Take 2

2025-07-24

```
## Load the data

# all g scores
all_gScores <- read_csv("epimex_g_10june2025.csv")
```

```
## Rows: 2680 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): studyid
## dbl (1): g
##
## 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.
```

```
# all cog tests
all_cogTests <- read_csv("epimex_gorilla_10june2025.csv")
```

```
## Rows: 2680 Columns: 20
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): studyid
## dbl (19): matrixreasoning_trials, matrixreasoning_correct, cvlt_correct, cvl...
##
## 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.
```

```
# all phenotype data
all_phenotype <-read_csv("epimex_12feb2025_sciddx_corrected_LMS.csv")
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 2280 Columns: 104
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (16): family_id2, info_yearsofedcomments, recruitment_site_other, info_d...
## dbl (87): studyid, studyid_2, relation_id, family_id, proband_2, control, co...
## lgl  (1): cohab_sib
##
## 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.
```

```r
# probands and controls phenotype file
probands <- read.table("epimex_probands.fam", header = FALSE, sep = "", stringsAsFactors = FALSE)
colnames(probands) <- c("FamilyID", "studyid", "PaternalID", "MaternalID", "sex", "phenotype")
probands$group <- "Proband"

controls <- read.table("epimex_controls.fam", header = FALSE, sep = "", stringsAsFactors = FALSE)
colnames(controls) <- c("FamilyID", "studyid", "PaternalID", "MaternalID", "sex", "phenotype")
controls$group <- "Control"

## Filter data

# filter phenotype data down to ID and age
all_ages <- all_phenotype[, c("studyid", "age_2")]
all_ages <- all_ages %>%
  rename(
    age = age_2
  )

# filter proband and control data down to ID, sex, and phenotype
probands <- probands[, c("studyid", "sex", "phenotype", "group")]
controls <- controls[, c("studyid", "sex", "phenotype", "group")]

# combine proband and control data
combined_phenotype <- bind_rows(controls, probands)

# combine data with age
combined_phenotype <- combined_phenotype %>%
  left_join(all_ages, by = "studyid")

# filter g scores and cog data down to the probands and controls
filtered_gScores <- all_gScores[all_gScores$studyid %in% combined_phenotype$studyid, ]
filtered_cogTests <- all_cogTests[all_cogTests$studyid %in% combined_phenotype$studyid, ]
```

Clean cognitive tests data

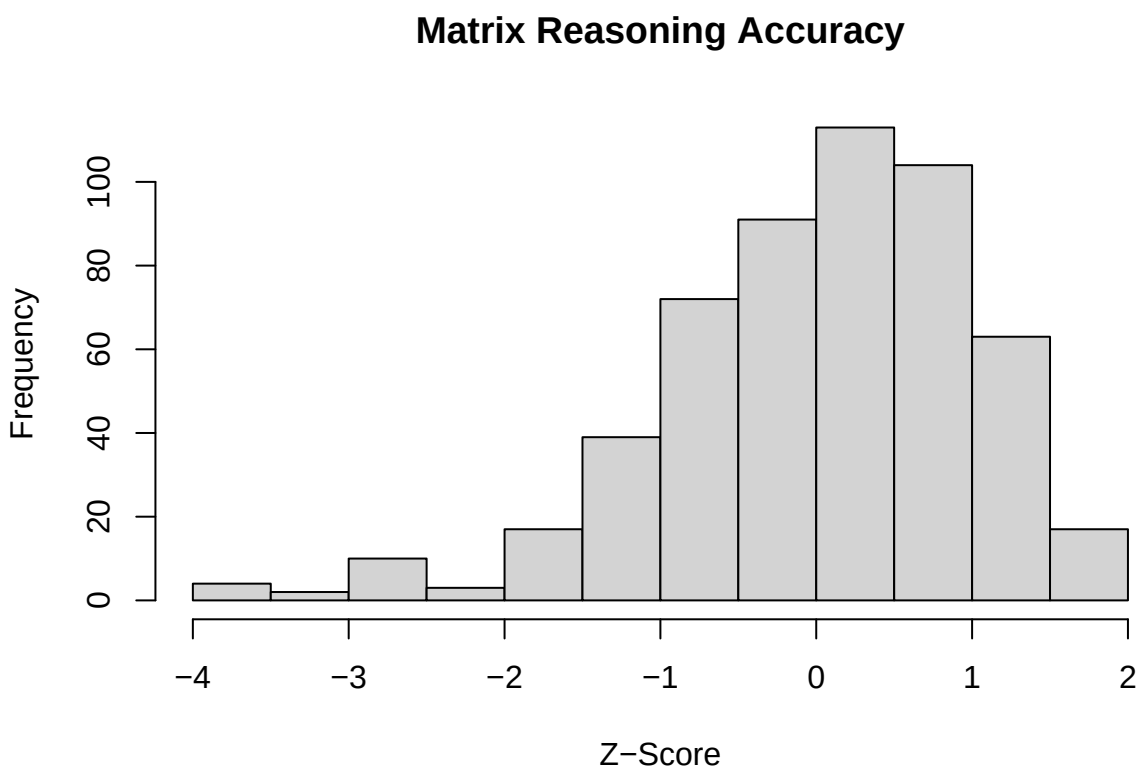
```r
## Matrix Reasoning Test

# calculate accuracy
filtered_cogTests$matrixreasoning_accuracy <- filtered_cogTests$matrixreasoning_correct / filtered_cogT

# standardize accuracy score (z-score)
filtered_cogTests$matrixreasoning_z <- scale(filtered_cogTests$matrixreasoning_accuracy)

# check work and visualize distribution
summary(filtered_cogTests$matrixreasoning_accuracy)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1667  0.5904  0.6970  0.6788  0.7812  0.9429
```

```r
summary(filtered_cogTests$matrixreasoning_z)
```

```
##        V1
##  Min.   :-3.8359
##  1st Qu.:-0.6622
##  Median : 0.1359
##  Mean   : 0.0000
##  3rd Qu.: 0.7671
##  Max.   : 1.9774
```

```r
hist(filtered_cogTests$matrixreasoning_z, main = "Matrix Reasoning Accuracy", xlab = "Z-Score")
```



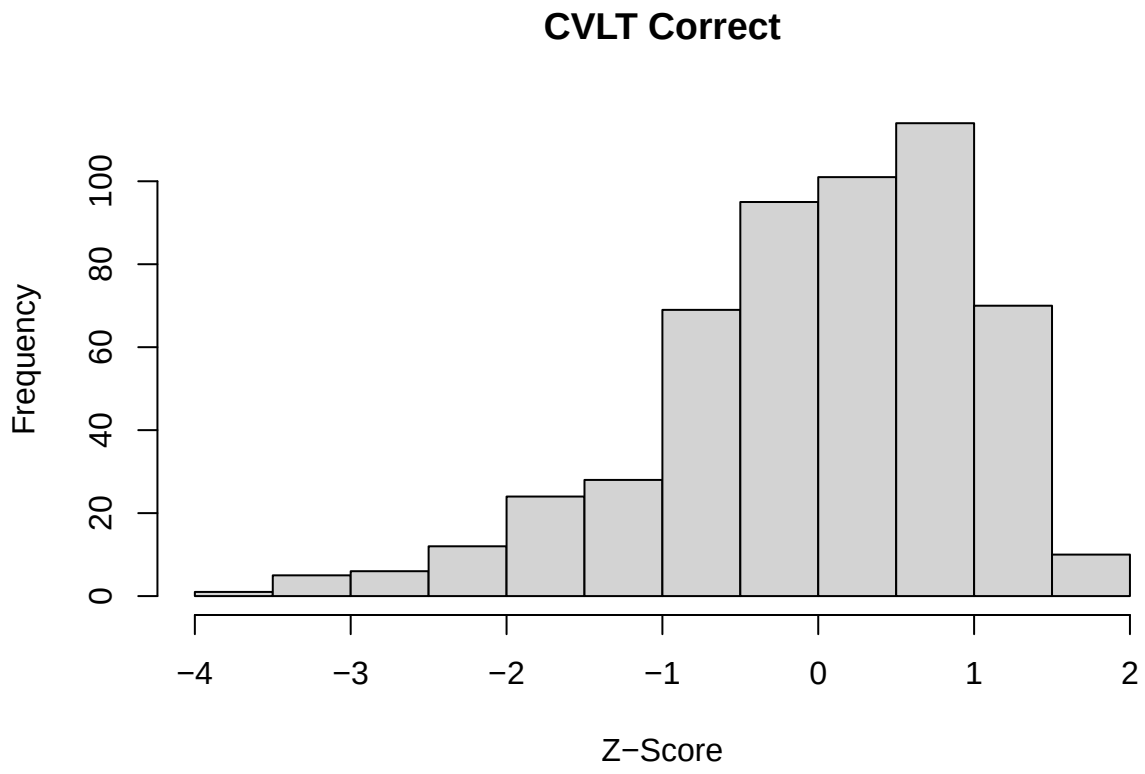

```r
## CVLT (Califronia Verbal Learning Test)
# cvlt_correct - total number of correctly recalled words
# cvlt_dprime - ability to distinguish targets from distractors

# standardize scores (z-score)
filtered_cogTests$cvlt_correct_z <- scale(filtered_cogTests$cvlt_correct)
filtered_cogTests$cvlt_dprime_z <- scale(filtered_cogTests$cvlt_dprime)

# check summaries and visualize distribution
summary(filtered_cogTests$cvlt_correct)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.00   48.00   56.00   54.85   64.00   78.00
```

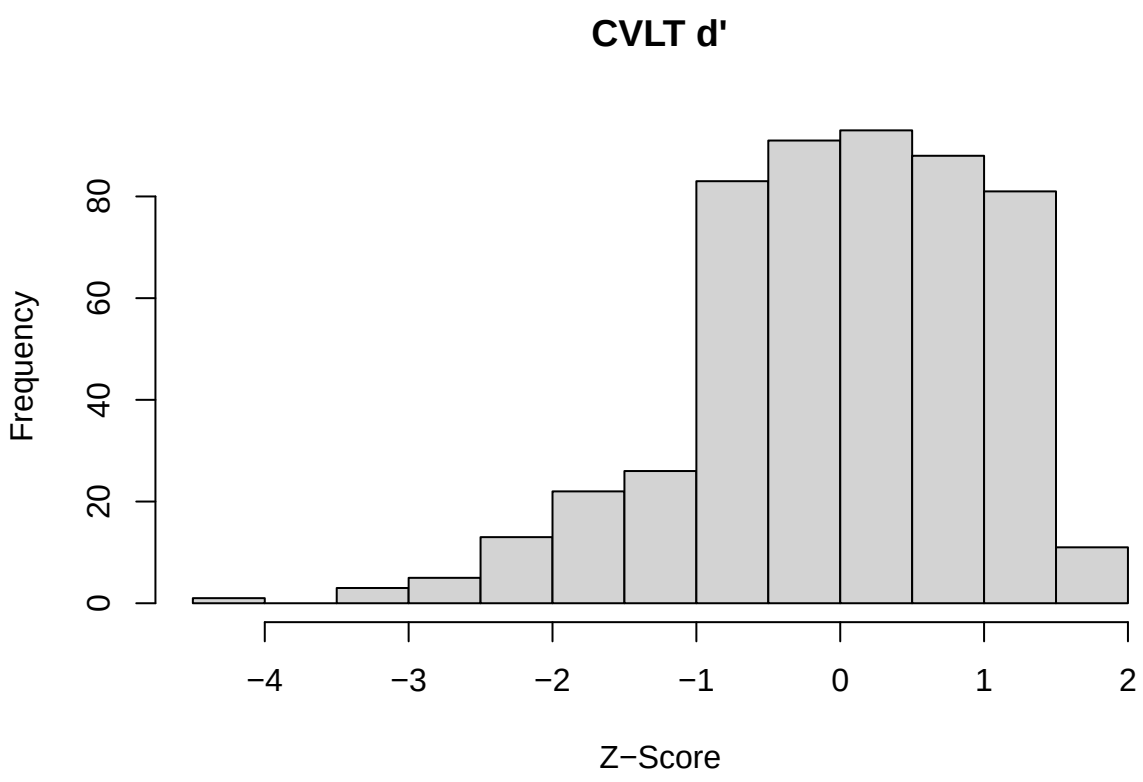
```r
summary(filtered_cogTests$cvlt_dprime)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.5431  2.1536  2.4585  2.4207  2.7702  3.1787      18
```

```r
hist(filtered_cogTests$cvlt_correct_z, main = "CVLT Correct", xlab = "Z-Score")
```




```r
hist(filtered_cogTests$cvlt_dprime_z, main = "CVLT d'", xlab = "Z-Score")
```



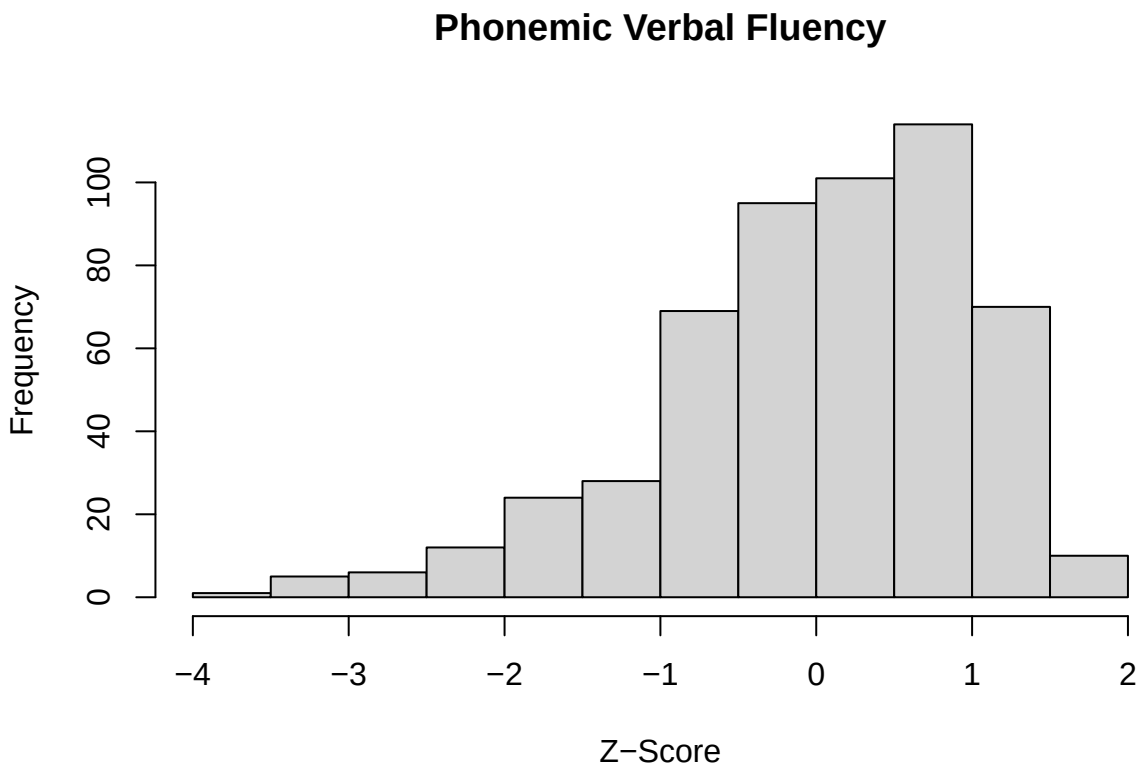

```
## Verbal Fluency
# pmr - phonemic fluency
# animal - semantic fluency

# check summaries
summary(filtered_cogTests$verbalfluency_es_pmr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    3.00   22.00   29.00   29.81   37.00   70.00       1
```

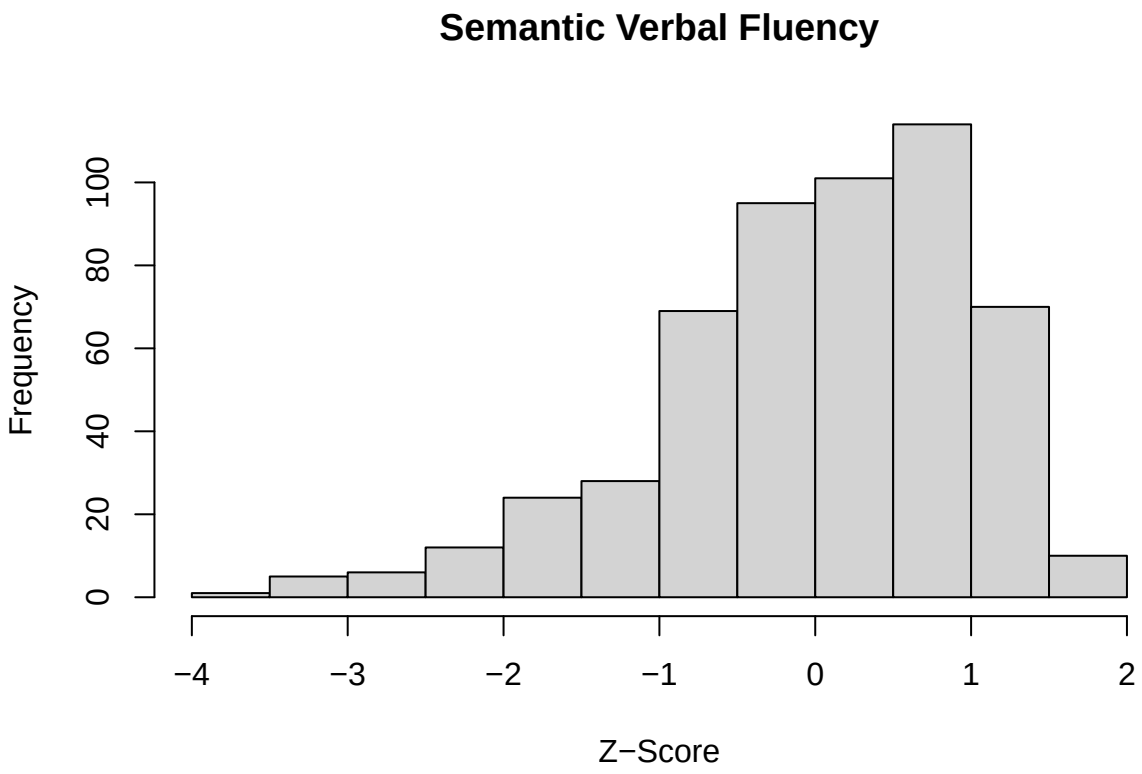
```
summary(filtered_cogTests$verbalfluency_es_animal)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   17.00   21.00   20.48   25.00   42.00       1
```

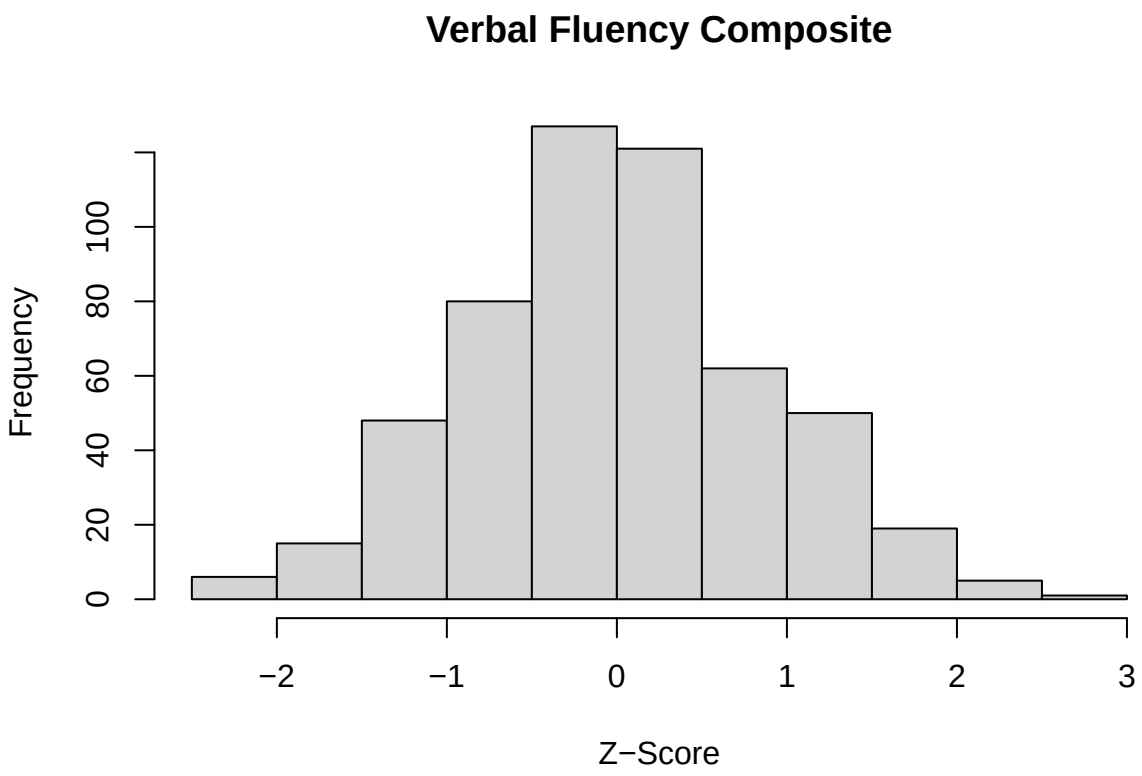
```
# standardize (z-score)
filtered_cogTests$verbalfluency_es_pmr_z <- scale(filtered_cogTests$verbalfluency_es_pmr)
filtered_cogTests$verbalfluency_es_animal_z <- scale(filtered_cogTests$verbalfluency_es_animal)

# visualize
hist(filtered_cogTests$cvlt_correct_z, main = "Phonemic Verbal Fluency", xlab = "Z-Score")
```




```
hist(filtered_cogTests$cvlt_correct_z, main = "Semantic Verbal Fluency", xlab = "Z-Score")
```



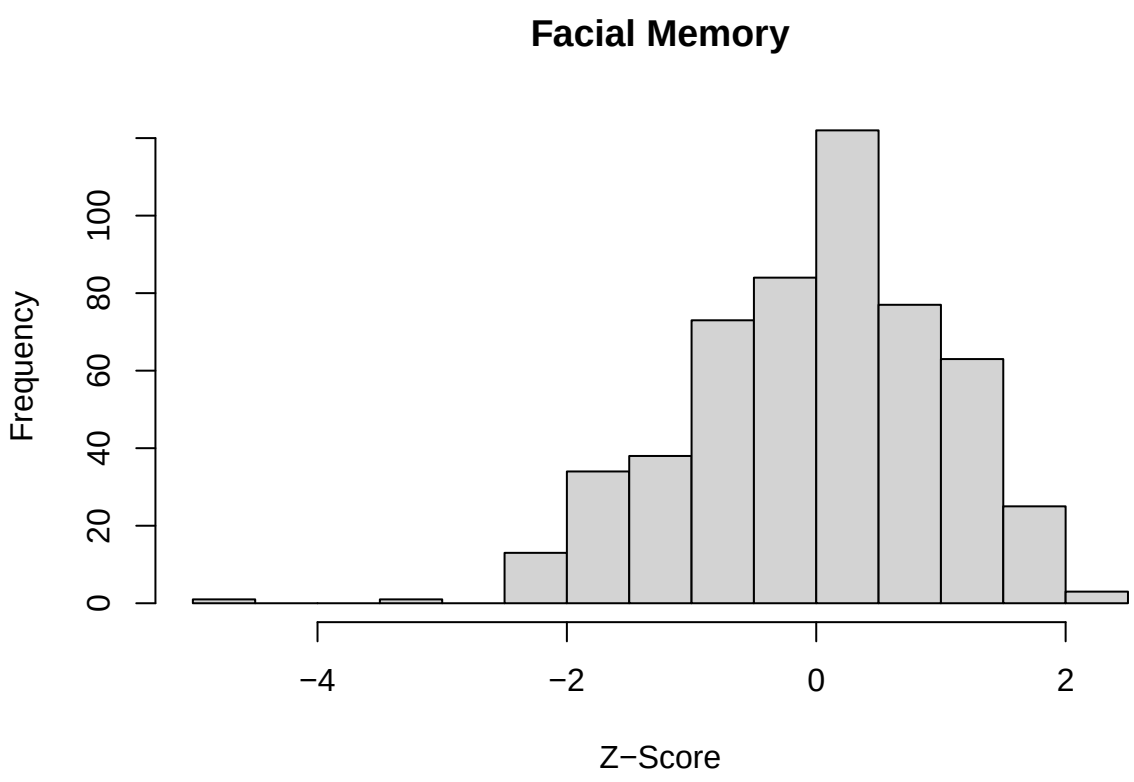

```r
# composite verbal fluency (takes avg per person across both columns)
filtered_cogTests$verbalfluency_composite <- rowMeans(
  filtered_cogTests[, c("verbalfluency_es_pmr_z", "verbalfluency_es_animal_z")],
  na.rm = TRUE
)
hist(filtered_cogTests$verbalfluency_composite, main = "Verbal Fluency Composite", xlab = "Z-Score")
```




```
## Facial Memory

# check summaries
summary(filtered_cogTests$facialmemory_correct)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   11.00   51.00   57.50   56.75   64.00   77.00       1
```

```
# standardize (z-score)
filtered_cogTests$facialmemory_z <- scale(filtered_cogTests$facialmemory_correct)

# visualize
hist(filtered_cogTests$facialmemory_z, main = "Facial Memory", xlab = "Z-Score")
```




```
## Verbal Working Memory
# forward_mns - simple span
# backward_mns - reversed span
# lns_mns - letter number sequencing

# summaries
summary(filtered_cogTests$verbalworkingmemory_forward_mns)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   4.000   4.500   4.578   5.000   7.500
```

```
summary(filtered_cogTests$verbalworkingmemory_backward_mns)
```

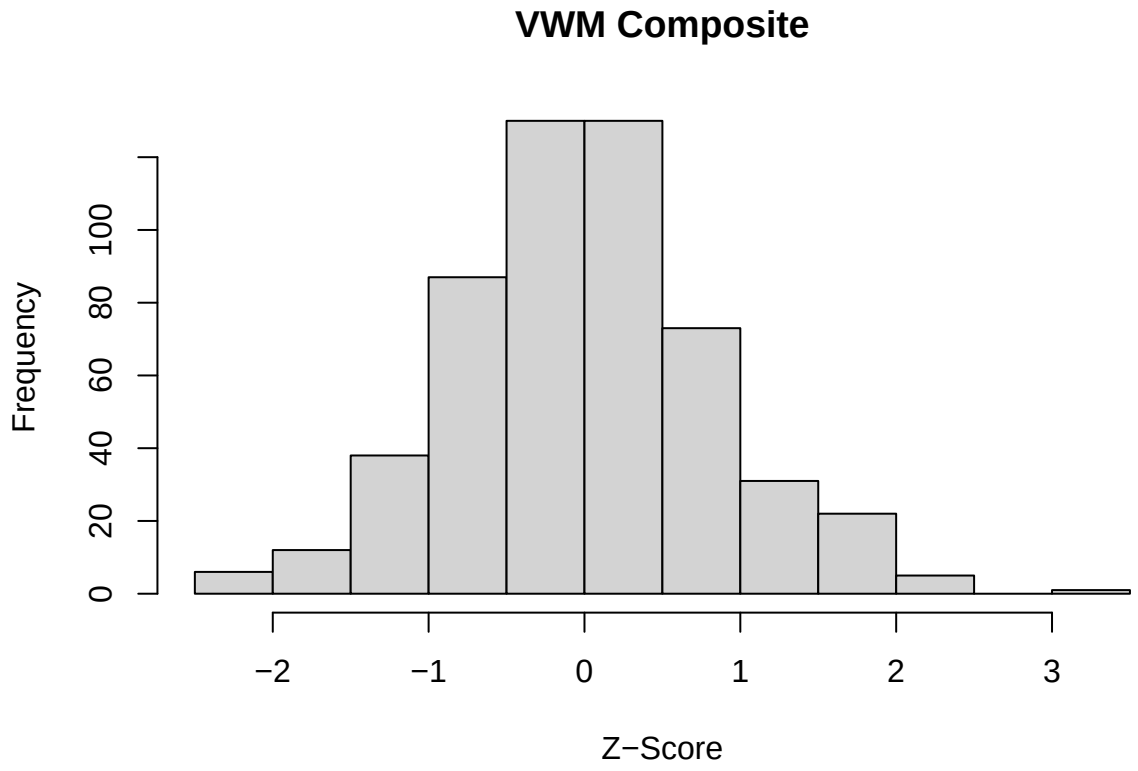
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    2.00    3.00    3.50    3.34    3.50    7.00      24
```

```
summary(filtered_cogTests$verbalworkingmemory_lns_mns)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   2.667   3.333   3.197   3.667   5.667       2
```

```
# scale
filtered_cogTests$vwm_forward_z <- scale(filtered_cogTests$verbalworkingmemory_forward_mns)
```

```r
filtered_cogTests$vwm_backward_z <- scale(filtered_cogTests$verbalworkingmemory_backward_mns)
filtered_cogTests$vwm_lns_z <- scale(filtered_cogTests$verbalworkingmemory_lns_mns)

# composite scores
filtered_cogTests$vwm_composite <- rowMeans(
  filtered_cogTests[, c("vwm_forward_z", "vwm_backward_z", "vwm_lns_z")],
  na.rm = TRUE
)

hist(filtered_cogTests$vwm_composite, main = "VWM Composite", xlab = "Z-Score")
```




```r
summary(filtered_cogTests$vwm_composite)
```

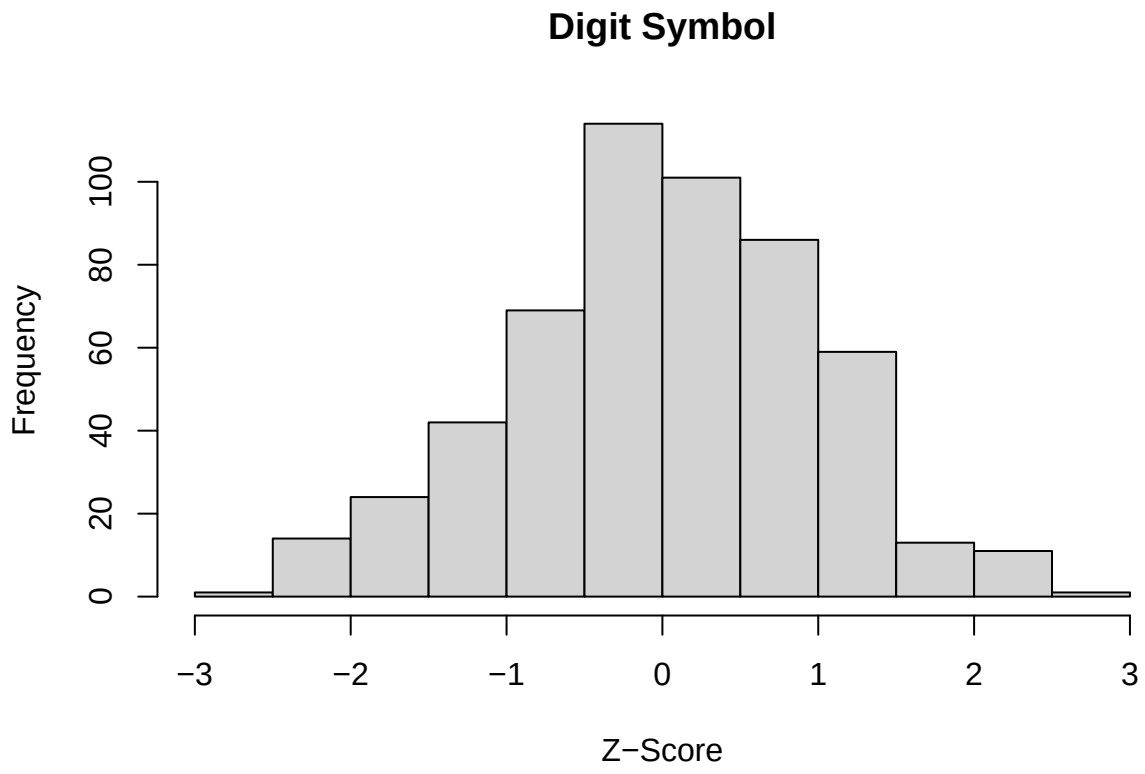
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.460126 -0.548849 -0.033826 -0.000664  0.495027  3.049111
```

```r
## Digit Symbols

# standardize
filtered_cogTests$digitsymbol1_score_z <- scale(filtered_cogTests$digitsymbol1_score)
filtered_cogTests$digitsymbol2_score_z <- scale(filtered_cogTests$digitsymbol2_score)
filtered_cogTests$digitsymbol_score_z <- scale(filtered_cogTests$digitsymbol_score)

# composite score
```

```r
filtered_cogTests$digitsymbol_composite <- rowMeans(
  filtered_cogTests[, c("digitsymbol1_score_z", "digitsymbol2_score_z", "digitsymbol_score_z")],
  na.rm = TRUE
)
summary(filtered_cogTests$digitsymbol_composite)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.53553 -0.62410  0.01985  0.00000  0.65322  2.51523
```

```r
hist(filtered_cogTests$digitsymbol_composite, main = "Digit Symbol", xlab = "Z-Score")
```



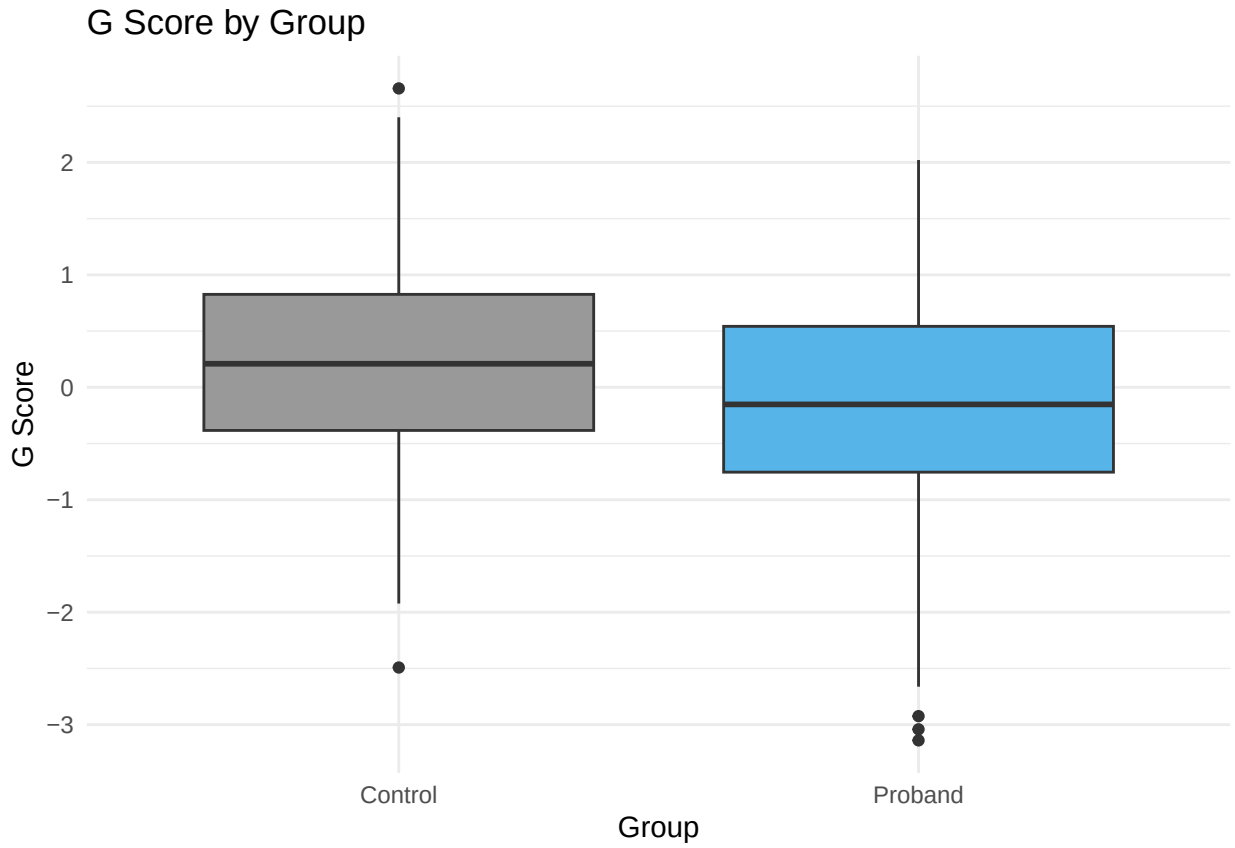

```r
# Combine all composite scores into new data table
composite_cogTests <- filtered_cogTests %>%
  select(studyid, digitsymbol_composite, facialmemory_z, verbalfluency_composite, cvlt_correct_z, cvlt_
view(composite_cogTests)
# combine g scores with tests
final_cogData <- composite_cogTests %>%
  left_join(filtered_gScores, by = "studyid")
view(final_cogData)

final_cogData$studyid <- as.character(final_cogData$studyid)
combined_phenotype$studyid <- as.character(combined_phenotype$studyid)
```

```r
# Combine cog and phenotype data
pheno_cog <- final_cogData %>%
  left_join(combined_phenotype, by = "studyid")
view(pheno_cog)
```

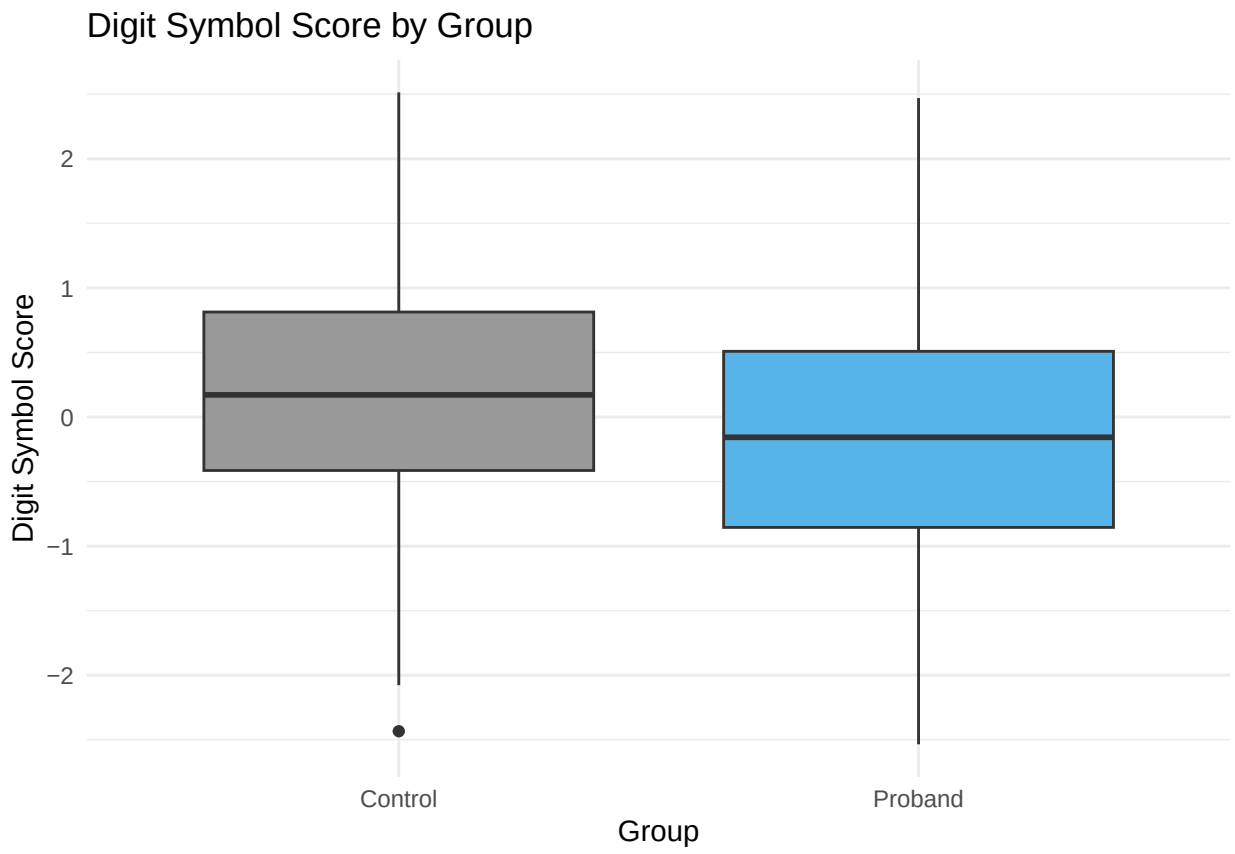
Correlation Testing Probands/Controls & Cog Data

```r
# Box plot comparing the Cognitive Scores for probands vs controls.

ggplot(pheno_cog, aes(x=factor(group), y = g)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Group", y = "G Score", title = "G Score by Group") +
  theme_minimal()
```

```
## Warning: Removed 46 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```

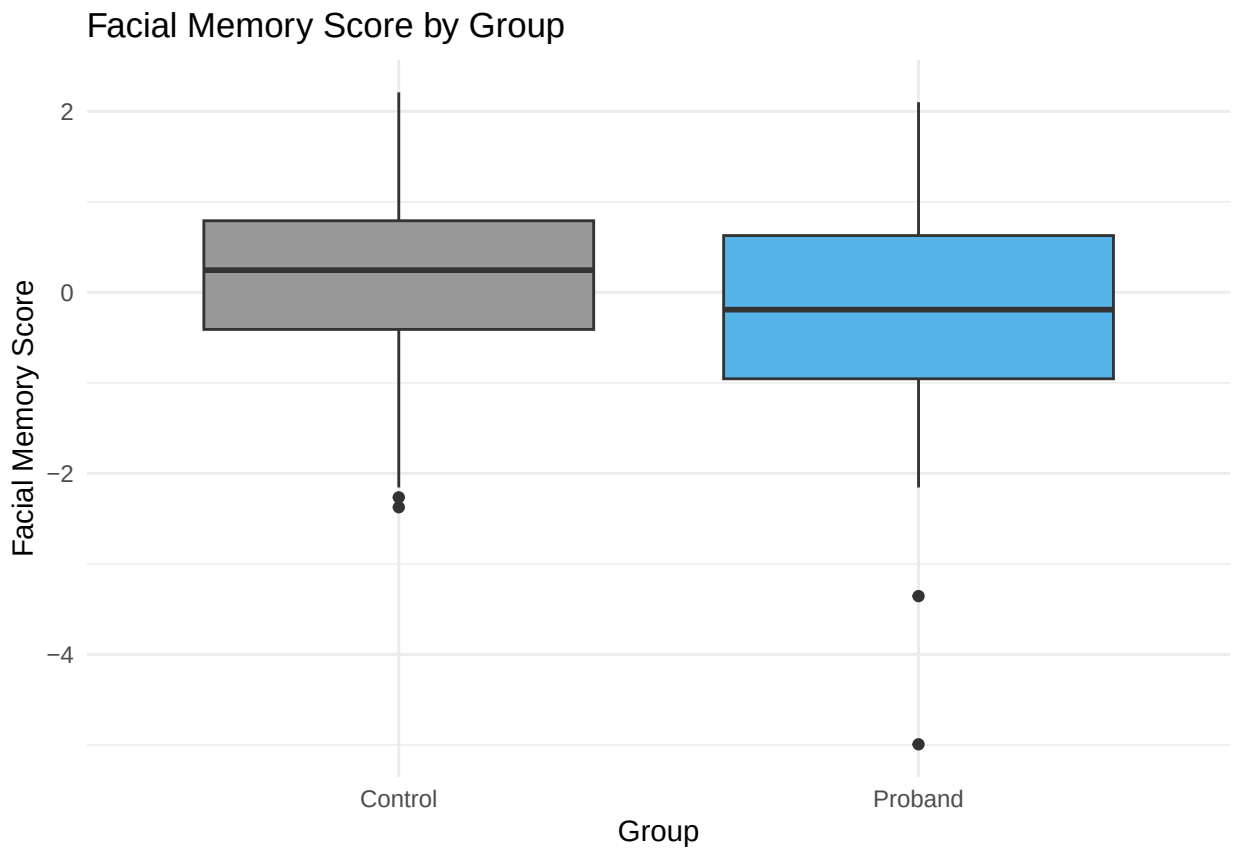



```r
ggplot(pheno_cog, aes(x=factor(group), y = digitsymbol_composite)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Group", y = "Digit Symbol Score", title = "Digit Symbol Score by Group") +
  theme_minimal()
```



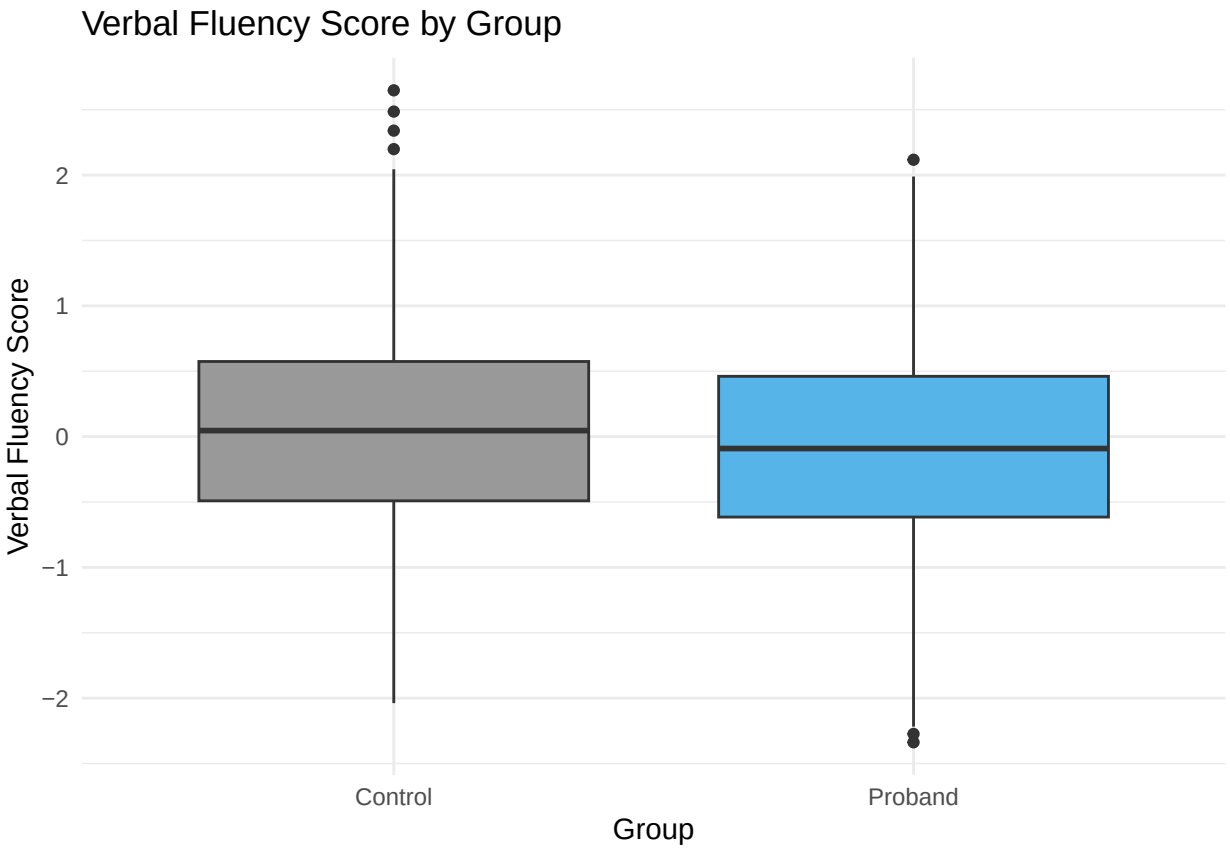

```
ggplot(pheno_cog, aes(x=factor(group), y = facialmemory_z)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Group", y = "Facial Memory Score", title = "Facial Memory Score by Group") +
  theme_minimal()
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```

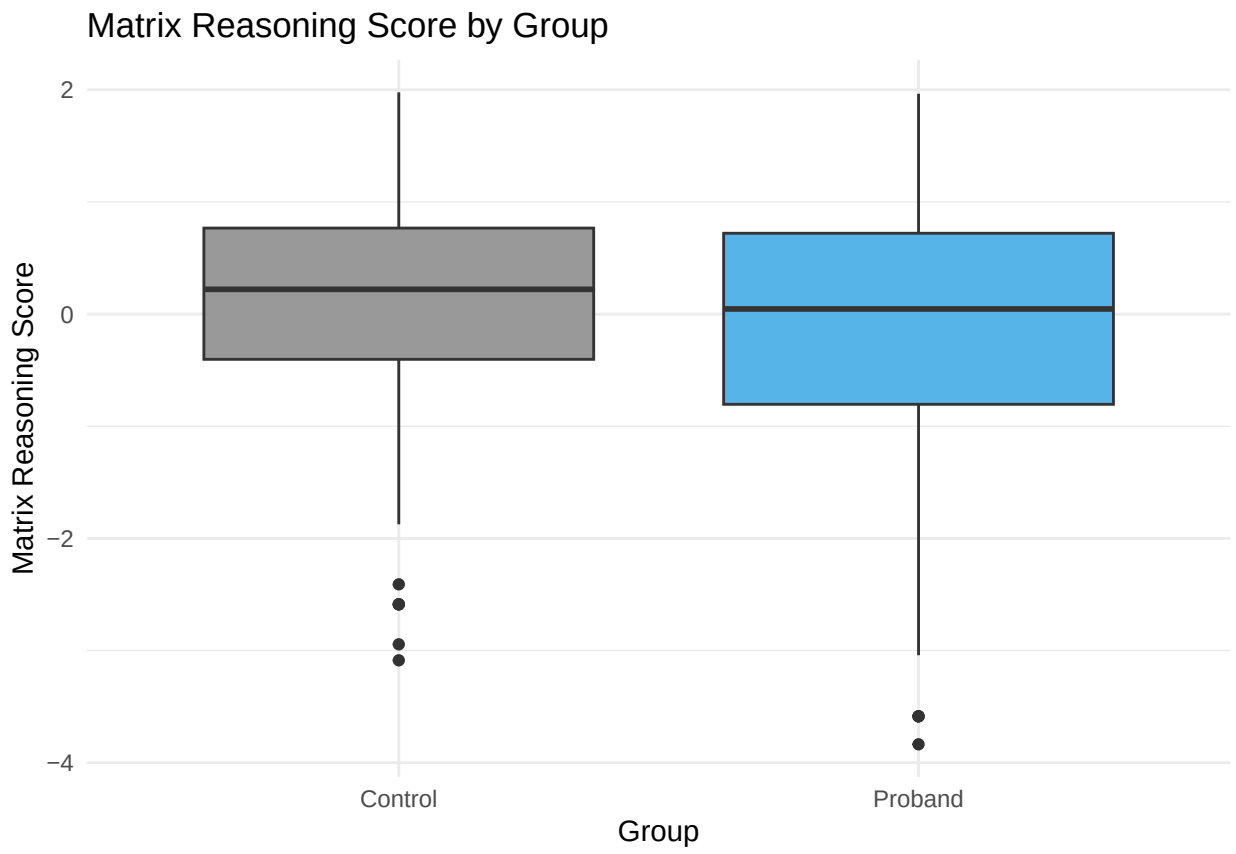



```
ggplot(pheno_cog, aes(x=factor(group), y = verbalfluency_composite)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Group", y = "Verbal Fluency Score", title = "Verbal Fluency Score by Group") +
  theme_minimal()
```

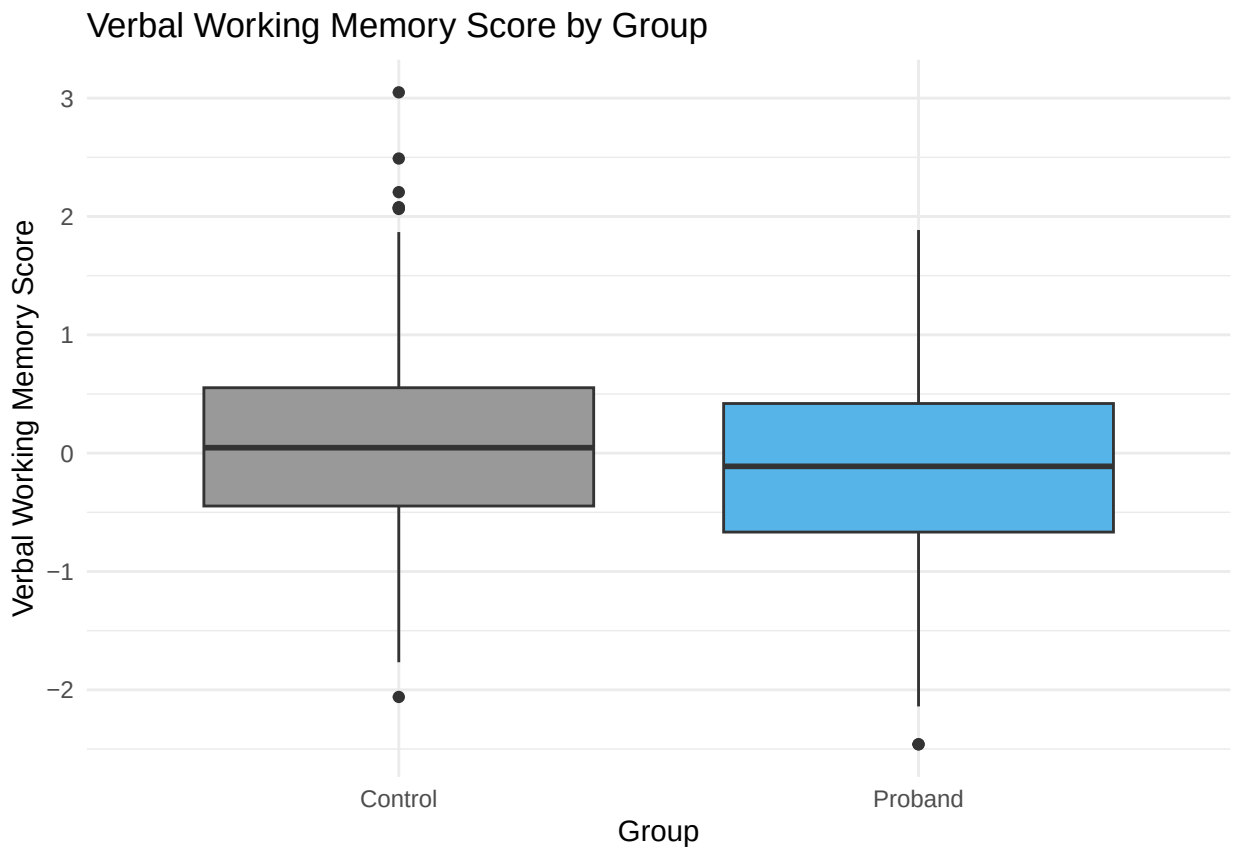
```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```




```
ggplot(pheno_cog, aes(x=factor(group), y = matrixreasoning_z)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Group", y = "Matrix Reasoning Score", title = "Matrix Reasoning Score by Group") +
  theme_minimal()
```

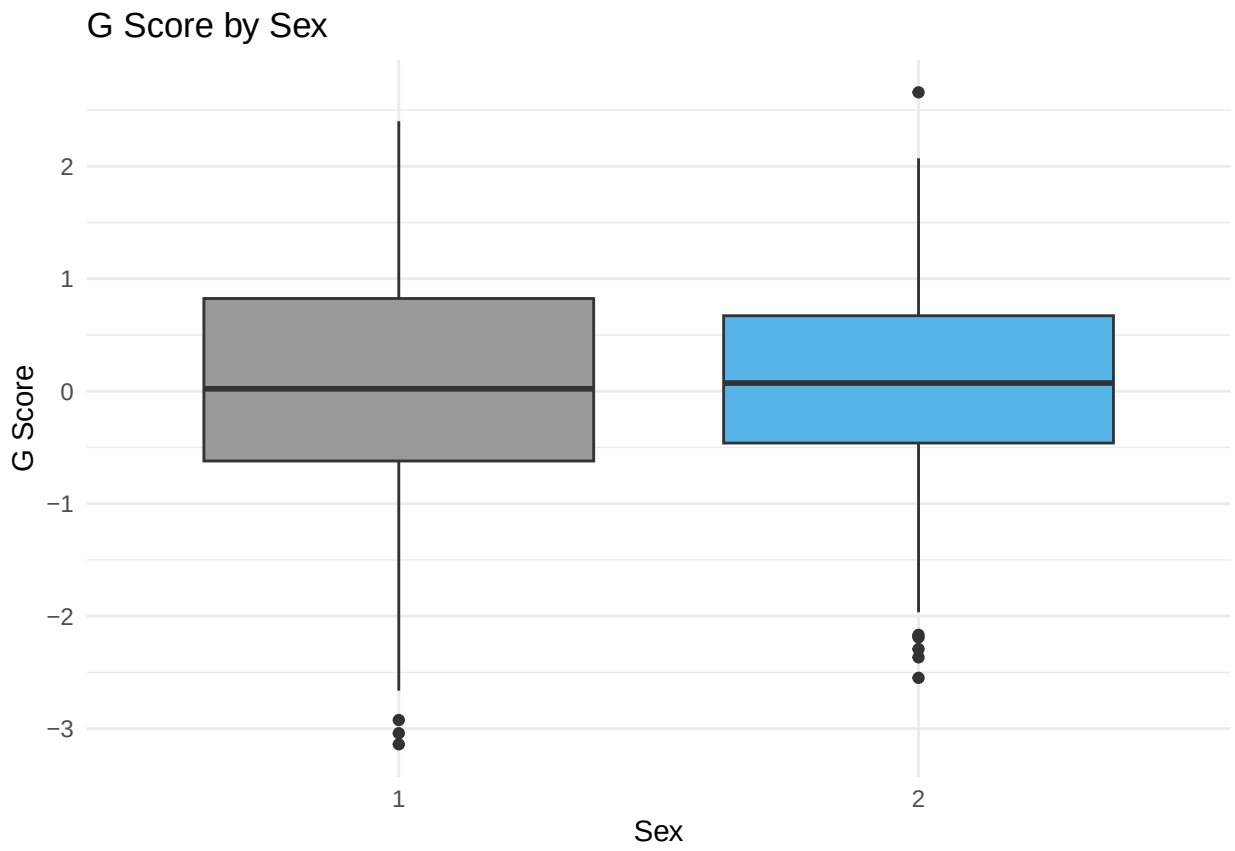



```
ggplot(pheno_cog, aes(x=factor(group), y = vwm_composite)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Group", y = "Verbal Working Memory Score", title = "Verbal Working Memory Score by Group") +
  theme_minimal()
```




```
ggplot(pheno_cog, aes(x=factor(sex), y = g)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Sex", y = "G Score", title = "G Score by Sex") +
  theme_minimal()
```

```
## Warning: Removed 46 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```



SNV and CNV data

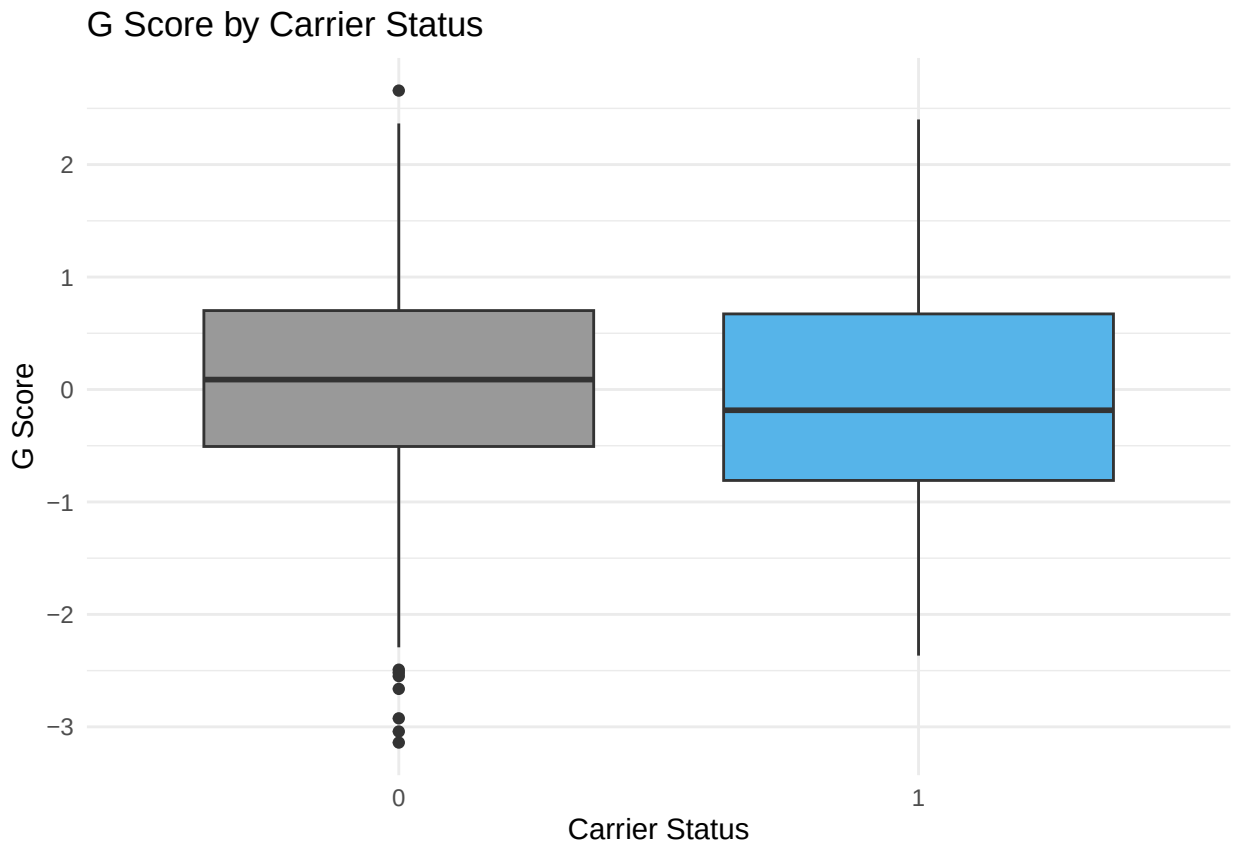
```
carriers <- read_csv("~/Downloads/cnv_snv_carriers(Sheet1).csv")
```

```
## Rows: 159 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): gene, variant_type, group, carrier_type
## dbl (3): studyid, is_cnv_carrier, is_snv_carrier
##
## 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.
```

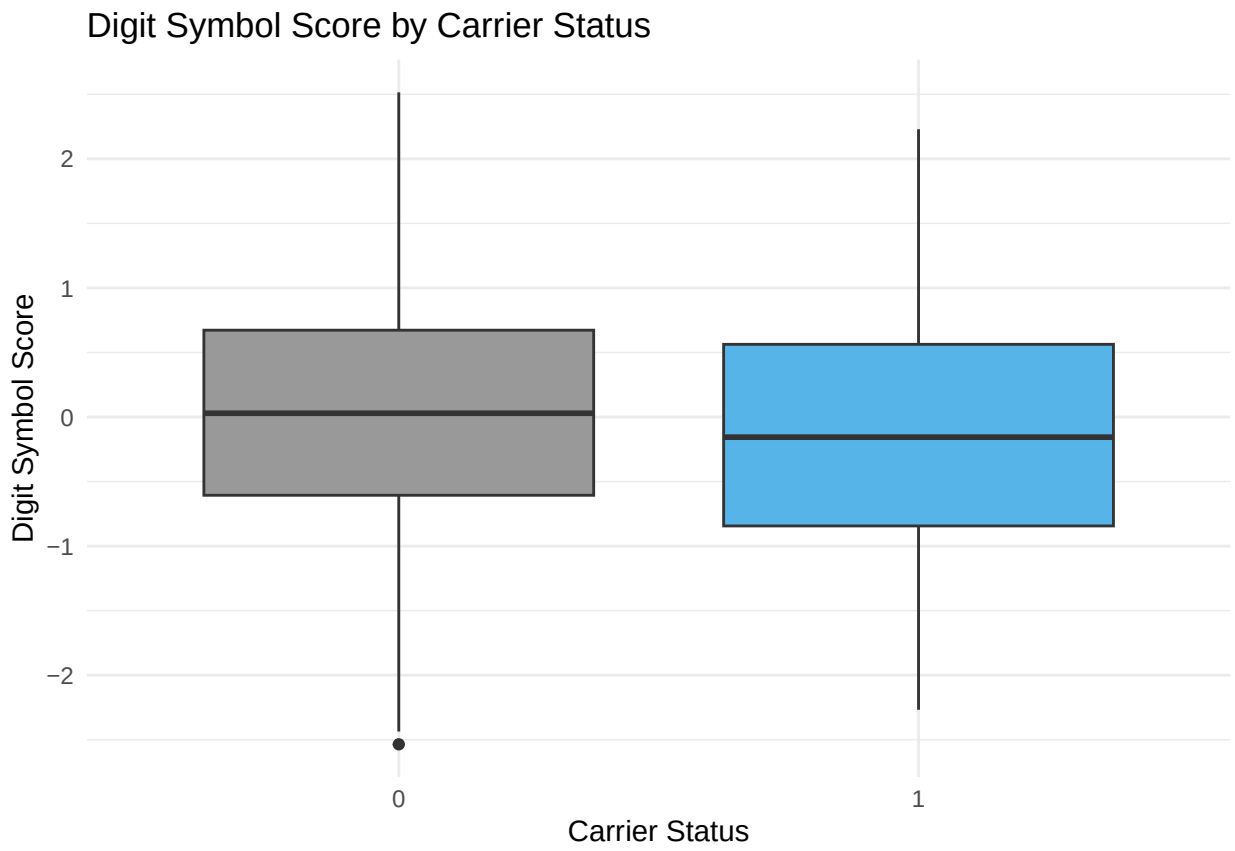
```
carriers <- carriers %>%
  mutate(
    is_carrier = 1
  )

filtered_carriers <- carriers[carriers$studyid %in% combined_phenotype$studyid, ]

filtered_carriers$studyid <- as.character(filtered_carriers$studyid)
pheno_cog$studyid <- as.character(pheno_cog$studyid)
```

```r
n_occur <- data.frame(table(filtered_carriers$studyid))

# add carrier status to pheno and cog data
carriers_combined <- pheno_cog %>%
  left_join(filtered_carriers, by = "studyid")

carriers_combined$is_carrier[is.na(carriers_combined$is_carrier)] <- 0

ggplot(carriers_combined, aes(x=factor(is_carrier), y = g)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "G Score", title = "G Score by Carrier Status") +
  theme_minimal()
```

```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

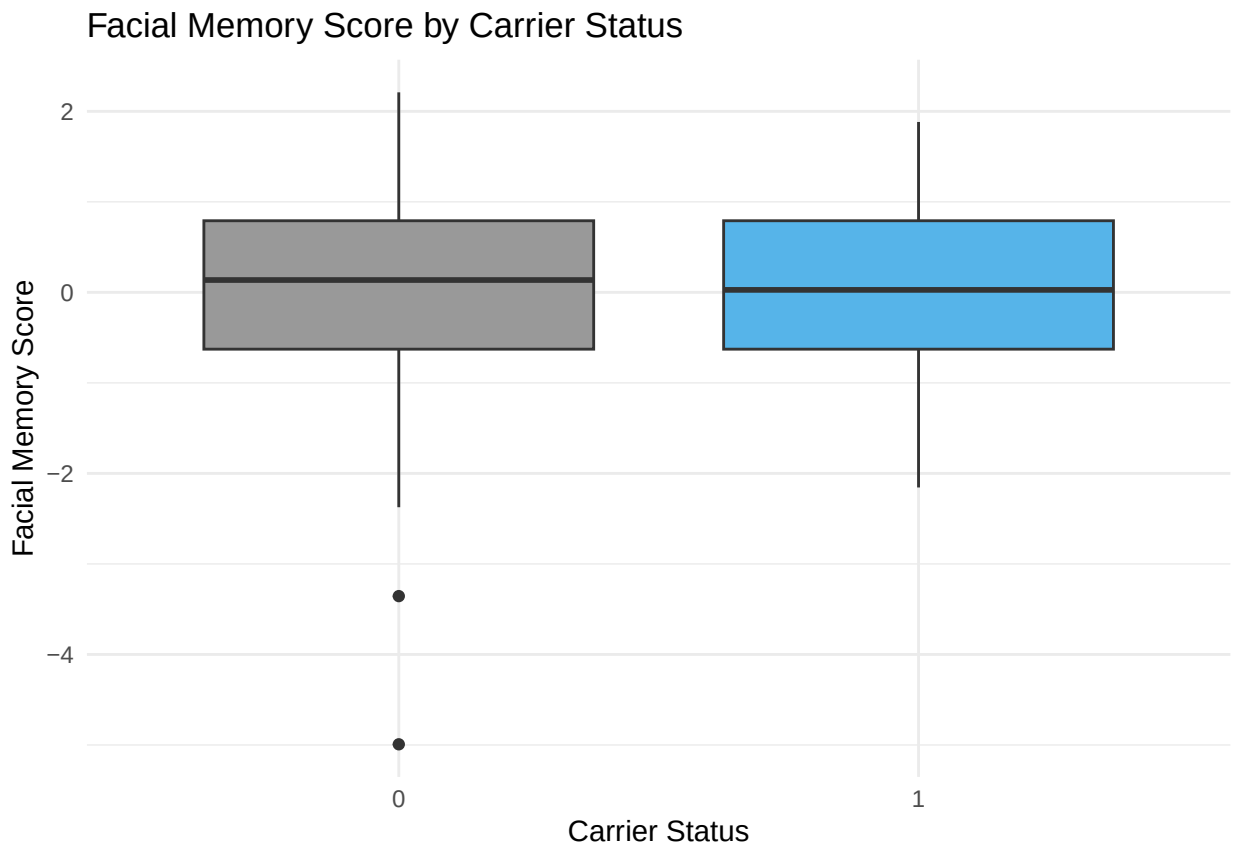



```r
ggplot(carriers_combined, aes(x=factor(is_carrier), y = digitsymbol_composite)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "Digit Symbol Score", title = "Digit Symbol Score by Carrier Status") +
  theme_minimal()
```



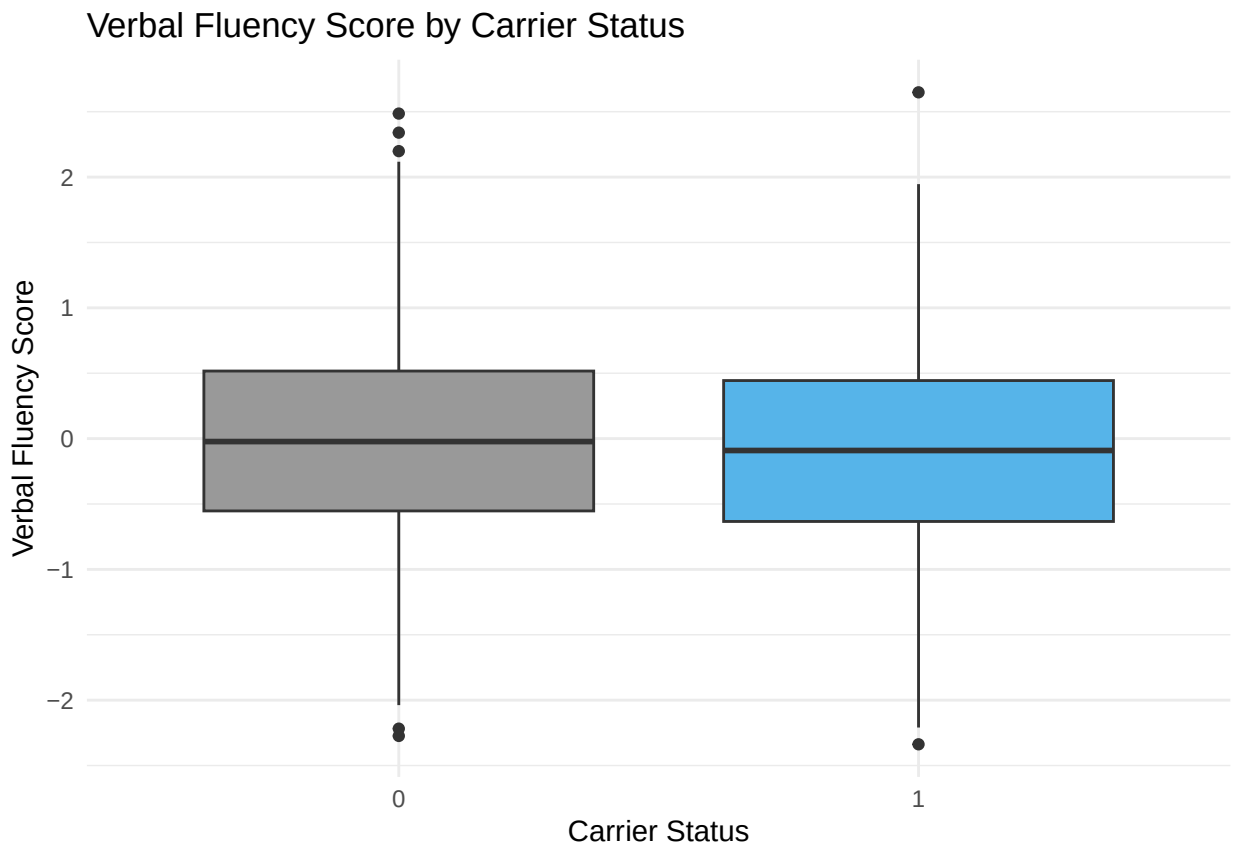

```
ggplot(carriers_combined, aes(x=factor(is_carrier), y = facialmemory_z)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "Facial Memory Score", title = "Facial Memory Score by Carrier Status")
  theme_minimal()
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```

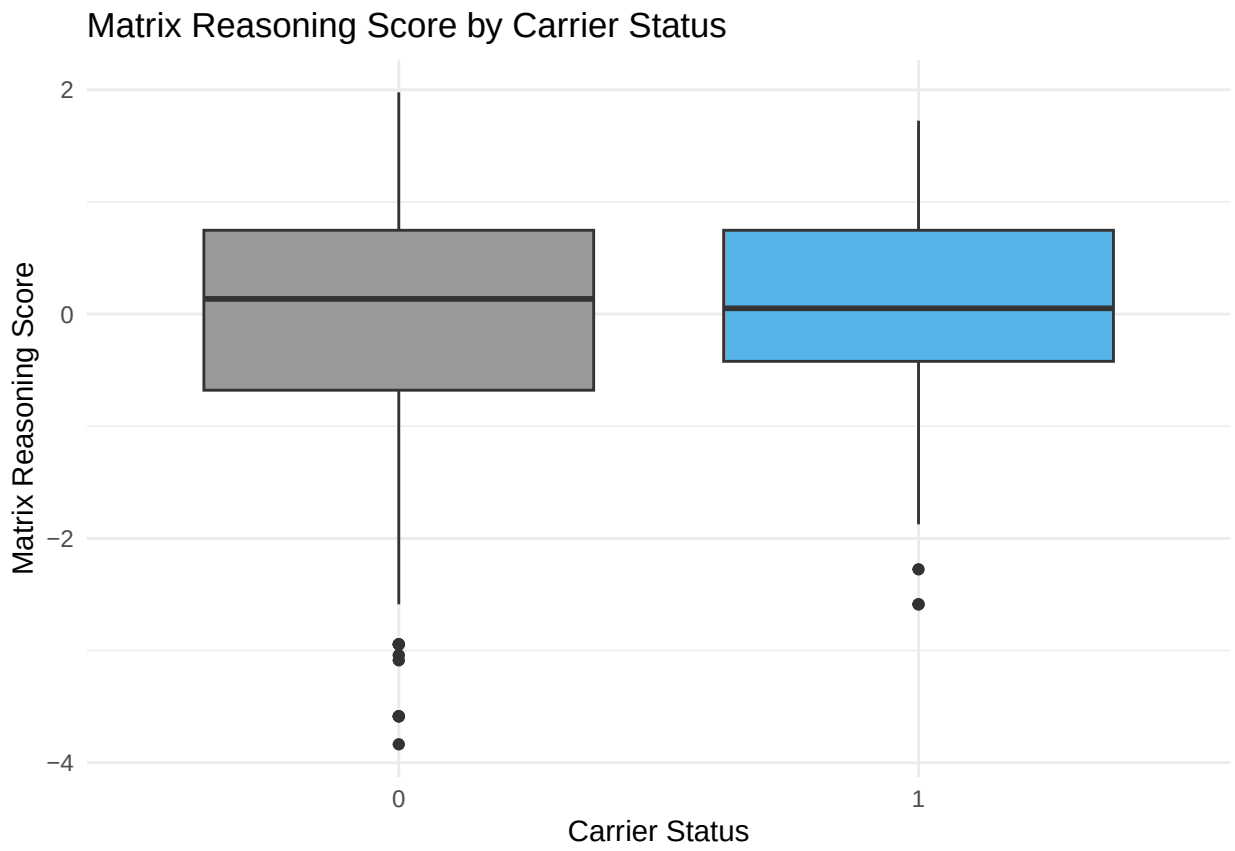



```
ggplot(carriers_combined, aes(x=factor(is_carrier), y = verbalfluency_composite)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "Verbal Fluency Score", title = "Verbal Fluency Score by Carrier Status
  theme_minimal()
```

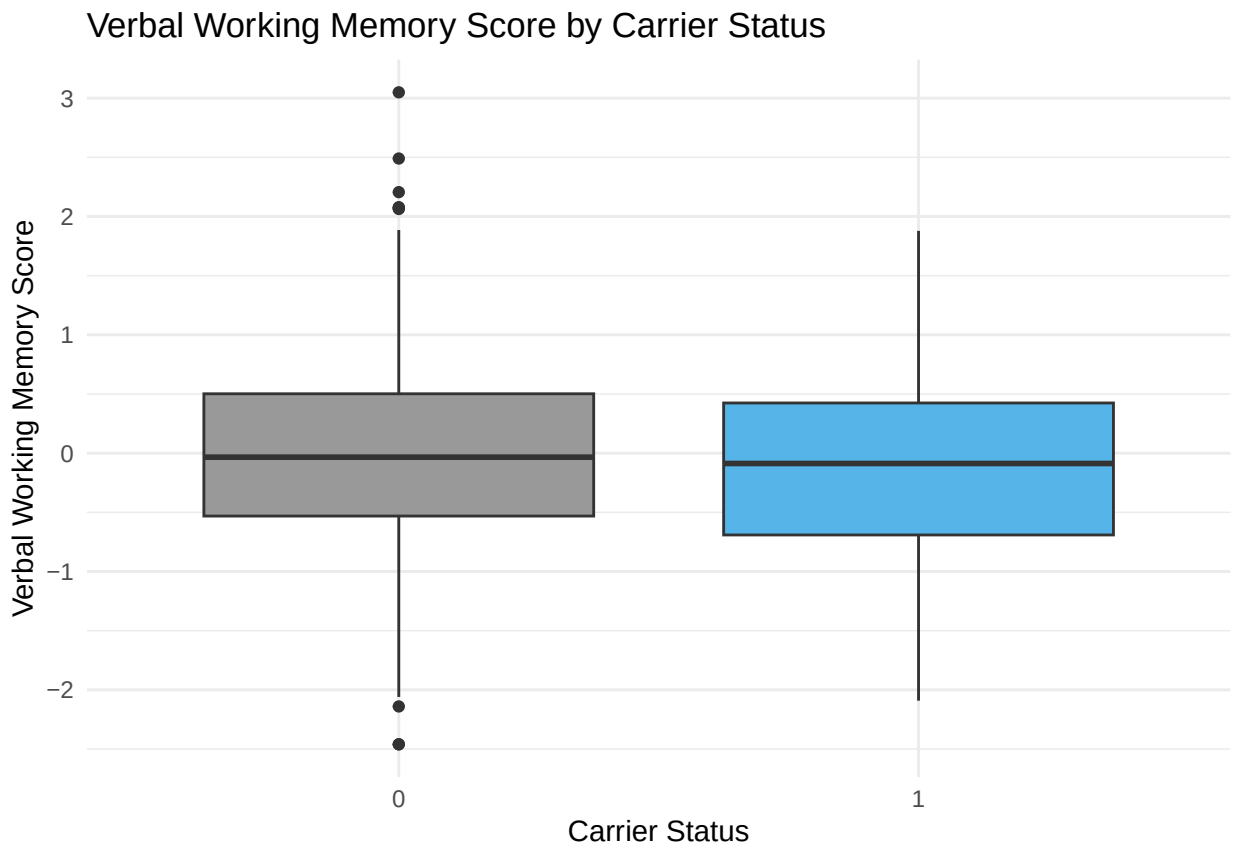
```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```



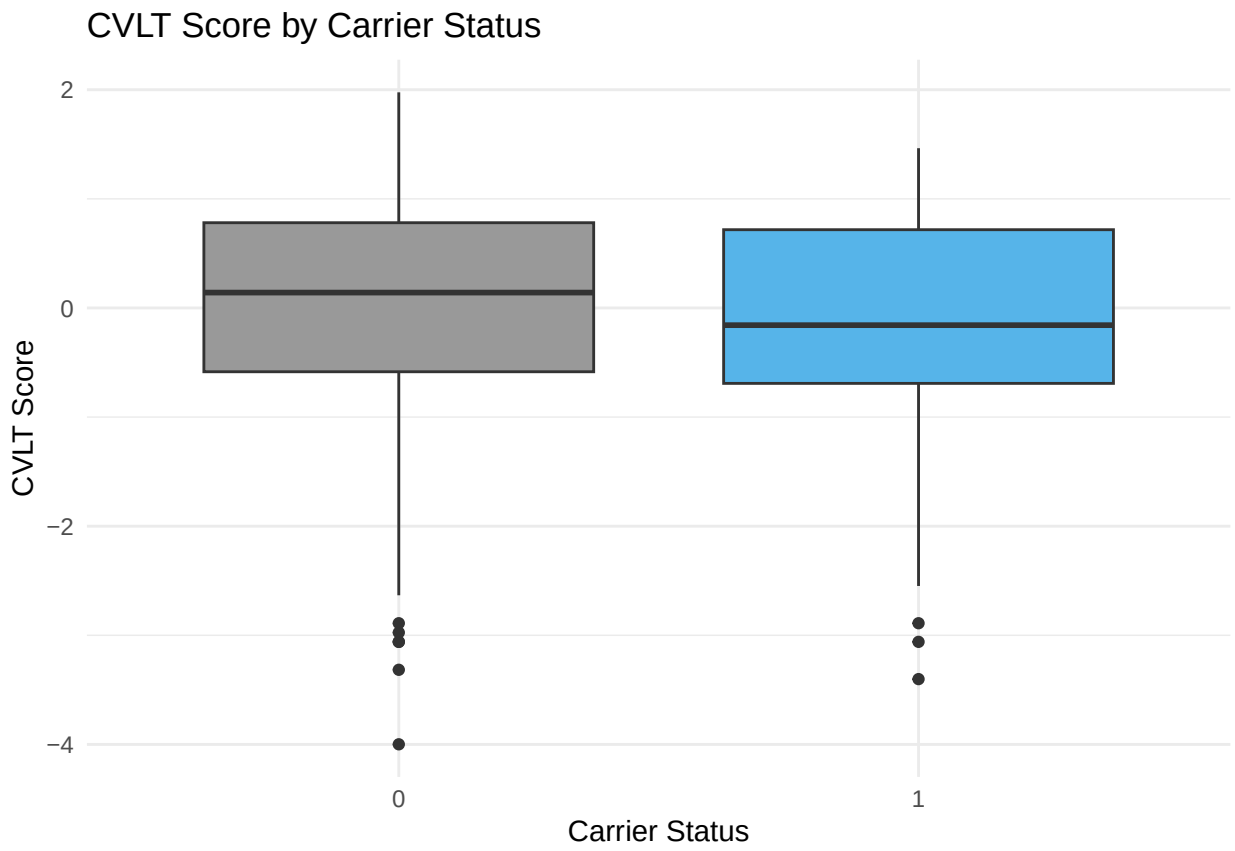

```
ggplot(carriers_combined, aes(x=factor(is_carrier), y = matrixreasoning_z)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "Matrix Reasoning Score", title = "Matrix Reasoning Score by Carrier St
  theme_minimal()
```




```
ggplot(carriers_combined, aes(x=factor(is_carrier), y = vwm_composite)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "Verbal Working Memory Score", title = "Verbal Working Memory Score by
  theme_minimal()
```

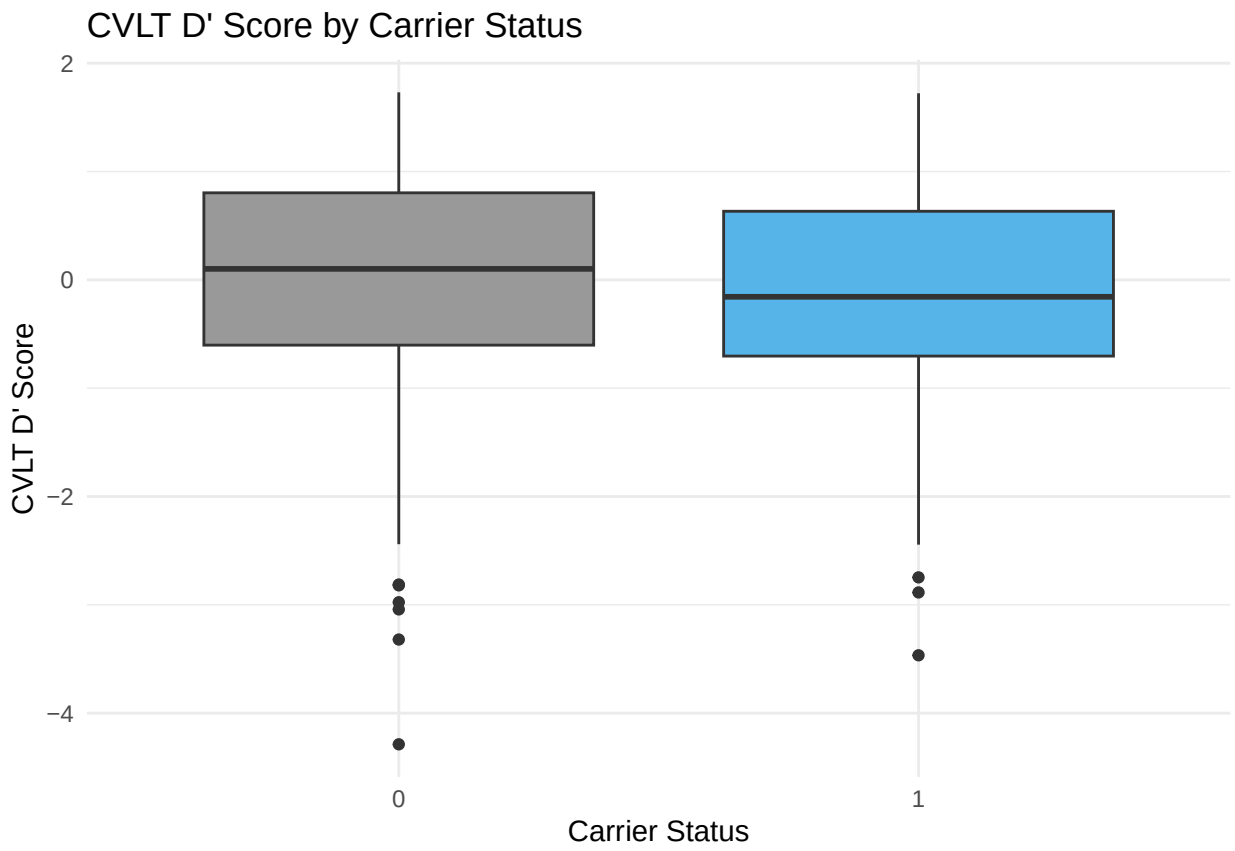



```
ggplot(carriers_combined, aes(x=factor(is_carrier), y = cvlt_correct_z)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "CVLT Score", title = "CVLT Score by Carrier Status") +
  theme_minimal()
```




```
ggplot(carriers_combined, aes(x=factor(is_carrier), y = cvlt_dprime_z)) +
  geom_boxplot(fill = c("#999999", "#56B4E9")) +
  labs(x = "Carrier Status", y = "CVLT D' Score", title = "CVLT D' Score by Carrier Status") +
  theme_minimal()
```

```
## Warning: Removed 19 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```




```
table(filtered_carriers$variant_type)
```

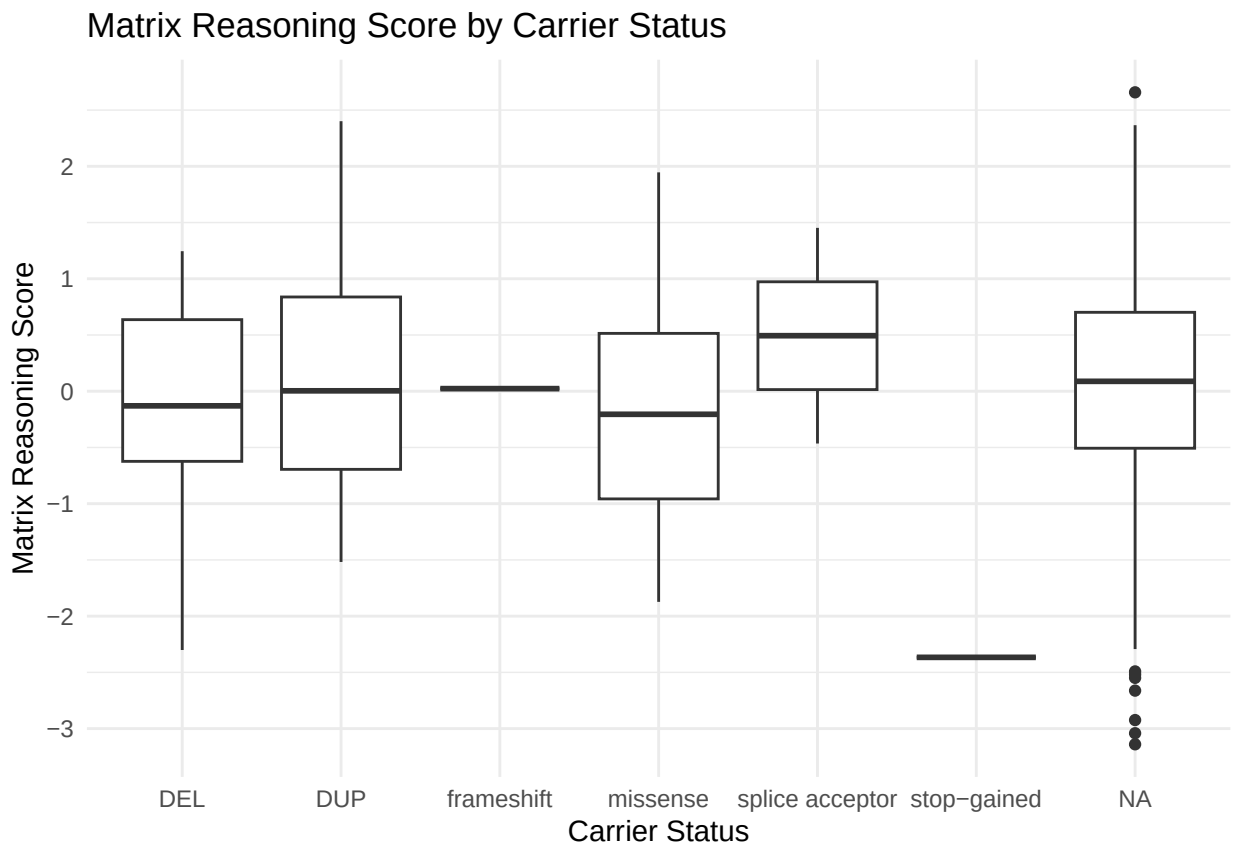
```
##
##             DEL             DUP      frameshift        missense splice acceptor
##              42              25               1              41               2
##     stop-gained
##               1
```

```
table(filtered_carriers$gene)
```

```
##
##  AKAP11   ATP9A CACNA1B CACNA1G CAGNA1G   CDK13    CUL1  GRIN2A   HDAC9   HERC1
##       1       3      16       3       2       2       3       1      10      15
##  JARID2   MAGI2    NBEA   NGLN2   NLGN2   PSMA3   SETD1     SP4   STAG1   TOP2B
##       4       4      16       2       2       1       1       7       9       3
##    TRIO  ZNF318
##       5       2
```

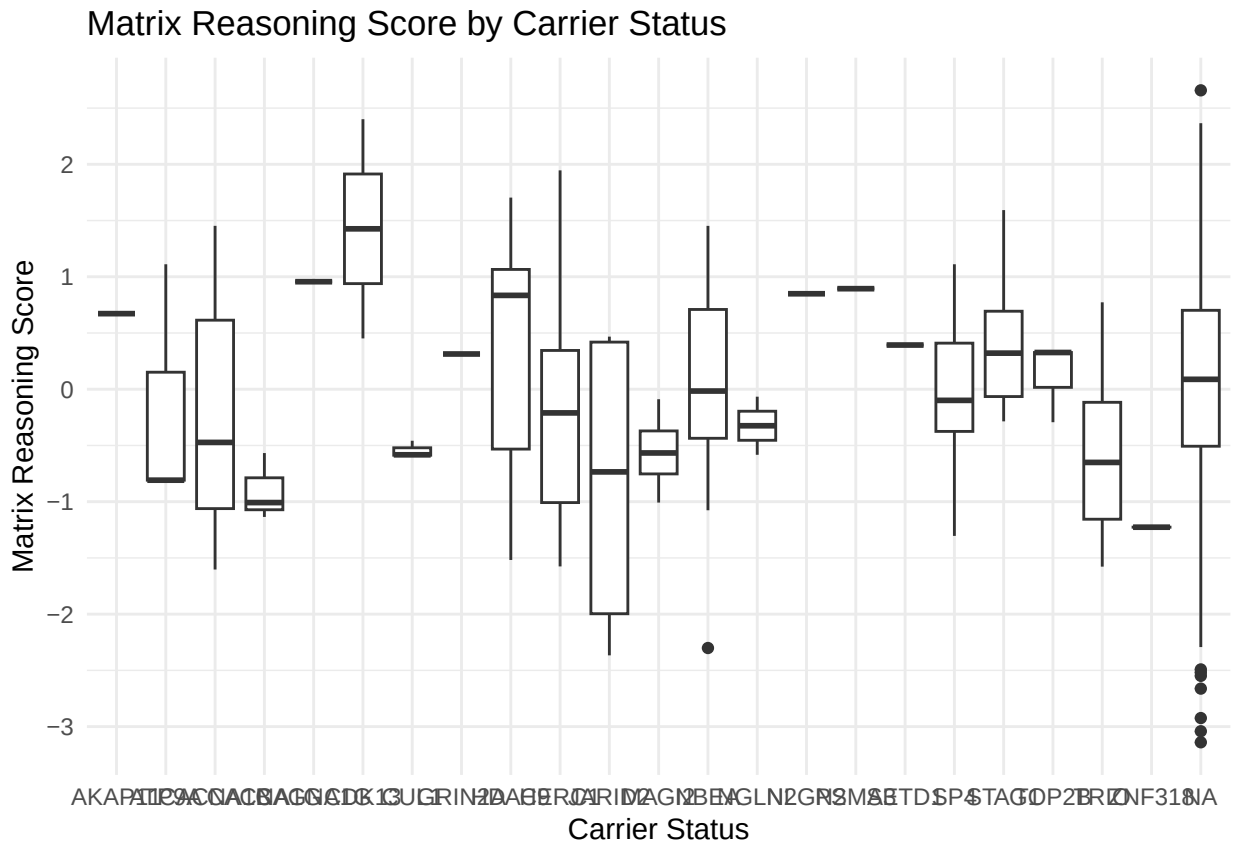
```
ggplot(carriers_combined, aes(x=factor(variant_type), y = g)) +
  geom_boxplot( ) +
  labs(x = "Carrier Status", y = "Matrix Reasoning Score", title = "Matrix Reasoning Score by Carrier S
  theme_minimal()
```

```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```



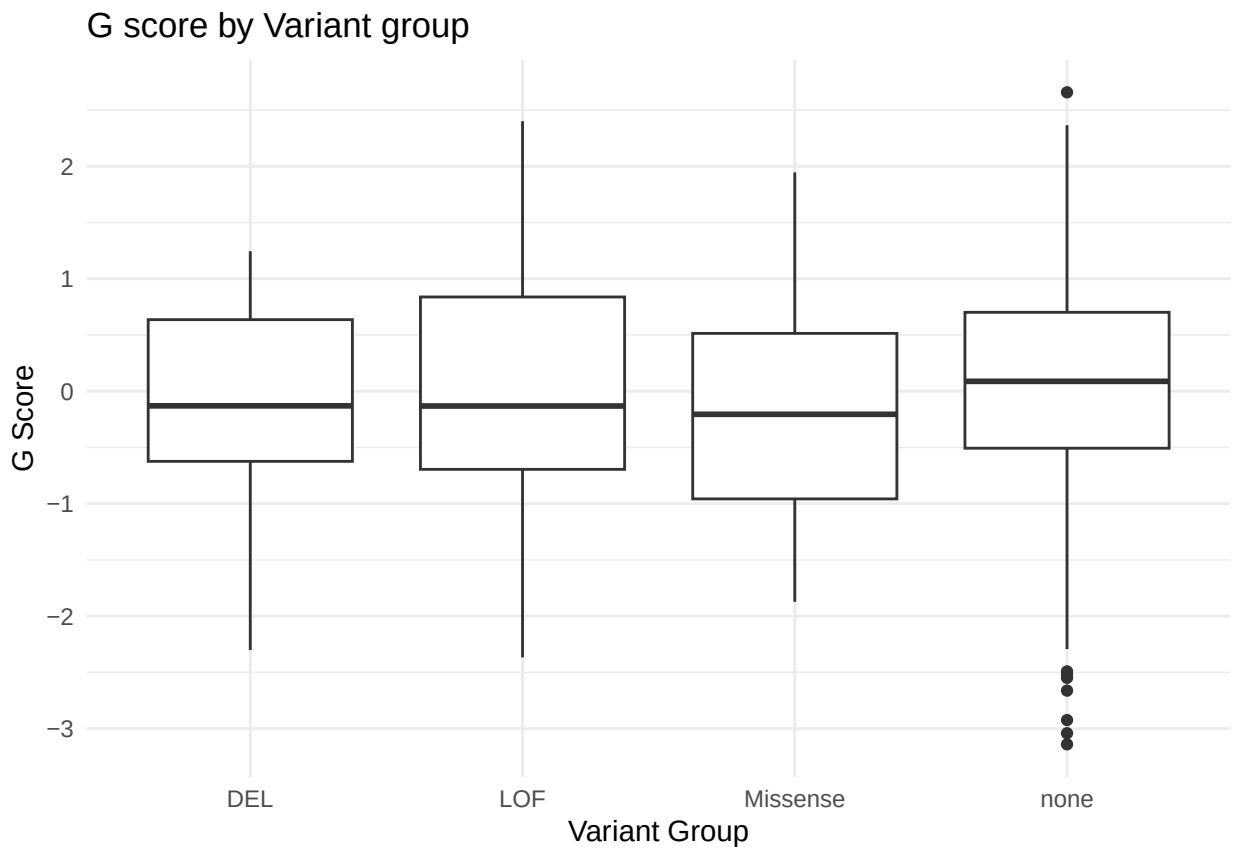

```
ggplot(carriers_combined, aes(x=factor(gene), y = g)) +
  geom_boxplot( ) +
  labs(x = "Carrier Status", y = "Matrix Reasoning Score", title = "Matrix Reasoning Score by Carrier St
  theme_minimal()
```

```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```




```
# Group variant type: DUP + LOF, DEL, Missense
grouped_types <- carriers_combined %>%
  mutate(
    variant_group = case_when(
      variant_type %in% c("stop-gained", "frameshift", "splice acceptor", "DUP") ~ "LOF",
      variant_type %in% c("missense") ~ "Missense",
      variant_type %in% c("DEL") ~ "DEL",
      TRUE ~ "none"
    )
  )

ggplot(grouped_types, aes(x=factor(variant_group), y = g)) +
  geom_boxplot( ) +
  labs(x = "Variant Group", y = "G Score", title = "G score by Variant group") +
  theme_minimal()
```

```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```

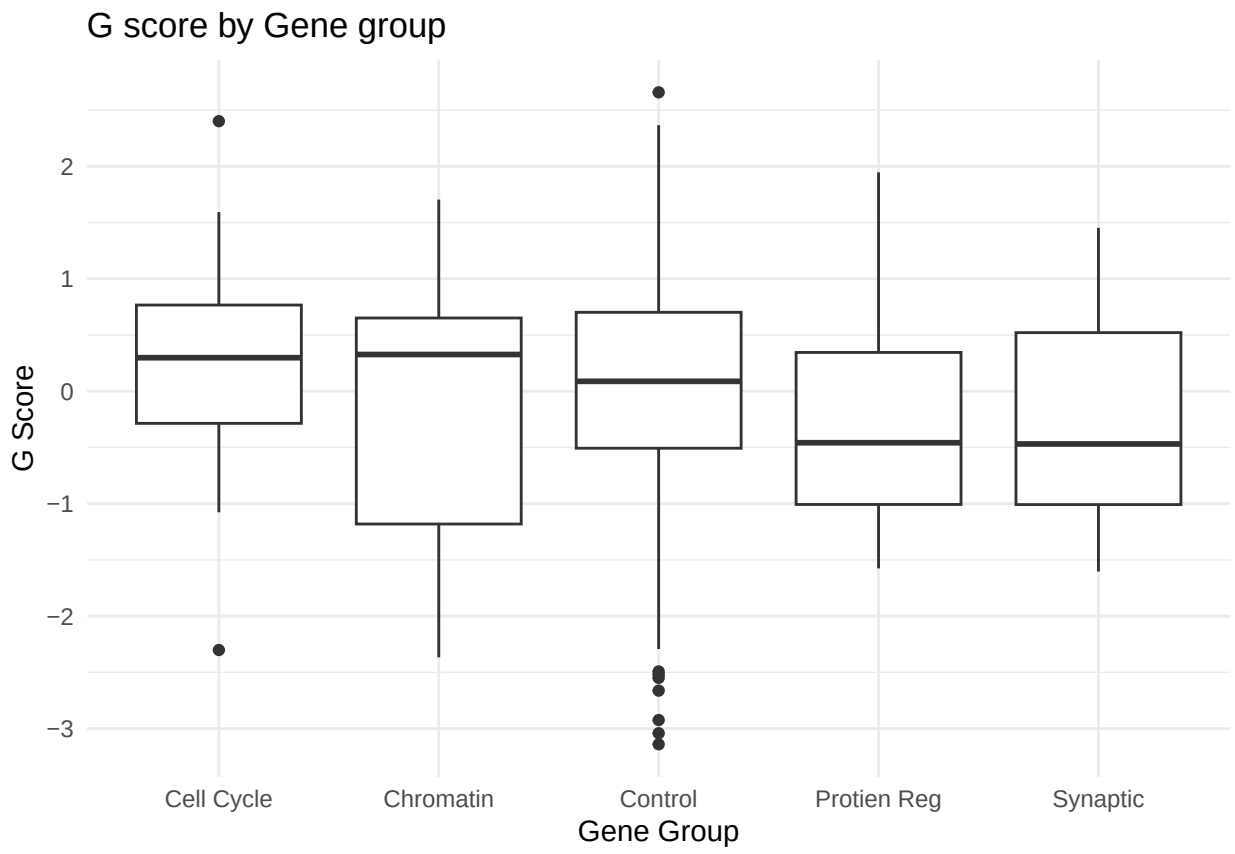



```
table(grouped_types$variant_group)
```

```
##
##      DEL      LOF Missense     none
##       42       29       41      436
```

```
# Group individual genes by biological function
grouped_types <- carriers_combined %>%
  mutate(
    gene_group = case_when(
      gene %in% c("CACNA1B", "CACNA1G", "CAGNA1G", "GRIN2A", "MAGI2", "NGLN2", "NLGN2", "SP4", "TRIO") ~
      gene %in% c("CUL1", "HERC1", "PSMA3") ~ "Protien Reg",
      gene %in% c("CDK13", "NBEA", "STAG1", "AKAP11", "ATP9A") ~ "Cell Cycle",
      gene %in% c("SETD1", "HDAC9", "JARID2", "TOP2B", "ZNF318") ~ "Chromatin",
      TRUE ~ "Control"
    )
  )

ggplot(grouped_types, aes(x=factor(gene_group), y = g)) +
  geom_boxplot( ) +
  labs(x = "Gene Group", y = "G Score", title = "G score by Gene group") +
  theme_minimal()
```

```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```




```
table(grouped_types$gene_group)
```

```
##
##  Cell Cycle   Chromatin     Control Protien Reg    Synaptic
##          31          20         436          19          42
```

```
carriers_cog <- grouped_types %>%
  filter(is_carrier == 1)

# pairwise t test
pairwise.t.test(
  x = carriers_cog$g,
  g = carriers_cog$gene_group,
  p.adjust.method = "bonferroni"
)
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  carriers_cog$g and carriers_cog$gene_group
##
##             Cell Cycle Chromatin Protien Reg
## Chromatin   1.00       -         -
## Protien Reg 0.64       1.00      -
## Synaptic    0.30       1.00      1.00
```

```
##
## P value adjustment method: bonferroni
```

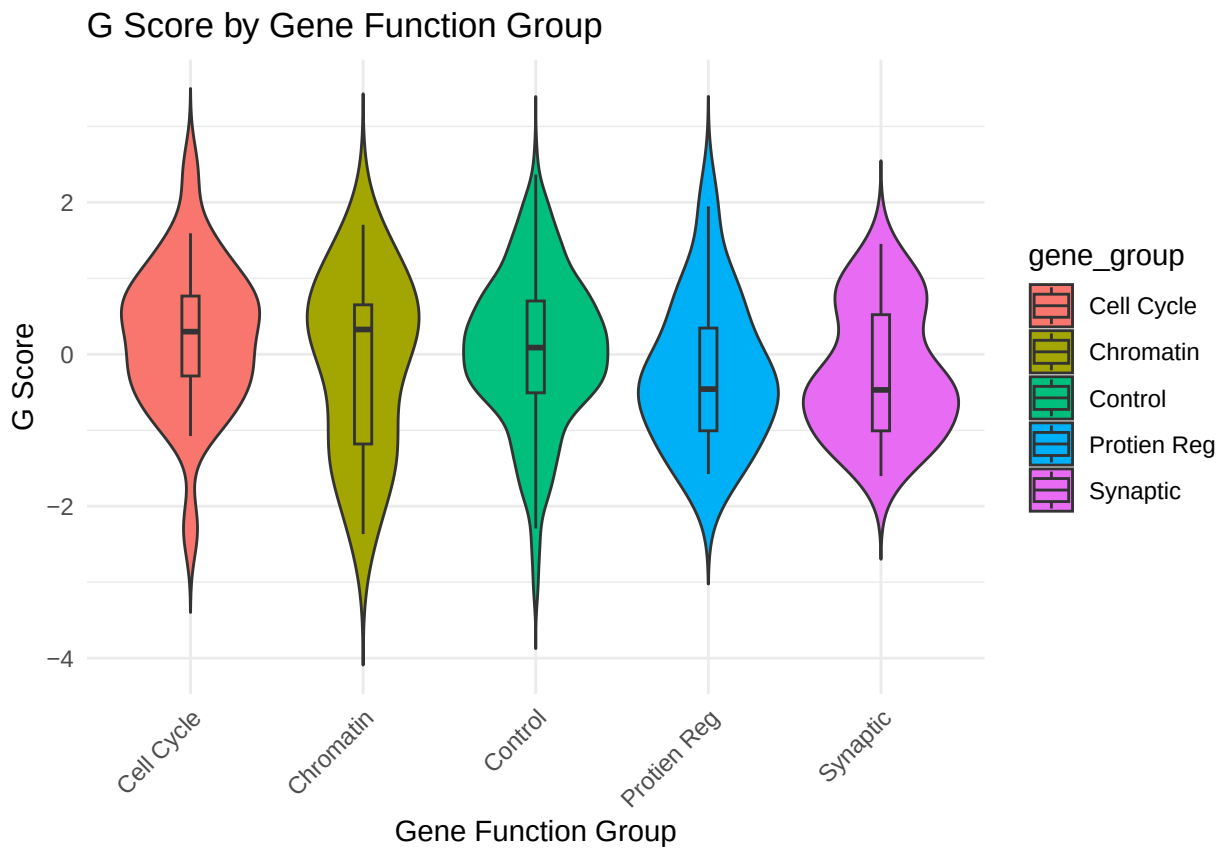
```
pairwise.t.test(
  x = carriers_cog$digitsymbol_composite,
  g = carriers_cog$gene_group,
  p.adjust.method = "bonferroni"
)
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  carriers_cog$digitsymbol_composite and carriers_cog$gene_group
##
##             Cell Cycle Chromatin Protien Reg
## Chromatin   1          -         -
## Protien Reg 1          1         -
## Synaptic    1          1         1
##
## P value adjustment method: bonferroni
```

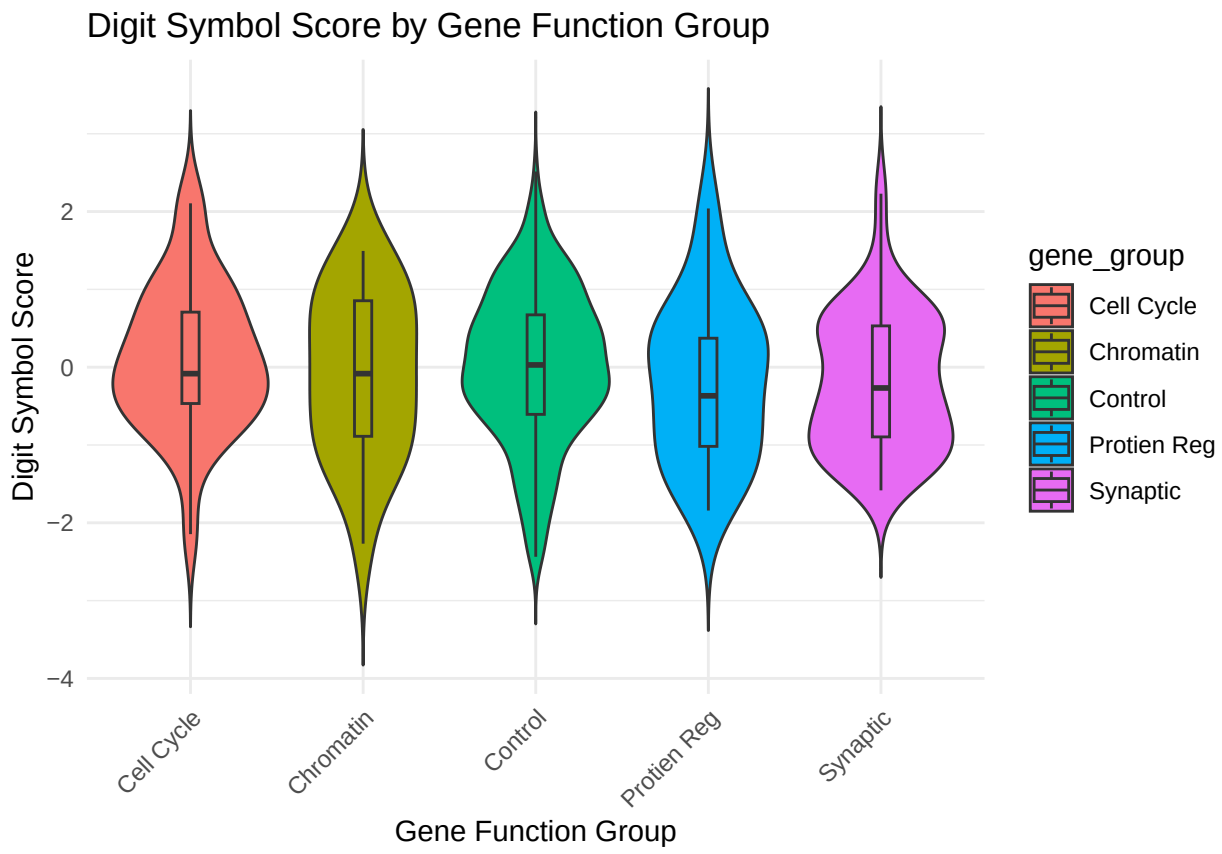
```
# Violin plot
ggplot(grouped_types, aes(x= gene_group, y = g, fill = gene_group)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, outlier.shape = NA) +
  labs(title = "G Score by Gene Function Group",
       x = "Gene Function Group",
       y = "G Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (`stat_ydensity()`).
```

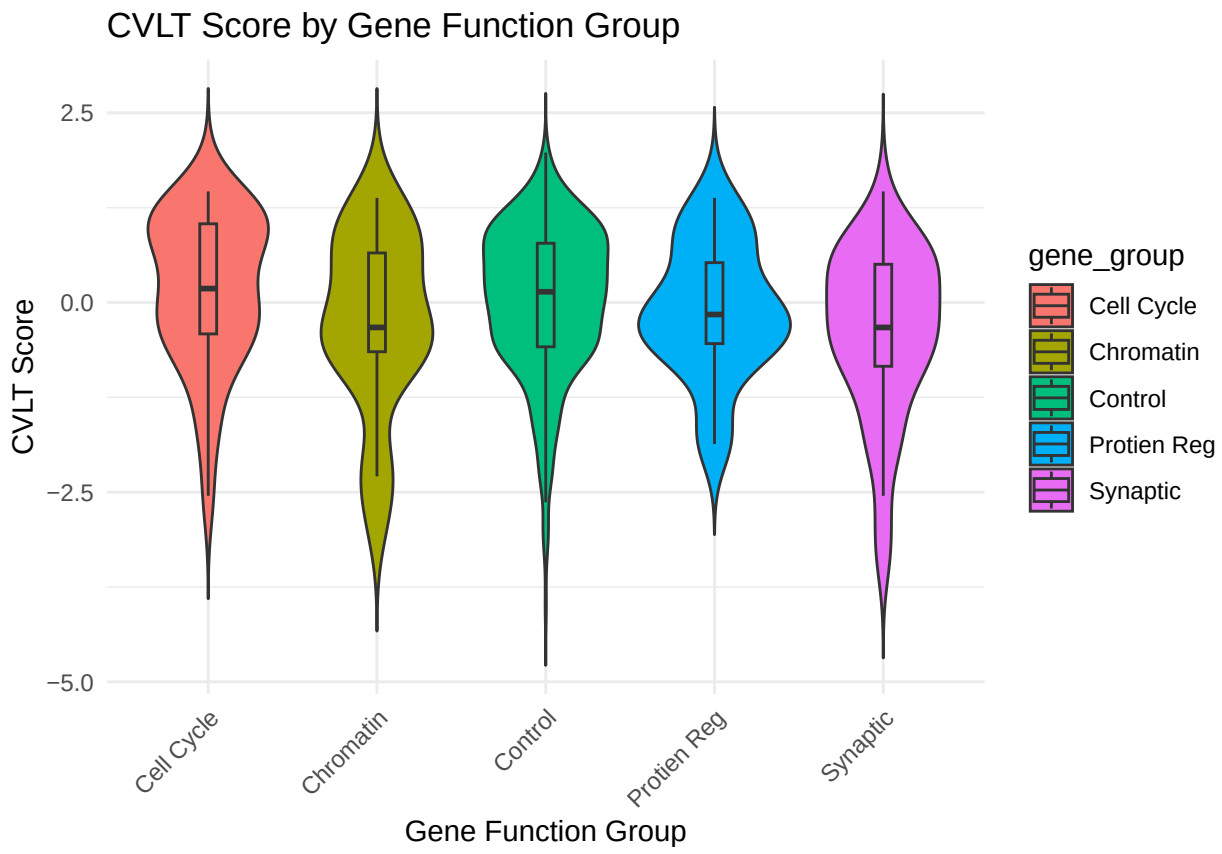
```
## Warning: Removed 47 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



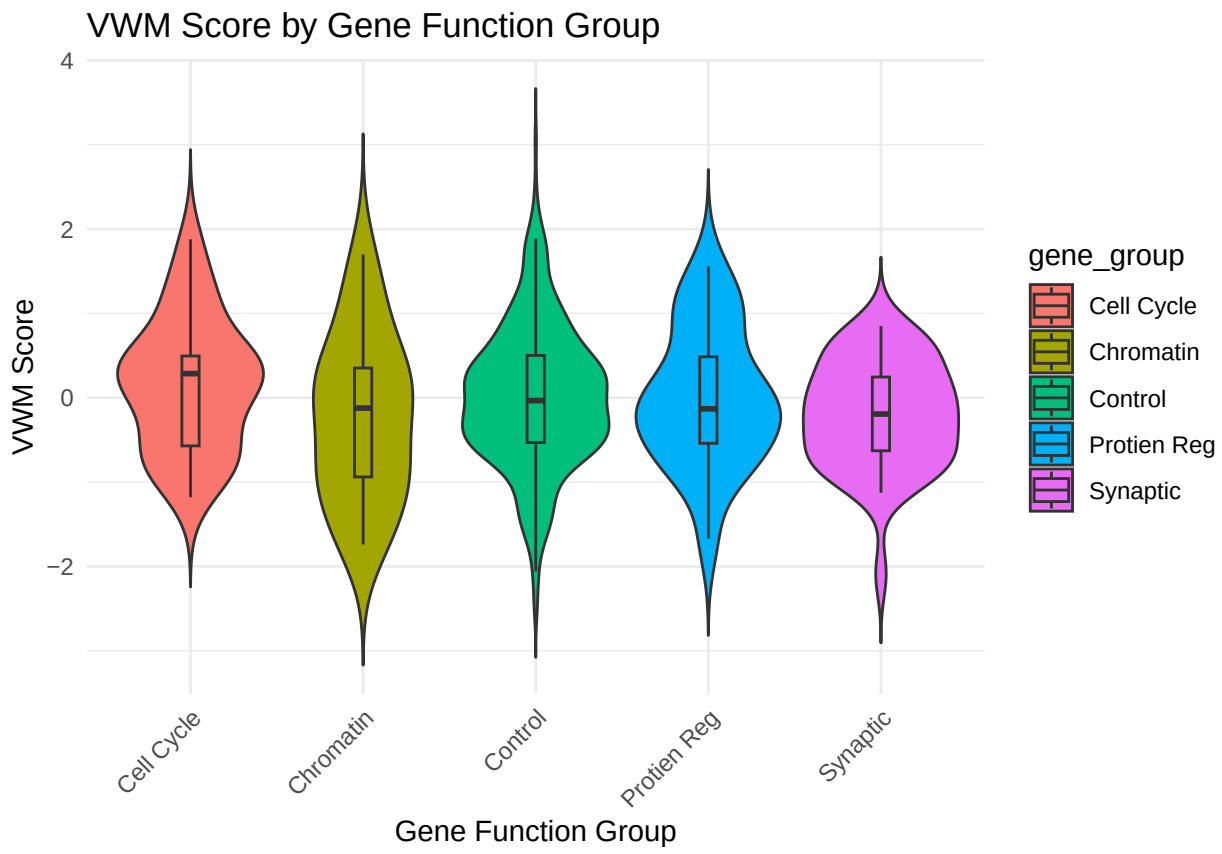

```
ggplot(grouped_types, aes(x= gene_group, y = digitsymbol_composite, fill = gene_group)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, outlier.shape = NA) +
  labs(title = "Digit Symbol Score by Gene Function Group",
       x = "Gene Function Group",
       y = "Digit Symbol Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(grouped_types, aes(x= gene_group, y = cvlt_correct_z, fill = gene_group)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, outlier.shape = NA) +
  labs(title = "CVLT Score by Gene Function Group",
       x = "Gene Function Group",
       y = "CVLT Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(grouped_types, aes(x= gene_group, y = vwm_composite, fill = gene_group)) +
  geom_violin(trim = FALSE) +
  geom_boxplot(width = 0.1, outlier.shape = NA) +
  labs(title = "VWM Score by Gene Function Group",
       x = "Gene Function Group",
       y = "VWM Score") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
model <- lm(g ~ is_carrier + age + sex, data = grouped_types)
summary(model)
```

```
##
## Call:
## lm(formula = g ~ is_carrier + age + sex, data = grouped_types)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4037 -0.5673  0.0516  0.6540  2.3745
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.10920    0.25988  -8.116 3.81e-15 ***
## is_carrier  -0.13698    0.10241  -1.338    0.182
## age          0.13635    0.01483   9.194  < 2e-16 ***
## sex          0.05516    0.08242   0.669    0.504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9195 on 497 degrees of freedom
##   (47 observations deleted due to missingness)
## Multiple R-squared:  0.1496, Adjusted R-squared:  0.1445
## F-statistic: 29.15 on 3 and 497 DF,  p-value: < 2.2e-16
```

```r
model <- lm(digitsymbol_composite ~ is_carrier + age + sex, data = grouped_types)
summary(model)
```

```
##
## Call:
## lm(formula = digitsymbol_composite ~ is_carrier + age + sex,
##     data = grouped_types)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.78527 -0.55237  0.02812  0.68976  2.48793
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.55759    0.24990  -6.233 9.19e-10 ***
## is_carrier  -0.10165    0.09785  -1.039    0.299
## age          0.11121    0.01427   7.793 3.33e-14 ***
## sex         -0.08333    0.07921  -1.052    0.293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9236 on 544 degrees of freedom
## Multiple R-squared:  0.1032, Adjusted R-squared:  0.09825
## F-statistic: 20.87 on 3 and 544 DF,  p-value: 8.278e-13
```

strong positive effect - older individuals score higher

carriers score ~0.14 points lower on g but this is not significant

Model is overall statistically significant - age adds menaingful predictive power

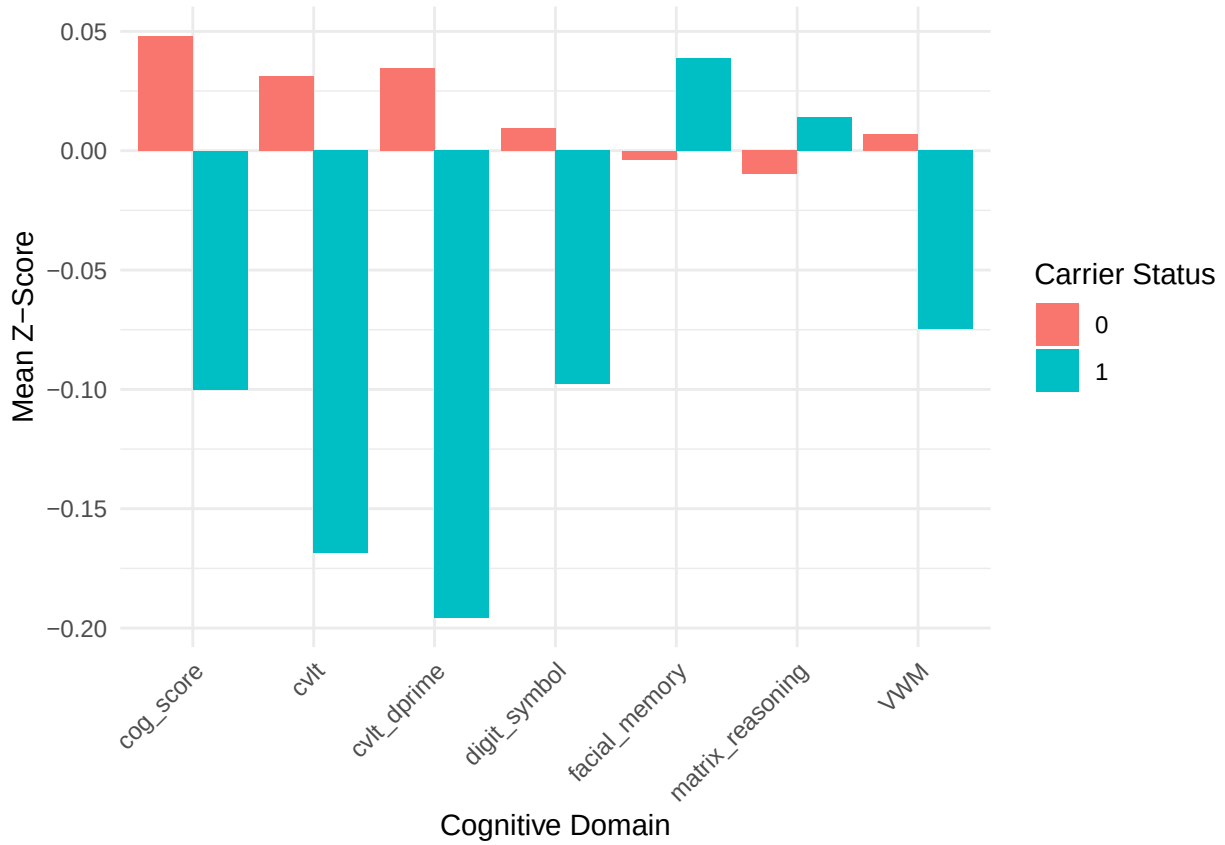
sec has no signfiicant effect on g

```r
model <- lm(g ~ gene_group + age + sex, data = grouped_types)
summary(model)
```

```
##
## Call:
## lm(formula = g ~ gene_group + age + sex, data = grouped_types)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4089 -0.5716  0.0676  0.6398  2.3710
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.89465    0.31025  -6.107 2.06e-09 ***
## gene_groupChromatin     -0.40542    0.27554  -1.471   0.1418
## gene_groupControl       -0.21740    0.18255  -1.191   0.2343
## gene_groupProtien Reg -0.57022    0.28407  -2.007   0.0453 *
```

```
## gene_groupSynaptic    -0.48395    0.23106  -2.095   0.0367 *
## age                    0.13723    0.01480   9.271  < 2e-16 ***
## sex                    0.04839    0.08253   0.586   0.5579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.917 on 494 degrees of freedom
##   (47 observations deleted due to missingness)
## Multiple R-squared:  0.1594, Adjusted R-squared:  0.1492
## F-statistic: 15.62 on 6 and 494 DF,  p-value: < 2.2e-16
```

```r
# grouped bar chart for carrier status by cog domains

grouped_data <- grouped_types %>%
  group_by(is_carrier) %>%
  summarise(
    digit_symbol = mean(digitsymbol_composite, na.rm = TRUE),
    facial_memory = mean(facialmemory_z, na.rm = TRUE),
    cvlt = mean(cvlt_correct_z, na.rm = TRUE),
    cvlt_dprime = mean(cvlt_dprime_z, na.rm = TRUE),
    matrix_reasoning = mean(matrixreasoning_z, na.rm = TRUE),
    VWM = mean(vwm_composite, na.rm = TRUE),
    cog_score = mean(g, na.rm = TRUE),
  )

long_data <- grouped_data %>%
  pivot_longer(cols = c(digit_symbol, facial_memory, cvlt, cvlt_dprime, matrix_reasoning, VWM, cog_score
               names_to = "domain", values_to = "score")

ggplot(long_data, aes(x = domain, y = score, fill = factor(is_carrier))) +
         geom_bar(stat = "identity", position = "dodge") +
         labs(x = "Cognitive Domain", y = "Mean Z-Score", fill = "Carrier Status") +
         theme_minimal() +
         theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# radar chart for cog domain and z score by carrier status
## install.packages("fmsb")
##library(fmsb)

##radar_data <- grouped_data %>%
  ##column_to_rownames("is_carrier")

##radar_data <- rbind(
 ## rep(3, ncol(radar_data)),
 ## rep(-3, ncol(radar_data)),
 ## radar_data
#)

# heat map for cog domain and z score by gene group
```