Aim: How do we use existing data to infer unknown data?

Do Now: Guess (in writing) how many small triangles are in Mr. Ramanathan's image (at the board). Average everyone's guess together.

Key point: The average value of a large number of approximations is often a better approximation than any single approximation taken alone.

Before beginning, google data and/or big data and look at their definitions.
Big Data video (optional): https://www.youtube.com/watch?v=bMrDHTGHFR4

Google Regression and/or Outlier/anomaly if you didn't learn them in math class.

5. A number of parents have volunteered their children to participate in a developmental study administered by a local child psychologist. The following chart summarizes the results of the psychologist's assessments.

![Developmental Test Scores Chart](chart.jpg)

Which of the following statistical techniques can the psychologist use to determine the developmental score of a typical 4-year-old child despite the fact that no 4-year-old children participated in the study?

a. Classification  
b. Clustering  
c. Outlier / anomaly detection  
d. Regression

6. A snack company is starting an advertising campaign for its new line of tortilla chips. Rather than target specific demographic groups in its commercials, the company has decided to perform market research to determine common characteristics of patrons who prefer their chips. This is an example of what type of data mining strategy?

a. Cluster analysis  
b. Data classification  
c. Association rule mining  
d. Outlier / anomaly detection

Association rule mining: https://www.techopedia.com/definition/30306/association-rule-mining
25. Similar to a Google Fusion Table, a company wants to organize their data in a relational database so that their data can be easily narrowed according to certain categories. This is done using what?
   a. Association rule mining
   b. SQL (Structured Query Language) filters
   c. Cluster analysis
   d. HTTP (Hypertext Transfer Protocol) indexing

26. A teacher noticed that they have had a lot of students in their classes that did not do very well on homework and classwork assignments, but still seem to do about the same as other students on the final exam. In order to check their claim, a statistics teacher helped them to create the following graph from a data set that compares a student’s final exam grade (FE_Grade) to their marking period average (Average) in the class. What the teacher was observing and what the graphs shows depicts what data analysis technique?

   ![Graph](image)

   a. Cluster analysis
   b. Regression
   c. Association rule mining
   d. Data classification

29. There was a large study conducted on a random sample of 500 students from the United Kingdom, South Africa and Australia. The graph displays a comparison of each student’s height and age. Four data points are represented by stars on the graph. Upon further inspection, it was discovered that the 4 students had rare medical conditions.

   ![Graph](image)

   The process above best describes what data mining strategy?
Data exploration
b. Classification
c. Cluster analysis
d. Anomaly detection

30. After reviewing the service records at a car dealership, the CEO (chief executive officer) discovered that people that scheduled service for a transmission fluid exchange and differential fluid exchange also typically schedule for an oil change. This conclusion is done using what data analysis technique?
   a. Classification
   b. Association rule mining
   c. Clustering
   d. Regression

33. The highest word frequency method, the TF*IDF method and the topic sentence concatenation method are all techniques that will produce a nice summary of commonly used words in a large group of text. This can then be displayed in a nice visual format like a word cloud. All of these methods are forms of what type of data mining strategy?
   a. Data observation
   b. Classification
   c. Association rule mining
   d. Automated summarization

35. There is a popular site, similar to Netflix and Hulu, where you are allowed to stream different television shows and movies over the internet. There is a vast assortment of shows and movies in their database that users can choose from. In order to find the show or movie of choice, the most popular way is to locate it by genre of movie or by the television station that the show is on. Each month, this site adds new movies and tv shows to their database and assign them to already existing categories. This process of adding new movies best describes which data analysis technique?
   a. Clustering
   b. Classification
   c. Association rule mining
   d. Regression