Análise Avançada de Dados Quantitativos 2019-1

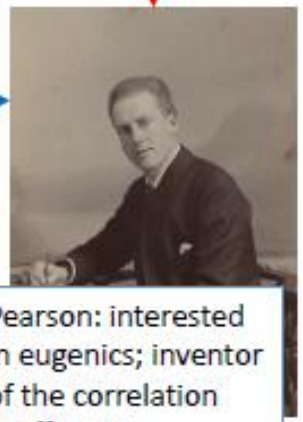
Prof. Dr. Wagner de Lara Machado

Founders of statistics

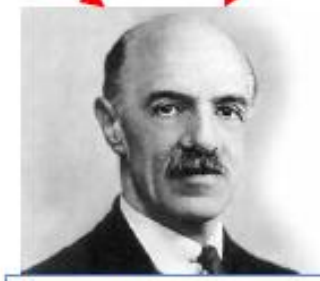
Galton: interested in individual differences, and the heretability of intelligence; founder of eugenics.

Wundt: interested in the general mind; used N=1 research, with replication and introspection

Fisher: realized we need to take sample fluctuation into account; F-test, ANOVA; hypothesis testing.



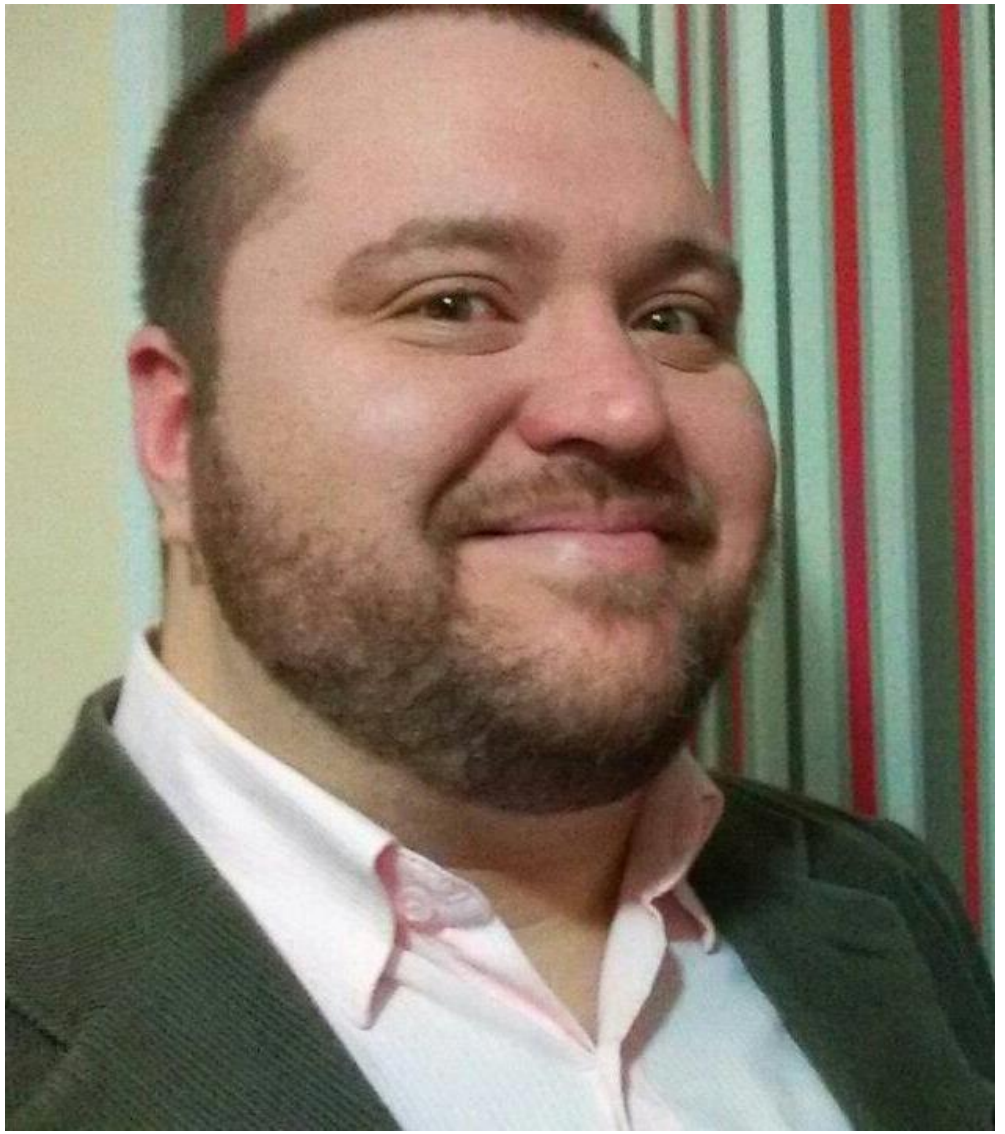
Pearson: interested in eugenics; inventor of the correlation coefficient.



Spearman: realized measurements contain measurement error; developed factor analysis to correct for this.

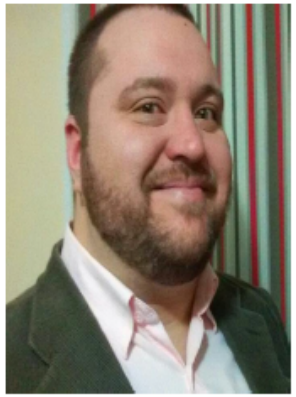
Gosset/Student: t-test

Quem sou eu?



Formação acadêmica e atuação profissional

Nível	**Tema**	**Ano de conclusão**
Graduação em Psicologia (ULBRA)	Prevalência de ansiedade e depressão em docentes	2007
Mestrado acadêmico (UFRGS)	Adaptação da Escala de Bem-estar Psicológico	2010
Doutorado (UFRGS)	Saúde mental positiva Adaptação de medidas de saúde mental	2013
Estágio de Pós-doutorado (UFRGS)	Saúde mental positiva Centro de análise de dados (CAD)	2015
Professor da graduação e pós-graduação *stricto sensu* Psicologia	Grupo de pesquisa Avaliação Psicológica do Potencial Humano	2015 a 2018



GRUPO DE PESQUISA
AVALIAÇÃO EM BEM-ESTAR E SAÚDE MENTAL

Grupo de Pesquisa Avaliação em Bem-Estar e Saúde Mental

PQ	h	i10	M
-	11	42	Membro da Diretoria da Associação Brasileira de Psicologia Positiva

Links

- https://scholar.google.com.br/citations?user=fH6qCDoAAAAJ&hl=en
- https://github.com/wagnerLM

temperament
relationship
diversity
organizational
version
teachers
undergraduate
adaptation
development
mini
form
depression
responsiveness
simulation
justice
regulatory
level
sexual
validity
prosocial
statistical
people
factorial
continuum
partner
inventory
styles
evidences
evidence
stigma
criterion
effects
validation
study
battery
attempted
risk
construct
student
pss
inferences
officers
adulthood
adolescents
item
scale
personality
teaching
assessing
od
increased
related
psychological
brazil
low
gender
short
items
big
factor
theory
models
inf
mental
making
correlates
stress
group
brazilian
health
work
sample
school
assessment
questionnaire
interactions
neuropsychological
trait
principal
engagement
experience
prejudice
young
metamemory
socioeconomic
perceived
structure
neupsilin
ghq
analysis
markers
mediators
anxiety
child
behaviors
teacher
convergent
dimensionality
confirmatory
model
positive
definition
difficulties
construction
response
portuguese
suicide
processes
revised

Análise Avançada de Dados Quantitativos

- https://github.com/wagnerLM/quantia

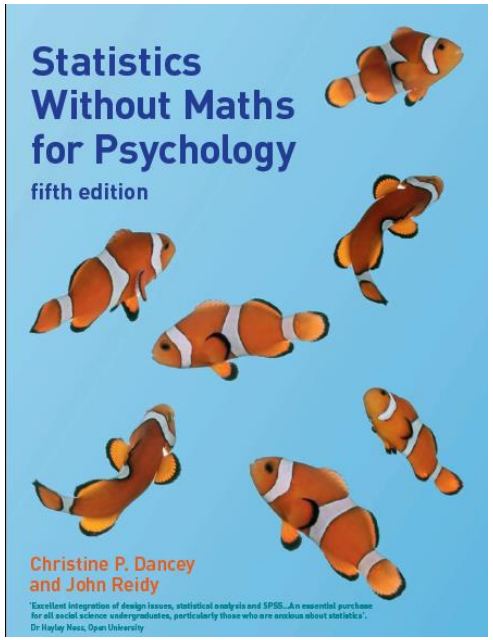
Livro texto base da disciplina

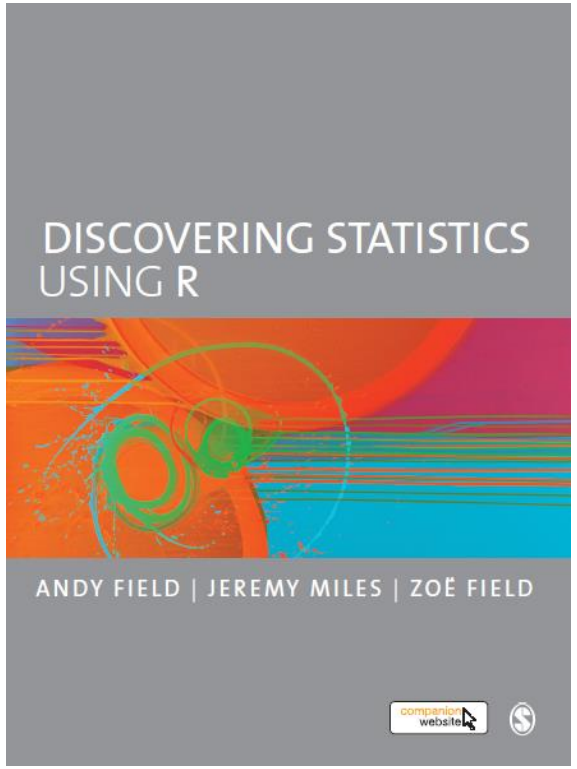
- Ver link no github
- https://raw.githubusercontent.com/wagnerLM/quantia/master/linklivro

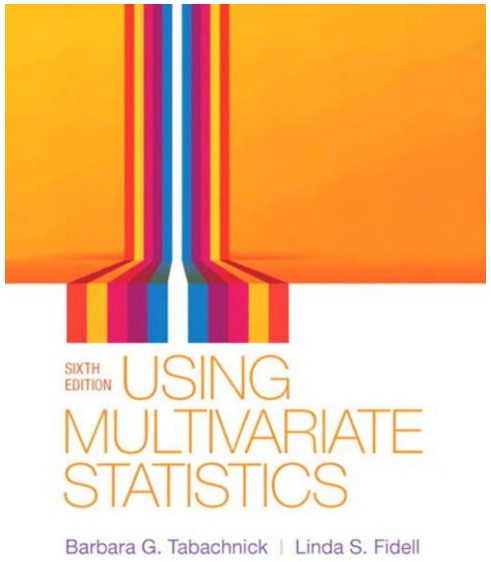
Livros auxiliares sobre R e estatística

- Acesse o link:

https://www.dropbox.com/sh/4kc8cjy51nkhuvr/AACLK0vJonv66OJSG3ZHId6Ra?dl=0







Tipos de variáveis

QUALITATIVAS		QUANTITATIVAS	
NOMINAIS OU CATEGÓRICAS	ORDINAIS	DISCRETAS	CONTÍNUAS
Sexo Cor do olho Profissão	NSE Nivel educacional Escalas likert	NÚMEROS NATURAIS (nº DE OBJETOS)	Altura Idade Tempo de reação

Níveis de mensuração

Nominal	Ordinal	Intervalar	Razão
Podemos atribuir números como rótulo, porém sem manter as propriedades dos números. Ex. Sexo 1 – Masculino 2 – Feminino	É possível estabelecer uma ordem de grandeza ou magnitude, porém a distância não é conhecida. Ex. Dureza dos materiais Nível educacional 1 – Fundamental 2 – Médio 3 – Superior 4 – Pós-graduação	A distância entre as unidades de medida é conhecida, podem ser realizadas a maioria das operações matemáticas. Ex. Escores modelados de instrumentos psicométricos	Escala intervalar com um zero conhecido e não arbitrário. Ex. Quando zero significa ausência do atributo

Questões sobre método

- Tipos (papéis) de variáveis
  - Confundidora
  - Controle
  - Dependente
  - Independente
  - Binária ou dicotômica
  - Covariável
  - Critério
  - “Dummy”
  - Endógena
- Exógena
- Latente
- Observada
- Mediadora
- Moderadora
- Preditora
- Desfecho
- Instrumental

Questões sobre método quantitativo

FREQUENTISTA	BAYESIANA
P(D|H)	P(H|D)

- Inferência estatística, teste de hipótese

Sendo D = dados e H = hipótese

- Interpretação do “p” ou Bayes Factor 0,4

Software R

- R Project

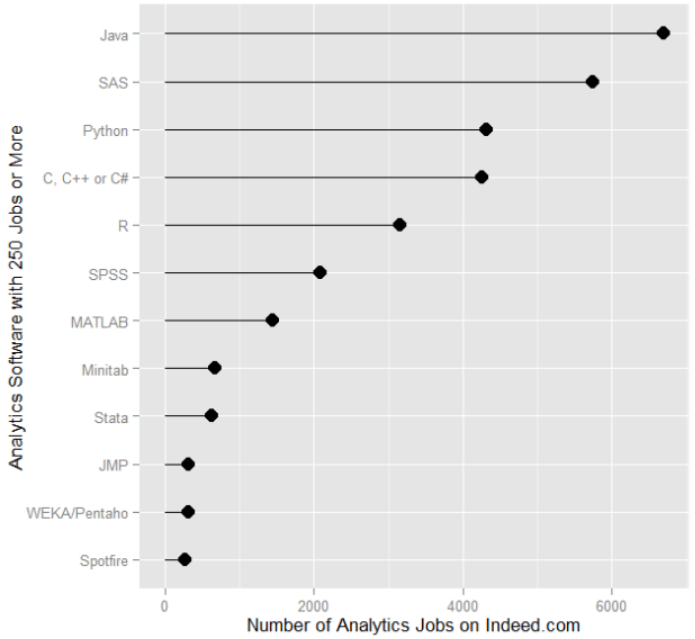
- John Chambers
- Software livre, colaborativo
- Possui um repositório CRAN
- Bibliotecas e pacotes (códigos e algoritmos)
- 100% FREE!!!

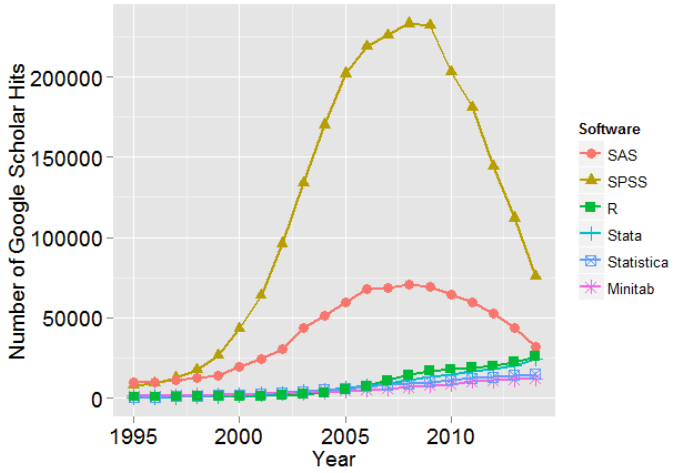
O quê é a linguagem r?

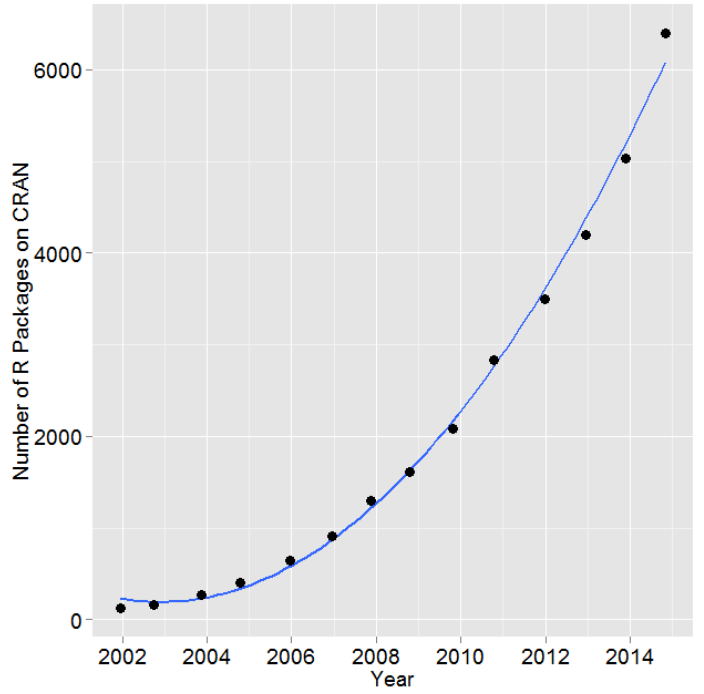
- R é um “dialeto” da linguagem S
- 1993 R é lançado ao público
- 1995 General Public License
- 1997 o R Core Group é formado – Controla a fonte de códigos do R
- 2000 R version 1.0.0 é lançado
- 2015 version 3.2.1

Porquê usar a linguagem r?

- Capacidades gráficas muito sofisticadas e melhores que muitos softwares
- Linguagem de programação que possibilita o desenvolvimento de novas ferramentas
- Comunidade de usuários muito ativa e participativa







Porquê usar a linguagem r?

Porquê usar a linguagem r?

stackoverflow

GitHub

Personality Project

Quick-R
powered by DataCamp

DataCamp

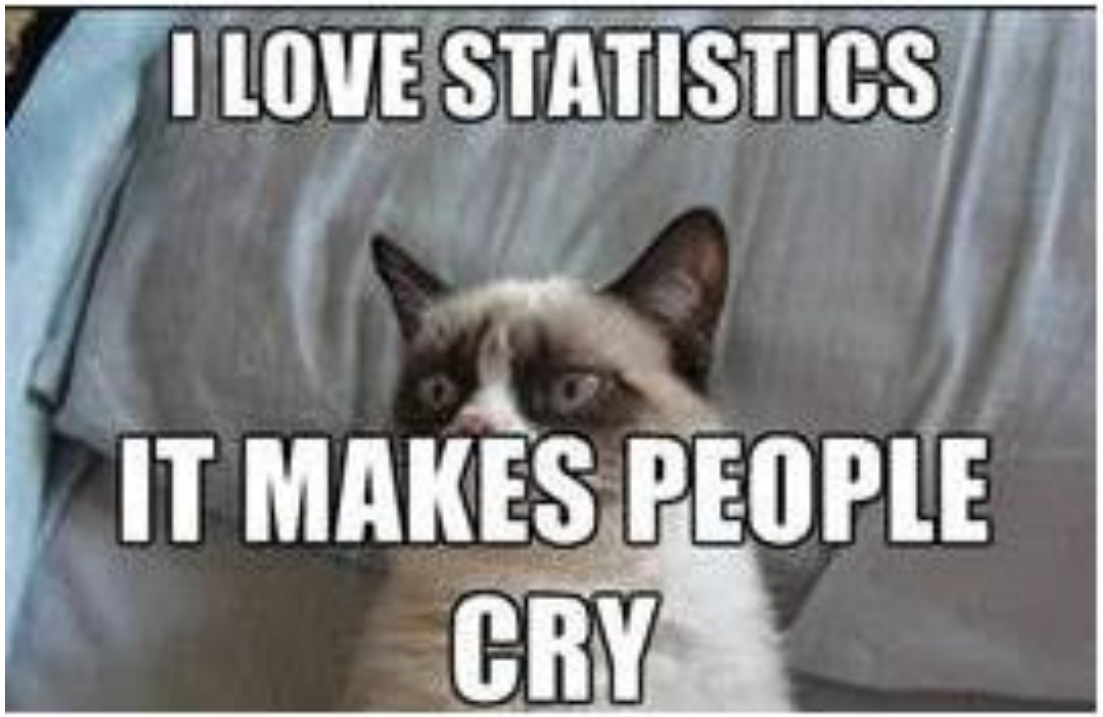
Passos inicias

- É o sistema básico – feito de forma colaborativa – Possui funções básicas que podem ser combinadas em pacotes mais avançados

- R Studio – Uma interface mais amigável para o uso do R. Permite análise, escrita (tem gente que escreve a tese por aqui), e publicações.

R desing

- O sistema R é divido em duas partes:
  - A “base” do sistema R que pode-se fazer o download no CRAN
  - Todo o resto – Funcionalidades chamadas de packages
  - Atualmente existem mais de 6000 packages



gibbs.cpp

Source on Save | Source

```cpp
#include <Rcpp.h>
using namespace Rcpp;

// [[Rcpp::export]]
NumericMatrix gibbs(int N, int thin) {

  NumericMatrix mat(N, 2);
  double x = 0, y = 0;

  for(int i = 0; i < N; i++) {
    for(int j = 0; j < thin; j++) {
      x = R::rgamma(3.0, 1.0 / (y * y + 4));
      y = R::rnorm(1.0 / (x + 1), 1.0 / sqrt(2 * x + 2));
    }
    mat(i, 0) = x;
    mat(i, 1) = y;
  }

  return(mat);
}
```

19:15 | gibbs(int N, int thin): NumericMatrix | C/C++

https://www.r-project.org/

[Home]

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Development Site
Conferences
Search

R Foundation

Foundation
Board
Members
Donors
Donate

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

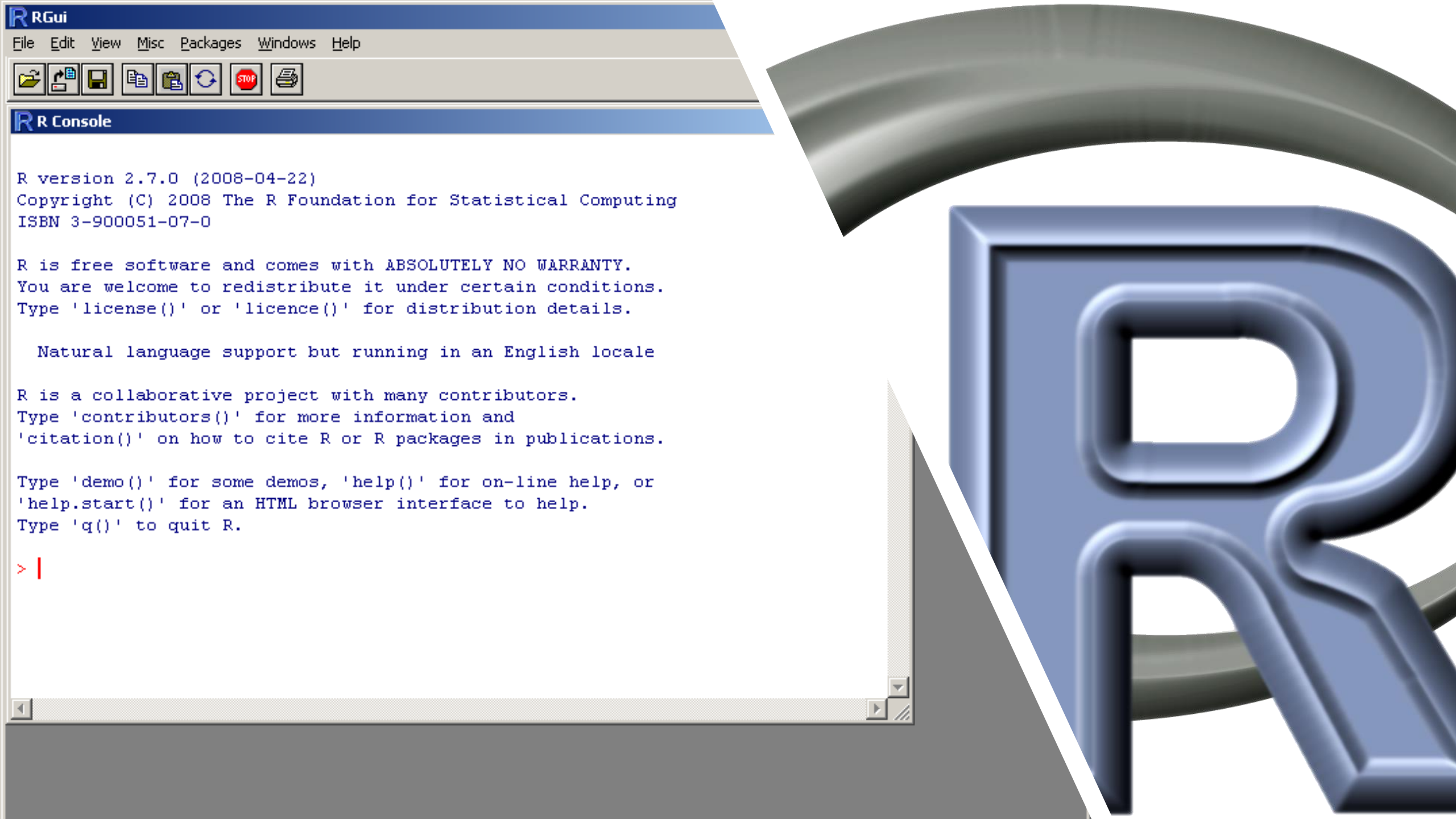
News

- **R version 3.5.1 (Feather Spray)** has been released on 2018-07-02.
- The R Foundation has been awarded the Personality/Organization of the year 2018 award by the professional association of German market and social researchers.
- **R version 3.5.0 (Joy in Playing)** has been released on 2018-04-23.

News via Twitter

The R Foundation Retweeted

Peter Dalgaard
@pdalgd
#rstats 3.5.1 "Feather Spray" is released (source version)


RGui
File Edit View Misc Packages Windows Help
R Console
R version 2.7.0 (2008-04-22)
Copyright (C) 2008 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
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

A (very) short introduction to R

Paul Torfs & Claudia Brauer

Hydrology and Quantitative Water Management Group

Wageningen University, The Netherlands

16 April 2012

https://www.dropbox.com/s/vppgdkx1u909r3n/Torfs%20Brauer%202012%20-%20introR.pdf?dl=0

https://swirlstats.com/

```
R version 3.4.3 (2017-11-30) -- "Kite-Eating Tree"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

>
```


RStudio
Products
Resources
Pricing
RStudio
Open source and enterprise-ready
professional software for R
Dow
D
Discover RS
Discove

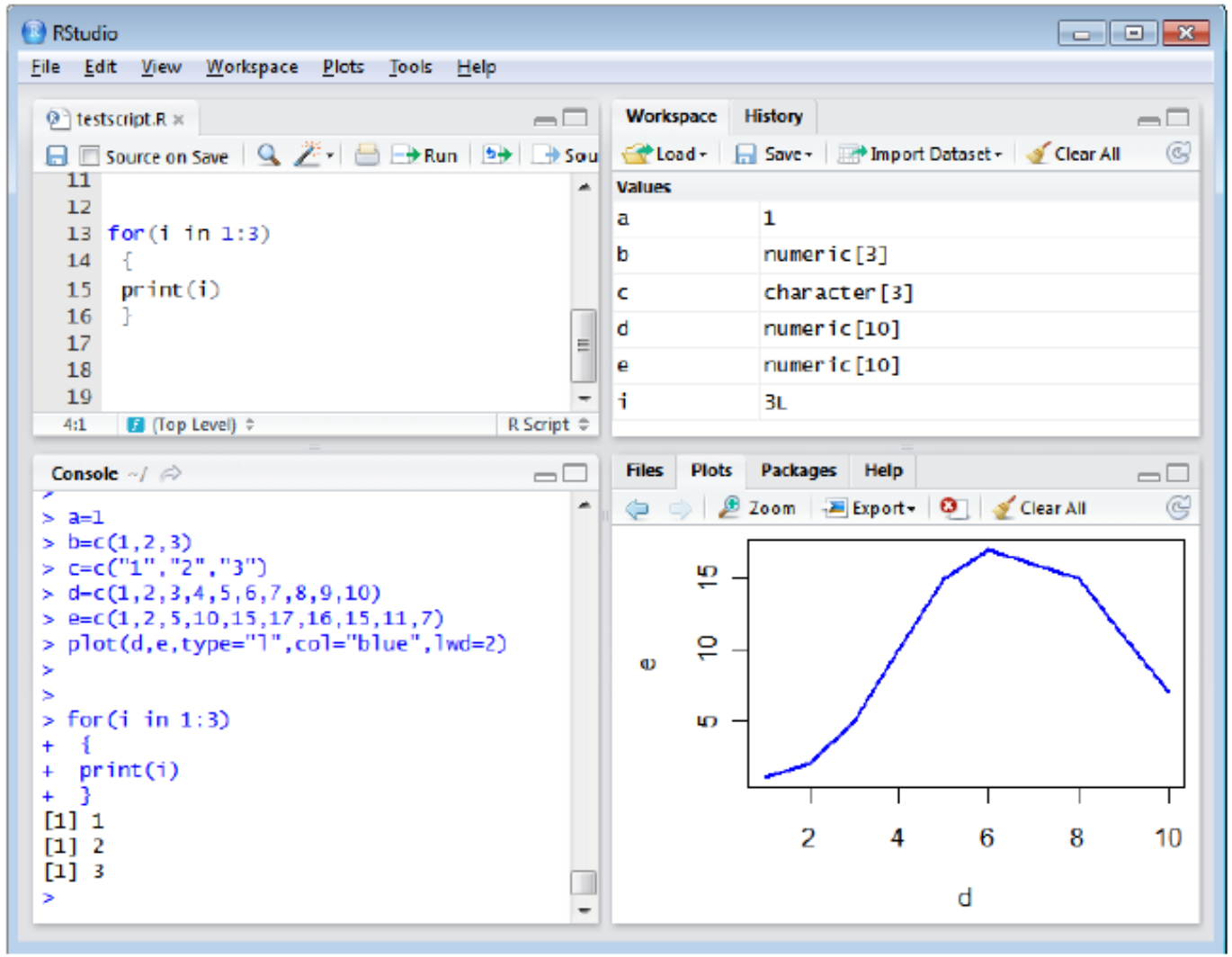


Figure 1 The editor, workspace, console and plots windows in RStudio.

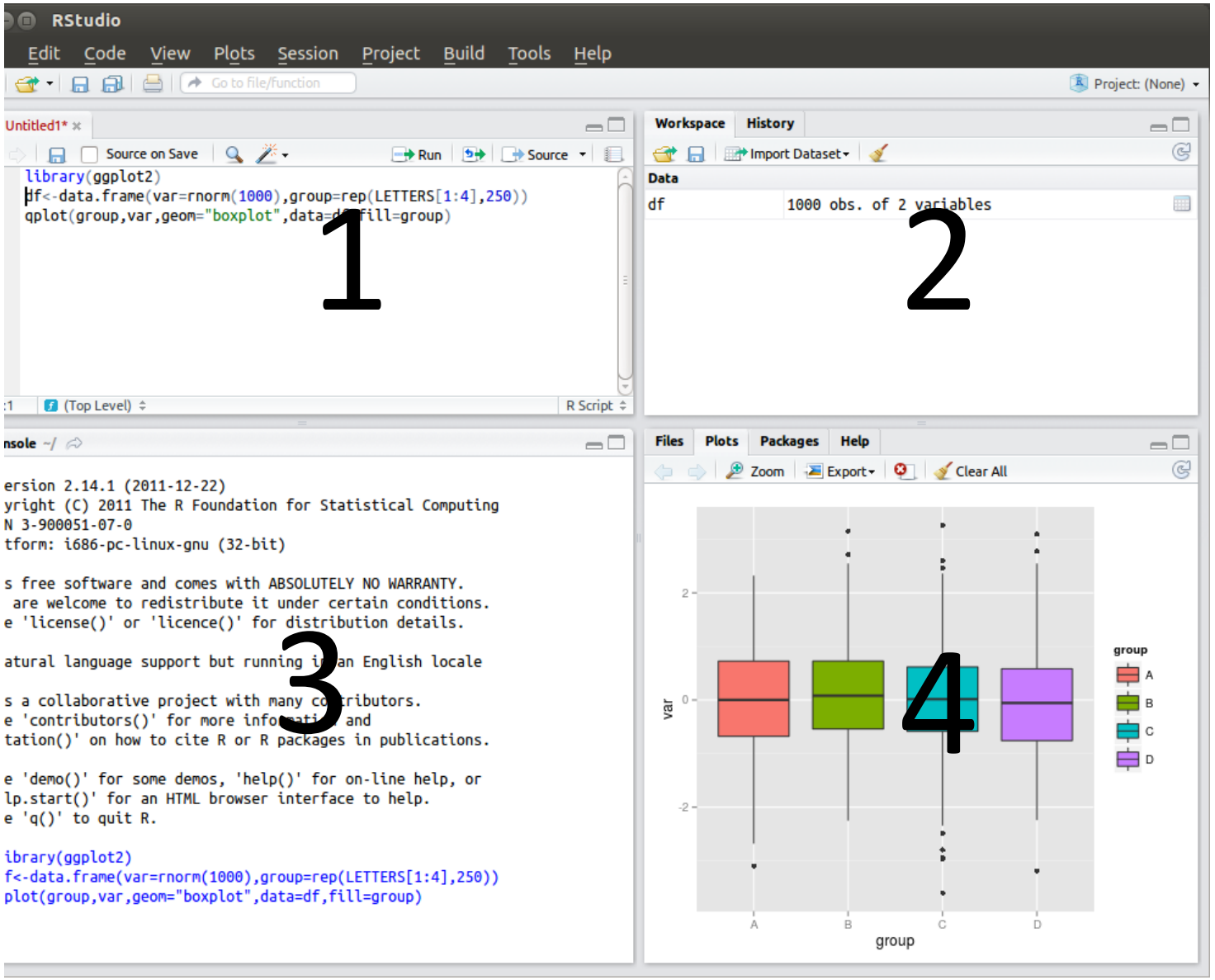
RStudio

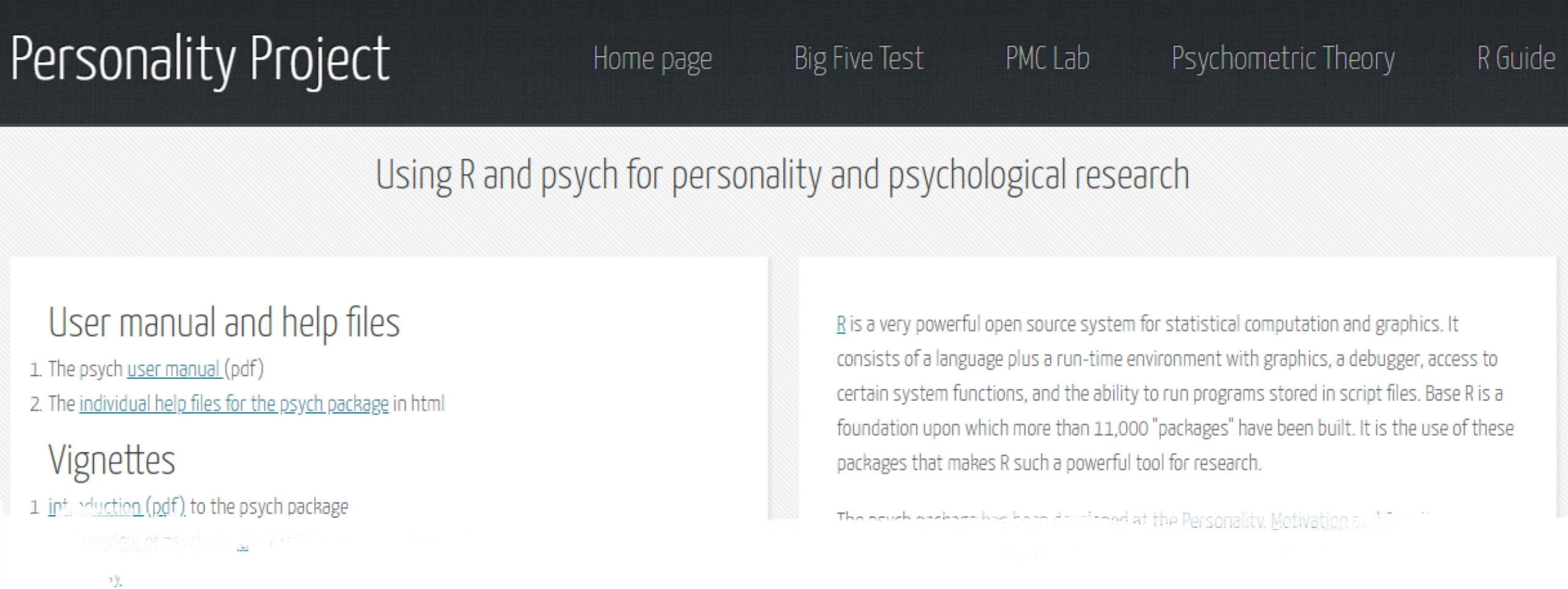
1 – script ou comandos; visualização de bancos de dados ou texto

2 – repositório de objetos e funções

3 – log de atividade

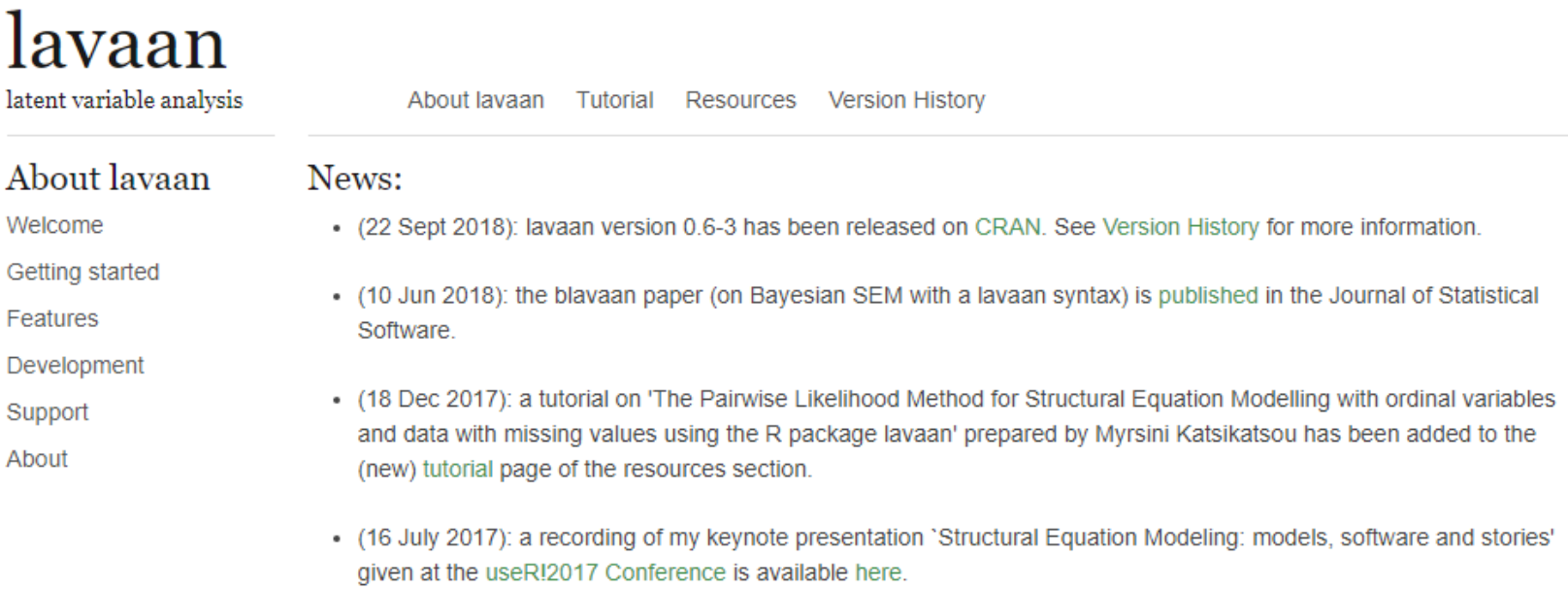
4 – gráficos e documentação



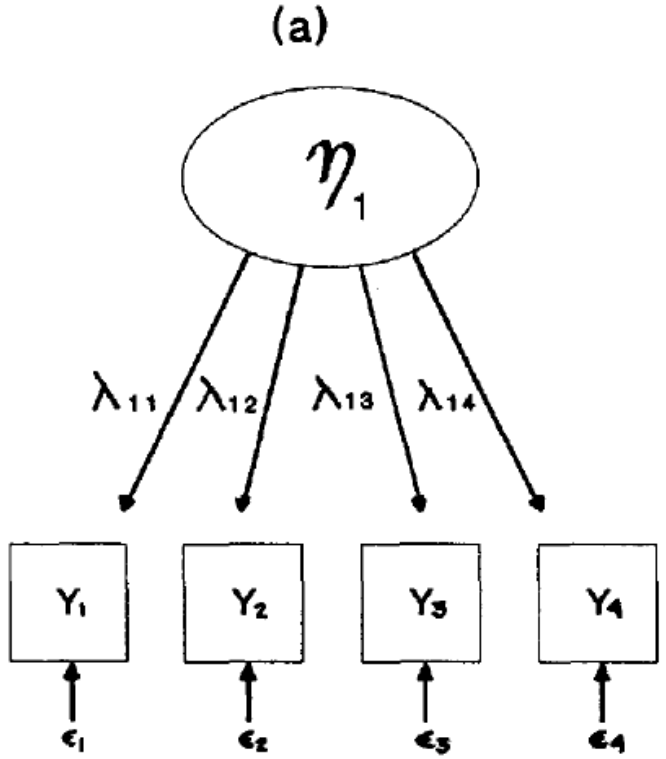


Conteúdos

- Análise fatorial exploratória
- http://personality-project.org/r/psych/



- Análise fatorial confirmatória
- Análise de trajetórias e modelagem estrutural de equações
- http://lavaan.ugent.be/



Psychological Bulletin
1991, Vol. 110, No. 2, 305–314

Conventional Wisdom on Measurement:
A Structural Equation Perspective

Kenneth Bollen
Sociology Department
University of North Carolina at Chapel Hill

Richard Lennox
Institute for Research in Social Science
University of North Carolina at Chapel Hill

Análise fatorial confirmatória

Journal of Statistical Software

May 2012, Volume 48, Issue 6. *http://www.jstatsoft.org/*

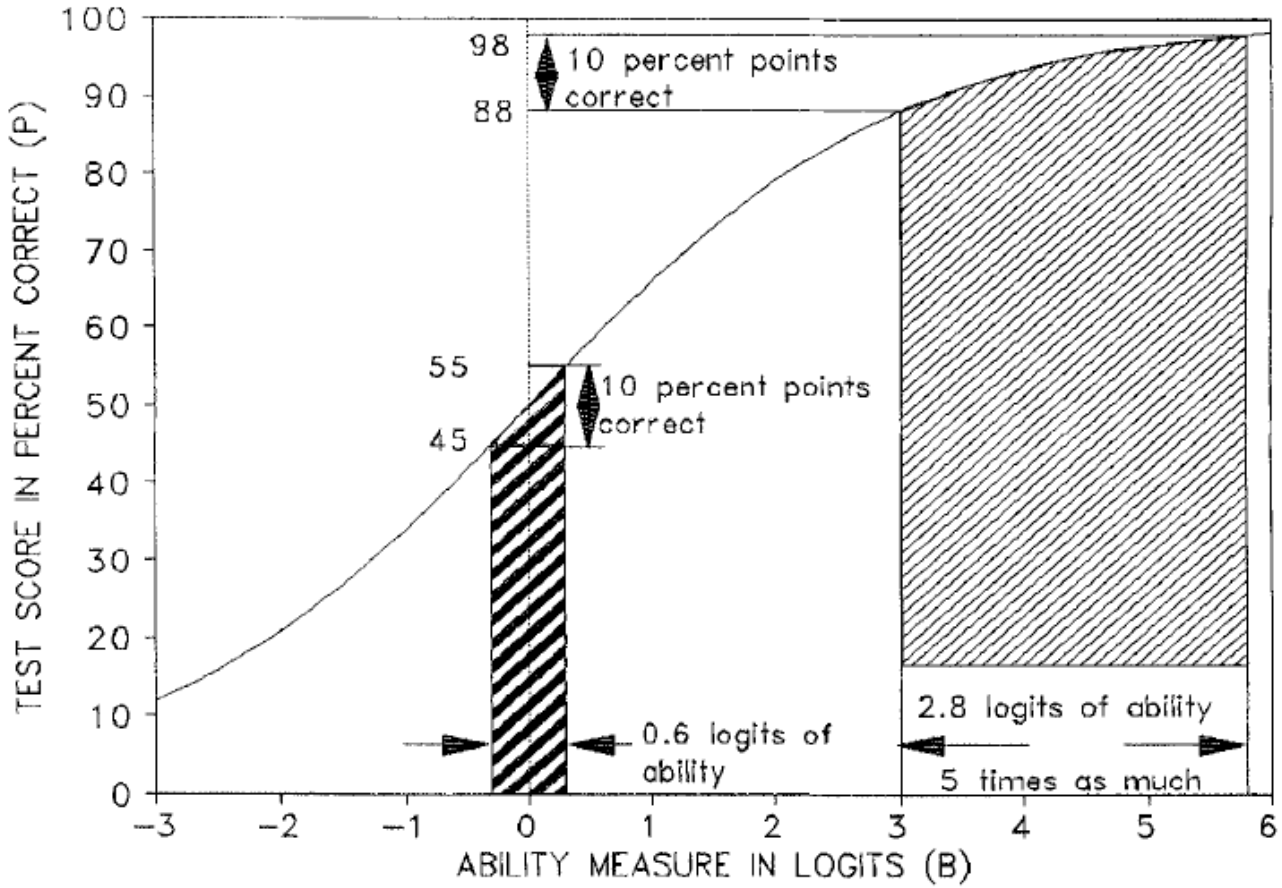
mirt: A Multidimensional Item Response Theory Package for the R Environment

R. Philip Chalmers
York University

- Teoria de Resposta ao Item
- https://www.rdocumentation.org/packages/mirt/versions/1.30
- https://www.jstatsoft.org/article/view/v048i06

A History of Social Science Measurement

Benjamin D. Wright
University of Chicago, MESA Psychometric Laboratory



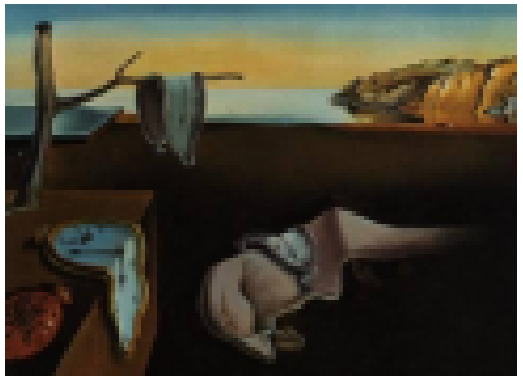
Análise de Rede

- Pacote qgraph

http://psychosystems.org/files/Literature/EpskampEtAl2012.pdf

- Capítulo de livro

- **MACHADO, W. L.**; VISSOCI, J. R. N. ; EPSKAMP, S. (2015). Análise de rede aplicada à psicometria e à avaliação psicológica. In C. S. Hutz; D. R. Bandeira & C. M. Trentini (Eds.). *Psicometria* (p. 125-146). Porto Alegre: Artmed.



Journal of Statistical Software

May 2012, Volume 48, Issue 4. http://www.jstatsoft.org/

qgraph: Network Visualizations of Relationships in Psychometric Data

Sacha Epskamp
University of Amsterdam

Angélique O. J. Cramer
University of Amsterdam

Lourens J. Waldorp
University of Amsterdam

(vii)
```
qgraph(E3,directed=FALSE)
qgraph(E3,layout="circle")
qgraph(E3)
qgraph(E3,directed=FALSE,layout="spring")
```

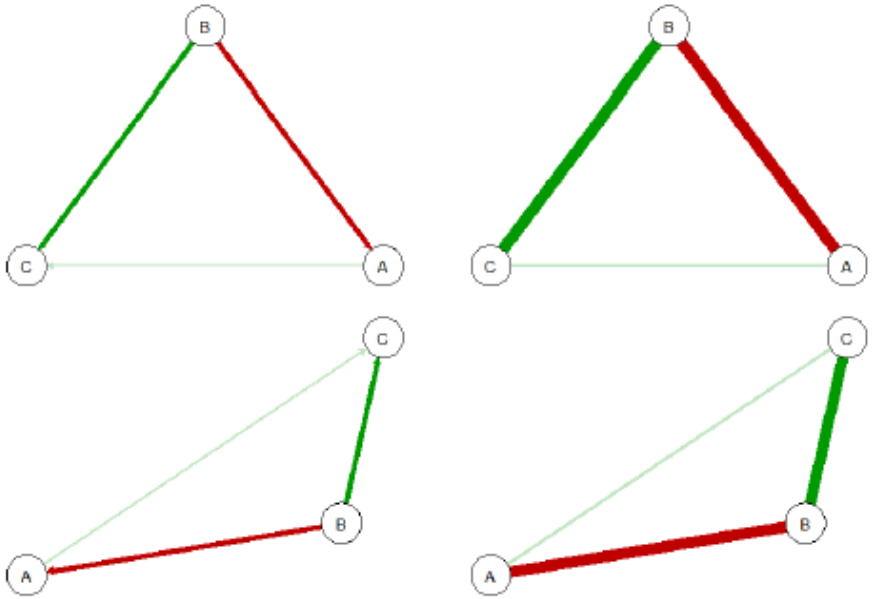


Figura 5. Redes ponderadas, direcionais e não-direcionais, sem (acima) e com (abaixo) o emprego do algoritmo de posicionamento.

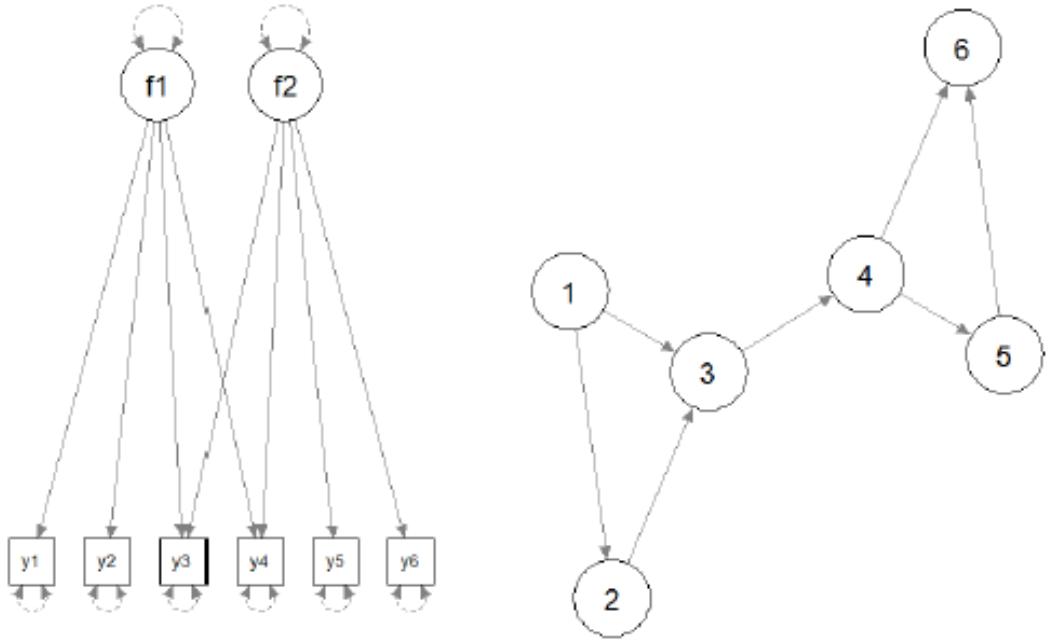


Figura 13. Modelo de traço latente (esquerda) e de rede (direita) da comorbidade.

- Sacha Epskamp
- Assistant Professor in Psychological Methods and Psychometrics at the University of Amsterdam

- João Vissoci
- Pesquisador na divisão de Emergency Medicine do departamento de Cirurgia, e na divisão Duke Global Neurosurgery and Neuroscience (DGNN) do departamento de Neurocirurgia, na Duke University

Análise de rede aplicada à psicometria e a avaliação psicológica

Wagner de Lara Machado

Pontifícia Universidade Católica de Campinas

João Ricardo Nickenig Vissoci

Faculdade Ingá e Duke University

Sacha Epskamp

Universiteit van Amsterdam

Psico-USF, Bragança Paulista, v. 20, n. 2, p. 259-274, mai./ago. 2015

Positive Mental Health Scale: Validation of the *Mental Health Continuum – Short Form*

Wagner de Lara Machado – Pontifícia Universidade Católica de Campinas, Campinas, São Paulo, Brasil
Denise Ruschel Bandeira – Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

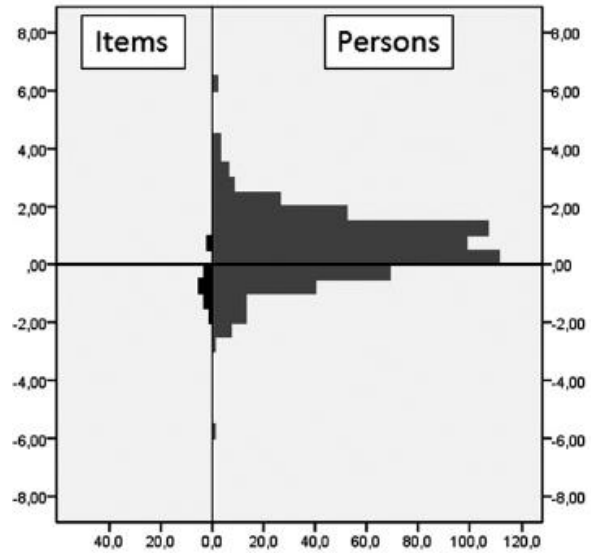


Figure 1. Map of items and persons. The vertical axis indicates the scale in *logits*. The distribution was obtained by fixing the contrary measure so the mean would be equal to zero.

Table 2
Item Factor Loadings and Reliability Measures of the MHC-SF in the Bifactor Model

Item (summarized content)	Factor loading			
	General factor	EWB	SWB	PWB
1 – Happy	.70	.65		
2 – Interested	.78	.32		
3 – Satisfied	.78	.36		
4 – Contribute to society	.69		.07	
5 – Belong to community	.67		.23	
6 – Society is becoming a better place	.62		.63	
7 – People are good	.62		.41	
8 – Way society works makes sense	.57		.56	
9 – Likes own personality	.79			.21
10 – Manages responsibility well	.70			.19
11 – Relationships with others	.72			.24
12 – Grow and become a better person	.62			.39
13 – Confident to express own ideas	.68			.52
14 – Life has direction or meaning	.83			.12
Average extracted variance	.49	.22	.19	.10
Composite reliability	.93	.43	.47	.34

Note. EWB = emotional well-being, SWB = social well-being, PWB = psychological well-being.

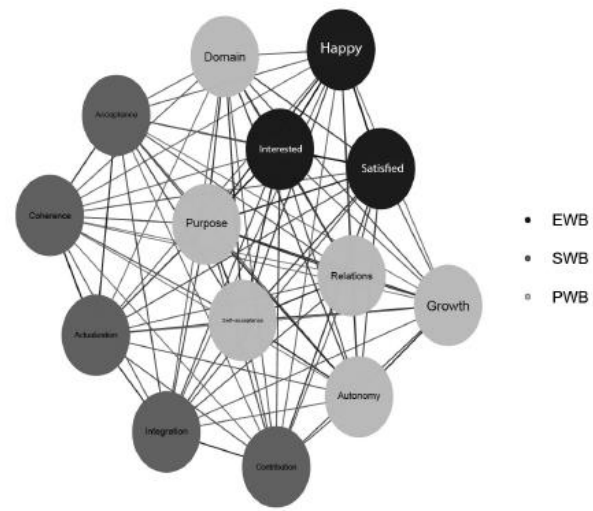


Figure 2. Network of positive mental health indicators. The covariance structure indicates that the MHC-SF items are strongly associated with the others, resulting in a dense component. The items in the emotional (EWB), social (SWB) and psychological (PWB) well-being subscales exhibit moderate to strong crossed associations. Purpose in life and self-acceptance are the central nodes of the system, meaning that they are more strongly associated with the remainder of the items. The stronger line represents the correlation between "happy" and "satisfied" ($r_{1\text{-}3}$ = .78); the fainter line represents the correlation between "relations" and "coherence" ($r_{8\text{-}11}$ = .35).

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 February 2017
doi: 10.3389/fpsyg.2017.00192

The Experience of Sexual Stigma and the Increased Risk of Attempted Suicide in Young Brazilian People from Low Socioeconomic Group

Angelo Brandelli Costa[1]*, *Andrew Pasley*[2], *Wagner de Lara Machado*[3], *Ernesto Alvarado*[4], *Luciana Dutra-Thomé*[4] *and Silvia Helena Koller*[4]

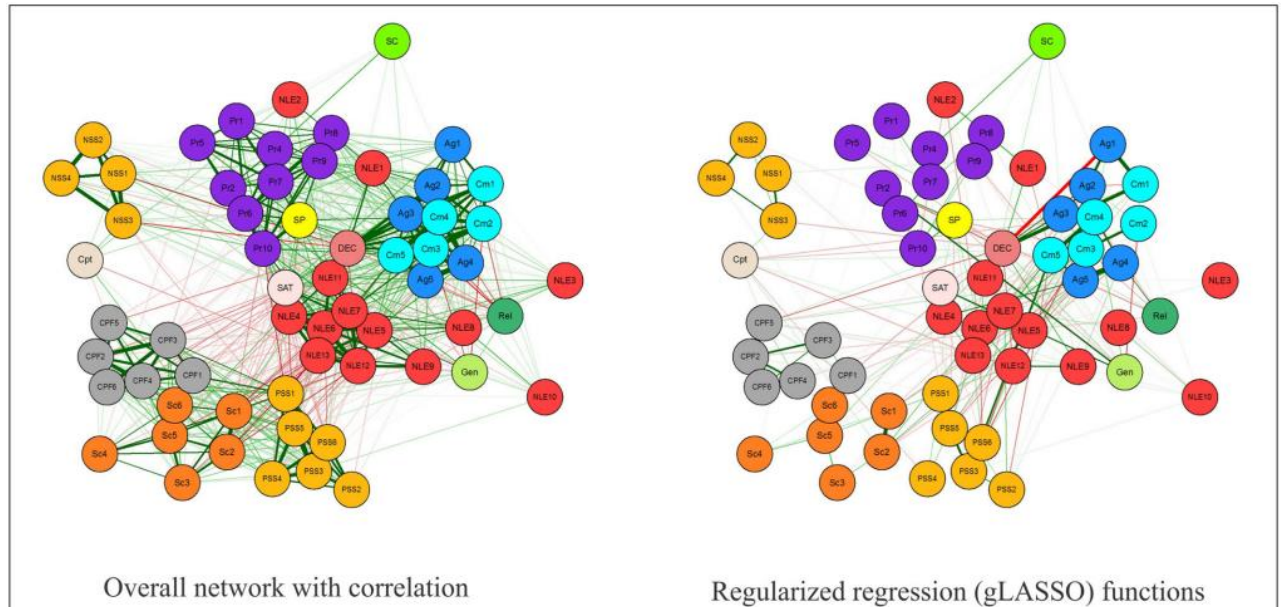


FIGURE 1 | Overall network with correlation (left) and regularized regression (gLASSO) functions **(right)**. Variables names match those in **Table 1**.

This study was intended to analyze the intersection of experience of sexual stigma low-socioeconomic status, and suicide attempt amongst young Brazilians (11–24 years old). In each of the data collection periods (2004–2006: $n = 7185$; 2010–2012: $n = 2734$), participants completed a questionnaire-based instrument. Network analysis provided support for a Minority Stress Model, oriented around whether participants had experienced sexual stigma. Although suicide attempts decreased by 20% for participants

Psychometric Perspectives on Diagnostic Systems

▼

Denny Borsboom (2008)

University of Amsterdam

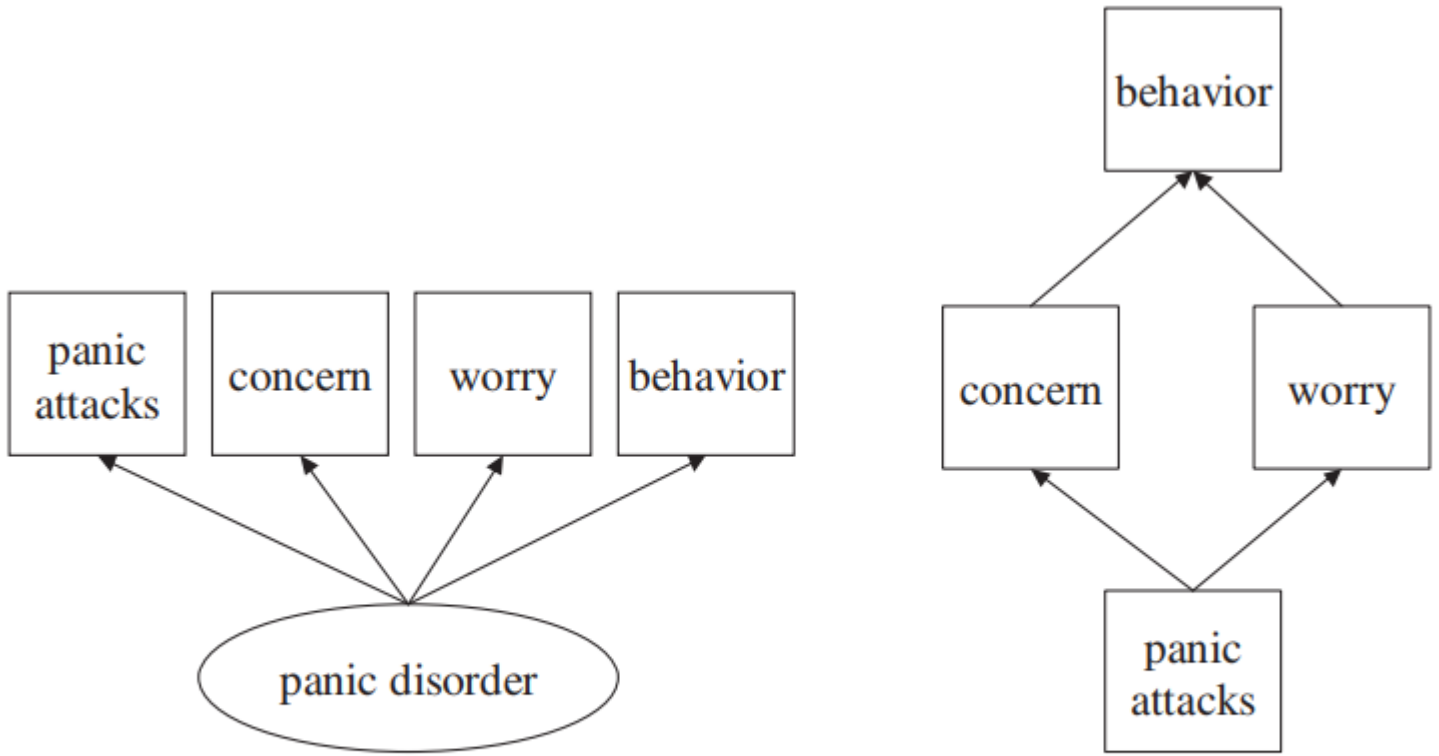


Figure 1. The left panel shows the relation between panic disorder and its symptoms from a latent variable modeling point of view. The right panel shows a representation of these symptoms as a causal system.

BEHAVIORAL AND BRAIN SCIENCES (2010) **33**, 137–193
doi:10.1017/S0140525X09991567

Comorbidity: A network perspective

Angélique O. J. Cramer
Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands
A.O.J.Cramer@uva.nl
www.aojcramer.com

Lourens J. Waldorp
Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands
L.J.Waldorp@uva.nl
http://users.fmg.uva.nl/lwaldorp

Han L. J. van der Maas
Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands
H.L.J.vanderMaas@uva.nl
http://users.fmg.uva.nl/hvandermaas/

Denny Borsboom
Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands
D.Borsboom@uva.nl
http://sites.google.com/site/borsboomdenny/dennyborsboom

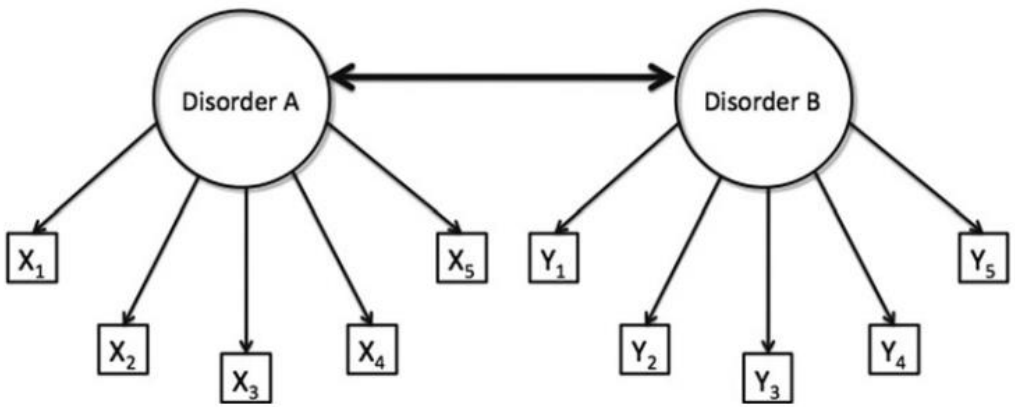


Figure 1. A model of comorbidity between disorders A and B, under the standard assumptions of latent variable modeling. The *circles* represent the disorders (i.e., latent variables) and the *rectangles* represent the observable core symptoms of those disorders (i.e., $X_1 - X_5$ for disorder A, and $Y_1 - Y_5$ for disorder B). In this model, comorbidity is viewed as a correlation between the latent variables, visualized by the *thick bidirectional edge* between disorders A and B.

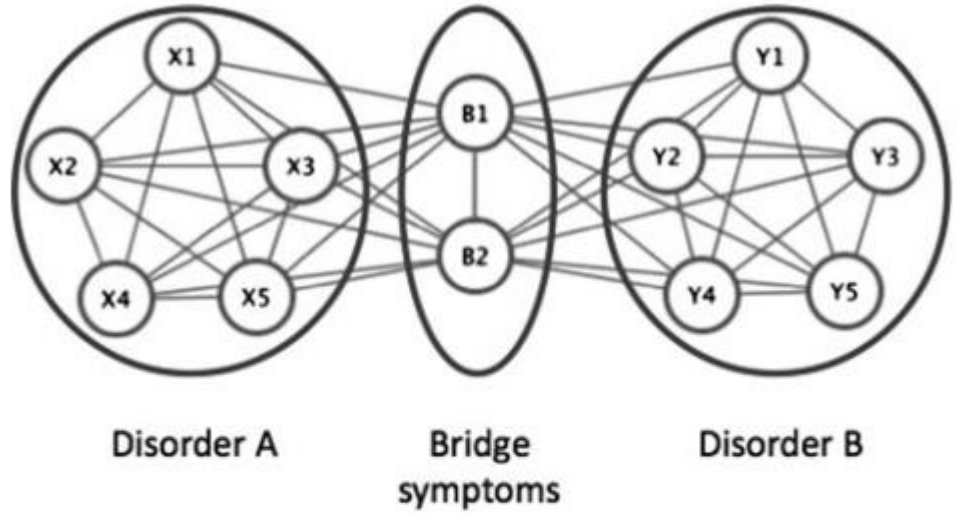
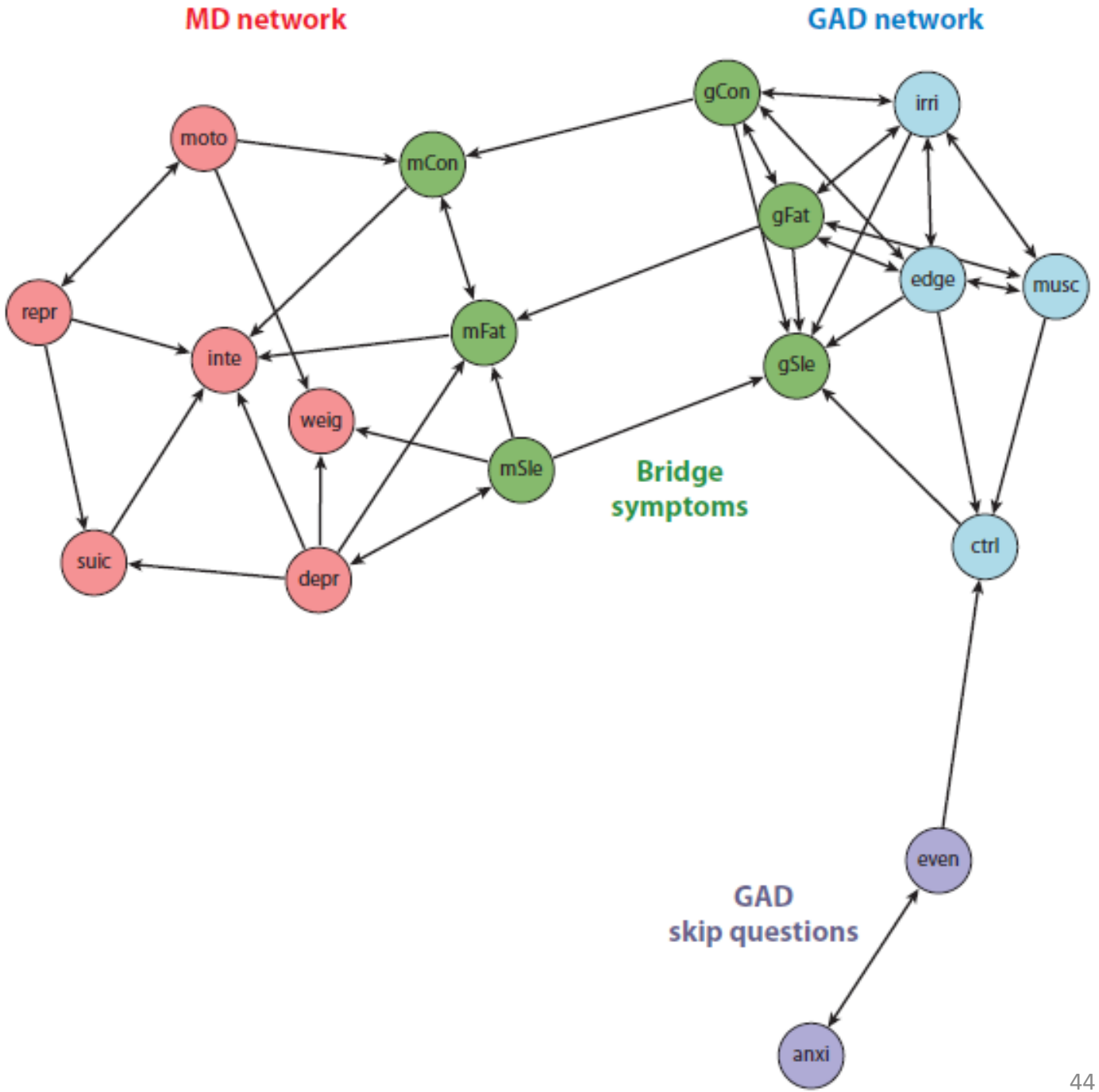


Figure 2. Comorbidity under a network approach. Disorder A consists of bidirectionally related symptoms $X_1 - X_5$, and disorder B consists of symptoms $Y_1 - Y_5$. Symptoms B_1 and B_2 are *bridge symptoms* that overlap between disorders A and B. In this model, comorbidity arises as a result of direct relations between the bridge symptoms of two disorders.

Network Analysis: An Integrative Approach to the Structure of Psychopathology

Denny Borsboom and Angélique O.J. Cramer

Department of Psychology, University of Amsterdam, Amsterdam 1018 XA, The Netherlands; email: D.Borsboom@uva.nl



- Redes Bayesianas
- http://www.bnlearn.com/