minimaxir.com

# Analyzing the Patterns of Numbers in 10 Million Passwords

5-7 minutes

The primary purpose of a password is to serve as an unique verification identifier for a given user. Ideally, the password for a given website or service should be both random and unique; if the letters and/or numbers in the password follow any patterns, then they might be easier to guess by an intruder. For example, someone may put their birth year such as "1987" or "1988" in their password, which makes the passwords easier to remember, but consequently easier to break.

A few weeks ago, security researcher Mark Burnett released a list of 10 million passwords compiled from various sources over the years. Reddit user jalgroy posted a histogram of the years used in these passwords, which I've verified using my own scripts:
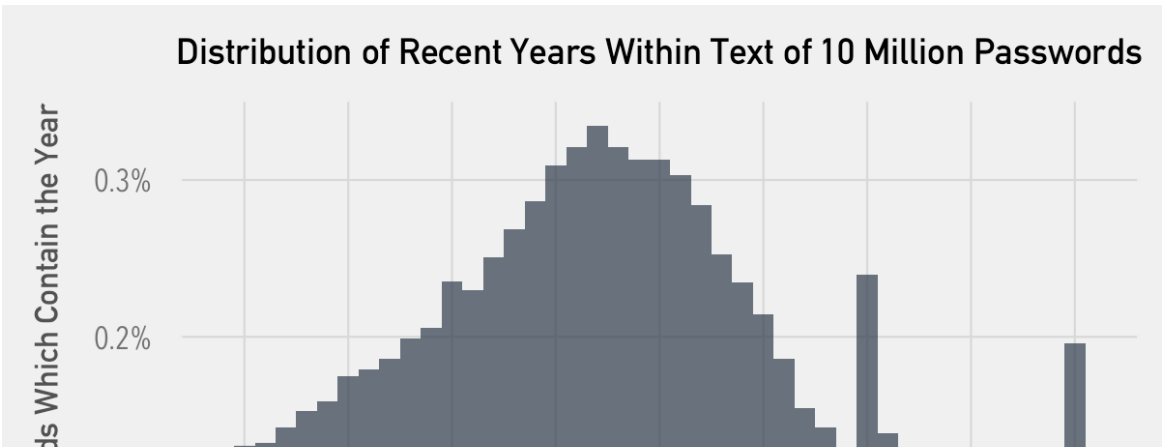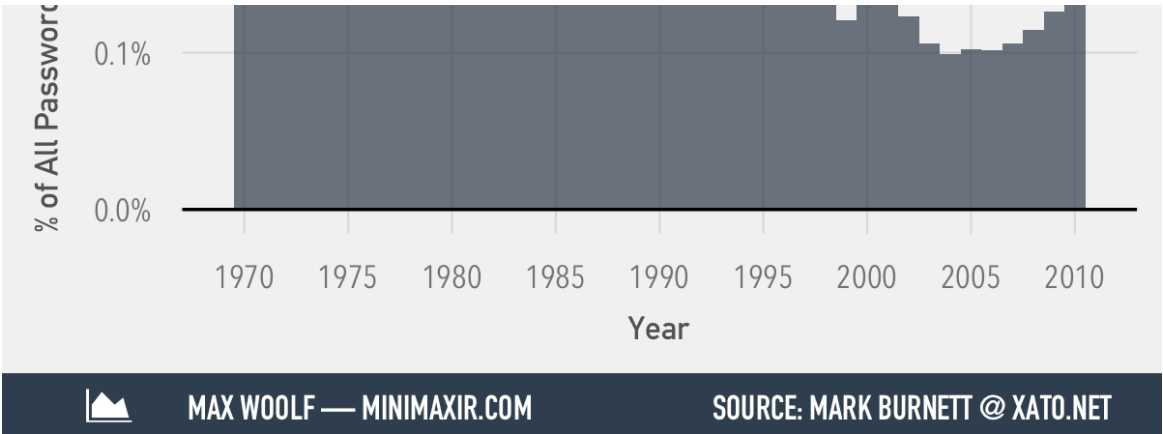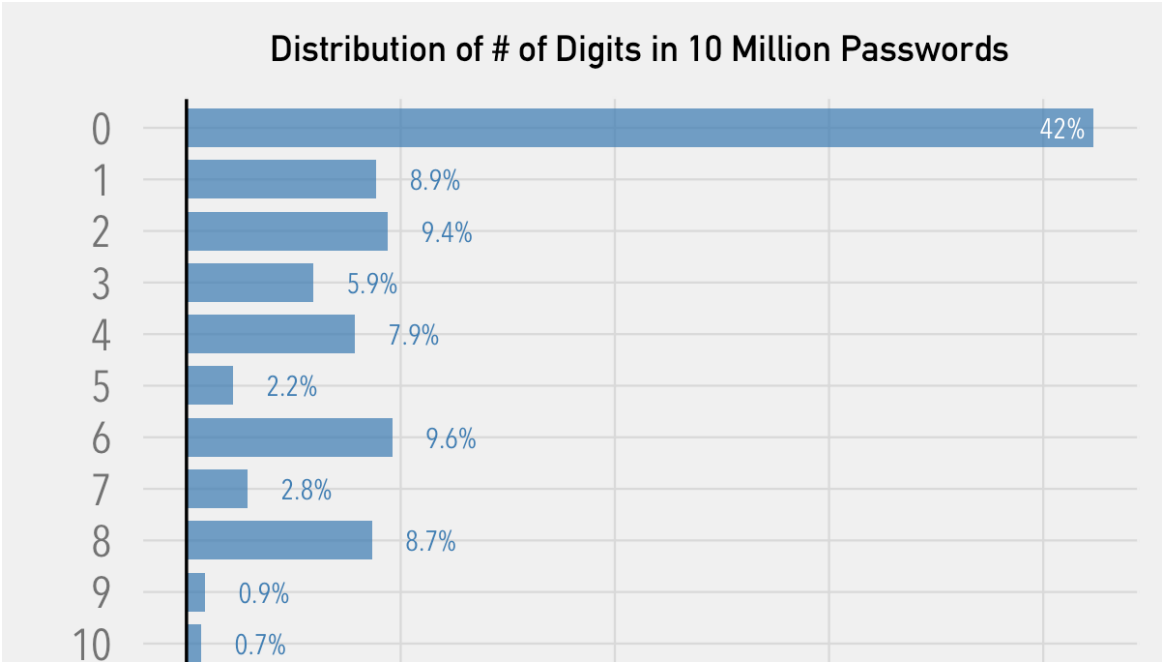
There is a clear maximum at 1987 which implies a current age of about 28. This makes sense, as internet users in their 20's are generally considered to be very attuned to internet usage. The spike at 2000 is likely not because it's a birth year, but because 2000 is a kewl number.

There are actually many similar patterns for numbers in passwords, which involve surprising yet intuitive logic.

Digit Behavior

The distribution of the number of digits in passwords varies significantly.

0%	10%	20%	30%	40%

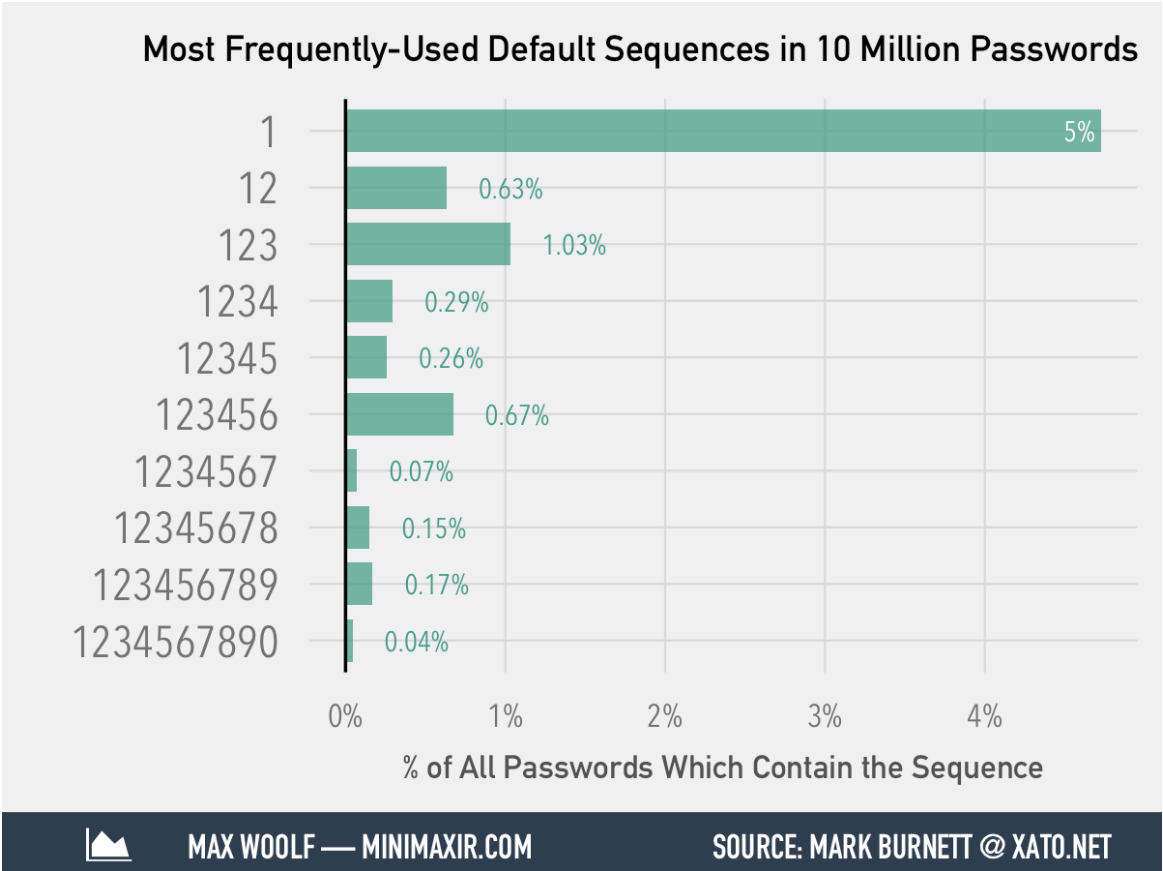**% of All Passwords Which Contain the # of Digits**

MAX WOOLF — MINIMAXIR.COM   SOURCE: MARK BURNETT @ XATO.NET
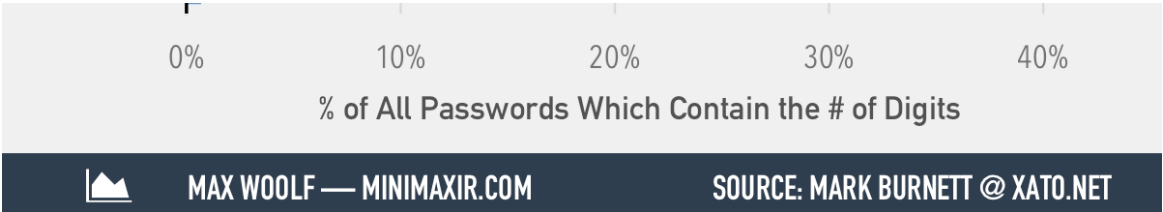
42% of passwords have zero numerical digits, which implies that 58% of passwords have atleast one digit. However, the local maxima in number of digits in a password all occur at *even* numbers of digits, which may imply that humans have an easier time of remembering even amounts of numbers.

If you look at a typical keyboard, you'll note that the default sequence of numbers is **1234567890**. If the user wants a number in their password that is easy to type, drawing from this sequence of numbers might be a good idea.

Most Frequently-Used Default Sequences in 10 Million Passwords

Sequence	%
1	5%
12	0.63%
123	1.03%
1234	0.29%
12345	0.26%
123456	0.67%
1234567	0.07%
12345678	0.15%
123456789	0.17%
1234567890	0.04%

0%	1%	2%	3%	4%

% of All Passwords Which Contain the Sequence

MAX WOOLF — MINIMAXIR.COM   SOURCE: MARK BURNETT @ XATO.NET

Note that the length of the sequence is uncorrelated with the number of occurrences of the sequence. Many more people use
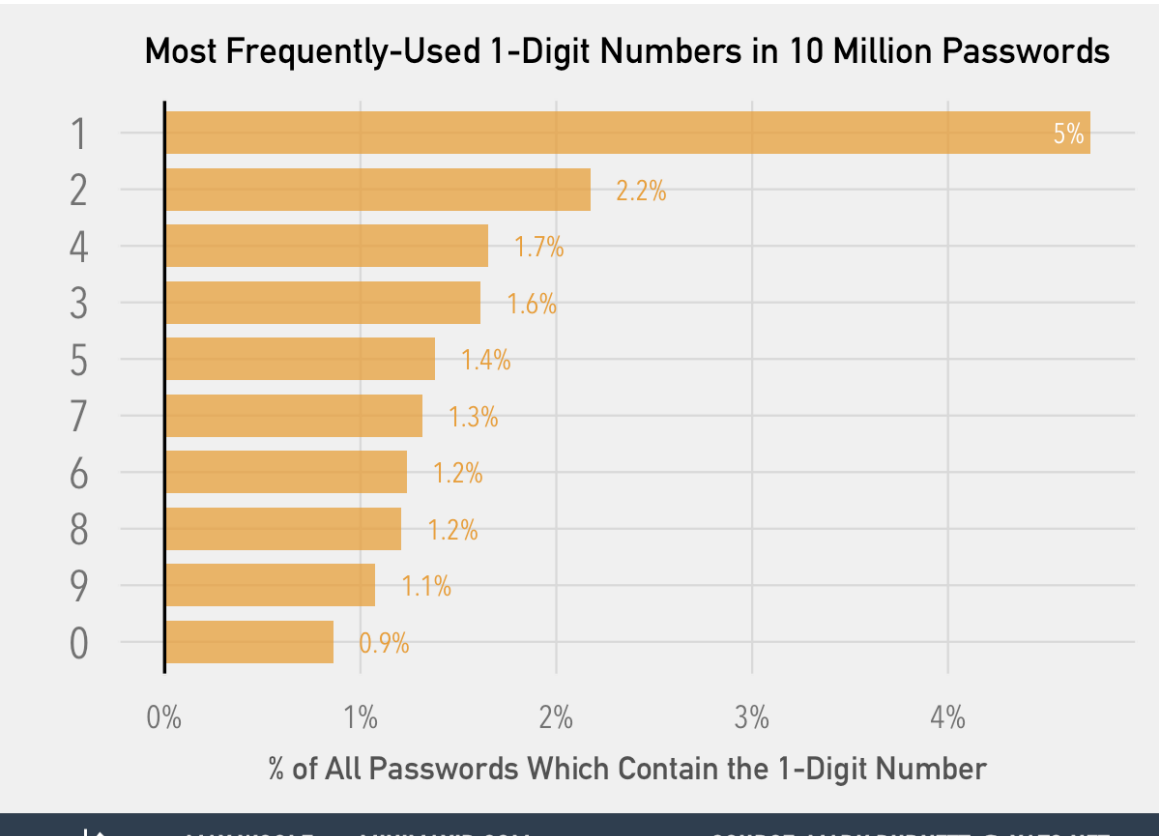
123 in a password than just 12, even though it's longer. 123, as a triplet of numbers, may be easier to remember by the average person than a pair of numbers. However, that contradicts the logic above that even numbers may be easier to remember, which suggest that another factor may be involved.

Sequences of numbers are popular, but are some sequences of numbers more popular than others? Let's look at the order and composition of 1-digit, 2-digit, and 3-digit numbers in these 10 million passwords.

More on Digit Patterns

Note: all number patterns are distinct number patterns, e.g. 2-digit numbers analyzed are not subsets of 3-digit or larger numbers.

Take a look at the most used single-digit numbers:

Most Frequently-Used 1-Digit Numbers in 10 Million Passwords

Number	%
1	5%
2	2.2%
4	1.7%
3	1.6%
5	1.4%
7	1.3%
6	1.2%
8	1.2%
9	1.1%
0	0.9%

% of All Passwords Which Contain the 1-Digit Number

1 is by far the most-used single-digit number, which may be due to the fact that it is the left-most number on the keyboard and therefore an easy press for services that force the inclusion of a digit in the password. Relatedly, 9 and 0 are the least-used single-digit numbers. That's intuitive enough. But does that hold for more complex patterns?

Let's look at the most-used 2-digit patterns, including numbers with 0 as a leading digit:

Most Frequently-Used 2-Digit Numbers in 10 Million Passwords

2-Digit Number	%
12	0.63%
11	0.37%
69	0.33%
01	0.28%
22	0.27%
13	0.27%
99	0.26%
10	0.24%
23	0.23%
21	0.20%
00	0.18%
77	0.17%
88	0.16%
33	0.15%
19	0.15%

% of All Passwords Which Contain the 2-Digit Number
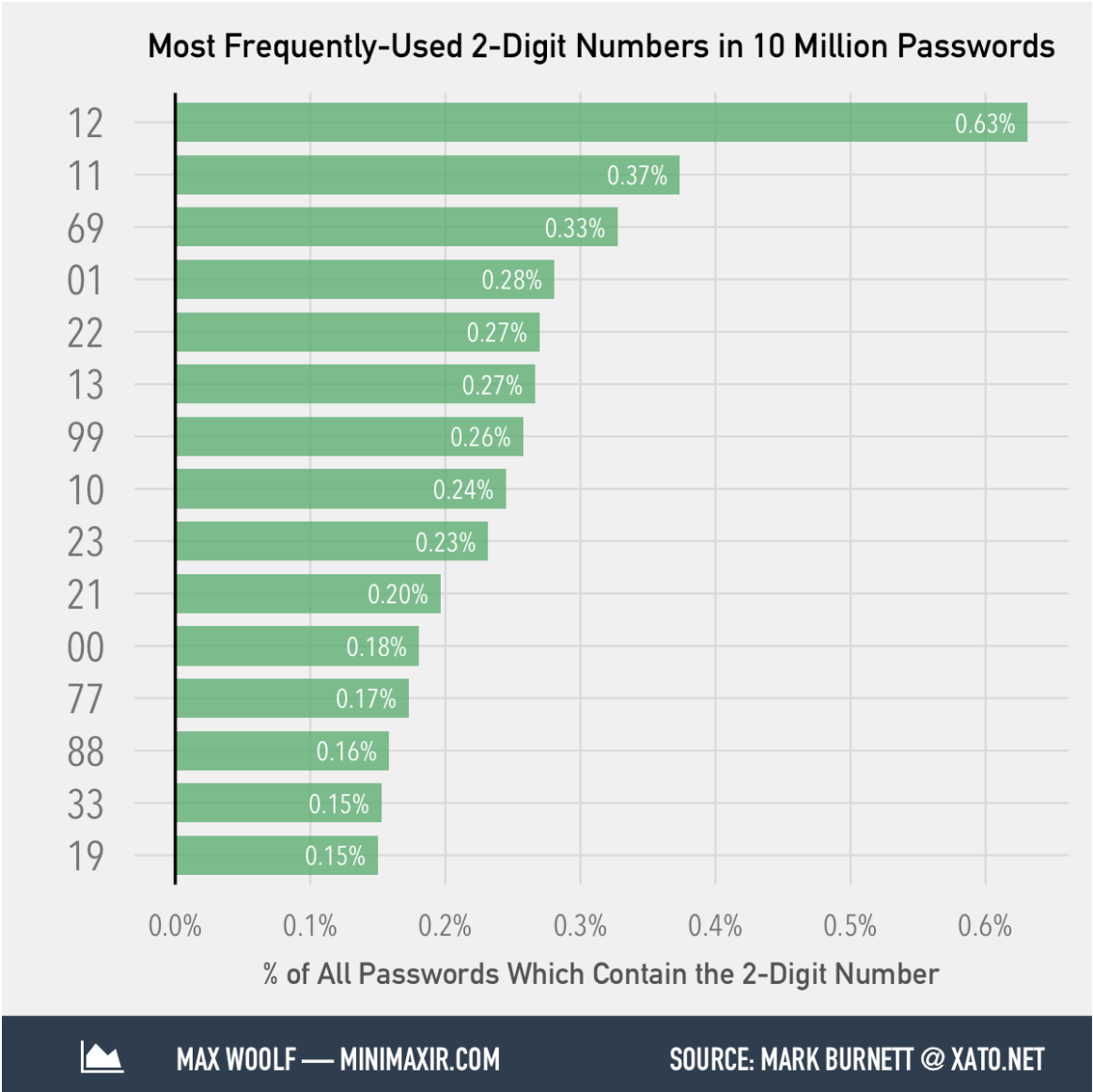
12 and 11, a sequential pattern and a repeating pattern

respectively, are by far the most-used 2-digit numbers. Many repeating patterns such as 22 and 99 are prominent. But why is 69 in third place? (besides the obvious non-family-friendly reason)
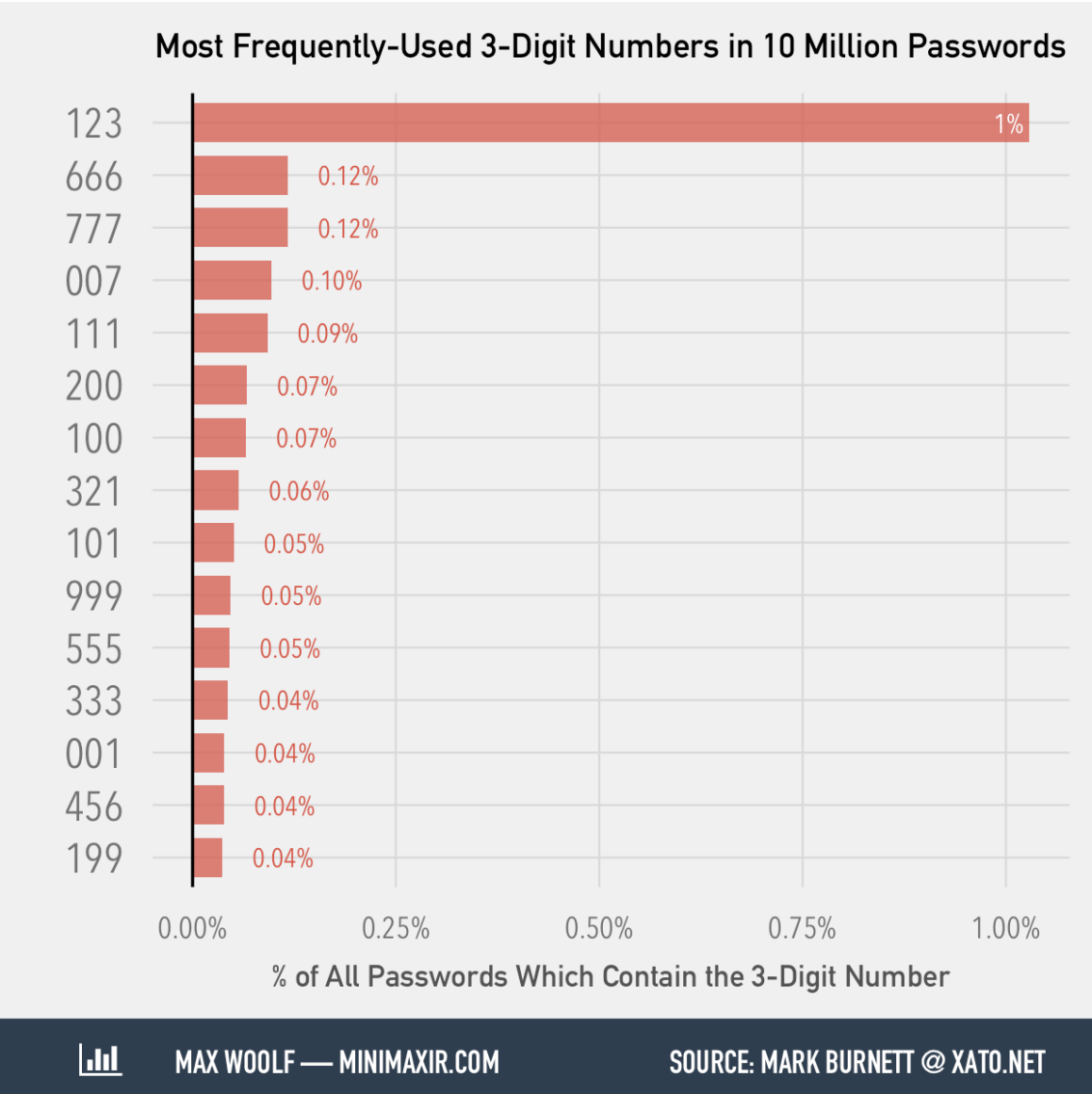
It may be helpful to look at a heat map of all possible 2-digit numbers to see if there are any observable patterns.
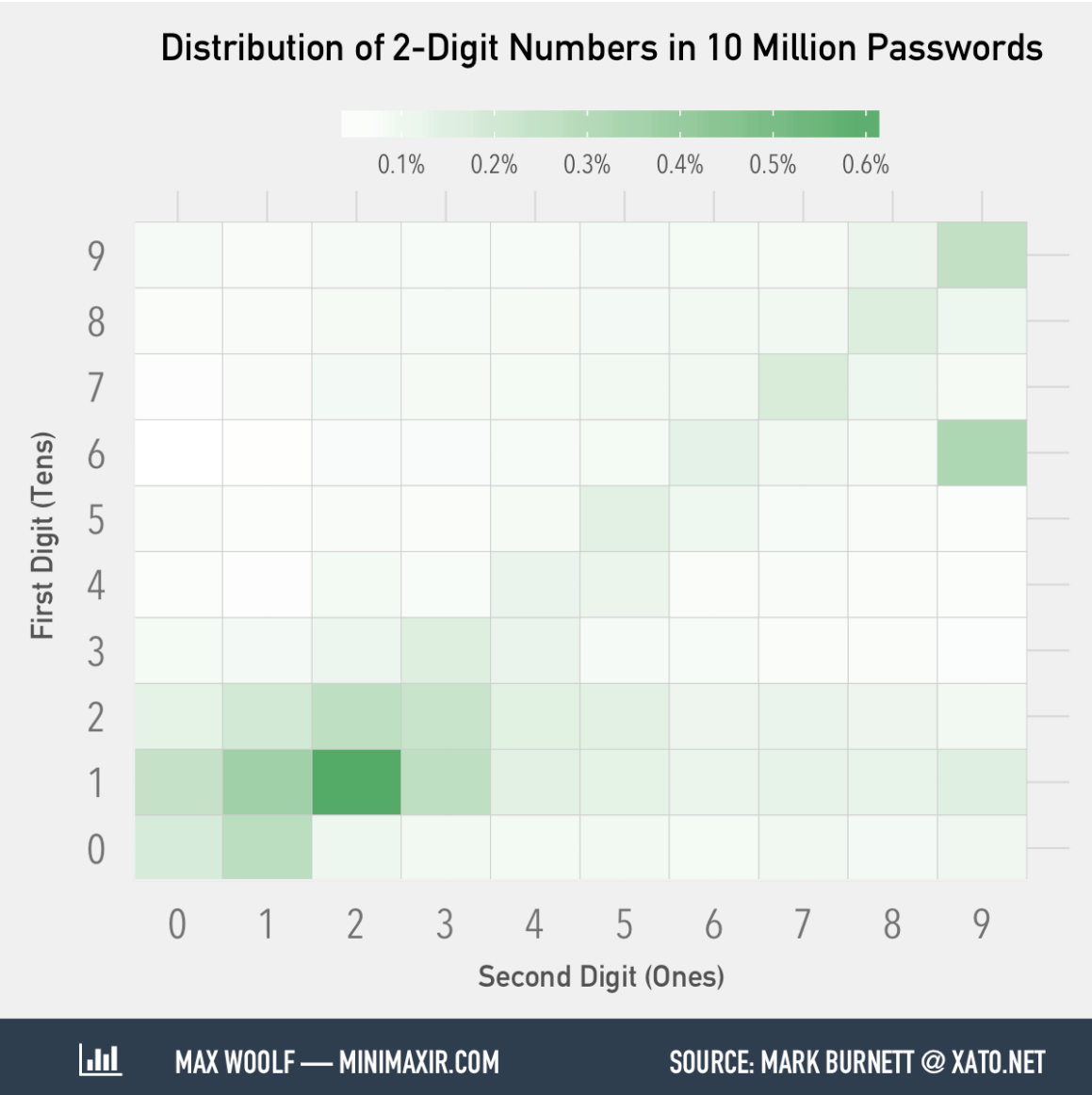
There are a couple distinct patterns: numbers beginning with a 1 or 2 are used the most frequently, and both repeating and sequential digits are used the most frequently.

Almost all 2-digit numbers outside of those patterns are unused (the exception is 69, of course) The intersection of both of these patterns is at 11/12, which is the reason both have high usage.

Do 3-digit numbers follow similar patterns? Here's a list of the most-used 3-digit numbers in passwords:

Most Frequently-Used 3-Digit Numbers in 10 Million Passwords

Number	%
123	1%
666	0.12%
777	0.12%
007	0.10%
111	0.09%
200	0.07%
100	0.07%
321	0.06%
101	0.05%
999	0.05%
555	0.05%
333	0.04%
001	0.04%
456	0.04%
199	0.04%

% of All Passwords Which Contain the 3-Digit Number

MAX WOOLF — MINIMAXIR.COM   SOURCE: MARK BURNETT @ XATO.NET

Yes and no. Here, there appear to be more instances of special numbers, such as 321 and 007 which deviate from the patterns above. Of note, 3-digit numbers ending in 00 appears as a new pattern.

This can be confirmed by looking at a faceted heat map for each possible combination.

Distribution of 3-Digit Numbers in 10 Million Passwords

By far the most popular pattern for a 3-digit number is a repetition pattern, followed by a sequential pattern (the sequential pattern is always located one tile up and two tiles right from the repetition pattern). There are very few outliers which deviate from this schema aside from the ones mentioned previously. (420 is not as significant of an outlier for 3-digit numbers as 69 is for 2-digit numbers)

The patterns of numbers in passwords can offer some insight to human psychology. However, if possible, I recommend you avoid using such patterns in your passwords since it introduces a vulnerability. It's a good idea to use a password manager instead, such as 1Password or KeePass, which offer advantages including the generation of both truly random and unique passwords.

All charts were made using R and ggplot2.

You can download the aggregate data used to create the charts in this Google Sheet.

If you liked this blog post, I have set up a Patreon to fund my machine learning/deep learning/software/hardware needs for my future crazy yet cool projects, and any monetary contributions to the Patreon are appreciated and will be put to good creative use.