OHIA VDI Desktops

User Guide

Version 2
Revised September 22, 202

TABLE OF CONTENTS

1. Introduction

A virtual desktop is a secured and compliant environment that provides access to UCLA's healthcare data and analytical tools. In this environment one can ingest, store, access and analyze all forms of healthcare data. A user is often asked to work on a virtual desktop when they are accessing and manipulating confidential healthcare data, which they cannot store or access on their laptops/desktops.

The Office of Health Informatics (OHIA) currently offers access to 6 different types of Virtual Desktop Interfaces (VDIs) based on the role of an individual in UCLA Health.

1. **xDR CX7 Desktop:** Used by Self-Service Centers of Excellence Developers
2. **Tableau Desktop:** Used by Self-Service Centers of Excellence Tableau Developers
3. **OHIA xDR VDI Desktop:** Used by OHIA Developers and other particular use cases
4. **OHIA Report Dev Desktop:** Used by OHIA Business Intelligence Developers
5. **OHIA App Dev Desktop:** Used by OHIA Developers
6. **OHIA Power BI VDI:** Used by OHIA Power BI Developers

Depending on the specific data and applications requirements of an individual, they will be given access to a particular VDI. Pages 3 – 6 of this document list all the software applications that come with each of these VDIs. This list is up-do-date as of May 20, 2020, and will be updated regularly.

Please be aware that in order to protect the privacy and security of healthcare data, certain activities will be strictly regulated on virtual desktops:

1. Access to the internet is not allowed, unless a specific website is whitelisted by OHIA
2. Transfer of data in and out of a virtual desktop has to take place via OHIA specified tools
3. Ability to collaborate via messaging software such as Skype or Teams is limited
4. Users shall not have admin access to the virtual desktops
5. Users shall not have the ability to install software on virtual desktops. They must request OHIA staff to do install software.

2. Software & Applications on VDIs

The Office of Health Informatics (OHIA) currently offers access to 5 different types of VDIs. These virtual desktops also provide a number of programming software and analytical tools as listed below.

Software versions are included in respective columns

	xDR VDI	OHIA Report Dev VDI	OHIA App Dev VDI	Power BI VDI	Legacy xDR CX7 Desktop	Legacy Tableau Desktop
7Zip		18.05	18.05	Yes	Yes	
Adobe Acrobat		Yes	Yes		Yes	Yes
Anaconda3 / Python 3.7.3	2019.03	2019.03	2019.03	2019.03	Yes	
Anypoint Studio 6.5.2			Yes			
Character Map					Yes	
Citrix	7.15	7.15	7.15	7.15	Yes	
Cortana	Yes	Yes	Yes	Yes		
cTakes	Yes	Yes	Yes	Yes		
dfrgui					Yes	
Dotfuscator and Analytics Community Edition 5.20					Yes	
Duo Authentication for Windows Logon x64	Yes	Yes	Yes	Yes	Yes	
ECAT Agent					Yes	
Google Chrome					Yes	
Google Update Helper					Yes	
Internet Explorer					Yes	
IIS		10.0 Express	10.0 Express			
JD Edwards					Yes	
JAVA	Java8 update 201	Java8 update 201	Java8 update 201	Yes		
Jupyter					Yes	
Link Mapping to File Transfer Server	Yes	Yes	Yes	Yes		
Link mMapping to Tableau File Server	Yes	Yes	Yes			
Math Input Panel					Yes	
Microsoft .Net framework	4.5, 4.5.1, 4.5.2	4.5, 4.5.1, 4.5.2, 4.6, 4.6.1, 4.6.2	4.5, 4.5.1, 4.5.2, 4.6, 4.6.1, 4.6.2	4.5, 4.5.1, 4.5.2	Yes	
Microsoft .Net Core SDK		3.1.100	2.2.109, 3.1.100	3.1.100		
Microsoft .Net Core Version Manager (x64)			Yes			

	xDR VDI	OHIA Report Dev VDI	OHIA App Dev VDI	PowerBI VDI	Legacy xDR CX7 Desktop	Legacy Tableau Desktop
Microsoft Analysis Management Objects	Yes	Yes	Yes	Yes		
Microsoft Analyis Services	Yes	Yes	Yes	Yes	Yes	
Microsoft Azure Services					Yes	
Microsoft Azure Authoring Tools		2.9.5.2, 2.9.6	2.9.5.2, 2.9.6			
Microsoft Azure Compute Emulator		2.9.5.2, 2.9.6	2.9.5.2, 2.9.6		2.9	
Microsoft Azure HPC Scheduler					Yes	
Microsoft Azure Libraries for .NET		2.9	2.9			
Microsoft Azure Mobile App SDK		2	2			
Microsoft Azure Powershell		April 2018, September 2016	April 2018, September 2016			
Microsoft Azure Storage Emulator		2015, 2017, 2019			4.6	
Microsoft Application Error Reporting					Yes	
Microsoft Blend for Visual Studio		5.0, 5.1	2015, 2017, 2019		2015	
Microsoft Build Tools					Yes	
Microsoft Connector for Oracle Attunity	Yes	4.0, 5.0	4.0, 5.0		Yes	
Microsoft Edge	Yes	Yes	Yes	Yes		
Microsoft Expressions			Yes		Yes	
Microsoft Help Viewer	2.0, 2.2	2.0, 2.2	2.0, 2.2	2.0, 2.2	Yes	
Microsoft Office Professional (MS Excel, Word, Access, PowerPoint)	2016	2016	Yes	2016	2010	
Microsoft One Drive	Yes	Yes	Yes	Yes		
Microsoft Power BI				May, 2020		
Microsoft Report Viewer	2012	2012	2012	2012	Yes	
Microsoft Silverlight	Yes	Yes	Yes	Yes	Yes	
Microsoft SQL Server	2008, 2012, 2014, 2016, 2017	2008, 2012, 2014, 2016, 2017	2008, 2012, 2014, 2016, 2017	2008, 2012, 2014, 2016, 2017	2008 2012, 2014, 2017	
Microsoft SQL Server Data Tools	2012	2012	2012, 2015	2012	2015	
Microsoft SQL Server	Yes	Yes	Yes	Yes	Yes	

Data-Tier Application Framework						
	xDR VDI	**OHIA Report Dev VDI**	**OHIA App Dev VDI**	**PowerBI VDI**	**Legacy xDR CX7 Desktop**	**Legacy Tableau Desktop**
Microsoft SQL Server Data Tools for Visual Studio 2012	Yes	Yes	Yes			
Microsoft SQL Server Integration Services (SSIS)	Connectors v4.0, v5.0 for Oracle	Connectors v4.0,v5.0, for Oracle	Connectors v4.0,v5.0 for Oracle	Connectors v4.0 for Oracle	2012, 2014, 2016	
Microsoft SSIS – SQL Server Data Tools (Ver 10 SP1 \ .NET Ver 4.6)	Yes	2015	Yes	2015		
Microsoft SQL Server Management Studio (SSMS)	(Ver 17.9.1)	(Ver 17.9.1)	(Ver 17.9.1)	(Ver 17.9.1)	Yes	
Microsoft SQL Server Tools			17		17	
Microsoft System CLR Types for SQL Server	Yes	Yes	Yes	Yes	2012, 2014, 2016, 2017	
Microsoft Team Foundation Server					2013	
Microsoft Test Manager			2015		2015	
Model Language Pack					Yes	
Microsoft SQL Sentry			Yes			
Microsoft Visio						
Microsoft Visual C++	2008, 2010, 2013, 2015	2008, 2010,2012, 2013, 2015, 2017	2008, 2010,2012, 2013, 2015, 2017	2008, 2010, 2013, 2015	2005, 2008, 2010, 2012, 2013, 2015	
Microsoft Visual Studio	2012, 2015	2012, 2015, 2017, 2019	2012, 2015, 2017, 2019	2012, 2015	2010, 2012, 2013, 2015	
Microsoft Visual Studio Applications for Tools		2012, 2015	2012, 2015	2012, 2015		
Microsoft Visual Studio Code	Yes	Yes	Yes			
Microsoft Visual Studio Data Tools (SSRS and SSAS)		Yes	Yes			
Microsoft Web Deploy		3.6, 4.0	3.6, 4.0		Yes	
Microsoft Web Platform					Yes	

Installer						
MySQL			Yes			
MySQL Workbench			8		8	
	xDR VDI	**OHIA Report Dev VDI**	**OHIA App Dev VDI**	**PowerBI VDI**	**Legacy xDR CX7 Desktop**	**Legacy Tableau Desktop**
NMAP			7.8			
Notepad ++	Yes	Yes	Yes		Yes	
Npcap			0.9982			
Open SSL			Yes			
Oracle Client 12 Home1			Yes		Yes	
Oracle SQL Developer	Yes	Yes	Yes	Yes		Yes
					Yes	
Peoplesoft					Yes	
Postman	Yes	Yes	Yes			
PowerShell		Yes	Yes		Yes	Yes
PuTTY	0.7	0.7	0.7	0.7	Yes	
Python			Yes	Yes		
Roslyn Language Services						
Rserve - for Tableau Integration						
Rstudio	Yes	Yes	Yes	Yes	Yes	3.6.2
SAP Business View Manager		Yes				
SAP Crystal Reports		2016			2016	
SAP Crystal Reports 2016 Service Pack 5		Yes				
SAP Information Design Tool		Yes				
SAP Universe Design Tool		Yes				
SAP Universe/Information Design Tool		4.2 – 14.2.3.2132				
SAP Web Intelligence Rich Client		Yes				
SentryOne Plan Explorer			Yes			
Siebel					Yes	
Skype					2016	
SMS Client Setup Bootstrap					Yes	
Snipping Tool					Yes	
Sound Recorder					Yes	
Spyder					Yes	
Sublime Text	Build 3176	Build 3176	Build 3176	Build 3176		

Symantec Endpoint Protection		Yes	Yes		Yes	
Task UnZip for SSIS x64					Yes	
TypeScript Power Tool					Yes	
Tableau Desktop						2019.1
	xDR VDI	**OHIA Report Dev VDI**	**OHIA App Dev VDI**	**PowerBI VDI**	**Legacy xDR CX7 Desktop**	**Legacy Tableau Desktop**
TabPy - for Tableau Integration						
UpmVDAPlugin					Yes	
Visual F# 3					Yes	
Visual Studio Enterprise		2019	2019			
VMware Tools	Yes	Yes	Yes	Yes	Yes	
WCF RIA Services V SP2					Yes	
WEBI	Yes	Yes	Yes			
WinSCP			5.9.3			
Windows App Certification Kit					Yes	
Windows Calculator					Yes	
Windows Espc Resource Package					Yes	
Windows Media Player					Yes	
Windows Phone 8.0, 8.1 SDK					Yes	
Windows Runtime Intellisense Content - en-us					Yes	
Windows Software Development Kit		Yes			Yes	
Wordpad					Yes	
X11 Server	Yes					
Xming	Yes				Yes	

3. Getting Help

Having an issue with a VDI platform

If you are facing any of the VDI issues mentioned below, please send an email to the Self Service Analytics team at SelfServiceAnalytics@mednet.ucla.edu

1. Not able to launch a VDI from your laptops/desktops
2. Login trouble
3. Slowness of applications in the VDI environment
4. Stability and consistent availability of the VDI environment
5. Unable to launch application within the VDI environment

In this email please provide the following:

a) Name of the VDI
b) Brief description of the issue
c) Frist time and date of occurrence of the issue
d) How often it has been occurring

Screenshots of the error message you see when you encounter this issue

Request New Software

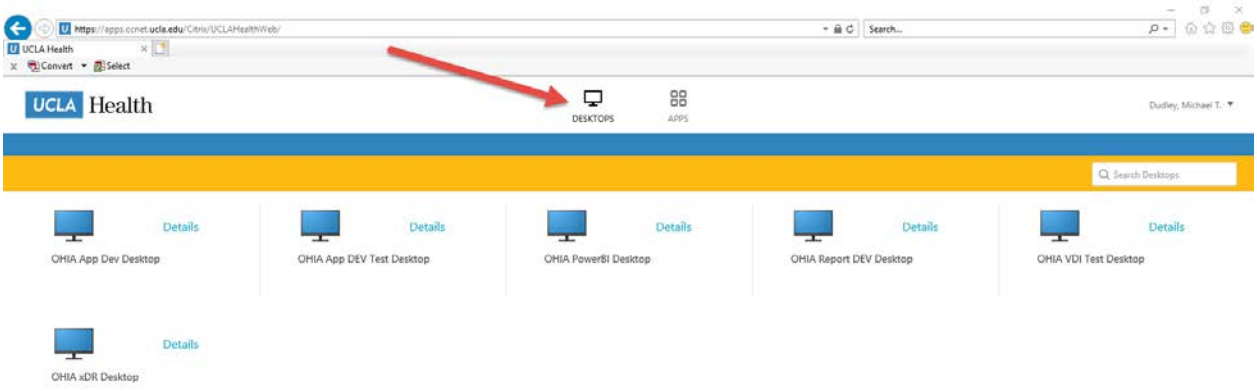
If you would like new software to be installed on a particular VDI**,** please send an email to the Self Service Analytics team at SelfServiceAnalytics@mednet.ucla.edu. In this email please provide the following:

a) Name of the VDI
b) How and where we can procure this software
c) Installation instructions for this software if you happen to have any

4. Accessing the VDI Environment

Once you are given access to a particular VDI, an icon for that VDI as shown below will be placed on your computer desktop. If you do not see it within a day of given access, please restart your computer as that may place the icon on the desktop. You can click this icon on your computer desktop to launch that VDI.

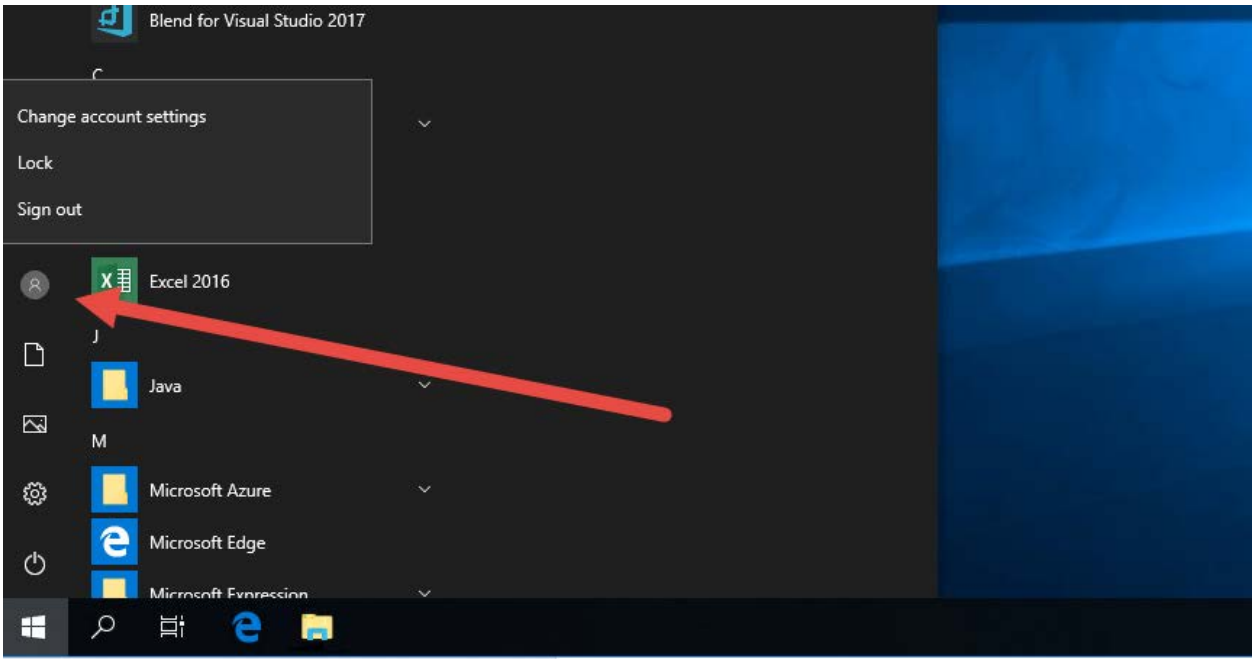
Another option is to use the Web Browser by going to the following URL, Logging in and clicking on Desktop Icon at the top, center of your screen:
https://apps.ccnet.ucla.edu/Citrix/UCLAHealthWeb/



Once you enter your username and password, you will be required to approve your DUO Authentication request.

5. Exiting the VDI environment

From the start menu, the “Power” button is disabled. You will need to use the “Sign-out” button instead. It is important to sing-out and let it fully sign-out in order for the system to change any user configurations you may have set.



6. Internet Access

Internet access from the environment is prohibited except for specific sites needing Internet access, such as WEBI. You may still use the Internet on your computer, you just cannot access the Internet using the web browser inside the VDI environment.

7. Copy and Paste

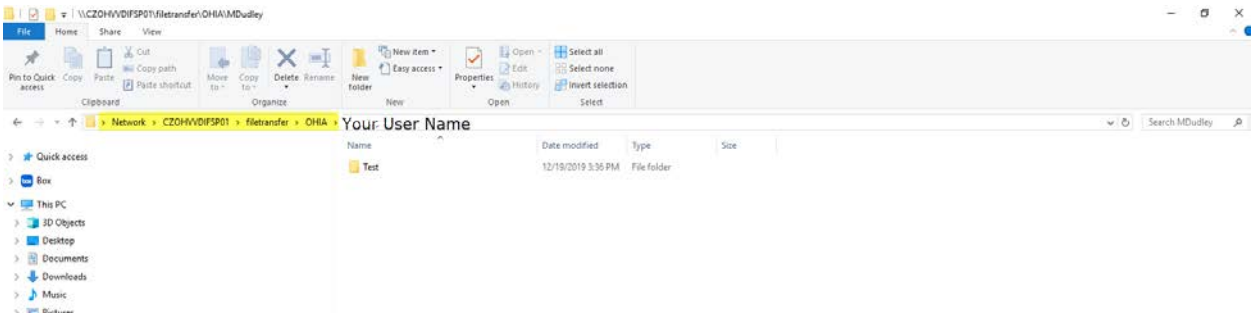
Coping and pasting from outside the environment (i.e. from your computer desktop) and inside the environment is blocked. You can copy and paste between applications within the environment, you just can't copy and paste between the environment and your computer. Any files you need to move from inside/outside the environment should be done using the Transferring Files option.

8. Dual Monitors

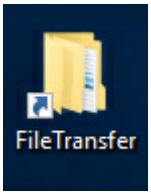
Unfortunately, Citrix doesn't currently support the ability to cast your session to dual monitors the way that you might normally do with your desktop/laptop computer.

9. Transferring Files into/out of the environment

Transferring files into or out of the environment should be done using your File Transfer folder. The folder should only be used to move files into or out of the environment. Any files left on your files transfer folder will automatically be deleted after 5 days and cannot be restored. The file transfer folder is accessible both from within the VDI environment and also from the Mednet network. All files moved through the environment are logged and may be audited by Compliance.
To access your file transfer folder from the Mednet network (i.e. from your laptop computer logged onto the network) simply use File Explorer and go to: \\CZOHVVDIFSP01\filetransfer\OHIA*yourloginname*

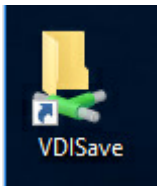


You might want to map a network drive for yourself to make it easier to access your folder. To access your file transfer folder from within the VDI environment, simply click on the File Transfer icon on the VDI desktop.

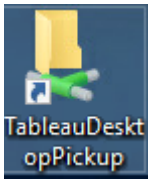


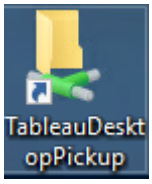
10. Where to save files within the environment

Please do not save files to the C:\ Drive or D:\ Drive in the environment. These are for configurations only. You may save files in you VDI Save folder using the icon on the VDI Desktop



If you have access to the Tableau Save and Pickup folders, these are also accessible from within the VDI environment.

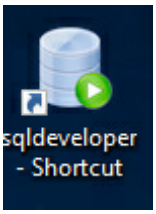




11. Oracle SQL Developer

a. Accessing Oracle SQL Developer

To access Oracle SQL Developer, only use the icon on the VDI Desktop. There is a version in the start menu, but it is an old version that will not work.



b. Configuring Oracle SQL Developer

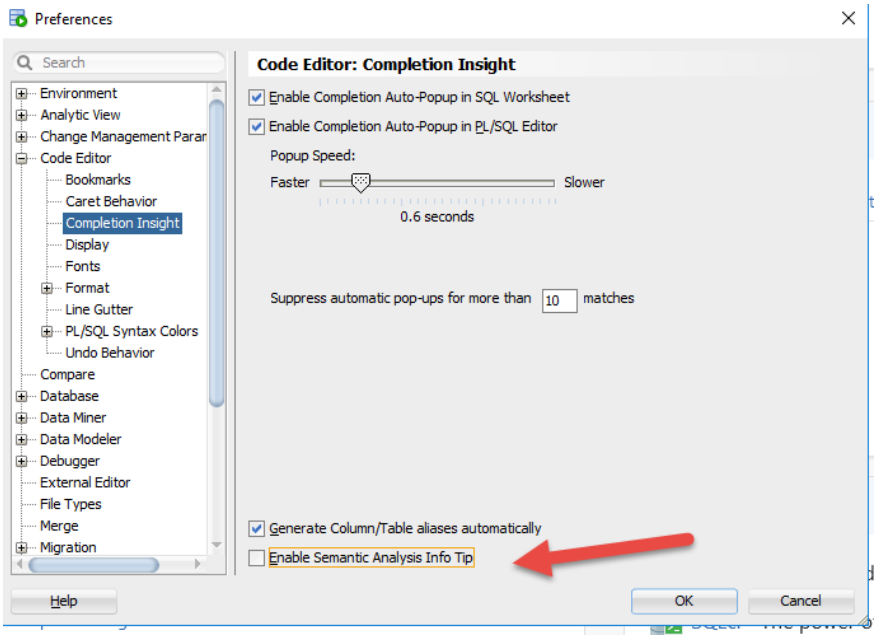
There are a few configurations that you will need to make when you first open Oracle SQL Developer. These are very important settings, so please make sure to set them up.

TNS Name File Location setting
Occasionally the tns names file is updated. To make sure that you always use the latest version you will need to set the location of the Oracle SQL Developer application looks for the tns names file.

From the Tools Menu, select Preferences. Then expand Database and click on Advanced. At the bottom of the options you will see “Tnsnames Directory” browse to:
C:\app\client32\product\12.2.0\client\network\admin
Select and click on OK to save the configuration.

Semantic Analysis Info Tip
Changing this setting is very important. The default setting in Oracle SQL Developer can cause your image to crash in certain instances.

From the Tools Menu, select Preferences. Then expand the Code Editor section and click on Completion Insight. **Uncheck** the last option “Enable semantic analysis info tip”. Then click OK to save your preferences.


Preferences
Search
Environment
Analytic View
Change Management Parar
Code Editor
Bookmarks
Caret Behavior
Completion Insight
Display
Fonts
Format
Line Gutter
PL/SQL Syntax Colors
Undo Behavior
Compare
Database
Data Miner
Data Modeler
Debugger
External Editor
File Types
Merge
Migration
Code Editor: Completion Insight
Enable Completion Auto-Popup in SQL Worksheet
Enable Completion Auto-Popup in PL/SQL Editor
Popup Speed:
Faster
Slower
0.6 seconds
Suppress automatic pop-ups for more than
10
matches
Generate Column/Table aliases automatically
Enable Semantic Analysis Info Tip
Help
OK
Cancel

12. Visual Studio 2015, 2017, and 2019

a. Required license

In order to use Visual Studio 2015 or 2017 you must have an MSDN Account with Visual Studio Professional license associated with it. If you purchase your license through Software Central, when they provide you with your license information, you will need to reply back to them to have the license associated with your MSDN account. Please remember to provide them with your MSDN account username.
Then when you open Visual Studio for the first time, remember to login using your MSDN account.

b. Azure Dev Ops

Always check-in your code before exiting Visual Studio or the environment. If you have log-in issues, your profile might be reset and you could lose your changes.

c. Source Control Drive Mapping

Due to the way that Microsoft Visual Studio saves Source Control Mapping information, it may not retain your drive mapping between sessions. You will need to re-enter your Source Control mapping each time for a while.

Basically, Microsoft stores the information relative to the machine name that you are on. When you log on to VDI you may or may not be on the same physical machine that you were on the last time. If you are, then the mapping will appear, if not, you will have to re-enter it. Once you have been on every VDI server and made your mapping, then it will show up every time you login. The Citrix VDI randomly assigns you to an available server so there is no way to insure that you will be logged onto the same server each time.

13. Appendix

a. Python Libraries

This list of libraries was last updated in May 2020

Python Library	Category	Description
Pandas	core & Statistics	Use this for data manipulation and analysis. In particular, it contains high-level data structures and tools designed for fast and easy data analysis operations. Pandas is built on NumPy and makes it easy to use in NumPy-centric applications, such as data structures with labelled axes. Explicit data alignment prevents common errors that result from misaligned data coming in from different sources.It is also easy to handle missing data using Python. Pandas is the best tool for doing data munging.
Statsmodels	core & statistics	Allows users to explore data, estimate statistical models, and perform statistical tests. An extensive list of descriptive statistics, statistical tests, plotting functions, and result statistics are available for different types of data and each estimator.
NumPy	core & statistics	An open source extension module for Python. The module NumPy provides fast precompiled functions for numerical routines. It adds support to Python for large, multi-dimensional arrays and matrices. Besides that it supplies a large library of high-level mathematical functions to operate on these arrays.
SciPy	core & statistics	SciPy is widely used in scientific and technical computing. SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering.
matplotlib	data visualization	matplotlib is a plotting library for NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like wxPython, Qt, or GTK+.

ggplot	data visualization	ggplot is based on ggplot2, an R plotting system. ggplot operates differently than matplotlib: it lets you layer components to create a complete plot. Seasoned matplotlib users might need time to adjust to this new mindset. According to the creator, ggplot isn't designed for creating highly customized graphics. It sacrifices complexity for a simpler method of plotting. ggplot is tightly integrated with pandas, so it's best to store your data in a DataFrame when using ggplot.
Seaborn	data visualization	Harnesses the power of matplotlib to create beautiful charts in a few lines of code. The key difference is Seaborn's default styles and color palettes, which are designed to be more aesthetically pleasing and modern. Since Seaborn is built on top of matplotlib, you'll need to know matplotlib to tweak Seaborn's defaults.
missingno	data visualization	Dealing with missing data is a pain. missingno allows you to quickly gauge the completeness of a dataset with a visual summary, instead of trudging through a table. You can filter and sort data based on completion or spot correlations with a heatmap or a dendrogram.
Bokeh	data visualization	Like ggplot, Bokeh is based on The Grammar of Graphics, but unlike ggplot, it's native to Python, not ported over from R. Its strength lies in the ability to create interactive, web-ready plots, which can be easily output as JSON objects, HTML documents, or interactive web applications. Bokeh also supports streaming and real-time data.
Theano	deep learning	Theano is a Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently.
scikit-learn	machine learning	An open source library for the Python. It features various classification, regression and clustering algorithms including support vector machines, logistic regression, naive Bayes, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.
XGBoost	machine learning	Convenient impletmentation of gradient boosting. These libraries provide highly optimized, scalable and fast implementations of gradient boosting, which makes them extremely popular.

Eli5	machine learning	Often the results of machine learning models predictions are not entirely clear, and this is the challenge that eli5 library helps to deal with. It is a package for visualization and debugging machine learning models and tracking the work of an algorithm step by step. It provides support for scikit-learn, XGBoost, LightGBM, lightning, and sklearn-crfsuite libraries and performs the different tasks for each of them.
Mlpy	machine learning	A machine learning library built on top of NumPy/SciPy, the GNU Scientific Library. mlpy provides a wide range of machine learning methods for supervised and unsupervised problem.mlpy is multi platform, it works with Python 2 and 3.
PyBrain	machine learning	PyBrain is an acronym for “Python-Based Reinforcement Learning, Artificial Intelligence, and Neural Network”. It is an open source library mainly used for neural networks, reinforcement learning and unsupervised learning. Neural network forms the basis for this library, making it a powerful tool for real-time analytics.
NLTK	natural language processing	The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs statistical natural language processing (NLP) for the Python. NLTK includes graphical demonstrations and sample data.NLTK has been used successfully as a platform for prototyping and building research systems.
Gensim	topic modeling	Gensim is a ython library for topic modeling. It is built on Numpy and Scipy.

b. R Libraries

This list of libraries was last updated in May 2020

ESSENTIALS		
tidyverse	An ecosystem of useful data science packages	An opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures.
readr	data import	*Contained in the tidyverse.* Base R handles most of these functions; but if you have huge files, this is a speedy and standardized way to read tabular files such as CSVs into R data frames, as well as plain text files into character strings with read_file. CRAN.
readxl	data import	*Contained in the tidyverse.* Fast way to read Excel files (.xls and .xlsx) in R, without dependencies such as Java. CRAN.
haven	data import, data wrangling	*Contained in the tidyverse.* Haven enables R to read and write various data formats used by other statistical packages by wrapping the fantastic ReadStat C library written by Evan Miller. CRAN.
xml2	data import, data wrangling	*Contained in the tidyverse.* Work with XML files using a simple, consistent interface. Built on top of the 'libxml2' C library.
httr	data import, data wrangling	*Contained in the tidyverse.* Useful tools for working with URLs and HTTP. An R interface to http protocols; useful for pulling data from APIs.
jsonlite	data import, data wrangling	Contained in the tidyverse. Parse json within R or turn R data frames into json. CRAN.
rvest	data scraping	*Contained in the tidyverse.* Easily Harvest (scrape) Web Pages
feather	data sharing	*Contained in the tidyverse.* Feather: fast, interoperable binary data frame storage for Python, R, and more powered by Apache Arrow. This binary data-file format can be read by both Python and R, making data interchange easier between the two languages. It's also built for I/O speed. CRAN.

tidyr	data wrangling	*Contained in the tidyverse.* While I still prefer the older reshape2package for some general re-arranging, tidyr won me over with specialized functions like fill (fill in missing columns from data above) and replace_na. Its main purpose is helping you change data row and column formats from "wide" to "long". CRAN.
purrr	data wrangling	Contained in the tidyverse. purrr makes it easy to apply a function to each item in a list and return results in the format of your choice. It's more complex to learn than the older plyr package, but also more robust. And, its functions are more standardized than base R's apply family -- plus it's got functions for tasks like error-checking. CRAN.
tibble	data wrangling	*Contained in the tidyverse.* Tibbles are data.frames that are lazy and surly: they do less (i.e. they don't change variable names or types, and don't do partial matching) and complain more (e.g. when a variable does not exist). This forces you to confront problems earlier, typically leading to cleaner, more expressive code. Tibbles also have an enhanced print() method which makes them easier to use with large datasets containing complex objects.
stringr	data wrangling	*Contained in the tidyverse.* Numerous functions for text manipulation. Some are similar to existing base R functions but in a more standard format, including working with regular expressions. Some of my favorites: str_pad and str_trim. CRAN.
forcats	data wrangling	*Contained in the tidyverse.* Provides a suite of useful tools that solve common problems with factors.
hms	data wrangling	*Contained in the tidyverse. Pretty time of Day (*working with times). CRAN.
lubridate	data wrangling	*Contained in the tidyverse.* Everything you ever wanted to do with date arithmetic, although understanding & using available functionality can be somewhat complex. Date-time data can be frustrating to work with in R. Lubridate makes life better. CRAN.
dplyr	data wrangling, data analysis	*Contained in the tidyverse. The* essential data-munging R package when working with data frames. Especially useful for operating on data by categories. CRAN.
modelr	data modelling	*Contained in the tidyverse.* Functions for modelling that help you seamlessly integrate modelling into a pipeline of data manipulation and visualisation.
broom	data modelling	*Contained in the tidyverse.* Takes the messy output of built-in functions in R, such as lm, nls, or t.test, and turns them into tidy data frames.

ggplot2	data visualization	*Contained in the tidyverse.* Powerful, flexible and well-thought-out dataviz package following 'grammar of graphics' syntax to create static graphics, but be prepared for a steep learning curve. CRAN.
magrittr	data wrangling	*Contained in the tidyverse.* This package gave us the %>% symbol for chaining R operations, but it's got other useful operators such as %<>% for mutating a data frame in place and and . as a placeholder for the original object being operated upon. CRAN.
PACKAGE INSTALLATION		
devtools	package development, package installation	devtools has a slew of functions aimed at helping you create your own R packages, such as automatically running all example code in your help files to make sure everything works. Requires Rtools on Windows and XCode on a Mac. On CRAN.
remotes	package installation	If you want to install R packages from GitHub, devtools was long the go-to. However, it has a ton of other functions and some hefty dependences. remotes is a lighter-weight alternative if all you want is to install packages from GitHub as well as Bitbucket and some other sources. CRAN. (ghit is another option, but is GitHub-only.)
githubinstall	package installation	Do you want to install a package from GitHub without typing out the GitHub user name along with the repo name? Whether because you can't remember a package's GitHub owner's name, that name is long/complex to type out, or you just want to save yourself a little typing, this package is a handy option. Simply run githubinstall("packagename") and the package will suggest an account; then you respond Y to install or n if it's the wrong one. It even includes fuzzy matching if you misspell a package name!
DATA IMPORT		
rio	data import, data export	rio has a good idea: Pull a lot of separate data-reading packages into one, so you just need to remember 2 functions: import and export. CRAN.

datapasta	data import	Data copy and paste: Meet reproducible research. If you've copied data from the Web, a spreadsheet, or other source into your clipboard, datapasta lets you paste it into R *as an R object, with the code to reproduce it*. It includes RStudio add-ins as well as command-line functions for transposing data, turning it into markdown format, and more. CRAN.
googleAuthR	data import	If you want to use data from a Google API in an R project and there's not yet a specific package for that API, this is the place to turn for authenticating CRAN.
fst	data import, data export	Another alternative for binary file storage (R-only), fst was built for fast storage and retrieval, with access speeds above 1 GB/sec. It also offers compression that doesn't slow data access too much, as well as the ability to import a specific range of rows (by row number). CRAN.
cloudyR project	data import, data export	This is a *collection* of packages aimed at making it easier for R to work with cloud platforms such as Amazon Web Services, Google and Travis-CI. Some are already on CRAN, some can be found on GitHub.
DATA WRANGLING		
sqldf	data wrangling, data analysis	Do you know a great SQL query you'd use if your R data frame were in a SQL database? Run SQL queries on your data frame with sqldf. CRAN.
splitstackshape	data wrangling	It's rare that I'd recommend a package that hasn't been updated in years, but the cSplit() function solves a rather complex shaping problem in an astonishingly easy way. If you have a data frame column with one *or more* comma-separated values (think a survey question with "select all that apply"), this is worth an install if you want to separate each item into its own new *data frame row.*. CRAN.
validate	data wrangling	Intuitive data validation based on rules you can define, save and re-use. CRAN.
car	data wrangling	car's recode function makes it easy to bin continuous numerical data into categories or factors. While base R's cut accomplishes the same task, I find recode's syntax to be more intuitive - just remember to put the entire recoding formula within double quotation marks. dplyr's case_when() function is another option worth considering. CRAN.

glue	data wrangling	Main function, also glue, evaluates variables and R expressions within a quoted string, as long as they're enclosed by {} braces. This makes for an elegant paste() replacement. CRAN.
scales	data wrangling	While this package has many more sophisticated ways to help you format data for graphing, it's worth a download just for the comma(), percent() and dollar() functions. CRAN.
data.table	data wrangling	Popular package for heavy-duty data wrangling. While I typically prefer dplyr, data.table has many fans for its speed with large data sets. CRAN.
janitor	data wrangling, data analysis	Basic data cleaning made easy, such as finding duplicates by multiple columns, making R-friendly column names and removing empty columns. It also has some nice tabulating tools, like adding a total row, as well as generating tables with percentages and easy crosstabs. CRAN.
diffobj	data analysis	Base R's identical() function tells you whether or not two objects are the same; but if they're not, it won't tell you why. diffobj gives you a visual representation of how two R objects differ. CRAN.
Hmisc	data wrangling, data analysis	There are a number of useful functions in here. Two of my favorites: describe, a more robust summary function, and Cs, which creates a vector of quoted character strings from unquoted comma-separated text. Cs(so, it, goes) creates c("so", "it", "goes"). CRAN.
tidytext	text mining	Elegant implementation of text mining functions using Hadley Wickham's "tidy data" principles. CRAN.
DMwR2	data mining	Functions and data accompanying the second edition of the book ``Data Mining with R, learning with case studies'' by Luis Torgo, published by CRC Press.
Matrix	computation	A rich hierarchy of matrix classes, including triangular, symmetric, and diagonal matrices, both dense and sparse and with pattern, logical and numeric entries. Numerous methods for and operations on these matrices, using 'LAPACK' and 'SuiteSparse' libraries.
DATA EXPLORATION		

zoo	data wrangling, data analysis	Robust package with a slew of functions for dealing with time series data; I like the handy rollmean function with its align=right and fill=NA options for calculating moving averages. CRAN.
gmodels	data wrangling, data analysis	There are several functions for modeling data here, but the one I use, CrossTable, simply creates cross-tabs with loads of options -- totals, proprotions and several statistical tests. CRAN.
prophet	forecasting	I don't do much forecasting analysis; but if I did, I'd start with this package. CRAN.
summarytools	data summary	Provides tools to neatly and quickly summarize data.
Amelia	missing data imputation	Performs multiple imputation for analyzing incomplete multivariate data.
questionr	data processing	Want to make the processing and analysis of surveys easier? Use the package. Interactive shiny apps and addins for data recoding, contingency tables, dataset metadata handling, and several convenience functions.
MACHINE LEARNING		
mlr	machine learning	Provides a unified interface for machine learning tasks as classification, regression, cluster analysis and survival analysis in R.
Xgboost	machine learning	Extreme Gradient Boosting, which is an efficient implementation of the gradient boosting framework from Chen & Guestrin (2016). The package includes efficient linear model solver and tree learning algorithms.
caret	machine learning	Misc functions for training and plotting classification and regression models.
gbm	machine learning	An implementation of extensions to Freund and Schapire's AdaBoost algorithm and Friedman's gradient boosting machine. Includes regression methods for least squares, absolute loss, t-distribution loss, quantile regression, logistic, multinomial logistic, Poisson, Cox proportional hazards partial likelihood, AdaBoost exponential loss, Huberized hinge loss, and Learning to Rank measures (LambdaMart).
prophet	machine learning	Implements a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects.

randomForest	machine learning	This package implements Breiman and Cutler's Random Forests for Classification and Regression
missForest	machine learning	Nonparametric Missing Value Imputation using Random Forest
e1071	machine learning	Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. Functions for latent class analysis, short time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier, ...
rpart	machine learning	Recursive partitioning for classification, regression and survival trees. An implementation of most of the functionality of the 1984 book by Breiman, Friedman, Olshen and Stone
igraph	machine learning	Routines for simple graphs and network analysis. It can handle large graphs very well and provides functions for generating random and regular graphs, graph visualization, centrality methods and much more.
nnet	machine learning	Software for feed-forward neural networks with a single hidden layer, and for multinomial log-linear models
kernlab	machine learning	Kernel-based machine learning methods for classification, regression, clustering, novelty detection, quantile regression and dimensionality reduction. Among other methods 'kernlab' includes Support Vector Machines, Spectral Clustering, Kernel PCA, Gaussian Processes and a QP solver.
glmnet	machine learning	Extremely efficient procedures for fitting the entire lasso or elastic-net regularization path for linear regression, logistic and multinomial regression models, Poisson regression and the Cox model. Two recent additions are the multiple-response Gaussian, and the grouped multinomial regression.
ROCR	machine learning	ROC graphs, sensitivity/specificity curves, lift charts, and precision/recall plots are popular examples of trade-off visualizations for specific pairs of performance measures.
pROC	machine learning	Tools for visualizing, smoothing and comparing receiver operating characteristic (ROC curves).
party	machine learning	A computational toolbox for recursive partitioning.
tree	machine learning	Classification and regression trees.

klaR	machine learning	Miscellaneous functions for classification and visualization, e.g. regularized discriminant analysis, sknn() kernel-density naive Bayes, an interface to 'svmlight' and stepclass() wrapper variable selection for supervised classification, partimat() visualization of classification rules and shardsplot() of cluster results as well as kmodes() clustering for categorical data, corclust() variable clustering, variable extraction from different variable clustering models and weight of evidence preprocessing.
ipred	machine learning	Improved predictive models by indirect classification and bagging for classification, regression and survival problems as well as resampling based estimators of prediction error.
lars	machine learning	Efficient procedures for fitting an entire lasso sequence with the cost of a single least squares fit.
earth	machine learning	Build regression models using the techniques in Friedman's papers ``Fast MARS'' and ``Multivariate Adaptive Regression Splines'' (The term ``MARS'' is trademarked and thus not used in the name of the package).
CORElearn	machine learning	A suite of machine learning algorithms written in C++ with the R interface contains several learning techniques for classification and regression.
mboost	machine learning	Functional gradient descent algorithm (boosting) for optimizing general risk functions utilizing component-wise (penalised) least squares estimates or regressiontrees as base-learners for fitting generalized linear, additive and interaction models to potentially high-dimensional data.
mlbench	machine learning	A collection of artificial and real-world machine learning benchmark problems, including, e.g., several data sets from the UCI repository.
MODELLING		
car	data analysis, modelling	advanced utilities for regression modeling
gvlma	data analysis, modelling	Performs a global validation of linear model assumptions as well separate evaluations of skewness, kurtosis, and heteroscedasticity
mass	data analysis, modelling	Functions and datasets to support Venables and Ripley, ``Modern Applied Statistics with S'' (4th edition, 2002).
lme4	data analysis, modelling	Fit linear and generalized linear mixed-effects models. Nice.

jtools	data analysis, modelling	This package's purpose is to promote efficient understanding and sharing the results of (primarily) regression analyses. There are a number of functions focused specifically on the interpretation and presentation of interactions.
DAAG	data analysis, modelling	K-fold cross-validation
bootstrap	data analysis, modelling	Software (bootstrap, cross-validation, jackknife) and data for the book ``An Introduction to the Bootstrap'' by B. Efron and R. Tibshirani, 1993, Chapman and Hall. This package is primarily provided for projects already based on it, and for support of the book. New projects should preferentially use the recommended package ``boot''.
boot	data analysis, modelling	Functions and datasets for bootstrapping from the book ``Bootstrap Methods and Their Application'' by A. C. Davison and D. V. Hinkley (1997, CUP), originally written by Angelo Canty for S.
leaps	data analysis, modelling	Regression subset selection, including exhaustive search.
relaimpo	data analysis, modelling	Relative importance of each predictor
mcmc	data analysis, modelling	Simulates continuous distributions of random vectors using Markov chain Monte Carlo (MCMC).
coda	data analysis, modelling	Provides functions for summarizing and plotting the output from Markov Chain Monte Carlo (MCMC) simulations, as well as diagnostic tests of convergence to the equilibrium distribution of the Markov chain.
MCMCPack	data analysis, modelling	Contains functions to perform Bayesian inference using posterior simulation for a number of statistical models
biglasso	data analysis, modelling	Extend lasso and elastic-net model fitting for ultrahigh-dimensional, multi-gigabyte data sets that cannot be loaded into memory. It's much more memory- and computation-efficient as compared to existing lasso-fitting packages like 'glmnet' and 'ncvreg', thus allowing for very powerful big data analysis even with an ordinary laptop.
InformationValue	data analysis, modelling	Provides companion function for analysing the performance of classification models.

pscl	data analysis, modelling	The Political Science Computational Laboratory is not just for political science people. Bayesian analysis of item-response theory (IRT) models, roll call analysis; computing highest density regions; maximum likelihood estimation of zero-inflated and hurdle models for count data; goodness-of-fit measures for GLMs; data sets used in writing and teaching at the Political Science Computational Laboratory; seats-votes curves.
bayesm	data analysis, modelling	You don't need to an econometrician to use this package which covers many important models used in marketing & econometrics. The packages covers: Multinomial Logit (MNL) and Multinomial Probit (MNP), Multivariate Probit to name a few. FYI, bayesm stands for: Bayesian Inference for Marketing/Micro-Econometrics (that's why there's an M).
BayesianTools	data analysis, modelling	General-purpose MCMC and SMC samplers, as well as plot and diagnostic functions for Bayesian statistics, with a particular focus on calibrating complex system models.
lmtest	data analysis, modelling	A collection of tests, data sets, and examples for diagnostic checking in linear regression models. Furthermore, some generic tools for inference in parametric models are provided.
sandwich	data analysis, modelling	Model-robust standard error estimators for crosssectional, time series, clustered, panel, and longitudinal data.
mvtnorm	data analysis, modelling	Computes multivariate normal and t probabilities, quantiles, random deviates and densities.
DATA VISUALIZATION		
ggExtra	data visualization	Collection of functions and layers to enhance ggplot2. You can add marginal histograms/boxplots/density plots to 'ggplot2' scatterplots.
cowplot	data visualization	Some helpful extensions and modifications to the 'ggplot2' package. In particular, this package makes it easy to combine multiple 'ggplot2' plots into one and label them with letters, e.g. A, B, C, etc.
effects	data visualization	Graphical and tabular effect displays, e.g., of interactions, for various statistical models with linear predictors.
patchwork	data visualization	Easily combine ggplot2 plots and keep the new, merged plot a ggplot2 object. plot_layout() adds ability to set columns, rows, and relative sizes of each component graphic. GitHub.

RColorBrewer	data visualization	Not a designer? RColorBrewer helps you select color palettes for your visualizations. CRAN. Note: For even more palettes, check out packages viridis for colors that print well in greyscale and are easier to read if you're color blind, pals, rcartcolor for map colors, colorr for sports-team colors, nord for "Northern-themed Color palettes," and wesanderson for color schemes used by director Web Anderson.
corrplot	data visualization	A graphical display of a correlation matrix, confidence interval. It also contains some algorithms to do matrix reordering. In addition, corrplot is good at details, including choosing color, text labels, color labels, layout, etc.
lattice	data visualization	A powerful and elegant high-level data visualization system inspired by Trellis graphics, with an emphasis on multivariate data. Lattice is sufficient for typical graphics needs, and is also flexible enough to handle most nonstandard requirements.
vcd	data visualization	Visualization techniques, data sets, summary and inference procedures aimed particularly at categorical data. Special emphasis is given to highly extensible grid graphics.
ggiraph	data visualization	Make ggplot2 plots interactive with this extension's new geom functions such geom_bar_interactive and arguments for tooltips and JavaScript onclicks. CRAN.
plotly	data visualization	R interface to the Plotly JavaScript library that was open-sourced in late 2015. Basic graphs have a distinctive look which may not be for everyone, but it's full-featured, relatively easy to learn (especially if you know ggplot2) and includes a ggplotly() function to turn graphs created with ggplot2 interactive. CRAN.
shiny	data visualization	Turn R data into interactive Web applications. I've seen some nice (if sometimes sluggish) apps and it's got many enthusiasts. CRAN.
flexdashboard	data visualization	If Shiny is too complex and involved for your needs, this package offers a simpler (if somewhat less robust) solution based on R Markdown. CRAN.

geofacet	data visualization, mapping	"geofacets" -- maps with same-sized blocks in geospatially appropriate locations. However, this package is so cool that I had to include it. Geofaceting is best understood by looking at an example. The package lets you create your own geofacet visualizations using ggplot2 and built-in grids such as US states, EU countries and San Francisco Bay Area counties. Even more impressive, it comes with design-your-own geofacet grid capabilities. CRAN.
sf	mapping, data wrangling	This package makes it much easier to do GIS work in R. Simple features protocols make geospatial data look a lot like regular data frames, while various functions allow for analysis such as determining whether points are in a polygons. A GIS game-changer for R. CRAN.
tidycensus	mapping, data wrangling	Want to analyze and map U.S. Census Bureau data from 5-year American Community Surveys or 10-year censuses? This makes it easy to download numerical and geospatial info in R-ready format. CRAN.
ggmap	mapping	Go-to for geocoding up to 2,500 addresses with the Google Maps API with its geocode and mutate_geocode functions. CRAN.
tmap & tmaptools	mapping	These package offer an easy way to read in shape files and join data files with geographic info, as well as do some exploratory mapping. Recent functionality adds support for simple features, interactive maps and creating leaflet objects. Plus, tmaptools::palette_explorer() is a great tool for picking ColorBrewer palettes. CRAN.
grid	data visualization	A rewrite of the graphics layout capabilities, plus some support for interaction.
gridBase	data visualization	Integration of base and grid graphics
gridExtra	data visualization	Miscellaneous Functions for ``Grid'' Graphics
gains	data visualization	Constructs gains tables and lift charts for prediction algorithms. Gains tables and lift charts are commonly used in direct marketing applications. The method is described in Drozdenko and Drake (2002), ``Optimal Database Marketing'', Chapter 11.
DOCUMENTATION		
rmarkdown	documentation	Convert R Markdown documents into a variety of formats.

knitr	data display	Add R to a markdown document and easily generate reports in HTML, Word and other formats. A must-have if you're interested in reproducible research and automating the journey from data analysis to report creation. CRAN.
MISC		
doParallel	computing	Parallel computing is easy to use in R thanks to packages like doParallel.
parallel		
officeR	data display	Import and edit Microsoft Word and PowerPoint documents, making it easy to add R-generated analysis and visualizations to existing as well as new reports and presentations. CRAN.
DT	data display	Create a sortable, searchable table in one line of code with this R interface to the jQuery DataTables plug-in. CRAN & GitHub rstudio/DT.
reticulate	programming	If you know Python as well as R, this package offers a suite of tools for calling Python from within R, as well as "translating" between R and Python objects such as Pandas data frames and R data frames. CRAN.
profvis	programming	Is your R code sluggish? This package gives you a visual representative of your code line by line so you can find the speed bottlenecks. CRAN.
testthat	programming	Package that makes it easy to write unit tests for your R code. CRAN.
RODBC	connectivity	ODBC Database Access
installr	misc	Windows only: Update your installed version of R from within R. On CRAN.
reinstallr	misc	Seeks to find packages that had previously been installed on your system and need to be re-installed after upgrading R. CRAN.
here	misc	This package has one function with a single, useful purpose: find your project's working directory. Surprisingly helpful if you want your code to run on more than one system. CRAN.

beepr	misc	This is pretty much pure fun. This package has one purpose: To make it easy to play notification sounds on whatever platform you are on. Yes, getting an audible notification when code finishes running or encounters an error could be useful; but here, the available sounds include options like a fanfare flourish, a Mario Brothers tune, and even a scream. CRAN.