

APIs & R

An example using HIBPwnd: an R wrapper for the Have I Been Powned API

Ruben Chavez

28 November 2018

MSDS 6303, Unit 4.3

Package Information

- **Package Website:** <https://github.com/lockedata/HIBPwned>
- **Reference Manual:** <https://cran.r-project.org/web/packages/HIBPwned/HIBPwned.pdf>
- **API Website:** <https://haveibeenpwned.com/API/v2>
- **Vignette:** <https://cran.r-project.org/web/packages/HIBPwned/vignettes/HIBP.html>
OR R: `vignette("HIBP")`
- **Install Command:** `install.packages("HIBPwned")`
- **Package function:** Check whether email addresses and/or user names have been present in a publicly disclosed data breach.

Command Example: account_breaches("email", truncate = Boolean)

- The command returns the record of data breaches on File
- Setting truncate to FALSE gives the full record displayed below
- Setting Truncate to TRUE simply returns the email address follow by a list of breaches by title

```
> account_breaches('████████████████████', truncate = FALSE)
$
  Name Title Domain BreachDate AddedDate
1 MrExcel MrExcel mrexcel.com 2016-12-05 2017-01-22T07:39:17Z
  ModifiedDate PwnCount
1 2017-01-22T07:39:17Z 366140

Description
1 In December 2016, the forum for the Microsoft Excel tips and solutions site <a href="http://www.mrexcel.com/details-of-data-breach-at-mrexcel-com/" target="_blank" rel="noopener">Mr Excel suffered a data breach</a>. The hack of the vBulletin forum led to the exposure of over 366k accounts along with email and IP addresses, dates of birth and salted passwords hashed with MD5. The owner of the MrExcel forum subsequently self-submitted the data to HIBP.

LogoPath
1 https://haveibeenpwned.com/Content/Images/PwnedLogos/MrExcel.png

DataClasses
1 Dates of birth, Email addresses, IP addresses, Passwords, Social connections, User names, website activity
  IsVerified IsFabricated IsSensitive IsRetired IsSpamList
1 TRUE FALSE FALSE FALSE FALSE
```

Additional Functionality

- User can also search by 'breached_site("breachname")' OR 'breached_sites("domain")' to return the list of data breaches for the domain including the data classes vulnerable to the breach
- Data_classes() will return a list of the available data classes used by the api

Additional Information about account_breaches()

- The data is returned as a dataframe with the related breaches on record for the email
- Column names are

```
[1] "Name"      "Title"      "Domain"      "BreachDate"  "AddedDate"  
[6] "ModifiedDate" "PwnCount"    "Description"  "LogoPath"    "DataClasses"  
[11] "IsVerified"  "IsFabricated" "IsSensitive"  "IsRetired"    "IsSpamList"
```

- Any advanced use should make use of the Boolean Values found in the last five columns as some data breaches are known fabricated or unverified.
- Using `unique(unlist(breach$emailaddress[, "DataClasses"]))` will return a highly usable vector of personal information data classes involved in reported data breaches