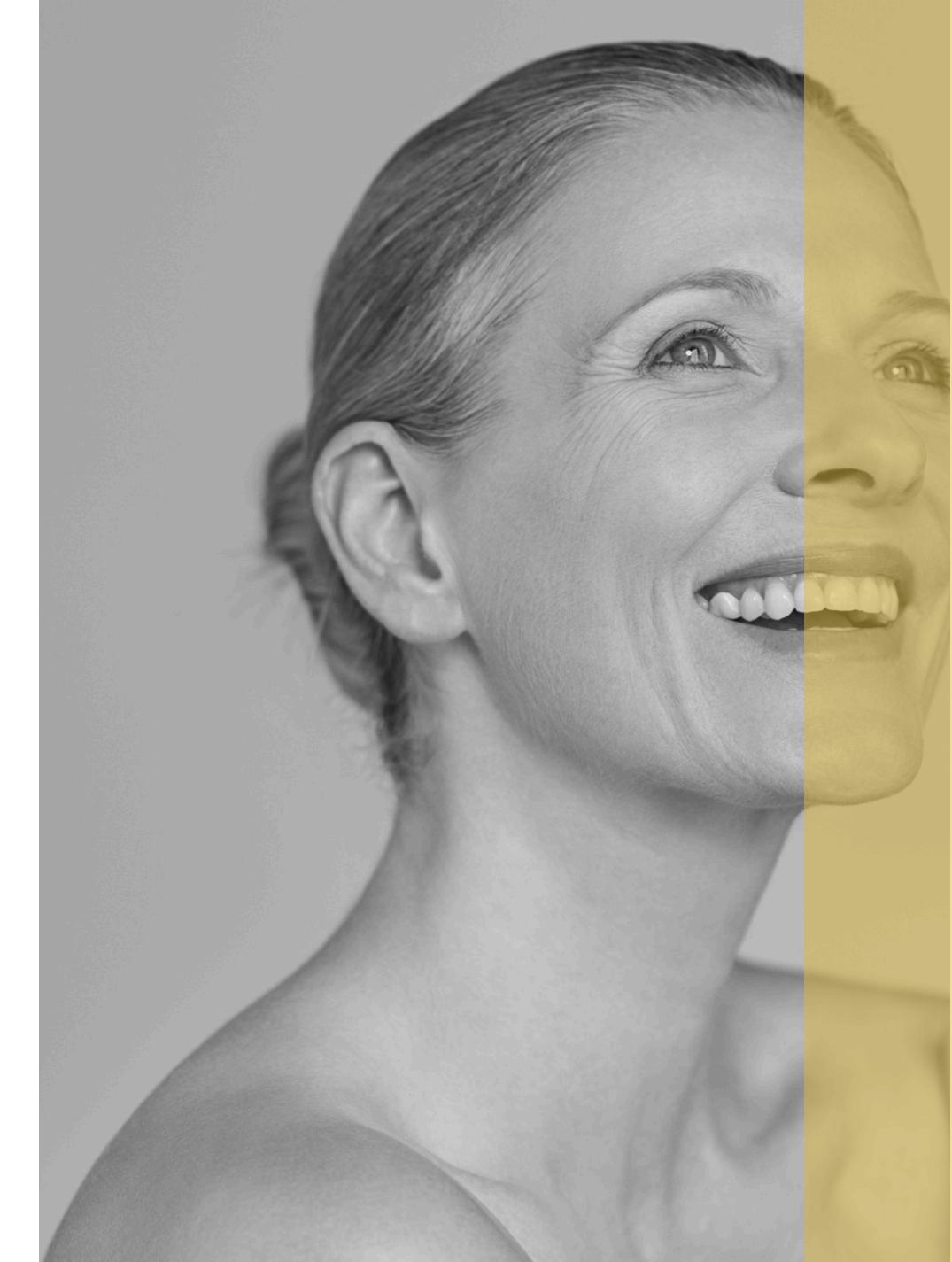


BY GIANG NGUYEN

# Understanding *Happiness*

Happiness Level of Big Cities  
Data Analysis



# Data Understanding

```
> str(city)
'data.frame': 44 obs. of 12 variables:
 $ City                  : chr "Amsterdam" "Sydney" "Vienna" "Stockholm"
 ...
 $ Rank                 : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Sunshine.hours.City.  : chr "1858" "2636" "1884" "1821" ...
 $ Cost.of.a.bottle.of.water.City. : chr "£1.92" "£1.48" "£1.94" "£1.72" ...
 $ Obesity.levels.Country.   : chr "20.40%" "29.00%" "20.10%" "20.60%" ...
 $ Life.expectancy.years...Country. : num 81.2 82.1 81 81.8 79.8 80.4 83.2 80.6 82.
2 81.7 ...
$ Pollution.Index.score...City. : chr "30.93" "26.86" "17.33" "19.63" ...
$ Annual.avg..hours.worked    : chr "1434" "1712" "1501" "1452" ...
$ Happiness.levels.Country.  : num 7.44 7.22 7.29 7.35 7.64 7.8 5.87 7.07 6.
4 7.23 ...
$ Outdoor.activities.City.    : int 422 406 132 129 154 113 35 254 585 218
...
$ Number.of.take.out.places.City. : int 1048 1103 1008 598 523 309 539 1729 2344
788 ...
$ Cost.of.a.monthly.gym.membership.City.: chr "£34.90" "£41.66" "£25.74" "£37.31" ...
```

With the dataset containing information about data **representing a healthy lifestyle of 44 cities**, it is divided into 2 categories:

- **Variables affecting mental health:** Annual avg. hours worked, Number of take out places, Sunshine hours
- **Variables affecting physical health:** Cost of Bottle water, Pollution, Outdoor activities, Gym membership

# Objective

I am interested in **the Happiness Level** variable here, whether it is influenced more by variables that affect mental or physical health, so the target question is **which variables will have an influence on the Happiness Level of a city.**

## My assumption:

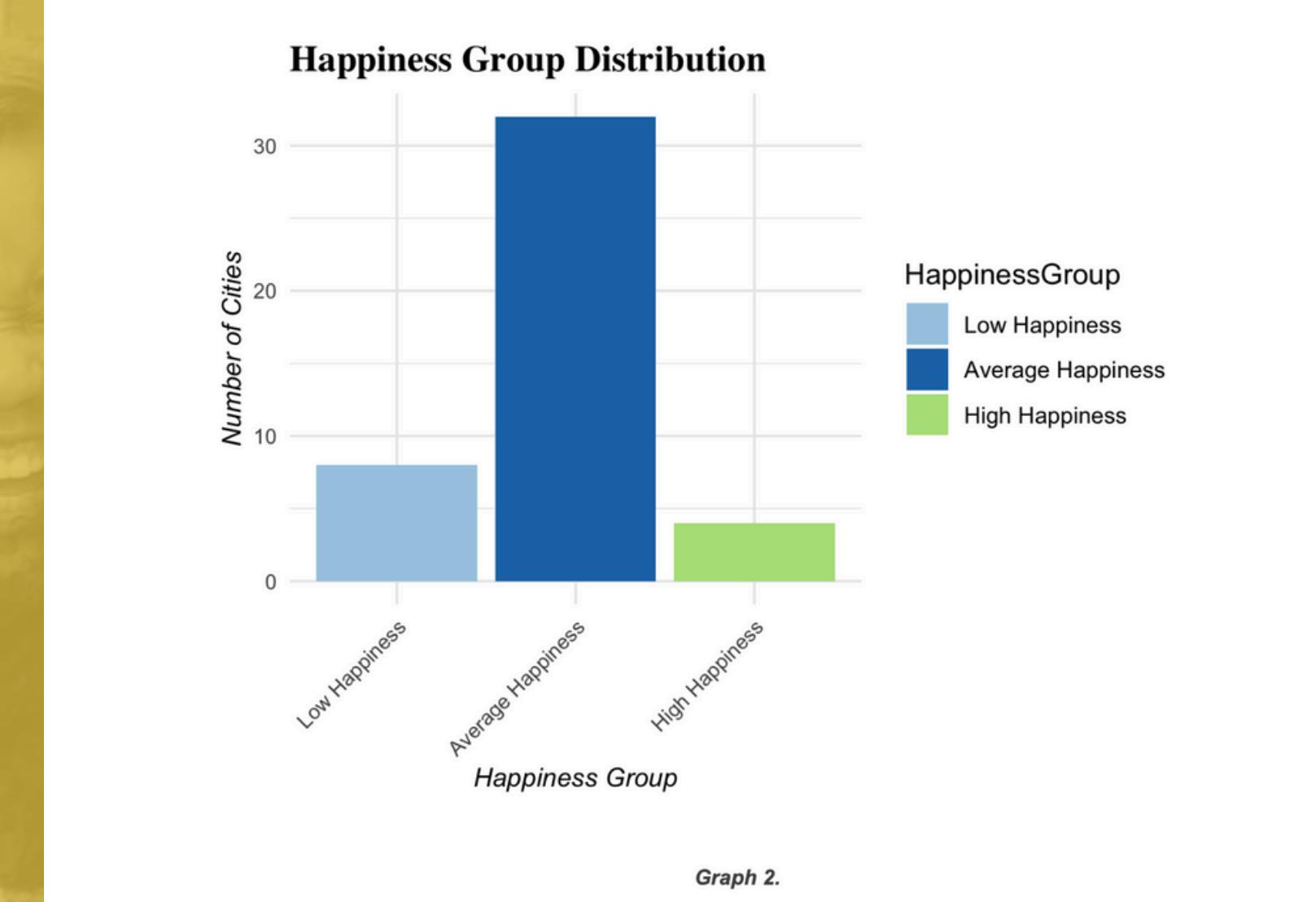
- **Outdoor activities** and **SunHr** are the two variables that have the greatest influence on Happiness Level
- **Gymcost** and **Waterbottlecost** are 2 variables that have a positive correlation with Obesity Level.



# Another assumption?

## **Assumption for a healthy lifestyle within a city**

- Great amount of SunHr and Outdoor Act
- Low Pollution Index
- High Life Expectancy
- Maybe low cost on gym and water bottle



The average\_happiness group has the highest number of cities (more than 30), followed by the low\_happiness group. So the number of cities with high happiness level is very low.

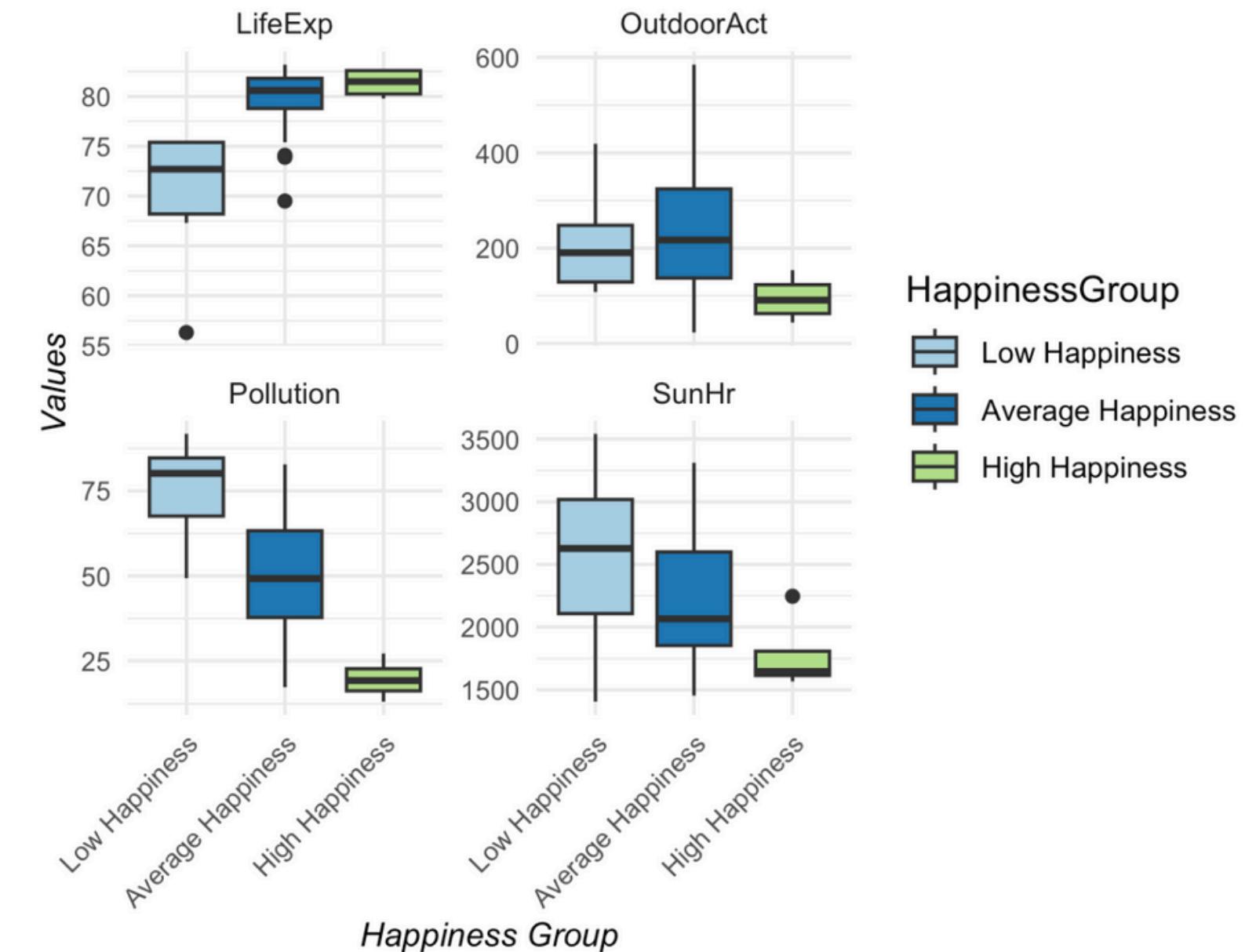
There is an observation that **the lower the Happiness Level, the lower the Life Expectancy**, assuming Happiness Level has a latent effect on Life Expectancy.

The **higher the Happiness Level, the lower their Pollution Index and Sun Hours**.

**The Low Happiness group has the highest number of sunshine hours**, ranging from **1500-3500 hours**, with the Median above around 2500 hours.

This shows that **High sunshine hours may not mean high happiness**.

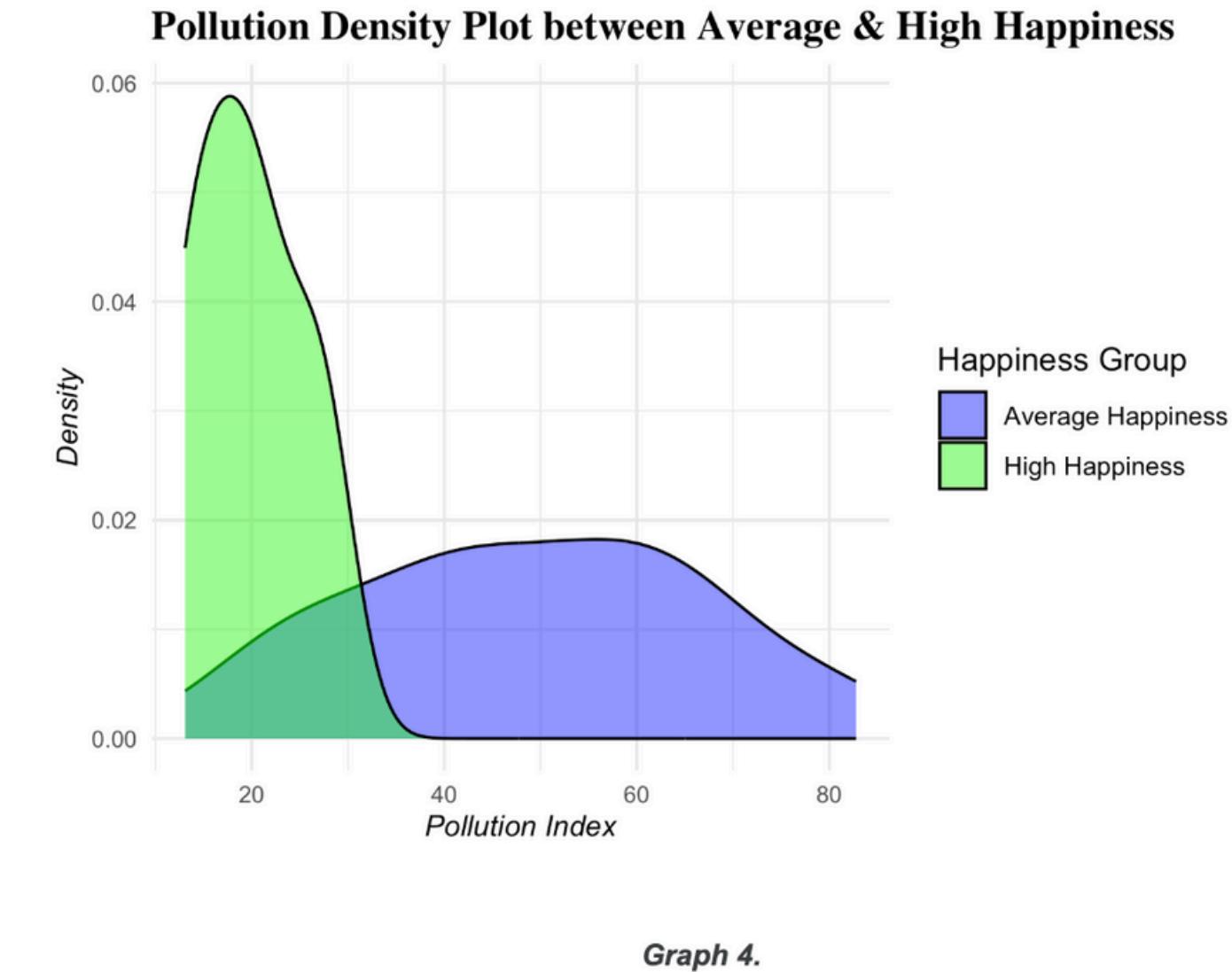
Compare factors between groups Happiness Group



Graph 3.

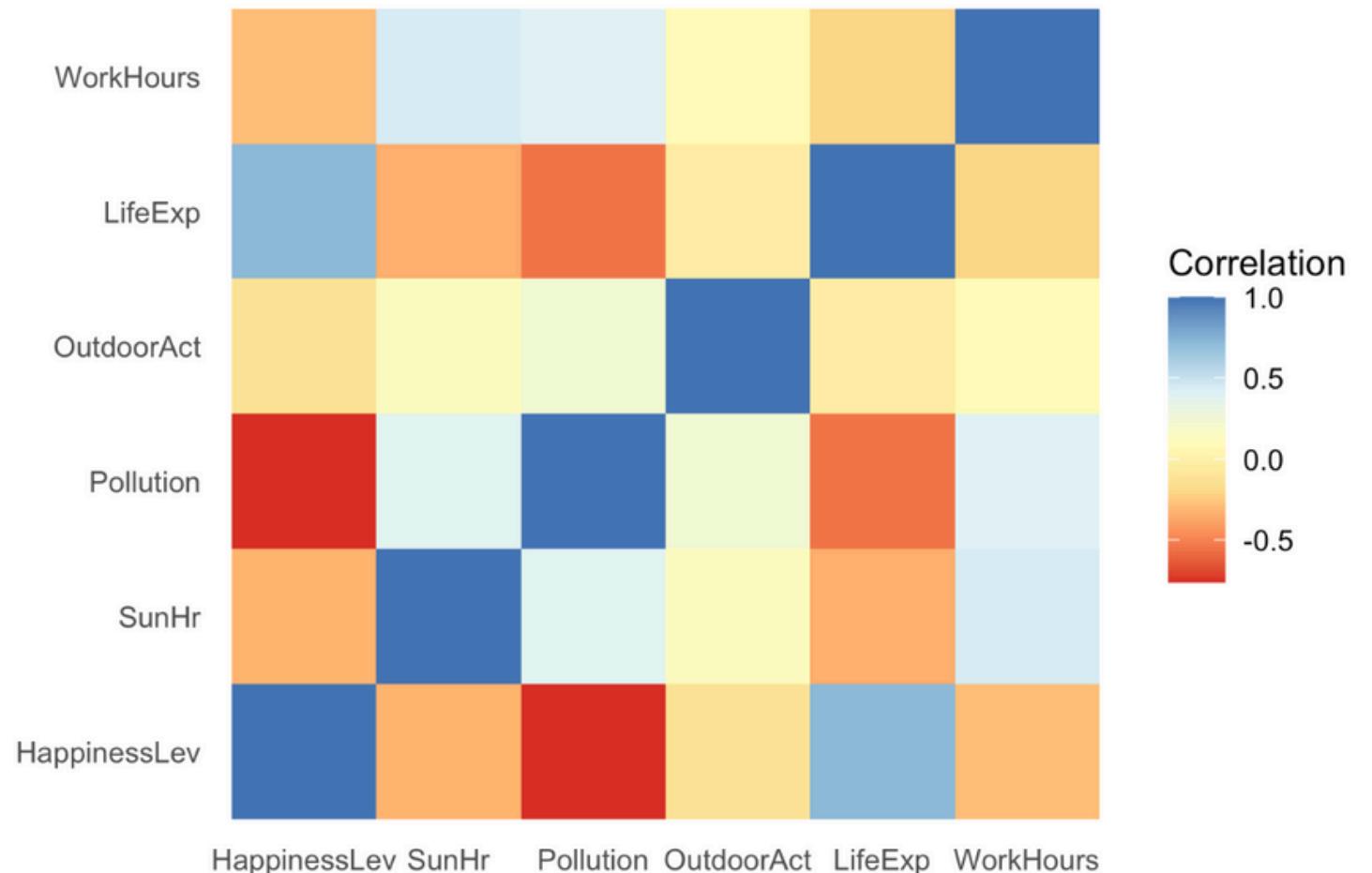
**The Mode of Average is high and narrow**, indicating that the data is less volatile, **cities with high levels of happiness often have similar Pollution index and are around 20**, as well as the left-skewed curve shows that the Pollution index of these cities has a lower value than the common level.

**The Mode of Avg. Group is around the pollution index value of 50**, the spread also shows that the data of this group is more diverse, reflecting the cities with very different environmental conditions.

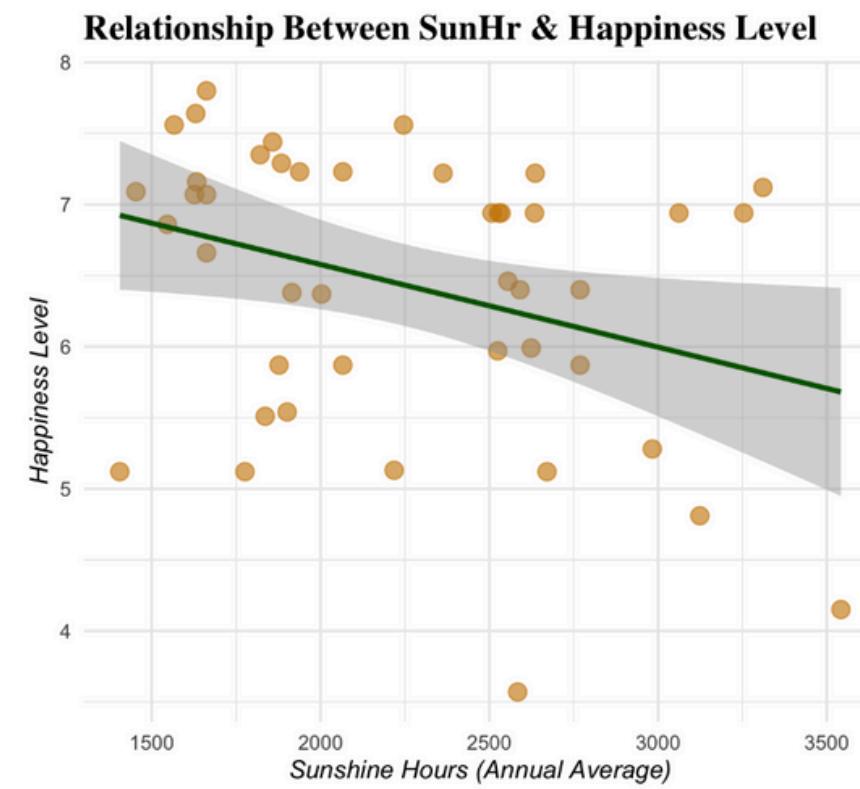


- **Positive Factors Affecting Happiness:**
  - LifeExp: Higher life expectancy is often associated with higher happiness levels.
- **Negative Factors Affecting Happiness:**
  - Pollution: Cities with high pollution levels often have lower happiness levels.
  - SunHr
  - WorkHours: More working hours can reduce happiness.
- **Low Impact Factors:**
  - OutdoorAct: not important determinants of happiness.

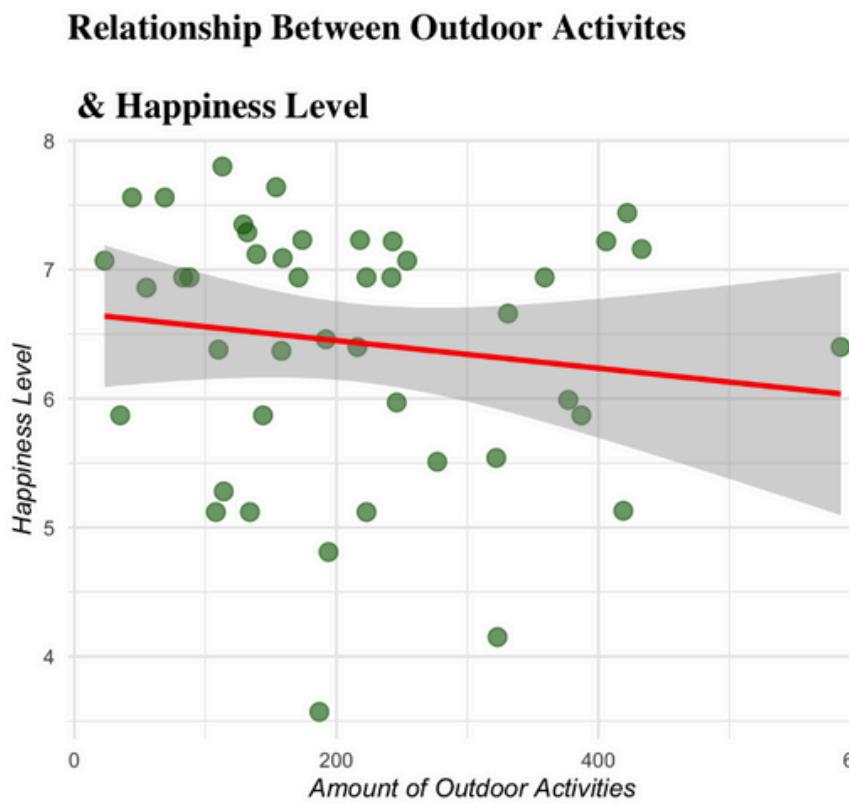
**Matrix Correlation between features and Happiness Level**



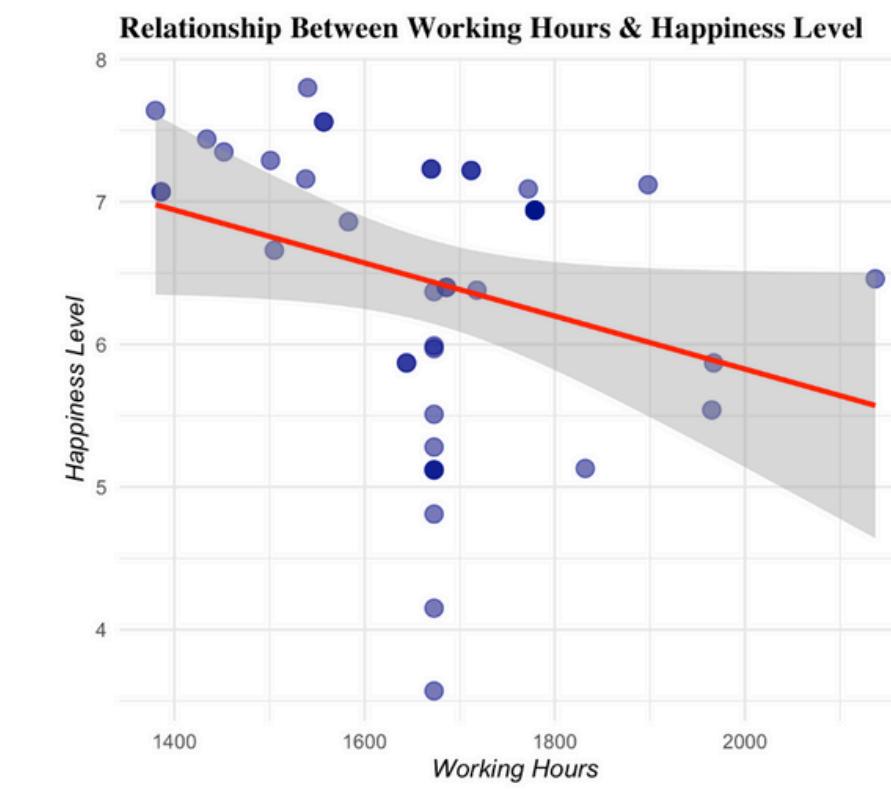
**Graph 5.**



Graph 6.



Graph 7.



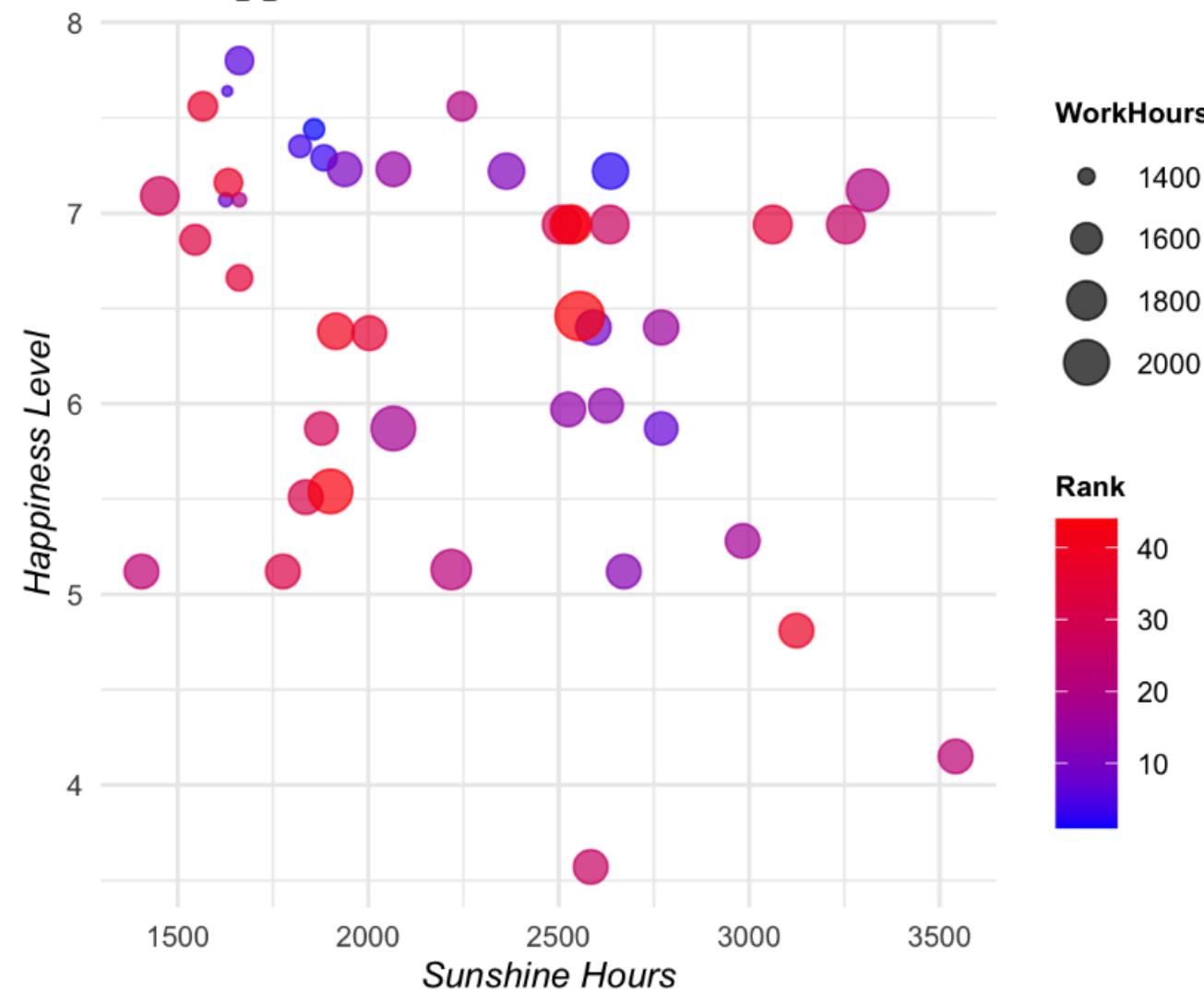
Graph 8.

All three trend lines have **negative slopes**, indicating that both SunHr and Outdoor Activities as well as Working Hours have a **negative impact on Happiness Level**. SunHr and Working Hours have stronger slopes and scatter points closer to the trend line, so they have a stronger impact than OutdoorAct.

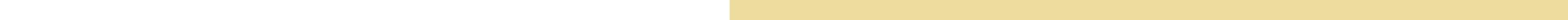


Large bubbles represent high work hours, which are typically in the Happiness Level 5-7 range, while low work hours are typically blue and fall in the Happiness Level 7 and above range. The chart shows that **cities with high healthy lifestyle rankings tend to have low work hours.**

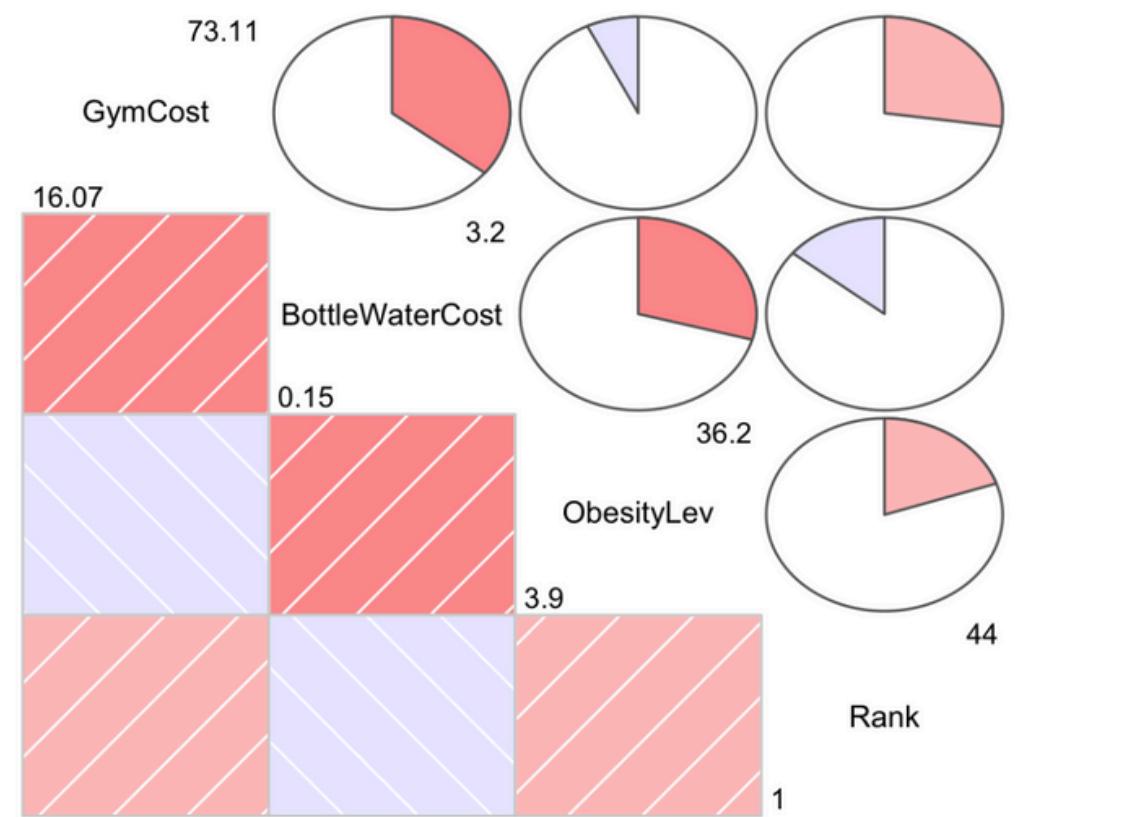
**Bubble Chart: SunHr, WorkHours vs Happiness Level**



**Graph 9.**



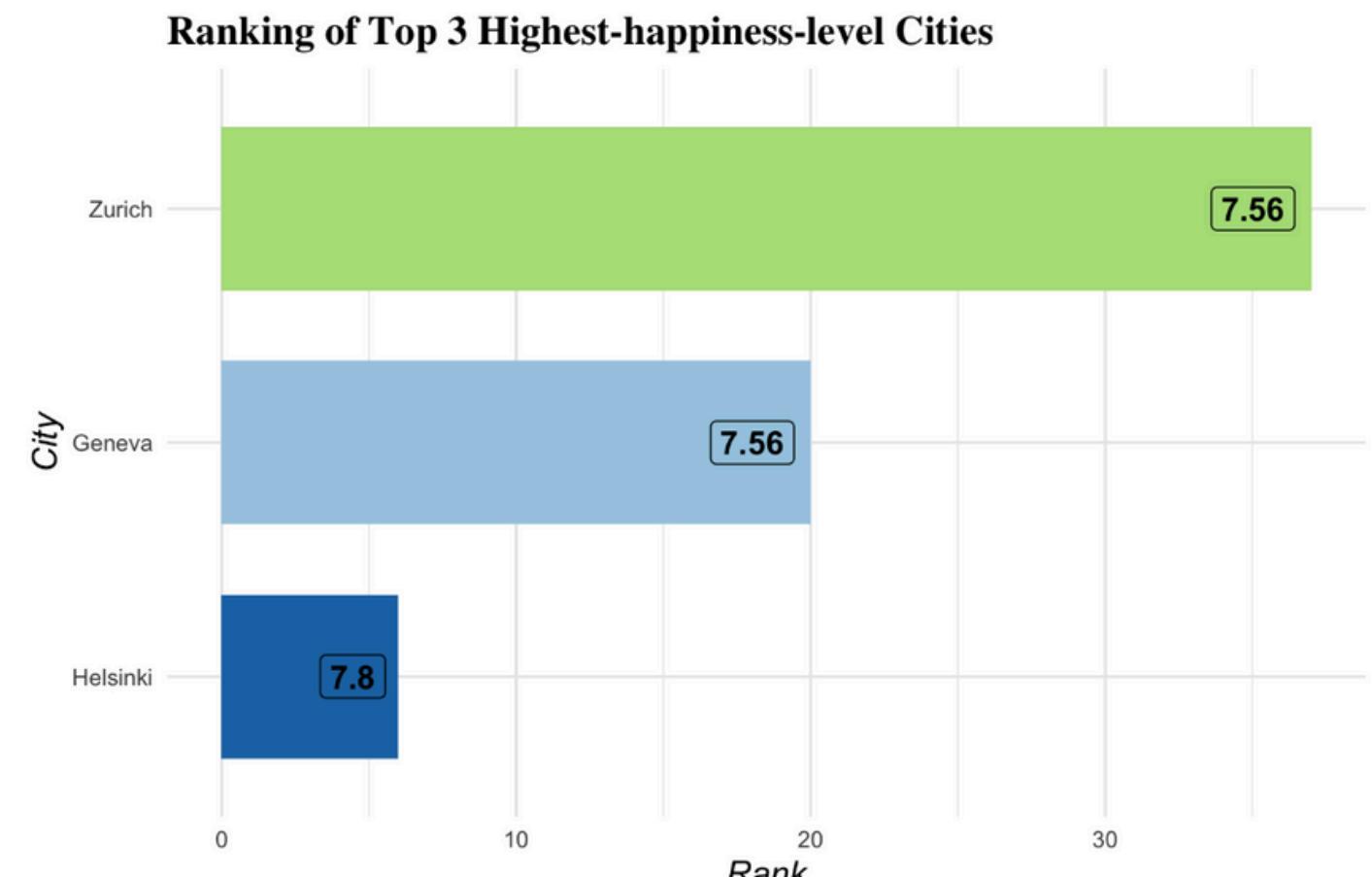
### Correlogram of Selected Variables



Graph 10.

The positive correlation value shows that as gym cost (GymCost) increases, so does bottle water cost (BottleWaterCost). **This may suggest that both variables reflect the high standard of living in cities where the general cost of living is high.**

# Conclusion



Graph 11.

Although **Helsinki has the highest happiness index, it is not the top city for healthy living.**

This result emphasizes that, although a healthy lifestyle plays an important role in improving quality of life, it is not the only factor that determines happiness. Other factors such as community connection, personal freedom, culture, and mental health also contribute to satisfaction and happiness.

**This encourages a balanced approach, combining physical and mental aspects of health, to achieve overall happiness.**