# Chapter 1: Numbers Do Not Speak for Themselves

January 28, 1986 was supposed to be a milestone in space exploration. On the crew of the Challenger space shuttle was Christa McAuliffe, a social studies teacher and the first civilian to go to space. NASA had wanted to find a talented teacher who could share her journey with students and increase public interest in the Space Shuttle program at a time when the agency was under continuous pressure to find financial support. McAuliffe was chosen out of more than 11,000 applicants.

With schoolchildren all over the country watching, including students at McAuliffe's school in Concord, New Hampshire, the Challenger launched at 11:39am from the Kennedy Space Center in Cape Canaveral, Florida. That morning was unusually cold for Florida, with a temperature of 18° F. A mere 73 seconds after take-off, the shuttle exploded over the Atlantic Ocean, killing all seven members of the crew. The Challenger disaster put NASA into a public-relations crisis from which it has yet to fully recover.

**Figure 1.1: Smoke plume after the break up of the Challenger Space Shuttle**

A task force found that the cause of the Challenger disaster was a faulty O-ring seal on the solid rocket boosters. These boosters were developed by Morton Thiokol International (MTI), whose engineers were aware of the O-ring seal problems. MTI engineers had warned their superiors about the seals and even estimated a one in one hundred probability of flight failure and loss of life on the Challenger. Moreover, engineers knew that this risk would increase with low temperatures. One MTI engineer was reported to have said in frustration "what business does anyone even have thinking about [launching in] 18 degrees? We're in no man's land'" (Bergin, 2007).

Why was the Challenger allowed to launch under these conditions? The task force noted that while engineers estimated a 1 in 100 probability of disaster, their (?) managers at MTI and NASA estimated a 1 in 100,000 probability. These managers had been warned about the faulty O-ring seal, but did not appreciate the significance of the data. When officials at NASA were later queried about a memo describing the problems of the O-ring seal, they replied:

> "I don't know if anybody at that time understood the joint well enough to realize that the data was crucial"
>
> "There were a whole lot of people who weren't smart enough to look behind the veil and say, 'Gee, I wonder what this means.'"
>
> "I didn't realize the data's significance"
>
> "It sounded like old news"

Figure 1.2 below shows a memo MTI engineers sent to NASA officials. See if you can figure out why this memo did not sufficiently alarm those in charge of the launch.

Per your request, this letter contains the answers to the two questions you asked at the July Problem Review Board telecom:

1. *Question:* If the field joint secondary seal lifts off the metal mating surfaces during motor pressurization, how soon will it return to a position where contact is re-established?

   *Answer:* Bench test data indicated that the o-ring resiliency (its capability to follow the metal) is a function of temperature and the rate of case expansion. MTI measured the force of the O-ring against Instron platens, which simulated the nominal squeeze on the o-ring and approximated the case expansion distance and rate.

   At 100$^{o}$F the O-ring maintained contact. At 75$^{o}$F the O-ring lost contact for 2.4 seconds. At 50$^{o}$F the O-ring did not reestablish contact for ten minutes at which time the test was terminated.

   The conclusion is that the secondary sealing capability in the SRM field joint cannot be guaranteed.

2. *Question:* If the primary o-ring does not seal, will the secondary seal seat in sufficient time to prevent joint leakage?

   *Answer:* MTI has no reason to suspect that the primary seal would ever fail after pressure equilibrium is reached, i.e., after the ignition transient. If the primary o-ring were to fail from 0 to 170 milliseconds, there is a very high probability that the secondary o-ring would hold pressure since the case has not expanded appreciably at this point. If the primary seal were to fail from 170 to 330 milliseconds, the probability of the secondary seal holding is reduced. From 330 to 660 milliseconds the chance of the secondary seal holding is small. This is a direct result of the o-ring's slow response compared to the metal case segments as the joint rotates.

**Figure 1.2: Memo from MTI engineers to NASA, names deleted (**Winsor, 1988)

The managers at NASA were technologically literate—yet they were not able to see the implications of the numbers MTI engineers presented. They read the data, but they did not understand its significance. Consequently, they did not take actions that would have saved lives.

The O-ring memo illustrates a central argument of this book: data does not speak for itself. Many "numbers people" tend to think they can just present data and likeminded people will understand what those numbers mean. Such seamless communication, unfortunately, is the exception rather than the norm.

To be effective technical, scientific, or business professionals, we need not just hard data to support our claims, but also methods for communicating the significance of that data to our audience. Numbers do not speak for themselves—they need to be interpreted and placed in context. We need words to understand what they mean. This book teaches you strategies for helping your audience understand what your data means.

**For Discussion: The Challenger memo**
Look at the memo in Figure 1.2 and locate the places where the MTI engineers signal problems with the O-ring seal. Why do you think NASA officials failed to realize the importance of this data? What could the engineers have done to better communicate problems with the seal without sounding alarmist or emotional?

## Data reporting involves argument

If we were to draw a continuum with emotion at one end and logical reasoning at the other, most people would place data on the logical reasoning end. Data is associated with facts, evidence and cold, hard, logic. Yet there is an emotional component to how we present data.

To illustrate this emotional component, imagine someone is trying to sell you a raffle ticket. Which of the following arguments will the person trying to sell the ticket use?

You have a one-in-twenty chance of winning
You have a 5% chance of winning
You have a 95% likelihood of losing

If we were listening with only the logical part of our brain, it shouldn't matter which of these statements we heard: all three arguments are mathematically equivalent. Yet, someone selling a ticket will inevitably use the first and never the last. The first way of presenting the argument makes the audience think "Twenty is a relatively low number. Heck, I've probably bought 20 raffle tickets over the past few years. Maybe this one is my ticket."

By contrast, telling someone they have a 5% chance of winning makes them focus not on the other raffle tickets they have bought in the past, but on the probability of *this* current ticket winning. And saying "you have a 95% likelihood of losing" suggests that only a fool would buy a ticket.

Such arguments affect all kinds of decisions ordinary people make. For instance, a couple I know became worried after reading that one in fifty pregnancies by older couples will result in an abnormal fetus. But when they shared their fears with their doctor, they were reassured by hearing there was a 97% probability that everything would be fine.

The problem with this reasoning, as my friends eventually realized, is that the number quoted by the doctor was actually a *worse* statistic than the one that had alarmed them. One-in-fifty translates to two-in-one hundred abnormalities, leaving us with a 98% likelihood that everything will be fine. Thus, the doctor's reassurance that they had a 97% probability of a normal pregnancy was slightly worse than the number that had worried them.

Why did my friends find one number alarming and another, slightly worse, number reassuring? One-in-fifty, the statistic that alarmed them, is easy to visualize. We all know 50 people; it is easy to imagine one of these fifty experiencing a tragic event. By contrast, 97% is abstract and much harder to visualize. It is also a number that years of school have conditioned us to equate with success: a grade of 97% is an occasion for self-congratulation.

In other words, the same number presented two different ways can have very different effects on the audience.

Even when we are discussing data collected for scientific research, we are still selecting and presenting data to persuade readers. We want readers to believe our results are valid and our findings interesting. We want them to believe we have made a significant contribution to scientific knowledge. Our discussions about data inevitably involve argument.

## Data-based arguments depend on purpose, audience, and credibility

Thus, we have choices about how we present and argue about data. These presentation choices involve three interrelated—yet sometimes competing—concerns. Figure 1.3 uses a triangle to illustrate how these three concerns depend upon one another.

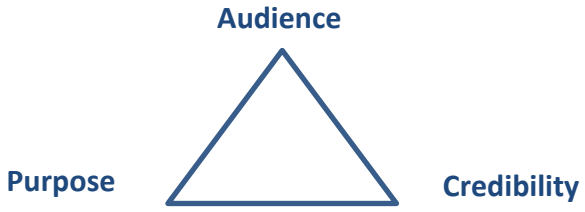


**Figure 1.3: How we present data depends upon our purpose, audience, and the need to demonstrate credibility**

At one end of the triangle is our **purpose**—the actions, thoughts, or emotions we want to inspire in our audience. Thus, someone selling a lottery ticket will want to make the ticket seem like a good opportunity while a counselor for gamblers anonymous will want to emphasize the folly of playing the lottery. These two individuals will present the same data in very different ways.

As an example, a few years ago, my two-year-old dog needed a $4,000 operation. Rather than presenting the total cost to me, my spouse, arguing on the dog's behalf, calculated the cost of the operation per year the dog was likely to live ($400 a year – a negligible figure given our other dog-related expenses). This strategy successfully diverted my attention from the large price tag, while simultaneously reminding me of the number of years left with our pet. Thus, by doing a simple per year calculation, my spouse successfully achieved his purpose of persuading me the operation was worthwhile.

How we present data is also shaped by the knowledge and needs of our **audience.** Thus, if our purpose is to persuade readers to improve a particular traffic intersection, we may present our data in one way for an engineer and a different way for a politician. A traffic engineer will want data on the number of accidents per vehicle using the intersection and how that intersection compares with similar ones throughout the country. By contrast, a politician charged with allocating public funds may only care about the accidents per person living in that area (in other words, the probability that a resident will experience an accident).

Finally, we must demonstrate our **credibility** by presenting data ethically and accurately: we can emphasize certain points but if we are perceived as distorting the data, our message will be rejected. This is not as simple as it may first seem. For instance, the statement *Campus crime rose by 50% in the past year* could be completely accurate and support our purpose to implement additional safety measures at our school. However, if our readers then learned that this 50% increase represents a jump from two to three crimes, they might likely dismiss anything else we had to say. Presenting the phrase "rose by 50%" without the context to realize that 50% is an increase of one erodes our credibility. Thus, providing enough data to gain an accurate picture is as important to earning readers' trust as presenting the data accurately.

These three considerations of purpose, audience, and credibility are intimately connected and decisions we make about one of these considerations will affect the other two. We will emphasize these three considerations throughout this book. You will learn how to select data representations for specific purposes and audiences while reporting your data with enough context and clarity to maintain your credibility—all without drowning your audience in unnecessary detail. Practicing these data reporting skills, in turn, will make you a much more critical and competent reader of others' data-based arguments. By better understanding common ways data writers emphasize or highlight certain interpretations while obscuring others, you will improve your ability to critique the statistics and data that you encounter in the news, your profession, or your daily life.

**For Discussion: Statistics on depression.**

Consider the following statistic: "*21.3% of women and 12.7% of men have experienced depression in their lifetime*" and consider how you might revise it to support a different purpose or audience.

1. How might you rewrite this statistic to encourage a woman's organization to support depression research?
2. What if you wanted to encourage more research into men's experiences of depression?
3. How might you rewrite it to minimize the importance of depression?
4. Are any of your revisions more or less trustworthy than others? Do any cast doubt on your credibility?

## Numbers can be manipulated—just like words

While some people place too much trust in numbers as facts that speak for themselves, others have just the opposite perspective and dismiss all data as untrustworthy because it can be manipulated for different purposes or audiences. You may be familiar with the witticism "There are lies, damn lies, and statistics." This statement suggests that numbers are inherently deceitful because they can be easily manipulated while retaining the appearance of truth.

However, such an attitude ignores the extent to which words and images can also be distorted. Changing the numbers in the statement "there is a one-in-twenty chance of winning" to "there is a 5% chance of winning" is similar to replacing the words in the following sentences.

*Steve told me the test was difficult.*
*Steve whined about the test being hard.*

In both cases, we have Steve communicating that the test was difficult, but the two versions produce very different impressions of Steve—just as the two different ways to describe the likelihood of winning the raffle produce different impressions of one's odds.

The fact that we can present data differently depending on our purpose and audience does not mean that we can say anything we want with it. Not all arguments about data are equal: some are deceptive and unethical while others usefully shape data to help readers make informed decisions.

The figure below illustrates these different ways of thinking about data. On the far left end is the belief that numbers offer a direct lens to the truth; on the far right end lies the belief that numbers are inherently untrustworthy. Our goal lies somewhere in the middle: we have to argue for what our data means, but we have to do so in a way that is ethical and credible.

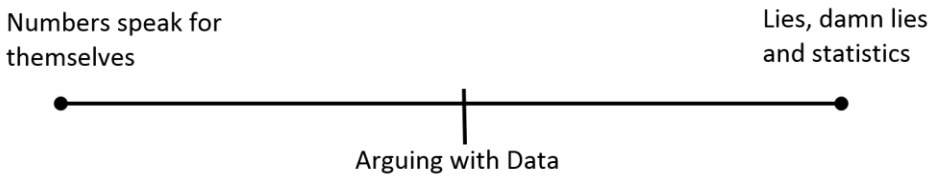


**Figure 1.4: Ethical argument lies on a continuum between no argument and pure relativism where all arguments are equal and equally suspect**

One reason that numbers and statistics have such a bad reputation may be that people simply are not practiced at using them in arguments. As young children, we learn to manipulate words to hide or highlight particular information. This ability carries over into adulthood as we become adept at scrutinizing others' words for hidden biases or purposes. However, most people have comparatively little practice with using numbers to hide and highlight information.

Consequently, these people often *feel* that numbers are biased but cannot identify what is wrong with them. The more experienced you become with making your own arguments with data, the more skilled you will become in identifying biases in other people's arguments.

## Data can be qualitative as well as quantitative

It is also worth noting that data is not just about numbers and measurements. Data also includes words, observations, and experiences. It can be qualitative as well as quantitative, encompassing information such as user comments on a website, observations of behavior in public places, or a detailed description of a medical symptom. The exact words that a dissatisfied user chooses to

characterize her experience with a website can often be more meaningful than numerical data on the number on failed submissions or the number of views.

Qualitative data can be as important as quantitative numbers in fields such as business and medicine when practitioners encounter rare or unique situations. For instance, a medical researcher might write up a detailed case study of an individual with a highly rare disease, describing the symptoms and attempted treatments so that other doctors can benefit from this knowledge. The data here is a blend of rich description and commentary as well as facts and figures.

Chapter ___ of this book focuses on qualitative data and making arguments when your data consists of words.

## Is this book for me?

This book assumes that you have mastery over working with averages, percentages, and ratios. However, unlike traditional math classes, this book challenges you to work with these concepts in real-world contexts. In addition, this book assumes you can use—or are willing to learn—a program for creating graphs, such as Microsoft Excel. If you have knowledge of statistics and know how to calculate significance tests, you will receive advice on how to present this information. However, knowledge of statistics is neither expected nor necessary in order to understand this book.

The skills you will learn in this book are fundamental to professional, personal, and civic life. Any professional will need to work with data in some way. Teachers need to report the percentages of proficient students; managers need to predict budgets and expenditures; artists and musicians may be asked to make arguments about the number of people their creations will reach or their projected impact on society.

Even if you manage to find a profession that is not numbers-intensive, basic citizenship in today's society requires understanding data. At some point, we all will need to review medical research to make an informed decision about healthcare treatments, understand the financing terms of a mortgage, or decide which politician has the better policies. The amount of data we encounter on a daily basis is only increasing and we need to prepare to use it wisely.

Thus, the concepts in this book have a wide application to your future professional, academic and personal life. The concepts require no technical expertise or background but only a willingness to learn and try out new strategies. If you approach the chapters and exercises with an open mind, you may find yourself rethinking what "writing" means.

**Is the word *data* singular or plural? Should I write *The data is…* or *The data are…*?**

In Latin, the word *data* is a plural noun and the singular form is *datum.* Thus, many people believe that *data* should be treated as a plural noun. However, others argue that when English speakers use the word *data* they are referring to an English word that has been in common usage for centuries. These people

believe we do not need to follow Latin rules and *data* should be treated as singular.

Both forms are commonly used in the English language. However, a Google search for the plural "data are" returns $6 \times 10^7$ hits while the singular "data is" returns $1.6 \times 10^8$ hits, suggesting that the singular form is becoming more popular. If we restrict our search just to news sites, we find an even bigger difference in favor of the singular form. However, a search restricted to only scholarly articles favors the plural form. These findings suggest that in most popular contexts, data is treated as a singular noun, but in academic contexts, it is still plural.

The English language is a living language affected by speakers from a wide range of cultures who influence what we consider standard. Because of these influences, words shift in meaning and grammatical forms undergo change (just compare a Shakespearean play to any contemporary writing). This book assumes that the usage of the word "data" is shifting and thus treats it as a singular noun. You should, however, be aware that there are individuals who have strong opinions on this issue and believe the singular usage (i.e., "data is") is incorrect.

## Summary

Numbers do not speak for themselves. Instead, good writers shape how they talk about numbers to fit their purpose and audience and to demonstrate their credibility. The fact that the same numbers can be used for different purposes does not mean that they should be discounted any more than we should discount words because people lie with them. Good writers use numbers ethically to make arguments that help others make good decisions.

## Works Cited

Chris Bergin (January 28, 2007). "Remembering the mistakes of Challenger". *nasaspaceflight.com*. Retrieved August 5, 2011.

D. Winsor, "Communication Failures contributing to the challenger accident: An example for technical communicators," *IEEE Transactions on Professional Communication*, vol. 31, pp. 101-107, 1988.