1 Review for DATA 301 Midterm

This is what the midterm looks like last year. I am not bound to the exact percentages, but they should give you a rough idea on how to study.

1.1 Previous Format

Time limit: 75 minutes
Total marks: 30

- $\sim$ 5 multiple choice (MC) questions (10 minutes total – 5 marks)
- $\sim$ 2 short answer questions (SA) (with parts) (10 minutes total – 5 marks)
- $\sim$ 2 long answer questions (LA) - about 20 minutes each (40 minutes total – 20 marks)

Topic Breakdown

- 13% - Data Representation: 2 MC, 1 SA
- 37% - Excel: 1 MC, 10 marks LA
- 3% - Excel VBA: 1 MC
- 37% - Databases: 1 MC, 10 marks LA
- ~~10% - Command line: 1 bonus MC, 1 SA~~ *since we haven't covered this the 10% will be distributed across the previous topics*

2 List of Topics

Table 1: Key

***	Extremely important
**	Assignment question or major topic
*	Important topic which probably should be tested
-	(no stars) topic covered but probably won't be tested
~~strikethrough~~	items will (definitely) not be covered

2.1 Introduction (01Intro)

* what is data analysis? what does a data analyst do?
* importance of data analytics

2.2 Data Representation (02DataRepresentation)

* Define: computer, software, memory, data, memory size/data size, cloud
* Explain "Big Data" and describe data growth in the coming years.
* Compare and contrast: digital versus analog
- Briefly explain how integers, doubles, and strings are encoded.
* Convert integer into unsigned binary

  ~~Convert real number into float~~
* Understand why ASCII table is required for character encoding.
* Explain why Unicode is used in certain situations instead of ASCII.
** Explain the role of metadata for interpreting data.
* Define: file, file encoding, text file, binary file
- ~~Encode using the NATO broadcast alphabet~~
- hexidecimal
- ~~Discuss the time-versus-space tradeoff.~~

2.3 Excel (03Excel part 1 and 2)

* Explain what a spreadsheet is and its usefulness
** spreadsheet cell addressing (range notation using :)
- selecting cells in a spreadsheet
- filling, hiding
* Define and explain: formula, function, argument, concatenation
** Using functions, eg. concatenate, lookup, index
*** compare absolute vs. relative addresses ; use absolute addresses
*** use aggregate functions
** use conditional formatting, format painter
** data and type formats
** Use sorting and filtering.
* Create and edit charts and use chart features: trendlines, sparklines
* Explain the usefulness of: what-if scenarios, goal seek, solver
*** Use and create pivot tables and charts.
** Evaluate and create conditions. Use IF() to make decisions.

2.4 Excel VBA (04 Excel VBA)

*** Explain how to create and use macros and macro recorder

** Explain the security issues with macros and how to handle them

- Create and use Excel variables

* Be able to read/manipulate VBA code

~~Write VBA code from scratch~~

* Explain how a collection is different from a typical variable

* Use/understand If/Then/Else syntax to make decisions

* Use For loop for repetition

** Explain how to create user-defined functions and use them in formulas

** Difference between subroutine and functions

- Objected-oriented definitions (object, class, property, method) and objects in Excel

- List some typical user interface controls

2.5 Relational Databases (05 Databases)

*** Given a small database write simple queries in SQL.

** Define: database, database system, schema, metadata

*** Define: relation, attribute, tuple, domain, degree, cardinality

* SQL properties: reserved words, case-insensitive, free-format

*** Be able to create a table using CREATE TABLE command

~~GUI commands in Microsoft Access and LibreOfficeBase (i.e Design View)~~

** Explain what a primary key is and what it is used for.

** Use DROP TABLE to delete a table and its data.

** Use INSERT/UPDATE/DELETE to add/update/delete rows of a table

* ALTER TABLE for adding columns

*** Execute queries using SQL SELECT

** SELECT DISTINCT for returning only unique values

** Sort rows using ORDER BY. Use LIMIT/TOP to keep only the first (top) N rows.

** Use GROUP BY and aggregation functions for calculating summary data.

** HAVING for filtering after GROUP BY.