

Preparing data for analysis

Introduction to KNIME



Co-funded by the
Erasmus+ Programme
of the European Union

Contents

- Let's get some data
- The KNIME environment
- Do exploratory data analysis.
 - “*Descriptive analytics*”





First, let's get some data

- In three formats.
 - CSV, TSV, XLS.
- Eurostat.
 - By theme
 - Let's choose some tables.
 - Life expectancy (?)
 - Educational level (?)
 - Greenhouse Gas emissions (?)
 - Let's download them.
 - CSV, TSV, XLS.
 - No Flags, no footnotes, separate tables, no delimiters

We assume our data comes from disparate sources



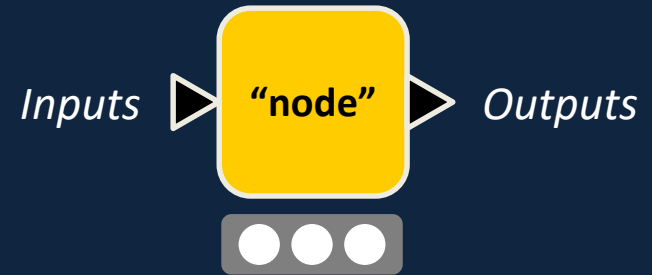
The working environment

& some useful settings.

- The Workspace.
 - Workflow space – your canvas
 - Nodes & Node info – your palette
 - Console – your feedback
- Workspace
 - Any folder
 - You may switch between workspaces
- File → Preferences
 - General → Web browser
 - Choose your favorite / installed
 - Knime GUI → *Workflow editor*
 - Grid : 10 x 10
 - Curved connections
 - R
 - Path to R home. (*folder above bin*)
 - Palladian → Location Extractors
 - [Register first – then add the entry]



The KNIME “logic”



- 1 Each node does one thing. **Only.**
- 2 A node may have input(s) and output(s)
- 3 To use a node you need to have an input, configure it, or both. *It depends on the node.*
- 4 All node actions are accessible via right click.
- 5 To do anything meaningful, you simply connect two or more nodes together.
- 6 Node process status is denoted by traffic lights:



Not configured yet



*Processing completed,
with warnings.*



Not properly configured

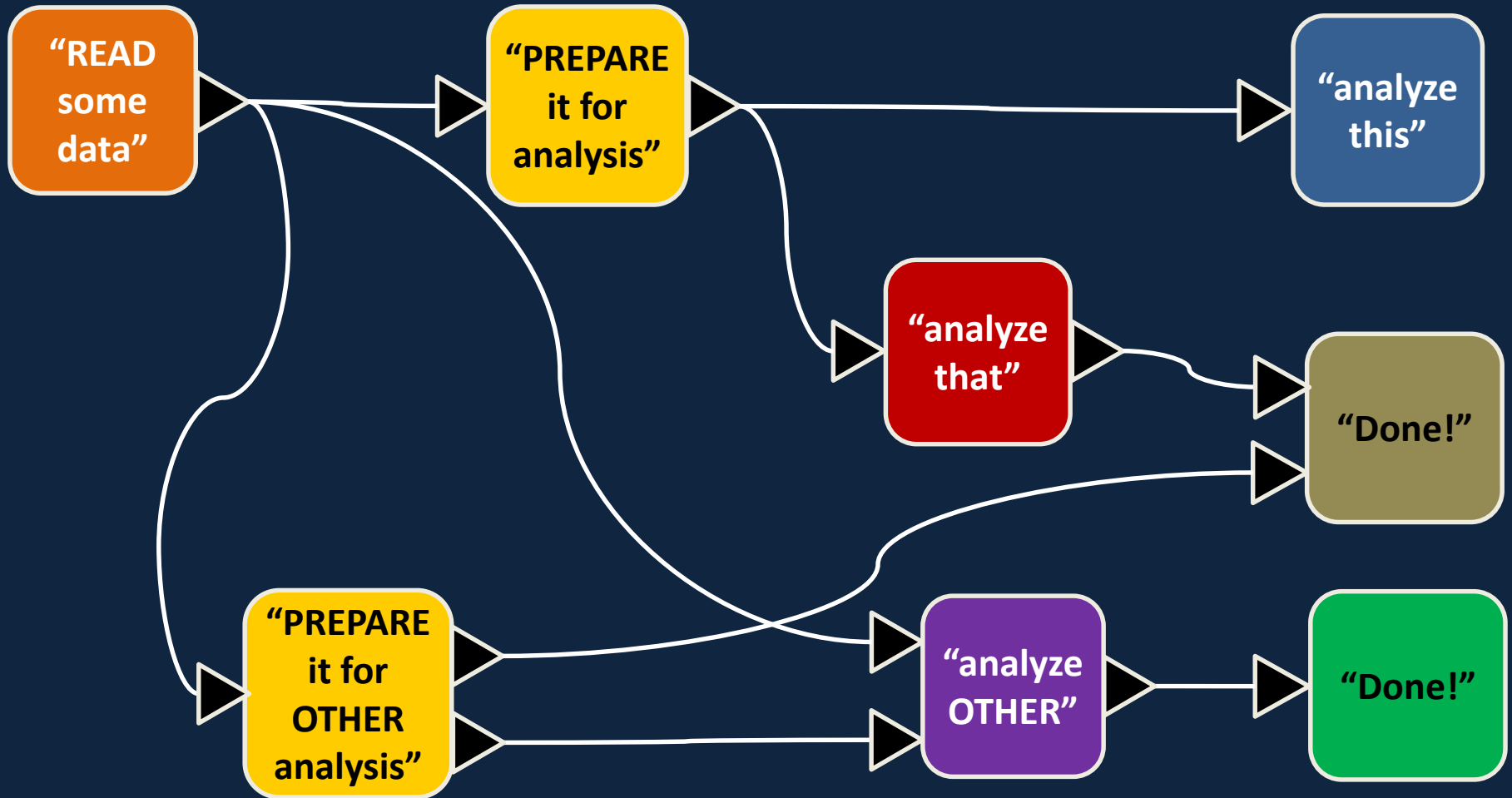


Processing completed.



Ready to process

'Meaningful' in KNIME is ...



Working with data tables

- Data preparations and analysis
 - **Table** operations
 - Combining, comparing, changing orientation, ...
 - **Row** operations
 - Filtering, adding, removing, *calculating* ...
 - **Column** operations
 - Filtering, adding, removing, changing data type, *calculating*, ...
 - **Cell** operations
 - Changing, *calculating*, ...



Let's read do some *exploratory data analysis*

- ***Objective:***
 - *Stats and charts for our downloaded data.*
- Read data
 - File reader
 - TSV
 - Delimited
 - CSV reader
 - XLS reader
- *And let's work our way through...*
 - *Prepare:* Make sure data is in proper form and type
 - *Analyze:* 'Do the math'. Or two.
 - *Visualize:* Show the output.



Summary

- Let's get some data
- The KNIME environment
- Do exploratory data analysis.
 - *“Descriptive analytics”*

