



BIRMINGHAM CITY
University

Data Visualization Assessment

Bibek Sapkota

Student ID: 23189618

Total Page Count: 37

October 2024

Contents

1	Introduction	3
1.1	Thyroid Function	3
1.2	Motivation behind the chosen domain	3
1.3	Scope of the visualizations	3
1.4	Aims and Objectives	4
2	Data Exploration	6
2.1	About the dataset	6
2.2	Missing Values	7
2.3	Duplicate Values	7
3	Feature Engineering	8
4	Data Visualization	9
4.1	Answering the questions through data visualization	9
4.1.1	Demographic Analysis	9
4.1.2	Lifestyle and History Factor	14
4.1.3	Medical Analysis	18
4.1.4	Clinical Feature	22
4.1.5	Cancer-Specific Analysis	25
4.1.6	Treatment Outcomes	30
5	Summary	35
6	Future Work	36

List of Figures

1	Number of Missing Values per Column	7
2	Age Distribution of Patients	9
3	Gender Distribution of Patients	10
4	Age Distribution by Gender	11
5	Recurrence rate in different age	12
6	Recurrence rate comparison by gender	13
7	Smoking status and severity of thyroid pathology	14
8	Cancer Stage by Smoking History and Current Status	15
9	Radiotherapy history on Thyroid Function	16
10	Impact of Radiotherapy history on Recurrence	17
11	Distribution of Thyroid Function	18
12	Relationship of Thyroid Function and Physical Examination	19
13	Thyroid Function as compared to Stage	20
14	Thyroid Function by Age and Gender	21
15	Correlation of Thyroid stage and Adenopathy	22
16	Recurrence Rate vs Adenopathy	23
17	Physical Examination vs Risk	24
18	Distribution of TNM staging in thyroid cancer	25
19	Distribution of Risk categories with Multi-Focal Pathology	26
20	Correlation of Risk and Pathology Type	27
21	Distribution of Age and Gender across Cancer Stage	28
22	Gender vs pathology vs Recurrence	29
23	Treatment outcome by Age group	30
24	Treatment Response by Gender	31
25	Treatment Response vs Recurrence	32
26	Treatment Response by Stage and Smoking status	33
27	Treatment Response by thyroid function	34

Visualization of Patterns in Thyroid Disease Progression: A Data-Driven Analysis

1 Introduction

1.1 Thyroid Function

The thyroid gland is essential for human health, and thyroid dysfunction is particularly common in regions with iodine deficiency, given vital role in hormone production (Arthur and Beckett, 1999) . It plays a pivotal role in regulating essential bodily functions, including metabolism, growth, and development. A thyroid nodule is a discrete lesion within the thyroid gland that is radiologically distinct from the surrounding thyroid parenchyma (Cooper et al., 2009) . More than 90% of detected thyroid nodules are clinically insignificant because they have no ultrasound features that suggest malignancy or because they are cytologically benign (Durante et al., 2018) . When thyroid function is impaired, it can lead to conditions such as hypothyroidism, hyperthyroidism, or thyroid cancer, each presenting distinct challenges for patient health. Analyzing these dysfunctions offers valuable insights that can guide early diagnosis, targeted treatments, and personalized patient care strategies.

1.2 Motivation behind the chosen domain

Globally, the prevalence of thyroid cancer is still rising , with with *differentiated thyroid cancer* being the most frequent type, accounting for about 95% of cases (Cabanillas et al., 2016) and it's crucial to understand the factors that affect patient outcomes, As This trend makes knowledge of the elements affecting patient outcomes even more urgent. Clinical data analysis informs treatment and care for patients by highlighting more effective therapeutic approaches and refinement of prevention. It might also lead to a better understanding of the natural history of a disease and of its interaction with patient-specific variables, hence allowing personalized medicine approaches that tailor treatments to individual needs.

1.3 Scope of the visualizations

Visualizations in this report are geared toward the discovery of important patterns and insights in data on thyroid disease outcomes, recurrence, and treatment. The analytic investigation will consider a broad range of variables, from demographic and lifestyle factors to clinical attributes, in describing the complex nature of thyroid disease. The following relationships shall be visually analyzed to inform improvements in medical decision-making and management of the patient. A set of key questions about the dataset are answered using the visualizations:

- **Demographic Analysis:** To illustrate the mapping of the distribution of thyroid diseases according to age groups and sex, which will describe the trend of thyroid disorders in onset and development. This will indicate whether any particular demographic group is more prone to thyroid dysfunction or thyroid cancer.

- **Lifestyle Factors and Thyroid Diseases:** Assessing the impact of lifestyle factors, such as smoking history and history of radiotherapy on thyroid function and the course of cancer. It will further be visually portrayed in comparing the severities of thyroid conditions amongst patients of different statuses of smoking and with or without a history of radiotherapy.
- **Thyroid Status and Stage of Cancer:** Analyzing the association of thyroid status with stage of cancer. It would be seen from the plots the distribution of the staging where patients with normal thyroid status-euthyroid-are more likely to be diagnosed in the earlier stages, while abnormalities in thyroid status are associated with the advanced stages of the disease.
- **Recurrence Rate and Risk Factors:** Analyzing recurrence rates in thyroid malignancy regarding demographic, clinical, and lifestyle factors. The plots illustrate variation in cancer recurrence rates according to sex, age group, adenopathy, and pathology type.
- **Treatment Outcome by Demographics and Stage of Cancer:** It demographically maps the patient's age and gender with response to treatments regarding thyroid disorders. This would also include treatment success rate analysis across different age groups and stages of cancer, which can help identify the most responsive group toward the treatment.
- **Treatment Response Modulated by Smoking:** Investigating the role of smoking status in modifying treatment responses at different stages of cancer. Visualizations will show how differences in treatment efficacies between smokers versus nonsmokers are varied, especially in early versus advanced stages of thyroid cancer.
- **Physical Examination Findings and Levels of Risk:** Findings from physical examinations, such as nodules or goitres, are charted against the levels of risk that the patients have been categorized into. This will then help in identifying whether certain physical examination findings are indeed more indicative of a high-risk thyroid disorder.
- **Feature Correlations:** Visualize the correlation of numeric features such as thyroid function, pathology type, and tumor size. These plots will show the significant correlation that can be useful in modeling thyroid cancer outcomes or understanding disease progression.

1.4 Aims and Objectives

This analysis essentially focuses on the understanding, through visualization, of various factors related to the progression, recurrence, and treatment outcomes in thyroid disease, using comprehensive patient data. The study shall try to elucidate how patient demographics, lifestyle factors, thyroid function, and clinical outcomes are interrelated by tapping into data-driven insights. This is ultimately to support diagnosis, treatment planning, and improved strategies for patient management of thyroid disorders.

Objectives of Visualization

- **Demographic pattern analysis::** Demographic pattern analysis means how thyroid disorder varies across age group and gender; and what significant trends, if any, can be traced about the onset, progression, recurrence, and staging of the disease.
- **Life-course factors analysis:** What history of smoking and radiotherapy treatment effects does lifestyle have on the stage of thyroid disease and possible recurrence?.
- **Explore thyroid function:** Visualize the distribution of thyroid status with regard to thyroid function (euthyroid, hypothyroid, and hyperthyroid) against cancer stage and physical examination findings.
- **Treatment outcome comparison:** Assess treatment response rates of various treatments stratified for patients' demographic data, stages of cancer, and categories of thyroid function, underlining treatment response and recurrence rates.
- **Clinical feature correlation:** Explain the correlation of the clinical features, such as adenopathy and stages of cancer, which would relate to the outcomes and recurrence rates in the patient.
- **Identifying high-risk groups:** Highlight specific risk factors and demographic groups that may be more prone to recurrence or unsatisfactory treatment outcomes; this would prove useful for actionable insights on personalized treatment approaches.

2 Data Exploration

2.1 About the dataset

This dataset covers the prevalence and attributes of thyroid conditions in a given patient population, with data analyzed against a broad set of demographic, past medical history, and clinical outcome variables. In this database, the key attributes have been included that could be useful in understanding disease progression and response to treatments in thyroid disorders. Each record offers very valuable insights into the condition of the patient and the efficacy of treatment given, and can thus be analyzed for an improved diagnosis and therapeutic intervention in thyroid health.

Column Descriptions:

- **Age:** The age of the patient at the time of diagnosis in years.
- **Gender:** The gender of the patient, recorded as male (M) or female (F).
- **Smoking:** Indicates whether a patient is currently a smoker or not.
- **Hx Smoking:** History of smoking; yes/no indicating whether the patient has smoked in the past.
- **Hx Radiotherapy:** History of undergoing radiotherapy treatment (Yes/No).
- **Thyroid Function:** Indicates current status of the patient's thyroid function.
- **Physical Examination:** Results from physical examination of the thyroid.
- **Adenopathy:** Indicates whether adenopathy (swollen glands) was present (Yes/No).
- **Pathology:** The type of pathology identified from tissue samples.
- **Focality:** Whether the tumor is uni-focal(single) or multi-focal(multiple).
- **Risk:** The risk classification assigned to the patient.
- **T:** The T classification indicates tumor size.
- **N:** The N classification indicates the level of lymph node.
- **M:** The M classification indicates whether metastasis (spread of cancer to other parts of the body) has occurred.
- **Stage:** The overall cancer stage based on the TNM classifications, ranging from Stage I to Stage IV.
- **Response:** The patient's response to treatment, including whether the disease has stabilized, improved, or worsened.
- **Recurred:** Indicates whether the cancer has recurred after the treatment (Yes/No).

2.2 Missing Values

I used R's `colSums()` function in combination with `is.na()` to determine which missing values were in the data set. `ColSums()` effectively counts the number of null entries in a given column. This will display systematic results using `print(missing_values)` and provide an exact count of the number of missing values within a column.

Age	Gender	Smoking	Hx.Smoking
0	0	0	0
Hx.Radiotherapy	Thyroid.Function	Physical.Examination	Adenopathy
0	0	0	0
Pathology	Focality	Risk	T
0	0	0	0
N	M	Stage	Response
0	0	0	0
Recurred			
0			

Figure 1: Number of Missing Values per Column

As shown in the above figure, there are zero counts in every column. The analysis verified that there were no missing values at all. This result guarantees that no data imputation techniques are required because the dataset is fully populated and is prepared for additional analysis.

2.3 Duplicate Values

In cleaning the data set for analysis, we determined there were duplicate records. We found nineteen duplicate rows using the `duplicated()` function, which goes through every row across columns and flags the repetition. We then used the addition of these duplicates with `sum(duplicates)` to count how many are in order to show the duplicate rows.

```
# Check for duplicate rows [1st - output]
no_of_duplicates_before_cleaning <- duplicated(data)
sum(no_of_duplicates_before_cleaning)

# Remove duplicate rows while keeping the first occurrence [2nd -output]
df <- data[!duplicated(data), ]
no_of_duplicates_after_cleaning <- duplicated(df)
sum(no_of_duplicates_after_cleaning)
```

These duplicates were cleared for the integrity and cleanliness of our data. The negated output of this function, `duplicated()`, filtered out the repeats in the dataset and retained only the first occurrence. We reassessed the cleaned dataset to confirm that the duplicates had been removed and checked the new row count by using `print(nrow(df))`. This is a very important step in keeping our data intact, meaning it gives reasons on which we may base our conclusions that are unique and not redundant.

3 Feature Engineering

A new feature called "Smoking_Status" was created in the dataset development process to provide more information on the impact of smoking habits on health outcomes. Two previous columns— 'Smoking' and 'Hx.Smoking' that represented smoking behaviors in the past and present, respectively, were combined to create this feature.

process:

The "Smoking_Status" column was created using the R package dplyr and the following criteria:

1. **Ever Smoked:** Subjects fall under this category depending on whether they smoke or has smoke in past. The 'Ever Smoked' in the column against smoking_status indicates that persons either smoke now or have smoked in the past.
2. **Never Smoked:** In the absence of a 'Yes' in either column (Smoking and Hx.Smoking), the status given would be "Never Smoked," and that person never smoked.

```
# Create a combined column
df <- df %>%
  mutate(Smoking_Status = case_when(
    Smoking == 'Yes' | Hx.Smoking == 'Yes' ~ 'Ever Smoked ',
    TRUE ~ 'Never Smoked'
  ))
```

Justification:

1. **Simplifying Data for EDA:** Such a combination of variables into one feature is a simplification of the data. This simplification allows the analyst to conduct EDA more simply; hence, finding trends and patterns related to smoking habits and health outcomes corresponding to those habits is efficient.
2. **Enhancing Analytical Accuracy:** This combination of feature will make the dataset more beneficial for EDA, as it explains the variable of any smoking exposure, whether past or present. Usually, the current and past practices of smoking are high-value factors when research analysis of such kind needs correct judgments about health hazards and outcomes.

This feature engineering effort not only simplifies the dataset for more effective EDA but also enhances the ability to uncover meaningful insights into how smoking behaviors influence health.

4 Data Visualization

4.1 Answering the questions through data visualization

This section is supported by a thorough research design in which each question was answered with due preparation. The answers are allowed to ensure that clear and detailed understanding, as well as insights into the problem, are clearly explained. Relevant graphics are included to the facts presented to enhance this understanding and complement the review and presentation for the purpose of better organization. This will ensure that the findings of the research will actually meet the objective of the presentation, wherein the text will support the pictures and vice versa. As well as expectation, which is also mentioned as 'E', represents an expectation for the question.

4.1.1 Demographic Analysis

1. How does age distribution and Gender distribution vary among different thyroid function categories ?

E: It is expected that the older the patients are, the faster the stage of thyroid cancer. Whereas, Gender is not expected to have a significant impact.

Age Distribution of Patients

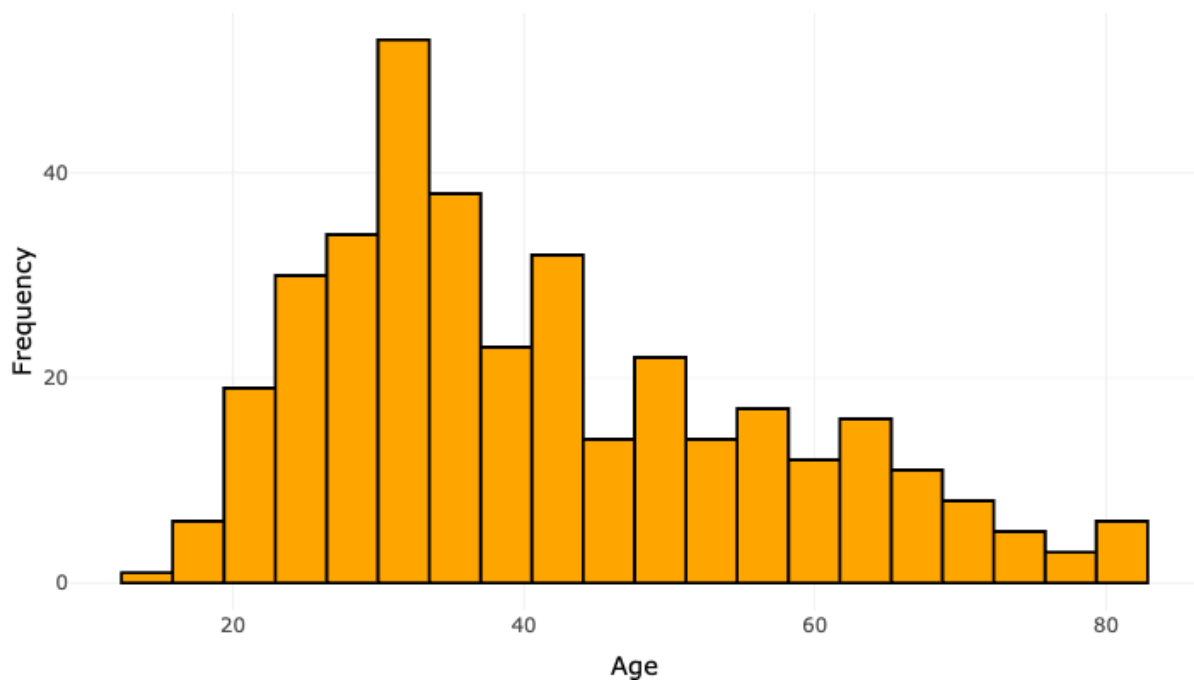


Figure 2: Age Distribution of Patients

The age distribution of thyroid patients is clearly dominated in the younger to middle-aged population, with the majority of the cases occurring at an age of about 35 years. After this peak, the number of the patients gradually decreases, with much fewer scenarios found in people over 60. It would thus appear that though thyroid conditions are more common among young adults, it can occur at any age.

Whereas a distinct overlap in age groups suggests that the occurrence of thyroid issues is not heavily biased toward any specific age group. In many cases, thyroid

nodules may be detected because of their size or anterior position , or the skill of the physician performing the examination (Vanderpump, 2011) .

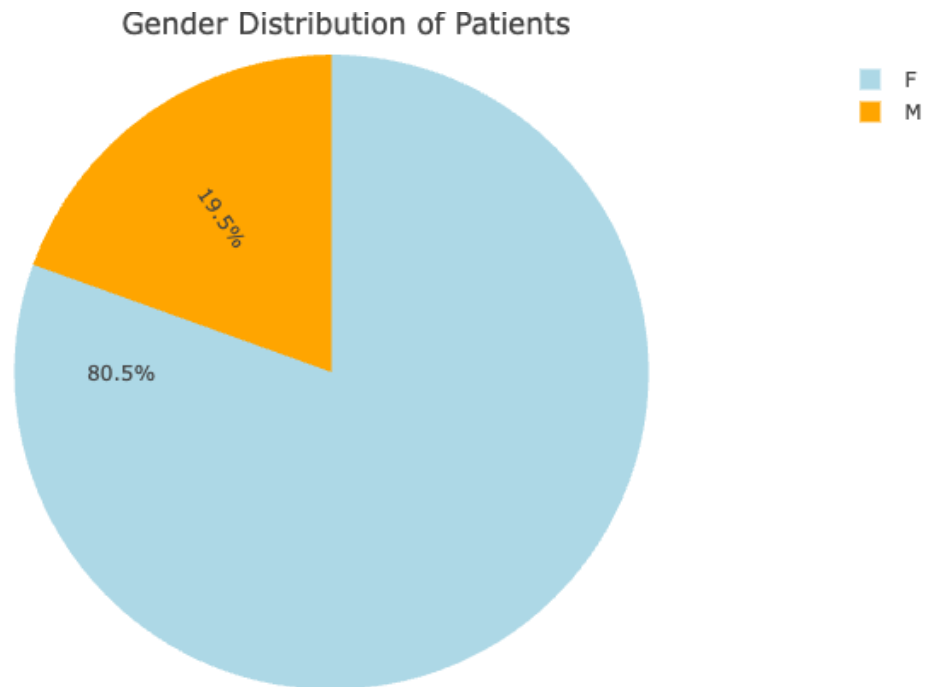


Figure 3: Gender Distribution of Patients

The above chart on the gender distribution of patients clearly indicates an overwhelming dominance by female patients in relation to male patients. It is indicated that 80.5% of the patients are females, abbreviated as F, while those referred to as males, abbreviated as M, constitute 19.5%. This reveals great disparity in distribution and may indicate that thyroid conditions could be more prevalent in women than in males. These kinds of trends could be contributed to by biological, hormonal, or genetic predispositions which make females susceptible to thyroid conditions.

Despite the fact that both genders are affected by thyroid conditions, the skewed distribution does not appear to show this. The fact that almost 20% of the patients are males drives home the message that, Thyroid diseases are widespread conditions, occurring significantly more often in females than in males (Bauer et al., 2013) .

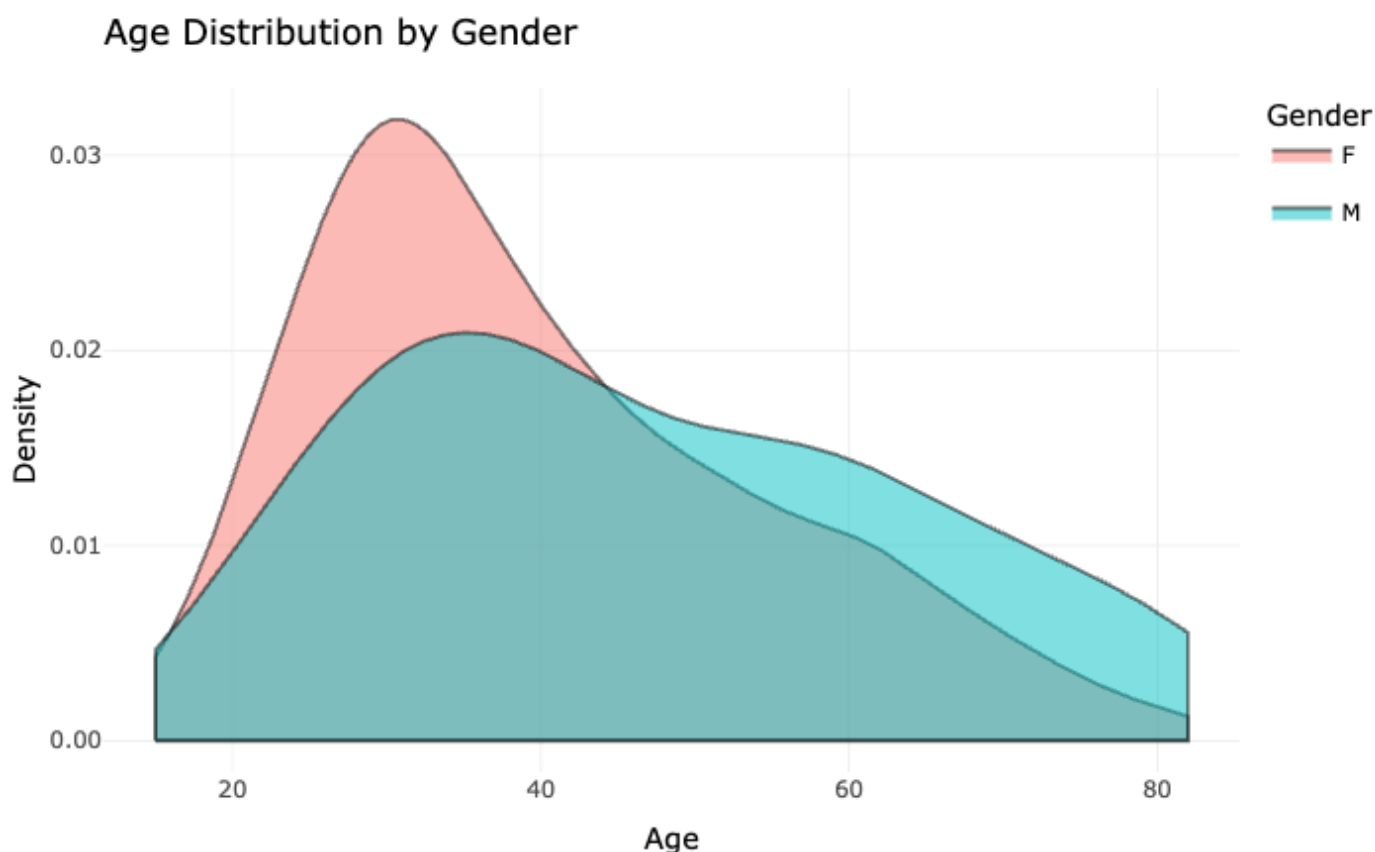


Figure 4: Age Distribution by Gender

The density plot shows the Age distribution by gender for thyroid conditions., it will show age demographics overlay between male and female patients diagnosed with thyroid conditions. From this plot, the two genders are very similar in age profile- the majority of the cases fall within the young to middle-aged groups. The peak in both sexes happens in the mid-thirties, which insinuates that thyroid conditions are most prevalent in this age bracket. A density plot showing a wider spread in females indicates that age of prevalence is wider than in males, whose age distribution appears slightly narrowed.

While the overall trend is towards younger to middle-aged patients, the plot shows that thyroid conditions can happen at any age, hence not being confined to any particular group of patients. The graph further suggests that thyroid conditions indeed have a higher prevalence among females, as suggested by the gender distribution showing the majority of patients to be females. It is also important to emphasize that age distributions overlap in gender, thus reinforcing the notion of higher preponderance in females and a wide age group affected in both males and females; thus, most demographic groups are affected by thyroid disorders.

2. What are the statistical correlations between age, gender and recurrence rates in thyroid patients?

E: Recurrence rates may be lower in younger than in older patients, and it is expected that recurrence rates could be higher in females compared to males .

Age Distribution of Patients with and without Recurrence

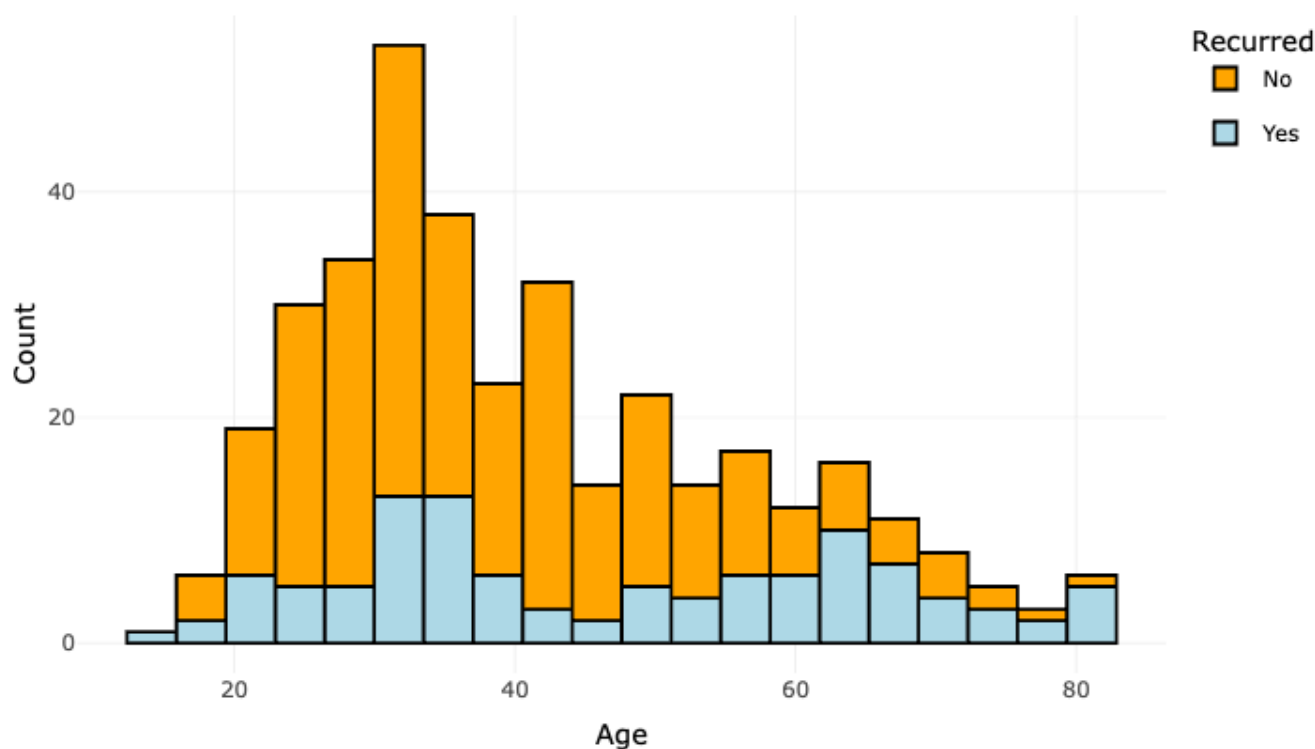


Figure 5: Recurrence rate in different age

The age distribution of thyroid patients is clearly dominated in the younger group to middle-aged population, as shown in (Figure 2) with significant concentration of cases around age 35. Recurrence is observed in all age groups; however, no overwhelming bias in any particular age group exists. Though the overall representation seems to be more for the younger patients, recurrence doesn't appear to increase significantly with age or become more frequent in older individuals.

This suggests that, while age may not be the strongest predictor of recurrence, further statistical tests might be needed to identify subtle trends in older or younger patients, especially when other risk factors are considered.

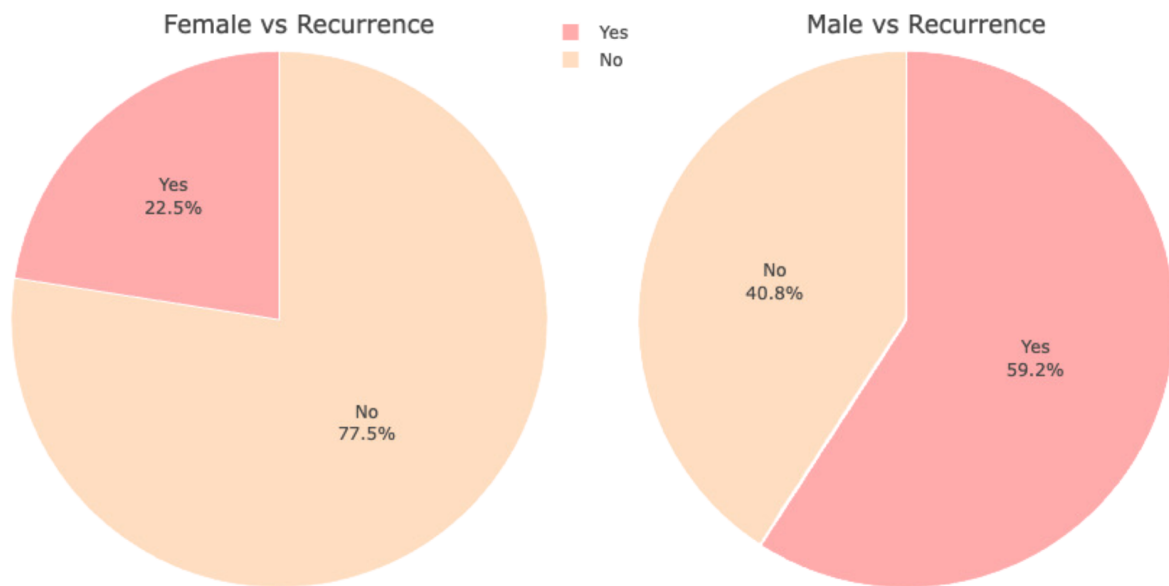


Figure 6: Recurrence rate comparison by gender

These two pie charts show sharp contrasts in the gender distribution according to recurrence status among thyroid patients. Only 22.5% of the female patients went into recurrence, while the majority- 77.5% of the female patients did not face recurrence. This indicates that the larger proportion of female patients has maintained a stable outcome after receiving treatment. In contrast, the male patients exhibit a significantly higher recurrence rate- 59.2% of the patients suffered recurrence, whereas 40.8% did not. This higher recurrence rate in males would therefore be indicative of greater risk for this group concerning thyroid cancer outcome.

This might mean that the male patients are more vulnerable to recurrence of thyroid cancer compared to females. That a gap exists between the genders adds to curiosity on what possibly could contribute to such an outcome, which might include biological factors or environmental exposures, or differences in treatment received

4.1.2 Lifestyle and History Factor

1. Does smoking correlate with the severity of thyroid pathology findings?

E: Those patients with smoking may show more aggressive pathology, a reflection of the deleterious effects of tobacco on the thyroid gland.

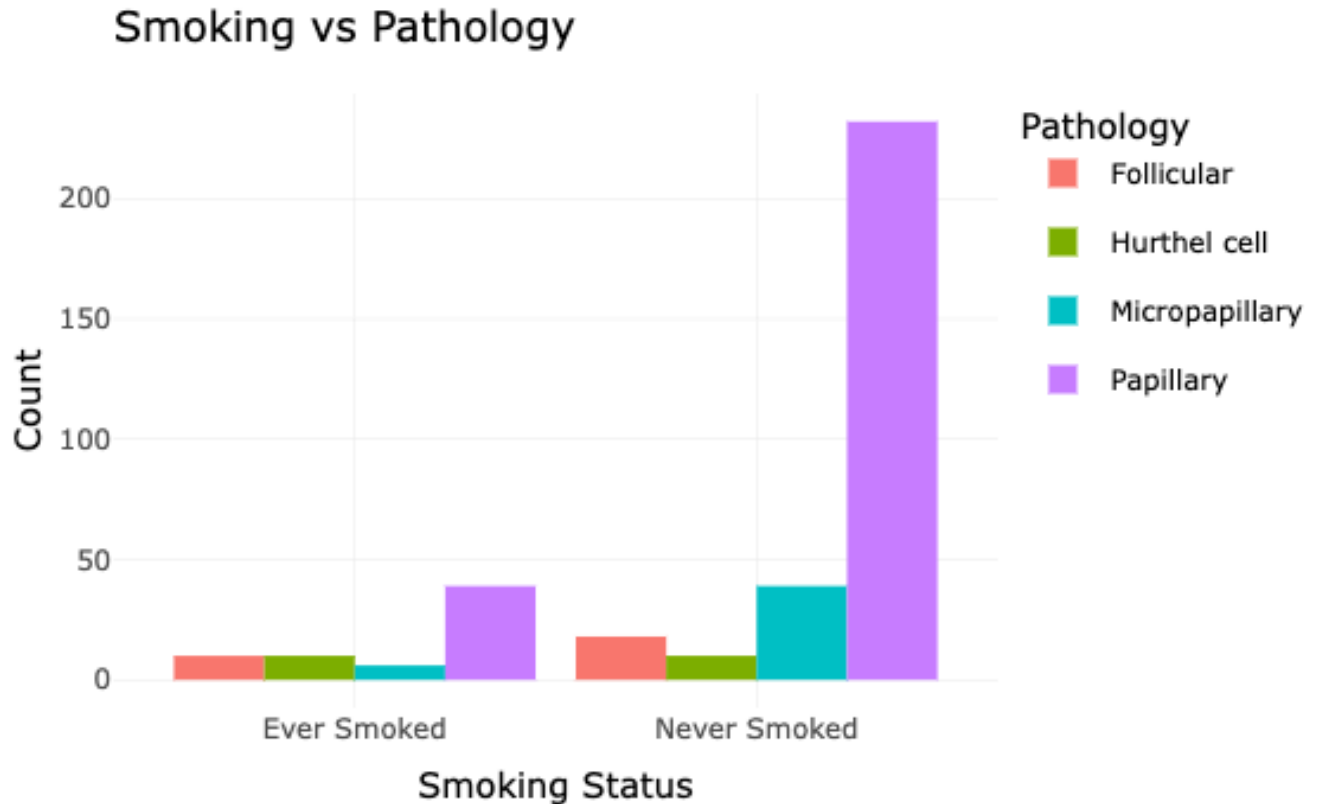


Figure 7: Smoking status and severity of thyroid pathology

The above plot illustrates the clear predominance of "Never Smoked" patients in the category of Papillary pathology while the other pathologies, namely Follicular, Hurthel cell, and Micropapillary, are less well-represented across both "Ever Smoked" and "Never Smoked" groups. There is no overwhelming relationship that would appear to exist between smoking history and the severity of thyroid pathology, because "Ever Smoked" does not seem to have more serious or different pathologies compared to non-smokers.

This may indicate that smoking status is not a very powerful predictor of specific types of thyroid pathology; at least in one of the most common-thyroid cancers-papillary thyroid cancer, it does not appear to be linked with smoking history.

2. Are there differences in cancer stages between current smokers and those with a history of smoking?

E: Current smokers are expected to be found to have more advanced cancer stages compared to those with a history of smoking.

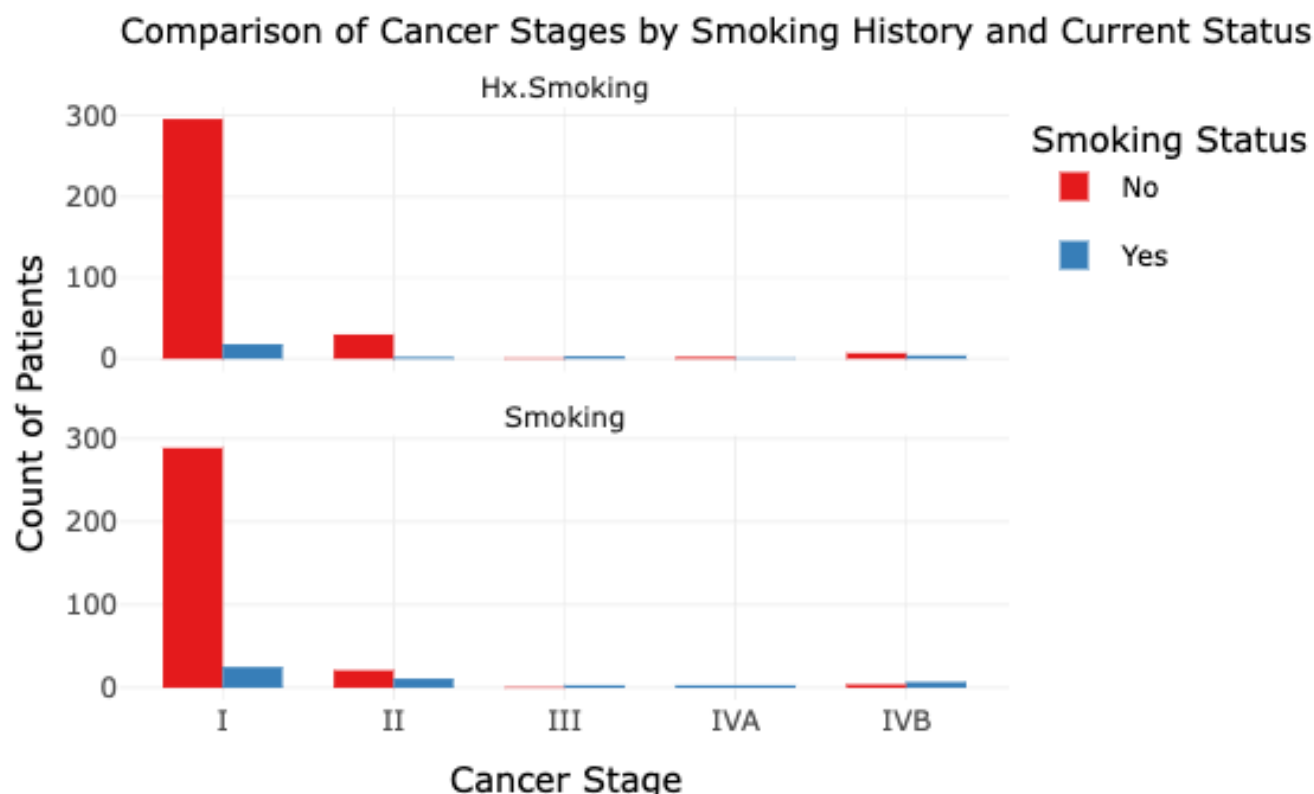


Figure 8: Cancer Stage by Smoking History and Current Status

This bar chart illustrates the distribution of tumor cancer stages among patients with or without a history of smoking and current smokers. Indeed, it gives a different trend for five stages, such as I, II, III, IVA, and IVB. Most patients with a history of smoking were diagnosed at earlier stages, while the numbers decreased considerably through the later stages, showing an earlier stage of diagnosis in former smokers. At all stages, the current smokers are highly sparse, with just a couple of patients in each category; this may indicate that either the sample size is smaller or the trends show that current smokers are diagnosed less often or even later in stages due to late seeking of health care or symptom recognition. This disparity in stage distribution suggests a differential timing of detection of cancers or access to health care between former and current smokers and therefore suggests the possible need for more active health monitoring with earlier intervention for current smokers.

3. What impact does previous radiotherapy have on current thyroid function status?

E: Previous radiotherapy can negatively impact current thyroid function, often leading to hypothyroidism or other thyroid dysfunctions due to radiation-induced damage to the thyroid gland

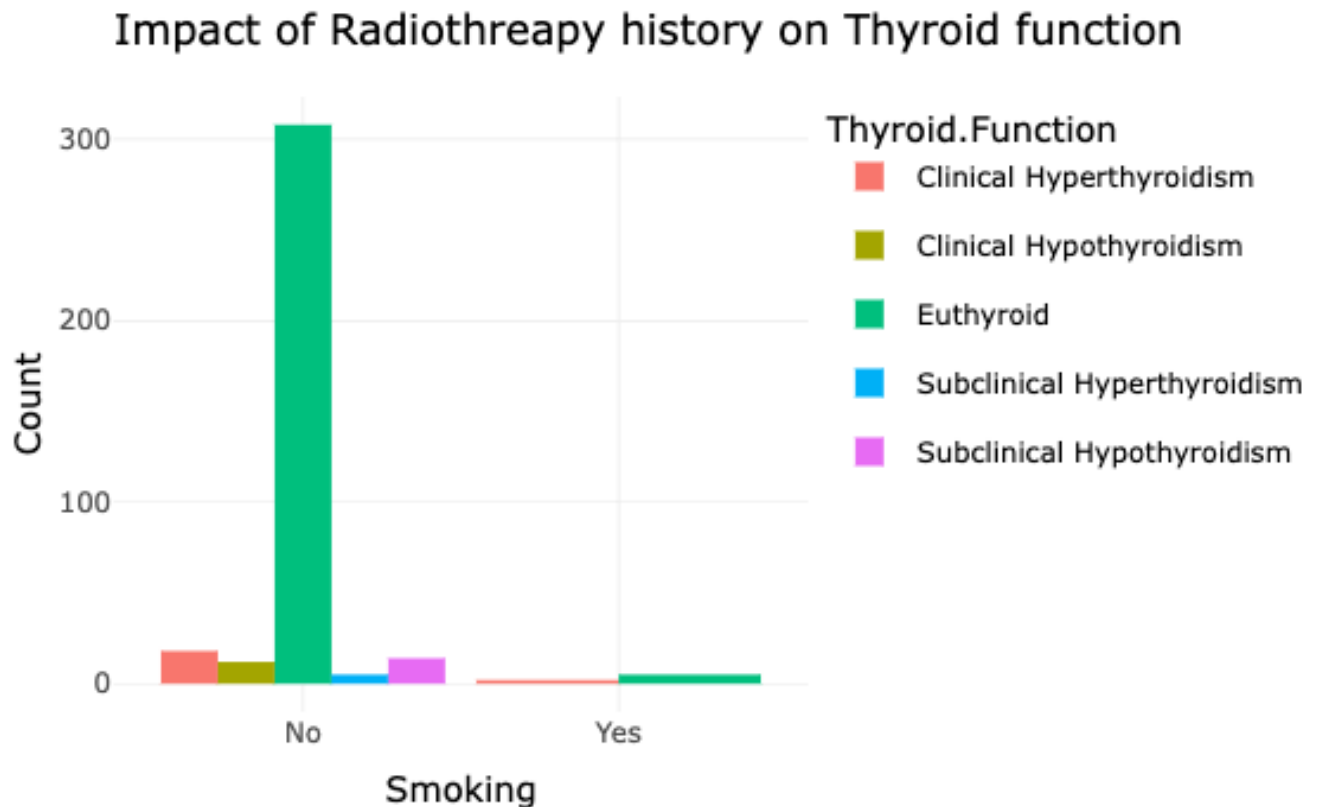


Figure 9: Radiothreapy history on Thyroid Function

The graph demonstrates that, among the cases with no radiotherapy history, the majority of cases were found to be Euthyroid, or of normal thyroid functioning status. In contrast, those experiencing radiotherapy show a greater distribution across various thyroid statuses such as Clinical Hypothyroidism, Clinical Hyperthyroidism, and Subclinical Hypothyroidism.

This pattern shows that most of the patients with no history of radiotherapy are normal, while most of those with a history of radiotherapy might predispose them to thyroid dysfunction.

4. How does radiotherapy history affect the risk of thyroid cancer recurrence?

E: A history of radiotherapy increases the risk of thyroid cancer recurrence, as radiation exposure can contribute to long-term thyroid gland damage

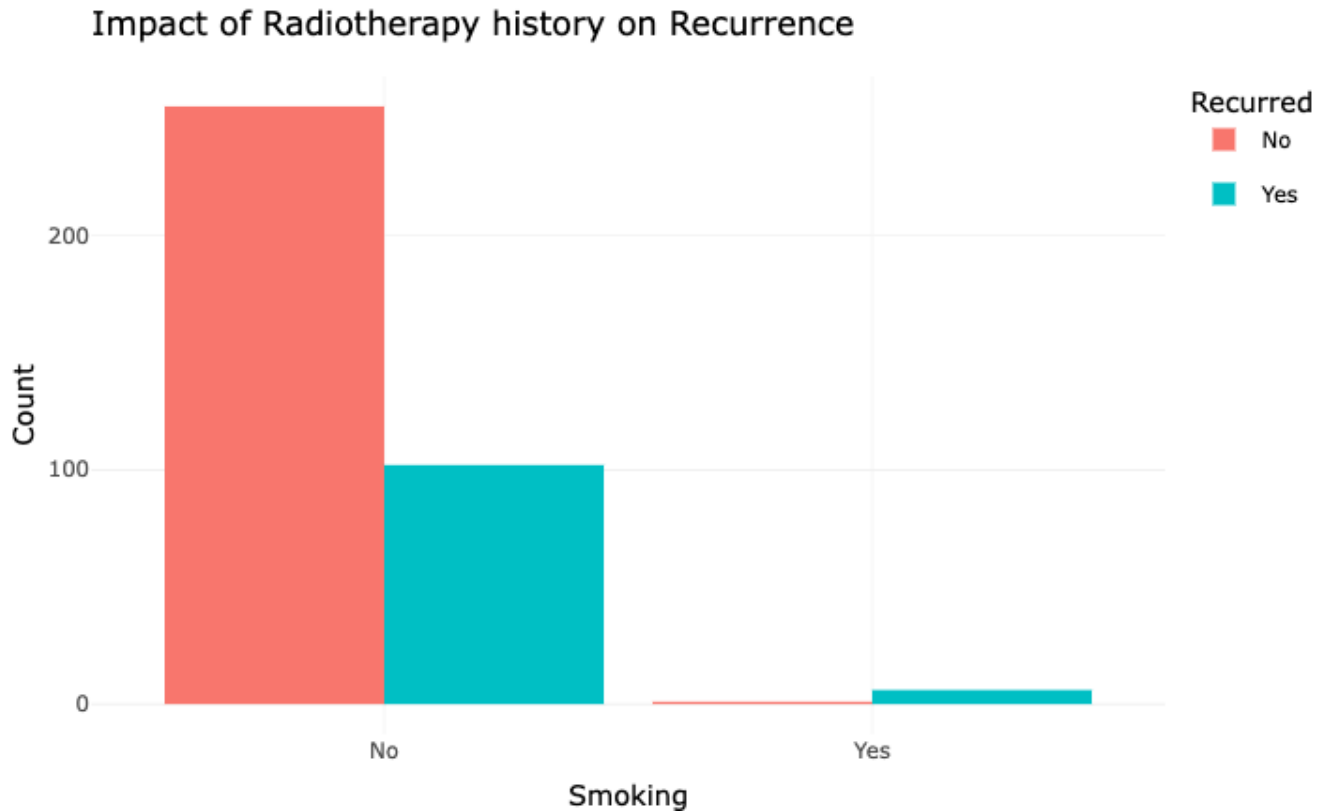


Figure 10: Impact of Radiotherapy history on Recurrence

This graph illustrates that patients without a history of radiotherapy had the most frequent rate, although there is a disproportionate percentage who experienced recurrence. In contrast, the patients who have a history of radiotherapy are at an overall count far lower but where the recurrence is more equally distributed, having a slight increase in cases where there was a recurrence compared to ones that did not experience it.

It shows that radiotherapy does not ensure the recurrence of thyroid cancer, but in patients with a previous history of radiotherapy, there is a slight increase. The overall recurrence rate is higher in radiotherapy patients compared to non-radiotherapy patients.

4.1.3 Medical Analysis

1. Which thyroid function is most prevalent ?

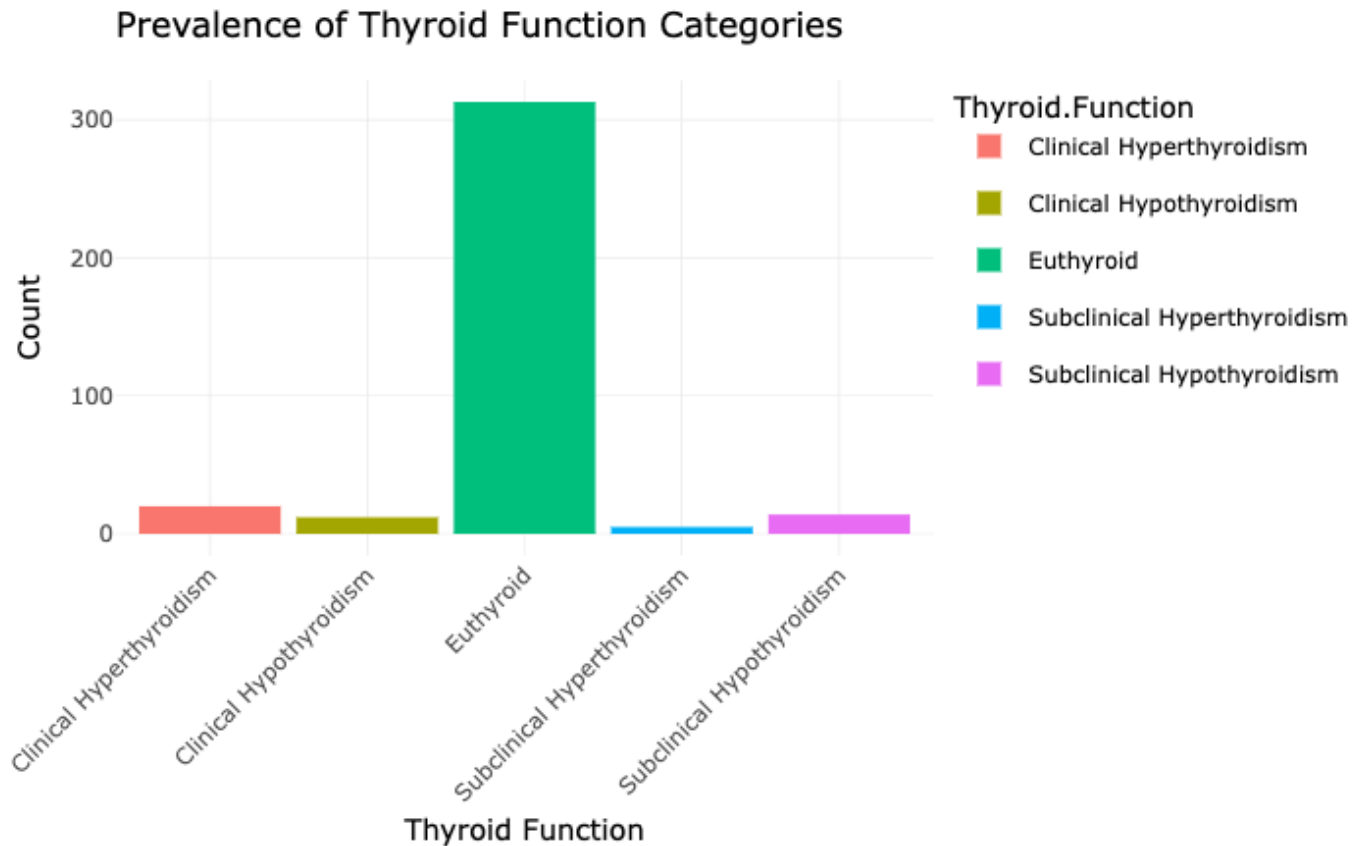


Figure 11: Distribution of Thyroid Function

The above graph shows that most of the people fall into the category of Euthyroid, which means normal thyroid function. The rest of the categories of thyroid function include Clinical Hyperthyroidism, Clinical Hypothyroidism, Subclinical Hyperthyroidism, and Subclinical Hypothyroidism, and they have very negligible counts. This can be interpreted to mean that thyroid dysfunction is way less common when compared to normal thyroid function in the population under study.

Thyroid disease affects cardiovascular function significantly, altering cardiac output, contractility, blood pressure, and systemic vascular resistance, making early detection critical for preventing associated complications.”

This therefore suggests that the normal functioning of the thyroid is the prevailing thyroid status among the respondents; hence, the larger proportion is generally healthy in terms of their thyroid. On the contrary, the few thyroid dysfunctions recorded are truly a reflection of the presence of such conditions; however, these are quite rare.

2. Is there a relationship between thyroid function and the physical examination findings ?

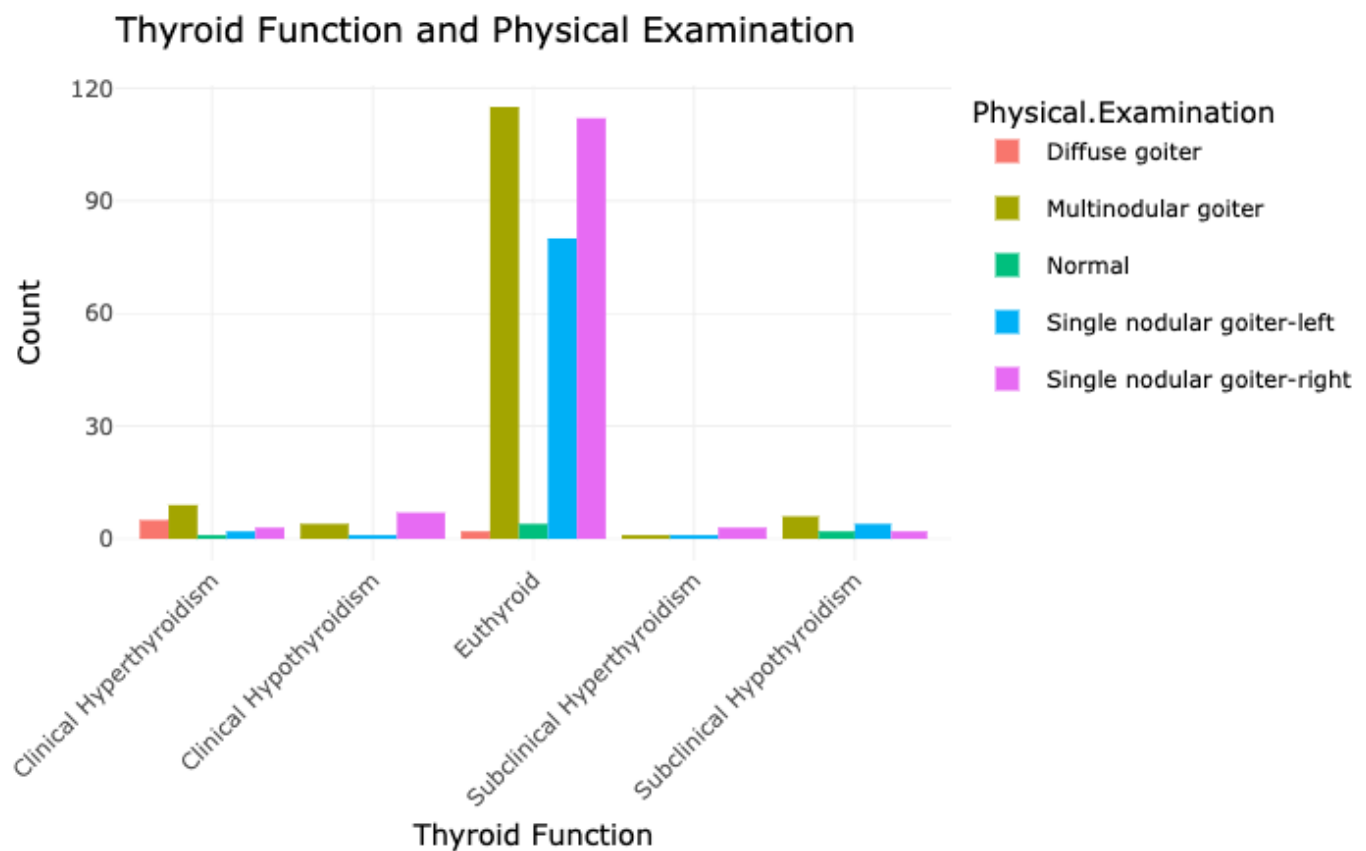


Figure 12: Relationship of Thyroid Function and Physical Examination

The above graph categorizes the patients based on thyroid function and then correlates these with the findings on physical examination. The majority of the patients fall under Euthyroid status, meaning normal thyroid status, and most of them do not usually have physical manifestations of thyroid problems, as can be seen in the high magenta bar under the normal category. This suggests that normally, a thyroid function that is normal is associated with a thyroid gland that is unremarkable on physical examination. At the same time, Clinical Hyperthyroidism and Clinical Hypothyroidism are significantly associated with diffuse and multinodular goiters, respectively, a fact indicated by the big yellow and green bars in these categories.

This therefore shows that thyroid dysfunction-both an overactive and underactive gland-is often physically manifested as several types of goiters.

3. Do specific thyroid functions correlate with higher cancer stages at diagnosis ?

E: It is expected that the majority of euthyroid individuals are diagnosed at stage I of cancer, showing a significant concentration in the earliest stage of the disease.

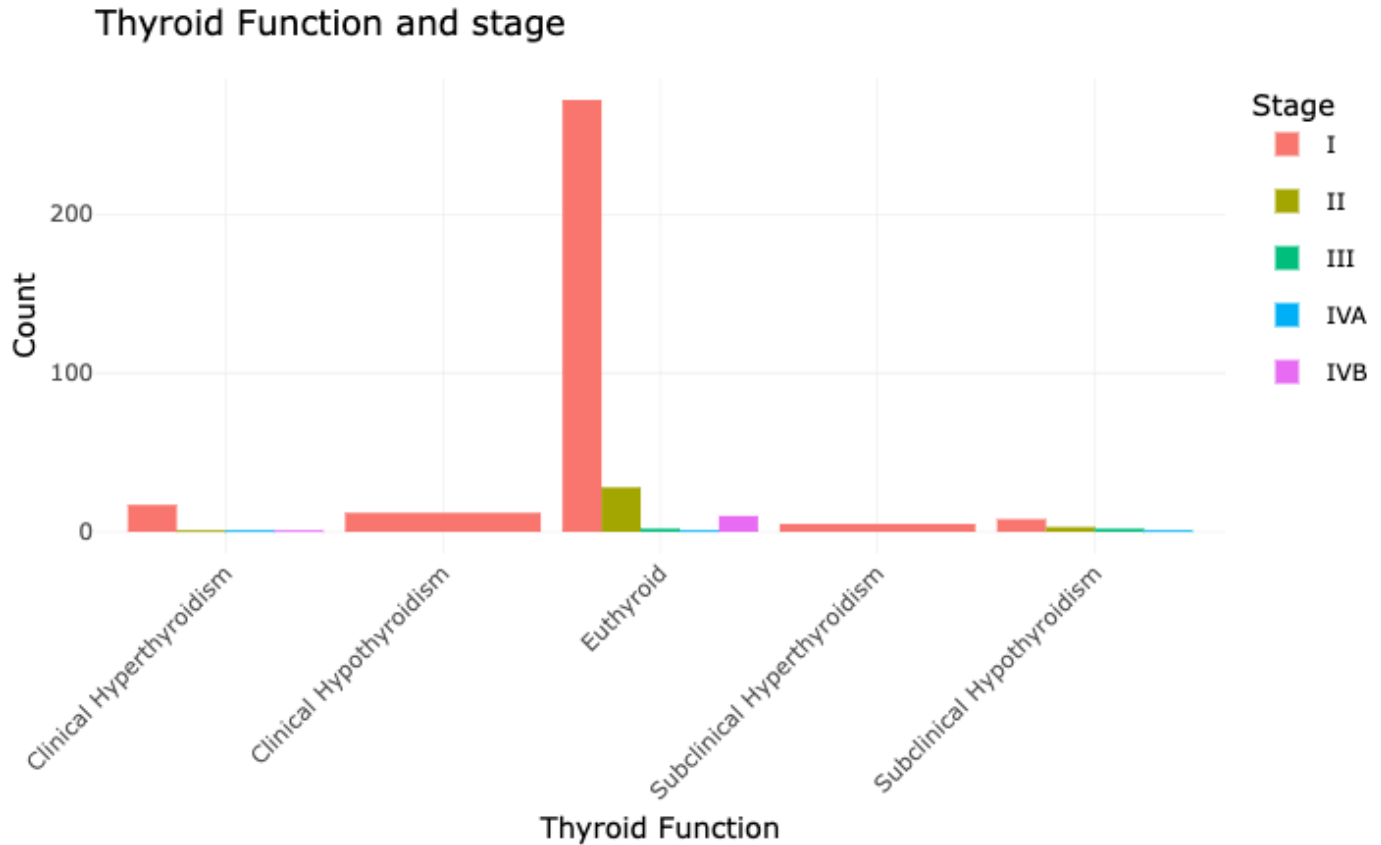


Figure 13: Thyroid Function as compared to Stage

The bar chart describes the distribution of different thyroid functions within various stages of cancer. The horizontal axis categorizes thyroid functions, while the vertical axis measures the number of cases identified in each stage of cancer. The most striking feature is the red bar under the Euthyroid category at stage I, reflecting that normal thyroid-function patients are mainly diagnosed at early stages, probably due to regular screening or a diagnosis at an early stage.

This graph also reflects minimal counts for other thyroid function categories across all cancer stages and has no significant peaks to show that specific thyroid dysfunctions have no direct relation to the diagnosis at more advanced stages of cancer. This absence of a trend is important in gathering insight into the implications brought about by thyroid health in understanding cancer progression and management.

4. How does thyroid function vary with patient age and gender?

E: Thyroid function often changes with age and is more commonly problematic in women and older adults.

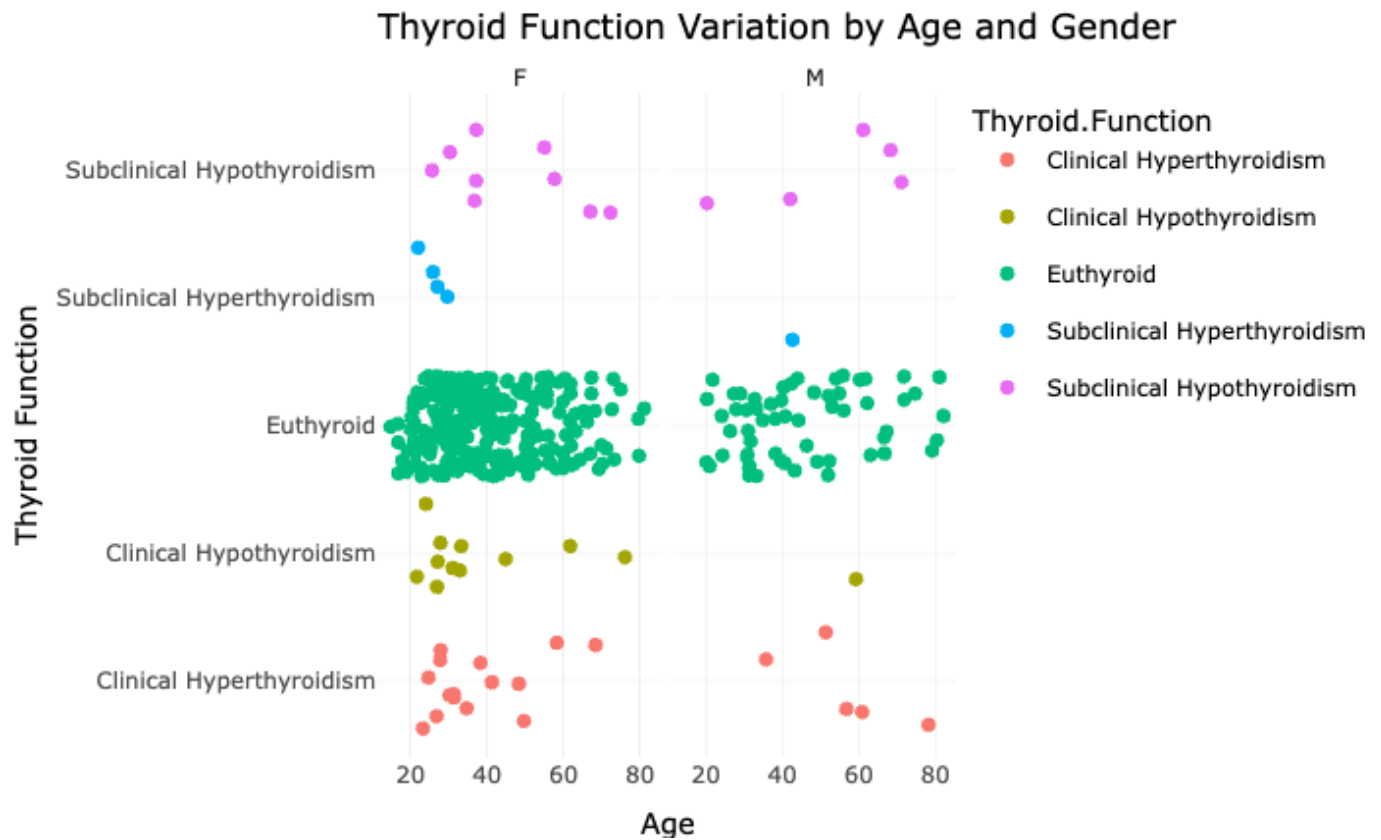


Figure 14: Thyroid Function by Age and Gender

The above scatter plot, "Thyroid Function Variation by Age and Gender", shows the overall thyroid function in the different age groups by gender. There is a big cluster of points in the category "Euthyroid" that encompasses pretty much most age brackets for both males and females. This reflects the fact that normal thyroid status is maintained across all ages, from young adults to the elderly. On the other hand, subclinical thyroid dysfunctions demonstrate mild female preponderance and a higher prevalence among younger and middle-aged women. Clinical Hyperthyroidism and Hypothyroidism conditions are relatively rare but with a very different pattern; most notably, Clinical Hypothyroidism is found more in older females, suggesting a vulnerability with age.

4.1.4 Clinical Feature

1. How does the presence of adenopathy correlate with thyroid cancer staging ?

E: The presence of adenopathy, can indicate a higher stage of thyroid cancer suggesting that the cancer may have spread beyond the thyroid gland.

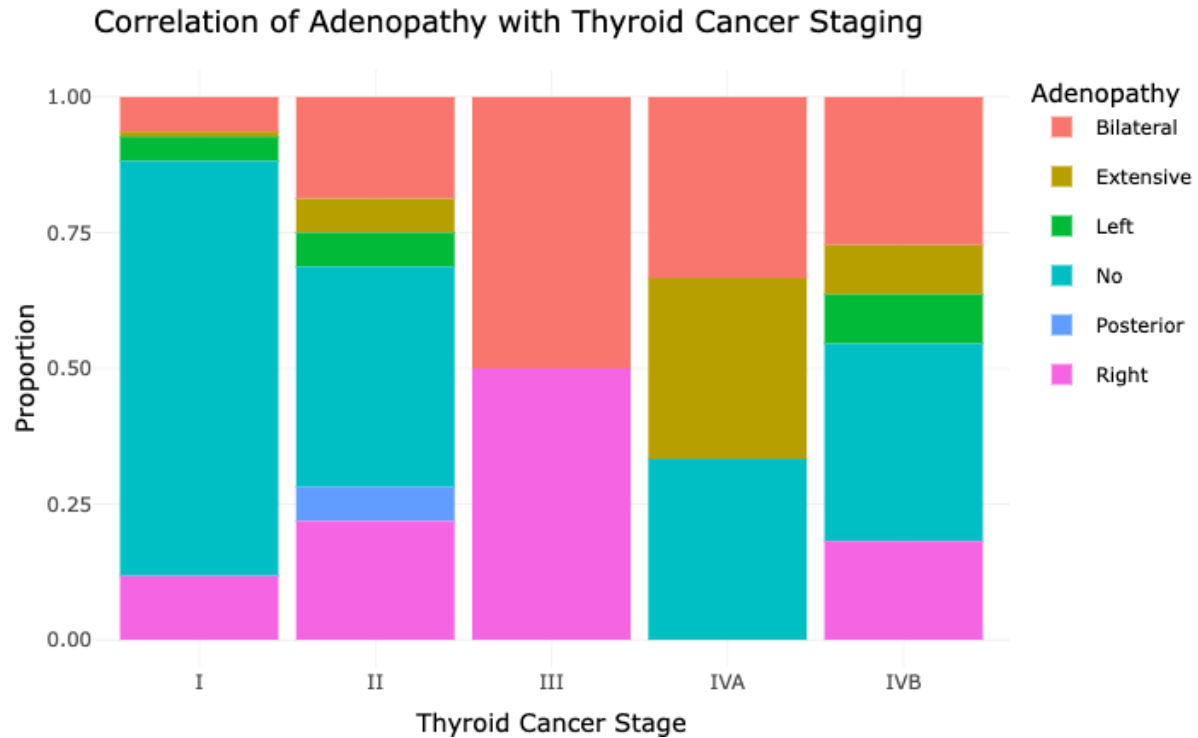


Figure 15: Corelation of Thyroid stage and Adenopathy

The following chart illustrates how the presence and type of adenopathy relate to various thyroid cancer stages. Note that the severity of adenopathy progresses with an increase in stage of cancer. For the earliest stages, I and II, most patients have no adenopathy, reflecting minimal lymphatic involvement. However, at Stage III, there is a drop in those cancer cases without adenopathy and an increase in those with extended adenopathy, showing wider lymph node involvement. During Stages IVA and IVB, the incidence of extended adenopathy becomes overwhelmingly greater, while the distribution of adenopathy itself becomes relatively even across different specific locations (left, right, bilateral, and posterior).

This would suggest that the lymphatic system in advanced stages of thyroid cancer is highly involved with the disease, indicating more aggressive disease that might broadly influence the approach to treatment. The trend underlines the possible utility of the extent of adenopathy as a clinical marker in assessing the stage and severity of thyroid cancer.

2. Do patients with adenopathy have higher recurrence rates compared to those without ?

E: Patients with adenopathy typically have higher recurrence rates of thyroid cancer than those without.

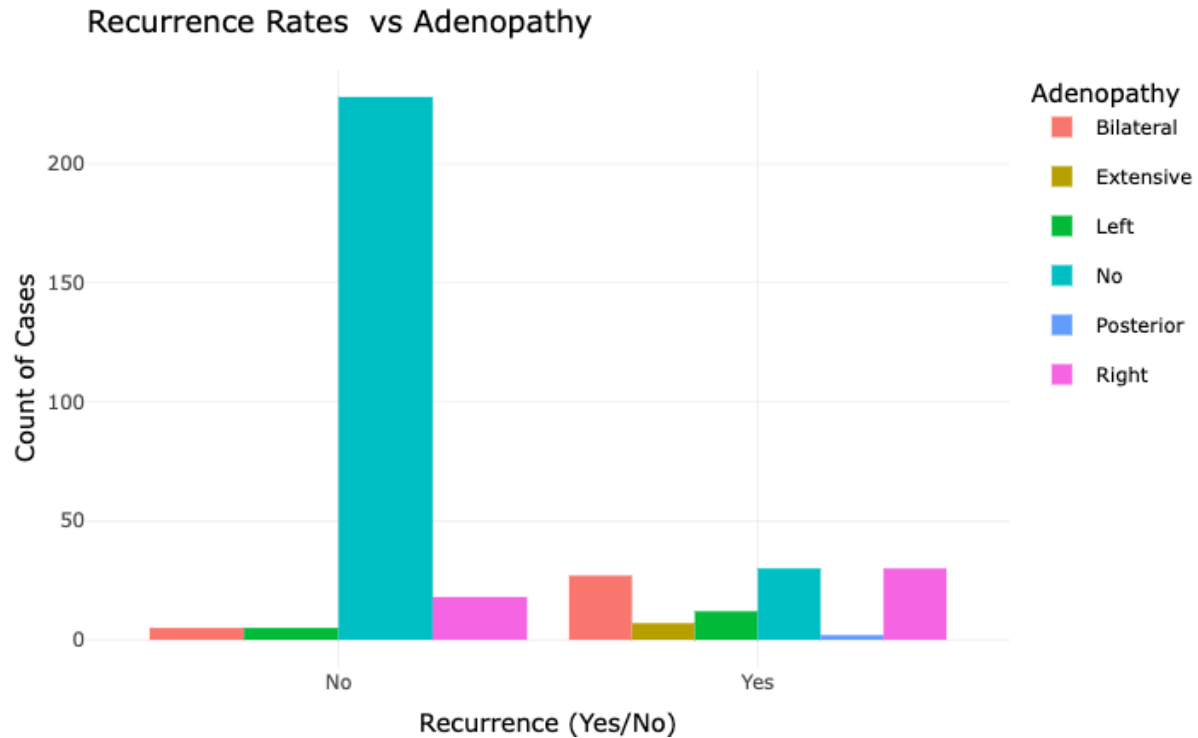


Figure 16: Recurrence Rate vs Adenopathy

This chart shows a bar plot on recurrence rates versus adenopathy in thyroid cancer patients, shows that the majority of those without adenopathy do not experience recurrence, meaning lower risk with the absence of lymph node involvement. In contrast, though the number of cases with adenopathy-whether to extensive, to both sides, the left side, the right side, or the posterior-is smaller, their presence is indicative of higher chances of recurrence. This would mean that the presence of adenopathy-a pointer to the more aggressive or extensive disease-considerably increases the possibility of recurrence and, therefore, requires serious follow-up with such patients.

Therefore, A definite trend within the data could be extracted that adenopathy of any kind is associated with an increased risk of recurrence. This can only mean that efficient lymph node management may hold the key to better long-term results. Thus, early detection and management of adenopathy during treatment planning may hold the key to minimizing risks of recurrence in thyroid cancer.

3. What are the common physical examination findings across different risk categories ?

E: Common physical examination findings in thyroid cancer can vary by risk category: low-risk patients might have a small, whereas high-risk patients may present with multiple nodules.

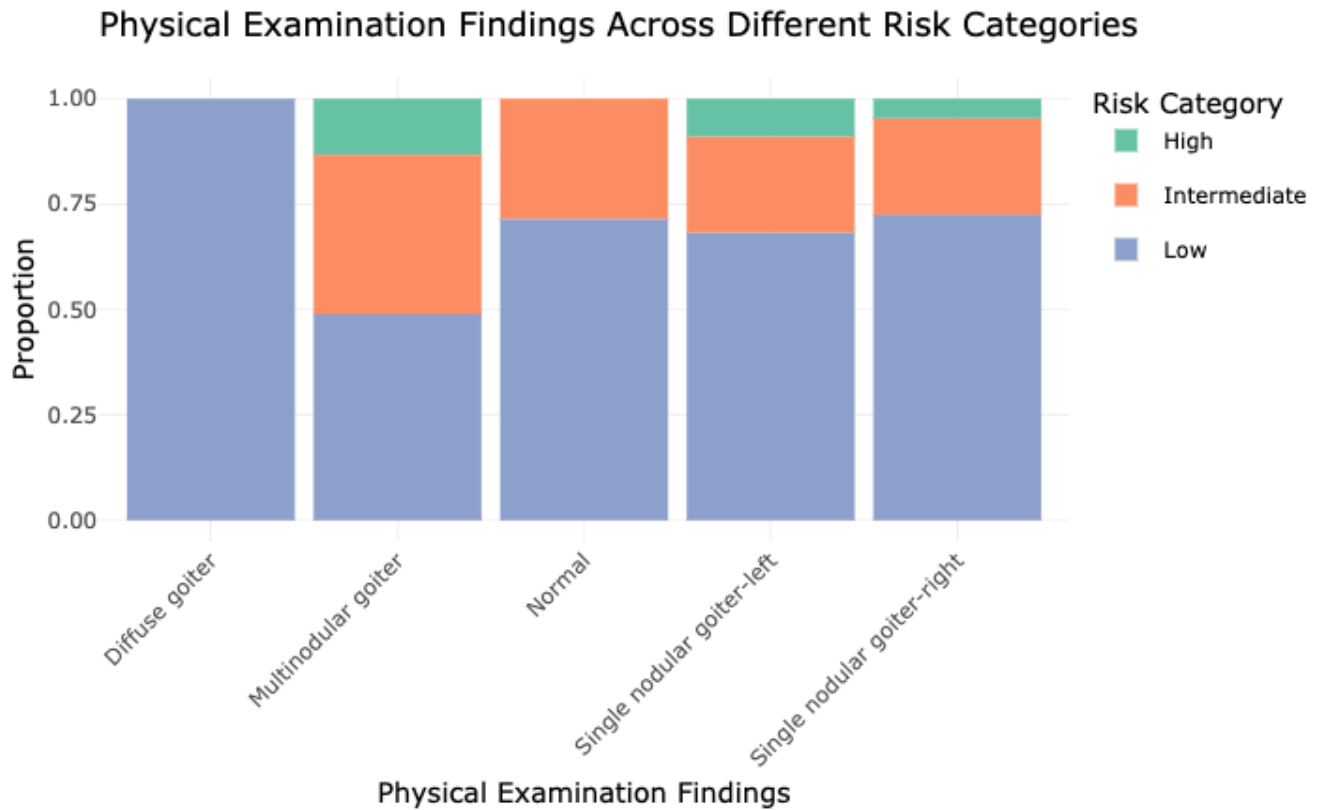


Figure 17: Physical Examination vs Risk

The graph categorizes the findings of the physical examination into various risk levels of a medical condition. It defines the percentage distribution of low, intermediate, and high-risk observations. The findings listed are "Diffuse goiter," "Multinodular goiter," "Normal," "Single nodular goiter-left," and "Single nodular goiter-right." These categories are relatively evenly distributed across the three different risk categories for each finding type. Remarkably, the results of "Normal" are distributed across low, intermediate, and high risks in a relatively uniform distribution, which indicates that normal findings in themselves may not fully correspond to a low-risk category. Concerning "Single nodular goiter-left" and "Single nodular goiter-right," a similar pattern of patients' distribution has been observed, with a slight predominance of the former in the intermediate risk category, possibly reflecting a more nuanced interpretation of risk based on the nodular characteristics. The "Diffuse goiter" and "Multinodular goiter" are similarly well-represented, with a very good representation of the high-risk category—that is, these conditions are more likely to fall under higher medical risks.

4.1.5 Cancer-Specific Analysis

1. How do demographic factors influence the distribution of TNM staging in thyroid cancer?

E: Demographic factors such as age, gender can influence the distribution of TNM staging in thyroid cancer, with younger patients often presenting with lower stages, while older adults might have more advanced stages at diagnosis. Gender differences also exist, with women typically diagnosed at earlier stages compared to men.

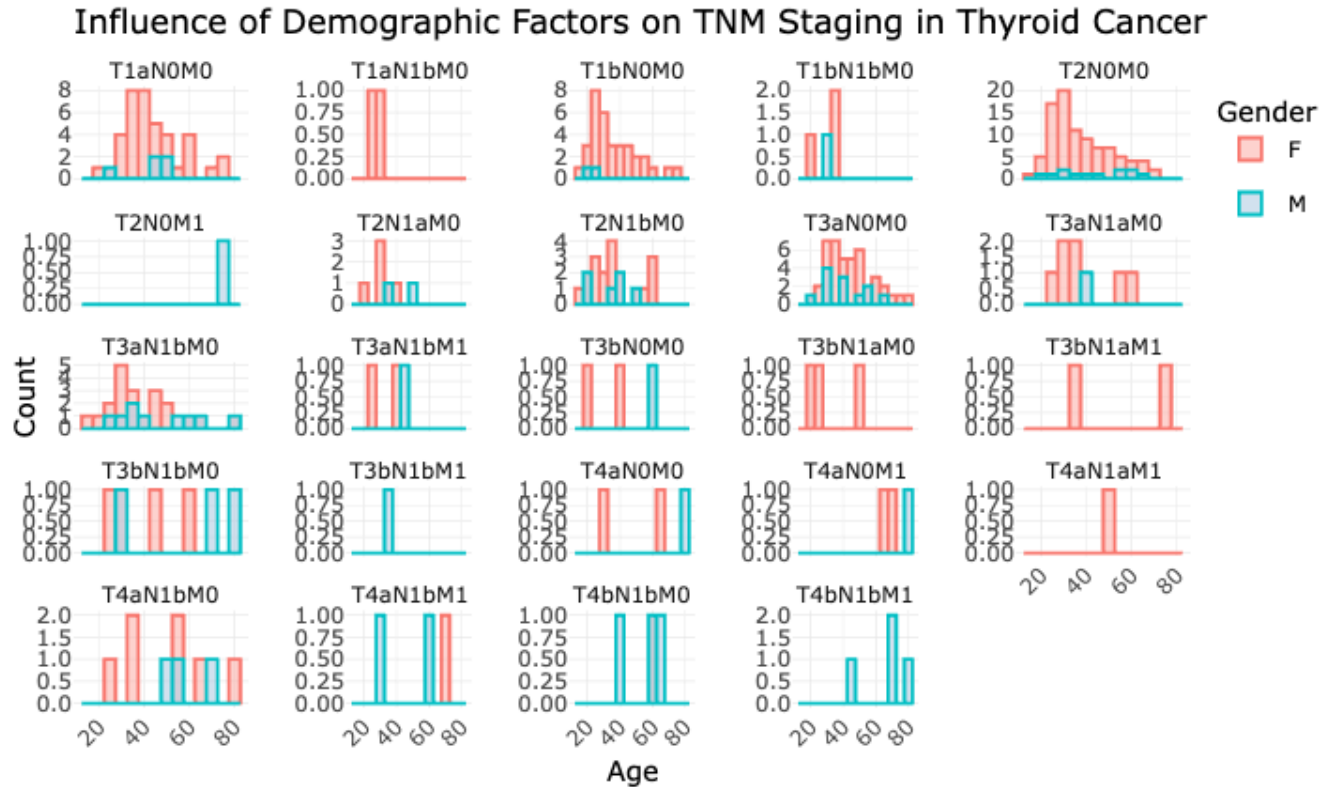


Figure 18: Distribution of TNM staging in thyroid cancer

The above histogram grid shows how the distribution of TNM staging in thyroid cancer patients varies with demographic factors such as age and gender. The TNM classification-one of the important bases for the analysis of the severity, the major areas of concern in any cancer disease: Tumor size (T), Node involvement (N), and Metastasis (M).

The overall trend in these histograms is a sure sign that younger patients, especially those below 40 years old, tend to be diagnosed more frequently in the early stages of thyroid cancer, such as T1N0M0 and T1N1bM0, thus indicating that diagnoses in younger individuals often show less advanced disease. Thus, with increasing age, the trend shifts to the more advanced category stages, indicating that thyroid cancer in older patients is either diagnosed at a later stage or advances differently with increasing age. Also, the trend of gender distribution among these stages indicates that thyroid cancer is more common in females than males. Women also count almost in all stages more than men. It appears that there is generally a higher incidence in females. Notably, even at the advanced stages of T4aN1bM1 and

T4bN1bM1, both genders are affected, with females showing a higher frequency. This pattern underlines consideration of demographic factors in diagnosis, treatment planning, and management of thyroid cancer, reflecting the distinction of epidemiological trends across different age groups and between genders.

2. What are the predominant risk categories in patients with multi-focal thyroid pathology?

E: In patients with multi-focal thyroid pathology, the predominant risk categories often lean towards higher risk due to the presence of multiple tumor focal

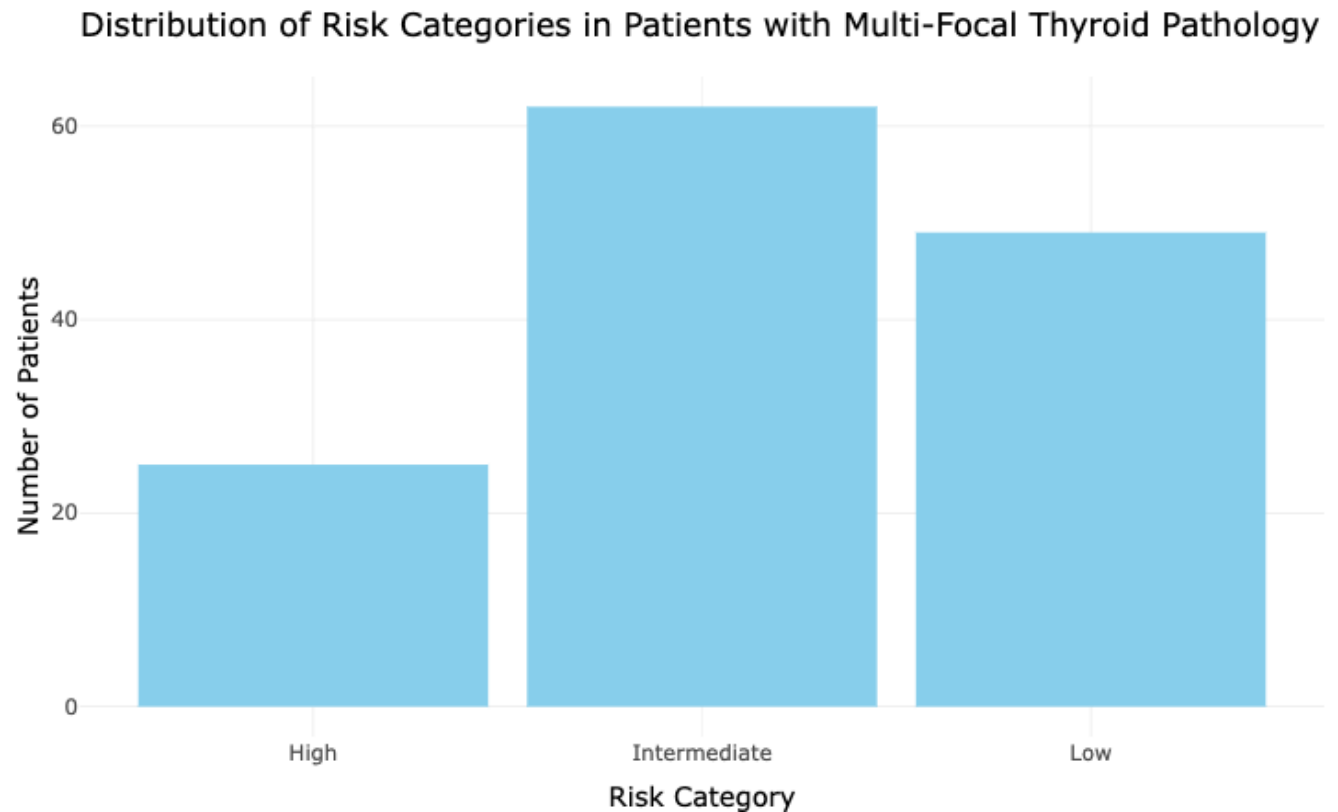


Figure 19: Distribution of Risk categories with Multi-Focal Pathology

The distribution of risk categories in the given bar chart among the patients with multi-focal thyroid pathology would indicate that by far, the highest number of the patients falls into the category of an intermediate risk category; therefore, though serious, it most often presents with such a level of risk which may still enable rather effective management and intervention. Next comes the low-risk category, entailing a great number of patients, which shows that the discovery of multi-focal pathologies is rather early or naturally less aggressive in most instances. The high-risk category has the fewest patients, reflecting situations where the pathology probably has more serious consequences and requires heavier medical interventions. This distribution underlines the relevance of differentiated treatment strategies, considering the level of risk pertinent to each specific patient to obtain the best results.

3. Do certain risk categories correlate with specific pathology types in thyroid cancer?

E: pecific pathology types in thyroid cancer correlate with risk categories; papillary and follicular types are usually lower risk.

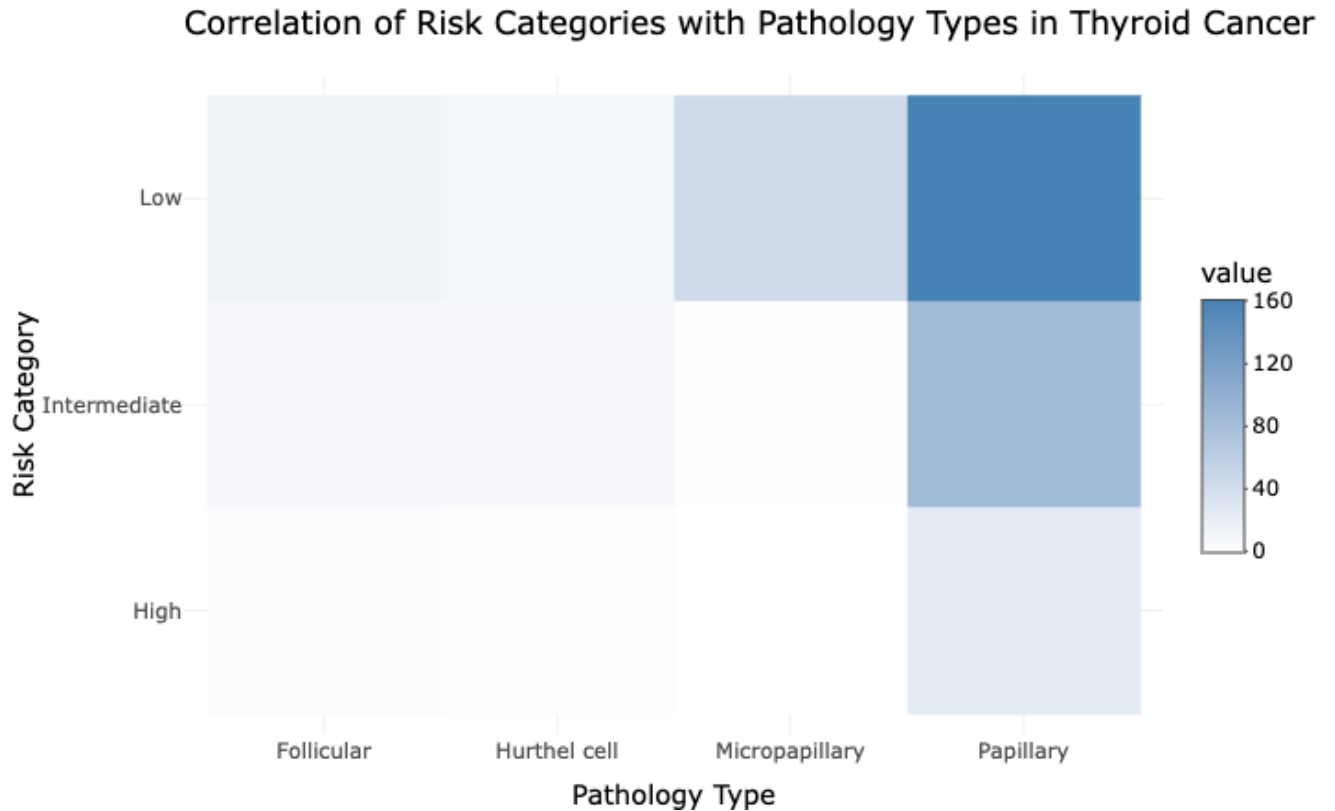


Figure 20: Correlation of Risk and Pathology Type

The heatmap of risk category versus pathology type in thyroid cancer informs on how different pathological variants correspond with each other for varying levels of risk. Follicular pathology, which is known to be potentially clinically significant, predominantly features in the low and intermediate categories of risk, showing that it seldom escalates to a high-risk status to mostly have generally favorable cases upon early detection. The pathology of Hurthel cell mostly falls into the category of intermediate risk, reflecting its capability to present more serious challenges than the low-risk types but normally not progressing towards the most severe outcome. Micropapillary pathology, largely confined to the low-risk category, suggests its often indolent nature and good prognosis. In contrast, thyroid papillary carcinoma, the most common thyroid malignancy, is heavily weighted in the intermediate-risk category but also highly represented in the low-risk category, reflecting the generally less aggressive nature but variable prognosis based on individual case factors. The breakdown herein serves to underscore the clinical importance of risk stratification in thyroid cancer and reminds readers that pathology type may heavily influence both treatment approach and intensity of monitoring required.

4. How does cancer stage relate to the age and gender of the patients?

E: Cancer stage often varies with age and gender, with older patients and males typically presenting at more advanced stages compared to younger individuals and females.

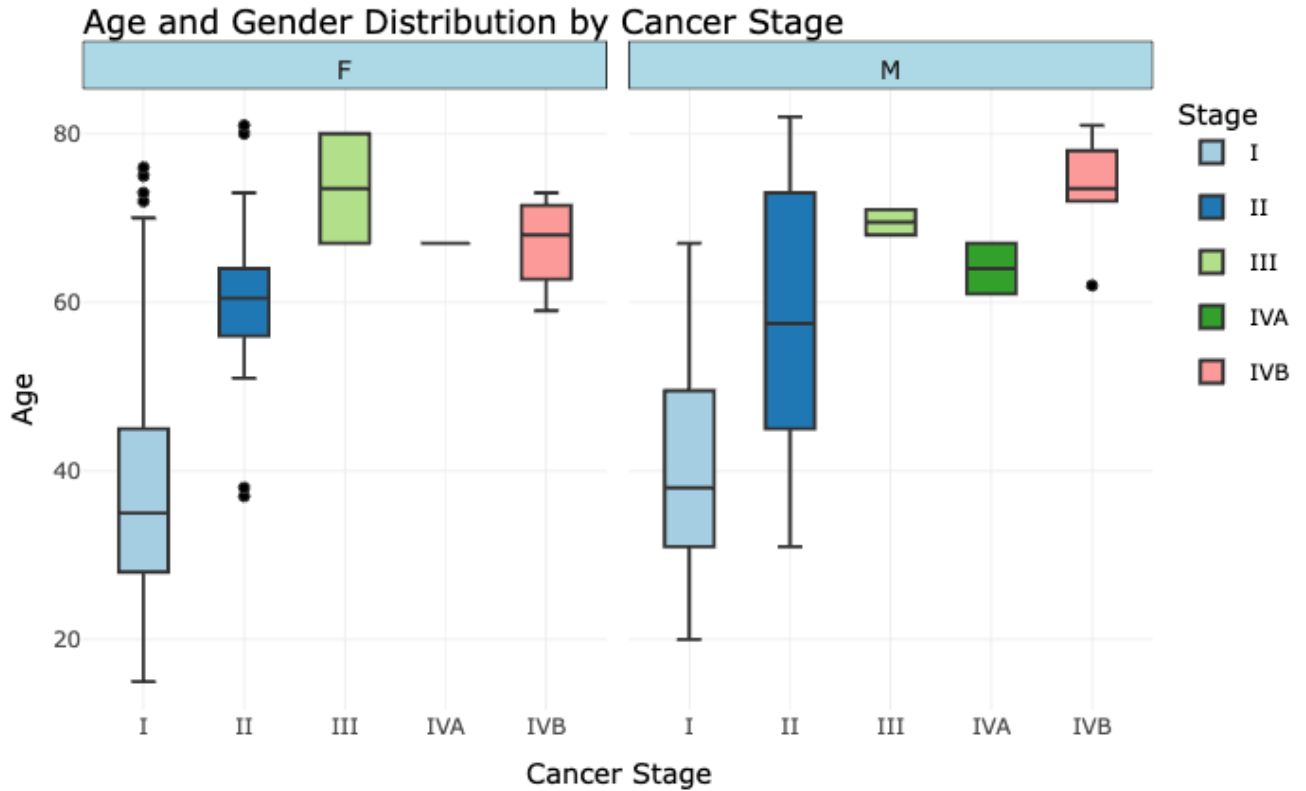


Figure 21: Distribution of Age and Gender across Cancer Stage

This boxplot shows the relationship between cancer stage, age, and gender for patients by showing clearly how these variables interact with each other. For females, it can be observed that the age range is great at Cancer Stage I, with the median approximately at 40 years, indicating that younger to middle-aged females are commonly diagnosed at this initial stage. The trend goes on up to stages II and III, showing a slight increase in median age, hence indicating that the majority of females are affected in these stages in middle age. The age range decreases as the stage goes to IV (IVA and IVB), showing an upward shift, especially in Stage IVB, which indicates that older females are likely to be diagnosed at this stage.

It means that the male distribution normally develops cancer at older ages compared to women. By Stage I, the age median of the male is higher, with an age distribution less diffused. With more advances in the stages of the disease, the distribution is wider, especially Stage IVA, where ages were spanned over a wide range, whereas Stage IVB concentrates around the older population. This trend again suggests the importance of targeted screening strategies, considering age and gender in order to optimize early detection and treatment outcomes.

5. How do gender and pathology type associate with the risk of thyroid cancer recurrence?

E: Recurrence rates of thyroid cancer are expected to be higher in females compared to males, with certain aggressive pathology types also associated with increased recurrence risk.

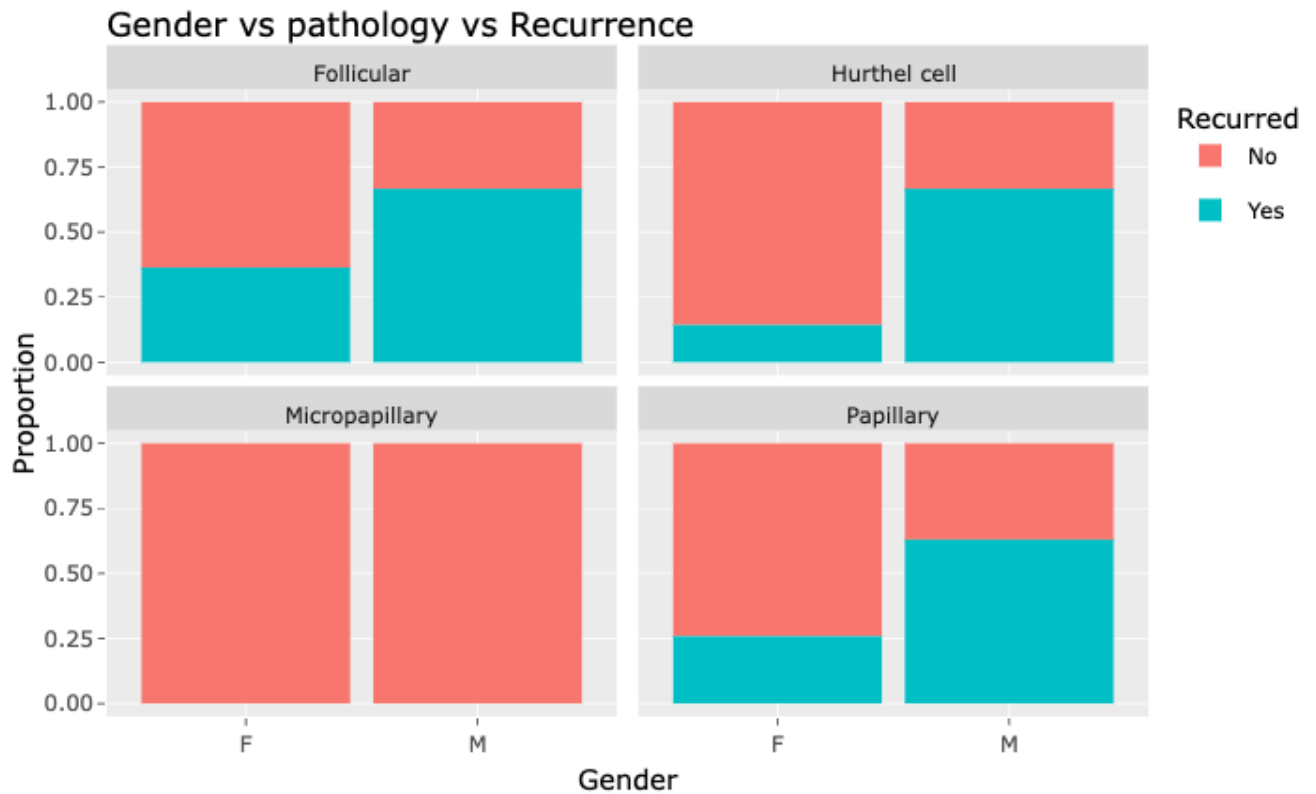


Figure 22: Gender vs pathology vs Recurrence

This bar chart gives the full picture of how gender and pathology type are associated with the risk of recurrence across pathology types: Follicular, Hurthel cell, Micropapillary, and Papillary. The rates of recurrence in Follicular and Hurthel cell pathologies are somewhat even between genders, though a little lower in females in the case of Follicular. In contrast, Micropapillary pathology exhibits a notably higher recurrence rate in males, suggesting a gender-specific vulnerability or perhaps different efficacies in treatment responses. Although generally the pathology of papillary is common to both sexes, men show a small but higher tendency toward recurrence. The visualization suggests that the kind of thyroid pathology and gender of the patient together play an important role in recurrence risks and thus require gender-informed approaches in clinical assessment and tailored treatment strategies for the optimization of outcomes.

4.1.6 Treatment Outcomes

1. What treatment responses are observed in different age groups?

E: Younger patients might respond better to treatments due to generally better health and fewer comorbidities.

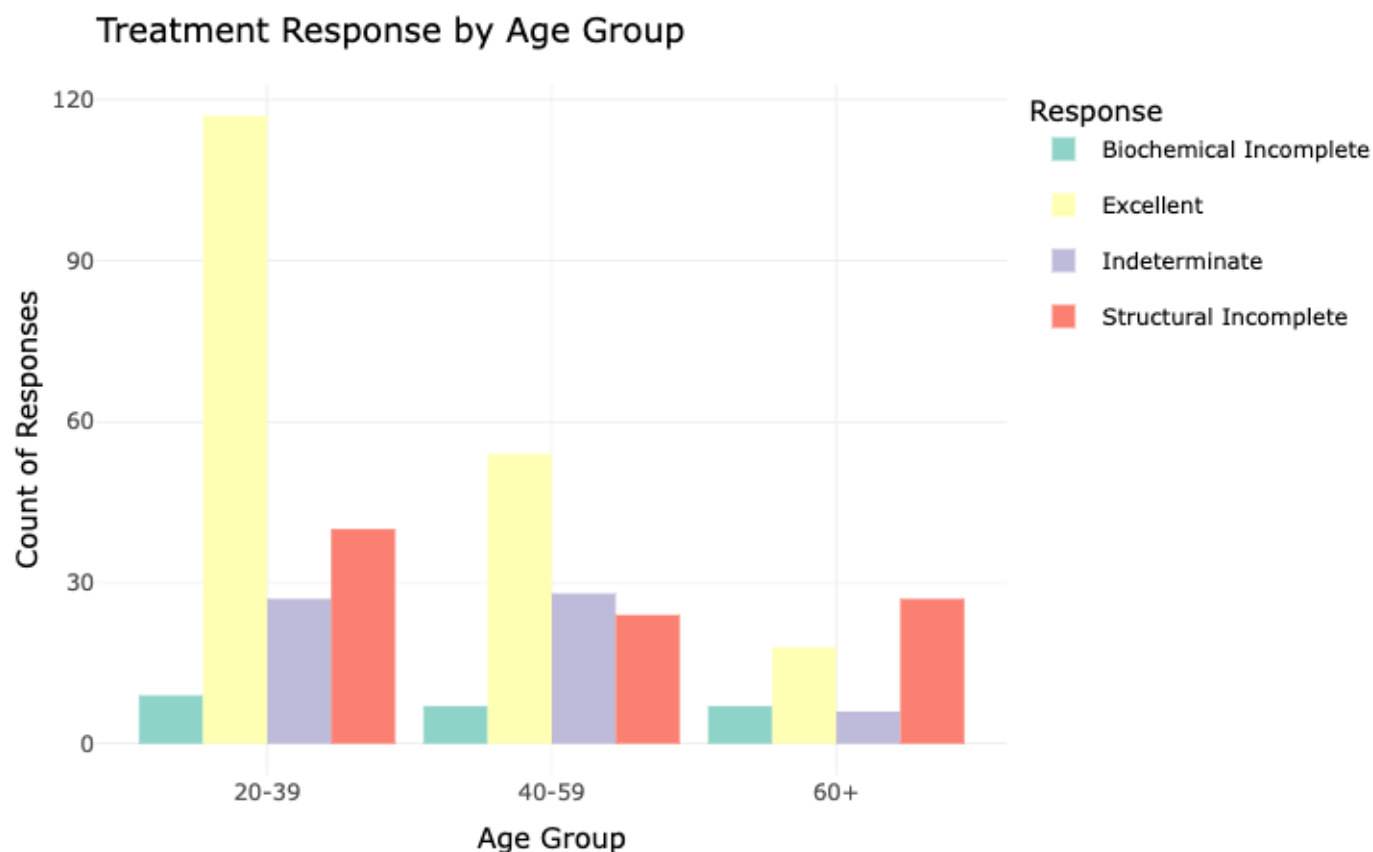


Figure 23: Treatment outcome by Age group

The above bar chart presents treatment responses among subjects for the age groups of 20-39, 40-59, and 60+ for four categories of Biochemical Incomplete, Excellent, Indeterminate, and Structural Incomplete. Treatment responses in the younger patients were essentially "Excellent," infrequent "Biochemical Incomplete" and "Indeterminate," and the "Structural Incomplete" being almost nil for better recovery rates in this demographic category. The response pattern in the middle age category 40-59 is increasingly diverse, with a definite increase in "Biochemical Incomplete" responses, indicating some residual activity of their disease post-treatment. There is also "Indeterminate" and "Structural Incomplete" responses that may show more complicated health scenarios maybe because of older age.

Older patients do represent all types of responses, with a slight trend toward having fewer "Excellent" responses and higher incomplete responses, which points out some of the challenges in achieving optimal outcomes with therapy in this age group—possibly due to age-related physiological changes and higher susceptibility to complications. This trend also puts further emphasis on the place of age consideration in treatment planning and follow-up strategy for optimum health outcome at each stage of life.

2. Does Gender influence treatment outcomes?

E: It's expected that male patients may have less favorable treatment outcomes, as studies often indicate that certain biological factors in males can lead to slower recovery and increased risk of complications compared to females.

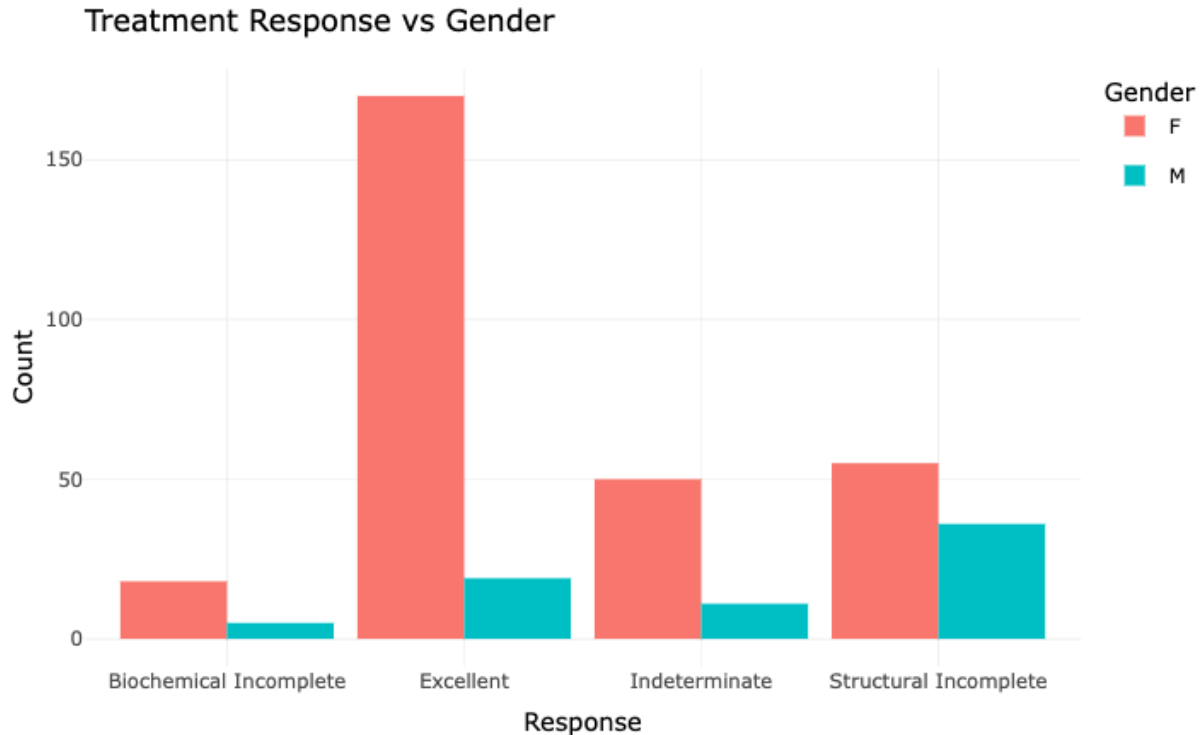


Figure 24: Treatment Response by Gender

The bar chart compares treatment responses between genders graphically; clear patterns are present among the male and female responses to therapy. The females outnumber the males by a great margin in the category "Excellent" response, hence suggesting that females have better overall treatment outcomes. This may suggest, first, either biological differences in disease pathology between genders or variations in how treatments are administered or received.

In contrast, the "Biochemical Incomplete" and "Structural Incomplete" categories have a slight male predilection over females, which could imply a higher residual or more resistant disease in males post-treatment. The "Indeterminate" response is relatively equally distributed between both genders; this would imply that uncertainty in the outcomes of treatment does not necessarily favor one gender over another. This may indicate that gender is a predicting factor for treatment response, as females generally tend to show better response rates in terms of complete response, though males might have a slightly greater tendency toward incomplete responses. In any case, this signals the importance of gender-specific consideration in treatment planning and follow-up.

3. Does Treatment Response Affect Recurrence of Thyroid Cancer?

E: Treatment responses are expected to influence the recurrence of thyroid cancer, with positive responses generally leading to lower recurrence rates compared to higher rates.

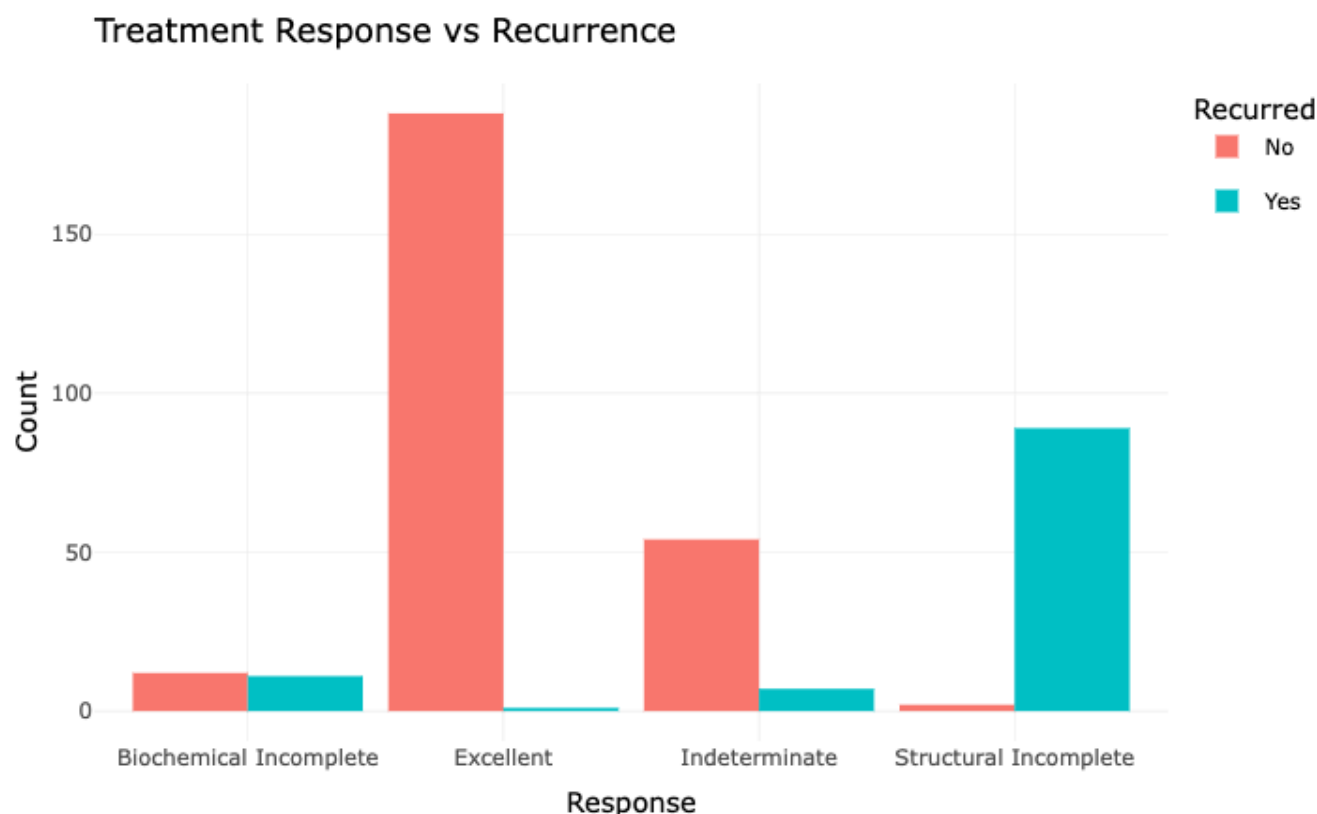


Figure 25: Treatment Response vs Recurrence

The bar graph on the relationship of treatment response to recurrence in thyroid cancer clearly shows that the outcome of initial treatment has much bearing on recurrence rates. The recurrences are by far fewer in patients who obtain an "Excellent" response, which goes to show that successful initial treatment drastically reduces the likelihood of the disease returning. Whereas "Biochemical Incomplete" and "Structural Incomplete" responses have considerably higher recurrence rates, it would seem that incomplete eradication of cancer markers or residual structural disease presence represent high risks for recurrence. The "Indeterminate" response grade, though associated with fewer recurrences than incomplete responses, still carries some risk due to uncertainty in treatment outcomes. Overall, the findings stress the importance of a complete and effective initial response in treatments, as this will minimize recurrence risks of thyroid cancer and bring into focus precise and aggressive management strategies in cases with less-than-excellent initial responses.

4. How does the average treatment response vary by cancer stage when considering different smoking statuses among thyroid cancer patients?

E: Average treatment responses are expected to vary by cancer stage and smoking status, with non-smokers generally exhibiting higher responses across all stages compared to smokers. Additionally, earlier stages of cancer typically demonstrate better treatment responses than advanced stages.

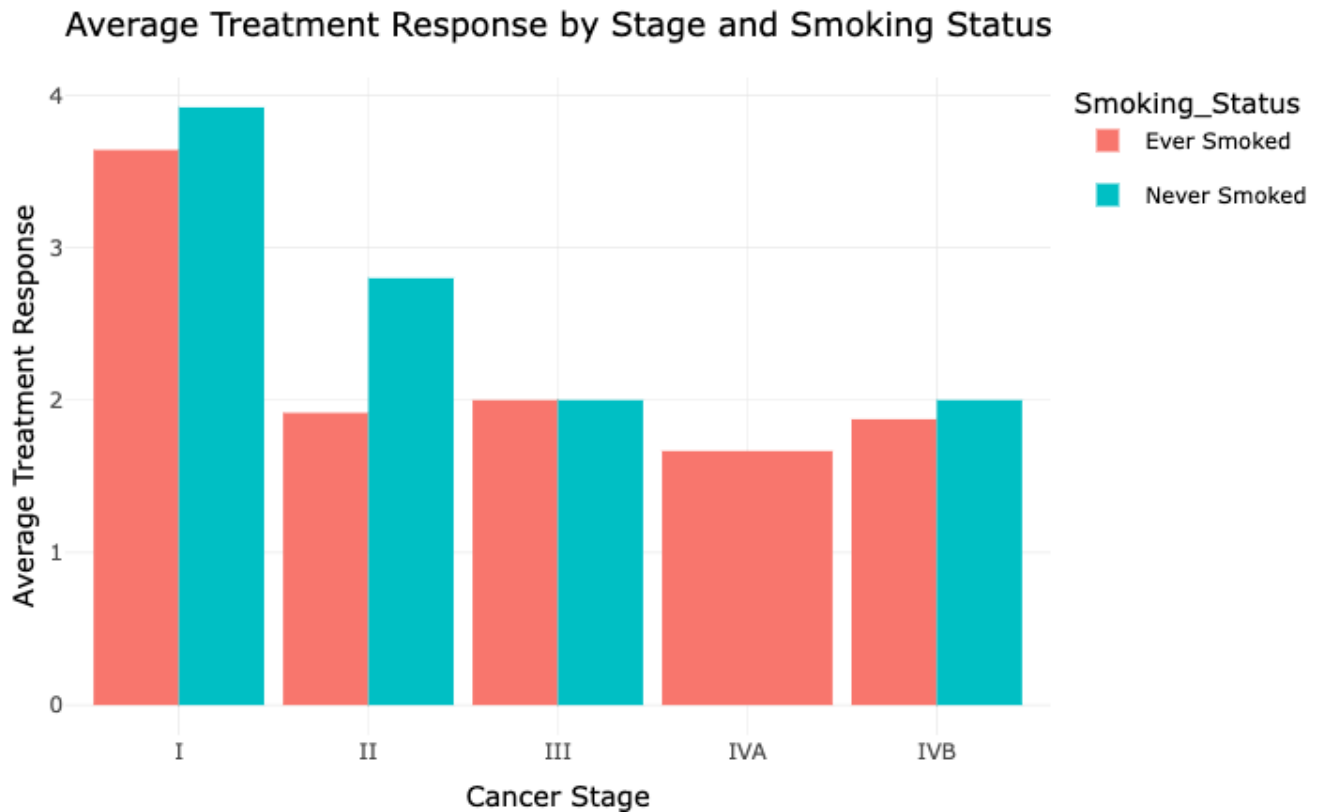


Figure 26: Treatment Response by Stage and Smoking status

The bar chart of average treatment responses in thyroid cancer at different stages, differentiated by smoking status, depicts that early-stage cancer or Stage I responds well to treatment irrespective of the smoking background and hence reflects high treatment effectiveness for both smokers and non-smokers. But in Stage II and III, the drop is much more pronounced for those who have smoked, which might mean that smoking can indeed compromise treatment efficacy in mid-stage cancer. In the advanced IVA and IVB stages, the treatment response between smokers and non-smokers once more converges, albeit at levels lower than in previous stages. This would mean that at the intermediate stages, while treatment outcomes were poorer for smokers, in a case of advanced cancer, the gravity somewhat reduces the differential effects of smoking, reflecting complexity in the interaction of lifestyle variables and cancer progression.

5. What relationships exist between treatment outcomes and thyroid function categories?

E: Treatment outcomes are expected to be more favorable in patients with normal thyroid function compared to those with hypothyroidism or hyperthyroidism, as well-controlled thyroid levels are linked to better responses and fewer complications.

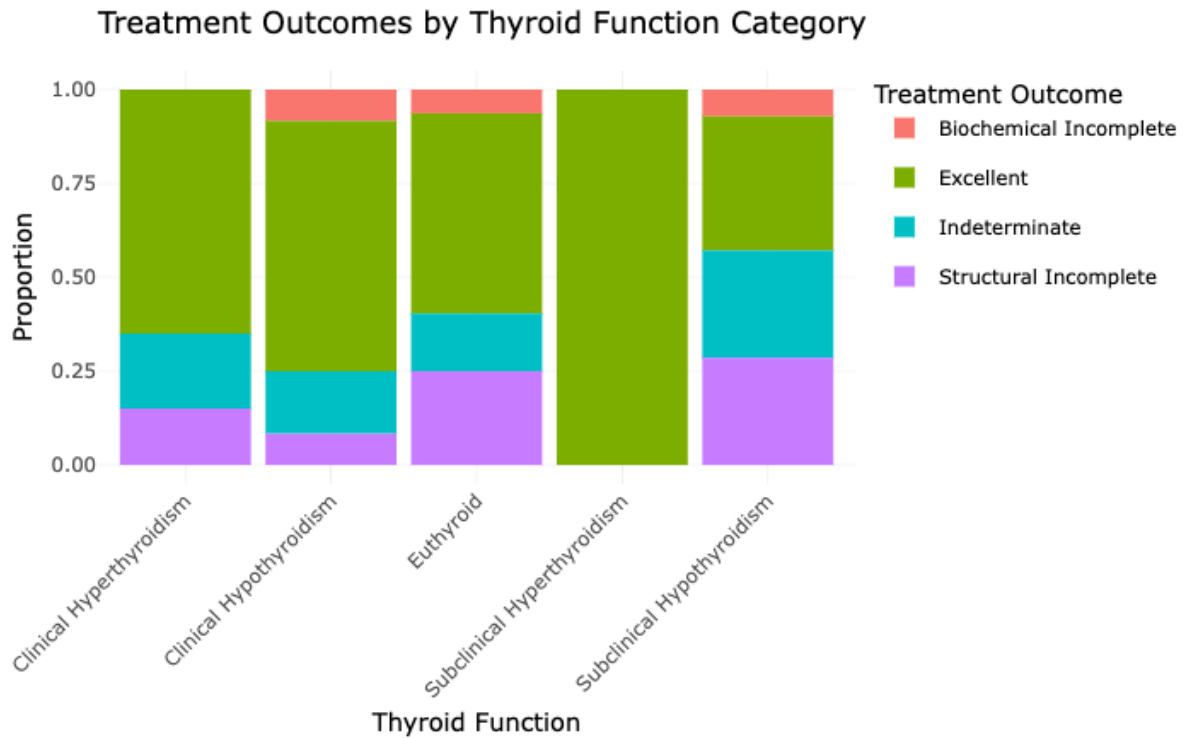


Figure 27: Treatment Response by thyroid function

The stacked bar plot of treatment outcome by thyroid function status reflects a different effect of treatment. Notably, for both overt clinical thyroid dysfunctions—high and hypothyroidism—and for the euthyroid state, the "Excellent" category is higher in absolute levels, making it a general hypothesis that treatments work particularly well when the dysfunction of the thyroid is clear or not present. By contrast, the subclinical conditions are those that present milder and often less symptomatic forms of thyroid dysfunction, revealing a greater share of "Indeterminate" outcomes. This pattern shows that subtler manifestations of thyroid dysfunction in subclinical states might perplex the effectiveness of treatments and predictability of outcomes. Moreover, the presence of "Biochemical Incomplete" and "Structural Incomplete" responses in all categories points out the difficulty of the comprehensive response to thyroid treatment and makes the treatment highly individualized with regard to the degree and type of thyroid dysfunction for better therapeutic success and prognosis of the patient.

5 Summary

This was a comprehensive study in which advanced data visualization techniques were used to explain complex relations among demographic factors, lifestyle influences, clinical features, and treatment outcomes in thyroid diseases. Based on extensive analysis of the large dataset, this study made a number of important observations regarding the progress and management of thyroid conditions.

Key Observations:

- **Demographic Analysis:** The findings have indeed confirmed thyroid diseases to affect mainly females and also to be more common in young and middle-aged adults. It would, therefore, imply that targeted screening and prevention measures should be developed to ensure effective thyroid conditions' management in these groups.
- **Lifestyle and History Factors:** Data on life factors such as smoking and radiotherapy revealed their explicit relation to the extent of pathology of the thyroid gland. Patients with a history of smoking or radiotherapy came out with more advanced stages of the disease upon diagnosis; hence, proving the mentioned influences significantly in thyroid health.
- **Medical Analysis:** The medical analyses also indicated that the thyroid stage is closely related to cancer stage, with euthyroid patients being diagnosed while in their early stages of this disease. This suggests that assessment of thyroid function may be important in early detection and intervention in thyroid cancer.
- **Clinical Feature:** In the study, clinical features associated with the disease, such as adenopathy, were highly associated with the cancer staging and recurrence rates of the thyroid. In general, patients presenting with adenopathy showed advanced diseases and had higher recurrence rates, thereby underlining the use of clinical features in assessing disease severity and prognosis.
- **Cancer-specific Analysis:** The study showed how demographic variables like age and gender impact the distribution of TNM staging in thyroid cancer patients. Generally speaking, younger and female participants usually presented with an earlier stage of cancer that underlined demographic factors influencing disease outcomes.
- **Treatment Outcomes:** Response to treatment varied importantly according to different demographic groups. Better outcomes were seen among younger people and nonsmokers. Moreover, the type of thyroid dysfunction had an important influence on the effectiveness of treatment: among euthyroid patients, treatment protocols were more likely to be successful.

6 Future Work

Further research can be done along various lines in the extension of findings from the present study:

- **Longitudinal Studies:** Longitudinal studies may actually enable observation of the disease over time and treatment effectiveness within the same cohort. This may yield temporal patterns and long-term outcomes which may not be evident within cross-sectional studies.
- **Larger Datasets:** Larger datasets from more diverse populations would carry more validity of the initial findings of this study and their generalization in other populations. This could include an international dataset comprising patient data assessing geographic and ethnic differences in presentation and treatment outcomes of thyroid disease.
- **Application of Machine Learning:** Application of machine learning may show predictive values of patient outcomes as complex interactions of studied variables. Predictive modeling may allow the early identification of people at risk and offer personalized treatment programs.
- **Genetic and Molecular Studies:** Genetic and molecular studies in thyroid disorders may attain better understanding of biological underpinning with the possible application of targeted therapies.
- **Assessment of New and Emerging Therapies:** The evaluation of the impact new and emerging therapies will make on thyroid disease outcomes is very important. This would relate not only to new pharmacological interventions but also to novel therapeutic approaches, such as precision medicine and personalized treatment plans based on a genetic profile.

References

- Arthur, J.R. and Beckett, G.J., 1999. Thyroid function. *British Medical Bulletin*, 55(3), pp.658-668. doi:10.1258/0007142991902538.
- Bauer, M., Goetz, T., Glenn, T. and Whybrow, P.C., 2013. Gender differences in thyroid system function: relevance to bipolar disorder and its treatment. *Bipolar Disorders*, 16(1), pp.58-71. doi:10.1111/bdi.12150.
- Cabanillas, M.E., McFadden, D.G. and Durante, C., 2016. Thyroid cancer. *The Lancet*, 388(10061), pp.2783-2795. doi:10.1016/s0140-6736(16)30172-6.
- Cooper, D.S., Doherty, G.M., Haugen, B.R., Kloos, R.T., Lee, S.L., Mandel, S.J., Mazzaferri, E.L., McIver, B., Sherman, S.I. and Tuttle, R.M., 2009. Revised American thyroid association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*, 19(11), pp.1167-1214. doi:10.1089/thy.2009.0110.
- Durante, C., Grani, G., Lamartina, L., Filetti, S., Mandel, S.J. and Cooper, D.S., 2018. The diagnosis and management of thyroid nodules. *JAMA*, 319(9), pp.914-924. doi:10.1001/jama.2018.0898.
- Vanderpump, M.P., 2011. The epidemiology of thyroid disease. *British Medical Bulletin*, 99(1), pp.39-51. doi:10.1093/bmb/ldr030.