Package 'rIO'

November 4, 2018

Type Package

Title Read/Write any format from anywhere

Version 0.1.0

Author Himanshu Sikaria [aut, cre]

Maintainer Himanshu Sikaria <himanshu.sikaria@socialcops.com>

Description Perform input, outout of files in R from any data source like Google Cloud Storage, AWS S3 or local drive.

URL https://github.com/socialcopsdev/rIO

BugReports https://github.com/socialcopsdev/rIO/issues

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Depends R (>= 3.1.2)

Imports stringr, assertthat, googleCloudStorageR, aws.s3, utils, tools

NeedsCompilation no

R topics documented:

`rioAuth` *Authenticate rIO*

Description

Authenticate any of the cloud storage platforms to perform any I/O

Usage

```
rioAuth(auth_list = "", data_source = rioGetDataSource(),
  scope = "https://www.googleapis.com/auth/devstorage.full_control")
```

Arguments

`auth_list`	path to the json file or the system environment name in case of GCS. For S3 a vector for access_key, secret_access_key, region (optional; default us-east-1) and session_id (optional); this could also be a single comma separated string.
`data_source`	default to local. Possible options : gcs, s3, local. Case insensitive
`scope`	the scope of the auth if gcs. Default https://www.googleapis.com/auth/devstorage.full_control

Examples

```
rioSetDataSource("local")
rioAuth()
```

`rioFileDownload` *Download file from cloud to local system*

Description

Save a single file from the cloud to your local drive

Usage

```
rioFileDownload(bucketpath, localfile, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source), overwrite = TRUE, ...)
```

Arguments

bucketpath	path of file in the bucket
localfile	path where the file needs to be downloaded. The filename and extension also needs to be present, if not the current file name will be considered
data_source	the name of the data source, if not set globally, gcs or s3
bucket	the name of the bucket, if not set globally
overwrite	logical. If the files should be overwritten if already present
...	other parameters for gcs_get_object or save_object

Value

the filename and path of the object saved to local

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioFileDownload("mtcars.csv", "~/Downloads/dsada.csv", overwrite = T)

## End(Not run)
```

---

rioFileExists *Check if a file exists*

Description

Check if a file exists

Usage

```
rioFileExists(path, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source))
```

Arguments

path	the entire path for the file
data_source	the name of the data source, gcs, s3 or local; if not set globally
bucket	the name of the bucket, if not set globally

Value

logical. if the file exists or not

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioFileExists(path = "tests/mtcars.csv")

## End(Not run)
```

rioFileUpload *Upload a file from the local system to cloud*

Description

Write a local file to the cloud, S3 or GCS

Usage

```
rioFileUpload(localfile, bucketpath, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source), ...)
```

Arguments

localfile	path of the file to be uploaded
bucketpath	path there the file needs to be uploaded, the filename can or cannot be present
data_source	the name of the data source, if not set globally. gcs or s3
bucket	the name of the bucket, if not set globally
...	other parameters for gcs_upload

Value

the filename and path of the file in the bucket

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioFileUpload("~/Downloads/dsada.csv", "tests/mtcars.csv")

## End(Not run)
```

rioGetBucket *Get global bucket name for rIO*

Description

Get global bucket name to be used for all the function in rIO

Usage

```
rioGetBucket(data_source = rioGetDataSource())
```

Arguments

data_source	the data source used for IO. Default chooses the data source set using rioSetDataSource()

Details

if the data source is local, then an empty string is returned

Value

the string - bucket name stored

Examples

```
# first setting the bucket for a data source
rioSetBucket(bucket = "socialcops-test", data_source = "S3")
# retrieving the bucket for S3
rioGetBucket(data_source = "S3")
```

---

rioGetDataSource *Get global Data Source name for rIO*

Description

Get global data source name to be used for all the function in rIO. Returns the value stored using rioSetDataSource

Usage

```
rioGetDataSource()
```

Value

the string - DataSource name stored

Examples

```
# first setting the data source
rioSetDataSource("s3")
# getting the data source
rioGetDataSource()
```

---

rioListFiles *List the Files in a Directory/Folder*

Description

list the files in cloud or locally - similar to list.files()

Usage

```
rioListFiles(path = "", pattern = NULL, recursive = FALSE,
  ignore.case = FALSE, full.names = TRUE,
  data_source = rioGetDataSource(), bucket = rioGetBucket(data_source))
```

Arguments

path	the folder for which the files needs to be listed
pattern	an optional regular expression. Only file path names which match the regular expression will be returned.
recursive	logical. Should the listing recurse into directories?
ignore.case	logical. Should pattern-matching be case-insensitive?
full.names	logical. Should the entire path be returned or only after the path inputed.
data_source	the name of the data source, gcs, s3 or local; if not set globally
bucket	the name of the bucket, if not set globally

Value

a vector of full file names

Examples

```
## Not run:
rioSetDataSource("s3")
rioSetBucket("socialcops-test")
gcsListFiles(path = "tests/", pattern = ".*csv")

## End(Not run)
```

---

rioReadRaster *Read raster files*

Description

Read raster data from anywhere using a function defined by you

Usage

```
rioReadRaster(file, FUN = raster::raster,
  data_source = rioGetDataSource(), bucket = rioGetBucket(data_source),
  ...)
```

Arguments

file	path of the file to be read
FUN	the function using which the file is to be read
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally
...	other parameters for the FUN function defined above

Value

the output of the FUN function

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
t = rioReadRaster("tests/testras.tif", raster)

## End(Not run)
```

`rioReadRda` *Read RDA file*

Description

Read RData or rda file from anywhere

Usage

```
rioReadRda(file, FUN = load, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source), ...)
```

Arguments

`file`	path of the file to be read
`FUN`	the function using which the file is to be read
`data_source`	the name of the data source, if not set globally. s3, gcs or local
`bucket`	the name of the bucket, if not set globally
`...`	other parameters for the FUN function defined above

Value

the output of the FUN function

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
load("tests/googletest.rda")

## End(Not run)
```

rioReadRds *Read RDS file*

Description

Read R data - RDS file from anywhere

Usage

```
rioReadRds(file, FUN = readRDS, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source), ...)
```

Arguments

file	path of the file to be read
FUN	the function using which the file is to be read
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally
...	other parameters for the FUN function defined above

Value

the output of the FUN function

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioReadRds("tests/googletest.rds", readRDS)

## End(Not run)
```

rioReadShp *Read shapefiles*

Description

Read shapefiles data from anywhere using a function defined by you

Usage

```
rioReadShp(dsn, layer, FUN = rgdal::readOGR, dsnlayerbind = F,
  data_source = rioGetDataSource(), bucket = rioGetBucket(data_source),
  ...)
```

Arguments

dsn	path of the file to be read
layer	the name of the shapefile without extension
FUN	the function using which the file is to be read
dsnlayerbind	if the FUN needs dsn and layer binded or not
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally
...	other parameters for the FUN function defined above

Value

the output of the FUN function

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
t = rioReadShp("tests/shptest/", "testshp", FUN = readOGR, dsnlayerbind = F)
t = rioReadShp("tests/shptest/", "testshp", FUN = shapefile, dsnlayerbind = T)

## End(Not run)
```

---

rioReadTable *Read csv, Excel files, txt*

Description

Read tabular data from anywhere using a function defined by you

Usage

```
rioReadTable(file, FUN = read.csv, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source), ...)
```

Arguments

file	path of the file to be read
FUN	the function using which the file is to be read
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally
...	other parameters for the FUN function defined above

Value

the output of the FUN function

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioReadTable("tests/googletest.xlsx", read_excel)

## End(Not run)
```

`rioSetBucket` *Set global bucket name for rIO*

Description

Set global bucket name to be used for all the function in rIO

Usage

```
rioSetBucket(bucket, data_source = rioGetDataSource())
```

Arguments

`bucket`	the bucket name to be set
`data_source`	the data source used for IO. Default chooses the data source set using rioSetDataSource()

Value

stores the bucket name in a global environment under rioBucketGcs or rioBucketS3

Examples

```
rioSetBucket(bucket = "socialcops-test", data_source = "S3")
```

`rioSetDataSource` *Set global DataSource name for rIO*

Description

Set global DataSource name to be used for all the function in rIO

Usage

```
rioSetDataSource(data_source)
```

Arguments

`data_source`	the DataSource name to be set

Value

stores the DataSource name in a global environment under rioDataSource

Examples

```
rioSetDataSource("local")
```

`rioWriteRaster` *Write raster*

Description

Write raster

Usage

```
rioWriteRaster(x, file, FUN = raster::writeRaster,
  data_source = rioGetDataSource(), bucket = rioGetBucket(data_source),
  ...)
```

Arguments

`x`	variable name
`file`	path of the file to be written to
`FUN`	the function using which the file is to write
`data_source`	the name of the data source, if not set globally. s3, gcs or local
`bucket`	the name of the bucket, if not set globally
`...`	other parameters for the FUN function defined above

Value

No output

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioWriteRaster(t, "tests/testras.tif", writeRaster, format = "GTiff")

## End(Not run)
```

rioWriteRda *Write RDA files*

Description

Write R data RDA file to anywhere from R

Usage

```
rioWriteRda(..., file, FUN = save, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source))
```

Arguments

...	R pbjects need to be saved
file	path of the file to be written to
FUN	the function using which the file is to write
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally

Value

No output

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioWriteRda(iris, mtcars, "tests/iris.rda")

## End(Not run)
```

rioWriteRds *Write RDS files*

Description

Write R data RDS file to anywhere from R

Usage

```
rioWriteRds(x, file, FUN = saveRDS, data_source = rioGetDataSource(),
  bucket = rioGetBucket(data_source), ...)
```

Arguments

x	variable name
file	path of the file to be written to
FUN	the function using which the file is to write
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally
...	other parameters for the FUN function defined above

Value

if FUN returns anything

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioWriteRds(iris, "tests/iris.rds", saveRDS)

## End(Not run)
```

---

rioWriteShp *Write shapefiles*

Description

Write shapefiles

Usage

```
rioWriteShp(obj, dsn, layer, FUN = rgdal::writeOGR, dsnlayerbind = F,
  data_source = rioGetDataSource(), bucket = rioGetBucket(data_source),
  ...)
```

Arguments

obj	R object to be written
dsn	path of the file to be read
layer	the name of the shapefile without extension
FUN	the function using which the file is to be read
dsnlayerbind	if the FUN needs dsn and layer binded or not
data_source	the name of the data source, if not set globally. s3, gcs or local
bucket	the name of the bucket, if not set globally
...	other parameters for the FUN function defined above

Value

output of the FUN function if any

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioWriteShp(t, "tests/shptest/", "new", driver = "ESRI Shapefile", overwrite = T)

## End(Not run)
```

`rioWriteTable` *Write csv, Excel files, txt*

Description

Write csv, Excel files, txt

Usage

```
rioWriteTable(x, file, FUN = write.csv,
  data_source = rioGetDataSource(), bucket = rioGetBucket(data_source),
  ...)
```

Arguments

`x`	variable name
`file`	path of the file to be written to
`FUN`	the function using which the file is to write
`data_source`	the name of the data source, if not set globally. s3, gcs or local
`bucket`	the name of the bucket, if not set globally
`...`	other parameters for the FUN function defined above

Value

No output

Examples

```
## Not run:
rioSetDataSource("gcs")
rioSetBucket("socialcops-test")
rioWriteTable(iris, "tests/iris.csv", write.csv)

## End(Not run)
```

Index