

The Secret Lives of Numbers

*by Golan Levin, with Martin Wattenberg, Jonathan Feinberg,
Shelly Wyncoop, David Becker and David Elashoff.*

<http://www.turbulence.org/Works/nums/>

<http://numbers.tmemain.org/>

The Secret Lives of Numbers

CONTRIBUTORS

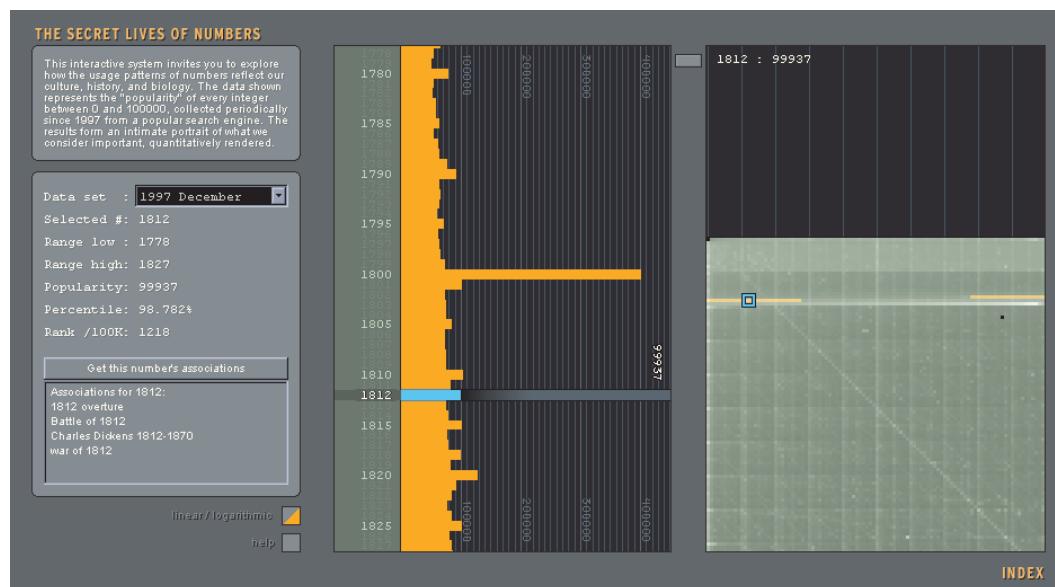
The Secret Lives of Numbers was created by:

Golan Levin [USA, concept / direction / interface design & programming], with
Martin Wattenberg [USA, interface & information design]
Jonathan Feinberg [USA, database & CGI programming]
David Becker [USA, web programming]
David Elashoff [USA, statistics consulting]
Shelly Wynecoop [USA, essay / research]

The Secret Lives of Numbers is a commission of New Radio and Performing Arts, Inc., for its Turbulence.org web site, and was developed with support from the Greenwall Foundation. Conceived late 1997 at Interval Research Corporation, the present interactive visualization was launched in early February 2002.

CONTENTS

Overview	3
Artists' Statement	4
Technical Realization	8
Contributor Biographies	9
Related Bibliography	10



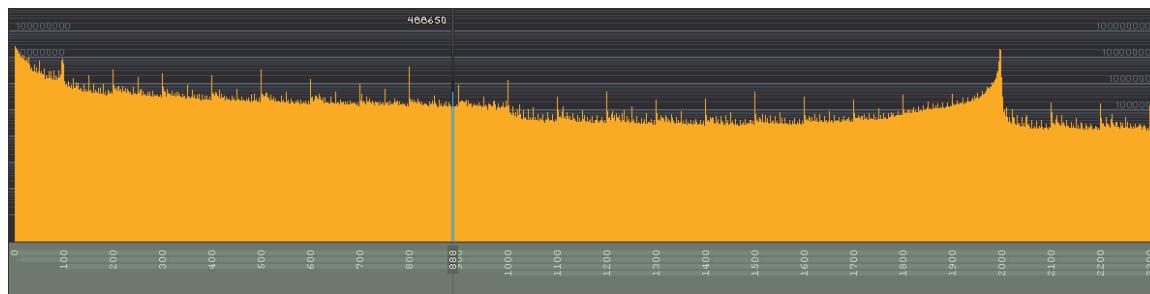
OVERVIEW

The authors conducted an exhaustive empirical study, with the aid of custom software, public search engines and powerful statistical techniques, in order to determine the relative popularity of every integer between 0 and one million. The resulting information exhibits an extraordinary variety of patterns which reflect and refract our culture, our minds, and our bodies. *The Secret Lives of Numbers* is an online interactive visualization system which allows users to explore this interesting view on our world.

For example, certain numbers, such as 212, 486, 911, 1040, 1492, 1776, 68040, or 90210, occur more frequently than their neighbors because they are used to denote phone numbers, tax forms, computer chips, famous dates, or television programs that figure prominently in our culture. Regular periodicities in the data, located at multiples and powers of ten, mirror our cognitive preference for round numbers in our biologically-driven base-10 numbering system. Certain numbers, such as 12345 or 8888, appear to be more popular simply because they are easier to remember.

Humanity's fascination with numbers is ancient and complex. Our present relationship with numbers reveals both a highly developed tool and a highly developed user, working together to measure, create, and predict both ourselves and the world around us. But like every symbiotic couple, the tool we would like to believe is separate from us (and thus objective) is actually an intricate reflection of our thoughts, interests, and capabilities. One intriguing result of this symbiosis is that the numeric system we use to describe patterns, is actually used in a patterned fashion to describe.

We surmise that our dataset is a numeric snapshot of the collective consciousness. Herein we return our analyses to the public in the form of an interactive visualization, whose aim is to provoke awareness of one's own numeric manifestations.



Here, the applet shows a section of data in a logarithmic view. Note the regular peaks at multiples and powers of ten, and the marked spike between the "years" 1500-2000.

ARTISTS' STATEMENT

PREFACE

When we choose to say there are “a lot” of gorillas, or “17 gorillas,” or “around 20,” we are making a choice based on our interests and abilities. Those interests and abilities are systematically expressed in the numbers we choose to send out into the world. The Internet provides us with a large and diverse database of the stuff people have sent out into the world, and it naturally includes large numbers of numbers. Since 1997, we have collected at intervals a novel set of data on the popularity of numbers: by performing a massive automated Internet search on each of the integers from 0 to 1,000,000 and counting the number of pages which contained each, we have obtained a picture of the Internet community’s numeric interests and inclinations. The interactive visualization which accompanies this statement attempts to make some of the more striking trends visible. Our data can be explored point by point, or viewed in larger sets, and may lend some insight into the cognitive structure of numeracy, culture, and memory.

Certain patterns are made readily visible in the data browser, such as people’s preference for multiples of 10, or reduplicative numbers such as 1010, 1111, 1212, etc. While American zip codes do not present a similarly coherent visual pattern, they can nevertheless be quite prominent—and unlike viewing cities from space, where only the larger cities are visible, here, both the larger and the more interesting places are brighter. Further highlights in the data reveal more fleeting reflections of our activities: bright spots such as those for 80486 and 68040 reveal our interest not only in technology, but tell us about the state of that technology at the time. Other points indicate our interest, or lack of interest, in history. And popular culture inevitably makes its presence felt, but perhaps less than we might expect.

Linguists have noted that the ideal construct of proper “Language” diverges considerably from its actual use, suggesting that the linguistic “ideal” may be a pale abstraction of a far more nuanced and textured practice. Although we like to think of the use of numbers as objective and removed from our personal lives, it appears they also display elements of “practice.” The denizens of the number line are not the mere automatons or corporate tools we have made them out to be: each has a personality, talents, communities, and sometimes a little *je ne sais quoi*. They reflect us. This unusual reflection is the focus of this project.

NUMBERS ARE TOOLS

In learning how to abstract, we learn that all information is potentially expressible in numbers. The ability to abstract from perceived phenomenon (such as a group of cows, or the effects of gravity) to descriptions of the physical world (such as 23 or $9.8s^2$) has allowed us to see commonalities in phenomena that may first have appeared to be distinct.

One consequence of abstraction is that we must ignore the individual characteristics of the entities we abstract. As a result, the numbers we use to codify these abstractions must also lack character. Twenty-three cows may be better (or worse) than three cows, but "23," is not better than "3." Both numbers are simply descriptors, which inherit their meaning solely from taking part in fixed systems of fixed relations with other numbers. Apart from the existence of the numeric system (and the numbers' participation in it) individual numbers have no meaning.

Thus, our number system is seen as an objective tool—a tool that does not reflect human preference, emotion, or inconsistency. As such it is a tool used not to express ourselves, but is reserved only to describe the world around us. We do not write poetry with numbers, nor do we express our personal doubts or prejudices through them ...except as our humanity is projected onto the emotionless toil of mathematical proof, ledger balances, or pedagogical exercises. But like every symbiotic couple, the tool we would like to believe is separate from us (and thus objective) actually provides an intricate reflection of our thoughts, interests, and capabilities. One intriguing result of this symbiosis is that the numeric system we use to describe patterns, is actually used in a patterned fashion to describe.

We are imperfect users of our perfect tool. Buildings often skip the 13th floor, there is no year 0, and our only contact with very large numbers comes from government debt, numbers which remain unreal to us for their very size. We spend most of our time using numbers not for calculating, or even measuring, but in acts of remembering, guessing, and simplifying. The secret lives of numbers presented here, deals less with the apparently inviolate laws we have contrived for the staid little number, than with our own natures expressed through their quixotic use. This secret life tells us not only about the number system we have fashioned, but also about our cognitive, professional, and creative liaisons with the various inhabitants of the number line.

THE DATA ITSELF

The characteristics of the data we describe below include aspects of the population as a whole, patterns of smaller groups, and individual charm. All tell us something about our culture as expressed through numbers.

The Population

The first and most striking characteristic of the data is the overall distribution of the numbers' occurrences. Instead of the uniform distribution one might expect if every number were equally useful, we see an exponential drop-off in popularity beginning with the number 1. These earliest, and most popular individuals aren't a glamorous set, but instead see their popularity rise from their accessibility. They are the first numbers we learn, and are the easiest to understand and use. For these reasons, with every increase in magnitude along the number line, the numbers see a sharp drop in this kind of basic popularity.

Prominent Families

There are, however, certain numbers further down the line that enjoy great popularity in spite of their greater number of digits. These numbers comprise some of the basic "royal families" of the number line: the base-2, base-10, base-12, and base-60 families. While most of these families have their niches in technology, time-based media and the English measurement system, the most prominent of these families is the one which clearly reflects our biology. The 10 family can be seen everywhere: numbers at multiples of 10, (and powers of ten, like 100, 1000, etc.) enjoy a popularity far greater than their neighbors throughout the data. Our biases for "rounding" suggest that most of these numbers' high standing comes at the direct expense of their nearest neighbors. The positively ignored 49949 is one such serf, apparently yielding its worldly recognition to its more prominent neighbor 50000.

In addition to the population at large, and the most prominent families within it, some numbers enjoy a certain degree of popularity owing to their occupation. Here are just a few examples; many more can yet be found in the data:

The Journalists

90-99, and 1990-2002 (no time like the present)

The Stars

90210 (the television show)

The Nerds & Techies

68040, 68030, 68000 (Macintosh)

286, 386, 486, 8086, 80286, 80386, 80486 (its Intel competitor)

2,4,16,32,64,128,256,512,1024 (base-2, now RAM sizes)

2400, 4800, 9600, 19200, 38400 (baud rates)

8859 (from the ISO-8859 character set)

The Responsible Citizens

1040, 1041 (Uncle Sam loves you)

10036, 26161, 13131, 77058... (American zip codes)

800, 888, 877 (toll-free phone number prefixes)

52062, 52064, 52066 (German postal codes)

The Salesmen

98, 99 (why don't things ever just cost \$1.00 even?)

900 (sex sells)

A LOOK AT OURSELVES

Moving up a level from the individual points and patterns covered in the data, what can we deduce about ourselves by examining the kinds of interests we display?

The explosion of occurrences of numbers in the range from 1990-2002 points not only to the growth of the Web during this period, but also to a kind of temporal narcissism. We are most interested in the year in which we live, and are less interested in events in the past regardless of their import.

Furthermore, historical years before the 1990s do not have magnitudes that reflect an atemporal vantage point, but appear instead to be talked about less, the further they fall from the present. (Perhaps this trend would not be visible if we were to do these searches on Web sites devoted only to history.) These phenomena point to the possibility of measuring the longevity of our memory, or the degree to which we care about how historical events may have shaped our present lives. In the slope of our curve between 1600 and 2002, we see an image of our cultural rate of forgetting.

FUTURE DIRECTIONS

The discussion above appears to suggest some degree of historical ignorance on the Web. Can we observe any awareness of historical context—a context which is actually present in our collective consciousness—on the Web? In the periodic gathering of this data, a vernacular history suggests itself. From the information we have gathered, we can construct a time line of historic individuals searched for by year of birth, death, or significant event, to create a history that consists solely of individuals holding the public interest in any given year. This history could be reconstructed every year with hopes of observing shifts in this expressed histori-

cal context over time. A preliminary look at the names that accompany number searches tells us that we may not only retrieve biographical (Sartre, 1905-1980), and historical information (Columbus, 1492) but also glimpses into how people are feeling about the individuals singled out (Bill Gates, 666).

Further comparisons over time are possible, and are likely to yield more interesting results than a look at any single slice of time can provide. Numeric searches on a defined subset of Web pages, such as those devoted to medicine, history, physics, or literature, may generate further insights. Differentiating country of origin may also prove interesting. On the most basic level, arranging the data using different parameters may shed light on patterns not visible in the current arrangements.

The numeric system has helped document the regularity and periodicity inherent in our environments and ourselves for millenia. In allowing us to examine our own patterns of use, we hope this data will be used to shed some light on our cultural biases and numeric capacities. We also hope to underscore the influence technology has had in changing the set of numbers we can and cannot imagine.

TECHNICAL REALIZATION

The Secret Lives of Numbers is implemented as a Java applet and is best viewed in Internet Explorer 5 on a fast Windows2000 PC.

The visualization's dataset was collected at regular intervals from the Altavista search engine. A Perl script was written to automatically collect this data, stripping out the number of responses from web pages such as the following:



CONTRIBUTOR BIOGRAPHIES

Golan Levin [golan@flong.com, b.1972] is an artist, designer and composer interested in developing artifacts and experiences which explore supple new modes of audiovisual expression. His work has focused on the design of systems for the creation and performance of simultaneous image and sound, as part of a more general examination of communications protocols for individual engagement and social dialogue. Levin is the recipient of an Award of Distinction in the Prix Ars Electronica 2000 for his *Audiovisual Environment Suite* (AVES) interactive software and its accompanying audiovisual performance, *Scribble*. Levin received undergraduate and graduate degrees from the MIT Media Laboratory, where he studied with John Maeda in the Aesthetics and Computation Group. Prior to this, he worked as a research scientist and designer at Interval Research Corporation for four years. He currently resides in New York City

Martin Wattenberg is an artist and designer interested in the expressive content latent in pure data. As Director of Interactive Media at SmartMoney.com, Martin developed the highly popular MarketMap visualization tool, for which he received an honorable mention in the 1998 Prix Ars Electronica.

Shelly Wynecoop has a background which combines research in philosophy, linguistics, cognitive psychology, and statistics. She is currently Associate Curator of Digital Media at the American Museum of the Moving Image; prior to this, she completed graduate studies in epistemology at the University of Chicago.

Jonathan Feinberg is a New York City-based software engineer and musician, and has collaborated on a number of interactive artworks with Martin Wattenberg, Marek Walczac, and Golan Levin.

David Elashoff is a Ph.D. candidate in the Statistics department at Stanford University.

David Becker is a graduate of the MIT Media Laboratory and a programmer at Amova.com.

RELATED BIBLIOGRAPHY

- Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78:551.
- Boyle, J. 1994. An application of the Fourier series to the most significant digit problem. *American Mathematical Monthly* 101(November):879.
- Hill, T.P. 1998. The first digit phenomenon. *American Scientist* 86(July-August):358.
- Newcomb, S. 1881. Note on the frequency of the use of digits in natural numbers. *American Journal of Mathematics* 4:39.
- Raimi, R.A. 1976. The first digit phenomenon. *American Mathematical Monthly* 83:521.