

Advances in the Dataset R package

Emmanuel Rousseaux

Institute for Demographic and Life Course Studies

University of Geneva

1211 Geneva 4, Switzerland

`emmanuel.rousseau@unige.ch`

Outline

Introduction

The Dataset project

A short demonstration

Conclusion

Future work

Outline

Introduction

The Dataset project

A short demonstration

Conclusion

Future work

Teaching and Research Assistant at the Departement of Economics, SES, Unige

PhD directed by

- ▶ Gilbert Ritschard

Institute for Demographic and Life Course Studies, University of Geneva

- ▶ Michel Léonard

Institute of Services Science, University of Geneva

Teaching and Research Assistant at the Department of Economics, SES, Unige

PhD directed by

▶ Gilbert Ritschard

Institute for Demographic and Life Course Studies, University of Geneva

▶ Michel Léonard

Institute of Services Science, University of Geneva

Teaching and Research Assistant at the Department of Economics, SES, Unige

PhD directed by

- ▶ Gilbert Ritschard

Institute for Demographic and Life Course Studies, University of Geneva

- ▶ Michel Léonard

Institute of Services Science, University of Geneva

Teaching and Research Assistant at the Departement of Economics, SES, Unige

PhD directed by

- ▶ Gilbert Ritschard

Institute for Demographic and Life Course Studies, University of Geneva

- ▶ Michel Léonard

Institute of Services Science, University of Geneva

(some) Research interests

- ▶ Decision Trees
- ▶ Association rules mining
- ▶ Nature-based optimization algorithms
- ▶ Health sociology
- ▶ Cognitive psychology

(some) Research interests

- ▶ Decision Trees
- ▶ Association rules mining
- ▶ Nature-based optimization algorithms
- ▶ Health sociology
- ▶ Cognitive psychology

(some) Research interests

- ▶ Decision Trees
- ▶ Association rules mining
- ▶ Nature-based optimization algorithms
- ▶ Health sociology
- ▶ Cognitive psychology

(some) Research interests

- ▶ Decision Trees
- ▶ Association rules mining
- ▶ **Nature-based optimization algorithms**
- ▶ Health sociology
- ▶ Cognitive psychology

(some) Research interests

- ▶ Decision Trees
- ▶ Association rules mining
- ▶ Nature-based optimization algorithms
- ▶ Health sociology
- ▶ Cognitive psychology

(some) Research interests

- ▶ Decision Trees
- ▶ Association rules mining
- ▶ Nature-based optimization algorithms
- ▶ Health sociology
- ▶ Cognitive psychology

NCCR LIVES "Overcoming vulnerability: life course perspectives"

I work within the IP14 "Measuring life sequences and the disorder of lives" led by Gilbert Ritschard

Aims at providing ad hoc methods for life course analysis in order to have more insight about dynamics of vulnerability

NCCR LIVES "Overcoming vulnerability: life course perspectives"

I work within the IP14 "Measuring life sequences and the disorder of lives" led by Gilbert Ritschard

Aims at providing ad hoc methods for life course analysis in order to have more insight about dynamics of vulnerability

NCCR LIVES "Overcoming vulnerability: life course perspectives"

I work within the IP14 "Measuring life sequences and the disorder of lives" led by Gilbert Ritschard

Aims at providing ad hoc methods for life course analysis in order to have more insight about dynamics of vulnerability

Goals pursued in my thesis project

First part: Increase a scientist's time available to concentrate on her research question

- ▶ Reduce the time needed to prepare data for analysis
- ▶ Automatically verify the quality of models computed
- ▶ Format analysis outputs to allow immediate interpretation of results

Goals pursued in my thesis project

First part: Increase a scientist's time available to concentrate on her research question

- ▶ Reduce the time needed to prepare data for analysis
- ▶ Automatically verify the quality of models computed
- ▶ Format analysis outputs to allow immediate interpretation of results

Goals pursued in my thesis project

First part: Increase a scientist's time available to concentrate on her research question

- ▶ Reduce the time needed to prepare data for analysis
- ▶ Automatically verify the quality of models computed
- ▶ Format analysis outputs to allow immediate interpretation of results

Goals pursued in my thesis project

First part: Increase a scientist's time available to concentrate on her research question

- ▶ Reduce the time needed to prepare data for analysis
- ▶ **Automatically verify the quality of models computed**
- ▶ Format analysis outputs to allow immediate interpretation of results

Goals pursued in my thesis project

First part: Increase a scientist's time available to concentrate on her research question

- ▶ Reduce the time needed to prepare data for analysis
- ▶ Automatically verify the quality of models computed
- ▶ Format analysis outputs to allow immediate interpretation of results

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided

- ▶ Decision trees for the discovery of vulnerable profiles
- ▶ Multi-channel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided

- ▶ *Decision trees for the discovery of vulnerable profiles*
- ▶ *Multichannel association rules mining for life courses*

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided

- ▶ *Dataset* tools for the discovery of vulnerable subgroups
- ▶ Multichannel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided
 - ▶ Decision trees for the discovering of vulnerable profiles
 - ▶ Multi-channel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ **Two new methods will be provided**
 - ▶ Decision trees for the discovering of vulnerable profiles
 - ▶ Multi-channel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided
 - ▶ Decision trees for the discovering of vulnerable profiles
 - ▶ Multi-channel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided
 - ▶ Decision trees for the discovering of vulnerable profiles
 - ▶ **Multi-channel association rules mining for life courses**

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided
 - ▶ Decision trees for the discovering of vulnerable profiles
 - ▶ Multi-channel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Goals pursued in my thesis project

Second part: Help the scientist discover social determinants of a specific social behavior

Key point: This job is harder when working on vulnerable populations (rare events)

- ▶ Mining rare but very correlated life events/subsequences
- ▶ Two new methods will be provided
 - ▶ Decision trees for the discovering of vulnerable profiles
 - ▶ Multi-channel association rules mining for life courses

More information here: <http://www.lives-nccr.ch/sites/default/files/pdf/page/poster-rousseau.pdf>

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ Currently R does not offer a robust framework to handle survey data
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ Currently R does not offer a robust framework to handle survey data
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ Currently R does not offer a robust framework to handle survey data
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
 - ▶ Much time is needed to prepare the data
 - ▶ A lot of state-of-the-art methods are provided on R only
 - ▶ Currently R does not offer a robust framework to handle survey data
 - ▶ Especially for panel survey data and network survey data
- ⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
 - ▶ Much time is needed to prepare the data
 - ▶ A lot of state-of-the-art methods are provided on R only
 - ▶ Currently R does not offer a robust framework to handle survey data
 - ▶ Especially for panel survey data and network survey data
- ⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ **Currently R does not offer a robust framework to handle survey data**
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ Currently R does not offer a robust framework to handle survey data
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ Currently R does not offer a robust framework to handle survey data
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Today's presentation: The Dataset project

Starting points

- ▶ Population studies strongly rely on survey data
- ▶ Much time is needed to prepare the data
- ▶ A lot of state-of-the-art methods are provided on R only
- ▶ Currently R does not offer a robust framework to handle survey data
- ▶ Especially for panel survey data and network survey data

⇒ Need for a specific software framework in R

Outline

Introduction

The Dataset project

A short demonstration

Conclusion

Future work

Overview

- ▶ Started about 1 year ago
- ▶ Aims to provide a framework for handling complex survey data
- ▶ Efficient
- ▶ Secure

Overview

- ▶ Started about 1 year ago
- ▶ Aims to provide a framework for handling complex survey data
- ▶ Efficient
- ▶ Secure

Overview

- ▶ Started about 1 year ago
- ▶ Aims to provide a framework for handling complex survey data
- ▶ Efficient
- ▶ Secure

Overview

- ▶ Started about 1 year ago
- ▶ Aims to provide a framework for handling complex survey data
- ▶ **Efficient**
- ▶ Secure

Overview

- ▶ Started about 1 year ago
- ▶ Aims to provide a framework for handling complex survey data
- ▶ Efficient
- ▶ **Secure**

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ Accepts user-defined missing values
- ▶ Natively accounts for weights
- ▶ Generates summary views directly in PDF format
- ▶ Automatic data consistency checks
- ▶ Automatic "loss of representativeness" checks

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ Accepts user-defined missing values
- ▶ Natively accounts for weights
- ▶ Generates summary views directly in PDF format
- ▶ Automatic data consistency checks
- ▶ Automatic "loss of representativeness" checks

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ **Accepts user-defined missing values**
- ▶ Natively accounts for weights
- ▶ Generates summary views directly in PDF format
- ▶ Automatic data consistency checks
- ▶ Automatic "loss of representativeness" checks

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ Accepts user-defined missing values
- ▶ **Natively accounts for weights**
- ▶ Generates summary views directly in PDF format
- ▶ Automatic data consistency checks
- ▶ Automatic "loss of representativeness" checks

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ Accepts user-defined missing values
- ▶ Natively accounts for weights
- ▶ **Generates summary views directly in PDF format**
- ▶ Automatic data consistency checks
- ▶ Automatic "loss of representativeness" checks

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ Accepts user-defined missing values
- ▶ Natively accounts for weights
- ▶ Generates summary views directly in PDF format
- ▶ **Automatic data consistency checks**
- ▶ Automatic "loss of representativeness" checks

Key functionalities

- ▶ Allows to store meta-information about the data
- ▶ Accepts user-defined missing values
- ▶ Natively accounts for weights
- ▶ Generates summary views directly in PDF format
- ▶ Automatic data consistency checks
- ▶ Automatic "loss of representativeness" checks

Tools for preparing data

- ▶ Search for specific variables across the whole database
- ▶ Specify the measure (scale, nominal, ordinal, ...)
- ▶ Turn a missing value to valid case and vice-versa
- ▶ Easy to use/remember recoding methods
- ▶ Detailed frequency tables

Tools for preparing data

- ▶ Search for specific variables across the whole database
- ▶ Specify the measure (scale, nominal, ordinal, ...)
- ▶ Turn a missing value to valid case and vice-versa
- ▶ Easy to use/remember recoding methods
- ▶ Detailed frequency tables

Tools for preparing data

- ▶ Search for specific variables across the whole database
- ▶ Specify the measure (scale, nominal, ordinal, ...)
- ▶ Turn a missing value to valid case and vice-versa
- ▶ Easy to use/remember recoding methods
- ▶ Detailed frequency tables

Tools for preparing data

- ▶ Search for specific variables across the whole database
- ▶ Specify the measure (scale, nominal, ordinal, ...)
- ▶ Turn a missing value to valid case and vice-versa
- ▶ Easy to use/remember recoding methods
- ▶ Detailed frequency tables

Tools for preparing data

- ▶ Search for specific variables across the whole database
- ▶ Specify the measure (scale, nominal, ordinal, ...)
- ▶ Turn a missing value to valid case and vice-versa
- ▶ **Easy to use/remember recoding methods**
- ▶ Detailed frequency tables

Tools for preparing data

- ▶ Search for specific variables across the whole database
- ▶ Specify the measure (scale, nominal, ordinal, ...)
- ▶ Turn a missing value to valid case and vice-versa
- ▶ Easy to use/remember recoding methods
- ▶ Detailed frequency tables

Tools for panel data

- ▶ Automatically check for missings values/valids cases across years
- ▶ Extract a whole trajectory in one step
- ▶ Switch missing/valid values across years in one step
- ▶ Perform recoding operation across years in one step
- ▶ Export to sequence objects ready to be analysed with the TraMineR toolbox (Gabadinho et al., 2011)

Tools for panel data

- ▶ Automatically check for missings values/valids cases across years
- ▶ Extract a whole trajectory in one step
- ▶ Switch missing/valid values across years in one step
- ▶ Perform recoding operation across years in one step
- ▶ Export to sequence objects ready to be analysed with the TraMineR toolbox (Gabadinho et al., 2011)

Tools for panel data

- ▶ Automatically check for missings values/valids cases across years
- ▶ **Extract a whole trajectory in one step**
- ▶ Switch missing/valid values across years in one step
- ▶ Perform recoding operation across years in one step
- ▶ Export to sequence objects ready to be analysed with the TraMineR toolbox (Gabadinho et al., 2011)

Tools for panel data

- ▶ Automatically check for missings values/valids cases across years
- ▶ Extract a whole trajectory in one step
- ▶ **Switch missing/valid values across years in one step**
- ▶ Perform recoding operation across years in one step
- ▶ Export to sequence objects ready to be analysed with the TraMineR toolbox (Gabadinho et al., 2011)

Tools for panel data

- ▶ Automatically check for missings values/valids cases across years
- ▶ Extract a whole trajectory in one step
- ▶ Switch missing/valid values across years in one step
- ▶ **Perform recoding operation across years in one step**
- ▶ Export to sequence objects ready to be analysed with the TraMineR toolbox (Gabadinho et al., 2011)

Tools for panel data

- ▶ Automatically check for missings values/valids cases across years
- ▶ Extract a whole trajectory in one step
- ▶ Switch missing/valid values across years in one step
- ▶ Perform recoding operation across years in one step
- ▶ Export to sequence objects ready to be analysed with the TraMineR toolbox (Gabadinho et al., 2011)

Development and availability

The toolbox comes as a series of R packages, freely available

- ▶ Cross-sectional: first stable version available on R-Forge
- ▶ Longitudinal: beta version will be released in two weeks
- ▶ Network: development started

Website: <https://r-forge.r-project.org/projects/dataset/>

Development and availability

The toolbox comes as a series of R packages, freely available

- ▶ **Cross-sectional: first stable version available on R-Forge**
- ▶ Longitudinal: beta version will be released in two weeks
- ▶ Network: development started

Website: <https://r-forge.r-project.org/projects/dataset/>

Development and availability

The toolbox comes as a series of R packages, freely available

- ▶ Cross-sectional: first stable version available on R-Forge
- ▶ Longitudinal: beta version will be released in two weeks
- ▶ Network: development started

Website: <https://r-forge.r-project.org/projects/dataset/>

Development and availability

The toolbox comes as a series of R packages, freely available

- ▶ Cross-sectional: first stable version available on R-Forge
- ▶ Longitudinal: beta version will be released in two weeks
- ▶ **Network: development started**

Website: <https://r-forge.r-project.org/projects/dataset/>

Development and availability

The toolbox comes as a series of R packages, freely available

- ▶ Cross-sectional: first stable version available on R-Forge
- ▶ Longitudinal: beta version will be released in two weeks
- ▶ Network: development started

Website: <https://r-forge.r-project.org/projects/dataset/>

Development and availability

The toolbox comes as a series of R packages, freely available

- ▶ Cross-sectional: first stable version available on R-Forge
- ▶ Longitudinal: beta version will be released in two weeks
- ▶ Network: development started

Website: <https://r-forge.r-project.org/projects/dataset/>

Dataset's cross-sectional package usage

- ▶ About 40 users
- ▶ Used in two Masters courses at the University of Geneva
- ▶ Two databases collected within the NCCR LIVES will be released in the Dataset format

Positive feedback from users

Dataset's cross-sectional package usage

- ▶ About 40 users
- ▶ Used in two Masters courses at the University of Geneva
- ▶ Two databases collected within the NCCR LIVES will be released in the Dataset format

Positive feedback from users

Dataset's cross-sectional package usage

- ▶ About 40 users
- ▶ Used in two Masters courses at the University of Geneva
- ▶ Two databases collected within the NCCR LIVES will be released in the Dataset format

Positive feedback from users

Dataset's cross-sectional package usage

- ▶ About 40 users
- ▶ Used in two Masters courses at the University of Geneva
- ▶ Two databases collected within the NCCR LIVES will be released in the Dataset format

Positive feedback from users

Dataset's cross-sectional package usage

- ▶ About 40 users
- ▶ Used in two Masters courses at the University of Geneva
- ▶ Two databases collected within the NCCR LIVES will be released in the Dataset format

Positive feedback from users

Outline

Introduction

The Dataset project

A short demonstration

Conclusion

Future work

Installing the package

```
install.packages('Dataset', repos='http://r-forge.r-project.org')
```

Loading the package

```
library(Dataset)
```

Importing an SPSS file

Here we use data from the Swiss Household Panel (Voorpostel et al., 2011)

```
shp.w2006 <- get.spss.file(  
  file = 'SHP06_P_USER.sav',  
  datadir = '/SHP-oct2011/SHP-Data-W1-W12-SPSS/W8_2006',  
  name = 'SHP wave 2006'  
)
```

Getting a codebook of the database

```
summaryToPDF(shp.w2006)
```

Preparing data for an analysis

First we set the weights

```
shp.w2006$wp06t1s <- wvar(shp.w2006$wp06t1s)  
weighting(shp.w2006) <- 'wp06t1s'
```

Retrieving variables of interest

```
health.var <- contains("health", shp.w2006)
```

```
##                                     Description
## p06c01                             Health status
## p06c02                             Satisfaction with health status
## p06c03                             Improvement in health: Last 12 months
## p06c04a                             Health problems: Back problems: Last 4 weeks
## p06c05a                             Health problems: Weakness, weariness: Last 4 weeks
## p06c06a                             Health problems: Sleeping problems: Last 4 weeks
## p06c07a                             Health problems: Headaches: Last 4 weeks
## p06c08                             Health impediment in everyday activities: Extension
## p06c19a                             Chronic illness or long-term health problem
## p06c11 Number of days affected by health problems: Last 12 months
## x06c05                             Assessment of health status
## x06c06                             Suffering from health problems
## x06c07                             Cause of health problems
## x06c09                             Days of suffering from health problems: Days
```


Live demonstration

Outline

Introduction

The Dataset project

A short demonstration

Conclusion

Future work

Conclusion

- ▶ The toolbox provides an efficient and secure framework for handling complex survey data
- ▶ Encouraging feedback from users
- ▶ Longitudinal and network versions forthcoming

Conclusion

- ▶ The toolbox provides an efficient and secure framework for handling complex survey data
- ▶ Encouraging feedback from users
- ▶ Longitudinal and network versions forthcoming

Conclusion

- ▶ The toolbox provides an efficient and secure framework for handling complex survey data
- ▶ Encouraging feedback from users
- ▶ Longitudinal and network versions forthcoming

Outline

Introduction

The Dataset project

A short demonstration

Conclusion

Future work

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ **Writing a full starting guide**
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ **Survival analysis**
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ **Structural equation modeling**

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Future work

- ▶ Reach out to survey data management experts for feedback
- ▶ Be compatible with DDI3 specification
 - importing a DDI3 file, and generating the DDI3 file for a specific database
- ▶ Facilitate export of data and analysis outputs in csv/tsv
- ▶ Writing a full starting guide
- ▶ Add front-ends for other popular methods, especially:
 - ▶ Survival analysis
 - ▶ Structural equation modeling

To request features: dataset-requests@lists.r-forge.r-project.org

Bibliography

Bibliography I

- [Gabadinho et al.] Gabadinho, A., Ritschard, G., Müller, N.S. & Studer, M. Analyzing and visualizing state sequences in R with TraMineR *Journal of Statistical Software*. Vol. 40(4), pp. 1-37.
- [Voorpostel et al.] Voorpostel, M., Tillmann, R, Lebert, F., Weaver, B., Kuhn, U., Lipps, O., Ryser, V.-A., Schmid, F., Rothenbühler, M., and Wernli, B. *Swiss Household Panel Userguide (1999-2010), Wave 12*. Lausanne: FORS.(October 2011).

I would be happy to provide a more specific bibliography to anyone interested

Thank you for your attention

Any question?