Learning Goals

In this activity, you will:

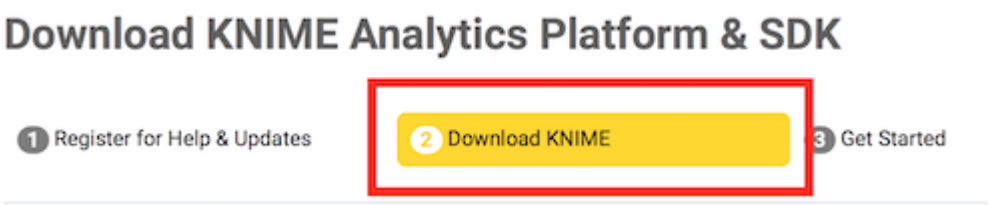
- Download and Install KNIME.
- Set up a KNIME Workspace.
- Navigate the KNIME Workbench.
- Use Nodes in a Workflow

Download and Install KNIME

Please use the following instructions to download and install the KNIME Analytics Platform before proceeding to The KNIME Workbench section.

NOTE: Download and install KNIME on your computer and not on the Cloudera VM.

1. **Download KNIME:** Go to https://www.knime.org/downloads/overview?quicktabs_knimed=1#quicktabs-knimed

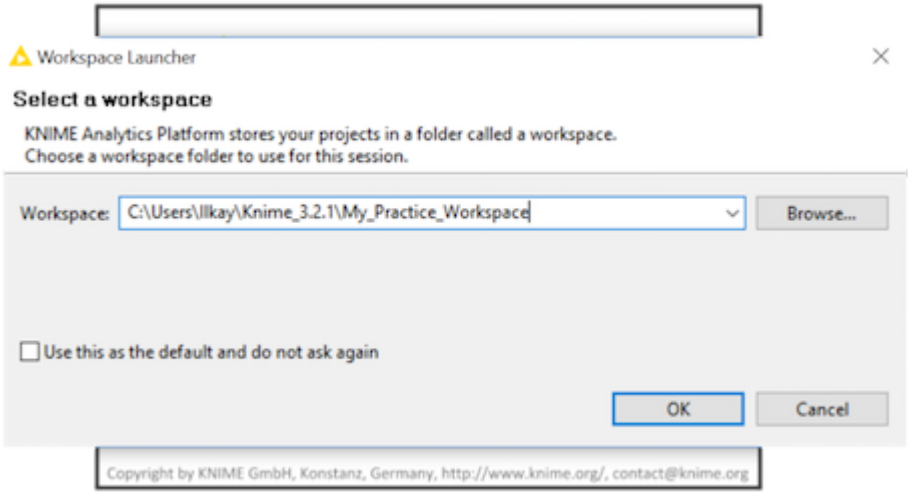


- **Windows:** Select 32-bit or 64-bit for *"KNIME Analytics Platform + all free extensions for Windows (installer)"*
- **Mac:** Select 64-bit for *"KNIME Analytics Platform + all free extensions for Mac OSX (10.7 and above)"*
- Accept the Terms and Conditions checkbox and click *Download*.

2. **Install KNIME** once the download is complete:

- **Windows:** Double click the KNIME installer exe file and complete the install wizard
- **Mac:** Double click the KNIME dmg file then drag the KNIME app file to the Applications folder

3. **Launch KNIME** (from Start Menu on Windows, from Applications folder on Mac) and it will prompt you to select a Workspace. Choose an appropriate location on your hard drive and click *OK*. This is the folder where your Workflow files will be stored.



This tutorial will use the Example Workflow provided in the KNIME installation. Recent versions of this example workflow may not use the Iris dataset used in this tutorial. Here is the dataset if you are unable to locate it in your KNIME installation.

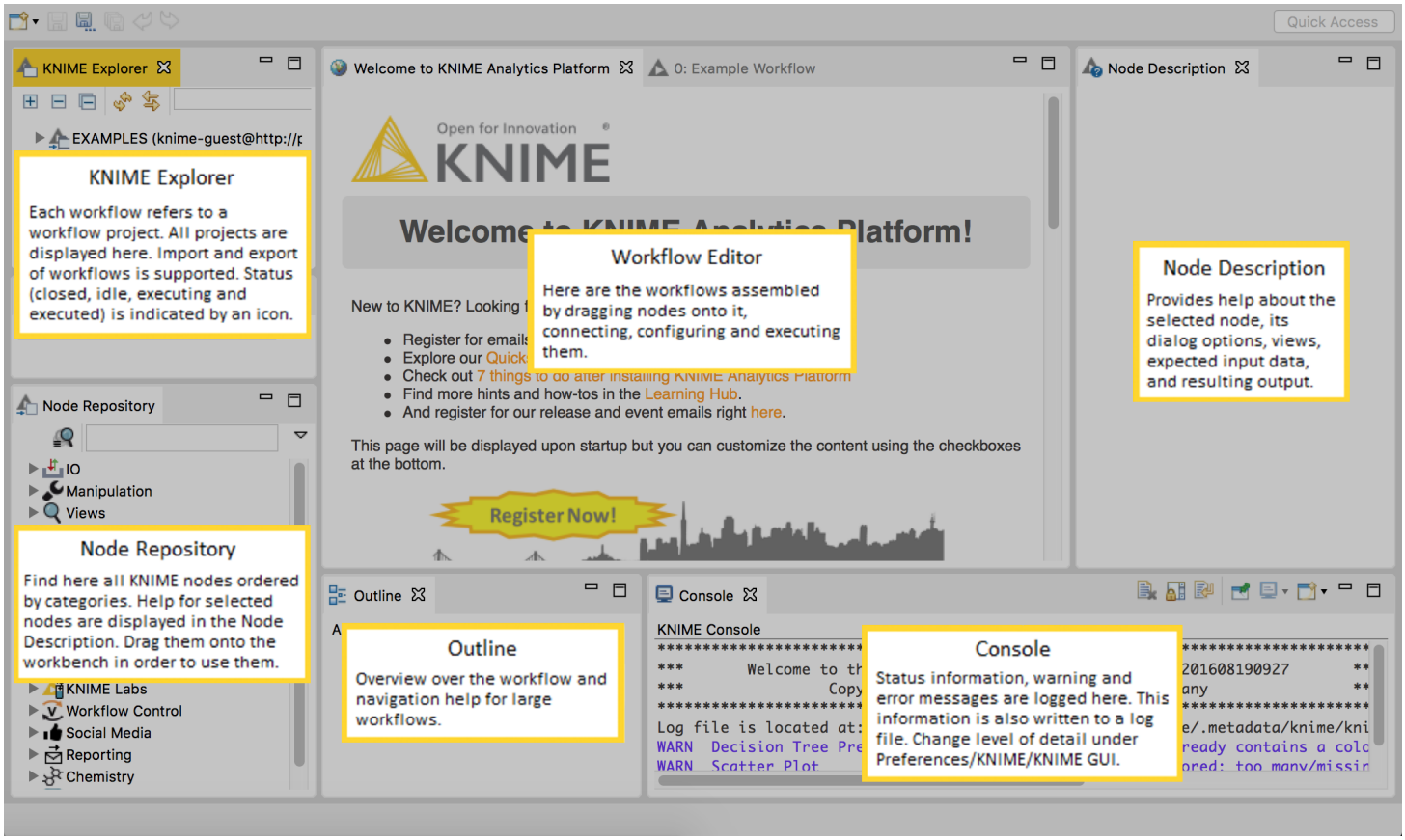
iris.csv

If you need the complete KNIME Example Workflow, download it here:

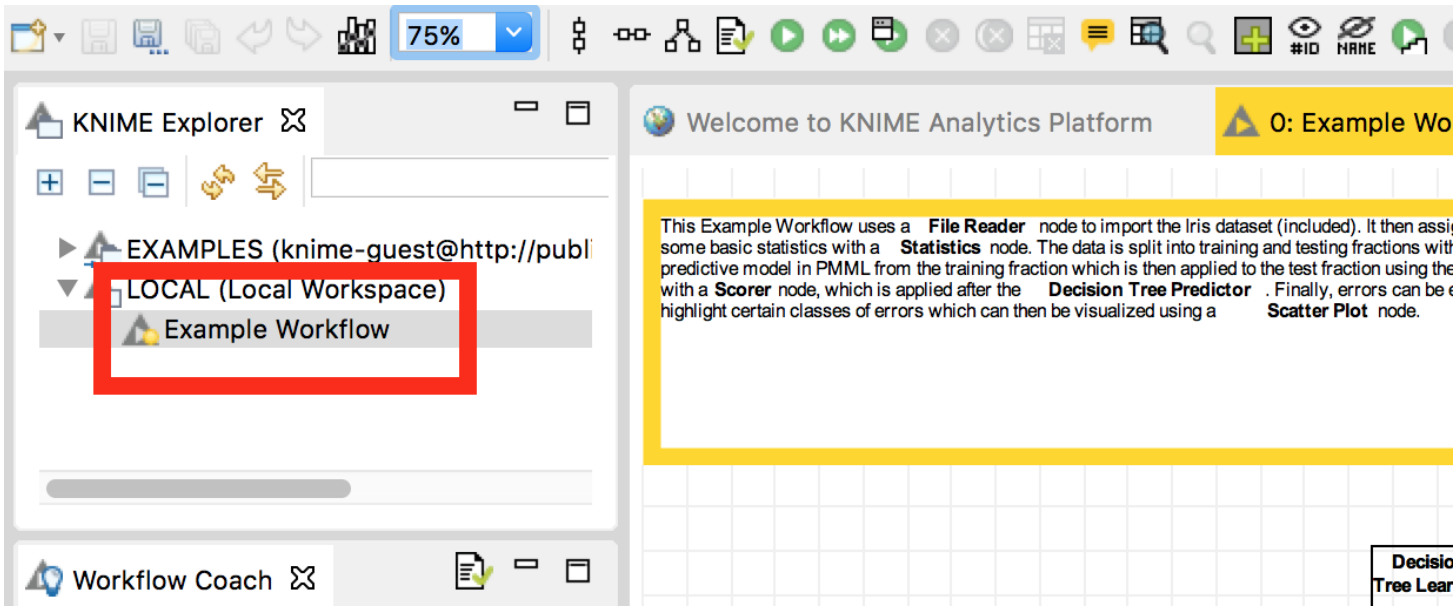
Example Workflow.zip

The KNIME Workbench

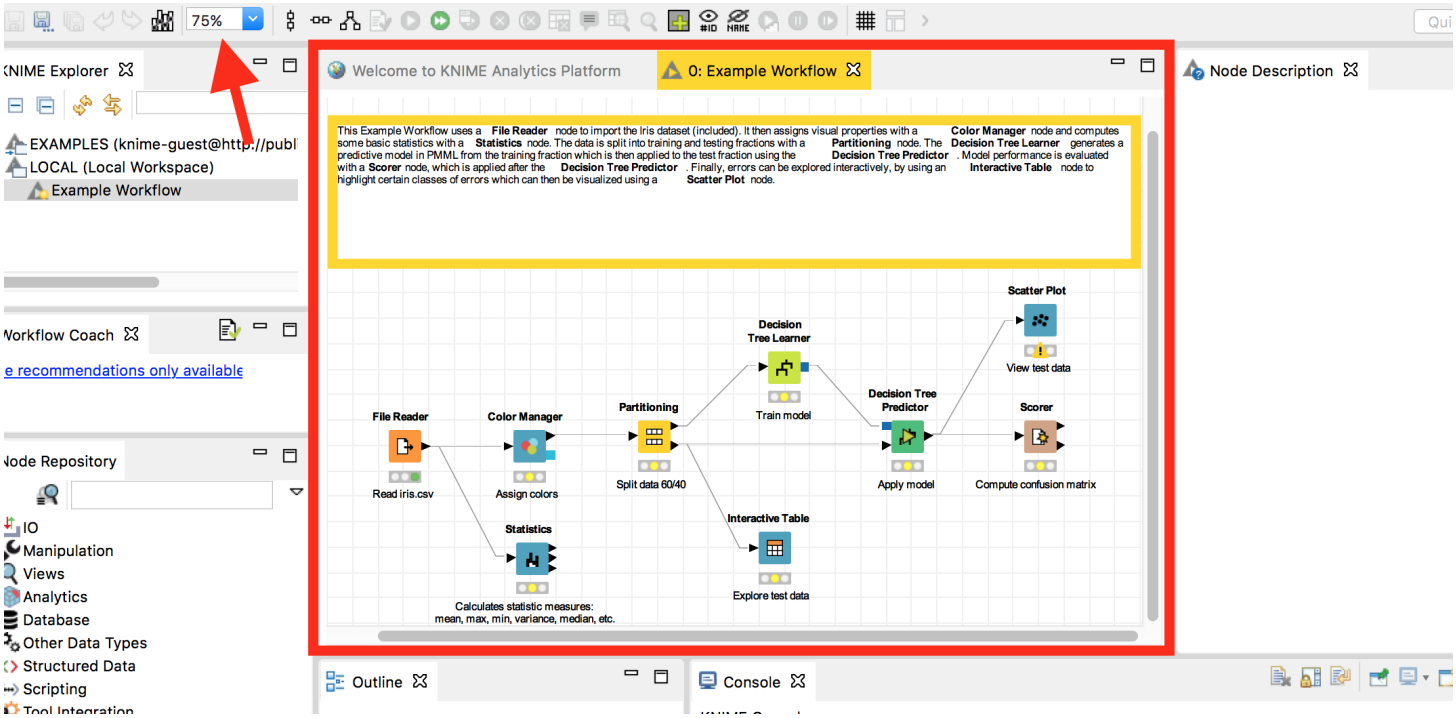
This is the Workbench. It is made up of different sections. Let’s see what each one is used for.



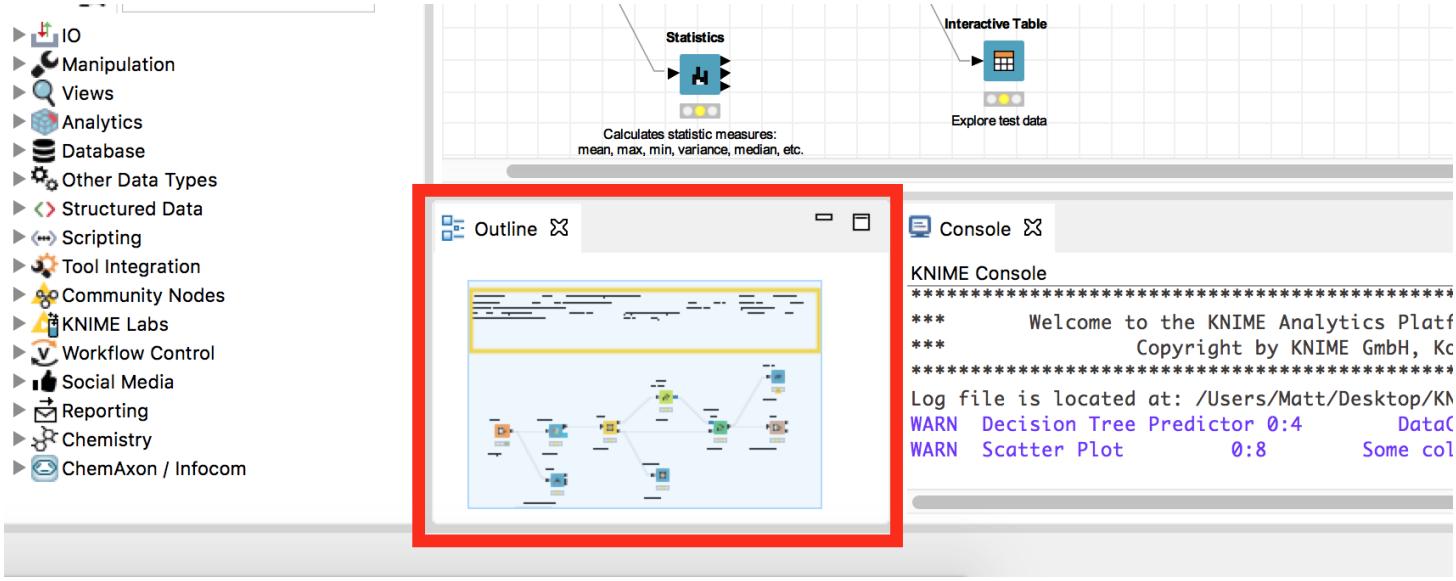
1. **KNIME Explorer:** This is where you can access and organize your saved workflow projects. They can be Created or Imported here via right-click, or Exported for sharing by right-clicking the project you wish to share. Open the Example Workflow in the KNIME Explorer by double-clicking it.



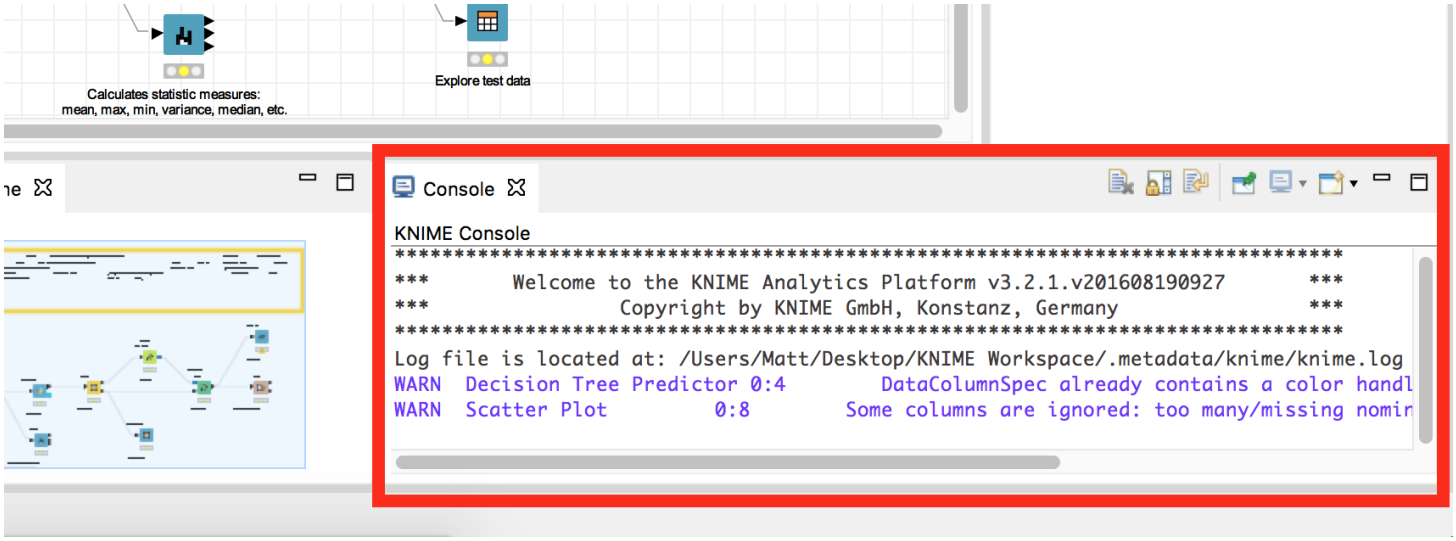
2. **Workflow Editor:** This is the main section of the window. It is where you drag new nodes, make connections between them, configure and execute them to perform operations and analysis on your data. Adjust the ZOOM using the drop-down menu indicated by the arrow.



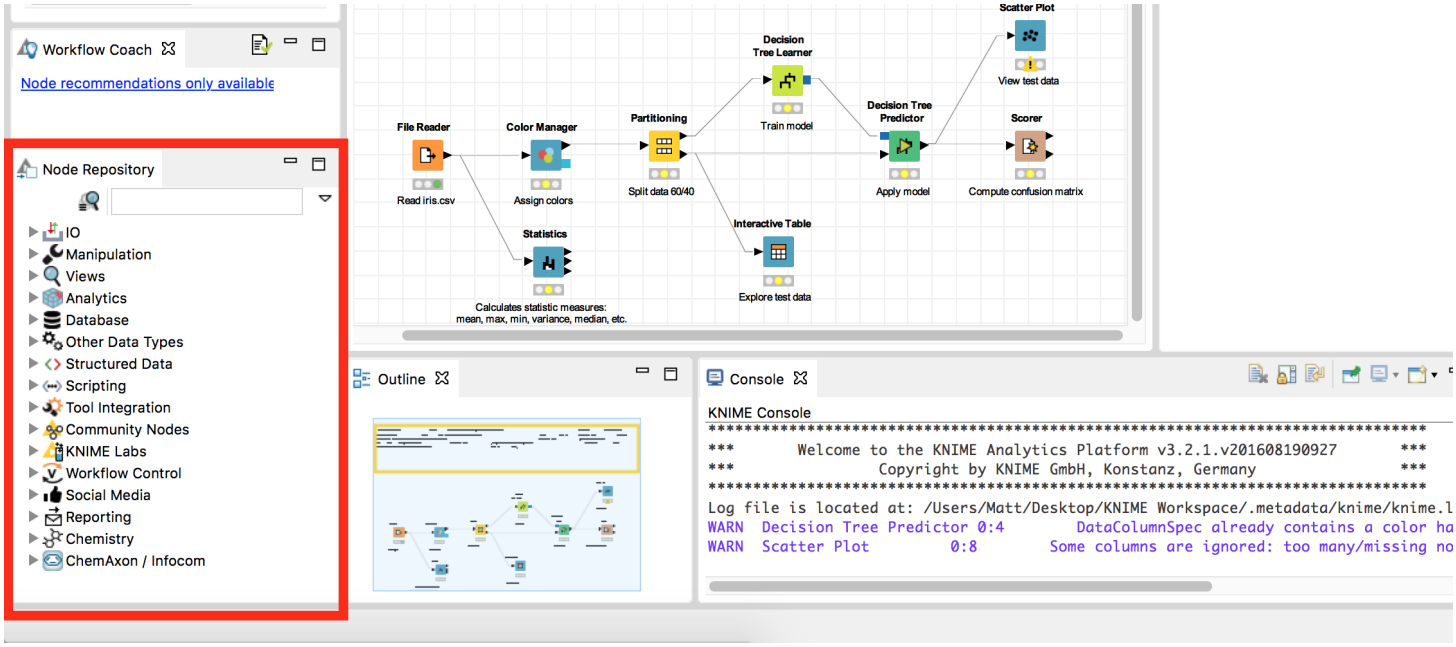
3. **Outline:** This is a wide overview of your workflow. You can click-and-drag the blue square to easily navigate a large workflow.



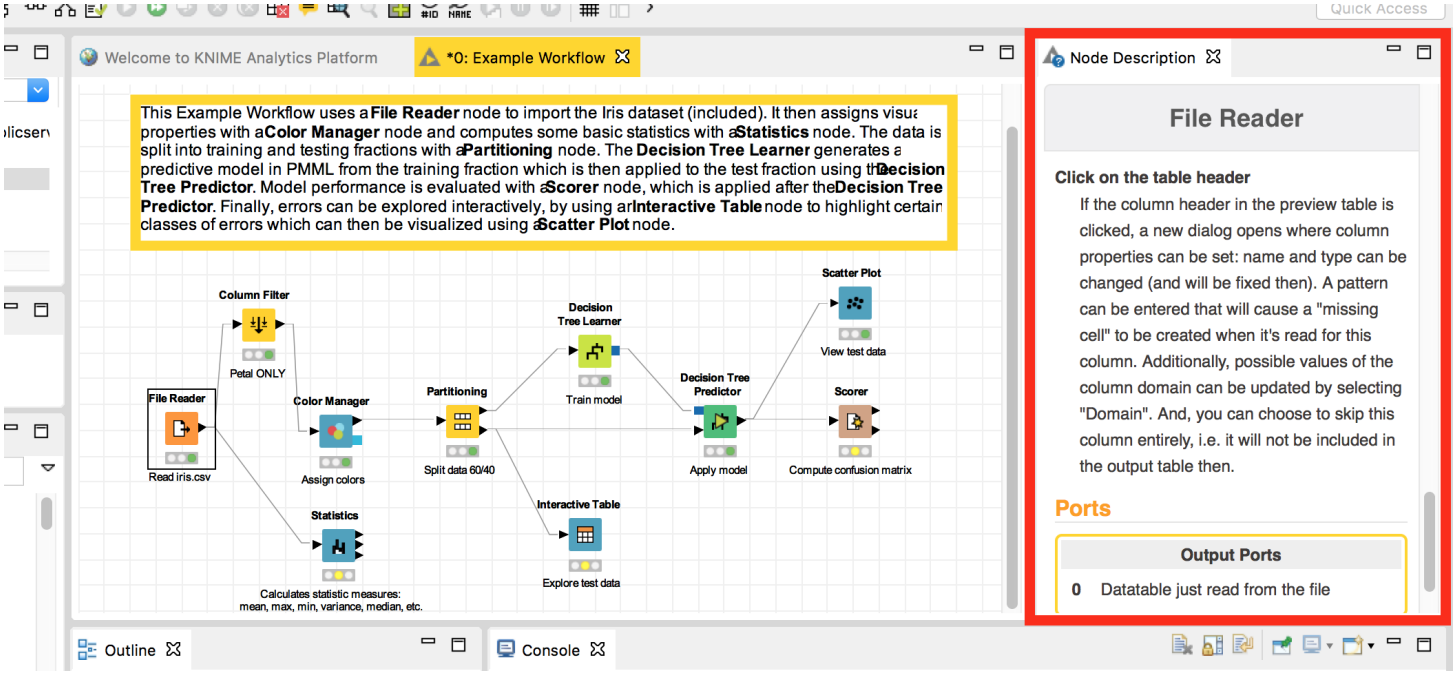
4. **Console:** This logs relevant information about the Status, Warnings, or Errors occurring with your nodes and data so you can fix possible issues. You can change the level of detail under Preferences>KNIME>KNIME GUI.



5. **Node Repository:** Find all the KNIME nodes organized into categories. Search for a particular node by name using the search box. To use a node, click-and-drag it onto the Workflow Editor. Just above it is the Workflow Coach, which is optional. It can suggest the next most likely node for your workflow based on community recommendations.



6. **Node Description:** When a node is selected, either in the Node Repository or the Workflow Editor, this will display the documentation about the given node, along with descriptions of each port on the node.

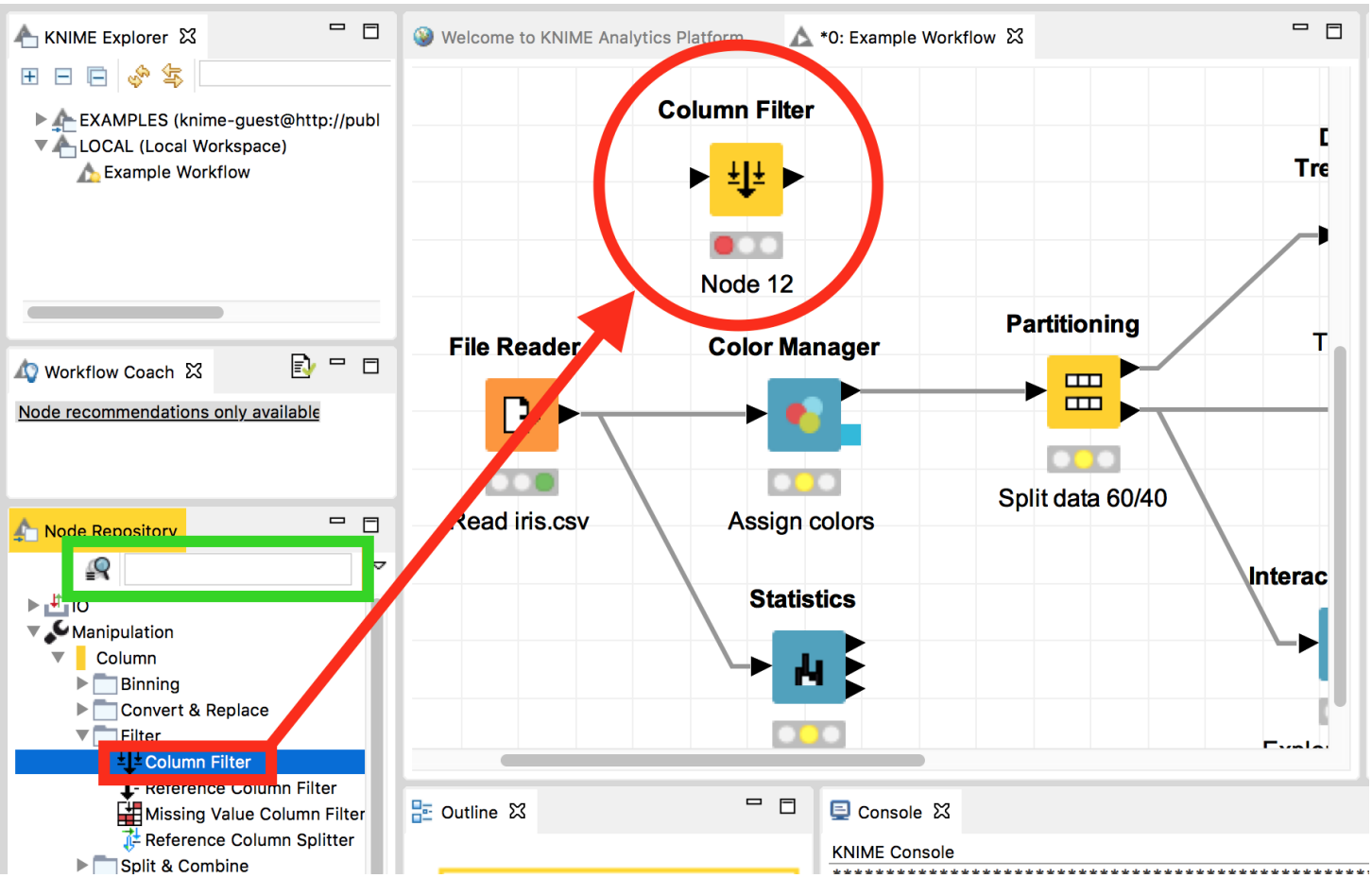


Using Nodes

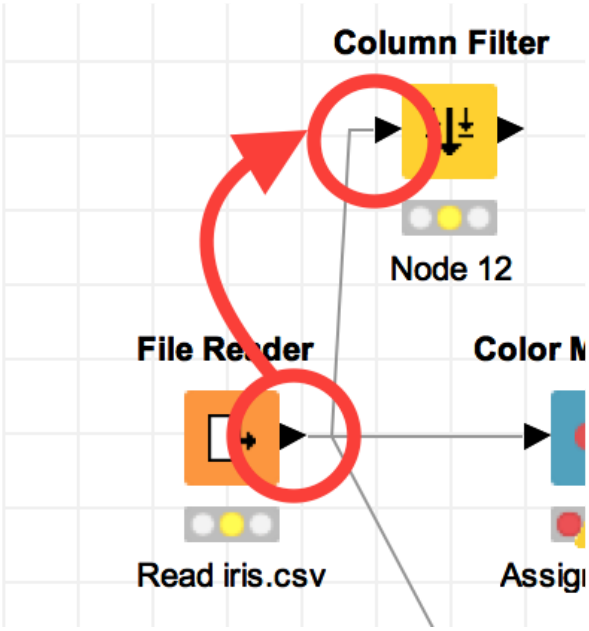
We are going to do a series of steps that will show you the fundamental user operations when working with nodes in KNIME. We are using the Example Workflow project to do this. There is a description of what this workflow does written in the editor.

> This Example Workflow uses a **File Reader** node to import the Iris dataset (included). It then assigns visual properties with a **Color Manager** node and computes some basic statistics with a **Statistics** node. The data is split into training and testing fractions with a **Partitioning** node. The **Decision Tree Learner** generates a predictive model in PMML from the training fraction which is then applied to the test fraction using the **Decision Tree Predictor**. Model performance is evaluated with a **Scorer** node, which is applied after the **Decision Tree Predictor**. Finally, errors can be explored interactively, by using an **Interactive Table** node to highlight certain classes of errors which can then be visualized using a **Scatter Plot** node.

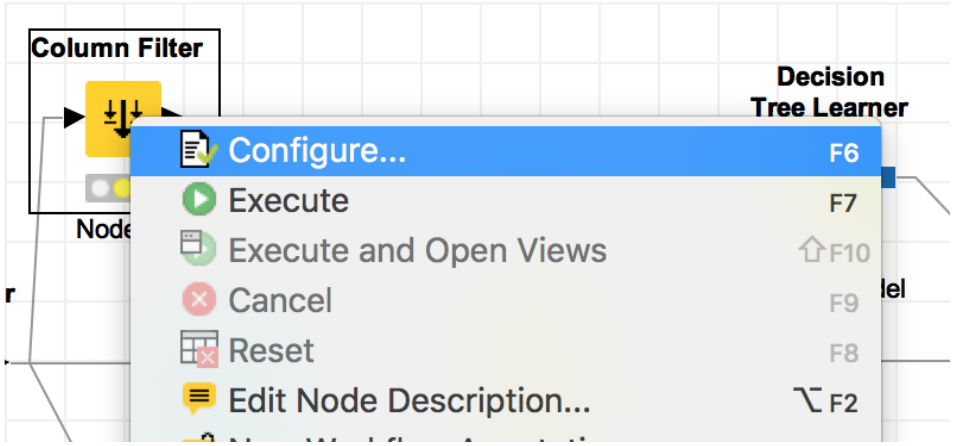
1. A node is a single processing unit in a workflow. First, let's create a new node. Make sure the Workflow Editor is currently showing the *Example Workflow*. In the **Node Repository**, choose category **Manipulation** > **Column** > **Filter**, and **drag** the Column Filter node onto the **Workflow Editor**. You can also find this node using the search bar highlighted by the green rectangle. You can search by name, such as "filter".



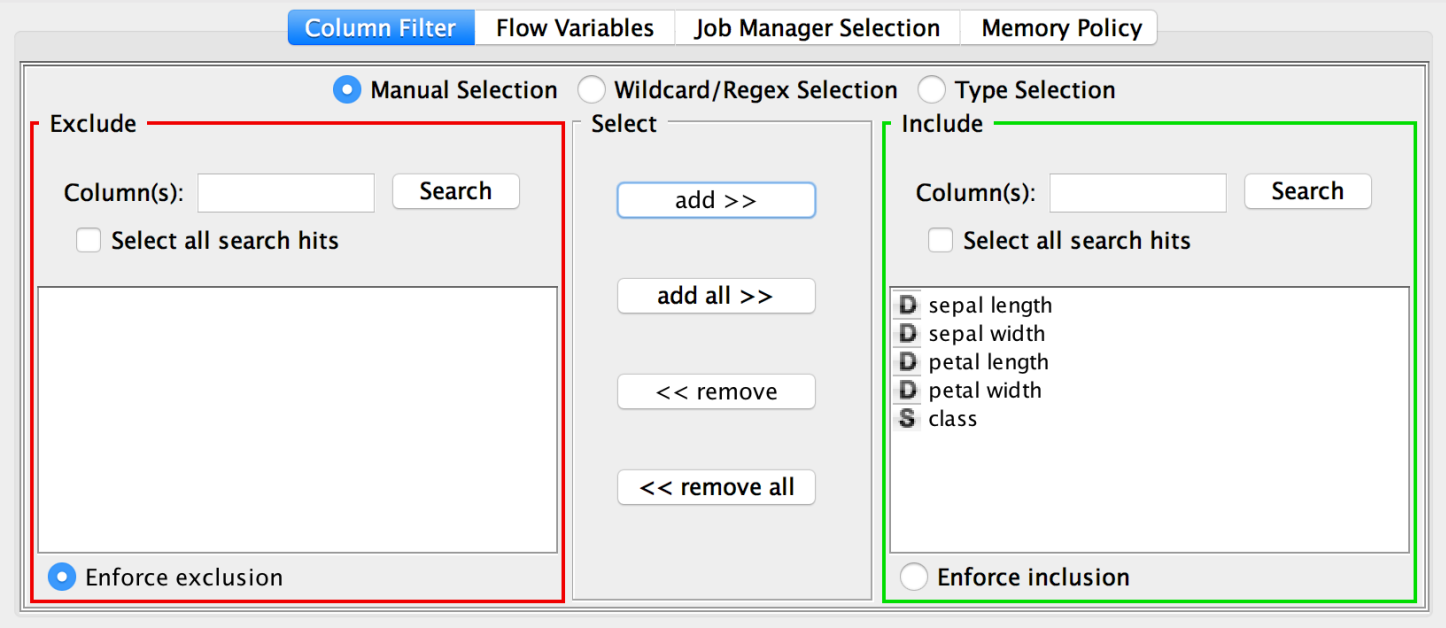
2. The File Reader node is what reads in a data set, and delivers it through the output port. To connect the output of the File Reader node to the input of the Column Filter node, **drag the mouse** cursor from the black arrow on the right side of the File Reader and drag it to the black arrow on the left side of the Column Filter.



3. Let's configure our new node. You can open the configure window for the Column Filter node by either **double-clicking** the icon, or **right-clicking** it and selecting "**Configure…**".

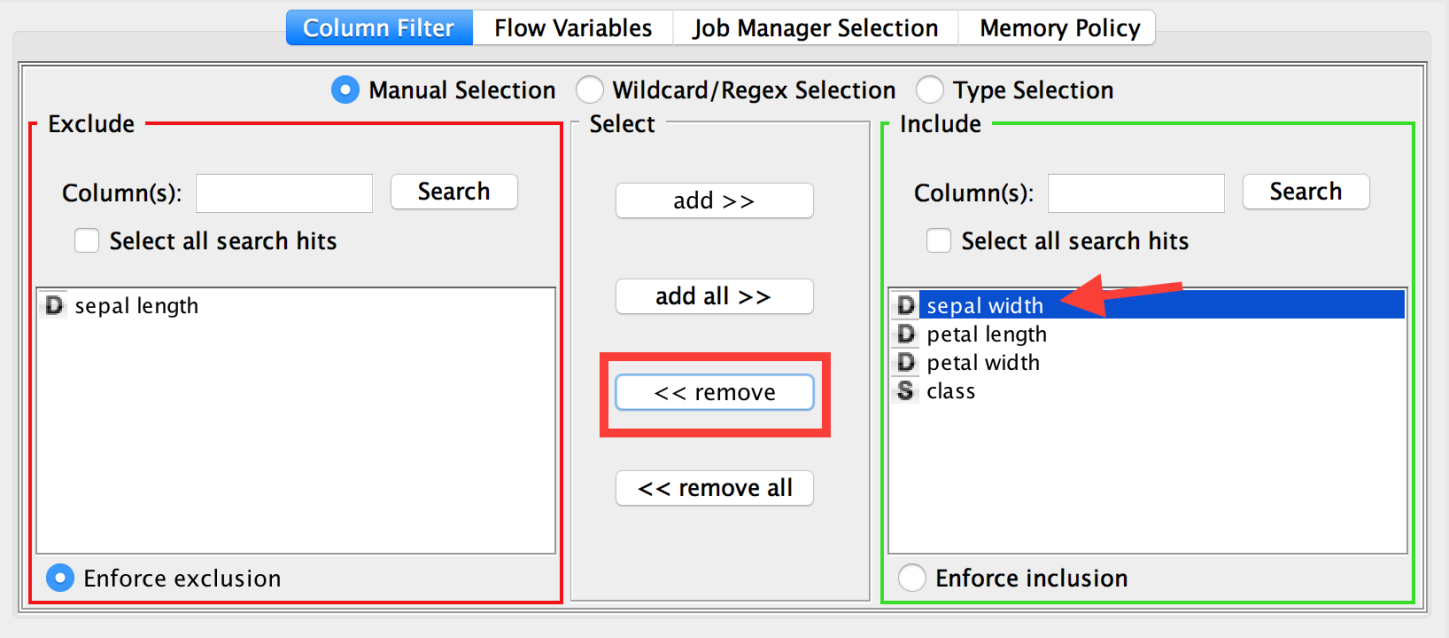


Every node-type has a different configure window, but Column Filter lets you select which column(s) of data you want to exclude from passing through to the pipeline to the next node.



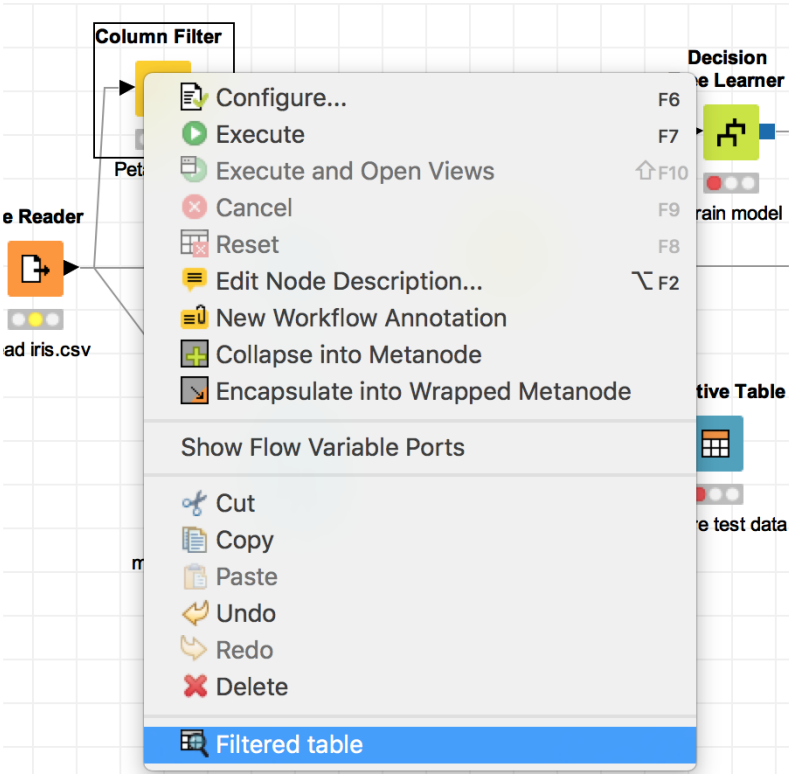
You can do this by clicking on a column name and selecting the “add >>” or “<< remove” button. To make it so only “petal width”, “petal length” and “class” are passed along the pipeline, do the following:

- select "sepal length", then select "<< remove"
- select "sepal width", then select "<< remove"

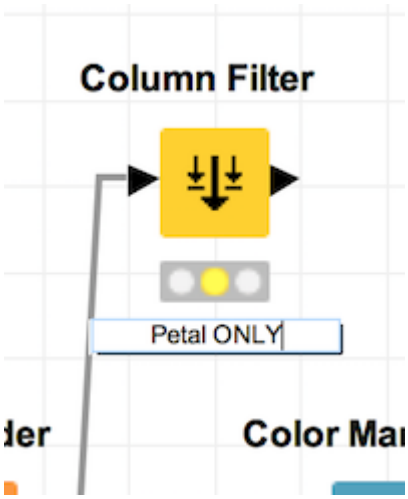


Then click OK at the bottom of the window. It will say “Warning, reset node(s)” every time you make changes to a node’s configuration. Click OK.

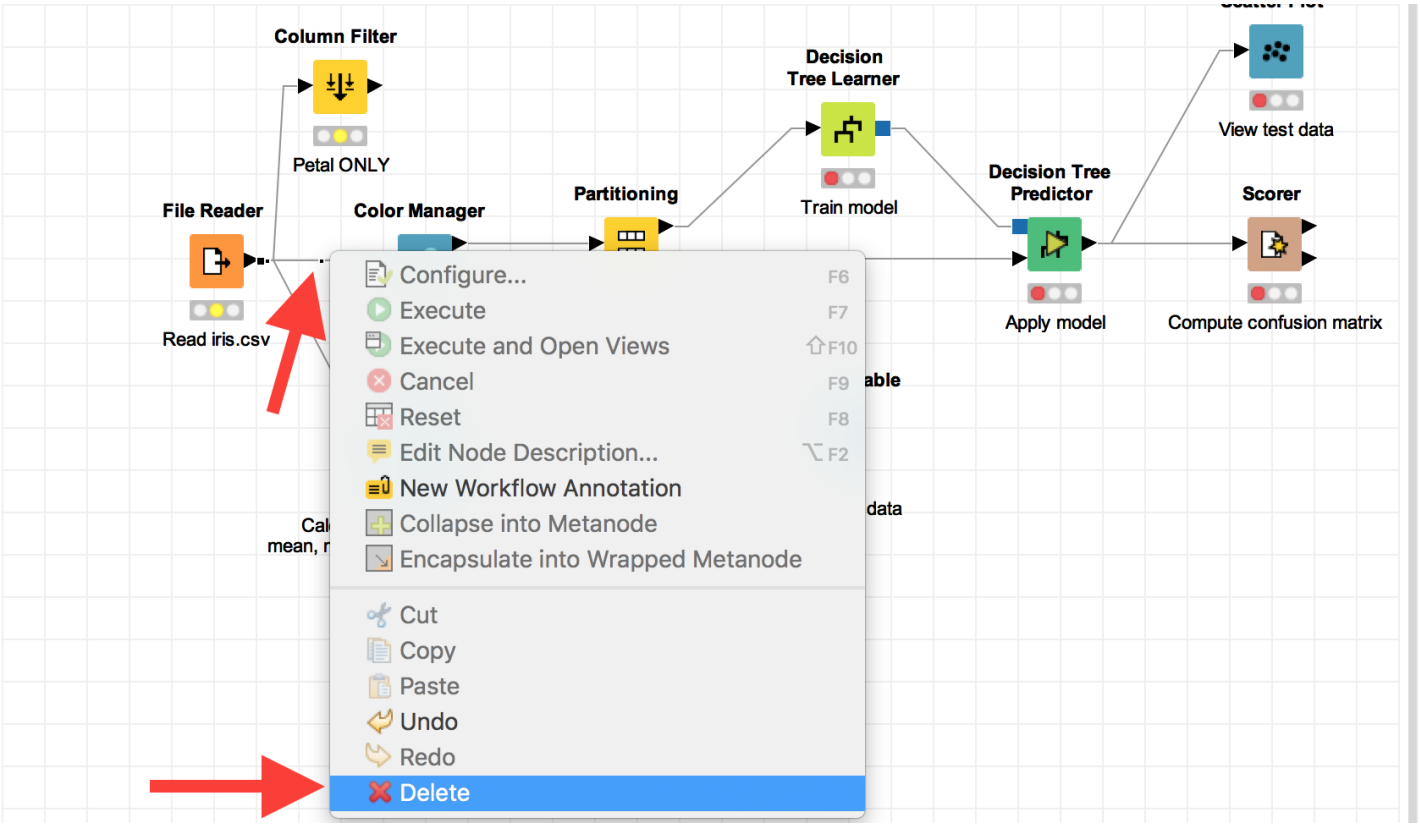
You can view the altered data by **right-clicking** the node and selecting the magnifying glass icon.



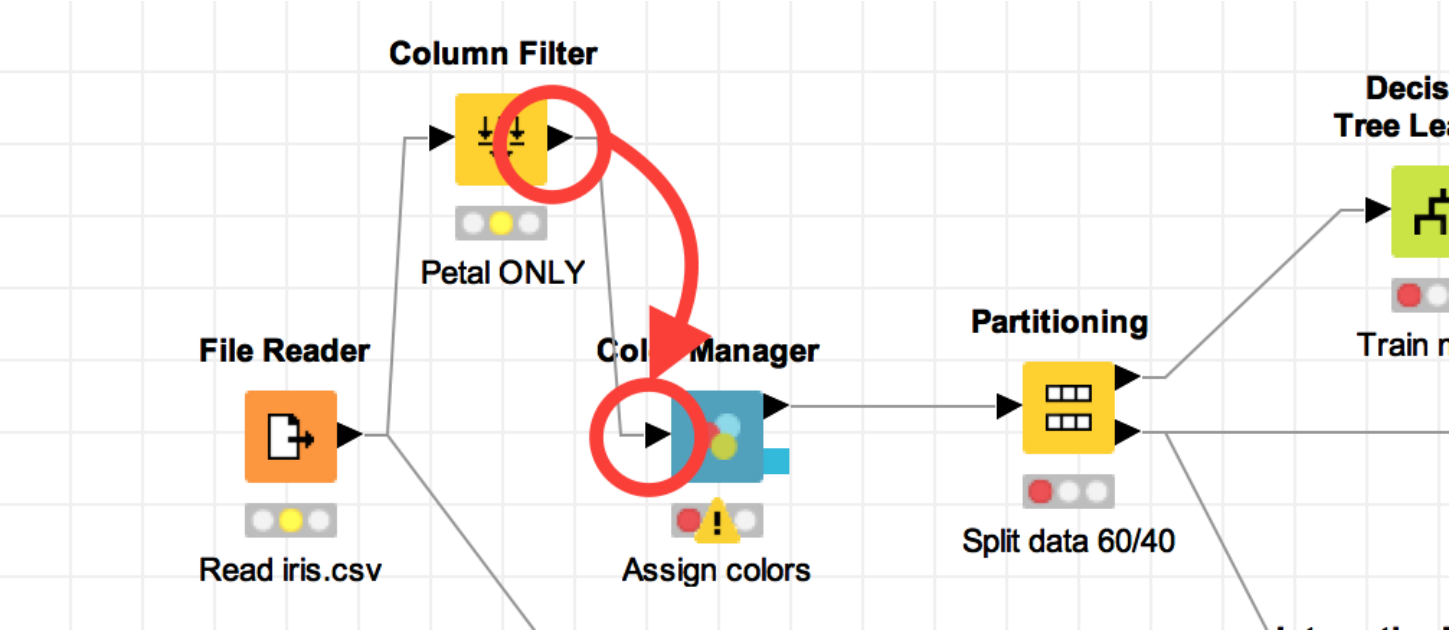
4. Let’s rename this node for easier reading of the pipeline. To do this, **double click the text** below the icon and start typing something like “Petal ONLY”, then click anywhere outside the text box.



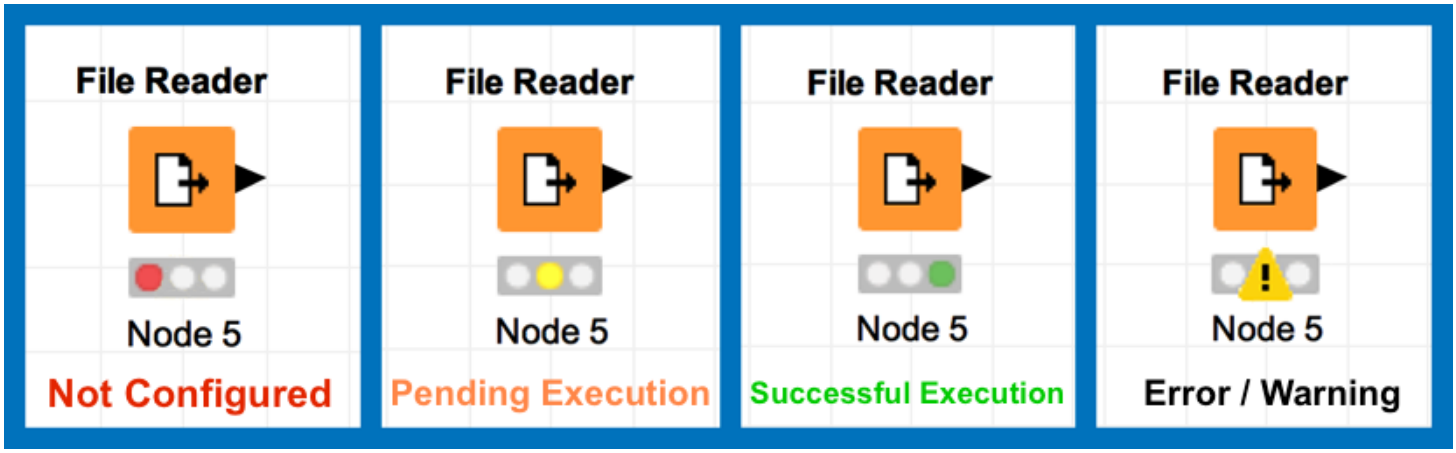
5. Now we are going to delete the connection currently leading to the Color Manager node and replace it with the output of the Column Filter node. Do this by **right-clicking** the connection and selecting **Delete**



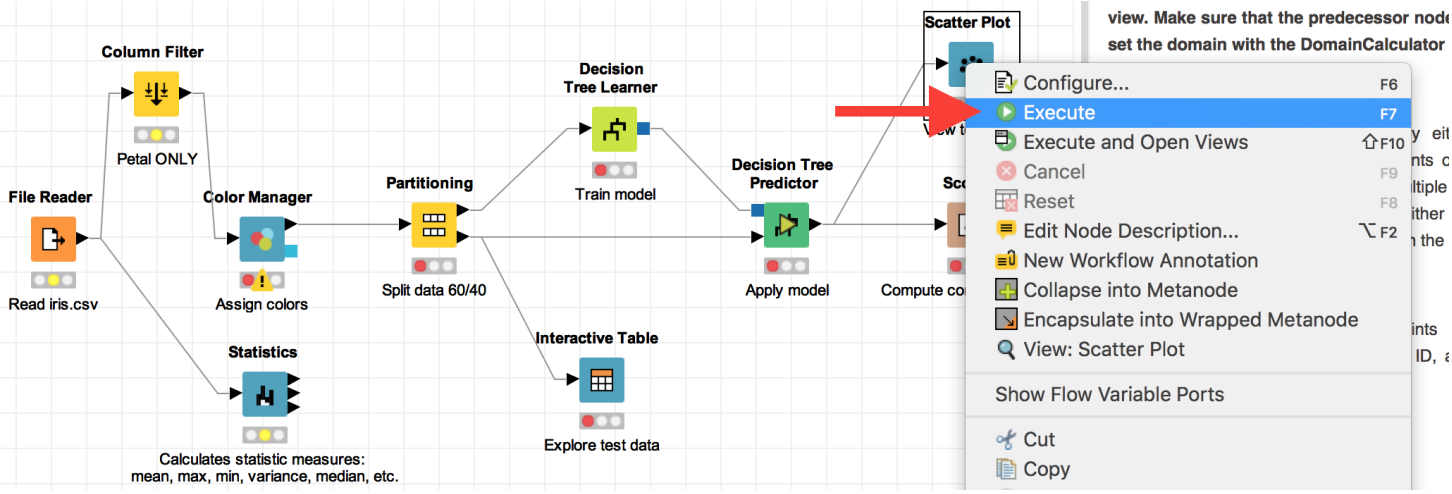
Then drag the mouse cursor from the output arrow on the right side of the Column Filter node to the input arrow of the Color Manager node. This just applied the filter node to the data all the way down the pipeline.



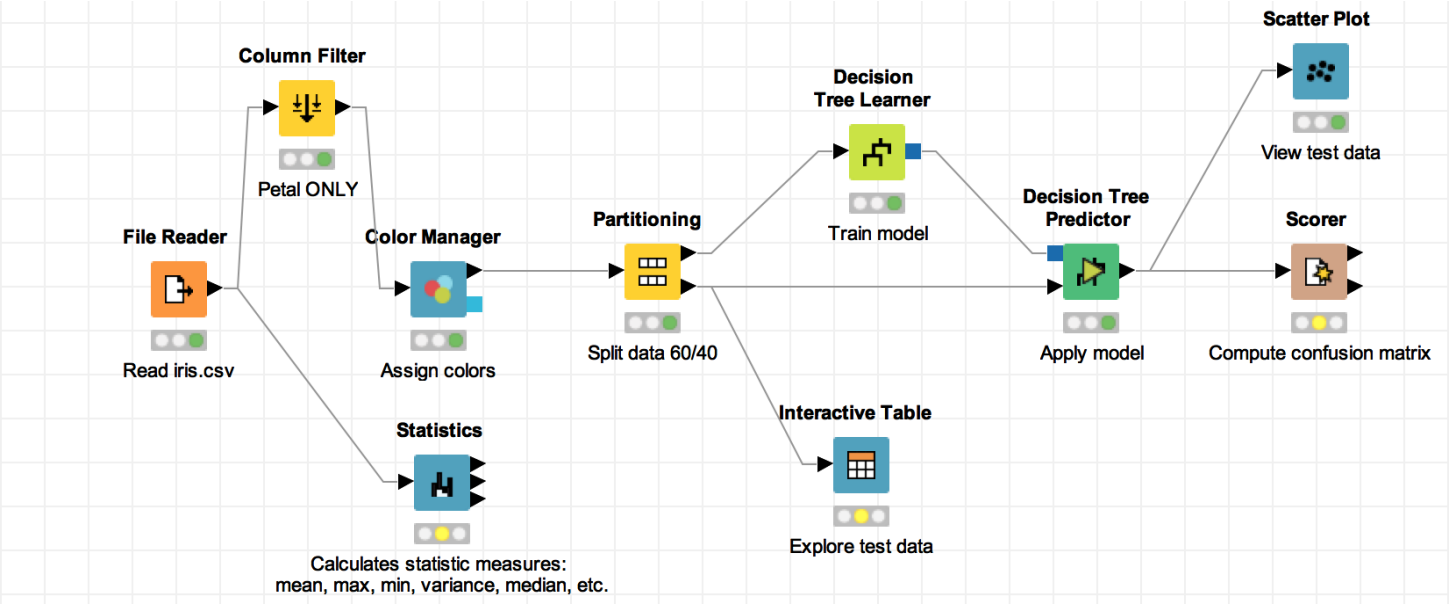
6. **Node States:** Notice the 3 circles below each node. They show the status of the node.



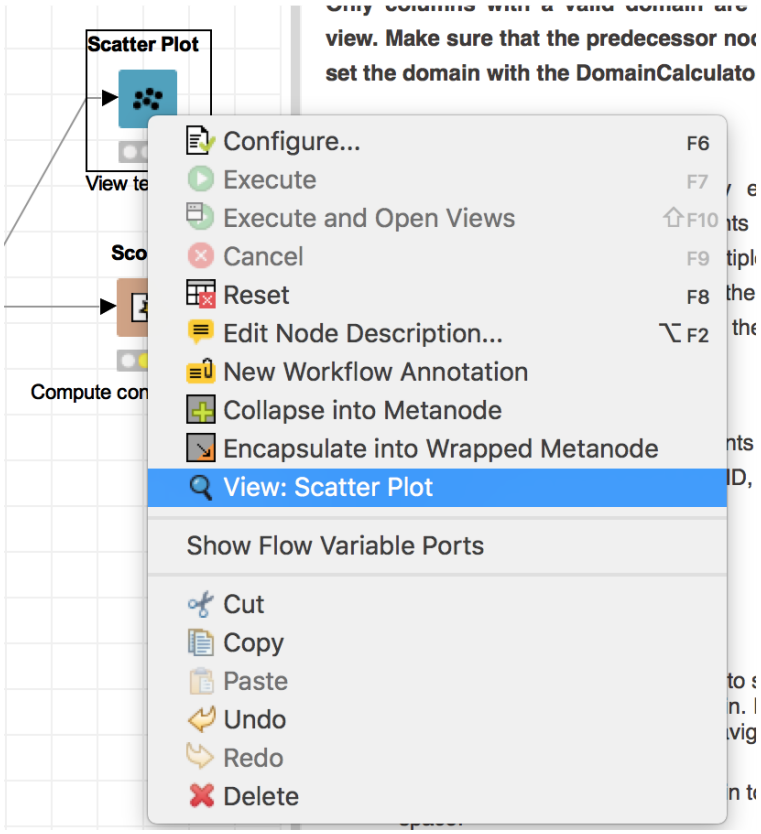
Go to the Scatter Plot node at the end of the workflow, right-click, select “Execute”.



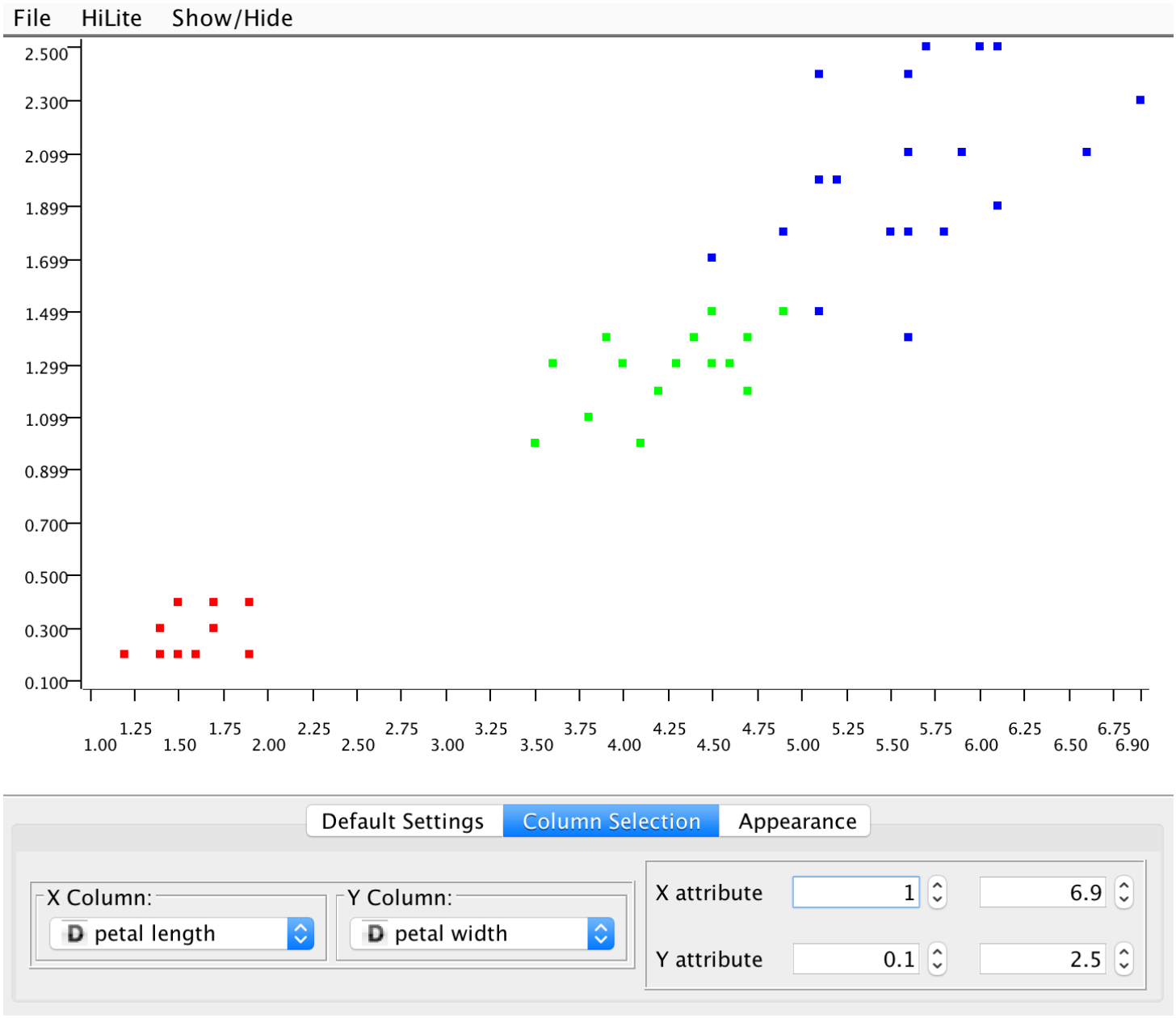
You should see all the nodes that Scatter Plot depends on turned green, meaning they all executed successfully.



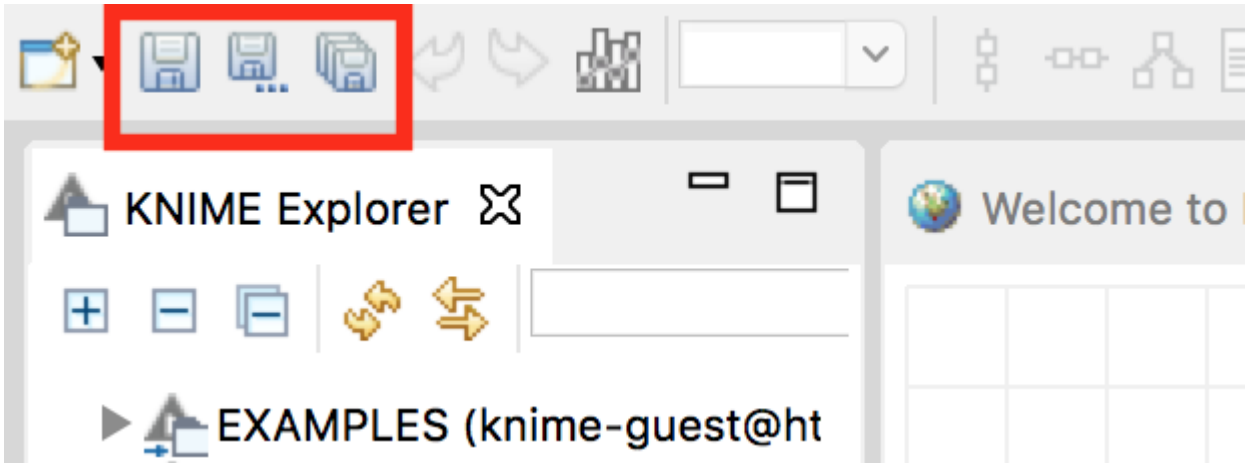
7. To view the data, right-click the Scatter Plot node and select “View: Scatter Plot”.



Select the “Column Selection” tab to see the axis settings for the plot.



8. To Save a workflow, either choose the Save, Save As, or Save All icon in the upper toolbar.



KNIME Resources

- KNIME Website
- KNIME Getting Started
- KNIME Online Self-Training
- KNIME Quickstart Guide
- KNIME Node Documentation
- KNIME Learning Hub
- KNIME Community
- KNIME Labs