The AP CS Principles Exam Reference Sheet
Starts at page 125

Other Resources
Two resources come highly recommended by some of our 2016–17 teachers:

- Fast Track to a 5: Preparing for AP Computer Science Principles Examination ($19.50 from National Geographic Learning)
- Albert.io: Sample exam questions for AP CS Principles, organized by the “Big Ideas” of the framework, are available at: http://tiny.cc/dwulgy.

Multiple Choice
1. Which Big Data analysis technique involves the examination of previously collected data sets in an attempt to discover patterns and other knowledge hidden within the data?
   a. Data Mining
   b. ReCAPTCHA
   c. Crowdsourcing
   d. Screen Scraping

2. An infographic displays the relative frequencies of the 100 most common emoji used in text messaging for each of the last 12 months. Which of the following conclusions cannot be drawn from such a representation of emoji usage?
   a. You can determine the growth or decline in popularity of a particular emoji.
   b. You can determine what percentage of text messages contains a particular emoji.
   c. You can determine how long the most popular emoji has held the #1 position.
   d. You can determine the average age of emoji users based on emoji use.

3. Consider the following relational database that contains the following census data.

<table>
<thead>
<tr>
<th>City</th>
<th>State</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albany</td>
<td>New York</td>
<td>98,566</td>
</tr>
<tr>
<td>Anaheim</td>
<td>California</td>
<td>346,997</td>
</tr>
<tr>
<td>Austin</td>
<td>Texas</td>
<td>912,791</td>
</tr>
<tr>
<td>Charlotte</td>
<td>North Carolina</td>
<td>809,958</td>
</tr>
<tr>
<td>Dallas</td>
<td>Texas</td>
<td>1,281,047</td>
</tr>
<tr>
<td>New York</td>
<td>New York</td>
<td>8,491,079</td>
</tr>
<tr>
<td>San Antonio</td>
<td>Texas</td>
<td>1,436,697</td>
</tr>
<tr>
<td>Tempe</td>
<td>Arizona</td>
<td>172,816</td>
</tr>
<tr>
<td>Utica</td>
<td>New York</td>
<td>61,332</td>
</tr>
</tbody>
</table>
Which of the following would be the result if a user were to query this database for any “City” with a population between 100,000 and 1,000,000?

a. Austin, Dallas, San Antonio
b. Albany, Dallas, New York, San Antonio, Utica
c. Anaheim, Austin, Charlotte, Tempe
d. Dallas, New York, San Antonio

4. Which of the following is NOT a benefit of making digital information and scientific databases openly available across the internet?

a. Innovations in medicine, business and science can be developed from the increased knowledge gained from large data sets.
b. Inaccurate and misleading data can be more easily disseminated to scientific researchers.
c. Scientific researchers can more easily share data and collaborate on related research projects.
d. Data scientists can discover previously unnoticed trends and patterns hidden within large data sets.

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.

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

7. Sarah works part-time as a babysitter for a number of families in her neighborhood. In order to coordinate all of her babysitting jobs, she has created a website that parents can access to check her availability and reserve a night when Sarah can watch their kids. Using the website, parents can see Sarah’s schedule,
including which nights she is booked as well as comments and ratings from other parents indicating their level of satisfaction with her services for recent babysitting jobs.
In order to do this, the website stores an online database of all of her babysitting appointments, including information on each job as well as personal details about the parents and their children. For each job, the database stores the location (i.e., the family’s home address), date and time, and the rating or comments that Sarah received for her work. For the parent information, Sarah’s database stores the parents’ names, home address and contact information (home/work/cell phone numbers, email address, etc.). For the children, Sarah’s database stores their names, ages, list of allergies, list of medications and interests/hobbies. Because of the potentially sensitive and personal nature of the data that is stored in Sarah’s database, which of the following factors does she NOT need to be concerned with, from a security standpoint?

a. The database should not be publicly available and should only be accessed by an authorized user (i.e. Sarah) using a secure username and password interface.
b. The database should not store any sensitive or personal information that Sarah does not actually need in order to schedule a job (i.e. parents’ political affiliations, previous addresses, etc.).
c. If parents provide only partial information when reserving a babysitting job (e.g., they do not enter data for one or more of the items stored in the database), the entire database record will be insecure and vulnerable to attackers.
d. The information stored in the database should be encrypted so that any unauthorized individual (i.e. anyone other than Sarah) cannot read the data.

8. Which of the following is a risk of obtaining information through the use of crowdsourcing?

a. Crowdsourcing requires private information to be publicly shared with competitors.
b. The combination of multiple perspectives cannot produce greater insight and knowledge that can be obtained when working alone.
c. Crowdsourcing can only address small-scale problems, but cannot be used to solve large-scale problems.
d. Unless independently verified, the results of crowdsourcing may be inaccurate.

9. Which of the following can be used to extract structured information from unstructured data?

a. Create frameworks for information
b. Identify patterns in data
c. Organize data into information
d. All of the above

10. A certain social media web site allows users to post messages and to comment on other messages that have been posted. When a user posts a message, the message itself is considered data. In addition to the data, the site stores the following metadata.

- time the message was posted
- name of the user who posted the message
- names of any users who comment on the message and the times the comments were made

For which of the following goals would it be more useful to analyze the data instead of the metadata?

a. To determine the users who post message most frequently
b. To determine the time of day that the site is most active
c. To determine the topics that many users are posting about
d. To determine which posts from a particular user have received the greatest number of comments

11. The table below shows the time a computer system takes to complete a specified task on the customer data of different-sized companies.
Based on the information in the table, which of the following tasks is likely to take the longest amount of time when scaled up for a very large company of approximately 100,000 customers?

a. Backing up data  
b. Deleting entries from data  
c. Searching through data  
d. Sorting data

12. Biologists often attach tracking collars to wild animals. For each animal, the following geolocation data is collected at frequent intervals.

- time  
- date  
- location of the animal

Which of the following questions about a particular animal could NOT be answered using only the data collected from the tracking collars?

a. Approximately how many miles did the animal travel in one week?  
b. Does the animal travel in groups with other tracked animals?  
c. Do the movement patterns of the animal vary according to the weather?  
d. In what geographic locations does the animal typically travel?

13. Which of the following tasks best shows an example of using searching and sorting techniques of big data in order to find a useful pattern.

a. Tallying how many pencils and pens you use throughout a school year so you know how many to buy for the start of the next school year to insure you will have enough  
b. Creating a seating chart for a classroom based on an alphabetized list of student Names  
c. Keeping track of all employees’ email use to see how many personal or work-related emails are sent during work time to check for productivity  
d. Recording the amount of time it takes a student to travel from one class to another class in order to find the average

14. A hospital that has its own pharmacy keeps track of the following information.
At the end of the week, all of the data is summarized into a database that is accessible by financial analysts of the hospital that can be sorted by any column in ascending or descending order. Below is a portion of this database.

<table>
<thead>
<tr>
<th>Data</th>
<th>Patient Name</th>
<th>Room #</th>
<th>Medication</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/13/2016</td>
<td>Joe Smith</td>
<td>110</td>
<td>Sulfamethoxazole</td>
<td>$30.50</td>
</tr>
<tr>
<td>11/14/2016</td>
<td>Steve Randall</td>
<td>342</td>
<td>Lipitor</td>
<td>$62.99</td>
</tr>
<tr>
<td>11/16/2016</td>
<td>Ashley Tamb</td>
<td>208</td>
<td>Tylenol</td>
<td>$1.27</td>
</tr>
<tr>
<td>11/20/2016</td>
<td>Sue Dennison</td>
<td>145</td>
<td>Insulin</td>
<td>$150.42</td>
</tr>
</tbody>
</table>

Which of the following cannot be determined from the information in the database?

a. How many unique prescriptions were filled for a patient in a certain room?
b. How many patients were in the hospital on a given day?
c. What was the average cost of the medication, Lipitor, paid by patients?
d. What day did the pharmacy filled the most prescriptions?

15. A large data set contains information on all registered republicans in the United States. The following information is recorded.

- Name
- Age
- Gender
- Home address
- Whether they voted or not in 2016 presidential election

Which of the following questions could not be answered based solely on the information in this data set?

a. What percent of registered republican voters are male?
b. Which state had the most registered republican voters vote in the 2016 presidential election?
c. What is the average age of registered republican voters?
d. How many registered republicans voted for Gary Johnson in the 2016 presidential election

16. Let’s assume when a user texts another person, the phone company keeps track of not only who the sender and receiver of the message was, but also keeps track of the content of the text being sent. Which of the following could fall under the category of metadata?

a. How many words were in the message?
b. Was there was an emoji used?
c. What time was the message received?
d. All of the above

17. A student in a history class is creating an infographic about the civil war. He includes information on where battles took place, how long the battles lasted (on average), how many soldiers were involved from both sides, etc. What type of statistical analysis is this student using?

a. Descriptive Analytics
b. Predictive Analytics  
c. Prescriptive Analytics  
d. Reductive Analytics

18. Temporal scan thermometers are popular tools in order to take a baby’s temperature. It uses sensors that when slowly moved across the forehead provides a temperature reading in either degrees Celsius or Fahrenheit. Which of the following has to deal with this type of thermometer’s usefulness and not usability?  
a. The thermometer outputs the temperature with a digital display that uses a backlight.  
b. The thermometer (when used properly) is accurate to within 0.2 degrees.  
c. The thermometer is reasonably priced, so most families can afford it.  
d. The thermometer is ergonomically designed to be comfortable to use.

19. Both small and big businesses can benefit from using big data in their organization. Which of the following is true about how businesses could use big data to their advantage?  
a. Big data can be easily collected and analyzed in real time regarding customer behavior to then implement real time solutions.  
b. Big data can help keep track of how well a promotional offer worked for a company and can determine if it would be worthwhile to run the promotion again.  
c. Big data can be used to give the company a competitive edge by looking at data that can tell them where they are lagging behind in comparison to other companies, allowing them to make changes accordingly.  
d. All of the above

20. What best describes what happens when we take unstructured data and organize it into structured data?  
a. When we extract the data from an unstructured source and organize it into a structured data set, no data is lost, it is just in a more usable format.  
b. When unstructured data is organized into structured data there is some loss of data, but the data is in a much more usable format.  
c. When data is taken from an unstructured data source and organized into a structured set of data, some data is lost, but the data is now much more useful.  
d. When unstructured data is reorganized into a structured format there is no loss of data, it just simply becomes more useful in the new format.

21. What is the conversion of data formatted for human use to a more easily used format that can be used by automated computer processes?  
a. Automated Summarization  
b. ReCAPTCHA  
c. Screen Scraping  
d. Data Mining

22. There are programs designed so that when hard copies of things such as receipts, business cards and recipes are scanned into a computer, items from those paper copies are represented electronically. In order for the program to understand the difference between a zip code and a phone number or between a price of an item and the total bill for the shopping trip, certain rules must be put into place. If the program was trying to identify what part of a receipt is the date of purchase, which of the follow probably would NOT be included in part of the rules?  
a. Identify if there are any month names on the receipt (January, February, March, etc...).  
b. Look for the following format of numbers -- 19__ or 20__ -- where the underscores represent numerical values.  
c. Search for the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 anywhere on the receipt.
d. Locate any sets of digits with the following format -- ___/___/___ -- where the underscores represent numerical values.

23. The World Wide Web is full of unstructured data. Search engines like Google, Bing and Yahoo have been doing a good job of allowing users to search by key term in order to quickly locate links to websites about that particular topic. In order to do this, these search engines use what tool in order to help index and find these results?
   a. Spiderbots
   b. Query Robots
   c. Antcrawlers
   d. Index Finders

24. A popular restaurant likes to keep track of what food their patrons are ordering so that they can be better informed about what items they need to order in preparation for the next week. What would be the best way for the restaurant to organize this data with that end goal in mind?
   a. Record the exact name of the meal that was ordered (Dave’s Famous Wings, Southwest Bacon Burger, Grilled Chicken Panini, etc).
   b. Keep track of what was ordered by the category that they fall under (Entree, Dinner, Appetizer, Salad, etc).
   c. List what meat, vegetable and fruit each customer ordered (steak - corn - apples, fish - peas - peaches, etc).
   d. Record the cost of each meal purchased and whether the customer got the meal as-is or substituted some items.

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)
27. Social media sites like Facebook, Twitter and Instagram collect large amounts of data from their users every single day. Even after a user decides to leave the social media world and delete their account the database still maintains a record of the user. This phenomenon is known to computer scientists as what?
   a. Data breaches
   b. Cloud computing
   c. Filter bubble
   d. Data persistence

28. Different kinds of data analytics have differing levels of utility and confidence. In general, as the utility level increases, what happens with the confidence levels?
   a. The confidence level also increases.
   b. The confidence level stays the same.
   c. The confidence level decreases.
   d. The confidence level has no correlation to the utility level.

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 of height vs. age with stars indicating outliers](image)

The process above best describes what data mining strategy?
   a. 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

31. By using your own data, search engines and other sites try to make your web experience more personalized. However, by doing this, certain information is being hidden from you. This process of choosing to show you only certain, custom information over others puts you in a what?
a. Clustered circle  
b. Indexed environment  
c. Filter bubble  
d. Relational table  

32. Two parents are trying to figure out how tall their child might be by using a formula that was created based on studying large amounts of parents and children. This formula takes into account the heights of the parents along with other key factors. This is an example of what kind of analytics?  
a. Descriptive  
b. Predictive  
c. Prescriptive  
d. Correlative  

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  

34. A local high school recently won the girls volleyball championships and have been rewarded by the school with $5,000 to purchase merchandise for the team. The coach is trying to surprise the players with the merchandise at the banquet, so instead of asking them what sizes they want, he is trying to figure that out based on the information from the volleyball program. The following information is stored in the program for each player. 

- Name  
- Age  
- Grade  
- Height  
- Weight  
- Jersey number  
- Position  

What would be the best way for the coach to use this information in order to order sizes that work for the majority of the team?  
a. Filter the data by grade and order larges for the juniors and seniors, mediums for the sophomores and smalls for the freshmen.  
b. Sort the data by jersey number, ordering smaller sizes for the lower numbers and larger sizes for higher numbers.  
c. Filter the data by age and height, where the older and taller girls will get larger sizes and the younger and shorter girls will get smaller sizes.  
d. Sort the data by height and weight and order smaller sizes for the girls that are shorter and weigh less and order larger sizes for the girls that are taller and weigh more.  

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

Multiple Response

36. Which of the following are examples of unstructured data?  
Select two answers.  
a. Closed-circuit security footage of a bank lobby  
b. Digital Image scans of store receipts  
c. “Contact Lists” in smartphones  
d. Job application forms for a fast food restaurant

37. Which of the following is true when it comes to usefulness and usability of a data set?  
Select two answers.  
a. Data that is useful is always usable  
b. Data that is usable is always useful  
c. Data can be useful but not usable  
d. Data can be usable but not useful

38. Which of the following are examples of structured data?  
Select two answers.  
a. The glossary and index in the back of a textbook  
b. A video taken of a violent protest  
c. Letters written by young kids to Santa asking for Christmas gifts  
d. An address book filled with family members names and addresses

39. Which of the following are examples of how people give up some of their privacy in order to gain something in return (utility)?  
Select two answers.  
a. Students make sure to use the “Incognito” tab in a web browser so that none of the browsing history is saved and all cache is cleared at the end of the browsing session.  
b. People only use cash to pay for products since they do not have any credit cards.  
c. Customers sign up for “rewards” programs for different grocery stores so that they can get discounts on different items throughout the store.  
d. People enable GPS on their phone so that Apps can locate nearby stores, restaurants and hotels.

40. Students are using data collected from a non-profit organization in order to try to convince the school board that their school should be year-round with several week long breaks as opposed to the usual 9 months on and 3 months off. Information that was collected by this organization was as follows.
They decided to make an infographic in order to try to easily display the data they have analyzed. Which of the following would be the best information to put on their infographic to try to convince the school board of the change to the schedule?
Select two answers.

a. A cluster analysis of where year-round schools and regular schools are located geographically.
b. A regression analysis of standardized tests scores comparing the two different types of schooling.
c. An association rule showing links between motivation and happiness levels to the type of schooling students were receiving.
d. An automated summarization of student handbooks from year round schools, displayed in a word cloud.