, скриншоты дашборда и ссылка на дашборд.

NOTE! Можно выполнять эту работу «вручную», создав дашборд на python, с использованием, например, библиотеки dash

Темы индивидуальных проектов:

Для укрепления изученного материала предусмотрено выполнение индивидуального проекта в рамках часов самостоятельной работы. Проект может быть выполнен как индивидуально, так и в мини-группе (2-3 чел.), при условии, что объем работы также будет увеличен. В конце семестра по каждому проекту представляется мини-презентация о результатах работы.

Тематика индивидуального проекта связана с темой ВКР магистранта. Цель работы – использование методов визуализации в своей научной работе.

Темы опросов на занятиях:

Связаны с материалом предыдущих лекций, а также личным опытом студентов. Студенты могут предлагать варианты решений поставленной преподавателем задачи, а также инструменты решения.

Примеры вопросов:

1) Какой тип визуализации лучше всего подойдет для задачи прогнозирования?

2) Как вы обоснуете выбор библиотеки для визуализации?

3) Как еще можно визуализировать представленную зависимость в данных?

3.2. Типовые задания для проведения промежуточной аттестации по дисциплине «Представление знаний и визуализация данных»

Зачет выставляется на основе представления и защиты индивидуального проекта.

Студент выполняет презентацию, а также демонстрирует программный код. Вопросы по результатам могут задавать все студенты группы, не только преподаватель.

Таблица критериев выставления зачета

Оценка	Критерии
Зачтено	Студент активно работал в течение семестра, выступил с презентацией индивидуального проекта, посещал лекционные занятия, лабораторные

	работы сданы в срок.
Не зачтено	Студент не работал во время семестра, не выступал с презентацией индивидуального проекта, не посещал лекционные занятия, лабораторные работы не сданы или сданы на менее чем 10 баллов.

4. Методические материалы, определяющие процедуры оценивания образовательных результатов обучения

Рейтинговая система для оценки текущей успеваемости обучающихся

Таблица – Балльные оценки для элементов контроля

Элементы учебной деятельности	Максимальный балл с начала семестра	Оцениваемая компетенция
Подготовка к лабораторным занятиям и защита отчета по лабораторной работе	15*4=60	ПК-4
Защита индивидуальных проектов	40	ПК-4
Зачет		

Пересчет баллов в оценки промежуточной успеваемости

Баллы на дату контрольной точки	Оценка
≥ 90% от максимальной суммы баллов	5 (зачтено)
От 70% до 89% от максимальной суммы баллов	4 (зачтено)
От 60% до 69% от максимальной суммы баллов	3 (зачтено)
< 60% от максимальной суммы баллов	2 (незачтено)