

# Introduction to Data Science

Instructor: Daniel D. Gutierrez

## MODULE 5 QUIZ

### Question 1

Which of the following are typically part of the data access phase of a data science project?

- (a) Discovery process to find applicable data sources
- (b) Accessing data in more than one format from a variety of sources
- (c) Interacting with enterprise domain experts to determine available and reliable data sources
- (d) All of the above

### Question 2

In a data science project, the following are ways to describe a "clean" data set EXCEPT which one?

- (a) Well defined columns (feature variables)
- (b) Well defined rows (observations)
- (c) Missing values are permissible
- (d) Variable names (first row in the data set)

### Question 3

Which is not typically considered an aspect of the definition of "data"?

- (a) SQL used to select variables from a data set
- (b) Variables – a measurement or characteristic of an item
- (c) Qualitative item – city, gender, department, etc.
- (d) Quantitative item – home price, square feet, annual sales, etc.

### Question 4

What is not a typical data format used in data science projects?

- (a) Comma separated values

- (b) Microsoft Word
- (c) JavaScript Object Notation
- (d) SQL database

**Question 5**

What are the ways you can use SQL in R to process data?

- (a) Establish a connection to a SQL database where the data resides
- (b) Use the `sqldf` package in R to access data frames
- (c) Use one of many “SQL equivalents” in R
- (d) All of the above