Data Transformation

Project Horse

Lynna King, Luis PV, Anh Do, Ethan Ding

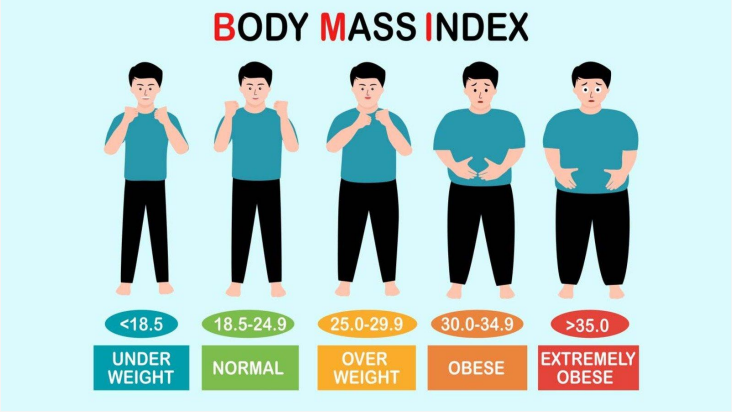
March 26, 2026

Research Question

Among U.S. adults, how is body mass index (BMI) associated with sedentary behavior (e.g. TV hours), and does this relationship differ by smoking status or age group?

Hypothesis

- Expect that individuals with **higher sedentary behavior** (measured by more TV hours) and **shorter sleep duration** will have higher BMI
- Relationships may differ by smoking status and age group.



Motivation

- Obesity is a **leading** driver of preventable chronic disease in the United States (contributor to type 2 diabetes, cardiovascular disease, certain cancers)
- Understand **health effects** of socioeconomic disadvantage
- **Rise** of diet and exercise culture
- Ozempic going **generic** in 2026 for certain countries

Introduction to NHANES

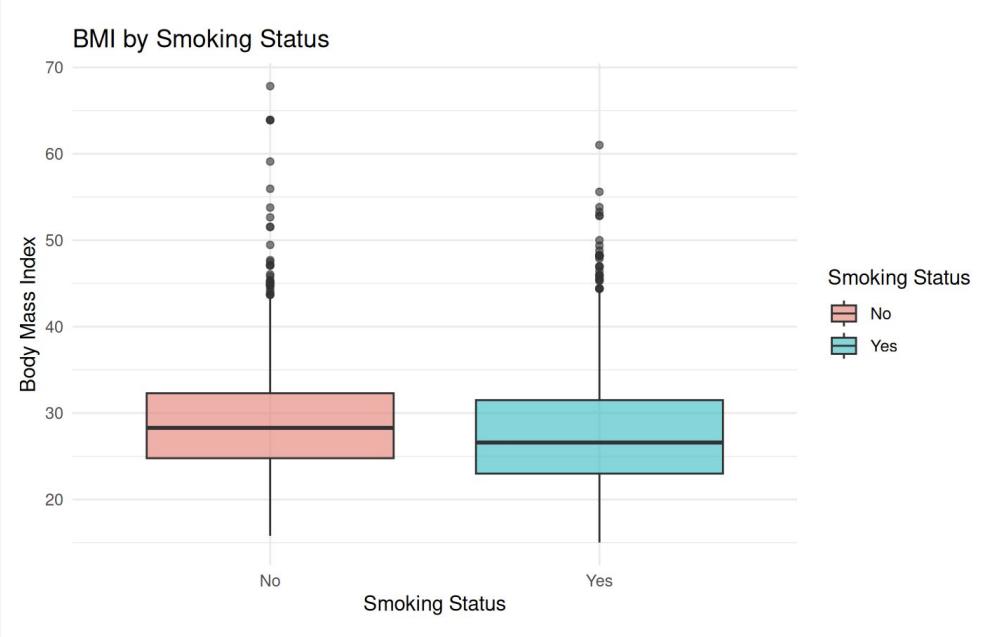
About NHANES, aka National Health and Nutrition Examination Survey

- Our specific dataset is published by the CDC Division of Nutrition, Physical Activity, Obesity
- Combines in-person interviews, physical exams, and lab tests to assess the health of U.S. adults
- Sample: ~ 10,000 U.S. adults from 2009–2012, selected using a complex stratified random sampling design to be nationally representative
- Study type: Cross-sectional, a snapshot in time, not tracking individuals over time
- Body mass index (BMI) estimates body fat → divides a person's weight in kilograms by their height in meters squared

Key Variables

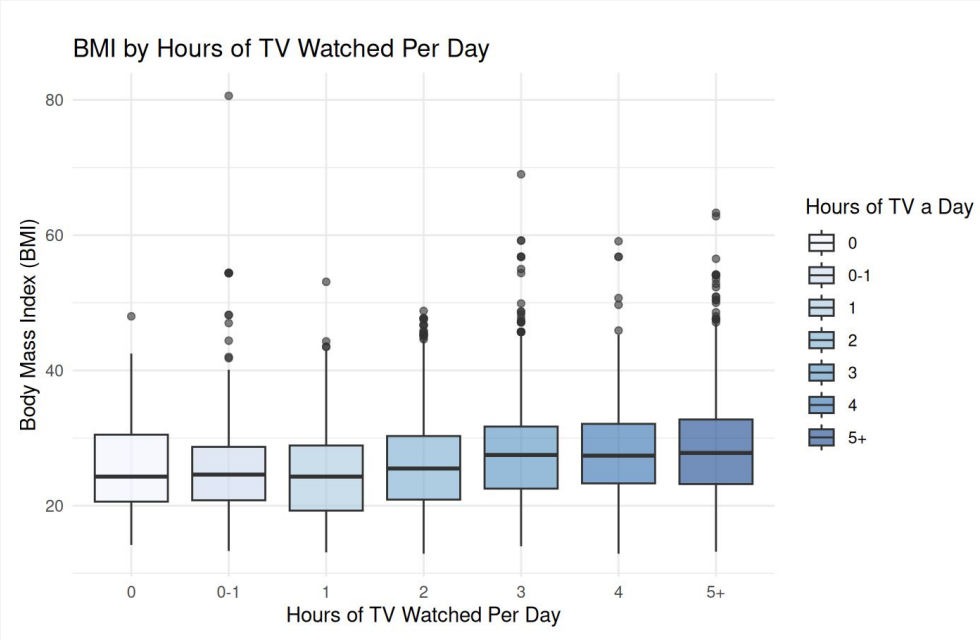
- bmi — continuous, clinician-measured
- tv_hrs_day — daily TV hours (proxy for sedentary behavior: 0, 0–1, 1, 2, 3, 4, 5+)
- sleep_hrs_night — self-reported hours of sleep per night
- smoke_now — current smoking status (Yes/No)
- age_decade — age grouped by decade (e.g., 20s, 30s, 40s...)

BMI vs Smoking



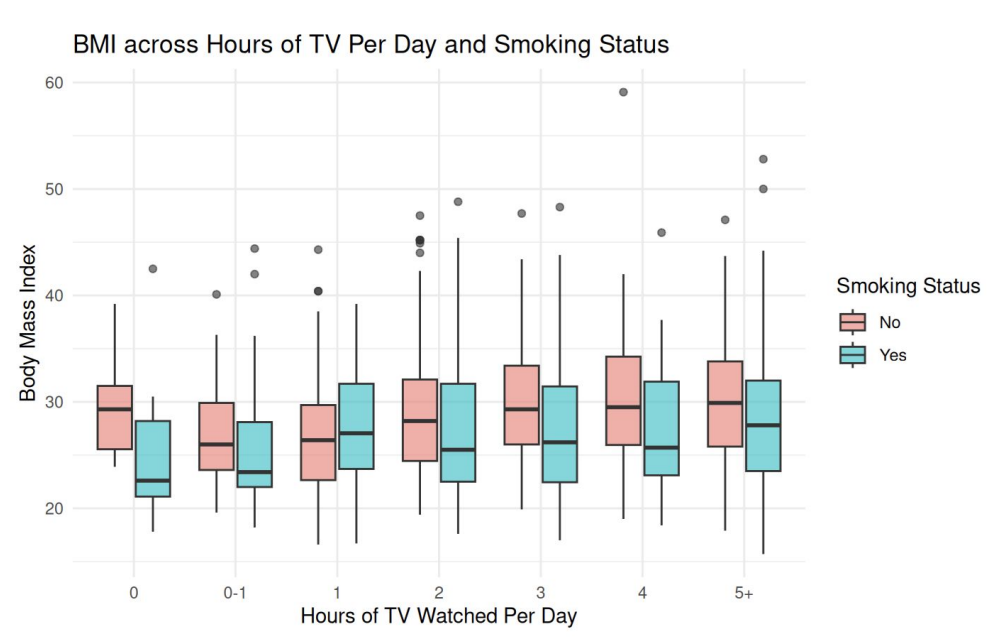
- Smokers have lower median BMI than non-smokers
- Distribution for smokers is slightly shifted downward
- Substantial overlap between the groups Outliers exist in both groups, indicating variability
- Importance
  - Lower BMI among smokers may be related to metabolic effects of nicotine However, the overlap suggests smoking is not a strong predictor of BMI on its own
- Smoking has a statistically visible but relatively weak relationship with BMI

BMI vs TV



- Median BMI increases steadily as TV hours increase
- Individuals watching 4–5+ hours have noticeably higher BMI than those watching 0–1 hours
- Spread of BMI also increases at higher TV levels, suggesting greater variability
- Pattern suggests a positive association between sedentary behavior and BMI
- Importance
  - Increased sedentary time may contribute to higher BMI through reduced physical activity and energy expenditure

Stratified Analysis



- BMI increases with TV hours for both smokers and non-smokers
- Parallel upward trends suggest the relationship is consistent across groups
- Smokers have slightly lower BMI at each TV level, but the gap remains small
- No evidence that smoking changes the direction of the relationship
- Interpretation
    - Sedentary behavior appears to influence BMI independently of smoking status Smoking shifts BMI slightly but does not alter the overall pattern
- TV viewing remains the dominant factor associated with BMI, regardless of smoking status

The -tion and -tions

Finding	Interpretation	Limitations
TV hours ↑ BMI Consistent across all TV levels	Sedentary behavior independently predicts higher BMI	TV ≠ all sedentary behavior; desk work & phone use not captured
Smokers: slightly lower BMI Small gap, much overlap	Smoking is a confounder, not a modifier — doesn't change the pattern	Self-reported smoking subject to social desirability bias
TV–BMI trend holds for smokers and non-smokers alike	TV time is a modifiable target for obesity prevention	Observational data — cannot establish causation