

Визуализация данных

Алексей Смагин | 2022

**Зачем мы
визуализируем данные?**



fx

год

1 год месяц браков

2 2006 январь 55 509

3 2006 февраль 62 449

4 2006 март 70 798

5 2006 апрель 86 055

6 2006 май 35 960

7 2006 июнь 111 409

8 2006 июль 127 475

9 2006 август 149 120

10 2006 сентябрь 151 116

11 2006 октябрь 95 192

12 2006 ноябрь 86 480

13 2006 декабрь 82 101

14 2007 январь 59 495

15 2007 февраль 68 255

16 2007 март 68 173

17 2007 апрель 106 800

18 2007 май 39 331

19 2007 июнь 119 012

20 2007 июль 147 253

21 2007 август 163 630

Зачем нужна визуализация?

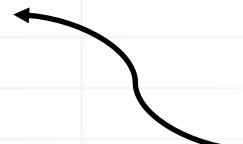


fx | год

	A	B	C	D	E
1	год	месяц	браков		
2	2006	январь	55 509		
3	2006	февраль	62 449		
4	2006	март	70 798		
5	2006	апрель	86 055		
6	2006	май	35 960		
7	2006	июнь	111 409		
8	2006	июль	127 475		
9	2006	август	149 120		
10	2006	сентябрь	151 116		
11	2006	октябрь	95 192		
12	2006	ноябрь	86 480		
13	2006	декабрь	82 101		
14	2007	январь	59 495		
15	2007	февраль	68 255		
16	2007	март	68 173		
17	2007	апрель	106 800		
18	2007	май	39 331		
19	2007	июнь	119 012		
20	2007	июль	147 253		
21	2007	август	163 630		

Зачем нужна визуализация?

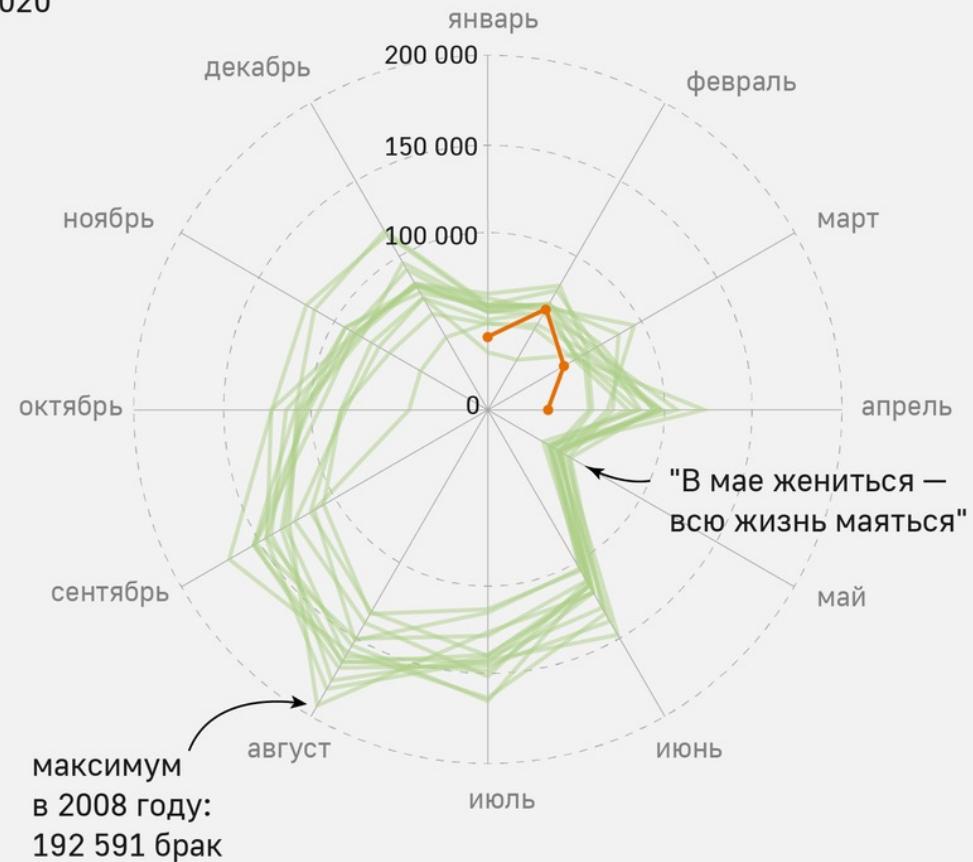
Таблица идеально подходит,
если нужно рассмотреть
каждое значение



Число браков в России

— 2006–2019

— 2020



**визуализация помогает донести
идею, которая содержится в данных**

Хороший график:

- Основан на хороших данных
- Не искажает данные
- Понятен
- Эстетически приятен

Самый страшный грех

прямой эфир

Динамика выявленных случаев заражения COVID-19 в России



Общее число заражённых за весь период эпидемии – **353 427**

Источник: Оперативный штаб

■ РБК 14:01 мск 3 тыс. специалистов [rbc.ru](#) Мишустин: выплаты медраб
EUR ЦБ 78,44 +0,83% [Финтный рынок](#) RUONIA 5,32 -0,746% [Мосбиржа](#) АЛРО

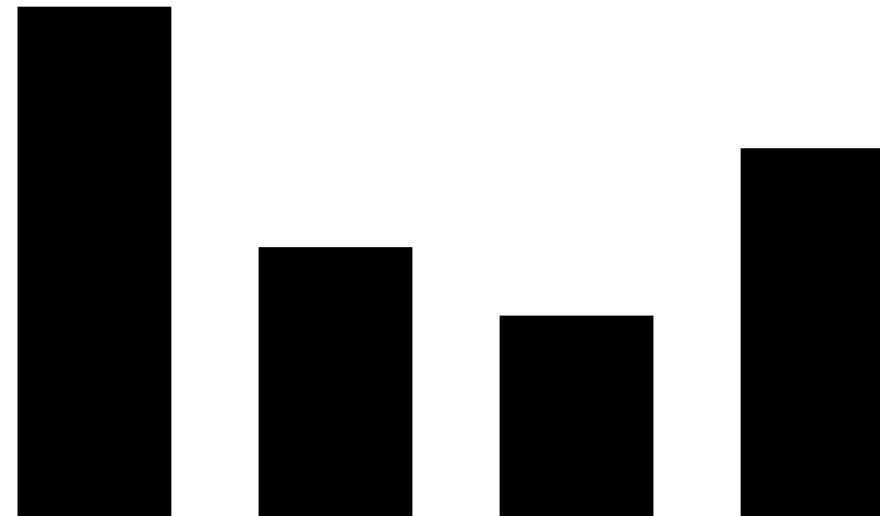
Как кодируется информация на графике?

Допустим, каждый прямоугольник – это страна



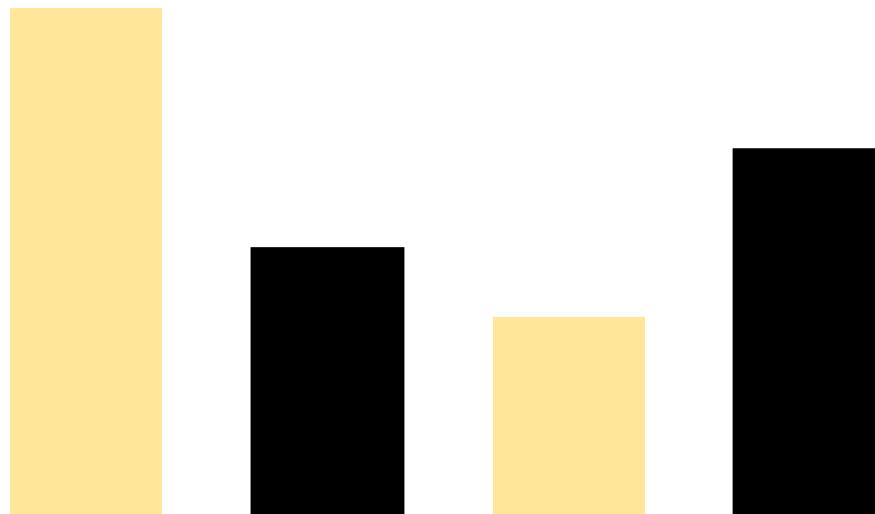
Как кодируется информация на графике?

Пусть размер ВВП – это высота
прямоугольника



Как кодируется информация на графике?

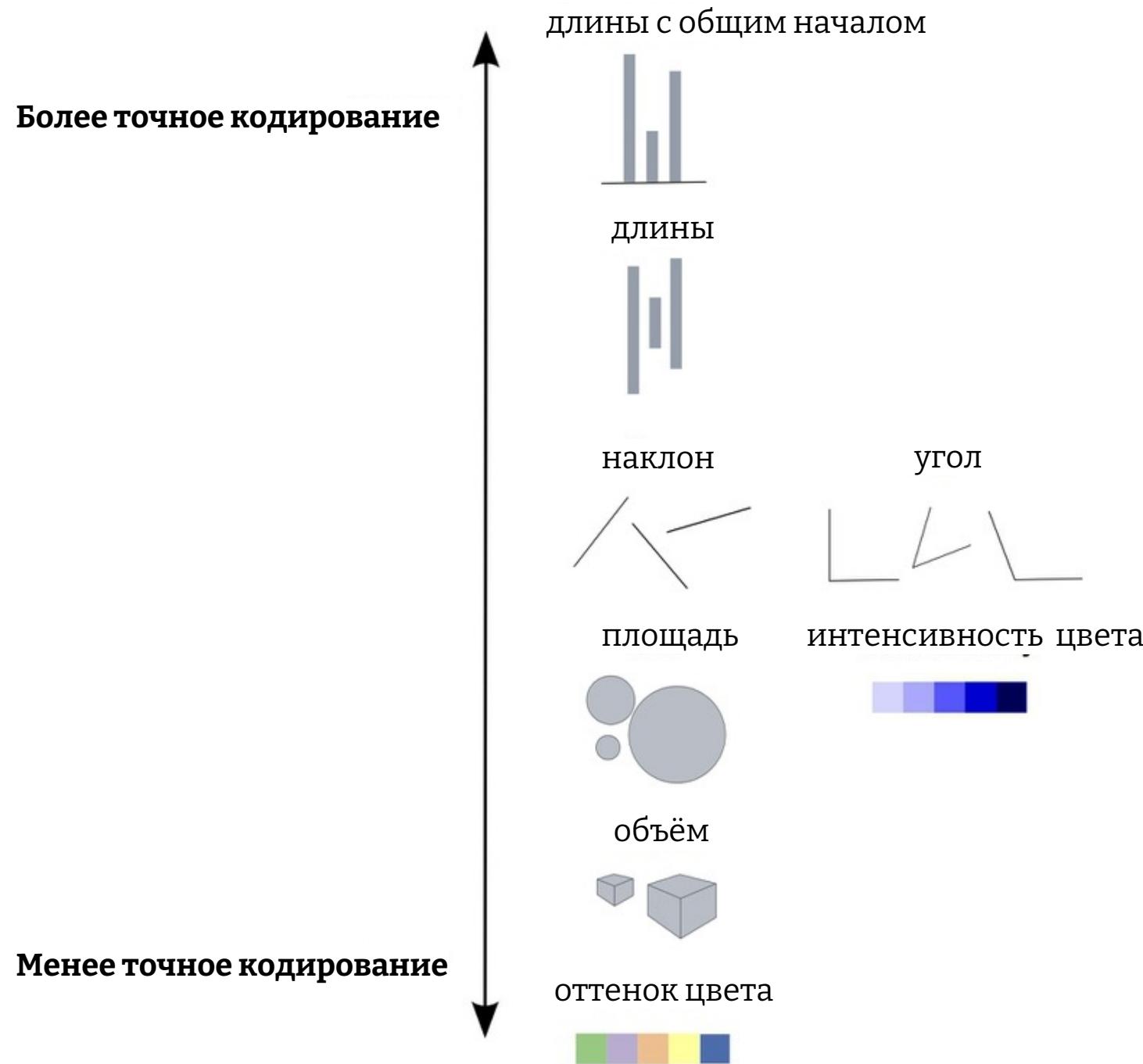
Обозначим страны СНГ – жёлтым



Как кодируется информация на графике?

Справа – страны, где глава государства –
мужчина, слева – где женщина



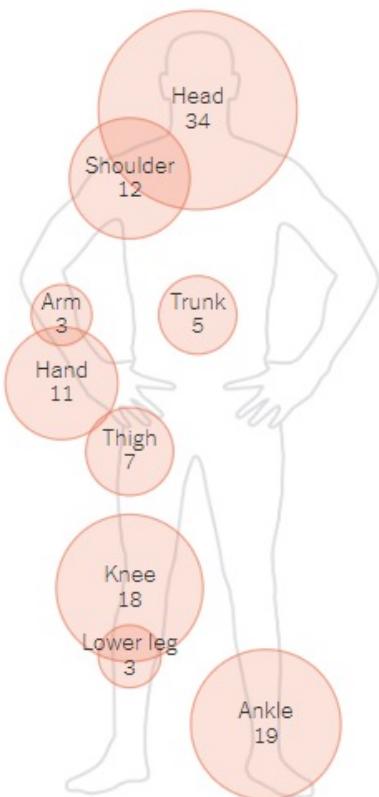


Как закодирована информация?

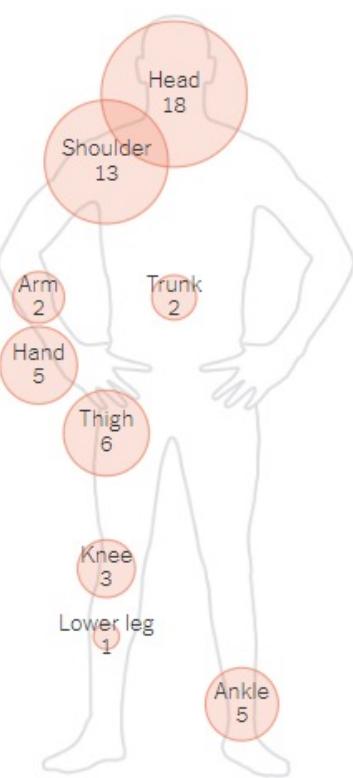
Common injuries for boys among popular high school sports

Injuries per 10,000 competition plays

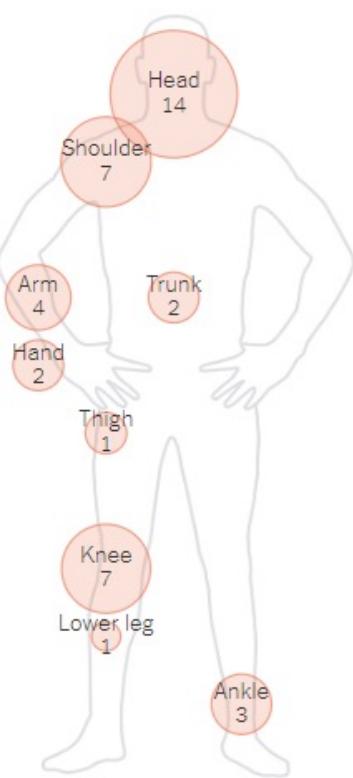
Football



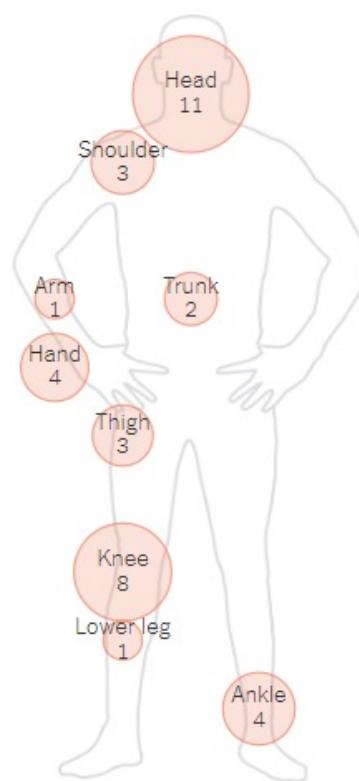
Hockey



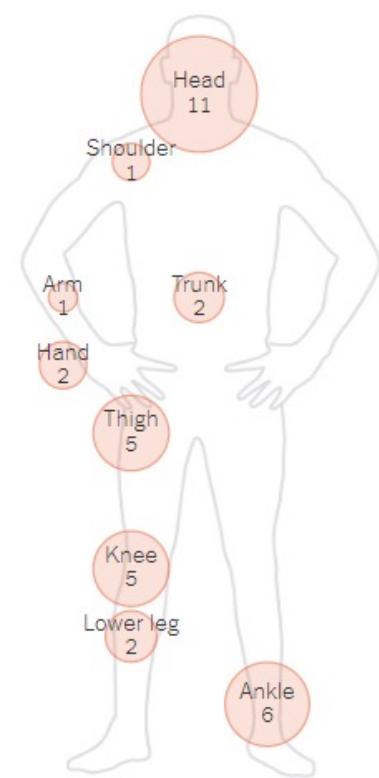
Wrestling



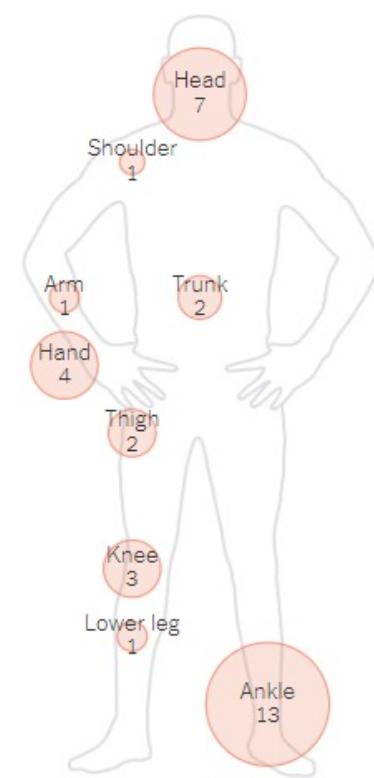
Lacrosse



Soccer

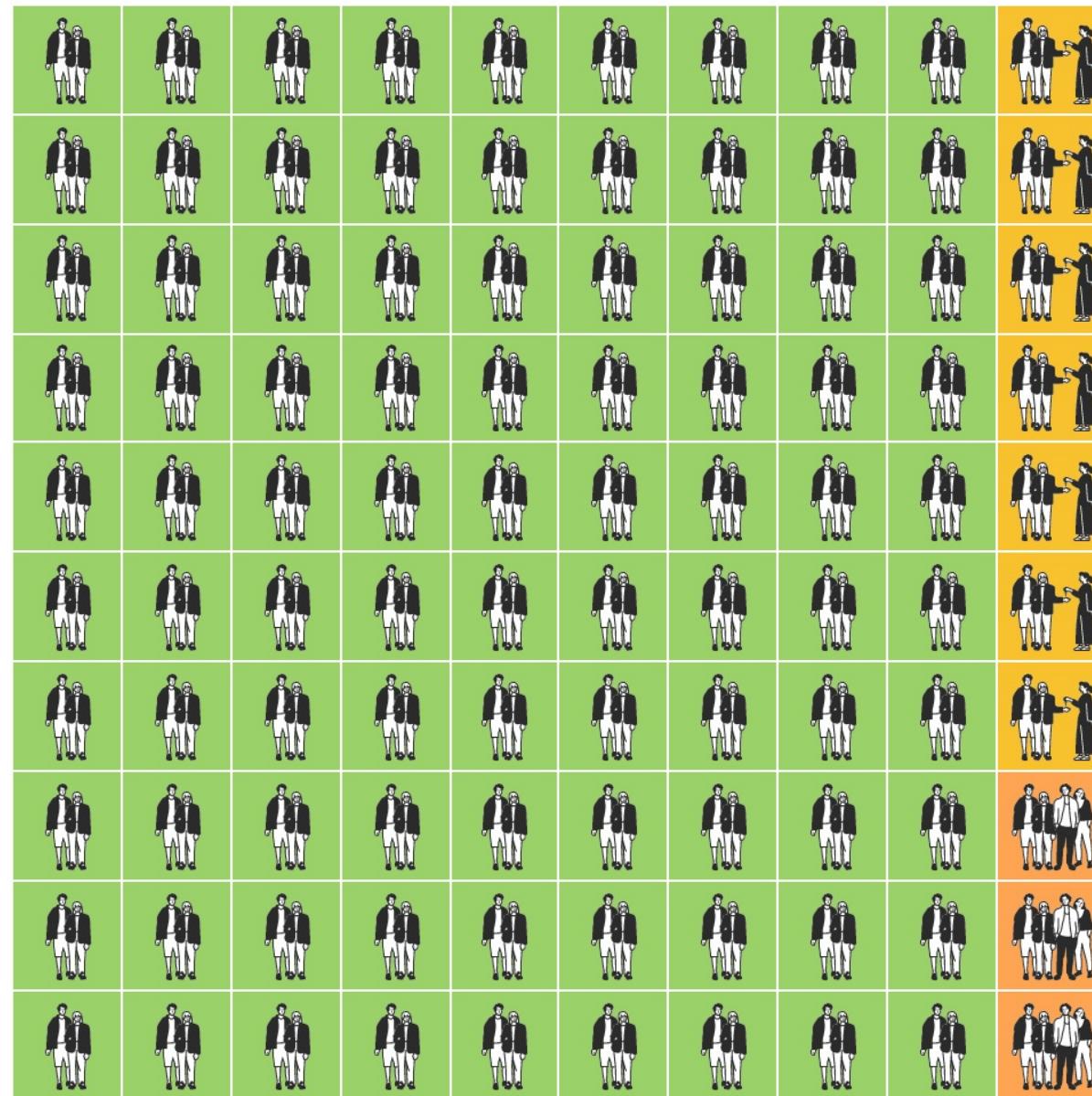


Basketball



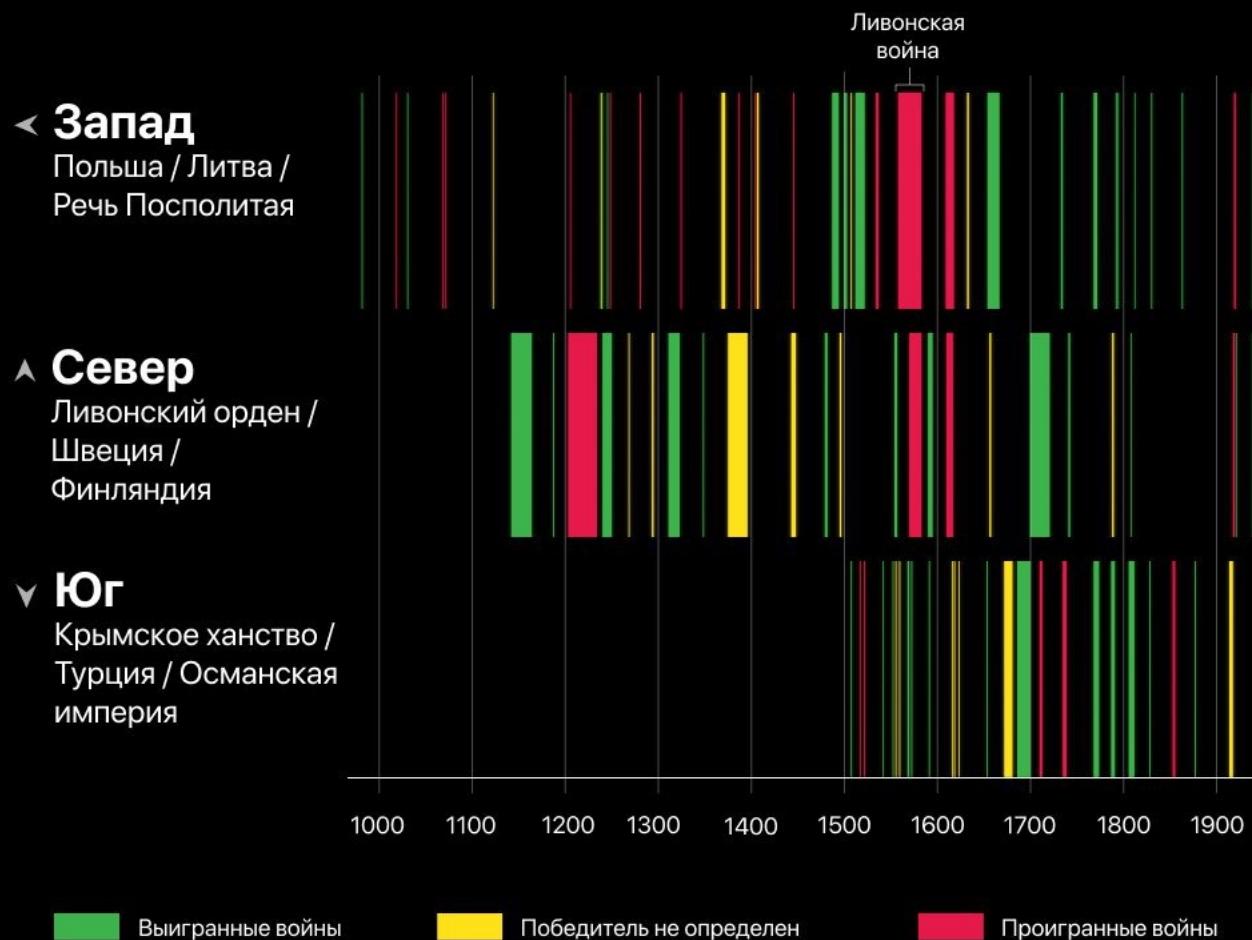
Где живут российские семьи ⓘ

Собственное жильё Аренда Общежитие



Войны России

с главными европейскими конкурентами, 982–1941

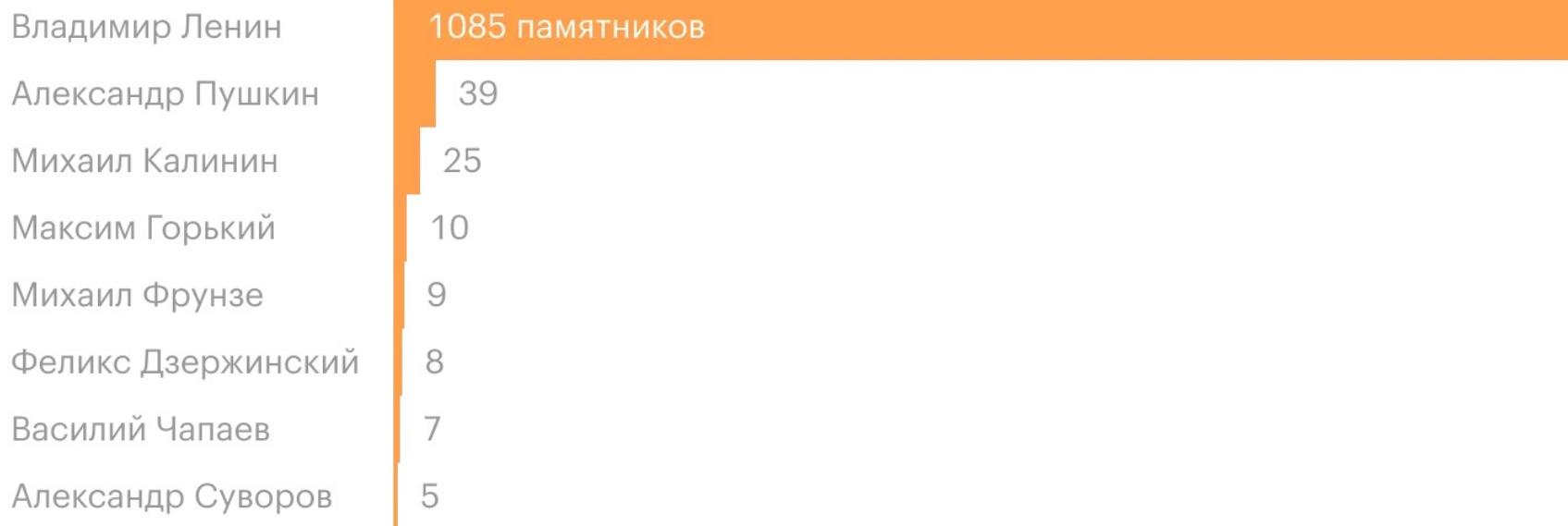


Правопреемники, датировка, участники и победители войн — по историографии, сложившейся в русской Википедии летом 2019 года.

Самые популярные виды графиков и как их правильно использовать

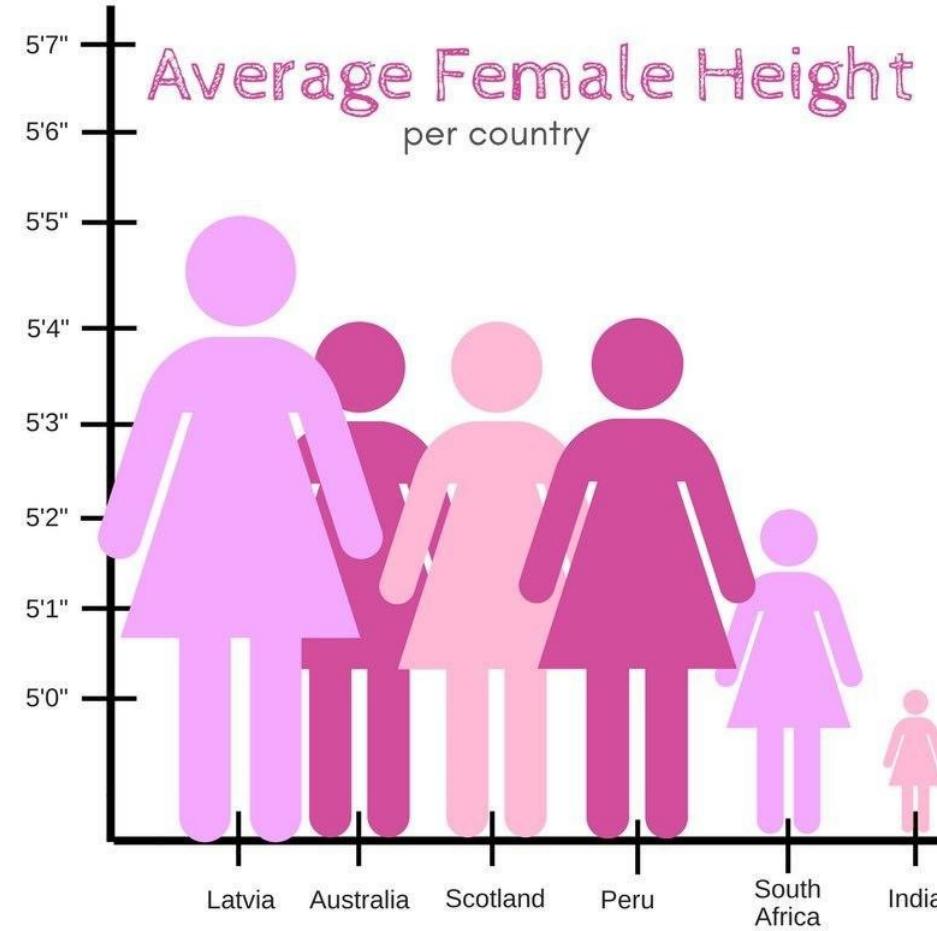
Барчарт

Кому посвятили больше всего памятников в России



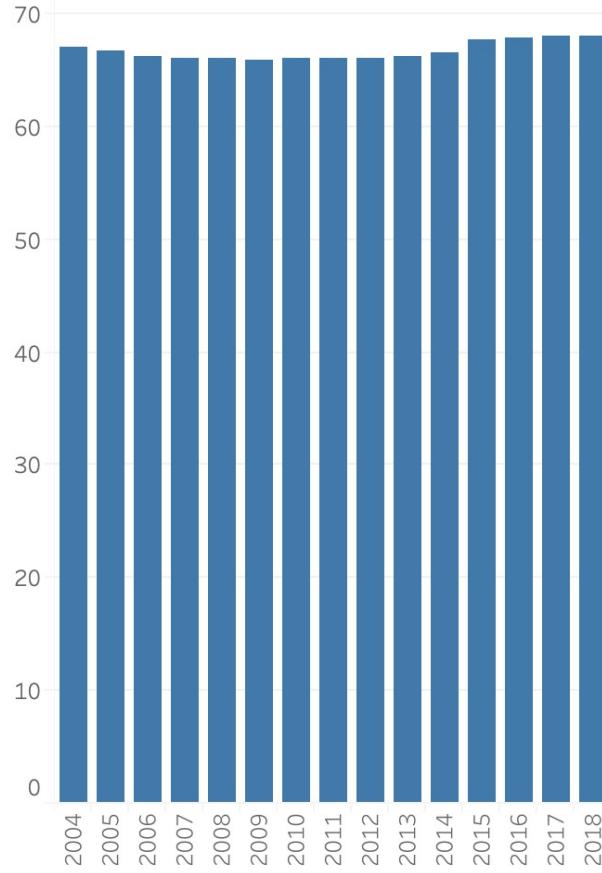
Источник: [Минкульт](#)

Всё ли в порядке?



Как показать изменение

Количество мужчин в РФ, млн. чел

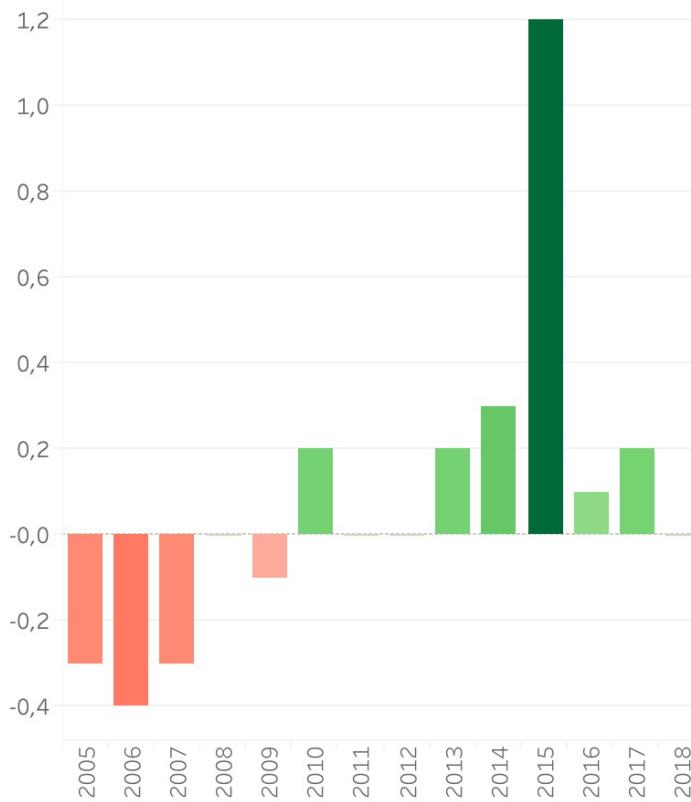


Не очень:

Изменение величины мало по сравнению с самой величиной и плохо воспринимается

Как показать изменение

Как менялось количество мужчин по сравнению с прошлым годом, млн. чел

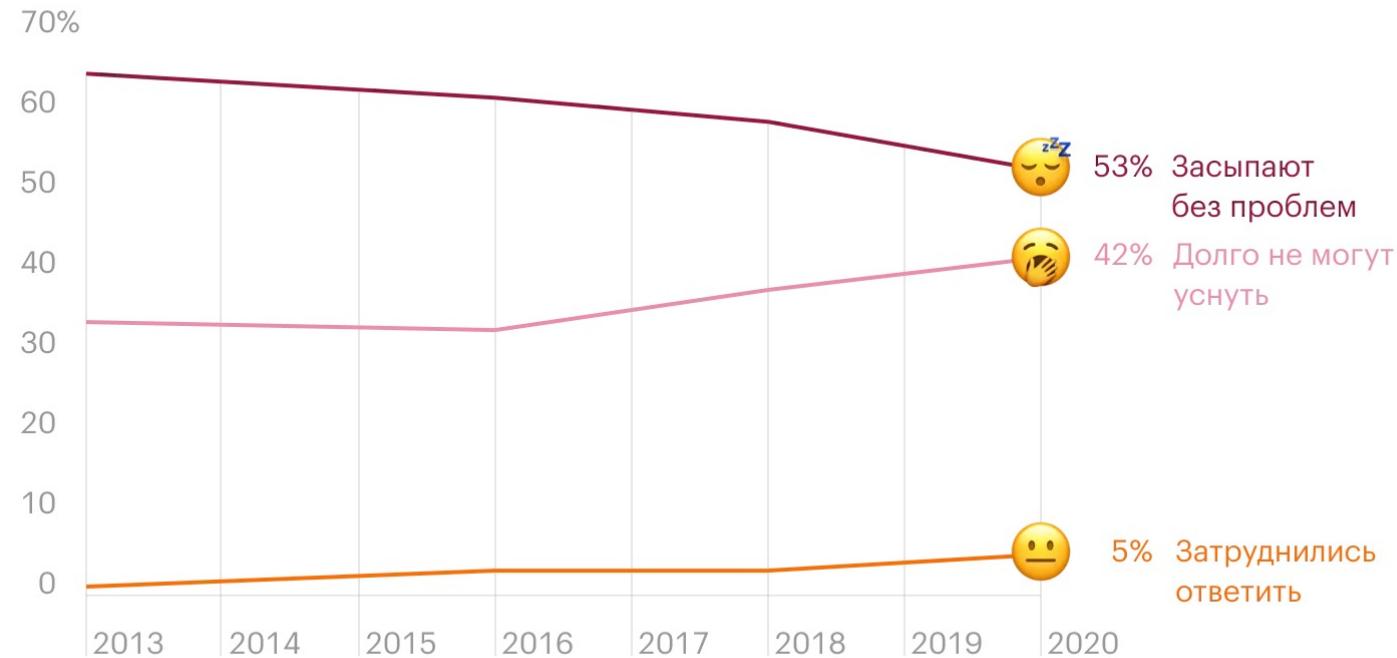


Да:

В высоту бара можно закодировать не саму величину, а её **изменение**

Линейный график

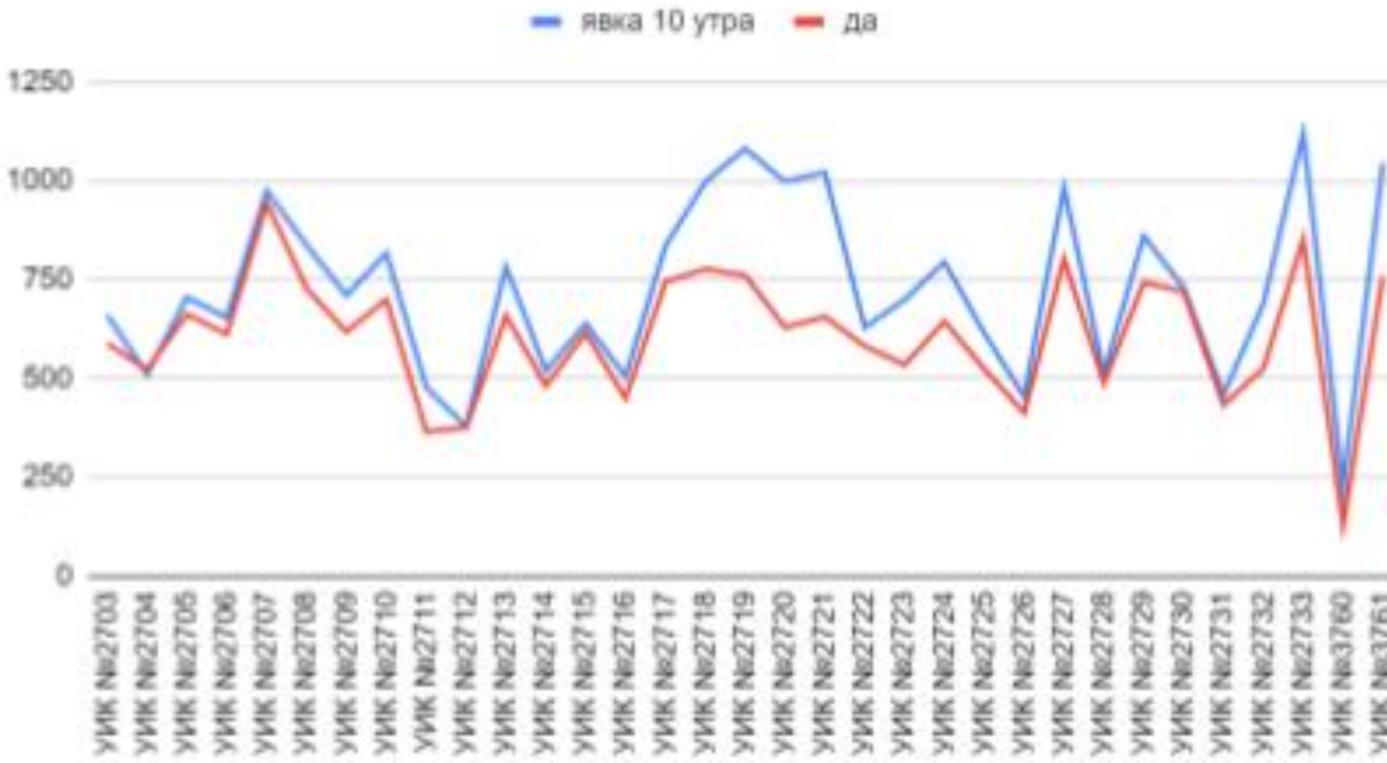
Как меняется доля россиян, которые хорошо спят



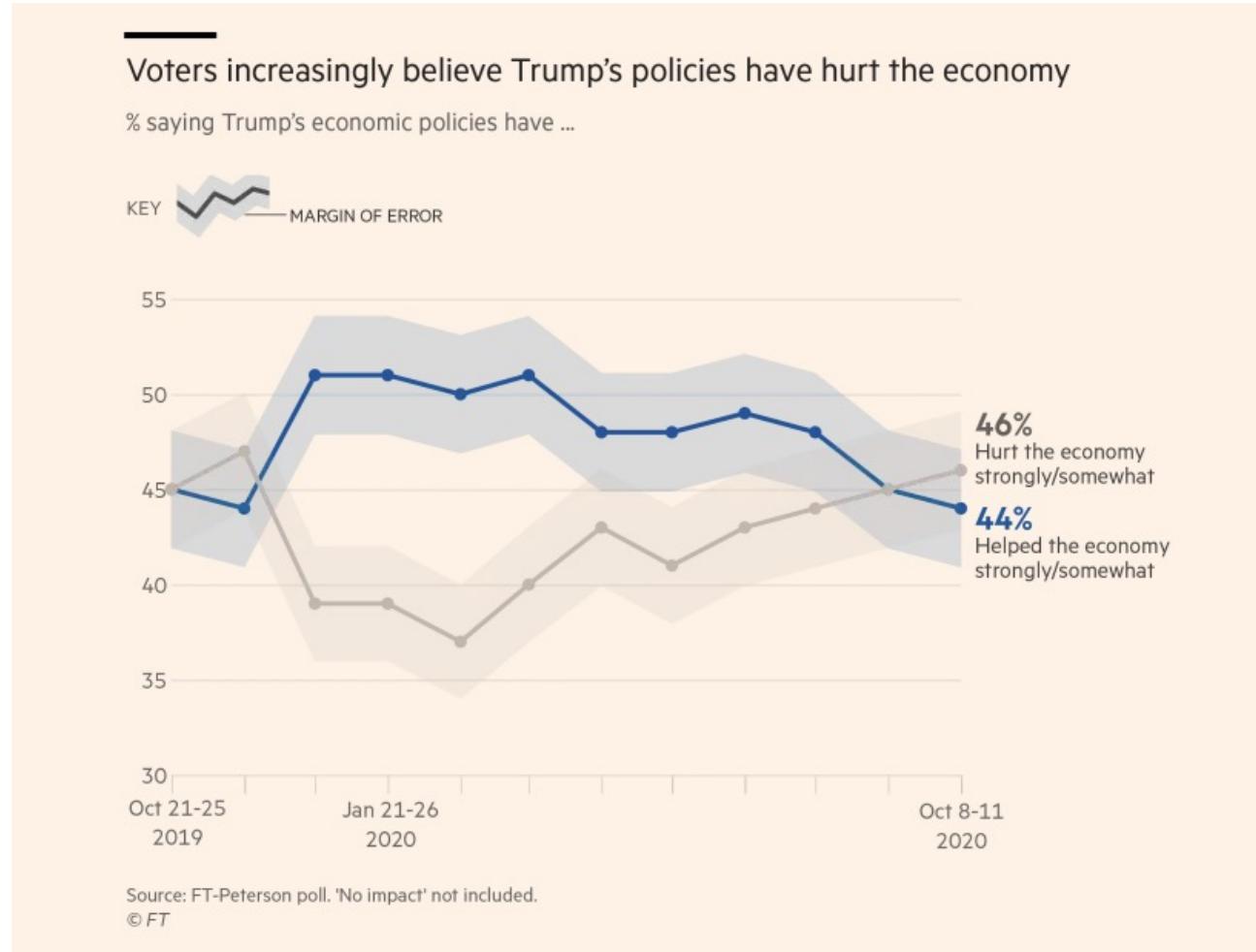
Источник: [ВЦИОМ](#)

Только для времени

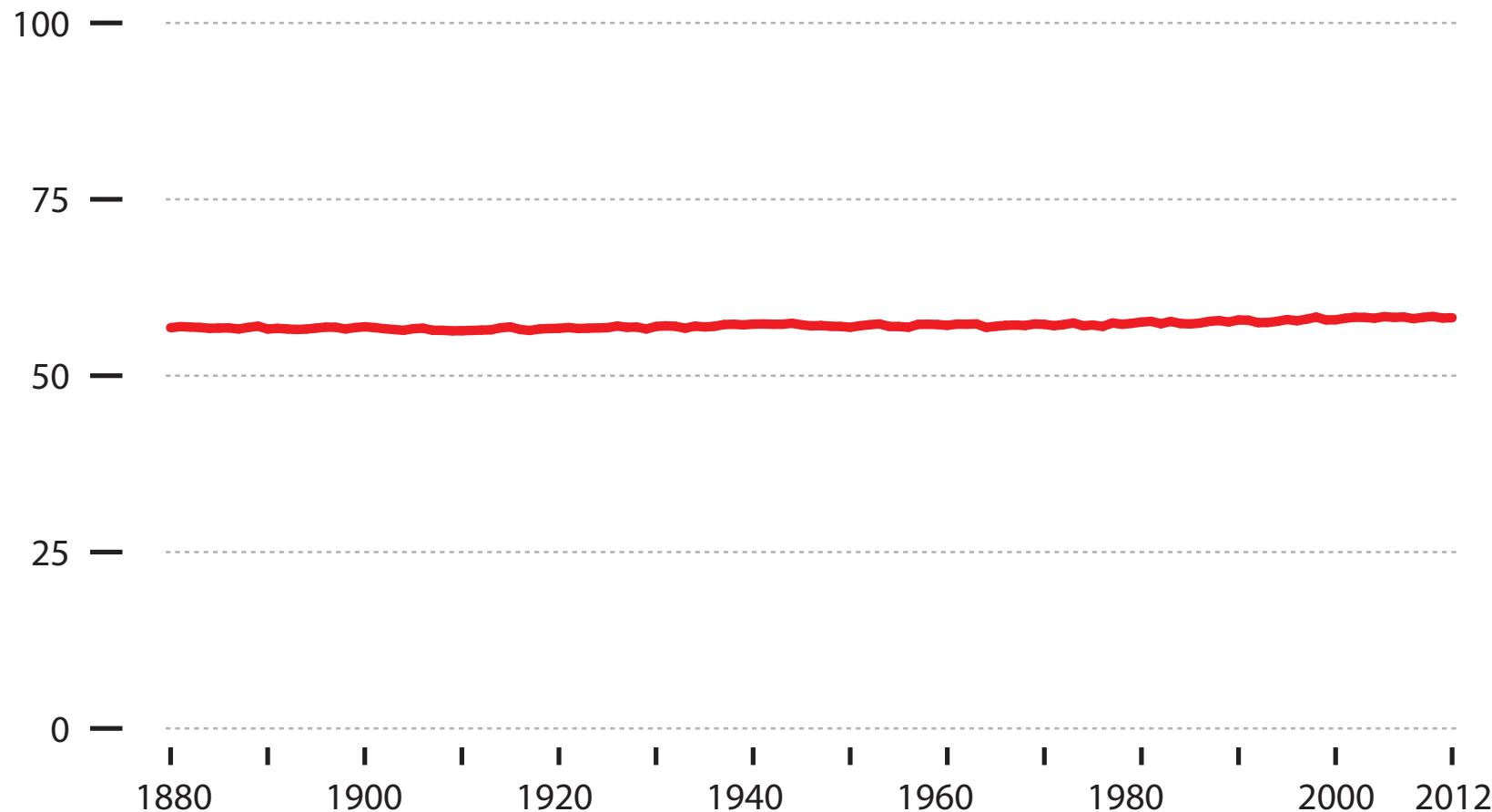
Очаково



Можно не от нуля

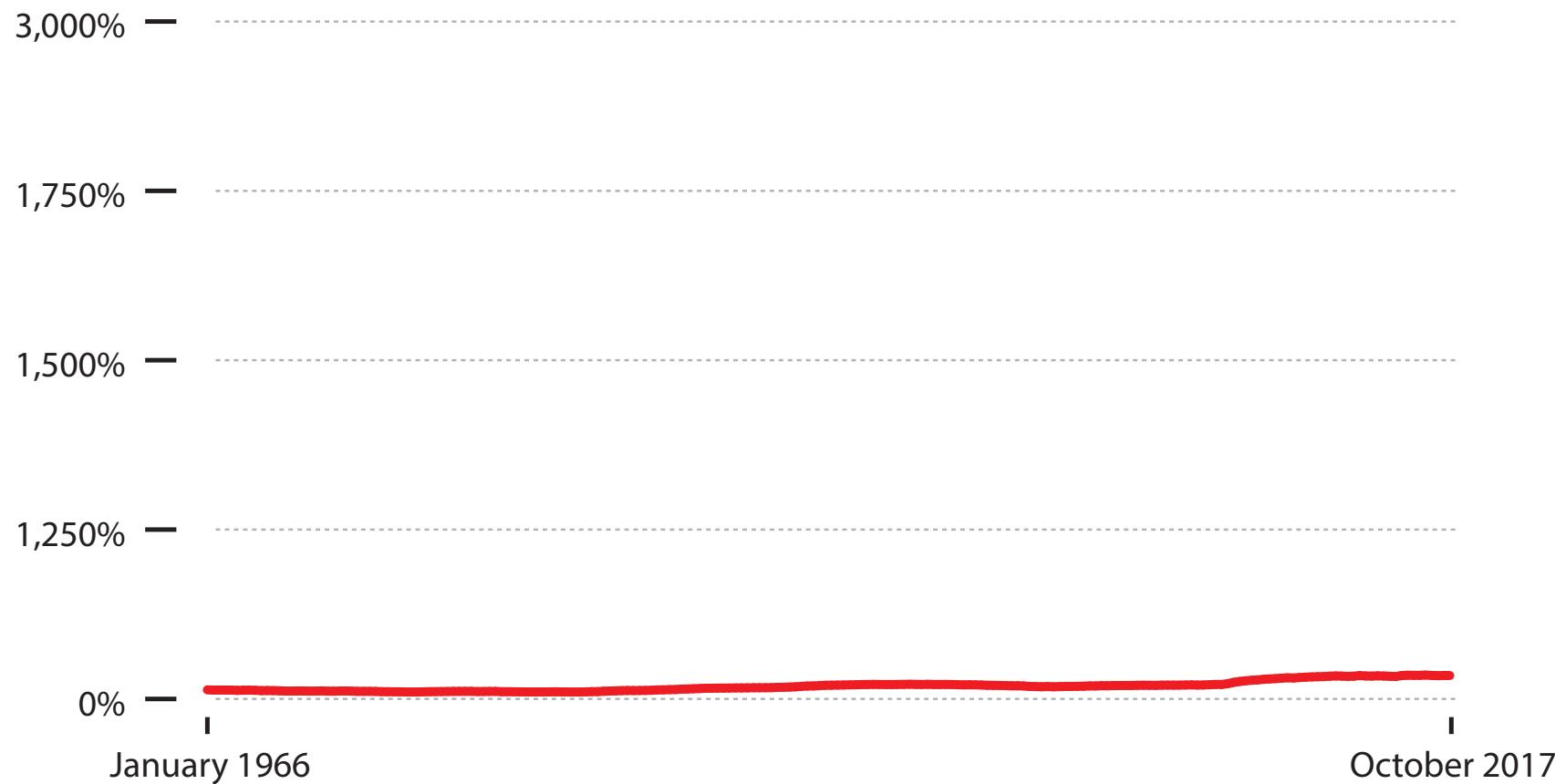


Average annual global temperature in degrees Fahrenheit

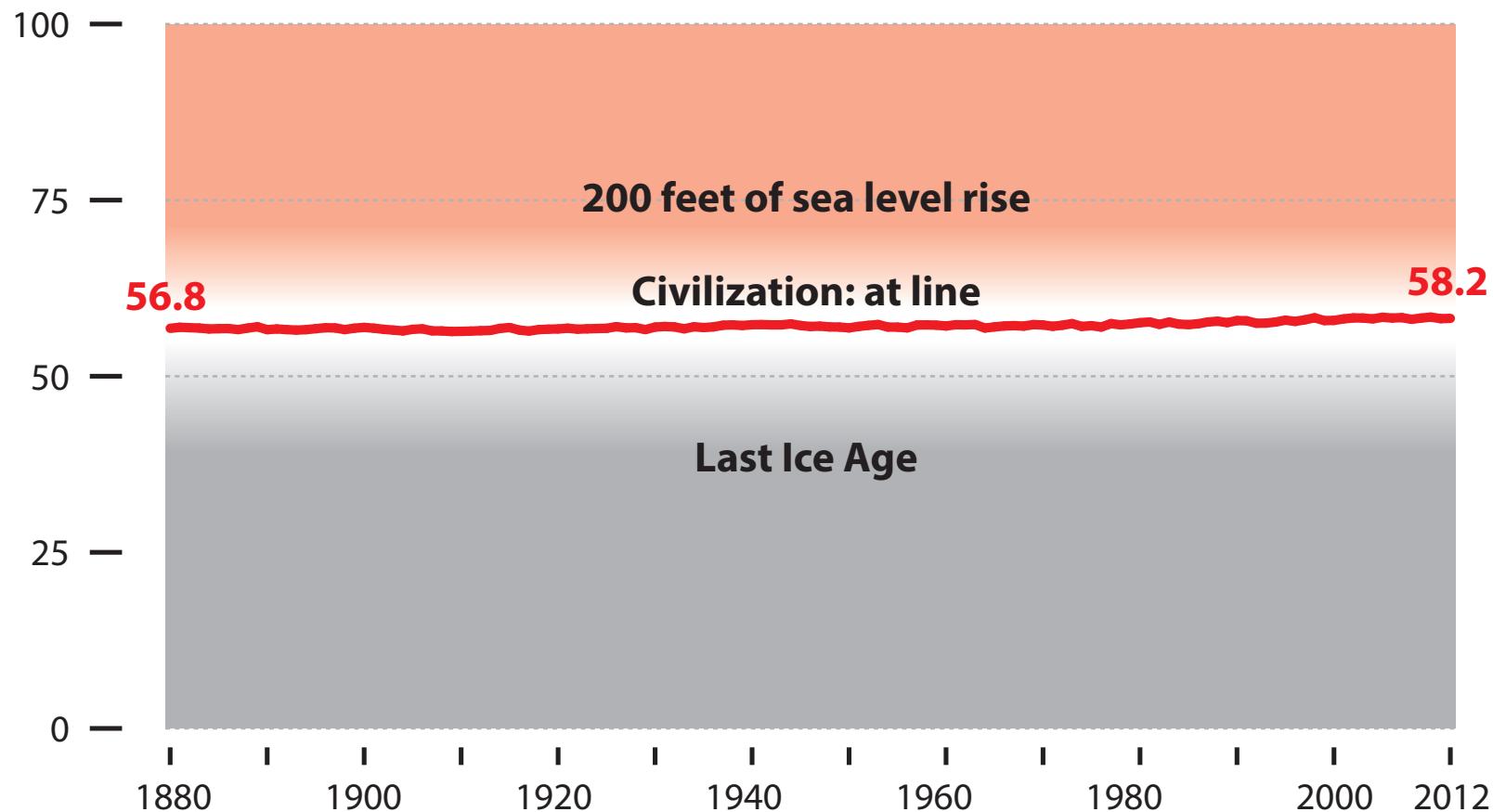


Alberto Cairo | How Charts Lie

Federal debt: total public debt as percent of gross domestic product



Average annual global temperature in degrees Fahrenheit



Average annual global temperature in degrees Fahrenheit

59 —

58 —

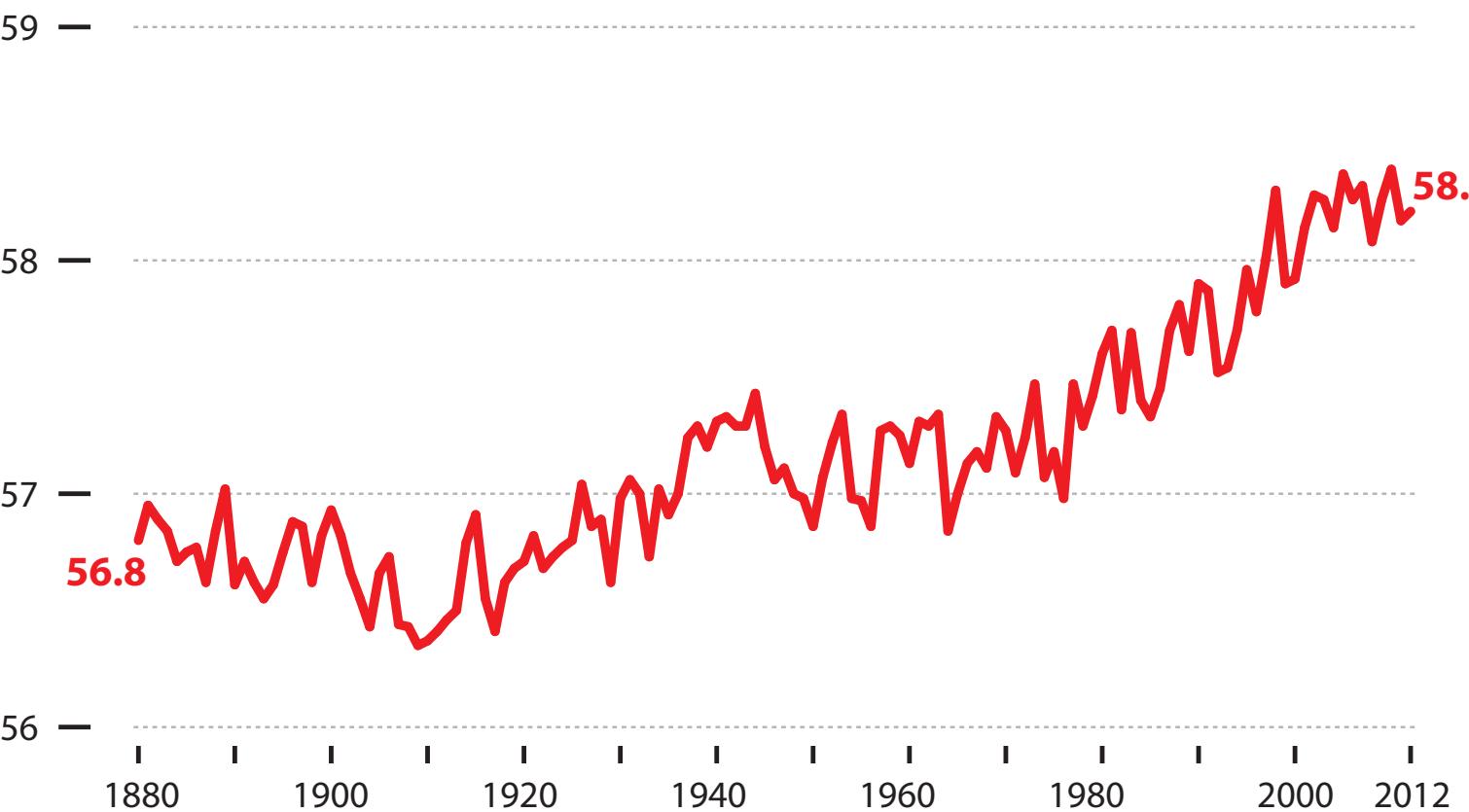
57 —

56 —

1880 1900 1920 1940 1960 1980 2000 2012

56.8

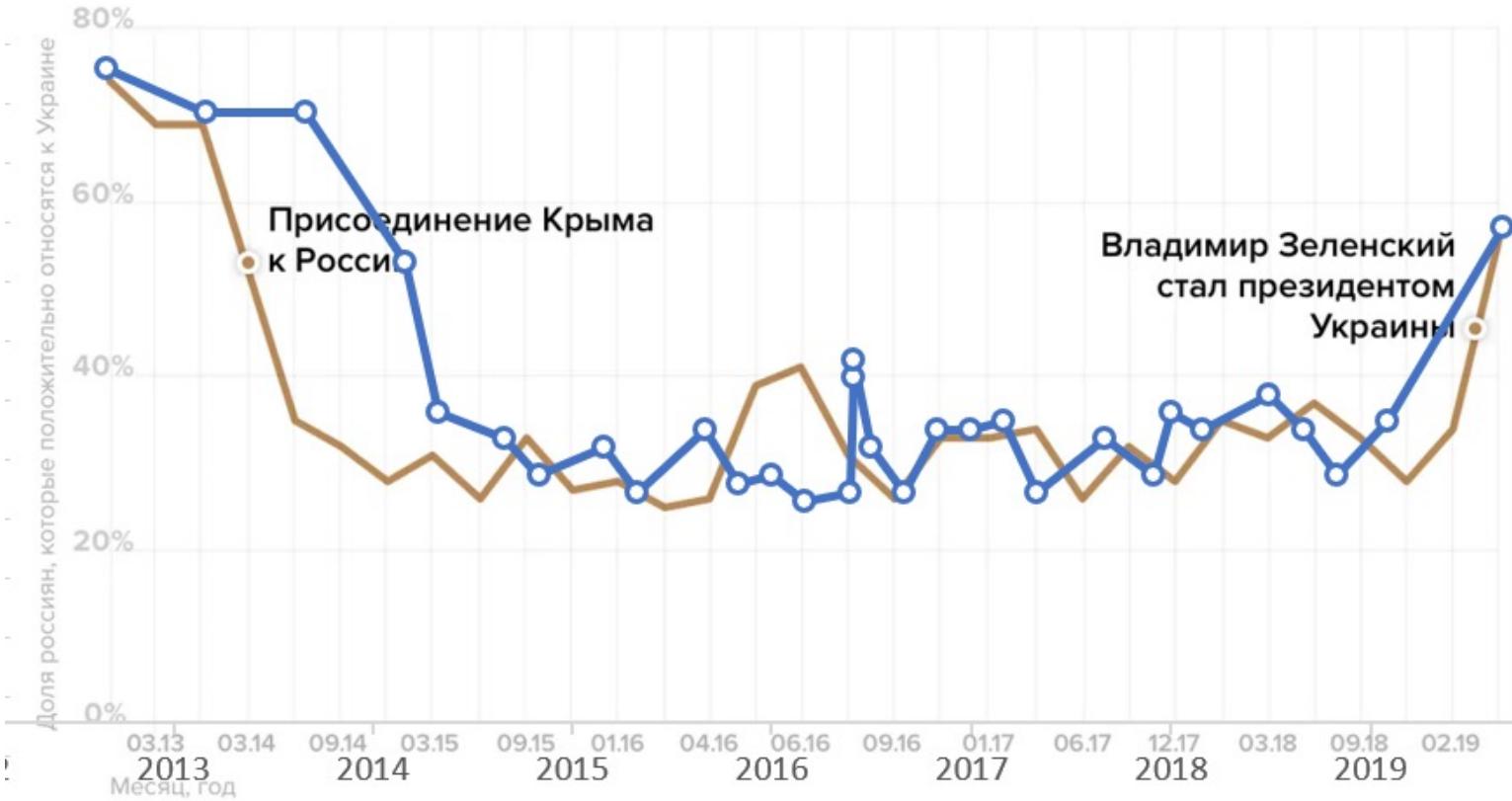
58.2



Что не так с этим графиком?



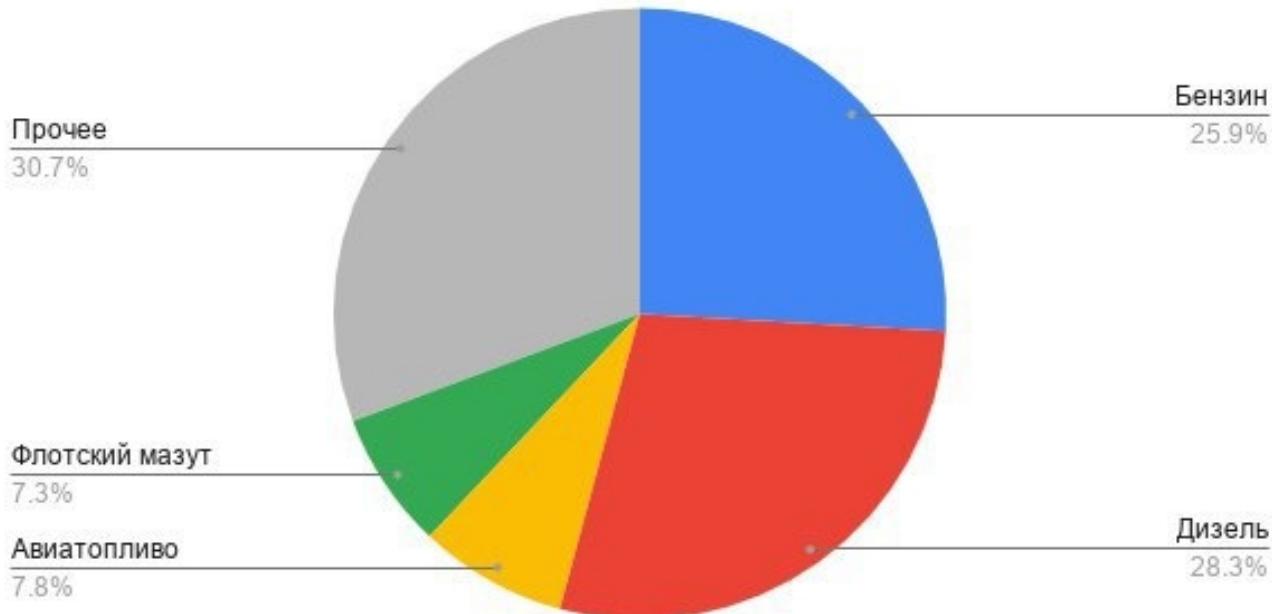
Ось времени нельзя деформировать



Пайчарт

Что Земля делает из нефти?

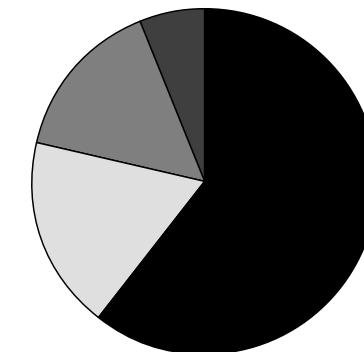
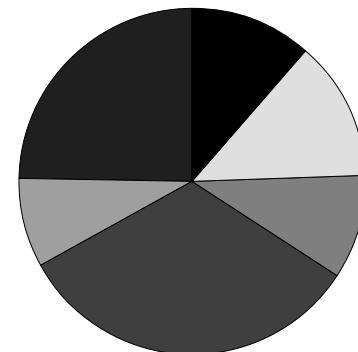
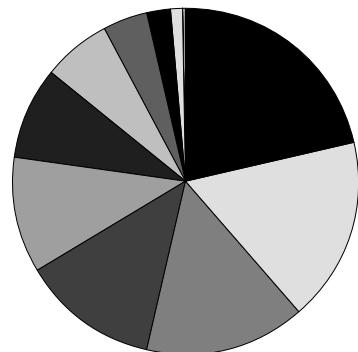
BP Statistical Review of World Energy June 2019



**Контекст применения – часть от
целого!**

Правила пайчарта

1. Сектора пайчарта – это части целого и в сумме **всегда** должны давать 100%
2. Лучше не использовать больше 5-6 секторов (можно объединять мелкие сектора в «другое»)
3. Сектора должны идти по часовой стрелке от большего к меньшему. Сектор «другое» всегда последний.



Когда использовать пайчарт

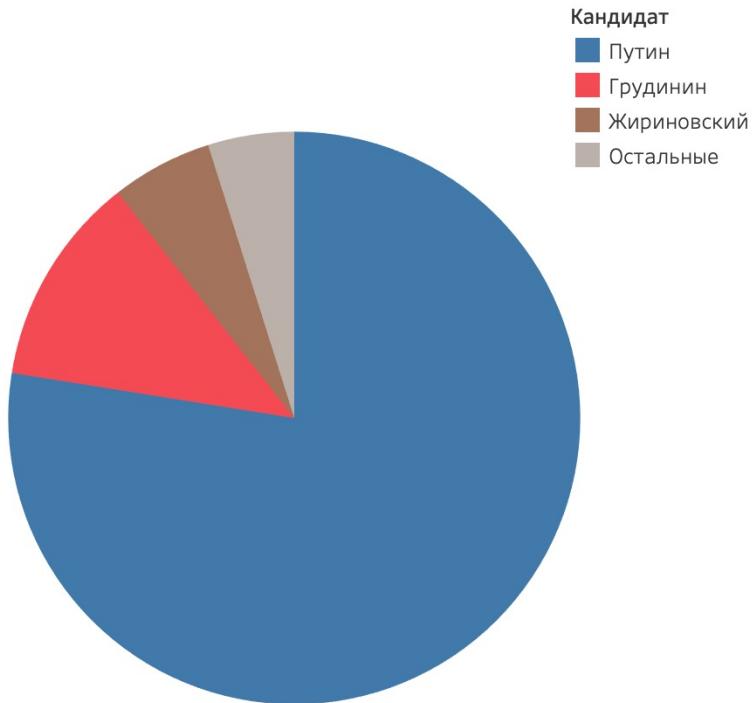
Сколько голосов набрали кандидаты в президенты
2018 года



Если нужно **сравнивать** доли голосов за разных кандидатов, лучше использовать бары

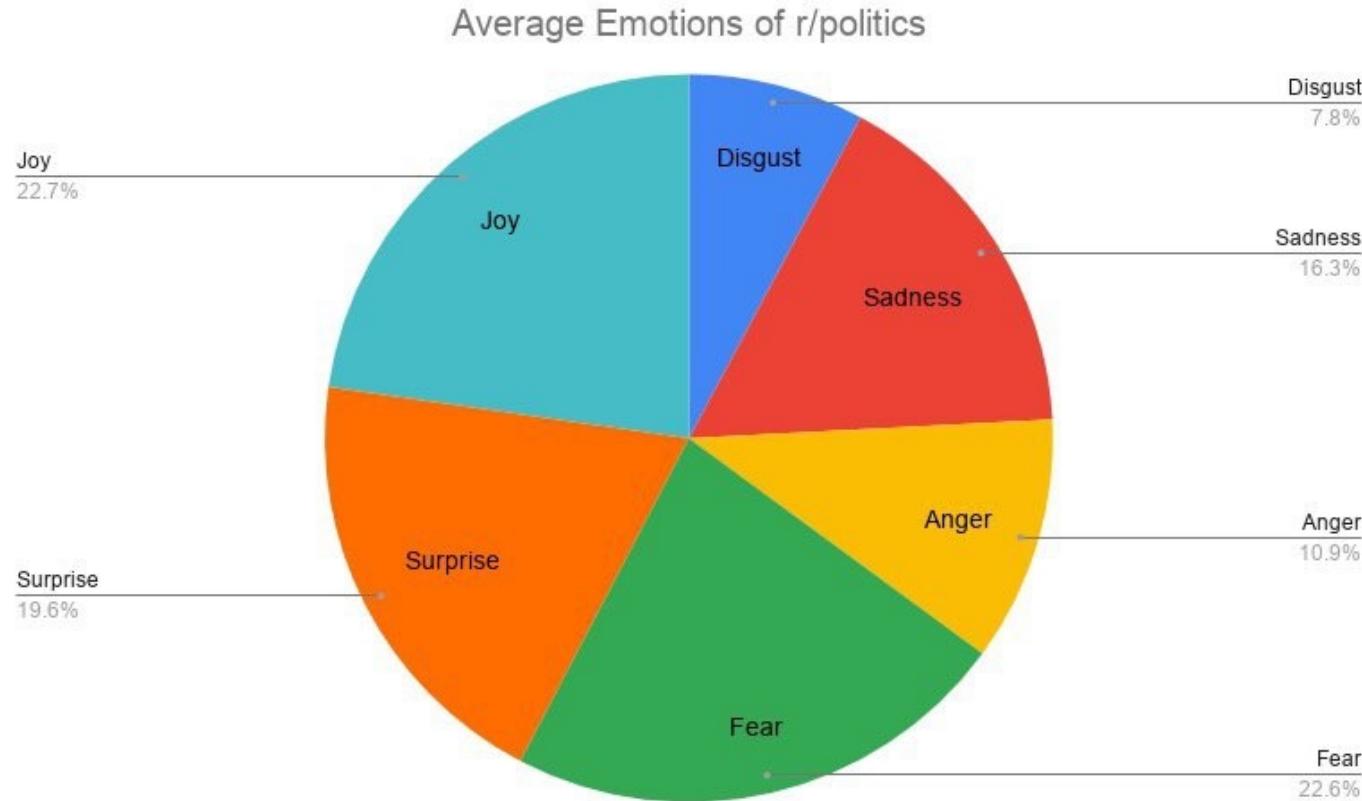
Когда использовать пайчарт

Путин набрал 76% голосов на выборах президента 2018 года



Если показать, что Путин набрал очень много голосов в общей доле – лучше пайчарт

Когда использовать пайчарт



Помимо того, что тут нарушены все правила расположения секторов, пайчарт использовать бессмысленно, потому что сравнить величины крайне сложно

Можно показывать только 100%



Да

Национальный состав класса:
русские – 76%, украинцы – 15%,
татары – 6%, остальные – 2%

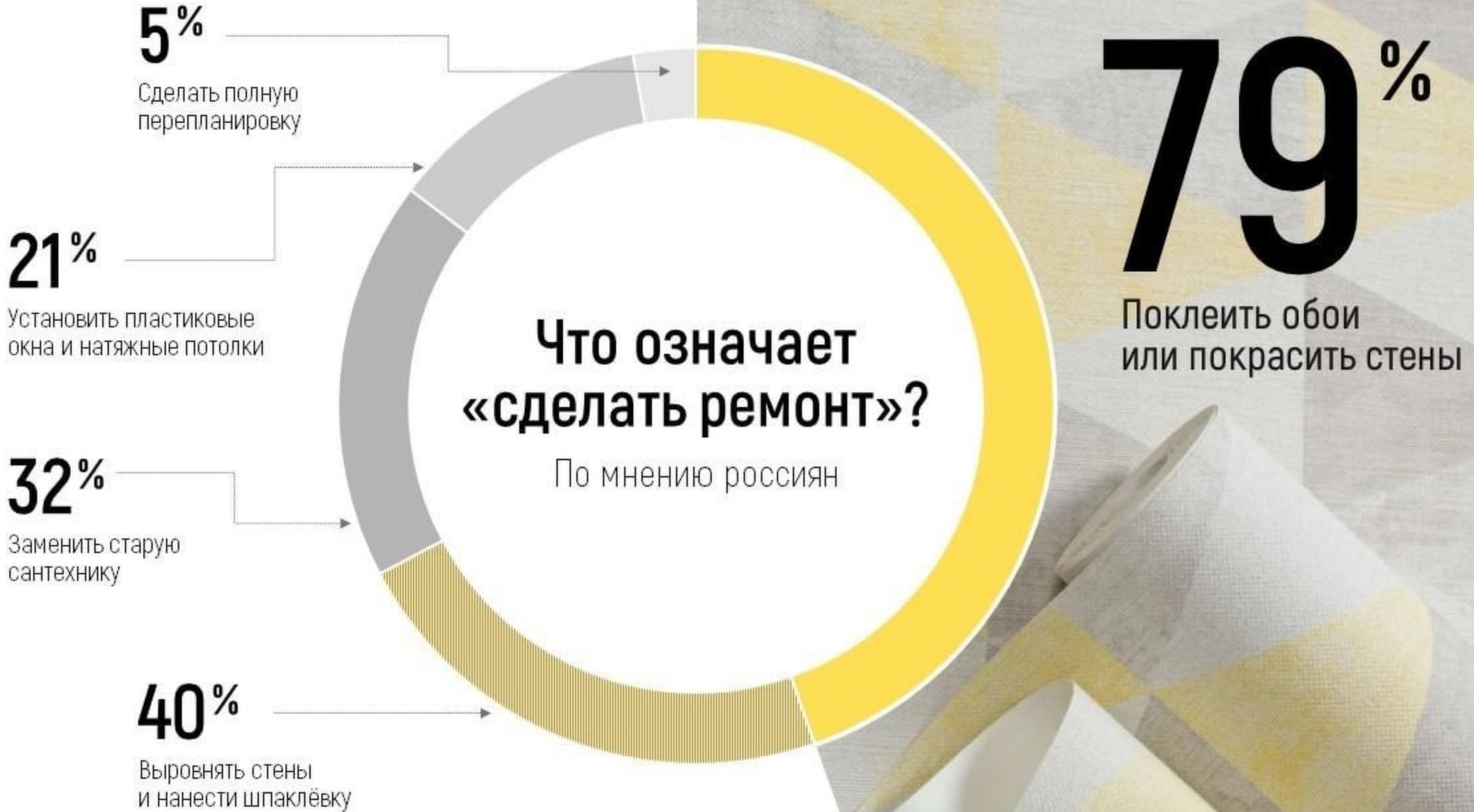


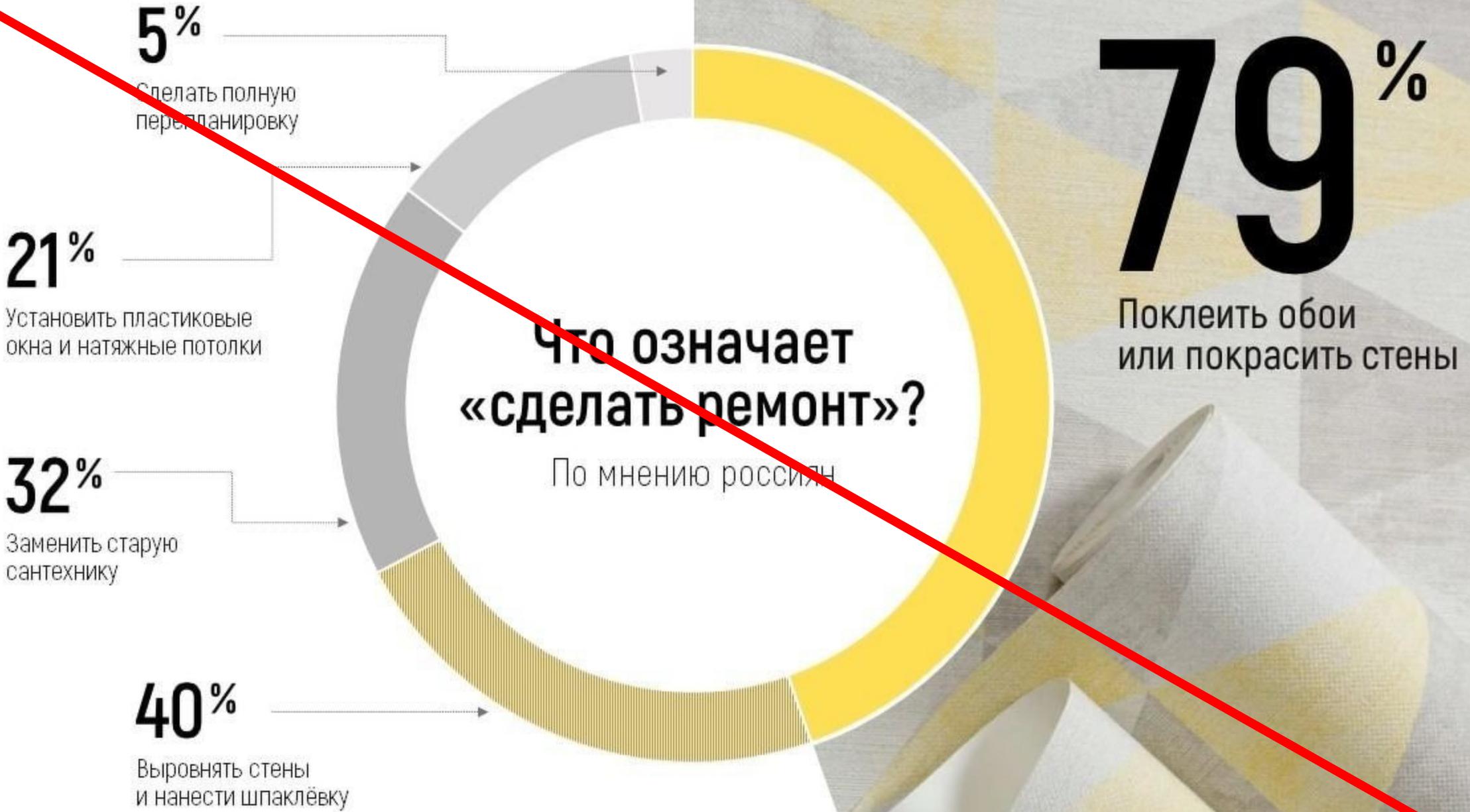
Нет

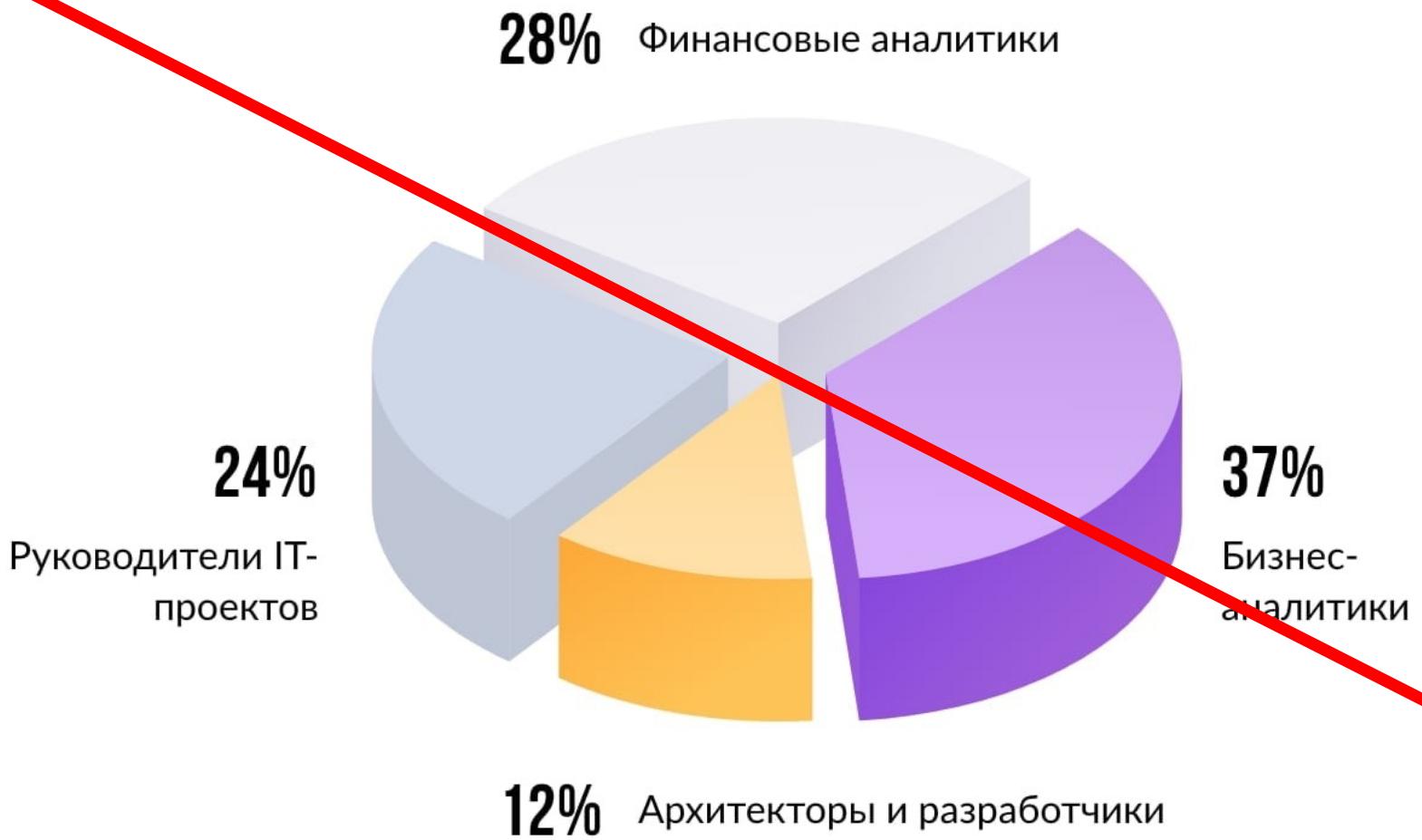
Доля людей, одобряющих действия
правительства: среди лиц, имеющих
высшее образование – 15%, среди
лиц, имеющих среднеспециальное
образование – 30%, среди лиц без
образования – 55%

Результаты выборов президента
России: Путин – 77%, Грудинин –
12%, Жириновский – 6%, остальные
– 5%

Топ-3 кандидата на выборах
президента России: Путин – 77%,
Грудинин – 12%, Жириновский –
6%







**Как сделать график
понятным?**

Принцип простоты

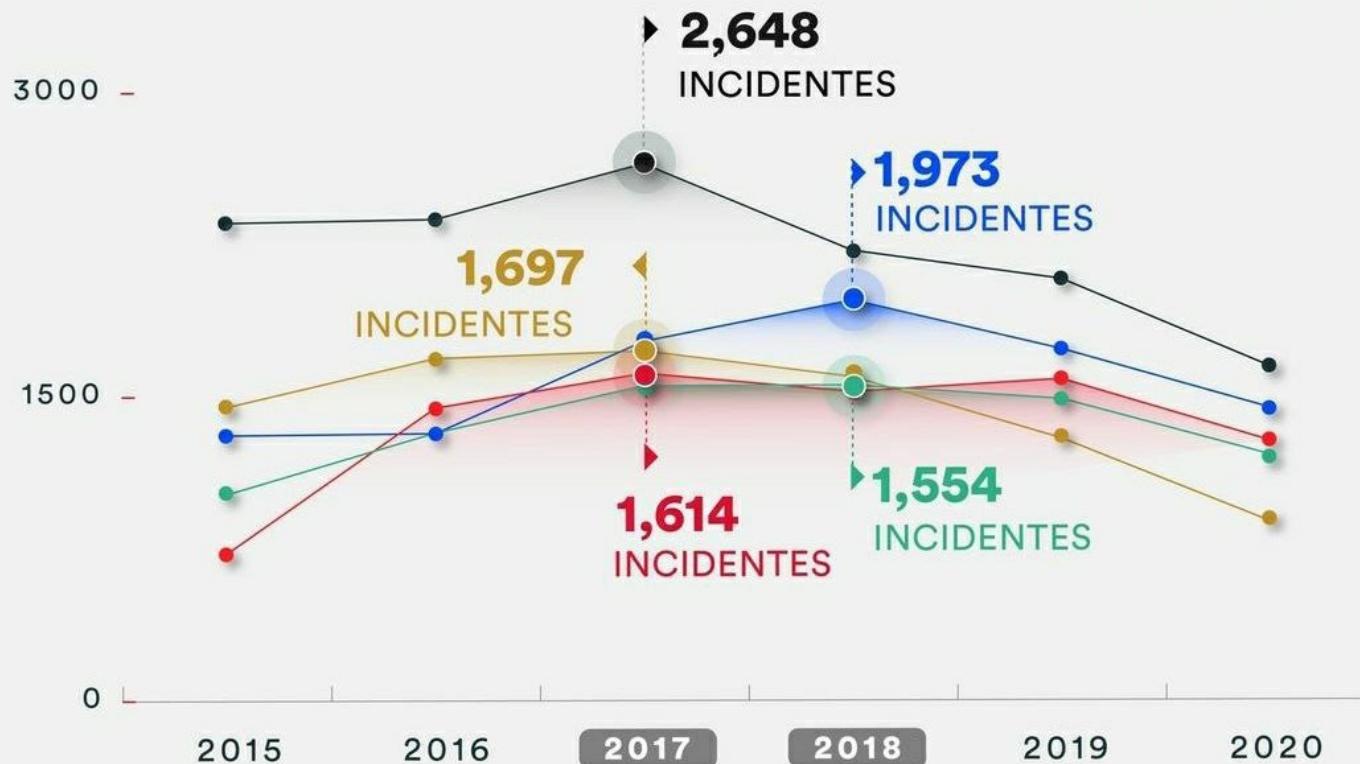
- Графики не должны содержать элементов, которые бессмысленно затрудняют восприятие
- Несколько сложных мыслей лучше разносить на несколько графиков
- Упорядоченность = удобство

HISTÓRICO:

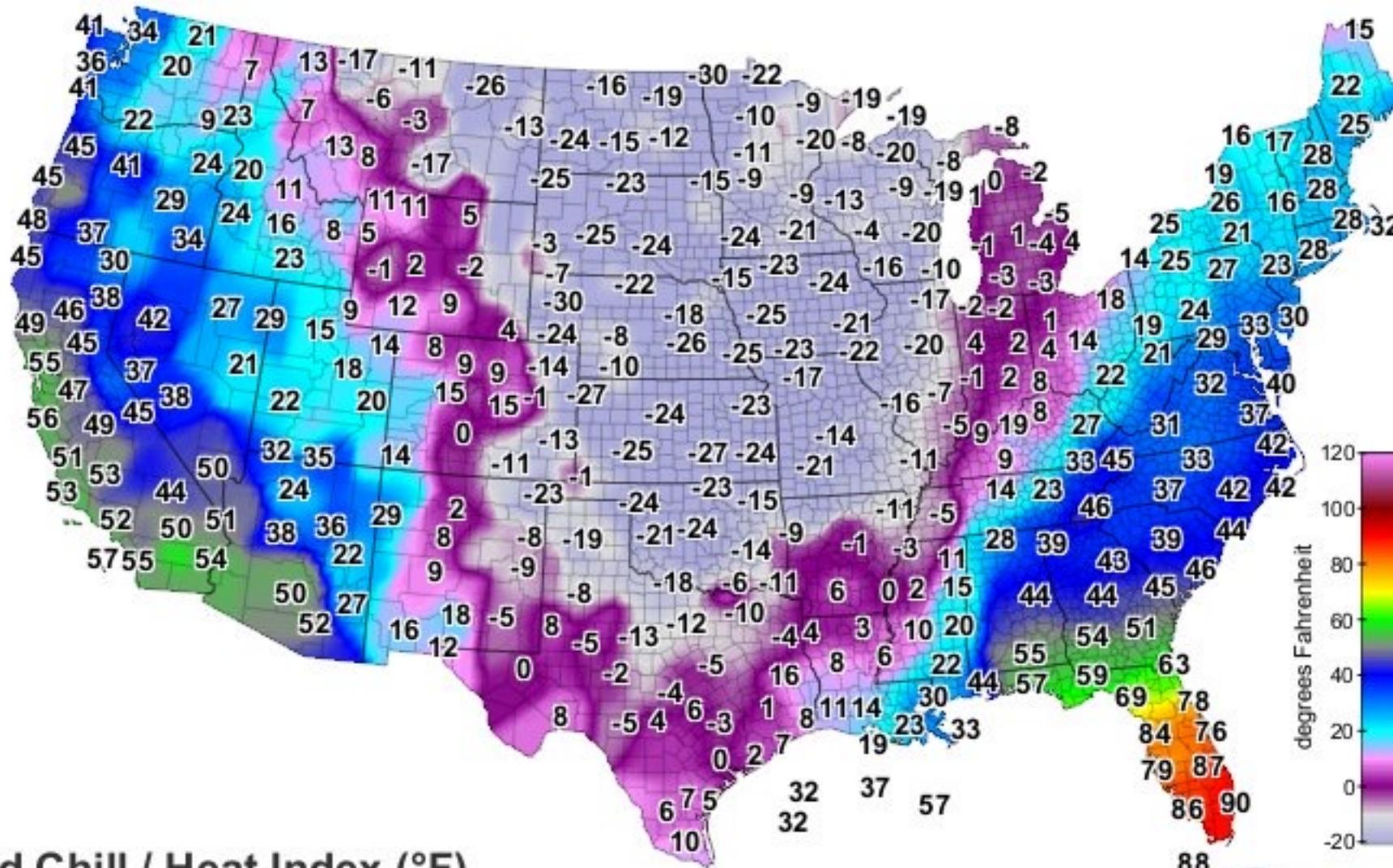
de los estados más violentos del 2015-2020



- Baja California ●
- Aguascalientes ●
- Tabasco ●
- Colima ●
- Querétaro ●



Тени, точки, хаотично
налепленные выноски
значительно
затрудняют восприятие



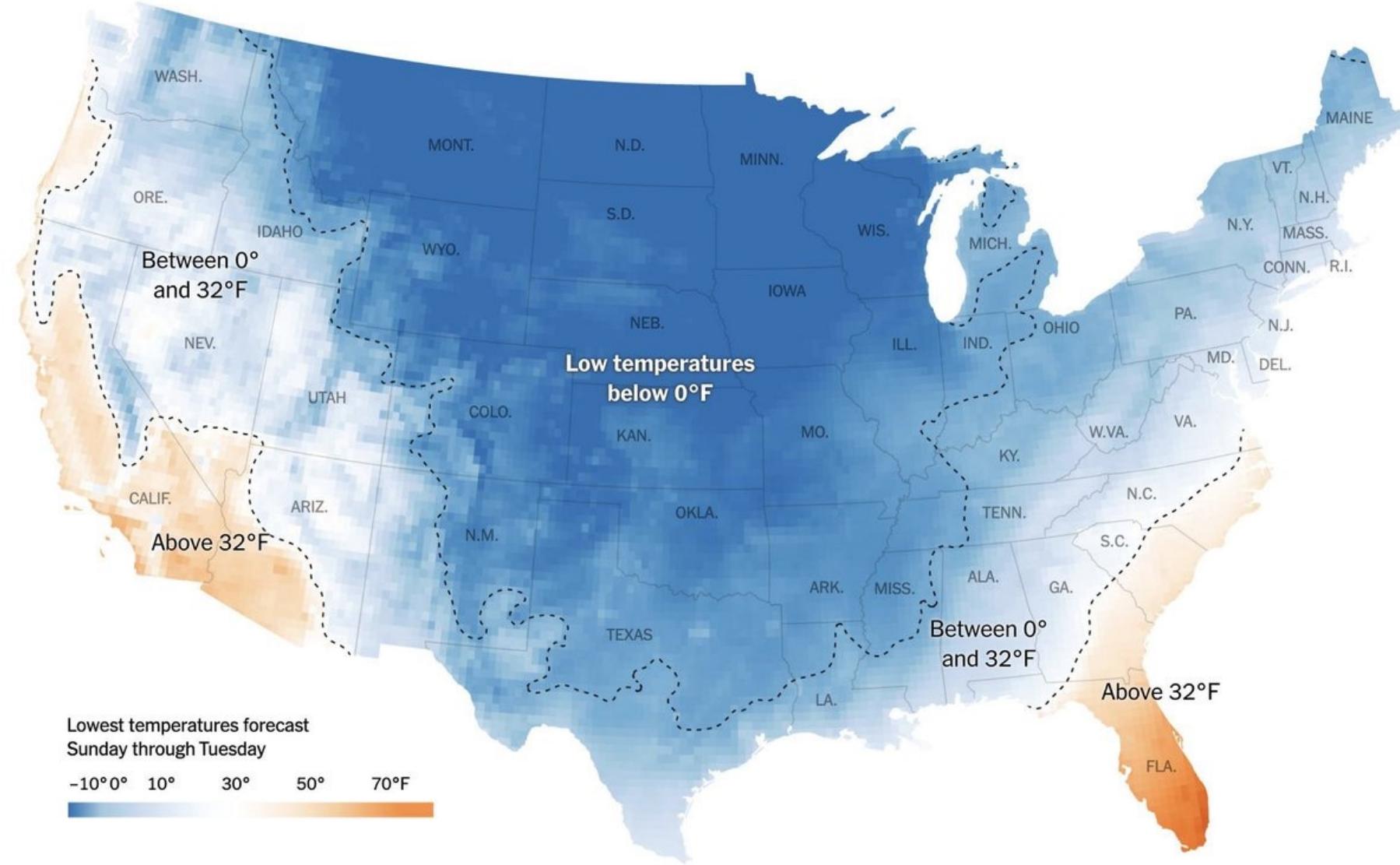
Wind Chill / Heat Index (°F)

11:00 AM February 15, 2021 CST

Data provided by NOAA's National Weather Service. Created 11:10:34 AM February 15, 2021 CST. © Copyright 2021

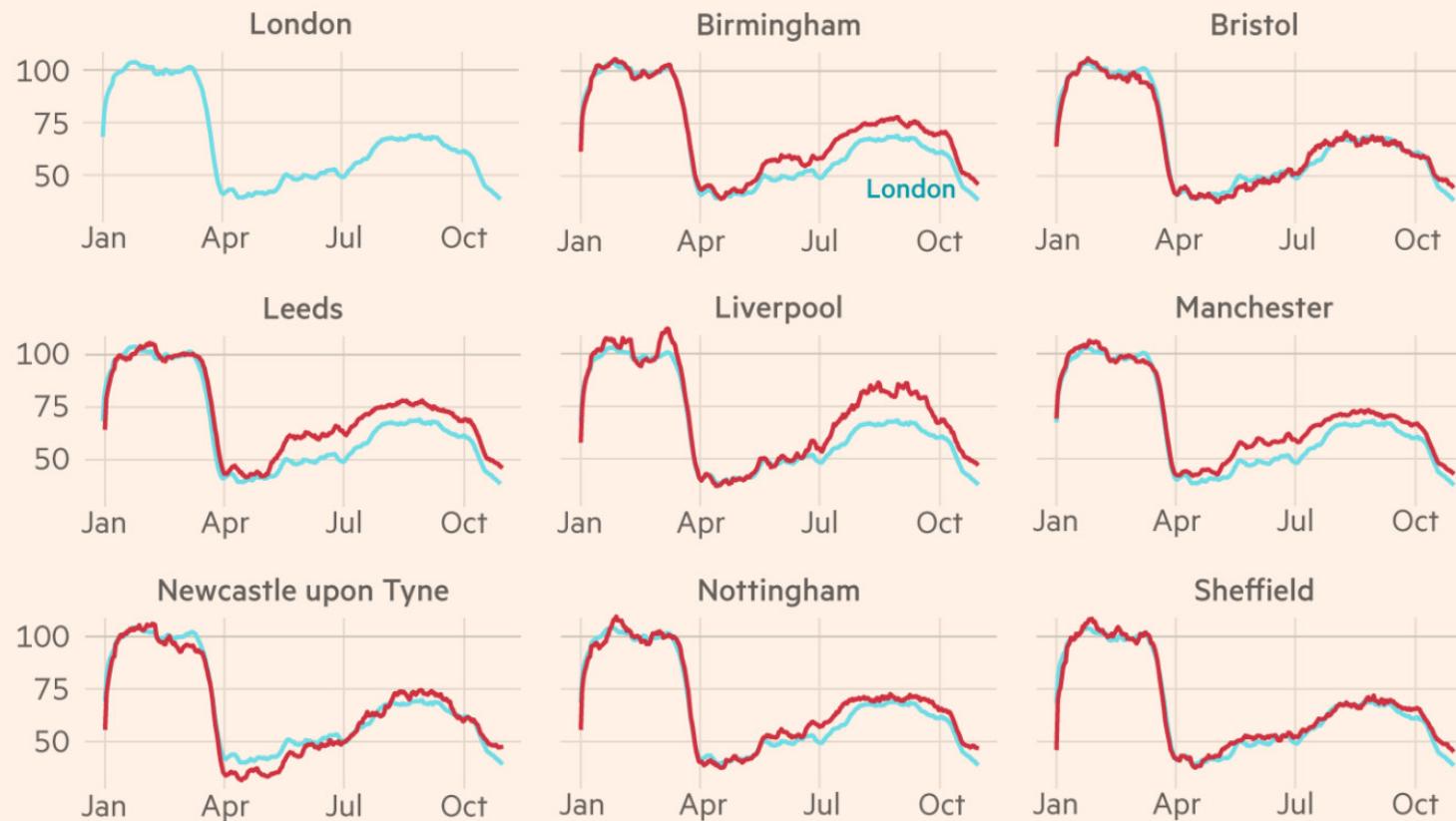


Mesonet



Staff in London have avoided a return to the office

Daily count of visits to offices indexed to their equivalent levels pre-lockdown
(rebased, 100 = Jan-Feb weekday median)



Includes industrial units, science and business parks, administrative centres. Daily count is indexed to the median value for the equivalent day of week during Jan-Feb 2020 and displayed using a 7-day rolling average

Source: Huq

© FT

Если бы все линии
были на одном
графике, то случился
бы хаос

Countries That Spend The Most On International Tourism

\$ Billion



Из-за того, что
значения не
отсортированы,
трудно понять, кто
лидирует по числу
трат на туризм

Принцип близости

- Элементы, которые расположены ближе друг к другу, проще обрабатываются мозгом
- Если мы хотим что-то сравнить в первую очередь – стоит расположить это ближе
- Все подписи стоит делать ближе к самим элементам

Сколько времени тратят на домашние дела

Мужчины и женщины 15 лет и старше, живущие с собственными детьми.

■ Мужчины ■ Женщины

На бытовые услуги, мин/сутки



На уход за детьми, мин/сутки



Проще всего
сравнить
мужчин и
женщин

© ТАСС, 2020. Источник: Выборочное наблюдение использования
суточного фонда времени населением 2019, Росстат.

Сколько времени тратят на домашние дела

Мужчины и женщины 15 лет и старше, живущие с собственными детьми.

■ Мужчины ■ Женщины

На бытовые услуги, мин/сутки



На уход за детьми, мин/сутки



Сложнее – будни
и выходные

© ТАСС, 2020. Источник: Выборочное наблюдение использования
суточного фонда времени населением 2019, Росстат.

Сколько времени тратят на домашние дела

Мужчины и женщины 15 лет и старше, живущие с собственными детьми.

■ Мужчины ■ Женщины

На бытовые услуги, мин/сутки



На уход за детьми, мин/сутки

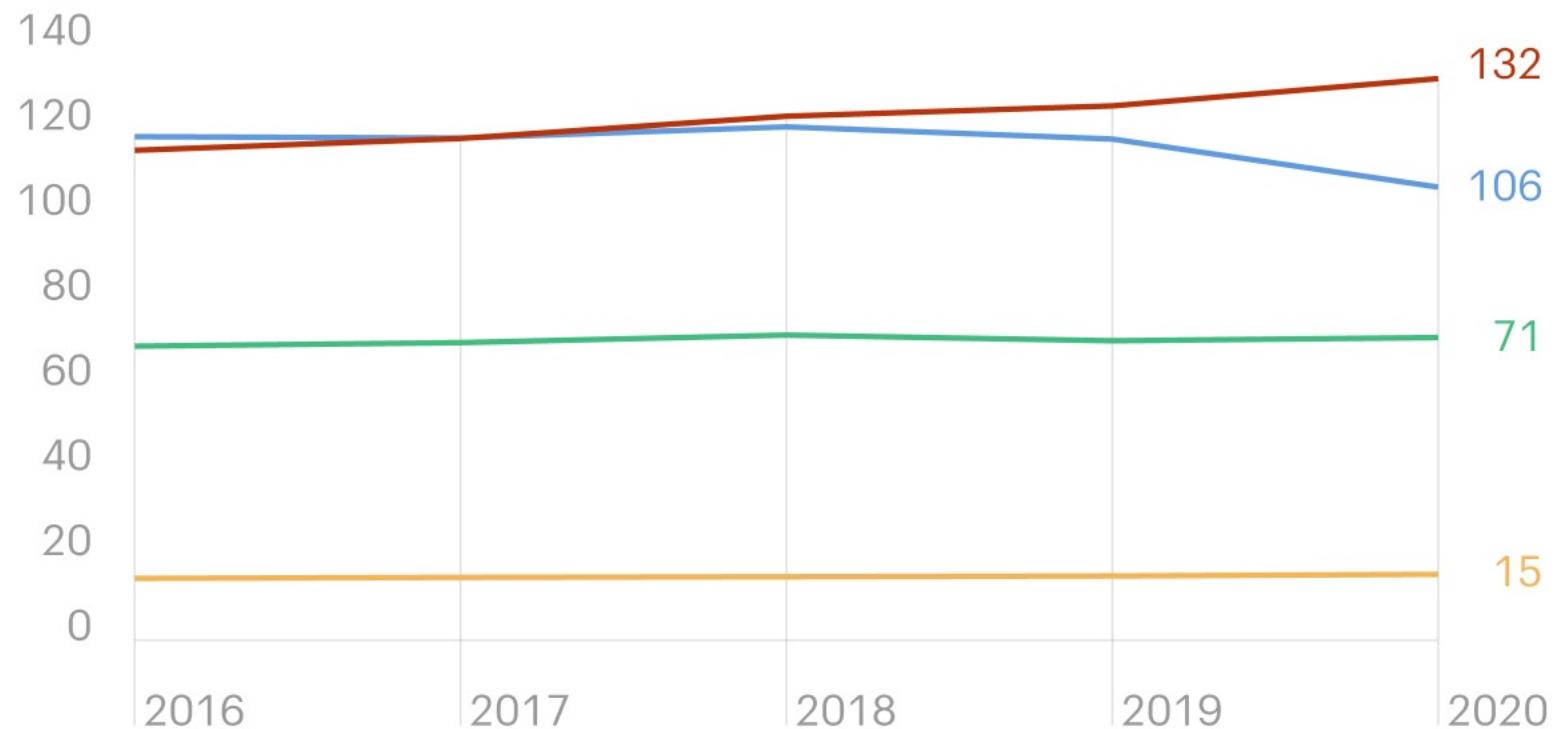


Максимально
сложнo: вид
занятий

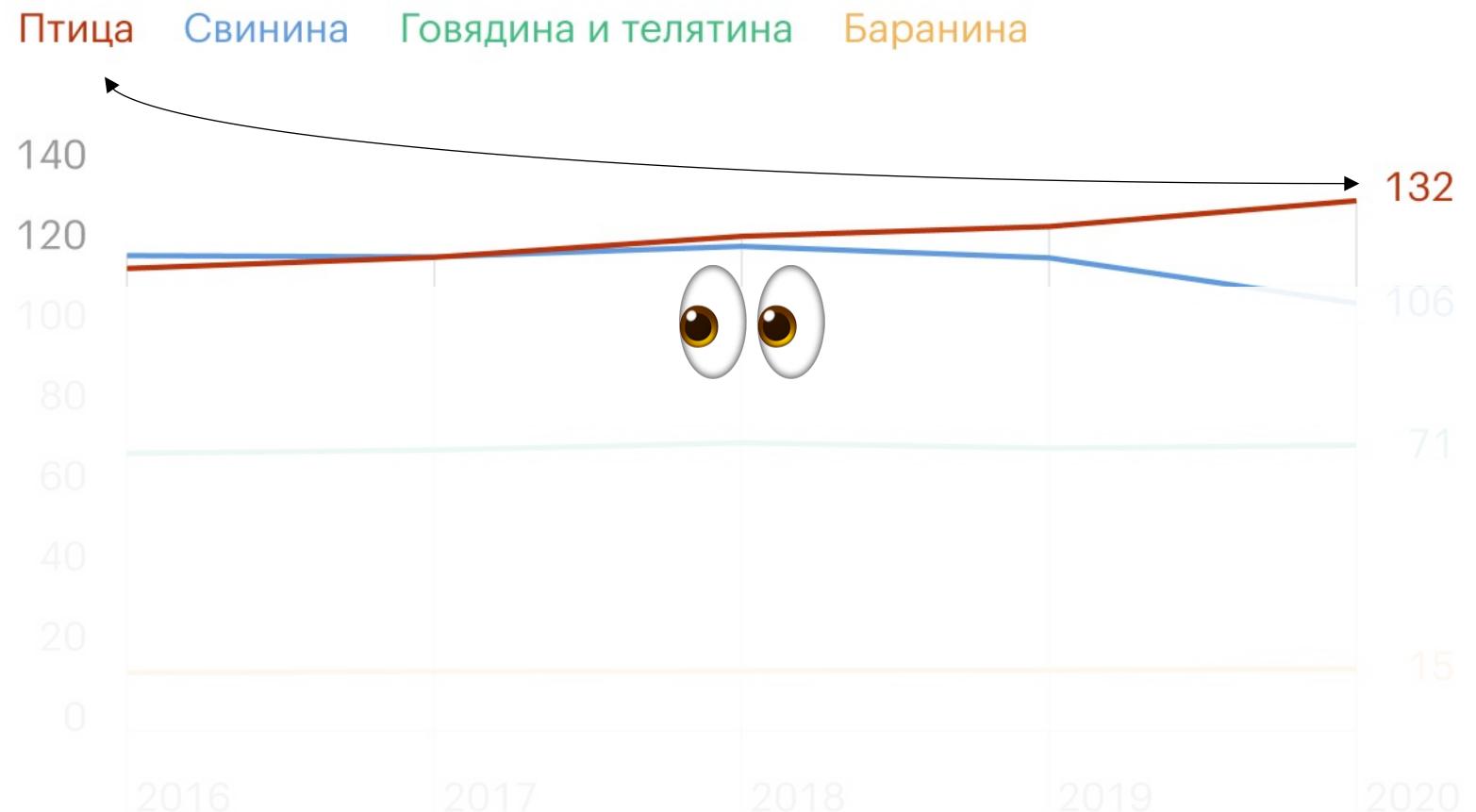
© ТАСС, 2020. Источник: Выборочное наблюдение использования
суточного фонда времени населением 2019, Росстат.

Сколько мяса покупают во всем мире, млн тонн

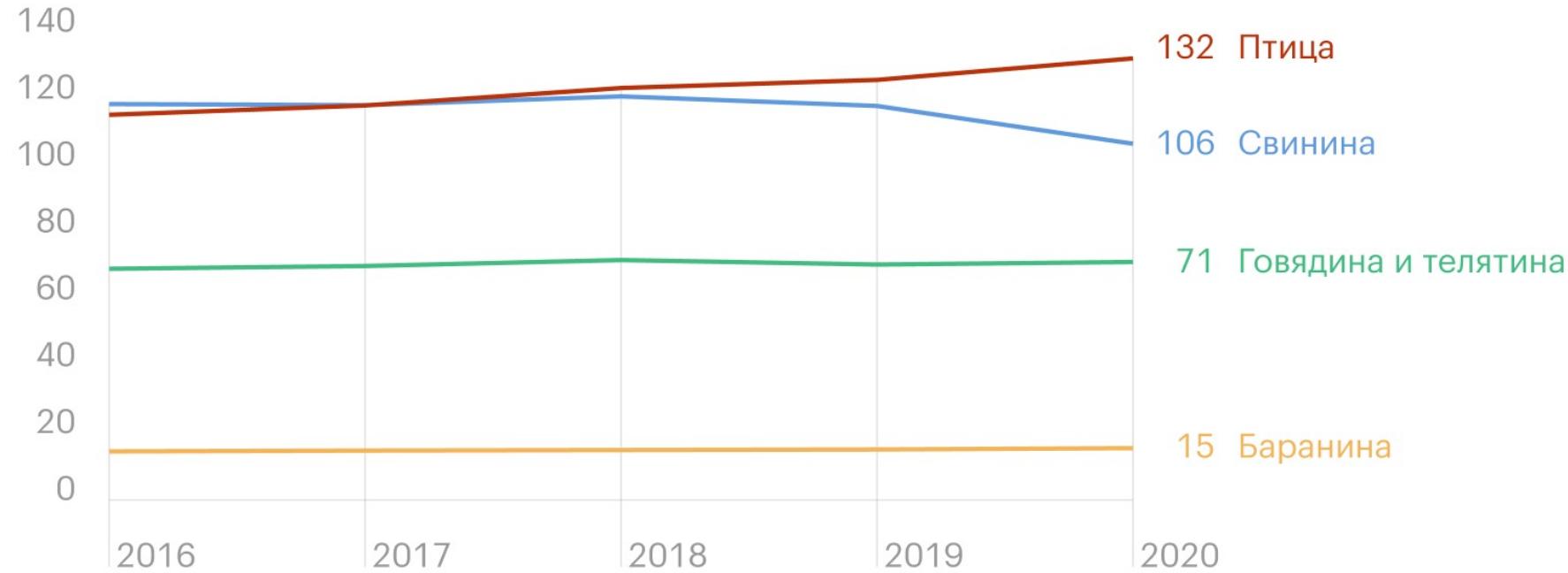
Птица Свинина Говядина и телятина Баранина



Сколько мяса покупают во всем мире, млн тонн



Сколько мяса покупают во всем мире, млн тонн



Какие места занимали российские города в рейтинге самых загруженных дорог в мире в 2020 году



Какие места занимали российские города в рейтинге самых загруженных дорог в мире в 2020 году



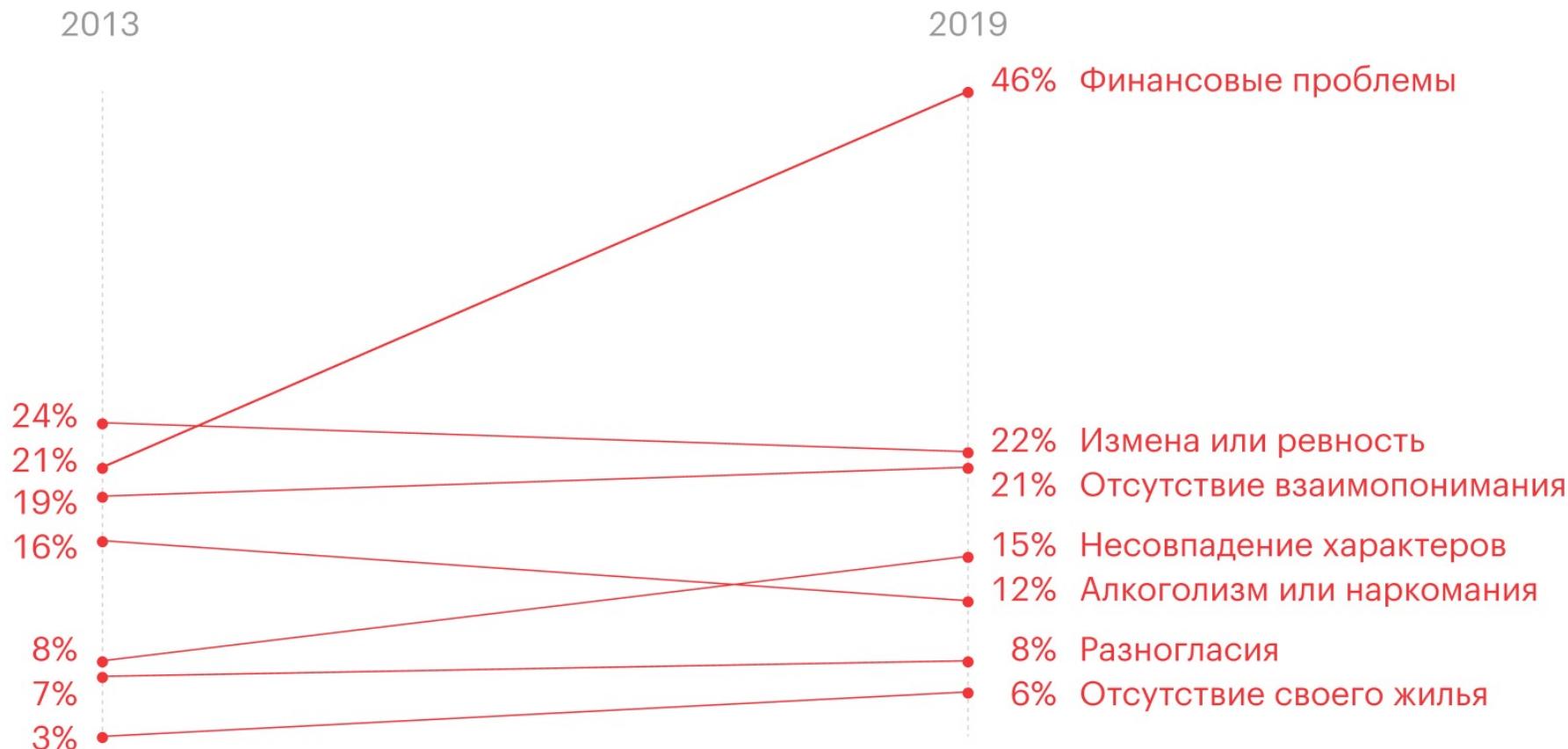
Принцип фокуса

0458944691234075284924642
9023978841034859830292877
5682907643189503782828389

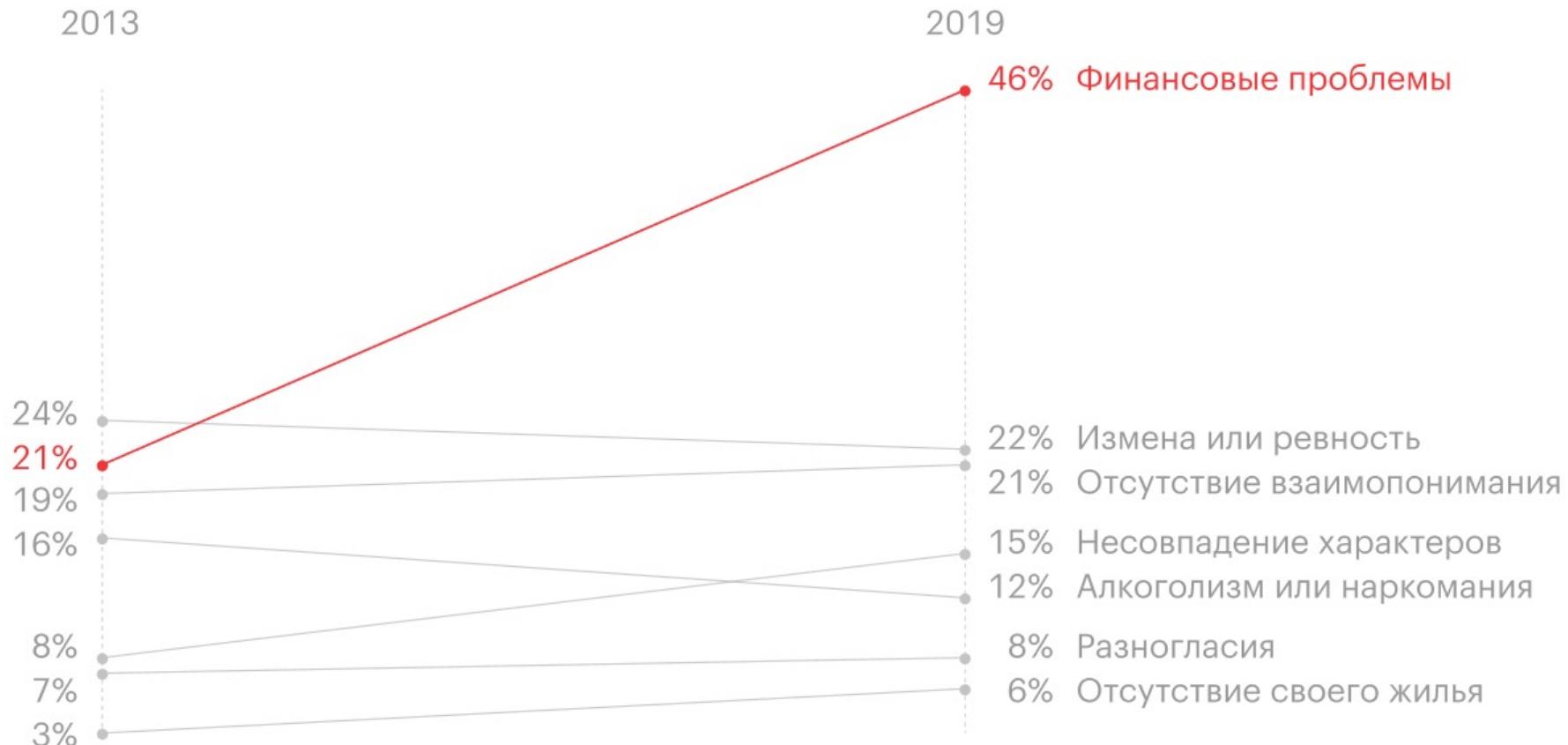
Принцип фокуса

0 4 5 8 9 4 4 6 9 1 2 3 4 0 7 5 2 8 4 9 2 4 6 4 2
9 0 2 3 9 7 8 8 4 1 0 3 4 8 5 9 8 3 0 2 9 2 8 7 7
5 6 8 2 9 0 7 6 4 3 1 8 9 5 0 3 7 8 2 8 2 8 3 8 9

Что вынуждает людей к разводам, по мнению россиян

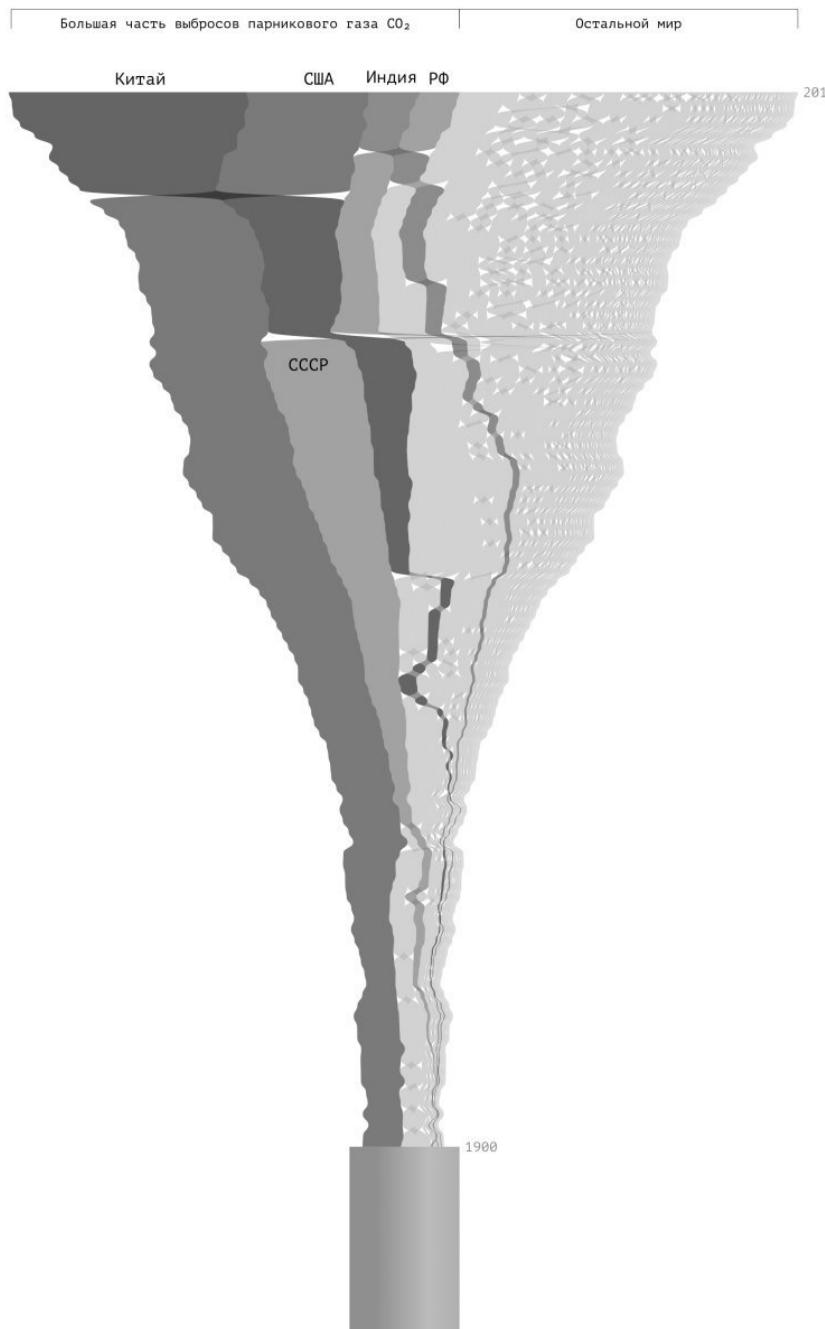


Что вынуждает людей к разводам, по мнению россиян



Кто виноват в изменении климата?

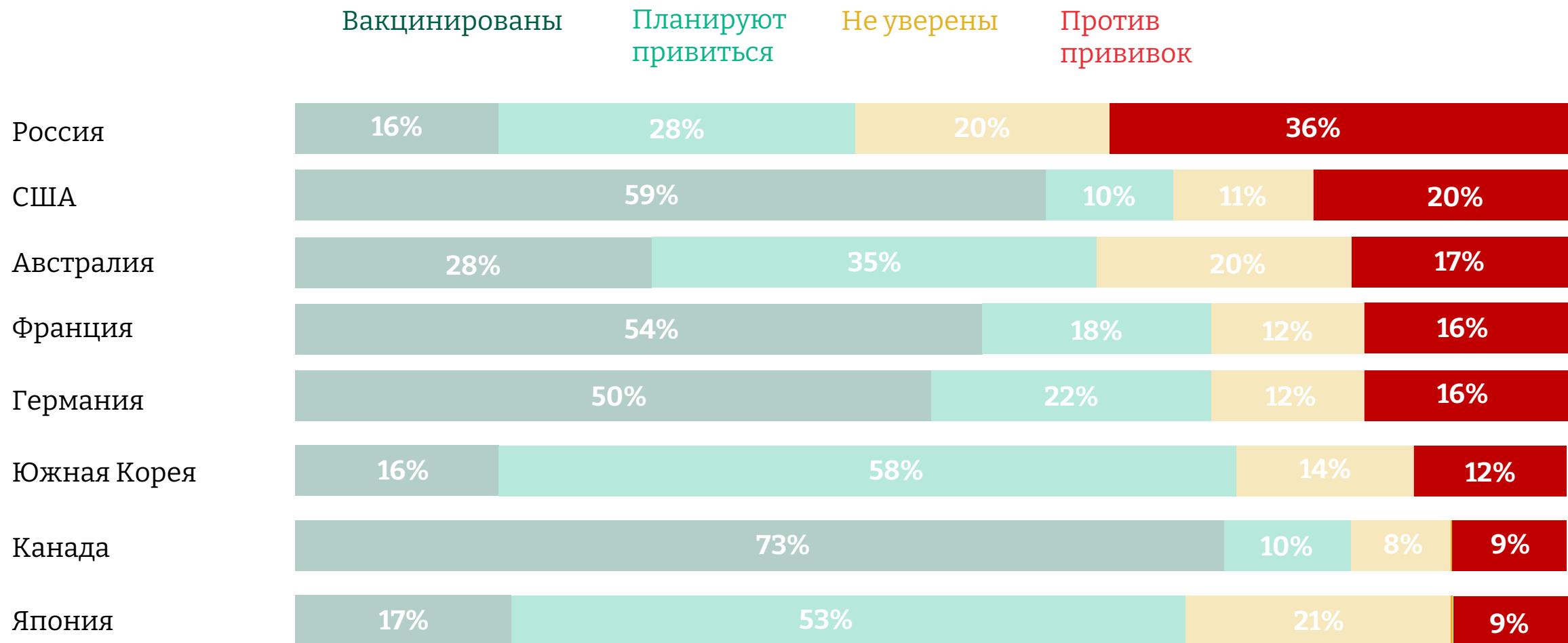
Кто украл снег на Новый год?



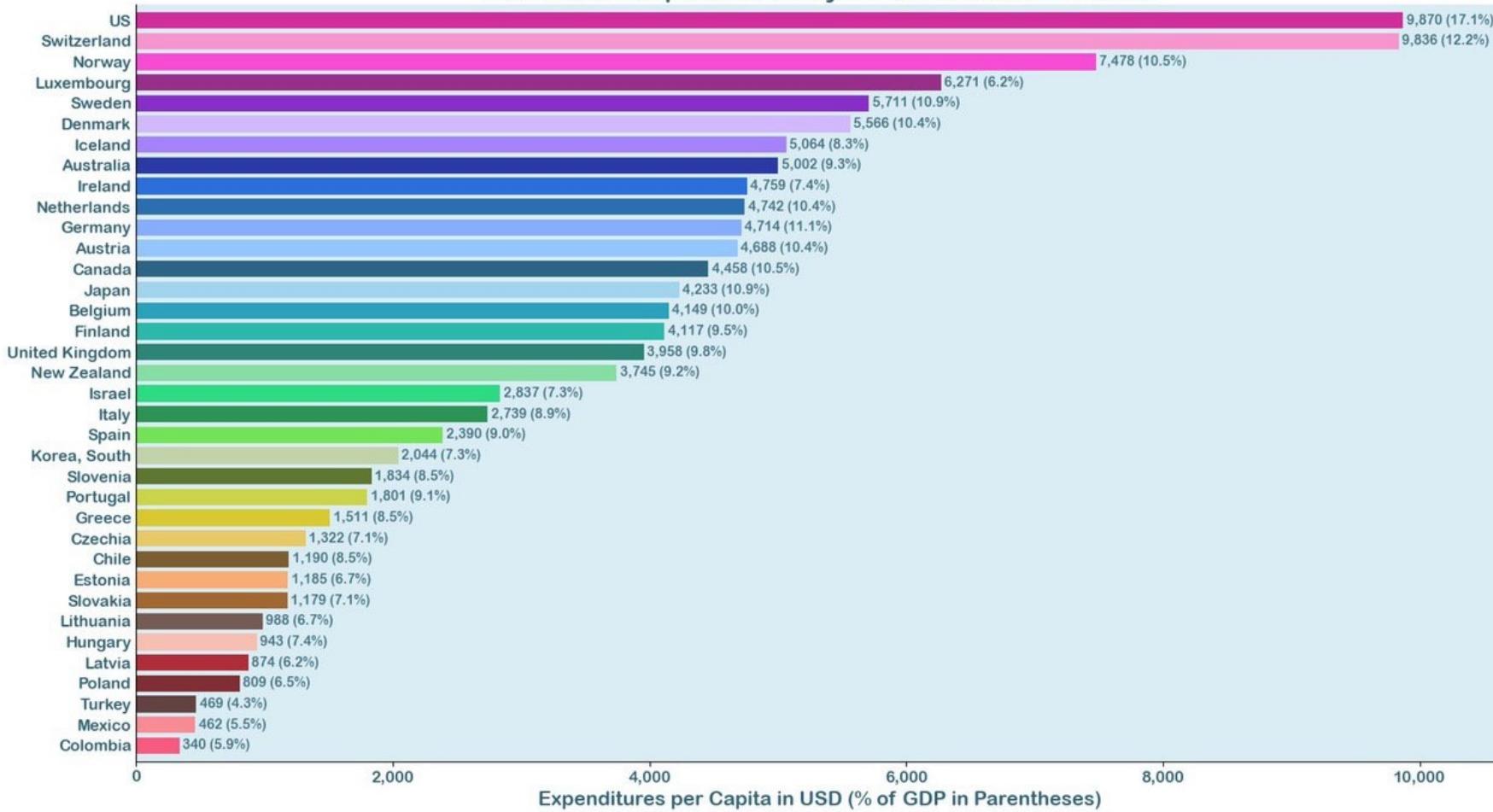
Россия – лидер по доле населения, отрицающего прививки



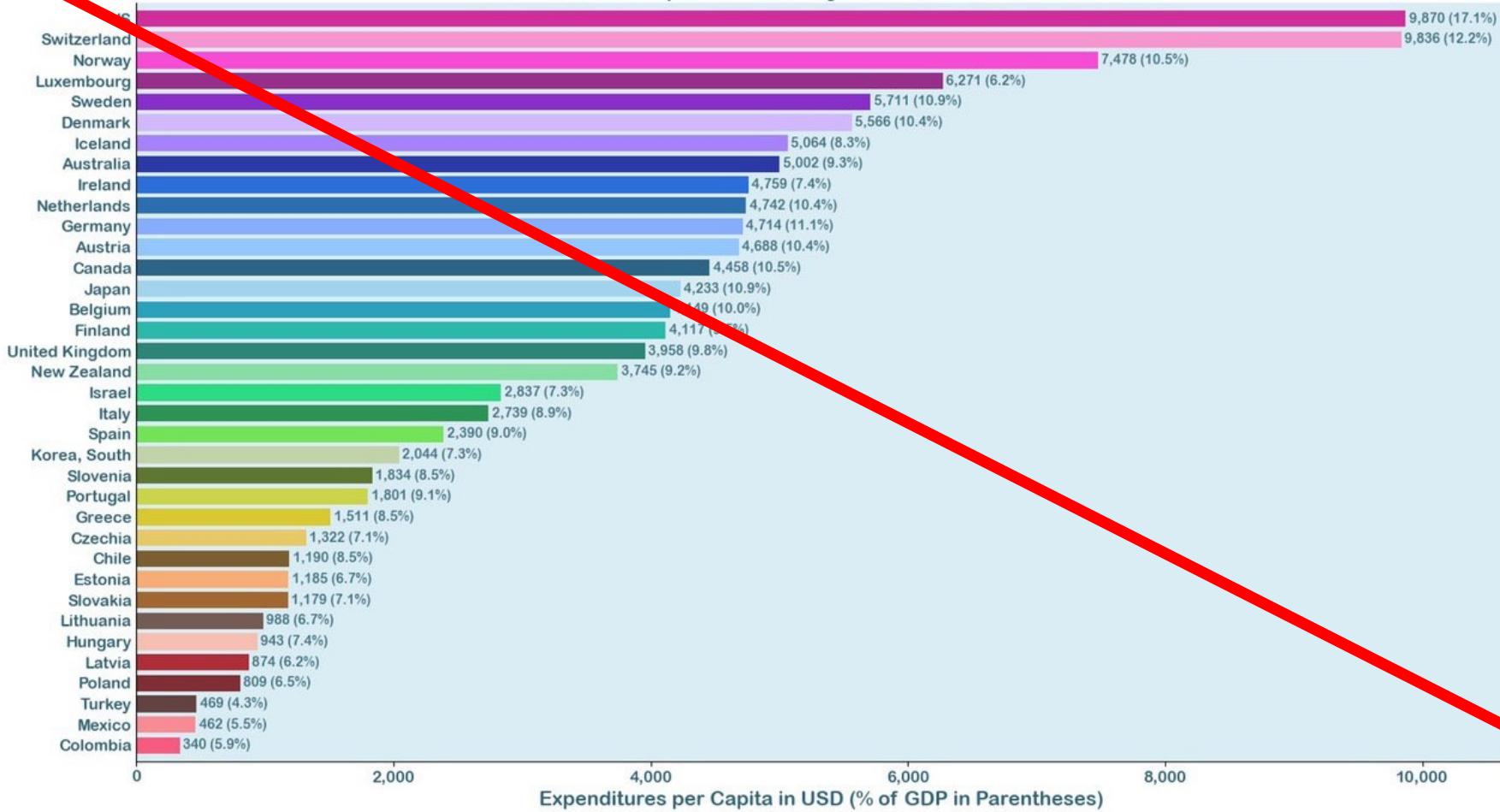
Россия – лидер по доле населения, отрицающего прививки



Healthcare Expenditures by OECD Countries in 2016



Healthcare Expenditures by OECD Countries in 2016



Принцип сходства

- Объекты одинакового цвета или формы воспринимаются нами как части одного целого





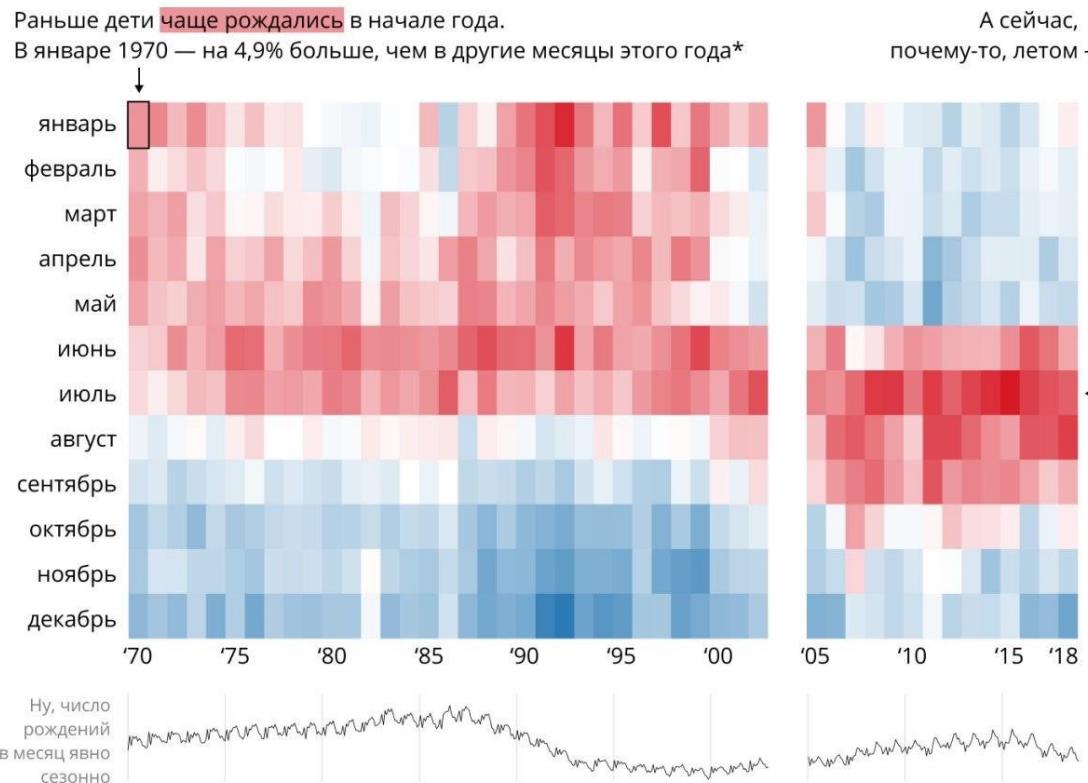
Топ-5 гомеопатических средств по количеству упоминаний



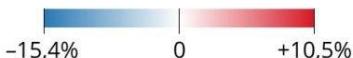
В какие месяцы больше дней рождения?

Раньше дети **чаще рождались** в начале года.

В январе 1970 — на 4,9% больше, чем в другие месяцы этого года*



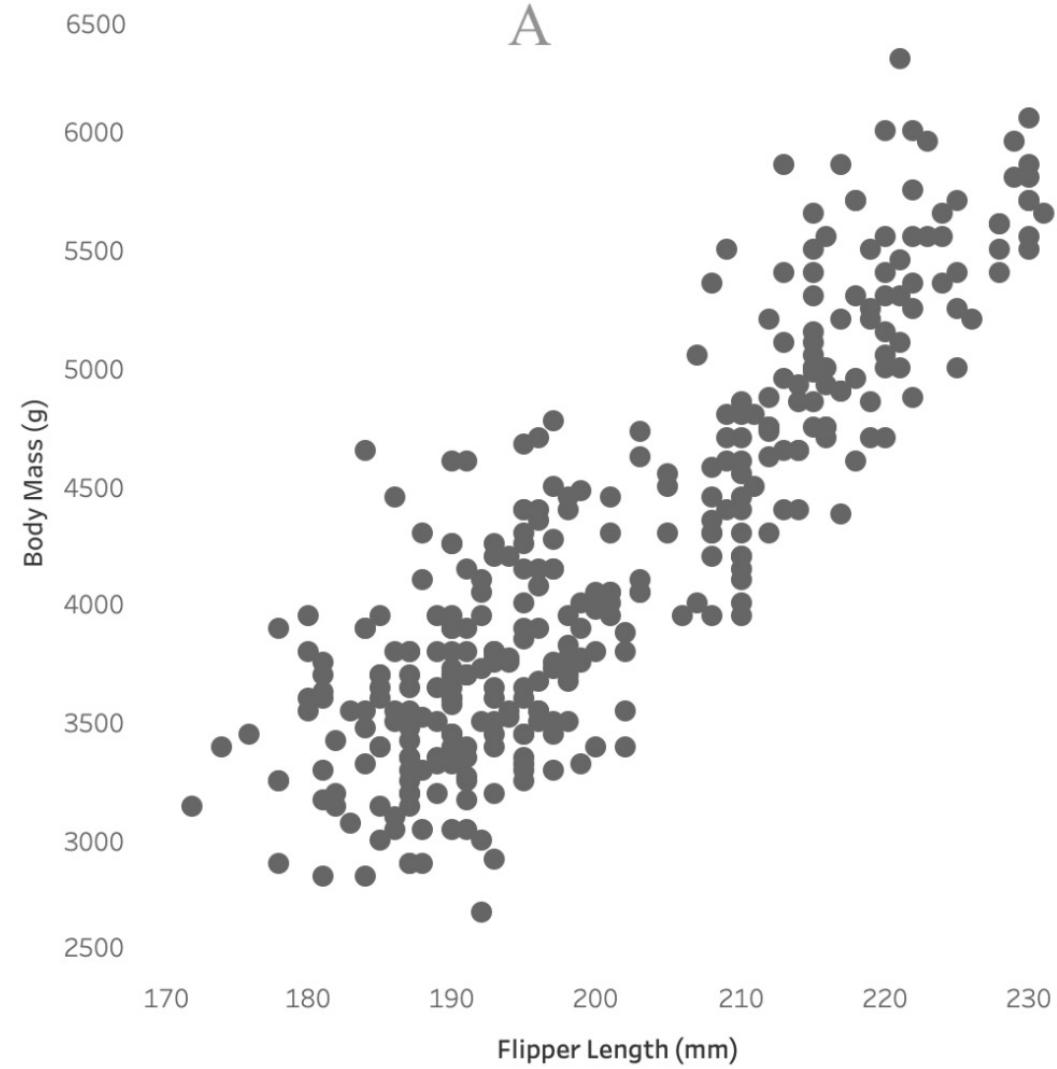
* Известно сколько в России родилось людей в каждом месяце каждого (кроме 2003–2004) года. Считаем среднее за день по каждому месяцу. Потом отклонение этого числа от среднего за день по соответствующему году (с учетом високосных лет и числа дней в месяцах). Данные — «Демографические ежегодники» Росстата и приложения «Демоскопа».



«Графики и жизнь», 9.20

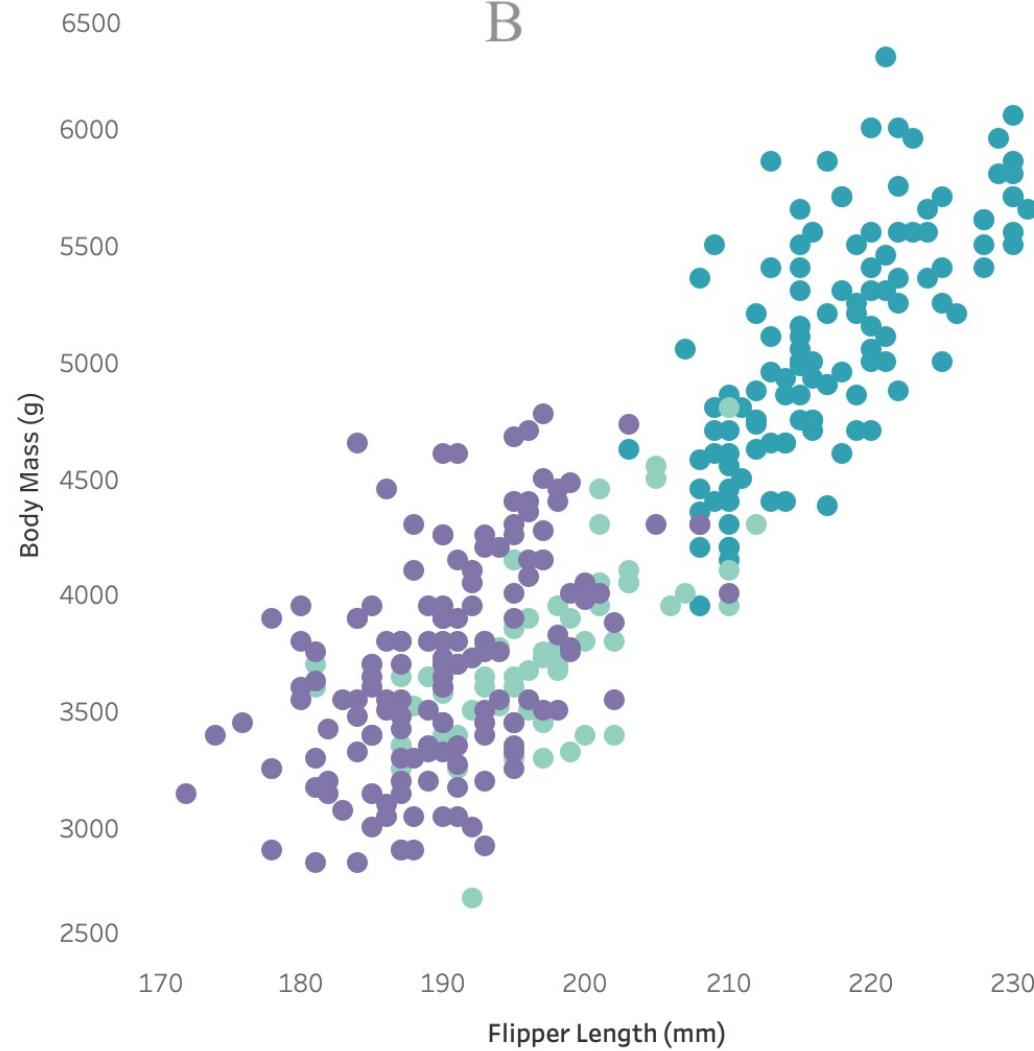
Body Mass & Flipper Length

A



Body Mass & Flipper Length

B



«Умные» заголовки



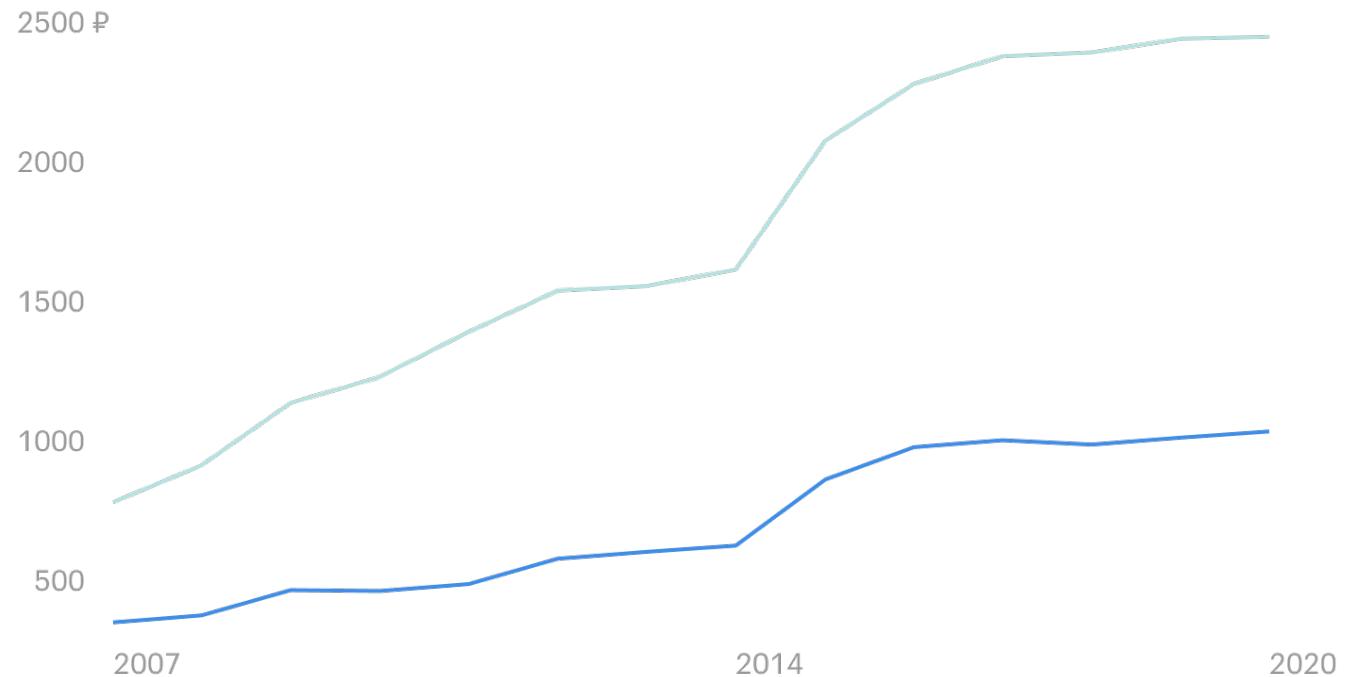
Не очень*:

Цены на растворимый и
зерновой кофе



Да:

Цены на растворимый кофе
растут быстрее, чем на
зерновой



«Умные» заголовки



Не очень*:

15 самых распространенных
уличных топонимов на 2020
год



Да:

Улица Ленина на 14 месте
среди самых «популярных»
улиц
15 самых распространенных
уличных топонимов на 2020 год

Место	Улица	Количество
1	Центральная	25 860
2	Молодежная	17 727
3	Лесная	16 470
4	Школьная	15 913
5	Садовая	14 441
6	Новая	13 792
7	Советская	12 766
8	Набережная	11 065
9	Заречная	10 695
10	Полевая	10 542
11	Луговая	9 972
12	Зеленая	9 749
13	Мира	9 127
14	Ленина	7 652
15	Октябрьская	6 919

Человечные подписи



В каких городах самый высокий
показатель загруженности дорог



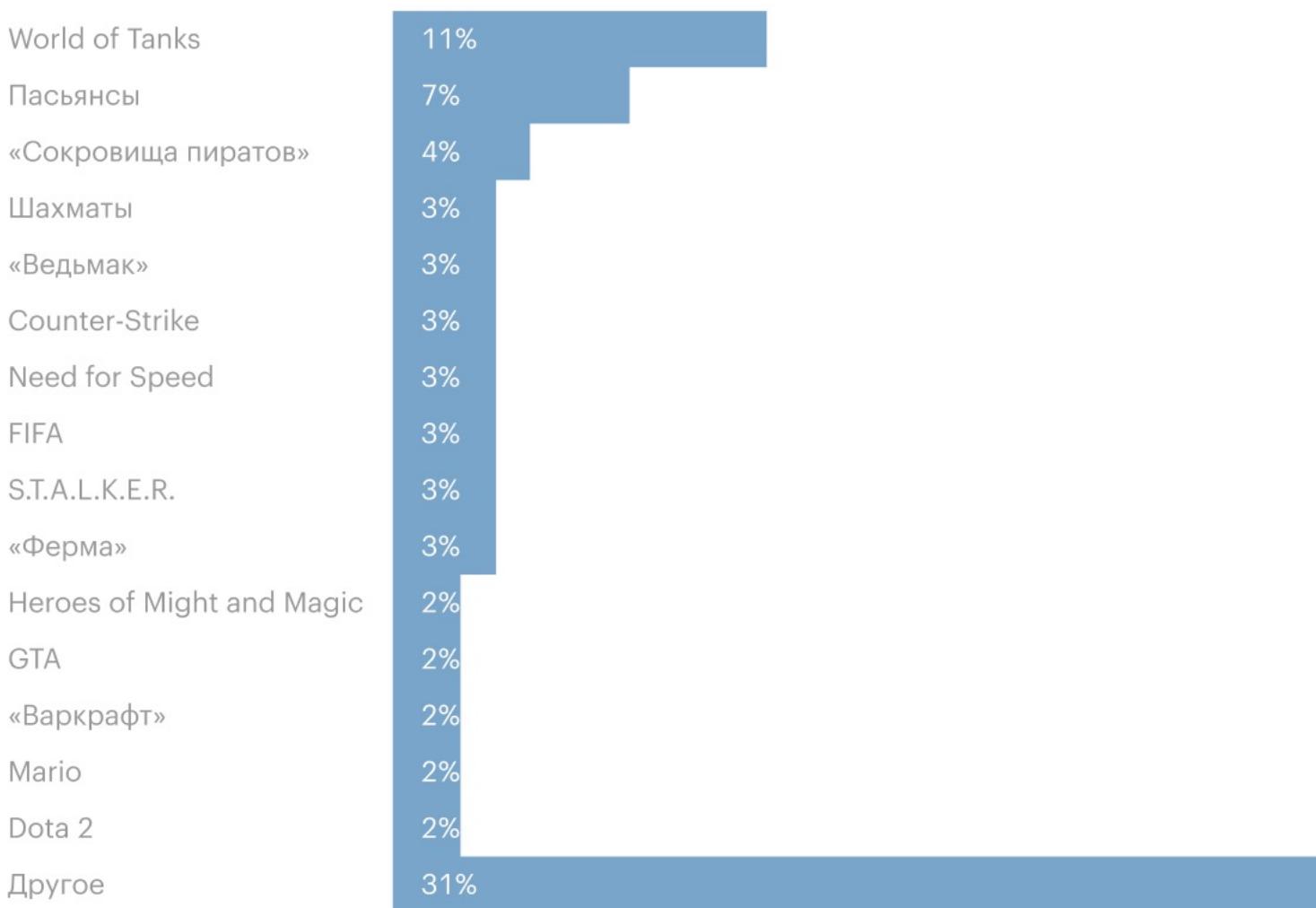
Человечные подписи

В каких городах самый высокий
показатель загруженности дорог



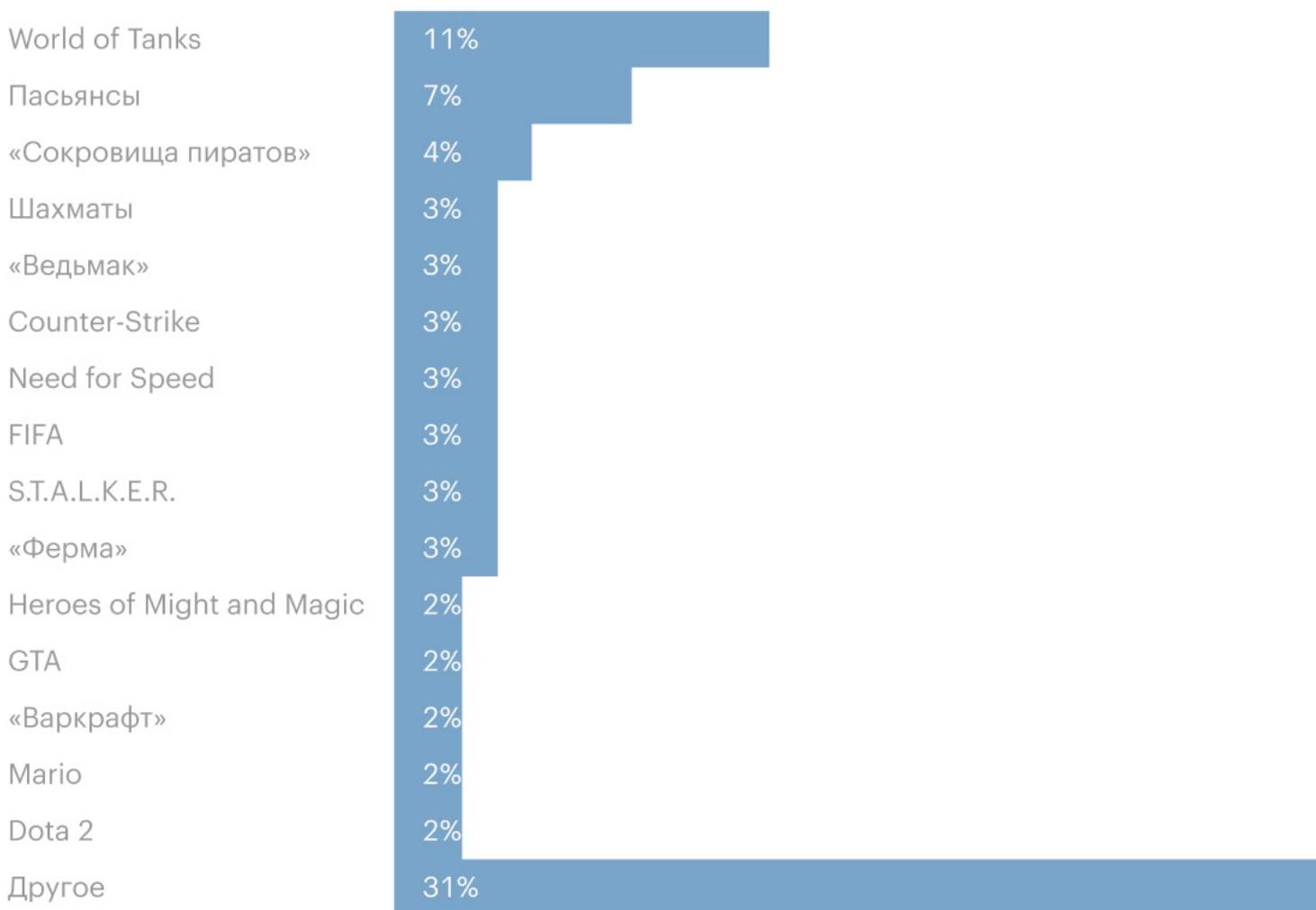
Всё должно быть подписано!

Самые популярные игры у российских геймеров



Самые популярные игры у российских геймеров

Процент от числа опрошенных

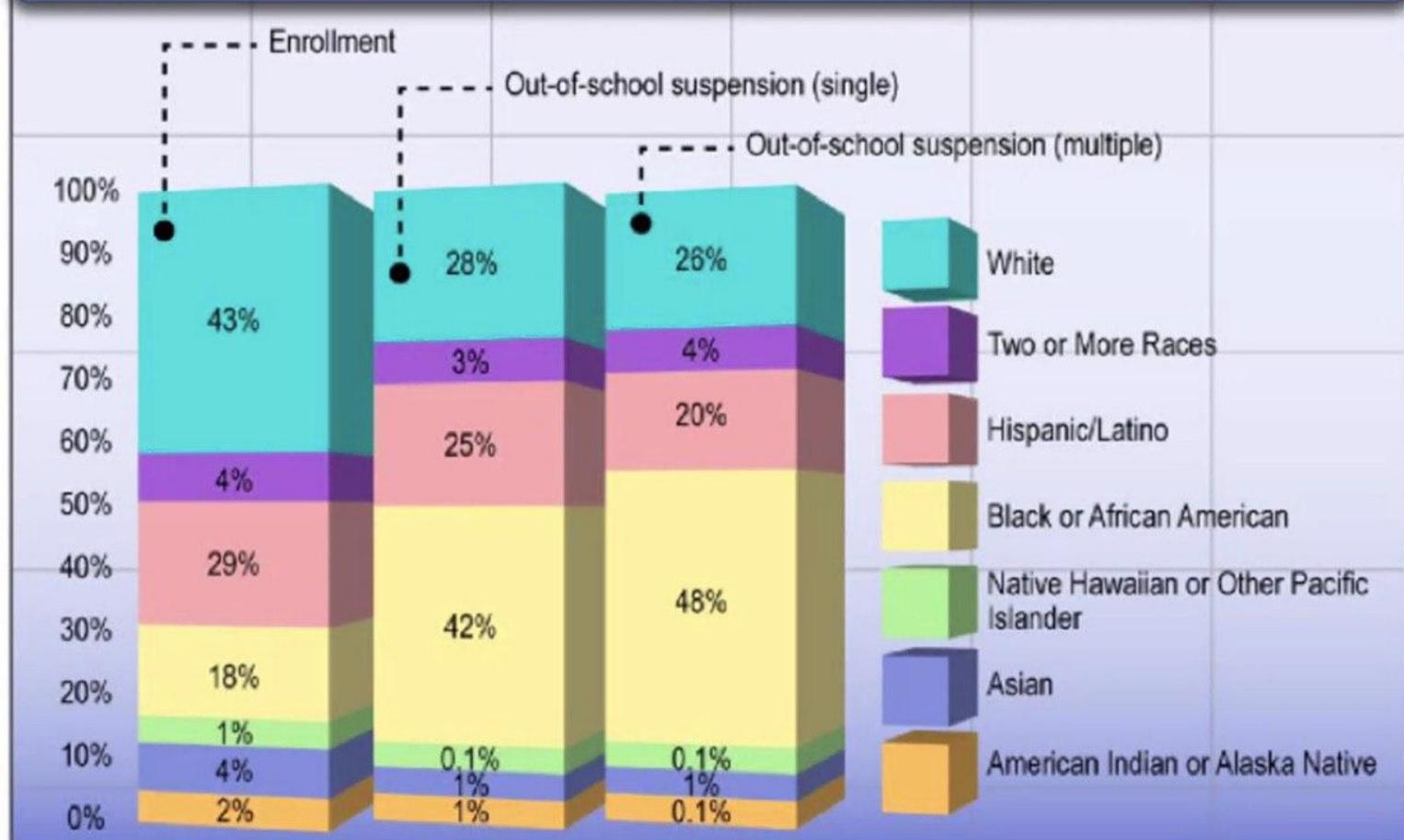


Чек-лист понятного графика

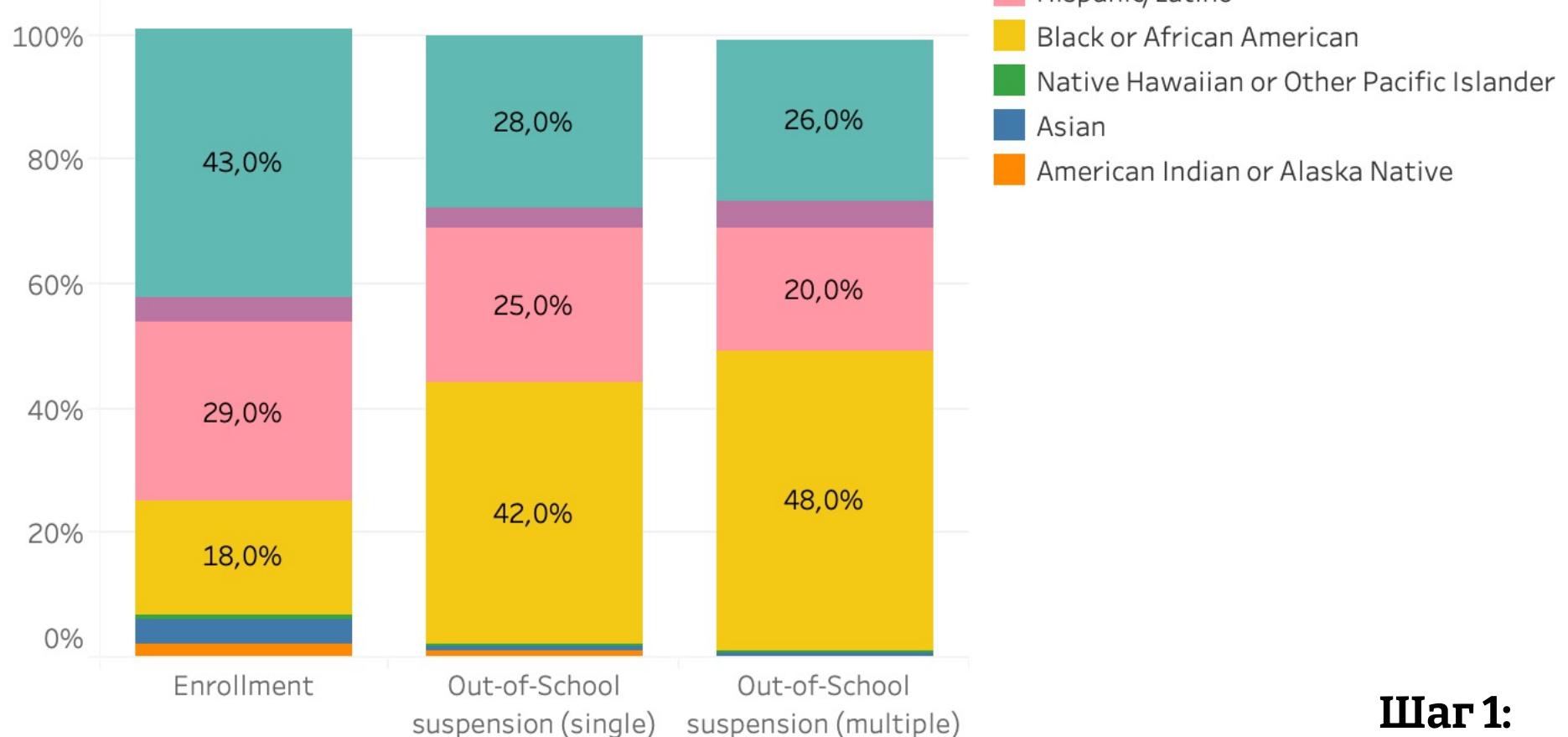
- ✓ Соблюдён принцип простоты
- ✓ Соблюдён принцип близости
- ✓ Соблюдён принцип фокуса
- ✓ Соблюдён принцип сходства
- ✓ Заголовок помогает основной идее
- ✓ Подписи удобно читать
- ✓ Всех подписей хватает, не осталось неочевидных моментов

Применяем на практике

Preschool students receiving suspensions, by race and ethnicity

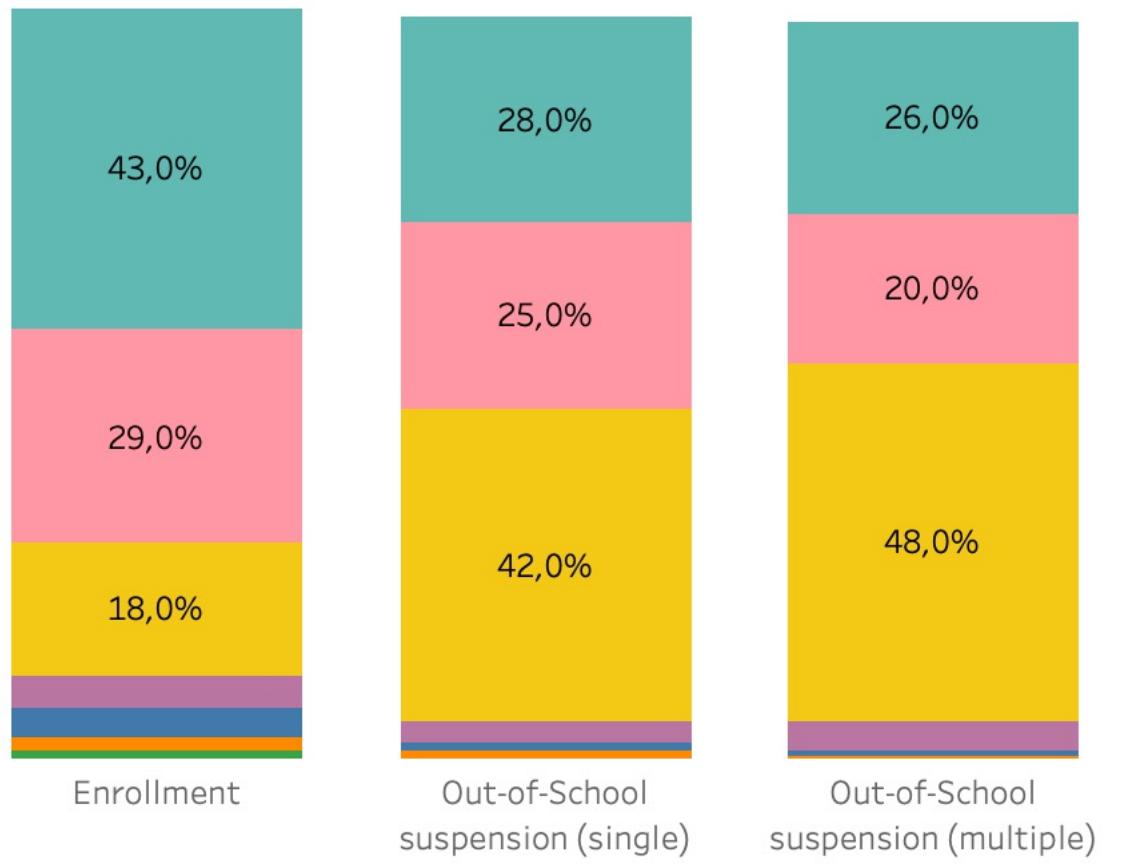


Preschool students receiving suspensions, by race and ethnicity



Шаг 1:
Убираем трёхмерность и
градиентный фон

Preschool students receiving suspensions, by race and ethnicity

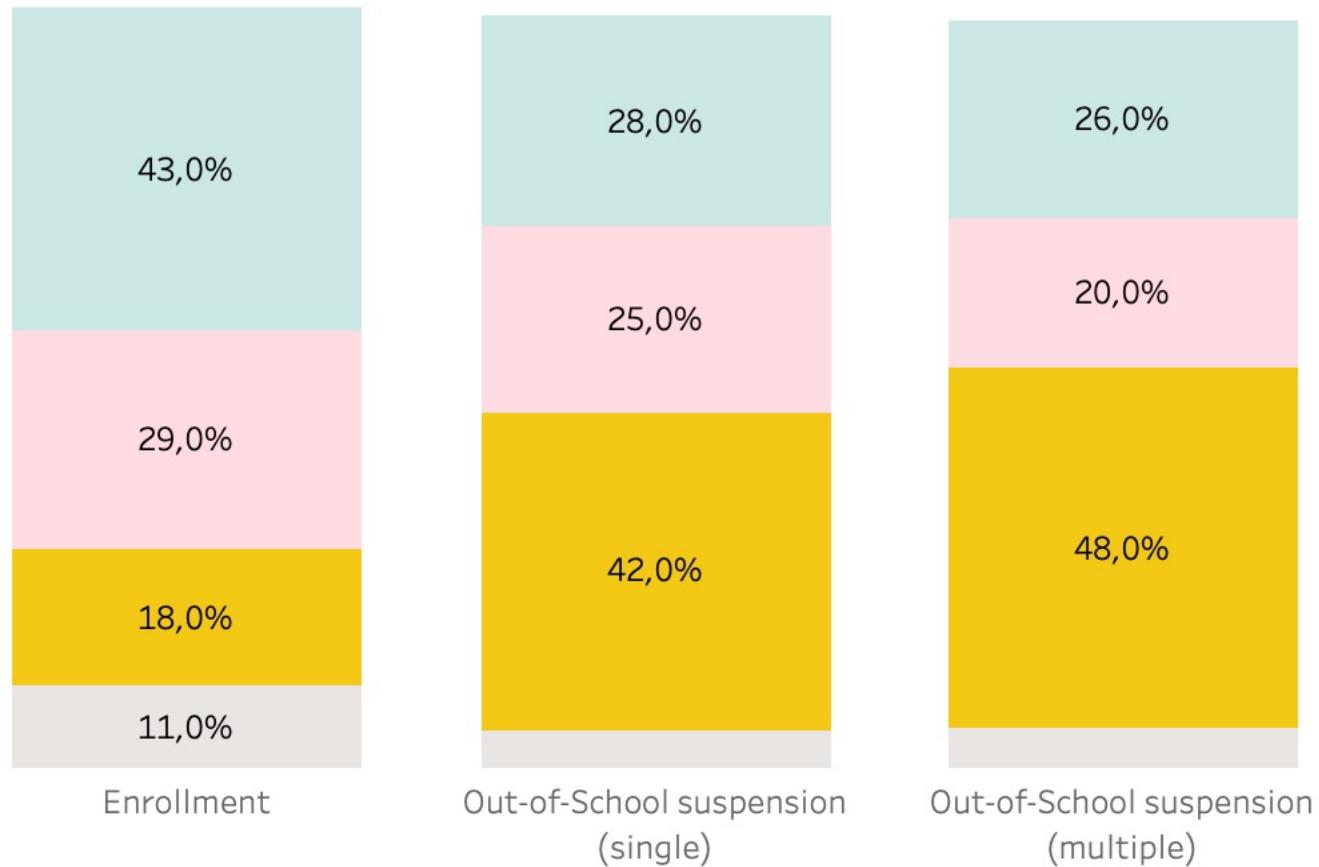


Race

- White
- Hispanic/Latino
- Black or African American
- Two or more races
- Asian
- American Indian or Alaska Native
- Native Hawaiian or Other Pacific Islander

Шаг 2:
Делаем осмысленную
сортировку, убираем
ось и сетку

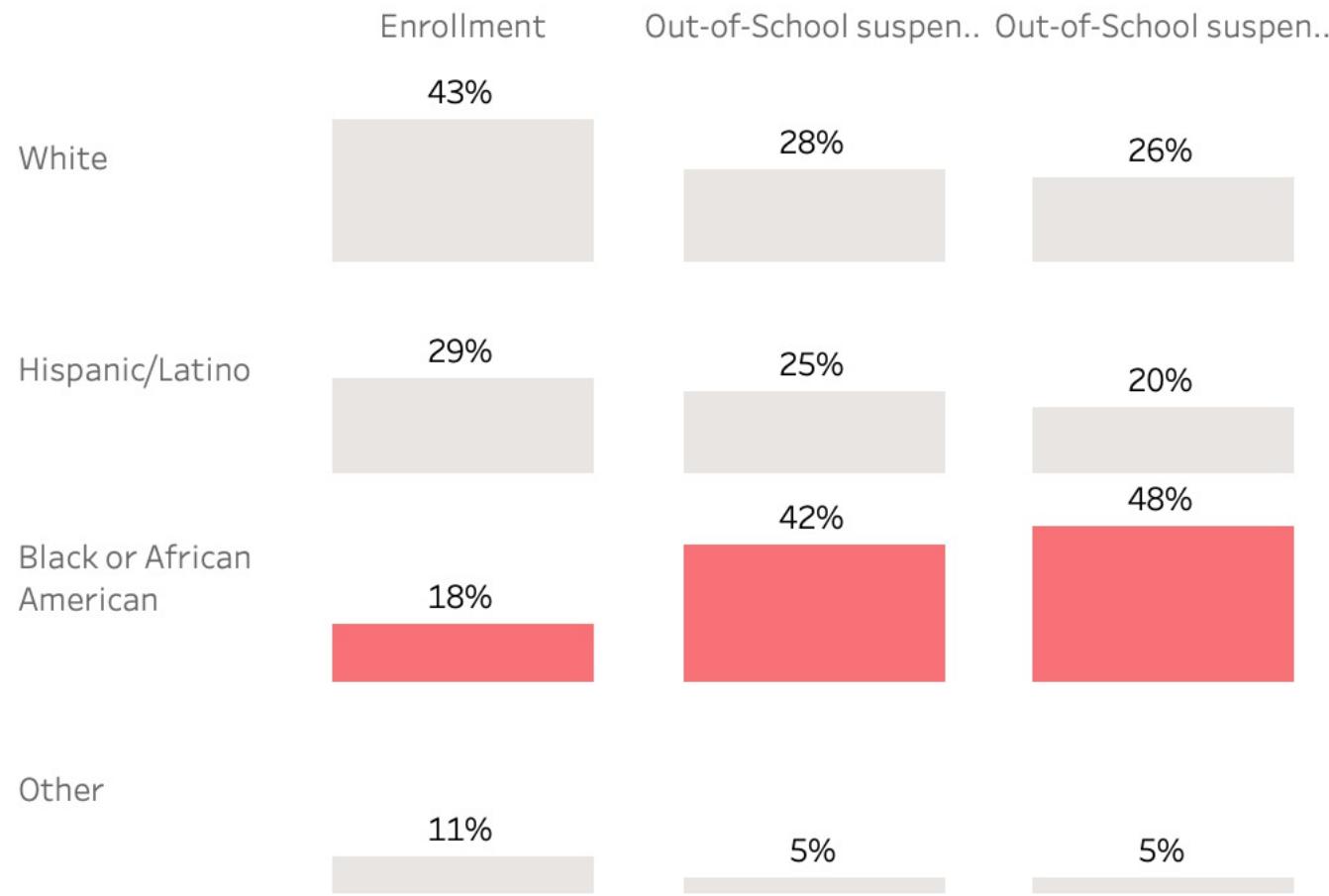
Black preschool students are more likely suspended than other children



Race
White
Hispanic/Latino
Black or African American
Other

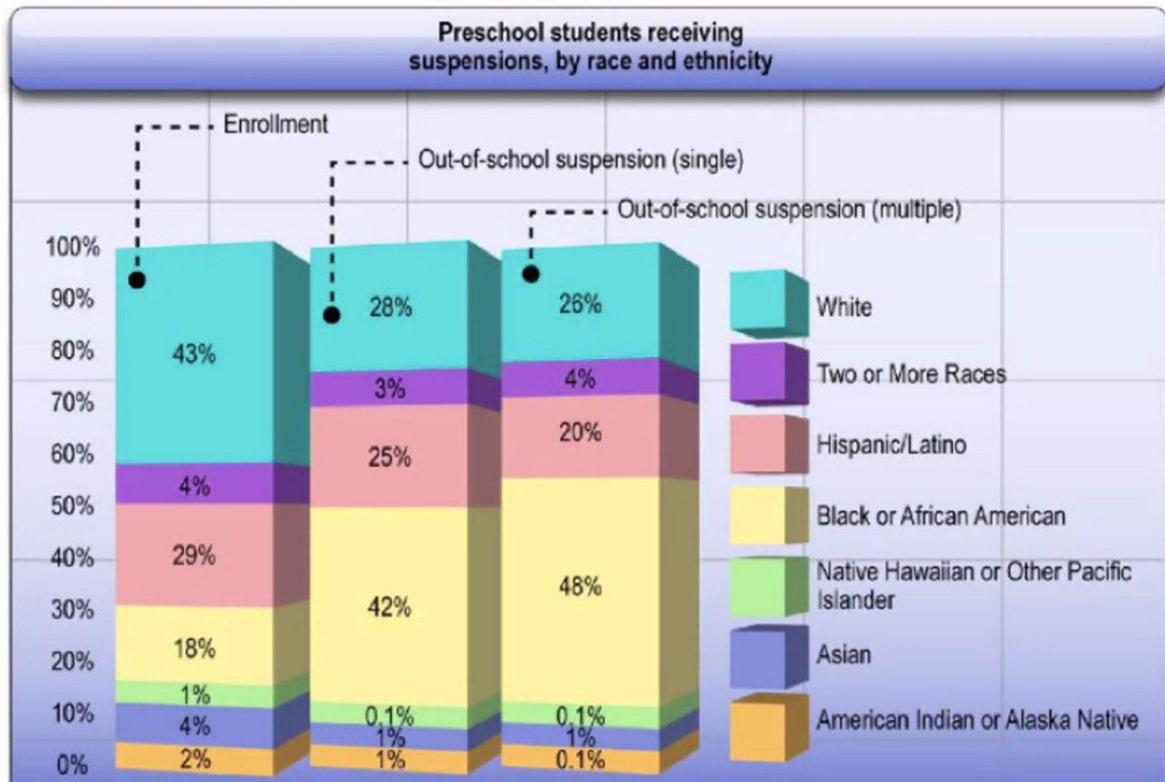
Шаг 3:
Группируем
малозначимые группы,
делаем осмысленный
заголовок, выделяем
ярким ключевую идею

Black preschool students are more likely suspended than other children



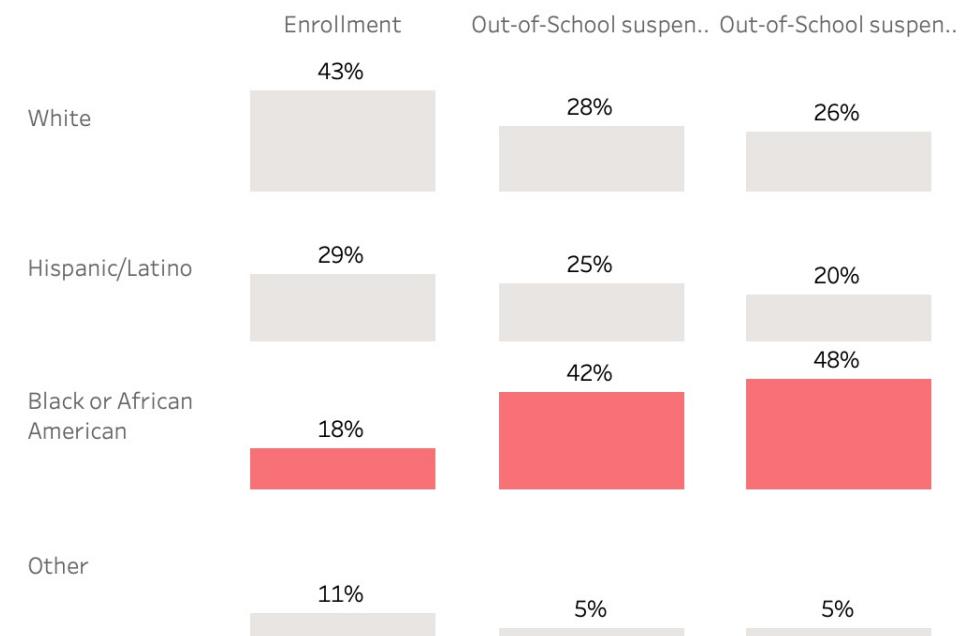
Шаг 4:
Выравниваем каждую
категорию по
горизонтали,
уменьшаем количество
цветов

До



После

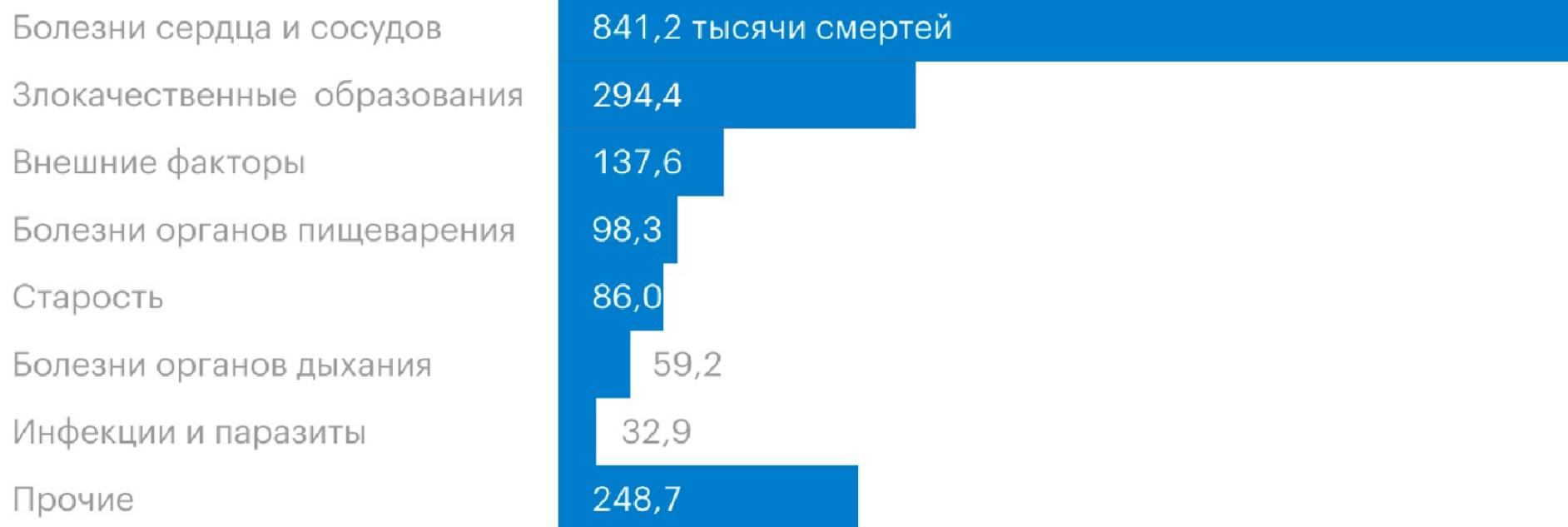
Black preschool students are more likely suspended than other children



**Как сделать графики
более интересными?**

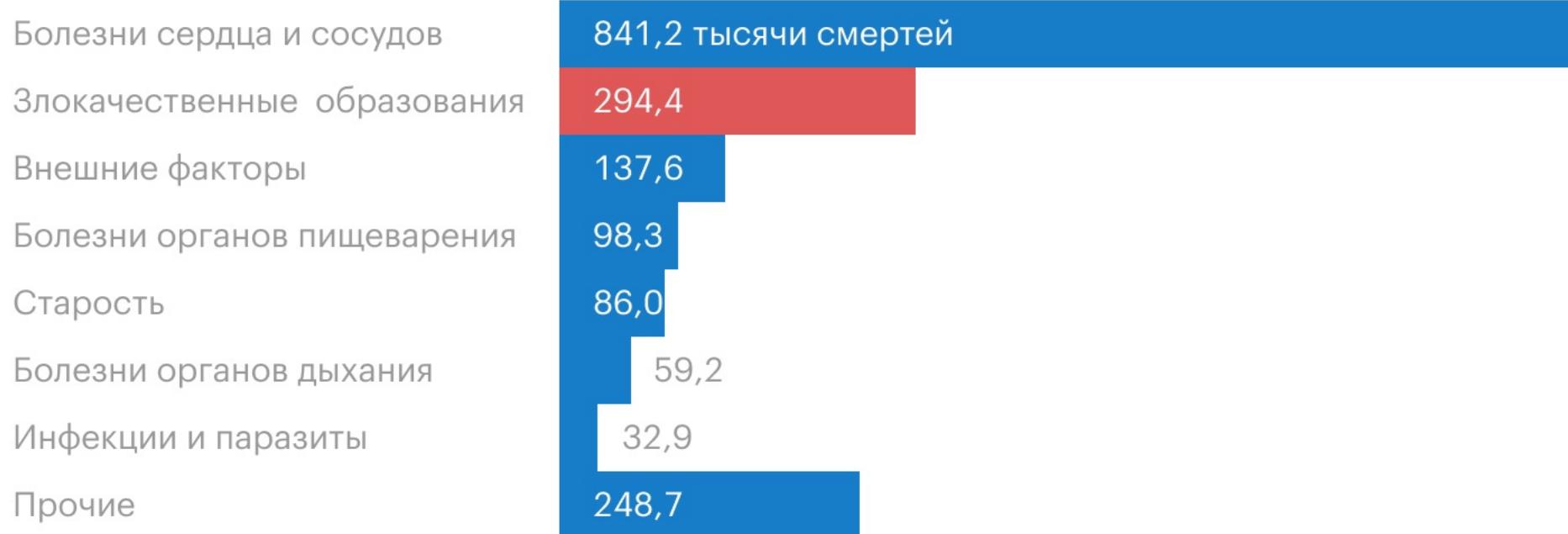
1. Искать истории

От чего умирали Россияне в 2019 году



Источник: [НИИ им. Герцена](#)

В 2019 году рак стал причиной каждой шестой смерти в России



Источник: [НИИ им. Герцена](#)

**2. Рассказывать, что
происходит на графике**

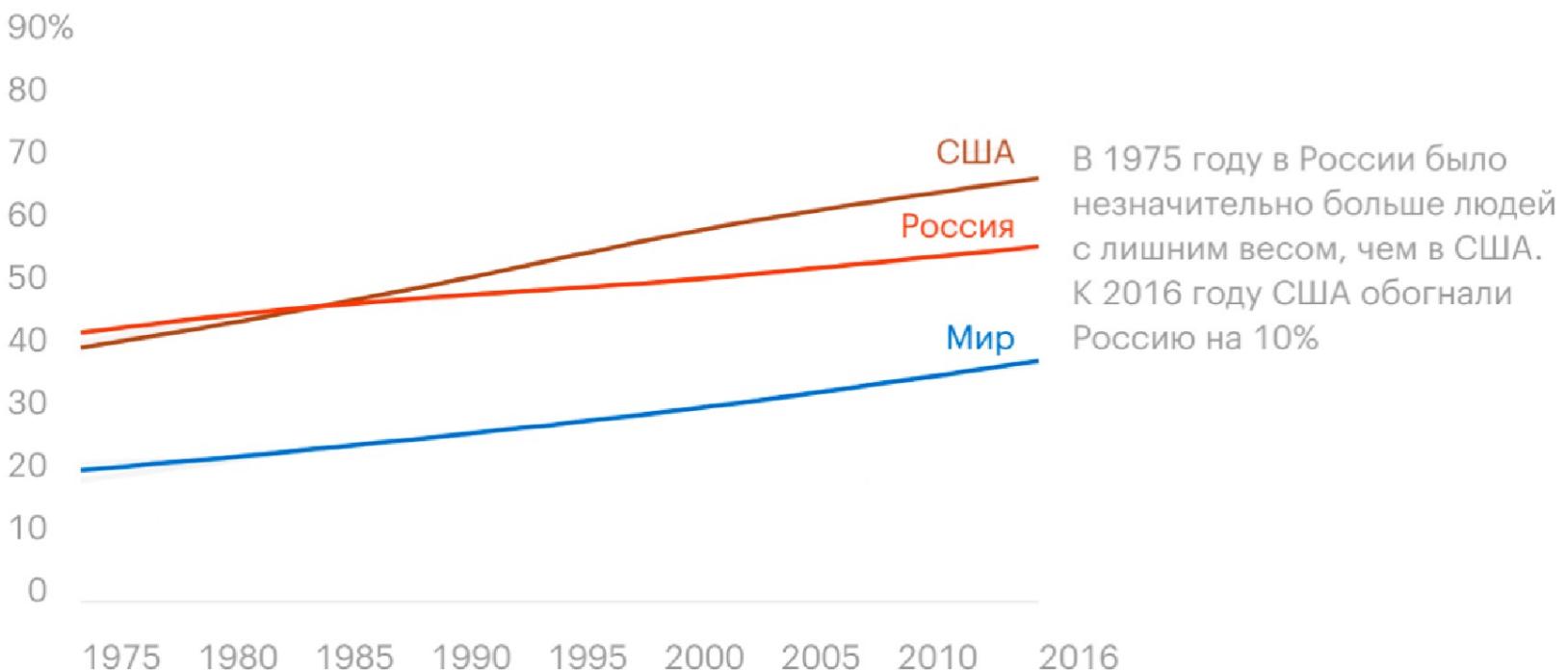
Стоимость нефти марки Brent



© ТАСС, 2020. Источник: [investing.com](https://www.investing.com), 29.10.2020 15:20 мск.

3. Не скучиться на данные

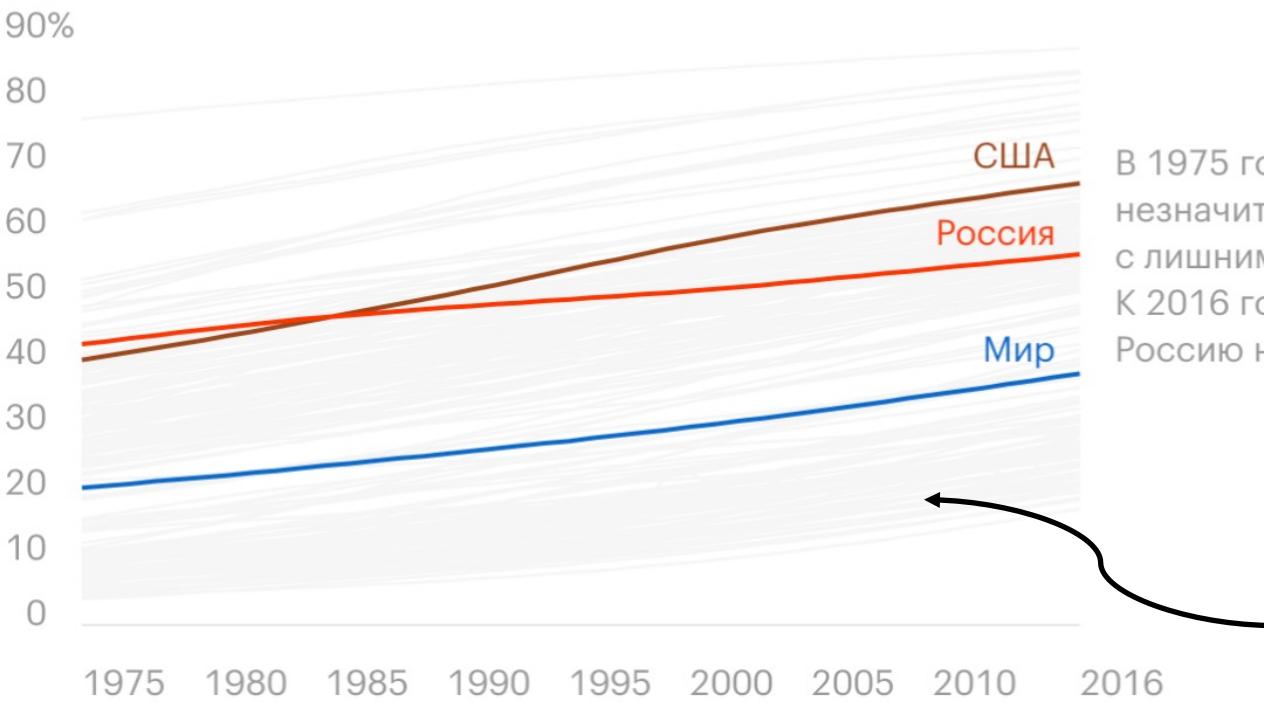
Как менялась доля людей с лишним весом и ожирением



В 1975 году в России было незначительно больше людей с лишним весом, чем в США. К 2016 году США обогнали Россию на 10%

Источник: ВОЗ

Как менялась доля людей с лишним весом и ожирением

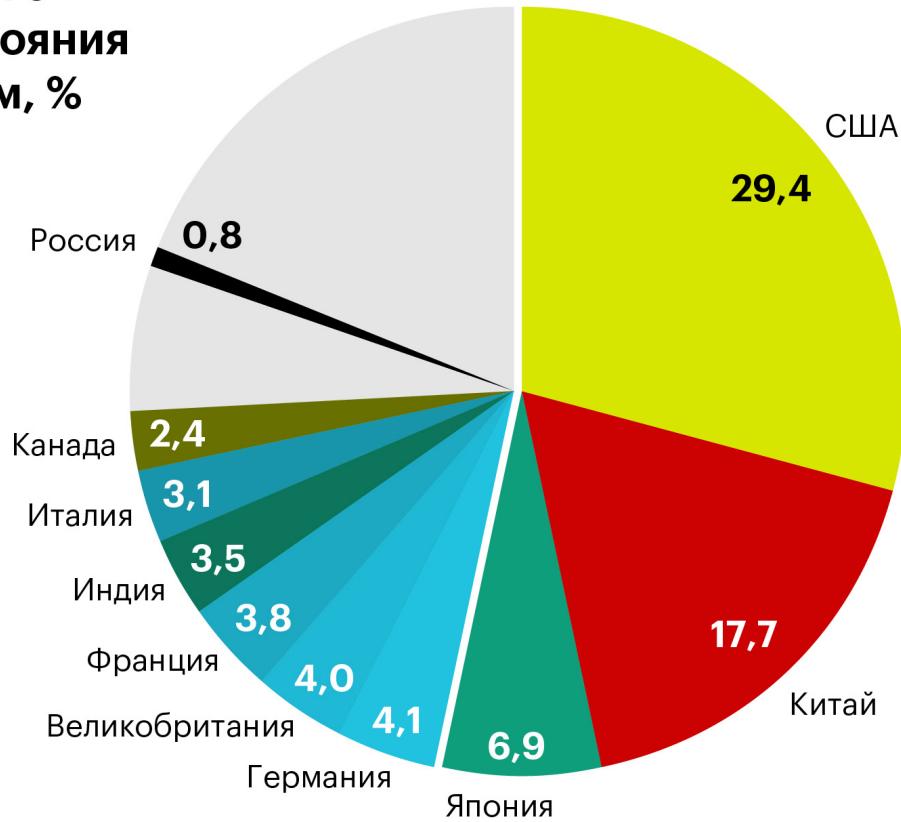


В 1975 году в России было незначительно больше людей с лишним весом, чем в США. К 2016 году США обогнали Россию на 10%

Серые линии дают понять, где находится Россия относительно других стран

Источник: ВОЗ

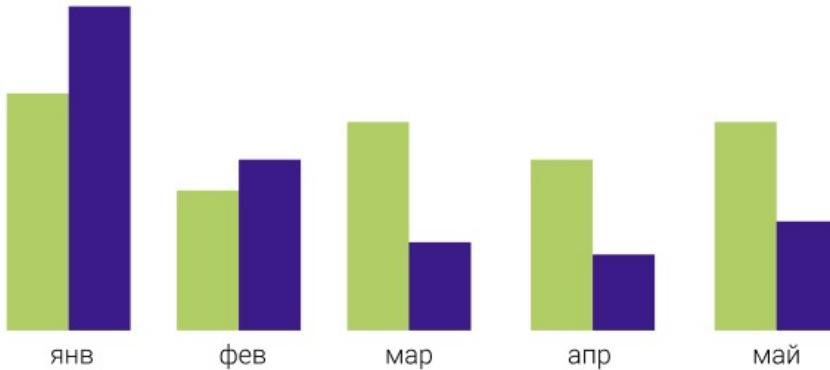
**Распределение
глобального
благосостояния
по странам, %**



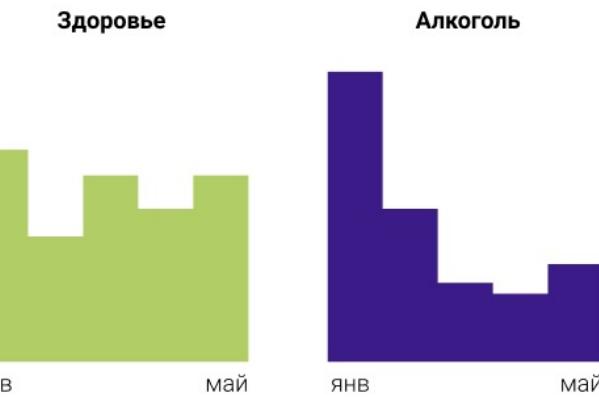
Алгоритм работы над графиком

«Что я хочу сказать?»

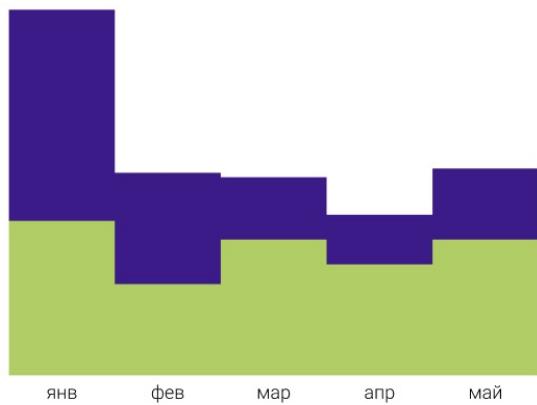
Траты на здоровье и алкоголь в 2021 году



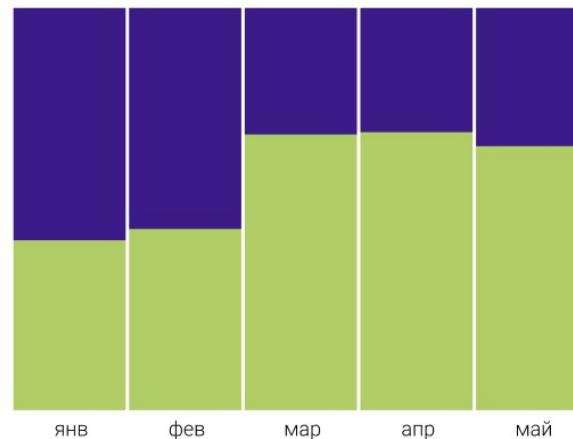
Мои траты в 2021 году



Траты на здоровье и алкоголь в 2021 году



Доля трат на здоровье и алкоголь в 2021 году



ВВП в странах мира в 2019 году

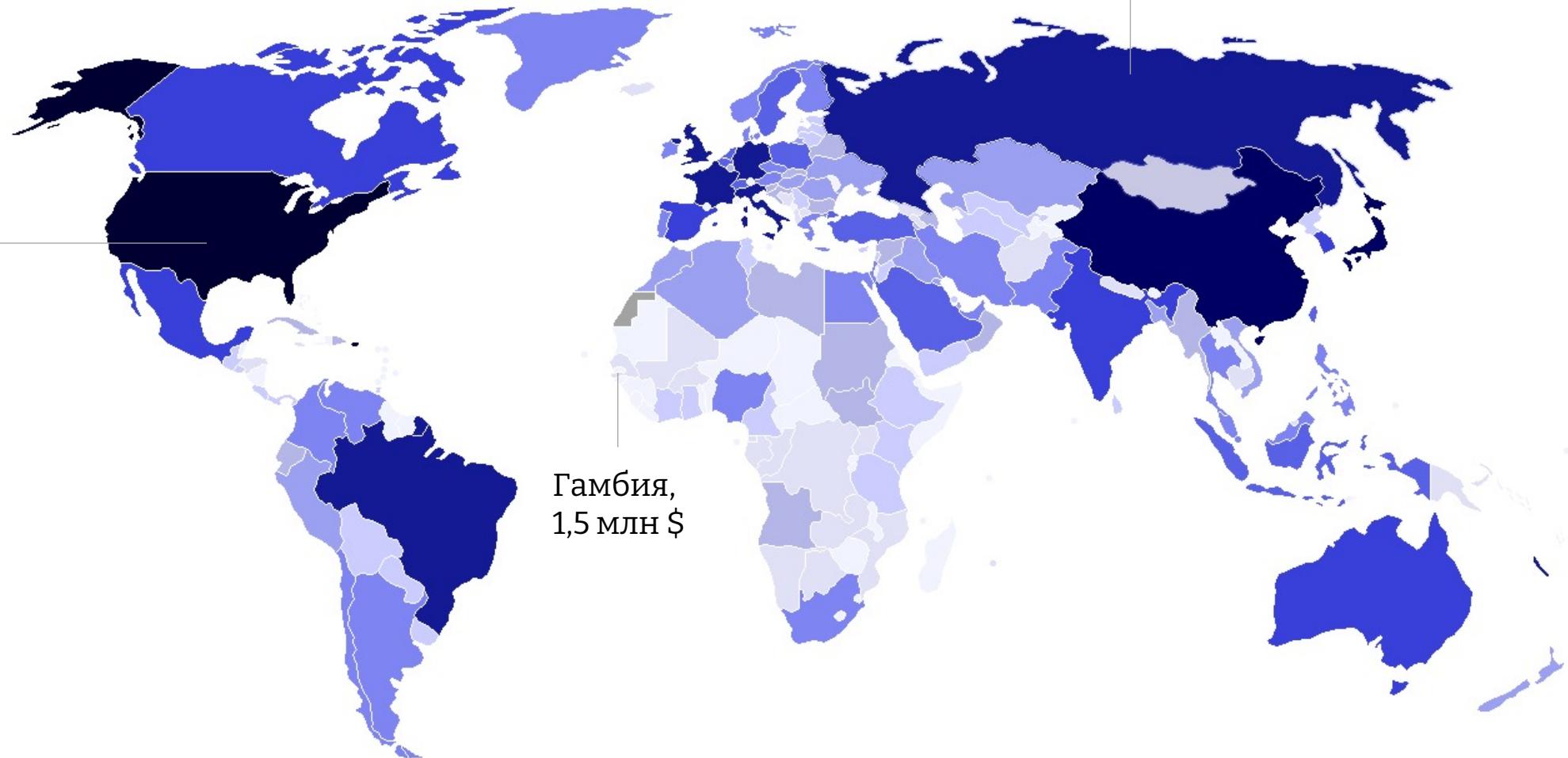


10 трлн \$

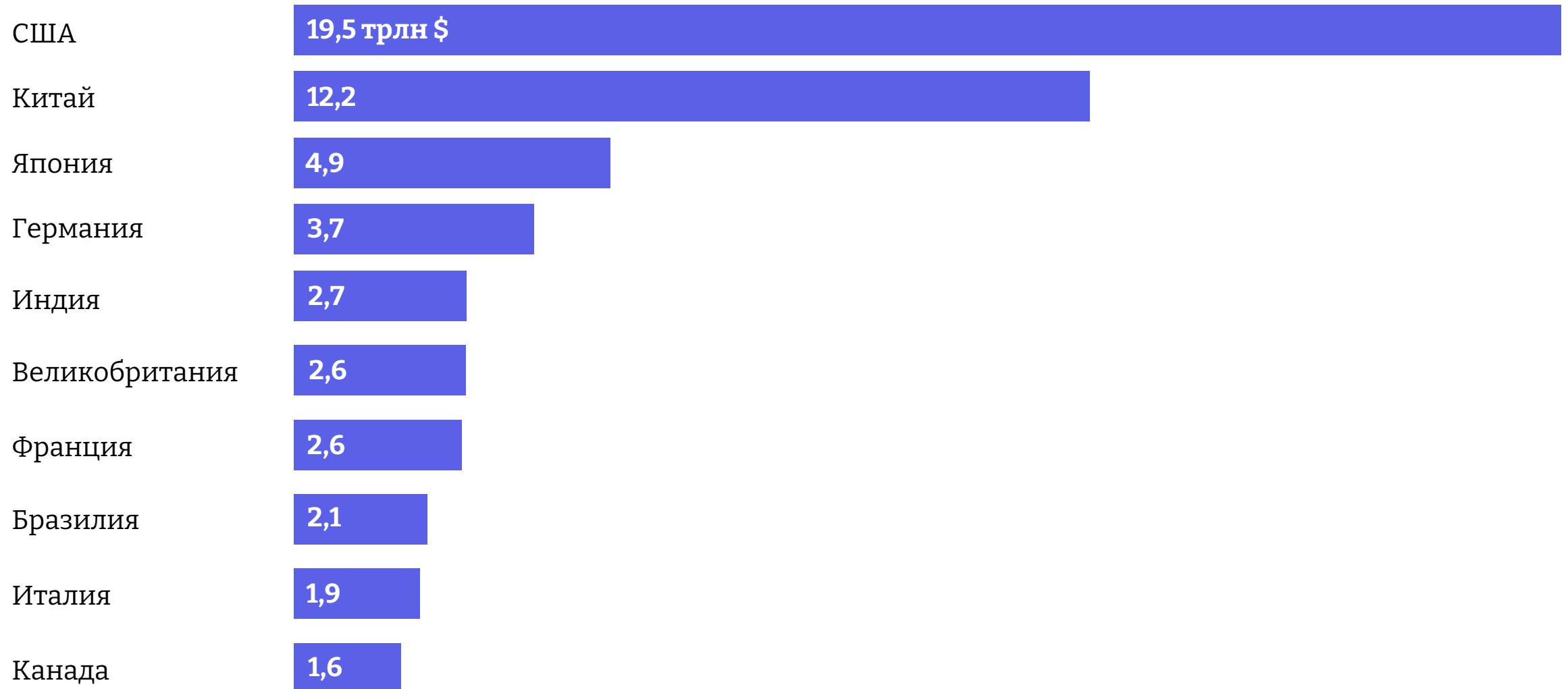
США,
19,5 трлн \$

Гамбия,
1,5 млн \$

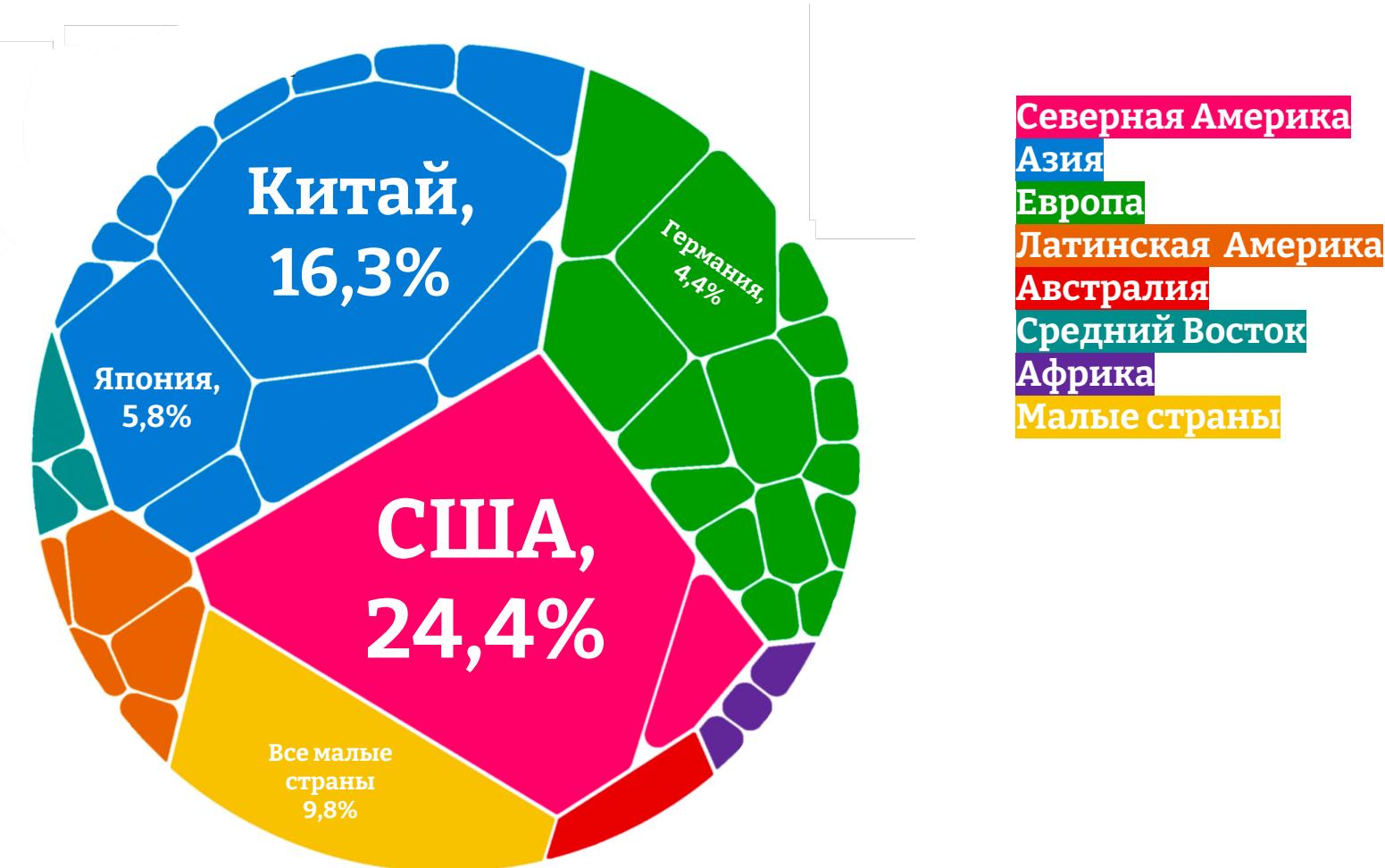
Россия,
1,6 трлн \$



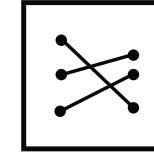
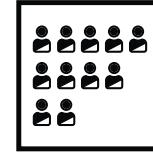
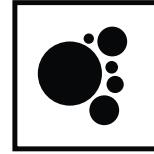
Топ-10 стран по ВВП в 2019 году



Как распределяется ВВП по странам мира в 2019 году



Выбрать из многообразия графиков



- [FT Visual Vocabulary](#)
- [Постер «Графики, которые убеждают всех»](#)
- [Data Viz Project](#)

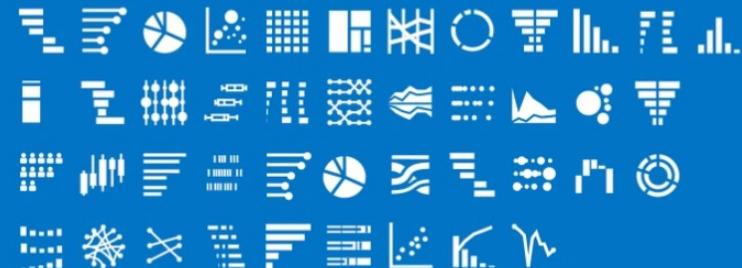
Где строить сложные визуализации

- [Datawrapper](#)
- [Flourish](#)
- [RawGraphs](#)

**Главное —
насмотренность**

Графики, которые убеждают всех

Александр Богачев





data.csv

https://t.me/data_csv

