How To: Install ${\sf R}$ and the psych package

William Revelle Department of Psychology Northwestern University

May 14, 2016

Contents

1	Ove	erview of this and related documents	2
2	Inst	all R and relevant packages	2
	2.1	Install R for the first time	3
		2.1.1 Install R	4
		2.1.2 Install relevant packages	9
	2.2	Seeing what packages are installed and active	12
3	Rea	ding in the data for analysis	12
	3.1	Find a file and read from it	12
	3.2	Copy the data from another program using the copy and paste commands	
		of your operating system	13
	3.3	Import from an SPSS or SAS file	14
4	Son	ne simple descriptive statistics before you start	15

1 Overview of this and related documents

To do basic and advanced personality and psychological research using R is not as complicated as some think. This is one of a set of "How To" to do various things using R (R Core Team, 2016), particularly using the *psych* (Revelle, 2016) package.

The current list of How To's includes:

- 1. Installing R and some useful packages (this document)
- 2. Using R and the *psych* package to find *omega_h* and ω_t .
- 3. Using R and the *psych* for factor analysis and principal components analysis.
- 4. Using the score.items function to find scale scores and scale statistics.
- 5. An overview (vignette) of the *psych* package

By following these simple guides, you soon will be able to do such things as find ω_h by issuing just three lines of code:

```
library(psych)
my.data <- read.clipboard()
omega(my.data)</pre>
```

The resulting output will be both graphical and textual.

This guide helps the naive R user to issue those three lines. Be careful, for once you start using R, you will want to do more.

2 Install R and relevant packages

To use R obviously requires installing R on your computer. This is very easy to do (see section 2.1.1) and needs to be done once.

The power of R is in the supplemental *packages*. There are at least 8,300 packages that have been contributed to the R project. To do any of the analyses discussed in these "How To's", you will need to install the package *psych* (Revelle, 2016). To do factor analyses or principal component analyses you will also need the *GPArotation* (Bernaards and Jennrich, 2005) package. With these two packages, you will be be able to find ω_h using Exploratory Factor Analysis. If you want to find to estimate ω_h using Confirmatory Factor Analysis, you will also need to add the *sem* (Fox et al., 2016) package. To use *psych* to create simulated data sets, you also need the *mnormt* (Azzalini and Genz, 2016) package. For a more complete installation of a number of psychometric packages, you can install and activate a package (ctv) that installs a large set of psychometrically relevant packages. As is true for R, you will need to install packages just once.

2.1 Install R for the first time

- 1. Download from R Cran (http://cran.r-project.org/) (see section 2.1.1)
 - Choose appropriate operating system and download compiled R
- 2. Install R (current version is 3.3.0)
- 3. Start R.

Note that the R prompt > starts off every line. This is R's way of indicating that it wants input. In addition, note that almost all commands start and finish with parentheses.

4. Add useful packages (just need to do this <u>once</u>) (see section 2.1.2)

1		i i	
	install.packages("psych",dependencies=TRUE) #the minimum requirement or		
	<pre>install.packages(c("psych", "GPArotation"), dependencies=TRUE) #required for f</pre>	actor	analysis

(a) or if you want to do CFA

<pre>install.packages(c("psych", "sem"),</pre>	dependencies=TRUE)

(b) or if you want to install the psychometric task views

install.packages("ctv") #th	is downloads the task view package
library(ctv)	#this activates the ctv package
install.views("Psychometrics")	#among others

R code

R code

- 5. Take a 5 minute break while the packages are loaded.
- 6. Activate the package(s) you want to use (e.g., psych)

R CODE	
library(psych) #Only need to make psych active once a session	

psych will automatically activate the other packages it needs, as long as they are installed. Note that psych is updated roughly quarterly, the current version is 1.6.4

7. Use R

2.1.1 Install R

First go to the Comprehensive R Archive Network (CRAN) at http://cran.r-project. org: (Figure 1)

Choose your operating system and then download and install the appropriate version

For a PC: (Figure 2)

Download and install the appropriate version – Mac, PC or Unix/Linux

Starting R on a PC.

Once you have installed R you probably will want to download and install the R Studio program. It is a very nice interface for PCs and Macs that combines four windows into one screen.

When using a PC, RStudio is very helpful. (Many like it for Macs as well).

	🔒 cran.r-project.org
	The Comprehensive R Archive Network
R	Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:
CRAN Mirrors What's new? Task Views Search	 <u>Download R for Linux</u> <u>Download R for (Mac) OS X</u> <u>Download R for Windows</u> R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.
About R	Source Code for all Platforms
R Homepage The R Journal Software R Sources R Binaries	Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!
Packages Other	 The latest release (2016-05-03, Supposedly Educational) <u>R-3.3.0.tar.gz</u>, read <u>what's new</u> in the latest version.
Documentation Manuals	• Sources of <u>R alpha and beta releases</u> (daily snapshots, created only in time periods before a planned release).
FAQs Contributed	• Daily snapshots of current patched and development versions are <u>available</u> <u>here</u> . Please read about <u>new features and bug fixes</u> before filing corresponding feature requests or bug reports.
	• Source code of older versions of R is <u>available here</u> .
	Contributed extension <u>packages</u>
	Questions About R
	• If you have questions about R like how to download and install the software, or what the license terms are, please read our <u>answers to frequently asked</u> <u>questions</u> before you send an email.
	What are R and CRAN?
provid	GNU S', a freely available language and environment for statistical computing and graphics which es a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical ime series analysis, classification, clustering, etc. Please consult the <u>R project homepage</u> for furthe nation.
	I is a network of ftp and web servers around the world that store identical, up-to-date, versions of ocumentation for R. Please use the CRAN <u>mirror</u> nearest to you to minimize network load.

5

Figure 1: The basic CRAN window allows you choose your operating system. Comprehensive R Archive Network (CRAN) is found at http://cran.r-project.org:

		🔒 cran.r-projec	ct.org	C	
			for Wind	lows	
	Subdirectories:				
	base	Binaries for base distribut want to install R for the		y Duncan Murdoch). This is what you
CRAN Mirrors	<u>contrib</u>	Binaries of contributed C Ligges). There is also info Windows services and co	ormation on third	d party software ava	ailable for CRAN
<u>What's new?</u> <u>Task Views</u> Search	old contrib	Binaries of contributed C 2.11.x; managed by Uwe	RAN packages f		
About R	<u>Rtools</u>	Tools to build R and R pa you want to build your ov			
<u>R Homepage</u> The R Journal	Please do not sub	mit binaries to CRAN. Pack case of questions / suggesti	cage developers i	might want to conta	act Duncan Murdoch or Uwe
Software R Sources		nt to read the <u>R FAQ</u> and <u>R</u>			
<u>R Binaries</u> Packages Other		s some checks on these bina downloaded executables.	aries for viruses,	but cannot give gua	arantees. Use the normal
Documentation Manuals					
FAQs Contributed					

Figure 2: On a PC you want to choose the base system ${6 \atop 6}$

	R-3.3.0 for Windows (32/64 bit)
	Download R 3.3.0 for Windows (62 megabytes, 32/64 bit) Installation and other instructions New features in this version
Mirrors What's new? Task Views Search	If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the $md5sum$ of the .exe to the <u>true</u> <u>fingerprint</u> . You will need a version of md5sum for windows: both <u>graphical</u> and <u>command line versions</u> are available.
About R <u>R Homepage</u> <u>The R Journal</u>	Frequently asked questions Does R run under my version of Windows?
Software R Sources	 <u>How do I update packages in my previous version of R?</u> <u>Should I run 32-bit or 64-bit R?</u>
<u>R Binaries</u> <u>Packages</u> <u>Other</u>	Please see the <u>R FAQ</u> for general information about R and the <u>R Windows FAQ</u> for Windows-specific information.
Documentation	Other builds
<u>Manuals</u> <u>FAQs</u> <u>Contributed</u>	 Patches to this release are incorporated in the <u>r-patched snapshot build</u>. A build of the development version (which will eventually become the next major release of R) is available in the <u>r-devel snapshot build</u>. <u>Previous releases</u>
	Note to webmasters: A stable link which will redirect to the current Windows binary release is < <u>CRAN MIRROR>/bin/windows/base/release.htm</u> .
	Last change: 2016-05-04, by Duncan Murdoch

Figure 3: Download the Windows version

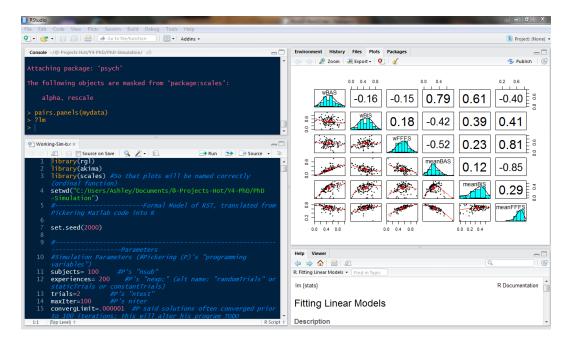


Figure 4: Using R Studio on a PC.

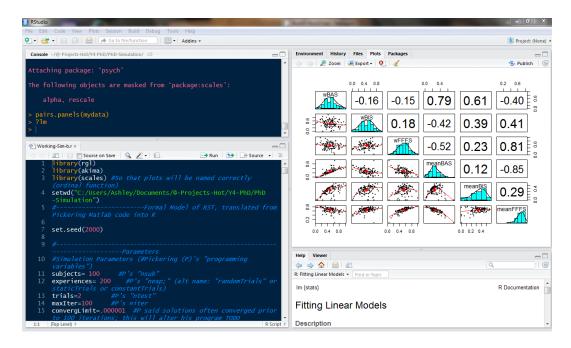


Figure 5: Using R Studio on a PC.

Start up R and get ready to play (Mac version).

```
R version 3.3.0 (2016-05-03) -- "Supposedly Educational"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
 Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[R.app GUI 1.68 (7202) x86_64-apple-darwin13.4.0]
[Workspace restored from /Users/revelle/.RData]
[History restored from /Users/revelle/.Rapp.history]
```

>

2.1.2 Install relevant packages

Once R is installed on your machine, you still need to install a few relevant "packages". Packages are what make R so powerful, for they are special sets of functions that are designed for one particular application. In the case of the *psych* package, this is an application for doing the kind of basic data analysis and psychometric analysis that psychologists and many others find particularly useful.

You may either install the minimum set of packages necessary to do the analysis using an Exploratory Factor Analysis (EFA) approach (recommended) or a few more packages to do both an EFA and a CFA approach. It is also possible to add many psychometrically relevant packages all at once by using the "task views" approach. A particularly powerful package is the *lavaan* (Rosseel, 2012) package for doing structural equation modeling. Another useful one is the *sem* package (Fox et al., 2016).

Install the minimum set This may be done by typing into the console or using menu options (e.g., the Package Installer under<u>neath the Packages and Data menu</u>).

	R code	
install.packages("psych", depende	encies = TRUE)}	

		cran.r-project.org	Ċ	
		R for Mac	OS X	
R	This directory contains binaries for above). Mac OS 8.6 to 9.2 (and M release of R for these systems (wh X 10.5) and PowerPC Macs can be	ac OS X 10.1) are no longich is R 1.7.1) <u>here</u> . Relea	ger supported but ye uses for old Mac OS	ou can find the last supported
<i>CRAN</i> <u>Mirrors</u> What's new?	Note: CRAN does not have Mac Oprecautions when assembling bina			
Task Views Search	As of 2016/03/01 package binaries archive so users of such versions s			
About R	R 3.3.0 "Suppo	sedly Educational	" released on	n 2016/05/03
<u>R Homepage</u> The R Journal	Please check the MD5 checksum of corrupted during the mirroring pro-		to ensure that it has	s not been tampered with or
Software <u>R Sources</u> <u>R Binaries</u> Packages	md5 R-3.3.0.pkg in the <i>Terminal</i> application to prin later you can also validate the sign pkgutilcheck-signature R-	nature using	he R-3.3.0.pkg ima	age. On Mac OS X 10.7 and
Other		Files:		
Documentation <u>Manuals</u> <u>FAQs</u> Contributed	R-3.3.0.pkg MD5-hash: 871f274d0f9a12731d328125774c639a SHA1- hash: a0ae68233cb826d228d7791e994fc634a888cc48 (ca. 71MB)	R 3.3.0 binary for Mac C signed package. Contain in 64-bit for Intel Macs, 5.2. The latter two compommitted when choosing you want to use the tclt documentation from sou	s R 3.3.0 framewor Tcl/Tk 8.6.0 X11 lii onents are optional g "custom install", i k R package or bui	k, R.app GUI 1.68 braries and Texinfo and can be t is only needed if
		Note: the use of X11 (ind installed since it is no low XQuartz when upgrading	iger part of OS X.	Always re-install
	R-3.2.1-snowleopard.pkg MD5-hash: 58fe9d01314d9eb75ff80ccfb914fd65 SHA1- hash: be6e91db12bac22a324f0cb51c7efa9063ece0d0 (ca. 68MB)	R 3.2.1 legacy binary for 10.8 (Mountain Lion), si framework, R.app GUI 1 This package contains th Tcl/Tk 8.6.0 X11 librarie NOT included (needed i sources that contain FOF <u>directory</u> . NOTE: the binary suppo phased out, we do not ex	gned package. Con .66 in 64-bit for Im e R framework, 64- s and Texinfop 5.2. f you want to comp (TRAN code) pleas rt for OS X before 1	tains R 3.2.1 tel Macs. -bit GUI (R.app), . GNU Fortran is tile packages from the see <u>the tools</u> Mavericks is being

Figure 6: For the Mac, you want to choose the latest version which includes the GUI as well as the 32 and 64 bit versions. 10

0• 🕣 • 🔒 🔒 🚔	Go to file/function							🛝 Projec	ct: (None)
Console ~/@·Projects·Hot/Y4-	PhD/PhD-Simulation/ 🔗		Enviro	nment History	Files	Plots Packages			
		*	OL In	stall 🕜 Update			Q		6
	'psych'		1	Name	D	escription		Version	
				n Library Ikima	1-	terpolation of Irregularly and Regul	and Connered Dates	0.5-12	
alpha, rescale				poot		ootstrap Functions (Originally by Ar		1.3-18	0
	Install Packages			lass		unctions for Classification	igelo canty for 5)	7.3-14	0
pairs.panels(myd	Install Packages					inding Groups in Data": Cluster An	1.1.5.1.1.1	2.0.4	
· ?1m	Install from: Configuring Repositories	-		luster		ousseeuw et al.	alysis Extended	2.0.4	8
•	Repository (CRAN, CRANextra)	-		odetools	C	ode Analysis Tools for R		0.2-14	0
				olorspace	C	olor Space Manipulation		1.2-6	0
Working-Sim-b.r ×	Packages (separate multiple with space or comma):			ompiler	TI	he R Compiler Package		3.3.0	Θ
🗅 🖒 🙇 📄 🖸 Sour	psych	ource 🔹 🖻		latasets	TI	he R Datasets Package		3.3.0	0
1 library(rgl) 2 library(akir	Install to Library:	-		dichromat	C	olor Schemes for Dichromats		2.0-0	0
3 library(sca (ordinal fu	C:/Program Files/R/R-3.3.0/library [Default]		E f	oreign		ead Data Stored by Minitab, S, SAS, Base,	SPSS, Stata, Systat, Weka	, 0.8-66	0
4 setwd("C:/U:	Install dependencies	/PhD	<u> </u>	graphics	TI	he R Graphics Package		3.3.0	0
-Simulation			<u> </u>	grDevices	TI	he R Graphics Devices and Support	for Colours and Fonts	3.3.0	Θ
5 #	Install Cancel	ed from		grid	TI	he Grid Graphics Package		3.3.0	Θ
Pickering Ma				KernSmooth		unctions for Kernel Smoothing Sup 995)	oorting Wand & Jones	2.23-15	٢
7 set.seed(200				abeling	A	xis Labeling		0.3	0
				attice	Tr	rellis Graphics for R		0.20-33	0
	Parameters		_						-
			Help	Viewer					
variables")			()	🔶 🏠 i 🚔 i 🖉			Q,		
11 subjects= 10 12 experiences=		inle" on	R: Fitti	ng Linear Models 👻	Find in	Topic			
		lais or	lm {s	tats}				R Document	tation
13 trials=2									
14 maxIter=100			Fitt	ing Linea	r Mo	dels			
15 convergLimit		ed prior							
1:1 (Top Level) \$	tions; this will alter his program TODO	R Script ÷		cription					

Figure 7: Installing packages using R studio on a PC. Use the install menu option.

Install a few more packages If you want some more functionality for some of the more advanced statistical procedures (e.g., omegaSem) you will need to install a few more packages (e.g., sem.

<pre>install.packages(c("psych", "GPArotation", "sem"), dependencies=TRUE)</pre>

Install a "task view" to get lots of packages If you know that there are a number of packages that you want to use, it is possible they are listed as a "task view". For instance, about 50 packages will be installed at once if you install the "psychometrics" task view. You can Install all the psychometric packages from the "psychometrics" task view by first installing a package ("ctv") that in turn installs many different task views. To see the list of possible task views, go to https://cran.r-project.org/web/views/. _ R code

<pre>install.packages("ctv") }</pre>	#this downloads the task view package
library(ctv)	<pre>#this activates the ctv package</pre>
<pre>install.views("Psychometrics")</pre>	#one of the many Taskviews

Take a 5 minute break because you will be installing about 50 packages.

Ē

Make the *psych* package active. You are almost ready. But first, to use most of the following examples you need to make the *psych* package active. You only need to do this once per session.

	R code	·
library (psych)		

(If you want to automate this last step, you can create a special command to be run every time you start R.

	R code	
<pre>.First <- function() {library(psych)}</pre>		

Do this when you first start R. Then quit with the save option. Then restart R. You will now automatically have loaded the *psych* package every time you start R.)

2.2 Seeing what packages are installed and active

To see what packages are active (as a way of telling which version of R you have, and which version of relevant packages are loaded):

sessionInfo()

```
> sessionInfo()
R Under development (unstable) (2016-05-10 r70594)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.11.4 (El Capitan)
locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
attached base packages:
[1] stats
             graphics grDevices utils
                                            datasets methods
                                                                base
other attached packages:
[1] psych_1.6.4
loaded via a namespace (and not attached):
[1] tools_3.4.0
                  parallel_3.4.0 mnormt_1.5-4
>
```

3 Reading in the data for analysis

3.1 Find a file and read from it

There are of course many ways to enter data into R. Reading from a local file using read.table is perhaps the most preferred. You first need to find the file and then read it. This can be done with the file.choose and read.table functions:

```
      R code

      file.name <- file.choose() #note the open and closing parentheses</td>

      my.data <- read.table(file.name)</td>
```

file.choose opens a search window on your system just like any open file command does. It doesn't actually read the file, it just finds the file. The read.table command is also necessary. It assumes that the first row of your table has labels for each column. If this is not true, specify names=FALSE, e.g.,

R code

```
file.name <- file.choose()
my.data <- read.table(file.name, names = FALSE)</pre>
```

3.2 Copy the data from another program using the copy and paste commands of your operating system

However, many users will enter their data in a text editor or spreadsheet program and then want to copy and paste into R. This may be done by using read.table and specifying the input file as "clipboard" (PCs) or "pipe(pbpaste)" (Macs). Alternatively, the read.clipboard set of functions are perhaps more user friendly:

read.clipboard is the base function for reading data from the clipboard.

- read.clipboard.csv for reading text that is comma delimited.
- read.clipboard.tab for reading text that is tab delimited (e.g., copied directly from an Excel file).
- read.clipboard.lower for reading input of a lower triangular matrix with or without a diagonal. The resulting object is a square matrix.

read.clipboard.upper for reading input of an upper triangular matrix.

read.clipboard.fwf for reading in fixed width fields (some very old data sets)

For example, given a data set copied to the clipboard from a spreadsheet, just enter the command

This will work if every data field has a value and even missing data are given some values (e.g., NA or -999). If the data were entered in a spreadsheet and the missing values were just empty cells, then the data should be read in as a tab delimited or by using the read.clipboard.tab function.

	R code
<pre>my.data <- read.clipboard(sep="\t")</pre>	#define the tab option, or
<pre>my.tab.data <- read.clipboard.tab()</pre>	#just use the alternative function

For the case of data in fixed width fields (some old data sets tend to have this format), copy to the clipboard and then specify the width of each field (in the example below, the first variable is 5 columns, the second is 2 columns, the next 5 are 1 column the last 4 are 3 columns).

	R code
<pre>my.data <- read.clipboard.fwf(widths=c()</pre>	<u>5,2,rep(</u> 1,5),rep(3,4))

3.3 Import from an SPSS or SAS file

To read data from an SPSS, SAS, or Systat file, you must use the *foreign* package. This should come with Base R need to be loaded using the library command.

read.spss reads a file stored by the SPSS save or export commands.

The **read.spss** function has many parameters that need to be set. In the example, I have used the parameters that I think are most useful.

file Character string: the name of the file or URL to read.

use.value.labels Convert variables with value labels into R factors with those levels?

- to.data.frame return a data frame? Defaults to FALSE, probably should be TRUE in most cases.
- **max.value.labels** Only variables with value labels and at most this many unique values will be converted to factors if use.value.labels = TRUE.
- trim.factor.names Logical: trim trailing spaces from factor levels?
- trim_values logical: should values and value labels have trailing spaces ignored when matching for use.value.labels = TRUE?
- **use.missings** logical: should information on user-defined missing values be used to set the corresponding values to NA?

The following is an example of reading from a remote SPSS file and then describing the data set to make sure that it looks ok (with thanks to Eli Finkel).

library(: datafile eli <- re	name	<-	"http		-	y-project						
describe	(eli	, ske	ew=FAL	SE)								
	var	n	mean	sd	median	trimmed	mad	min	max	range	se	
USER*	1	69	35.00	20.06	35	35.00	25.20	1	69	68	2.42	
HAPPY	2	69	5.71	1.04	6	5.82	0.00	2	7	5	0.13	
SOULMATE	3	69	5.09	1.80	5	5.32	1.48	1	7	6	0.22	

6.70 0.00

0.39 0.00 0 1

7

2

5 0.12

1 0.06

4 Some simple descriptive statistics before you start

7

0

ENJOYDEX

UPSET

4 68 6.47 1.01

5 69 0.41 0.49

Although you probably want to jump right in and do a factor analysis or find $\boldsymbol{\omega}$, you should first make sure that your data are reasonable. Use the **describe** function to get some basic descriptive statistics. This next example takes advantage of a built in data set (sat.act) in the *psych* package.

<pre>my.data <- sat.act #built in example replace with your data pairs.panels(sat.act,pch='.') describe(my.data)</pre>													
	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
gender	1	700	1.65	0.48	2	1.68	0.00	1	2	1	-0.61	-1.62	0.02
education	2	700	3.16	1.43	3	3.31	1.48	0	5	5	-0.68	-0.07	0.05
age	3	700	25.59	9.50	22	23.86	5.93	13	65	52	1.64	2.42	0.36
ACT	4	700	28.55	4.82	29	28.84	4.45	3	36	33	-0.66	0.53	0.18
SATV	5	700	612.23	112.90	620	619.45	118.61	200	800	600	-0.64	0.33	4.27
SATQ	6	687	610.22	115.64	620	617.25	118.61	200	800	600	-0.59	-0.02	4.41

In addition to simple descriptives, it is always helpful to graphically examine your data using the pairs.panels function.

There are, of course, all kinds of things you could do with your data at this point, but read about them in the vignette for the *psych* package http://cran.r-project.org/web/packages/psych/vignettes/overview.pdf.

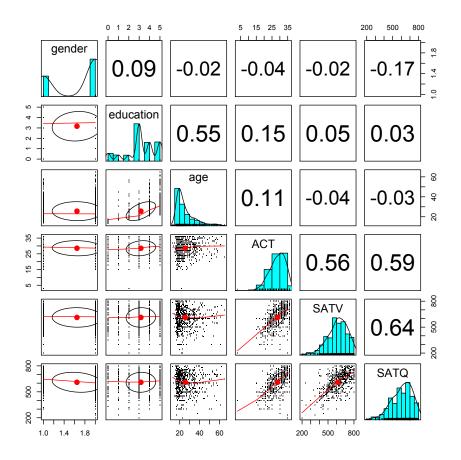


Figure 8: A scatter plot matrix (splom) of a data set is a a powerful way to examine a data set. Elements on the diagonal show the histograms and densities of the data, lower off diagonal elements are the pairwise scatter plots, upper off diagonal elements are the pairwise correlations.

References

- Azzalini, A. and Genz, A. (2016). The R package mnormt: The multivariate normal and t distributions (version 1.5-4).
- Bernaards, C. and Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65(5):676–696.
- Fox, J., Nie, Z., and Byrnes, J. (2016). sem: Structural Equation Models. R package version 3.1-7.
- R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W. (2016). psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, http://cran.r-project.org/web/packages/psych/. R package version 1.6.4.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36.