

Obesity Causes Full Analysis

Abeer Gaber

2026-01-02

Contents

1. Introduction	2
2. Data Overview	2
3. Numeric Variables Correlation	3
4. Adult Obesity Distribution	4
5. Child Obesity Distribution	5
6. Scatter: Sugar Consumption vs Adult Obesity	6
7. Scatter: Fast Food Sales vs Adult Obesity	7
8. Scatter: Physical Activity vs Adult Obesity	8
9. Scatter: Animal Protein Intake vs Adult Obesity	9
10. Fruit & Vegetable Consumption vs Adult Obesity	10
11. Boxplot: Obesity by Income Level	11
12. Regression: Adult Obesity	11
13. Regression: Child Obesity	12
14. Death Rate Change vs Adult Obesity	14
15. Diabetes Prevalence vs Obesity	15
16. Random Forest Importance	16

17. Sugar & Fast Food vs Child Obesity	17
18. Age vs Obesity	18
19. Gender Differences in Obesity	19
20. Animal-based Food vs Obesity	20
21. Min & Max Per Capita Food Consumption	21
22. Years Span vs Obesity	22
23. Boys vs Girls Obesity	23
24. Age vs Child Obesity	24
25. Income Level vs Child Obesity	25
26. Conclusion & Recommendations	25

1. Introduction

This report provides a deep analysis of global obesity patterns across more than 190 countries using data from **Our World in Data**.

The analysis focuses on both **adult and child obesity**, examining dietary intake, physical activity, income level, age structure, and gender differences.

Research Objective: This study aims to evaluate whether commonly cited dietary and lifestyle factors—such as sugar consumption, fast food sales, and physical activity can meaningfully explain cross country differences in obesity prevalence.

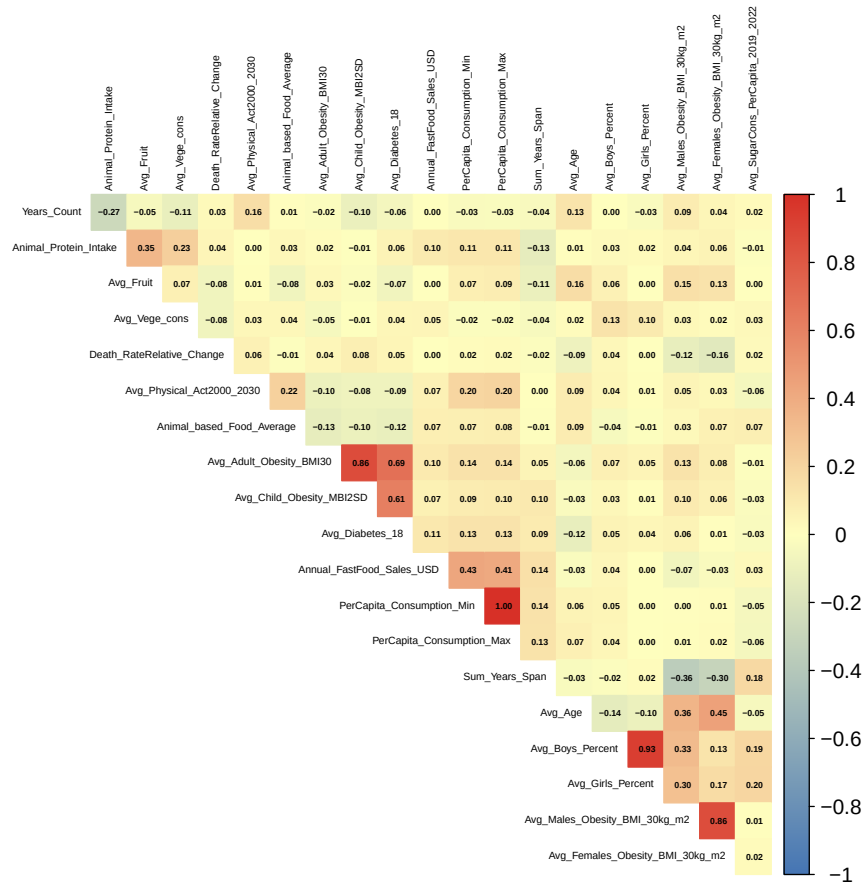
**** Conclusion:****

Results indicate that **no single factor** can account for global obesity variation. Instead, obesity reflects **complex, long-term structural, social, and environmental influences** beyond simple consumption measures.

2. Data Overview

- Data covers over **190 countries**.
- Source: **Our World in Data**.
- Variables include adult and child obesity rates, sugar consumption, fast food sales, animal-based food intake, fruit and vegetable consumption, physical activity, income level, age, gender distribution, and health indicators.

3. Numeric Variables Correlation



Food Types vs Obesity Correlation

Interpretation:

Obesity indicators (adult, child, male, and female obesity rates) show **strong positive correlations**, indicating that obesity prevalence is consistent across age and gender groups.

Fast food sales and **diabetes prevalence** exhibit **moderate positive correlations** with obesity, suggesting a notable association between unhealthy dietary patterns and obesity-related diseases.

Animal protein intake shows a **mild to moderate** relationship with obesity, while **overall animal-based food consumption** remains **weakly correlated**.

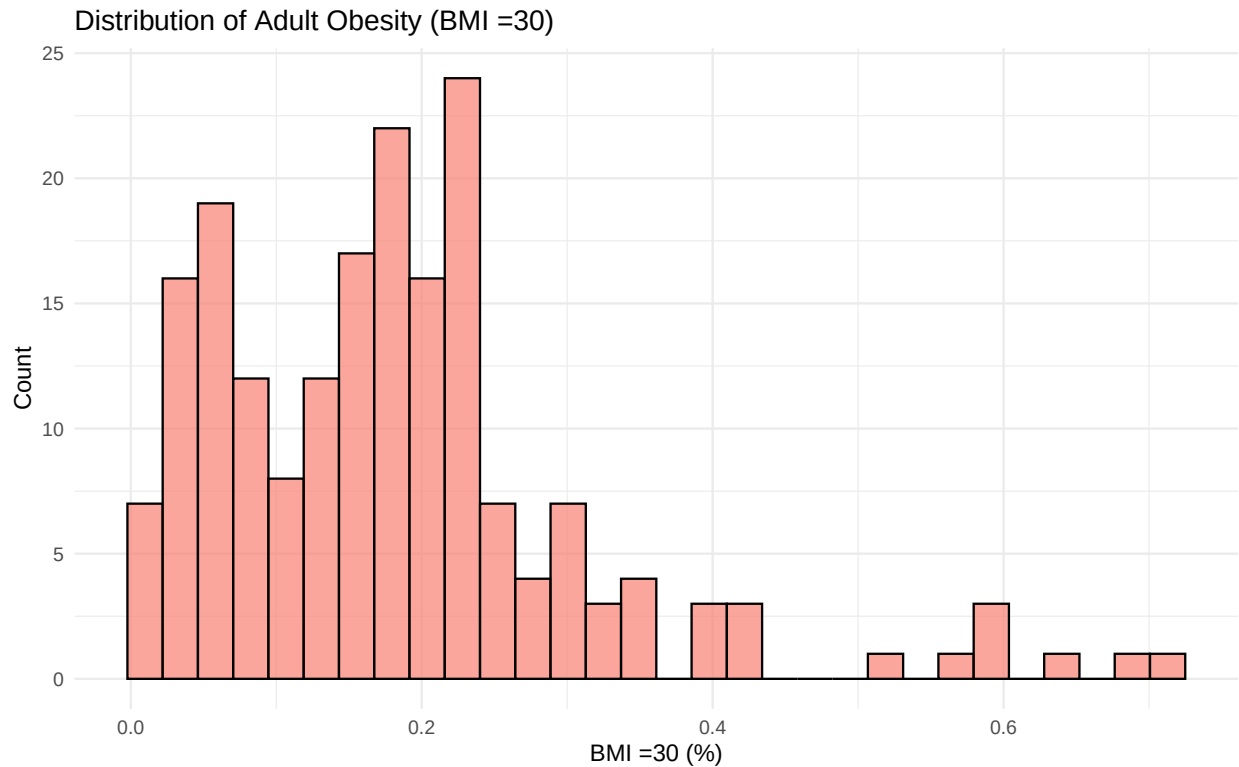
Fruit and vegetable consumption displays **weak or negligible correlations**, implying that these factors alone do not explain cross-country obesity differences.

Physical activity has a **weak negative correlation** with obesity, indicating a limited but protective effect.

Conclusion:

The matrix highlights that structural and lifestyle-related factors are more **strongly associated with obesity** than individual food categories.

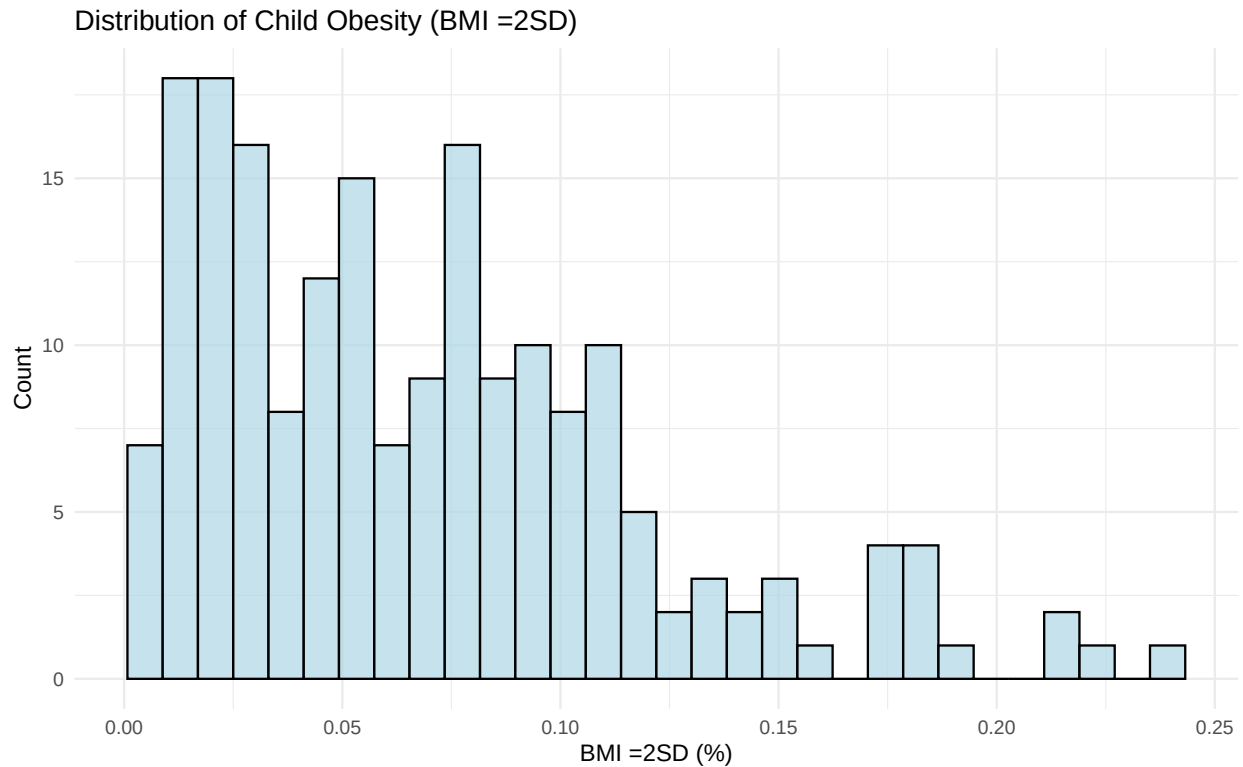
4. Adult Obesity Distribution



Interpretation:

Adult obesity rates follow a **left-skewed distribution**, where a few countries exhibit exceptionally high values, indicating that **extreme obesity prevalence is not widespread** but linked to specific national-level factors.

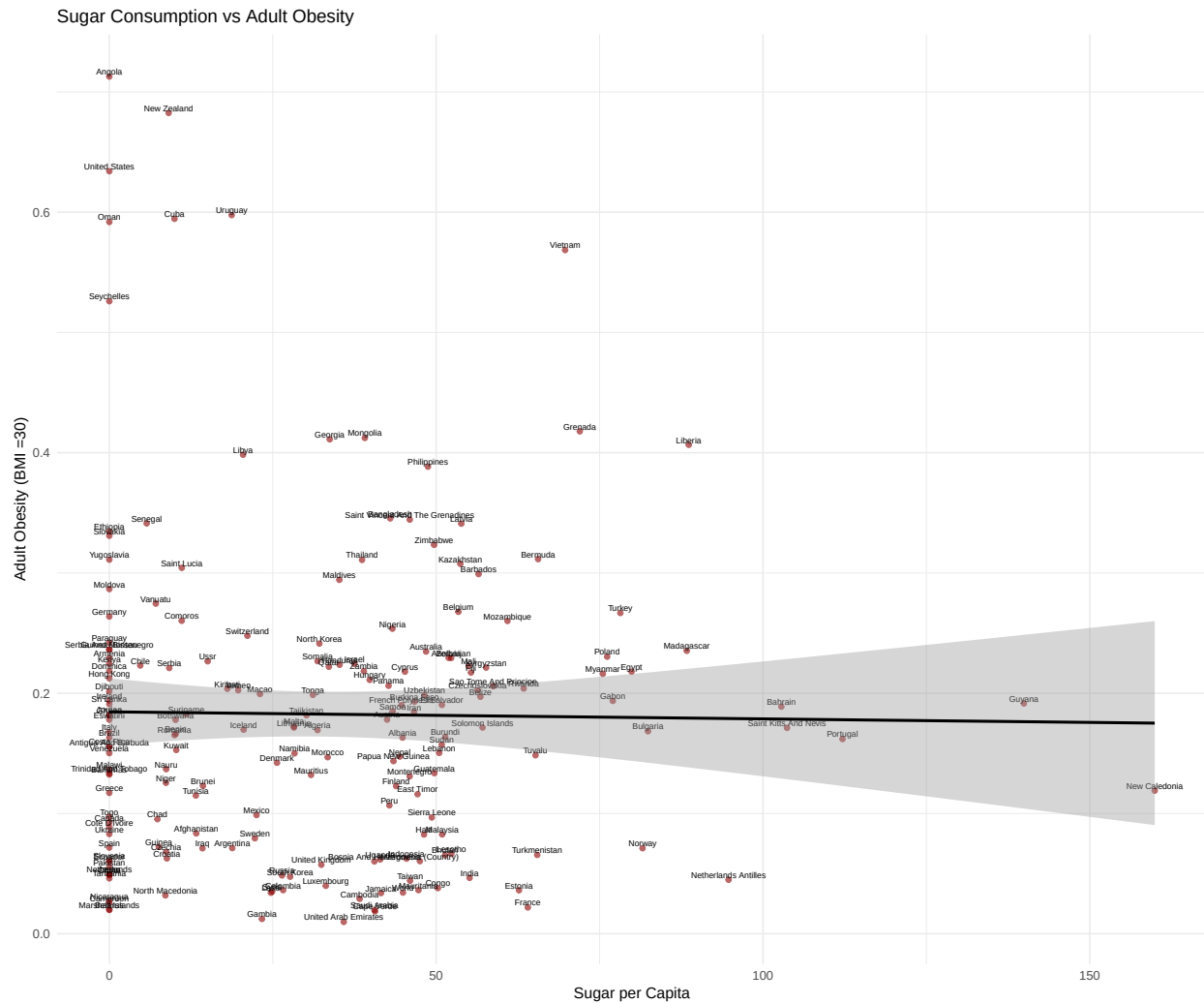
5. Child Obesity Distribution



Interpretation:

Child obesity follows a **left-skewed distribution**, and although it aligns with adult obesity trends in many countries, the observed variation indicates the influence of age-specific lifestyle and environmental factors.

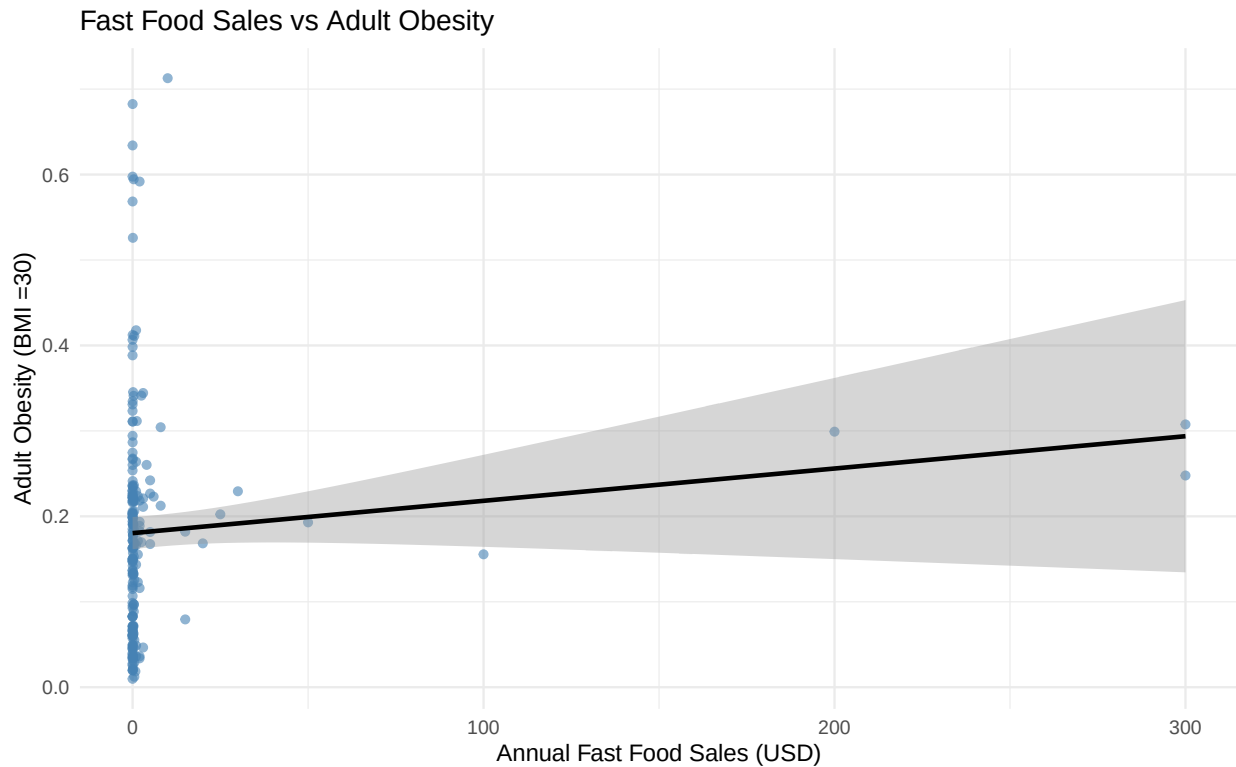
6. Scatter: Sugar Consumption vs Adult Obesity



Interpretation:

Despite a nearly **flat trend line**, the clustering of countries at low sugar consumption levels with highly variable obesity rates suggests that **sugar intake alone** is a **weak predictor** of adult obesity.

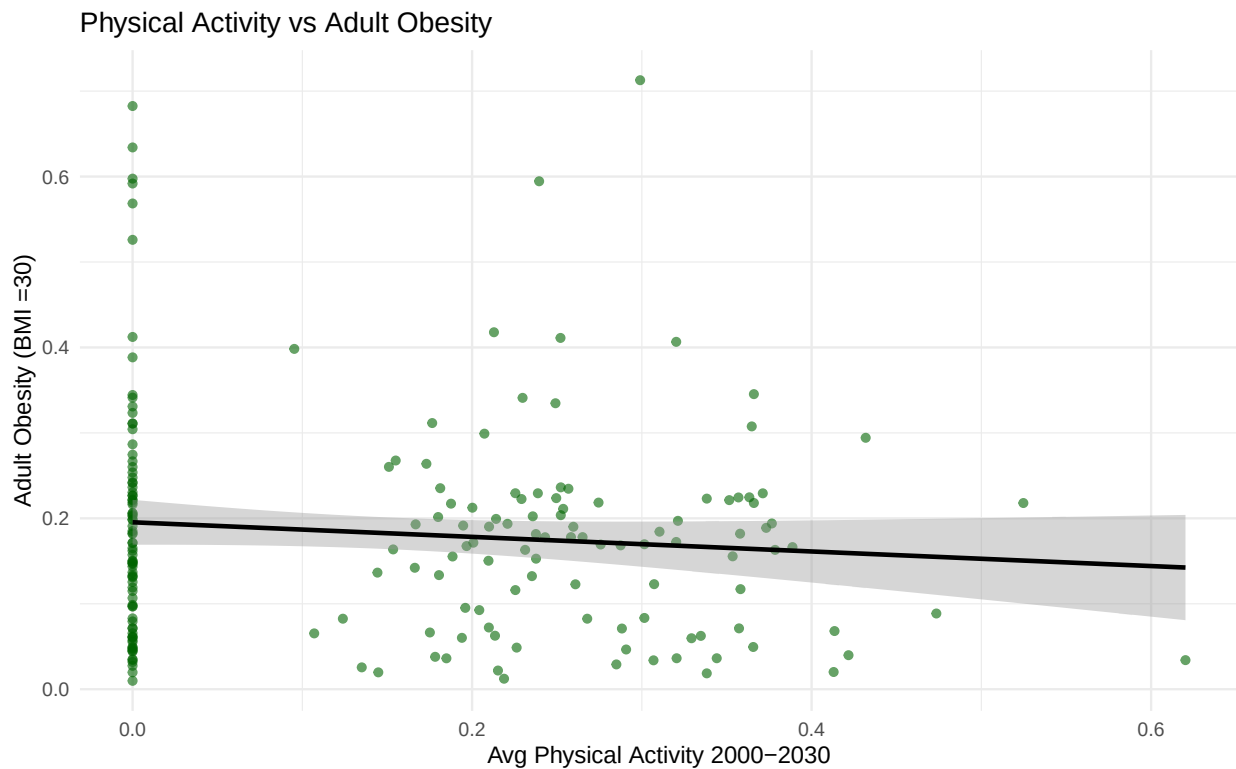
7. Scatter: Fast Food Sales vs Adult Obesity



Interpretation:

This plot reflects the association between fast food sales and adult obesity rates. Data are **right-skewed**, with most observations at low sales levels and few high-sales outliers. A **slight positive trend** is visible, but **fast food consumption alone** does not fully explain obesity variation.

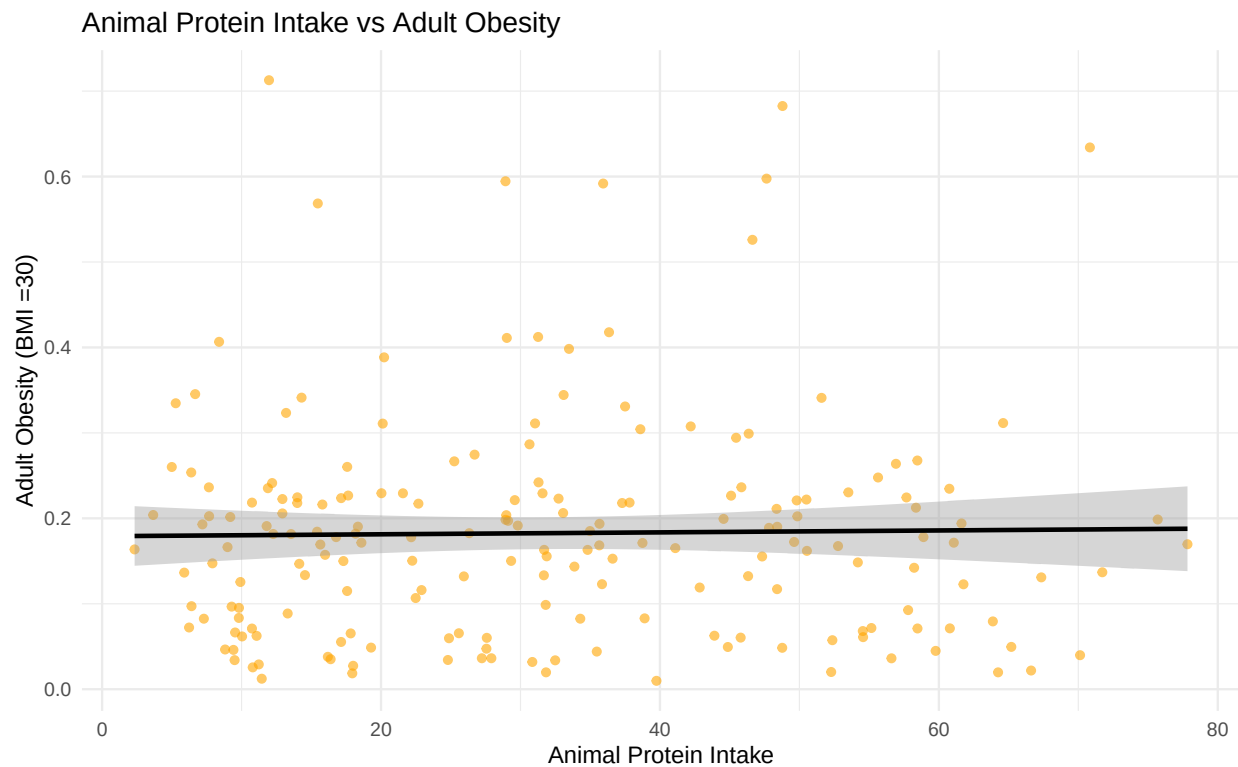
8. Scatter: Physical Activity vs Adult Obesity



Interpretation:

Most countries cluster at **moderate physical activity and obesity levels**, with some showing very low obesity rates. **Higher activity** tends to relate to **lower obesity**, but other factors also play a significant role.

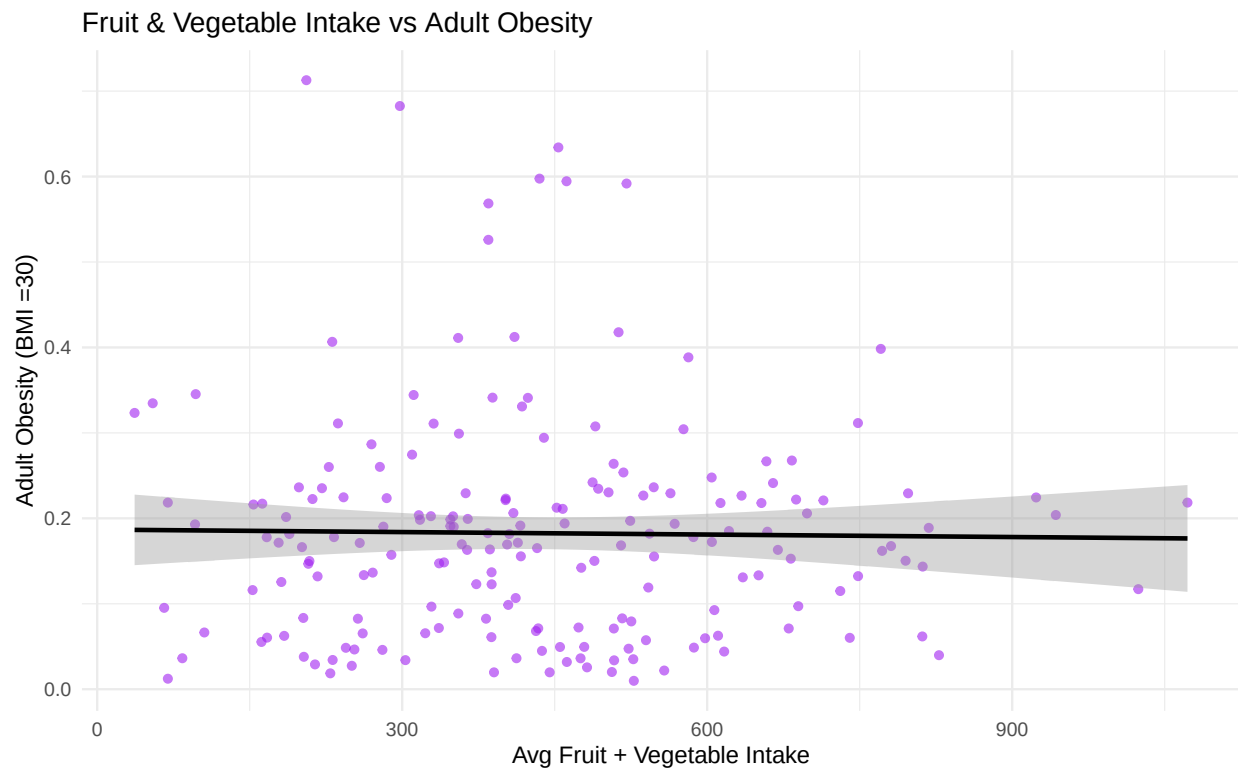
9. Scatter: Animal Protein Intake vs Adult Obesity



Interpretation:

Most countries are evenly distributed around a nearly **straight trend line**, with only a few at the extreme right. This suggests that **protein intake alone** does not strongly determine obesity.

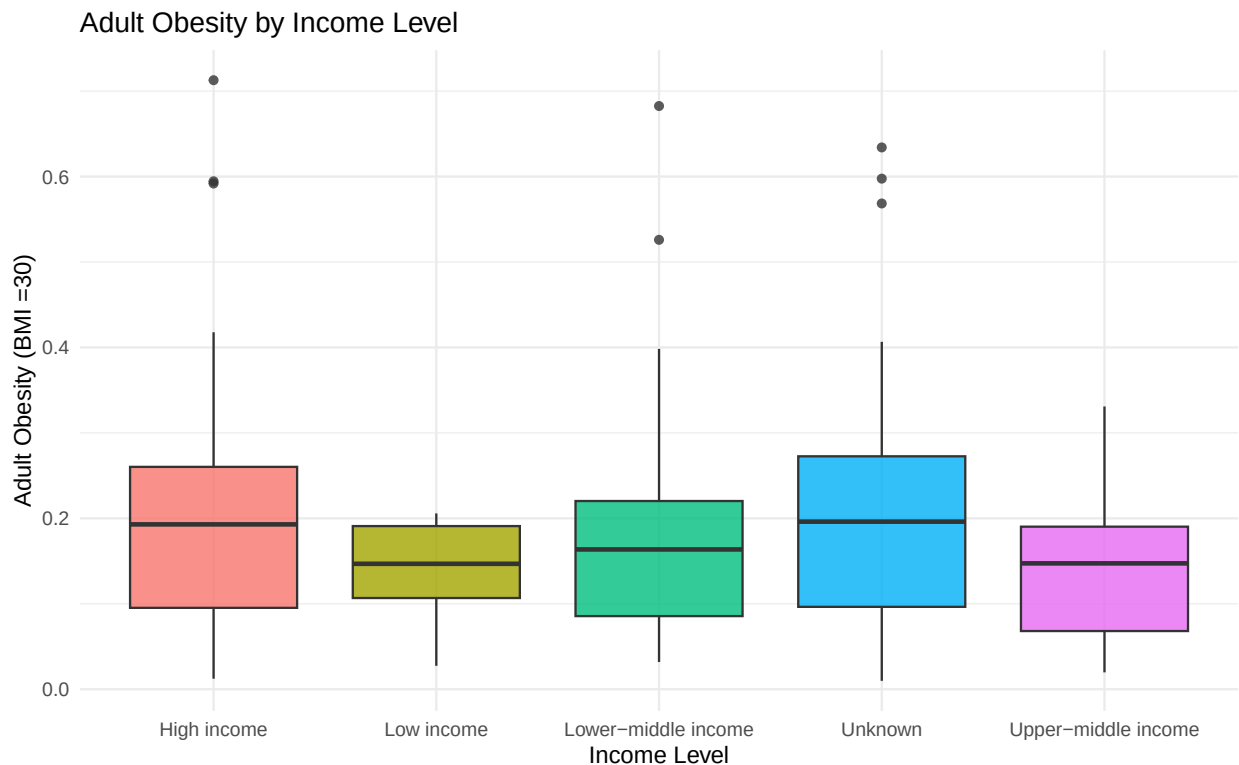
10. Fruit & Vegetable Consumption vs Adult Obesity



Interpretation:

Most countries are evenly distributed around a nearly **straight trend line**, with only a few at the extreme right. This suggests that **fruit and vegetable intake alone** does not strongly determine obesity.

11. Boxplot: Obesity by Income Level



Interpretation:

The **heavy overlap across income groups** shows that income level by itself does not explain obesity differences between countries.

Some **high-income countries appear as extreme cases**, driven by food environments and lifestyles that promote **over consumption rather than by income alone**.

At the same time, **low-income countries can show high obesity** because diets are often calorie-rich but nutritionally poor, reflecting access constraints rather than abundance.

Overall, the figure illustrates that **obesity is shaped less by income and more by how food is produced, priced, and consumed within each society**.

12. Regression: Adult Obesity

```
##  
## Call:  
## lm(formula = Avg_Adult_Obesity_BMI30 ~ Avg_SugarCons_PerCapita_2019_2022 +  
##   Annual_FastFood_Sales_USD + Animal_Protein_Intake + Avg_Fruit +  
##   Avg_Vege_cons + Avg_Physical_Act2000_2030 + Income_Level +  
##   Avg_Age, data = data)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20432 -0.08000 -0.00716  0.05458  0.51249
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.301e-01  3.610e-02   6.375  1.5e-09 ***
## Avg_SugarCons_PerCapita_2019_2022 -7.164e-05  3.270e-04  -0.219   0.8268
## Annual_FastFood_Sales_USD      3.563e-04  2.770e-04   1.286   0.2000
## Animal_Protein_Intake      3.338e-06  5.521e-04   0.006   0.9952
## Avg_Fruit      2.329e-05  7.766e-05   0.300   0.7646
## Avg_Vege_cons     -2.477e-05  6.841e-05  -0.362   0.7177
## Avg_Physical_Act2000_2030     -6.735e-02  6.730e-02  -1.001   0.3183
## Income_LevelLow income     -7.519e-02  4.706e-02  -1.598   0.1119
## Income_LevelLower-middle income -2.783e-02  2.674e-02  -1.041   0.2993
## Income_LevelUnknown      6.702e-03  2.895e-02   0.231   0.8172
## Income_LevelUpper-middle income -6.236e-02  2.557e-02  -2.439   0.0157 *
## Avg_Age      -8.521e-04  1.109e-03  -0.768   0.4435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1295 on 180 degrees of freedom
## Multiple R-squared:  0.07526,    Adjusted R-squared:  0.01875
## F-statistic: 1.332 on 11 and 180 DF,  p-value: 0.21
```

Interpretation:

The model has **very low explanatory power** (Adjusted $R^2 = 0.019$) and is **not statistically significant overall** ($F(11,180) = 1.33$, $p = 0.21$), indicating that the covariates explain little of the variation in obesity rates.

Only **upper-middle income** is **statistically significant** ($\beta = -0.062$, $p = 0.016$), while all dietary, activity, and demographic variables have p -values > 0.20 and near-zero coefficients.

This suggests that cross-country obesity differences are **driven more by structural and contextual factors than by average consumption or behavior measures**.

Conclusion: The regression highlights both the **weak predictive power of macro-level indicators** and the **importance of unobserved social, economic, and environmental mechanisms**.

13. Regression: Child Obesity

```
##
## Call:
## lm(formula = Avg_Child_Obesity_MBI2SD ~ Avg_SugarCons_PerCapita_2019_2022 +
##      Annual_FastFood_Sales_USD + Animal_Protein_Intake + Avg_Fruit +
##      Avg_Vege_cons + Avg_Physical_Act2000_2030 + Income_Level +
```

```

##      Avg_Age, data = data)
##
## Residuals:
##      Min        1Q      Median        3Q        Max
## -0.071749 -0.038169 -0.007224  0.030161  0.158751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.858e-02  1.403e-02   6.313 2.07e-09 ***
## Avg_SugarCons_PerCapita_2019_2022 -4.862e-05  1.271e-04  -0.382  0.7026
## Annual_FastFood_Sales_USD      8.067e-05  1.077e-04   0.749  0.4546
## Animal_Protein_Intake    -2.712e-05  2.146e-04  -0.126  0.8996
## Avg_Fruit    -1.062e-05  3.019e-05  -0.352  0.7254
## Avg_Vege_cons      5.338e-06  2.659e-05   0.201  0.8411
## Avg_Physical_Act2000_2030    -2.874e-02  2.616e-02  -1.099  0.2734
## Income_LevelLow income    -2.866e-02  1.829e-02  -1.567  0.1189
## Income_LevelLower-middle income -1.319e-02  1.039e-02  -1.269  0.2062
## Income_LevelUnknown    -4.189e-03  1.125e-02  -0.372  0.7102
## Income_LevelUpper-middle income -2.067e-02  9.939e-03  -2.080  0.0389 *
## Avg_Age    -1.382e-04  4.312e-04  -0.320  0.7490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05032 on 180 degrees of freedom
## Multiple R-squared:  0.04735,    Adjusted R-squared:  -0.01087
## F-statistic: 0.8133 on 11 and 180 DF,  p-value: 0.6268

```

Interpretation:

The model explains virtually **none of the variation in child obesity** (Adjusted $R^2 = -0.011$) and is **not statistically significant** overall ($F(11,180) = 0.81$, $p = 0.63$).

Only **upper-middle income** is **statistically significant** ($\beta = -0.021$, $p = 0.039$), while all other predictors have p-values above 0.20 and very small coefficients.

This indicates that **average dietary intake, physical activity, and demographics at the country level** have **limited power** to explain cross-country differences in child obesity.

Conclusion: Child obesity appears to be driven primarily by factors not captured in this model, such as **household behavior, urban environment, health systems, and early-life conditions**.

14. Death Rate Change vs Adult Obesity



Interpretation:

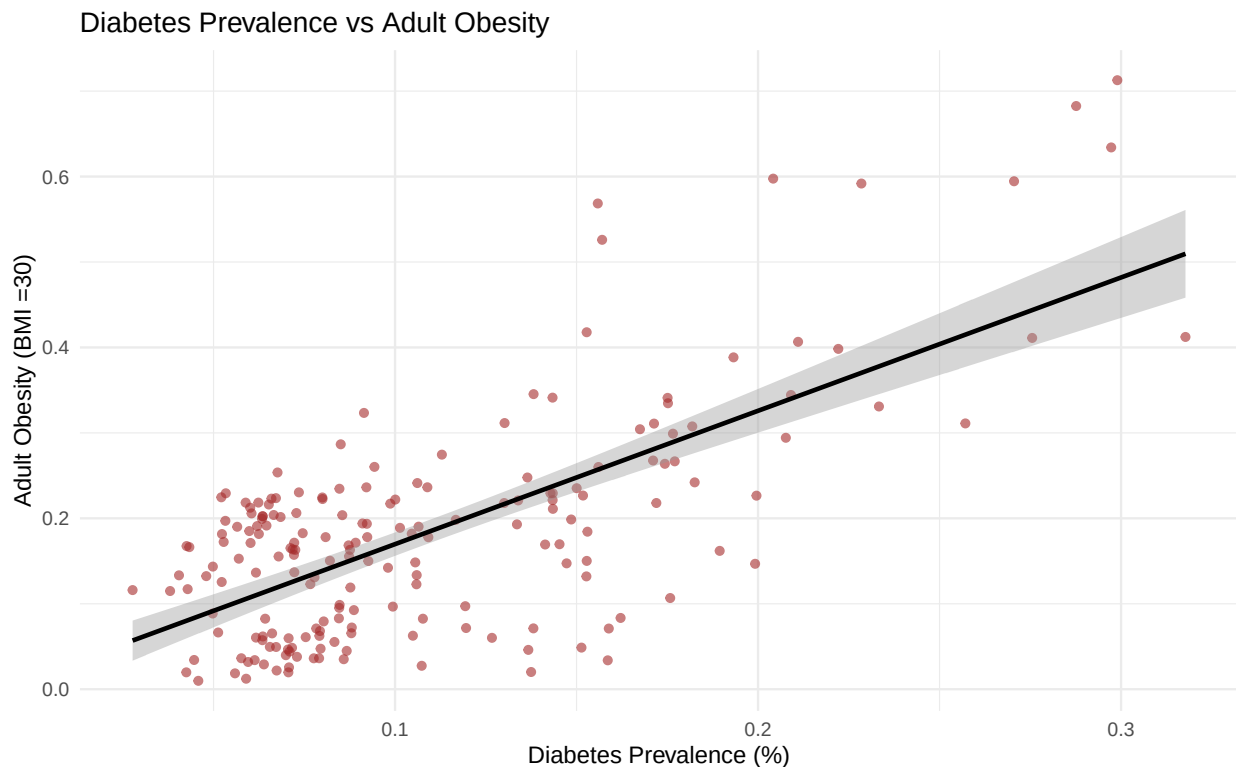
The **near-flat regression line** and wide dispersion indicate **no meaningful linear relationship** between changes in **mortality and adult obesity** across countries.

This suggests that **short-term changes in death rates** are driven by **acute factors** (healthcare shocks, epidemics, conflicts, policy changes) rather than by chronic conditions like obesity.

The **few high-obesity or high-mortality observations** behave as **context-specific outliers**, not as part of a general pattern.

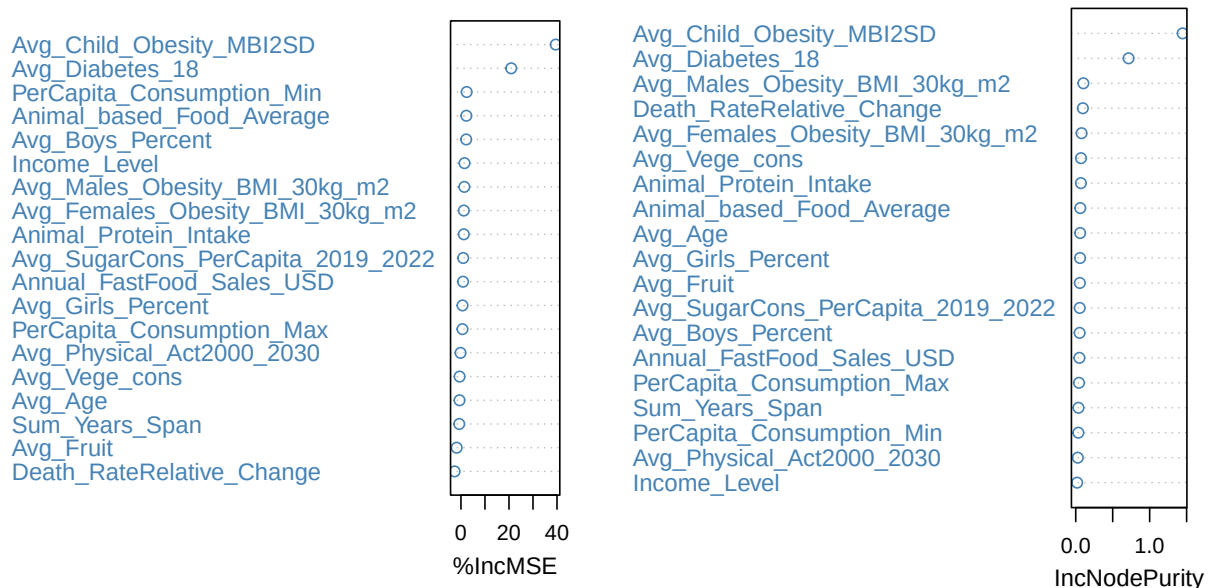
Conclusion: Adult obesity is a **long-run structural risk factor**, while changes in **death rates reflect short-run dynamics** making a direct cross-sectional relationship between them inherently weak.

15. Diabetes Prevalence vs Obesity



16. Random Forest Importance

Variable Importance – Random Forest



Interpretation:

This plot shows which variables the model relied on most to predict obesity.

%IncMSE (left): Shows how much including a variable improves prediction accuracy.

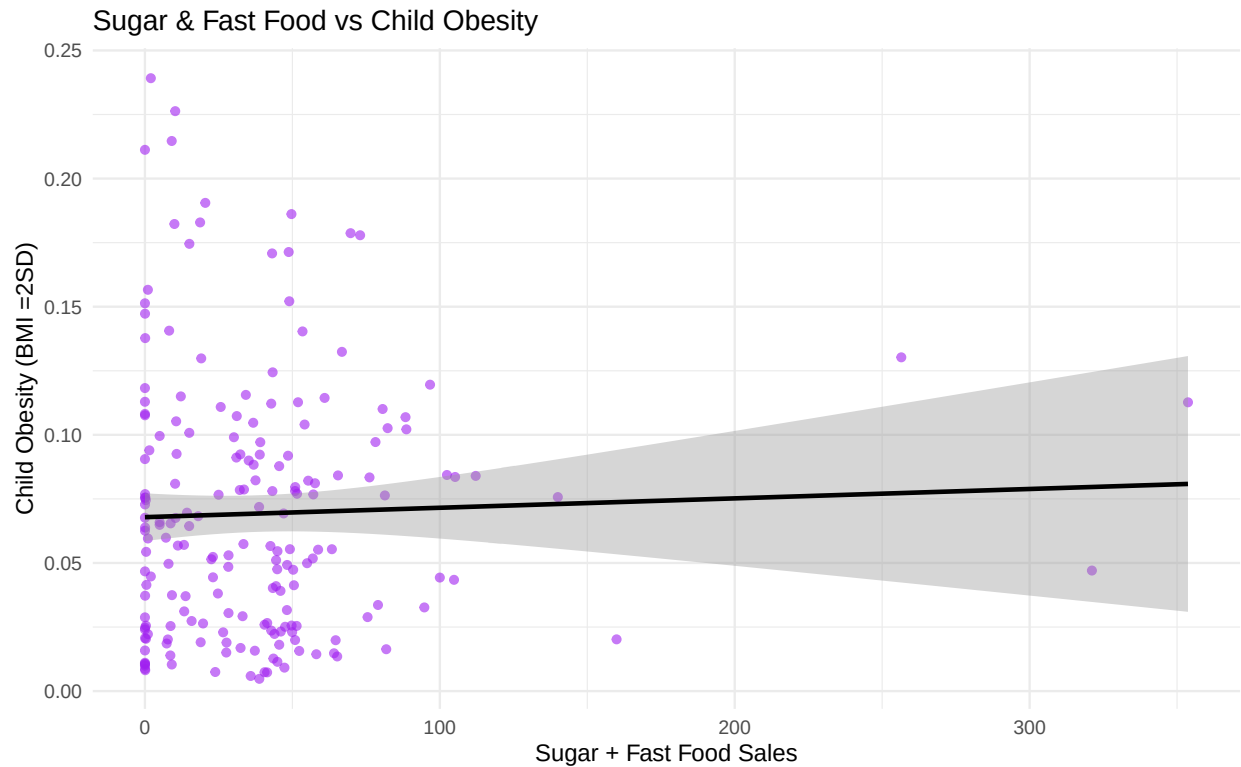
IncNodePurity (right): Shows how well a variable helps the model separate the data into clear groups.

Top predictors include Avg_Child_Obesity_MBI2SD, Avg_Diabetes_18, and Avg_Males_Obesity_BMI_30kg_m2.

Variables like **Income_Level** or **Physical Activity** appear **lower**, likely due to overlap with other variables.

Conclusion: This reflects the model's perspective on importance, not necessarily the real-world effect of each variable.

17. Sugar & Fast Food vs Child Obesity



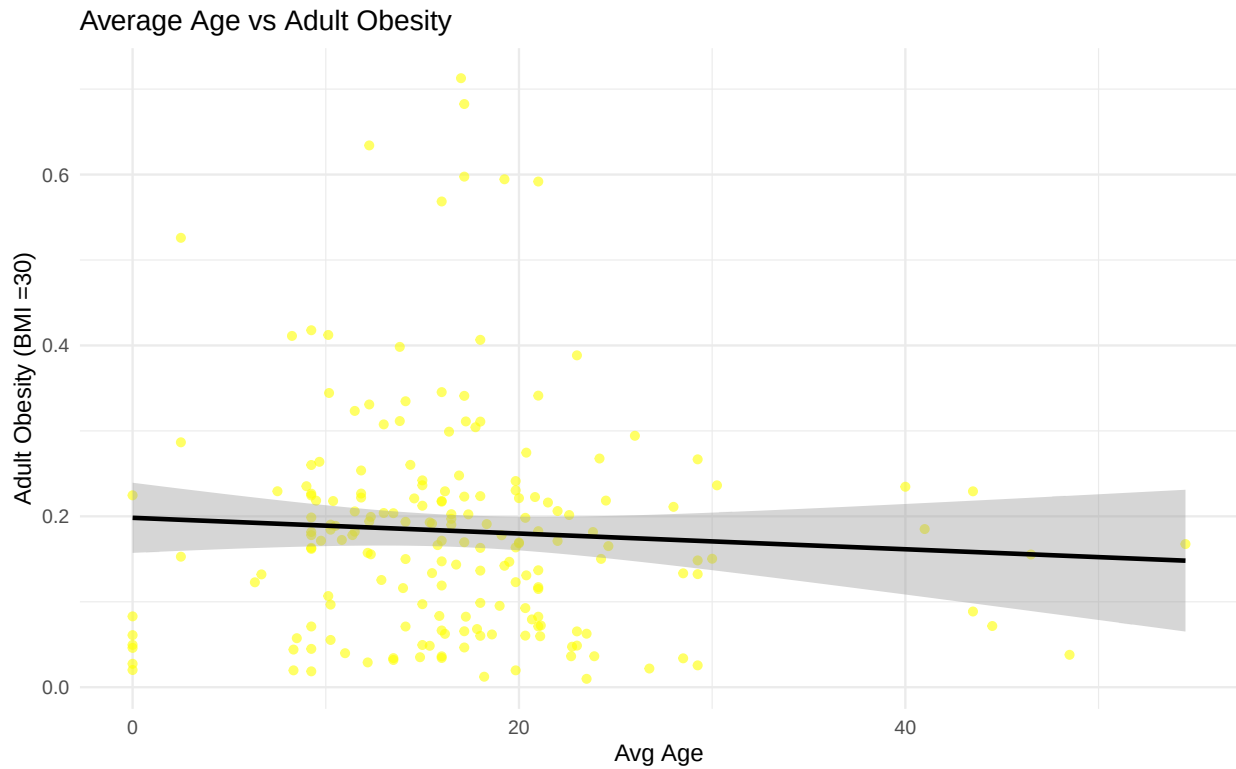
Interpretation:

Most countries have **low sugar and fast food consumption**, and the **trend line is nearly flat**.

This shows a **very weak link between sugar/fast food intake and child obesity**.

Conclusion: Other factors are likely more important in explaining child obesity.

18. Age vs Obesity



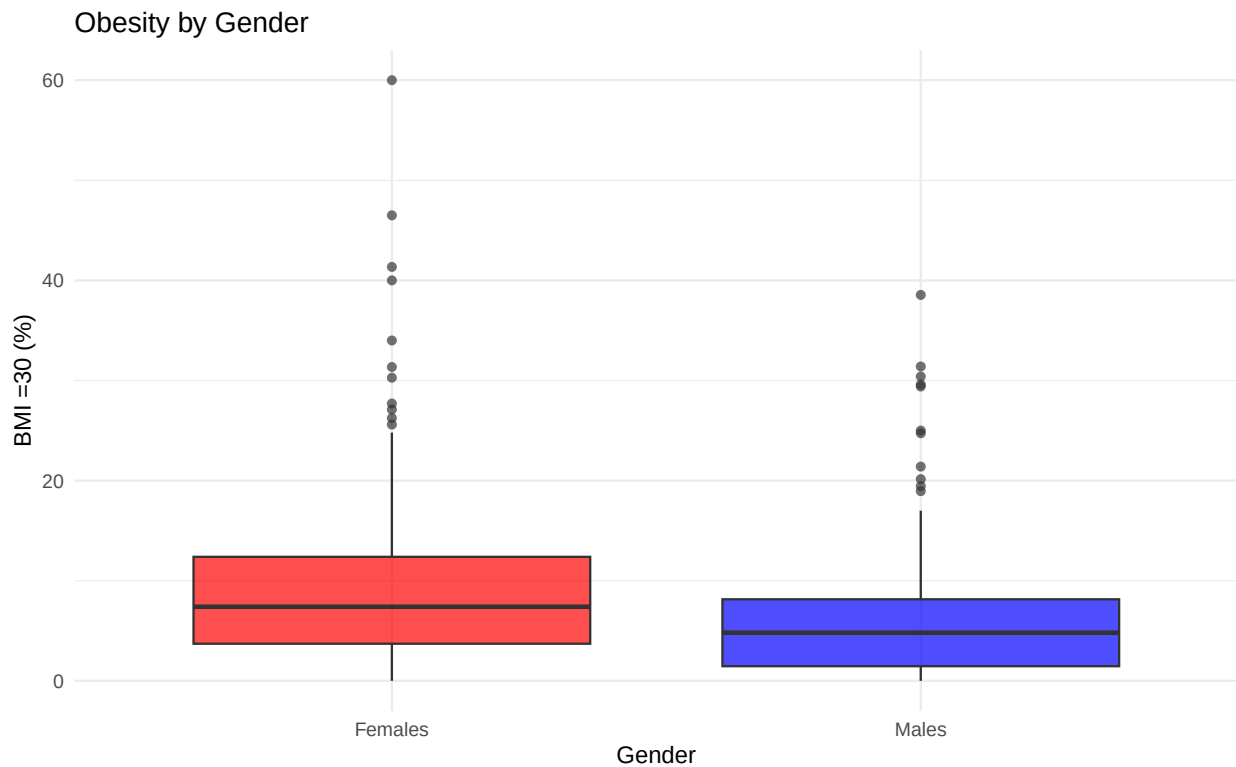
Interpretation:

Countries cluster away from the very low ages, and **the trend line slopes slightly downward.**

This shows a **weak decrease in obesity with higher average age.**

Conclusion: Age alone does not strongly explain obesity differences across countries

19. Gender Differences in Obesity



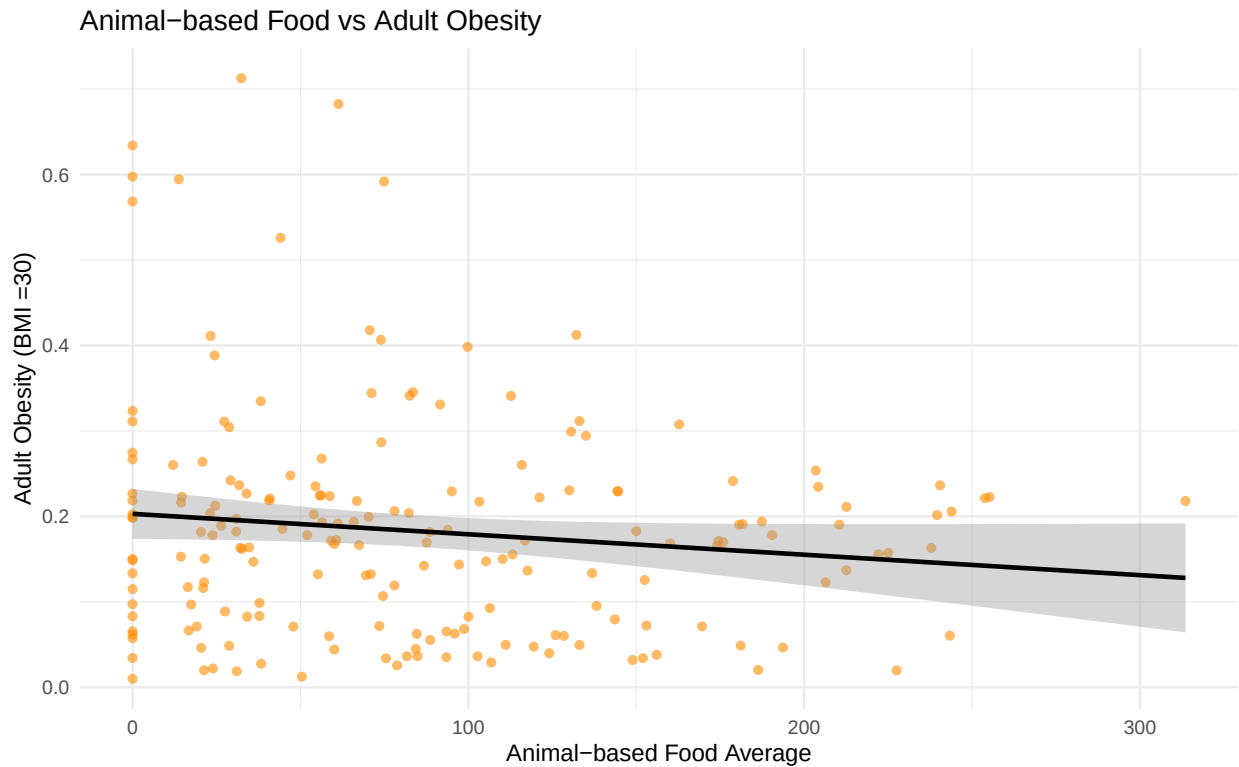
Interpretation:

Males have generally **lower obesity**, and their range (**whiskers**) is **shorter**, showing **fewer extreme values**.

Females tend to have **higher and more variable obesity rates**, suggesting gender differences that could be important for targeted interventions.

Conclusion: Females show **higher** and more variable obesity, possibly due to **differences in hormones, metabolism, or lifestyle factors**.

20. Animal-based Food vs Obesity



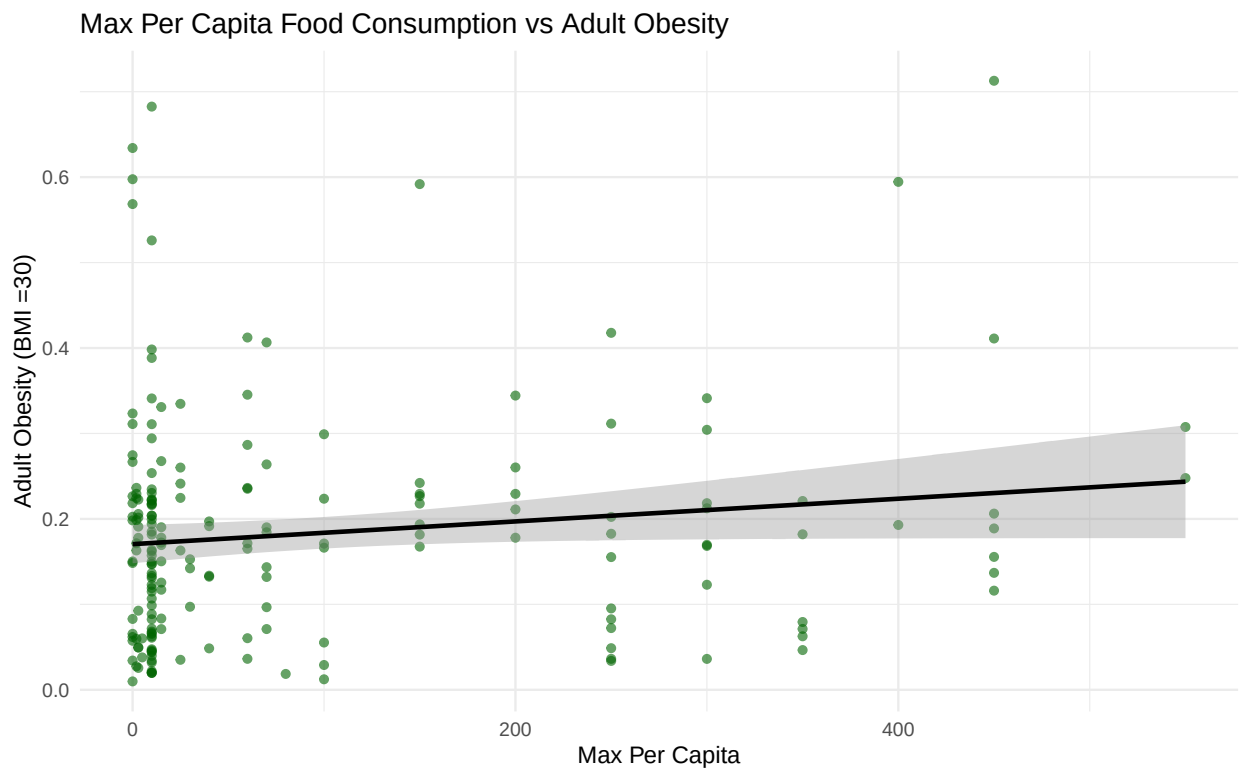
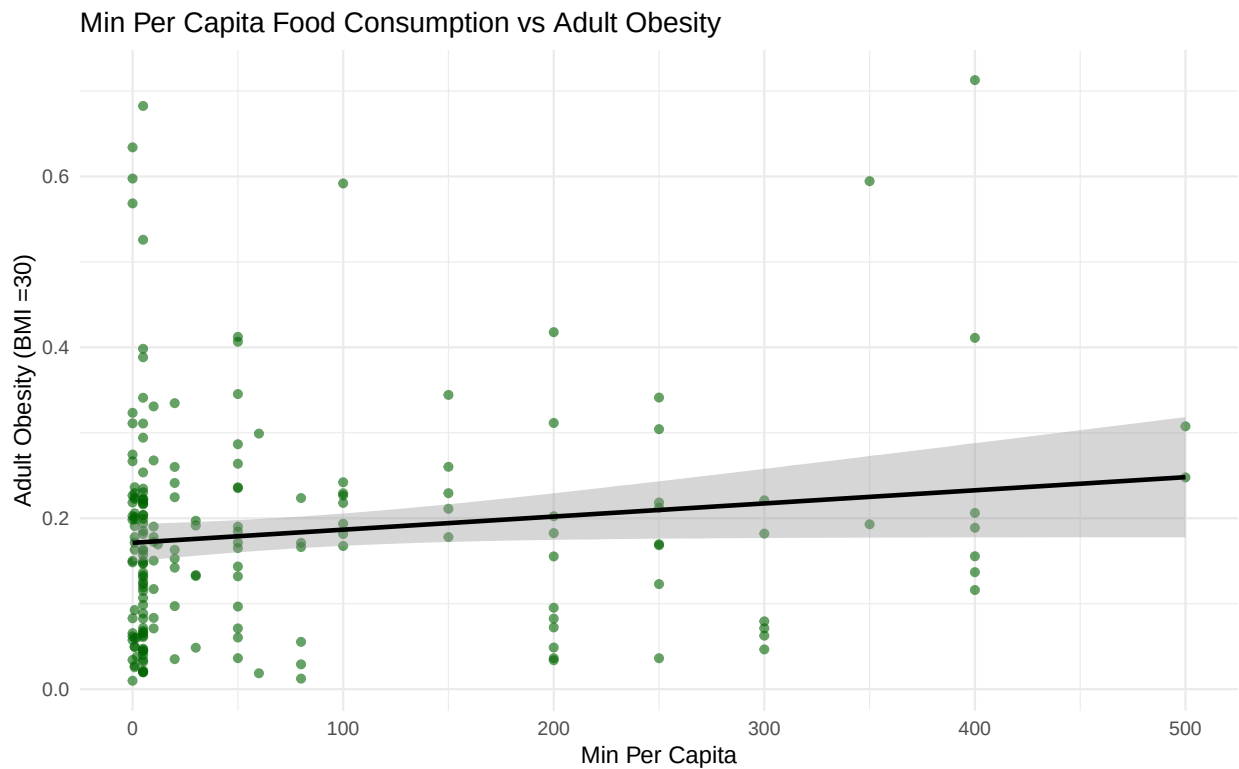
Interpretation:

Most countries cluster on the low consumption side, slightly below the trend line, while fewer countries above the line are more spread out.

The **trend line slopes slightly downward**, indicating a **very weak inverse relationship** between animal-based food consumption and obesity.

Conclusion: Animal-based food alone is **not a strong driver of adult obesity**, and other factors likely play a bigger role.

21. Min & Max Per Capita Food Consumption



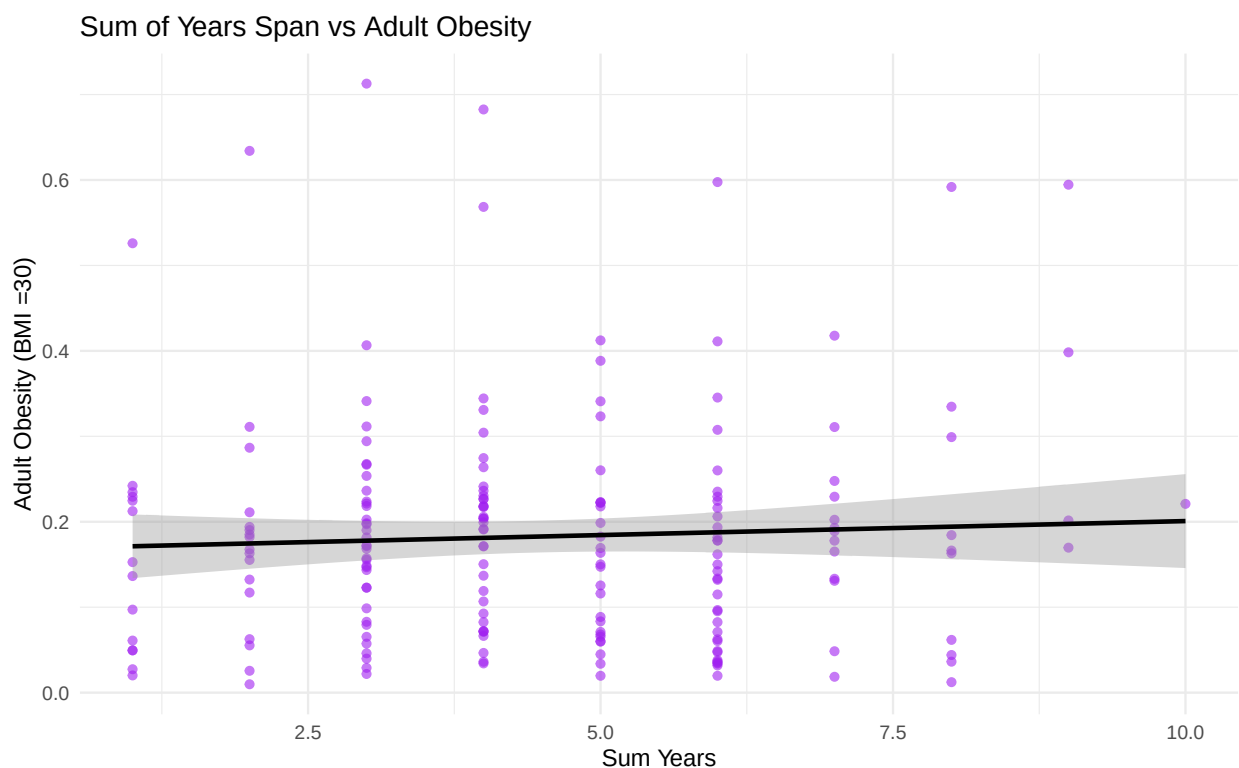
Interpretation:

Most countries are clustered at **lower consumption and lower obesity**, while fewer countries with higher consumption show slightly higher obesity.

The upward trend is weak, showing that **food consumption alone** does not fully explain obesity differences.

Conclusion: Higher per capita consumption may contribute to obesity, but other factors are likely more important.

22. Years Span vs Obesity



Interpretation:

The regression line shows a **very slight positive slope**, which indicates a **weak positive association** between the number of years included in the analysis and the estimated adult obesity rate. This suggests that **extending the time span does not materially change the overall obesity level**.

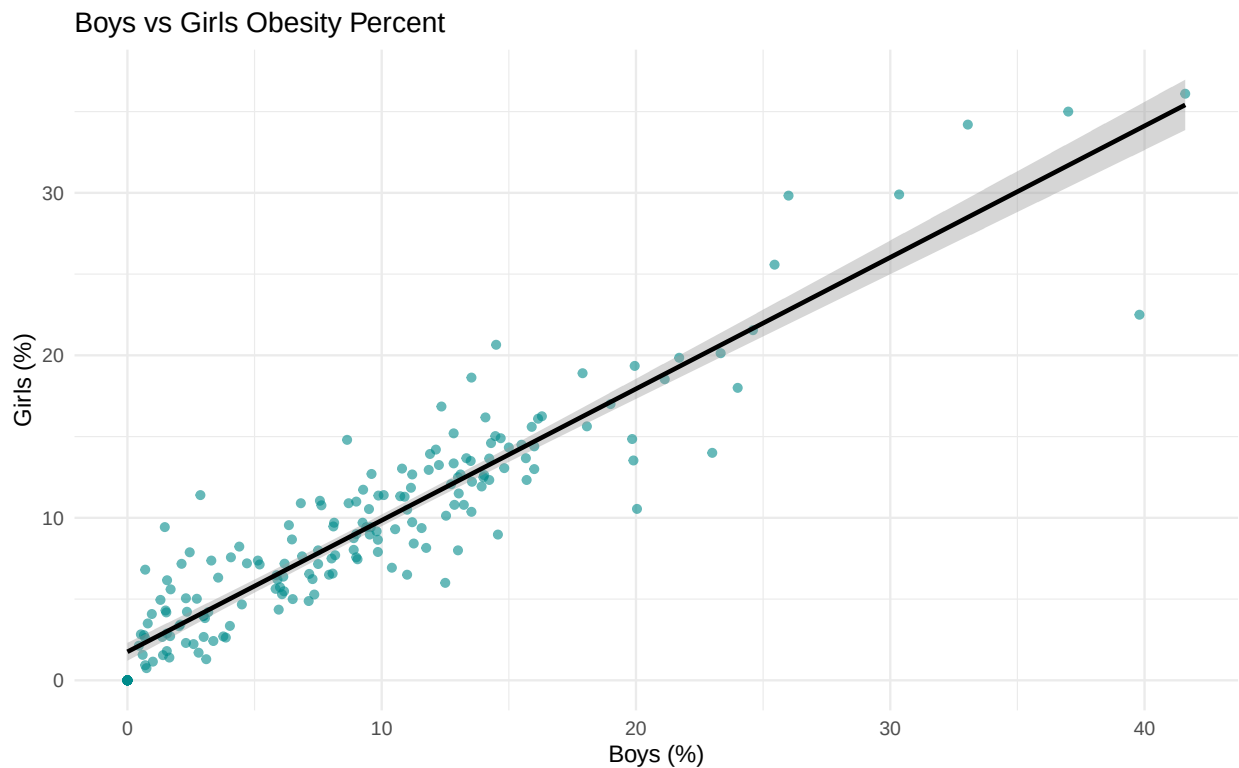
The wide dispersion of points around the fitted line reflects **substantial variability across countries and years**, meaning that differences in obesity are driven more by structural or regional factors than by the length of the time window itself.

The confidence band around the regression line is relatively wide compared to the slope, **reinforcing that the observed relationship is statistically weak** and should not be interpreted as a meaningful temporal effect.

Conclusion:

Overall, adult obesity estimates are stable across different time spans, which indicates that the results are robust to the choice of temporal aggregation and that the analysis is not biased by the selected period length.

23. Boys vs Girls Obesity



Interpretation:

The near-linear pattern indicates that obesity risk **scales similarly across genders**.

The regression slope shows **girls' obesity increases** at a **slightly lower rate** than boys as overall obesity rises.

Gender differences are minimal at low obesity levels but widen modestly in high-obesity regions. The narrow confidence interval across most of the range suggests **strong model stability and low estimation risk**.

Increased dispersion at higher obesity levels points to **localized environmental or policy factors amplifying outcomes**.

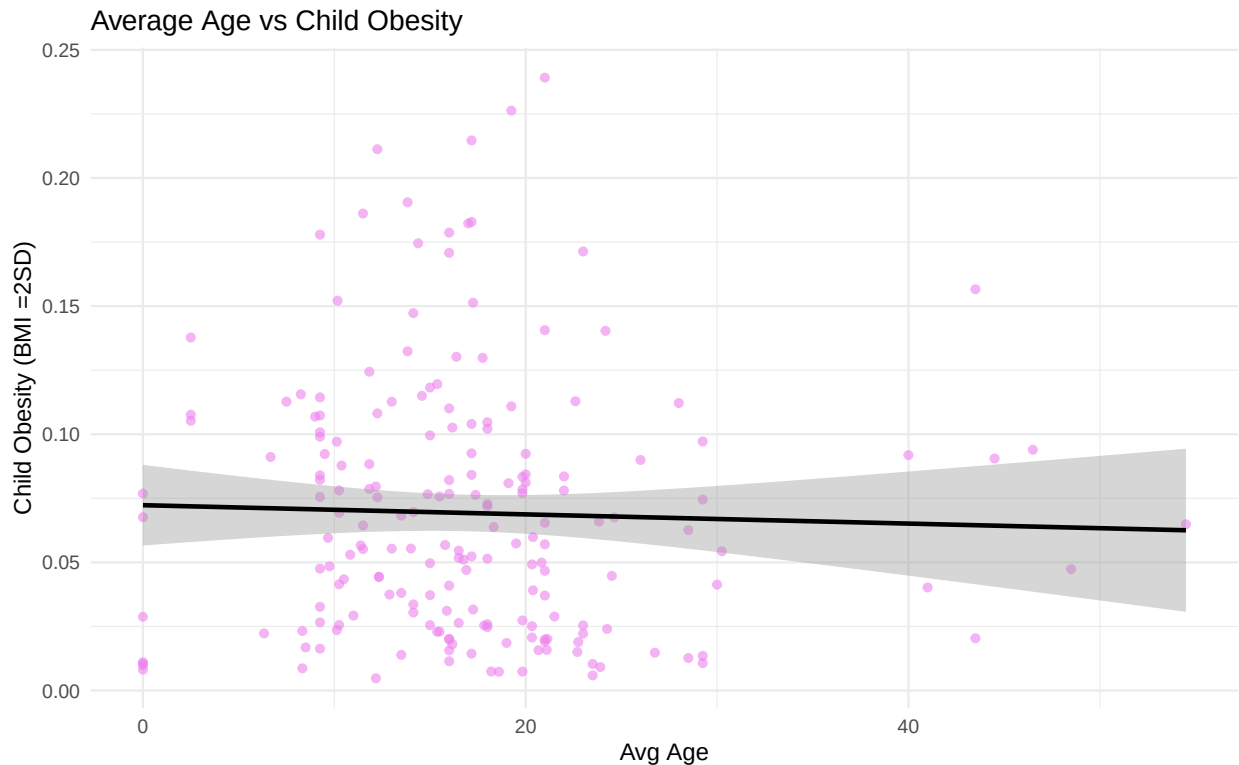
Implication:

Obesity should be addressed as a **system-level issue**, with intervention intensity adjusted by prevalence rather than gender alone.

Gender-specific strategies may refine impact, but population-wide policies will drive the largest gains.

This insight is based on **trend behavior, slope interpretation, and variance analysis** using R-based statistical modeling.

24. Age vs Child Obesity



Interpretation:

The trend indicates a **weak negative relationship** between average age and child obesity.

Obesity rates show **high variability**, especially at lower average ages.

The widening confidence interval at higher ages suggests **limited data support** in that range.

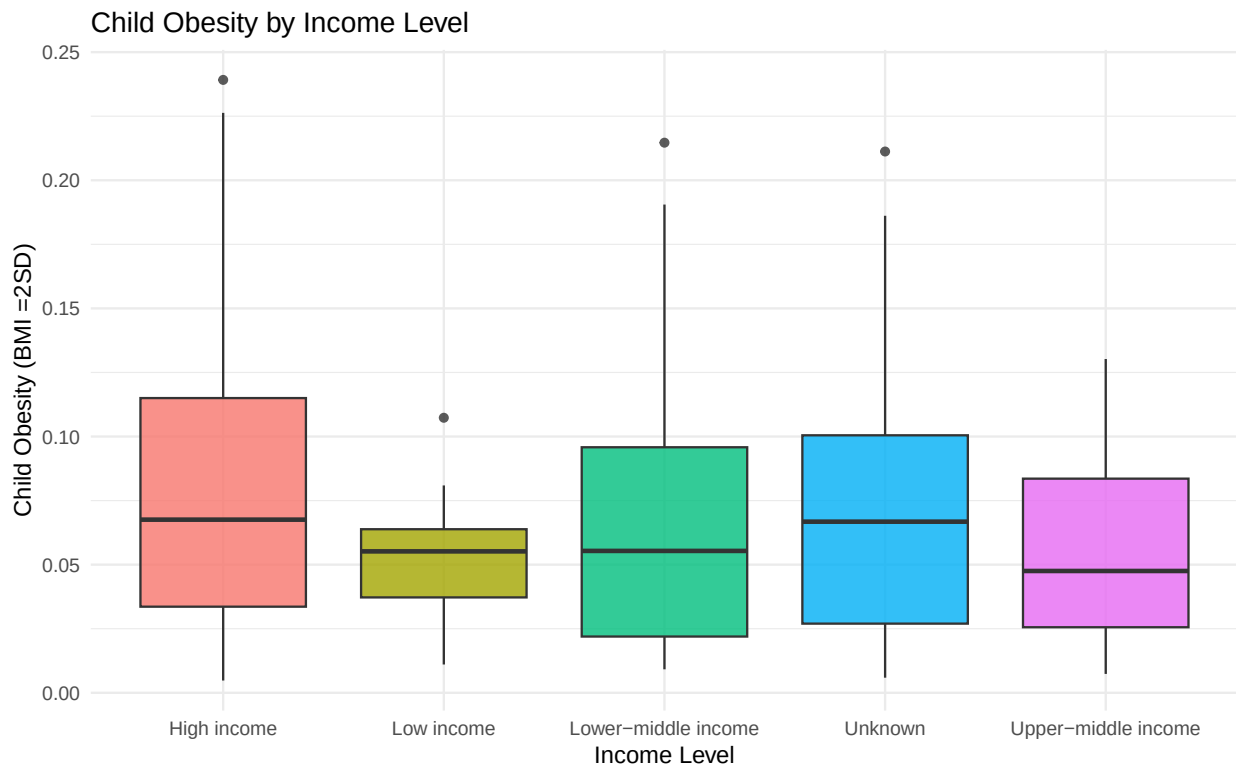
Overall, average age is **not a strong driver** of child obesity outcomes.

Implication:

Child obesity is influenced more by **behavioral and environmental factors** than population age structure.

Policy focus should prioritize **modifiable risk factors** over demographic characteristics.

25. Income Level vs Child Obesity



Interpretation:

Median child obesity rates are **higher in high-income and unknown-income groups** than in low-income groups.

Low-income countries show a **narrower spread**, indicating more uniform outcomes.

High-income and middle-income groups exhibit **greater variability**, suggesting uneven exposure to risk factors.

The presence of high outliers across income levels indicates that **income alone does not fully explain obesity risk**.

Implication:

Child obesity is **not strictly inversely related to income level**.

Structural and behavioral factors likely outweigh income classification alone.

Policy responses should be **context-specific**, not income-tier driven.

26. Conclusion & Recommendations

Obesity is a multifaceted issue shaped by systemic factors, not just individual food choices or income levels. Key health risks like diabetes rise sharply in high-obesity populations, highlighting the urgency for targeted

action.

Recommendations:

- Avoid one-size-fits-all solutions focusing on single foods or nutrients.
- Improve food environments: availability, affordability, marketing, and urban planning.
- Focus on high-prevalence areas with stronger, tailored interventions.
- Address early-life and household factors, especially for child obesity.
- Use data-driven approaches to identify priority regions and groups.

Key message:

Obesity is driven by how societies influence lifestyle choices over time. Effective strategies must tackle these broader systems not just diets to create lasting change.

Data source: Our World in Data