Visual Stratification for Epidemiological Analysis

José Matute
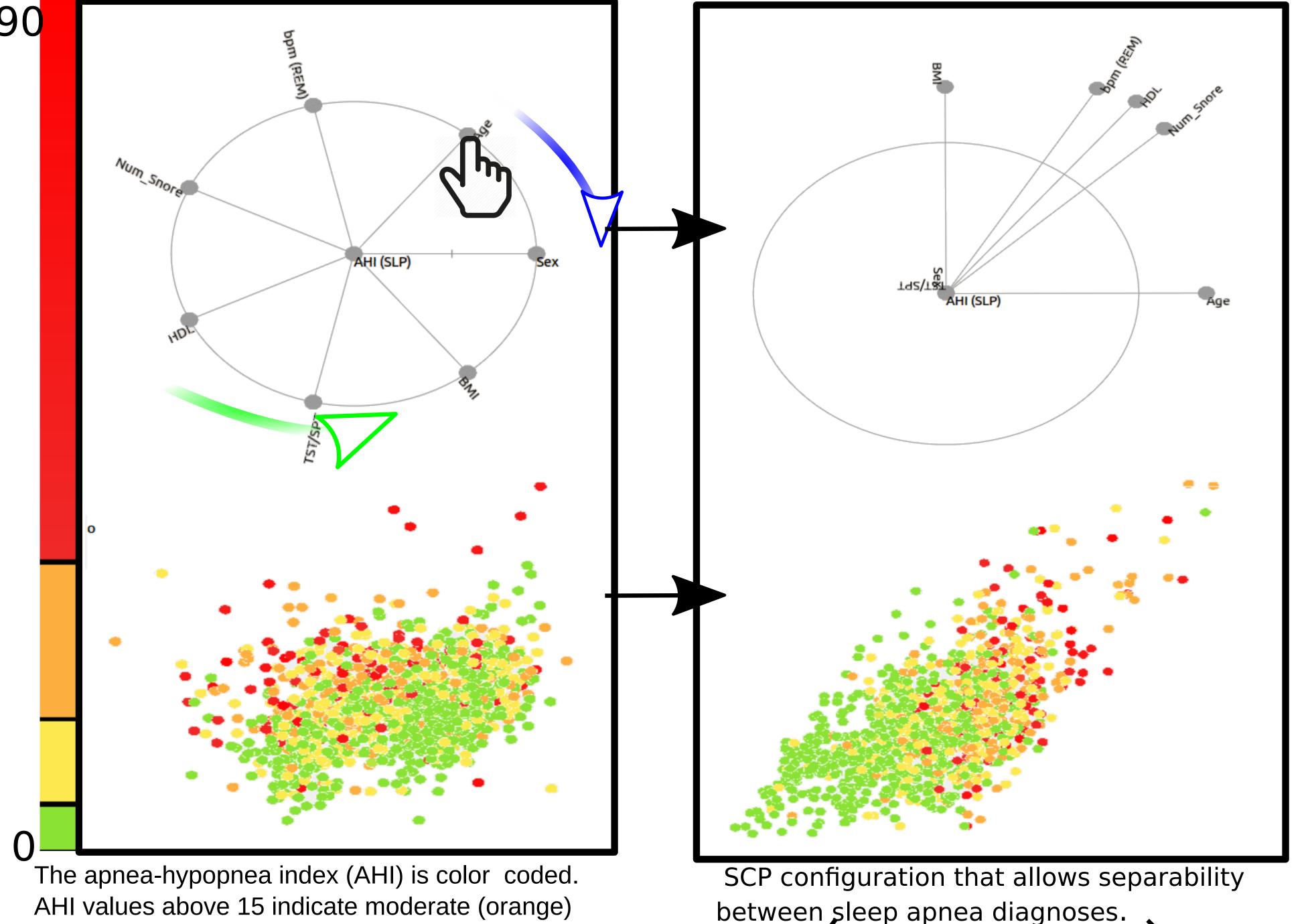Westfälische Wilhelms-Universität Münster

Lars Linsen
Westfälische Wilhelms-Universität Münster

Abstract

Investigating etiology of a disease depends on the combination of tacit medical knowledge and multivariate analysis on a wide array of collected data. Confounding variables may generate a bias when exploring disease determinants, thus, reducing the predictive capabilities of risk factors. Stratified analysis is widely used in epidemiological settings to reduce the effect of confounding factors. We propose a stratified visual analysis approach based on linear projections and interactions in a Star Coordinates Plot (SCP), where the segregation power of dimensions in multiple strata can be explored interactively. We apply our approach to gain insight into three epidemiological results using stratified analysis regarding the prevalence of sleep apnea within age and gender strata and the segregating power of well-defined epidemiological risk factors.

The apnea-hypopnea index (AHI) is color coded. AHI values above 15 indicate moderate (orange) and above 30 severe (red) apnea.

SCP configuration that allows separability between sleep apnea diagnoses.

- Star Coordinate Plots (SCP) allow for the determination of factors to segregate between diseased and non-diseased population.
- Stratified analysis reduces confounding bias.
- Multi-level stratification allow for exploring the confounding effect of multiple factors.
- Visual stratification allows for the exploration of the segregation power of factors in multiple strata.
- We explore well-defined risk factors in the segregation of moderate to severe sleep apnea

Male population has a higher risk of moderate to severe sleep apnea reflected by the higher concentration of red and orange samples.
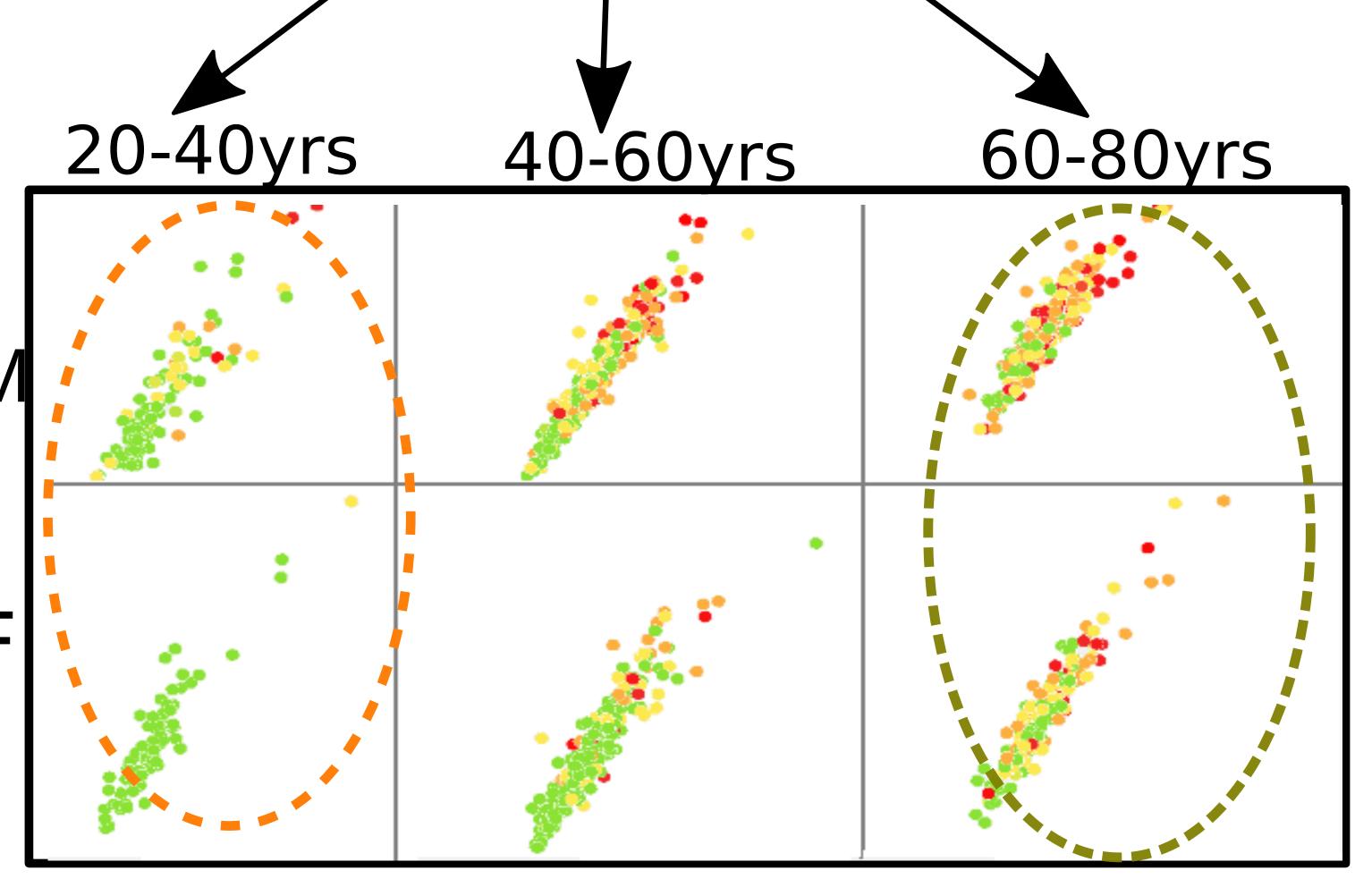
First- level gender based stratification.

The configuration is able to separate more clearly mild to severe cases in females than males.

Younger population, both males and females, display lower prevalence of apnea.

Older population display higher prevalence of sleep apnea.

Future Challenges

We will combine statistical and visual analysis for the validation of epidemiological hypotheses.

Multi-level stratification. Gender (M,F) and age (from 20 to 80, 20 years interval)