

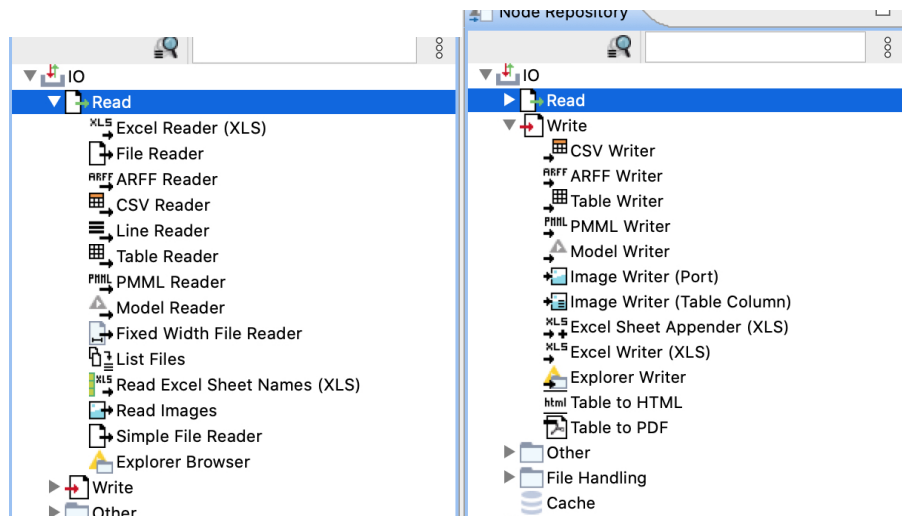
1. Objetivos

Uno de los primeros pasos en el proceso de análisis de datos es la visualización. En muchos casos, la visualización de datos es útil para explorarlos y crear gráficos. Por ejemplo, aquellos que se utilizan en los informes para describir los datos y el sistema subyacente. Knime proporciona muchos nodos para la visualización de datos, incluyendo diagramas de dispersión, gráficos circulares, gráficos de caja, histogramas, así como las nubes de etiquetas y visualizaciones de redes.

En esta práctica veremos el uso básico de Knime mediante la construcción de diferentes gráficas a partir de un conjunto de datos. El objetivo es familiarizarse con el programa y conocer algunos métodos simples de visualización de datos.

2. Nodos para entrada/salida de datos

Knime cuenta con una serie de nodos para entrada y salidas en la carpeta IO del repositorio de nodos.

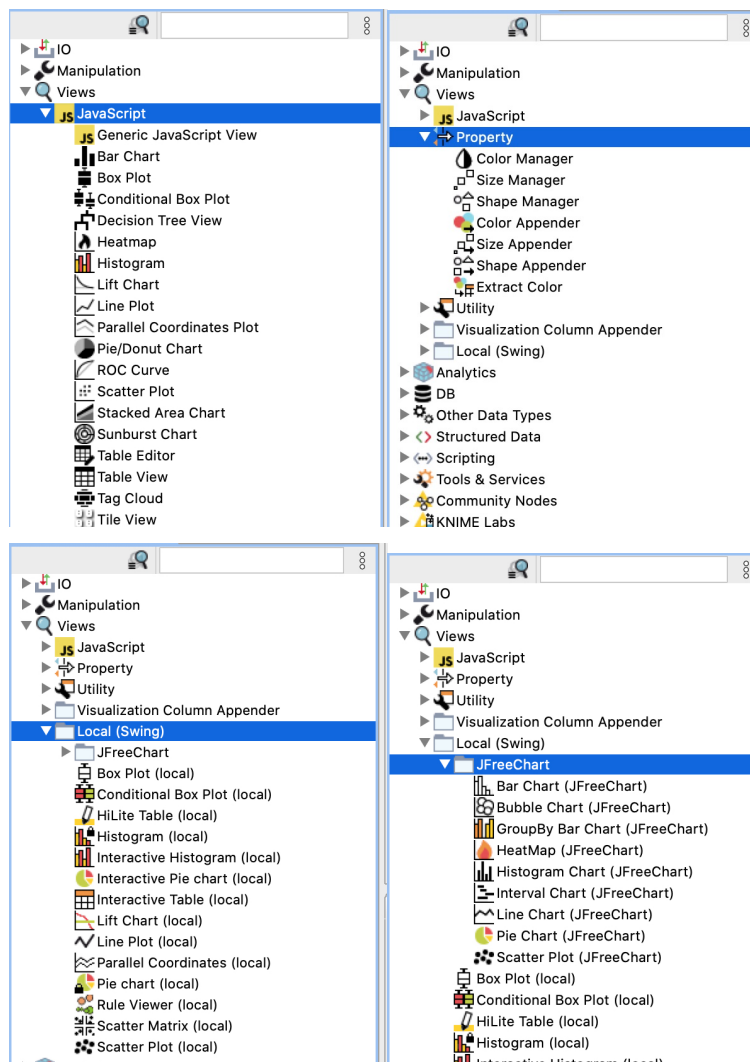
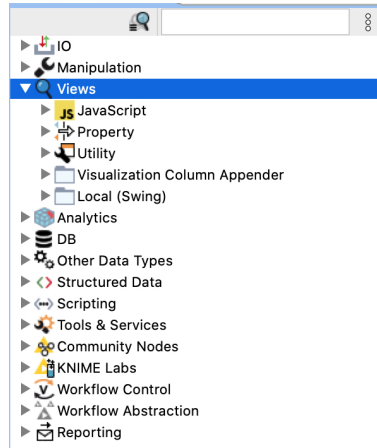


En general, para utilizarlos, simplemente se debe configurar el nodo añadiendo la ruta del archivo que contiene los datos; aunque dependiendo del tipo/formato de archivo se puede necesitar alguna configuración extra. Algunos de los nodos más comunes son:

- **File Reader.** Este nodo puede ser utilizado para leer datos de un archivo ASCII o de una ubicación URL. Se puede configurar para leer varios formatos. Cuando se abre el diálogo de configuración del nodo y se proporciona un nombre de archivo, este trata de adivinar el formato mediante el análisis del contenido del archivo. Una versión más sencilla se encuentra en el nodo **Simple File Reader**.
- **Excel Reader (XLS).** Este nodo lee una hoja de cálculo y la proporciona en su puerto de salida. Lee sólo los datos de una hoja (por defecto la primera). Puede leer sólo los datos numéricos (o strings), pero no diagramas, imágenes, u otros artículos. Knime soporta actualmente datos de tipo String, Double, e Int. La lectura de archivos de gran tamaño necesita mucho tiempo y utiliza una gran cantidad de memoria (especialmente los archivos en formato .xlsx).

3. Nodos de visualización

Knime proporciona muchos nodos para la visualización de datos, incluyendo diagramas de dispersión, gráficos circulares, gráficos de caja, histogramas, así como las nubes de etiquetas y visualizaciones de redes. Los nodos clásicos están disponibles en la carpeta **Views**.



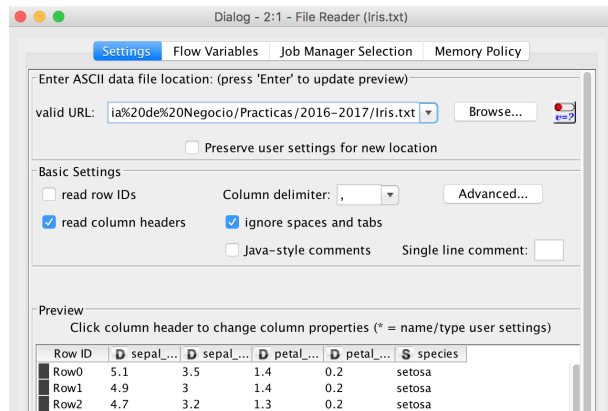
Un ejemplo de uso:

1. Crea un nuevo flujo de trabajo y carga el conjunto de datos del archivo `iris.txt`.

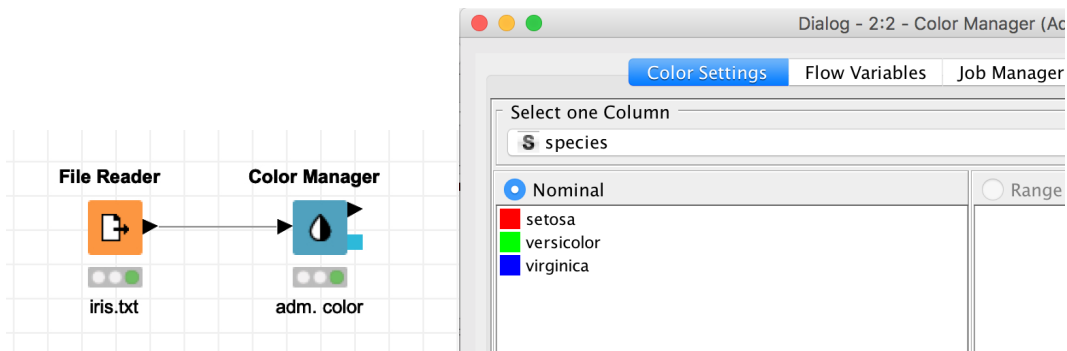
File Reader



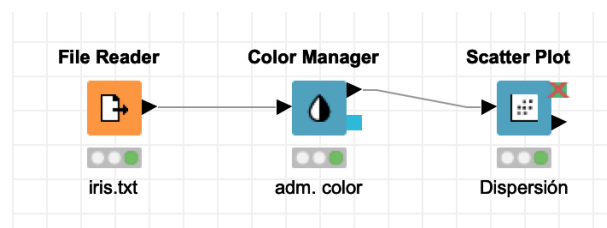
iris.txt

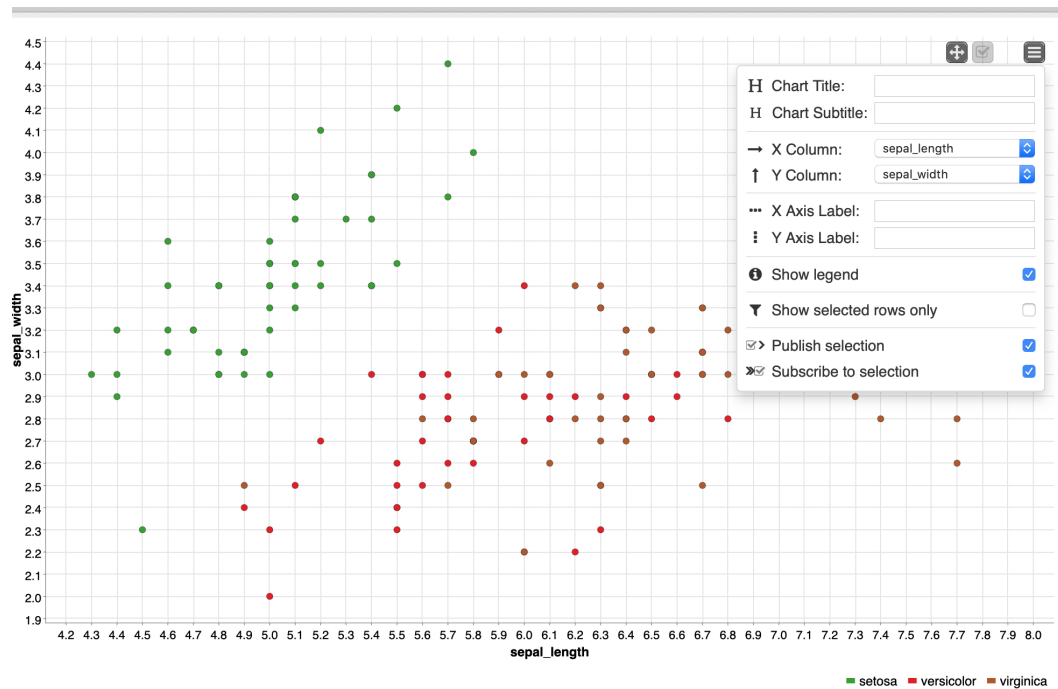


2. Anexar colores basados en la columna de la especie mediante el Administrador de color (hay que ejecutar antes el nodo de lectura del fichero).



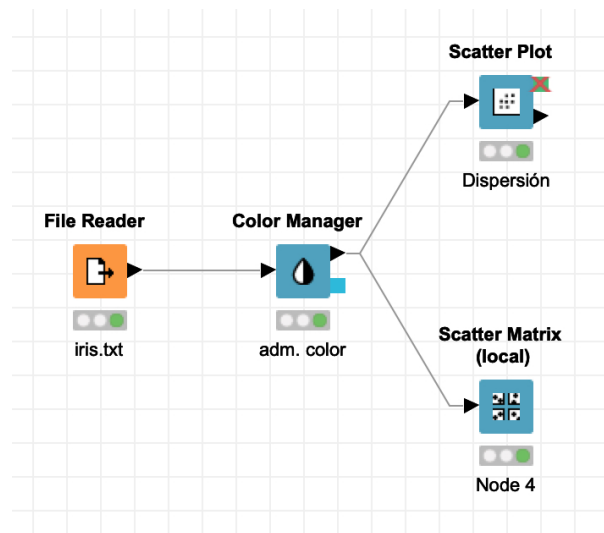
3. Visualizar los datos utilizando un gráfico de dispersión **Scatter Plot** en Views/JavaScript.

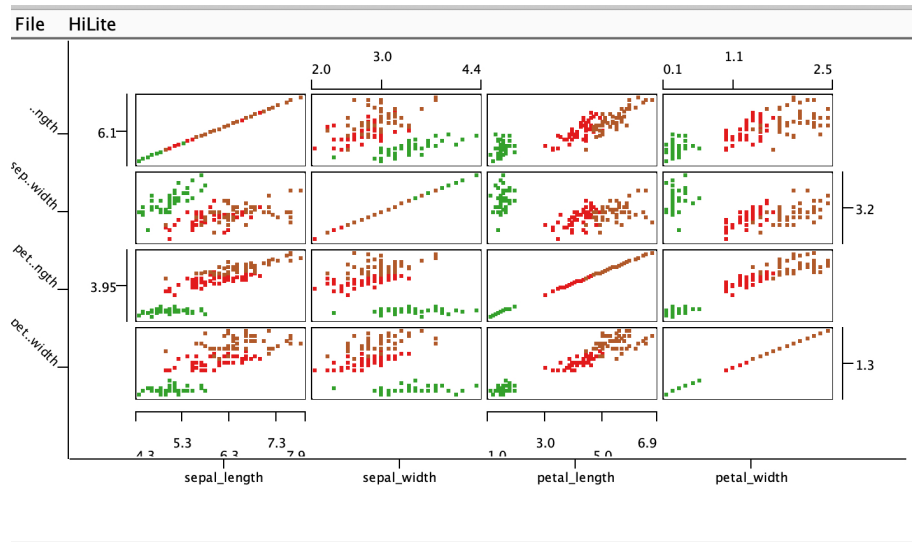




Con el botón en la esquina superior izquierda del diagrama de dispersión podemos cambiar la configuración del diagrama.

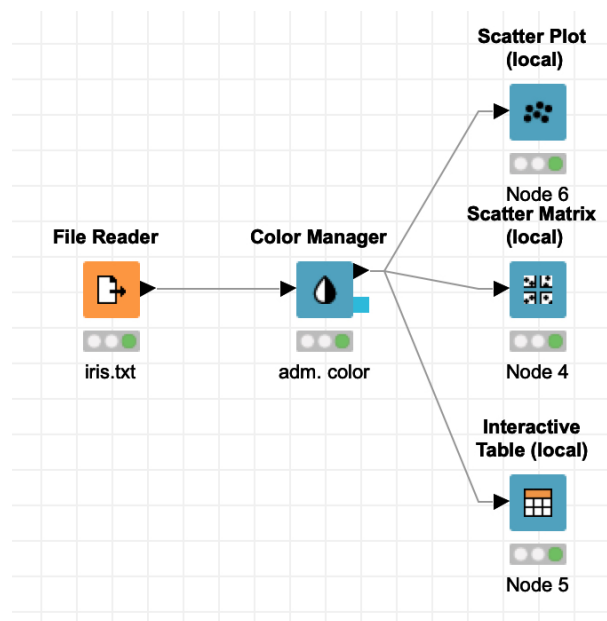
- Visualizar los datos utilizando una matriz de dispersión Scatter Matrix.



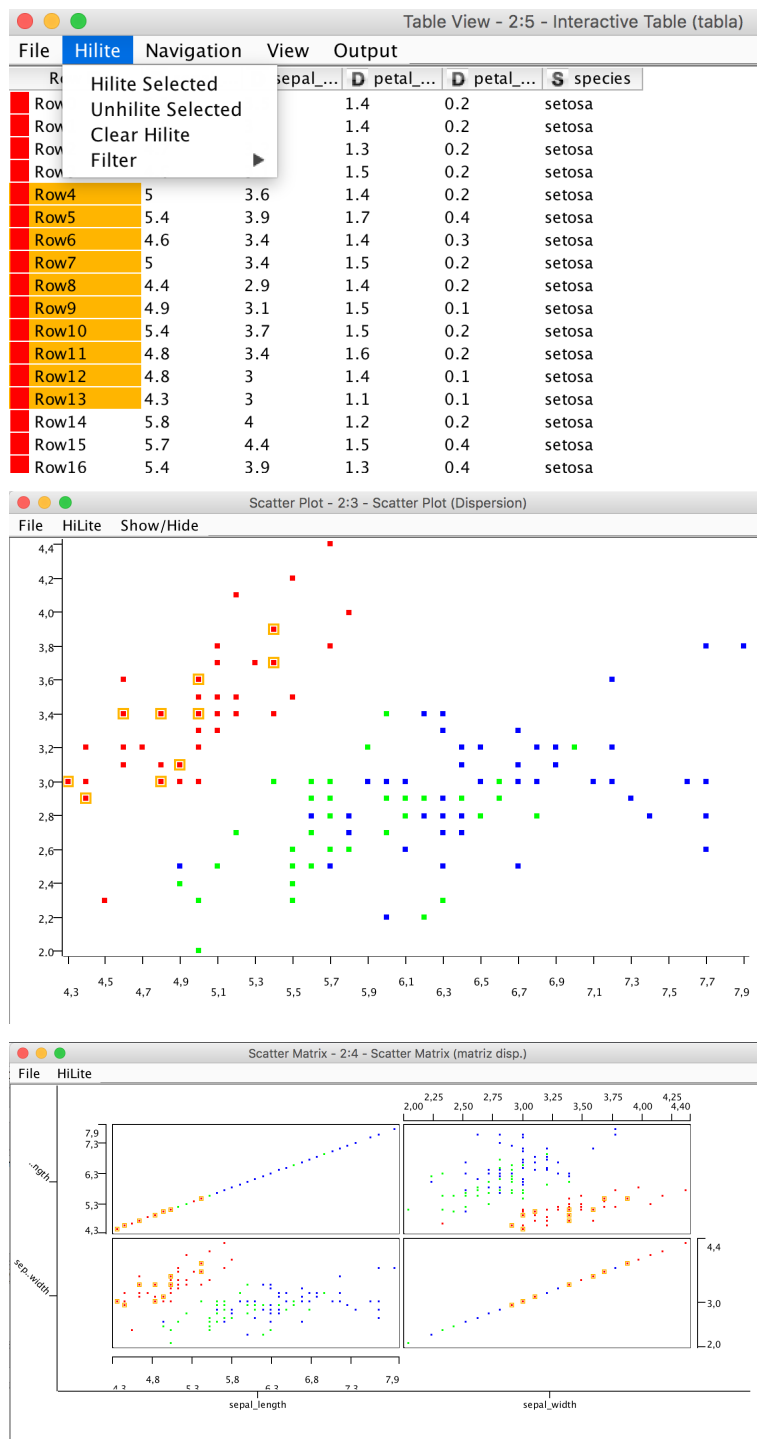


Añade todos los atributos en la matrix, ¿puedes deducir algo de la visualización de las matrices?

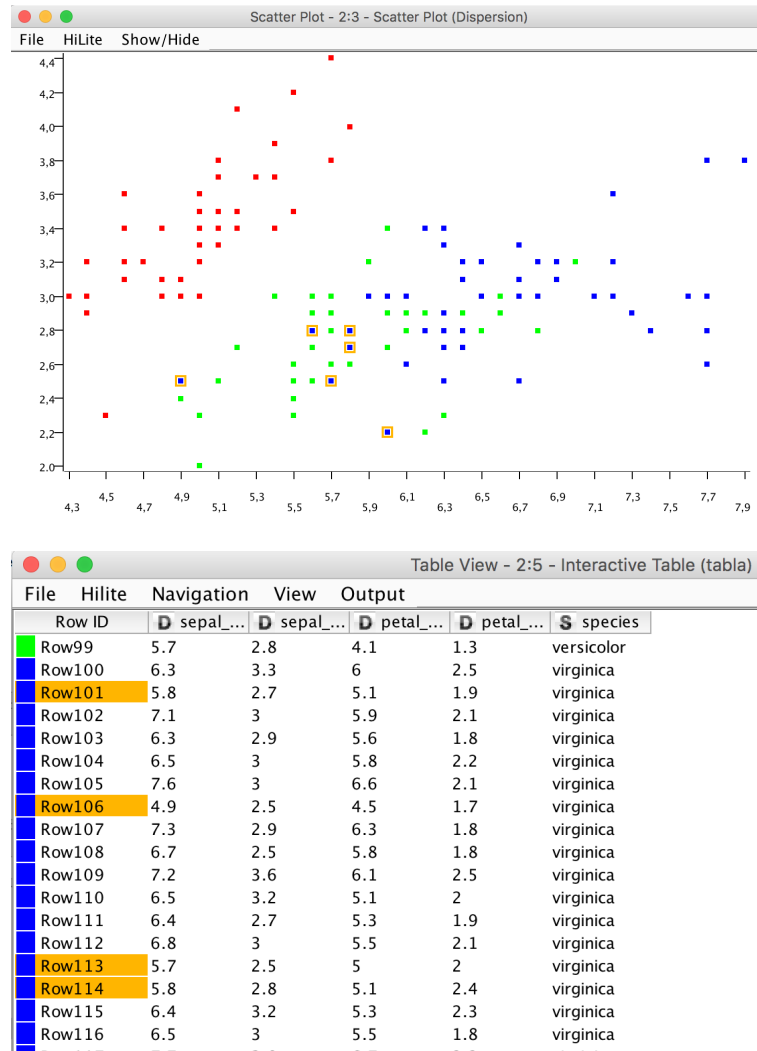
- Visualizar los datos utilizando el nodo **Interactive Table**, utilizamos también **Scatter Plot** en Views/Local.



- Selecciona algunos puntos en la tabla interactiva (Hilite), observa como se muestran en el gráfico de dispersión y en la matriz de dispersión.



- Realiza el proceso contrario, selecciona puntos en el gráfico de dispersión y observa qué puntos son en la tabla interactiva.



También puedes encontrar más nodos en la carpeta **KNIME Labs/JavaScript Views (Labs)** Para realizar estadísticas básicas (media, desviación típica, máximo, mínimo, etc) puedes utilizar los nodos en la carpeta **Analytics/Statistics** y **KNIME Labs/Statistics**.

4. Tareas a realizar

Si los datos no provienen de un fichero, insértalos manualmente en un uno para la lectura desde algún nodo de Knime (utiliza el formato que prefieras).

1. Se considera el siguiente registro de llamadas a una oficina

Hora	Día 1	Día 2	Día 3	Día 4	Día 5
9 : 00	23	12	34	34	23
9 : 30	12	12	23	21	23
10 : 00	43	12	15	20	23
10 : 30	54	23	65	21	23
11 : 00	12	23	78	24	12
12 : 00	1	23	45	57	56
13 : 00	13	23	23	56	34
13 : 30	0	12	41	60	34

6. Considera los datos de la tabla `prestamo.xls`. Realiza un histograma para representar el número de personas que tienen tarjeta de crédito y que no. Realiza un diagrama de burbujas para analizar el número de miembros de la familia de las personas recogidas en la tabla. Calcula las estadísticas básicas de las variables. Realiza una gráfica de dispersión para analizar visualmente los campos edad y años de experiencia. Estudia las matrices de dispersión de algunas variables. ¿Algunos atributos parecen relacionados? ¿Existen personas con algún atributo, o relación entre atributos, que se salga fuera de lo común? Estudia la matriz de correlación entre dichas variables.