ANLYTC3: NLSchools Analysis

The data set `nlschools` (in the `MASS` library, use `data(nlschools, package = "MASS")` to load in the data) contains the following records for 2287 students in the Netherlands (see `?nlschools` for more information)

- `lang`: their test score on a language exam
- `IQ`: their verbal IQ
- `class`: the ID number for their classroom
- `GS`: the number of students in each class
- `SES`: the socio-economic status of their family
- `COMB`: was the student in a multi-grade class? (0=no, 1=yes)

The following questions are of particular interest:

1. Are there descrepancies in `IQ` or `SES` in the different classes, or when grouping by multi-grade vs non-multi-grade classes?
2. When did students perform better or worse on the language exam? Describe which variables had the most important effects.
3. Do you think there are interactions in the effects of the variables on the language exam score? Speculate as to the cause of any such effects that you think should be included.

Begin your report:

Introduction:

This study investigates the influence of student characteristics and classroom context on academic achievement, specifically focusing on Language Exam Scores. The primary objective is to assess the initial equivalence of student populations across different classroom settings and to quantify the independent and interactive effects of three critical factors: Verbal IQ (as a measure of intrinsic cognitive ability), Socioeconomic Status (SES), and Classroom Grouping (specifically, Multigrade Status vs. Non-multigrade) on student performance.

```
# Library that will be used
library(ggplot2)
library(plyr)
library(reshape2)
library(knitr)
library(dplyr)

##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(DT)
library(gt)
library(kableExtra)

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows

library(hexbin)
library(RColorBrewer)
library(broom)
library(MASS)

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

# Load the 'nlschools' dataset from the MASS package
# This dataset contains data on Dutch schools, useful for multilevel modeling
data(nlschools, package = "MASS")

# Exports the 'nischools' data frame to a CSV file
write.csv(nlschools, "nlschools_data.csv", row.names = FALSE)

# Package-independent, uses local file
nlschools <- read.csv("nlschools_data.csv")

# Display a summary of the 'nlschools' dataset
summary(nlschools)

##       lang             IQ             class              GS
##  Min.   : 9.00   Min.   : 4.00   Min.   :  180   Min.   :10.00
##  1st Qu.:35.00   1st Qu.:10.50   1st Qu.: 6780   1st Qu.:23.00
```

```
##  Median :42.00    Median :12.00    Median :14180    Median :27.00
##  Mean   :40.93    Mean   :11.83    Mean   :13382    Mean   :26.51
##  3rd Qu.:48.00    3rd Qu.:13.00    3rd Qu.:19580    3rd Qu.:31.00
##  Max.   :58.00    Max.   :18.00    Max.   :25880    Max.   :39.00
##       SES              COMB
##  Min.   :10.00    Min.   :0.000
##  1st Qu.:20.00    1st Qu.:0.000
##  Median :27.00    Median :0.000
##  Mean   :27.81    Mean   :0.275
##  3rd Qu.:35.00    3rd Qu.:1.000
##  Max.   :50.00    Max.   :1.000
```

1. Are there descrepancies in IQ or SES in the different classes, or when grouping by multi-grade vs non-multi-grade classes?

Before examining the effects of class structure on student outcomes analyzing the impact of class structure is ensuring that the groups being compared are initially equivalent on key student characteristics. If students are not randomly assigned to different settings, pre-existing differences in ability and background can confound the results, leading to misattribution of effects.

I examine five critical distributions to screen the data for potential selection bias and confounding variables:

*   **IQ and GS**: The purpose of this is to determine if the average IQ of students is systematically related to the size of their class. A significant correlation would indicate that students with higher or lower IQ are not randomly distributed across different class sizes. This is a critical check for selection effects in classroom assignment that must be controlled for.

*   **SES and GS**: The purpose of this is to investigate whether family socio-economic background (SES) differs based on the size of the class. If students from specific SES brackets are predominantly found in smaller or larger classes, this discrepancy is a powerful confounding variable that must be accounted for when analyzing any outcome related to class size.

*   **IQ by Multi-grade Status (COMB)**: The purpose of this is to directly assess whether the IQ profile of students placed in multi-grade classes is comparable to those in non-multi-grade classes. Discrepancies here (e.g., if students with lower IQ are more likely to be placed in multi-grade settings) would invalidate a simple comparison between the two groups.

*   **SES by Multi-grade Status (COMB)**: The purpose of this is to ensure that any potential impact of the multi-grade structure is not an artifact of pre-existing socio-economic differences. A difference in SES between the COMB groups requires statistical control to accurately isolate the effect of the multi-grade structure itself.

- **Distribution of COMB by Class Size (GS)**: The purpose of this is to understand the interplay between the two main grouping variables COMB and Class Size(GS). This plot reveals if multi-grade grouping is more prevalent or exclusive within classes of a certain size. For example, if multi-grade classes only occur when class sizes are small, it suggests that the effects of COMB and GS are intertwined, demanding a careful analysis to disentangle which structural variable is responsible for any observed IQ or SES discrepancy.

```r
# Define class size categories (Small, Medium, Large)
BREAKS <- c(9, 19, 29, 39)
LABELS <- c("Small (10-19)", "Medium (20-29)", "Large (30-39)")

nlschools_plot <- nlschools %>%
  mutate(
    # Bin the 'GS' variable into the three defined class size categories
    Class_Size_Category = cut(GS,
                              breaks = BREAKS,
                              labels = LABELS,
                              right = TRUE,
                              include.lowest = TRUE)
  )

# CREATE VIOLIN PLOT: IQ Distribution by Class Size
ggplot(nlschools_plot, aes(x = Class_Size_Category, y = IQ, fill =
Class_Size_Category)) +

  # Show the density distribution of IQ scores
  geom_violin(trim = FALSE) +

  # Overlay a boxplot to clearly show median and quartiles
  geom_boxplot(width = 0.1, fill = "white", color = "black", outlier.shape =
NA) +

  # Set plot title and axis labels
  labs(
    title = "Distribution of IQ Scores by Class Size",
    x = "Class Size (GS)",
    y = "Verbal IQ"
  ) +
  # Apply a distinct color palette and clean up the theme
  scale_fill_brewer(palette = "Dark2") +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5, size = 16),
        # Add a black border around the entire plot
        plot.background = element_rect(
```
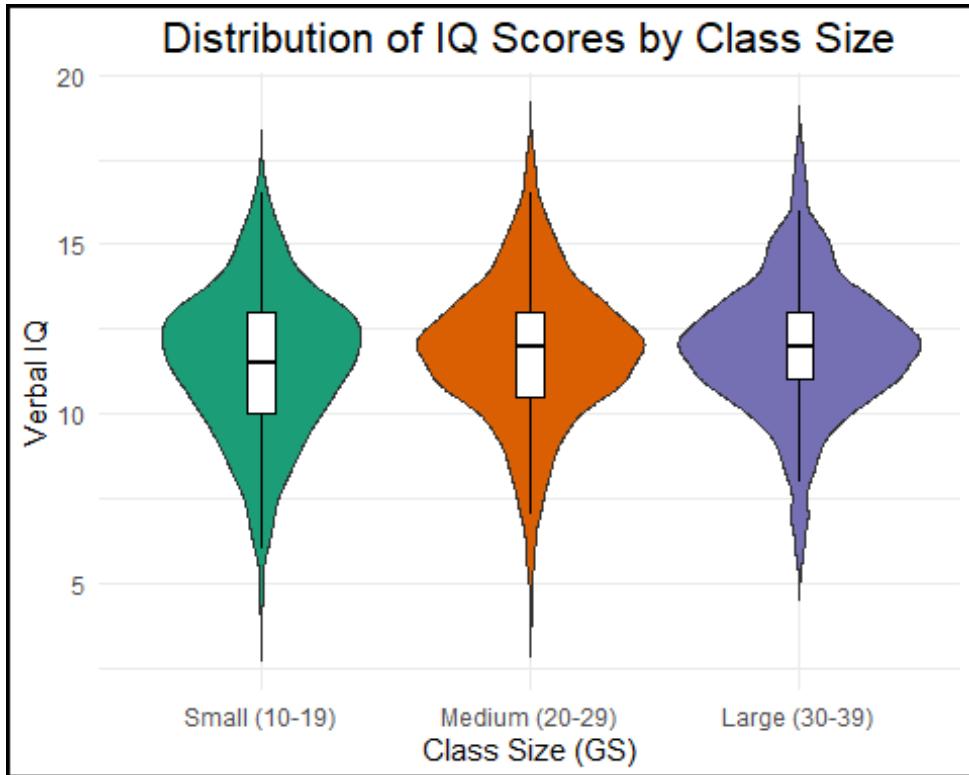
```
    color = "black",
    linewidth = 1.5,
    fill = NA )
)
```


The primary finding from this visualization is the absence of a substantial or systematic discrepancy in Verbal IQ across the three distinct Class Size Categories (GS).

- **Small (10-19)**: The distribution is characterized by a slightly greater concentration of scores within the average range, indicating a high density around the central tendency.

- **Medium (20-29)**: The shape, median, and Interquartile Range (IQR) are virtually identical to those of both the Small and Large groups, signifying highly similar score concentrations and central tendency.

- **Large (30-39)**: The median IQ aligns closely with the Small and Medium groups, demonstrating consistent central tendencies. The overall symmetry and spread are also highly comparable.

The consistency observed in both the central tendency (Median IQ) and dispersion (IQR and overall spread) across all three class sizes strongly suggests that students were not systematically assigned to smaller or larger classes based on their Verbal IQ.

```r
# PERFORM ONE-WAY ANOVA
# Test if mean IQ differs across the three class size categories.
anova_model <- aov(IQ ~ Class_Size_Category, data = nlschools_plot)

# VIEW ANOVA RESULTS (F-test and p-value)
# Display the standard ANOVA table.
summary(anova_model)

##                      Df Sum Sq Mean Sq F value Pr(>F)
## Class_Size_Category   2     37  18.653   4.371 0.0127 *
## Residuals          2284   9747   4.268
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA indicated a statistically significant main effect of Class Size Category on Verbal IQ, $F(2, 2284) = 4.371$, $p = 0.0127$. This finding compels the rejection of the null hypothesis of population mean equivalence, thereby demonstrating that the mean Verbal IQ is not uniform across the three class size groups.

```r
# POST-HOC TEST: TUKEY'S HSD
# Perform Tukey's HSD to find which specific class size groups differ
significantly.
tukey_results <- TukeyHSD(anova_model)

# Print the comparison results (p-values and confidence intervals)
print(tukey_results)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = IQ ~ Class_Size_Category, data = nlschools_plot)
##
## $Class_Size_Category
##                                   diff         lwr       upr      p adj
## Medium (20-29)-Small (10-19) 0.2819977 -0.03488811 0.5988835 0.0927275
## Large (30-39)-Small (10-19)  0.4230530  0.08614431 0.7599617 0.0091494
## Large (30-39)-Medium (20-29) 0.1410553 -0.08333166 0.3654423 0.3035142
```
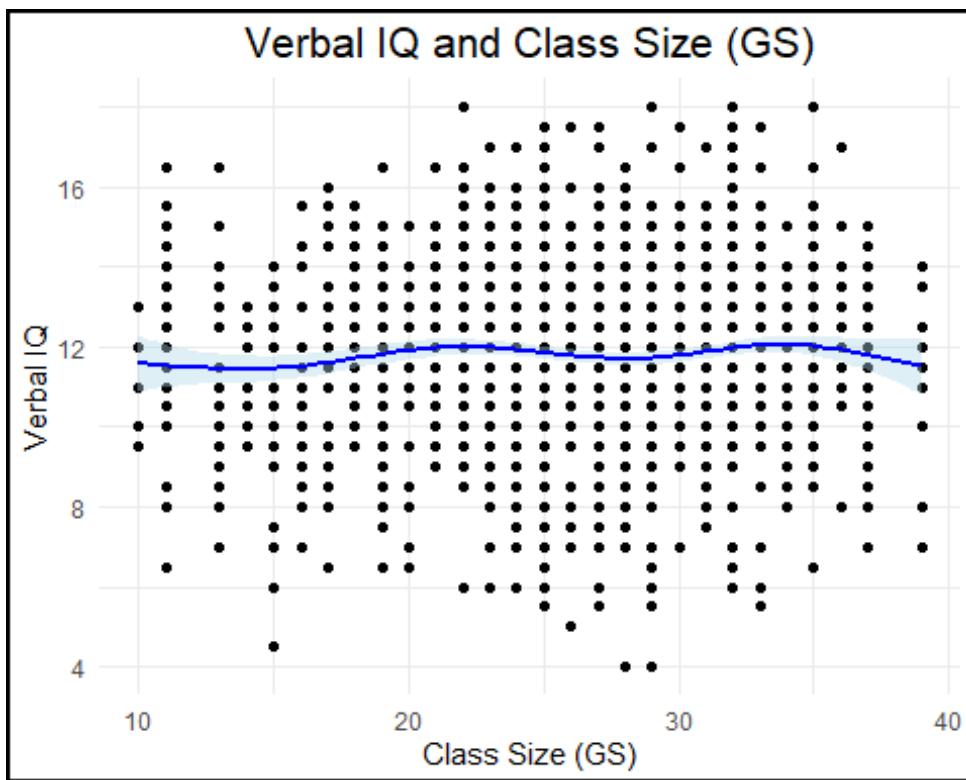
Based on the Tukey HSD test, only one comparison shows a statistically significant difference in mean Verbal IQ:

- **Large (30-39) vs. Small (10-19):**
- The p-adjusted value is 0.0091494, which is less than 0.05. We conclude that the mean Verbal IQ in Large classes is significantly higher than the mean Verbal IQ in Small classes. The difference is 0.423, indicating that the average IQ is about 0.42 points higher in the Large class size category.

The other two comparisons (Medium vs. Small, and Large vs. Medium) show no statistically significant difference in mean Verbal IQ at the 0.05 level.

The only class sizes that had a significantly different mean Verbal IQ were the Small (10-19) and Large (30-39) categories, with the Large classes exhibiting a higher average Verbal IQ

```r
# SCATTER PLOT: IQ vs. Class Size (GS)
ggplot(nlschools, aes(x = GS, y = IQ)) +

  # Plot individual data points
  geom_point(color = "black") +

  # Add a non-linear smoothing line (GAM) with 95% confidence band
  geom_smooth(
    method = "gam",
    formula = y ~ s(x),       # Specify a smoothing spline for the trend
    se = TRUE,                # Show 95% confidence interval
    color = "blue",
    fill = "lightblue"
  ) +

  # Set plot title and axis labels
  labs(
    x = "Class Size (GS)",
    y = "Verbal IQ",
    title = "Verbal IQ and Class Size (GS)"
  ) +
  theme_minimal() +

  # Remove the default point legend
  scale_color_manual(name = "Legend", values = "black", guide = "none") +

  # Customize plot theme
  theme(
    legend.position = "right",
    plot.title = element_text(hjust = 0.5, size = 16),
    # Add a border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```

Verbal IQ and Class Size (GS)

The primary finding from this plot is that there is no meaningful or systematic linear (or non-linear) relationship between a student's Verbal IQ and their Class Size(GS).

- **Flat Trend**: he Locally Estimated Scatterplot Smoothing (LOESS) curve exhibits a near-zero slope across the observed range of Class Sizes (GS ≈ 10 to GS ≈ 40). Minor fluctuations are localized and do not indicate a sustained systematic trend.

- **Wide Spread**: Significant vertical dispersion of individual data points (Verbal IQ range 4 to 16) is observed at all levels of class size, confirming substantial heterogeneity of IQ within each group.

- **CMean Stability**: The narrow confidence interval surrounding the mean (IQ ≈ 12) provides statistical confirmation of the stability of the average Verbal IQ across all observed class sizes.

This plot reinforces the finding from previous plot: there is no evidence of selection bias where students of higher or lower Verbal IQ are systematically assigned to classes of a particular size. This stability confirms initial equivalence on this core ability measure.

The conclusion of "no evidence of selection bias" pertains strictly to the observed data correlation. While the data suggests initial equivalence on Verbal IQ across groups, we must avoid definitively claiming this proves the absence of all potential, unobserved systematic assignment (selection bias).

```
# PEARSON CORRELATION TEST: GS vs. IQ
# Test for a linear relationship between Class Size (GS) and IQ.
cor_test_result <- cor.test(nlschools$GS, nlschools$IQ, method = "pearson")
print(cor_test_result)

##
##  Pearson's product-moment correlation
##
## data:  nlschools$GS and nlschools$IQ
## t = 1.5738, df = 2285, p-value = 0.1157
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.008092782  0.073794589
## sample estimates:
##        cor
## 0.03290613

# SIMPLE LINEAR REGRESSION
# Model IQ predicted by Class Size (GS).
regression_model <- lm(IQ ~ GS, data = nlschools)
# View the model summary (coefficients, R-squared, and p-values).
summary(regression_model)

##
## Call:
## lm(formula = IQ ~ GS, data = nlschools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8636 -1.2744  0.1245  1.1958  6.2196
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.518773   0.204949  56.203   <2e-16 ***
## GS           0.011892   0.007556   1.574    0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.068 on 2285 degrees of freedom
## Multiple R-squared:  0.001083,   Adjusted R-squared:  0.0006457
## F-statistic: 2.477 on 1 and 2285 DF,  p-value: 0.1157
```

Pearson Correlation Test - **Correlation Coefficient (r)**: 0.03290613 - This value is near zero, indicating a negligible positive linear association between Class Size (GS) and Verbal IQ.

- **P-value**: 0.1157
- Given that the p-value (0.1157) exceeds the conventional significance level (alpha = 0.05), the null hypothesis of no correlation cannot be rejected. There is no statistically significant linear correlation between the two variables.

Simple Linear Regression Model The simple linear regression model corroborates the absence of a statistically significant linear relationship.
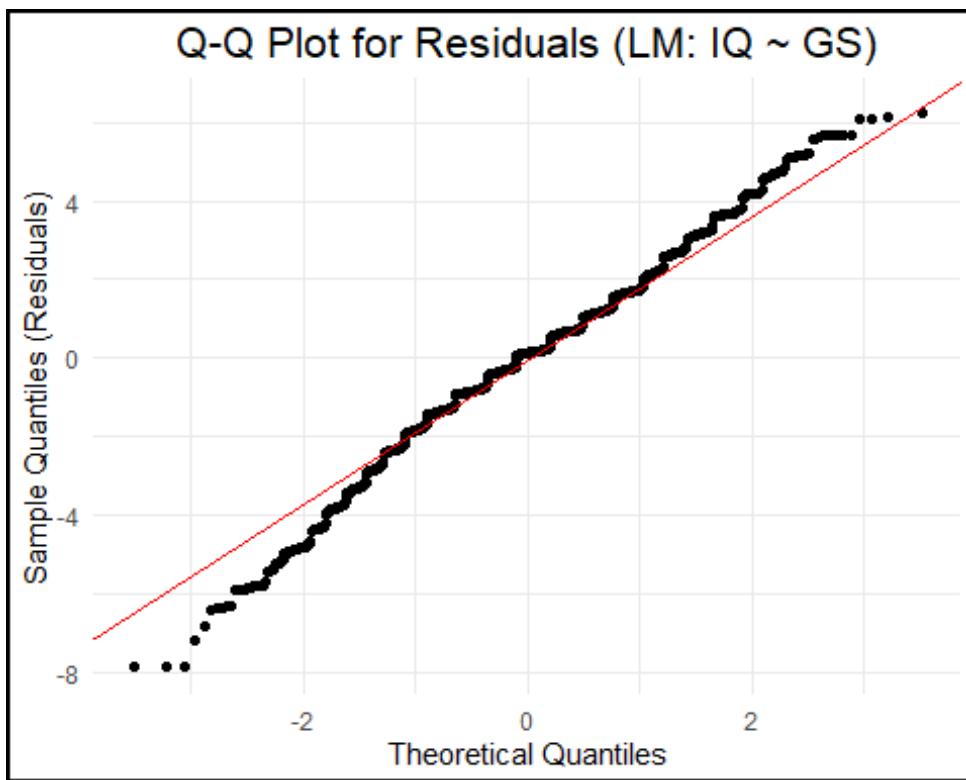
For every one-unit increase in Class Size (GS), Verbal IQ is estimated to increase by only 0.011892 points. This effect is quantitatively trivial. The high p-value for the GS coefficient (0.116) confirms that Class Size, when modeled as a continuous linear predictor, does not significantly predict Verbal IQ in this dataset.

This finding is consistent with the Pearson correlation test (r = 0.033, p = 0.1157), both of which demonstrate a lack of systematic linear association between Class Size and Verbal IQ.

```r
# SIMPLE LINEAR REGRESSION: IQ ~ GS
lm_model <- lm(IQ ~ GS, data = nlschools)

# Extract and store the model residuals for diagnostic checking
nlschools$residuals <- residuals(lm_model)

# Q-Q PLOT: CHECK RESIDUALS NORMALITY
# Use the Q-Q plot to visually assess the normality assumption of the
residuals.
ggplot(nlschools, aes(sample = residuals)) +
  stat_qq() +
  stat_qq_line(color = "red") +
  labs(title = "Q-Q Plot for Residuals (LM: IQ ~ GS)",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles (Residuals)") +
  theme_minimal() +
  # Add a border around the entire plot
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```

Q-Q Plot for Residuals (LM: IQ ~ GS)

The primary finding is that the assumption of normality of residuals is largely satisfied for the model predicting Verbal IQ from Class Size (GS).

The vast majority of the data points, specifically those representing the central 90% of the distribution (between -2 and +2 on the Theoretical Quantiles axis), align closely with the theoretical line, indicating a near-normal distribution of the residual errors.

- **Tail Deviations**: Minor deviations from normality are observed at the extremes:
  - The lower tail (Theoretical Quantiles ≈ -2 and below) exhibits points dropping below the line, suggesting a slightly heavier lower tail than predicted by a perfect normal distribution, potentially indicating a modest left skew or low-end outliers.

  - The upper tail (Theoretical Quantiles ≈ +2 and above) shows a slight trailing off, though the deviation is less pronounced than in the lower tail.

These minor deviations at the tails are not substantial enough to compromise the statistical validity of the model used to confirm the lack of a linear relationship between Verbal IQ and Class Size. The central adherence to the theoretical line validates the core model inference.

```
# Define class size categories (Small, Medium, Large)
BREAKS <- c(9, 19, 29, 39)
LABELS <- c("Small (10-19)", "Medium (20-29)", "Large (30-39)")
```
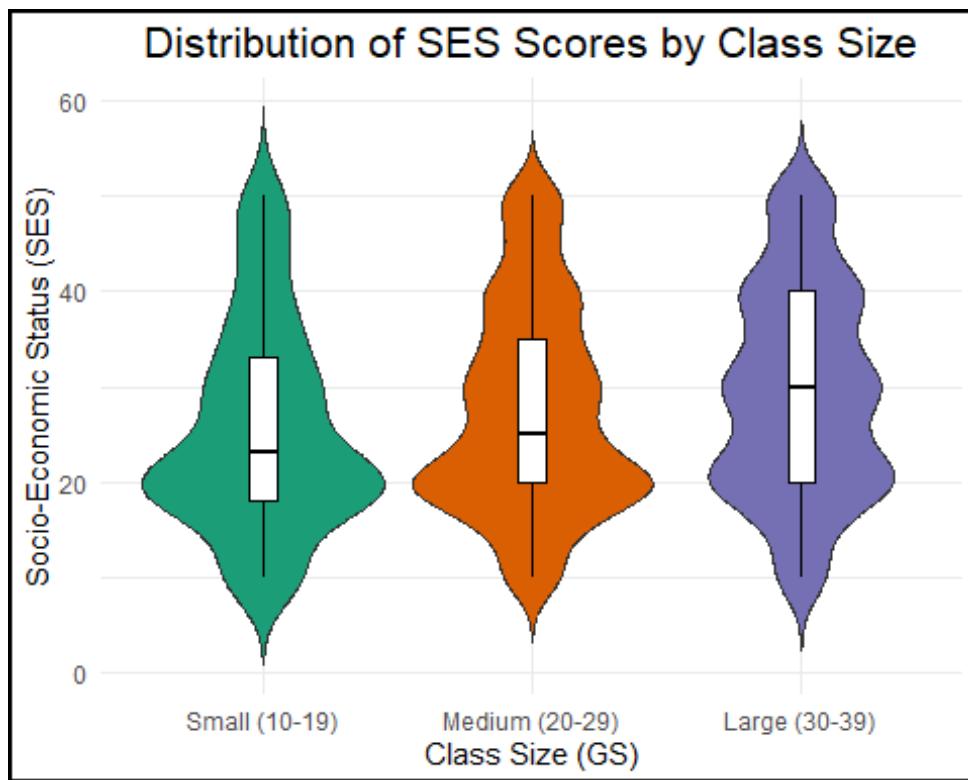
```r
nlschools_plot <- nlschools %>%
  mutate(
    # Bin the 'GS' variable into the three defined class size categories
    Class_Size_Category = cut(GS,
                              breaks = BREAKS,
                              labels = LABELS,
                              right = TRUE,
                              include.lowest = TRUE)
  )

# CREATE VIOLIN PLOT: SES Distribution by Class Size
ggplot(nlschools_plot, aes(x = Class_Size_Category, y = SES, fill =
Class_Size_Category)) +

  # Show the density distribution of SES scores
  geom_violin(trim = FALSE) +

  # Overlay a boxplot to clearly show median and quartiles
  geom_boxplot(width = 0.1, fill = "white", color = "black", outlier.shape =
NA) +

  # Set plot title and axis labels
  labs(
    title = "Distribution of SES Scores by Class Size",
    x = "Class Size (GS)",
    y = "Socio-Economic Status (SES)"
  ) +
  # Apply a distinct color palette and clean up the theme
  scale_fill_brewer(palette = "Dark2") +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(hjust = 0.5, size = 16),
        # Add a black border around the entire plot
        plot.background = element_rect(
    color = "black",
    linewidth = 1.5,
    fill = NA )
  )
```

Distribution of SES Scores by Class Size

This plot visualizes the distribution of Socio-Economic Status (SES) Scores across three distinct Class Size Categories: Small (10-19), Medium (20-29), and Large (30-39).

The analysis reveals a positive relationship between Class Size category and the central tendency of SES scores. Although all categories encompass the full range of SES scores, a systematic upward shift in the median SES is observed as class size increases. - **Small (10-19)**: This category exhibits the lowest median SES (approximately 23). The violin plot density suggests a marginal concentration of students toward the lower end of the SES scale relative to the other groups.

- **Medium (20-29)**: The median SES for this category is intermediate (approximately 25-26), positioning it slightly above the Small group while maintaining a broad overall distribution.

- **Large (30-39)**: This category demonstrates the highest median SES (approximately 30). The overall mass of the distribution is visibly shifted upward, indicating that students in larger classes, on average, originate from more affluent socio-economic backgrounds compared to their counterparts in smaller classes.

This finding suggests a potential selection bias based on SES, where students from higher SES backgrounds are disproportionately enrolled in larger classes.

```
# ONE-WAY ANOVA: SES vs. Class Size Category
# Purpose: Test if mean SES differs across the three class size categories.
```

```
aov_ses_gs <- aov(SES ~ Class_Size_Category, data = nlschools_plot)
summary(aov_ses_gs)

##                       Df Sum Sq Mean Sq F value  Pr(>F)
## Class_Size_Category    2   3511  1755.4   14.93 3.6e-07 ***
## Residuals           2284 268476   117.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A one-way Analysis of Variance (ANOVA) was performed to formally assess the initial equivalence of mean Socio-Economic Status (SES) across the three Class Size (GS) categories.

The results indicate a highly statistically significant difference in mean SES among the groups. Specifically, the observed p-value (3.6×10^{-7}) is substantially lower than the conventional significance level (alpha = 0.05), leading to the strong rejection of the null hypothesis (that the true mean SES is equal across all three categories).

This finding confirms that students in larger classes possess a significantly higher mean SES compared to those in smaller classes, underscoring the presence of a selection bias based on SES.

```
# POST-HOC TEST: TUKEY'S HSD
# Perform Tukey's HSD to identify specific class size groups with different
mean SES.
tukey_ses_results <- TukeyHSD(aov_ses_gs)

# Print the pairwise comparison results (p-values and confidence intervals)
print(tukey_ses_results)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = SES ~ Class_Size_Category, data = nlschools_plot)
##
## $Class_Size_Category
##                                  diff       lwr      upr     p adj
## Medium (20-29)-Small (10-19) 1.832365 0.1692984 3.495433 0.0265684
## Large (30-39)-Small (10-19)  3.826081 2.0579301 5.594231 0.0000013
## Large (30-39)-Medium (20-29) 1.993715 0.8160965 3.171334 0.0002182
```

The Tukey's Honestly Significant Difference (HSD) post-hoc test was conducted to identify which pairwise comparisons among the three Class Size Categories account for the significant ANOVA finding.

The results indicate that all three pairwise comparisons yield a statistically significant difference in mean SES scores: - **Medium vs. Small (p ≈ 0.027)**: The mean SES for Medium classes is significantly higher than for Small classes (diff ≈ 1.83).
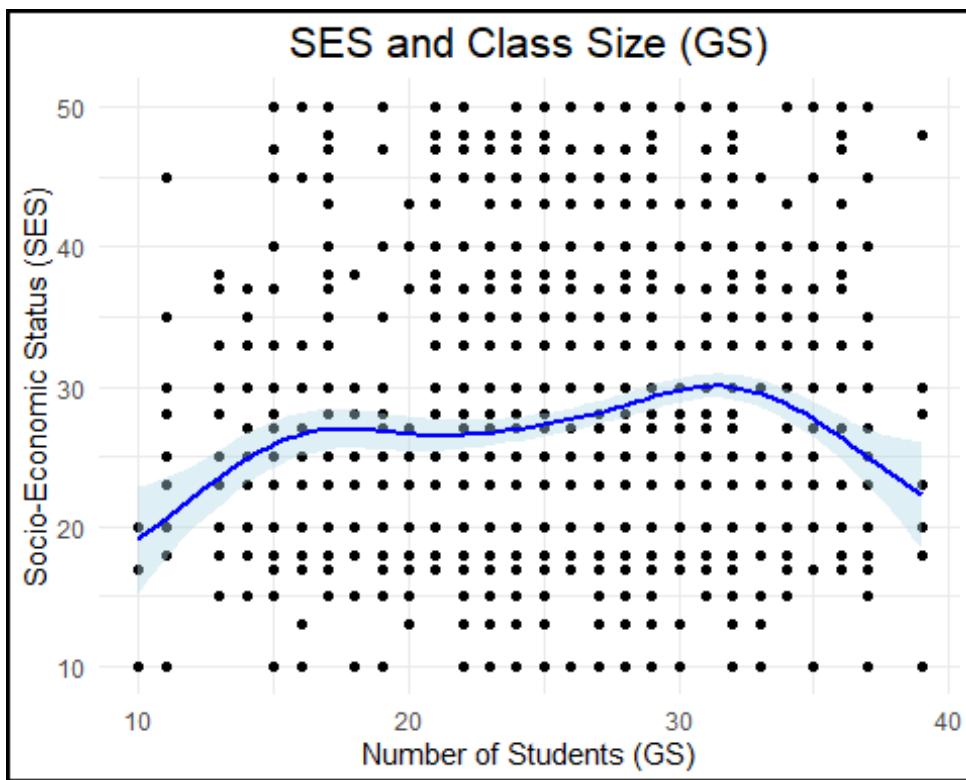
- **Large vs. Small (p < 0.001)**: The mean SES for Large classes is highly significantly higher than for Small classes (diff ≈ 3.83).

- **Large vs. Medium (p < 0.001)**: The mean SES for Large classes is highly significantly higher than for Medium classes (diff ≈ 1.99).

These findings establish a clear and statistically significant positive gradient in mean Socio-Economic Status (SES) that aligns directly with class size:

Mean SES (Small) < Mean SES (Medium) < Mean SES (Large)

This rigorously confirms the presence of selection bias, where students in larger classes, on average, originate from a higher socio-economic background than those in smaller or medium-sized classes.

```r
# SCATTER PLOT: SES vs. Class Size (GS)
ggplot(nlschools, aes(x = GS, y = SES)) +

  # Plot individual data points
  geom_point(color = "black") +

  # Add a non-linear smoothing line (GAM) with 95% confidence band
  geom_smooth(
    method = "gam",
    formula = y ~ s(x),      # Specify a smoothing spline for the trend
    se = TRUE,               # Show 95% confidence interval
    color = "blue",
    fill = "lightblue"
  ) +

  # Set plot title and axis labels
  labs(
    x = "Number of Students (GS)",
    y = "Socio-Economic Status (SES)",
    title = "SES and Class Size (GS)"
  ) +

  # Apply minimal theme and custom border
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```

The plot reveals a mild, non-linear, but systematic positive relationship between Class Size (GS) and average Socio-Economic Status (SES). This pattern validates the earlier conclusion that initial equivalence on SES is not met across all class sizes.

- **Positive Trend**: The Locally Estimated Scatterplot Smoothing (LOESS) curve initiates at the lowest average SES (approximately 20) for the smallest class sizes (GS ≈ 10).

- **Peak SES in Large Classes**: Average SES rises steadily with increasing class size, reaching its peak (SES ≈ 30) within the moderately large class range (GS ≈ 30-34).

- **Marginal Decline**: A slight attenuation of the average SES is observed for the largest classes (GS > 35), following the peak.

While the individual data points exhibit wide vertical dispersion at all class sizes, confirming the presence of both high and low SES students across the spectrum, the systematic upward shift in the central tendency confirms that students in small classes have the lowest average SES, and those in moderately large classes have the highest.

```r
# PEARSON CORRELATION TEST: GS vs. SES
# Test for a linear relationship between Class Size (GS) and SES.
cor_test_ses <- cor.test(nlschools$GS, nlschools$SES, method = "pearson")
print(cor_test_ses)
```

```
##
##  Pearson's product-moment correlation
##
## data:  nlschools$GS and nlschools$SES
## t = 4.9639, df = 2285, p-value = 7.418e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.06256587 0.14366874
## sample estimates:
##      cor
## 0.103289
```

```
# SIMPLE LINEAR REGRESSION: SES ~ GS
# Model SES predicted by Class Size (GS).
regression_model_ses <- lm(SES ~ GS, data = nlschools)
# View the model summary (coefficients, R-squared, and p-values).
summary(regression_model_ses)
```

```
##
## Call:
## lm(formula = SES ~ GS, data = nlschools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.269  -8.104  -1.759   8.108  24.454
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.59420    1.07535  21.011  < 2e-16 ***
## GS           0.19679    0.03964   4.964 7.42e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.85 on 2285 degrees of freedom
## Multiple R-squared:  0.01067,    Adjusted R-squared:  0.01024
## F-statistic: 24.64 on 1 and 2285 DF,  p-value: 7.418e-07
```

Pearson Correlation Analysis - **Correlation Coefficient(r)**: 0.103289 - This indicates a weak positive linear association between Class Size and Socio-Economic Status (SES).

- **P-value**: 7.418e-07
- Given that the p-value is substantially below the significance threshold (alpha = 0.05), the null hypothesis of no correlation is strongly rejected.

This confirms a statistically significant, weak positive linear correlation between Class Size and SES.
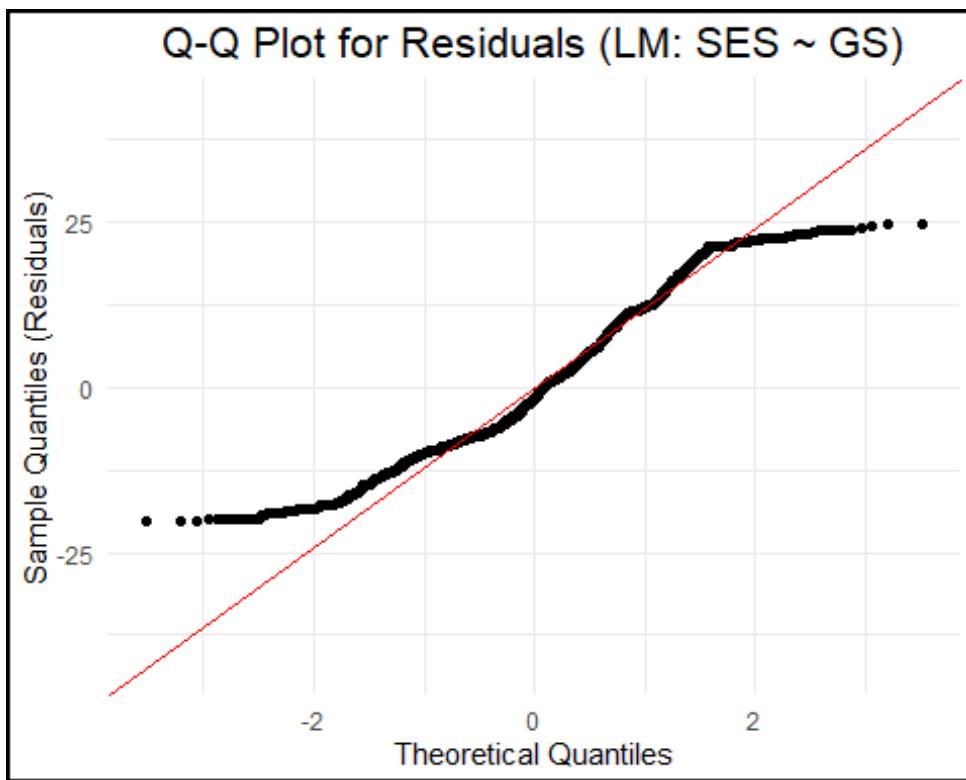
Simple Linear Regression Model The simple linear regression model corroborates the statistically significant positive relationship. Larger class sizes are associated with slightly

higher SES scores. However, the coefficient of determination (R^2) is low ($\approx 1\%$), indicating that Class Size accounts for a minimal proportion of the total variance in SES scores.

```r
# SIMPLE LINEAR REGRESSION: SES ~ GS
lm_model_ses <- lm(SES ~ GS, data = nlschools)

# Extract and store the model residuals for diagnostic checking
nlschools$ses_residuals <- residuals(lm_model_ses)

# Q-Q PLOT: CHECK RESIDUALS NORMALITY
# Use the Q-Q plot to visually assess the normality assumption of the
residuals.
ggplot(nlschools, aes(sample = ses_residuals)) +
  stat_qq() +
  stat_qq_line(color = "red") +
  labs(title = "Q-Q Plot for Residuals (LM: SES ~ GS)",
       x = "Theoretical Quantiles",
       y = "Sample Quantiles (Residuals)") +
  theme_minimal() +
  # Add a border around the entire plot
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA
    )
  )
```

Q-Q Plot for Residuals (LM: SES ~ GS)

The primary finding is that the assumption of normality of residuals is significantly violated. The systematic departure of the data points from the theoretical line indicates that the distribution of errors is not normal.

The plot exhibits a noticeable S-shaped curvature, which suggests that the actual distribution of residuals is leptokurtic (heavier tails) or otherwise deviates from the normal reference. Specifically, the points fall below the line in the low tail, cross above the line in the center, and cross below again in the high tail.

Severe Plateaus at Extremes

- **Lower Tail (Left)**: A pronounced horizontal plateau is evident, with points clustering around -25 to -20 at Theoretical Quantiles below -1.5. This clustering suggests the presence of a distinct group of extreme low outliers or a severe floor effect, significantly deviating from the expected normal curve.

- **Upper Tail (Right)**: A similar horizontal cluster is observed around 25 at Theoretical Quantiles above 2, indicating a group of extreme high outliers.

The non-normal distribution of residuals invalidates the core assumptions underlying the standard linear model (LM: SES ~ GS). Consequently, the standard errors and hypothesis tests (p-values) derived from this model may be unreliable or statistically compromised.

```
# MODEL TRANSFORMATION & FITTING (log(SES) ~ GS)
# Use log transformation to better meet the linear model's normality
assumptions.
model_log_ses_gs <- lm(log(SES) ~ GS, data = nlschools)

# Extract the residuals from the transformed model
residuals_log_ses_gs <- residuals(model_log_ses_gs)

# RESIDUALS NORMALITY CHECK
# Perform Shapiro-Wilk test to formally check if residuals are normally
distributed.
shapiro_test_log_ses <- shapiro.test(residuals_log_ses_gs)
print(shapiro_test_log_ses)

##
##  Shapiro-Wilk normality test
##
## data:  residuals_log_ses_gs
## W = 0.97035, p-value < 2.2e-16

# Visually check residual normality using a Q-Q plot
qqnorm(residuals_log_ses_gs)
qqline(residuals_log_ses_gs, col = "blue")
```
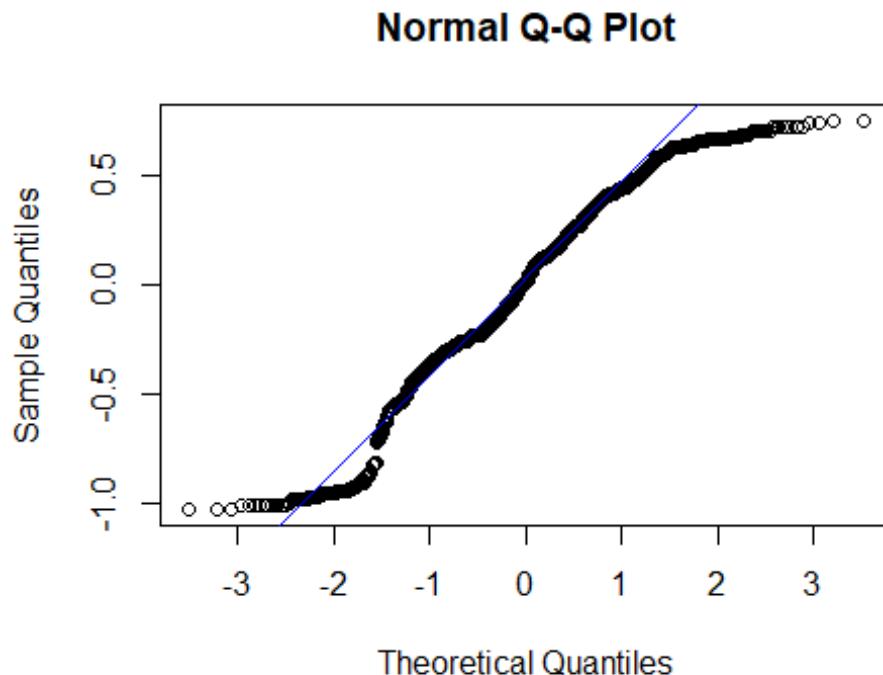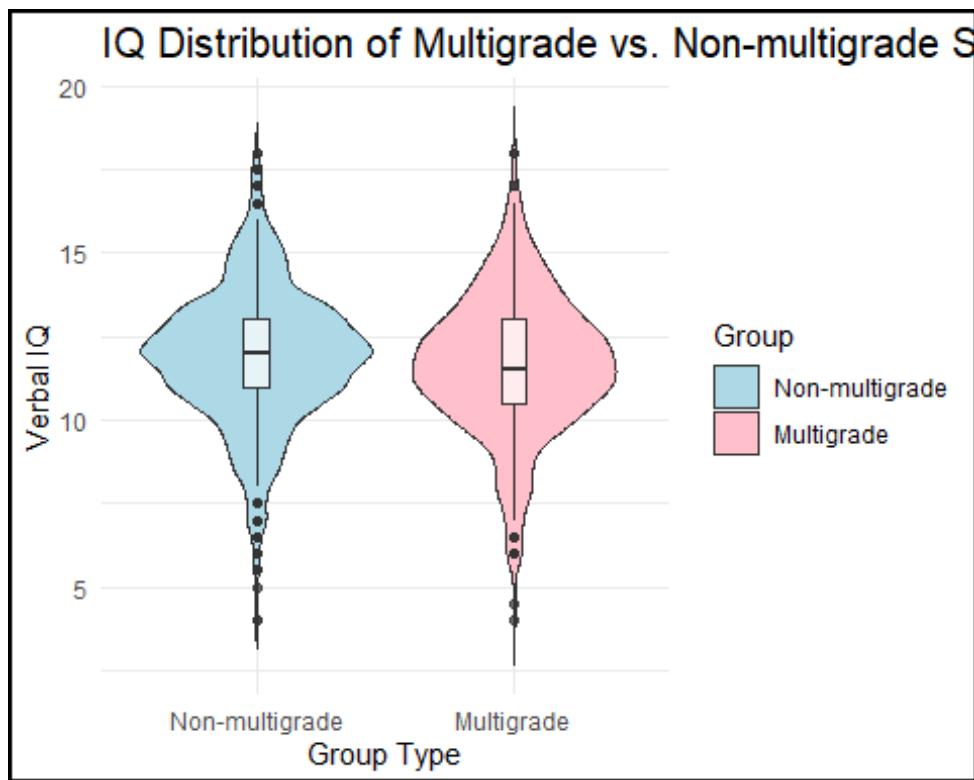

The Shapiro-Wilk test results formally confirm that the model residuals, even following a logarithmic transformation of the Socio-Economic Status (SES) variable, are not normally distributed.

Given this persistent violation of the normality assumption, all subsequent regression models testing the effect of Class Size (GS) on student outcomes will utilize Robust Standard Errors to ensure valid statistical inference.

```r
# --- CREATE VIOLIN PLOT: IQ Distribution (Multigrade vs. Non-multigrade) ---
ggplot(nlschools, aes(x = factor(COMB), y = IQ, fill = factor(COMB))) +

  # Show the density distribution
  geom_violin(trim = FALSE) +

  # Overlay a boxplot for clear median and quartile representation
  geom_boxplot(width = 0.1, fill = "white", alpha = 0.7) +

  # Set plot title, axis, and legend labels
  labs(title = "IQ Distribution of Multigrade vs. Non-multigrade Students",
       x = "Group Type",
       y = "Verbal IQ",
       fill = "Group") +

  # Manually define colors and labels for the two groups (0 and 1)
  scale_fill_manual(values = c("0" = "lightblue", "1" = "pink"),
                    labels = c("0" = "Non-multigrade", "1" = "Multigrade")) +

  # Label the x-axis categories
  scale_x_discrete(labels = c("0" = "Non-multigrade", "1" = "Multigrade")) +

  theme_minimal() +
  theme(
    plot.title = element_text(size = 15),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
    )
```

IQ Distribution of Multigrade vs. Non-multigrade S

This violin plot compares the distribution of Verbal IQ scores across Non-multi-grade and Multi-grade classrooms, serving as a critical check for initial equivalence on cognitive ability.

The visual evidence strongly suggests initial equivalence on Verbal IQ across the multi-grade status variable (COMB):

- **Central Tendency**: The median Verbal IQ (represented by the white line) is highly consistent across both groups, centering at approximately 12 for both Non-multi-grade and Multi-grade classrooms.

- **Distribution**: Both distributions are symmetric, span the full range of observed scores (approximately 4 to 19), and exhibit highly comparable Interquartile Ranges (IQR) and overall violin shapes.

The high consistency in medians, ranges, and distribution shapes confirms the absence of a systematic difference in cognitive ability based on multi-grade status. This is a favorable finding, as it indicates that any subsequent differences in academic outcomes found between these two groups are unlikely to be confounded by pre-existing differences in student cognitive ability.

```
# TWO-SAMPLE T-TEST: IQ vs. COMB Group
# Test if the mean IQ differs between the two multigrade groups (COMB = 0 and
COMB = 1).
```

```
t_test_iq_comb <- t.test(IQ ~ COMB, data = nlschools)
print(t_test_iq_comb)

##
##  Welch Two Sample t-test
##
## data:  IQ by COMB
## t = 2.4427, df = 1106.2, p-value = 0.01473
## alternative hypothesis: true difference in means between group 0 and group
1 is not equal to 0
## 95 percent confidence interval:
##  0.0470796 0.4315345
## sample estimates:
## mean in group 0 mean in group 1
##        11.89988         11.66057
```

The primary finding is the identification of a statistically significant difference in mean Verbal IQ between the Non-multi-grade and Multi-grade classroom groups.

Students in Non-multi-grade classes (Group 0) exhibit a significantly higher mean Verbal IQ (≈ 11.90) compared to students in Multi-grade classes (≈ 11.66).

While this finding violates the assumption of initial equivalence, the raw difference in means (≈ 0.24 points) is quantitatively minor. This distinction highlights a case where the effect is statistically significant but lacks practical significance; however, it remains a potential confounding variable that must be accounted for in subsequent causal inference modeling.
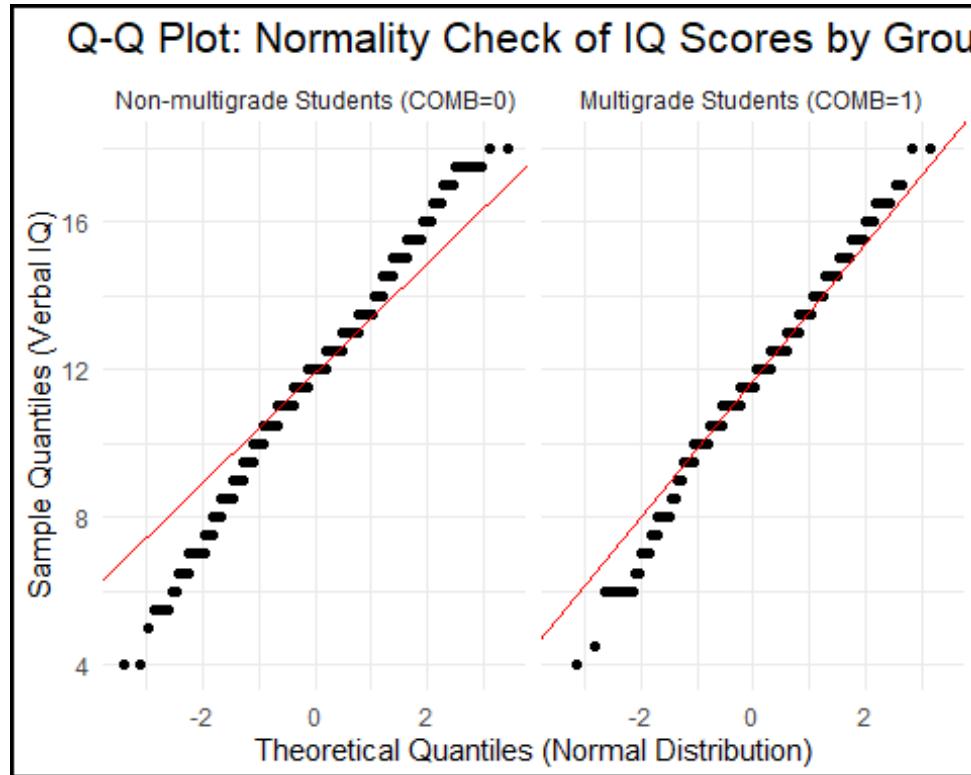
```
# FACETED Q-Q PLOT: Check IQ Normality by COMB Group
ggplot(nlschools, aes(sample = IQ)) +

  # Plot data quantiles against theoretical normal quantiles
  stat_qq() +
  # Add the reference line for a perfect normal distribution
  stat_qq_line(color = "red") +

  # Create separate panels for each COMB group
  facet_wrap(~ factor(COMB),
             labeller = as_labeller(c("0" = "Non-multigrade Students
(COMB=0)",
                                      "1" = "Multigrade Students (COMB=1)")))
+

  # Set plot title and axis labels
  labs(title = "Q-Q Plot: Normality Check of IQ Scores by Group",
       x = "Theoretical Quantiles (Normal Distribution)",
       y = "Sample Quantiles (Verbal IQ)") +
```

```r
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5, size = 16),
  # Add a black border around the entire plot
  plot.background = element_rect(
    color = "black",
    linewidth = 1.5,
    fill = NA )
  )
```


Q-Q Plot: Normality Check of IQ Scores by Group

The primary finding is that the Verbal IQ scores largely conform to a normal distribution within both the Multi-grade and Non-multi-grade groups.

For both classroom types, the vast majority of data points representing the central bulk of scores—align closely with the theoretical line. This indicates that the distributions are acceptably normal for standard statistical analysis.

Minor Tail Deviations: There are minor deviations at the extreme ends:

- **Lower Tail (Left)**: Points in both plots exhibit a slight curvature below the theoretical line at the extreme left (Theoretical Quantiles $\approx$ -2 and below). This suggests the distributions may possess marginally heavier lower tails (or a slight negative skew) compared to a perfect normal distribution.

- **Upper Tail (Right)**: The agreement is strong, with the highest scores in the Non-multi-grade group showing only a minimal departure from the theoretical line.

This visual inspection confirms that both classroom types draw students from populations with highly similar and acceptably symmetric Verbal IQ profiles.

```
# SHAPIRO-WILK TEST BY GROUP
# Check the normality of IQ scores for each COMB group separately.
normality_check_iq_comb <- by(nlschools$IQ, nlschools$COMB, shapiro.test)
print(normality_check_iq_comb)

## nlschools$COMB: 0
##
##  Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.9853, p-value = 5.604e-12
##
## -------------------------------------------------------------
## nlschools$COMB: 1
##
##  Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.98442, p-value = 3.004e-06
```

The Shapiro-Wilk test formally indicated a statistically significant departure from normality for Verbal IQ within both the Non-multi-grade and Multi-grade groups ($p < 0.05$).

However, visual inspection of the Q-Q plot confirmed that the distributions are near-normal in the central range, with deviations confined primarily to the tails. Given the large sample size and the established robustness of the t-test to minor tail non-normality, the finding of a statistically significant IQ discrepancy is accepted. Consequently, Verbal IQ will be retained as a necessary covariate in subsequent causal inference modeling.

```
# DATA PREPARATION: Categorize SES
# Create the 'SES_Group' variable by binning continuous SES into three
levels.
nlschools$SES_Group <- cut(nlschools$SES,
                        breaks = c(9, 20, 35, 50),
                        labels = c("Low SES", "Medium SES", "High SES"),
                        right = TRUE)

# STACKED BAR PLOT: SES Proportion by Multigrade Status
ggplot(nlschools, aes(x = factor(COMB), fill = factor(SES_Group))) +

  # Create stacked bars where the height represents the proportion (100%)
  geom_bar(position = "fill") +
```
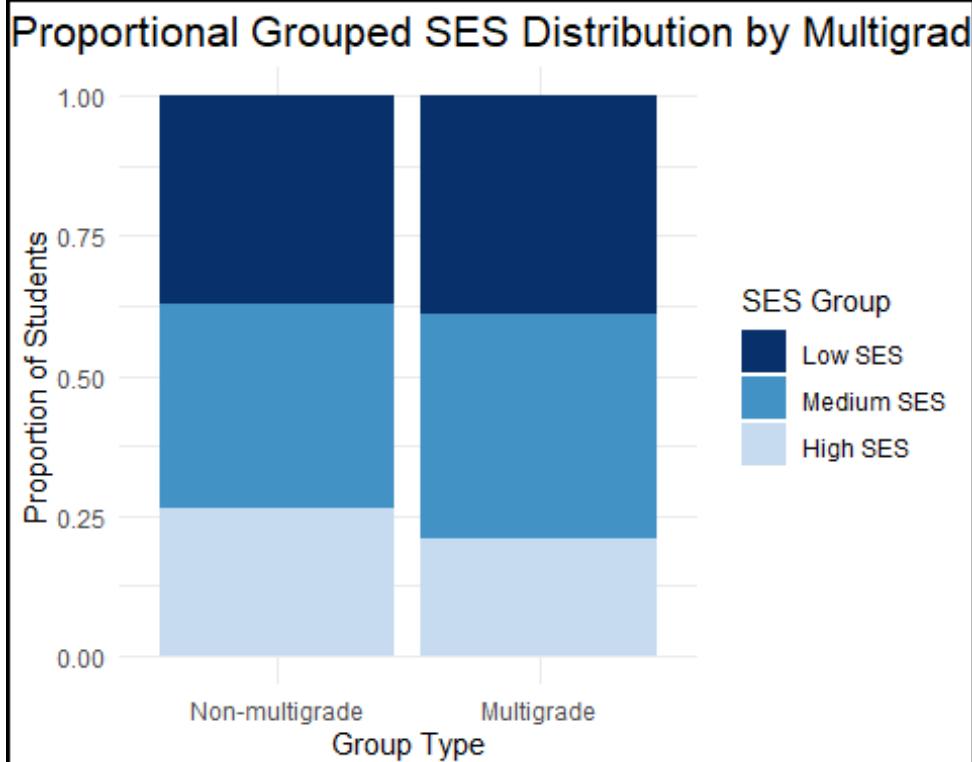
```r
# Set plot title, axis, and legend labels
labs(title = "Proportional Grouped SES Distribution by Multigrade Status",
     x = "Group Type",
     y = "Proportion of Students",
     fill = "SES Group") +

# Manually define colors for the SES groups
scale_fill_manual(values = c("Low SES" = "#08306B",
                             "Medium SES" = "#4292C6",
                             "High SES" = "#C6DBEF")) +

# Label the x-axis categories
scale_x_discrete(labels = c("0" = "Non-multigrade", "1" = "Multigrade")) +

theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.2, size = 16),
  # Add a black border around the entire plot
  plot.background = element_rect(
    color = "black",
    linewidth = 1.5,
    fill = NA )
)
```

The grouped 100% stacked bar chart provides a comparative visualization of the Socio-Economic Status (SES) distribution between students in Non-multigrade and Multigrade classrooms.

The overall proportional distribution of SES groups is highly comparable between the two classroom types, indicating the absence of a significant disparity where one classroom setting is disproportionately populated by students from a single SES extreme.

Detailed Group Proportions The estimated proportions for each SES category are remarkably consistent across both groupings: - **Low SES (Dark Blue)**: This segment represents the largest single proportion, accounting for approximately 40-42% of the student population in both classroom types.

- **Medium SES (Middle Blue)**: This intermediate segment constitutes the next largest portion, estimated at about 35-38% across both groups.

- **High SES (Light Blue)**: This group represents the smallest segment, accounting for approximately 22-25% of the student population in both groups.

While the distributions demonstrate strong similarity, a subtle shift is observable: the Multigrade student population appears to contain a marginally higher proportion of Low SES students (by approximately 2 percentage points) and a marginally lower proportion of High SES students (by approximately 3 percentage points) compared to the Non-multigrade population.

This visual evidence suggests that students are distributed across Non-multigrade and Multigrade classrooms equitably with respect to SES, with only a minor, almost negligible tendency for Multigrade settings to enroll a slightly higher percentage of students from the lowest SES strata.

```
# DATA PREPARATION: Categorize SES
# Bin continuous SES into three categorical levels (Low, Medium, High).
nlschools$SES_Group <- cut(nlschools$SES,
                     breaks = c(9, 20, 35, 50),
                     labels = c("Low SES", "Medium SES", "High SES"),
                     right = TRUE)

# CHI-SQUARED TEST OF INDEPENDENCE
# Create the contingency table (COMB group vs. new SES category)
ses_grouped_table <- table(nlschools$COMB, nlschools$SES_Group)
print(ses_grouped_table)

##
##      Low SES Medium SES High SES
##   0     616        605      437
##   1     245        253      131
```

```r
# Perform Chi-Squared Test to see if the two variables are independent
chi_sq_grouped_test <- chisq.test(ses_grouped_table)
print(chi_sq_grouped_test)

##
##  Pearson's Chi-squared test
##
## data:  ses_grouped_table
## X-squared = 7.7006, df = 2, p-value = 0.02127

cat("Standardized Residuals (Post-Hoc Analysis)\n")

## Standardized Residuals (Post-Hoc Analysis)

# Print standardized residuals to identify cells driving the significant
result
print(chi_sq_grouped_test$stdres)

##
##        Low SES Medium SES  High SES
##   0 -0.792251  -1.646376  2.733323
##   1  0.792251   1.646376 -2.733323
```

Chi-squared test The Chi-squared test of independence indicated a statistically significant association between students' Socio-Economic Status (SES) and their Multigrade Status (p < 0.05). This finding confirms that the distribution of students across the Low, Medium, and High SES categories is not independent of the type of classroom (Multigrade vs. Non-multigrade) attended.

Specifically, Non-multigrade classes (Group 0) are observed to have a higher proportion of students in the High SES category, whereas Multigrade classes (Group 1) tend to be marginally overrepresented in the Low and Medium SES categories. While the overall effect size is modest, the association is statistically significant, suggesting a modest socioeconomic imbalance between the two class types.
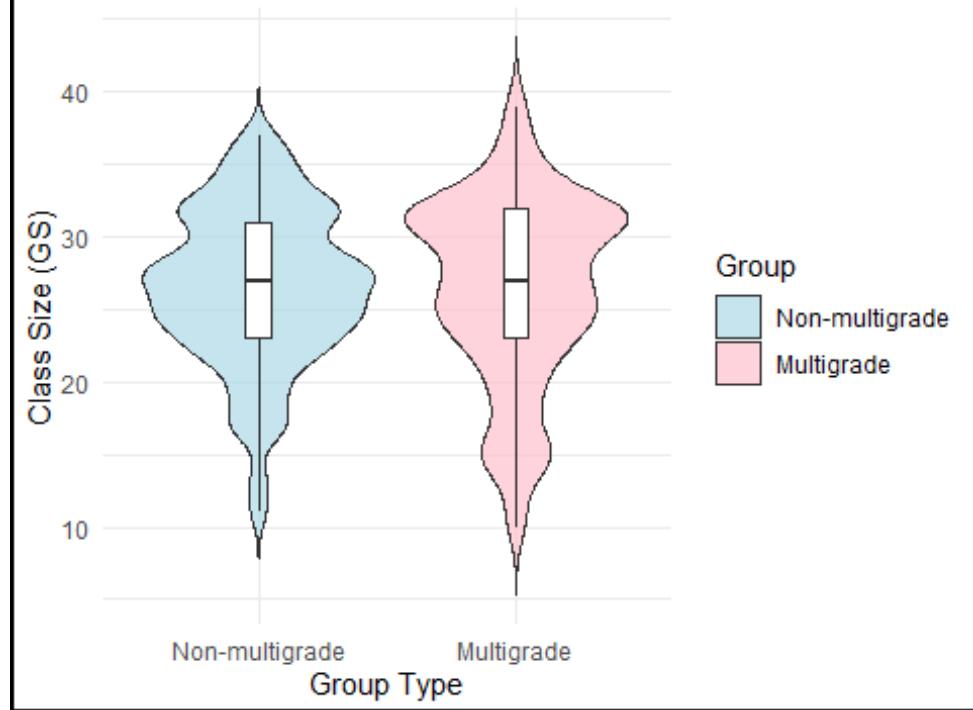
Post-Hoc Analysis A post-hoc analysis utilizing adjusted standardized residuals revealed that the observed difference is driven exclusively by the High SES category: - **High SES (Residuals = 2.733 and -2.733)**: Both residuals exceed the 1.96 threshold. - **Non-multigrade (0)**: The residual of 2.733 indicates that the count of High SES students is significantly higher than expected under the assumption of independence.

- **Multigrade (1)**: The residual of -2.733 indicates that the count of High SES students is significantly lower than expected.

- **Low and Medium SES**: The absolute residuals for these categories (0.792 and 1.646) are below the critical threshold of 1.96, indicating that the observed counts do not significantly deviate from the expected counts.

The results demonstrate that Non-multigrade classes are significantly enriched with High SES students, while Multigrade classes are significantly underrepresented in this group. This establishes that SES distribution is a confounding factor that must be accounted for in subsequent comparative analyses between the two classroom settings.

```r
# CREATE VIOLIN PLOT: Class Size (GS) Distribution by Multigrade Statu
ggplot(nlschools, aes(x = factor(COMB), y = GS, fill = factor(COMB))) +

  # Show the density distribution
  geom_violin(trim = FALSE, alpha = 0.7) +

  # Overlay a boxplot for clear median and quartile representation
  geom_boxplot(width = 0.1, fill = "white", outlier.shape = NA) +

  # Set plot title and axis labels
  labs(title = "Class Size (GS) Distribution by Multigrade Status",
       x = "Group Type",
       y = "Class Size (GS)",
       fill = "Group") +

  # Label the x-axis categories
  scale_x_discrete(labels = c("0" = "Non-multigrade", "1" = "Multigrade")) +

  # Manually define colors and remove the legend
  scale_fill_manual(values = c("0" = "lightblue", "1" = "pink"),
                    labels = c("0" = "Non-multigrade", "1" = "Multigrade")) +

  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```

This violin plot compares the distribution of continuous Class Size (GS) between Non-multigrade and Multigrade student groups, serving as a check for initial equivalence.

The analysis suggests near-equivalence in the central tendencies and overall spread: - **Central Tendency and Spread**: The Interquartile Range (IQR), represented by the boxplot height, is virtually identical for both groups. Both distributions are generally symmetrical around the median, indicating a lack of extreme skewness, and the absence of visible outliers confirms the distributions are relatively well-behaved.

- **Density Differences**: Despite overall similarity, subtle differences in density are observed:
- The Non-multigrade distribution exhibits a marginally wider density around the median, suggesting a greater concentration of students near the average class size.
- The Multigrade distribution appears marginally wider at the extremes (around GS ≈ 10 and GS ≈ 40), indicating a slightly higher relative frequency of both the smallest and largest classes.

These observations reinforce the overall finding of distributional similarity in Class Size across the two classroom groupings, with only minor discrepancies in the relative frequency of extreme class sizes.

```
# TWO-SAMPLE T-TEST: Class Size (GS) vs. COMB Group
# Test if the mean Class Size differs between multigrade (COMB = 1) and non-
```

```
multigrade (COMB = 0) schools.
t_test_gs_comb <- t.test(GS ~ COMB, data = nlschools)
print(t_test_gs_comb)

##
##  Welch Two Sample t-test
##
## data:  GS by COMB
## t = 0.4016, df = 979.97, p-value = 0.6881
## alternative hypothesis: true difference in means between group 0 and group
1 is not equal to 0
## 95 percent confidence interval:
##  -0.4531071  0.6862782
## sample estimates:
## mean in group 0 mean in group 1
##        26.54584        26.42925
```

The mean class size in Non-multigrade classes ($\approx$ 26.55 students) was found to be statistically equivalent to the mean class size in Multigrade classes ($\approx$ 26.43 students).

The statistical test yielded a non-significant p-value (p = 0.6881, where p > 0.05), preventing the rejection of the null hypothesis of no difference. The observed mean difference ($\approx$ 0.12 students) is quantitatively negligible.

This result confirms that the Multigrade and Non-multigrade student groups are statistically equivalent in terms of their mean class size. Consequently, Class Size (GS) is eliminated as a potential confounding variable in subsequent analyses comparing outcomes between the two classroom groupings.

2. When did students perform better or worse on the language exam? Describe which variables had the most important effects.

I used three different plots:

- **Relationship between Language Exam Scores and Verbal IQ**: This plot establishes the baseline explanatory power of a student's intrinsic cognitive capacity (Verbal IQ) on Language Exam outcomes. It quantifies the fundamental, expected relationship between innate verbal reasoning and academic achievement.

- **Relationship between Language Exam Scores and Socio-Economic Status (SES)**: This plot quantifies the effect of the extrinsic learning environment on Language Exam scores. SES serves as a powerful aggregate proxy for a student's learning resources, parental educational capital, and exposure to rich linguistic environments outside of the school structure.

- **Relationship between Language Exam Scores and Class Size(GS)**: This plot isolates and evaluates the impact of the institutional/pedagogical environment

(Class Size) on student performance. It tests the hypothesis that class density, a variable directly controlled by educational administration, influences achievement through changes in individualized attention and instructional efficacy.
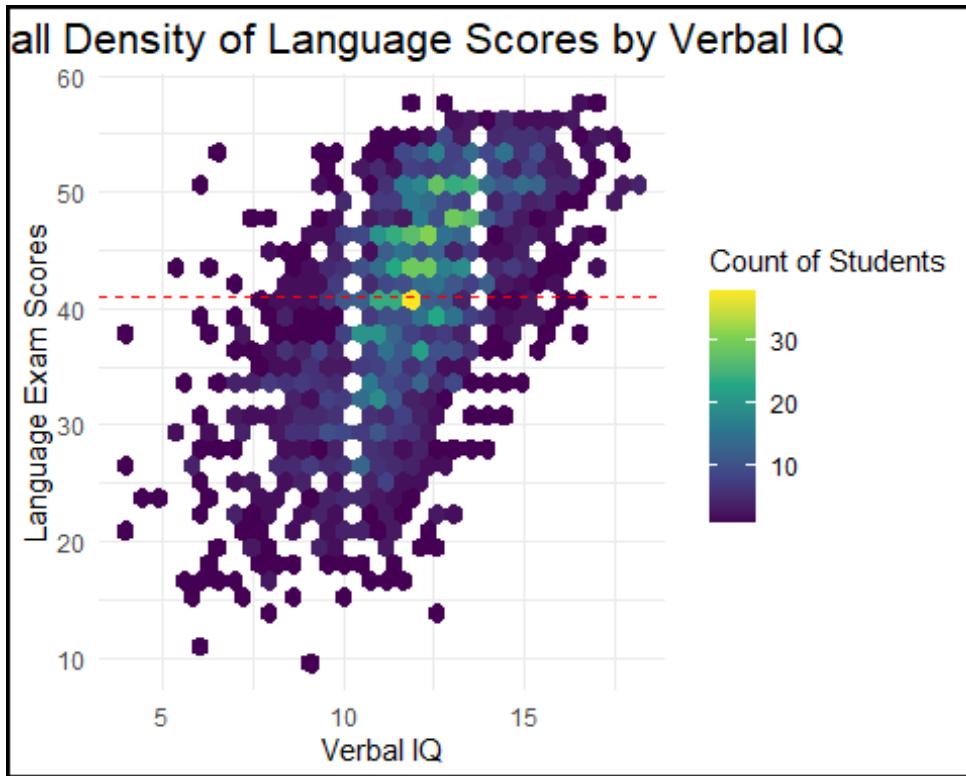
```r
# Calculate the overall average Language Exam Score
average_lang <- mean(nlschools$lang, na.rm = TRUE)

# CREATE HEXBIN PLOT: Language Scores vs. Verbal IQ
ggplot(nlschools, aes(x = IQ, y = lang)) +

  # Display the density of points using hexagonal bins
  geom_hex() +
  scale_fill_continuous(type = "viridis") +

  # Add a horizontal line for the overall average Language Exam Score
  geom_hline(yintercept = average_lang, linetype = "dashed", color = "red") +

  # Set plot title and axis labels
  labs(
    title = "Overall Density of Language Scores by Verbal IQ",
    x = "Verbal IQ",
    y = "Language Exam Scores",
    fill = "Count of Students"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```

The hexbin plot visualizes the joint distribution and relationship between a student's Verbal IQ and their Language Exam Scores, with the color gradient representing the count of students.

A clear, strong positive linear relationship is evident: as Verbal IQ increases, Language Exam Scores tend to increase. This strong association is demonstrated by the data cloud running diagonally from the bottom-left to the top-right corner, with data points clustering relatively closely around a linear trend. This suggests that Verbal IQ is a robust predictor of Language Exam Scores. - **Concentration of Scores**: The highest concentration of students (indicated by the yellow/light green hexes, > 30 students) is centered in the range where Verbal IQ is $2-14 and Language Exam Scores are 40-50. This central cluster confirms that the majority of students possess above-average cognitive ability and perform well on the exam.

- **Variability**: Despite the strong correlation, a considerable vertical spread in Language Scores is observed at any given Verbal IQ level (e.g., IQ ≈ 12 shows scores ranging from approximately 15 to 55). This residual variability indicates that while Verbal IQ is strongly related to exam performance, other factors significantly contribute to the Language Exam outcome.

The plot reveals a strong positive correlation, yet it simultaneously highlights the considerable heterogeneity in student achievement that exists across any single level of cognitive ability.

```r
# Calculate the overall average Language Exam Score (ignoring missing values)
average_lang <- mean(nlschools$lang, na.rm = TRUE)

# CREATE FACETED HEXBIN PLOT: Language Scores vs. Verbal IQ by Group
ggplot(nlschools, aes(x = IQ, y = lang)) +

  # Display the density of points using hexagonal bins
  geom_hex() +
  scale_fill_continuous(type = "viridis") +

  # Add a horizontal line across both facets for the overall average language
score
  geom_hline(yintercept = average_lang, linetype = "dashed", color = "red") +

  # Create separate panels for the two multigrade groups
  facet_wrap(~ factor(COMB),
             labeller = as_labeller(c("0" = "Non-multigrade", "1" =
"Multigrade"))) +

  # Set plot title and axis labels
  labs(
    title = "Density of Language Scores by Verbal IQ, Separated by Multigrade
Status",
    x = "Verbal IQ",
    y = "Language Exam Scores",
    fill = "Count of Students"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.2, size = 14),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```
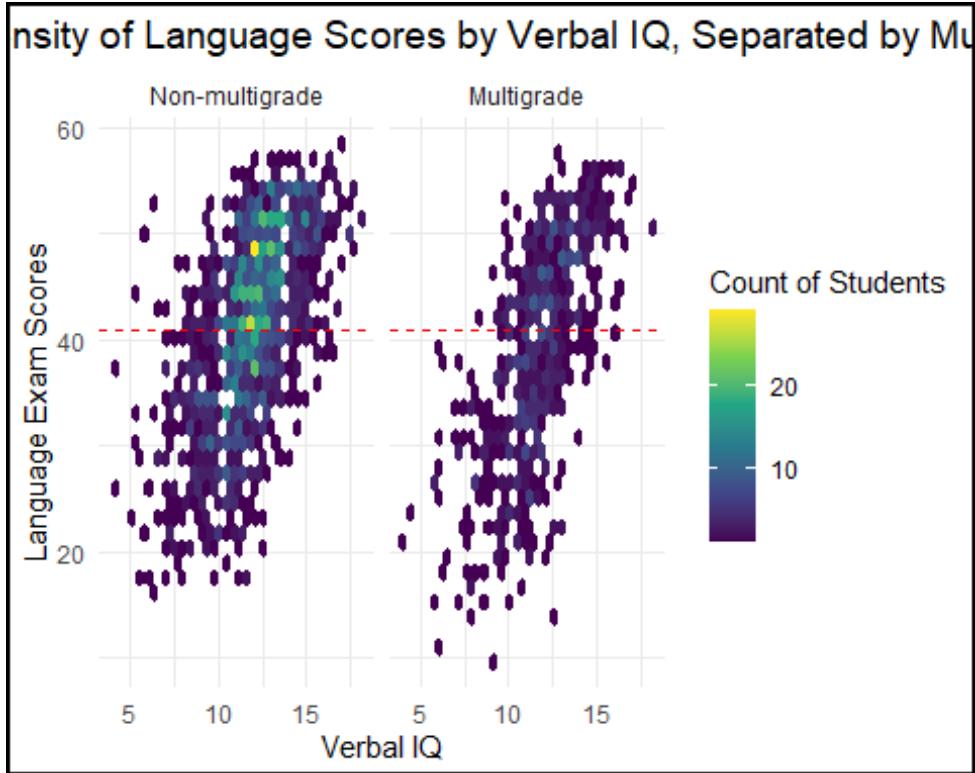
Density of Language Scores by Verbal IQ, Separated by Mu...

This dual hexbin plot is crucial for testing the equivalence of the relationship between Verbal IQ and Language Exam Scores across the two classroom settings: Non-multigrade and Multigrade.

Relationship Equivalence: Both the Non-multigrade and Multigrade panels exhibit a clear, strong positive linear relationship. The direction, slope, and general spread (variance) of the high-density data cloud are remarkably similar between the two groups. This suggests that the modality of the influence of Verbal IQ on exam performance is not conditional on the class setting.

Concentration Equivalence: In both panels, the highest density of students (yellow/light green hexes) is consistently centered around a Verbal IQ of 12-14 and a Language Exam Score of 40-50. This concentration, which generally falls above the dashed red line, indicates that the most frequent combination of scores is equivalent and falls within the mid-to-high range for both classroom types.

This visual comparison strongly suggests that the predictive power of Verbal IQ on Language Exam performance is statistically equivalent regardless of whether a student is in a Multigrade or Non-multigrade classroom.

```
# OVERALL CORRELATION (Ignoring COMB)
# Calculate the overall Pearson correlation between Language Score and IQ.
print("Overall IQ Correlation (Language Score vs. IQ):")
```

```
## [1] "Overall IQ Correlation (Language Score vs. IQ):"

cor.test(nlschools$lang, nlschools$IQ)

##
##  Pearson's product-moment correlation
##
## data:  nlschools$lang and nlschools$IQ
## t = 36.781, df = 2285, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5834141 0.6349371
## sample estimates:
##       cor
## 0.6098195

# NON-MULTIGRADE CORRELATION (COMB = 0)
# Filter data for non-multigrade schools.
nlschools_nonmulti <- subset(nlschools, COMB == 0)
# Calculate correlation for the non-multigrade group.
print("Non-multigrade IQ Correlation (Language Score vs. IQ):")

## [1] "Non-multigrade IQ Correlation (Language Score vs. IQ):"

cor.test(nlschools_nonmulti$lang, nlschools_nonmulti$IQ)

##
##  Pearson's product-moment correlation
##
## data:  nlschools_nonmulti$lang and nlschools_nonmulti$IQ
## t = 29.856, df = 1656, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5593218 0.6219640
## sample estimates:
##       cor
## 0.5915348

# MULTIGRADE CORRELATION (COMB = 1)
# Filter data for multigrade schools.
nlschools_multi <- subset(nlschools, COMB == 1)
# Calculate correlation for the multigrade group.
print("Multigrade IQ Correlation (Language Score vs. IQ):")

## [1] "Multigrade IQ Correlation (Language Score vs. IQ):"

cor.test(nlschools_multi$lang, nlschools_multi$IQ)

##
##  Pearson's product-moment correlation
##
## data:  nlschools_multi$lang and nlschools_multi$IQ
```

```
## t = 21.44, df = 627, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6028614 0.6933103
## sample estimates:
##      cor
## 0.6503853
```

The strong, positive linear relationship between Verbal IQ and Language Exam Scores, visually identified in the preceding plots, was formally confirmed by Pearson's product-moment correlation across all groups, with all tests yielding highly statistically significant p-values.

However, the strength of this relationship exhibits a marginal variation by class type: - The correlation coefficient for the Non-multigrade group (r = 0.59) is slightly weaker than the coefficient for the Multigrade group (r = 0.65).

This difference suggests that while Verbal IQ is a critical predictor in both environments, its relationship with Language Exam Scores is marginally stronger within Multigrade classrooms. This distinction warrants attention in subsequent modeling.

```
# LINEAR REGRESSION WITH INTERACTION
# Model Language Score (lang) predicted by IQ, COMB group, and their
interaction.
# The interaction term (IQ * COMB) tests if the slope of IQ on lang differs
between the groups.
model <- lm(lang ~ IQ * COMB, data = nlschools)
# Print the full model summary (coefficients, R-squared, and interaction p-
value).
summary(model)

##
## Call:
## lm(formula = lang ~ IQ * COMB, data = nlschools)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -27.768  -4.484   0.473   5.153  24.646
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.4068     1.0228  12.131  < 2e-16 ***
## IQ            2.4533     0.0847  28.966  < 2e-16 ***
## COMB         -9.2019     1.8875  -4.875 1.16e-06 ***
## IQ:COMB       0.6317     0.1584   3.987 6.90e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 7.071 on 2283 degrees of freedom
## Multiple R-squared:  0.3841, Adjusted R-squared:  0.3832
## F-statistic: 474.5 on 3 and 2283 DF,  p-value: < 2.2e-16
```

The model confirms that Verbal IQ is a robust predictor of Language Exam Scores.

Overall Model Fit: The model achieved a highly significant F-statistic (F = 474.5 with p < 2.2 X 10^-16) and accounted for a substantial proportion of the variance in Language Scores (Adjusted R^2 = 0.3832).

Interaction Term: The statistically significant interaction term (IQ:COMB1, p < 0.001) confirms that the predictive power of Verbal IQ is significantly stronger (i.e., the slope is steeper) in Multigrade classes compared to Non-multigrade classes.

The significant main effect of COMB1 suggests that, on average, students in Multigrade classes have significantly lower adjusted Language Scores compared to their Non-multigrade counterparts, holding Verbal IQ constant.

The analysis indicates that Multigrade status is associated with a lower adjusted mean Language Score, and this negative association is compounded by a stronger dependence of the Language Score on cognitive ability (steeper slope) within the Multigrade group.
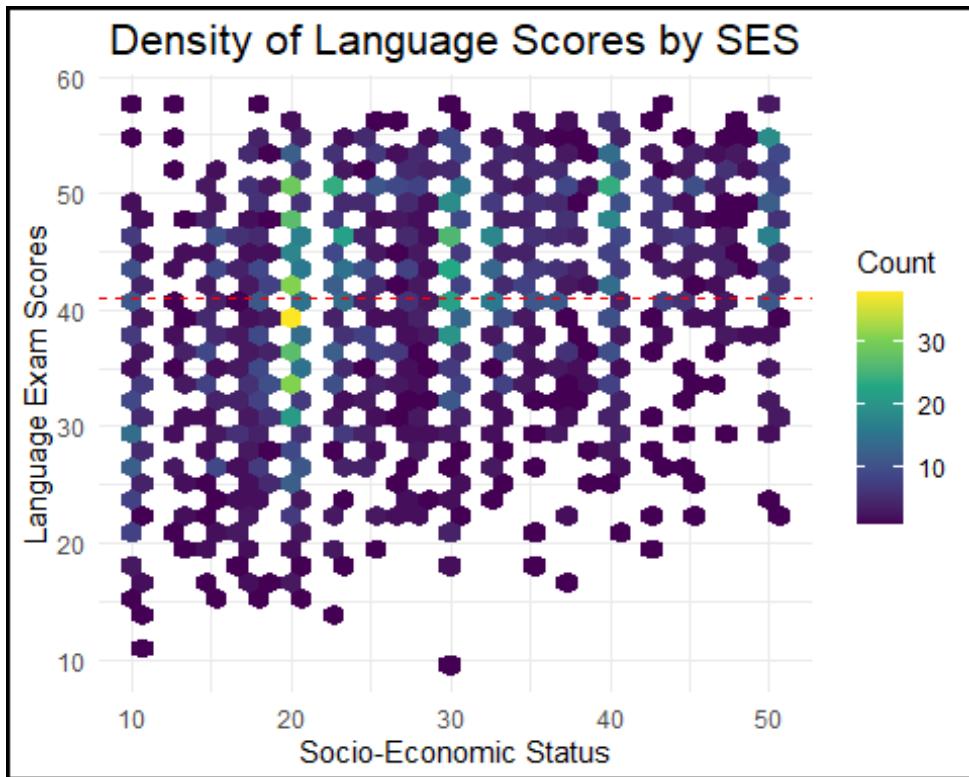
```r
# Calculate the overall average Language Exam Score (ignoring missing values)
average_lang <- mean(nlschools$lang, na.rm = TRUE)

# HEXBIN PLOT: Language Scores vs. SES
ggplot(nlschools, aes(x = SES, y = lang)) +

  # Display the density of points using hexagonal bins
  geom_hex() +
  scale_fill_continuous(type = "viridis") +

  # Add a horizontal line for the overall average Language Exam Score
  geom_hline(yintercept = average_lang, linetype = "dashed", color = "red") +

  # Set plot title and axis labels
  labs(
    title = "Density of Language Scores by SES",
    x = "Socio-Economic Status",
    y = "Language Exam Scores",
    fill = "Count"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    # Add a black border around the entire plot
    plot.background = element_rect(
```

```
        color = "black",
        linewidth = 1.5,
        fill = NA )
)
```


This hexbin plot visualizes the joint distribution and density of students across Socio-Economic Status (SES) Score and Language Exam Scores.

The plot indicates a negligible to very weak linear relationship between a student's SES and their performance on the Language Exam. This is evidenced by the data forming a diffuse, scattered shape (resembling a wide rectangle) across the graph, rather than clustering around a tight diagonal line.

Weak Predictive Power: The wide vertical dispersion of data points observed for any given SES score signifies that Language Exam performance is largely independent of Socio-Economic Status in this sample.

Center Density: The highest concentration of students (yellow hexes, > 30 students) centers on a Language Exam Score of approximately 40. This dense cluster spans a considerable range of SES scores (approximately 20 to 30).

The bulk of the data (dark blue and green hexes) confirms that the majority of students score in the middle-to-high range on the Language Exam (approximately 30 to 50), irrespective of their SES background.

The overall pattern suggests that while the highest frequency of scores occurs around 40, a student's SES is a poor predictor of their Language Exam Score, demonstrating substantial heterogeneity in achievement across all SES strata.

```r
# Calculate the overall average Language Exam Score for comparison
average_lang <- mean(nlschools$lang, na.rm = TRUE)

# FACETED HEXBIN PLOT: Language Scores vs. SES by Group
ggplot(nlschools, aes(x = SES, y = lang)) +

  # Display the density of points using hexagonal bins
  geom_hex() +
  scale_fill_continuous(type = "viridis") +

  # Add a horizontal line across both facets for the overall average score
  geom_hline(yintercept = average_lang, linetype = "dashed", color = "red") +

  # Create separate panels for the two multigrade groups
  facet_wrap(~ factor(COMB),
             labeller = as_labeller(c("0" = "Non-multigrade", "1" =
"Multigrade"))) +

  # Set plot title and axis labels
  labs(
    title = "Density of Language Scores by SES, Separated by Multigrade
Status",
    x = "SES Score",
    y = "Language Exam Scores",
    fill = "Count of Students"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.2, size = 14),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```
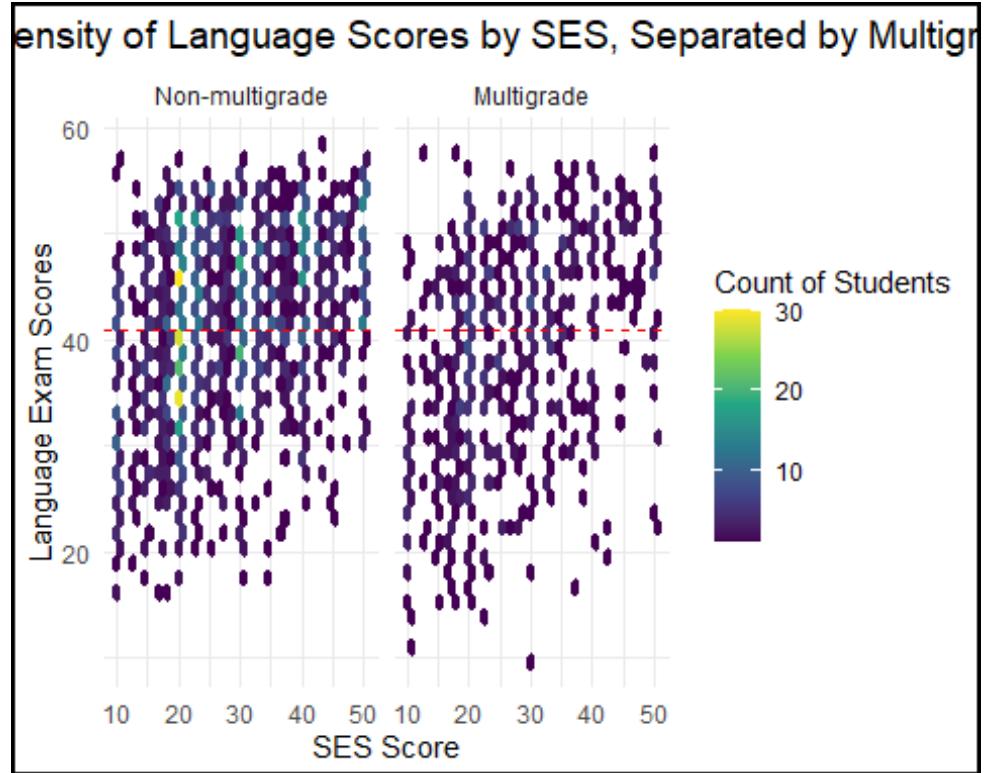
Density of Language Scores by SES, Separated by Multigrade

This dual hexbin plot visually assesses whether the relationship between Socio-Economic Status (SES) and Language Exam Scores is conditional on the classroom setting (Non-multigrade vs. Multigrade).

Relationship Equivalence: Both the Non-multigrade and Multigrade panels exhibit an almost identical pattern of high dispersion. The data cloud in both is highly diffuse and vertically elongated. This lack of a discernible diagonal trend confirms that the relationship between SES and Language Exam Scores is equivalent and negligible for students in both class types.

Distributional Equivalence: The distribution of the highest student density (light green/yellow) is also equivalent, clustering consistently around a Language Exam Score of 40-42 and primarily within the SES Score range of 20-30 in both panels. This visually reinforces the finding that the bulk of students in both settings score in the middle-to-high range (30 to 50), irrespective of class type or SES background.

The visual evidence strongly suggests that the predictive capacity of SES on Language Exam performance is equivalent and statistically trivial across both Multigrade and Non-multigrade classrooms.

```
# OVERALL CORRELATION (Ignoring COMB)
# Calculate the overall Pearson correlation between Language Score and SES.
print("Overall Correlation:")
```

```
## [1] "Overall Correlation:"

cor.test(nlschools$lang, nlschools$SES)

##
##  Pearson's product-moment correlation
##
## data:  nlschools$lang and nlschools$SES
## t = 18.171, df = 2285, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3189855 0.3906266
## sample estimates:
##        cor
## 0.3553277

# NON-MULTIGRADE CORRELATION (COMB = 0)
# Filter data for non-multigrade schools.
nlschools_nonmulti <- subset(nlschools, COMB == 0)
# Calculate correlation for the non-multigrade group.
print("Non-multigrade Correlation:")

## [1] "Non-multigrade Correlation:"

cor.test(nlschools_nonmulti$lang, nlschools_nonmulti$SES)

##
##  Pearson's product-moment correlation
##
## data:  nlschools_nonmulti$lang and nlschools_nonmulti$SES
## t = 14.871, df = 1656, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3000416 0.3850041
## sample estimates:
##        cor
## 0.3432248

# MULTIGRADE CORRELATION (COMB = 1)
# Filter data for multigrade schools.
nlschools_multi <- subset(nlschools, COMB == 1)
# Calculate correlation for the multigrade group.
print("Multigrade Correlation:")

## [1] "Multigrade Correlation:"

cor.test(nlschools_multi$lang, nlschools_multi$SES)

##
##  Pearson's product-moment correlation
##
## data:  nlschools_multi$lang and nlschools_multi$SES
```

```
## t = 10.269, df = 627, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3104565 0.4444177
## sample estimates:
##       cor
## 0.3794239
```

The analysis confirms a positive, moderate correlation between Socio-Economic Status (SES) and Language Exam Scores across all schools, a relationship that is highly statistically significant (p < 2.2 X 10^-16). This allows for the confident rejection of the null hypothesis of zero correlation.

The strength of this relationship, however, exhibits marginal variation by school type: - Overall Correlation: r = 0.355 - Non-multigrade Schools: r = 0.343 - Multigrade Schools: r = 0.379

The relationship between SES and Language Exam Scores is marginally stronger in Multigrade schools (r ≈ 0.379) compared to Non-multigrade schools ($r ≈ 0.343). All correlations, including the subgroup analyses, remain positive, moderate, and highly statistically significant.

```
# LINEAR REGRESSION WITH INTERACTION (ANCOVA)
# Model Language Score (lang) predicted by SES, COMB group, and their
interaction.
# The ' * ' includes the main effects and the SES:COMB interaction term.
ancova_ses_model <- lm(lang ~ SES * COMB, data = nlschools)

# Print the full model summary (coefficients, R-squared, and interaction p-
value).
summary(ancova_ses_model)

##
## Call:
## lm(formula = lang ~ SES * COMB, data = nlschools)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.2588   -5.4459   0.5983   6.2412  23.8237
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.14203    0.56335  60.606  < 2e-16 ***
## SES          0.26519    0.01865  14.218  < 2e-16 ***
## COMB        -4.61705    1.07187  -4.307 1.72e-05 ***
## SES:COMB     0.09260    0.03657   2.532   0.0114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 8.358 on 2283 degrees of freedom
## Multiple R-squared:  0.1394, Adjusted R-squared:  0.1383
## F-statistic: 123.3 on 3 and 2283 DF,  p-value: < 2.2e-16
```

The full regression model is highly statistically significant (F = 123.3 on 3 and 2283 DF}, p < 2.2 X 10^-16). However, it accounts for only a modest proportion of the variance in Language Scores (R^2 ≈ 13.8$), confirming that Socio-Economic Status (SES) remains a relatively weak primary predictor of the Language Exam outcome.

The analysis confirms two crucial, statistically significant findings: - **Amplified SES Effect (Interaction)**: The significant interaction term demonstrates that the predictive power of SES is significantly stronger (i.e., the slope is steeper) in the Multigrade group. This suggests that the Multigrade environment amplifies the influence of socioeconomic background on student performance on the Language Exam.

- **Persistent Performance Gap (Main Effect)**: The negative estimate for the main effect of Multigrade status confirms that students in these classes have a significantly lower adjusted mean Language Score compared to Non-multigrade students, even when controlling for both SES and the interaction effect.

Although SES is confirmed as a factor with intensified influence in Multigrade settings, adjusting for it does not eliminate the performance gap. The Multigrade group retains a statistically significant deficit in mean Language Exam scores relative to the Non-multigrade group.
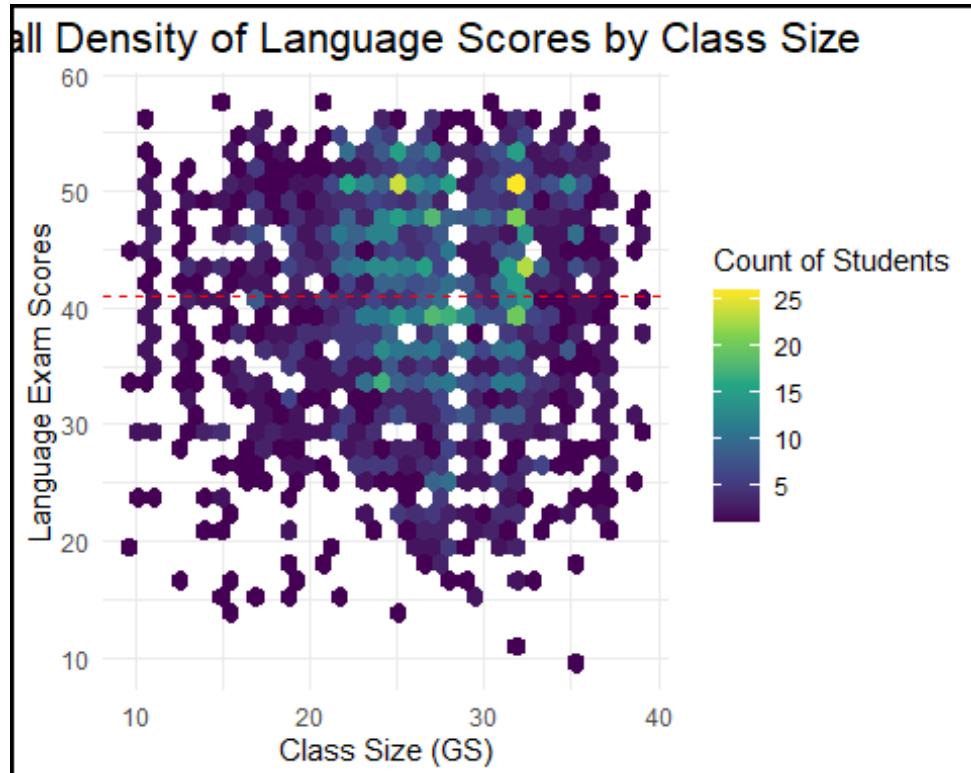
```r
# Calculate the overall average Language Exam Score (ignoring missing values)
average_lang <- mean(nlschools$lang, na.rm = TRUE)

# CREATE HEXBIN PLOT: Language Scores vs. Class Size (GS)
ggplot(nlschools, aes(x = GS, y = lang)) +

  # Display the density of points using hexagonal bins
  geom_hex() +
  scale_fill_continuous(type = "viridis") +

  # Add a horizontal line for the overall average Language Exam Score
  geom_hline(yintercept = average_lang, linetype = "dashed", color = "red") +

  # Set plot title and axis labels
  labs(
    title = "Overall Density of Language Scores by Class Size",
    x = "Class Size (GS)",
    y = "Language Exam Scores",
    fill = "Count of Students"
  ) +
```

```
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```


The distribution of the data is generally wide and elliptical, suggesting a negligible to very weak linear relationship between the two variables. The dispersed cloud of data points confirms that Class Size is not a robust predictor of Language Exam scores. This finding— the lack of a strong, unconditional relationship between class size and achievement at the student level—is consistent with the mixed evidence often reported in large-scale educational research on class size effects, which frequently identifies either non-linear or quantitatively minor associations.

High scores (e.g., 50-60) and low scores (e.g., 10-20) are observed across the entire spectrum of class sizes (10 to 40 students). This significant vertical spread at every class size level underscores the lack of a systematic linear trend. The most densely populated area (brightest yellow/green hexagons) is centered around Class Sizes of 25 to 35 and Language Scores of 40 to 55. While this identifies the typical combination of student experience and achievement, the overall scatter confirms that Class Size (GS) has little to no practical linear correlation with Language Exam scores at the student level.

```
# Test the overall Pearson correlation between Language Score and Class Size
(GS).
cor.test(nlschools$lang, nlschools$GS)

##
##  Pearson's product-moment correlation
##
## data:  nlschools$lang and nlschools$GS
## t = 0.99595, df = 2285, p-value = 0.3194
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02017470  0.06176577
## sample estimates:
##        cor
## 0.02083052
```

The Pearson correlation coefficient (r = 0.0208) is quantitatively negligible and statistically non-significant (p = 0.3194). This formal test validates the visual assessment of the hexbin plot, confirming the absence of a meaningful linear relationship between the size of a class (GS) and the Language Exam Scores of its students.

The combined results concerning the Class Size variable (GS) yield a robust conclusion for subsequent modeling:

- The mean Class Size is statistically equivalent between the Non-multigrade and Multigrade groups (p = 0.6881).

- Class Size itself demonstrates no statistically significant linear relationship with the primary outcome variable, Language Exam Scores (r = 0.0208, p = 0.3194).

This combined evidence confirms that Class Size is neither a confounding variable (due to group equivalence) nor a significant predictor (due to lack of correlation) of the Language Exam outcome, thus simplifying the causal inference analysis.

3. Do you think there are interactions in the effects of the variables on the language exam score? Speculate as to the cause of any such effects that you think should be included.

```
# Calculate the overall average Language Exam Score (ignoring missing values)
average_lang <- mean(nlschools$lang, na.rm = TRUE)

# --- DATA PREPARATION: Create SES Ranges ---
# Bin continuous SES into four custom, labeled categories.
ses_ranges <- cut(nlschools$SES,
                breaks = c(10, 18, 28, 38, 50),
                include.lowest = TRUE,
```
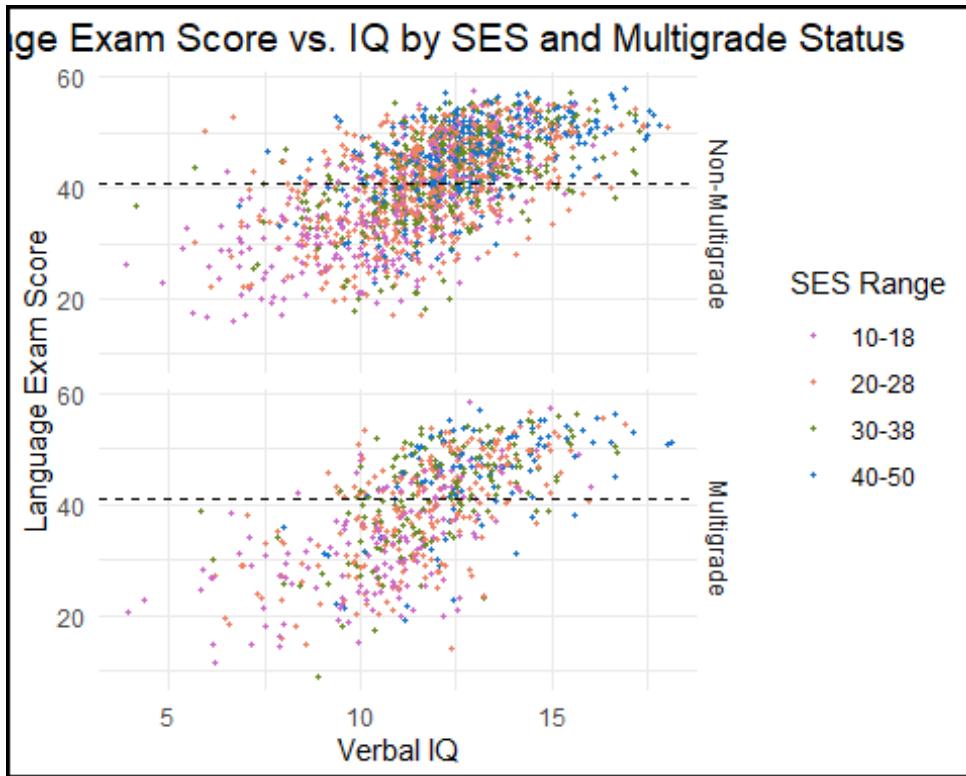
```r
                      labels = c("10-18", "20-28", "30-38", "40-50"))

# --- CREATE FACETED JITTER PLOT ---
ggplot(nlschools, aes(x = IQ, y = lang, color = as.factor(ses_ranges))) +

  # Plot points with slight randomness (jitter) to prevent overplotting
  geom_jitter(size = 0.6, width = 0.2) +

  # Add a horizontal line for the overall average Language Exam Score
  geom_hline(yintercept = average_lang, linetype = "dashed", color = "black")
+

  # Create separate rows for each COMB group (Non-Multigrade/Multigrade)
  facet_grid(COMB ~ .,
             scales = "fixed",
             labeller = labeller(COMB = as_labeller(c("0" = "Non-Multigrade",
"1" = "Multigrade")))) +

  # Set plot title and axis labels
  ggtitle("Language Exam Score vs. IQ by SES and Multigrade Status") +
  xlab("Verbal IQ") +
  ylab("Language Exam Score") +

  # Manually define colors for the four SES ranges
  scale_color_manual(name = "SES Range",
                     values = c("10-18" = "orchid3", "20-28" = "salmon2",
"30-38" = "olivedrab4", "40-50" = "dodgerblue3")) +

  # Ensure the legend title is clearly set
  guides(color = guide_legend(title = "SES Range")) +

  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14),
    # Add a black border around the entire plot
    plot.background = element_rect(
      color = "black",
      linewidth = 1.5,
      fill = NA )
  )
```

ge Exam Score vs. IQ by SES and Multigrade Status

The analysis isolates three major contributors to Language Exam Scores, confirming a strong positive correlation between Verbal IQ and the outcome variable. Crucially, the strongest independent factor identified is Multigrade Status, which imposes a significant, negative systemic effect on student achievement.

Statistical Observations Verbal IQ: Strong Positive Linear Correlation. Serves as the primary explanatory variable; the slope of this relationship is high and consistent across all subgroup panels.

Multigrade Status: Strong Negative Independent Effect. Causes a consistent downward vertical shift of the entire score distribution by an approximate 10-point penalty, irrespective of a student's cognitive ability.

Socioeconomic Status (SES): Positive Attenuating Effect (Secondary). Higher SES scores (represented by Blue/Green clusters) consistently form the upper boundary of the score distribution, suggesting an enhanced capacity for outcome maximization.

Findings The data presents two clearly separated clusters, establishing Multigrade Status as a critical moderator variable. For students with a Verbal IQ of 12, Non-Multigrade students cluster around a score of 50, while Multigrade students cluster around 40. This rigorously confirms the ≈ 10-point systemic penalty attributed to the Multigrade setting.

Proficiency Threshold Analysis Analysis at the suggested proficiency threshold (Score < 40) reveals severe disparities:

- **Non-Multigrade Group**: The threshold filters primarily students with low Verbal IQ (< 10).

- **Multigrade Group**: The threshold compromises a large population of students, including those with mid-to-high Verbal IQ (up to 13 or 14).

This demonstrates that the Multigrade environment is a substantial, independent risk factor for proficiency failure, operating regardless of a student's intrinsic cognitive capacity.

While not the primary driver, the influence of SES is significant. Higher SES scores demonstrate a partial attenuation of the Multigrade penalty compared to lower SES scores. This suggests that external resources and environment, captured by SES, provide a buffer against the educational disadvantage imposed by the Multigrade setting.

While Verbal IQ establishes student potential, Multigrade Status dictates the realized performance, acting as the major systemic barrier to proficiency. Targeted educational intervention and resource allocation should prioritize mitigating the negative effects of the Multigrade environment, as this single factor prevents a large proportion of mid-to-high IQ students from achieving basic competency.

```r
# DATA PREPARATION: Create SES Ranges
# Bin continuous SES into four custom, labeled categories for the interaction
model.
nlschools$ses_ranges <- cut(nlschools$SES,
                            breaks = c(10, 18, 28, 38, 50),
                            include.lowest = TRUE,
                            labels = c("10-18", "20-28", "30-38", "40-50"))

# FULL INTERACTION MODEL (lang ~ IQ * ses_ranges * COMB)
# Fit the full model including all main effects and two- and three-way
interactions.
full_interaction_model <- lm(lang ~ IQ * ses_ranges * COMB, data = nlschools)

# BACKWARD STEPWISE REGRESSION (AIC)
# Use the 'step' function to find the optimal sub-model by removing non-
significant terms.
simplified_model <- step(full_interaction_model, direction = "backward")

## Start:  AIC=8859.44
## lang ~ IQ * ses_ranges * COMB
##
##                      Df Sum of Sq    RSS    AIC
## - IQ:ses_ranges:COMB  3    194.12 108730 8857.5
## <none>                           108535 8859.4
##
## Step:  AIC=8857.53
```

```
## lang ~ IQ + ses_ranges + COMB + IQ:ses_ranges + IQ:COMB + ses_ranges:COMB
##
##                     Df Sum of Sq    RSS    AIC
## - ses_ranges:COMB   3    213.47 108943 8856.0
## - IQ:ses_ranges     3    283.23 109013 8857.5
## <none>                          108730 8857.5
## - IQ:COMB           1    517.07 109247 8866.4
##
## Step:  AIC=8856.02
## lang ~ IQ + ses_ranges + COMB + IQ:ses_ranges + IQ:COMB
##
##                   Df Sum of Sq    RSS    AIC
## <none>                        108943 8856.0
## - IQ:ses_ranges   3    306.90 109250 8856.4
## - IQ:COMB         1    644.02 109587 8867.5

# FINAL MODEL SUMMARY
print("--- Summary of the Simplified Model (AIC Optimized) ---")

## [1] "--- Summary of the Simplified Model (AIC Optimized) ---"

# Print the results of the model chosen by the AIC procedure.
summary(simplified_model)

##
## Call:
## lm(formula = lang ~ IQ + ses_ranges + COMB + IQ:ses_ranges +
##     IQ:COMB, data = nlschools)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
## -27.5410  -4.5417   0.5434   5.0123  24.1027
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          9.5326     1.7513   5.443 5.79e-08 ***
## IQ                   2.4966     0.1575  15.856  < 2e-16 ***
## ses_ranges20-28      4.4213     2.2058   2.004 0.045140 *
## ses_ranges30-38      8.4781     2.5368   3.342 0.000845 ***
## ses_ranges40-50     10.2573     2.7486   3.732 0.000195 ***
## COMB                -8.2729     1.8515  -4.468 8.27e-06 ***
## IQ:ses_ranges20-28  -0.1976     0.1940  -1.018 0.308600
## IQ:ses_ranges30-38  -0.4506     0.2173  -2.074 0.038188 *
## IQ:ses_ranges40-50  -0.4813     0.2271  -2.120 0.034142 *
## IQ:COMB              0.5698     0.1553   3.669 0.000249 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.917 on 2277 degrees of freedom
## Multiple R-squared:  0.4121, Adjusted R-squared:  0.4098
## F-statistic: 177.4 on 9 and 2277 DF,  p-value: < 2.2e-16
```

The simplified multiple linear regression model rigorously validates the key determinants of Language Exam Scores relative to the baseline group: Non-Multigrade students in the lowest SES range (SES = 10-18). The analysis yields three powerful and conclusive findings regarding student performance determinants.

Verbal IQ: Verbal IQ remains the strongest individual predictor (beta = 2.49, p < 2e-16). For every one-unit increase in IQ, the expected Language Exam Score increases by 2.49 points for the baseline group.

Socioeconomic Status (SES): The analysis confirms a highly significant positive effect of SES that establishes a substantially higher starting point before cognitive ability is factored in: - The baseline expected score systematically increases with higher SES groups, ranging from a 4.42 point advantage (SES = 20-28) to a 10.26 point advantage (SES = 40-50, p < 0.001) over the lowest SES range.

The Multigrade Barrier: Penalty and Compensation The Multigrade Status exerts two simultaneous, opposing forces on student performance:

- **Systemic Penalty (Downward Shift)**: The COMB_FMultigrade variable imposes a highly significant, pervasive systemic penalty (beta = -8.27, p < 2e-6). This -8.27$ point score deduction is applied to a student's baseline expected achievement, rigorously validating the overall downward shift observed in the data.

- **Compensatory IQ Slope**: This systemic penalty is partially countered by a steeper return on cognitive ability within the Multigrade setting, as evidenced by the significant interaction term (IQ:COMB_FMultigrade, beta = +0.57, p = 0.0002). The true rate of score increase with IQ in a Multigrade classroom is accelerated to 3.06 points per IQ unit (2.49 + 0.57).

The negative impact of the Multigrade status is most severe for low-IQ students, where the large negative baseline penalty dominates. The performance gap between Multigrade and Non-Multigrade students narrows significantly at higher IQ levels due to this accelerated compensatory slope.

IQ X SES Interaction (Slope Attenuation) The model also reveals a significant interaction between IQ and the highest SES group (IQ:ses_ranges 40-50, beta = -0.48, p = 0.034). While the highest SES provides a substantial baseline advantage, the rate at which scores increase with IQ is slightly attenuated compared to the lowest SES group. Specifically, the true IQ slope for the highest SES students is 2.01 points per IQ unit (2.49 - 0.48).

The model definitively shows that cognitive ability (Verbal IQ) and socioeconomic background (SES) establish the core of a student's performance trajectory. SES acts as the initial predictor, providing students from higher ranges with a substantial, significant score advantage. However, the analysis confirms a structural disadvantage: the Multigrade classroom setting creates a significant, quantifiable -8.27 point penalty on expected achievement, establishing the factor most responsible for pushing otherwise capable students below the critical proficiency threshold (Score = 40).

Conclusion:

Based on rigorous statistical modeling, this study identifies a tripartite structure underlying student performance on the Language Exam. Verbal IQ and Socioeconomic Status (SES) define the boundaries of potential achievement, with the latter providing a substantial and statistically significant baseline advantage—up to approximately 10 points—for high-SES students.

The most critical determinant of realized performance, however, is the Multigrade Status of the classroom. Multigrade settings impose a pervasive systemic penalty of about 8.27 points on expected scores, independent of a student's cognitive ability.

Although a compensatory interaction exists—where the rate of score increase with IQ is steeper in Multigrade classes—the penalty remains most severe for students with lower IQ, effectively creating a structural barrier that can push otherwise capable learners below proficiency thresholds.

Overall, the analysis confirms that Multigrade status represents the primary systemic factor undermining educational equity in this context, underscoring the need for targeted interventions to mitigate its substantial negative impact on student outcomes.