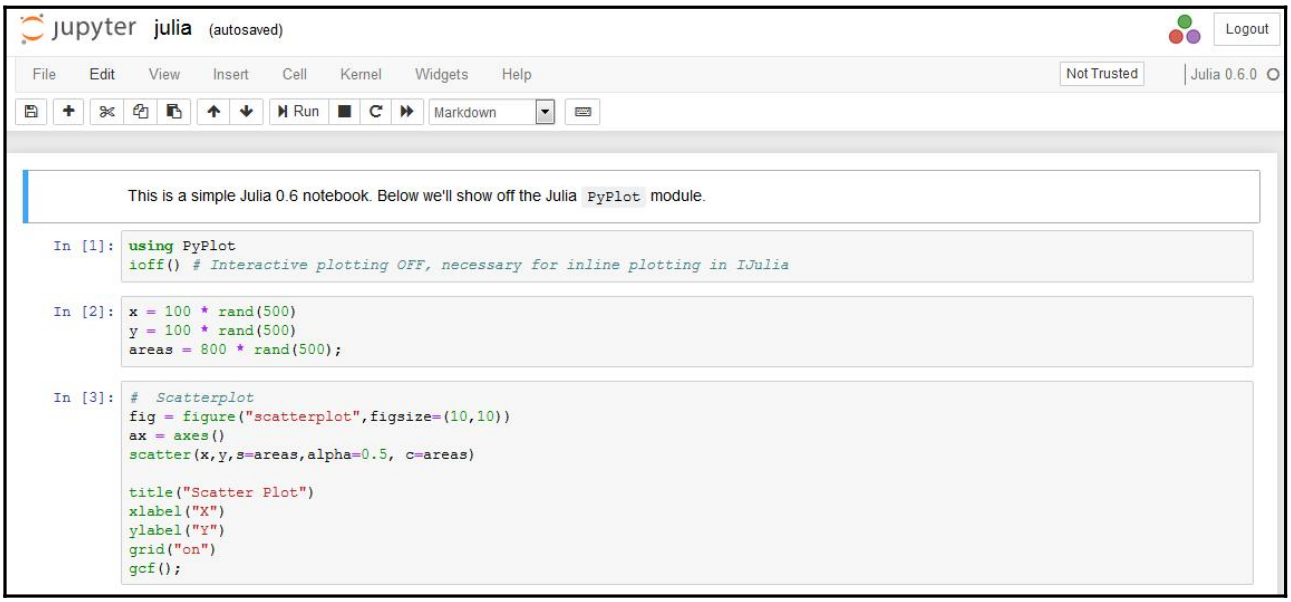
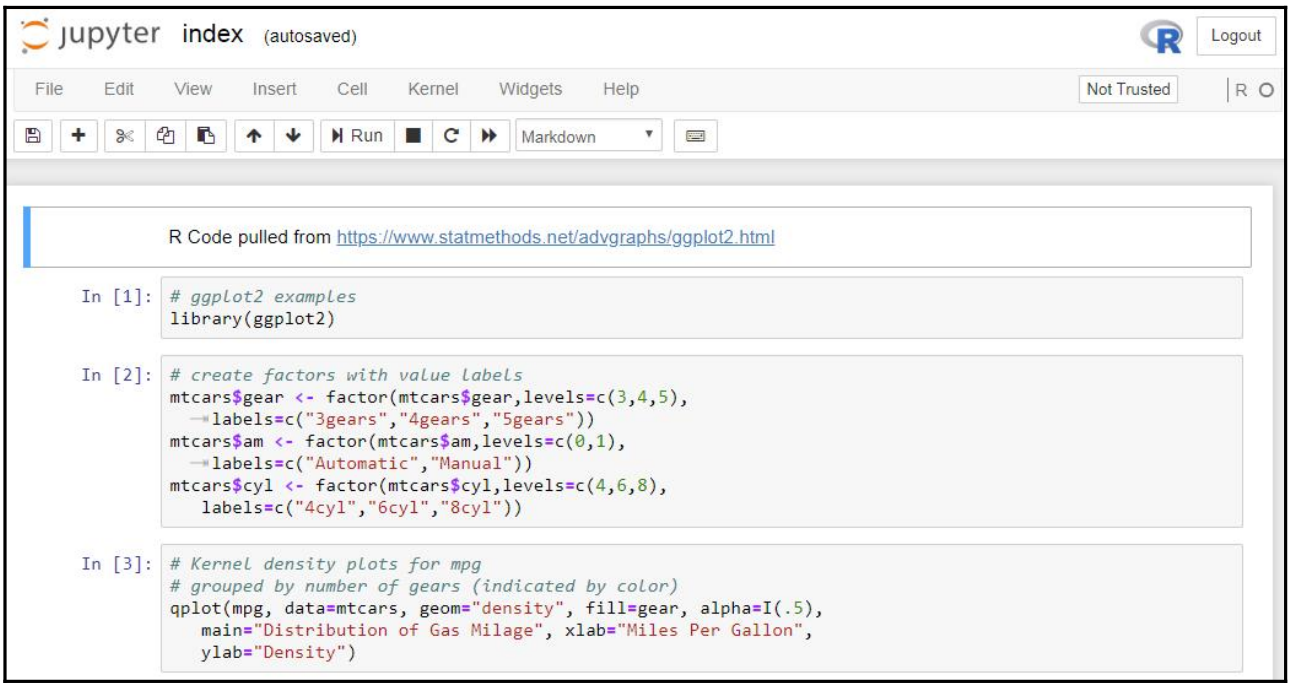
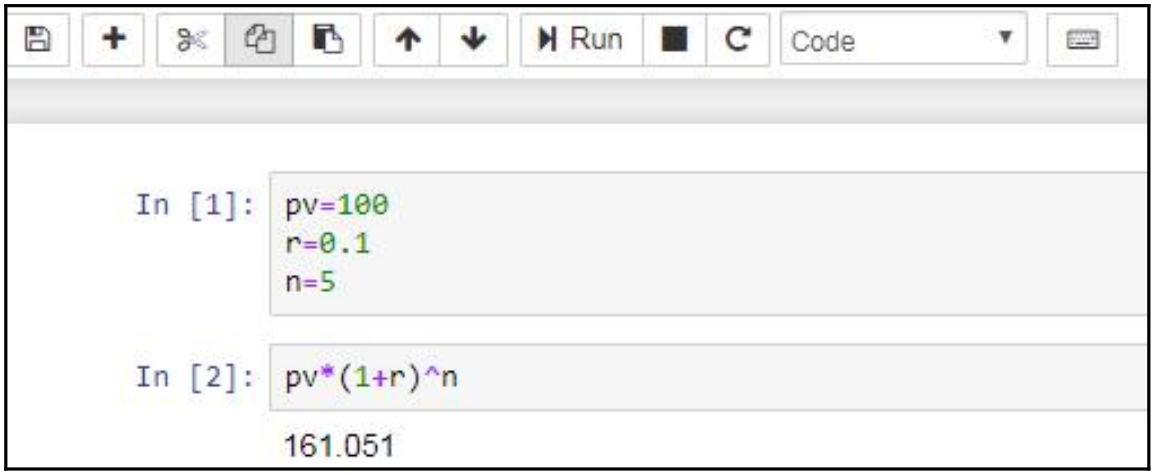
1

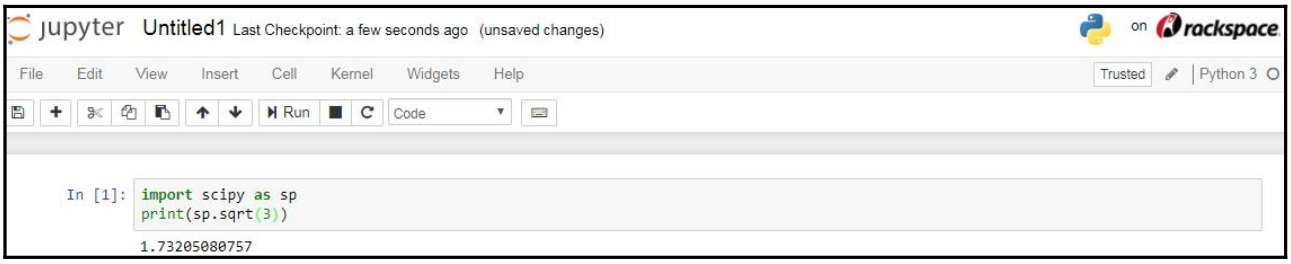
Chapter 01: Ecosystem of Anaconda



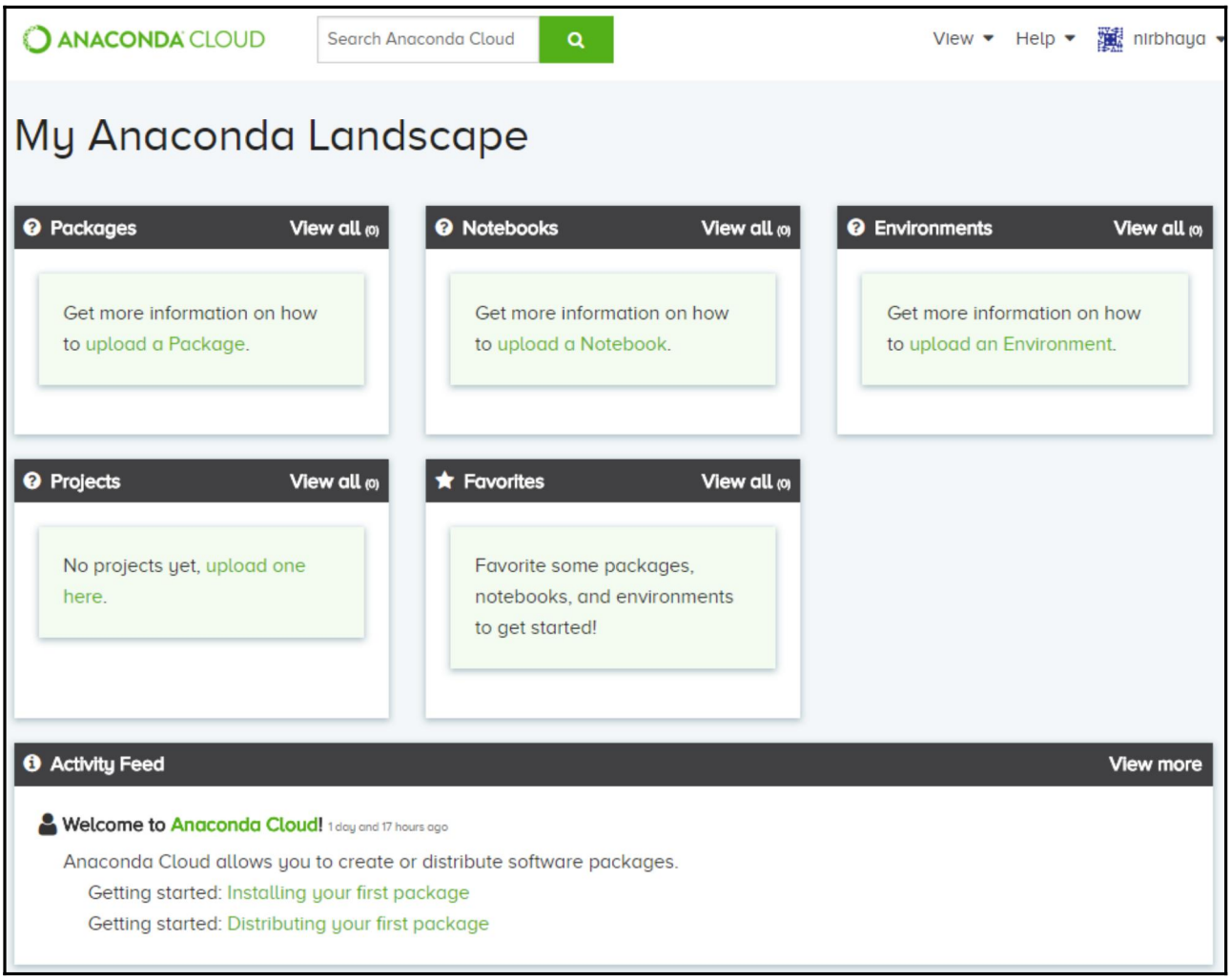

jupyter index (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Run
Markdown
R Code pulled from https://www.statmethods.net/advgraphs/ggplot2.html
In [1]:
# ggplot2 examples
library(ggplot2)
In [2]:
create factors with value labels
mtcars$gear <- factor(mtcars$gear,levels=c(3,4,5),
labels=c("3gears","4gears","5gears"))
mtcars$am <- factor(mtcars$am,levels=c(0,1),
labels=c("Automatic","Manual"))
mtcars$cyl <- factor(mtcars$cyl,levels=c(4,6,8),
labels=c("4cyl","6cyl","8cyl"))
In [3]:
Kernel density plots for mpg
grouped by number of gears (indicated by color)
qplot(mpg, data=mtcars, geom="density", fill=gear, alpha=I(.5),
main="Distribution of Gas Milage", xlab="Miles Per Gallon",
ylab="Density")

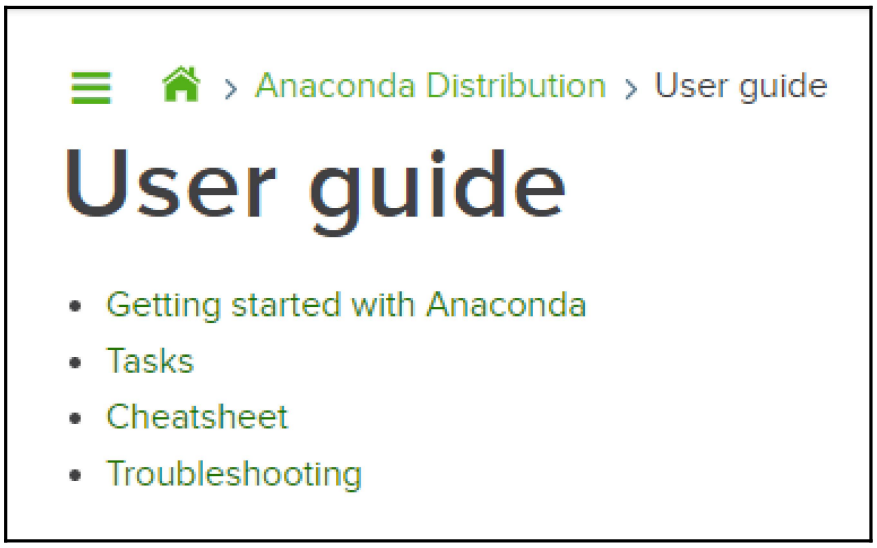

upyter Untitled Last Checkpoint: a few seconds ago (unsaved changes)
on rackspace
File Edit View Insert Cell Kernel Widgets Help
Trusted
Run
Code
In [ ]:
pv=100
r=0.1
n=5

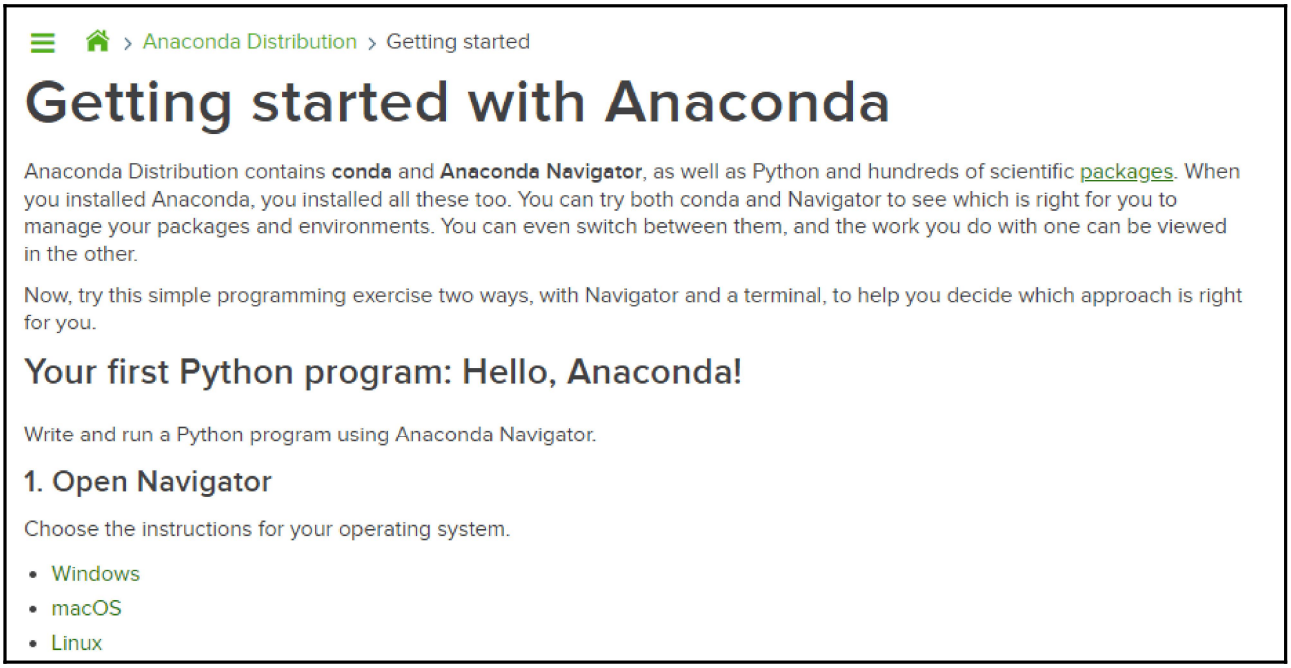

Run
Code
In [1]:
pv=100
r=0.1
n=5
In [2]:
pv*(1+r)^n
161.051

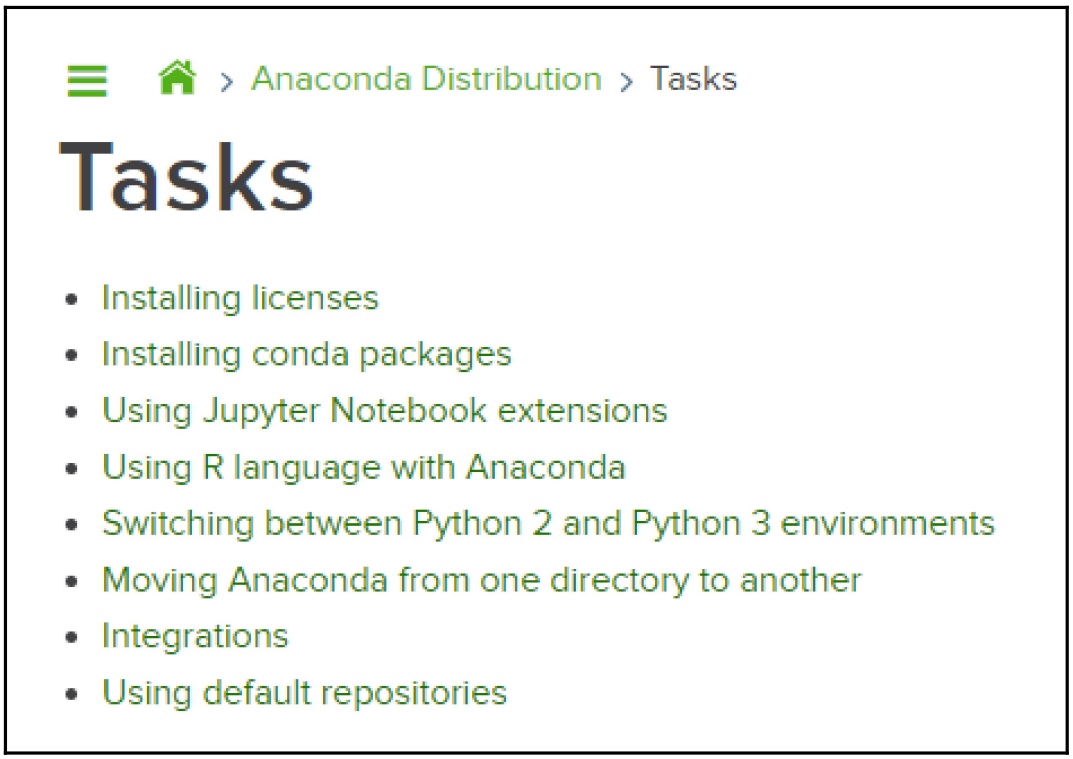

jupyter Untitled1 Last Checkpoint: a few seconds ago (unsaved changes)
on rackspace
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
Run Code
In [1]:
import scipy as sp
print(sp.sqrt(3))
1.73205080757


jupyter Untitled3 Last Checkpoint: a few seconds ago (unsaved changes)
on rackspace
File Edit View Insert Cell Kernel Widgets Help
Trusted Julia 0.6.0
Run Code
In [1]:
function sphere_vol(r)
return 4/3*pi*r^3
end
Out[1]: sphere_vol (generic function with 1 method)
In [2]:
sphere_vol(2.5)
Out[2]: 65.44984694978736


ANACONDA CLOUD
Search Anaconda Cloud
View
Help
nirbhaya
My Anaconda Landscape
Packages
View all (0)
Get more information on how to upload a Package.
Notebooks
View all (0)
Get more information on how to upload a Notebook.
Environments
View all (0)
Get more information on how to upload an Environment.
Projects
View all (0)
No projects yet, upload one here.
Favorites
View all (0)
Favorite some packages, notebooks, and environments to get started!
Activity Feed
View more
Welcome to Anaconda Cloud! 1 day and 17 hours ago
Anaconda Cloud allows you to create or distribute software packages.
Getting started: Installing your first package
Getting started: Distributing your first package


Anaconda Distribution > User guide
User guide
Getting started with Anaconda
Tasks
Cheatsheet
Troubleshooting


Anaconda Distribution > Getting started
Getting started with Anaconda
Anaconda Distribution contains conda and Anaconda Navigator, as well as Python and hundreds of scientific packages. When you installed Anaconda, you installed all these too. You can try both conda and Navigator to see which is right for you to manage your packages and environments. You can even switch between them, and the work you do with one can be viewed in the other.
Now, try this simple programming exercise two ways, with Navigator and a terminal, to help you decide which approach is right for you.
Your first Python program: Hello, Anaconda!
Write and run a Python program using Anaconda Navigator.
1. Open Navigator
Choose the instructions for your operating system.
Windows
macOS
Linux



> Anaconda Distribution > Tasks

Tasks

- Installing licenses
- Installing conda packages
- Using Jupyter Notebook extensions
- Using R language with Anaconda
- Switching between Python 2 and Python 3 environments
- Moving Anaconda from one directory to another
- Integrations
- Using default repositories

> Anaconda Distribution > Installing conda packages

Installing conda packages

For more information about using the conda package manager in Anaconda Prompt (Terminal on Linux or macOS), see the conda documentation.

You can also use the graphical interface Anaconda Navigator to install conda packages with just a few clicks.

Open an Anaconda Prompt (Terminal on Linux or macOS) and follow these instructions.

Installing a conda package

Enter the command:

```
conda install package-name
```

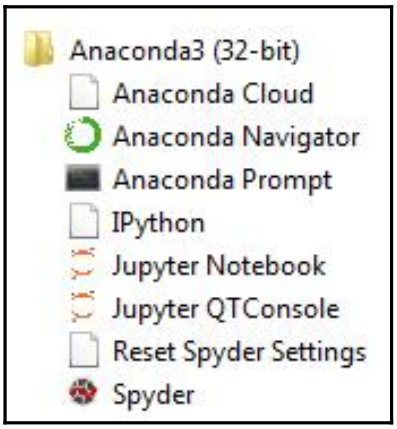
NOTE: Replace `package-name` with your package name.

Installing specific versions of conda packages

Include the desired version number or its prefix after the package name:

```
conda install package-name=2.3.4
```

Chapter 02: Anaconda Installation



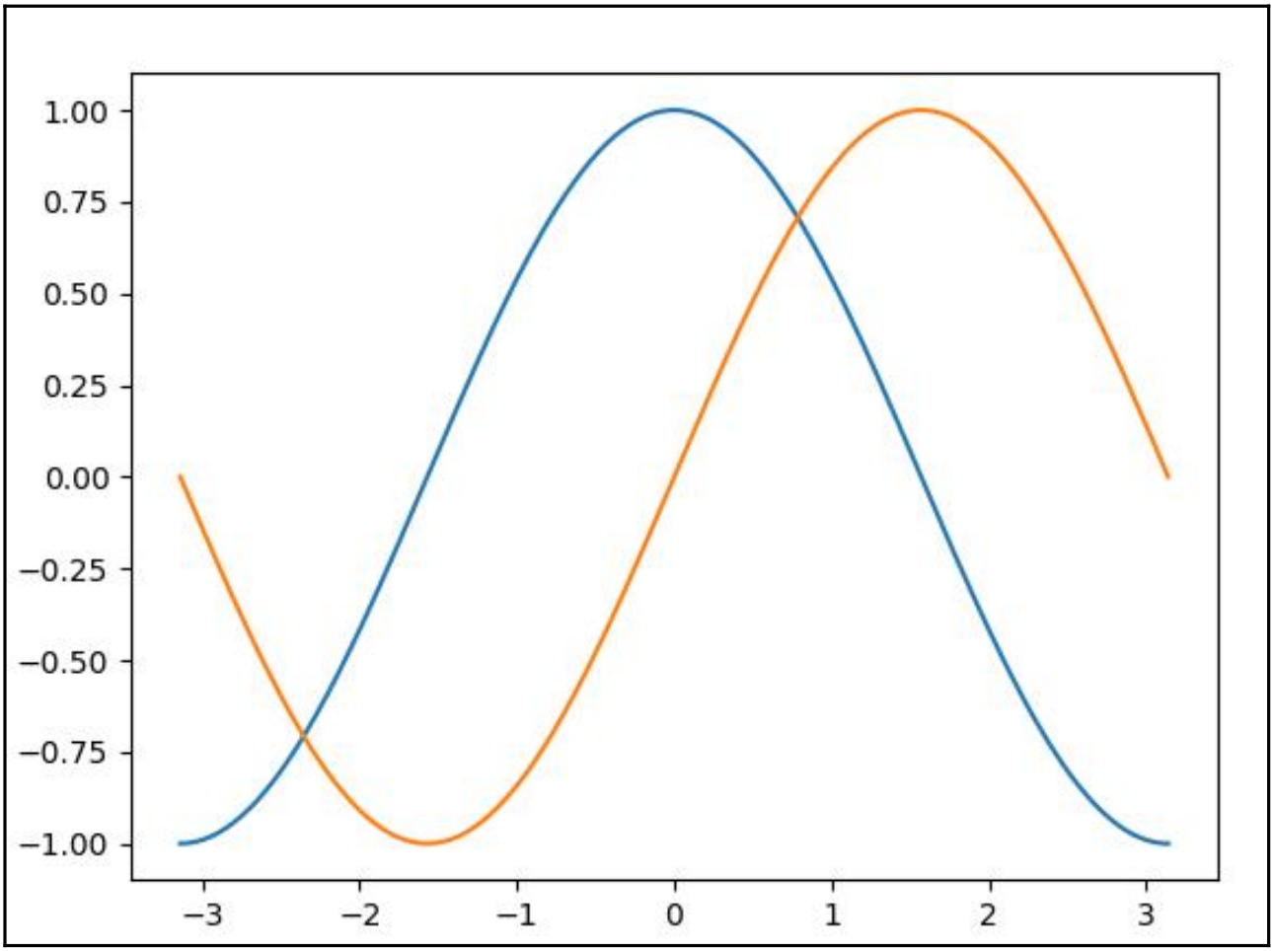
Anaconda Prompt

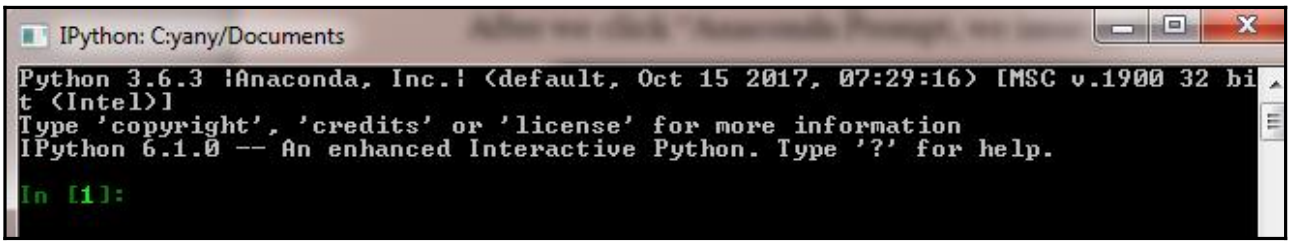
```
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>conda info
Current conda install:

               platform : win-32
          conda version : 4.3.30
       conda is private : False
      conda-env version : 4.3.30
    conda-build version : 3.0.27
         python version : 3.6.3.final.0
       requests version : 2.18.4
       root environment : C:\Users\yany\AppData\Local\Continuum\anaconda3  (writ
able)
    default environment : C:\Users\yany\AppData\Local\Continuum\anaconda3
       envs directories : C:\Users\yany\AppData\Local\Continuum\anaconda3\envs
                          C:\Users\yany\AppData\Local\conda\conda\envs
                          C:\Users\yany\.conda\envs
          package cache : C:\Users\yany\AppData\Local\Continuum\anaconda3\pkgs
                          C:\Users\yany\AppData\Local\conda\conda\pkgs
           channel URLs : https://repo.continuum.io/pkgs/main/win-32
                          https://repo.continuum.io/pkgs/main/noarch
                          https://repo.continuum.io/pkgs/free/win-32
                          https://repo.continuum.io/pkgs/free/noarch
                          https://repo.continuum.io/pkgs/r/win-32
                          https://repo.continuum.io/pkgs/r/noarch
                          https://repo.continuum.io/pkgs/pro/win-32
                          https://repo.continuum.io/pkgs/pro/noarch
                          https://repo.continuum.io/pkgs/msys2/win-32
                          https://repo.continuum.io/pkgs/msys2/noarch
            config file : None
              netrc file : None
           offline mode : False
             user-agent : conda/4.3.30 requests/2.18.4 CPython/3.6.3 Windows/7 W
indows/6.1.7601
          administrator : False
```

```
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>
```

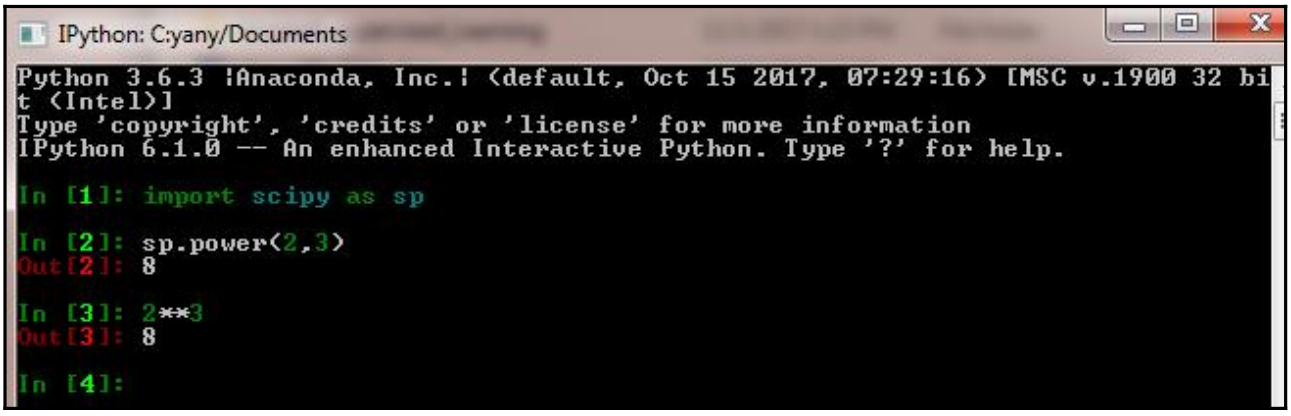
```
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>python
Python 3.6.3 |Anaconda, Inc.| (default, Oct 15 2017, 07:29:16) [MSC v.1900 32 bi
t (Intel)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

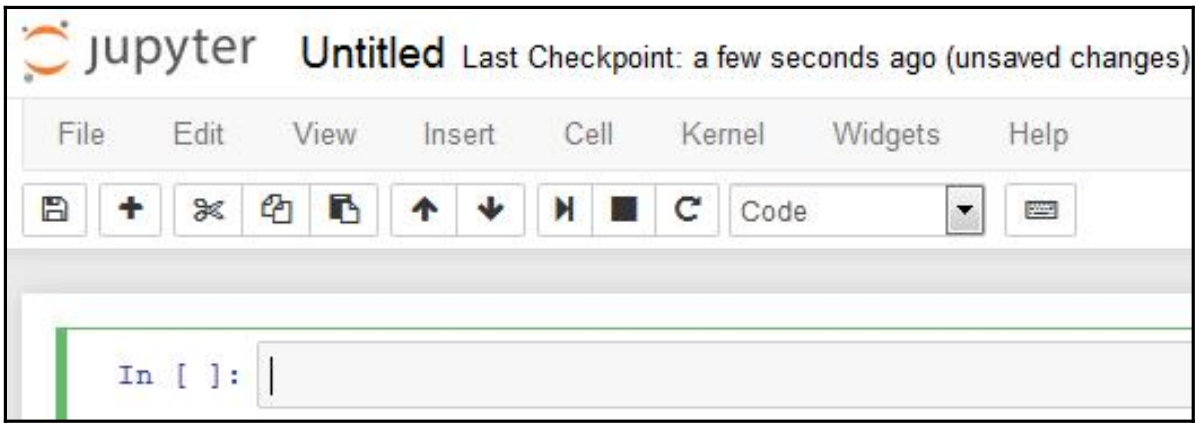
```
>>> import scipy as sp
>>> sp.sqrt(3)
1.7320508075688772
>>>
```





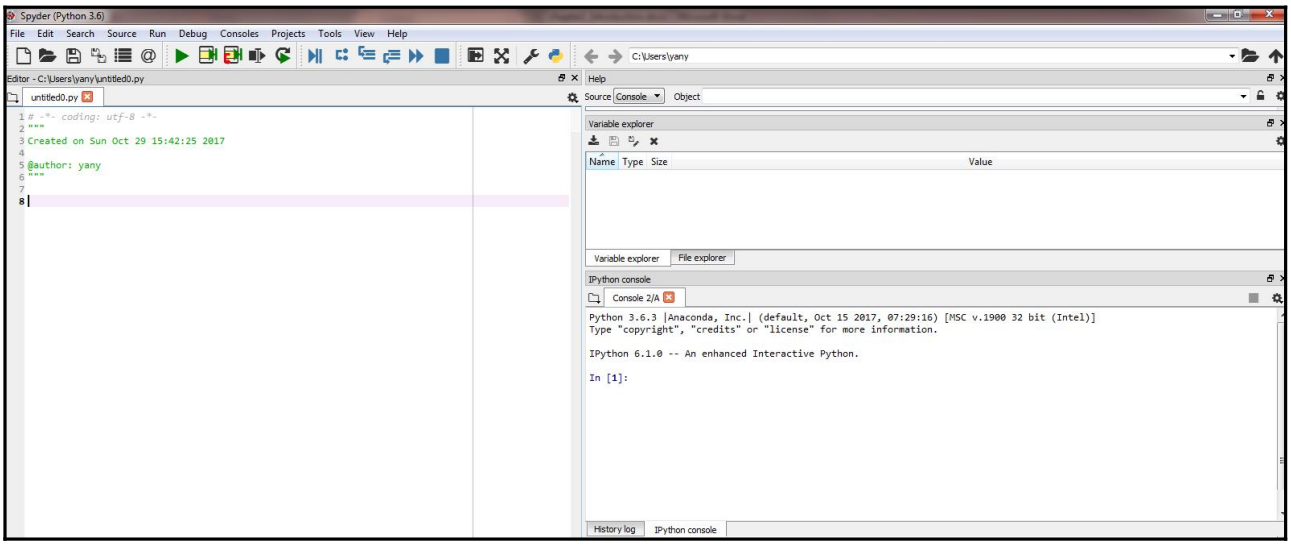

```
In [1]: pv=100

In [2]: pv
Out[2]: 100

In [3]: r=0.1

In [4]: n=5

In [5]: pv*(1+r)^n
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
<ipython-input-5-e52348b6e68d> in <module>()
----> 1 pv*(1+r)^n

TypeError: unsupported operand type(s) for ^: 'float' and 'int'

In [6]: pv*(1+r)**n
Out[6]: 161.05100000000004
```

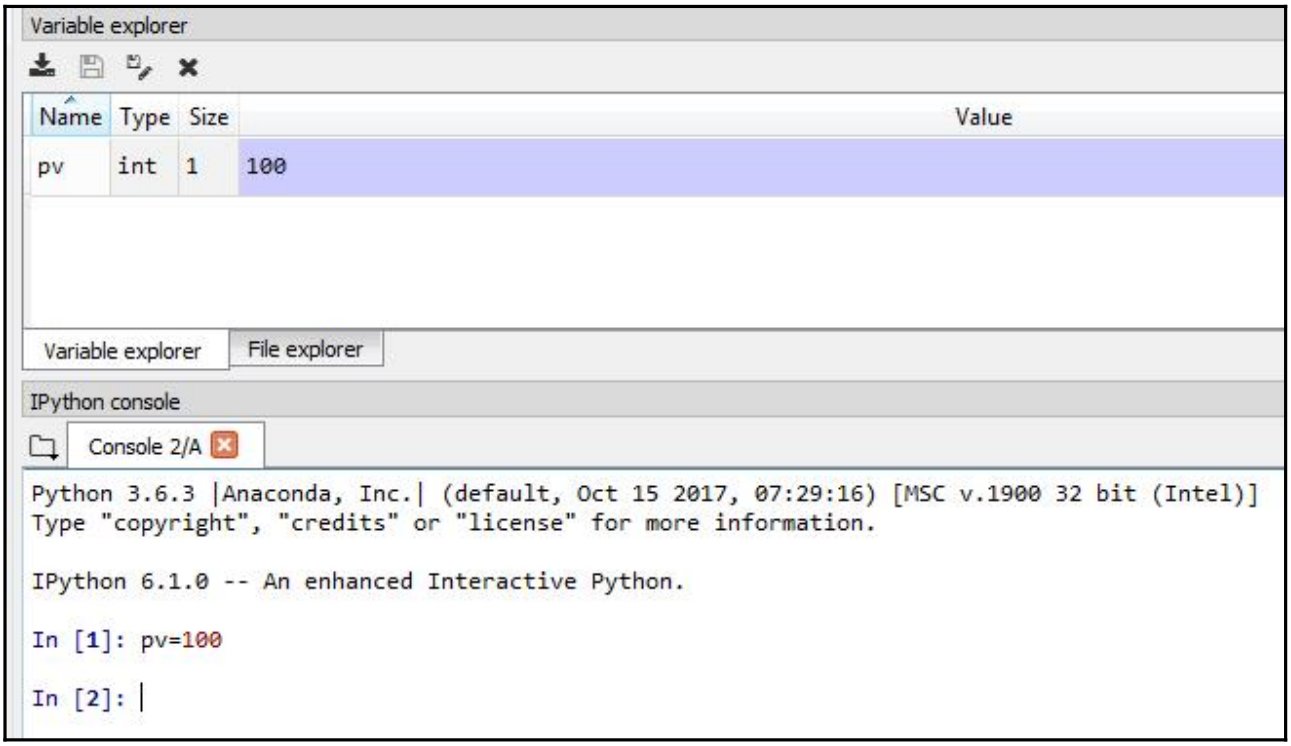


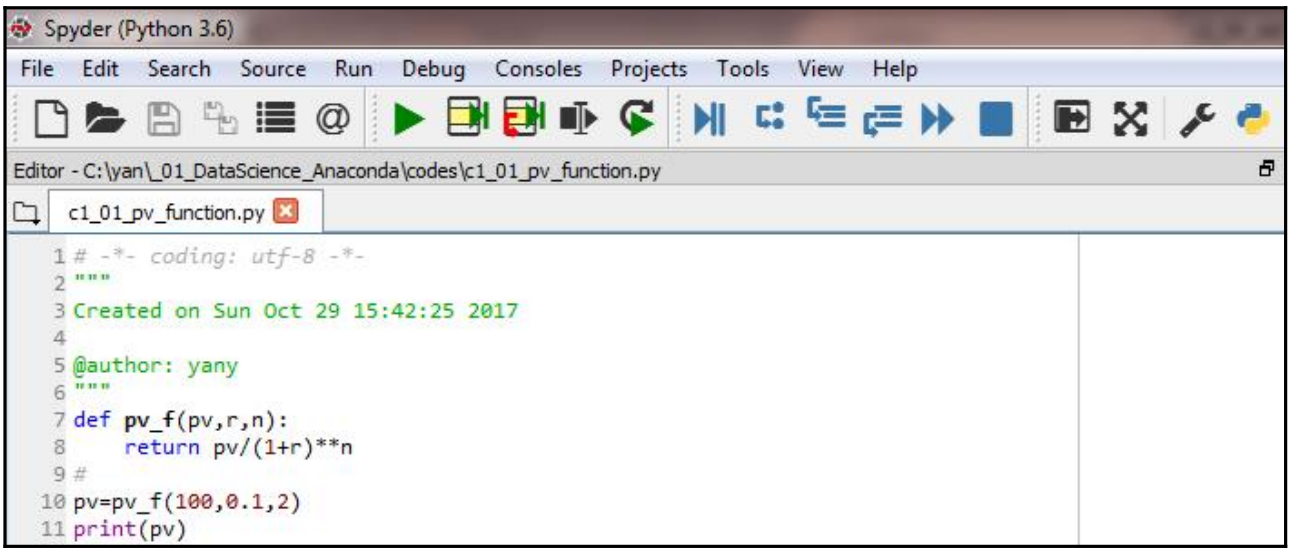


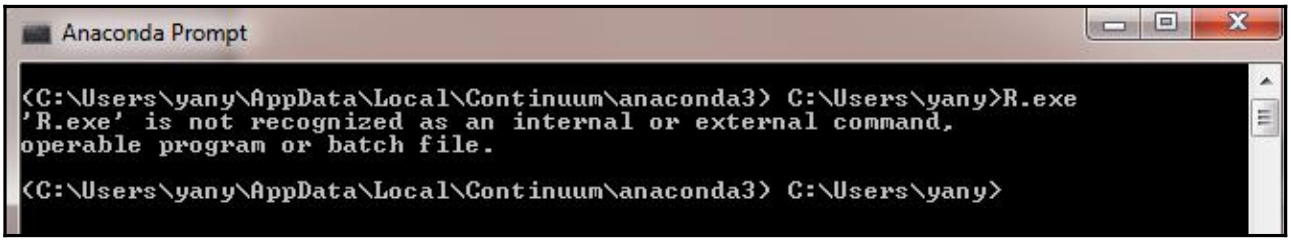

```
In [1]: pv=10

In [2]: pv
Out[2]: 10
```

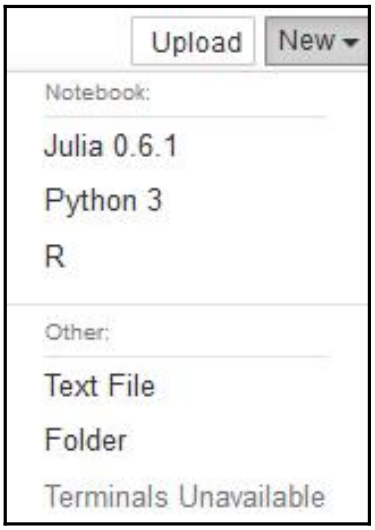
```
In [4]: import scipy as sp
        sp.sqrt(3)
Out[4]: 1.7320508075688772
```




Variable explorer
Name
Type
Size
Value
pv
int
1
100
Variable explorer
File explorer
IPython console
Console 2/A
Python 3.6.3 |Anaconda, Inc.| (default, Oct 15 2017, 07:29:16) [MSC v.1900 32 bit (Intel)]
Type "copyright", "credits" or "license" for more information.
IPython 6.1.0 -- An enhanced Interactive Python.
In [1]: pv=100
In [2]:


Spyder (Python 3.6)
File
Edit
Search
Source
Run
Debug
Consoles
Projects
Tools
View
Help
Editor - C:\yan_01_DataScience_Anaconda\codes\c1_01_pv_function.py
c1_01_pv_function.py
-*- coding: utf-8 -*-
"""
Created on Sun Oct 29 15:42:25 2017
@author: yany
"""
def pv_f(pv,r,n):
return pv/(1+r)**n
#
pv=pv_f(100,0.1,2)
print(pv)

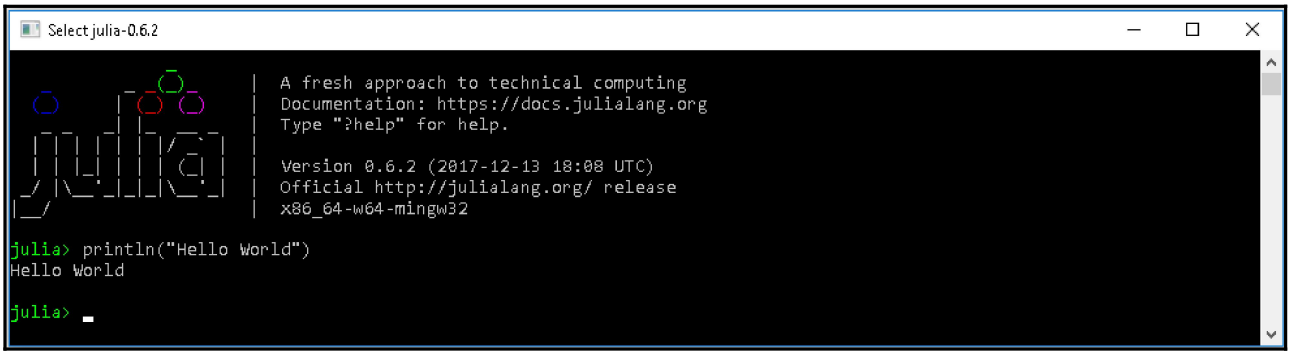

Anaconda Prompt
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>R.exe
'R.exe' is not recognized as an internal or external command,
operable program or batch file.
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>


Upload
New
Notebook:
Julia 0.6.1
Python 3
R
Other:
Text File
Folder
Terminals Unavailable

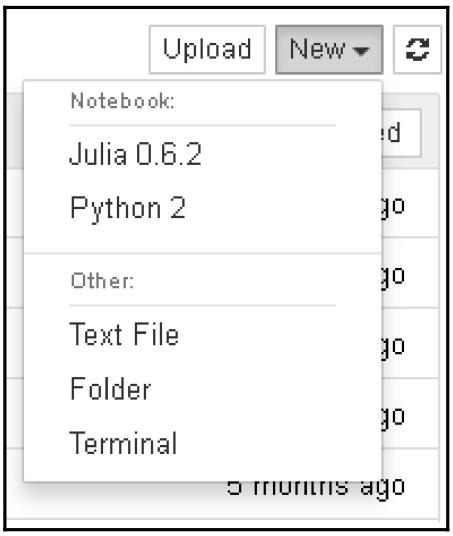

jupyter
Untitled4
Last Checkpoint: a minute ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Code
In [1]:
pv_f<-function(fv,r,n)fv/(1+r)^n
In [2]:
pv_f(100,0.1,1)
90.9090909090909

Julia (command line version)

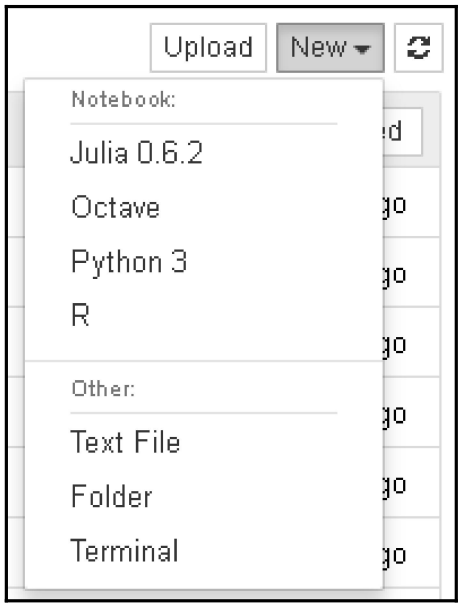
Windows Self-Extracting Archive (.exe) [help]	32-bit	64-bit	
macOS Package (.dmg) [help]	10.8+ 64-bit		
Generic Linux Binaries for x86 [help]	32-bit (GPG)	64-bit (GPG)	
Generic Linux Binaries for ARM	32-bit (armv7-a hard float) (GPG)	64-bit (armv8-a) (GPG)	
Source	Tarball (GPG)	Tarball with dependencies (GPG)	GitHub

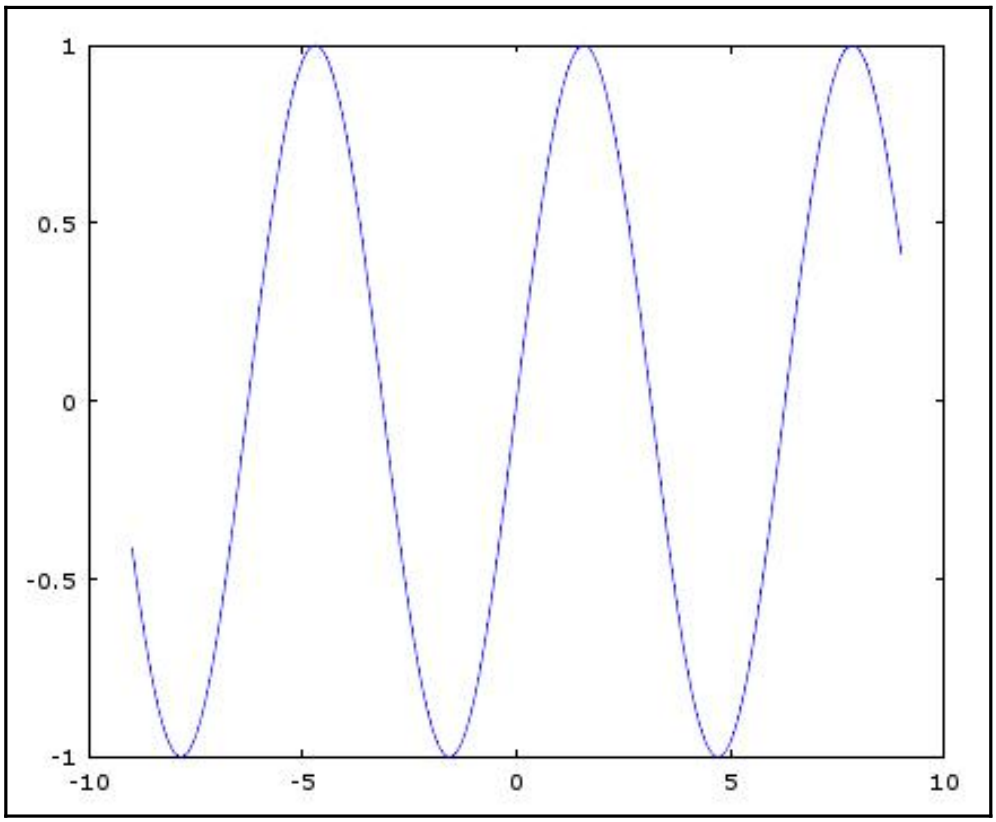



```
julia> Pkg.add("IJulia")
INFO: Initializing package repository C:\Users\yany\.julia\v0.6
INFO: Cloning METADATA from https://github.com/JuliaLang/METADATA.jl
INFO: Cloning cache of BinDeps from https://github.com/JuliaLang/BinDeps.jl.git
INFO: Cloning cache of BufferedStreams from https://github.com/BioJulia/Buffered
Streams.jl.git
```


Upload
New
Notebook:
Julia 0.6.2
Python 2
Other:
Text File
Folder
Terminal

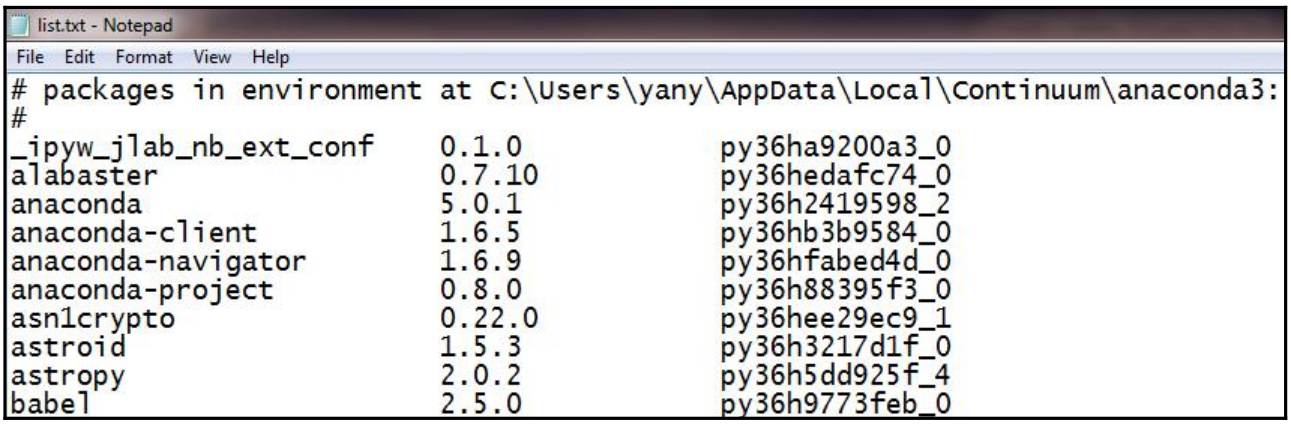

jupyter
Untitled
Last Checkpoint: 11 minutes ago
(unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Julia 0.6.2
Run
Code
In [1]:
println("This is great!")
This is great!
In [ ]:


Upload
New
Notebook:
Julia 0.6.2
Octave
Python 3
R
Other:
Text File
Folder
Terminal


1
0.5
0
-0.5
-1
-10
-5
0
5
10


Anaconda Distribution
User guide
User guide
Getting started
Tasks
Cheatsheet
Troubleshooting

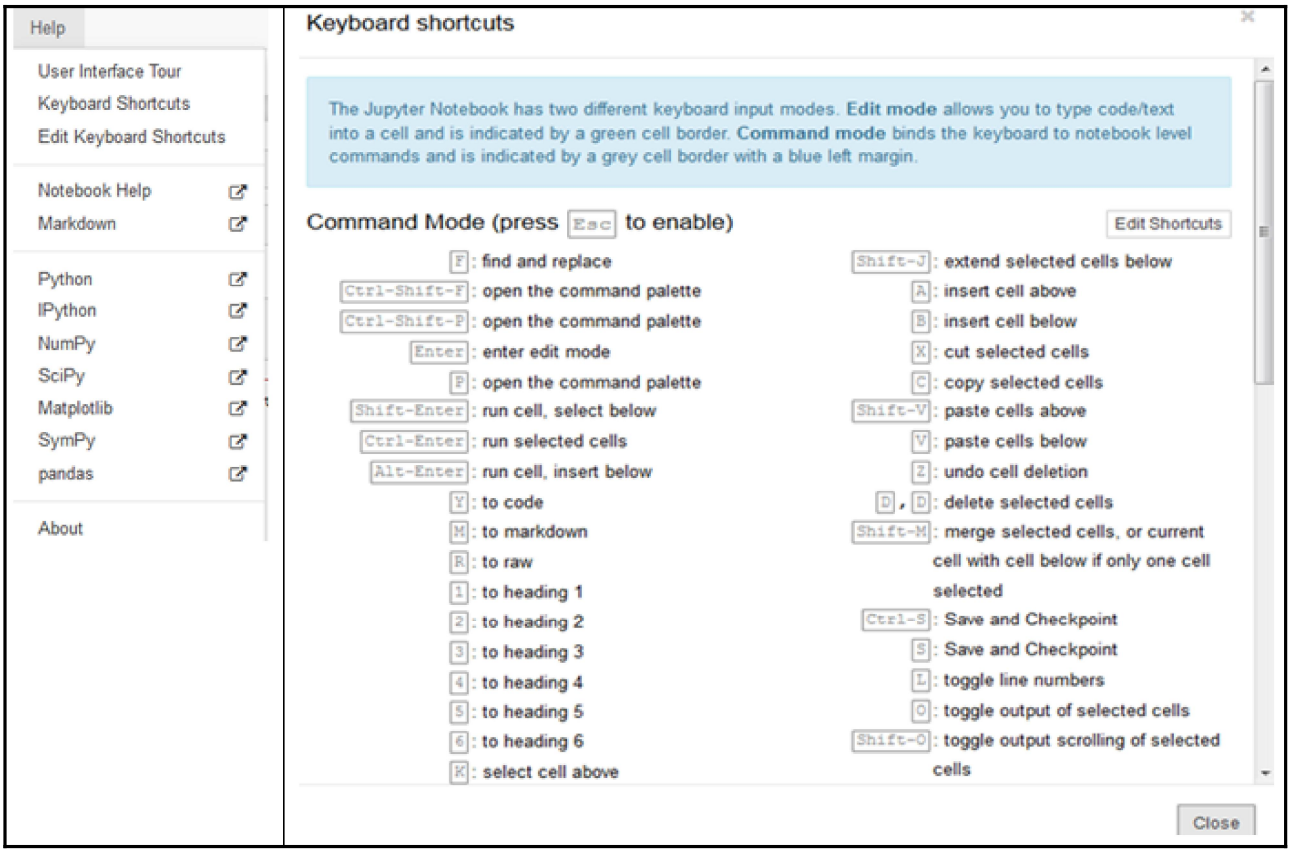

```
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>conda list >c:/t
emp/list.txt
```



list.txt - Notepad

File Edit Format View Help

```
# packages in environment at C:\Users\yany\AppData\Local\Continuum\anaconda3:
#
_ipyw_jlab_nb_ext_conf    0.1.0            py36ha9200a3_0
alabaster                 0.7.10           py36hedafc74_0
anaconda                  5.0.1            py36h2419598_2
anaconda-client           1.6.5            py36hb3b9584_0
anaconda-navigator        1.6.9            py36hfabed4d_0
anaconda-project          0.8.0            py36h88395f3_0
asn1crypto                0.22.0           py36hee29ec9_1
astroid                   1.5.3            py36h3217d1f_0
astropy                   2.0.2            py36h5dd925f_4
babel                     2.5.0            py36h9773feb_0
```



Help

User Interface Tour
Keyboard Shortcuts
Edit Keyboard Shortcuts

Notebook Help
Markdown

Python
IPython
NumPy
SciPy
Matplotlib
SymPy
pandas

About

Keyboard shortcuts

The Jupyter Notebook has two different keyboard input modes. **Edit mode** allows you to type code/text into a cell and is indicated by a green cell border. **Command mode** binds the keyboard to notebook level commands and is indicated by a grey cell border with a blue left margin.

Command Mode (press Esc to enable) — Edit Shortcuts

F : find and replace
Ctrl-Shift-F : open the command palette
Ctrl-Shift-P : open the command palette
Enter : enter edit mode
P : open the command palette
Shift-Enter : run cell, select below
Ctrl-Enter : run selected cells
Alt-Enter : run cell, insert below
Y : to code
M : to markdown
R : to raw
1 : to heading 1
2 : to heading 2
3 : to heading 3
4 : to heading 4
5 : to heading 5
6 : to heading 6
K : select cell above

Shift-J : extend selected cells below
A : insert cell above
B : insert cell below
X : cut selected cells
C : copy selected cells
Shift-V : paste cells above
V : paste cells below
Z : undo cell deletion
D , D : delete selected cells
Shift-M : merge selected cells, or current cell with cell below if only one cell selected
Ctrl-S : Save and Checkpoint
S : Save and Checkpoint
L : toggle line numbers
O : toggle output of selected cells
Shift-O : toggle output scrolling of selected cells

Close

Chapter 03: Data Basics

Most Popular Data Sets (hits since 2007):

1631325:  Iris

1052490:  Adult

801975:  Wine

Iris Data Set

Download: Data Folder, Data Set Description

Abstract: Famous database; from Fisher, 1936

Data Set Characteristics:	Multivariate	**Number of Instances:**	150	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	4	**Date Donated**	1988-07-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	1631334

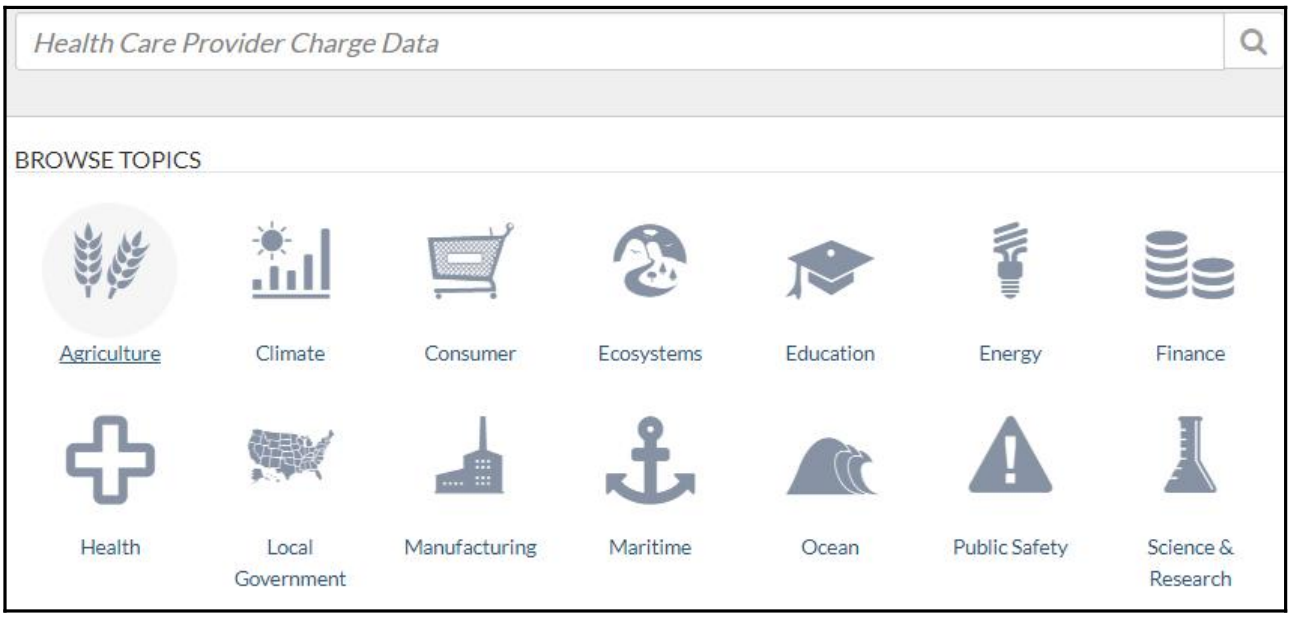
Source:

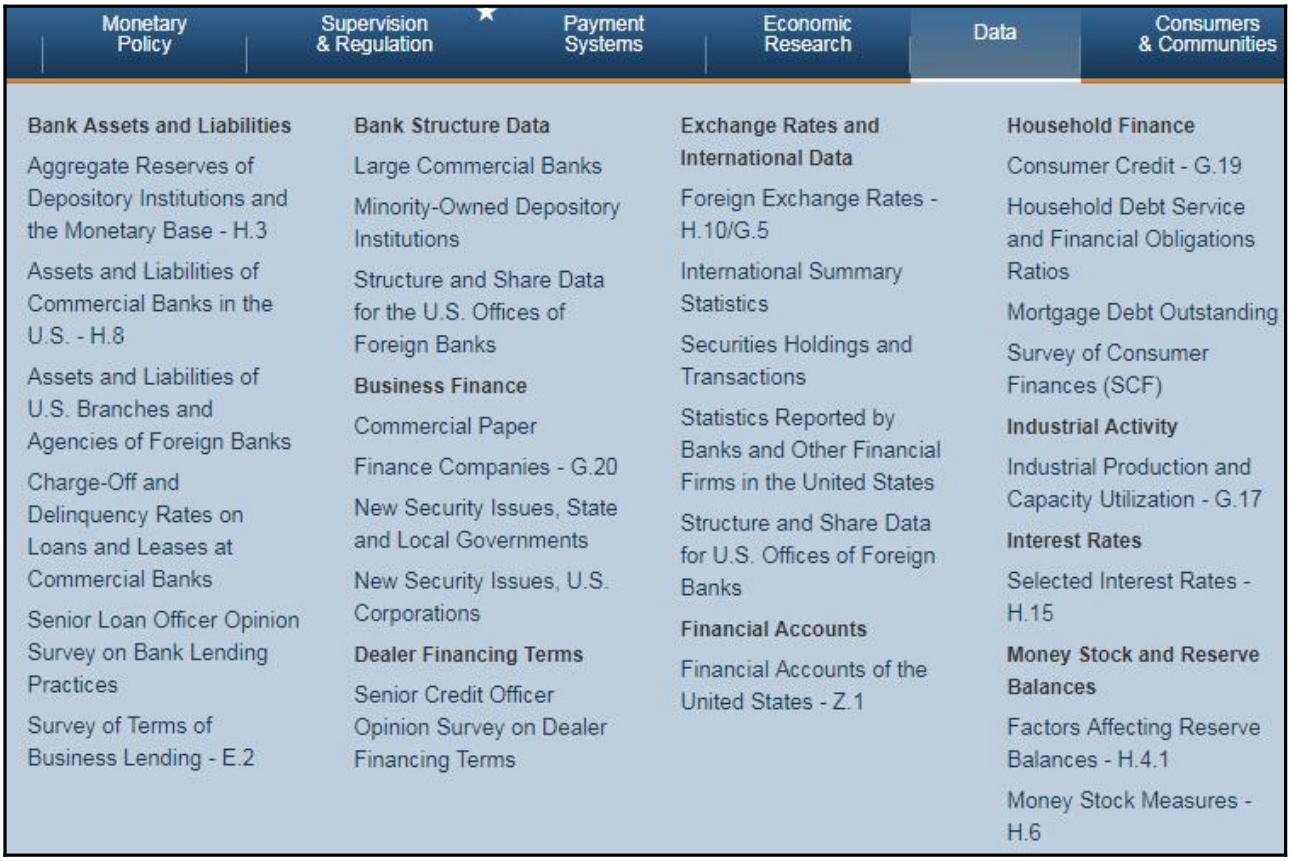
Creator:

R.A. Fisher

Donor:

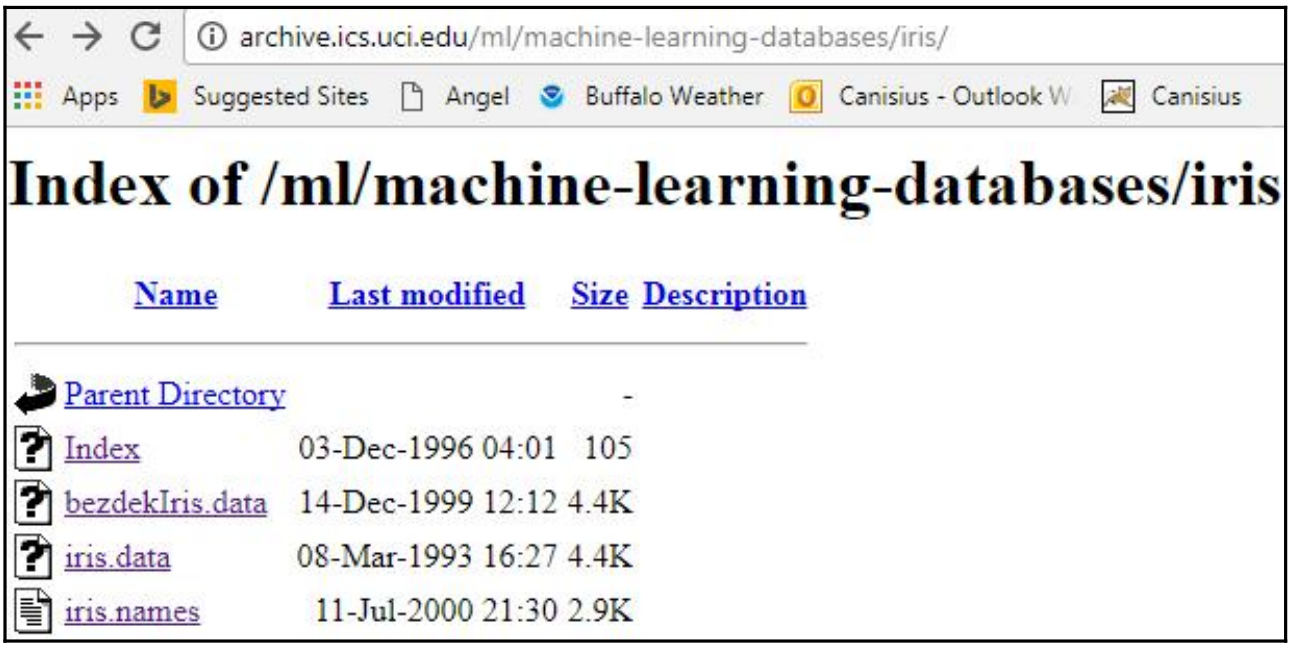
Michael Marshall (MARSHALL%PLU '@' io.arc.nasa.gov)


Health Care Provider Charge Data
BROWSE TOPICS
Agriculture
Climate
Consumer
Ecosystems
Education
Energy
Finance
Health
Local Government
Manufacturing
Maritime
Ocean
Public Safety
Science & Research


Monetary Policy
Supervision & Regulation
Payment Systems
Economic Research
Data
Consumers & Communities
Bank Assets and Liabilities
Aggregate Reserves of Depository Institutions and the Monetary Base - H.3
Assets and Liabilities of Commercial Banks in the U.S. - H.8
Assets and Liabilities of U.S. Branches and Agencies of Foreign Banks
Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks
Senior Loan Officer Opinion Survey on Bank Lending Practices
Survey of Terms of Business Lending - E.2
Bank Structure Data
Large Commercial Banks
Minority-Owned Depository Institutions
Structure and Share Data for the U.S. Offices of Foreign Banks
Business Finance
Commercial Paper
Finance Companies - G.20
New Security Issues, State and Local Governments
New Security Issues, U.S. Corporations
Dealer Financing Terms
Senior Credit Officer Opinion Survey on Dealer Financing Terms
Exchange Rates and International Data
Foreign Exchange Rates - H.10/G.5
International Summary Statistics
Securities Holdings and Transactions
Statistics Reported by Banks and Other Financial Firms in the United States
Structure and Share Data for U.S. Offices of Foreign Banks
Financial Accounts
Financial Accounts of the United States - Z.1
Household Finance
Consumer Credit - G.19
Household Debt Service and Financial Obligations Ratios
Mortgage Debt Outstanding
Survey of Consumer Finances (SCF)
Industrial Activity
Industrial Production and Capacity Utilization - G.17
Interest Rates
Selected Interest Rates - H.15
Money Stock and Reserve Balances
Factors Affecting Reserve Balances - H.4.1
Money Stock Measures - H.6

```
In [1]: import pandas as pd

In [2]: x=dir(pd)

In [3]: print(x)
['Categorical', 'CategoricalIndex', 'DataFrame', 'DateOffset', 'DatetimeIndex', 'ExcelFile', 'ExcelWriter', 'Expr',
'Float64Index', 'Grouper', 'HDFStore', 'Index', 'IndexSlice', 'Int64Index', 'MultiIndex', 'NaT', 'Panel', 'Panel4D',
'Period', 'PeriodIndex', 'RangeIndex', 'Series', 'SparseArray', 'SparseDataFrame', 'SparseList', 'SparseSeries',
'SparseTimeSeries', 'Term', 'TimeGrouper', 'TimeSeries', 'Timedelta', 'TimedeltaIndex', 'Timestamp', 'WidePanel',
'__builtins__', '__doc__', '__docformat__', '__file__', '__name__', '__package__', '__path__', '__version__', '_join',
'_np_version_under1p10', '_np_version_under1p11', '_np_version_under1p12', '_np_version_under1p8', '_np_version_under1p9',
'_period', '_sparse', '_testing', '_version', '_window', 'algos', 'api', 'bdate_range', 'compat', 'computation', 'concat',
'core', 'crosstab', 'cut', 'date_range', 'datetime', 'datetools', 'describe_option', 'eval', 'ewma', 'ewmcorr', 'ewmcov',
'ewmstd', 'ewmvar', 'ewmvol', 'expanding_apply', 'expanding_corr', 'expanding_count', 'expanding_cov', 'expanding_kurt',
'expanding_max', 'expanding_mean', 'expanding_median', 'expanding_min', 'expanding_quantile', 'expanding_skew',
'expanding_std', 'expanding_sum', 'expanding_var', 'factorize', 'fama_macbeth', 'formats', 'get_dummies', 'get_option',
'get_store', 'groupby', 'hashtable', 'index', 'indexes', 'infer_freq', 'info', 'io', 'isnull', 'json', 'lib', 'lreshape',
'match', 'melt', 'merge', 'merge_asof', 'merge_ordered', 'msgpack', 'notnull', 'np', 'offsets', 'ols', 'option_context',
'options', 'ordered_merge', 'pandas', 'parser', 'period_range', 'pivot', 'pivot_table', 'plot_params', 'pnow', 'qcut',
'read_clipboard', 'read_csv', 'read_excel', 'read_fwf', 'read_gbq', 'read_hdf', 'read_html', 'read_json', 'read_msgpack',
'read_pickle', 'read_sas', 'read_sql', 'read_sql_query', 'read_sql_table', 'read_stata', 'read_table', 'reset_option',
'rolling_apply', 'rolling_corr', 'rolling_count', 'rolling_cov', 'rolling_kurt', 'rolling_max', 'rolling_mean',
'rolling_median', 'rolling_min', 'rolling_quantile', 'rolling_skew', 'rolling_std', 'rolling_sum', 'rolling_var',
'rolling_window', 'scatter_matrix', 'set_eng_float_format', 'set_option', 'show_versions', 'sparse', 'stats', 'test',
'timedelta_range', 'to_datetime', 'to_msgpack', 'to_numeric', 'to_pickle', 'to_timedelta', 'tools', 'tseries', 'tslib',
'types', 'unique', 'util', 'value_counts', 'wide_to_long']
```

```
In [78]: help(pd.to_pickle)
Help on function to_pickle in module pandas.io.pickle:

to_pickle(obj, path, compression='infer')
    Pickle (serialize) object to input file path

    Parameters
    ----------
    obj : any object
    path : string
        File path
    compression : {'infer', 'gzip', 'bz2', 'xz', None}, default 'infer'
        a string representing the compression to use in the output file

        .. versionadded:: 0.20.0
```



archive.ics.uci.edu/ml/machine-learning-databases/iris/

Apps | Suggested Sites | Angel | Buffalo Weather | Canisius - Outlook W | Canisius

Index of /ml/machine-learning-databases/iris

Name	Last modified	Size	Description
Parent Directory		-	
Index	03-Dec-1996 04:01	105	
bezdekIris.data	14-Dec-1999 12:12	4.4K	
iris.data	08-Mar-1993 16:27	4.4K	
iris.names	11-Jul-2000 21:30	2.9K	

read.fwf {utils}

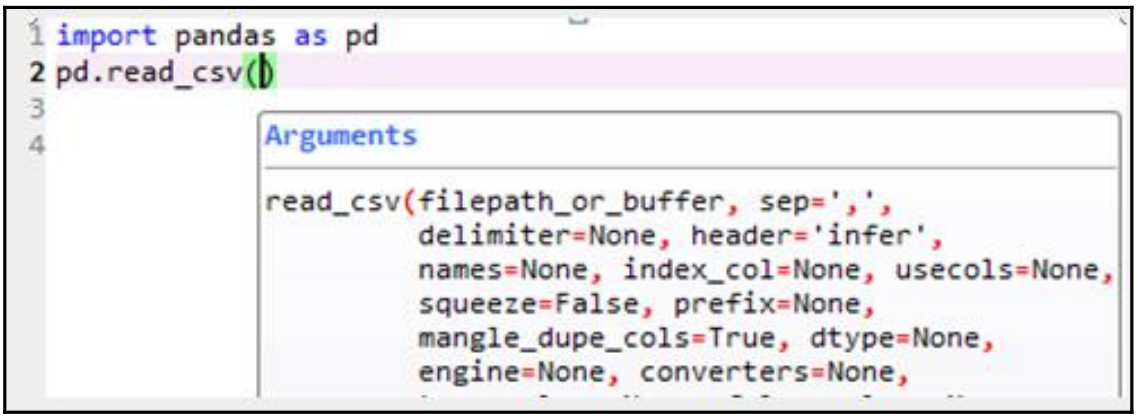
Read Fixed Width Format Files

Description

Read a table of **f**ixed **w**idth **f**ormatted data into a `data.frame`.

Usage

```
read.fwf(file, widths, header = FALSE, sep = "\t",
         skip = 0, row.names, col.names, n = -1,
         buffersize = 2000, fileEncoding = "", ...)
```




```
In [4]: help(pd.read_sas)
Help on function read_sas in module pandas.io.sas.sasreader:

read_sas(filepath_or_buffer, format=None, index=None, encoding=None, chunksize=None, iterator=False)
    Read SAS files stored as either XPORT or SAS7BDAT format files.
```

```
>  head(y,2)
        Date Pre_Close  Open  High   Low Close Adj_Pre_Close Adj_Open
1 2018-01-10     30.25 30.18 30.21 29.78 30.12         30.25    30.18
2 2018-01-09     29.65 29.70 30.30 29.55 30.25         29.65    29.70
  Adj_High Adj_Low Adj_Close Turnover_Volume Turnover_Value Deal_Amount
1    30.21   29.78     30.12         16383506      491497957       21220
2    30.30   29.55     30.25         27625443      830430487       33053
  Circulation_Market_Value Market_Value Turnover_Rate Adj_Factor Adj_Reason
1             627204723600 851332914600        0.0008         NA         NA
2             629911782500 855007326250        0.0013         NA         NA
```

```
In [36]: y.head(2)
Out[36]:
            Open   High    Low  Close  Volume  Ex-Dividend  Split Ratio
Date
1972-08-25  32.5  33.13  32.50  33.00  4900.0          0.0          1.0
1972-08-28  33.0  33.25  32.88  32.88  1900.0          0.0          1.0

            Adj. Open  Adj. High  Adj. Low  Adj. Close  Adj. Volume
Date
1972-08-25   0.043162   0.043999  0.043162    0.043826    2508800.0
1972-08-28   0.043826   0.044158  0.043667    0.043667     972800.0

In [37]: y.tail(1)
Out[37]:
             Open    High     Low   Close     Volume  Ex-Dividend  \
Date
2018-01-09  102.03  102.35  100.31  100.39  7296463.0          0.0

            Split Ratio  Adj. Open  Adj. High  Adj. Low  Adj. Close  \
Date
2018-01-09          1.0     102.03     102.35    100.31      100.39

            Adj. Volume
Date
2018-01-09    7296463.0
```

```
> dim(nyseListing)
[1] 3387    4
> head(nyseListing)
  Symbol                          Name MarketCap Exchange
1      A   Agilent Technologies, Inc. $12,852.3     NYSE
2     AA                  Alcoa Inc. $28,234.5     NYSE
3    AAI       AirTran Holdings, Inc.    $156.9     NYSE
4    AAP       Advance Auto Parts Inc  $3,507.4     NYSE
5    AAR              AMR CORPORATION     $81.7     NYSE
6    AAV ADVANTAGE ENERGY INCOME FUND  $1,674.4     NYSE
```

```
> x<-nyseListing[order(nyseListing$Name),]
> head(x)
     Symbol                          Name MarketCap Exchange
2017    MMM                    3M Company $48,398.7     NYSE
557     CFD 40/86 Strategic Income Fund     $56.8     NYSE
1721    KDE   4Kids Entertainment, Inc.     $99.7     NYSE
2164    NDN        99 CENTS ONLY STORES    $432.3     NYSE
87      AHC       A.H. Belo Corporation    $107.4     NYSE
1242    GFW    AAG Holding Company Inc.      <NA>     NYSE
```

```
> y
  ID  RET Data1
1  1  0.1   100
2  3  0.3   300
3  1 -0.4    30
```

```
> z
  ID  RET Data1
3  1 -0.4    30
1  1  0.1   100
2  3  0.3   300
```

```
   X  Y
0  8  3
1  8  2
2  1 -1
   X  Y
2  1 -1
0  8  3
1  8  2
   X  Y
2  1 -1
1  8  2
0  8  3
```

```
  inner  merge
   FirmA  FirmB  YEAR  FirmC  SP500
0   -0.3   0.00  2011   0.12   0.10
1   -0.2   0.23  2013   0.23   0.17

  outter merge
   FirmA  FirmB  YEAR  FirmC  SP500
0   0.20   0.10  2010    NaN    NaN
1  -0.30   0.00  2011   0.12   0.10
2   0.13   0.05  2012    NaN    NaN
3  -0.20   0.23  2013   0.23   0.17
4    NaN    NaN  2014   0.11  -0.05
5    NaN    NaN  2015  -0.10   0.13

  left  merge
   FirmA  FirmB  YEAR  FirmC  SP500
0   0.20   0.10  2010    NaN    NaN
1  -0.30   0.00  2011   0.12   0.10
2   0.13   0.05  2012    NaN    NaN
3  -0.20   0.23  2013   0.23   0.17

  right  merge
   FirmA  FirmB  YEAR  FirmC  SP500
0   -0.3   0.00  2011   0.12   0.10
1   -0.2   0.23  2013   0.23   0.17
2    NaN    NaN  2014   0.11  -0.05
3    NaN    NaN  2015  -0.10   0.13
```

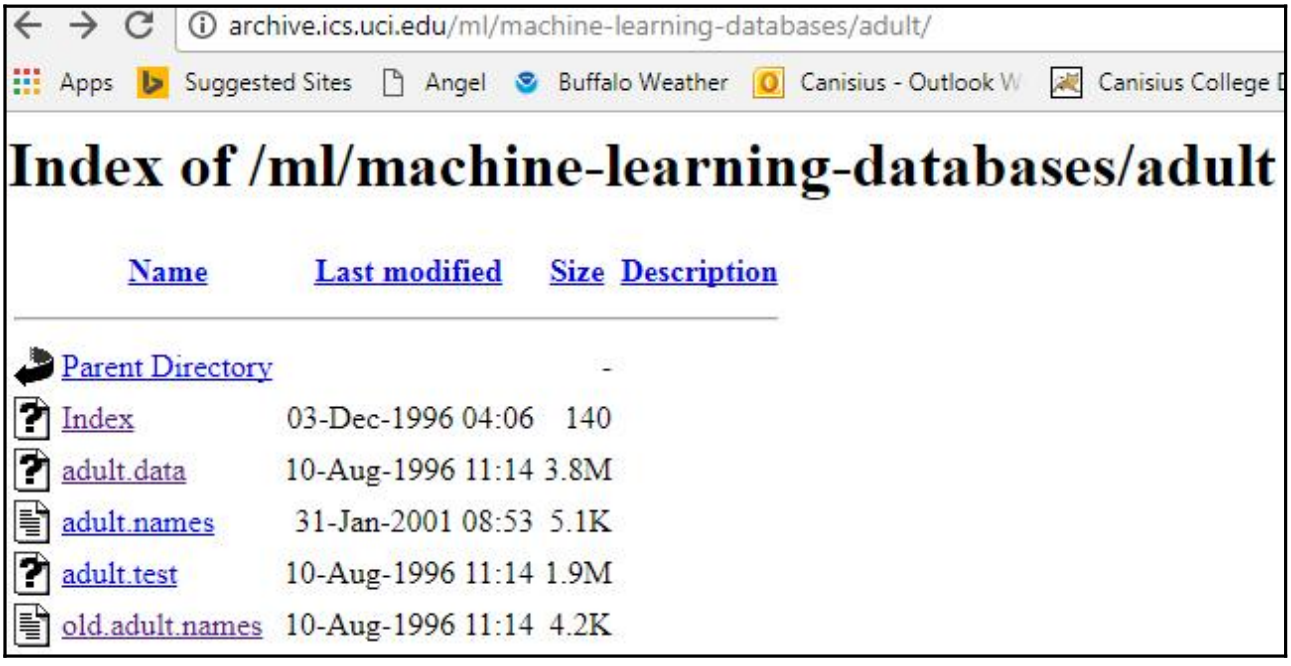
```
Retrieving data from table '82070ENG'
Done!
  CaribbeanNetherlands  EmployedLabourForceInternatDef_1  \
0              Bonaire                            8837.0
1        St. Eustatius                            2099.0
2                 Saba                            1045.0
3              Bonaire                            6344.0
4        St. Eustatius                            1443.0

   EmployedLabourForceNationalDef_2                 Gender  ID Periods
0                            8559.0  Total male and female   0    2012
1                            1940.0  Total male and female   1    2012
2                             971.0  Total male and female   2    2012
3                            6241.0  Total male and female   3    2012
4                            1362.0  Total male and female   4    2012

                PersonalCharacteristics
0         Total personal characteristics
1         Total personal characteristics
2         Total personal characteristics
3  Employment status: permanent contract
4  Employment status: permanent contract
Caribbean Netherlands; employed labour force characteristics 2012
```

```
In [188]: import datadotworld as dw

In [189]: x=dir(dw)

In [190]: print(x)
['ChainedConfig', 'DataDotWorld', 'EnvConfig', 'FileConfig', 'InlineConfig', 'UriParam', '__builtins__',
'__cached__', '__doc__', '__file__', '__instances', '__loader__', '__name__', '__package__', '__path__',
'__spec__', '__version__', '_get_instance', 'absolute_import', 'api_client', 'client', 'config',
'datadotworld', 'files', 'load_dataset', 'models', 'open_remote_file', 'query', 'util', 'weakref']
```

```
RuntimeError: Configuration file not found at C:\Users\yany/.dw/config.To fix this issue, run dw
configure
```

```
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>dw configure
API token (obtained at: https://data.world/settings/advanced): c:/Users/yany/.dw
/
```

```
   DataDotWorldBBallStats.name  pointspergame  assistspergame  \
0                          Jon           20.4             1.3
1                          Rob           15.5             8.0
2                       Sharon           30.1            11.2
3                         Alex            8.2             0.5
4                      Rebecca           12.3            17.0
5                       Ariane           18.1             3.0
6                        Bryon           16.0             8.5
7                         Matt           13.0             2.1

   DataDotWorldBBallTeam.name  height handedness
0                          Jon    6'5"      Right
1                          Rob  6'7.5"       Left
2                       Sharon    6'3"      Right
3                         Alex    6'2"      Right
4                      Rebecca      7'      Right
5                       Ariane    5'8"       Left
6                        Bryon      7'      Right
7                         Matt    5'5"      Right
```



```
In [66]: adult.head()
Out[66]:
   age          workclass  fnlwgt   education  education-num  \
0   39          State-gov   77516   Bachelors             13
1   50   Self-emp-not-inc   83311   Bachelors             13
2   38            Private  215646     HS-grad              9
3   53            Private  234721        11th              7
4   28            Private  338409   Bachelors             13

        marital-status          occupation    relationship    race      sex
0        Never-married        Adm-clerical   Not-in-family   White     Male
1   Married-civ-spouse     Exec-managerial         Husband   White     Male
2             Divorced   Handlers-cleaners   Not-in-family   White     Male
3   Married-civ-spouse   Handlers-cleaners         Husband   Black     Male
4   Married-civ-spouse      Prof-specialty            Wife   Black   Female

   capital-gain  capital-loss  hours-per-week  native-country   class
0          2174             0              40   United-States   <=50K
1             0             0              13   United-States   <=50K
2             0             0              40   United-States   <=50K
3             0             0              40   United-States   <=50K
4             0             0              40            Cuba   <=50K
```

```
In [21]: import pandas as pd

In [22]: help(to_numeric)
Traceback (most recent call last):

  File "<ipython-input-22-a1c2c20cd180>", line 1, in <module>
    help(to_numeric)

NameError: name 'to_numeric' is not defined
```

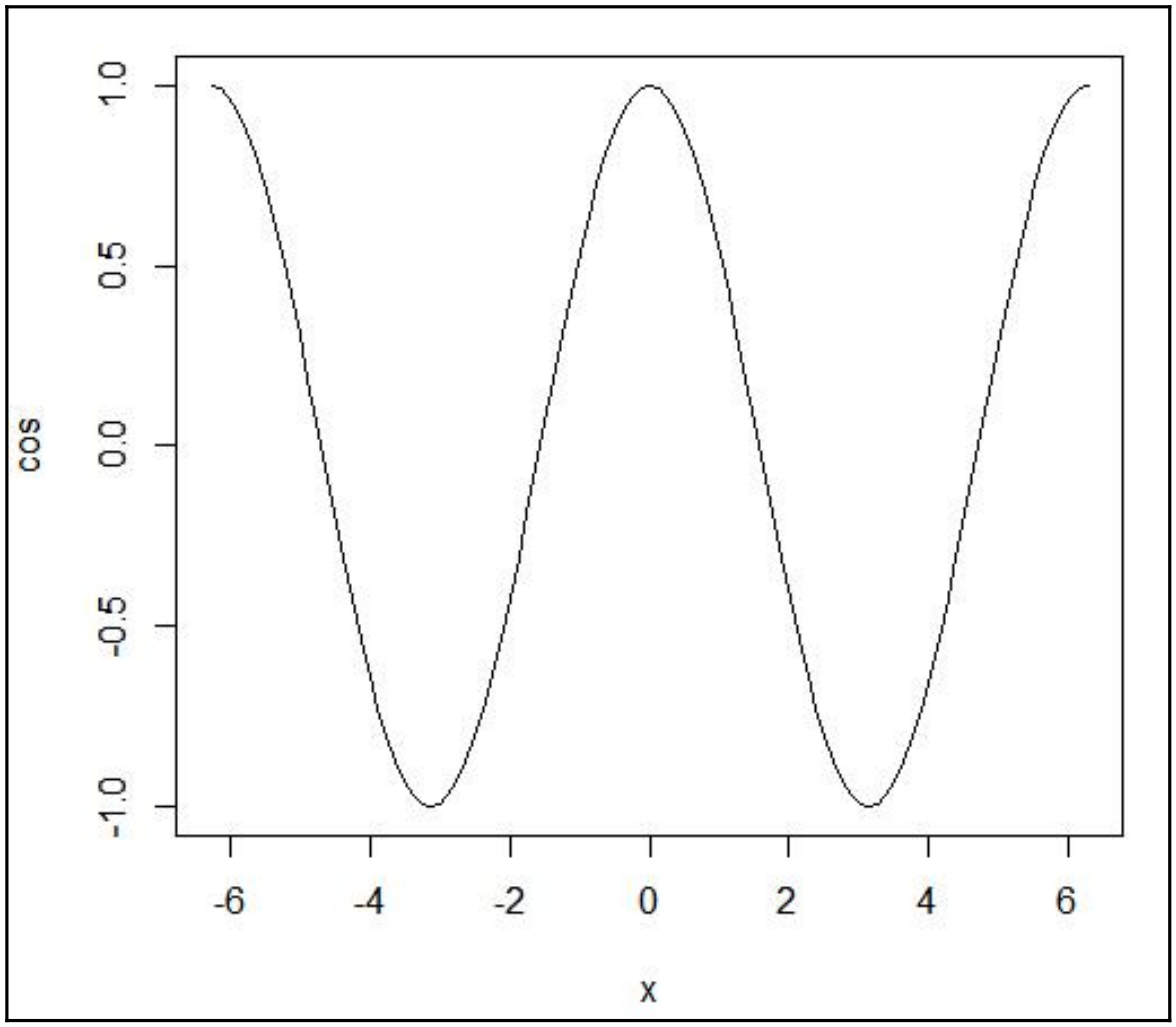
```
> x<-read.table("c:/temp/bank-full.csv",sep=";",header=T)
> head(x,3)
  age          job marital education default balance housing loan contact day month
1  58   management married  tertiary      no    2143     yes   no unknown   5   may
2  44   technician  single secondary      no      29     yes   no unknown   5   may
3  33 entrepreneur married secondary      no       2     yes  yes unknown   5   may
  duration campaign pdays previous poutcome  y
1      261        1    -1        0  unknown no
2      151        1    -1        0  unknown no
3       76        1    -1        0  unknown no
>
```

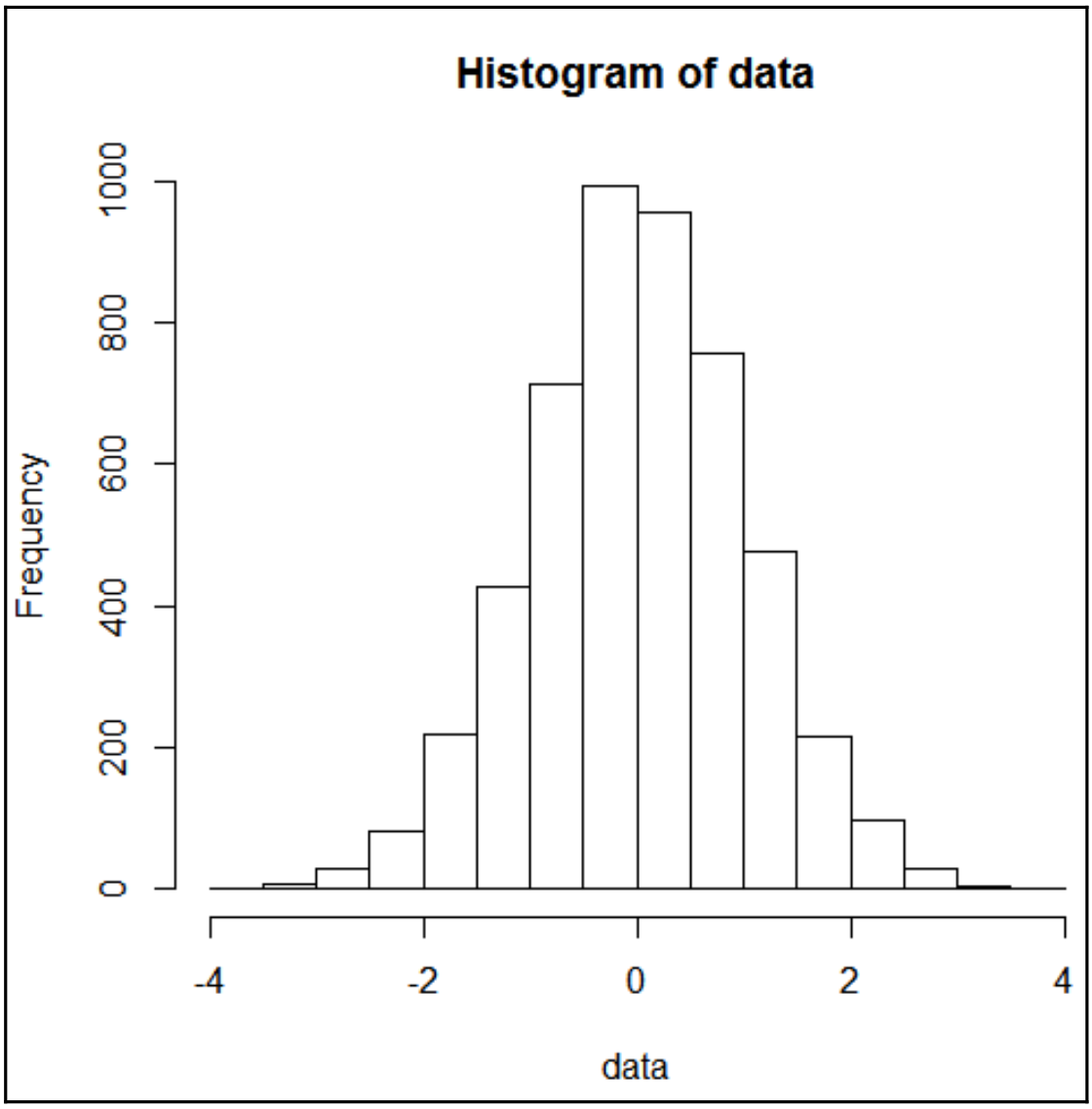
```
          0         1         2         3         4         5         6  \
0 -0.560476  0.253319 -0.710407  0.787739  2.198810 -0.375603 -0.715242
1 -0.230177 -0.028547  0.256884  0.769042  1.312413 -0.561876 -0.752689
2  1.558708 -0.042870 -0.246692  0.000000 -0.265145  0.000000 -0.938539
3  0.070508  1.368602 -0.347543 -1.008377  0.543194  0.090497 -1.052513
4  0.129288 -0.225771 -0.951619 -0.119453 -0.414340  1.598509 -0.437160

          7         8         9
0  1.014943 -0.073556  1.430402
1 -1.992748 -1.168651  1.046629
2 -0.427279  0.000000  0.435289
3  0.000000 -0.028842  0.715178
4 -0.893208  0.000000  0.917175
1    10
1    10
2     9
3     8
4    12
5    12
dtype: int64
```

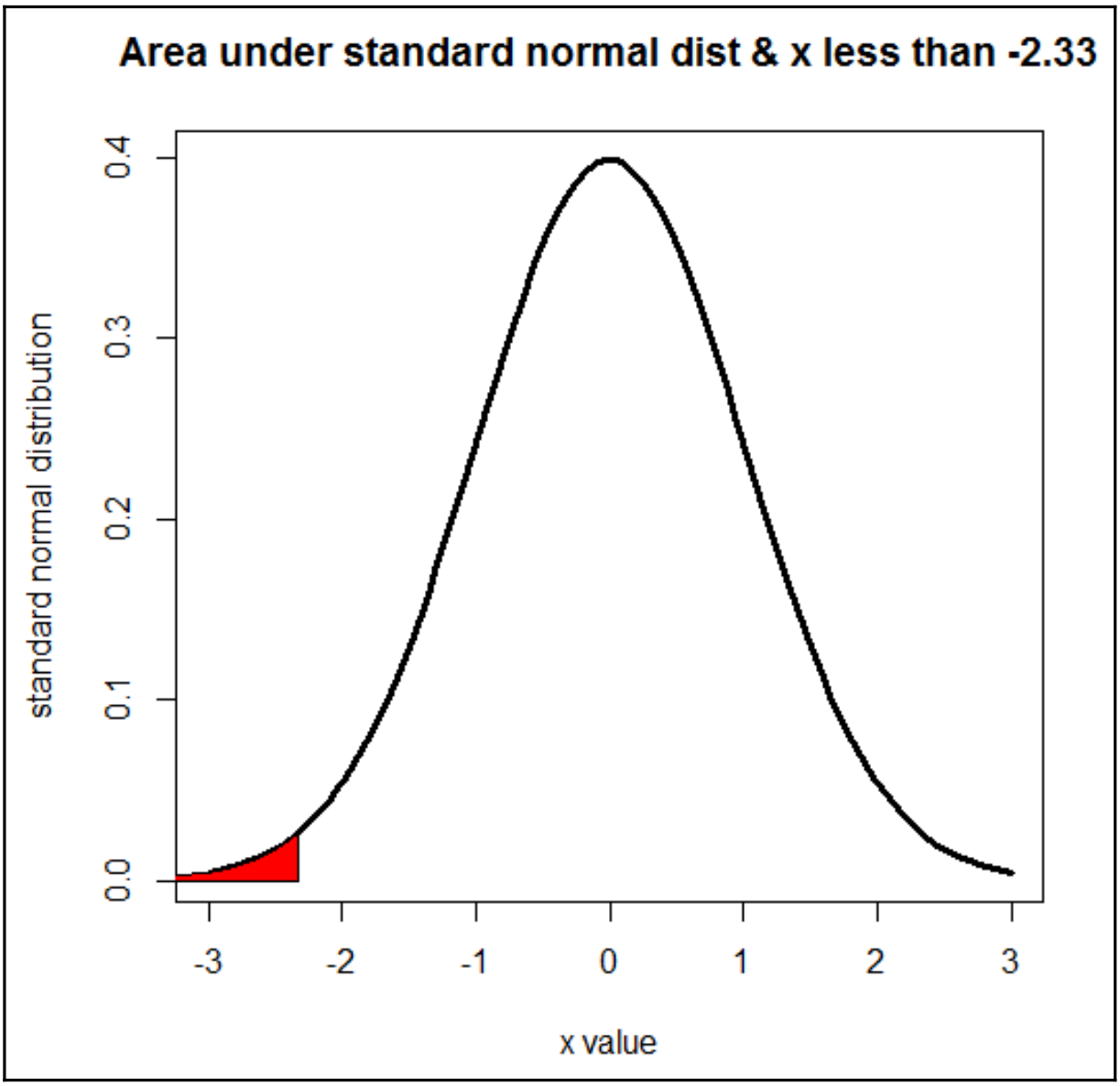
```
In [14]: x2=x
    ...: x2[[1,2,3,4,5]] = x2[[1,2,3,4,5]].replace(0, sp.NaN)
    ...: print(x2.head())
          0         1         2         3         4         5         6  \
0 -0.560476  0.253319 -0.710407  0.787739  2.198810 -0.375603 -0.715242
1 -0.230177 -0.028547  0.256884  0.769042  1.312413 -0.561876 -0.752689
2  1.558708 -0.042870 -0.246692       NaN -0.265145       NaN -0.938539
3  0.070508  1.368602 -0.347543 -1.008377  0.543194  0.090497 -1.052513
4  0.129288 -0.225771 -0.951619 -0.119453 -0.414340  1.598509 -0.437160

          7         8         9
0  1.014943 -0.073556  1.430402
1 -1.992748 -1.168651  1.046629
2 -0.427279  0.000000  0.435289
3  0.000000 -0.028842  0.715178
4 -0.893208  0.000000  0.917175
```

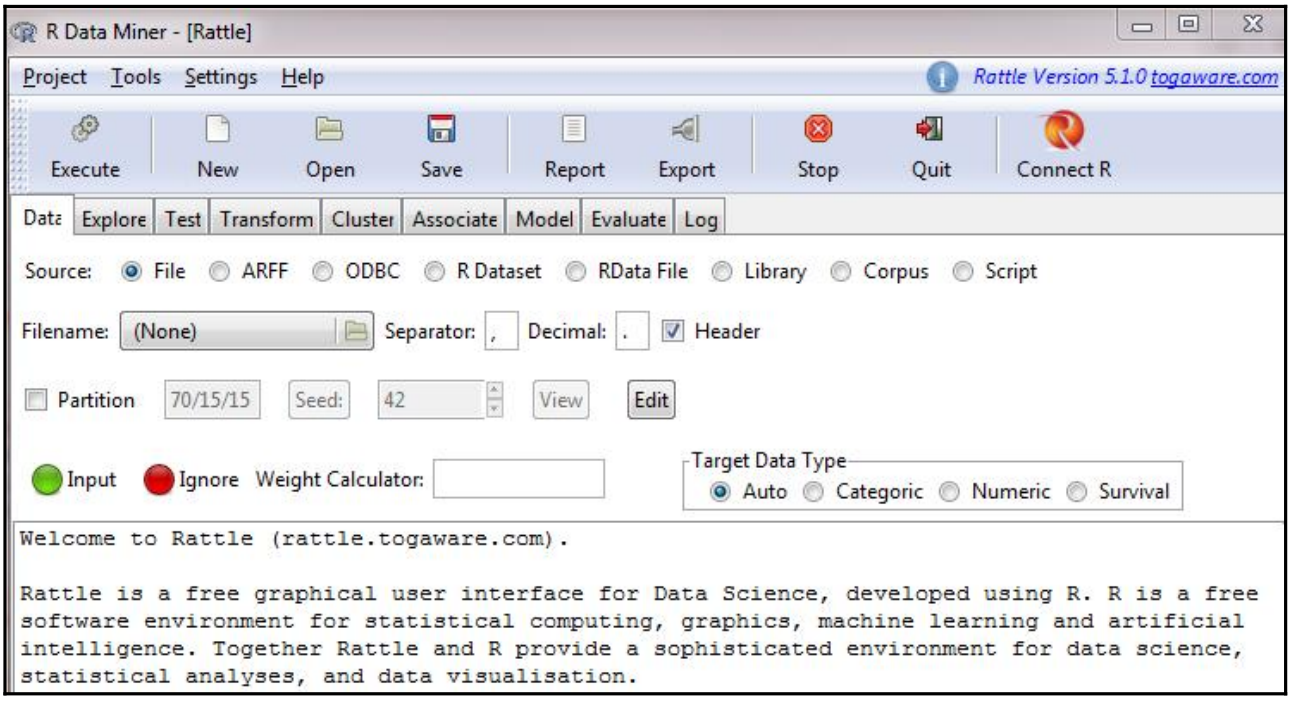
```
In [15]: x3=x2
    ...: x3.fillna(x3.mean(), inplace=True)
    ...: print(x3.head())
          0         1         2         3         4         5         6  \
0 -0.560476  0.253319 -0.710407  0.787739  2.198810 -0.375603 -0.715242
1 -0.230177 -0.028547  0.256884  0.769042  1.312413 -0.561876 -0.752689
2  1.558708 -0.042870 -0.246692 -0.004695 -0.265145  0.274805 -0.938539
3  0.070508  1.368602 -0.347543 -1.008377  0.543194  0.090497 -1.052513
4  0.129288 -0.225771 -0.951619 -0.119453 -0.414340  1.598509 -0.437160

          7         8         9
0  1.014943 -0.073556  1.430402
1 -1.992748 -1.168651  1.046629
2 -0.427279  0.000000  0.435289
3  0.000000 -0.028842  0.715178
4 -0.893208  0.000000  0.917175
```

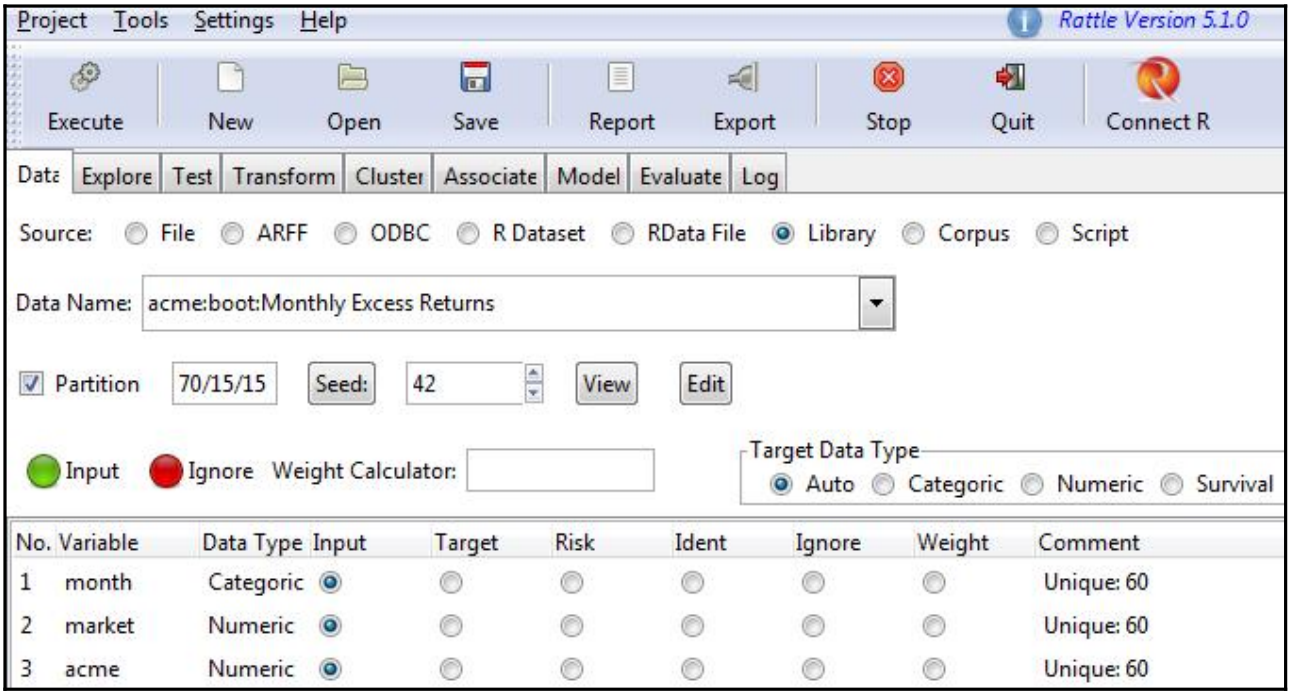
Chapter 04: Data Visualization




Histogram of data
Frequency
0
200
400
600
800
1000
-4
-2
0
2
4
data


Area under standard normal dist & x less than -2.33
standard normal distribution
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x value


R Data Miner - [Rattle]
Project Tools Settings Help
Rattle Version 5.1.0 togaware.com
Execute New Open Save Report Export Stop Quit Connect R
Data Explore Test Transform Cluster Associate Model Evaluate Log
Source: File ARFF ODBC R Dataset RData File Library Corpus Script
Filename: (None) Separator: , Decimal: . Header
Partition 70/15/15 Seed: 42 View Edit
Input Ignore Weight Calculator:
Target Data Type
Auto Categoric Numeric Survival
Welcome to Rattle (rattle.togaware.com).
Rattle is a free graphical user interface for Data Science, developed using R. R is a free software environment for statistical computing, graphics, machine learning and artificial intelligence. Together Rattle and R provide a sophisticated environment for data science, statistical analyses, and data visualisation.

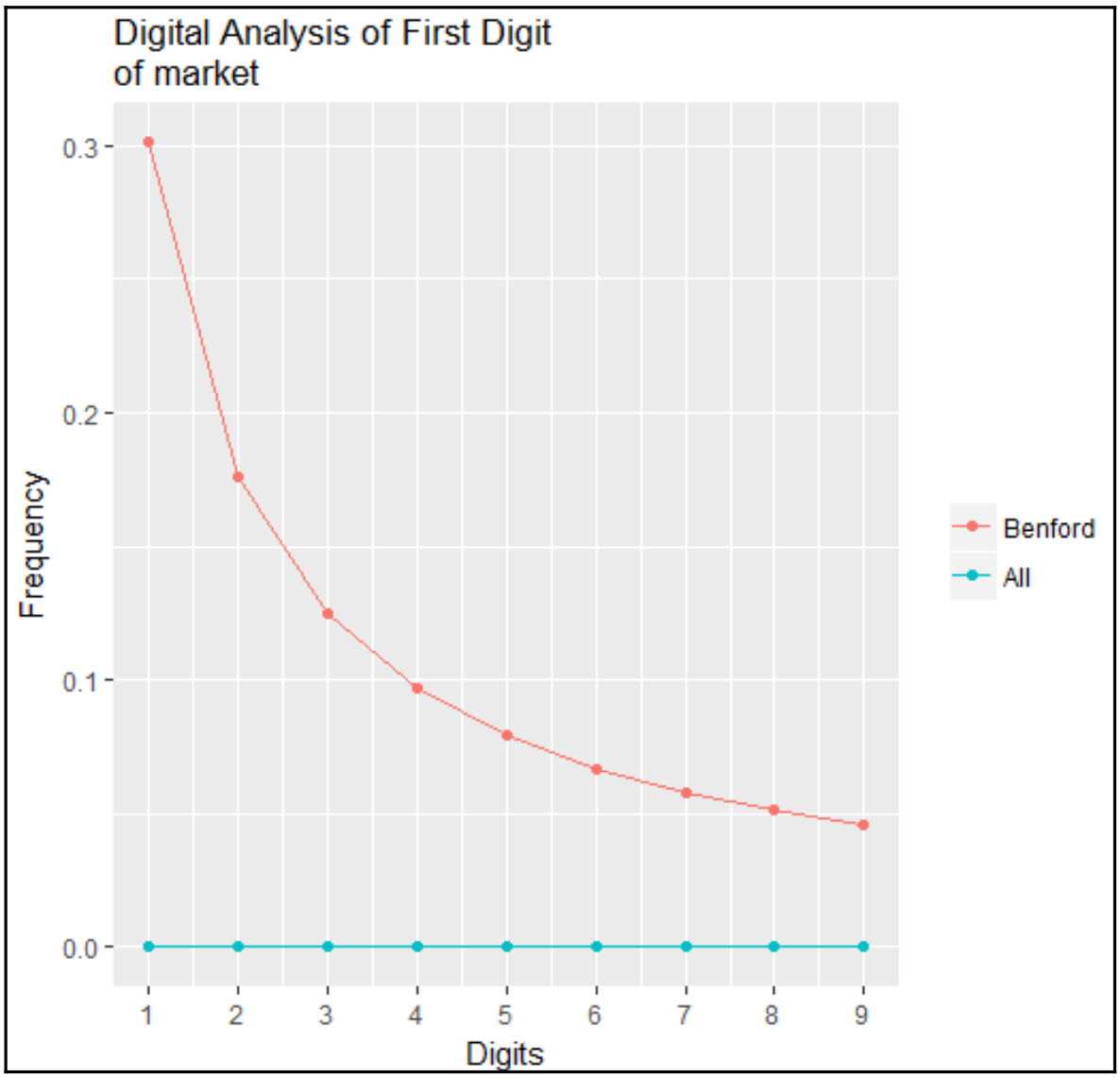

Project Tools Settings Help
Rattle Version 5.1.0
Execute New Open Save Report Export Stop Quit Connect R
Data Explore Test Transform Cluster Associate Model Evaluate Log
Source: File ARFF ODBC R Dataset RData File Library Corpus Script
Data Name: acme:boot:Monthly Excess Returns
Partition 70/15/15 Seed: 42 View Edit
Input Ignore Weight Calculator:
Target Data Type
Auto Categoric Numeric Survival
No. Variable Data Type Input Target Risk Ident Ignore Weight Comment
1 month Categoric Unique: 60
2 market Numeric Unique: 60
3 acme Numeric Unique: 60

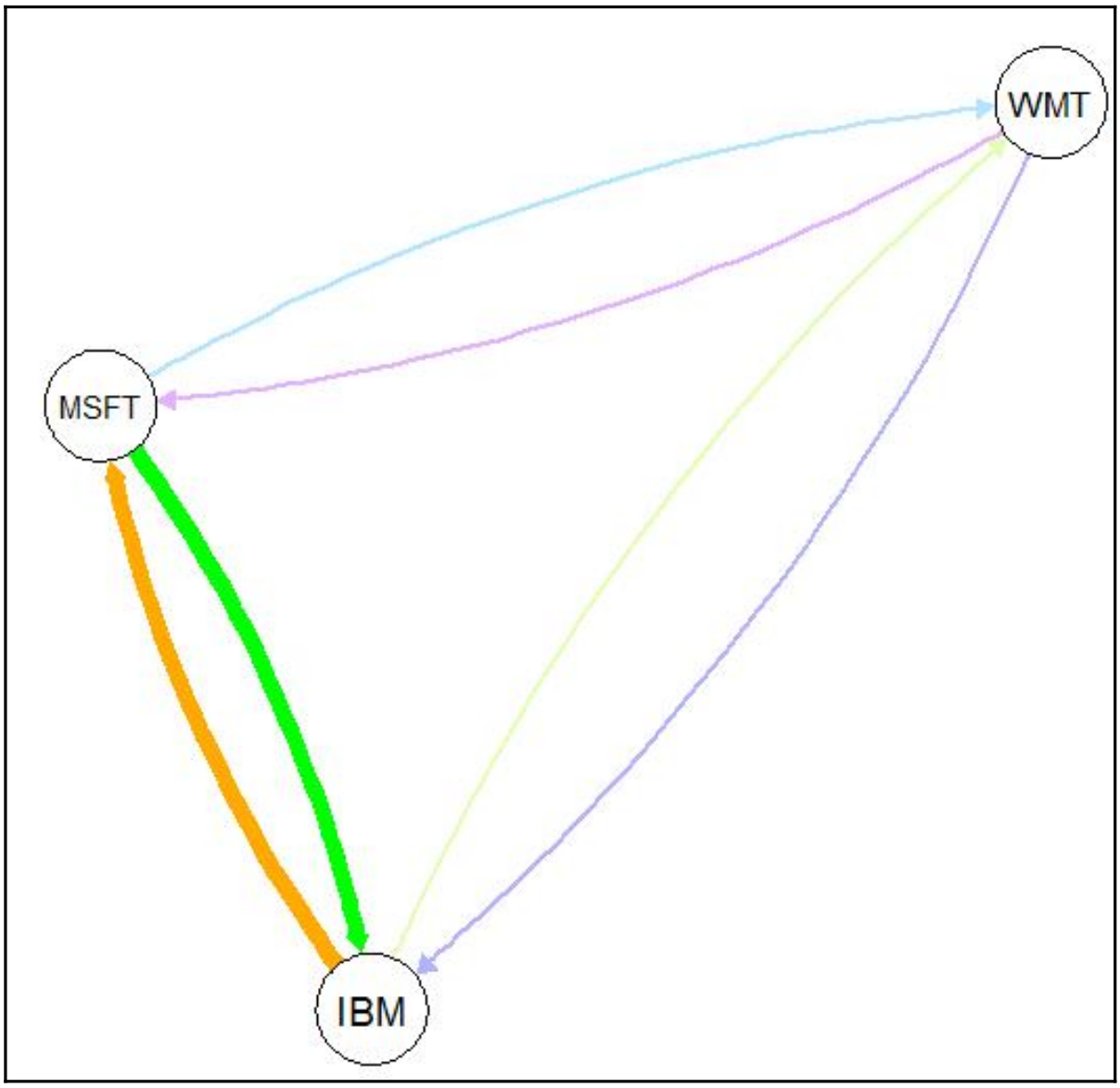
Type: Summary Distributions Correlation Principal Components Interactive

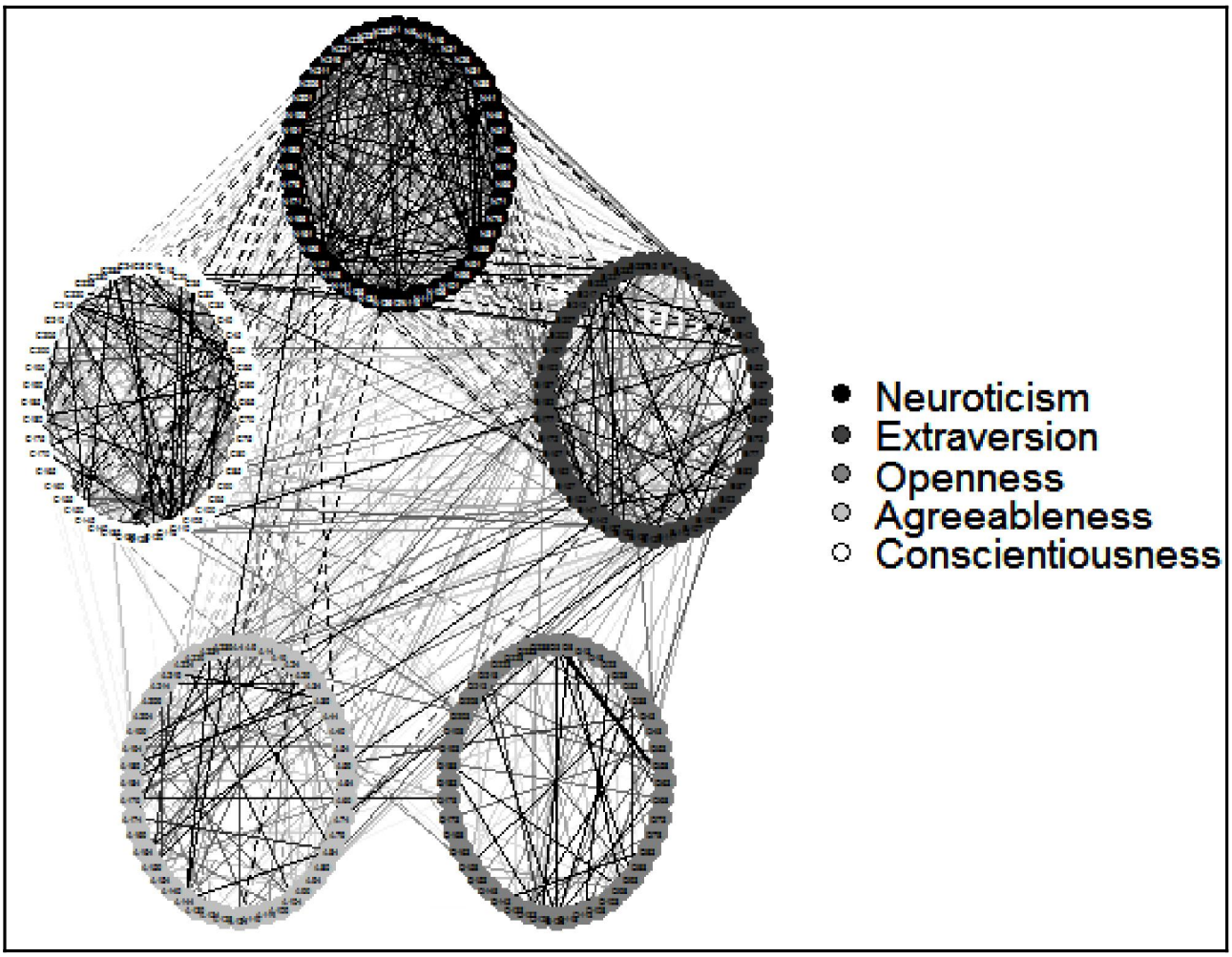
Numeric: Annotate **Group By:**

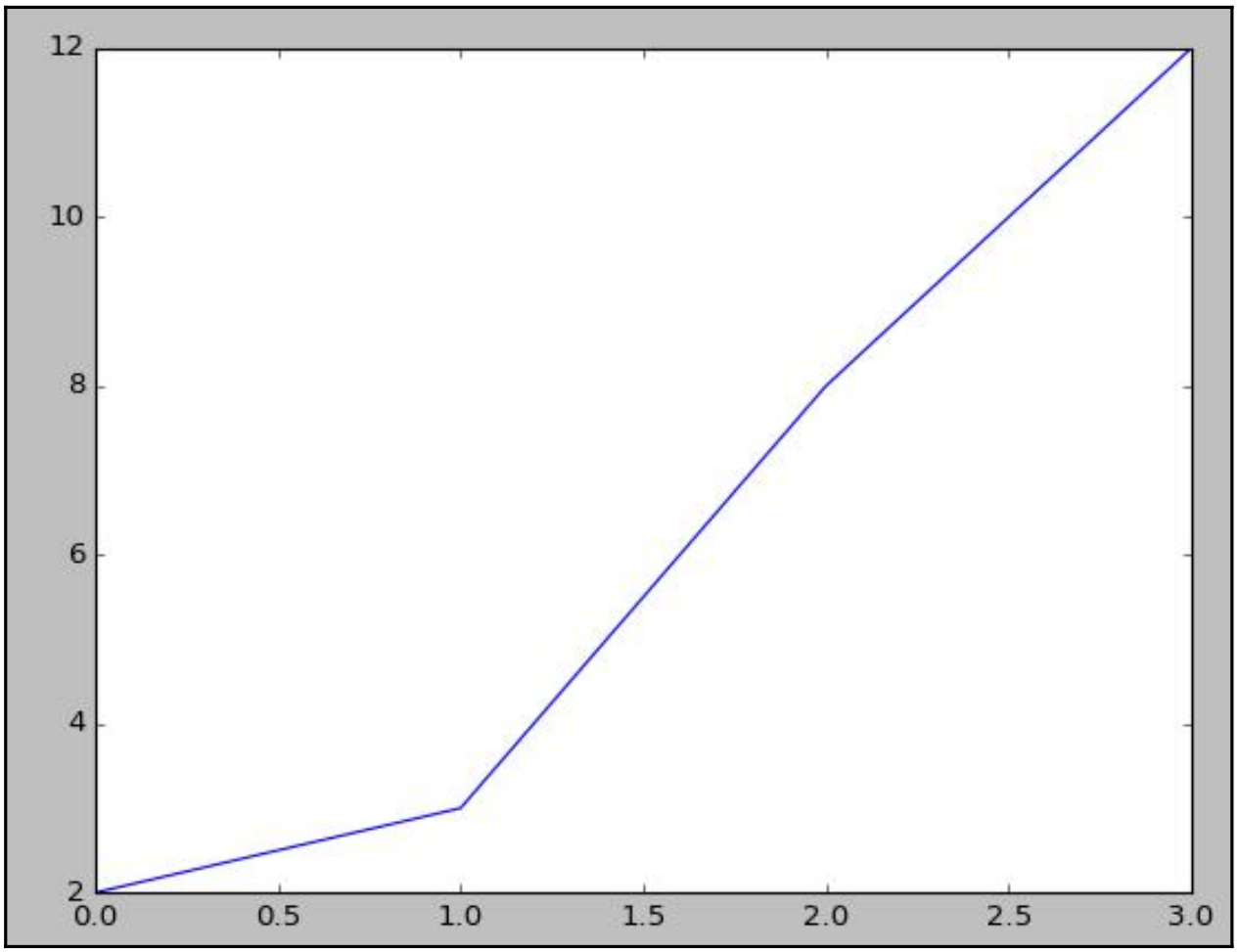
Benfords: Bars Starting Digit: 1 Digits: 1 abs +ve -ve

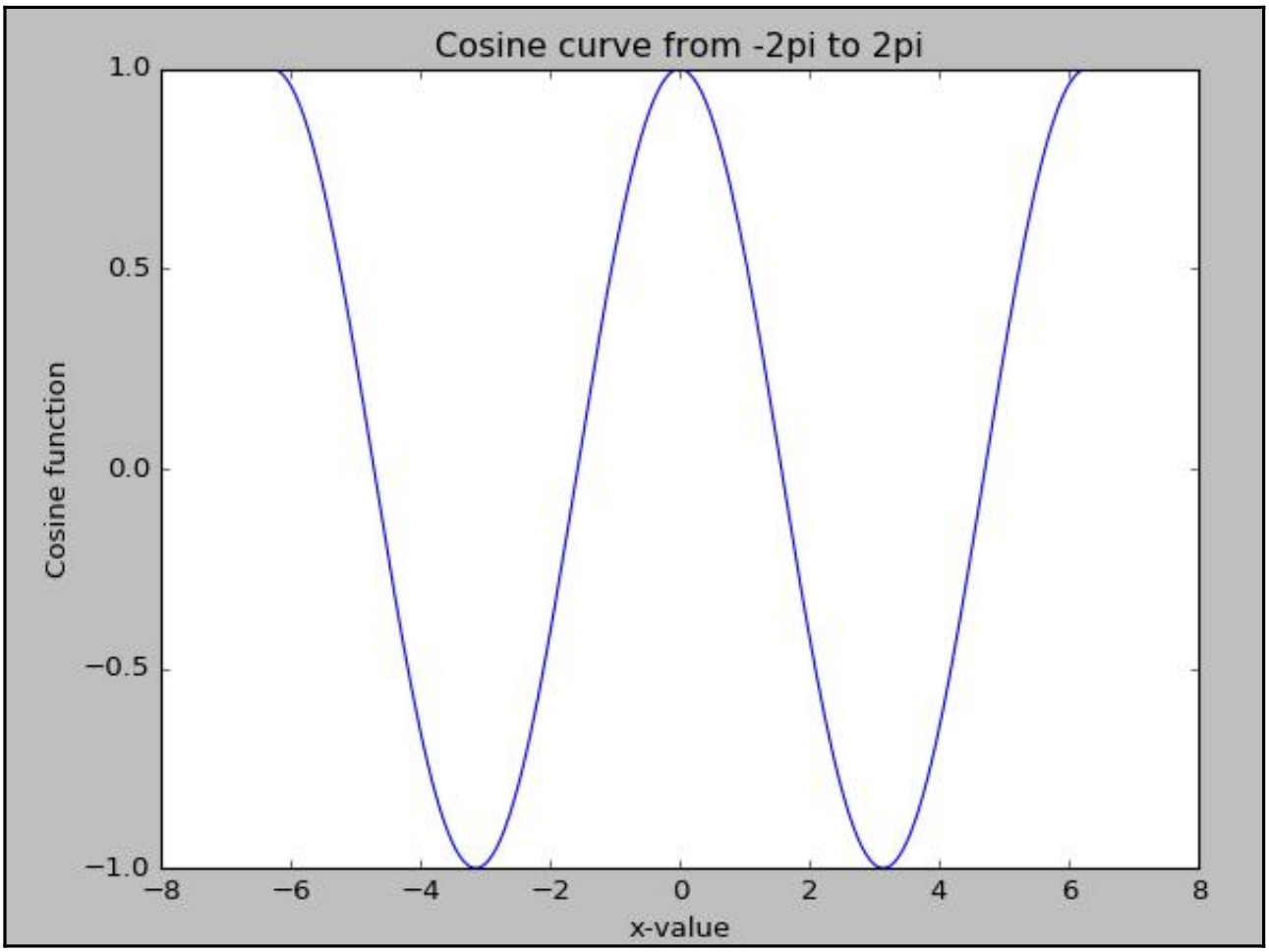
No.	Variable	Box Plot	Histogram	Cumulative	Benford	Pairs	Min; Median/Mean; Max
2	market	☐	☐	☐	☑	☐	-0.26; -0.04/-0.05; 0.07
3	acme	☐	☐	☐	☐	☐	-0.28; -0.09/-0.07; 0.24

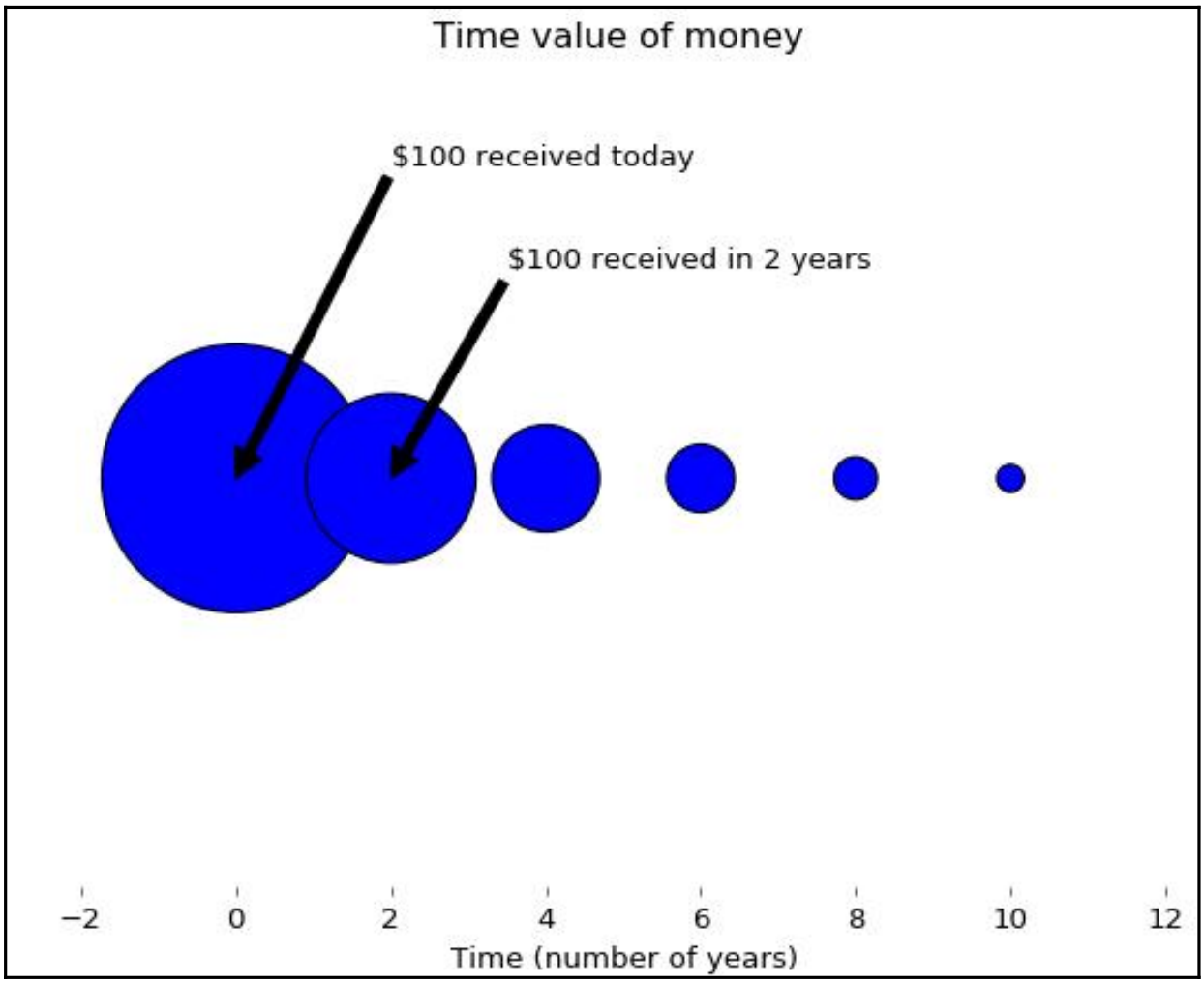


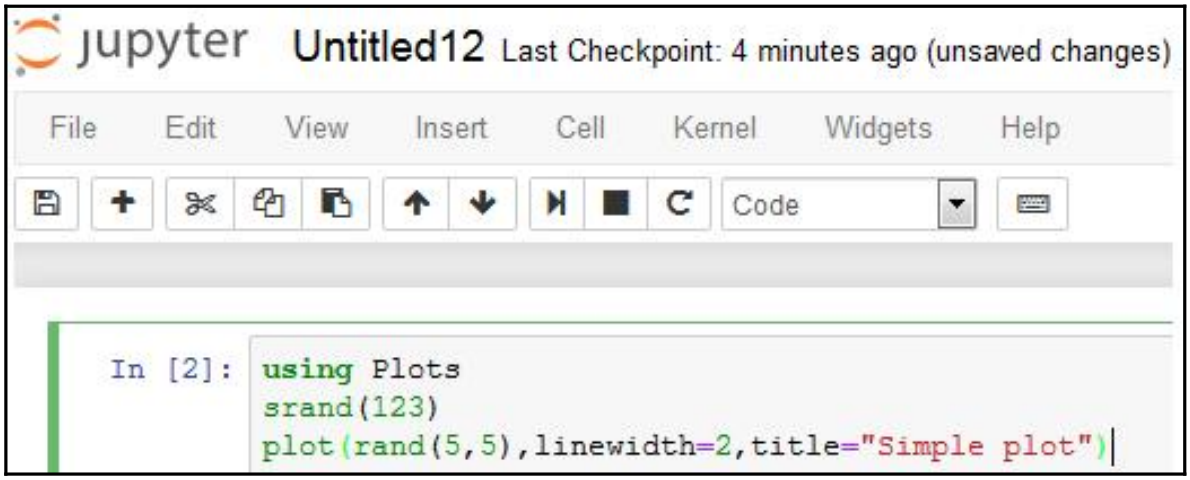

WMT
MSFT
IBM

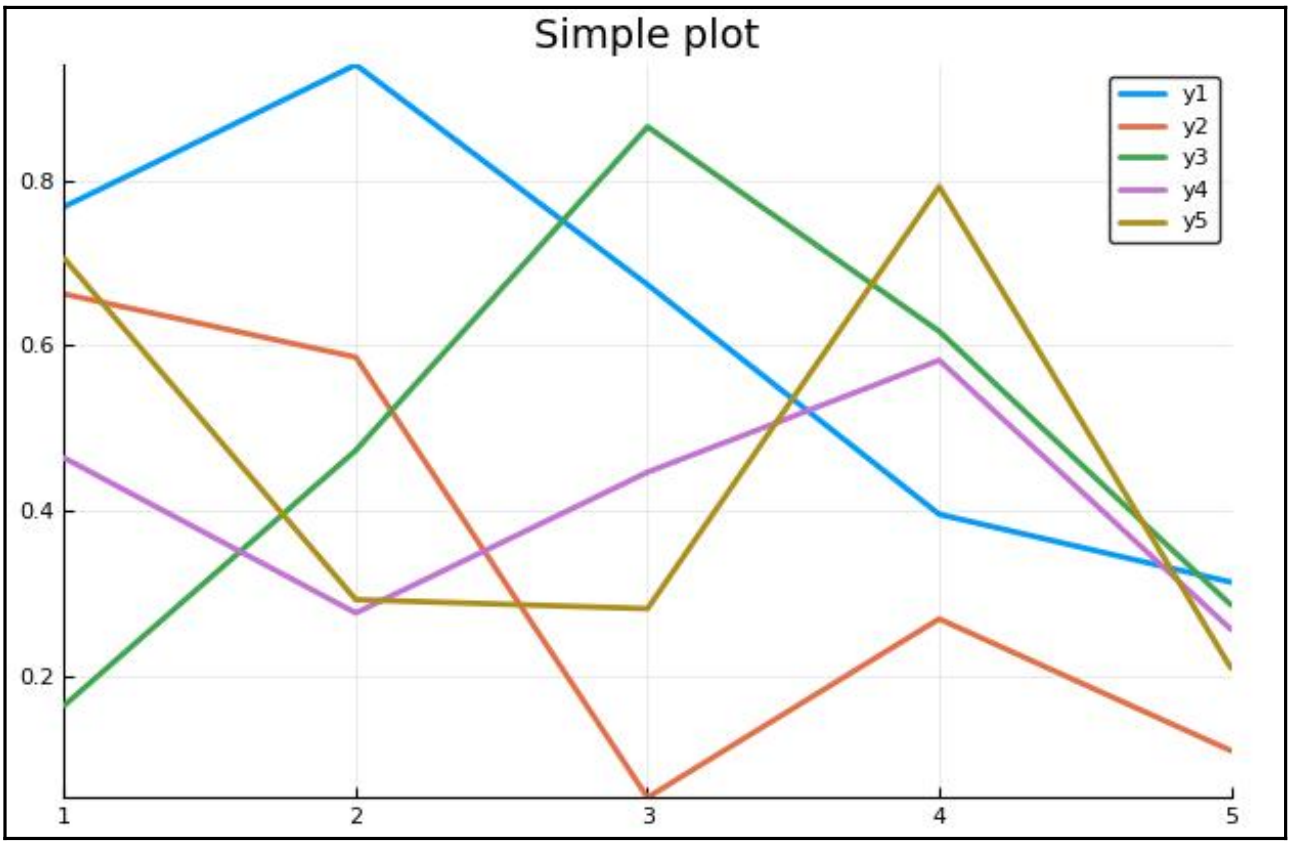

Neuroticism
Extraversion
Openness
Agreeableness
Conscientiousness

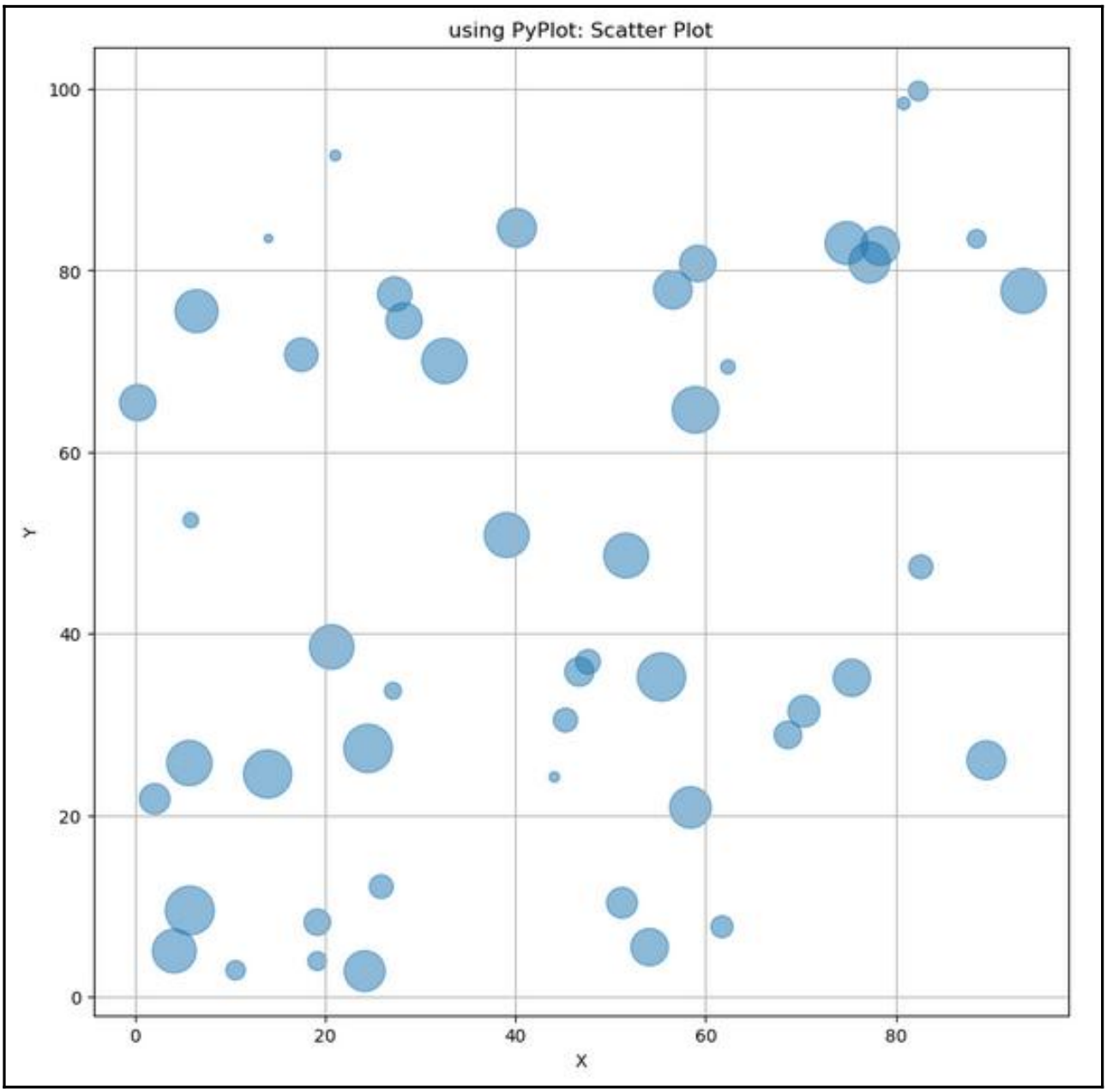

12
10
8
6
4
2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

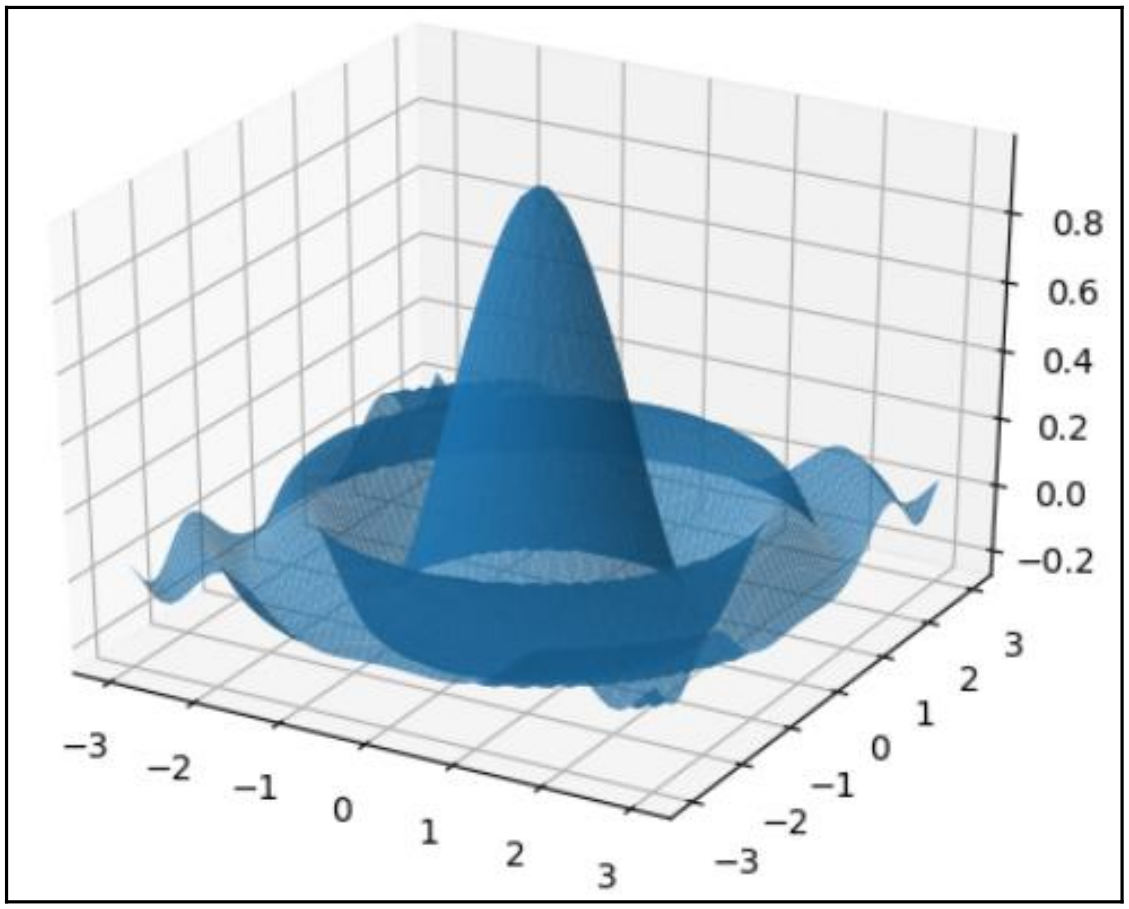

Cosine curve from -2pi to 2pi
Cosine function
x-value
1.0
0.5
0.0
−0.5
−1.0
−8
−6
−4
−2
0
2
4
6
8

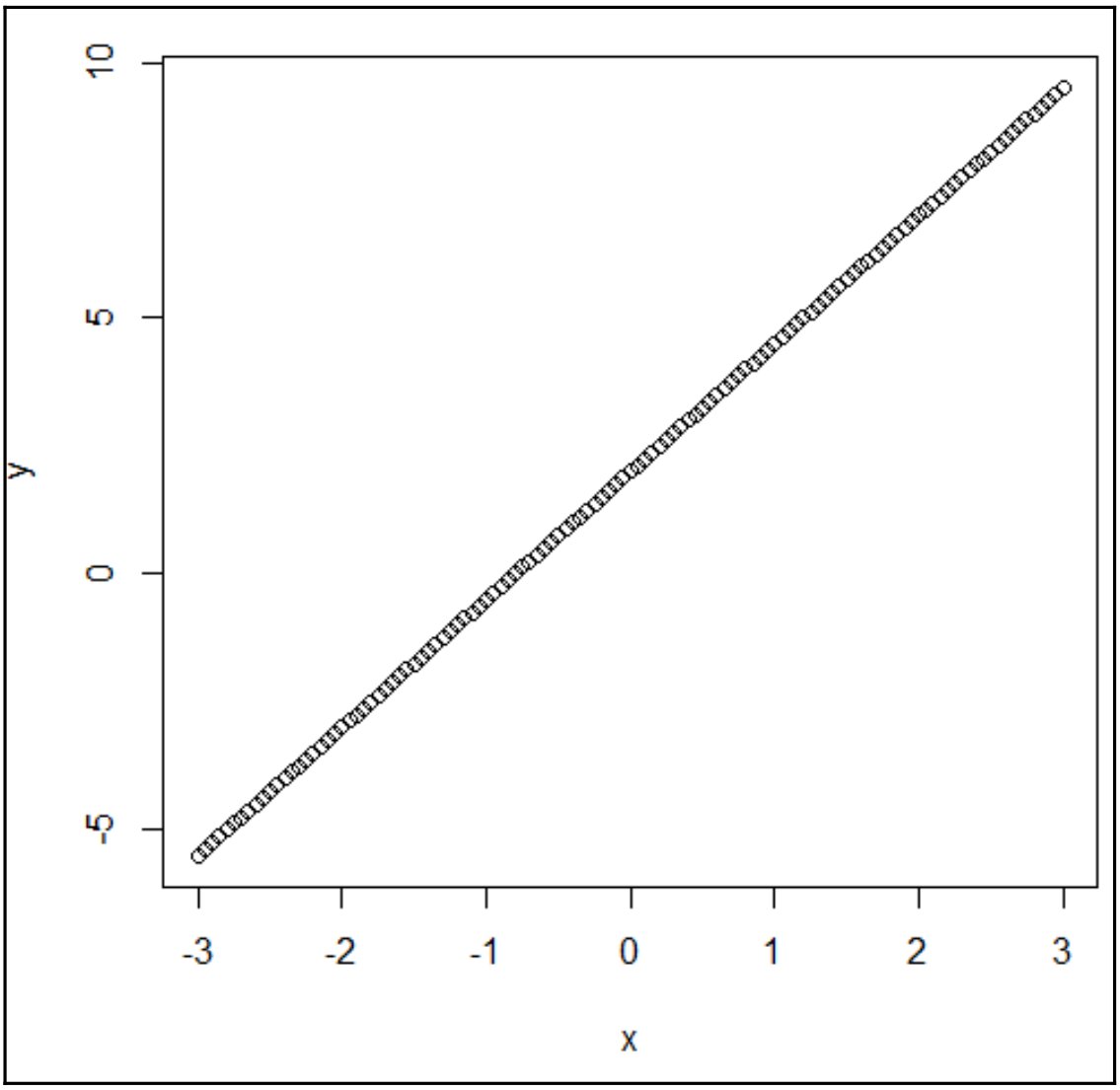

Time value of money
$100 received today
$100 received in 2 years
−2
0
2
4
6
8
10
12
Time (number of years)

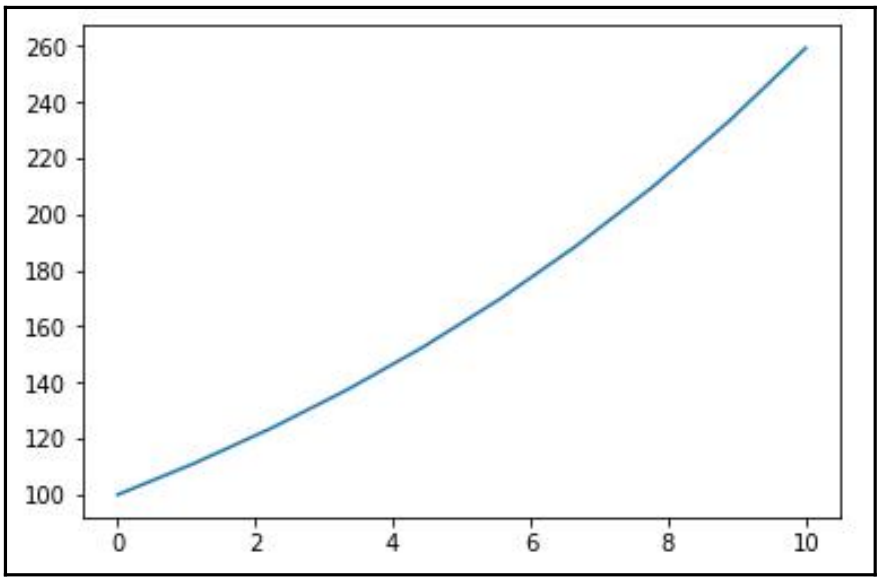

jupyter
Untitled12
Last Checkpoint: 4 minutes ago (unsaved changes)
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Code
In [2]:
using Plots
srand(123)
plot(rand(5,5),linewidth=2,title="Simple plot")

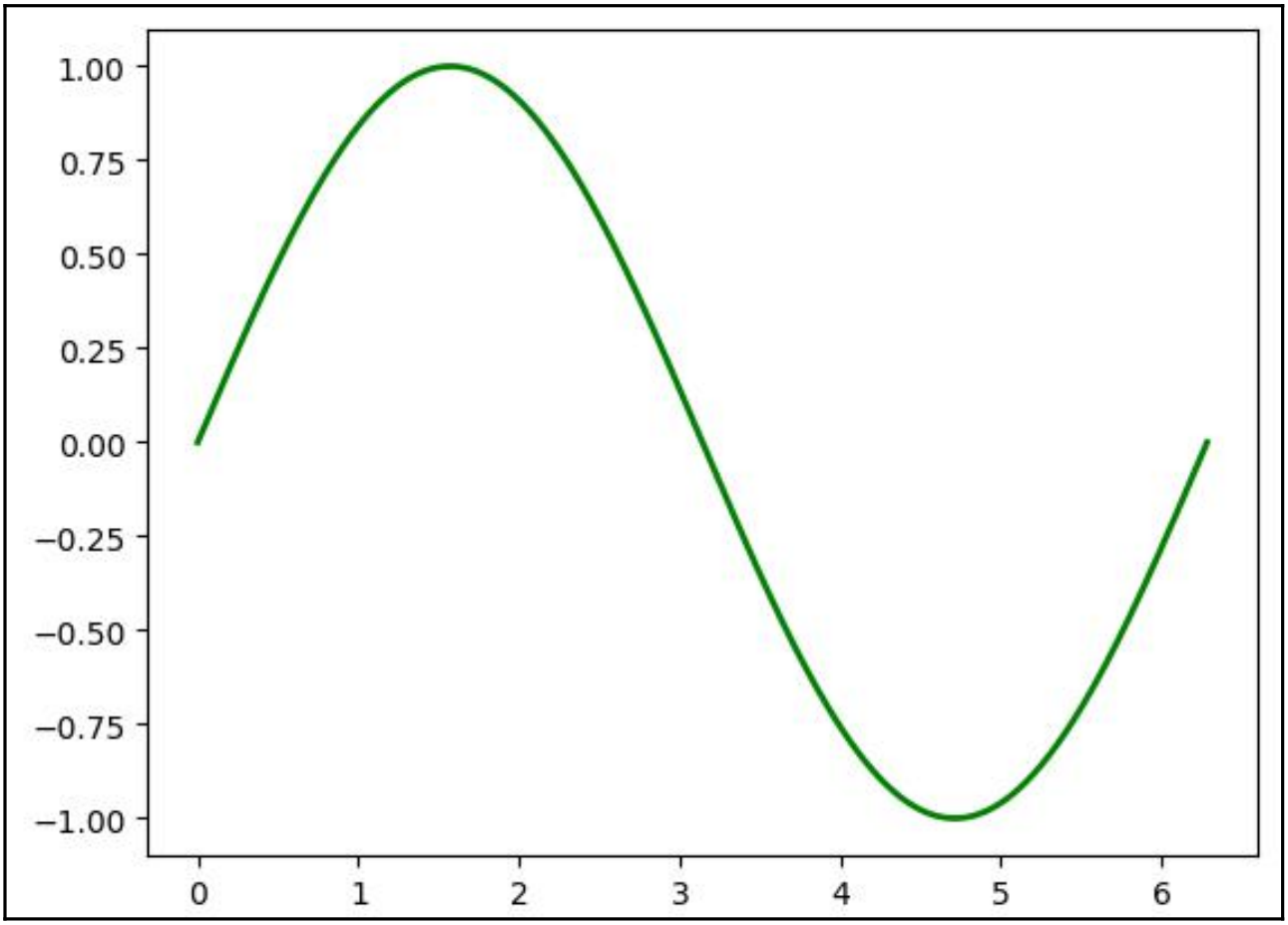

Simple plot
y1
y2
y3
y4
y5
0.8
0.6
0.4
0.2
1
2
3
4
5

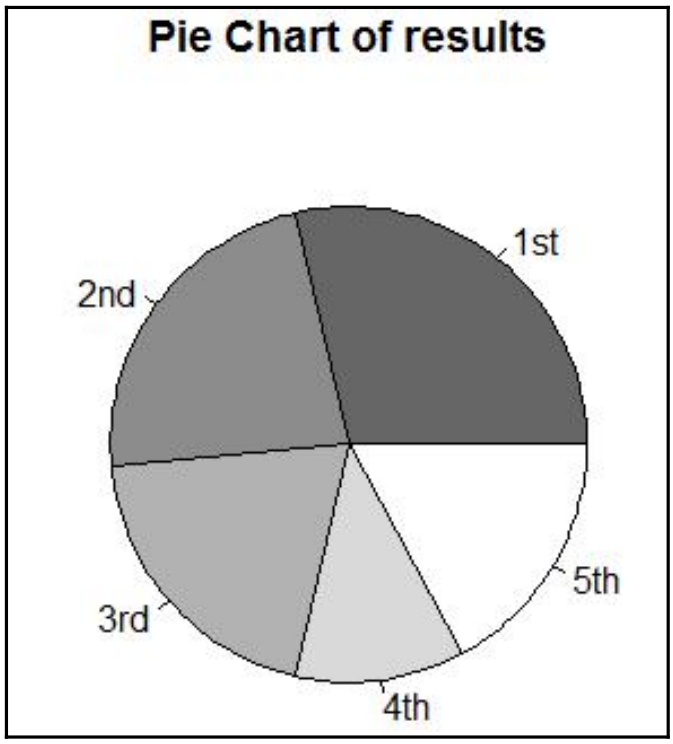

using PyPlot: Scatter Plot
100
80
60
40
20
0
Y
0
20
40
60
80
X

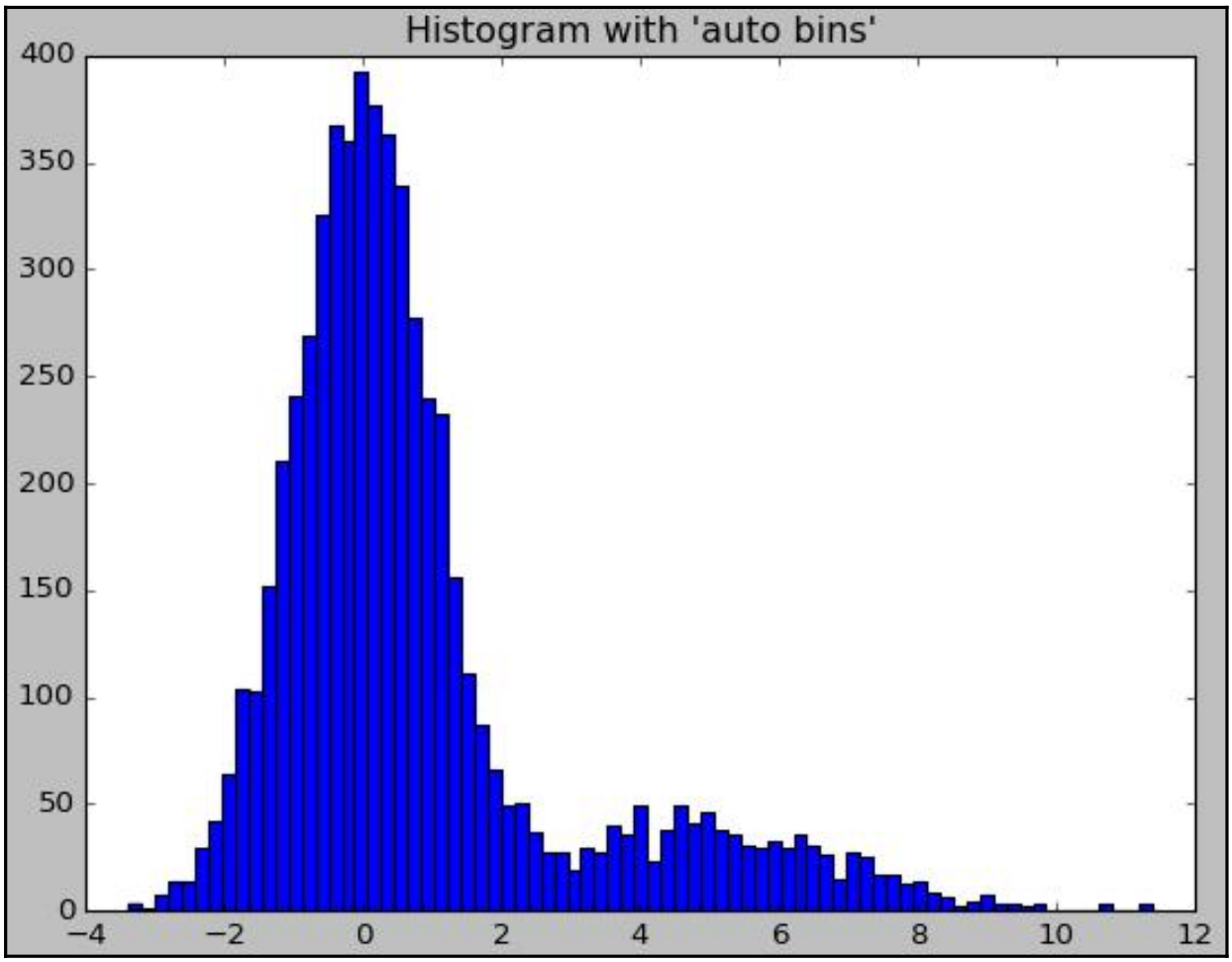

0.8
0.6
0.4
0.2
0.0
−0.2
−3
−2
−1
0
1
2
3
−3
−2
−1
0
1
2
3

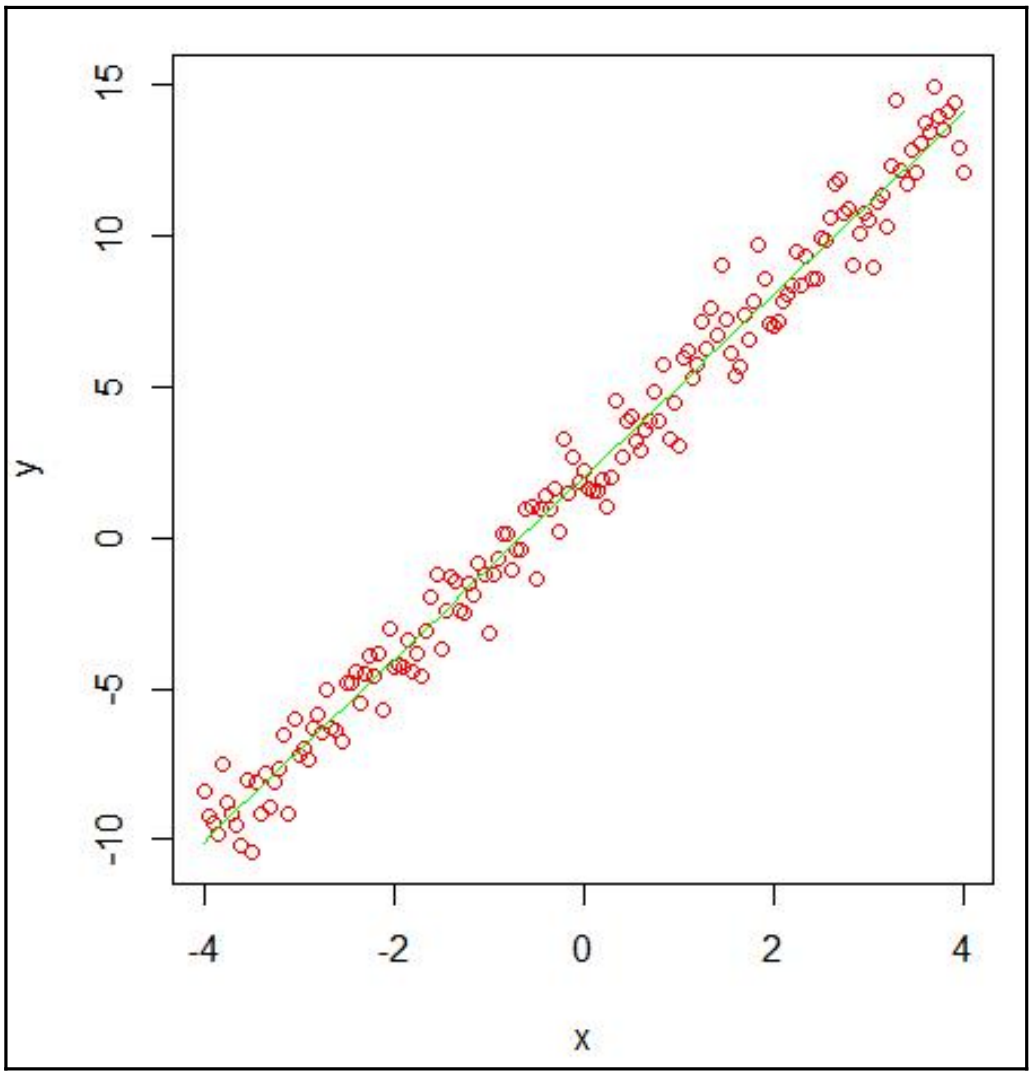

y
10
5
0
-5
-3
-2
-1
0
1
2
3
x

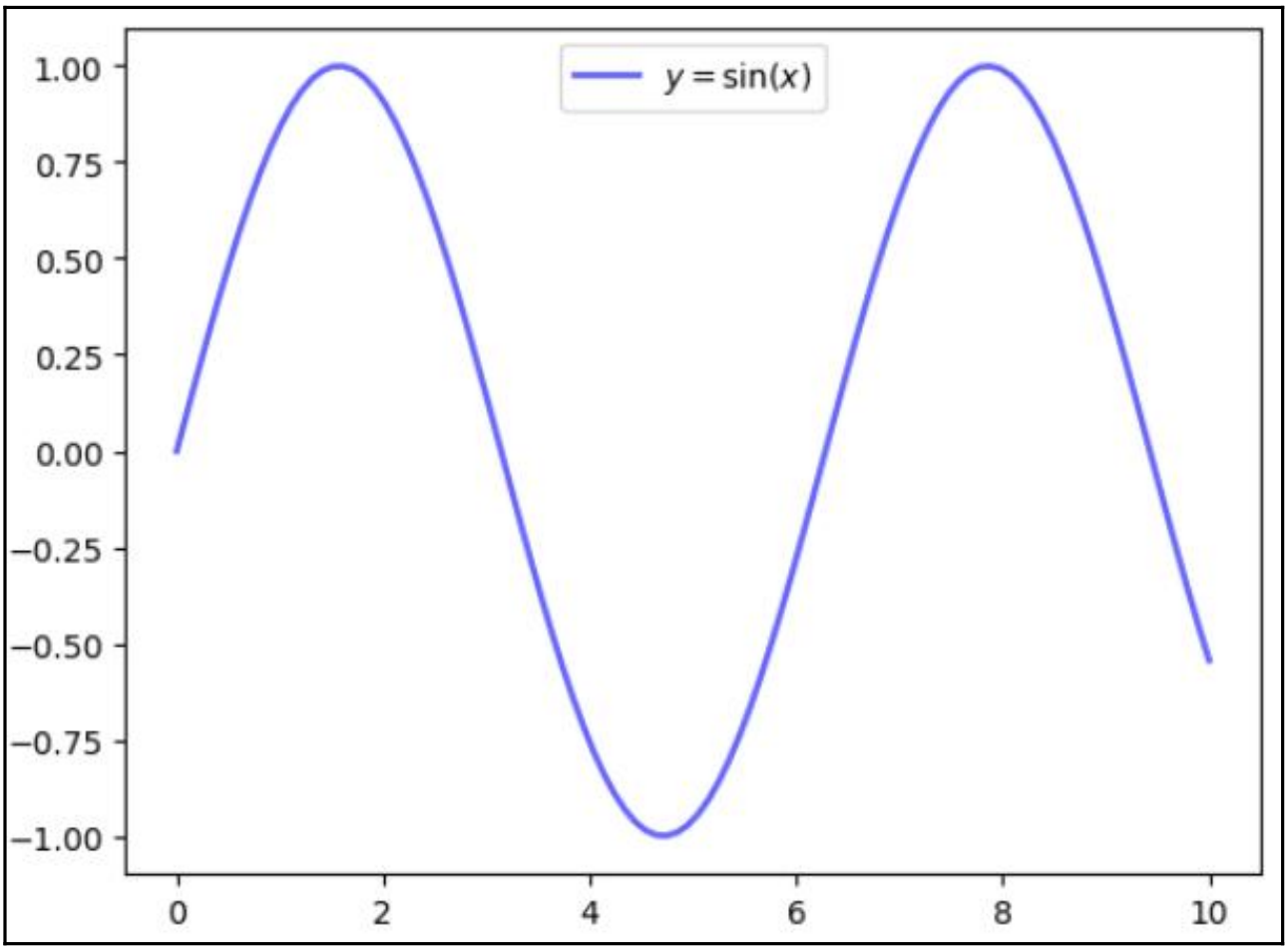

260
240
220
200
180
160
140
120
100
0
2
4
6
8
10

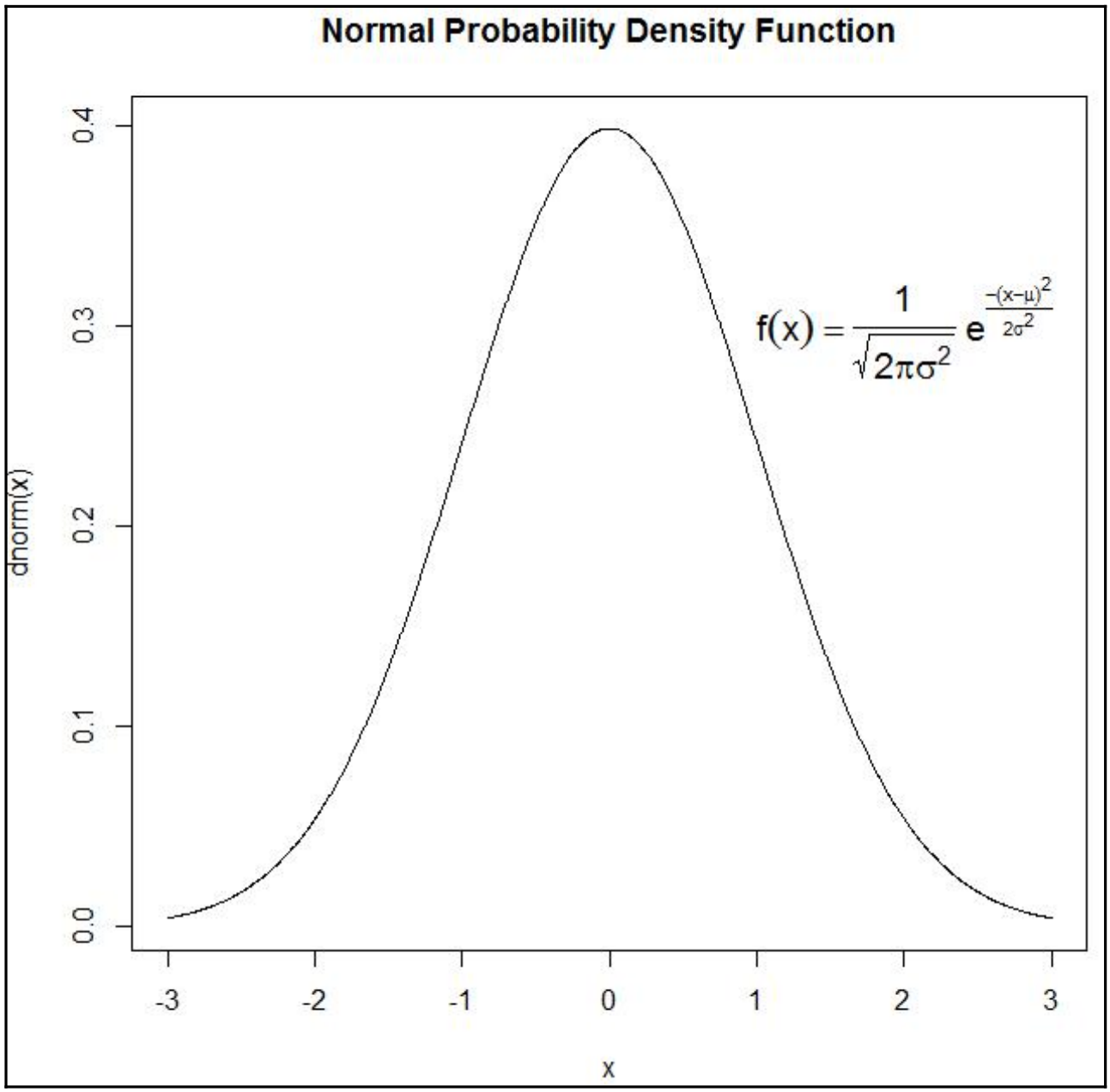

1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
1
2
3
4
5
6


Pie Chart of results
1st
2nd
3rd
4th
5th


Histogram with 'auto bins'
400
350
300
250
200
150
100
50
0
−4
−2
0
2
4
6
8
10
12


y
15
10
5
0
-5
-10
-4
-2
0
2
4
x


y = sin(x)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
2
4
6
8
10


Normal Probability Density Function
$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
dnorm(x)
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
x

$$c = S_0 N(d_1) - K e^{-rT} N(d_2)$$

$$d_1 = \frac{ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

Documentation for package 'ggplot2'

- DESCRIPTION file.
- User guides, package vignettes and other documentation.

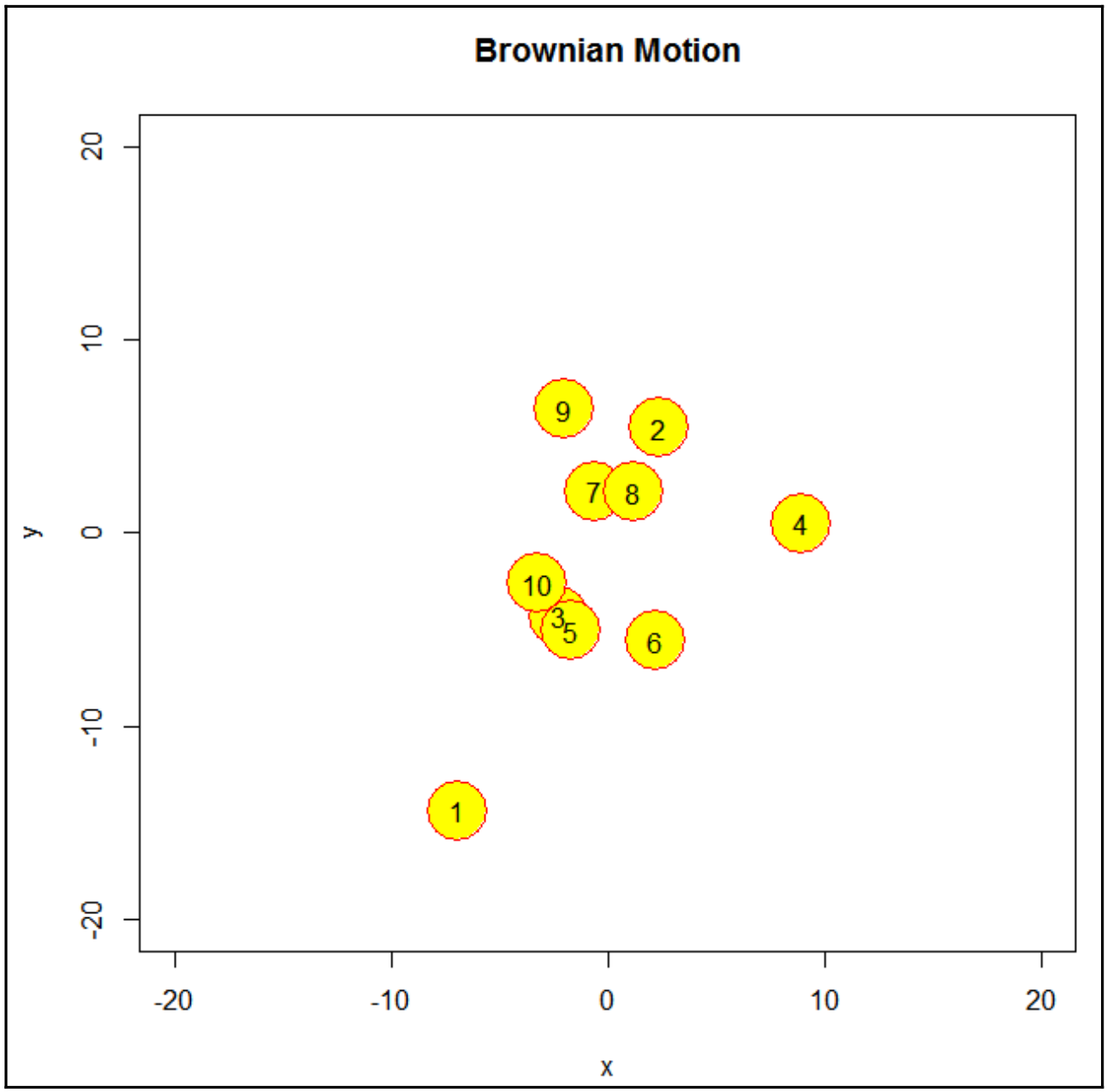
Help Pages

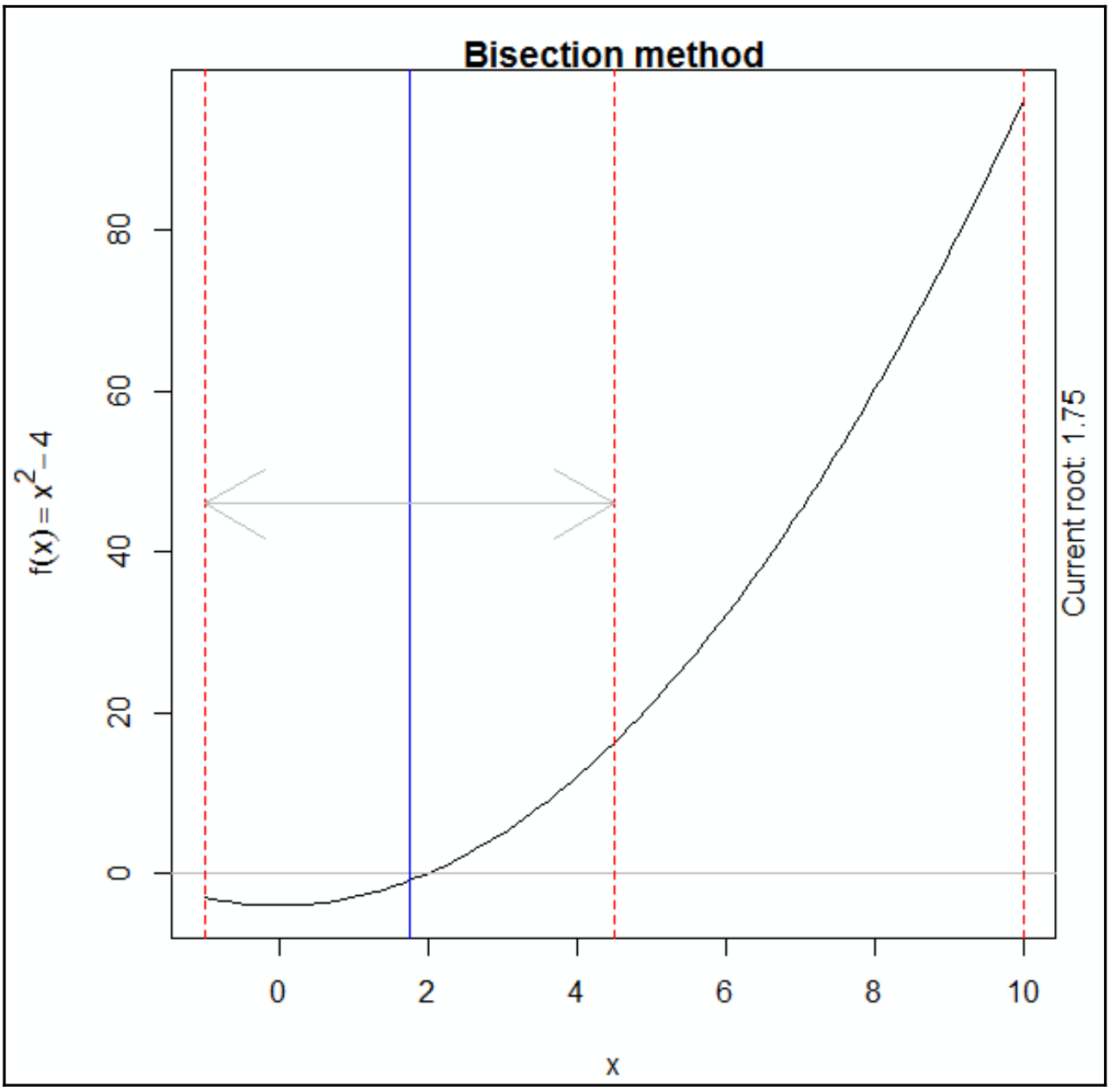
A B C D E F G H I L M P Q R S T X Y misc

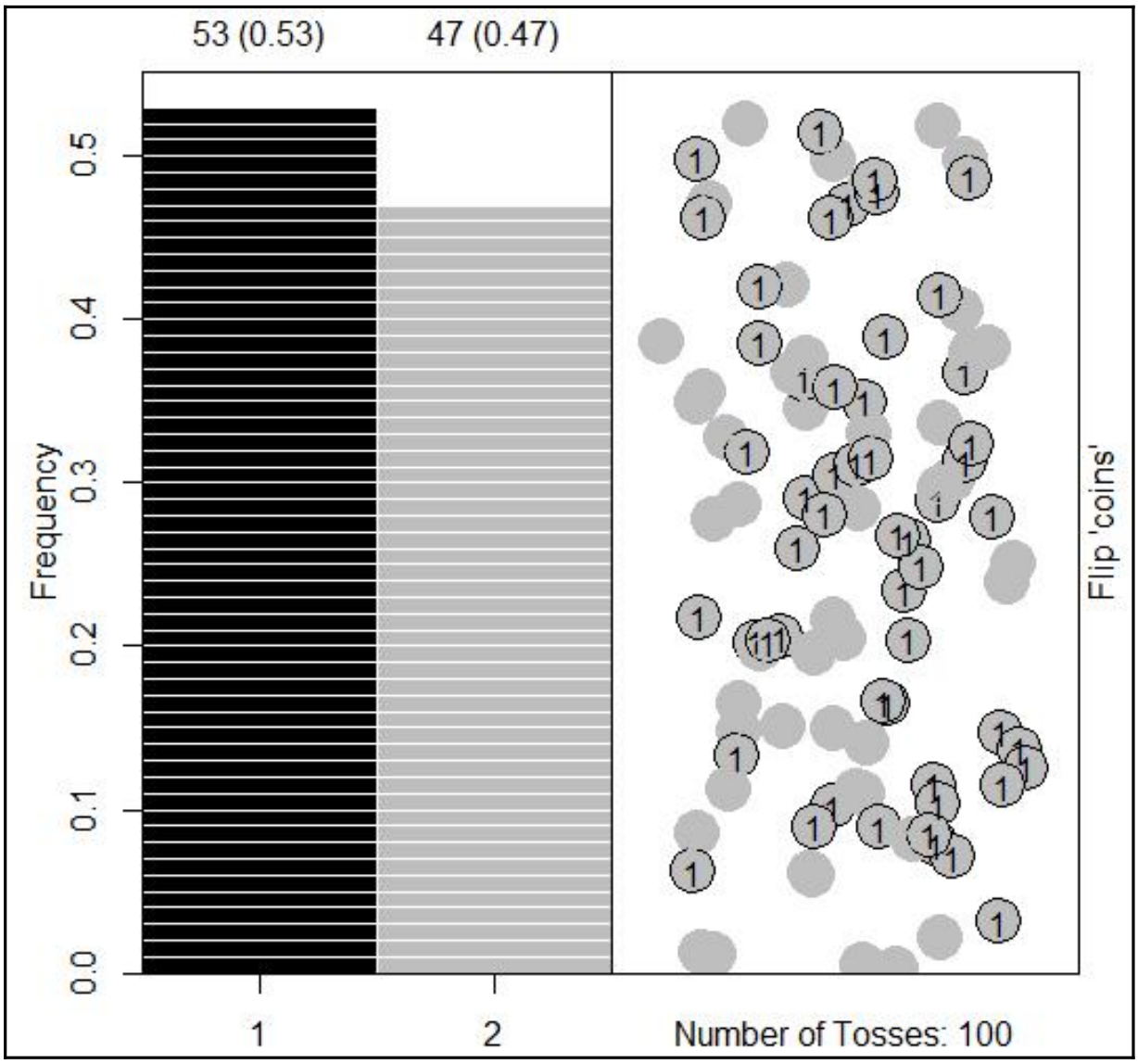
-- A --

aes	Construct aesthetic mappings
aes_	Define aesthetic mappings programatically
aes_colour_fill_alpha	Colour related aesthetics: colour, fill and alpha

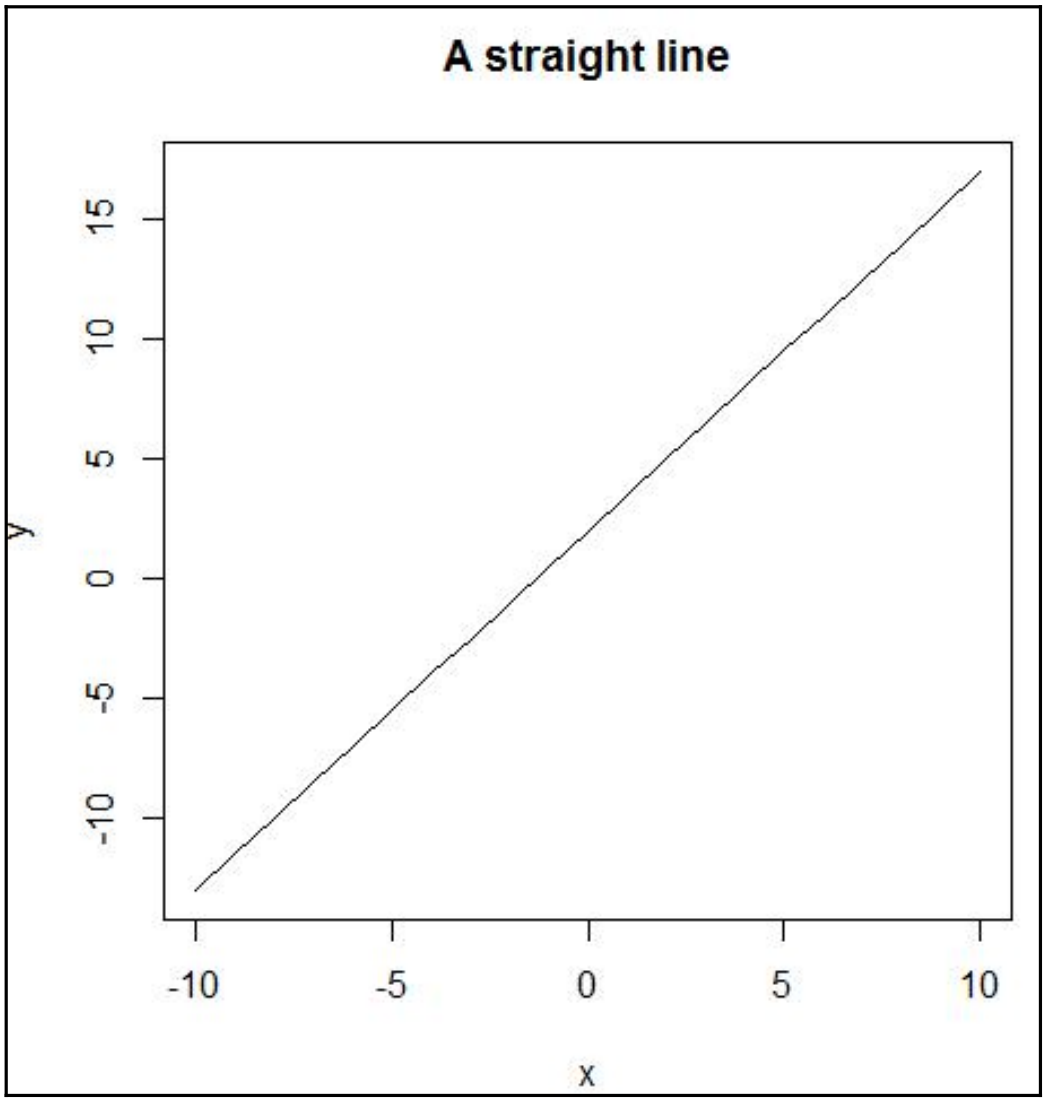
```
In [7]: print(x)
['MutableMapping', 'RcParams', 'URL_REGEX', 'Verbose', '_DATA_DOC_APPENDIX', '__bibtex__', '__builtins__', '__cached__',
'__doc__', '__file__', '__loader__', '__name__', '__package__', '__path__', '__spec__', '__version__', '__version__numpy__',
'_all_deprecated', '_backports', '_cm', '_cm_listed', '_cntr', '_color_data', '_contour', '_create_tmp_config_dir',
'_decode_filesystem_path', '_deprecated_ignore_map', '_deprecated_map', '_deprecated_set', '_error_details_fmt',
'_get_cachedir', '_get_config_or_cache_dir', '_get_configdir', '_get_data_path', '_get_data_path_cached', '_get_home',
'_get_xdg_cache_dir', '_get_xdg_config_dir', '_image', '_init_tests', '_is_writable_dir', '_mathtext_data', '_obsolete_set',
'_open_file_or_url', '_path', '_png', '_preprocess_data', '_pylab_helpers', '_python27', '_python34', '_qhull',
'_rc_params_in_file', '_replacer', '_tri', '_url_lines', '_use_error_msg', '_version', 'absolute_import', 'afm', 'artist',
'axes', 'axis', 'backend_bases', 'backend_tools', 'backends', 'bezier', 'blocking_input', 'category', 'cbook',
'checkdep_dvipng', 'checkdep_ghostscript', 'checkdep_inkscape', 'checkdep_pdftops', 'checkdep_ps_distiller', 'checkdep_tex',
'checkdep_usetex', 'checkdep_xmllint', 'cm', 'collections', 'colorbar', 'colors', 'compare_versions', 'compat', 'container',
'contextlib', 'contour', 'cycler', 'dates', 'dateutil', 'dedent', 'defaultParams', 'default_test_modules', 'distutils',
'division', 'docstring', 'dviread', 'figure', 'font_manager', 'fontconfig_pattern', 'ft2font', 'functools', 'get_backend',
'get_cachedir', 'get_configdir', 'get_data_path', 'get_home', 'get_label', 'get_py2exe_datafiles', 'gridspec', 'image',
'inspect', 'interactive', 'io', 'is_interactive', 'is_url', 'itertools', 'legend', 'legend_handler', 'lines', 'locale',
'markers', 'mathtext', 'matplotlib_fname', 'mlab', 'mplDeprecation', 'numpy', 'offsetbox', 'os', 'patches', 'path',
'print_function', 'projections', 'pyparsing', 'pyplot', 'quiver', 'rc', 'rcParams', 'rcParamsDefault', 'rcParamsOrig',
'rc_context', 'rc_file', 'rc_file_defaults', 'rc_params', 'rc_params_from_file', 'rcdefaults', 'rcsetup', 're', 'reload',
'sanitize_sequence', 'scale', 'six', 'spines', 'stackplot', 'streamplot', 'style', 'subprocess', 'sys', 'table', 'tempfile',
'test', 'texmanager', 'text', 'textpath', 'ticker', 'tight_bbox', 'tk_window_focus', 'transforms', 'tri', 'unicode_literals',
'units', 'urlopen', 'use', 'validate_backend', 'verbose', 'warnings', 'widgets']
```

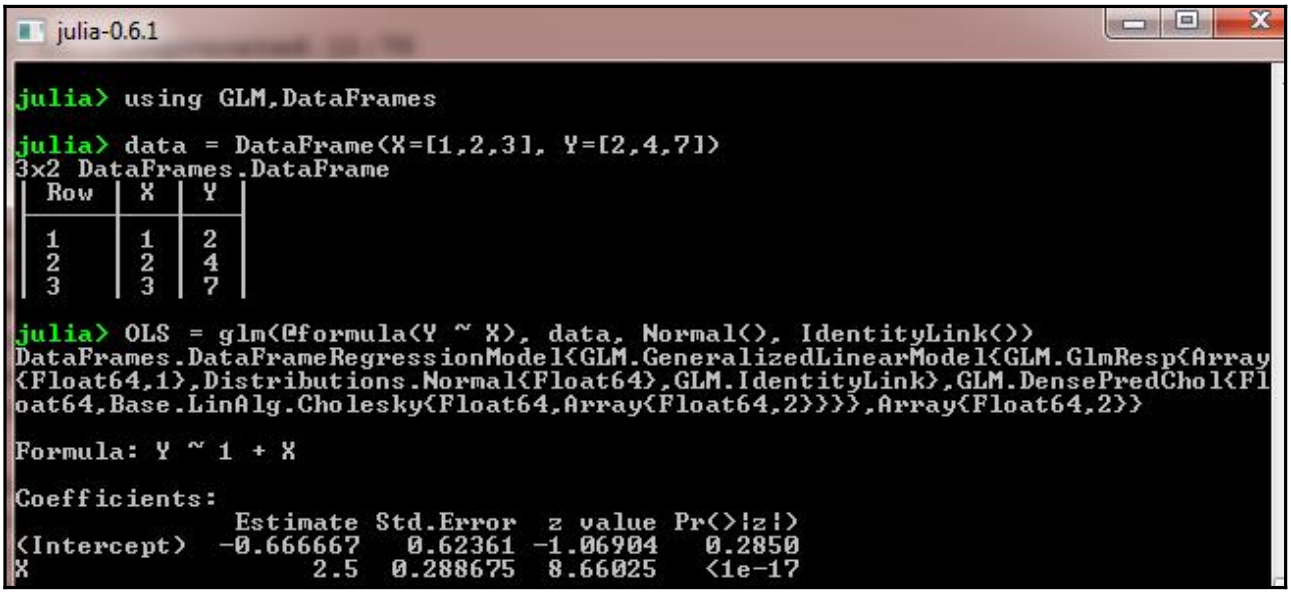

Brownian Motion
y
x
20
10
0
-10
-20
-20
-10
0
10
20
9
2
7
8
4
10
3
5
6
1


Bisection method
f(x) = x^2 − 4
Current root: 1.75
0
20
40
60
80
0
2
4
6
8
10
x

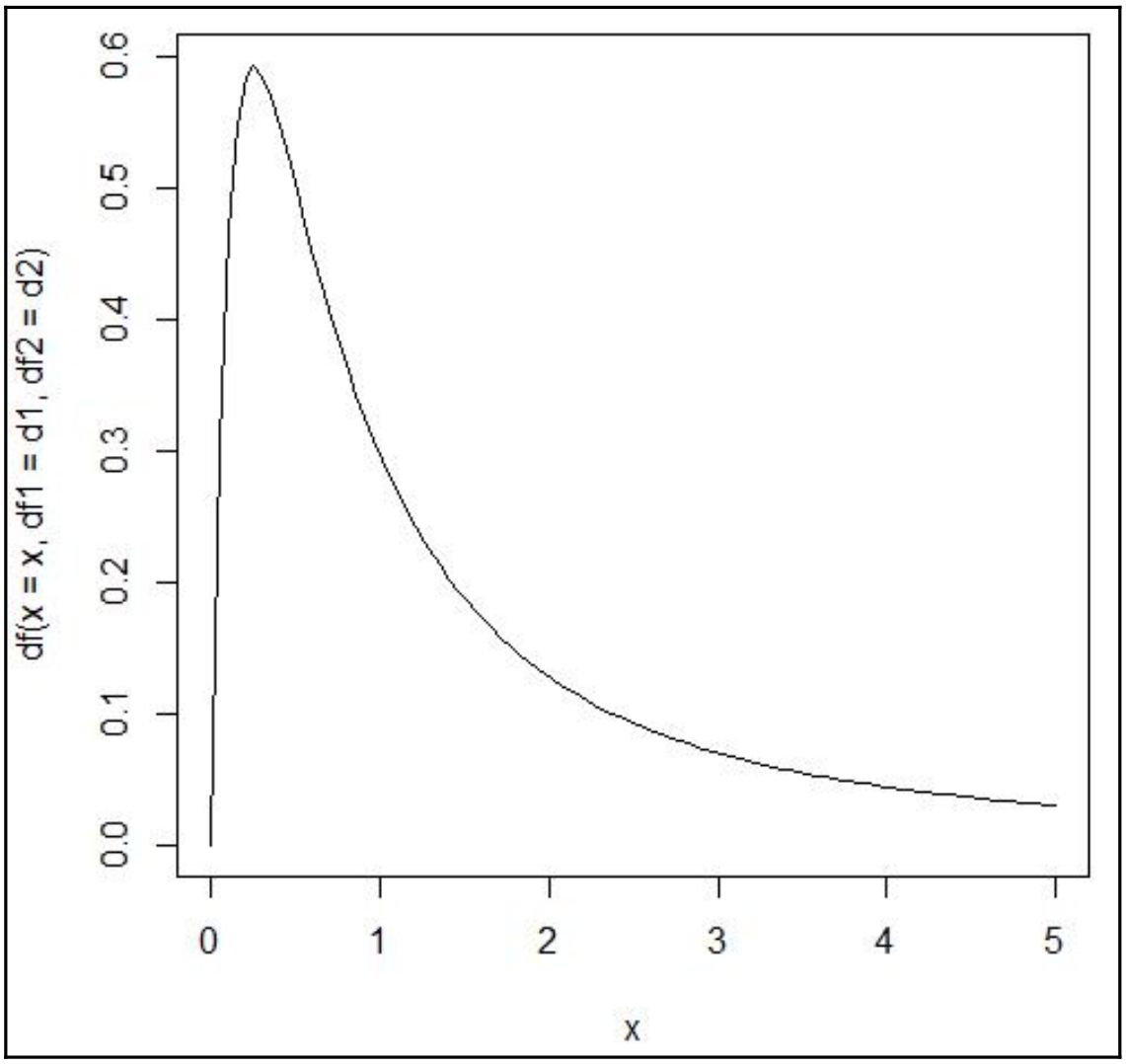

53 (0.53)
47 (0.47)
Frequency
0.0
0.1
0.2
0.3
0.4
0.5
1
2
Number of Tosses: 100
Flip 'coins'

Chapter 05: Statistical Modeling in Anaconda




julia-0.6.1
julia> using GLM,DataFrames
julia> data = DataFrame(X=[1,2,3], Y=[2,4,7])
3x2 DataFrames.DataFrame
Row	X	Y
1	1	2
2	2	4
3	3	7
julia> OLS = glm(@formula(Y ~ X), data, Normal(), IdentityLink())
DataFrames.DataFrameRegressionModel{GLM.GeneralizedLinearModel{GLM.GlmResp{Array{Float64,1},Distributions.Normal{Float64},GLM.IdentityLink},GLM.DensePredChol{Float64,Base.LinAlg.Cholesky{Float64,Array{Float64,2}}}},Array{Float64,2}}
Formula: Y ~ 1 + X
Coefficients:
Estimate Std.Error z value Pr(>|z|)
(Intercept) -0.666667 0.62361 -1.06904 0.2850
X 2.5 0.288675 8.66025 <1e-17


A fresh approach to technical computing
Documentation: https://docs.julialang.org
Type "?help" for help.
Version 0.6.1 (2017-10-24 22:15 UTC)
Official http://julialang.org/ release
i686-w64-mingw32
julia> Pkg.add("GLM")
INFO: Cloning cache of GLM from https://github.com/JuliaStats/GLM.jl.git
INFO: Installing GLM v0.8.1
INFO: Building SpecialFunctions
INFO: Building Rmath
INFO: Package database updated
INFO: METADATA is out-of-date - you may not have the latest version of GLM
INFO: Use `Pkg.update()` to get the latest versions of your packages
julia> Pkg.update()
INFO: Updating METADATA...


df(x = x, df1 = d1, df2 = d2)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3
4
5
x

```
> summary(lm(final$RET~final$MKTRET))

Call:
lm(formula = final$RET ~ final$MKTRET)

Residuals:
      Min        1Q    Median        3Q       Max
-0.261101 -0.031148 -0.002537  0.029451  0.281473

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.00277    0.00217   1.277    0.202
final$MKTRET  0.96442    0.05060  19.061   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0555 on 667 degrees of freedom
Multiple R-squared:  0.3526,    Adjusted R-squared:  0.3517
F-statistic: 363.3 on 1 and 667 DF,  p-value: < 2.2e-16
```

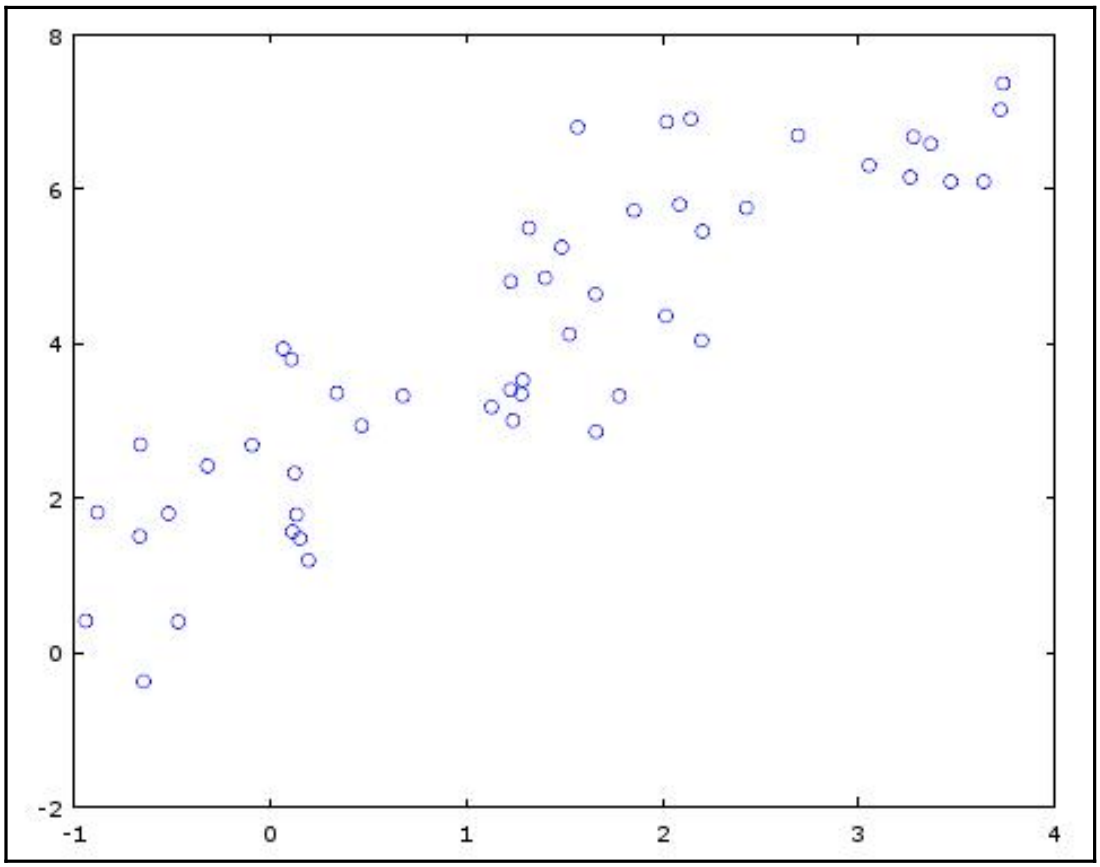
```
In [1]: %run "c:\users\yany\appdata\local\temp\tmphevjxh.py"
             Open   High    Low  Close   Volume  Ex-Dividend  Split Ratio  \
Date
1962-01-02  578.5  578.5  572.0  572.0  19360.0          0.0          1.0
1962-01-03  572.0  577.0  572.0  577.0  14400.0          0.0          1.0

            Adj. Open  Adj. High  Adj. Low  Adj. Close  Adj. Volume
Date
1962-01-02  15.270839  15.270839  15.099257   15.099257     387200.0
1962-01-03  15.099257  15.231243  15.099257   15.231243     288000.0
              Open    High     Low   Close     Volume  Ex-Dividend  \
Date
2018-02-06  150.29  155.49  149.11  155.34  9823824.0          0.0
2018-02-07  154.17  155.34  153.28  153.85  6069258.0          0.0

            Split Ratio  Adj. Open  Adj. High  Adj. Low  Adj. Close  \
Date
2018-02-06          1.0     150.29     155.49    149.11      155.34
2018-02-07          1.0     154.17     155.34    153.28      153.85

            Adj. Volume
Date
2018-02-06    9823824.0
2018-02-07    6069258.0
```

```
    year,     alpha,   beta,R_value, P_value
(u'1972', 0.00073605, 0.054, 0.19, 0.081888902299272023)
(u'1973', -0.00040998, 0.092, 0.3, 1.2421671282048366e-06)
(u'1974', -0.00120897, 0.11, 0.261, 2.5687411810107753e-05)
(u'1975', 0.00080998, 0.072, 0.22, 0.00042128685673863711)
(u'1976', 0.00064678, 0.076, 0.253, 4.5885040919840052e-05)
```

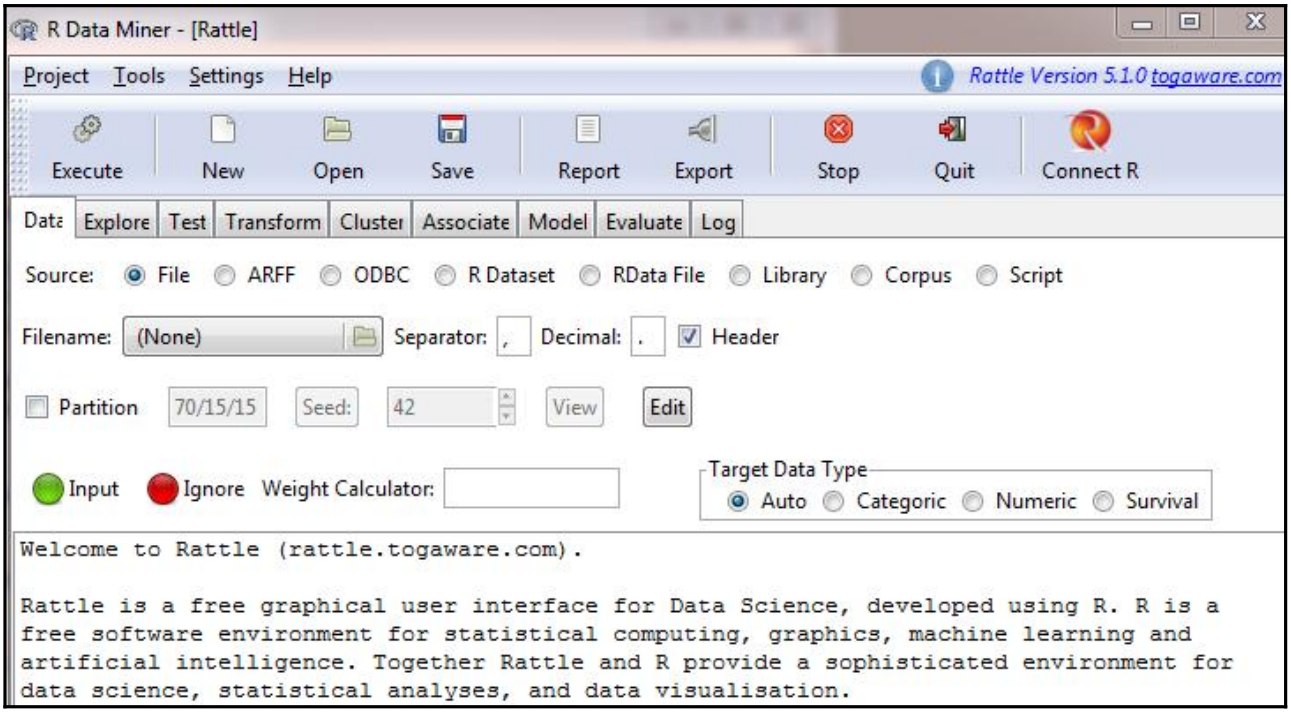
```
julia> OLS = glm(@formula(Y ~ X), data, Normal(), IdentityLink())
DataFrames.DataFrameRegressionModel{GLM.GeneralizedLinearModel{GLM.GlmResp{Array
{Float64,1},Distributions.Normal{Float64},GLM.IdentityLink},GLM.DensePredChol{Fl
oat64,Base.LinAlg.Cholesky{Float64,Array{Float64,2}}}},Array{Float64,2}}

Formula: Y ~ 1 + X

Coefficients:
             Estimate  Std.Error  z value Pr(>|z|)
(Intercept)  0.00617165 0.00863561 0.714674   0.4748
X              0.145727   0.282945 0.515036   0.6065
```

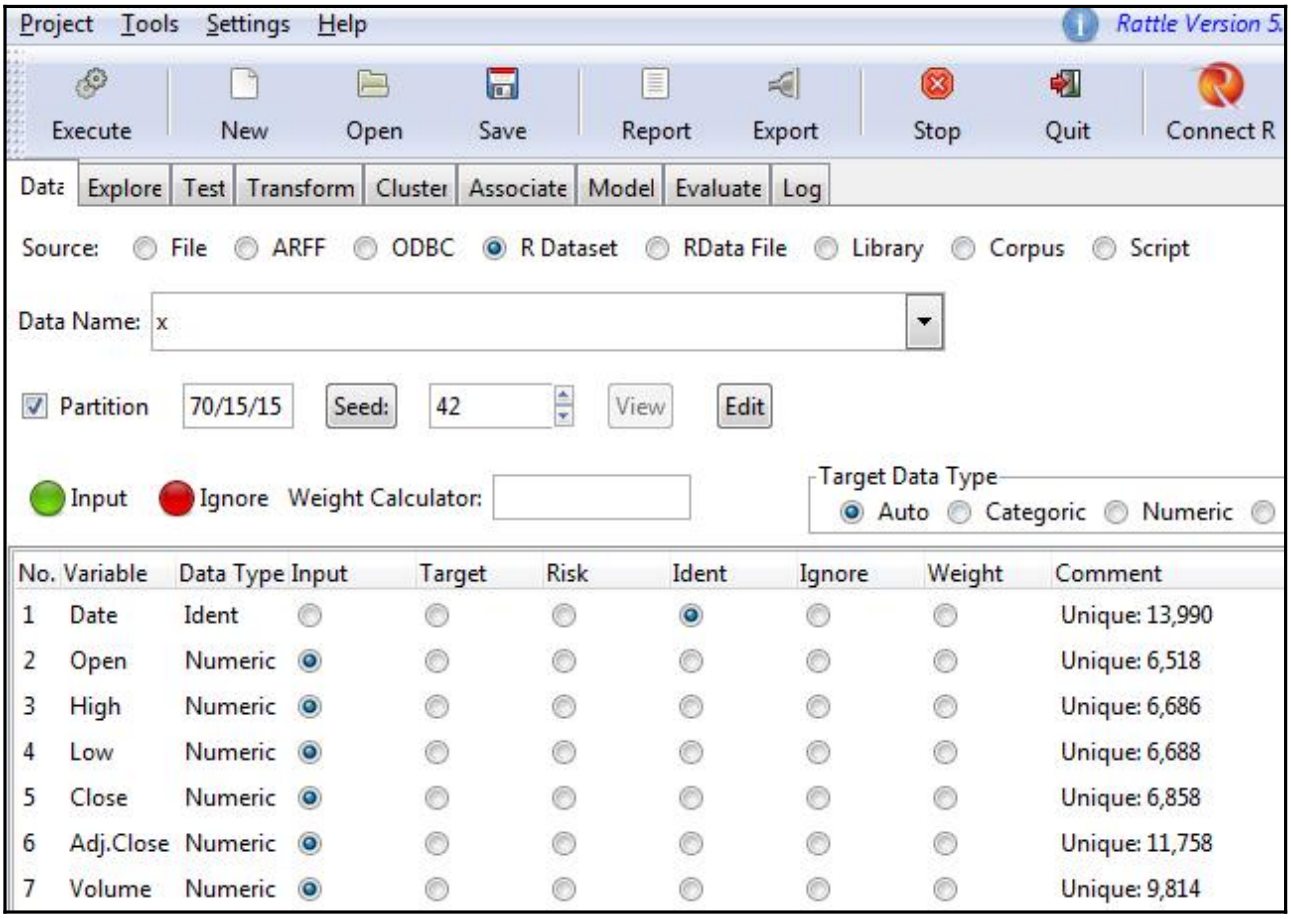
Fama/French 3 Factors TXT CSV Details
Fama/French 3 Factors [Weekly] TXT CSV Details
Fama/French 3 Factors [Daily] TXT CSV Details

```
> summary(lm(y~x))

Call:
lm(formula = y ~ x)

Residuals:
      Min        1Q    Median        3Q       Max
-0.268710 -0.032789 -0.001212  0.029722  0.282335

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.005664   0.002216   2.557 0.010794 *
xMKT_RF      0.888114   0.052521  16.910  < 2e-16 ***
xSMB        -0.179172   0.074995  -2.389 0.017168 *
xHML        -0.278756   0.081042  -3.440 0.000619 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05589 on 660 degrees of freedom
Multiple R-squared:  0.3469,    Adjusted R-squared:  0.344
F-statistic: 116.9 on 3 and 660 DF,  p-value: < 2.2e-16
```


8
6
4
2
0
-2
-1
0
1
2
3
4

Chapter 06: Managing Packages

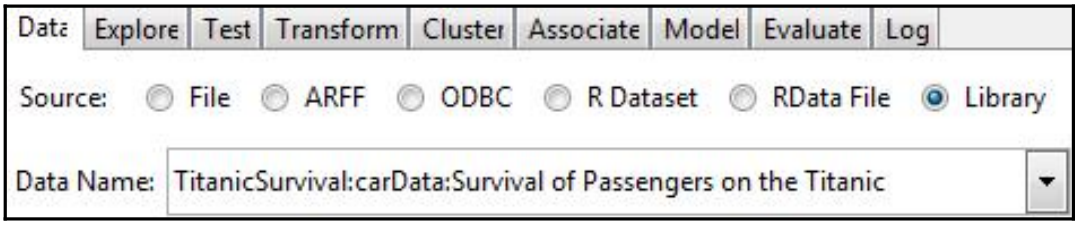


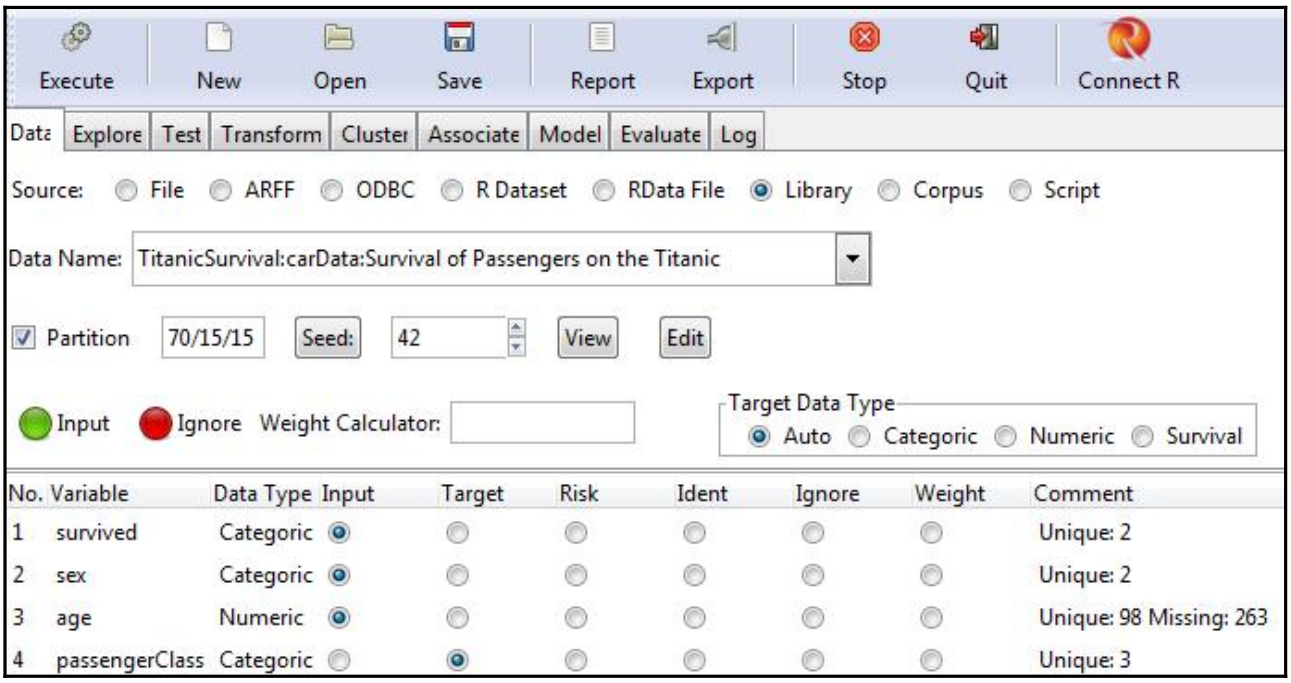

Project Tools Settings Help
Rattle Version 5.
Execute New Open Save Report Export Stop Quit Connect R
Data Explore Test Transform Cluster Associate Model Evaluate Log
Source: File ARFF ODBC R Dataset RData File Library Corpus Script
Data Name: x
Partition 70/15/15 Seed: 42 View Edit
Input Ignore Weight Calculator:
Target Data Type
Auto Categoric Numeric
No. Variable Data Type Input Target Risk Ident Ignore Weight Comment
1 Date Ident Unique: 13,990
2 Open Numeric Unique: 6,518
3 High Numeric Unique: 6,686
4 Low Numeric Unique: 6,688
5 Close Numeric Unique: 6,858
6 Adj.Close Numeric Unique: 11,758
7 Volume Numeric Unique: 9,814

Type: Summary Distributions Correlation Principal Components Interactive

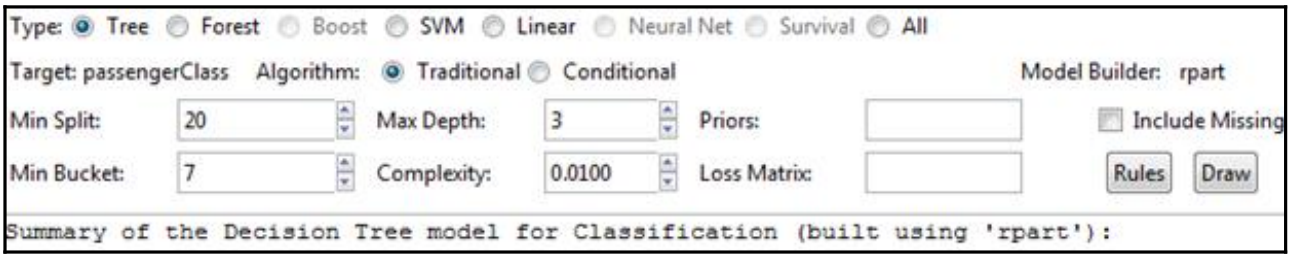
Summary Describe Basics Kurtosis Skewness Show Missing Cross Tab

```
The data is limited to the training dataset.


Data frame:crs$dataset[crs$sample, c(crs$input, crs$risk, crs$target)]  9793 observations and 6 variables    Maximum # NAs:0


          Storage
Open       double
High       double
Low        double
Close      double
Adj.Close  double
Volume    integer

For the simple distribution tables below the 1st and 3rd Qu.
refer to the first and third quartiles, indicating that 25%
of the observations have values of that variable which are
less than or greater than (respectively) the value listed.

      Open             High              Low              Close
 Min.   :  4.08   Min.   :  4.287   Min.   :  4.00   Min.   :  4.08
 1st Qu.: 15.84   1st Qu.: 15.963   1st Qu.: 15.73   1st Qu.: 15.84
 Median : 27.56   Median : 27.812   Median : 27.34   Median : 27.56
 Mean   : 56.43   Mean   : 56.961   Mean   : 55.93   Mean   : 56.45
 3rd Qu.: 92.22   3rd Qu.: 93.000   3rd Qu.: 91.25   3rd Qu.: 92.27
 Max.   :215.38   Max.   :215.900   Max.   :214.30   Max.   :215.80
   Adj.Close           Volume
 Min.   :  1.189   Min.   :       0
 1st Qu.:  5.797   1st Qu.: 1214400
 Median : 15.834   Median : 4127800
 Mean   : 43.126   Mean   : 4866490
 3rd Qu.: 71.146   3rd Qu.: 6947600
 Max.   :190.226   Max.   :69444700

Rattle timestamp: 2018-02-15 15:41:35 yany
```

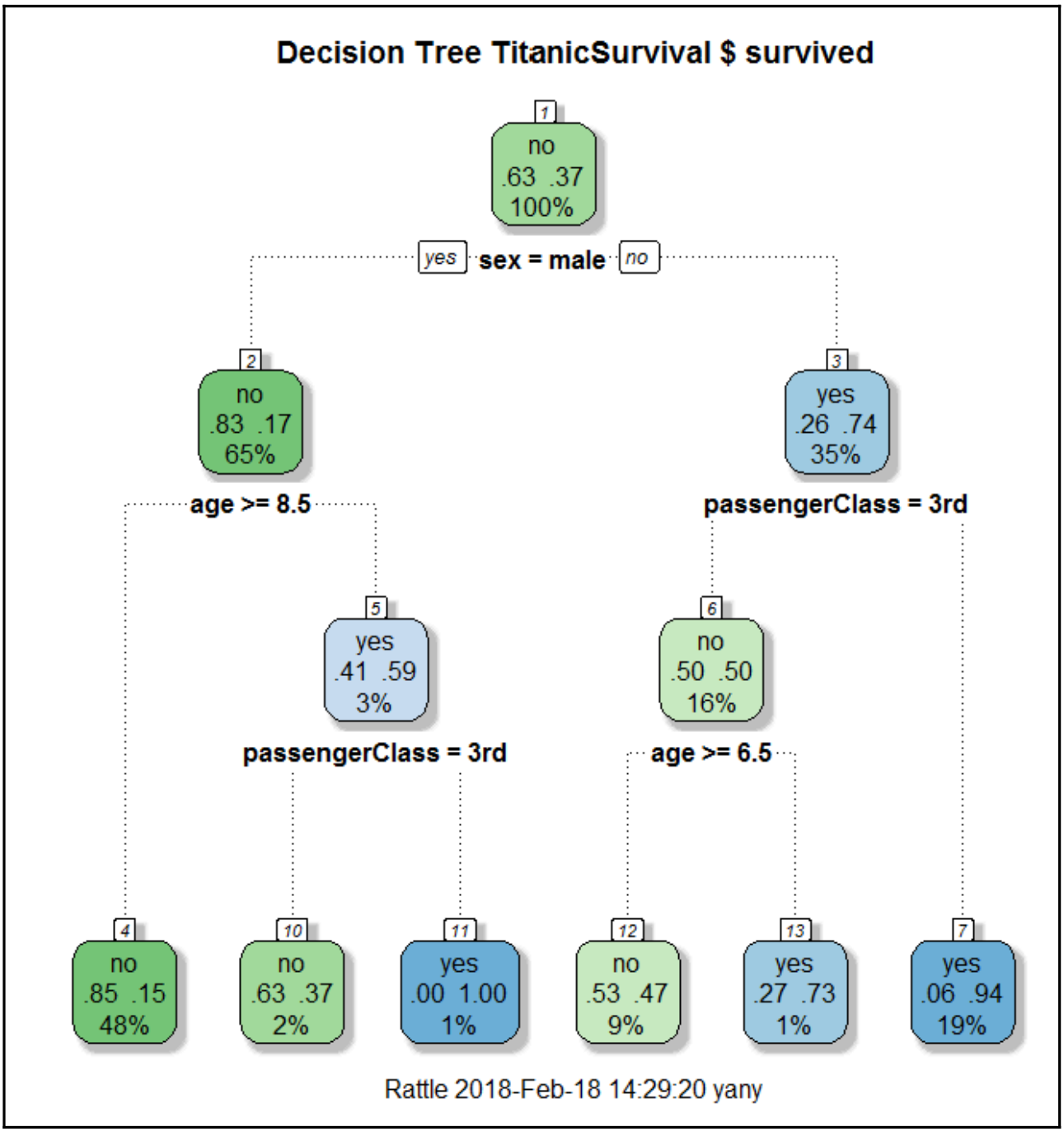



Execute
New
Open
Save
Report
Export
Stop
Quit
Connect R
Data
Explore
Test
Transform
Cluster
Associate
Model
Evaluate
Log
Source: File ARFF ODBC R Dataset RData File Library Corpus Script
Data Name: TitanicSurvival:carData:Survival of Passengers on the Titanic
Partition 70/15/15 Seed: 42 View Edit
Input Ignore Weight Calculator:
Target Data Type
Auto Categoric Numeric Survival
No. Variable Data Type Input Target Risk Ident Ignore Weight Comment
1 survived Categoric Unique: 2
2 sex Categoric Unique: 2
3 age Numeric Unique: 98 Missing: 263
4 passengerClass Categoric Unique: 3


No. Variable Data Type Input Target Risk Ident Ignore Weight Comment
1 survived Categoric Unique: 2
2 sex Categoric Unique: 2
3 age Numeric Unique: 98 Missing: 263
4 passengerClass Categoric Unique: 3


Type: Tree Forest Boost SVM Linear Neural Net Survival All
Target: passengerClass Algorithm: Traditional Conditional
Model Builder: rpart
Min Split: 20
Max Depth: 3
Priors:
Include Missing
Min Bucket: 7
Complexity: 0.0100
Loss Matrix:
Rules
Draw
Summary of the Decision Tree model for Classification (built using 'rpart'):

```
Summary of the Decision Tree model for Classification (built using 'rpart'):

n= 916

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 916 337 no (0.63209607 0.36790393)
   2) sex=male 598 102 no (0.82943144 0.17056856)
     4) age>=8.5 439  67 no (0.84738041 0.15261959) *
     5) age< 8.5 29  12 yes (0.41379310 0.58620690)
      10) passengerClass=3rd 19   7 no (0.63157895 0.36842105) *
      11) passengerClass=1st,2nd 10   0 yes (0.00000000 1.00000000) *
   3) sex=female 318  83 yes (0.26100629 0.73899371)
     6) passengerClass=3rd 145  72 no (0.50344828 0.49655172)
      12) age>=6.5 86  40 no (0.53488372 0.46511628) *
      13) age< 6.5 11   3 yes (0.27272727 0.72727273) *
     7) passengerClass=1st,2nd 173  10 yes (0.05780347 0.94219653) *
```


Decision Tree TitanicSurvival $ survived
1
no
.63 .37
100%
yes
sex = male
no
2
no
.83 .17
65%
3
yes
.26 .74
35%
age >= 8.5
passengerClass = 3rd
5
yes
.41 .59
3%
6
no
.50 .50
16%
passengerClass = 3rd
age >= 6.5
4
no
.85 .15
48%
10
no
.63 .37
2%
11
yes
.00 1.00
1%
12
no
.53 .47
9%
13
yes
.27 .73
1%
7
yes
.06 .94
19%
Rattle 2018-Feb-18 14:29:20 yany

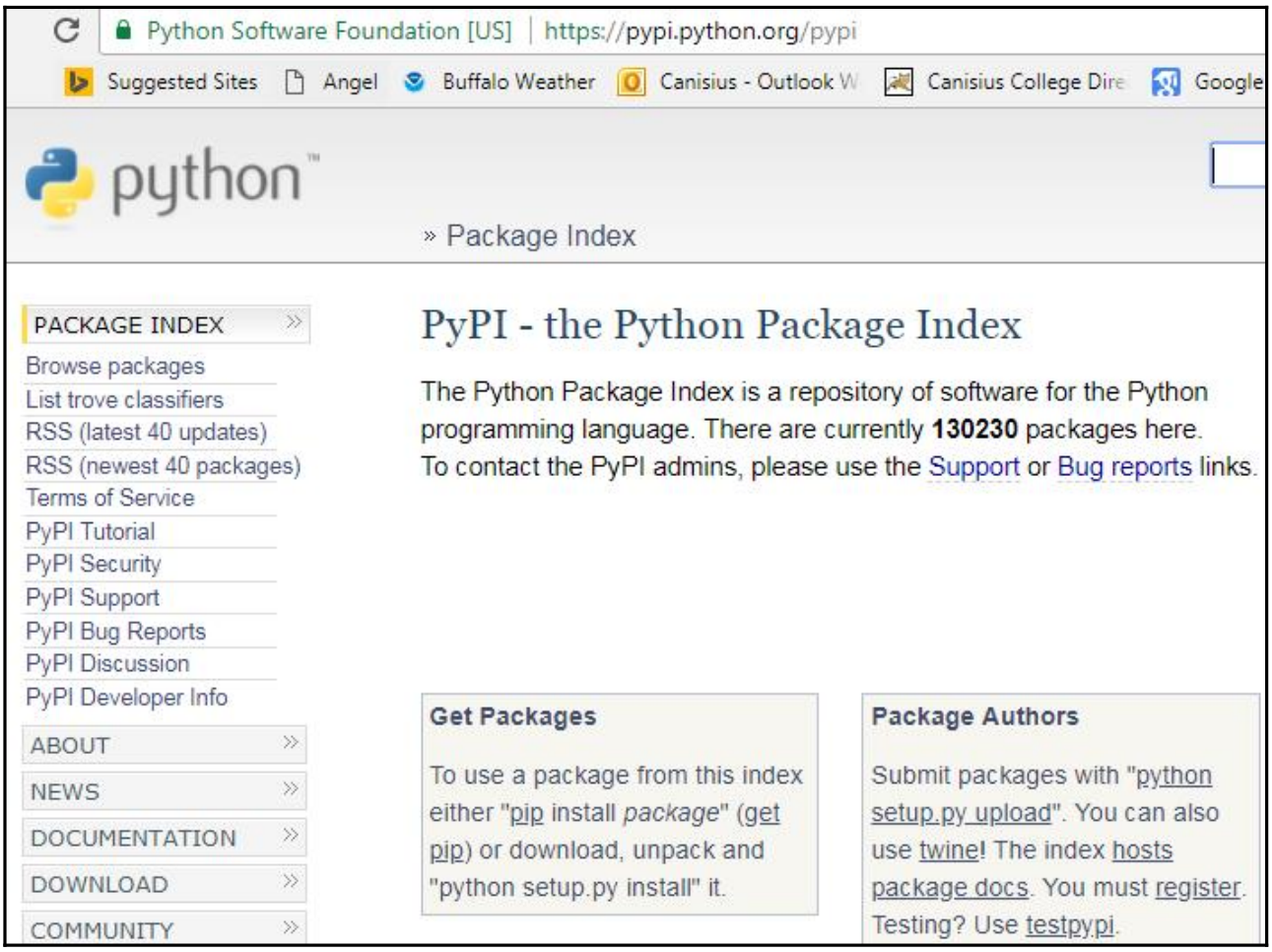
Available Packages

Currently, the CRAN package repository features 12173 available packages.

Table of available packages, sorted by date of publication

Table of available packages, sorted by name

Date	Package	Title
2018-02-22	anipaths	Animation of Observed Trajectories Using Spline-Based Interpolation
2018-02-22	BayesXsrc	R Package Distribution of the BayesX C++ Sources
2018-02-22	profvis	Interactive Visualizations for Profiling R Code
2018-02-22	suncalc	Compute Sun Position, Sunlight Phases, Moon Position and Lunar Phase
2018-02-22	wbsts	Multiple Change-Point Detection for Nonstationary Time Series


Python Software Foundation [US] | https://pypi.python.org/pypi
Suggested Sites
Angel
Buffalo Weather
Canisius - Outlook W
Canisius College Dire
Google
python
» Package Index
PACKAGE INDEX
Browse packages
List trove classifiers
RSS (latest 40 updates)
RSS (newest 40 packages)
Terms of Service
PyPI Tutorial
PyPI Security
PyPI Support
PyPI Bug Reports
PyPI Discussion
PyPI Developer Info
ABOUT
NEWS
DOCUMENTATION
DOWNLOAD
COMMUNITY
PyPI - the Python Package Index
The Python Package Index is a repository of software for the Python programming language. There are currently 130230 packages here.
To contact the PyPI admins, please use the Support or Bug reports links.
Get Packages
To use a package from this index either "pip install package" (get pip) or download, unpack and "python setup.py install" it.
Package Authors
Submit packages with "python setup.py upload". You can also use twine! The index hosts package docs. You must register. Testing? Use testpypi.

Index of Packages Matching 'data science'

Not Logged In

Login
Register
Lost Login?
Login with OpenID
Login with Google

Status

Spam on PyPI: monitoring

Package	Weight*	Description
jpl.mcl.site.sciencedata 0.0.5	17	Science data for the MCL site
data-science 0.1	16	Full Data Science lib
datascience 0.10.4	14	A Jupyter notebook Python library for introductory data science
dataone.scimeta 2.4.2	13	Utilities for processing Science Metadata
jr-data-science 0.1.1	13	Quelques functions pour la data science.
data-science-library 1.0-dev	12	Mnubo Data Science Library
datapack 0.0	12	Data Packages for Science
datascience_tools 0.0.3	12	Tools for data science
datasciencebox 0.3	12	Data Science Box
datafuzz 0.1.0a0	11	A data-science library built for testing cleaning, schema validation and model robustness. It messes up your data so you can test your data engineering and data science code (before it breaks in production).
datakit-core 0.3.0	11	A pluggable command-line tool for custom data science workflows.
data-pack 0.0	10	Data packages
data-utilities 1.2.9	10	A data analysis and visualization helper module.

Listing all 1725 registered packages for the Julia programming language.
Made with PackageEvaluator.jl.

Last updated 2018-02-21 — Package ecosystem pulse

Packages tested on Julia versions:
v0.5.2 (unmaintained release) — **v0.6.2 (current release)** — v0.7.0-DEV (unstable)

ACME

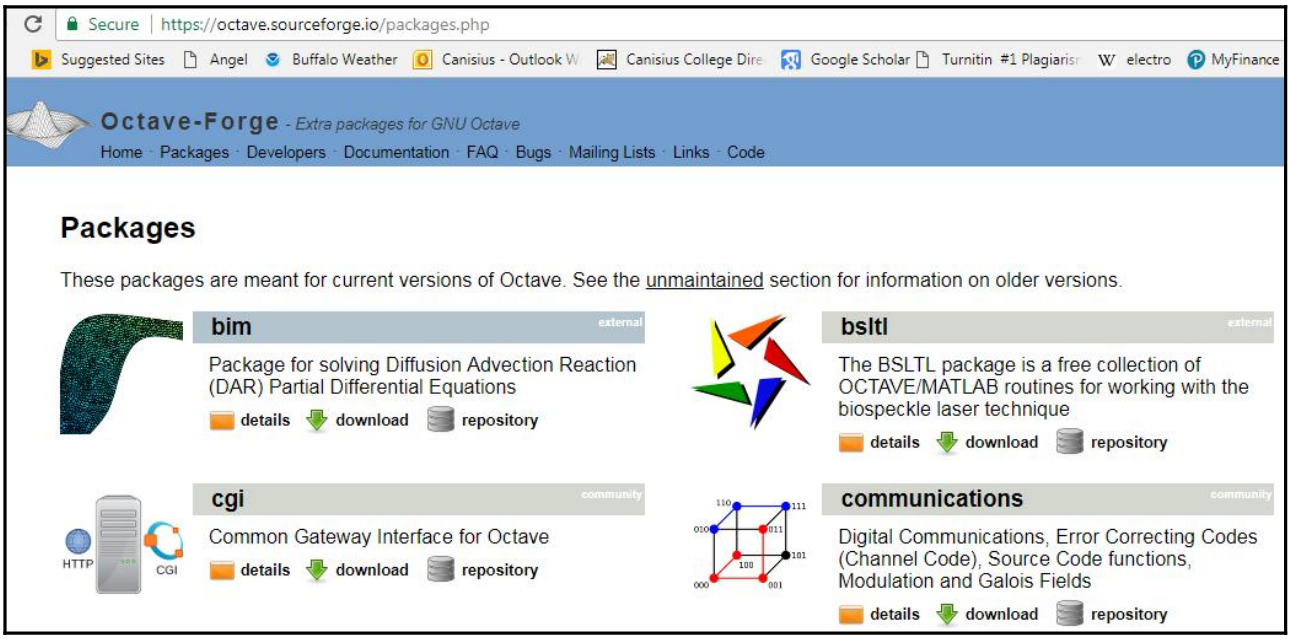
ACME.jl - Analog Circuit Modeling and Emulation for Julia

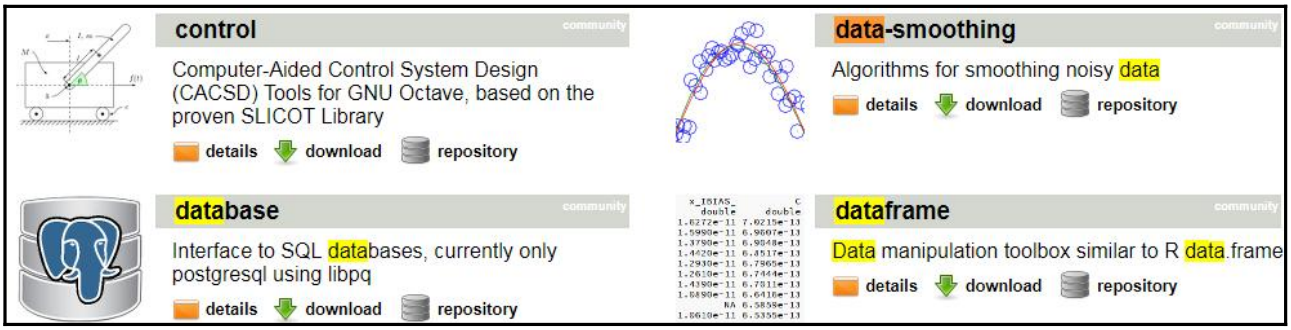
MIT license / Owner: HSU-ANT / permalink
26 ★ / 4 ↧ / 0 ↥
Julia v0.6: 0.6.2 (2 days ago) / ■ Tests fail.
Julia v0.7: 0.6.2 (2 days ago) / ■ Tests fail.

AbaqusReader

AbaqusReader.jl is a parse for ABAQUS FEM models. It's capable of parsing the geometry accurately, including surface sets, node sets, and other relevant geometrical data used in FEM calculations. Other option is to parse whole model, including boundary conditions, material data and load steps.

MIT license / Owner: JuliaFEM / permalink
3 ★ / 1 ↧ / 1 ↥
Julia v0.6: 0.1.0 (6 months ago) / Tests pass.





CRAN Task Views

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
WebTechnologies	Web Technologies and Services
gR	gRaphical Models in R

CRAN Task View: Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization

Maintainer: Nicholas Lewin-Koh
Contact: nikko at hailmail.net
Version: 2015-01-07
URL: https://CRAN.R-project.org/view=Graphics

R is rich with facilities for creating and developing interesting graphics. Base R contains functionality for many plot types including coplots, mosaic plots, biplots, and the list goes on. There are devices such as postscript, png, jpeg and pdf for outputting graphics as well as device drivers for all platforms running R. lattice and grid are supplied with R's recommended packages and are included in every binary distribution. lattice is an R implementation of William Cleveland's trellis graphics, while grid defines a much more flexible graphics environment than the base R graphics.

csv	10/25/2017 3:41 PM	File folder	
doc	10/25/2017 3:41 PM	File folder	
etc	10/25/2017 3:41 PM	File folder	
help	10/25/2017 3:41 PM	File folder	
html	10/25/2017 3:41 PM	File folder	
Meta	10/25/2017 3:41 PM	File folder	
odt	10/25/2017 3:41 PM	File folder	
po	10/25/2017 3:41 PM	File folder	
R	10/25/2017 3:41 PM	File folder	
CITATION	10/25/2017 3:41 PM	File	1 KB
DESCRIPTION	10/25/2017 3:41 PM	File	4 KB
INDEX	10/25/2017 3:41 PM	File	3 KB
MD5	10/25/2017 3:41 PM	File	3 KB
NAMESPACE	10/25/2017 3:41 PM	File	2 KB
NEWS	10/25/2017 3:41 PM	File	48 KB

Python Packaging User Guide

Welcome to the *Python Packaging User Guide*, a collection of tutorials and references to help you distribute and install Python packages with modern tools.

This guide is maintained on GitHub by the Python Packaging Authority. We happily accept any contributions and feedback. ☺

Note: Looking for guidance on migrating from legacy PyPI to pypi.org? Please see Migrating to PyPI.org.

statistics

community

Package Version:	1.3.0
Last Release Date:	2016-10-09
Package Author:	various authors
Package Maintainer:	Arno Onken <asnelt@asnelt.org>
License:	GPLv3+, public domain

Download Package

Repository

Older versions

Function Reference

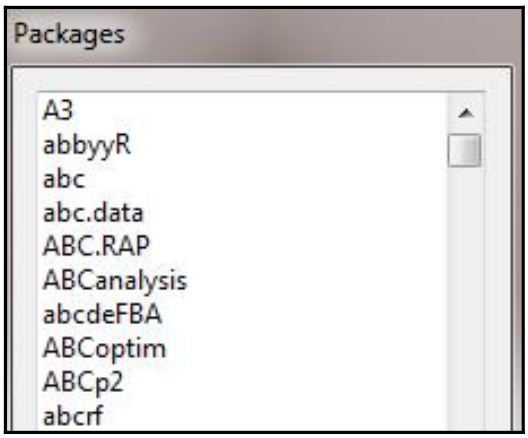
NEWS

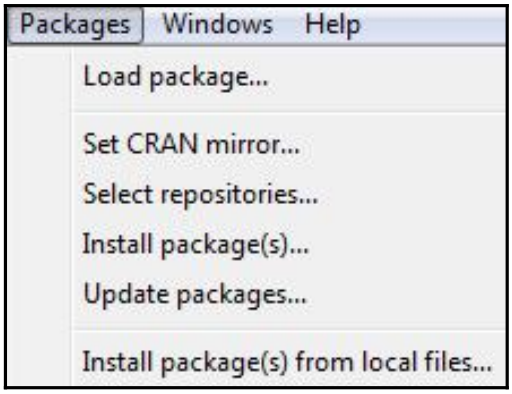
Description

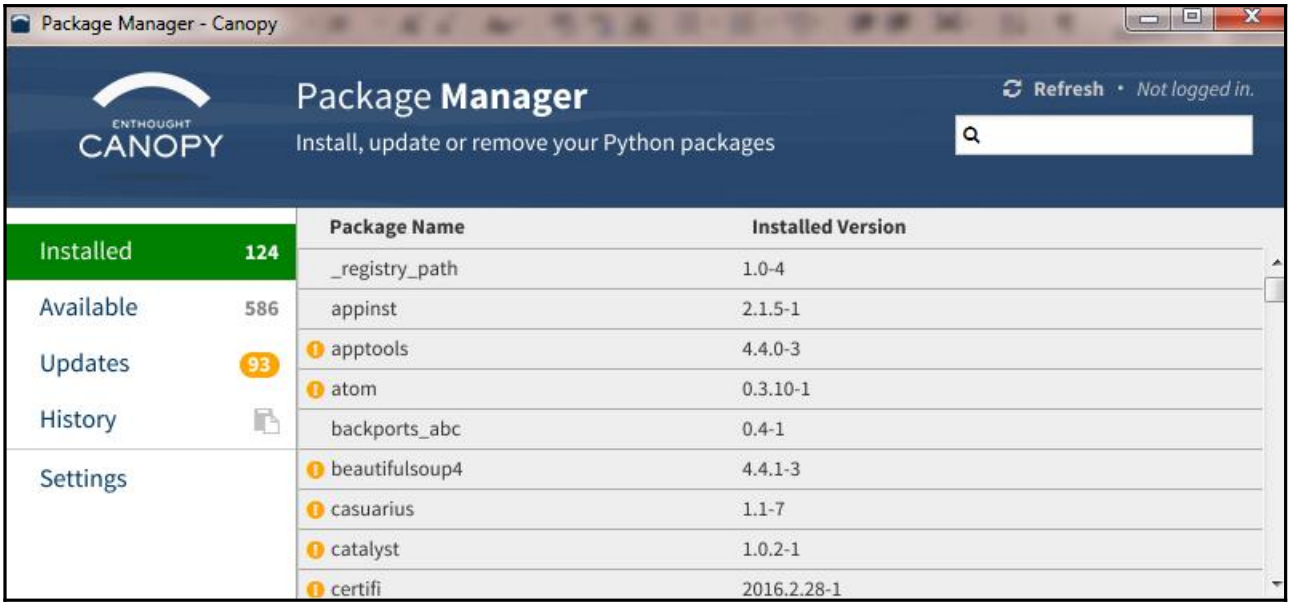
Additional statistics functions for Octave.

Details

Dependencies: Octave >= 4.0.0 io >= 1.0.18


Packages
A3
abbyyR
abc
abc.data
ABC.RAP
ABCanalysis
abcdeFBA
ABCoptim
ABCp2
abcrf


Packages
Windows
Help
Load package...
Set CRAN mirror...
Select repositories...
Install package(s)...
Update packages...
Install package(s) from local files...


Package Manager - Canopy
ENTHOUGHT
CANOPY
Package Manager
Install, update or remove your Python packages
Refresh
Not logged in.
Installed 124
Available 586
Updates 93
History
Settings
Package Name
Installed Version
_registry_path 1.0-4
appinst 2.1.5-1
apptools 4.4.0-3
atom 0.3.10-1
backports_abc 0.4-1
beautifulsoup4 4.4.1-3
casuarius 1.1-7
catalyst 1.0.2-1
certifi 2016.2.28-1

```
In [4]: import matplotlib as mat

In [5]: x=dir(mat)

In [6]: print(x)
['RcParams', 'URL_REGEX', 'Verbose', '_DATA_DOC_APPENDIX', '__builtins__', '__doc__', '__file__', '__name__',
'__package__', '__path__', '__version__', '__version__numpy__', '_all_deprecated', '_cm', '_cm_listed',
'_cntr', '_contour', '_create_tmp_config_dir', '_decode_filesystem_path', '_deprecated_ignore_map',
'_deprecated_map', '_error_details_fmt', '_get_cachedir', '_get_config_or_cache_dir', '_get_configdir',
'_get_data_path', '_get_data_path_cached', '_get_home', '_get_xdg_cache_dir', '_get_xdg_config_dir',
'_image', '_is_writable_dir', '_mathtext_data', '_obsolete_set', '_open_file_or_url', '_path', '_png',
'_pylab_helpers', '_python26', '_qhull', '_rc_params_in_file', '_replacer', '_tri', '_url_lines',
'_use_error_msg', '_version', 'absolute_import', 'afm', 'artist', 'axes', 'axis', 'backend_bases',
'backend_tools', 'backends', 'bad_pyparsing', 'bezier', 'blocking_input', 'cbook', 'chain',
'checkdep_dvipng', 'checkdep_ghostscript', 'checkdep_inkscape', 'checkdep_pdftops', 'checkdep_ps_distiller',
'checkdep_tex', 'checkdep_usetex', 'checkdep_xmllint', 'cm', 'collections', 'colorbar', 'colors',
'compare_versions', 'compat', 'container', 'contextlib', 'contour', 'converter', 'cycler', 'dates',
'dateutil', 'dedent', 'default', 'defaultParams', 'default_test_modules', 'distutils', 'division',
'docstring', 'dviread', 'externals', 'f', 'figure', 'finance', 'font_manager', 'fontconfig_pattern',
'ft2font', 'functools', 'get_backend', 'get_cachedir', 'get_configdir', 'get_data_path', 'get_example_data',
'get_home', 'get_label', 'get_py2exe_datafiles', 'gridspec', 'image', 'inspect', 'interactive', 'io',
'is_interactive', 'is_string_like', 'is_url', 'key', 'legend', 'legend_handler', 'lines', 'locale', 'major',
'markers', 'mathtext', 'matplotlib_fname', 'minor1', 'minor2', 'mlab', 'mplDeprecation', 'numpy',
'offsetbox', 'os', 'patches', 'path', 'print_function', 'projections', 'pylab', 'pyparsing', 'pyplot',
'quiver', 'rc', 'rcParams', 'rcParamsDefault', 'rcParamsOrig', 'rc_context', 'rc_file', 'rc_file_defaults',
'rc_params', 'rc_params_from_file', 'rcdefaults', 'rcsetup', 're', 'reload', 's', 'scale', 'six', 'spines',
'stackplot', 'streamplot', 'style', 'subprocess', 'sys', 'table', 'tempfile', 'test', 'texmanager', 'text',
'textpath', 'ticker', 'tight_bbox', 'tk_window_focus', 'tmp', 'transforms', 'tri', 'unicode_literals',
'units', 'unpack_labeled_data', 'urlopen', 'use', 'validate_backend', 'verbose', 'verify_test_dependencies',
'warnings', 'widgets']
```

```
julia> Pkg.status()
11 required packages:
 - AbstractTables                0.0.1
 - Calculus                      0.2.2
 - Clustering                    0.9.1
 - DataFrames                    0.10.1
 - GLM                           0.8.1
 - Gadfly                        0.6.4
 - IJulia                        1.7.0
 - Plots                         0.15.0
 - PyPlot                        2.3.2
 - QuantEcon                     0.14.1
```

```
julia> Pkg.installed()
Dict{String,VersionNumber} with 91 entries:
  "CodecZlib"                => v"0.4.2"
  "MacroTools"               => v"0.4.0"
  "Reexport"                 => v"0.1.0"
  "CoupledFields"            => v"0.0.1"
  "DualNumbers"              => v"0.3.0"
  "Measures"                 => v"0.1.0"
  "Plots"                    => v"0.15.0"
  "NLopt"                    => v"0.3.6"
  "CommonSubexpressions"     => v"0.0.1"
  "KernelDensity"            => v"0.4.1"
  "IJulia"                   => v"1.7.0"
  "Compat"                   => v"0.53.0"
  "PlotThemes"               => v"0.2.0"
  "FixedPointNumbers"        => v"0.4.6"
  "DiffRules"                => v"0.0.3"
  "SpecialFunctions"         => v"0.3.8"
  "PositiveFactorizations"   => v"0.1.0"
  "StatsBase"                => v"0.19.5"
  "Primes"                   => v"0.2.0"
  "AbstractFFTs"             => v"0.2.1"
  ?                          => ?
```

statistics

community

Package Version:	1.3.0
Last Release Date:	2016-10-09
Package Author:	various authors
Package Maintainer:	Arno Onken <asnelt@asnelt.org>
License:	GPLv3+, public domain

Download Package

Repository

Older versions

Function Reference

NEWS

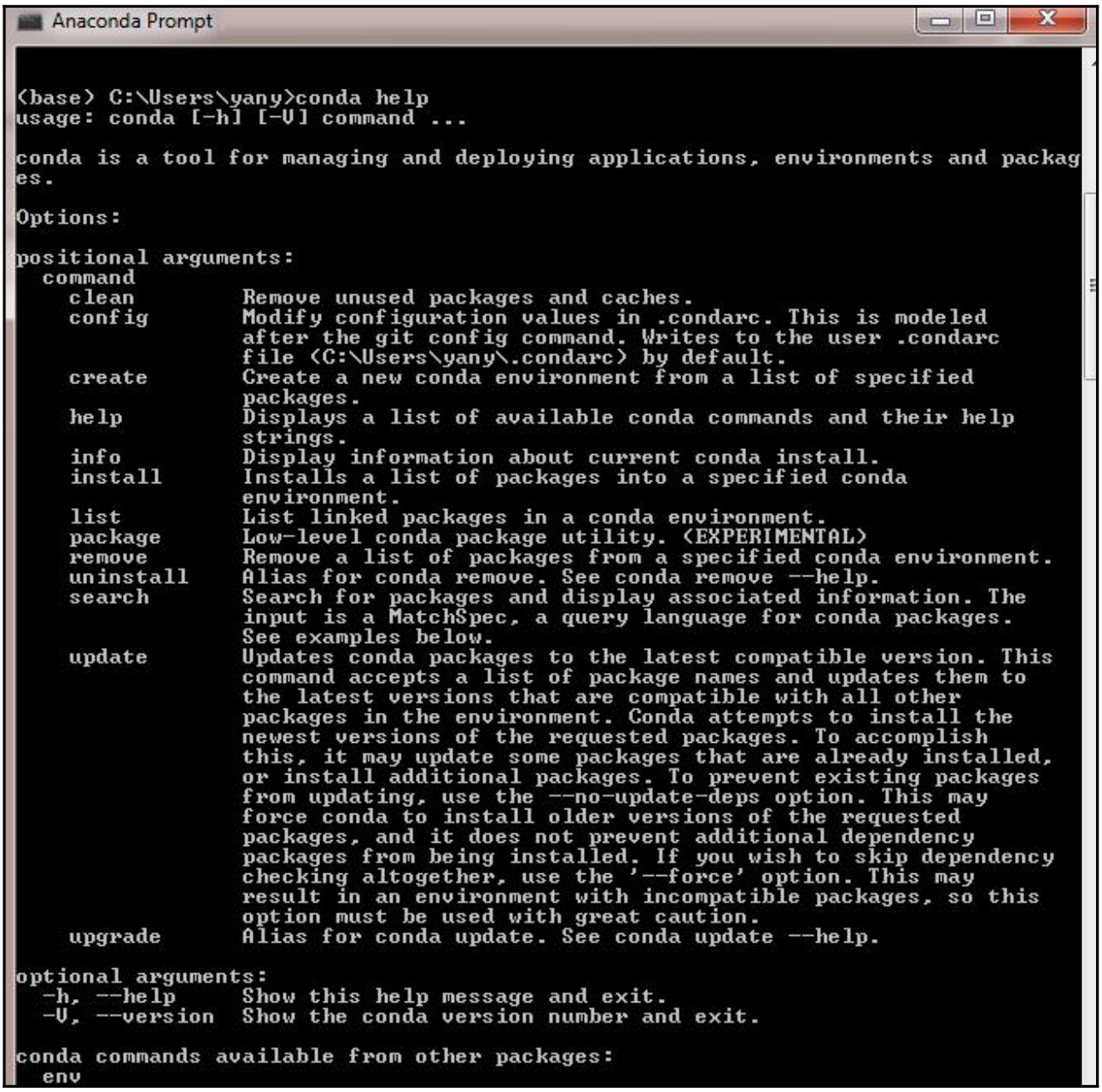
Description

Additional statistics functions for Octave.

Details

Dependencies: Octave >= 4.0.0 io >= 1.0.18

```
>> news statistics
Summary of important user-visible changes for statistics 1.3.0:
------------------------------------------------------------------

 ** The following functions are new:

      bbscdf bbsinv bbspdf bbsrnd
      binotest
      burrcdf burrinv burrpdf burrrnd
      gpcdf gpinv gppdf gprnd
      grp2idx
      mahal
      mvtpdf
      nakacdf nakainv nakapdf nakarnd
      pdf
      tricdf triinv tripdf trirnd
      violin
```


Anaconda Prompt
(base) C:\Users\yany>conda help
usage: conda [-h] [-V] command ...
conda is a tool for managing and deploying applications, environments and packag
es.
Options:
positional arguments:
command
clean Remove unused packages and caches.
config Modify configuration values in .condarc. This is modeled after the git config command. Writes to the user .condarc file (C:\Users\yany\.condarc) by default.
create Create a new conda environment from a list of specified packages.
help Displays a list of available conda commands and their help strings.
info Display information about current conda install.
install Installs a list of packages into a specified conda environment.
list List linked packages in a conda environment.
package Low-level conda package utility. (EXPERIMENTAL)
remove Remove a list of packages from a specified conda environment.
uninstall Alias for conda remove. See conda remove --help.
search Search for packages and display associated information. The input is a MatchSpec, a query language for conda packages. See examples below.
update Updates conda packages to the latest compatible version. This command accepts a list of package names and updates them to the latest versions that are compatible with all other packages in the environment. Conda attempts to install the newest versions of the requested packages. To accomplish this, it may update some packages that are already installed, or install additional packages. To prevent existing packages from updating, use the --no-update-deps option. This may force conda to install older versions of the requested packages, and it does not prevent additional dependency packages from being installed. If you wish to skip dependency checking altogether, use the '--force' option. This may result in an environment with incompatible packages, so this option must be used with great caution.
upgrade Alias for conda update. See conda update --help.
optional arguments:
-h, --help Show this help message and exit.
-V, --version Show the conda version number and exit.
conda commands available from other packages:
env

```
(C:\Users\yany\AppData\Local\Continuum\anaconda3) C:\Users\yany>conda install cb
sodata
Fetching package metadata .............

PackageNotFoundError: Packages missing in current channels:

  - cbsodata

We have searched for the packages in the following channels:

  - https://repo.continuum.io/pkgs/main/win-32
  - https://repo.continuum.io/pkgs/main/noarch
  - https://repo.continuum.io/pkgs/free/win-32
  - https://repo.continuum.io/pkgs/free/noarch
  - https://repo.continuum.io/pkgs/r/win-32
  - https://repo.continuum.io/pkgs/r/noarch
  - https://repo.continuum.io/pkgs/pro/win-32
  - https://repo.continuum.io/pkgs/pro/noarch
  - https://repo.continuum.io/pkgs/msys2/win-32
  - https://repo.continuum.io/pkgs/msys2/noarch
```

```
(base) C:\Users\yany>
(base) C:\Users\yany>conda install numpy
Solving environment: done


==> WARNING: A newer version of conda exists. <==
  current version: 4.4.7
  latest version: 4.4.10

Please update conda by running

    $ conda update -n base conda
```

```
(base) C:\Users\yany>conda search matplotlib
Loading channels: done
Name                       Version                 Build  Channel
matplotlib                 1.2.0              np16py27_0  defaults
matplotlib                 1.2.0              np16py27_1  defaults
matplotlib                 1.2.0              np17py27_1  defaults
matplotlib                 1.2.0              np17py33_1  defaults
matplotlib                 1.2.1              np17py27_0  defaults
matplotlib                 1.2.1              np17py27_1  defaults
matplotlib                 1.2.1              np17py33_0  defaults
matplotlib                 1.2.1              np17py33_1  defaults
```

Name	Date modified	Type	Size
myPackage.pyc	2/27/2018 2:07 PM	Compiled Python ...	19 KB
myPackage.py	2/27/2018 2:03 PM	PY File	21 KB

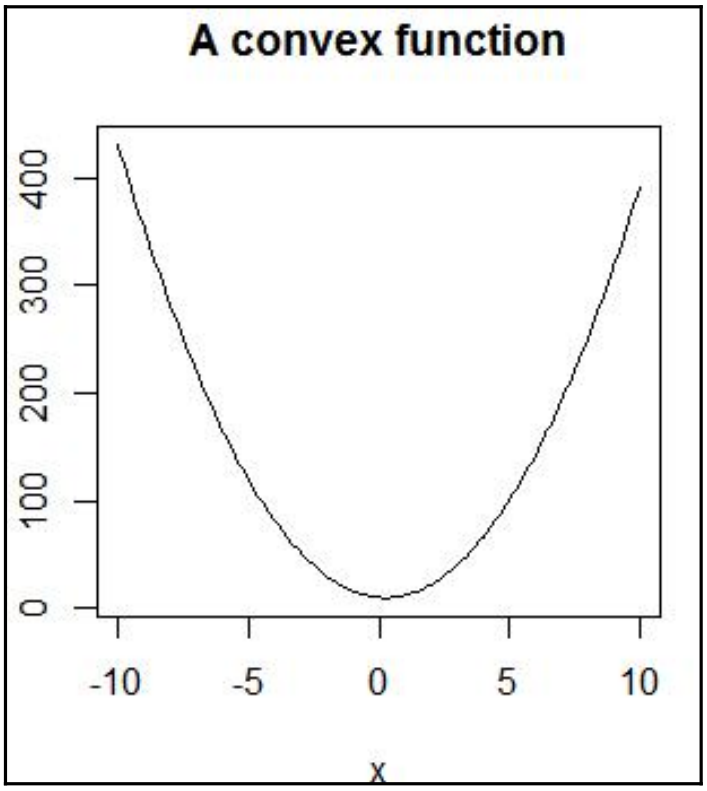
```
In [19]: sys.path.append("c:/temp/")

In [20]: import myPackage as my

In [21]: x=dir(my)

In [22]: print(x)
['CND', 'EAR_f', 'IRR_f', 'IRRs_f', 'NPER', 'PMT', 'Rc_f', 'Rm_f', '__builtins__', '__doc__',
'__file__', '__name__', '__package__', 'binomial_grid', 'bond_price', 'bs_call', 'bs_call_old',
'bs_put', 'delta_call', 'delta_put', 'durationBond', 'fv_annuity', 'fv_f', 'mean', 'n_annuity',
'npv_f', 'payback_', 'payback_period', 'pvValueNperiodModel', 'pv_annuity',
'pv_annuity_k_period_from_today', 'pv_excel', 'pv_f', 'pv_grow_perpetuity', 'pv_growing_annuity',
'pv_perpetuity', 'pv_perpetuity_due', 'r_continuous', 'sign']
```

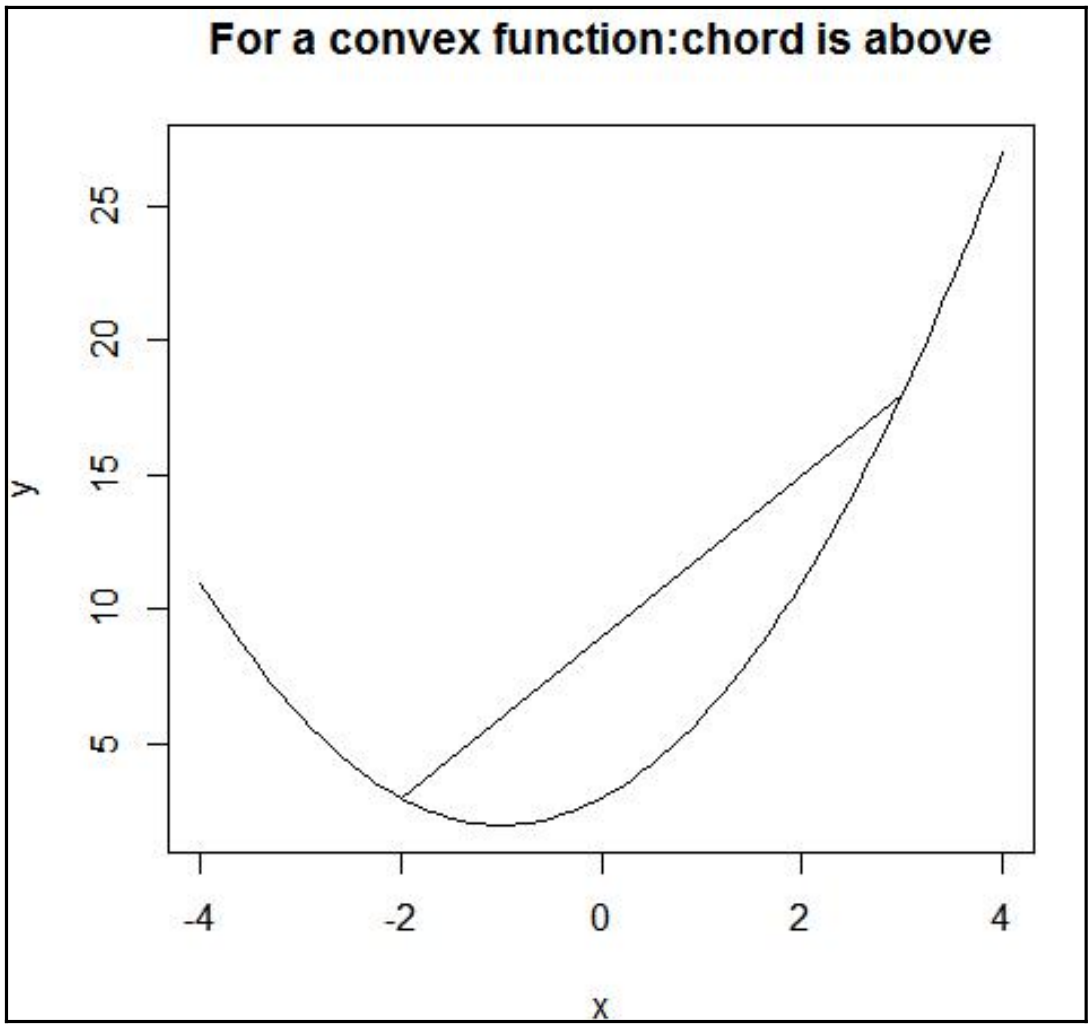
```
> Sys.getenv()
ALLUSERSPROFILE             C:\ProgramData
APPDATA                     C:\Users\yany\AppData\Roaming
CommonProgramFiles          C:\Program Files\Common Files
COMPUTERNAME                IC-22866
ComSpec                     C:\windows\system32\cmd.exe
CP_USER_NAME                yany
EMC_AUTOPLAY                C:\Program Files\Common Files\Roxio Shared\
f95include                  C:\Program Files\Silverfrost\FTN95\include
FP_NO_HOST_CHECK            NO
GFORTRAN_STDERR_UNIT        -1
GFORTRAN_STDOUT_UNIT        -1
HOME                        C:\Users\yany\Documents
```

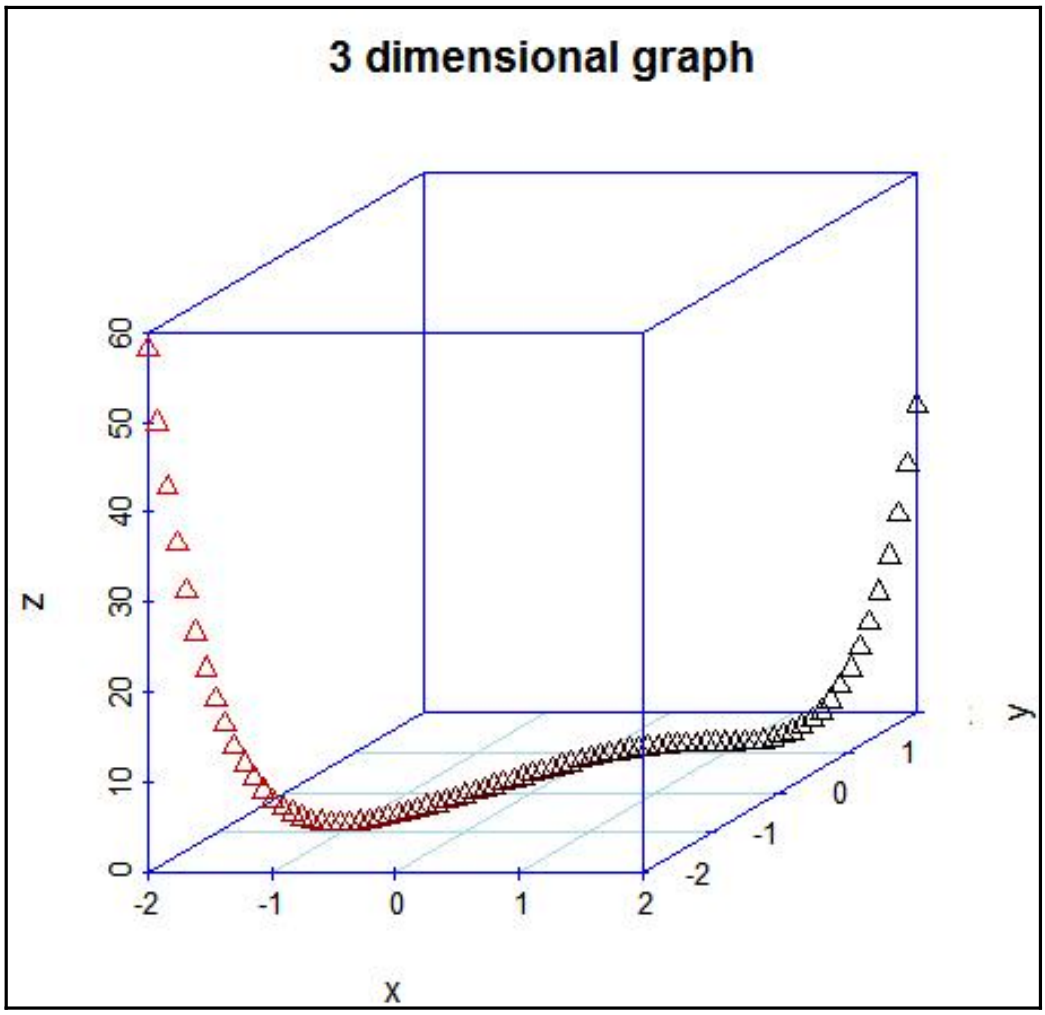
```
In [8]: sys.path
Out[8]:
['',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\Scripts\\python27.zip',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\DLLs',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\plat-win',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\lib-tk',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User',
```

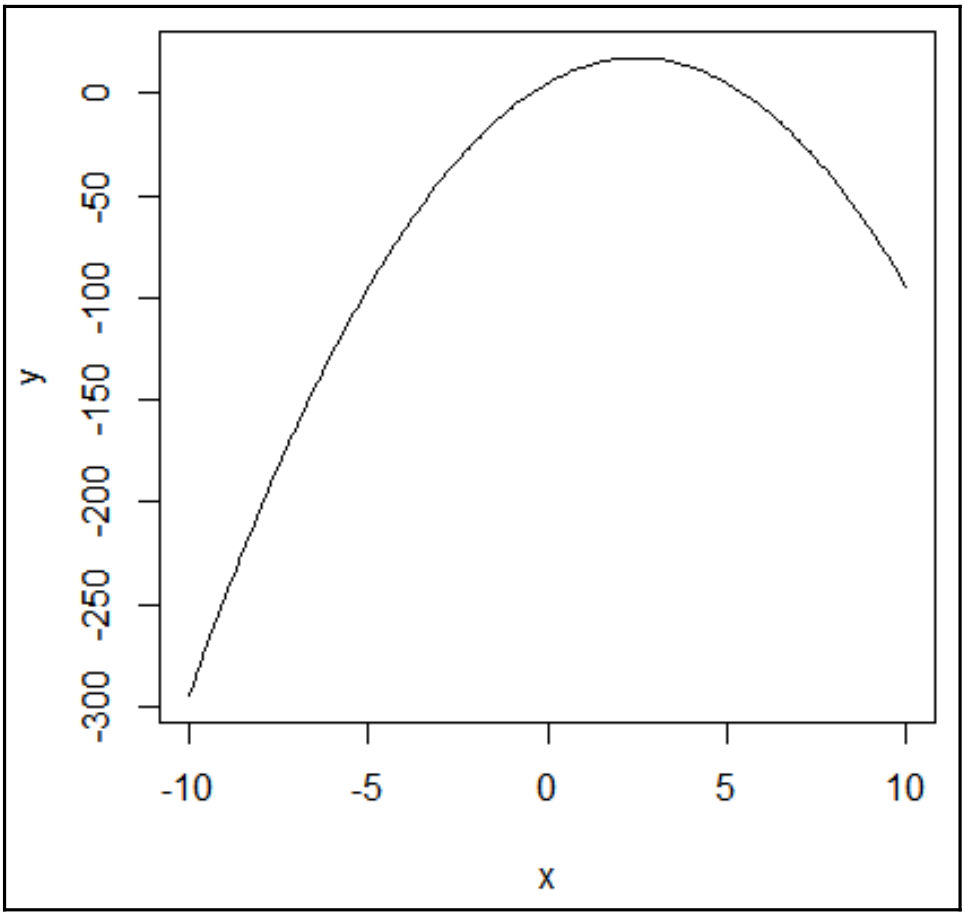
```
julia> ENV
Base.EnvHash with 47 entries:
  "ALLUSERSPROFILE"      => "C:\\ProgramData"
  "APPDATA"              => "C:\\Users\\yany\\AppData\\Roaming"
  "CommonProgramFiles"   => "C:\\Program Files\\Common Files"
  "COMPUTERNAME"         => "IC-22866"
  "ComSpec"              => "C:\\windows\\system32\\cmd.exe"
  "CP_USER_NAME"         => "yany"
  "EMC_AUTOPLAY"         => "C:\\Program Files\\Common Files\\Roxio Shared\\"
  "f95include"           => "C:\\Program Files\\Silverfrost\\FTN95\\include"
  "FP_NO_HOST_CHECK"     => "NO"
  "HOMEDRIVE"            => "H:"
  "HOMEPATH"             => "\\"
  "HOMESHARE"            => "\\\\filer\\faculty\\yany"
  "JULIA_EDITOR"         => "vim"
  "LOCALAPPDATA"         => "C:\\Users\\yany\\AppData\\Local"
  "LOGONSERVER"          => "\\\\AD-CANISIUS0"
  "mod_path"             => "C:\\Program Files\\Silverfrost\\FTN95\\include"
  "MOZ_PLUGIN_PATH"      => "C:\\Program Files\\Foxit Software\\Foxit Reader\\p.

  "NUMBER_OF_PROCESSORS" => "4"
  "OPENBLAS_MAIN_FREE"   => "1"
  "OS"                   => "Windows_NT"
  ?                      => ?
```

Chapter 07: Optimization in Anaconda




For a convex function:chord is above
y
5
10
15
20
25
-4
-2
0
2
4
x


3 dimensional graph
z
0
10
20
30
40
50
60
x
-2
-1
0
1
2
y
-2
-1
0
1




```
> head(.ff5industries,3)
    date VW_Cnsmr VW_Manuf VW_HiTec VW_Hlth VW_Other EW_Cnsmr EW_Manuf EW_HiTec EW_Hlth EW_Other
1 192607     5.43     2.73     1.83    1.77     2.16     1.69     1.42     2.01    2.45     0.21
2 192608     2.76     2.33     2.41    4.25     4.38     2.87     2.27     2.27    6.11     4.53
3 192609     2.16    -0.44     1.06    0.69     0.29    -0.90    -0.33     0.72    0.80    -0.53
> 
```

```
In [4]: import scipy as sp
   ...: x=dir(sp.optimize)
   ...: print(x)
   ...:
['LbfgsInvHessProduct', 'OptimizeResult', 'OptimizeWarning', 'Tester', '__all__', '__builtins__', '__doc__',
'__file__', '__name__', '__package__', '__path__', '_basinhopping', '_cobyla', '_differentialevolution',
'_group_columns', '_hungarian', '_lbfgsb', '_linprog', '_lsq', '_minimize', '_minpack', '_nnls', '_numdiff',
'_root', '_slsqp', '_spectral', '_trustregion', '_trustregion_dogleg', '_trustregion_ncg', '_zeros',
'absolute_import', 'anderson', 'approx_fprime', 'basinhopping', 'bench', 'bisect', 'bracket', 'brent',
'brenth', 'brentq', 'broyden1', 'broyden2', 'brute', 'check_grad', 'cobyla', 'curve_fit', 'diagbroyden',
'differential_evolution', 'division', 'excitingmixing', 'fixed_point', 'fmin', 'fmin_bfgs', 'fmin_cg',
'fmin_cobyla', 'fmin_l_bfgs_b', 'fmin_ncg', 'fmin_powell', 'fmin_slsqp', 'fmin_tnc', 'fminbound', 'fsolve',
'golden', 'lbfgsb', 'least_squares', 'leastsq', 'line_search', 'linear_sum_assignment', 'linearmixing',
'linesearch', 'linprog', 'linprog_verbose_callback', 'lsq_linear', 'minimize', 'minimize_scalar', 'minpack',
'minpack2', 'moduleTNC', 'newton', 'newton_krylov', 'nnls', 'nonlin', 'optimize', 'print_function', 'ridder',
'root', 'rosen', 'rosen_der', 'rosen_hess', 'rosen_hess_prod', 's', 'show_options', 'slsqp', 'test', 'tnc',
'zeros']
```

```
Help on function minimize in module scipy.optimize._minimize:

minimize(fun, x0, args=(), method=None, jac=None, hess=None, hessp=None,
bounds=None, constraints=(), tol=None, callback=None, options=None)
    Minimization of scalar function of one or more variables.

    In general, the optimization problems are of the form::

        minimize f(x) subject to

        g_i(x) >= 0,  i = 1,...,m
        h_j(x)  = 0,  j = 1,...,p

    where x is a vector of one or more variables.
    ``g_i(x)`` are the inequality constraints.
    ``h_j(x)`` are the equality constrains.

    Optionally, the lower and upper bounds for each element in x can also be
    specified using the `bounds` argument.

    Parameters
    ----------
```

```
Optimization terminated successfully.
         Current function value: 0.000000
         Iterations: 339
         Function evaluations: 571
[ 1.  1.  1.  1.  1.]
```

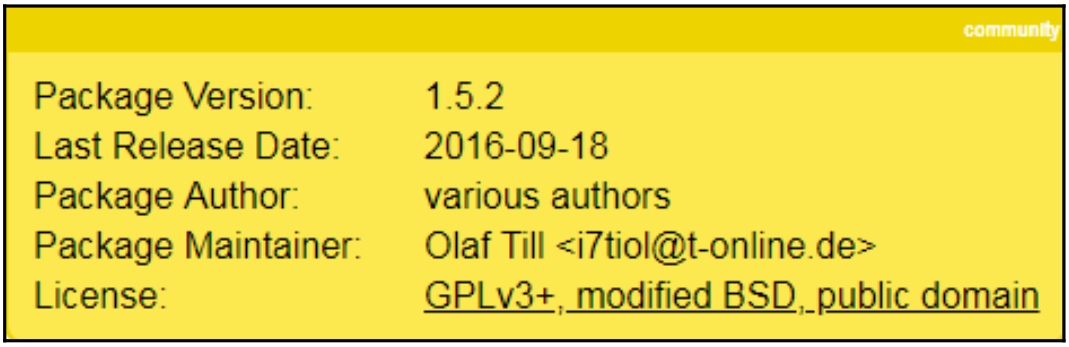
```
julia> print(m)
Max 5 x + 3 y
Subject to
 x + 5 y <= 3
 0 <= x <= 2
 0 <= y <= 30
```

```
julia> status = solve(m)

ECOS 2.0.2 - (C) embotech GmbH, Zurich Switzerland, 2012-15. Web: www.embotech.com/ECOS

It     pcost       dcost      gap   pres   dres    k/t    mu     step   sigma     IR    |   BT
 0  -1.252e+01  -2.013e+02  +5e+02  3e-01  3e-01  1e+00  8e+01    ---    ---    1  1  -  |  -  -
 1  -9.764e+00  -1.482e+01  +1e+01  1e-02  8e-03  2e-01  2e+00  0.9703  2e-03   0  0  0  |  0  0
 2  -1.080e+01  -1.123e+01  +1e+00  7e-04  7e-04  2e-02  2e-01  0.9005  4e-03   0  0  0  |  0  0
 3  -1.060e+01  -1.061e+01  +3e-02  2e-05  2e-05  1e-03  6e-03  0.9800  5e-03   0  0  0  |  0  0
 4  -1.060e+01  -1.060e+01  +4e-04  2e-07  2e-07  1e-05  7e-05  0.9890  1e-04   1  0  0  |  0  0
 5  -1.060e+01  -1.060e+01  +4e-06  2e-09  2e-09  1e-07  8e-07  0.9890  1e-04   1  0  0  |  0  0
 6  -1.060e+01  -1.060e+01  +5e-08  2e-11  2e-11  1e-09  8e-09  0.9890  1e-04   1  0  0  |  0  0

OPTIMAL (within feastol=2.1e-11, reltol=4.4e-09, abstol=4.7e-08).
Runtime: 0.045317 seconds.

:Optimal

julia> println("Objective value: ", getObjectiveValue(m))
Objective value: 10.600000000335104

julia> println("x = ", getValue(x))
x = 2.000000000720887

julia> println("y = ", getValue(y))
y = 0.19999999891022338
```

Solver	Julia Package	solver=	License	LP	SOCP	MILP	NLP	MINLP	SDP
Artelys Knitro	KNITRO.jl	KnitroSolver()	Comm.				X	X	
BARON	BARON.jl	BaronSolver()	Comm.				X	X	
Bonmin	AmplNLWriter.jl	BonminNLSolver() *	EPL	X		X	X	X	
CoinOptServices.jl	OsilBonminSolver()								
Cbc	Cbc.jl	CbcSolver()	EPL			X			
Clp	Clp.jl	ClpSolver()	EPL	X					
Couenne	AmplNLWriter.jl	CouenneNLSolver() *	EPL	X		X	X	X	
CoinOptServices.jl	OsilCouenneSolver()								
CPLEX	CPLEX.jl	CplexSolver()	Comm.	X	X	X			
ECOS	ECOS.jl	ECOSSolver()	GPL	X	X				
FICO Xpress	Xpress.jl	XpressSolver()	Comm.	X	X	X			
GLPK	GLPKMath...	GLPKSolver[LP|MIP]()	GPL	X		X			
Gurobi	Gurobi.jl	GurobiSolver()	Comm.	X	X	X			
Ipopt	Ipopt.jl	IpoptSolver()	EPL	X			X		
MOSEK	Mosek.jl	MosekSolver()	Comm.	X	X	X	X		X
NLopt	NLopt.jl	NLoptSolver()	LGPL				X		
SCS	SCS.jl	SCSSolver()	MIT	X	X				X

```
julia> using Optim

julia> f(x) = (1.0 - x[1])^2 + 200.0 * (x[2] - x[1]^2)^2
f (generic function with 1 method)

julia> optimize(f, [0.0, 0.0])
Results of Optimization Algorithm
 * Algorithm: Nelder-Mead
 * Starting Point: [0.0,0.0]
 * Minimizer: [0.9999855177658813,0.9999718307694201]
 * Minimum: 3.361490e-10
 * Iterations: 66
 * Convergence: true
   *  √(Σ(y?-y-)²)/n < 1.0e-08: true
   * Reached Maximum Number of Iterations: false
 * Objective Calls: 126
```

```
Results of Optimization Algorithm
 * Algorithm: Fminbox with Gradient Descent
 * Starting Point: [2.0,2.0]
 * Minimizer: [1.6557274658535348,2.74270195452600074]
 * Minimum: 4.303003e-01
 * Iterations: 3
 * Convergence: true
   * |x - x'| ≤ 1.0e-32: true
     |x - x'| = 0.00e+00
   * |f(x) - f(x')| ≤ 1.0e-32 |f(x)|: true
     |f(x) - f(x')| = 0.00e+00 |f(x)|
   * |g(x)| ≤ 1.0e-08: false
     |g(x)| = 4.64e-01
   * Stopped by an increasing objective: true
   * Reached Maximum Number of Iterations: false
 * Objective Calls: 322
 * Gradient Calls: 322
```

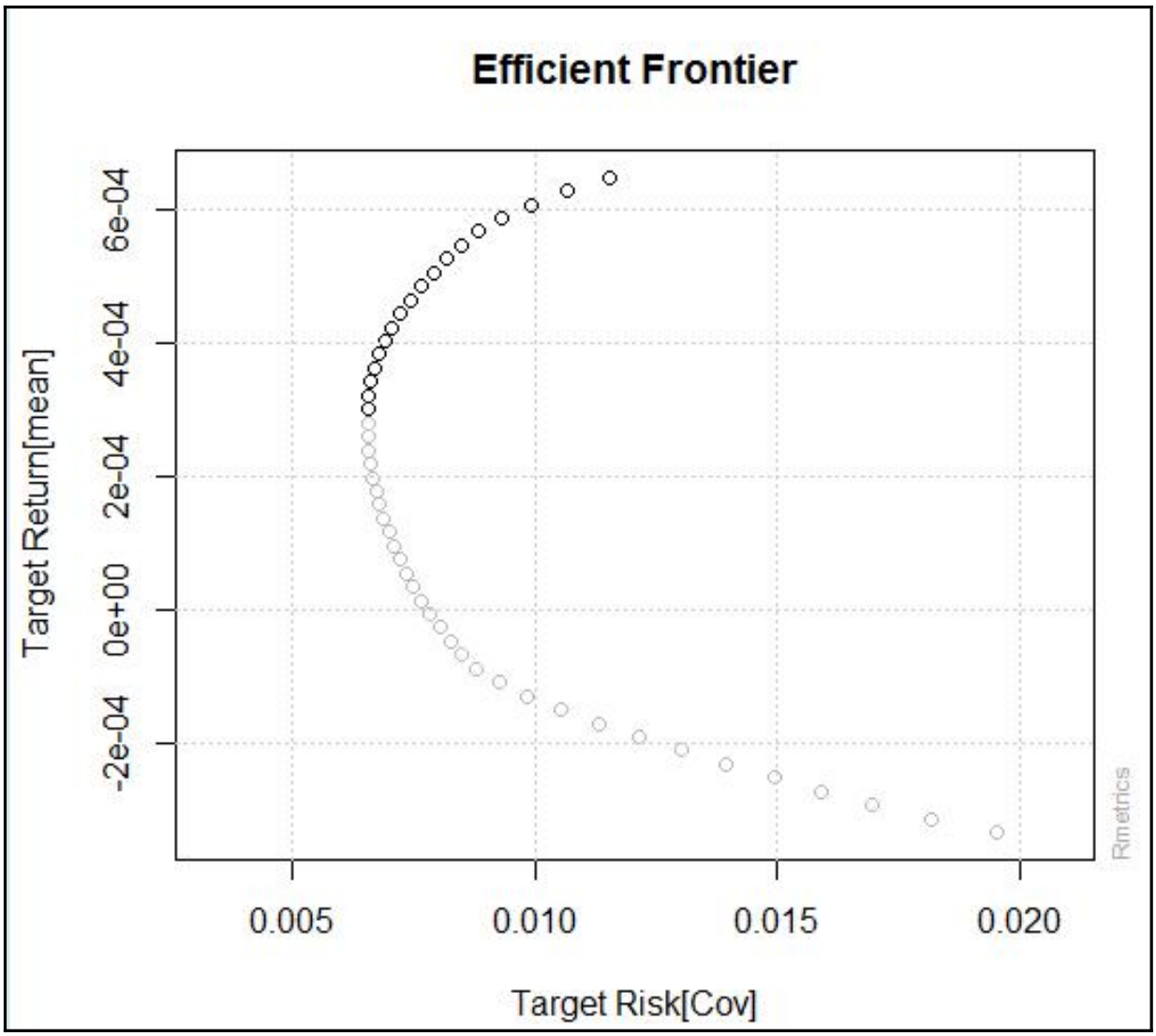


```
Package name:
        optim
Version:
        1.5.2
Short description:
        Non-linear optimization toolkit.
Status:
        Loaded
---
Provides:
Optimization
        nelder_mead_min
        nrm
        line_min
        powell
        fmins
        adsmax
        mdsmax
        nmsmax
        bfgsmin
        samin
        battery
        cg_min
        de_min
        nonlin_min
        brent_line_min
Data fitting
        lsqlin
        expfit
```

```
>> help fminsearch
'fminsearch' is a function from the file C:\Octave\Octave-4.0.0\share\oc

 -- Function File: X = fminsearch (FUN, X0)
 -- Function File: X = fminsearch (FUN, X0, OPTIONS)
 -- Function File: [X, FVAL] = fminsearch (...)
     Find a value of X which minimizes the function FUN.

     The search begins at the point X0 and iterates using the Nelder &
     Mead Simplex algorithm (a derivative-free method).  This algorithm
     is better-suited to functions which have discontinuities or for
     which a gradient-based search such as `fminunc' fails.

     Options for the search are provided in the parameter OPTIONS using
     the function `optimset'.  Currently, `fminsearch' accepts the
     options: "TolX", "MaxFunEvals", "MaxIter", "Display".  For a
     description of these options, see `optimset'.

     On exit, the function returns X, the minimum point, and FVAL, the
     function value thereof.

     Example usages:

          fminsearch (@(x) (x(1)-5).^2+(x(2)-8).^4, [0;0])

          fminsearch (inline ("(x(1)-5).^2+(x(2)-8).^4", "x"), [0;0])

     See also: fminbnd, fminunc, optimset.
```

```
>> [x,fval] = fminsearch(@fun2,x0,OPTIONS)
f(x0) = 1.8125e+000
Iter.  1,  how = initial    nf =   3,  f = 1.8125e+000  (0.0%)
Iter.  2,  how = contract,  nf =   5,  f = 1.8125e+000  (0.0%)
Iter.  3,  how = reflect,   nf =   6,  f = 1.8125e+000  (0.0%)
Iter.  4,  how = reflect,   nf =   8,  f = 4.7390e+000  (161.5%)
```

```
Iter. 33,  how = contract,  nf =  64,  f = 6.8781e+000  (0.0%)
Iter. 34,  how = contract,  nf =  66,  f = 6.8781e+000  (0.0%)
Simplex size 9.4335e-005 <= 1.0000e-004...quitting
x =

  -0.16954  -0.50866

fval = -6.8781
```



CRAN Task View: Optimization and Mathematical Programming

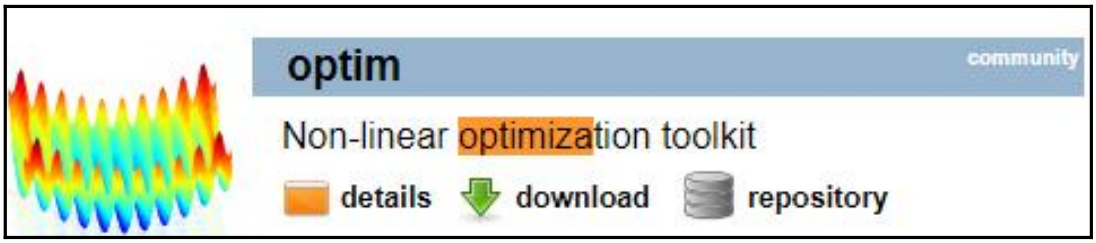
Maintainer: Stefan Theussl and Hans W. Borchers
Contact: R-optimization at mailbox.org
Version: 2018-03-02
URL: https://CRAN.R-project.org/view=Optimization

This CRAN task view contains a list of packages which offer facilities for solving optimization problems. Although every regression model in statistics solves an optimization problem they are not part of this view. If you are looking for regression methods, the following views will contain useful starting points: Multivariate, SocialSciences, Robust among others. The focus of this task view is on Optimization Infrastructure Packages , General Purpose Continuous Solvers , Mathematical Programming Solvers , and Specific Applications in Optimization . Packages are categorized in these four sections.

Many packages provide functionality for more than one of the subjects listed at the end of this task view. E.g., mixed integer linear programming solvers typically offer standard linear programming routines like the simplex algorithm. Therefore following each package description a list of abbreviations describes the typical features of the optimizer (i.e., the problems which can be solved). The full names of the abbreviations given in square brackets can be found at the end of this task view under Classification According to Subject .

If you think that some package is missing from the list, please let us know.

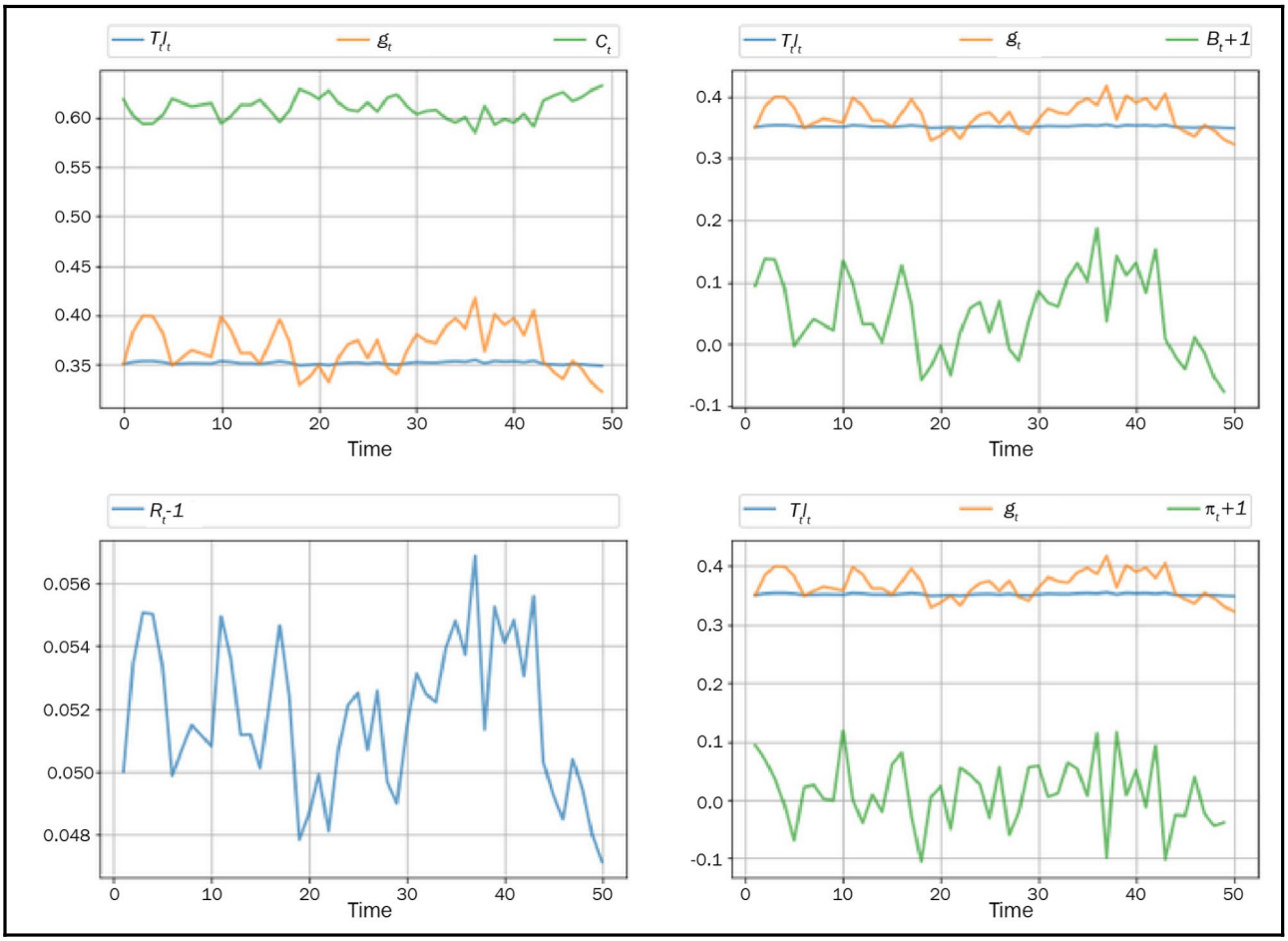
Optimization Infrastructure Packages



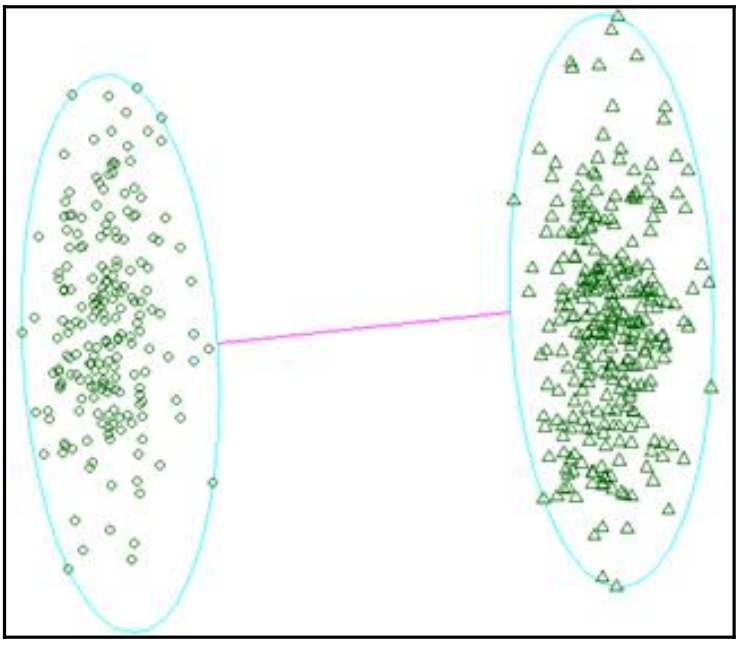
JuMP

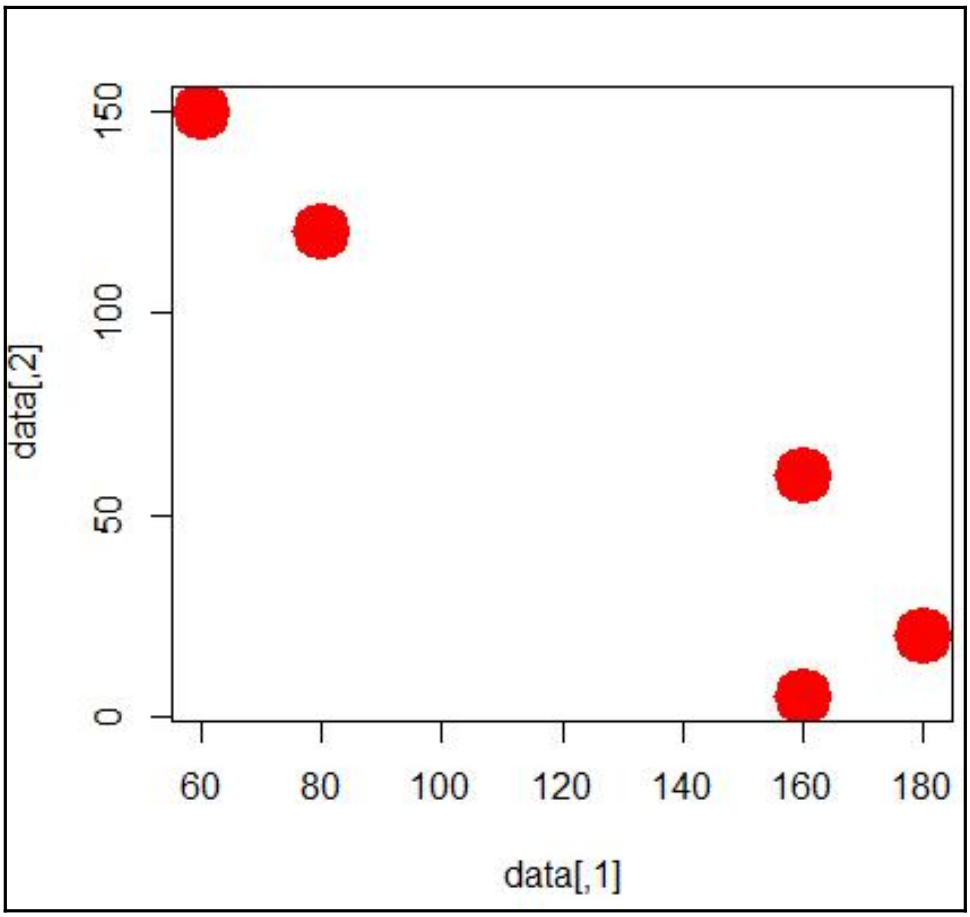
Modeling language for Mathematical Optimization (linear, mixed-integer, conic, semidefinite, nonlinear)

MPL v2 license / Owner: JuliaOpt / permalink
456 ★ / 15 ↧ / 40 ↥
Julia v0.6: 0.18.0 (7 months ago) / Tests pass.
Julia v0.7: 0.18.0 (7 months ago) / Tests fail.

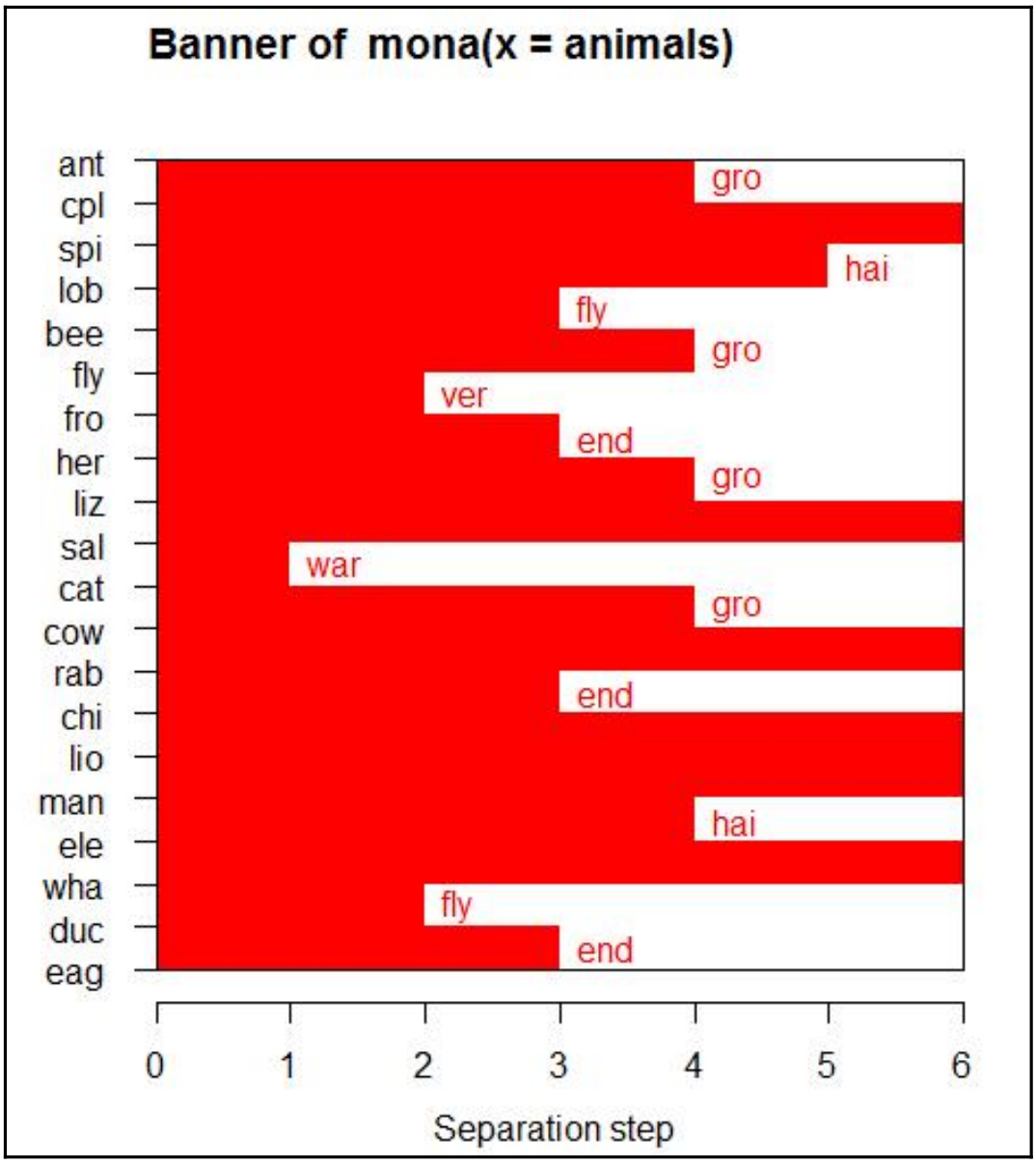


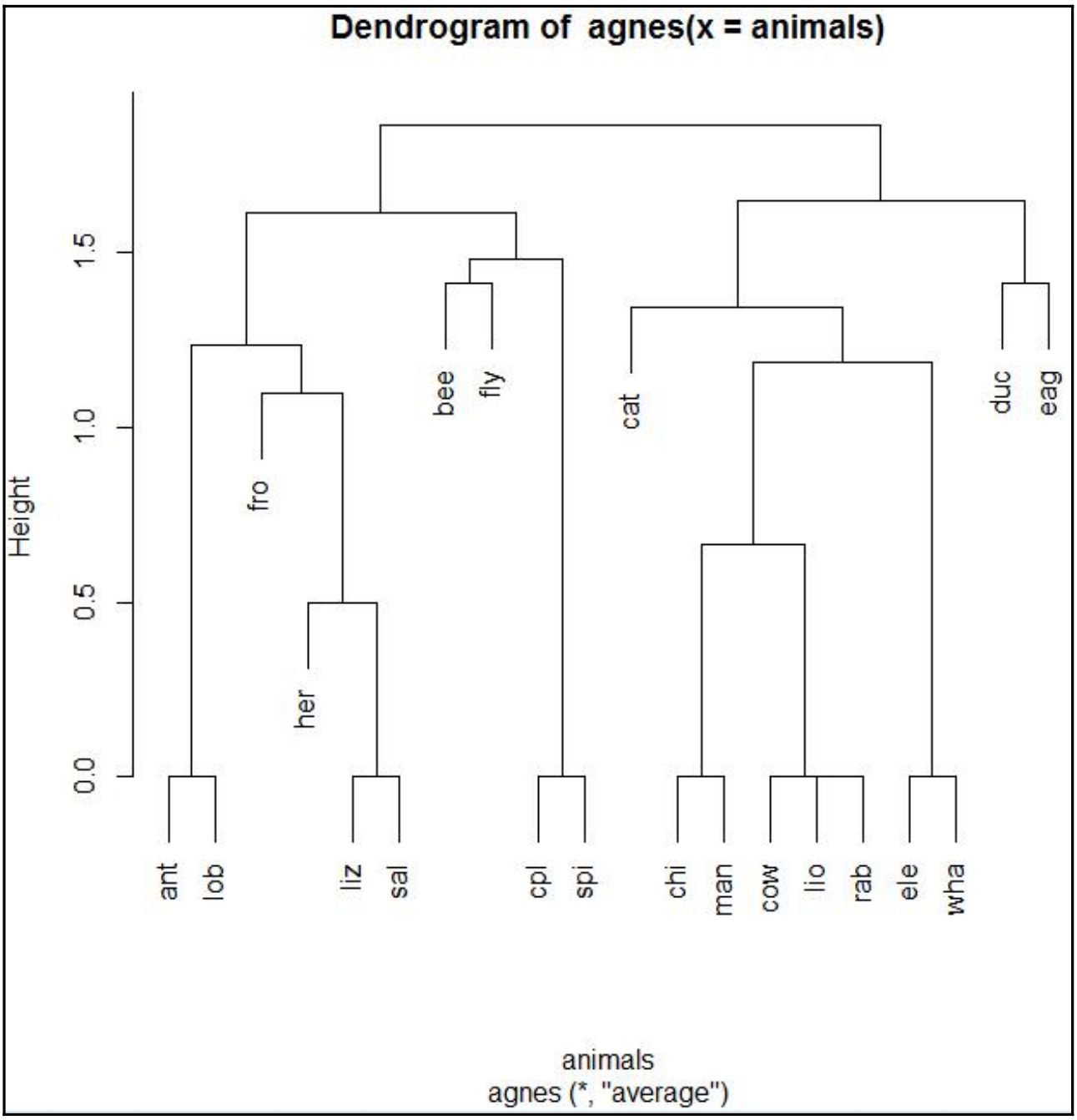
Chapter 08: Unsupervised Learning in Anaconda

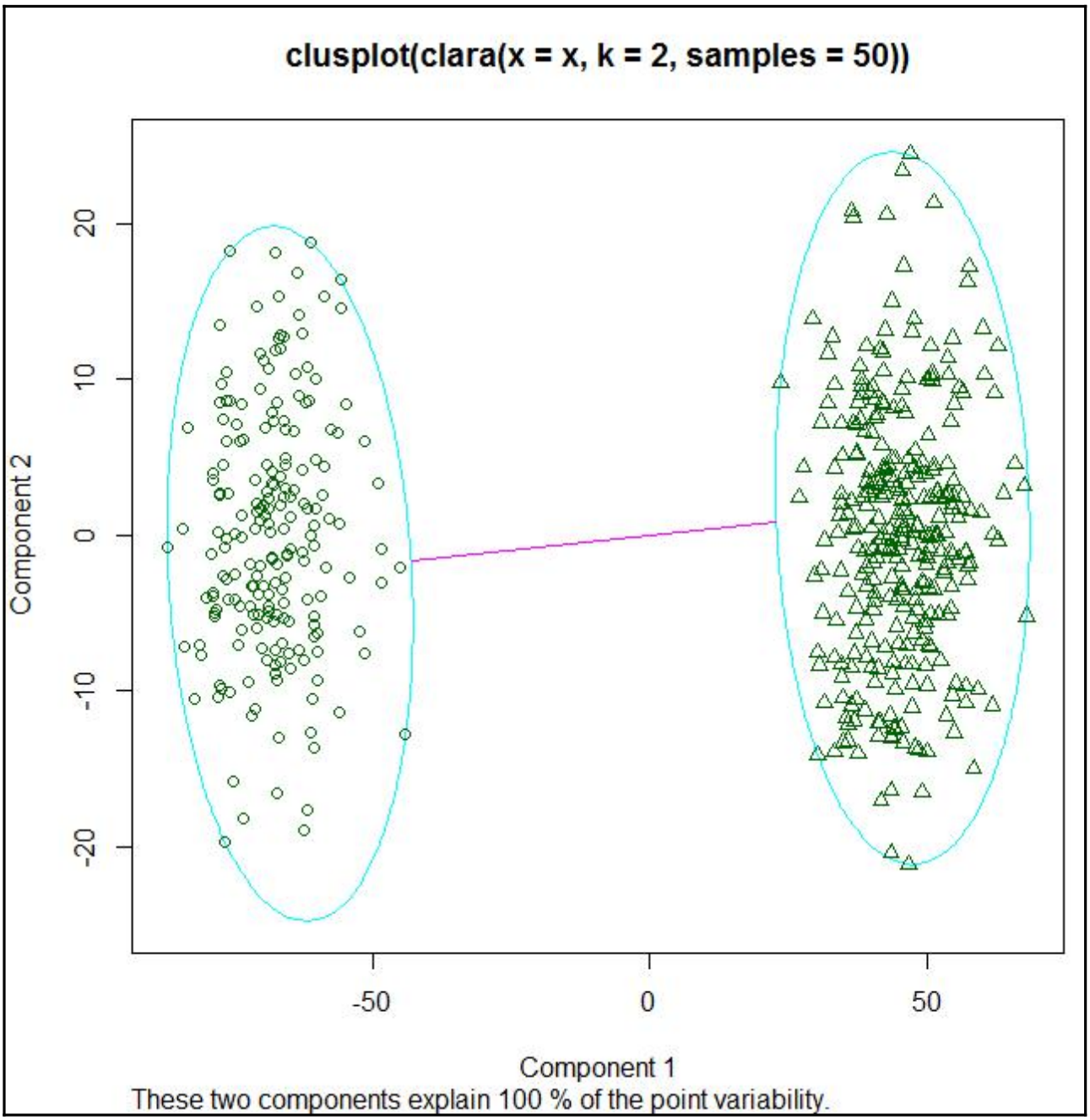


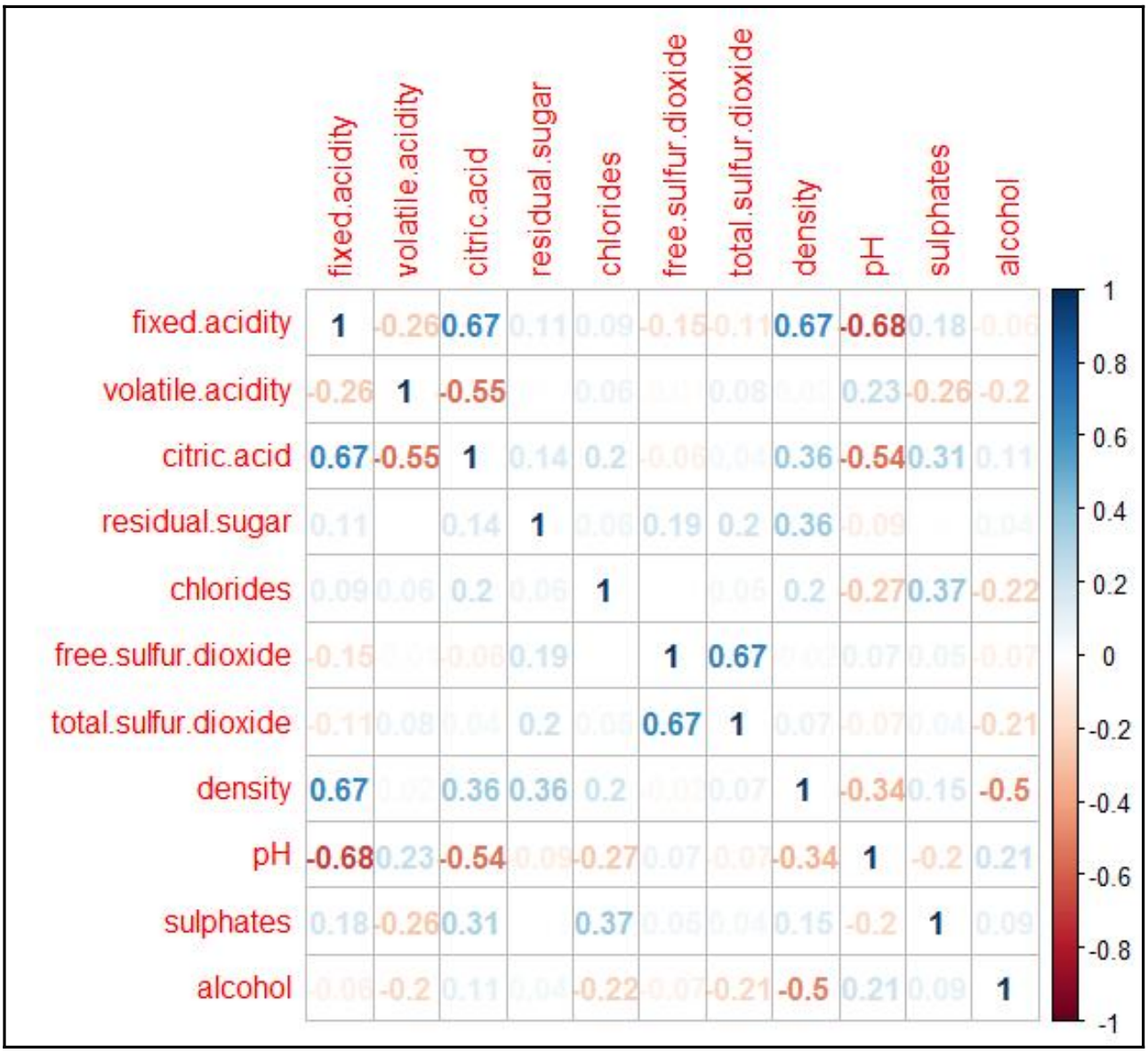

data[,2]
150
100
50
0
60
80
100
120
140
160
180
data[,1]

```
> ma
Revised data:
    war fly ver end gro hai
ant   0   0   0   0   1   0
bee   0   1   0   0   1   1
cat   1   0   1   0   0   1
cpl   0   0   0   0   0   1
chi   1   0   1   1   1   1
cow   1   0   1   0   1   1
duc   1   1   1   0   1   0
eag   1   1   1   1   0   0
ele   1   0   1   1   1   0
fly   0   1   0   0   0   0
fro   0   0   1   1   0   0
her   0   0   1   0   1   0
lio   1   0   1   1   1   1
liz   0   0   1   0   0   0
lob   0   0   0   0   0   0
man   1   0   1   1   1   1
rab   1   0   1   0   1   1
sal   0   0   1   0   0   0
spi   0   0   0   0   0   1
wha   1   0   1   1   1   0
Order of objects:
 [1] ant cpl spi lob bee fly fro her liz sal cat cow rab chi lio man
[17] ele wha duc eag
Variable used:
 [1] gro  NULL hai  fly  gro  ver  end  gro  NULL war  gro  NULL end
[14] NULL NULL hai  NULL fly  end
Separation step:
 [1] 4 0 5 3 4 2 3 4 0 1 4 0 3 0 0 4 0 2 3

Available components:
[1] "data"      "order"     "variable"  "step"      "call"
[6] "order.lab"
```

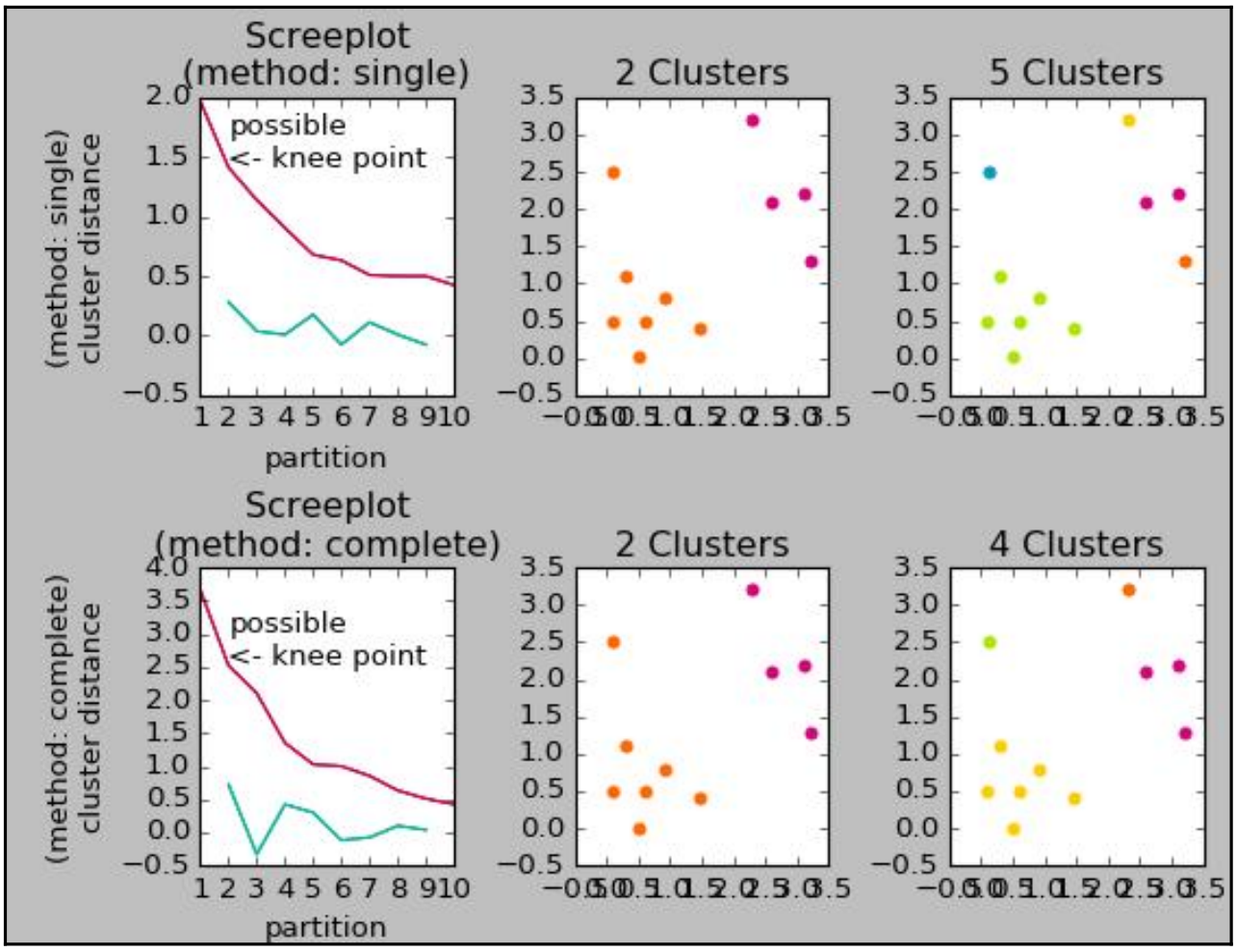

Banner of mona(x = animals)
ant
cpl
spi
lob
bee
fly
fro
her
liz
sal
cat
cow
rab
chi
lio
man
ele
wha
duc
eag
gro
hai
fly
gro
ver
end
gro
war
gro
end
hai
fly
end
0
1
2
3
4
5
6
Separation step


Dendrogram of agnes(x = animals)
Height
0.0
0.5
1.0
1.5
ant
lob
fro
her
liz
sal
bee
fly
cpl
spi
cat
chi
man
cow
lio
rab
ele
wha
duc
eag
animals
agnes (*, "average")


clusplot(clara(x = x, k = 2, samples = 50))
Component 2
20
10
0
-10
-20
-50
0
50
Component 1
These two components explain 100 % of the point variability.




```
In [13]: %run
"C:/yan/_now/_01_DataScience_Anaconda/codes/c8_05_dir_scipy_cluster.py"
['Tester', '__all__', '__builtins__', '__doc__', '__file__', '__name__',
'__package__', '__path__', '_hierarchy', '_vq', 'absolute_import',
'division', 'hierarchy', 'print_function', 'test', 'vq']

In [13]:
```




```
(base) C:\Users\yany>pip install contrastive
Collecting contrastive
  Downloading contrastive-0.1.6.tar.gz
Requirement already satisfied: numpy in c:\users\yany\.julia\v0.6\conda\deps\usr
\lib\site-packages (from contrastive)
Collecting sklearn (from contrastive)
  Downloading sklearn-0.0.tar.gz
```

```
In [3]: from contrastive import cluster
        x=dir(cluster)
        print(x)
```

```
['AffinityPropagation', 'AgglomerativeClustering', 'Birch',
'DBSCAN', 'FeatureAgglomeration', 'KMeans', 'MeanShift', 'M
iniBatchKMeans', 'SpectralBiclustering', 'SpectralClusterin
g', 'SpectralCoclustering', '__all__', '__builtins__', '__d
oc__', '__file__', '__name__', '__package__', '__path__', '
_dbscan_inner', '_feature_agglomeration', '_hierarchical',
'_k_means', '_k_means_elkan', 'affinity_propagation', 'affi
nity_propagation_', 'bicluster', 'birch', 'dbscan', 'dbscan
_', 'estimate_bandwidth', 'get_bin_seeds', 'hierarchical',
'k_means', 'k_means_', 'linkage_tree', 'mean_shift', 'mean_
shift_', 'spectral', 'spectral_clustering', 'ward_tree']
```

```
In [6]: import sklearn as sk
        x=dir(sk)
        print(x)
```

```
['_ASSUME_FINITE', '__SKLEARN_SETUP__', '__all__', '__built
ins__', '__check_build', '__doc__', '__file__', '__name__',
'__package__', '__path__', '__version__', '_contextmanager'
, '_isotonic', 'base', 'clone', 'cluster', 'config_context'
, 'decomposition', 'exceptions', 'externals', 'feature_sele
ction', 'get_config', 'isotonic', 'linear_model', 'logger',
'logging', 'manifold', 'metrics', 'model_selection', 'neigh
bors', 'os', 'preprocessing', 'random_projection', 're', 's
et_config', 'setup_module', 'svm', 'sys', 'tree', 'utils',
'warnings']
```

```
In [5]: from sklearn import cluster
        x=dir(cluster)
        print(x)
```

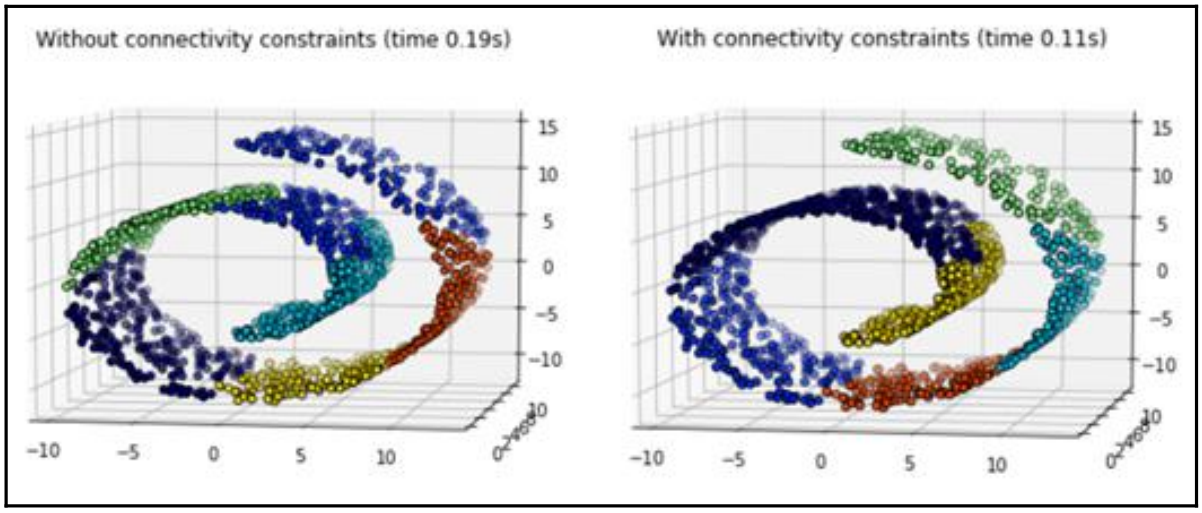
```
['AffinityPropagation', 'AgglomerativeClustering', 'Birch',
'DBSCAN', 'FeatureAgglomeration', 'KMeans', 'MeanShift', 'M
iniBatchKMeans', 'SpectralBiclustering', 'SpectralClusterin
g', 'SpectralCoclustering', '__all__', '__builtins__', '__d
oc__', '__file__', '__name__', '__package__', '__path__', '
_dbscan_inner', '_feature_agglomeration', '_hierarchical',
'_k_means', '_k_means_elkan', 'affinity_propagation', 'affi
nity_propagation_', 'bicluster', 'birch', 'dbscan', 'dbscan
_', 'estimate_bandwidth', 'get_bin_seeds', 'hierarchical',
'k_means', 'k_means_', 'linkage_tree', 'mean_shift', 'mean_
shift_', 'spectral', 'spectral_clustering', 'ward_tree']
```

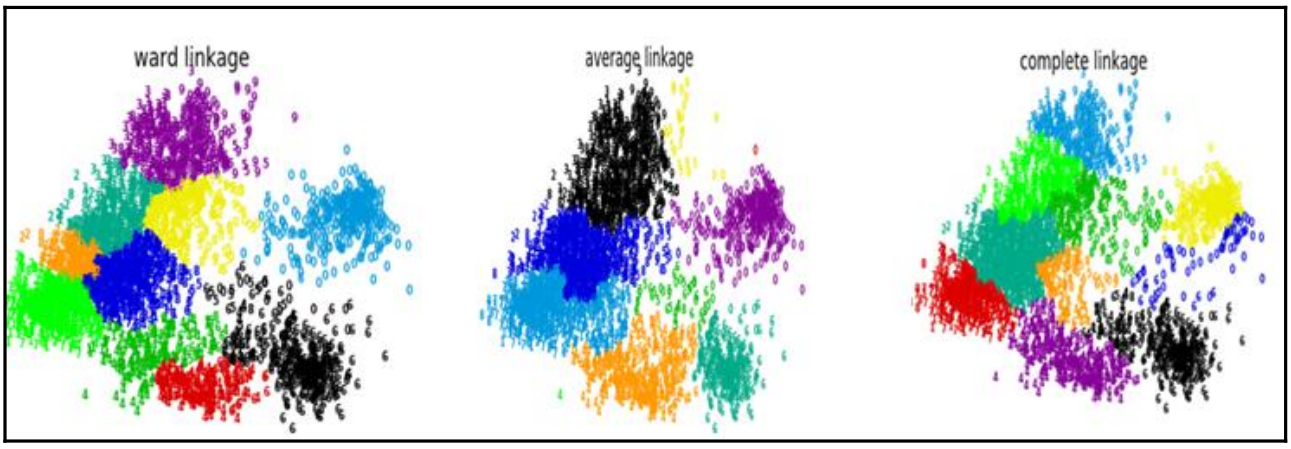
```
print(x)
```

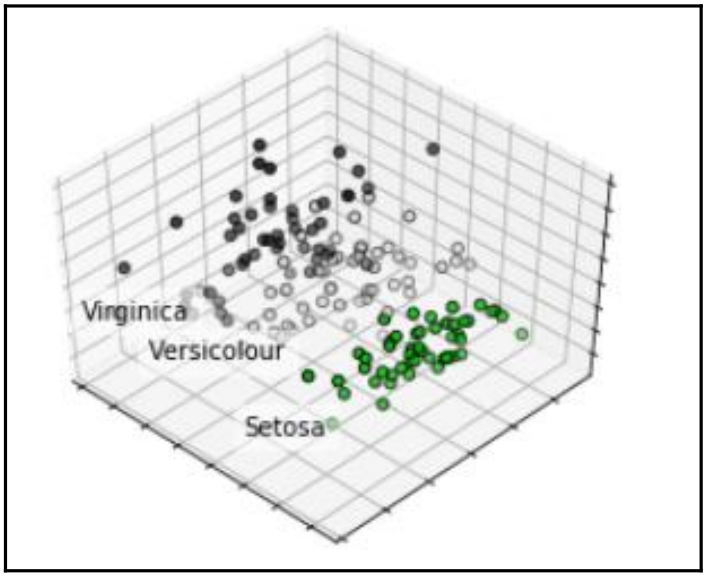
```
['__all__', '__builtins__', '__doc__', '__file__', '__name__', '__package
__', '__path__', '_svmlight_format', 'base', 'california_housing', 'clear
_data_home', 'covtype', 'dump_svmlight_file', 'fetch_20newsgroups', 'fetc
h_20newsgroups_vectorized', 'fetch_california_housing', 'fetch_covtype',
'fetch_kddcup99', 'fetch_lfw_pairs', 'fetch_lfw_people', 'fetch_mldata',
'fetch_olivetti_faces', 'fetch_rcv1', 'fetch_species_distributions', 'get
_data_home', 'kddcup99', 'lfw', 'load_boston', 'load_breast_cancer', 'loa
d_diabetes', 'load_digits', 'load_files', 'load_iris', 'load_linnerud', '
load_mlcomp', 'load_sample_image', 'load_sample_images', 'load_svmlight_f
ile', 'load_svmlight_files', 'load_wine', 'make_biclusters', 'make_blobs'
, 'make_checkerboard', 'make_circles', 'make_classification', 'make_fried
man1', 'make_friedman2', 'make_friedman3', 'make_gaussian_quantiles', 'ma
ke_hastie_10_2', 'make_low_rank_matrix', 'make_moons', 'make_multilabel_c
lassification', 'make_regression', 'make_s_curve', 'make_sparse_coded_sig
nal', 'make_sparse_spd_matrix', 'make_sparse_uncorrelated', 'make_spd_mat
rix', 'make_swiss_roll', 'mlcomp', 'mldata', 'mldata_filename', 'olivetti
_faces', 'rcv1', 'samples_generator', 'species_distributions', 'svmlight_
format', 'twenty_newsgroups']
```

```
import numpy as np
from sklearn import datasets
x= datasets.load_iris()
print(x)
```

```
{'target_names': array(['setosa', 'versicolor', 'virg
inica'], dtype='|S10'), 'data': array([[5.1, 3.5, 1.4
, 0.2],
       [4.9, 3. , 1.4, 0.2],
       [4.7, 3.2, 1.3, 0.2],
       [4.6, 3.1, 1.5, 0.2],
       [5. , 3.6, 1.4, 0.2],
       [5.4, 3.9, 1.7, 0.4],
       [4.6, 3.4, 1.4, 0.3],
```







Source: File ARFF ODBC R Dataset RData File Library Corpus Script

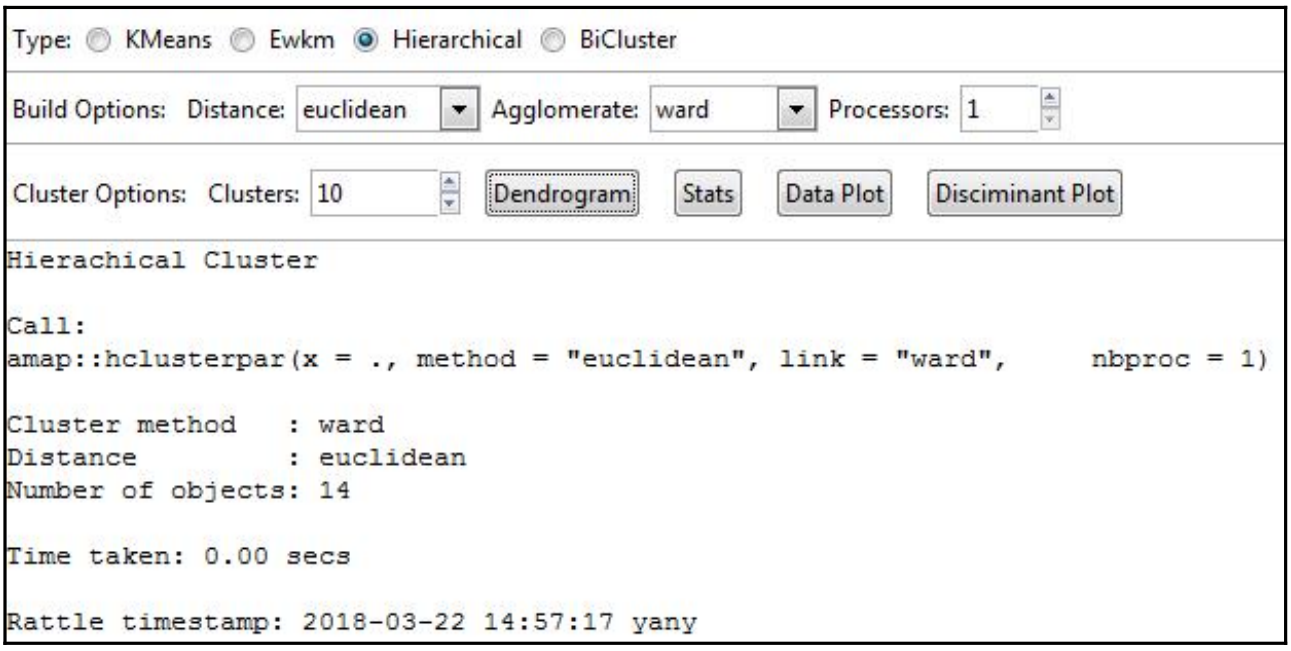
Data Name: x

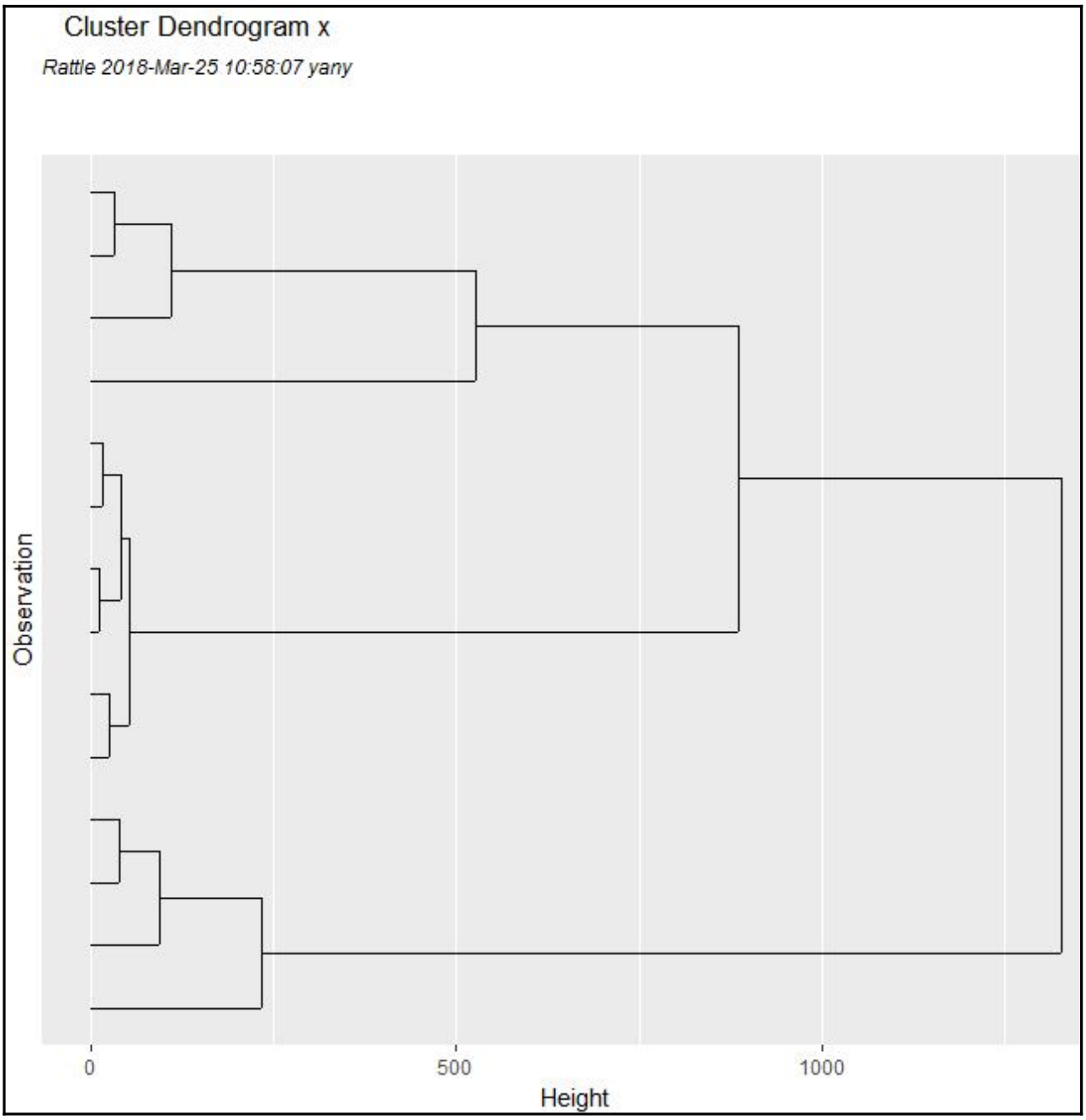
Partition 70/15/15 Seed: 42 View Edit

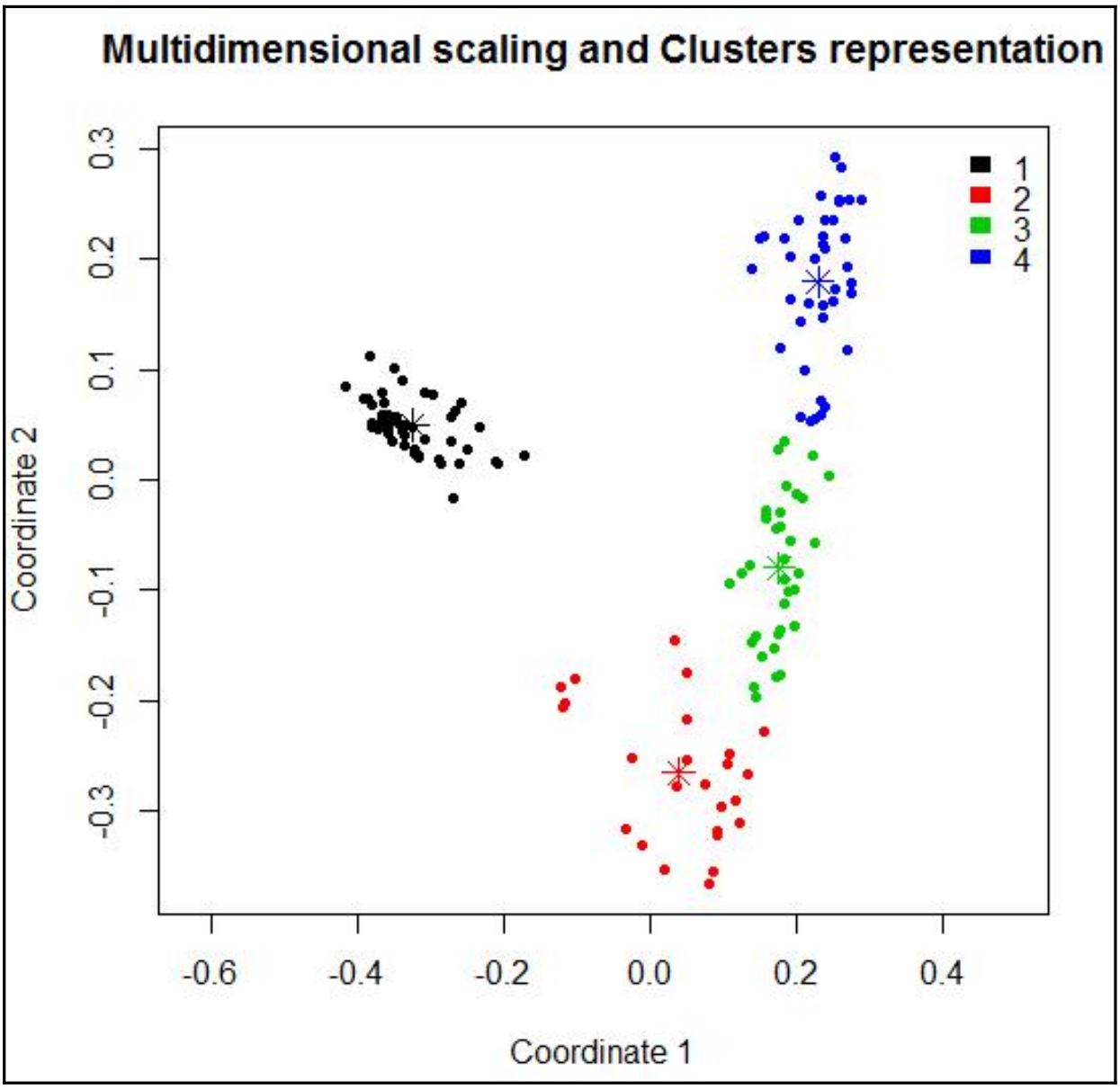
Input Ignore Weight Calculator:

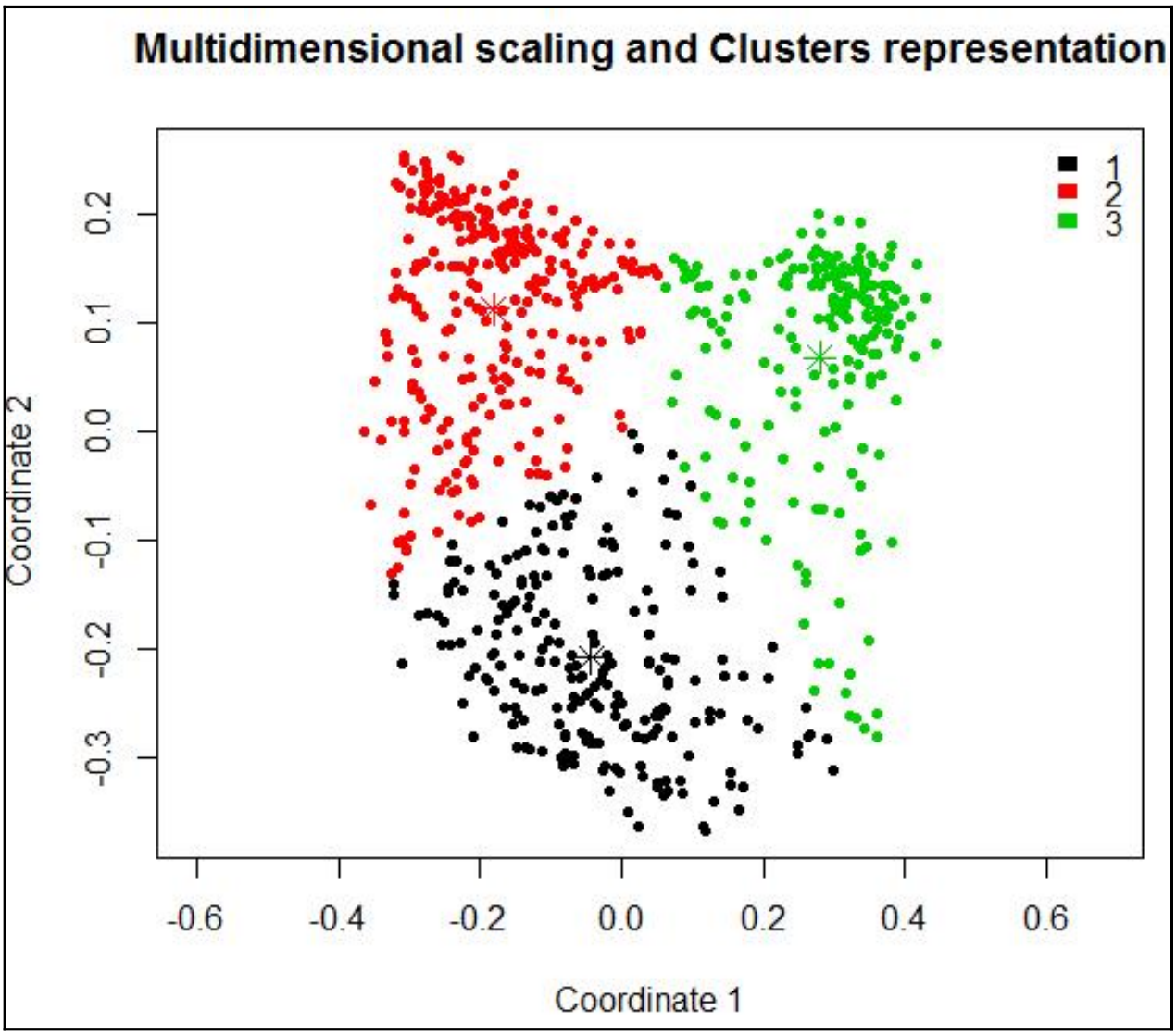
Target Data Type: Auto Categoric Numeric Survival

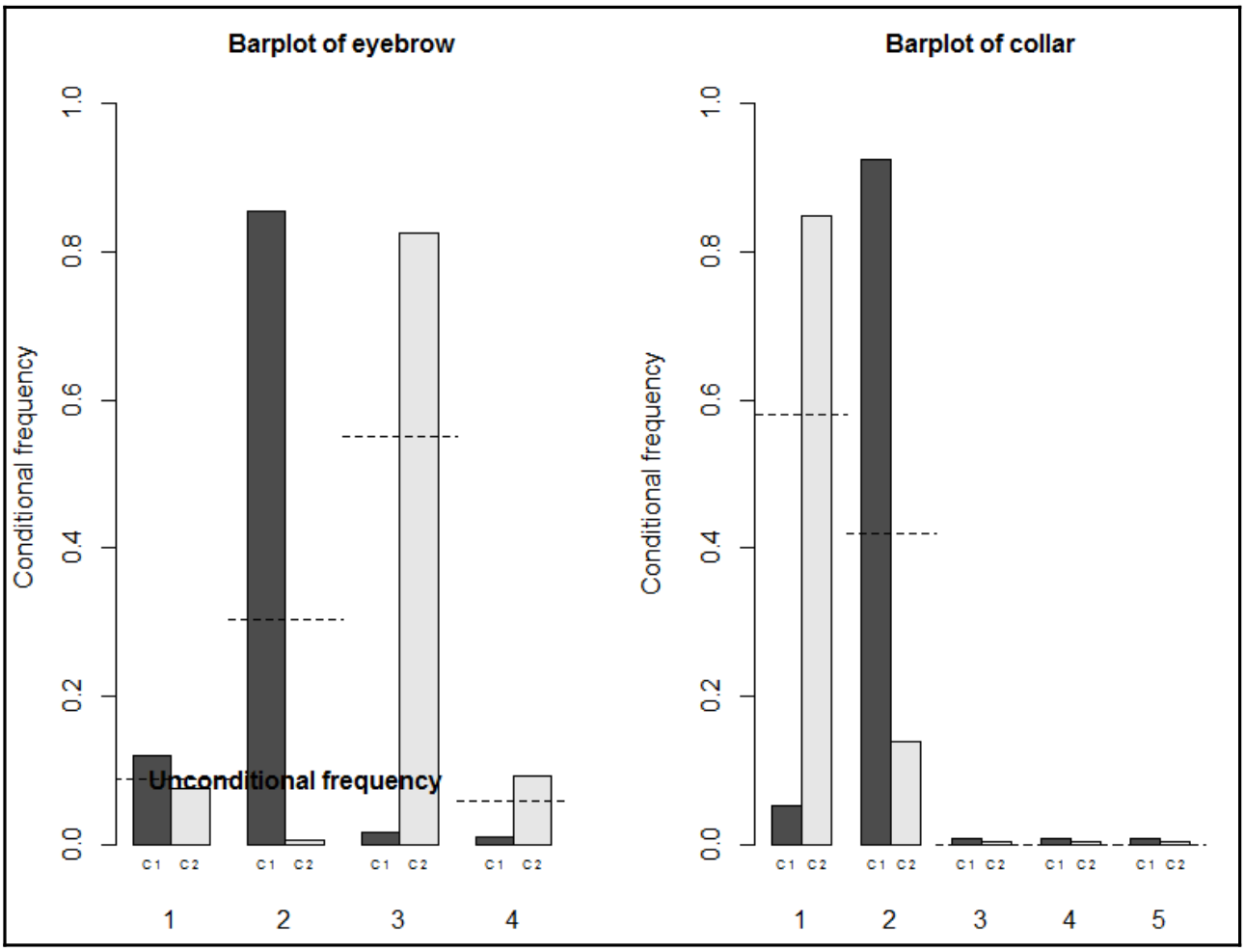
No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	Type	Categoric		●					Unique: 1
2	Alcohol	Numeric	●						Unique: 20


Type: KMeans Ewkm Hierarchical BiCluster
Build Options: Distance: euclidean Agglomerate: ward Processors: 1
Cluster Options: Clusters: 10 Dendrogram Stats Data Plot Disciminant Plot
Hierachical Cluster
Call:
amap::hclusterpar(x = ., method = "euclidean", link = "ward", nbproc = 1)
Cluster method : ward
Distance : euclidean
Number of objects: 14
Time taken: 0.00 secs
Rattle timestamp: 2018-03-22 14:57:17 yany


Cluster Dendrogram x
Rattle 2018-Mar-25 10:58:07 yany
Observation
0
500
1000
Height


Multidimensional scaling and Clusters representation
Coordinate 2
Coordinate 1
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.6
-0.4
-0.2
0.0
0.2
0.4
1
2
3
4


Multidimensional scaling and Clusters representation
1
2
3
Coordinate 2
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Coordinate 1




```
julia> R = affinityprop(S)
Clustering.AffinityPropResult([11, 26, 55, 56, 102, 111, 114, 123, 125, 126  …
401, 402, 406, 429, 434, 448, 451, 452, 486, 495], [50, 15, 29, 44, 38, 12, 51,
53, 1, 52  …  39, 34, 36, 41, 56, 20, 48, 30, 28, 35], [10, 19, 10, 8, 8, 10, 6,
 16, 8, 7  …  7, 15, 6, 5, 9, 11, 6, 8, 10, 7], 68, true)
```

CRAN Task View: Cluster Analysis & Finite Mixture Models

Maintainer: Friedrich Leisch and Bettina Gruen
Contact: Bettina.Gruen at jku.at
Version: 2018-03-09
URL: https://CRAN.R-project.org/view=Cluster

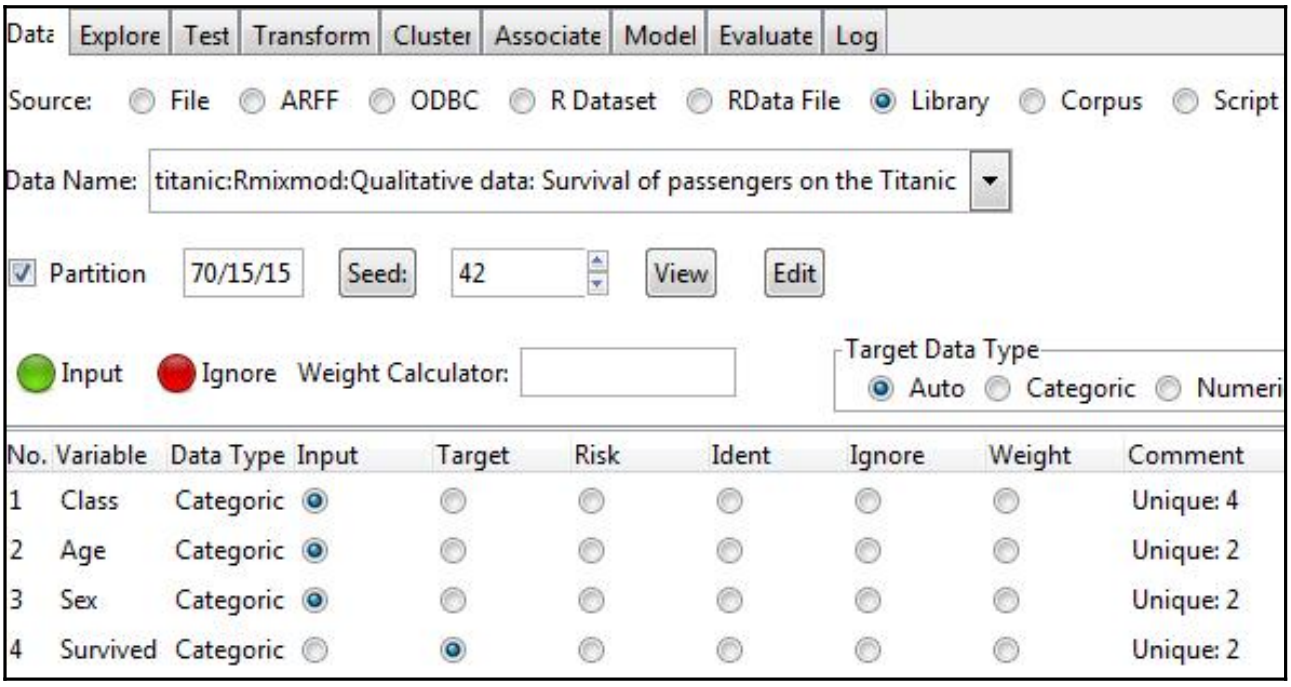
This CRAN Task View contains a list of packages that can be used for finding groups in data and modeling unobserved cross-sectional heterogeneity. Many packages provide functionality for more than one of the topics listed below, the section headings are mainly meant as quick starting points rather than an ultimate categorization. Except for packages stats and cluster (which ship with base R and hence are part of every R installation), each package is listed only once.

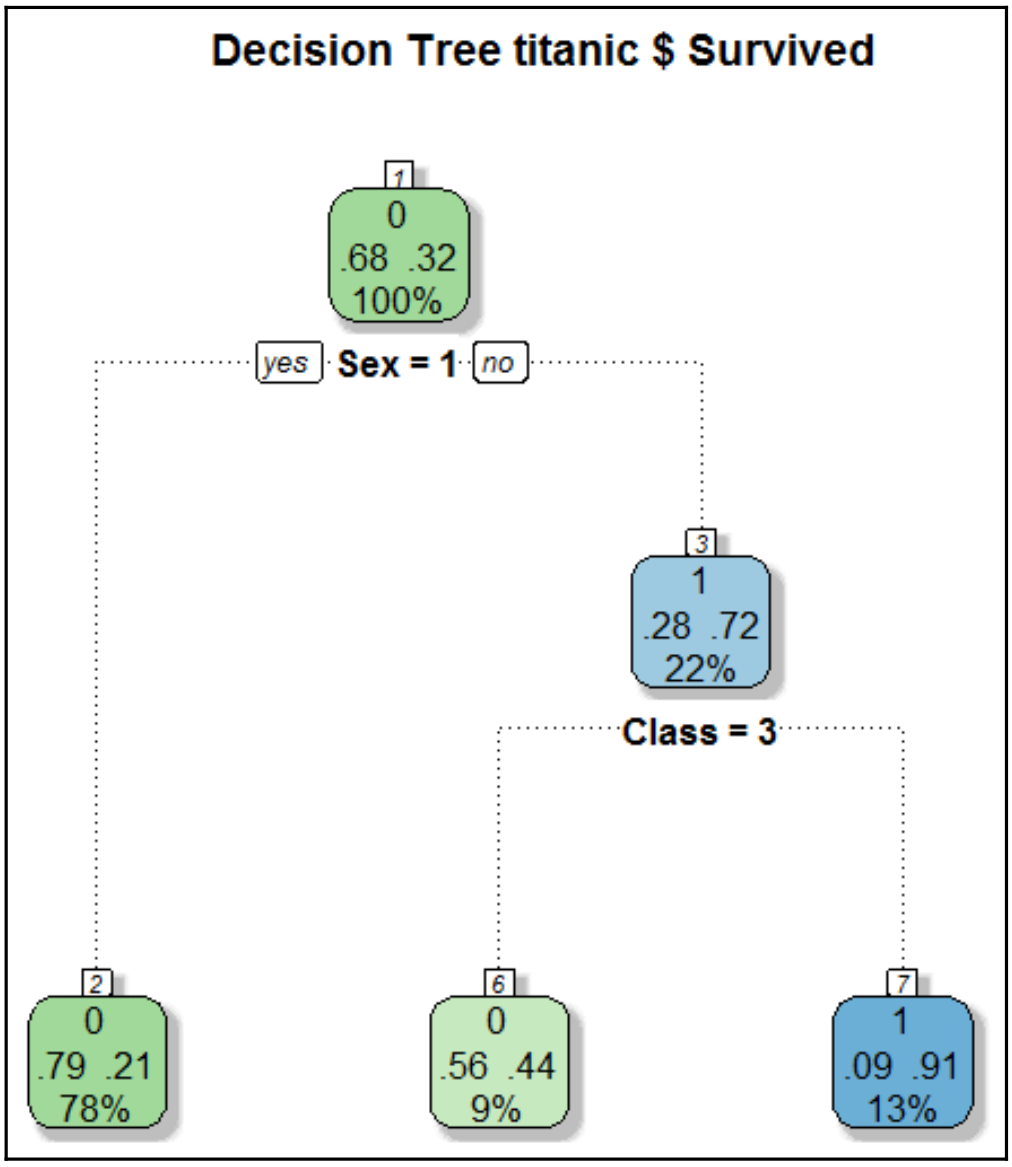
Most of the packages listed in this CRAN Task View, but not all are distributed under the GPL. Please have a look at the DESCRIPTION file of each package to check under which license it is distributed.

Hierarchical Clustering:

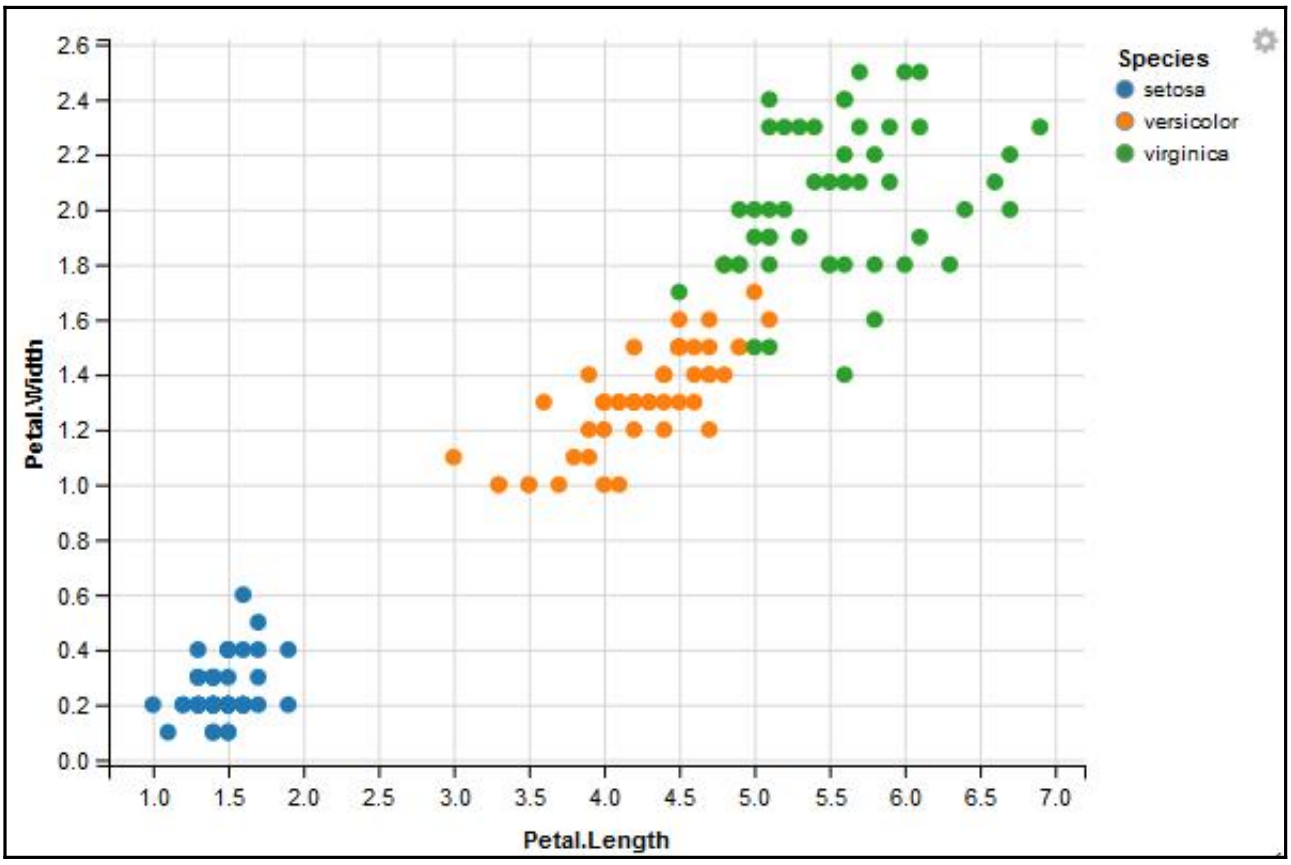
- Functions `hclust()` from package stats and `agnes()` from cluster are the primary functions for agglomerative hierarchical clustering, function `diana()` can be used for divisive hierarchical clustering. Faster alternatives to `hclust()` are provided by the packages fastcluster and flashClust.

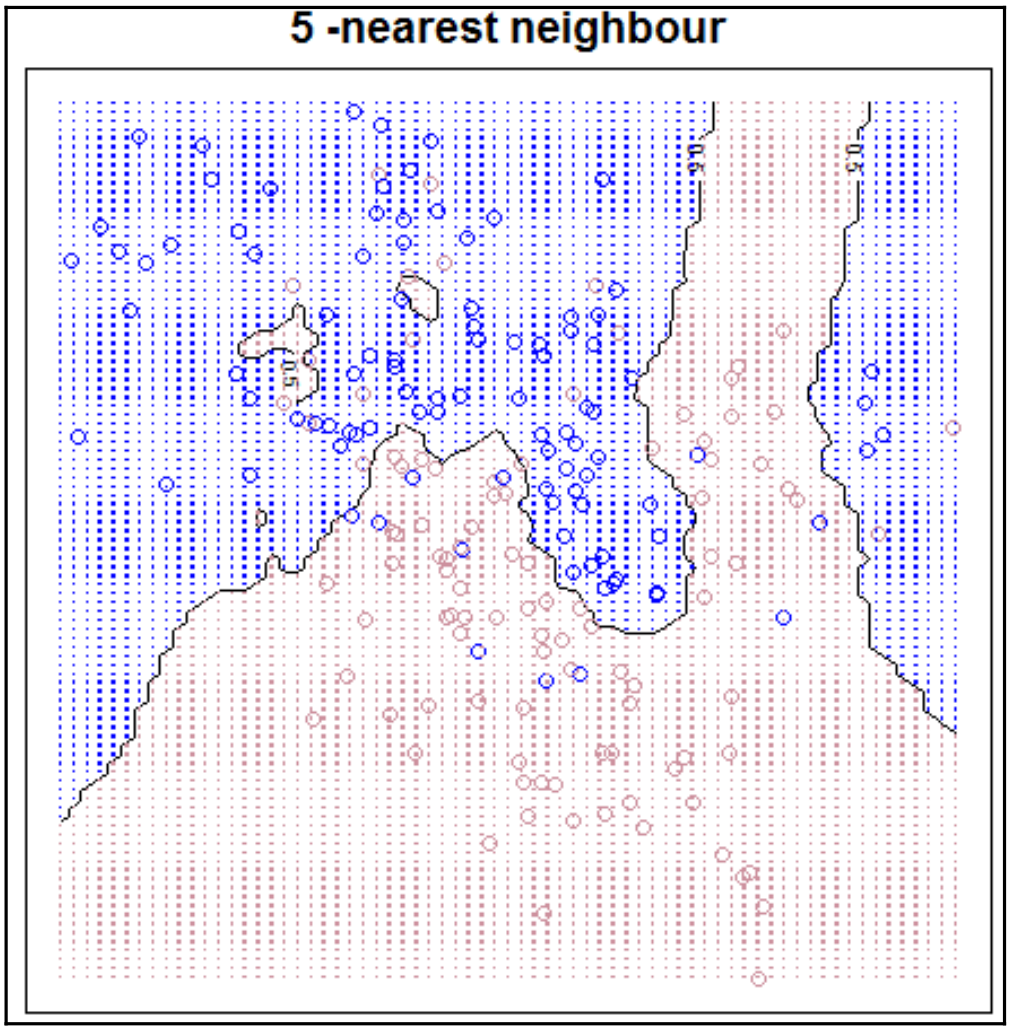
Chapter 09: Supervised Learning in Anaconda



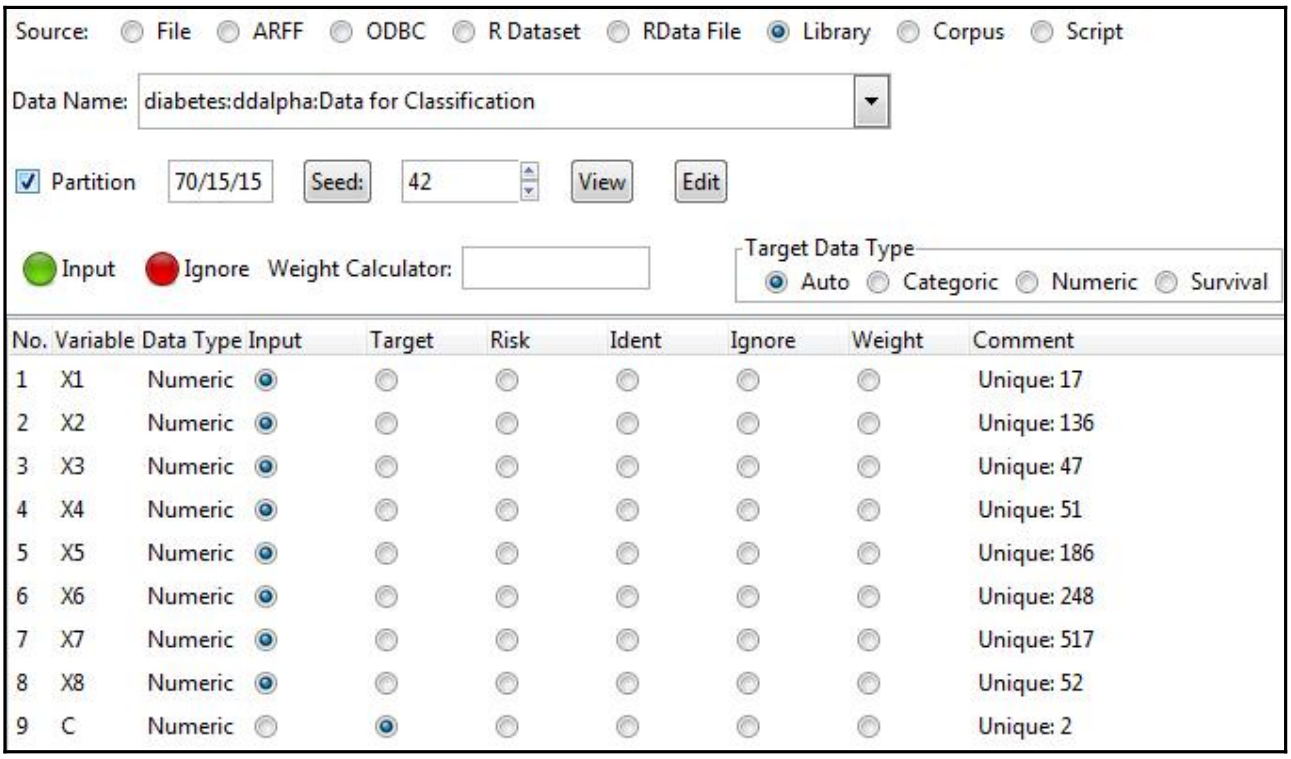



```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
```

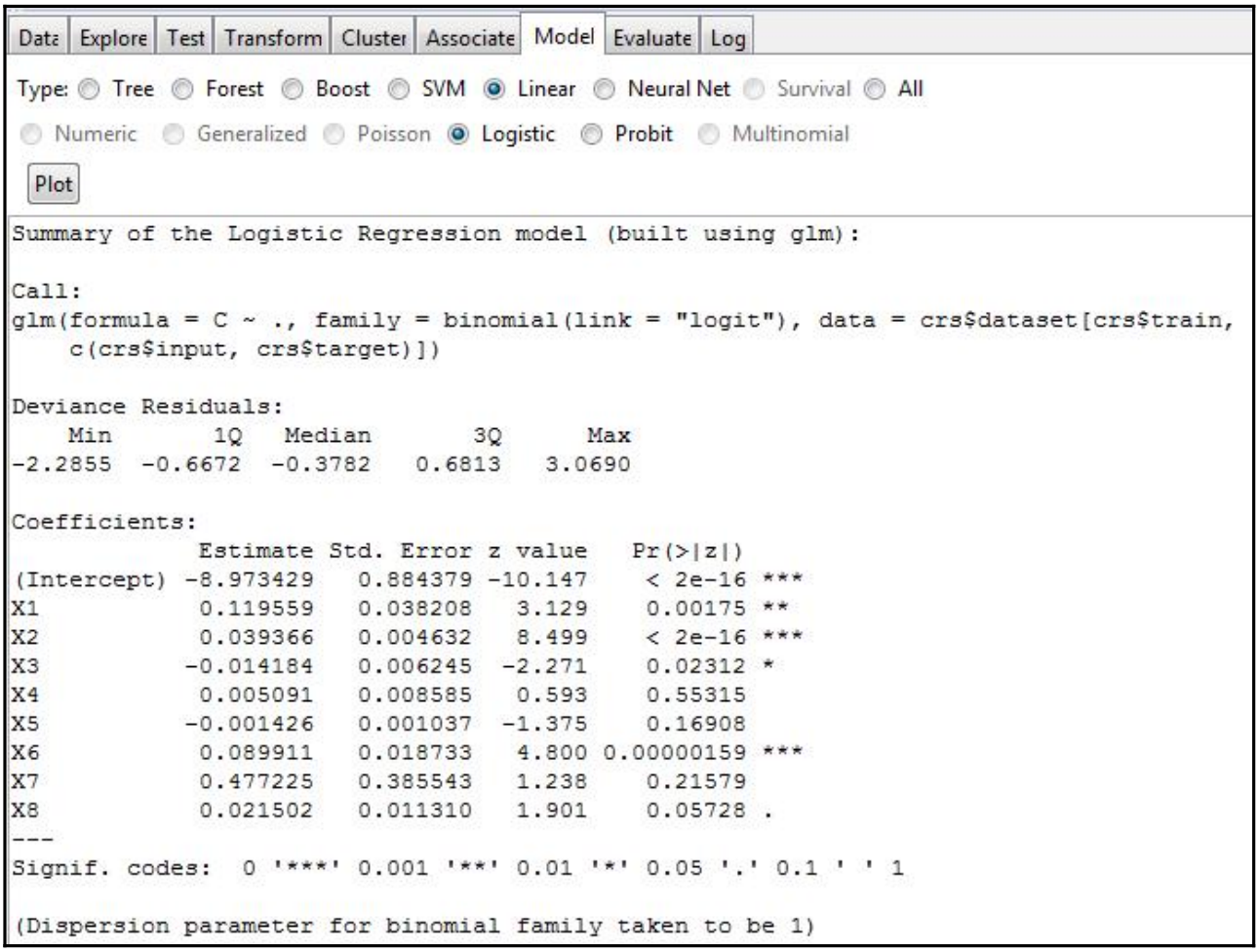

Species
setosa
versicolor
virginica
Petal.Width
Petal.Length
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0



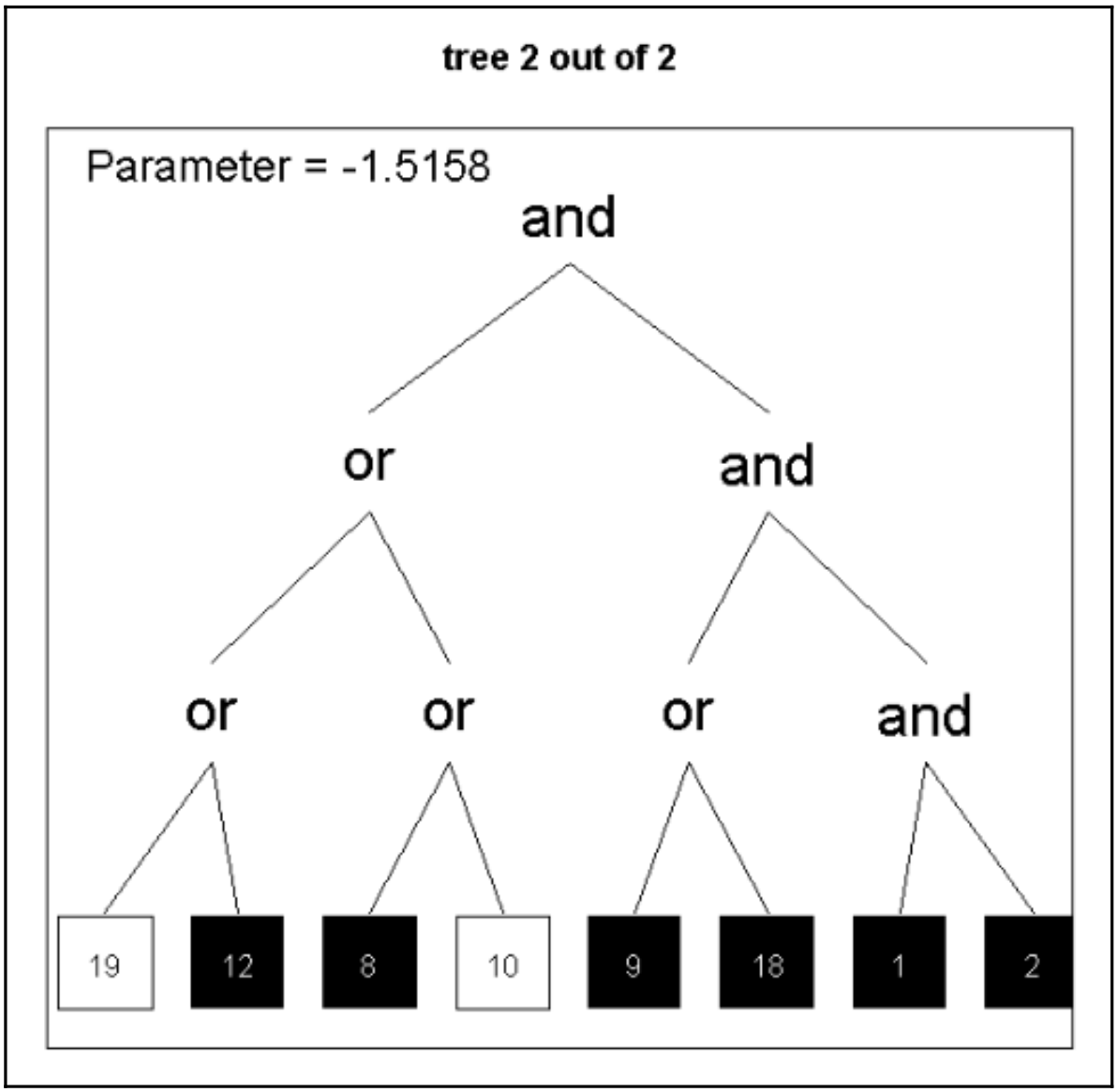

```
> head(HouseVotes84)
       Class   V1 V2 V3   V4   V5 V6 V7 V8 V9 V10  V11  V12 V13 V14 V15  V16
1 republican    n  y  n    y    y  y  n  n  n   y <NA>    y   y   y   n    y
2 republican    n  y  n    y    y  y  n  n  n   n    n    y   y   y   n <NA>
3   democrat <NA>  y  y <NA>    y  y  n  n  n   n    y    n   y   y   n    n
4   democrat    n  y  y    n <NA>  y  n  n  n   n    y    n   y   n   n    y
5   democrat    y  y  y    n    y  y  n  n  n   n    y <NA>   y   y   y    y
6   democrat    n  y  y    n    y  y  n  n  n   n    n    n   y   y   y    y
> |
```



No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	X1	Numeric	●	○	○	○	○	○	Unique: 17
2	X2	Numeric	●	○	○	○	○	○	Unique: 136
3	X3	Numeric	●	○	○	○	○	○	Unique: 47
4	X4	Numeric	●	○	○	○	○	○	Unique: 51
5	X5	Numeric	●	○	○	○	○	○	Unique: 186
6	X6	Numeric	●	○	○	○	○	○	Unique: 248
7	X7	Numeric	●	○	○	○	○	○	Unique: 517
8	X8	Numeric	●	○	○	○	○	○	Unique: 52
9	C	Numeric	○	●	○	○	○	○	Unique: 2




```
Summary of the Logistic Regression model (built using glm):

Call:
glm(formula = C ~ ., family = binomial(link = "logit"), data = crs$dataset[crs$train,
    c(crs$input, crs$target)])

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2855  -0.6672  -0.3782   0.6813   3.0690

Coefficients:
             Estimate Std. Error z value   Pr(>|z|)
(Intercept) -8.973429   0.884379 -10.147    < 2e-16 ***
X1           0.119559   0.038208   3.129    0.00175 **
X2           0.039366   0.004632   8.499    < 2e-16 ***
X3          -0.014184   0.006245  -2.271    0.02312 *
X4           0.005091   0.008585   0.593    0.55315
X5          -0.001426   0.001037  -1.375    0.16908
X6           0.089911   0.018733   4.800 0.00000159 ***
X7           0.477225   0.385543   1.238    0.21579
X8           0.021502   0.011310   1.901    0.05728 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```


tree 2 out of 2
Parameter = -1.5158
and
or
and
or
or
or
and
19
12
8
10
9
18
1
2

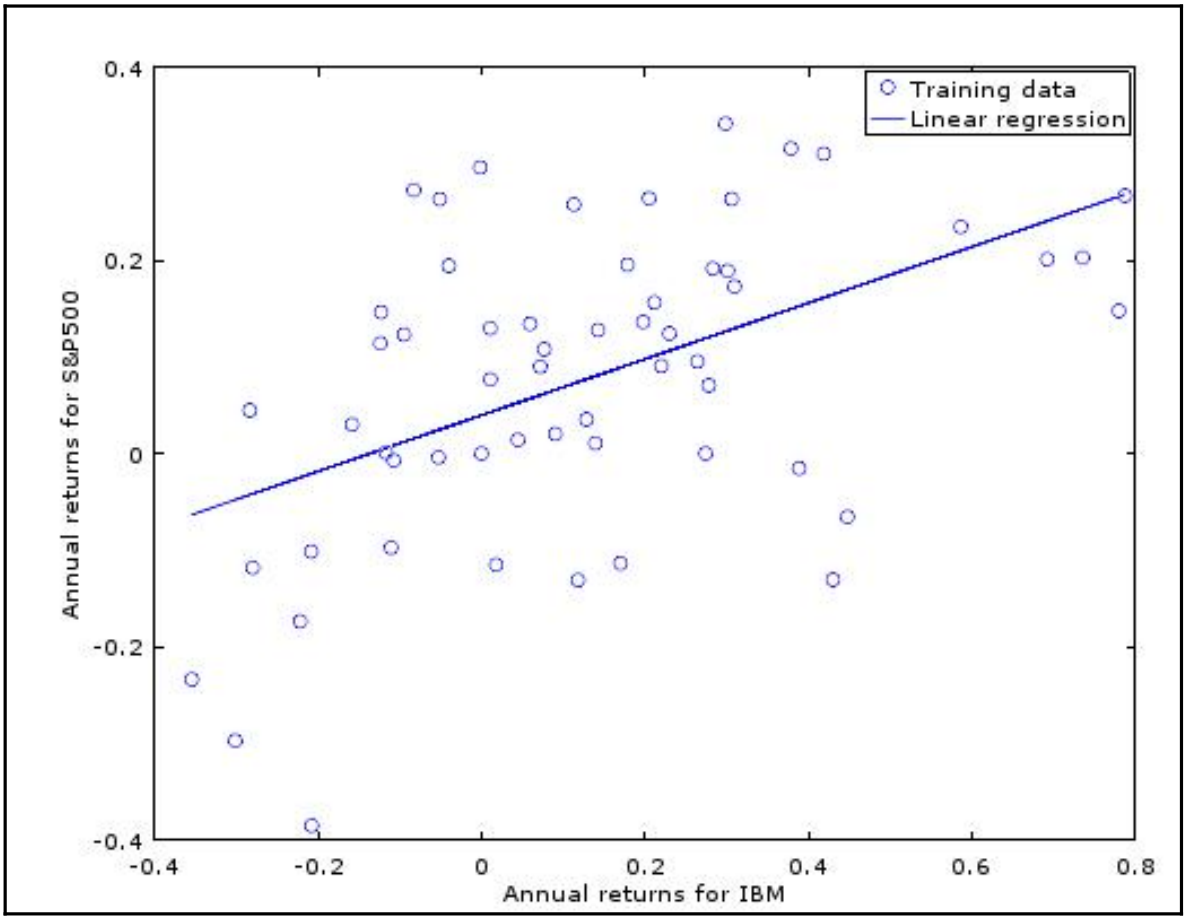
```
> head(data)
     Article_ID      Date                                                    Title
892       33741 14-Jan-99     Removal of Healthy Breasts Is Found to Cut Cancer Risk
2443       7711 17-Apr-04           In Afghanistan, U.S. Envoy Sits In Seat of Power
1268      19020 25-Apr-00              Iran Reformers Feeling Pressed By Hard-Liners
2735      27021 25-Jun-05            Arbitration Points to 10% Raise for City Police
2912      37154 25-Mar-06 New York City Meets Stranger From the West: the Brush Fire
141       42842 22-Jun-96                 Serb Gangs Rule in Last-Chance Fief of U.N.
                                                                          Subject Topic.Code
892                 At risk women take drastic measures to prevent breast cancer          3
2443 Key role of U.S. Ambassador to Afghanistan in the new government's policy.         16
1268                                                      Iranian reform politics       19
2735                                                      Salary disputes of NYPD       12
2912                                                          brush fires in NYC        27
141                                          UN involvement in Eastern Slavonia         19
```

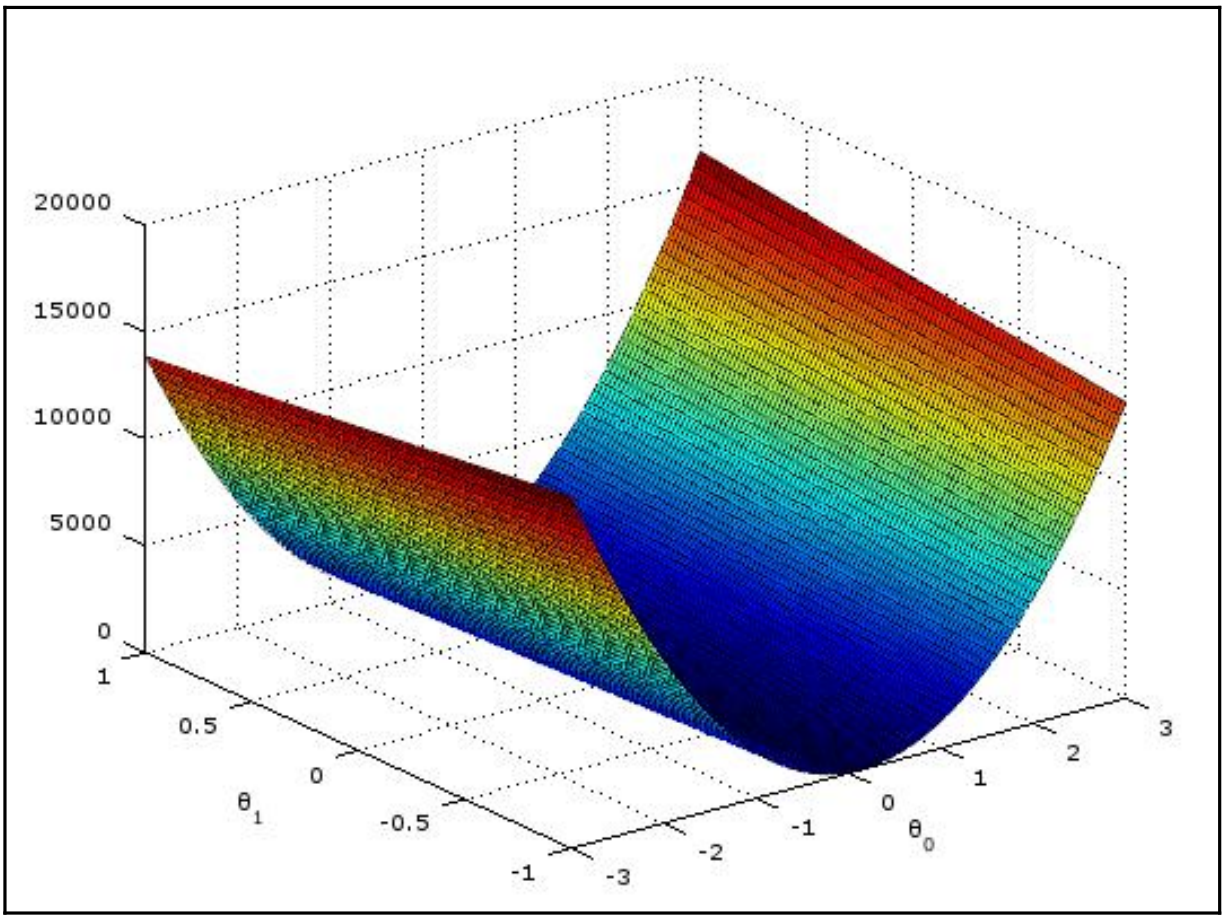
	MKT_RF	SMB	HML	Rf	MOM
1927-01-01	-0.0006	-0.0056	0.0483	0.0025	0.0044
1927-02-01	0.0418	-0.0010	0.0317	0.0026	-0.0201
1927-03-01	0.0013	-0.0160	-0.0267	0.0030	0.0359
1927-04-01	0.0046	0.0043	0.0060	0.0025	0.0419
1927-05-01	0.0544	0.0141	0.0493	0.0030	0.0301
	MKT_RF	SMB	HML	Rf	MOM
2016-06-01	-0.0005	0.0061	-0.0149	0.0002	0.0428
2016-07-01	0.0395	0.0290	-0.0098	0.0002	-0.0317
2016-08-01	0.0050	0.0094	0.0318	0.0002	-0.0316
2016-09-01	0.0025	0.0200	-0.0134	0.0002	-0.0052
2016-10-01	-0.0202	-0.0440	0.0415	0.0002	0.0058

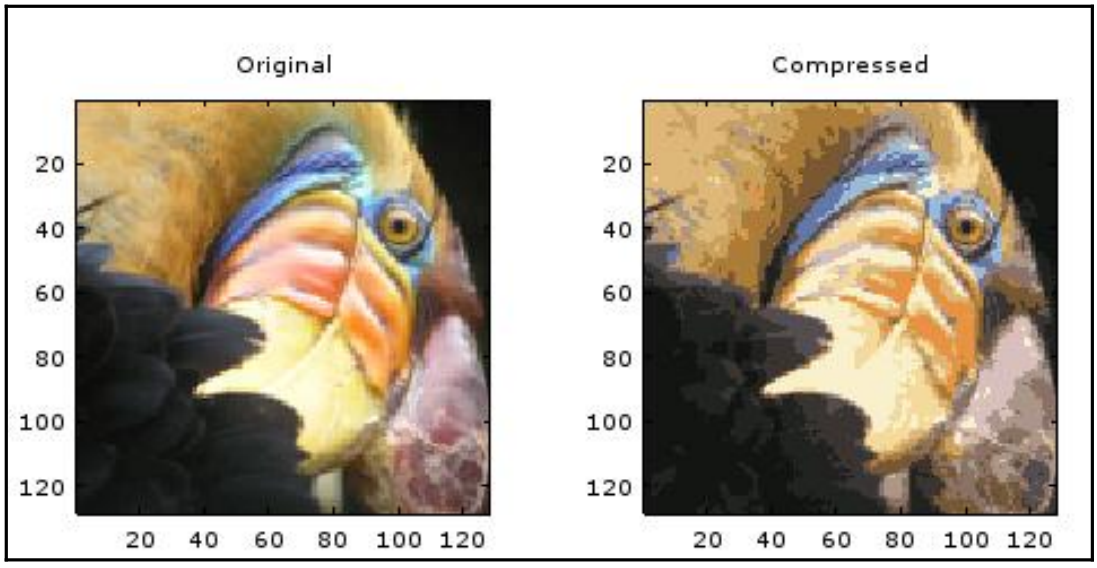
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    RET   R-squared:                       0.373
Model:                            OLS   Adj. R-squared:                  0.373
Method:                 Least Squares   F-statistic:                     2805.
Date:                Wed, 18 Apr 2018   Prob (F-statistic):               0.00
Time:                        10:15:11   Log-Likelihood:                 41169.
No. Observations:               14135   AIC:                        -8.233e+04
Df Residuals:                   14131   BIC:                        -8.230e+04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
const          0.0002      0.000      1.642      0.101     -3.52e-05     0.000
MKT_RF         0.9217      0.012     79.182      0.000         0.899     0.945
SMB           -0.4164      0.022    -19.361      0.000        -0.459    -0.374
HML           -0.4337      0.023    -18.962      0.000        -0.479    -0.389
==============================================================================
Omnibus:                    16154.987   Durbin-Watson:                   1.988
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         19838982.354
Skew:                          -5.038   Prob(JB):                         0.00
Kurtosis:                     186.258   Cond. No.                         214.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

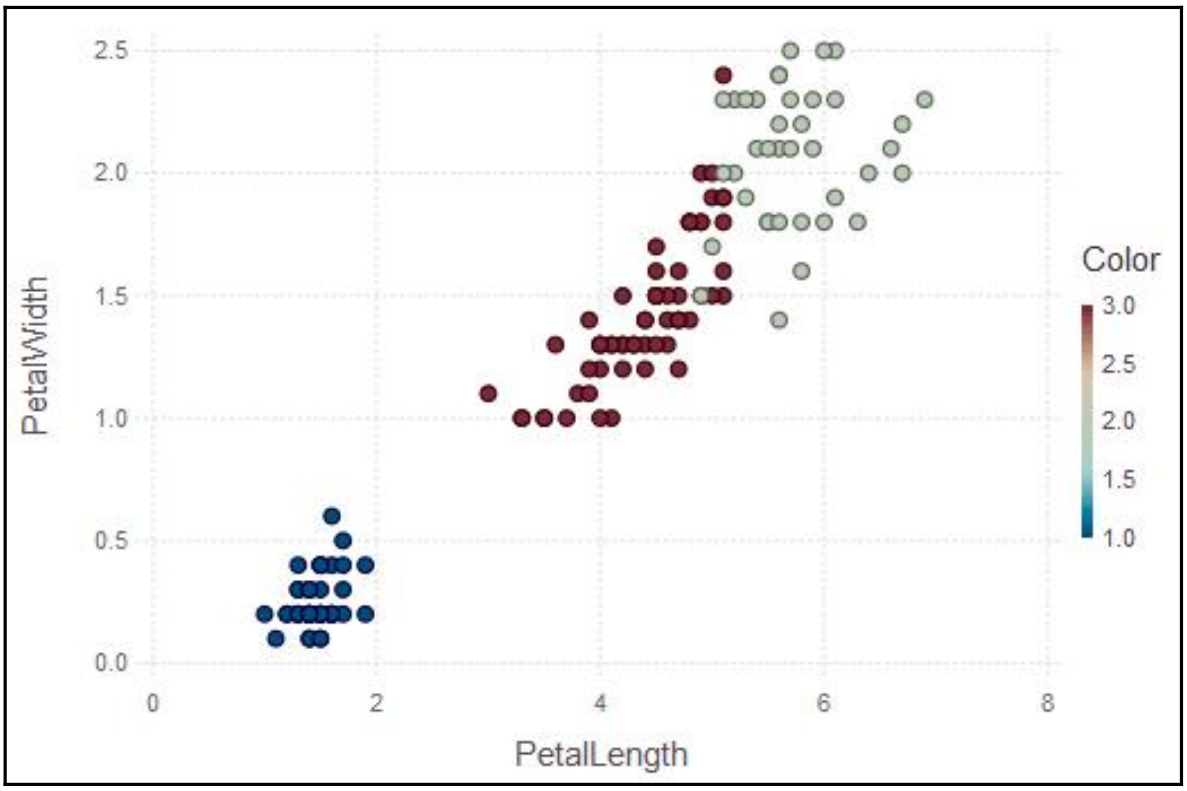
```
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=None,
            splitter='best')
------ output below ---------

             precision    recall  f1-score   support

          0       1.00      1.00      1.00        50
          1       1.00      1.00      1.00        50
          2       1.00      1.00      1.00        50

avg / total       1.00      1.00      1.00       150

[[50  0  0]
 [ 0 50  0]
 [ 0  0 50]]
```

```
Classification report for classifier SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma=0.001, kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False):
             precision    recall  f1-score   support

          0       1.00      0.99      0.99        88
          1       0.99      0.97      0.98        91
          2       0.99      0.99      0.99        86
          3       0.98      0.87      0.92        91
          4       0.99      0.96      0.97        92
          5       0.95      0.97      0.96        91
          6       0.99      0.99      0.99        91
          7       0.96      0.99      0.97        89
          8       0.94      1.00      0.97        88
          9       0.93      0.98      0.95        92

avg / total       0.97      0.97      0.97       899
```




20000
15000
10000
5000
0
1
0.5
0
-0.5
-1
θ1
-3
-2
-1
0
1
2
3
θ0


Original
Compressed
20
40
60
80
100
120
20 40 60 80 100 120




```
julia> x
5x1000 Array{Float64,2}:
 0.590845  0.854147  0.648882   0.112486    .  0.853632  0.325751  0.707301
 0.766797  0.200586  0.0109059  0.276021       0.151614  0.798595  0.608648
 0.566237  0.298614  0.066423   0.651664       0.862541  0.115767  0.0761294
 0.460085  0.246837  0.956753   0.0566425      0.52499   0.401829  0.901916
 0.794026  0.579672  0.646691   0.842714       0.932708  0.994582  0.348597

julia> clusters
5x20 Array{Float64,2}:
 0.731216  0.80513   0.306281  0.322454  .  0.745429  0.727124  0.278213
 0.799427  0.212387  0.191713  0.280681     0.708182  0.709777  0.216027
 0.808221  0.293112  0.733413  0.200435     0.663792  0.192699  0.654714
 0.309802  0.713249  0.799372  0.181764     0.755552  0.792906  0.329358
 0.595867  0.33937   0.266841  0.450164     0.719329  0.395103  0.259146
```

CRAN Task View: Machine Learning & Statistical Learning

Maintainer: Torsten Hothorn
Contact: Torsten.Hothorn at R-project.org
Version: 2018-02-28
URL: https://CRAN.R-project.org/view=MachineLearning

Several add-on packages implement ideas and methods developed at the borderline between computer science and statistics - this field of research is usually referred to as machine learning. The packages can be roughly structured into the following topics:

- *Neural Networks and Deep Learning* : Single-hidden-layer neural network are implemented in package nnet (shipped with base R). Package RSNNS offers an interface to the Stuttgart Neural Network Simulator (SNNS). An interface to the FCNN library allows user-extensible artificial neural networks in package FCNN4R. rnn implements recurrent neural networks. Packages implementing deep learning flavours of neural networks include deepnet (feed-forward neural network, restricted Boltzmann machine, deep belief network, stacked autoencoders), RcppDL (denoising autoencoder, stacked denoising autoencoder, restricted Boltzmann machine, deep belief network) and h2o (feed-forward neural network, deep autoencoders). An interface to tensorflow is available in tensorflow.

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
```

Chapter 10: Predictive Data Analytics – Modeling and Validation

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Browse Through: **427** Data Sets

Default Task
Classification (314)
Regression (82)
Clustering (74)
Other (54)

Attribute Type
Categorical (37)
Numerical (274)
Mixed (55)

Data Type
Multivariate (325)
Univariate (19)
Sequential (45)
Time-Series (79)
Text (44)
Domain-Theory (23)
Other (21)

Area
Life Sciences (97)
Physical Sciences (47)
CS / Engineering (150)
Social Sciences (24)
Business (26)
Game (10)
Other (69)

Attributes
Less than 10 (98)
10 to 100 (195)
Greater than 100 (78)

Instances
Less than 100 (25)
100 to 1000 (148)
Greater than 1000 (222)

Format Type
Matrix (293)
Non-Matrix (134)

Name	Data Types
Abalone	Multivariate
Adult	Multivariate
Annealing	Multivariate
Anonymous Microsoft Web Data	
Arrhythmia	Multivariate
Artificial Characters	Multivariate
Audiology (Original)	Multivariate
Audiology (Standardized)	Multivariate
Auto MPG	Multivariate
Automobile	Multivariate

UCI

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Abalone Data Set

Download: Data Folder, Data Set Description

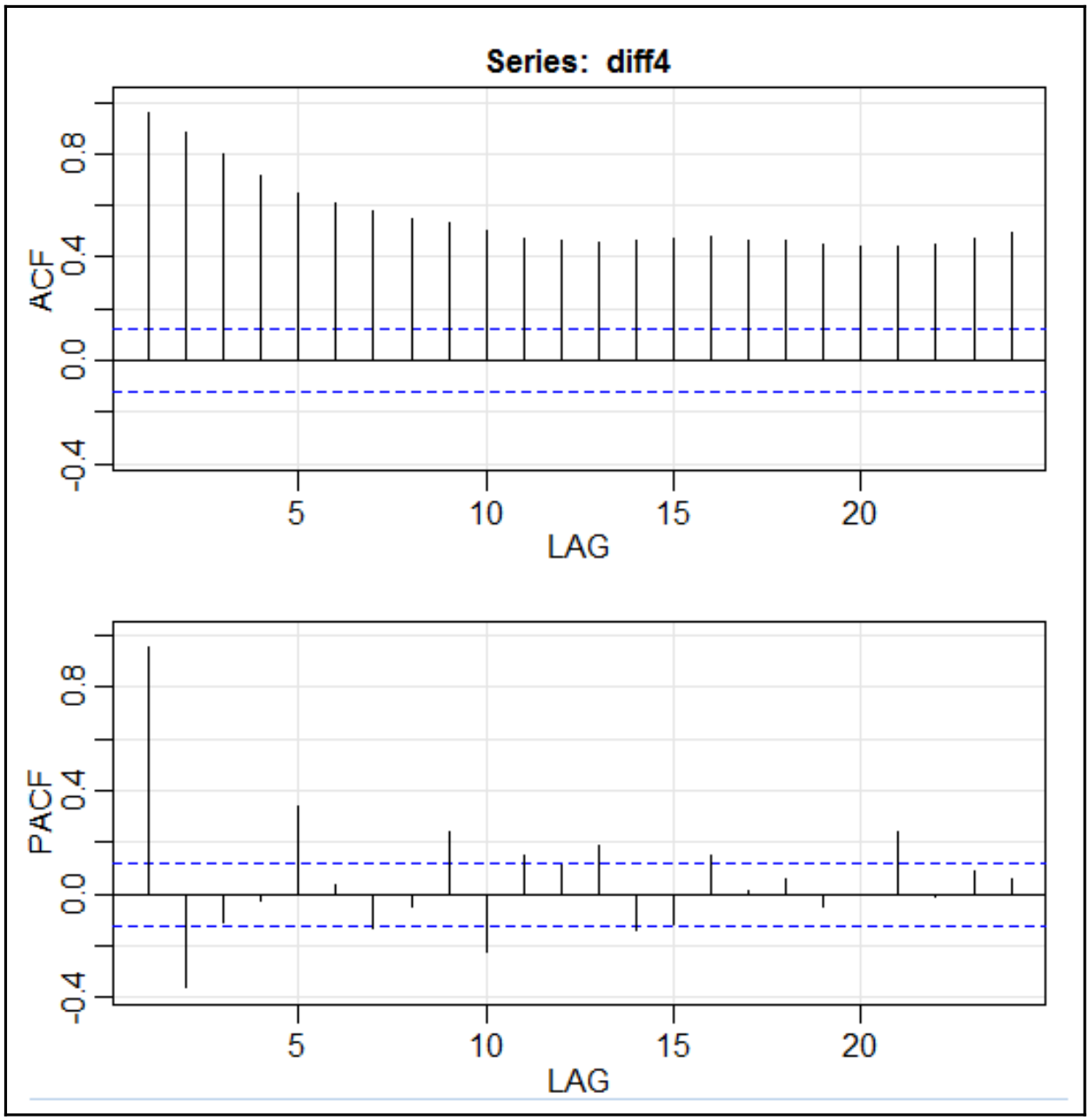
Abstract: Predict the age of abalone from physical measurements

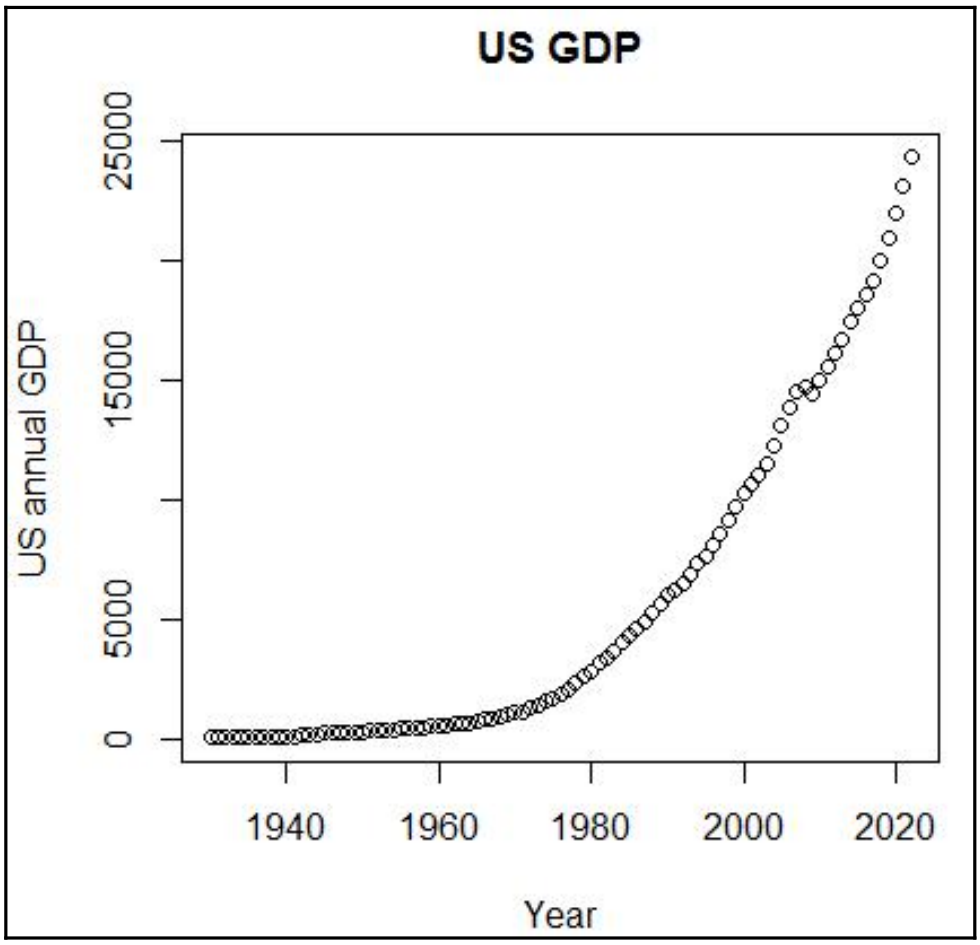
Data Set Characteristics:	Multivariate	**Number of Instances:**	4177	**Area:**	Life
Attribute Characteristics:	Categorical, Integer, Real	**Number of Attributes:**	8	**Date Donated**	1995-12-01
Associated Tasks:	Classification	**Missing Values?**	No	**Number of Web Hits:**	577549

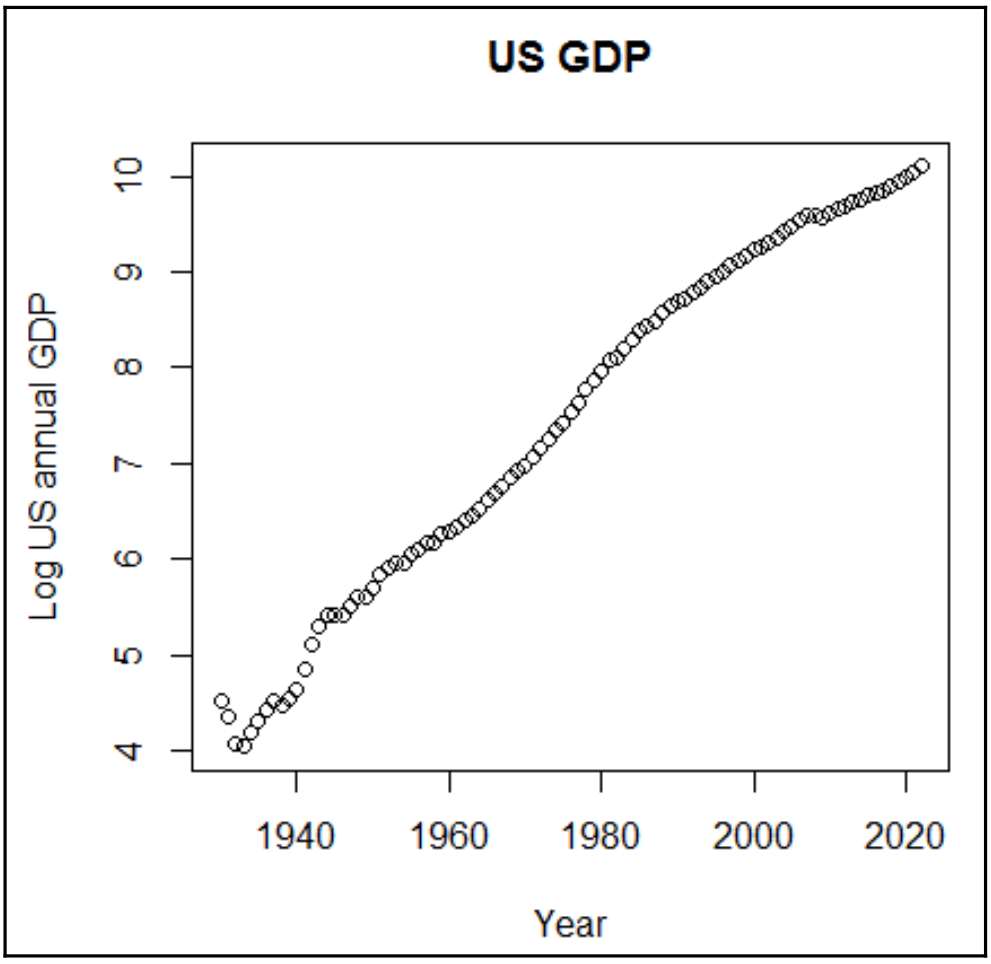
```
> dim(.UCIdatasets)
[1] 427   7
> head(.UCIdatasets)
                        Name   Data_Types        Default_Task            Attribute_Types N_Instances N_Attributes Year
1                    Abalone Multivariate      Classification Categorical, Integer, Real        4177            8 1995
2                      Adult Multivariate      Classification       Categorical, Integer       48842           14 1996
3                  Annealing Multivariate      Classification Categorical, Integer, Real         798           38   NA
4 Anonymous Microsoft Web Data             Recommender-Systems                Categorical       37711          294 1998
5                 Arrhythmia Multivariate      Classification Categorical, Integer, Real         452          279 1998
6       Artificial Characters Multivariate      Classification Categorical, Integer, Real        6000            7 1992
```

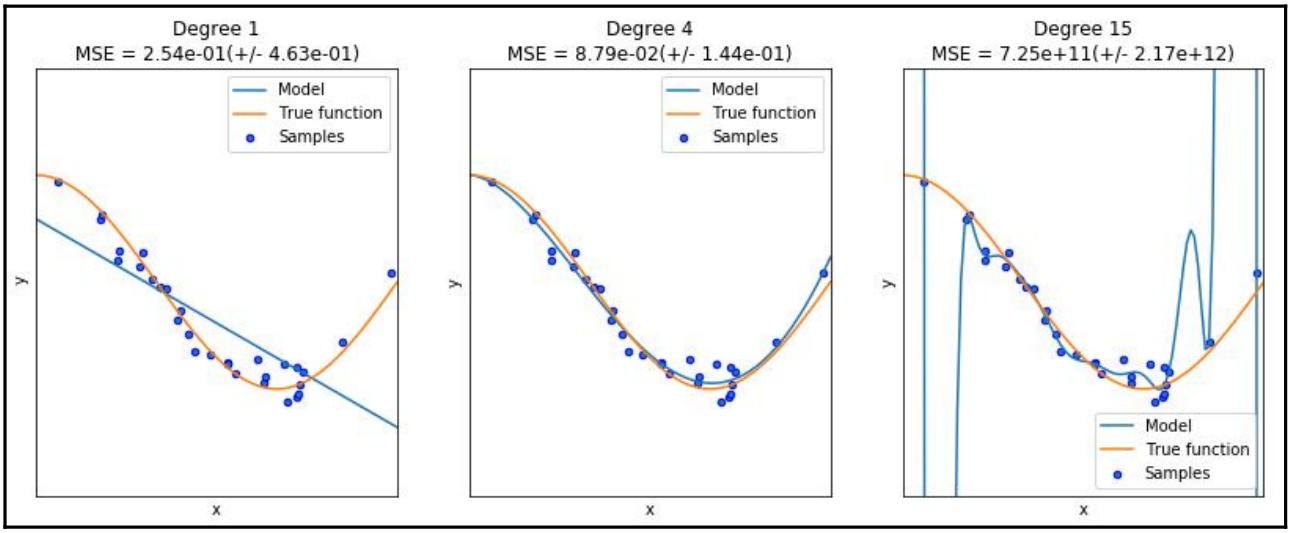
```
> unique(.UCIdatasets$Default_Task)
 [1] Classification                                  Recommender-Systems
 [3] Regression                                      Classification, Regression
 [5]                                                 Function-Learning
 [7] Relational-Learning                             Clustering
 [9] Regression, Description                         Classification, Clustering
[11] Causal-Discovery                                Classification, Causal-Discovery
[13] Regression, Clustering, Causal-Discovery        Classification, Regression, Clustering
[15] Regression, Clustering                          Regression, Classification
[17] Classification, Regression, Clustering, Causa   Classification, Clustering, Causal-Discovery
[19] Clustering, Causal-Discovery
```

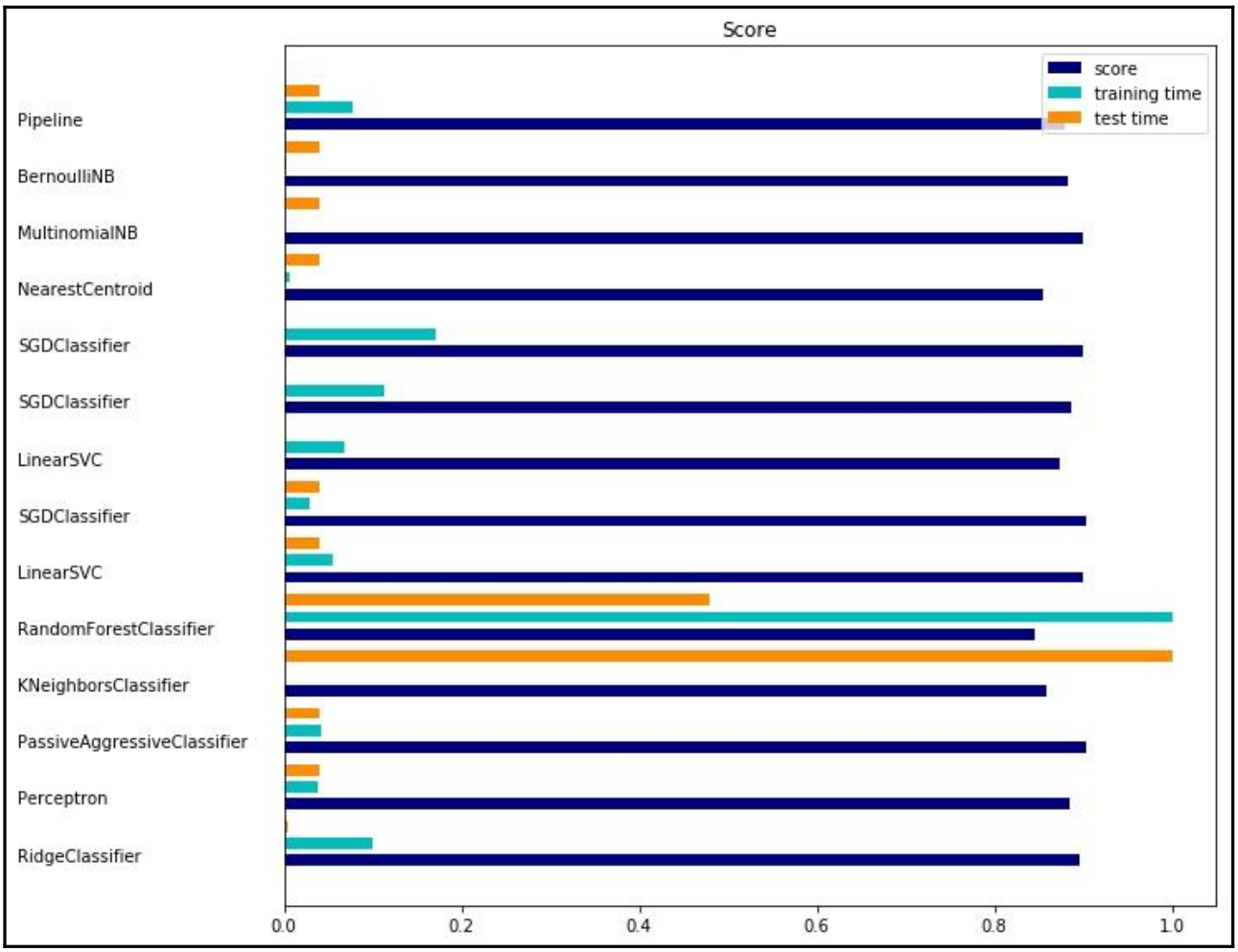
```
> dim(abalone)
[1] 4177    9
> head(abalone)
  Type LongestShell Diameter Height WholeWeight ShuckedWeight VisceraWeight ShellWeight Rings
1    M        0.455    0.365  0.095      0.5140        0.2245        0.1010       0.150    15
2    M        0.350    0.265  0.090      0.2255        0.0995        0.0485       0.070     7
3    F        0.530    0.420  0.135      0.6770        0.2565        0.1415       0.210     9
4    M        0.440    0.365  0.125      0.5160        0.2155        0.1140       0.155    10
5    I        0.330    0.255  0.080      0.2050        0.0895        0.0395       0.055     7
6    I        0.425    0.300  0.095      0.3515        0.1410        0.0775       0.120     8
>
```


Series: diff4
ACF
0.8
0.4
0.0
-0.4
5
10
15
20
LAG
PACF
0.8
0.4
0.0
-0.4
5
10
15
20
LAG

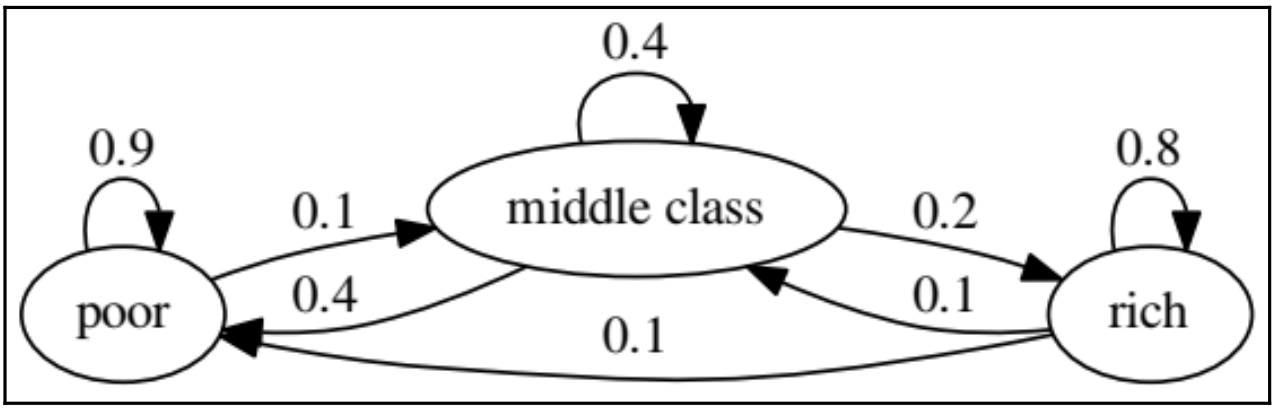

US GDP
US annual GDP
0
5000
15000
25000
1940
1960
1980
2000
2020
Year

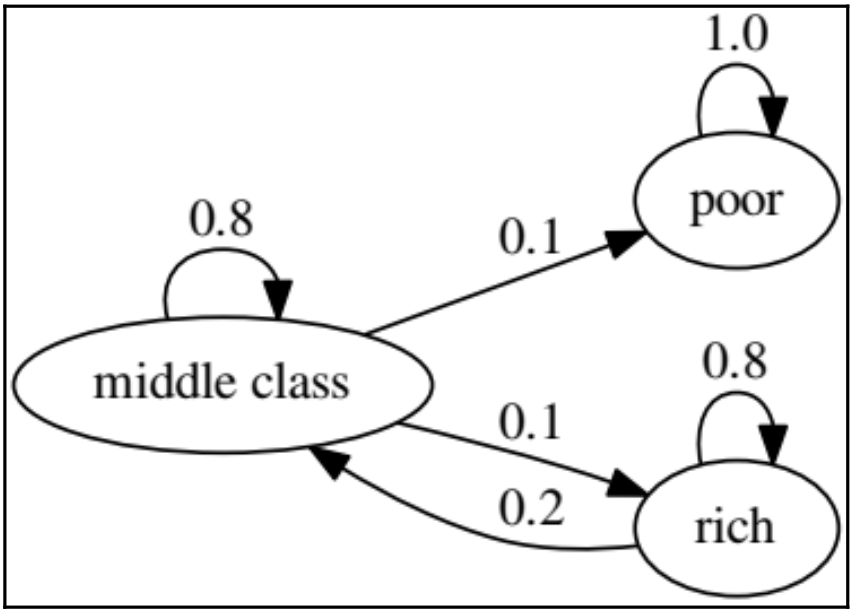

US GDP
Log US annual GDP
4
5
6
7
8
9
10
1940
1960
1980
2000
2020
Year

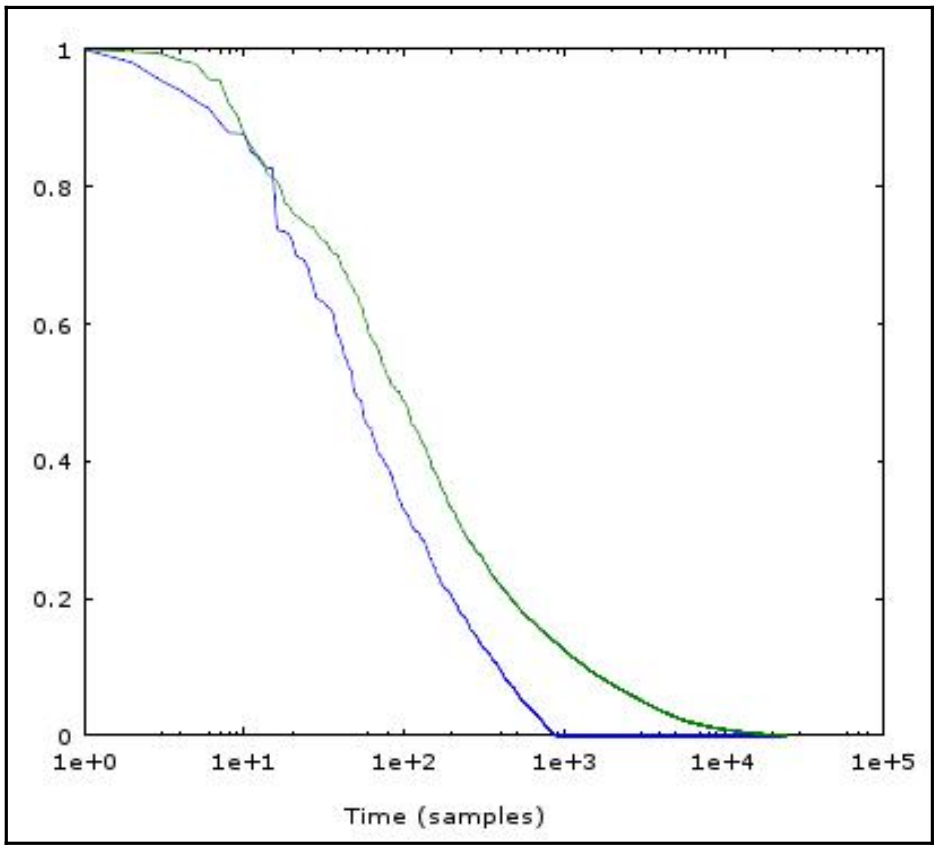

Degree 1
MSE = 2.54e-01(+/- 4.63e-01)
Degree 4
MSE = 8.79e-02(+/- 1.44e-01)
Degree 15
MSE = 7.25e+11(+/- 2.17e+12)
Model
True function
Samples
x
y



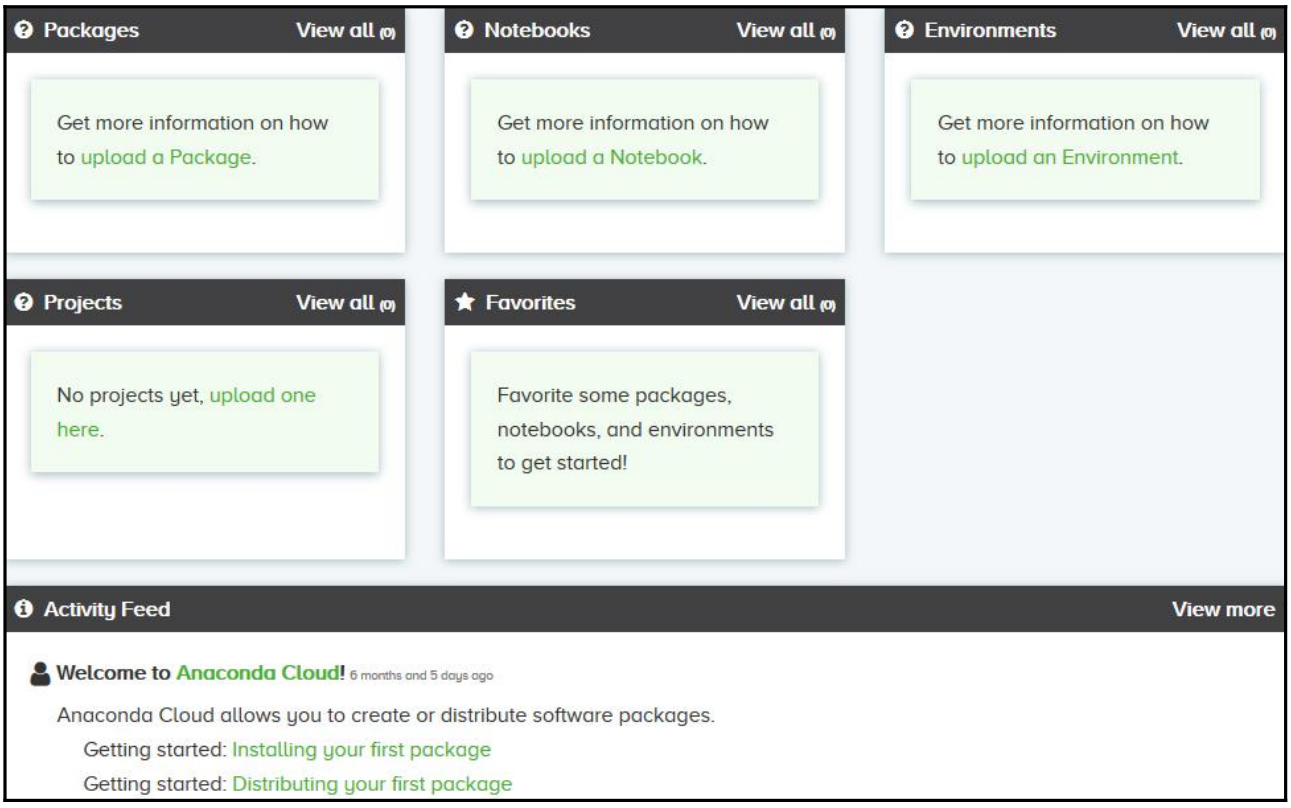

```
julia> mean(x .== 1)
0.24978

julia> #

julia> mc2 = MarkovChain(P, ["employed", "unemployed"])
Discrete Markov Chain
stochastic matrix of type Array{Float64,2}:
[0.4 0.6; 0.2 0.8]

julia> simulate(mc2, 4)
4-element Array{String,1}:
 "unemployed"
 "unemployed"
 "unemployed"
 "employed"
```

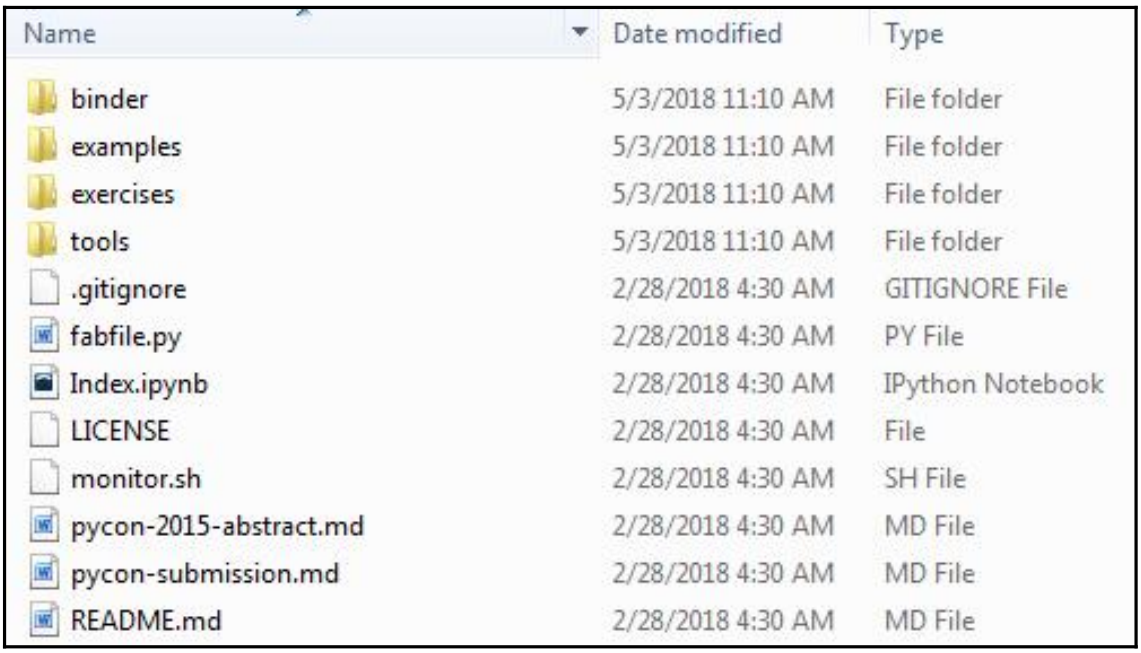

0.4
0.9
0.8
0.1
middle class
0.2
poor
0.4
0.1
rich
0.1


1.0
poor
0.8
0.1
middle class
0.8
0.1
0.2
rich


1
0.8
0.6
0.4
0.2
0
1e+0
1e+1
1e+2
1e+3
1e+4
1e+5
Time (samples)

Chapter 11: Anaconda Cloud





Name	Date modified	Type
binder	5/3/2018 11:10 AM	File folder
examples	5/3/2018 11:10 AM	File folder
exercises	5/3/2018 11:10 AM	File folder
tools	5/3/2018 11:10 AM	File folder
.gitignore	2/28/2018 4:30 AM	GITIGNORE File
fabfile.py	2/28/2018 4:30 AM	PY File
Index.ipynb	2/28/2018 4:30 AM	IPython Notebook
LICENSE	2/28/2018 4:30 AM	File
monitor.sh	2/28/2018 4:30 AM	SH File
pycon-2015-abstract.md	2/28/2018 4:30 AM	MD File
pycon-submission.md	2/28/2018 4:30 AM	MD File
README.md	2/28/2018 4:30 AM	MD File

Factoring Polynomials with SymPy

Here is an example that uses SymPy to factor polynomials.

In [ ]:
```
1 from ipywidgets import interact
2 from IPython.display import display
```

In [ ]:
```
1 from sympy import Symbol, Eq, factor, init_printing
2 init_printing(use_latex='mathjax')
```

In [ ]:
```
1 x = Symbol('x')
```

In [ ]:
```
1 def factorit(n):
2     display(Eq(x**n-1, factor(x**n-1)))
```

Notice how the output of the `factorit` function is properly formatted LaTeX.

In [ ]:
```
1 factorit(12)
```

In [ ]:
```
1 interact(factorit, n=(2,40));
```

In [6]:
```
1 interact(factorit, n=(2,40));
```

n 8

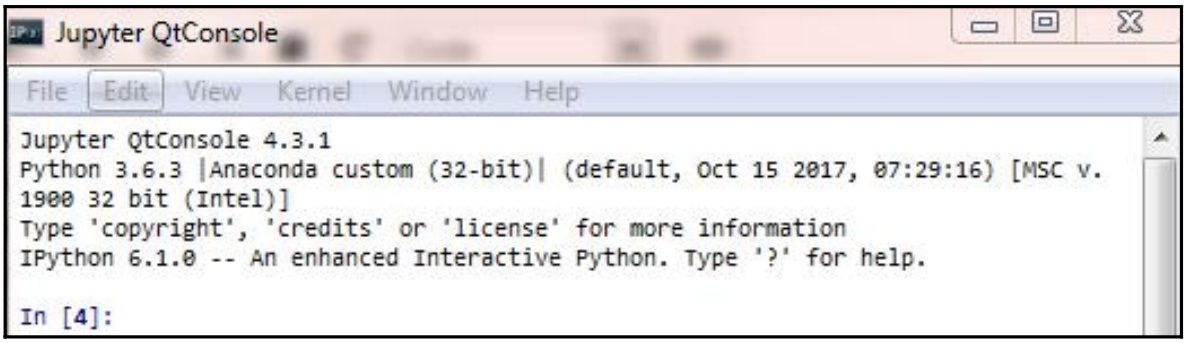
$$x^8 - 1 = (x-1)(x+1)(x^2+1)(x^4+1)$$

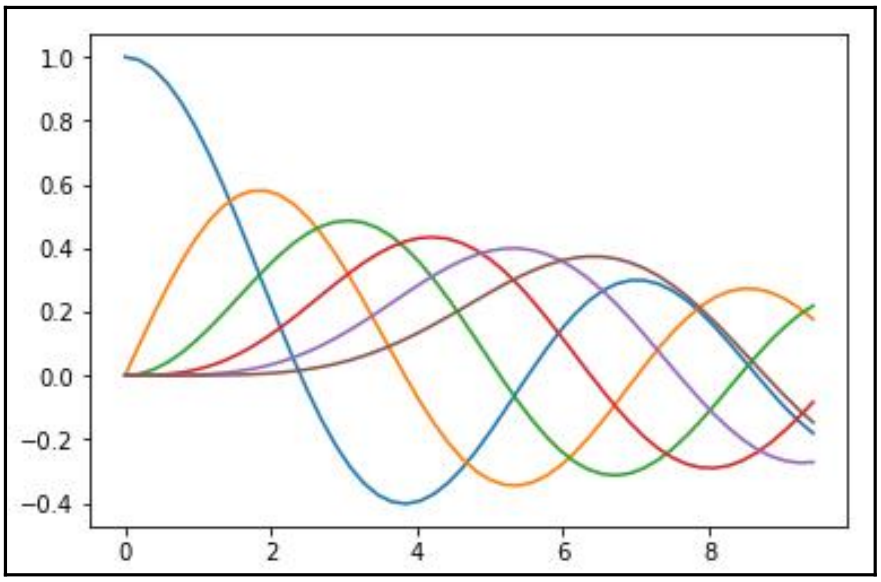
In [6]:
```
1 interact(factorit, n=(2,40));
```

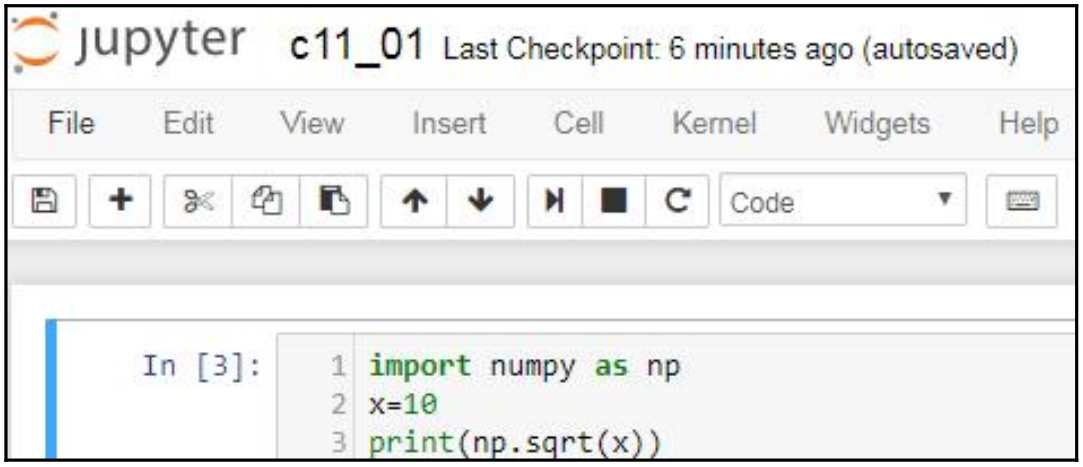
n 20

$$x^{20} - 1 = (x-1)(x+1)(x^2+1)(x^4 - x^3 + x^2 - x + 1)(x^4 + x^3 + x^2 + x + 1)(x^8 - x^6 + x^4 - x^2 + 1)$$

```
In [1]: 1 %connect_info
{
  "shell_port": 55225,
  "iopub_port": 55226,
  "stdin_port": 55227,
  "control_port": 55228,
  "hb_port": 55229,
  "ip": "127.0.0.1",
  "key": "5a2f3863-38c1ceb64da3b92b9077cc67",
  "transport": "tcp",
  "signature_scheme": "hmac-sha256",
  "kernel_name": ""
}
```

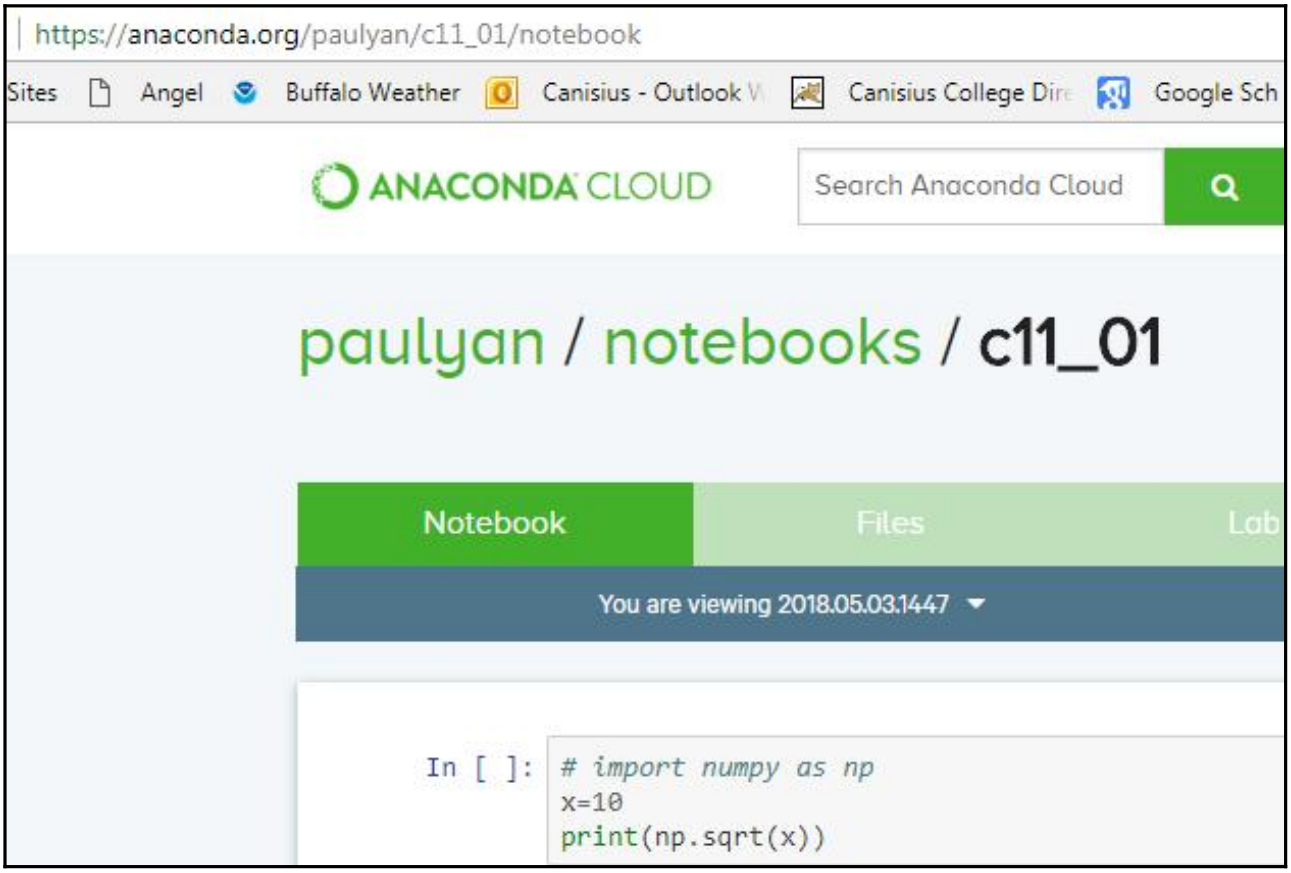




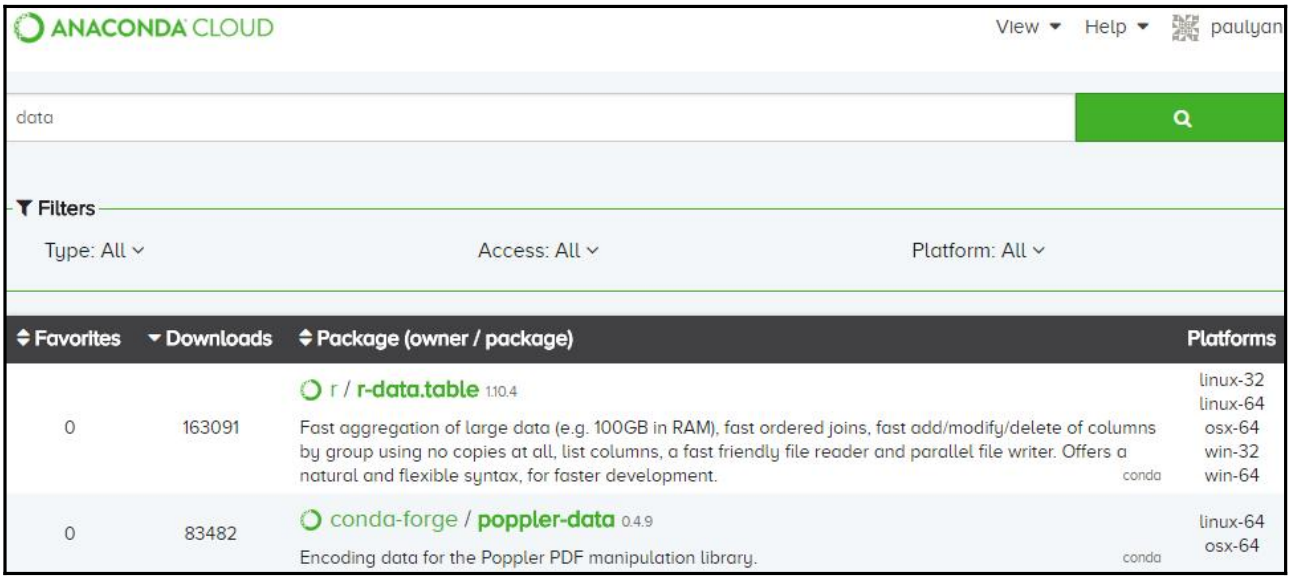


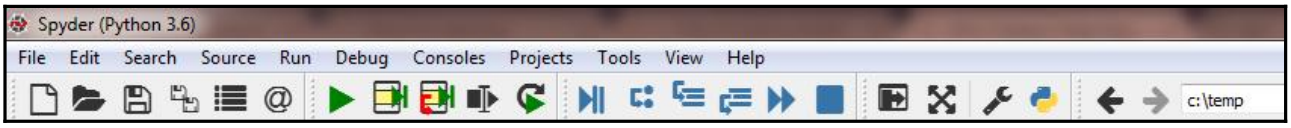

```
1 import numpy as np
2 x=10
3 print(np.sqrt(x))
```

```
(base) C:\Users\yany\Downloads>anaconda upload c11_01.ipynb
Using Anaconda API: https://api.anaconda.org
detecting package type ...
ipynb
extracting package attributes for upload ...
done

Uploading file paulyan/c11_01/2018.05.03.1447/c11_01.ipynb ...
 uploaded 4 of 4Kb: 100.00% ETA: 0.0 minutes


Upload(s) Complete

Package located at:
https://anaconda.org/paulyan/c11_01
```




```
(base) C:\Users\yany\Downloads>cd c:/temp

(base) c:\temp>anaconda-project init --directory project01
Create directory 'c:\temp\project01'? y
Project configuration is in c:\temp\project01\anaconda-project.yml
```






```
In [4]: import sys
   ...: sys.path
   ...:
Out[4]:
['',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\Scripts\\python27.zip',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\DLLs',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\plat-win',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\lib-tk',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages\\win32',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages\\win32\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages\\Pythonwin',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\win32',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\win32\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\Pythonwin',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\IPython\\extensions'
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages',
 'C:\\Users\\yany\\.ipython']
```

```
In [13]: sys.path
Out[13]:
['',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\Scripts\\python27.zip',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\DLLs',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\plat-win',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\lib-tk',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages\\win32',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages\\win32\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\User\\lib\\site-packages\\Pythonwin',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\win32',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\win32\\lib',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\Pythonwin',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages\\IPython\\extensions',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86',
 'C:\\Users\\yany\\AppData\\Local\\Enthought\\Canopy32\\App\\appdata\\canopy-1.7.4.3348.win-x86\\lib\\site-packages',
 'C:\\Users\\yany\\.ipython',
 'c:\\temp']
```

```
In [17]: import fincal
    ...: x=dir(fincal)
    ...: print(x)
    ...:
['CND', 'EBITDA_value', 'IRR_f', 'IRRs_f', 'NPER', 'PMT', 'Rc_f', 'Rm_f', '__builtins__', '__doc__', '__file__', '__name__',
'__package__', '__request', 'bondPrice', 'bsCall', 'convert_B_M', 'duration', 'exp', 'fincalHelp', 'fvAnnuity', 'fvAnnuityDue', 'fv_f',
'get_200day_moving_avg', 'get_50day_moving_avg', 'get_52week_high', 'get_52week_low', 'get_EBITDA', 'get_all', 'get_avg_daily_volume',
'get_book_value', 'get_change', 'get_dividend_per_share', 'get_dividend_yield', 'get_earnings_per_share', 'get_historical_prices',
'get_market_cap', 'get_price', 'get_price_book_ratio', 'get_price_earnings_growth_ratio', 'get_price_earnings_ratio',
'get_price_sales_ratio', 'get_short_ratio', 'get_stock_exchange', 'get_volume', 'log', 'market_cap', 'mean', 'modified_duration',
'n_annuity', 'npv_f', 'payback_', 'payback_period', 'pi', 'pvAnnuity', 'pvAnnuityDue', 'pvAnnuity_k_period_from_today',
'pvGrowingAnnuity', 'pvGrowingPerpetuity', 'pvPerpetuity', 'pvPerpetuityDue', 'pv_excel', 'pv_f', 'r_continuous', 'rateYan', 're', 'sign',
'sqrt', 'urllib']

In [18]:
```

Chapter 12: Distributed Computing, Parallel Computing, and HPCC

Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning

CRAN Task View: High-Performance and Parallel Computing with R

Maintainer: Dirk Eddelbuettel
Contact: Dirk.Eddelbuettel at R-project.org
Version: 2016-01-28

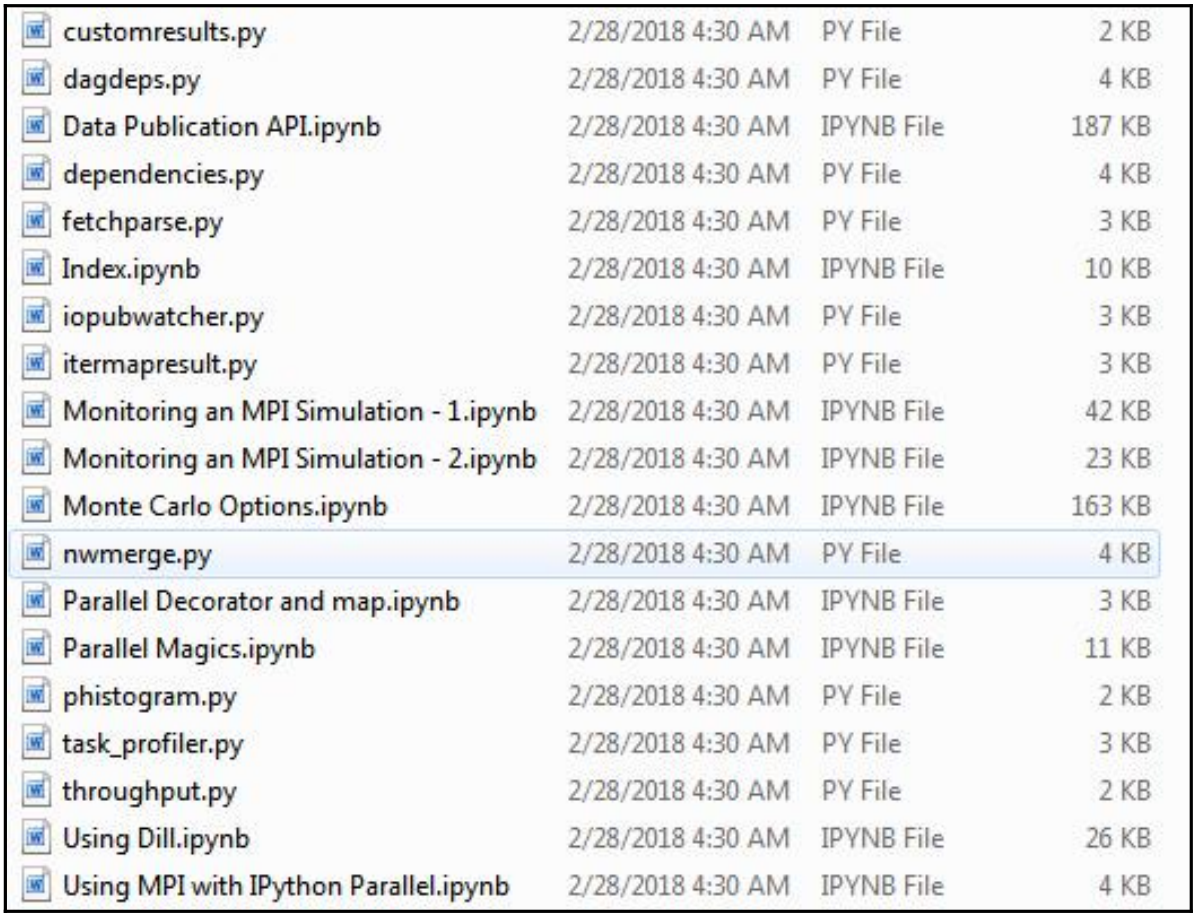
This CRAN task view contains a list of packages, grouped by topic, that are useful for high-performance computing (HPC) with R. In this context, we are defining 'high-performance computing' rather loosely as just about anything related to pushing R a little further: using compiled code, parallel computing (in both explicit and implicit modes), working with large objects as well as profiling.

Unless otherwise mentioned, all packages presented with hyperlinks are available from CRAN, the Comprehensive R Archive Network.

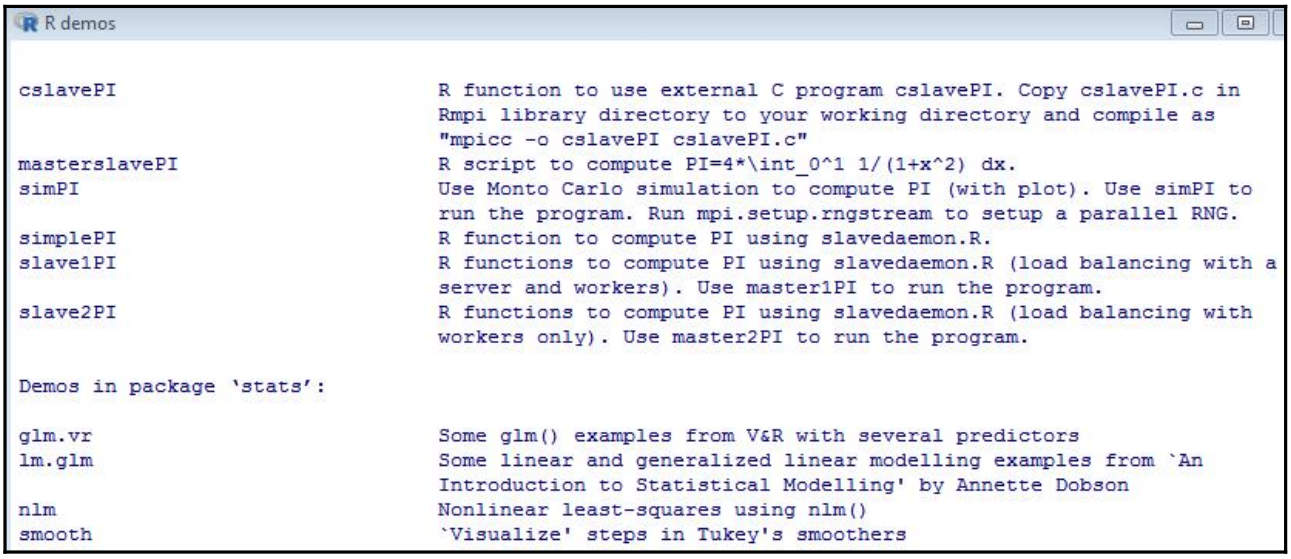
Several of the areas discussed in this Task View are undergoing rapid change. Please send suggestions for additions and extensions for this task view to the task view maintainer .

Suggestions and corrections by Achim Zeileis, Markus Schmidberger, Martin Morgan, Max Kuhn, Tomas Radivoyevitch, Jochen Knaus, Tobias Verbeke, Hao Yu, David Rosenberg, Marco Enea, Ivo Welch, Jay Emerson, Wei-Chen Chen, Bill Cleveland, Ross Boylan, Ramon Diaz-Uriarte, and Mark Zeligman (as well as others I may have forgotten to add here) are gratefully acknowledged.

daVinci Word Count	5/11/2018 1:41 PM	File folder
interengine	5/3/2018 11:10 AM	File folder
pi	5/3/2018 11:10 AM	File folder
rmt	5/3/2018 11:10 AM	File folder
wave2D	5/3/2018 11:10 AM	File folder
workflow	5/3/2018 11:10 AM	File folder

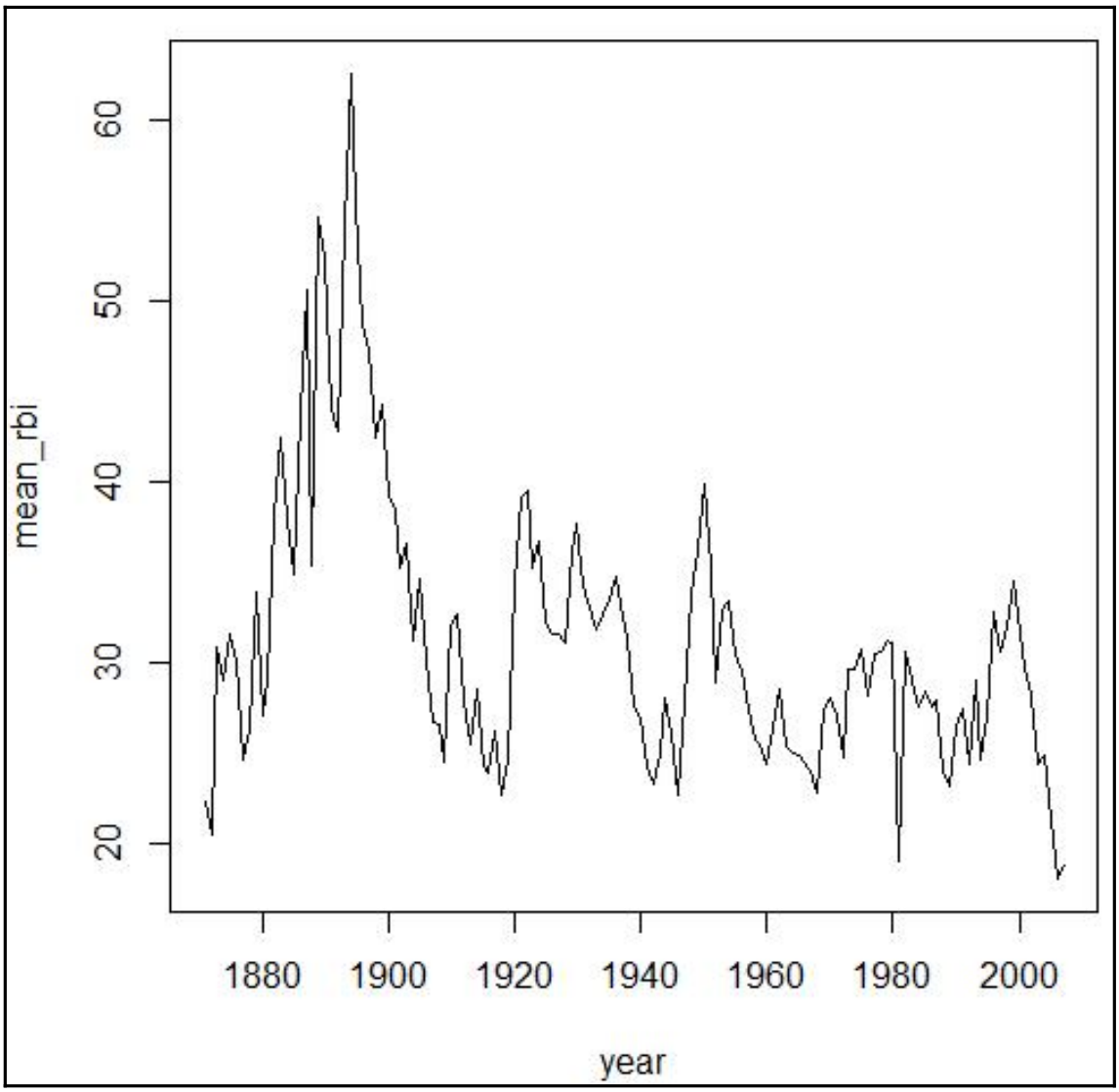


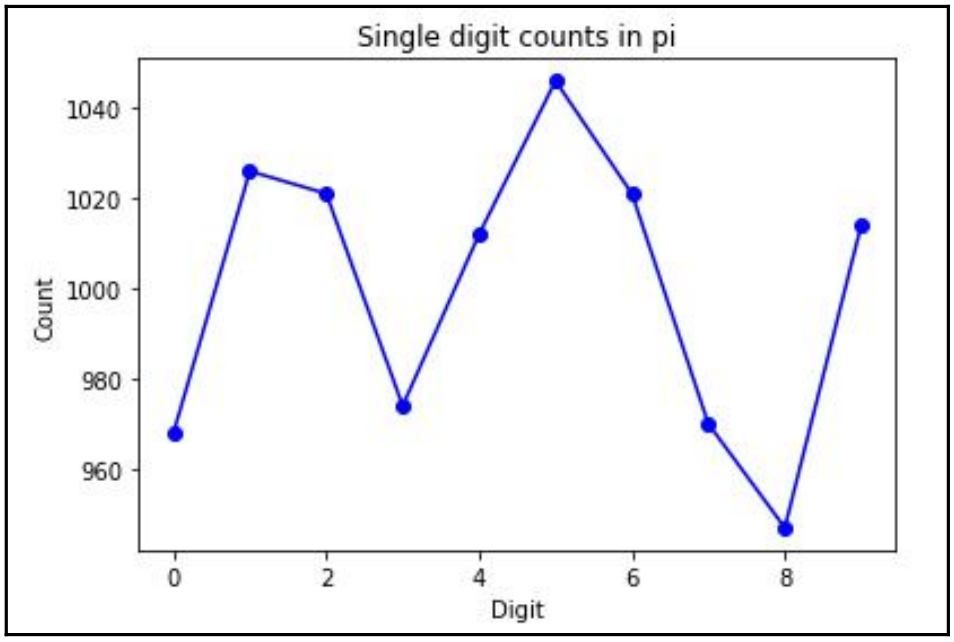
customresults.py	2/28/2018 4:30 AM	PY File	2 KB
dagdeps.py	2/28/2018 4:30 AM	PY File	4 KB
Data Publication API.ipynb	2/28/2018 4:30 AM	IPYNB File	187 KB
dependencies.py	2/28/2018 4:30 AM	PY File	4 KB
fetchparse.py	2/28/2018 4:30 AM	PY File	3 KB
Index.ipynb	2/28/2018 4:30 AM	IPYNB File	10 KB
iopubwatcher.py	2/28/2018 4:30 AM	PY File	3 KB
itermapresult.py	2/28/2018 4:30 AM	PY File	3 KB
Monitoring an MPI Simulation - 1.ipynb	2/28/2018 4:30 AM	IPYNB File	42 KB
Monitoring an MPI Simulation - 2.ipynb	2/28/2018 4:30 AM	IPYNB File	23 KB
Monte Carlo Options.ipynb	2/28/2018 4:30 AM	IPYNB File	163 KB
nwmerge.py	2/28/2018 4:30 AM	PY File	4 KB
Parallel Decorator and map.ipynb	2/28/2018 4:30 AM	IPYNB File	3 KB
Parallel Magics.ipynb	2/28/2018 4:30 AM	IPYNB File	11 KB
phistogram.py	2/28/2018 4:30 AM	PY File	2 KB
task_profiler.py	2/28/2018 4:30 AM	PY File	3 KB
throughput.py	2/28/2018 4:30 AM	PY File	2 KB
Using Dill.ipynb	2/28/2018 4:30 AM	IPYNB File	26 KB
Using MPI with IPython Parallel.ipynb	2/28/2018 4:30 AM	IPYNB File	4 KB



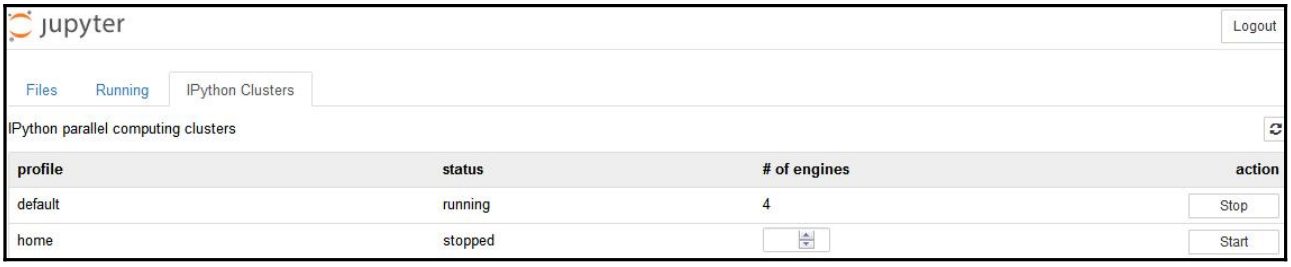

```
> output
[[1]]
 [1]    0.8414710    0.9092974    0.1411200   -0.7568025   -0.9589243
 [6]   -0.2794155    0.6569866    3.0000000    6.0000000   11.0000000
[11]   18.0000000   27.0000000   38.0000000   51.0000000    1.0000000
[16]    6.0000000    7.0000000   -2.0000000  -27.0000000  -74.0000000
[21] -149.0000000
```

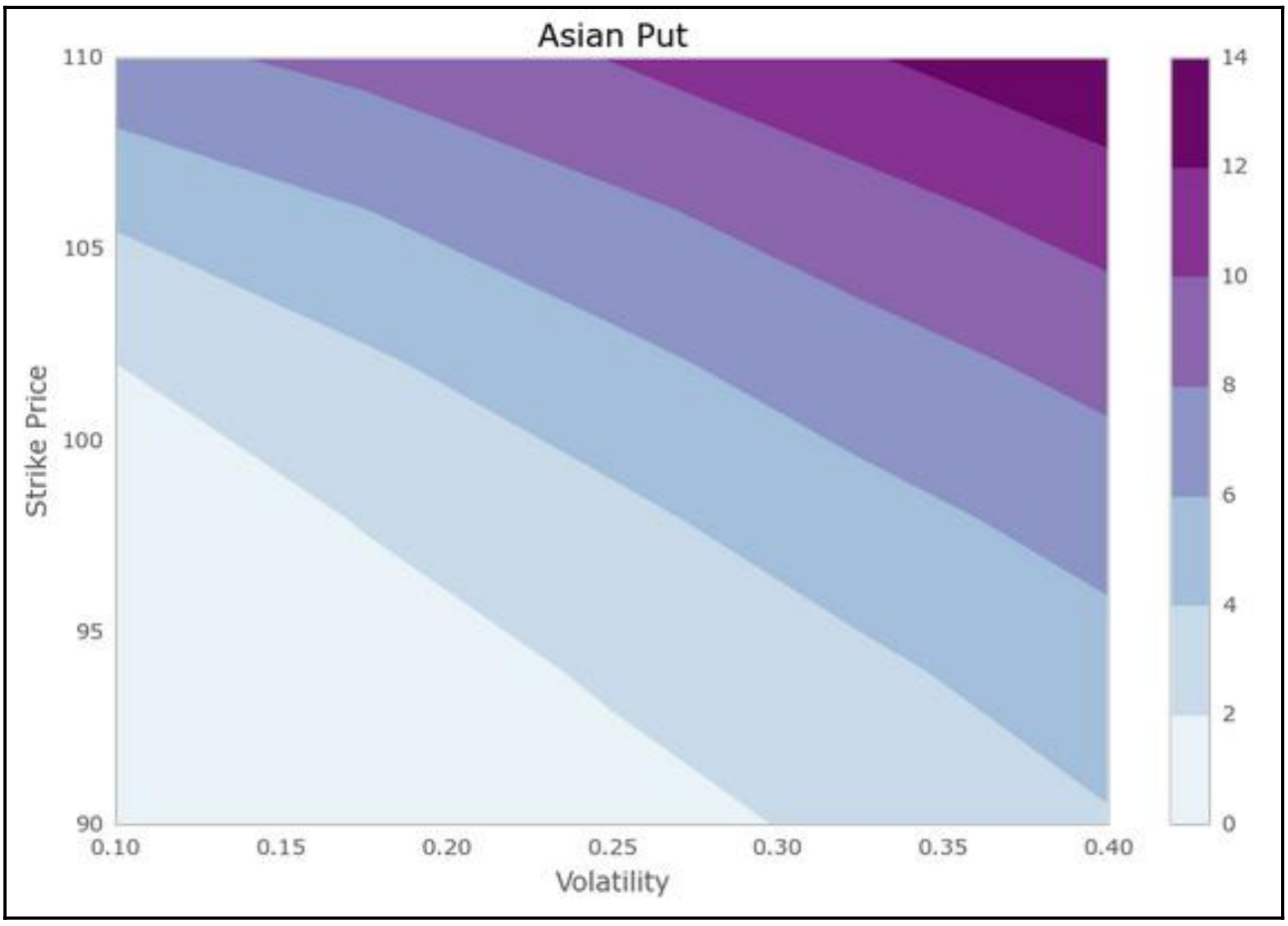
```
> demo(cslavePI)


        demo(cslavePI)
        ---- ~~~~~~~~

Type  <Return>   to start :

> cslavePI <- function (n)
+ {
+       if (mpi.comm.size(1) > 1)
+               stop ("It seems some slaves running on comm 1.")
+         mpi.comm.spawn("cslavePI")
+       mpi.intercomm.merge(2,0,1)
+         mpi.bcast(as.integer(n),1)
+         out <-mpi.reduce(0)
+         mpi.comm.free()
+         out
+ }
>
```

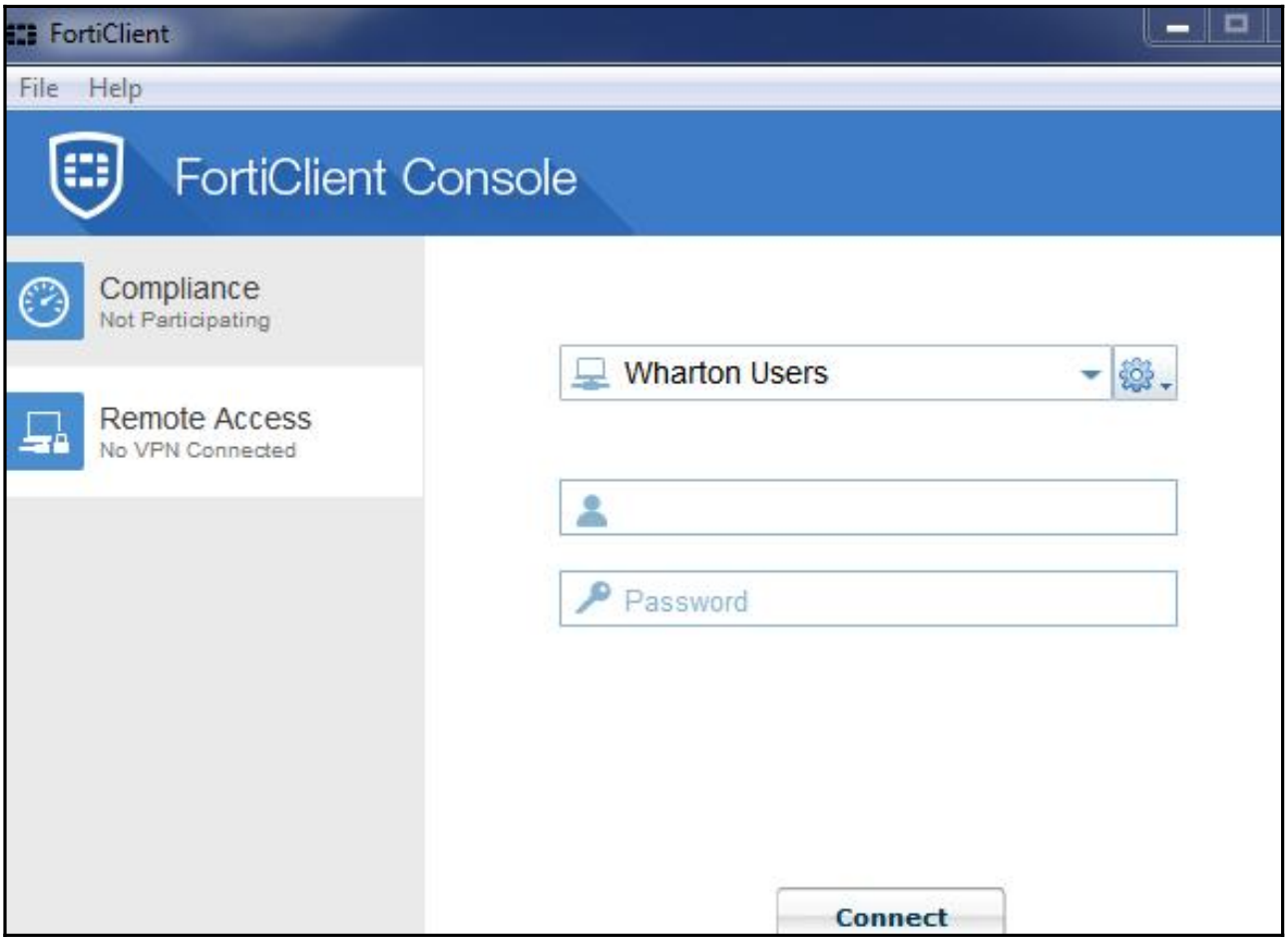

mean_rbi
20
30
40
50
60
1880
1900
1920
1940
1960
1980
2000
year

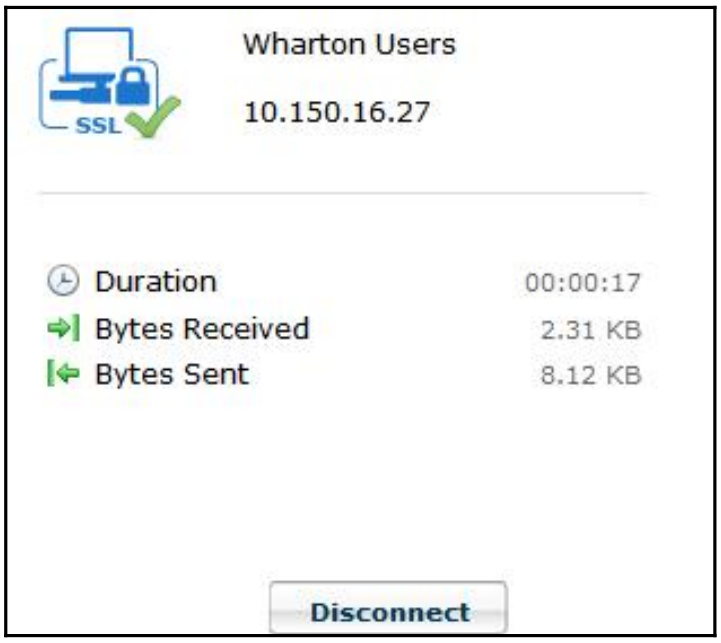

Single digit counts in pi
Count
1040
1020
1000
980
960
0
2
4
6
8
Digit


jupyter
Logout
Files
Running
IPython Clusters
IPython parallel computing clusters
profile
status
# of engines
action
default
stopped
Start
home
stopped
Start

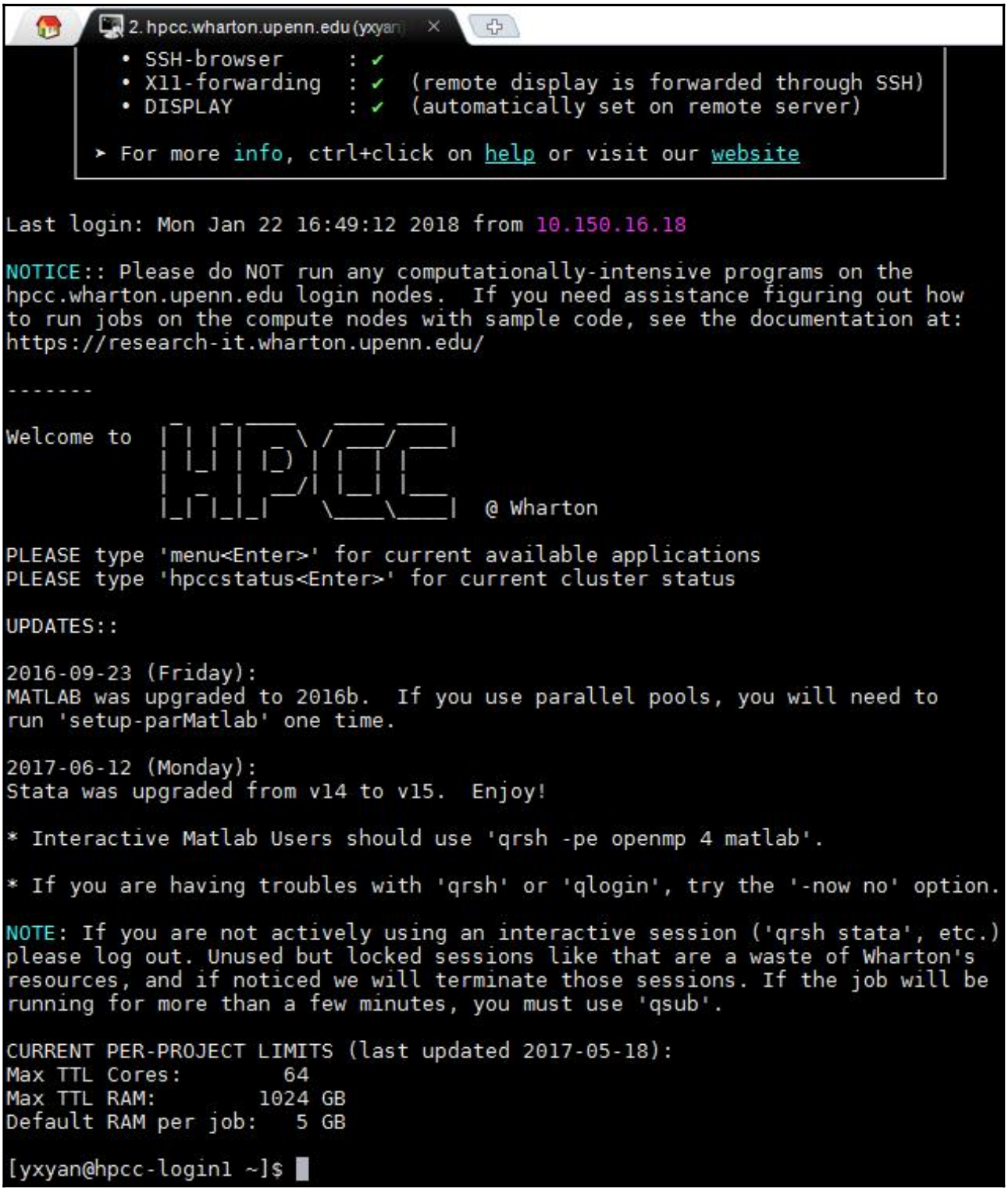

jupyter
Logout
Files
Running
IPython Clusters
IPython parallel computing clusters
profile
status
# of engines
action
default
running
4
Stop
home
stopped
Start


Asian Put
Strike Price
110
105
100
95
90
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Volatility
14
12
10
8
6
4
2
0


FortiClient
File Help
FortiClient Console
Compliance
Not Participating
Remote Access
No VPN Connected
Wharton Users
Password
Connect


Wharton Users
10.150.16.27
SSL
Duration
00:00:17
Bytes Received
2.31 KB
Bytes Sent
8.12 KB
Disconnect


mobaXterm
Start local terminal
Recover previous sessions
Find existing session or server name...


2. hpcc.wharton.upenn.edu (yxyan)
• SSH-browser : ✔
• X11-forwarding : ✔ (remote display is forwarded through SSH)
• DISPLAY : ✔ (automatically set on remote server)
➤ For more info, ctrl+click on help or visit our website
Last login: Mon Jan 22 16:49:12 2018 from 10.150.16.18
NOTICE:: Please do NOT run any computationally-intensive programs on the hpcc.wharton.upenn.edu login nodes. If you need assistance figuring out how to run jobs on the compute nodes with sample code, see the documentation at: https://research-it.wharton.upenn.edu/
Welcome to @ Wharton
PLEASE type 'menu<Enter>' for current available applications
PLEASE type 'hpccstatus<Enter>' for current cluster status
UPDATES::
2016-09-23 (Friday):
MATLAB was upgraded to 2016b. If you use parallel pools, you will need to run 'setup-parMatlab' one time.
2017-06-12 (Monday):
Stata was upgraded from v14 to v15. Enjoy!
* Interactive Matlab Users should use 'qrsh -pe openmp 4 matlab'.
* If you are having troubles with 'qrsh' or 'qlogin', try the '-now no' option.
NOTE: If you are not actively using an interactive session ('qrsh stata', etc.) please log out. Unused but locked sessions like that are a waste of Wharton's resources, and if noticed we will terminate those sessions. If the job will be running for more than a few minutes, you must use 'qsub'.
CURRENT PER-PROJECT LIMITS (last updated 2017-05-18):
Max TTL Cores: 64
Max TTL RAM: 1024 GB
Default RAM per job: 5 GB
[yxyan@hpcc-login1 ~]$