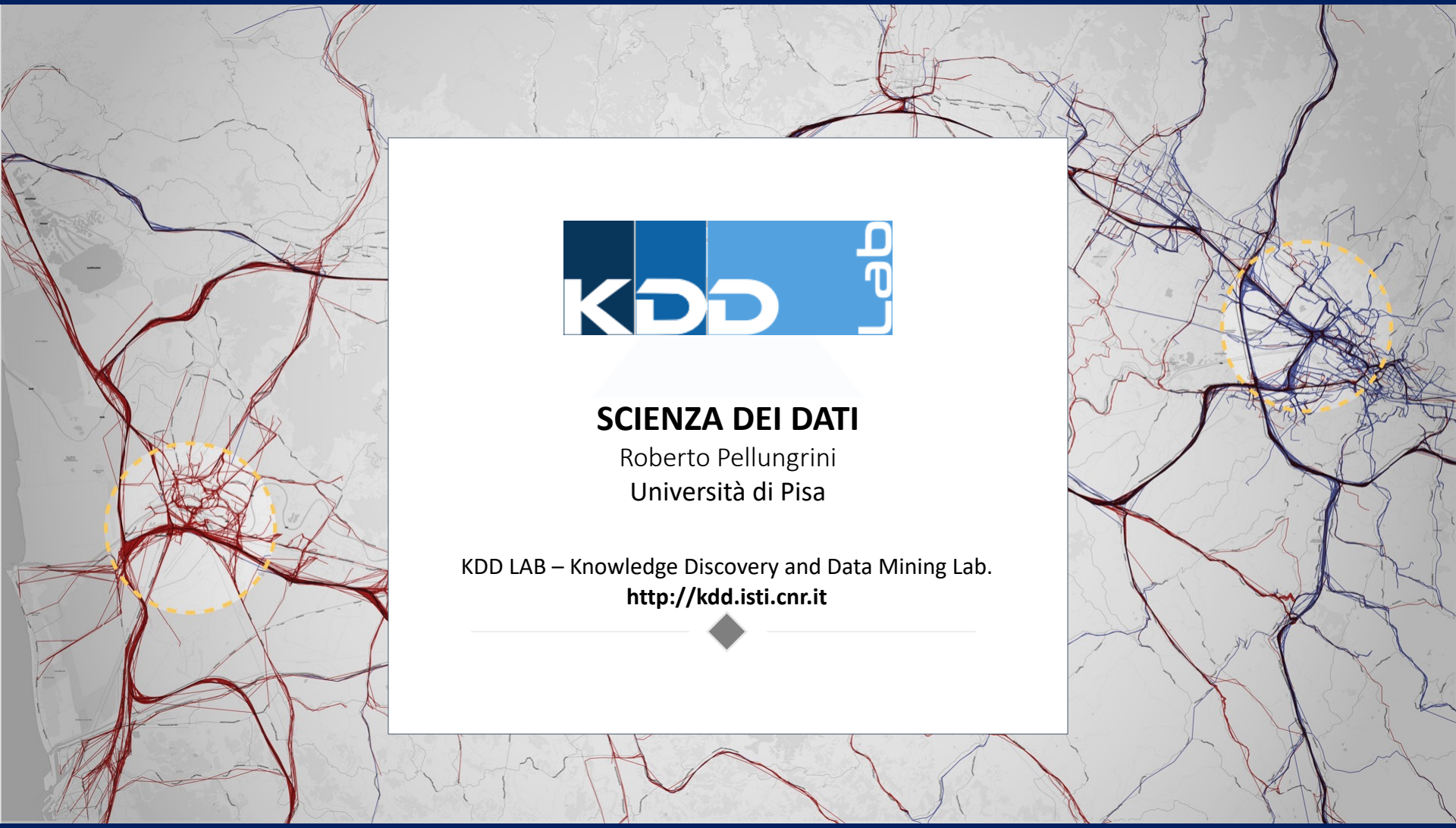


SCIENZA DEI DATI

Roberto Pellungrini
Università di Pisa

KDD LAB – Knowledge Discovery and Data Mining Lab.
http://kdd.isti.cnr.it

Cosa vedremo oggi

- Introduzione a KNIME
- Principali componenti
- Aprire ed esplorare i dati
- Clusterizzazione
- Classificazione

Cosa è KNIME?

- KNIME = Konstanz Information Miner
- Sviluppato inizialmente presso l'università di Konstanz in Germania
- Versione completamete gratuita per desktop
- Piattaforma modulare per la data science basata su **workflows a nodi.**
- Mette a disposizione funzioni standard per **data mining, analisi e manipolazione dei dati**
- Si possono installare varie estensioni per integrare nuove funzionalità

Risorse per KNIME

- Pagina web principale e documentazione
  https://www.knime.com
- Downloads
  https://www.knime.com/downloads/download-knime
- Apprendimento

  https://www.knime.com/learning

Download the latest KNIME Analytics Platform for Windows, Linux, and macOS: **4.4.2**. This version is intended for end users and provides everything needed to immediately begin using KNIME as well as extend KNIME with extension packages developed by others.

Windows

KNIME Analytics Platform for Windows (installer) *The installer adds an icon to the desktop and suggests suitable memory settings*	**Download** (548 MB)
KNIME Analytics Platform for Windows (self-extracting archive) *The self-extracting archive only creates a folder holding the KNIME installation*	**Download** (553 MB)
KNIME Analytics Platform for Windows (zip archive)	**Download** (677 MB)

Linux

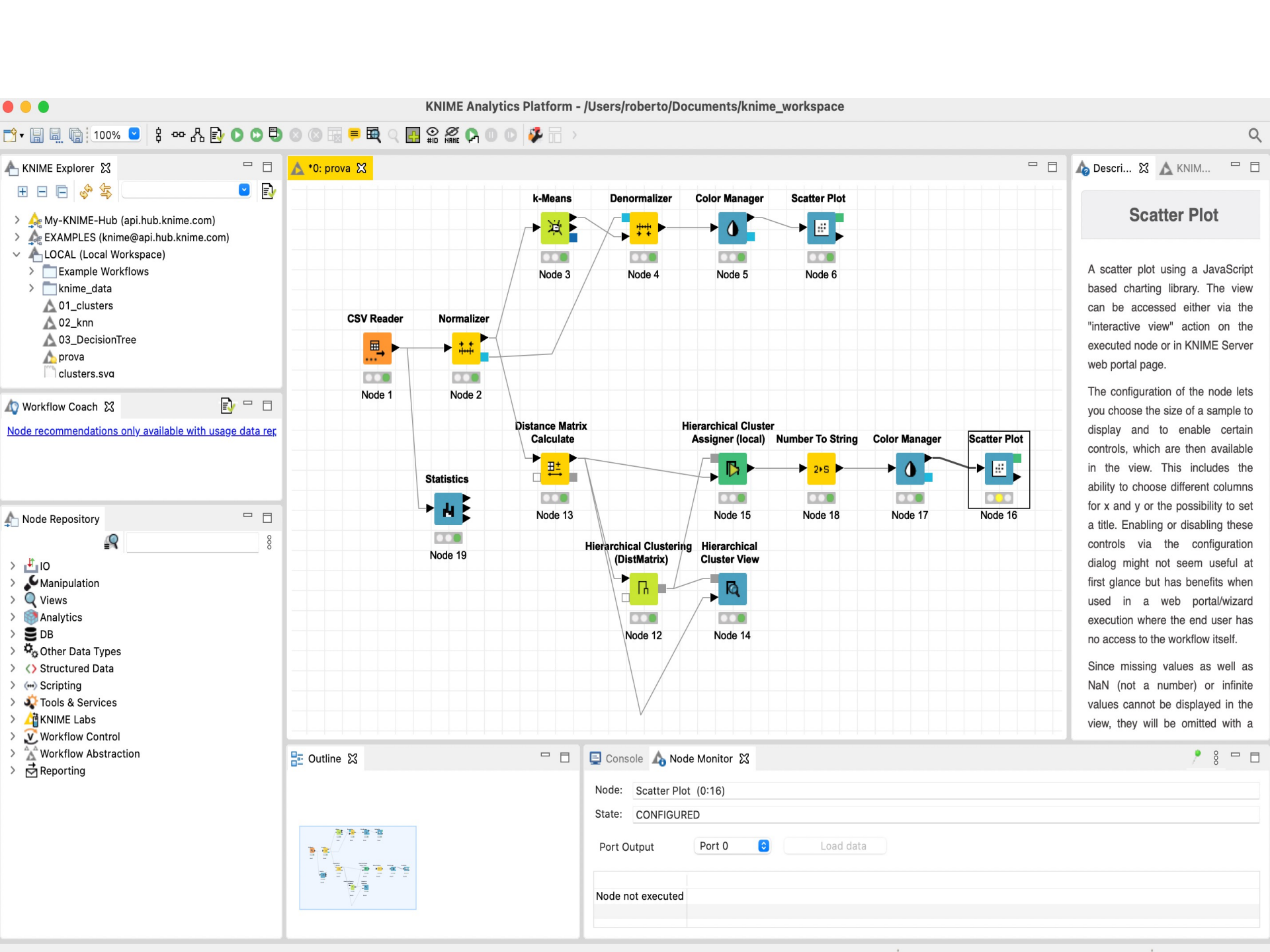
KNIME Analytics Platform for Linux	**Download** (712 MB)

Mac

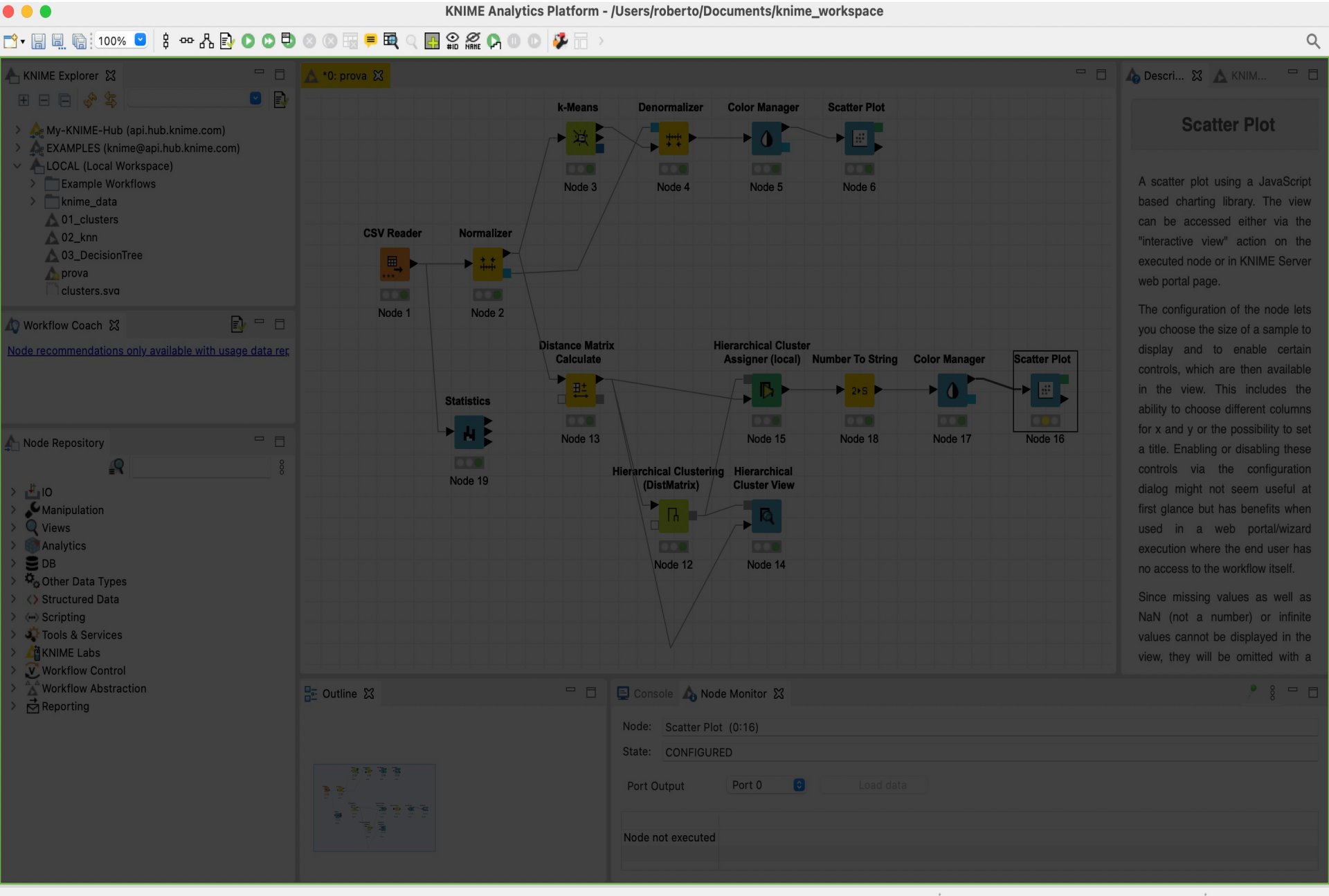
KNIME Analytics Platform for macOS (10.13 and above)	**Download** (556 MB)

Find out what's new in the latest KNIME 4.4 release here.

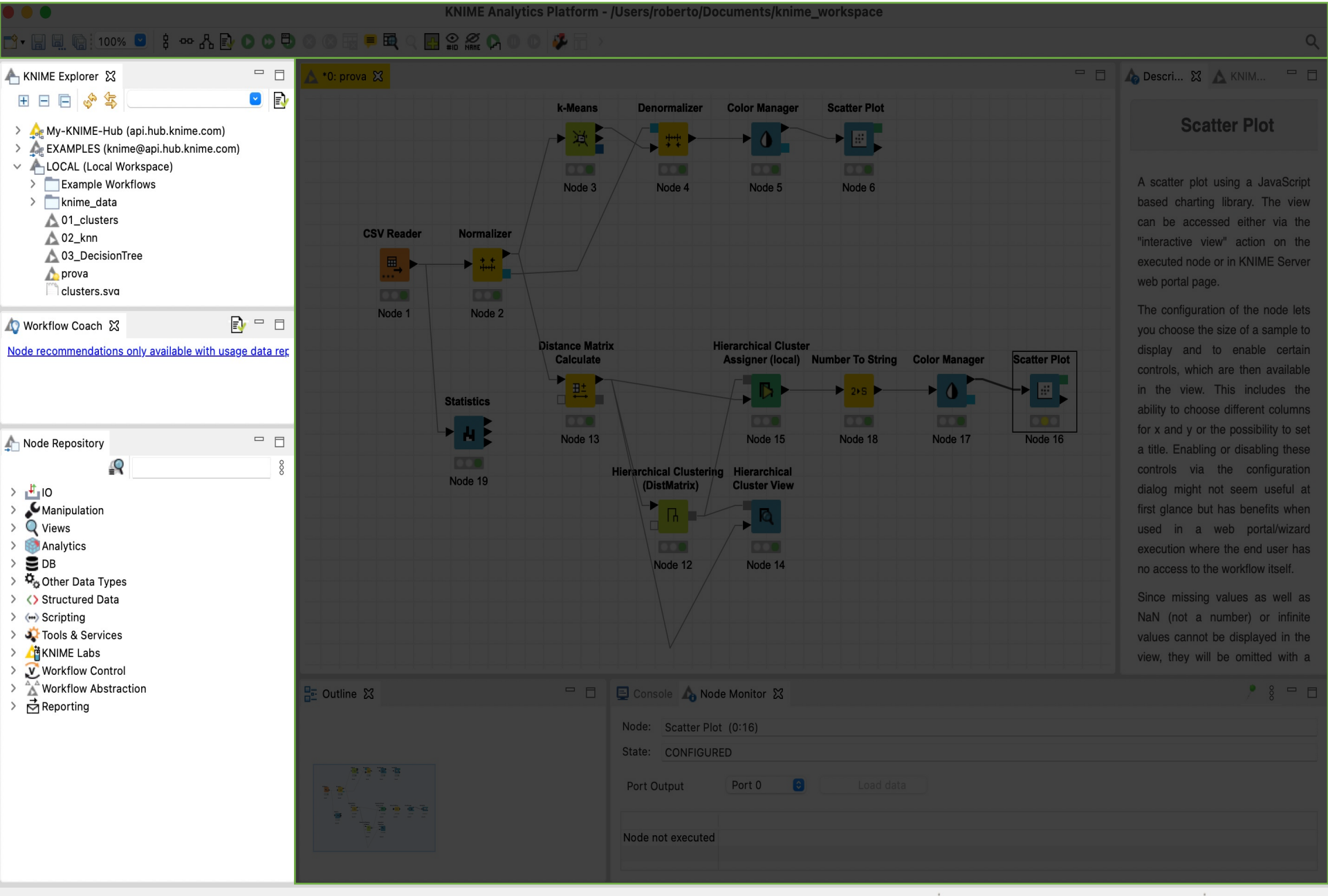
KNIME Analytics Platform - /Users/roberto/Documents/knime_workspace



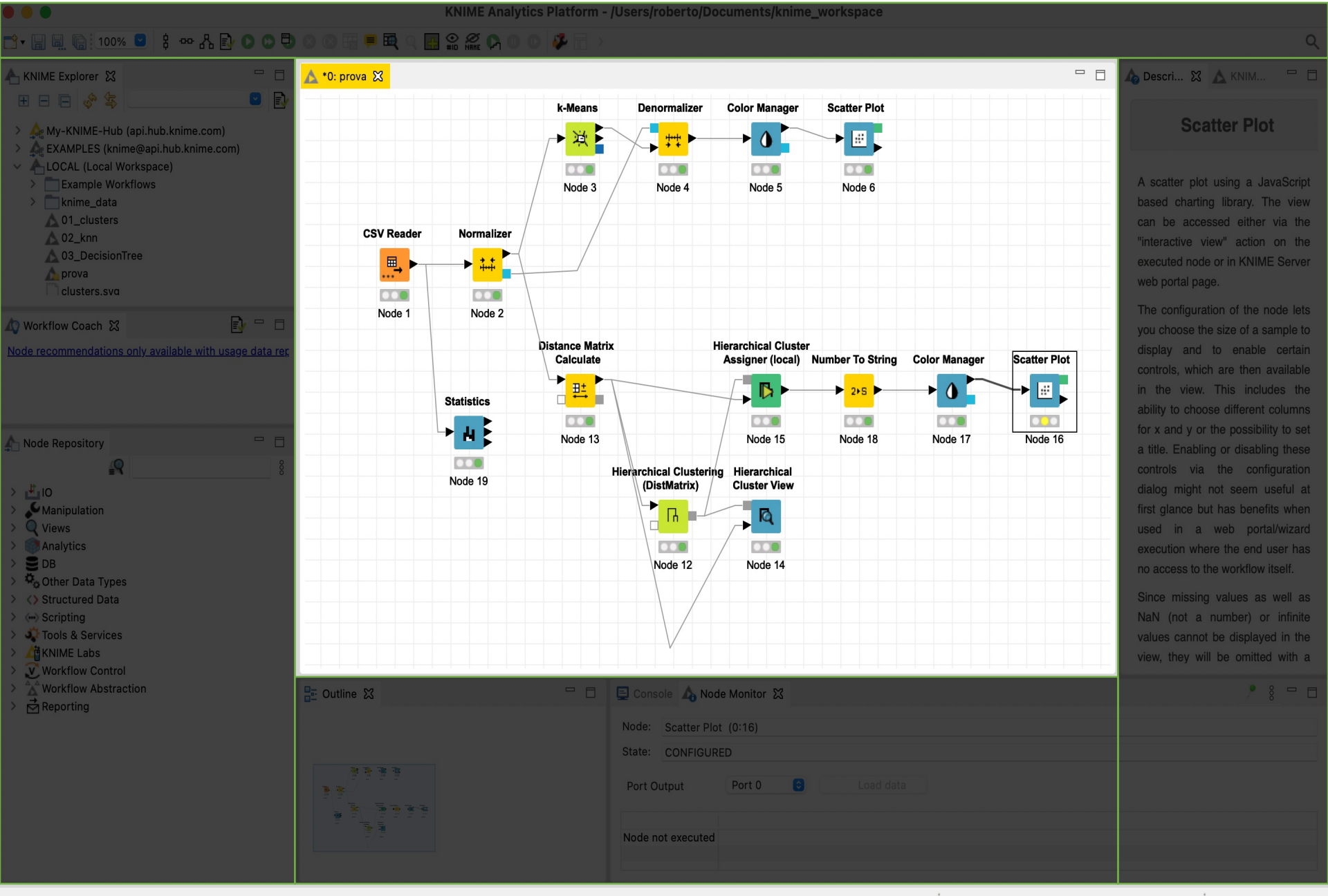
Barra dei comandi, bottoni di esecuzione e layout



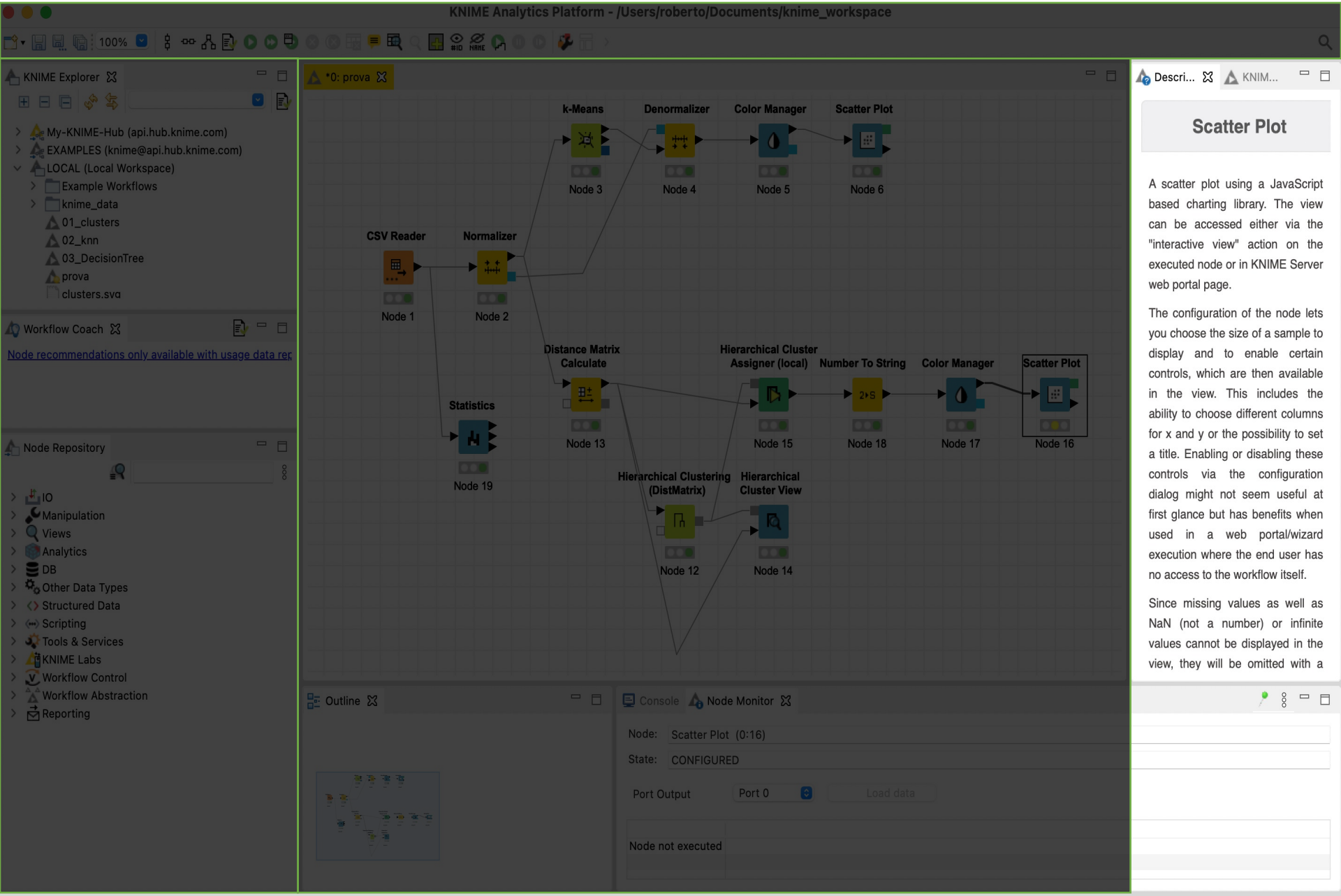
Barra laterale del workspace e selezione dei nodi



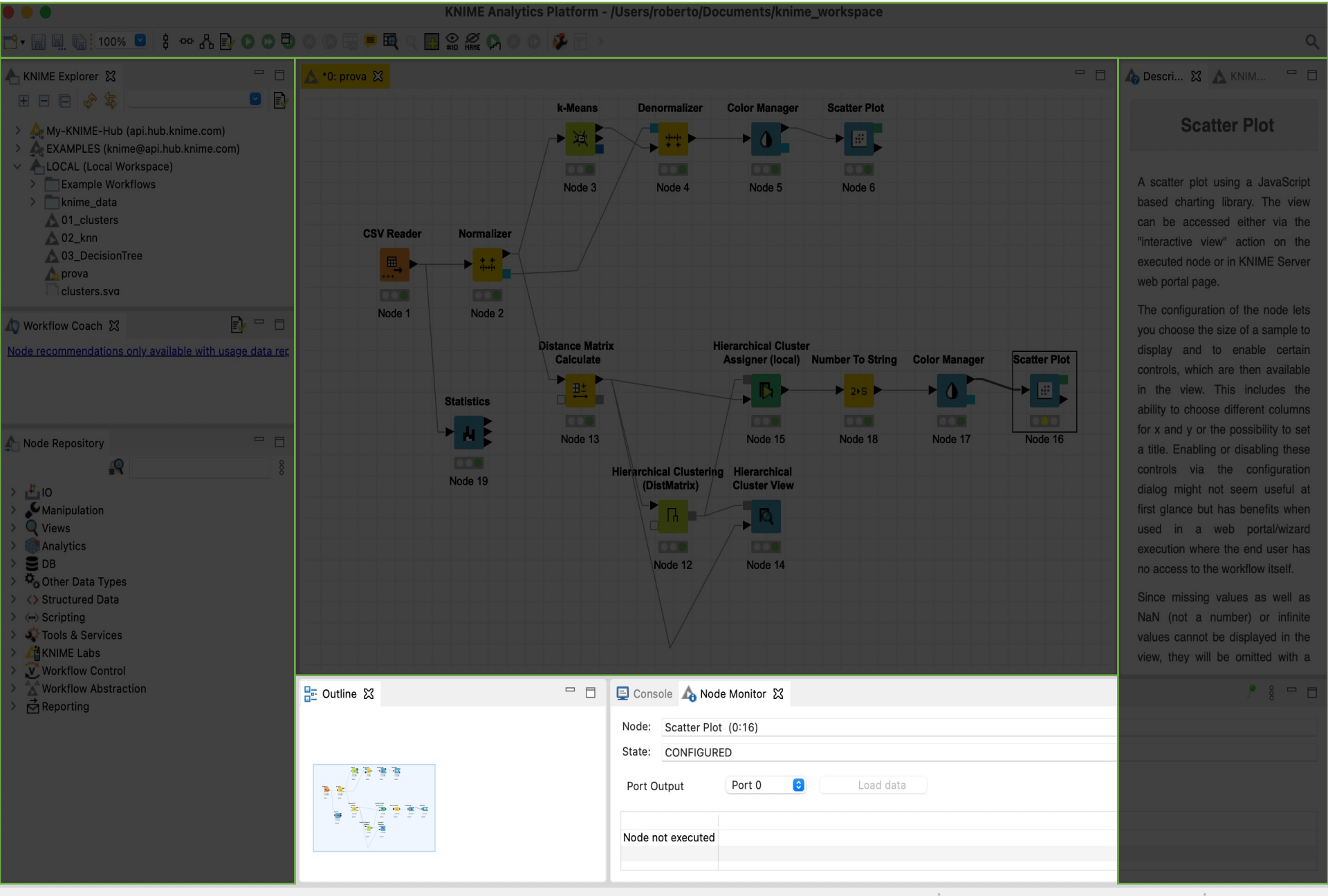
Workspace principale



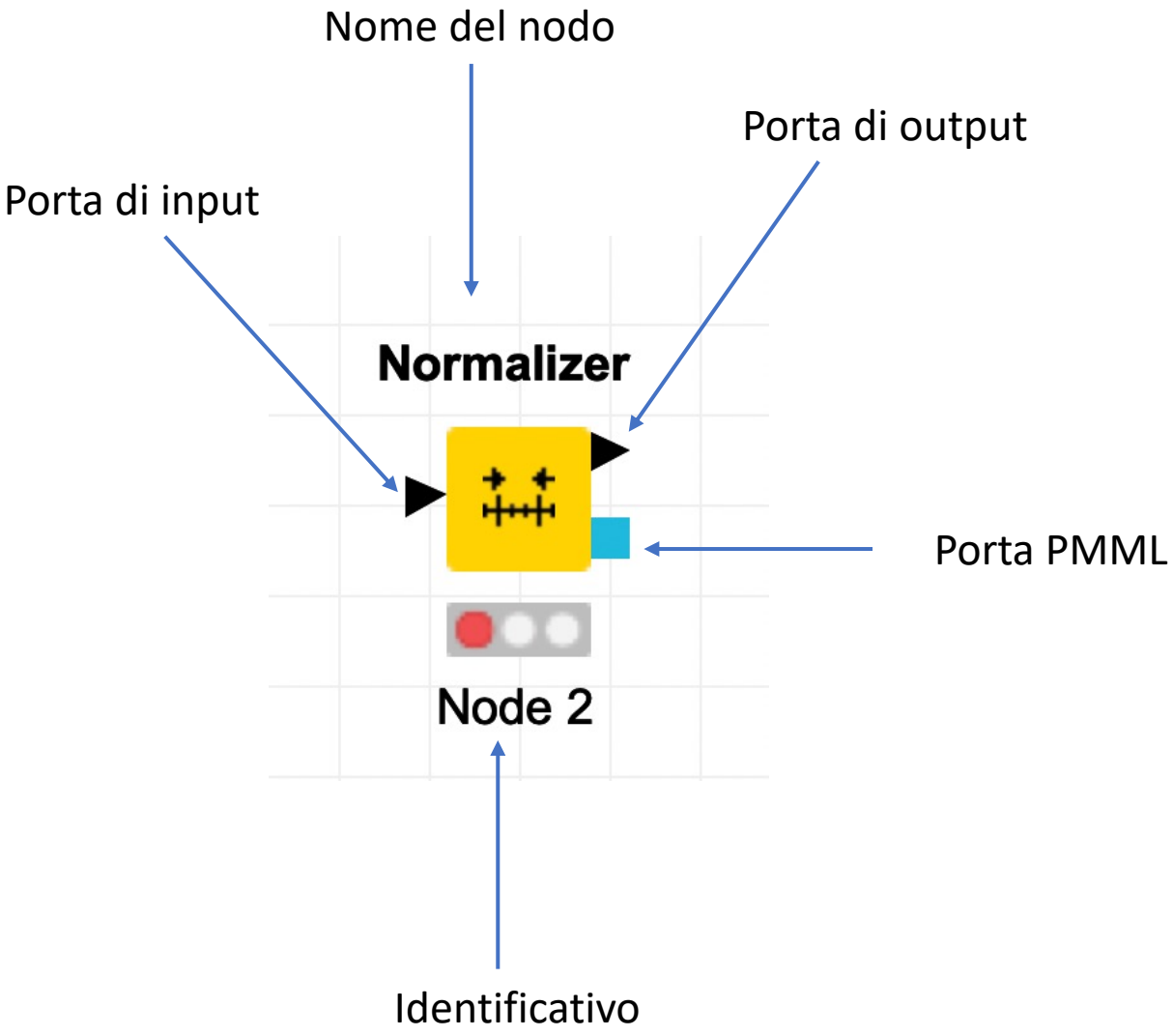
Descrizione dei nodi e del workspace



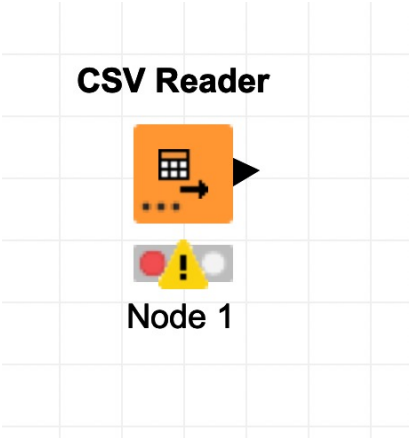
Outline e console



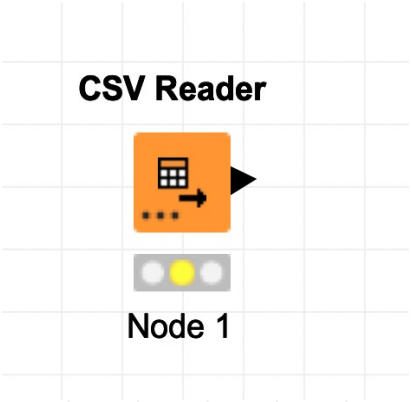
Struttura di un nodo



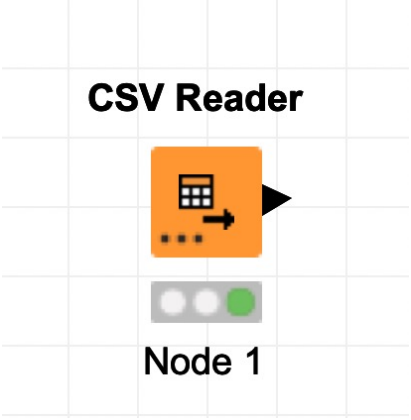
Stato del nodo



Da configurare

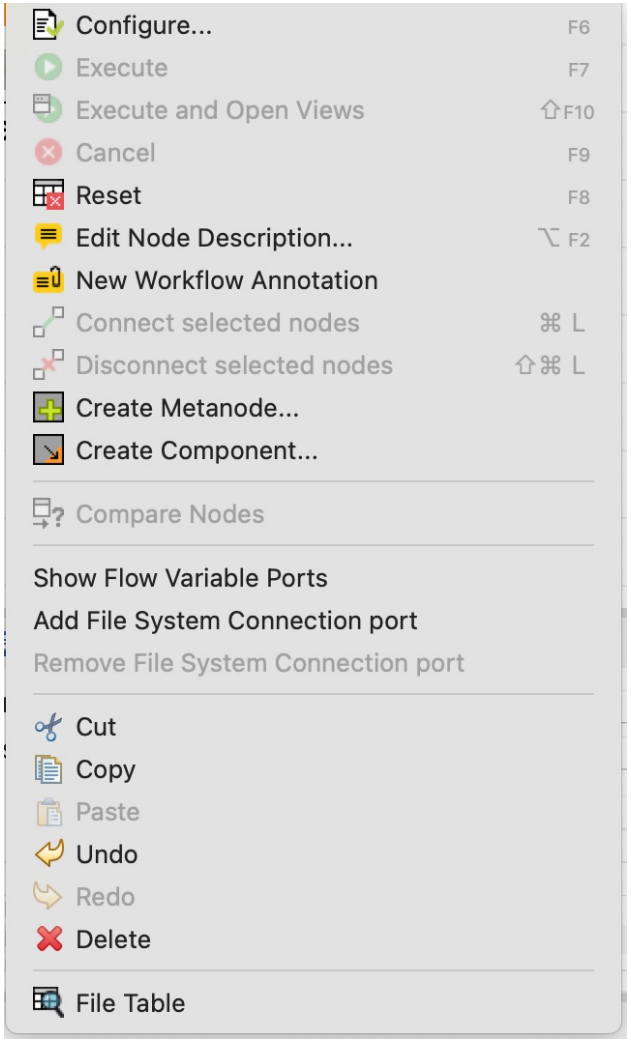


Pronto all'esecuzione

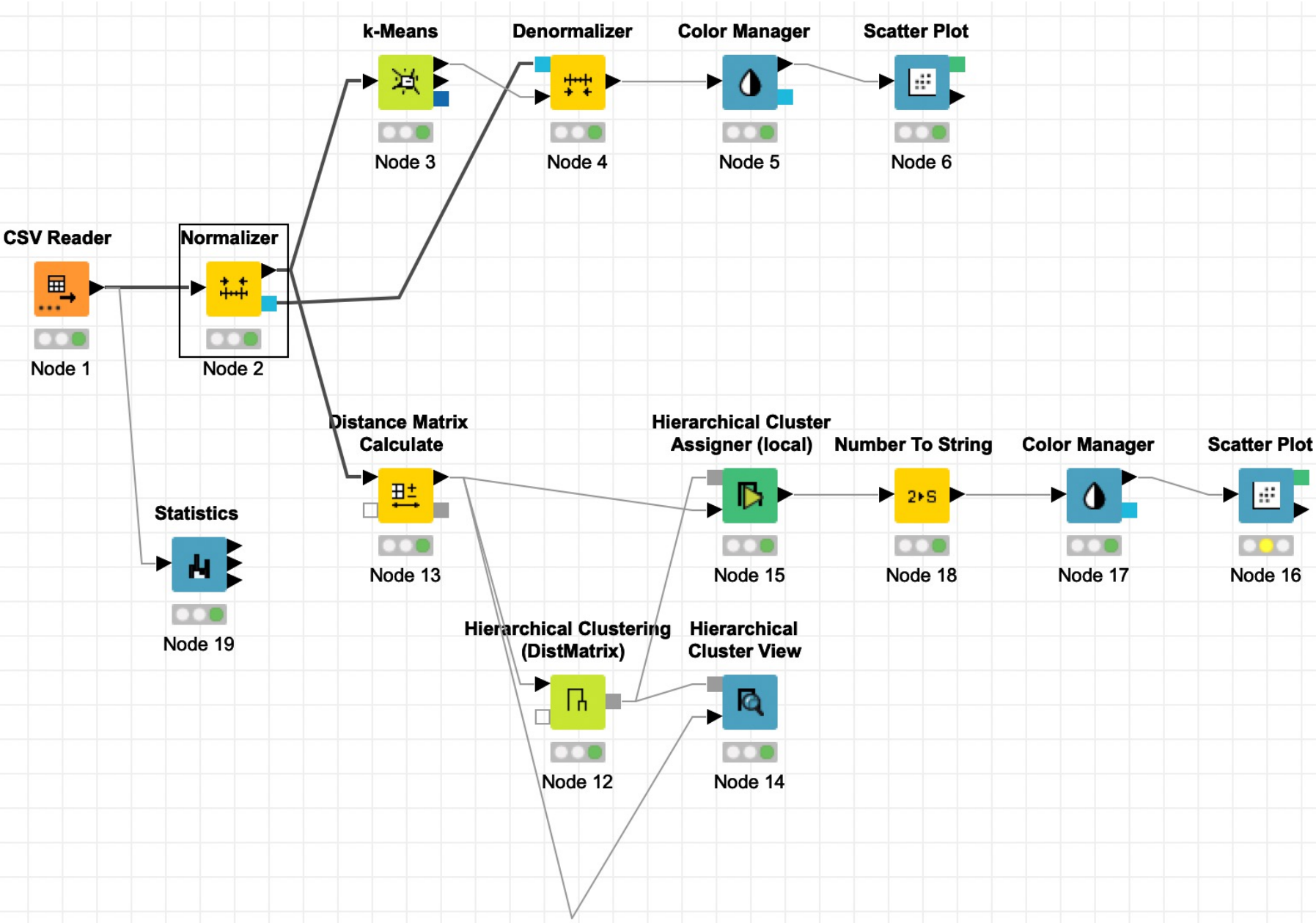


Eseguito

Menù contestuale:
tasto destro del mouse



Un workflow di esempio



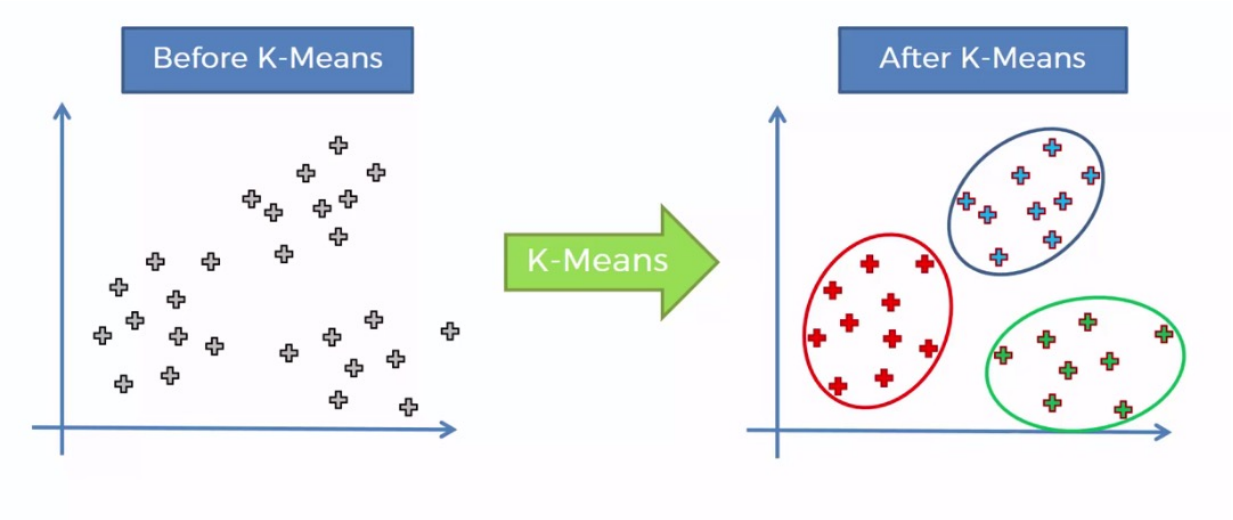
Normalizer

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Normalizer

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

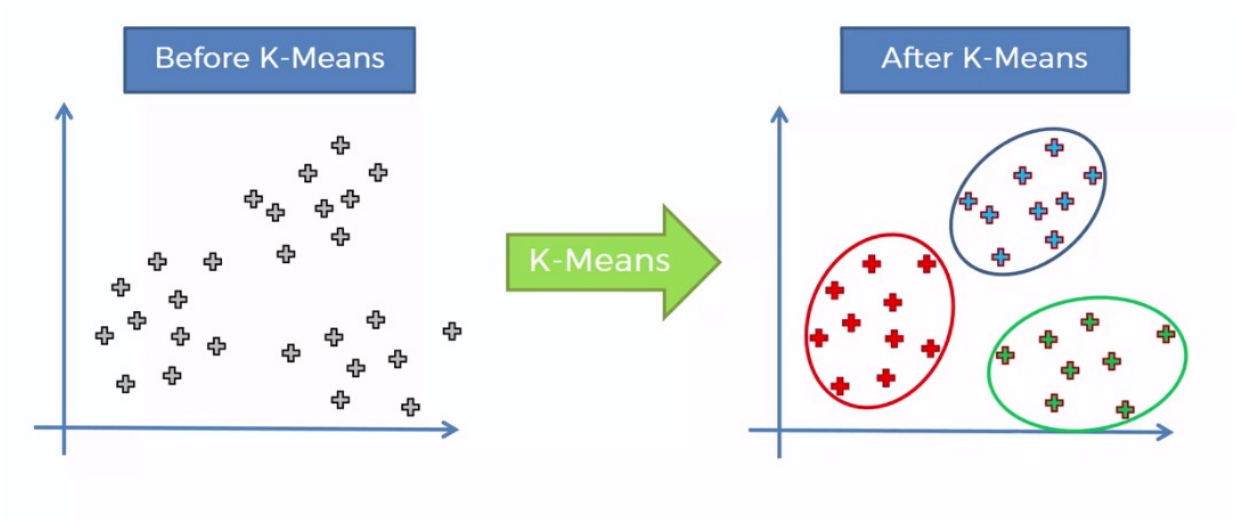
K-means



Normalizer

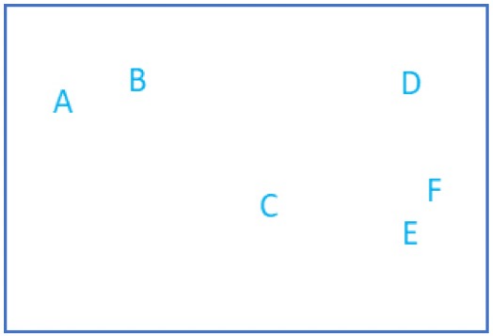
$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

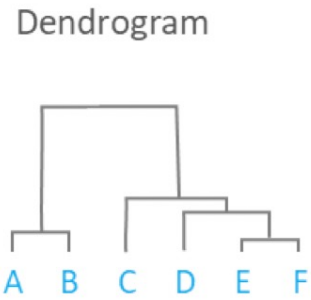
K-means



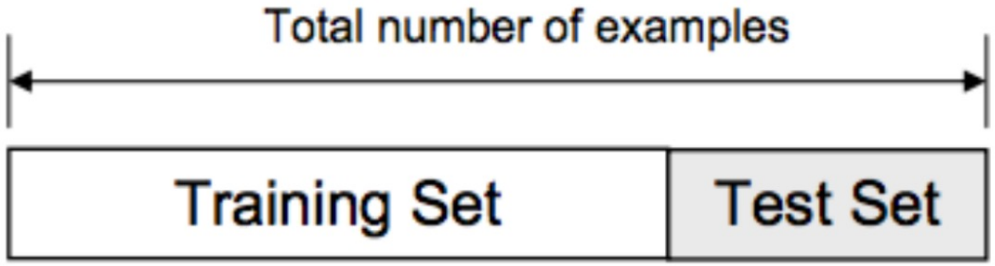
Cluster Gerarchico

	a	b	c	d	e	f
a	0	184	222	177	216	231
b	184	0	45	123	128	200
c	222	45	0	129	121	203
d	177	123	129	0	46	83
e	216	128	121	46	0	83
f	231	200	203	83	83	0





Train e test



Knn