



Прикладные задачи анализа данных

ИСКУССТВО ВИЗУАЛИЗАЦИИ

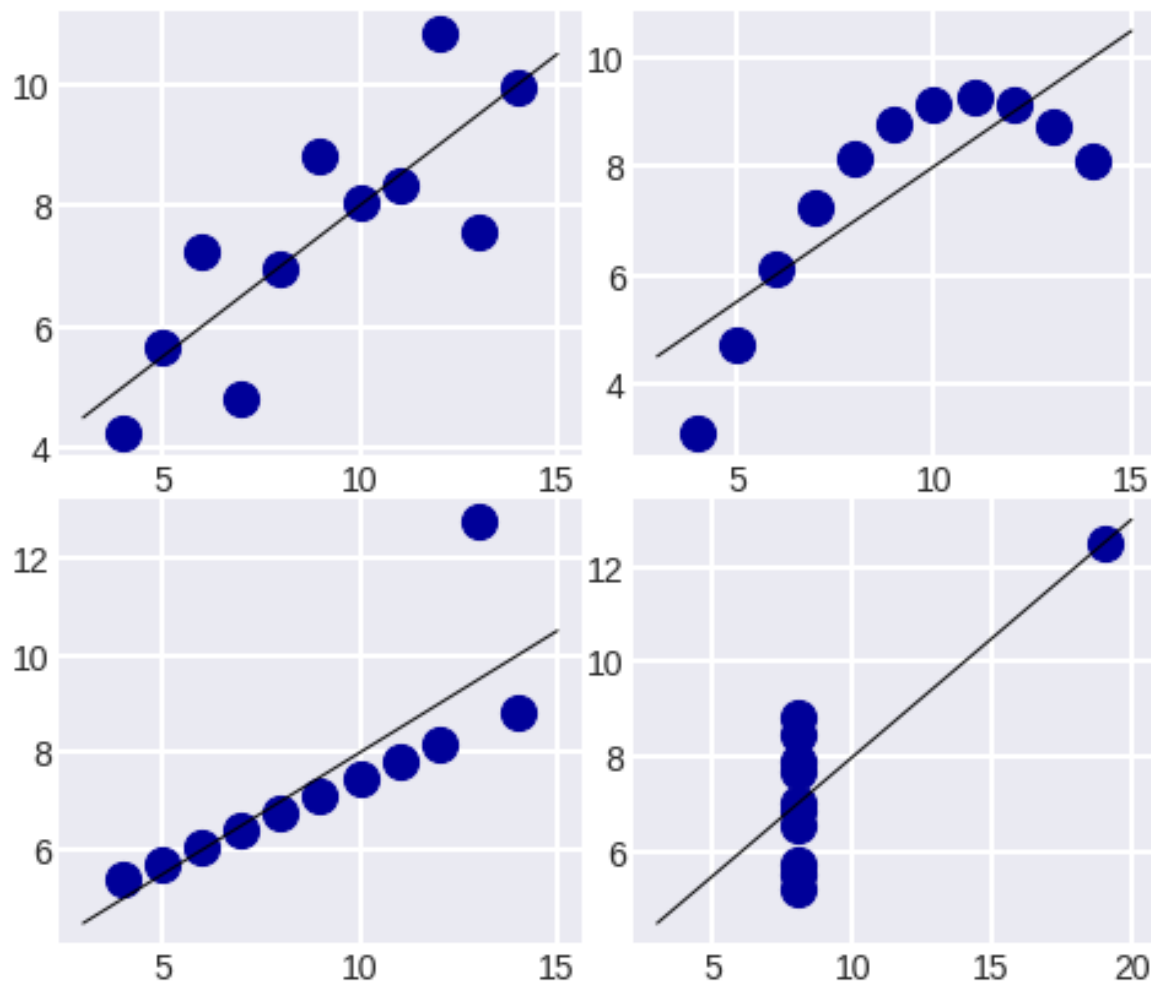
Часть 1. Историческая

Дьяконов А.Г.

**Московский государственный университет
имени М.В. Ломоносова (Москва, Россия)**

Зачем смотреть на данные?

Наборы данных имеют идентичные статистические характеристики, но их графики существенно различаются.



Характеристика	Значение
Среднее значение переменной X	9,0
Дисперсия переменной X	10,0
Среднее значение переменной Y	7,5
Дисперсия переменной Y	3,75
Корреляция между переменными	0,816
Прямая линейной регрессии	$Y=3+X/2$

F.J. Anscombe Graphs in Statistical Analysis // American Statistician, 27 (February 1973), 17-21.

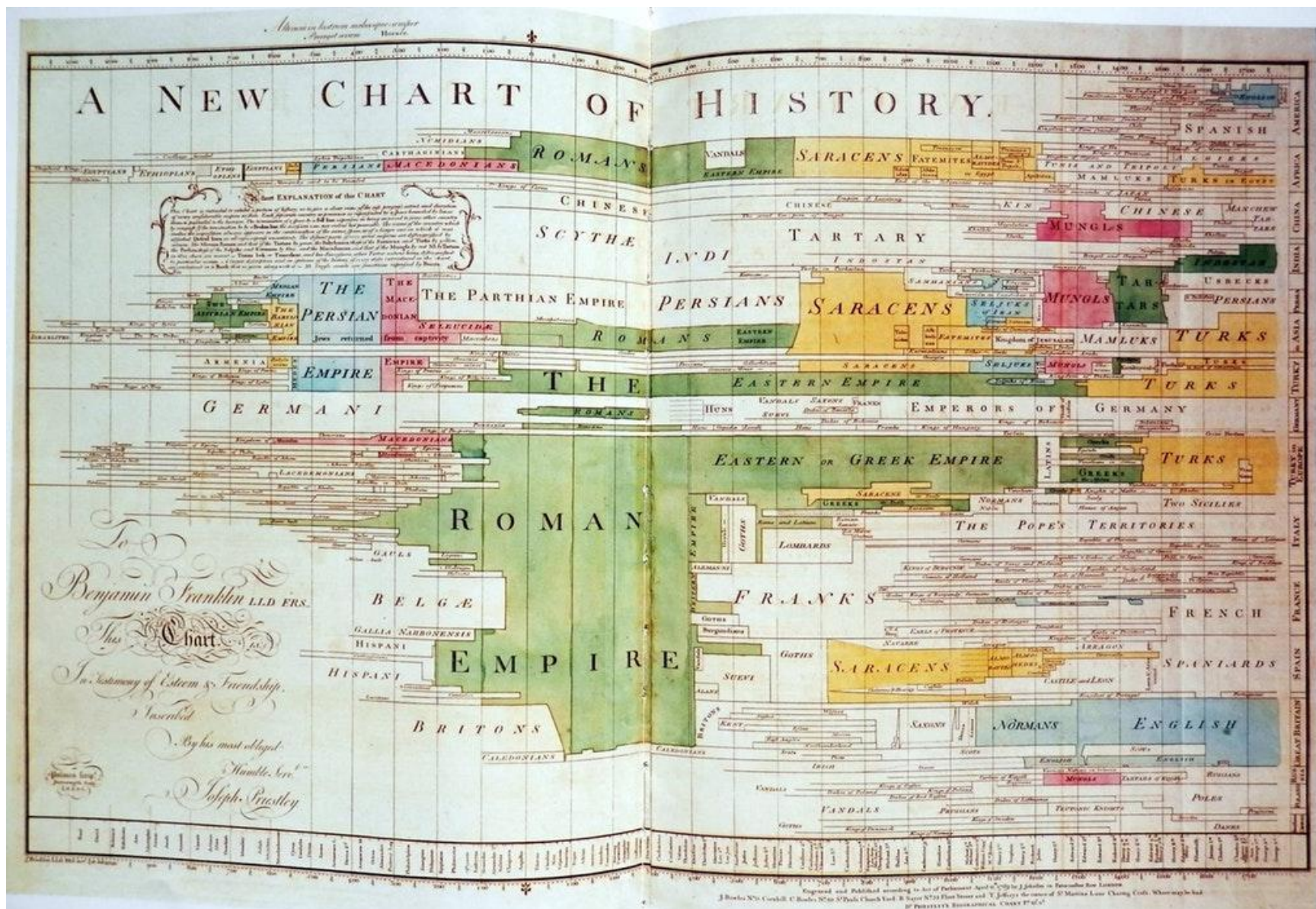
Немного об истории визуализации

**инфографика
виды графиков
графический анализ**

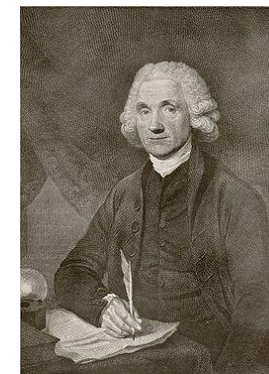
18 век – зарождение инфографики

19 век – протоанализ данных

18 век Джозеф Пристли



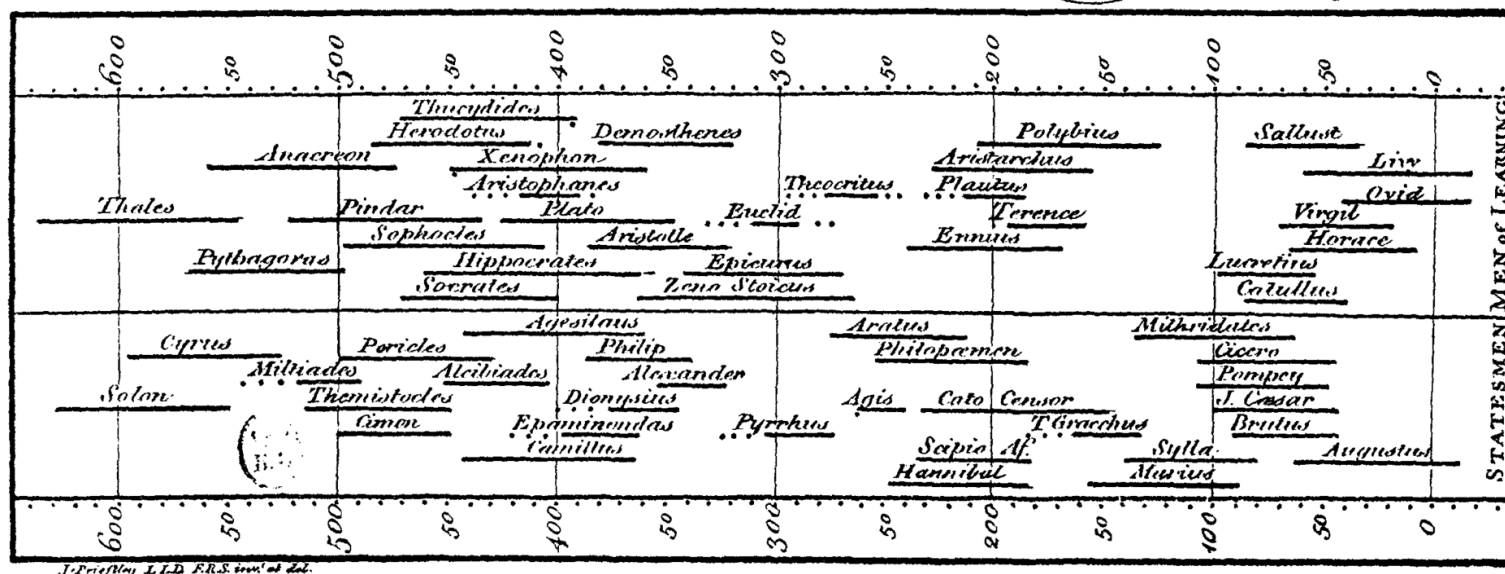
18 век Джозеф Пристли



Joseph Priestley

(13.03.1733 – 6.02.1804) британский священник, естествоиспытатель, философ. Открыл кислород.

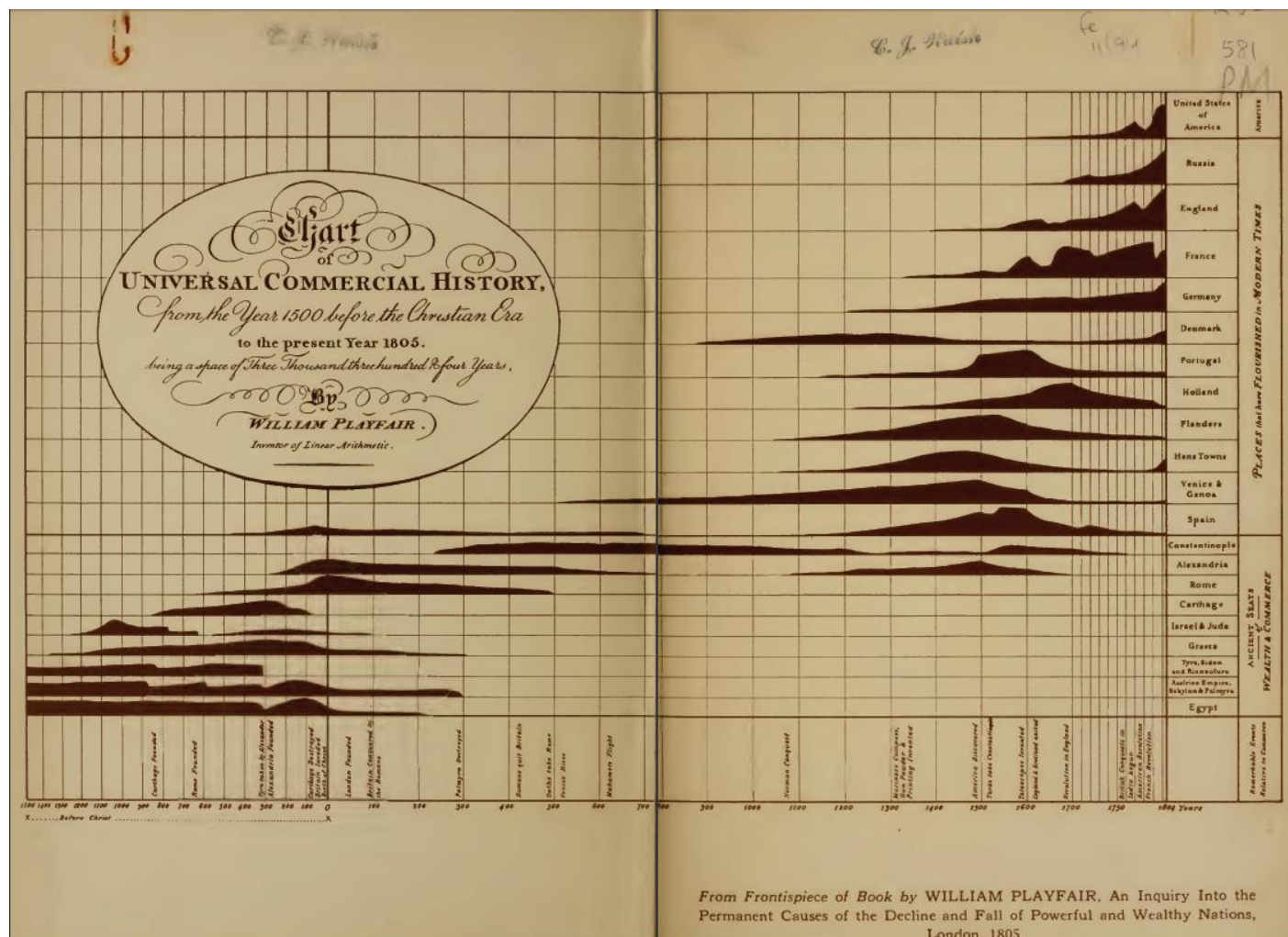
A Specimen of a Chart of Biography.



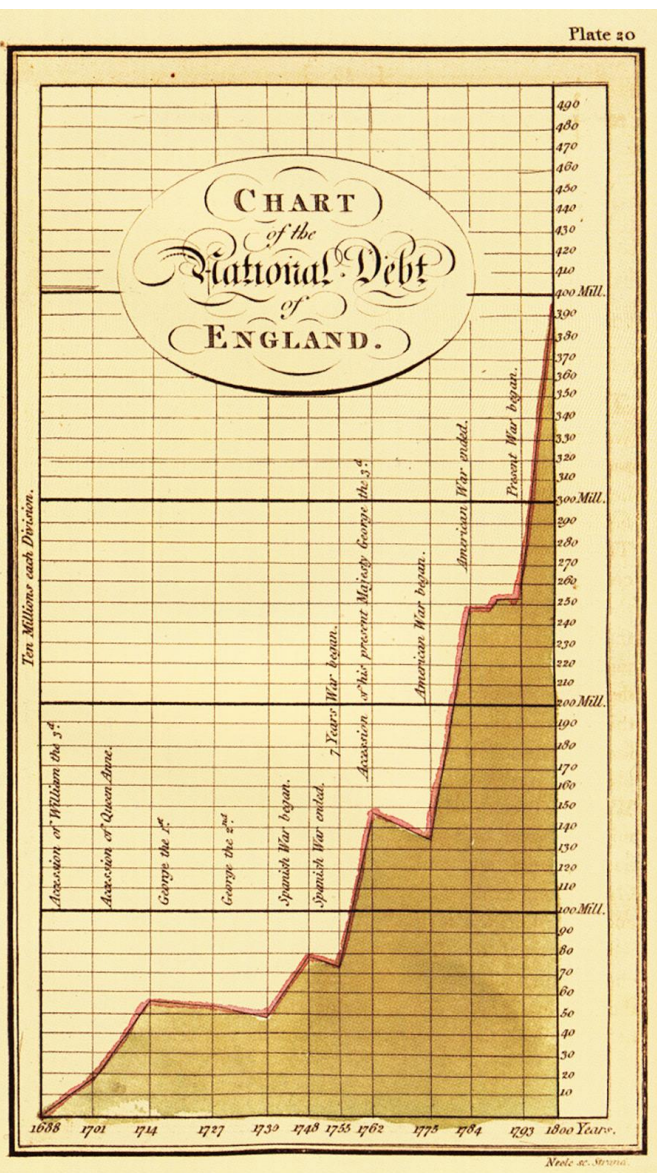
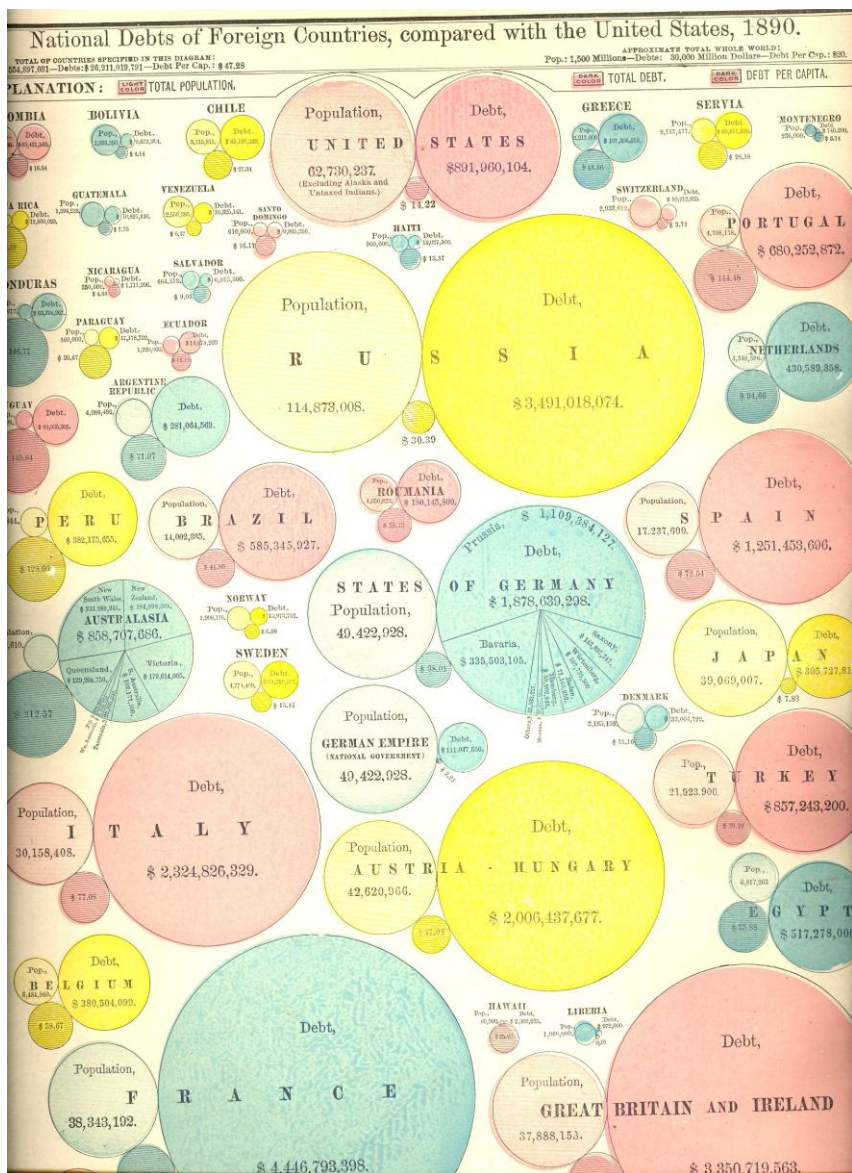
J. Priestley L.L.D. F.R.S. inv. et del.

18 век Уильям Плейфэр

- 1786 – линейчатый график и гистограммы
- 1801 – секторная диаграмму в круге и круговая диаграмма

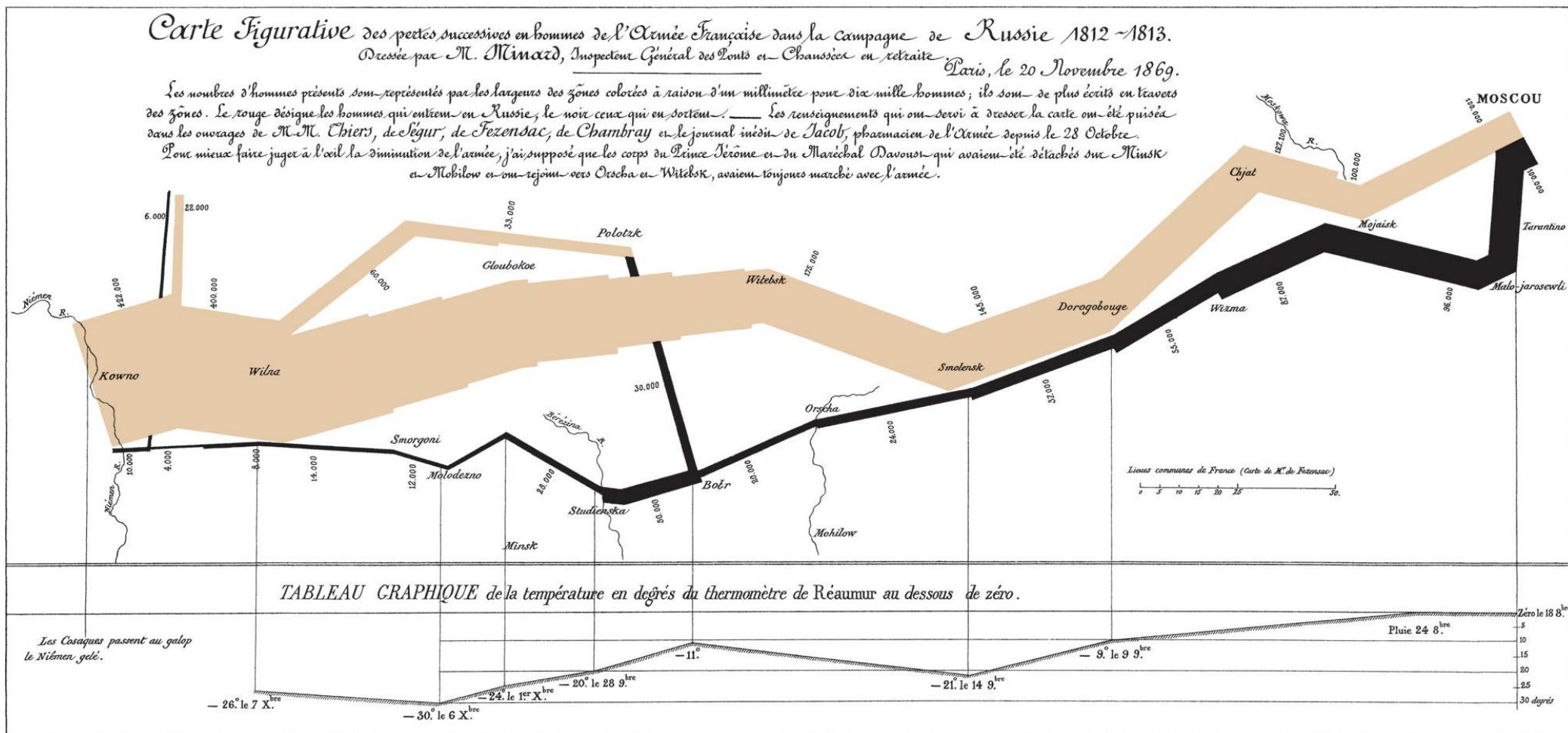


18 век Уильям Плейфэр

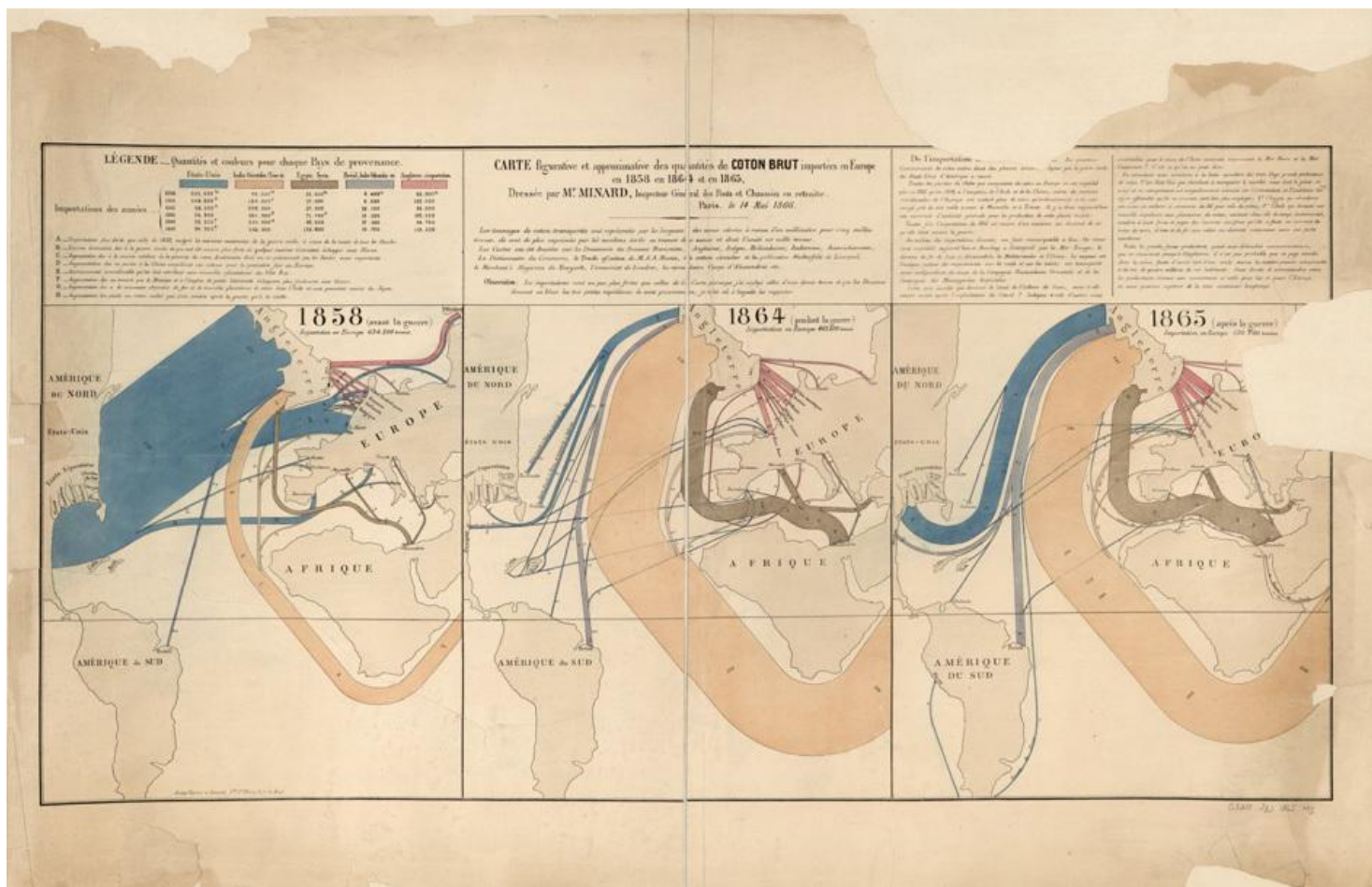


22.09.1759 – 11.02.1823
шотландский инженер,
основатель
графических методов
статистики.

19 век Шарль Жозеф Минар



19 век Шарль Жозеф Минар

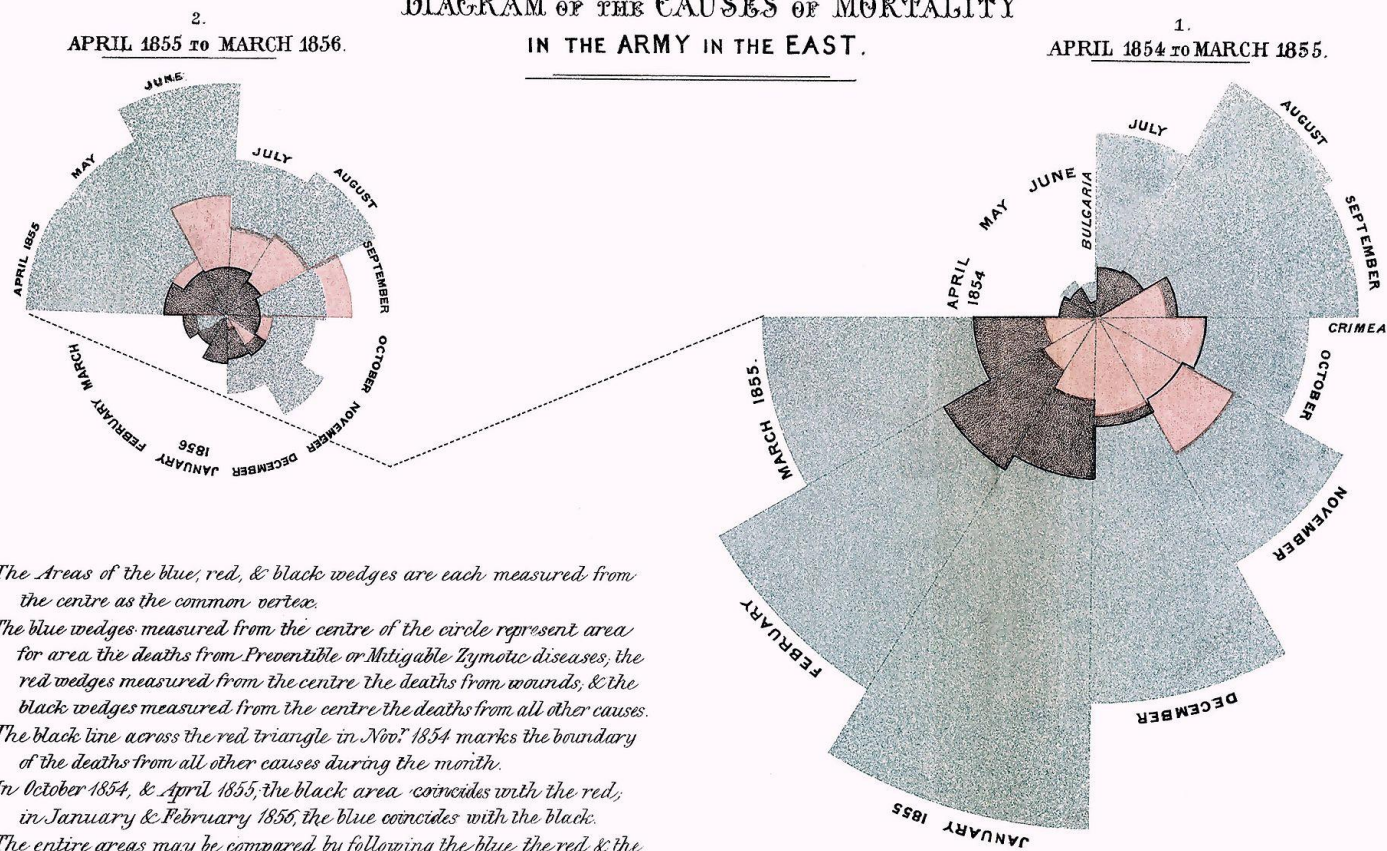


Шарль Жозеф Минар
27.03.1781 – 24.10.1870
 французский инженер,
 топограф, пионер в
 области графических
 методов анализа и
 представления
 информации в области
 инженерных наук и
 статистики

19 век – Флоренс Найтингейл

«Петушинный гребень»

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.



Florence Nightingale

**12.05.1820 –
13.09.1910**

**сестра милосердия
и общественная
деятельница
Великобритании.**

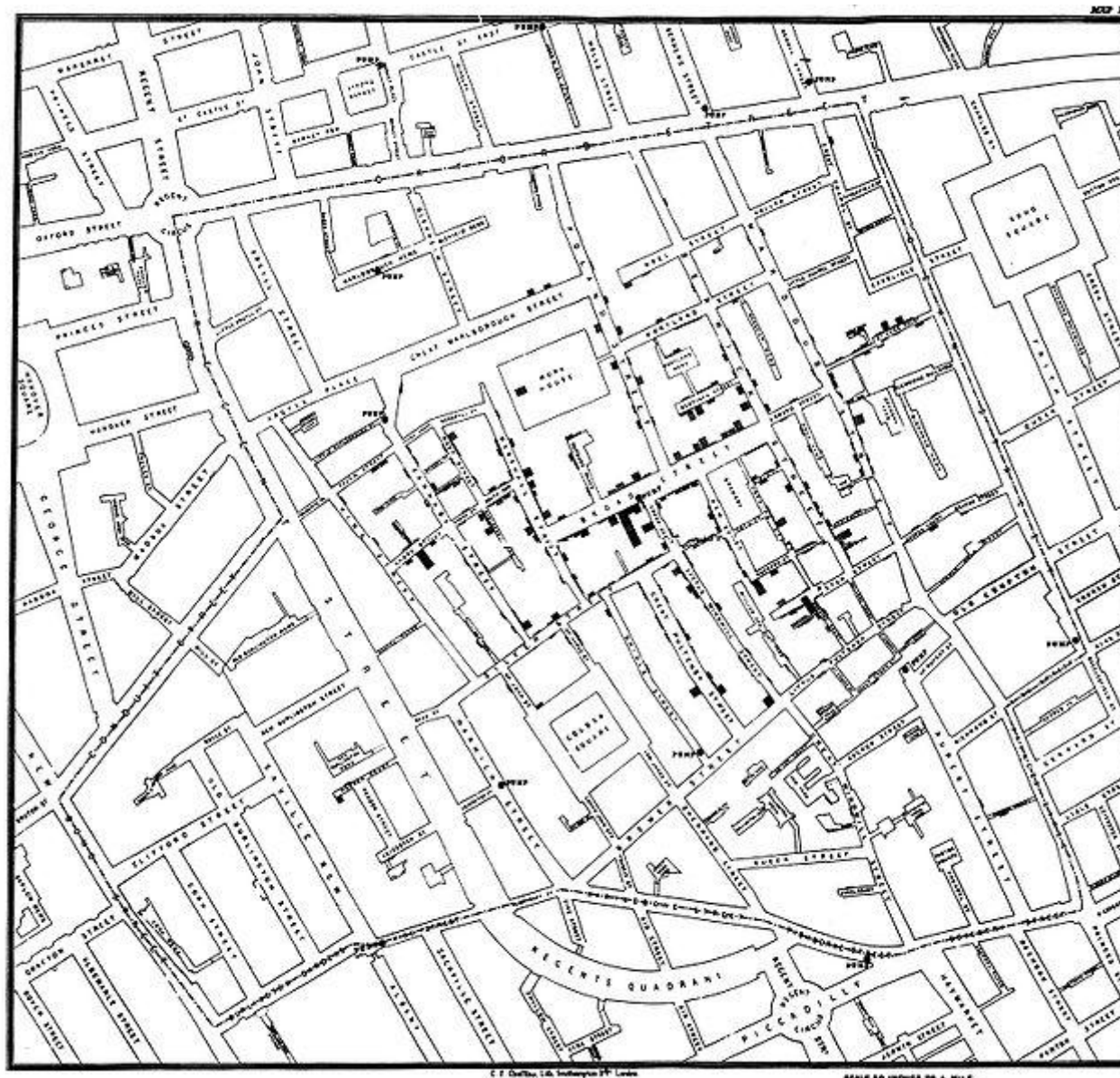
Смертность солдат во время крымской войны.

Площадь каждого сектора пропорциональна смертности.

Голубой – смертность от болезней, красный – от ран, и коричневый слой – от других причин.

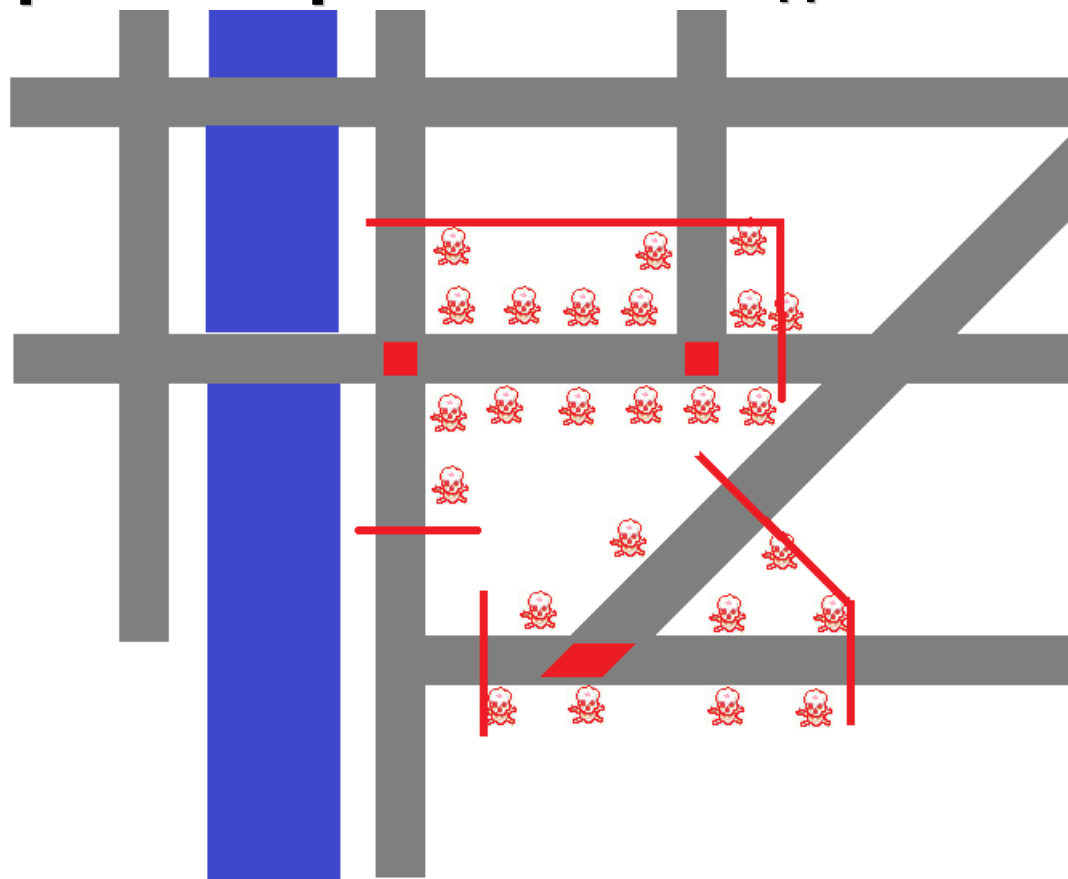
Вспышка холеры на Брод-стрит в 1854 году

См. Википедию



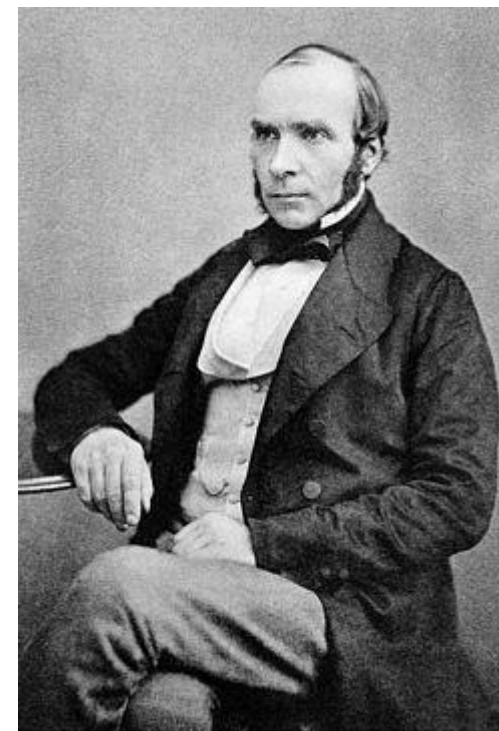
Центры эпидемии – колодцы!

Диаграмма Вороного была видна на карте...



**Нечистоты сливались в Темзу,
в результате была заражена
местная система водоснабжения.**

Джон Сноу



John Snow

(15.03.1813 — 16.06.1858)
британский врач, один из пионеров
массового внедрения анестезии и
медицинской гигиены

Что это за данные?



Что это за данные?



Ответ**Плотность внимания на Интернет-странице****Измеряется по числу и продолжительности фиксаций**

**L. A. Granka, H. A. Hembrooke, G. Gay, M. K. Feusner Correlates of Visual Salience and Disconnect:
An Eye-tracking Evaluation**

**A. Santella D. DeCarlo Robust Clustering of Eye Movement Recordings for Quantification of Visual
Interest**

Viewer 1:



Viewer 2:



POR data

Gaze clusters with $\sigma_x = 100$, $\sigma_y = \frac{1}{3}$ Regions-of-interest with $\sigma_x = 100$

Что за данные?

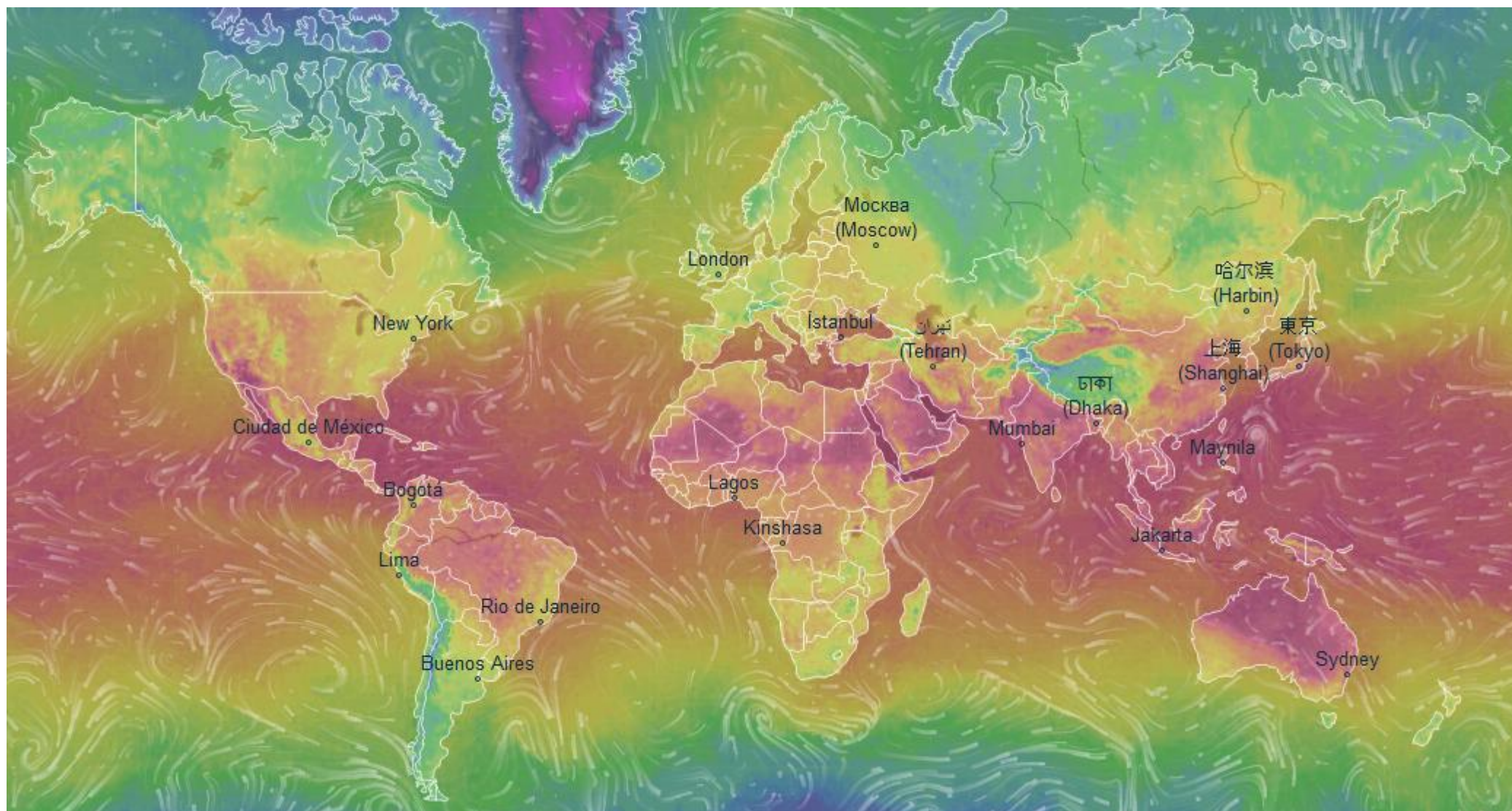


Что за данные?

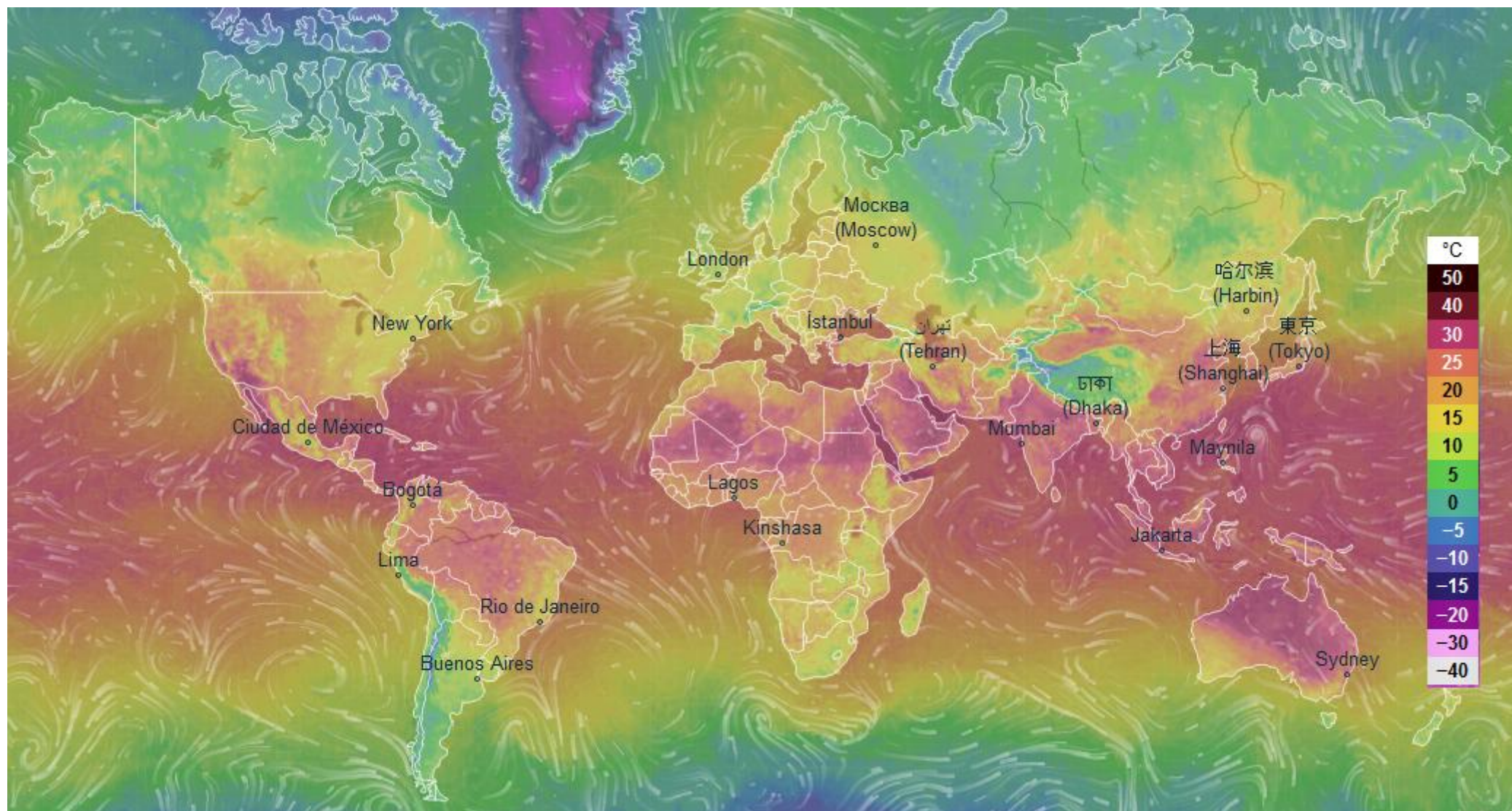


Трафик авиарейсов США

Что за данные?

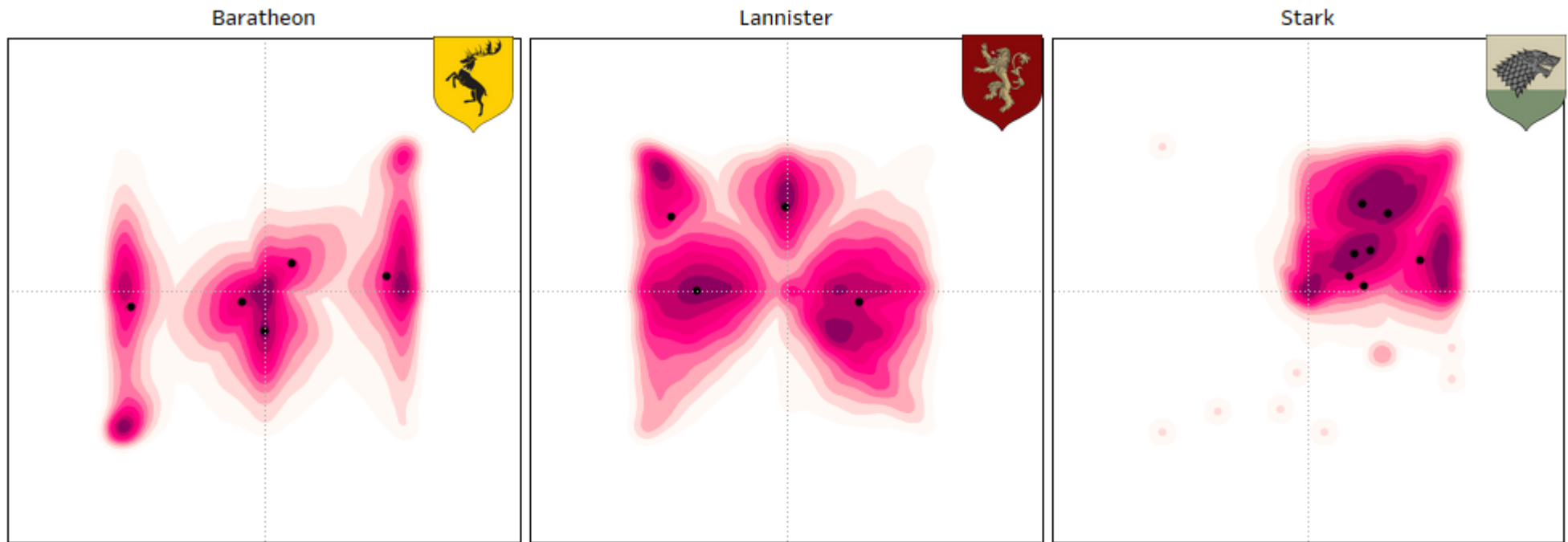


Что за данные?

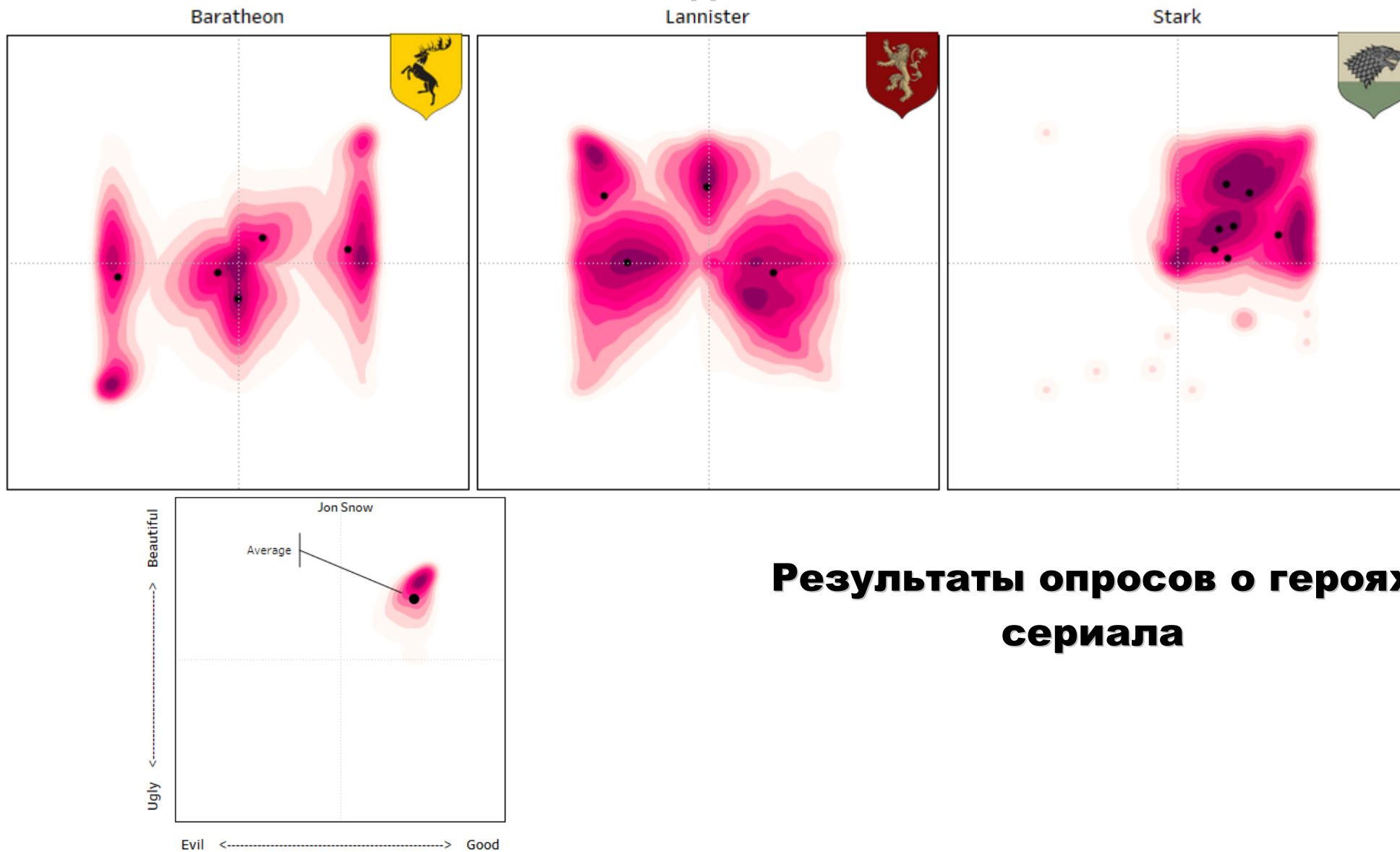


Температура (2м над землёй) <https://www.ventusky.com/>

Что за данные?



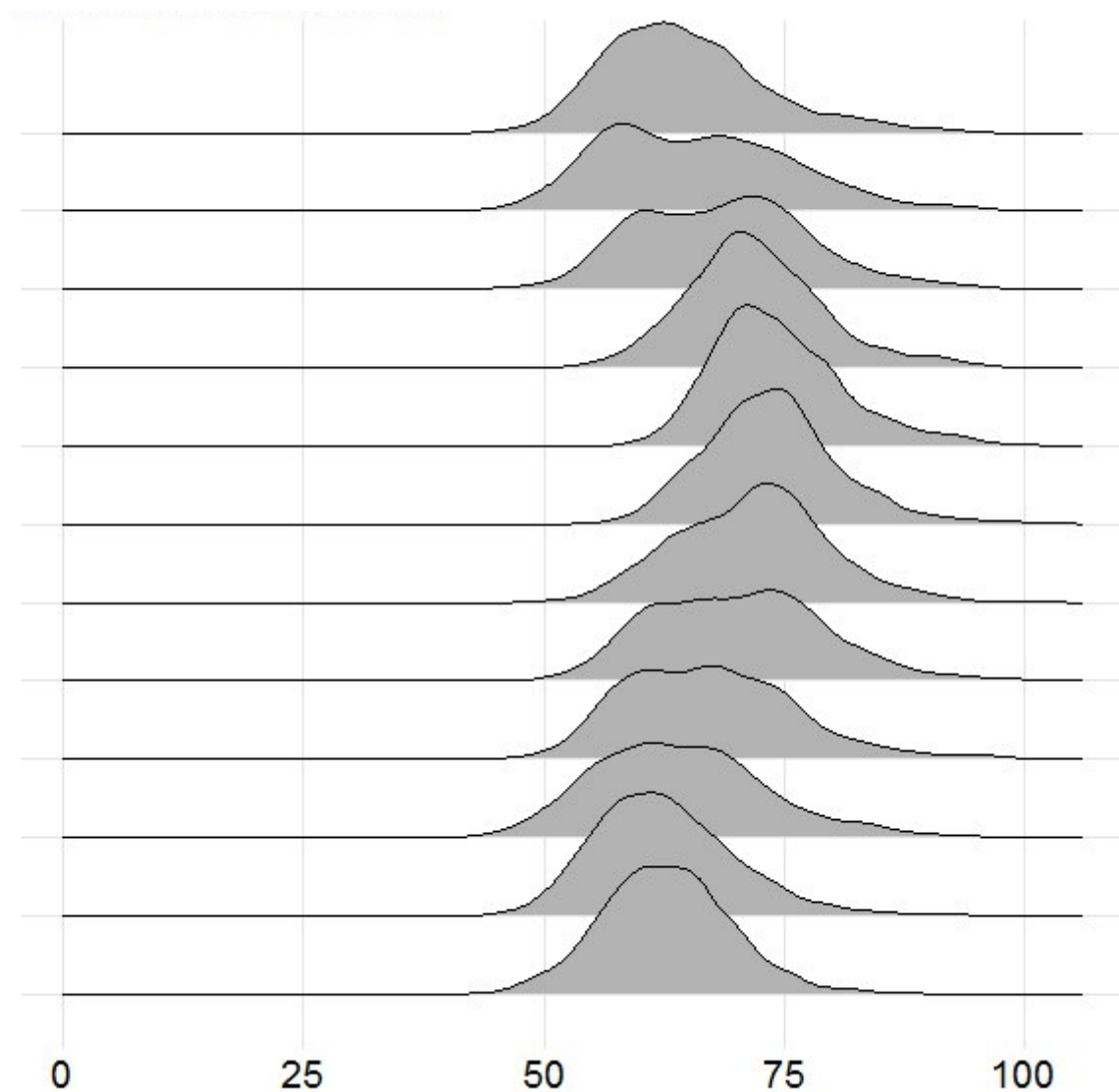
Что за данные?



Результаты опросов о героях сериала

<https://public.tableau.com/profile/nicco.cirone#!/vizhome/GameofThronesastudyofthemaincharacters/GameOfThronesAstudyoncharacters>

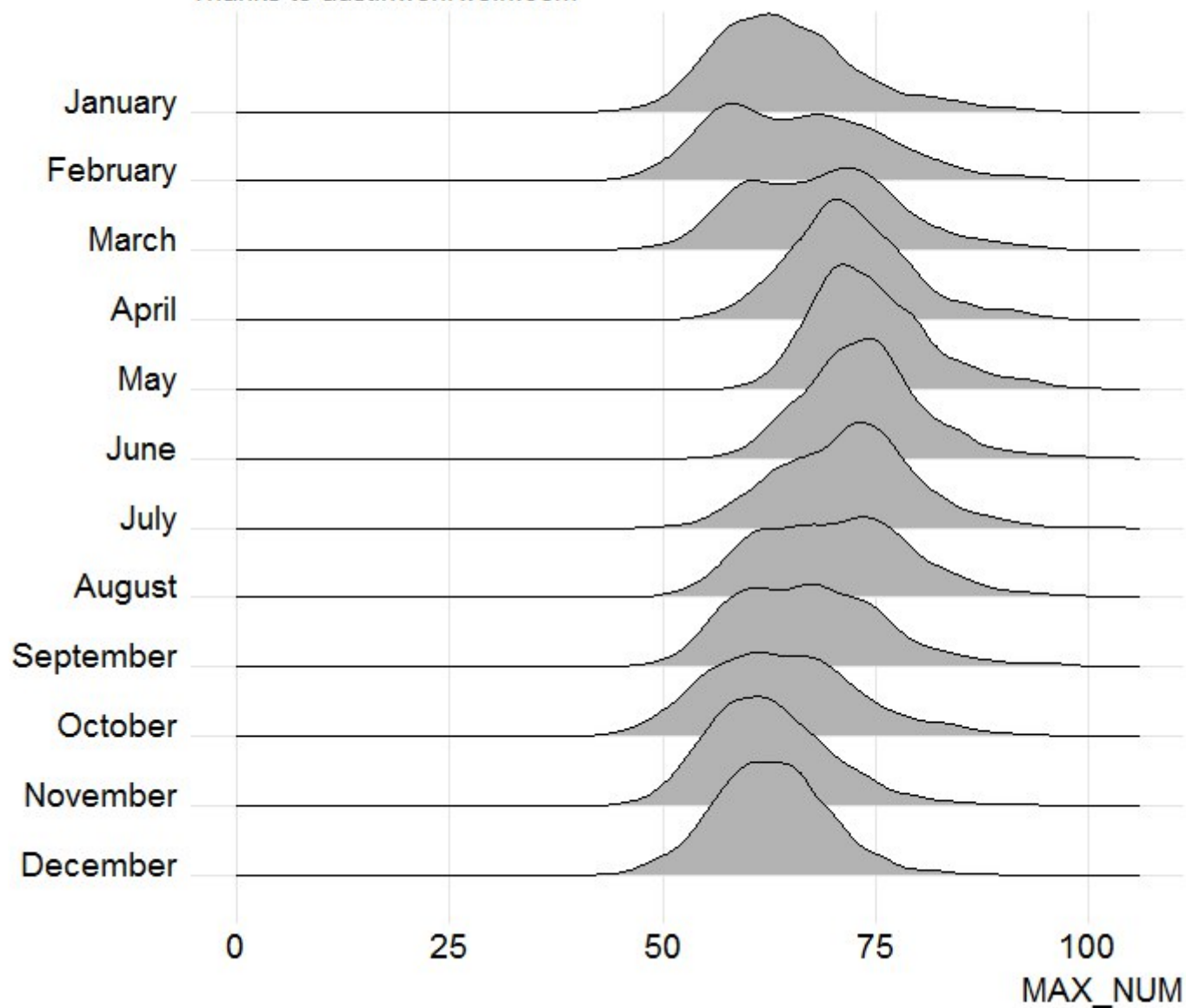
Что это за данные?



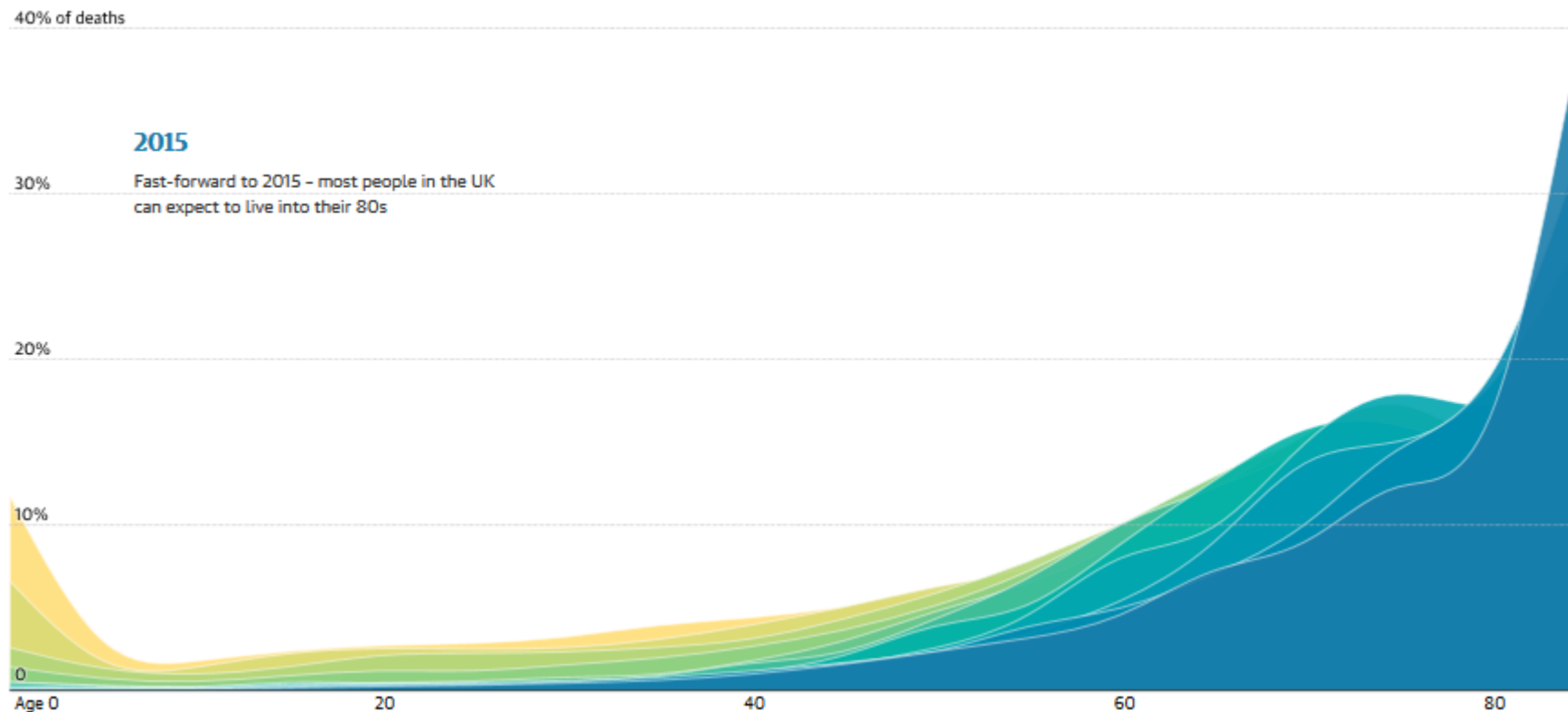
Что это за данные?

MAX Daily Temperatures in Mountain View, CA (1950-2016)

Thanks to austinwehrwein.com



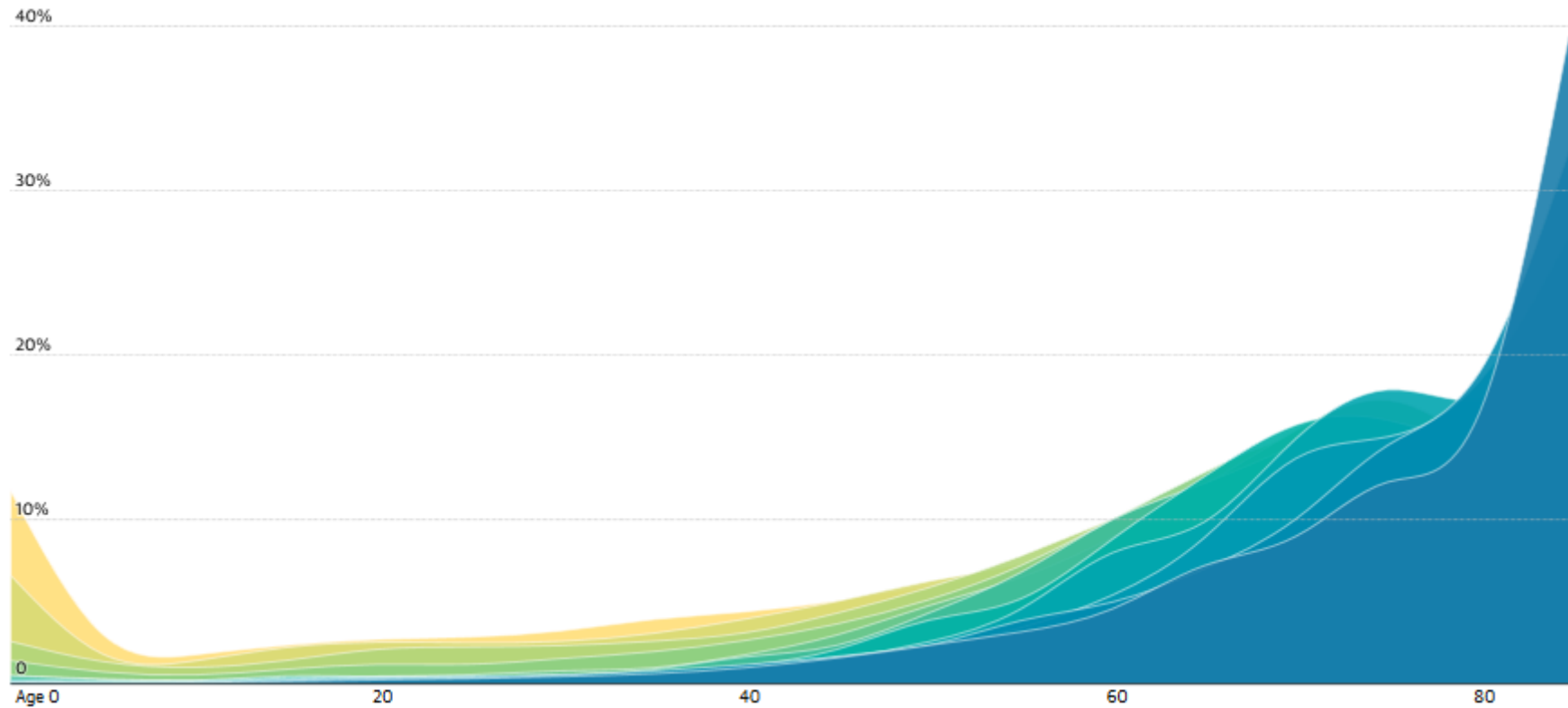
Что за данные?



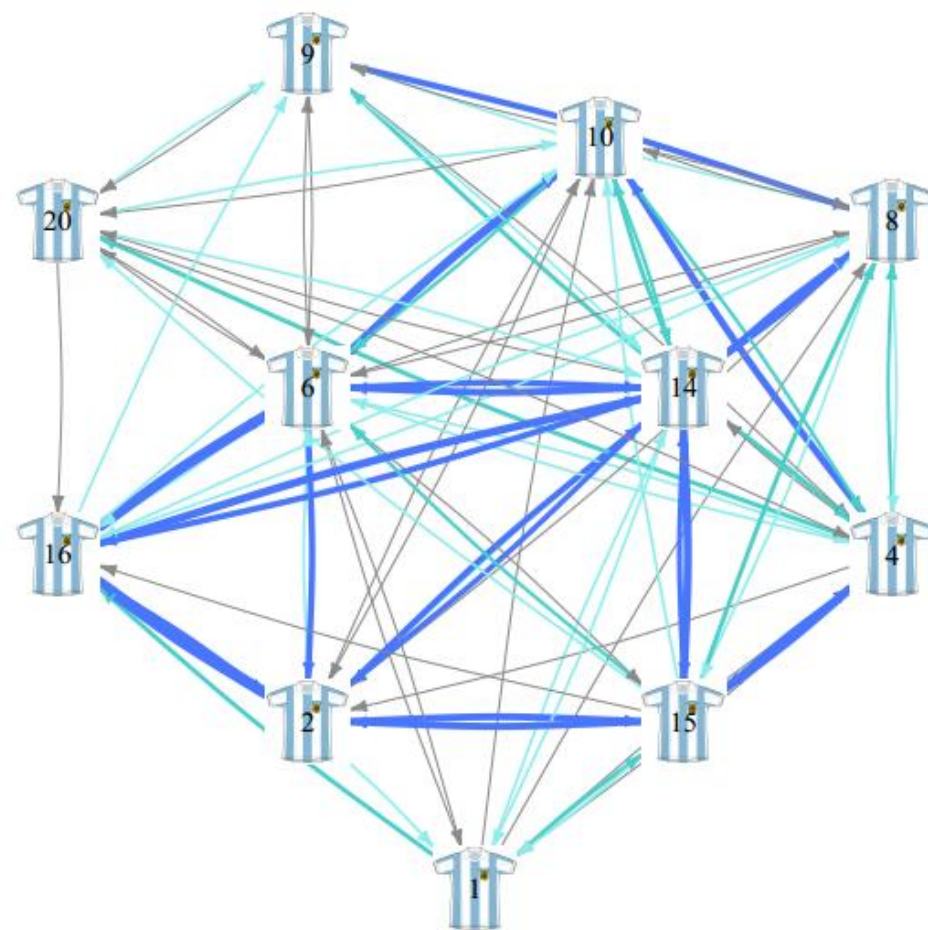
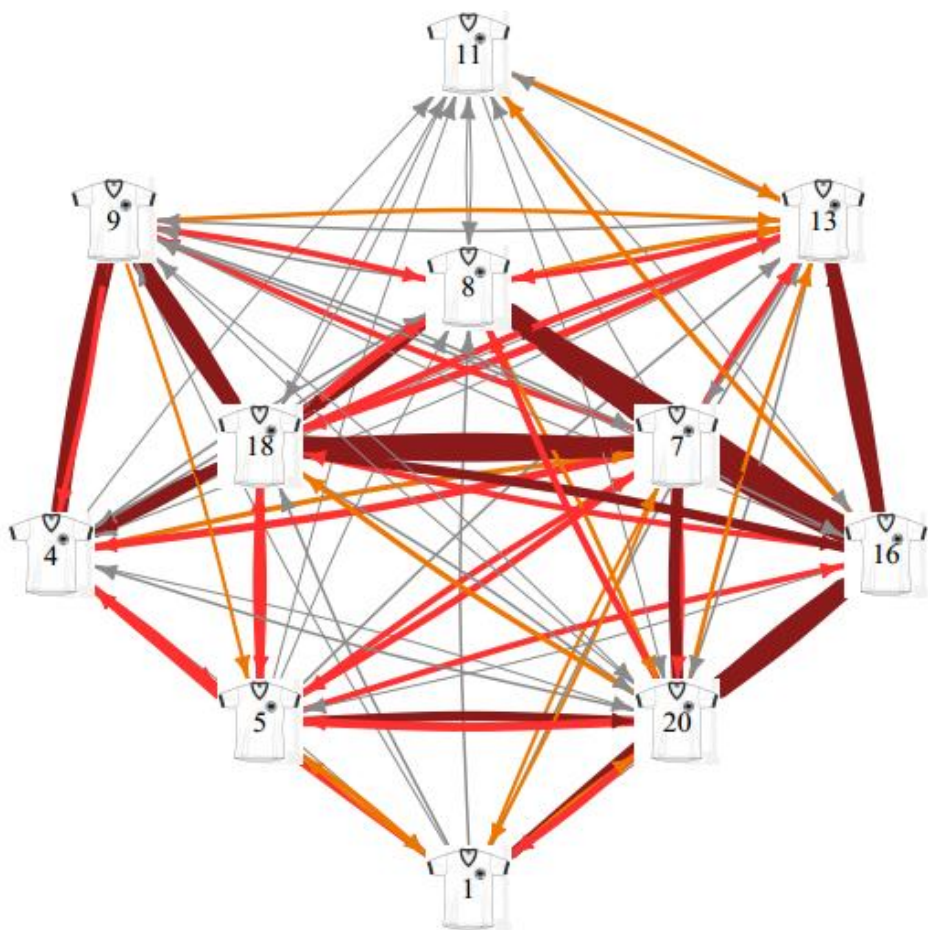
Плотность распределения смертности по возрасту в разные годы в Великобритании

<https://www.theguardian.com/lifeandstyle/ng-interactive/2017/sep/18/how-death-has-changed-over-100-years-in-britain>

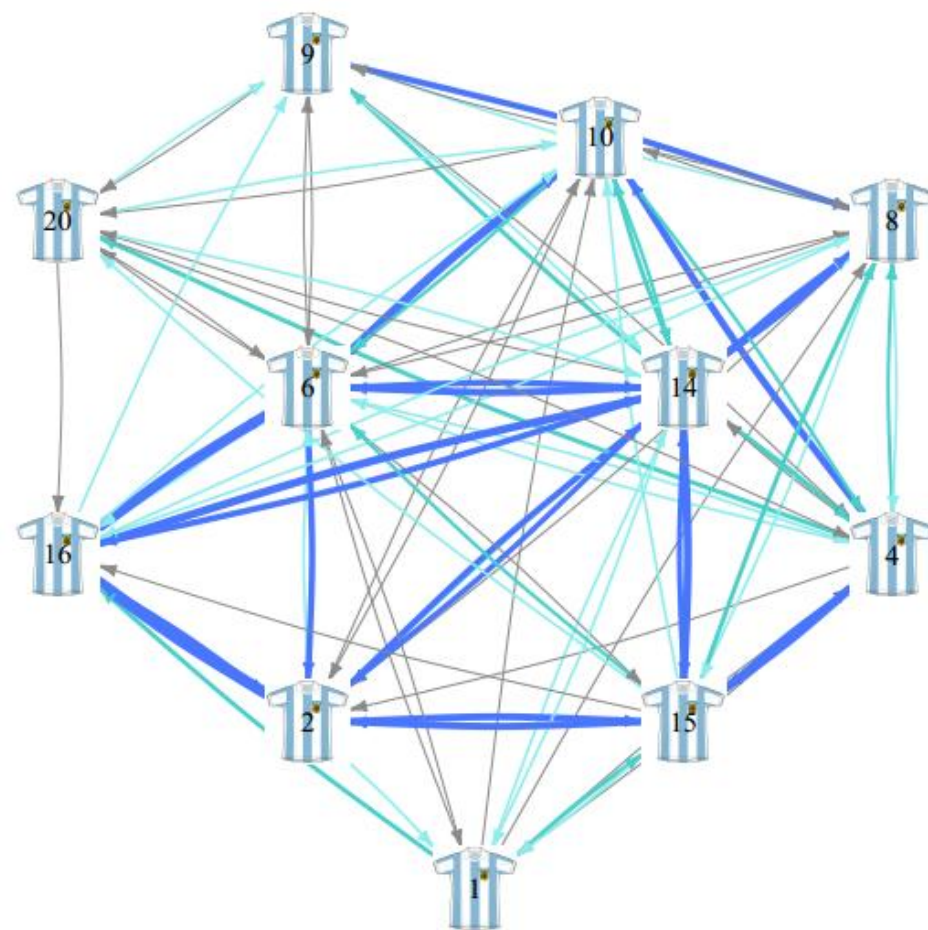
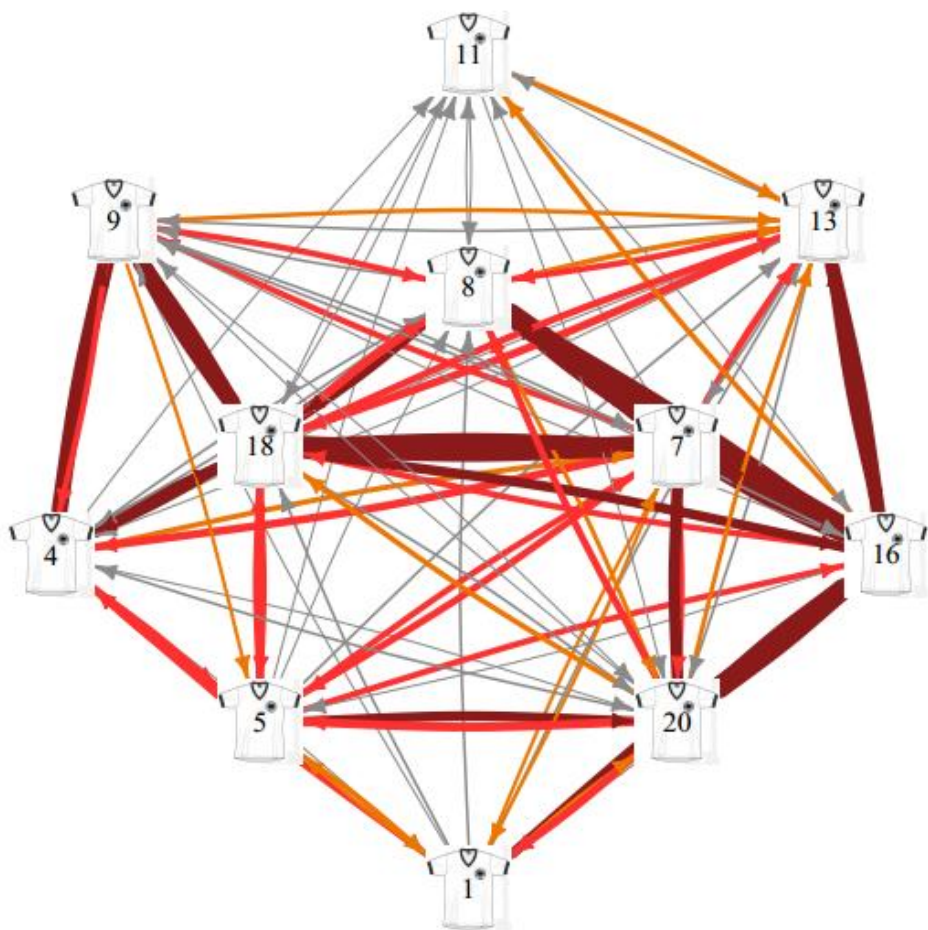
Что за данные?



Что за данные?

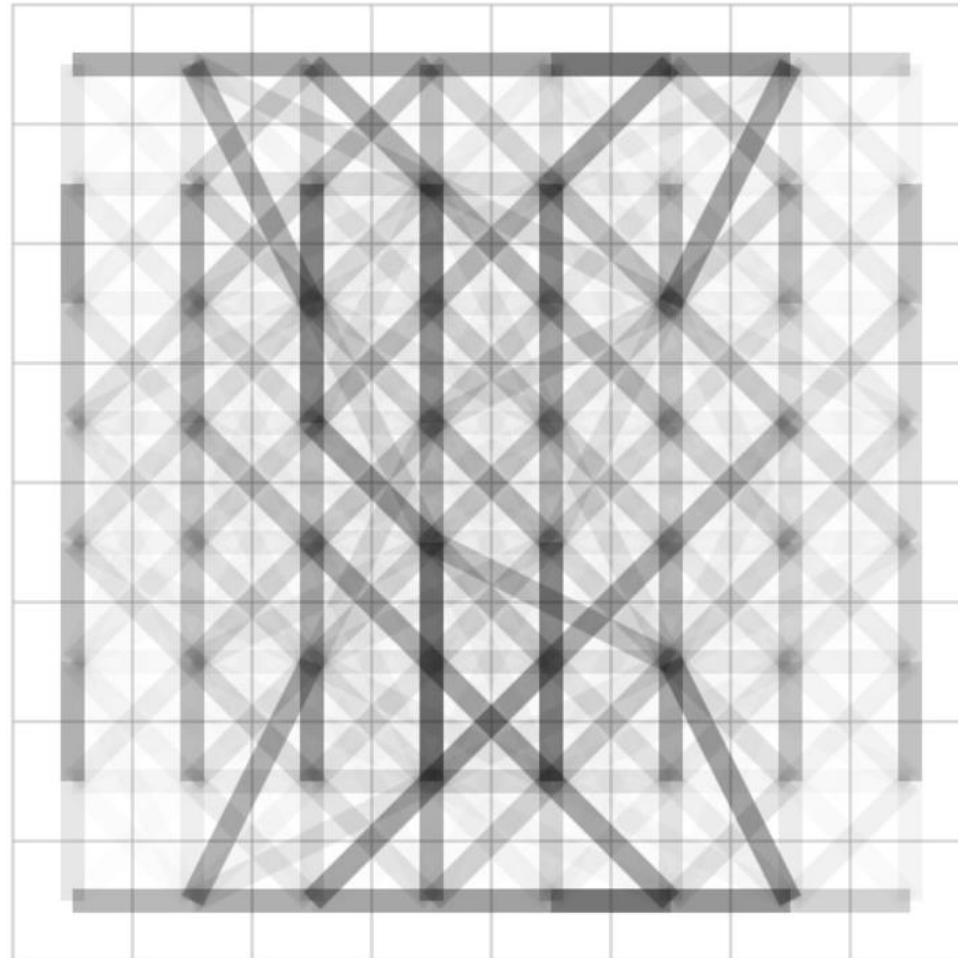


Что за данные?

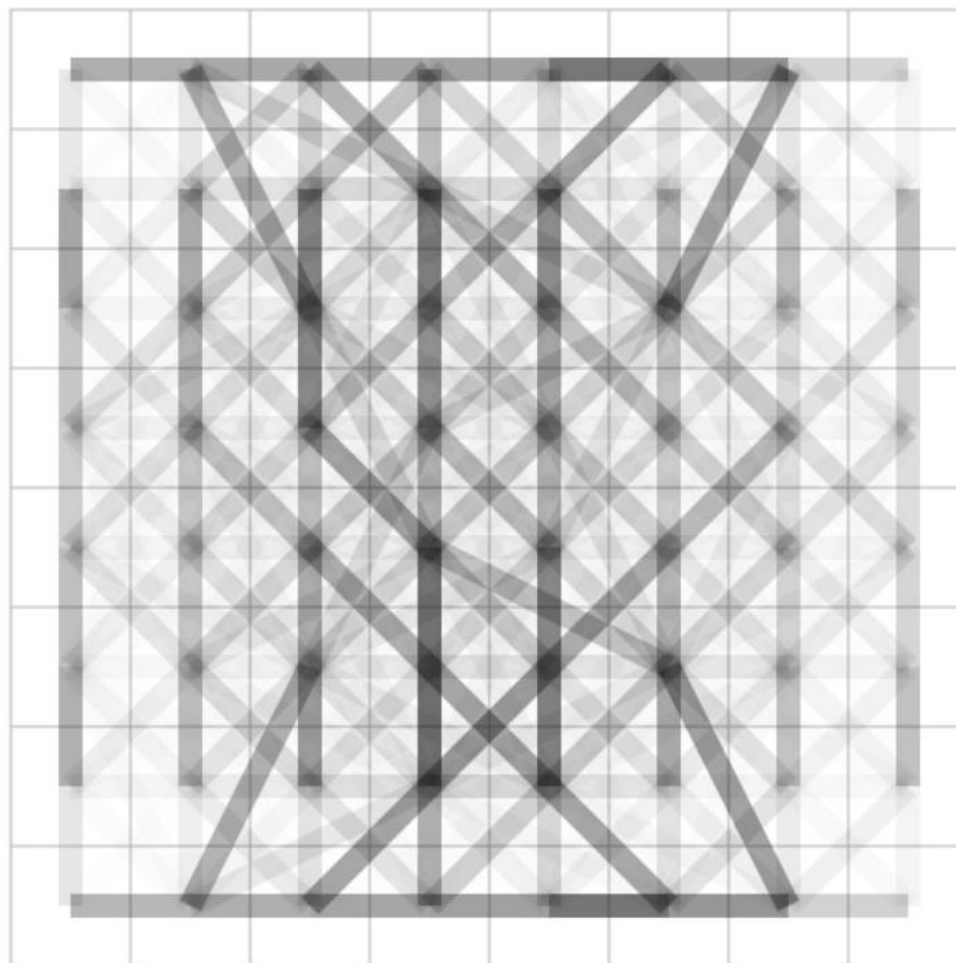


Передачи в финале ЧМ-2014 Германия – Аргентина

Что за данные?

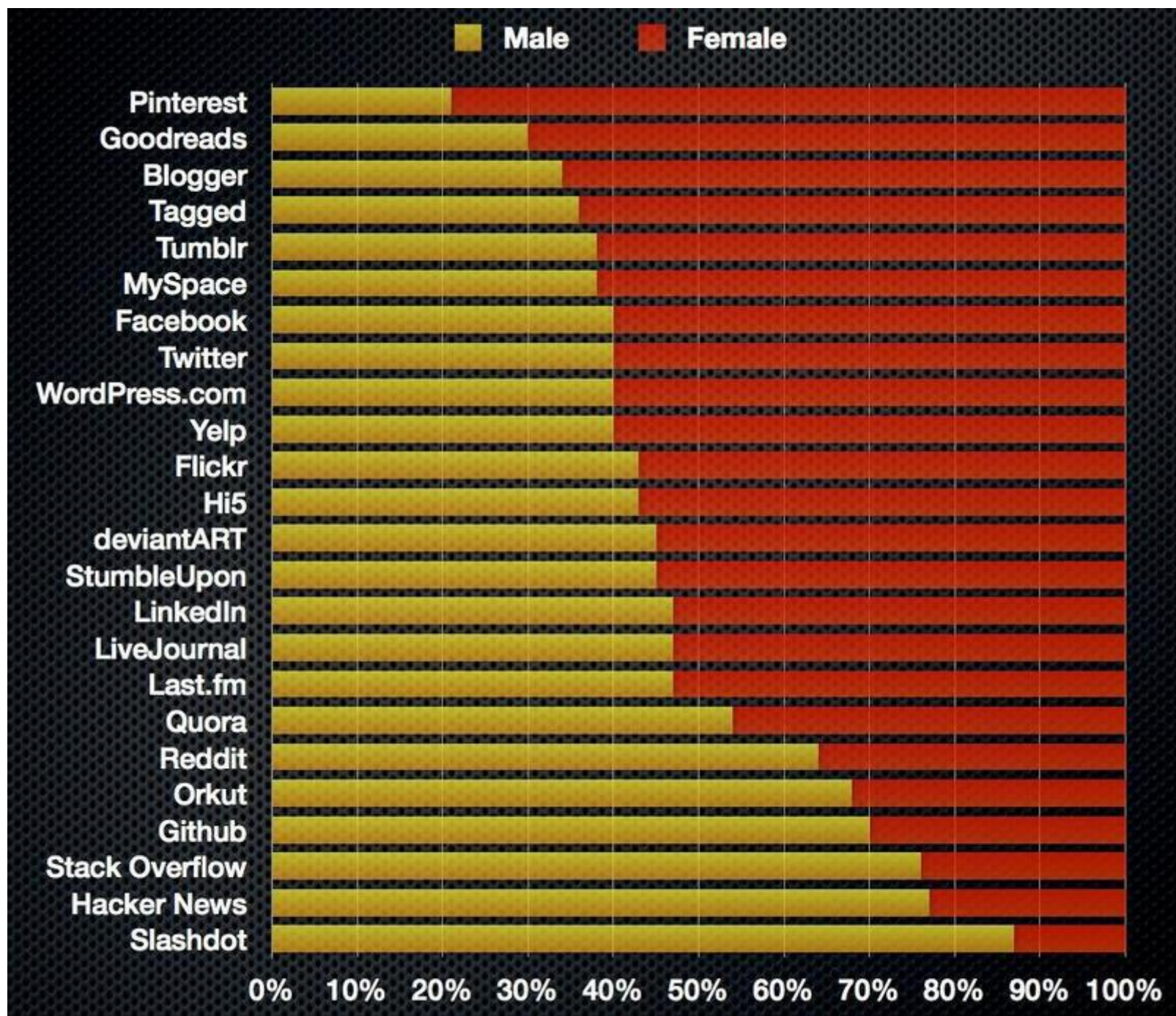


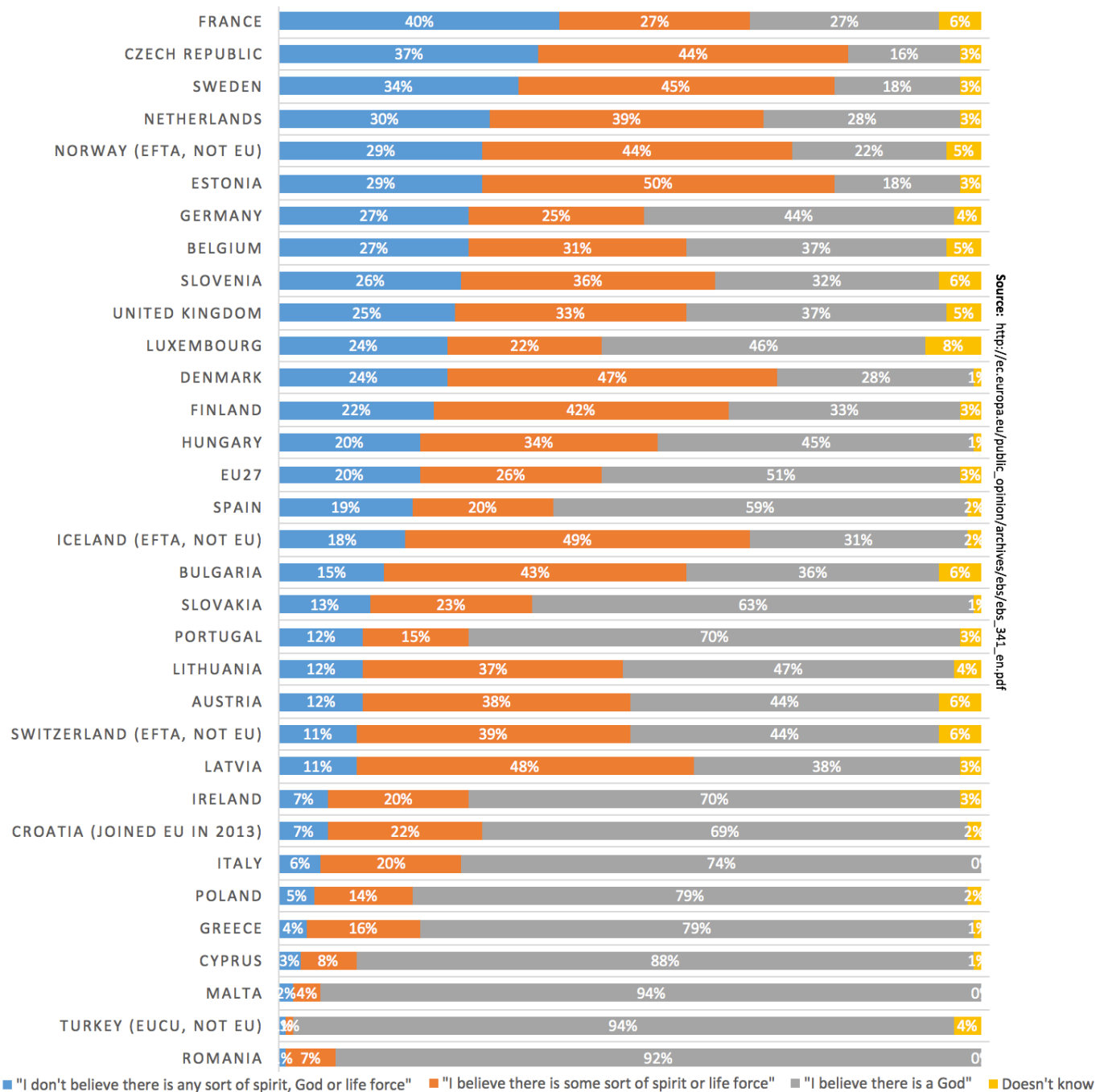
Что за данные?



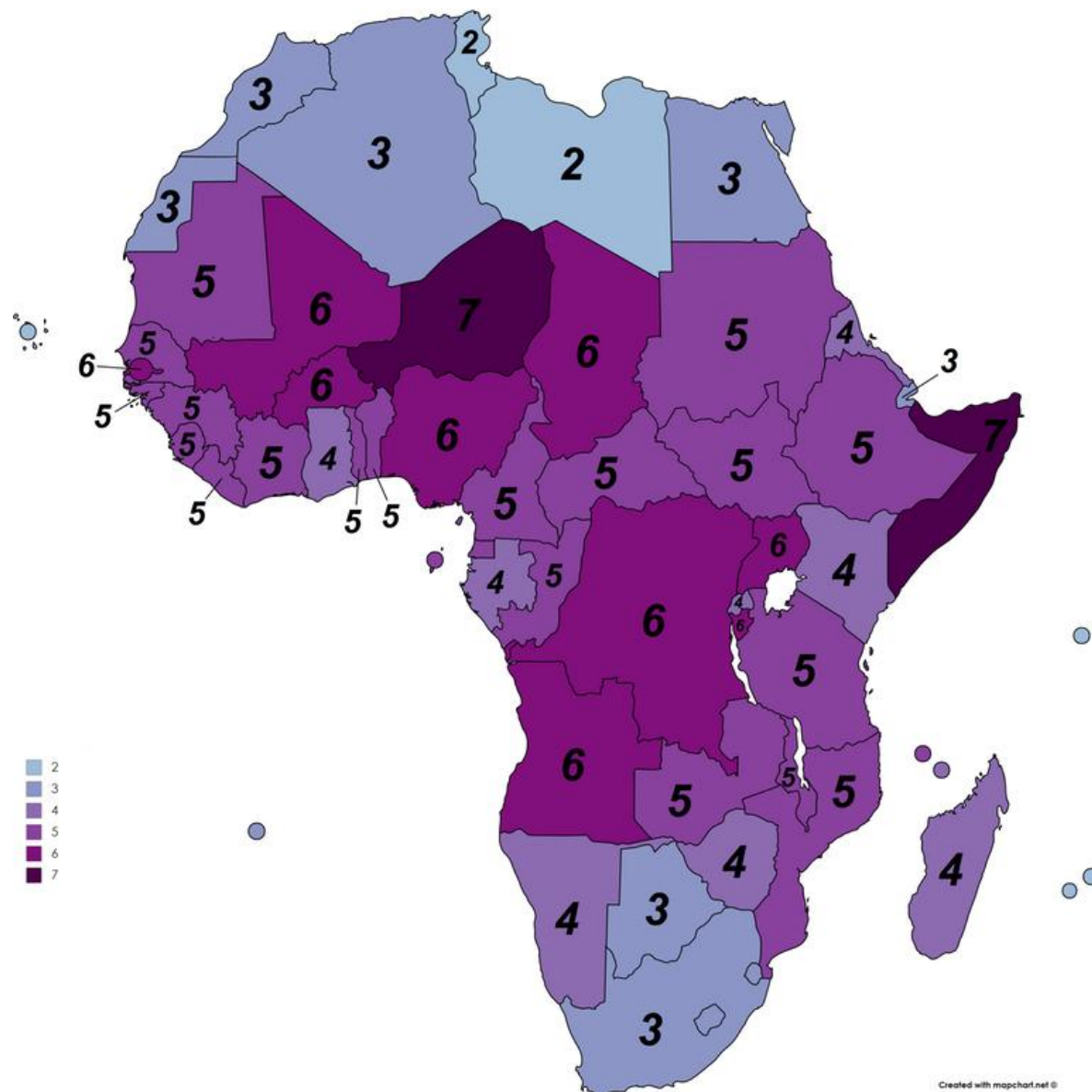
Перемещение фигур в шахматных партиях

<https://github.com/timhutton/chessviz>



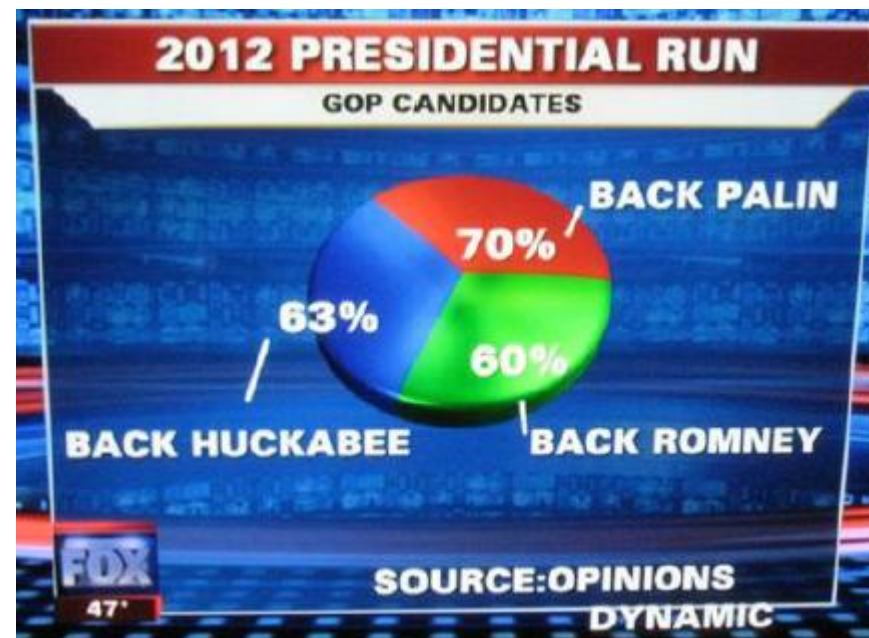
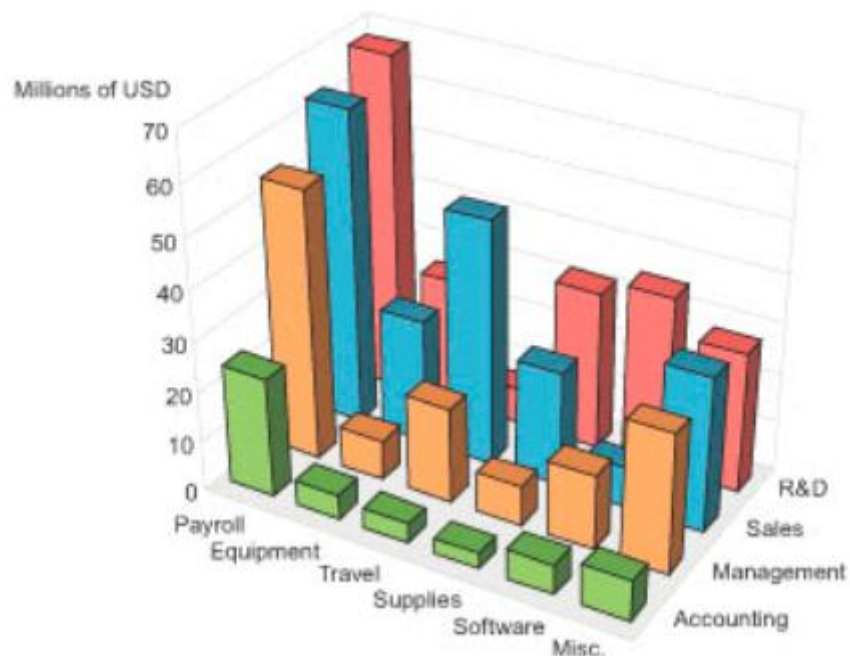


Среднее число детей на одну женщину



Основное правило

3D – плохо!



Домашнее задание

Найти интересные нетривиальные визуализации