Time Series Database Interface: R ODBC (TSodbc)

May 6, 2009

1 Introduction

The code from the vignette that generates this guide can be loaded into an editor with edit(vignette("TSodbc")). This uses the default editor, which can be changed using options(). It should be possible to view the pdf version of the guide for this package with print(vignette("TSodbc")).

WARNING: running these example will overwrite tables in the ODBC "test" database on the server.

Once R is started, the functions in this package are made available with

```
> library("TSodbc")
```

This will also load required packages TSdbi, DBI, RODBC, methods, and tframe. Some examples below also require zoo, and tseries.

The ODBC user, password, hostname, etc, should be set in ODBC client configuration file (/.odbc.ini on Linux/Unix systems) before starting R. An example of this file is provided in the final section of this vignette. Alternatively, this information can be set with environment variables ODBC_USER and ODBC_PASSWD. The variable ODBC_HOST does not seem to work for passing the ODBC connection, so a properly setup ODBC configuration file is also needed, but the environment variables will override the user and passwd setting in that file. (An environment variable ODBC_DATABASE can also be set, but "test" is specified below.) Below, the environment variable ODBC_USER is used to determine which of these methods is being used. If this environment variable is empty then it is assumed the configuration file will be used.

```
> user <- Sys.getenv("ODBC_USER")
> if ("" != user) {
    passwd <- Sys.getenv("ODBC_PASSWD")
    if ("" == passwd)
        passwd <- NULL
}</pre>
```

The next small section of code is necessary to setup database tables that are used in the examples below. It needs to be done only once for a database and might typically be done by an administrator setting up the database, rather than by an end user.

More detailed description of the instructions for building the database tables is given in the vignette for the *TSdbi* package. Those instruction show how to build the database using database utilities rather than R, which might be the way a system administrator would build the database.

2 Using the Database - TSdbi Functions

This section gives several simple examples of putting series on and reading them from the database. (If a large number of series are to be loaded into a database, one would typically do this with a batch process using the database program's utilities for loading data.) The first thing to do is to establish a connection to the database:

TS connect uses odbc Connect from the RODBC package, but checks that the database has expected tables, and checks for additional features. (It cannot be used before the tables are created, as done in the previous section.)

This puts a series called vec on the database and then reads is back

```
> z <- ts(rnorm(10), start = c(1990, 1), frequency = 1)
> seriesNames(z) <- "vec"
> if (TSexists("vec", con)) TSdelete("vec", con)
> TSput(z, con)
> z <- TSget("vec", con)</pre>
```

If the series is printed it is seen to be a "ts" time series with some extra attributes.

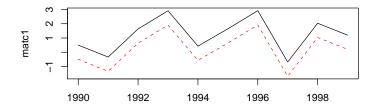
TSput fails if the series already exists on the con, so the above example checks and deletes the series if it already exists. TSreplace does not fail if the series does not yet exist, so examples below use it instead. Several plots below show original data and the data retrieved after it is written to the database. One is added to the original data so that both lines are visible.

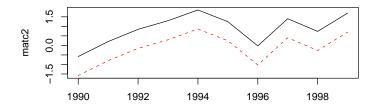
And now more examples:

```
> z \leftarrow ts(matrix(rnorm(20), 10, 2), start = c(1990, 1), frequency = 1)
> seriesNames(z) <- c("matc1", "matc2")</pre>
> TSreplace(z, con)
[1] TRUE
> TSget("matc1", con)
Time Series:
Start = 1990
End = 1999
Frequency = 1
                      2
                                   3
1.27826034 -1.88898480 0.25883226 -1.00823293 -1.17018760 1.42957759
               8
                                   9
1.12900034 -0.50176713 -0.05916423 -0.05562778
attr(,"seriesNames")
[1] matc1
attr(,"TSmeta")
An object of class âĂIJTSmetaâĂ
Slot "TSdescription":
[1] NA
Slot "TSdoc":
[1] NA
Slot "TSlabel":
[1] NA
Slot "serIDs":
[1] "matc1"
Slot "conType":
[1] "TSodbcConnection"
attr(,"package")
[1] "TSodbc"
Slot "DateStamp":
[1] NA
Slot "dbname":
[1] "test"
Slot "hasVintages":
[1] FALSE
```

```
Slot "hasPanels":
[1] FALSE
> TSget("matc2", con)
Time Series:
Start = 1990
End = 1999
Frequency = 1
                                                       5
                    2
                                3
                                            4
         1
-0.0259557 \ -1.0261141 \quad 0.9021059 \ -0.3810351 \ -1.1217003 \ -0.1093234 \quad 0.2313553
         8
                   9
                               10
-0.1748460 -0.4724316 1.5809247
attr(,"seriesNames")
[1] matc2
attr(,"TSmeta")
An object of class âĂIJTSmetaâĂ₹
Slot "TSdescription":
[1] NA
Slot "TSdoc":
[1] NA
Slot "TSlabel":
[1] NA
Slot "serIDs":
[1] "matc2"
Slot "conType":
[1] "TSodbcConnection"
attr(,"package")
[1] "TSodbc"
Slot "DateStamp":
[1] NA
Slot "dbname":
[1] "test"
Slot "hasVintages":
[1] FALSE
Slot "hasPanels":
[1] FALSE
> TSget(c("matc1", "matc2"), con)
```

```
Time Series:
Start = 1990
End = 1999
Frequency = 1
           matc1
                      matc2
1990 1.27826034 -0.0259557
1991 -1.88898480 -1.0261141
1992 0.25883226 0.9021059
1993 -1.00823293 -0.3810351
1994 -1.17018760 -1.1217003
1995 1.42957759 -0.1093234
1996 1.12900034 0.2313553
1997 -0.50176713 -0.1748460
1998 -0.05916423 -0.4724316
1999 -0.05562778 1.5809247
attr(,"TSmeta")
An object of class âĂIJTSmetaâĂ록
Slot "TSdescription":
[1] NA
Slot "TSdoc":
[1] NA
Slot "TSlabel":
[1] NA
Slot "serIDs":
[1] "matc1" "matc2"
Slot "conType":
[1] "TSodbcConnection"
attr(,"package")
[1] "TSodbc"
Slot "DateStamp":
[1] NA
Slot "dbname":
[1] "test"
Slot "hasVintages":
[1] FALSE
Slot "hasPanels":
[1] FALSE
```





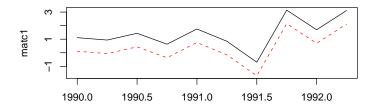
```
> z \leftarrow ts(matrix(rnorm(20), 10, 2), start = c(1990, 1), frequency = 4)
> seriesNames(z) \leftarrow c("matc1", "matc2")
> TSreplace(z, con)
```

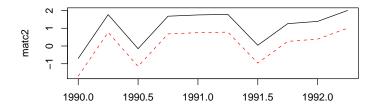
[1] TRUE

> TSget(c("matc1", "matc2"), con)

		matc1	matc2
1990	Q1	0.3048342	-0.24186957
1990	Q2	-0.9594408	-1.78877242
1990	QЗ	-1.2996596	-0.23239201
1990	Q4	2.0922763	-0.48742393
1991	Q1	0.4863951	0.48535148
1991	Q2	-0.5950624	0.50611924
1991	QЗ	1.0801153	0.29084720
1991	Q4	-0.2647560	0.34712160
1992	Q1	0.9799473	-0.05276098
1992	Q2	0.5551654	-0.87436259
attr(,"TSmeta")			

```
An object of class âĂIJTSmetaâĂ₹
Slot "TSdescription":
[1] NA
Slot "TSdoc":
[1] NA
Slot "TSlabel":
[1] NA
Slot "serIDs":
 [1] "matc1" "matc2"
Slot "conType":
[1] "TSodbcConnection"
attr(,"package")
[1] "TSodbc"
Slot "DateStamp":
 [1] NA
Slot "dbname":
 [1] "test"
Slot "hasVintages":
[1] FALSE
Slot "hasPanels":
[1] FALSE
> tfplot(z + 1, TSget(c("matc1", "matc2"), con), lty = c("solid", lty = c("solid", lty = c("solid", lty = 
                                    "dashed"), col = c("black", "red"))
```

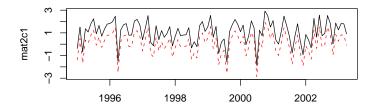


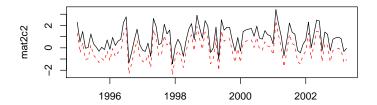


```
> z <- ts(matrix(rnorm(200), 100, 2), start = c(1995, 1), frequency = 12) > seriesNames(z) <- c("mat2c1", "mat2c2") > TSreplace(z, con)
```

[1] TRUE

> tfplot(z + 1, TSget(c("mat2c1", "mat2c2"), con), lty = c("solid", "dashed"), col = c("black", "red"))





The following extract information about the series from the database, although not much information has been added for these examples.

- > TSmeta("mat2c1", con)
- > TSmeta("vec", con)
- > TSdates("vec", con)
- > TSdescription("vec", con)
- > TSdoc("vec", con)

Below are exampoles that make more use of TSdescription and codeTSdoc. Often it is convenient to set the default connection:

> options(TSconnection = con)

and then the con specification can be omitted from the function calls unless another connection is needed. The con can still be specified, and some examples below do specify it, just to illustrate the alternative syntax.

> z <- TSget("mat2c1")
> TSmeta("mat2c1")

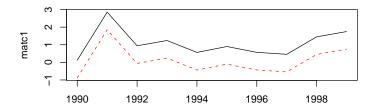
An object of class âĂIJTSmetaâö Slot "TSdescription":

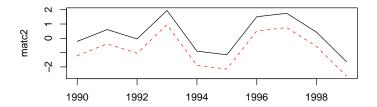
[1] NA

```
Slot "TSdoc":
[1] NA
Slot "TSlabel":
[1] NA
Slot "serIDs":
[1] "mat2c1"
Slot "conType":
[1] "TSodbcConnection"
attr(,"package")
[1] "TSodbc"
Slot "DateStamp":
[1] NA
Slot "dbname":
[1] "test"
Slot "hasVintages":
[1] FALSE
Slot "hasPanels":
[1] FALSE
```

Data documentation can be in two forms, a description specified by TSdescription or longer documentation specified by TSdoc. These can be added to the time series object, in which case they will be written to the database when TSput or TSreplace is used to put the series on the database. Alternatively, they can be specified as arguments to TSput or TSreplace. The description or documentation will be retrieved as part of the series object with TSget only if this is specified with the logical arguments TSdescription and TSdoc. They can also be retrieved directly from the database with the functions TSdescription and TSdoc.

```
> zz <- TSget("Series1", con, TSdescription = TRUE, TSdoc = TRUE)
> start(zz)
 [1] 1990
                                             1
> end(zz)
 [1] 1999
                                             1
> TSdescription(zz)
 [1] "short rnorm series"
> TSdoc(zz)
 [1] "Series created as an example in the vignette."
> TSdescription("Series1", con)
 [1] "short rnorm series"
> TSdoc("Series1", con)
[1] "Series created as an example in the vignette."
> z \leftarrow ts(rnorm(10), start = c(1990, 1), frequency = 1)
> seriesNames(z) <- "vec"
> TSreplace(z, con)
 [1] TRUE
> zz <- TSget("vec", con)
> z \leftarrow ts(matrix(rnorm(20), 10, 2), start = c(1990, 1), frequency = 1)
> seriesNames(z) <- c("matc1", "matc2")</pre>
> TSreplace(z, con)
 [1] TRUE
> tfplot(z + 1, TSget(c("matc1", "matc2"), con), lty = c("solid", lty = c("solid", lty = c("solid", lty = c("solid", lty = l
                        "dashed"), col = c("black", "red"))
```

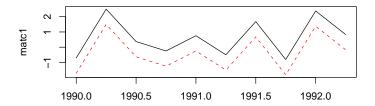


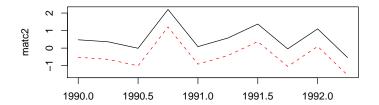


```
> z \leftarrow ts(matrix(rnorm(20), 10, 2), start = c(1990, 1), frequency = 4) > seriesNames(z) \leftarrow c("matc1", "matc2") > TSreplace(z, con)
```

[1] TRUE

> tfplot(z + 1, TSget(c("matc1", "matc2"), con), lty = c("solid", "dashed"), col = c("black", "red"))

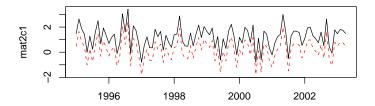


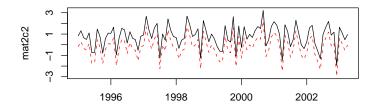


```
> z <- ts(matrix(rnorm(200), 100, 2), start = c(1995, 1), frequency = 12) > seriesNames(z) <- c("mat2c1", "mat2c2") > TSreplace(z, con)
```

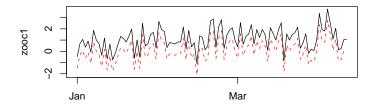
[1] TRUE

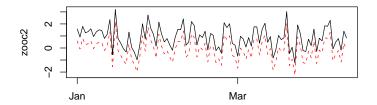
> tfplot(z + 1, TSget(c("mat2c1", "mat2c2"), con), lty = c("solid", "dashed"), col = c("black", "red"))

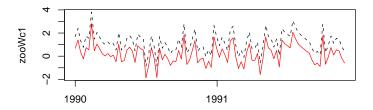


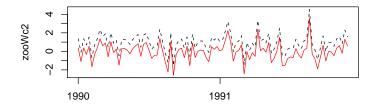


The following examples use dates and times which are not handled by ts, so the zoo time representation is used.









> dbDisconnect(con)

3 Examples Using Web Data

This section illustrates fetching data from a web server and loading it into the database. This would be a very slow way to load a database, but provides examples of different kinds of time series data. The fetching is done with TShistQuote which provides a wrapper for get.hist.quote from package tseries to give syntax consistent with the TSdbi.

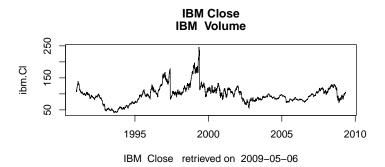
Fetching data may fail due to lack of an Interenet connection or delays. First establish a connection to the database where data will be saved:

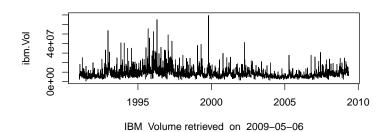
Now connect to the web server and fetch data:

```
> require("TShistQuote")
> Yahoo <- TSconnect("histQuote", dbname = "yahoo")
> x <- TSget("^gspc", quote = "Close", con = Yahoo)
> plot(x)
> tfplot(x)
> TSrefperiod(x)
```

```
[1] "Close"
> TSdescription(x)
[1] "^gspc Close from yahoo"
> TSdoc(x)
[1] "^gspc Close from yahoo retrieved 2009-05-06 22:13:58"
> TSlabel(x)
[1] "^gspc Close"
  Then write the data to the local server, specifying table B for business day
data (using TSreplace in case the series is already there from running this ex-
ample previously):
> TSreplace(x, serIDs = "gspc", Table = "B", con = con)
[1] TRUE
  and check the saved version:
> TSrefperiod(TSget(serIDs = "gspc", con = con))
[1] 1
> TSdescription("gspc", con = con)
[1] "^gspc Close from yahoo"
> TSdoc("gspc", con = con)
[1] "^gspc Close from yahoo retrieved 2009-05-06 22:13:58"
> TSlabel("gspc", con = con)
[1] NA
> tfplot(TSget(serIDs = "gspc", con = con))
```

```
> x <- TSget("ibm", quote = c("Close", "Vol"), con = Yahoo)
> TSreplace(x, serIDs = c("ibm.Cl", "ibm.Vol"), con = con, Table = "B",
      TSdescription. = c("IBM Close", "IBM Volume"), TSdoc. = paste(c("IBM Close"))
          "IBM Volume retrieved on "), Sys.Date()))
[1] TRUE
> z <- TSget(serIDs = c("ibm.Cl", "ibm.Vol"), TSdescription = TRUE,
      TSdoc = TRUE, con = con)
> TSdescription(z)
[1] "IBM Close"
                  "IBM Volume"
> TSdoc(z)
[1] "IBM Close
                  retrieved on 2009-05-06"
[2] "IBM Volume retrieved on 2009-05-06"
> tfplot(z, xlab = TSdoc(z), Title = TSdescription(z))
> tfplot(z, Title = "IBM", start = "2007-01-01")
```





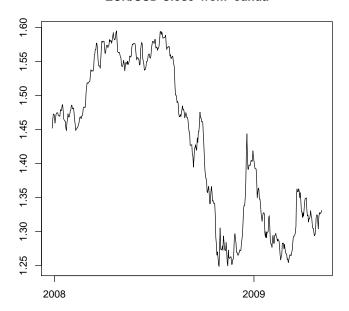
Oanda has maximum of 500 days, so the start date is specified here so as to not exceed that.

```
> Oanda <- TSconnect("histQuote", dbname = "oanda")
> x <- TSget("EUR/USD", start = Sys.Date() - 495, con = Oanda)
> TSreplace(x, serIDs = "EUR/USD", Table = "D", con = con)
```

[1] TRUE

Then check the saved version:

EUR/USD Close from oanda



- > dbDisconnect(con)
- > dbDisconnect(Yahoo)
- > dbDisconnect(Oanda)

3.1 Examples Using TSdbi with ets

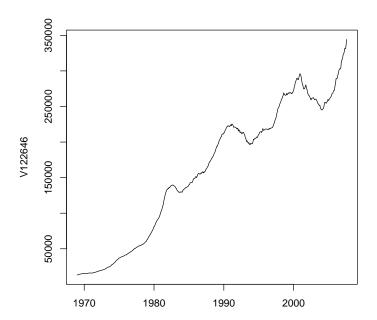
These examples use a database called "ets" which is available at the Bank of Canada. This set of examples illustrates how the programs might be used if a larger database is available. Typically a large database would be installed using database scripts directly rather than from R with *TSput* or *TSreplace*.

The following are wrapped in *if* (!inherits(conets, "try-error")) so that the vignette will build even when the database is not available. This seems to require an explicit call to print(), but that is not usually needed to display results below. Another artifact of this is that results printed in the if block do not display until the end of the block.

```
"M.CAN.CCUSMA02.ST", "CAN/USD exchange rate", "M.MEX.CCUSMA02.ST",
         "MEX/USD exchange rate", "M.JPN.CCUSMA02.ST", "JPN/USD exchange rate",
         "M.EMU.CCUSMA02.ST", "Euro/USD exchange rate", "M.OTO.CCUSMA02.ST",
         "OECD /USD exchange rate", "M.G7M.CCUSMA02.ST", "G7 /USD exchange rate",
         "M.E15.CCUSMA02.ST", "Euro 15. /USD exchange rate"),
         2, 8))
     print(TSdates(EXCH.IDs[, 1]))
     z <- TSdates(EXCH.IDs[, 1])</pre>
     print(start(z))
     print(end(z))
     tfplot(TSget(serIDs = "V122646", conets))
An object of class âĂIJTSmetaâĂ
Slot "TSdescription":
[1] "Special Drawing Right---Currency Conversions/US$ exchange rate/Average of daily rates/N
Slot "TSdoc":
[1] "Special Drawing Right---Currency Conversions/US$ exchange rate/Average of daily rates/N
Slot "TSlabel":
[1] NA
Slot "serIDs":
[1] "M.SDR.CCUSMA02.ST"
Slot "conType":
[1] "TSodbcConnection"
attr(,"package")
[1] "TSodbc"
Slot "DateStamp":
[1] NA
Slot "dbname":
[1] "ets"
Slot "hasVintages":
[1] FALSE
Slot "hasPanels":
[1] FALSE
     [,1]
[1,] "M.SDR.CCUSMA02.ST from 1960 1 to 2007 9 1
```

```
[2,] "M.CAN.CCUSMA02.ST from 1960 1 to 2007 9 1
[3,] "M.MEX.CCUSMA02.ST from 1963 1 to 2007 9 1
                                                   NA
[4,] "M.JPN.CCUSMA02.ST from 1960 1 to 2007 9 1
                                                   NA
[5,] "M.EMU.CCUSMA02.ST from 1979 1 to 2007 9 1
                                                   NA
[6,] "M.OTO.CCUSMA02.ST not available"
[7,] "M.G7M.CCUSMA02.ST not available"
[8,] "M.E15.CCUSMA02.ST not available"
[[1]]
[1] 1960
[[2]]
[1] 1960
            1
[[3]]
[1] 1963
            1
[[4]]
[1] 1960
[[5]]
[1] 1979
            1
[[6]]
[1] NA
[[7]]
[1] NA
[[8]]
[1] NA
[[1]]
[1] 2007
[[2]]
[1] 2007
            9
[[3]]
[1] 2007
[[4]]
[1] 2007
            9
[[5]]
[1] 2007
            9
```

```
[[6]]
[1] NA
[[7]]
[1] NA
[[8]]
[1] NA
```

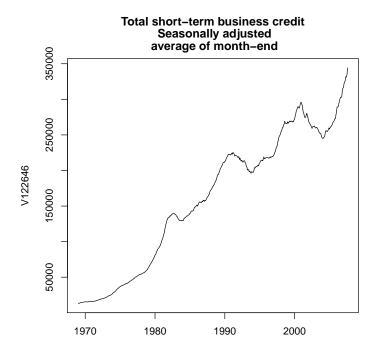


```
> if (!inherits(conets, "try-error")) {
    print(TSdescription(TSget("V122646", TSdescription = TRUE)))
    print(TSdescription("V122646"))
    print(TSdoc(TSget("V122646", TSdoc = TRUE)))
    print(TSdoc("V122646"))
    tfplot(TSget("V122646", names = "V122646", conets))
}

[1] "Total short-term business credit, Seasonally adjusted, average of month-end"
[1] "Total short-term business credit, Seasonally adjusted, average of month-end"
[1] "Same as B171"
[1] "Same as B171"
```

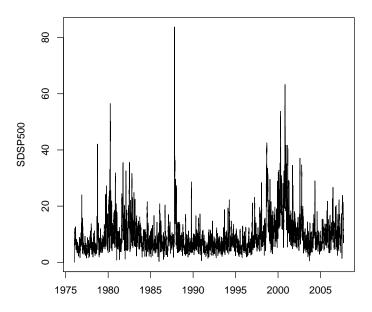
```
0000 92 0000 1900 0000 1900 2000 1900 2000 1900 2000
```

```
> if (!inherits(conets, "try-error")) {
    z <- TSget("V122646", TSdescription = TRUE)
    tfplot(z, Title = strsplit(TSdescription(z), ","))
}</pre>
```

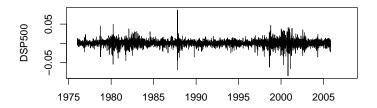


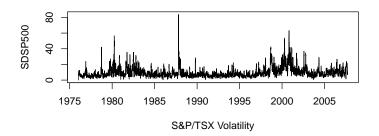
```
> if (!inherits(conets, "try-error")) {
    z <- TSget("SDSP500", TSdescription = TRUE)
    tfplot(z, Title = TSdescription(z))
    plot(z)
}</pre>
```

S&P/TSX Volatility



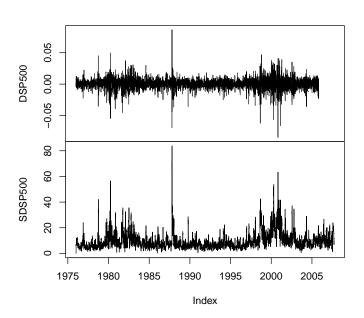
```
> if (!inherits(conets, "try-error")) {
    z <- TSget(c("DSP500", "SDSP500"), TSdescription = TRUE)
    tfplot(z, xlab = TSdescription(z))
}</pre>
```

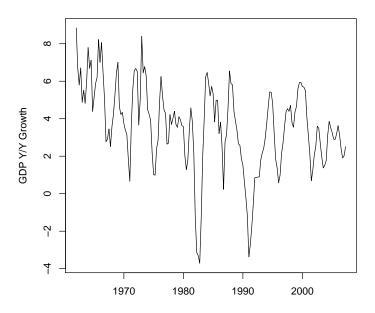




```
> if (!inherits(conets, "try-error")) {
     plot(z)
}
```

z





```
> if (!inherits(conets, "try-error")) {
     dbDisconnect(options()$TSconnection)
     options(TSconnection = NULL)
}
```

4 Examples Using DBI and direct SQL Queries

The following examples are queries using direct SQL queries. They should not often be needed to access time series, but may be useful to get at more detailed information, or formulate special queries. Some databases support special calls to access database or table information, but the following try to use generic SQL.

If schema queries are supported then table information can be obtained in a generic SQL way, but on some systems this will fail because users do not have read priveleges on the INFORMATION_SCHEMA table, so the following are wrapped in tryCatch(). (SQLite does not seem to support this at all.)

NULL

> tryCatch(dbGetQuery(con, paste("SELECT COLUMN_NAME, COLUMN_DEFAULT, COLLATION_NAME, DATA_".

"CHARACTER_SET_NAME, CHARACTER_MAXIMUM_LENGTH, NUMERIC_PRECISION",

"FROM INFORMATION_SCHEMA.Columns WHERE TABLE_SCHEMA='test' AND table_name='A';")))

NULL

> tryCatch(dbGetQuery(con, paste("SELECT COLUMN_NAME, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH, I
"FROM INFORMATION_SCHEMA.Columns WHERE TABLE_SCHEMA='test' AND table_name='M';")))

NULL

Finally, to disconnect gracefully, one should

- > dbDisconnect(con)
- > options(TSconnection = NULL)
- > odbcCloseAll()
- > dbUnloadDriver(m)

5 Example ODBC configuration file

Following is an example ODBC configuration file I use in Linux (so the file is in my home directory and called ".odbc.ini") to connect to a remote PostgreSQL server:

[test]

Description = test DB (Postgresql)

Driver = Postgresql

Trace = No

TraceFile = /tmp/test_odbc.log

 Database
 = test

 Servername
 = some.host

 UserName
 = paul

 Password
 = mySecret

 Port
 = 5432

 Protocol
 = 6.4

 ReadOnly
 = No

RowVersioning = No ShowSystemTables = No ShowOidColumn = No FakeOidIndex = No ConnSettings =

[ets]

Description = ets DB (Postgresql)

Driver = Postgresql

Trace = No

TraceFile = /tmp/test_odbc.log

Database = ets Servername = some.host UserName = paul Password = mySecret Port = 5432 Protocol = 6.4 ReadOnly = No = No RowVersioning ShowSystemTables = No = No ShowOidColumn FakeOidIndex = No ConnSettings