

На примере анализа и перепроектировании модели визуализации рассмотрим инфографику, представленную в отчёте Программы международной оценки (Pisa) за 2012 год. Данная инфографика, представлена на [рисунке 1](#). Она показывает результаты прохождения детьми 15-ти лет из различных стран и регионов тестов по математике, навыкам чтения и естественным наукам. Таким образом, входными данными для её построения являются 4 параметра: название страны, а так же средние результаты по трём различным тестам. Согласно варианту задания, задача состоит в том, чтобы определить, какие страны справились лучше остальных, какие хуже, а также как они сравниваются по каждому из предметов. Кроме того, в заголовке инфографики указано, что с другими странами сравнивается Соединённое Королевство Великобритании и Северной Ирландии. Поэтому акцент делается именно на этой стране.

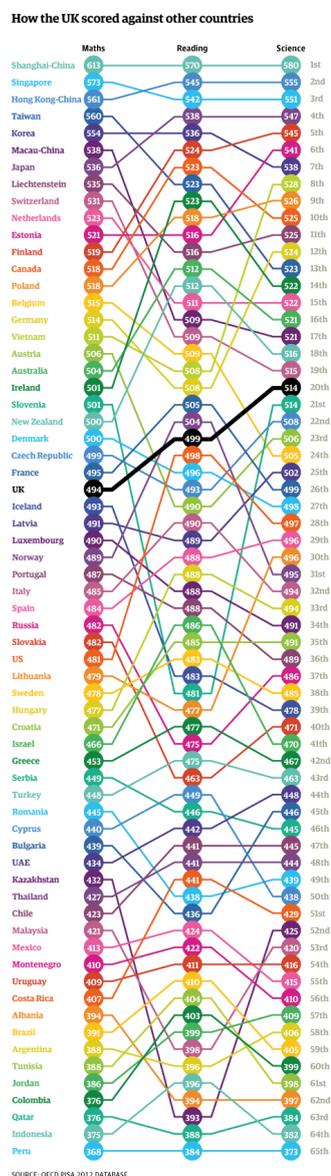


Рисунок 1. Результаты Pisa 2012 - как соотносятся результаты по странам и предметам

Графическим символом являются круги, выстроенные в три колонки. Цветом они связаны с находящимися слева названиями соответствующих стран. Так же три соответствующие одной стране круга связаны ломаной. Внутри круга белым шрифтом подписан результат по соответствующему колонке предмету. Страны отсортированы по результатам математического тестирования. Так же справа приведена позиция страны в общем рейтинге в числовом выражении. Интересующая нас страна выделена увеличенной толщиной соединительной ломаной и контрастирующим чёрным цветом.

Инфографика нарушает принцип выразительности, так как она не отображает отношение порядка результата за тест относительно среднего значения. Также сортировка по результатам математического теста не даёт удобно сравнить страны по двум другим дисциплинам, чем тоже нарушается принцип выразительности. Кроме того, из-за большого количества стран, палитра цветов получается слишком большой, что в совокупности с запутанным изображением связей затрудняет читаемость визуализации и целостное восприятие, что можно классифицировать как нарушение принципа эффективности. Сравнение всех трёх категорий тестовых данных на одном графике привело к загромождению визуализации, что также ухудшает её.

Таким образом, попытаемся решить рассмотренные выше проблемы, не повредив эффективности модели.

Финальный результат отображён на [рисунке 2](#), а далее попробуем описать и обосновать процесс построения исправленной визуализации. Во-первых, для того, чтобы уменьшить зашумлённость модели, разнесём каждый предмет на отдельный график. По горизонтали будем отмечать страны, а по вертикали оценку за тест. Ввиду большого количества стран и ограниченного места по горизонтали, применим для обозначения стран общепринятые двухбуквенные коды (согласно стандарту ISO 3166-1). Так же под ними подпишем их рейтинг для данного предмета.

Для изображения оценки применим столбчатую диаграмму. Проградуируем вертикальную шкалу с шагом 50. В выборке отсутствуют значения ниже 350, поэтому целесообразно начинать ось с него. Таким образом лучше заметна разница между значениями. Добавим в каждый из графиков горизонтальную прямую для отображения средней оценки за тест по данному предмету. На вертикальной оси отметим жирным шрифтом эту величину.

Упорядочим страны по среднему значению по трём тестам, а так же введём четвёртый график для отображения этого значения аналогично трём описанным выше. Благодаря этому горизонтальные оси графиков согласованы и их расположение один под другим позволяет удобнее сравнивать результаты тестирования по разным предметам в пределах одной страны.

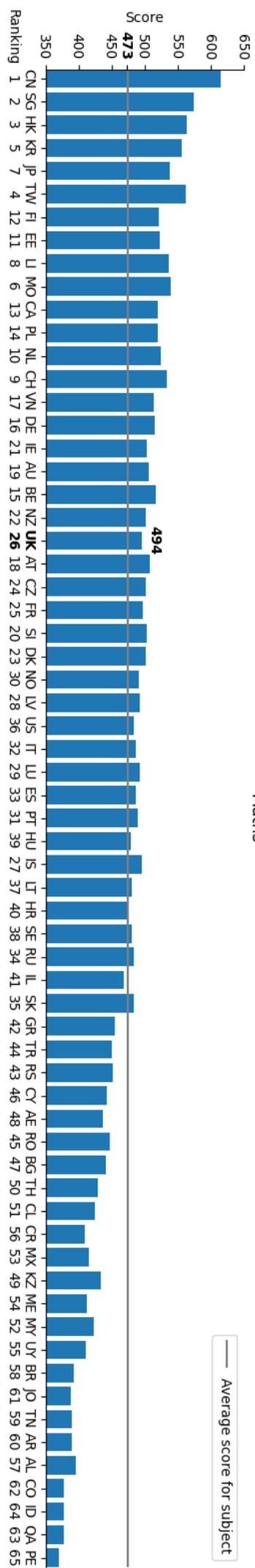
Для облегчения нахождения столбца с результатами Соединённого Королевства, выделим его код и рейтинг жирным шрифтом и укажем над столбцом численную оценку. Так можно более точно определить, насколько количественно Великобритания обошла среднее значение по каждому предмету. В заключение, добавим подписи, поясняющие элементы визуализации.

Несмотря на то, что чтение текста — это когнитивное действие, в случае данной визуализации, подписи и градуировка осей, а так же легенда для пояснения значения горизонтальной линии необходимы. Такое решение оправдано, так как они облегчают восприятие многомерного набора данных большой мощности и при этом хорошо различимы.

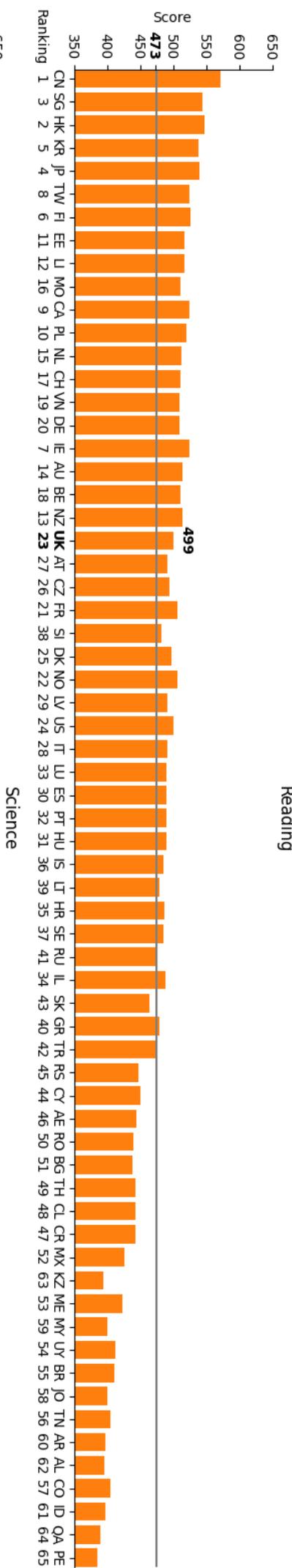
В [таблице 1](#) кратко представлены результаты оценки предложенной модели визуализации.

Pisa 2012 test scores

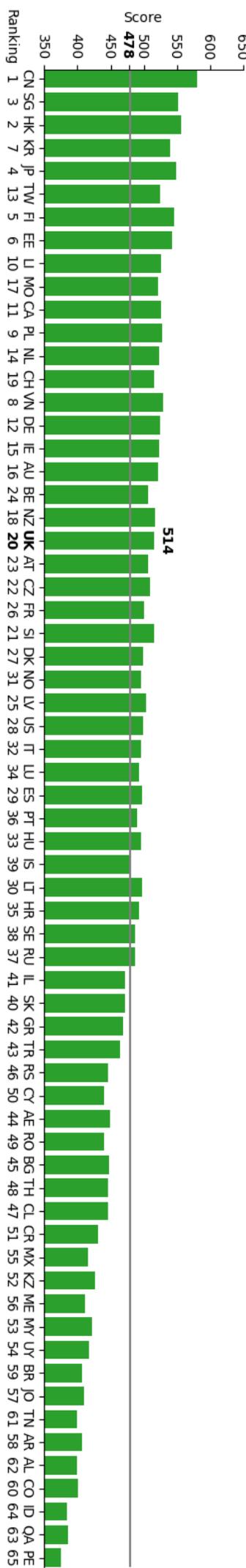
Maths



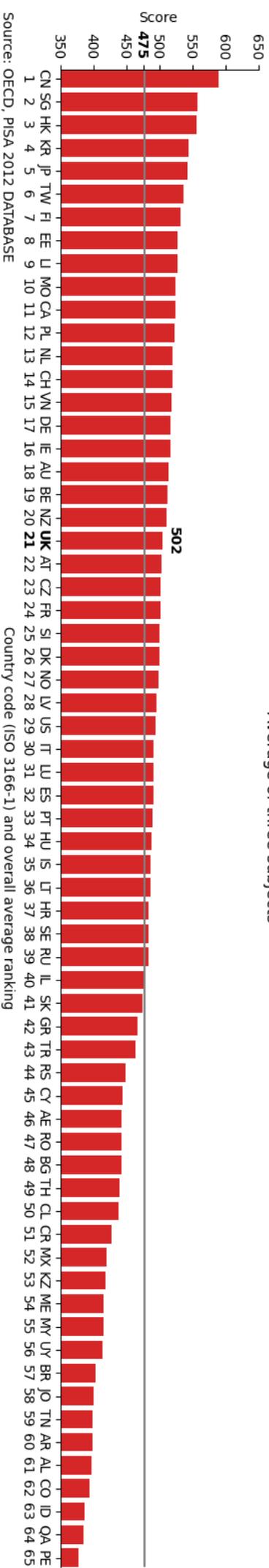
— Average score for subject



Science



Average of three subjects



Source: OECD, PISA 2012 DATABASE

Country code (ISO 3166-1) and overall average ranking

Рисунок 2. Результаты Pisa 2012 - как соотносятся результаты по странам и предметам (исправленный вариант, повернуто на 90°)

Таблица 1. Результаты анализа методики визуализации

Анализ исходной модели визуализации	
Критерий оценки	Вопросы
1. Аналитическая задача	1.1. Анализ соотношения результатов теста по странам и предметам, сравнение показателей Соединённого Королевства с другими странами.
2. Исходные данные	2.1 Источником данных является тестирование Pisa, проводимое Организацией экономического сотрудничества и развития (OECD)
	2.2. Категориальный параметр: предмет по которому проводится тест. Номинальный параметр: страна. Количественная переменная: оценка за прохождения теста
	2.3. Данные отсортированы по убыванию результата за тест по математике, однако, это не соответствует задаче сравнения результатов по всем предметам
3. Способ кодирования	3.1. Множество ломаных с узлами выравненными по трём столбцам в форме кругов со вписанным в их текстом
	3.2. Столбец определяет предмет тестирования, цвет и ломаная страна, численное значение в узле — оценку за соответствующий тест
Оценка корректности выбранной схемы кодирования	
4. Оценка корректности выбранных визуальной схемы кодирования	4.1. В общем, типы данных соответствуют выбранным визуальным переменным. Однако, принцип эффективности нарушается из-за чрезмерно раздутой палитры цветов, что не спасает от неоднозначного её декодирования.
	4.2. Числовые значения приведены верно. 3-мерное представление не используется. Однако, отсутствие данных об среднем значении и отношении порядка относительно него для других оценок нарушает принцип выразительности.
5. Оценка соответствия визуализации и поставленное задачи	5.1. В некоторой степени задача выполняется, однако, отсутствие части исходных данных, а так же перегруженность не дают визуализации полностью выполнить поставленную задачу.
6. Оценка других графических элементов	6.1. Да
	6.2 Не требуется
	6.3 Да
7. Общая	7.1 Лишних графических элементов нет, отсутствуют

эстетическая оценка	некоторые данные, необходимые для анализа, однако, и без того, схема перегружена
	7.2 В целом, визуализация действительно сложна для восприятия, визуальный шум, создаваемый разными цветами и обилием ломаных напоминает лабиринт, поэтому не нравится.