

Майнор

«Обработка и анализ данных»

electives.hse.ru/minor_data_spb/

Руководитель

Александр Владимирович Сироткин

www.hse.ru/staff/avsirotkin

avsirotkin@hse.ru



Темы и методы

- **“Computational / data thinking”**
- **Как превратить мысли и идеи в код**
- **Что вообще такое интеллектуальный анализ данных (data mining), анализ социальных сетей (social network analysis), и анализ текста (text mining)**
- **Что такое сервисы, основанные на данных (data-driven services)**



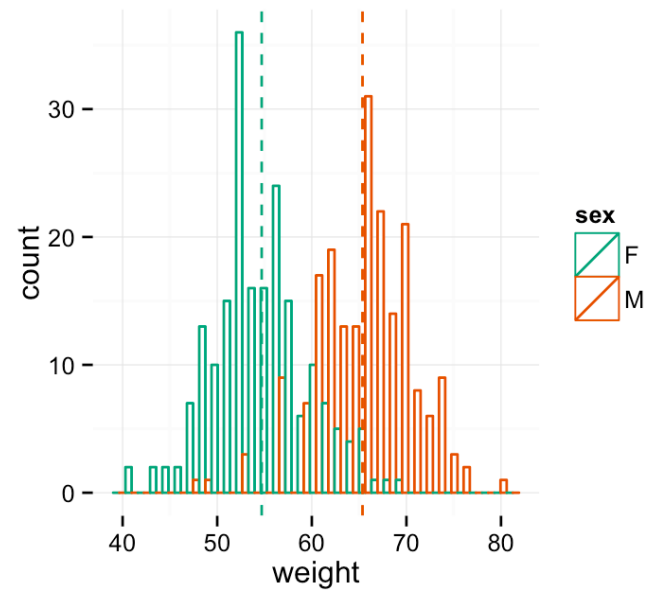
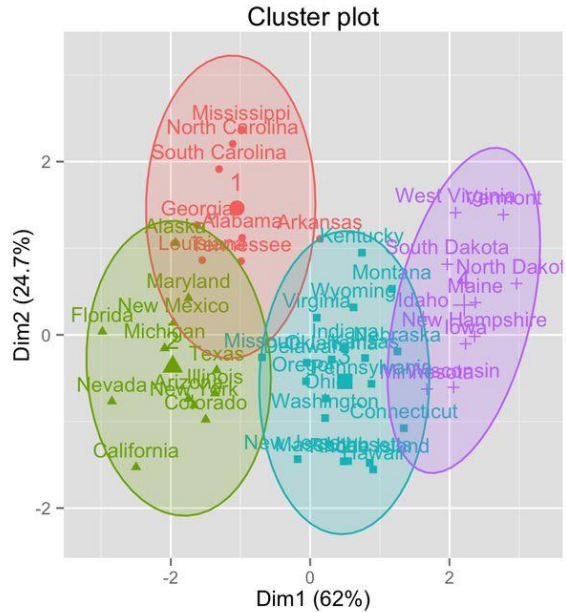
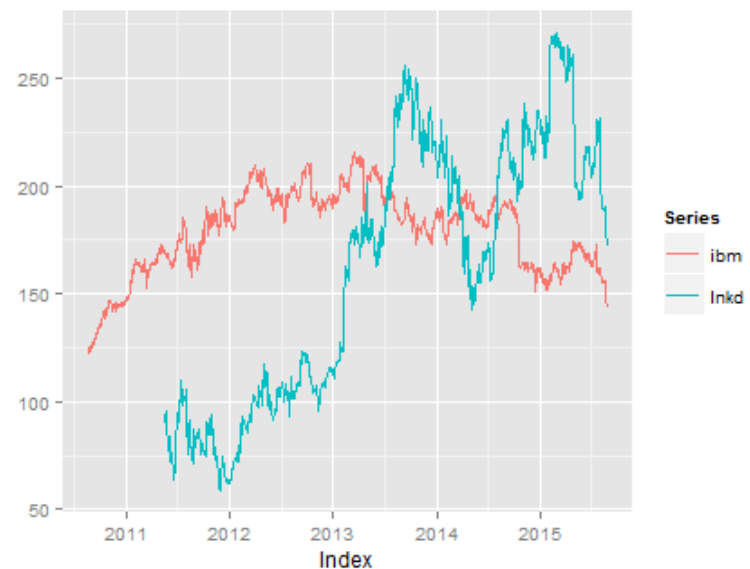
Этапы анализа данных

1. Постановка и уточнение вопроса.
2. Исследование данных (разведка).
3. Построение формальных статистических моделей.
4. Интерпретация результатов.
5. Распространение результатов (презентация).

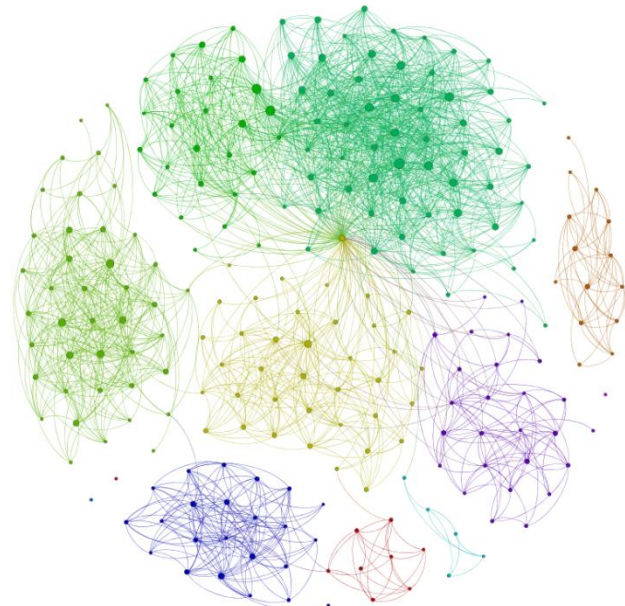
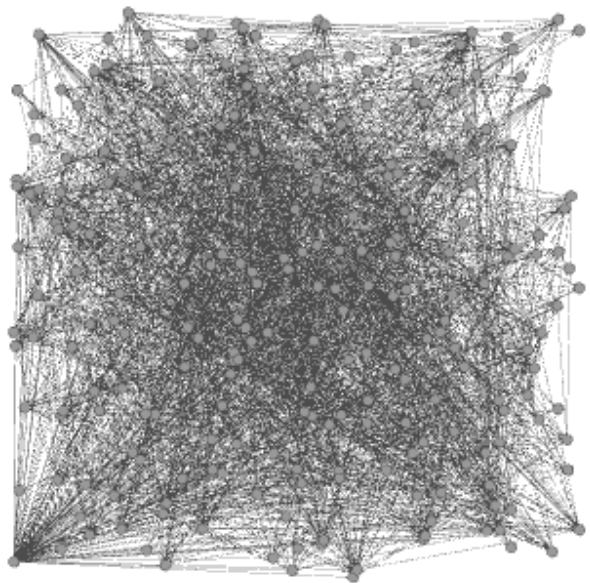


Обобщение и визуализация

1	2	3	4	5	6	7	8		
1988	67.0	Mass	2	33,543	1973	38	Mass	14	17,637
1976	41.7	Mass	5	33,447	1974	34.3	Mass	16	45,675
1997	26.2	Mass	14	22,188	1975	15.3	Mass	17	38,747
1998	26.1	Mass	15	21,798	1976	12.6	Mass		33,240
1999	27.1	Mass	16	20,934	1977	17.4	Mass	1	24,340
1990	8.4	Mass	17	15,729	1978	97.5	Mass	2	48,984
1991	2.7	Mass		27,707	1979	119.4	Mass		103,074
1992	5.1	Mass	1	21,308	1980	154.6	Mass	1	100,812
1993	24.4	Mass	2	24,666	1991	148.5	Mass	2	59,774
1994	47.1	Mass	3	22,857	1992	119.9	Mass	3	68,487
1995	67.3	Mass		16,594	1993	66.8	Mass	4	78,736
1996	54.1	Mass	3	19,158	1994	45.7	Mass	3	65,575
1997	51.7	Mass	2	19,219	1995	18	Mass	16	68,636
1998	48.9	Mass	5	18,990	1996	15.4	Mass		37,168
1999	41.9	Mass	4	18,169	1987	73.1	Mass	1	48,017
1918	18.0	Mass	9	11,410	1988	198.2	Mass	2	24,968
1911	7.7	Mass	16	11,410	1999	17.9	Mass		14,786
1912	1.9	Mass	17	10,874	1999	147.3	Mass	1	42,914
1913	1.4	Mass		77,683	1991	119.7	Mass	7	44,768
1914	9.6	Mass	1	10,664	1992	64.3	Mass	1	32,681
1915	47.4	Mass	11	14,513	1993	54.6	Mass	14	27,684
1916	21	Mass	17	11,032	1994	70.9	Mass	7	35,266
1917	184.7	Mass		38,788	1995	17.4	Mass	16	38,435
1918	80.8	Mass	3	71,108	1996	36	Mass		47,180
1919	61.5	Mass	2	27,499	1997	21.3	Mass	1	28,181
1920	27.8	Mass	5	26,158	1998	64.3	Mass	2	18,467
1921	26.1	Mass	4	21,877	1999	93.5	Mass	5	38,934
1922	14.2	Mass	19	27,721	2000	119.6	Mass		48,138
1923	2.8	Mass		18,578	2001	111	Mass	1	32,635
1924	16.1	Mass	1	10,787	2002	104	Mass	2	32,924
1925	44.4	Mass	7	21,873	2003	65.7	Mass	3	37,093
1926	61.9	Mass	1	24,132	2004	49.4	Mass	4	47,494
1927	68.8	Mass	4	19,111	2005				
1928	27.8	Mass		16,158	2006				
1929	44.8	Mass	3	17,437	2007				
1930	15.8	Mass	7	16,833	2008				
1931	21.1	Mass	8	10,898	2009				
1932	21.2	Mass	4	15,678	2010				
1933	7.5	Mass		14,103	2011				
1934	8.7	Mass	1	17,669	2012				
1935	36.1	Mass	1	16,718	2013				
1936	79.1	Mass	1	18,612	2014				
1937	114.4	Mass		19,448	2015				
1938	189.6	Mass	1	18,474	2016				



Поиск скрытых структур и закономерностей



Create a Pre-Assessment: Put the certification test at the beginning of your course. Make it a very comprehensive assessment so that you can truly identify their knowledge and skill level. If the user passes the test, he jumps to the end and is certified. If the user doesn't pass, then you direct him to the course where he can get remedial training and additional assistance.

Empower the User: The first idea is to create the assessment up front and then direct the user based on the assessment result. While it is a simple approach and easy to design, this can be intimidating for some users. Here's a way to soften it up and empower them at the same time. Addition, some customers just aren't comfortable with this type of approach where the user can self-navigate and choose when to take the assessment. They don't like the fact that people can test out. Instead, they want them exposed to something that resembles a course.

Break the course content into distinct sections.: At the beginning of each section, give the user a choice to assess or go through the content. At the end of all of the sections, do a final assessment. You can still capture some time savings because a knowledgeable person can go through each section and test out quickly. However, by breaking it into sections you can be more specific in the assessment process and catch areas where people might not be as fluent.



Курсы майнора

- 1. Программирование для анализа данных и воспроизводимые исследования – Основы R, построение графиков и отчетов, агрегация данных.**
- 2. Анализ данных и технологии работы с данными – Основы обработки структурированных и не структурированных данных, работа с текстам и сетями.**
- 3. Интеллектуальный анализ данных и основы машинного обучения – изучение продвинутых моделей.**
- 4. Приложения и практика анализа данных – командные проекты.**



Что мы будем делать

- **находить скрытые связи**
- **объяснять**
- **предсказывать**
- **рекомендовать**
- **...и многое другое**

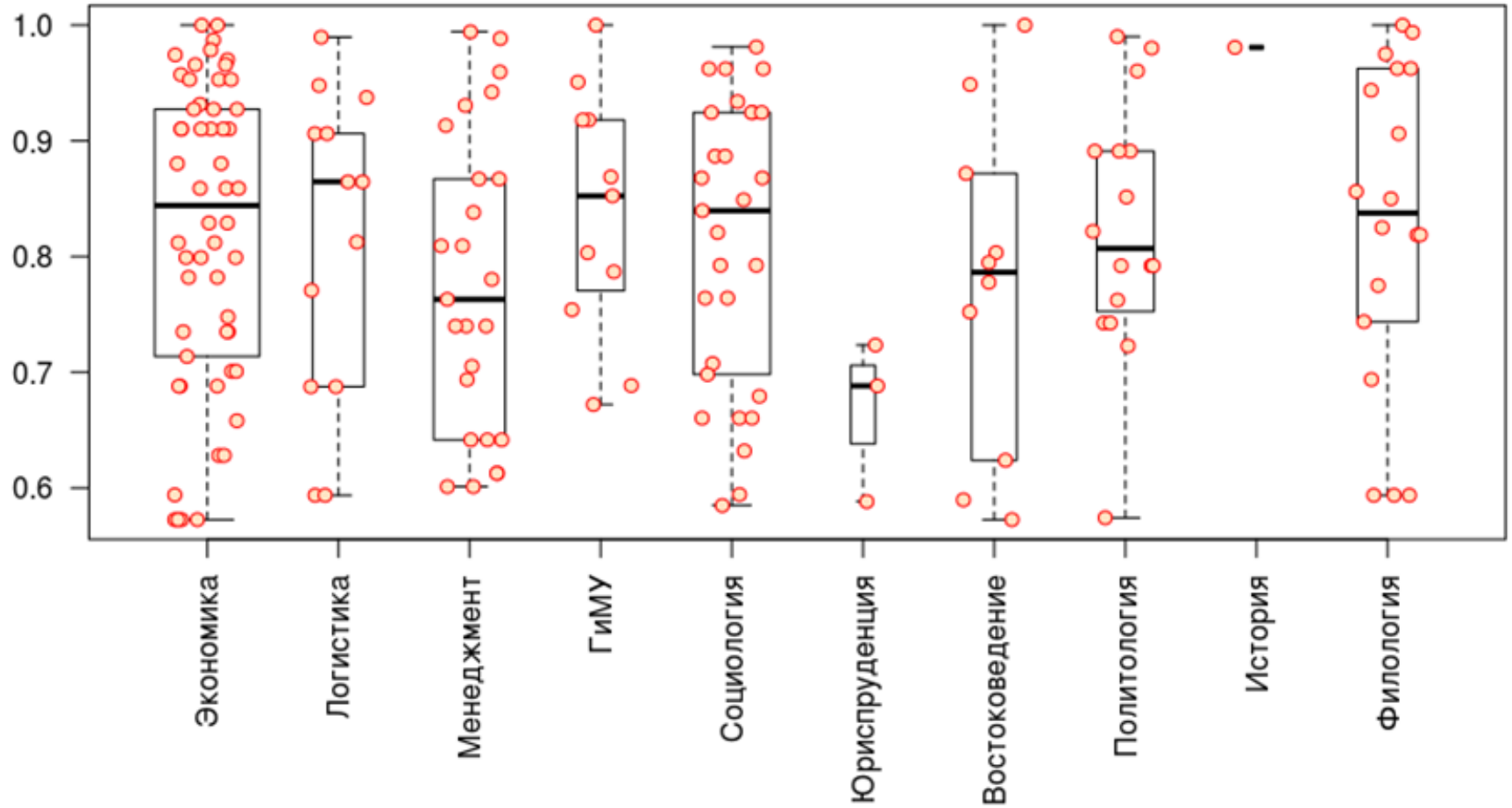


Благодаря нам студенты научились

- **выбрать дорамы**
- **находить лучшую шаверму в Санкт-Петербурге**
- **искать лучшую страну в мире**
- **советовать книги и фильмы**
- **читать твиттер и долго думать**
- **понимать, как отзывы связаны с оценками отелей**
- **выбирать китайский ресторан в Китае**
- **и многое другое**



Распределение студентов



Майнор

«Обработка и анализ данных»

electives.hse.ru/minor_data_spb/

Руководитель

Александр Владимирович Сироткин

www.hse.ru/staff/avsirotkin

avsirotkin@hse.ru

