DATA VISUALIZATION

ASSIGNMENT - 7

- Rohith Reddy Vangala (016762109)

Question :

Use the data set of you choice – or create a data set. The data set should include at least three dimensions.

- Using the tool of your choice, produce an excellent table showing the data. The table should employ at least one Gestalt principle to clearly present the data relationships.
- Include a separate comment describing your Gestalt principle use
- Next, using the tool of your choice, create a chart that clearly shows at least three dimensions of data.
- You will be graded on full use the the design principles we have discussed.

Solution :

The Dataset that I am taking is called as customers.csv and it is from kaggle (src = https://www.kaggle.com/datasets/datascientistanna/customers-dataset), the tool that I am using is Tableau.

Customers

CustomerID	Gender	Age	Annual Income ($)	Spending Score (1-100)	Profession	Work Experience	Family Size
1	Male	19	15000	39	Healthcare	1	4
2	Male	21	35000	81	Engineer	3	3
3	Female	20	86000	6	Engineer	1	1
4	Female	23	59000	77	Lawyer	0	2
5	Female	31	38000	40	Entertainment	2	6
6	Female	22	58000	76	Artist	0	2
7	Female	35	31000	6	Healthcare	1	3
8	Female	23	84000	94	Healthcare	1	3
9	Male	64	97000	3	Engineer	0	3
10	Female	30	98000	72	Artist	1	4
11	Male	67	7000	14	Engineer	1	3
12	Female	35	93000	99	Healthcare	4	4
13	Female	58	80000	15	Executive	0	5
14	Female	24	91000	77	Lawyer	1	1
15	Male	37	19000	13	Doctor	0	1
16	Male	22	51000	79	Healthcare	1	2
17	Female	35	29000	35	Homemaker	9	5

Here with the help of tableau I have generated a table, by grouping the data by profession, gender and taking the averages of income, age, family size and spending score(0-100) to make it more readable.

Vangala-Assignment-7

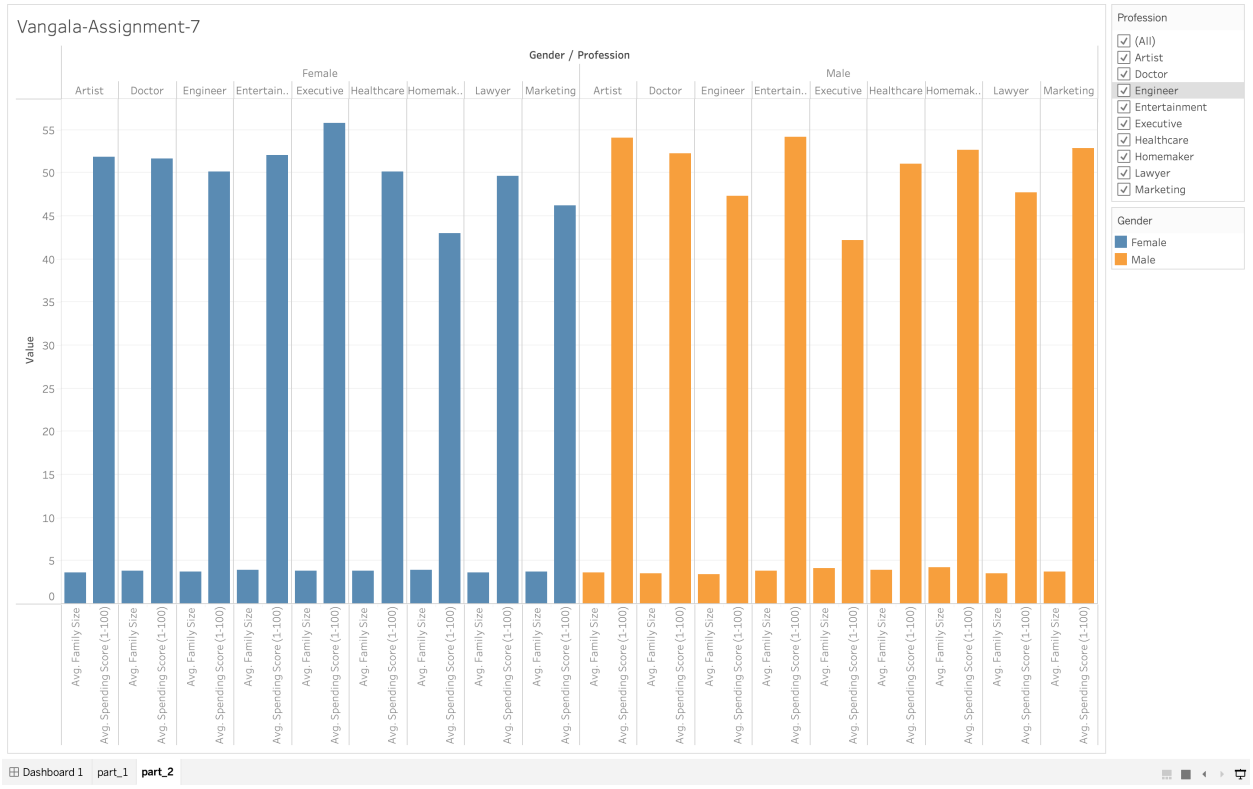
Gender	Profession	Avg. Age	Avg. Annual Income ($)	Avg. Family Size	Avg. Spending Score (1-100)
Female	Artist	49	108,102	4	52
	Engineer	55	108,498	4	50
	Entertainment	50	108,907	4	52
	Lawyer	48	109,004	4	50
	Doctor	47	109,463	4	52
	Homemaker	46	110,100	4	43
	Marketing	46	111,247	4	46
	Healthcare	47	114,215	4	50
	Executive	49	118,801	4	56
Male	Marketing	45	102,607	4	53
	Homemaker	45	106,268	4	53
	Executive	43	107,139	4	42
	Artist	49	109,881	4	54
	Healthcare	49	110,326	4	51
	Entertainment	53	112,945	4	54
	Lawyer	48	114,055	4	48
	Doctor	46	114,181	4	52
	Engineer	55	114,771	3	47

This table follows some of the important interactive principles such as column sorting, fixed headers, filter by profession (I have attached the twbx file to this assignment please check), we can sort columns in ascending or descending based on our requirements, this also follows gestalt principles likes proximity, figure ground.

1. Proximity : The Law of Proximity, also known as the Gestalt principle of proximity, states that objects that are closer to each other tend to be perceived as a group or a single unit. The proximity principle is evident in the way that the data is organized. All the information for each individual column is grouped together, making it easy to see the relationship between the different variables for each person. Additionally, the columns are spaced evenly apart, which helps to visually separate them from each other and reinforce the idea that each row represents a distinct group of data.
2. figure ground : The "law of figure and ground" is a design principle in data visualization that involves creating a clear distinction between the main data elements (the "figure") and the background or surrounding elements (the "ground"). This contrast helps to focus the viewer's attention on the main data being presented, making it easier to interpret and understand. In the above table, there is a high-contrasting color for every even row to increase readability and added a white background. This use of high contrast for even rows and white background provides a clear distinction between the foreground (even rows) and the background (odd rows), making it easier for

the reader to distinguish between them. By doing so, I have created a clear visual separation between the different rows, making it easier for the reader to focus on the individual rows and the information they contain.

For creating a chart that clearly shows at least three dimensions of data. I have used side by side Barchart and encoded the color based on the gender. The dimensions that I have used are Gender, Profession, avg family size, avg spending score.

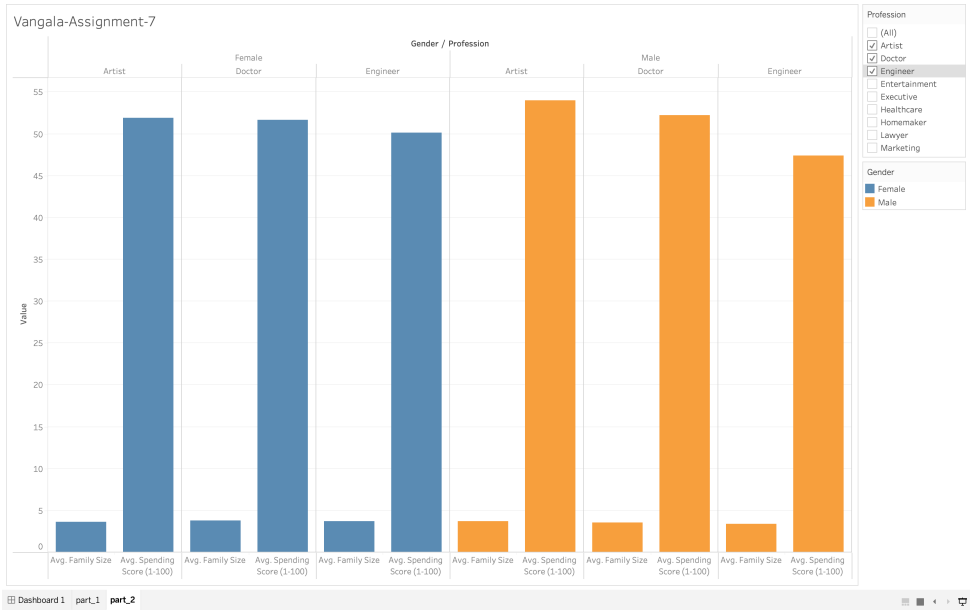


The above chart follows all the required Design principles

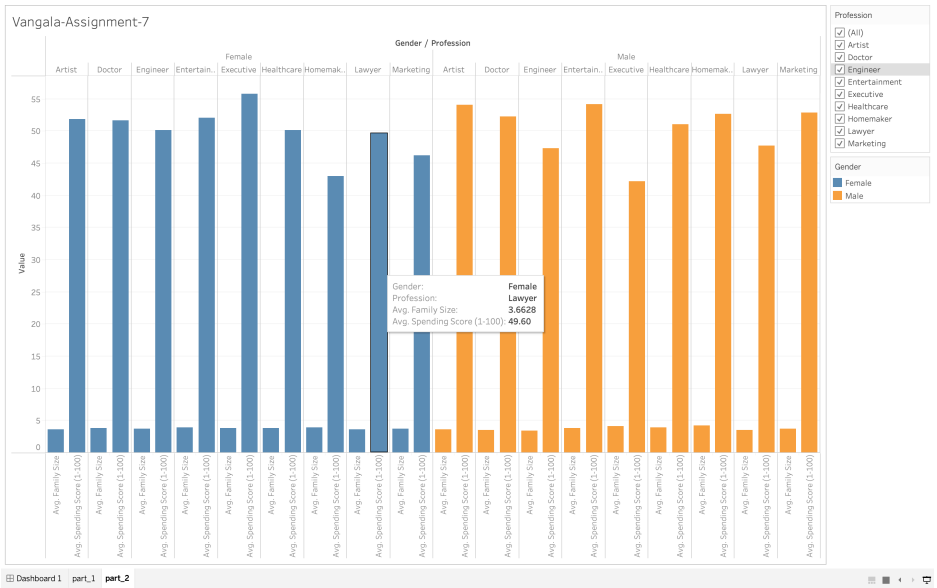
- The side-by-side bar chart effectively simplifies the presentation of complex data, making it easy to understand at a glance.
- The use of color and size consistently highlight the most important data points (gender), making it easy to draw attention to key insights and trends.
- The chart provides appropriate contextualization by comparing the data to relevant benchmarks, allowing the viewer to understand the significance of the data.
- The consistent use of font styles and labeling ensures coherence and readability throughout the chart.

- The chart's accessibility features, including appropriate color schemes and text sizes, make it accessible to a wide audience.
- The accuracy of the data is ensured through clear communication of any assumptions or limitations, making the chart a reliable source of information.

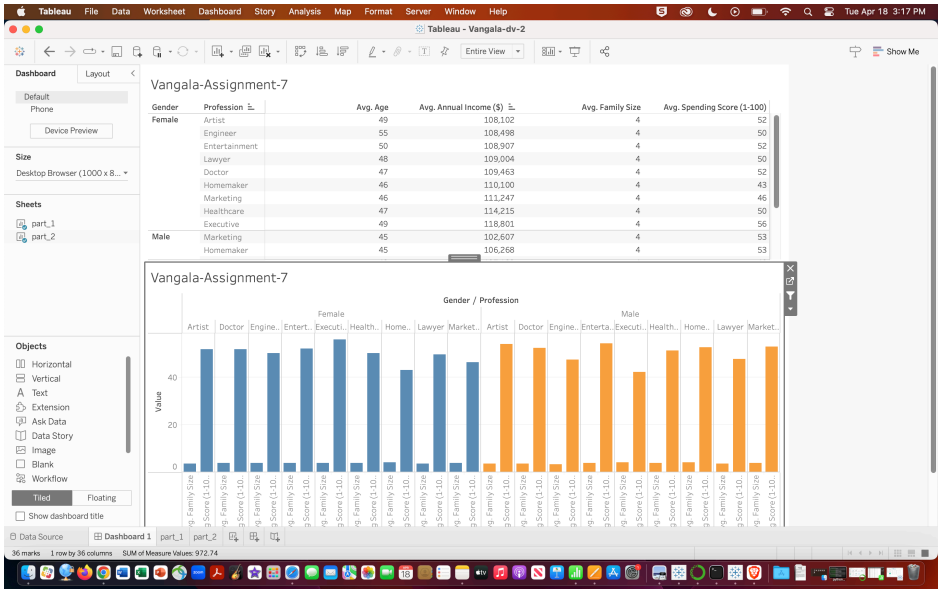
The visualization also follows some of the important interactive principles To make it more interactive, I have added filter by profession to the above Barchart, we can see the output based on 3 selected professions below.

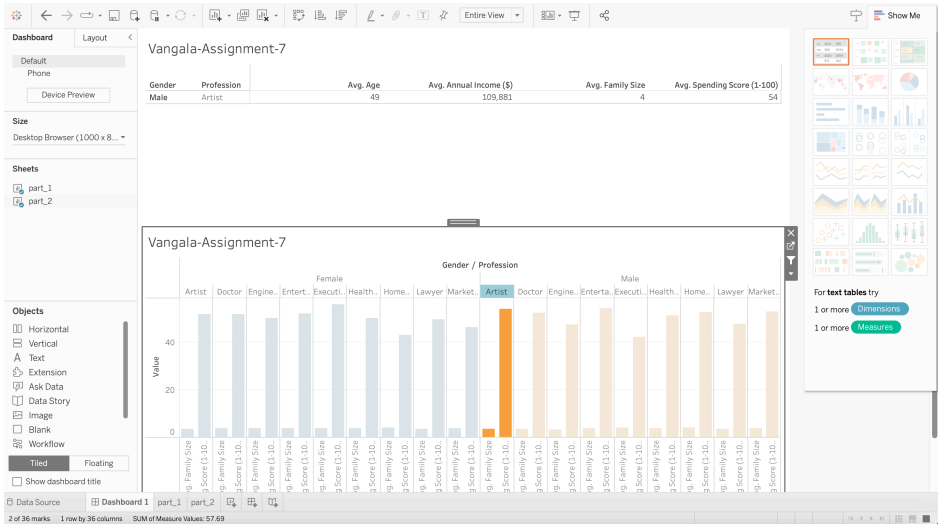


The visualization also contains hover actions that displays the data that bar wants to show.



I have also created a interactive dashboard that displays specific data in the table when selected in the side by side Barchart, to make the visualization interactive.





THE END