



imputeTS

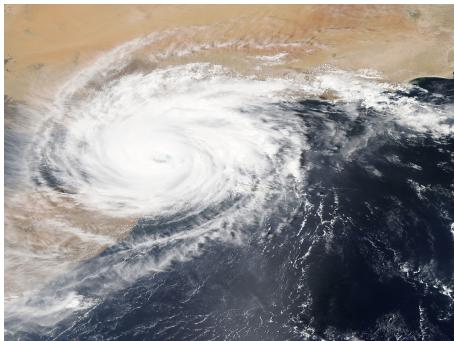
Umgang mit fehlenden Werten in Zeitreihen

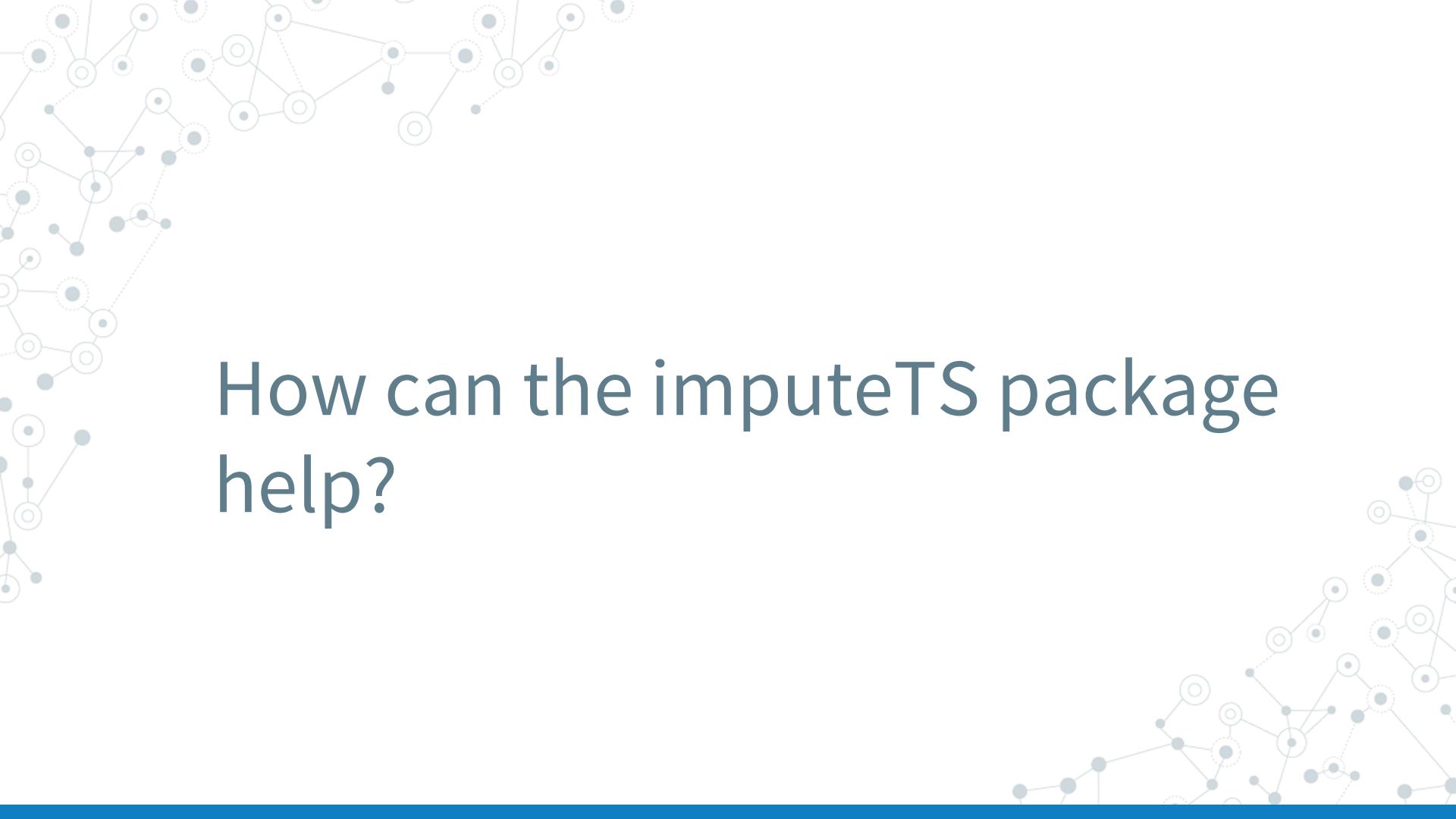
Steffen Moritz

Institut für Data Science, Engineering, and Analytics, TH Köln
steffen.moritz10@gmail.com

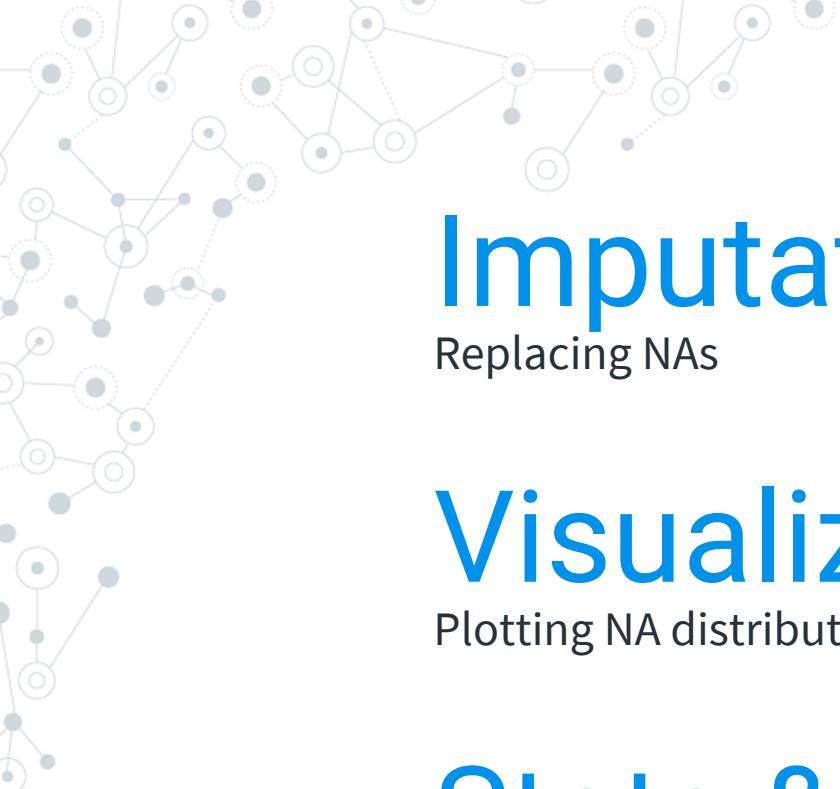
Time Series Data

- Missing data is quite common, especially with sensor data
- Typical problems are:
 - Data Recording
 - Data Transmission
 - Data Processing





How can the imputeTS package help?



Imputation

Replacing NAs



Visualization

Plotting NA distributions

Stats & Data

Statistics about the NAs

Features **imputeTS** package

Visualization

Multiple functions for visualization of time series missing values.

For example:

```
ggplot_na_intervals()  
ggplot_na_imputations()
```

Imputation

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

For example:

```
na_interpolation()  
na_kalman()
```

Stats & more

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

For example:

```
statsNA()  
tsNH4, tsHeating
```



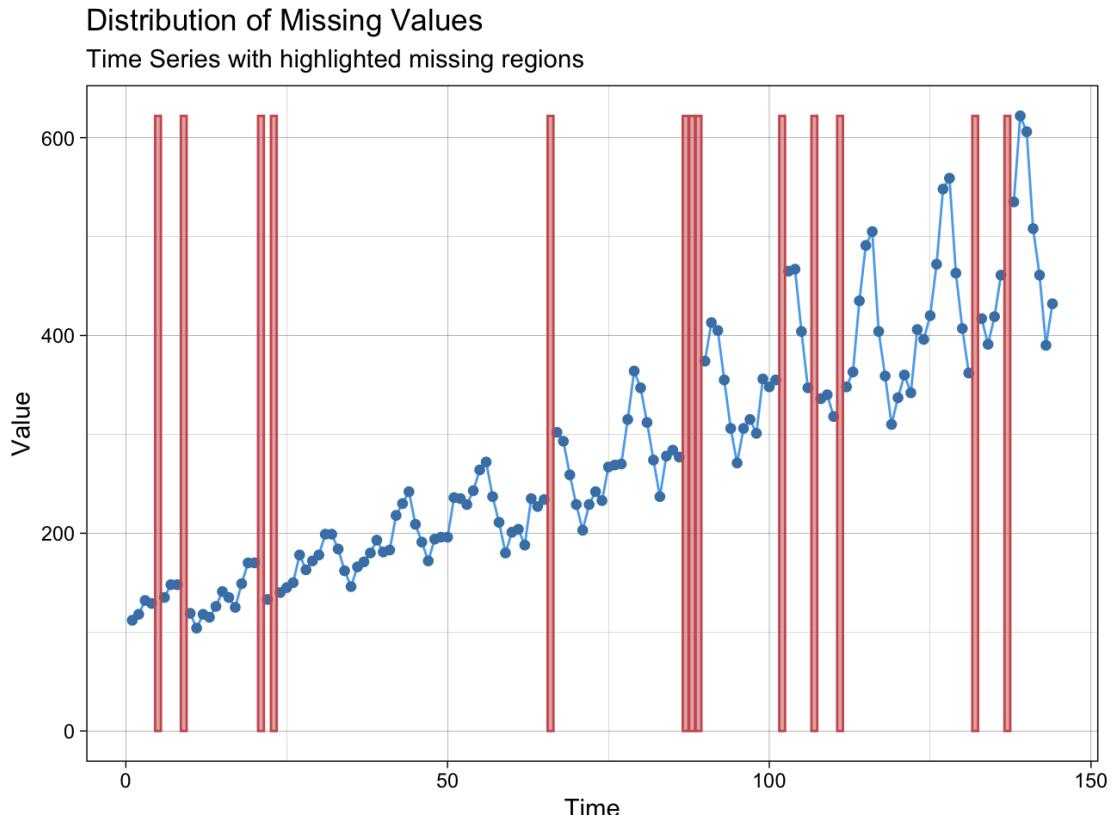
VISUALIZATION DATA EXPLORATION

Visualization of NA distribution

`ggplot_na_distribution()`

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

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



Discovering Patterns

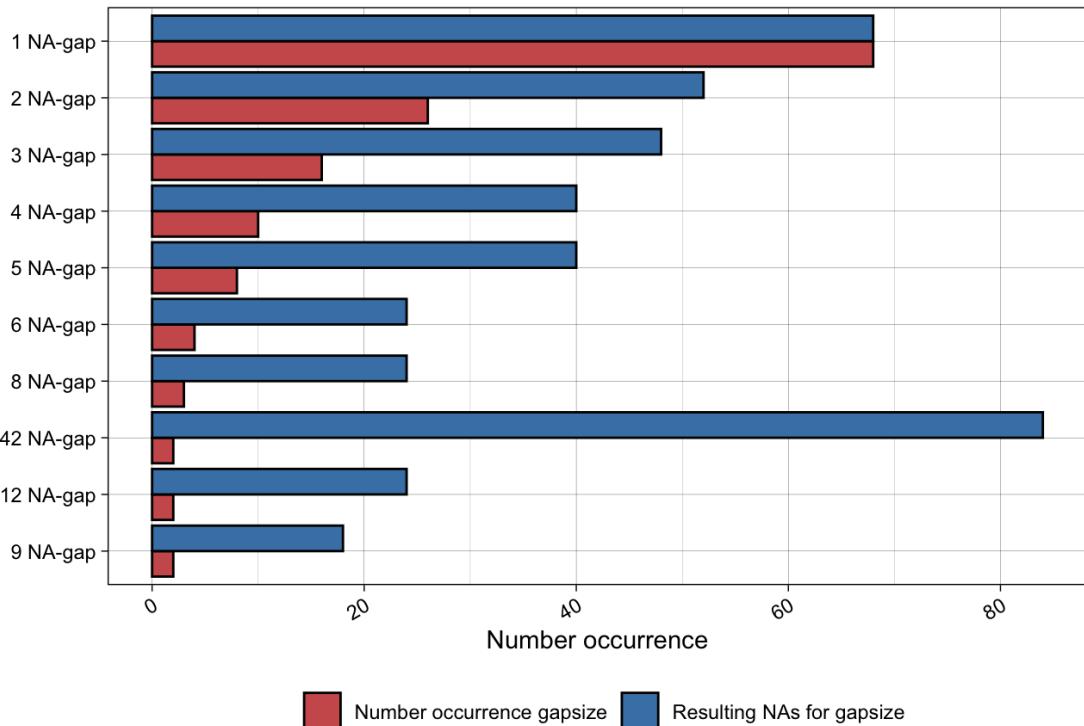
`ggplot_na_gapsize()`

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



statsNA(tsNH4)

Length of time series:

4552

Number of Missing Values:

883

Percentage of Missing Values:

19.4%

Number of Gaps:

155

Average Gap Size:

5.696774

Longest NA gap (series of consecutive NAs)

157 in a row

Most frequent gap size (series of consecutive NA series)

1 NA in a row (occurring 68 times)

Gap size accounting for most NAs

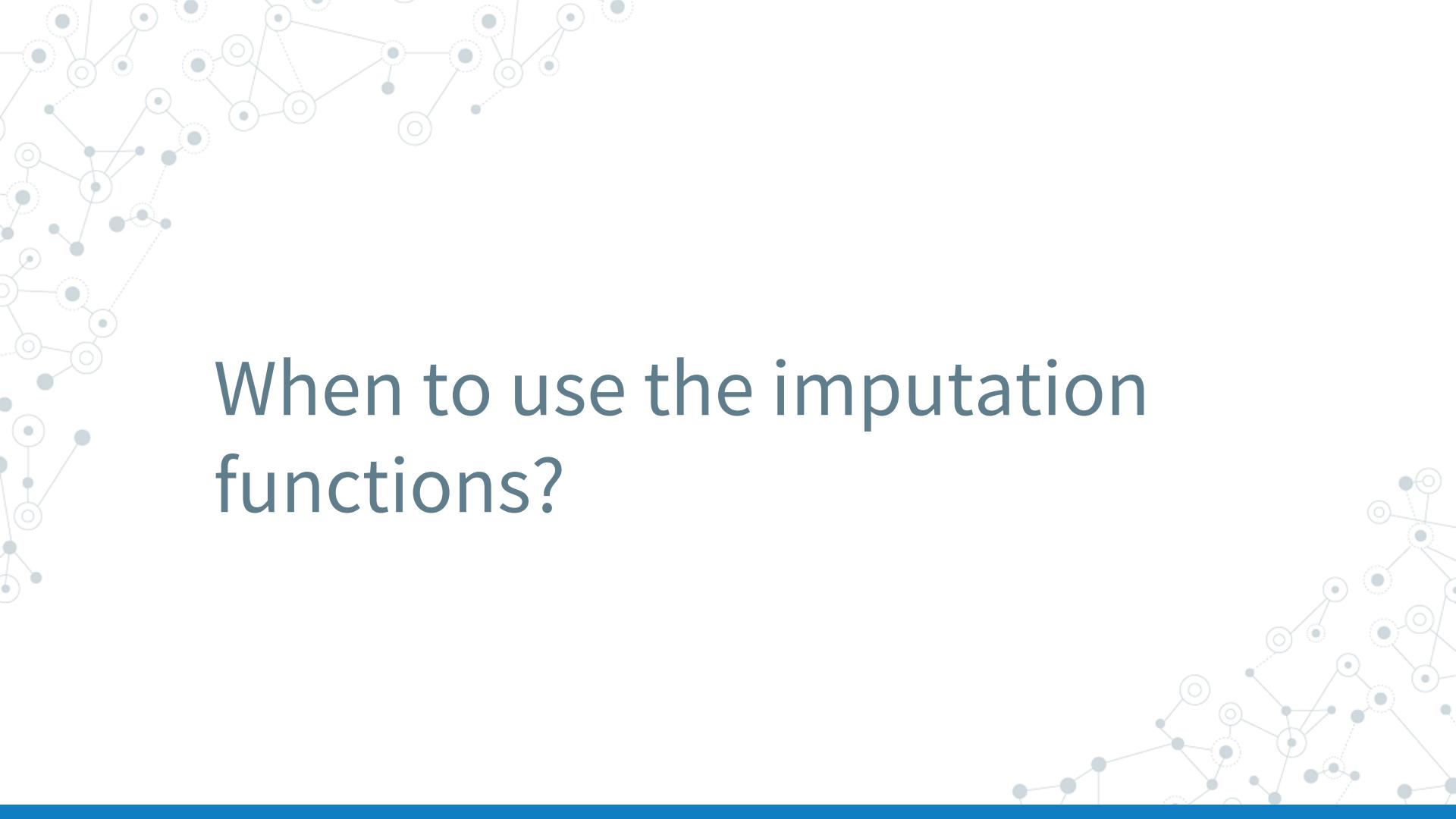
157 NA in a row (occurring 1 times, making up for overall 157 NAs)

Overview NA series

1 NA in a row: 68 times"

2 NA in a row: 26 times"

...



When to use the imputation functions?

Different missing data problems

V1	V2	V3	V4
91	91	91	91
NA	13	13	13
14	14	14	14
55	55	55	55
19	19	19	19
32	32	32	32
23	23	23	23
27	27	27	27
67	67	67	67



Cross Sectional

inter-variable

Time	V1	V2	V3
t1	13	33	15
t2	13	34	NA
t3	13	35	15
t4	13	36	16
t5	13	37	16
t6	14	38	16
t7	14	39	16
t8	14	40	17
t9	14	41	17



TS Cross Sectional

inter-variable + inter-time

Time	V1
t1	12
t2	12
t3	NA
t4	13
t5	13
t6	13
t7	14
t8	14
t9	14



Time Series

inter-time

All Variables Missing Simultaneously

Time	V1	V2	V3
t1	13	33	15
t2	NA	NA	NA
t3	NA	NA	NA
t4	13	36	16
t5	NA	NA	NA
t6	NA	NA	NA
t7	14	39	16
t8	14	40	17
t9	NA	NA	NA



Problem:

**Only whole observations are missing
(V1,V2,V3 at one point in time)**

**This is often common for transmission
problems**

**Thus inter-variable correlation can not
be sufficiently employed**

TS Cross Sectional

--> Pure time series imputation needed



IMPUTATION REPLACING MISSING VALUES

Available Imputation Functions

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



Usage Example without and with pipe

1. `na_kalman(tsAirgap)`

2. `tsAirgap %>%
na_seadec() %>%
ets() %>%
forecast(h=36)`



Imputation functions take all kinds of inputs:

ts, mts, data.frame, zoo, xts, vector, tibble, tsibble

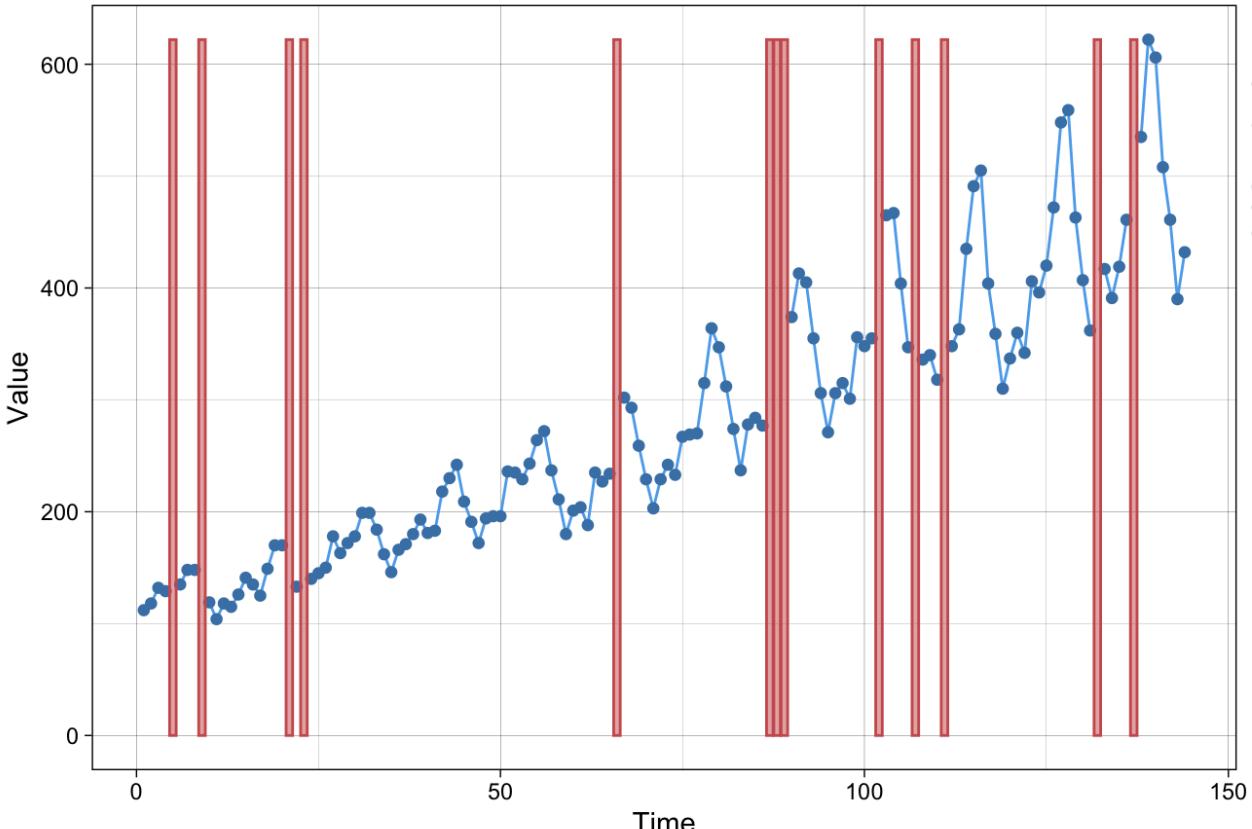


`na_‘algorithmname’(yourInput, add. param)`

Similar syntax also used by other packages like zoo, forecast

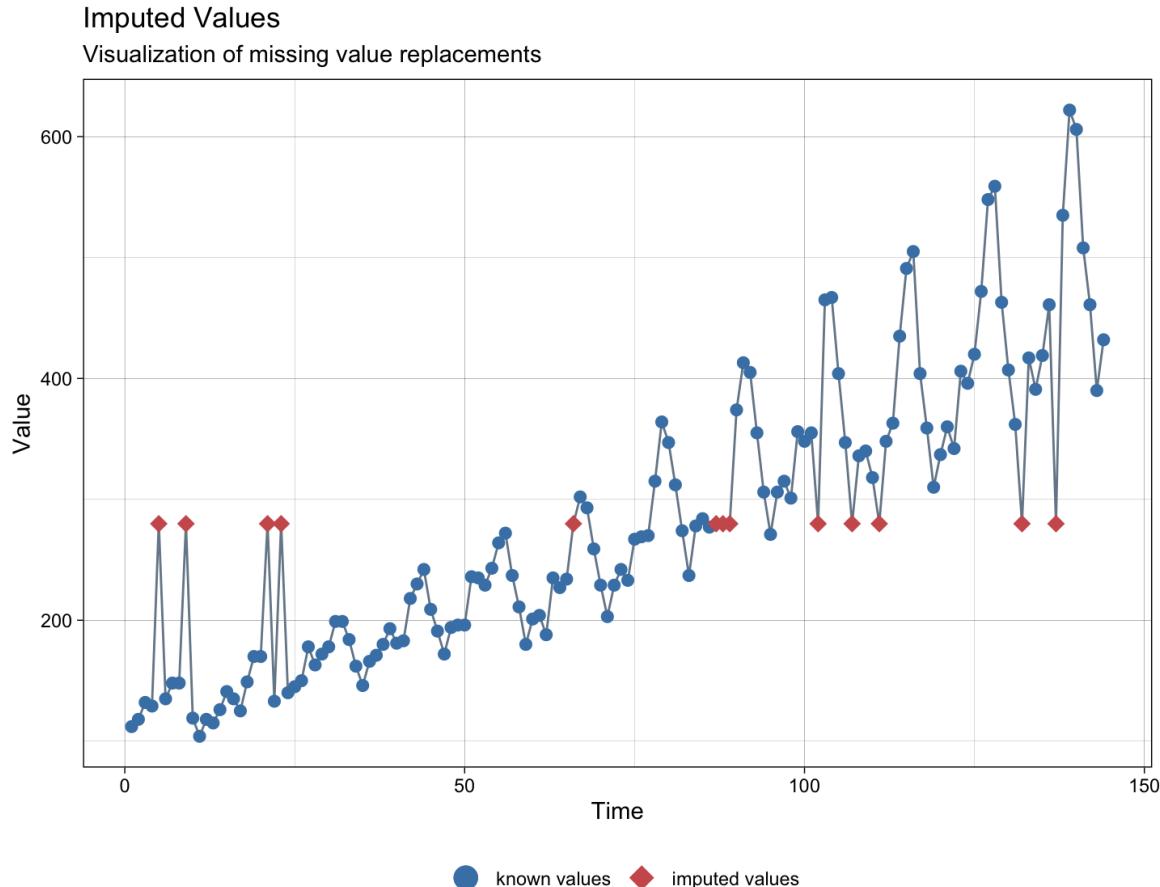
Examples with tsAirgap series

Distribution of Missing Values
Time Series with highlighted missing regions



Usage example na_mean

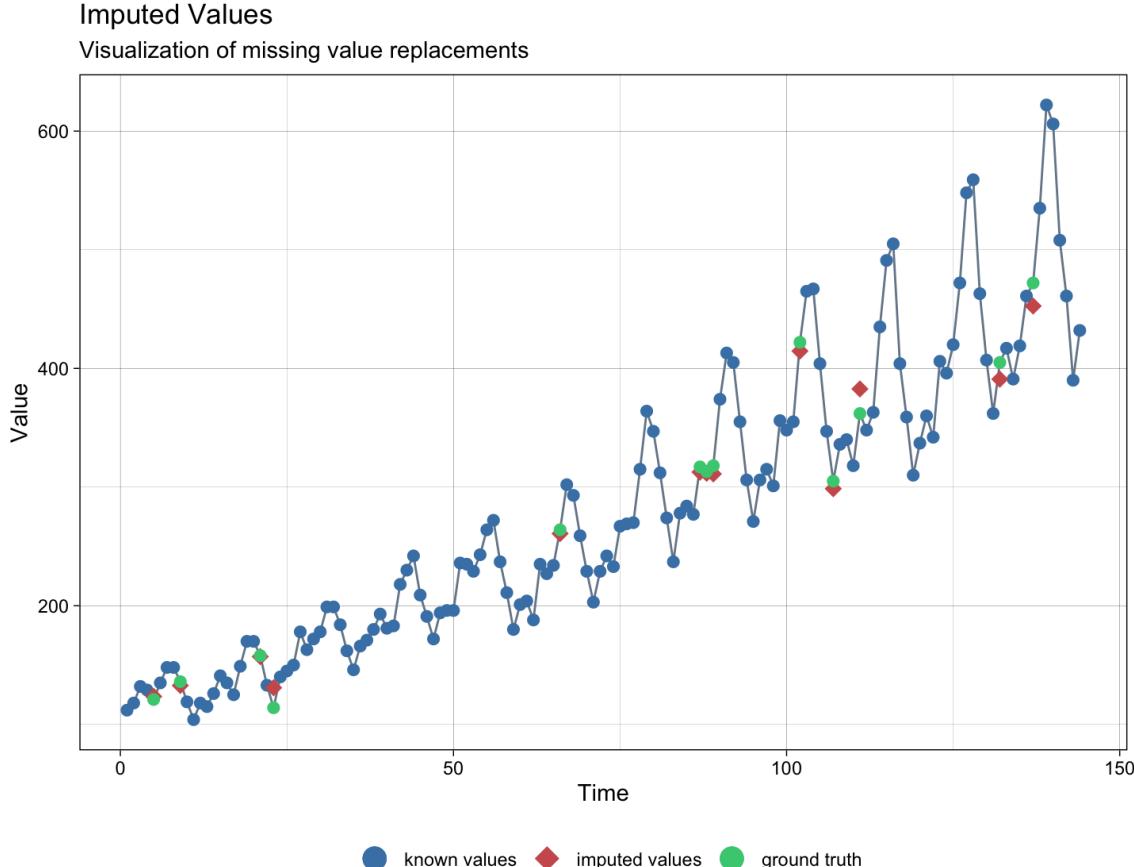
```
imp <- na_mean(tsAirgap)  
ggplot_na_imputations(tsAirgap, imp)
```



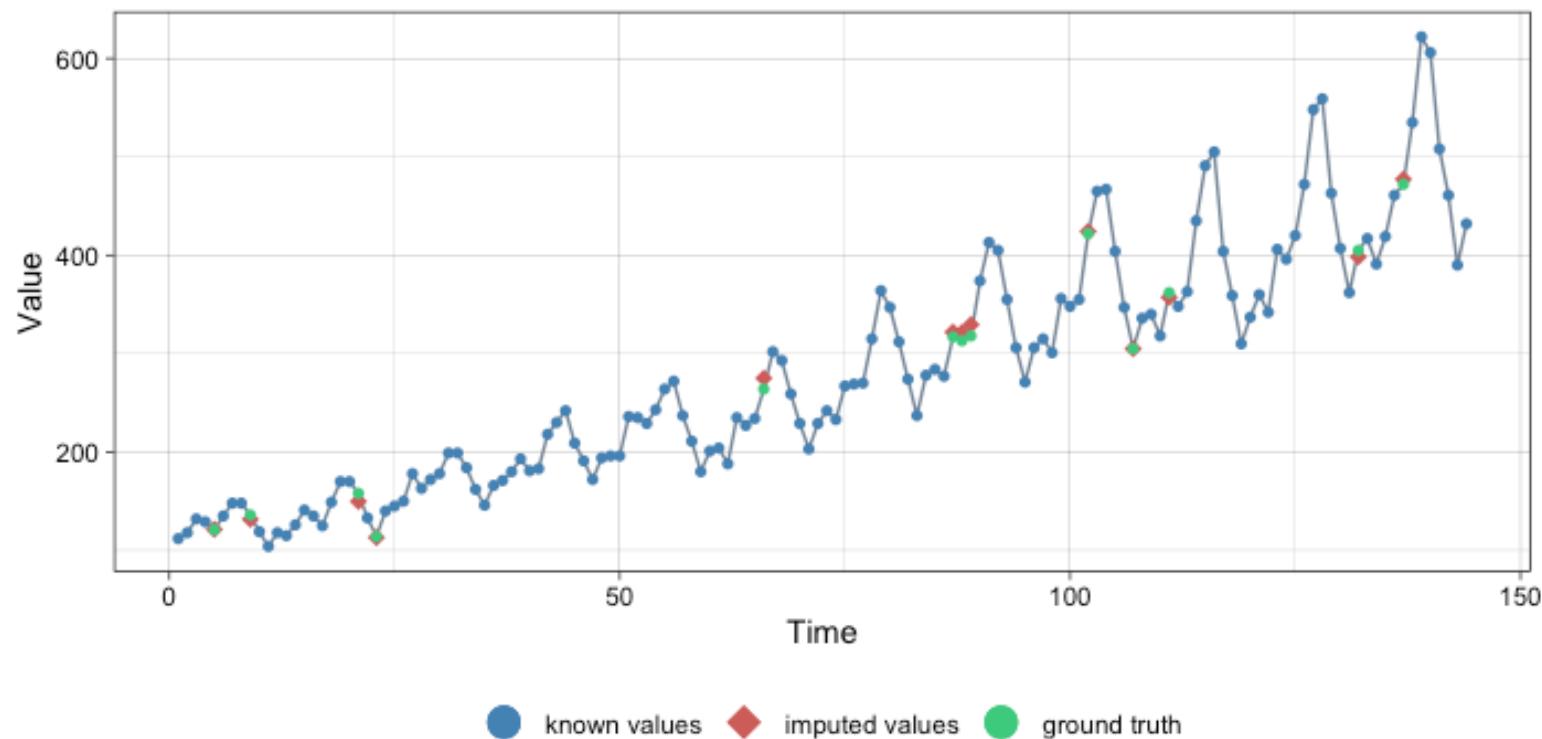
Usage example na_kalman

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

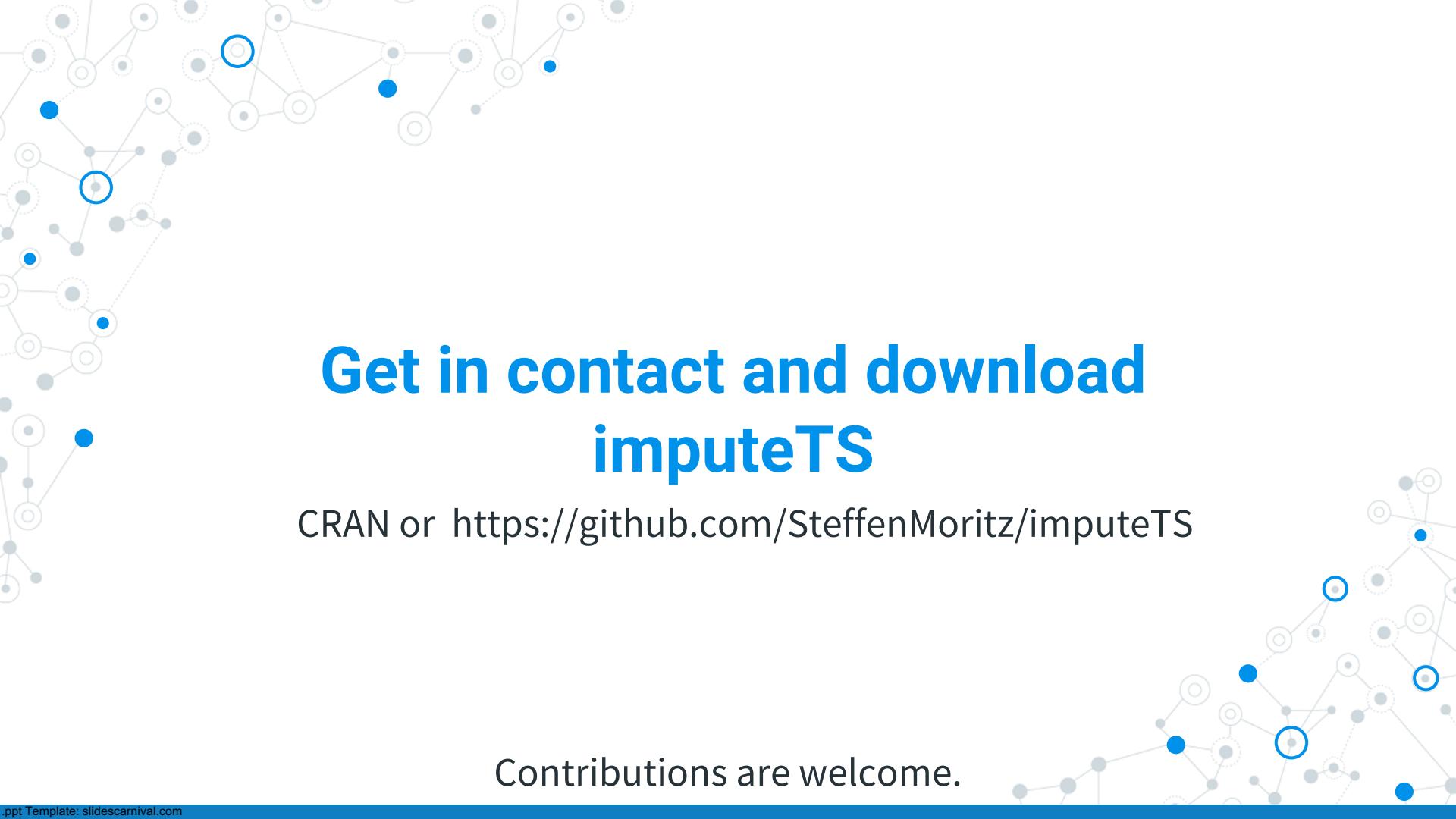
```
ggplot_na_imputations(tsAirgap,  
                      imp, tsAirgapComplete)
```



Seasonally Decomposed Missing Value Imputation (na_seadec)



```
ggplot_na_imputations(tsAirgap, na_seadec(tsAirgap), tsAirgapComplete)
```



Get in contact and download imputeTS

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

Contributions are welcome.