

**Project: Studying Factors Affecting Obesity in Jordan**

**Technical Report**

**Course Name**: Data visualization  
 **Course Instructor:** Dr. Ameera Al-Amery

**Prepared by:**

Saja Salameh 20210562

Dalia Tamimi 20200288

Yumna Al-Natsheh 20221019

**Detailed Description of Each Member Work**

### **Yumna Al-Natsheh (2022109)**

### 1) Explored and cleaned Kessler Psychological Distress Scale (K10+) and did the code for it.

### 2) Wrote the technical report.

**Saja Salameh (20210562)**

### 1) Cleaned and explored the Pittsburgh Sleep Quality Index (PSQI) and did the code for it.

### 2) Visualized the data in R.

### **Dalia Tamimi (20200288)**

### 1) Explored and cleaned the Resilience Scale (RS) and Demographic and Lifestyle Factors and did the code for it.

### 2) Wrote the analytical report

**It is worth mentioning that most processes worked on in the project were done collaboratively. The whole group participated in online meetings and worked together on everything.**

**1.Abstract and introduction**

1.1 Abstract

Our study focused on factors related to obesity in Jordan. 900 people from had participated in the survey. Five scales were used to measure various factors affecting obesity in Jordan as follow:

1. The Pittsburgh Sleep Quality Index (PSQI) where the global score ranges (0–4) as “good” sleep and (5–21) as “poor” sleep.
2. Resilience Scale (RS) where scores (6 – 13) is classified as “low resilience” ,(14 – 21) as “normal resilience” and (22 – 30) as ”high resilience”.
3. Kessler Psychological Distress Scale (K10+) where scores (10 – 19) likely to be “well” ,(20 – 29) have a “moderate disorder” and (30-50) have a “severe disorder”.
4. Demographic and Lifestyle Factors such as (gender, age, family history with overweight, smoking status, transportation and city of residence) are used to measure effect of these factors on obesity rates.

By identifying these connections, we aim to develop best strategies to overcome the obesity issue.

1.2 Introduction

Addressing the obesity issue is a key priority for Jordan, and understanding the factors that contribute to this issue is important for developing effective measures. In our study, we collected data from 900 induvial to investigate how various factors influence obesity rates. The survey collected detailed information on various factors (details in the abstract section).

Throughout this report, we present insightful visualizations and analyze the data to reveal trends and patterns. Our goal is to identify key drivers of obesity, understand their interactions, and provide recommendations.

**2 Methodology**

2.1Data Overview

The analysis of the dataset was conducted using Python. The process began with a thorough examination of the metadata to understand the meaning and purpose of each column.

2.2 Data Exploration and Cleaning:

* **Exploration:** Initial exploration involved identifying and addressing missing values and inconsistencies within the dataset.
* **Handling Missing Values:** Missing values were managed by replacing them with appropriate statistics such as the mode or mean, depending on the column type.
* **Data Inconsistencies:** Any inconsistencies in the data were addressed to ensure accuracy and reliability.

2.3 Data Transformation**:**

* **Conversion to Numeric:** Columns requiring calculation were transformed from categorical to numerical formats based on the mappings provided in the metadata.

2.4 Data Aggregation**:**

**Combining and Summarizing:** Multiple columns were aggregated into single columns or summary statistics were created as needed. For instance, columns with related information were summed up for each record. This transformation was guided by standards and metadata provided.

2.5 Column Renaming:

**Improving Readability:** Columns were renamed to enhance readability and facilitate easier analysis.

2.6 Using R Code for Data Visualization and Analysis

**2.6.1 Data Exploration and Preparation**

**Convert Variables to Factors** because the dealing with ordered levels.

2.6.2 Count **and** **Visualize** Each Variable

**Count and Bar Plot for Each Categorical Variable**

**Mosaic Plot Between Two Categorical Variables**

using bar plots to compare the count of different categories within each obesity group.

**2.6.3Discretize *Numerical Variables***

Convert *Age* into categories based on chart available on public domain *(refrence:https://www.researchgate.net/figure/Discretization-functions-for-age\_fig8\_255686833”)* to apply chi-square tests and visualize the distribution.

***2.6.4 Additional Data Transformations***

Transform Mental Distress Level into categorical categories to apply chi-square tests and visualize the distribution

2.6.5 Using chi test and Fisher's Exact Test (for imbalanced data). Based on information in public domain*(https://www.pharmasug.org/proceedings/2014/CC/PharmaSUG-2014-CC19.pdf)*

To test if the two variables are statistically significant

**3 Result**

**3.1 Gender association**

The visual representations did not provide a very clear association between gender as shown in **Appendix 4.1**. However, the Chi-Square test indicated that there is statistically significant association between them(and the male has the highest).

**3.2 Smoking association**

The visual representations did not provide a very clear association between smoking and obesity as shown in **Appendix 4.2** However, the Fisher's Exact Test indicated that there is statistically significant association between them.

**3.3 Age association**

The visual representations provide a very clear association between age and obesity particularly in adults as shown in **Appendix 4.3**. However, the Chi-Square test indicated that there is statistically significant association between them.

**3.4  Mental Distress Level association**

The visual representations provided a very clear association between Mental Distress Level and obesity as shown in **Appendix 4.4**. However, the Chi-Square test indicated that there is not statistically significant association between them.

**3.5 Resilience Level association**

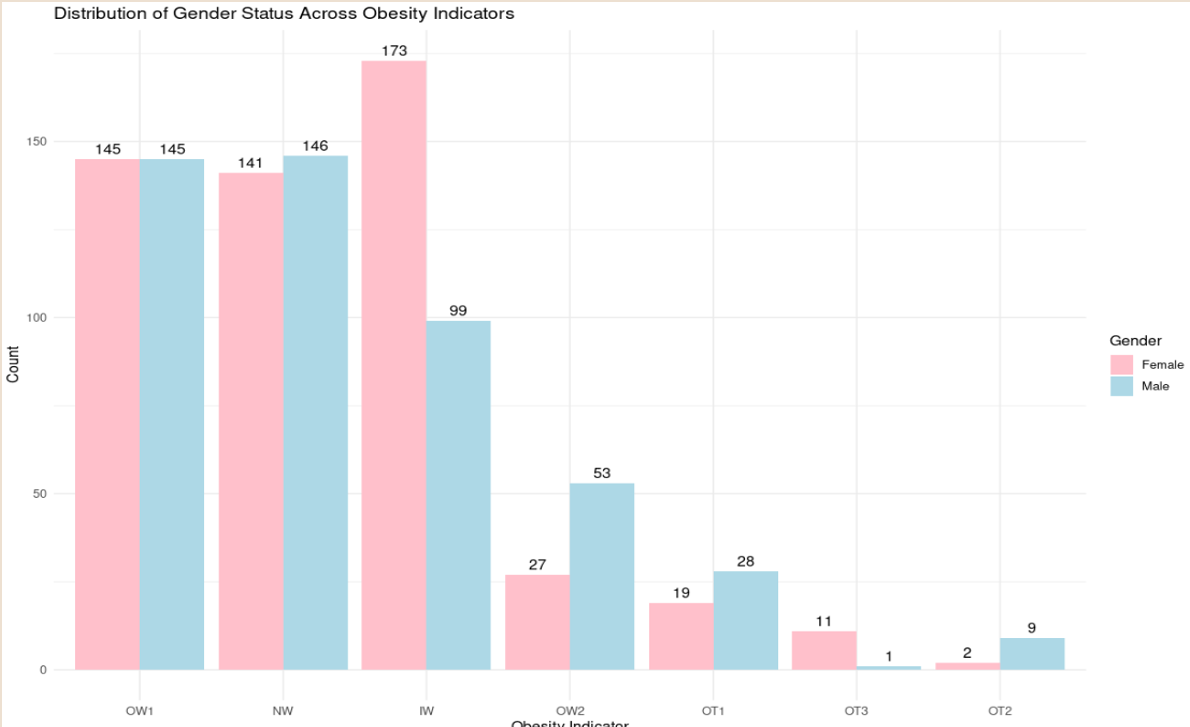
The visual representations did not provide a very clear association betweenResilience Level association and obesity as shown in **Appendix4.5.** However, the Chi-Square test indicated that there is not statistically significant association between them.

**3.6 Pittsburgh Sleep Quality Index (PSQI)**

The visual representations provided a very clear association between PSQI and obesity as shown in **Appendix 4.6.** However, the Chi-Square test indicated that there is not statistically significant association between them

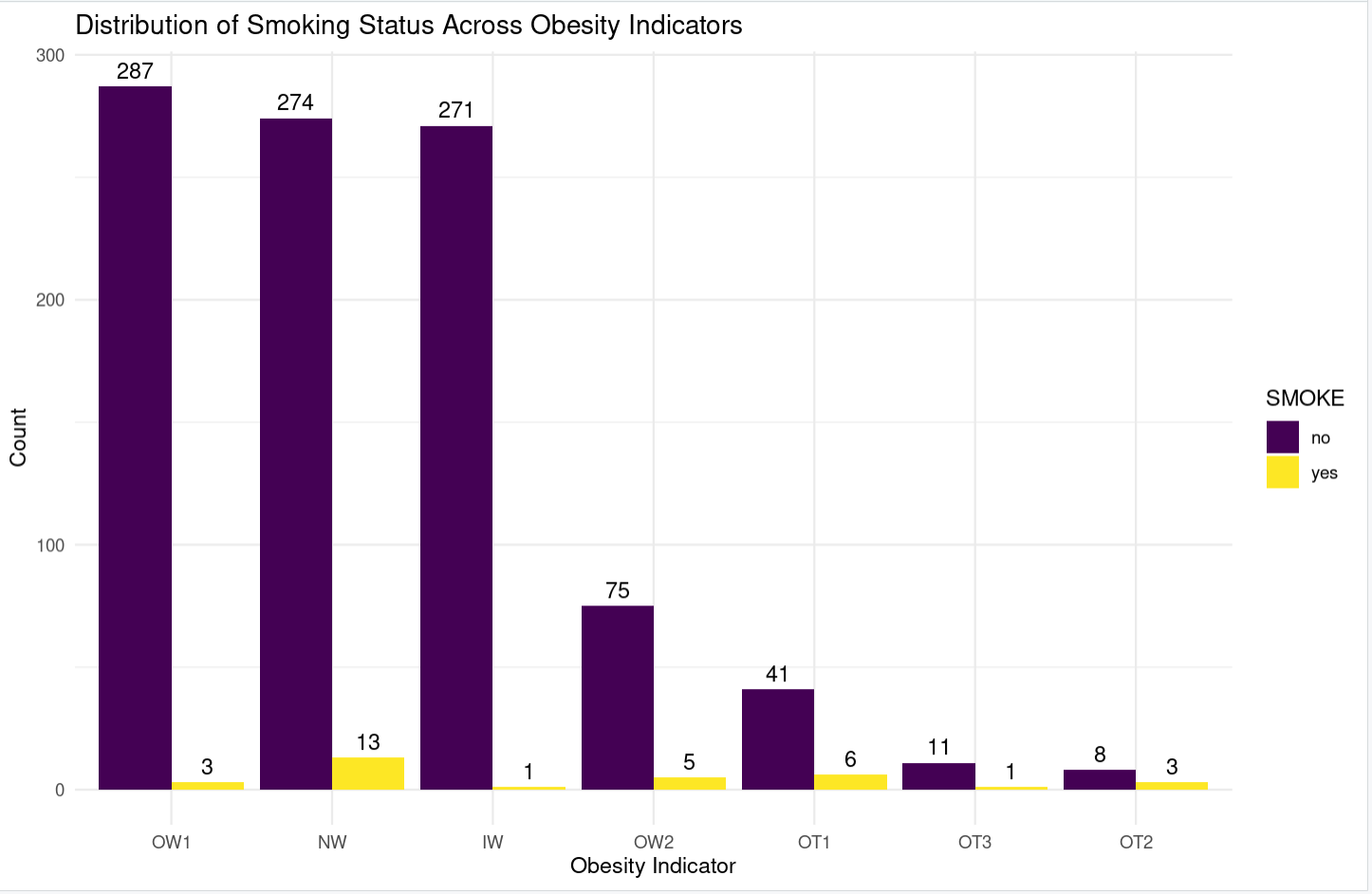
**4 Appendix Inclusion**

***Appendix 4.1 : Gender association***

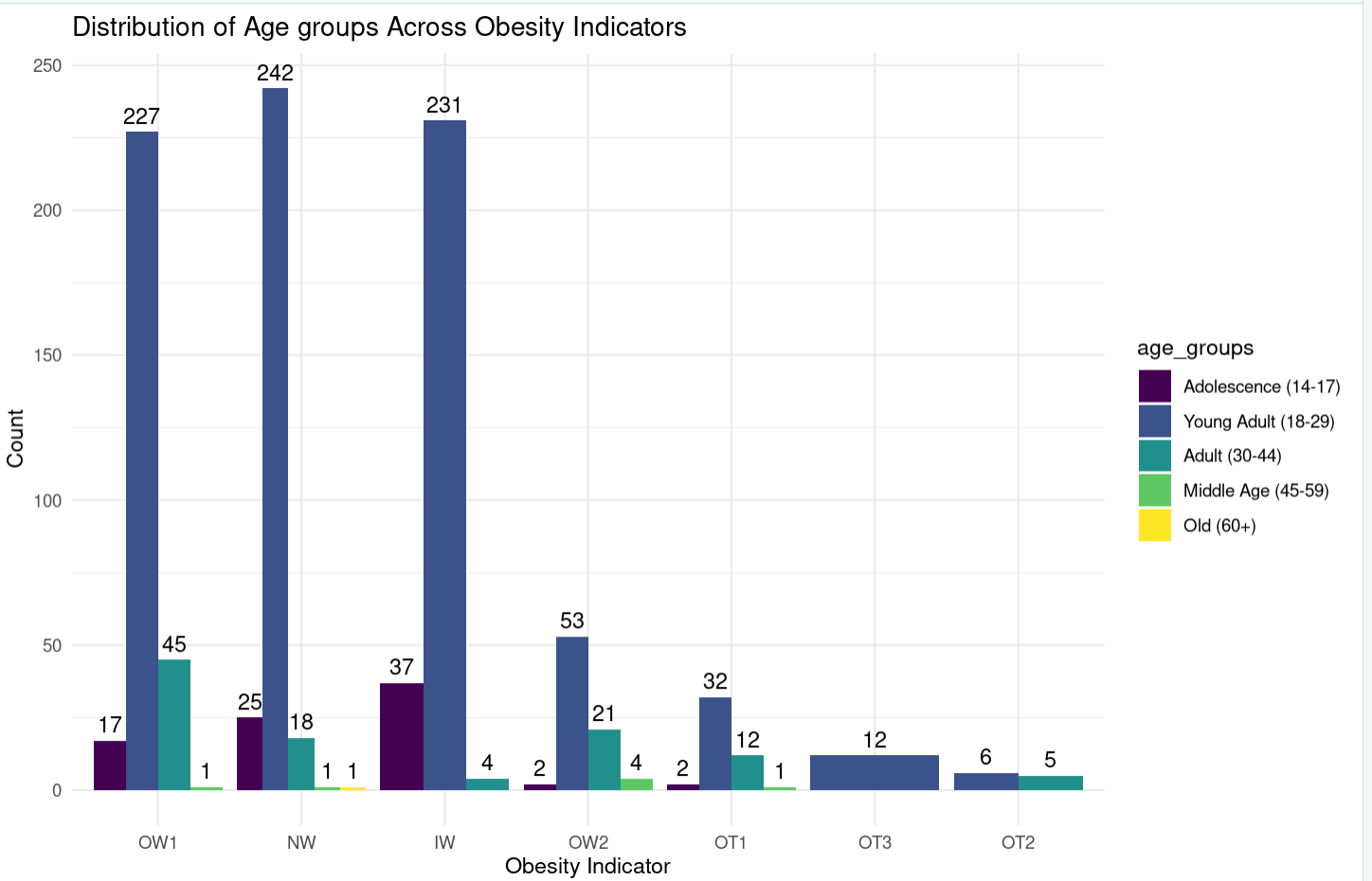


.

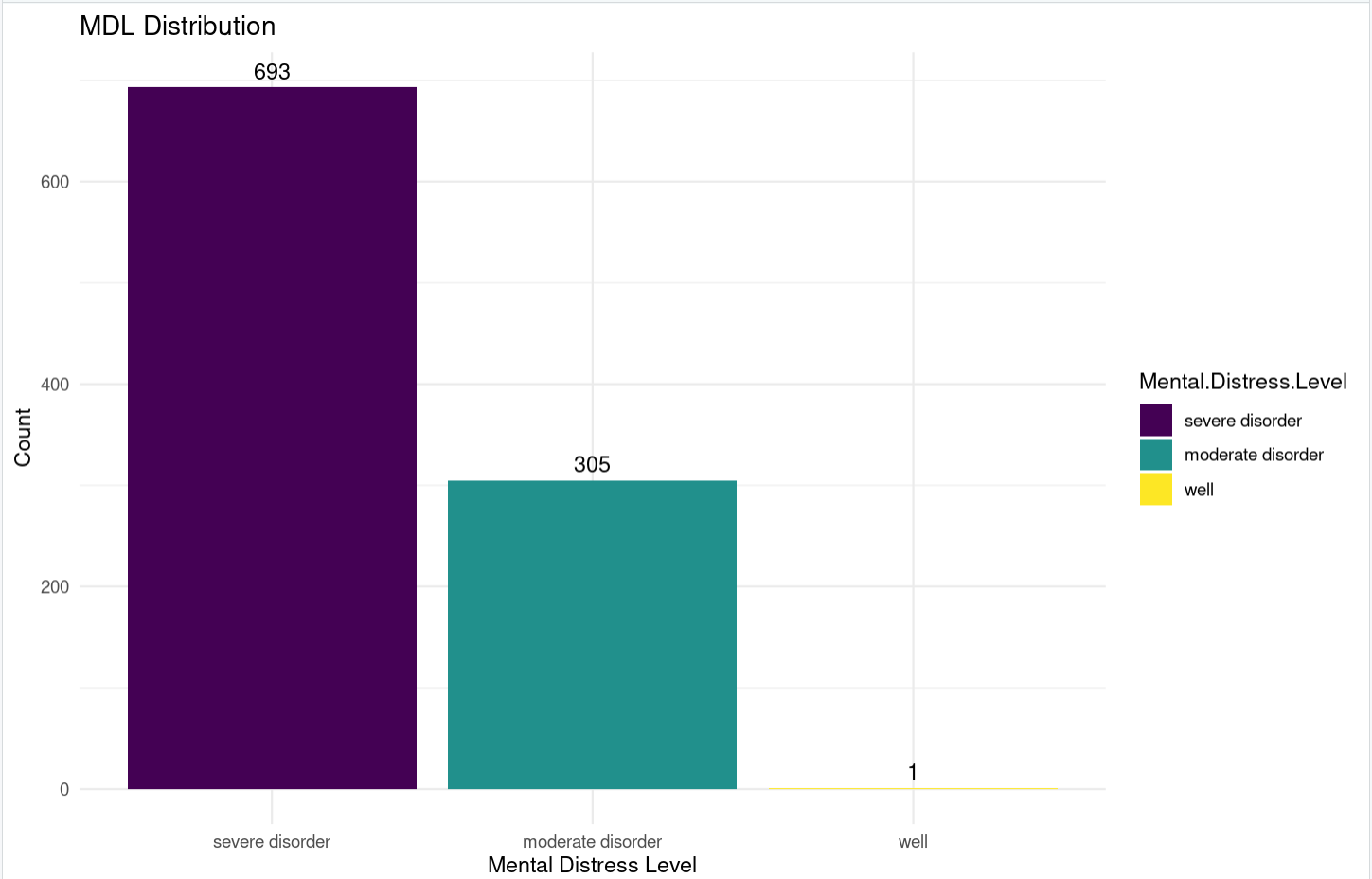
***Appendix 4.2: Smoking association***

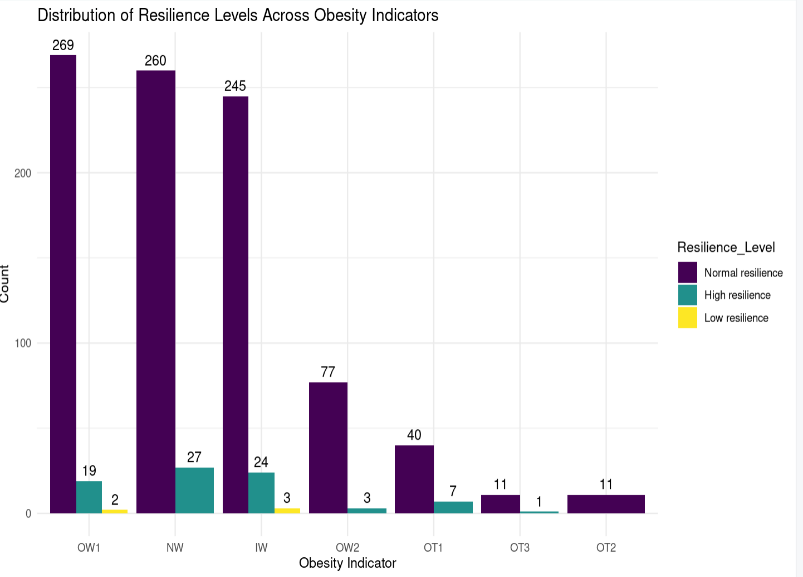
******

***Appendix 4.3:* Age association**

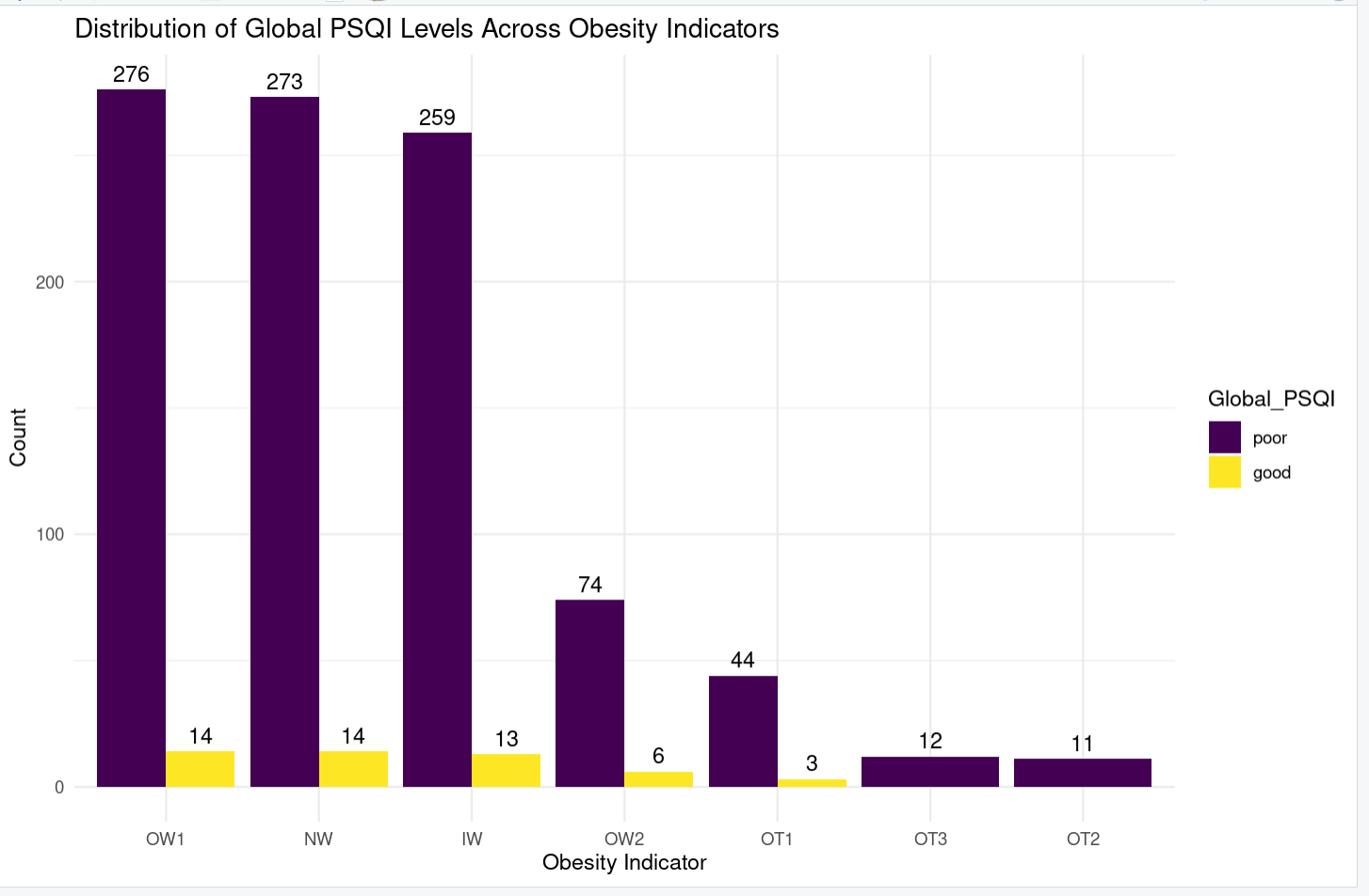


***Appendix 4.3 : Mental Distress Level association***

******

***Appendix 4.5:* Resiliehnce Level association**

***Appendix 4.6:* Pittsburgh Sleep Quality Index (PSQI)**

****