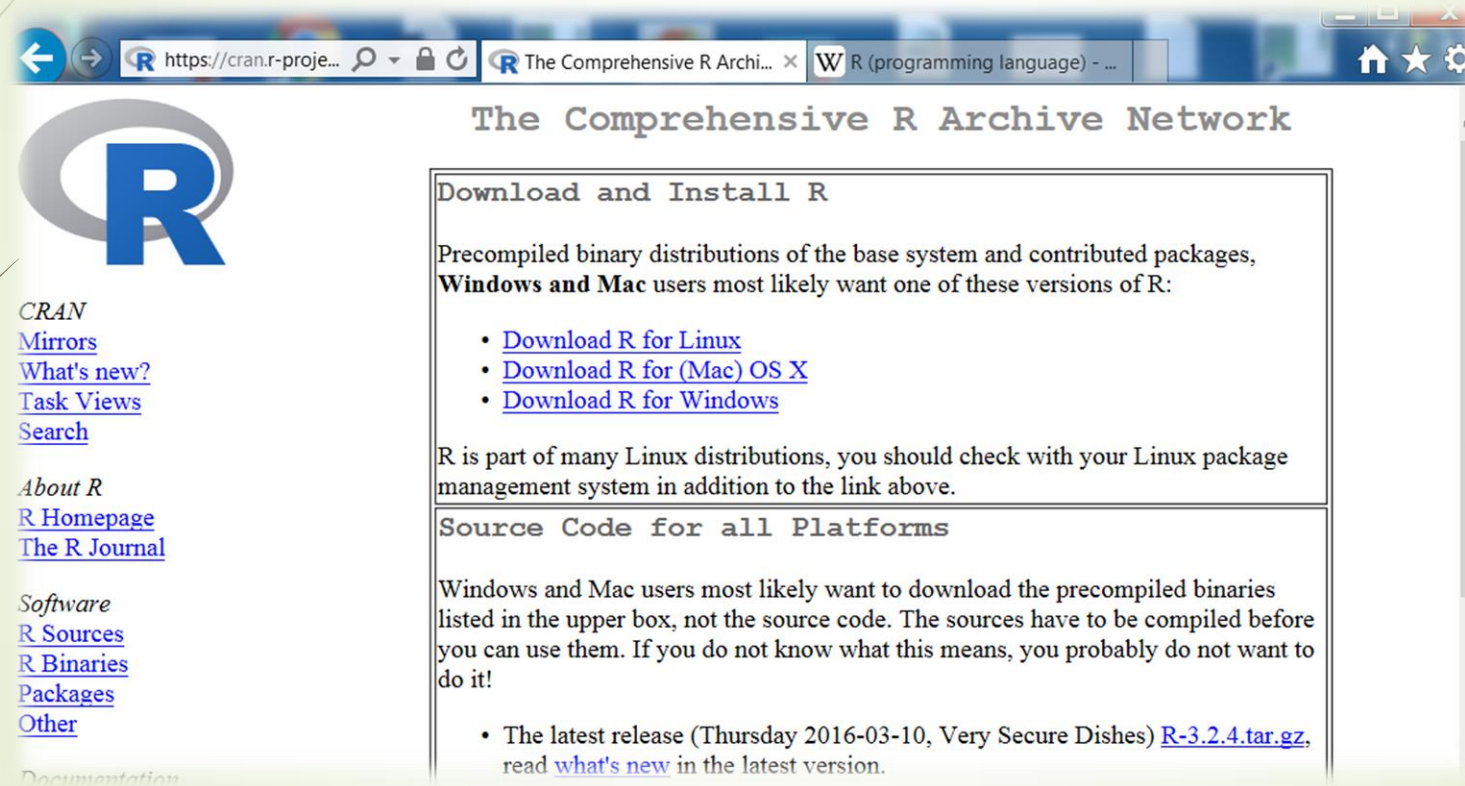# Session Topics

- R History
- R Primer
    - Purposes
    - Basic Commands
    - Command Line and Batch Mode
    - Data Structures
    - How to get Data into R
- Ways of Connecting Directly to Oracle from R
    - RODBC
    - RJDBC
    - ROracle
- Open Source R and Enterprise versions (Oracle R)
- Demo

What is R?

- R is a popular open source programming language and environment used by data scientists, data miners and statisticians to analyze data

- Supported by the R Foundation

- Functional Programming (FP) language ; Object oriented

- GNU License

- Evolved from the "S" Programming language

- R was created by **Ross Ihaka and Robert Gentleman**

- Runs on multiple platforms: MAC / Windows / UNIX

- You can get it here: *https://cran.r-project.org/*

CRAN – "Comprehensive R Archive Network"

*https://**cran**.r-project.org/*

There are thousands of packages that you can download from CRAN.
R is highly extensible.

R Console

R Studio

https://www.rstudio.com/products/rstudio/download/
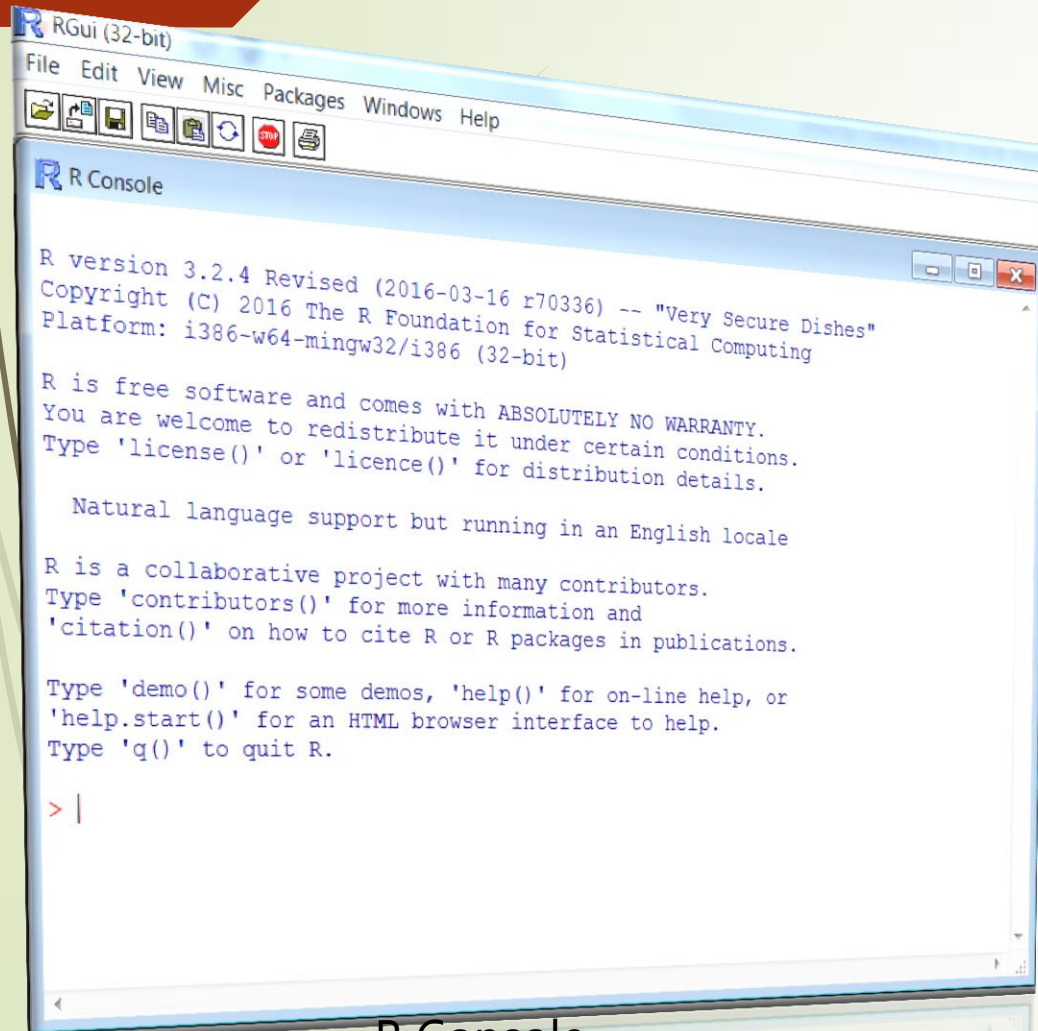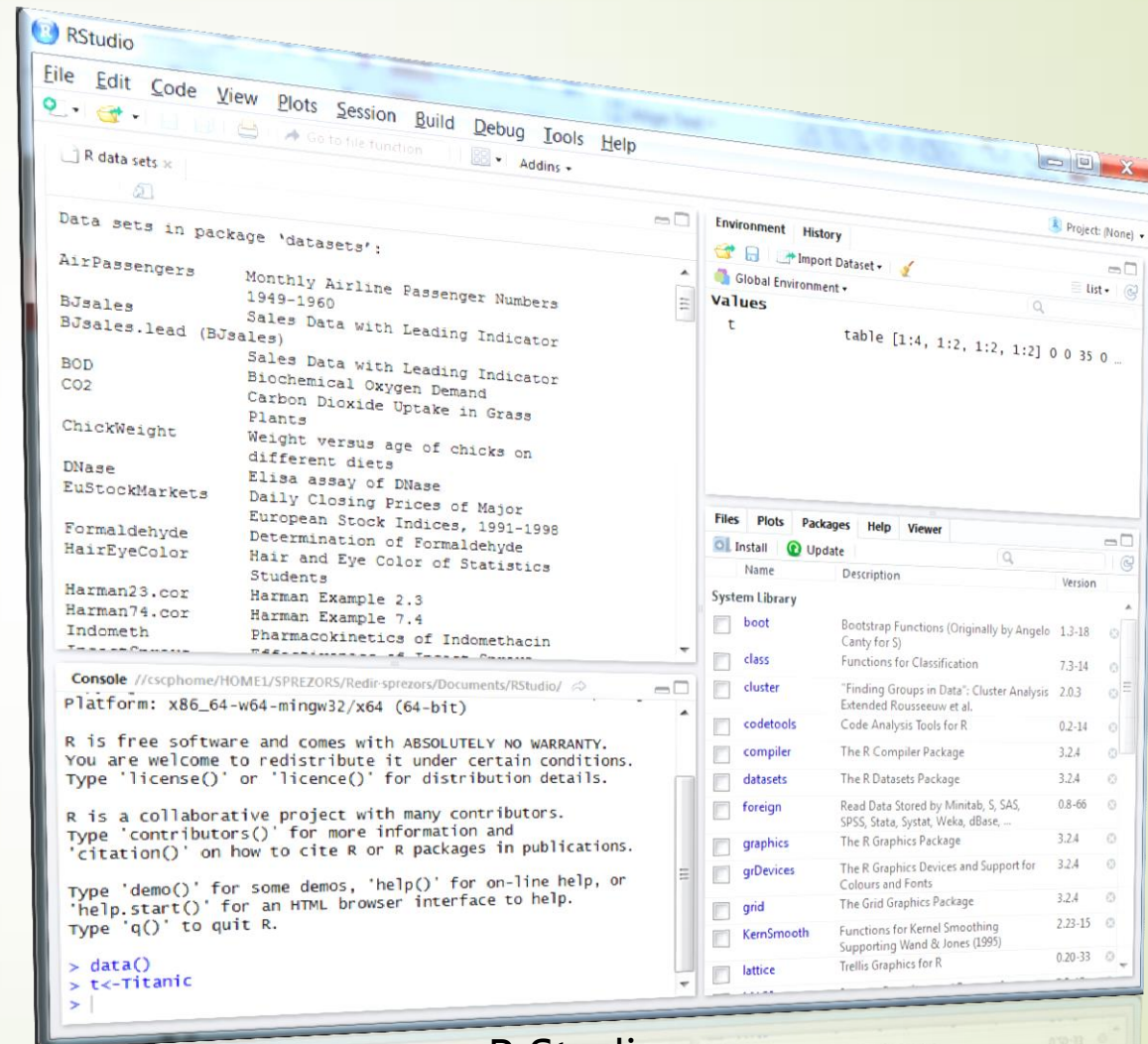
What can I do with R?

- Data Analysis
- Data Visualizations
- Descriptive and Predictive Modeling
- Data Mining
- Non-Statistical Uses

A Topographic Map of Maunga Whau
10 Meter Contour Spacing

Edgar Anderson's Iris Data
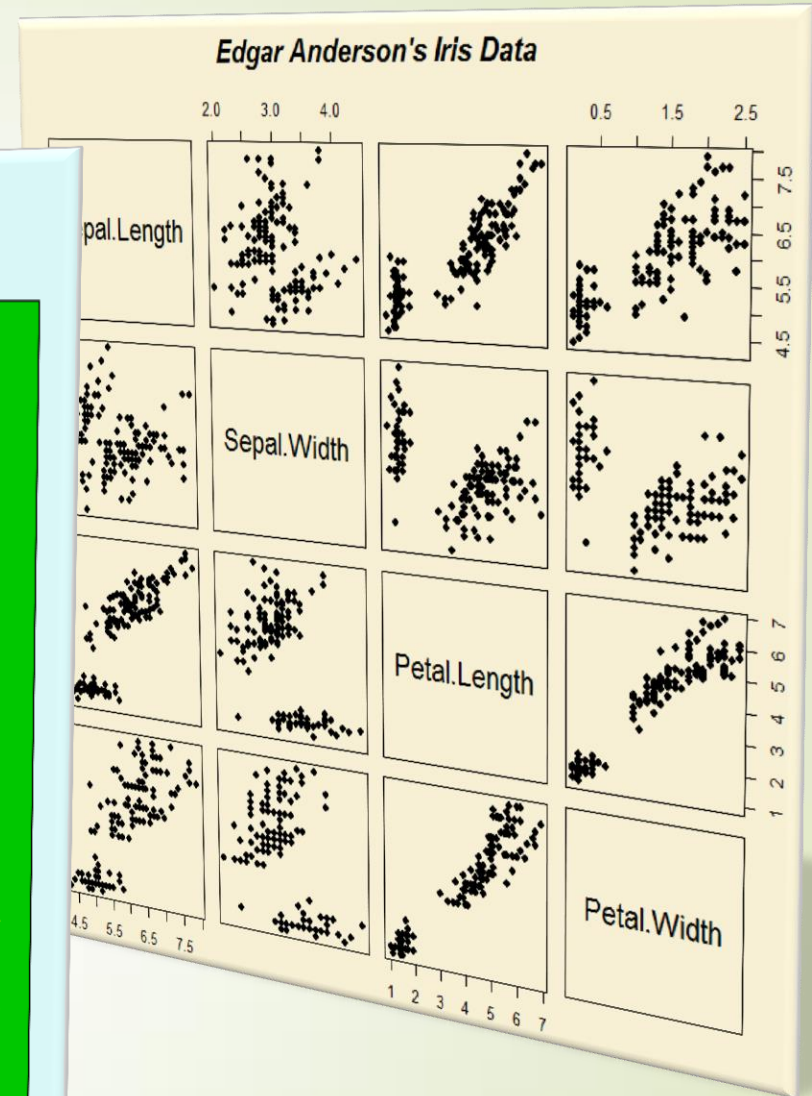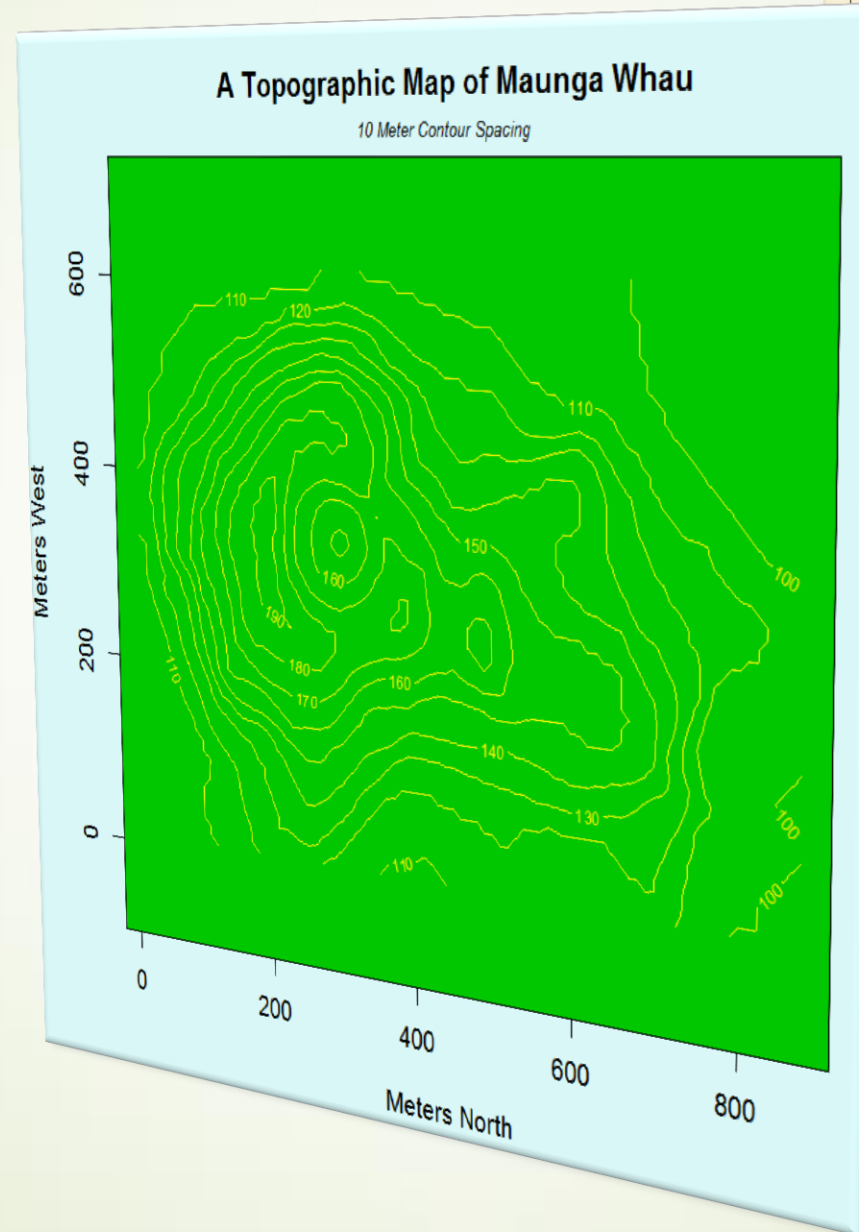
Graphs produced by
Demo package included with R.
demo()

R Basic Commands
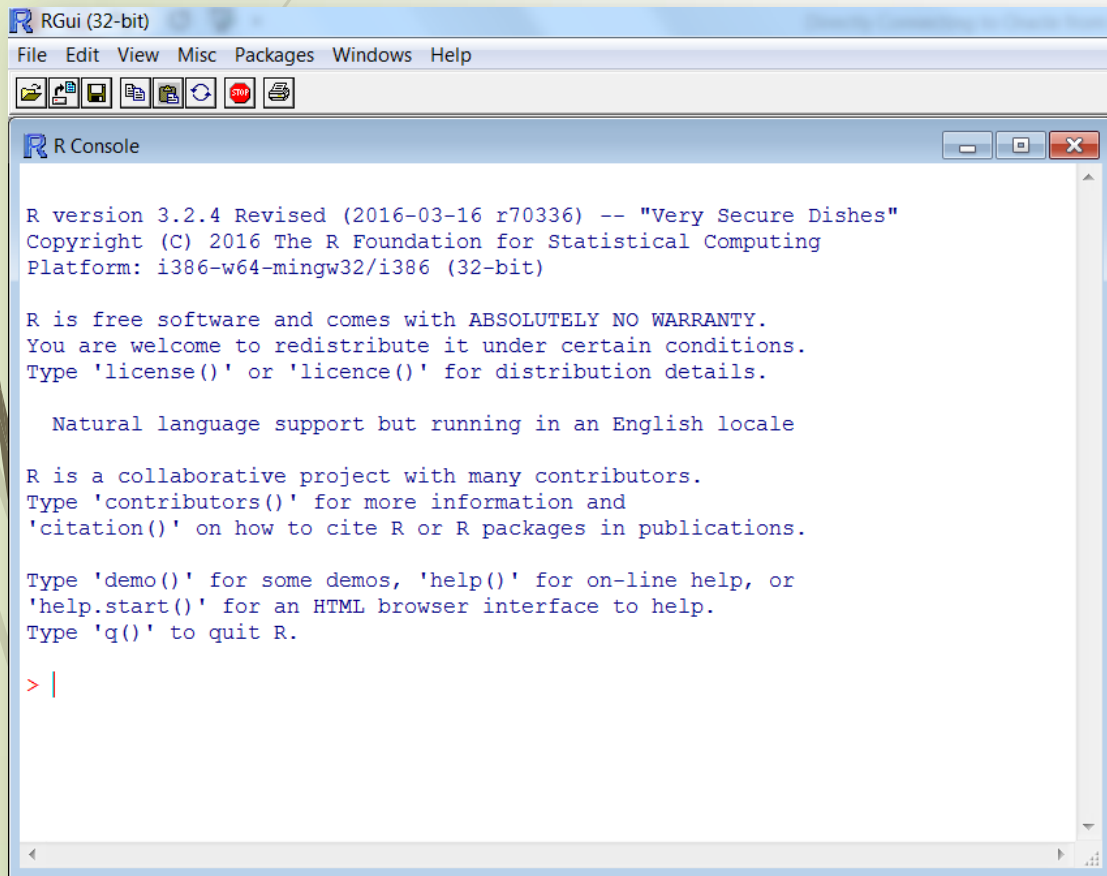
"Everything that exists in an object ; Everything that happens is a function call." - John Chambers

Command	Purpose
ls()	Lists objects in a workspace
help()	Get help on a function, include function name: ex. help(version)
getwd()	Get working directory
setwd("c:/temp")	
install.packages("RODBC")	How to install a package
rm()	Remove an object from your workspace
library(RODBC)	Use a library once it is installed
summary()	Show basic stats on an object
q()	Leave R session

Caution: R is case sensitive!

R – Interactive Mode and Batch Mode

- R has an interactive mode (CLI) and a mode where you can create R scripts and run programs.

A Few Basic R Data Structures : Vectors & Data Frames

- A vector is a one dimensional array of one data type.
  You can have for example a vector of numeric data or string, but not a mixture.

  Some examples:
  MyFavoriteColors<-c("blue","purple","red")
  LuckyLotteryNumbers<-c(14,32,64)

Environment | History

Import Dataset | Clear

Global Environment

values

LuckyLotteryNumbers	num [1:3] 14 32 64
MyFavoriteColors	chr [1:3] "blue" "purple" "red"

- A data frame is a collection of vectors that have all the same length

	tourney_id	tourney_name	surface	draw_size	tourney_level	tourney_date	match_num	winner_id	winner_seed	winner_entry	winn
1	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	1	201468	NA	Q	Yaro
2	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	2	200002	NA		Mirj
3	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	3	201521	7		Carl
4	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	4	201493	3		Ange
5	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	5	201611	NA	Q	Dari
6	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	6	202494	NA		Elin
7	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	7	201548	NA	WC	Ajla
8	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	8	201485	NA	LL	Alla
9	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	9	201662	NA		Karo
10	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	10	201478	NA		Varv
11	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	11	201619	NA		Madi
12	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	12	201466	NA		Kaia
13	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	13	201483	NA	Q	Madi
14	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	14	201457	NA	WC	Jarm
15	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	15	201345	1		Mari
16	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	16	201521	7		Carl
17	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	17	201493	3		Ange
18	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	18	202494	NA		Elin
19	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	19	201485	NA	LL	Alla
20	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	20	201478	NA		Varv
21	2015-W-PR-AUS-01A-2015	Brisbane International	Hard	30	P	20150105	21	201466	NA		Kaia

> str(wta_tennis_matches_2016) 'data.frame':
2651 obs. of 49 variables:

One Method to Import data into R

→ There are a variety of ways to load data into R depending on the format of the data. Here is how to import data in a common .csv file format:

wta_tennis_matches_2016=**read.csv**("c:/nyoug/wta_tennis_data.csv")

The data is loaded into a data frame in the R work space.

```
> ls()
 [1] "cnames"        "dat"                  "doc"        "fileURL"
 [5] "g"             "LuckyLotteryNumbers"  "my_char"    "my_data"
 [9] "my_matrix"     "my_matrix2"           "my_na"      "my_name"
[13] "my_vector"     "MyFavoriteColors"     "num_vect"   "patients"
[17] "q"             "q1"                   "rootNode"   "simple"
[21] "tf"            "the_zipcodes"         "url"        "vect"
[25] "vect2"         "wta_tennis_matches_2016" "y"        "z"
[29] "zc"
```

Connecting R to an Oracle Database Directly

R Packages for Oracle Connections

RODBC | RJDBC | ROracle

Using R to Connect to a database using the RODBC package

- \> install.packages("RODBC")

- \> library(RODBC)

- \> myconn <-odbcConnect("demo", uid="demousr", pwd="demopwd")

  \> demodat <- sqlFetch(myconn, "USER_TABLES")

ODBC Data Source Name

"demodat" is an R object, a data frame.
The entire USER_TABLES table is now
loaded in R.

ODBC Data Source Administrator (64-bit)

User DSN | System DSN | File DSN | Drivers | Tracing | Connection Pooling | About

User Data Sources:

Name	Platform	Driver
dBASE Files	64-bit	Microsoft Access dBASE Driver (*.dbf, *.ndx, *.mdx)
demo	64-bit	Oracle in OraDB12Home1
Excel Files	64-bit	Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlsb)
MS Access Database	64-bit	Microsoft Access Driver (*.mdb, *.accdb)

An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you and can only be used on this computer.

OK | Cancel | Apply | Help

RODBC Functions Overview

- Connect to Oracle Database: **odbcConnect**()

- Query data via SELECT statements
sqlquery(channel,query)

- Insert sqlSave() / or update

- Drop: **sqlDrop**(channel, sqtable, errors = TRUE)

- Set Autocommit

- Can copy data from one database to another
sqlCopy()

- Close the connection when done
odbcClose()

RODBC: ODBC Database Access

An ODBC database interface.

Version:	1.3-12
Depends:	R (≥ 3.0.0)
Imports:	stats
Published:	2015-06-29
Author:	Brian Ripley [aut, cre], Michael Lapsley [aut] (1999 to Oct 2002)
Maintainer:	Brian Ripley <ripley at stats.ox.ac.uk>
License:	GPL-2	GPL-3
NeedsCompilation:	yes
SystemRequirements:	An ODBC3 driver manager and drivers.

https://cran.r-project.org/web/packages/RODBC/RODBC.pdf
 Reference Manual or **RShowDoc**("RODBC", package="RODBC")

Using R to Connect to a database using the ROracle Package

- > library("ROracle")

- > drv<-dbDriver("Oracle")

                               User id   password

- > con<-dbConnect(drv,"SYSTEM","ROracle123")
  > demodat <- dbGetQuery(con,"select table_name from user_tables")

"demodat" is an R object, a data frame.
The results of the query above are
sent into the demodat object into the
R workspace.

http://www.oracle.com/technetwork/database/database-technologies/r/roracle/overview/index.html

The ROracle package is created and maintained by Oracle. You can download the package from Oracle.

Where to get the ROracle package

ROracle Functions Overview

- Connect to Oracle Database:
  dbConnect

- Query data via SELECT statements
  dbSendQuery(conn,query) / fetch()

- **dbGetQuery**(conn,query)

- Drop: **dbRemoveTable**(conn,<Table>)

- **dbWriteTable**(conn, <table name>,<obj>)

- Close the connection when done
  dbDisconnect(conn, …)

ROracle: OCI Based Oracle Database Interface for R

Oracle Database interface (DBI) driver for R. This is a DBI-compliant Oracle driver based on the OCI.

Version:	1.2-2
Depends:	methods, DBI ($\geq$ 0.2-5)
Imports:	utils
Published:	2016-02-17
Author:	Denis Mukhin, David A. James and Jake Luciani
Maintainer:	Rajendra S. Pingte <rajendra.pingte at oracle.com>
License:	LGPL-2 | LGPL-2.1 | LGPL-3 [expanded from: LGPL]
URL:	http://www.oracle.com
NeedsCompilation:	yes
SystemRequirements:	Oracle Instant Client or Oracle Database Client
Materials:	NEWS INSTALL
CRAN checks:	ROracle results

https://cran.r-project.org/web/packages/ROracle/ROracle.pdf Reference

RODBC V.S. RORACLE Packages

➤ RODBC implements Open Database Connectivity

➤ OCI, or "Oracle Call Interface", out performs ODBC, so RODBC will be slower than ROracle, which uses OCI and uses a native connection.

➤ Use ROracle for better performance. Performance lift increase depends on the data type of the data and other factors.

Performance Test

RODBC V.S. RORACLE

```
roracle_time<-function()
{
  #Function to time how long it takes to write a data frame into Oracle database
  library("ROracle")
  drv<-dbDriver("Oracle")

  # Start timing
  ptm <- proc.time()
  #Establish the connection...
  conn<-dbConnect(drv,"SYSTEM","ROracle123")

  #Write out wta_tennis_matches_2016 data frame in the R workspace into Oracle as table called: WTA_TENNIS
  dbWriteTable(conn,"WTA_TENNIS", wta_tennis_matches_2016)

  # Stop the clock
  timing<-proc.time() - ptm

  #Close the connection
  dbDisconnect(conn)

  #Print the timing results to the console:
  timing

}
```

```
> source('C:/nyoug/roracle_time.r')

> roracle_time()

   user system elapsed
>  0.11   0.01    0.55
```

Performance Test

RODBC V.S. RORACLE

```r
rodbc_time<-function()
{
  #Function to time how long it takes to write a data frame into Oracle database
  library(RODBC)

  # Start timing
  ptm <- proc.time()

  myconn<-odbcConnect("demo",uid="SYSTEM",pwd="ROracle123")

  sqlSave(myconn, wta_tennis_matches_2016)

   # Stop the clock
  timing<-proc.time() – ptm

  odbcClose(myconn)

  timing

}
```
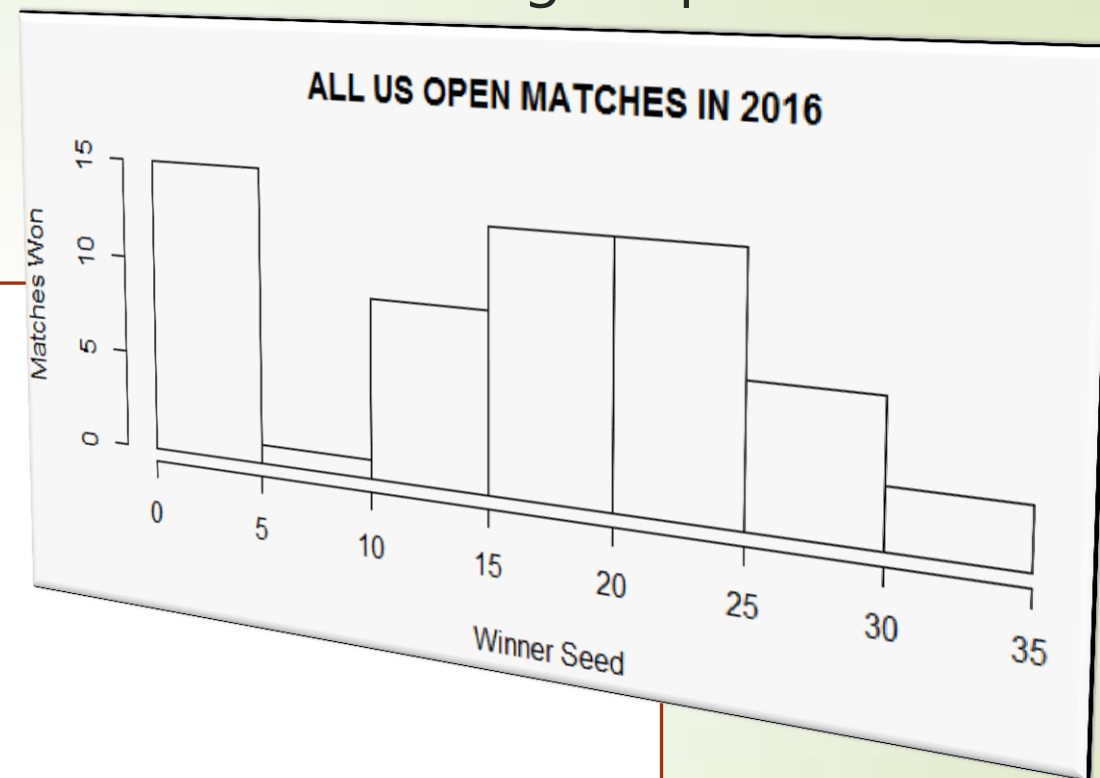
> source('~/rodbc_time.r')
> rodbc_time()
>
> user system elapsed
> 0.44 0.41   44.80

ALL US OPEN MATCHES IN 2016

```
hist_top_seeded_usopen<-function()

{

  library("ROracle")

  drv<-dbDriver("Oracle")

  #Establish the connection...

  conn<-dbConnect(drv,"SYSTEM","ROracle123")

  #Execute and fetch query and return to Q data frame:

  Q<-dbGetQuery(con,"select winner_seed from TENNIS_MATCHES_2016 where tourney_name='US Open' AND winner_seed is not null")

  dbDisconnect(conn)

  #Create a histogram of the data now inside R, give title name and axis titles:

  h<-hist(Q$WINNER_SEED,main="ALL US OPEN MATCHES IN 2016",xlab="Winner Seed",ylab="Matches Won")

  }
```

Oracle R Offerings:

Oracle R Distribution: Oracle created their own distribution of the R package.

Oracle R Enterprise: Integration of R

Oracle R Advanced Analytics for Hadoop

ROracle: package for high performance connection from R to Oracle databases

http://www.oracle.com/technetwork/database/database-technologies/r/r-technologies/r-offerings-1566363.html

Oracle R Enterprise

- R objects are stored in the oracle database, not in memory in the R workspace

- Offers parallel processing

- Helps resolve in-memory limitations and single threadness of open source R

- http://www.oracle.com/technetwork/database/database-technologies/r/r-enterprise/overview/index.html

DEMO Using ROBDC / ROracle

Environment:
>   Oracle 12c Enterprise Version / R 3.2.4 / Windows 8

```
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application Testing opt
ions
```

```
R version 3.2.4 Revised (2016-03-16 r70336) -- "Very Secure Dishes"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

Session Summary

➤ R is an open source language for statistics but can also be used as a general purpose data processing language.

➤ There are Enterprise versions available, including Oracle R Enterprise

➤ There are a variety of ways to get data into R.

➤ I can connect directly to an Oracle database by using RODBC / ROracle / RJDBC

➤ Use: RSEEK (rseek.org) : a specialized search engine to get more information.