Data Visualization with Tableau

Data and analysis

1. **What**
   1.1.Dataset type:

The type of the dataset is Table. Tables are made up of columns and rows. Each column represents an attribute and row represent the item. Cell is the intersection point of a column and a rom which contains its value.

   1.2.**Data type:**

The data type of the dataset is items and attributes. Dataset is given in table form which means column in the data represent attributes whereas, rows represent items which is also known as a discrete entity.

   1.3.Attribute type:

The following attributes are in the given dataset

country, sex, year, BMI, diabetes, raised blood pressure

   Categorical Attributes: country, sex

   Ordered attributes (Ordinal): year

   Quantitative: BMI, diabetes, raised blood pressure

2. **The 'Why'**
   2.1.Actions
   1. **Consume**
      The purpose of this coursework is to consume information from the given dataset.
      Discover:
      In consume, there is need to discover new knowledge which is not known previously, as the questions are given to find pattern and relationship for between the measurement over time and across different regions.
   2. **Search**
      The search is whether the location and target are known for questions or not.
      Explore:
      Explore is a search type in which user is unknown about location and target. The given questions are asked to find the pattern and relationship, but the location is not known where to look for the pattern. Target is also unknown because there is no information which measurement pattern and which measurement has relationship with other measurement.
   3. **Query**
      Identify: (Question 1)
      For first question there is need to identify any pattern for each measurement. It can identify which measurement has more/less prevalence of BMI, raised blood pressure and diabetes.

Compare: (Question 2)

Compare is used for multiple targets. In the second question, there is need to check whether a relationship exist between two or three measurement over time and across different regions.

2.2.Targets

1. Data

Trend: Both questions are asked to find pattern and relationship individually and between two measurements which means there is need to find trend. Trend can be shown over time and across different regions.

2. Attributes

One-Distribution (Question 1)

For question 1, there is need to find pattern for each measurement which means how values of the BMI, raised blood pressure and diabetes are distributed over time and across regions.

Many-Correlation (Question 2)

In the question 2, there is need to find relation between two or three measurement over time and across regions, so correlation has been used to show relationship. Correlation check whether the values of second attribute are tied to first attribute or not.

The Quality of Findings

Question No. 1:

Is there any pattern for each of the measurements?

1. Over time

Finding 1.1: pattern for each measurement over time

The line graph illustrates average prevalence of three measurements (prevalence of BMI>=30kg/m<=obesity, Age-standardized diabetes prevalence, prevalence of raised blood pressure) over 36 years from 1978-2014.

    i.   Overall, the average prevalence of BMI and age-standardized diabetes prevalence for both sex (color channel) increased over the given period, while prevalence of raised blood pressure, it decreased.

    ii.   Size of the line represents the rate of prevalence. The thicker the line, higher the value of prevalence. Although the prevalence of the raised blood pressure reduced over time, it remained higher than prevalence of BMI and diabetes.

    iii.   Over the time, men prevalence for raised blood pressure remained higher than women.

    iv.   For BMI, women have more prevalence than men over the time

For diabetes, women had high prevalence in the start of the period. However, men prevalence increased showing higher prevalence at the end of period in 2014.
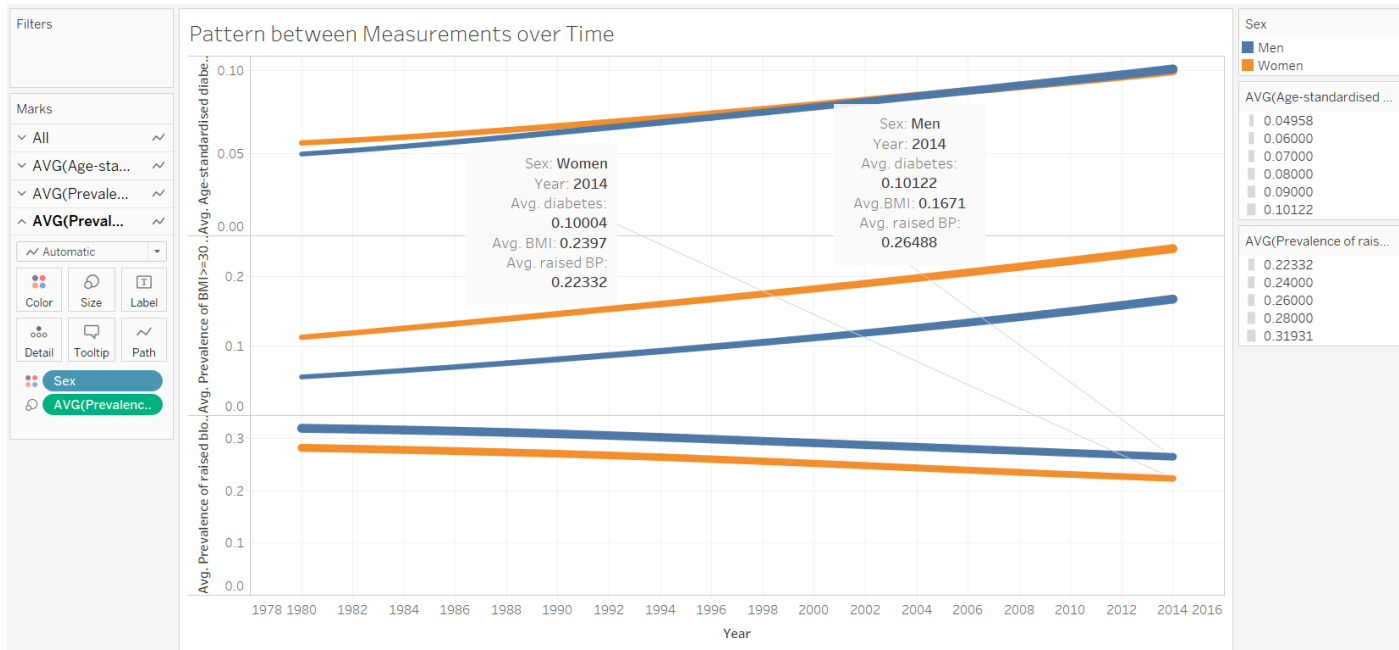
Finding 1.1: Pattern Over Time

3. The Effectiveness of the Visulaization Design(Finding 1.1)
3.1. Marks and Channels:

Mark: Line is one dimensional (1D) channel used to represent pattern over time. Line chart maps 'year' attribute which is ordinal which means it is in order but cannot peform numerical computation. To show trend for continuous attributes line chart is suitable.

Channel:

- Color: It represnets the categorical attribute 'Sex' to represent Women and Men.
- Size: To represent how smaller/bigger value of prevalence of BMI, diabetes and blood pressure which all are quantative attributes.

3.2. Effectiveness of channels and Visulaization

Channels

- It is more effective to use Color identity channel to represent categorical attributes.
- Size channel is used for quantative attribute because this channel is best suitable for magnitude.

Visual Mapping

- The questions is supposed to find the pattern between measurements. Line chart visual mapping has been used to find the trend relationship between quantative measures BMI, diabetes and raised blood pressure.
- Data: it uses one quantative attribute and one ordered key attribute.
- Scale for the key attribute can be mapped to hundred of levels.
- Arrangment Layout: Rectilinear spatial
- Line chart is used for ordered keys only

3.3.Alternative Design :

Dot Chart:

- Alternative could be Dot chart which also uses one quantative value and one ordered key
- It uses Point mark which is zero dimensional (0D) and expresses value of quantative attribute in the vertical position of ordered key attribute but it doesn't connect point which is unable to show trends. This is the reason dot chart can not be suitable for this question.

2. Across different countries/regions?

The quality of Findings

Finding 1.2: Prevalence of BMI>=30 across different regions

- The prevalence of BMI>=30 is lower for most of the countries in Africa and Asia. However, North America, Europe, south America and Oceania shows higher prevalence of BMI>=30.

Finding 1.2: BMI across Region

Finding 1.2: Prevalence of age-standardized diabetes across different regions

- Canada and Greenland and some countries in Africa shows less prevalence of diabetes, whereas Europe, South America and Oceania shows average prevalence.
- Moreover, some countries of Africa and Asia shows highest prevalence.

Finding 1.2 Diabetes across Regions

Finding 1.2: Prevalence of raised blood pressure across different regions

- North America shows less prevalence of raised blood pressure except Greenland whereas, Europe, Oceania, Africa, Asia excluding china and South America has more prevalence of raised blood pressure.

Finding 1.2: Blood Pressure across region

The Effectiveness of the visualization design

Marks:

Area(2D): To find the pattern of each measurement over the region.

Channels:

Color: which represents the prevalence value of each measurement. The darker the color higher the value of prevalence. Lighter color represents lower value.

Effectiveness of Visual Mapping:

Choropleth map (filled gaps) is used to find trend across all the regions for BMI, raised blood pressure and diabetes. This is suitable because filled color on map can indicate if a region has darker color, it has higher prevalence of measurements and lighter color means lower prevalence.

Alternative Visual Design:

Symbol map (proportional) can be used as alternative for the above visualization. However, symbol map is more suitable to find quantitative values for individual country/area. But, the purpose of the question was to find pattern across all region. So, symbol map is not good option to show pattern.

Question No 2:

Is there any relationship between two or three measurements?

    2.1 Over time

The quality of finding

Finding 2.1: Correlation relationship between BMI, Diabetes and raised Blood Pressure

- BMI and diabetes have strong positive correlation over time which means higher the BMI or obesity there are more chances people getting diabetes.
- BMI and Raised blood pressure show negative correlation which means increase in BMI level doesn't affect the prevalence of blood pressure.
- Diabetes have also negative relationship with raised blood pressure which means it's not necessary people who have diabetes will suffer from raised blood pressure. Because there is strong negative relationship.
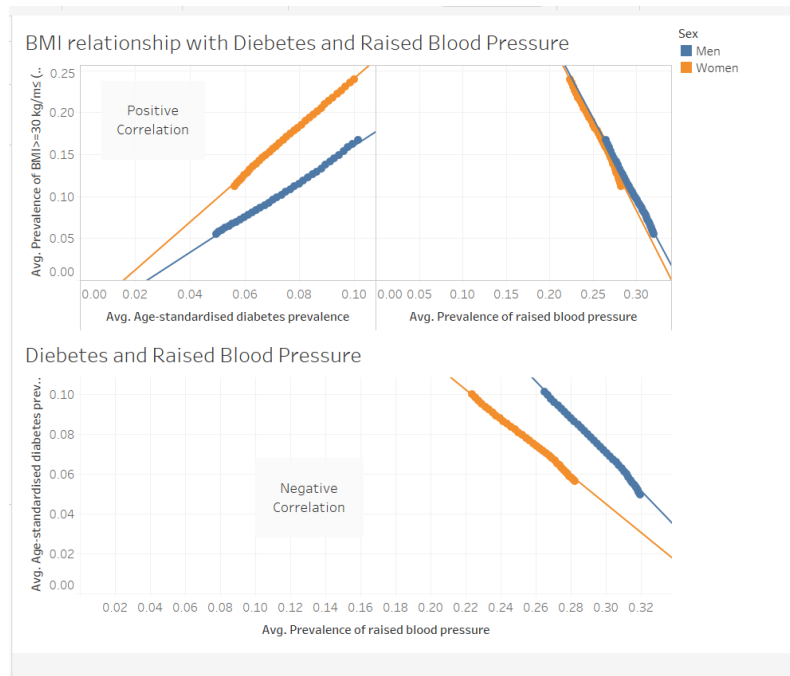
Finding 2.1: Relationship over time

Marks:

Point: which represents a 'Year' attribute with horizontal position and vertical position and quantitative attributes prevalence of BMI, diabetes and raised blood pressure are used.
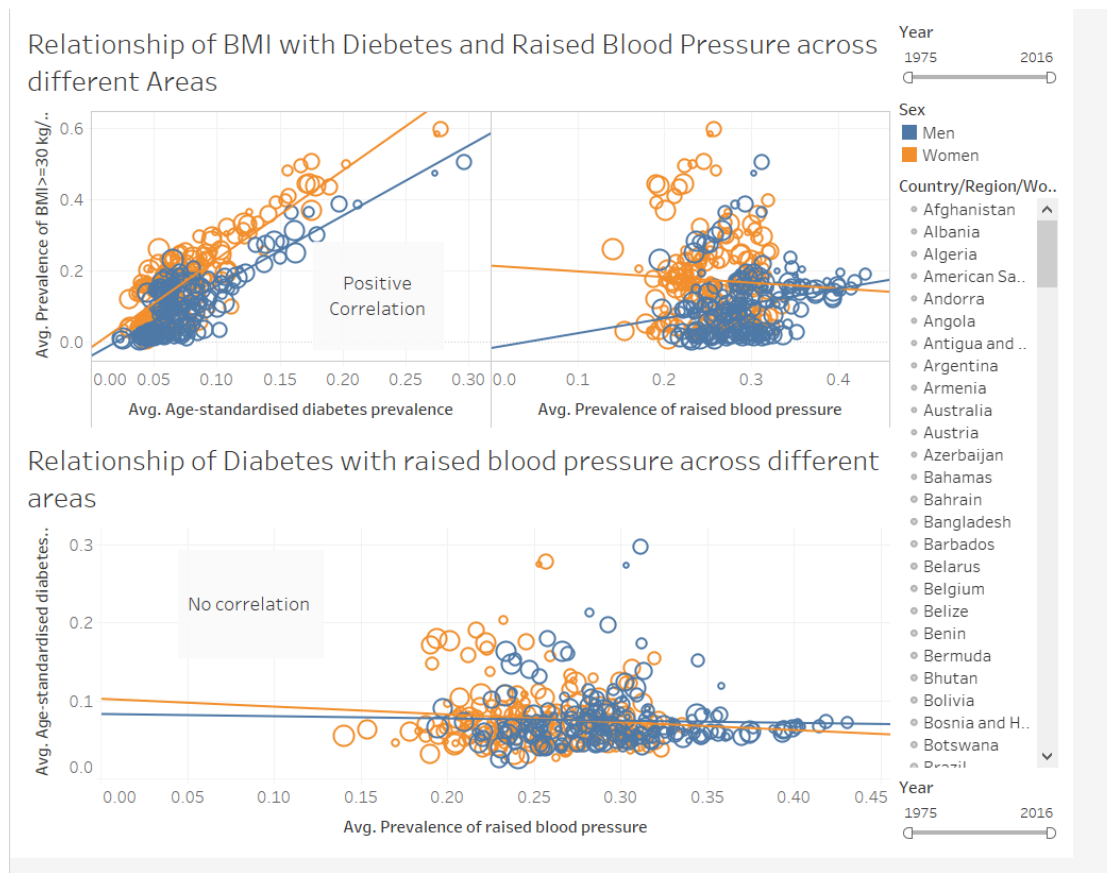
Channel:

Color: Color channel has been used for categorical attribute sex to represent one color for women and one for men. Color has more effectiveness than motion and shape in the ranking.

2.2 Across different area?

Finding 2.2: Correlation between BMI, Diabetes and Raised Blood Pressure

- From the visualization, can see BMI and diabetes have strong correlation for women and men over the different region which means higher obesity level can cause diabetes. This trend is found all over the world.
- However, it can be clearly seen that BMI and diabetes doesn't have any relationship with raised blood pressure across regions. It reflects that higher obesity level or diabetes has no relationship with raised blood pressure for all the regions.

Finding 2.2: Relationship across Regions

Mark (Finding 2.2):

Point (0D): which represents 'country' attribute which is categorical attribute and two quantitative attributes on x-axis and y-axis.

Channels (Finding 2.2)

Size: Size channel has been used for categorical attribute. Bigger the size of circle means bigger the size of region.

Color: It represents categorical attribute 'sex' to differentiate Male and Female.

Effectiveness of Visual Mapping: (Finding 2.1, Finding 2.2)

For visualization, rectilinear layout scatter plot has been used to find correlation between three measurements. It uses two quantitative attributes, and values are mapped with horizontal and vertical spatial position. It is effective because question was to find relationship, so correlation can be visualized using scatterplot. Scatterplot is suitable for finding trends, outliers, distribution and correlations. It can visualize hundreds of items.

Alternative visual design: (Finding 2.1, Finding 2.2)

Parallel coordinates: It is parallel layout to arrange table data. This plot is suitable for more than two attributes to find correlation. The problem is it is very difficult to understand parallel coordinates when there is dozen thousands of items. Lines in the mapping gets overlapped which makes it difficult to interpret.

Use of Advanced Features: (For both questions)

1. Story (Finding 1.2)
   For Finding 1.2, story has been used to show three visualization each representing BMI, diabetes and raised blood pressure pattern across different regions. Story has been used to show single map. Dashboard was not effective because map might not be visible in the dashboard.
2. Dashboard (Finding 2.1, Finding 2.2)
   For Finding 2.1 and 2.2 , Dashboard has been used to merge two visualization in one dash board to depict relationship between all three measurements. Two visualization with scatterplot has been merged.
3. Filtering (Finding 2.2)
   Filtering has been used in Finding 2.2 to check correlation between different regions for prevalence of BMI, diabetes and raised blood pressure over time.