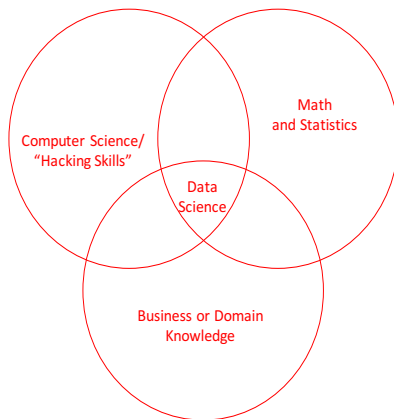


1. **(2 pts.)** Draw the data science Venn diagram: (Include at least the three major labels and the location of data science in the diagram.)

A: Accept *only* Venn diagram from instructor notes. ½ point each.



Order around Venn diagram does not matter, but data science must be in the middle.

2. **(1 pt.)** Describe, or give an example of, structured data.

A: Rows and columns of neatly ordered numeric and simple character data. Like a plain excel spreadsheet or the Adult data set from the lecture materials.

3. **(1 pt.)** What is the definition of “data mining” according to Tan’s *Introduction to Data Mining*?

A: Non-trivial extraction of implicit, previously unknown and potentially useful information from data; exploration & analysis, by automatic means of large quantities of data in order to discover meaningful patterns

4. **(2 pts.)** What are the two major types of machine learning commonly used in data mining tasks?

A: Supervised learning and unsupervised learning.

5. **(1 pt.)** According to Tan’s *Introduction to Data Mining*, name one challenge faced by practitioners of data mining?

A: One of: Scalability, dimensionality, complex and heterogeneous data; data quality; data ownership and distribution; privacy preservation; streaming data

6. **(1pt.)** According to SAS' *Advanced Business Analytics*, what is one common business application of data mining?

A: One of: New customer acquisition; customer loyalty; cross-sell/up-sell; pricing tolerance; supply optimization; staffing optimization; financial forecasting; product placement; churn; insurance rate setting; fraud detection

7. **(1pt.)** According to *A Short History of Data Science*, in what decade was the phrase "data science" first officially used?

A: 1970s

8. **(1pt.)** What is the name of the statistician who is often credited with inspiring the current field of data science with his seminal paper, "The Future of Data Analysis?"

A: John W. Tukey/John Tukey/Tukey