

MultiMorbidity Tool

Globaal is dit document opgebouwd uit de MMT handleiding en een aantal bijlagen. De bijlagen betreffen:

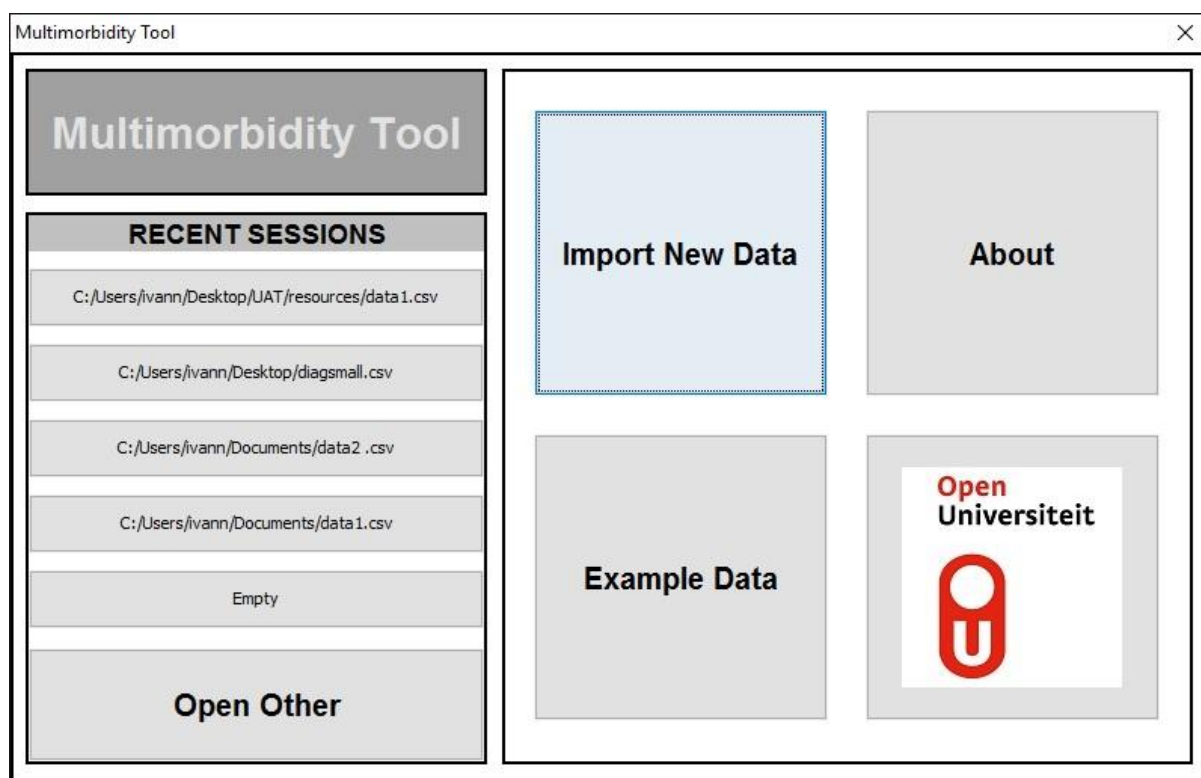
- Bijlage 1: MMT installatiehandleiding
- Bijlage 2: R installatiehandleiding
- Bijlage 3: JRI installatiehandleiding
- Bijlage 4: Korte R handleiding

MMT Handleiding

Getting Started

De MMT handleiding bestaat uit drie onderdelen: het startscherm, het hoofdscherm en de afzonderlijke views.

1 Het Startscherm



Afbeelding 1. Het startscherm.

Als je de applicatie opstart verschijnt er eerst een startscherm (afb.1). Dit scherm bestaat uit 3 delen (afb.2).

Deel **1** is een lijst van 'recent history'. Als je de applicatie voor het eerst opstart, is deze lijst nog helemaal leeg en de buttons hebben nog geen functie. Maar iedere keer als men een bestand opent, wordt het pad naar dit bestand hier opgeslagen. Als men een al eerder geopend bestand

opnieuw wil gebruiken, kan men op de desbetreffende button klikken. De laatste 5 al eerder geopende bestanden worden hier getoond.

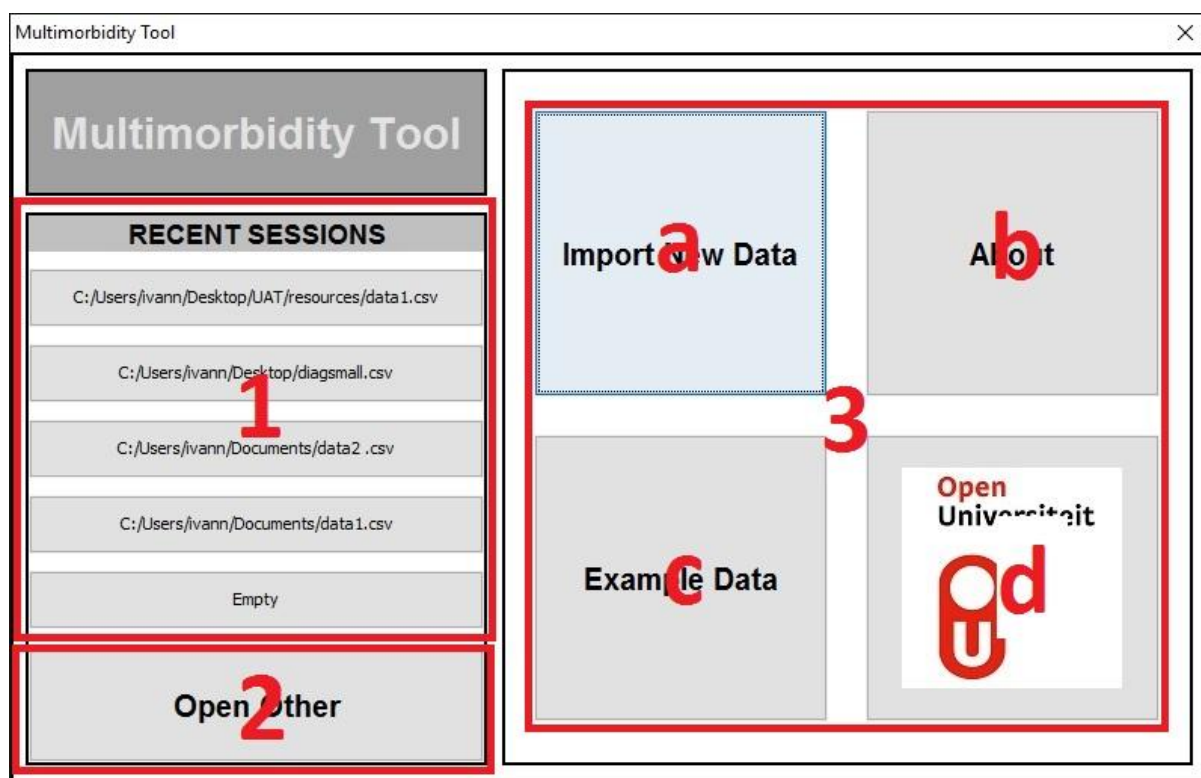
Deel **2** is een button 'Open Other' waarmee een gebruiker 'MMT' en 'RES' bestanden kan laden. Deze bestanden worden door de applicatie gemaakt als een gebruiker bepaalde gegevens opslaat. Door gebruik te maken van de button "Open Other" kan een gebruiker deze bestanden weer openen.

Deel **3** toont 4 buttons. De belangrijkste button is button **a**. Met de 'Import New Data' kan een gebruiker CSV bestanden importeren.

Button **b** toont een scherm met extra algemene informatie met betrekking tot het project.

Als een gebruiker de tool wil testen, maar niet over een dataset beschikt, dan kan hij gebruik maken van button **c** met 'Example Data'. Met deze knop kan een gebruiker voorbeeld data importeren en zo de functionaliteit van de applicatie bekijken.

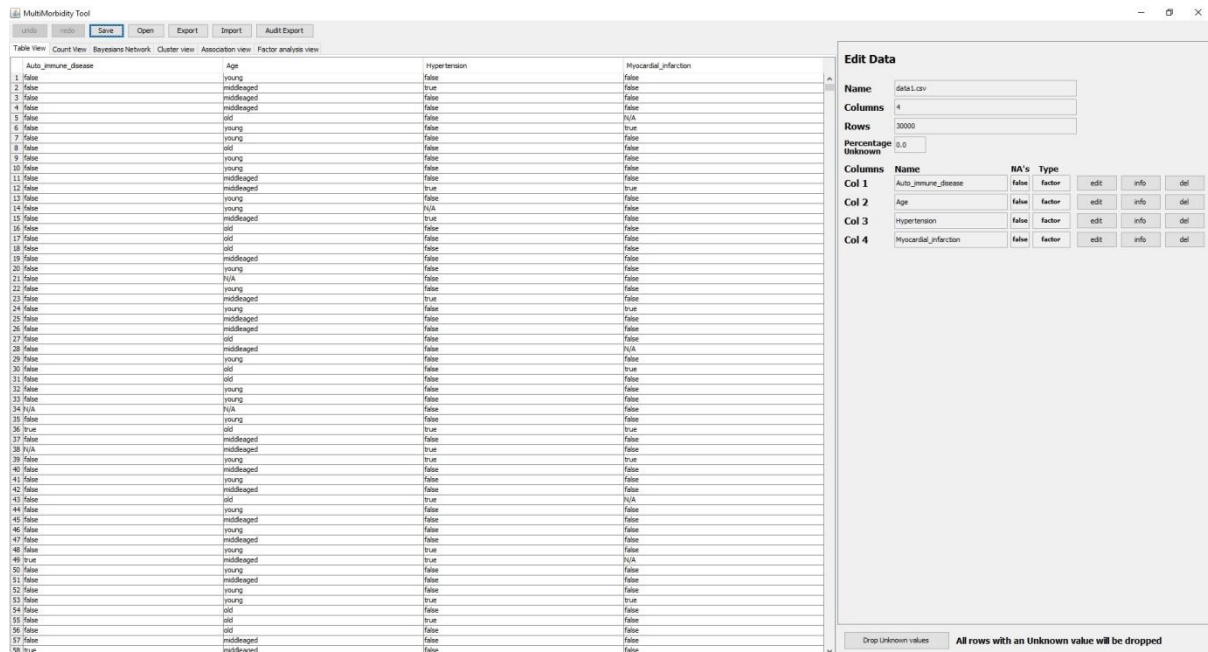
Op het startscherm is er tevens een Open Universiteitsknop (button **d**), deze knop start de default internet browser van de gebruiker en opent de homepage van de Open Universiteit.



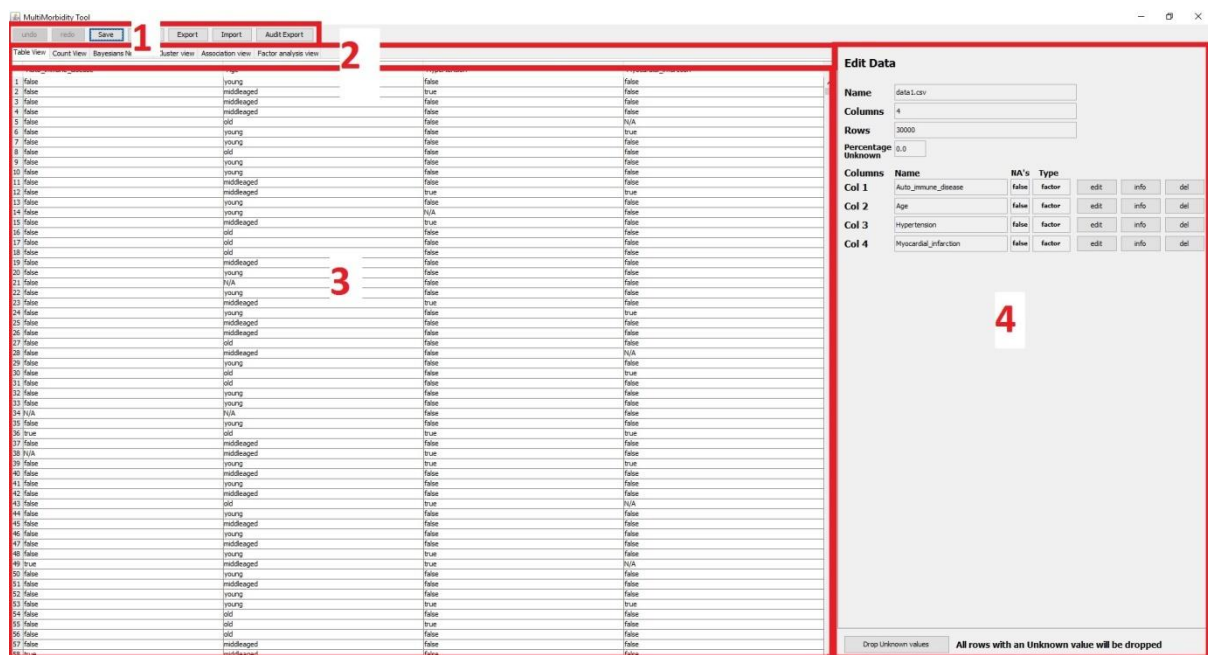
Afbeelding 2. Indeling startscherm.

2 Het Hoofdscherm

Nadat er gegevens zijn ingelezen verschijnt het hoofdscherm (afb. 3). Dit scherm bestaat uit 4 onderdelen (afb. 4).



Afbeelding 3. Hoofdscherm.



Afbeelding 4. Indeling hoofdscherm.

Onderdeel 1 is het hoofdmenu (afb. 5). Dit menu bestaat uit 7 knoppen.

De eerste twee knoppen hebben betrekking op de undo functionaliteit. Als deze knoppen actief zijn, kan een gebruiker een actie op de dataset ongedaan maken.

Knoppen 3 tot en met 6 hebben betrekking op het opslaan en openen van gegevens. Er is een onderscheid tussen 'Save' en 'Export'. Door te save worden de gegevens als MMT bestand

opgeslagen. Door te exporteren wordt de dataset als CSV bestand opgeslagen. Een CSV bestand bevat minder metadata dan een MMT. Dezelfde dichotomie bestaat tussen de 'Open' en 'Import' knoppen. Met behulp van 'Open' worden MMT bestanden geopend. Met 'Import' worden CSV bestanden gelezen.

De laatste knop betreft 'Audit Export'. Hiermee wordt een R Script bestand gecreëerd, waarmee een gebruiker deze direct in een R console kan inlezen.



Afbeelding 5. Hoofdmenu.

Onder het hoofdmenu bevindt zich onderdeel **2**, dat is het navigatiemenu in de vorm van tabs (afb. 6). Door een andere tab te kiezen kan de gebruiker tussen de verschillende analyse tools navigeren (onderdeel **3**). De afzonderlijke tabpanels worden later behandeld.



Afbeelding 6. Navigatie menu.

Onderdeel **4** is de EditPane (afb. 7). Met dit panel kan een gebruiker enkele eigenschappen en ook metadata van zijn dataset bewerken. Bovenin het panel worden enkele eigenschappen van de data afgebeeld: de naam van de dataset, het aantal kolommen, het aantal rijen en het percentage onbekende waarden in de dataset.

Vervolgens wordt er per kolom een aantal zaken afgebeeld in een lijst:

- **Name**, dit is de naam van de kolom.
- **NA's**, als de kolom een NA-waarde bevat dan wordt hier true afgebeeld, anders false.
- **Type**, in R kunnen kolommen verschillende typen hebben, waaronder Factor, Character, Logical, Numeric, Integer en Date. Het type van de kolom wordt hier getoond.

Hierachter zijn 3 knoppen afgebeeld:

- **edit**, met die knop verschijnt een editscherm (afb. 8) waarmee de gebruiker de kolom kan wijzigen.
- **info**, deze knop is alleen actief bij factor kolommen. Er verschijnt een scherm met enkele statistische gegevens met betrekking tot de kolom.
- **del**, met deze knop wordt de hele kolom verwijderd.

Onderin het panel bevindt zich nog een knop. Met deze 'Drop Unkown values' button, worden alle rijen in de hele dataset die een NA bevatten verwijderd.

Edit Data

Name

Columns

Rows

Percentage Unknown

Columns	Name	NA's	Type			
Col 1	Auto_immune_disease	false	factor	edit	info	del

All rows with an Unknown value will be dropped

Afbeelding 7. EditPane.

Edit Column Values

Column

Name

Type

Classification

Median

Replace Unknown values

☒ No Action

☐ Median

☐ Other

Levels

Level 1	<input type="text" value="middleaged"/>	<input type="button" value="Delete"/>
Level 2	<input type="text" value="N/A"/>	<input type="button" value="Delete"/>
Level 3	<input type="text" value="old"/>	<input type="button" value="Delete"/>
Level 4	<input type="text" value="young"/>	<input type="button" value="Delete"/>

Afbeelding 8. Editscherm.

Met het Editscherm kan men een aantal eigenschappen van een kolom wijzigen.

- **Name**, in dit veld kan een nieuwe naam gezet worden. Let op, een naam mag geen nummer zijn, mag geen whitespace bevatten en kan alleen uit letters en nummers bestaan.
- **Type**, met dit dropdown menu kan je het type van de kolom wijzigen.
- **Classification**, met dit dropdown menu kan je de classification van de kolom wijzigen.
- **Replace Unknown values**, met deze radio button kan je aangeven wat er met de NA waarden moet gebeuren. Met 'No Action' blijven ze staan, met 'Median' worden ze door de meest voorkomende waarde in de kolom vervangen of met 'Other' krijgen ze een andere waarden.

Om deze wijzigingen te bevestigen, moet men na het aanbrengen van de wijzigingen op de 'Save & Continue' knop drukken.

Met 'Cancel' worden de wijzigingen niet aangebracht.

Factor type

Bij een Factor kolom worden ook de afzonderlijke factors/levels afgebeeld. De gebruiker kan hier de namen wijzigen, of de level verwijderen.

Bij een Factor kolom is er ook een extra knop zichtbaar, de 'Convert to Columns' knop. Met deze knop worden de afzonderlijke levels omgezet naar aparte kolommen.

Numeric type

Bij Numeric en Integer type kolommen, kan een gebruiker de 'Group' button gebruiken om de waarden te grouperen en om te zetten naar levels.

3 Views

Met gebruik van het navigatiemenu kan een gebruiker tussen de verschillende views navigeren en de afzonderlijk views bekijken.

3.1 Count view

The screenshot displays the 'Countview' interface. On the left side, there are two control panels, both highlighted with red rectangles. The top panel, 'Count Controls', includes a 'Select Group ID (Select One)' dropdown menu with 'Auto_immune_disease' selected, a 'Select Disease Columns(Multi Select)' list, and a 'Calculate' button. Below it, instructions read: 'Step 1. Select an ID Column', 'Step 2. Select Diseases', 'Only logical Columns can be selected', and 'Use Edit Menu to cast Columns'. The bottom panel, 'GLM Controls', features a 'Select Dependent' dropdown menu with 'Auto_immune_disease' selected, a 'Select Other Columns (Multi Select)' list containing 'Auto_immune_disease', 'Age', 'Hypertension', and 'Myocardial_infarction', and a 'Calculate' button. The right side of the interface contains three large, empty table areas. The top-left table is titled 'Total Observations per Disease Count', the top-right table is titled 'Total Observations per Group and Disease Count', and the bottom-left table is titled 'Generalized Linear Models (GLM)'. The bottom-right area is a large, empty grey rectangle.

Afbeelding 9. Countviewer

De counterview heeft twee controls: een count Controller en een GLM Controller (afb. 9).

- **Count Controller.** Met deze berekeningen wordt een aantal statische kenmerken berekend. In de eerste tabel worden de geselecteerde waarden geteld. In de tweede tabel worden ze gegroepeerd aan de hand van een de waarden van een geselecteerde kolom.

In de JList selecteert de gebruiker de kolom waarvan de waarden geteld moeten worden. En in de dropdownlijst de waarde waaraan deze waarden gegroepeerd moeten worden.

- **GLM Controller.** Hiermee wordt een GLM berekening gemaakt. In de dropdown moet men een afhankelijke kolom selecteren en in de JList de kolommen waarvan de afhankelijkheden berekend moet worden.

3.2 Bayesian Network

Met behulp van het Bayesian Network panel kan een gebruiker een bayesiaans netwerk creëren, bewerken en onderzoeken.

Stap 1.

De gebruiker moet eerst de kolommen selecteren die deel uit zullen maken van het netwerk. Let op, er kunnen alleen factoriale kolommen gekozen worden die geen onbekende waarden bevatten.

Stap 2.

Met de dropdown lijst moet er een algoritme gekozen worden. Als men vervolgens op 'calculate' drukt, wordt het netwerk gecreëerd. Rechts verschijnt er een netwerk graaf, met daaronder de afkorting van het algoritme. Als het netwerk compleet is, worden de CPTS tabellen gevuld. Men kan deze zien door met het nieuwe tabmenu bovenin tussen de twee views te navigeren.

Een netwerk is compleet als:

- er geen cycles bestaan.
- er maximaal 1 arc tussen twee verschillende nodes bestaat.

Stap 3. (Niet verplicht)

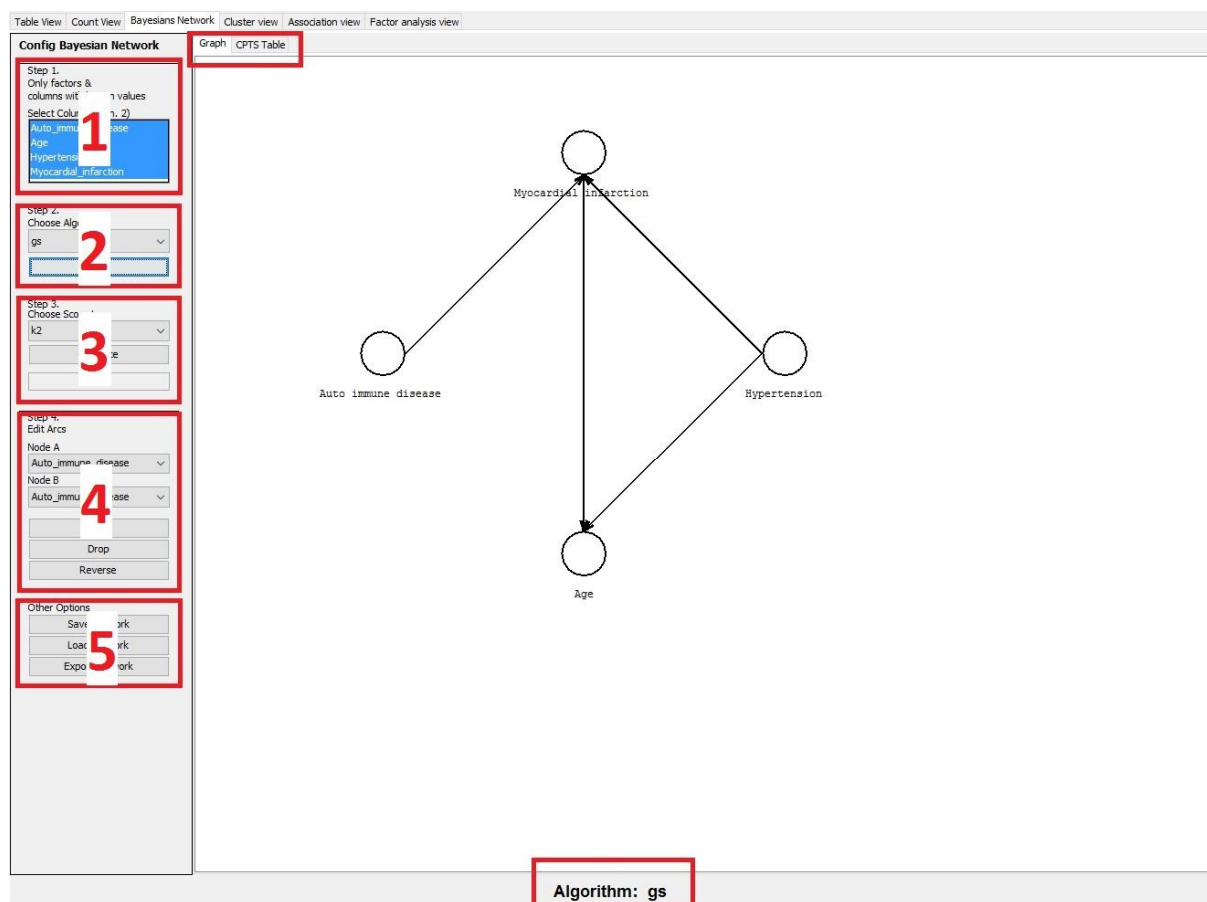
Men kan de score van het netwerk berekenen. Kies een algoritme en druk op calculate.

Stap 4. (Niet verplicht)

Met de dropdown menu's kunnen er twee nodes geselecteerd worden. Met de drie knoppen kunnen de 'Arcs' tussen deze nodes worden gewijzigd.

De opties zijn:

- **Set**, er wordt een nieuwe arc gecreëerd als er geen arc bestaat.
- **Drop**, als er een arc tussen de twee nodes bestaat dan wordt hij verwijderd.
- **Reverse**, als er een arc tussen de twee nodes bestaat dan wordt hij omgekeerd.



Afbeelding 10. Bayesian Network Panel.

Stap 5

Met de volgende knoppen kunnen resultaten worden opgeslagen of geladen:

- **Save Network**, met deze optie wordt een RES bestand opgeslagen.
- **Load Network**, met deze optie kan een RES bestand worden gelezen.
- **Export Network**, met deze optie wordt er een R Net bestand opgeslagen, dit bestand kan alleen met R gelezen worden.

3.3 Association analysis

Met behulp van de 'Association analysis' kan een gebruiker relaties tussen variabelen op een aantal aspecten onderzoeken.

Het associatie-analyse scherm bestaat uit een linker gedeelte (menu panel) en een rechter gedeelte (resultaat panel).

Association Analysis

Step 1.
Only Integer and Numeric Columns can be selected.
If comboboxes are empty use Edit Data Menu

Select first variable:
LungCap

Select second variable:
Height

Step 2.

Select completeness
complete.obs

Select method
kendall

Correlation

Step 3.

Select alternative
not equal

Input confidence level: (value >0&<1)
0.95 Help

Correlation test

Odds Ratio

Step 1.
Only Factor Columns can be selected.
If comboboxes are empty use Edit Data Menu

Select first variable:

Select second variable:

Step 2.

Select method
cohort.count

Input confidence Level (value >0&<1)
Help

Charts

Scatter plot showing the relationship between LungCap (x-axis) and Height (y-axis). The plot displays a positive correlation between the two variables.

Summary

LungCap		Age		Height		Smoke		Gender	
Min.	: 0.507	Min.	: 3.00	Min.	: 45.30	no	: 648	female	: 358
1st Qu.	: 6.150	1st Qu.	: 9.00	1st Qu.	: 59.90	yes	: 77	male	: 367
Median	: 8.000	Median	: 13.00	Median	: 65.40				
Mean	: 7.863	Mean	: 12.33	Mean	: 64.84				
3rd Qu.	: 9.800	3rd Qu.	: 15.00	3rd Qu.	: 70.30				
Max.	: 14.675	Max.	: 19.00	Max.	: 81.80				
Caesarean									
no : 561									
yes : 164									

Menu panel Result panel Edit Panel

Association analyse bevat twee soorten analyses:

1. Correlatie tussen twee **numerieke variabelen** bepalen

Stap 1: Selecteer de te onderzoeken numerieke variabelen. Men krijgt in het resultaat panel [Charts] de frequentie grafiek te zien of de relatie grafiek tussen de twee variabelen. Op hetzelfde panel [Summary] krijgt men de samenvatting van de dataset te zien.

Step 1.
Only Integer and Numeric Columns can be selected.
If comboboxes are empty use Edit Data Menu

Select first variable:
LungCap

Select second variable:
Age

LungCap	Age	Height	Smoke	Gender
Min. : 0.507	Min. : 3.00	Min. :45.30	no :648	female:358
1st Qu.: 6.150	1st Qu.: 9.00	1st Qu.:59.90	yes: 77	male :367
Median : 8.000	Median :13.00	Median :65.40		
Mean : 7.863	Mean :12.33	Mean :64.84		
3rd Qu.: 9.800	3rd Qu.:15.00	3rd Qu.:70.30		
Max. :14.675	Max. :19.00	Max. :81.80		
Caesarean				
no :561				
yes:164				

Stap 2: Na het selecteren van twee numerieke variabelen, **Variable1** en **Variable2**, kan men de zogenaamde “**Use**” kiezen die aangeeft hoe het programma de onbekende waarden moet afhandelen. Daarna kan men kiezen voor een **correlatie methode** en op de **correlatie knop** drukken om uiteindelijk de correlatie tussen de geselecteerde variabelen te bepalen. De correlatie waarde wordt in het resultpanel [Summary] getoond.

Step 2.

Select completeness
complete.obs

Select method
kendall

Correlation

Stap 3: Het is ook mogelijk om een **correlatie test** uit te voeren, gegeven een **alternatieve hypothese** met een **confidence level** (default is 0.95 ofwel 95%).

Step 3.

Select alternative
not equal

Input confidence level: (value >0&<1)
0.95 Help

Correlation test

Pearson's product-moment correlation

data: cdata[, "LungCap"] and cdata[, "Age"]
t = 38.476, df = 723, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7942660 0.8422217
sample estimates:
cor
0.8196749

- Hoge P waarden: het is waarschijnlijker dat de nul hypothese correct is.
- Lage P waarden: het is waarschijnlijker dat de nul hypothese niet correct is.

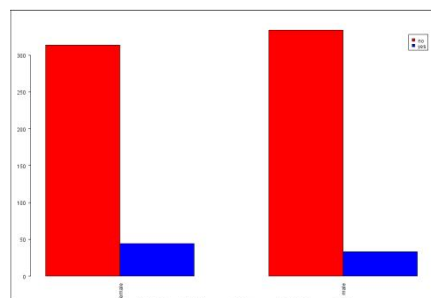
2. Relative risk and Odd Ratio tussen twee (categorical) variabelen bepalen

Stap 1: Selecteer de te onderzoeken (categorical) variabelen. Men krijgt in de result panel [Charts] de frequentie grafiek te zien of de relatie grafiek tussen de twee variabelen.

Step 1.
Only Integer and Numeric Columns
can be selected.
If comboboxes are empty use Edit Data Menu

Select first variable:
LungCap

Select second variable:
Age



Stap 2: Voor het bepalen van de Relative risk and Odd Ratio moet men twee variabelen, **Variable 1 (Categorical)** en **Variable 2 (Categorical)** selecteren. Nadat de relaties geselecteerd zijn, kan men een **methode** selecteren met een **confidence level** (default is 0.95 ofwel 95%). Wanneer men op de knop “RR/Odd Ratio” drukt, wordt de Relative risk and Odd Ratio bepaald en op het resultspanel getoond.

Step 2.

Select method
cohort.count

Input confidence Level (value >0&<1)
0.99

Help

RR/ Odds Ratio

	Outcome +	Outcome -	Total	Inc risk *	Odds
Exposed +	314	334	648	48.5	0.940
Exposed -	44	33	77	57.1	1.333
Total	358	367	725	49.4	0.975

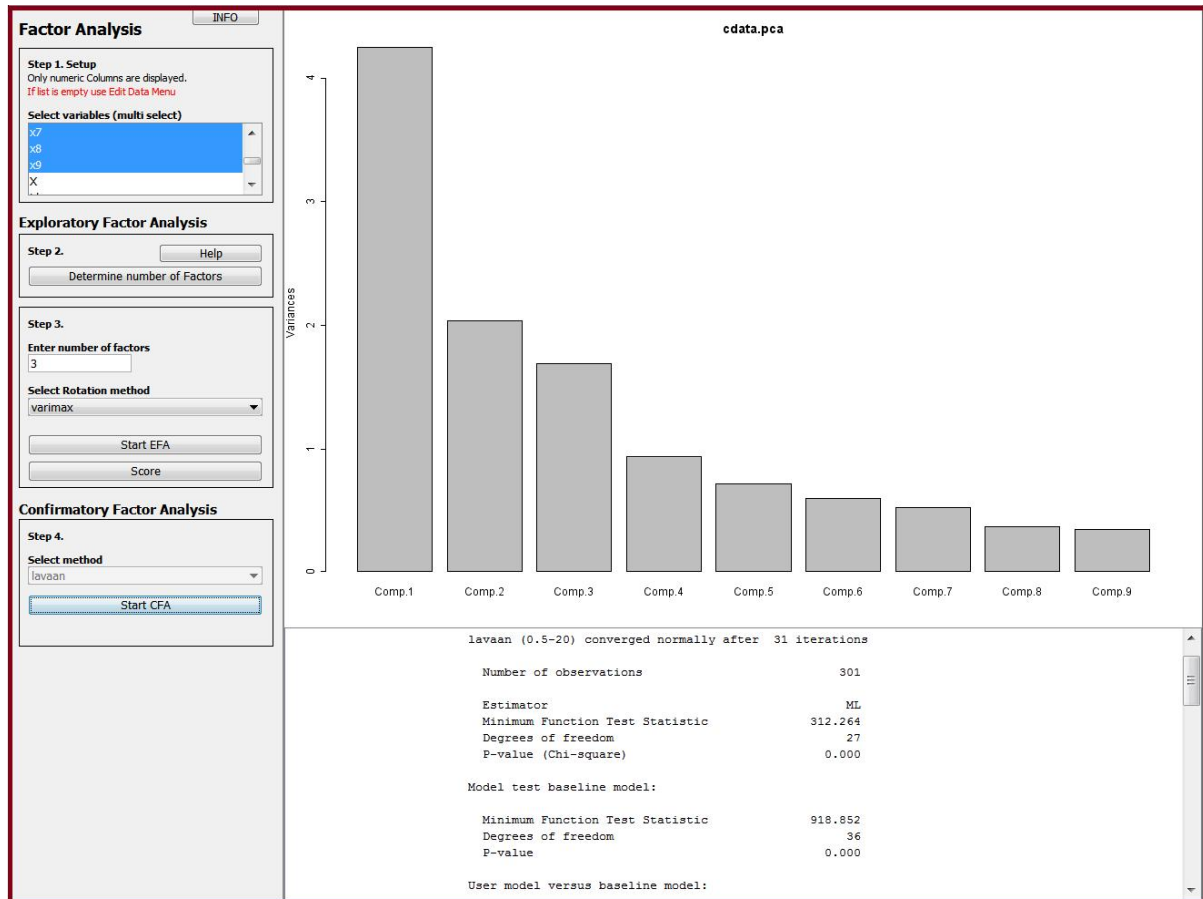
Point estimates and 99 % CIs:

Inc risk ratio (W)	0.85 (0.64, 1.12)
Odds ratio (W)	0.71 (0.38, 1.32)
Attrib risk (W) *	-8.69 (-24.07, 6.70)
Attrib risk in population (W) *	-7.76 (-23.06, 7.53)
Attrib fraction in exposed (%)	-17.93 (-55.22, 10.41)
Attrib fraction in population (%)	-15.72 (-47.27, 9.07)

X2 test statistic: 2.077 p-value: 0.15
W: Wald confidence limits

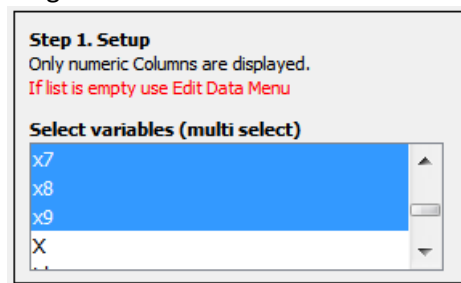
3.4 Factor analysis

Factoranalyse wordt gebruikt om na te gaan of er in een reeks variabelen een gemeenschappelijke factor zit. Met andere woorden, bij factoranalyse wil de onderzoeker nagaan of de afzonderlijke variabelen te herleiden zijn naar één of een beperkt aantal aspecten (factoren). Zo ja, dan kan men een samengestelde variabele maken, de zogenaamde latent variable.



Stap 1.

De gebruiker moet eerst de kolommen selecteren die tijdens de factor analyse gebruikt zullen worden.



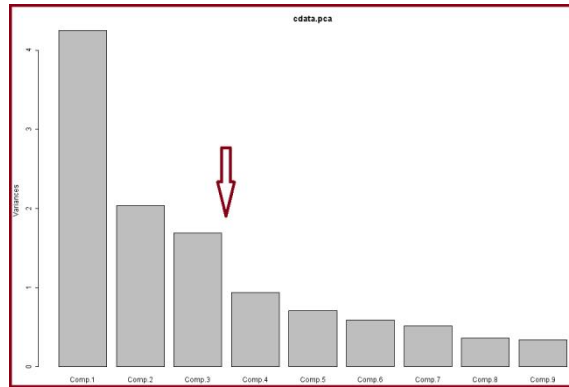
Stap 2: Helpen bij het bepalen van het aantal factoren dat voor de factornalyse als invoer gebruikt kan worden. Men krijgt een grafiek. Het beste aantal factoren ligt rond het punt op de grafiek vanaf het moment dat er een stijging begint te vormen.

Exploratory Factor Analysis

Step 2.

Help

Determine number of Factors



Step3: Nadat de gebruiker het aantal factoren, de rotation-method heeft gekozen en op “start EFA” drukt, wordt de EFA gestart. Het resultaat van EFA wordt op het resultpanel getoond:

Step 3.

Enter number of factors
3

Select Rotation method
varimax

Start EFA

Score

```
factanal(x = cdata, factors = 3, rotation = "varimax")
```

Uniquenesses:

	x1	x2	x3	x4	x5	x6	x7	x8	x9
	0.513	0.749	0.543	0.279	0.243	0.305	0.502	0.469	0.543

Loadings:

	Factor1	Factor2	Factor3
x1	0.277	0.623	0.151
x2	0.105	0.489	
x3		0.663	0.130
x4	0.827	0.165	
x5	0.861		
x6	0.801	0.212	
x7			0.696
x8		0.162	0.709
x9	0.132	0.406	0.524

Wanneer men op de button “score” drukt, verschijnen op het resultpanel de scores van alle observaties, geprojecteerd op de gekozen factoren.

	Factor1	Factor2	Factor3
[1,]	0.082471519	-0.725861635	-0.001591841
[2,]	-1.212737200	0.516199927	0.822875023
[3,]	-1.768413075	-0.247687769	-1.088698544
[4,]	-0.044497068	0.766687125	-0.769782192
[5,]	-0.066667378	-0.530656200	0.344127073
[6,]	-1.530295463	0.349964568	1.148005610
[7,]	0.907038627	-1.382074196	0.367721892
[8,]	0.012680971	-0.086960055	-0.838373167
[9,]	0.515935167	-0.284137603	0.220987133
[10,]	0.427257419	-1.315367285	-0.589241384

Stap4: Nadat de gebruiker de kolomen heeft geselecteerd die bij een Latent Factor horen en op de knop “Start CFA” drukt, wordt de CFA gestart. Het resultaat wordt op het resultpanel getoond.

Note: CFA method is default “lavaan”

Confirmatory Factor Analysis

Step 4.

Select method

lavaan

Start CFA

```
lavaan (0.5-20) converged normally after 22 iterations

Number of observations              301

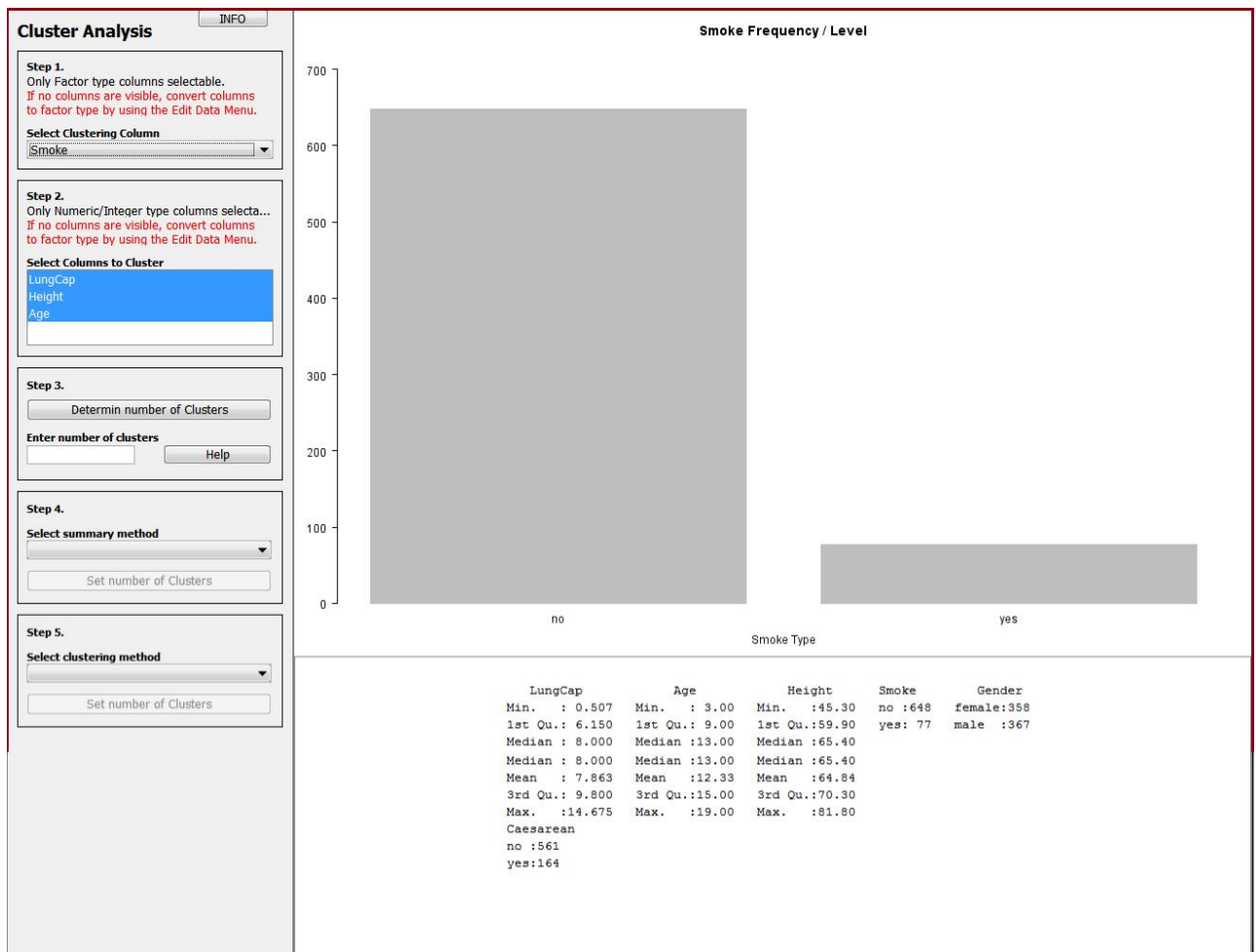
Estimator                          ML
Minimum Function Test Statistic    11.798
Degrees of freedom                  2
P-value (Chi-square)               0.003

Model test baseline model:

Minimum Function Test Statistic    157.076
Degrees of freedom                  6
P-value                            0.000
```

3.5 Cluster Analysis

Clusteranalyse is een multivariate analyse die groepen of "clusters" van voorwerpen (waarnemingspunten in ons geval) probeert te vormen die met elkaar "vergelijkbaar" zijn maar wel verschillen tussen clusters.



Stap 1: Selecteer een (categorical) variable waarop clustering wordt gedaan

Step 1.
Only Factor type columns selectable.
If no columns are visible, convert columns to factor type by using the Edit Data Menu.

Select Clustering Column

ziektes

Stap 2: Selecteer de numerieke variabelen op basis waarmee clustering wordt uitgevoerd.

Step 2.
Only Numeric/Integer type columns selecta...
If no columns are visible, convert columns to factor type by using the Edit Data Menu.

Select Columns to Cluster

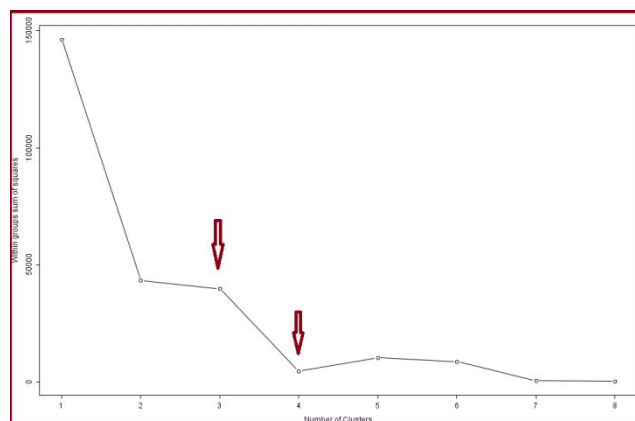
med1
med3
med5
med6

Step 3: Help bij het bepalen van het aantal clusters. Een grafiek wordt in het resultpanel[Chart] getoond. Het beste aantal clusters ligt rond het punt op de grafiek vanaf het moment dat er een stijging begint te vormen.

Step 3.

Help

Determin number of Clusters



Step 4: Nadat de gebruiker het aantal clusters en de methode (clusterattribute) heeft gekozen, wordt de samenvatting over dit attribuut in het resultpanel getoond die bij het meegegeven aantal clusters hoort.

Step 4.

Enter number of clusters

4

Select summary method

Cluster assignment

View Clustering Summary

Select summary method

Cluster assignment

cluster
size
withinss
centers
Cluster assignment

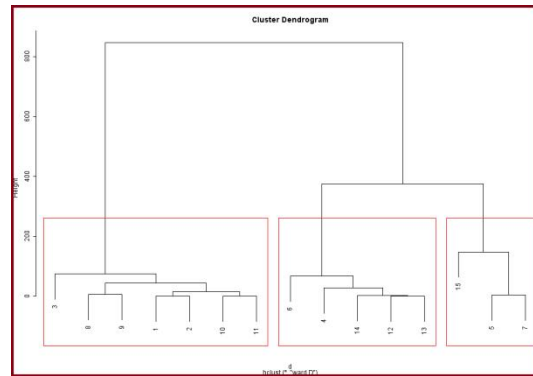
Stap 5: Nadat de gebruiker een cluster methode heeft gekozen en op "start clusteren" drukt, wordt het clusteren uitgevoerd en het resultaat op het resultpanel getoond. Het resultaat is een grafiek die aangeeft welke clusters aangemaakt zijn.

Step 5.

Select clustering method

Hierarchical clustering-kmeans ▼

Start Clustering



Het is nog steeds mogelijk om de samenvatting van de attributen van het clustering proces op te vragen: step 4.

MMT Installatiehandleiding

MultiMorbidity Tool

versie 1

Stap 1: installatie van R
Stap 2: installatie van de vereiste packages voor MMT
Stap 3: de installatie van MMT en configuratie
Stap 4: start MMT

Stap 1: R installeren

⇒ Zie R installatiehandleiding

Volg de instructies. Let op, om eventuele onvoorziene problemen te voorkomen, installeer R in het volgende pad: C:/R-3.2.3.

Stap 2: De vereiste packages installeren

Start R en installeer via de R console de volgende packages:

R Package name	Used for:	How to install in R
rJava	Interface tussen R en Java	install.packages("rJava")
bnlearn	Bayesian netwerken	install.packages("bnlearn")
JavaGD	R plots in Java tonen	install.packages("JavaGD")
reshape2	Data set manipulatie	install.packages("reshape2")
lavaan	Factoranalyse	install.packages("lavaan")
epiR	Relative risk en Odds ratio	install.packages("ePir")

Controleer of de bestanden die gedownload worden in de <R installatie folder>\library map terechtkomen. Soms heeft R hiervoor geen rechten en worden de bestanden in een andere folder geplaatst. Zo ja, dan moet je ze handmatig naar de hierboven vermelde folder verplaatsen. Pas op, sommige packages downloaden tegelijk meerdere mappen.

Stap 3: de installatie van MMT (MultiMorbidity Tool) en configuratie

Maak een folder aan en pak het MMT zip bestand uit naar deze folder.

Navigeer naar de folder en wijzig eventueel het JRI.properties bestand in de configuration folder. Dit moet je doen als je R in een andere folder hebt geïnstalleerd dan hierboven aangegeven. Gebruik voor het pad geen backslashes maar wel forwardslashes.

Stap 4: start MMT

Navigeer naar de folder waar de MMT is geïnstalleerd en click mmt.jar of de snelkoppeling om de applicatie te starten:

⇒ Zie handleiding Multimorbidity Tool

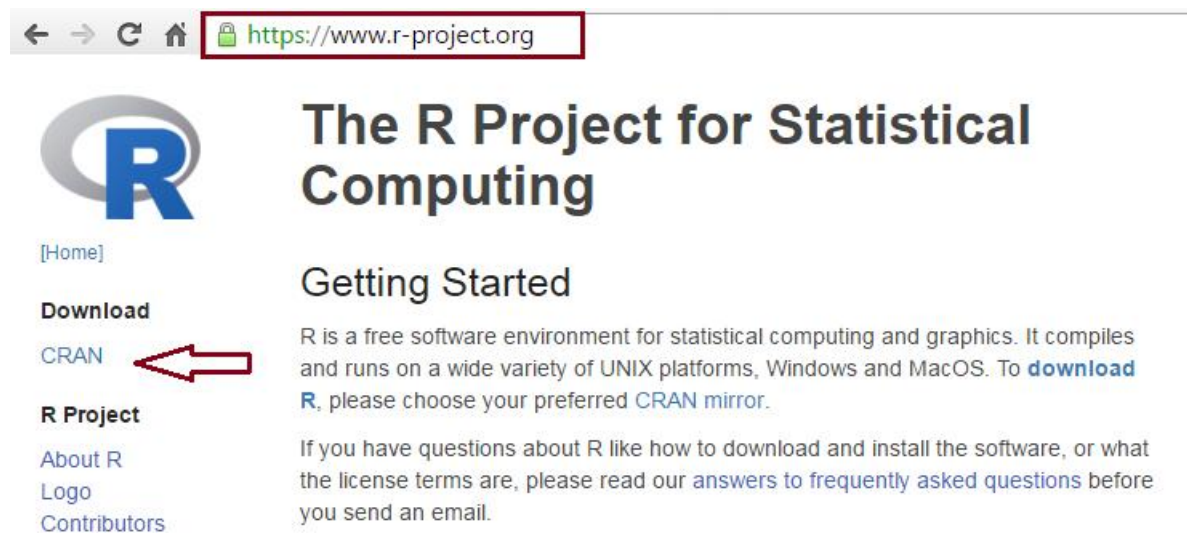
Voor meer informatie over de installatie, zie <https://cran.r-project.org/doc/manuals/r-release/R-admin.html#Obtaining-R>

R Installatie Handleiding

1. R downloaden
2. R installeren
3. Packages installeren
4. Aantal R commando's
5. More info

1. R downloaden

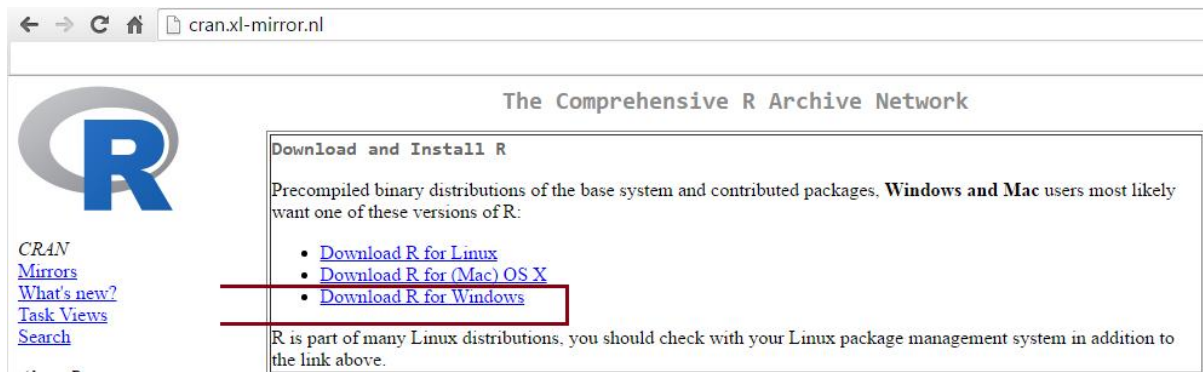
Navigeer naar de website CRAN (Comprehensive R Archive network) : <http://www.r-project.org>
Click op CRAN aan linker menu op de pagina



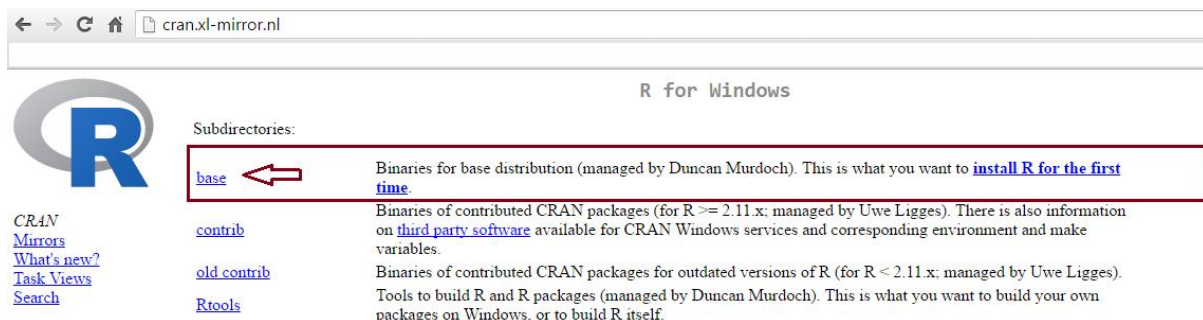
Kies een locatie van waaruit de download gestart kan worden:



Kies het platform waarop R geïnstalleerd moet worden. In ons geval: download R voor windows.



Download de R base en sla deze op (<download-directory>)

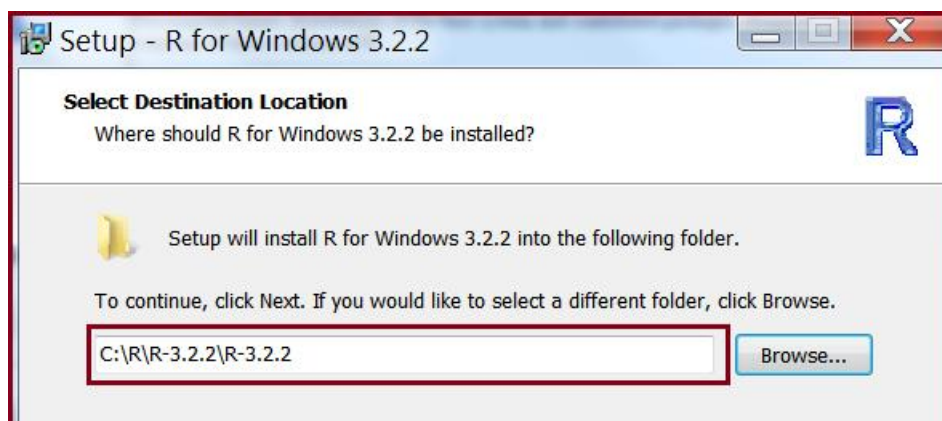


2. R installeren

Navigeer naar de <download-directory> en dubbel click op R-3.2.2-win.exe, in ons geval gaat het om versie R-3.2.2.

- Kies de taal en druk op de knop OK
- Kies C:\R\R-3.2.2 als installatiepad

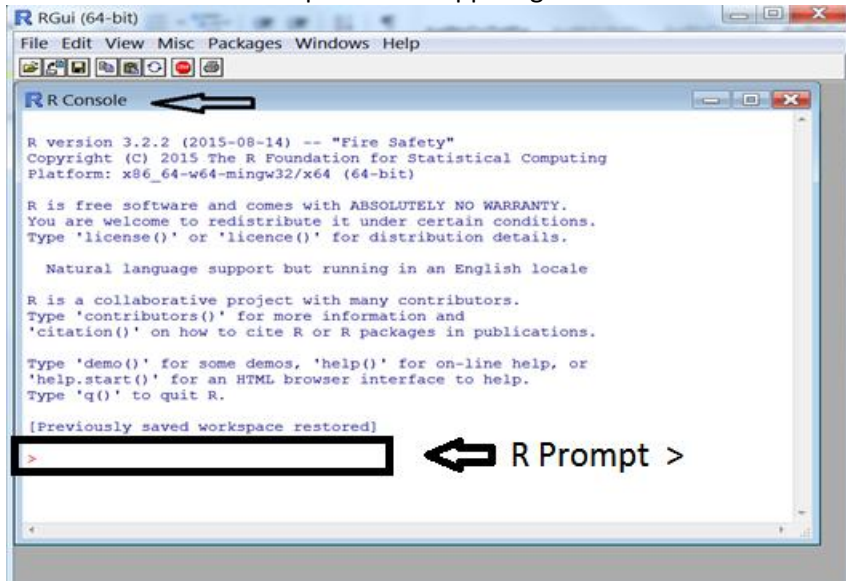
NOTE: Om enkele installatie problemen te voorkomen, wordt er aangeraden om R niet in de Program Files of Program Files(x86) folder te installeren. Graag een folder zonder whitespaces gebruiken.



- Na de een succesvol installatie, wordt een aantal snelkoppelingen aan het start menu toegevoegd.



- Start de R console door op een snelkoppeling te klikken.



3. R packages installeren

Na de installatie van R, zijn er een aantal packages automatisch voor je geïnstalleerd. Dat zijn de R base packages.

Voor een uitgebreide lijst en informatie, zie:

<https://stat.ethz.ch/R-manual/R-devel/library/base/html/00Index.html>

De installatie van een package gebeurt als volgt. Achter de R prompt, moet men het volgende R statement invoeren en daarna op Enter drukken.

- `install.packages("NaamPackage")`
waarbij "NaamPackage" de naam van de package die men will installeren.

Voorbeeld package "lavaan":

- `install.packages("lavaan")`

Wil men een overzicht van alle tot nu toe geïnstalleerd packages:

- `installed.packages()`

4. Aantal R commando's

R-code	Beschrijving
<code>getwd()</code>	get working dir
<code>setwd("c:/workspace")</code>	set working dir
<code>projectDir <- "C:\projectdir"</code>	

setwd(projectDir)	
dir()	list the contents of the current directory
rm(list=ls())	clear working space
save.image(firstproject.Rdata)	save workspace image
load("firstproject.Rdata")	Load a saved workspace
ls()	Listing workspace
load(file.choose())	File chooser dialog
source("myfile.R")	load and execute a script of R commands
sink("myfile.txt")	redirect console output to a file
sink()	restore output to the screen
pdf("mygraph.pdf")	subsequent graphical output will go to a PDF
png("mygraph.png")	subsequent graphical output will go to a PNG
jpeg("mygraph.jpeg")	subsequent graphical output will go to a JPEG
bmp("mygraph.bmp")	subsequent graphical output will go to a BMP
postscript("mygraph.ps")	subsequent graphical output will go to a PostScript file

5. More info

[a\) http://www.cyclismo.org/tutorial/R/types.html](http://www.cyclismo.org/tutorial/R/types.html)

[b\) http://www.r-tutor.com/r-introduction](http://www.r-tutor.com/r-introduction)

[c\) http://www.r-tutor.com/](http://www.r-tutor.com/)

https://www.youtube.com/watch?v=cX532N_XLIs&list=PLqzoL9-eJTNBDdKgJzaQcY6OXmsXAHU&index=1

JRI Installatiehandleiding (5 stappen)

Stap 1. R installeren

Stap 2. rJava installeren

Stap 3. Eclipse configureren

3a. JVM argument

3b. Systeem variabelen

3c. Argument en variabelen instellen in Eclipse.

Stap 4. Het toevoegen van Jars aan de Buildpath.

Stap 5. Resterende packages in R installeren.

Stap1. R installeren

Een JRI applicatie (bijvoorbeeld de Multimorbidity Tool) werkt samen met R. Het is daarom belangrijk dat R ook geïnstalleerd is op de computers waarop de JRI applicatie draait. Om enkele installatie problemen te voorkomen, wordt er aangeraden om R niet in de Program Files of Program Files(x86) folder te installeren. Graag een folder zonder whitespaces gebruiken.

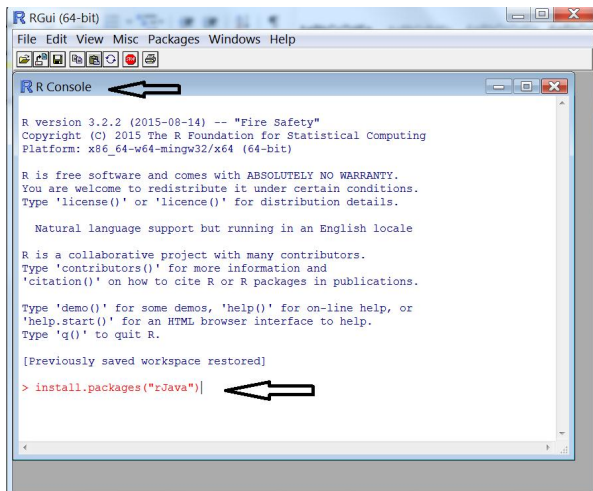
⇒ **Zie handleiding installatie R**

Wij hanteren in deze installatie de map : C:\R\R-3.2.2, waarin R is geïnstalleerd. Na een succesvolle installatie, wordt een snelkoppeling aan het start menu toegevoegd:



Stap 2. rJava installeren

- Start de R console door de R snelkoppeling te kiezen (R X64 3.2.2) of de C:\R\R-3.2.2\bin\x64\Rgui.exe te starten.
- Voer het volgende commando uit:
install.packages("rJava")



Stap 3. Eclipse configuratie

Er moeten drie zaken worden ingesteld in Eclipse bij de Run Configurations van de klasse met de main methode van het project. De default klasse met de main methode in ons project is de klasse StartApplication. Deze zaken betreffen twee systeem variabelen en één Java Virtual Machine (JVM) argument.

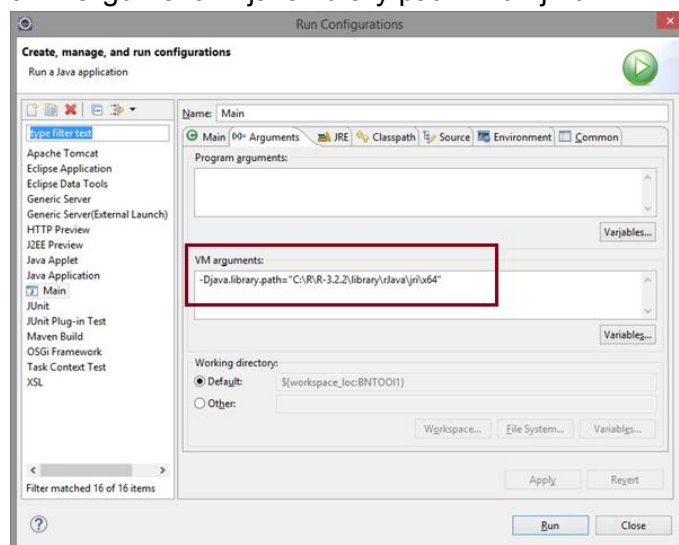
Er bestaat ook een plugin voor Eclipse om deze variabelen in te stellen. Voor meer informatie en om de plugin te downloaden, ga naar:

<http://www.studytrails.com/RJava-Eclipse-Plugin/>

Stap 3a. JVM Argument

Het is ook mogelijk om de R_HOME en PATH variabelen met behulp van Control Panel van Windows default aan de path systeem variabele toe te voegen.

JVM argument: `-Djava.library.path="<dir jri.dll>"`



Bij het installeren van rJava wordt er een dynamic library bestand geïnstalleerd. De locatie hiervan moet als JVM argument doorgegeven worden aan de Java Virtual Machine. De 64bit versie is te vinden in de R installatie folder in de submappen "library\rJava\jri\x64".

Er is ook een 32bit versie; deze is te vinden in de subfolder library\rJava\jri\x86".

Een voorbeeld van de <dir jri.dll> pad is: C:\R\R-3.2.2\library\rJava\jri\x64". Let op: je hoeft alleen de locatie van het bestand door te geven.

Stap 3b. Systeem variabelen

PATH variabel: <dir r.dll>

Bijvoorbeeld:

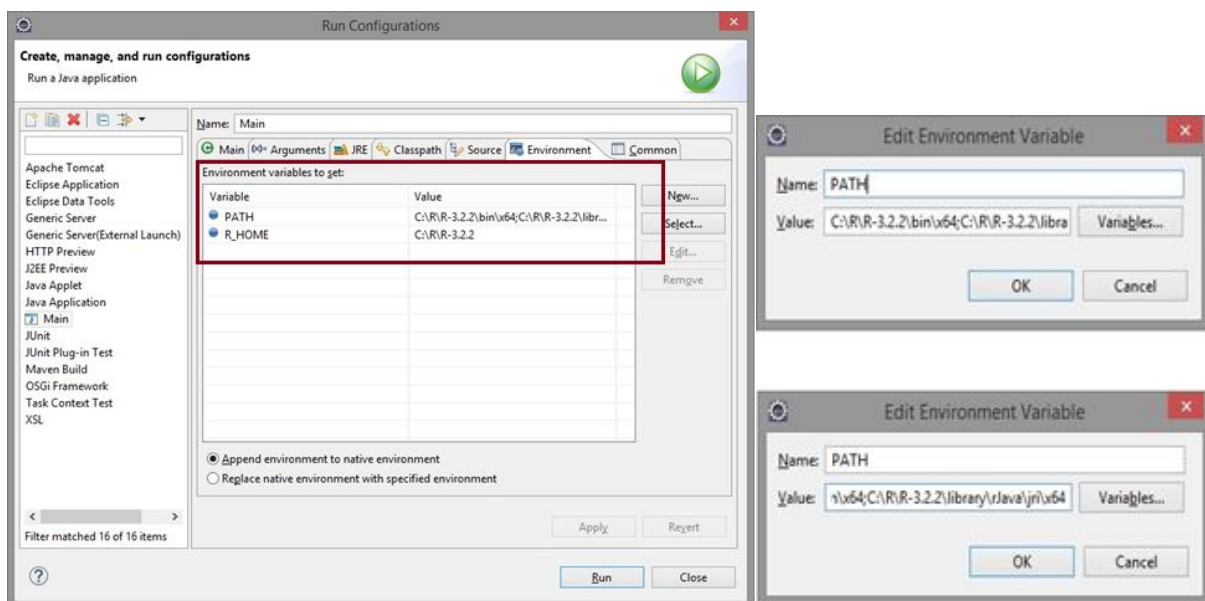
C:\R\R-3.2.2\bin\x64;

R_HOME variabel: <dir R.exe>

Bijvoorbeeld:

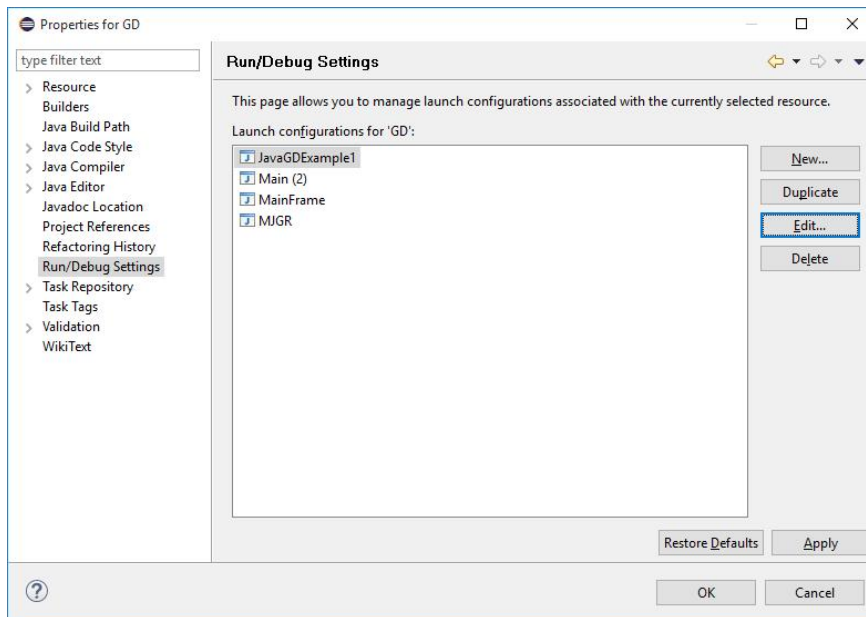
C:\R\R-3.2.2\bin\x64;

Pas op! Er bestaat van de r.dll een 32bit en 64bit versie. Gebruik deze versies niet door elkaar.



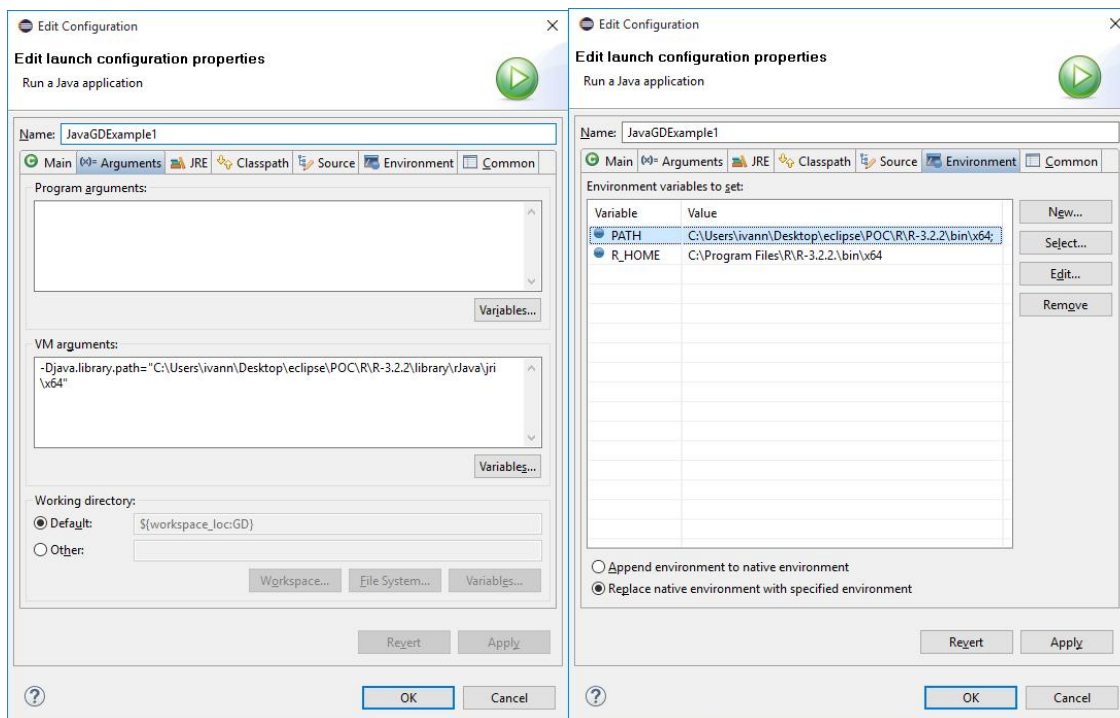
Stap 3c. Argument en variabelen instellen in Eclipse.

Variabelen en argumenten kunnen op de volgende manier in Eclipse worden ingevoerd:
Via file -> properties



Selecteer “Run/Debug”. Vervolgens selecteer je je JRI java project en klik daarna op “Edit”.

Selecteer het tabblad (x)= Arguments en vul bij “VM arguments” de desbetreffende JVM argumenten in.



Selecteer daarna het tabblad “Envioment”. Gebruik “New” om twee nieuwe variabelen te maken. Maak een PATH variabele aan met als waarde de locatie van het jri.dll bestand en een R_HOME variabele met als waarde de locatie van de R.exe die gebruikt moet worden.

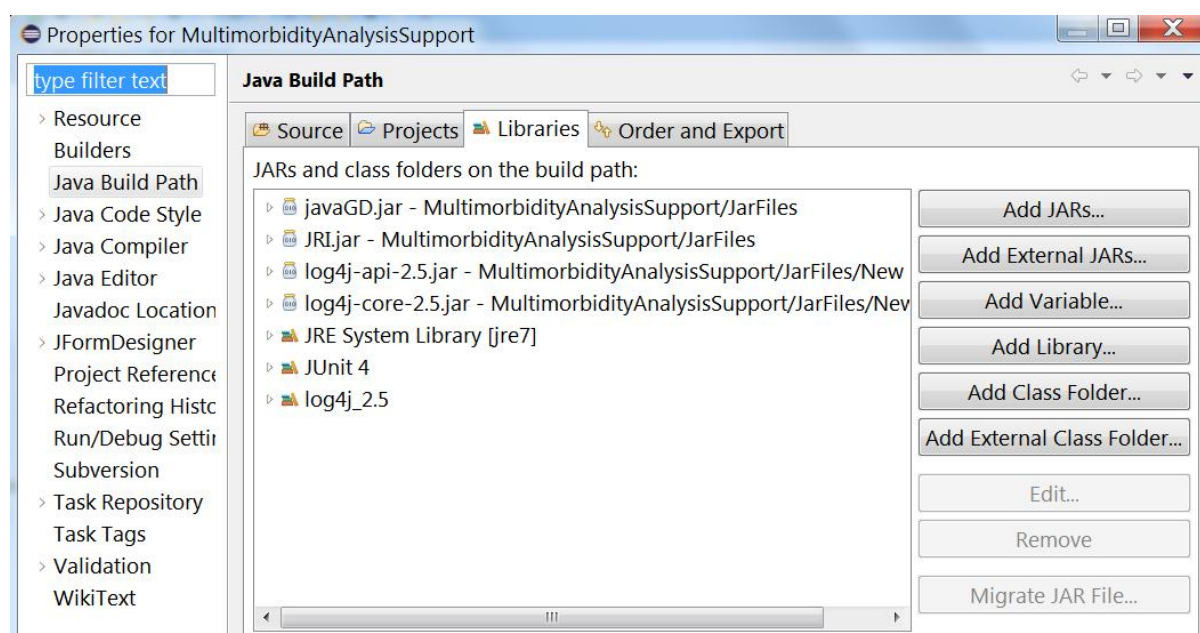
Men kan er voor kiezen de variabelen te appenden of om de huidige systeemvariabelen te vervangen. Dat doe je door gebruik te maken van de onderste radiobuttons.

Stap 4. Het toevoegen van Jars aan de Buildpath.

De volgende Jars moeten worden toegevoegd aan de buildpath van het project:

-JRI.jar
 -javaGD.jar (nodig als je gebruik maakt van javaGD)
 -----Voorbeeld jars voor logging en JUnit -----
 -log4j-core-2.5.jar (voor logging)
 -log4j-api-2.5.jar
 -junit.jar

JRI.jar wordt gebruikt als API om een koppeling te maken tussen Java en R. De javaGD.jar wordt gebruikt om R plots binnen Java te tonen. De log4j jars zijn nodig voor de logger en de junit.jar is nodig voor de Junit test klassen.



Stap 5. Resterende packages in R installeren.

Het is ook nodig dat de volgende libraries in R gedownload/geïnstalleerd worden. Dit kan je doen door de R console te gebruiken.

R Package name	Used for:	How to install in R
rJava	Interface tussen R en Java	<code>install.packages("rJava")</code>
JavaGD	R plots in Java tonen	<code>install.packages("JavaGD")</code>

Let op, de gedownloade packages moeten in de map "<R installatie_pad>\library" geplaatst worden. Hetzelfde geldt voor afhankelijke libraries. Vaak worden er meerdere libraries met een enkel commando gedownload. Zorg ervoor dat al deze mappen in de "library" folder van R zitten.

(beknopte) R-Handleiding

Inhoud

1	De programmeertaal R	2
1.1	De R omgeving	2
1.2	De R console	2
1.3	De R werkruimte	3
1.4	De R packages	3
2	Basis R	3
2.1	Operaties en Data types	3
2.2	Objecten en variabelen	4
2.2	Vectoren	5
2.3	Matrices en arrays	5
2.4	Dataframes	6
2.5	Lijsten	6
2.6	Karakter Expressies en Indexeren van een lijst	6
2.7	Logische Expressies en vergelijkingsoperatoren	7
2.8	Converteren van data types	7
2.9	Functies	8
2.10	De controles en iteraties	9
3	R tools met een grafische interface	11
3.1	RStudio	11
3.2	R Commander	12
3.3	Imputation met de Amelia package	12

1 De programmeertaal R

R is een gratis, open source, interactieve technische ‘computing software’ die een grote verzameling van statistische hulpmiddelen en grafische componenten bevat. Verschillende sites zijn gewijd aan deze programmeertaal. De site <http://www.r-project.org/> biedt een uitgebreide beschrijving van de taal R, toegang tot de verschillende bibliotheekfuncties en de documentatie. De gecompileerde versies R zijn beschikbaar voor Linux, Windows en Mac OS X.

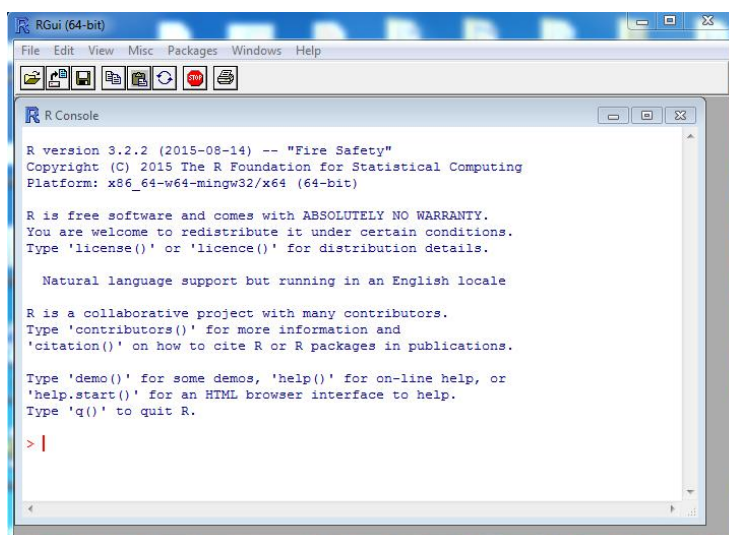
De R programmeertaal is een uitloper van een programmeertaal taal genaamd S¹. Het werd ontwikkeld door Ross Ihaka en Robert Gentleman van de Universiteit van Auckland, Nieuw-Zeeland. Het was in de eerste plaats door statistici aangenomen en is nu een van de standaarden voor de statistische computing². De taal R is een van de programmeertalen die voor de statistische analyse, gebruikt wordt, vooral in de wetenschappelijke en academische gemeenschap.

1.1 De R omgeving

R is eenvoudig te installeren³ op Windows, Linux, en Apple computers die op het internet zijn aangesloten. Ga naar de webpagina <http://cran.r-project.org/>, klik op de link van het besturingssysteem en volg de instructies voor het downloaden en installeren van de nieuwste versie van R. Nieuwe versies van R worden regelmatig uitgebracht. De taal R, wordt uitgebreid met vele packages die door de R gemeenschap van ontwikkelaars geschreven zijn.

1.2 De R console

Wanneer R eenmaal is geïnstalleerd, kan R gestart worden. Achter de opdrachtprompt, aangegeven door “>”, kan men instructies invoeren en vervolgens uitvoeren door op Enter te drukken.



R is een geïnterpreteerde taal. Wanneer men een syntactische volledige instructie achter de opdrachtprompt “>” invoert en op <Enter> drukt, zal R de instructie onmiddellijk evalueren en krijgt de gebruiker de opdrachtprompt terug. Wanneer men een “+” terugkrijgt in plaats van een “>”, dan

¹ S is a statistical programming language

² <http://www.inside-r.org/why-use-r>

³ Zie de bijlage R-installatie handleiding

betekent dit dat de instructie niet volledig is. R wacht in dit geval op het vervolg van de instructie. Vaak zal dit gebeuren als men per ongeluk een afsluitend haakje weglaat.

Voor uitgebreidere informatie over R en de geschiedenis, refereer ik naar:

<http://www.r-project.org/> en [http://en.wikipedia.org/wiki/R_\(programming_language\)](http://en.wikipedia.org/wiki/R_(programming_language)).

1.3 De R werkrumte

De R werkrumte bestaat uit objecten die door de gebruiker zijn gedefinieerd (variabelen, vectoren, matrices, gegevensframes). Hieronder zijn er een aantal standaard commando's in de tabel opgenomen die nuttig zijn voor het beheer van de werkrumte:

Command	Description
➤ <code>getwd()</code>	# print the current working directory
➤ <code>setwd(new.dir)</code>	# set the working directory
➤ <code>ls()</code>	# list the objects in the workspace
➤ <code>help(options)</code>	# list of available options
➤ <code>load("myfile.RData")</code>	# load a saved workspace into the current session
➤ <code>q(save="no")</code>	# quit R > q()
➤ <code>save.image()</code>	# save the workspace to the file ".RData" (default)

1.4 De R packages

De packages bestaan uit een reeks van voorgeprogrammeerde functies en zijn ontwikkeld voor het uitvoeren van specifieke taken. Er zijn twee soorten packages:

- packages die met de basis installatie van R komen (R base).
- packages die te downloaden zijn en handmatig geïnstalleerd dienen te worden.

```
# Installeren van een package
> install.Packages(bnlearn)

# Inladen van bv. de package "bnlearn"
> Library(bnlearn)
> require(bnlearn)

# Help over een package aanroepen
> help(bnlearn)
```

2 Basis R

2.1 Operaties en Data types

We kunnen R gewoon als rekenmachine gebruiken. Met behulp van een rekenmachine kunnen we waarden bewerken door gebruik te maken van operatoren, zoals machtsverheffen (^), vermenigvuldigen (*), delen (/), optellen (+), en aftrekken (-). In tabel 1 wordt een overzicht gegeven van operatoren die R ondersteunt.

De basis aritmetische en de logische operaties, met de rangorde van hoog naar laag:

Operatie	Beschrijving - Prioriteit van Hoog naar laag	Associativiteit
\wedge	# Exponent	Right to Left
$-x, +x$	# Unary minus, Unary plus	Left to Right
:	# Sequence	Left to Right
%naam%	# Modulus	Left to Right
*, /	# Multiplication, Division	Left to Right
+, -	# Addition, Subtraction	Left to Right
<, >, <=, >=, ==, !=	# Comparisons	Left to Right
!	# Logical NOT	Left to Right
&, &&	# Logical AND	Left to Right
,	# Logical OR	Left to Right
->, ->>	# Rightward assignment	Left to Right
<-, <<-	# Leftward assignment	Right to Left
=	# Leftward assignment	Right to Left

Tabel 1. Prioriteitsvolgorde van R-operatoren van Hoog naar Laag

Het is mogelijk dat verschillende operaties dezelfde rangorde hebben. In een dergelijk geval wordt de volgorde van uitvoering door de associativiteit bepaald zoals in de tabel 2 is aangegeven.

2.2 Objecten en variabelen

Een variabele in R is een object dat verschillende typen kan aannemen: Numeric, Complex, Logical en Character. Een variabele is hoofdlettergevoelig en begint altijd met een letter. Met behulp van een pijltje '<-' of '=' kunnen we een waarde aan een variabele toekennen.

Met behulp van een pijltje '<-' of '=' kunnen we een waarde toekennen aan een variabele.

```
> x = 5
> y <- 10
> y + x
> x
```

- De '**mode**' volgens de basisstructuur van het object. (Mutual exclusive classification)
 - numeric, complex, character, logical.
- De '**classe**' van een object dat bepaalt hoe generieke functies hiermee opereren.
 - vector, matrix, data.frame, array, list. (Property assigned)

Gegeven een multidimensionaal object M die numerieke waarden bevat. De class van M is 'matrix' en de mode is 'numeric'.

Voorbeeld: creëer een simpele vector en controleer zijn "mode" en "class":

```
> x <- 1:4
> mode(x)
[1] "numeric"
> class(x)
[1] "integer"
```

```
> x <- 1:4
> dim(x) <- c(2,2)
> mode(x)
[1] "numeric"
> class(x)
[1] "matrix"
```

Note: Als er geen specifieke klasse aan een object is toegewezen, dan is default gelijk aan de mode. Bijvoorbeeld het gewone getal 5 (geen vector) is numeriek. Het getal 5 behoort niet tot een 'classe' en hierdoor is de 'classe' ook 'numeric'.

2.2 Vectoren

Degenen onder ons die een cursus lineaire algebra hebben gevolgd, hebben geleerd dat een getal (bv. 5) op zichzelf staand kan voorkomen. We spreken van een *scalair*. Getallen kunnen ook gegroepeerd zijn als een lijst van getallen. Dit kan op twee manieren: een *vector* of een *matrix*. Het onderscheid tussen een ‘scalair’ en ‘vector’ wordt niet gehandhaafd in R. Een scalair is een vector met lengte 1. Denk aan een vector met lengte n als een stapel van n cellen die gevuld worden met allerlei objecten of waarden.

Vectoren zijn het belangrijkste onderdeel van R. De elementen van een vector moeten dezelfde *mode* hebben.

Voor het maken van een vector maken we gebruik van de volgende drie functies:

- `c` van ‘concatenate’, neemt een groep van objecten als invoer en zet ze in een vector.
- `seq` van ‘Sequence Generation’ of
- `rep` (Replicate Elements of Vectors and Lists)

```
> x1 <- c(1,2,3,4)           # Vector van getallen
> y1 <- c("man" , "vrouw")   # Vector van karakters/strings

> eenTot10_a <- seq(from=1, to=10, by=1)
> eenTot10_b <- rep(1, times=10)
> eenTot3_tweekeer <- rep(seq(from=1, to=3, by=1), times=2)
```

Als twee vectoren evenveel waarden hebben, kunnen wij kolomsgewijs operaties uitvoeren:

```
> eenTot10_a * 10             # Elk element vermenigvuldigen met 10
> eenTot10_a + eenTot10_b     # Sommeren van elementen in vectors
> eenTot10_a[3]               # Het derde element
> eenTot10_a[-3]              # Alle elementen behalve het derde element
```

Als twee vectoren in een operatie zijn betrokken en niet dezelfde lengte hebben, zal R toch de operatie uitvoeren door de kortste vector te repliceren tot de lengte van de langste vector. Dit wordt **Recycling genoemd**. R geeft in dit geval wel een waarschuwing.

```
> a <- 1:10
> a + 1
[1] 2 3 4 5 6 7 8 9 10 11

> a <- 1:10
> a + c(1,2,3)
[1] 2 4 6 5 7 9 8 10 12 11
longer object length is not a multiple of shorter object length
```

2.3 Matrices en arrays

Een array is een multidimensionale uitbreiding van een vector. Zoals vectoren, moeten alle elementen van een array van dezelfde *modus* (mode) zijn. Arrays hebben een attribuut genaamd ‘*dim*’ (**dimension**). De meest voorkomende arrays die in R worden gebruikt is de *matrix*. Het is een 2-dimensionale array. Matrix heeft twee extra attributen dan een vector: het aantal rijen en het aantal kolomen.

```
> matrix(c(1:9))           # c(1:9) is een vector 1:9
> matrix(c(1:9), nrow=3, byrow=TRUE) # vul de rij eerst
> matrix(c(1:9), nrow=3, byrow=FALSE) # vul de kolom eerst
> matrix(c(1:9), nrow=3, ncol=3)      # Default byrow=TRUE
```

Als twee matrices evenveel waarden hebben, kunnen wij zoals bij vectoren aritmetische operaties uitvoeren.

Matrices worden opgeslagen intern als een vector. De kolommen zijn op elkaar, van boven opgesteld.

2.4 Dataframes

Het is op zich wel mogelijk om tekst en getallen samen in de vorm van een matrix op te slaan, maar dat noemen we in R dan een data frame. Bijvoorbeeld een data matrix zoals we die ook buiten R zouden kunnen maken met Excel of in NotePad of een andere platte tekst editor, met in de eerste rij de namen van de variabelen, en in de overige rijen de data waaruit die variabelen bestaan:

```
> x <- c("lengte", "hoogte", "breedte")
> y <- c(4, 5, 6)
> z <- c(7, 8, 9)
> d <- data.frame(x, y, z)
> d
```

LET OP: de vectoren x, y, en z worden kolomsgewijs aan elkaar geplakt tot enkele data frames. De elementen in de kolommen van een data frame moeten van het zelfde type zijn (vector). Dus in de eerste kolom staan alleen maar woorden, en in de overige kolommen alleen maar getallen. Je kunt in één en dezelfde kolom niet getallen en tekst mengen, om dezelfde reden waarom dat in een vector ook mis gaat. De getallen worden dan namelijk automatisch naar tekst omgezet.

Data frames worden vaak gebruikt om data in een matrix vorm op te slaan omdat ze verschillende ‘mode’ toelaten.

2.5 Lijsten

Lijsten in R hebben veel met vectoren gemeen. Ze kunnen wel objecten van verschillende types bevatten, zoals dat bij vectoren niet mogelijk is. Ze zijn generieke vectoren die elementen van verschillende ‘mode’ toelaten.

```
> y <- list(naam="Piet", geslacht="M", bedrijf="ING")
> y
      # $naam      [1] "Piet"
      # $geslacht  [1] "M"
      # $bedrijf   [1] "ING"

> y$naam      # $naam      [1] "Piet"
```

2.6 Karakter Expressies en Indexeren van een lijst

Wij hebben gezien hoe vectoren werken. Vectoren kunnen cijfers of logische resultaten bevatten. Een andere vorm van gegevens die we vaak met R karakter moeten verwerken, is data. We duiden karakter van gegevens door het plaatsen van het karakter string (woord(en)) tussen aanhalingstekens.

```
> namen <- c("Piet", "Jan", "Kees", "Jaap")
> namen
```



```
[1] "Piet" "Jan" "Kees" "Jaap"

> studenten <- c("Piet",2012)
> studenten
[1] "Piet" "2012"
```

Merk op hoe de 2012 door R is omgezet in “2012”. R realiseert dat de twee elementen zijn van verschillende types. Het is veel makkelijker om 2012 naar een tekenreeks om te zetten dan “Piet” naar een nummer te converteren. Het benaderen van de elementen van een lijst gebeurt via “subscript []”:

```
> studenten[2]
[1] "2012"
```

2.7 Logische Expressies en vergelijkingsoperatoren

Een aanzienlijk deel van ‘computer programming’ omvat hoe bepaalde voorwaarden worden berekend of eraan voldoen. Vaak worden beslissingen genomen door het vergelijken van hoeveelheden, aantallen...etc. Denk aan de volgende expressie en het resultaat:

```
> 4 > 3
[1] TRUE
```

Dergelijke evaluaties kunnen ook worden gevectoriseerd. Neem het volgende voorbeeld.

```
> x <- -2 : 5          # x tussen -2 en 5
> x
[1] -2 -1 0 1 2 3 4 5
> x > 0                # x > 0
[1] FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE
```

De gelijkheid wordt met de operator ‘==’ geëvalueerd. Gegeven de definitie van ‘x’ hier bovenaan:

```
> y <- x == 2
> y
[1] FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE
```

De “ongelijk” operator is !=.

```
> x != 3
[1] TRUE TRUE TRUE TRUE TRUE FALSE TRUE TRUE
```

2.8 Converteren van data types

R objecten kunnen naar een andere ‘mode’ of ‘class’ geconverteerd worden om het gedrag van een een R object te veranderen. De functies om conversies uit te voeren beginnen met de prefix “as.”

Voorbeelden: as.numeric(), as.character(), as.matrix(), as.data.frame()

```
> n <- 10:14
> n
#> [1] 10 11 12 13 14
```

```
# Numeric to Character
> c <- as.character(n)

# Numeric to Factor
> f <- factor(n)
# 10 11 12 13 14
# Factor to Character

> as.character(f)
#> [1] "10" "11" "12" "13" "14"
```

2.9 Functies

Eén van de sterke punten van de R is dat de gebruiker in staat is om nieuwe functies toe te voegen. In R zijn er veel functies die feitelijk functies van functies zijn. De structuur van een functie volgt hieronder.

```
myfunction1 <- function(arg1, arg2, ... ){
  statements
}

# VOORBEELD:
# function printHallo - krijgt een naam als parameter (naam)
# en print Hello + <naam>

printHallo <- function(naam) {
  print "Hallo" & " " & naam
}
```

Functie aanroep: argument of named argument

```
> msg <- printHallo("Piet")
> msg <- printHallo(naam = "Piet")
> msg
[1] "Hallo Piet"
```

Functie met return waarde

In veel gevallen zullen wij van een functie eisen om verwerkingen uit te voeren en het resultaat terug te geven. Dit wordt gerealiseerd door de functie `return()`. De geretourneerde waarde kan elke geldige type van R zijn.

```
myfunction2 <- function(arg1, arg2, ... ){
  statements
  return(object)
}

# VOORBEELD:
# function som - krijgt twee getallen en retourneert de som.

som <- function(x, y) {
  return (x+y)
}
> som
```

Functie met functionele programmeertaal style

In R kan men ook een aantal constructies/stylen gebruiken zoals bij functionele programmeertalen. Een voorbeeld hiervan zijn de *recursieve functies en infix notatie*:

```
recursive.factorial <- function(x) {
  if (x == 0)      return (1)
  else             return (x * recursive.factorial(x-1))
}
```

De meeste operatoren die wij in R gebruiken zijn binaire operatoren (hebben twee operanden). Ze zijn ook infix operatoren genoemd en worden gebruikt tussen twee operanden.

```
> 5 + 4      # of wel:
> `+` (5,4)
```

Men kan eigen infix operatie maken. Dit is gedaan door de naam van de functie tussen “%” symbool te plaatsen.

```
`%IsDeelbaar%` <- function(x,y)
{
  if (x%y ==0) return (TRUE)
  else        return (FALSE)
}
```

```
> 10 %IsDeelbaar% 3
[1] FALSE
```

Een aantal , dat men in overweging moet nemen bij het aanmaken van een infix functie:

- Ze moeten starten en eindigen met %
- Ze moeten aangesloten worden door “`” (back tick `)
- Geen speciale symbolen gebruiken. Gebruik escape om deze te kunnen gebruiken.

Variabelen bereik

Objecten in een functie zijn lokaal gedeclareerd. Het object kan elke waarde teruggeven die tot de data typen van R behoort.

Functie code bekijken

In R, kan de code van een functie bekeken worden door het intikken van de naam van de functie zonder het (), voorbeeld `printHallo`.

2.10 De controles en iteraties

Hoewel R is opgebouwd rond operaties op matrices en vectoren, bevat het nog steeds een krachtige set van tools genaamd loops en statements die je zult tegenkomen in veel andere programmeertalen. Dit zijn de fundamentele controle-flow constructies van de R taal.

if statement, if...else statement en de geneste if...else statement

Als de test `test_expression` WAAR is, wordt de instructie “statement” uitgevoerd anders gebeurt er niets.

```
if (test_expression) {  
  statement  
}
```

Als de test `test_expression` WAAR is, wordt de `statement1` uitgevoerd anders wordt `statement2` uitgevoerd.

```
if (test_expression) {  
  statement1  
} else {  
  statement2  
}
```

Alleen één statement wordt uitgevoerd gelang het resultaat van de `test_expressie`:

```
if ( test_expression1) {  
  statement1  
} else if ( test_expression2) {  
  statement2  
} else if ( test_expression3) {  
  statement3  
} else  
  statement4
```

for loop

```
for (value in sequence) {  
  statement  
}
```

De “sequence” is een vector en de “value” neemt in elke iteratie een waarde aan. In elke iteratie wordt de instructie “statement” geëvalueerd.

while loop

```
while (test_expression) {  
  statement  
}
```

De “test_expression” wordt geëvalueerd en indien deze WAAR is, wordt de body uitgevoerd. Na de uitvoering van de instructie “statement”, wordt de controle weer aan While teruggegeven en wordt de test_expressie weer geëvalueerd. Dit wordt herhaald totdat de test_expressie ONWAAR wordt.

repeat loop

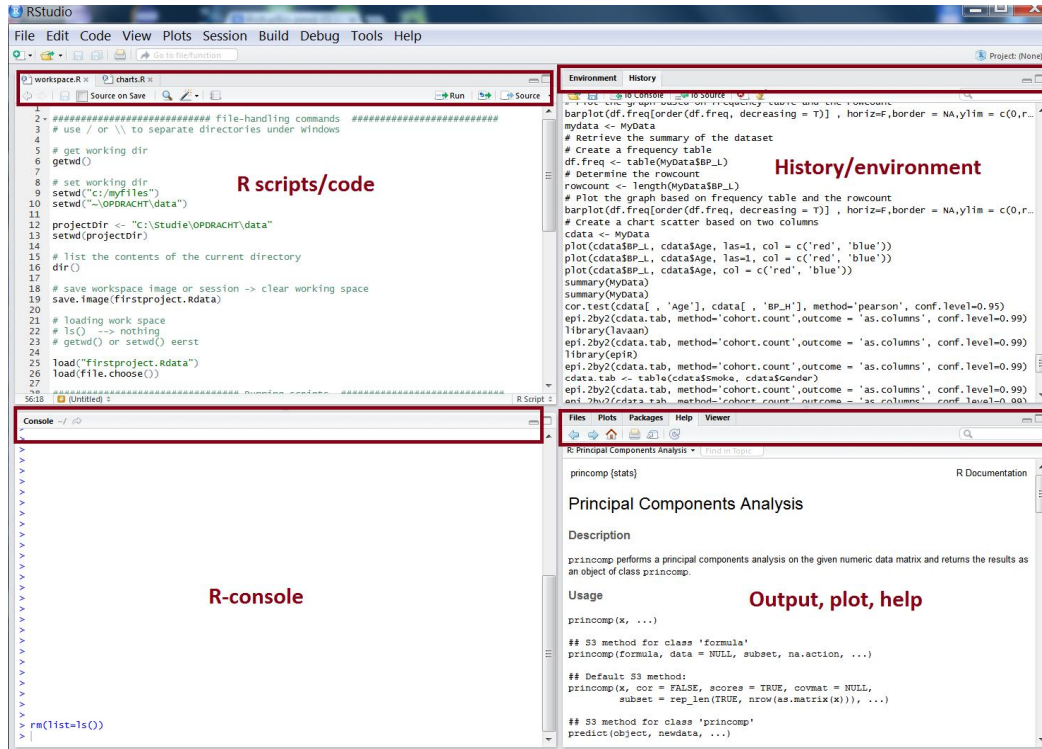
```
repeat {  
  statement  
  if (test_conditie){  
    break  
  }  
}
```

De “statement” wordt herhaald, totdat een conditie waar wordt. In dit geval moet de “repeat” gestopt worden door gebruik te maken van “break”.

3 R tools met een grafische interface

3.1 RStudio

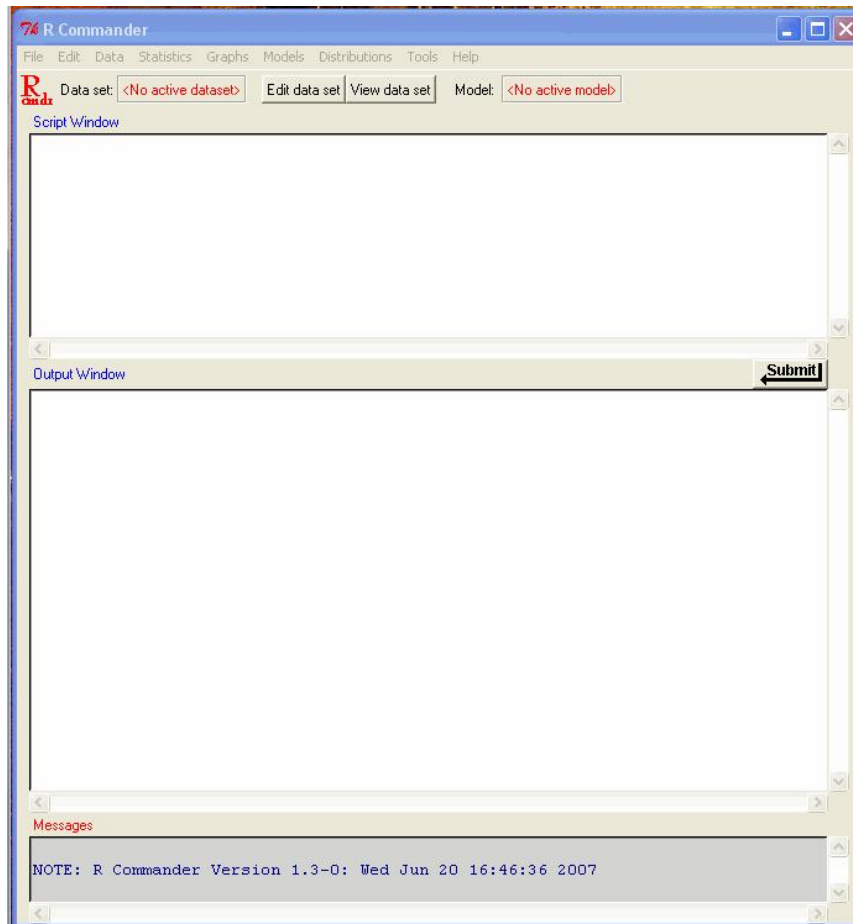
Voor meer informatie: <https://www.rstudio.com/>



3.2 R Commander

Voor meer informatie: <http://socserv.mcmaster.ca/ifax/Misc/Rcmdr/>

- 1) `install.packages("Rcmdr")`
- 2) Start Rcmdr: `library(Rcmdr)`



3.3 Imputation met de Amelia package

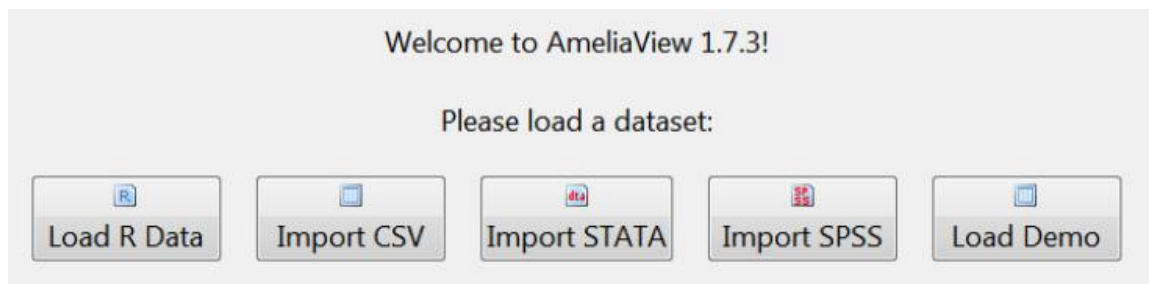
Voor meer informatie: <https://cran.r-project.org/web/packages/Amelia/Amelia.pdf>

Imputing betekent dat de ontbrekende waarden in een dataset door een schatting van de werkelijke waarde automatisch worden vervangen. Amelia is een R package die dit proces faciliteert. Men moet eerst de package in R installeren, inladen en dan starten:

- ```
> install.packages("Amelia")
> library(Amelia)
> AmeliaView()
> Voorbeeld dataset
```



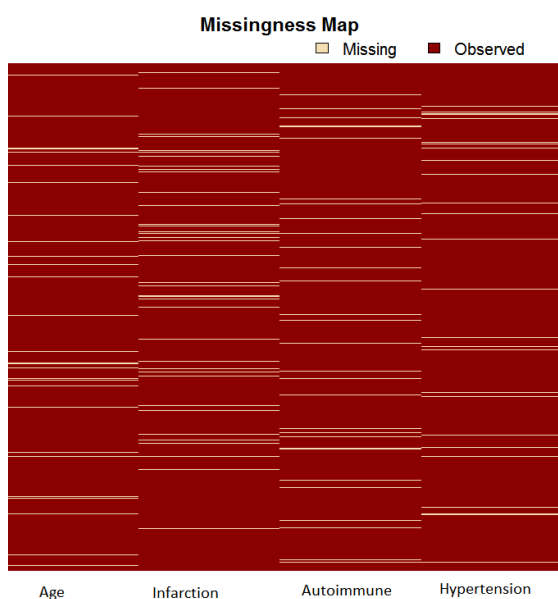
data1.csv



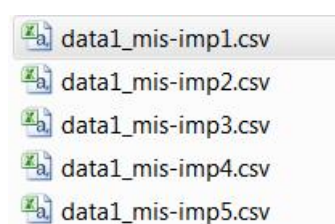
Amelia geeft een samenvatting van missing values: In dit geval de NA-waarden voor

- Auto\_immune disease
- Hypertension
- Myocardial\_infarction

| Variable              | Min      | Max | Mean    | SD     | Missing    |
|-----------------------|----------|-----|---------|--------|------------|
| Auto immune disease   | 0        | 1   | 0.04985 | 0.2176 | 1515/30000 |
| Age                   | (factor) |     |         |        | 1546/30000 |
| Hypertension          | 0        | 1   | 0.2621  | 0.4398 | 1458/30000 |
| Myocardial_infarction | 0        | 1   | 0.1595  | 0.3661 | 1534/30000 |

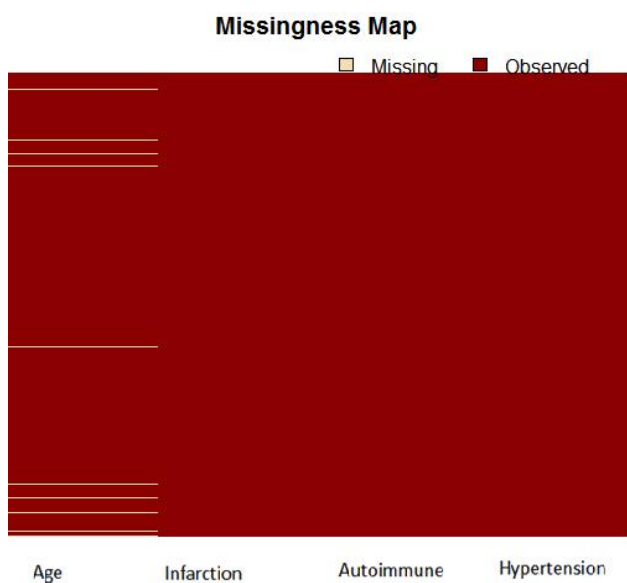


Wanneer men de "Impute" knop drukt en de imputation is met succes afgerond, krijgt men vijf mogelijke imputations/iteraties te zien als vijf excel bestanden:



Open de data1\_mis\_imp5.csv in Amelia:

| Load Session       | Save Session | Plot Histogram | Edit Data | Missingness Map | Impute!    |
|--------------------|--------------|----------------|-----------|-----------------|------------|
| Variable           | Min          | Max            | Mean      | SD              | Missing    |
| Auto immune di     | 0            | 1              | 0.04864   | 0.2151          | 7/30000    |
| Age                | (factor)     | ...            | ...       | ...             | 1546/30000 |
| Hypertension       | 0            | 1              | 0.2625    | 0.44            | 7/30000    |
| Myocardial_infarct | 0            | 1              | 0.1608    | 0.3674          | 7/30000    |



Na de imputation, kan men de “diagnostics” over een specifieke variabele bekijken:

