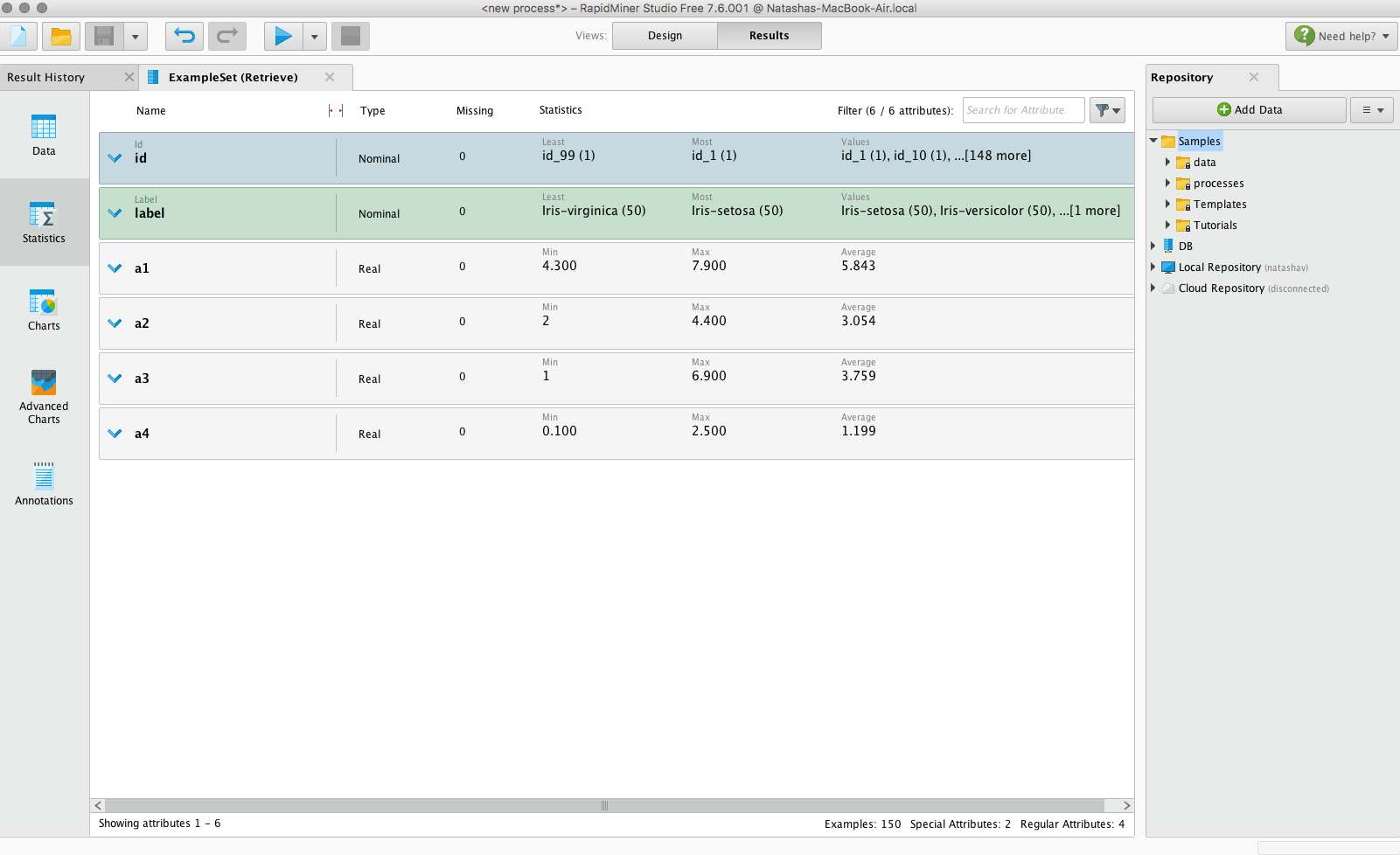
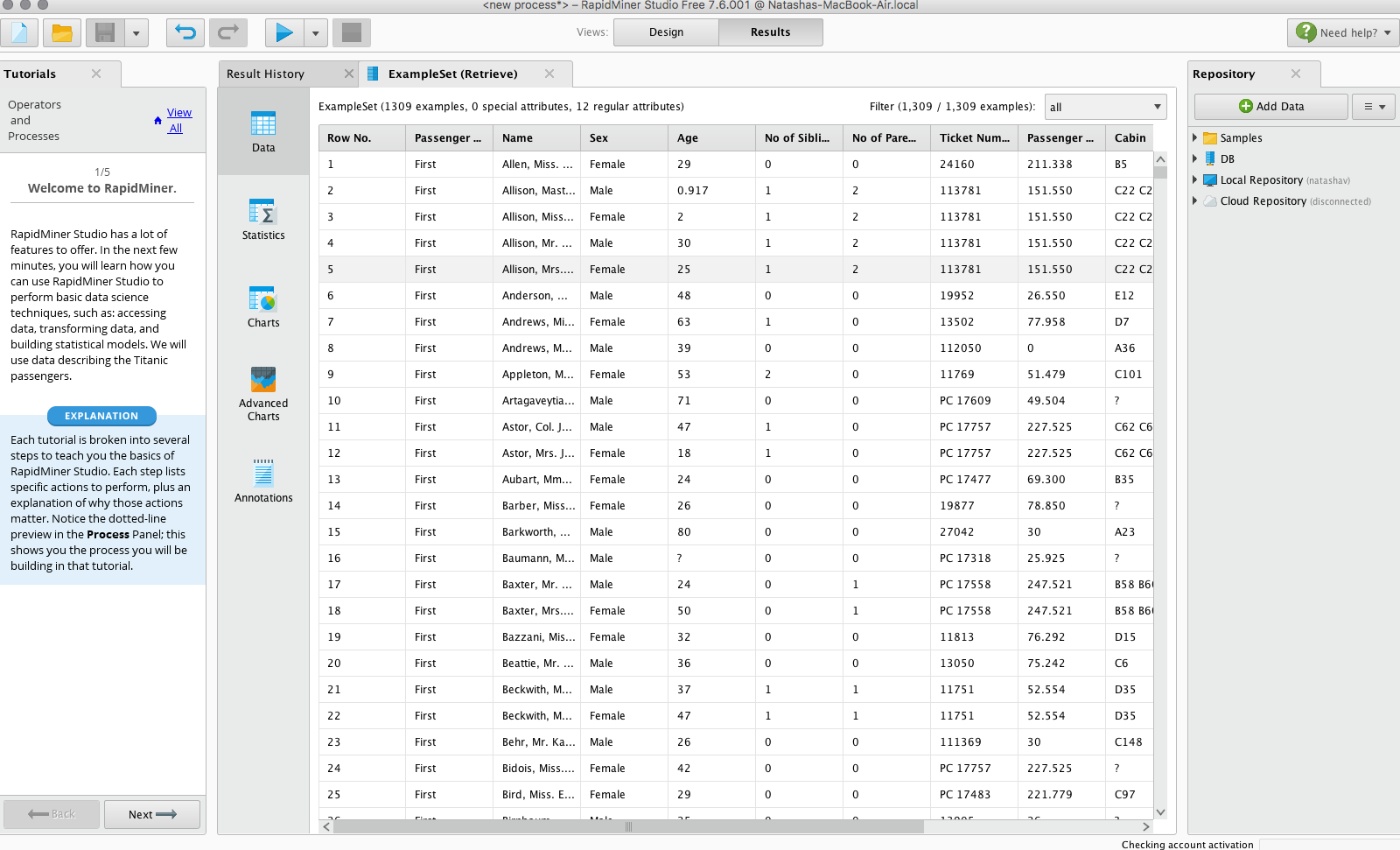
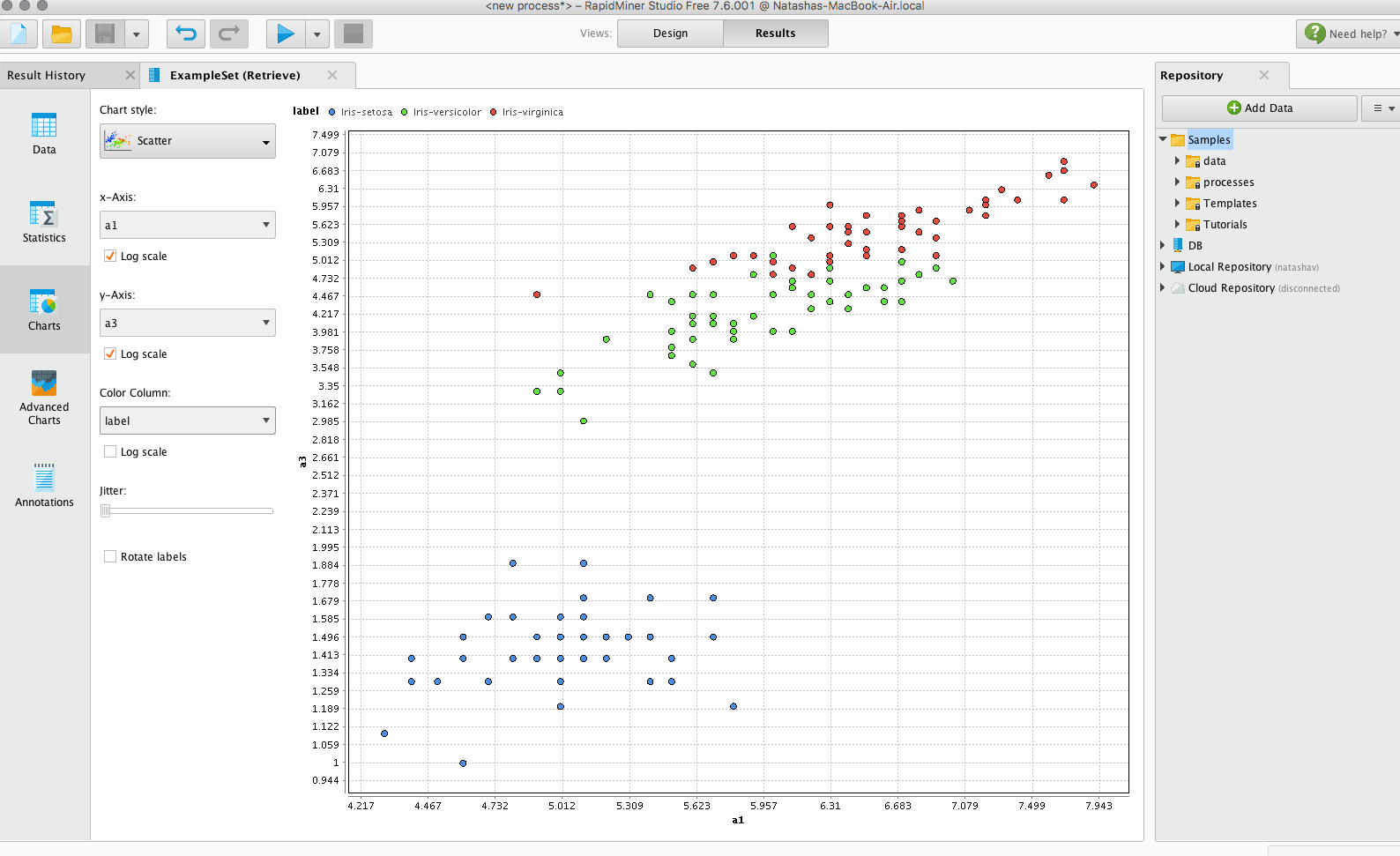
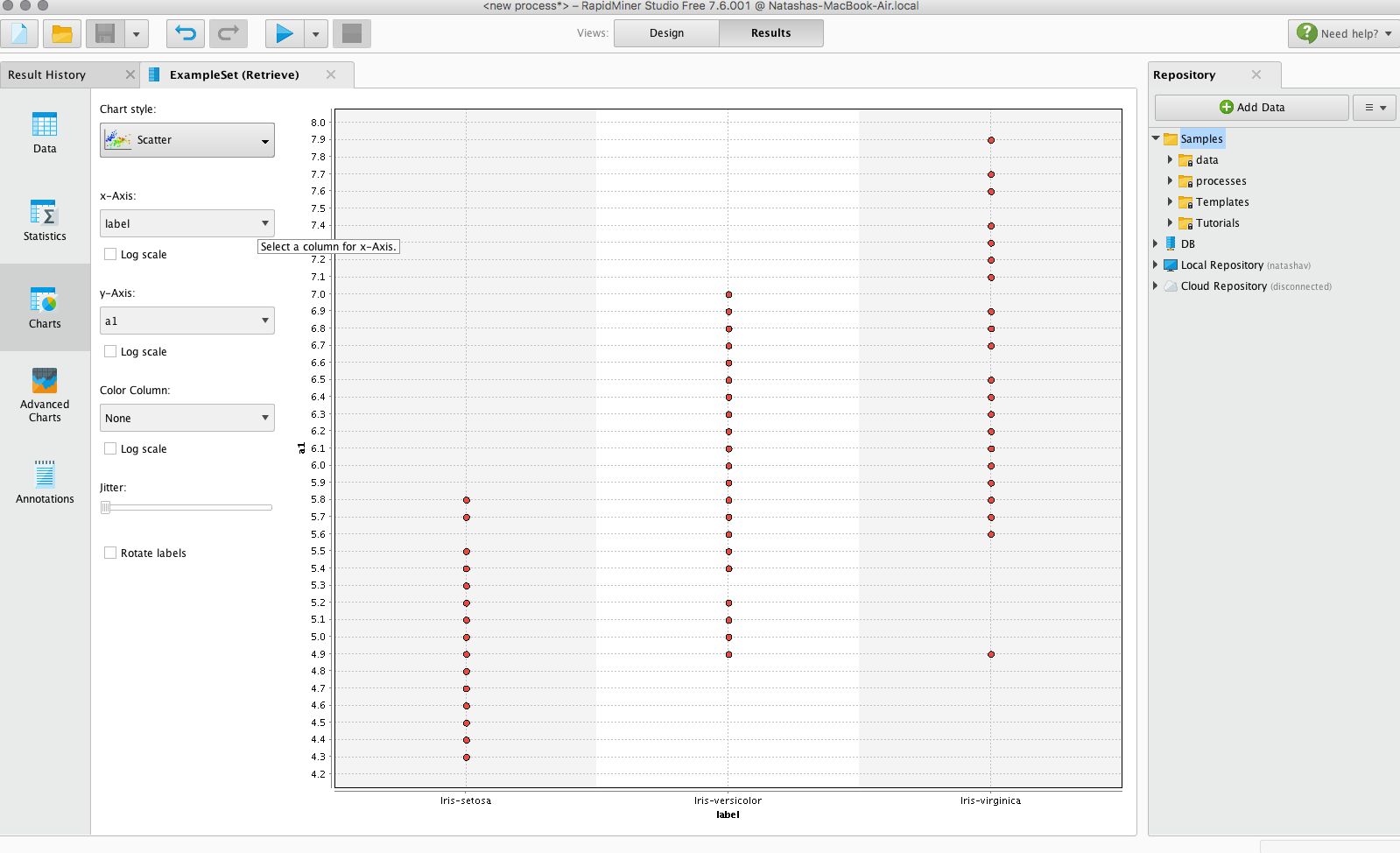
Lab 8

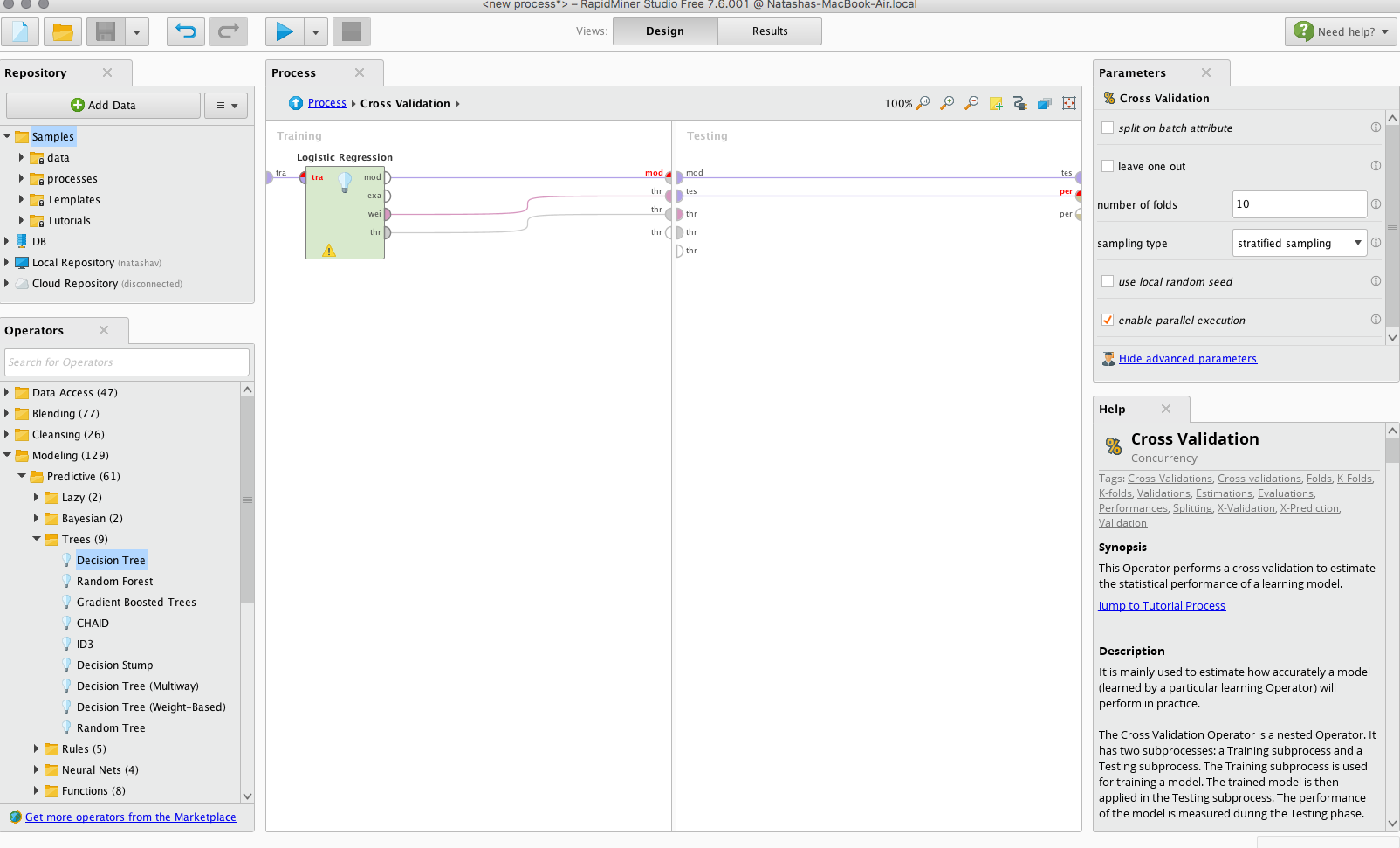
In downloading and using RapidMiner, it was clear that it is very similar to KNIME, with a somewhat cleaner GUI. Thus, I was very excited to get started using it. Unfortunately, it seems that the RapidMiner website has undergone some re-design and it was not very straightforward to find the introduction videos. Additionally, some introduction videos stated that there were links to data so that users could follow along with the video, but there were no links to data. Thus, I followed along using different, pre-uploaded data to follow along. Below are some screenshots of me following along with the videos to get an idea of how to use RapidMiner.

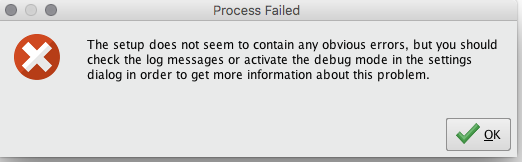
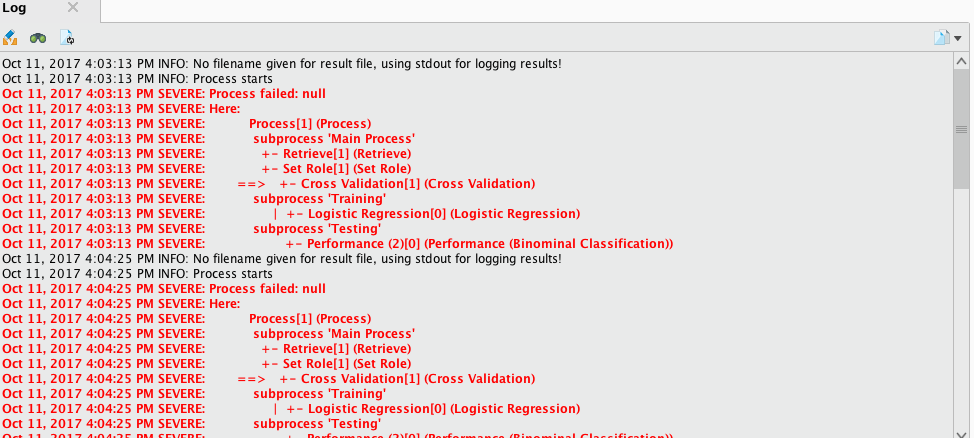


Below are screenshots I took while following along with how to use the built-in data visualization capabilities.

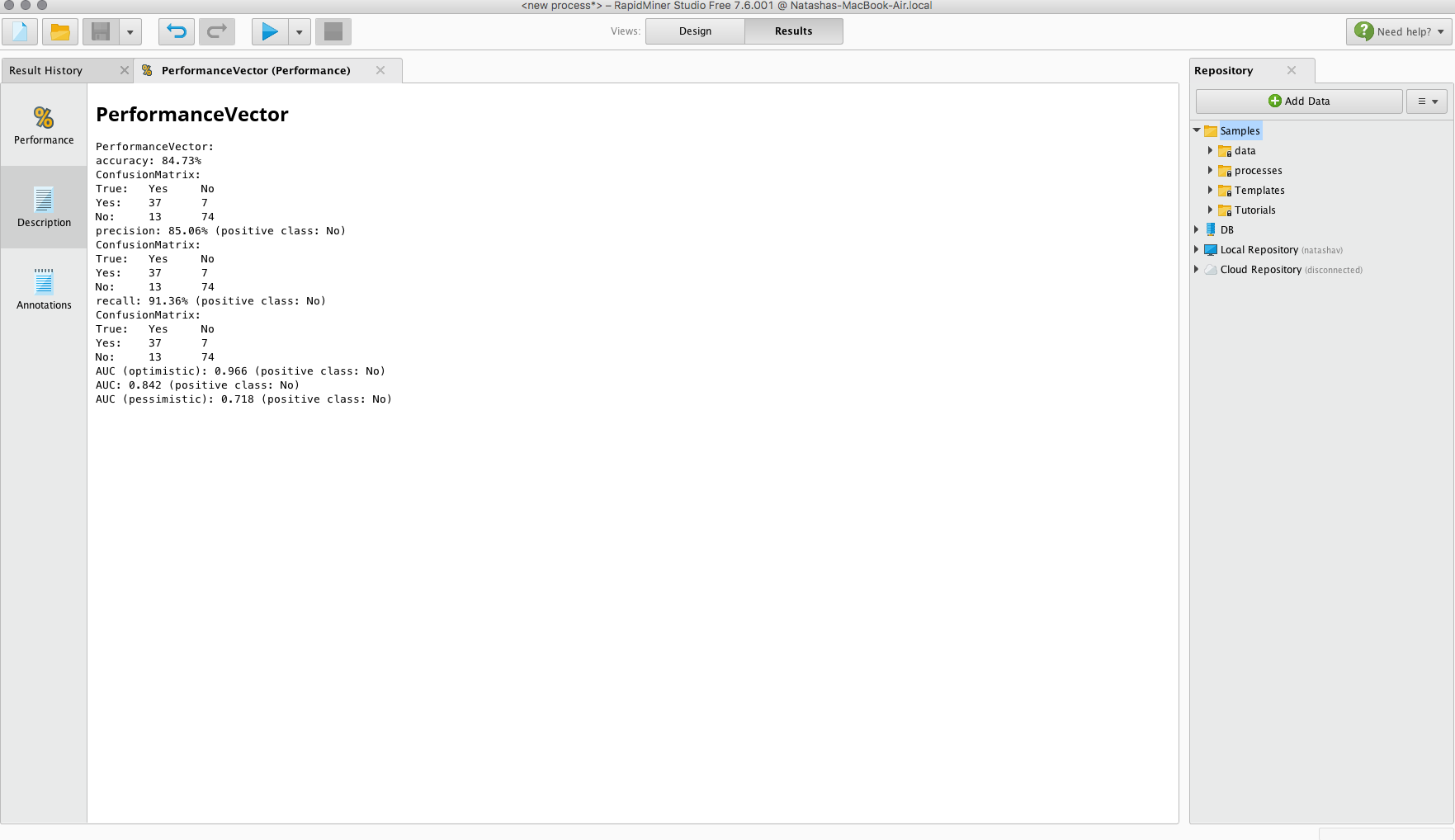
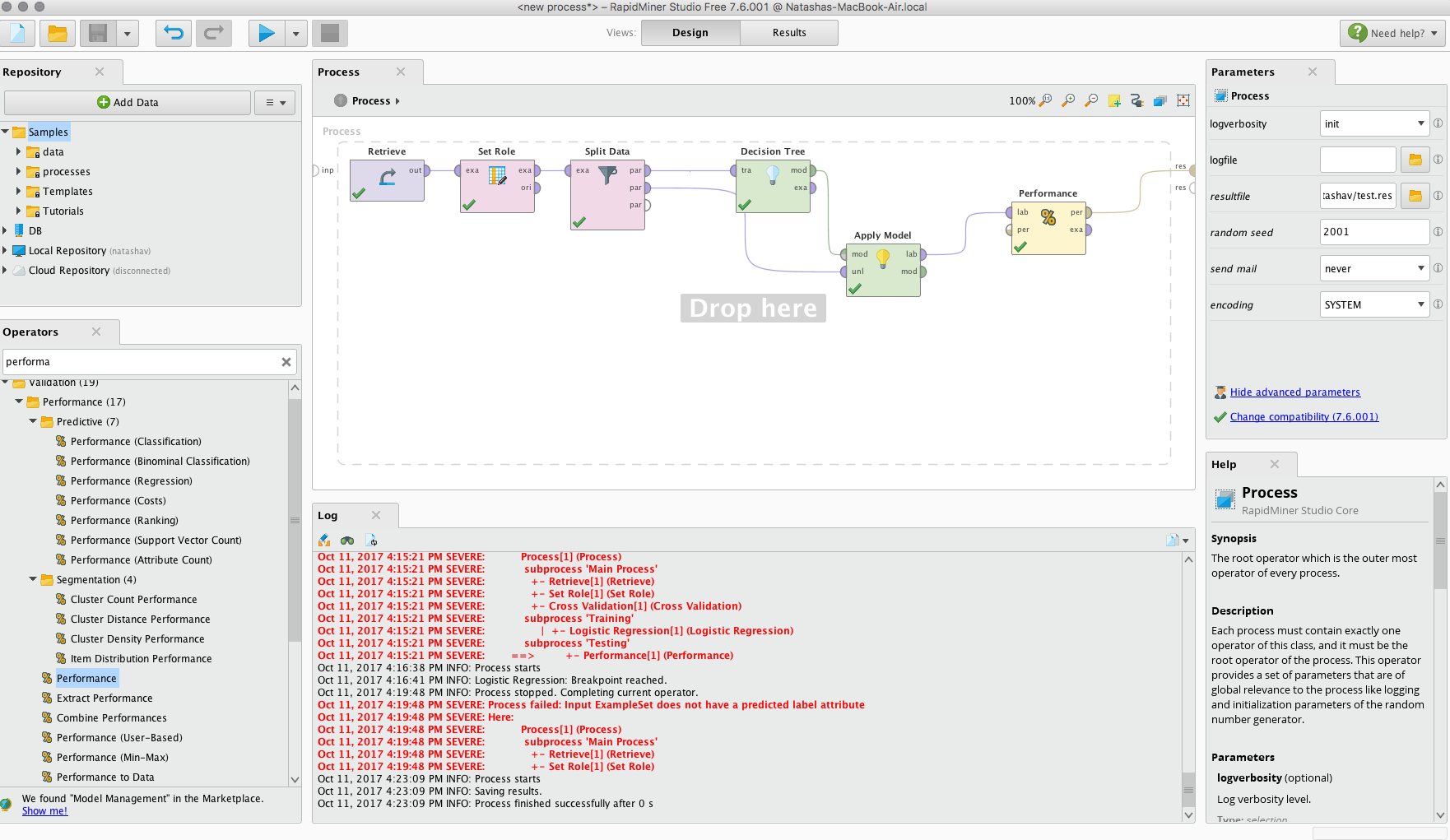


After following along with the videos, I wanted to try my hand at doing my own mini-project in RapidMiner. My goal was to build a model to predict who would survive the Titanic (a classic introductory data analytics project). I attempted to use the cross-validation technique to train and test the model. RapidMiner has a built-in module to help set up cross-validation model-building. However, the instructions/parameters on how to use the model were completely unclear, and RapidMiner did not have a good method for specifying why a module isn’t working. In the screenshot below, you can see that the module splits up the training and testing of the model. However, it was unclear when you added the cross-validation module that you needed to add modules in here, and it wasn’t clear what types of modules you needed! Further, in the next screenshot, you can see the completely useless error message and error log that RapidMiner was giving me even when I turned on the debug mode.





So, I gave up on the cross-validation method and decided to use a module to split the data, train on the first split and test on the second split. The setup of that model and the performance of that model are featured in the screenshots below.



Ultimately, though RapidMiner and KNIME seem very similar, I think I prefer KNIME thus far due to the lack of clarity and lack of error handling provided by RapidMiner. Thus, I would need to learn more about the specific capabilities of RapidMiner if I were to switch over from using KNIME.