KNIME - Data Science and Reporting

Pulling big data, csv/excel/google docs, real time data, and just about anything else for big decision making!

Michael Chrisco - Nov. 23rd 2022

What is KNIME?

What is KNIME?

in a nutshell

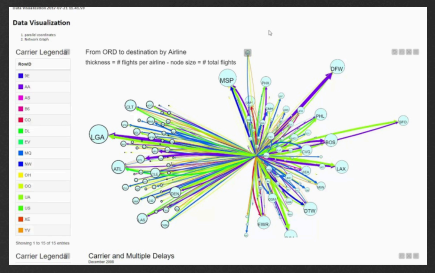
- KNIME Analytics Platform is the open source software for creating data science.
- Good at prototyping with lots of data
- Pull from many inputs at the same time
- Can create multiple reports off of the same data
- Can filter, sort, Groupby (all the data science things)
- Contains many community based resources such as AI modules, ODBCs, and many more!

What is KNIME?

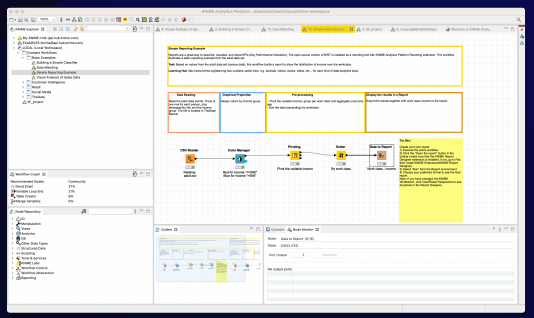
But what can it *really* do?

Data Science is all about answering the big questions…with Data!

- How is the business doing regarding finance? https://www.knime.com/blueprints-for-finance-analysis
- Find out what the housing market look like?
- How is your team doing?
- Deal with big data IE: large, hard-to-manage volumes of data – both structured and unstructured – that businesses use on a day-to-day basis.

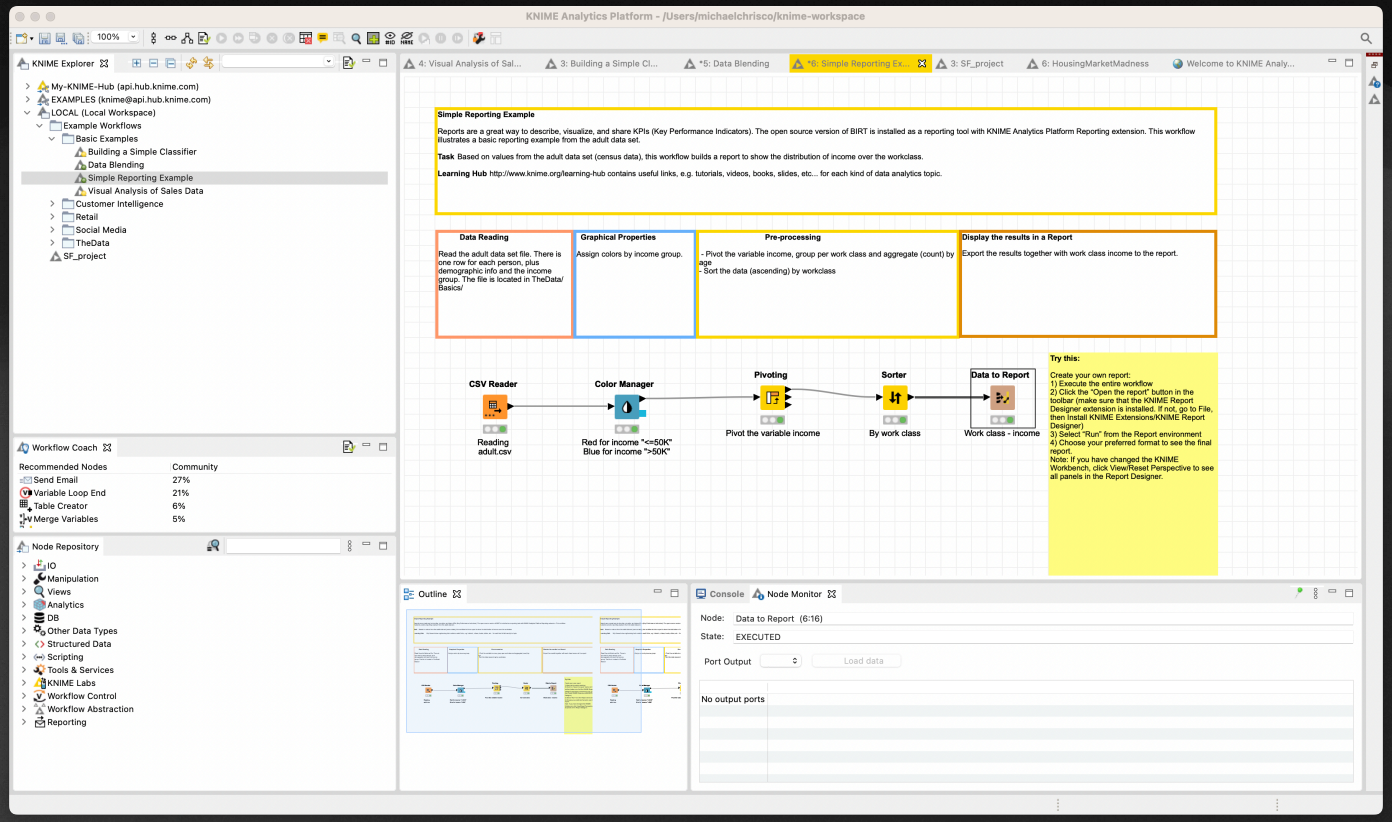


What does KNIME look like?



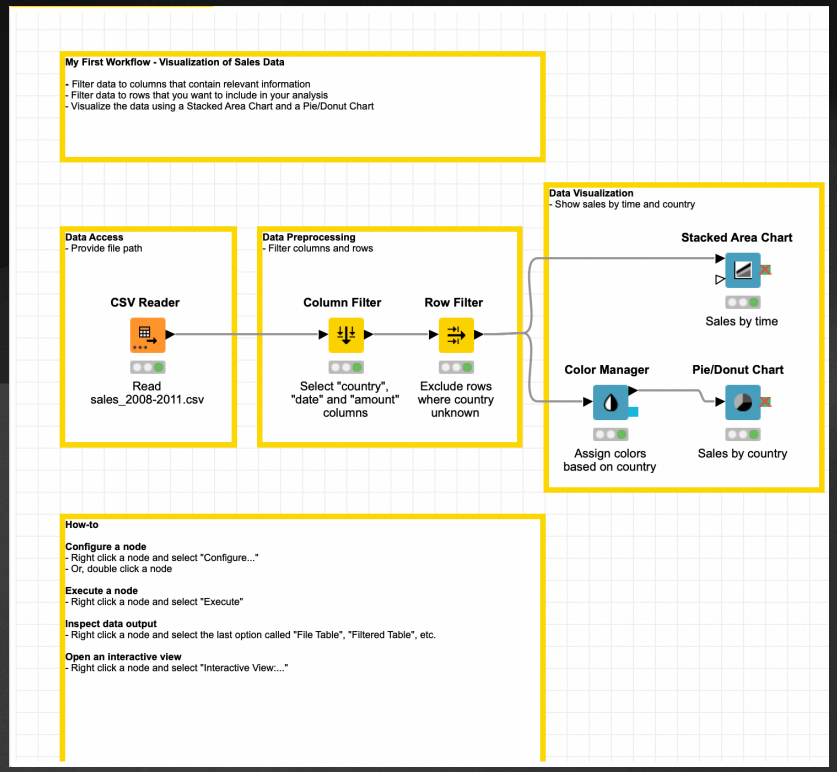
Example - Simple Reporting Example

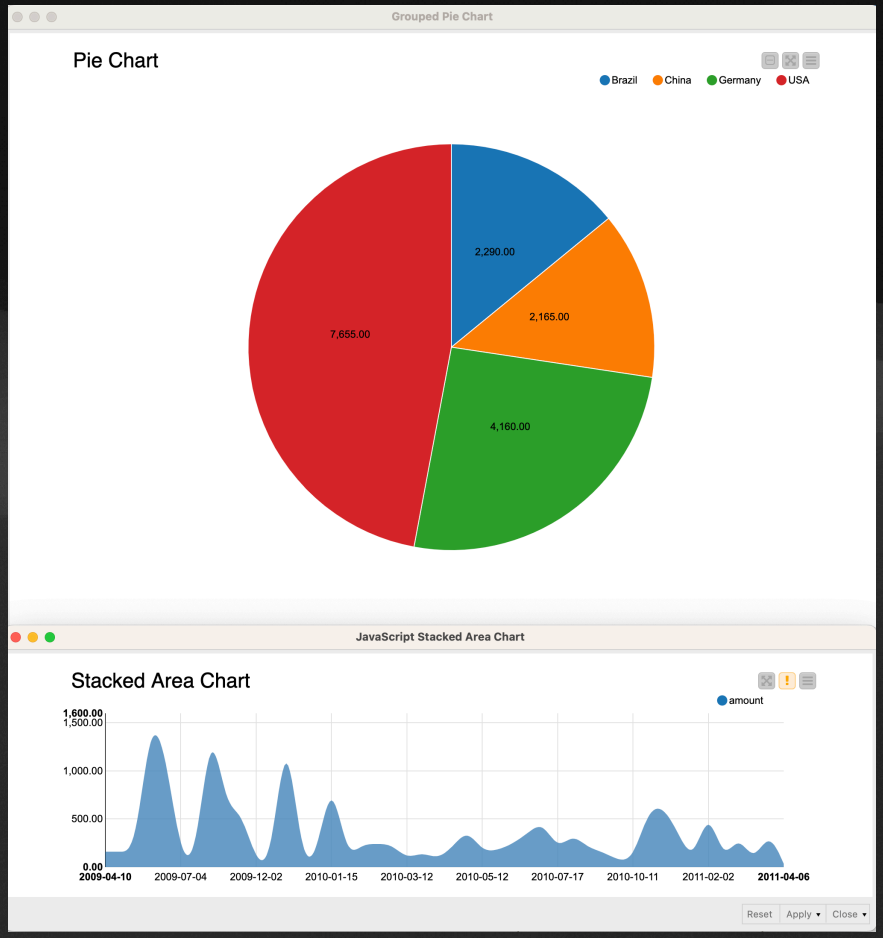
Single Input, multiple nodes, single Output



Example - Sales Data

Single input, filters, multiple outputs





Example - Big Data

Big Data Connectors

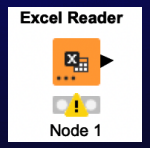
- KNIME Big Data Connectors allow easy access to Apache Hadoop data from within KNIME Analytics Platform and KNIME Server. This extension offers a set of KNIME nodes for accessing Hadoop/HDFS via Hive or Impala and ships with all required libraries.
- Move data between KNIME Analytics Platform or KNIME Server and Hive/ Impala
- Write Hive/Impala SQL queries using the standard KNIME Database Query node
- Process SQL queries directly in Hive and Impala using standard KNIME database nodes

KNIME - Real examples

- Top 3 Housing Regions with most drastic price changes in a year.
- Going over specific Medical Procedures Prices

Components of a KNIME workflow

Components


Excel Reader
Node 1

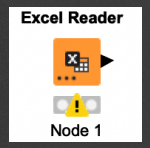

PostgreSQL Connector
Node 3

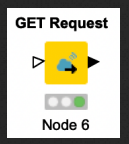

GET Request
Node 6

Workflow

Inputs

- File formats such as excel, css, json, xml, etc…
- ODBC (database connections) to PostgreSQL, MySQL, MSSQL, and many others.
- REST web services (GET/POST/PUT/PATCH/DELETE)
- Community made input nodes are also available





Workflow

Outputs

- Save within File formats such as excel, css, json, xml, etc…
- Push back into PostgreSQL, MySQL, MSSQL, and many others.
- REST web services (GET/POST/PUT/PATCH/DELETE)
- Interactive Graphs, figures, movies, etc…

Workflow

Business logic (the messy stuff in-between)

- Filtering, pivots, GroupBy, Joins, Unions, etc…
- Anything/everything SQL
- Full integrations with Java, Python, R, Tableau, Deep Learning, Apache spark, Cluster connections, and many more community made nodes.

Important links

- For more info: https://www.knime.com/
- Download now! https://www.knime.com/downloads
- Getting Started: https://www.knime.com/getting-started-guide
- Big Data Example: https://www.knime.com/knime-big-data-connectors
- Docs: https://docs.knime.com/


Questions?

Thanks for listening!