CS102 Sample Midterm Problems

*The dataset used by most of the problems on the exam is provided at the **end** of the exam. Two copies are included; feel free to tear off the last two pages.*

The midterm will be in-class on Tuesday, October 31st at the time and place our class usually meets. You may bring 3 double sided 8.5"x11" pages of notes to the midterm with any information you'd like on them.

Your exam will cover material and assignments through the fourth week of class, including the Overview material in week 1 but not including the guest lecture in week 4. This sample is based off of exams from previous iterations of the class so does not include questions on all of the topics we've covered. Note for example there are no sample questions that ask you to write code, which may be something we ask you to do on the midterm (on paper, not executed).

Additionally, you should expect the actual midterm to be much longer than this set of problems. We do not expect everyone to finish the midterm within the allotted time. You should come prepared to manage your time wisely, focusing on the problems you are more comfortable with first.

Problem 1 (5 points) - *Spreadsheets*

Suppose the sample data is inserted into a spreadsheet in columns A (*name*), B (*year*), and C (*course*), including the header. The following formula is put into cell D2 (i.e., the top cell of column D below the header row):

```
=unique(A2:A17)
```

Then the following formula is put into cell E2 and extended to the entire column:

```
=countif(A2:A17,D2)
```

Show the contents of columns D and E.

D	E

Problem 2 (5 points) - *Spreadsheets*

Suppose the sample data is inserted into a spreadsheet as in Problem 1, then a Pivot table is created. In the Pivot table, field *course* is added to Rows, field *year* is added to Columns, and field *name* is added to Values summarized by COUNTA. “Show totals” are turned off. Show the structure and contents of the Pivot table.

Problem 3 (5 points) - *SQL*

Suppose the sample data is loaded into a relational database table *Took*(*name, year, course*), and the following query is run:

```
Select course, count(*)
From Took
Group By course
Order By count(*) desc, course
```

What is the query result?

Problem 4 (5 points) - *SQL*

Now suppose the following query is run on the original table:

```
Select Distinct name
From Took C1
Where Not Exists
 (Select * From Took C2
  Where C1.name <> C2.name and C1.year = C2.year and C1.course = C2.course)
```

What is the query result?

Problem 5 (5 points) - *Correlation*

In the sample data, does there appear to be a correlation between values for *year* and *course*? Answer a definite YES or NO, then justify your answer quantitatively using techniques from class/homework. *Hint:* Your answer for Problem 2 above may be useful in your analysis.

Problem 6 (10 points) - *Python*

Suppose the sample data is stored in a CSV file called *Midterm.csv* (with the header line), and the following Python program is executed. After each set of print statements, which are in bold font for your convenience, put the program's output in the box provided. The program's variables continue from section to section, and not that the problem spans two pages. Feel free to omit the single-quotes that Python prints around string values in dictionaries and lists.

```
import csv # standard reading from CSV file into list of dictionaries
took = []
with open('Midterm.csv', 'rU') as file:
    data = csv.DictReader(file)
    for d in data:
        took.append(d)
print took[0]
print took[len(took)-1]
```

```
names = []
for c1 in took:
    for c2 in took:
        if c1['name'] == c2['name'] and c1['course'] == '102' \
        and c2['course'] == '107':
            names.append(c1['name'])
for n in names: print n
```

```
juniors = []
seniors = []
for c in took:
    if c['year'] == 'junior': juniors.append(c['name'])
    elif c['year'] == 'senior': seniors.append(c['name'])
print 'JUNIORS:', juniors
print 'SENIORS:', seniors
```

Problem 7 (5 points) - *Tableau*

The following dataset contains eight items representing colored points on the x-y plane:

x	y	color
1	1	red
1	3	green
2	5	blue
3	5	green
4	1	blue
4	4	red
5	3	blue
5	4	green

Suppose you import this data into Tableau, then:

- Add `color` to Columns
- Add `MIN(x)` to Rows
- Add `MIN(y)` to Rows
- Select chart type `side-by-side bars`

Show a sketch of Tableau's visualization:

The following sample data is used for all of the problems on the exam except the last one. It contains information about five students (assume their names are unique), what year they are, and what courses they've taken so far. There are 16 rows altogether.

We're including two copies of the same data for your convenience.
Feel free to tear off these two pages.

name	year	course
Amy	junior	101
Amy	junior	102
Amy	junior	106A
Amy	junior	106B
Bob	junior	106B
Cam	junior	102
Cam	junior	106A
Cam	junior	107
Dan	senior	101
Dan	senior	102
Dan	senior	106A
Dan	senior	106B
Dan	senior	107
Eve	senior	102
Eve	senior	106A
Eve	senior	106B

The following sample data is used for all of the problems on the exam except the last one. It contains information about five students (assume their names are unique), what year they are, and what courses they've taken so far. There are 16 rows altogether.

We're including two copies of the same data for your convenience.
Feel free to tear off these two pages.

name	year	course
Amy	junior	101
Amy	junior	102
Amy	junior	106A
Amy	junior	106B
Bob	junior	106B
Cam	junior	102
Cam	junior	106A
Cam	junior	107
Dan	senior	101
Dan	senior	102
Dan	senior	106A
Dan	senior	106B
Dan	senior	107
Eve	senior	102
Eve	senior	106A
Eve	senior	106B