Group 14

TIME SERIES MONITORING DATA

DATA SCIENECE FOR THE HEALTH SCIENCES

INTRODUCTION

According to the study of Chen et al. (2023), time series data are sequences of information organized by time, crucial for understanding and predicting various real-world phenomena, from financial markets and weather patterns to traffic flows. This data reflects how something changes over time and can include not only traditional numerical data but also video, audio, and voice recordings. Key characteristics of time series data include trend (long-term increase or decrease), periodicity (regular but unfixed rising or falling patterns), seasonality (fixed, recurring patterns tied to specific seasons or time periods), and randomness (irregular fluctuations that still adhere to statistical rules). Additionally, time series can be classified as stationary (consistent mean and variance over time) or nonstationary, with conversion to stationarity often improving forecasting accuracy.

This study explores the **"4_Time_series_Mointoring_data_Group_014.csv"** dataset, which contains time series data on average steps, stress level, and BMI for multiple patients.. Analyzing these three metrics will help us understand the relationships and temporal patterns within these health metrics.

METHODS

The dataset was first imported into the R environment and cleaned using the tidyverse and skimr packages.

Key steps included:

Data Cleaning: Using function glimpse() and summary () to understand its structure.

Type Conversion: colSums(is.na()) is used for checking missing values. rename_() is to rename these columns particularly patient , using fill() operation to address these related columns.

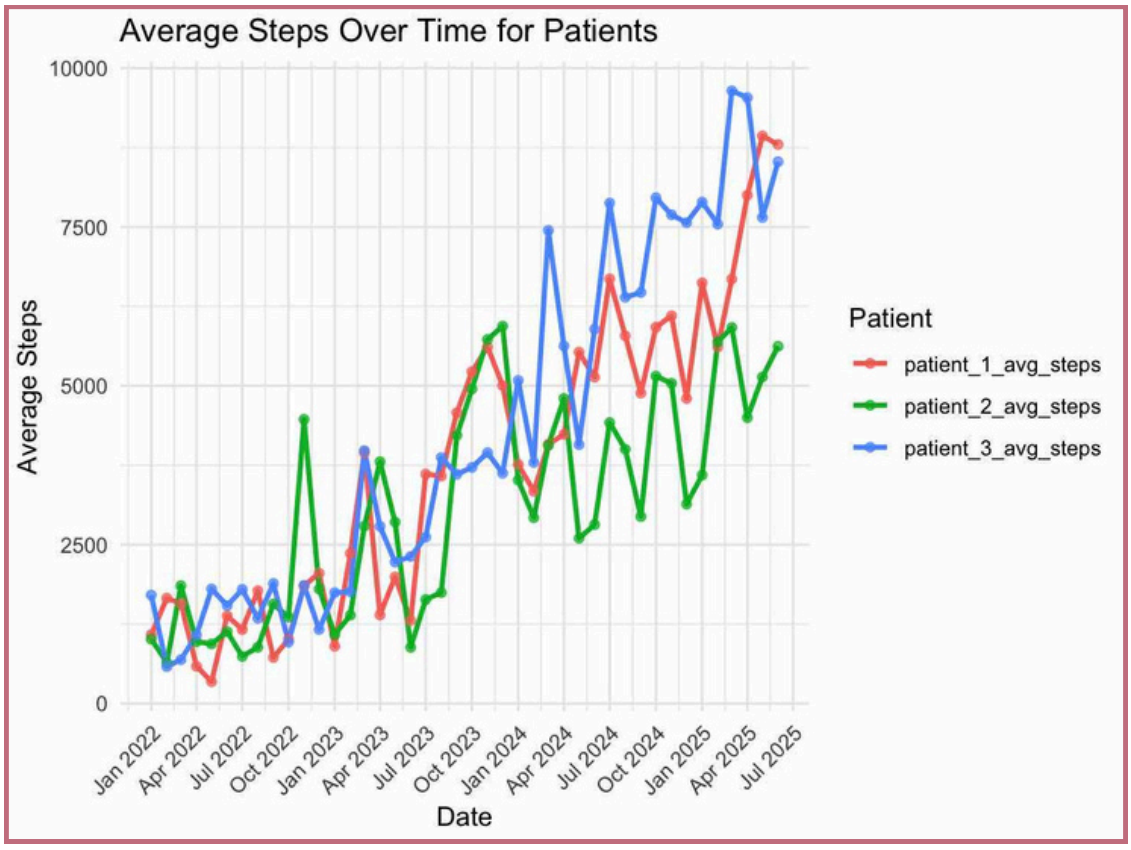
Formatting Data Types: using mutate ( ) to convert the data to month_name using as.Date to a date format, patient_ using as. numeric–using mutate to add new variables or modify the ones in a data frame.

Descriptive Statistics: Summary metrics including mean, median, quartiles and standard deviation were computed for all numeric variables using select (), summary () and skim().

Visual Analysis: A variety of plots were created using ggplot2 for visualizations of the selected metrics, including line plot for the average steps, Stress level, and BMI over time for all patients. Box plot of stress levels across patients. Scatter plot of patient 1 stress level vs. patient 1 average steps (over time)

A time series decomposition for patient_1_avg_steps to identify trend, seasonal, and random components. It also computes a cross-correlation function (CCF) between patient_1_avg_steps and patient_1_stress_level to explore their lead-lag relationships.

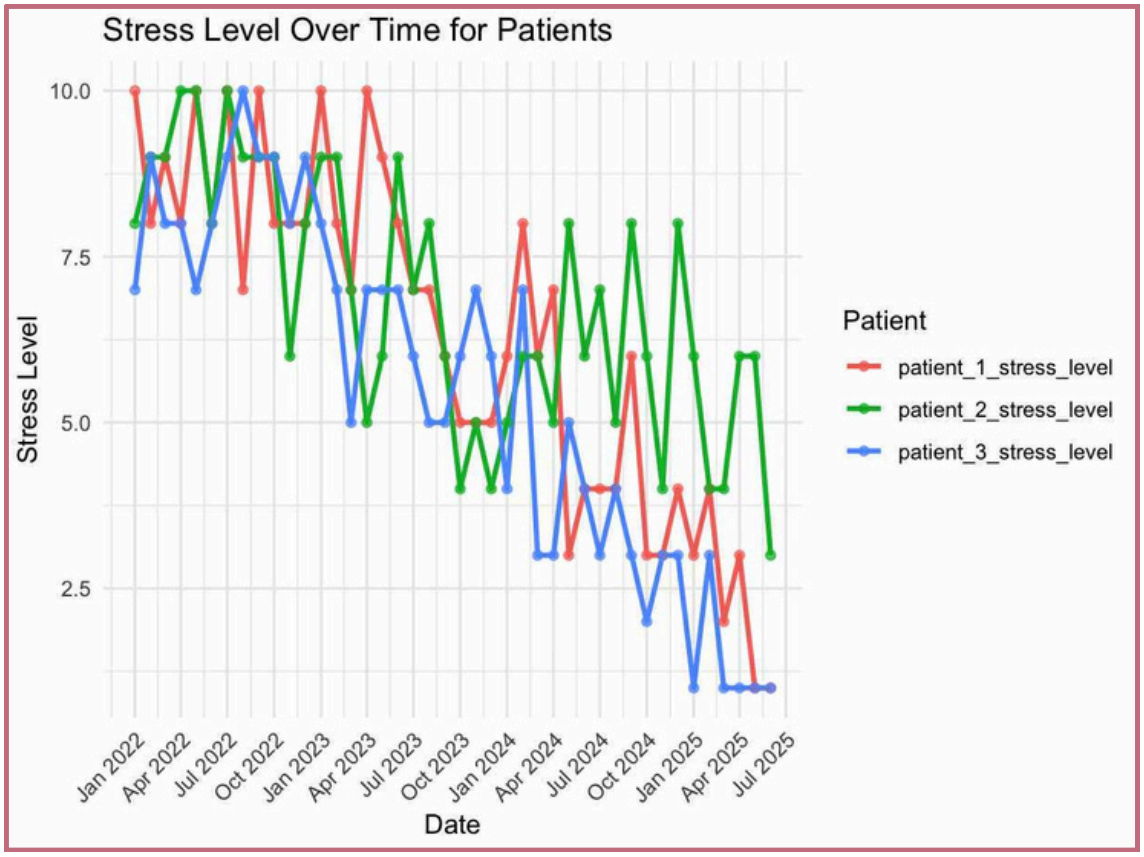
RESULTS AND FIGURES



Average Steps Over Time for Patients

The line plots show three distinct lines that are representing different patient's average steps over time, January 2022 to July 2025. While all patients showed an upward trend in their average steps, their individual steps show monthly variation. Patient 3 has higher average steps that is approaching ten thousand steps. While both patient 1 and 2 demonstrated increasing step counts, patient 1 shows a more consistent rise towards July 2025.

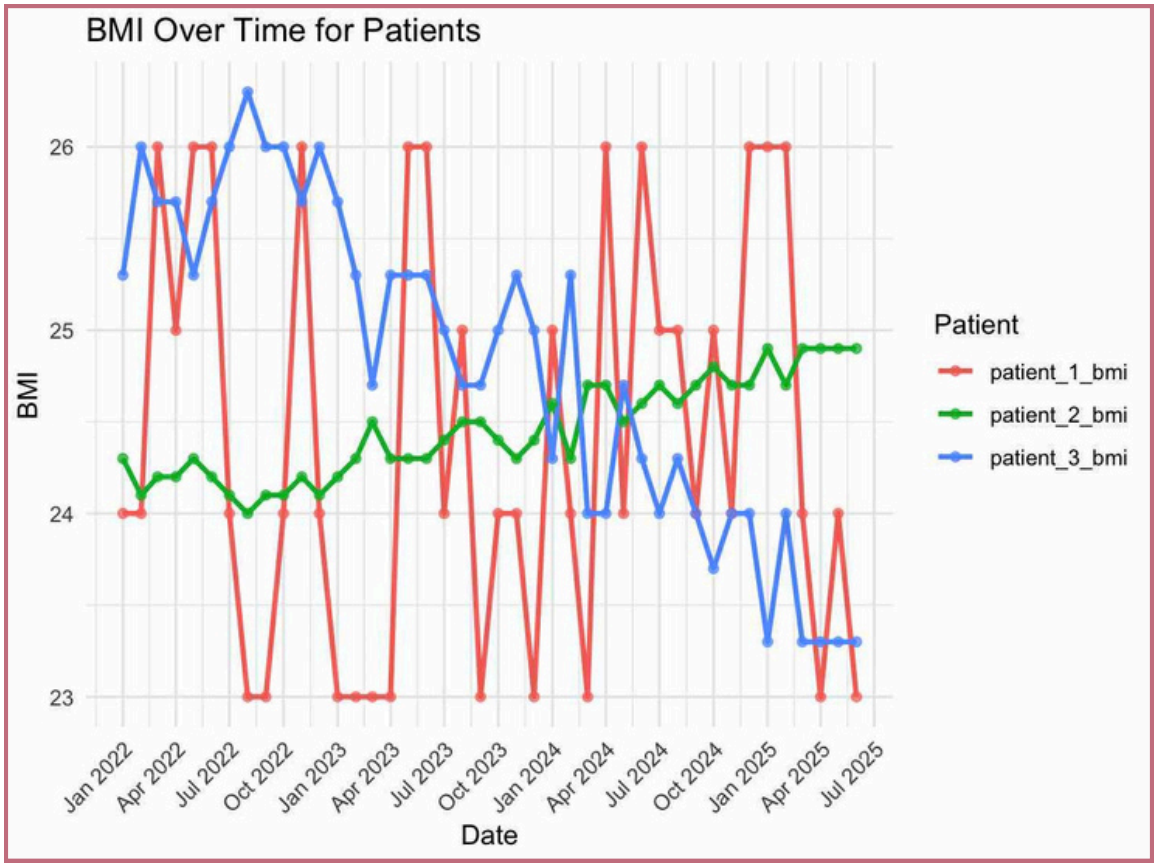
RESULTS AND FIGURES



Stress Level Over Time for Patients

The line plots show that all patients showed a high and fluctuating stress levels, reaching 7.5 to 10 of stress level in early 2022. As time passes there is a downward trend in stress levels for all patients. By reaching mid 2024 and into 2205, stress levels of the patients have dropped below 5. Patient 3 stress level have stabilized at a lower point in 2025, while patients 1 and 2 also show that there is a reduction of stress level by the end of the observed period but still continuous fluctuations.

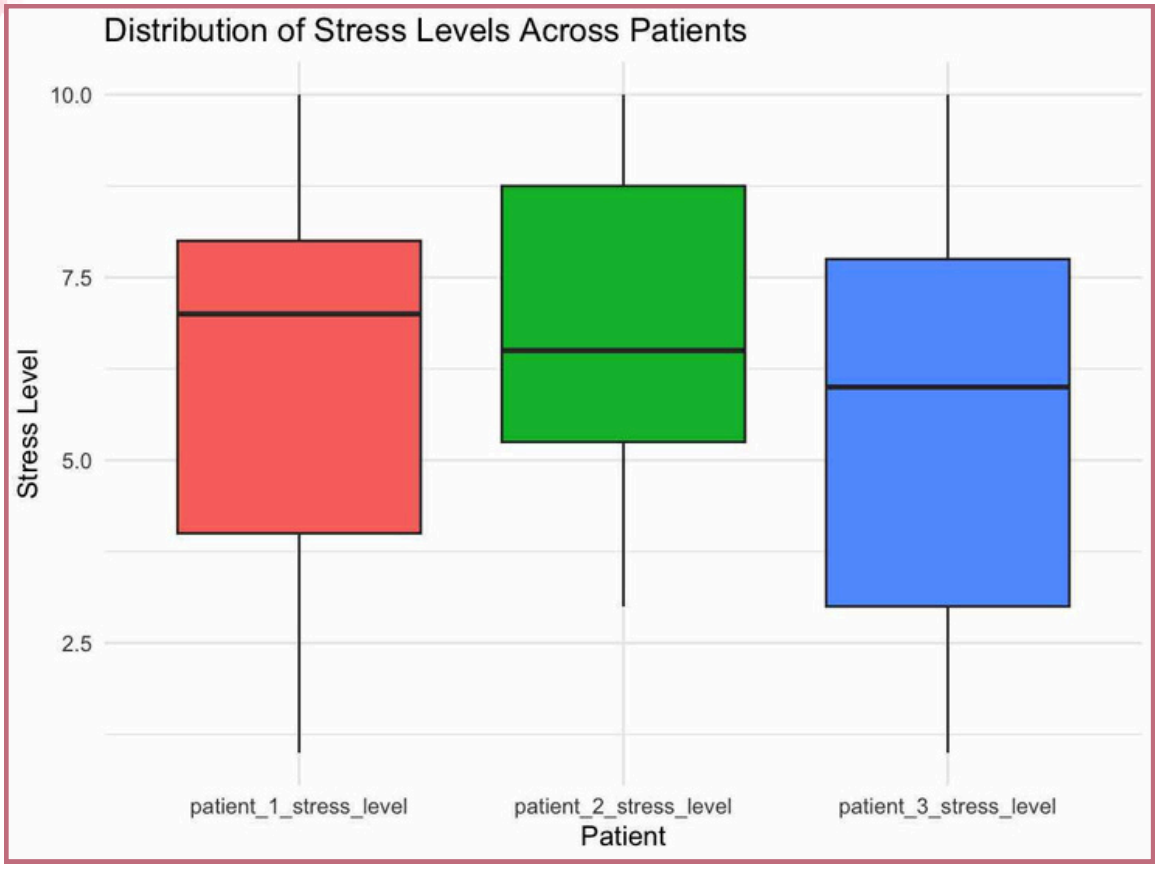
RESULTS AND FIGURES



BMI Over Time for Patients

The line plot shows three different patients BMI trends as time passes. A significant fluctuation in line is patient 1's BMI throughout the period, moving between 23 and 26. Patient 2 is relatively stable, around 24 to 25 with a mild increase towards the end of the observed period. Patient 3's BMI line is initially higher and then decreasing towards 23-24 by early year of 2025.

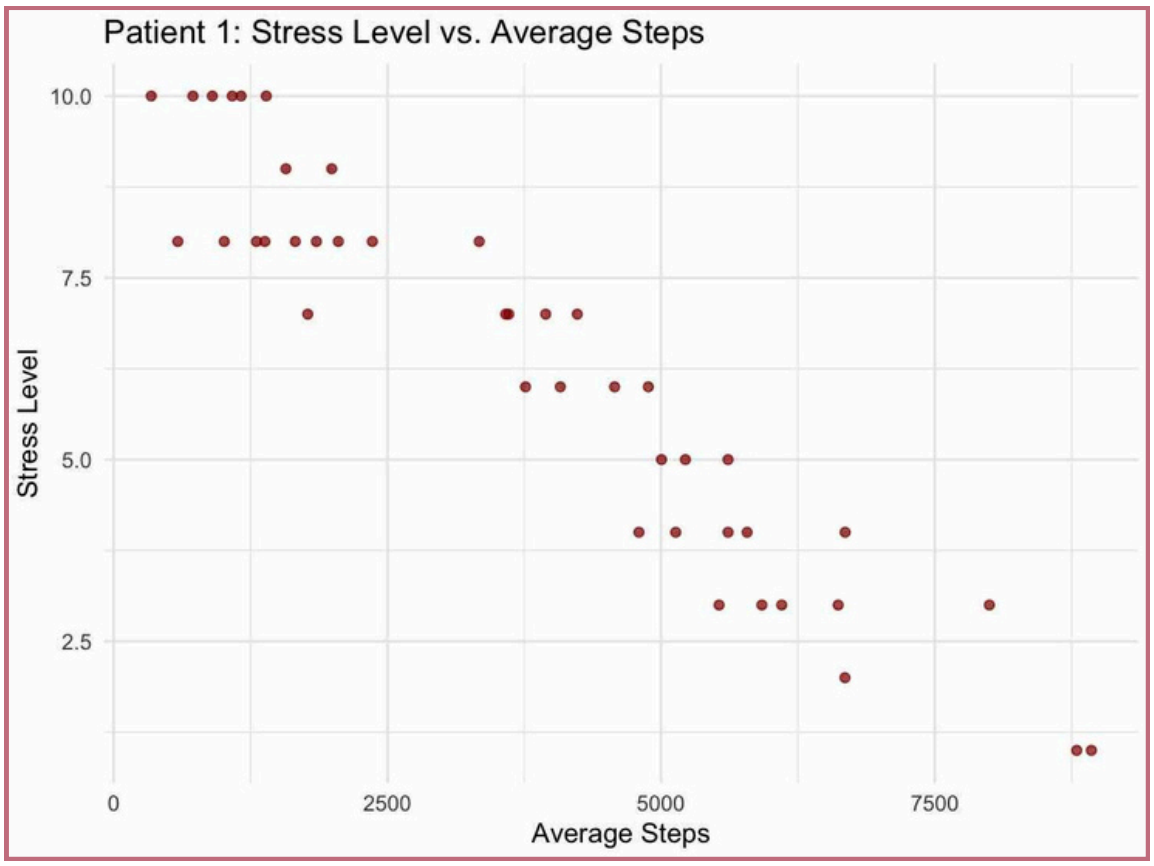
RESULTS AND FIGURES



Distribution of Stress Levels Across Patients

The box plot shows the distribution of stress levels of three different patients. Patient 1's stress level shows whiskers extend from just above 2.5 to 10, as its median stress level is around 7. Patient 2's stress level whiskers is from around 3.0 to, its median is around 6.5. Patient 3's stress level whiskers is from around 2.5 to 10, its median level is around 6. All three patients exhibit a wide range of stress level, with patient 1 having a slightly higher median stress levels, a difference to patient 2 and 3.

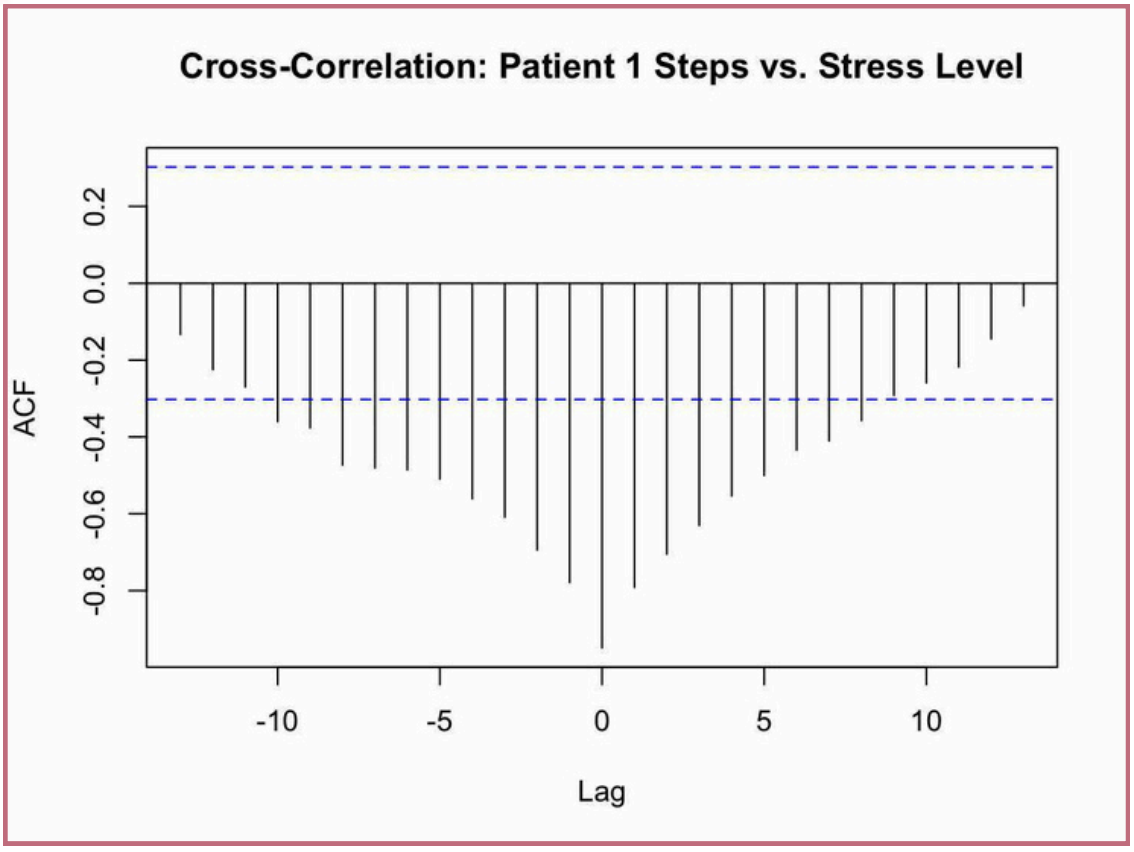
RESULTS AND FIGURES



Patient 1: Stress Level vs. Average Steps

The scatter plot shows an inverse relationship between average steps and stress level of patient 1.A downward-sloping trend from left to right. As the average steps increase, the stress level tends to decrease. Based on the plot the patient's stress level decreases as their average steps increases, we can say that the patient is in a calm state.

RESULTS AND FIGURES



Cross-Correlation : Patient 1 Steps vs. Stress Level

The Cross-Correlation Function (CCF) plot shows how steps taken relate to stress levels over time. The x-axis indicates time differences (lags) from about -10 to 10. The y-axis displays the correlation coefficient between -0.8 and 0.2. There is a strong negative correlation at lag 0. This means that when steps are high, stress levels are low, and the opposite is true. This points to a significant reverse relationship between the two variables, particularly at the same time.

CONCLUSION

The analysis of average steps, stress levels, and BMI from January 2022 to July 2025 shows clear trends for each patient. All patients generally increased their average steps, with Patient 3 often leading in step count. Stress levels started high but decreased over time, particularly for Patient 3, while Patients 1 and 2 also saw reductions despite still experiencing fluctuations. BMI trends varied; Patient 1 had significant changes, Patient 2 was stable, and Patient 3's initially high BMI dropped gradually. Box plots indicated a wide range of stress levels, with Patient 1 usually having a higher median. Notably, for Patient 1, there was a strong inverse relationship between average steps and stress levels, confirmed by scatter plots and cross-correlation analysis showing that more steps correlated with lower stress. This finding is particularly significant as it supports the widely recognized benefits of physical activity on mental well-being and could inform personalized health recommendations for stress management

REFERENCE

Chen, Z., Ma, M., Li, T., Wang, H., & Li, C. (2023). Long sequence time-series forecasting with deep learning: A survey. Information Fusion, 97, 101819. https://doi.org/10.1016/j.inffus.2023.101819