

Rense Nieuwenhuis

Accessing World Fertility Survey-data using Read.ISI

Radboud University Nijmegen





Introduction

- Read.ISI: R-Project package for accessing old survey data
 - Technological change
- Fertility Project
- World Fertility Surveys
 - Problem: the ISI codebook
- Rationale behind Read.ISI
 - Options and usage of the package
- Future Development



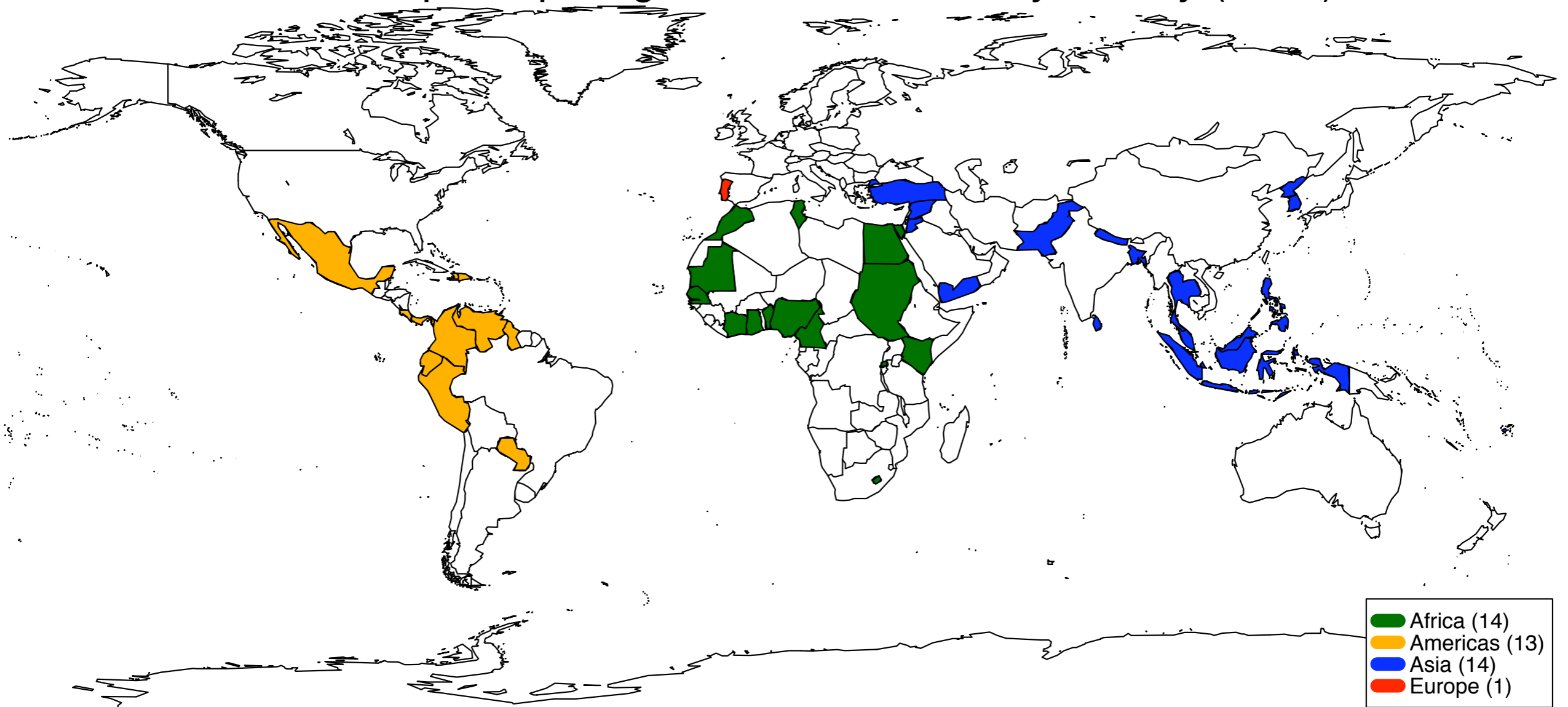
Fertility Project

- Abortion, contraception and assisted reproduction: technological innovations and the role of religion and education
 - Dr. Ariana Need
 - 2 PhD students
 - 2 Research Assistants
- PhD Project of Mark Levels, MSc.
 - *Explaining Abortion - The Rationality of Ethical Choices*
 - Internationally comparative, longitudinal perspective
- World Fertility Surveys



Fertility Project: World Fertility Survey

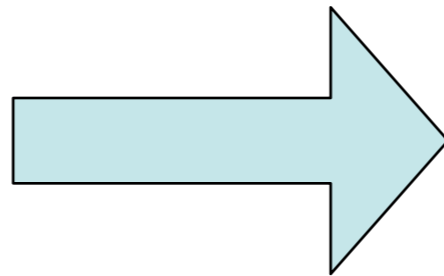
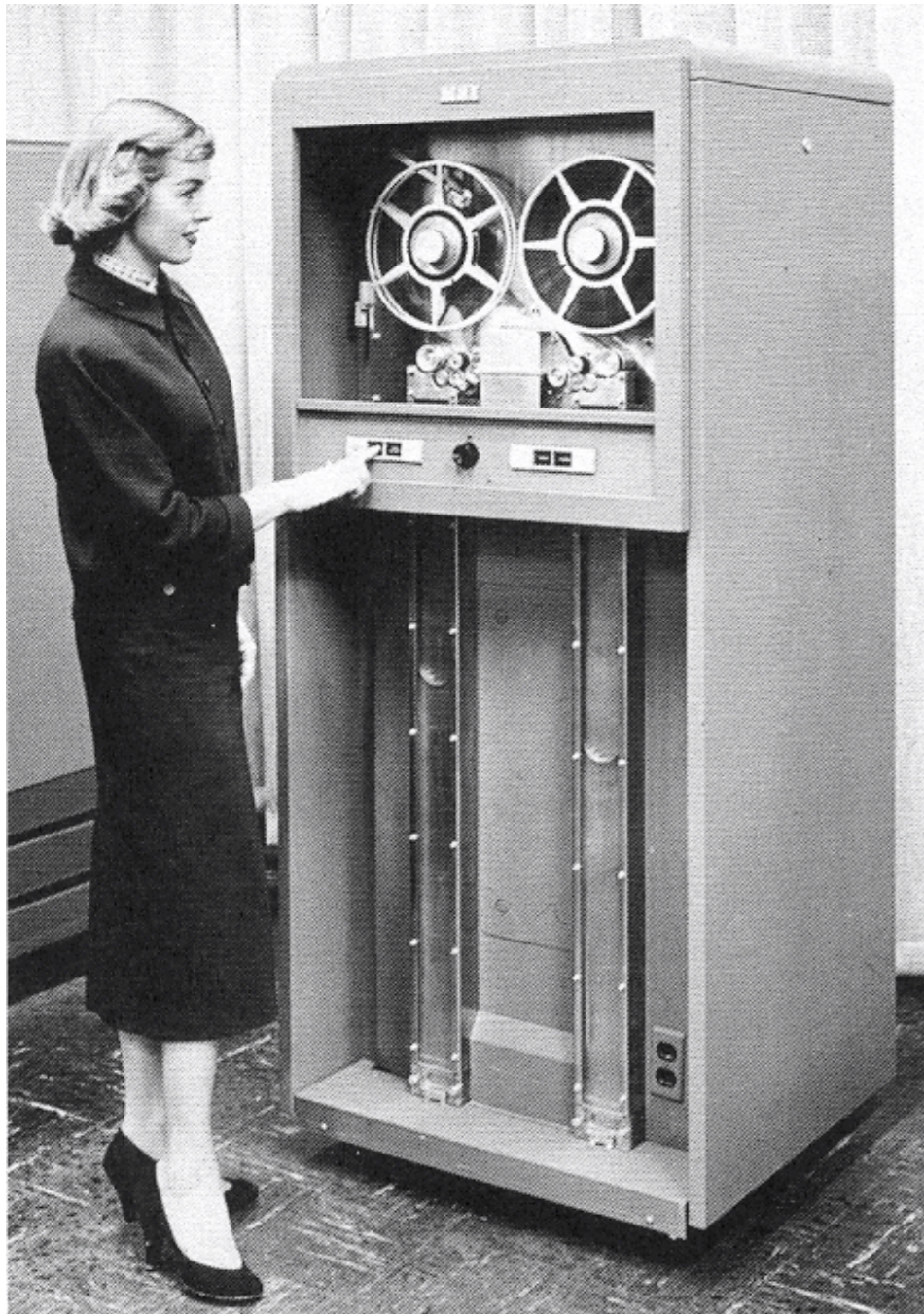
Countries participating in the World Fertility Survey (WFS)



- Number of countries: 41
- Late 70's, early 80's
- Fertility Calendars
- Data as provided
 - Fixed width data files
 - ISI formatted code-books



On Technological Change





Approach

6	4	2	1	6	4	30	6
V104	137	2	1	6	88	Status of first relationship 1 Married 2 Common law 3 Visiting 4 Was married 5 Was common law 6 Was visiting 88 Never in rel.	
V105	139	2	0	1	88	First relationship dissolved 0 No 1 Yes 88 Never in rel.	
V106	141	2	0	1	88	Has had 2nd or later rel. 0 No 1 Yes 88 Not dissolved	
V107	143	2	1	6	88	Current marital status	V104

Variable

Start & number
positions

Missings

Variable labels

Value labels

Label Reference



Approach: Converting the Codebook

- Read the code-book:
 - `read.fwf(input.file, c(6,3,4,1,2,9,4,1,4,1,30,1,6))`
- Two Matrices:
 - converted.codebook
 - variable name, variable label
 - start position, number of positions
 - missings
 - label reference
 - converted.labels
 - variable name
 - value
 - label
- Returned as list:
 - `converted.result <- list(converted.codebook, converted.labels)`



Approach: Reading the Data to R-Project

- Reading the data
 - `isi.data <- read.fwf(dat.file, width = isi.widths, col.names = isi.names, ...)`
- Missing Values
 - Selecting values matching with indicated missing labels
- Value labels
 - Converting to `factor()`
 - Factor `levels()` based on iteratively matching variable name and value labels



Approach: Creating SPSS-syntax

- Using R matrices to store SPSS syntax
- Calling the `get data` function in SPSS
 - `file.header[1] <- "GET DATA /TYPE = TXT"`
- `data.positions`: matrix with on each row:
 - variable name
 - start & end positions
 - type of variable (F)
- Further sections:
 - Variable labels
 - Missing values
 - Value labels
- Matrices are written to text-file
 - `write.table(file.sps, append=TRUE)`



Package Read.ISI

- **read.codebook.isi**
 - workhorse function
- **read.isi**
 - Read data into R-Project, based on ISI code-book
- **convert.isi**
 - Convert ISI code-book into SPSS executable syntax
- **clean**
 - Helper function



Available Options

- **input.file**
 - Location of the ISI-formatted .dct code-book file.
- **dat.file**
 - Location of the fixed-width data-file to load.
- **add.missings**
 - Should value labels indicating missing values be transformed to NA?
Defaults to `TRUE`.
- **add.value.labels**
 - Convert variables with value labels to factors
- Further arguments passed on to `read.fwf ()`
 - **N**
 - **skip**



Future Development of read.ISI

- Speed
- Efficient reading of large files
- Read selections of variables



Questions?

- For more information:
 - Download read.ISI
 - Available from a CRAN server near you
 - Conference paper
 - www.rensenieuwenhuis.nl/r-project/read.isi/
 - contact@rensenieuwenhuis.nl