

**Online-Zertifikatslehrgang** 

## Data Analyst IHK

Die neue Generation digitaler IHK-Weiterbildungen



IHK■Die Weiterbildung



## **INHALT**

KNIME Knoten
und ihre Konfiguration
Modul 1

Data Analyst<sub>(ІНК)</sub>



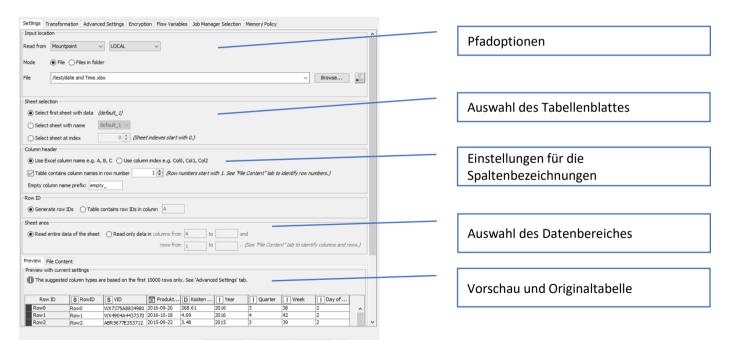


# Importknoten: Beispiel Excelreader

## Settings:

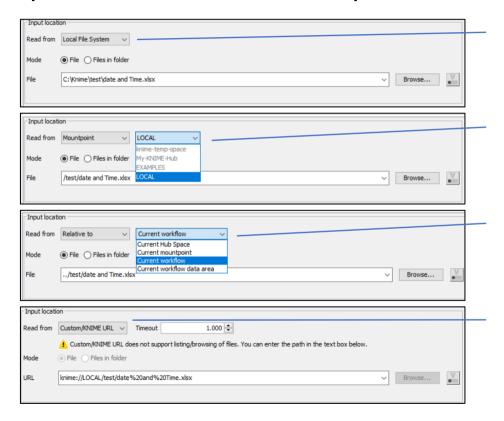
#### **Excel Reader**





Cancel

# Importknoten: Pfadoptionen



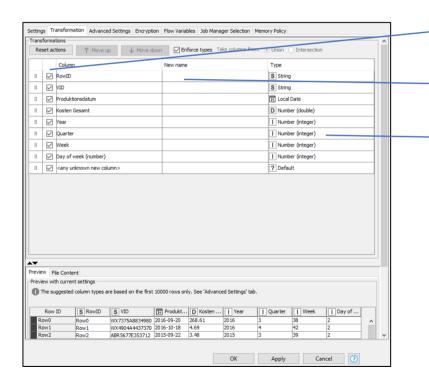
**Local File System**: Der Pfad wird entsprechend des lokalen Betriebssystems (Windows, Mac) geschrieben

**Mountpoint**: Der Pfad beginnt am gewählten KNIME-Mountpoint. In der Regel ist dies LOCAL.

**Releative to**: Der Pfad wird in einem relativen Ausdruck zu einem Bezugspunkt angegeben. Häufig ist dies der Workflow

**Custom/KNIME URL**: Der Pfad wird als URL formuliert.

# Importknoten: Transformation

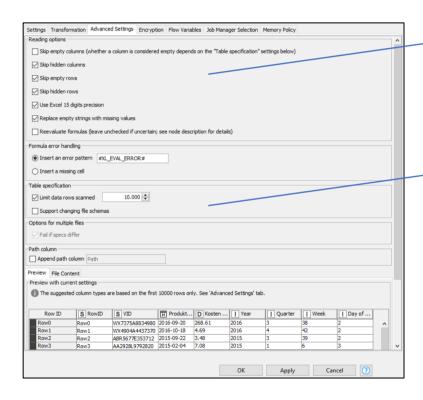


Auswahl der zu importierenden Spalten

Änderung des Spaltennamens

Auswahl des Datentyps

# Importknoten: Advanced Settings



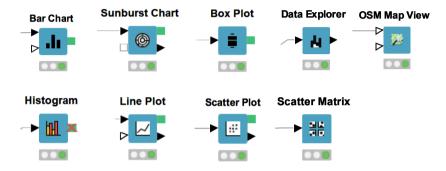
Einstellungen für leere oder verborgene Zeilen und Spalten

Begrenzung der Vorschau der Daten

## Knoten zur Datenvisualisierung

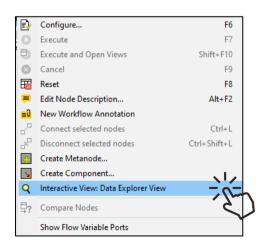
#### Meist gebräuchlich:

JavaScript Nodes



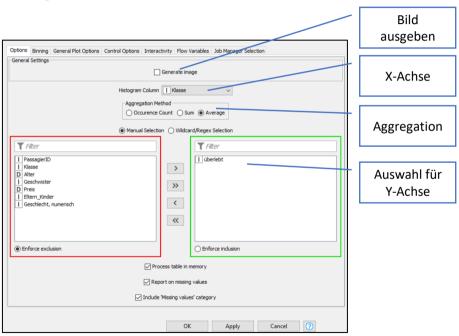
## Visualisierungsknoten haben interaktive Ansichten:

- a) Rechtsklick auf Knoten → Interactive View
- b) Oder F10 Taste



## Konfiguration: Options

#### Histogramm

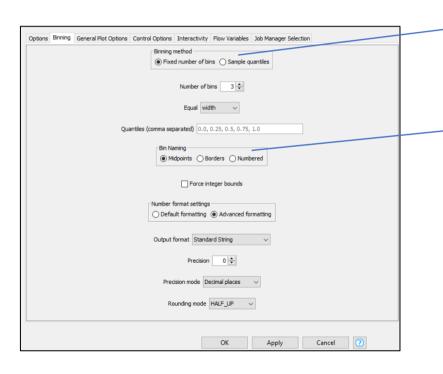


#### Scatter Plot



# Konfiguration: Binning (Histogramm)

Mit "Binning" werden Zahlendaten zu Blöcken zusammengefasst

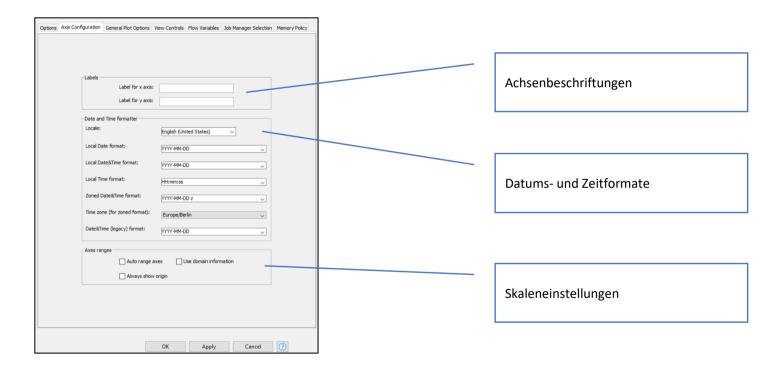


Binning-Methode (Anzahl der Säulen):

- Vorgegebene Anzahl
- Anzahl der einzigartigen Werte

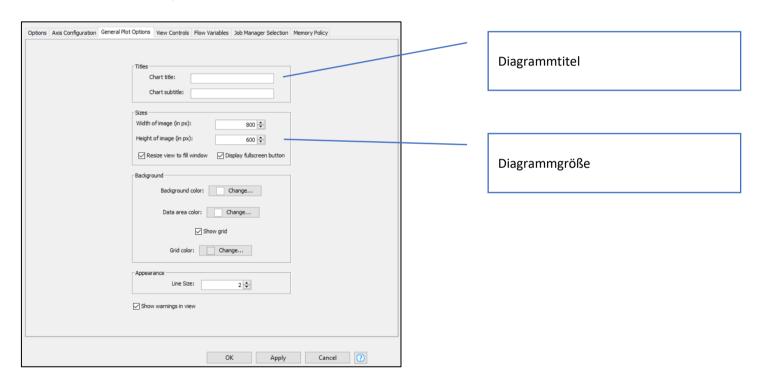
Position der Säulenbezeichnungen

## Konfiguration: Axis Configuration (Scatter, Line, etc.)



# Konfiguration: General Plot Options

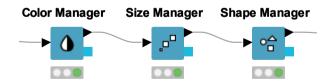
Ist in allen Java Script Views enthalten

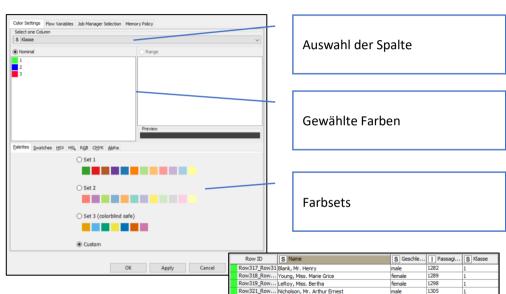


## Color Manager, Shape Manager, Size Manager

Datenpunkte erhalten mit dem View Manager Knoten :

- Farbe
- Größe
- Form





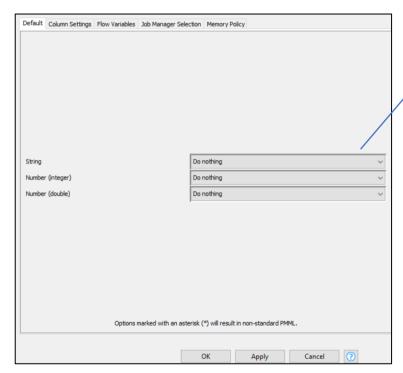
Die Konfiguration wird mit der Tabelle verknüpft im Workflow abwärts weitergegeben.





## Knoten Missing Value: Default



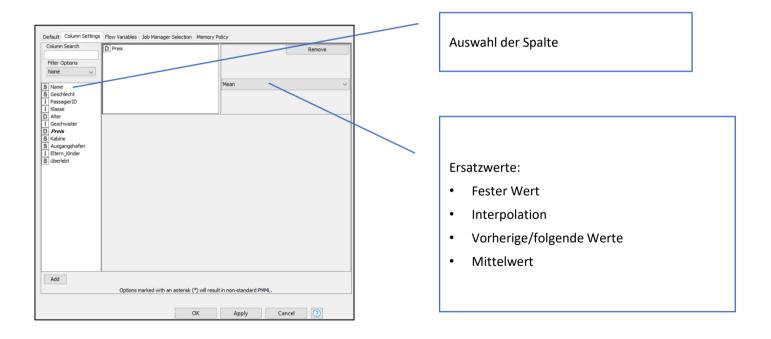


Einstellung für Ersatzwerte über die ganze Tabelle hinweg:

- Fester Wert
- Interpolation
- Vorherige/folgende Werte
- Mittelwert



## Knoten Missing Value: Column Settings

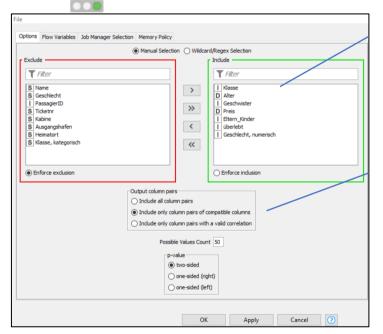


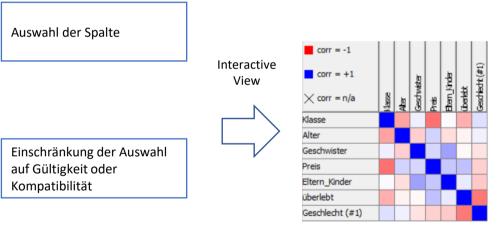
## **Knoten Linear Correlation**

#### **Linear Correlation**



• Berechnet die Korrelationskoeffizienten zwischen numerischen Variablen



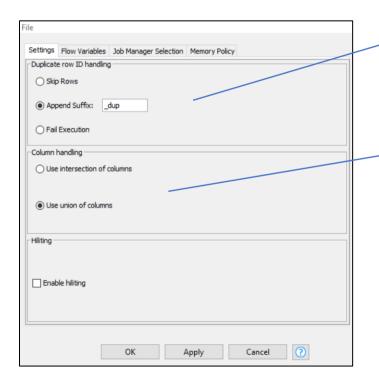






## Knoten Concatenate





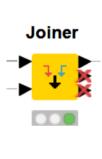
Einstellung für doppelt vorkommende Row-IDs

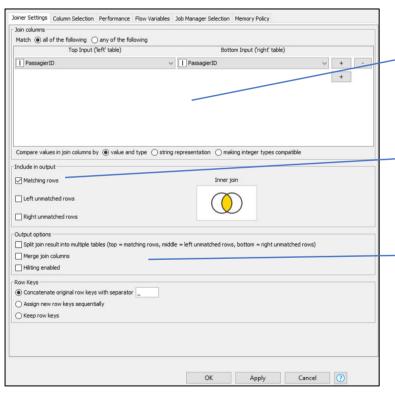
**Intersection**: Nur Spalten, die in allen Tabellen auftauchen werden übernommen.

**Union**: Übernahme aller Spalten aus allen Tabellen.

In einer Tabelle nicht vorhandene Zellen werden als "?" angezeigt.

## Knoten Joiner: Joiner Settings





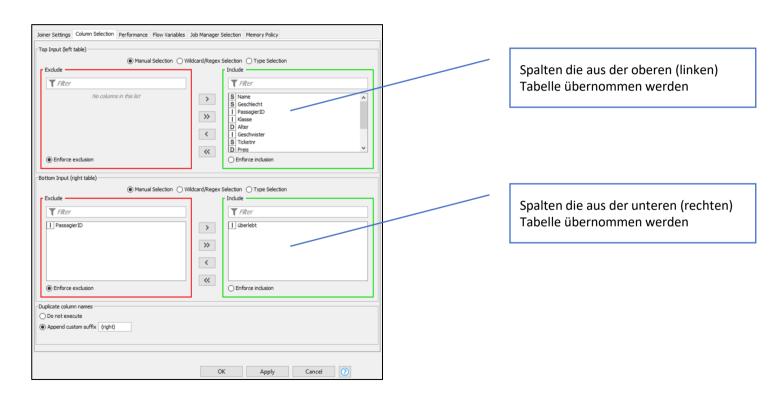
Auswahl der Schlüssel zum Verbinden der Tabellen. Es können auch mehrere verwendet werden.

Join-Methode: Inner Join, Left Outer Join, Right Outer Join und Full Outer Join

Einstellungen für den Output:

- 1. Aufteilung bei Outer Joins in mehrere Tabellen
- 2. Daten aus gleichen Spaltenbezeichnungen zusammenlegen

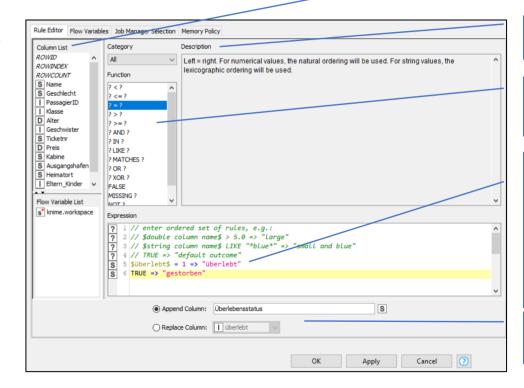
## Knoten Joiner: Column Selection



# Knoten Rule Engine

#### **Rule Engine**





Auswahl der Spalten für die Regeln.

Beschreibung der Funktionen

Mögliche Funktionen für die Abfrage

Liste der erstellten Regeln:

#### Beispiel:

\$überlebt\$ = 1 => "überlebt"

#### **Bedeutung:**

Wenn in der Spalte "überlebt" der Wert gleich 1 ist, soll in der Ausgabespalte der Wert "überlebt" ausgegeben werden.

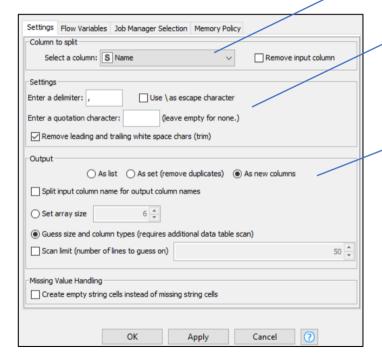
Wahl der Ausgabespalte

# Knoten Cell Splitter

#### **Cell Splitter**



Teilt Strings anhand von Trennzeichen auf



Auswahl der Spalte in der die Inhalte aufgeteilt werden sollen

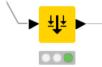
#### Auftrennungseinstellung:

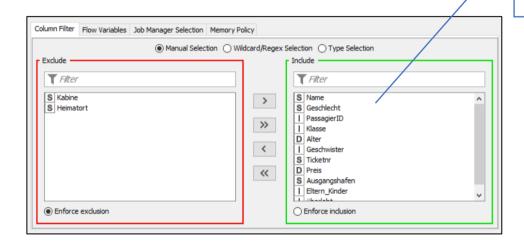
- Trennzeichen, bspw. ein Komma, Semikolon oder Leerzeichen
- Ausnahmen und Kommentare

Ausgabe: Als Liste oder neue Spalten

## Knoten Column Filter

#### Column Filter



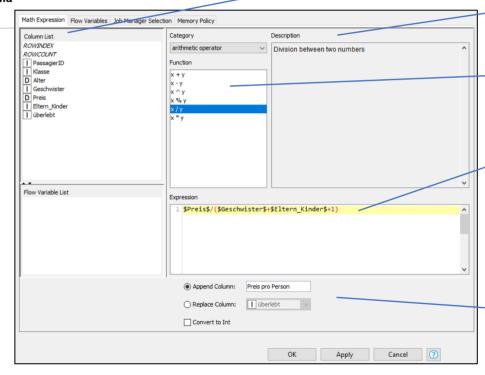


Spalten, die in der Tabelle erhalten bleiben sollen.

## Knoten Math Formula

#### **Math Formula**





Auswahl der Spalten für die Formeln.

Beschreibung der Funktionen

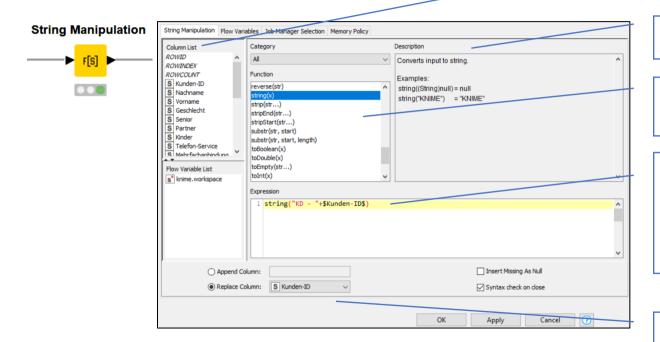
Mögliche Funktionen für die Berechnung

Formel: Die Formeln werden gemäß der Syntax von Java zusammengestellt

Zu den jeweiligen Funktionen finden sich Beschreibungen und Beispiele, wie sie aufgebaut werden

Wahl der Ausgabespalte

# **Knoten String Manipulation**



Auswahl der Spalten für die Formeln.

Beschreibung der Funktionen

Mögliche Funktionen für die Berechnung

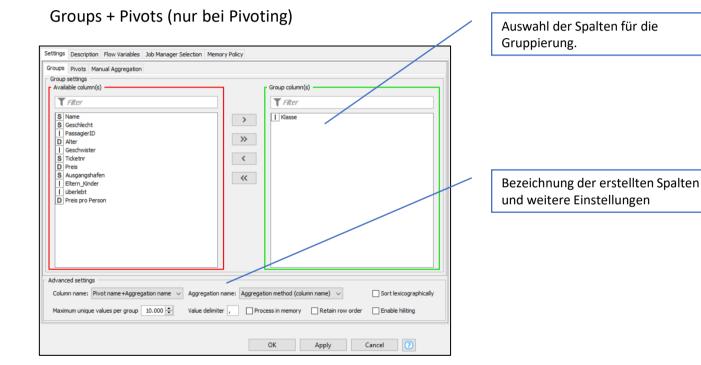
Formel: Die Formeln werden gemäß der Syntax von Java zusammengestellt

Zu den jeweiligen Funktionen finden sich Beschreibungen und Beispiele, wie sie aufgebaut werden

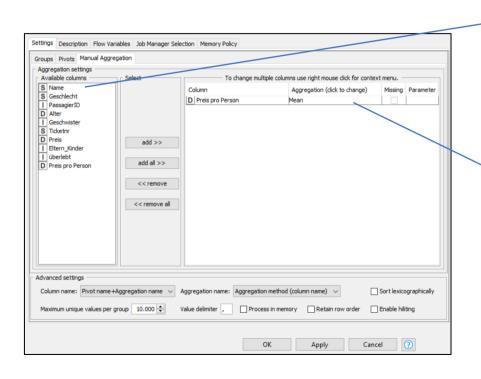
Wahl der Ausgabespalte

# Aggregationen: GroupBy und Pivoting

# GroupBy Pivoting



# GroupBy und Pivoting: Manual Agregation



Auswahl der Spalte

#### Aggregationsmethode (abhängig vom Datentyp)

- Liste
- Häufigster Wert
- Frster Wert
- Mittelwert
- Summe
- •

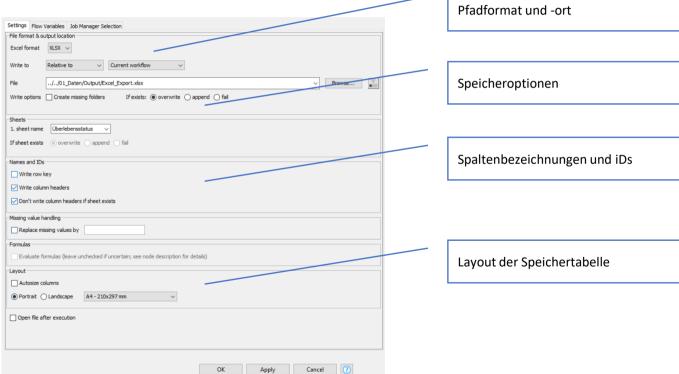




# Speicherknoten: Beispiel Excel Writer

#### **Excel Writer**









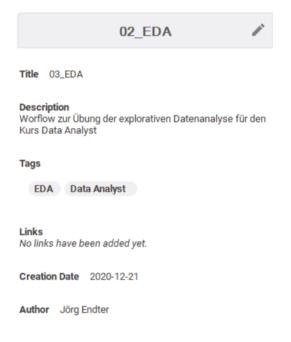
## Dokumentation in visuellen Analytics Anwendungen

#### Workflow Metainformation

Die grundlegendste Form der Dokumentation sind die begleitenden Informationstexte zu einem Workflow. Diese sind in der Regel als "Beschreibung", "Inhalt" oder "Workflow-Information" im Menü des Workflows abrufbar.

Je nach Gestaltung können hier alle wichtigen Informationen beschrieben und dokumentiert werden.

- Klick auf Workflow im Explorer
- In Description Fenster editieren



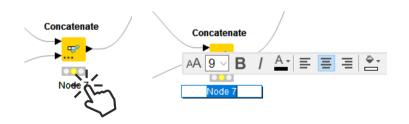
Beschreibung des Workflows in einer Infobox

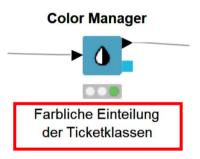
## Dokumentation in visuellen Analytics Anwendungen

## Beschreibung der Knoten

Jeder Bearbeitungsschritt sollte eine kurze Information enthalten, was in ihm durchgeführt wird und welches Ziel erreicht werden soll.

In KNIME kann man dafür die Textboxen unterhalb der Knoten verwenden.



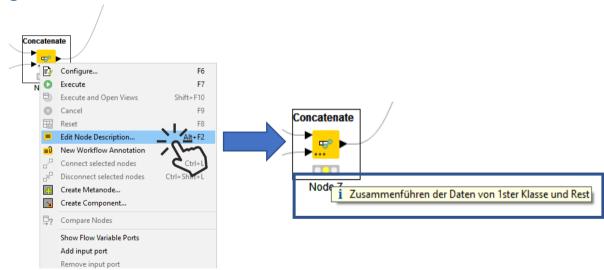


Funktionsbeschreibungen von Knoten

# Dokumentation in visuellen Analytics Anwendungen

## Zusätzliche Knoteneschreibungen mit MouseOver Funktion

- Rechtsklick auf Knoten
- Edit Node Description

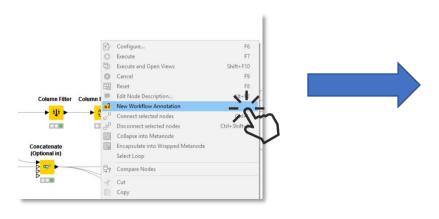


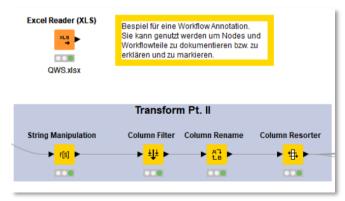
# Dokumentation in visuellen Analytics Anwendungen

Notizen und Beschreibungen zu Knotengruppen

## "Workflow Annotation"

- Rechtsklick im Editor
- "New Workflow Annotation"





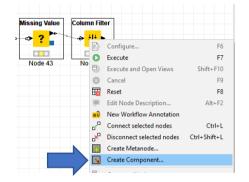
## Workflow Organisation

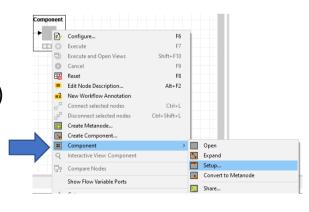
## Erstellen von Komponenten:

- Knoten markieren,
- Rechtsklick auf markierten Knoten
  - → Create Component

## Ändern von Komponenten:

- Rechtsklick auf Komponente → Component
  - Setup (Ändern der Ports und des Namens)
  - > Open (Knoten in der Komponente bearbeiten)
  - > Expand (Komponente wieder entfalten)
  - Share (Speichern und Teilen)







# Workflow Organisation

## Beschreibung in "Component Description":

## Öffnen der Komponente

Beschreibung der Komponente und Ports ändern

### Doppelklick auf Anfang/Endknoten im Inneren

Konfigurieren der Übergabe" von Flow Variables an den Workflow

