

Arshdeep Benipal

100591622

I, Arshdeep Benipal, certify that this work is my own, submitted for CSCI 4210U in compliance with the Academic Integrity Policy .

Question 1

- If the list was organized by amount of potassium per serving from largest to smallest, 65% of the top 20 cereals contain the word bran in them
- The Nabisco Manufacturer provides the overall healthiest cereal as it has the highest average of fiber, protein, and potassium, while also having the lowest average calories, fat, sodium, and sugars.
- General Mills Manufacturer provides the overall unhealthiest cereal as it has the highest average sodium, sugars, and calories (second highest in fat) while having the lowest average fiber and potassium

2												
3		Data										
4	Manufacturer	Average of Carbohydrates	Average of Fiber	Average of Calories	Average of Protein	Average of Fat	Average of Sodium	Average of Sugars	Average of Potassium	Average of Vitamins	Average of Cups	
5	A	16	0	100	4	1	0	3	95	25	-1	
6	G	14.72727273	1.272727273	111.3636364	2.318181818	1.363636364	200.4545455	7.954545455	85.22727273	35.22727273	0.875	
7	K	15.13043478	2.739130435	108.6956522	2.652173913	0.608695652	174.7826087	7.565217391	103.0434783	34.7826087	0.622173913	
8	N	16	4	86.66666667	2.833333333	0.166666667	37.5	1.833333333	120.6666667	8.333333333	0.111666667	
9	P	13.22222222	2.777777778	108.8888889	2.444444444	0.888888889	146.1111111	8.777777778	113.8888889	25	0.714444444	
10	Q	10	1.3375	95	2.625	1.75	92.5	5.25	74.375	12.5	0.32375	
11	R	17.625	1.875	115	2.5	1.25	198.125	6.125	89.25	25	0.37125	

I loaded the data in to Excel, then I applied filters to all the headers. I then played around with the filters by sorting the data based off of different columns. As I played around with the filters, I noticed the first insight I provided above. I then took the average of each column based on each cereal and noticed my other two insights I provided based on this new table I used. I provided the table above as well.

Some issues I ran into was that I was unsure what to do with the values that were '-1' as I was unsure whether this was missing data or just data that would be unused. The data did not have any kind of consistent measurement between the cereals. So it was harder to traverse the as it was now compared to if all the cups were set to 1. Also I was unsure what the measurements were being used or what the column meant, like vitamins. Vitamins could either be all the vitamins together or just a certain few. Also unsure if they are measured in grams or percent daily values.

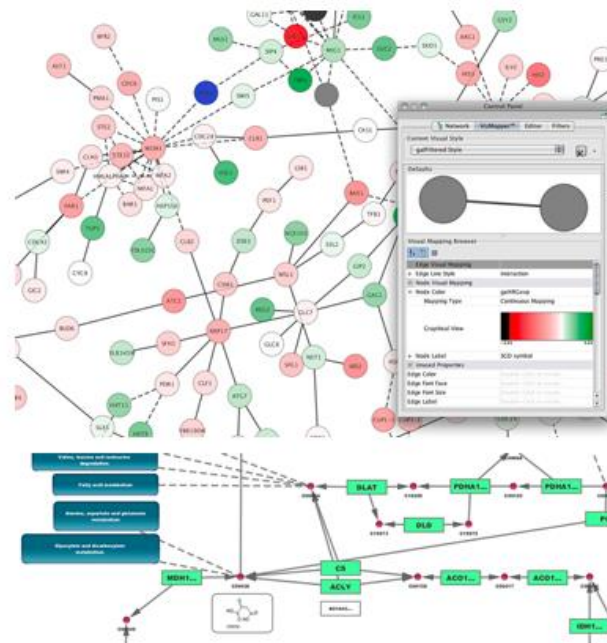
Question 2

What Can You Do With Cytoscape?

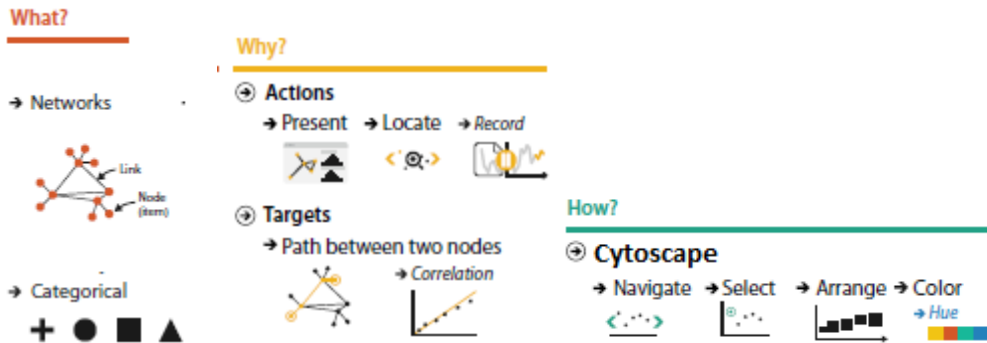
Biology

Cytoscape supports many use cases in molecular and systems biology, genomics, and proteomics:

- Load molecular and genetic [interaction data sets](#) in [many standards formats](#)
- Project and integrate global datasets and functional annotations
- Establish powerful visual mappings across these data
- Perform advanced analysis and modeling using [Cytoscape Apps](#)
- Visualize and analyze human-curated pathway datasets such as [WikiPathways](#), [Reactome](#), and [KEGG](#).



Cytoscape is a platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. The core distribution provides a basic set of complex network analysis and visualization. The dataset is a network. With the dataset availability being a static while attributes can be gene composition or a model of the gene. The attribute type is categorical. Actions that are supported are present, record, or locate and some targets could be correlation and path between two nodes. The networks can be navigated through with your mouse by using the scroll wheel to zoom in and out. Also if you keep the left mouse button pressed and then move the mouse it will allow you to move around the network. You can select and move or arrange a node however you may like. There are also features that allow the user to record different phases of the network by using snapshots and compiling the snapshots together. For example you can take a snapshot of the initial state, then move around or zoom in to a specific area and take another snapshot, then change the hue of the nodes.



Question 3

- I noticed that for team based sports the medal counts were not based off of country but each player so instead of the US having 1 gold medal it has 23 gold medals because that was how many players were on the team
- United States has had the highest medal count every Olympics except for one in which Canada won
- There is significantly less amount of medals given in the winter Olympics compared to the summer Olympics

I used the show me function to create a map in which we can view the total medals and medal breakdown. Then I created a bar chart that lists the total medals for each country in each year. I then created a stacked bar chart that showed the gold medals for each country in each sport. Which was stacked by different colour for a different year. I noticed on this graph my first insight.

Data analysis with Tableau was harder compared to Excel because you can notice a lot more with the data itself. Some things that frustrated me was missing data on both data sets, such as the null and -1 values. As I am not very proficient with Tableau compared to Excel, I have a bias towards Excel. My previous experience with Excel made Excel much easier to use. Tableau can provide a much clearer visualization of the data compared to Excel. As tableau can create a map from the data that was provided as well as apply filters and Marks. You can also create a dashboard that is a customized compilation of the graphs you created. The dashboard allows one to convey your ideas to someone else.