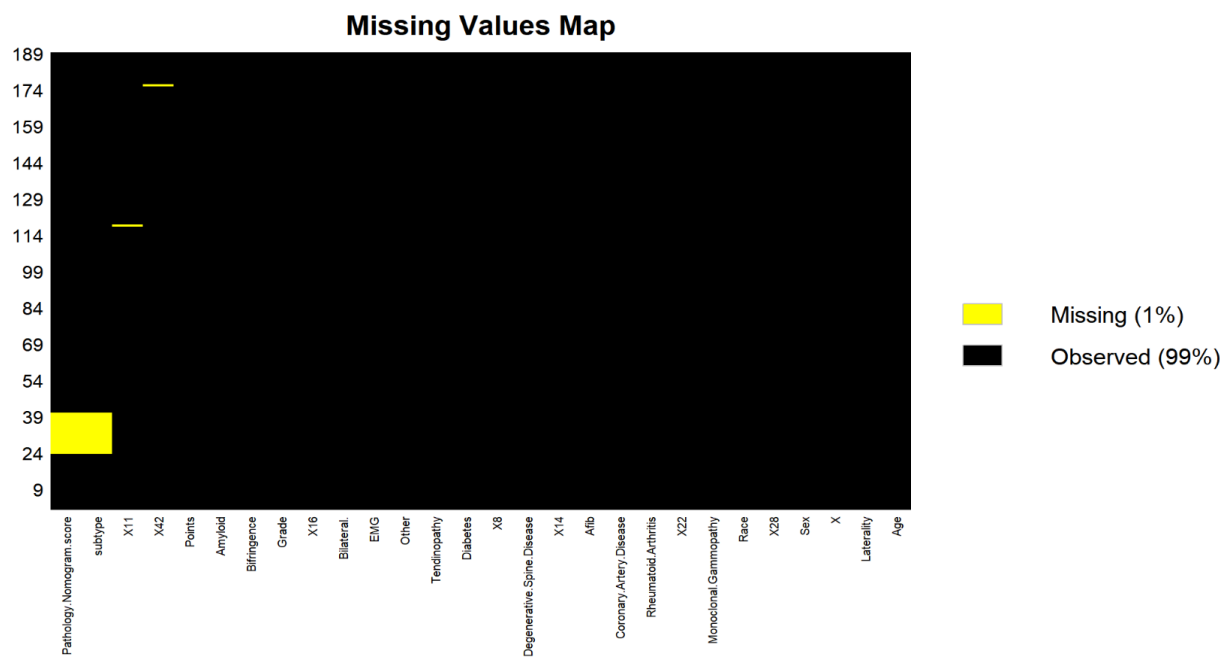
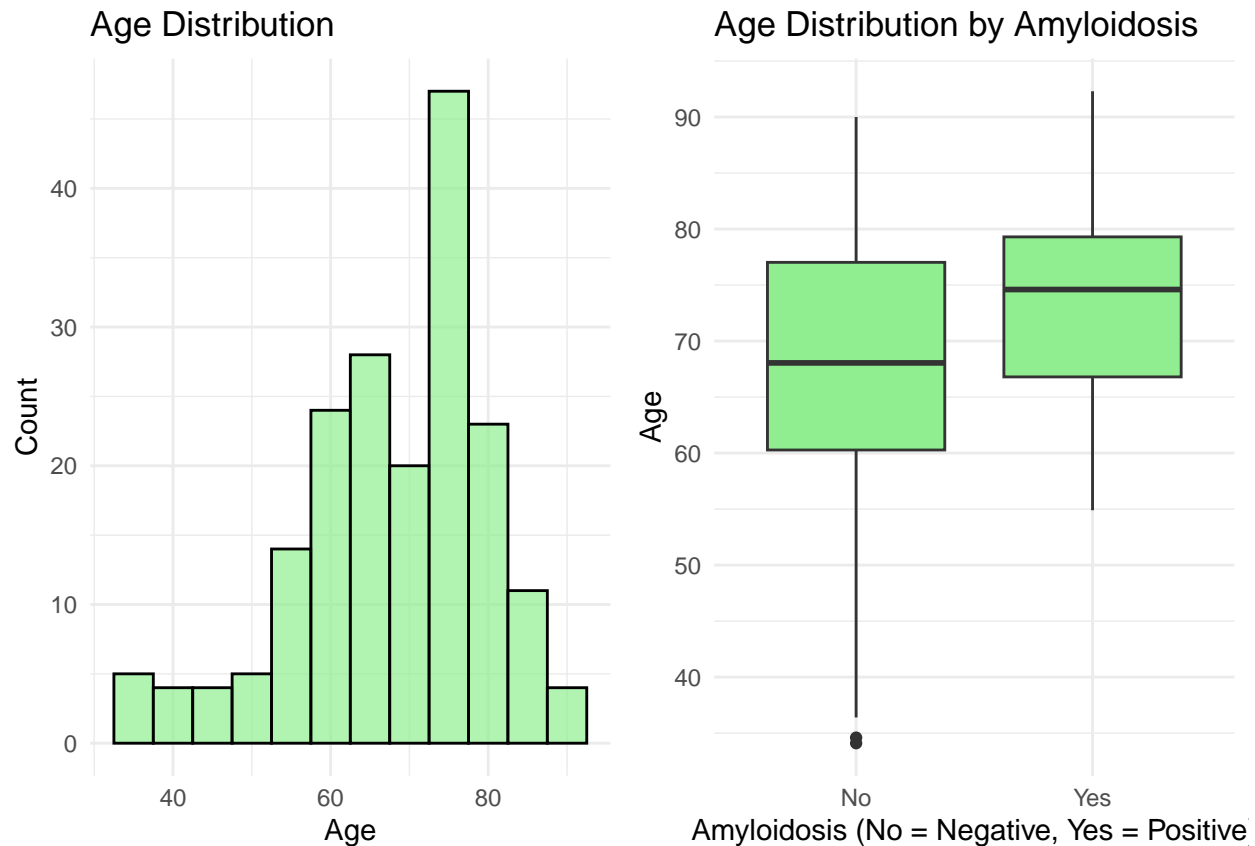


EDA

1 Visualizing missing values



2 Age distribution



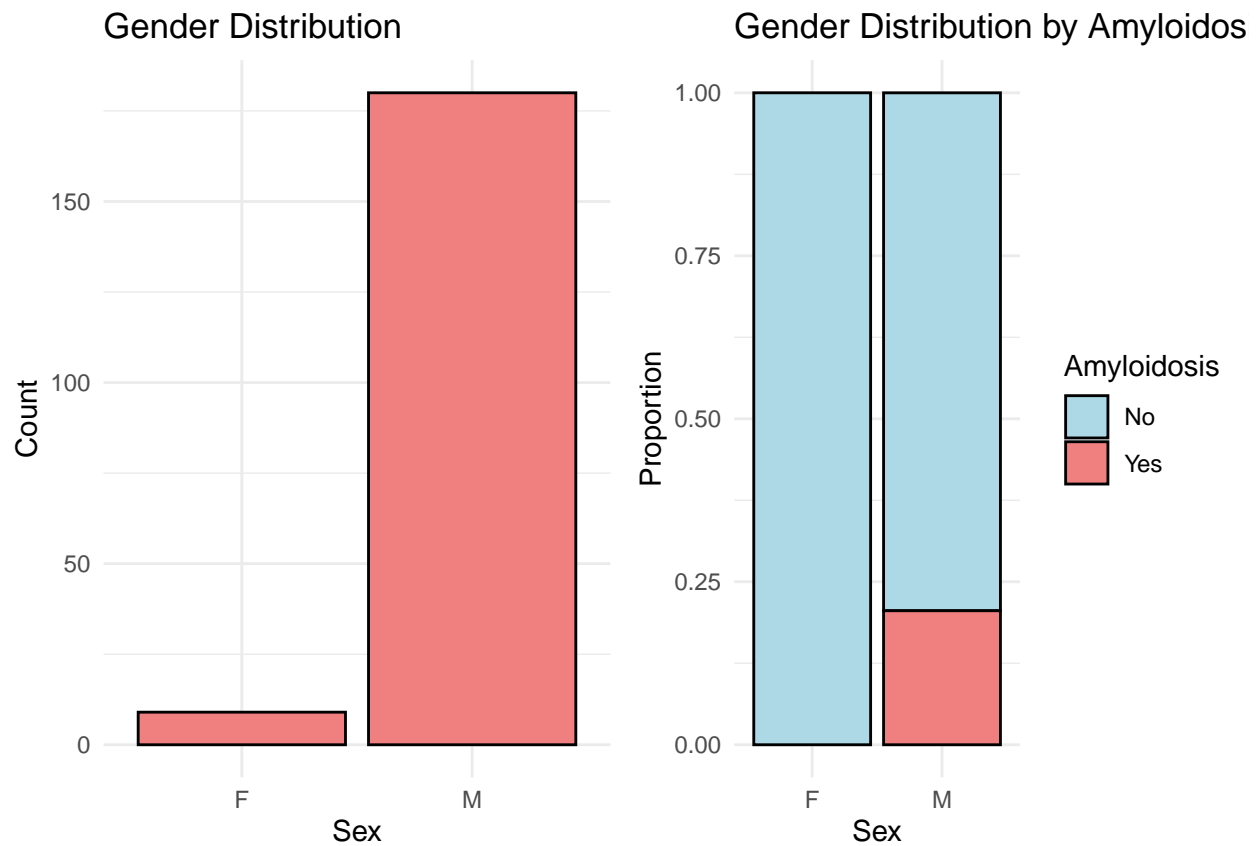
1. Age distribution (left histogram)

- Age is mainly concentrated between 60 and 80 years old, especially around 70 years old, and the data volume is relatively dense. Most patients are between 60 and 80 years old, showing a left-skewed distribution trend. This means that older patients occupy more samples than younger patients.

2. Age box plot grouped by Amyloid status (right box plot)

- The age distribution of the **Amyloid Negative group** is relatively concentrated, close to around 70 years old, and there are some extreme values (below 40 years old), showing a left-skewed distribution. This means that most patients in the negative group are relatively young, and there are some young patients.
- The median age of the **Amyloid Positive group** is slightly higher than that of the negative group, concentrated around 70 years old. The IQR (interquartile range) range of the positive group is narrower, showing an approximate normal distribution. Most of the patients in this group were around 70 years old, and the age distribution was relatively uniform without significant skewness, indicating that the age of amyloid-positive patients was relatively uniform and concentrated in the older age group.

3 Sex distribution



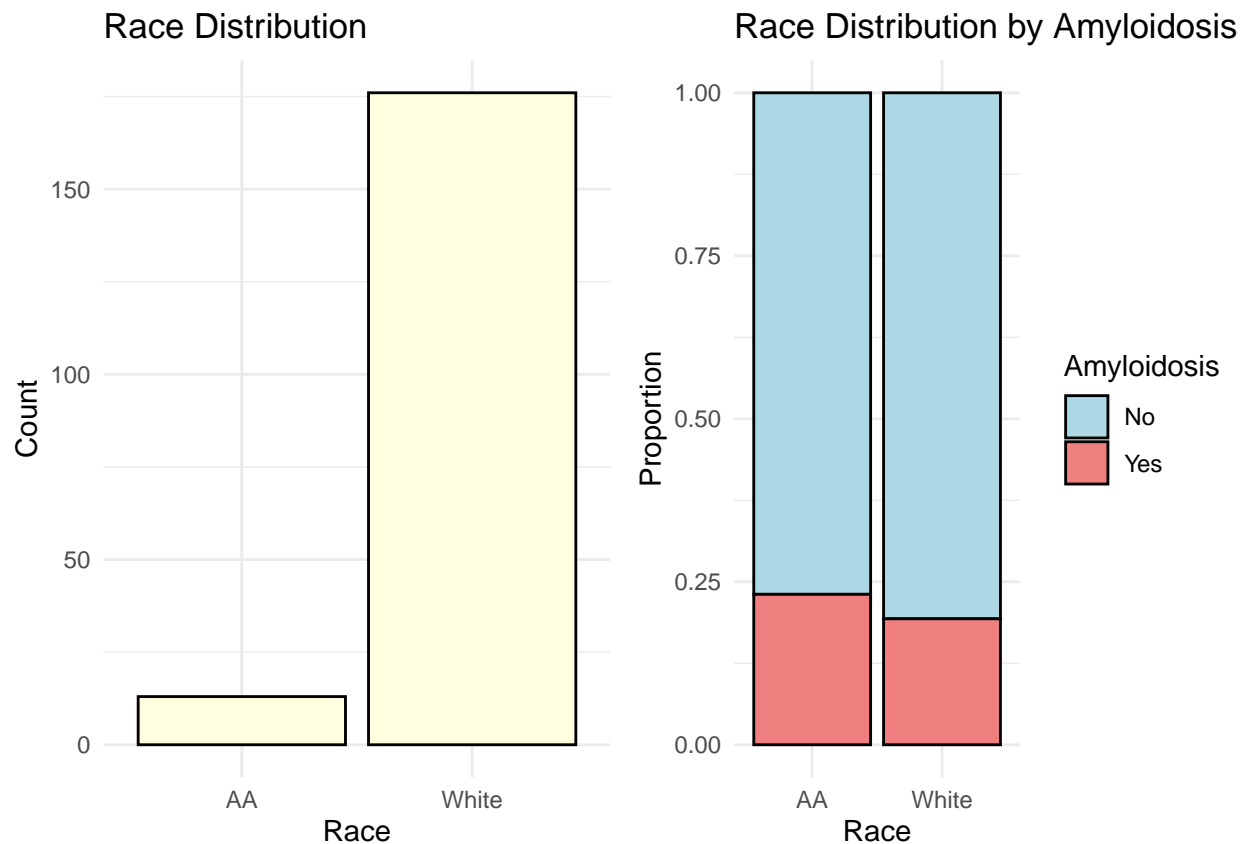
1. Gender distribution (left bar chart)

- The vast majority of patients in the dataset are male (“M”), with a number of 180, while there are only 9 females (“F”). This shows that the gender ratio in this dataset is seriously unbalanced, and most patients are male.

2. Gender distribution by amyloidosis status (right stacked bar chart)

- For **amyloid negative** (Amyloid = 0, blue), the gender distribution shows a relatively balanced state, with 143 males and 9 females, indicating that all female patients in the data belong to the negative group.
- For **amyloid positive** (Amyloid = 1, red), all patients are males, and females do not appear in the positive group at all. This result further shows that **amyloid positive** patients are composed only of males.

4 Race distribution



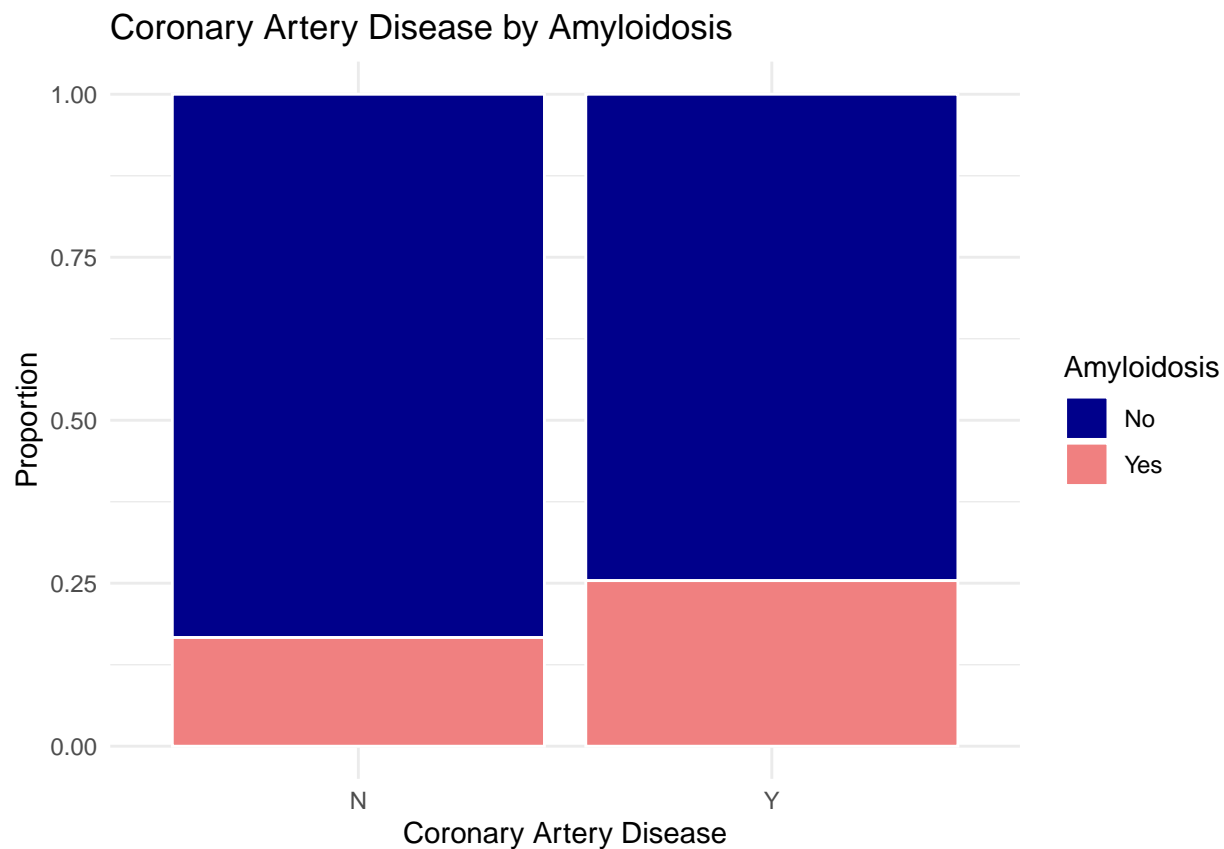
1. Racial distribution (left bar chart)

- This chart shows the distribution of the “Race” variable in the dataset. It can be seen that **White** patients account for the vast majority of the data, with a total of 176 people, while the proportion of **African Americans (AA)** is very small, with only 13 people.
- **White**: 176
- **African Americans (AA)**: 13

2. Racial distribution by amyloid status (right stacked bar chart)

- For **Amyloid negative** (Amyloid = 0, blue), the proportions of **White** and **African Americans** are almost equal, and these two races account for most of the data, specifically:
- **White**: 142
- **African Americans**: 10
- For **Amyloid positive** (Amyloid = 1, red), among the positive cases, **White** has 34 people and **African Americans** only has 3 people, and the proportion difference is large.
- **White**: 34
- **African American**: 3

5 The relationship between coronary artery disease and amyloid



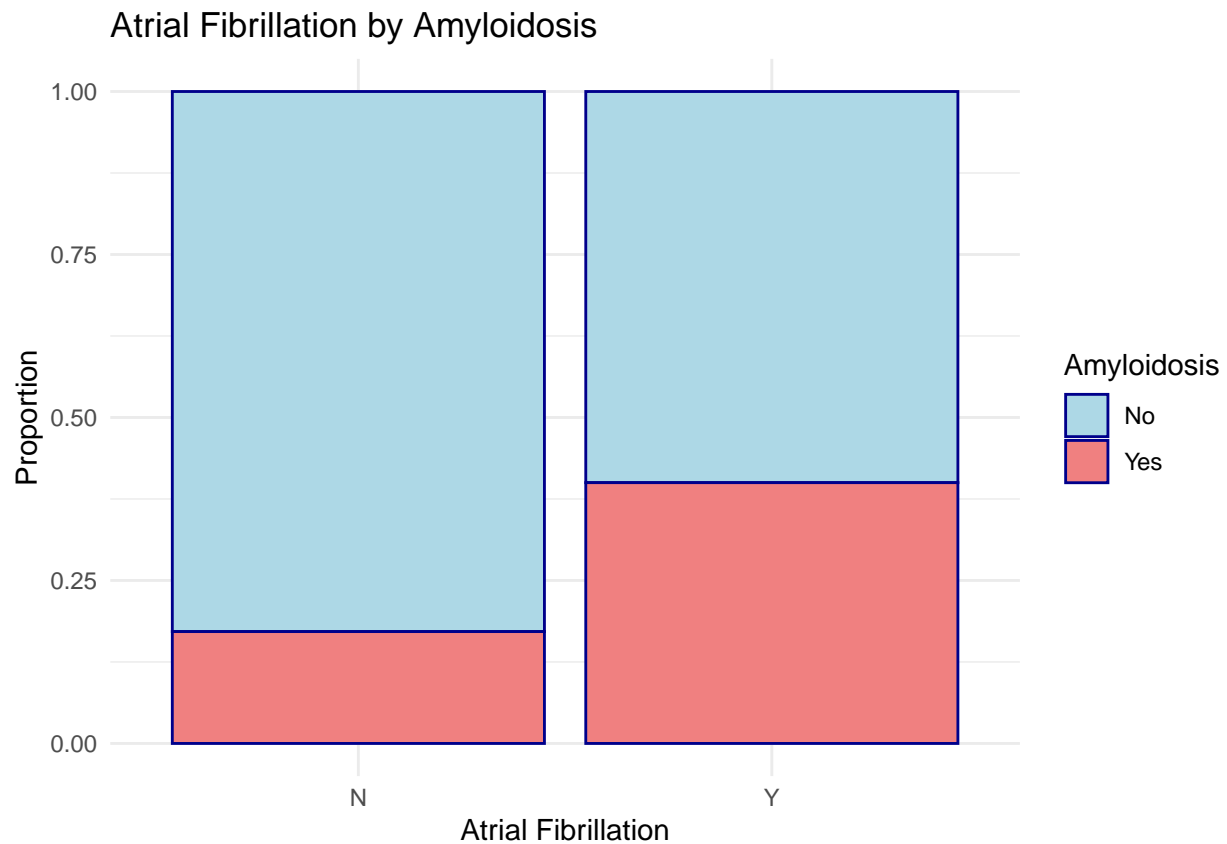
Coronary Artery Disease Negative (CAD = N)

- In the **Amyloid Negative group (Amyloid = 0)**, approximately **31%** of patients had coronary artery disease, while **69%** of patients did not have coronary artery disease.
- In the **Amyloid Positive group (Amyloid = 1)**, approximately **43%** of patients had coronary artery disease, while **57%** of patients did not have coronary artery disease.

Coronary Artery Disease Positive (CAD = Y)

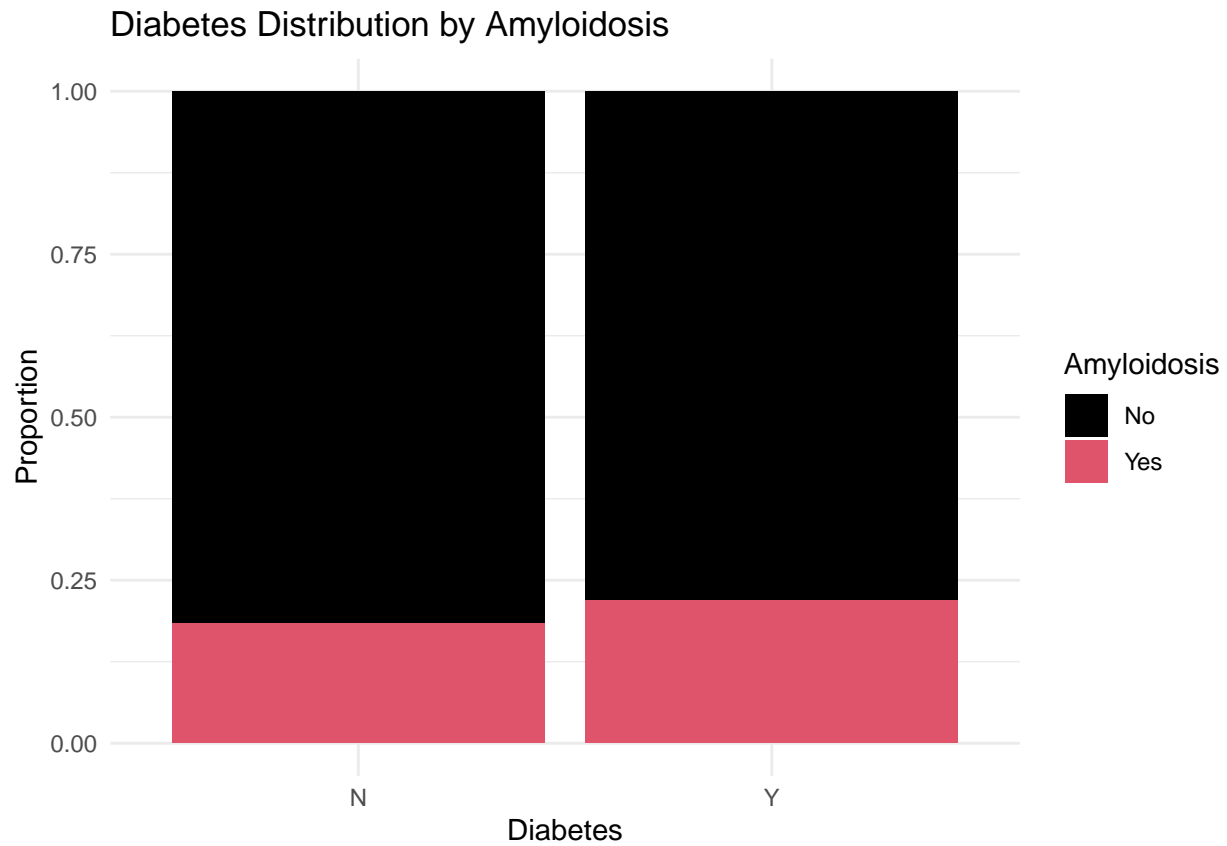
- In the **Amyloid Negative group (Amyloid = 0)**, approximately **31%** of patients had coronary artery disease, while **69%** of patients did not have coronary artery disease.
- In the **Amyloid Positive group (Amyloid = 1)**, **43%** of patients had coronary artery disease, while **57%** of patients did not have coronary artery disease.

6 The relationship between atrial fibrillation (Afib) and amyloid



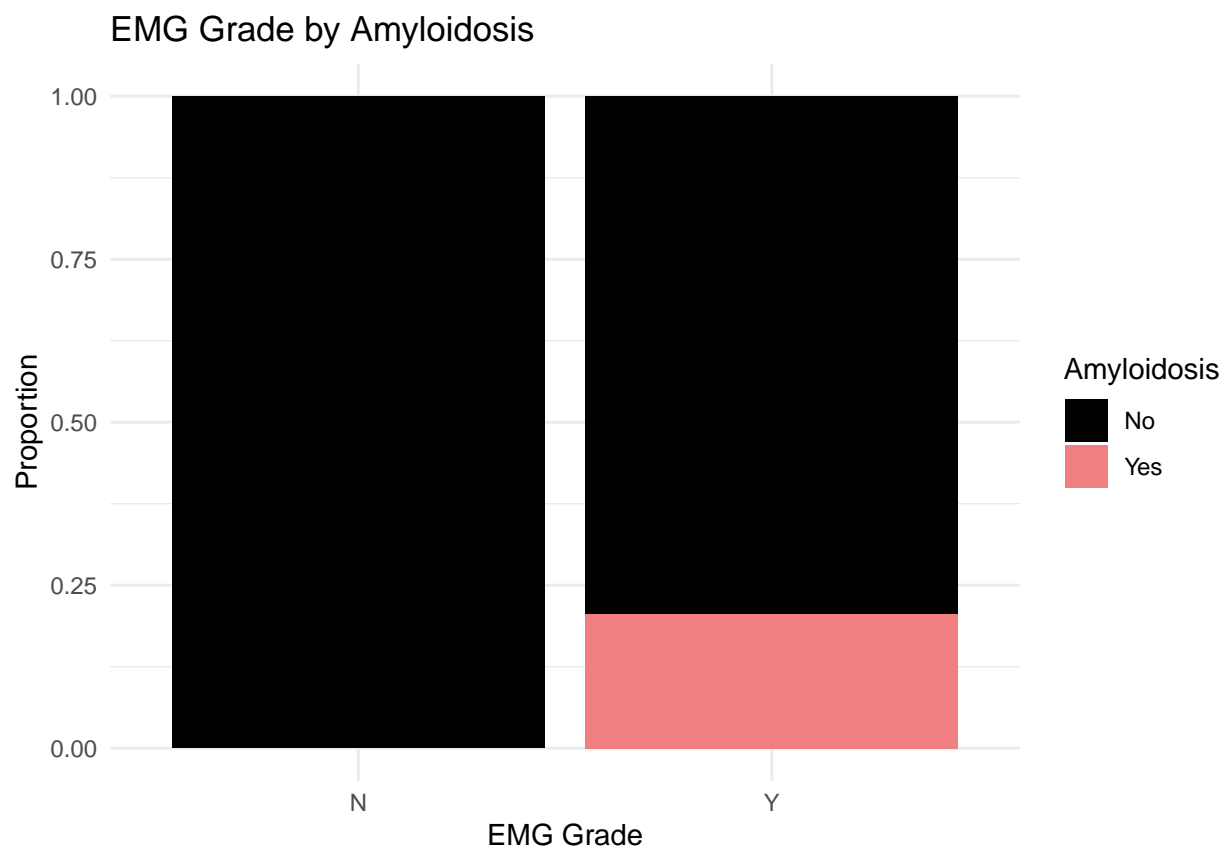
- In the **Amyloid negative group (Amyloid = 0)**, most patients did not have atrial fibrillation (92.1%), and only a few patients had atrial fibrillation (7.9%).
- In the **Amyloid positive group (Amyloid = 1)**, the incidence of atrial fibrillation was significantly higher, about **21.6%** of patients had atrial fibrillation, while **78.4%** of patients did not have atrial fibrillation.

7 The relationship between diabetes and amyloid



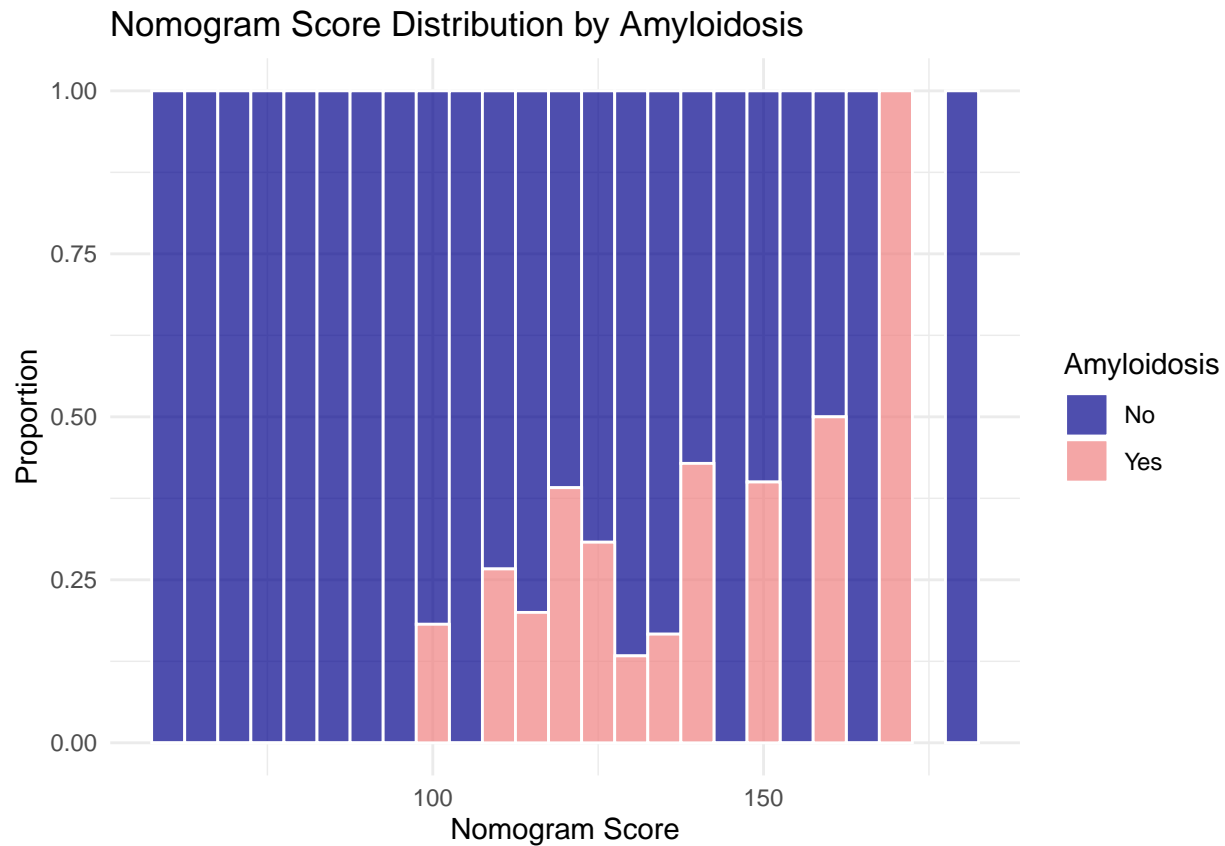
- **Diabetes negative (N)** group: Most patients did not have amyloidosis (black bars), accounting for about **67.1%**; while the proportion of amyloid positive (pink bars) was **32.9%**.
- **Diabetes positive (Y)** group: Although the proportion of diabetic patients was small, the proportion of amyloid positive (pink bars) and negative (black bars) in diabetic patients was almost equal, with amyloid positive being **37.8%** and negative being **62.2%**.

8 Relationship between EMG and Amyloid



- **EMG Grade 0** group: In this group, most patients did not have amyloidosis (black bars), accounting for about **94.1%**, while the proportion of amyloid positive patients (light coral bars) was **5.9%**.
- **EMG Grade 1** group: All patients in this group had amyloidosis (light coral bars), with **100%** of the patients being amyloid positive.

9 Relationship between Nomogram Score and Amyloid



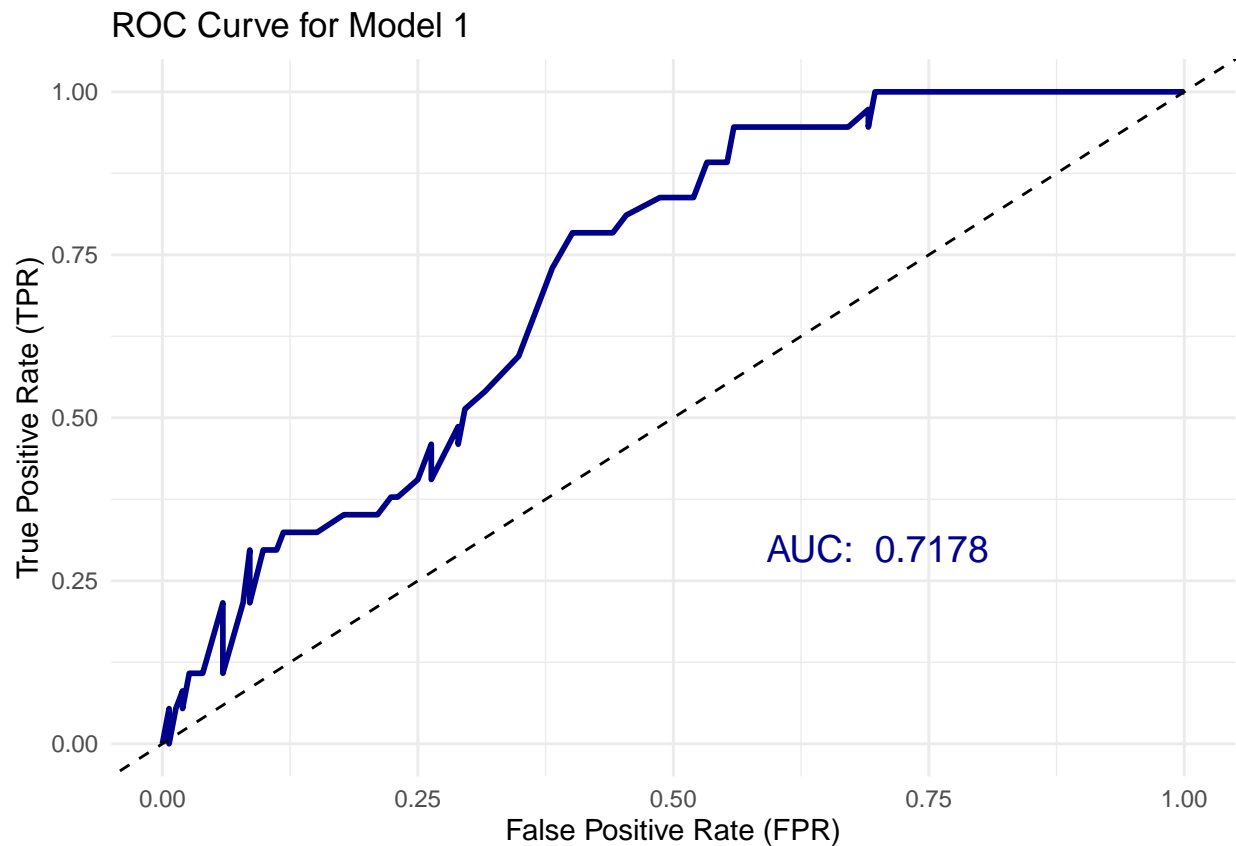
- Most patients with low Nomogram scores (close to 0 points)** were not amyloid positive (blue bars). As the score increased, the number of patients with **amyloid positive (red bars)** gradually increased.
- Among patients with high Nomogram scores (over 150 points), the proportion of amyloid positive patients increased significantly, while most patients with low scores were amyloid negative.

Model

Model1 Amyloidosis ~ Nomogram

```
##
## Call:
## glm(formula = Amyloid ~ Points, family = binomial(link = "logit"),
##      data = Amyloid_data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.715152   1.156982  -4.94 7.82e-07 ***
## Points       0.035890   0.009202   3.90 9.60e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

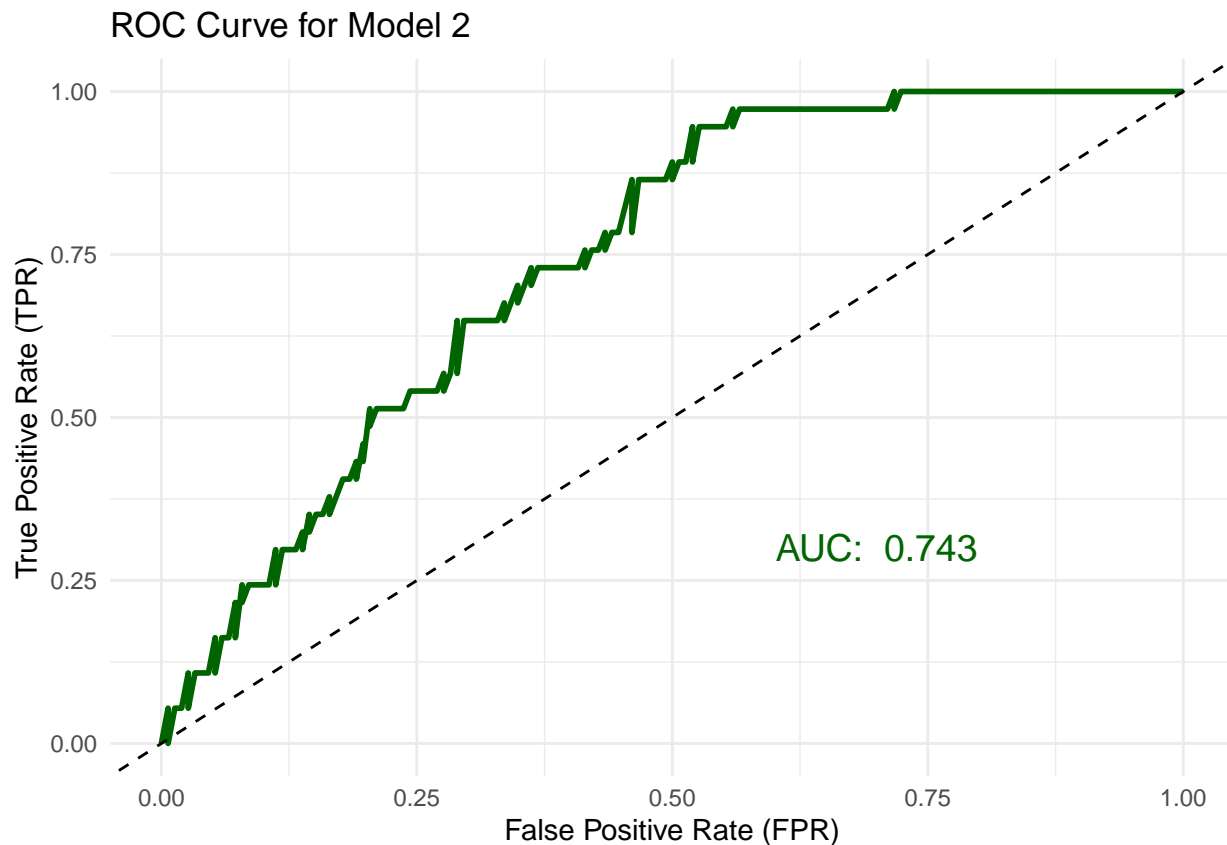
```
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 186.91 on 188 degrees of freedom
## Residual deviance: 169.65 on 187 degrees of freedom
## AIC: 173.65
##
## Number of Fisher Scoring iterations: 4
```



Model2 Amyloidosis ~ Risk Factors

```
##
## Call:
## glm(formula = Amyloid ~ Age + Sex + Monoclonal.Gammopathy + Afib +
## Diabetes + Bilateral., family = binomial(link = "logit"),
## data = Amyloid_data)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -21.42951 1289.99450 -0.017 0.98675
## Age 0.03982 0.01979 2.012 0.04425 *
## SexM 15.85893 1289.99384 0.012 0.99019
## Monoclonal.GammopathyY 0.34281 0.96372 0.356 0.72206
## AfibY 0.85336 0.54135 1.576 0.11494
## DiabetesY 0.13716 0.41594 0.330 0.74157
## Bilateral.Y 1.57060 0.57283 2.742 0.00611 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 186.91  on 188  degrees of freedom
## Residual deviance: 161.61  on 182  degrees of freedom
## AIC: 175.61
##
## Number of Fisher Scoring iterations: 16
```



Model3 Amyloidosis ~ Risk Factors + Severity

```
##
## Call:
## glm(formula = Amyloid ~ Age + Sex + Monoclonal.Gammopathy + Afib +
##      Diabetes + Bilateral. + Grade, family = binomial(link = "logit"),
##      data = Amyloid_data)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -37.51244 3803.27440  -0.010   0.9921
## Age              0.02874   0.02159   1.331   0.1831
## SexM            16.76737 1970.98680   0.009   0.9932
## Monoclonal.GammopathyY  0.08158   0.98217   0.083   0.9338
```

```
## AfibY          0.51953    0.56463    0.920    0.3575
## DiabetesY      0.09292    0.44021    0.211    0.8328
## Bilateral.Y    1.09498    0.59936    1.827    0.0677 .
## Grademild      0.19806 3585.50777    0.000    1.0000
## Grademoderate  15.70345 3252.70753    0.005    0.9961
## Gradesevere    17.10533 3252.70753    0.005    0.9958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 186.91  on 188  degrees of freedom
## Residual deviance: 147.39  on 179  degrees of freedom
## AIC: 167.39
##
## Number of Fisher Scoring iterations: 17
```

