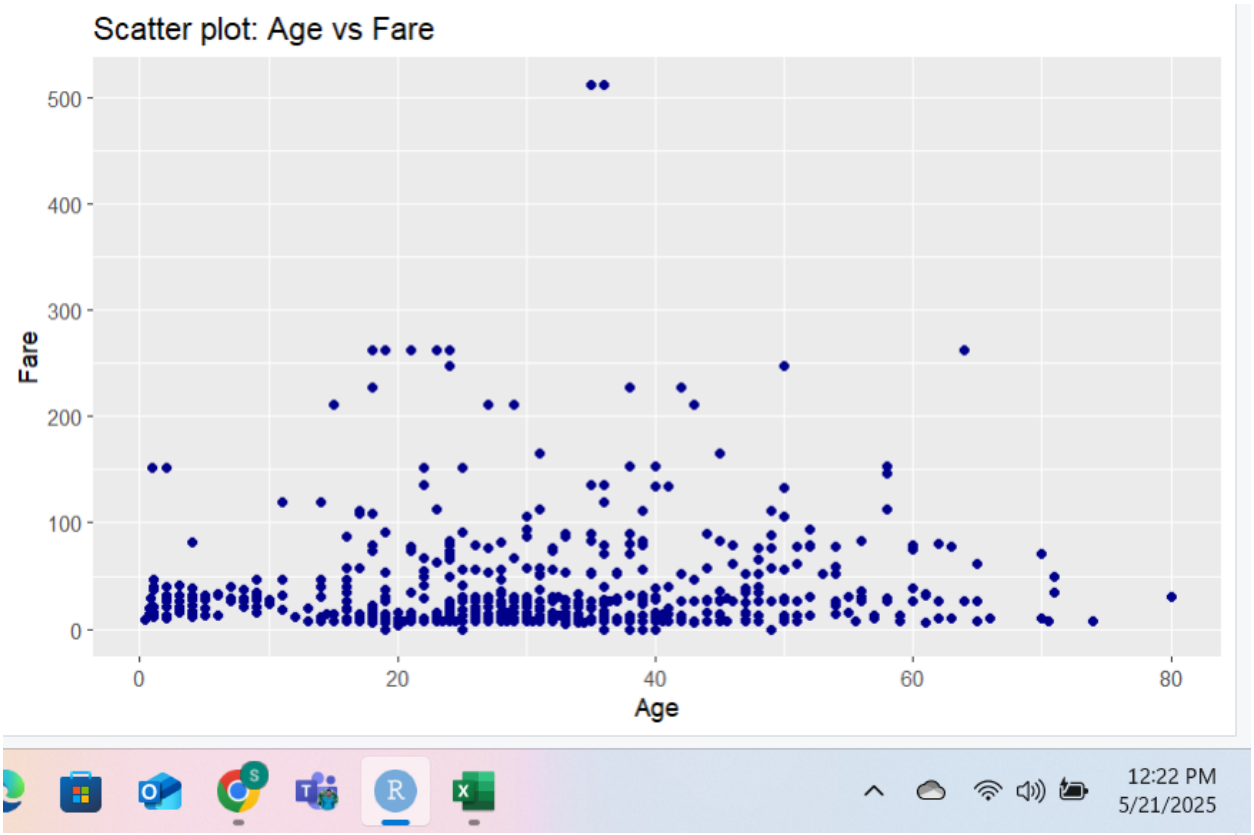
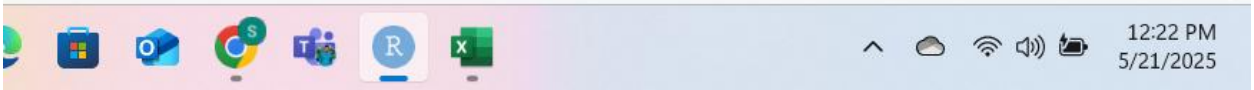
Task-1

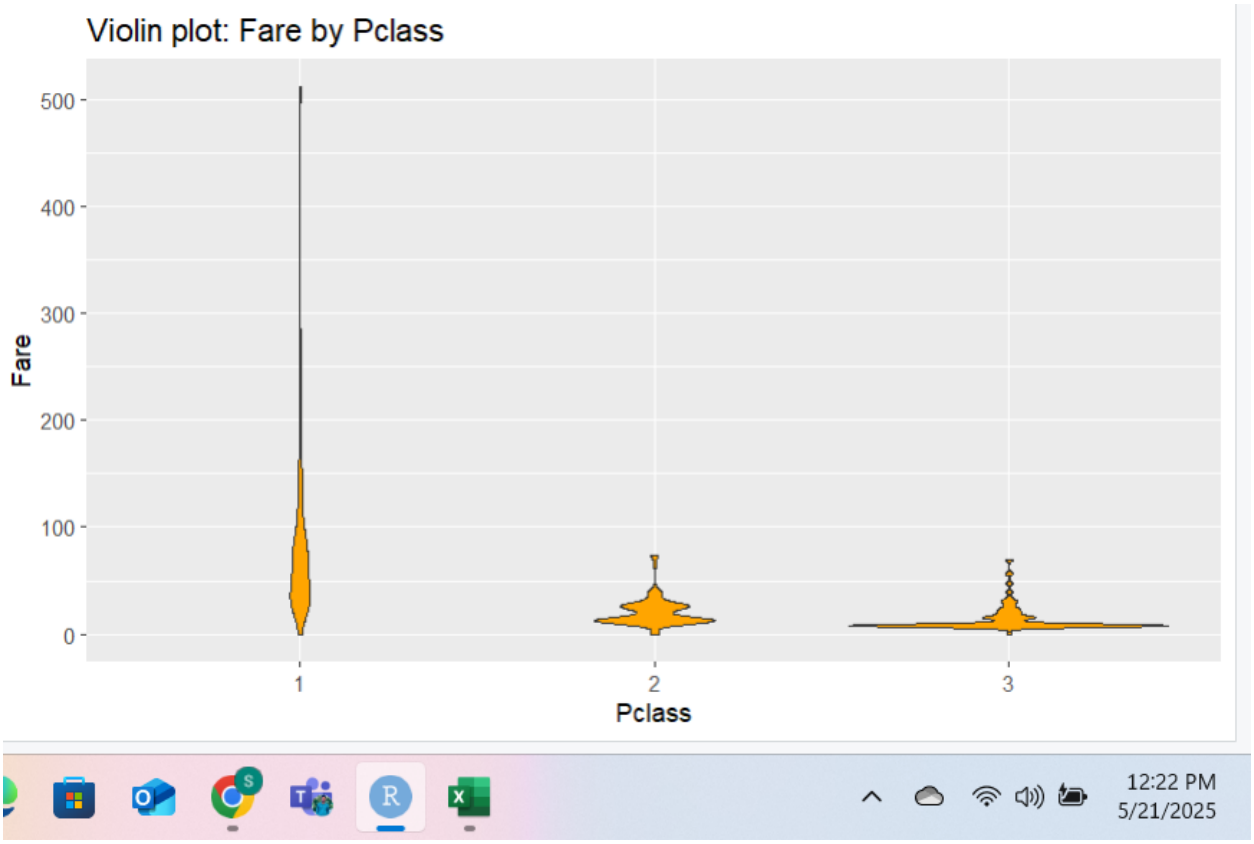

Histogram of Fare
count
Fare
0
100
200
300
400
500
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Bar Chart of Sex
count
Sex
female
male
600
400
200
0
12:22 PM
5/21/2025


Scatter plot: Age vs Fare
Fare
500
400
300
200
100
0
0
20
40
60
80
Age

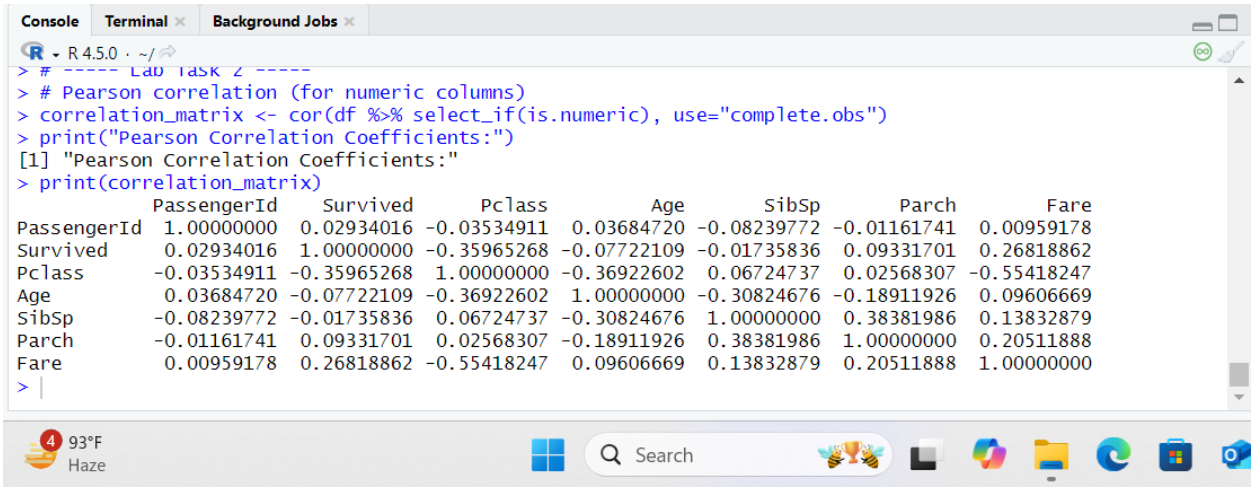
12:22 PM
5/21/2025

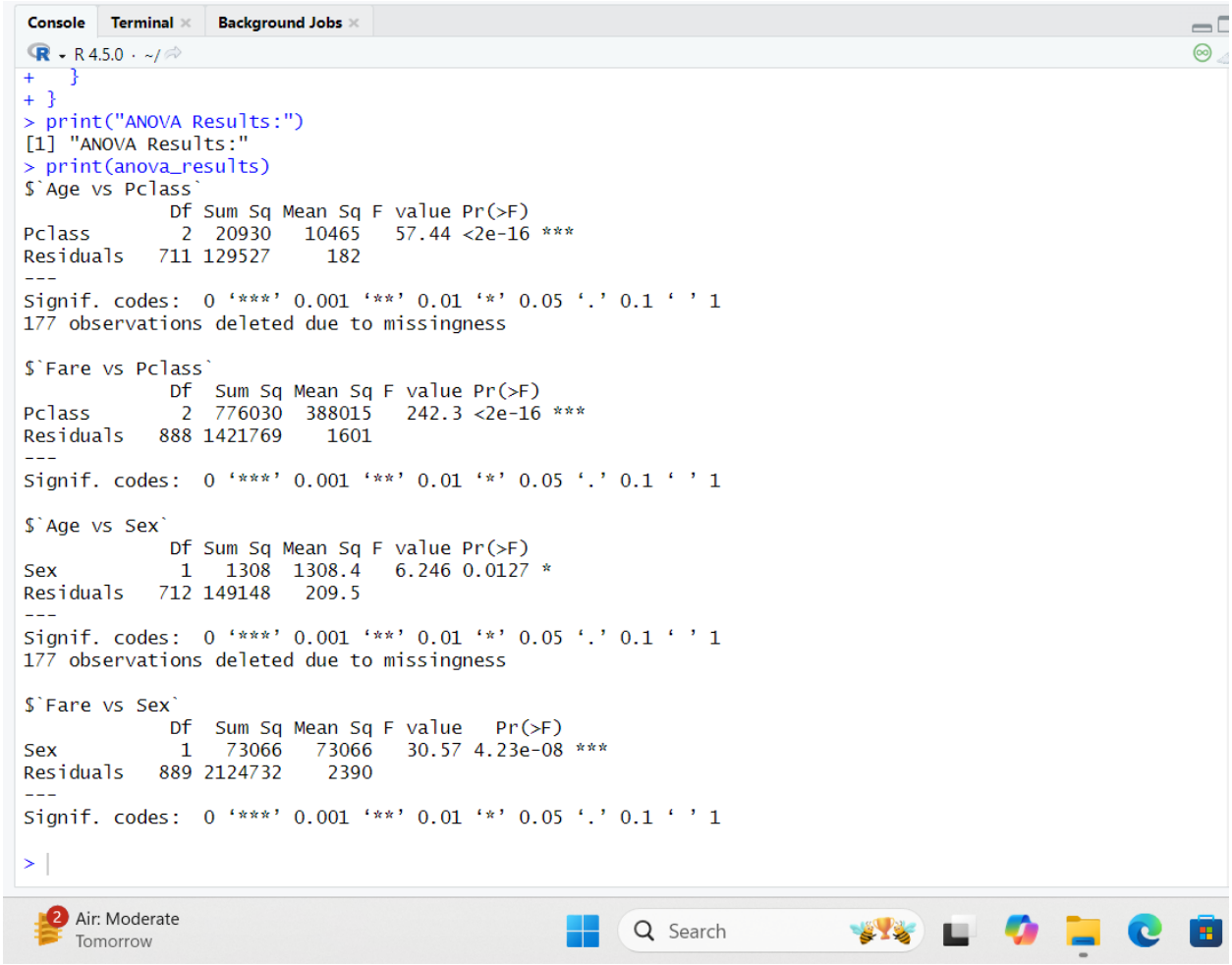

Violin plot: Age by Sex
Age
80
60
40
20
0
female
male
Sex


12:22 PM
5/21/2025


Violin plot: Fare by Pclass
Fare
500
400
300
200
100
0
1
2
3
Pclass
12:22 PM
5/21/2025

Task-2


Console
Terminal
Background Jobs
R 4.5.0 · ~/
> # ----- Lab Task 2 -----
> # Pearson correlation (for numeric columns)
> correlation_matrix <- cor(df %>% select_if(is.numeric), use="complete.obs")
> print("Pearson Correlation Coefficients:")
[1] "Pearson Correlation Coefficients:"
> print(correlation_matrix)
PassengerId 1.00000000 0.02934016 -0.03534911 0.03684720 -0.08239772 -0.01161741 0.00959178
Survived   0.02934016 1.00000000 -0.35965268 -0.07722109 -0.01735836 0.09331701 0.26818862
Pclass   -0.03534911 -0.35965268 1.00000000 -0.36922602 0.06724737 0.02568307 -0.55418247
Age   0.03684720 -0.07722109 -0.36922602 1.00000000 -0.30824676 -0.18911926 0.09606669
SibSp   -0.08239772 -0.01735836 0.06724737 -0.30824676 1.00000000 0.38381986 0.13832879
Parch   -0.01161741 0.09331701 0.02568307 -0.18911926 0.38381986 1.00000000 0.20511888
Fare   0.00959178 0.26818862 -0.55418247 0.09606669 0.13832879 0.20511888 1.00000000
>
93°F
Haze
Search

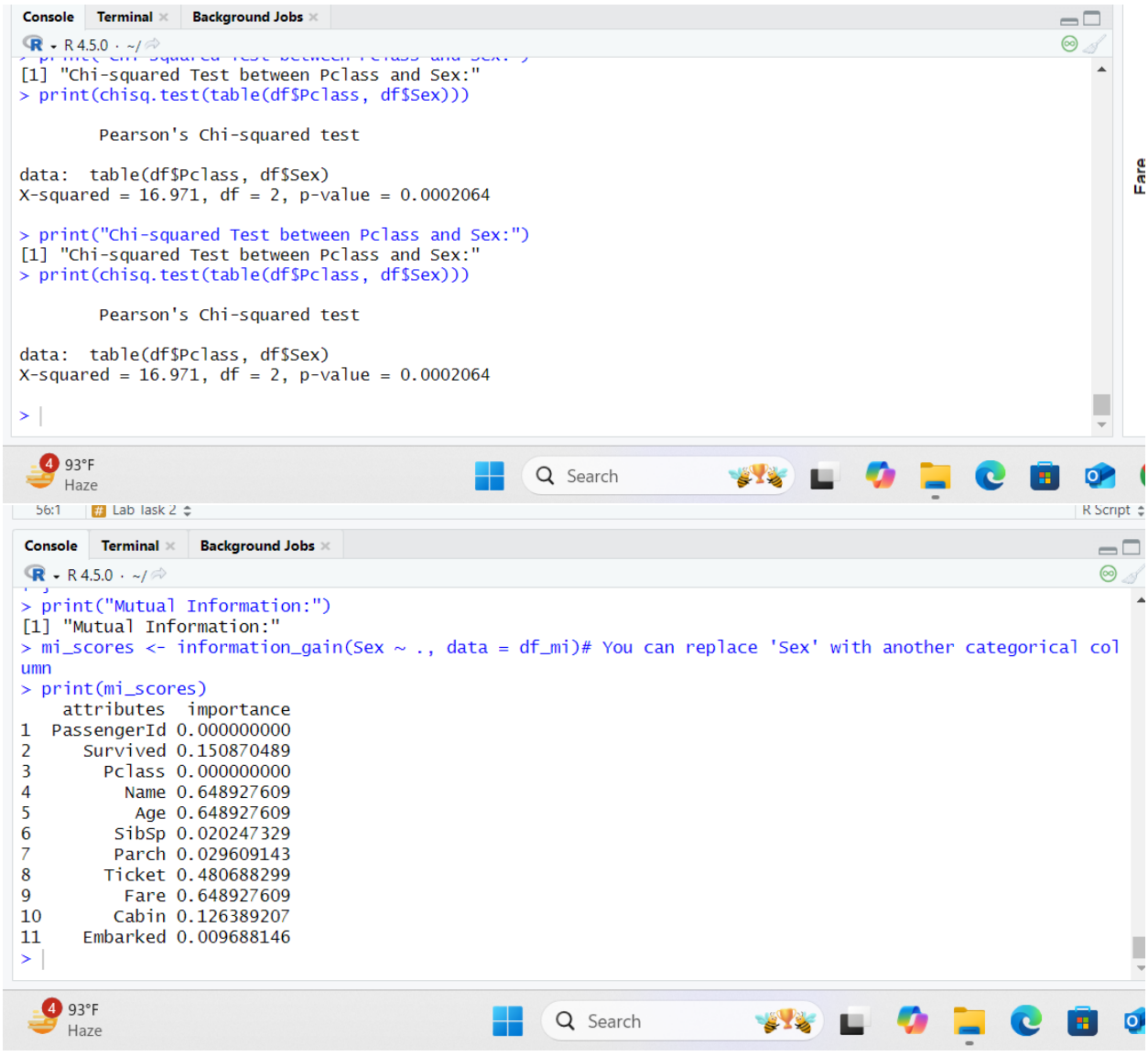

Console
Terminal
Background Jobs
R 4.5.0 · ~/
+ }
+ }
> print("ANOVA Results:")
[1] "ANOVA Results:"
> print(anova_results)
$`Age vs Pclass`
Df Sum Sq Mean Sq F value Pr(>F)
Pclass 2 20930 10465 57.44 <2e-16 ***
Residuals 711 129527 182

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
177 observations deleted due to missingness
$`Fare vs Pclass`
Df Sum Sq Mean Sq F value Pr(>F)
Pclass 2 776030 388015 242.3 <2e-16 ***
Residuals 888 1421769 1601

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
$`Age vs Sex`
Df Sum Sq Mean Sq F value Pr(>F)
Sex 1 1308 1308.4 6.246 0.0127 *
Residuals 712 149148 209.5

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
177 observations deleted due to missingness
$`Fare vs Sex`
Df Sum Sq Mean Sq F value Pr(>F)
Sex 1 73066 73066 30.57 4.23e-08 ***
Residuals 889 2124732 2390

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
Air: Moderate
Tomorrow
Search


Console
Terminal
Background Jobs
R 4.5.0 · ~/
[1] "Chi-squared Test between Pclass and Sex:"
> print(chisq.test(table(df$Pclass, df$Sex)))
Pearson's Chi-squared test
data: table(df$Pclass, df$Sex)
X-squared = 16.971, df = 2, p-value = 0.0002064
> print("Chi-squared Test between Pclass and Sex:")
[1] "Chi-squared Test between Pclass and Sex:"
> print(chisq.test(table(df$Pclass, df$Sex)))
Pearson's Chi-squared test
data: table(df$Pclass, df$Sex)
X-squared = 16.971, df = 2, p-value = 0.0002064
93°F
Haze
Search
Console
Terminal
Background Jobs
R 4.5.0 · ~/
> print("Mutual Information:")
[1] "Mutual Information:"
> mi_scores <- information_gain(Sex ~ ., data = df_mi)# You can replace 'Sex' with another categorical col
umn
> print(mi_scores)
attributes importance
1 PassengerId 0.000000000
2 Survived 0.150870489
3 Pclass 0.000000000
4 Name 0.648927609
5 Age 0.648927609
6 SibSp 0.020247329
7 Parch 0.029609143
8 Ticket 0.480688299
9 Fare 0.648927609
10 Cabin 0.126389207
11 Embarked 0.009688146
93°F
Haze
Search