

# Package ‘raw’

August 8, 2017

**Title** R Actuarial Workshops

**Description** In order to facilitate R instruction for actuaries, we have organized several sets of publicly available data of interest to non-life actuaries. In addition, we suggest a set of packages, which most practicing actuaries will use routinely. Finally, there is an R markdown skeleton for basic reserve analysis.

**Version** 0.1.5

**Maintainer** Brian A. Fannin <captain@PirateGrunt.com>

**LazyData** true

**License** CC0

**Depends** R (>= 3.0)

**Suggests** dplyr, tidyr, ggplot2, lubridate, devtools, actuar, ChainLadder, knitr, readr, readxl, rmarkdown, mondate, FinCal, scales, nlme, maps, fitdistrplus, tweedie, nycflights13, XML, RColorBrewer, Lahman

**VignetteBuilder** knitr

**Type** Package

**RoxygenNote** 6.0.1

**URL** [http://pirategrunt.com/raw\\_package/](http://pirategrunt.com/raw_package/)

**NeedsCompilation** no

**Author** Brian A. Fannin [aut, cre]

**Repository** CRAN

**Date/Publication** 2017-08-08 14:54:09 UTC

## R topics documented:

CasColNames . . . . .	2
COTOR2 . . . . .	3
Hurricane . . . . .	4
InstallSuggestions . . . . .	4
MultiTri . . . . .	5

NAIC . . . . .	6
NJM_WC . . . . .	7
PackageDependencies . . . . .	8
PackageIsInstalled . . . . .	8
PackageNeedsUpdating . . . . .	9
PPA . . . . .	10
raw . . . . .	11
RegionExperience . . . . .	12
StateExperience . . . . .	12
<b>Index</b>	<b>14</b>

---

CasColNames	<i>CAS Column Names</i>
-------------	-------------------------

---

## Description

This will change the column names of a data frame obtained from the CAS database of reserve triangles. This data was procured and modified by Glenn G. Meyers and Peng Shi and is hosted by the CAS. More information on the data is available here: [http://www.casact.org/research/index.cfm?fa=loss\\_reserves\\_data](http://www.casact.org/research/index.cfm?fa=loss_reserves_data)

## Usage

```
CasColNames(df, restore = TRUE)
```

## Arguments

df	A data frame of loss reserving data
restore	Boolean indicating whether to restore the names defined by Meyers and Shi. The default is true.

## Value

A data frame with different column names.

## Examples

```
## Not run:
data(NJM_WC)
NJM_WC <- CasColNames(NJM_WC, restore = TRUE)

## End(Not run)
```

**Description**

The **Casualty Actuarial Society's Committee on Theory of Risk (COTOR)** issued several challenges as part of their work. These involved artificial constructed loss data. Actuaries were invited to construct models to estimate appropriate premiums for risk transfer.

Note that there is no COTOR1.

**COTOR2**

COTOR2 is a vector containing 250 observations of claim values.

**COTOR round 2****COTOR3**

COTOR3 is a data frame with 490 observations of 2 variables. The values were generated by the great Stuart Klugman.

- Year: an integer indicating the year in which the claim occurred.
- Amount: the amount of the claim.

**COTOR round 3****COTOR4**

COTOR4 is a data frame with 2500 rows and 2 columns.

- Year: an integer indicating the year in which the claim occurred.
- Amount: the amount of the claim.

**COTOR round 4****COTOR5**

COTOR5 is a data frame with 4849 rows and 5 columns

- Risk: A single letter indicating the individual risk to which the claim belongs.
- Year: an integer indicating the year in which the claim occurred.
- MSA: simulated metropolitan statistical area.
- Amount: the amount of the claim.

**COTOR round 5****Usage**

```
data(COTOR2)
```

---

Hurricane

*Hurricane data*


---

### Description

Listing of the position and windspeed of hurricanes from 1970 through 2011.

Data is taken from the [National Hurricane Center](#), a division of the National Oceanic and Atmospheric Association. The data may be obtained here: [NOAA ftp](#). A description of the data as well as data for storms in other regions may be found here: [IBTrACS](#).

The fields "SerialNumber", "Basin", "SubBasin", "Nature", "Pressure", "Center", "WindPercentile", "PressurePercentile" and "TrackType" have been removed.

- Year: Integer from 1971 through 2011
- Number: Integer used to designate the hurricanes in a particular year
- Name: The name of the hurricane
- ISO\_Time: The time the observation was made
- Latitude: Latitude where the observation was made
- Longitude: Longitude where the observation was made
- Wind: windspeed measured knots (kt)
- Pressure: atmospheric pressure measured in millibars (mb)

### Usage

```
data(Hurricane)
```

### Format

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 17494 rows and 8 columns.

---

InstallSuggestions

*InstallSuggestions*


---

### Description

This function will install the full set of suggested packages.

### Usage

```
InstallSuggestions(pkg = "raw", quiet, ...)
```

**Arguments**

pkg	Name of the package for which you want to install suggestions
quiet	Include additional messages while installing?
...	Additional named arguments passed to 'install.packages'

**Value**

A character vector indicating which packages were installed.

**Examples**

```
## Not run:
  InstallSuggestions()

## End(Not run)
```

---

MultiTri

*MultiTri*


---

**Description**

Complete triangle of losses for the same 4 private passenger auto and workers comp companies. Data is taken from the CAS.

- GroupCode: NAIC company code
- Company
- AccidentYear
- DevelopmentYear
- Lag: Integer indicating the development lag in years.
- CumulativeIncurred
- CumulativePaid
- IBNR: Bulk and IBNR reserves
- DirectEP: Direct and assumed earned premium, gross of reinsurance.
- CededEP: Earned premium ceded to reinsurers.
- NetEP: Premium earned net of reinsurance.
- Single: Either 1 or 0. 1 indicates a single entity, 0 indicates a group entity.
- Reserve1997: Posted reserves as of year-end 1997.

**Source**

[http://www.casact.org/research/reserve\\_data/ppauto\\_pos.csv](http://www.casact.org/research/reserve_data/ppauto_pos.csv)  
[http://www.casact.org/research/reserve\\_data/wkcomp\\_pos.csv](http://www.casact.org/research/reserve_data/wkcomp_pos.csv)  
[http://www.casact.org/research/reserve\\_data/othliab\\_pos.csv](http://www.casact.org/research/reserve_data/othliab_pos.csv)

**See Also**

[NAIC](#)

---

NAIC

*NAIC*

---

**Description**

This data represents ten complete years of Schedule P development for many NAIC reporting companies. The data was prepared by Glenn Meyers and Peng Shi and is available from the CAS [http://www.casact.org/research/index.cfm?fa=loss\\_reserves\\_data](http://www.casact.org/research/index.cfm?fa=loss_reserves_data).

NOTE that the columns have been given more user friendly names. The original column names may be restored by using the CasColNames function.

- GroupCode: NAIC company code
- Company
- AccidentYear
- DevelopmentYear
- Lag: Integer indicating the development lag in years.
- CumulativeIncurred
- CumulativePaid
- IBNR: Bulk and IBNR reserves
- DirectEP: Direct and assumed earned premium, gross of reinsurance.
- CededEP: Earned premium ceded to reinsurers.
- NetEP: Premium earned net of reinsurance.
- Single: Either 1 or 0. 1 indicates a single entity, 0 indicates a group entity.
- Reserve1997: Posted reserves as of year-end 1997.

**Source**

[http://www.casact.org/research/reserve\\_data/ppauto\\_pos.csv](http://www.casact.org/research/reserve_data/ppauto_pos.csv)  
[http://www.casact.org/research/reserve\\_data/wkcomp\\_pos.csv](http://www.casact.org/research/reserve_data/wkcomp_pos.csv)  
[http://www.casact.org/research/reserve\\_data/comauto\\_pos.csv](http://www.casact.org/research/reserve_data/comauto_pos.csv)  
[http://www.casact.org/research/reserve\\_data/medmal\\_pos.csv](http://www.casact.org/research/reserve_data/medmal_pos.csv)  
[http://www.casact.org/research/reserve\\_data/prodliab\\_pos.csv](http://www.casact.org/research/reserve_data/prodliab_pos.csv)  
[http://www.casact.org/research/reserve\\_data/othliab\\_pos.csv](http://www.casact.org/research/reserve_data/othliab_pos.csv)

**Description**

This data represents ten complete years of Schedule P workers comp development for the New Jersey Manufacturing Company. The data was prepared by Glenn Meyers and Peng Shi and is available from the CAS [http://www.casact.org/research/index.cfm?fa=loss\\_reserves\\_data](http://www.casact.org/research/index.cfm?fa=loss_reserves_data). The columns have been given more user friendly names. The original column names may be restored by using the `CasColNames` function.

This data has been included because it is the triangle used by Greg Taylor and Grainne McGuire in their monograph "Stochastic Loss Reserving Using Generalized Linear Models". <http://www.casact.org/pubs/monographs/monograph03>

This is a data frame with 100 observations of 13 variables.

- `GroupCode`: NAIC company code
- `Company`
- `AccidentYear`
- `DevelopmentYear`
- `Lag`: Integer indicating the development lag in years.
- `CumulativeIncurred`
- `CumulativePaid`
- `IBNR`: Bulk and IBNR reserves
- `DirectEP`: Direct and assumed earned premium, gross of reinsurance.
- `CededEP`: Earned premium ceded to reinsurers.
- `NetEP`: Premium earned net of reinsurance.
- `Single`: Either 1 or 0. 1 indicates a single entity, 0 indicates a group entity.
- `Reserve1997`: Posted reserves as of year-end 1997.

**Usage**

```
data(NJM_WC)
```

**Format**

An object of class `tbl_df` (inherits from `tbl`, `data.frame`) with 100 rows and 13 columns.

PackageDependencies     *PackageDependencies*

---

### Description

PackageDependencies will examine the DESCRIPTION file of a package and return a vector of dependencies.

### Usage

```
PackageDependencies(x, type = c("Depends", "Imports", "Suggests", "Enhances",  
"LinkingTo"))
```

### Arguments

x	The name of the packages
type	What sort of dependency are you looking for? Permissible values are "Depends", "Imports", "Suggests", "Enhances" and "LinkingTo"

### Examples

```
## Not run:  
PackageDependencies("ggplot2", "Depends")  
  
## End(Not run)
```

---

PackageIsInstalled     *PackageIsInstalled*

---

### Description

This function will indicate whether a package has been installed. It uses the installed.packages function for this.

### Usage

```
PackageIsInstalled(pkgs)
```

### Arguments

pkgs	Character vector of package names.
------	------------------------------------

### Value

A logical vector indicating whether the packages have been installed.



**Examples**

```
## Not run:  
  PackageIsInstalled("ggplot2")  
  
## End(Not run)
```

---

`PackageNeedsUpdating` *PackageNeedsUpdating*

---

**Description**

This function will indicate whether a package needs to be updated. It uses the `old.packages` function for this.

**Usage**

```
PackageNeedsUpdating(pkgs)
```

**Arguments**

`pkgs`            Character vector of package names.

**Value**

A logical vector indicating whether the packages need to be updated.

**Examples**

```
## Not run:  
  PackageNeedsUpdating("ggplot2")  
  
## End(Not run)
```

**Description**

This is data taken from Appendix A of the "**Basic Ratemaking**" study note by Werner and Modlin.

**Rate change**

This is a data frame with 7 observations of 5 variables.

- RateGroup: Letters from A-G. Note that rate group "A" does not have any rate changes.
- EffectiveDate: The effective date of the rate change.
- RateChange: A number indicating the rate change. Note that this is NOT done as a percent, i.e. .05 = 5
- RateLevelIndex: This is one plus the rate change.
- CumulativeRateLevel: This is the cumulative product of the RateLevelIndex. Note that this varies slightly from the figures shown in the study note. I believe this is due to rounding in the study note.

**Accident year**

This is a data frame with 5 observations of 4 variables.

- AccidentYear: The accident year.
- EarnedExposure: The exposure earned for the accident year.
- EarnedPremium: The premium earned for the accident year.
- ReportedLoss: The reported loss and paid ALAE for the accident year.

**Premium trend**

Data frame with 25 observations of 3 values. The premium is shown at current level. This may be used to determine premium or exposure trend.

- QuarterEnding: The ending quarter as a Date type.
- WrittenPremium: Premium written during the quarter at current rate level.
- WrittenExposure: Exposure written during the quarter.

**Loss development**

Data frame with 25 observations of 3 values. Note that this loss development data is stored in "long" rather than "wide" format.

- AccidentYear: The accident year as an integer.
- Lag: The lag in months as an integer.
- ReportedLoss: The reported loss and paid ALAE for that accident year and lag.

**Loss trend**

Data frame with 20 observations of 7 values. The data may be used to estimate loss trend.

- YearEndingQuarter: The end of a one-year period as a Date.
- EarnedExposure: The exposure earned during the period.
- ClosedClaimCount: The number of claims closed during the period.
- PaidLosses: The amount of loss paid during the period.
- Frequency: The closed claim count divided by the exposure.
- Severity: The paid losses divided by the closed claim count.
- PurePremium: The paid losses divided by the earned exposure.

**ULAE**

A data frame with 3 observations of 4 values.

- CalendarYear: The calendar year as an integer.
- PaidLoss\_ALAE: The paid loss and ALAE for the calendar year.
- PaidULAE: The ULAE paid in the calendar year.
- ULAE\_Ratio: The ratio of ULAE to the paid loss and ALAE.

**Usage**

```
data(PPA)
```

---

raw

*raw - R Actuarial Workshops*

---

**Description**

This is a package which stores data used in R workshops sponsored by the Casualty Actuarial Society.

**Details**

For more information, read the package vignette. To access the package vignette, type in the code shown in the Examples section (the line without the hash marks).

**Examples**

```
## Not run:  
vignette(topic = "raw")  
  
## End(Not run)
```

RegionExperience      *Region data*

---

**Description**

Consolidation of state data into various regions. The state data was generated randomly as a Poisson sample.

- Region: The name of the region
- PolicyYear: An integer between 2001 and 2010
- NumPolicies: Integer giving the number of policies in that region for that policy year
- NumClaims: Integer giving the number of claims in that region for that policy year

**Usage**

```
data(RegionExperience)
```

**Format**

An object of class `data.frame` with 40 rows and 4 columns.

**See Also**

[StateExperience](#)

---

StateExperience      *State data*

---

**Description**

Random illustrative data

- Region: The name of the region to which the state belongs
- State: The name of the state
- Postal: Two character postal code for the state
- PolicyYear: An integer between 2001 and 2010.
- NumPolicies: Integer giving the number of policies in that region for that policy year
- NumClaims: Integer giving the number of claims in that region for that policy year

**Usage**

```
data(StateExperience)
```

**Format**

An object of class `data.frame` with 500 rows and 6 columns.

**See Also**

[RegionExperience](#)

# Index

## \*Topic **datasets**

- Hurricane, [4](#)
  - NJM\_WC, [7](#)
  - RegionExperience, [12](#)
  - StateExperience, [12](#)
- CasColNames, [2](#)
- comauto (NAIC), [6](#)
- COTOR (COTOR2), [3](#)
- COTOR1 (COTOR2), [3](#)
- COTOR2, [3](#)
- COTOR3 (COTOR2), [3](#)
- COTOR4 (COTOR2), [3](#)
- COTOR5 (COTOR2), [3](#)
- Hurricane, [4](#)
- InstallSuggestions, [4](#)
- medmal (NAIC), [6](#)
- MultiTri, [5](#)
- NAIC, [6](#), [6](#)
- NJM\_WC, [7](#)
- othliab (NAIC), [6](#)
- PackageDependencies, [8](#)
- PackageIsInstalled, [8](#)
- PackageNeedsUpdating, [9](#)
- PPA, [10](#)
- PPA\_AccidentYear (PPA), [10](#)
- PPA\_LossDevelopment (PPA), [10](#)
- PPA\_LossTrend (PPA), [10](#)
- PPA\_PremiumTrend (PPA), [10](#)
- PPA\_RateChange (PPA), [10](#)
- PPA\_ULAE (PPA), [10](#)
- ppauto (NAIC), [6](#)
- prodliab (NAIC), [6](#)
- raw, [11](#)
- raw-package (raw), [11](#)
- RegionExperience, [12](#), [13](#)
- StateExperience, [12](#), [12](#)
- wkcomp (NAIC), [6](#)