Open for Innovation
KNIME

KNIME Interactive R Statistics Integration Installation Guide

KNIME AG, Zurich, Switzerland
Version 4.6 (last updated on 2022-05-02)

Table of Contents

Introduction 1
KNIME Interactive R Statistics Integration installation 2
R installation 4
R packages installation 4
    Windows 5
    MacOS 8
    Linux Ubuntu 9
KNIME Interactive R Statistics Integration usage 10
    Configure and export R environments 10

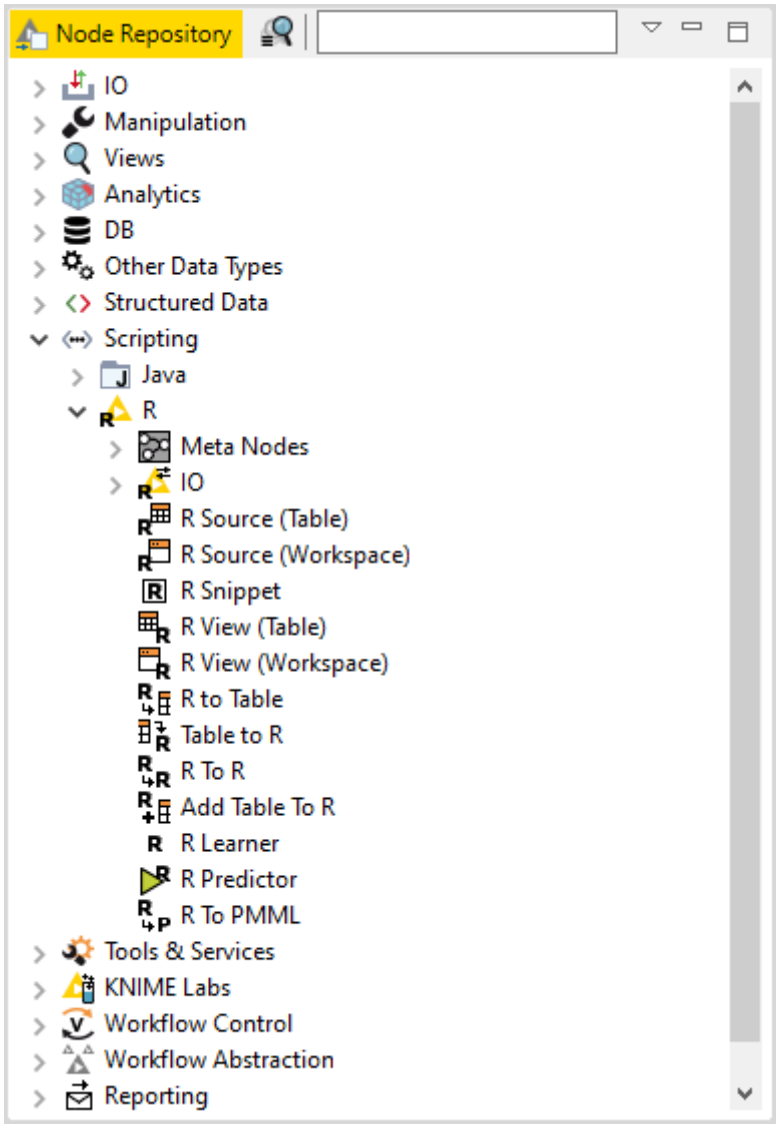
Introduction

This guide describes how to install the KNIME Interactive R Statistics Integration to be used with KNIME Analytics Platform.

The KNIME Interactive R Statistics Integration allows to write and execute R scripts by interacting with an external R™ installation.

This integration comprises a handful of nodes which can be used to write R scripts and execute them in KNIME Analytics Platform. Besides providing the ability to read data from R into KNIME Analytics Platform, these nodes allow to manipulate data and create views using R, as well as to learn and apply models, trained in R, to the data.

These nodes can be found in the node repository under *Scripting* → *R*.



This integration, in order to communicate with R, depends on additional R packages which need to be installed. In this guide we describe how to configure the KNIME Interactive R

© 2022 KNIME AG. All rights reserved.

Statistics Integration, as well as how to install R and the necessary packages.

KNIME Interactive R Statistics Integration installation

The following steps are required to install and use the KNIME Interactive R Statistics Integration.

1. First, install the KNIME Interactive R Statistics Integration.

   In KNIME Analytics Platform, go to *File* → *Install KNIME Extensions…*. The KNIME Interactive R Statistics Integration can be found under *KNIME & Extensions* or by entering *R integration* into the search box.

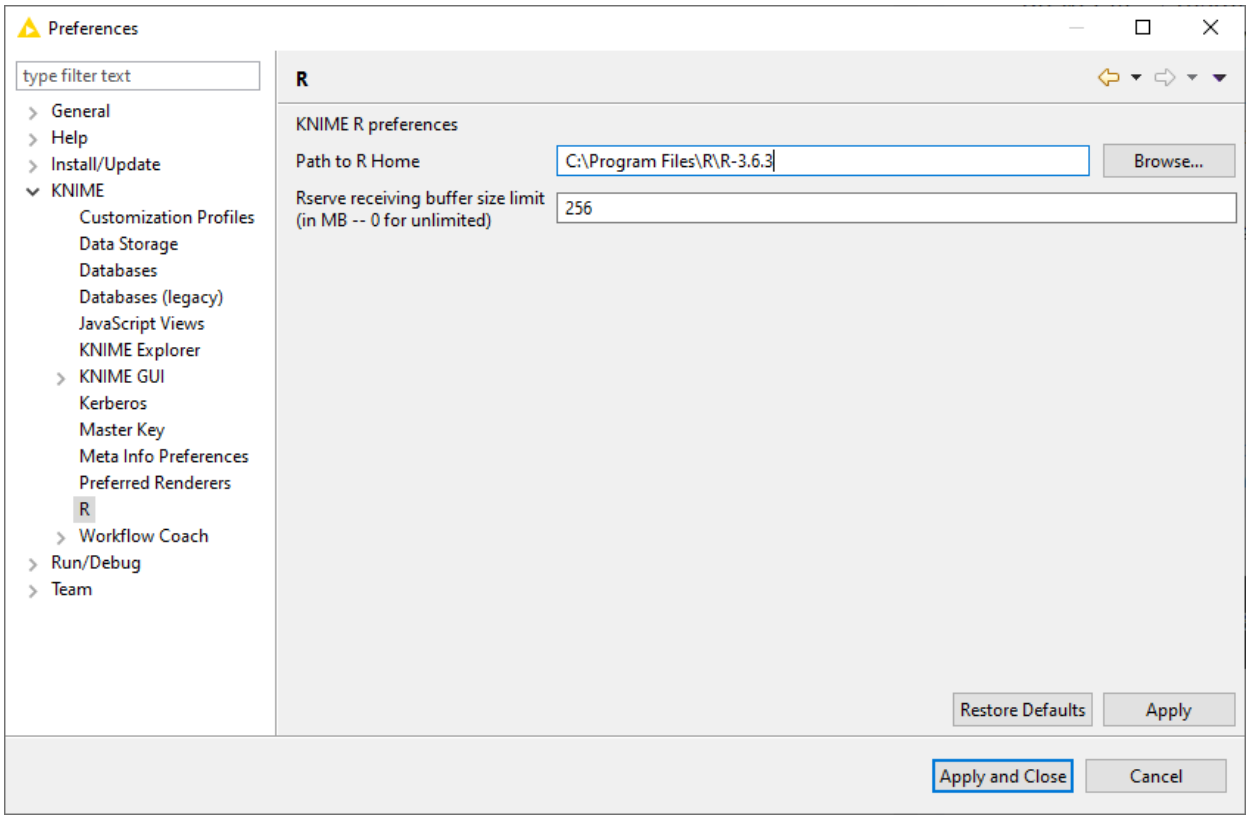
2. Download and install R and all required packages as described in the R installation and R packages installation Sections.

   > Please note that in case you are proceeding with the installation on a Windows machine you can also install the KNIME R Statistics Integration (Windows Binaries) which contains the platform specific binaries of the R integration to KNIME Analytics Platform. By doing this you can skip all the next installation and configuration steps described in this guide.

3. Finally, configure the KNIME Interactive R Statistics Integration.

   In KNIME Analytics Platform go to *File* → *Preferences*. From the list on the left, select *R* under *KNIME* and the following dialog opens:

© 2022 KNIME AG. All rights reserved.



In *Path to R Home* enter the path to identify the location where R is installed and click *Apply* to make the changes effective. This path can be found from within `R` by typing `R.home()`.

© 2022 KNIME AG. All rights reserved.

R installation

The installation files as well as all the necessary information about R installation can be found on one of the mirror sites of the Comprehensive R Archive Network (CRAN). R installation can be completed safely using the default settings. The destination location where R will be installed will have to be specified in KNIME Analytics Platform and can be found from within `R` by typing `R.home()`. The 4.0.x versions of R are also supported.

- For R installation on Windows the default location is recommended (e.g. usually `C:\Program Files\R\R-<version>`).
- On macOS systems the R installation folder location is usually `/Library/Frameworks/R.framework/Versions/Current/Resources/`.
- On Linux, instead, the installation folder is usually located in `/usr/lib/R`.

We recommend using R via the command-line interface and to avoid using `RStudio` or other tools as this might lead to unexpected behavior.

R packages installation

`Rserve` is an R package that allows other applications to talk to R using TCP/IP or local sockets sending binary requests. It requires different installation steps and different packages depending on the operating system in use.

© 2022 KNIME AG. All rights reserved.

Windows

1. Proceed with the installation of `Rserve`.

   From within R run the command:

   ```
   install.packages('Rserve')
   ```

   This will install the version 1.7-3.1 of `Rserve`.

   If the library folder is not writable a warning message will appear, together with the request to use a personal library:

   ```
   Warning in install.packages("Rserve") :
     'lib = "C:/Program Files/R/R-3.6.3/library"' is not writable
   Would you like to use a personal library instead? (yes/No/cancel) yes
   ```

   answer "yes" and a personal library will be installed and used.

   > Please test the installation with KNIME Analytics Platform and in case a problem is encountered follow the instructions at point 2.

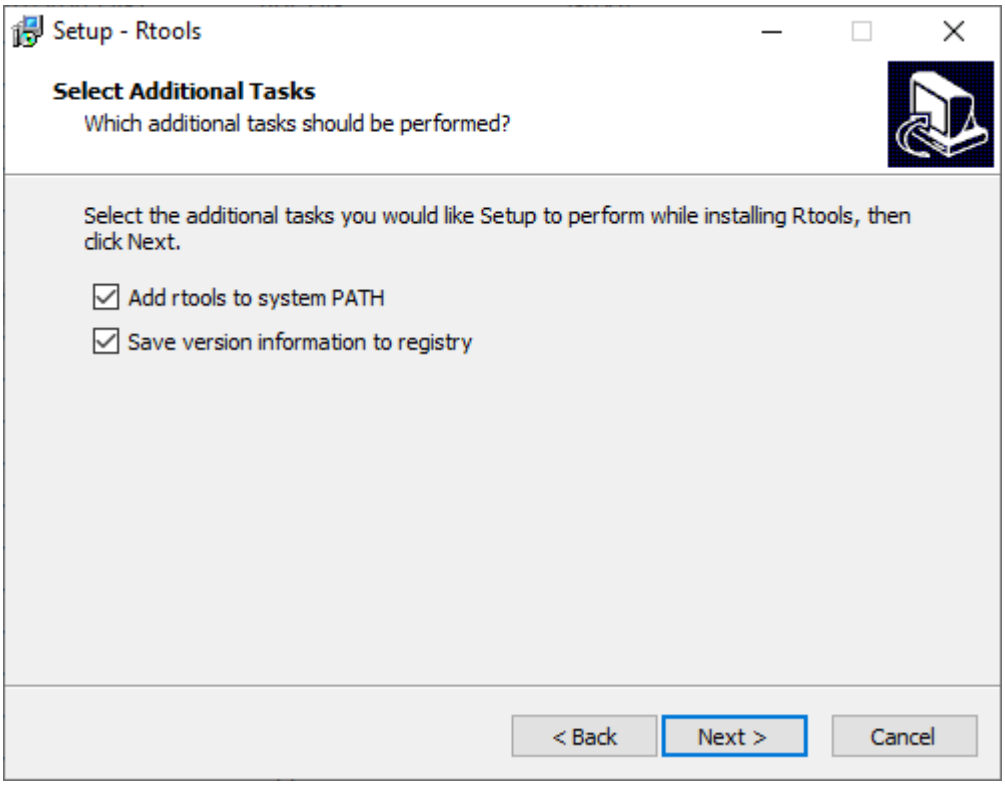
2. If the procedure at point 1 did not work properly, the installation of the latest version of `Rserve` might be necessary.

   a. First, from within `R`, remove any previously installed `Rserve` version running the command:

      ```
      remove.packages("Rserve")
      ```

   b. In this case, also `Rtools`, a collection of tools necessary for building R packages on Windows, needs to be installed. The installer is available here. In this guide `Rtools` is assumed to be installed under the default destination (e.g. usually `C:\Rtools\`).

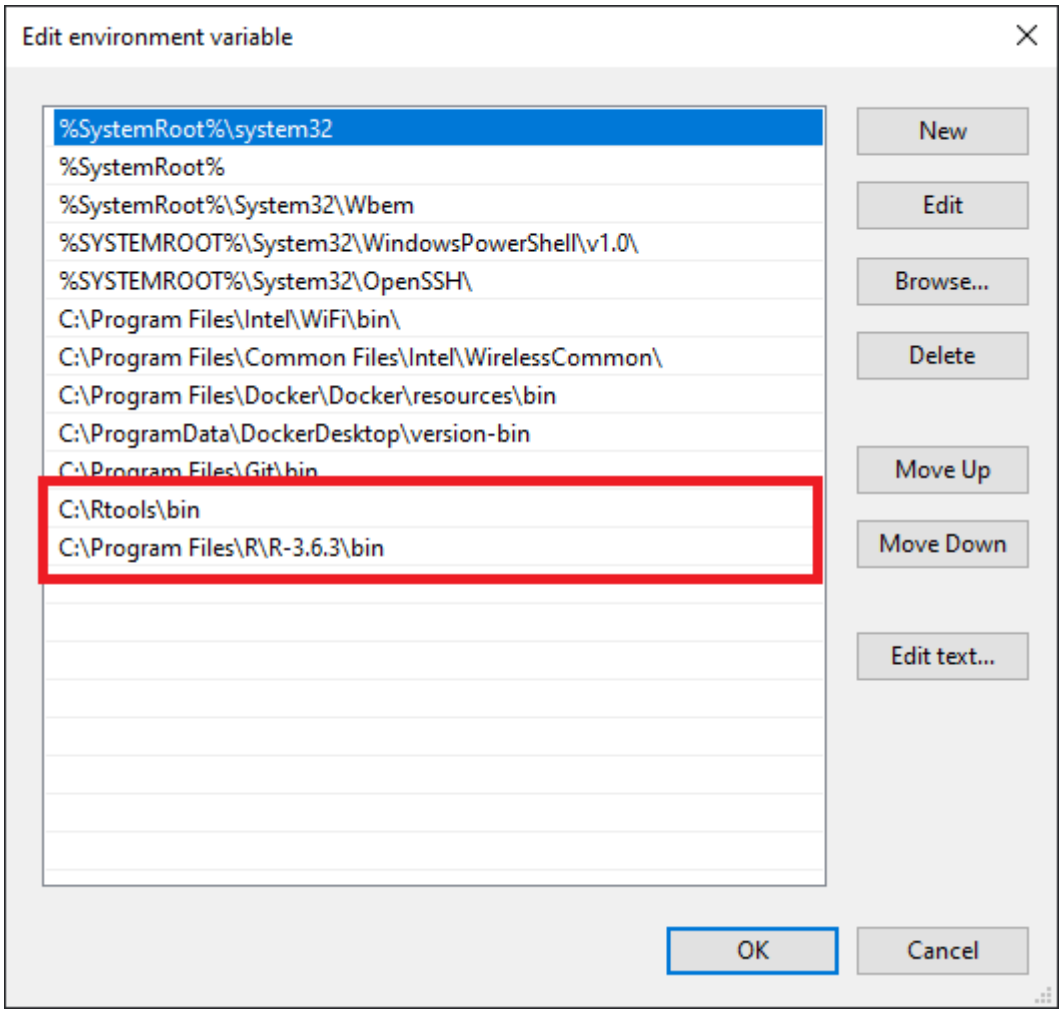
© 2022 KNIME AG. All rights reserved.



Check the box *Add rtools to system PATH* as shown in the above picture.

c. Please make sure that both `R` bin path and `Rtools` path are added to the Windows Path variable. From the *Start* menu open the *Control Panel* and click *System*. In the dialog that opens, click *Advanced system settings* in the left column. This opens the *System Properties* window. Here, click *Environment Variables…*. Then in the *System variables* section click *Edit…* to inspect and, if required, change the variable *Path*. Add the paths to `R` and `Rtools` bin folders, if they are not yet listed.

© 2022 KNIME AG. All rights reserved.



For the latest version Rtools40 the `bin` folder is located in `C:\Rtools40\usr\bin`.

d. Finally, download and install the latest version of `Rserve`. From within R type the command:

```
install.packages('Rserve',,"http://rforge.net/",type="source")
```

© 2022 KNIME AG. All rights reserved.

MacOS

1. First, `Clang` and `GNU Fortran` are needed. If they are not already installed, the installation packages can be downloaded from this link.
2. Another program that needs to be installed is `XQuartz`, which is available for download at this link.

   > A script is also available that can be used to install all of the necessary packages for R integration in KNIME Analytics Platform with macOS >= 10.14. After installing the KNIME Interactive R Statistics Integration the script can be found in `KNIME/plugins/org.knime.r_<knime_version>/scripts/` or at this link, only for KNIME Analytics Platform version 4.2 or higher. To allow for the execution of the script, please open a new *Terminal* and from the folder containing the script run `chmod u+x <name_of_the_script>.sh`.

3. Then the R package `Rserve` can be installed. Run `R` from *Terminal* and type the command:

   ```
   install.packages('Rserve',,"http://rforge.net/",type="source")
   ```

   The following error might appear when installing `Rserve`:

   ```
   *** Rserve requires R (shared or static) library.                      ***
   *** Please install R library or compile R with either --enable-R-shlib  ***
   ```

   If your macOS is < 10.14 (Mojave), open a new *Terminal* window and type the following command:

   ```
   xcode-select --install
   ```

   otherwise please make use of the previously mentioned script.

   Now continue with installing the `Rserve` package.

4. Finally, the R package `Cairo` needs to be installed. From within R type the command:

   ```
   install.packages('Cairo')
   ```

© 2022 KNIME AG. All rights reserved.

Linux Ubuntu

To install `Rserve`, run `R` from *Terminal* and type the command:

```
install.packages('Rserve',,"http://rforge.net/",type="source")
```

If the following error message appears:

```
/usr/bin/ld: cannot find -lssl
collect2: error: ld returned 1 exit status
```

open a new *Terminal* window, type the following command:

```
sudo apt install libssl-dev
```

and continue with installing `Rserve`.

© 2022 KNIME AG. All rights reserved.

KNIME Interactive R Statistics Integration usage

Once you install the KNIME Interactive R Statistics Integration the included nodes will be available in the node repository under *Scripting* > *R*.

Configure and export R environments

Besides setting up R Home for your entire KNIME workspace via the Preference page, you can also use the Conda Environment Propagation node to set up specific conda environments to propagate the environment to downstream R scripting nodes. This node also allows you to export the specific conda environment together with your workflows.

This node is also useful to make workflows that contain R scripting nodes more portable by allowing to recreate the conda environment used on the source machine (for example your personal computer) on the target machine (for example a KNIME Server instance).

Before proceeding you will need to install the KNIME Python Integration which contains the Conda Environment Propagation node. You can drag and drop the Conda Environment Propagation node or the KNIME Python Integration from KNIME Hub or install the extension by going to *File* > *Install KNIME Extensions…*, typing Python in the search field of the window that opens, selecting the KNIME Python Integration extension under KNIME & Extensions category, and clicking *Next* >.

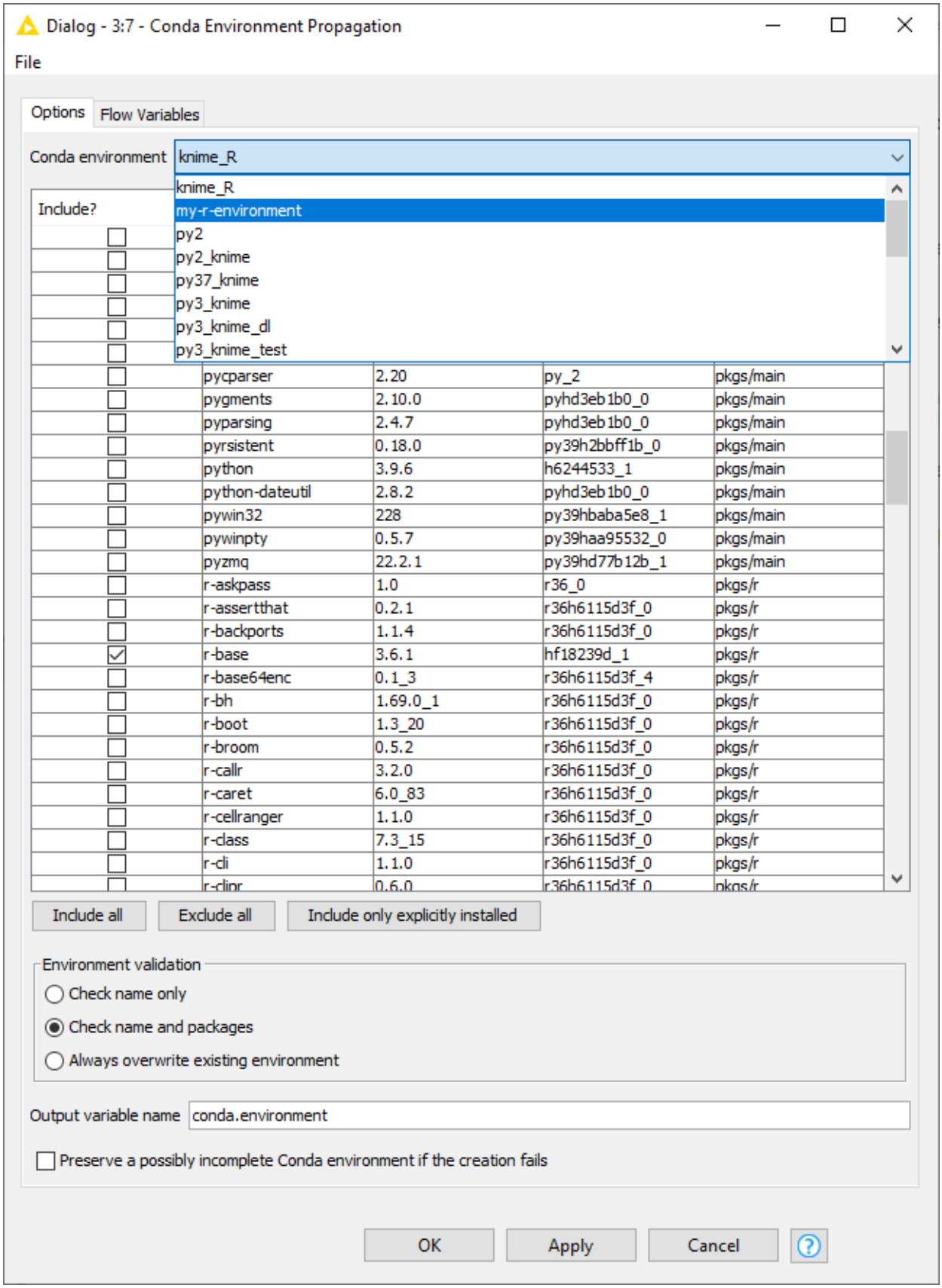
Configure the R environment with Conda Environment Propagation node

To configure the Conda Environment Propagation node follow these steps:

1. On your local machine, you need to have conda set up and configured in the *Preferences* of the KNIME Python Integration as described in the Anaconda Setup section of the KNIME Python Integration Guide
2. Create the desired conda R environment on your local machine. You will need to have installed at least the following three packages in order for KNIME Analytics Platform to be able to use R in the R scripting nodes:
    a. `r-base=3.6.1`
    b. `r-rserve=1.8_7`
    c. `r-essentials=3.6.0`

© 2022 KNIME AG. All rights reserved.

3. Open the node configuration dialog and select the conda environment you want to propagate and the packages to include in the environment in case it will be recreated on a different machine

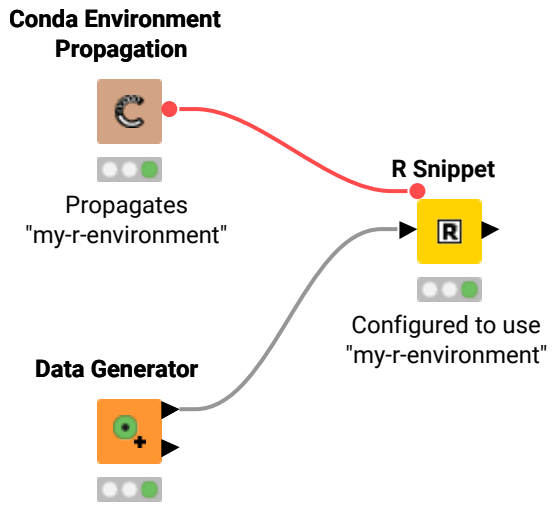


© 2022 KNIME AG. All rights reserved.

4. The Conda Environment Propagation node outputs a flow variable which contains the necessary information about the R environment (i.e. the name of the environment and the respective installed packages and versions). The flow variable has `conda.environment` as default name but you can specify a custom name. In this way you can avoid name collisions that may occur when employing multiple Conda Environment Propagation nodes in a single workflow.

In order for any R node in the workflow to use the environment you just created you need to:

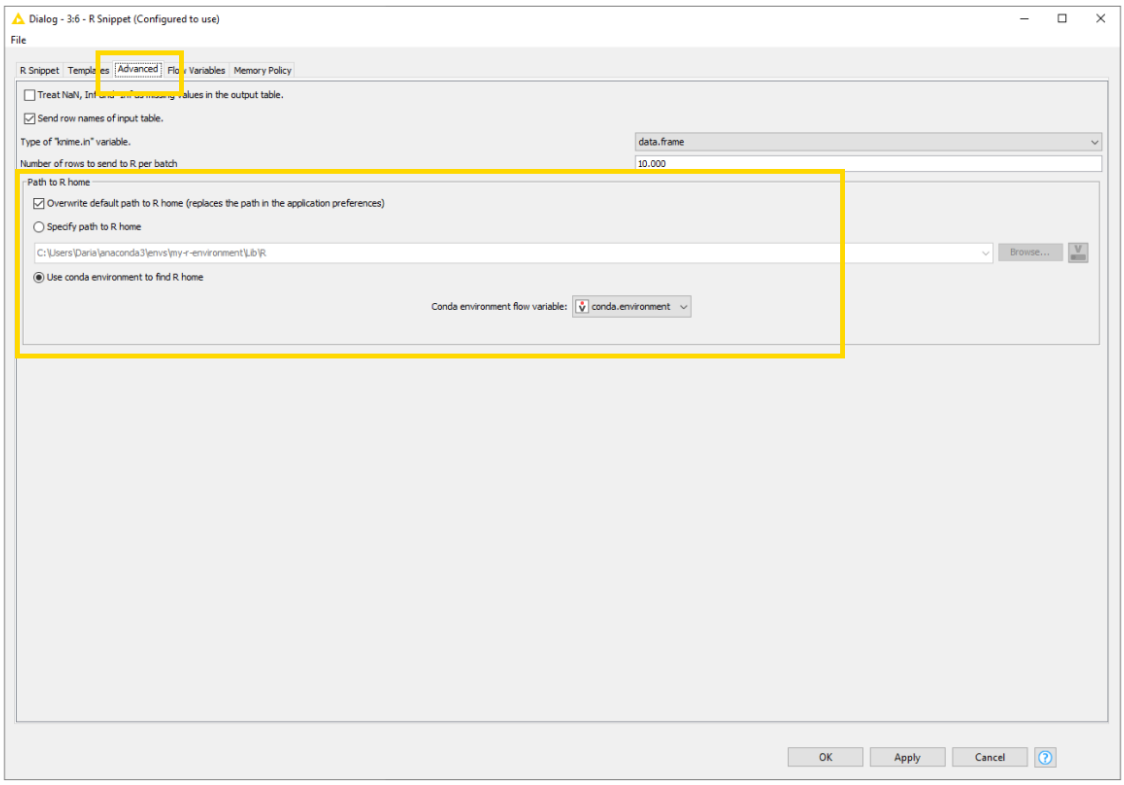
1. Connect the flow variable output port of Conda Environment Propagation node to the input flow variable port of the R node



Please note that, since flow variables are propagated also through connections that are not flow variable connections, the flow variable propagating the conda environment you created with the Conda Environment Propagation node will be available also for all the downstream nodes.

2. Successively open the configuration dialogue of the R nodes in the workflow that you want to make portable, go to the *Advanced* tab, and under *Path to R home* select *Overwrite default path to R home (replaces the path in the application preferences)*. This will activate the fields below where you are then able to select:
    a. *Specify path to R home*: You can specify a path to an R installation on your local machine different from the one you specified in the application preferences to be used by this specific node.
    b. *Use conda environment to find R home* and select the conda environment flow variable you want to use from the dropdown menu.

© 2022 KNIME AG. All rights reserved.



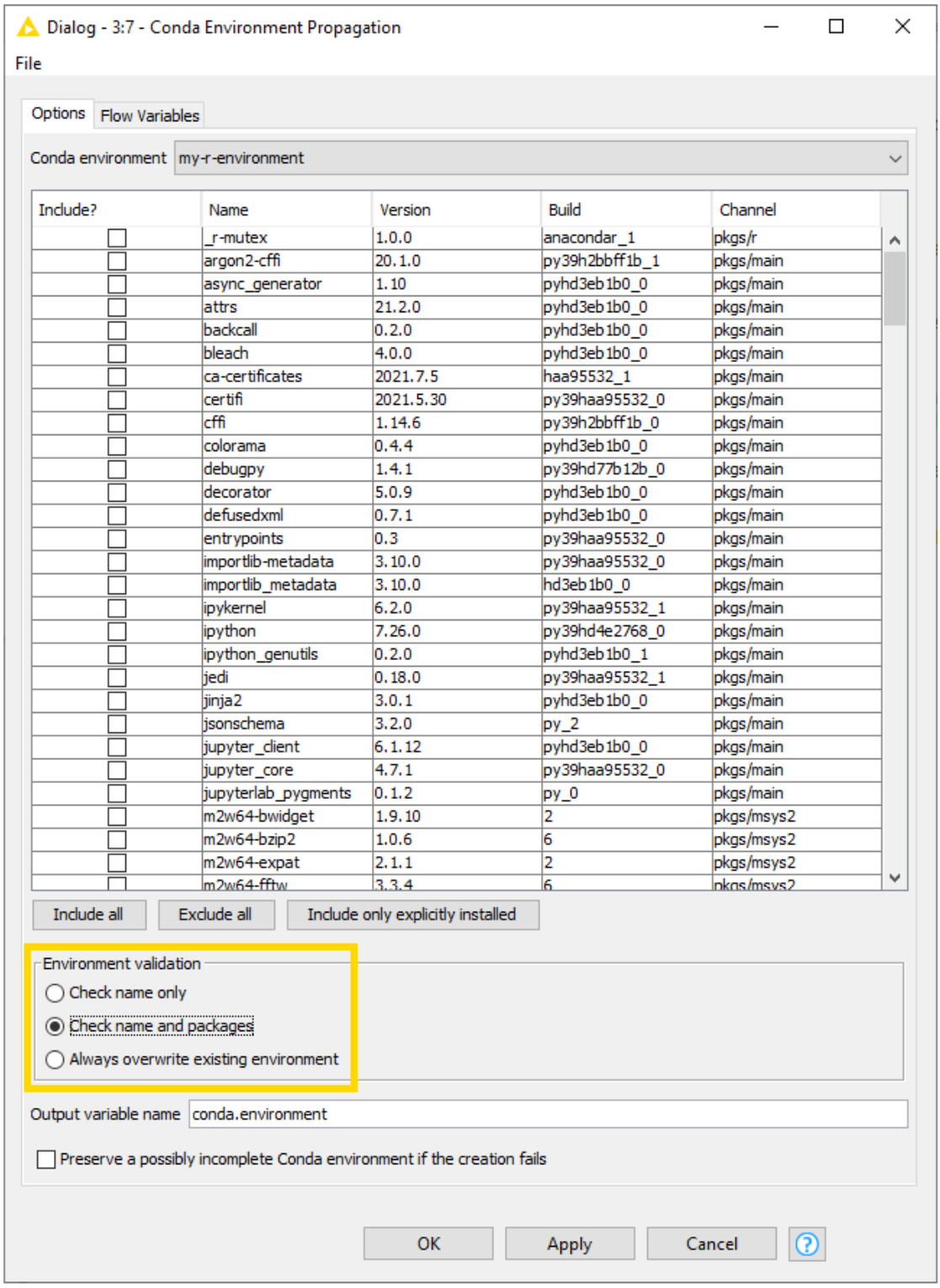
Export an R environment with a workflow

Once you configured the Conda Environment Propagation node and set up the desired workflow, you might want to run this workflow on a target machine, for example a KNIME Server instance.

1. Deploy the workflow by uploading it to the KNIME Server, sharing it via the KNIME Hub, or exporting it. Make sure that the Conda Environment Propagation node is reset before or during the deployment process.
2. On the target machine, conda must also be set up and configured in the Preferences of the KNIME Python Integration. If the target machine runs a KNIME Server, you may need to contact your server administrator and/or refer to the Server Administration Guide in order to do this.
3. During execution (on either machine), the node will check whether a local conda environment exists that matches its configured environment. When configuring the node you can choose which modality will be used for the conda environment validation on the target machine. *Check name only* will only check for the existence of an environment with the same name as the original one, *Check name and packages* will check both name and requested packages to correspond, while *Always overwrite existing environment* will disregard the existence of an equal environment on the target machine and will recreate it.

© 2022 KNIME AG. All rights reserved.

This option will affect the speed of execution of the node as conda requires an increasing amount of time if the check of the environment is based only on the name of the environment, or if a packages checks is also requested.



© 2022 KNIME AG. All rights reserved.

Please be aware that exporting R environments between systems that run different Operating Systems might cause some libraries to conflict.

© 2022 KNIME AG. All rights reserved.

KNIME AG
Talacker 50
8001 Zurich, Switzerland
www.knime.com
info@knime.com

The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.