**LungCap**: This variable represents the measured lung capacity (closing capacity) of everyone, serving as a critical indicator of respiratory health and function.

**Age**: Indicates the age of everyone, providing insight into how lung capacity may vary across various stages of life.

**Height**: Reflects the height of everyone, offering valuable information on potential associations between lung capacity and physical stature.

**Smoke:** This binary variable categorizes individuals into smokers and non-smokers, allowing for the exploration of the impact of smoking habits on lung capacity.

**Gender**: Classifies individuals into male or female categories, enabling the examination of potential gender-based differences in lung capacity.

**Cesarean**: This binary variable indicates whether individuals were born via Cesarean section, providing additional context on potential prenatal factors influencing lung capacity.

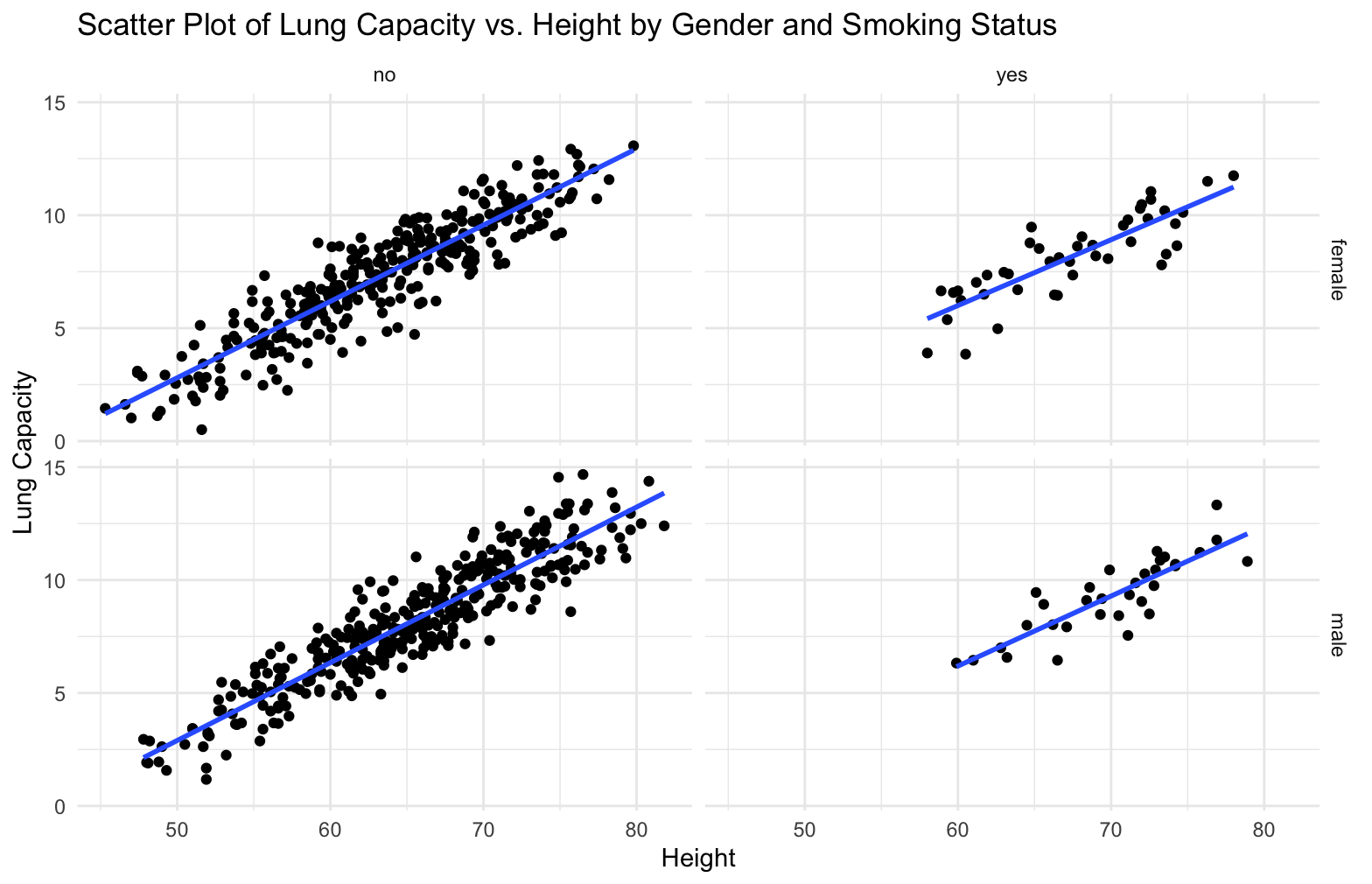
**Objective**: By analyzing this dataset, we aim to uncover patterns and relationships regarding lung capacity among smokers and non-smokers, considering age, gender, and height as key factors. This exploration will contribute to a better understanding of respiratory health and may inform interventions aimed at promoting lung health and preventing respiratory conditions.

1. Initial Research Question: Is there a correlation between lung capacity and height in both smokers and non-smokers, and does this relationship differ based on gender? Refinement: To achieve meaningful and relevant outcomes, we can refine this question to focus on specific aspects:

|  |  |
| --- | --- |
| cor\_male\_smokers | 0.8485134 |
| cor\_female\_smokers | [1] 0.8413888 |
| cor\_non\_smoking\_males | [1] 0.9185245 |
| cor\_non\_smoking\_females | [1] 0.9107652 |
| cor\_male\_smokers | [1] 0.8485134 |
| cor\_male\_non\_smokers | [1] 0.9185245 |
| cor\_female\_smokers | [1] 0.8413888 |
| cor\_female\_non\_smokers | [1] 0.9107652 |

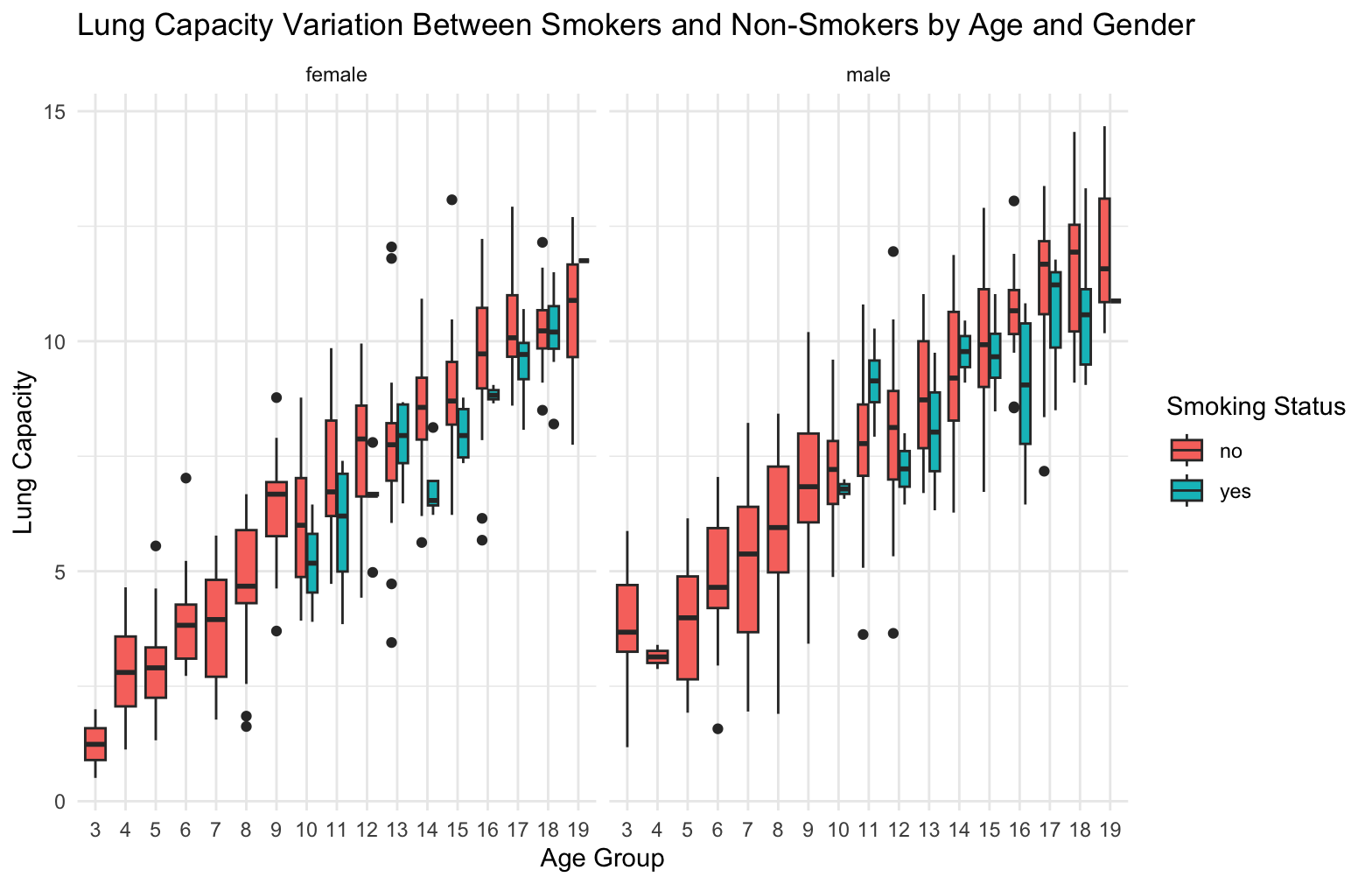
* 1. Is there a significant correlation between lung capacity and height among male smokers, and does this relationship differ from that observed among female smokers?
  2. Do non-smoking males exhibit a stronger correlation between lung capacity and height compared to non-smoking females?
  3. Are there differences in the strength of the correlation between lung capacity and height between smokers and non-smokers within each gender group?

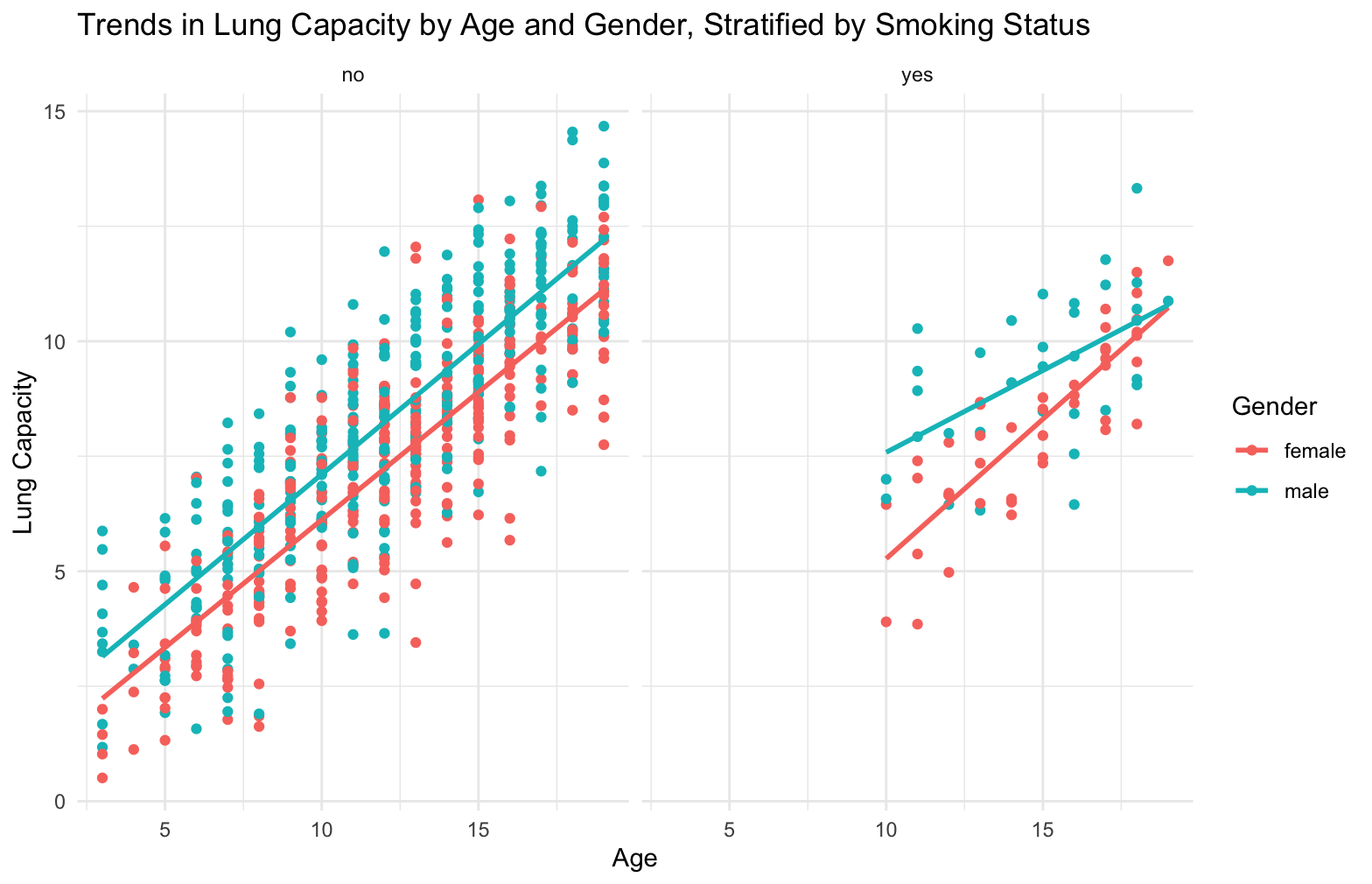
**Analysis:** Relationship between height and lung capacity: a. Taller male smokers appear to have higher lung capacities due to a strong positive correlation (0.8485134) between height and lung capacity. When compared to smokers who are female, this connection is marginally less (0.8413888). b. The correlation between height and lung capacity in non-smoking males is even higher (0.9185245) than it is in non-smoking females (0.9107652), suggesting that there is a stronger association between height and lung capacity in non-smoking males. c. Nonsmokers show a greater link between lung capacity and height within each gender group than do smokers, indicating that smoking may somewhat attenuate this relationship.



1. Initial Research Question: How does lung capacity vary between smokers and non-smokers across different age groups, and what are the gender-specific differences within each group? Refinement: To achieve meaningful and relevant outcomes, we can refine this question to delve deeper into specific demographic dynamics:
   1. Are there significant differences in lung capacity between male smokers and male non-smokers within each age group, and how do these differences evolve over time?
   2. Similarly, how does lung capacity differ between female smokers and female non-smokers across various age brackets, and are there any notable trends observed?
   3. Among smokers, is there a particular age group where the gender disparity in lung capacity is most pronounced, and what factors might contribute to this difference?

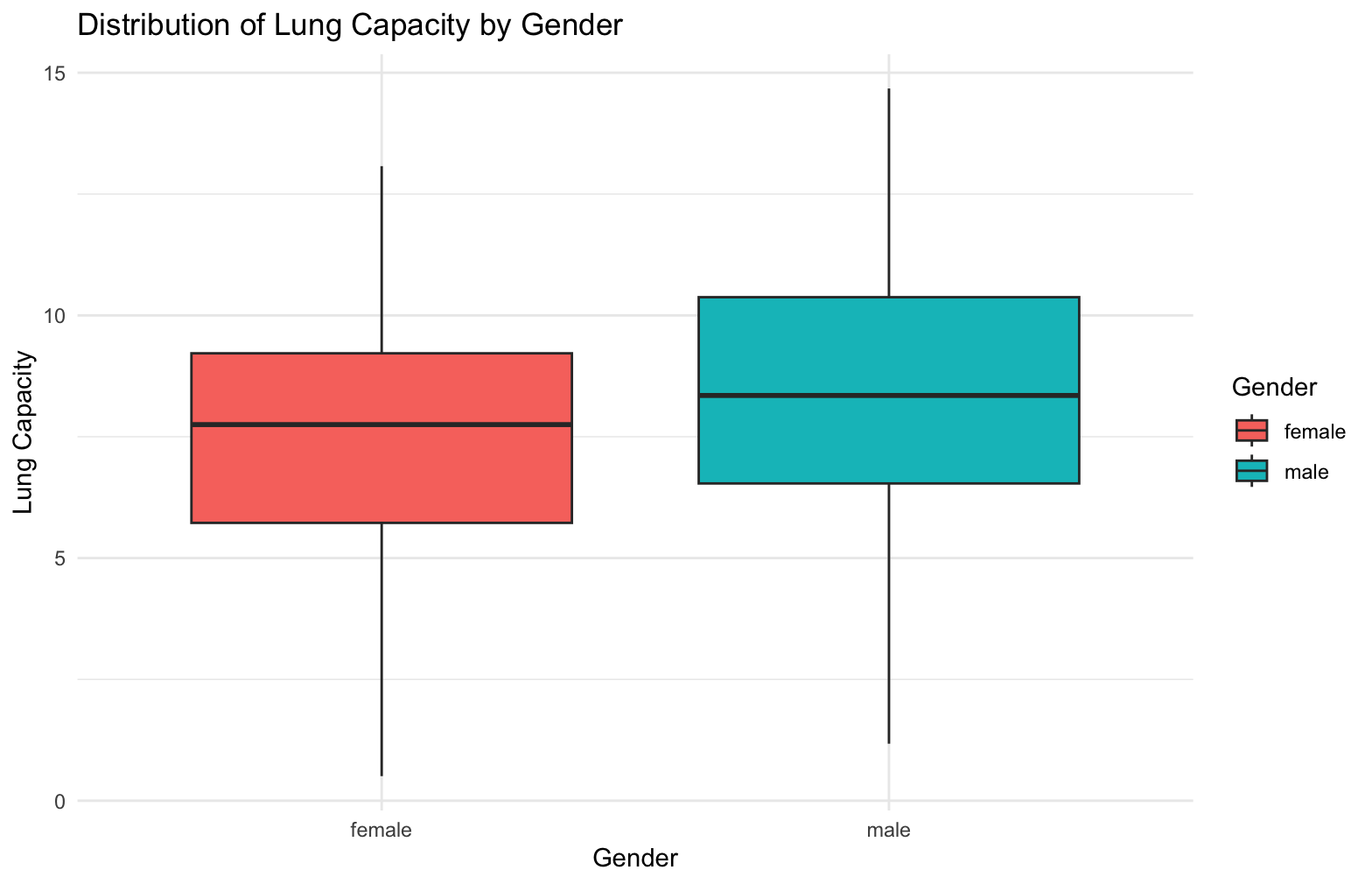
**Analysis**: Variation in lung capacity between smokers and non-smokers in different age groups: a. The graph indicates that male smokers have a negative effect on lung capacity, with male non-smokers consistently having a higher lung capacity than male smokers throughout all age groups. b. In a similar vein, females who do not smoke generally have greater lung capacity than female smokers, with a few exceptions where the difference is insignificant or slightly in favor of female smokers. c. Among smokers, the gender difference in lung capacity appears to be most noticeable in the 15–18 age range, where male smokers' lung capacities are noticeably greater than those of female smokers. This discrepancy could be explained by variables such variations in smoking behaviors, hormonal effects, or developmental phases.

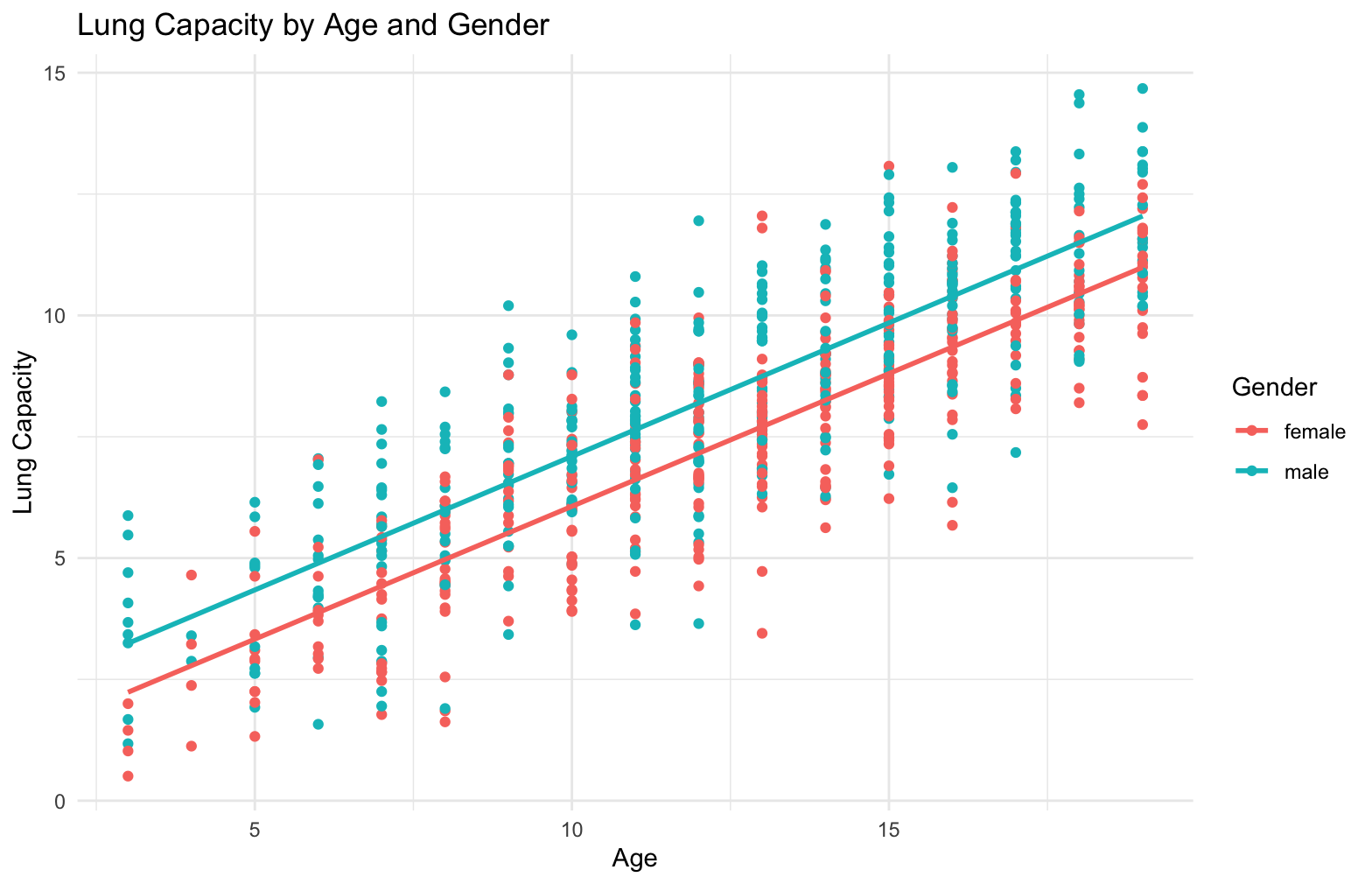


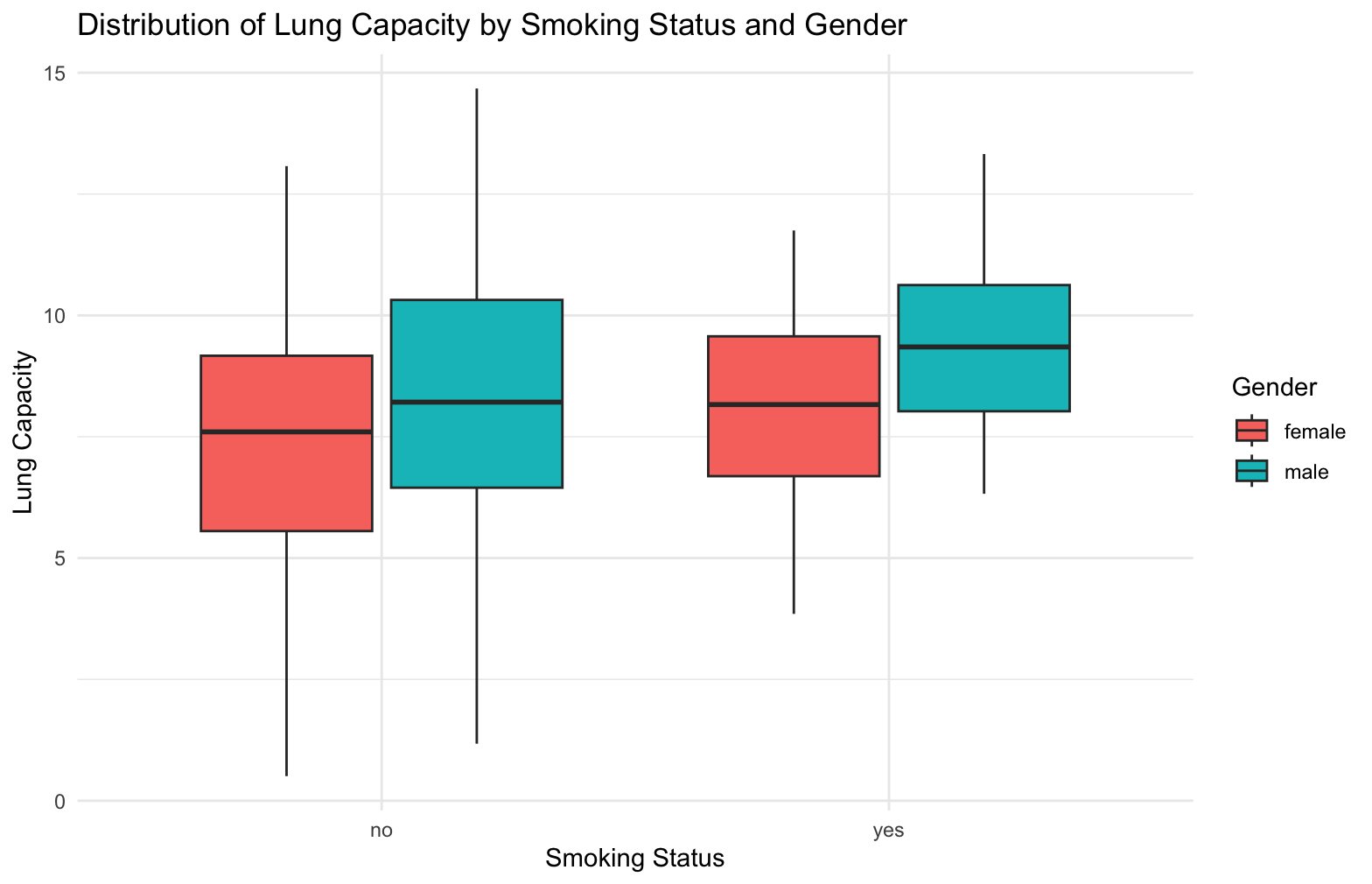


1. Initial Research Question: How does gender influence the difference in lung capacity between males and females, and what factors contribute to any observed disparities? Refinement: To achieve meaningful and relevant outcomes, we can refine this question to explore specific aspects of urbanization and its impacts:
   1. What anatomical and physiological differences between males and females, such as lung size, respiratory muscle strength, or airway dimensions, contribute to variations in lung capacity?
   2. How do hormonal factors, including sex hormones such as testosterone and estrogen, impact lung development, function, and capacity differently in males and females, and to what extent do these hormonal differences contribute to disparities in lung capacity?
   3. Are there behavioral and lifestyle factors, such as smoking habits, physical activity levels, or occupational exposures, that differ between males and females and influence lung capacity, and how do these factors interact with gender to contribute to observed disparities in lung capacity?

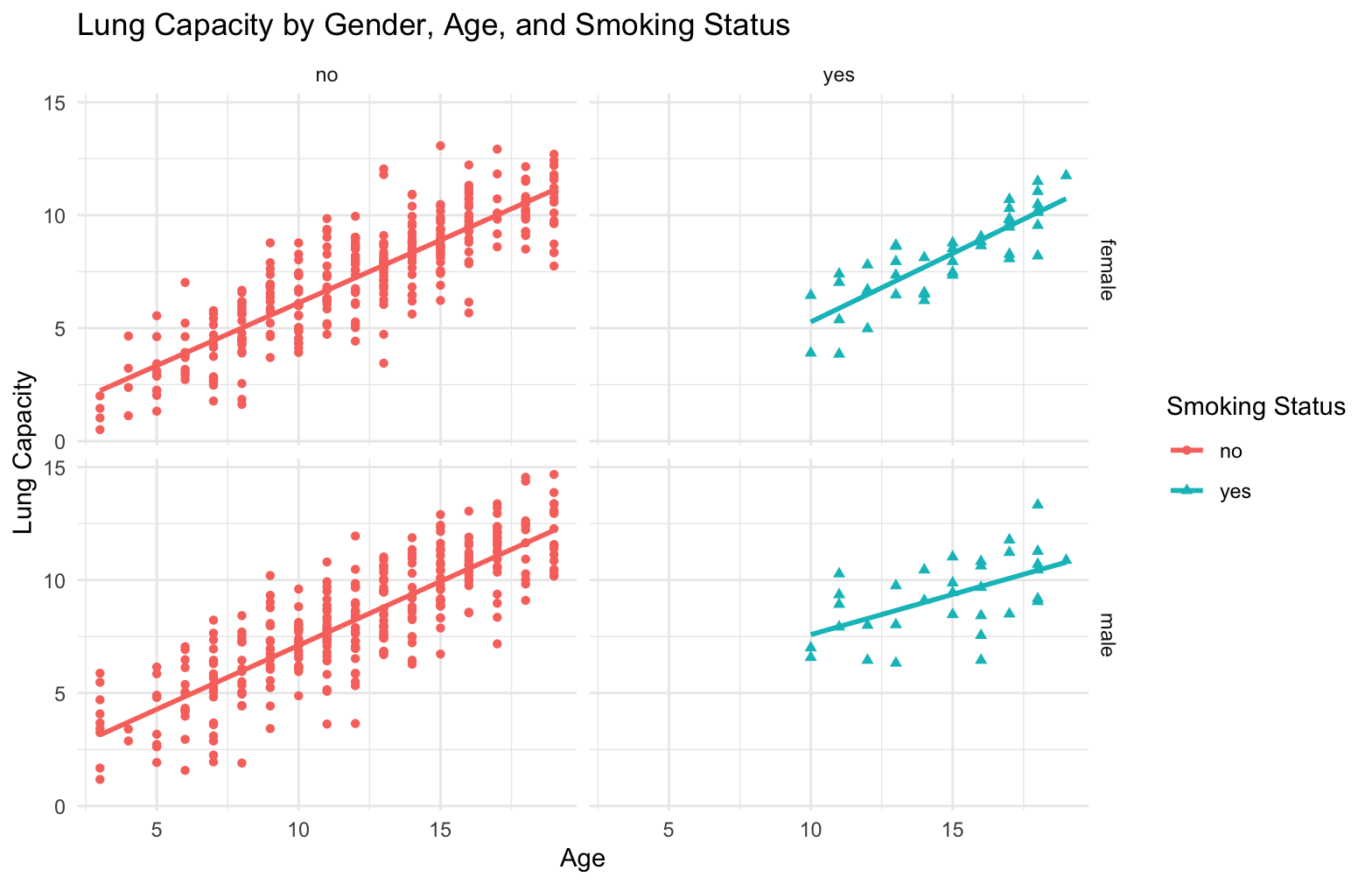
**Analysis:** a. The box plots indicate that, generally speaking, men tend to have larger lungs than women, maybe as a result of anatomical and physiological variations in things like lung size, respiratory muscle strength, or airway dimensions. b. The observed differences in lung capacity between males and females may be partially explained by hormonal factors such as testosterone and estrogen, which may have an impact on lung growth and function. c. Since behavioral and lifestyle factors can affect lung health and capacity differently in males and females, they may also be responsible for the observed gender differences in lung capacity. Examples of these factors include smoking behaviors, levels of physical activity, and occupational exposures.







All of the analysis in 1 graph:



Reference: <https://www.kaggle.com/datasets/radhakrishna4/lung-capacity>