Plain Numbers

A Presentation by Beatrice Baker

for the Editors' Association of Canada

2007 Annual Conference

Contents

1 About Plain Numbers

> "Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write."
>
> ——*H. G. Wells*

Like plain language, plain numbers is reader-focused. Numerical information should be accurate, easy to understand, and memorable.

Some common problems

People often don't know

- what information to include: uncertain of which numbers are important to their presentation, they compensate by throwing them all in.
- which tools to use: text, table, or chart?
- how to choose between tables and charts: decisions to use a table or a pie chart or a bar graph may be made on which one looks "more impressive" rather than on its ability to convey information.
- how to use software appropriately: just because software comes with multiple bells and whistles doesn't mean that it's a good idea to use them all.
- how to label information: titles, captions, column and row stubs can mislead the reader if they don't accurately reflect the data.
- how to link the visual presentation with the text: they may use the text to parrot every figure shown in a table, or, conversely, limit themselves to a succession of "See Table X".

Questions to ask

- What is the purpose of the document?
- Who is your audience, and why do they want the information?
- What do they already know about the subject? How much more information do they need?
- Are they number literate? That is, do they understand how to read tables and charts? Interpret percentages?

Choosing the right tools

Your choice of tools (text, table and/or chart) will depend on

- how much numerical data you have to present,

- how much time your reader has to absorb the information, and
- how precise the numbers must be

Selecting the numbers

Your numbers should

- Be timely: using the most up-to-date figures that you can get makes your work more relevant to your audience, more convincing, and more interesting.
- Be accurate: you will destroy your credibility if you let errors slip through. Remember to credit your sources. If your sources are credible, they reinforce your credibility.

2 Guidelines for Displaying Numbers

Arrange numbers in a logical order

When numbers are organized in a logical order, the reader can see if there are patterns or exceptions.

Usually we use size order: listing numbers from largest to smallest or smallest to largest. The former is more intuitive because it is easier for people to subtract smaller numbers from larger numbers.

There may be exceptions to this rule. Imagine how the reader will use the information. It is possible that an alphabetical, chronological, or other type of order will be more useful. Organize the data to suit the needs of the reader.

Create focus by classifying numbers

Classify numbers helps readers focus on the message in the numbers. The most common classifications are totals, averages, and percentages.

Although common, totals are not always necessary. However, some situations demand totals: budgets, for example. They often give the reader a good overview.

Averages collapse information so that it can more easily be compared.

While totals and averages help the reader see the volume or total quantities involved, percentages show the reader proportions.

In general, it is better to show only one classification at a time. If you try to put two or more types of classifications into a table you may simply make the table too complex for the message to appear. If your audience needs to know both amounts and their proportions, arrange the information in two simple tables instead of one larger, more complex one.

Here are some Before and After examples.

Before

Spending Analysis FY 2001-2002 to FY 2003-2004

	Airline	Lodging	Restaurant	Retail	Vehicle	Cash Advance	Other	Total
FY 2001-2002								
Amount	287.73	25,101.47	3,963.33	1,056,155.22	0.00	0.00	10,495.20	1,096,002.95
Count	2	45	24	3,523	0	0	42	3,636
FY 2002-2003								
Amount	11,344.37	42,068.05	14,892.96	2,474,518.01	54.47	0.00	46,914.18	2,589,792.04
Count	49	70	72	12,002	1	0	151	12,345
FY 2003-2004								
Amount	16,713.56	19,695.09	8,027.14	2,208,231.81	1,430.18	0.00	49,810.66	2,303,908.44
Count	51	31	38	11,826	4	0	140	12,090
Total Amount	28,345.66	86,864.61	26,883.43	5,738,905.04	1,484.65	0.00	107,220.04	5,989,703.43
Total Count	102	146	134	27351	5	0	333	28,071

Better—but still needs work

Table 2 Departmental acquisition card purchases by BMO category, fiscal years 2001–04

	Airline	Lodging	Restaurant	Retail	Vehicle	Cash Advance	Other	Total dollars
2001–2002	288	25,101	3,963	1,056,155	0	0	10,495	1,096,002
2002–2003	11,344	42,068	14,893	2,474,518	54	0	46,914	2,589,791
2003–2004	16,714	19,695	8,027	2,208,232	1,430	0	49,811	2,303,909
Total dollars	28,346	86,864	26,883	5,738,905	1,484	0	107,220	5,989,702

Better—but could still be improved

Table 2 Departmental acquisition card purchases by BMO category, fiscal years 2001–04[a]

BMO Category	2001–2002 FY		2002–2003 FY		2004–2005 FY		Total dollars
Airline	288	(2)	11,344	(49)	16,714	(51)	28,346
Lodging	25,101	(45)	42,068	(70)	19,695	(31)	86,864
Restaurant	3,963	(24)	14,893	(72)	8,027	(38)	26,883
Retail	1,056,155	(3,523)	2,474,518	(12,002)	2,208,232	(11,826)	5,738,905
Vehicle	0	(0)	54	(1)	1,430	(4)	1,484
Other	10,495	(42)	46,914	(151)	49,811	(140)	107,220
Total dollars	1,096,002		2,589,791		2,303,909		5,989,702

[a] The number of statements examined is shown in parentheses.
Note: While BMO also has a Cash Advance category, there were no Cash Advance transactions during fiscal years 2001–04.

Highlight comparisons

Use relevant comparisons that give your audience a context. This year's sales figures all by themselves say very little. This year's sales figures compared to last year's will demonstrate an increase, a decrease, or a static state. Items that you want the reader to compare should be visually close to one another to help the reader make the connection. The following Before and After examples illustrate how readability can be improved by keeping comparisons visually close.

Before

During the period audited, the department received rebates from both banks. The rebates were:

- FY 2001-2002 $7,551.74
- FY 2002-2003 $10,437.88
- FY 2003-2004 $8497.22
- Total Rebates Received $26,486.84

Using the information obtained from the BMO system and assuming that the department paid BMO electronically within seven days of receiving the statement, the auditors

calculated that the department could have received the following rebates if the suggested processing system had been used instead of the current system in use:

- FY 2001-2002 $7,124.00
- FY 2002-2003 $19,423.00
- FY 2003-2004 $17,279.00
- Total Estimated Rebates $43,826.00

After

As shown in Table 4, the Department's rebates from both banks for this three-year period could have been substantially greater if the suggested processing system had been used instead of the current system in use. The possible rebates were calculated by the auditors using the information obtained from the BMO system.

Table 4 Actual and possible rebates for FYs 2001–2004

	Actual $	**Possible $***
FY 2001–2002	7,551.74	7,124.00
FY 2002–2003	10,437.88	19,423.00
FY 2003–2004	8,497.22	17,279.00
Total rebates	26,486.84	43,826.00

*Estimates were based on paying within seven days of receiving statements.

Example of Substantive Edit

Read the following appendix to a Transportation Safety Board of Canada (TSB) accident report. The objective of this exercise is to rearrange the numerical information so that the reader can more easily make comparisons and so that the author gets the message across more clearly.

Background information. The purpose of TSB reports is not to assign fault or determine civil or criminal liability, but to advance transportation safety. With that in mind, the reports must examine what actually happened and research what could have happened under different conditions to assess what risks were involved and if they could have been avoided.

In this instance, a cattle truck collided with a moving train. In brief, the circumstances were: a driver who was fatigued, driving late at night in thick fog on a route he had never before travelled. Road conditions were good. The cattle truck was composed of two sections: a tractor (the cab where the driver sits) and a trailer that holds the cattle. The brakes on the tractor were working, but the brakes on the trailer were not. The driver was unaware that the trailer brakes were not working. Because of the fog, the truck was not travelling very fast.

Some of the circumstances related to this particular accident could have been changed so that the accident would not have happened. For example, the driver could have stopped for the night instead of continuing to drive when fatigued. In accordance with regulations, the driver could have checked all the truck brakes when they stopped for supper; he didn't.

Rearrange the data for easier comparison. If the message of the report is to emphasize safety factors under the control of the driver, is it necessary to report all of the data from TSB's simulation?

Appendix 1: TSB Truck Stopping Distance Analysis

The TSB Engineering Branch determined, at various speeds, the critical point that a driver must see the train, react to start braking and brake the truck to a stop. The report studied scenarios with all the truck's brakes working or with only the tractor brakes working on good, typical or poor pavement. Most drivers perceive a problem and react within a range of 3.0 seconds to 4.5 seconds in situations of complex or inconspicuous stimuli. For this occurrence, a conservative figure of 3.75 seconds was used, however, it was likely longer as it was after midnight and the driver was likely fatigued, the road and the crossing were unfamiliar and the train and signals were obscured by fog.

All drivers stated that Highway 23 was in good driving condition to the crossing. The RCMP and Alberta Transportation inspectors found that there were no skid marks evident on the pavement from the semi-trailer truck approaching the crossing.

The second driver slowed down on the highway grade and concentrated on the road. He did not look at his speedometer but stated that he reduced speed to a slow speed as the fog was very thick so he could see the road's pavement line in the low beam headlights. The TSB lab determined the stopping distances for speeds of 10 to 60 km/ph on good to poor pavement. The TSB lab determined the following total stopping distances for the accident vehicle with full braking and trailer brake failure for a 2.5 second and a 3.75 second reaction time as follows:-

Trailer Brake failure - 3.75 seconds perception and reaction

SPEED	**GOOD**	**POOR**
60 km/h	100m (330')	127m (416')
50 km/h	79m (258')	97m (317')
40 km/h	59m (192')	70m (230')
30 km/h	41m (134')	47m (155')
20 km/h	25m (82')	28m (92')
10 km/h	11.5m (38')	12m (40')

With a slow perception and reaction time and trailer brake failure for the driver to have

stopped clear of the crossing at a speed of 20 km/ph he would have to have seen the train a minimum of 25 metres (82 feet) away from the crossing on good pavement; 28 metres (92') on poor pavement.

Trailer Brake Failure - 2.5 seconds perception and reaction

SPEED	**GOOD**	**POOR**
60 km/h	80m (262	106m (347')
50 km/h	61m (201')	79m (260')
40 km/h	45m (147')	56m (185')
30 km/h	30m (100')	37m (121')
20 km/h	18m (60')	21m (69')
10 km/h	8m (26')	9m (29')

With a normal perception and reaction time and trailer brake failure for the driver to have stopped clear of the crossing at a speed of 20 km/h he would have to have seen the train a minimum of 18 metres (60 feet) away from the crossing on good pavement; 21 metres (69 feet) on poor pavement.

Full Braking - 3.75 seconds perception and reaction

SPEED	**GOOD**	**POOR**
60 km/h	81m (265')	93m (304')
50 km/h	65m (213')	73m (240')
40 km/h	50m (164')	55m (181')
30 km/h	36m (118')	39m (127')
20 km/h	23m (75')	24m (79')
10 km/h	11m (36')	11m (37')

With a slow perception and reaction time and full braking for the driver to have stopped clear of the crossing at a speed of 20 km/h he would have to have seen the train a minimum of 23 metres (75 feet) on good pavement; 24 metres (79 feet) on poor pavement.

Full Braking - 2.5 seconds perception and reaction

SPEED	GOOD	POOR
60 km/h	60m (197')	72m (236')
50 km/h	48m (156')	56m (183')
40 km/h	36m (118')	41m (135')
30 km/h	25m (83')	28m (93')
20 km/h	16m (52')	17m (57')
10 km/h	8m (25')	8m (26')

With a normal perception and reaction time and full braking for the driver to have stopped clear of the crossing at a speed of 20 km/h he would have to have seen the train a minimum of 16 metres (52 feet) on good pavement; 17 metres (57 feet) on poor pavement.

Decide how many digits the reader needs

Very few people can remember exact numbers beyond four places: 246,785 is difficult to remember; 250,000 isn't. Rounded numbers are easier to compare. Mentally subtracting 246.8 from 459.3 is difficult. Subtracting 250 from 460 is much easier.

Rules for Rounding Numbers

1. **Don't round twice.** That is, don't round the numbers, then add them, then round the total. It introduces unacceptable mathematical errors.

2. **Don't adjust rounded numbers so that they total correctly.** Add a footnote: "Deviations in totals are due to rounding."

3. **Do not round in succession.** If you are converting 45 lb. into kilograms, you divide 45 by 2.2, and the calculator shows 20.454545. Rounded correctly it produces: 20. If you round in succession you will get 21, which is incorrect.

4. **Don't round numbers until the document is finished.**

3 Tables

Tables simplify, explain, and condense information. They excel at presenting

- data for which precise numeric values are important
- a large number of values in a compact form
- a summary of information
- information too complex to be easily or concisely explained in words

Some guidelines for designing tables

Remember that a table should be "stand alone". It should be sufficiently complete that a reader can understand it without reference to the text. That is, the reader should be able to understand the following:

- the purpose of the table
- the context of the data (the W's)
- the location of specific variables in the table
- the units of measurement used for every number in the table
- the data source
- definitions of relevant terms and abbreviations

It should be as simple as possible (remember: clear, concise, correct) and logically organized.

The units, symbols, and data for the table must be consistent with those used in the text.

Tables that contain similar types of information should have similar formats.

Do not present the same data in both a table and a chart.

Don't use a table if the information can be conveyed in a sentence or two.

Number and title

Even if there is only one table in a document, it must have a number and title. Every table in a document must have a unique number and title. Table titles should be concise, but informative. It should be a phrase, not a sentence. It should not simply list the column headings, but specify a category that includes the variables in the table. For example, if the headings are "Catholic", "Protestant", "Jewish", and "Pagan", the title could use "by religion".

4 Charts

Charts are best used to demonstrate large differences rather than fine details; to get across a specific message rather than a general one. While they are not as versatile as tables, they can be visually stunning and, thus, memorable and persuasive. Each type of chart has its strengths and weaknesses. These are the most common types of charts.

Bar charts

Bar charts use bars or columns to compare differences between two or more variables.

Paired bar charts compare two or more coupled items. They are very commonly used, for example, to divided information by men and women, or by age groups.

Component bar charts (sometimes called stacked bar charts) are used to show parts of the whole. Component bar charts can be difficult to understand and can too easily distort data because when the lower segments vary in size, the upper segments begin at varying points.

Line charts

The strength of line charts is their ability to show changes over time. They are usually easy to understand and visually quite effective. By careful not to use more than five lines. At five or above, readers have difficulty sorting out the lines.

Remember that two or three points in time are not sufficient data to show a pattern over time. If you have only two or three numbers, describe them in the text. If you have only three or four numbers, use a table.

Pie charts

Pie charts use a circle to show parts of a whole. Limit the slices to five or fewer. Arrange the slices in an order, usually from largest to smallest, beginning at 12 o'clock. Finally, indicate the total volume or quantity.

Pictographs

Pictographs are bar charts that use images, symbols, or some type of picture to represent quantities rather than bars. They are often seen as a way to enliven text, to make the subject more interesting. However, they have some inherent difficulties. If not done with great care, they distort the data: do not use an image in different sizes; use multiple of the same size image. Remember that images can communicate only approximate numbers.

Eliminating Chart Junk

Software that can create tables and charts offers the user an incredible range of options that include colour, shadowing (3-D effects), background designs, etc. Unfortunately,

because they are there people want to use them whether they are appropriate or not. In most cases, they aren't appropriate. When creating charts, most software offers an overwhelming amount of chart junk. Don't be tempted! Stick to the simplest charts available in your software. Make sure that any words are large enough to read. Above all, do not place labels horizontally, sideways, or in stacked letters.

Checklist for choosing a chart

✓ Useful
✓✓ Excellent

For data that shows	Recommended type of chart			Comments
	Bar chart	Line chart	Pie chart	
Parts of a whole	✓✓		✓	
Changes over time	✓	✓✓		Line charts are excellent for showing changes over time
Comparisons	✓	✓		Two or more pie charts should not be used to draw comparisons

Table or chart?

Use the nature of your data and your purpose to decide whether to use a table or a chart, while remembering the needs of your reader.

Use a table for	Use a chart for
Precise numbers	Trends and relationships
Large quantities of numbers as in reference material	Changes over time
Comparisons	Comparisons
Parts of a whole	Parts of a whole
Great range between the largest and smallest figures	Explaining a point vividly

If you aren't certain of which to use, it will always be safer to choose a table over a chart. Tables are also easier and faster to design.

5 Guidelines for Writing About Numbers

Give the reader a context

All information needs context. The journalist's "W's" work perfectly: who, what, when, and where. Compare these sentences:

Sufficient context: "Between 1981 and 2001, there were 42 million deaths in Canada attributable to smoking cigarettes."

In addition to providing context, you can give the reader a reference point for comparison: "Between 1981 and 2001, there were 42 million deaths in Canada attributable to smoking cigarettes. That represents 32 per cent of all deaths in Canada during that period."

Use easy-to-grasp examples

One simple and plausible example can embed your message in the reader's mind. You could telling someone that a healthy portion of meat at a meal is about four ounces. Then you could say, "That's about the size of a deck of cards." The image of the example is much more accessible. Make sure that the reader understands whether your example is typical or atypical, normative or extreme.

Compare these statements.

> "In 2006, the average precipitation in the Ottawa area was 110 inches."
>
> "In 2006, the average precipitation in the Ottawa area was 110 inches, 25 inches below normal."
>
> "In 2006, the average precipitation in the Ottawa area was 110 inches, 25 inches below normal, making it the third driest year on record."

Match your vocabulary to your audience

Define your terms, acronyms, and symbols. Will your audience understand and be comfortable with your terms—even if you define them? Don't be afraid to use widely understood phrases in place of numbers: everyone knows that a dozen means 12, a century means 100 years, etc. But be careful about unspecified terms: "Generation X" in general means the generation following the baby boom, but has been given various cut-off years.

Give the reader explanations

Presenting the numbers, whether in text, table, or chart, is only a first step. While readers should be able to draw their own conclusions, if they are unfamiliar with the topic you need to point out the comparisons, the generalizations, and the exceptions, and explain the implications. This is part of using numbers to tell a story.

Describe the size and direction of associations

Variables can have a positive or direct association (as the value of one variable increases, the value of the other variable also increases) or a negative or inverse association (as the value of one variable increases, the value of the other variable decreases). You should also tell the reader if the change between variables is large or small.

Understand your variables

Some ways of presenting numbers only work if variables are measured in continuous units, others work only for variables that are classified in categories.

Continuous variables are measured in continuous units such as age, metres, dollars, etc.

Categorical variables classify information into discrete categories, such as gender, smoking or non-smoking, etc. There are two types of categorical variables:

> Ordinal variables have categories that can be ranked according to the values of those categories, such as letter grades (A+, A, A–).
>
> Nominal categories have no inherent order. Some examples are race and religion.

Some continuous variables can be classified into categories. We usually treat age and income as categorical variables: 15–19, 20–24, 25–29. This helps simplify information.

Specify and define units of measurement

The unit of analysis identifies the level at which numbers are reported. Poverty can be measured in a number of ways: how many individuals have income below a certain threshold or how many families have income below a certain threshold. Make sure that you use comparable units of analysis.

The unit of measurement has two aspects: scale and system of measurement. Scale refers to multiples of units, as in numbers of people or number of thousands of people, etc. Because we use so many different ways to measure things, you must specify the system of measurement being used: imperial or metric? standard time or daylight saving time?

For measures that compare parts to a whole (percentages, fractions, proportions), make sure that the reader knows what the "whole" is. Compare these statements.

> "In the 2004 federal election, voter turnout was 60.9 per cent."
>
> "In the 2004 federal election, of all citizens registered to voter, 60.9 per cent cast ballots."

Identify standards

Standards include cut-off points, recognized patterns, and records that define the highest and lowest observed values.

Some cut-off point are a fact of nature: water freezes at a certain temperature and boils at another. Other cut-off points are social: our laws define who is old enough to vote or drink.

Some standards are empirically derived, like the average height of 16-year old boys computed from a national sample. Many of these are set by various industries, like the life insurance industry. Some standards are set by decades of record keeping: weather statistics, for example.

Put it all together

Jane Miller in *The Chicago Guide to Writing About Numbers,* emphasizes the importance of summarizing the numerical information and showing how it relates to the substantive topic. She recommends using the GEE formula: generalization, example, and exceptions.

Generalization: is there a trend? Is it rising or falling? Is it stable over time? Is there a pattern? Wordings for generalizations: in general, typically, by and large. Phrases such as "virtually all, in the majority of cases, or roughly XX number of", enhance the summary by indicating magnitude.

Having described a generalized pattern, illustrate it with numbers from the table or chart. This ties the text and number presentation together.

There's almost always an exception! Sometimes they are tiny variations that simply illustrate the unpredictability of life. Sometimes they stand out and need explanation. To introduce exceptions, use phrases such as "an exception (to) or on the other hand". If the exception is literally the opposite of the generalization, use word like "on the contrary or conversely".

Resources

Books

Presenting Numbers, Tables, and Charts, Sally Bigwood and Melissa Spore (Oxford University Press, 2003).

The Visual Display of Quantitative Information, Edward R. Tufte (Graphs Press, 1983). Tufte is recognized as the leading expert in his field. He now has several books on this subject.

The Non-Designer's Design Book, Robin Williams (Peachpit Press, 1994).

The Chicago Guide to Writing About Numbers, Jane E. Miller (Chicago: University of Chicago Press, 2003).

Scientific Style and Format, 7th ed. (The Council of Science Editors, 2006). This is one of the best references for information about designing tables and charts.

Websites

All websites were last accessed in March 2007.

www.rdg.ac.uk/ssc/publications/guides/toptgs.html
Statistical Services Centre (The University of Reading) "Informative Presentation of Tables, Graphs and Statistics"

www.robertniles.com/stats/
"Robert Niles' Journalism Help: Statistics Every Writer Should Know"

www.edwardtufte.com
Edward R. Tufte is *the* acknowledge world expert on the visual display of quantitative information. This website includes a discussion board.

www.statcan.ca/english/edu/power/toc/contents.htm
Statistics: Power from Data! offered by Statistics Canada, this is an educational site that was written for students and the general public.

