**Q1:**

Defining insights for visual analytics talks about the exploration of understanding how people learn. Primarily the author mentions the differences and similarities between two fields of thought discussing the definition of insight. First mentioned is cognitive science’s definition of insight, second is insight defined from a visual analytics perspective. Those on the cognitive science side refer to insight as an “aha” moment in the learning process, it can even be observed using modern scientific instruments pointed at a being’s neural activity. The visual analytics side of things argues that visualizations can be crucial in helping people to hit this “eureka” moment.

The two fields of thought differ on what exactly an insight is. On the cognitive science side, the definition for an insight is far more rigid whereas visual analytics tends to have a looser definition. The author uses the term spontaneous insight to define the insight from the cognitive science side of things. Visual analytics argues that insight is more than just the moment someone flips a switch from not understanding to understanding. Visual analytics aims to define the moments before someone has that insight. The thought is that with the aide of visual analytics we can help people to have more of these revelations.

Experimenters are able to give users tasks that are known to provide spontaneous insight while connected to an EEG and fMRI. An example of one of the tasks is the 9 dots task where a user must connect a grid of dots without lifting the pen. During this task it is shown that when users are relaxed or have exhausted all the options they can think of to solve it do they have a jolt of understanding. The author argues that with spontaneous insight a user must have no structure or definition of how exactly to learn a problem, the user must gain insight through knowledge or assumptions they make on their own. Knowledge based insight lets people follow a set of instructions in order to understand the already solved issue or insight.

For both types of insight, spontaneous and knowledge based, it is crucial the person trying to have an insight has more knowledge about what they are trying to learn. As they get more and more the chance for them to have an insight gets greater. This paper came out in 2009 and is only more relevant now with the rise of artificial intelligence and machine learning. Not to say AI is anything new, it’s been around since the 1960s, however breakthroughs in machine learning have led us away from the AI winter. Many argue that AI doesn’t yet have a clear definition, “insight” is much the same, those in the field of visual analytics argue that it does not yet have a clear definition.

Visual analytics is supposed to help foster learning and lead to insight. Although we have different definitions for insight I am going to choose the cognitive sciences simple definition of a “eureka” moment. I would argue that visual analytics is just as important as it was in 2009. Although advances in machine learning and data mining have drastically increased web crawlers and scrapers were definitely around in 2009 and prior. Visual analytics have been important in aiding understanding and insight since the time of only pen and paper.

Although insights and visual analytics are not anything new, I would argue they are more commonplace and accessible. To both those viewing and interacting with visual analytics and those creating visual analytics. My hypothesis would be the net amount of data visualizations is increasing due to the increase in tools lowering the barrier to entry. Microsoft Excel being one of the simpler tools and D3.js being one of the more complex tools. Once again, many spreadsheet applications have been around pre 2009 probably allowing the creation of simple line charts, bar charts, and scatter plots. After all, for many problems those do the trick just fine.

More complex visual analytics are possible and those that know how to properly display the data can have a relatively simple time creating in depth and interactive tools. Unfortunately, it also means it is far easier for people who are less knowledgeable to portray data falsely and misinform others due to the internet and ease of sharing. For example, in class we spoke of how difficult it can be when trying to plot data on a map. Although the task is rather simple to implement it is challenging to convey a message without skewing what is meant to be conveyed.

Visual analytics has not necessarily changed. It has however become more accessible and susceptible to improper usage. Insights are more plentiful than ever due to visual analytics. Looking forward the importance of ensuring the quality of those visual analytics increases is growing more drastically.

**Q2:**

The four columns I chose were: Cylinders, Acceleration, MPG, and Horsepower. I chose those columns for analysis because I was interested to see what properties each car group would have. The cylinders was a good categorical feature that clustered the cars. Each car’s cylinder count makes it belong to a cluster. I used teal for 8 cylinder cars, yellow for 6 cylinder cars, red for 5 cylinder cars, purple for 4 cylinder cars, and orange for 3 cylinder cars. The largest cluster was purple or the 4 cylinder cars with a grand total of 201 instances. The size of the node shows how good of a MPG the car gets. It looks like the 4 cylinder cars get the best gas mileage. They generally have a lower horsepower it seems in exchange for the great MPG. It also looks like the cars can have varying acceleration rates.

I didn’t find any cars that were falsely clustered. I’m trusting that the data was gathered properly but if I had more time it would’ve been useful to examine the data more. Something else I’m not so sure of was the fact I scaled the data to make it easier to examine. I’m not sure if this was the correct thing to do as it may be misleading and further cement my answer to Q1 that it’s hard to make a good data visualization. That being said, it may be possible the data is improperly clustered. I did multiply the Horsepower by 2 to make the data look better, I also multiplied the acceleration by 30, and I bumped the MPG values down by ¼ in order to make it easier to see the surrounding nodes.