



TAX MANAGEMENT  
ASSOCIATES, INC

# DATA ANALYTICS WITH KNIME

v.3.4.0



# WHO

# WE ARE

Tax Management Associates is a professional services firm that has served the interests of state and local government since 1979.

## QUALIFICATIONS & EXPERIENCE

- ▶ 38 years of providing professional services to state and local taxing officials
- ▶ TMA works exclusively with government partners
- ▶ TMA is composed of 150+ employees in five main offices across the United States
- ▶ Our main focus is on revenue enhancement services for state and local jurisdictions and property tax compliance efforts

# KNIME POWERED CUSTOM ANALYTICS

- ▶ TMA is a proud KNIME Trusted Consulting Partner.



**Visit:** [www.knime.org/knime-trusted-partners](http://www.knime.org/knime-trusted-partners)

- ▶ Successful analytics solutions:
  - Fraud Detection (Michigan Department of Treasury)
  - Entity Discovery (multiple counties)
  - Data Aggregation (Louisiana State Tax Commission)

# KNIME POWERED CUSTOM ANALYTICS

- ▶ KNIME is an open source data toolkit
- ▶ Active development community and core team
- ▶ GUI based with scripting integration
  - Easy adoption, integration, and training
- ▶ Data ingestion, transformation, analytics, and reporting

# FEATURES & TERMINOLOGY

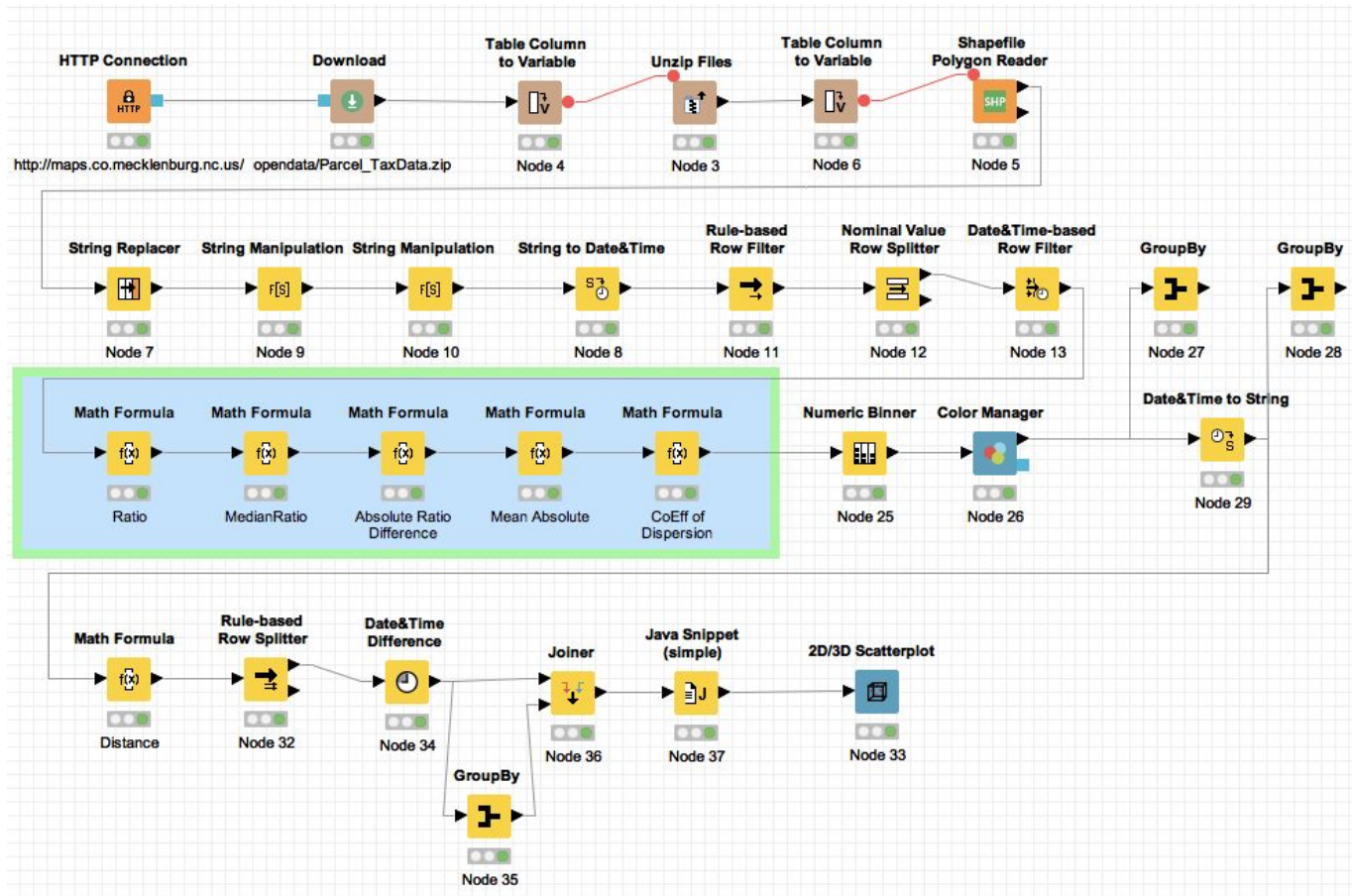
# KNIME WORKBENCH

The screenshot displays the KNIME Analytics Platform interface with a workflow titled "ReAppraisal" (Node 3). The workflow is composed of several interconnected nodes:

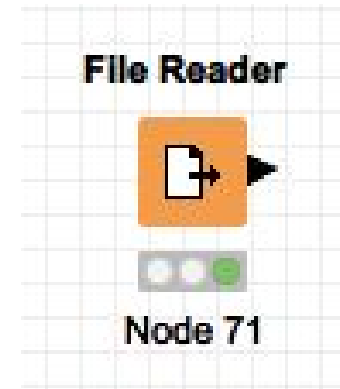
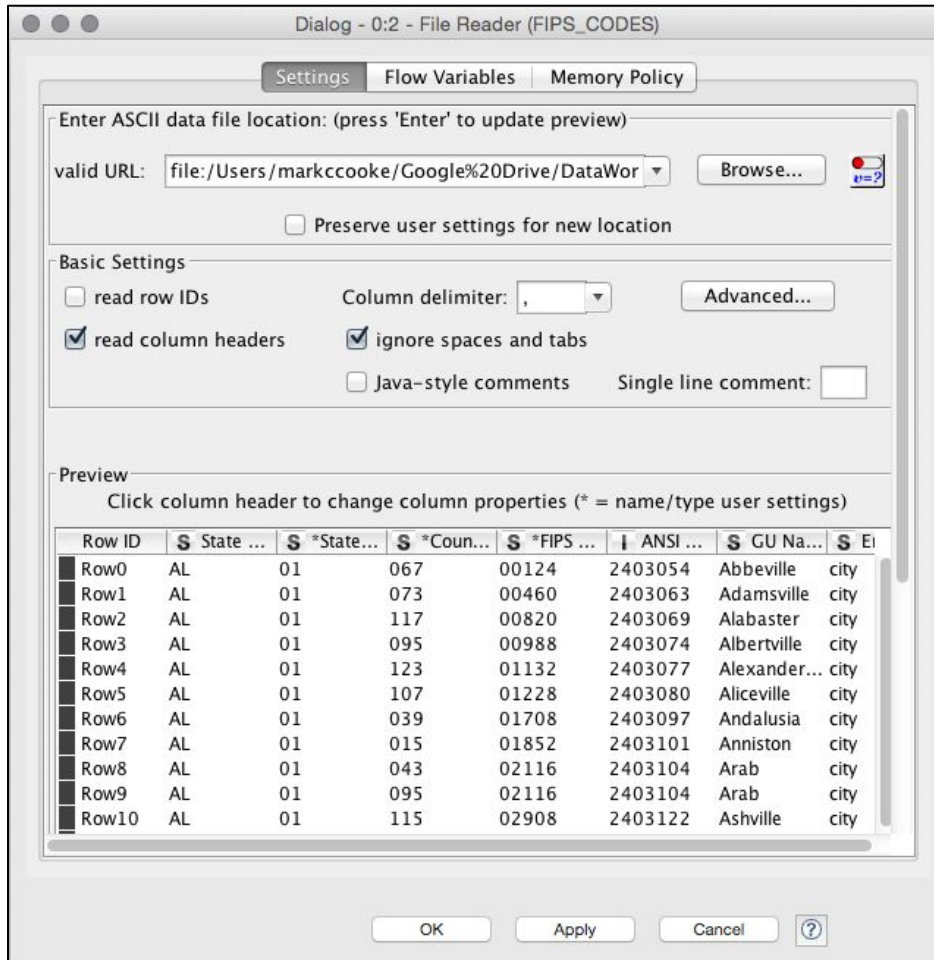
- HTTP Connection** (Node 1): Connects to the URL `http://maps.co.mecklenburg.nc.us/opendata/Parcel_TaxData.zip`.
- Download** (Node 2): Downloads the data from the HTTP connection.
- Table Column to Variable** (Node 4): Extracts data from the downloaded file.
- Unzip Files** (Node 3): Unzips the downloaded data.
- Table Column to Variable** (Node 6): Further processes the data.
- Shapefile Polygon Reader** (Node 5): Reads the data as a shapefile.
- String Replacer** (Node 7): Replaces specific strings in the data.
- String Manipulation** (Node 9): Performs string manipulation operations.
- String Manipulation** (Node 10): Another string manipulation step.
- String to Date&Time** (Node 8): Converts string data to date and time format.
- Rule-based Row Filter** (Node 11): Filters rows based on defined rules.
- Nominal Value Row Splitter** (Node 12): Splits rows based on nominal values.
- Date&Time-based Row Filter** (Node 13): Filters rows based on date and time.
- GroupBy** (Node 27): Groups data by a specific column.
- GroupBy** (Node 28): Another grouping operation.
- Math Formula** (Node 24): A row of five math formula nodes: **Ratio**, **MedianRatio**, **Absolute Ratio Difference**, **Mean Absolute**, and **CoEff of Dispersion**.
- Numeric Binner** (Node 25): Bins numeric data into categories.
- Color Manager** (Node 26): Manages colors for data visualization.
- Date&Time to String** (Node 29): Converts date and time back to string format.
- Math Formula** (Node 31): **Distance** calculation.
- Rule-based Row Splitter** (Node 32): Splits rows based on rules.
- Date&Time Difference** (Node 34): Calculates differences between dates and times.
- Joiner** (Node 36): Joins data from different sources.
- Java Snippet (simple)** (Node 37): Executes a simple Java snippet.
- 2D/3D Scatterplot** (Node 33): Visualizes the data as a scatterplot.
- GroupBy** (Node 35): Final grouping operation.

The interface includes a **KNIME Explorer** on the left showing the project structure, a **Node Repository** for selecting components, and a **Console** at the bottom for monitoring the workflow's execution.

# KNIME WORKFLOW



# KNIME NODES



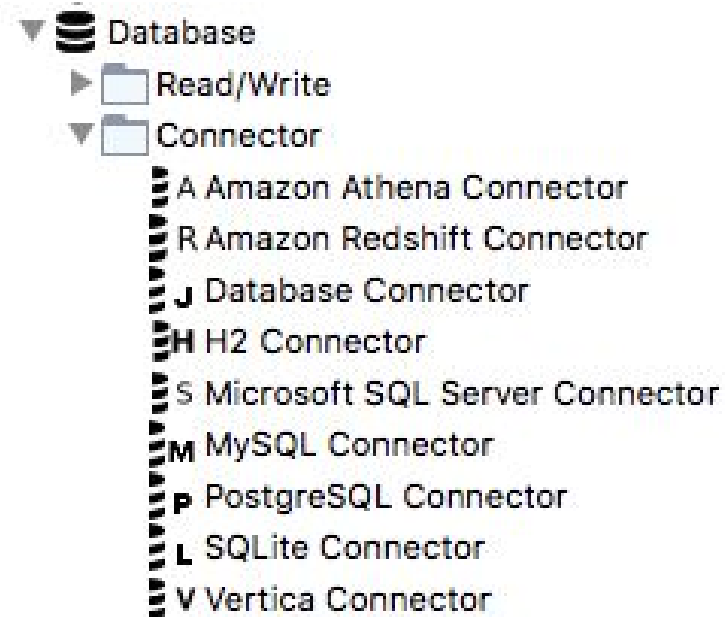
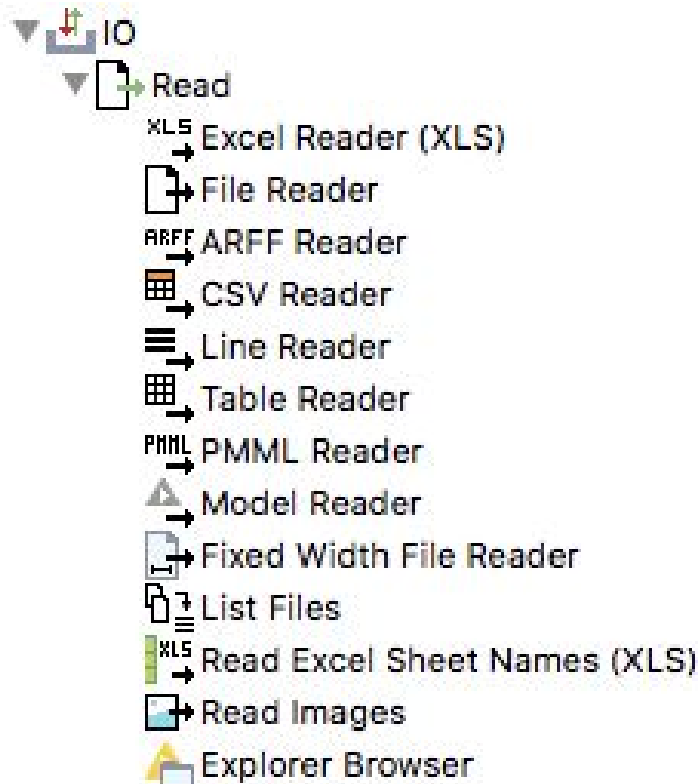


# DATA TYPES & SOURCES

# DATA AGNOSTIC

- ▶ Flat Files
- ▶ Xls/x Reader
- ▶ Fixed Width
- ▶ Text Files
- ▶ Image Files
- ▶ XML
- ▶ Shapefiles
- ▶ HTTP Requests
- ▶ RSS Feeds
- ▶ Custom API's/Curl
- ▶ Standard API's
- ▶ JSON

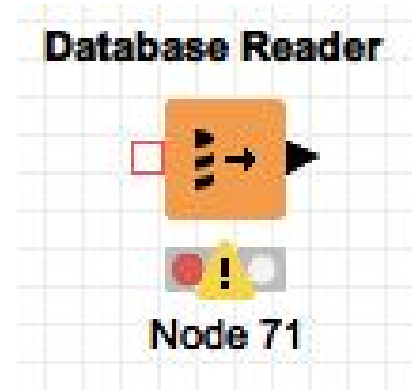
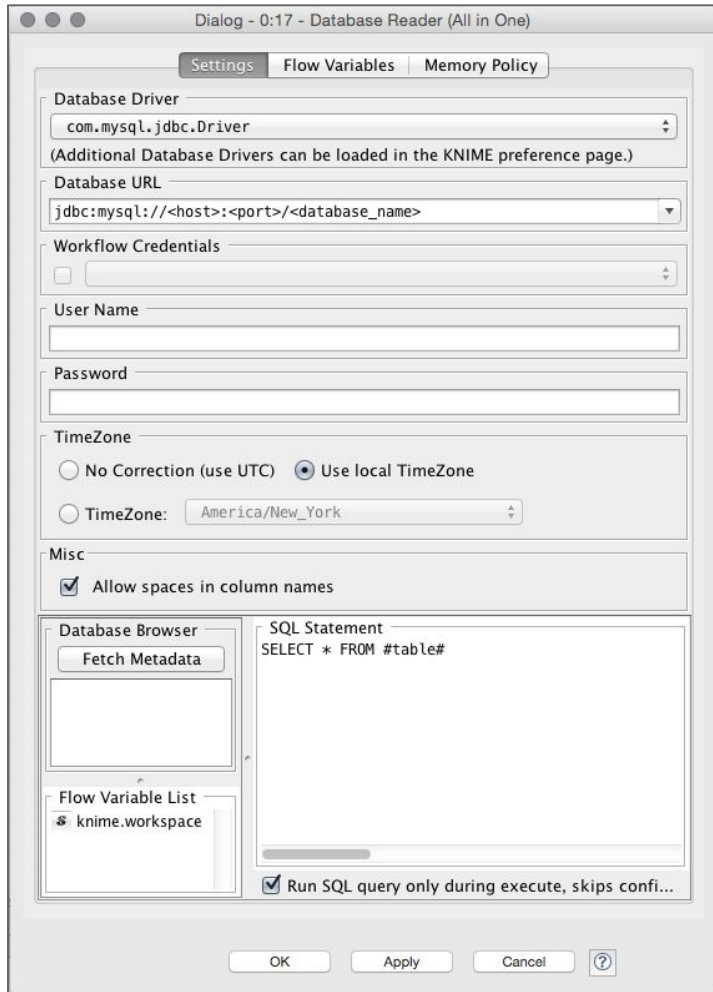
# KNIME DATA NODES



# DATABASE AGNOSTIC

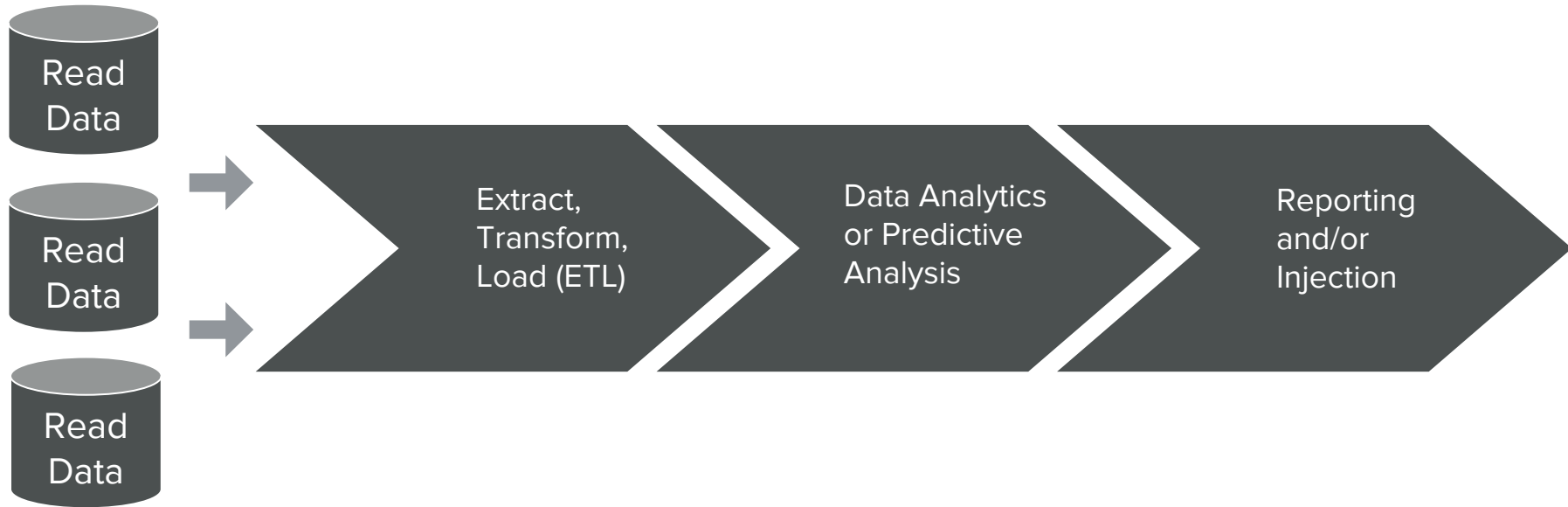
- ▶ Microsoft SQL
- ▶ MySQL
- ▶ Postgres
- ▶ SQLite
- ▶ Oracle
- ▶ IBM DB2
- ▶ Hadoop
- ▶ Any JDBC driver

# KNIME DATABASE NODES

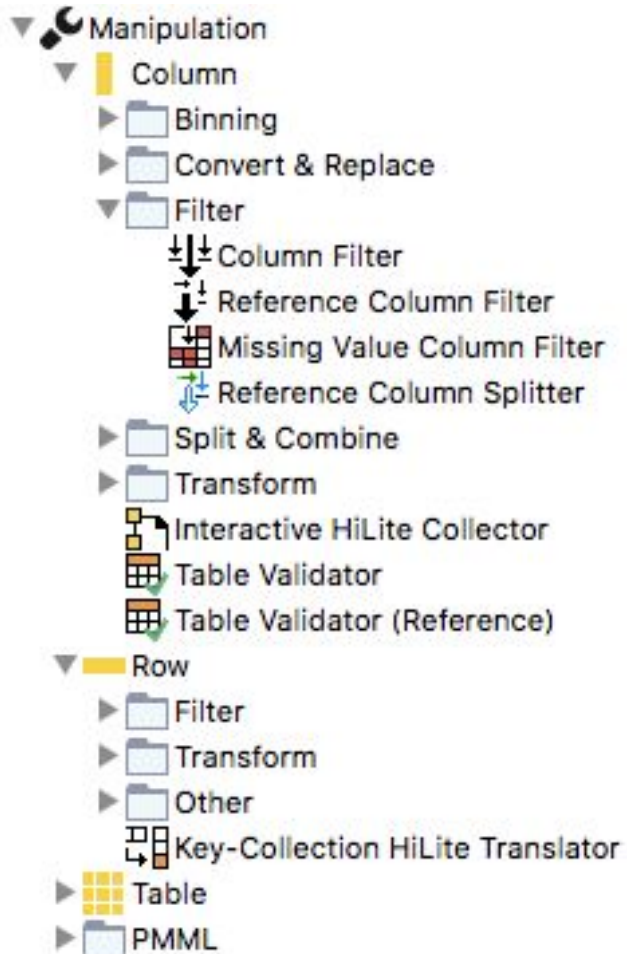


# CORE DATA ANALYTICS FEATURES

# KNIME DATA ANALYTICS LIFECYCLE



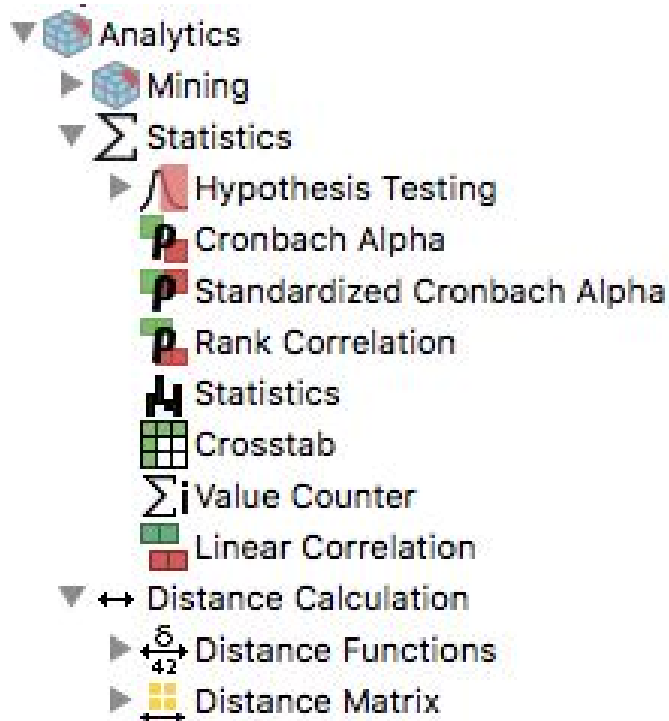
# KNIME ETL FEATURES



- ▶ Rule-based filtering and transformation
- ▶ Support for REGEX style replacements
- ▶ Map style reduction functions including statistical collapse
- ▶ Logical joins

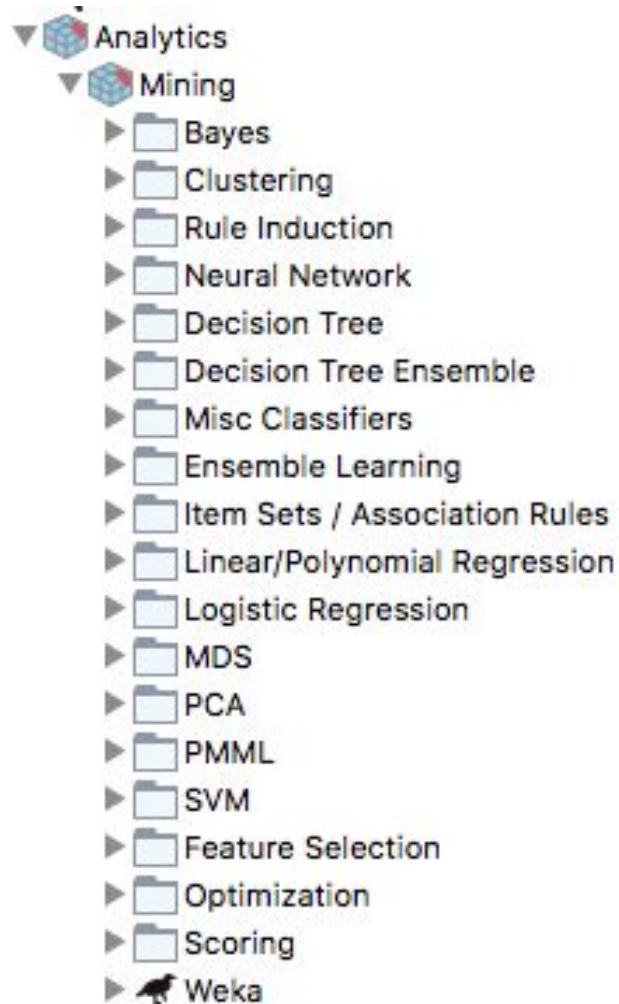


# KNIME STATISTICS



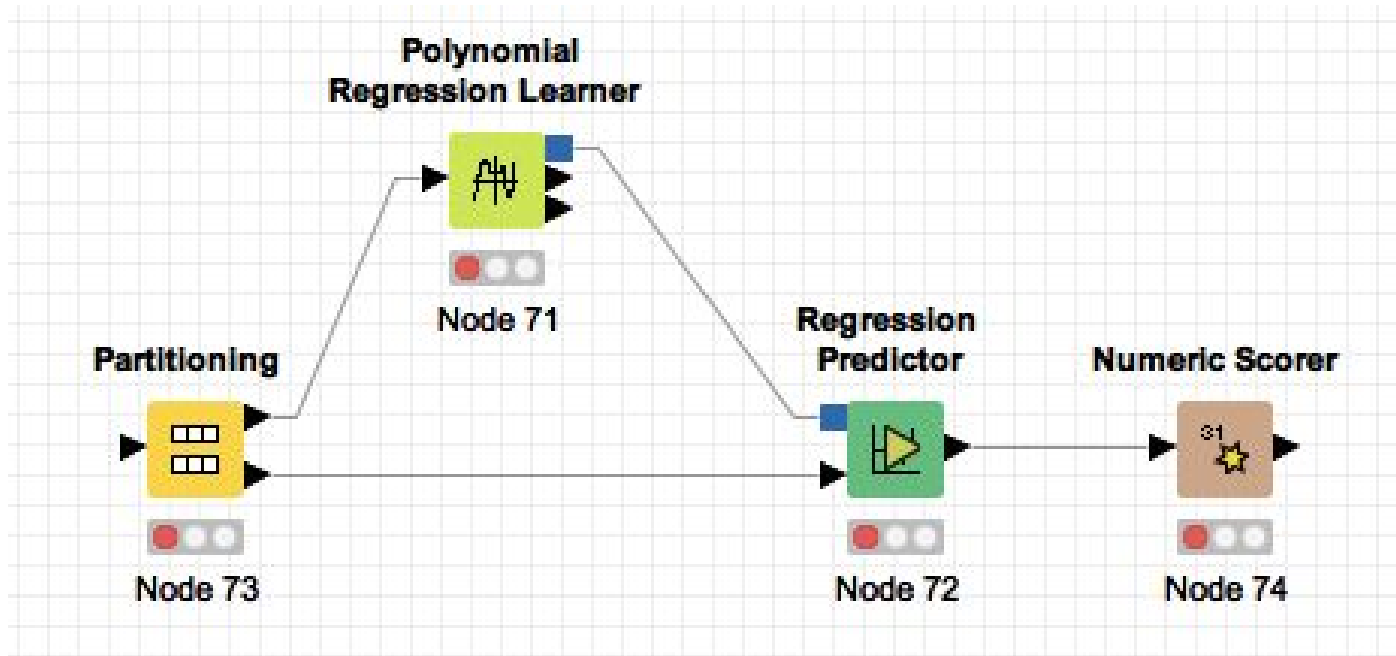
- ▶ “Statistics” node has base measures of distribution
- ▶ Linear correlation and dependency measures
- ▶ Many nodes also support statistical standards such as count, sum, mean, etc.

# KNIME MACHINE LEARNING



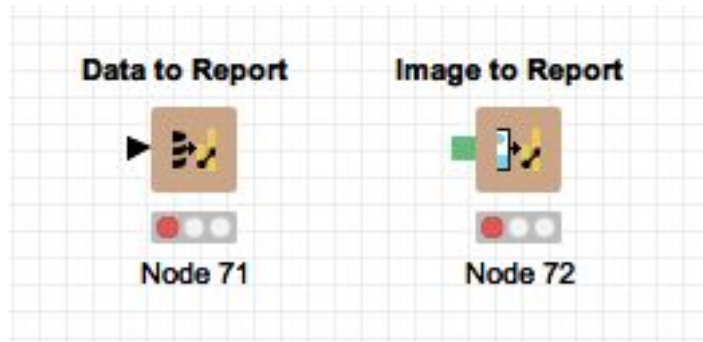
- ▶ Base KNIME supports most machine learning algorithms
- ▶ These are extended through partner implementations and scripting languages (R, Python, Weka, etc.)
- ▶ Data partitioning and multiple folds

# KNIME Machine Learning



# KNIME REPORTING

- ▶ Generates reports in office document formats, PDF, and HTML
- ▶ BIRT Tool as part of the Eclipse framework
  - Native part of the KNIME workbench
- ▶ Extends data visualization capabilities
- ▶ Auto-distribute by email, or publish to websites

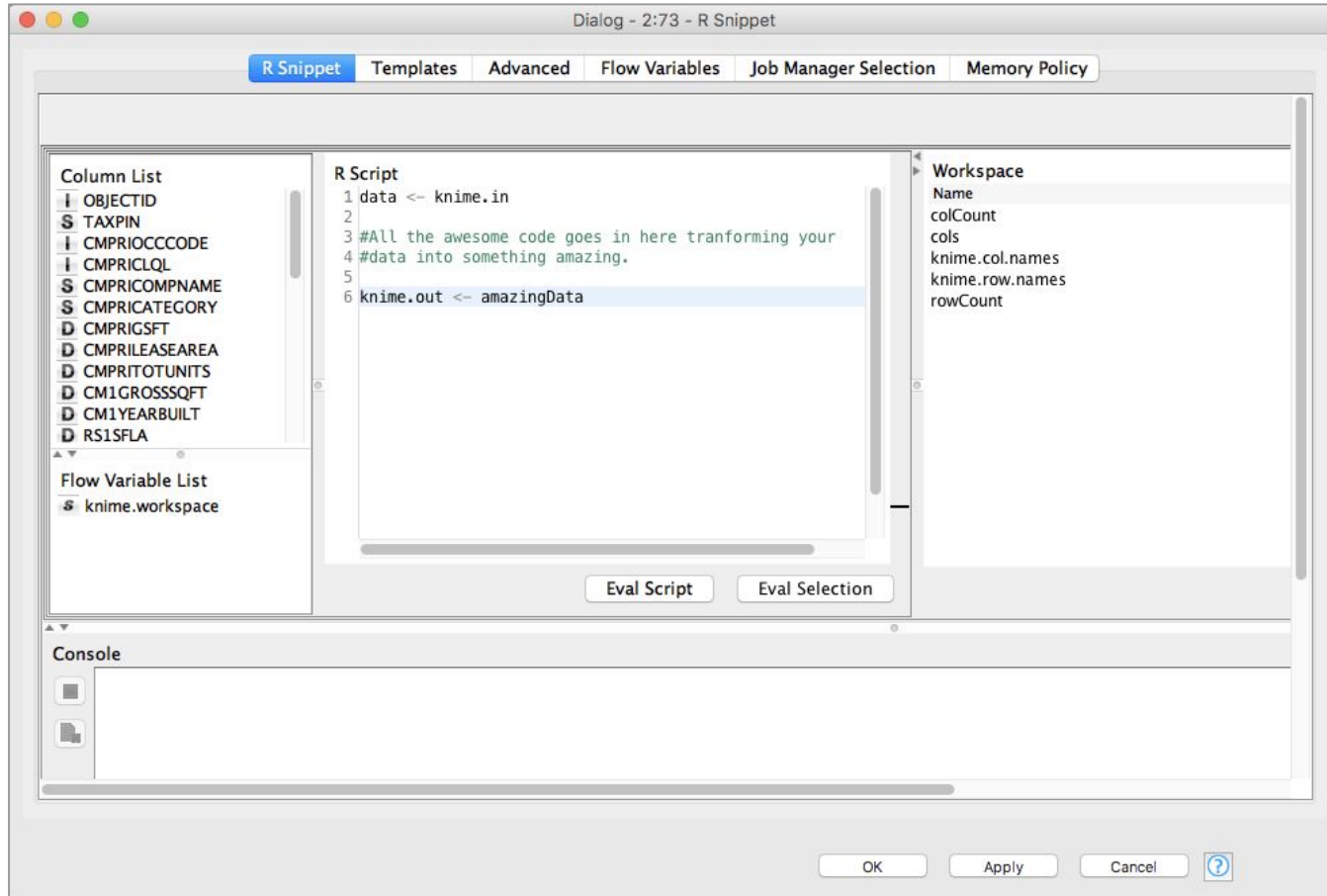


# ENHANCEMENTS & EXTENSIONS

# KNIME Enhancements

- ▶ Community Contributions
  - Third party approved KNIME nodes
- ▶ Integration for Popular Languages
  - Python, P, Java, Javascript, Perl

# KNIME R INTEGRATION



# WHAT IS R ANYWAY?

R is a programming language developed principally for data analytics and statistics.



**Visit:** <https://www.r-project.org>

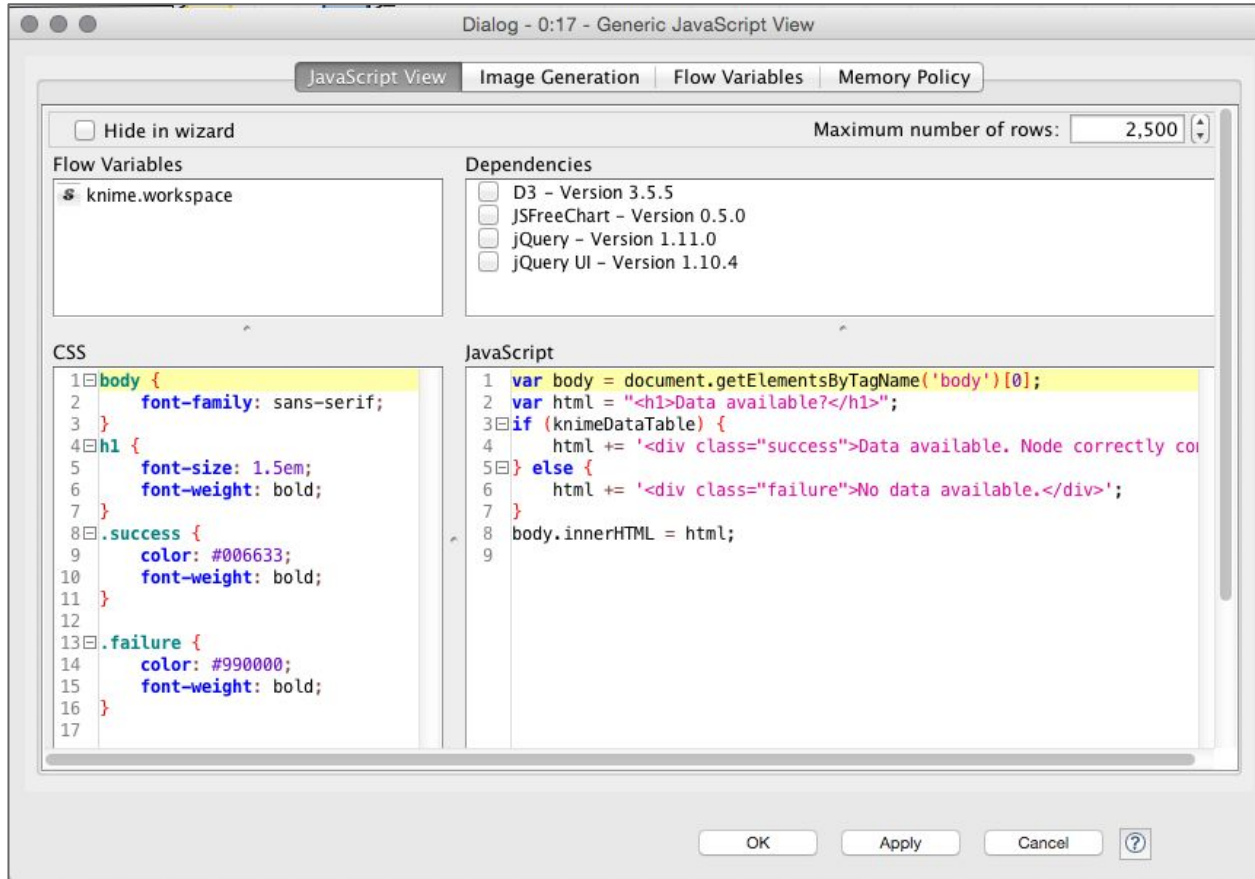
- ▶ R supports graphical representations as well, and has been extended to reporting and web applications

## **EXAMPLES:**

- R Markdown
- Shiny



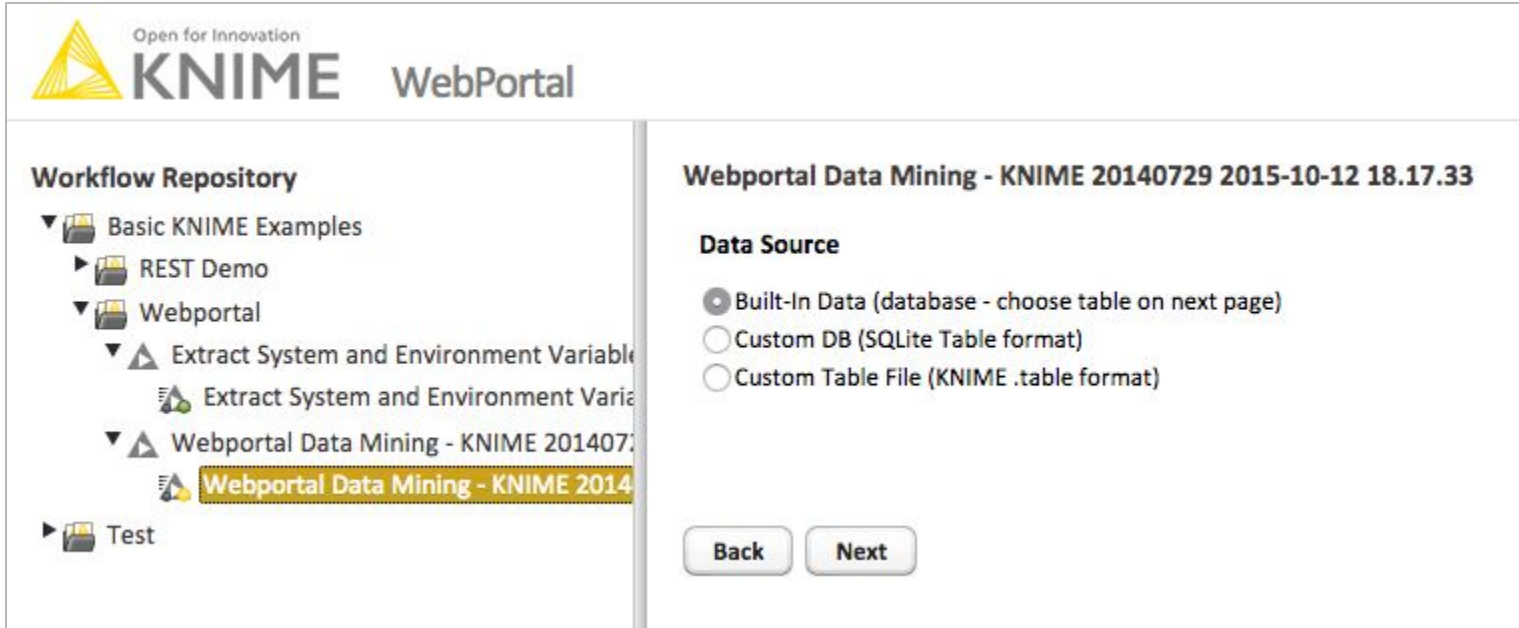
# KNIME JAVASCRIPT INTEGRATION



# COMMODITY EXTENSIONS

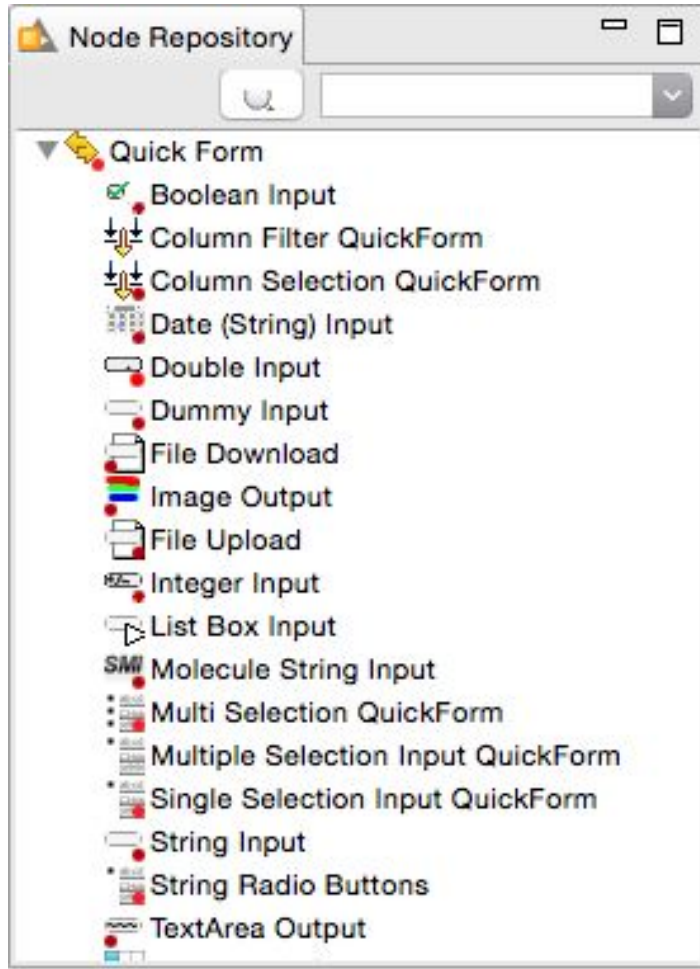
# KNIME SERVER

- ▶ Teamspace: collaboration and shared routines
- ▶ “Chron jobs” and remote execution
- ▶ Inbuilt web-services for remote access, configuration, and execution (Parameterized URL execution calls)



The screenshot displays the KNIME WebPortal interface. At the top left is the KNIME logo with the tagline "Open for Innovation" and the text "WebPortal". The main content is divided into two panels. The left panel, titled "Workflow Repository", shows a tree view of workflows. Under "Webportal", the workflow "Webportal Data Mining - KNIME 20140729" is expanded, and its sub-workflow "Webportal Data Mining - KNIME 20140729" is highlighted. The right panel, titled "Webportal Data Mining - KNIME 20140729 2015-10-12 18.17.33", shows configuration options for the "Data Source". Three radio buttons are present: "Built-In Data (database - choose table on next page)" is selected, "Custom DB (SQLite Table format)", and "Custom Table File (KNIME .table format)". At the bottom of the right panel are "Back" and "Next" buttons.

# KNIME WEBPORTAL



- ▶ End-User initiated workflows
- ▶ Generates a webpage to represent the workflow and options
- ▶ End-User selections configurable from inside workflows

# INNOVATION & DATA ANALYTICS

## USE CASE

# DATA ANALYTICS USE CASES

## DATA AGGREGATION

- ▶ Combine data from different sources, local or remote
- ▶ ETL data into a single repository for querying/analytics

## BUSINESS INTELLIGENCE

- ▶ Data intelligence and reporting over large aggregated datasets
- ▶ Automated reusable workflows for standardized reporting

# DATA ANALYTICS USE CASES

## PREDICTIVE ANALYTICS

- ▶ Discover patterns across time or entities
- ▶ Ability for insight across very large datasets

## KNIME ANALYTICS

- ▶ Advantage of being a data agnostic aggregator
- ▶ Ability to work through very large datasets with little hardware
- ▶ Access to complex algorithms with easy tools

# THANK YOU FOR WATCHING

Tax Management Associates appreciates your time.

**Please Contact:**



Mark C. Cooke, Ph.D.



704.847.1234



Mark.Cooke@tma1.com

**Visit Our Website:** [www.tma1.com](http://www.tma1.com)

**Download KNIME:** [www.knime.org](http://www.knime.org)



TAX MANAGEMENT  
ASSOCIATES, INC