

Майнор

«Обработка и анализ данных» v2.0

http://electives.hse.ru/2015/minor_data_spb/
http://electives.hse.ru/minor_data_spb/

А.В. Сироткин

Data Science

Наука о том, что делать с данными

- Где брать?
- Что брать?
- Как анализировать и делать выводы?
- Как проверять выводы?

- Кому все это надо?

Мы будем учиться:

- Собирать данные:
 - Работать с базами
 - Грабить интернет
- Анализировать данные
 - Ну тут очень много чего...
- Рассказывать о данных и наших выводах
 - Рисовать красивые картинки и графики
 - Объяснять, почему это круто

Курсы майнора

- Программирование для анализа данных и воспроизводимые исследования
- Анализ данных и технологии работы с данными
- Интеллектуальный анализ данных и основы машинного обучения
- Приложения и практика анализа данных

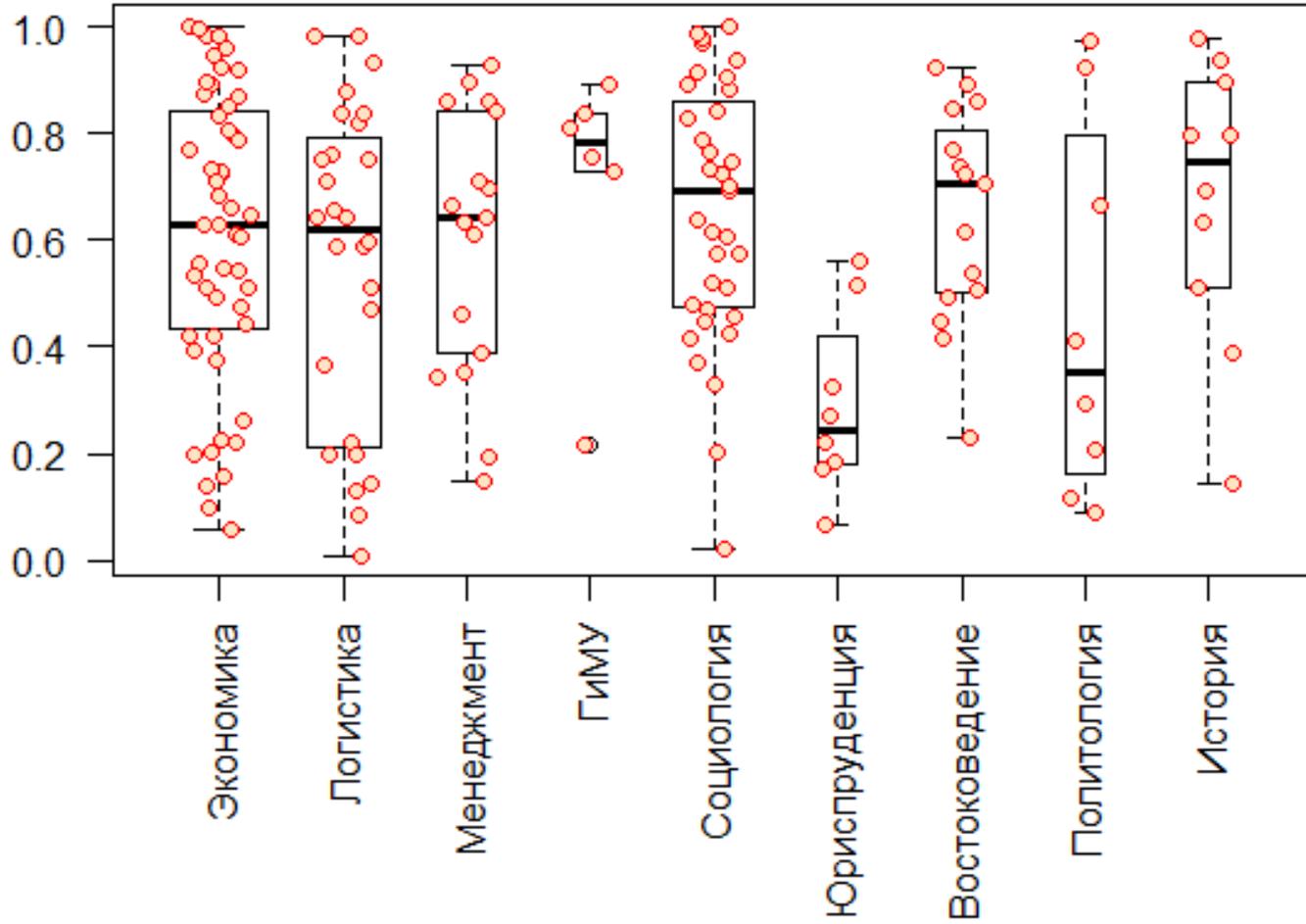
Проектная работа

- Мы будем делать проекты
- Уже делались или делаются:
 - Музыкальные предпочтения студентов Вышки и студентов ПТУ
 - Анализ фанфиков
 - Анализ поведения игроков в онлайн играх
 - Что обсуждают в кавказских группах Вконтакте
 - Предсказание победителя в Dota 2
 - и много чего еще...
- Можно придумать свой проект и сделать его.
- В последнем семестре никто не уйдет без проекта

Требования к студентам

- Хотеть изучать анализ данных
- Не испытывать непреодолимого страха перед компьютером
- Быть готовым поработать
- Немножко знать английский

Кто пришел в прошлом году



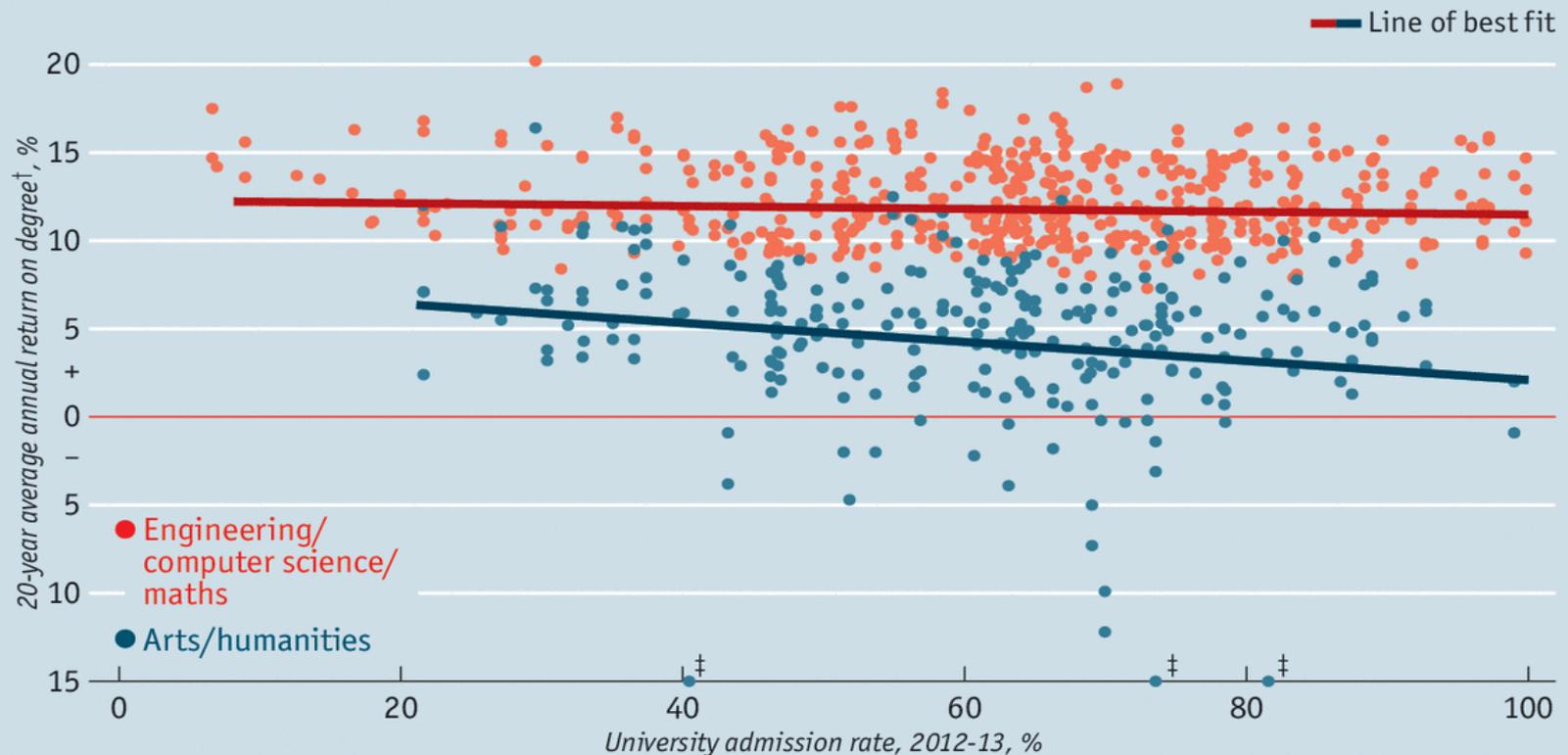
А теперь немного веселых историй про анализ данных

Картинки из разных исследований с комментариями

Точные науки против гуманитарных

It's what you learn, not where

American universities*, selectivity and returns

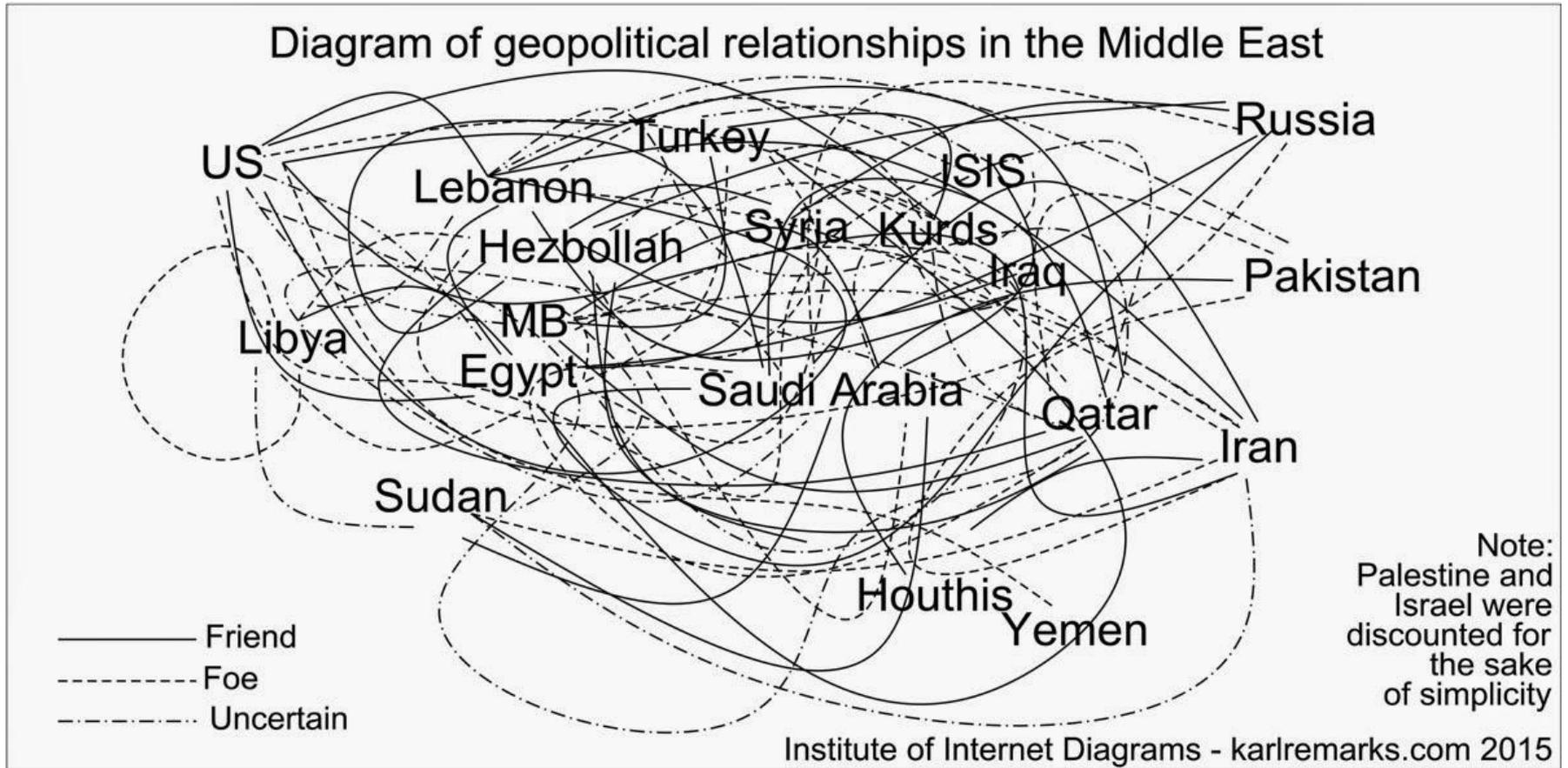


Sources: PayScale; National Centre for Education Statistics

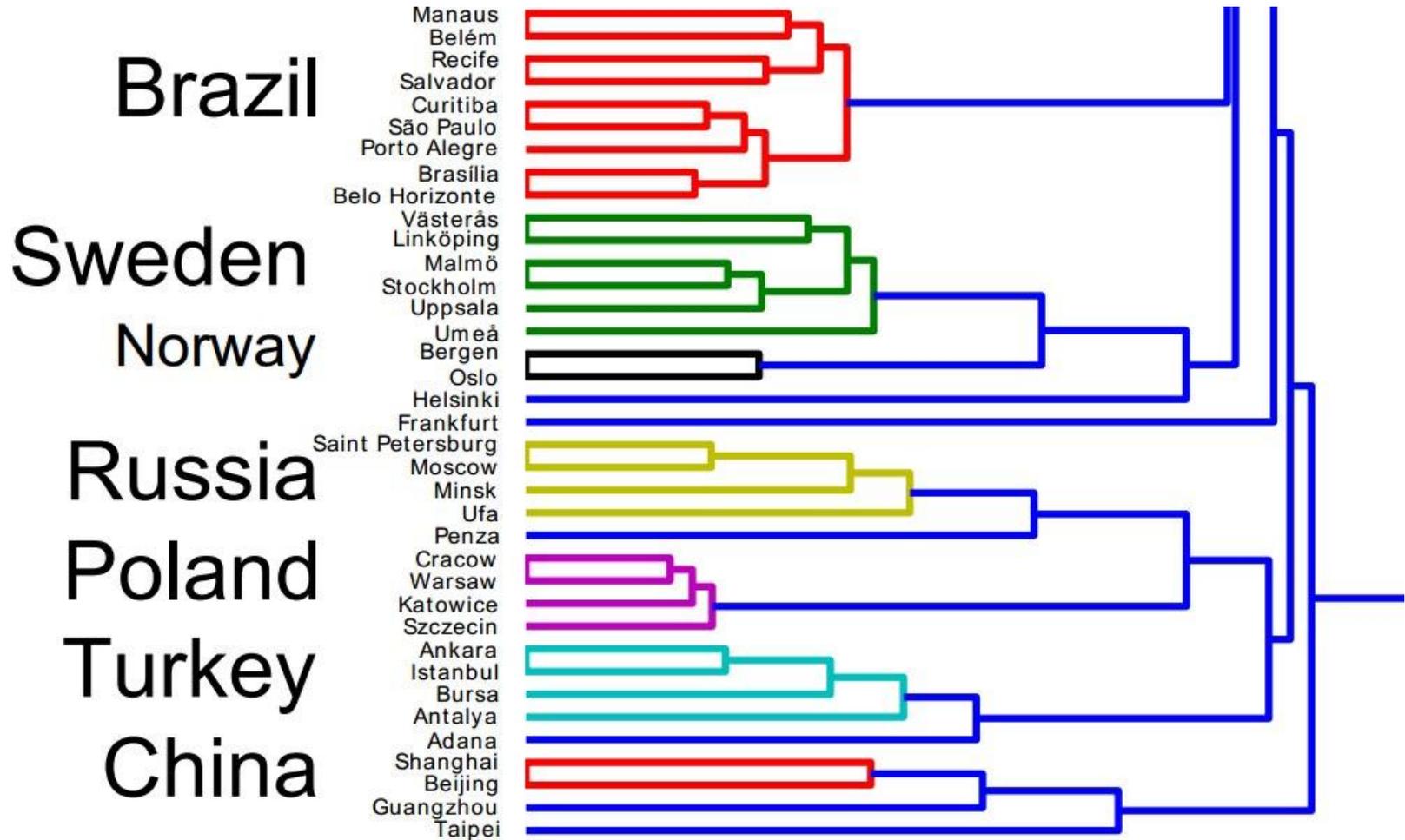
*Sample of 240 institutions †Based on 2012-13

‡Returns of minus 15% or worse

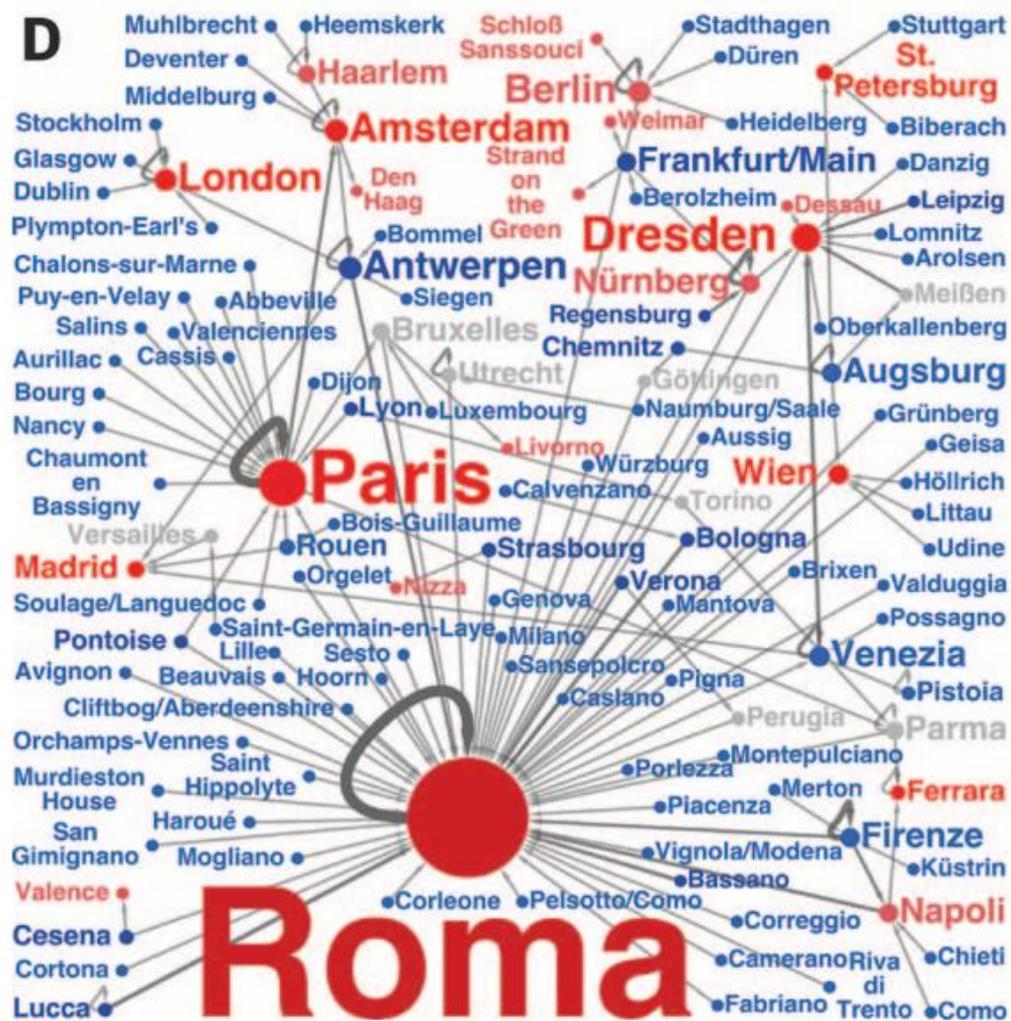
Все сложно?...



Схожесть музыкальных вкусов



Где умирали знаменитости в 18 веке



From “A network framework of cultural history”

Science 1 August 2014:
Vol. 345 no. 6196
pp. 558-562
DOI: 10.1126/science.1240064

DIY:Justice

- 33 млн судебных решений различных судов РФ стали открытыми данными
- 150Gb данных в сжатом виде
- Их можно анализировать
- Их можно классифицировать
- И, возможно, что-то даже предсказывать

Анализ данных и законы

Tasks	Research goals	Methodologies	Relevant works
Analysis of legal systems	<ul style="list-style-type: none"> Analysis of (hidden) structural and dynamic properties of legal systems in order to characterize and compare different legal corpora or juridical cultures Study of the evolution of legal systems over time 	<ul style="list-style-type: none"> Automated information extraction Complexity modelling Social network analysis 	Pagallo (2010); Katz <i>et al.</i> (2011)
Analysis of normative texts	<ul style="list-style-type: none"> Analysis of legal concepts and of their relations Study of the evolution of norms over time. 	<ul style="list-style-type: none"> Automated information extraction Social network analysis 	Bommarito and Katz (2009, 2010); Boulet, Mazzega and Bourcier (2010)
Case law analysis	<ul style="list-style-type: none"> Analysis of the relations between precedents. Study of the evolution of case law over time. Analysis of the relevance of precedents in case law 	<ul style="list-style-type: none"> Automated information extraction Social network analysis 	Chandler (2005); Fowler <i>et al.</i> (2007); Pagallo (2007); Katz and Stafford (2010); Vining and Wilhelm (2011)

Спасибо за внимание!

- Свежая информация про майнор:

http://electives.hse.ru/2015/minor_data_spb/

http://electives.hse.ru/minor_data_spb/

- Истории социологов про курсы и проекты:

<http://nosoc.io>

Ждем вас!