

Использование современных методов статистического анализа в образовательном процессе

Е.Н. Матеров

26 сентября 2018

ФГБОУ ВО «Сибирская пожарно-спасательная академия» ГПС МЧС России
кафедра физики, математики и информационных технологий

- **Наука о данных** (Data Science) — междисциплинарная область, изучающая проблемы обобщения, анализа, алгоритмизации и представления данных в численной и визуальной формах.
- Наука о данных опирается на инструменты эмпирических наук, статистики, отчетности, анализа, визуализации, бизнес-аналитики, экспертных систем, машинного обучения, баз данных, хранения данных, интеллектуального анализа данных и больших данных.

Инструменты для обработки данных



а) Microsoft Excel



б) Statistica



в) IBM SPSS



г) Matlab



д) Python

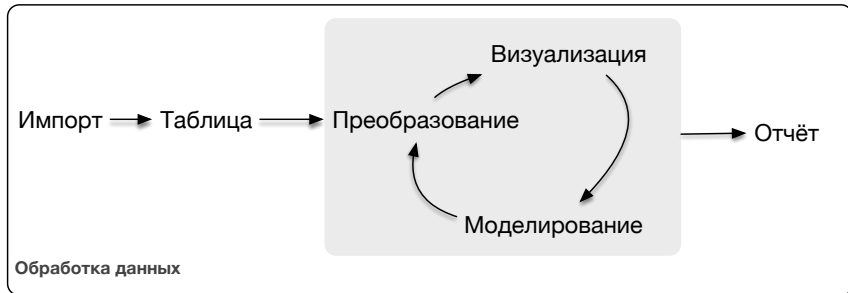


е) R

Преимущества языков программирования для анализа данных

- Бесплатность
- Возможность обработки больших объёмов данных
- Автоматизация вычислений, воспроизводимость результата
- Использование современных актуальных алгоритмов
- Графические возможности
- Организация проектов, работа с Git
- Формирование интерактивных отчётов

Анализ данных



Hadley Wickham, Garrett Grolemund

R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.
O'Reilly, 2016.



**МАТЕМАТИЧЕСКИЕ МЕТОДЫ И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ В НАУЧНЫХ ИССЛЕДОВАНИЯХ**

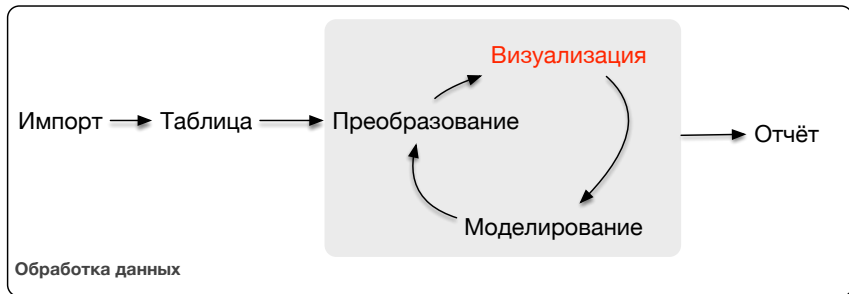
Учебное пособие

Железногорск
2018

Оглавление

Введение	5
1 Описательная статистика	8
1.1 Характеристики средней тенденции данных	11
1.2 Показатели вариации данных относительно среднего	14
1.3 Показатели формы эмпирических распределений	17
1.4 Программные средства для описательных статистик	19
1.5 Проверка одномерного распределения на нормальность	22
1.6 Способы графического представления данных	29
2 Элементы корреляционного анализа	55
2.1 Линейный коэффициент корреляции Пирсона	55
2.2 Коэффициент корреляции Спирмена	68
2.3 Коэффициент корреляции Кенделла	72
2.4 Связь коэффициентов корреляции	74
2.5 Программная реализация корреляционного анализа	75
3 Основы регрессионного анализа	78
3.1 Парная линейная регрессия	80
3.2 Множественная линейная регрессия	93
3.3 Программная реализация линейной регрессии	97
4 Временные ряды	113
4.1 Общие представления о временном ряде	113
4.2 Примеры временных рядов	121
4.3 Статистические функции временного ряда	126
4.4 Стационарные временные ряды	131
4.5 ARIMA-модель временного ряда	139
4.6 Программный анализ временных рядов	152

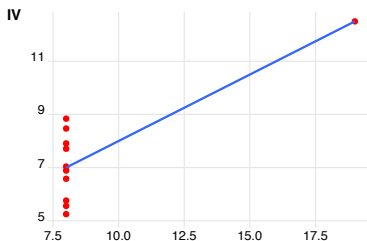
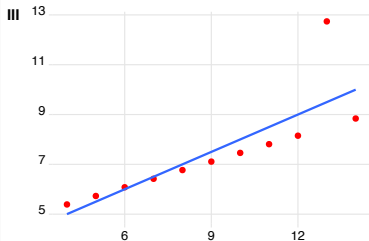
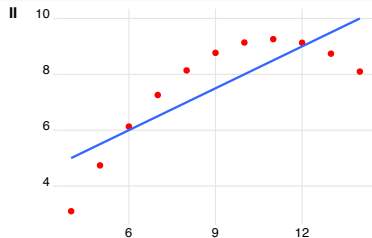
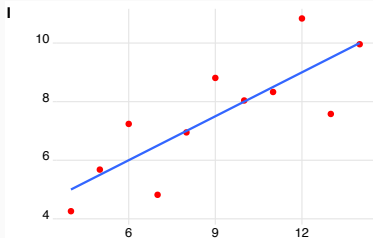
Визуализация данных



“ Самая большая ценность графика — это когда он заставляет нас замечать то, чего мы никогда не ожидали увидеть! ”

Джон Тьюки (американский статистик)

Зачем визуализировать данные?



Общие рекомендации по визуализации

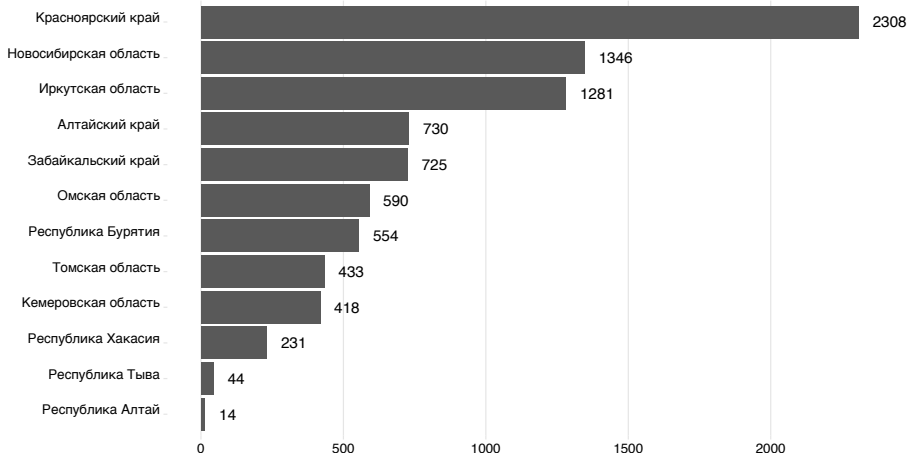
1. Используйте правильный **тип графика**
2. Располагайте данные в **порядке**, согласно иерархии
3. Используйте **панелирование**, цвет, форму и площадь для упрощения
4. У графика должно быть **название** и полная **легенда**
5. Используйте правильный выбор **масштаба**
6. Используйте максимально **простой** дизайн графиков

Классификация графиков по типам данных

переменная (x)	отклик (y)	тип графика
численная		гистограмма
		график плотности
категорная		столбиковая диаграмма
численная	численный	диаграмма рассеяния
хронологическая	численный	линейный график
категорная	численный	диаграмма размаха
категорная	категорный	мозаичная диаграмма

Столбиковая диаграмма

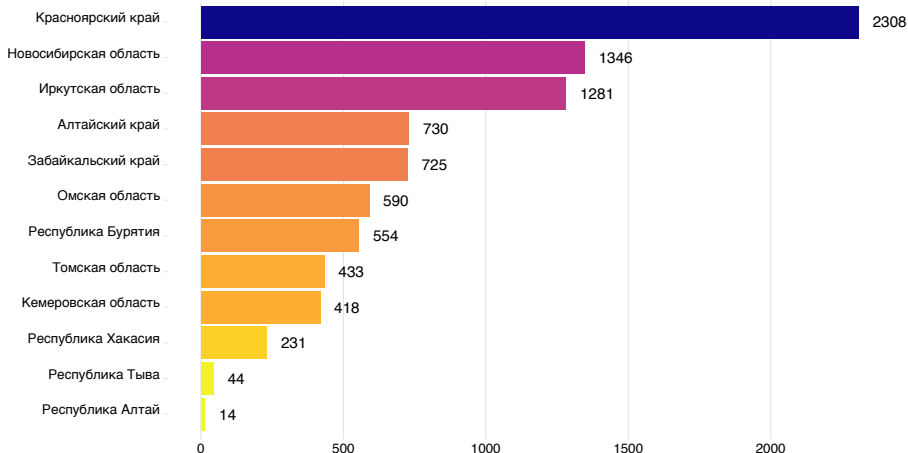
Количество спасённых на пожарах за 12 месяцев 2017 года
по Сибирскому федеральному округу



Количество спасённых, чел.
по данным ФГБУ ВНИИПО МЧС России

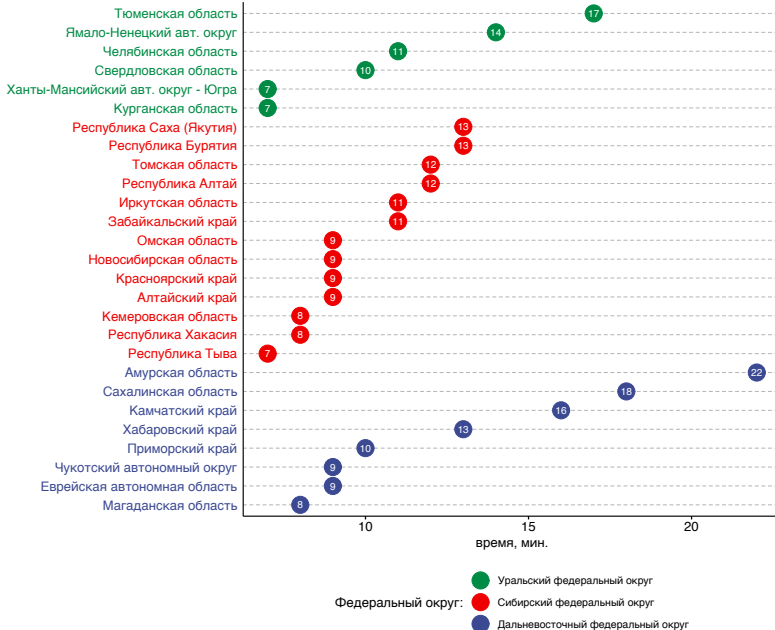
Столбиковая диаграмма

Количество спасённых на пожарах за 12 месяцев 2017 года
по Сибирскому федеральному округу



Количество спасённых, чел.
по данным ФГБУ ВНИИПО МЧС России

Среднее время тушения пожаров в городе в 2016 году
данные сгруппированы по трём федеральным округам

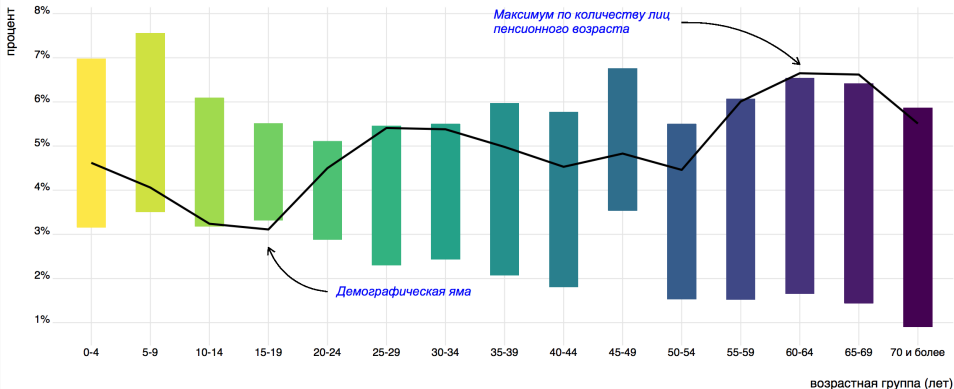


Пример анализа динамики

Распределение населения по возрастным группам в России с 1927 по 2018 год (в процентном соотношении)

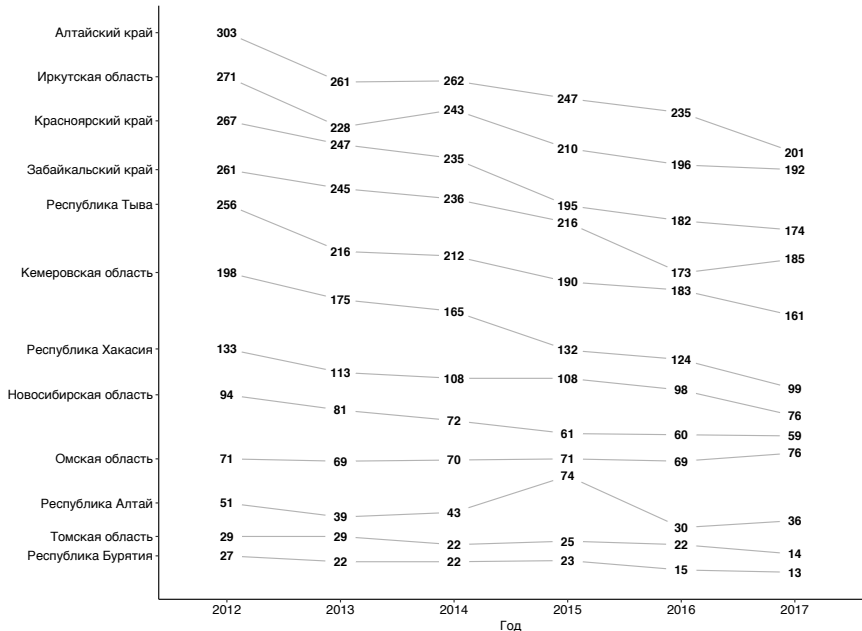
по данным Федеральной службы государственной статистики

линия соответствует 2018 году



Количество погибших на пожарах в Сибирском федеральном округе, чел.

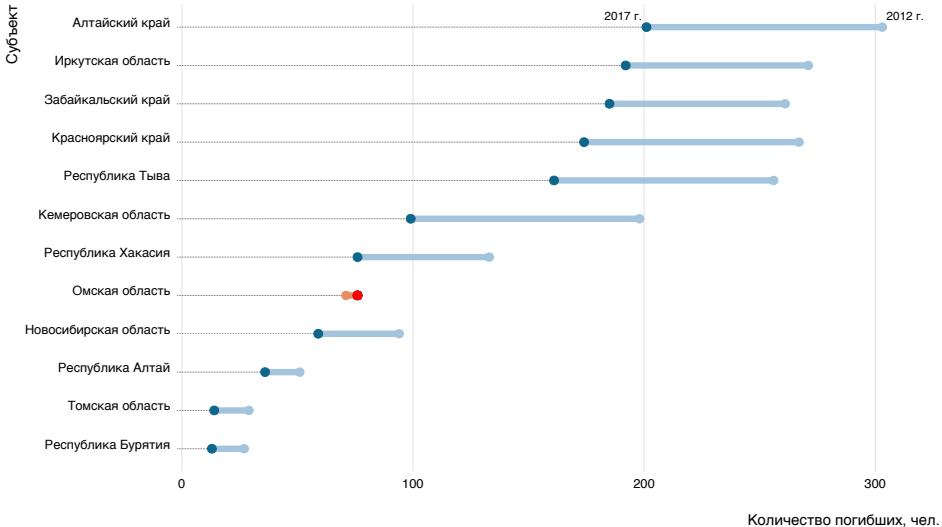
Динамика с 2012 г. по 2017 г.



по данным ФГБУ ВНИИПО МЧС России

Количество погибших на пожарах в Сибирском федеральном округе, чел.

Динамика с 2012 г. по 2017 г.



по данным ФГБУ ВНИИПО МЧС России

Пример табличных данных

Пожары и пожарная безопасность в 2015 г.

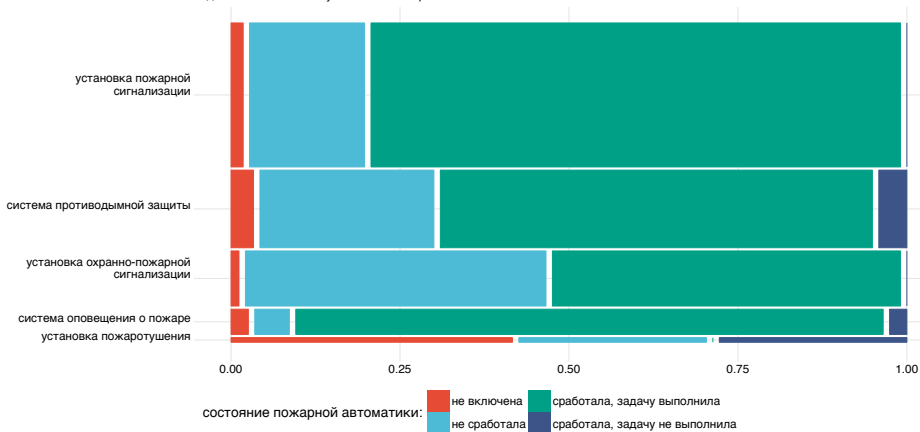
Таблица 62

Эффективность работы пожарной автоматики при пожарах в зданиях жилого назначения в 2014-2015 гг.

Вид пожарной автоматики	В С Е Г О		Сработала, задачу выполнила		Сработала, задачу не выполнила		Не сработала		Не включена	
	2014	2015	2014	2015	2014	2015	2014	2015	2014	2015
	Количество пожаров, ед. Прямой ущерб, тыс. руб. Спасено материальных ценностей, тыс. руб. Погибло людей, чел. Травмировано людей, чел.									
1	2	3	4	5	6	7	8	9	10	11
В С Е Г О	450	387	304	277	11	8	127	96	20	11
	20876	44789	13213	22760	1879	15	5732	22036	53	557
	75032	129382	57202	116832	100	0	13805	11550	4525	1000
	39	30	22	18	1	1	17	11	1	0
	69	106	55	91	1	3	13	11	2	3
Установка охранно-пожарной сигнализации	43	83	35	44	-	-	4	38	4	1
	2648	15614	2248	4617	-	-	400	10997	0	0
	5950	53450	4950	43450	-	-	0	0	1000	0
	3	8	1	6	-	-	2	2	0	0
	1	7	0	4	-	-	1	3	0	0

Мозаичная диаграмма

**Эффективность работы пожарной автоматики в РФ в 2015 году
при пожарах в зданиях жилого назначения**
в долях от количества участия на пожарах

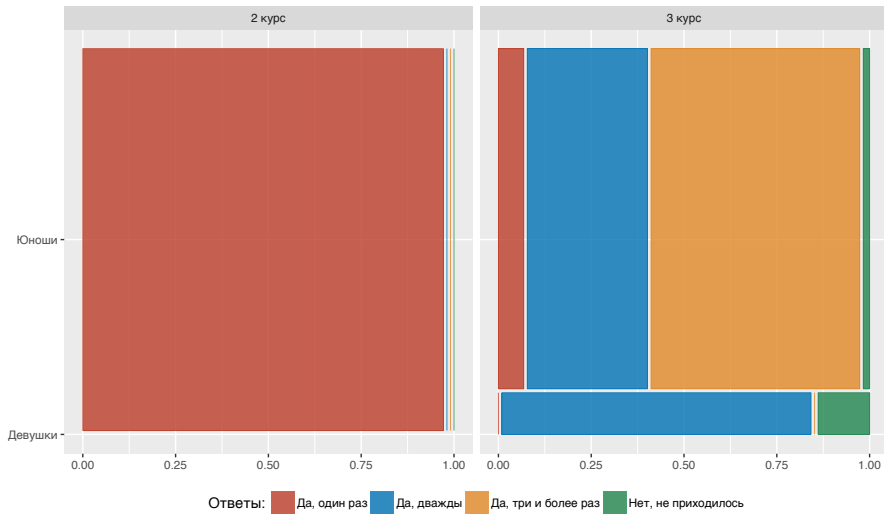


по данным ФГБУ ВНИИПО МЧС России

Мозаичная диаграмма

Вопрос 1: Приходилось ли Вам проходить стажировку в Институте развития?

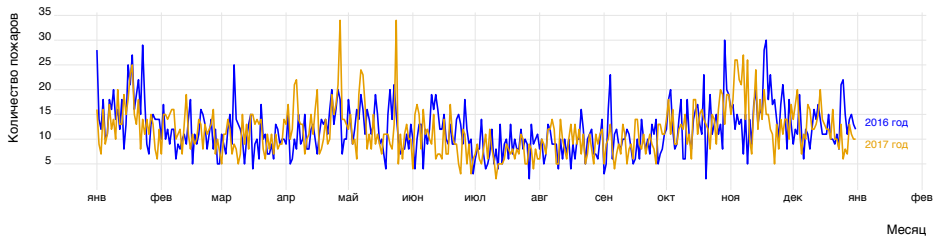
Указаны соотношения



Тепловые карты

Количество пожаров в Красноярском крае в 2016 и 2017 годах

синий цвет - 2016 год, оранжевый - 2017 год



Календарь количества пожаров в сутки в Красноярском крае за 2016 и 2017 год

Тепловая карта



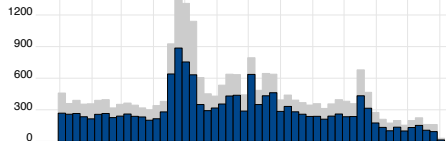
Панелирование

Количество пожаров и загораний в неделю в Красноярском крае в 2013 году

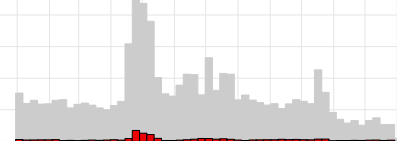
разбиение сделано по типам населённых пунктов

Количество пожаров

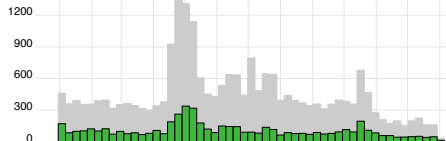
Город



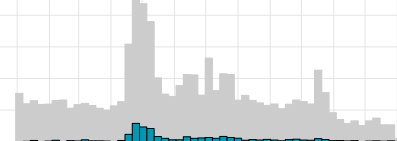
Населенный пункт городского типа



Сельский населенный пункт



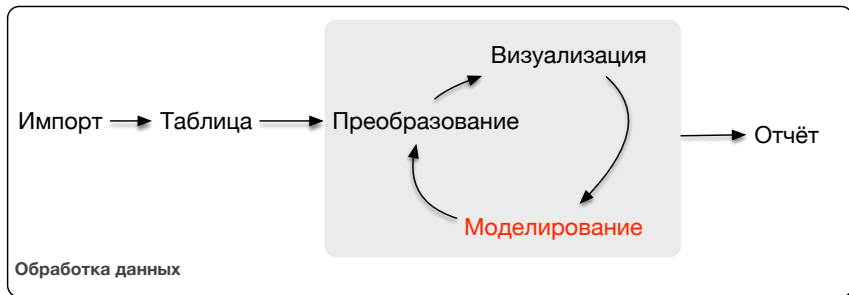
Вне территории населенного пункта



Месяц

источник данных: ГУ МЧС России по Красноярскому краю

Моделирование





Пример данных: электронная карточка учёта пожаров в Красноярском крае за 2017 год

F2	F6	F9	F11	F20
1 Енисейский район	2017-01-04	0	Город	Обособленная свалка бытовых и промышленных отхо...
2 Ирбейский район	2017-01-02	0	Сельский населенный пункт	Надворная постройка
3 Сухобузимский район	2017-01-01	0	Сельский населенный пункт	Многоквартирный жилой дом
4 Сухобузимский район	2017-01-07	0	Сельский населенный пункт	Полосы отчуждения и обочины дорог, луга, пустыри (в...
5 Сухобузимский район	2017-01-10	0	Сельский населенный пункт	Надворная постройка
6 Сухобузимский район	2017-01-17	0	Сельский населенный пункт	Многоквартирный жилой дом
7 Рыбинский район	2017-01-08	0	Сельский населенный пункт	Многоквартирный жилой дом
8 Рыбинский район	2017-01-07	0	Город	Полосы отчуждения и обочины дорог, луга, пустыри (в...
9 Рыбинский район	2017-01-11	0	Город	Полосы отчуждения и обочины дорог, луга, пустыри (в...
10 Туруханский район	2017-01-07	0	Сельский населенный пункт	Легковой автомобиль
11 Туруханский район	2017-01-09	0	Сельский населенный пункт	Многоквартирный жилой дом
12 Туруханский район	2017-01-09	0	Сельский населенный пункт	Неэксплуатируемое здание (сооружение)
13 Туруханский район	2017-01-16	0	Город	Склад со смешанным или универсальным ассортимент...
14 Бирилюсский район	2017-01-07	0	Сельский населенный пункт	Легковой автомобиль
15 Бирилюсский район	2017-01-15	0	Сельский населенный пункт	Многоквартирный жилой дом
16 Большеулуйский район	2017-01-16	0	Сельский населенный пункт	Одноквартирный жилой дом
17 Большеулуйский район	2017-01-18	0	Сельский населенный пункт	Баня, сауна на территории домовладения
18 Краснотуранский район	2017-01-06	0	Сельский населенный пункт	Одноквартирный жилой дом

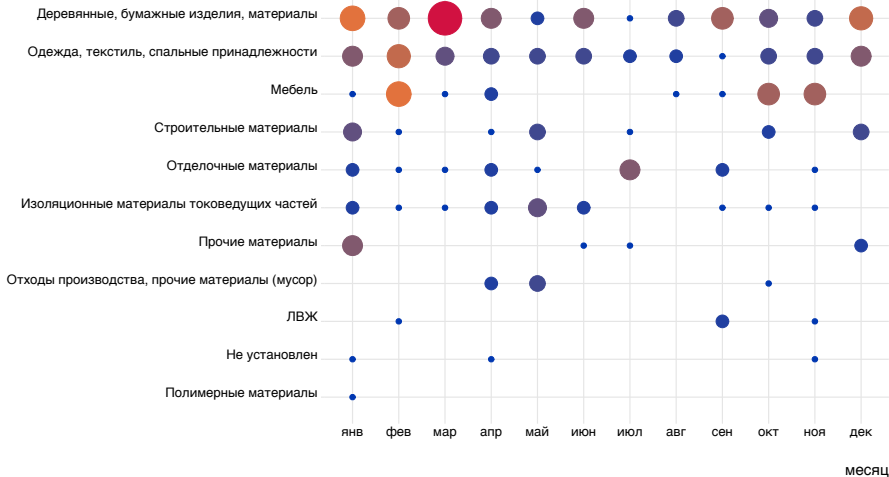
Преобразованные данные

	Дата	Количество погибших	Деревянные, бумажные изделия, материалы	Строительные материалы	Отделочные материалы	Одежда, текстиль, спальные принадлежности	Теплоизоляционный, гидроизоляционные материалы	Мебель
1	2017-01-01	1	2	0	1	2	0	2
2	2017-01-02	0	2	0	2	2	0	0
3	2017-01-03	0	2	1	0	0	0	0
4	2017-01-04	1	2	0	3	2	1	0
5	2017-01-05	0	1	1	0	1	0	1
6	2017-01-06	0	5	0	1	0	1	0
7	2017-01-07	1	4	1	4	1	0	1
8	2017-01-08	0	5	3	0	0	1	0
9	2017-01-09	0	4	1	2	0	0	0
10	2017-01-10	0	3	1	2	0	0	0
11	2017-01-11	2	3	3	8	0	0	2
12	2017-01-12	0	2	2	2	1	0	0
13	2017-01-13	1	5	2	2	0	0	0
14	2017-01-14	1	5	3	1	3	0	0
15	2017-01-15	0	4	2	1	1	1	0
16	2017-01-16	0	4	4	2	0	0	0

Количество погибших на пожарах в Красноярском крае в 2017 году в зависимости от материала горения

показан календарь событий по месяцам

материал горения

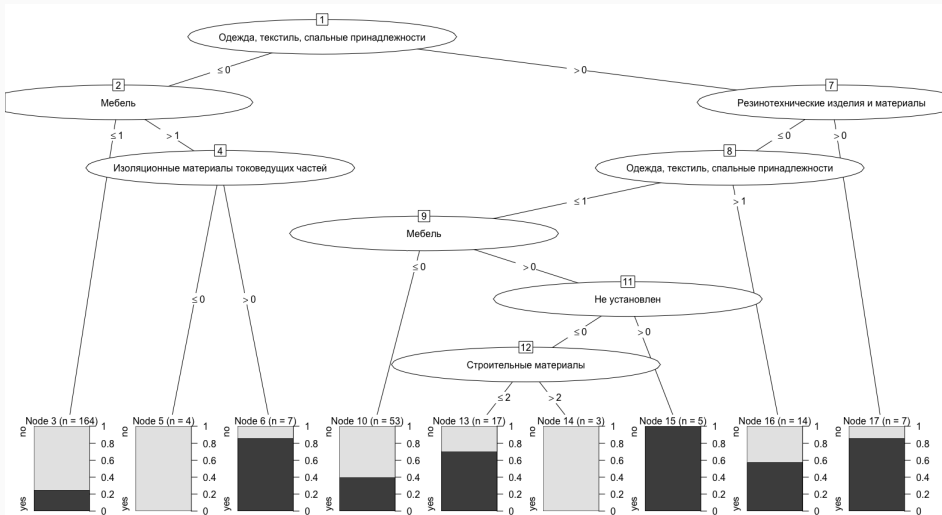


количество погибших

4 8 12 16

месяц

Модель («дерево решений»)



Возможности использования компьютерного анализа данных

- Иллюстрация современных статистических методов анализа данных и машинного обучения в учебном процессе
- Использование в социологических исследованиях
- Обработка экспериментальных данных
- Исследование статистики пожаров
- Оценка рисков, анализ выживаемости
- Работа с геоданными
- ...

Спасибо за внимание!