W01 – Tools in Your Career

CASE STUDY: A look at how your career may make use of databases.

So why do you need this class for your career?

Data Science or Business Analysts majors might be thinking I'm good with R or Tableau or PowerBI and that is where I'll analyze data. Why do I need to know about databases? Or if you are an Accounting, Business, or Economics major you might be very knowledgeable about Spreadsheets and wonder what databases can add. Or maybe Geospatial Science majors might know that arcGIS is going to be the tool they use. How do databases work with that? CIT, CS or Software Engineering wonder how databases fit with what they are learning about with programming languages like Python or PHP.

Let's look at how databases fit with a few of these careers.

A tool **Geospatial Science** majors will use is ArcGIS, which is a mapping software that allows analysis to better understand data. But in the background of this tool are relational databases feeding the data into ArcGIS. You may or may not have to deal directly with this database but understanding where the data is coming from and being able to collect or combine data you receive into a relational database is very important.

Here is a look at ArcGIS:
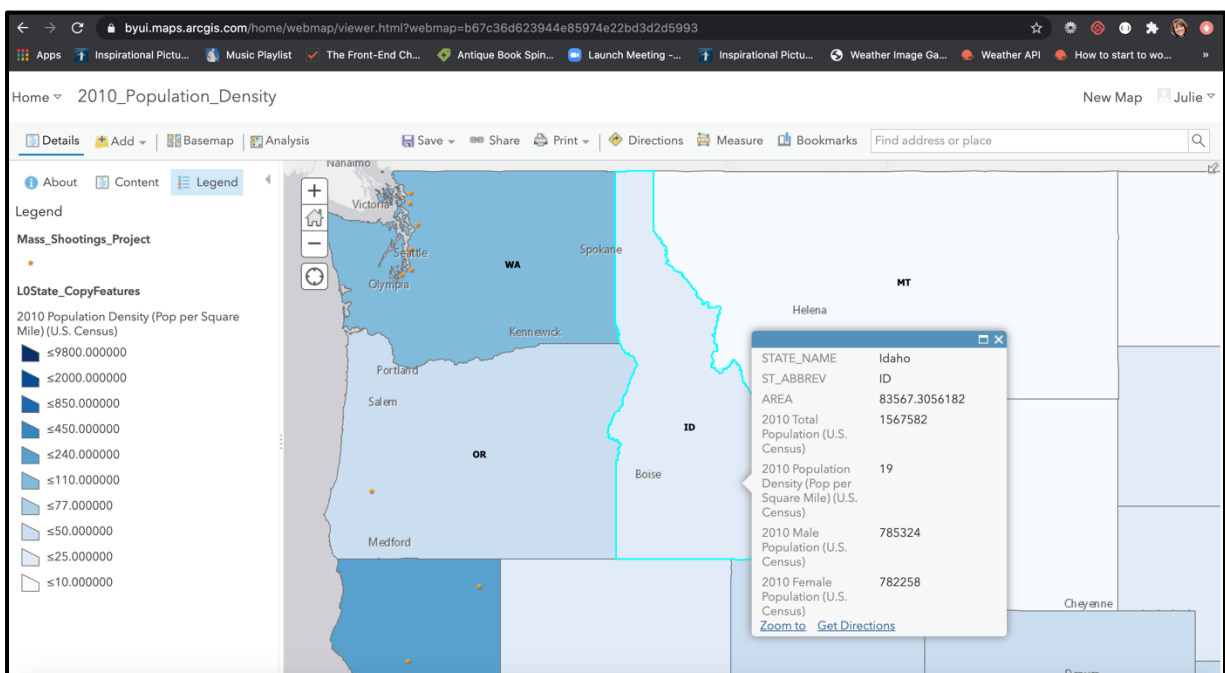
For **Data Science**, there are great analysis tools such as R, Tableau or PowerBI that many of you will use in your career. They do a great job of visually representing the data being analyzed and building statistical models for prediction. The data you use to analyze in these tools may very well come from a database. Knowing how the data is structured will be very beneficial. Also knowing how to get the data sets out of a database is important. Once you retrieve the data you need from a very large database you can take that data set and use it in one of these tools to analyze it. Even with R libraries that do similar data manipulations as SQL, a data science system may not have the memory resources to be able to handle these vast amounts of data. It is much more efficient to use SQL for the power hungry and tedious data operations and is quicker in most cases.

"The best strategy I keep seeing with many R users is to fine tune the SQL queries to extract the data at a manageable size for your PC's memory either by filtering, aggregating, or sampling, then import into R instance in memory so that you can quickly and iteratively explore and analyze the data with all the statistical horse powers." Kan Nishida

Here's a look at R:
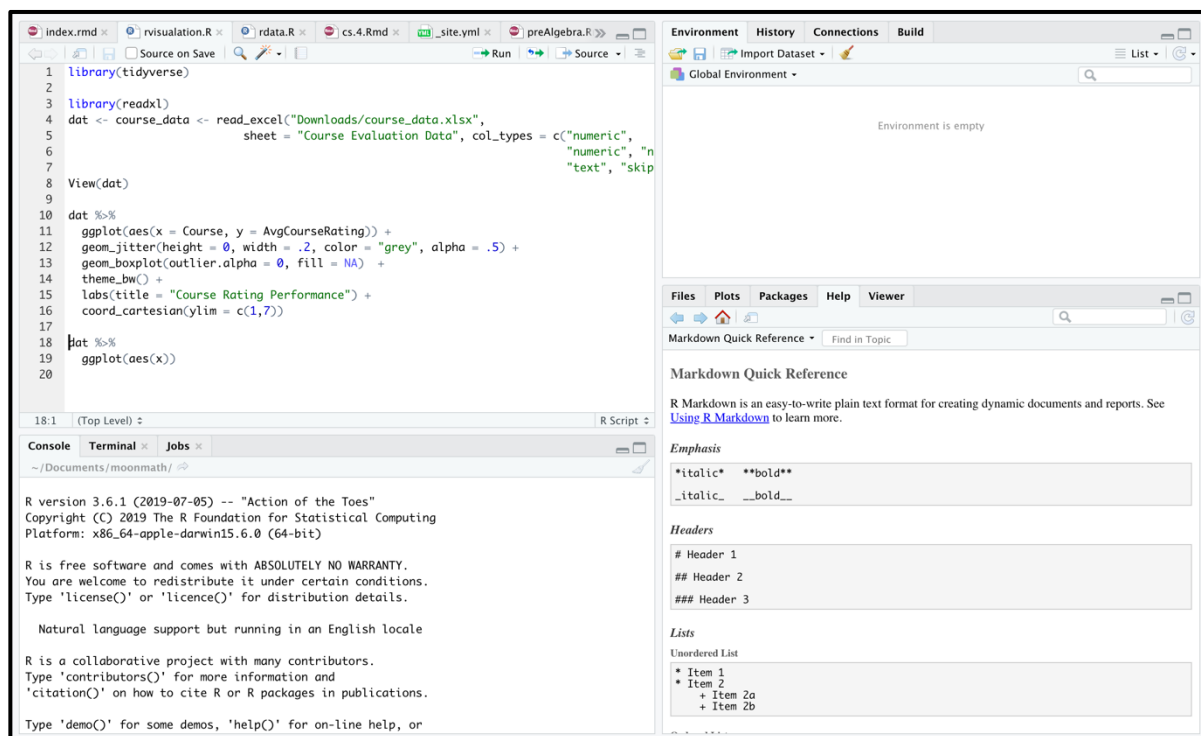
Spreadsheets are a great tool for many in **business** areas of focus. When it comes to Spreadsheets vs. Databases, you have a video to watch this week that will explain the difference. But they are two very different tools, and each has their purpose, but spreadsheets are not meant for large amounts of data storage, they are for number crunching, not data retrieval from a large amount of data. You will need both.

At times you will be taking data from spreadsheets to place in a database or vice versa. Here is a CSV (comma separated values) file that can transfer easily back and forth from a spreadsheet program to a database.

```
Users > user > Downloads >  biostats.csv
 1   "Name",     "Sex", "Age", "Height (in)", "Weight (lbs)"
 2   "Alex",      "M",    41,       74,          170
 3   "Bert",      "M",    42,       68,          166
 4   "Carl",      "M",    32,       70,          155
 5   "Dave",      "M",    39,       72,          167
 6   "Elly",      "F",    30,       66,          124
 7   "Fran",      "F",    33,       66,          115
 8   "Gwen",      "F",    26,       64,          121
 9   "Hank",      "M",    30,       71,          158
10   "Ivan",      "M",    53,       72,          175
11   "Jake",      "M",    32,       69,          143
12   "Kate",      "F",    47,       69,          139
13   "Luke",      "M",    34,       72,          163
14   "Myra",      "F",    23,       62,           98
15   "Neil",      "M",    36,       75,          160
16   "Omar",      "M",    38,       70,          145
17   "Page",      "F",    31,       67,          135
18   "Quin",      "M",    29,       71,          176
19   "Ruth",      "F",    28,       65,          131
```

Computer Information Technology, Computer Science and Software Engineers

will need to be able to use databases with their systems they create. Whether it's a web-based application or a stand-alone system, you will need to understand the database structure and how to retrieve the right information from the database so the users of your application can also see that information.

Here is some code that is making a connection to a database so it can then be used in the PHP programming language:

```php
1   <?php
2   function dbConnect($usertype) {
3           $host = 'byuiwebdevcom.ipagemysql.com';
4           $db = 'art';
5           if ($usertype == 'read') {
6                   $user = 'artread';
7                   $pwd = "Artuser20!";
8           } elseif ($usertype === 'write') {
9               $user = 'artwrite';
10              $pwd = 'ArtG20!';
11          } else {
12              exit("Unrecognized user");
13          }
14
15          $conn = new mysqli($host, $user, $pwd, $db);
16
17          if (mysqli_connect_errno()) {
18                  echo "<p class='error'>Could not connect to database.<br/>
19                      Please try again later.</p>";
20          }
21
22          return $conn;
23  }
```

Here's a look at a SQL query inside of PHP code.

```php
1   <?php
2   require '../includes/title.php';
3   if (isset($_GET['go'])) {
4       require_once "../includes/connection.php";
5       $conn = dbConnect('read');
6       //perform a db query
7       $query ="SELECT title, file, year, period, type, location, donated, keyword
8          FROM artwork
9          JOIN keybridge ON artwork.artwork_id = keybridge.artwork_id
10         JOIN keyword ON keyword.keyword_id = keybridge.keyword_id
11         WHERE keyword LIKE ?";
12      $searchterm = '%'.$_GET['search'].'%';
13      $stmt = $conn->stmt_init();
14      $stmt = $conn->prepare($query);
15      $stmt->bind_param('s', $searchterm);
16      $stmt->execute();
17      $stmt->bind_result($title, $file, $year, $period, $type, $location, $donated, $keyword);
18      $stmt->store_result();
19      $numRows = $stmt->num_rows;
20  }
21  ?>
22  <!DOCTYPE html>
23  <html>
24    <head>
25      <meta charset="UTF-8">
26      <meta name="viewport" content="width=device-width, initial-scale=1.0">
27      <title>Artitorium - <?php echo $title; ?></title>
28      <link rel="stylesheet" type="text/css" href="../styles/styles.css">
29    </head>
30    <body>
31    <?php
32    if (isset($error)){
33      echo "<p>$error</p>";
34    }
35    ?>
36      <div id="wrapper">
```

The users can then input what they need from the database, the code takes their input and retrieves what they need from the database and gives it to them in a user-friendly manner.

Databases will definitely play a part in each of your careers and are important for you to understand. You may not create the database, your business may already have a database in place, or someone else may create it, but knowing how the data is structured and how to retrieve the data you need is crucial.

Mark Evans, the Director of Enterprise Strategy, Innovation and Transformation Center for the church came to our class and here are some of his thoughts:

- SQL is more important than R or PowerBi because they have their limits. (Analyzing millions of rows of data doesn't work in R)

- All data will eventually go back to a database, (or come from a database)

- SQL is your most important class. SQL is king with any job having to do with data. Data Science, Data Analysis, Business Analytics, etc.

- And problem-solving skills are even more important than SQL. You can google any syntax or functions to use with SQL but you can't always google the actual problem you have to solve.