

# **CIS4560-01 Term Project Tutorial: Team 2**

#### **Authors: Journey Mendoza; Anthony Gonzalez; Kevin Danao; Jonathan Diaz**

#### **Instructor:** [**Jongwook Woo**](https://www.linkedin.com/in/jongwook-woo-7081a85)

#### **Date: 05/23/2021**

**Lab Tutorial**

[jmend161@calstatela.edu](mailto:jmend161@calstatela.edu), [agonza230@calstatela.edu](mailto:agonza230@calstatela.edu), [kdanao@calstatela.edu](mailto:kdanao@calstatela.edu), [jdiaz231@calstatela.edu](mailto:jdiaz231@calstatela.edu)

05/23/2021

**COVID-19 Vaccination Data & Hesitation**

**Objectives**

In this hands-on lab, you will learn how to:

* Filter and Sort Data in Pig
* Utilize Excel to Import CSV Files
* Visualize GeoSpatial Data with Excel 3D Maps/Power BI
* Create Charts Based on GeoMap
* Analyze Results

**Platform Spec**

· Korean On-Premise Server

· CPU Speed: 3GHz

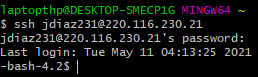
· # of CPU cores: 12

· # of nodes: 3

· Total Memory Size: 192GB

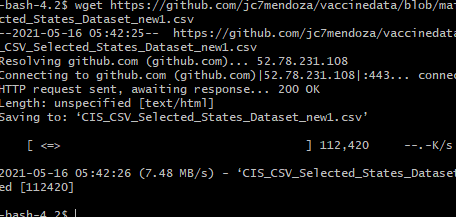
**1. Pig Tutorial to order by ascending and descending**

1. Open a shell and ssh into the server.



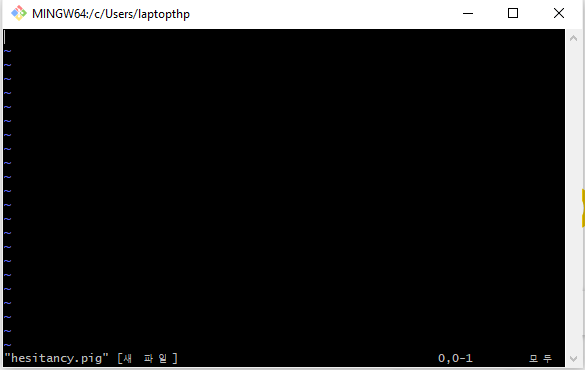
ssh yourusername@220.116.230.21

2. Download date into the server using either SCP or WGET.



wget https://github.com/jc7mendoza/vaccinedata/blob/main/CIS\_CSV\_Selected\_States\_Dataset\_new1.csv

3. Using the vi editor we will create a pig script file to write code in step 12.



vi pig\_script.pig

4. Insert the following into the text editor by pressing (i) on your keyboard to enable editing in the vi editor.

hesitancy\_dirty = LOAD '/user/jdiaz231/project1/CIS\_CSV\_Selected\_States\_Dataset\_new1.csv' using PigStorage(',');

hesitancy\_clean = FOREACH hesitancy\_dirty generate (chararray) $1 as County, (chararray) $2 as State, (double) $3 as EstimatedHesitancy, (double) $4 as StrongHesitancy;

StateHesitancy\_Descending = ORDER hesitancy\_clean BY EstimatedHesitancy DESC;

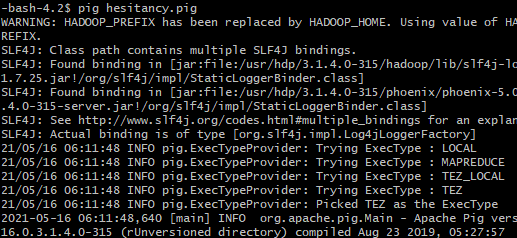
DUMP StateHesitancy\_Descending LIMIT 5;

StateHesitancy\_Ascending = ORDER hesitancy\_clean BY EstimatedHesitancy ASC;

DUMP StateHesitancy\_Ascending LIMIT 5;

Edit the file path to your username and file path of the CSV file.

5. Run the pig script.

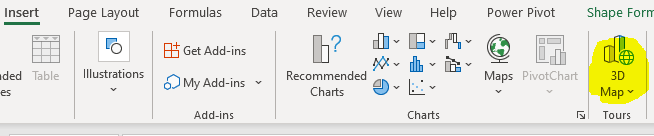


The script should produce 5 rows of States vaccine hesitation rate in descending order followed by 5 rows of states in ascending order of rate.

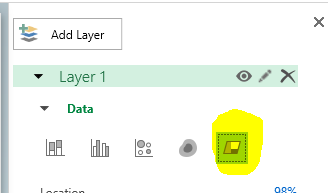
**2. Visualization: Geomap Tutorial in Excel 3D Map/Powermap**

1. Open the vaccine hesitancy CSV file with Excel. Excel will prompt you to save it as a workbook which we did.

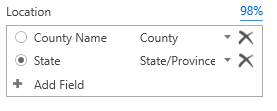
2. Select the “3D Map” option highlighted in yellow within the insert tab in the menu bar.



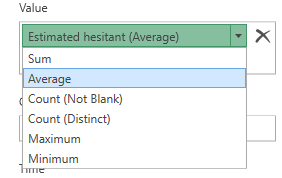
3. In the 3D Maps editor which is now open, look to the far right and select the data visualization to the region option indicated by the highlighted option below.



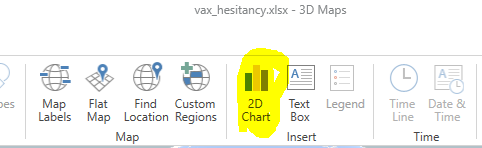
4. In the same pane, drag “State” from the Field List into the Location box or add the field and select it. 3D Maps may preload County and State into the box. Make sure “State” is selected.



5. Drag or add the “estimated hesitant” field into the Value box. Then select the drop-down menu for the field and select average.

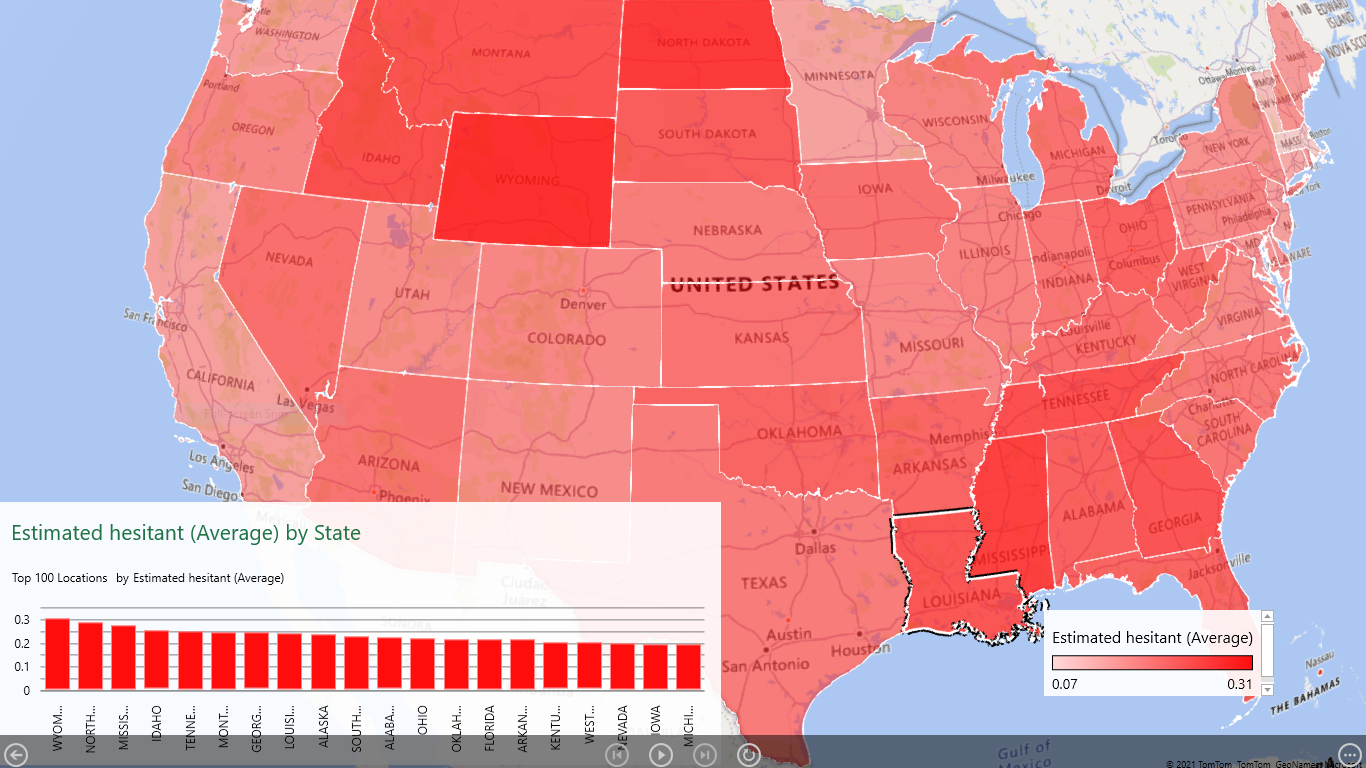


6. Now that the map is complete, we can add go to the menu or ribbon bar and select the 2D Chart option.



7. Like the previous step, select the Map Labels option to enable State names on the Map.

**Result:** The result shows the estimated hesitancy of each state by different shades or red. The darker shades indicate a higher hesitancy rate for that state. We can see the same information in the 2D Chart from highest hesitancy rate to least.

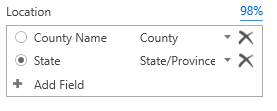


**Adding a Layer Showing Fully Vaccinated People By State**

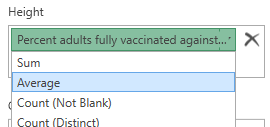
8. Return to the same Layer Pane worked with in the previous steps and select Add Layer.



9. Set the Location box to State once again for this new layer.

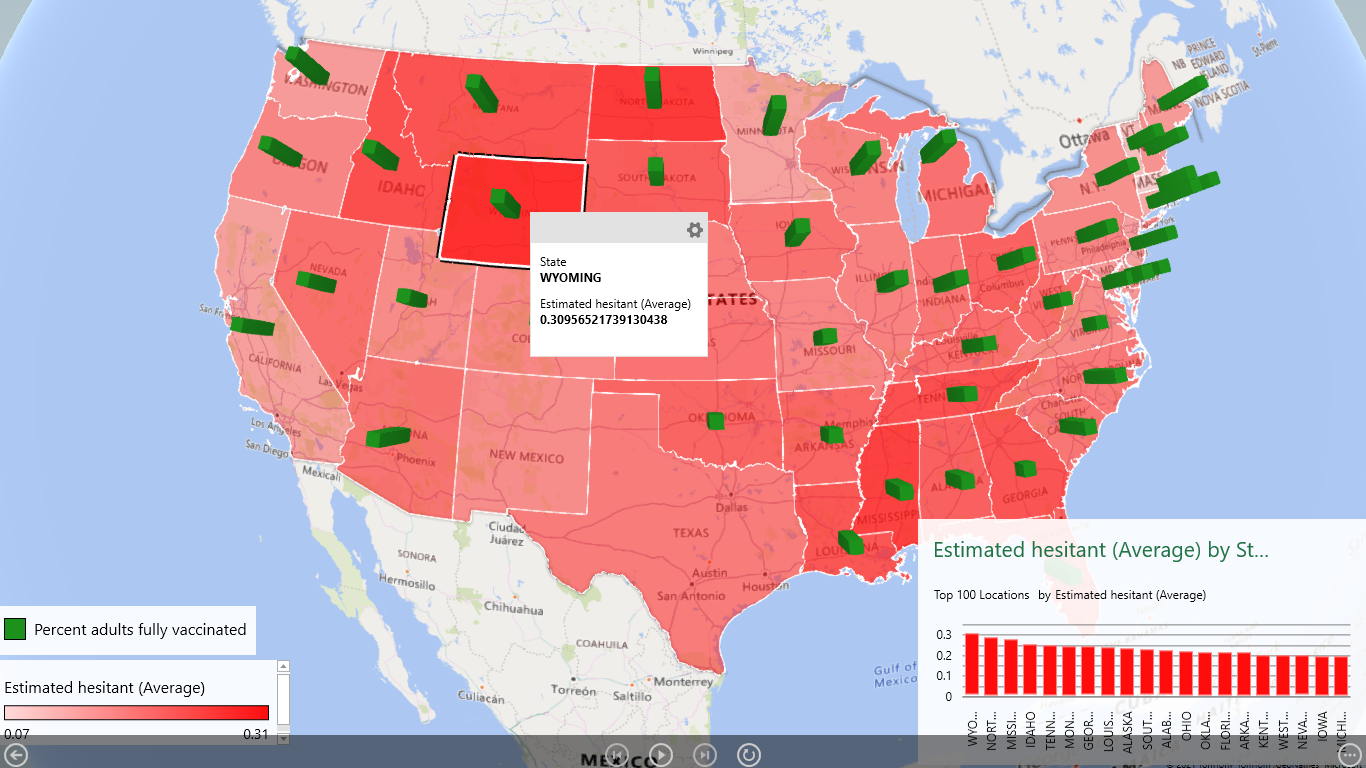


10. For the Height box, drag or add the “Percent Adults Fully Vaccinated Against COVID-19” field and then select the drop-down menu partially covered by its label. Set it to average.



**Final Result:** We have the first layer showing COVID-19 vaccination hesitancy rates by state, and have just added the bars with the fully vaccinated adult percentages. Hovering over the red

or green provides the rates or percentages depending on which you select.



*Note: Texas and New Mexico are missing relevant vaccination data.*