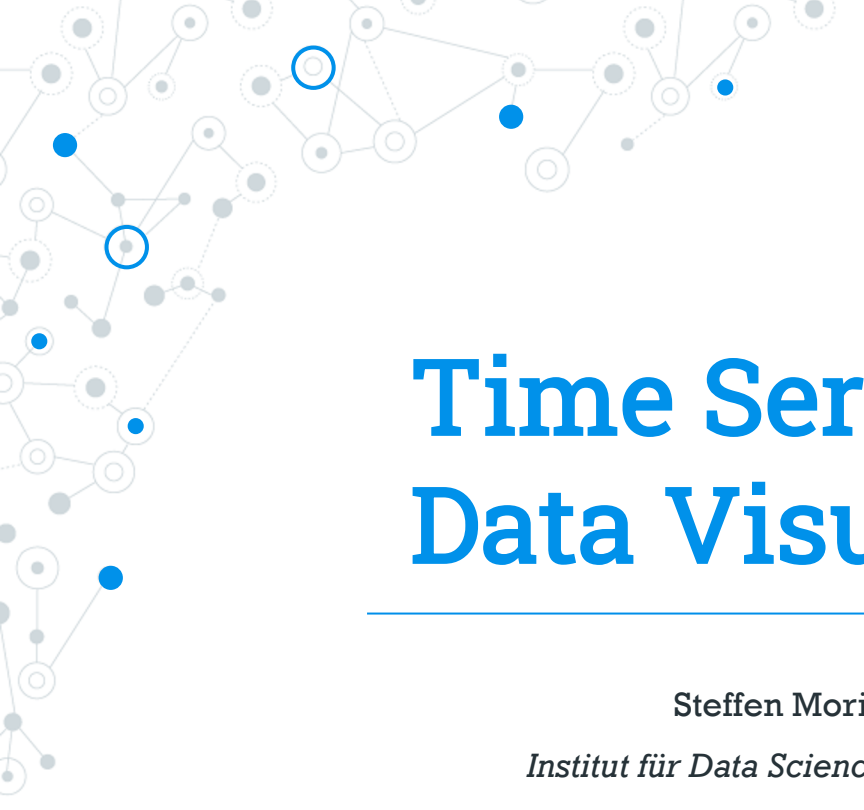




Time Series Missing Data Visualizations

Steffen Moritz, Thomas Bartz-Beielstein

Institut für Data Science, Engineering, and Analytics, TH Köln

steffen.moritz10@gmail.com

Hello!

I am Steffen Moritz

I am at eRum 2020,
because I love R :)

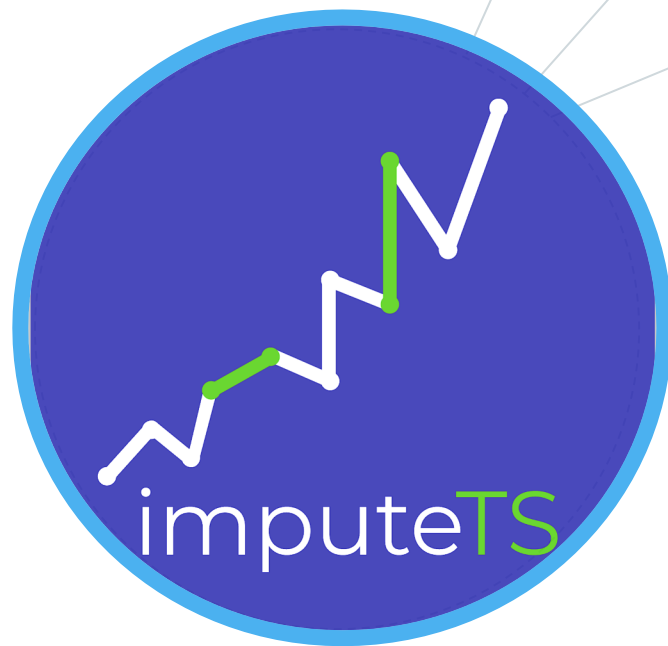


Time Series Missing Data

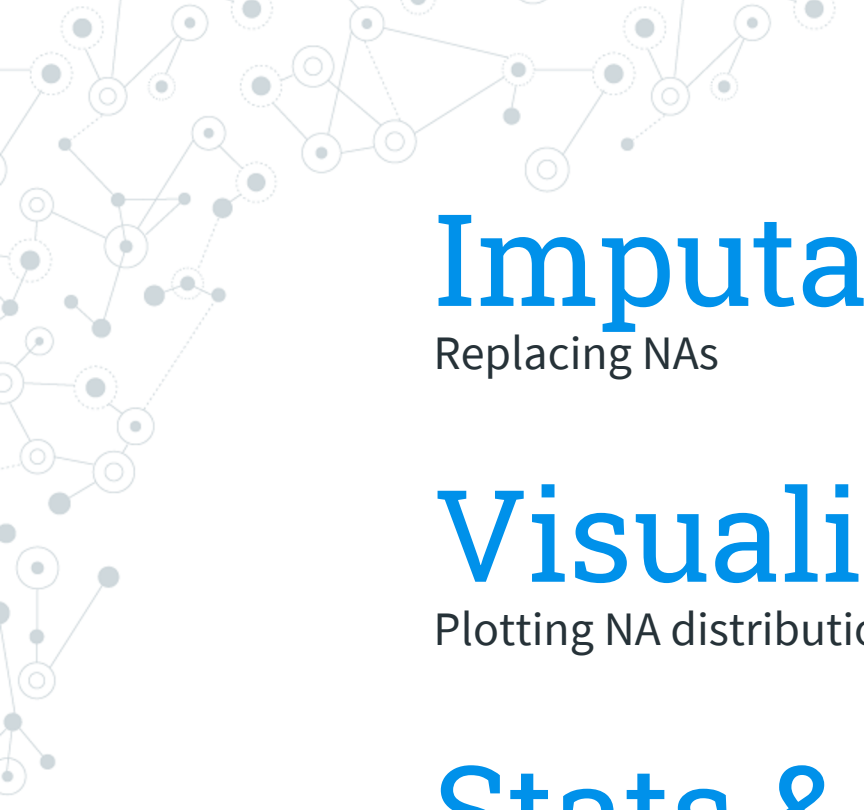
The **imputeTS** package helps with missing data in time series.

New plots with 3.1 Version!

(new version currently on Github soon on CRAN)



<https://github.com/SteffenMoritz/imputeTS>



Imputation

Replacing NAs

Visualization

Plotting NA distributions

Stats & Data

Statistics about the NAs



Features of the package

Imputation

Several functions for replacing missing data (imputation). Simple and advanced functions.

For example:

`na_interpolation()`

`na_kalman()`

Visualization

Multiple functions for visualization of time series missing values.
(updated in 3.1 version)

For example:

`na_intervals()`

`na_imputations()`

Stats & more

Printing missing data statistics and benchmark datasets for time series missing data.

For example:

`statsNA()`

`tsNH4`, `tsHeating`



Top 3

Time series missing data plots



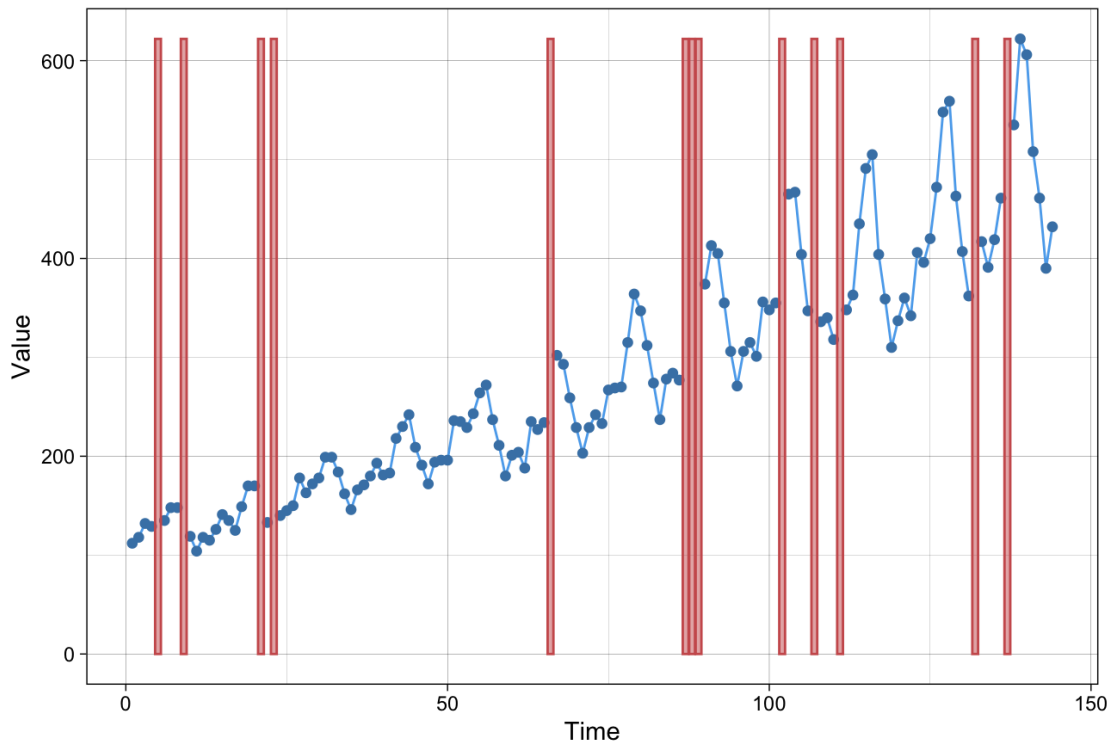
Visualization of NA distribution

`ggplot_na_distribution()`

Getting an overview, where in the time series NAs are occurring.

```
tsAirgap %>%  
ggplot_na_distribution()
```

Distribution of Missing Values
Time Series with highlighted missing regions



Discovering Patterns

`ggplot_na_gapsize()`

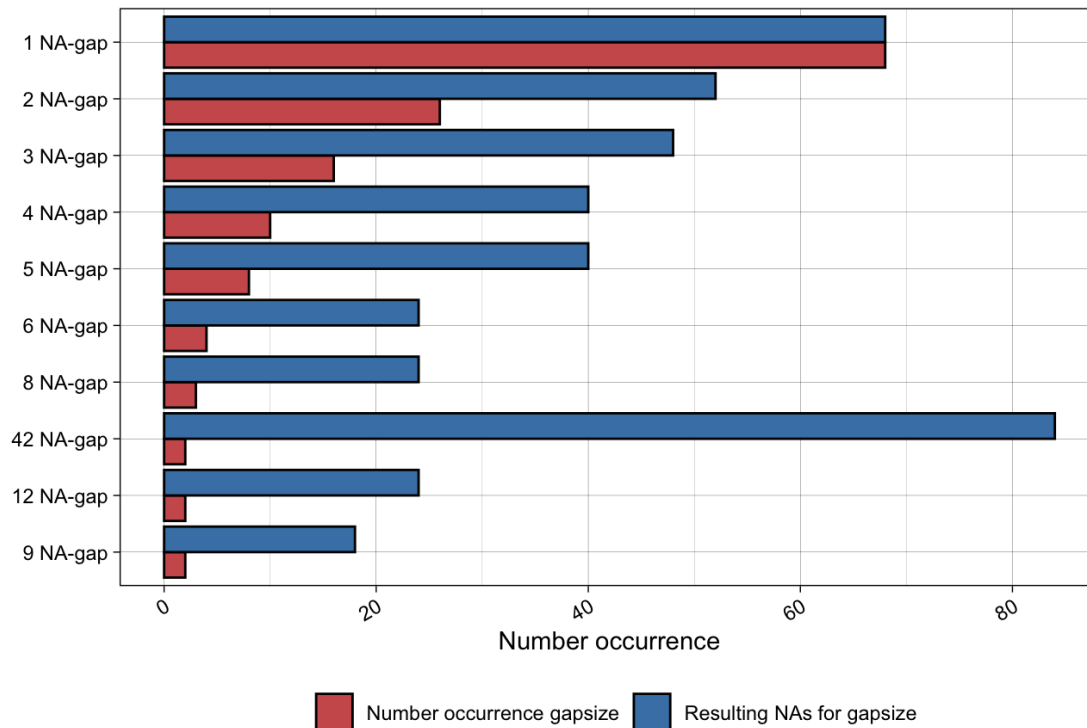
For discovering patterns in the occurrence of missing data.

`tsNH4 %>%`

`ggplot_na_gapsize()`

Occurrence of gap sizes

Gap sizes (NAs in a row) ordered by most common



Visualization of Imputations

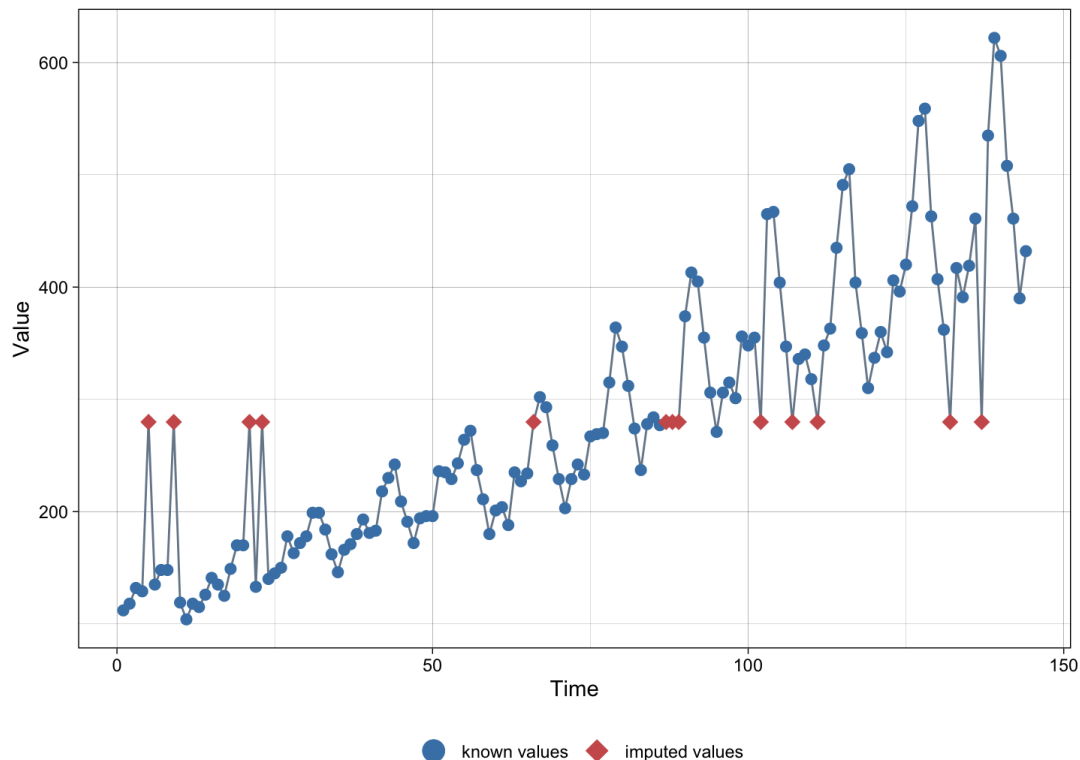
ggplot_na_imputations()

For assessing the performance of NA replacement (imputation) algorithms.

```
series <- tsAirgap  
imp <- na_mean(tsAirgap)
```

```
ggplot_na_imputations(series, imp)
```

Imputed Values
Visualization of missing value replacements



Visualization of Imputations

ggplot_na_imputations()

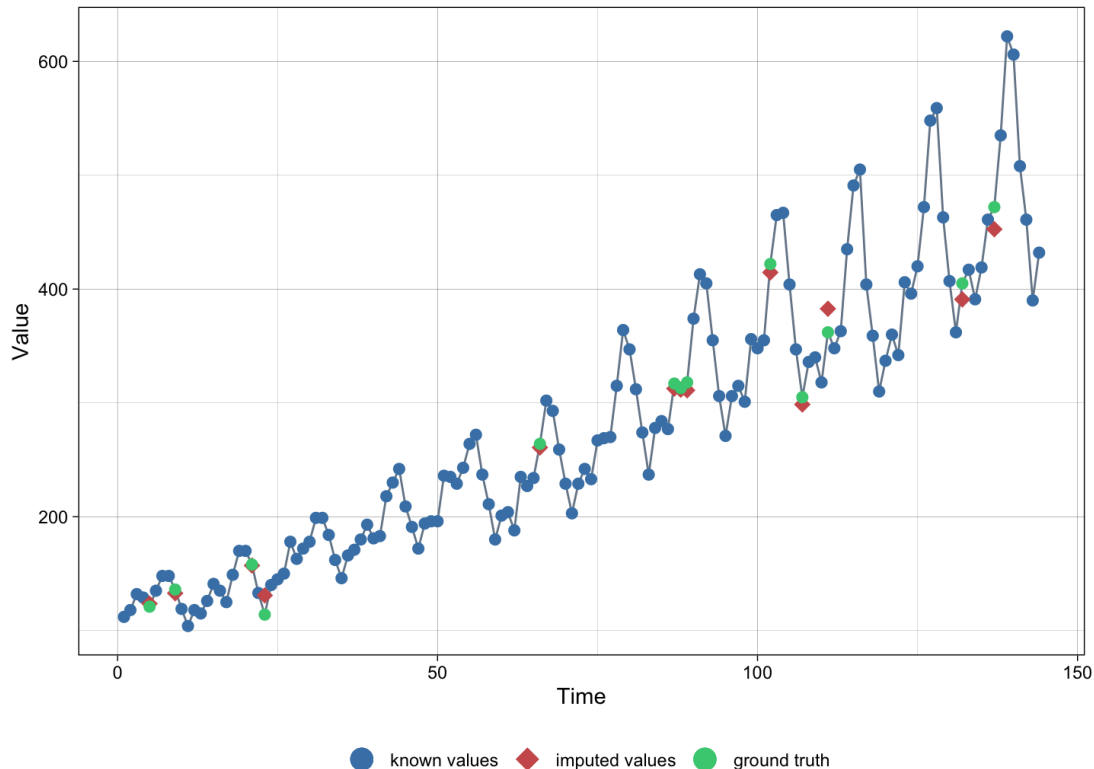
For assessing the performance of imputation algorithms.

```
series <- tsAirgap  
imp <- na_kalman(tsAirgap)
```

```
ggplot_na_imputations(series, imp,  
tsAirgapComplete)
```

Imputed Values

Visualization of missing value replacements



Interesting other packages about missing data visualization

naniar

Cross-Sectional Data, Tidy Style

Missing Data Task View

<https://cran.r-project.org/web/views/MissingData.html>

VIM

Cross-Sectional Data, Patterns, Distribution

visdat

General Data Exploration



A light gray background with a complex network of interconnected nodes and lines, resembling a molecular or data network structure. The nodes are represented by small circles, some solid and some hollow, connected by thin lines.

Thanks.

Finished! Feel free to contact me for questions and suggestions.

also visit <https://github.com/SteffenMoritz/imputeTS> for more information.