**Lab 2**

**Manipulating Data Using Access (and Excel)**

In this lab, you will practice how to carry out various kinds of queries using Access, with a focus on aggregating data to get key stats and proportions. You also get familiar with the ways to import and export data between Access and Excel, so as to leverage strength of each tool to best perform the tasks.

**Setting Up the Stage**

* Create a new Access file and import data (Beijing PM 2.5 2016) as a table.
  + Make the first row into column headers and number fields as Integer

**Preview Data and Check Quality Issues**

* Show all values in the table
* Show the following columns: Date, Value, and QC Name
* Apply the following condition: Value <0
* Sort result from high to low by Value
* Uncheck the Show box for Date to exclude it from the query, and switch to SQL View to review the SELECT statement as generate by Access
* Add the DISTINCT keyword after the SELECT keyword, and switch back to Datasheet view to see the results
* Save the query as DistinctNegativeValues

**Apply Aggregation and Scalar Functions**

* Select Month, QC Name, and ID,
  + Check results to make sure only three columns got selected, and then
  + Check the SELELCT statement as generated in SQL View
* In Design View, press the Totals (∑) button, and in the newly added Total line, select
  + Group by for Month and QC Name; Count for ID
  + Take a screenshot of the result and describe what the set shows

|  |  |
| --- | --- |
| Screenshot | Description |
|  | This shows the amount of missing and valid readings for each month. |

* Now copy the following expressions to the last two Fields (used to contain QC Name and ID)   
  ValidCount: Sum(IIf([QC Name]="Missing",0,1)), and  
  MissingCount: Sum(IIf([QC Name]="Missing",1,0)) respectively, and select Expression for the two columns in the Total line.
  + Take a screenshot of the result and describe what the set shows, especially on the purpose of the scalar function IIf (or Immediate If)

|  |  |
| --- | --- |
| Screenshot | Description |
|  | This shows the count of missing and valid readings for each month. The IIF statements test to see if the argument is true or false. If true, the first number is returned, if false the last number is returned. |

* Now add the following expressions to the next two Fields (after the MissingCount Field)   
  Total: [MissingCount] + [ValidCount] , and  
  %: [MissingCount]/[Total] respectively, and select Expression for the two columns in the Total line.
  + Take a screenshot of the result and describe what the set shows. Make sure to set Format to percent and Decimal places to 2 on the Properties Sheet for %.

|  |  |
| --- | --- |
| Screenshot | Description |
|  | This set shows the missing and valid counts, the totals for every month as well as the percentage of missing values for each month. |

**Comparing Data By Month and By Hour**

* Show Min, Max, Avg, and StDev for Value for the year 2016, excluding entries with a negative Value. Create two queries to
  + Show the stats by Month, and by Hour, respectively; and describe what you see.

|  |  |
| --- | --- |
| Screenshot | Description |
|  | This shows the min, max, average, and std deviation values by month. It appears these values are higher in the winter months compared to the spring/summer months. Perhaps due to heating homes. |
|  | This shows the min, max, average, and std deviation values for each hour from the data set. It appears these values are higher in the late afternoon and evening hours when more people are home. |

**Exporting to Excel**

* Export your queries for stats by Month and Hour into Excel. Format the tables properly.
* Use Bar charts as appropriate to visualize the results and support your conclusions.

**Due by Wed, 1/31, before class.**