

Lab 5

Tableau Practical

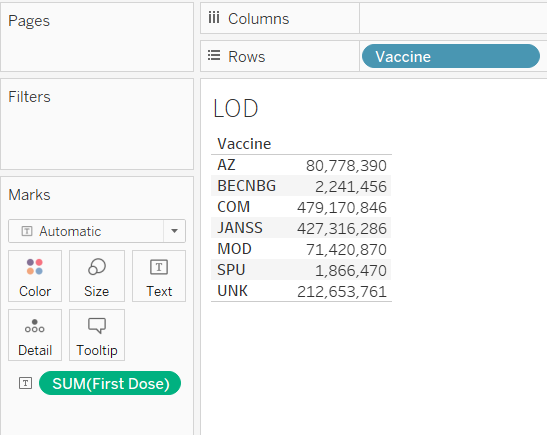


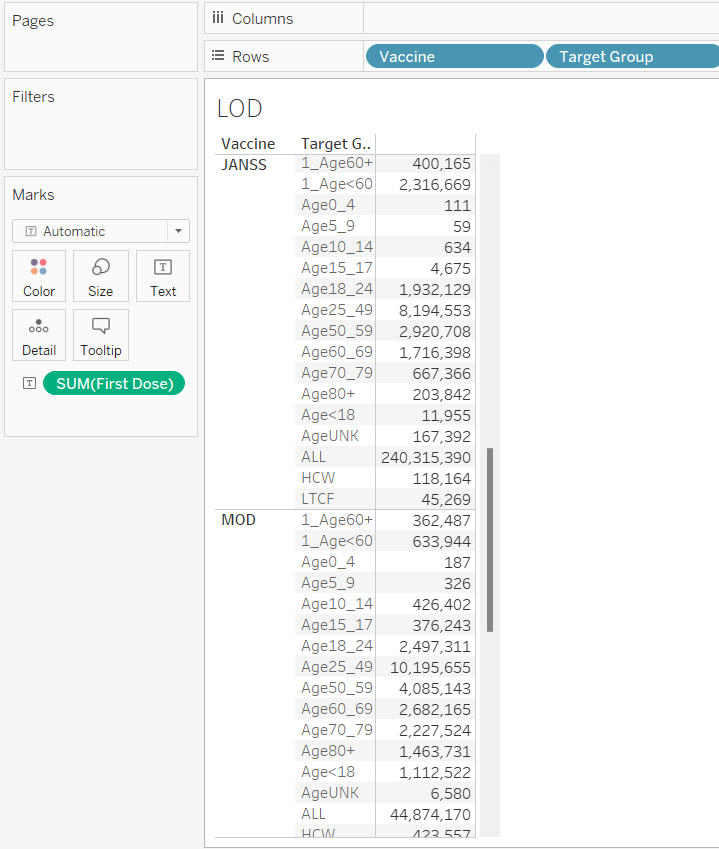
February 11, 2022

William Jackson, Julie nguyen, Matt wehnert

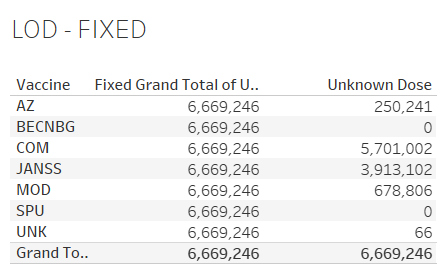
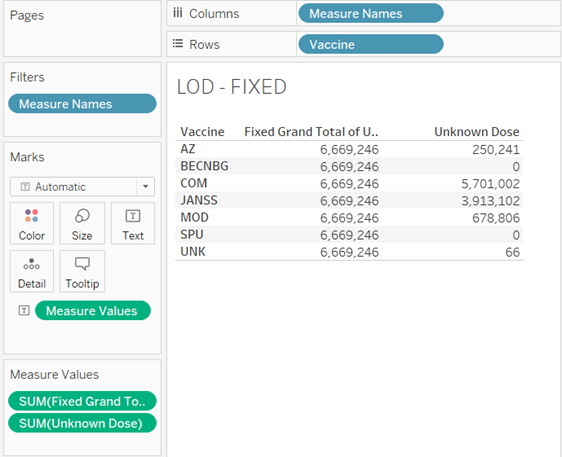
CSI 475

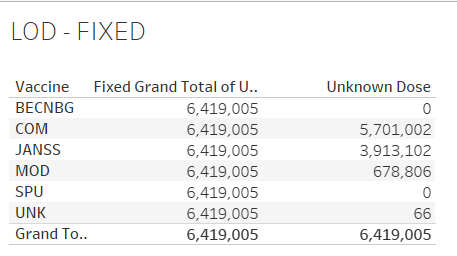
Procedure for LOD and Calculated Fields (Tableau)

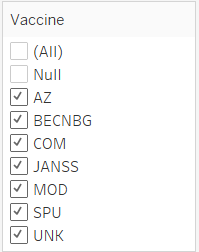
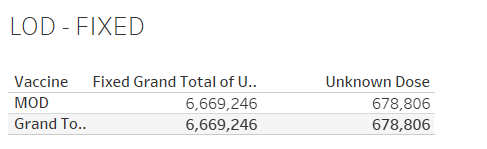
1. Boot up Tableau and continue from lab4.twbx from the previous lab.
2. Add a new worksheet, this will be sheet 3 and the Europa dataset will be used for this procedure. You can rename this as “LOD” for the header.
3. Add Vaccine to the Rows.
4. Drag First Dose into the little image button that says “Text” in the Marks box. You will then see a small chart detailing the population of those who have taken the specified vaccine as their first dose:
5. Next, drag Target Group into the rows. The numbers should become granular and more detail is added. This is considered to be the granular subtotal. The image below shows the granular values:



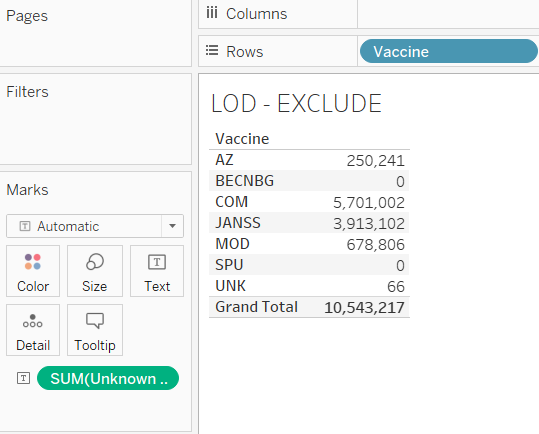
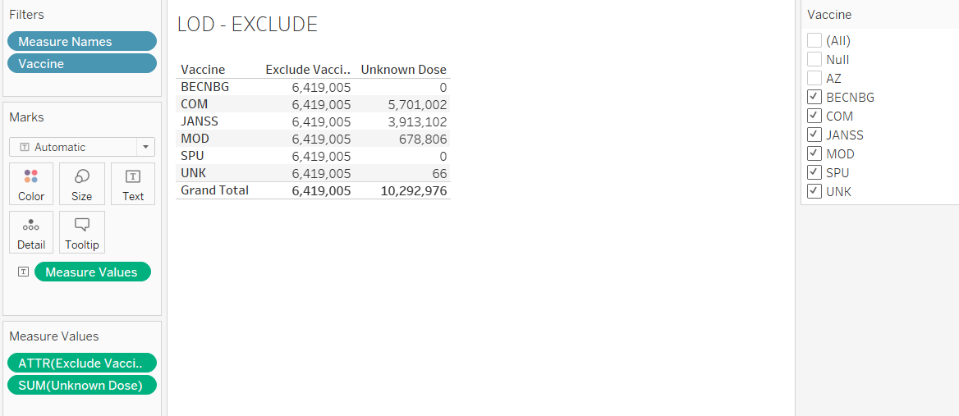
LOD Fixed

1. In this example the FIXED value will be looked at, create a new sheet. This will be sheet 4 and edit the header to “LOD – FIXED”
2. Drag Unknown Dose to Text.
3. Create a calculated field with the fixed expression {FIXED :SUM([Unknown Dose])} and save the calculation as “Fixed Grand Total of Unknown Dose”.
4. Apply the calculation and double-click the newly made calculation. You should see two identical values.
5. Next, drag Vaccine into the row. This should now reveal a column of identical numbers but next to that will have different values. See image below:
6. Under the Analytics tab, click totals. This should give a new row that gives the grand total:

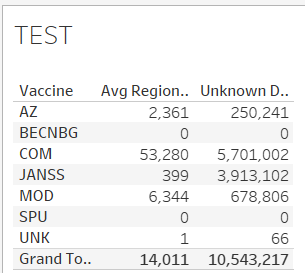


1. Go to Vaccine in the row, click the dropdown, and then click show filter. This will put the Vaccine filter in the Filters. This will then add a filter to the right of the application with a checklist of the specified vaccine:
2. Uncheck every vaccine except for Moderna. This will change the values as seen here due to dimension filters.
3. Click Add to Context under Vaccine which is located under Filters. The word Vaccine should now be grayed out. Check All to check all vaccines and then uncheck both AZ and null.

LOD EXCLUDE Procedure

1. Create a new sheet, this will be sheet 5. Rename the header to “LOD – EXCLUDE”.
2. Drag Unknown Dose to Text and Vaccine into rows. Create a grand total row from the Analytics tab. To accurately reflect the grand total right click SUM(Unknown Dose) in the Marks section and set the “Total using sum” to Sum instead of automatic. See image below for reference:
3. Create a calculated field called “Exclude Vaccine” with the following expression: {EXCLUDE [Vaccine]:SUM([Unknown Dose])}
4. Double click to apply the Exclude Vaccine calculated field.
5. Right click on the Vaccine in rows and click show filter. Uncheck the AZ and null vaccine and the numbers will update as shown here:
6. Drag Target Group into the rows. You will notice the values are peculiar but this will be fixed.
7. Right click on the “Exclude Vaccine” calculated field in the Tables section and click on Edit. Type in the following expression: {EXCLUDE [Vaccine],[Target Group]:SUM([Unknown Dose])}
8. Apply the new expression.

LOD INCLUDE Procedure

1. Create a new worksheet, this will be sheet 6 and rename the header to “LOD – INCLUDE”
2. Drag Unknown Dose to Text while Vaccine and Region are in rows.
3. Go to the analytics tab and click on totals to give a total row.
4. Change the “Total using” to average for the unknown dose in the Marks section. The figure below shows the total and grand total using the average:
5. Create a calculated field called “Avg Regional Vaccines” with the following expression: AVG({INCLUDE[Region]:SUM([Unknown Dose])})
6. Apply the calculation
7. Create a new sheet renaming the header to “Test” this will be sheet 7 and will be used to test the average calculated field made in the previous step.
8. Drag the Unknown Dose to text in the Marks section and drag the Vaccine field into rows. Then go to Analytics tab to create a grand total row.
9. Change the “Total Using Sum” of the Unknown Dose from automatic to sum to accurately reflect the grand total.
10. Double click the “Avg Regional Vaccine” calculated field that was made in step 27 – step 28. The output should look something like this:

LOD Map Procedure

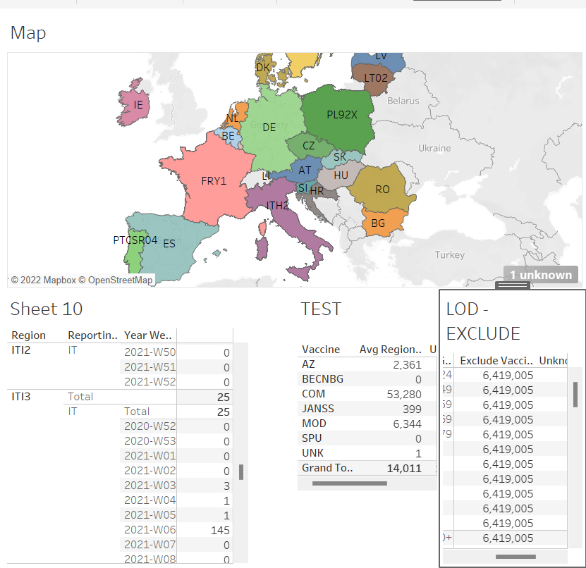
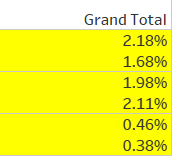
1. Create a new sheet, this will be sheet 8 and rename the header to “Map”.
2. Change the field type of region to “country/region” as this was previously set as text type. Make sure the Reporting Country field type is also set to country/region.
3. Double click region.
4. Drag and drop Reporting Country to color.
5. Put region to label.
6. In the marks section, under the dropdown menu select map.
7. Put the Unknown Dose in tooltip.
8. Create a calculated field called “Fixed Country Unknown Dose” with the following expression: {SUM([Unknown Dose])}
9. Apply the calculation and apply the calculated field to tooltip.
10. Make a new calculated field called “Regional Unknown Vaccines – EXCLUDE”.
11. Create a new sheet, this will be sheet 9 and rename the header to “Regional Unknown”. This sheet is used to double check the numbers.
12. Create a new sheet, this will be sheet 10 and rename the header to “Avg unknown doses in reporting country”.
13. Drag both Reporting Country, Year Week ISO, and Region into rows.
14. Drag the Unknown Dose to text.
15. Go to the Analytics tab, add totals.
16. Under analysis, go to totals and click “column totals to top”.
17. Change the “Total using sum” from automatic to average for Unknown Dose.
18. From the dropdown, create a calculated field “Avg Unknown Dose by Reporting Country” with the following expression: AVG({INCLUDE [Year Week ISO]: SUM([Unknown Dose])})
19. Click ok.
20. Go back to the map worksheet that was made previously and drag the “Avg Unknown Dose by Reporting Country” calculated field to tooltip. Do the same for the “Regional Unknown Vaccines – EXCLUDE”
21. Create a dashboard by clicking on the 4 square icon.
22. Drag sheet 8 (Map) and sheets 5,7,10 after. The result should be this:
23. Save the file as lab5.twb to save your work before moving on to the next part.

Table Calculations Procedure

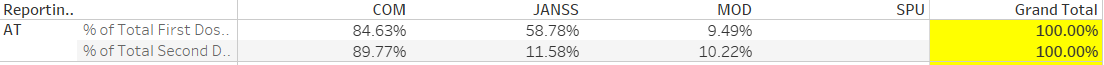
1. Create a new sheet, this will be sheet 11. Rename the header to “Table Calculations”.
2. Drag and drop Vaccine into column. This will create a text-based column.
3. Drag Reporting Country into the row.
4. Double click on both First Dose and Second Dose, this will put them as Measure Values.
5. Go to the Analytics tab and add totals.
6. Click Format and on the dropdown, click Shading.
7. Go to the Total section and select the pane color to yellow or any color you wish. Your grand total and total rows/columns should be highlighted the color you chose. Example image:
8. Right-click on SUM(First Dose) and hover the mouse cursor over quick table calculation, then click “percent of total”. Do the same thing for SUM(Second Dose). Example image:
9. Right-click on SUM(First Dose) and hover the mouse cursor over “compute using”, then click “table down”. Do the same thing for SUM(Second Dose). This puts the grand total of 100% at the bottom and the different percentage values to the right as columns. Image:

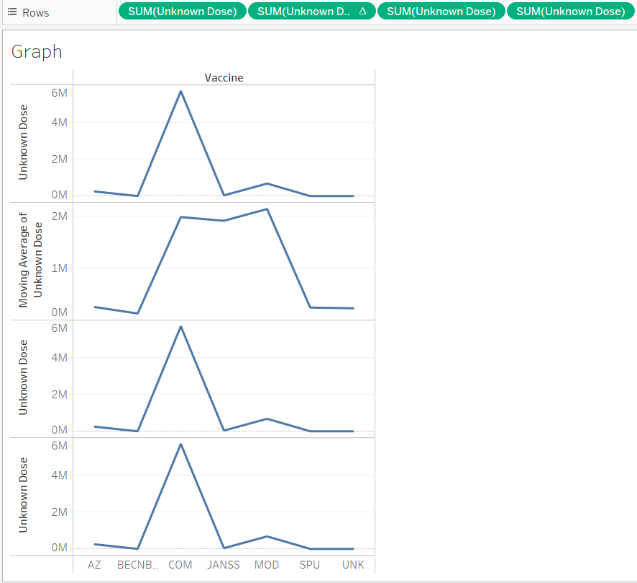
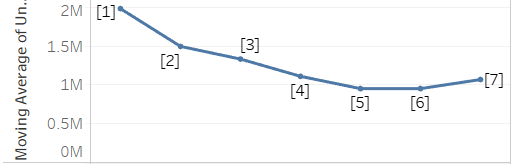
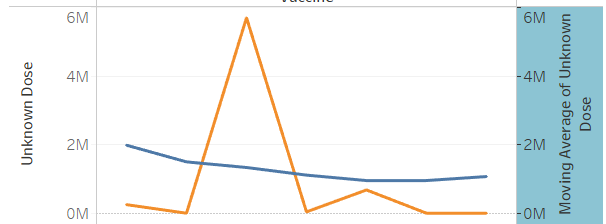
*Note: If the format does not feel right, you can revert back to table (across)*

1. Right-click on SUM(First Dose) and hover the mouse cursor over “compute using”, then click “pane down”. Do the same thing for SUM(Second Dose). This resets how the percent of total is calculated. Instead of the 100 being at the very bottom it is in each group.

***Warning: Following the YouTube tutorial and using this dataset will result in undesirable results. Step 66 resulted in each cell to be 100% which is incorrect. It is heavily advised to keep the format as table across to maintain the integrity of the data. Proceed to revert back to table across before moving on to step 67.***

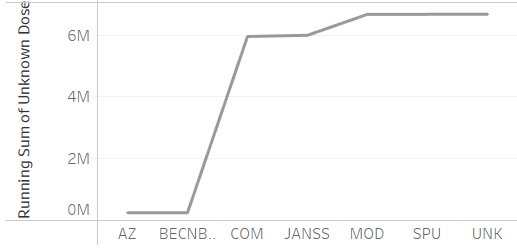
1. Add a filter on Vaccine. On the dropdown, select show filter.
2. Uncheck all but leave Moderna, COM, JANSS, and SPU. Result:



1. Double-click on First and Second Dose to add the original values to the table.
2. Create a new sheet, this will be sheet 12. Rename the header to “graph”.
3. Drag Vaccine into column and Unknown Dose into rows.
4. Copy the Unknown Dose using CTRL 3 times for a total of 4.
5. For the second Unknown Dose, add a Moving Average through Quick Table Calculation. This will create a new graph:
6. For the second Unknown Dose click Edit Table Calculation.
7. Click on the second drop down box under Calculation Type. Change the previous value to 5 and the next value to 2. This smooths out the volatility of the numbers to enable insight into patterns. Image:
8. For the second Unknown Dose, add a Dual Axis through the dropdown menu. Then right-click the right of the graph that was combined and click synchronize. This will adjust the curve:
9. To remove the header, right click the teal-colored area and uncheck show header.
10. For the third pill (the third Unknown Dose) under the dropdown, click percent difference. Then change the format to bar from line.
11. CTRL + Drag the third pill to color.

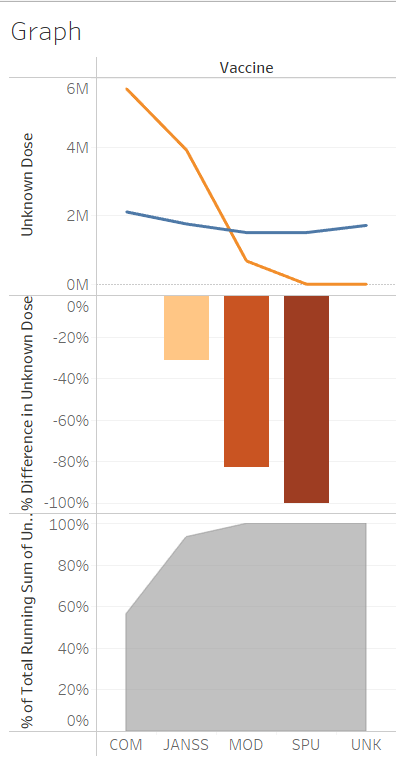
Note: You can hide the null indicator by right clicking and clicking hide indicator

1. On the fourth pill, right-click and add a running total. The graph gives the following:



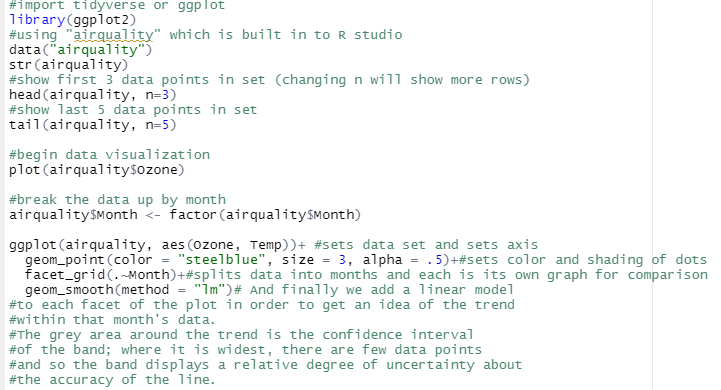
1. Change the mark to area.
2. On the fourth pill, click on the dropdown to Edit Table Calculation and check “add secondary calculation”. Then on Secondary Calculation Type, click percent of total.
3. Add Target Group as a filter.

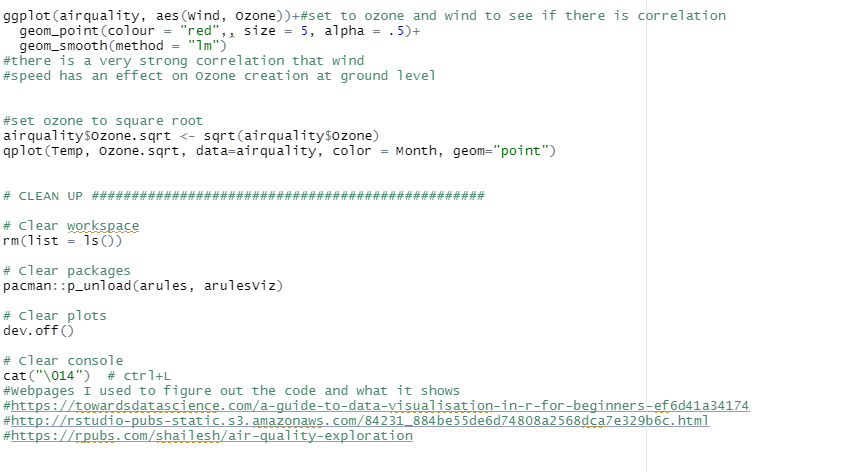
Note: By unchecking the filter of the target groups this drastically effects the data depicted. Histograms will be inverted and line graphs will be a different shape.

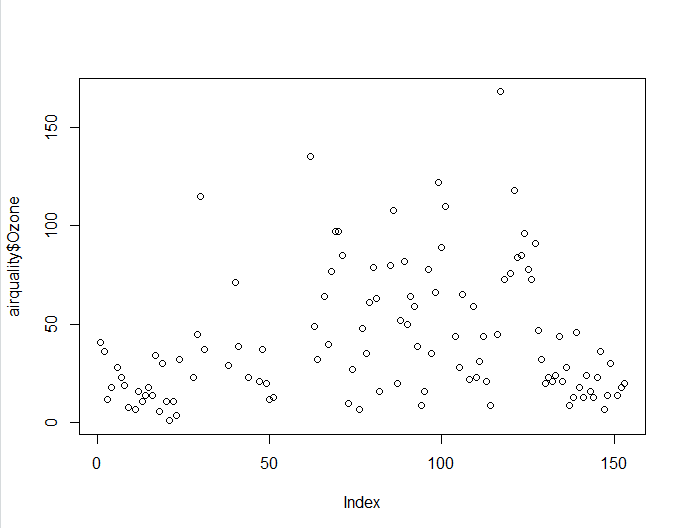
1. Remove the filters.
2. On the fourth pill (Unknown Dose), click show filter. The graph shape is affected by the percentages, here 49% is used as input:

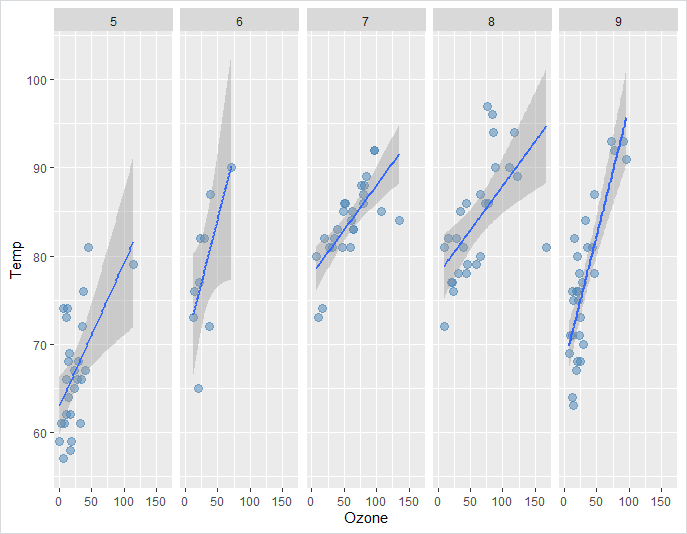
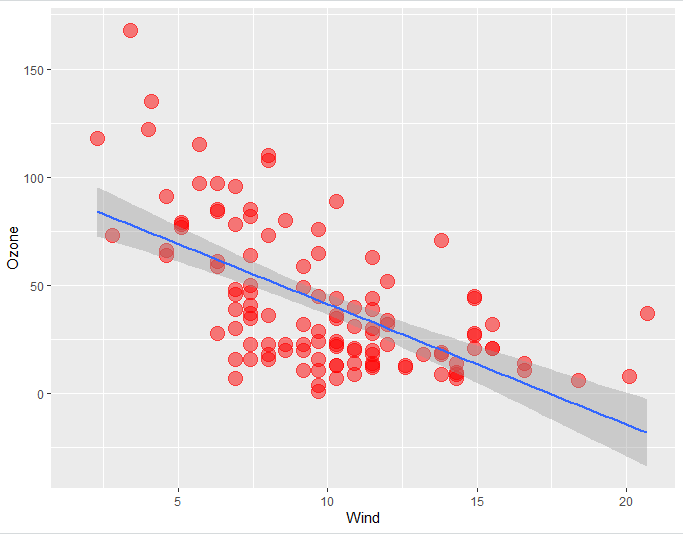
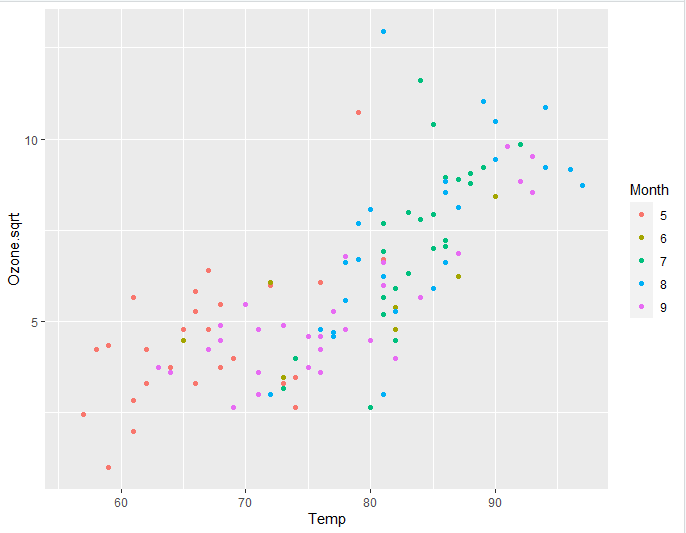
1. Save your work and keep the file in a safe location.
2. Close out of Tableau.

Procedure (R Studio)

1. Open up R Studio and create a new script file.
2. Import the library necessary to visualize the dataset provided by R Studio.
3. Type the following code provided in the screenshot:



1. Run each line of code until the first graph is given on line 12:

1. Run lines 17-20 for the second graph:
2. Following lines 28-30 depicts the third graph:
3. Following lines 36-37 provides the final fourth graph:
4. Save the file in a safe place and under a rememberable name.

Conclusion

Overall, my group and I applied concepts of LOD, calculated fields, and revisited R Studio. LOD provides the user the ability to compute values at the data source level and visualization level. We were able to create a map showcasing different values from the calculated fields applied to them.

# Challenges/Problems

1. Learning the concept took a while because it seemed like a large topic.
2. Unsure if calculations applied were correct.
3. Some queries took longer than expected.
4. Starting from scratch and rewatching the YouTube tutorials took up a lot of time.
5. Trying to open up another teammate’s Tableau file was impossible because it needed connection to their database. The lack of login credentials and security was concerning.
6. Keep table across to maintain integrity of the data.
7. R was difficult but a few articles were able to help us.

# Insights/Resolutions

1. We learned how to make a map based on the data and field provided by the dataset.
2. Watching the YouTube tutorial helped us understand how Tableau is processing the data based on their order of operations and the steps taken behind in the background.
3. While there were technical difficulties, we decided to work off of on
4. We learned about FIXED, INCLUDE, and EXCLUDE LOD expressions and how they can be useful in isolating which parts of the data we need or don’t need.
5. Calculated fields are useful in creating calculations to give an output.
6. Tables are useful in seeing which variables are important.
7. The YouTube tutorial was useful somewhat but the scenario used is different from ours. The tutorial is using sales while our dataset is about COVID-19.
8. Pane down causes every cell to be 100% which is incorrect, only solution is to maintain table across.
9. R has a built-in dataset that we could use.
10. In R, one of the graphs depict that there is a strong correlation that wind speed effects the ozone creation at ground level.