**Как с помощью диаграммы приукрасить действительность? или о факторе лжи Эдварда Тафти**

В предыдущей заметке о живом классике инфографики Эдворде Тафте я рассказал о его [принципе минимизации количества элементов диаграммы](http://baguzin.ru/wp/?p=2071). Сегодня мы познакомимся с введенным Эдвардом Тафтом фактором лжи, а также рассмотрим несколько ошибок, возникающих при отображении и интерпретации диаграмм.

Существуют три вида лжи: ложь, наглая ложь и статистика
*Высказывание приписывается премьер-министру
Великобритании Бенджамину Дизраэли*

Я помню, как однажды, работая в издательстве, хотел продемонстрировать сокращение числа ошибок в рекламных объявлениях. Исходные данные выглядели неплохо (рис. 1), но самое первое значение (год 2002-й, I кв.) явно не вписывалось в обнаруженную тенденцию… и я его просто… отбросил (рис. 2).



Рис. 1. Доля объявлений, вышедших с ошибками



Рис. 2. Доля объявлений, вышедших с ошибками; после коррекции

Такого рода искажения диаграмм – явное жульничество, но встречаются и более тонкие ошибки, возникающие, наверное, непреднамеренно (хотя… кто знает? ☺)

Эдвард Тафти предлагает использовать *фактор лжи* – показатель правдивость отображения информации. «Честная» диаграмма должна отражать исходные данные без искажения, то есть иметь фактор лжи равный единице. Чем больше визуальное искажение, тем выше фактор лжи – отношение размера эффекта, показанного на графике к размеру эффекта, присущего данным:

Фактор лжи = $\frac{размера эффекта, показанного на графике}{размер эффекта, присущего данным}$

Размер эффекта = $\frac{| второе значение – первое значение |}{первое значение}$

В числителе второй формулы используется разность двух значений, взятых по модулю.

Рассмотрим, как работают формулы на примере [инфографики](http://infographer.ru/lie_factor/) (рис. 3). Дальняя от нас линия, пересекающая дорогу, соответствует 1978 году и показывает, что 1 [галлон](http://ru.wikipedia.org/wiki/%D0%93%D0%B0%D0%BB%D0%BB%D0%BE%D0%BD) бензина расходовался за 18 миль, в 1985 году галлона хватало уже на 27,5 миль[[1]](#footnote-1).



Рис. 3. Инфографика, искажающая восприятие

Размер эффекта, присущего данным = (27,5 – 18) / 18 = 0,53. Для измерения эффекта, показанного на графике, надо измерить длину линий (ширину шоссе) соответствующих 78-му и 85-му годам. Это измерение зависит от разрешения экрана или ширины линий на распечатке. У меня вышло 12 и 103 мм. Таким образом, размер эффекта, показанного на графике = (103 – 12) / 12 = 7,6. Эффект лжи = 7,6 / 0,53 = 14,4. Не мало!

Типичные ошибки в построении графиков, приводящие к эффекту лжи, возникают при неправильном выборе масштаба. На рис. 4 приведены данные измерения веса женщины, которая очень хочет похудеть ☺ Глядя на график, можно утверждать, что вес снизился с 68 до 66 кг. Но… если вы обратитесь к Excel-файлу, то увидите, что данные получены путем генерации случайных чисел между 66 и 68! Я просто несколько раз для пересчета формул нажимал F9, чтобы получить задуманный визуальный эффект. Более того, при сохранении файла формулы пересчитались еще раз, так что сохранить эффект в файле мне не удалось. Он запечатлелся только на статичных рисунках.



Рис. 4. «Снижение» веса

«Честный» график приведен на рис. 5. На нем никакого видимого эффекта снижения веса конечно же обнаружить не удастся… Определим фактор лжи рис. 4. Размер эффекта, присущего данным = (68 – 66) / 66 = 0,03. Расстояние наибольшего значения (68 кг) от оси абсцисс = 38 мм, расстояние наименьшего значения (66 кг) от оси абсцисс = 12 мм. Размер эффекта, показанного на графике = (38 – 12) / 12 = 2,2. Эффект лжи = 2,2 / 0,03 = 72



Рис. 5. Колебания веса между 66 и 68 кг

1. 1 Галлон в США = 3,78 л, а 1 миля = 1,6 км; то есть расход 1978 года соответствует 7,6 км/л, 1985 г. – 11,6 км/л [↑](#footnote-ref-1)