Data 301: Final Review Questions

Note that these questions are sample questions based on the material covered in class only, and not necessarily representative of what you might see on your final.

We also have all of our previous material covering Excel, SQL, and Python in more depth on our website.

1. What is Data Analysis? What does a data analyst do? Why is data analytics important?

2. What is this in decimal in 1011 1110? In hexadecimal?

3. Which is bigger?
    a. 10 TB
    b. 100 GB
    c. 1,000,000,000,000 bytes
    d. 1 PB

4. Explain ”Big Data” and describe data growth in the coming years.

5. Compare and contrast: digital versus analog

6. Explain the role of metadata for interpreting data

7. How many of the following are true?
    - It is possible to have data without metadata.
    - Growth rates of data generation are decreasing.
    - It is possible to represent all decimal numbers precisely on a computer.
    - A character encoded in Unicode uses twice as much space as ASCII.

    A) 0 B) 1 C) 2 D) 3 E) 4

8. How many of the following statements about Excel VBA are true?
    - A macro is a recorded set of actions that can be executed repeatedly.
    - User defined functions can be used in formulas like built in Excel functions.
    - Macros are only created with VBA code.
    - Macros are only saved in *.xlsm* files.

    A) 0 B) 1 C) 2 D) 3 E) 4

9. Explain the difference between absolute and relative addressing. How do you use absolute addresses?

10. Cell A1 contains the following: =$B2+D$4. What is the formula if the cell is copied to cell D3?
    - error

- =$B2+D$4
- =$B4+F$4
- =$B4+G$4

11. Write an Excel formula to compute the average of the cells in the range E4 to E98.

12. When concatenating in Excel, what symbol do you use? & or + ?

13. Write an Excel formula that counts the number of cells in the range A2 to A 39 with a negative value.

14. How many of the following are true?
    - CONCATENATE function can take 3 arguments.
    - There is an Excel function that has 0 arguments.
    - =INDEX({1,3,5},2) returns 5.
    - =LOOKUP(5,{1,3,5},{"a","b","c"}) returns "c".

    A) 0 B) 1 C) 2 D) 3

The SQL Queries below use the database below. Note that your code should work if more rows are added to the tables below.

`emp` Table

eno	ename	bdate	title	salary	supereno	dno
E1	J. Doe	01-05-75	EE	30000	E2	null
E2	M. Smith	06-04-66	SA	50000	E5	D3
E3	A. Lee	07-05-66	ME	40000	E7	D2
E4	J. Miller	09-01-50	PR	20000	E6	D3
E5	B. Casey	12-25-71	SA	50000	E8	D3
E6	L. Chu	11-30-65	EE	30000	E7	D2
E7	R. Davis	09-08-77	ME	40000	E8	D1
E8	J. Jones	10-11-72	SA	50000	null	D1

`workson` Table

eno	pno	resp	hours
E1	P1	Manager	12
E2	P1	Analyst	24
E2	P2	Analyst	6
E3	P3	Consultant	10
E3	P4	Engineer	48
E4	P2	Programmer	18
E5	P2	Manager	24
E6	P4	Manager	48
E7	P3	Engineer	36

`proj` Table

pno	pname	budget	dno
P1	Instruments	150000	D1
P2	DB Develop	135000	D2
P3	Budget	250000	D3
P4	Maintenance	310000	D2
P5	CAD/CAM	500000	D2

`dept` Table

dno	dname	mgreno
D1	Management	E8
D2	Consulting	E7
D3	Accounting	E5
D4	Development	null

15. Write a query selecting all of the employees from the emp table with a salary greater than $20,000.

16. Write a SQL query that return all projects who have an employee working on them whose title is 'EE' with the following data. Assume eno and pno in workson are foreign keys to emp and proj respectively.

17. What is the output of the following query on the database above?

```
SELECT dno, MAX(salary)
FROM Emp
```

GROUP BY title;

18. Which is the correct query to find the average score per user from table Users
    a. SELECT AVERAGE(score)
       FROM Users
       GROUPBY score
    b. SELECT user, AVERAGE(score)
       FROM Users
       GROUPBY user
    c. SELECT AVERAGE(score)
       FROM Users
    d. SELECT AVERAGE(user)
       FROM Users
       GROUPBY score

19. Describe why R is a useful language.

20. Compare the syntax in R and Python. Note the assignment operators in R.

21. What is the difference between quantitative and qualitative data?

22. How do you calculate mean, median, variance, standard deviation, range?

23. How many of the following will successfully read in the "data.csv" file into your R program?

```
name = 'data.csv'
data = read.csv(name)
data <- read.csv('data.csv')
data = read.csv('data.csv')
attach(read.csv('data.csv'))
```

A) 0 B) 1 C) 2 D) 3 E) 4

24. What command can you use to view a dataframe in R?

25. Explain when you might use the R command na.omit() vs when you might use is.na().

26. How do you display the five number summary numerically in R? Graphically?

27. Which of the following is NOT true?
    a. Variance is always positive (or zero)
    b. If x < y, then quantile(x) <= quantile(y)
    c. The five number summary provides the median of the dataset.
    d. The standard deviation is always less than the variance.

28. Describe how you can concatenate a string and a numeric in R. Is this different from python? Why or why not?

29. Write an R program that changes the working directory to "*crocs*" (assume this is a directory inside your current working directory), and reads in "*crocs.csv*". There are 2 vectors in this csv, "*blue_crocs*" and "*green_crocs*", where each is a vector of rating out of ten. This data has been collected in a study to see if green crocs receive higher reviews than blue crocs. Conduct an appropriate t test and print the results.

30. Explain when you might use a paired versus unpaired t test.

31. What is a pvalue?

32. Write an R program that fits a linear model for the response variable *stress* as a function of the response variable *n_finals* and then plots the relationship. Assume that the vectors are already loading into your R program.

33. Write equivalent code to the one shown below using subset:

```
conditon = data$variable == 'keep'
new_data = data[conditon,]
```

34. How many of the following statements are true?
    a. Vectors in R are indexed from 0.
    b. 1:10 creates a vector of ten numbers.
    c. A vector may have elements of different data types.
    d. If data <- 1:5 then data[-1] returns 5.

35. How many of the following statements are true?
    a. Elements in a list may be of different types.
    b. In a matrix, the number of rows and columns must be the same.
    c. Given matrix m, m[2] would return an error.
    d. Given matrix m, m[,2] would return all data in column 2.

36. Define a matrix, vector, and scalar in R.

37. What command and arguments (this includes optional arguments) should you use to select certain columns of a dataframe?

38. The *transform()* function in R is similar to what SQL command? Why? Provide an example of both.

39. Describe an apply function and when you might use it.

40. Describe Open Science and why it is important.

41. Provide two examples of open datasets.

42. Define a condition

43. What is Python? What are the differences between Python 2 and 3?

44. What is the Output of the following:

```
n = 10, v = 5

1.  print(n == 2)
2.  print(n > v)
3.  print((n > v) and ((n + 2) == 13))
4.  print((n > v) or ((n + 2) == 13))
5.  print((n < v) and ((n + 3) == 13))
6.  print((not n < v) and (not n == v))
7.  print((not n < v) and (not (n + 1) != (v + 6))
```

45. Write a print statement for the following lines:
    "Data 301 covers a lot of subjects"
    "Thats okay though because I will rock the exam"
    "Thanks QSCU for these questions"
    BONUS: write it all in one print statement.

46. Write a python program that asks the user their name and their favourite kind of crocs. Then append their name and choice to a file called "kat_needs_this_for_science.txt". This should not overwrite any previous responses and take a new line for every entry, otherwise this data is useless to science.
    Example output:
    Cam - Fuzzy and blue
    Sam - pink floral print

47. Write a python program that asks in a loop for user input with the prompt ":)?". If the user inputs ":(" or ":/" then raise an Exception with the message of the user's input. Your program should continue to run if an exception occurs, but it should exit the loop and print "That's rough buddy" to the consol.

48.

You guys are going to rock this exam! :)