library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readxl)  
  
Stress\_and\_Coping\_Strategies\_of\_Nursing\_Students\_September\_26\_2024\_12\_12 <- read\_csv("Stress and Coping Strategies of Nursing Students\_September 26, 2024\_12.12.csv")

## Rows: 88 Columns: 71  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (71): StartDate, EndDate, Status, IPAddress, Progress, Duration (in seco...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

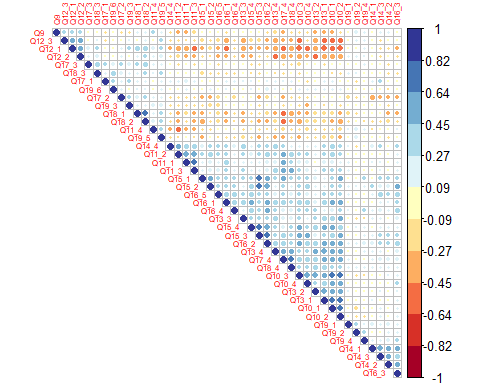
Stress\_and\_Coping1 <-  
 Stress\_and\_Coping\_Strategies\_of\_Nursing\_Students\_September\_26\_2024\_12\_12 %>% select(Q9:Q19\_6)  
  
Stress\_and\_Coping2 <- na.omit(Stress\_and\_Coping1)  
  
Stress\_and\_Coping2 <-  
 Stress\_and\_Coping2[3:85,]  
  
Stress\_and\_Coping2 <-   
 Stress\_and\_Coping2[1:72,]  
  
  
  
  
for (i in seq\_len(ncol(Stress\_and\_Coping2))){  
 for (j in seq\_len(nrow(Stress\_and\_Coping2))){  
 if (Stress\_and\_Coping2[j,i] == '1st' || Stress\_and\_Coping2[j,i] == 'Never'){  
 Stress\_and\_Coping2[j,i] <- '1'  
 }else if (Stress\_and\_Coping2[j,i] == '2nd' || Stress\_and\_Coping2[j,i] == 'Rarely'){  
 Stress\_and\_Coping2[j,i] <- '2'  
 }else if (Stress\_and\_Coping2[j,i] == '3rd' || Stress\_and\_Coping2[j,i] == 'Sometimes'){  
 Stress\_and\_Coping2[j,i] <- '3'  
 }else if (Stress\_and\_Coping2[j,i] == '4th' || Stress\_and\_Coping2[j,i] == 'Frequently'){  
 Stress\_and\_Coping2[j,i] <- '4'  
 }else if (Stress\_and\_Coping2[j,i] == '5th' || Stress\_and\_Coping2[j,i] == 'Always'){  
 Stress\_and\_Coping2[j,i] <- '5'  
 }  
 }  
}  
  
  
# Convert all columns to numeric  
Stress\_and\_Coping2 <- Stress\_and\_Coping2 %>%  
 mutate(across(everything(), ~ as.numeric(as.character(.))))  
  
  
summary(lm(data = Stress\_and\_Coping2, formula = Q16\_1 ~ Q9 + Q10\_1 + Q10\_2 + Q10\_3 + Q10\_4))

##   
## Call:  
## lm(formula = Q16\_1 ~ Q9 + Q10\_1 + Q10\_2 + Q10\_3 + Q10\_4, data = Stress\_and\_Coping2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.48050 -0.41566 0.03823 0.43343 1.36239   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.86083 0.46013 1.871 0.0658 .  
## Q9 -0.03052 0.06908 -0.442 0.6601   
## Q10\_1 0.21307 0.15120 1.409 0.1635   
## Q10\_2 0.28787 0.12943 2.224 0.0296 \*  
## Q10\_3 -0.06554 0.15295 -0.428 0.6697   
## Q10\_4 0.18029 0.11134 1.619 0.1101   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7151 on 66 degrees of freedom  
## Multiple R-squared: 0.3749, Adjusted R-squared: 0.3276   
## F-statistic: 7.918 on 5 and 66 DF, p-value: 6.855e-06

library(corrplot)

## corrplot 0.94 loaded

## corrplot 0.92 loaded  
library(RColorBrewer)  
Corr\_matrix <- cor(Stress\_and\_Coping2, use = 'everything', method = 'pearson')  
corrplot(Corr\_matrix, type="upper", order="hclust",  
col=brewer.pal(n=11, name="RdYlBu"), tl.cex = .5)



Columns 10 - 15: stress  
column 9: semester  
column 16-19: coping

Never = 1  
Rarely = 2  
Sometimes = 3  
Frequently = 4  
Always = 5

For visualization

## Anova

Analysis of Variance (ANOVA) is a fundamental statistical method used to determine whether there are significant differences between the means of three or more independent (unrelated) groups. It helps in understanding if the observed variations in data are due to actual differences between groups or merely due to random chance.

State the Hypotheses:

Null Hypothesis (H₀): All group means are equal.  
Alternative Hypothesis (H₁): At least one group mean is different.

If different groups have different values (for our example, if the semester dictate how much stress they have or how they cope with it)

Anova on Stress Total

Stress\_and\_Coping2$Total\_Stress <- rowSums(Stress\_and\_Coping2[,2:24])  
Stress\_and\_Coping2$Q9\_Group <- as.factor(Stress\_and\_Coping2$Q9)   
  
Stress\_anova <- aov(formula = Total\_Stress ~ Q9\_Group, data = Stress\_and\_Coping2)  
  
summary(Stress\_anova)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Q9\_Group 4 981 245.19 2.788 0.0333 \*  
## Residuals 67 5893 87.95   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

TukeyHSD(Stress\_anova)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Total\_Stress ~ Q9\_Group, data = Stress\_and\_Coping2)  
##   
## $Q9\_Group  
## diff lwr upr p adj  
## 2-1 -3.1868132 -13.313329 6.939703 0.9024696  
## 3-1 4.6103896 -5.982721 15.203500 0.7397493  
## 4-1 -5.8214286 -14.427312 2.784455 0.3293004  
## 5-1 -0.2380952 -13.066989 12.590799 0.9999983  
## 3-2 7.7972028 -2.973684 18.568090 0.2634570  
## 4-2 -2.6346154 -11.458404 6.189173 0.9180853  
## 5-2 2.9487179 -10.027357 15.924793 0.9684376  
## 4-3 -10.4318182 -19.787400 -1.076236 0.0213268  
## 5-3 -4.8484848 -18.191880 8.494910 0.8458987  
## 5-4 5.5833333 -6.244322 17.410989 0.6777546

For Conclusion on Anova and Total Stress, since the F statistic is really low, it shows that the groups have different means. (Semester and Total Stress). The group value dictates the value of total stress. Additionally, reject the null hypothesis in favor of the alternative hypothesis that there is a difference in group means.

1- Mean  
2 - mean  
etc.

Anova on Coping Total

Stress\_and\_Coping2$Total\_Coping <- rowSums(Stress\_and\_Coping2[,25:43])  
Coping\_anova <- aov(formula = Total\_Coping ~ Q9\_Group, data = Stress\_and\_Coping2)  
  
summary(Coping\_anova)

## Df Sum Sq Mean Sq F value Pr(>F)  
## Q9\_Group 4 100.5 25.12 0.593 0.669  
## Residuals 67 2836.6 42.34

TukeyHSD(Coping\_anova)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = Total\_Coping ~ Q9\_Group, data = Stress\_and\_Coping2)  
##   
## $Q9\_Group  
## diff lwr upr p adj  
## 2-1 -2.2527473 -9.278648 4.773153 0.8963208  
## 3-1 -1.2597403 -8.609370 6.089889 0.9888894  
## 4-1 -2.6785714 -8.649439 3.292296 0.7176722  
## 5-1 0.4523810 -8.448462 9.353224 0.9999048  
## 3-2 0.9930070 -6.479966 8.465980 0.9958028  
## 4-2 -0.4258242 -6.547876 5.696228 0.9996689  
## 5-2 2.7051282 -6.297831 11.708088 0.9163366  
## 4-3 -1.4188312 -7.909848 5.072186 0.9725866  
## 5-3 1.7121212 -7.545689 10.969932 0.9852147  
## 5-4 3.1309524 -5.075220 11.337124 0.8213565

For Conclusion on Anova with Semester and Total COping, since the F statistic is not really low, it shows that the groups have similar means (Semester and Total Coping). The group value does not dictate the value of total stress. Additionally, do not reject the null hypothesis in favor of the alternative hypothesis that there is a difference in group means.

## Linear Regression

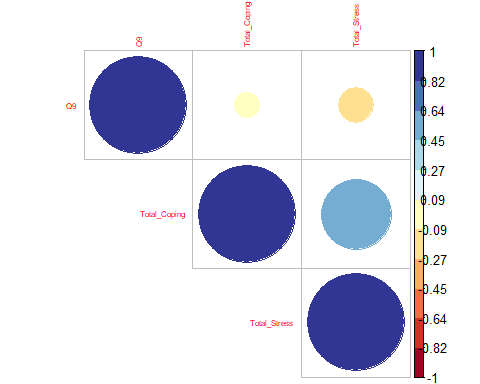
summary(lm(formula = Total\_Coping ~ Q9 + Total\_Stress, data = Stress\_and\_Coping2))

##   
## Call:  
## lm(formula = Total\_Coping ~ Q9 + Total\_Stress, data = Stress\_and\_Coping2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.2439 -3.6954 -0.1994 3.6911 16.5467   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 38.35790 4.98944 7.688 7.44e-11 \*\*\*  
## Q9 0.01165 0.51053 0.023 0.982   
## Total\_Stress 0.34206 0.06773 5.050 3.45e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.561 on 69 degrees of freedom  
## Multiple R-squared: 0.2735, Adjusted R-squared: 0.2524   
## F-statistic: 12.99 on 2 and 69 DF, p-value: 1.633e-05

Total\_Coping is the response variable in our linear regression problem and Q9(Semester) and Total\_Stress are independent variables. Essentially do the values of the independent dictate values for the response variable. Look at the p value being less than .05. The p value for Q9 is high so we do not reject the null hypothesis. P value for total stress was very low so we reject the null hypothesis in favor that total stress and total coping have a significant relationship.

Null Hypothesis: Slope for the independent variable is equal to zero.  
Alternative Hypothesis:  
Slope for the independent variable is not equal to zero

Stress\_and\_Coping4 <- Stress\_and\_Coping2 %>% select(Q9,Total\_Coping, Total\_Stress)  
  
Corr\_matrix2 <- cor(Stress\_and\_Coping4, use = 'everything', method = 'pearson')  
corrplot(Corr\_matrix2, type="upper", order="hclust",  
col=brewer.pal(n=11, name="RdYlBu"), tl.cex = .5)



summary1 <- group\_by(Stress\_and\_Coping2, Q9)  
summary1 <- summarise(summary1, mean(Total\_Stress), mean(Total\_Coping))  
print(summary1)

## # A tibble: 5 × 3  
## Q9 `mean(Total\_Stress)` `mean(Total\_Coping)`  
## <dbl> <dbl> <dbl>  
## 1 1 68.6 62.7  
## 2 2 65.4 60.5  
## 3 3 73.2 61.5  
## 4 4 62.8 60.0  
## 5 5 68.3 63.2