



Insights from common password data sets

By: Banaag, Estoque, and Rivera

Problem



Securing Passwords

Finding ways to secure passwords from hackers



Preventing data breaches

Finding ways to secure passwords of developers



Preventing getting black mailed by hackers

Stopping hackers from using access to our personal data to threaten us

Data Set 1

10000 Most Common Passwords

	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	num_syllables
0	123456	6	0	6	0	0	0	0	1
1	password	8	8	0	0	8	0	2	2
2	12345678	8	0	8	0	0	0	0	1
3	qwerty	6	6	0	0	6	0	1	3
4	123456789	9	0	9	0	0	0	0	1

Data Set 2

Top 200 common passwords by
country 2021

	country_code	country	Rank	Password	User_count	Time_to_crack	Global_rank	Time_to_crack_in_seconds
0	au	Australia	1	123456	308483	< 1 second	1.0	0
1	au	Australia	2	password	191880	< 1 second	5.0	0
2	au	Australia	3	lizottes	98220	3 Hours	NaN	10800
3	au	Australia	4	password1	86884	< 1 second	16.0	0
4	au	Australia	5	123456789	75856	< 1 second	2.0	0

Merged Data Set

Formula for bits of entropy = $\log(b^l)/\log(2)$

As stated in the source code for password_strength (https://github.com/kolypto/py-password-strength/blob/master/password_strength/stats.py), a strength of 0.33 refers to a weak password. Strong passwords have a strength of at least 0.66.

	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	pattern_number	pattern_qwerty	min_ascii	max_ascii
0	123456	6	0	6	0	0	0	0	1	0	49	54
1	password	8	8	0	0	8	0	2	0	0	97	119
2	12345678	8	0	8	0	0	0	0	1	0	49	56
3	qwerty	6	6	0	0	6	0	1	0	1	101	121
4	123456789	9	0	9	0	0	0	0	1	0	49	57

bits_of_entropy	strength
15.509775	0.172331
22.458839	0.249543
24.000000	0.266667
15.509775	0.172331
28.529325	0.316992

Evaluation of Password Strengths

Medium to Strong Strength Passwords

Medium is 0.33 to 0.66 (340 Passwords)
Strong is 0.66 to 1 (3 Passwords)

Medium Passwords

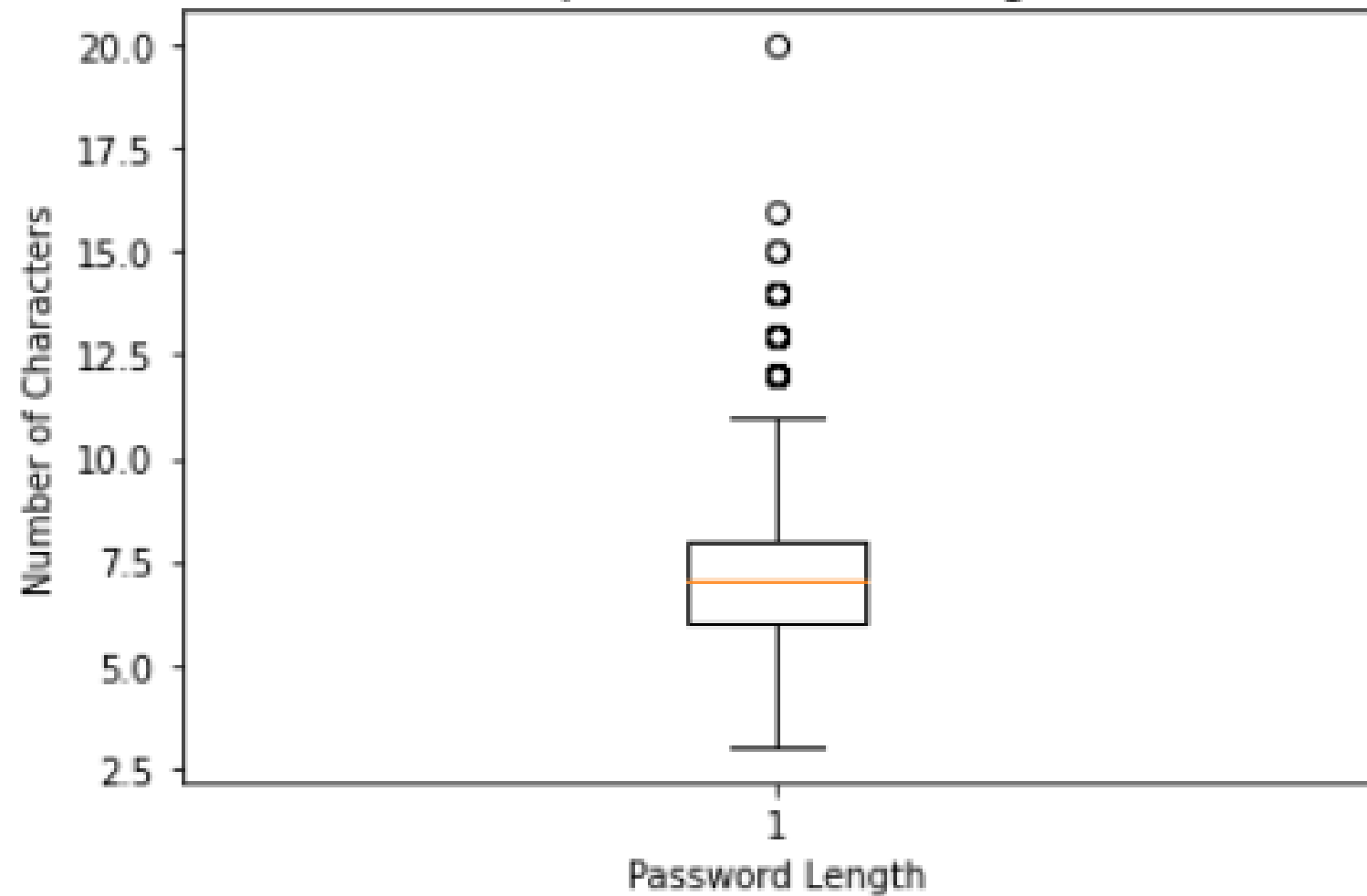
	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	bits_of_entropy	strength
11	wrongpassword	13	13	0	0	13	0	3	41.209025	0.517160
20	qwertyuiop	10	10	0	0	10	0	4	33.219281	0.392327
23	1234567890	10	0	10	0	0	0	0	33.219281	0.392327
119	q1w2e3r4t5	10	5	5	0	5	0	1	33.219281	0.392327
373	1q2w3e4r5t	10	5	5	0	5	0	1	33.219281	0.392327

Strong Passwords

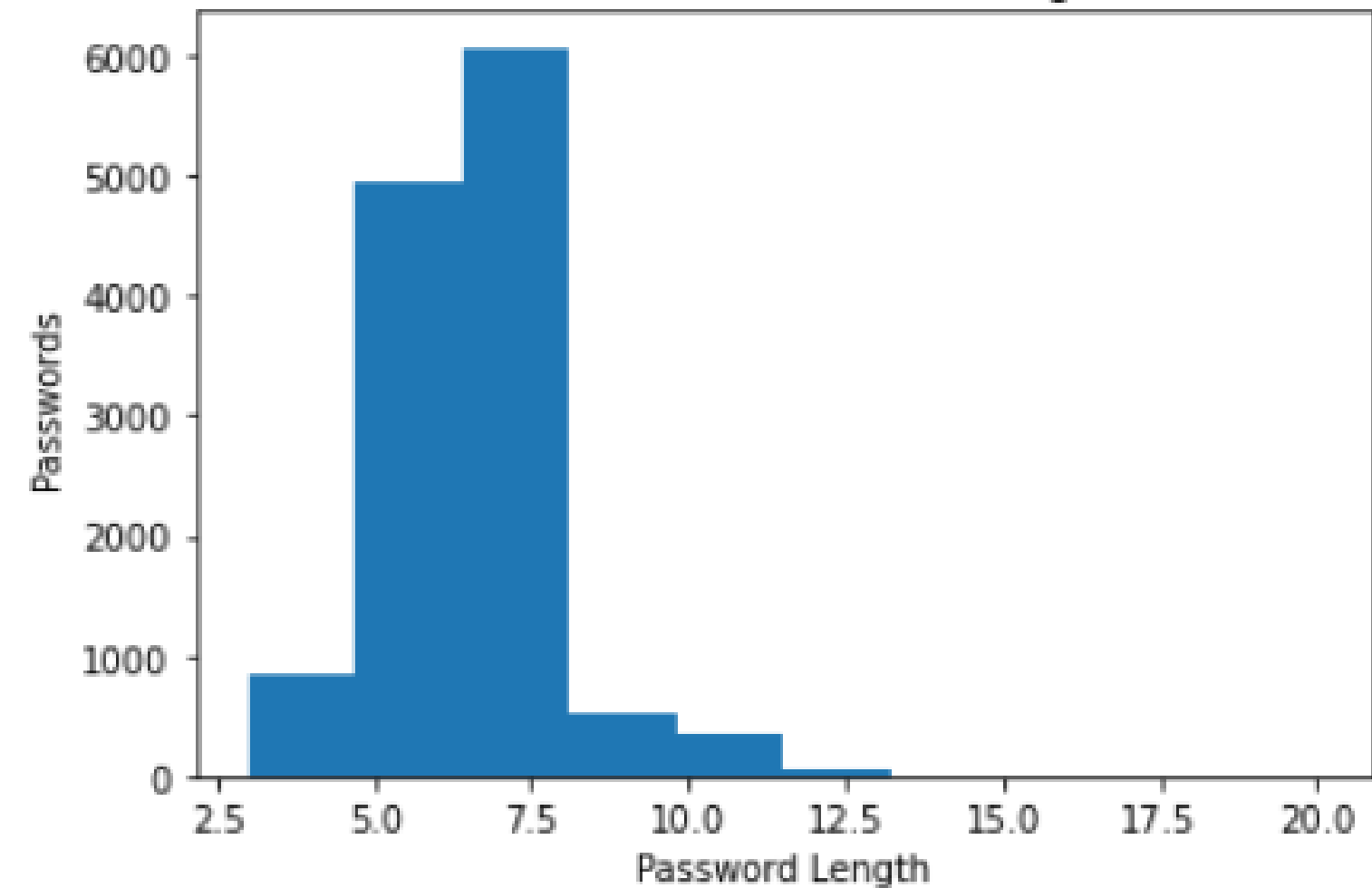
	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	bits_of_entropy	strength
2200	Mailcreated5240	15	11	4	1	10	0	5	55.506596	0.680041
4759	PolniyPizdec0211	16	12	4	2	10	0	4	59.207035	0.712366
12193	1q2w3e4r5t6y7u8i9o0p	20	10	10	0	10	0	4	86.438562	0.868639

Examination of Password Lengths

Boxplot of Password Length



Distribution of Password Length



Examination of Short and Long Passwords

The researchers separated their dataset into two dataframes with short (less than 8 characters) and long (at least 8 characters) passwords.

Longer passwords tend to be stronger. So, the researchers thought of finding ways to encourage making longer passwords.

Long Passwords

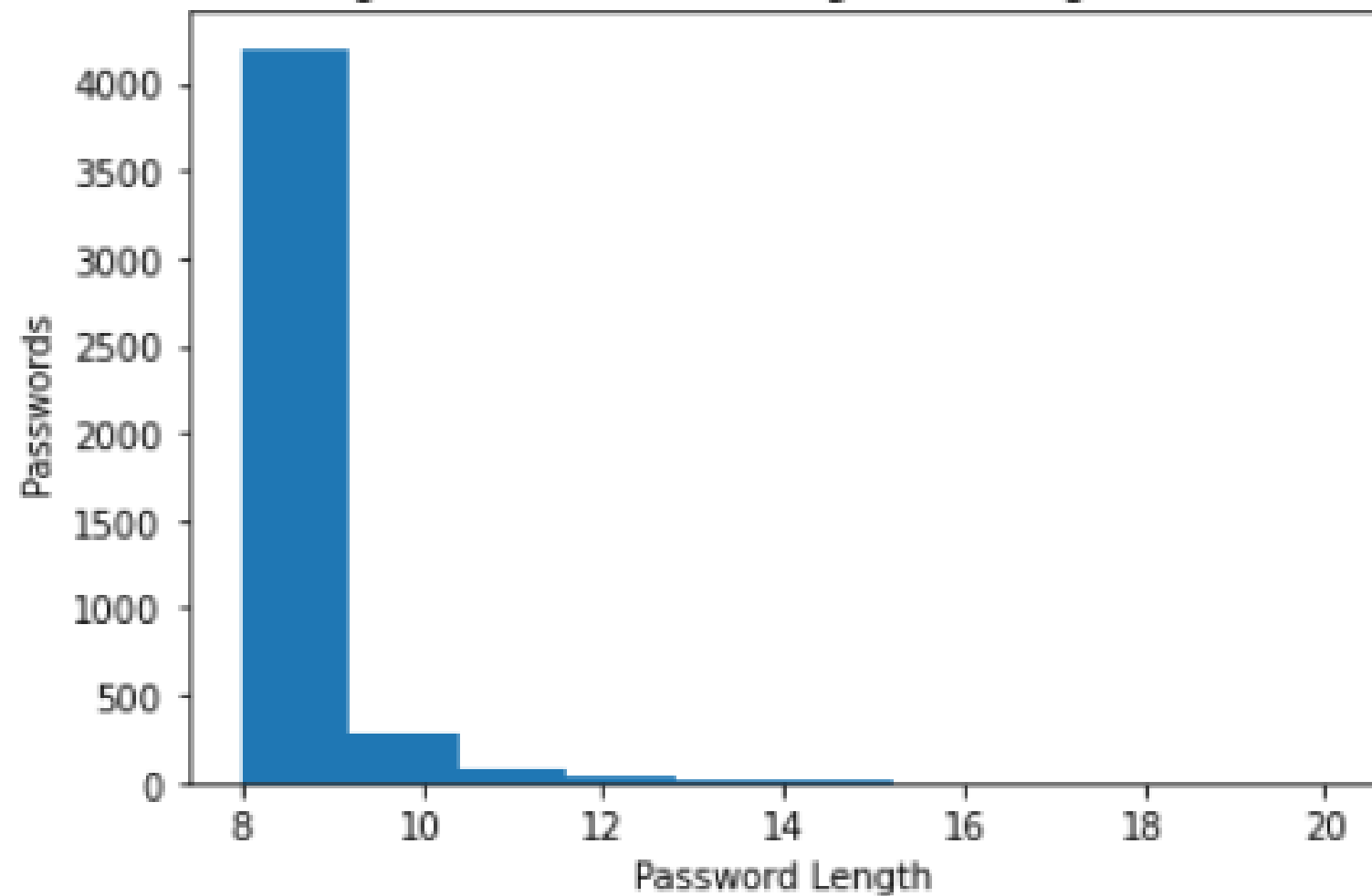
	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	bits_of_entropy	strength
1	password	8	8	0	0	8	0	2	22.458839	0.249543
2	12345678	8	0	8	0	0	0	0	24.000000	0.266667
4	123456789	9	0	9	0	0	0	0	28.529325	0.316993
11	wrongpassword	13	13	0	0	13	0	3	41.209025	0.517160
13	football	8	8	0	0	8	0	3	20.679700	0.229774

Short Passwords

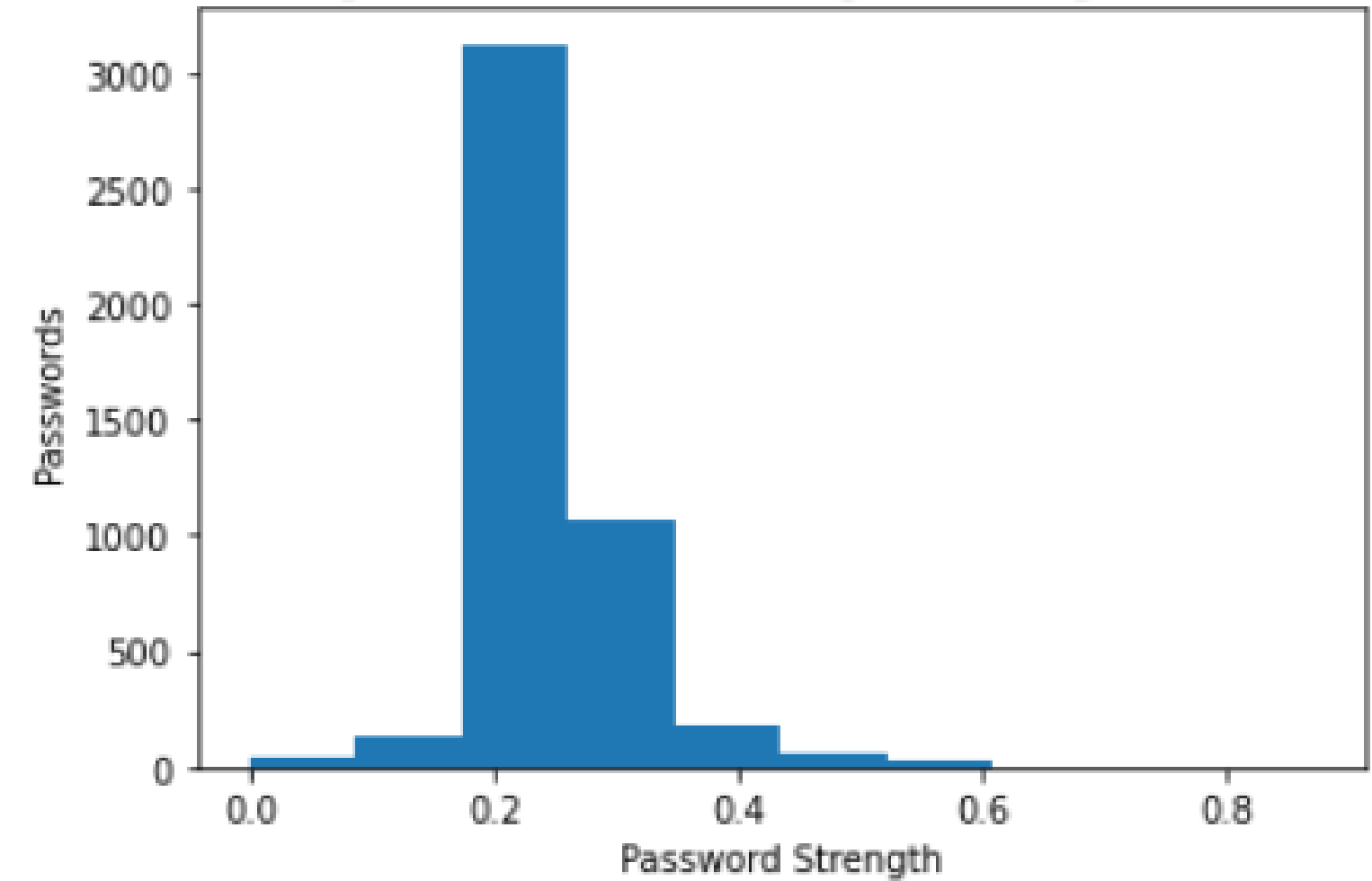
	password	length	num_chars	num_digits	num_upper	num_lower	num_special	num_vowels	bits_of_entropy	strength
0	123456	6	0	6	0	0	0	0	15.509775	0.172331
3	qwerty	6	6	0	0	6	0	1	15.509775	0.172331
5	12345	5	0	5	0	0	0	0	11.609640	0.128996
6	1234	4	0	4	0	0	0	0	8.000000	0.088889
7	111111	6	0	6	0	0	0	0	0.000000	0.000000

Examination of Long Passwords

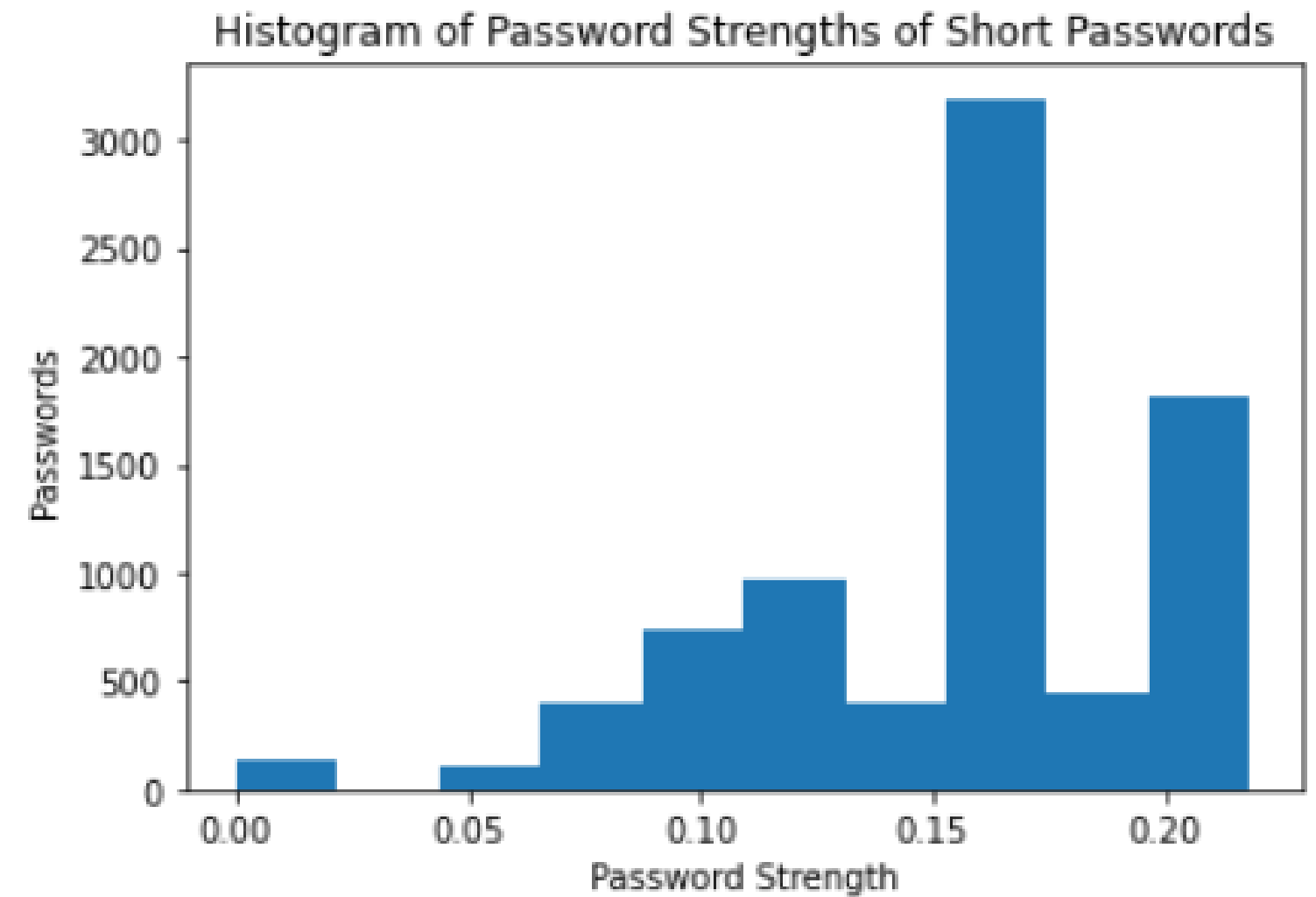
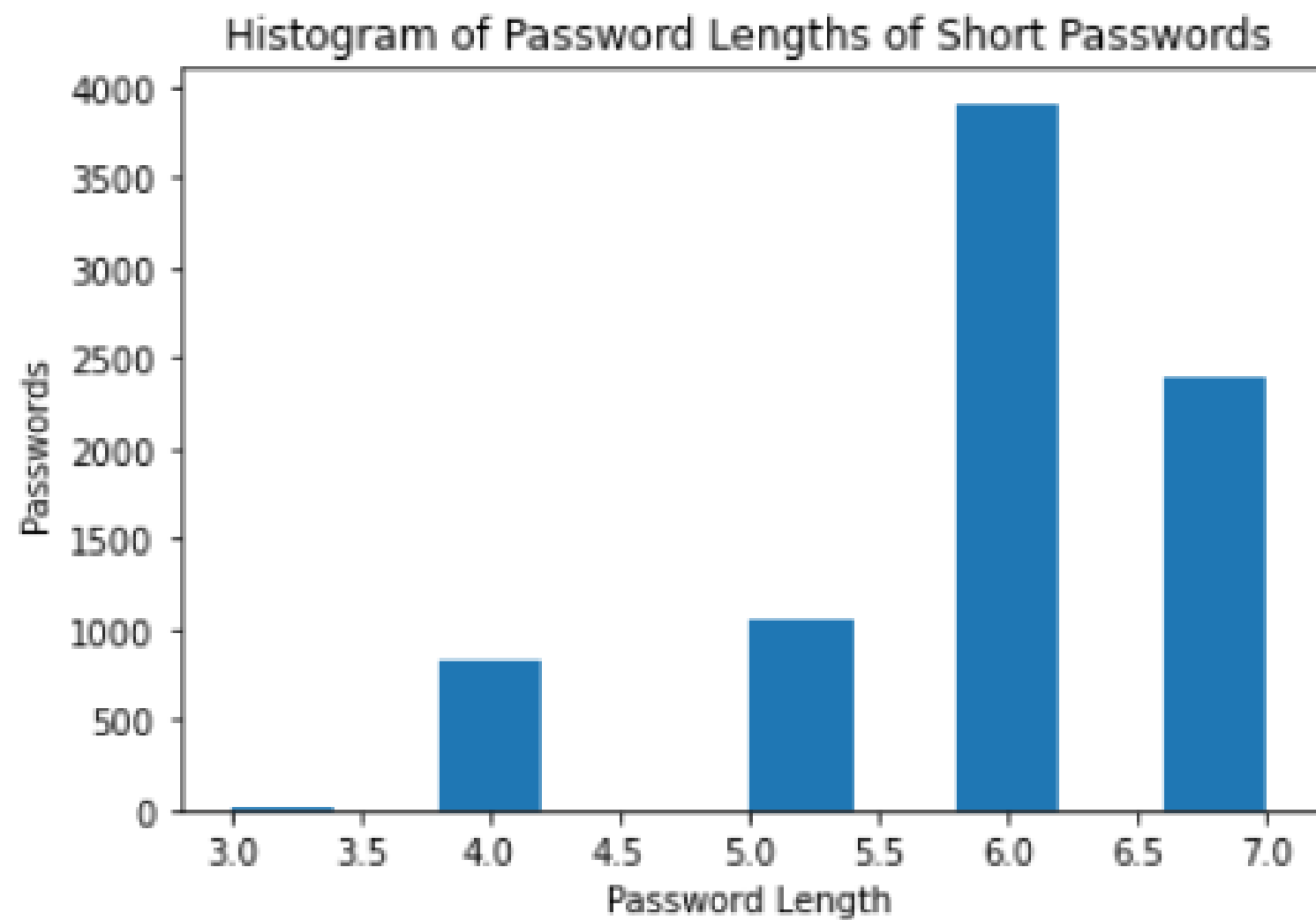
Histogram of Password Lengths of Long Passwords



Histogram of Password Strengths of Long Passwords



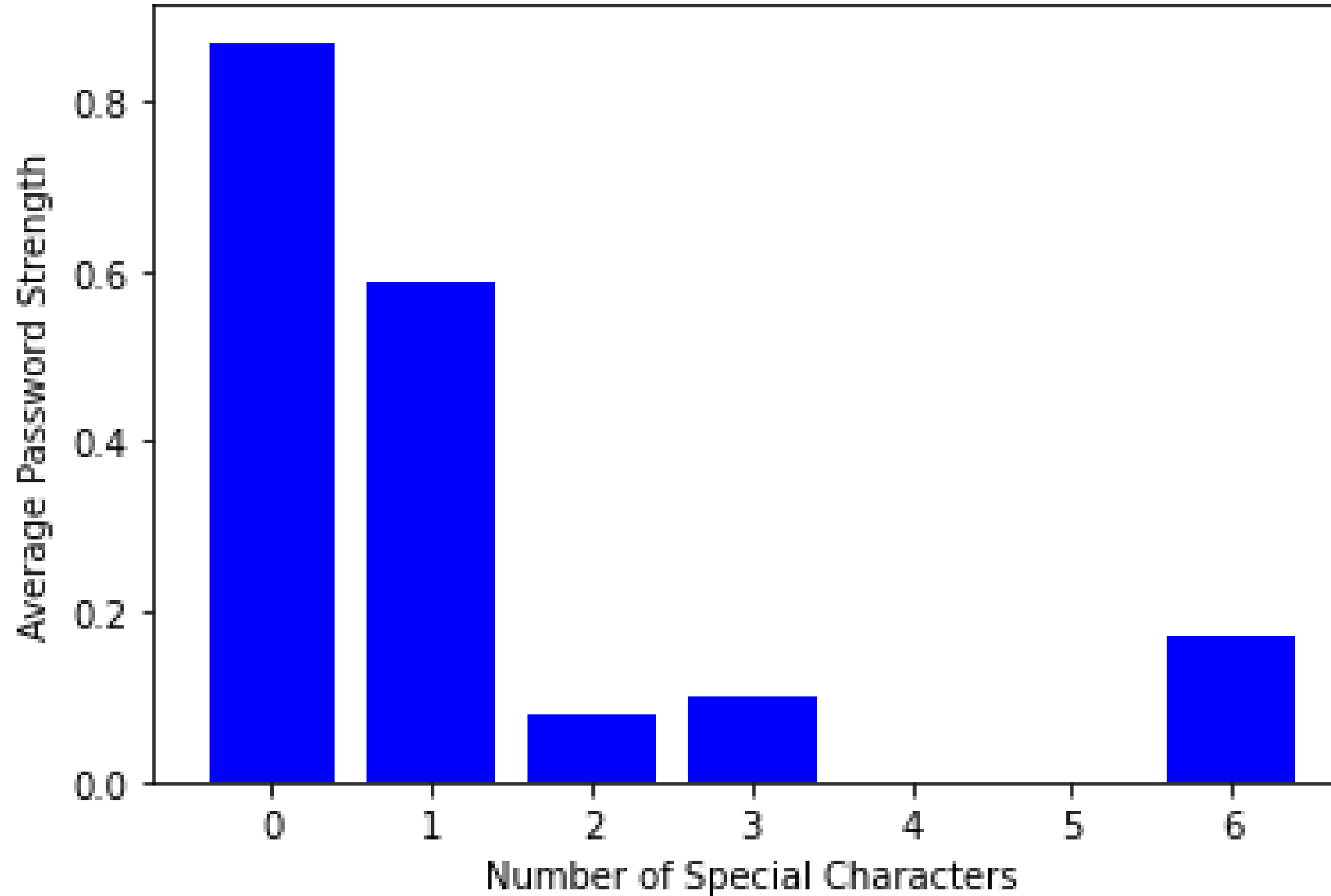
Examination of Short Passwords



Password Strength and Special Characters

The researchers found out that having many special characters in a password does not necessarily make the password strong.

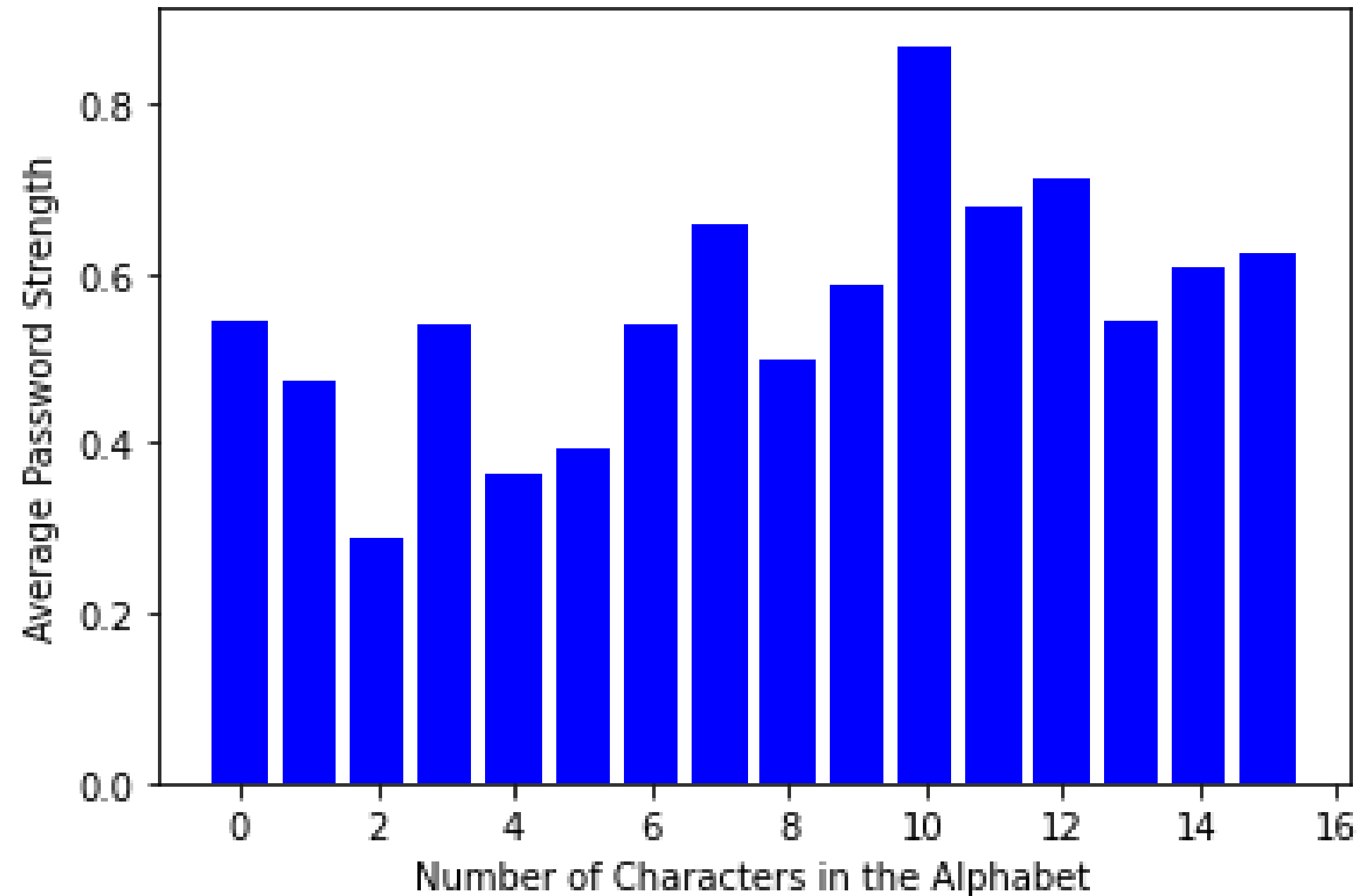
Barplot of Average Password Strengths based on Special Characters



Password Strength and Characters in the Alphabet

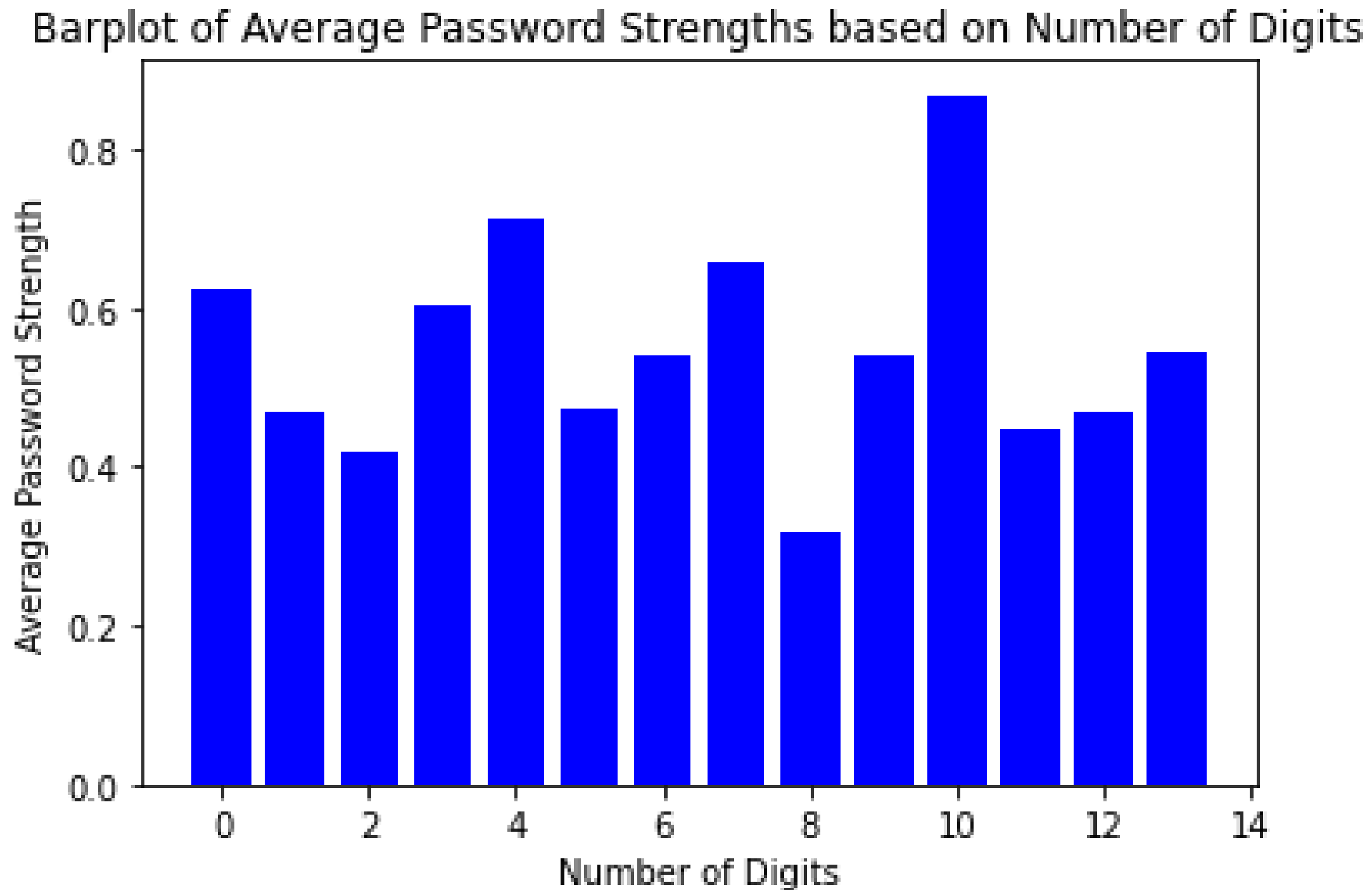
The researchers found out that having many characters from the alphabet in a password has a positive correlation to password strength. This phenomenon may be explained by the idea that characters in the alphabet are used to get make long passwords.

Barplot of Average Password Strengths based on Characters in the Alphabet



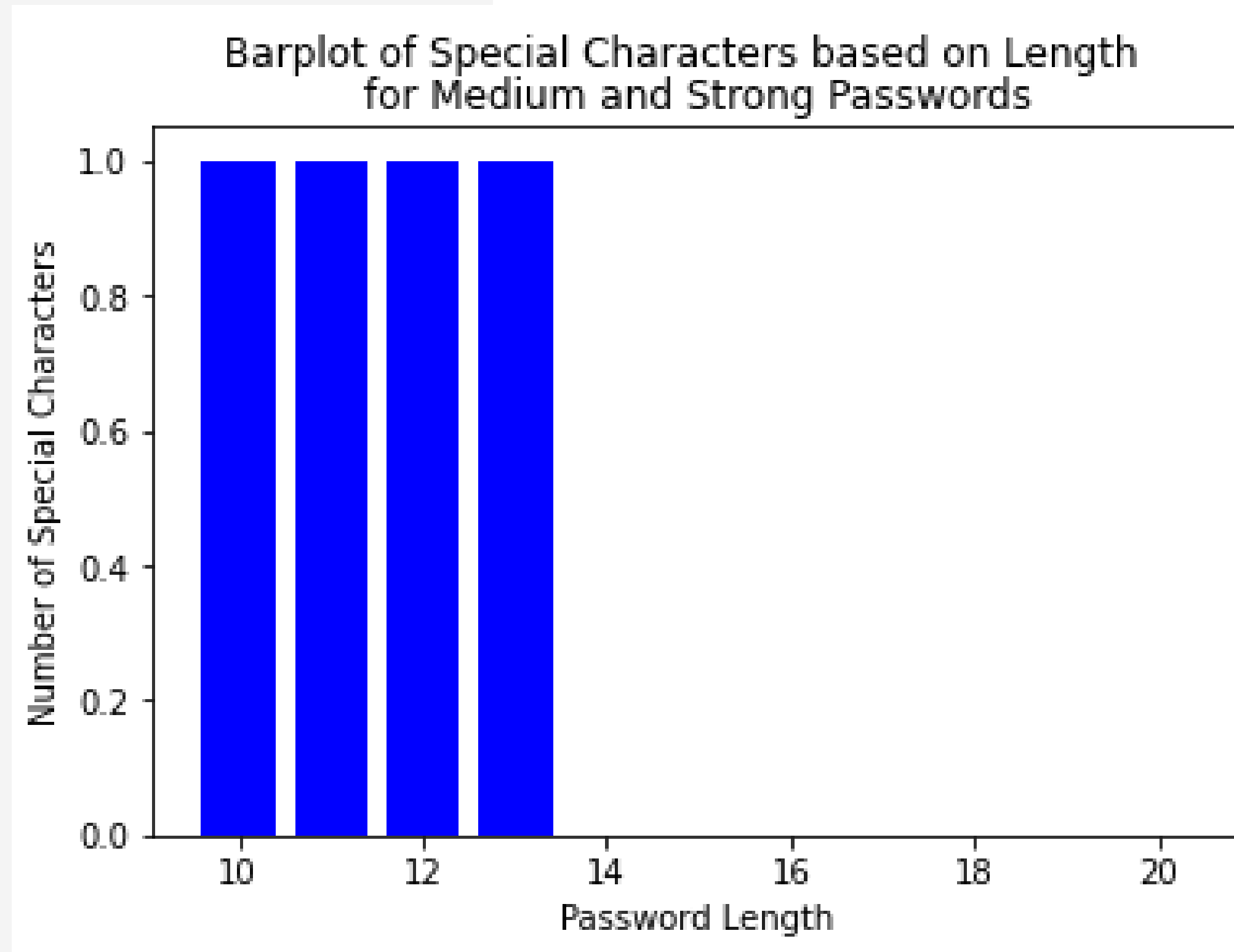
Password Strength and Digits

The researchers found out that having many digits in a password has a neglible effect on its strength.



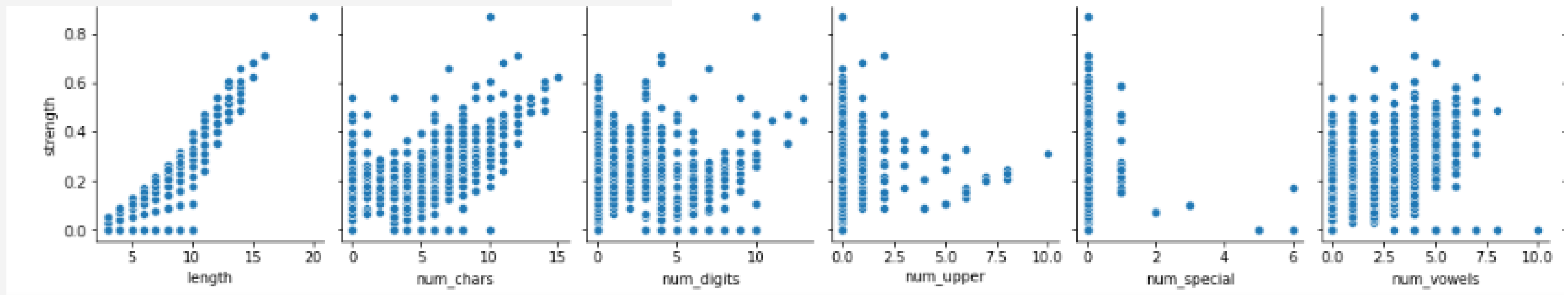
Special Characters and Password Length

The researchers found out that medium and strong passwords have 1 special character if they have any special characters.



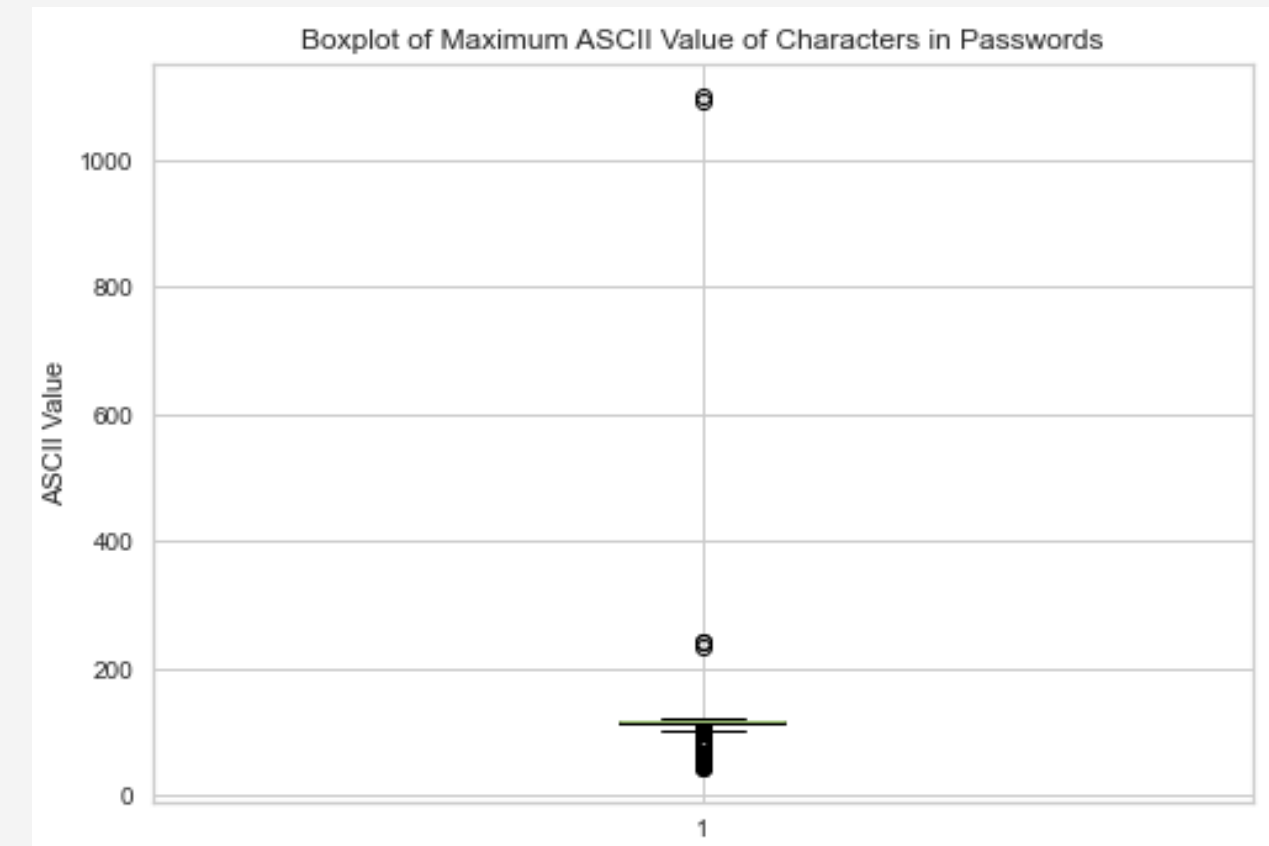
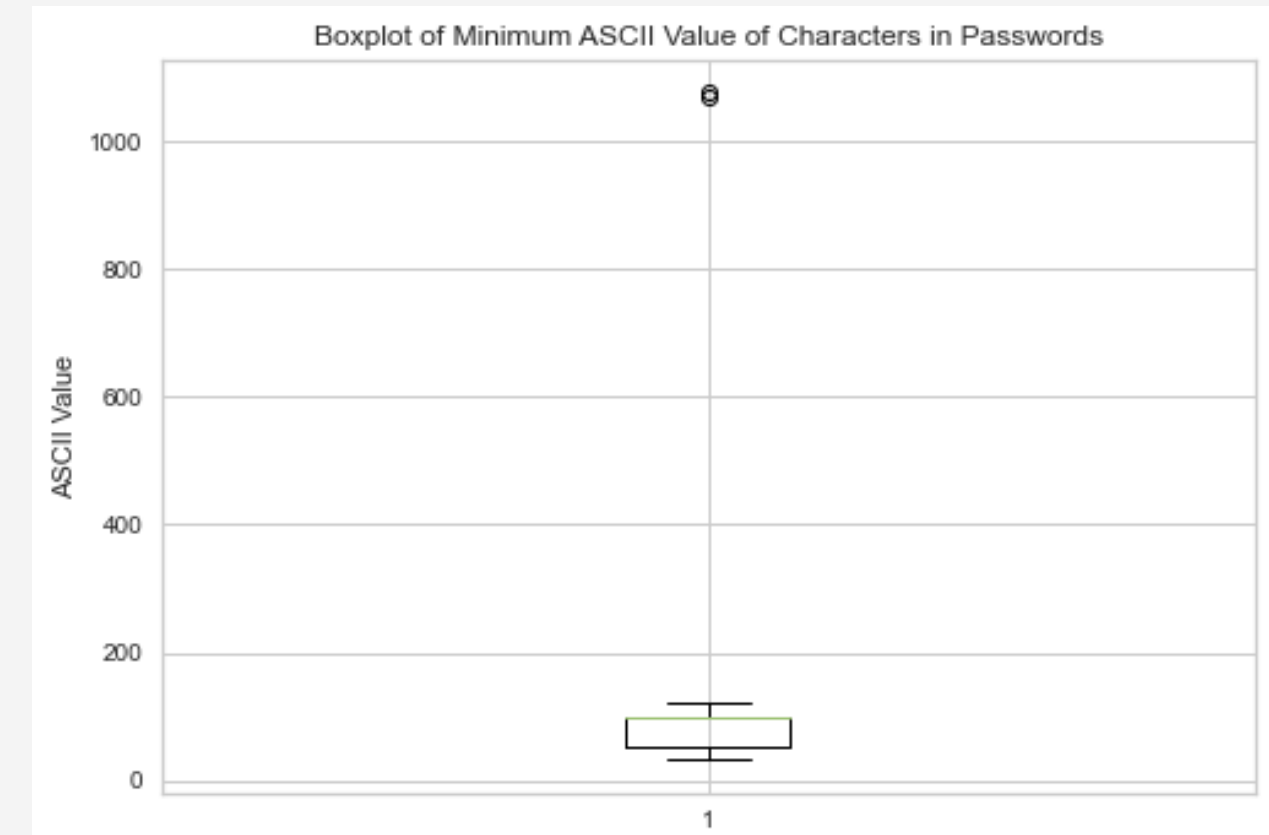
Correlogram of Features

The researchers found out that the features that are most correlated to the strength of passwords are length, number of characters in the alphabet, number of uppercase characters, and number of digits. They believe that the correlation of the number of vowels is merely a result of having many characters from the alphabet.



Evaluation of ASCII Codes

- Minimum of 32 (space key)
- Maximum
 - 117 (u) for most characters
 - 1100 (b) for entire dataset



Recommendations

Recommended Password Requirements

- Require uppercase characters
- Require digits
- Recommend 8 characters
- Do not put limits on passwords if applicable

Other Recommendations

- More datasets for studying passwords
- Use proprietary hashing techniques for storing passwords
- Use characters with ASCII codes above 117

**Passwords such as
1 year Latər**