**Санкт-Петербургский филиал федерального государственного
автономного образовательного учреждения высшего образования
"Национальный исследовательский университет**

**"Высшая школа экономики"**

Факультет Санкт-Петербургская школа социальных и гуманитарных наук Национального исследовательского университета «Высшая школа экономики»

Департамент социологии

**Рабочая программа дисциплины**

«Анализ данных и технологии работы с данными»

для майнора «Обработка и анализ данных»

уровень бакалавриат

Разработчики программы:

Мусабиров И.Л., MA (Digital Media and Information Society), магистр по направлению «Информатика и вычислительная техника», старший преподаватель департамента социологии, ilya@musabirov.info

Суворова А.В., кандидат физико-математических наук, доцент департамента прикладной математики и бизнес-информатики, asuvorova@hse.ru

Маслинский К.А., кандидат исторических наук, преподаватель департамента социологии

Утверждена академическим руководителем майнора

Сироткин А.В. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ [подпись]

«09» января 2018 г.

Санкт-Петербург, 2018

1. **Область применения и нормативные ссылки**

Настоящая рабочая программа дисциплины устанавливает минимальные требования к знаниям и умениям студента, а также определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов, изучающих дисциплину «Программирование для анализа данных и воспроизводимые исследования» в рамках майнора «Обработка и анализ данных».

Рабочая программа дисциплины разработана в соответствии с:

* Объединенным учебным планом майнора, утвержденным в 2017 г.
1. **Цели освоения дисциплины**

Целями освоения дисциплины «Анализ данных и технологии работы с данными» являются изучение технологий и методов обработки и анализа данных, включая социально-сетевой анализ, вычислительный анализ текста, а также развитие навыков программирования для анализа данных и построения моделей машинного обучения. Дисциплина предполагает участие в практическом проекте

1. **Компетенции обучающегося, формируемые в результате освоения дисциплины**

В результате освоения дисциплины студент должен:

* Знать
	+ ключевые понятия и задачи социально-сетевого анализа данных и вычислительного анализа текста;
	+ методологические основы цифровых исследований, использующих социально-сетевой анализ и вычислительный анализ текста;
	+ актуальные направления исследований в области анализа социальных сетей и вычислительного анализа текста
* Уметь
	+ выбирать метод анализа, соответствующий исследовательской задаче,
	+ интерпретировать полученные результаты;
	+ визуализировать социальные сети,
	+ анализировать свойства, структуру и особенности социальных сетей,
	+ применять методы вычислительного анализа текста, основанные на подходе bag of words,
	+ решать задачи машинного обучение с применением социально-сетевого анализа и вычисления текста,
	+ организовывать индивидуальные проекты по анализу данных, применяющие изученные методы анализа,
	+ организовывать и участвовать в работе проектной группы,
	+ презентовать в необходимой форме результаты работы, осознанно выбирая методы визуализации данных, текстовые и графические материалы
* Иметь навыки (приобрести опыт)
	+ чтения и анализа академической литературы по методам анализа данных,
	+ использования современного инструментария анализа данных и технологий работы с данными,
	+ работы в групповом проекте по анализу данных, включая самостоятельное планирование групповой работы

Уровни формирования компетенций:

 **РБ** - ресурсная база, в основном теоретические и предметные основы (знания, умения)

 **СД** - способы деятельности, составляющие практическое ядро данной компетенции

 **МЦ** - мотивационно-ценностная составляющая, отражает степень осознания ценности

компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает следующие компетенции:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Компетенция | Код по ОС ВШЭ | Уровень формирования компетенции | Дескрипторы – основные признаки освоения (показатели достижения результата) | Формы и методы обучения, способствующие формированию и развитию компетенции | Форма контроля уровня сформированности компетенции |
| Способен учиться, приобретать новые знания, умения, в том числе в области, отличной от профессиональной | УК-1 | РБ | Способен на основе рекомендаций и самостоятельного поиска осваивать новые методы анализа данных | Практические и самостоятельные занятия по написанию программ для сбора и обработки информации | Контрольные работыДомашнее заданиеГрупповой программный проект |
| Способен вести исследовательскую деятельность, включая анализ проблем, постановку целей и задач, выделение объекта и предмета исследования, выбор способа и методов исследования, а также оценку его качества | УК-6 | РБ | Способен выделять постановки задач для решения с использованием различных методов анализа данных, осознанно выбирать методы и инструментальные средства | Самостоятельная деятельность по анализу результатов проведенных исследований | Контрольные работыДомашнее заданиеГрупповой программный проект |
| Способен работать с информацией: находить, оценивать и использовать информацию из различных источников, необходимую для решения научных и профессиональных задач (в том числе на основе системного подхода) | УК-5 | РБ | Владеет навыками написания программ для автоматизированного сбора и анализа информации из различных источников в глобальных компьютерных сетях, социально-сетевого и текстового анализа | Практические занятия, программный проект | Контрольные работыДомашнее заданиеГрупповой программный проект |

1. **Место дисциплины в структуре образовательной программы**

Настоящая дисциплина относится к блоку дисциплин дополнительного профиля (майнора) «Обработка и анализ данных», обеспечивающих бакалаврскую подготовку.

Изучение данной дисциплины базируется на дисциплине «Программирование для анализа данных и воспроизводимые исследования» майнора «Обработка и анализ данных».

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин майнора.

1. **Тематический план учебной дисциплины**

ОБЪЕМ ДИСЦИПЛИНЫ - 5 зачетных единиц

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название раздела | Всего часов  | Аудиторные часы | Самостоятельная работа |
| Лекции | Практические работы | Семинары |
| 1 | Введение в анализ текста в рекомендательных системах | 60 | 2 | 18 |  | 40 |
| 2 | Анализ и визуализация сетей в R. Социальные сети в бизнесе, экономике, обществе | 57 | 3 | 12 |  | 30 |
| 3 | Проектная организация цифровых исследований | 73 | 3 | 18 |  | 64 |
| ИТОГО | **190** | 8 | 48 |  | 134 |

1. **Формы контроля знаний студентов**

|  |  |  |  |
| --- | --- | --- | --- |
| Тип контроля | Форма контроля | 1 год | Параметры |
| 1 | 2 | 3 | 4 |
| Текущий | Контрольная работа |  |  | \* | \* | Письменные работы на 70 минут |
| Домашнее задание |  |  | \* |  | Задание по анализу эгосети дружбы (письменный отчет) |
| Подготовка к практикумам и работа на них, включая групповую |  |  | \* | \* | Прохождение онлайн-упражненийРабота на практикумах, включая групповую |
|  | Групповой программный проект |  |  |  | \* | Индивидуальный рефлексивный дневник проектаПисьменный групповой отчет по финальному проекту и программный кодЗащита проекта и ответов на дополнительные вопросы по применяемым методам |

1. **Критерии оценки знаний, навыков**

**Критерии оценки группового программного проекта**

*Письменный отчет и программный код*

Внимание при оценке уделяется следующим критериям, пропорционально влияющим на оценку.

* корректность применения методов агрегации и анализа данных, написанного программного кода
* ориентированность на заказчика
* качество и целостность выдвинутой системы гипотез
* хорошо структурированная организация и содержательная наполненность отчета
* использование навыков оформления графического материала, полученного в ходе обучения на настоящем курсе, корректность применения визуализаций и таблиц

Проект, исходный код которого (в виде RMarkdown) не собирается корректно из-за ошибок, не может получить оценку выше 5 (удовлетворительно). Полученная оценка за отчет и программный код выставляется всем участникам проектной команды

*Индивидуальный рефлексивный дневник*

Дневник заполняется еженедельно с момента разделения на проектные команды (список контрольных точек доступен в виртуальном образовательном окружении). Представляет собой ответы на вопросы о роли в проекте и индивидуальном прогрессе его выполнения, текущей работе студента (включая написанный им программный код и другие артефакты работы, если требуется заданием). Заполнение каждого раздела возможно только до заранее определенного для каждого задания срока.

За индивидуальный дневник выставляется оценка «зачтено», «не зачтено» на основе следующих критериев:

* полнота заполнения
* соответствие ответов поставленным вопросам
* ясность и понятность изложения

При заполнении менее 70% индивидуального дневника студенту за групповой программный проект выставляется оценка 0. 70% порог применяется и в случае пропусков дедлайнов по болезни и другим уважительным обстоятельствам. В случае документально подтвержденного длительного (более 5 недель) отсутствия по уважительным причинам студенту предлагается воспользоваться опциональными заданиями по курсу для компенсации оценки.

*Защита проекта*

Защита проекта проходит в виде презентации проекта и ответов на вопросы. В презентации участвуют все члены проектной команды.

Критериями оценки при защите проекта являются:

* хорошо структурированная организация и содержательная наполненность презентации
* отражение всех обязательных частей проекта (описание идеи, данных, состава команды, распределения ролей, используемых моделей, результатов)
* корректность ответов на вопросы аудитории

Полученная за защиту проекта оценка выставляется всем участникам проектной команды, присутствующим на защите. При отсутствии на защите по уважительной причине, студенту предоставляется возможность получения дополнительных вопросов, предназначенных для компенсации оценки за защиту, в дату до показа работ по курсу.

Общая оценка за программный проект выставляется по формуле

Опроект = 0,5·Одневник + 0,25·Оотчет + 0,25·Озащита, где

Одневник – оценка за индивидуальный рефлексивный дневник проекта («10» если зачтено, «0», если не зачтено»)

Оотчет – оценка за письменный отчет («10» если зачтено, «0», если не зачтено»)

Озащита – оценка за защиту проекта и ответы на дополнительные вопросы (по десятибалльной шкале)

Участники, заполнившие менее 50% индивидуального дневника, считаются не участвовавшими в выполнении группового проекта, им ставится оценка 0 за групповой программный проект.

**Критерии оценки подготовки к практикумам и работы на них**

Практическая направленность курса предполагает высокую вовлечённость в индивидуальную и групповую работу на практикумах. Её оценка предполагает два компонента:

*Выполнение онлайн-упражнений* на закрепление и проверку усвоения материала на онлайн-платформе Stepik (stepik.org). Их прохождение играет формативную роль: количество попыток не ограничивается.

*Работа на практикумах* оценивается по: присутствию на практикуме и наличию активности в виртуальном образовательном окружении в компьютерной аудитории во время практикума (исключения и технические ошибки должны быть доведены до сведения преподавателя или учебного ассистента и отмечаются ими в электронной ведомости), ответам на вопросы и выполнению текущих заданий на практикуме.

Оценка за подготовку к практикумам и работу на них: зачтено/не зачтено.

 Оценка “зачтено” ставится при одновременном выполнении следующих условий:

* прохождение суммарно более 50% заданий в установленные и отображаемые на онлайн-платформе Stepik дедлайны, без ограничения числа попыток (компонент Выполнение онлайн-упражнений)
* и активность не менее, чем на 50% практикумов в 3 модуле и не менее чем на 80% практикумов в 4 модуле (компонент Работа на практикумах).

При получении оценки “зачтено”, в формулу накопленной оценки выставляется 10, при получении оценки “не зачтено” -- 0.

В исключительных документально подтвержденных учебным офисом ОП студента уважительных случаях длительного отсутствия студента на занятиях оценка за подготовку к практикумам и работу за них может быть выставлена студенту только по онлайн-компоненте, при более чем 90% выполнении онлайн-заданий.

**Критерии заданий, включающих элементы анализа данных (в составе домашней и контрольных работ, отчетов)**

• корректность применения методик анализа (в рамках знаний, полученных в курсе, смежных дисциплинах, домашнем чтении) и интерпретации результатов;

• уверенность использования языковых средств и структур данных, методов преобразования и агрегации данных в организации потока анализа данных, их ввода и вывода.

**Опциональные формы контроля**

Студенту предоставляется несколько возможностей для улучшения индивидуальной оценки. Участие в них является добровольным, возможно получение оценки «10» за курс без участия в дополнительных формах. Суммарный вклад всех опциональных форм контроля в накопленную оценку каждого студента не может превышать 2 баллов.

Опциональные формы ориентированы на вклад студентов в развитие сообщества практики и обмена знаниями в рамках майнора и цифровых инициатив НИУ ВШЭ Санкт-Петербург и поддерживаются в рамках исследования НУЛ «Социология образования и науки» по обмену знаниями в образовательных окружениях и структурным детерминантам академической успешности.

**Критерии оценки активности на форуме**

Учитывается полезность вопросов и ответов (полезные вопросы отмечаются модераторами – учебными ассистентами и преподавателями дисциплины), их количество и содержательность наполнения. В качестве руководства для студентов можно ориентироваться на советы StackExchange: <https://stackoverflow.com/help/how-to-answer>

При подведении итогов семестра публикуется упорядоченный по убыванию рейтинг активности студентов и соответствующие ему бонусные оценки.

**Критерии оценки тьюториала**

Тьюториалы создаются в виде отдельного урока по заданной преподавателями теме на платформе для онлайн-курсов Stepik.org. Каждый тьюториал должен содержать теоретическую часть и не менее 5 заданий для самостоятельной работы. Преподавателями и учебными ассистентами оценивается качество и соответствие тьюториала выбранной теме.

Перед созданием урока студенту требуется пройти вводный курс «Epic Guide to Stepik» [https://stepik.org/course/1/](https://stepik.org/course/1/syllabus).

Получение оценки «не зачтено» для тьюториала предполагает возможность его переписывания вплоть до дня последнего семинара, при этом допускается не более трёх попыток сдачи.

**Критерии оценки «рецептов»**

Короткие рецепты, содержащие полезные и интересные примеры использования новых функций и библиотек в R, студенты могут опубликовать в виртуальном образовательном окружении. Все рецепты будут доступны для просмотра другим студентам. Общим голосованием студентов будут определены лучшие рецепты. Темы и количество возможных рецептов, которые может создать один студент, не ограничиваются. По итогам периода проектной деятельности с учетом количества студенческих голосов и оценки качества рецептов преподавателями не менее 10% студентов, принявших участие в создании рецептов, получают бонусную оценку за рецепты.

На практических занятиях в виде материалов в виртуальном образовательном окружении освещаются принципы создания эффективных рецептов. Также можно ознакомиться с руководством StackExchange: <https://stackoverflow.com/help/how-to-answer>

 Создание рецептов, содержащих значительные элементы отчетных заданий по курсу с привязкой к наборам данных не допускается, такие рецепты будут удалены, бонусная оценка за них не выставляется.

1. **Содержание дисциплины**

**Раздел 1. Введение в анализ текста в рекомендательных системах**

Инфраструктура анализа текста в R. Классические пакеты и новые разработки. Обработка текста. Регулярные выражения. Схожесть текстов. Анализ тональности текста. Тематическое моделирование.

Литература по разделу

1. Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach*. O’Reilly Media.
2. Grolemund, Garrett, and Hadley Wickham. 2016. *R for Data Science*. http://r4ds.had.co.nz/.

Количество часов аудиторной работы – 20

Общий объем самостоятельной работы – 40, из них выполнение заданий текущего контроля 20, подготовка к практическим занятиям 20.

Формы и методы проведения занятий по разделу: практические занятия в компьютерном классе.

**Раздел 2. Анализ и визуализация сетей в R. Социальные сети в бизнесе, экономике и обществе**

Основные понятия сетевого анализа. Меры центральности. Модели формирования и эволюции сетей. Алгоритмы Community Detection. Сильные и слабые связи. Пакет igraph. Создание графов. Импорт данных. Визуализация графов. Меры центральности в igraph. Социальный капитал. Примеры сетей (сети организаций, сети аффилиаций, сети дружбы, сети финансовых отношений, сети политических связей)

Литература по разделу

1. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning: With Applications in R (Springer Texts in Statistics). 2013 edition. Springer, 2013.
2. Easley D., Kleinberg J. Networks, crowds, and markets: Reasoning about a highly connected world. – Cambridge University Press, 2010.
3. Newman M., Barabasi A. L., Watts D. J. The structure and dynamics of networks. – Princeton University Press, 2006.
4. Barabasi A. L. Network Science. Cambridge University Press, 2016. URL: http://barabasi.com/book/network-science
5. Adamic L. A., Glance N. The political blogosphere and the 2004 US election: divided they blog //Proceedings of the 3rd international workshop on Link discovery. – ACM, 2005. – P. 36-43.
6. Dunbar R. I. M. The social brain: mind, language, and society in evolutionary perspective //Annual Review of Anthropology. – 2003. – P. 163-181.
7. Luke, Douglas. 2015. A User’s Guide to Network Analysis in R. Use R! Cham: Springer International Publishing.
8. Kolaczyk E.D. Statistical Analysis of Network Data with R / E.D. Kolaczyk, G. Csárdi. – New York: Springer, 2014. – 207 p.
9. igraph R package. URL: http://igraph.org/r/
10. igraph documentation. URL: http://igraph.org/r/doc/igraph.pdf

Количество часов аудиторной работы – 15

Общий объем самостоятельной работы – 30, из них выполнение заданий текущего контроля 15, подготовка к практическим занятиям 15.

Формы и методы проведения занятий по разделу: практические занятия в компьютерном классе.

**Раздел 3. Рекомендательные системы. Проектная организация цифровых исследований**

Рекомендательные системы. Алгоритмы построение рекомендательных систем. Социальные рекомендательные системы. Введение в проектную организацию цифровых исследований. Стадии проекта. Выполнение учебного проекта. Презентация результатов.

Количество часов аудиторной работы – 21

Общий объем самостоятельной работы – 64, из них выполнение заданий текущего контроля 20, подготовка к практическим занятиям 30, подготовка к экзамену 14

Формы и методы проведения занятий по разделу: лекции, практические занятия в компьютерном классе.

Литература по разделу

* Grolemund, Garrett, and Hadley Wickham. 2016. R for Data Science. http://r4ds.had.co.nz/.
1. **Образовательные технологии**

Преподавание языковых средств R и концепций анализа текста и сетевого анализа осуществляется с использованием современных абстракций, с упором на понимание на концептуальном уровне и формальным введением по мере необходимости.

Объединяющим методическим компонентом курса является учебный проект, выполняемый группами студентов с применением спектра рассматриваемых в курсе методов. Проект задаёт минимальную технологическую планку, делая акцент на поиске дополнительных источников информации, осмысленном выборе методов, обсуждении и интерпретации результатов. При этом технически сложные элементы выполняются при помощи учебных ассистентов и/или преподавателей. Обсуждение технических приёмов и методов, найденных дополнительных источников данных, обмен информацией со студентами, работающими по смежным темам, и peer review составляют существенную часть работы.

* 1. **Методические указания студентам по освоению дисциплины**

Для обеспечения необходимого уровня уверенного владения инструментальными средствами (языком R и средой RStudio) предусмотрен сквозной компьютерный практикум по всем разделам курса. Кроме того, рабочая среда с веб-доступом позволяет прозрачно переносить работу между практикумом и самостоятельной работой студента. Поэтому для успешного освоения дисциплины студент должен пользоваться возможностью самостоятельной работы, дополнительными ресурсами, указанными в программе и на форуме курса. **Пользуйтесь возможностью самостоятельной работы!**

Свои знания по части материалов курса можно проверить, используя онлайн-тесты. Хотя порог их прохождения на оценку достаточно низок, мы советуем вам проходить все тесты и задавать вопросы на форуме.

Учебные ассистенты, преподаватели, ваши однокурсники и старшекурсники часто делятся дополнительными ресурсами на форуме курса. **Пользуйтесь этой возможностью, задавайте и отвечайте на вопросы других.** Помимо возможного бонуса к оценке, это позволяет более глубоко усвоить материал.

Для коммуникации с преподавателями по административным вопросам пользуйтесь почтой курса.

Проект в составе этого курса даёт вам первую возможность сделать что-то сравнительно сложное в новой для вас области. Воспользуйтесь этой возможностью, чтобы углубить свои знания! Помимо возможного бонуса к оценке, это позволяет более глубоко усвоить материал.

* 1. **Учебно-методическая литература для самостоятельной работы студентов**
* Grolemund, Garrett, and Hadley Wickham. 2016. *R for Data Science*. <http://r4ds.had.co.nz/>
* Easley, David, and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. 1 edition. New York: Cambridge University Press. (online version available: www.cs.cornell.edu/home/kleinber/networks-book/)
1. **Оценочные средства для текущего контроля и аттестации студента**
	1. **Вопросы для оценки качества освоения дисциплины**
2. Зачем нужны алгоритмы выявления сообществ
3. Приведите пример сети соавторства
4. Объясните понятие гомофилии на примере
5. Объясните принцип работы алгоритма тематического моделирования
6. Что такое облако слов
	1. **Пример заданий, входящих в тесты и контрольную работу**

Примените различные метрики центральности к сети персонажей из "Отверженных" используя функции пакета igraph. Выпишите вторых по популярности персонажей на основе этих метрик (Valjean везде самый популярный)

degree(g) # Degree centrality
betweenness(g) # Betweenness centrality
closeness(g) # Closeness centrality
page\_rank(g)$vector # The Page Rank algorithm

#Для упорядочивания персонажей можете использовать шаблон ниже

library(dplyr)

data.frame(v = V(g)$label, degree = degree(g)) %>%
  arrange(...)

Извлеките эго-сеть для персонажа Gavroche, чтобы остались только те, кто появлялся с ним в одной главе. Укажите, сколько узлов в получившейся сети.

* 1. **Пример домашнего задания**

Домашнее задание состоит в анализе сети дружбы студента согласно данным сети ВКонтакте. Код для сбора данных предоставляется преподавателем и разбирается в классе. Студенту необходимо применить изученные методы выявления сообществ, определить центральность вершин и интерпретировать результаты. Отчет предоставляется как в письменной форме, так и в виде 2-3 минутной презентации о наиболее интересных и важных выводах.

* 1. **Примерное задание для программного проекта**

Одним из важных моментов при формировании навыков и знаний в рамках майнора "Обработка и анализ данных" является формирование навыков совместной работы над общим проектом. Одним из путей формирования таких навыков могут являться групповые проекты, выполняемые в небольших коллективах из 3-5 студентов разных образовательных программ. Количество студентов, проходящих обучение в рамках майнора, составляет 180-190 человек, что требует формирования большого числа (30-40) сравнимых между собой проектов. Этого можно достичь, сформировав общее для всех групп задание, проанализировать по заданным критериям набор данных, при этом набор данных для каждой группы формируется индивидуально, как фрагмент одного большого набора данных. Одним из примеров такого набора данных является "Amazon product data" – база, содержащая 142.8 миллионов обзоров различных товаров, оставленных в период с мая 1996 года по июль 2014. Кроме непосредственно текста обзора, в данных так же содержится оценка товара, оценка полезности обзора, идентификационный номер пользователя, написавшего обзор, а так же описание товара. Такая база позволяет формировать разные по сложности и по содержанию задания на разных этапах курса, в частности, задания по анализу текстов отзывов и их связь с оценкой, анализ сетевой структуры потребителей, выделение "схожих" по оценкам пользователей продуктов, для формирования рекомендаций и другие.

В качестве единого проектного задания используется построение рекомендательной системы на основе этих данных.

**11 Порядок формирования оценок по дисциплине**

**Накопленная оценка по дисциплине** рассчитывается с помощью взвешенной суммы оценок за отдельные формы текущего контроля знаний следующим образом:

О*накопленная*= 0.1*\*Отекущий 1* + 0.15\**Отекущий 2* + 0.2\**Отекущий 3* + 0.4\**Отекущий 4* + 0.15\**Отекущий 5,* где

*Отекущий 1* – оценка за контрольную работу 1,

*Отекущий 2* – оценка за контрольную работу 2,

*Отекущий 3* – оценка за домашнее задание по сетям,

*Отекущий 4* – оценка за групповой программный проект,

*Отекущий 5* –оценка за работу на практикумах и прохождение онлайн-упражнений и тестов.

Способ округления накопленной оценки текущего контроля: арифметический.

**Результирующая оценка по дисциплине** рассчитывается следующим образом:

*Орезульт = Онакопл,* где

*Онакопл* – накопленная оценка по дисциплине

Участие в опциональных формах контроля может вносить вклад в накопленную оценку не более 2 баллов.

1. **Учебно-методическое и информационное обеспечение дисциплины**
	1. **Основная литература**
* Grolemund, Garrett, and Hadley Wickham. 2016. *R for Data Science*. <http://r4ds.had.co.nz/>
* Easley, David, and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. 1 edition. New York: Cambridge University Press. (online version available: www.cs.cornell.edu/home/kleinber/networks-book/)
	1. **Дополнительная литература**
* Barabási, Albert-László. 2016. Network Science. Cambridge University Press. (online version available: barabasi.com/book/network-science)
* Luke, Douglas. 2015. A User’s Guide to Network Analysis in R. Use R! Cham: Springer International Publishing.
* Cioffi-Revilla, Claudio. 2014. Introduction to Computational Social Science: Principles and Applications. 2014 edition. London ; New York: Springer.
* Mark EJ Newman. 2010. Networks: An Introduction. 1 edition. Oxford ; New York: Oxford University Press.
* Kolaczyk E.D. Statistical Analysis of Network Data with R / E.D. Kolaczyk, G. Csárdi. – New York: Springer, 2014. – 207 p.
* The SAGE Handbook of Social Network Analysis. 2011. London ; Thousand Oaks, Calif: SAGE Publications Ltd.
* Robins, Garry L. 2015. Doing Social Network Research: Network-Based Research Design for Social Scientists. SAGE Publications Ltd.
* Kolaczyk E.D. Statistical Analysis of Network Data with R / E.D. Kolaczyk, G. Csárdi. – New York: Springer, 2014. – 207 p.
	1. **Ресурсы информационно-телекоммуникационной сети “Интернет”**
1. igraph R package. URL: <http://igraph.org/r/>
2. stepik.org
3. datacamp.com
4. edx.org

Большое количество онлайн-курсов и дополнительных ресурсов по R и анализу данных во всем его многообразии можно найти в интернет и на основных площадках образовательных курсов. Их всегда можно обсудить с одно и старшекурсниками, учебными ассистентами и преподавателями на форуме.

Часть материалов курса доступна в виде интерактивных упражнений на платформе Степик.

* 1. **Программные средства**

RStudio. Пакет MS Office/OpenOffice.org

1. **Материально-техническое обеспечение дисциплины**

Аудитория с проектором для лекций, компьютерные аудитории с современными версиями браузеров согласно требованиям RStudio Server к клиентам. Клиент ssh для консольного доступа к серверу. Сервер для работы студентов (спецификация в зависимости от количества записавшихся на майнор) с Ubuntu Linux, RStudio Server и другими пакетами, в зависимости от специфики проектов.