Exploring Stress Among Teenagers

Christian Arndt & Reno Malanga

April 21, 2025

1 Introduction

2 Data

3 Methods

4 Wrapping Up

Choosing a Topic

Choosing a Topic

- BEHAVIORAL Data Science

Choosing a Topic

- BEHAVIORAL Data Science
- Focus on stress data

Choosing a Topic

- BEHAVIORAL Data Science
- Focus on stress data
  - Reno's specific research interest!

Choosing a Topic

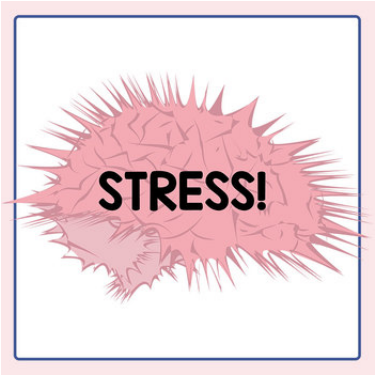
- BEHAVIORAL Data Science
- Focus on stress data
  - Reno's specific research interest!
- Look into the relationship between stress and lifestyle in teenagers

Choosing a Topic

- BEHAVIORAL Data Science
- Focus on stress data
  - Reno's specific research interest!
- Look into the relationship between stress and lifestyle in teenagers
  - Kaggle dataset

What’s up with teens?

What's up with teens?



The Data

The Data

- Demographic data

The Data

- Demographic data
  - Age, gender

The Data

- Demographic data
  - Age, gender
- Lifestyle data

The Data

- Demographic data
  - Age, gender
- Lifestyle data
  - Physical activity, academic performance, social media time, screen time, sleep time, support system

The Data

- Demographic data
  - Age, gender
- Lifestyle data
  - Physical activity, academic performance, social media time, screen time, sleep time, support system
- Stress data

The Data

- Demographic data
  - Age, gender
- Lifestyle data
  - Physical activity, academic performance, social media time, screen time, sleep time, support system
- Stress data
  - Self-reported (integer scale of 1-5), wearable stress metric (float scale 0-1)

The Data (cont.d)

The Data (cont.d)

- Some interesting statistics:

The Data (cont.d)

- Some interesting statistics:
  - Max social media hours per day: 10

The Data (cont.d)

- Some interesting statistics:
  - Max social media hours per day: 10
  - Min sleep hours per day: 4

The Data (cont.d)

- Some interesting statistics:
  - Max social media hours per day: 10
  - Min sleep hours per day: 4
  - Max screen time per day: 12

The Data (cont.d)

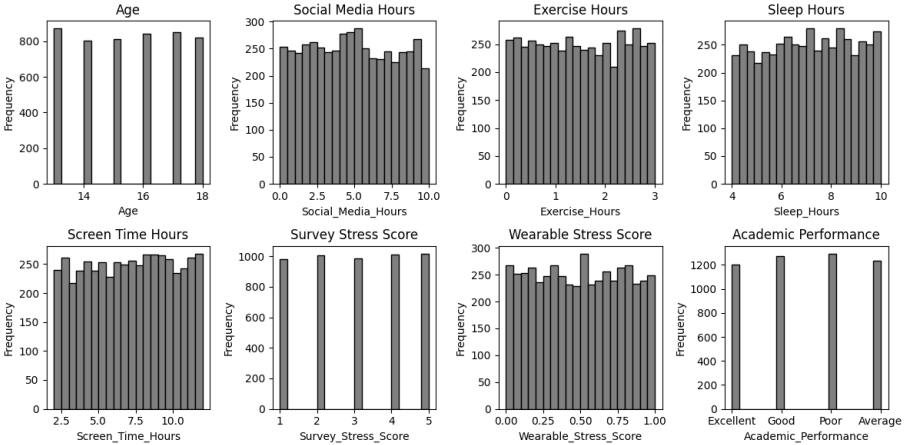
- Some interesting statistics:
  - Max social media hours per day: 10
  - Min sleep hours per day: 4
  - Max screen time per day: 12

  - These are all the same person

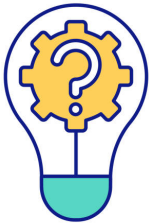
The Data (cont.d)

- Some interesting statistics:
  - Max social media hours per day: 10
  - Min sleep hours per day: 4
  - Max screen time per day: 12
  - These are all the same person
- Some of the teens are not OK

The Data (cont.d, cont.d)

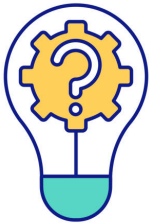


Methods



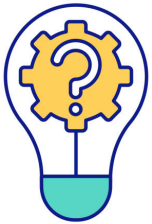
Methods

- Two experimental stages:



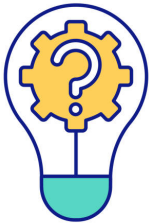
Methods

- Two experimental stages:
  - Distill lifestyle data



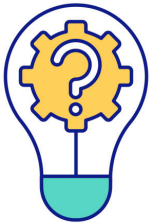
Methods

- Two experimental stages:
  - Distill lifestyle data
    - Factor analysis on lifestyle parameters



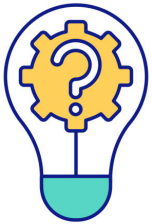
Methods

- Two experimental stages:
  - Distill lifestyle data
    - Factor analysis on lifestyle parameters
  - Relate lifestyle to stress



Methods

- Two experimental stages:
  - Distill lifestyle data
    - Factor analysis on lifestyle parameters
  - Relate lifestyle to stress
    - Regression on lifestyle and stress levels



Methods: Factor Analysis

Methods: Factor Analysis

- Factors represent abstract teenager lifestyle dimensions

Methods: Factor Analysis

- Factors represent abstract teenager lifestyle dimensions

Methods: Factor Analysis

- Factors represent abstract teenager lifestyle dimensions
- Concerns
  - Factor analysis is a dimensionality reduction

Methods: Factor Analysis

- Factors represent abstract teenager lifestyle dimensions
- Concerns
  - Factor analysis is a dimensionality reduction
    - We have 6 features related to lifestyle metrics; not very many dimensions to reduce

Methods: Factor Analysis

- Factors represent abstract teenager lifestyle dimensions
- Concerns
  - Factor analysis is a dimensionality reduction
    - We have 6 features related to lifestyle metrics; not very many dimensions to reduce
  - Potential alternative: clustering (where clusters represent lifestyles, instead of factors representing lifestyle dimensions)

Methods: Factor Analysis

- Factors represent abstract teenager lifestyle dimensions
- Concerns
  - Factor analysis is a dimensionality reduction
    - We have 6 features related to lifestyle metrics; not very many dimensions to reduce
  - Potential alternative: clustering (where clusters represent lifestyles, instead of factors representing lifestyle dimensions)
    - Can't use clustering output in regression

Methods: Regression

Methods: Regression

- Regression inputs depend on how well the data factorizes

Methods: Regression

- Regression inputs depend on how well the data factorizes
  - If there are obvious lifestyle factors, we can regress from those factors

Methods: Regression

- Regression inputs depend on how well the data factorizes
  - If there are obvious lifestyle factors, we can regress from those factors
  - If not, we regress from the original features

Methods: Regression

- Regression inputs depend on how well the data factorizes
  - If there are obvious lifestyle factors, we can regress from those factors
  - If not, we regress from the original features
- Fit a number of model types

Methods: Regression

- Regression inputs depend on how well the data factorizes
  - If there are obvious lifestyle factors, we can regress from those factors
  - If not, we regress from the original features
- Fit a number of model types
  - Linear, random forest, and neural network

Methods: Regression

- Regression inputs depend on how well the data factorizes
  - If there are obvious lifestyle factors, we can regress from those factors
  - If not, we regress from the original features
- Fit a number of model types
  - Linear, random forest, and neural network
- Goal: simplest (most explainable) model with highest performance

What's left

What's left

- We have prepared the data and produced summary statistics

What's left

- We have prepared the data and produced summary statistics
- What's left:

What's left

- We have prepared the data and produced summary statistics
- What's left:
  - Factor analysis

What's left

- We have prepared the data and produced summary statistics
- What's left:
  - Factor analysis
  - Regressor models

Thank you!