**Cyber Security Breaches**

**Description**

data.frame of cyber security breaches involving health care records of 500 or more humans reported to the U.S. Department of Health and Human Services (HHS) as of June 27, 2014.

**Usage**

data(breaches)

**Format**

A data.frame with 1055 observations on the following 24 variables:

Number

integer record number in the HHS data base

Name\_of\_Covered\_Entity

factor giving the name of the entity experiencing the breach

State

Factor giving the 2-letter code of the state where the breach occurred. This has 52 levels for the 50 states plus the District of Columbia (DC) and Puerto Rico (PR).

Business\_Associate\_Involved

Factor giving the name of a subcontractor (or blank) associated with the breach.

Individuals\_Affected

integer number of humans whose records were compromised in the breach. This is 500 or greater; U.S. law requires reports of breaches involving 500 or more records but not of breaches involving fewer.

Date\_of\_Breach

character vector giving the date or date range of the breach. Recodes as Dates in breach\_start and breach\_end.

Type\_of\_Breach

factor with 29 levels giving the type of breach (e.g., "Theft" vs., "Unauthorized Access/Disclosure", etc.)

Location\_of\_Breached\_Information

factor with 41 levels coding the location from which the breach occurred (e.g., "Paper", "Laptop", etc.)

Date\_Posted\_or\_Updated

Date the information was posted to the HHS data base or last updated.

Summary

character vector of a summary of the incident.

breach\_start

Date of the start of the incident = first date given in Date\_of\_Breach above.

breach\_end Date of the end of the incident or NA if only one date is given in Date\_of\_Breach above. year integer giving the year of the breach

**Details**

The data primarily consists of breaches that occurred from 2010 through early 2014 when the extract was taken. However, a few breaches are recorded including 1 from 1997, 8 from 2002-2007, 13 from 2008 and 56 from 2009. The numbers of breaches from 2010 - 2014 are 211, 229, 227, 254 and 56, respectively. (A chi-square test for equality of the counts from 2010 through 2013 is 4.11, which with 3 degrees of freedom has a significance probability of 0.25. Thus, even though the lowest number is the first and the largest count is the last, the apparent trend is not statistically significant under the usual assumption of independent Poisson trials.)

The following corrections were made to the file:

|  |  |  |
| --- | --- | --- |
| Number | Name of Covered Entity | Corrections |
|  |  |  |
| 45 | Wyoming Department of Health | Cause of breach was missing. Added "Unauthorized |
|  |  | Access / Disclosure" per [smartbreif.com/03/29/10](http://www.smartbrief.com/03/29/10/5-more-organizations-added-hhs-online-data-breach-list-0) |
| 55 | Reliant Rehabilitation Hospital North | Cause of breach was missing. Added "Unauthorized |
|  | Houston | Access / Disclosure" per Dissent. "Two Breaches |
|  |  | Involving Unauthorized Access Lead to Notification." |
|  |  | [PHIprivacy.net. N.p., 20 Apr. 2010.](http://www.phiprivacy.net/two-breaches-involving-unauthorized-access-lead-to-notification/) |
| 123 | Aetna | Cause of breach was missing. Added Improper |
|  |  | disposal per [Aetna.com/news/newsReleases/2010/0630](http://www.aetna.com/news/newsReleases/2010/0630_File_Cabinet_Final.html) |
| 157 | Mayo Clinic | Cause of breach was missing. Added Unauthorized |
|  |  | Access/Disclosure per Anderson, Howard. "Mayo Fires |
|  |  | "Employees in 2 Incidents: Both Involved |
|  |  | Unauthorized Access to Records." |
|  |  | [Data Breach Today. N.p., 4 Oct. 2010](http://www.databreachtoday.com/mayo-fires-employees-in-2-incidents-a-2974) |
| 341 | Saint Barnabas MedicL Center | Misspelled "Saint Barnabas Medical Center" |
| 347 | Americar Health Medicare | Misspelled "American Health Medicare" |
| 484 | Lake Granbury Medicl Ceter | Misspelled "Lake Granbury Medical Center" |
| 782 | See list of Practices under Item 9 | Replaced name as "Cogent Healthcare, Inc." checked |
|  |  | from XML and web documents |
| 805 | Dermatology Associates of Tallahassee | Had 00/00/0000 on breach date. This was crossed |
|  |  | check to determine that it was Sept 4, 2013 with 916 records |
| 815 | Santa Clara Valley Medical Center | Mistype breach year as 09/14/2913 corrected as 09/14/2013 |
| 961 | Valley View Hosptial Association | Misspelled "Valley View Hospital Association" |
| 1034 | Bio-Reference Laboratories, Inc. | Date changed from 00/00/000 to 2/02/2014 as |
|  |  | subsequently determined. |
|  |  |  |

**Source**

U.S. Department of Health and Human Services: Health Information Privacy: [Breaches Affecting 500 or More Individuals](https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf)

**See Also**

HHSCyberSecurityBreaches for a version of these data downloaded more recently. This newer version includes changes in reporting and in the variables included in the data.frame.

**Examples**

data(breaches)

quantile(breaches$Individuals\_Affected)

# confirm that the smallest number is 500

# -- and the largest is 4.9e6

# ... and there are no NAs

dDays <- with(breaches, breach\_end - breach\_start)

quantile(dDays, na.rm=TRUE)

# confirm that breach\_end is NA or is later than

# breach\_start