



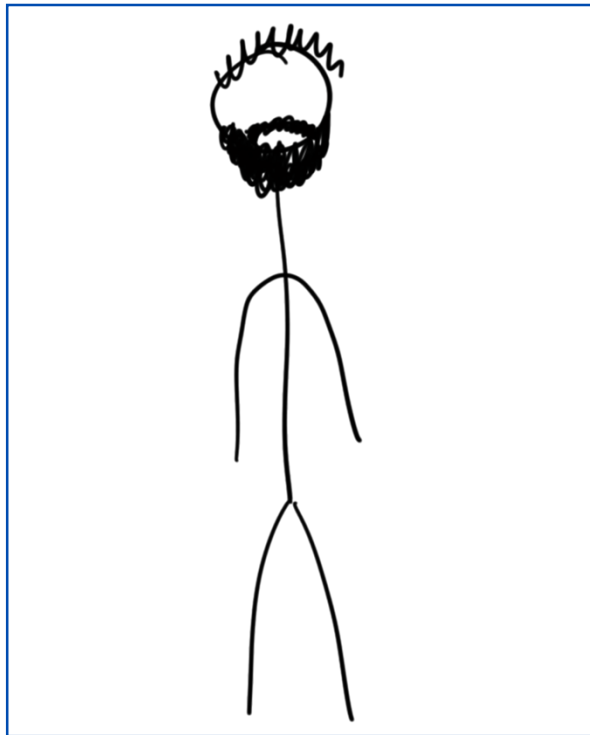
Making a Story From Your Data

Alexander Filatov

14 июня 2018

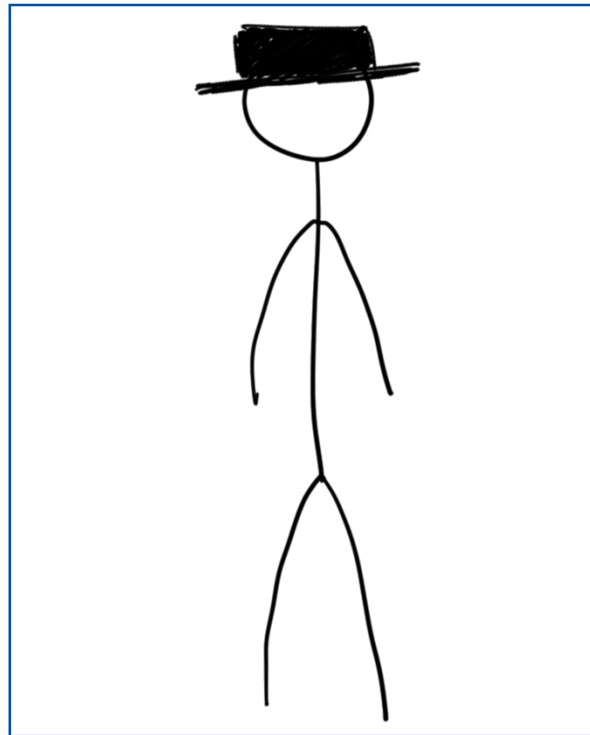
A. Storytelling Intro

We are not naturally good at storytelling in data

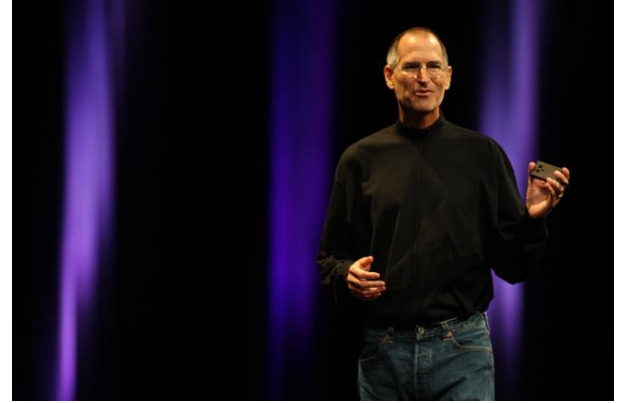
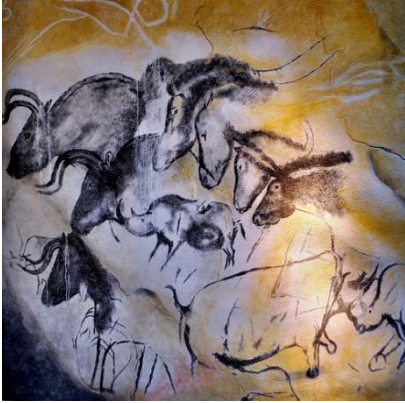


Мы хорошо
решаем проблемы

Другие хорошо
рассказывают
истории



История сказительства (storytelling)



http://bit.ly/jobs_presentation

Storytelling 101



Storytelling 101

Power Corrupts. PowerPoint Corrupts Absolutely.

— Edward Tufte

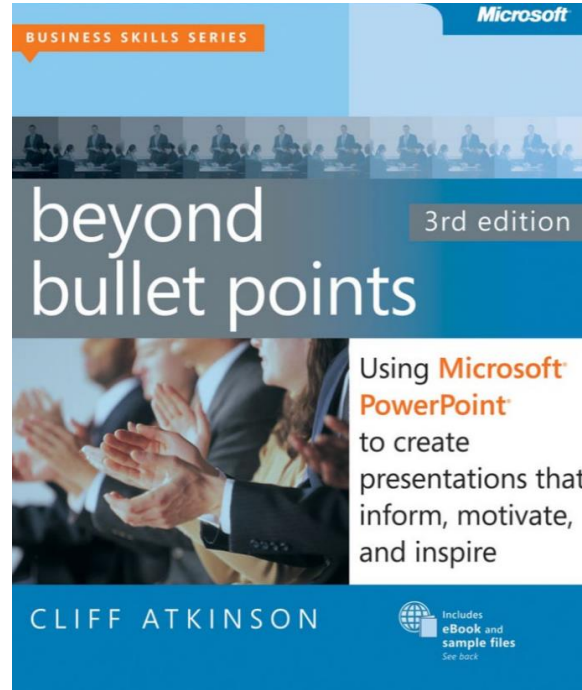
Storytelling 101

1. **Идея / черновик**
2. **Раскадровка**
3. **Материал (слайды, речь)**

Lessons in storytelling



Lessons in storytelling

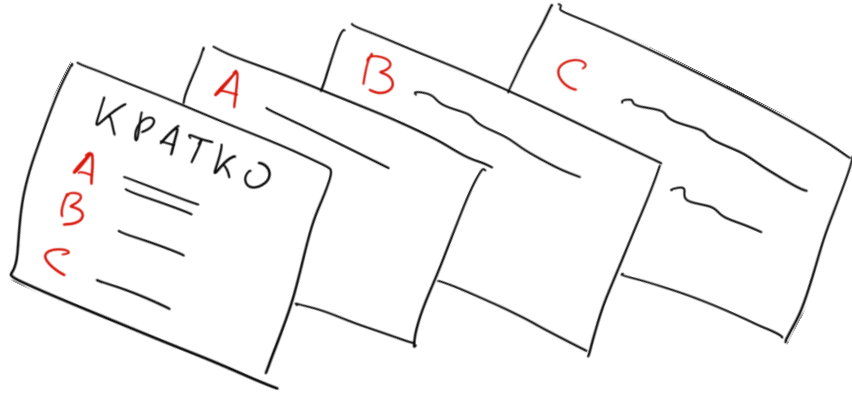


http://bit.ly/beyond_bullet_points

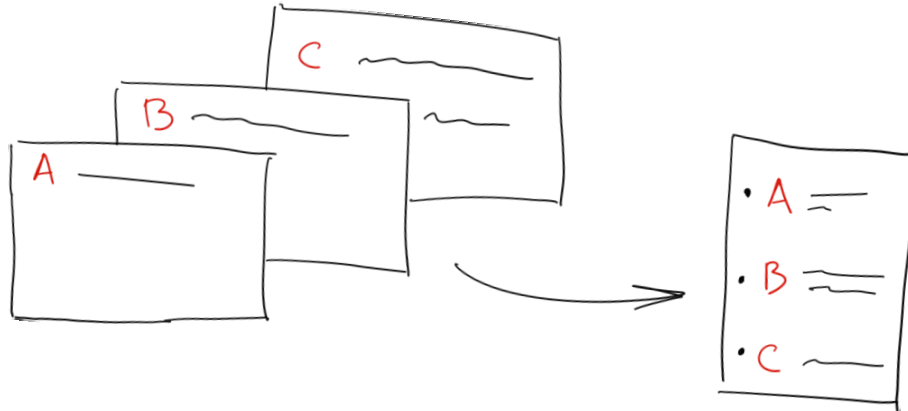
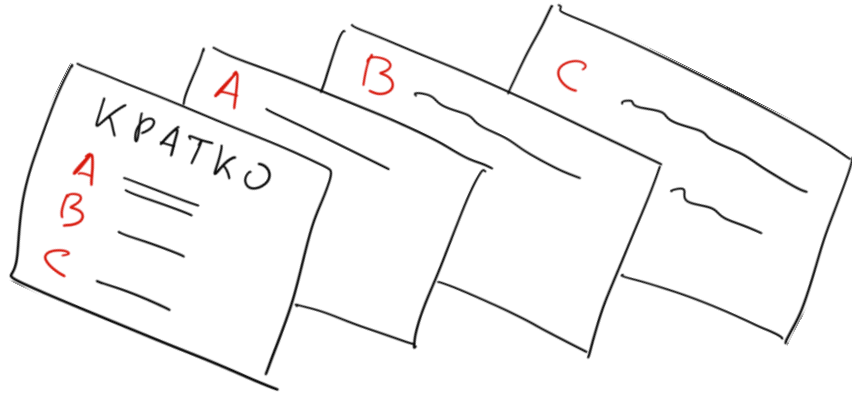
Lessons in storytelling



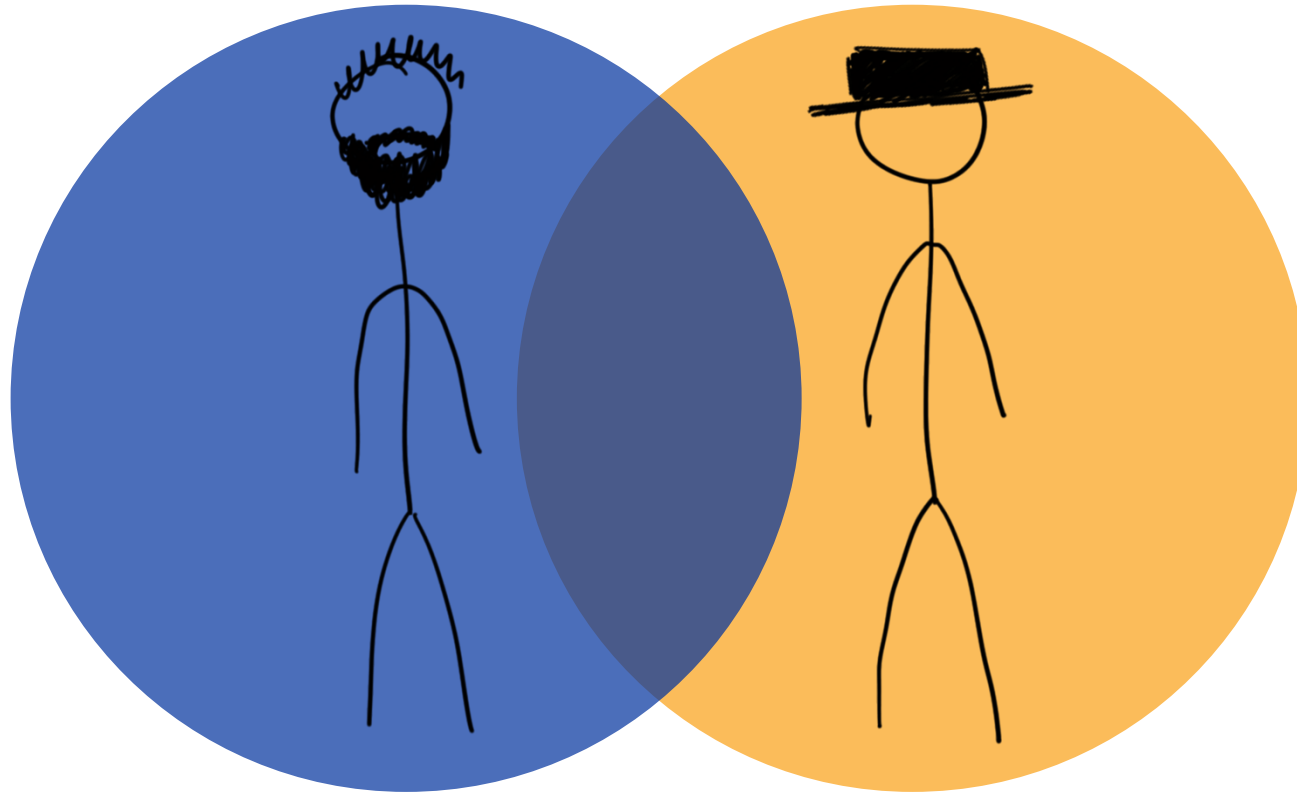
Lessons in storytelling



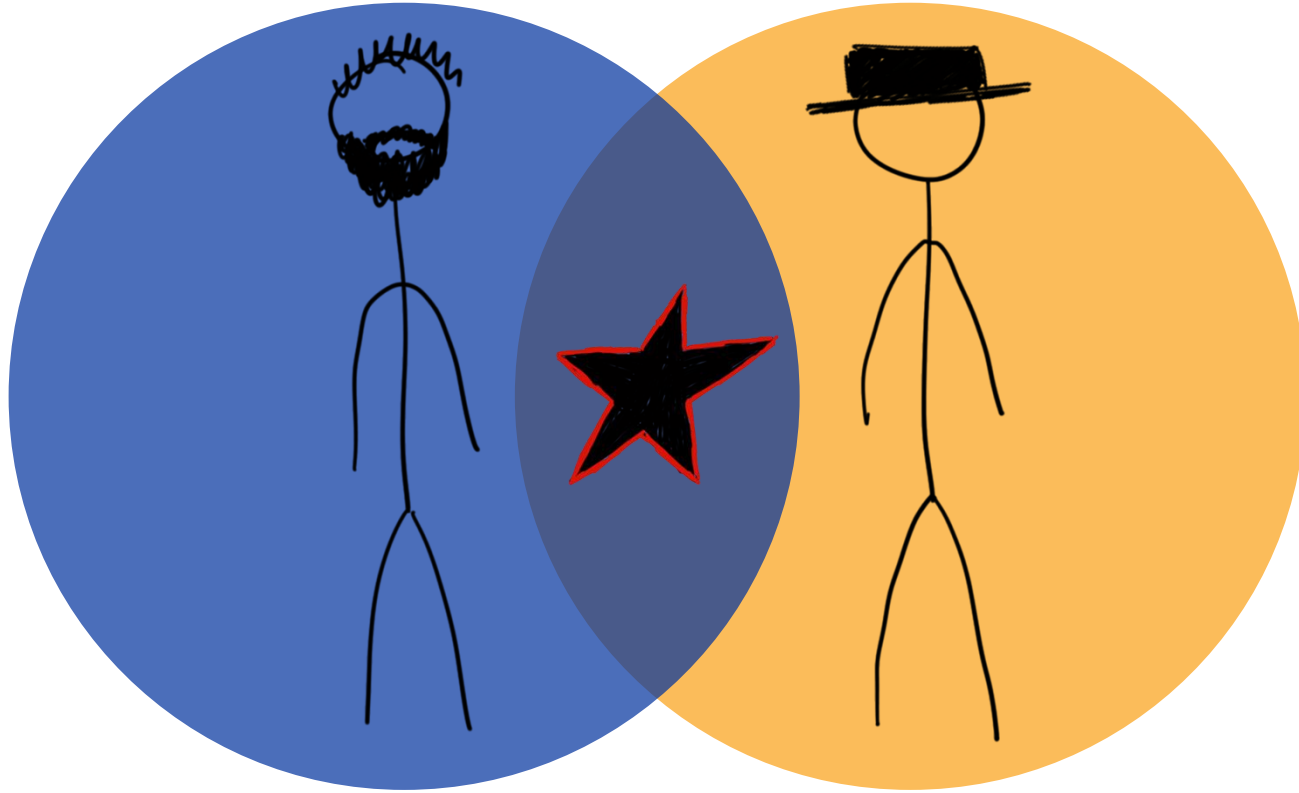
Lessons in storytelling



My Story



My Story



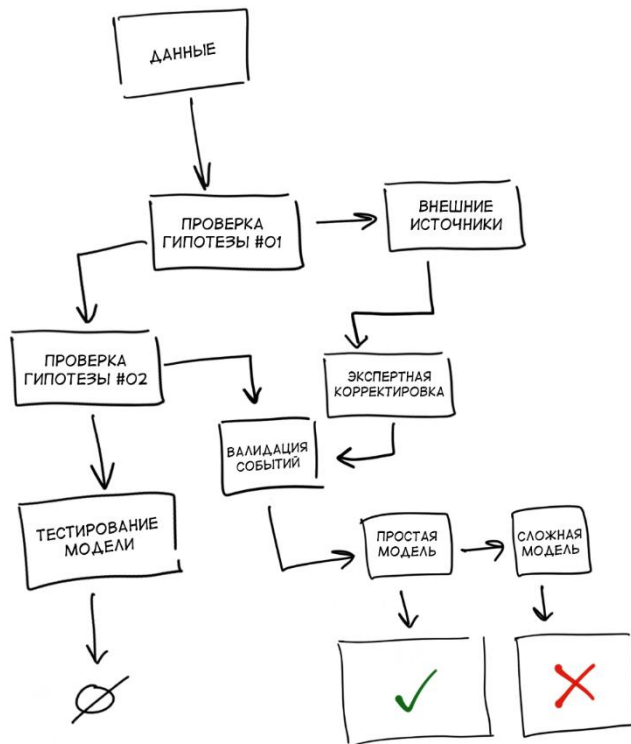
B. Storytelling with Data

План

1. Важность контекста
2. Эффективная визуализация
3. Проблемы от хаоса
4. Контроль внимания
5. Дизайнерские приёмы

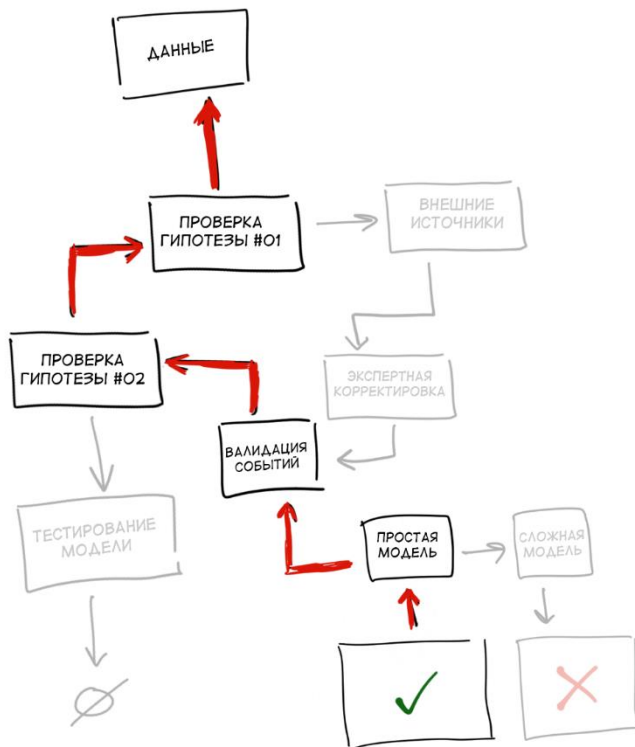
Контекст важен — два вида рассказа

Аналитик делает исследование (exploratory)

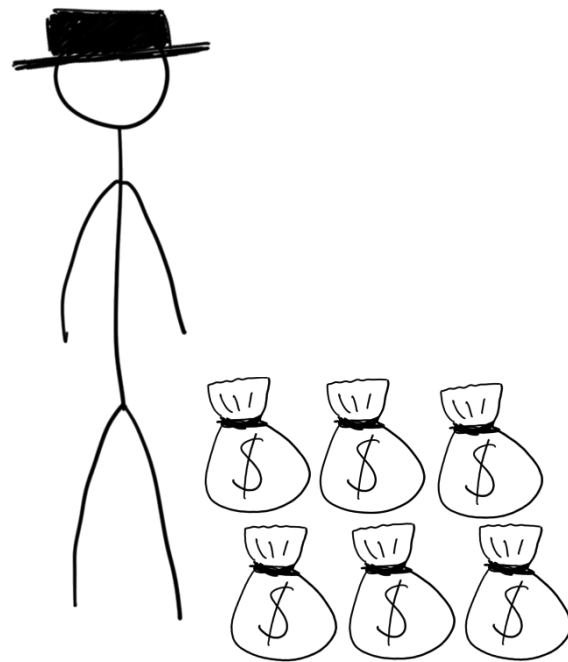


Контекст важен — два вида рассказа

Аналитик делает исследование (exploratory)



Бизнесу нужно объяснение (explanatory)



Контекст важен — два вида рассказа

Exploratory

http://bit.ly/exploratory_nobel

By Bill Shander

http://bit.ly/exploratory_ashland

By Michael Barry

Explanatory

http://bit.ly/exlanatory_boston

© BeehiveMedia

By Bill Shander

Контекст важен — кто, что и как (3W)?

Кто (Who)

Аудитория

- Все заинтересованные;
- Сотрудники компании;
- Члены бюджетного комитета, которые принимают решение о продлении контракта;
- Коллеги с 8ого этажа, пользующиеся кухней.

Подсказка: кто принимает решение (decision maker)?

Ты сам

- Взаимоотношение с аудиторией;
- Связанность с данными.

Что (What)

Действие

- Что аудитория должна сделать?
- Призыв к действию наталкивает на правильные вопросы и дискуссию.

Подсказка: почему их должно заботить то, что вы рассказываете?

Механизм

- Живая презентация;
- Отправленная почтой презентация.

Как (How)

Тон

- Празднование успеха?
- Серьёзная и важна тема?
- Завести аудиторию на немедленное действие?

Данные

- Последний вопрос!

Подсказка: какие данные доступны, чтобы доказать (проверить) моё утверждение?

Контекст важен — ещё 3W!

When

Время/ситуация коммуникации

- Одно и то же сообщение может быть воспринято совсем по-разному в зависимости от обстоятельств;
- Ситуация может как помогать, так и вредить докладу.

Подсказка: как текущие обстоятельства влияют восприятие коммуникации целевой аудиторией?

Where

Какое медиа используется

- On-line / off-line сильно влияет на контент и внимание;
- Презентация — не всегда лучший способ (альтернативы: демо, хакатон);
- Сильно зависит от аудитории и контента.

Подсказка: как выбранное медиа помогает в донесение основного сообщения до целевой аудитории?

Why

Предпосылка к коммуникации

- Проверка предыстории запланированного общения;
- Настройка докладчика на аудиторию, предпосылку и цель;
- Последняя проверка, уместны ли все предыдущие Ws.

Подсказка: насколько уместна данная коммуникация?

Контекст важен — и ещё 3W!

What is Wanting? (missings)

Данных всегда не хватает, есть только 3 выхода:

- Сохранять спокойствие

Лучше явно заявить об ограниченности данных в рассказе, это может дать развитие новому проекту.

- Возврат к источнику

Например, заказчик готов предоставить дополнительных данных, это поможет сделать историю законченной.

- Собрать больше данных

What in the World?

Сравнение — упрощает понимание

- Данные из внешних источников, исследования на смежные темы

Актуальные исследования, которые подтверждают гипотезу (или просто целостность данных).

- Добавить контекста, привести бэнчмарков

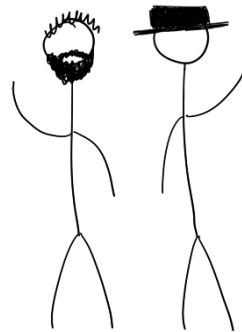
Лучшие мировые практики, опыт других клиентов.

- Сделать завершённую картину

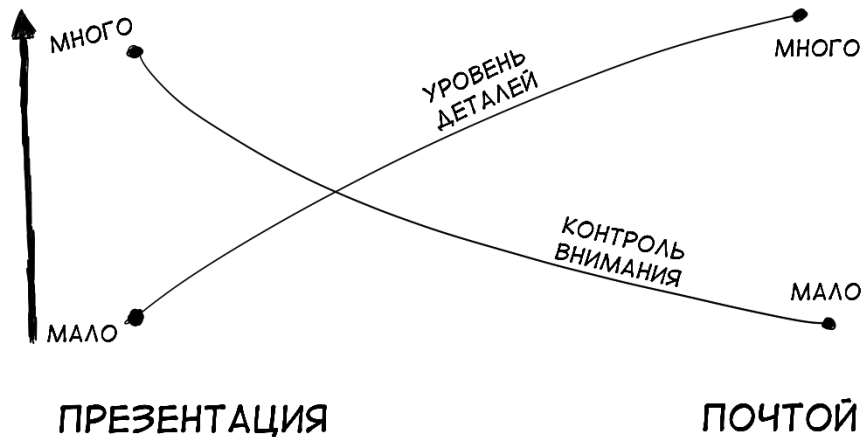
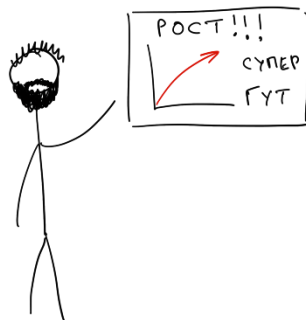
What is Wild?

Что-то, что будет напоминать о рассказе

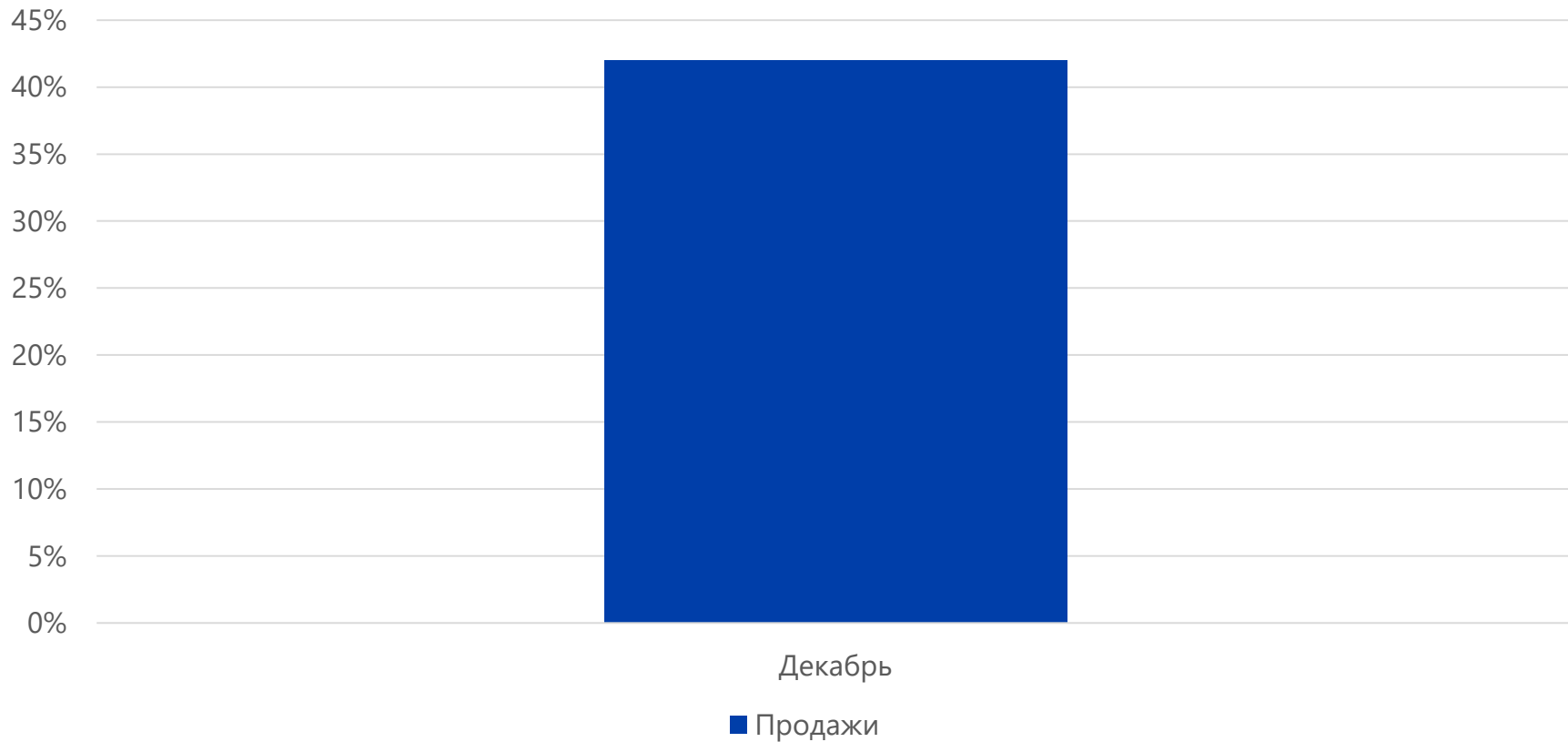
- Особенный визуальный подход
- Неожиданные контекст и выбор данных
- Интерактивность



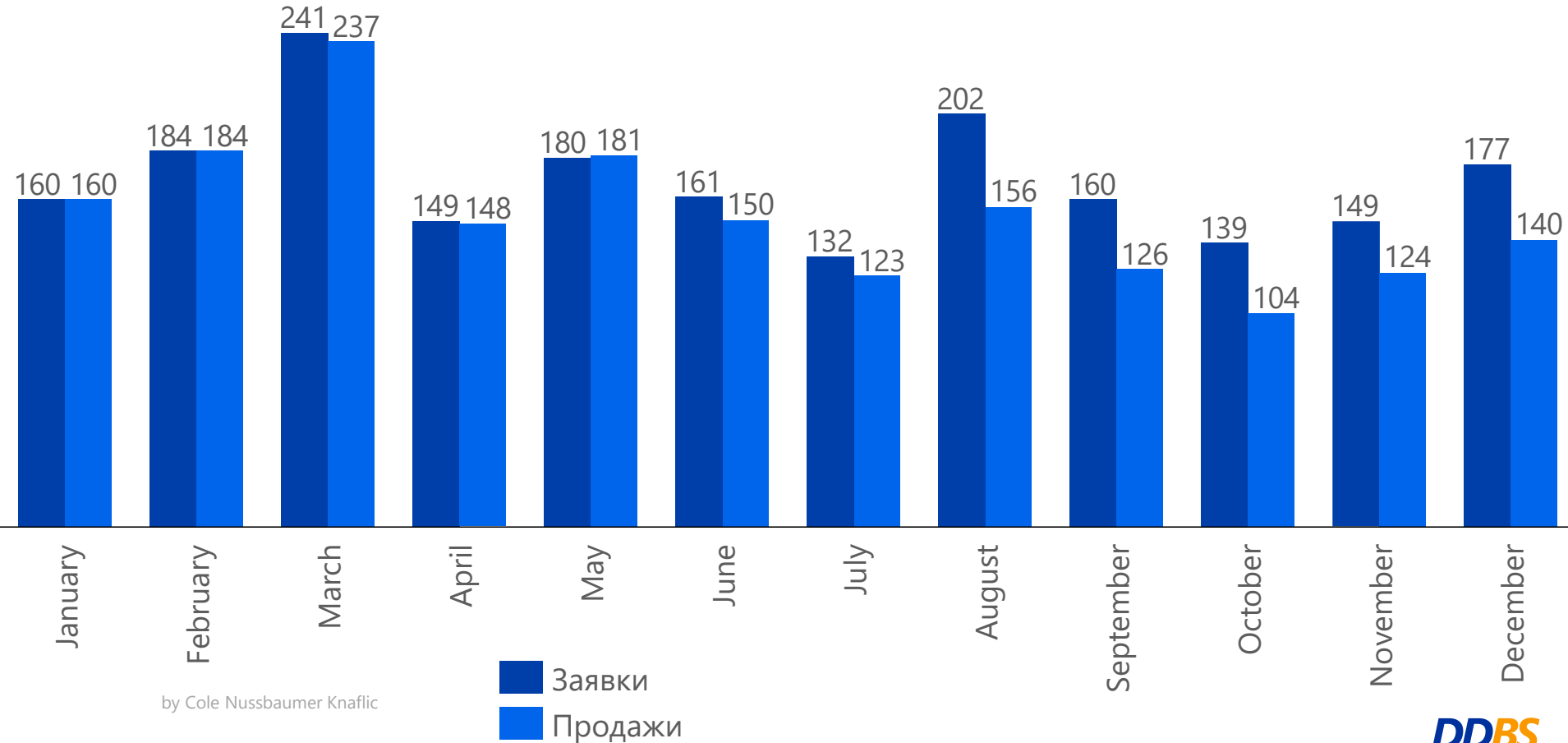
Эффективная визуализация



Динамика роста

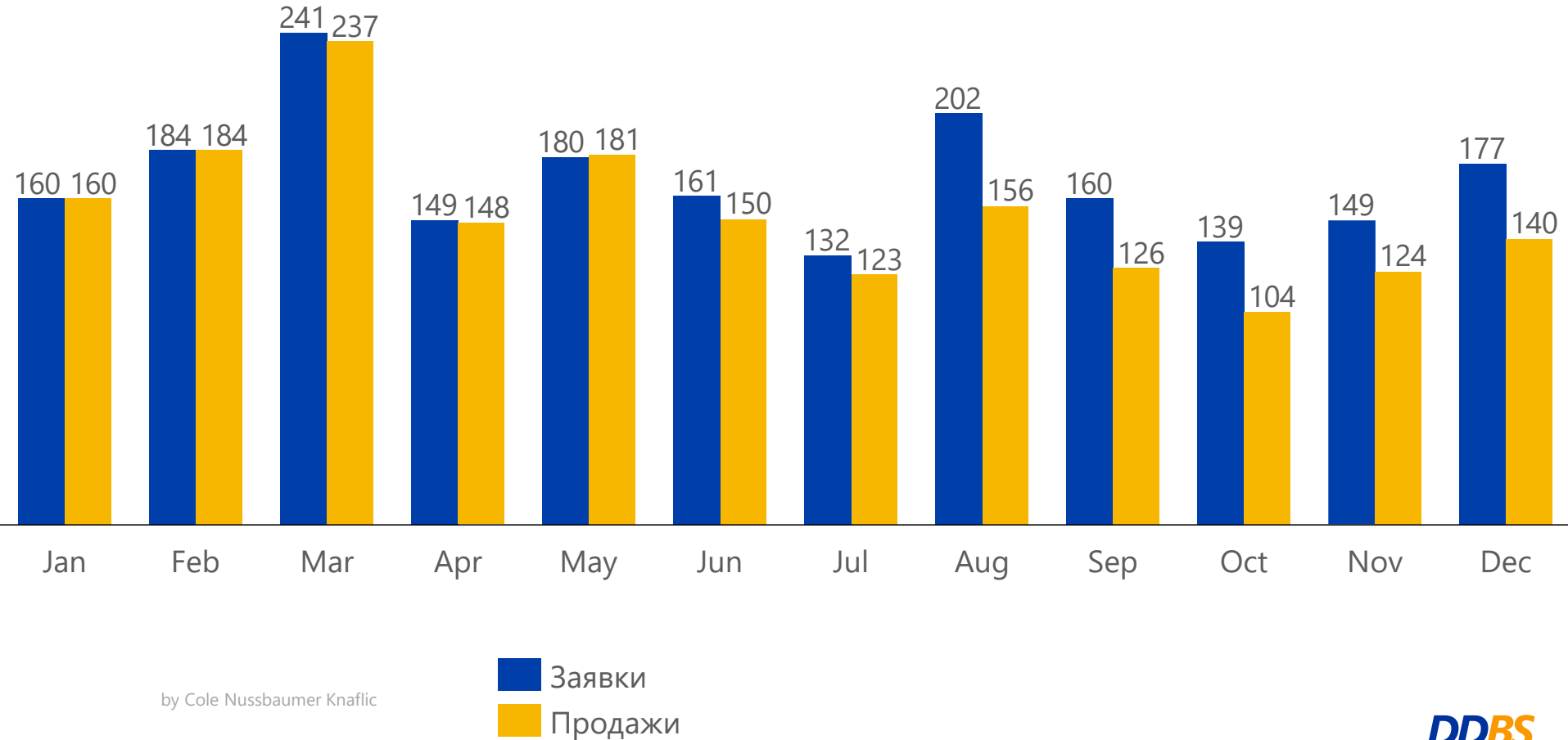


Эффективная визуализация

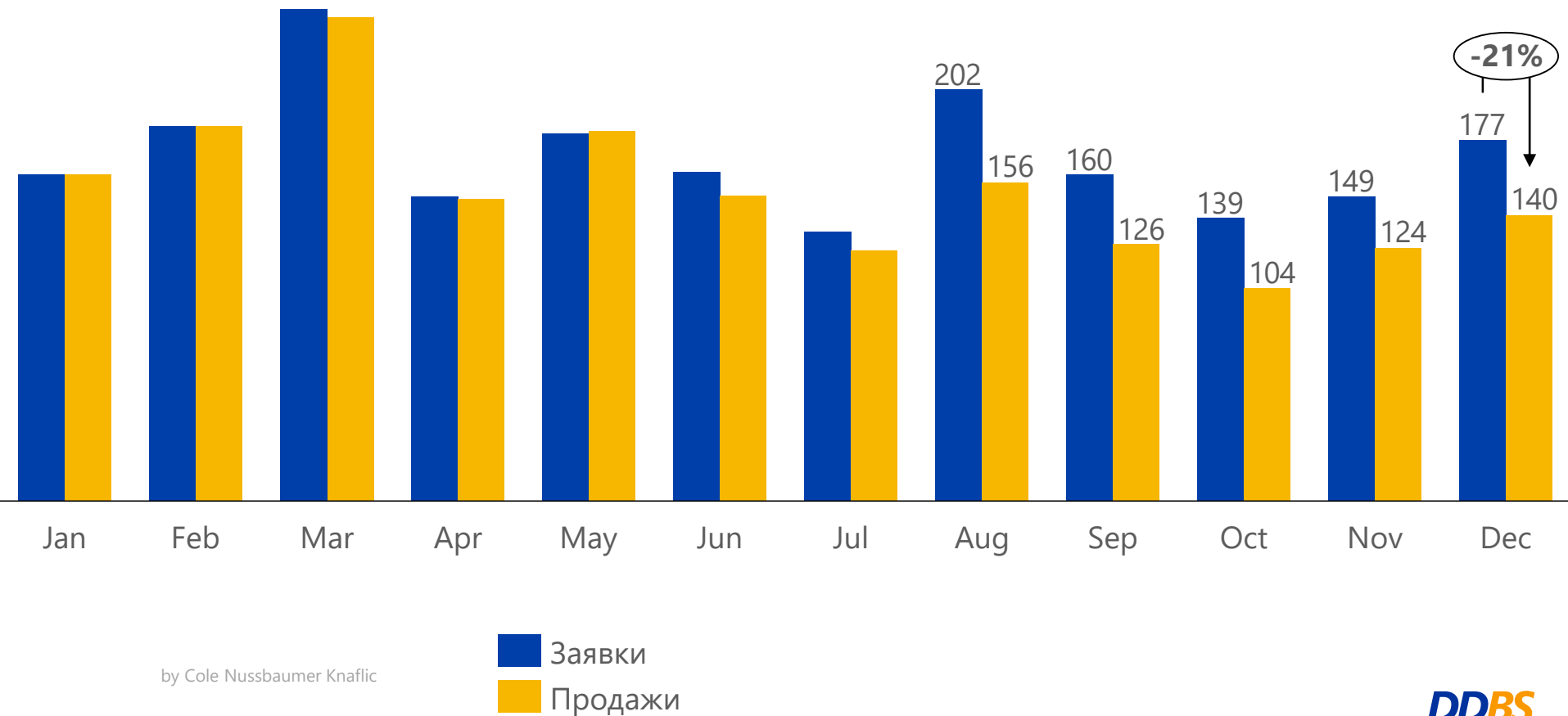


by Cole Nussbaumer Knaflic

Эффективная визуализация

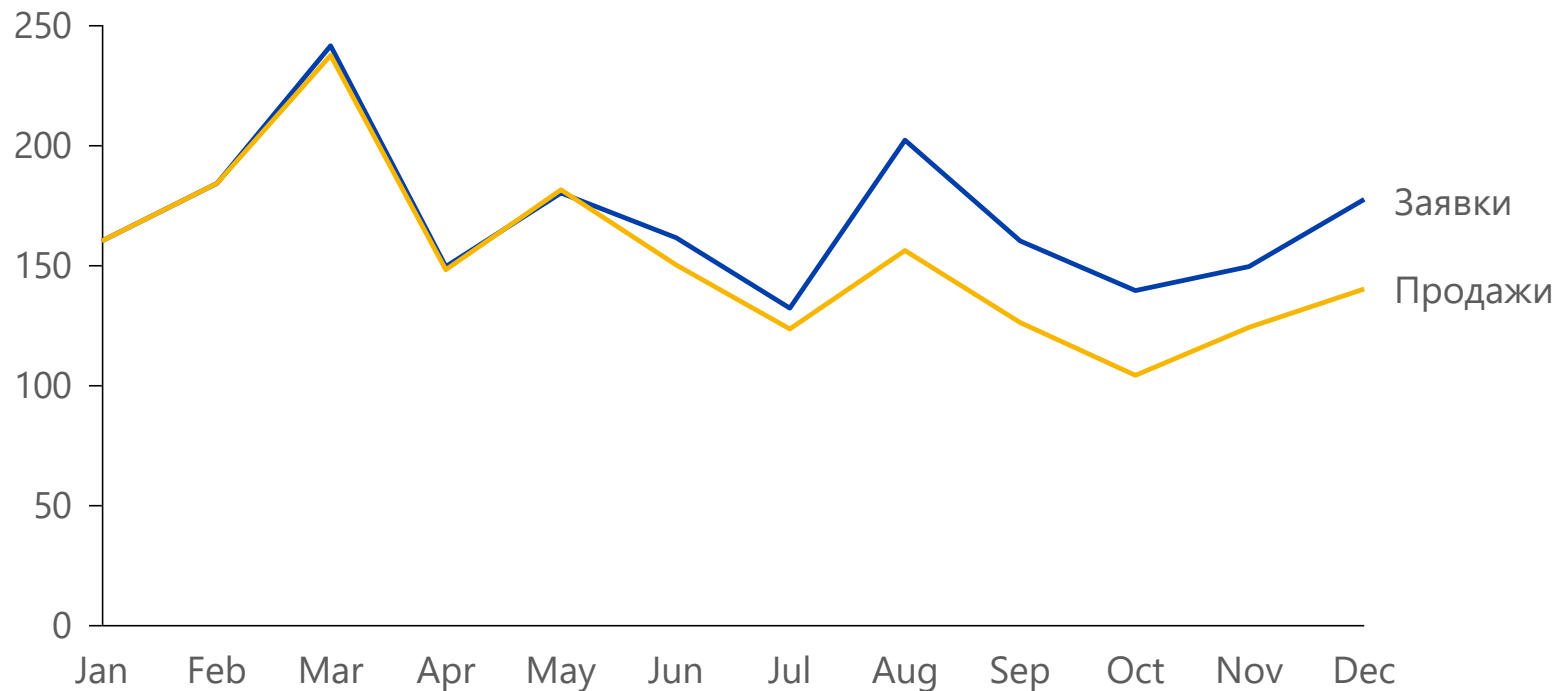


Эффективная визуализация



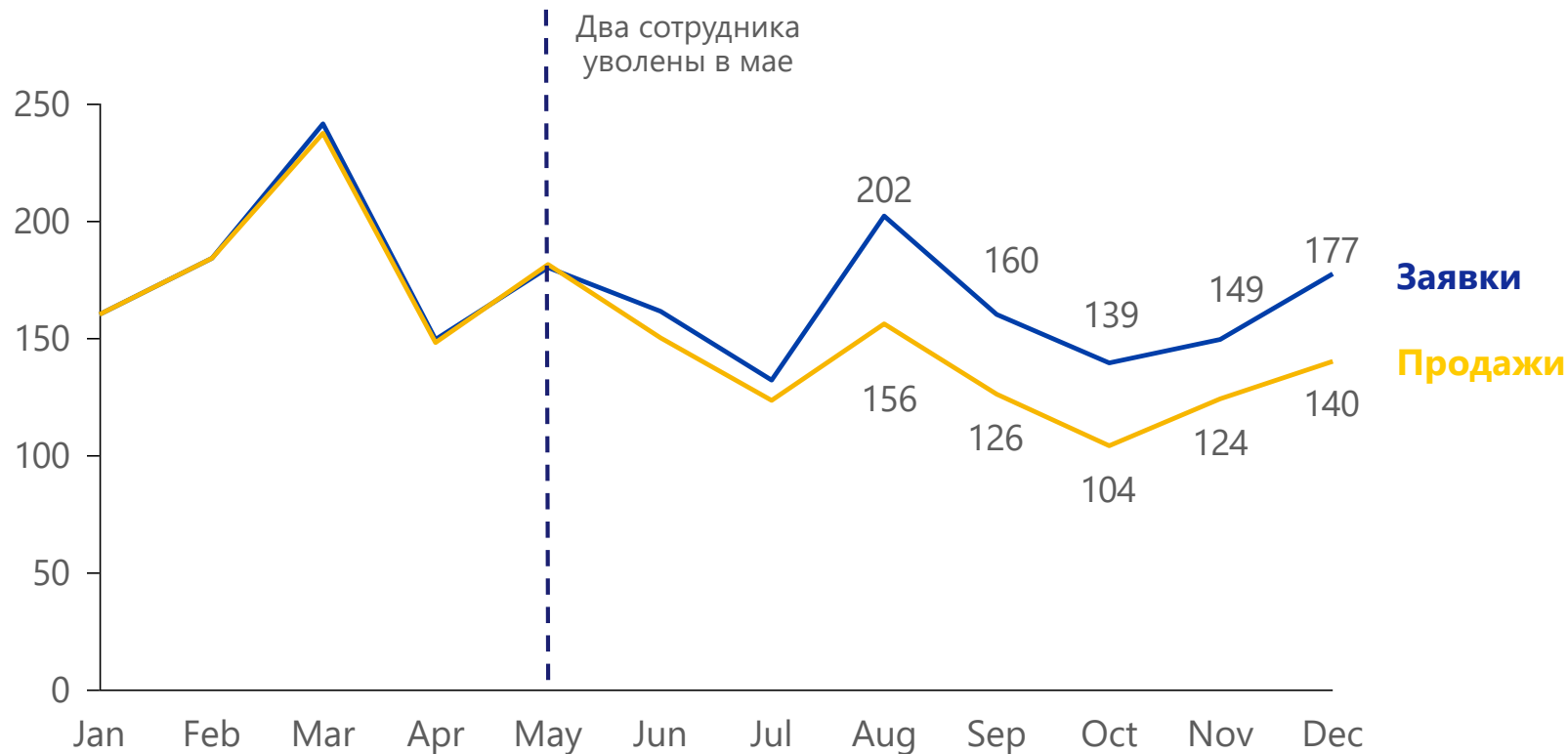
by Cole Nussbaumer Knaflic

Эффективная визуализация



by Cole Nussbaumer Knaflic

Эффективная визуализация

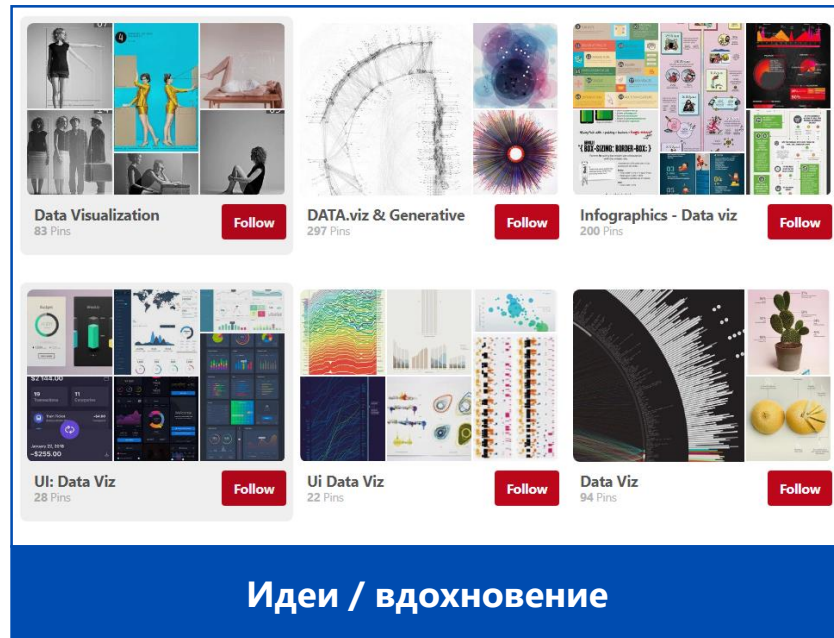


by Cole Nussbaumer Knaflic

Эффективная визуализация

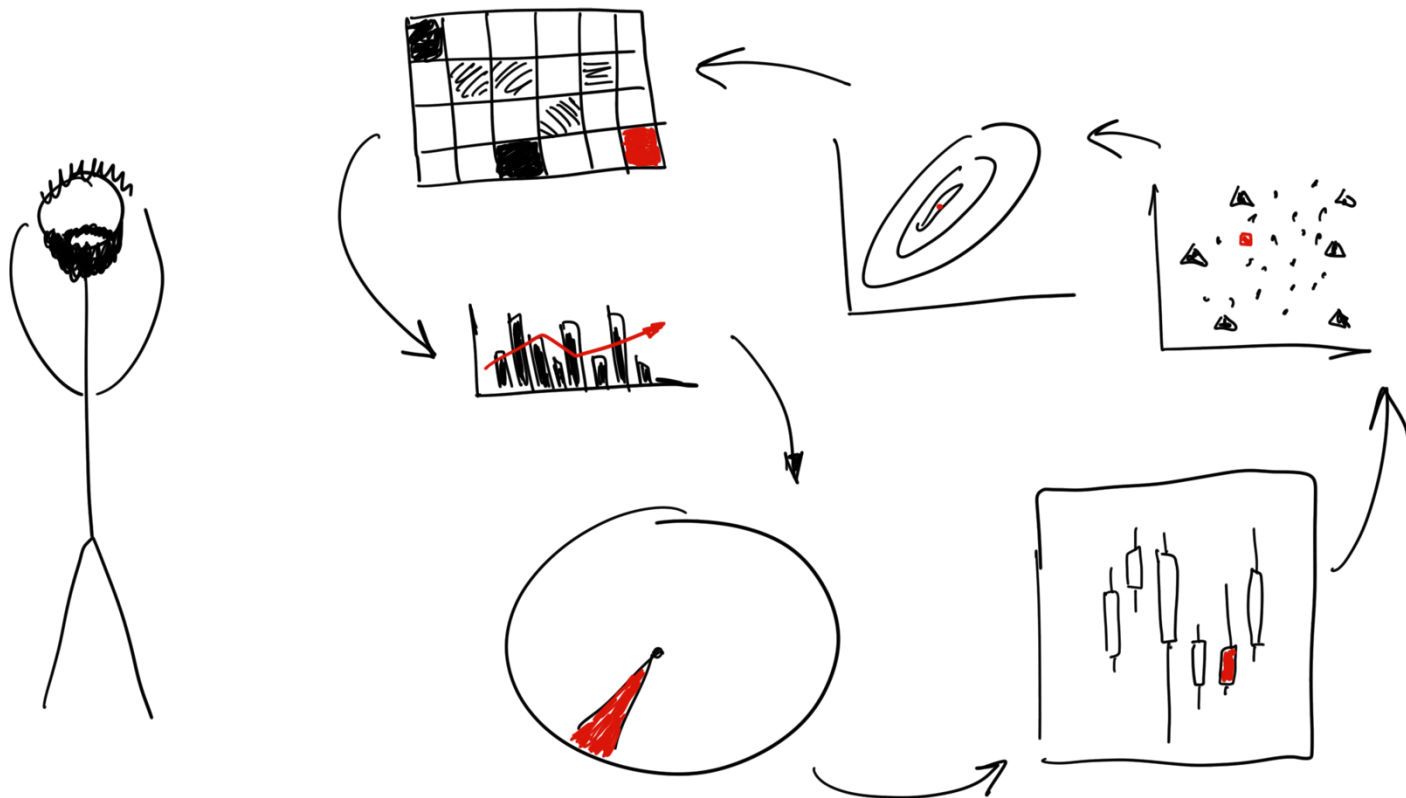


<https://matplotlib.org/gallery/index.html>
<https://seaborn.pydata.org/examples/index.html>
<https://python-graph-gallery.com/>
<https://www.mathworks.com/products/matlab/plot-gallery.html>
<https://www.think-cell.com/en/product/features.shtml>
<https://plot.ly/>



<https://www.reddit.com/r/dataisbeautiful/>
<https://www.reddit.com/r/Infographics/>
<https://www.reddit.com/r/dataisugly/>
<https://t.me/DataScienceology>
<https://pinterest.com/>

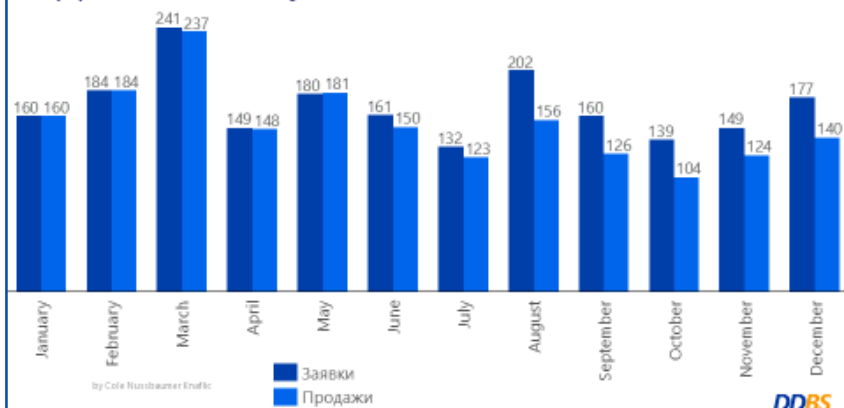
Хаос — ваш враг



Хаос — ваш враг

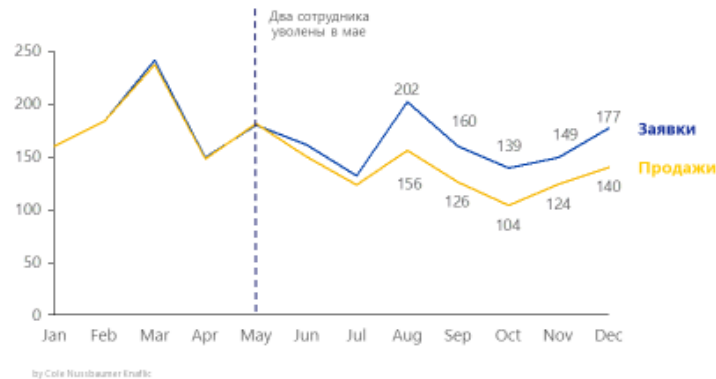
До

Эффективная визуализация

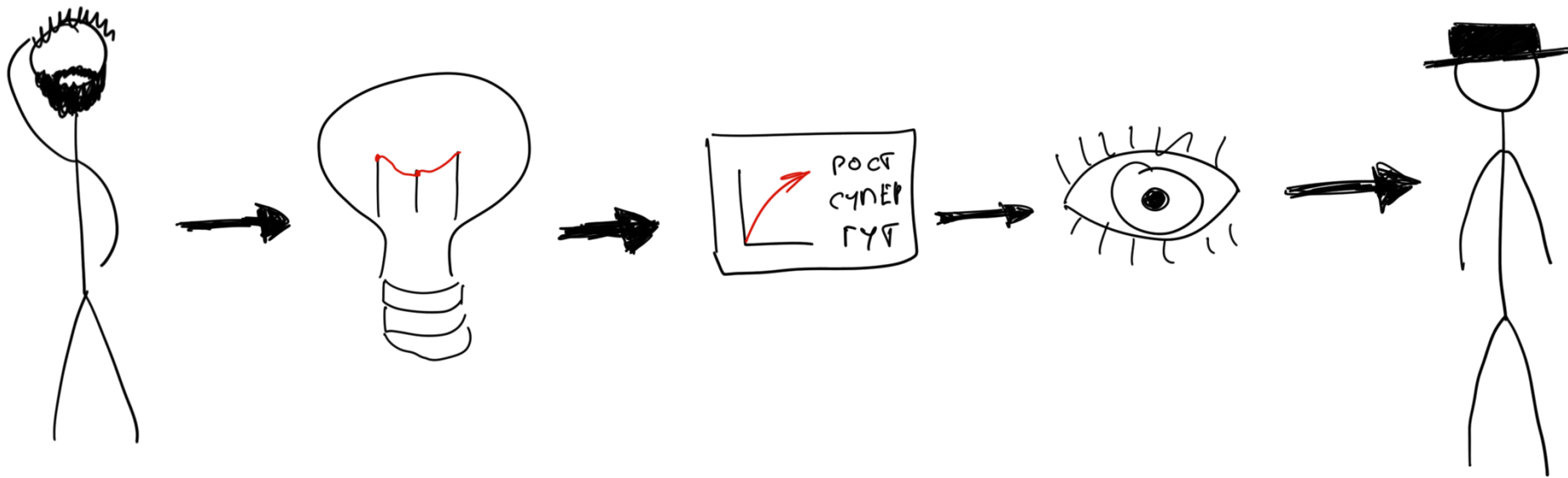


После

Эффективная визуализация



Контроль внимания аудитории



Контроль внимания аудитории

Iconic Memory

Якоря внимания

- Очень краткосрочная память;
- Почти мгновенно переходит в short-term, если внимание заостряется;
- Случай «запомнил, но не помню откуда».

Short-term Memory

Основная обработка информации

- Все выводы будут сделаны здесь;
- Не более 4-х частей информации (4 линии, 4 цвета, 4 простых графика);
- Увеличение когнитивной нагрузки → потеря внимания;
- До ~30 секунд.

Long-term Memory

Здесь то, что останется через неделю

- Из short-term информация либо исчезает, либо переходит в long-term;
- Изображения будут просто напоминать о содержании;
- Работает на ассоциативных связях;
- Отложится лучше, если показать в short-term несколько раз;
- Мы на неё влияем минимально;
- В соревнованиях по памяти часто используют связь с привычными объектами.

Контроль внимания аудитории

23493290751601
29385712093487
12098343190322

by Cole Nussbaumer Knaflic

Контроль внимания аудитории

2**3**49**3**290751601
29**3**8571209**3**487
12098**3**43190**3**22

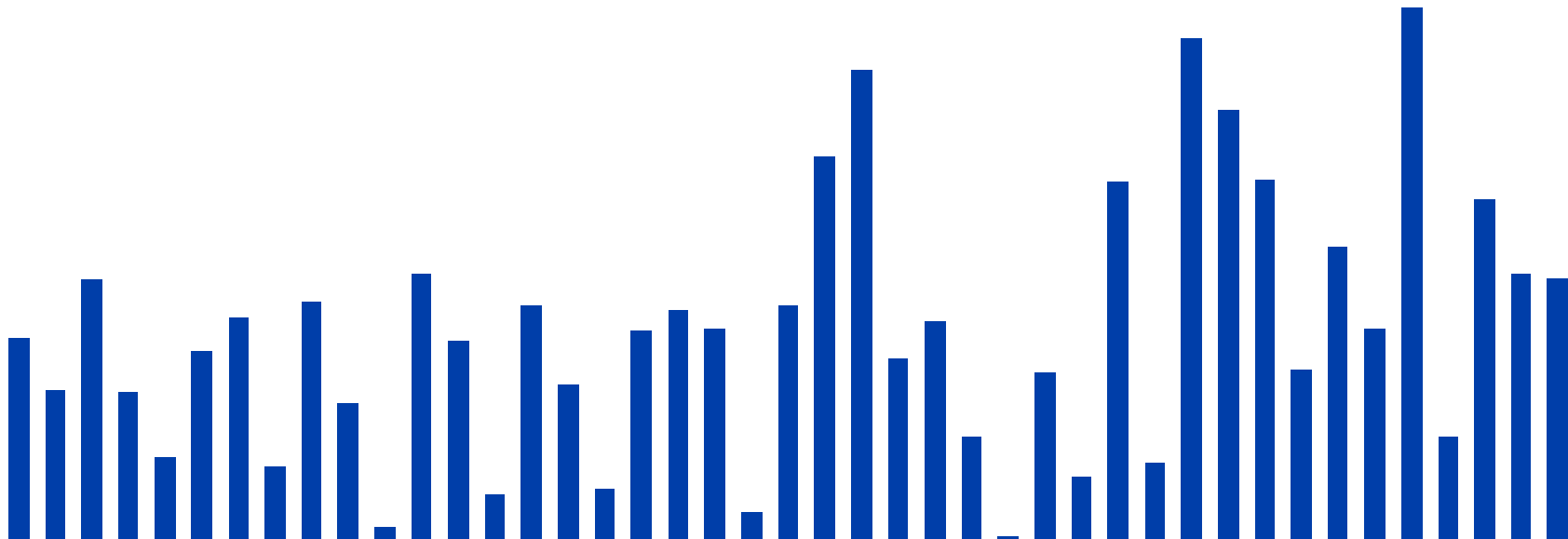
by Cole Nussbaumer Knaflic

Контроль внимания аудитории

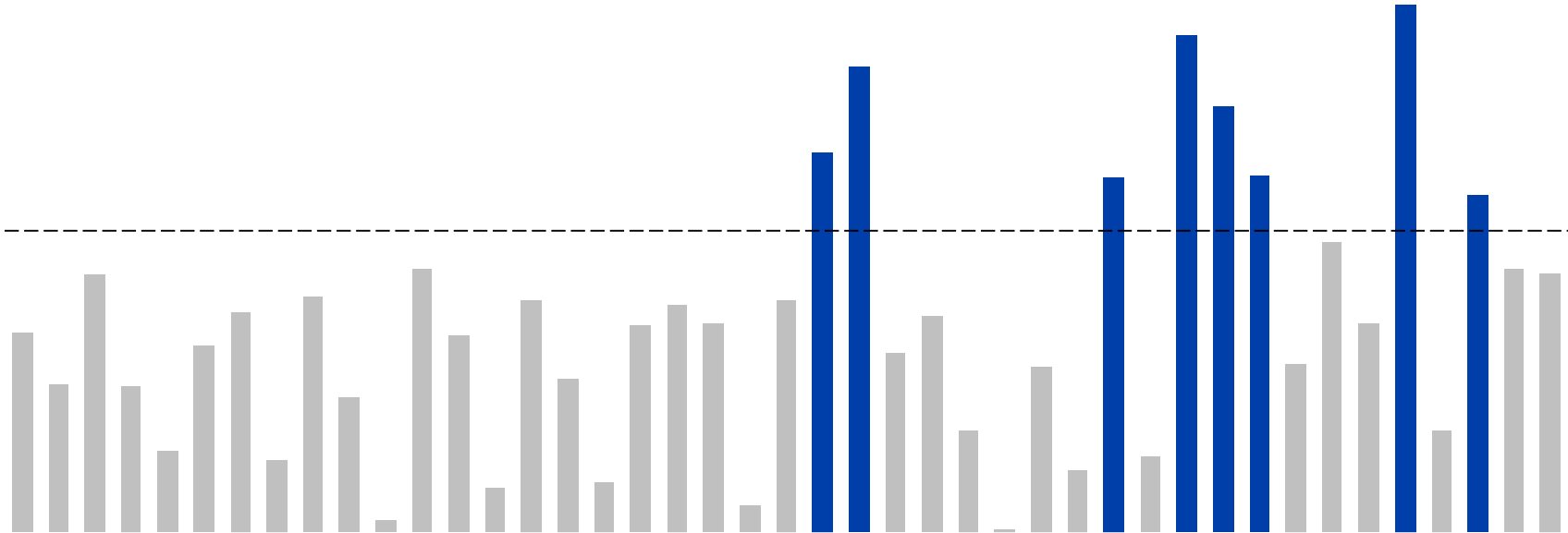
2**3**49**3**290751601
29**3**8571209**3**487
12098**3**4**3**190**3**22

by Cole Nussbaumer Knaflic

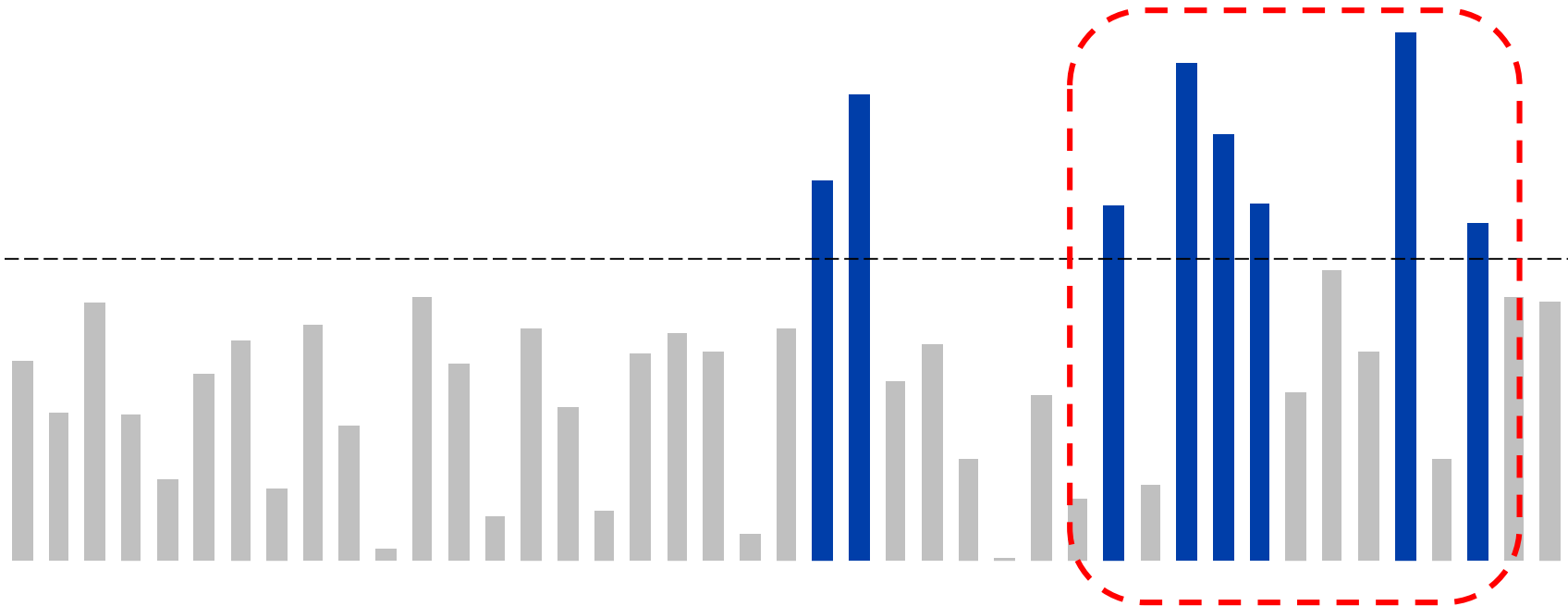
Контроль внимания аудитории



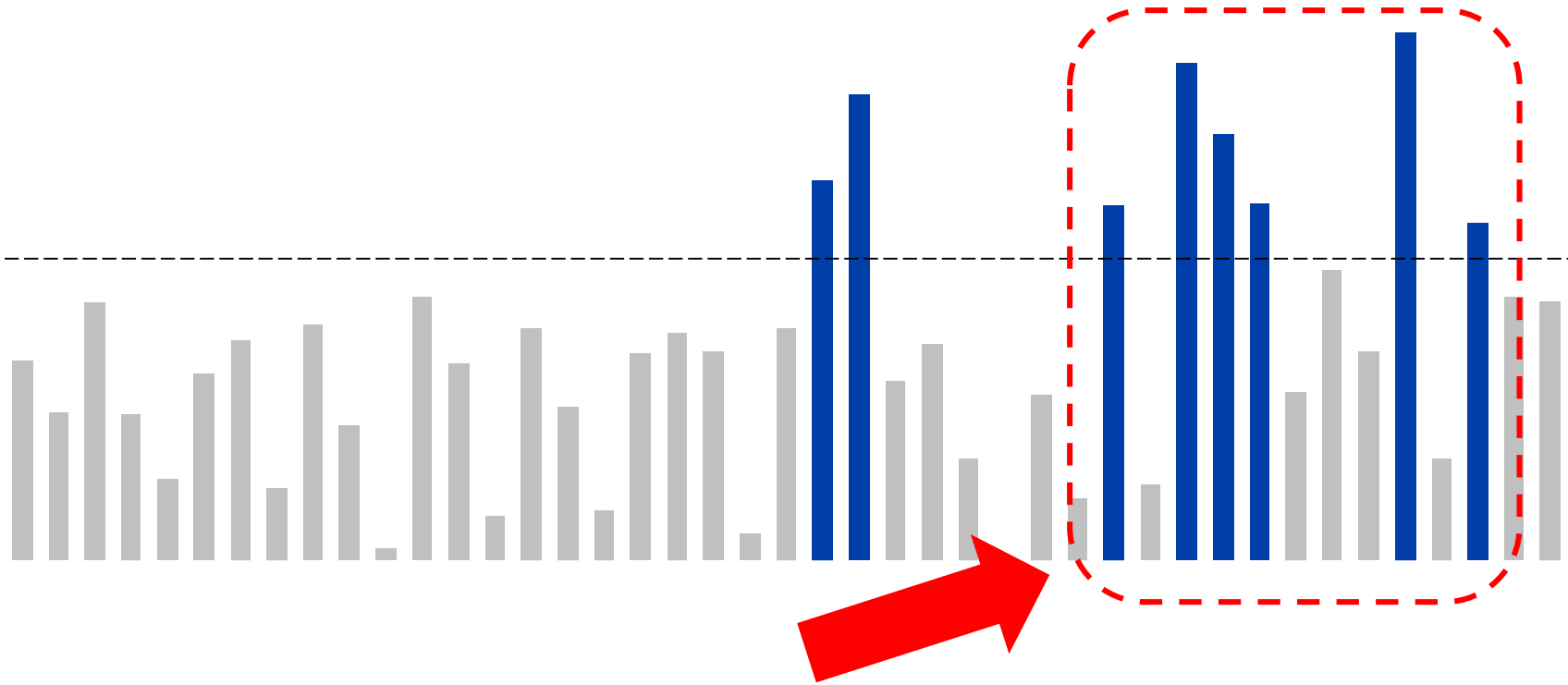
Контроль внимания аудитории



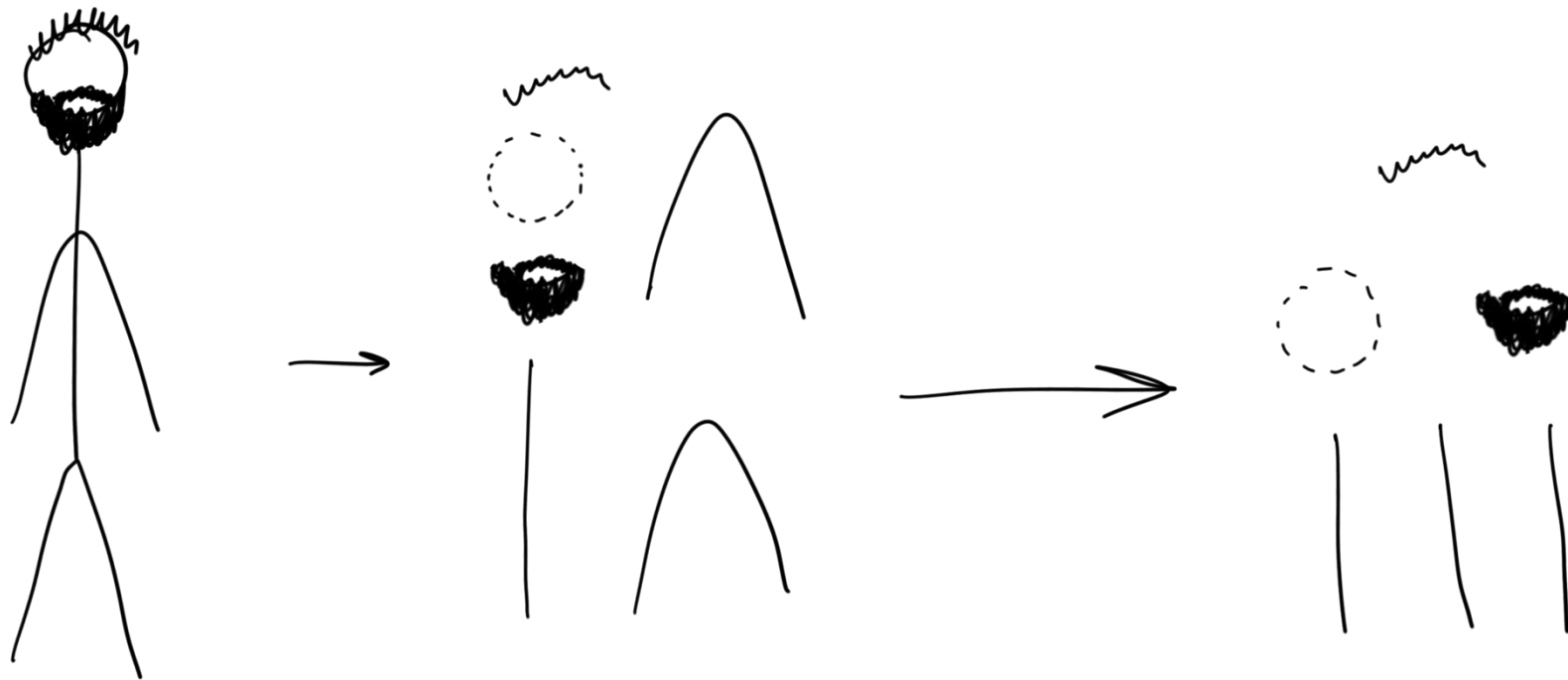
Контроль внимания аудитории



Контроль внимания аудитории



Думай как дизайнер



Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- КАПС или другой шрифт;
- **Цвет** — очень действенно;
- Д а ж е т а к или ТАК.

Удаление раздражителей

- *Не все данные одинаково важны;*
- ...

Создание иерархии

- Фокус;
- ...

Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- КАПС или другой шрифт;
- **Цвет** — очень действенно;
- Д а ж е т а к или **ТАК**.

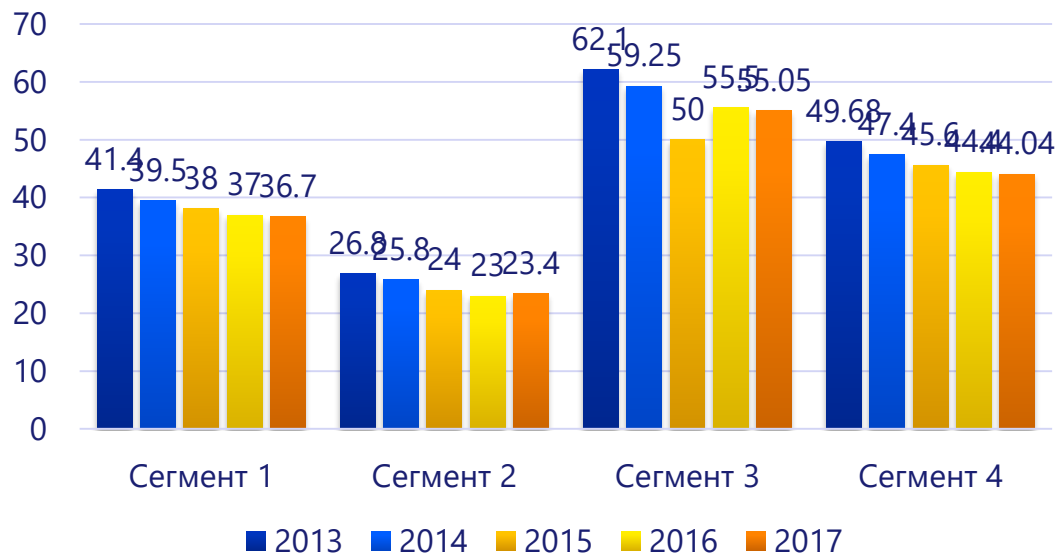
Удаление раздражителей

- *Не все данные одинаково важны;*
- ...

Создание иерархии

- Фокус;
- ...

Качество прогноза



Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- КАПС или другой шрифт;
- **Цвет** — очень действенно;
- Д а ж е т а к или **ТАК**.

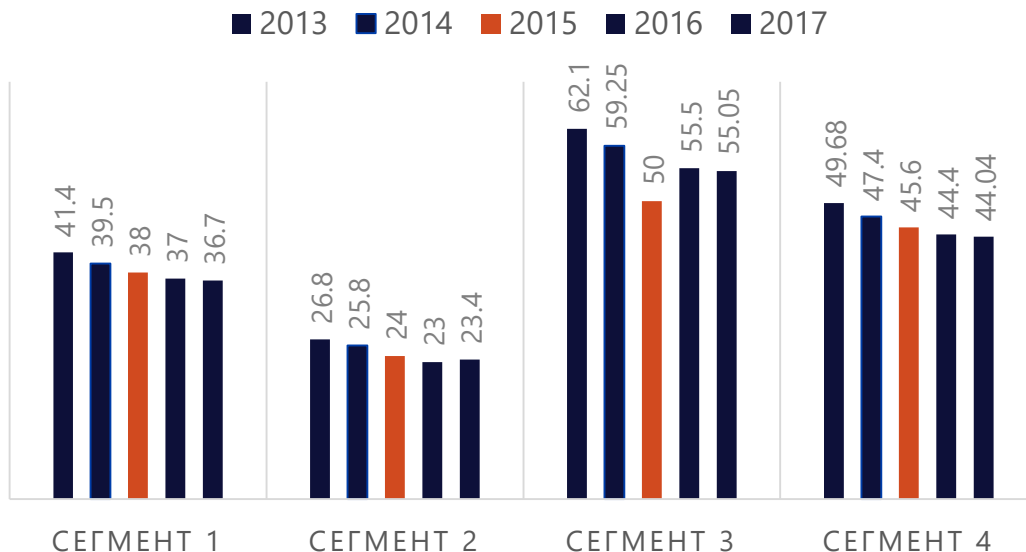
Удаление раздражителей

- *Не все данные одинаково важны;*
- ...

Создание иерархии

- *Фокус;*
- ...

Качество прогноза



Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- ...

Удаление раздражителей

- Не все данные одинаково важны;
- Обобщай, когда детали не важны;
- «Что изменится убрав часть?»;
- Серый цвет для «нецелвых».

Создание иерархии

- Фокус;
- ...

Качество прогноза



Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- ...

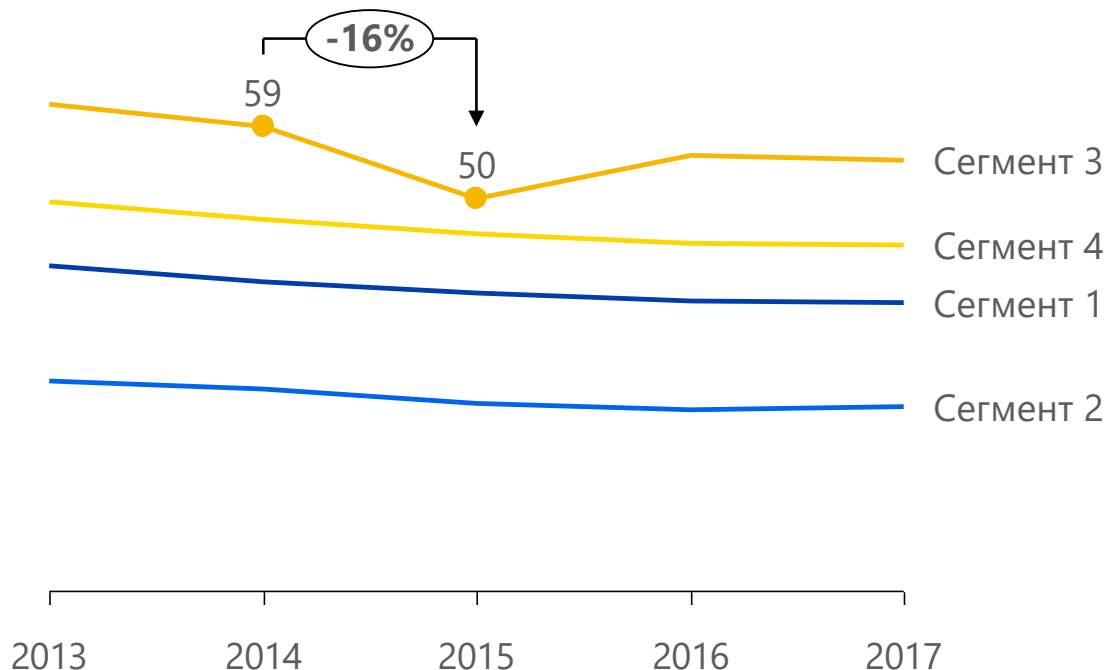
Удаление раздражителей

- Не все данные одинаково важны;
- Обобщай, когда детали не важны;
- «Что изменится убрав часть?»;
- Серый цвет для «нецелвых».

Создание иерархии

- Фокус;
- ...

Качество прогноза



Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- ...

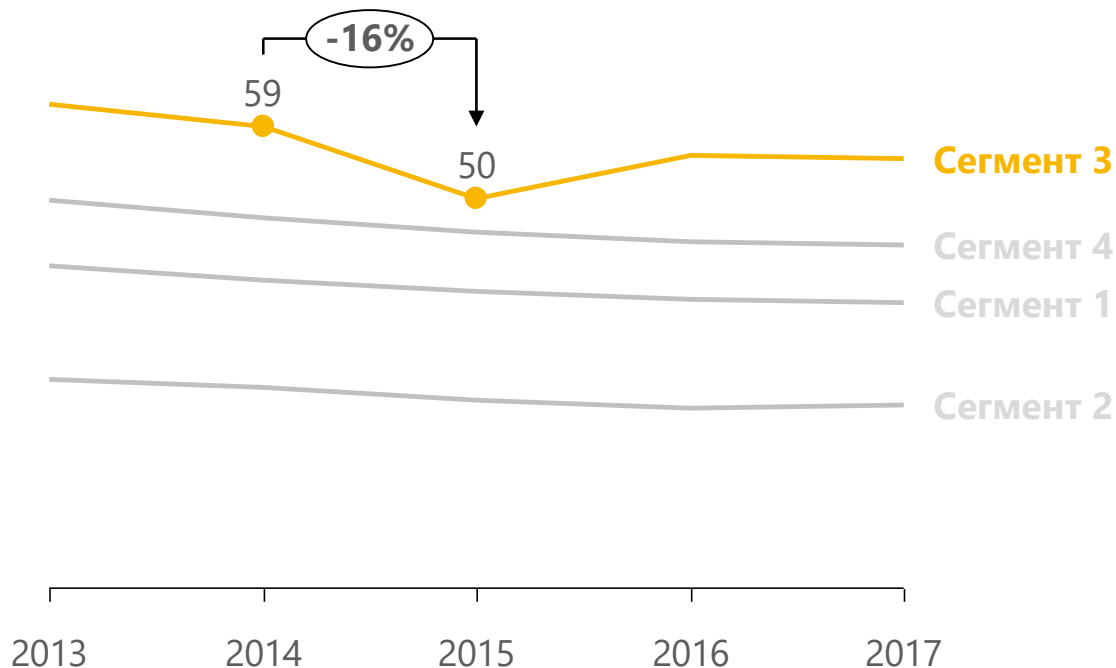
Удаление раздражителей

- Не все данные одинаково важны;
- Обобщай, когда детали не важны;
- «Что изменится убрав часть?»;
- Серый цвет для «нецелвых».

Создание иерархии

- Фокус;
- ...

Качество прогноза



Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- ...

Удаление раздражителей

- *Не все данные одинаково важны;*
- ...

Создание иерархии

- Фокус (передний и задний планы);
- Последовательность;
- Композиция;
- ...

Думай как дизайнер

Доступность

Выделение важного

- **Bold**, *italic*, underline;
- ...

Удаление раздражителей

- *Не все данные одинаково важны;*
- ...

Создание иерархии

- Фокус (передний и задний планы);
- Последовательность;
- Композиция;
- ...

15 GOLDEN PRINCIPLES OF VISUAL HIERARCHY



FOCAL POINT

Need to lay maximum strength on a particular aspect?
Go focal!

MOVEMENT

With this one, make your audience move with the flow.



GOLDEN RATIO

Need to add beauty to your work?
Go Golden.

BALANCE

Play safe with balance.
A little balance never made anyone cringe!



REPETITION

With repetition, make sure no one misses your callouts ever again!

WHITE SPACE

Add sophistication with whitespace—where less is more!



VISUAL TRIANGLE

Works best for limited data. Makes sure no one misses a point!

TEXTURE

Add texture to refract negative space.



TYPES

Say it loud and clear with a typeface. Leave no one in doubt!

RANDOM

At times, an absolute lack of symmetry



2

SIMPLIFICATION

Too much going on? Simplify it. Compositions with clutter can make it difficult to identify the focal point.

3

RULE OF ODDS

An odd number of elements can be perceived as being more natural than an even number of elements.

4

RULE OF THIRDS

Segment your composition into thirds both ways. A common technique is to place your focal point at one of the intersections. Also, as a general guideline, each of the segments should be different.

5

CREATE MOVEMENT

A stagnant composition may be boring. Create movement in your composition and use it to guide your viewer's eye around the composition.

6

DON'T CUT IN TWO

Avoid bisecting your composition, for example with the horizon.

7

PEOPLE LIKE TRIANGLES

Triangular compositions are thought to be more aesthetically pleasing. This is similar to the Rule of Odds.

www.drawpaintacademy.com

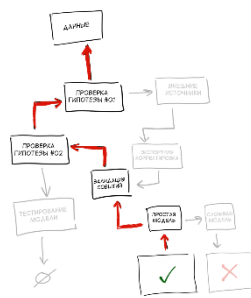
C. Live

Pulling it all together

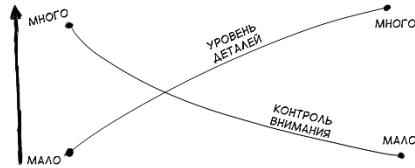
История



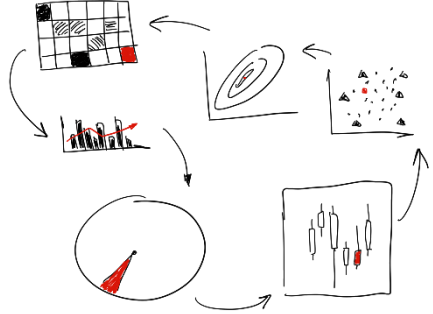
Контекст



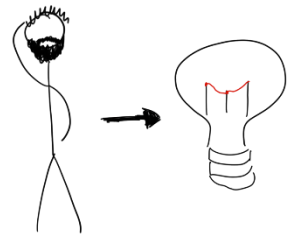
Визуализация



Чистка



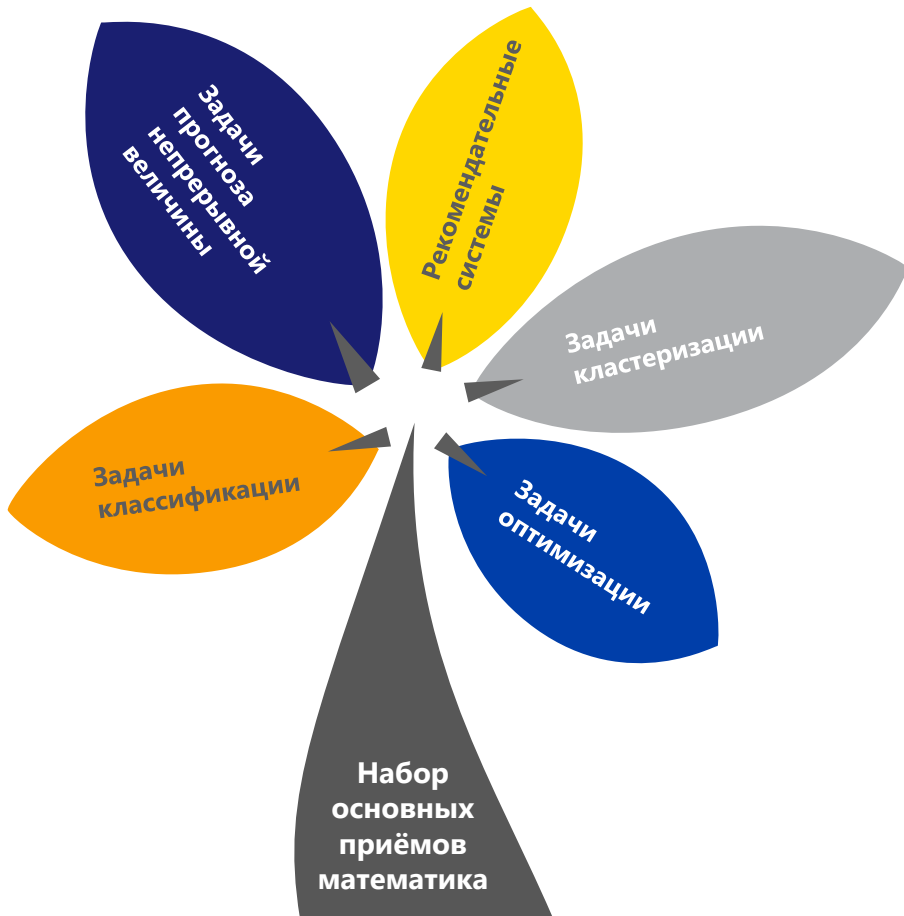
Внимание



Дизайн



Классические приёмы для решения математических задач



Классификация

Самые явные проблемы — проблемы с учителем. Они же приносят основной доход

Прогноз непрерывной величины

Такие задачи, как оценка времени ожидания клиента или утилизация кредитного лимита

Рекомендательные системы

Рекомендательные системы часто выделяют в отдельную группу задач, т.к. строится ранжирование сразу на группу сущностей

Кластеризация

Группа задач, зачастую применяется для уменьшения размерности или при изучении данных

Задачи оптимизации

Используется обыкновенно для максимизации прибыли используя ограничения/ёмкости текущих бизнес-процессов.

Кейс

БП: Повышение дохода портфеля кредитных карт

Бизнес проблема → постановка математической задачи



МП: Набор моделей, одна на проблему

Бизнес проблема → постановка математической задачи



Модели построены

Краткий отчёт

Построены модели

- Прогноз вероятности дефолта
- Прогноз утилизации
- Оптимизация ставки

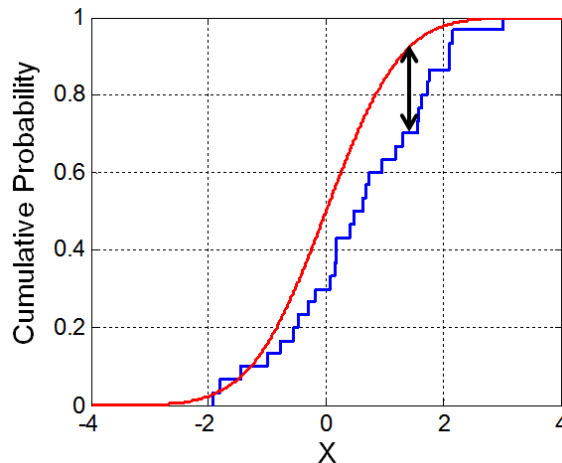
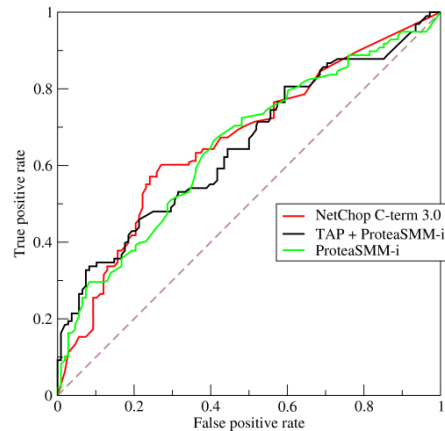
Основные допущения



Исключены клиенты с историей трат менее, чем 3 месяца



Доходы клиентов более чем 1 год назад скорректированы на инфляцию



Модели построены

Краткий отчёт

Построены модели

- Прогноз вероятности дефолта
- Прогноз утилизации
- Оптимизация ставки

Основные допущения

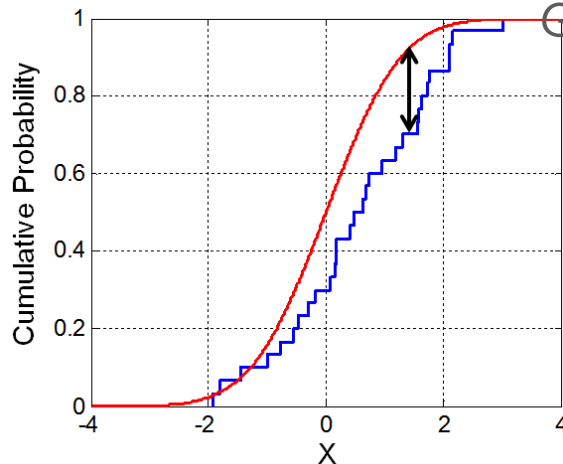
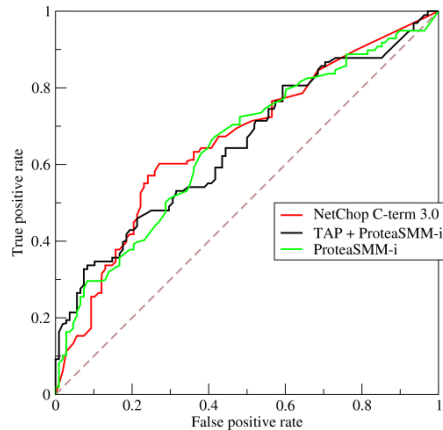


Исключены клиенты с историей трат менее, чем 3 месяца



Доходы клиентов более чем 1 год назад скорректированы на инфляцию

Train Gini: 0.73
Test Gini: 0.70
Lift on test top 10%: 5.6
Train R^2 : 0.69



Train K-S: 0.31
Test Gini: 0.29
Gain on test top 10%: 27%
Test R^2 : 0.63

Модели построены

Краткий отчёт

Построены модели

- Прогноз вероятности дефолта
- Прогноз утилизации
- Оптимизация ставки

Основные допущения

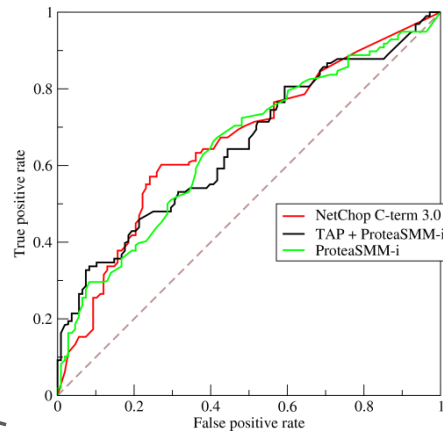


Исключены клиенты с историей трат менее, чем 3 месяца

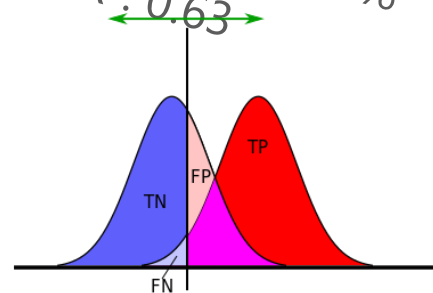
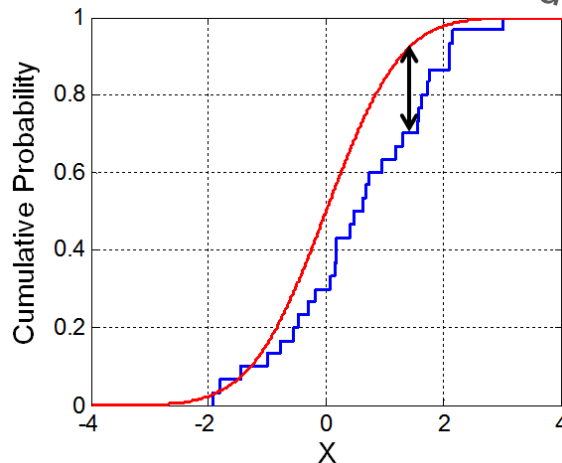


Доходы клиентов более чем 1 год назад скорректированы на инфляцию

Train Gini: 0.73
Test Gini: 0.70
Lift on test top 10%: 5.6
Train R^2 : 0.69



Train K-S: 0.31
Test Gini: 0.29
Gain on test top 10%: 27%
Test R^2 : 0.63



Модели построены

Краткий отчёт

Построены модели

- Прогноз вероятности дефолта
- Прогноз утилизации
- Оптимизация ставки

Основные допущения



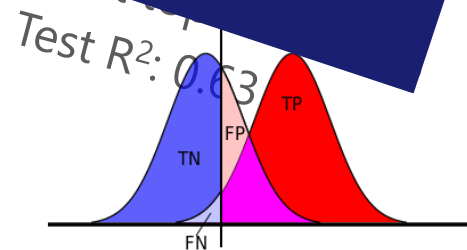
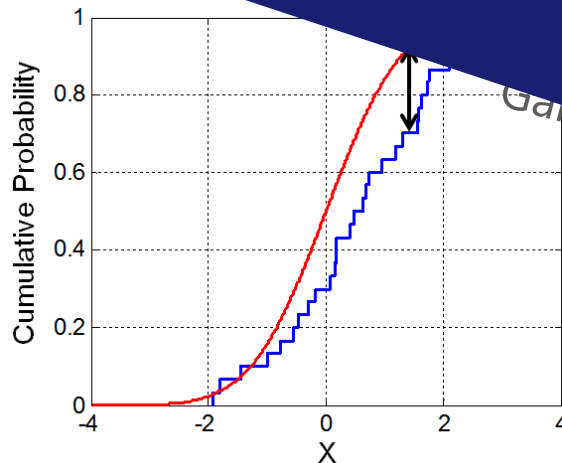
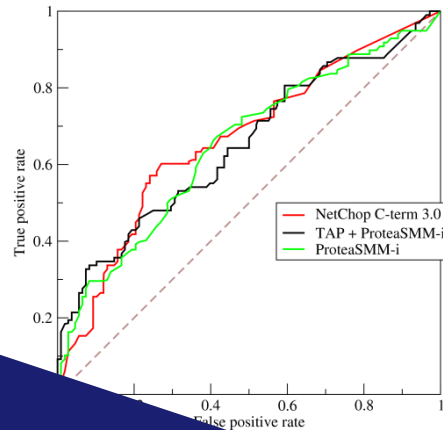
Исключены клиенты с историей трат менее, чем 3 месяца



Доходы клиентов более чем 1 год назад скорректированы на инфляцию

Train Gini: 0.73
Test Gini: 0.70
10%: 5.6





















ROI = 1.1



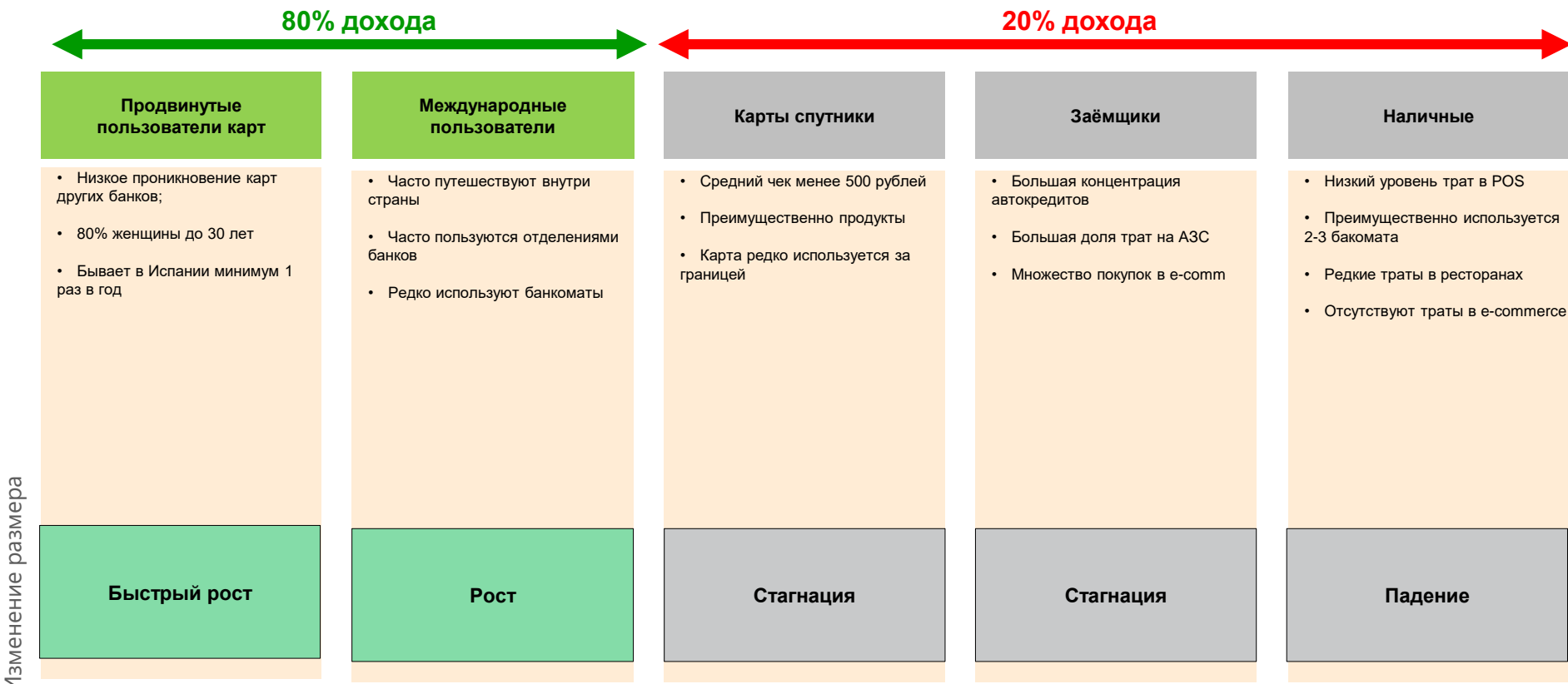
Test R²: 0.63

История

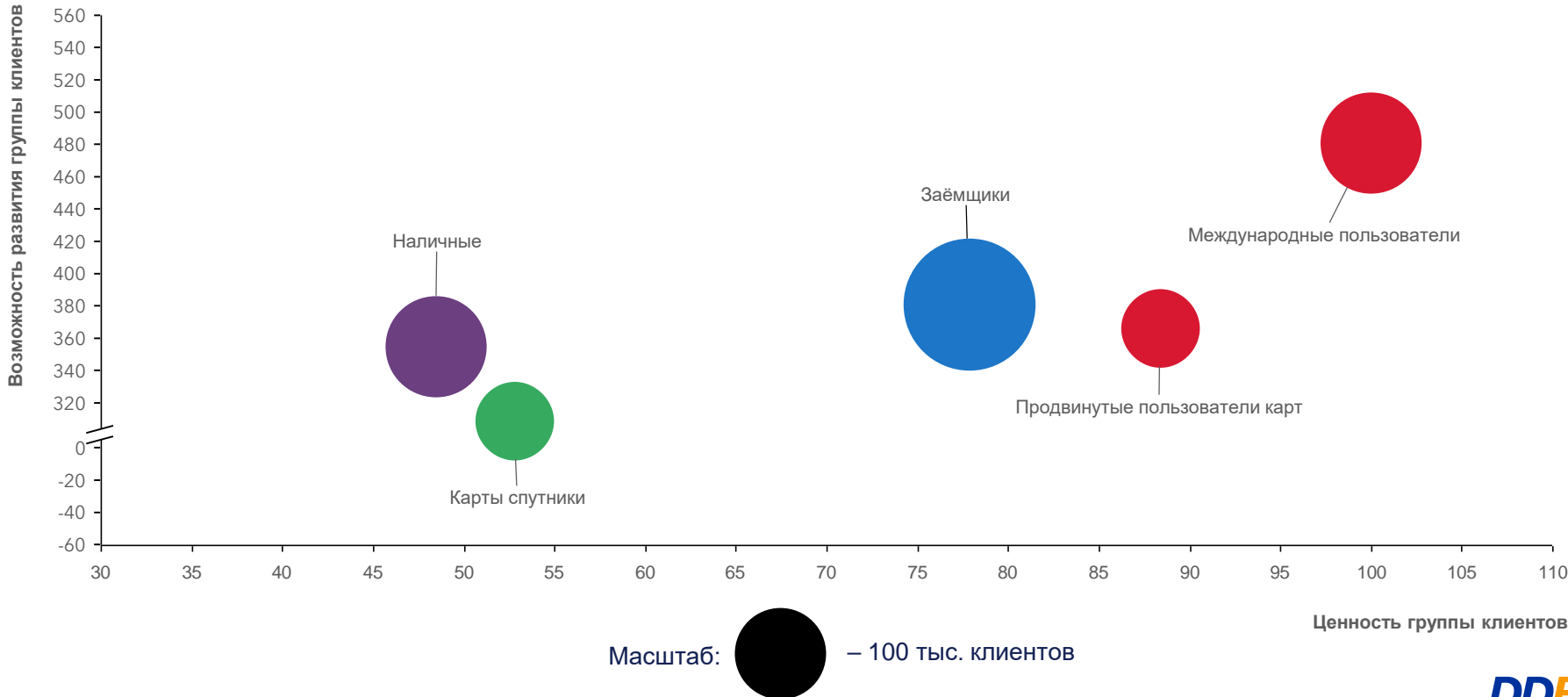
Все истории начинаются с описания действующих лиц

1. Имя кластера	2. Обобщенное имя кластера	Депозиты	Кредиты	Доля POS
Активные пользователи карт	 Продвинутые пользователи карт			
Пользователи более, чем одного банка				
Пользователи более, чем трёх банков				
Бизнесмены	 Карты спутники			
Держатели крупных депозитов				
Редкие крупные покупки				
Активные путешественники	 Международные пользователи			
Иностранные граждане				
Социальный сегмент				
Более 1 кредита наличными	 Заёмщики			
Наличие ипотеки				
Наличие автокредита				
Более трёх кредитных карт	 Наличные			
Используют только наличные				
Преобладают наличные				
Мелкие траты по карте				

Сюжет. Раскрытие особенностей групп клиентов

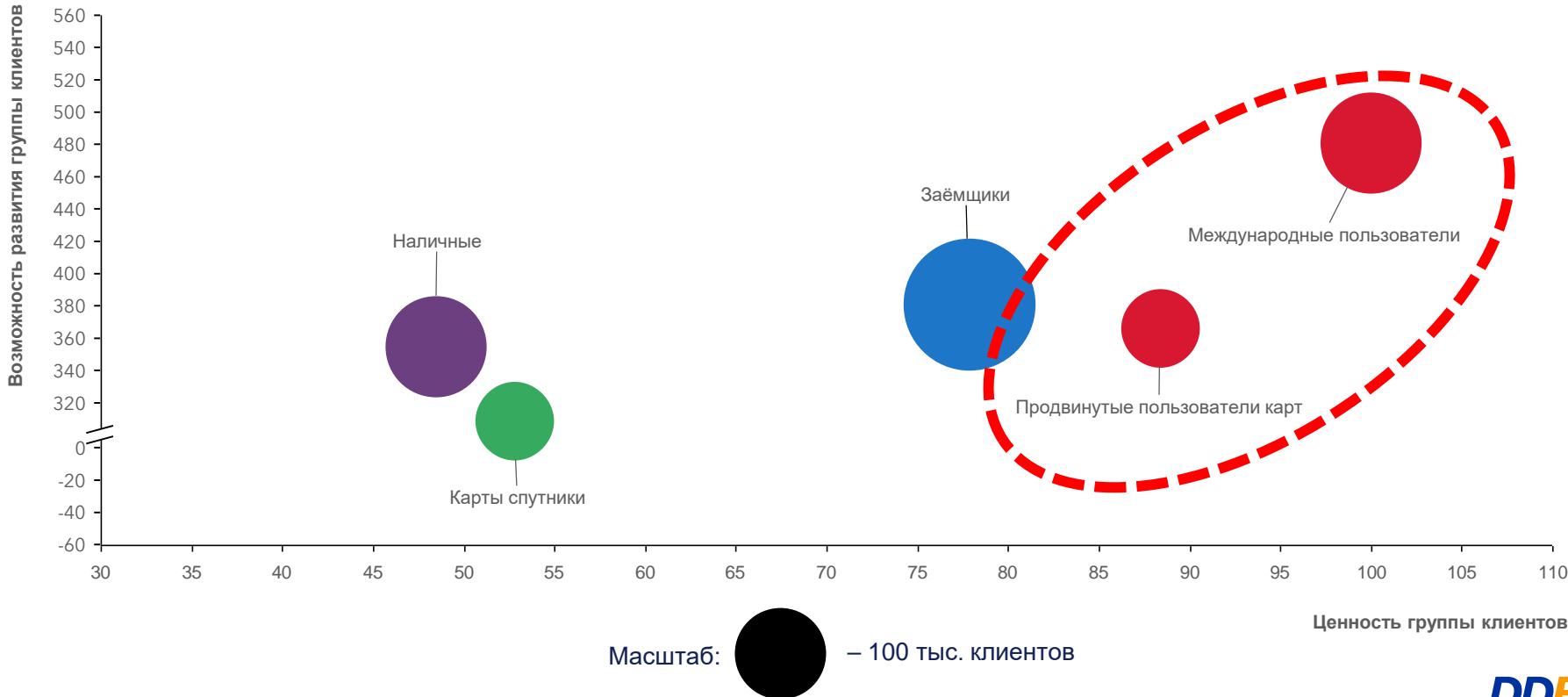


Сравнение поведения групп клиентов



Сравнение поведения групп клиентов

Целевая аудитория



Ключ к сердцу бизнеса

Действующие лица

- Использовать только понятные и бизнес-значимые показатели
- Описание должно быть сравнимо
- Все лица должны помещаться на 1 слайд с описанием

Сюжет

- Сравнить развитие лиц во времени между собой
- Описать взаимодействие лиц с бизнесом
- Описать влияние и отношение лица на бизнес

Сравнение

- Описать влияние лиц на итоговый результат
- Выбрать лиц наиболее близких к решению и сделать основной фокус на них
- И не забыть о финальном влиянии на бизнес

Final

Инструменты



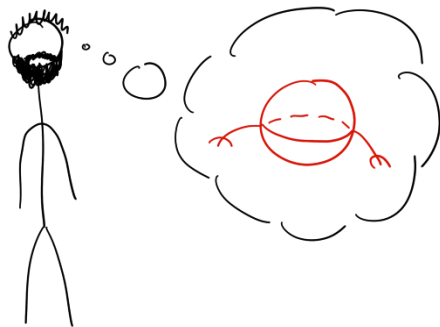
Обратная связь



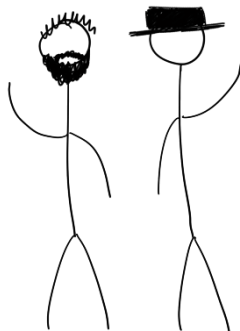
Время



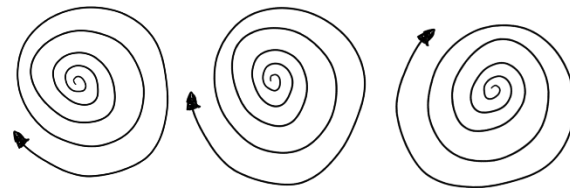
Вдохновение



Стиль



Повторение





Спасибо!

DDBS