

Визуализация данных

Е.Н. Матеров

7 февраля 2019

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

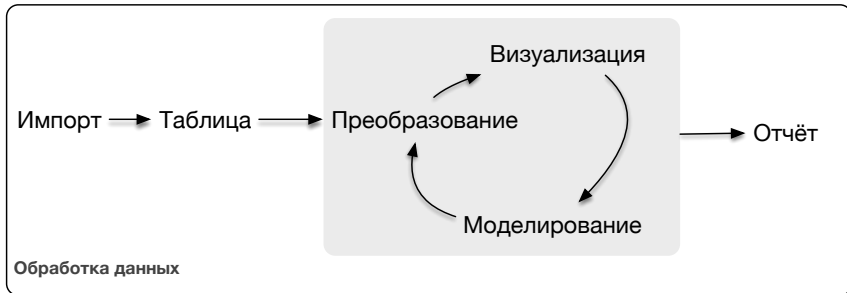


е) R

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**

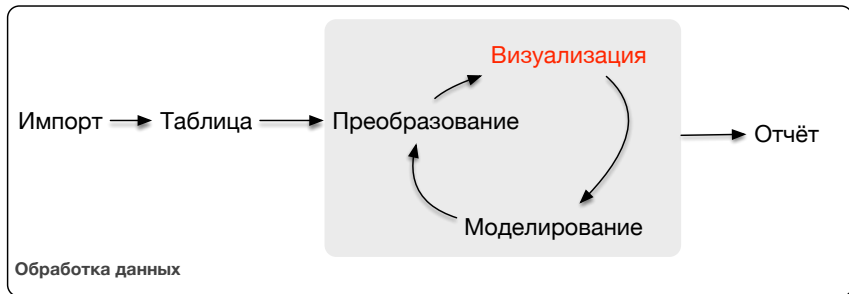
Учебное пособие

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)

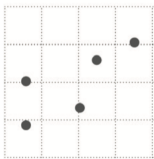
Характеристика восприятия (в порядке убывания)

1. Положение относительно общего масштаба
2. Положение относительно несимметричной шкалы
3. Длина, направление, угол
4. Площадь
5. Объем, кривизна
6. Затенение, насыщение цветом



William S. Cleveland, Robert McGill

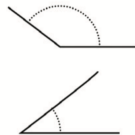
Graphical Preception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 1984. Vol. 79, No. 387, P. 531–534.



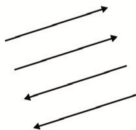
а) положение



б) длина



в) угол



г) направление



д) форма



е) площадь

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 типов данных

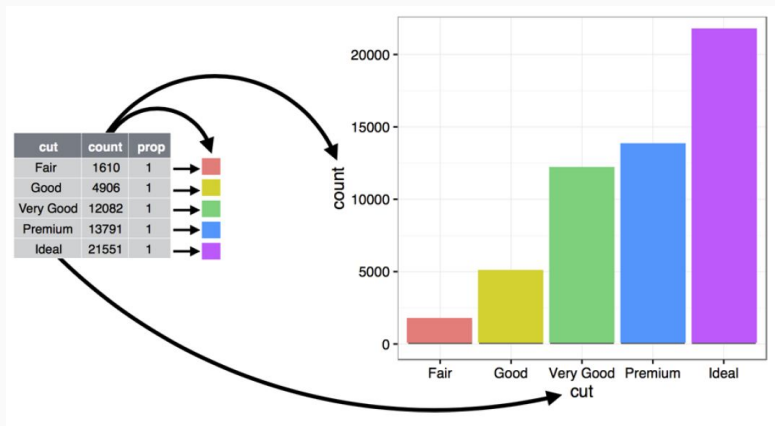
1. **Количественные** (дискретные и непрерывные) данные
2. **Категориальные** (номинальные и порядковые) данные
3. **Хронологические** данные
4. **Текстовые** данные
5. **Пространственные** (географические) данные

Основные виды графиков

классификация 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

Визуализация данных в R как отображение





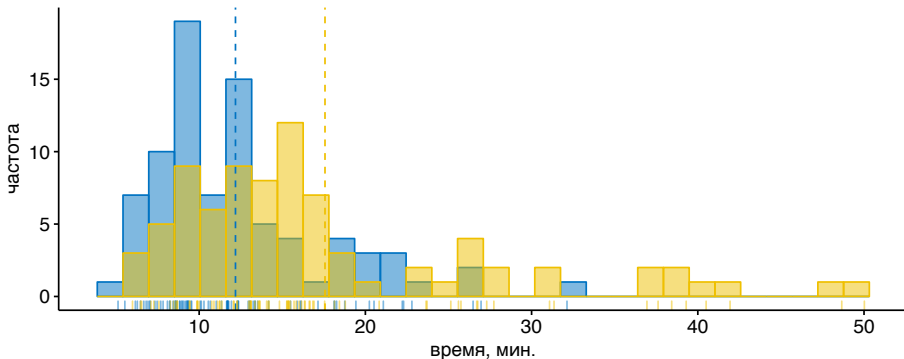
Виды графиков

Гистограмма

Среднее время ликвидации открытого горения, мин. на пожарах в РФ за 12 мес. в 2016 году

по видам населённых пунктов
пунктиром указаны средние значения

Вид населённого пункта:  город  село



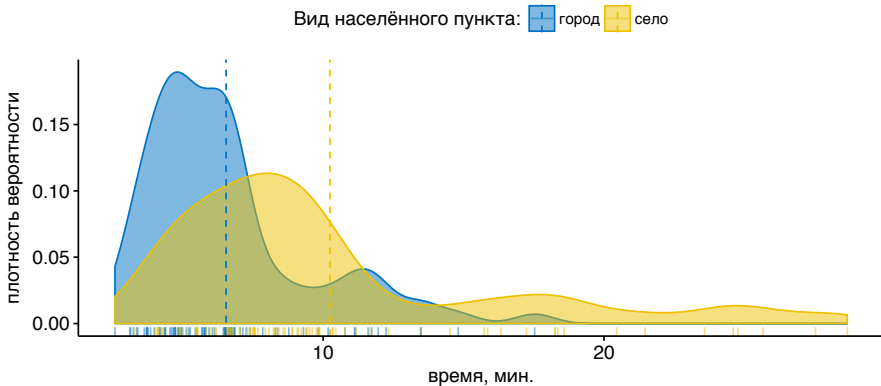
источник: ФГБУ ВНИИПО МЧС России

Виды графиков

График плотности

Среднее время ликвидации открытого горения, мин. на пожарах в РФ за 12 мес. в 2016 году

по видам населённых пунктов
пунктиром указаны средние значения

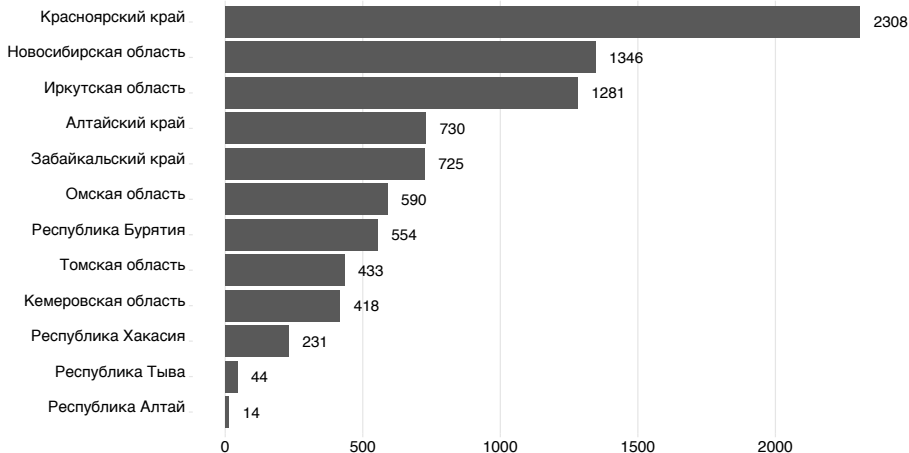


по данным ФГБУ ВНИИПО МЧС России

Виды графиков

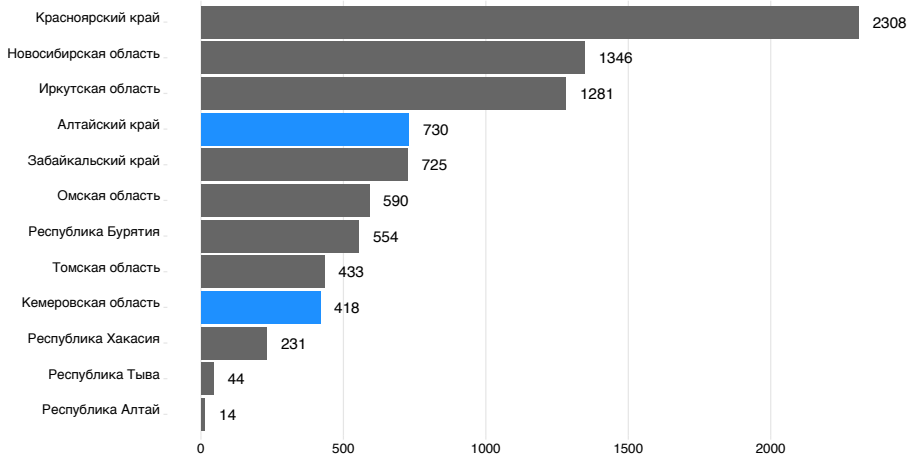
Столбиковая диаграмма

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



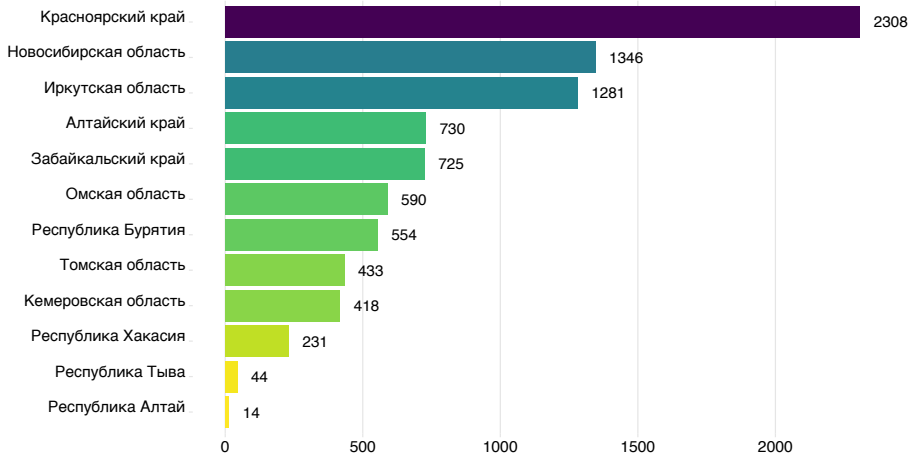
Количество спасённых, чел.
по данным ФГБУ ВНИИПО МЧС России

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



Количество спасённых, чел.
по данным ФГБУ ВНИИПО МЧС России

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



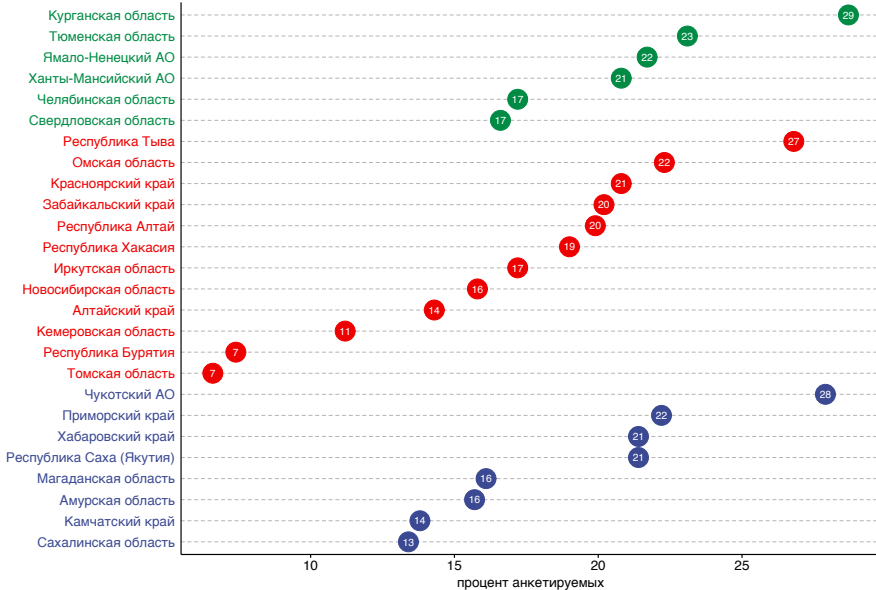
Количество спасённых, чел.
по данным ФГБУ ВНИИПО МЧС России

Виды графиков

Диаграмма Кливленда

Процент детей, которые ответили утвердительно на вопрос
Хотел бы ты стать пожарным или спасателем когда вырастешь?

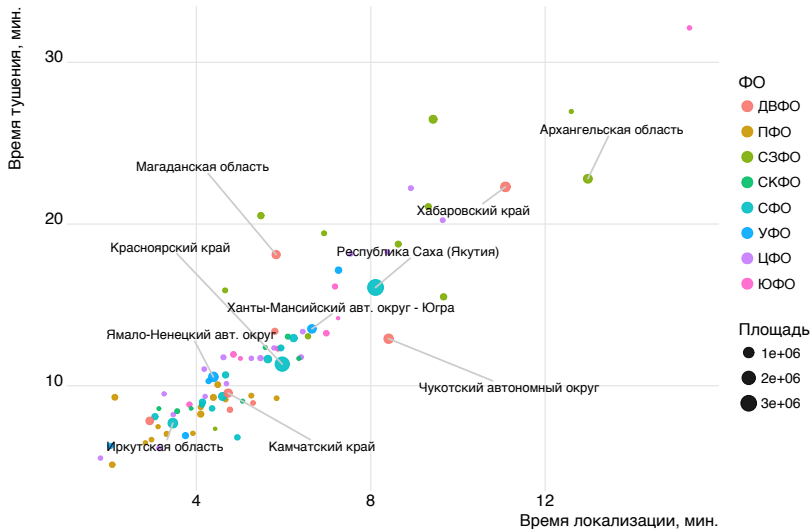
Данные сгруппированы по трём федеральным округам



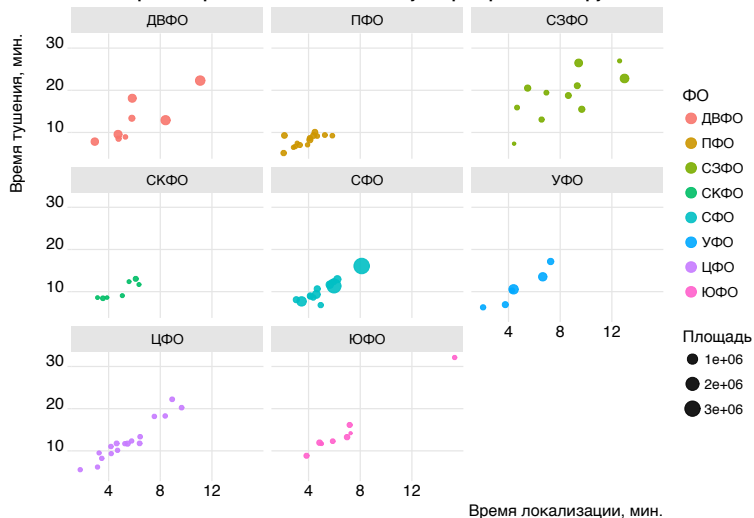
Виды графиков

Диаграмма рассеяния

Зависимость среднего времени тушения от среднего времени локализации, мин.
по пожарам в городе за 12 мес. в 2016 году по федеральным округам



**Зависимость среднего времени тушения от среднего времени локализации, мин.
по пожарам в городе за 12 мес. в 2016 году по федеральным округам**

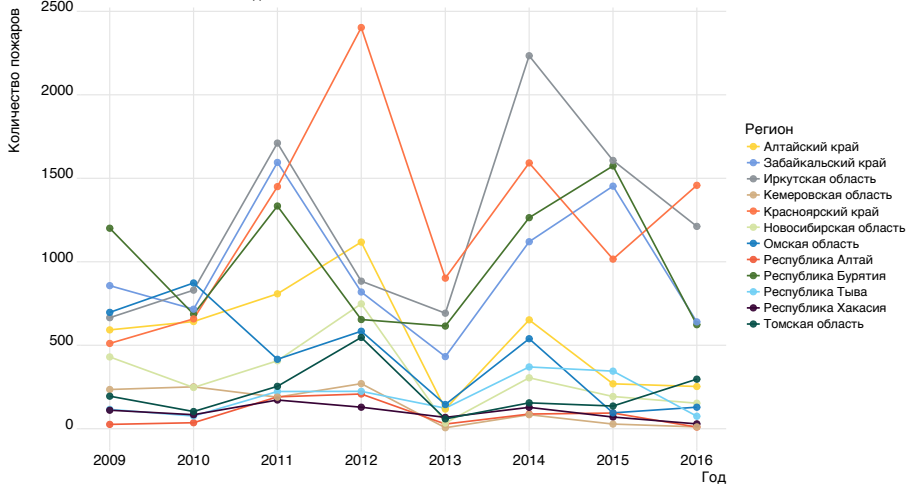


Виды графиков

Линейный график

Количество лесных пожаров в Сибирском федеральном округе

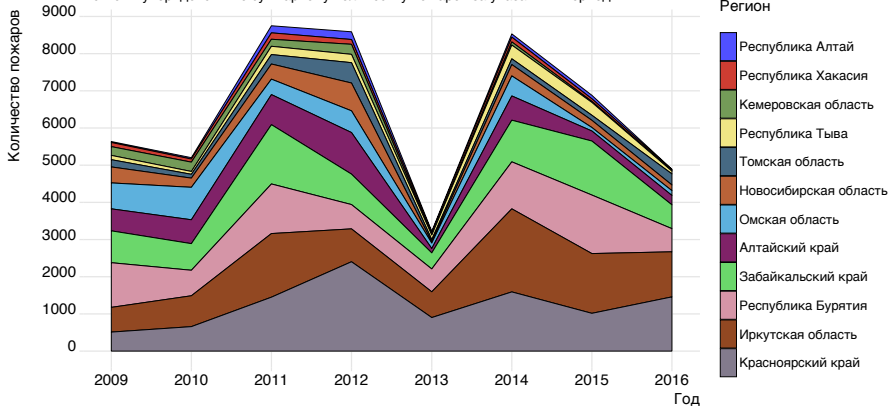
Статистика за 2009-2016 года



Источник: wiki-fire.org

Количество лесных пожаров в Сибирском федеральном округе в 2009-2016 годах

Регионы упорядочены по суммарному количеству пожаров за указанный период

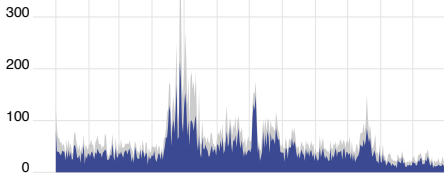


по данным wiki-fire.org

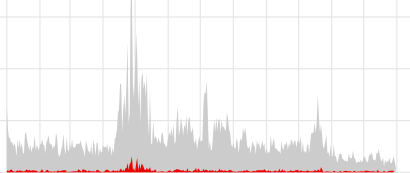
Количество пожаров (возгораний) зарегистрированных в сутки в Красноярском крае в 2016 году
 разбиение сделано по видам населённых пунктов

Количество пожаров

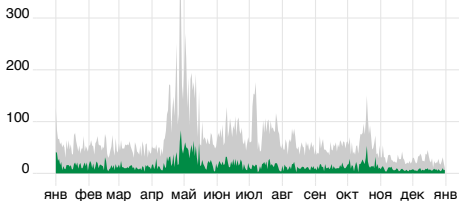
Город



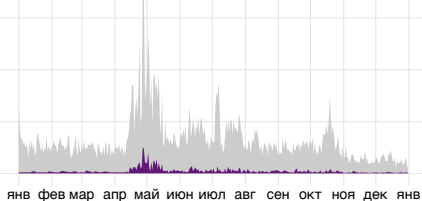
Населенный пункт городского типа



Сельский населенный пункт



Вне территории населенного пункта



янв фев мар апр май июн июл авг сен окт ноя дек янв

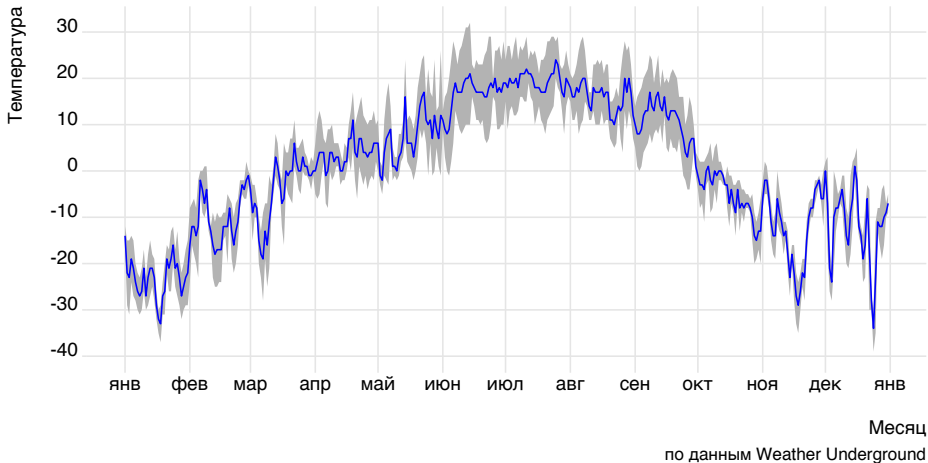
янв фев мар апр май июн июл авг сен окт ноя дек янв

Месяц

источник данных: ГУ МЧС России по Красноярскому краю

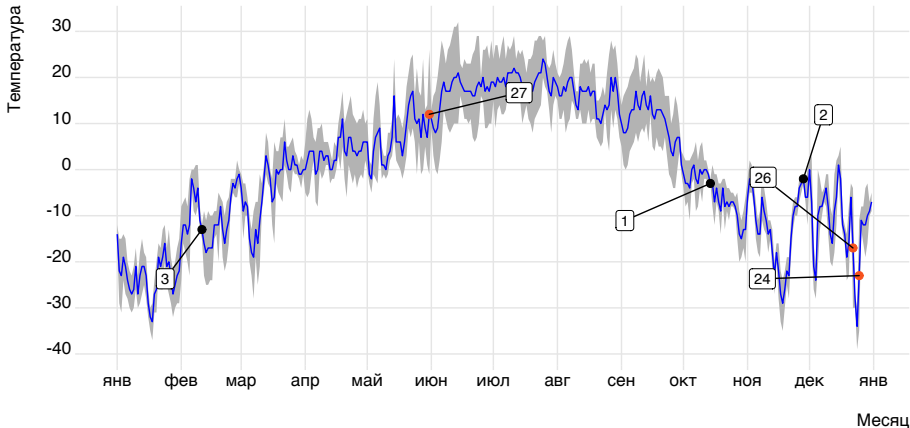
Динамика средней годовой температуры воздуха в Красноярске в 2016 году

Указаны: разброс минимальных и максимальных температур и средняя температура



Распределение средней годовой температуры в Красноярске в 2016 году

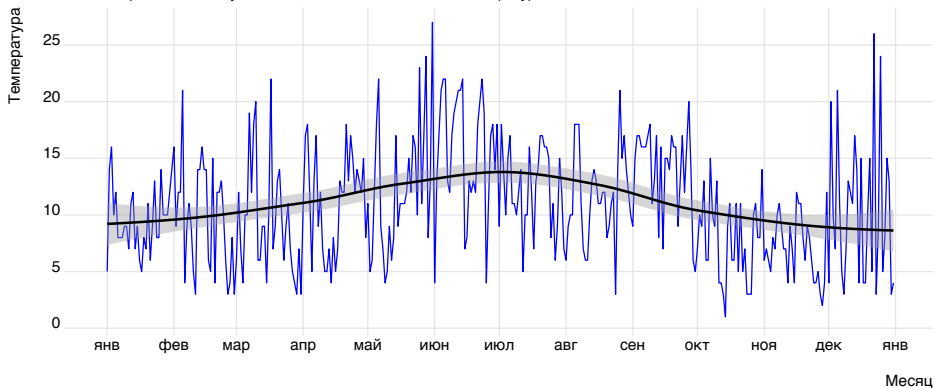
Указан разброс минимальных и максимальных температур



источник: Weather Underground

Распределение средней годовой температуры в Красноярске в 2016 году

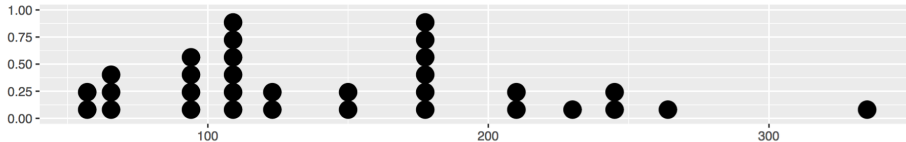
Указаны разности между максимальными минимальными температурами



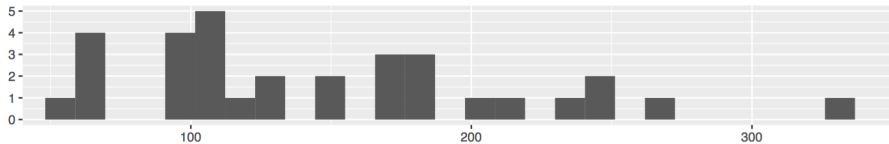
источник: Weather Underground

Виды графиков

Диаграмма размаха



Исходные данные



Гистограмма

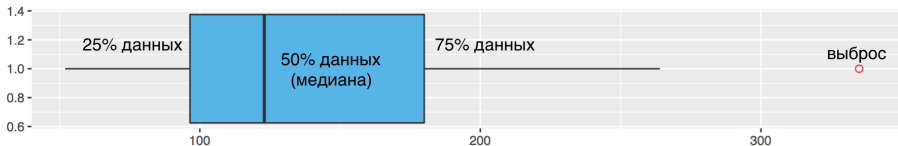
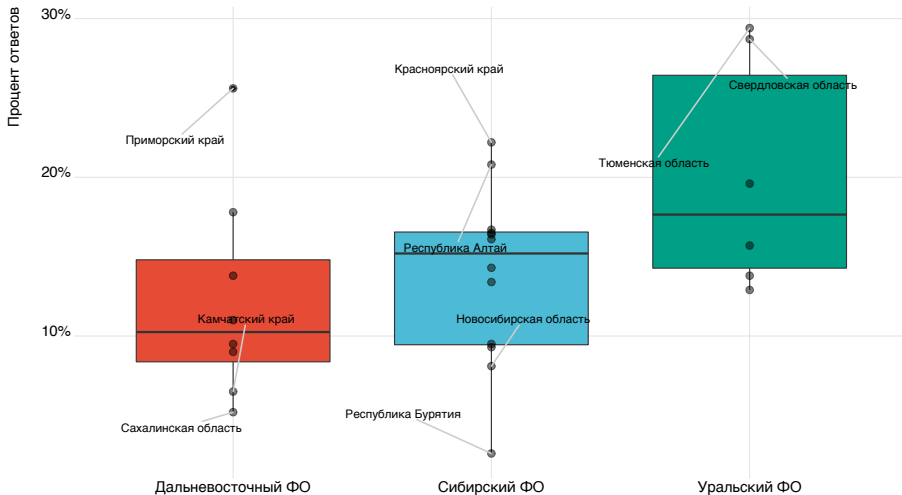


Диаграмма размаха (боксплот)

Процент детей, которые ответили утвердительно на вопрос Есть ли у вас в школе дружина юных пожарных?

Данные сгруппированы по трём федеральным округам

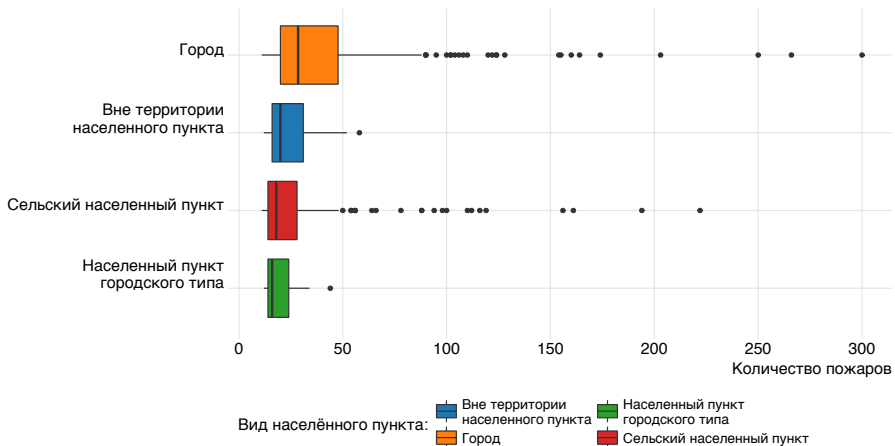


Федеральный округ

по данным социологического опроса Безопасность жизнедеятельности 2017

Количество пожаров (возгораний) регистрируемых в сутки в Красноярском крае в 2016 году

разбиение сделано по видам населённых пунктов



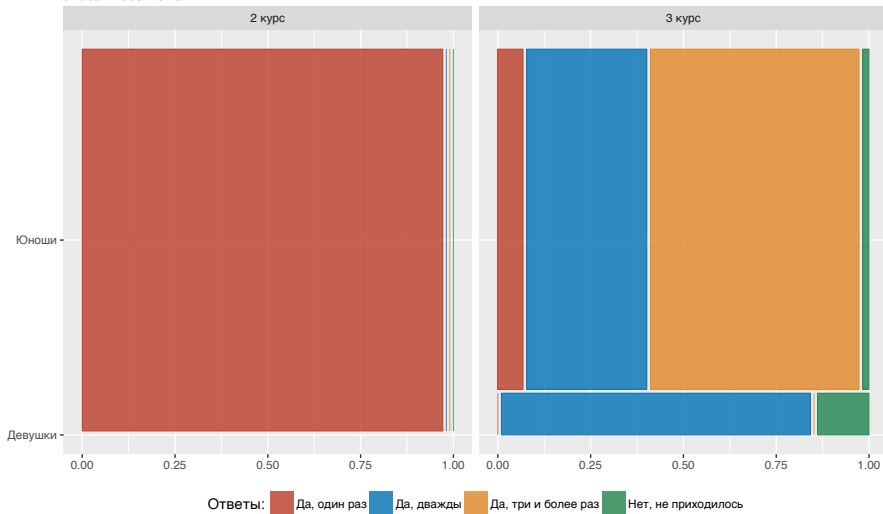
источник: ГУ МЧС России по Красноярскому краю

Виды графиков

Мозаичная диаграмма

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

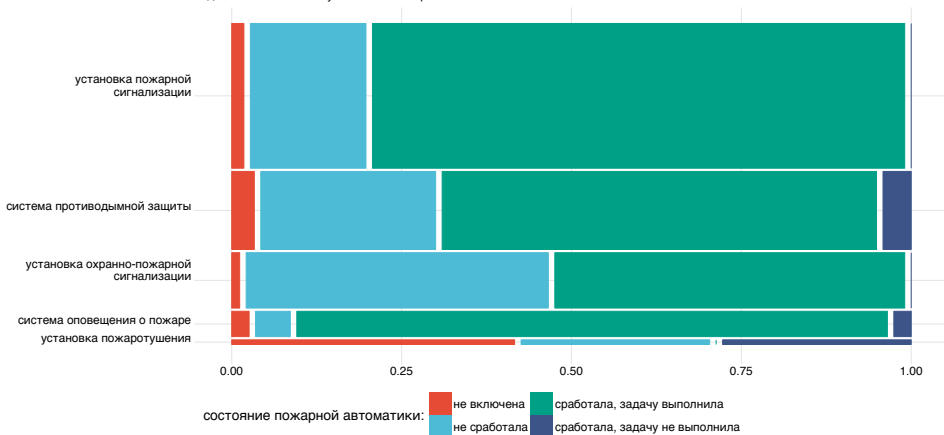
Указаны соотношения



По результатам опроса 2017 года

Эффективность работы пожарной автоматики в РФ в 2015 году при пожарах в зданиях жилого назначения

в долях от количества участия на пожарах

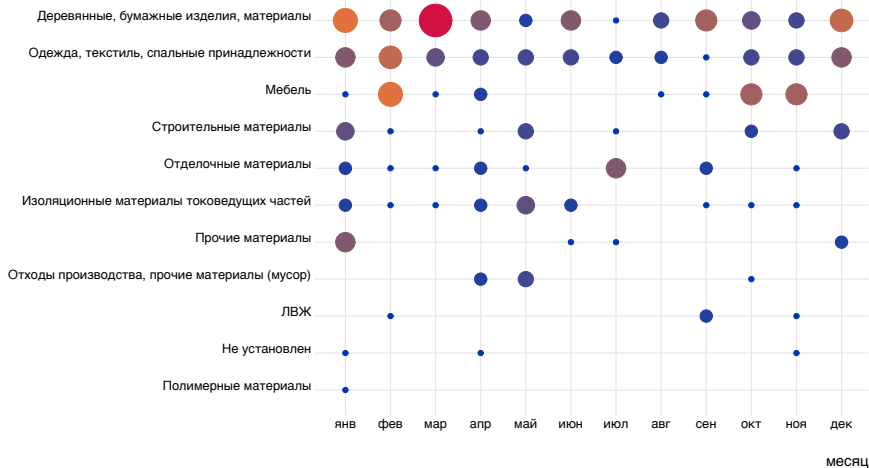


по данным ФГБУ ВНИИПО МЧС России

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

показан календарь событий по месяцам

материал горения



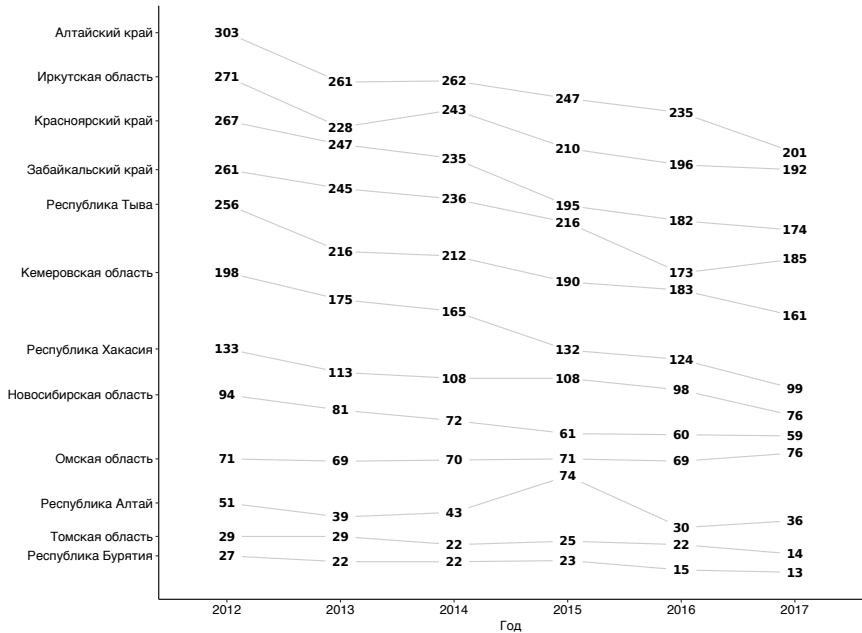
количество
погибших



Представление одних и тех же данных различными способами

Количество погибших на пожарах в Сибирском федеральном округе, чел.

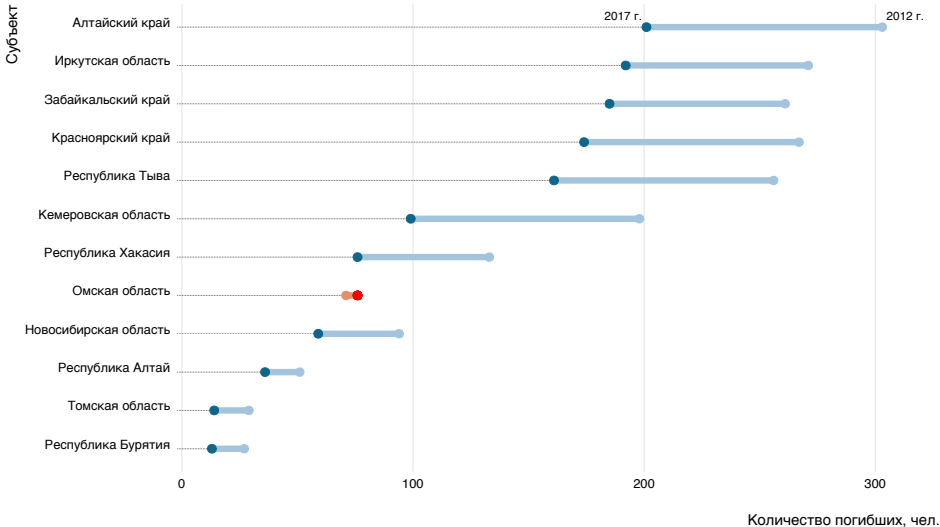
Динамика с 2012 г. по 2017 г.



по данным ФГБУ ВНИИПО МЧС России

Количество погибших на пожарах в Сибирском федеральном округе, чел.

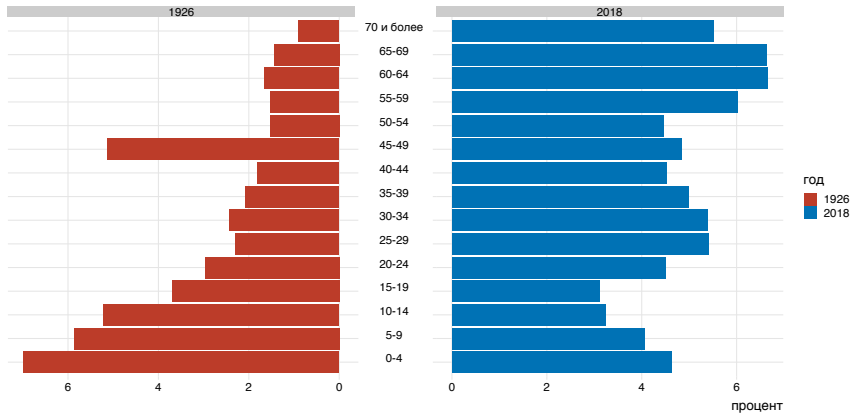
Динамика с 2012 г. по 2017 г.



по данным ФГБУ ВНИИПО МЧС России

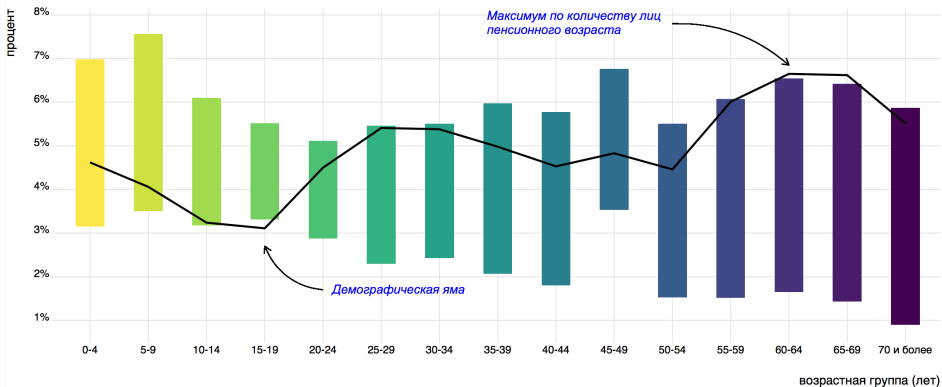
Распределение населения по возрастным группам в России в 1926 и 2018 году
по данным Федеральной службы государственной статистики

возрастная группа



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

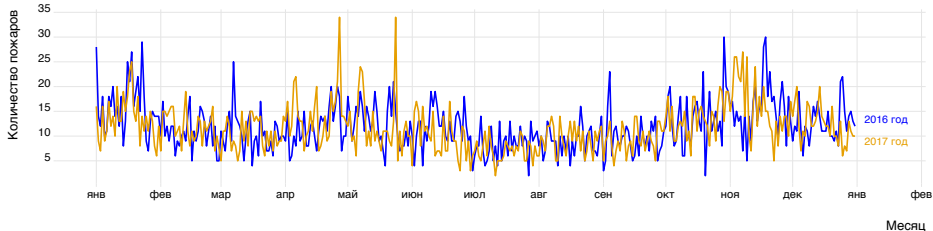


Специальные виды графиков

Тепловые диаграммы

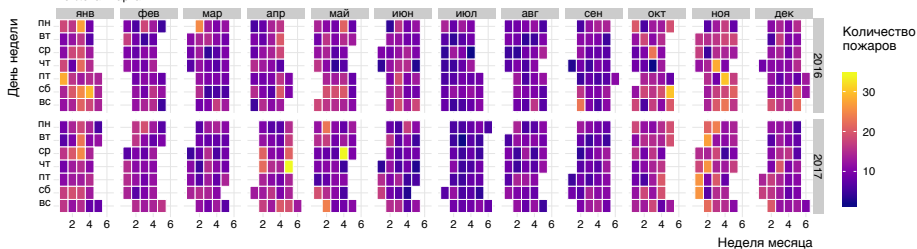
Количество пожаров в Красноярском крае в 2016 и 2017 годах

синий цвет - 2016 год, оранжевый - 2017 год



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

Тепловая карта



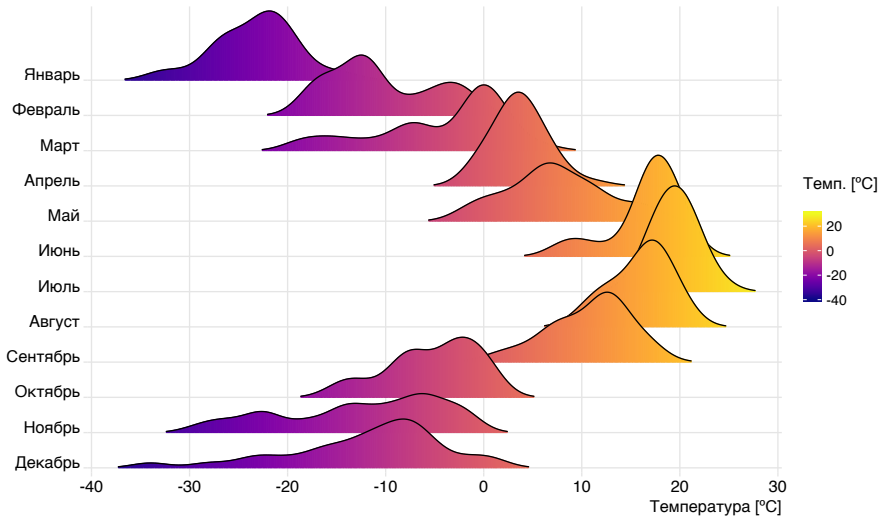
по данным ГУ МЧС России по Красноярскому краю

Специальные виды графиков

Риджлайны

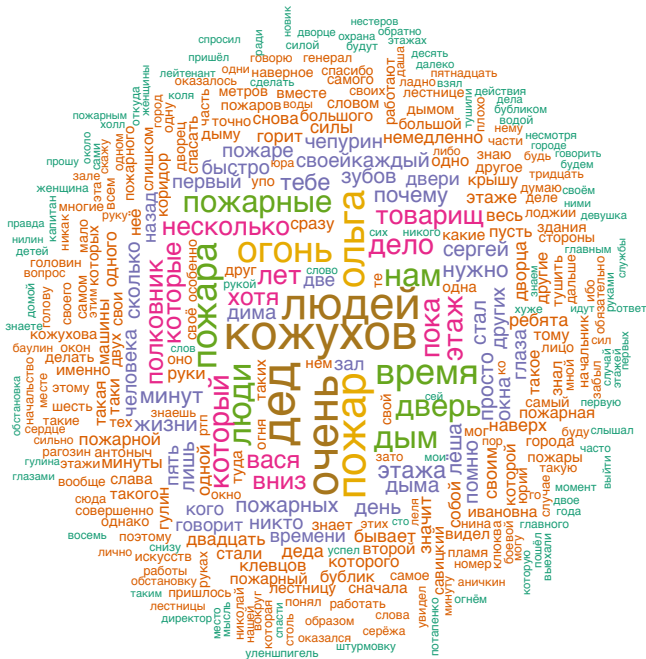
Распределение годовой температуры воздуха в Красноярске в 2016 году

Средняя температура (в градусах Цельсия) по месяцам в 2016 году
по данным сайта Weather Underground



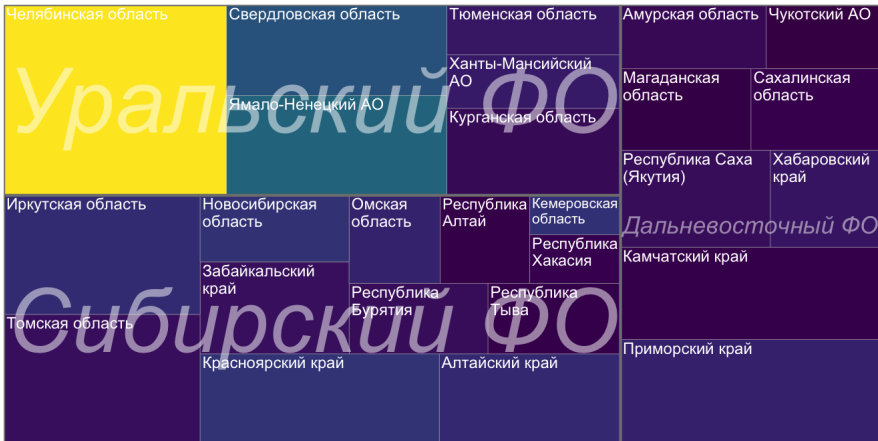
Специальные виды графиков

Облака слов



Специальные виды графиков

Картограмма



Спасибо за внимание!