|  |  |
| --- | --- |
| **Assignment D.4 - Milligan Chapter 10** | |
| **Student Name:** | Matt Schnorr |

Week 7 - Milligan Chapter 10

Welcome to your hands-on activity! This assignment invites you to work with Tableau using a mix of generated and real-world data. You will:

* Download the attached spreadsheets and upload those data sets into Tableau. If you need help loading your data sources, refer to Assignment A.2 for help.
* Go through this document and use Tableau to answer all the questions listed below. Where applicable, paste screenshots into the template below.
* When you are ready, complete the online quiz which verifies your homework. You are welcome to refer to this document when you complete the quiz.
* You are welcome to use the quiz results to update your homework file.
* When you have completed the online quiz, submit the Word document.
* Remember, you can always ask your instructor for help if needed.
* If you need to adjust the size of your visualizations to match the options in the questions, use the “Format”->”Cell size” options. For example, “Ctrl+Shift+B” on a Windows computer will make the visualization bigger and “Ctrl+Up” will make it taller.

The spreadsheet attached to this assignment contains seven datasets and six icon files:

1. D.4 NW National Park Visitation 2001-2020
2. D.4 Monthly Personal Budget (two worksheets-July and August)
3. D.4 Video game sales
4. D.4 2019 NYC Restaurant Inspections
5. D.4 Baltimore City Crime Data June 2021
6. D.4 University course completion rates
7. D.4 University pass rates
8. Icon files:
   1. Biology
   2. Chemistry
   3. Computer Programming
   4. Data Analytics
   5. Math
   6. Physics

**Question 1: Bump chart**

You have been asked to help the Northwest National Park system to find trends. Specifically, what do five-year visitation trends look like for the parks? Load the D.4 NW National Park Visitation 2001-2020 into your data sources. Drag “Year” into the columns and create an ad hoc calculation in the row: SUM(Visitor Count). Then drag “Park Type” and “Year” into the filters. Select ONLY the “National Park” park type and the years 2015-2019. We are purposely excluding 2020 due to the COVID19 pandemic shutdown of many parks. Drag “Park Name” onto Color in the Marks tab and “Park Name” onto Label. Make sure the marks type is set to “Line”. Edit the label so that it shows the “Line Ends”.

Graphical user interface, text, application

Description automatically generated

You should now have a bump chart that looks like this:

Chart, line chart

Description automatically generated

From this chart, we can see that Yosemite and Olympic are firmly the most visited with some ups and downs. Next, Joshua Tree is on an increase. The line that really sticks out is the Hawaii Volcanoes NP, it really took a big dip in 2018 and had a slow increase in 2019.

While this is helpful, the leadership has requested to see a chart with 20 years (2010-2019). They also want to see only National Parks with at least 500K visitors a year. Add both filters to your bump chart.

**Question 1A: For the Hawaii Volcanoes NP, was there an increase or decrease between 2010 and 2017?**

|  |  |
| --- | --- |
| **Question 1A Answer:** | Increase |

**A screenshot of a graph

Description automatically generated**

**Question 1B: What was the highest visitor count for Hawaii Volcanoes NP between 2010 and 2019?**

|  |  |
| --- | --- |
| **Question 1B Answer:** | 1995104 |

**A screenshot of a graph

Description automatically generated**

**Question 2: Waterfall**

Your friend is always complaining to you that they can’t save any money and their bank account is always going down. You ask your friend to write down all the expenses and incomes in a month and you record them in the D.4 Monthly Personal Budget- July. For each expense, you have tracked the date, type, whether it is income, fixed, or discretionary, and the amount (negative for expenses and positive for income).

Graphical user interface, text, application

Description automatically generatedCreate a waterfall chart of the July data by dragging the “Income/Fixed/Discretionary” and “Type” dimensions into the columns. Then drag “Amount” into the rows. Since we want to see the total as the month goes on, we will create a running total calculation using *table across*. To do this, first, right-click “Amount” and select “Add Table Calculation…”. This should bring up the following window, which you should change Calculation Type and Compute Using to match:

Graphical user interface, text, application

Description automatically generated

Drag the “Amount” measure onto Color, Size, and Label in the Marks tab.

Change the chart type to “Gantt Bar”.

Graphical user interface, text, application, chat or text message

Description automatically generated

Click on “SUM(Amount)” on Size and edit in the shelf. Add a negative sign in front of the SUM([Amount]).

Graphical user interface, text, application

Description automatically generated

Change the “SUM(Amount)” on Label to be a Running Total.

Graphical user interface

Description automatically generated with medium confidence

Your waterfall chart should be taking shape now. You should see something similar to the following:

Chart, waterfall chart

Description automatically generated

Now, we can see quickly that your friend spent $690 in the month of July just on discretionary spending. Then by the time they paid their rent, they had spent a total of $4,313 for the month. They make $3,600 from their regular job and another $700 from side gigs. This means that in July, your friend overspent their monthly income by $13 dollars. If they spend like that every month, that means they are overspending $156 a year.

Your friend would really like to start saving money so you help them go through expenses and try to spend less money. Now, they come back with their August expenses (another worksheet in the D.4 Monthly Personal Budget) and they want to see the same chart.

**Question 2: How much money did your friend save this month? Provide your answer rounded to the nearest dollar (no cents).**

|  |  |
| --- | --- |
| **Question 2 Answer:** | $528 |

**A screenshot of a graph

Description automatically generatedQuestion 3: Step lines**

For this exercise, you will be using the “D.4 Video game sales” data set to answer questions about how certain genres and performing by global sales in the last five years.

Drag “Genre” over to columns and “Global Sales” into Rows. Add a quick calculation of Running Total to the “Global Sales”.

Drag “Global Sales” onto Color in the Marks tab.

Filter by Years 2011 to 2016.

Change the Marks type to “Line” and then click on “Path” and choose “Step”.

Graphical user interface, application

Description automatically generated

Click on *Ctrl+Shift+B* (or *CMD+Shift+B* if you’re on a Mac) a few times to make the graph wider (it might help to read the bottom axis better).

Chart

Description automatically generated

We can see that **shooter games** had a significant step within the global sales between 2011-2016.

However, we can see that action starts around $140. Let’s modify this a bit.

Right-click on “Genre” in the columns area, click on “Sort” function. We want to sort by Field, Ascending, Field Name-Global Sales, and Aggregation type-Sum.

Graphical user interface, application

Description automatically generated with medium confidence

Chart

Description automatically generated

Now, we can see where the calculation starts in the bottom left corner with *Fighting* and slowly steps until we get to *Action* and then *Shooter* games where we see two big jumps. This is a better representation of the step-line chart.

Now, change the years to 2006-2016.

**Question 3: What is the bottom step, and what is the top step?**

Option A: Fighting and Shooter

Option B: Action and Shooter

Option C: Adventure and Action

Option D: Strategy and Shooter

|  |  |
| --- | --- |
| **Question 3 Answer:** | Option D |

**A screenshot of a graph

Description automatically generated**

**Question 4: Sparklines**

You are being asked to find trends about the New York City Health Department Restaurant Inspections. There are over 27,0000 restaurants in the five boroughs of New York City. You need to load the data source: D.4 2019 NYC Restaurant inspections. This is a rather large data source and may take a little longer to load the data.

When restaurants have violations, they receive scores for each violation. So, restaurants want to have low scores, and high scores are not desirable. Grades A-C are assigned to restaurants based on their scores. For restaurants with scores of 0 to 13 sanitary violations, they receive an “A”. If they get a score of 14-27, they receive a “B”. A “C” grade is 28 or more. If you would like to read more <https://a816-health.nyc.gov/ABCEatsRestaurants/#!/faq>.

Drag “Inspection Date” into columns and “Boro” into rows. Now pull “Score” into row and change the aggregation to “Average” instead of sum. Change the date from year to week number (right-click on the inspection date field in columns).

Graphical user interface, application

Description automatically generated

Right-click in the line graph area and click on format. Then click on the lines icon and turn off gridlines.

Graphical user interface, application

Description automatically generated

Widen the “Boro” column to make it easier to read the names of each borough.

Add the field “Inspection Type” as a filter. Use the wildcard filter to search for “compliance”- this will get any inspections that were related to compliance concerns.

Graphical user interface, application

Description automatically generated

Your sparkline chart should look like the following:

Chart, line chart

Description automatically generated

We can see that this is very sporadic data that doesn’t really show a clear pattern. Now, remove all the filters so that you are looking at all of the inspection types.

**Question 4a: Based on these sparklines, do you think there is a clear pattern of food inspection scores throughout the year for each borough (boro)?**

**Yes**

**No**

|  |  |
| --- | --- |
| **Question 4A Answer:** | No – Scores vary up and down with no pattern. |

**Question 4b: Using the sparklines, which borough (boro) had the biggest swings in averages (increases and decreases) over the year?**

**A: Queens**

**B: Staten Island**

**C: Manhattan**

**D: Brooklyn**

**E: Bronx**

|  |  |
| --- | --- |
| **Question 4B Answer:** | B – Staten Island |

A screenshot of a graph

Description automatically generated

**Question 5: Dumbbell**

In this exercise, you will be looking for patterns in crimes in the City of Baltimore that occur inside versus outside. To do this, we will use the dumbbell method. First, load the data source: D.4 Baltimore City Crime Data June 2021: All Districts into Tableau and start a new worksheet.

Drag “All districts (Count)” measure into columns. Repeat this again (you will have two next to each other). Now drag “District” into the rows. Move “Inside Outside” into filters and choose: Inside and Outside.

Right-click on the “CNT(All Districts)” column on the right and select “Dual Axis”.

Graphical user interface, text, application

Description automatically generated

In the Marks area, change the first CNT(All Districts) to be a line. Drag “Inside Outside” onto the path marks area.

Make sure the second CNT(All Districts) mark is a circle. Now, drag the “Inside Outside” onto the color inside that marks menu.

You should have dumbbells that look like the following.

A screenshot of a graph

Description automatically generated with low confidence

**Question 5: Which district had more inside incidents than outside incidents in June 2021?**

Option A: Central

Option B: Eastern

Option C: Western

Option D: Southern

|  |  |
| --- | --- |
| **Question 5 Answer:** | Option A - Central |

A screenshot of a graph

Description automatically generated

**Question 6: Unit/symbol**

We will be analyzing enrollment versus room capacity for a university. Load the dataset *D.4 University courses and completion rates*. Create a scatterplot by dragging “Student enrollment” and “Student Completed” onto columns and rows.

Now, we are going to make groups for each subject. Right-click on the “Course Number” dimension and

Graphical user interface, application, Word

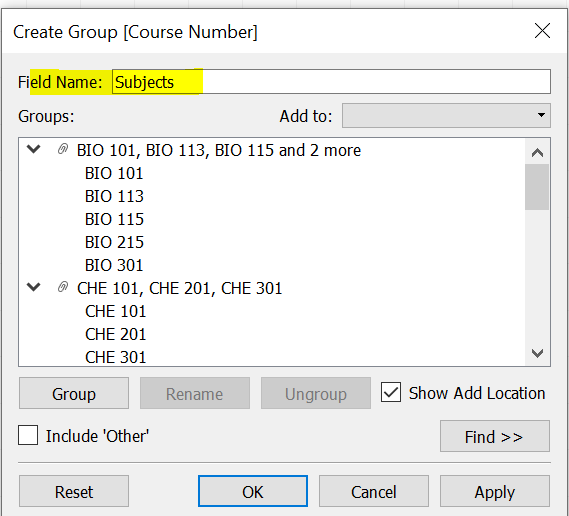
Description automatically generated

Highlight all the BIO classes and click Group.

Graphical user interface, application

Description automatically generated

Repeat this for CHE, CSI, DTA, MAT, and PHY. After you finished this, change the field name to “Subjects”.



Now, drag that new field, “Subjects” onto the “Shapes” mark.

We get a scatterplot that looks like this:

Chart, scatter chart

Description automatically generated

The problem is that our user has to look at the legend on the right to know what the subjects are. Let’s eliminate this by adding icons to make it easier for the user to read our graph.

Download the six icon image files that were provided to you in the assignment. You need to save them into your Tableau Repository file.

To find out where this is- go under File ->Repository Location

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Make a note of the location of where your Tableau Repository is located. In the image above, it is located in “Documents”. Go to the location of your Tableau Repository and create a new folder called “University subjects”. Then copy and paste the six images provided to you into this folder.

Graphical user interface

Description automatically generated

Return to the scatterplot you made. Click on the Shape mark.

Graphical user interface, text, application, chat or text message

Description automatically generated

Change the Shape Palette from Default to University subjects. If you don’t see them, click on “Reload Shapes”.

Graphical user interface, text

Description automatically generated

Now, you need to assign the data item on the left with the appropriate icon on the right. They are both in alphabetical order.

Graphical user interface, text, application

Description automatically generated

Click apply when you are finished.

Your scatterplot probably has very small icons now:

A picture containing text, white

Description automatically generated

Click the Size Marks and move the slider to the next to last position.

Graphical user interface

Description automatically generated

Hide the legend card. Now, your scatterplot should be easy to interpret.

**Question 6: Which subject has the highest number of students enrolled and completed in courses?**

**A: Math**

**B: Computer Science**

**C: Data Analytics**

**D: Chemistry**

|  |  |
| --- | --- |
| **Question 6 Answer:** | C: Data Analytics |

**A screenshot of a graph

Description automatically generated**

**Question 7: Marimekko**

The University we have been working with has asked for an analysis of the student pass/fail rates by major/nonmajor within six departments. Load the dataset: D.4 University pass rates.

Drag “Department”, “Major/NonMajor”, and “Pass/Fail” into rows. Drag “Students” onto the Label. You should see a simple table.

A picture containing graphical user interface

Description automatically generated

If we want to create a visualization across the three different dimensions, we can create a Marimekko chart.

We need to create two calculated fields that will help us to get the right level of detail calculation for this analysis:

1. “Students per column” Syntax: {EXCLUDE [Pass/Fail]: SUM([Students])}
2. “# of students” Syntax:

IF FIRST()==0 THEN

MIN([Students per column])

ELSEIF MIN([Department]) != LOOKUP(MIN([Department]),-1) THEN

PREVIOUS\_VALUE(0) + MIN([Students per column])

ELSEIF MIN([Major/NonMajor]) != LOOKUP(MIN([Major/NonMajor]),-1) THEN

PREVIOUS\_VALUE(0) + MIN([Students per column])

ELSE

PREVIOUS\_VALUE(0)

END

Move the three dimensions “Department”, “Major/NonMajor, “Pass/Fail” out of the rows and into the detail of the marks area.

Drag the new calculated field- "# of students” into columns. Right-click to create a table calculation using across the dimensions “Department”, “Major/NonMajor, “Pass/Fail”.

A picture containing application

Description automatically generated

Drag the measure “Students” up into rows. Right-click and add table calculation- **Percent of total** compute using **Pass/Fail**.

Graphical user interface, text, application

Description automatically generated

Calendar

Description automatically generated

Change the graph chart from automatic to bar.

Graphical user interface, text, application

Description automatically generated

Chart, bar chart

Description automatically generated

Drag “Students per column” into the size marks.

Click on the Size box- change it to **Fixed** and **Right** alignment.

Graphical user interface, text, application, chat or text message

Description automatically generated

Drag “Pass/Fail” onto the Color mark. You should now have a bar chart like this:

Chart, bar chart

Description automatically generated

Drag “Department” onto the labels.

Using the visualization you created, find the department that had the highest number of students majoring in their department that passed.

**Question 7: For that department, what % did the passing majors represent? (Hint: hover over the bars to look at the tooltips).**

|  |  |
| --- | --- |
| **Question 7 Answer:** | Biology – 93.8% |

**A screenshot of a graph

Description automatically generated**

**Question 8: Animated**

Note: The animated feature only works in versions of Tableau after 2020.1- if you don’t have the feature, you will need to upgrade.

For this animation, we will create a scatterplot using the D.4 Video game sales dataset. Create a new worksheet. Drag the measures “Global Sales” into columns and “Rank” into rows. Leave the aggregation as a SUM.

Drag “Publisher” into Color. Change the Marks type to Circle. Click the size button to increase the size.

Graphical user interface

Description automatically generated

Chart, scatter chart, bubble chart

Description automatically generated

Drag “Publisher” and “Platform” into the filters. Right-click and select “Show filter” for each. Click ok. Make sure you have a blank scatterplot by ensuring no filters are selected.

Graphical user interface, application

Description automatically generated

Go up to the Format Menu and select Animations.

Graphical user interface, application

Description automatically generated

In your filters, click on “DS” for the Platform and “Nintendo” for Publisher. Make a note of where the dot is located on the graph. What color is the dot? *Pink*

Now, click on “PS” for the Platform and “Sony Computer Entertainment” for Publisher.

**Question 8: What happened to the DS/Nintendo dot? You can unclick the filters and apply again to watch again.**

Option A: Nothing changed

Option B: The DS/Nintendo dot dropped in rank

Option C: The DS/Nintendo dot increased in rank

Option D: The PS/Sony Computer Entertainment dot has higher global sales

|  |  |
| --- | --- |
| **Question 8 Answer:** | B: Dropped in Rank |