DataGenetics

Home Blog About Us Work Content Contact Us

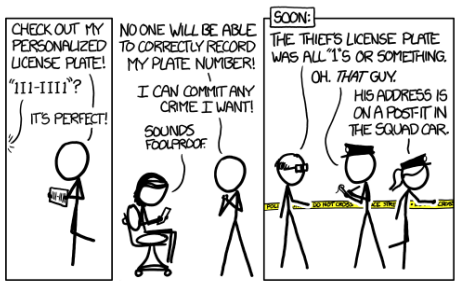
PIN analysis

A good friend of mine, Ian, recently forwarded me an internet joke. The headline was something like:

"All credit card PIN numbers in the World leaked"

The body of the message simply said 0000 0001 0002 0003 0004 …

Ian's messages made me chuckle. Then, later the same day, I read this XKCD cartoon. The merging of these two humorous topics created the seed for this article.



I love Randall's work. My favorite, to date, is this one. I have a signed copy of it on my office wall.

Like many of his creations, this cartoon is excellent at bifurcating readers; people read it, then either smile and chuckle, or stare blankly at it followed by a *"Huh? I don't get it!"* comment. Then you explain it, and get a reply *"Yeeaaaaaa…no, I still don't get it!"*

Esoteric humor in action.

You can be cool and buy his signed artwork too.

What is the least common PIN number?

There are 10,000 possible combinations that the digits 0-9 can be arranged to form a 4-digit pin code. Out of these ten thousand codes, which is the least commonly used?

Which of these pin codes is the least predictable?

Which of these pin codes is the most predictable?

If you were given the task of trying to crack a random credit card by repeatedly trying PIN codes, what order should you try guessing to maximize your chances of selecting the correct number in the shortest time?

If you had to make predication about what the least commonly used 4-digit PIN is, what would be your guess?

This tangentially relates to the XKCD cartoon. In Randall's cartoon, the perpetrator's plan backfired because his selected license plate was so unique that it was very memorable. What is the least memorable license plate? Ask any spy you know (snigger) what the best way to blend into a crowd is. Their answer will be not stand out, to appear "normal", and not be notable in any way.

People are notoriously bad at generating random passwords. I hope this article will scare you into being a little more careful in how you select your next PIN number.

Are you curious about what the least commonly used PIN number might be?

How about the most popular?

Read on …

DISCLAIMER

This article is **not** intended to be a hacker bible, or to be used as a utility, resource, or tool to help would-be thieves perform nefarious actions. I will only disclose data sufficient to make my points, and will try to avoid giving specific data outside of the obvious examples. I do not want to be an enabler for script-kiddies. Please do not email me asking for the database I used; if you do, you will be wasting your time as I'm not going to respond. I'm not going to sell, donate or release the source data – don't ask!

Source

Obviously, I don't have access to a credit card PIN number database. Instead I'm going to use a proxy. I'm going to use data condensed from released/exposed/discovered password tables and security breaches.

Soap Box – Password Database Exposures

Over the years, there have been numerous password table security breaches: Some very high profile, some low profile, but all embarrassing (and many exceedingly expensive; both in direct fines and indirect loss of business through erosion of trust and reputation).

Fool me once, well, no, even that's not really acceptable, but fool me twice … I'll go even further: **Any** developer who stores the password table of their database in clear text should be so mortified by this lack of security that they should not be sleeping at night until they fix it. Ignoring the fact that you should never have ever coded it this way, you have an obligation to learn from these past breaches.

If you work for a company and are knowledgeable that your customer database is *"protected"* by such lightweight security then run, don't walk, to your CEO/Presidents office, pound on the door and insist (s)he puts out a mandate to fix the matter with extreme prejudice. Don't leave until you get an affirmative response. Badger, badger then badger them again. Make yourself a proverbial thorn in their side.

I'm not trying to sell my services as a consultant here (though if you are interested, my rates are very reasonable compared to the cost of legal defense, potential FTC sanctions, class action suits, shareholder backlash, fines, loss of reputation and business …) There are plenty of security experts in the industry who can help you (if you need help filtering them and don't have referrals, someone who has CISSP qualifications is a good place to start).

Bottom line Security strengthens with layers, and the simple application of encryption on your database table can help protect your customer's data if this table is exposed. It does not defend against all possible attacks, but it does nothing but good things. What possible reason is there store things in clear-text?

Back to the data

By combining the exposed password databases I've encountered, and filtering the results to just those rows that are exactly four digits long [0-9] the output is a database of all the four digit character combinations that people have used as their account passwords.

Given that users have a free choice for their password, if users select a four digit password to their online account, it's not a stretch to use this as a proxy for four digit PIN codes.

The Data

I was able to find almost *3.4 million* four digit passwords. Every single one of the of the 10,000 combinations of digits from 0000 through to 9999 were represented in the dataset.

The most popular password is 1234 …

… it's *staggering* how popular this password appears to be. Utterly *staggering* at the lack of imagination …

… nearly 11% of the 3.4 million passwords are 1234 *!!!*

The next most popular 4-digit PIN in use is 1111 with over 6% of passwords being this.

In third place is 0000 with almost 2%.

A table of the top 20 found passwords in shown at the right. A staggering 26.83% of all passwords could be guessed by attempting these 20 combinations!

(Statistically, with 10,000 possible combination, if passwords were uniformly randomly distributed, we would expect the these twenty passwords to account for just 0.2% of the total, not the 26.83% encountered)

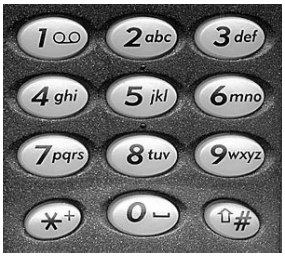
Looking more closely at the top few records, all the usual suspects are present 1111 2222 3333 … 9999 as well as 1212 and (snigger) 6969 .

It's not a surprise to see patterns like 1122 and 1313 occurring high up in the list, nor 4321 or 1010 .

2001 makes an appearance at #19. 1984 follows not far behind in position #26, and James Bond fans may be interested to know 0007 is found between the two of them in position #23 (another variant 0070 follows not much further behind at #28).

	PIN	Freq
#1	1234	10.713%
#2	1111	6.016%
#3	0000	1.881%
#4	1212	1.197%
#5	7777	0.745%
#6	1004	0.616%
#7	2000	0.613%
#8	4444	0.526%
#9	2222	0.516%
#10	6969	0.512%
#11	9999	0.451%
#12	3333	0.419%
#13	5555	0.395%
#14	6666	0.391%
#15	1122	0.366%
#16	1313	0.304%
#17	8888	0.303%
#18	4321	0.293%
#19	2001	0.290%
#20	1010	0.285%

The first "puzzling" password I encountered was 2580 in position #22. What is the significance of these digits? Why should so many people select this code to make it appear so high up the list?



Then I realized that 2580 is a straight down the middle of a telephone keypad!

(Interestingly, this is very compelling evidence confirming the hypothesis that a 4-digit password list is a great proxy for a PIN number database. If you look at the numeric keypad on a PC-keyboard you'll see that 2580 is slightly more awkward to type on the PC than a phone because the order of keys on a keyboard is the inverted. Cash machines and other terminals that take credit cards use a phone style numeric pads. It appears that many people have an easy to type/remember PIN number for their credit card and are re-using the same four digits for their online passwords, where the *"straight down the middle"* mnemonic no longer applies).

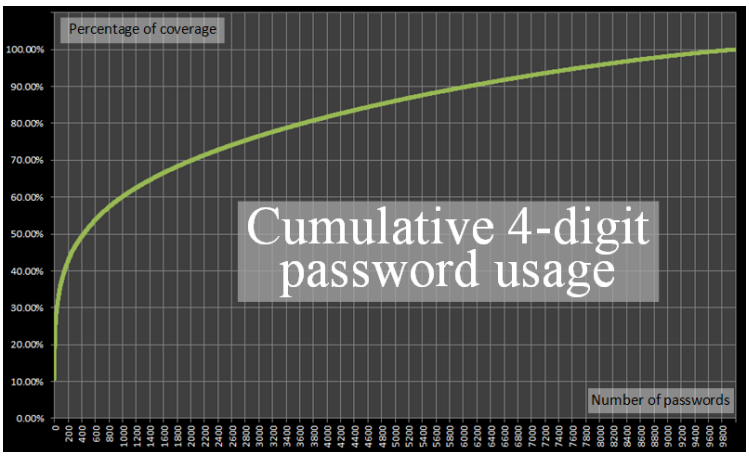
(Another fascinating piece of trivia is that people seem to prefer even numbers over odd, and codes like 2468 occur higher than a odd number equivalent, such as 1357).

Cumulative Frequency

As noted above, the more popular password selections dominate the frequency tables. The most popular PIN code of 1234 is more popular than the *lowest 4,200 codes combined!*

That's right, you might be able to crack over 10% of all codes with one guess! Expanding this, you could get 20% by using just five numbers!

Below is a cumulative frequency graph:



Statistically, *one third* of all codes can be guessed by trying just 61 distinct combinations!

The 50% cumulative chance threshold is passed at just 426 codes (far less than the 5,000 that a random uniformly distribution would

predict). Paranoid yet?

Bottom of the pile

OK, we've investigated most frequently used PINS and found they tend to be predictable and easy to remember, let's turn for a second to the bottom of the pile.

What are the least "interesting" (least used) PINS?

In my dataset the answer is 8068 with just 25 occurrences in 3.4 million (this equates to 0.000744%, far, far fewer than random distribution would predict, and five orders of magnitude behind the most popular choice).

To the right are the twenty least popular 4-digit passwords encountered.

Warning Now that we've learned that, historically, 8068 is (was?) the least commonly used password 4-digit PIN, please don't go out and change yours to this! Hackers can read too! They will also be promoting 8068 up their attempt trees in order to catch people who read this (or similar) articles.

Check out about the Nash Equilibrium

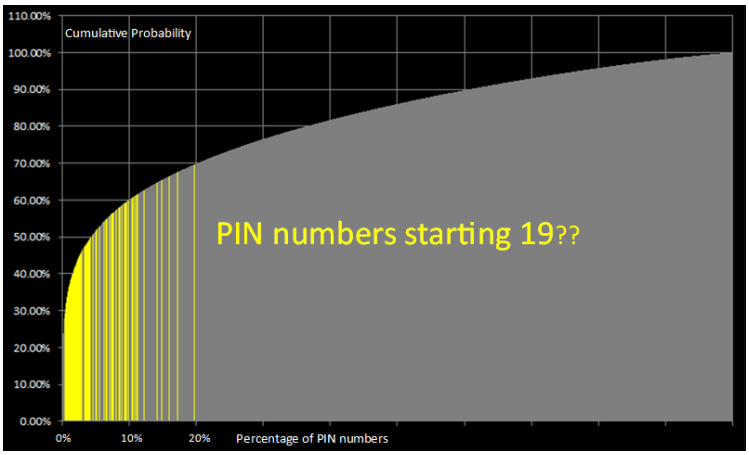
	PIN	Freq
#9980	8557	0.001191%
#9981	9047	0.001161%
#9982	8438	0.001161%
#9983	0439	0.001161%
#9984	9539	0.001161%
#9985	8196	0.001131%
#9986	7063	0.001131%
#9987	6093	0.001131%
#9988	6827	0.001101%
#9989	7394	0.001101%
#9990	0859	0.001072%
#9991	8957	0.001042%
#9992	9480	0.001042%
#9993	6793	0.001012%
#9994	8398	0.000982%
#9995	0738	0.000982%
#9996	7637	0.000953%
#9997	6835	0.000953%
#9998	9629	0.000953%
#9999	8093	0.000893%
#10000	8068	0.000744%

Memorable Years

Many of the high frequency PIN numbers can be interpreted as years, *e.g.* 1967 1956 1937 … It appears that many people use a year of birth (or possibly an anniversary) as their PIN. This will certainly help them remember their code, but it greatly increases its predictability.

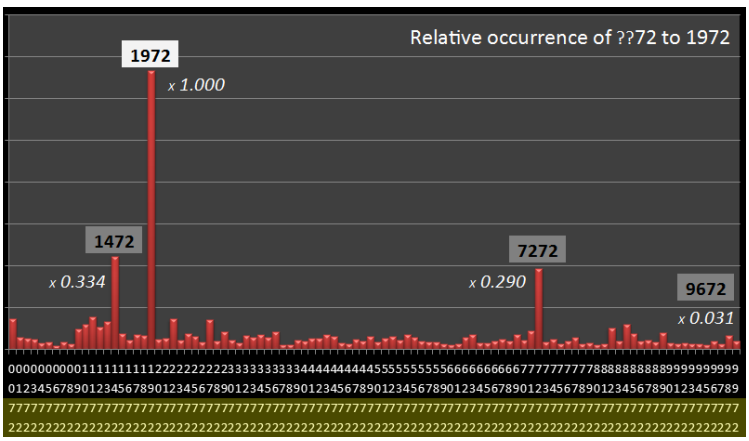
Just look at the stats: Every single 19?? combination can be found in the *top fifth* of the dataset!

Below is a plot of this in graphical format. In this chart, each yellow line represents a PIN number that starts 19??



If all the passwords were uniformly distributed, there should be no significant difference between the frequency of occurrence of, *for instance*, 1972 and any other PIN ending in seventy two ??72 . However, as we shall see, this is not the case at all.

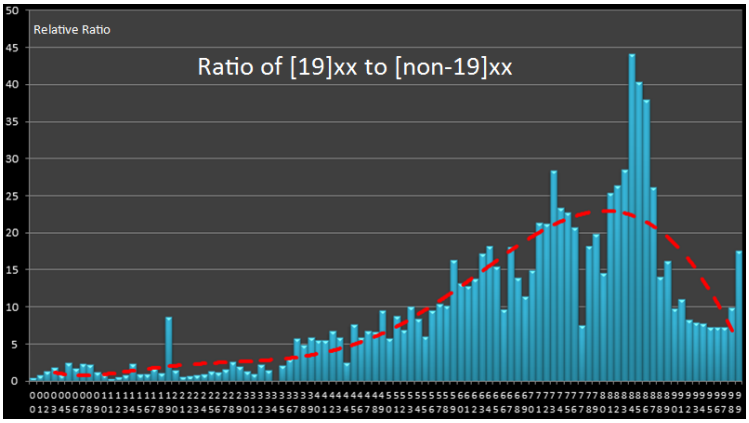
1972 occurs in ordinal position #76 (with a frequency 0.099363%). Here's a histogram for the occurrences of all ??72 probabilities.



You can clearly see the spike at 1972 (with smaller spikes at 7272 and 1472)

If you calculate the ratio of the peak of 1972 to the average of all the other ??72 PINS you get the ratio of 22:1

PINS starting with 19?? are much more likley to occur. Of course, it's not just 1972. Here is plot of the ratio of 19 to non-19 for all hundred combinations. Along the x-axis are all the combinations of last two digits –XX, and for each of these the ratio of the 19XX to average of all the other ??XX occurrences has been calculated. Here's the chart:



It's a pretty good approximation for a demographic chart! (suggested by the red-dashed trend line) which would probably allow a fair estimation of the ages (years of birth) of the people using the various websites. (Of course, hackers invert this strategy and use the age of a target to try and give information to guess a user's PIN. Looking at this graph, this might give them up to a *40x* advantage!)

Just about all the ratios are above 1.0. The noteable exceptions are ??34 and ??00 (which are easy to explain, since the massive popularity of 1234 and 0000 dwarf 1934 and 1900 respectively). Simiarly 33 44 55 66 … are lower than expected as the quad codes like 3333 mask out even the 1933 boost.

There are also spikes in the graph corresponding to the popular PINS of 1919 1984 and 1999

Patterns in data

I love pretty ways to graphically vizualize data. Pictures really do paint thousands of words.

Another interesting way to visualize the PIN data is in this grid plot of the distribution. In this heatmap, the x-axis depicts the left two digits from [00] to [99] and the y-axis depicts the right two digits from [00] to [99]. The bottom left is 0000 and the top right is 9999 .

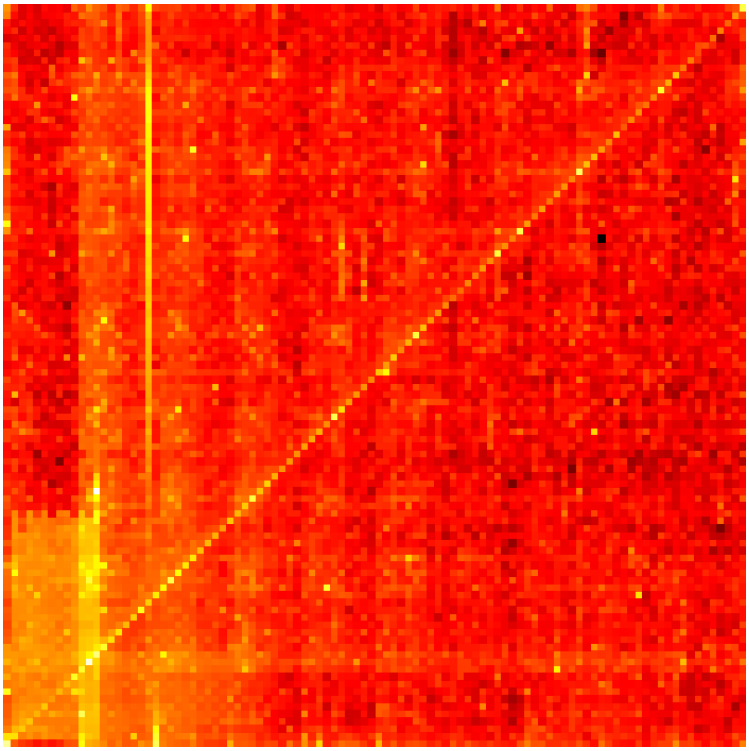
Color is used to represent frequency. The higher frequency occurences are yellow to white hot, and the lower frequency occurences are red, through dark red to black.

Geek Note The scaling is logarithmic.

You could look at this plot all day!

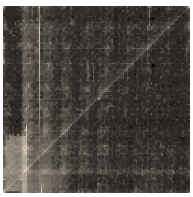
The bright line for the leading diagonal shows the repeated couplets that people love to use for their PIN numbers 0000 0101 0202 … 5454 5555 5656 … 9898 9999 .

Every eleventh dot on the leading diagonal is brighter corresponding to the quad numbers *e.g.* 4444 5555 . Here is a larger scale version:



Interesting things

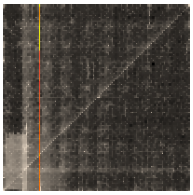
There are so many interesting things to learn from this heatmap. Here are just a couple:



The first is the interesting harmonics of shading (seen here more easily in a gray scale plot).

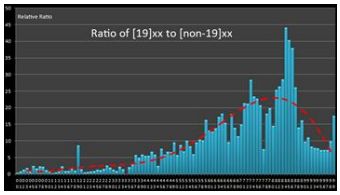
You can make out a *"grid pattern"* in the plot.

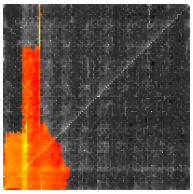
The lighter areas corresponding to couplets of numbers that are close to each other. For some reason, people don't like to select pairs of numbers that have larger numerical gaps between them. Combinations like 45 and 67 occur much more frequently than things like 29 and 37



Here we see the line corresponding to 19XX . The intensity the dots relates to the chart we plotted earlier

There are a large number of codes starting with 19, especially towards the higher end.

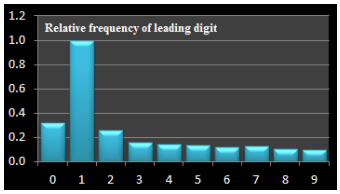


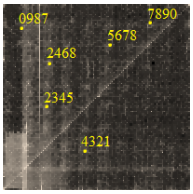


There is a strong bias towards the lower left quadrant. People love to start their PIN numbers with 0, and even more so with the digit 1.

The chart on the right shows the relative frequency of the first digit of 4-digit pin codes.

As you can see, the digit 1 dominates (and it's not all down to the 19XX phenomenon.)





Little bright specs dot the plot in places corresponding to numerical runs (both ascending and descending) such as 2345 , 4321 and 5678 .

I've highlighted just a couple on the plot to the left.

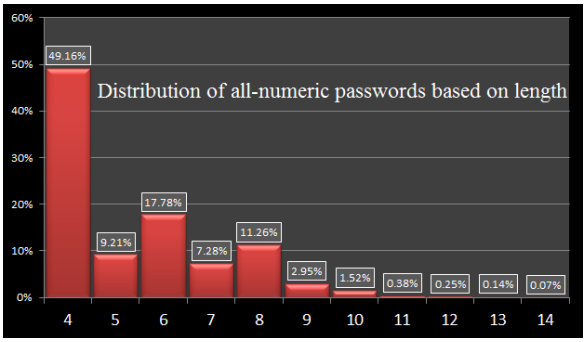
Jumps in steps of two are also visible *e.g.* 2468

Repeated-pair couplets of numbers are very common, such as XYXY

The hundred sets of repeating couplet pairs represent a staggering 17.8% of all observed PIN numbers.

More than four

The purpose of this posting was to investigate patterns and frequency of four digit PIN numbers. However, the database I collected also has *all-numeric* password of different lengths. It's worth taking a quick look at these too.



I found close to 7 million all-numeric passwords. Approximately half of these were the four-digit codes we've just examined.

Six digit codes are the next most popular length, followed eight.

I hope, hope that the people who have passwords of nine digits long are not using their Social Security Numbers!

Below are the top 20 passwords for the various lengths, along with their share of their same-size namespace.

#	5		6		7		8		9		10	
	PSWD	%	PSWD	%	PSWD	%	PSWD	%	PSWD	%	PSWD	%
#1	12345	22.802%	123456	11.684%	1234567	3.440%	12345678	11.825%	123456789	35.259%	1234567890	20.431%
#2	11111	4.484%	123123	1.370%	7777777	1.721%	11111111	1.326%	987654321	3.661%	0123456789	2.323%
#3	55555	1.769%	111111	1.296%	1111111	0.637%	88888888	0.959%	123123123	1.587%	0987654321	2.271%
#4	00000	1.258%	121212	0.623%	8675309	0.465%	87654321	0.815%	789456123	1.183%	1111111111	2.087%
#5	54321	1.196%	123321	0.591%	1234321	0.220%	00000000	0.675%	999999999	0.825%	1029384756	1.293%
#6	13579	1.112%	666666	0.577%	0000000	0.188%	12341234	0.569%	147258369	0.591%	9876543210	0.971%
#7	77777	0.618%	000000	0.521%	4830033	0.158%	69696969	0.348%	741852963	0.455%	0000000000	0.942%
#8	22222	0.454%	654321	0.506%	7654321	0.154%	12121212	0.320%	111111111	0.425%	1357924680	0.479%
#9	12321	0.412%	696969	0.454%	5201314	0.128%	11223344	0.293%	123454321	0.413%	1122334455	0.441%
#10	99999	0.397%	112233	0.417%	0123456	0.124%	12344321	0.275%	123654789	0.378%	1234512345	0.402%
#11	33333	0.338%	159753	0.283%	2848048	0.124%	77777777	0.262%	147852369	0.356%	1234554321	0.380%
#12	00700	0.261%	292513	0.250%	7005425	0.120%	99999999	0.223%	111222333	0.304%	5555555555	0.259%
#13	90210	0.244%	131313	0.235%	1080413	0.111%	22222222	0.219%	963852741	0.255%	1212121212	0.244%
#14	88888	0.217%	123654	0.228%	7895123	0.107%	55555555	0.205%	321654987	0.253%	9999999999	0.231%
#15	38317	0.216%	222222	0.212%	1869510	0.102%	33333333	0.176%	420420420	0.241%	2222222222	0.219%
#16	09876	0.185%	789456	0.209%	3223326	0.100%	44444444	0.165%	007007007	0.227%	7777777777	0.206%
#17	44444	0.179%	999999	0.194%	1212123	0.096%	66666666	0.160%	135792468	0.164%	3141592654	0.195%
#18	98765	0.169%	101010	0.190%	1478963	0.088%	11112222	0.140%	397029049	0.158%	3333333333	0.186%
#19	01234	0.160%	777777	0.188%	2222222	0.085%	13131313	0.131%	012345678	0.154%	7894561230	0.165%
#20	42069	0.154%	007007	0.186%	5555555	0.082%	10041004	0.127%	123698745	0.152%	1234567891	0.161%

Some interesting observations (and a little speculation)

For five digit passwords, users appear to have *even less* imagination in selecting their codes (22.8% select 12345). All the usual suspects occur, but a new addition is the puerile addition in position #20 of the concatenation of 420 and 69.

For six digit password, again 696969 appears highly. Also of note is 159753 (a "X" mark over the numeric keypad). James Bond returns with 007007.

For seven digits, the standby of 1234567 is a much lower frequency (though still the top). I speculate that this is because many people may be using their telephone number (without area code) as a seven digit password. Telephone numbers are fairly distinct, and already memorized, so when a seven digit code is needed, they spring to mind easily. The higher frequency of usage of telephone numbers reduces the need to use imagination (or lack thereof) and select something else.

Is Jenny there? The fouth most popular seven digit password is 8675309 (It's a popular 80's song).

Eight digit passwords are just as expected. Lots of pattern, and lots of repetition.

Common nine digit passwords also follow patterns and repetition. 789456123 appears as an easy *"Along the top, middle and bottom of the keypad"* 147258369 is related in the vertical direction (and other variants appear high up). Again we get a 420 moment with 420420420, and also the shaken, not stirred, but repeated 007007007 returns.

Interestingly for ten digits 1029384756 appears (alternating ascending/descending digits), as well as the odd/even 1357924680.

Hurrah for math! In position #17 of the ten digit password list we get 3141592654 (The first few digits of *Pi*)

Conclusions

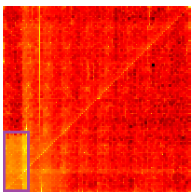
If you are a **developer** , **tester** or **executive** I hope you are sufficiently paranoid that you will *immediately* check to see that your systems do not store sensitive information, like passwords, unencrypted. The entire reason I was able to perform this analysis is because ~~dumb~~ stupid and lazy coders stored information in clear text. Your lazyness has the potential to impact millions.

If you are a **consumer** and your recognize any of the numbers I've used in this article to be your passwords/pins I hope you apply common sense and immediately change them to something a little less predictable. Alternatively, you could be lazy and not change things (In that case, at least the only person you are harming with this apathy is yourself.)

Updates

Since publishing this article, it's been brought to my attention that, of course, in addition to anniversary years, many people encapsulate dates in the format MMDD (such as birthdays …) for their PIN codes.

This clearly explains the lower left corner where, if you look at the heatmap, there is a huge contrast change at the height of around 30-31 (the number of days in a month), extending to 12 on the x-axis. (Thanks to **zero79** for first pointing this out).

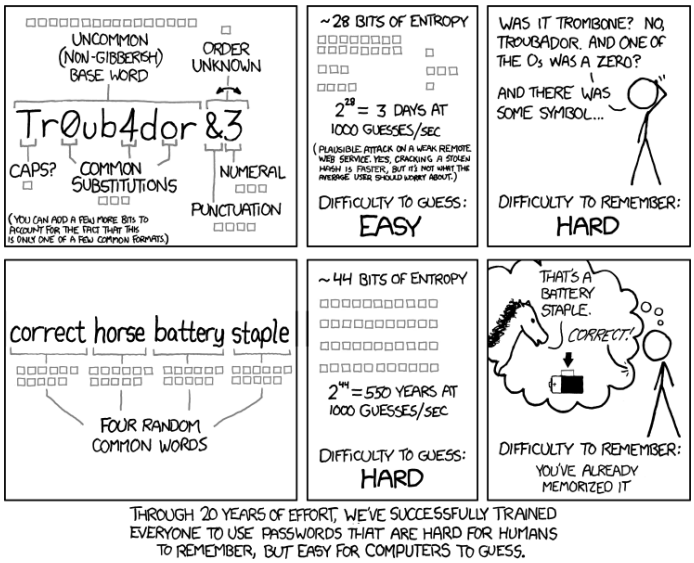


Many people also asked the significance of 1004 in the four character PIN table. This comes from Korean speakers. When spoken, "1004" is *cheonsa* (cheon = 1000, sa=4).

"Cheonsa" also happens to be the Korean word for *Angel*.

Another XKCD cartoon

It only seems appropriate to end with another XKCD cartoon. This one is Password Strength



Did Not Connect: Potential Security Issue

Firefox detected a potential security threat and did not continue to **datagenetics.com** because this website requires a secure connection.

Learn more...

Go Back (Recommended) Advanced...

You can find a complete list of all the articles here. Click here to receive email alerts on new articles.

Buy me a coffee