

Искажение информации в визуализации данных

Научный руководитель – Богуславская Олеся Владимировна

Дрыкова Иоланта Сергеевна

Студент (бакалавр)

Сибирский федеральный университет, Институт филологии и языковой коммуникации,
Красноярск, Россия

E-mail: fialki.iolanta@gmail.com

В связи с ускорившимся распространением информации и увеличением её объёмов меняется и медиапотребление: распространяется и утверждается клиповое мышление - способность человека воспринимать мир как ряд почти не взаимосвязанных событий и фактов. Благодаря клиповому мышлению люди молниеносно принимают решения. Также многие из них склонны полагаться на интуицию [1]. Соответственно, клиповая культура основана скорее на эмоциональном восприятии образов, чем на рациональных доводах [2]. Первое впечатление играет ключевую роль: неважно насколько корректно подобраны данные, важен источник и красивое лаконичное оформление. Это даёт основу для манипуляции с данными, что и происходит на Российском и зарубежном медиапространстве.

Целью работы было выявить основные неточности отображения информации в визуализациях, основанных на данных, сознательно или неосознанно вводящих читателей в заблуждение. В исследовании анализируются публикации российских медиа. Эмпирическую базу составляют визуализации Федеральной службы государственной статистики, Дождя, Новой Газеты, Медузы.

Каждый из отобранных материалов анализировался на предмет ошибок в визуализации. В ходе работы было выявлено 8 ошибок в визуализациях данных:

1. Использование 3D графиков: 3D модели визуально увеличивают одни зоны и уменьшают другие, в связи с чем складываются нереалистичные представления о соотношении элементов диаграммы/графика. Подобный пример мы можем встретить в презентации Федеральной службы государственной статистики «Новогодние ожидания россиян». В 3D диаграмме указано количество ценовых котировок, регистрируемых ежемесячно для расчёта ИПЦ. Диаграмма разделена на три части: «непродовольственные товары» расположена в верхней правой части, «продовольственные товары» слева и обе эти части выглядят больше, чем есть на самом деле из-за угла зрения; по этой же причине последняя часть «услуги», гораздо меньше относительно их, чем могла бы выглядеть на аналогичной 2D модели. Можно добавить, что в этой диаграмме также не указаны проценты.

2. Обрезание оси x: этот «приём» может уменьшить или увеличить масштабы. Например, в материале Дождя «10 лет протестов в цифрах. Как изменилась уличная активность и ее поддержка в обществе» мы встречаем линейный график, который начинается с отметки в 5% и заканчивается 35%. Поэтому складывается впечатление, что в 2012 году чуть ли не 100% россиян считали возможным протесты при действительных ~33% (точных цифр авторы нам не дают), а в 2018 около нуля, хотя на самом деле эта отметка означает чуть больше 5%.

3. Визуальная взаимосвязь при её отсутствии в действительности: в материале о протестах Дождя первая визуализация демонстрирует количество протестующих на определённые акции в течение 10 лет. Число отражается в виде полукругов, которые пересекаются, в связи с чем создаётся впечатление, что конкретные лица выходили на акции раз за разом и Дождь знает их количество.

4. Выбор неподходящего способа визуализации: основная проблема - такая визуализация не выполняет своей задачи и либо не говорит читателю ни о чём, либо заставляет его мучиться и искать способ как ей воспользоваться, хотя основная задача визуализации - упрощение. В материале Новой газеты «Я тебя и убью» на круговых диаграммах изображены статьи, по которым судили родителей, дети которых пострадали от домашнего насилия. Каждая из 6 диаграмм разделена на два части «другие дети в семье стали смягчающим обстоятельством» и «у подсудимых нет других детей кроме пострадавших». Сравнение круговых диаграмм не представляется возможным, в данной ситуации лучше бы подошли столбиковые диаграммы, которые легко сопоставить.

5. Переизбыток информации: большое количество данных оттолкнёт и запутает читателя. Пример - визуализация о коронавирусе, которую использовала Новая газета: вокруг цифрографики о количестве заболевших за пределами Китая выстроен огромный ряд стран, но не ясен их порядок. Страны выстроены не в алфавитном порядке и не в порядке количества заболевших, также не сразу становится ясным выбор двух цветов. Возможно, подбор иного способа визуализации, даже список в несколько колонок, исправил бы положение и сделал её понятнее. Также переизбыток данных можно увидеть в уже упомянутом материале Дождя о протестах: на графике «Повод для протеста» целых 14 причин, которые в разные годы пользовались большим «успехом». Однако их настолько много, что у некоторых причин даже не указан процент и их масштаб сложно оценить.

6. Вводящий в заблуждение подбор цветов (оттенки, заливка, штриховка): схожие оттенки цветов усиливают корреляцию, использование контрастных цветов указывает на противоположности. В материале Новой газеты «Если подойдешь, я воткну тебе вилы в горло» в «Судебном бинго» одни строчки белого цвета, другие заштрихованы, одна из них в том числе обведена, но это никак не отражает сути материала. Признаки, по которым выделены строчки не ясны.

7. Отображение неполной картины (использование недостаточного количества данных). Часто при анализе нужно использовать несколько пакетов данных, чтобы установить взаимосвязь событий, так как если этого не сделать, то получится иная картина. Но иногда данные просто пропадают, как, например, произошло в визуализации Медузы: сначала промежуток на оси уравнивается году, затем шести месяцам, а потом и вовсе уменьшается до 90 дней. Естественно график показывает нереальную картину.

8. Отсутствие пояснений. Некоторые диаграммы требуют комментариев: почему именно в этот период произошёл скачок или наоборот упадок чего-либо. Например, в визуализации Медузы об отношении россиян к Украине у читателей может возникнуть вопрос: почему в июне 2016 года 40% россиян выражали симпатию этой стране, а уже в сентябре того же года отношение резко ухудшилось? Не хватает комментария о произошедшем событии.

Таким образом, самые часто встречающиеся ошибки связаны скорее с невнимательностью авторов и их желанием сделать красиво и ёмко - в ущерб отображению действительности - чем с сознательной манипуляцией данными, хотя исключать этот вариант нельзя. Однако российская дата-журналистика развивается и в СМИ уже чаще можно найти качественную работу с данными.

Источники и литература

- 1) Симакова С.И. Визуализация в СМИ: вынужденная необходимость или объективная реальность? // Вестник Самарского университета. История, педагогика, филология. 2017. №1.2

- 2) Симакова С.И. Журналистика данных как объективное требование времени и её влияние на формирование визуальной журналистики // Знак: проблемное поле медиаобразования
- 3) Фролова М.А. История возникновения и развития инфографики // Вестник ПГГПУ. Информационные и компьютерные технологии в образовании