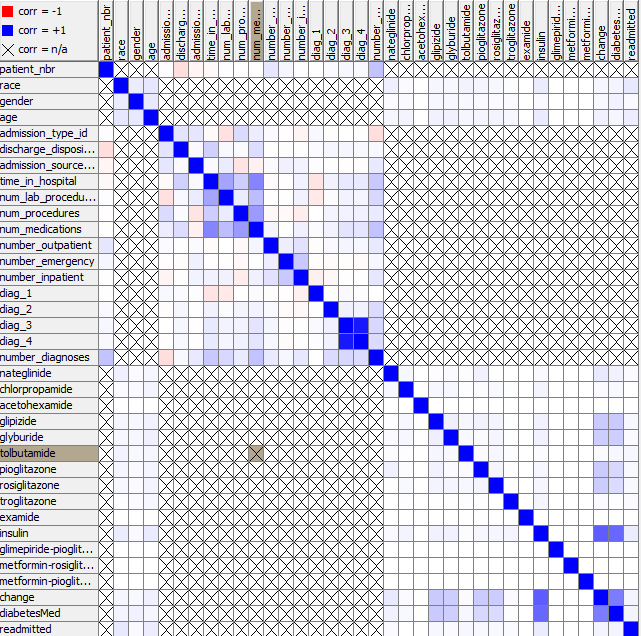
# Correlación de las variables

Lo vamos a hacer con KNIME.  
Al nodo FileReader le unimos el nodo LinearCorrelation, a través del cual podemos ver la siguiente matriz de correlación. En color azul oscuro salen los valores de correlación 1.



Podemos observar que haciendo un test de correlación entre las variables del dataset las variables diag\_3 y diag\_4 tienen una corrección de 1.  
Existen otras variables con un significativo valor de correlación, como son insuline y change, con un valor de 0,6402.

# Interpretación de los datos

Tenemos 37 variables y 5000 entradas en el dataset analizadas en la siguiente tabla:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nombre de la variable** | **Descripción** | **Tipo de variable** | **Valor mínimo** | **Valor máximo** | **Nº de nulos** |
| X1 | Identificador | Numérica | 0 | 4999 | 0 |
| patient\_nbr |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Búsqueda de nulos:

Variable : tolbutamide = ¿, admission\_type\_id= 6 y discharge\_disposition\_id= 18 y admission\_source\_id = 17

nulTolbu <- DataSet\_Group\_9\_O$tolbutamide[DataSet\_Group\_9\_O$tolbutamide == '?'] 4011 nulos de 5000 totales

nulAdmin -> 283  
nulDIschat -> 175  
nulASource ->334