CIS-4560 Term Project Tutorial

Authors: Michael Brumwell

Instructor: Jongwook Woo

Date: 12/12/2019

Kibana Lab Tutorial

Michael Brumwell (mbrumwe@calstatela.edu)

12/12/2019

Flu Clinic Personnel using Kibana

Objectives

List what your objectives are. In this hands-on lab, you will learn how to:

- Import Data to Kibana

- Visualize with Kibana

- Create Geographic map using Excel

Platform Spec

- Windows 10 (Local Host)
- CPU Speed: 3.6 GHz
- # of CPU cores: 8
- Total Memory Size: 21 GB

Importing Data to Kibana

*This lab assumes you are using the client version of Kibana and that you know how to run it

-First you will need to download the hc_flu_vaccination csv data

-Access Kibana

-On the Right-Side toolbar, Select Machine Learning (Between Maps and Infrastructure)

-Under "Import Data" Select Upload File

-Drag the data file to the Kibana Screen

From there you will be asked to confirm the data. Following that you will be asked to create a new index for the data. Give it the name **flu_hospital**. Now that we have uploaded and indexed our data, we will be able to use it for the following visualizations.

Bar Graph

For our Bar Graph we will be looking at the top 5 clinics with vaccinated personnel by county. We will only be using a select number of clinics in order to make the graphs look neat and easy to find the data we're looking for.

To start, click on **visualizations** in the right tool bar, then select **Vertical Bar Graph**. You will be asked to select an indexed data set, select the **flu_hospital** index you created at the start of this tutorial.

On the left side of the screen under Metrics, look at the fields under **Y-Axis**. Ensure te following fields are set correctly:

-**Aggregation** should be set to **Count**

*The Y axis is going to represent the total number of clinics, so we are going to give it the **Custom label** of **Number of Clinics**

Now we are going to organize our data into buckets. To do so, look at the **Buckets** section underneath the **Metrics** Section and click **add**. A menu will appear with 3 selections; **Select X – axis.**

The X-axis is what we're going to use to divide our clinics by county. For the Fields of our new **X-axis** ensure the following:

Aggregation – Terms ("-" ---> **is set to**)

Field – County

Order By – Metric: Number of clinics

Order – Descending

Size - 5

Now we currently have a bar graph with a Y-axis showing the number of clinics and the X-axis separated by **top 5** counties with the **greatest number of clinics**. To include the percentage of vaccinated personnel we will do the following:

-Under our X-axis, click **ADD** then click **Split Series**

-Ensure the following:

-Sub aggregation – Terms

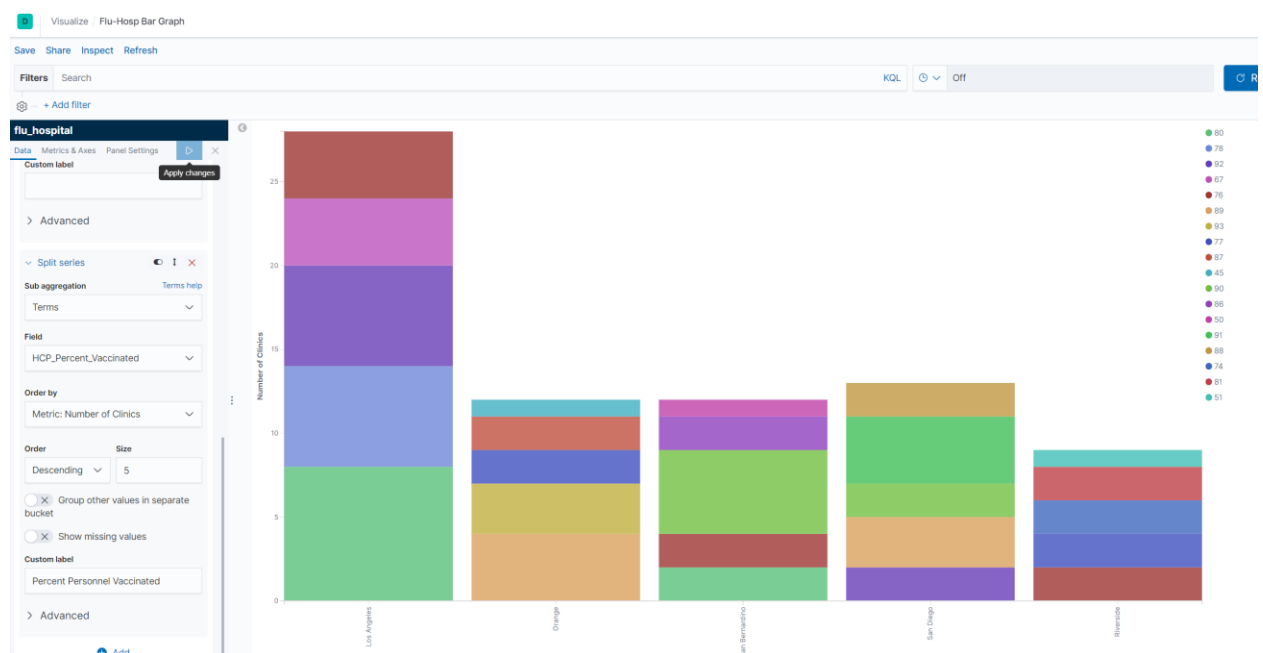-Field – HCP_Percent_Vaccinated

-Order by – Metric: Number of Clinics

-Order – Descending

-Size - 5

-Custom Label – Percent Personnel Vaccinated

-Save as Flu-Hosp Bar Graph

-results should look like something below

Pie Graph

For this pie graph, we are going to check the clinics with the lowest 5 percentage of vaccinated personnel.

-Create a new Pie Graph using our hospital index

-Under the Metrics block, ensure the following:

    -Under Slice Size:

        -Aggregation is set to Count and

        -Custom label; is set to Number of Clinics

Under buckets, we are going to divide our slices by selecting Split Slices. Then do the following:

-Aggregation – Terms

-Field – County

-Order By – Metric: Num of Clinics

-Order – Desc

Size – 7

Custom Label – Number of County Clinics

Now we have a slice with a single ring, representing the top 7 counties with the greatest number of flu clinics. Now we are going to further visualize our data to determine the clinics with the least amount of vaccinated personnel. To finish visualizing the data, we are going to do the following:

-Create a new sub-bucket using the Split Slices option

-Set Sub aggregation to Terms

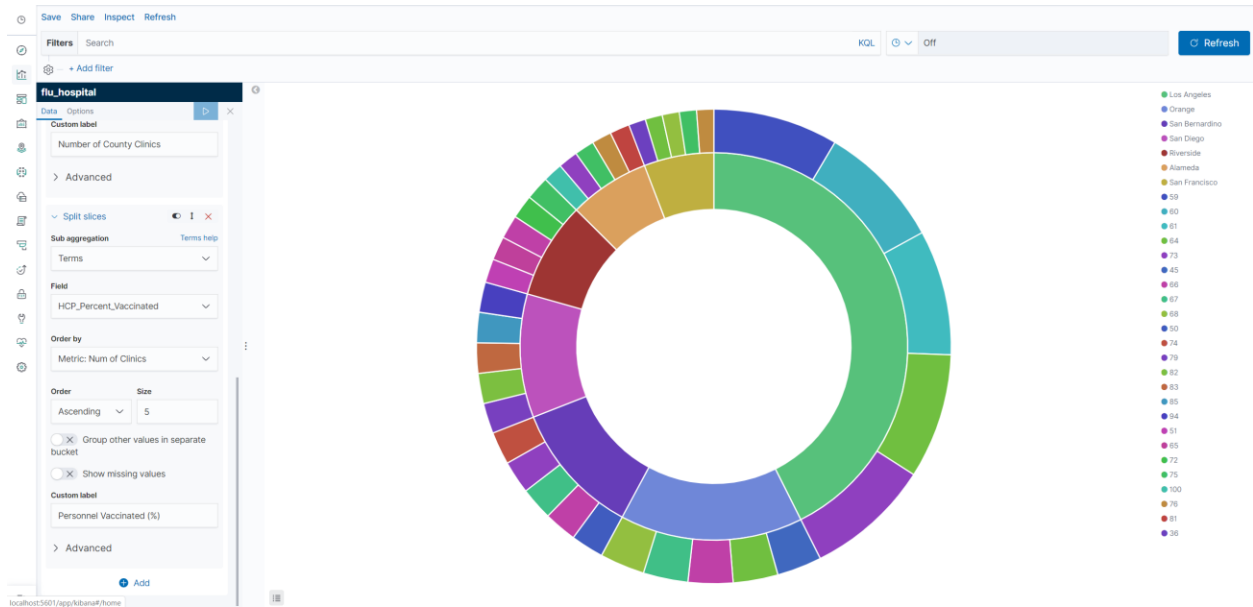-Field – HCP_Percent_Vaccinated

-Order by Metric:Num of Clinics

-Order – ASC

-Size 5

-Custom Label – Personnel Vaccinated (%)

The result should look like this:

Save the graph as Percent Vaccinated Personnel by County

Bar graph 2

Create a new Vertical Bar Graph using the hospital index

Under Y-axis, set the following:

        Aggregation – Count

        Custom Label – Total Number of Clinics

Now we separate our clinics by county by doing the following:

>   -Click the ADD button bellow our Y-axis and click X-axis

>   -Set the following:

>>   -Aggregation – Terms

>>   -Field – County

>>   -Order By – Metric: Total Number of Clinics

>>   -Order – DESC

>>   -Size 5

>>   -Custom Label - County

In order to further determine clinic eligability, we need to further organize our data by seperating the chart into parts.

>   -Add a new sub-bucket and click Split Chart

-In the row that says Rows and Columns, Select Rows

-Sub aggregation – Terms

-Field – On Track To Achieve 90% by 2020

-Order By - Metric: Total Number of Clinics

-Order – ASC

-Size 5

Now we have our graph seperated by county, and seperated into 3 graphs seperated by yes, no, and 'DNR' but we still want to see the percentage of people that have been vaccinated.
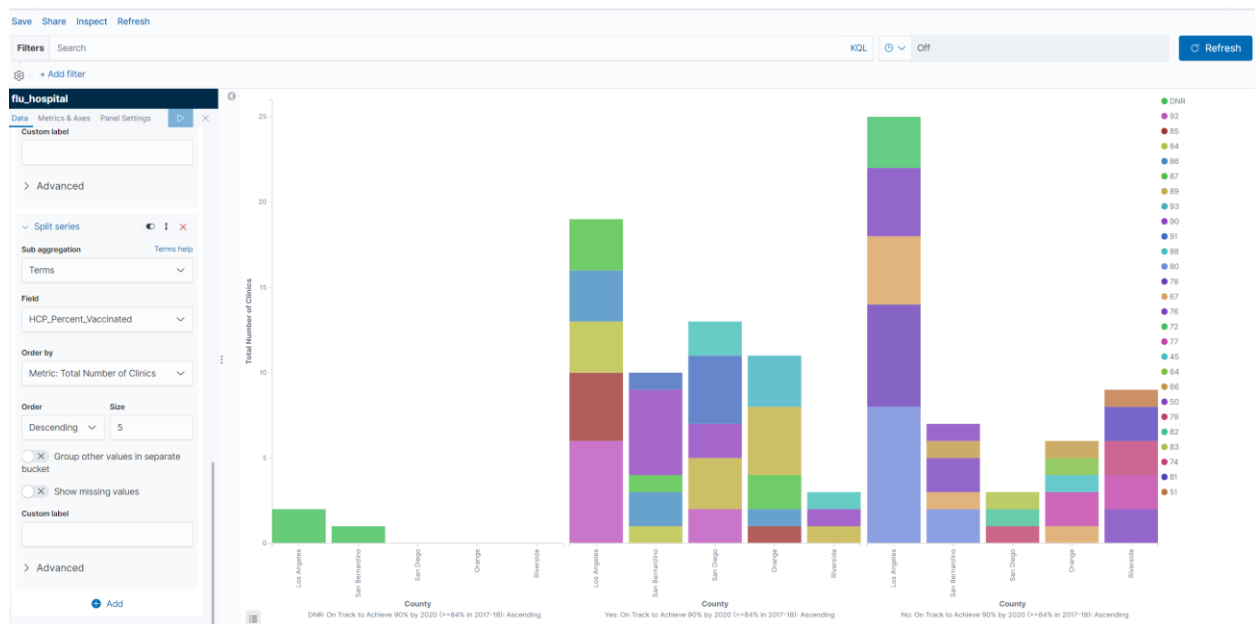
-Create a Split Series sub-bucket

-Sub-aggregation – Terms

-Field – HCP_Percent_Vaccinated

-Order By – Metric: Total Number of Clinics

-Order – DESC

-Size 5

Save the graph as 'clinics_on_track bar graph'

Dashboard

Now that we have our 3 graphs, we can use kibana to organize them in a dashboard for easy access.

-To create a dashboard, click the dashboard icon on the left toolbar of kibana.

-In the top left corner of the screen, click add to create a new dashboard

A list will appear on the right side of the screen showing existing visualizations

-Click the three visualizations we created earlier in the tutorial and **save** as hospital dashboard. Results should look like below:

Geospatial Map with Excel

Because the data we are using doesn't contain geo hash coordinates, we are unable to create a coordinate map using Kibana. Since the data includes county names in California, we can use that data to create a map of California.

-First, open the data in Excel

-Highlight the data by left-clicking the first field of the first column and dragging it until every record is highlighted.

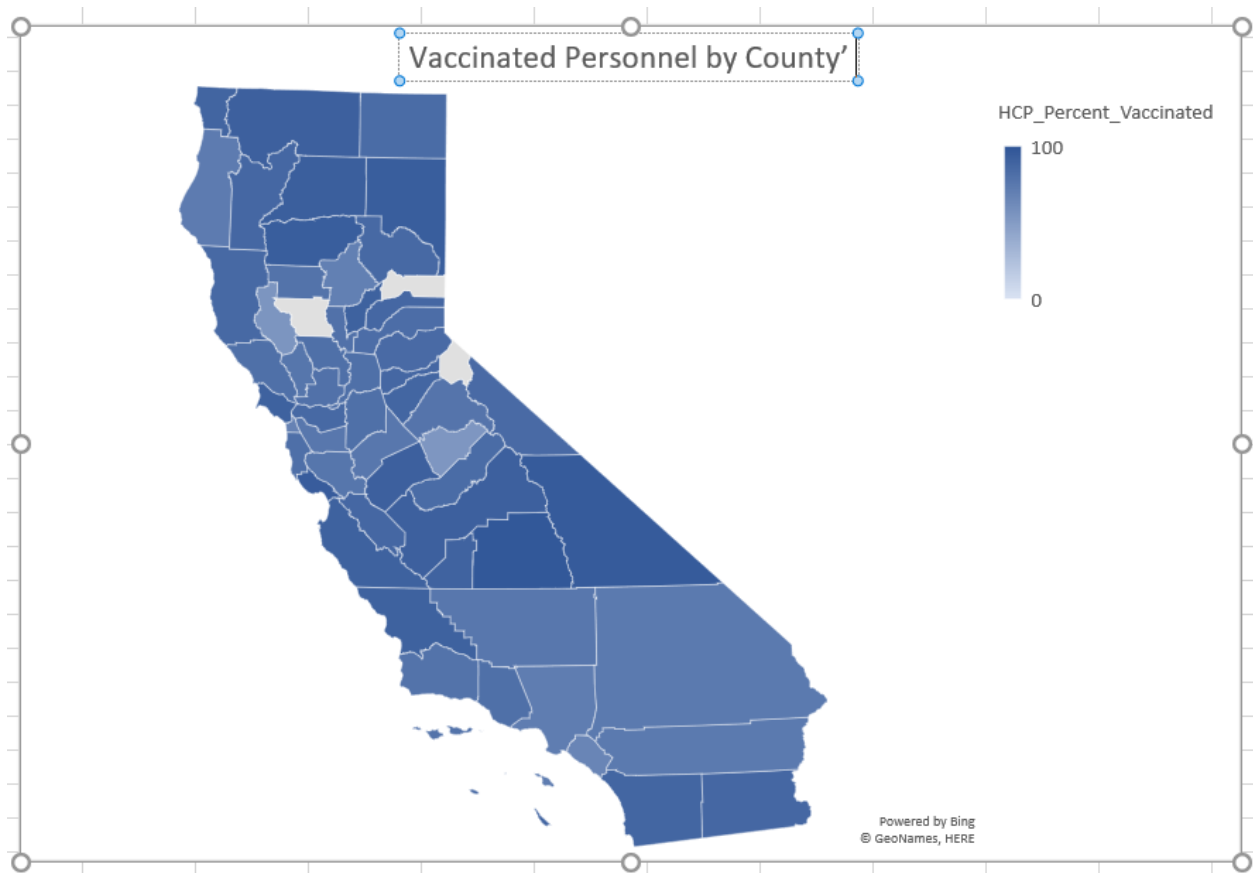-In the top left corner, Select **Insert,** then **Maps,** then **Filled Maps**

Now a map of California will appear with clinic locations, but we need to remove some extra data

-Right click the map and select **Select Data**

-In the left box, **Remove** Facility_name and Facility_ID

-Name the map "Vaccinated Personnel by County" where it says **Chart Title**

-Result should look like this:

<u>End of Tutorial</u>

References

URL of Data Source: https://data.ca.gov/group/health-human-services?q=flu&sort=score+desc%2C+metadata_modified+desc

URL of your Github: https://github.com/bulldog19999/DataProcessing_FluVaccinations