

MSc Data Science

Coursework 1

Data Visualization with Tableau

**Table of Contents**

[➢ Objectives 4](#_Toc150961331)

[Problem statements: 4](#_Toc150961332)

[Hypothesis: 4](#_Toc150961333)

[➢ Data Processing 5](#_Toc150961334)

[Data Collection 5](#_Toc150961335)

[Data Engineering 5](#_Toc150961336)

[1. Data Source Connection 5](#_Toc150961337)

[2. Data Blending 5](#_Toc150961338)

[➢ Findings 7](#_Toc150961339)

[Finding 1: *(Worksheets Colored Red- Sheet name & calculated measures prefixed with A)* 7](#_Toc150961340)

[1. Quality of Finding1 (A Scenario): 7](#_Toc150961341)

[2. What 8](#_Toc150961342)

[3. Why 9](#_Toc150961343)

[4. How 10](#_Toc150961344)

[Finding 2: (Worksheets Colored Blue - Sheet name & calculated measures prefixed with B ) 12](#_Toc150961345)

[1. Quality of Finding2: 12](#_Toc150961346)

[2. What 13](#_Toc150961347)

[3. Why 14](#_Toc150961348)

[4. How 17](#_Toc150961349)

[➢ Future Enhancements: 18](#_Toc150961350)

[➢ Research, Conclusion, and Recommendation: 18](#_Toc150961351)

[➢ Appendix: 21](#_Toc150961352)

[Data Cleansing 21](#_Toc150961353)

[Statistical Definitions 21](#_Toc150961354)

[Extra/Relevant Significant Insights Driving the Report findings: 22](#_Toc150961355)

[1. Storyboard A 22](#_Toc150961356)

[2. Storyboard B 24](#_Toc150961357)

**Table of Figures**

[Figure I Finding 1 Source – Left Join BMI and ContinentRegions Tables 6](#_Toc150945319)

[Figure II Finding 2 Source 6](#_Toc150945320)

[Figure III Top Average BMI over Countries/Years 7](#_Toc150945321)

[Figure IV Analyze – Action Figure V Analyze - Action 2 9](#_Toc150945322)

[Figure VI Sorting, Filtering, highlighting (Finding 1) 11](#_Toc150945323)

[Figure 7 Finding 1 - Dashboard 11](#_Toc150945324)

[Figure 8 Finding 2 - Dashboard 12](#_Toc150945325)

[Figure X Actions 2 15](#_Toc150945326)

[Figure XI B3 Z-score Formula 16](#_Toc150945327)

[Figure XII Sorting, Filtering, Highlighting (Finding 2) 18](#_Toc150945328)

# Objectives

The objective of this analysis is to comprehensively analyze three medical datasets, encompassing blood pressure, diabetes prevalence, and BMI, with a specific focus on uncovering longitudinal patterns over the last four decades (1975-2016) and identifying extreme health indicators worldwide. By examining these datasets, we aim to gain insights into global health trends, including variations, disparities, and potential risk factors.

Additionally, this analysis seeks to propose solutions and recommendations for public health interventions, informed by the identified patterns, to address and mitigate extreme health outcomes and enhance global well-being.

## Problem statements:

* Examine how health indicators, including BMI, diabetes, and blood pressure, vary across different demographic groups (gender, time, regions, countries, and continents).
* Examine trends in BMI over time and identify regions or demographics with increasing or decreasing BMI.
* Investigate the relationship between BMI and the prevalence of diabetes.
* Determine if there is a threshold BMI beyond which the risk of diabetes significantly increases.
* Analyze the distribution of blood pressure levels in different geographic continents.
* Explore how the prevalence of diabetes has changed over time.

## Hypothesis:

* Hypothesis 1: Countries with sedentary lifestyles have high BMI prevalence.
* Hypothesis 2: Diabetes and BMI are directly proportional to each other, meaning that they are positively correlated.
* Hypothesis 3: Men with diabetes have a higher probability of having a higher BMI and vice versa, compared to women.
* Hypothesis 4: There is a link between health issues like BMI in areas where atomic experiments were conducted in the past like the South Pacific region.
* Hypothesis 5: Changes in dietary habits are associated with variations in BMI and diabetes prevalence over time.
* Hypothesis 6: There is a gender-based difference in the relationship between BMI and blood pressure, with males showing a stronger correlation.
* Hypothesis 7: Countries with high obesity levels have high diabetic prevalence.

# Data Processing

## Data Collection

Elaboration below on how data blending is used to combine multiple data sources and multiple data sheets in Tableau.

* General Data provided in Coursework: **cw1 -dataset.xlsx**
* An additional data set Continent/Region is download and aggregate ContinentRegions.xlsx file as from:
  + ‘Statistic Times’. (United Nations, 2023)
* 2 data sources were created “**cw1-dataset” using**  **cw1 -dataset.xlsx** which loaded 3 tables of Blood Pressure, BMI, and Diabetes, and “**ContinentRegion“ using ContinentRegions.xlsx which** loaded one table**.**
* Two techniques were applied for data preparation and join:
  1. **VLOOKUP** function in Excel was used to join the main CW tables with the incorporated 2 additional fields (Continent, Region) resulting in a new dataset of ContinentRegions.
  2. **Table join** in Tableau was used to join the 3 tables of the cw dataset as illustrated below in addition to joining with the newly added dataset/data source of ContinentRegions including cw1 -dataset.xlsx and ContinentRegions.xlsx.

## Data Engineering

**Null** or **missing** values were identified during testing of the joins between the ContinentRegion data source and the CW tables, accordingly, data cleansing of completing those was conducted manually in the Excel source file**. (refer to the fixed records in the appendix~~)~~**

### Data Source Connection

Finding A - analyze one measure:

In Tableau click on "**Connect to Data**” and upload the **cw1 -dataset.xlsx** file. The first data source created is **“BMI (cw1-dataset)”**, and the BMI table is used for Finding 1 and the Continent/Region dataset from the ContinentRegions.xlsx file loaded in data source 2.

Finding B - analyze 2 or 3 measures: ­­­­

The same procedure is followed, however, Finding 2 will do an Inner Join among the 3 tables in Data source 1 (BMI is the main table as it contains the largest data volume). Also, join the main BMI table with Vlookup a table Continent/Regions from data source 2 using the Left Outer Join method.

### Data Blending

In Data Source 1, relationships are based on using **composite keys 'Country', 'Year' & and 'Sex’**, while joining with Data Source 2 a **single key** “**Country**” is used for joining with the **Continent/Region** data table**.**

Tableau automatically recognizes the common field having exact names within joined tables, and duplicate columns among tables are removed manually.

* + Finding A and B linking Countries with Continent/Region we are using Left outer join,
  + Findings B linking BMI as the main table using Inner join with BP and Diabetes tables from cw1 -dataset.xlsx.

Below are screenshots showing data sources and joining techniques for each finding:

Finding 1/A:

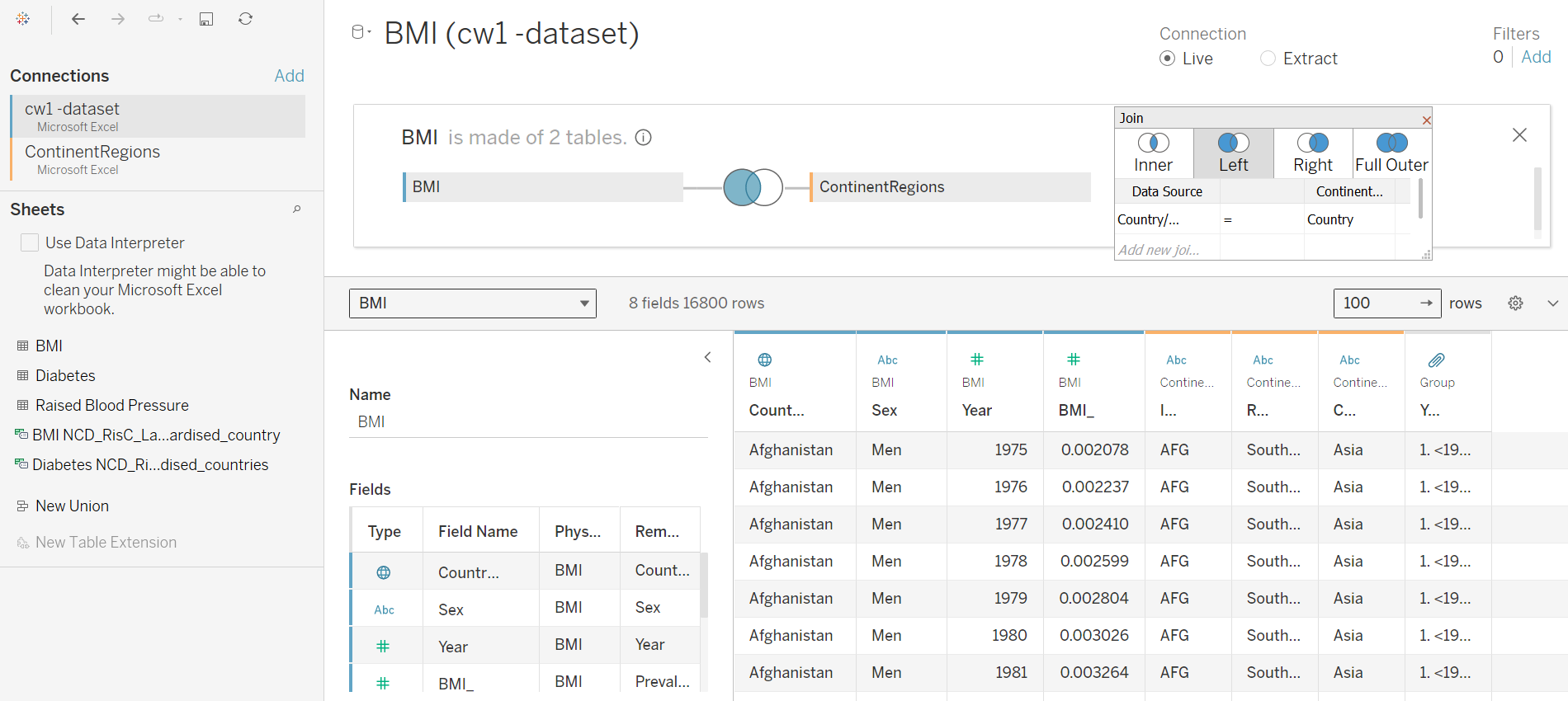
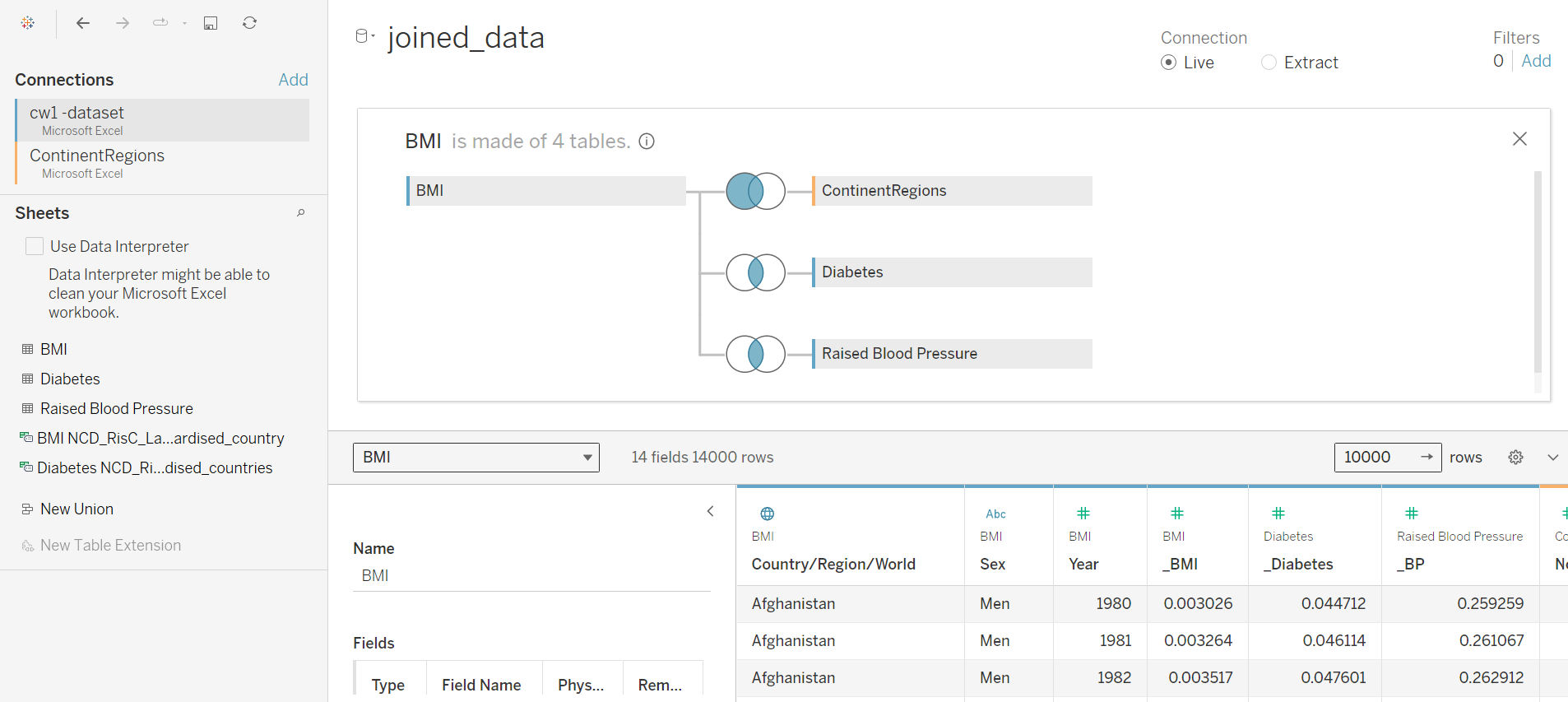


Figure I Finding 1 Source – Left Join BMI and ContinentRegions Tables

Finding 2/B:



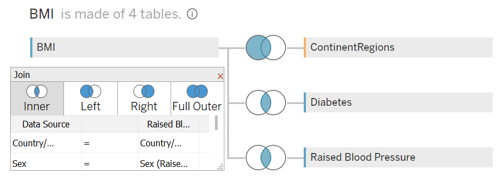
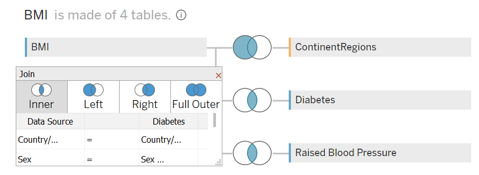
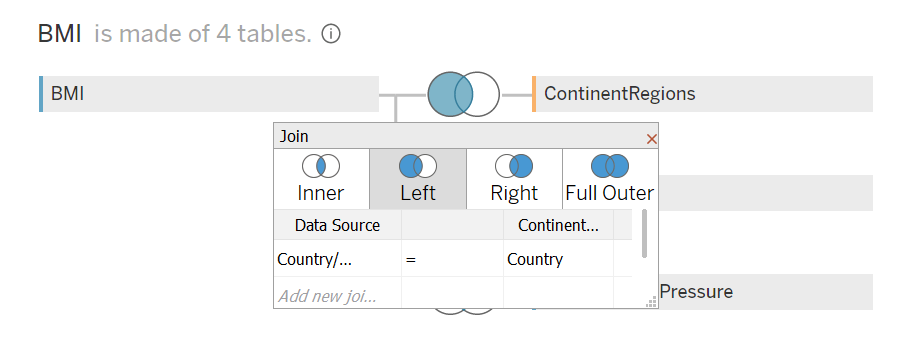


Figure II Finding 2 Source

# Findings

**Note**: For ease of segregation between the 2 findings visualizations in the tableau workbook, a nomenclature is adopted.

* The first finding worksheet label and calculated fields are prefixed with a letter (A) and the Second finding with a letter (B)
* Worksheets colored in Red for Finding 1 and Blue for Finding 2
* Calculated measures are prefixed with sequential numbers. example, B1 (measure) is mainly used in B1 visualization, and few may be used in the following sheets so B1 Can be used in visuals B2, B3, etc.

## Finding 1: *(Worksheets Colored Red- Sheet name & calculated measures prefixed with A)*

**South Pacific Islands** exhibit the **highest BMI** as compared to the rest of the world, with **Nauru** having the highest BMI level **worldwide in** **2016**.

### Quality of Finding1 (A Scenario):

The below visual illustrates that **Nauru** has the highestBMI spotted with an average of **0.54** worldwide from 1975 to 2016 across 200 countries. It was also observed that the top 10 countries with the highest BMI were South Pacific Islands.

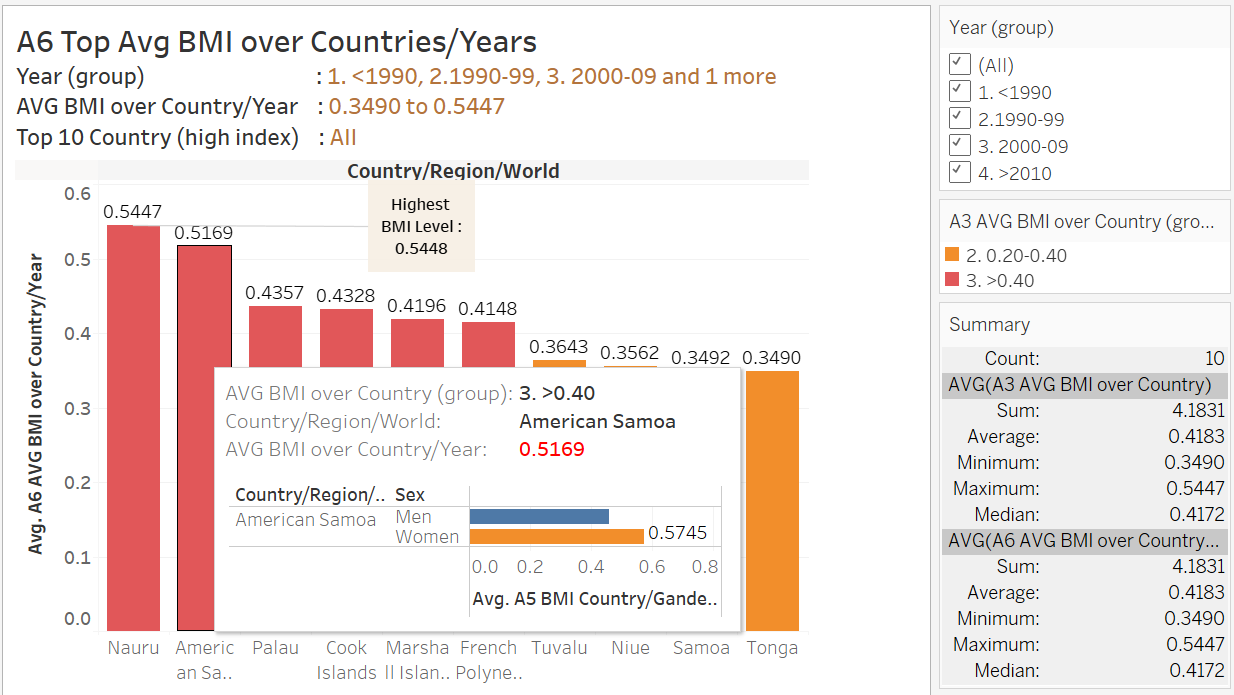


Figure III Top Average BMI over Countries/Years

**Factor/s**: Average BMI

According to the WHO organization, a Body Mass Index **over 30 kg/m2** is considered obese which means the Republic of **Nauru** has a severe health concern given that **94.5% are obese**. (World Health Organization, n.d.)

The island nation of Nauru in the South Pacific has suffered greatly because of extensive phosphate mining. Phosphate mining was the nation's primary source of income in the past, but it also severely degraded the ecosystem, causing biodiversity loss and soil erosion. Agriculture has been severely harmed by this environmental upheaval, which has decreased arable land and increased dependency on imported, processed commodities.

As a result, an increase in BMI levels has been noted. As phosphate reserves decreased, the country's economy grew increasingly reliant on phosphate exports, raising questions about sustainability over the long run. This economic reliance has contributed to Nauru's socioeconomic environment, impacting access to basic services and the distribution of resources. A multifaceted strategy that seeks to strike a balance between public welfare, environmental preservation, and commercial interests is being used to solve these issues.

### What

* **Dataset Type :**

The NCD\_RisC dataset type is a **Table/Tabular** of **items/records row-wise** in 3 tables/sheets of BMI Body Mass Index, Diabetes, and Blood Pressure. For this finding, we will use only the BMI dataset.

**Finding 1** analysis will use one **data source,** using the **full BMI** data set of **16800 rows**.

The dataset is a comprehensive health dataset including **attributes** on the prevalence of specific health conditions, organized and combined based on several key dimensions, including Continent, Region, Country, Sex, and Year.

1. **Health Prevalence Data:** This data set primarily contains information on the prevalence of specific health conditions in various populations. The specific health conditions of interest are:

* Prevalence of BMI > 30 kg/m²: contains information on the proportion of individuals in each group (defined by Continent, Region, Country, Sex, and Year) with a BMI greater than 30 kg/m², indicating obesity.

1. **Dimensions for Aggregation:** The dataset is organized and combined based on three key dimensions:

* Country: The data is aggregated by country, making it possible to compare health conditions among different nations.
* Sex: The data is segmented by gender, allowing for the analysis of health conditions in males and females separately.
* Year: The data is collected over different years (1975-2016), enabling the analysis of trends and changes in health conditions over time.
* Continent and/or Region: an additional dimension from the internet to further group on region and continent level. (United Nations, 2023)
* **Data Type:**
* The data types used are **categorical** and **numerical datasets**.
* **Sources** of the dataset, a - **BMI (cw1-dataset)**

|  |  |  |
| --- | --- | --- |
| **Variables/Fields** | **Data Types** | **Role in Tableau** |
| Country/Region/World | Geographic Value | Discrete Dimension |
| Sex | String Value | Discrete Dimension |
| Year | Number (Whole) | Continuous Dimensions |
| Prevalence of BMI>=30 kg/m≤ (obesity) | Number (Decimal) | Continuous Measure |
| Continent / Region | String Value | Discrete Dimension |

Table 1 Data Type

* **Attribute Type**
* **Dimensions** or qualitative - Continent, Region, Country, Sex, Year.
* **Measures** or continuous numbers - Average BMI
* **Calculated Measures** :

**Used in Main Finding Visual**

* + A3 AVG BMI over Country = {FIXED [Country/Region/World] : AVG([BMI\_]) }
  + A5 BMI Country/Gender Fixed = {INCLUDE [Year]: AVG([BMI\_]) }
  + A6 AVG BMI over Country/Year = {FIXED [Country/Region/World], [Year] : AVG([BMI\_]) }

**Used in Extra Finding Visuals/Dashboard/Story**

* + A2 AVG(BMI) over Years = {FIXED [Year]: AVG([BMI\_])}
  + A3 AVG BMI over Continent = {FIXED [Continent] : AVG([BMI\_]) }
  + A7 #Countries with more BMI Index = WINDOW\_SUM(IF AVG([A7 Selected Difference]) > 0 THEN 1 END)
  + A7 % BMI from Selected Difference = ([A3 AVG BMI over Country] / [A7 Selected Country BMI EXCLUDE]) - 1
  + A7 Selected Country BMI = IF [Country/Region/World] = [Param Country] THEN [A3 AVG BMI over Country] END
  + A7 Selected Country BMI EXCLUDE = { EXCLUDE [Country/Region/World]: AVG([A7 Selected Country BMI])}
  + A7 Selected Difference = [A3 AVG BMI over Country] - [A7 Selected Country BMI EXCLUDE]

### Why

* **What does the visualization aim to show?**

The visualization aims to illustrate high-risk areas and pinpoint the top 10 countries where population exhibits significantly higher average BMI Index. This information can assist public health authorities and policymakers in allocating appropriate resources and designing tailored preventive measures**.**

* **Describe the ‘Actionsʼ from all three aspects: ‘Analyseʼ, ‘Searchʼ, and ‘Queryʼ;**

**Analyze**:

The Analysis unfolds the t**op 10 countries** with the highest obesity levels worldwide using BMI percentage falls in South Pacific islands. In addition to disparities between male and female groups per country using bar-chart visualization. The Percentage of BMI is Produced by deriving the average percentage of BMI compared to the world average.

Filtration and sorting uncover various notable interesting findings, such as most countries fall within the first BMI Level group, characterized by a percentage below 0.20.

Countries that were observed to have extreme values on the charts clarify a significant finding about the problem statement defined accordingly these are relevant to keep in the dataset. While for future analysis involving machine learning the extreme measure might represent outlier values depending on the model context.

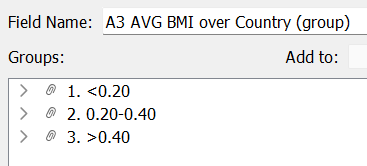
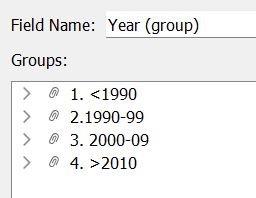
 

Figure IV Analyze – Action Figure V Analyze - Action 2

**Action 1**: create grouping field: A3 AVG BMI over Country (group): grouped obesity into three bands:

< 0.20, between 0.20 and 0.40 and greater than 0.40 based on average BMI Index:

In this way, we can filter data presentation on average BMI Country level and compare results.

**Action 2**: create grouping field: Year (group): represent decade groups of main Year field:

In this way, we can filter data presentation on Decade Year level and compare results.

Apart from this, the years from 1990 – 2010 were grouped where each year group represents a decade except for the first (15 years) and the last group for 6 years only. The new field created is a discrete dimension.

**Search** based on Year group; the data was grouped by obesity to browse for countries where the percentage is extreme compared to other countries.

**Query** seeks to analyze the percentage growth and decline across diverse countries and among various gender groups.

Filters can further uncover results and trends on various levels and combinations of Regions, Countries, Sex, Years, Year (group), or Avg BMI (group).

•A user can identify the results for one country.

•Compare the selection results of one country to another.

•Summarize the selection results across all countries.

* **Describe the ‘Targets’ from both the ‘Dataʼ and ‘Attributeʼ aspects.**

**Targets:** to depict the **distribution** of average BMI percentage. The utilized attributes represent extremes, elucidated through a bar chart visualization sorted in descending order for enhanced visual comprehension.

* **Trends:**

Trend lines graph used “to predict a continuation of a multiple variables’ trends. Additionally, identify correlations between two variables by observing the trend in both simultaneously.”[[1]](#footnote-2)

For instance:

*The trend of % Difference in Avg. AVG(BMI) over Years for Year.*

*R-Squared: 0.958624 (a measure of how well the data fits the linear model, gives us the percentage variation in y explained by x-variables.*

*The range is 0 to 1 (i.e., 0% to 100% of the variation in y can be explained by the x-variables.)*

*Equation: % Difference in Avg. AVG(BMI) over Years = -0.000296663\*Year + 0.619147*

*Conclusion here: That means in the next period of 4-5 years the trend will continue going down as a % of Difference in between years even the general trend will continue to have an intended factor of increase.*

* **Features:** we are usually working with AVG BMI disparities over different groups
* **Attributes:**

For Attributes as specific properties that are visually encoded, for Finding 1, we are using a single attribute. For example, find the extremes and Higher-level: the distribution of an attribute.

On the other hand, with finding 2 only, we are using other Target attributes such as correlation: the values of two attributes are tied together as a part of Multiple attributes.

### How

* **Visual mapping or encoding - Marks:** 1 Dimensional diagram, and color.
* **Channels: Bar chart distribution mapped** as the **Average BMI** attribute and **Color hue** as the Categorical Attribute to identify the **Year (group) and/or Avg BMI group levels** Attribute.
* **Features:**
  + **Filters: Top 10** Countries by field **[A3 AVG BMI over Country]** and Sort By **‘Nested’- Descending** order. also, a combination of **[A3 AVG BMI over Country (group)]** and **Year (group)**.
  + **Dashboard A (Fig 7 below)**
  + **Storyboard A:** presenting significant visuals elaborating Finding B (Appendix).
* **Why this Visual**: A **Bar Chart** is usually the best way of comparing dimensions and representing visuals for different attributes, especially categorical attributes, or tracking them over time. It is recommended to use when changes are noticeable. This design helps in identifying the percentage increase of **AVG BMI** among Countries, using **Year (group)** and **[A3 AVG BMI over Country (group)]** filter fields. Also, this visualization has an option of using **‘country mouse hover’ / tooltip** to show **gender disparities** on **the Country level**.
* **Sorting, Filtering, Highlighting**: The filters option help compare Countries Avg BMI over the Year (groups) and identify variations. For sorting purposes, the "Top 10" sorting is applied to average BMI across countries. Additionally, a prefix sequential number (e.g., 1., 2.,) as a label is incorporated in group names. This adjustment addresses the sorting challenge in Tableau, which currently employs ascending character sorting for group labels.

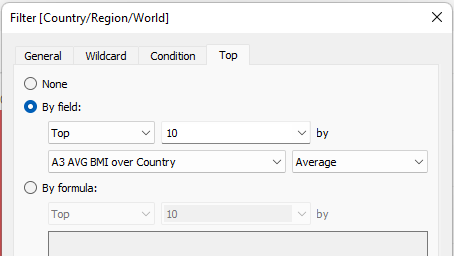
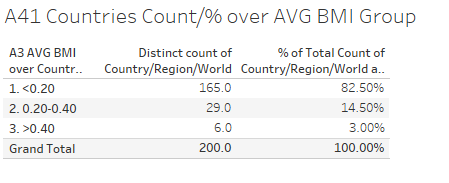
 

Figure VI Sorting, Filtering, highlighting (Finding 1)

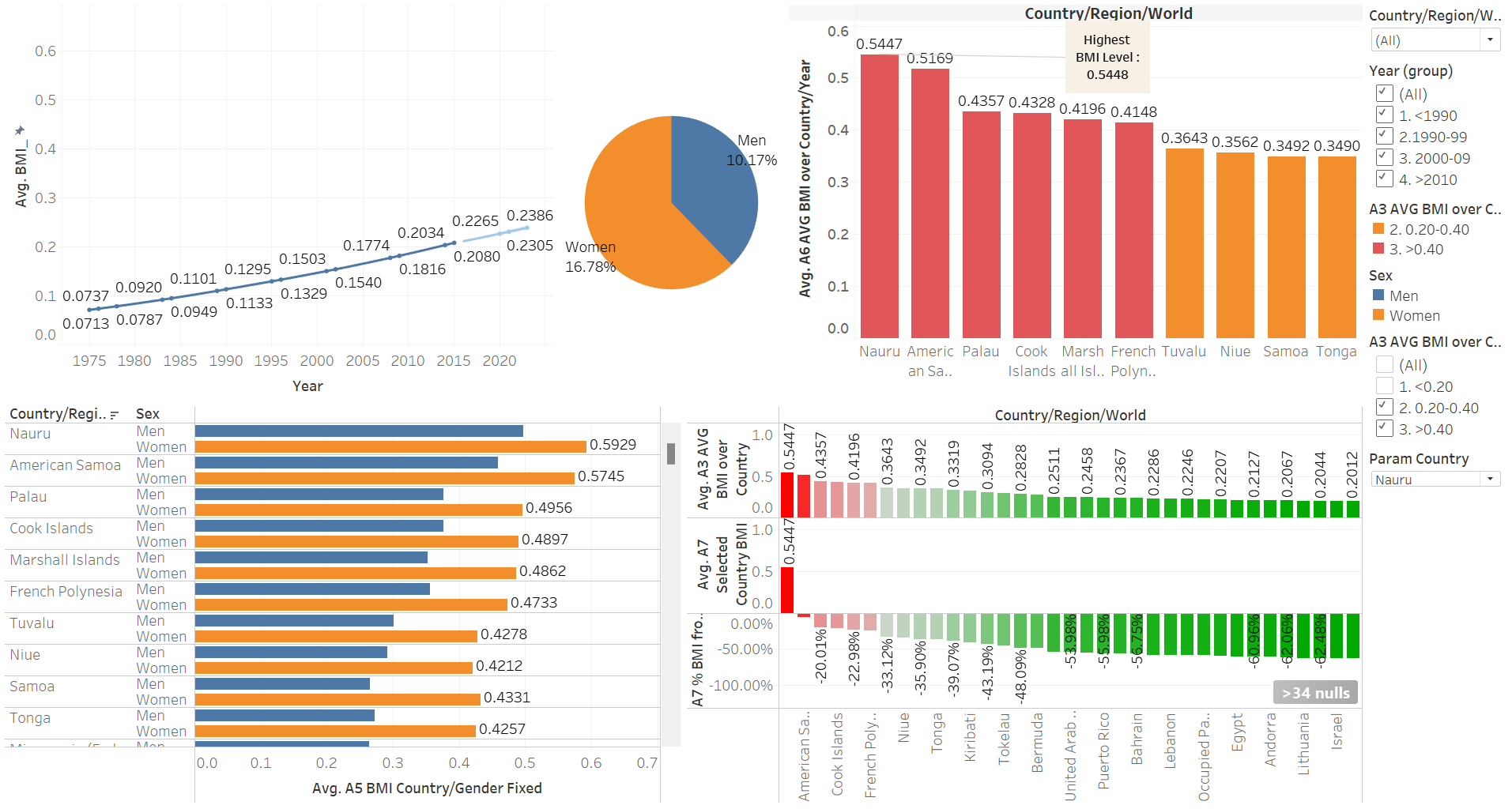


Figure 7 Finding 1 - Dashboard

**Dashboard visual1 illustrates:**

1. Increase in average BMI levels worldwide in the past 4 decades - hypothesis 1
2. Women exhibit higher average BMI than men - hypothesis 3
3. The top 10 average BMI countries are all in the South Pacific region - hypothesis 1,4
4. Nauru recorded the highest BMI average of 0.54 - hypothesis 1,4

## Finding 2: (Worksheets Colored Blue - Sheet name & calculated measures prefixed with B )

In the **Eastern Asia region,** **Vietnam** shows the lowest Diabetes and BMI Index worldwide across the last 4 decades. In **African** **regions, the lowest** prevalence of average **diabetes and BMI** index countries were perceived.

### Quality of Finding2:

The below visual illustrates that **Eastern Asia** has the lowest z-score of Diabetes + BMI index with an average of **-1.23 compared to the average** worldwide from 1975 to 2014 across 200 countries.

It was also observed that most of the top 10 countries, with the lowest prevalence of average Diabetes + BMI, were among African countries.

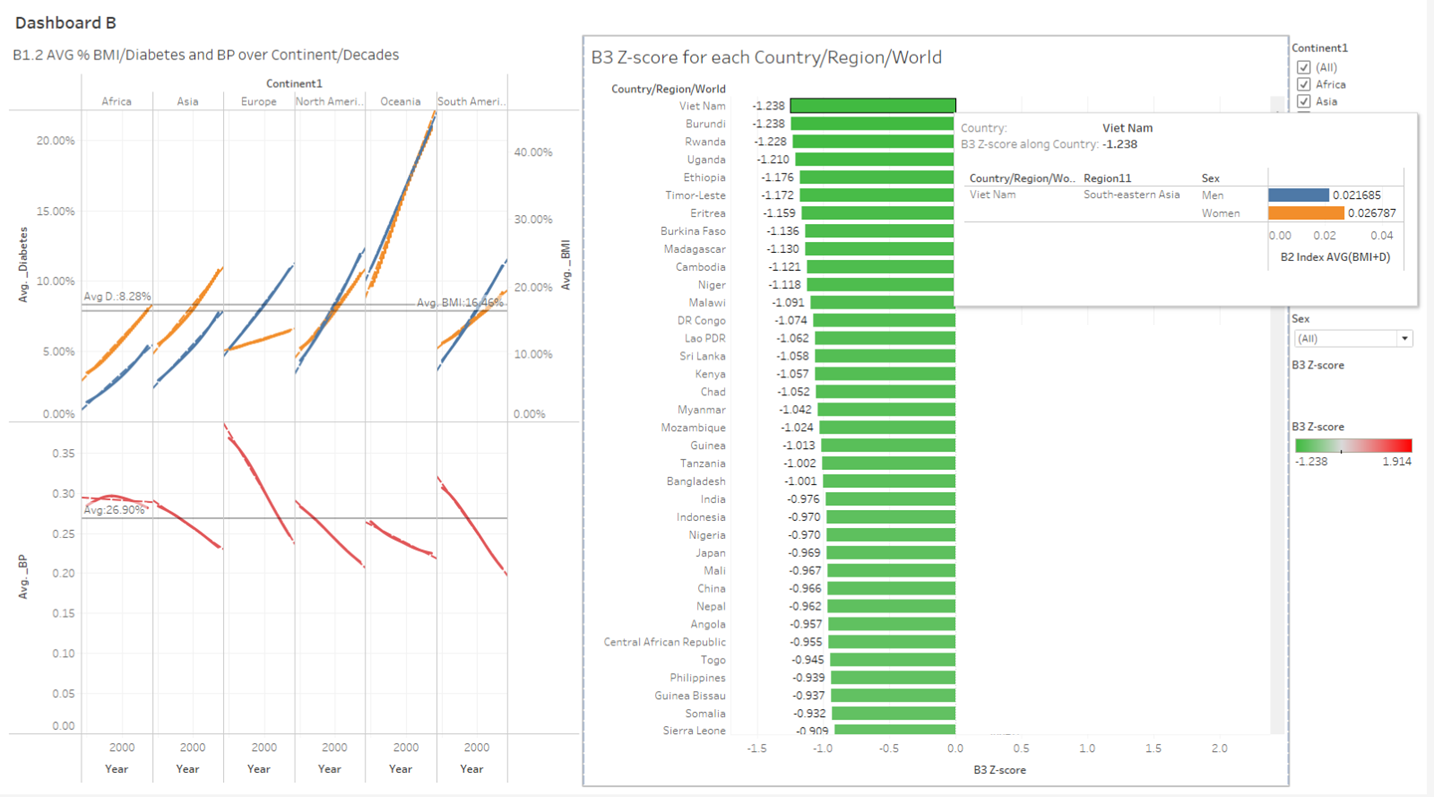


Figure 8 Finding 2 - Dashboard

**Factor/s**: Average (Diabetes+BMI), Average Blood Pressure

Blood pressure correlation with the Diabetes + BMI index varies significantly among continents. While in Africa BP levels are almost constant across the past 4 decades, there is a strongly inverse correlation observed in Europe, Asia, Oceania, North America, and South America.

One of the main factors why a lot of African countries have low Diabetes+BMI levels is due to their **higher prevalence of other diseases and lower life expectancy**. According to Mario’s article, around 60% of the patients in Africa die while in treatment (Mario Azevedo, Sridevi Alla, 2008). Additionally, foot issues are a significant public health concern in Africa, where they are the primary cause of extended hospital stays for those with diabetes and are linked to high rates of death.

Over fifty percent of diabetics are thought to pass quite quickly across the majority of Africa when their condition manifests outside of big cities, suggesting a life expectancy akin to that of pre-insulin Europe or North America. (FM Mwendwa, CF Otieno, JK Kayima, 2006)

For both **men and women**, the average BMI level and the prevalence of diabetes in 2014 exhibited a substantial positive correlation across all nations (correlation values of 0.86 and 0.88, respectively). Compared to the connections found in 1980, which were 0.88 and 0.91, respectively, this is marginally weaker. For women, there was a moderately strong correlation (r = 0.76), but for men, the correlation was weaker (r = 0.34), between the absolute change in BMI and the relative change in diabetes. (Prof. Andre Pascal Kengne, 2017)

### What

* **Dataset Type :**

The NCD\_RisC dataset type is a **Table/Tabular** of **items/records row-wise** in 3 tables/sheets of BMI Body Mass index, Diabetes, and Blood Pressure.

**Finding 2** analysis will use t**wo data sources,** by joining 3 tables using **inner join** resulting in **14000** records.

The dataset is a comprehensive health dataset including **attributes** on the prevalence of specific health conditions, organized and combined based on several key dimensions, including Continent, Region, Country, Sex, and Year.

1. **Health Prevalence Data:** This data set primarily contains information on the prevalence of specific health conditions in various populations. The specific health conditions of interest are:

* **B3 BMI+D Index**: a newly created measure **summing** the average of both the Prevalence of BMI > 30 kg/m² **and** the Prevalence of Diabetes.
* **\_BP** : post the join, a new variable is used in tableau containing the inner joined values of Prevalence of Blood Pressure within the 3 tables (the redundant column is hidden).
* **\_Diabetes**: post the join, a new variable used in tableau containing the inner joined values of Prevalence of Diabetes within the 3 tables (the redundant column is hidden).
* **\_BMI:** post the join, a new variable used in tableau containing the inner joined values of Prevalence of BMI > 30 kg/m² within the 3 tables (the redundant column is hidden).

1. **Dimensions for Aggregation:** The dataset is organized and combined based on three key dimensions:

* Sex: The data is segmented by gender, allowing for the analysis of health conditions in males and females separately. (sub chart in the tooltip)
* Year: The data is collected over different years (1975-2016), enabling the analysis of trends and changes in health conditions over time.
* Country: The data is aggregated by country, making it possible to compare health conditions among different nations.
* Continent and/or Region: an additional dimension from the internet to further group on region and continent level. (United Nations, 2023)
* **Data Type:**
* The data types used are categorical and numerical datasets.
* Source of dataset: joined\_data , ContinentRegions

|  |  |  |
| --- | --- | --- |
| **Variables/Fields** | **Data Types** | **Role in Tableau** |
| Country/Region/World | Geographic Value | Discrete Dimension |
| Sex | String Value | Discrete Dimension |
| Year | Number (Whole) | Continuous Dimensions |
| B3 BMI+D Index | Number (Decimal) | New Calc field : (BMI+D)/2 |
| \_BP | Number (Decimal) | Continuous Measure |
| \_Diabetes | Number (Decimal) | Continuous Measure |
| \_BMI | Number (Decimal) | Continuous Measure |
| Continent/Region | String Value | Discrete Dimension |

* **Attribute Type**
* **Dimensions** or qualitative- Continent, Region, Country, Sex, Year.
* **Measures** or continuous numbers- Average BMI+Average Diabetes Index, Average BP
* **Calculated Measures :**

**Used in Main Finding Visual**

* + - B2 Index AVG(BMI+D) = (AVG([\_BMI])+AVG([\_Diabetes]))/2
    - B3 AVG BMI+D Index = WINDOW\_AVG(SUM([B3 BMI+D Index]))
    - B3 BMI+D Index = ([\_BMI]+[\_Diabetes])/2
    - B3 STDEVP BMI+D Index = WINDOW\_STDEVP(SUM([B3 BMI+D Index]))
    - B3 Z-score = (SUM([B3 BMI+D Index]) - [B3 AVG BMI+D Index]) / [B3 STDEVP BMI+D Index]
    - B5 WIN AVG \_BMI = WINDOW\_AVG(SUM([\_BMI]))
    - B5 WIN AVG \_Diabetes = WINDOW\_AVG(SUM([\_Diabetes]))

**Used in Extra Finding Visuals/Story**

* + - B1 WCORR BMI-Diabetes = WINDOW\_CORR(SUM([\_BMI]), SUM([\_Diabetes]))
    - B2 BMI+D % Diff from Mean = { FIXED [Country/Region/World]: [B2 Index AVG(BMI+D)]/AVG([B2 Mean of BMI+D Index])-1 }
    - B2 Mean of BMI+D Index = { FIXED : SUM([\_BMI]+[\_Diabetes])/COUNT([BMI])}

### Why

* **What does the visualization aim to show**

The objective of the dashboard visualization is to depict relationships among diabetes, BMI, and blood pressure, examining observations across countries, regions, and continents. The focus is on identifying the lowest recorded levels of diabetes and BMI and understanding their relationship with blood pressure levels. The goal is to derive actionable recommendations based on the conclusions drawn from these observations.

* **Describe the Actions from all three aspects: ‘Analyze’, ‘Search’, and ‘Query’**

**Analyze**:

The Analysis reflects per continent the nature of the relation among the three indices. In Europe, the upward trend in average BMI was significantly steeper than the more gradual ascent observed in the average diabetes. In addition to that, Europe displayed a noticeable decline of 31.27% in average BP across all decades. On the other hand, the **African** continent recorded a **nearly constant BP index** across 4 decades of **27.14%.** **Oceania** exhibited notably **steep inclines** for **both average BMI and Diabetes**, standing out prominently when contrasted with the more **moderate slopes** characterizing the remaining graphs in the set as shown in Figure 7. To sum up, the overall trend shown was that average BMI and Diabetes are positively correlated while average BP was negatively correlated - inversely proportional to them.

In addition to disparities between male and female groups per country using bar chart visualization. The Percentage of BMI is Produced by deriving the average percentage of BMI compared to the world average.

• Comparing **1980** and **2016** indexes showcases that the Average increase of BMI (66% ) & Diabetes (50%) is higher for **Men** compared to **women's** BMI (53% ) & Diabetes (45%)

**Actions 1**: Obesity has been grouped into three bands based on the average BMI Index: less than 0.20, between 0.20 and 0.40, and greater than 0.40.

**Actions 2**: Year dimension is the source for a new field called [Year (group)] reflecting decade groups. It functions as a discrete dimension facilitating comparison, years spanning from 1990 to 2010 are grouped per decade except for the initial 15 years and above 2010 group covering the last 6 years.

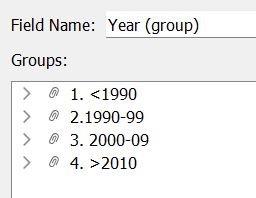


Figure IX Actions 2

**Z-Score technical elaboration:**

For the "B3 Z-score for each Country/Region/World" and "B3 Z-score for each Region," we employ the following calculated fields, which are integral to the Z-Score calculation process:

B3 BMI+D Index = ([\_BMI] +[\_Diabetes])/2

B3 AVG BMI+D Index = WINDOW\_AVG(SUM([B3 BMI+D Index]))

B3 STDEVP BMI+D Index = WINDOW\_STDEVP(SUM([B3 BMI+D Index]))

B3 Z-score = (SUM([B3 BMI+D Index]) - [B3 AVG BMI+D Index]) / [B3 STDEVP BMI+D Index]

The z-score, also known as the standard score, quantifies the number of standard deviations an observation is above or below the population mean. As a guiding principle, z-scores falling below -1.96 or exceeding 1.96 are deemed noteworthy and unusual. In statistical terms, such values are considered statistically significant outliers.

Z-score = (x-m)/d; x=sum(BMI+D Index); m=Mean; d=Standard Deviation

Dimensions:

Region1 has 21 members on this sheet.

Members: Australia and New Zealand; Eastern Asia; Middle Africa; Southern Africa; Southern Asia; ...

The B3 Z-score filter ranges from –1.96 to 1.96.

Measures:

B3 Z-score ranges from -1.201 to 1.132 on this sheet.

The filter associated with this field ranges from -1.96 to 1.96.

The formula is as follows:

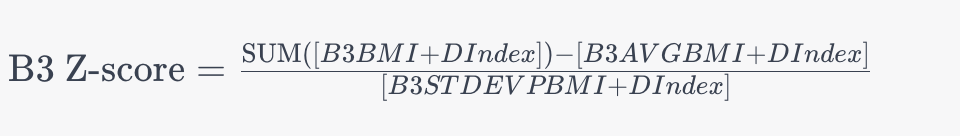


Figure X B3 Z-score Formula

**Search-**based: The averages for BMI percentage (Avg\_BMI%), diabetes percentage (AVG\_Diabetes%), and blood pressure (Avg\_BP) are calculated across Year (group) and Continent.

Second Visual: **Z-Score** (STDev (BMI+D) Index) over countries where the percentage is compared difference % to other countries.

**Query:**

In the first visual: Compare the percentage rise and fall among different continents over the Year (groups).

In the second visual: Assess the Z-Score for the Standard Deviation (STDev) of the (BMI+D) Index over regions for the entire period.

Filters can further uncover results and trends on various levels and combinations of Continents, Regions, Countries, Sex, Years, Year (group), or Avg BMI (group).

• The user can identify the results at both the continent level and among individual regional states.

•Compare the selection results of one Continent to another, or select multiple Continents

•Summarize the selection results across all Continents or Regions.

* **Describe the ‘Targetsʼ from both the ‘Dataʼ and ‘Attributeʼ aspect**

**Targets:** First visual:The Target is to show the correlation between 3 measures. Second Visual: to show the distribution of Z-Score between Regions.

The attributes used are **continuous measures that** can be visually understood through both line and bar chart visualization.

* **Features:** AVG \_BMI, AVG \_Diabetes, and AVG BP dispersion over different groups. Also, for Z-Score Second Visual AVG(BMI+D) Index which also creates dispersion over different groups such as Continent. Region, Gender.
* **Attributes:** Finding 2 compares 3 measures, hence Target attribute such as Correlation is used where values of two attributes are tied together. in the first visual “**B1.3 PC R using WINDOW\_CORR**” the Pearson Correlation method is presented over (**\_BMI and \_Diabetes measure**) using the calculated field:

**B1.3 WCORR BMI-Diabetes = WINDOW\_CORR(SUM([\_BMI]), SUM([\_Diabetes]))**

upon observing the results, a new native measure is calculated to ease data analysis:

**B3 BMI+D Index = ([\_BMI]+[\_Diabetes])/2**

**Presenting results:**  
In the Trend Lines Model (BMI vs. Diabetes), a linear trend model is calculated for \_Diabetes given \_BMI. The model's significance may be observed at p <= 0.05. The R-squared value is 0.635931, indicating that approximately 63.6% of the variation in \_Diabetes is explained by \_BMI (R-squared: 0.635931).

The square root of the correlation coefficient (r) is approximately 0.797452, which is roughly 0.80 or 80%. This represents how well the data fits the linear model, signifying that 80% of the variation in \_Diabetes can be explained by \_BMI.

The equation of the model is Diabetes = 0.289917\*BMI + 0.0348971.

The model formula is presented as (\_BMI + intercept ), and the number of modeled observations is 14000.

**Correlation:**

The trends of WCORR BMI-Diabetes are analyzed along the table (downward) for the count of Sheet1, BMI, and Diabetes. A correlation coefficient, denoted as *r*, is a singular numerical representation of the degree of relationship between two measures. The correlation coefficient adheres to the range -1 <= *r* <= 1.

A positive correlation implies a relationship between the x and y measures, where an increase in x corresponds to an increase in y. In this case, the correlation coefficient is +.70 or higher indicating a very strong positive relationship.

### How

* **Visual mapping or encoding -Marks:** 1 Dimensional diagram, and color
* **Channels:** 
  + **First Visual: Line Chart** mapped as theAVG \_BMI, AVG \_Diabetes, and AVG BPAttribute and **Color hue** as Categorical Attribute to identify the measuresgroup levels Attribute. The line chart is segregated into dual y-Axes and x-Axes for Year (group) decade.
  + **Second Visual: Bar chart,** distribution of Z-Score **(x**-Axes)over Region (y-Axes)
* **Features:**
  + **Filters** used on the Continent (or multiple choice), Region, Gender, and optional if we want to analyze only one of the measures – option for Measure filtering. Using Filters we can make a combination between Continent, Region, and Gender to filter them accordingly.
  + **StoryBoard B:** presenting significant visuals elaborating Finding B (Appendix).
* **Why this Visual**: This design helps in identifying the percentage increase of AVG BMI+D Index among Continents, Regions, and Countries over all 4 decades, and in the opposite way we can see AVG BP is going down and decreasing over the same attributes. Also, this visualization has an option of using ‘country mouse hover’ to show gender disparities on the Region level.

A line Chart is best for showing continuous attributes, and a Bar Chart is usually the best way of representing visuals for categorical attributes, good visual perception, and easy user understanding …

* **Sorting, Filtering, Highlighting**: Filters as an optional part, help us to compare main attributes. For example, on Z-Score visualization over Region, Soring is Region ascending over Z-Score. On each visualization we are using a filtering option to compare results, filtering data depending on Continents and Regions, Years (group), or using additional group fields such as Sex (Gender) to determine how results are reflected on some cumulative group level.

Countries’ Avg BMI is analyzed over Year (groups) to deduce differences between them. For sorting purposes, a "Top 10" sort is applied to the average BMI Index over countries. Additionally, a prefix sequential number (1., 2., 3., etc.) is used as a label to address Tableau's sorting challenge with group labels, as it defaults to ascending character sorting.

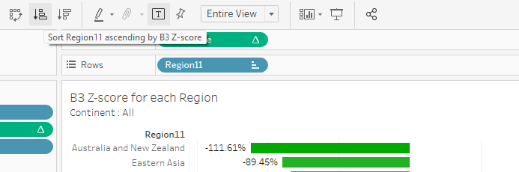
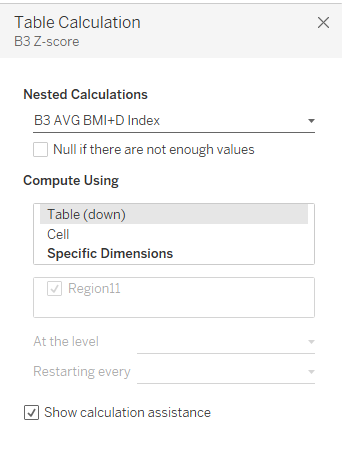
 

Figure XI Sorting, Filtering, Highlighting (Finding 2)

# Future Enhancements:

Further research and insights are to be conducted to uncover , confirm or negate assumptions around the health factors and metrics. hence propose action items to address risks and issues identified in relation to them especially around environmental issues, nuclear experiments and the impact on low income countries.

# Research, Conclusion and Recommendation:

Analyzing the **NCD-RisC** dataset uncovered a diversity of valuable insights around the complex interplay of Body Mass Index (BMI), diabetes, and blood pressure in relation to time and geography, and gender.

Several key conclusions and facts are further clarified using the insights presented, below are some of the key points.

* A clear correlation between BMI and the prevalence of diabetes was established. Accordingly, individuals with higher BMI values are at a higher risk of diabetes, weight management become a priority for diabetes prevention.
* Gender disparities in health outcomes were evident, hence understanding these disparities is crucial for tailoring health interventions to targeted population groups.
* **Nauru** has had to deal with the severe effects of extensive phosphate mining. Phosphate mining was the nation's primary source of income in the past, but it also severely degraded the ecosystem, causing soil erosion, and agriculture severely harmed, which increased dependency on imported, processed commodities. As a result, an increase in BMI levels has been noted. Nauru's socioeconomic environment, impacting access to basic services hence a multifaceted strategy seeking a balance between public welfare, environmental preservation, and commercial interests is being used to solve these issues.
* In the **Pacific area**, in the Marshall Islands, French Polynesia, and some nearby countries, **318 nuclear bombs were launched** by the US, UK, and France between 1946 and 1996. Fallout was found in the Cook Islands, Samoa, Tonga, and Tuvalu, and it was distributed throughout the region. Apart from nuclear explosions, the United States experimented with chemical weapons, explosives, and ballistic missiles in the Marshall Islands, while the United Kingdom carried out over 600 hazardous "minor trials" in Australia. A terrible legacy of illness, anxiety, radiation, and ecological destruction has been left behind by this. The thousands of military and civilian participants in the experiments from the USA, France, UK, Fiji, Australia, and New Zealand who were exposed to radiation also have a social influence beyond the communities most directly impacted. (INTERNATIONAL DISARMAMENT INSTITUTE NEWS, n.d.) (Atomic Heritage Foundation, n.d.)
* East African countries like Burundi and Uganda, demonstrated a low prevalence of Diabetes and BMI according to our data. This could be because the life expectancy is low meaning patients are not being cured, and primitive ways of diagnosis due to lack of access to health care systems.
* On Country level, Vietnam has the lowest BMI + Diabetes Index worldwide.
* [Vietnam has the least obesity rate due to factors such as](https://www.bing.com/ck/a?!&&p=f9bd7005bc2ae99fJmltdHM9MTcwMDAwNjQwMCZpZ3VpZD0wMTdhZTNjMC1kM2U1LTYzNTAtMzc2Ny1mMGQxZDJmNzYyNDQmaW5zaWQ9NTYxNQ&ptn=3&ver=2&hsh=3&fclid=017ae3c0-d3e5-6350-3767-f0d1d2f76244&psq=why+does+veitnam+has+the+lowest+rate+of+obesity+in+the+world&u=a1aHR0cHM6Ly9lbW9qaWN1dC5jb20va25vd2xlZGdlYmFzZS93aHktaXMtdmlldG5hbS1ub3Qtb2Jlc2U&ntb=1) healthy dietary focused on organic food, also [the low ranking could also be due to economic factors such as long working hours with workers having no time for exercise](https://www.bing.com/ck/a?!&&p=6e053a9a11d0900bJmltdHM9MTcwMDAwNjQwMCZpZ3VpZD0wMTdhZTNjMC1kM2U1LTYzNTAtMzc2Ny1mMGQxZDJmNzYyNDQmaW5zaWQ9NTYyMg&ptn=3&ver=2&hsh=3&fclid=017ae3c0-d3e5-6350-3767-f0d1d2f76244&psq=why+does+veitnam+has+the+lowest+rate+of+obesity+in+the+world&u=a1aHR0cHM6Ly9lbW9qaWN1dC5jb20va25vd2xlZGdlYmFzZS93aHktaXMtdmlldG5hbS1ub3Qtb2Jlc2U&ntb=1)
* On Continent level, Lowest Z-score was observed in East Asia and Africa
* For women, there was a positive moderately strong correlation, but for men, the positive correlation was weaker, between the absolute change in BMI and the relative change in diabetes.

**Recommendations to Governments and policymakers** **are to Build measures that:**

* Shifts dietary habits away from consumption of calorie-dense and processed foods, often high in sugars, unhealthy fats, and refined carbohydrates.
* Transition the population to a physically active lifestyle.
* Prove low-income populations access to affordable, nutritious foods.
* Fight and educate against aggressive marketing of high-calorie, low-nutrient foods and sugary beverages, especially targeted towards children and adolescents.
* Enforce the regulatory environment, including policies related to food labeling, advertising, and access to healthier food options.
* Improve access to healthcare services, including obesity management programs.
* Improving education about healthy nutrition and lifestyle choices can contribute to unhealthy behaviors.
* Enforce control over nuclear experiments.
* Nuclear experiments should have more control measures to avoid casualties and life-threatening consequences over time.

References

United Nations. (2023). *Standard country or area codes for statistical use (M49)*. Retrieved from https://unstats.un.org/unsd/methodology/m49/

*World Health Organization*. (n.d.). Retrieved November 13, 2023, from https://www.who.int/health-topics/obesity#tab=tab\_1

Prof. Andre Pascal Kengne. (2017). Trends in obesity and diabetes across Africa from 1980 to 2014: an analysis of pooled population-based studies. *International Journal of Epidemiology, 46*(5), 1421–1432.

Deepika Singh. (2020, July 16). *Discern the Three Types of Level of Detail (LOD) Calculations in Tableau*. Retrieved from https://www.pluralsight.com/guides/discern-the-three-different-types-of-level-of-detail-(level-of-details)-calculations-in-tableau

Mario Azevedo, Sridevi Alla. (2008). Diabetes in Sub-Saharan Africa: Kenya, Mali, Mozambique, Nigeria, South Africa and Zambia. *International Journal of Diabetes in Developing Countries, 28*(4), 101–108.

FM Mwendwa, CF Otieno, JK Kayima. (2006). Risk factor profile and the occurrence of microvascular complications in short term type 2 diabetes mellitus at Kenyatta National Hospital, Nairobi. *East African Medical Journal, 82*(12), S163-S172.

*INTERNATIONAL DISARMAMENT INSTITUTE NEWS*. (n.d.). Retrieved from Education and Research on Global Disarmament Policy: https://disarmament.blogs.pace.edu/nuclear-test-archive/#:~:text=From%201946%20to%201996%2C%20the,Atoll%20and%20Amchitka%20Island%2C%20Alaska

*Atomic Heritage Foundation*. (n.d.). Retrieved from National Museum of Nuclear Science & History: https://ahf.nuclearmuseum.org/ahf/location/marshall-islands/#:~:text=Between%201946%20and%201958%2C%20the,spread%20throughout%20the%20Marshall%20Islands.

**Ω**

# Appendix:

## Data Cleansing

* + ContinentRegions dataset cw1 -dataset
  + Bolivia\_ Bolivia
  + China, Hong Kong Special Administrative Region China (Hong Kong SAR)
  + Czech Republic Czechia
  + Congo (Democratic Republic of Congo) DR Congo
  + Guinea-Bissau Guinea Bissau
  + Iran (Islamic Republic of) Iran
  + Lao People's Democratic Republic Lao PDR
  + North Macedonia Macedonia (TFYR)
  + Republic of Moldova Moldova
  + Democratic People's Republic of Korea North Korea
  + State of Palestine Occupied Palestinian Territory
  + Republic of Korea South Korea
  + United Republic of Tanzania Tanzania
  + United Kingdom of Great Britain and Northern Ireland United Kingdom
  + Venezuela (Bolivarian Republic of) Venezuela
  + New records what missing:
  + Swaziland Southern Africa Africa
  + Taiwan Eastern Asia Asia
  + Côte d'Ivoire Western Africa Africa

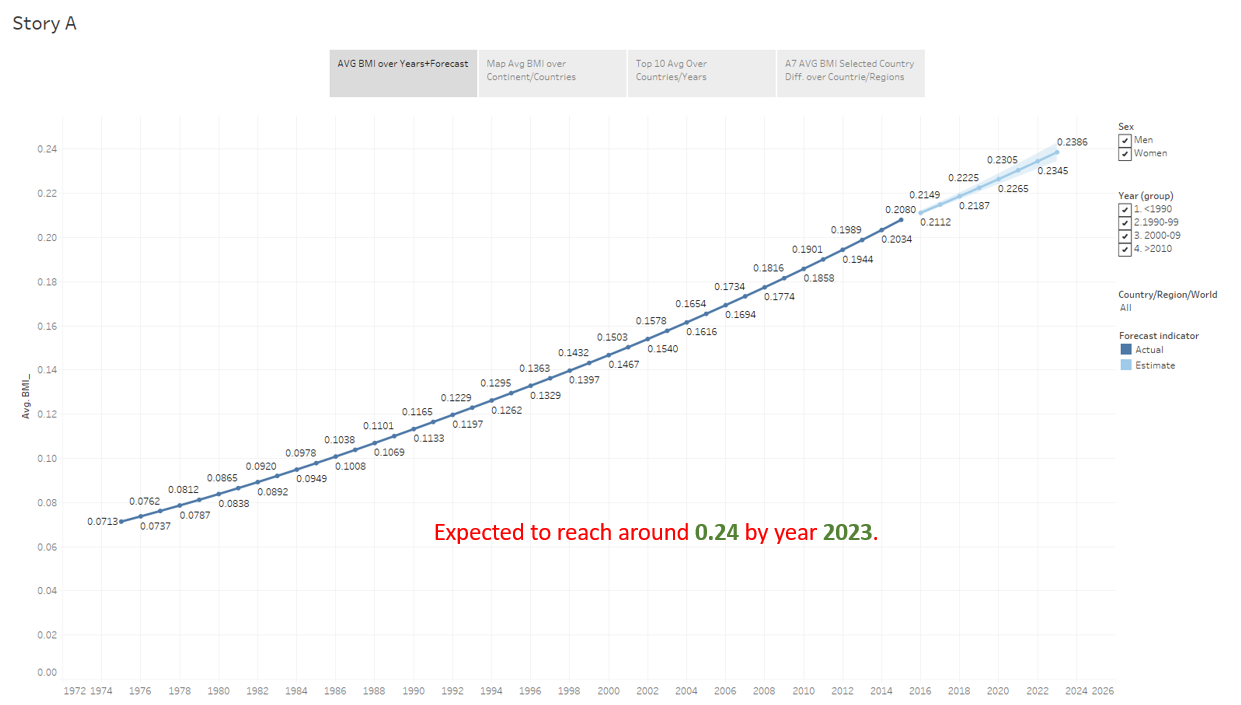
## Statistical Definitions

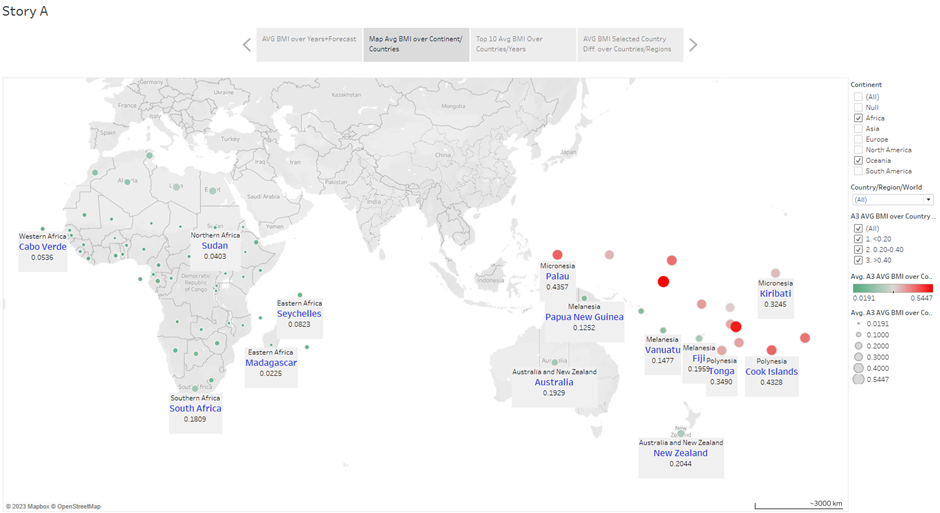
**LOD calculations explanation:**

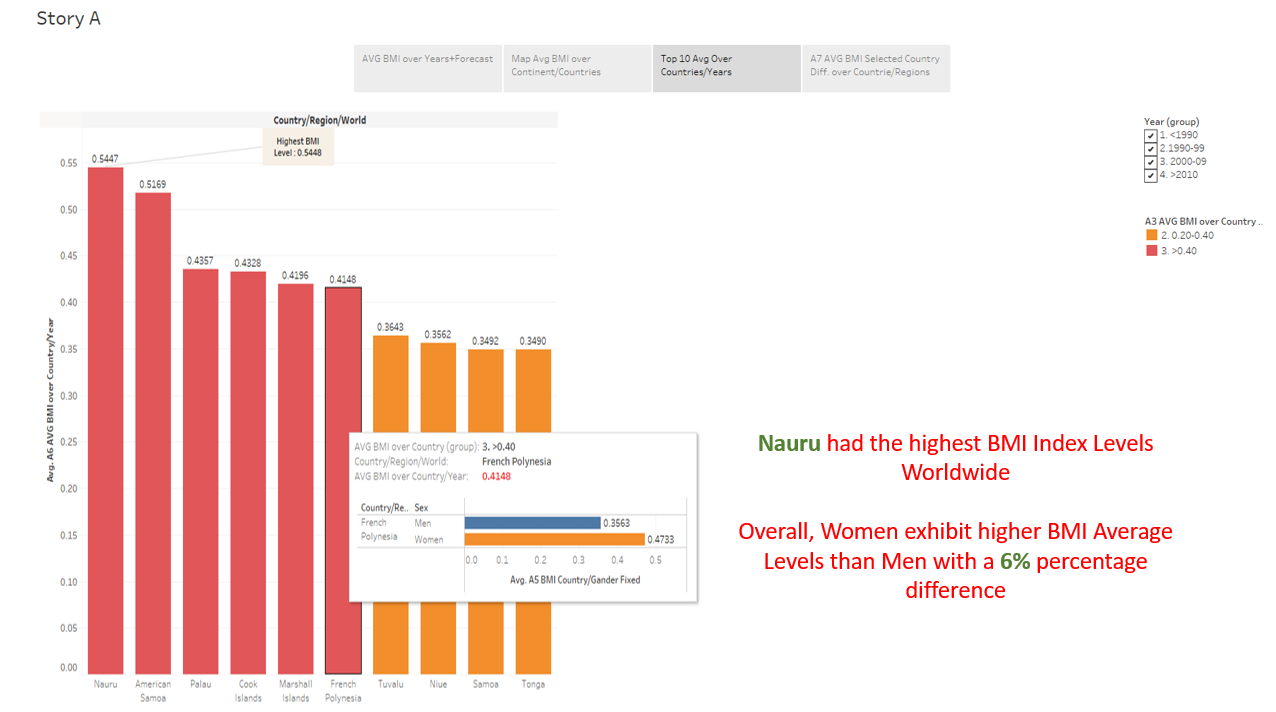
* + *“FIXED : Fixed Level of Detail : A fixed LOD expression is used to compute a value using specified dimensions without reference to the dimensions in the view.*
  + *INCLUDE : Include Level of Detail:An include LOD expression is used to compute values using specified dimensions in addition to the dimensions available in the view*
  + *EXCLUDE: Exclude Level of Detail:An Exclude LOD expression is used to remove unwanted information from the data by creating an EXCLUDE function to omit the dimension at which aggregation is not required.”[[2]](#footnote-3)*

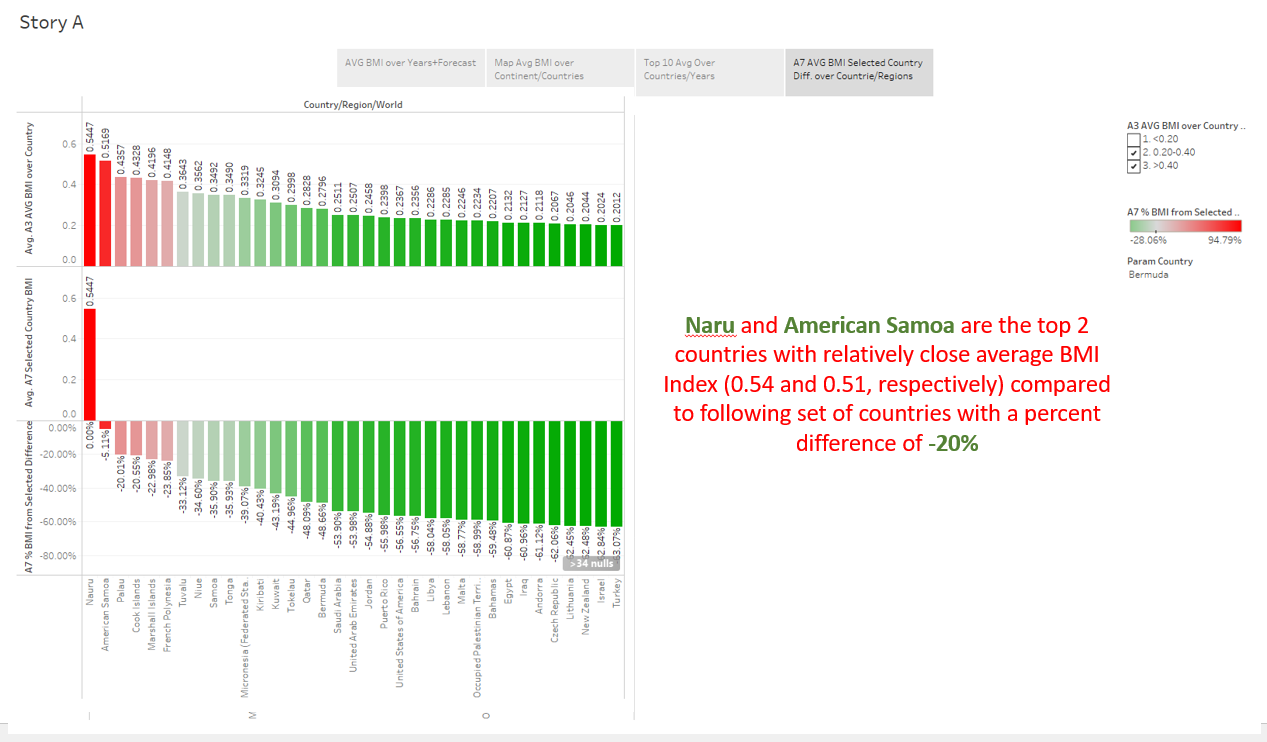
## Extra/Relevant Significant Insights Driving the Report findings:

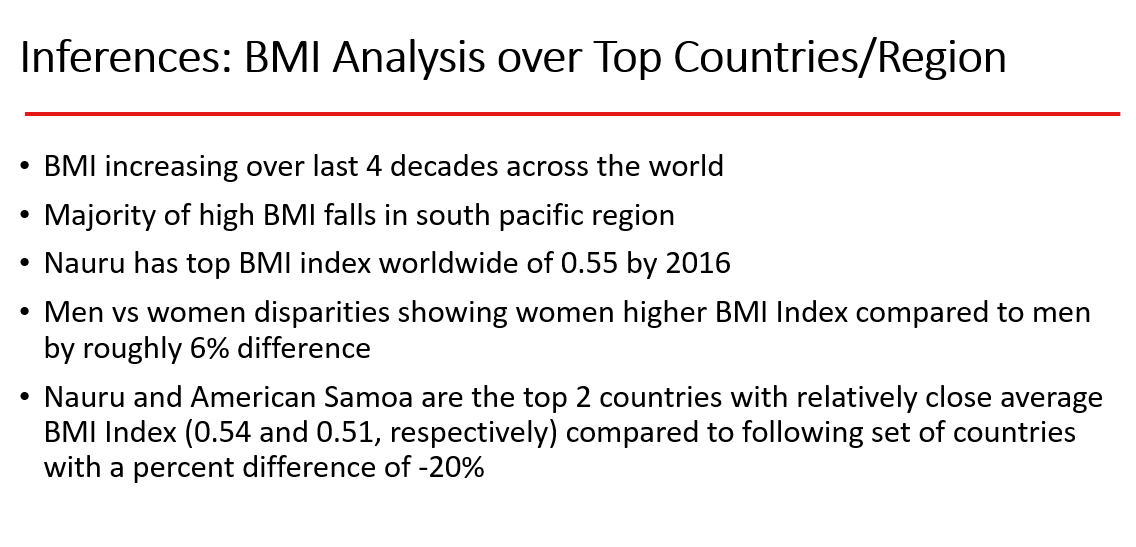
### Storyboard A



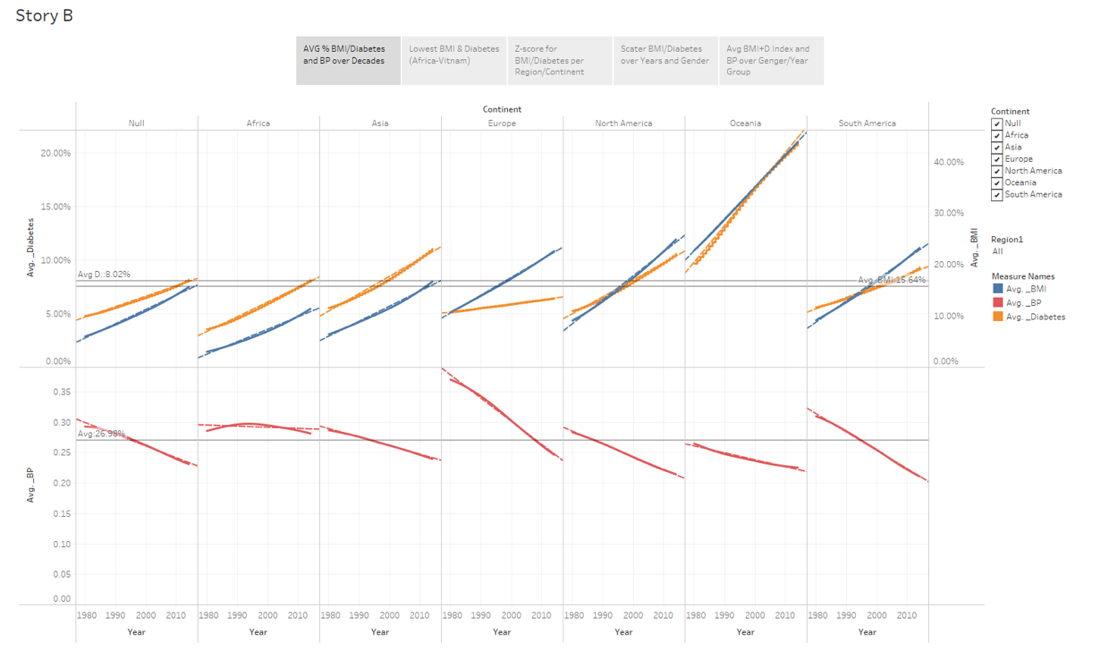


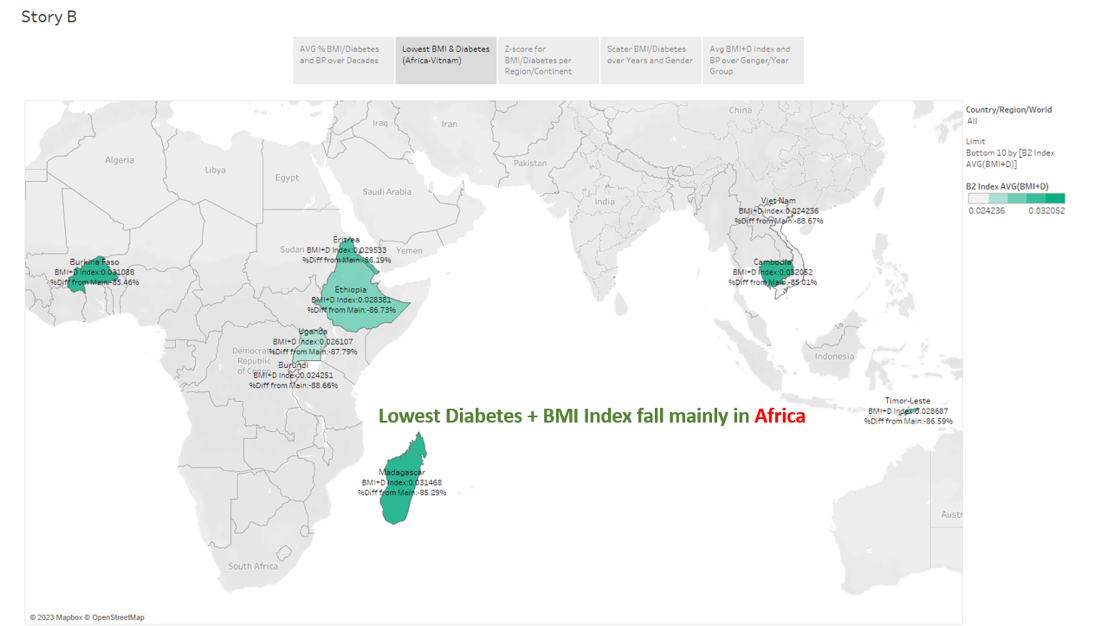




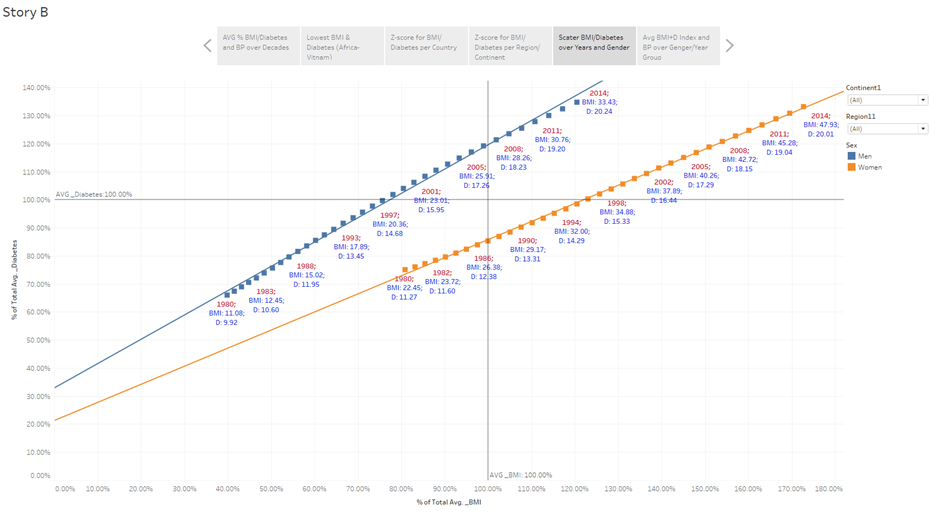


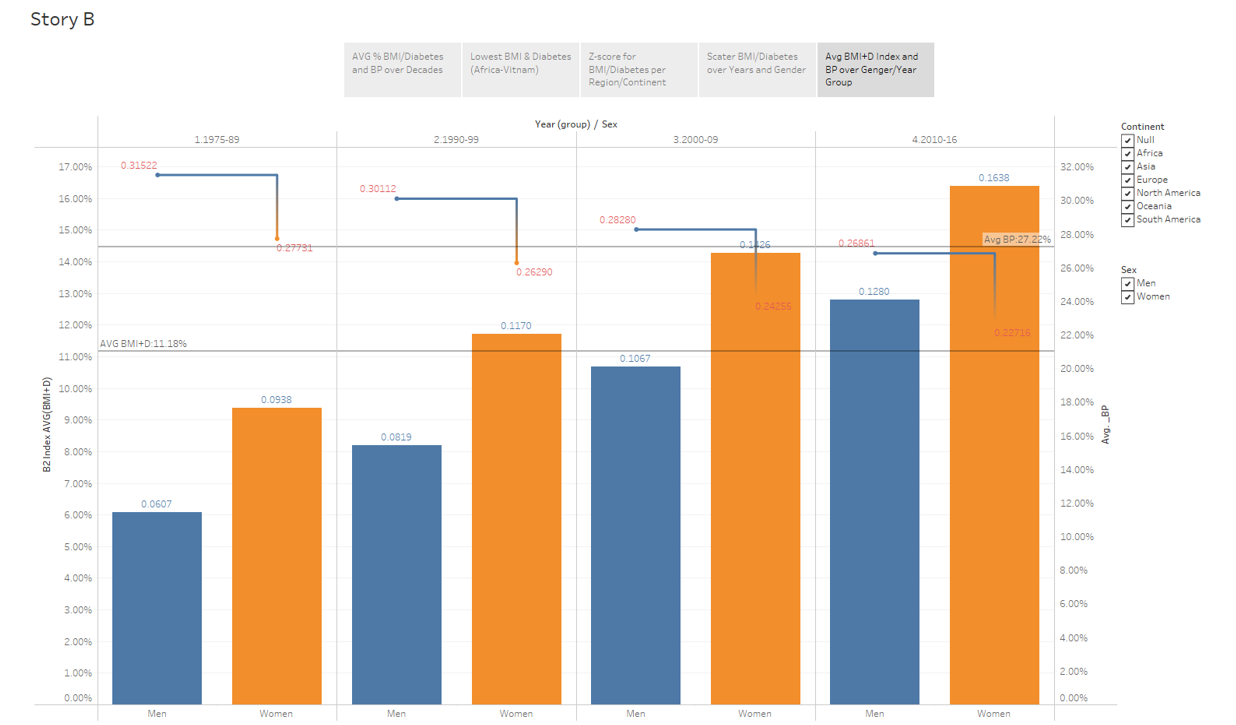
### Storyboard B

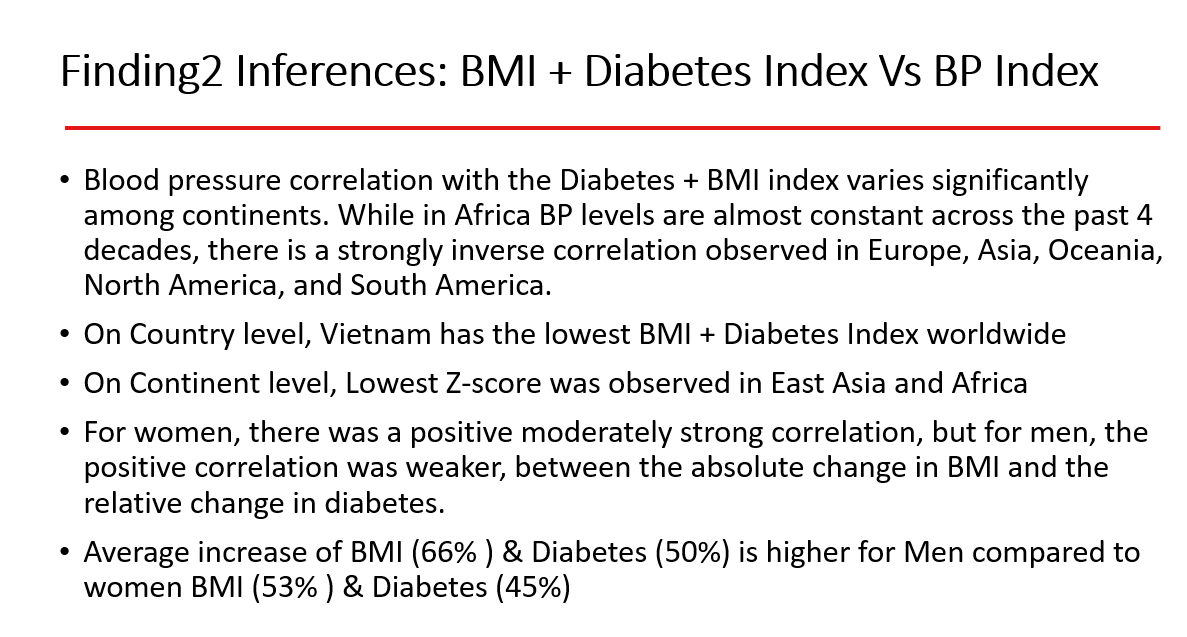












1. TutorialsPiont, Tableau - Trend Lines: https://www.tutorialspoint.com/tableau/tableau\_trend\_lines.htm#:~:text=Trend%20lines%20are%20used%20to,in%20both%20of%20them%20simultaneously. [↑](#footnote-ref-2)
2. Discern the Three Types of Level of Detail (LOD) Calculations in Tableau

   https://www.pluralsight.com/guides/discern-the-three-different-types-of-level-of-detail-(level-of-details)-calculations-in-tableau [↑](#footnote-ref-3)