

## Data Wrangling Project Proposal

### Introduction

Sleep is a fundamental necessity of life. Sleep loss impairs task performance, cognitive performance, and mood [1]. The Centers for Disease Control and Prevention (CDC) declared that insufficient sleep is a public health epidemic [2]. Self-reported short sleep duration in the United States varies by industry and occupation, and has increased over the past two decades [3]. According to several studies, more than half of healthcare professionals report symptoms of burnout, insomnia, depression, anxiety, and mental health illnesses [4,5]. Some studies have shown race and ethnic disparities in sleep patterns in the United States [6]. Sleep duration varies across the human lifespan and the ideal amount of sleep recommended each night may vary due to genetic factors [7]. Some studies have shown an inverse relationship between sleep with age and insufficient sleep has become widespread, especially among adolescents [7,8]. Our study will investigate the association between sleep and other covariates, and will highlight the consequences of sleep insomnia on mental health.

### Data Description

The selected dataset consists of three csv files: 1) “insomnia\_data.csv”; 2) “insomnia\_data\_dictionary.csv”; 3) “insomnia\_item\_level\_data.csv”. The main dataset (“insomnia\_data.csv”) features 95 records, each with 95 fields capturing questionnaire composite scores and z-scored values. Variables includes but not limiting to: ‘Sex’, ‘Age’, ‘White’, ‘Black’, ‘Asian’, ‘ISI\_total’ (Insomnia severity index), ‘ASHS\_SleepEnvirnmont’, ‘ASHS\_DaytimeSleep’, ‘ASHS\_substances’, ‘ASHS\_sleepStability’, ‘GCTI\_anxiety’, ‘GCTI\_worries’, ‘NEO\_openness’, ‘asq\_home’ (Stress of Home Life), ‘asq\_school’, ‘asq\_finance’, ‘asq\_peer’, ‘cope\_disengage\_mental’, and etcetera. The “insomnia\_data\_dictionary.csv” clarifies variables with 95 rows detailing variable short-names. The “insomnia\_item\_level\_data.csv” supplements psychosocial factors and questionnaires responses related to insomnia. Both serve as a vital resource for understanding the intricate relationship between sleep patterns and psychological factors in adolescent insomnia.

### Aim 1: Visualization of insomnia symptomatology in adolescents by subgroups

- 1.1. Visually identify distinct subgroups among adolescents based on demographic variables (e.g., Age, Race) and their distribution on the GCTI, ASHS, ASQ battery tests.
- 1.2. Utilize data visualization techniques (e.g., heatmap analysis) to depict the relationship between identified subgroups and psychological factors. Provide an intuitive representation of how diverse psychosocial factors contribute to varied sleep patterns in adolescents.

### Aim 2: Statistical Analysis of Sleep Quality and Habits in Adolescents

- 2.1. Use sleep quality to predict mental health and habits among adolescents with regression analyses.
- 2.2. Investigate if the interaction effect of significant predictor factor(s) above and the race factor is statistically significant.

### Concluding Remarks

This project aims to analyze a dataset gathered from self-reports and standardized questionnaires, focusing on exploring insomnia-related factors in adolescents. Our approach involves utilizing a combination of Excel, R, and potentially Python for data wrangling, analysis, and visualization. The objective is to discover patterns, trends, and distinctive clinical features of adolescent insomnia, aiming to visualize the severity and complexity of insomnia symptomatology across various behavioral and psychosocial factors.

## References

1. Troynikov O, Watson CG, Nawaz N. Sleep environments and sleep physiology: A review. *J Therm Biol.* 2018;78:192-203. doi:10.1016/j.jtherbio.2018.09.012
2. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of Healthy Sleep Duration among Adults--United States, 2014. *MMWR Morb Mortal Wkly Rep.* 2016;65(6):137-141. Published 2016 Feb 19. doi:10.15585/mmwr.mm6506a1
3. Luckhaupt SE, Tak S, Calvert GM. The prevalence of short sleep duration by industry and occupation in the National Health Interview Survey. *Sleep.* 2010;33(2):149-159. doi:10.1093/sleep/33.2.149
4. Prasad K, McLoughlin C, Stillman M, et al. Prevalence and correlates of stress and burnout among U.S. healthcare workers during the COVID-19 pandemic: a national cross-sectional survey study. *EClinicalMedicine* 2021;35:100879-100879.
5. Bryant-Genevier J, Rao CY, Lopes-Cardozo B, et al. Symptoms of depression, anxiety, post-traumatic stress disorder, and suicidal ideation among state, tribal, local, and territorial public health workers during the COVID-19 pandemic — United States, March–April 2021. *MMWR Morb Mortal Wkly Rep* 2021;70:1680-1685.
6. Johnson DA, Jackson CL, Williams NJ, Alcántara C. Are sleep patterns influenced by race/ethnicity - a marker of relative advantage or disadvantage? Evidence to date. *Nat Sci Sleep.* 2019;11:79-95. Published 2019 Jul 23. doi:10.2147/NSS.S169312
7. Chaput JP, Dutil C, Sampasa-Kanyinga H. Sleeping hours: what is the ideal number and how does age impact this?. *Nat Sci Sleep.* 2018;10:421-430. Published 2018 Nov 27. doi:10.2147/NSS.S163071
8. Matricciani L, Olds T, Petkov J. In search of lost sleep: secular trends in the sleep time of school-aged children and adolescents. *Sleep Med Rev.* 2012;16(3):203-211. doi:10.1016/j.smr.2011.03.005