

# ТЕХНИЧЕСКАЯ ЗАМЕТКА 01

## ТОНКОСТИ ИСПОЛЬЗОВАНИЯ ЧАСТОТНОГО СРАВНЕНИЯ

Вам кажется, что на некотором форуме из огромного числа участников, активно общаются всего несколько человек.

### Как это проверить и показать наглядно?

Очевидно, нам необходимы данные по статистике постов участников форума. Но какие именно? В данном случае поможет информация о числе пользователей форума, написавших одинаковое количество сообщений (или количество сообщений, попадающее в определенный интервал).

Т.е. мы должны выяснить сколько пользователей написали не более одного сообщения, сколько не более 10 сообщений и т.д. Тогда можно будет произвести *частотное* сравнение — оно покажет на какое количество пользователей приходится максимальное количество сообщений.

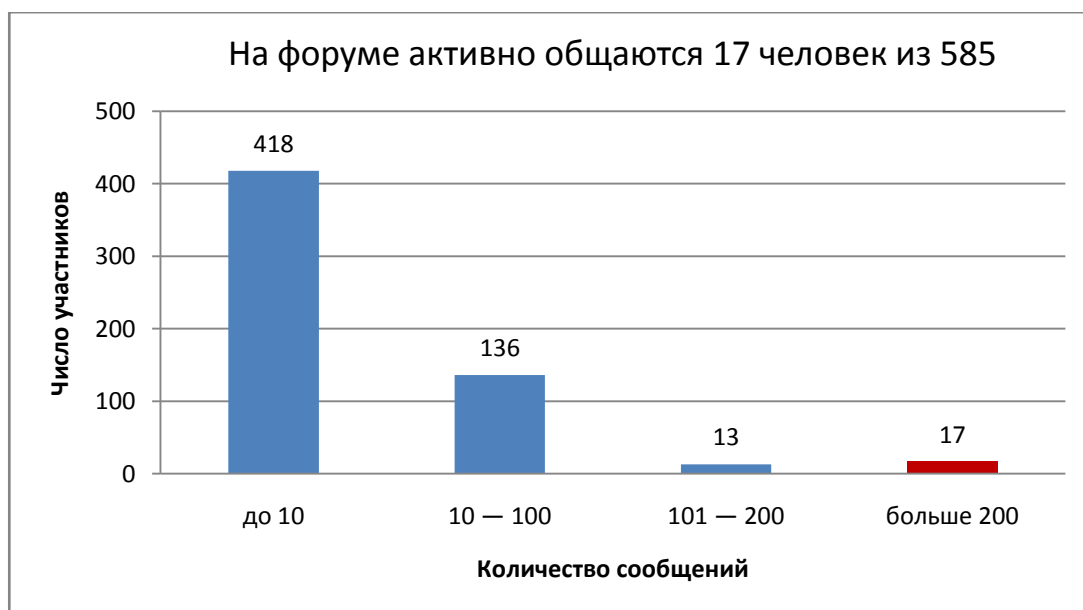
Результатом нашего анализа будет вот такая таблица:

Количество сообщений	Число участников	Всего сообщений	% от общего числа сообщений
до 10	418	45	0,4
10 — 100	136	2853	25,2
101 — 200	14	1481	13,1
больше 200	17	6951	61,4
	Всего участников: 585	Всего сообщений: 11330	

После изучения таблицы, становится понятно, что форум активно используют всего 17 человек.

### Как это показать наглядно?

Принято считать, что для визуального отображения частотного сравнения лучший тип диаграммы — *вертикальная гистограмма*. Для нашего случая она выглядит так:

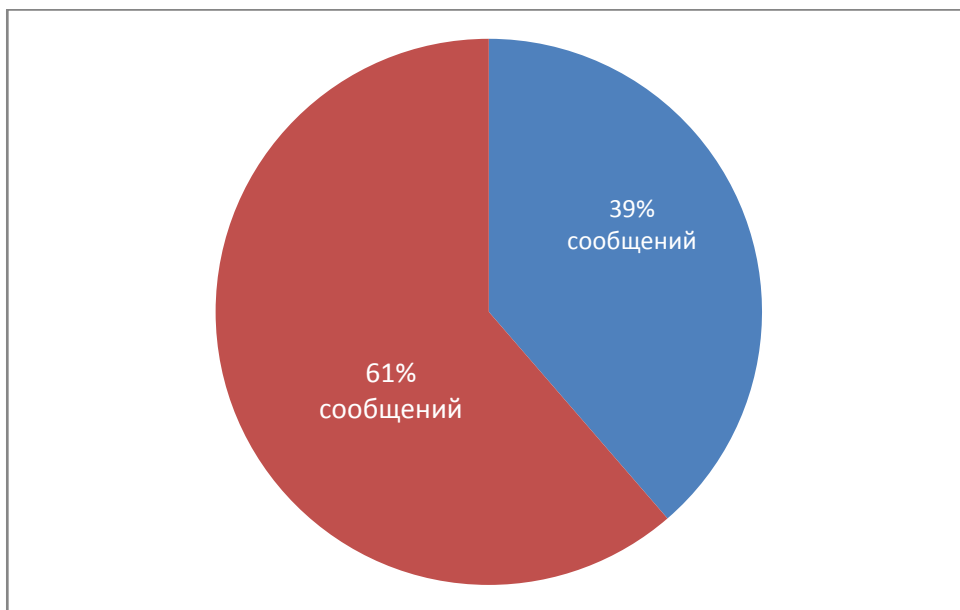


Однако получившаяся диаграмма не наглядна — низкий столбец гистограммы говорит о *наибольшем* числе сообщений. Зрителем это может восприниматься как противоречие. Для

понимания причин такого противоречия зрителю требуется время. Ему необходимо сопоставить подписи по осям, название и относительные размеры столбцов гистограммы.

**Можно ли как-то улучшить визуализацию?** Да можно. Нам следует графически отразить тезис о том, что на 17 человек приходится 61,4% сообщений.

Для этого необходимо построить вспомогательную круговую диаграмму:



Круговая диаграмма покажет соотношение между объемами сообщений пользователей форума.

*Возникает резонный вопрос — почему бы нам не ограничиться только круговой диаграммой?* Ведь из нее видно, что 17 человек генерируют более 61% сообщений? Ответ прост — круговая диаграмма не объясняет почему эти 17 человек наиболее активны. Из нее не видно, что каждый из этих семнадцати является автором более двухсот сообщений. Соответственно, необходимо совместить круговую диаграмму с построенной ранее гистограммой.

В результате получится вот такая совмещенная диаграмма. Для нашего случая она наиболее наглядна, проста для понимания и информативна.

