pwned: exposure from data breaches

gaurav and ken

the price of a data breach

- September 22, 2016

- September 22, 2016

    Yahoo! reveals that 500M accounts have been compromised.

- September 22, 2016

  Yahoo! reveals that 500M accounts have been compromised.

- December 14, 2016

- September 22, 2016

  Yahoo! reveals that 500M accounts have been compromised.

- December 14, 2016

  Yahoo! announces that data had been stolen from nearly 1 billion user accounts in a different breach.

- September 22, 2016

  Yahoo! reveals that 500M accounts have been compromised.

- December 14, 2016

  Yahoo! announces that data had been stolen from nearly 1 billion user accounts in a different breach.

- Verizon, in the middle of negotiating a takeover of *Yahoo!*, values the breaches at $350M, less than a quarter-cent per breached account.

- September 22, 2016

  Yahoo! reveals that 500M accounts have been compromised.

- December 14, 2016

  Yahoo! announces that data had been stolen from nearly 1 billion user accounts in a different breach.

- Verizon, in the middle of negotiating a takeover of *Yahoo!*, values the breaches at $350M, less than a quarter-cent per breached account.

- September 22, 2016

  Yahoo! reveals that 500M accounts have been compromised.

- December 14, 2016

  Yahoo! announces that data had been stolen from nearly 1 billion user accounts in a different breach.

- Verizon, in the middle of negotiating a takeover of *Yahoo!*, values the breaches at $350M, less than a quarter-cent per breached account.

- October 4, 2017

- September 22, 2016

  Yahoo! reveals that 500M accounts have been compromised.

- December 14, 2016

  Yahoo! announces that data had been stolen from nearly 1 billion user accounts in a different breach.

- Verizon, in the middle of negotiating a takeover of *Yahoo!*, values the breaches at $350M, less than a quarter-cent per breached account.

- October 4, 2017

  Yahoo! revises the estimate for the latter breach to 3 billion.

the consequence of a low price of a breach

Data
  - YouGov

Data
- YouGov
    - $n = 5,000$

Data
  - YouGov
      - $n = 5,000$
      - Matched sampling. Median Error is 1%.

Comparison Between YouGov and CPS 2018

	cps	yg	diff
Age			
18 to 25	0.14	0.13	0.01
26 to 35	0.18	0.18	0.00
36 to 50	0.25	0.23	0.02
51 to 65	0.25	0.26	-0.01
66 to 80+	0.19	0.18	0.01
Sex			
Male	0.48	0.51	-0.03
Female	0.52	0.49	0.03

Comparison Between YouGov and CPS 2018

	cps	yg	diff
Race			
White alone	0.78	0.64	0.14
Black or African American alone	0.13	0.12	0.01
American Indian and Alaska Native alone	0.01	0.01	0.00
Educational Attainment			
No high school diploma	0.11	0.07	0.04
High school or equivalent	0.29	0.33	-0.04
Some college, less than 4-yr degree	0.28	0.31	-0.03
Bachelor's degree or higher	0.32	0.29	0.03

Data

- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.

Data

- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.
- Have I Beeb Pwned (HIBP):

Data

- YouGov
    - $n = 5,000$
    - Matched sampling. Median Error is 1%.

- Have I Beeb Pwned (HIBP):
    - Clearinghouse of information about data breaches

Data
- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.
- Have I Beeb Pwned (HIBP):
  - Clearinghouse of information about data breaches
  - It currently carries data from 293 breaches covering 278 unique domains

Data

- ## YouGov
    - $n = 5,000$
    - Matched sampling. Median Error is 1%.

- ## Have I Beeb Pwned (HIBP):
    - Clearinghouse of information about data breaches
    - It currently carries data from 293 breaches covering 278 unique domains
    - Information about 5,235,843,322 breached accounts

Data

- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.

- Have I Beeb Pwned (HIBP):
  - Clearinghouse of information about data breaches
  - It currently carries data from 293 breaches covering 278 unique domains
  - Information about 5,235,843,322 breached accounts

- Query HIBP with the email ID on the file.

Data

- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.

- Have I Beeb Pwned (HIBP):
  - Clearinghouse of information about data breaches
  - It currently carries data from 293 breaches covering 278 unique domains
  - Information about 5,235,843,322 breached accounts

- Query HIBP with the email ID on the file.

Empirical Caveat

Data
- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.

- Have I Beeb Pwned (HIBP):
  - Clearinghouse of information about data breaches
  - It currently carries data from 293 breaches covering 278 unique domains
  - Information about 5,235,843,322 breached accounts

- Query HIBP with the email ID on the file.

Empirical Caveat
- Most breaches are never discovered

Data
- YouGov
  - $n = 5,000$
  - Matched sampling. Median Error is 1%.

- Have I Beeb Pwned (HIBP):
  - Clearinghouse of information about data breaches
  - It currently carries data from 293 breaches covering 278 unique domains
  - Information about 5,235,843,322 breached accounts

- Query HIBP with the email ID on the file.

Empirical Caveat
- Most breaches are never discovered

- We can only estimate the *lower bound*.

The horror! The horror!

The horror! The horror!

- 14,979 breaches are associated with the 5,000 emails on file.

The horror! The horror!

- 14,979 breaches are associated with the 5,000 emails on file.

- Or on average, there are three breaches per person. The median is also three.

The horror! The horror!

- 14,979 breaches are associated with the 5,000 emails on file.

- Or on average, there are three breaches per person. The median is also three.

- And at least 82.84% of Americans' accounts have been breached at least once.
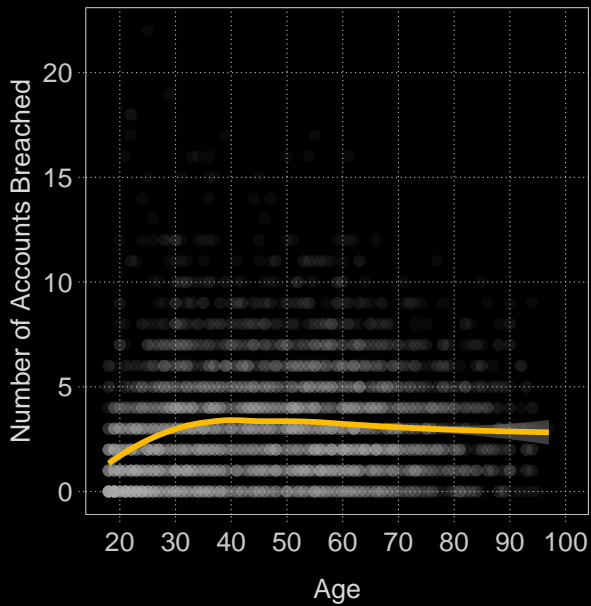
The horror! The horror!

- 14,979 breaches are associated with the 5,000 emails on file.

- Or on average, there are three breaches per person. The median is also three.

- And at least 82.84% of Americans' accounts have been breached at least once.

who is pwned?

the age of the pwned

Number of Breaches by Age

	Dependent variable:
	Number of Breaches
ns(age, 2)1	2.23***
	(.20)
ns(age, 2)2	−.90***
	(.21)
Constant	2.01***
	(.09)
Observations	5,000
R^2	.03
Adjusted R^2	.03
Note:	*p<0.1; **p<0.05; ***p<0.01

the race of the pwned

Average Number of Accounts Breached
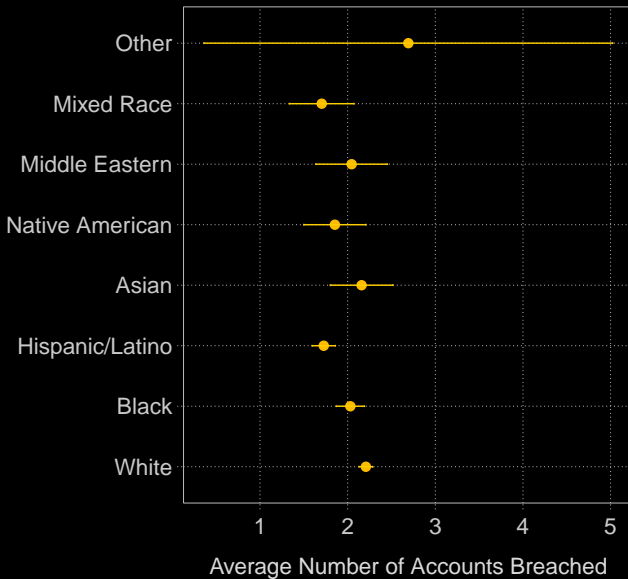
Number of Breaches by Race/Ethnicity

	Dependent variable:
	Number of Breaches
Black	.04
	(.12)
Hispanic/Latino	−.62***
	(.10)
Asian	−.30
	(.23)
Native American	−.16
	(.36)
Middle Eastern	−.46**
	(.24)
Mixed Race	−.67**
	(.29)
Other	−.20
	(.73)
Constant	3.12***
	(.05)
Observations	5,000
R^2	.01
Adjusted R^2	.01
Note:	*p<0.1; **p<0.05; ***p<0.01

the sex of the pwned

Average Number of Accounts Breached
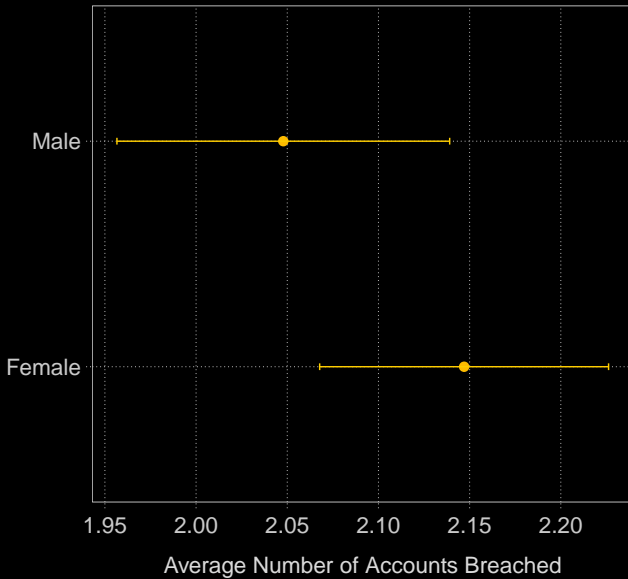
Number of Breaches by Sex

	Dependent variable:
	Number of Breaches
Male	−.35***
	(.07)
Constant	3.17***
	(.05)
Observations	5,000
R^2	.004
Adjusted R^2	.004

Note:	*p<0.1; **p<0.05; ***p<0.01

the education of the pwned

Average Number of Accounts Breached
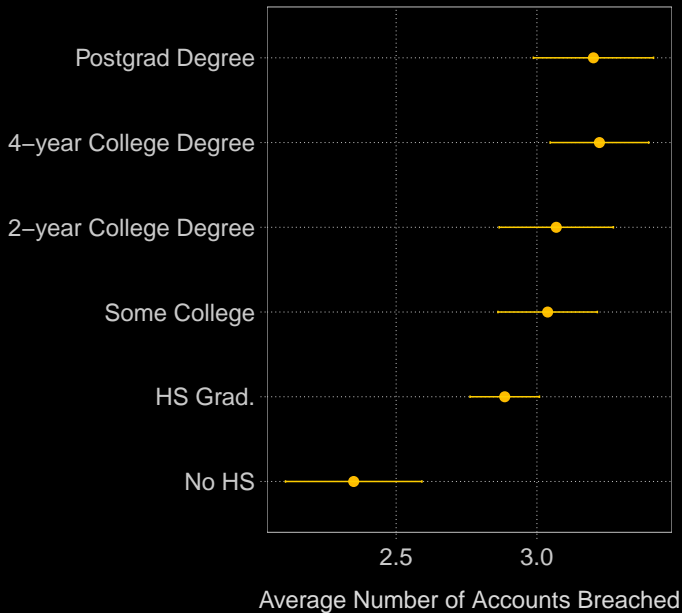
Number of Breaches by Education

	Dependent variable:
	Number of Breaches
HS Grad.	.54***
	(.16)
Some College	.69***
	(.16)
2-year College Degree	.72***
	(.18)
4-year College Degree	.87***
	(.17)
Postgrad Degree	.85***
	(.18)
Constant	2.35***
	(.14)
Observations	5,000
R^2	.01
Adjusted R^2	.01

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

who is most at blame?

or most frequently implicated domains

domain name	n
rivercitymediaonline.com	2,913
linkedin.com	1,089
modbsolutions.com	1,067
myspace.com	1,059
data4marketers.com	996
cashcrate.com	856
adobe.com	609
disqus.com	570
ticketfly.com	393
tumblr.com	340
dropbox.com	288
dailymotion.com	255
last.fm	248
evony.com	171
clixsense.com	150
cafemom.com	145
imesh.com	144
kickstarter.com	140
edmodo.com	130
zomato.com	112
neopets.com	108

domain name	n
rivercitymediaonline.com	2,913
linkedin.com	1,089
modbsolutions.com	1,067
myspace.com	1,059
data4marketers.com	996
cashcrate.com	856
adobe.com	609
disqus.com	570
ticketfly.com	393
tumblr.com	340
dropbox.com	288
dailymotion.com	255
last.fm	248
evony.com	171
clixsense.com	150
cafemom.com	145
imesh.com	144
kickstarter.com	140
edmodo.com	130
zomato.com	112
neopets.com	108

domain name	n
rivercitymediaonline.com	2,913
linkedin.com	1,089
modbsolutions.com	1,067
myspace.com	1,059
data4marketers.com	996
cashcrate.com	856
adobe.com	609
disqus.com	570
ticketfly.com	393
tumblr.com	340
dropbox.com	288
dailymotion.com	255
last.fm	248
evony.com	171
clixsense.com	150
cafemom.com	145
imesh.com	144
kickstarter.com	140
edmodo.com	130
zomato.com	112
neopets.com	108

Upshot

Upshot
  - at least 82.84% of Americans' accounts have
    been breached at least once

Upshot

- at least 82.84% of Americans' accounts have been breached at least once

- people who use online services more are somewhat more likely to have their accounts breached

Upshot

- at least 82.84% of Americans' accounts have been breached at least once

- people who use online services more are somewhat more likely to have their accounts breached

- **popular websites are to blame!**

thank you!

Data and scripts at:
https://github.com/themains/pwned/

exposure to the worst case

counting only verified, non-spamlist breaches.

	mean	se
Age		
(18,25]	1.63	0.10
(25,35]	2.44	0.08
(35,50]	2.37	0.07
(50,65]	2.16	0.06
(65,100]	1.78	0.05
Missing	0.91	0.13
Education		
No HS	1.53	0.09
HS Grad.	1.91	0.05
Some College	2.22	0.08
2-year College Degree	2.10	0.08
4-year College Degree	2.37	0.08
Postgrad Degree	2.30	0.08

	mean	se
Sex		
Female	2.15	0.04
Male	2.05	0.05
Race		
White	2.21	0.04
Black	2.03	0.08
Hispanic/Latino	1.73	0.07
Asian	2.16	0.18
Native American	1.85	0.18
Middle Eastern	2.05	0.21
Mixed Race	1.70	0.19
Other	2.69	1.19