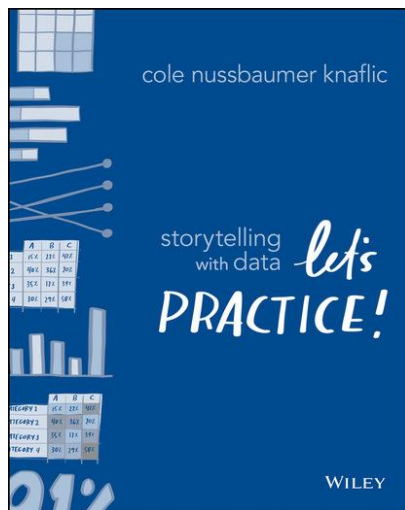


Говори на языке диаграмм: визуализация результатов исследования



Валько Данила Валерьевич
Южно-Уральский технологический университет
valkodv@inuesco.ru

*Владимир Савельев. Статистика и котики



Cole Nussbaumer Knaflic «Storytelling with Data»

URL: habr.com/ru/company/true_engineering/blog/422093



Желязны Джин «Говори на языке диаграмм»

URL: www.mann-ivanov-ferber.ru/books/mif/026

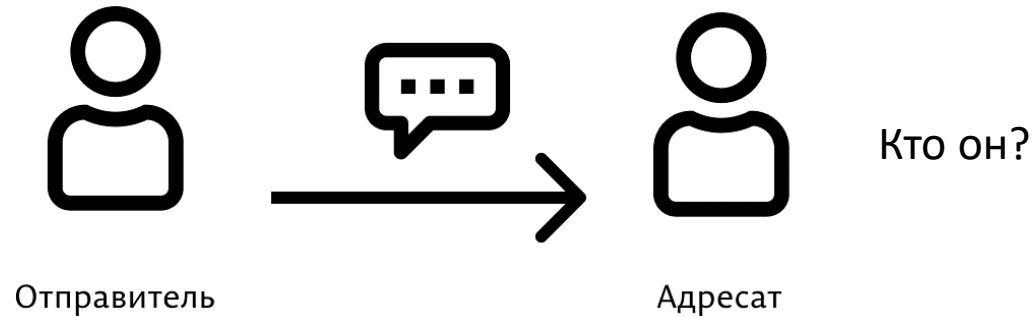


Александр Богачев «Графики, которые убеждают всех»

URL: visualthink.ru/book



Коммуникация



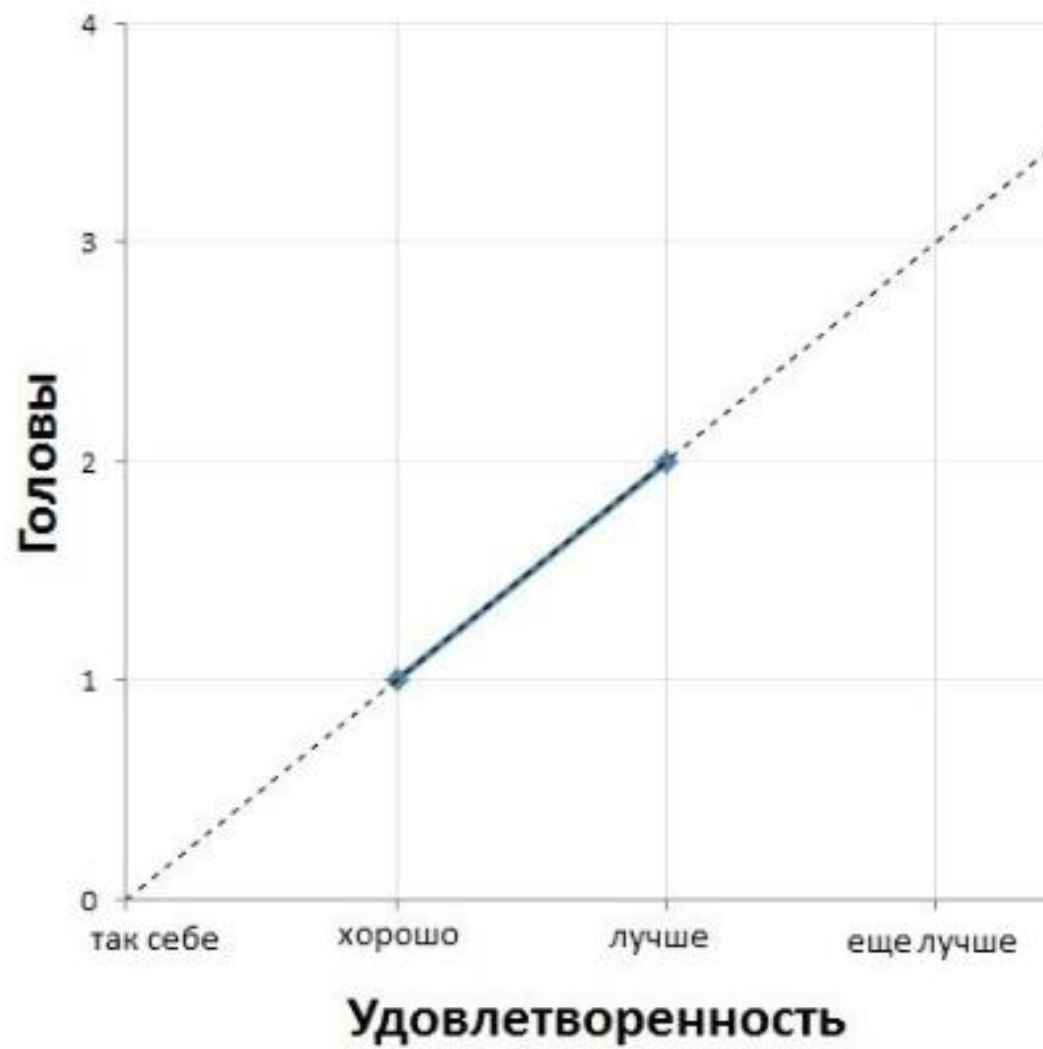
Поиск и исследование, компактное представление данных

- Технические графики/схемы, дашборды, научный отчёт
- Самостоятельный поиск трендов, тенденций, значимых изменений

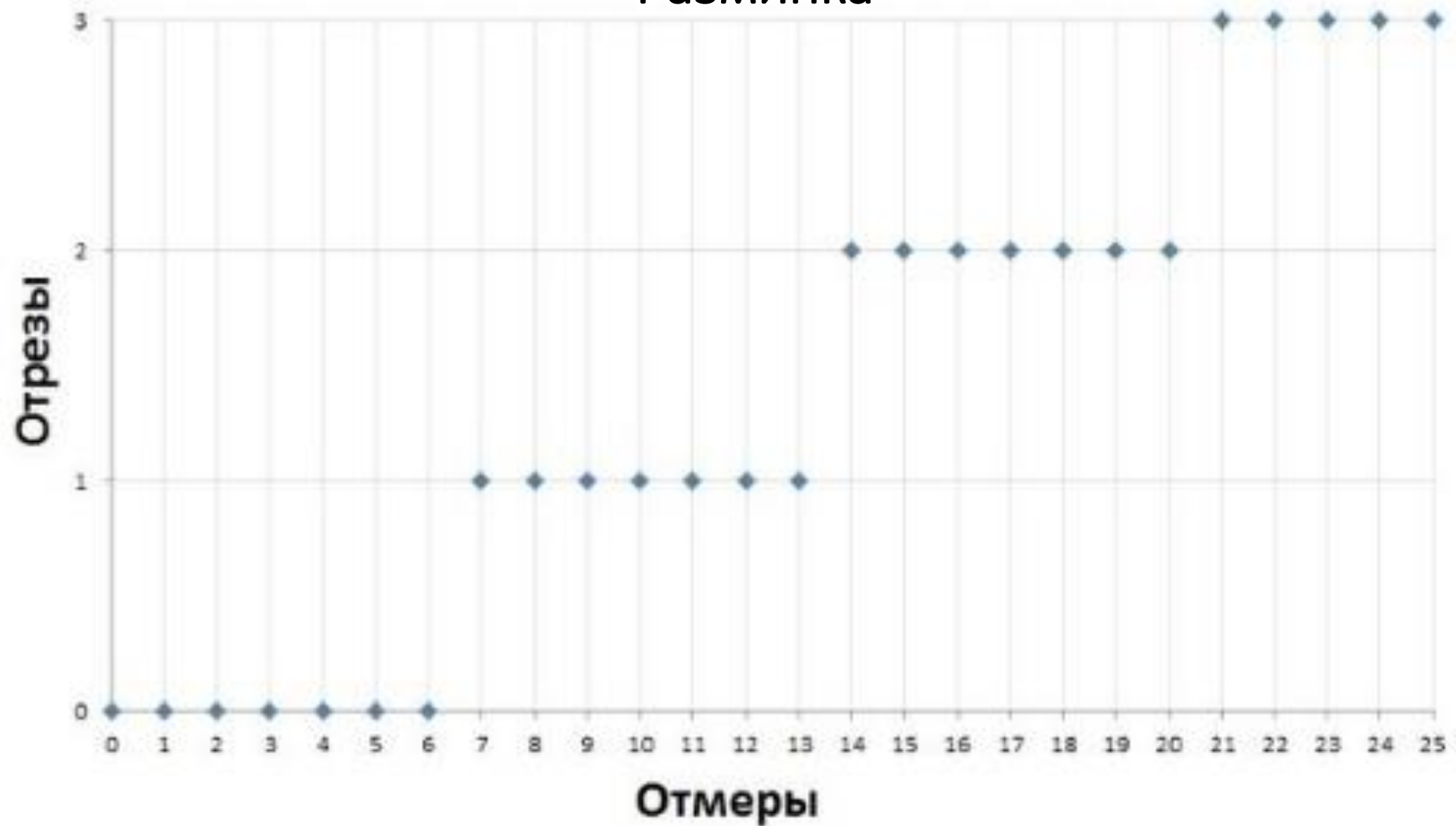
Объяснение и убеждение, аргументация данными

- Инфографика, бизнес-презентация, научный доклад
- Сообщение: готовая *идея*, факт, результат, который необходимо донести

Разминка

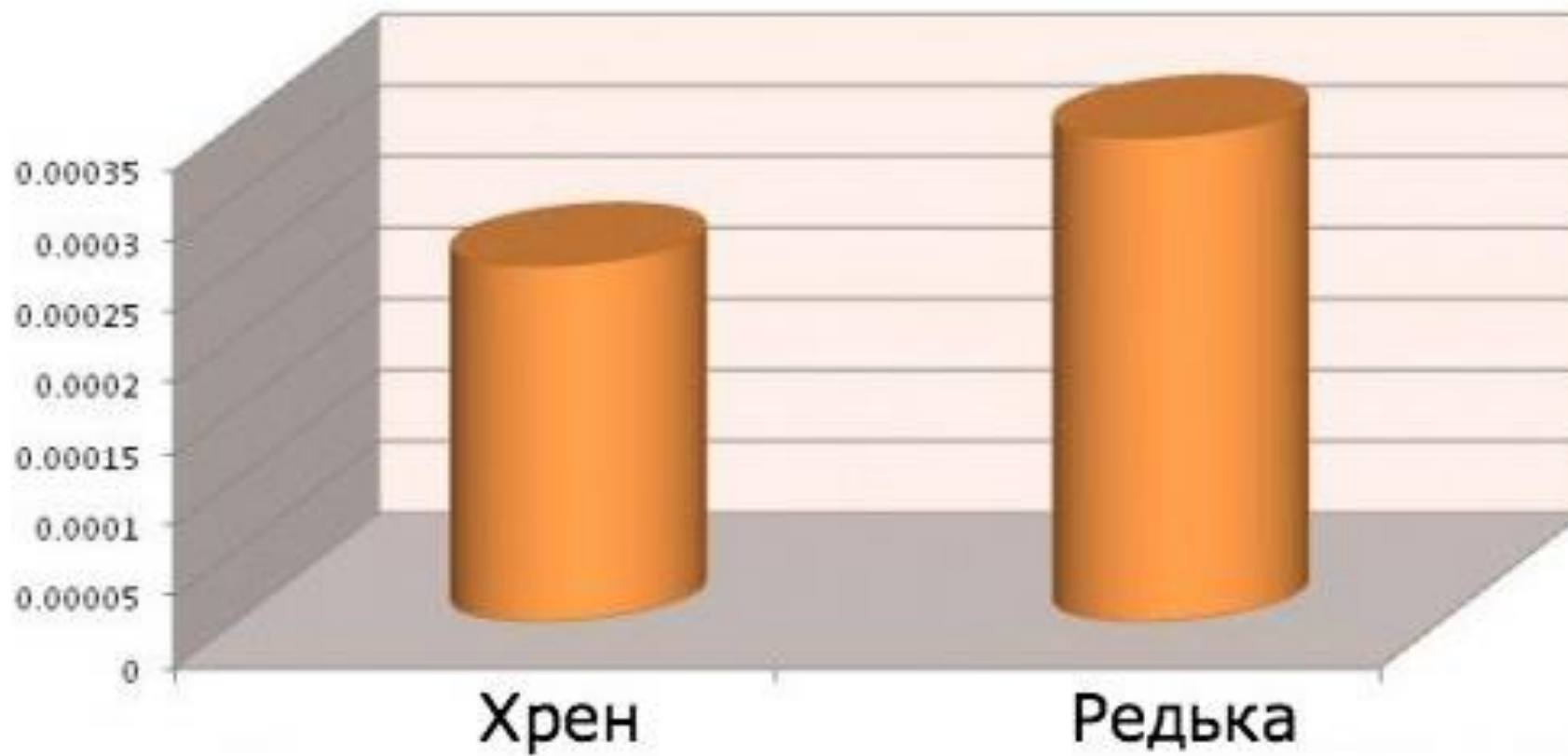


Разминка

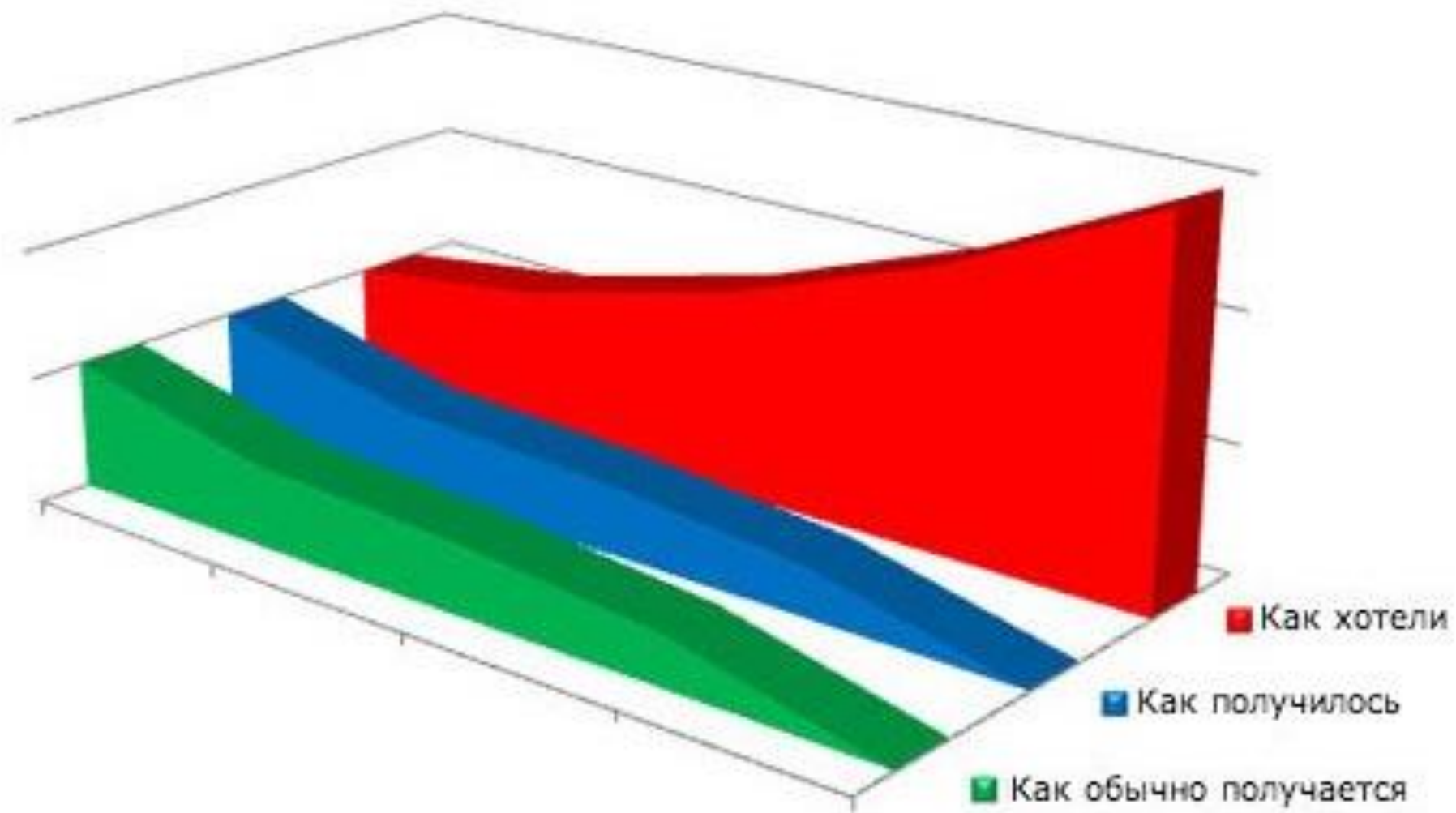


Разминка

Сравнительная характеристика овощной сладости

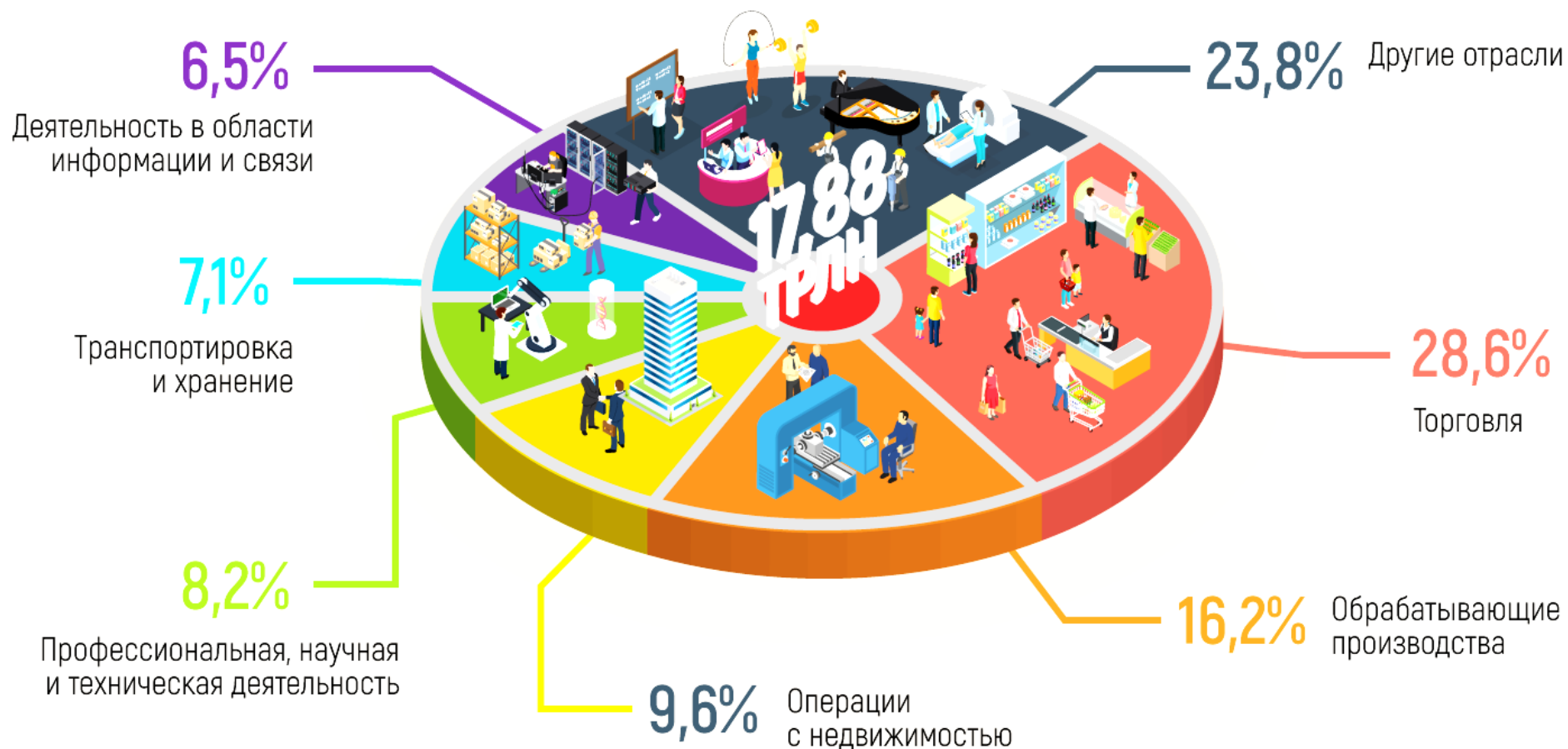


Разминка



Идеи?

СТРУКТУРА ВАЛОВОГО РЕГИОНАЛЬНОГО ПРОДУКТА МОСКВЫ



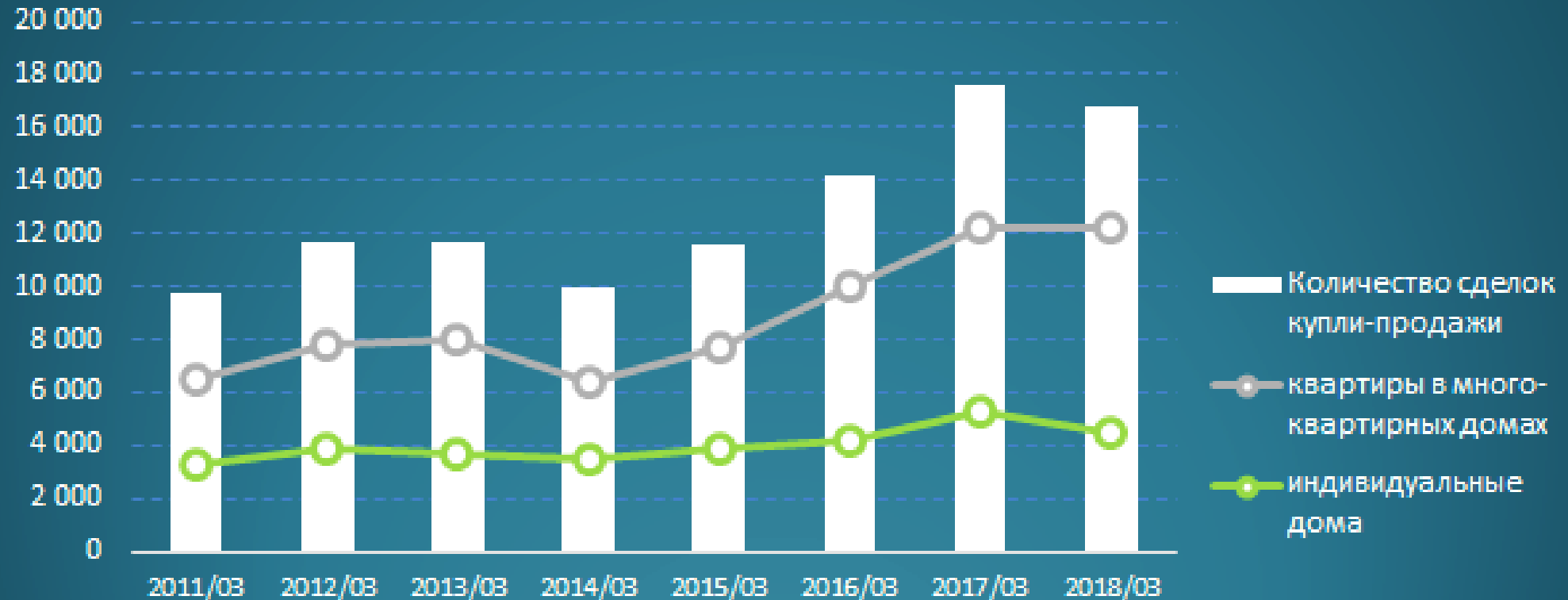
Информация о валовом региональном продукте Москвы за 2018 год

mos.ru

*URL: www.mos.ru/upload/newsfeed/newsfeed/namestofoto.png

Идеи?

Количество сделок купли-продажи | единиц |



Источник: КС МНЭ РК

 EnergyProm

*URL: www.energyprom.kz/storage/app/media/2018/04/17/2.png

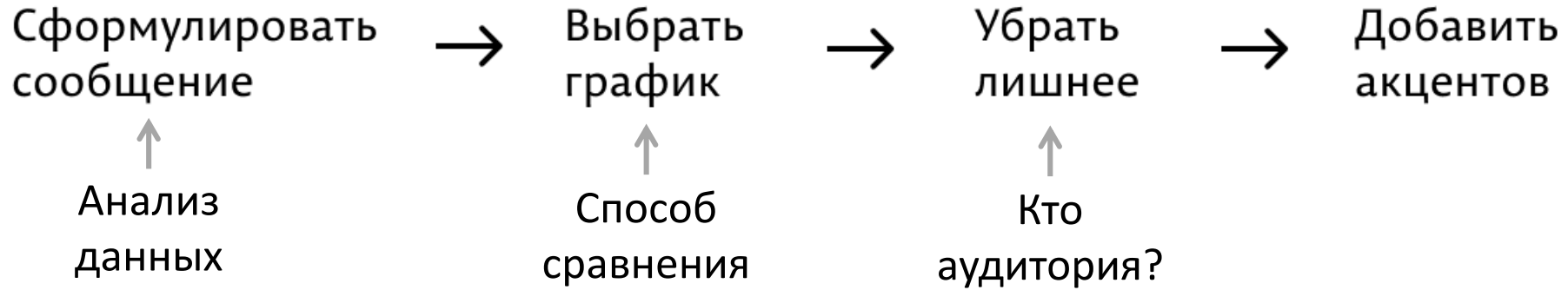
Ticket volume over time

Идеи?



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

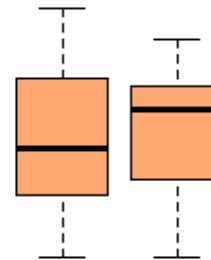
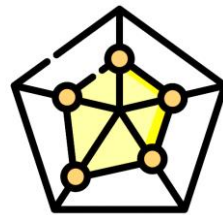
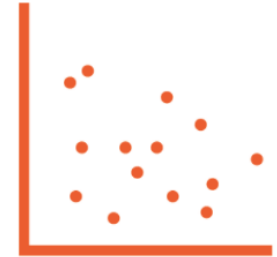
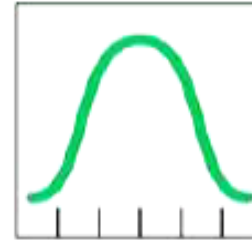
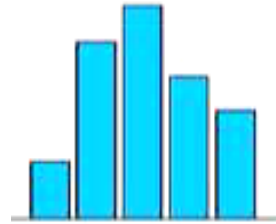
Алгоритм



Сообщение – то, что вы хотите сказать, тот смысл, который вы хотите в диаграмму вложить:

- Основывается на данных (убеждает не график, а данные)
- Возникает как результат анализа (сравнения/сопоставления) данных
- Формулируется в терминах трендов, паттернов, изменения динамики, характера распределения, связи одних данных с другими
- Содержит **ключевые слова** сравнения: повышение/снижение; сезонность, характер роста, сравнение относительно других; занимаемое место и т.д.
- Эффективность коммуникации зависит от характера сообщения

Виды диаграмм



Способ сравнения



Покомпонентное: процент от целого

Позиционное: расположение объектов

Временное: изменения во времени

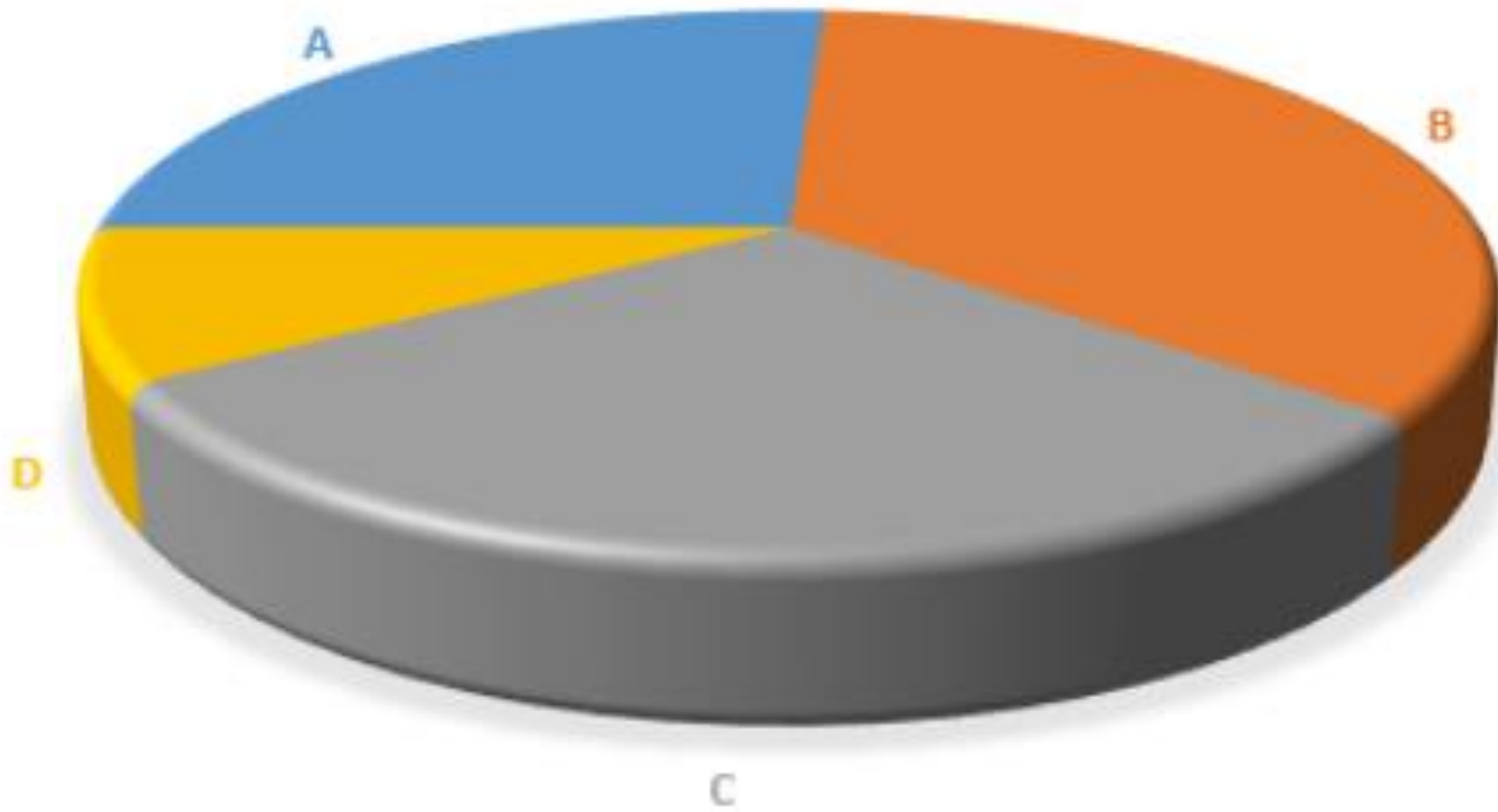
Частотное: объекты в интервалах

Корреляционное: зависимость между переменными

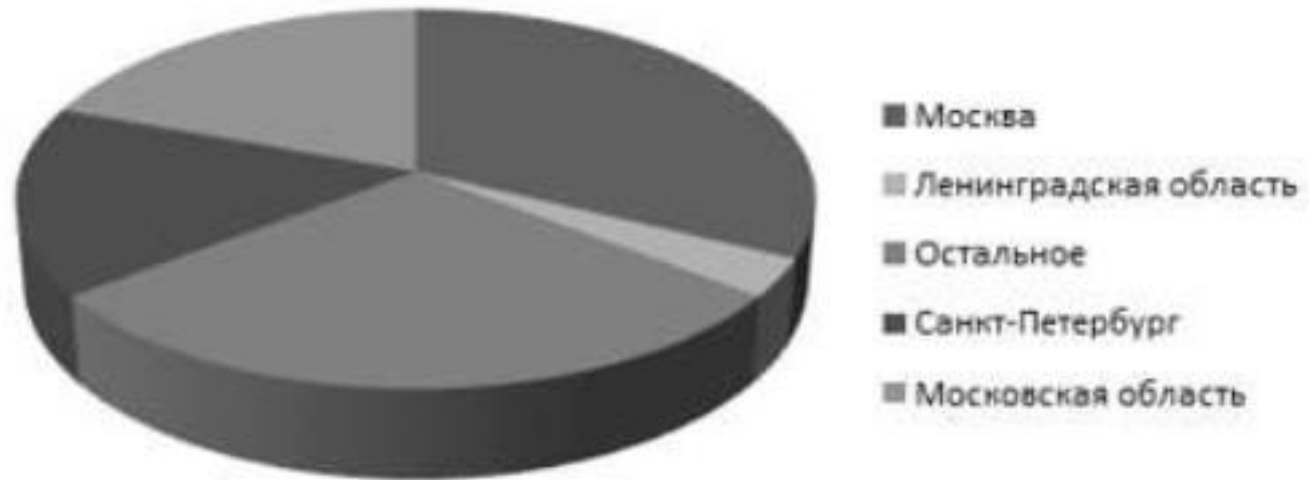
1. ПОКОМПОНЕНТНОЕ СРАВНЕНИЕ

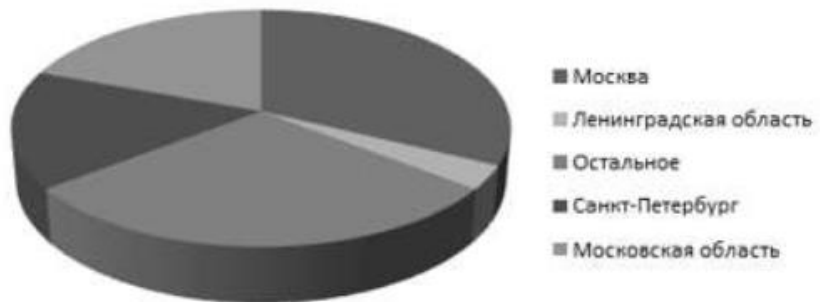
- При покомпонентном сравнении, мы, прежде всего, показываем размер каждого компонента *в процентах* от некоего целого.
- Ключевые слова: *доля, часть, процент от целого, составило X %*
 - ▶ *Доля услуг* в структуре ВВП 2010 г. составляет менее 10%.
 - ▶ В мае продажи продукции А составили наибольшую *долю* в общем объеме продаж компании.
 - ▶ Почти *половина* всех ресурсов привлечена из двух источников.

Насколько С больше В?



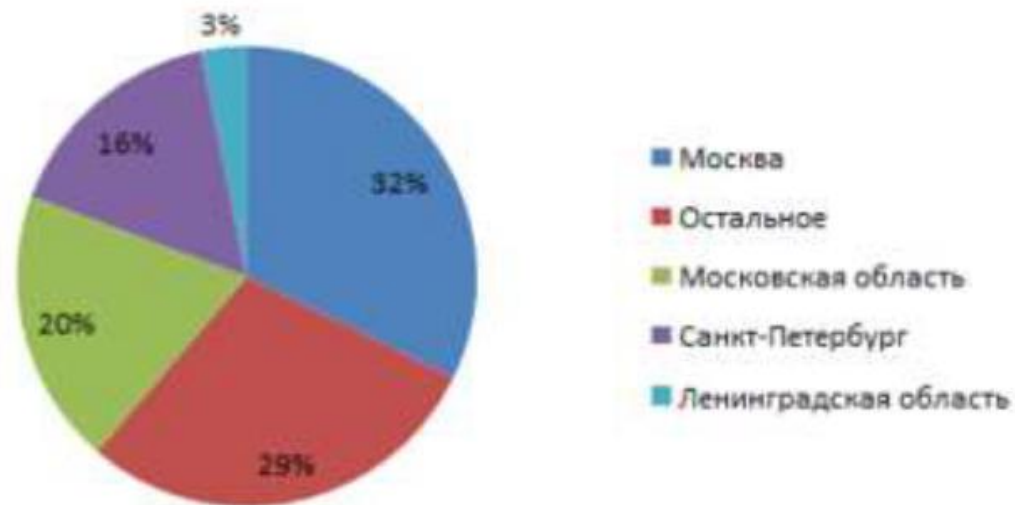
Что плохо?





Разбивка клиентов по регионам

- ? Нет заголовка
- ? Цвета плохо различимы
- ? Не отсортировано
- ? Не указаны ни числа, ни доли
- ? Может, лучше плоскую?





Покомпонентное: процент от целого

Позиционное: расположение объектов

Временное: изменения во времени

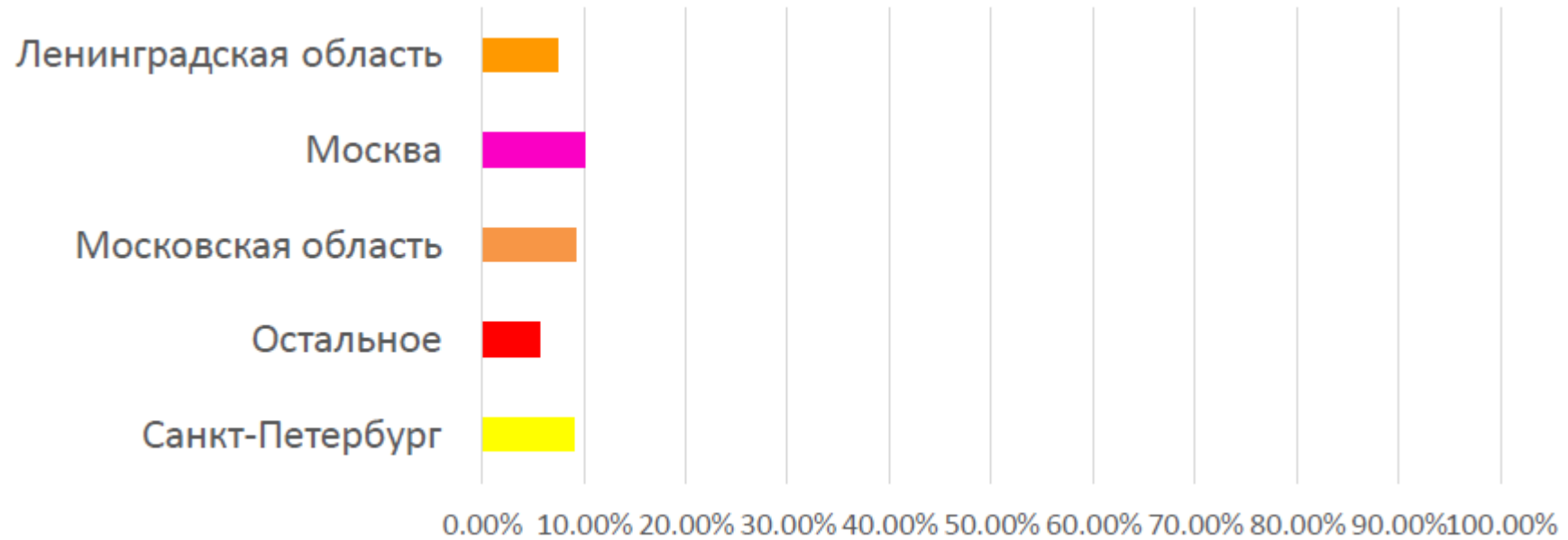
Частотное: объекты в интервалах

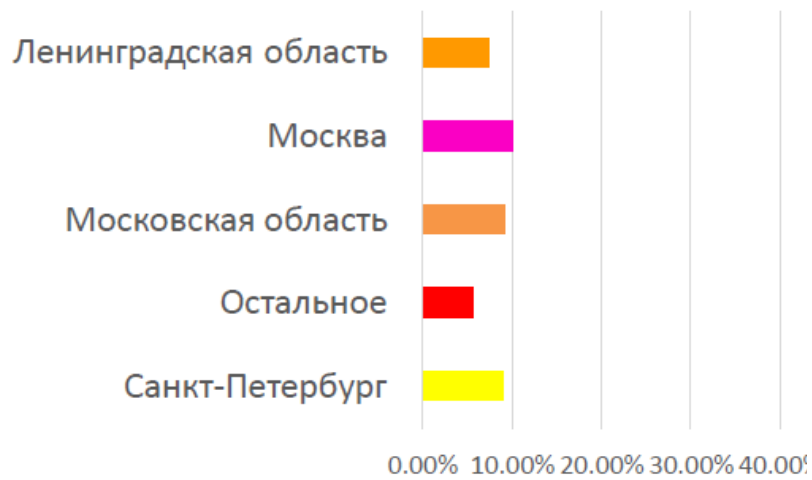
Корреляционное: зависимость между переменными

2. ПОЗИЦИОННОЕ СРАВНЕНИЕ

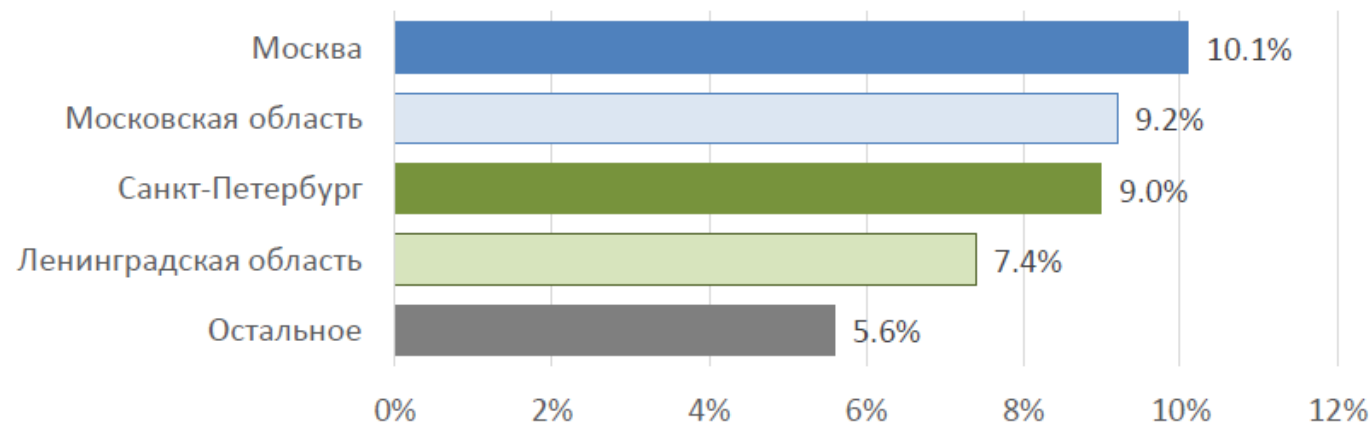
- При позиционном сравнении мы выявляем, как объекты *соотносятся* друг с другом - *одинаковы* ли они, *больше* или *меньше* других.
 - Ключевые слова: *больше чем, меньше чем, равно*
-
- ▶ Бюджетные расходы в регионе А в 2010 г. *превысили* совокупные бюджетные расходы регионов Б и В.
 - ▶ Выручка от продаж компании клиенту N *находится* на четвертом месте.
 - ▶ Текучесть кадров в шести подразделениях примерно *одинакова*.
 - ▶ Город А *сопоставим* с городом Б по плотности населения.

Что плохо?

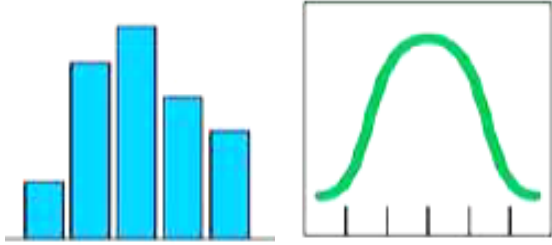




Доля клиентов, воспользовавшихся предложением, в разбивке по регионам



- ❓ Нет заголовка
- ❓ От выбора цветов вытекают глаза (отвлекается внимание)
- ❓ Не отсортировано
- ❓ Нет подписей данных
- ❓ Масштаб не позволяет наглядно сравнить группы



Покомпонентное: процент от целого

Позиционное: расположение объектов

Временное: изменения во времени

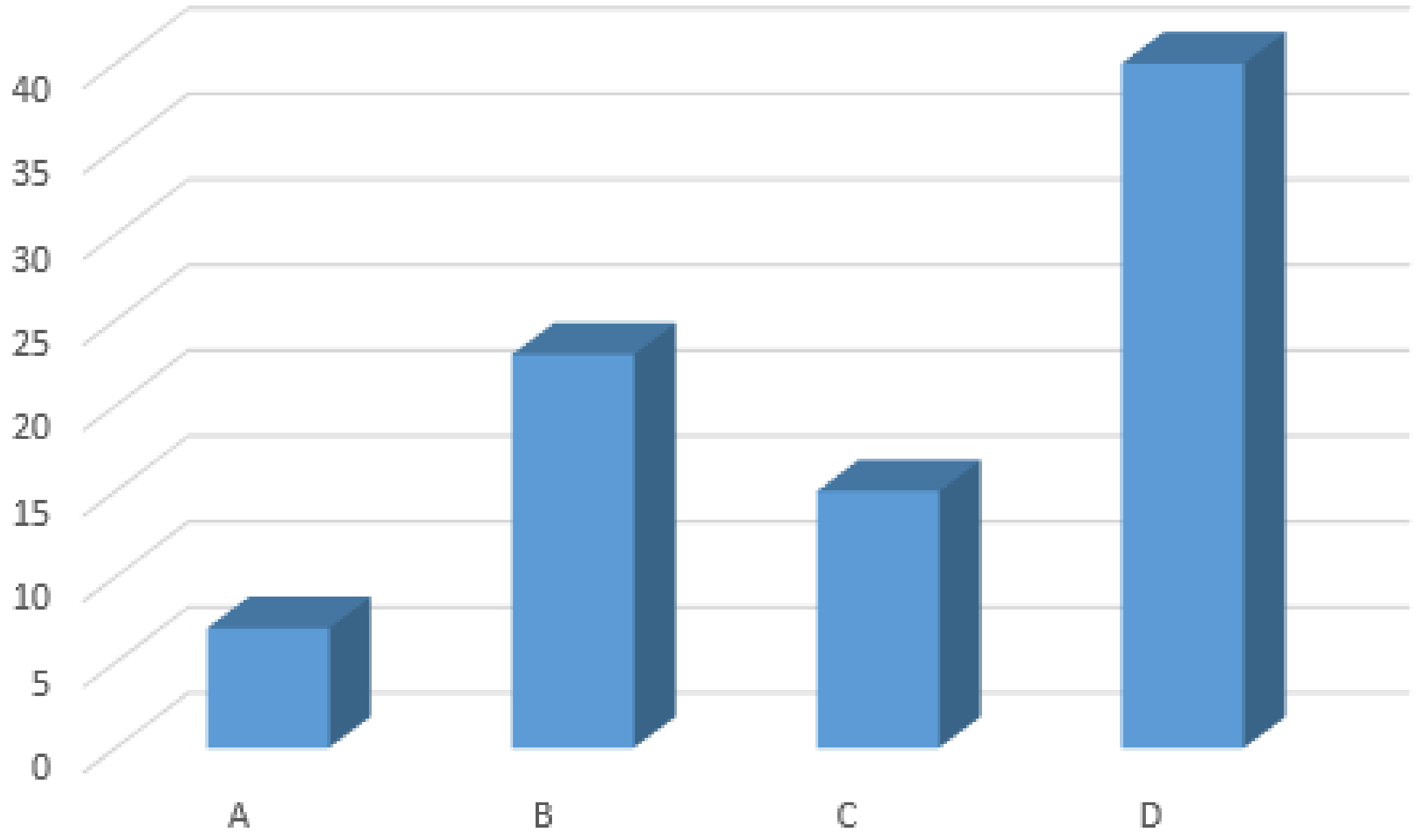
Частотное: объекты в интервалах

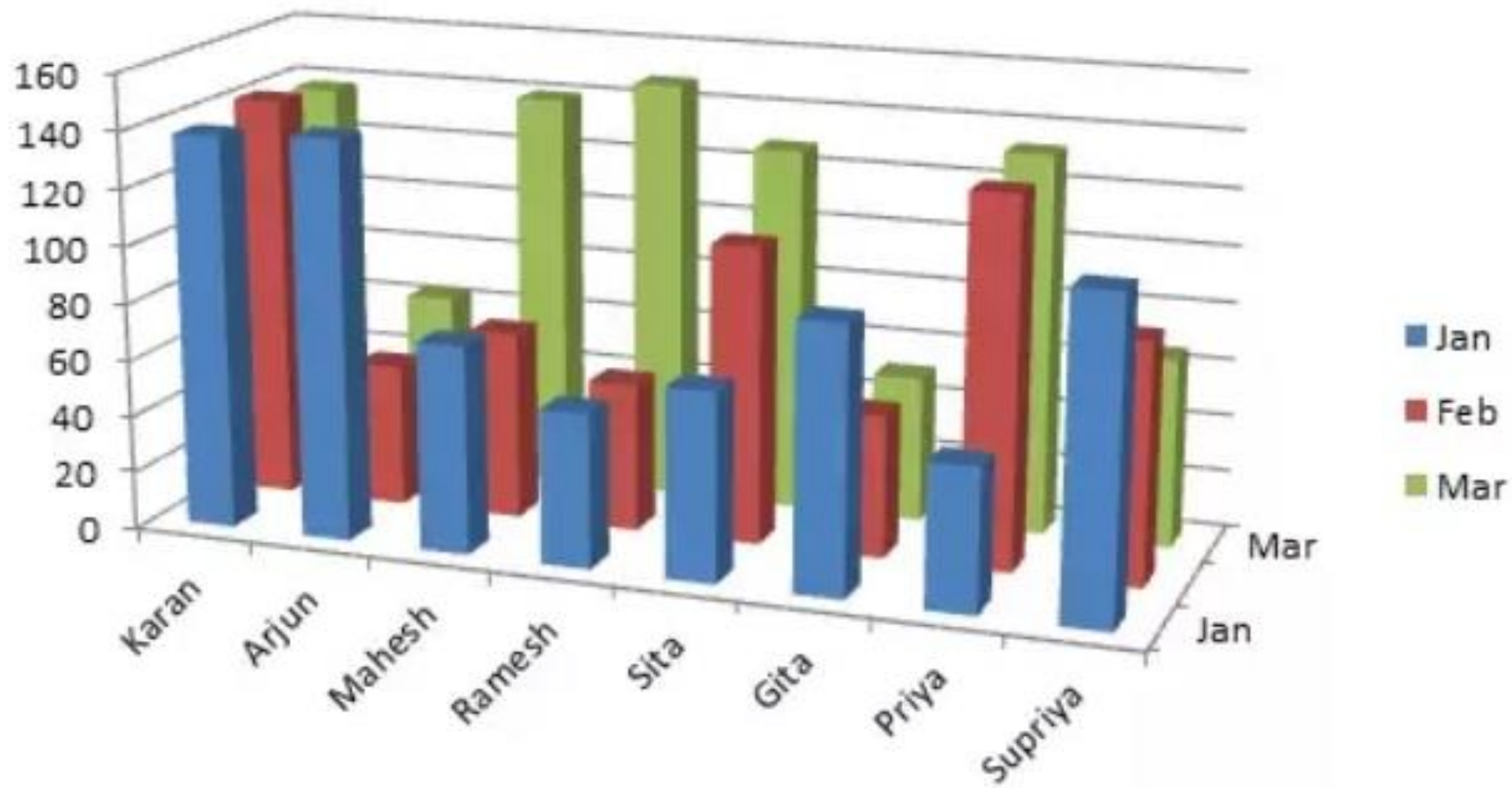
Корреляционное: зависимость между переменными

3. ВРЕМЕННОЕ СРАВНЕНИЕ

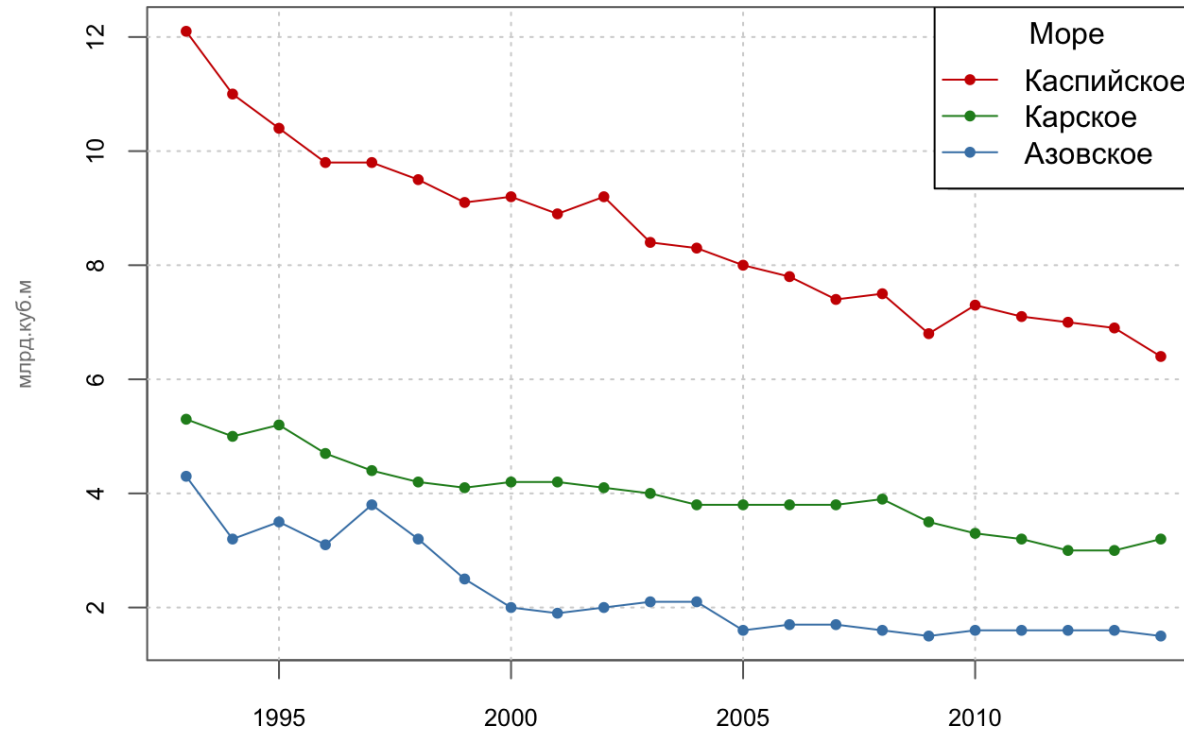
- В данном случае нас интересует не размер каждой доли в сравнении с целым, не соотношение долей, а то, как они *изменяются во времени* – что происходит с определенными показателями на протяжении недель, месяцев, кварталов, лет - *возрастают* ли они, *снижаются*, *колеблются* или *остаются неизменными*.
- Ключевые слова: *изменяться, расти, убывать, возрастая, снижаться, колебаться*
 - ▶ Средняя продолжительность жизни в стране А *снижается* в последние 10 лет.
 - ▶ Продажи в январе неуклонно *росли*.
 - ▶ Рентабельность инвестиций за последние пять лет резко *сократилась*.
 - ▶ Ставки процента в течение последних семи кварталов *колебались*.

Чему равно D?

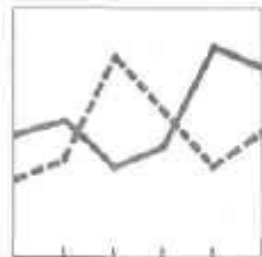




Объем сброса загрязненных сточных вод



- Внимание должно быть сосредоточено на линии тренда.
- Вертикальные линии координатной сетки можно использовать для того, чтобы разделить значения параметра за предыдущие периоды и данные прогноза или разбить временную ось по кварталам или пятилетним периодам.
- Горизонтальные линии помогают точнее сравнивать относительные величины.



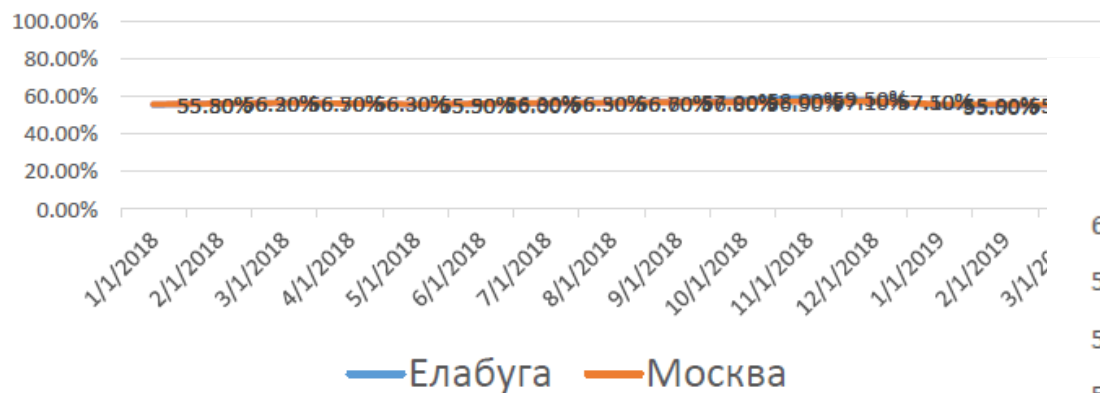
Совмещенный график показывает сравнение двух и более параметров. Для исследуемого объекта используйте контрастный цвет или самую жирную линию, а для остальных - менее яркие цвета и более тонкие или пунктирные линии.

Что плохо?

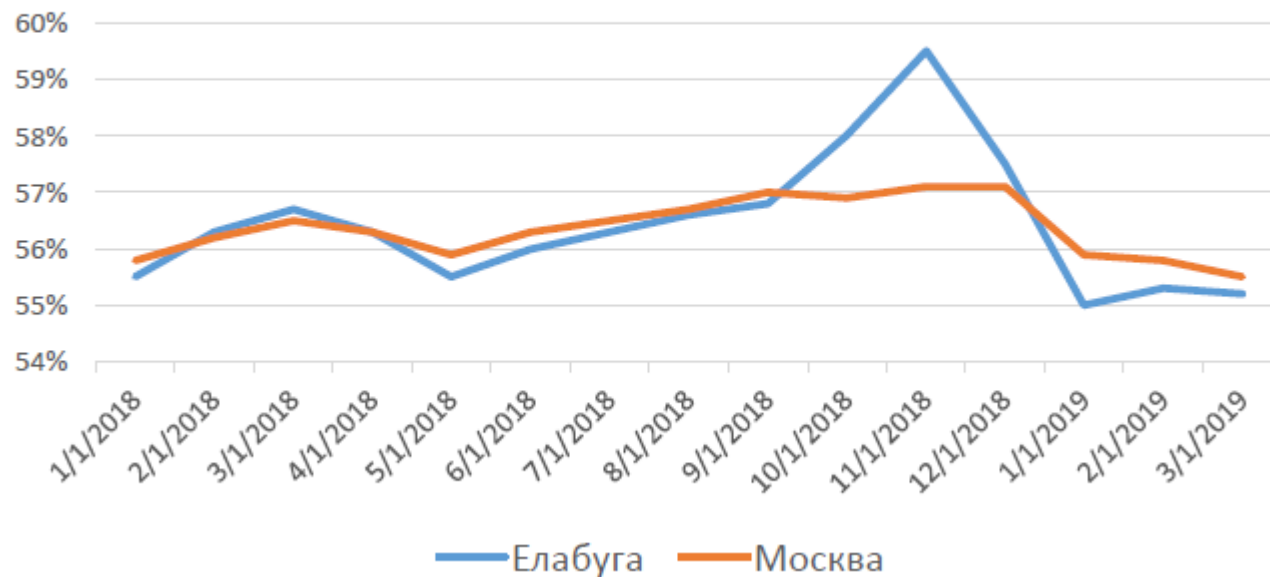
Утилизация дебетовых карт в зависимости от города клиента



Утилизация дебетовых карт в зависимости от города клиента

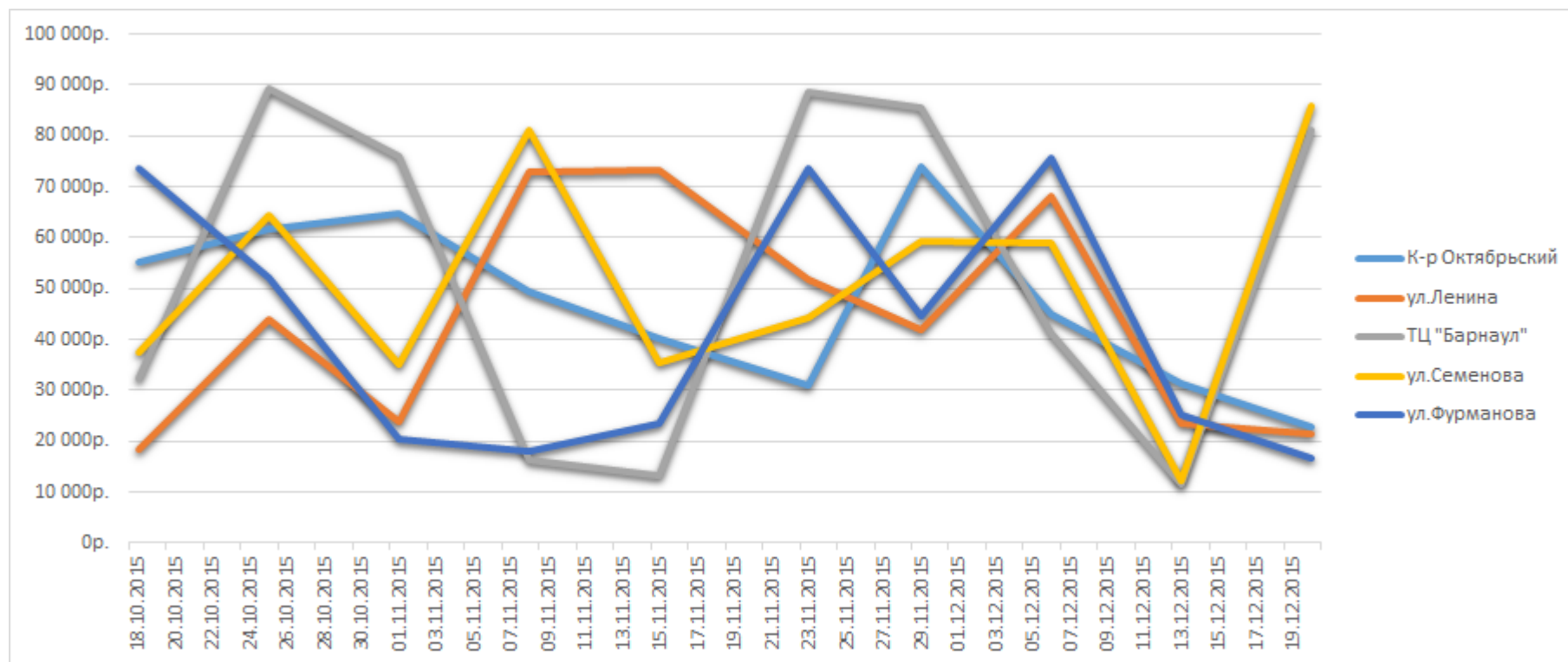


Утилизация кредитных карт зависимости от города клиента



- ? Мои глаза! Подписи явно лишние.
- ? Не видно различий между линиями из-за масштаба
- ? Не видно временных эффектов

Что плохо?



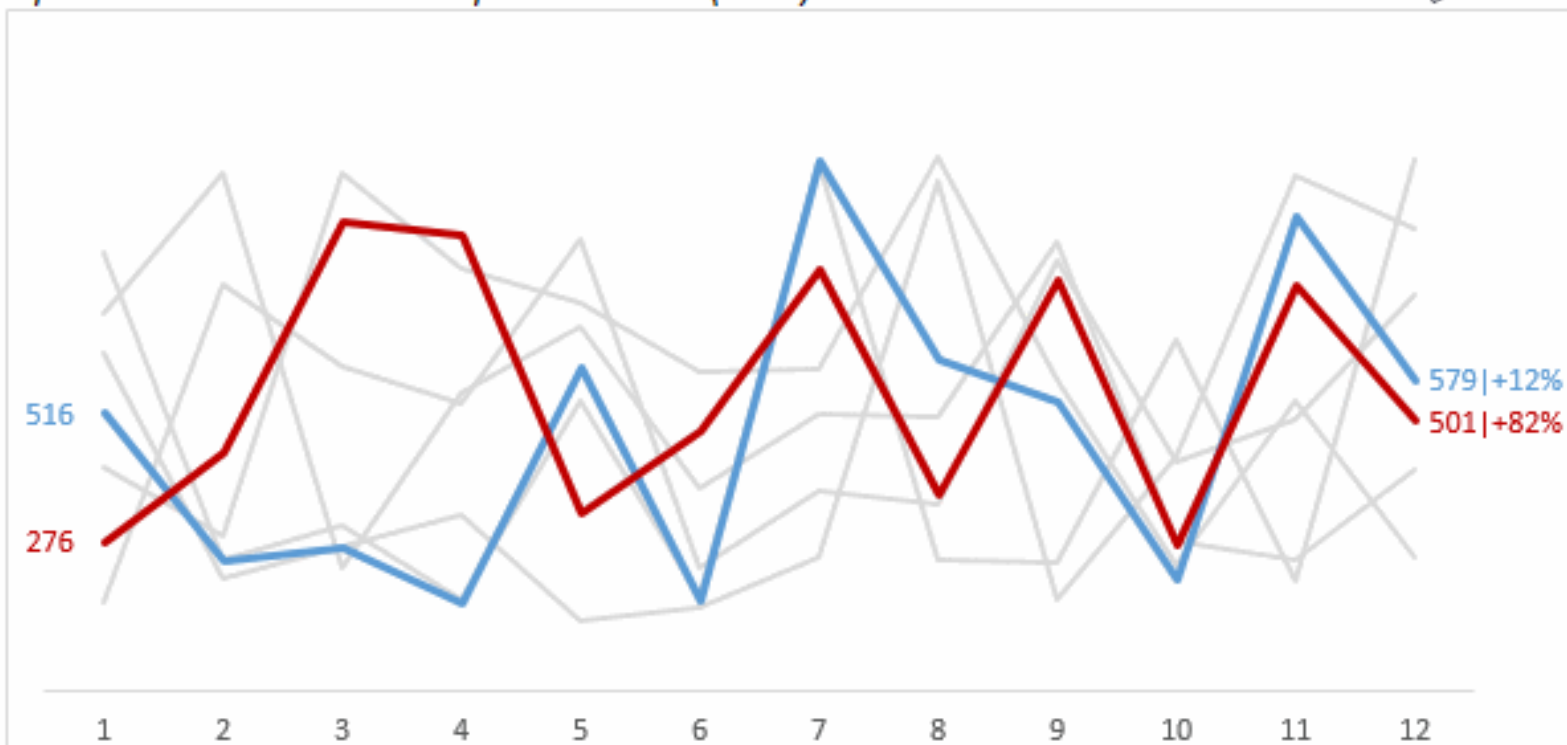
ВЫБЕРИ

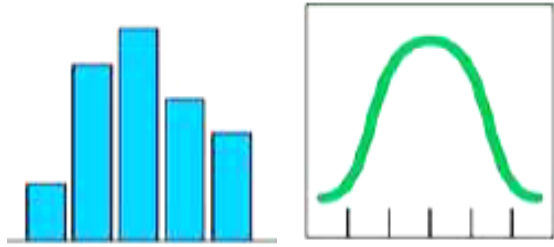
E

vs.

D

Сравнительный анализ продаж 2017 (шт.)





Покомпонентное: процент от целого

Позиционное: расположение объектов

Временное: изменения во времени

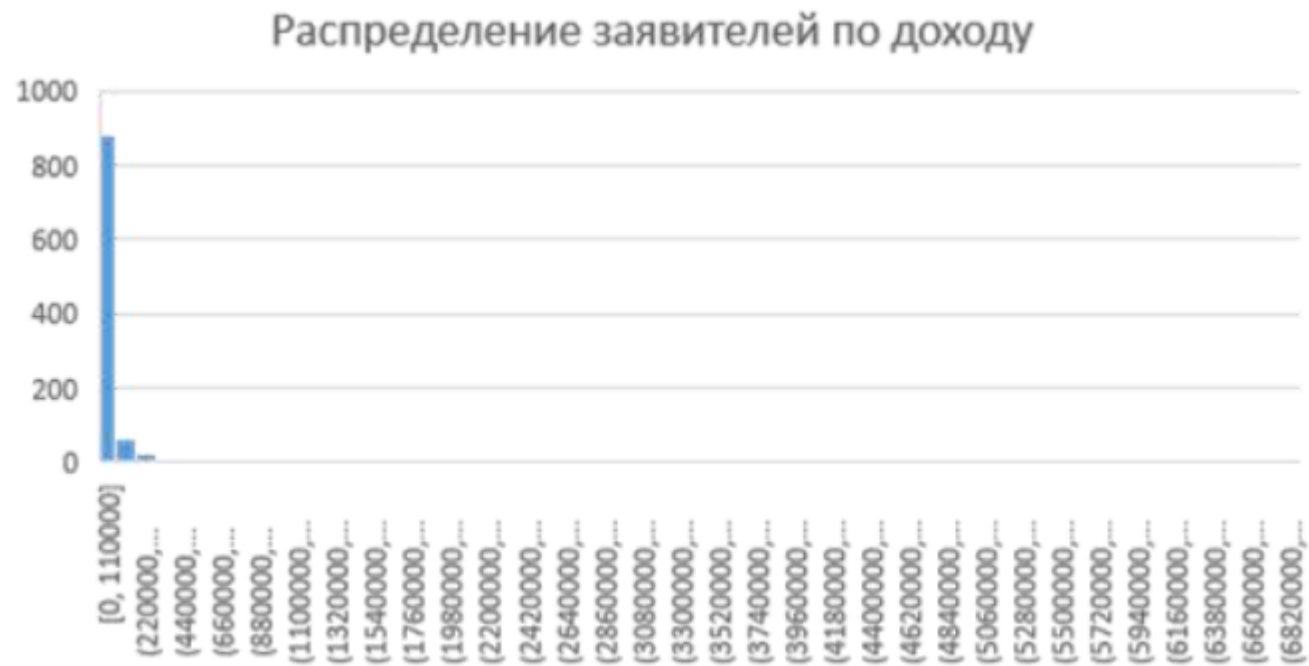
Частотное: объекты в интервалах

Корреляционное: зависимость между переменными

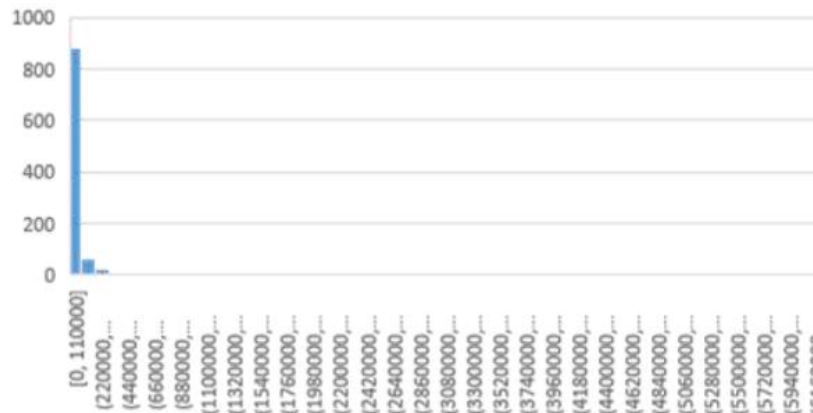
4. ЧАСТОТНОЕ СРАВНЕНИЕ

- Данный вид сравнения помогает определить, *сколько объектов попадает в определенные интервалы (диапазоны)*.
 - Ключевые слова: *в диапазоне от x до y; концентрация; частотность и распределение*
-
- ▶ *Диапазон* зарплат в регионе А составляет от 10 до 78 тыс. руб.
 - ▶ В мае сделки в основном заключались в *диапазоне* от одной до двух тысяч долларов.
 - ▶ Большая часть грузов была получена за *2-5 дней*.
 - ▶ Молодые люди *чаще* голосуют за независимых кандидатов.

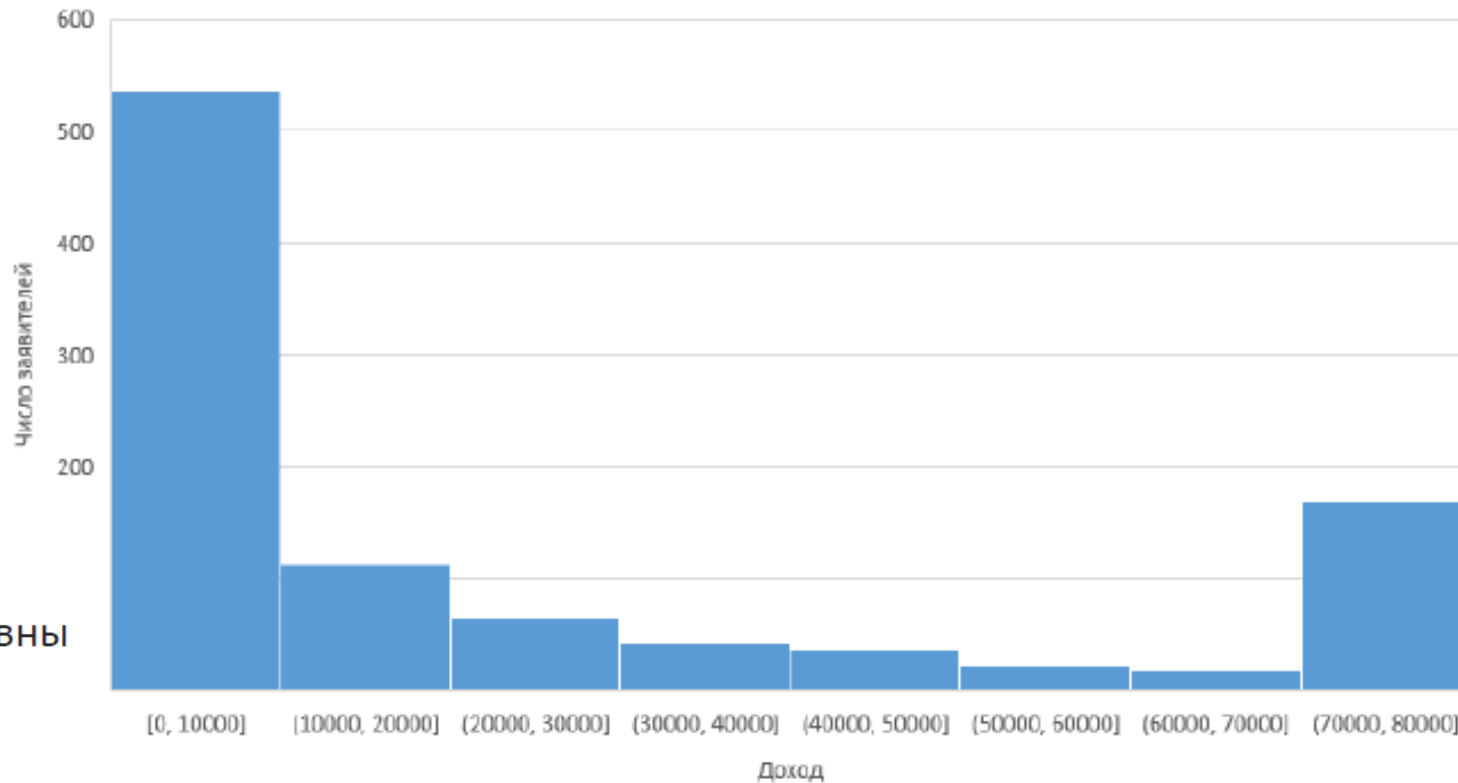
Что плохо?



Распределение заявителей по доходу



Распределение заявителей по доходу



- ? Не ясна структура
- ? Большое число столбиков не информативны
- ? Плохой выбор интервалов
- ? Многоточия? Загадочно..



Покомпонентное: процент от целого

Позиционное: расположение объектов

Временное: изменения во времени

Частотное: объекты в интервалах

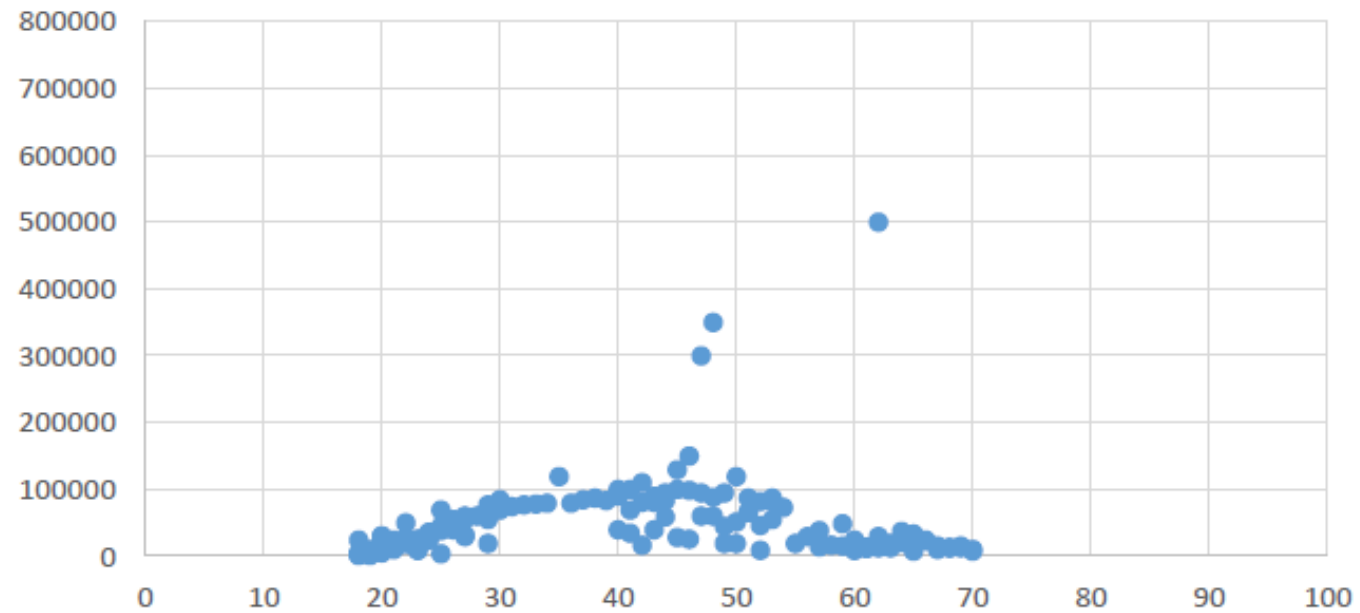
Корреляционное: зависимость между переменными

5. КОРРЕЛЯЦИОННОЕ СРАВНЕНИЕ

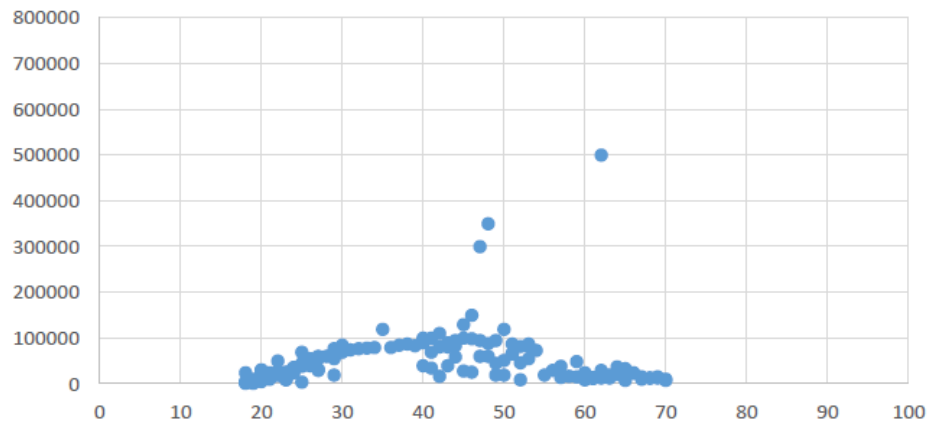
- Корреляционное сравнение показывает наличие (или отсутствие) зависимости между двумя переменными. Например, обычно ожидается, что при увеличении объемов продаж возрастает прибыль и т.п.
- Ключевые слова: *относится к, возрастает при (в случае), снижается при (в случае), меняется при (в случае)*
 - ▶ Зарплата исполнительных директоров не зависит от размера компании.
 - ▶ Уровень доходов нелинейно связан с уровнем образования.
 - ▶ Результаты продаж в мае демонстрируют отсутствие взаимосвязи между объемом продаж и опытом продавцов.
 - ▶ Страховая сумма возрастает при росте доходов индивидов.

Что плохо?

зависимость между доходом и возрастом клиента

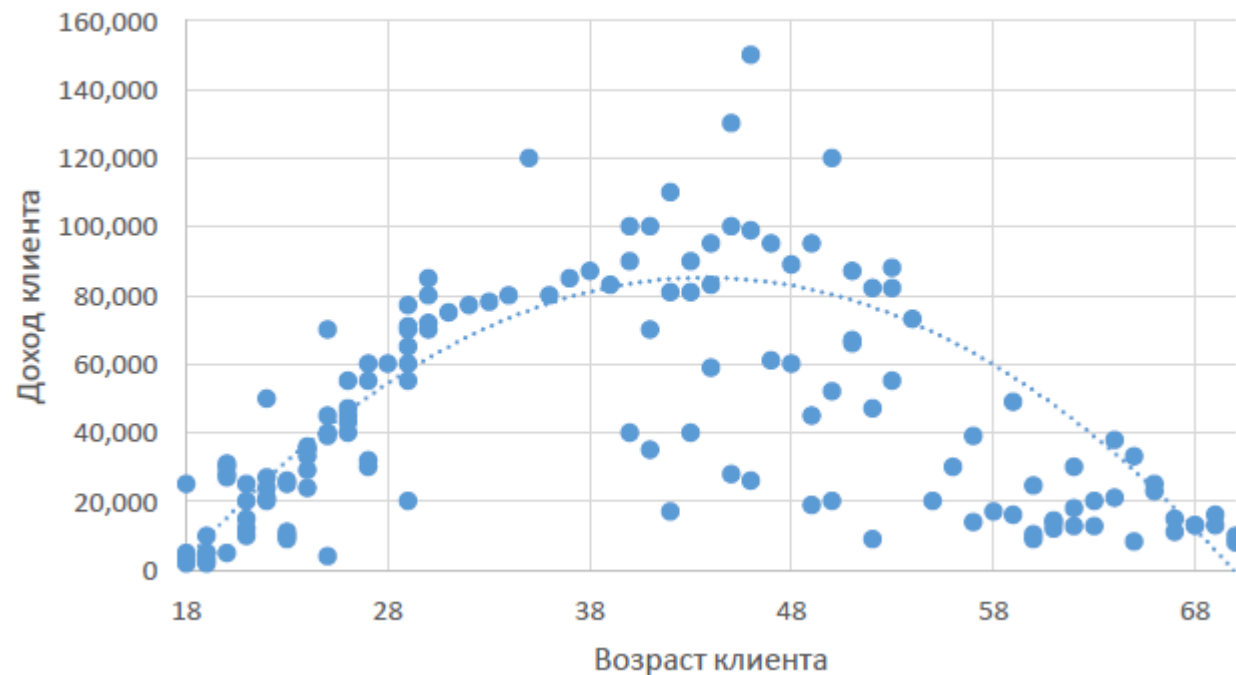


зависимость между доходом и возрастом клиента



- ? Не видно тренда
- ? Мешают выбросы
- ? Неудачный масштаб
- ? Не подписаны оси

зависимость между доходом и возрастом клиента



Покомпонентное: процент от целого

Позиционное: расположение объектов

Временное: изменения во времени

Частотное: объекты в интервалах

Корреляционное: зависимость между переменными

Идеи?

1. В течение следующих десяти лет прогнозируется увеличение объема продаж
2. Большинство сотрудников получает от 30 до 35 тыс. долл.
3. Повышение цены на отдельные сорта бензина не означает повышения их качества
4. В сентябре уровень текучести кадров в шести подразделениях был примерно одинаков
5. Менеджер по продажам проводит с клиентами лишь 15 % своего времени
6. Размер прибавки по результатам работы не зависит от выслуги лет
7. В прошлом году наибольшая текучесть кадров наблюдалась в возрастной группе от 30 до 35 лет
8. Центральный регион занимает последнее место по производительности
9. Доходность акций нашей компании сокращается
10. Наибольшая доля фондов задействована в производстве
11. Наблюдается связь между доходами и зарплатой
12. В августе два завода обогнали по производительности шесть других

Выбор диаграмм

Шаг 1: ФОРМУЛИРОВАНИЕ ИДЕИ

(от данных к идее)

- Для того, чтобы правильно выбрать тип диаграммы, вы в первую очередь должны четко сформулировать конкретную идею, которую вы хотите донести до аудитории при помощи диаграммы.

Шаг 2: ОПРЕДЕЛЕНИЕ ТИПА СРАВНЕНИЯ ДАННЫХ

(от идеи к сравнению)

- Сформулированная вами идея будет обязательно заключать в себе один из пяти основных типов сравнения данных: покомпонентное, позиционное, временное, частотное и корреляционное.

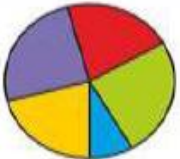
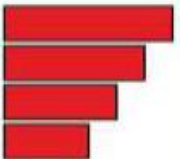
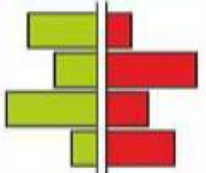
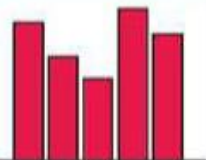

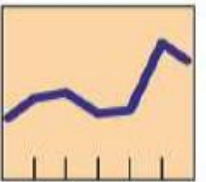
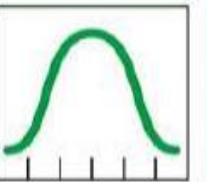
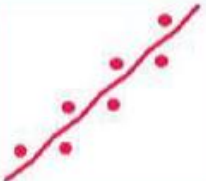
Шаг 3: ВЫБОР ТИПА ДИАГРАММЫ

(от сравнения к диаграмме)

- Каждому типу сравнения соответствует один из пяти видов диаграмм.

ТИПЫ СРАВНЕНИЯ

ОСНОВНЫЕ ТИПЫ ДИАГРАММ

	ПОКОМПОНЕНТНОЕ	ПОЗИЦИОННОЕ	ВРЕМЕННОЕ	ЧАСТОТНОЕ	КОРРЕЛЯЦИОННОЕ
КРУГОВАЯ					
ЛИНЕЙЧАТАЯ					
ГИСТОГРАММА					
ГРАФИК					
ТОЧЕЧНАЯ					

Убрать лишнее

Посчитайте количество цифр «3»

756395068473

658663037576

860372658602

846589107830

Убрать лишнее

Посчитайте количество цифр «3»

756**3**9506847**3**

65866**3**0**3**7576

860**3**72658602

8465891078**3**0

Убрать лишнее

- Выделите важные моменты. Рекомендуется выделять не больше 10% элементов визуализации, иначе теряется эффект
- Уточните масштаб и шкалы
- Устраните отвлекающие факторы
 - + Не все данные одинаково важны. Избавляйтесь от некритических компонентов.
 - + Когда детали не нужны, агрегируйте.
 - + Периодически спрашивайте себя: если это выкинуть, то что-нибудь изменится?
 - + Отправляйте второстепенные элементы на задний план. Используйте для этого серый цвет
- Создайте чёткую иерархию информации. Используя разные атрибуты, привлекающие внимание, покажите вашей аудитории, в какой последовательности надо получать информацию

Give them a break?

2019 revenue forecast, as % of total

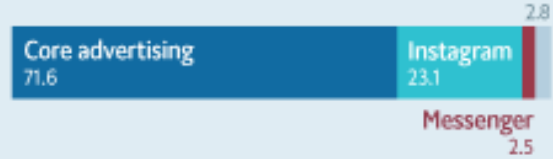
Alphabet (Google)

Total: \$132.2bn

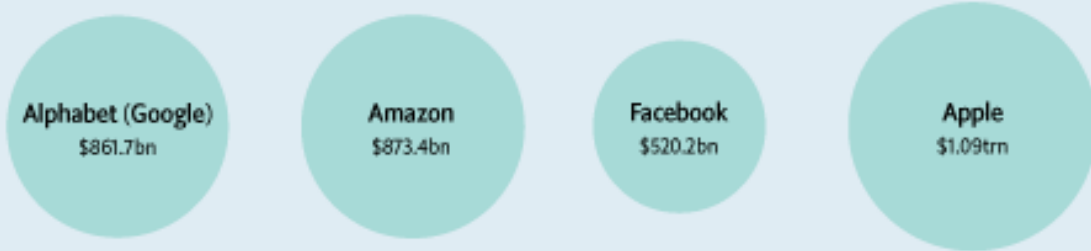


Facebook

Total: \$70.2bn



Market capitalisation, October 23rd 2019



Sources: Jefferies; UBS; Evercore ISI; Bloomberg

Amazon

Total: \$278.8bn



Apple

Total: \$259.0bn



This is how the biggest technology giants make their money

2019 revenue forecast, as % of total

Alphabet (Google)



Amazon



Facebook



Apple



Sources: Jefferies; UBS; Evercore ISI; Bloomberg

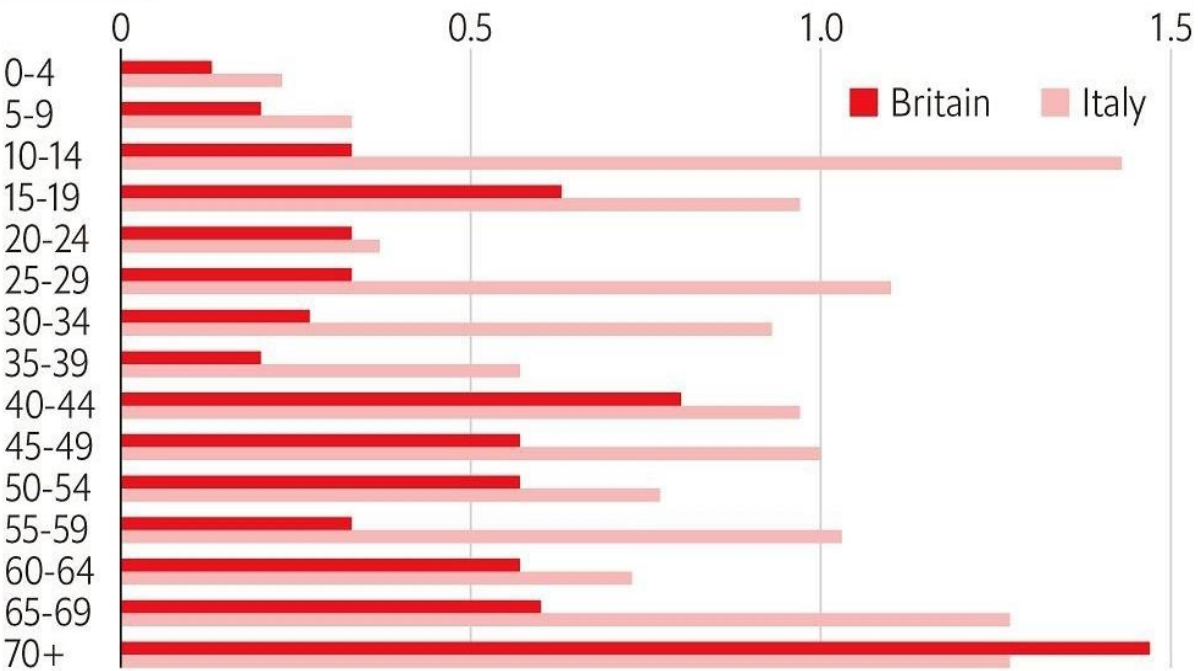
The Economist

*URL: medium.economist.com/charting-new-territory-7f5afb293270

How *The Economist's* data team makes bar charts better

Daily encounters people aged 70+ have with different age groups, average, May 2005-Sep 2006

Before

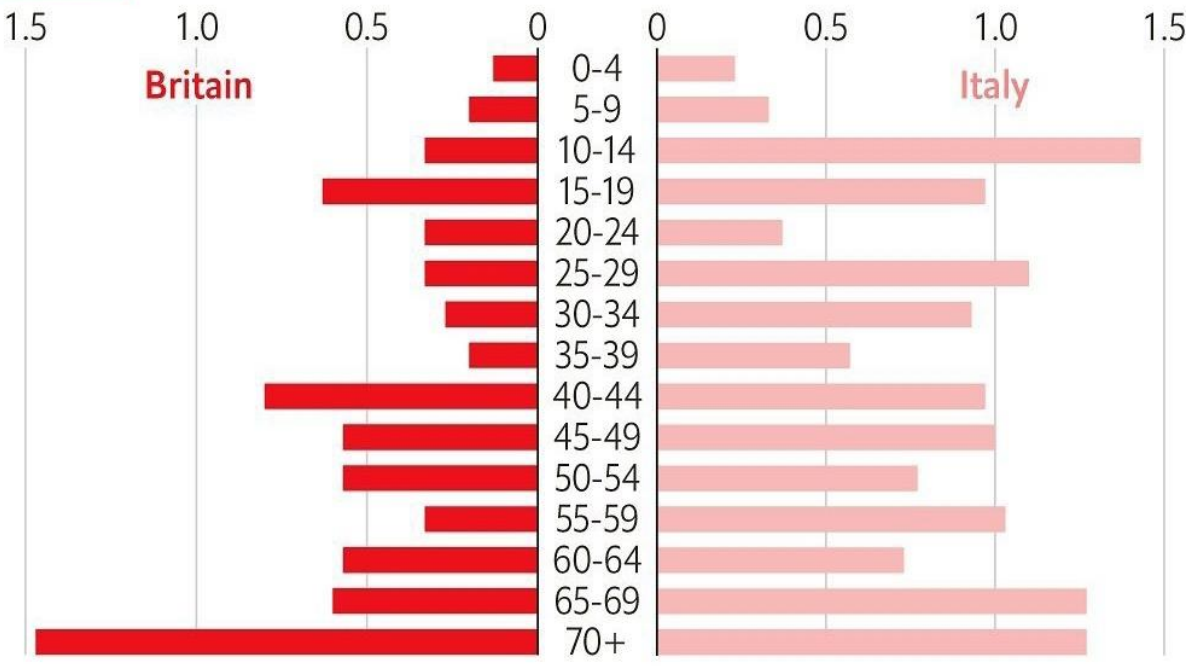


Source: "Social contacts and mixing patterns", J. Mossong et al. **The Economist**

How *The Economist's* data team makes bar charts better

Daily encounters people aged 70+ have with different age groups, average, May 2005-Sep 2006

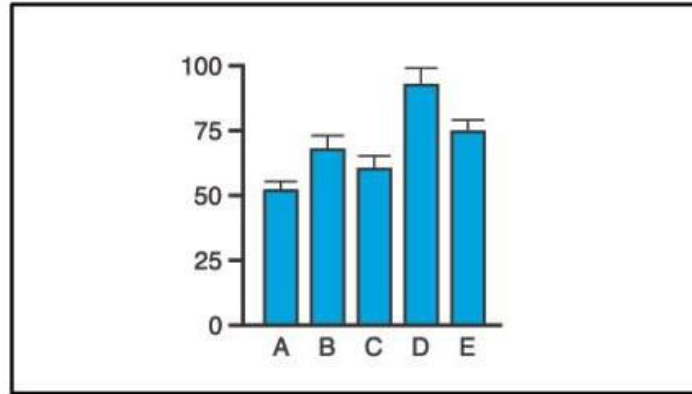
After



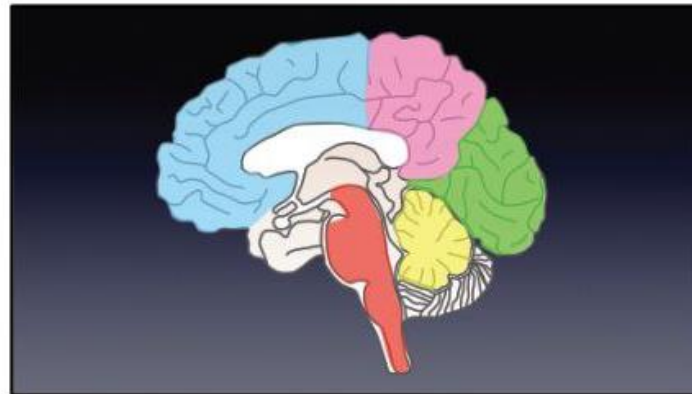
Source: "Social contacts and mixing patterns", J. Mossong et al. **The Economist**

*URL: medium.economist.com/charting-new-territory-7f5afb293270

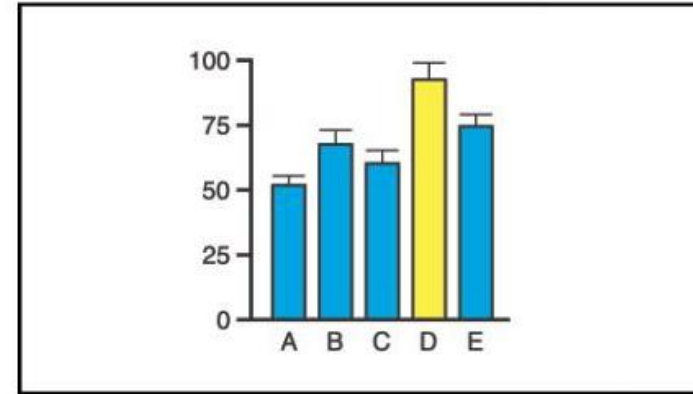
Before



We hypothesized that exposure to ethylene would increase the total number of fruit per tree.



After



We hypothesized that exposure to ethylene would **increase** the total number of fruit per tree.

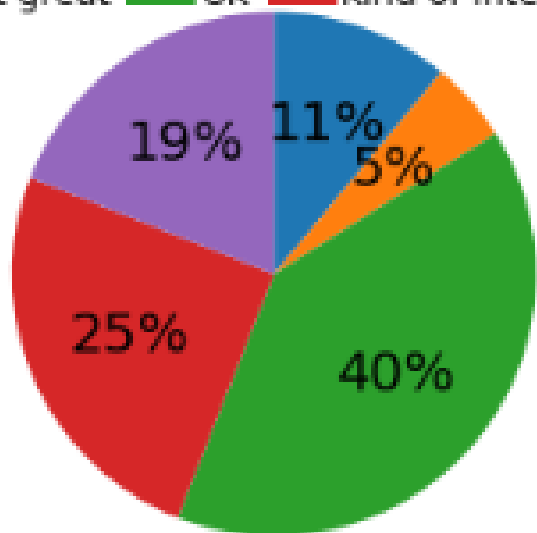


Survey Results

Было

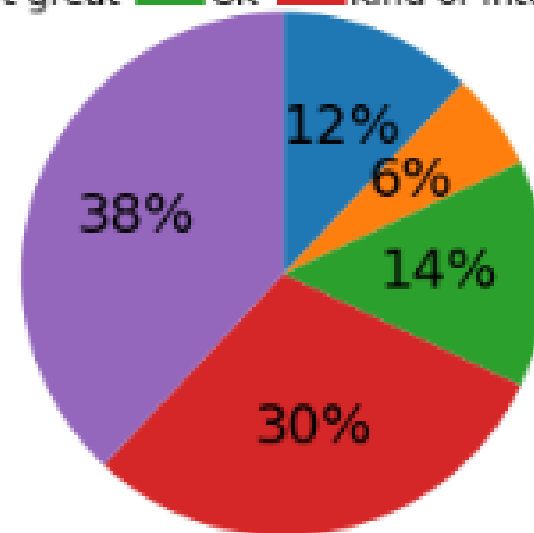
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



Ищем глазами проценты

Трудно сравнивать углы, дуги ещё сложнее

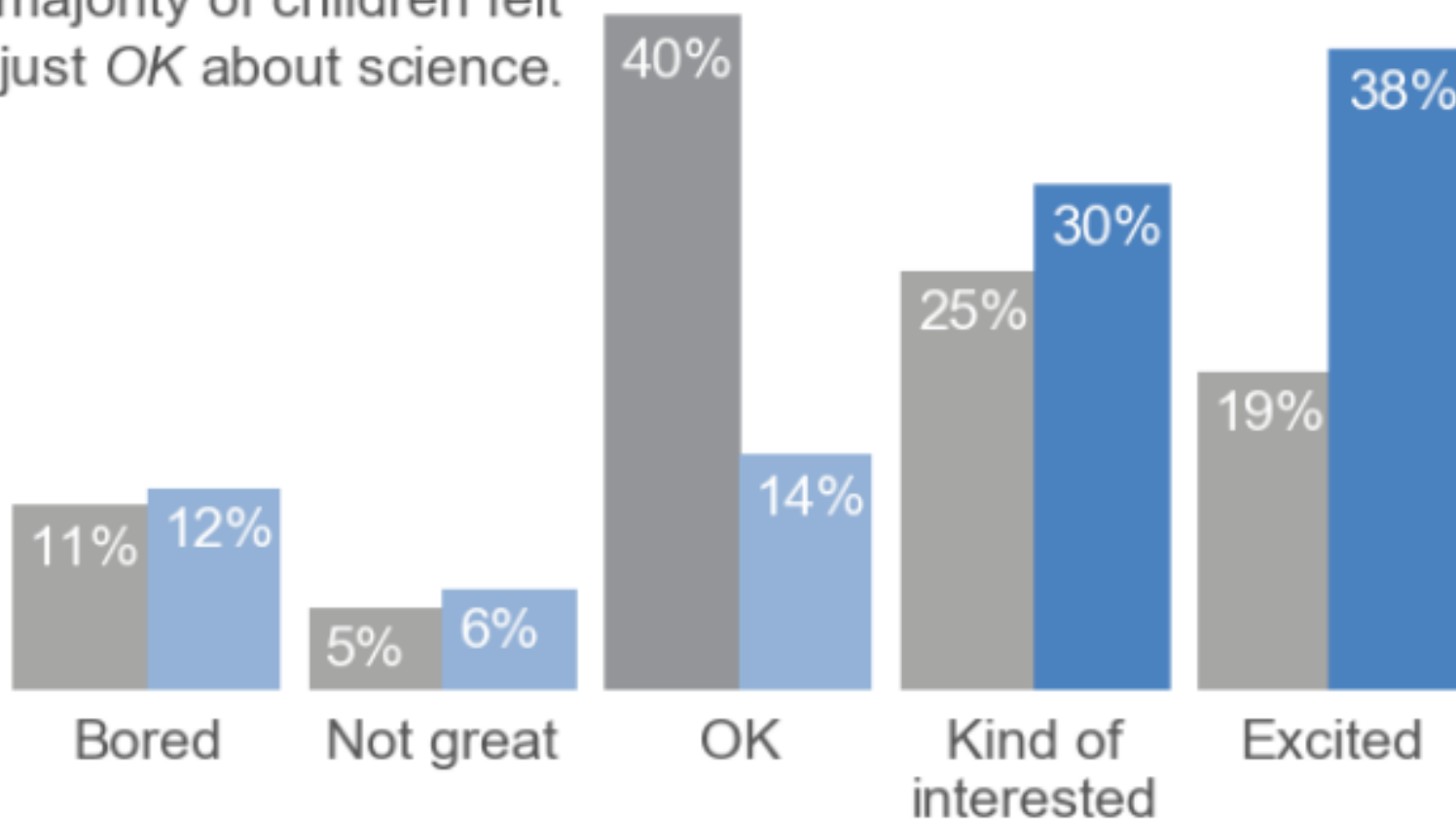
Рассматриваем по часовой стрелке

Pilot program was a success

Стало

How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science.



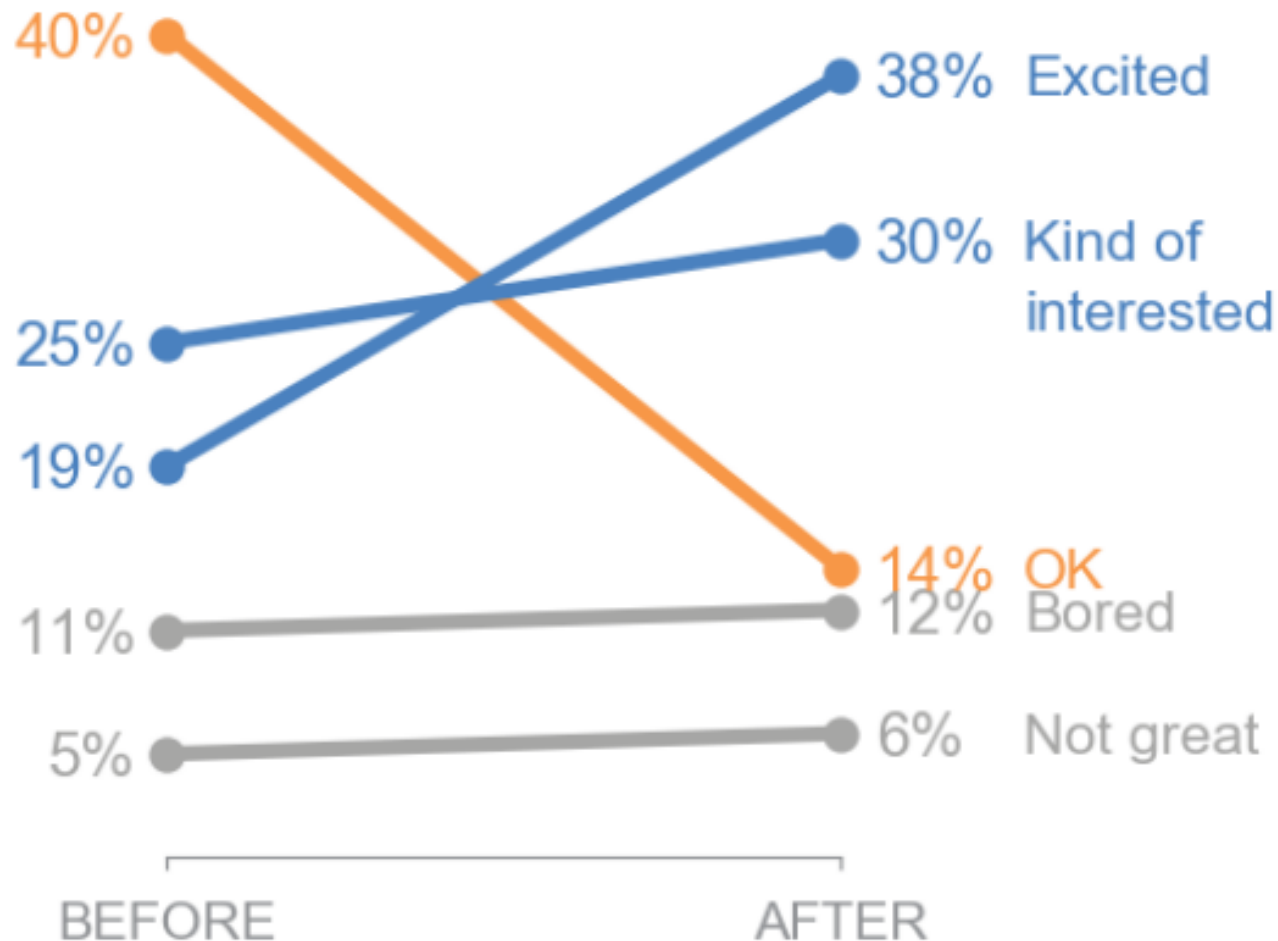
AFTER program, more children were *Kind of interested & Excited* about science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Pilot program was a success

Стало

How do you feel about science?



BEFORE program, the majority of children felt just *OK* about science.

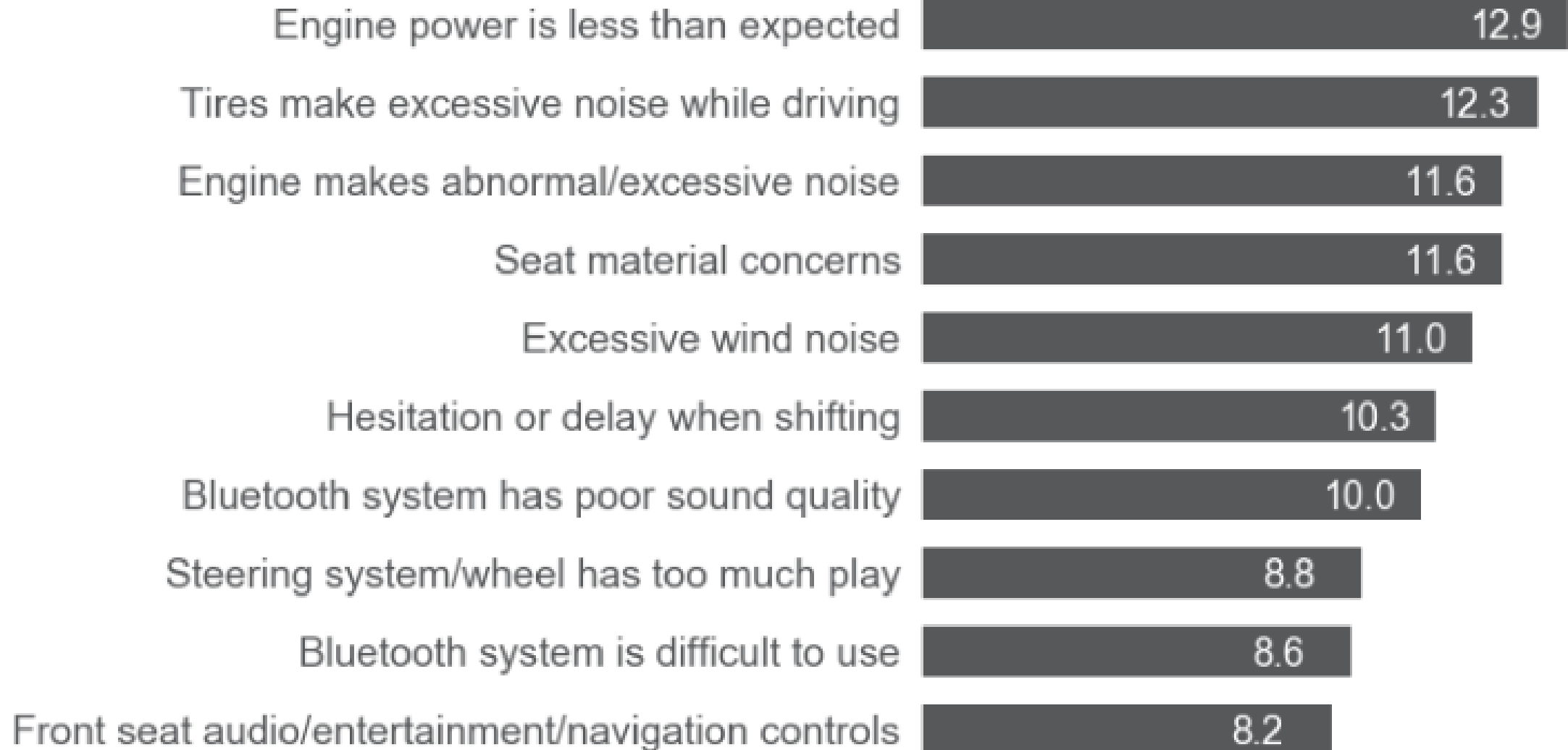
AFTER program, more children were *Kind of interested & Excited* about science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Как получить информацию с этого графика?

Top 10 design concerns

concerns per 1,000



Top 10 design concerns

concerns per 1,000

Engine power is less than expected



12.9

Tires make excessive noise while driving



12.3

Engine makes abnormal/excessive noise



11.6

Seat material concerns



11.6

Excessive wind noise



11.0

Hesitation or delay when shifting



10.3

Bluetooth system has poor sound quality



10.0

Steering system/wheel has too much play



8.8

Bluetooth system is difficult to use



8.6

Front seat audio/entertainment/navigation controls



8.2

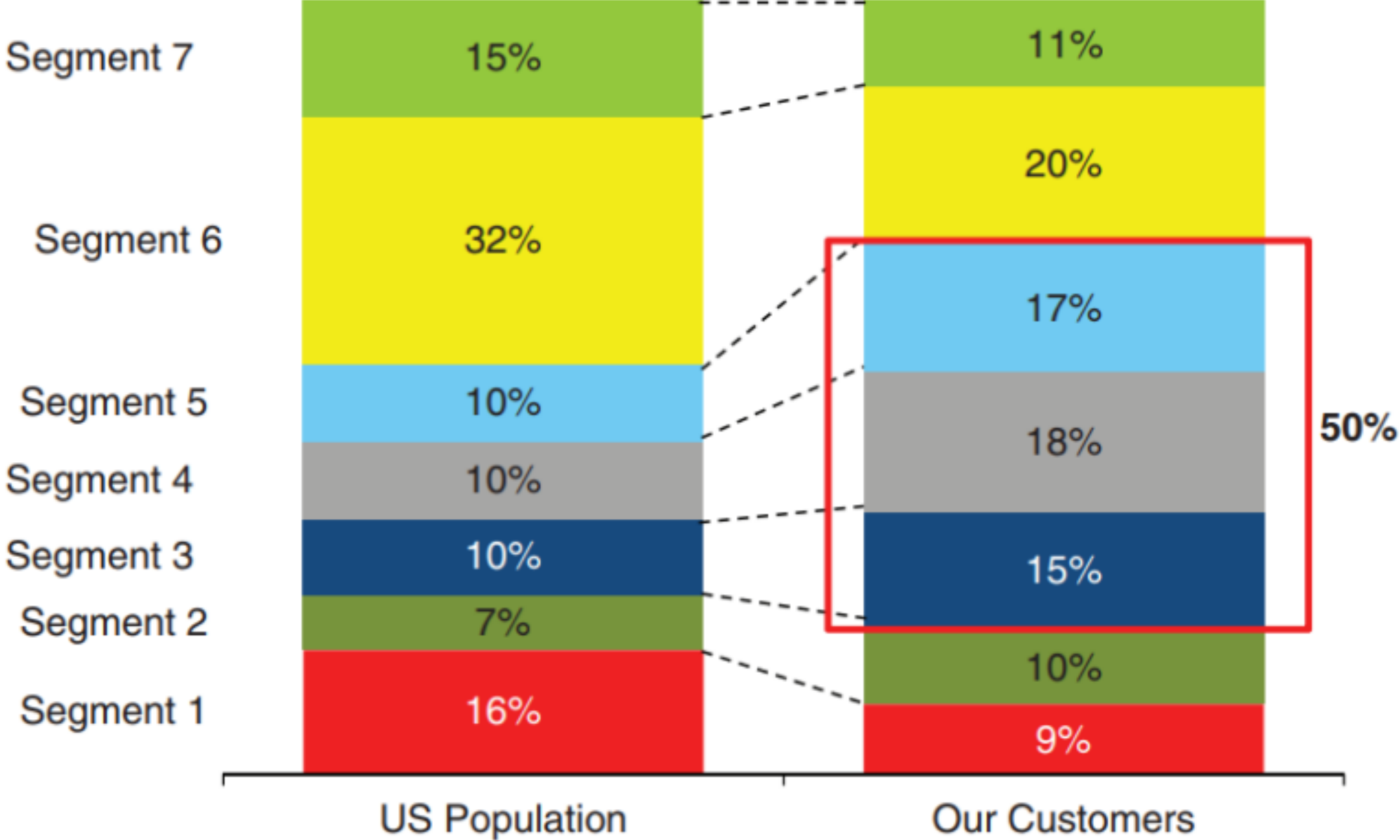
Comments indicate that **noisy tire issues** are most apparent **in the rain**.

Complaints about **engine noise** commonly cited **after the car had not been driven for a while**.

Excessive **wind noise** is noted primarily in **freeway driving at high speeds**.

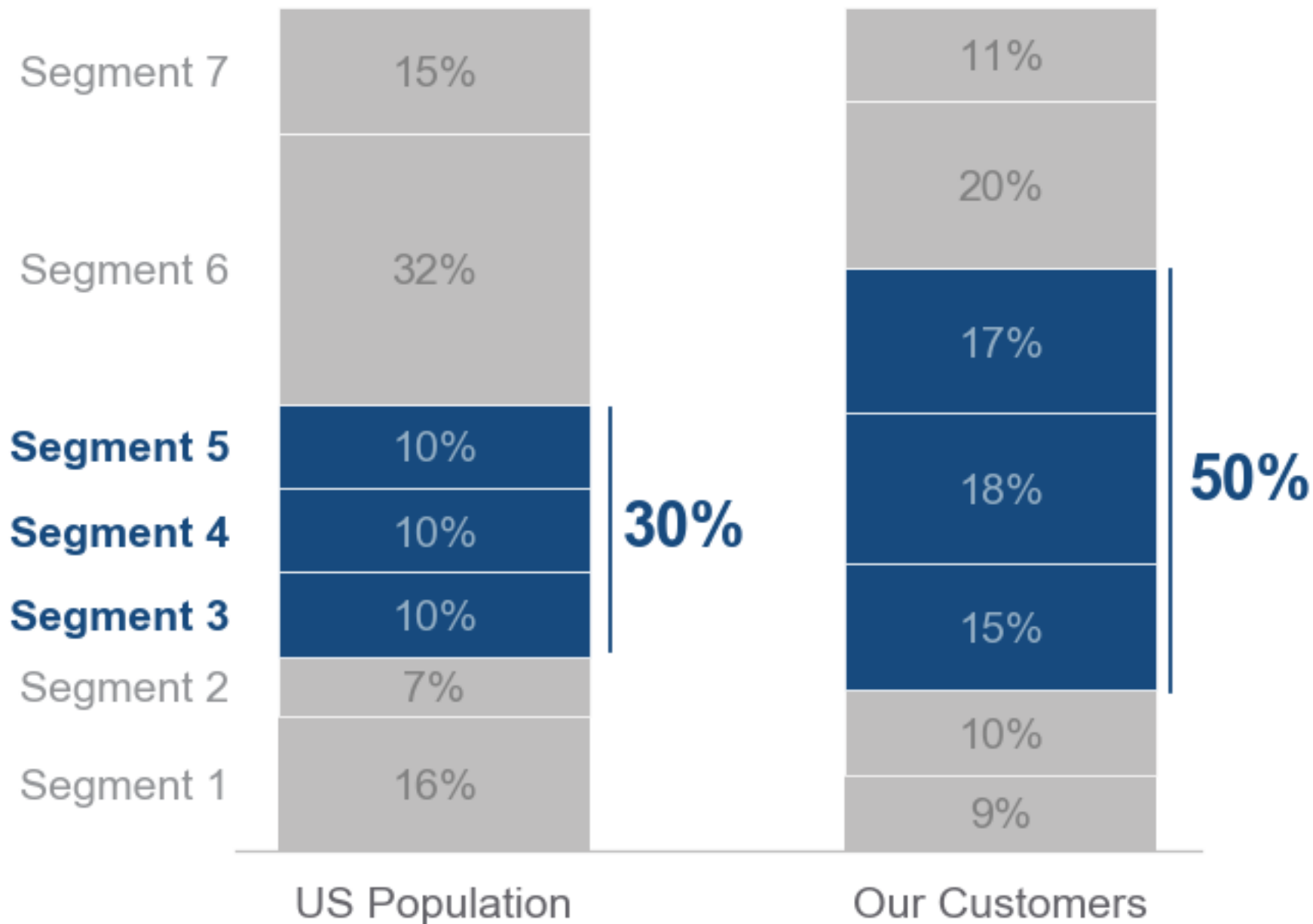
Distribution by customer segment

Было



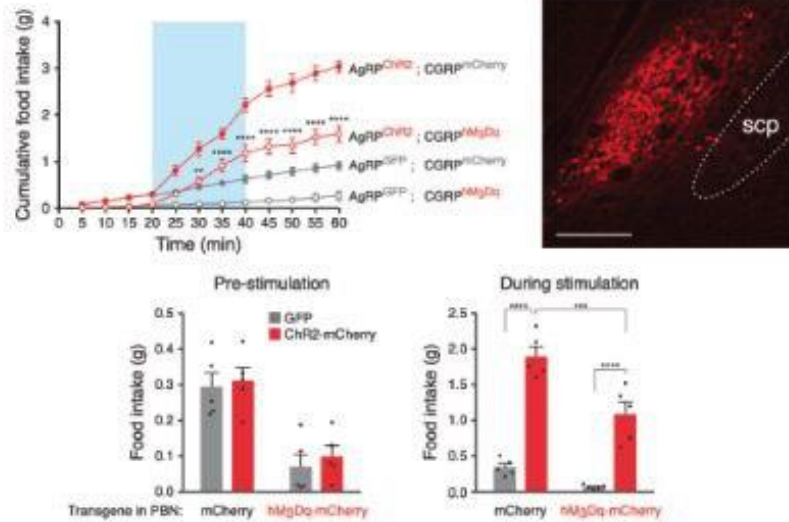
Distribution by customer segment

Стало



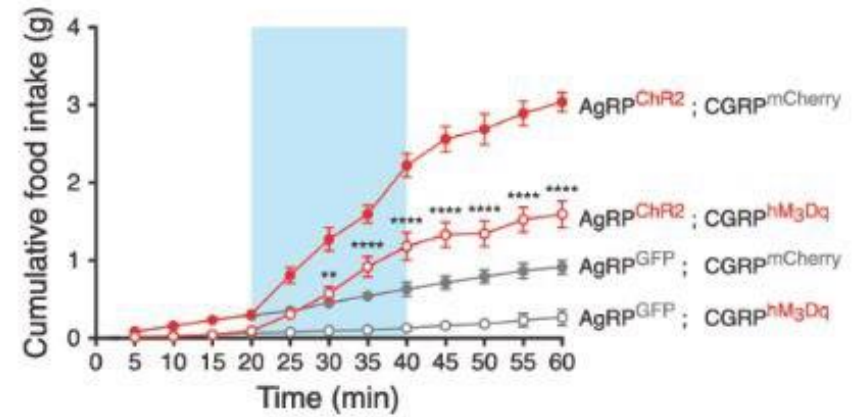
Before

Results



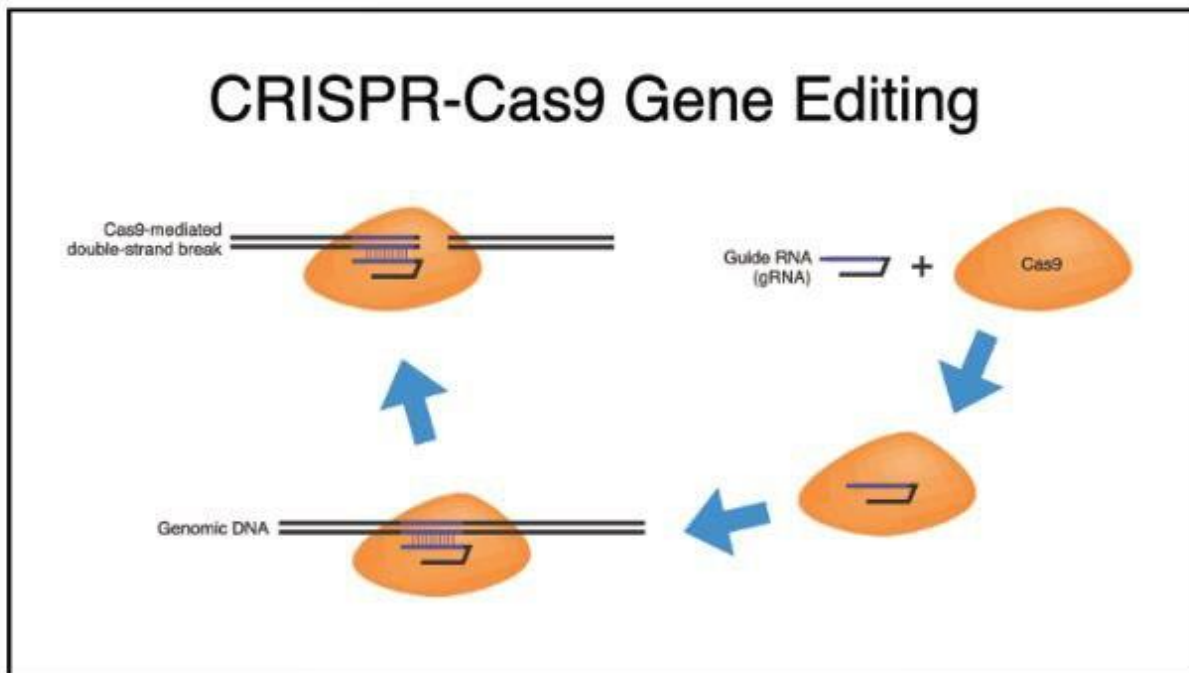
After

Stimulation of AgRP neurons increases feeding behavior

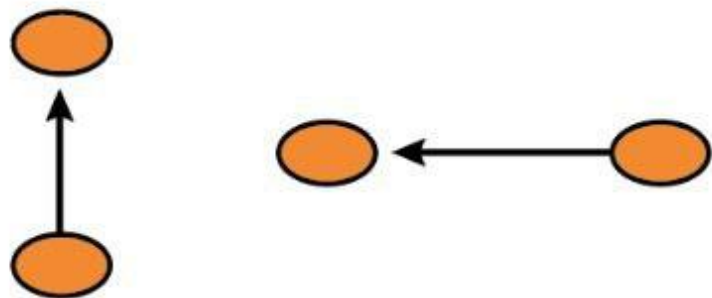
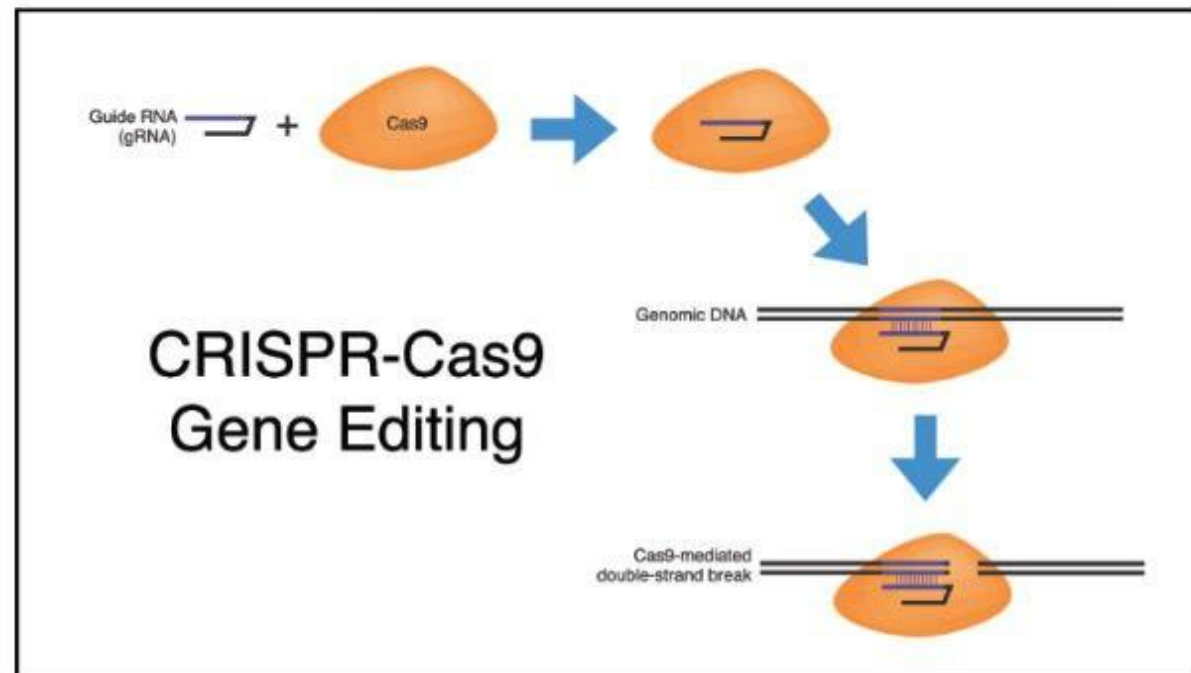


n=6 mice; **p<0.001; ****p<0.0001

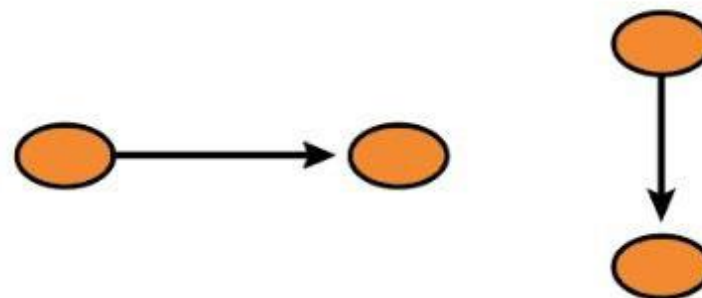
Before



After



Nonintuitive layout



Intuitive layout

Before

Dandelion growth cycle

Bud head



Flower head



Seed head



After

Dandelion growth cycle

Bud head



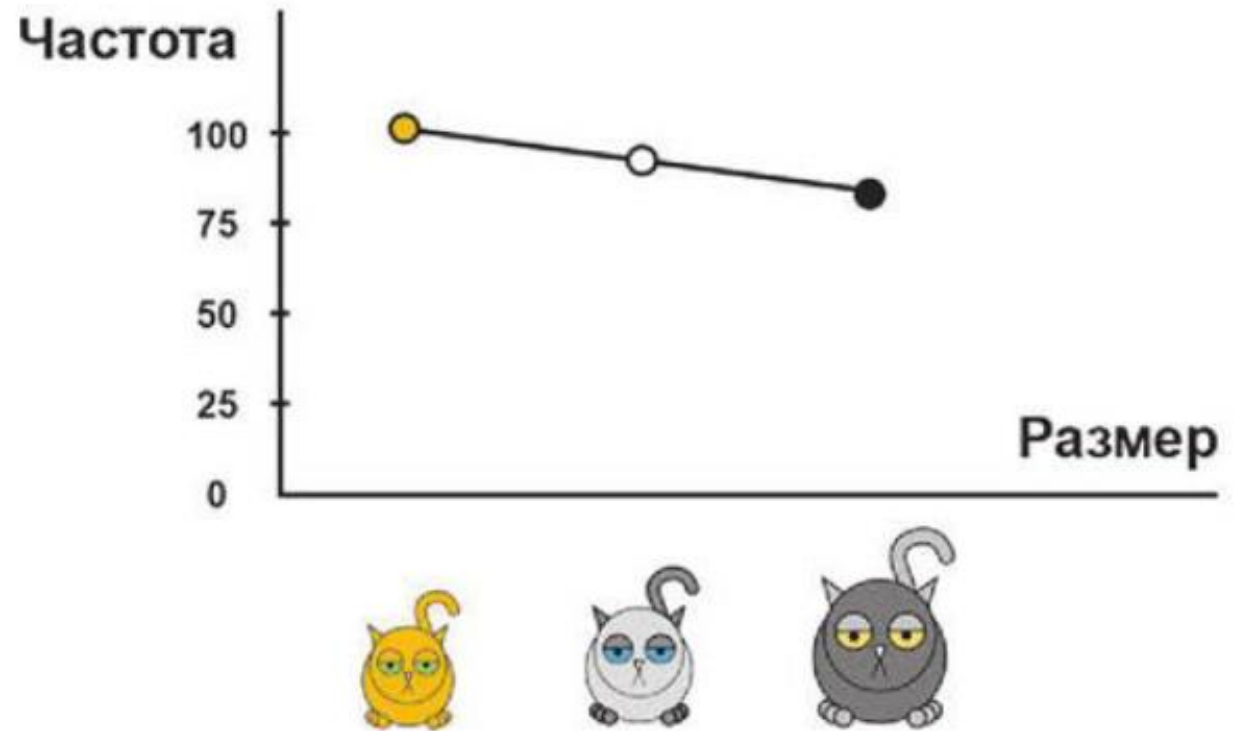
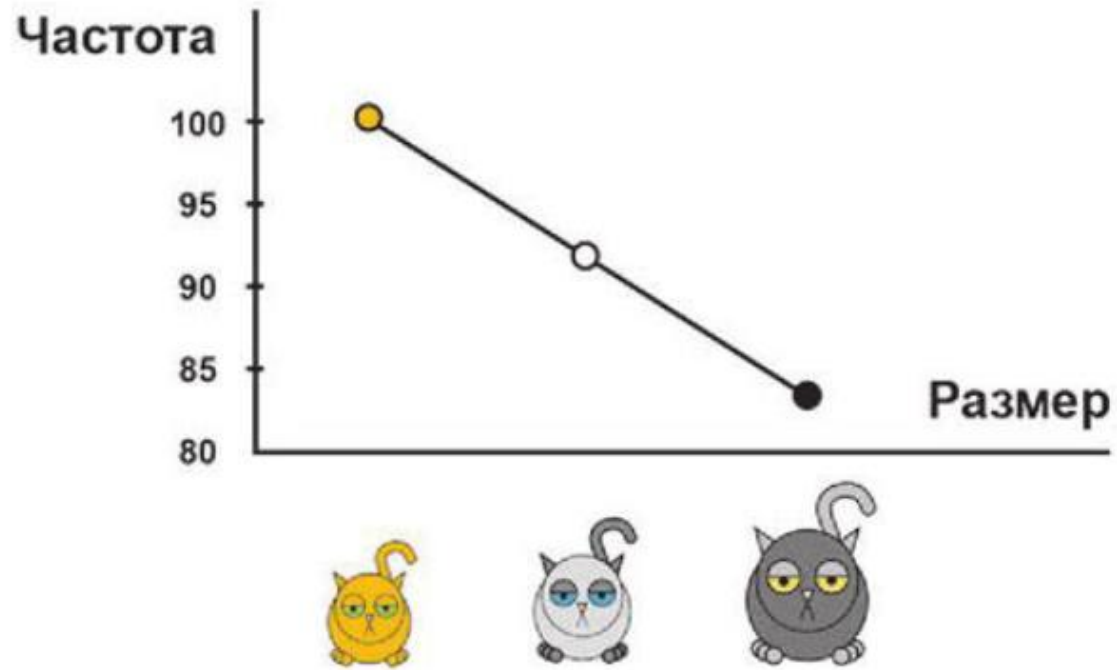
Flower head



Seed head



Трюки



Трюки

