

Average:

...5. transf. The distribution of the aggregate inequalities (in quantity, quality, intensity, etc.) of a series of things among all the members of the series, so as to equalize them, and ascertain their common or mean quantity, etc., when so treated; the determination or statement of an arithmetical mean; a medial estimate. ...

6. a. The arithmetical mean so obtained; the medium amount, the generally prevailing, or ruling, quantity, rate, or degree; the 'common run.'...

Excerpt from Oxford English Dictionary 2 reference

Description: Numbers for the Average

What is a typical income? What's the height of a typical person? What do I mean by "typical"? There are two different numbers in common use as the "typical" values: These are the "median" and the "mean," both of which are sometimes referred to as the "average" (or typical) value.¹ We have looked at the strategy that makes the median and the mean important. The definition of these two typical values is easy, and generally well known, but for completeness let me now define them and use them.

¹The language for these things is a little imprecise: Some sources will insist that "average" is the generic term, classifying the median and the mean as two different kinds of average. Other sources will use "average" as the specific term (for the mean). And, of course, it does no good to insist on one definition or the other — because you will encounter both usages in reputable sources. I describe this ambiguity to warn you. For myself, and when necessary, I will use average, the longer word, as the generic term and mean, the shorter word as the specific: The generic term average, will include the median, the mean, and in later chapters a host of additional representations for the central value of a set of data. When possible, I will simply dodge this ambiguity by using generic terms such as "typical value" or "central value," and I will attempt to rely on careful language, in context, to make it clear what is being done.

The Median

Assuming that people come with different incomes, with different heights and weights, assuming that anything we may measure offers a certain amount of variety, what is the typical value? The most useful number for this purpose is the median. (The mean is the number that everyone knows as the “average,” but we’ll get to that later.) The median of a set of numbers is the one in the middle: If you put numbers in rank order, from low to high, then *the median is the number in the middle: There as many values above the median as there are below it.* Specifically, when the number of elements is odd, the median is the number in the middle. When the number of elements is even, the median is the average of the two numbers in the middle.

For example, in the set of three numbers 1, 2, 3, the median is 2, the number in the middle: There is one number less than 2 and there is one number greater than 2. In the set of four numbers 1, 2, 3, 4, the median is 2.5, the average of 2 and 3: In this set of four numbers there are two numbers less than 2.5 and there are two numbers greater than 2.5.

The only trick, and it’s a small one, is to find the middle. Assuming that the numbers are in rank order, we count in to a certain “depth”, and that’s the median (or the two numbers whose average is the median). For a set of three numbers in rank order we count in to a “depth” of 2 — there’s the median. For a set of four numbers we count in to a depth of 2.5, and take the average of the second number and the third.

The formula for the depth is simple: Where “n” is the number of things in the set, the arithmetic is to compute the number $(n+1)/2$. If the result is a whole number, it identifies the median: If the result is a fraction, then it identifies two numbers whose average is the median.

Working it out, with three things

$n = 3$ implies the arithmetic $(n+1)/2 = 2$ So, the depth of the median is 2 and, in rank order, the median is the second number.

With four things

$n = 4$ implies the arithmetic $(n+1)/2 = 2.5$ So, the depth of the median is 2.5 and, using the rank order, the median is the mean of the second number and the third.

Going back to breakfast, there are 30 breakfast cereals listed in Figure __. So

$n = 30$ implies $(n+1)/2 = 15.5$ The depth of the median is 15.5 and therefore, in rank order, the median is the average of the fifteenth number and the sixteenth. The median number of grams of protein is _____ .

Among the four rice cereals

$n = 4$ implies $(n+1)/2 = 2.5$ The depth of the median is 2.5. In rank order, the median is the mean of the second number and the third. The median grams of protein among the rice cereals is _____ .

Among the ___ corn cereals

$n = ?$ implies $(n+1)/2 = ??$ In rank order, the median is the average of the _____ number and the _____. The median grams of protein among the rice cereals is _____ .

Looking back at those cereals and computing the median protein content you see one important property of the median: For the whole group of cereals, the one lone very high value has little effect on the median; the median lies pretty close to the middle as you see it. Similarly, for the group of corn cereals, the one odd case, has little effect on the median as a typical value for the group.

There are times when it is possible to be overly precise, and such is the case with the median. Let me use a somewhat more labored definition of the median and then note that a precise definition is not

always a specific definition. More precisely, the median is a number such that 50% or more of the data are less than or equal to the median while, at the same time, 50% or more of the data are greater than or equal to the median.

That is more precise, but the nasty fact is that it does not necessarily come up with a single number: In the set of four values, 1, 2, 3, and 4, any number between two and three satisfies the definition of the median. Now, in truth, that's reality: Any number between two and three is equally acceptable as the median of these four numbers. However, convention and simple expectation, "What is the median?" seem to be comfortable with a single number, so we improvise to pick one. Unfortunately, even here there is a choice of conventions, none of which is absolutely defensible (and none of which is unreasonable) because, in truth, the median is not always a single number. However, to see the kind of thing you will find, in use, consider the question: What is the median size of a household in the United States? From the U. S. Census Bureau, Current Population Survey, here are the numbers for 1991.

{Need two examples: The first one should have the median fall between two values — that's not what I've got here. Here I have the median category interpreted as an interval, requiring interpolation in the interval.}

Size of Household	Number (in millions) 1991	Percentage Distribution 1991	Cumulative Percentage Distribution (Low to High)	Cumulative Percentage Distribution (High to Low)
1 Person	23.6	25	25	100
2 People	30.2	32	57	75
3 People	16.1	17	74	43
4 People	14.6	15	89	26
5 People	6.2	7	96	11
6 People	2.2	2	98	4
7 or more	1.5	2	100	2

Table ____
Households, by Size of Household: 1991

From the *Statistical Abstract of the United States, 1992*, p. 47, No. 56. On the left, data as presented in the source. On the right, work sheet of cumulative percentages from low to high and from high to low.

The numbers on the left indicate the size of the household and the number of millions of households of that size: There were 23.6 million one-person households, which accounted for twenty-five percent of all households. On the right, you see my work sheet the numbers I need in order to talk about the median.

O.K., now: The median is a number such that 50% or more of the data are less than or equal to the median while, at the same time, 50% or more of the data are greater than or equal to the median.

The cumulative percentages show that the size of 57 percent of households is less than or equal to two while, at the same time, the size

of 75 percent of households is greater than or equal to two. So the median size of a household is two.

But if you want to worry about “something better” for this median, you can note it takes almost all of the 30.2 million households to add up to fifty percent (counting up from low to high). That has a feel to it, suggesting that the number is “almost” 3. Looking at it the other way, from high to low, it takes only a few of the 30.2 million households to add up to fifty percent. So, I may choose to interpolate a fraction that will give me a median that is greater than 2, and close to 3. Recognizing that we are dealing with fictions here, useful fictions, but fictions nonetheless: Let me act as if household sizes came in intervals, up to 1.5, 1.5 to 2.5, 2.5 to 3.5, 3.5 to 4.5, and so forth — Acting as if there were some continuous tendency toward small or large households that works itself out in the world as a simple integer, 1, 2, 3, or more.

So now, in this useful fiction, twenty-five percent of households have 1.5 or fewer members. And now, to get up to fifty percent, I need to take most of the next interval: I need another twenty-five percent of the population and the interval contains thirty-two percent of the population, so I need almost all of the population in this group (in order to add up to fifty percent).

So, what do I need? I need $25/32$ of the people in the this group. So I will go $25/32$ of the way from 1.5 to 2.5 — and call that the median:

Interpolated median (work sheet):

“Interval” Below the Median: 0 to 1.5

End of Interval: 1.5

Percent of Population in Interval Below the Median: 25%

Percent of Population Needed to Add Up to 50%: 25%

“Interval” Including the Median: 1.5 to 2.5

Width of Interval: 1

Percent of Population in Interval Including the Median: 32%

Fraction of this Population needed to Add Up to 50%: $25/32$

Interpolated Median:

End of lower Interval plus Interpolated Fraction of Width of Next Interval

= 1.5 plus $(25/32)$ times 1

= $1.5 + (.78)*1$

= 2.28

For the sake of increasing my own security with this answer, I check from the other direction: Do I get the same answer?

Interpolated median (work sheet) checking:

“Interval” Above the Median: 2.5 to 7 or more

End of Interval: 2.5

Percent of Population in Interval Above the Median: 43%

Percent of Population Needed to Add Up to 50%: 7%

“Interval” Including the Median: 1.5 to 2.5

Width of Interval: 1

Percent of Population in Interval Including the Median: 32%

Fraction of this Population needed to Add Up to 50%: $7/32$

Interpolated Median:

End of Higher Interval Minus Interpolated Fraction of Width of Next Interval

= 2.5 minus $(7/32)$ times 1

= $2.5 + (.22)*1$

= 2.28 check!

Result: The interpolated median size of household was 2.3 people.

And now, finally, a word of advice: Don't attempt to memorize these formulas — that's the hard way. Instead, think the problem through from first principles, asking yourself what it is that you are trying to accomplish. And then, just to be careful, check by doing it another way: Here I reasoned from low to high, got an answer, and then did it another way to check: Same answer. If there is any doubt about your method, because there are other acceptable methods, then briefly write it out so your reader will know what you've done. You can't just appeal to authority, saying: "Here it is! This is right." You can't do that because there are different authorities you could appeal to. Each one makes a slightly different interpretation of the problem and it's up to you to choose among them. So, you might as well take the responsibility from the beginning: Think clearly. It's up to you.

Exercise:

Return to the data for the Dow Industrials, showing change as a percentage of the earlier price. Use a stem and leaf diagram to place the industrials in rank order by amount of change and then use that rank ordering to determine the median. Report the “n” (how many industrials are there?) Report the location for the median. And report the median.

As above, apply the same technique to the rates of infant mortality among states of the United States. Write a brief report of your results.

Exercise: Table _ includes a collection of UNESCO data describing nations of the world, already rank ordered on each of the variables: For gross national product, what is the “n”? For population, what is the “n”. Noting that these “n’s” are unequal, prepare a very short report stating the median gross national product for nations. Begin at the beginning (although that may not be the first line in your report): Do you trust the data? Are these data complete? What’s missing? Do you trust your report of the median?

	Country	Expectation of Life at Birth (years) 1991	Infant Mortality Rate 1991 (Number of deaths of children under 1 year of age per 1,000 live births in a calendar year.)	Health Expenditures 1990: per capita on basis of GDP purchasing power parities		Health Expenditures 1990: % of gross Domestic product (preliminary estimates)	Health Expenditures - Public Health as % of gross domestic product	Health Expenditures - Public Health as % of total health expenditures
1	Australia	77.0	7.9	1,151		7.5	5.2	69.6
2	Austria	77.3	5.4	1,192		8.4	5.6	66.5
3	Belgium	77.1	5.5	1,087		7.4	6.1	82.5
4	Canada	77.5	7.2	1,795		9.0	6.7	74.1
5	Denmark	75.9	6.1	963		6.2	5.2	84.2
6	Finland	75.8	6.0	1,156		7.4	6.2	83.3
7	France	77.8	6.1	1,379		8.9	6.6	74.2
8	Germany	75.8	7.1	1,287		8.1	5.9	72.7
9	Greece	77.7	10.0	406		5.3	4.0	76.0
10	Ireland	75.5	6.2	693		7.1	5.8	82.0
11	Italy	78.1	6.0	1,138		7.7	5.9	75.9
12	Japan	79.2	4.4	1,145		6.5	4.6	71.4
13	Netherlands	77.8	6.9	1,182		8.0	5.8	72.6
14	New Zealand	75.5	9.5	853		7.2	5.9	81.7
15	Norway	77.1	6.7	1,281		7.2	6.9	95.7
16	Portugal	74.7	13.3	529		6.7	4.1	61.7
17	Spain	78.3	6.2	730		6.6	5.2	78.4
18	Sweden	77.8	5.9	1,421		8.7	7.8	89.3
19	Switzerland	79.1	4.7	1,436		7.7	5.2	68.1
20	Turkey	69.8	54.3	197		4.0	1.4	35.6
21	United Kingdom	76.5	7.2	932		6.2	5.2	84.5
22	USA	75.7	10.3	2,566		12.4	5.2	42.4

	Sources:							
		Statistic al Abstract of the United States 1991	Source: Stat Abstack of The United States, 1992, Table 1361, p. 824. "Soviet Union" as former S U "without" Independe nt Rep					
		Table 1361 p. 824						
		Original : U.S. Bureau of the Census, World Populati on Profile: 1991.		Original: U.S. Bureau of the Census, World Population Profile: 1991.		Original: U.S. Bureau of the Census, World Populati on Profile: 1991.	Original: : U.S. Bureau of the Census, World Populati on Profile: 1991.	Original: U.S. Bureau of the Census, World Populatio n Profile: 1991.
				Table 1368 p. 830		Table 1368 p. 829	Table 1368 p. 831	Table 1368 p. 832

				Originl Source Organizati on of Economic Cooperatio n and Developme nt, Paris, France, OECD Health Data 1991, and OECD Health Systems: Factgs and Trends, 1993		Originl Source Organiza tion of Economic Cooperat ion and Develop ment, Paris, France, OECD Health Data 1991, and OECD Health Systems: Factgs and Trends, 1992	Originl Source Organiz ation of Economic Coopera tion and Develop ment, Paris, France, OECD Health Data 1991, and OECD Health Systems: Factgs and Trends, 1994	Originl Source Organizat ion of Economic Cooperati on and Developm ent, Paris, France, OECD Health Data 1991, and OECD Health Systems: Factgs and Trends, 1995
--	--	--	--	---	--	---	---	---

0 Country	Expectation of Life at Birth (years) 1991	51 Syria	69.4	111 Nigeria	48.9
1 Japan	79.2	52 Korea North	69.0	112 Mozambique	47.4
2 Switzerland	79.1	53 Thailand	68.5	113 Central African Republic	47.1
3 Spain	78.3	54 Lebanon	68.4	114 Mali	46.1
4 Italy	78.1	55 Libya	68.1	115 Sierra Leone	44.8
5 France	77.8	56 Malaysia	68.1	116 Angola	44.3
6 Netherlands	77.8	57 Dominican Republic	67.2	117 Afghanistan	43.5
7 Sweden	77.8	58 Iraq	67.0	118 Guinea	42.8
8 Greece	77.7	59 Algeria	66.7	119 Chad	39.8
9 Canada	77.5	60 Ecuador	66.2	120 Andorra	
10 Austria	77.3	61 Honduras	66.0	121 Antigua and Barbuda	
11 Belgium	77.1	62 Saudi Arabia	65.9	122 Armenia	
12 Norway	77.1	63 El Salvador	65.5	123 Aruba	
13 Australia	77.0	64 Brazil	65.2	124 Bahamas	
14 Israel	77.0	65 Mongolia	65.1	125 Bahrain	
15 Costa Rica	76.8	66 Vietnam	64.7	126 Barbados	
16 United Kingdom	76.5	67 Morocco	64.6	127 Belize	
17 Denmark	75.9	68 Philippines	64.6	128 Bhutan	
18 Finland	75.8	69 Iran	64.5	129 Bosnia Herzegovina	
19 Germany	75.8	70 Peru	64.3	130 Botswana	
20 USA	75.7	71 South Africa	64.2	131 Brunei	
21 Cuba	75.6	72 Guatemala	63.2	132 Byelarus	
22 Ireland	75.5	73 Nicaragua	62.5	133 Cambodia-cf. Campuchia	
23 New Zealand	75.5	74 Zimbabwe	61.7	134 Cape Verde	
24 Albania	75.1	75 Bolivia	61.5	135 Comoros	
25 Singapore	74.8	76 Kenya	61.5	136 Croatia	
26 Portugal	74.7	77 Indonesia	61.0	137 Cyprus	
27 Venezuela	74.2	78 Egypt	60.8	138 Djibouti	
28 Panama	74.0	79 India	57.2	139 Dominica	
29 Jamaica	73.6	80 Oman	56.6	140 Equatorial Guinea	
30 Kuwait	73.6	81 Pakistan	56.6	141 Estonia	
31 Chile	73.4	82 Liberia	56.4	142 Fiji	
32 Yugoslavia	73.0	83 Zambia	56.4	143 Gabon	
33 Czechoslovakia	72.9	84 Somalia	55.9	144 Gambia	
34 Poland	72.9	85 Togo	55.6	145 Georgia	
35 Bulgaria	72.7	86 Papua New Guinea	55.4	146 Germany East	
36 Uruguay	72.6	87 Senegal	55.1	147 Germany West	
37 Mexico	72.2	88 Burma	54.9	148 Grenada	
38 Romania	71.9	89 Ghana	54.6	149 Guinea-Bissau	
39 Tunisia	71.9	90 Ivory Coast / Cote d'Ivoire	54.3	150 Guyana	
40 Hungary	71.6	91 Congo	54.2	151 Hong Kong	
41 Jordan	71.2	92 Zaire	53.9	152 Iceland	
42 Sri Lanka	71.1	93 Haiti	53.6	153 Kiribati	
43 Colombia	71.0	94 Bangladesh	53.0	154 Kyrgystan	
44 Argentina	70.9	95 Sudan	53.0	155 Lesotho	
45 UAR United Arab Emirates	70.9	96 Madagascar	52.6	156 Liechtenstein	
46 China / People's Republic of China / Mainland	70.0	97 Burkina	52.5	157 Lithuania	
47 Soviet Union	69.8	98 Rwanda	52.5	158 Latvia	
48 Turkey	69.8	99 Burundi	52.4	159 Luxembourg	
49 Korea South	69.7	100 Tanzania	52.0	160 Maldives	
50 Paraguay	69.7	101 Ethiopia	51.3	161 Malta	
		102 Cameroon	51.0	162 Mauritania	
		103 Niger	51.0	163 Mauritius	
		104 Uganda	51.0	164 Moldova	
		105 Nepal	50.6	165 Namibia	
		106 Benin	50.5	166 Puerto Rico	
		107 Laos	50.2	167 Qatar	
		108 Yemen	49.9	168 Russia	
		109 Kampuchia / Cambodia	49.3		
		110 Malawi	49.2		

- 169 Saint Kits and Nevis
- 170 Saint Vincent and the Grenadines
- 171 San Marino
- 172 Santa Lucia
- 173 Sao Tome and Principe
- 174 Serbia
- 175 Seychelles
- 176 Suriname
- 177 Swaziland
- 178 Taiwan / Republic of China
- 179 Tajikistan
- 180 Trinidad and Tobago
- 181 Turkmenistan
- 182 Tuvalu
- 183 Ukraine
- 184 Upper Volta
- 185 USSR
- 186 Uzbekistan
- 187 Venuatu
- 188 Vietnam North
- 189 Vietnam South
- 190 Western Somoa
- 191 Yemen (Aden)
- 192 Yemen (Sana)

Statistical Abstract of the United States 1991 Table 1361 p. 824 Original: U.S. Bureau of the Census

'World Population Profile: 1991.

Mean:

... 8. a. Math. [= F. moyenne, ellipt. for quantité moyenne.] The term (or, in plural, the terms) intermediate between the first and last terms (called the extremes) of a progression of any kind (distinctively, arithmetic(al), geometric(al), harmonic(al) mean). Also, in a wider sense, a quantity so related to a set of n quantities that the result of operating with it in a certain manner n times is the same as that of operating similarly with each of the set. In this sense the arithmetic(al) mean (commonly called simply the mean) of a set of n quantities is the quotient of their sum divided by n ; the geometric(al) mean is the n th root of their product. ...

(Excerpt from Oxford English Dictionary 2 -- CD -- exact reference??)

The Mean

What is a typical income? What is the size of a typical family? The second number commonly used to answer that question is the mean. Assuming that people come with different incomes, with different heights and weights, the mean is the sum of a set of numbers divided by number of elements in the set.

It may be that the simplicity of the mean and the ease of computation (when you are working without a computer) is what makes it the most commonly used measure of the center: What is the mean of 1 and 2? Add them up, the sum is 3, and divide by 2. The mean of 1 and 2 is 1.5. That's it. There is nothing more to the arithmetic: Compute the sum of the numbers and then divide the sum into equal parts. Going back to the data reporting protein content of 30 breakfast cereals, listed in Figure __: The sum of the 30 numbers reporting protein content is 89.1 grams of protein. Dividing the sum into 30 equal parts, the average is 89.1 grams of protein divided by 30: The mean is 2.97 grams of protein per serving.

Cereal Number	Grams of Protein
1	4.3
2	3.1
3	2.9
4	2.8
5	1.8
6	2.2
7	3.4
8	1.5
9	2.1
10	5.1
11	2.8
12	2.7
13	10.2
14	4.8
15	2
16	2.6
17	3
18	2.2
19	4.5
20	2.1
21	1.5
22	2.1
23	1.6
24	0.8
25	3.2
26	4.4
27	2.8
28	1.6
29	2.2
30	2.8
Sum:	89.1

Mean of 30:	2.97
------------------------	-------------

You can think of the mean as the answer to a “What if?” question: *What if* the total protein content were divided equally among cereals. *What if* the total income of families were equally divided among families? *What if* the total number of children were divided equally among families? *If* the total were divided into equal parts, the result would be the mean.

Exercises --- as for the median, substituting the mean.

Which Average, The Median or the Mean?

You now have two answers to the question, “What’s typical?, What’s the average?”. You have the median and the mean. And you may suspect, correctly, that there are more answers — which raises the question, “Which one?” When do you use the median? When do you use the mean?

In most cases there is no right answer, no right answer in the sense that you can look up the answer in the back of a book and be correct. There is no clear answer because, usually, the question itself is not clear: “What’s typical? What’s average?” What the person asking the question wants is a general description: How much money do these people have? How large is an average family? Does the average breakfast cereal provide much protein? Is the rate of infant mortality high or low? When you think about what you want from these data, that you want a description, the answer becomes clear: Either number will answer the request for an average that describes the data, as long as you attend to two concerns that will make the description useful: First, be specific about what you’ve done. Say “the median is 2” or “the mean is 3”. (Do not say “the average is 3” or “the average is 2” — that’s not specific.) Second, be consistent: If you or the person reading your work is going to compare the average to some other average, then use the same average. I ask you the average income of families in order to know whether it is higher or lower than it was ten years ago. What average did we use ten years ago? Use the same average now. I ask you the average protein content of rice-based cereals in order to compare rice-based cereals to wheat-based cereals. Use the same average for both.

Sometimes there is an easy answer to that question: What is the average life expectancy of people diagnosed to have a certain disease? Almost certainly, for this case, the correct average is the median because you would have to wait too long to get an answer for the mean: If the disease is detected in 100 people, then when the 51st person has died you will know the median. Happily for your people, but

unhappily for your research, you may have to wait much longer to collect all the data you need for the mean.

Sometimes the answer to the question is easy because the median and the average are almost the same. When you find that the stem and leaf diagram gives you a symmetrical result — with the low side of the distribution looking like a mirror image of the high side of the distribution, the median and the average will be almost the same. For example, here is the stem and leaf for infant mortality rates, by state, in the United States. Here, the median, at 9.7 deaths per thousand live births, and the average, at 9.75 deaths per thousand live births, are visually indistinguishable — implying that the difference between the median and the average is small compared to the overall range.

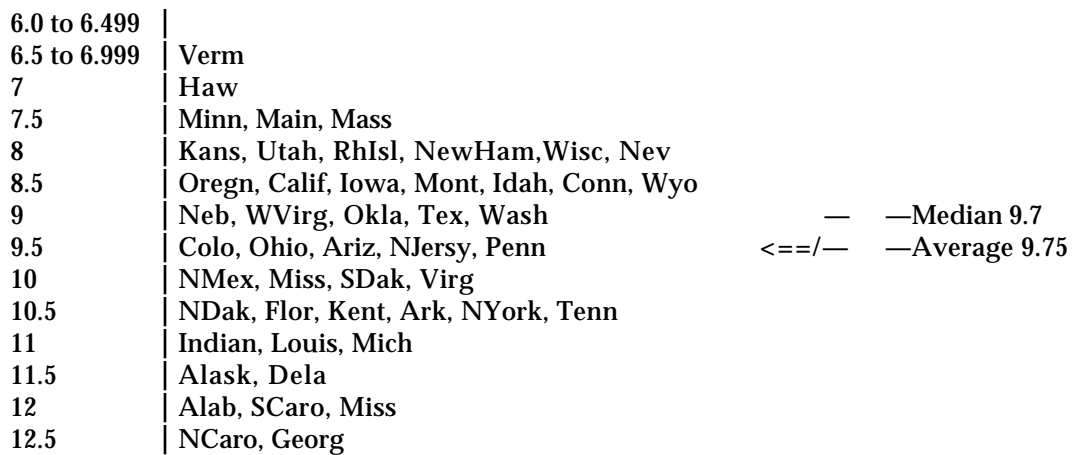


Figure _
Stem and Leaf of 1988 Infant Mortality Rates, Showing the Median and the Average.

Numbers shown are in deaths per 1,000 Live Births, see Figure __ of Chapter __.

(Draw the stuff on the right as two arrows pointing from separate labels into one stem.)

But most of the time the best answer is to begin at the beginning: At this stage of your data analysis you are creating a description, so begin with a picture of the thing that is going to be described by the average: Show the stem and leaf (or the histogram, section __). Then compute both averages, both the median and the mean, because the difference between the two, when they are different, improves a better description of the data, better than either one alone.

The result for your “rules” is that I have changed the initial question. The question was “should I use the median or the mean”. I’ve answered by saying, “use both” — because it serves your real purpose, which is to describe the data. And now the new question is “what do I learn by comparing the median to the mean?”

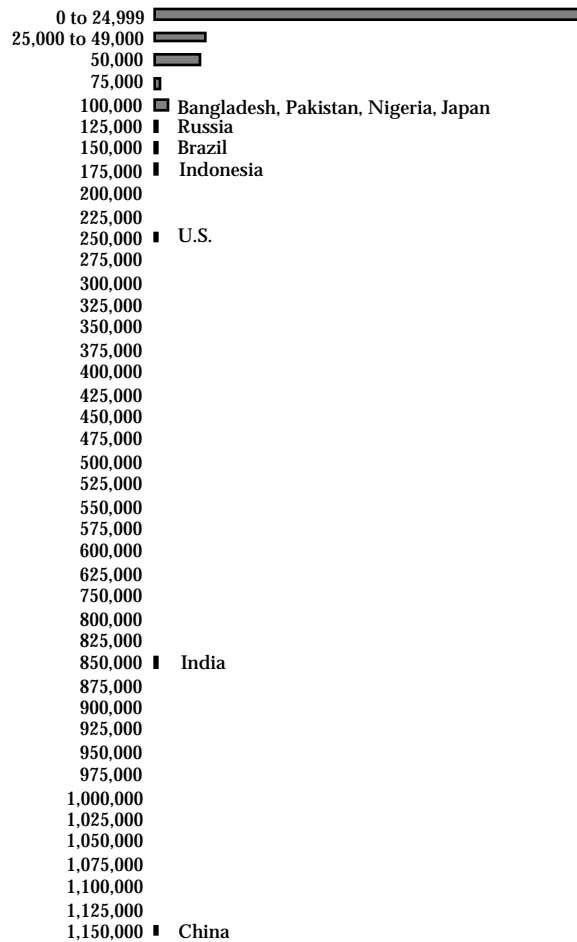
The median is, by definition, the middle value. Therefore, one special feature of the median is that it “ignores” the actual numbers that are above it or below it, as long as they preserve the order so that the same number is in the middle. Thus, for the thirty, breakfast cereals, the median is 2.75 grams of protein, the number in the middle. And if, by chance, by error, by whatever accident, the number for Gerber’s had shown up as 5.2 instead of 10.2? The number in the middle is still 2.75. And if the number for Gerber’s had been 20.2 instead of 5.2? The number in the middle is still 2.75. And if the number for Gerber’s had been 100 times larger than it already is? The number in the middle is still 2.75. That is one property of the median: The median is “robust”, with respect to changes in the extreme values.

Rank (Low to High)	Cereal Number (Referring to original alphabetical order)	Grams of Protein (Rank Ordered)	Grams of Protein (Altering the Extreme Value)	Grams of Protein (Altering the Extreme Value)	Grams of Protein (Altering the Extreme Value)
1	24	0.8			
2	8	1.5			
3	21	1.5			
4	23	1.6			
5	28	1.6			
6	5	1.8			
7	15	2.0			
8	9	2.1			
9	20	2.1			
10	22	2.1			
11	6	2.2			
12	18	2.2			
13	29	2.2			
14	16	2.6			
15	12	2.7			
16	4	2.8			
17	11	2.8			
18	27	2.8			
19	30	2.8			
20	3	2.9			
21	17	3.0			
22	2	3.1			
23	25	3.2			
24	7	3.4			
25	1	4.3			
26	26	4.4			
27	19	4.5			
28	14	4.8			
29	10	5.1			
30	13	10.2	5.1	20.4	1,020.0
@ rank 15.5	Median of 30	2.75	2.75	2.75	2.75
	Mean of 30	2.97	2.80	3.31	36.63

And thus, when the median is *different* from the mean that's a clue: The median "ignores" the extreme value; the mean uses all values. When they are different, look for extreme values on one side (either above or below) the average. In the case of the breakfast cereals the fact that the median is different from the mean, 2.75 versus

2.97, is a numerical match to what we saw in the stem and leaf — that there is one extreme value, much larger than all the others.

Showing exactly the same properties from a different perspective, consider the data for the gross national products of 150? states.



Rank by Population (Low to High)		Country	Total Population 1990 in thousands (From the Statistical Abstract of the United States, 1991 Table 1359)		
1	Tuvalu	9	48	Kuwait	2,124
2	San Marino	23	49	Mongolia	2,187
3	Liechtenstein	28	50	Congo	2,242
4	Saint Kitts and Nevis	40	51	UAR United Arab Emirates	2,254
5	Andorra	52	52	Panama	2,425
6	Antigua and Barbuda	64	53	Jamaica	2,469
7	Aruba	64	54	Liberia	2,640
8	Seychelles	68	55	Latvia	2,695
9	Kiribati	70	56	Singapore	2,721
10	Grenada	84	57	Central African Republic	2,877
11	Dominica	85	58	Costa Rica	3,033
12	Saint Vincent and the Grenadines	113	59	Uruguay	3,102
13	Sao Tome and Principe	125	60	Albania	3,273
14	Santa Lucia	150	61	Jordan	3,273
15	Vanuatu	165	62	New Zealand	3,296
16	Western Samoa	186	63	Lebanon	3,339
17	Maldives	218	64	Armenia	3,357
18	Belize	220	65	Ireland	3,500
19	Bahamas	249	66	Nicaragua	3,602
20	Barbados	254	67	Turkmenistan	3,658
21	Iceland	257	68	Togo	3,674
22	Djibouti	337	69	Lithuania	3,726
23	Malta	353	70	Papua New Guinea	3,823
24	Equatorial Guinea	369	71	Laos	4,024
25	Brunei	372	72	Sierra Leone	4,166
26	Cape Verde	375	73	Libya	4,223
27	Luxembourg	384	74	Norway	4,253
28	Suriname	397	75	Moldova	4,393
29	Comoros	460	76	Kyrgyzstan	4,394
30	Qatar	491	77	Israel	4,436
31	Bahrain	520	78	Bosnia Herzegovina	4,517
32	Cyprus	702	79	Paraguay	4,660
33	Fiji	738	80	Benin	4,674
34	Guyana	753	81	Croatia	4,686
35	Swaziland	837	82	Honduras	4,804
36	Gambia	848	83	Finland	4,977
37	Guinea-Bissau	999	84	Chad	5,017
38	Gabon	1,068	85	Denmark	5,131
39	Mauritius	1,072	86	El Salvador	5,310
40	Botswana	1,224	87	Tajikistan	5,342
41	Trinidad and Tobago	1,271	88	Georgia	5,479
42	Namibia	1,453	89	Burundi	5,646
43	Oman	1,481	90	Haiti	6,142
44	Bhutan	1,566	91	Somalia	6,654
45	Estonia	1,584	92	Switzerland	6,742
46	Lesotho	1,755	93	Bolivia	6,989
47	Mauritania	1,935			

94	Kampuchia / Cambodia	6,991
95	Dominican Republic	7,241
96	Guinea	7,269
97	Rwanda	7,609
98	Austria	7,644
99	Senegal	7,714
100	Niger	7,879
101	Tunisia	8,104
102	Mali	8,142
103	Zambia	8,154
104	Angola	8,449
105	Sweden	8,526
106	Bulgaria	8,934
107	Guatemala	9,038
108	Burkina	9,078
109	Malawi	9,197
110	Yemen	9,746
111	Serbia	9,883
112	Belgium	9,909
113	Greece	10,028
114	Byelarus	10,257
115	Portugal	10,354
116	Zimbabwe	10,394
117	Ecuador	10,507
118	Hungary	10,569
119	Cuba	10,620
120	Cameroon	11,092
121	Madagascar	11,801
122	Ivory Coast / Cote d'Ivoire	12,478
123	Syria	12,483
124	Chile	13,083
125	Mozambique	14,539
126	Netherlands	14,936
127	Ghana	15,130
128	Afghanistan	15,564
129	Czechoslovakia	15,683
130	Australia	17,037
131	Saudi Arabia	17,116
132	Sri Lanka	17,198
133	Malaysia	17,556
134	Uganda	18,016
135	Iraq	18,782
136	Nepal	19,146
137	Venezuala	19,698
138	Taiwan / Republic of China	20,435
139	Uzbekistan	20,569
140	Korea North	21,412
141	Peru	21,906
142	Romania	23,273
143	Kenya	24,342

144	Algeria	25,337
145	Morocco	25,630
146	Tanzania	25,971
147	Sudan	26,245
148	Canada	26,538
149	ARgentina	32,291
150	Colombia	33,076
151	Zaire	36,613
152	Poland	37,777
153	SPAN	39,269
154	South Africa	39,539
155	Burma	41,277
156	Korea South	42,792
157	Ethiopia	51,407
158	Ukraine	51,711
159	Egypt	53,212
160	Thailand	56,002
161	France	56,358
162	Iran	57,003
163	Turkey	57,285
164	United Kingdom	57,366
165	Italy	57,664
166	Germany West	63,232
167	Philippines	64,404
168	Vietnam	66,171
169	Germany	79,123
170	Mexico	88,010
171	Bangladesh	113,930
172	Pakistan	114,649
173	Nigeria	118,819
174	Japan	123,567
175	Russia	148,254
176	Brazil	152,505
177	Indonesia	190,136
178	USA	250,410
179	India	852,667
180	China / People's Republic of China / Mainland	1,133,683
	Median	6,398
	Mean	29,707

For these data the median country has 6 million people (6,398,000), like Haiti and Somalia. The mean is 29,707,000, like Argentina. The difference between this median of 6 million and this mean of 30 million is very large. The descriptive solution is to provide both, noting that the difference between the two corresponds to the picture in which we see that there is a "tail" at one end of the distribution. Technically, this is called a "skewed" distribution. And the difference between the median and the mean corresponds to the visual evidence that the population of the world is concentrated in a small number of heavily populated countries

Exercise:

--- Review of previous data, using the stem and leaf, as well as both the median and the mean. Note the differences between the median and the mean. Note the relation between the difference (or the absence of a difference) and the shape of the stem and leaf. Write it up, clearly.
