Visual Data Analysis

Assignment no - 01

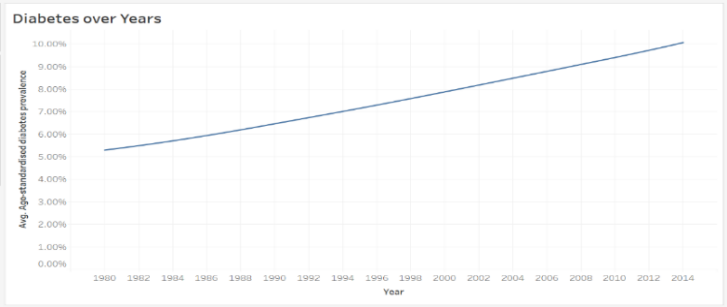
Group Members:

1. **Muhammad Khurram Shahzad – M00972373**
2. **Junaid Dawood Momin – M00904599**
3. **Mahmud Hossain – M00976273**
4. **Sabafatima Alibakar Sayyed – M00934401**
5. **Dipalee Meher – M00970691**

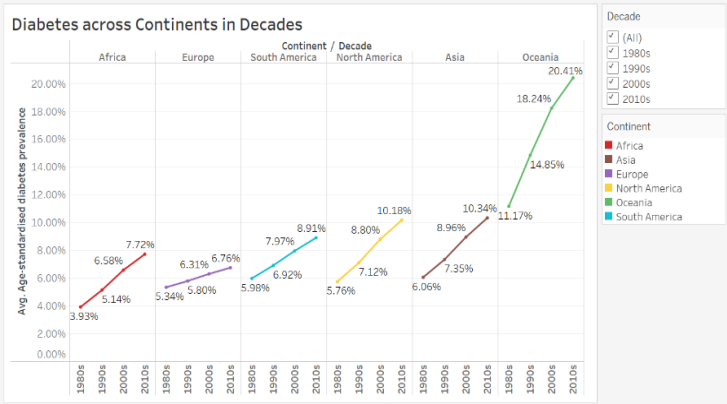
Finding 1

In the Oceania region, the key contributors to diabetes are actively managing the condition, while other contributors show a reduced commitment to preventive measures.

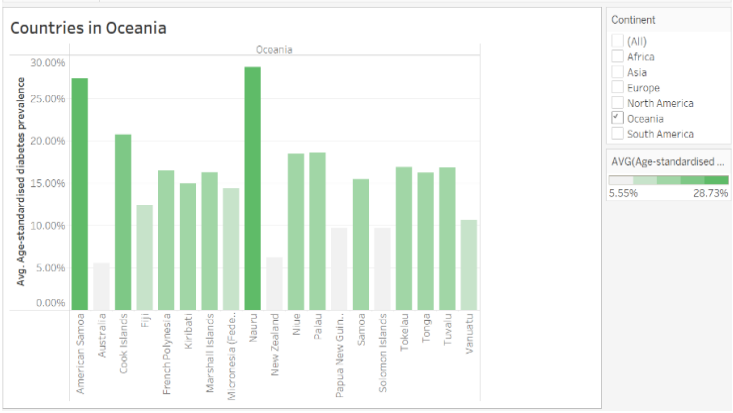
We are **analyzing** the diabetes dataset to see the overall trend of diabetes over the whole period. The Dataset holds the Diabetes Prevalence of Women and Men from 1980 to 2014 across different countries. The dataset type is Table with attributes Countries (Geographic), Years (Number), Sex (Categorical), Diabetes Prevalence (Number).



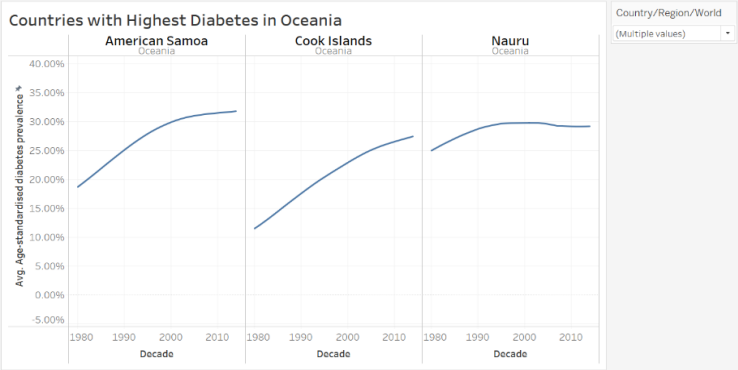
In this chart, Channel is line, and it is mapped to years(x-axis) and diabetes(y-axis). We have chosen line to show trend. The Diabetes Prevalence is converted to percentage for easy understanding, here in this Line chart, Diabetes is showing increasing trend over the years across all the countries.



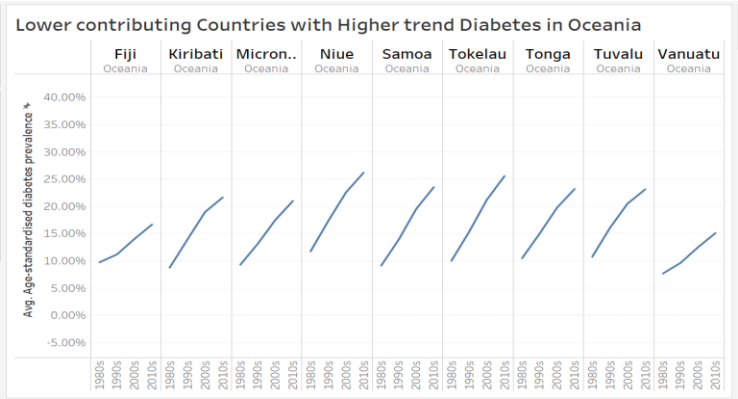
Here the line is used as channel and colors are used as marks. Points and labels are used for annotation. Line is used to show trends. It is mapped to diabetes on Y-axis and decade, continent on x-axis. Points are used to find each decade; label is used to show average diabetes in a decade and color is used to show each continent. We have not used the bar chart here as it does not clearly show the trend but shows only magnitude. We have converted different countries into continents to find which region has the highest trend percentage of diabetes. In this chart, it has been discovered that Oceania is making the largest contribution towards diabetes. Hence, we have chosen to narrow down our focus to this continent to **browse** which countries handle the notably high average diabetes rates.



Now, we are **searching** which countries are responsible for the increasing trend, particularly in Oceania. Here vertical line is used as channel and color luminance is used as mark. A bar chart has been used to show the countries and percentage of diabetes in Oceania. Here bar chart is used because we wanted to see the magnitude and color used to make those higher contributing countries stand out from the rest. From here it has been explored that “Nauru”, “American Samoa” and “Cook Island” countries appeared to be having the major contribution while "Fiji," "Kiribati," "Niue," "Samoa," "Tokelau," "Tonga," "Tuvalu," and "Vanuatu,” have lesser contribution as compared to higher contributors towards diabetes in Oceania.



Here, we are **identifying** the trend of the top contributing countries & **comparing** them with the other countries. Here the line is channel. Line chart is used to find out the trend over the decades in the that in "Nauru," "American Samoa," and "Cook Islands," which are the highest contributing countries in Oceania region. It's clear the diabetes trend is stabilizing in American Samoa and Cook Islands, and it is decreasing in Nauru.



Here the line is channel and Line chart is used to show the trend of lower contributing countries. It has been found that lower contributing countries are taking part in a higher rate from the past decades and still not slowing down.

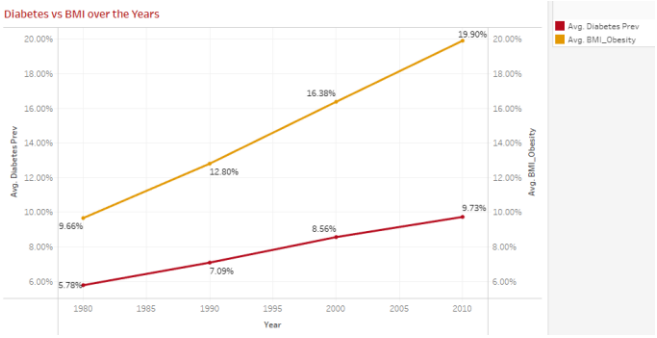
In **summary**, when examining the top three contributors, we saw that they are stabilizing and even displaying a decreasing trend. This finding supports our hypothesis that these major contributors are actively working to manage diabetes. On the other hand, when we looked up few other contributors in Oceania, such as "Fiji," "Kiribati," "Niue," "Samoa," "Tokelau," "Tonga," "Tuvalu," and "Vanuatu," it is apparent that their diabetes trends have been increasing over the decades and have not yet shown signs of slowing down. Consequently, it is highly likely that Oceania will continue to rank high in terms of diabetes contribution in the coming decades.

Finding 2:

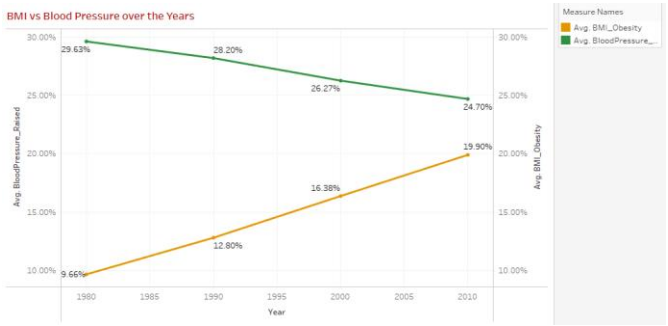
African's Health Puzzle: Connecting the Dots Between Rising BMI and Diabetes to know if female is the highest contributor in the world in terms of diabetes w.r.t BMI

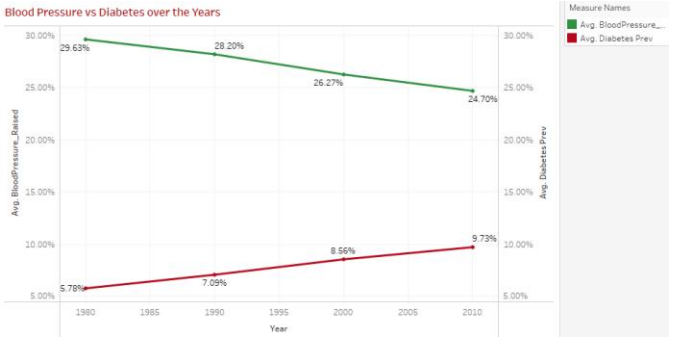
Quality of Finding: The purpose of this worksheet is to illustrate the percentage difference in average prevalence of diabetes & prevalence of high blood pressure versus BMI in various nations.
The visualization is displayed using a Line Chart view.

Dataset Type: The dataset type used as a Table in this case. This table has items such as countries and qualities such as year, avg prevalence of diabetes, and average prevalence of BP. Items and attributes are the datatypes that are used.



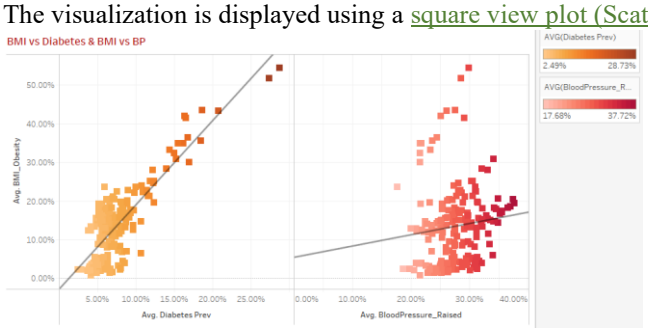
Type of Attribute: The attribute is found as a proportion of the avg BMI prevalence, avg prevalence diabetes & avg prevalence of blood pressure. The attribute is quantitatively ordered.
Sorting/Filtering/Highlighting: The filter here used is Decade.





Analysis used here is to figure out the increase or decrease in the % of BMI, % of diabetes & % of blood pressure. Target is to view the relation between the three - BMI, Diabetes & BP. Here, we have tried to view the value of the BMI correlates with the other two attributes (diabetes & blood pressure). In this chart, Channel is line, and it is mapped to years(x-axis) and diabetes & BMI (y-axis). We have chosen a line to show trends and the relation between the chosen attributes over a period. The marks here are the 1D line & color. This design helps in showing the % increase or decrease in BMI, diabetes & blood pressure. With search we are trying to find the relation between the avg prevalences.
With the above line chart, we can see that with an increase in BMI, there is an increase in diabetes while on other hand there is decreases in blood pressure w.r.t BMI & diabetes.

Quality of Finding: The purpose of this worksheet is to illustrate the relationship between the average prevalence of diabetes & prevalence of high blood pressure versus BMI in various nations across the globe.
The visualization is displayed using a square view plot (Scatter Plot).



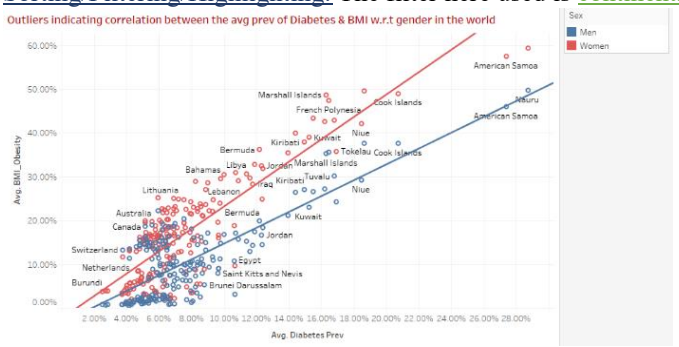
Dataset Type: The dataset type used as a Table in this case. This table has items such as countries and qualities such as year, avg prevalence of diabetes, and average prevalence of BP. Items and attributes are the datatypes that are used.
Type of Attribute: The attribute is found as a proportion of the avg BMI prevalence, avg prevalence diabetes & avg prevalence of blood pressure & countries. The attribute is quantitatively ordered.
Sorting/Filtering/Highlighting: The filter here used is Countries/nations.

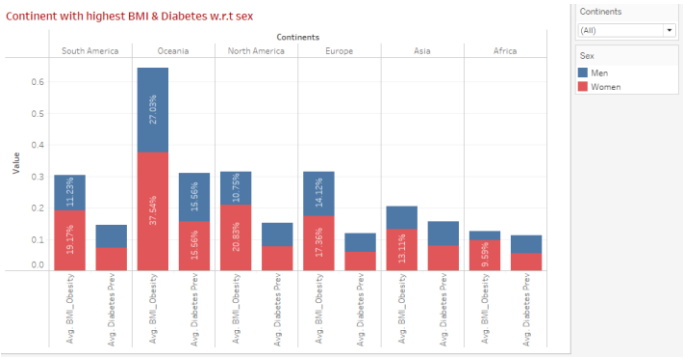
Analysis used here is to figure out the relationship between the % of BMI vs % of diabetes & % of blood pressure. Target is to view the relation between the three - BMI, Diabetes & BP. Here, we have tried to view how strong does the BMI correlates with the other two attributes (diabetes & blood pressure). In this chart, Channel is Color & vertical & horizontal position. We have chosen square

view (scatter) plot to show the relationship between the three attributes & how strong do they correlate by means of linear correlation. The marks here are the points. This design helps in finding the strong relation of the BMI with diabetes & blood pressure. The line crossing in between the chart is the linear regression of the attributes. With search we have tried to analyze the correlation between the prevalences.
With the above scatter plot chart, we can see that there is a strong relation between BMI & prevalence diabetes as compared to BMI & prevalence blood pressure.

Quality of Finding: The purpose of this worksheet is to illustrate the relationship between the average prevalence of diabetes versus BMI in various nations across the globe w.r.t gender over the years. The increase in one measure results in an increase in other measures w.r.t countries.
The visualization is displayed using a Shape Plot view (Scatter Plot) & Stack Bar chart.

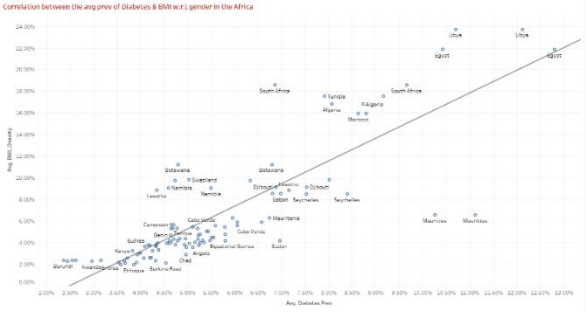
Dataset Type: The dataset type used as a Table in this case. This table has items such as countries and qualities such as year, sex, avg prevalence of diabetes, and average prevalence BMI. Items and attributes are the datatypes that are used.
Type of Attribute: The attribute is found as a proportion of the avg prevalence diabetes & avg prevalence BMI, countries & years. The attribute is quantitatively ordered, sex is the categorical attribute.
Sorting/Filtering/Highlighting: The filter here used is continents





Analysis used here is to plot the prevalence from the two datasets to find the outliers. Here, we have tried to analyze the correlation between diabetes & BMI over the country w.r.t sex. Further we have used stack bar chart to get the overall view of the scatter plot analysis. The search used here is to find the country where the % of prevalence is high or low as compared to other countries w.r.t sex. Query is to find the country with the highest/lowest prevalence for both the sex & later deep dive w.r.t continent to find which continent contributes the **lowest** in the world. The channel used here is vertical & horizontal position (both visual) & color (sex). The mark used here is Point & Line. The country's name & sex count is used as label.
We have used this visual representation which quickly helps us in finding the outliers with highest prevalence for sex w.r.t other countries with the lowest prevalence. It also shows the clear picture of the BMI prevalence of women across the globe is highest then men, while on the other hand, the prevalence of diabetes is higher is men around the world then the women. On the second chart, we can have a summarize overview of which continent contributes the highest prevalence of diabetes & BMI and which continent contributes the least w.r.t sex.

Quality of Finding: The purpose of this worksheet is to illustrate the Highest contributing countries in Africa w.r.t highest average prevalence of diabetes & BMI w.r.t sex. Since the continent has the lowest contribution in the world, we need to find the patterns in it. The overall countries in the continent do contribute to the prevalence of BMI & diabetes.



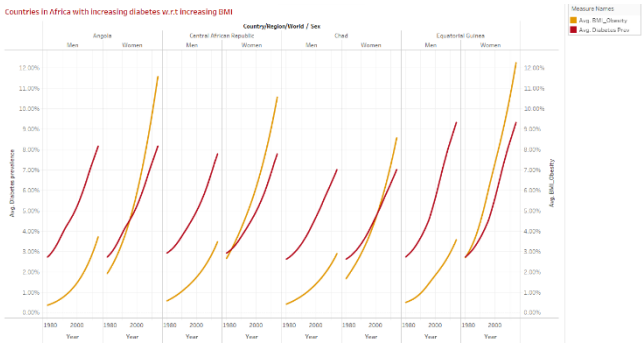
The visualization is displayed using a Tree map view.
Dataset Type: The dataset type used as a Table in this case. This table has items such as countries and qualities such as avg prevalence of diabetes, and average prevalence BMI.
Type of Attribute: The attribute is found as a proportion of the avg prevalence diabetes & avg prevalence BMI, & countries
Sorting/Filtering/Highlighting: The filter here used is continent

The analysis used here is to plot the prevalence from the two datasets to find the highest contributing countries in the Africa continent. The target here is based on strong correlation between the diabetes & obesity of the continent-Africa also resulting in finding the outliers. The search used here is to find the country where the % of prevalence is high compared to other countries. Query is to find the country with the highest prevalence for both the sex. The channel used here is position. The mark used here is Point. The country

We have used this visual representation which quickly helps us in finding the countries with highest prevalence in the Africa continent. We can easily find that the countries such as Central African Republic, Chad, Angola, Equatorial Guina, Ethiopia, Madagascar, Eritrea, Burundi and many other contribute a significant amount of prevalence to its continent.

Quality of Finding: The purpose of this worksheet is to illustrate the relationship between the average prevalence of diabetes versus BMI in the Africa's countries w.r.t sex over the years. The increase in one measure results in an increase in other measures w.r.t countries & sex over a period.
The visualization is displayed using a Line chart.

Dataset Type: Dataset type used as a Table in this case. This table has items such as countries and qualities such as avg prevalence of diabetes, and average prevalence BMI.
Type of Attribute: The attribute is found as a proportion of the avg prevalence diabetes & avg prevalence BMI, & countries. The attribute is quantitatively ordered, sex is the categorical attribute.
Sorting/Filtering/Highlighting: The filter here used is Continent>Countries.



Analysis used here is to plot the prevalence from the two datasets to find the relationship between highest contributing countries in the Africa w.r.t sex over a period despite being the least contributor in the world. The target here is to find the countries which have a ratio of contribution by either gender in the correlation of BMI & diabetes. The search used here is to find the relationship between the % of prevalence in countries of Africa & the relationship between the trends.

Visual representation quickly helps us in finding the relationship between the countries with highest prevalence in the African continent w.r.t sex over a period. We can find that even if there is variation in men, but the female shows a much greater contribution to the world's % prevalence. We can easily identify with the line chart that with the increase in one measure (BMI), there is significant increase in diabetes approximately.

Summary:

When we looked at the top four donors, we noticed that there is a rise in prevalence of diabetes with respect to increase in BMI. When we checked into a few added African contributors, such as "Algeria," "Angola," "Central African Republic," "Chad," "Burundi," "Equatorial Guinea," "DR Congo," "Eritrea," "Nigeria," and "Ethiopia," we discovered that their diabetes patterns have been growing throughout the decades and show no indications of slowing down. Both genders in the African region have a high contribution towards obesity & diabetes. We were able to see that in both genders there is a significant rise in diabetes over a period w.r.t increase in obesity level. We can also see that there is slightly low relation in men's obesity level & diabetes level, but the female has a high contribution. The contribution of the African population is higher as compared to other countries in the world. Thus, with this we can see that female are the most highly effected & in higher number of contributions of prevalence of diabetes in the world.

Thus, the hypothesis justifies that on an avg when the prevalence BMI is high for the female, the chances of them having diabetes is also high.