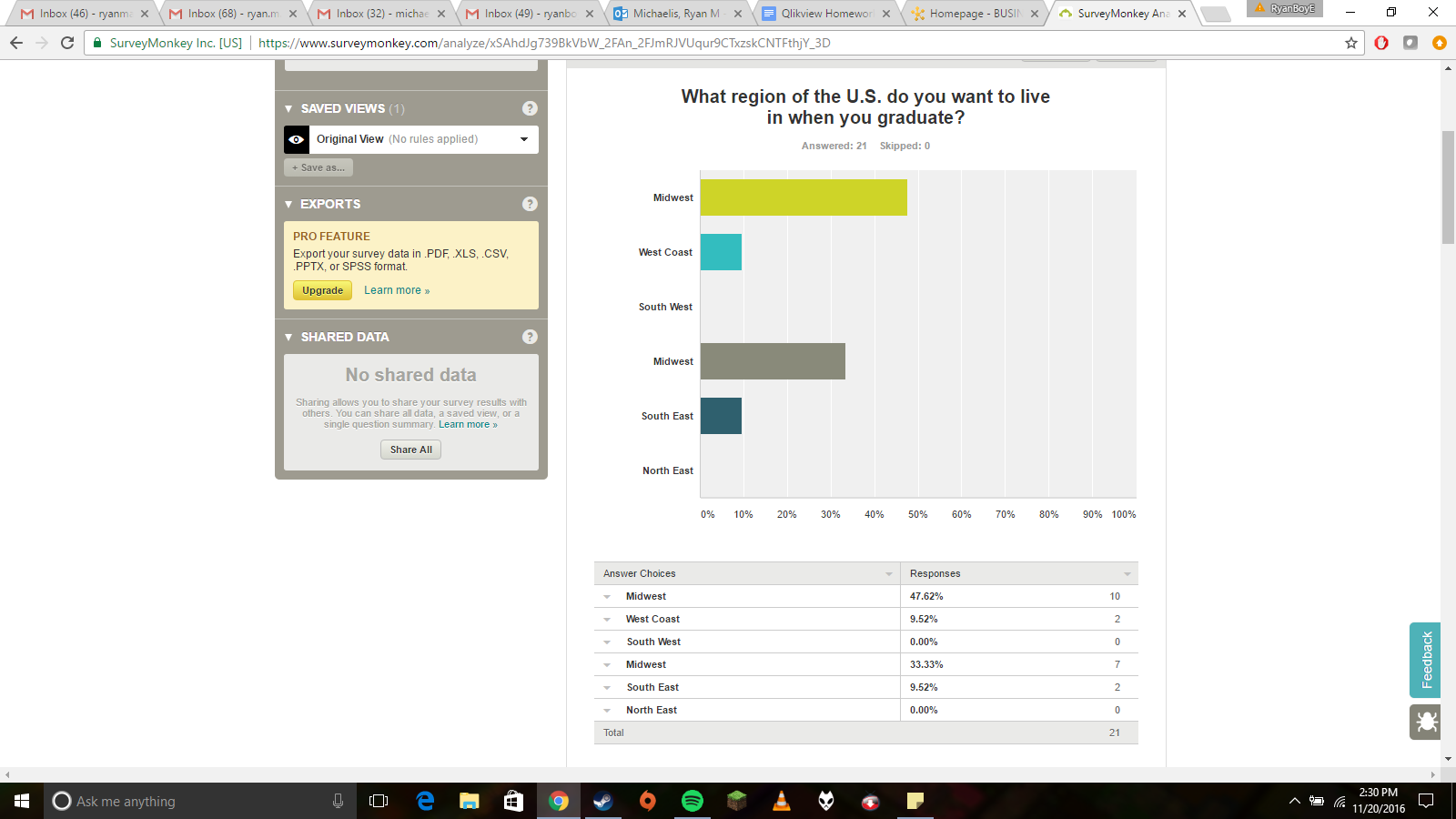
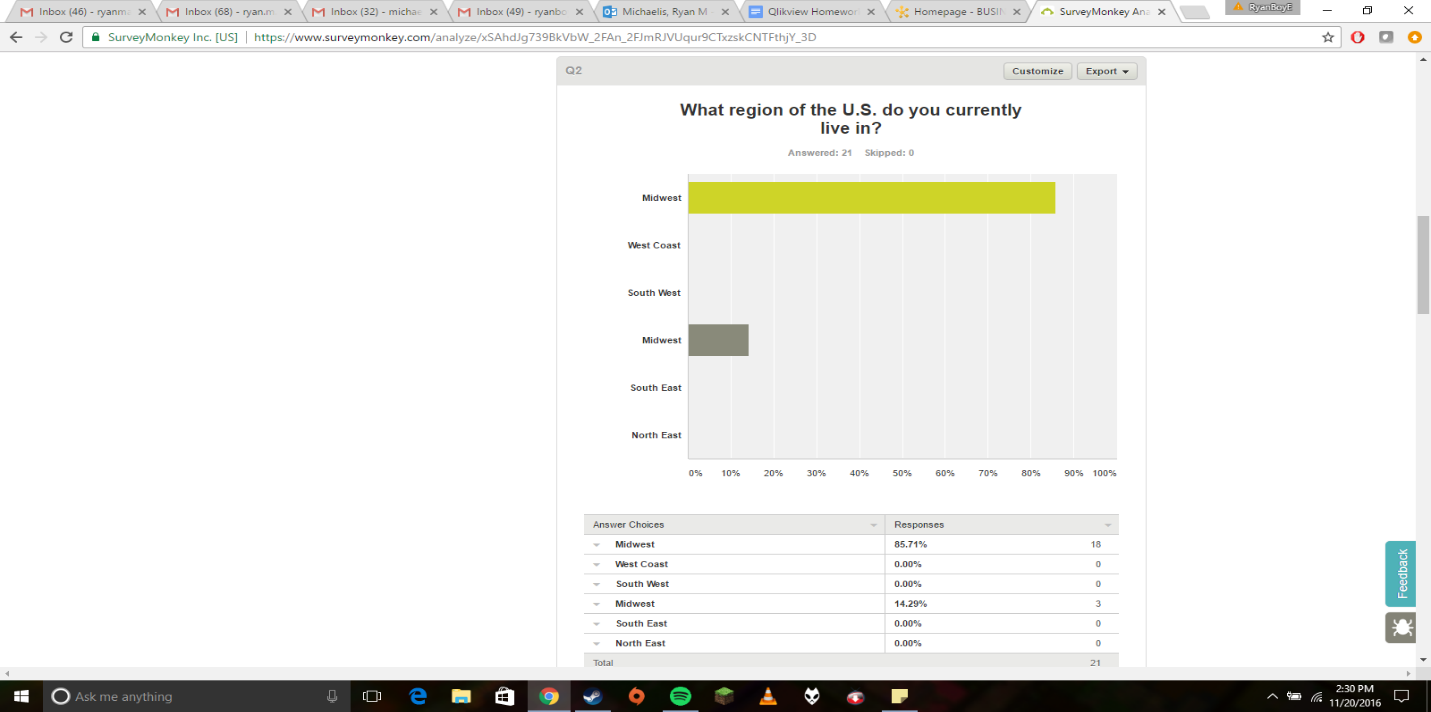
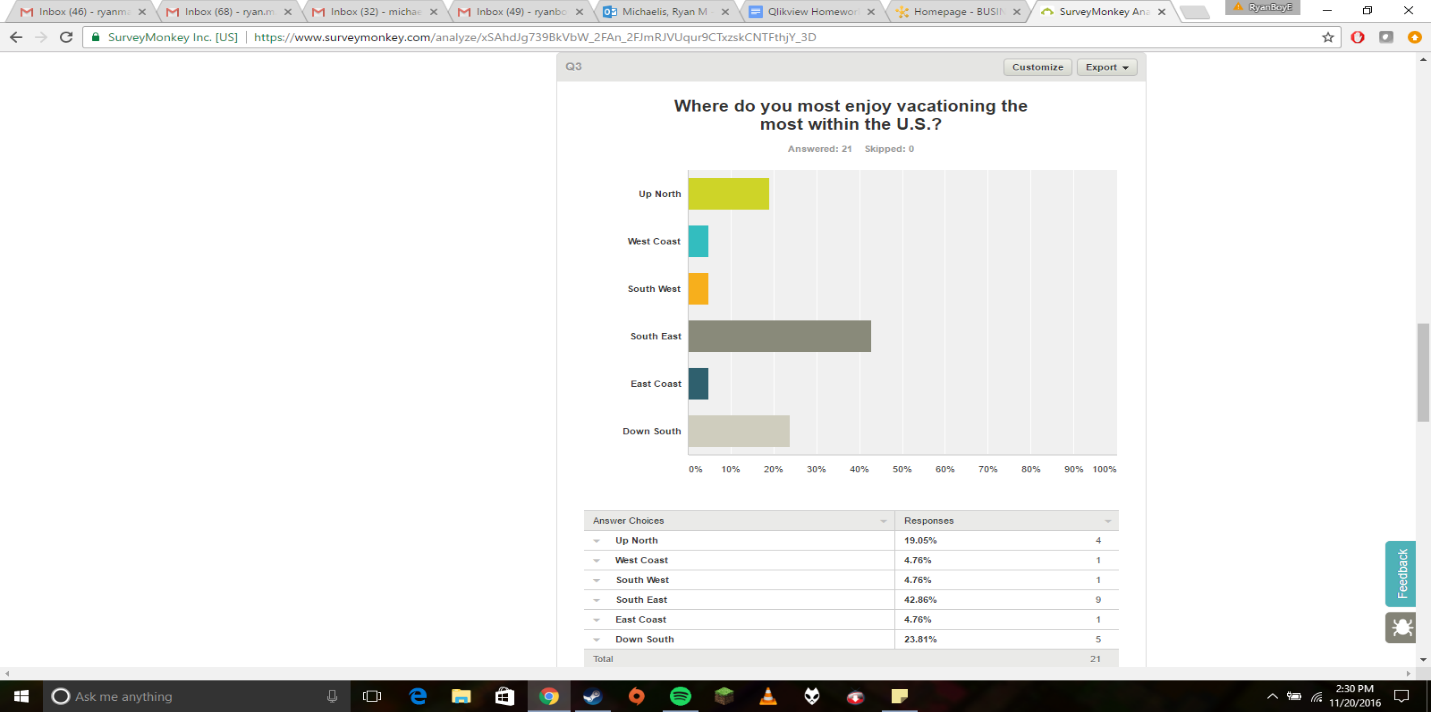
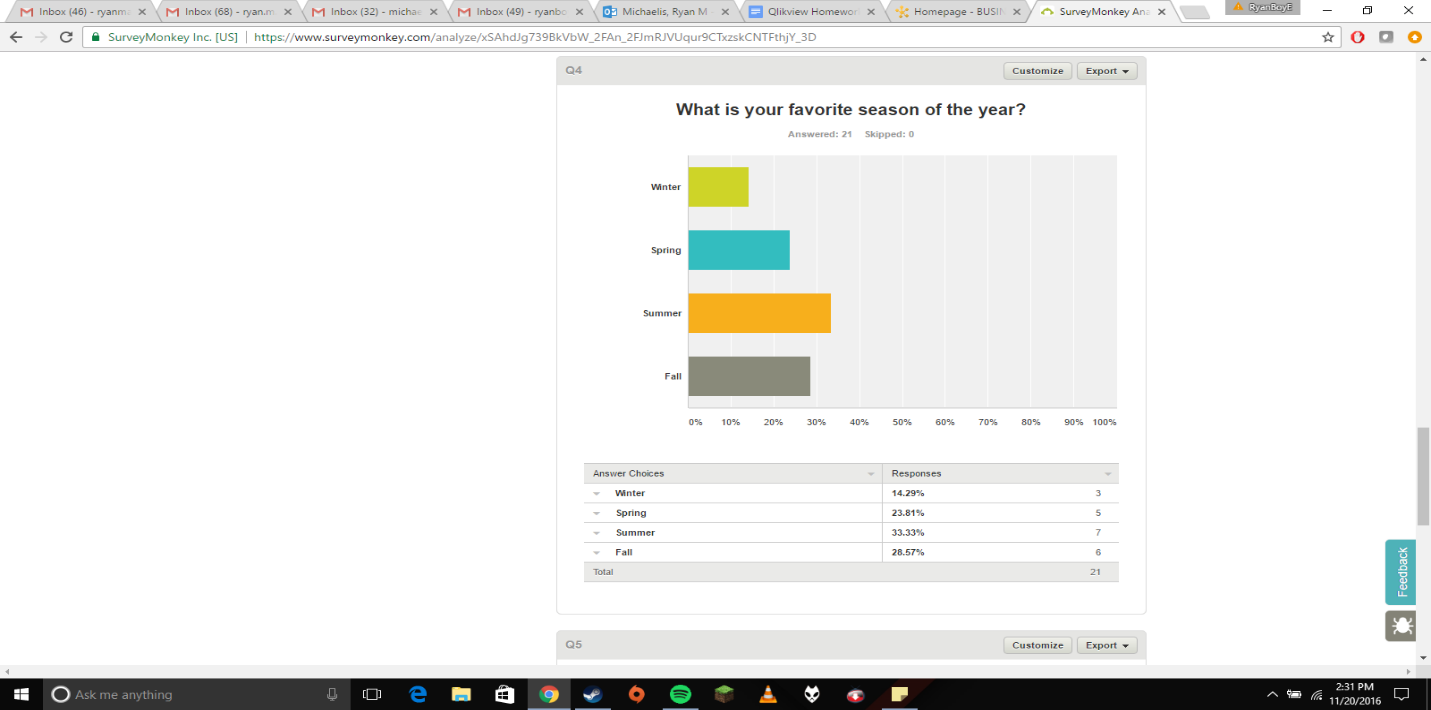
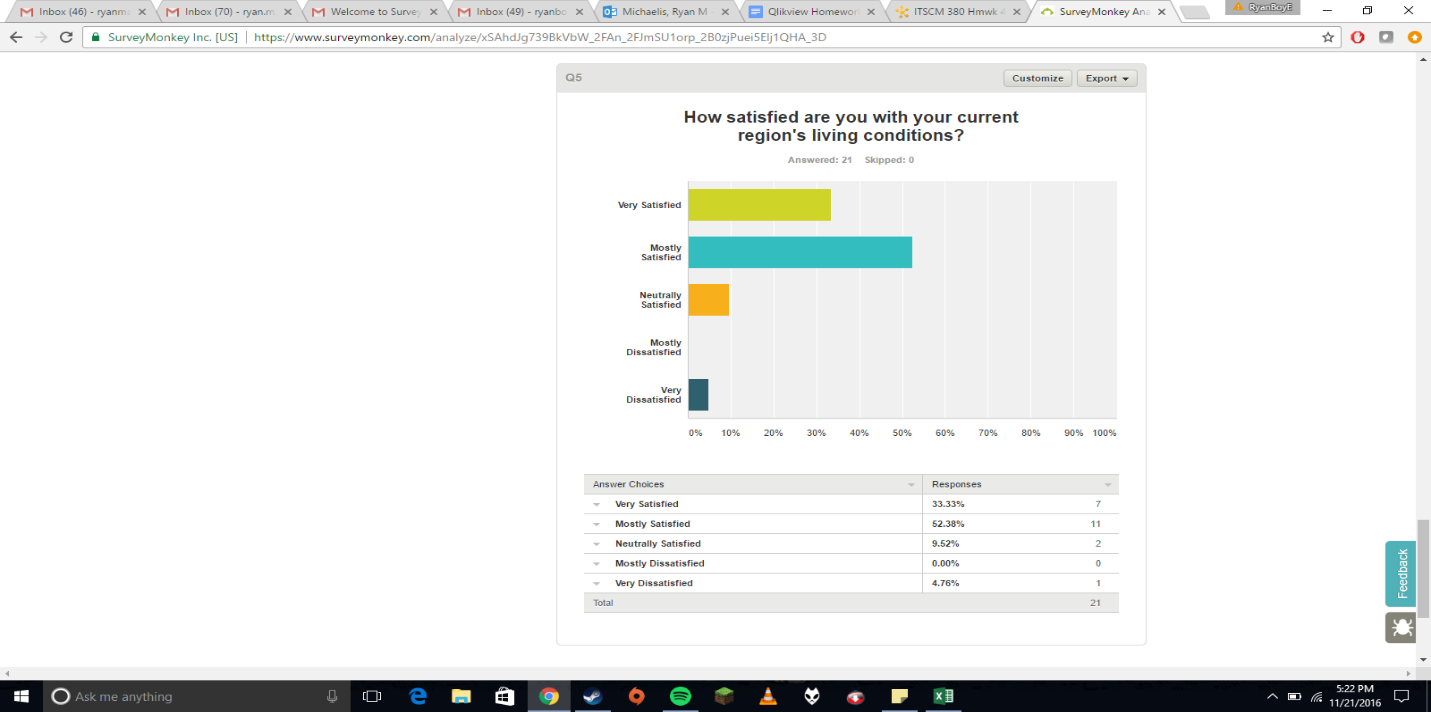
1. After going through Chapter 3, I feel more comfortable uploading data sets, calling them up, adding and connecting multiple operators, filtering values or attributes, and viewing results. Very helpful introduction for basic skills.
2. Chapters 5 and 6 taught me how to use “Select Attributes” operators to pinpoint useful data, “Numerical to Binominal” operators to make future organization of data easier, “Create Association Rules” operators to help interpret massive amounts of results while altering confidence amounts, and “Cluster” operators to filter and form data to better answer questions related to them.
3. Considering the datasets from the last homework assignment, I feel as though not a lot of preprocessing is required. All questions were required to be answered for a final entry – so no blank values, the options were straight forward and answers were finite – so no value corrections needed. I suppose I should’ve either used Google forms or signed up for SurveyMonkey premium to get the actual spreadsheets for easy importing. Other than that, I’d just use the “Create Association Rules” operators to determine the confidences of associations between all values from each of the questions and would most likely start with testing our hypothesis of a majority of respondents anticipating to stay in the Midwest, yet leaving for some period of time during the winter months. Survey Results Below:

****

****

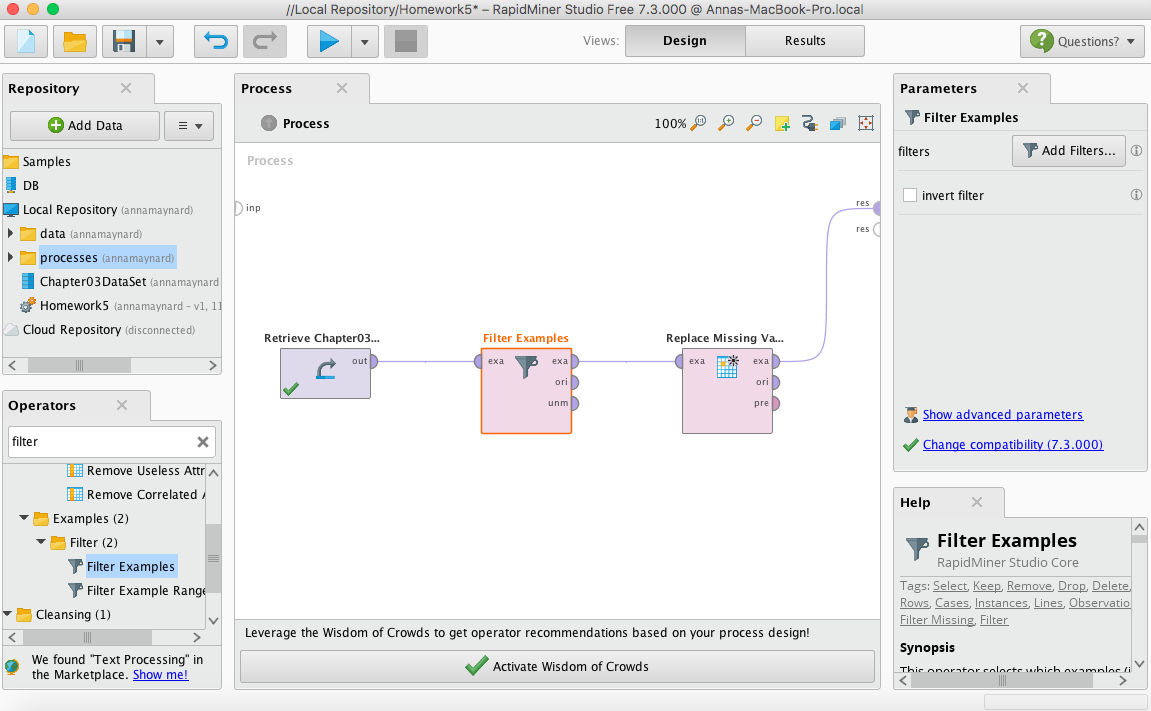
****

****

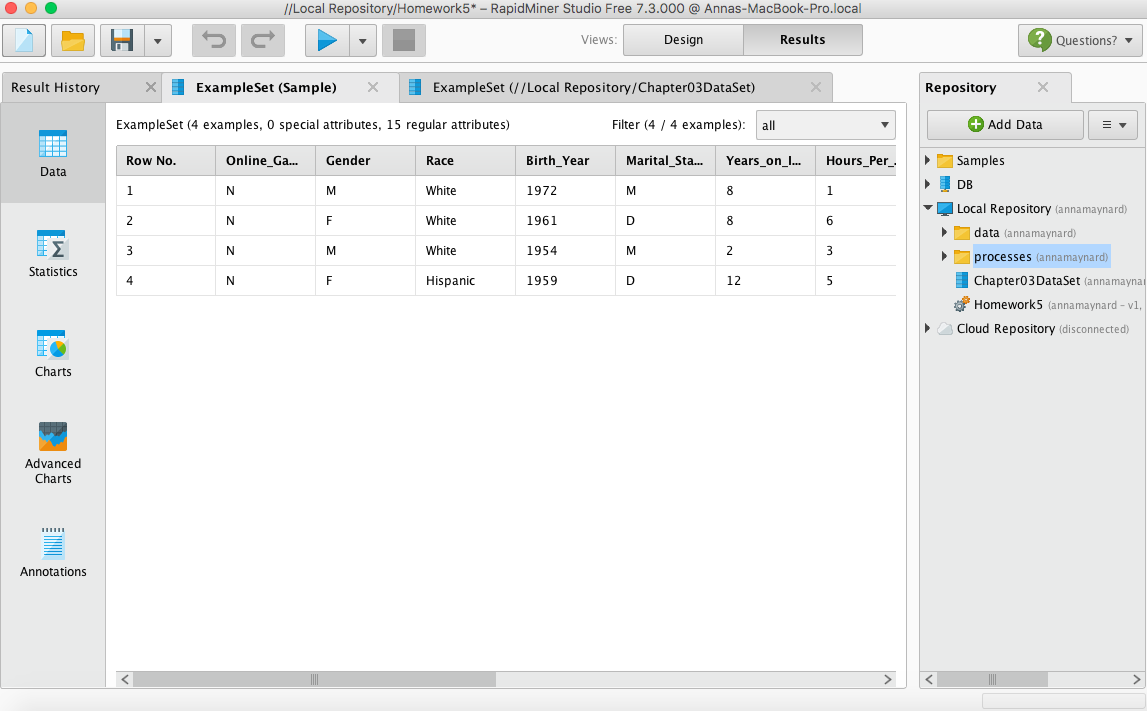
****

1. Association Rule usage must be a key element in product recommendations for browsing customers. You could use sales data to determine what complimentary products are bought when any certain item is purchased, maybe determine the likelihood of a complimentary purchase and display the recommendations with the highest confidence in the customer’s browser. This would help Amazon with their overall sales by convincing customers to spend more based-on data from their peers.

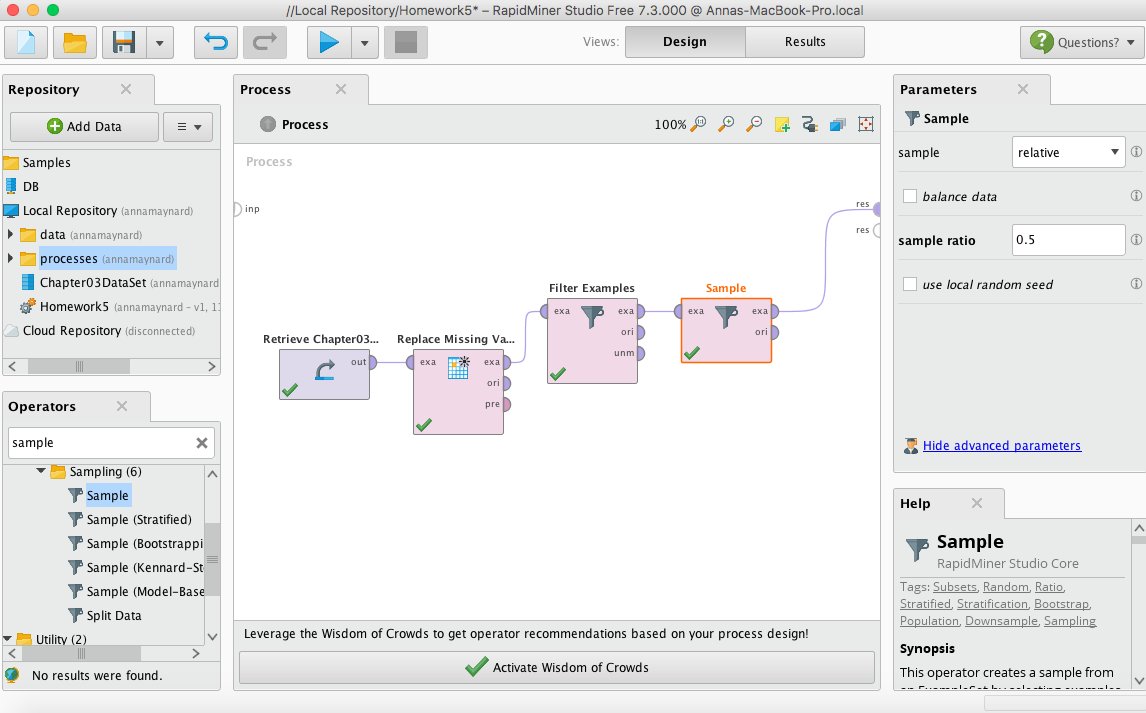
**Part 1 Chapter 3 Screenshots:**



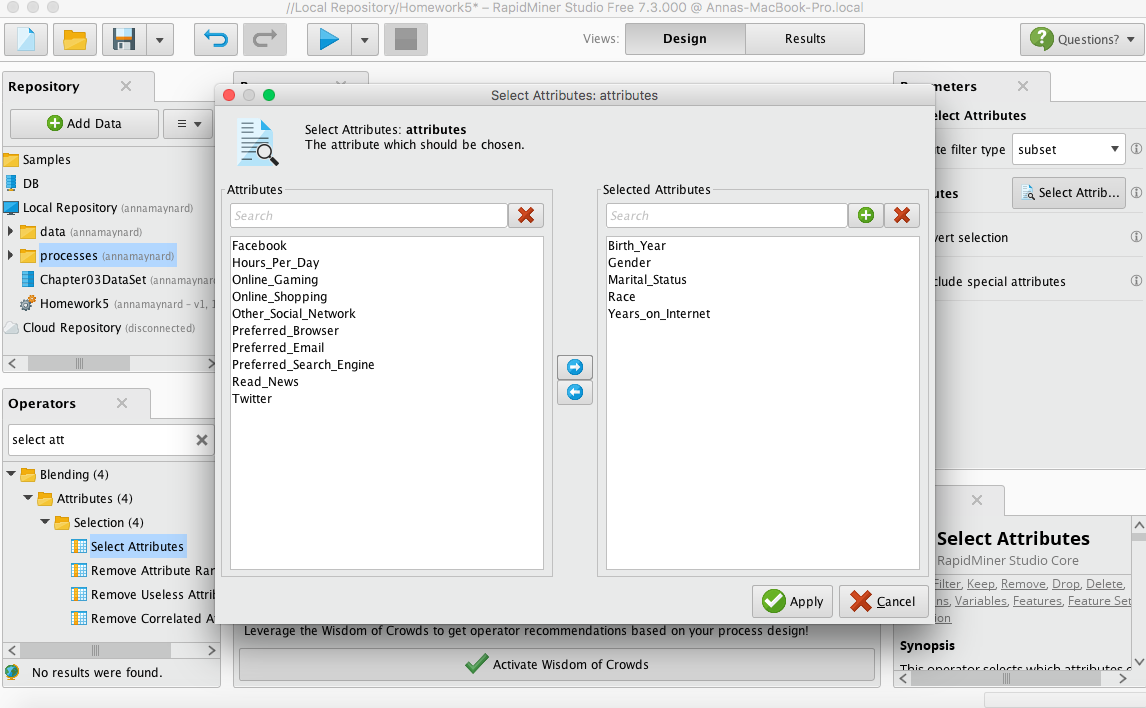
This shows that you can have multiple operator boxes going at once.



This shows that you can take a sample set of your data.

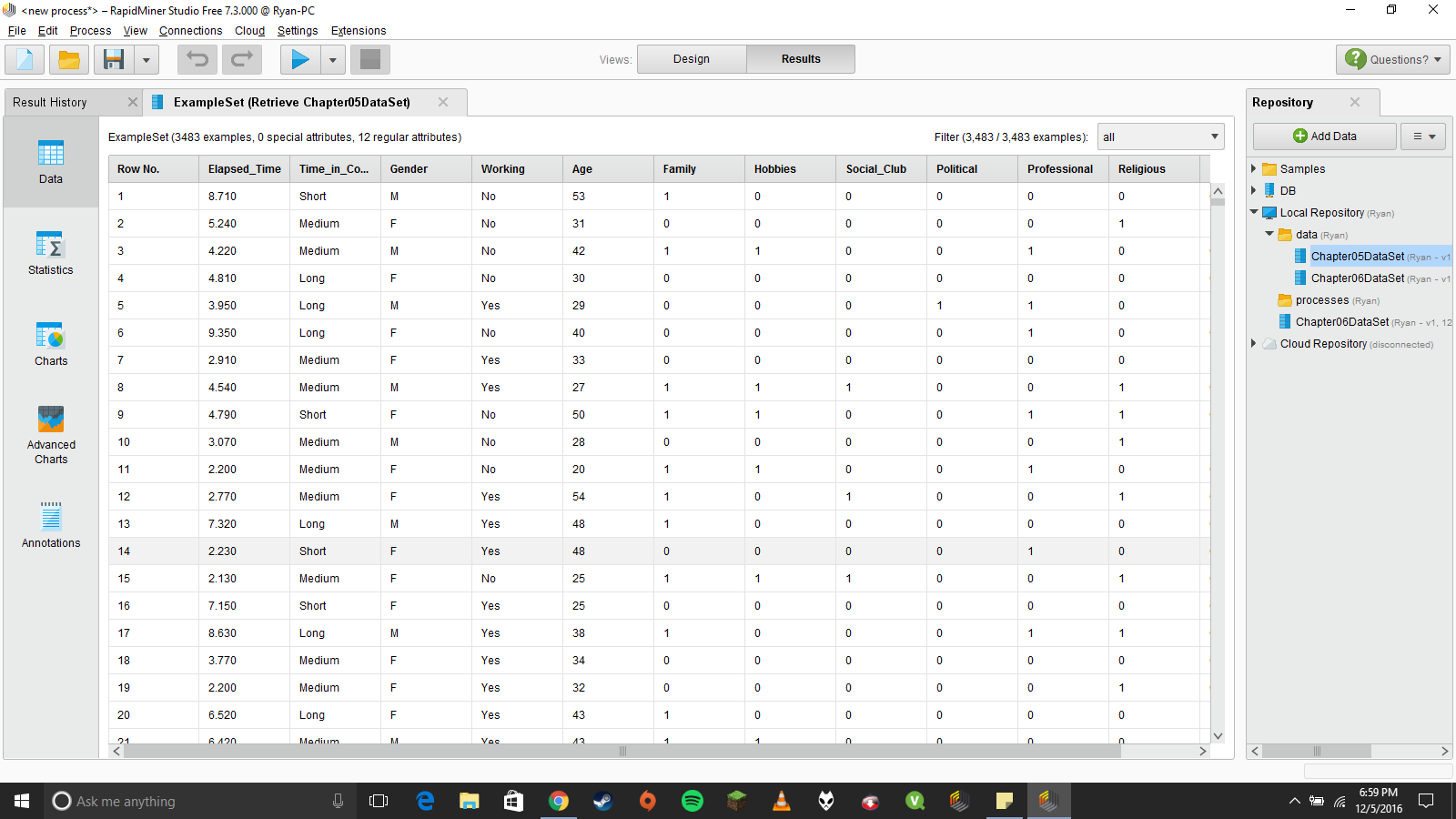


I included this screenshot to show that you can have many filters in play at once and then take a sample from it (shown on the right).

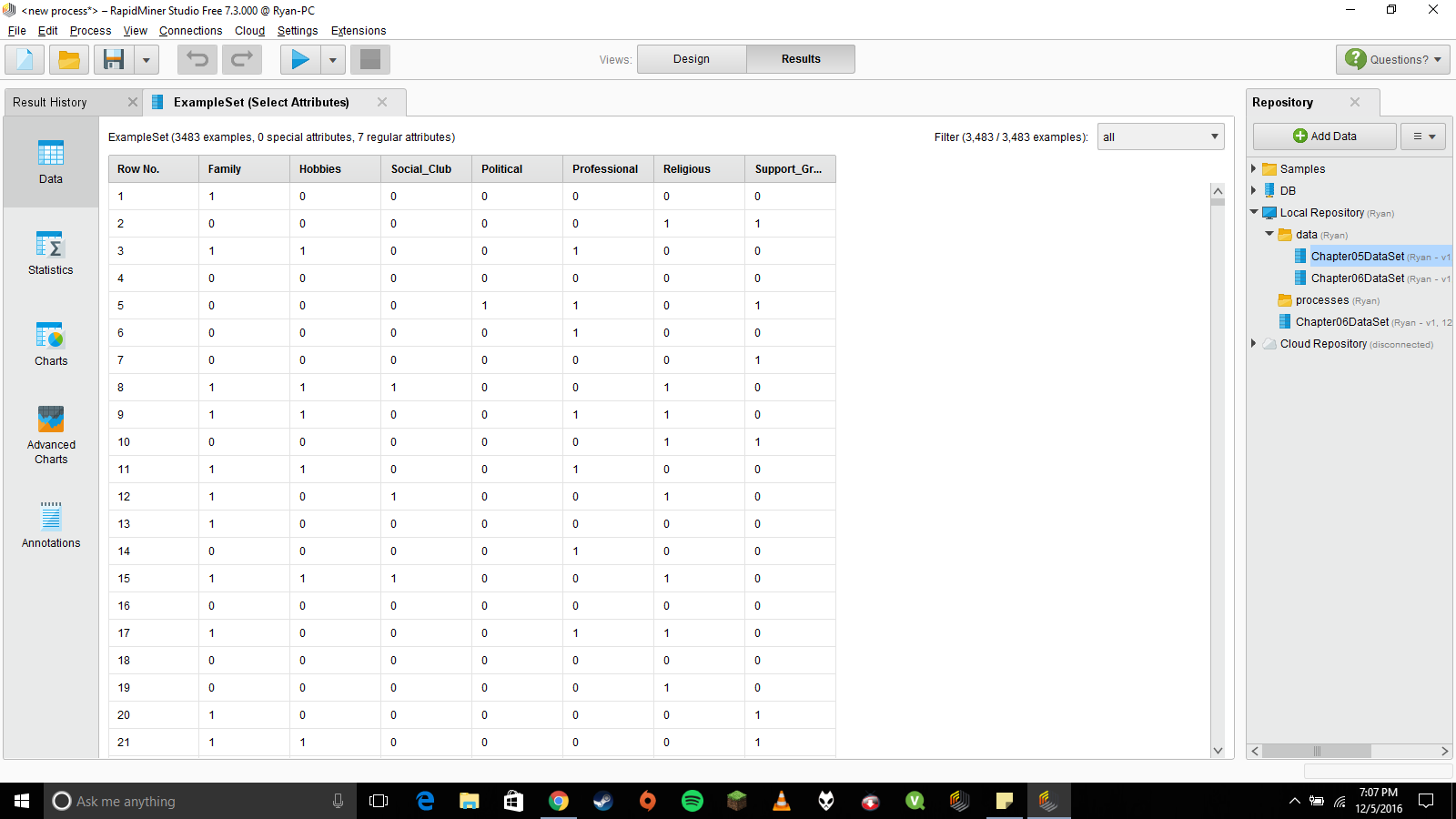


I added this screenshot to show that you can select certain attributes to focus on if you only care about a select few.

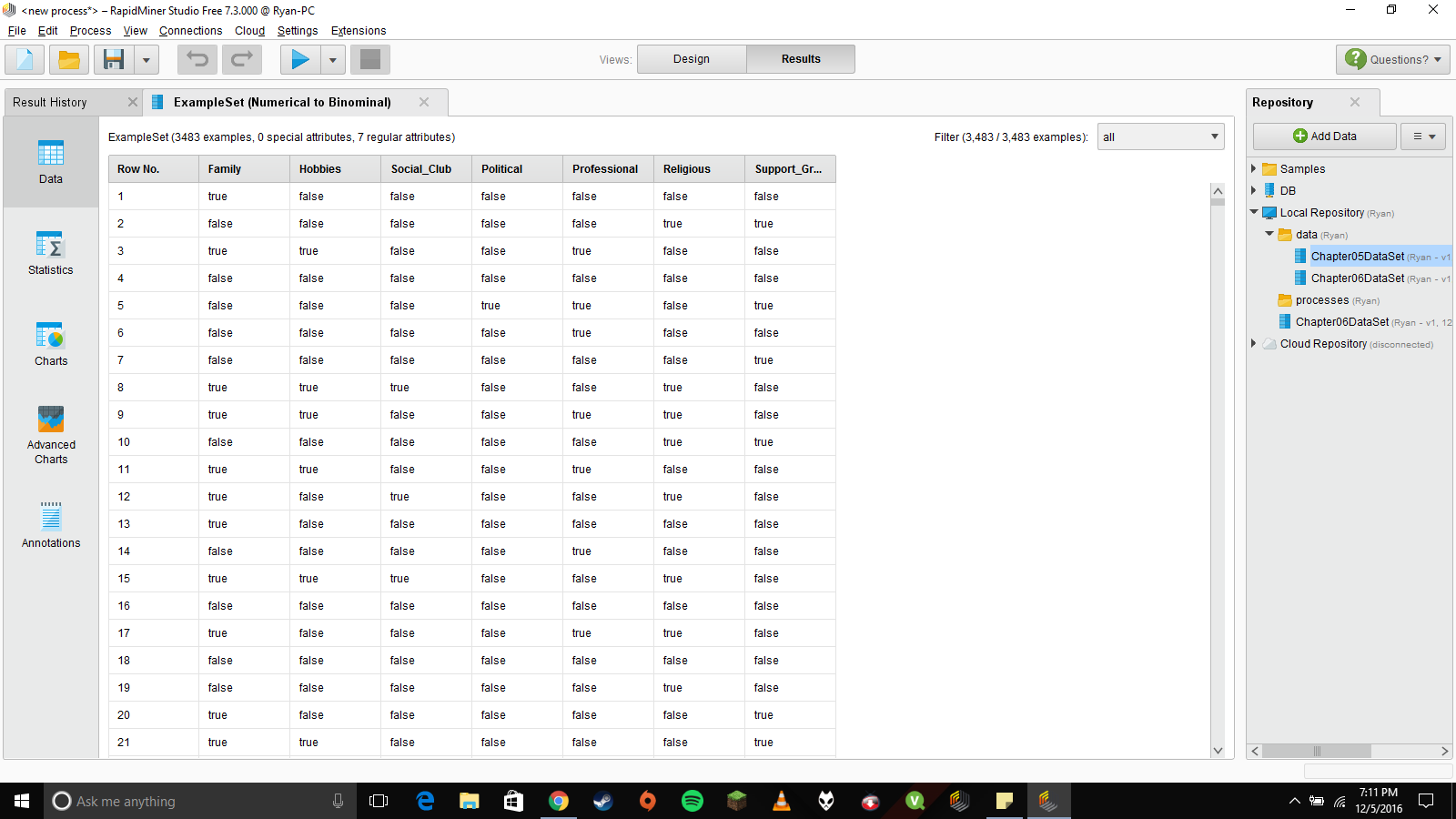
**Part 2 Chapter 5 Screenshots:**



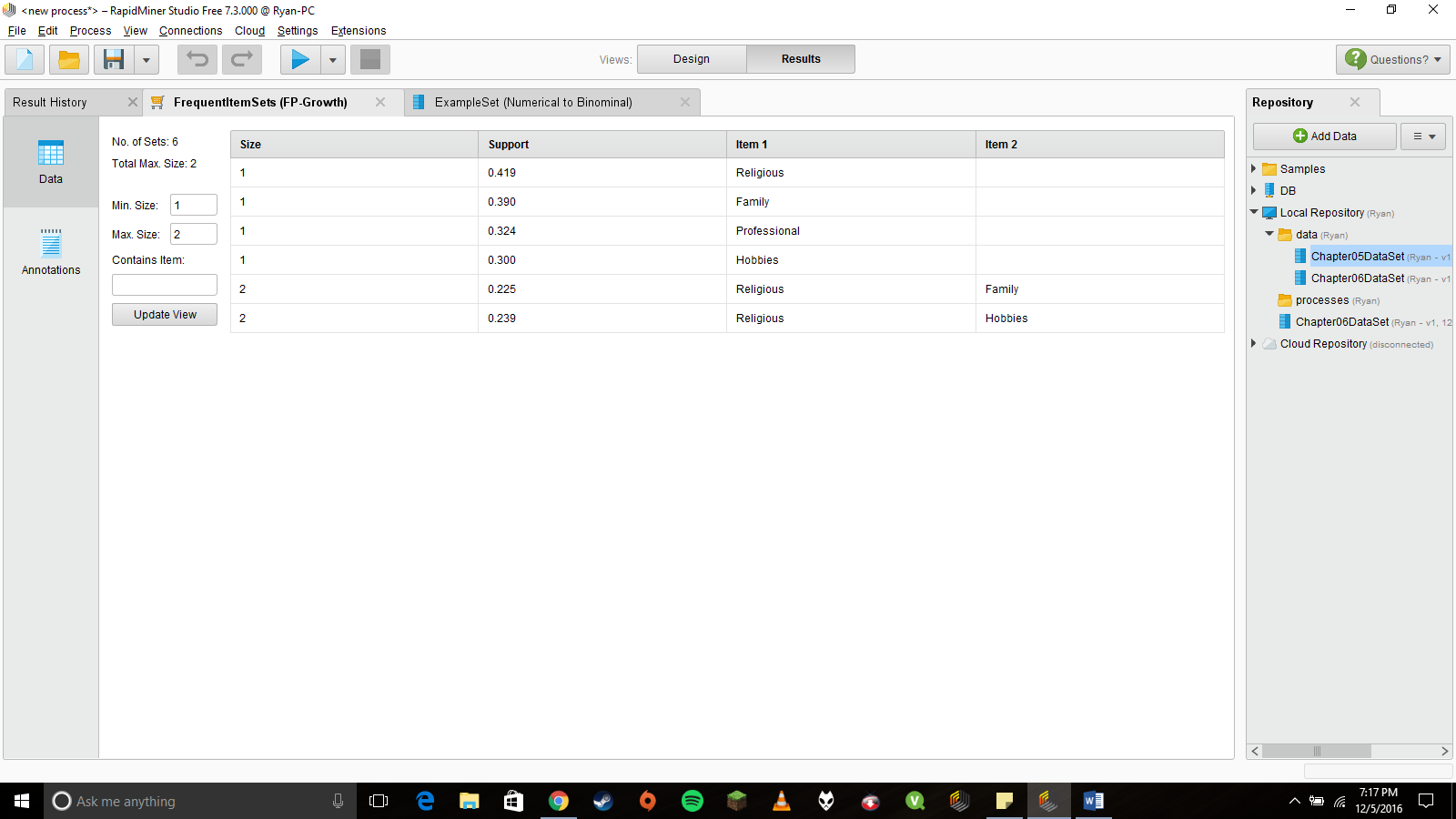
Metadata results from first, dry run.



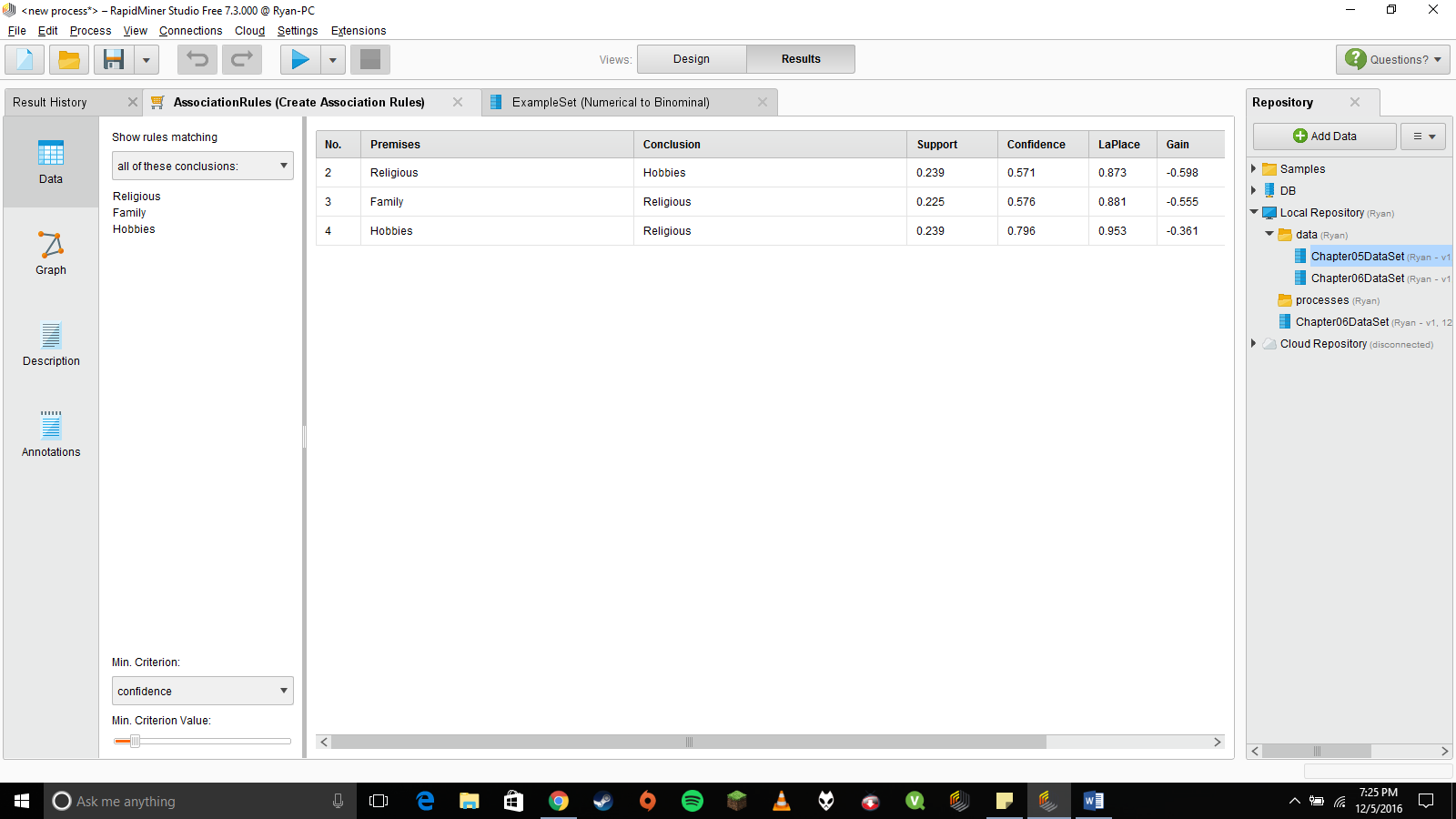
Filtering out some attributes with “Select Attributes” Operator.



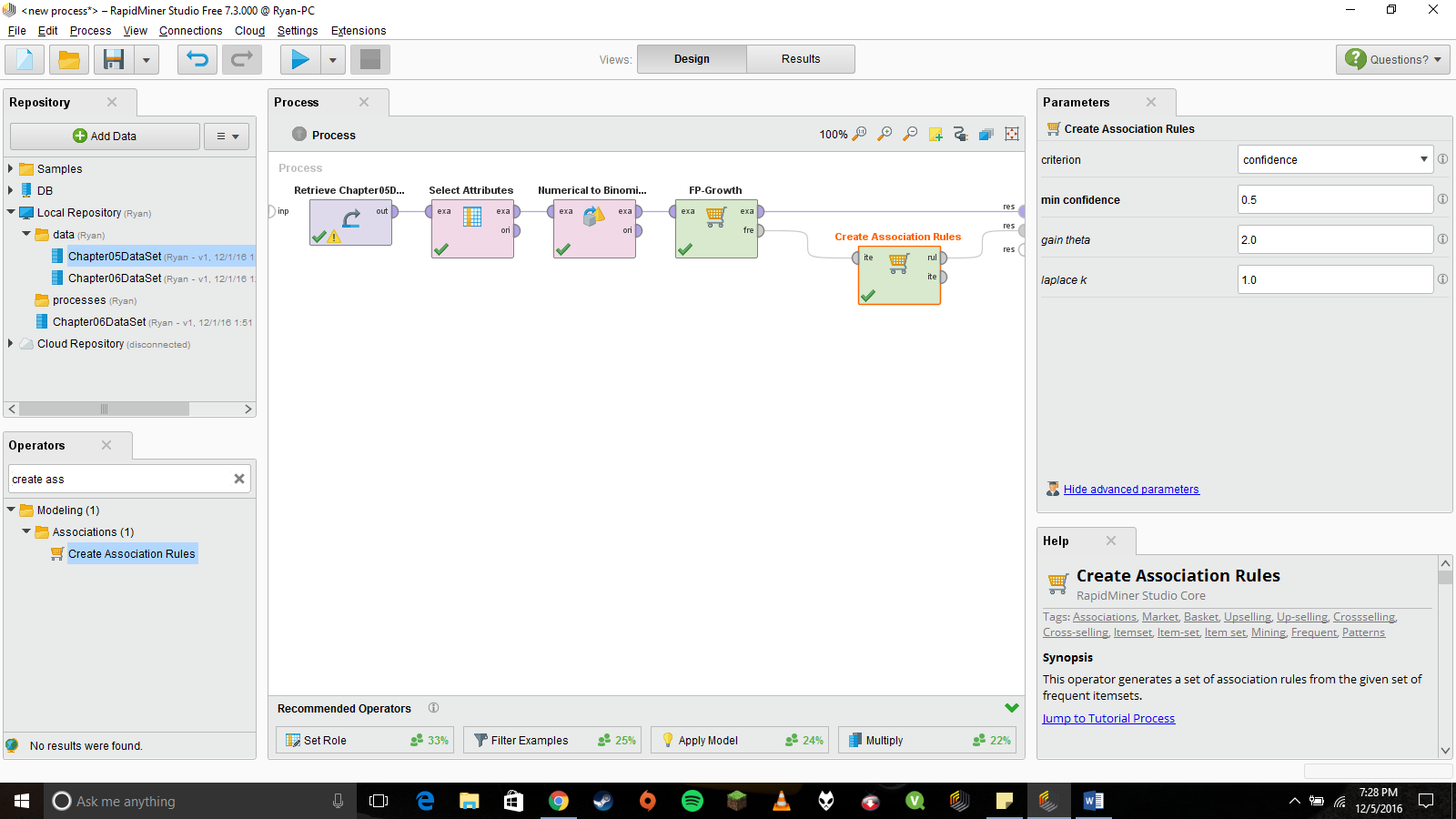
Converting values with the “Numerical to Binominal” Operator.



Sorting data with the “FP-Growth” Operator. Attribute patterns emerge between Religious Organizations and Family and Hobbies.

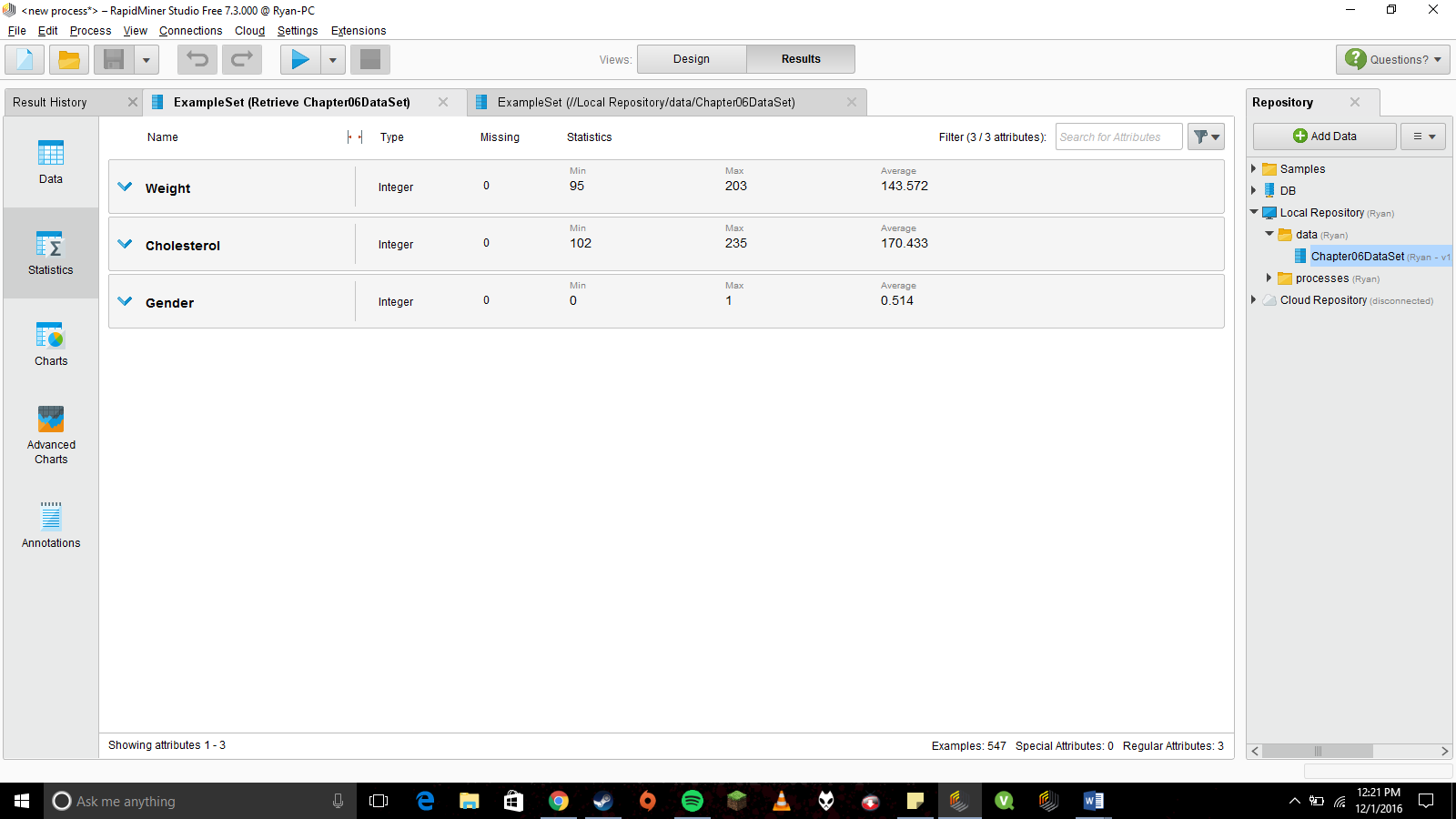


Data results after using the “Create Association Rules” Operator and changing the minimum confidence to 50% instead of 80%. Religious, Family, and Hobbies are in fact associated.

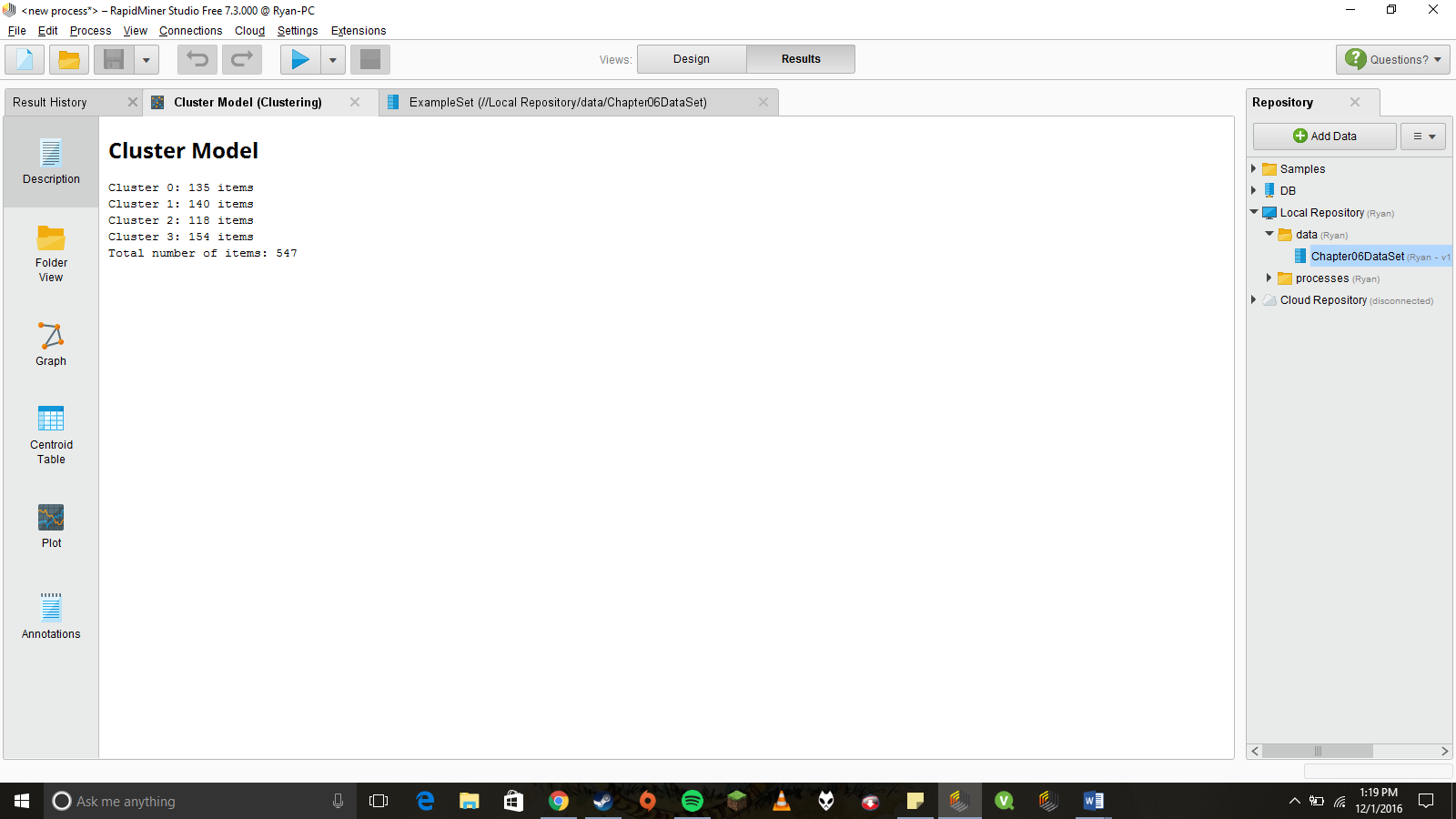


Final design.

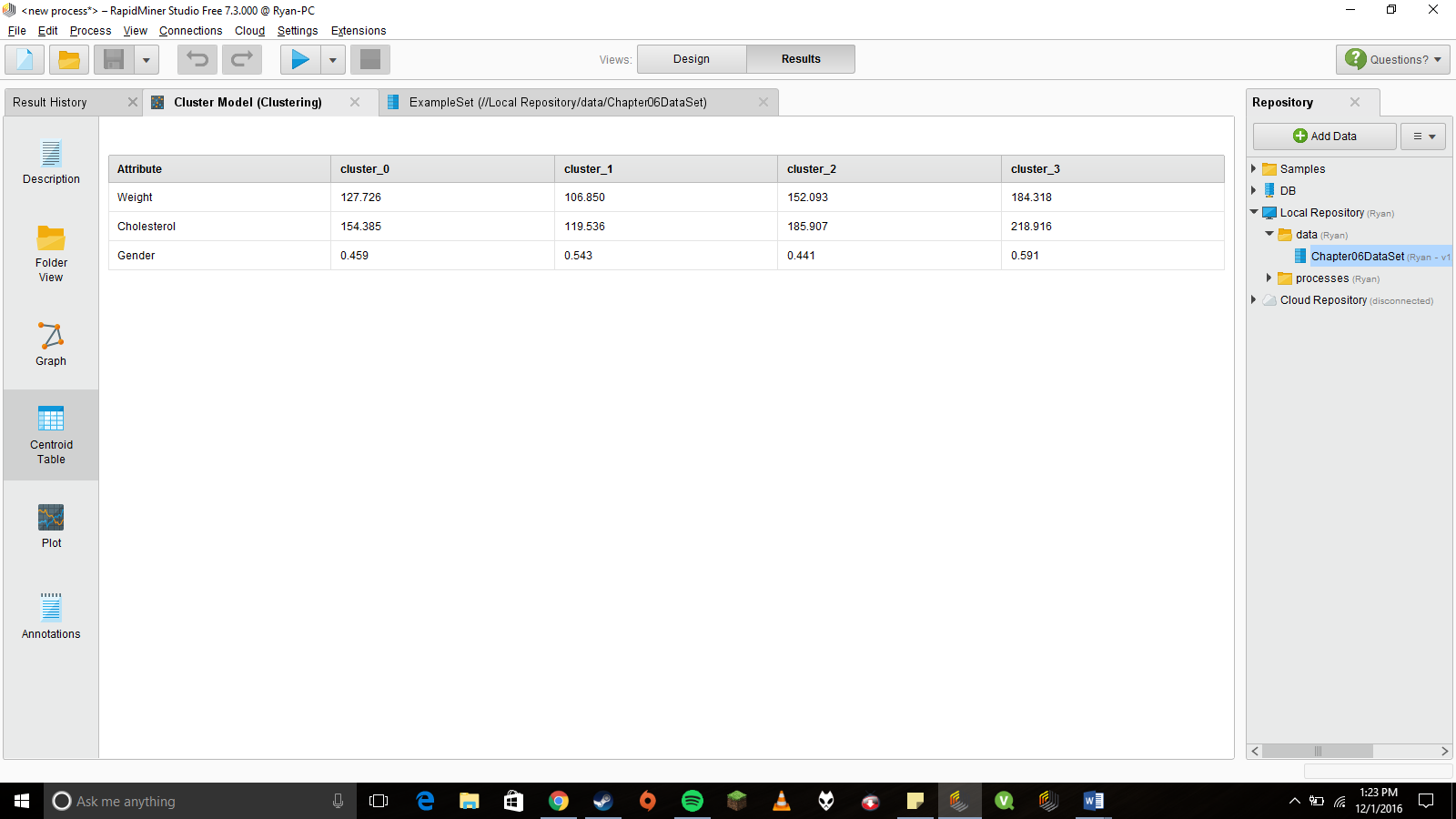
**Part 2 Chapter 6 Screenshots:**



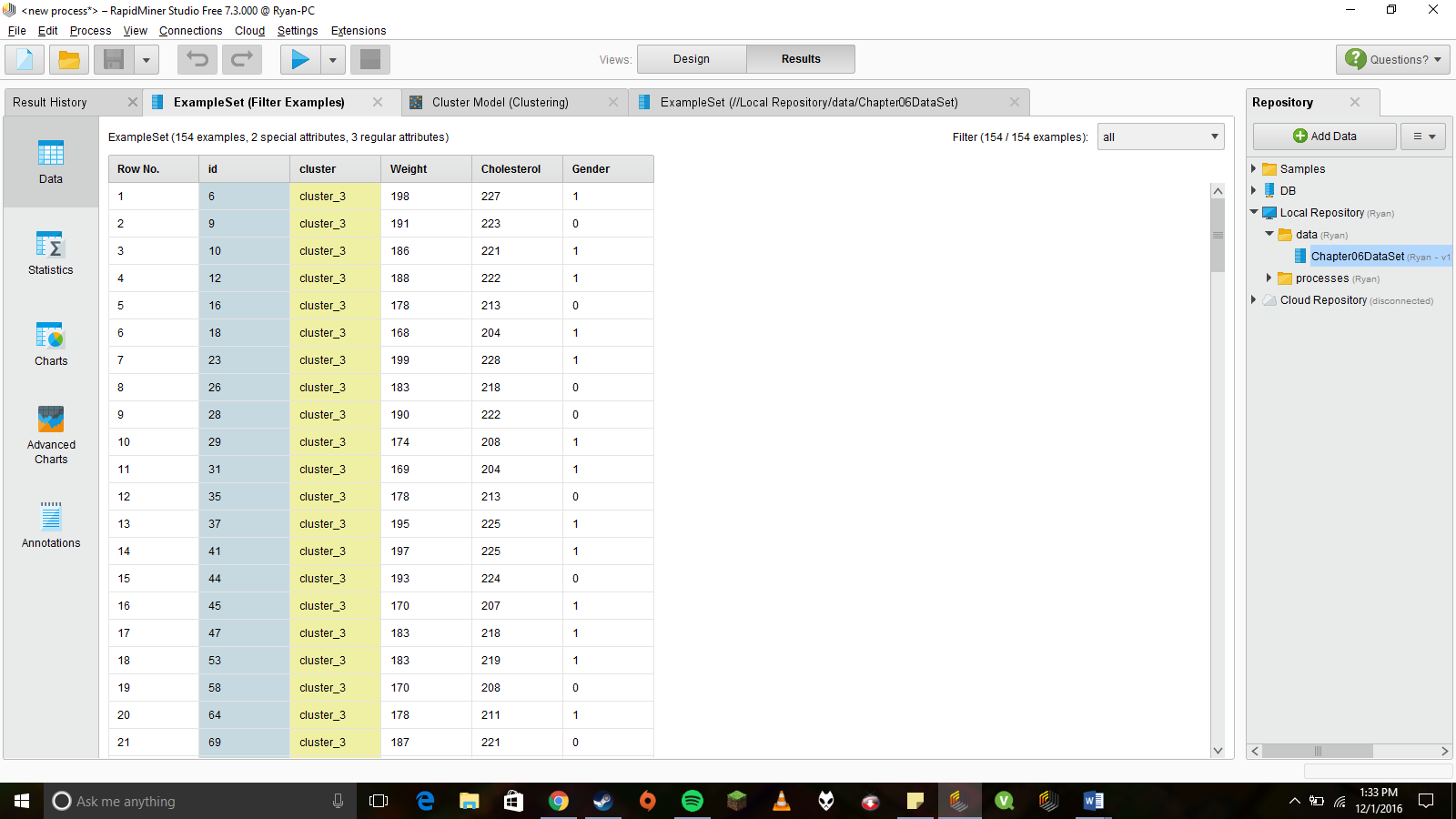
After connecting our data set to the computational tool, the analytical results were displayed on multiple tabs and presented in different formats to help us interpret the results no matter what we were looking for in the data set.



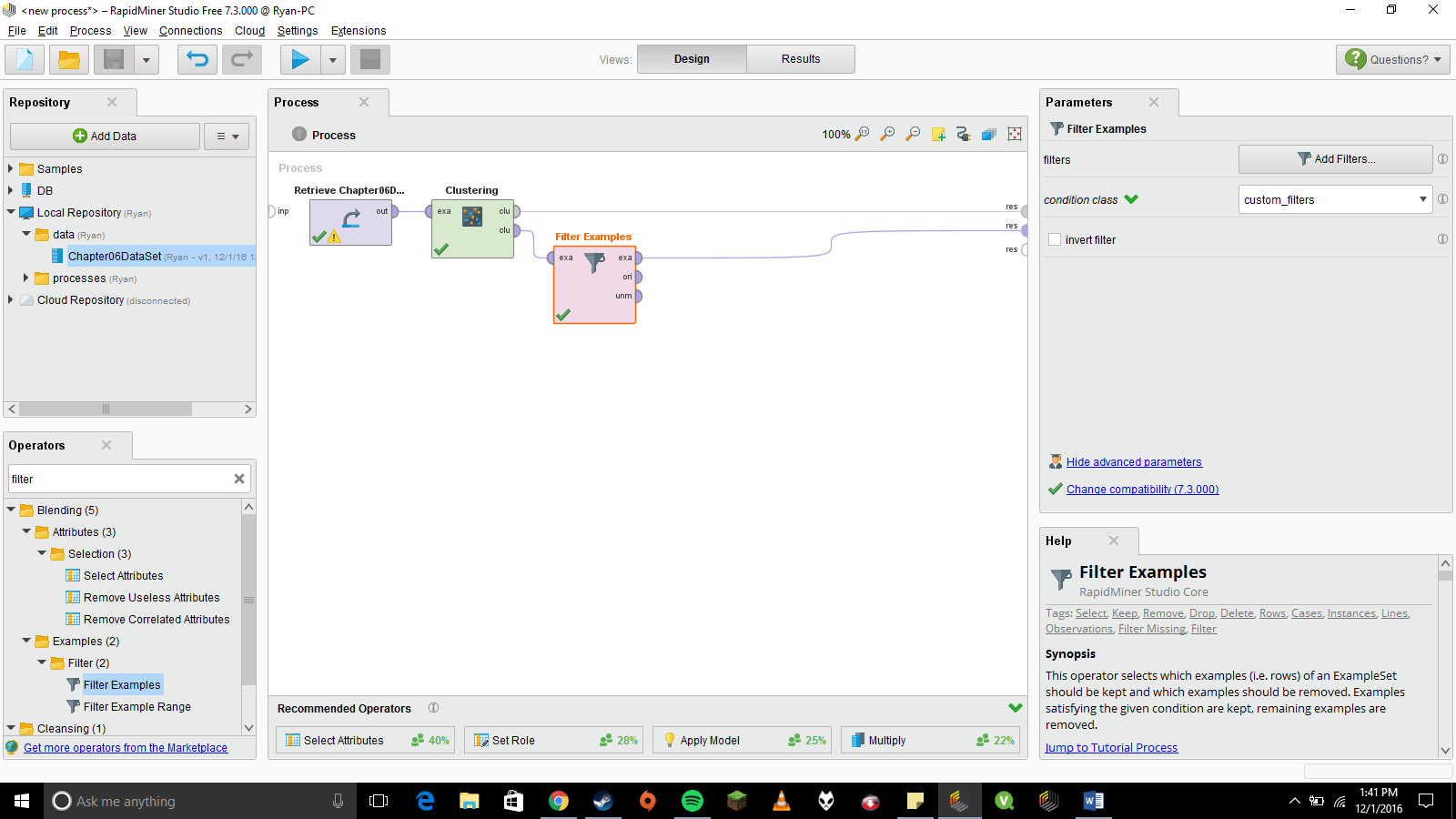
First cluster operator results.



Centroid Table view shows our current clusters. Cluster 3 has the highest values for Weight and Cholesterol. The Gender value of .591 indicates that a majority of the cluster is male since a value of 0 represents female respondents and 1 represents male.



Cluster 3’s individual results using the Filter Examples Operator. (Not sure how my clusters re organized from the guide but Cluster 3 is Sonia’s target cluster for health concerns)



Clustering design overview.