

Визуальное представление количественной информации

Мария Голикова, BI-разработчик

Визуализация данных – это наглядное представление массивов информации

Цель визуализации – коммуникация

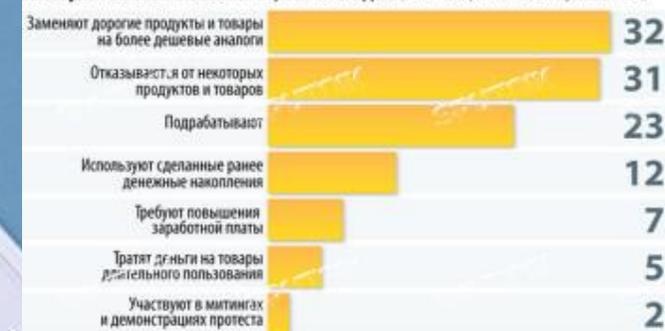
Инфографика, годовые отчеты, статьи



Как бороться с высокими ценами

Всероссийский центр изучения общественного мнения (ВЦИОМ) представил данные о том, что предпринимают россияне в условиях высоких цен на продукты, товары и услуги

Как противостоят высоким ценам сегодня (% от общего числа опрошенных)



Что планируют предпринимать в будущем (% от общего числа опрошенных)



Инициативный всероссийский опрос ВЦИОМ проведен 7-8 февраля 2009 г. Опрошено 1600 человек в 140 населенных пунктах в 42 областях, краях и республиках России. Статистическая погрешность не превышает 3,4%

Зачем представлять информацию
визуально?

Сколько здесь троек?

16247320125062

95830248639480

84459014811235

05379231602886

Сколько здесь троек?

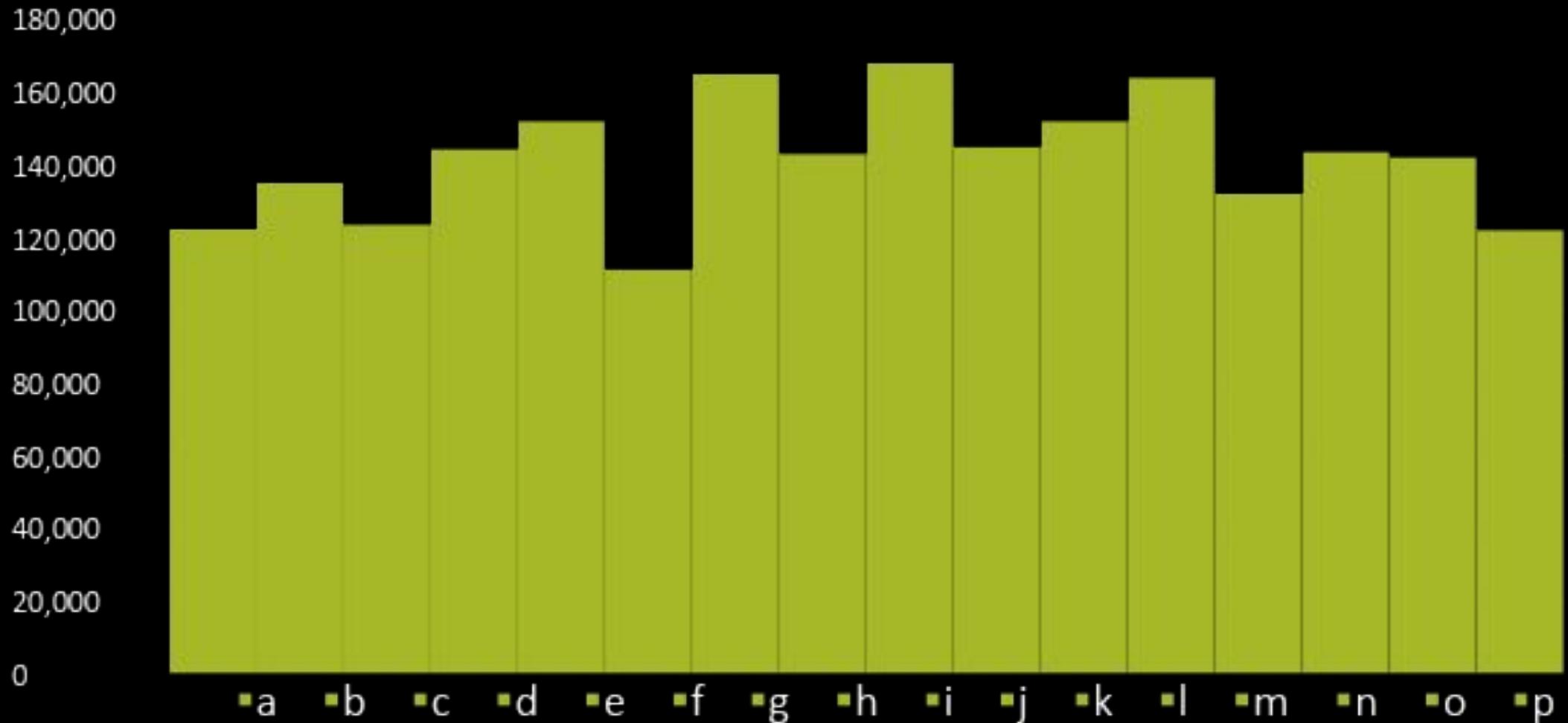
16247**3**20125062
958**3**02486**3**9480
844590148112**3**5
05**3**792**3**1602886

Какое число минимальное?

a	b	c	d	e	f	g	h
122 457	135 266	123 856	144 586	152 335	111 354	165 235	143 253

i	j	k	l	m	n	o	p
168 253	145 236	152 324	164 253	132 254	143 879	142 356	122 386

Какое число минимальное?



Немного фактов

Визуальные образы воспринимаются в 60 000 раз быстрее слов. Исследования Массачусетского Технологического Института (MIT) 2014 года показали, что человек может понять, что изображено на фотографии, посмотрев на нее всего 13 миллисекунд.



Немного фактов

Визуальные образы воспринимаются в 60 000 раз быстрее слов. Исследования Массачусетского Технологического Института (MIT) 2014 года показали, что человек может понять, что изображено на фотографии, посмотрев на нее всего 13 миллисекунд.

Визуальные образы держатся в памяти намного дольше, вызывают больший эмоциональный отклик, и самое главное - очень ёмкие

Немного фактов

Визуальные образы воспринимаются в 60 000 раз быстрее слов. Исследования Массачусетского Технологического Института (MIT) 2014 года показали, что человек может понять, что изображено на фотографии, посмотрев на нее всего 13 миллисекунд.

Визуальные образы держатся в памяти намного дольше, вызывают больший эмоциональный отклик, и самое главное - очень ёмкие

На интересном и заслуживающем внимания образе мозг способен концентрироваться сколь угодно долго

Немного фактов

Визуальные образы воспринимаются в 60 000 раз быстрее слов. Исследования Массачусетского Технологического Института (MIT) 2014 года показали, что человек может понять, что изображено на фотографии, посмотрев на нее всего 13 миллисекунд.

Визуальные образы держатся в памяти намного дольше, вызывают больший эмоциональный отклик, и самое главное - очень ёмкие

На интересном и заслуживающем внимания образе мозг способен концентрироваться сколь угодно долго

Физиологически, восприятие визуальной информации является основной для человека: наш мозг для обработки визуальных образов задействует гораздо больше нейронов, чем для всех других видов сенсорного восприятия, вместе взятых

Визуализация – дело
тонкое...

Этапы построения визуализации

1. **Идея, смысл:** что хотим донести до аудитории? Эту идею используем в качестве заголовка диаграммы

Этапы построения визуализации

1. **Идея, смысл:** что хотим донести до аудитории? Эту идею используем в качестве заголовка диаграммы
2. **Определение соотношения:** как нужно выразить идею? 5 видов соотношений: покомпонентное, позиционное, временное, частотное, корреляционное

Этапы построения визуализации

1. **Идея, смысл:** что хотим донести до аудитории? Эту идею используем в качестве заголовка диаграммы
2. **Определение соотношения:** как нужно выразить идею? 5 видов соотношений: покомпонентное, позиционное, временное, частотное, корреляционное
3. **Выбор диаграммы:** выбираем из 50+ разных типов, но основных всего 5 😊

Виды соотношений (этап 2)

1. **Покомпонентное** – процент от целого

Виды соотношений (этап 2)

1. **Покомпонентное** – процент от целого
2. **Позиционное** – расположение объектов

Виды соотношений (этап 2)

1. **Покомпонентное** – процент от целого
2. **Позиционное** – расположение объектов
3. **Временное** – изменение во времени

Виды соотношений (этап 2)

1. **Покомпонентное** – процент от целого
2. **Позиционное** – расположение объектов
3. **Временное** – изменение во времени
4. **Частотное** – объекты в интервалах

Виды соотношений (этап 2)

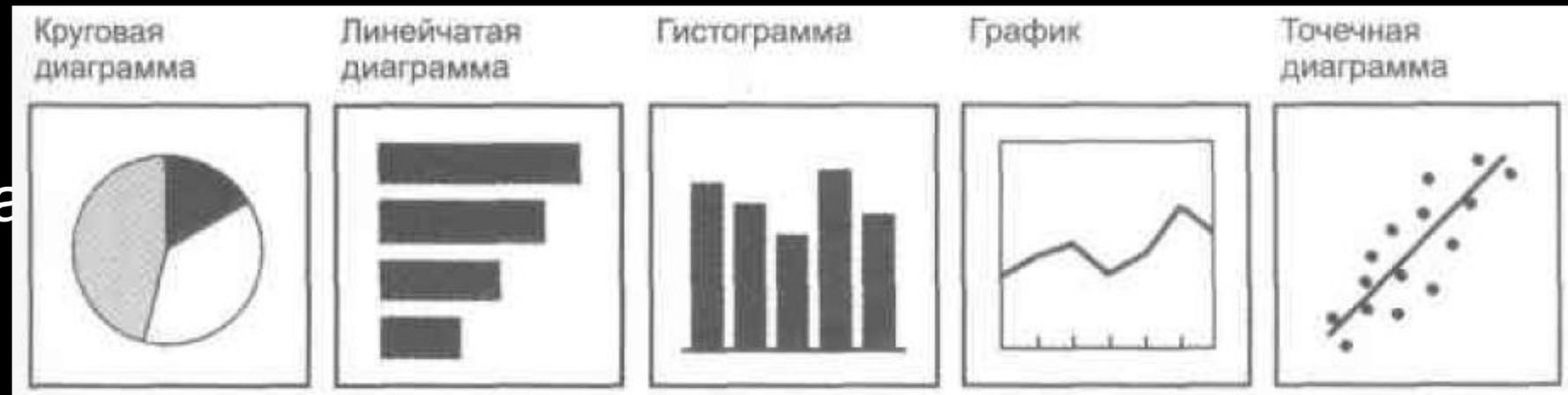
1. **Покомпонентное** – процент от целого
2. **Позиционное** – расположение объектов
3. **Временное** – изменение во времени
4. **Частотное** – объекты в интервалах
5. **Корреляционное** – зависимость между переменными

Виды соотношений (этап 2)

1. **Покомпонентное** – процент от целого
2. **Позиционное** – расположение объектов
3. **Временное** – изменение во времени
4. **Частотное** – объекты в интервалах
5. **Корреляционное** – зависимость между переменными
6. **Изменение на местности** – показатели на карте

Виды диаграмм (этап 3)

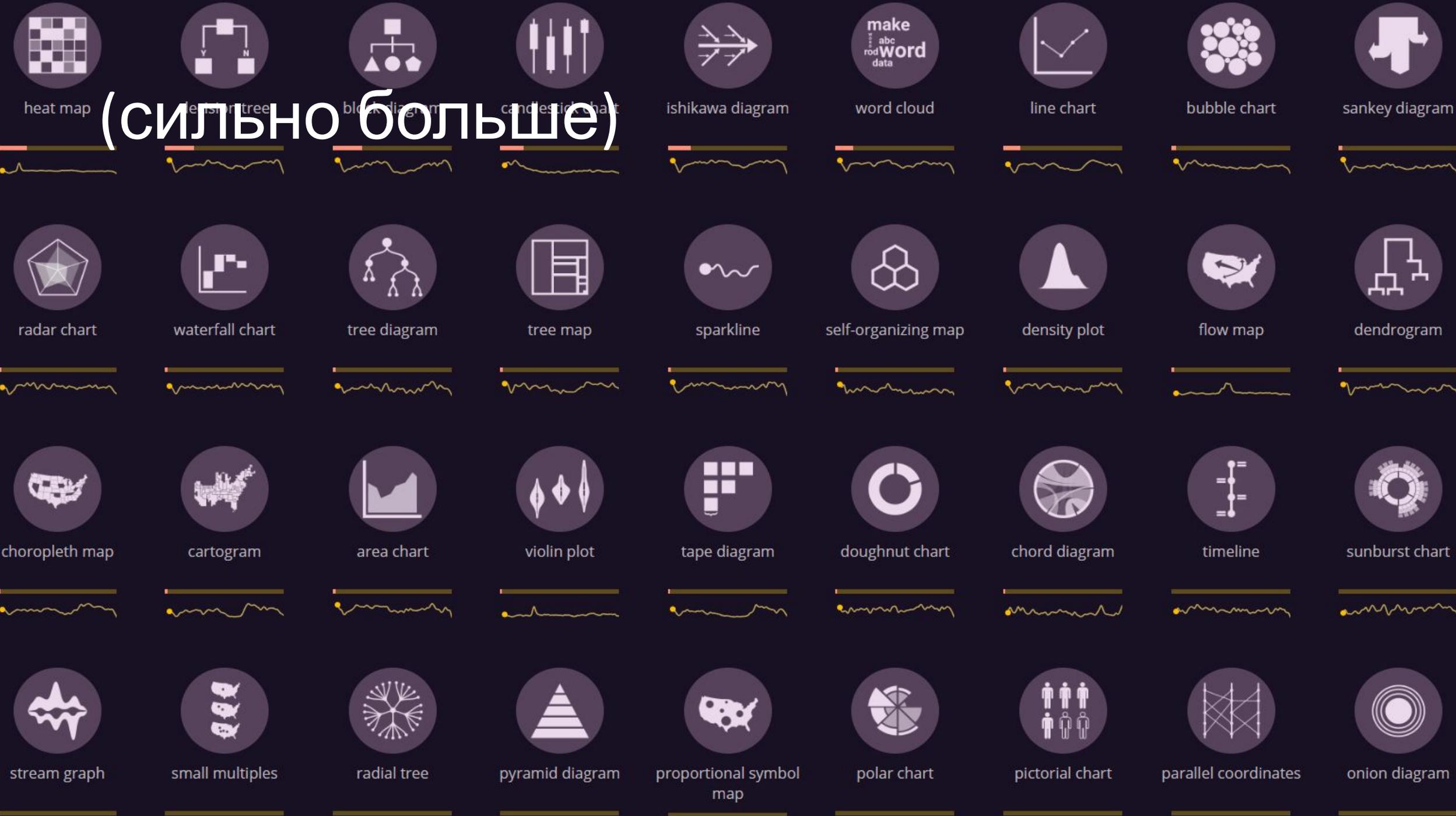
1. Круговая
2. Линейчатая
3. Гистограмма
4. График
5. Точечная
6. (Карта)



(на самом деле их больше)



(СИЛЬНО БОЛЬШЕ)



heat map

decision tree

block diagram

candlestick chart

ishikawa diagram

word cloud

line chart

bubble chart

sankey diagram

radar chart

waterfall chart

tree diagram

tree map

sparkline

self-organizing map

density plot

flow map

dendrogram

choropleth map

cartogram

area chart

violin plot

tape diagram

doughnut chart

chord diagram

timeline

sunburst chart

stream graph

small multiples

radial tree

pyramid diagram

proportional symbol map

polar chart

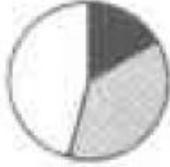
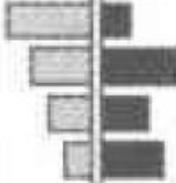
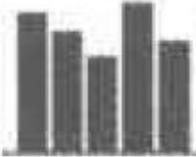
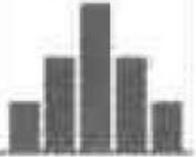
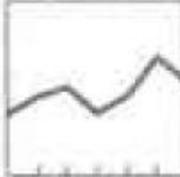
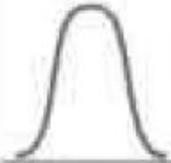
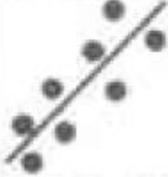
pictorial chart

parallel coordinates

onion diagram

ТИПЫ СРАВНЕНИЯ

ОСНОВНЫЕ ТИПЫ ДИАГРАММ

	ПОКОМПОНЕНТНОЕ	ПОЗИЦИОННОЕ	ВРЕМЕННОЕ	ЧАСТОТНОЕ	КОРРЕЛЯЦИОННОЕ
КРУГОВАЯ					
ЛИНЕЙЧАТАЯ					
ГИСТОГРАММА					
ГРАФИК					
ТОЧЕЧНАЯ					

Change v Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader

Examples of use

Share price movements, economic time series

Chart types

line



The standard way to show a changing time series. If data are irregular, consider markers to represent data points

column-timeline



Columns work well for showing change over time - but usually best with only one series of data at a time

column-line-timeline



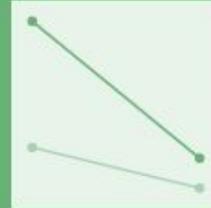
A good way of showing the relationship over time between an amount (columns) and a rate (line)

stock-price



Usually focused on day-to-day activity, these charts show opening/closing and hi/low points of each day

slope



Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story

area



Use with care. These are good at showing changes to total, but seeing change in components can be very difficult.

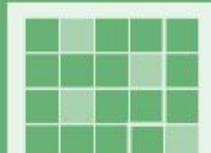
fan



scatterplot-line-timeline



calendar-heatmap



priestley-timeline



circles-timeline



seismogram



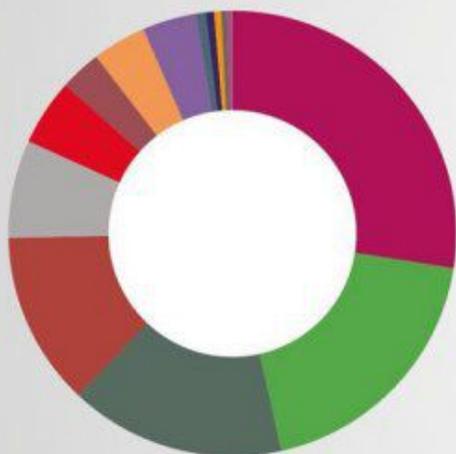
Покомпонентное сравнение

- Показать долю от целого удобно с помощью круговой диаграммы
- Используем не более 6 компонентов (если больше – ненужных делаем категорию Прочее)
- Глаз двигается по часовой стрелке, поэтому самое важное на линии 12 часов
- Наименее практичная
- Зачастую лучше заменить гистограммой с накоплением, нормированной на 100%

Не очень удачный пример...

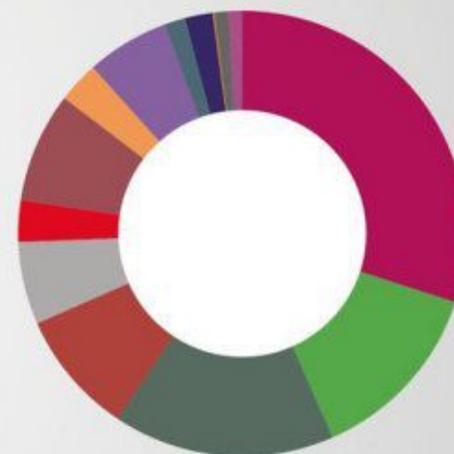
Структура федерального бюджета РФ

Как сейчас тратятся бюджетные деньги
в среднем за последние пять лет



27,5%	Пенсии и пособия	30%
19%	Армия и вооружение	13,4%
15,6%	Госинвестиции	16%
12,6%	Полиция и спецслужбы	9%
7,1%	Госаппарат	6%
4,8%	Дотации регионам	3%*
3%	Здравоохранение	8%
3,9%	Выплата долгов	3%
3,9%	Образование	6%
0,7%	ЖКХ	1,5%
0,6%	Культура	2%
0,5%	СМИ	0,1%
0,4%	Экология	1%
0,4%	Спорт (в т. ч. детский)	1%

Как должны тратиться
бюджетные деньги



* Основные средства в регионы должны прийти в виде собственных региональных доходов через изменение системы распределения налогов между центром, регионами и муниципалитетами.

Позиционное сравнение

- Лучше всего подходит линейчатая диаграмма или гистограмма
- Линейчатая удобна тем, что можно расположить категории по вертикали в нужной последовательности
- Контрастным цветом можно выделить идею
- Подписи категорий для гистограммы должны быть соответствующей ширины, иначе используем линейчатую

Неплохой пример

Соотношение налогов и расходов на благоустройство, млрд руб.

■ Имущественные налоги
■ Местные налоги
■ Расходы на благоустройство

За исключением Уфы все центральные города агломераций генерируют налоговые доходы от городской недвижимости, достаточные для покрытия расходов на благоустройство.

Однако такие расходы пока сохраняются на низком уровне



Временное сравнение

- Отражает динамику изменений
- Лучше всего подходят гистограмма или график
- Гистограмма строится всегда от нуля
- График можно строить не от нуля
- Нужно избегать «спагетти-графиков»

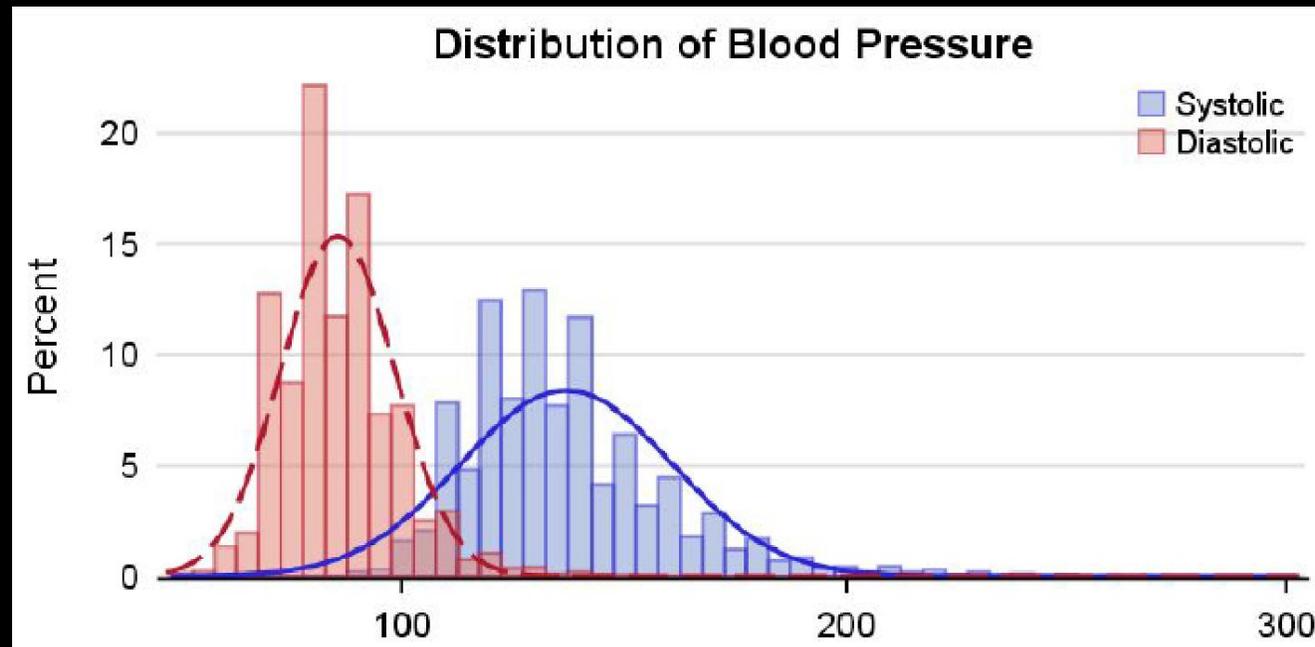
Послание президента, между прочим



Частотное сравнение

- Показывает, сколько значений данного параметра попадает в области числовых значений
- Обычно используется в аналитических или статистических целях
- Нужно подобрать такой размер диапазонов, чтобы он верно донес основную идею
- Интервалы лучше делать одинакового размера

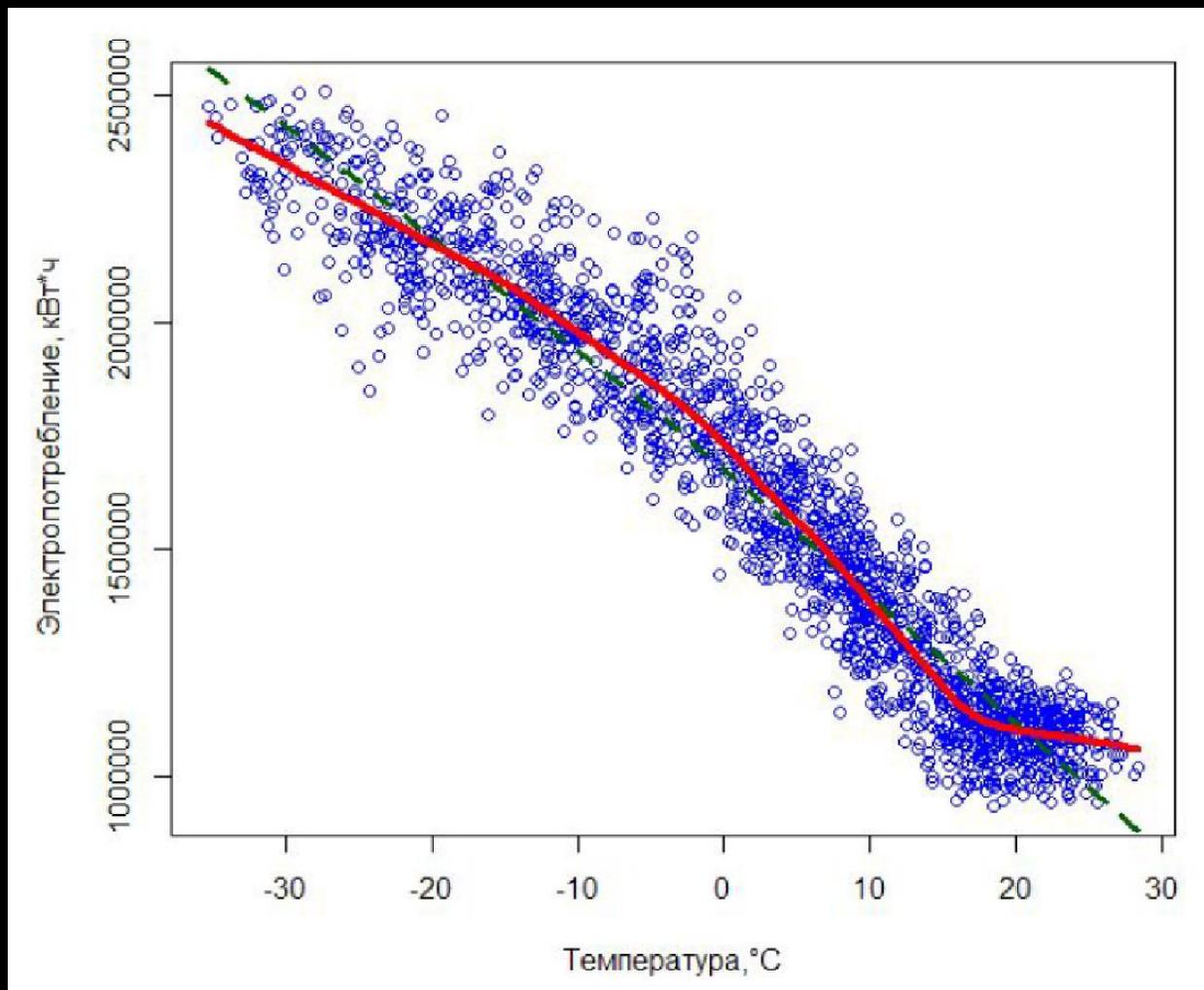
Вспоминаем статистику



Корреляционное сравнение

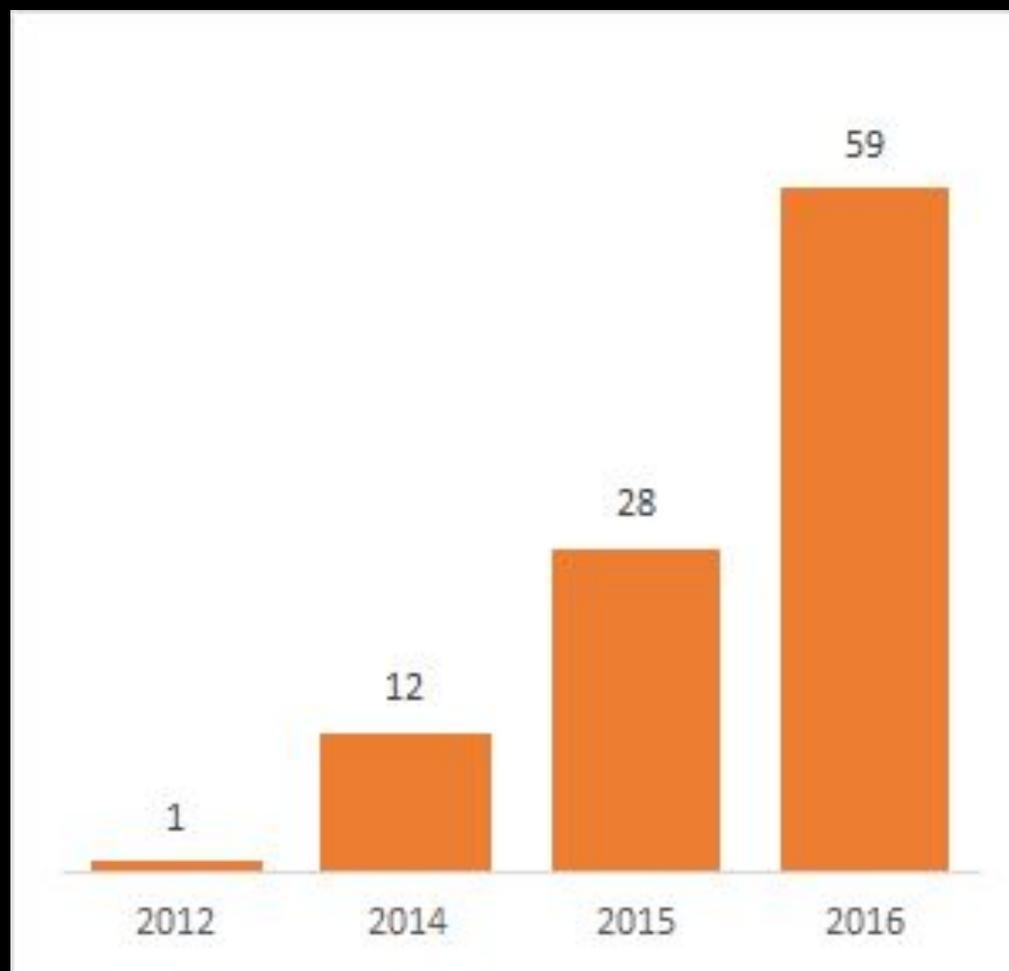
- Проверяет, соответствует ли соотношение двух величин ожидаемой зависимости
- Подходят точечная или двухсторонняя линейчатая диаграммы
- Пузырьковая содержит третье измерение – размер точки

Что за примеры у тебя, Маша???

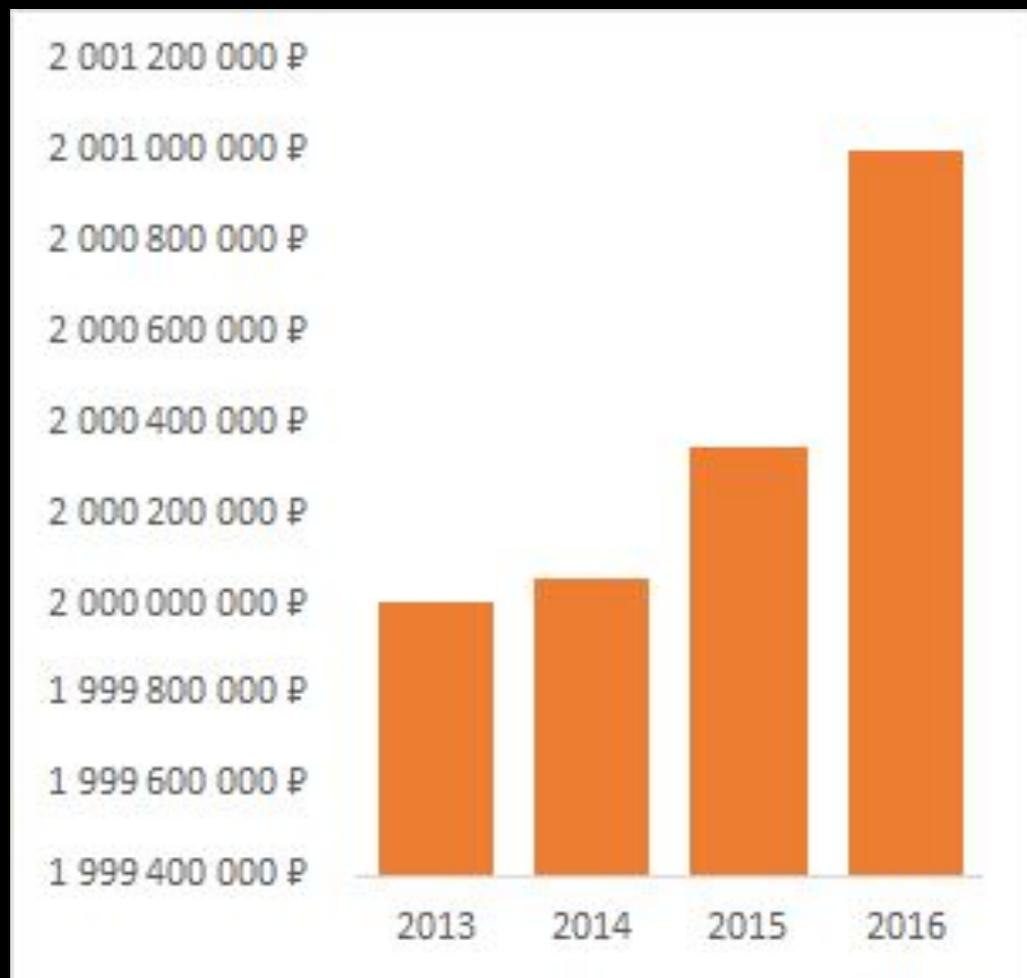


А теперь про ошибки

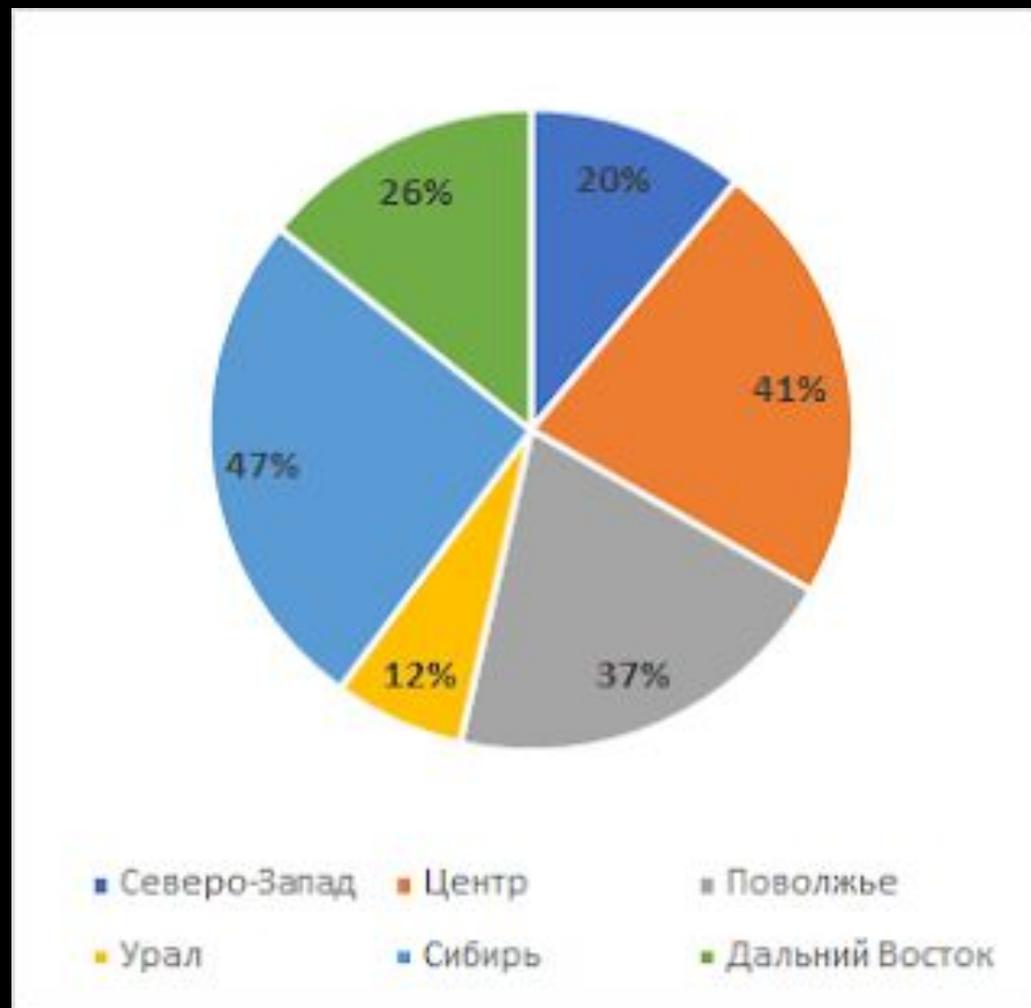
1. Проверьте данные



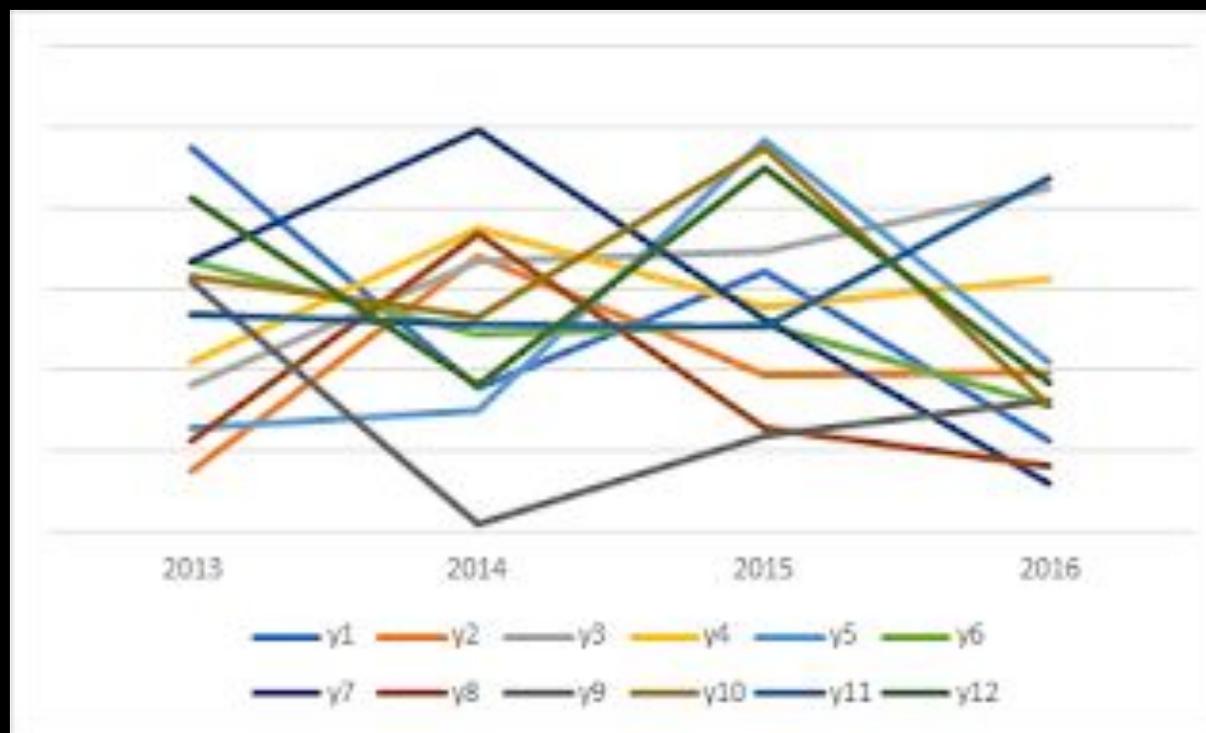
2. Правильно выбирайте начало оси



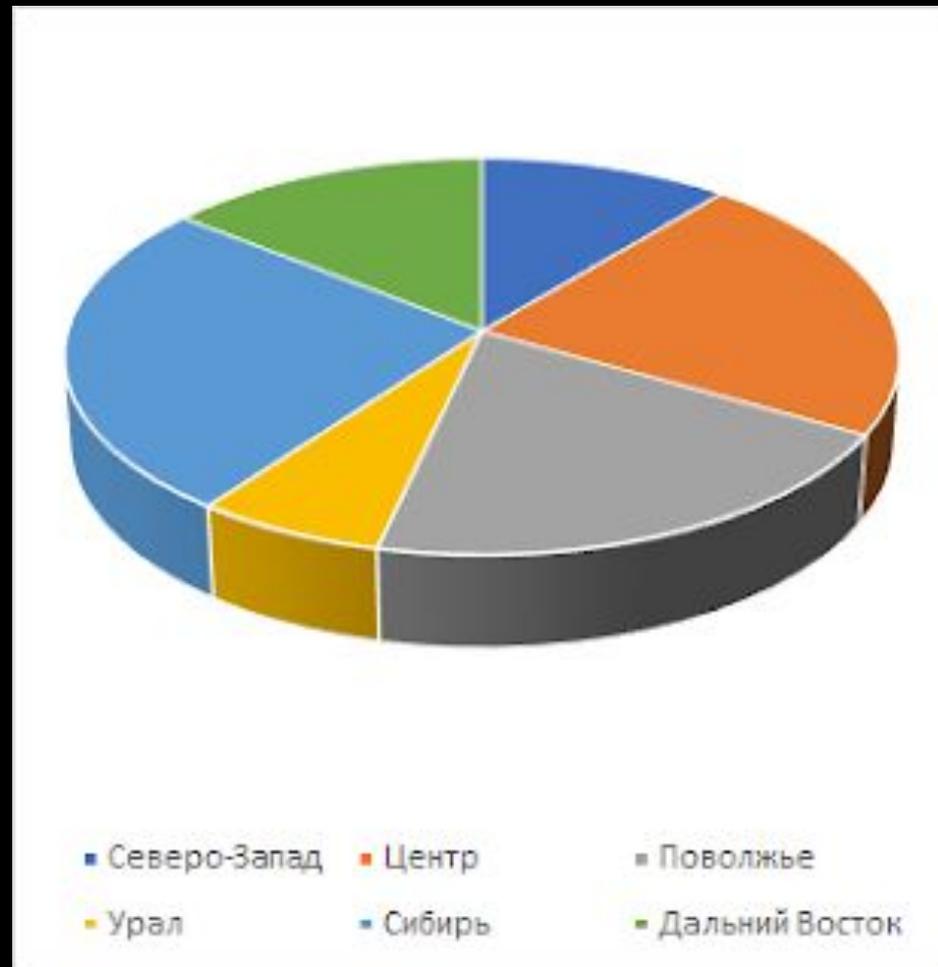
3. Сумма частей целого = 100%



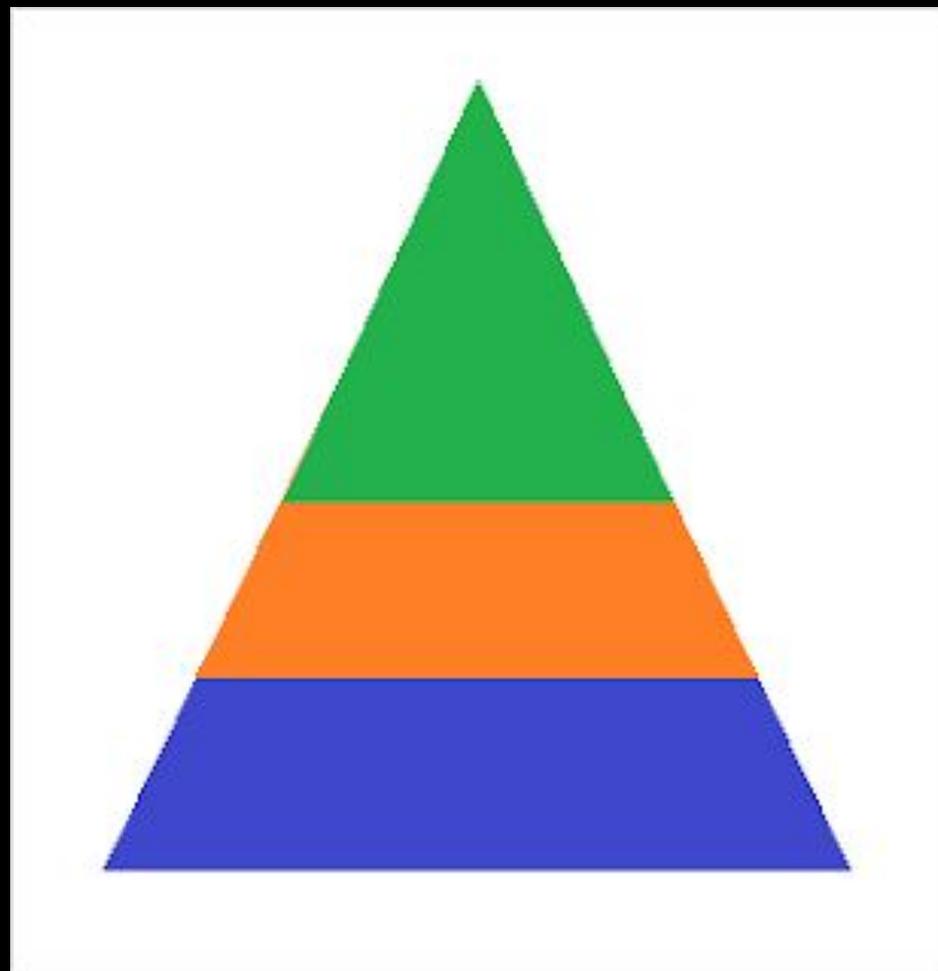
4. Избегайте визуального шума



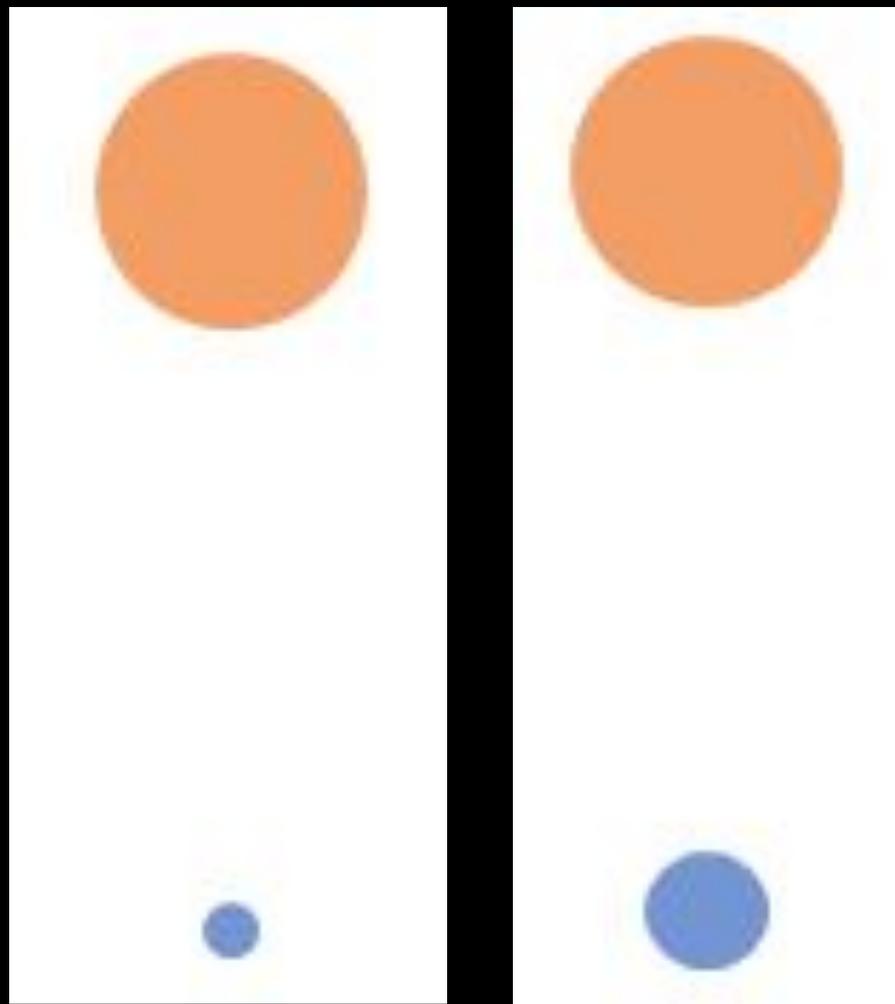
5. Забудьте про 3D...



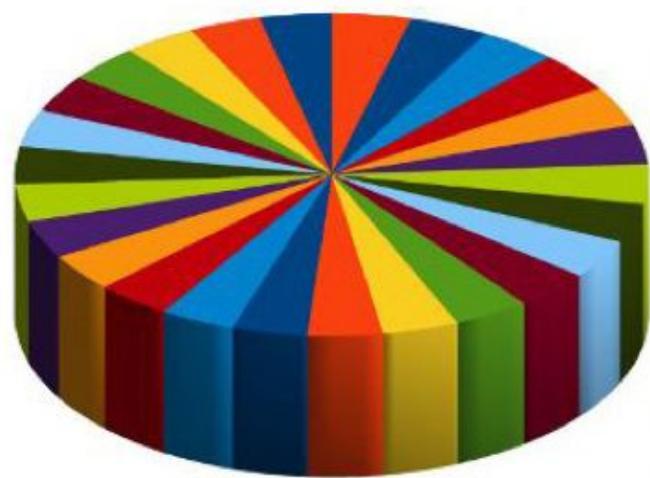
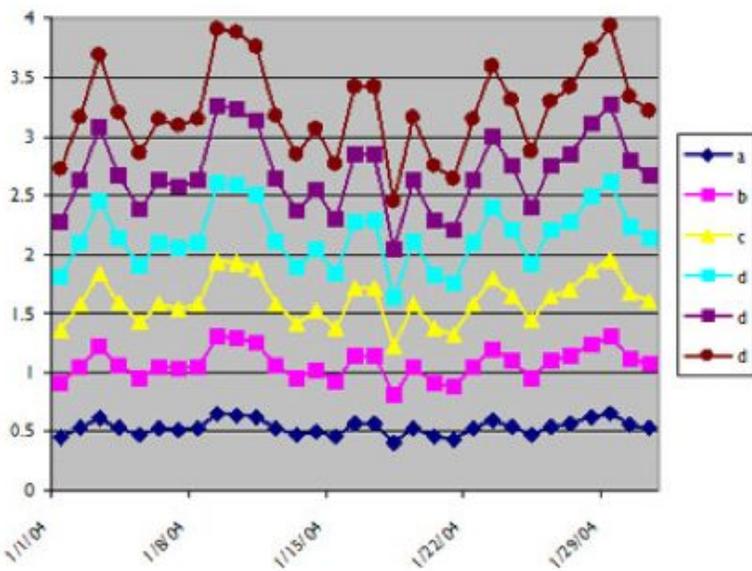
6. ...и нестандартные фигуры



7. Выразайте число через площадь фигуры

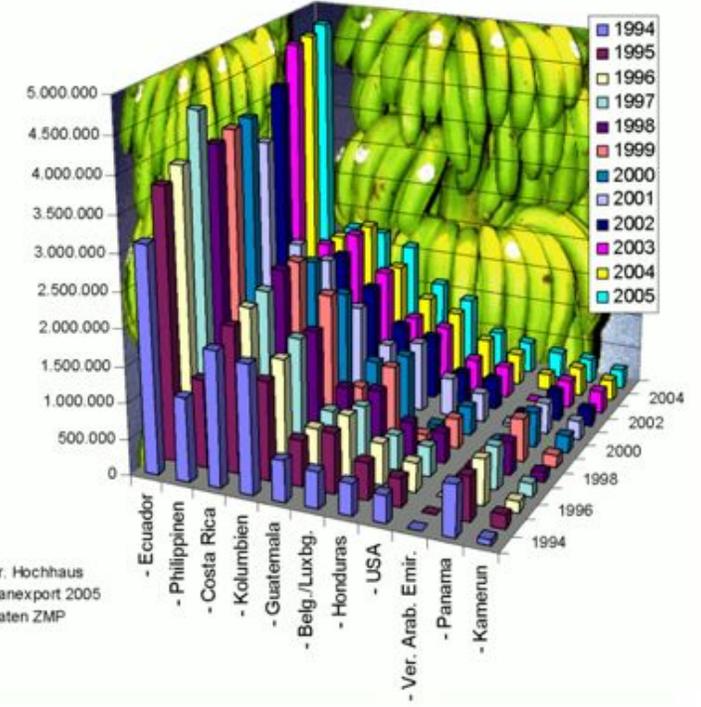


Советы



- | | |
|----|----|
| 1 | 14 |
| 2 | 15 |
| 3 | 16 |
| 4 | 17 |
| 5 | 18 |
| 6 | 19 |
| 7 | 20 |
| 8 | 21 |
| 9 | 22 |
| 10 | 23 |
| 11 | 24 |
| 12 | 25 |
| 13 | 26 |

Export von Bananen in Tonnen von 1994-2005



Dr. Hochhaus
Banexport 2005
Daten ZMP

Советы

1. Учитывайте аудиторию
2. Правильно выбирайте шкалу и масштаб
3. Делайте аннотации
4. Следите за читабельностью текста
5. Тщательно выбирайте палитру, учитывайте дальтонизм
6. Избегайте визуальный шум, 3D, стили по умолчанию
7. Пустое пространство не может быть больше, чем полезное
8. Перенесите информацию с осей и легенд на диаграмму, насколько это возможно

И еще раз про область диаграммы

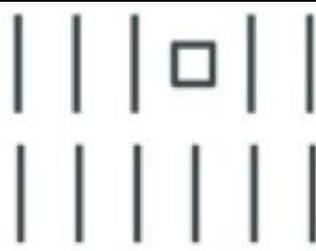
- ✓ Проверьте заголовок
- ✓ Проверьте названия осей
- ✓ Округлите цифры, дроби смотрятся плохо
- ✓ Отсортируйте информацию
- ✓ Сделайте сетку не бросающейся в глаза или уберите совсем
- ✓ Проверьте подписи данных, все должно легко читаться
- ✓ По возможности упростите или уберите легенду
- ✓ По возможности следите, чтобы категорий было не более 5-6, создайте категорию «Прочее»
- ✓ Выбирайте цвет осознанно
- ✓ Выделяйте важное

Визуальные атрибуты

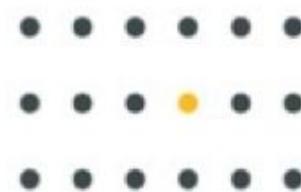
Помогают донести то, на что хотим обратить внимание аудитории



Положение



Форма



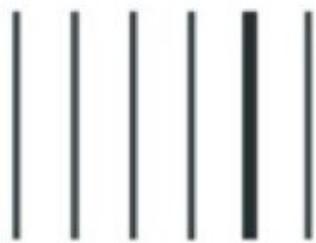
Цвет



Размер



Длина линии



Ширина линии

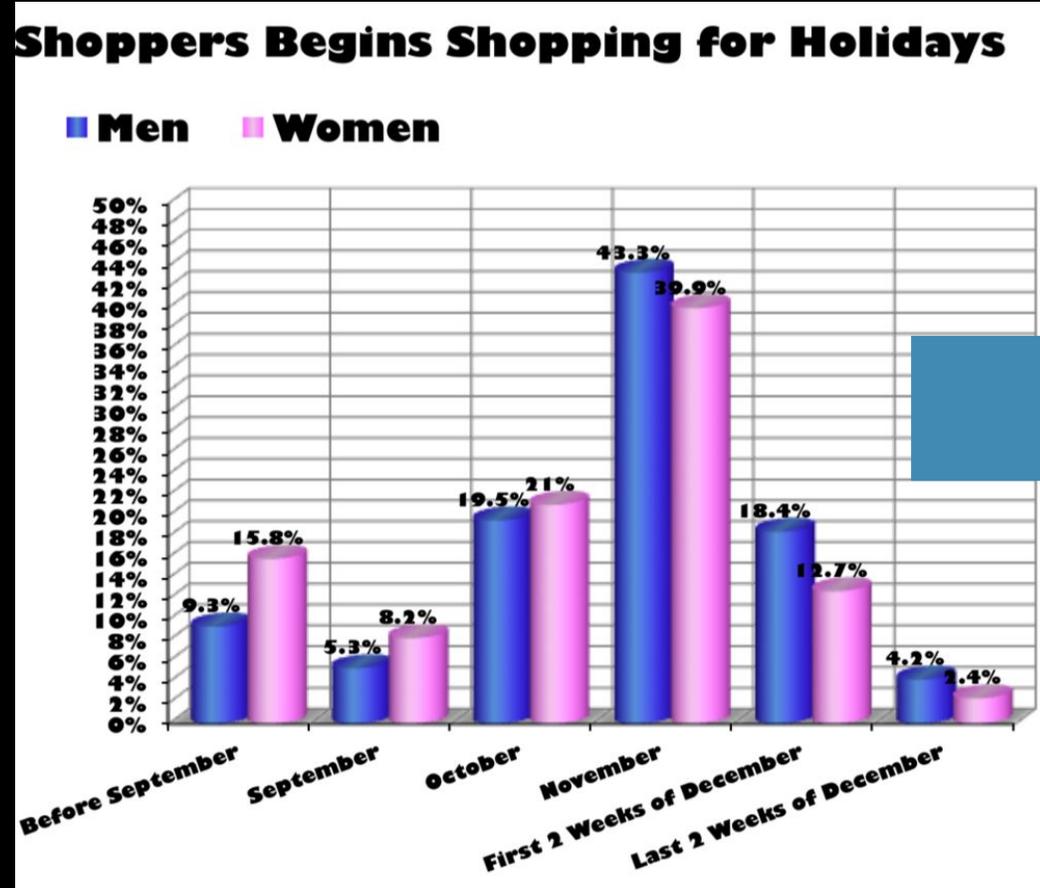


Насыщенность



Положение в пространстве

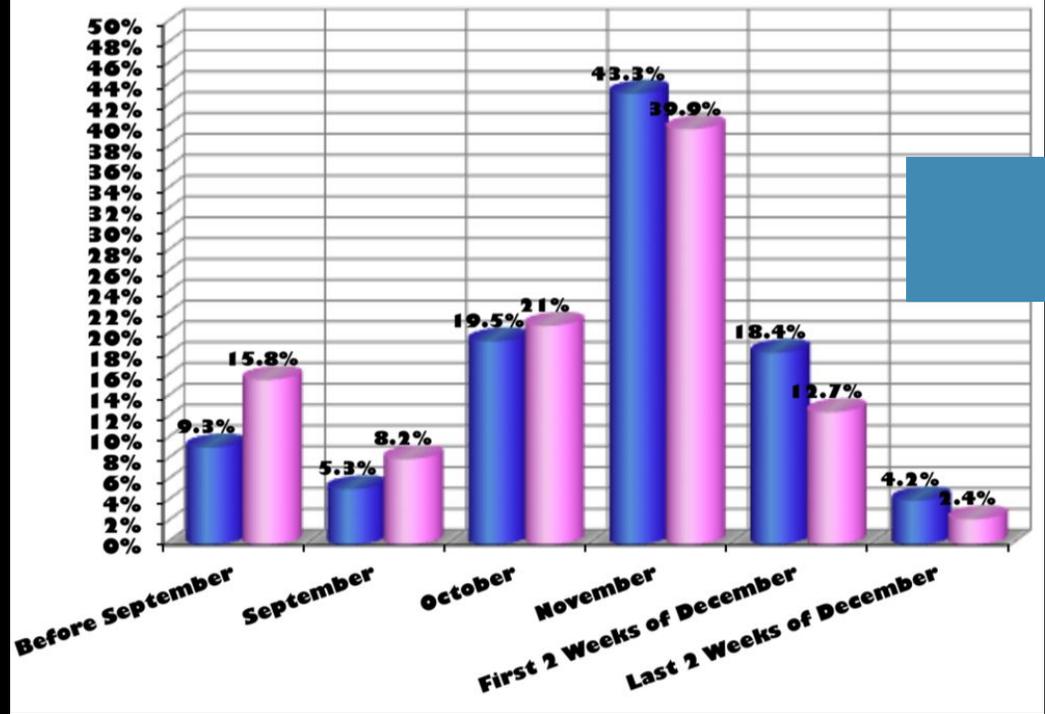
График превращается....



...превращается...

Shoppers Begins Shopping for Holidays

■ Men ■ Women

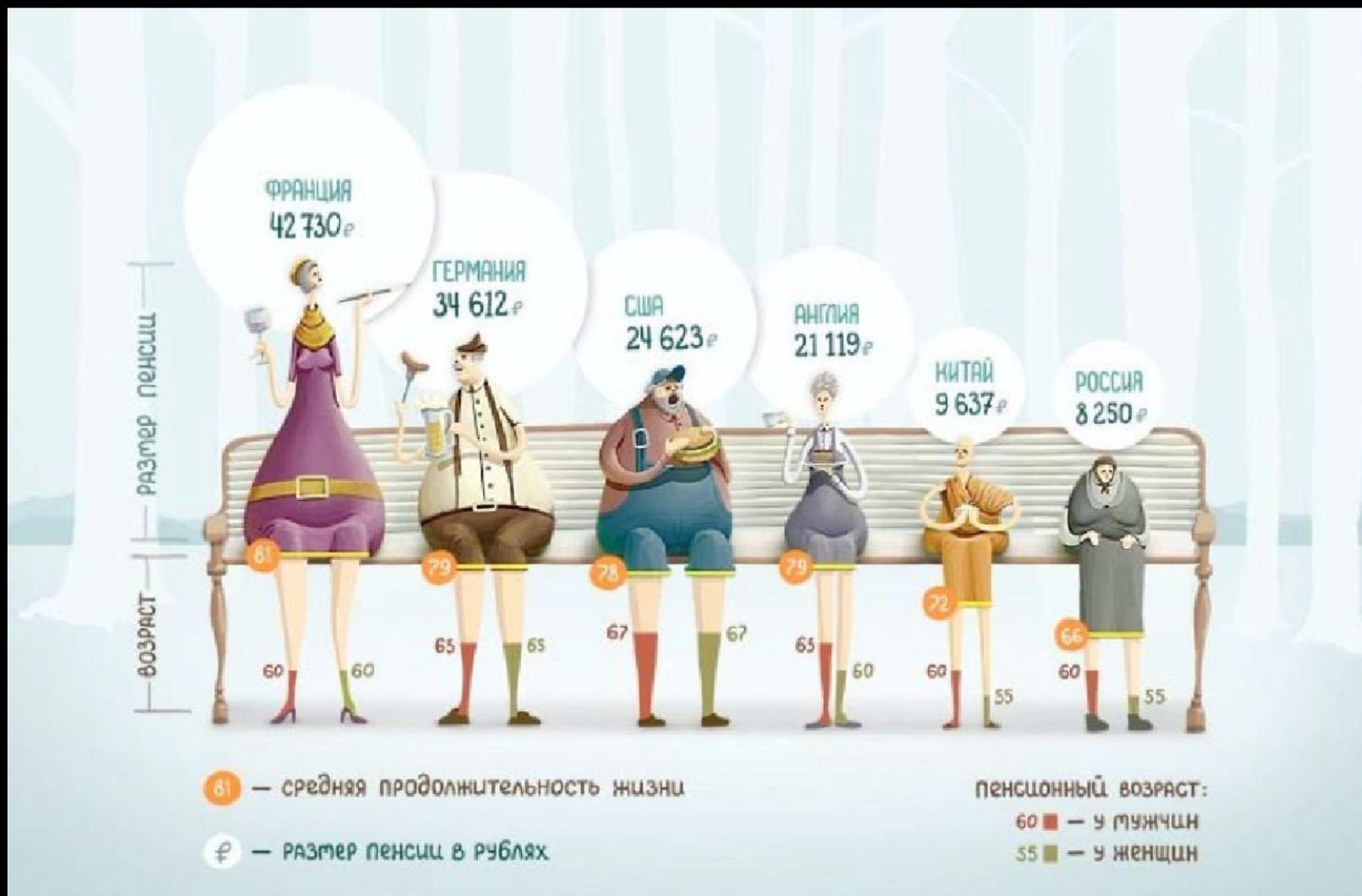


More women start their holiday shopping early

■ Men ■ Women
% OF TOTAL



А дальше можно креативить



А как же инструменты, книги, ссылки?

<https://ft-interactive.github.io/visual-vocabulary/> - ТОТ САМЫЙ ЧИТ

<http://visualizationuniverse.com/> - ТУТ ВСЕ