

Time Series Imputation with imputeTS: : CHEAT SHEET



Mission

Missing Data is nearly everywhere. Also in time series, especially in sensor recordings missing data is common.

imputeTS helps you with your missing data problems.

Features

The package provides easy to use functions in these areas:

1. Imputation Functions

Several algorithms for replacing NAs with reasonable values (imputation).



2. Missing Data Visualizations

Plots for analysis of the distribution of NAs, patterns and imputation performance.



3. Stats and Datasets

Functions for printing missing data stats and benchmarking datasets.



Scope

imputeTS specializes on univariate time series that are:

- numeric
- equally-spaced



Visualizations

There are multiple plots provided for analysing the missing data before and after imputation. All plotting functions start with `ggplot_na_` prefix.

Function	Description
<code>ggplot_na_distribution</code>	Getting a first overview of NAs
<code>ggplot_na_intervals</code>	Insights about NAs in specific periods
<code>ggplot_na_gapsizes</code>	Insights about occurring NA gapsizes
<code>ggplot_na_imputations</code>	Evaluating imputation quality

Imputation

The package offers multiple missing data replacement (imputation) functions, which are really easy to use.

```
na_interpolation(x, option = "spline")
```

Imputation Function

Your input time series

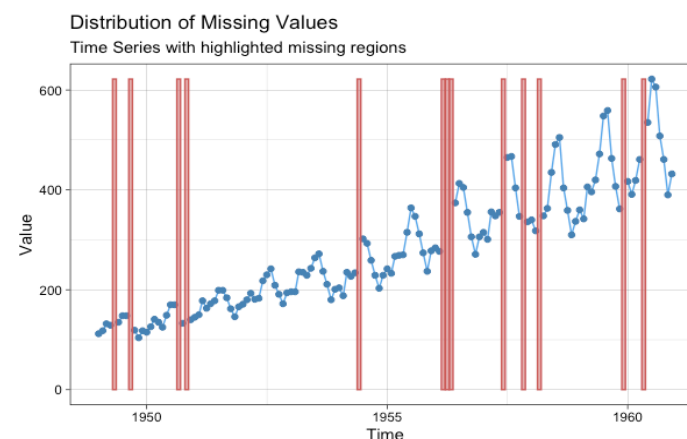
Additional Parameters

List of available Algorithms

Function	Description
<code>na_interpolation</code>	Imputation by Interpolation
<code>na_kalman</code>	Imputation by Kalman Smoothing
<code>na_locf</code>	Last Observation Carried Forward
<code>na_ma</code>	Imputation by Moving Average
<code>na_mean</code>	Imputation by Mean Value
<code>na_random</code>	Imputation by Random Sample
<code>na_remove</code>	Remove Missing Values
<code>na_replace</code>	Replace Missing Values by a Defined Value
<code>na_seadec</code>	Seasonally Decomposed Imputation
<code>na_seasplit</code>	Seasonally Split Imputation

Missing Data Overview Plots

The 'distribution', 'intervals' and 'gapsizes' plots can be used on new datasets to gain insights about missing data patterns and distribution.

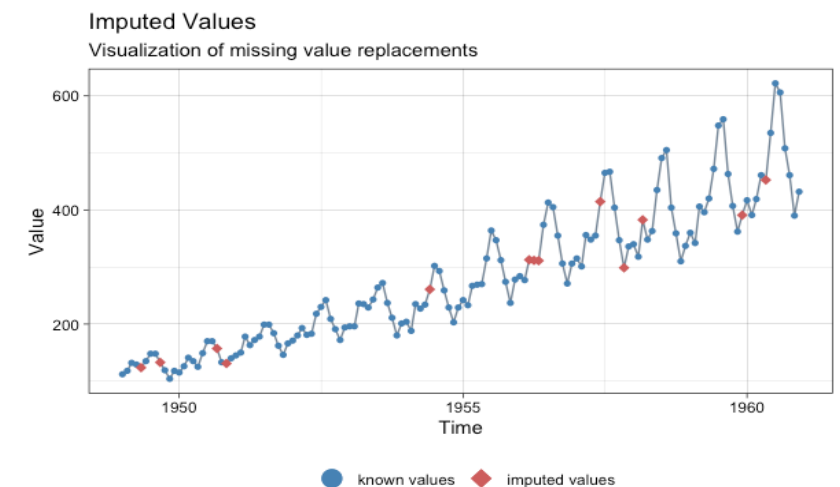


```
ggplot_na_distribution(tsAirgap)
```

example data with NAs

Imputation Analysis Plots

Imputation results can be visualized with the 'imputations' plot. Here first `na_kalman` is performed and then the results are plotted.



```
imp <- na_kalman(tsAirgap)
ggplot_na_imputations(tsAirgap, imp)
```

imputation

visualization

Workflows

The functions also work well in tidy style pipe workflows. Here an example of first using imputation and later forecasting and plotting.

```
library("forecast")
tsAirgap %>% na_interpolation() %>%
ets() %>% forecast(h=36) %>%
autoplot()
```

can be put in pipe workflows

a 36 step forecast is created and plotted

Datasets

The package includes three datasets for imputation experiments.

Function	Description
<code>tsAirgap</code>	Monthly totals of international airline passengers. 144 Observations / 13 NAs
<code>tsNH4</code>	NH4 concentration in a wastewater system. 3552 observations / 883 NAs
<code>tsHeating</code>	A heating systems supply temperature. 606837 observations / 57391 NAs

CITATION

You can cite imputeTS the following:

Moritz, Steffen and Bartz-Beielstein, Thomas. "imputeTS: Time Series Missing Value Imputation in R." R Journal 9.1 (2017). doi: 10.32614/RJ-2017-009.

