E-report on a Vital Signs Diagnosis Dataset

Presented by:

Matilla, Gabriel Jeri

Richardson, Xoli Anastasia

July 2025

Introduction

This report presents a descriptive and inferential analysis of a health dataset containing 905 adult participants (after data cleaning). The dataset includes various health-related metrics such as age, BMI, blood pressure, cholesterol, glucose levels, sleep, physical activity, stress, and lifestyle behaviors like smoking. The goal is to explore general health trends and identify significant relationships between risk factors using basic statistics and statistical tests. All analyses were conducted using R.

Methods

Descriptive statistics were computed for all variables to understand central tendencies and variability across the dataset. Measures like mean, median, standard deviation, and minimum values were used. Each health domain was reviewed for key insights such as average values, typical ranges, and clinical implications.

To examine potential associations between lifestyle behaviors and health risks, a Chi-square test of independence was performed. Specifically, we tested the relationship between Smoking Status (categorized from 0 = Non-smoker to 2 = Chain smoker) and Hypertension Category (scaled from 0 = Normal to higher levels indicating more severe hypertension).

Key Results and Figures

Descriptive Findings:

- Age: Mean of 53.8 years, ranging from 18 to 90 — majority are middle-aged or older.

- Blood Pressure: Systolic BP averaged 133 mmHg and Diastolic 81.8 mmHg, with a mean Hypertension Category of 2.94, indicating Stage 1 hypertension is common.

- BMI: Average of 26.7 kg/m² — suggests that many individuals fall into the overweight range. Reflected in Figure 1.

- Sleep: Mean of 5.39 hours per night — well below the recommended 7–9 hours, raising concern for chronic sleep deprivation.

- Physical Activity: Average of 8.49 hours/week — generally meets health guidelines but shows wide variation.

- Heart Rate: Average of 98.9 bpm — higher than normal resting heart rate, possibly linked to stress or cardiovascular strain.

- Glucose: Mean of 128 mg/dL — falls within the prediabetic/diabetic range.

- Cholesterol: Mean of 188 mg/dL — near the upper-normal threshold.

- Smoking Status: Mean score of 1.01 — indicates a mix of non-smokers, occasional, and chain smokers. Reflected in Figure 2.

- Stress Level: Average of 5.16 out of 10 — shows moderate stress levels.

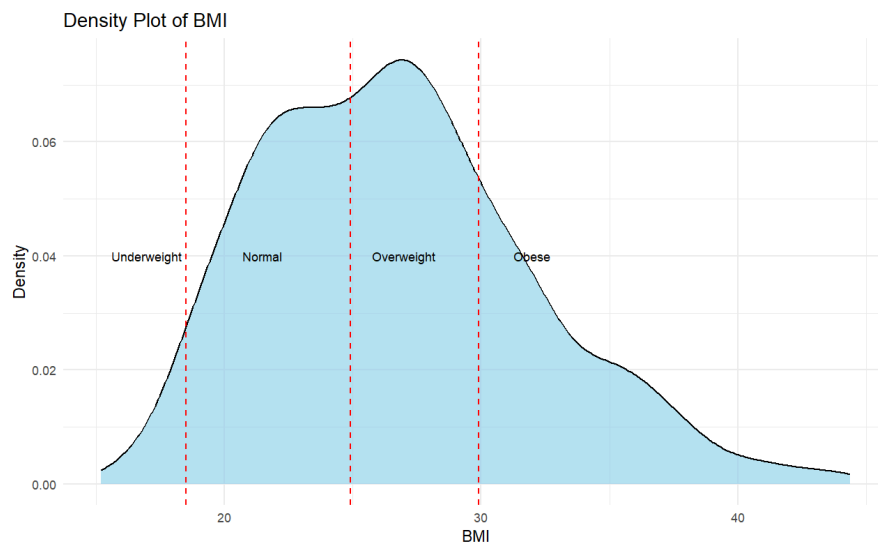- Sex Ratio: Roughly equal (50:50 split).

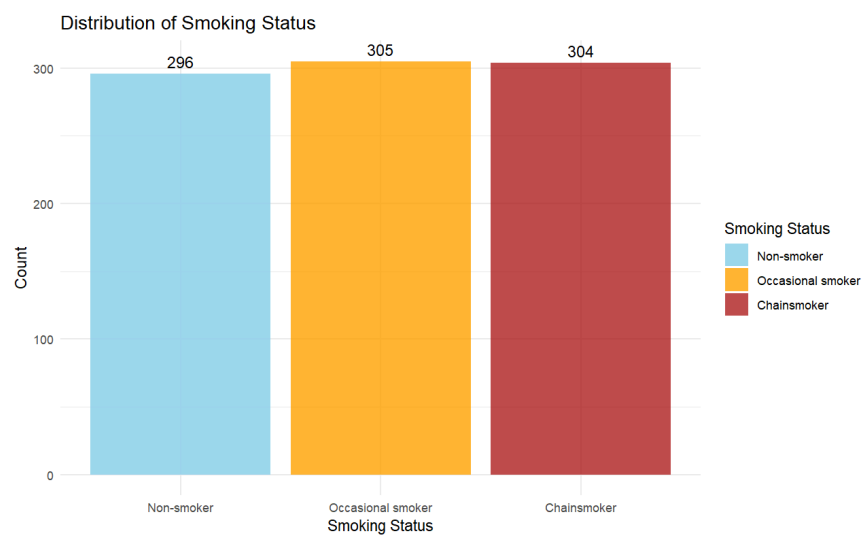**Figure 1.** Distribution of BMI values across patients

Figure 2. Distribution of patients by smoking status

Chi-square Test Result: Smoking vs Hypertension

- **Chi-square statistic** = 2715
- **Degrees of freedom** = 15
- **p-value** < 2.2e-16

This test result was highly statistically significant, providing strong evidence that smoking status is associated with hypertension severity. The large Chi-square value indicates that the distribution of hypertension categories differs substantially across smoking groups. This is reflected in Figure 3, where the majority of the heavily smoking population experiencing stage 2 hypertension.
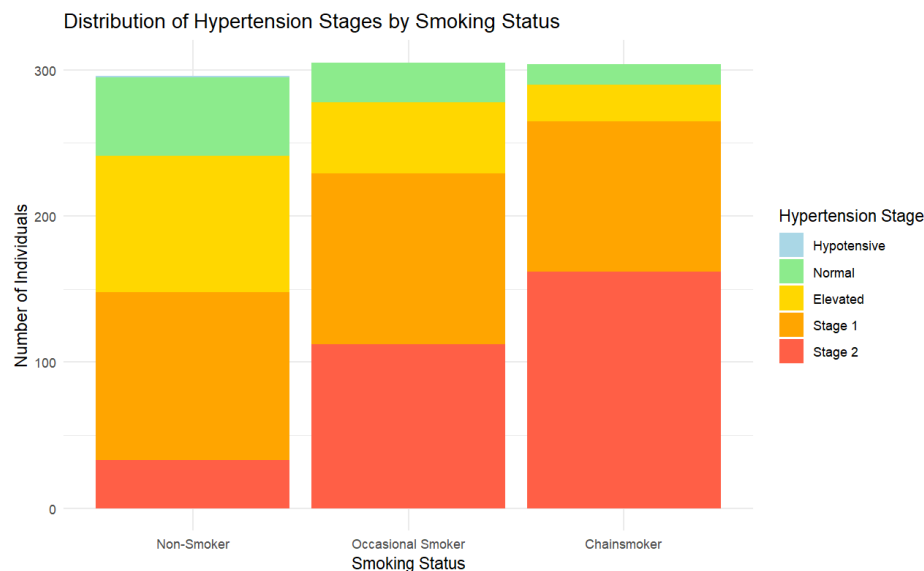

Figure 3. Distribution of hypertension stage by smoking status

Interpretation and Conclusion

Overall, the dataset suggests that the studied population faces several notable health risks. A significant portion is middle-aged or elderly, overweight, and suffers from elevated blood pressure, heart rate, and glucose levels. Sleep duration is below healthy standards, while stress is moderately high.

Most notably, the Chi-square test confirms a strong association between smoking and hypertension, aligning with known medical research on the cardiovascular dangers of tobacco use. These findings reinforce the importance of smoking cessation and blood pressure management in population health strategies.

This exploratory analysis provides a useful snapshot of population health and highlights areas for deeper research or targeted public health interventions.