

Analysing the data according to the mmse scale

Upload the libraries:

```
library(rstatix)
library(tidyr)
library(dplyr)
library(ggplot2)
library(patchwork)
library(ggpubr)
library(tidyverse)
library(ggpmisc)
```

Upload the data:

```
scales <- read.csv("scales_small_charlson.csv", header = FALSE,
  stringsAsFactors = FALSE)
lipids <- read.csv("lipids_small_charlson.csv", header = FALSE,
  stringsAsFactors = FALSE)
colnames(lipids) <- lipids[1, ]
lipids <- lipids[-1, ]
colnames(scales) <- scales[1, ]
scales <- scales[-1, ]
scales_lipids <- merge(scales, lipids, by = "MS ID")
```

Let's convert the data to the required type:

```
columns_to_convert <- which(names(scales_lipids) != "sex" & names(scales_lipids) !=
  "MS ID")
scales_lipids[, columns_to_convert] <- lapply(scales_lipids[,
  columns_to_convert], as.numeric)
```

We divide the patients' data into the group of patients with and without dementia according to the mmse scale:

Let's create a new dataset with the 50% worse values according to mmse. This group will include patients with dementia.

```
n_rows <- 495
sorted_mmse <- scales_lipids[order(scales_lipids$mmse), ]
worse_mmse <- sorted_mmse[1:n_rows, ]
worse_mmse$mmse_dementia <- rep("Dementia", 495)
```

Let's create a new dataset with the 50% best values. This group will include patients without dementia.

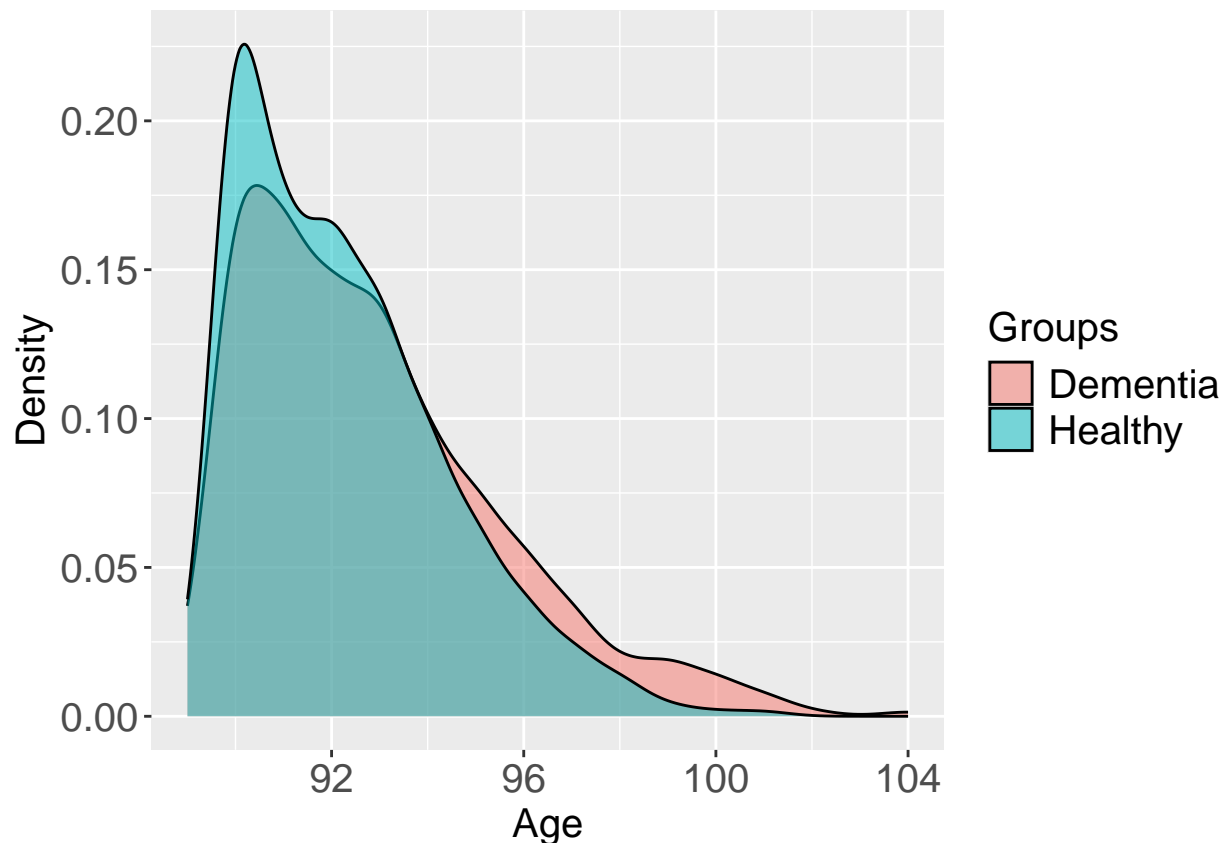
```
sorted_mmse_best <- scales_lipids[order(scales_lipids$mmse, decreasing = TRUE),
]
best_mmse <- sorted_mmse_best[1:n_rows, ]
best_mmse$mmse_dementia <- rep("Healthy", 495)
```

Then we combine both dataframes into one:

```
mmse <- rbind(worse_mmse, best_mmse)
mmse_long <- pivot_longer(mmse, cols = 10:230, names_to = "lipid_features",
  values_to = "values")
mmse_long$values <- as.numeric(mmse_long$values)
```

Let's look at the distribution of data in two groups:

```
ggplot(data = mmse, aes(x = age, fill = mmse_dementia)) + geom_density(alpha = 0.5) +
  labs(x = "Age", y = "Density", fill = "Groups") + theme(plot.title = element_text(hjust = 0.5,
    size = 15), axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15),
    legend.text = element_text(size = 15), legend.title = element_text(size = 15)) +
  theme(axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))
```



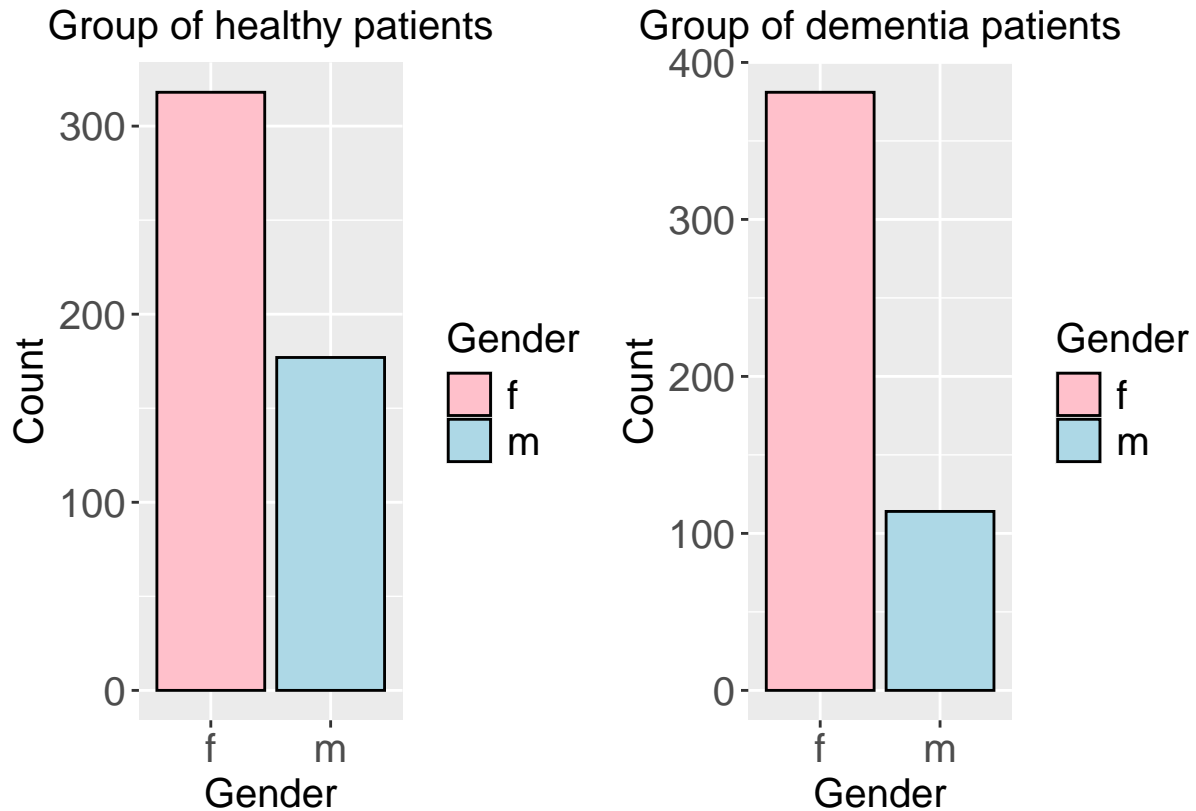
It can be seen that the age distribution of the data is similar in the dementia patient and healthy groups.

Let's look at the distribution of genders in the resulting patient groups

```
m1 <- ggplot(best_mmse, aes(x = factor(sex))) + geom_histogram(aes(fill = factor(sex)),
  color = "black", bins = 20, stat = "count") + scale_fill_manual(values = c(f = "pink",
  m = "lightblue")) + labs(x = "Gender", y = "Count", title = "Group of healthy patients",
  fill = "Gender") + theme(plot.title = element_text(hjust = 0.5,
  size = 15), axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15),
  legend.text = element_text(size = 15), legend.title = element_text(size = 15)) +
  theme(axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))

m2 <- ggplot(worse_mmse, aes(x = factor(sex))) + geom_histogram(aes(fill = factor(sex)),
  color = "black", bins = 20, stat = "count") + scale_fill_manual(values = c(f = "pink",
  m = "lightblue")) + labs(x = "Gender", y = "Count", title = "Group of dementia patients",
  fill = "Gender") + theme(plot.title = element_text(hjust = 0.5,
  size = 15), axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15),
  legend.text = element_text(size = 15), legend.title = element_text(size = 15)) +
  theme(axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))

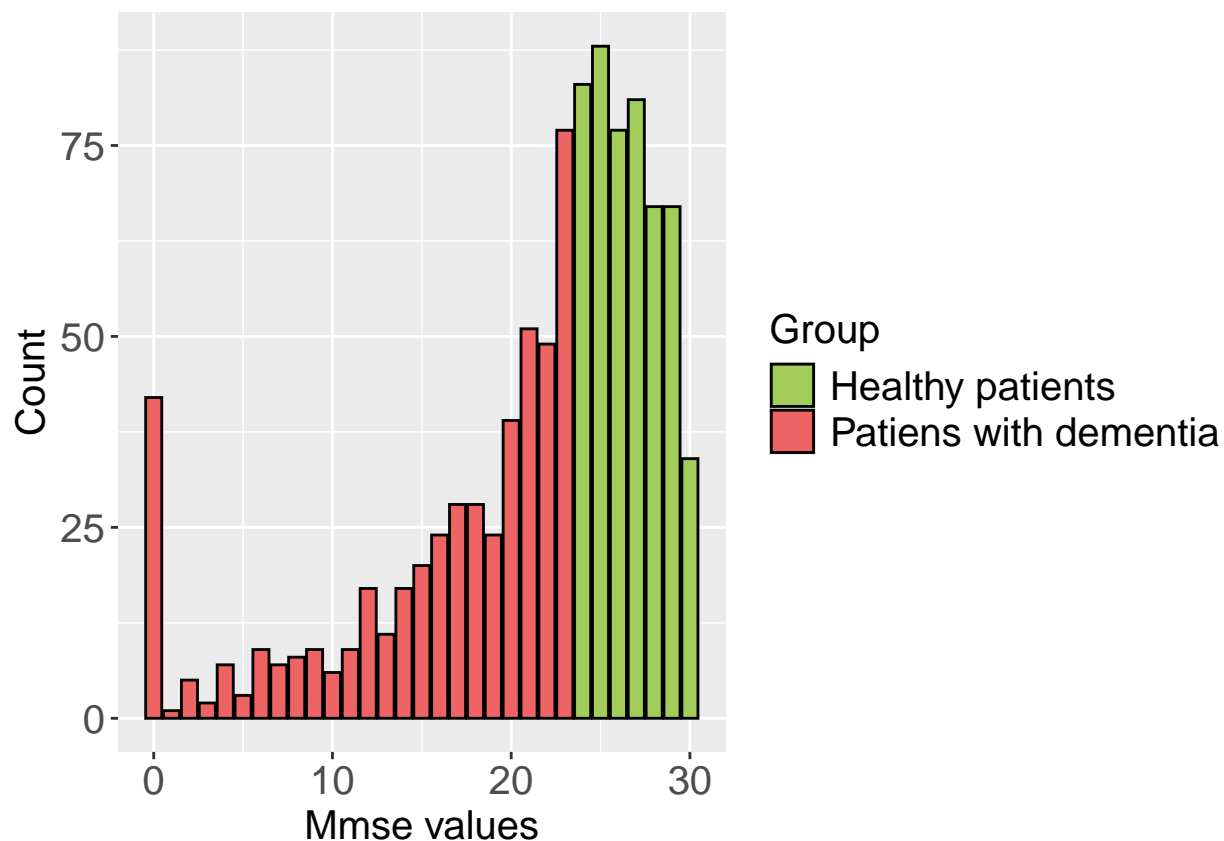
m1 + m2
```



There are more men in the group of healthy patients than in the group of patients with dementia

Let's look at the distribution of mmse scale values in the resulting groups:

```
ggplot(scales_lipids, aes(x = mmse)) + geom_histogram(aes(fill = ifelse(mmse %in%
  best_mmse$mmse, "Healthy patients", ifelse(mmse %in% worse_mmse$mmse,
    "Patients with dementia", "Data not included \n in analysis"))),
  color = "black", bins = 20, stat = "count") + scale_fill_manual(values = c("darkolivegreen3",
    "#EE6363")) + labs(x = "Mmse values", y = "Count", fill = "Group") +
  theme(axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15),
    legend.text = element_text(size = 15), legend.title = element_text(size = 15)) +
  theme(axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))
```



Let's perform t-test between the lipids of the 50% best and 50% worst samples according to the mmse scale

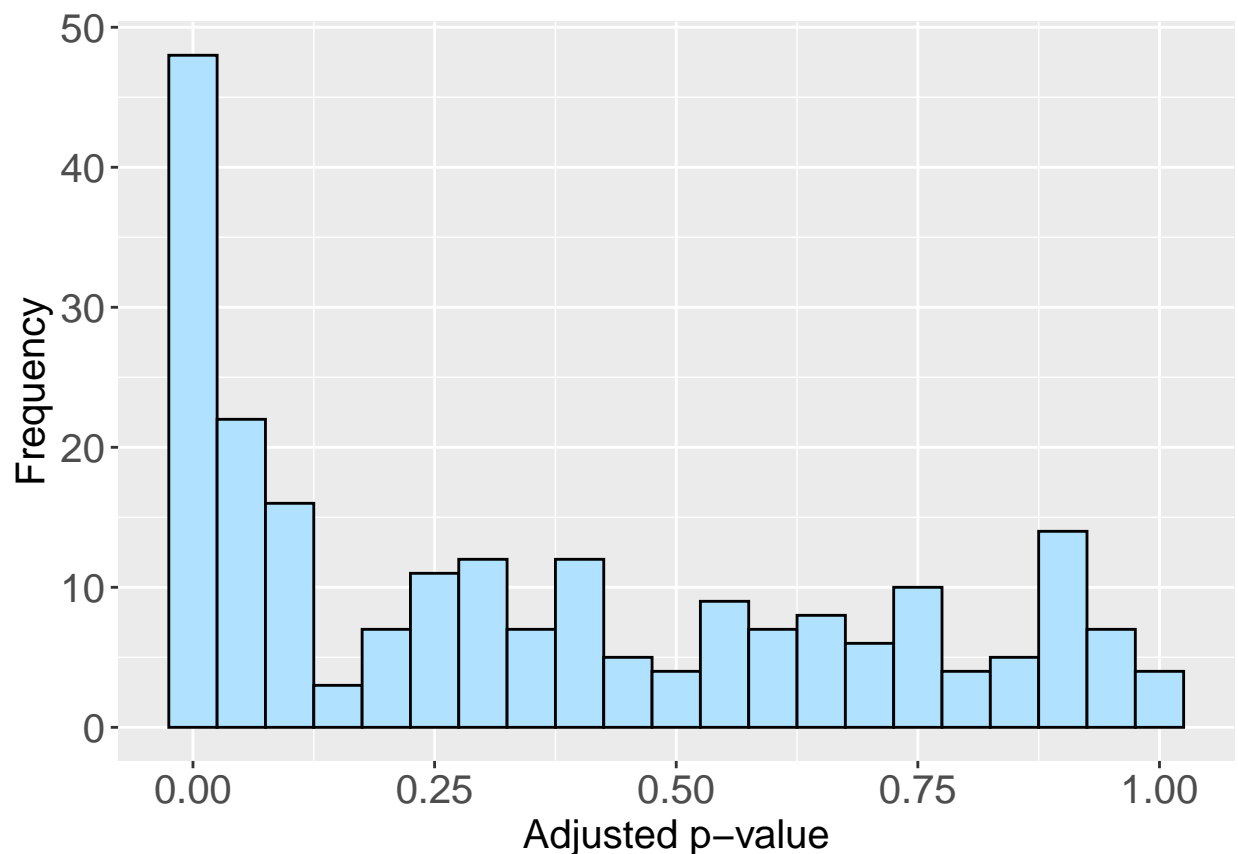
```
stat.test_mmse <- mmse_long %>%
  group_by(lipid_features) %>%
  t_test(values ~ mmse_dementia) %>%
  adjust_pvalue(method = "BH") %>%
  add_significance()
print(stat.test_mmse)
```

```
## # A tibble: 221 x 11
##   lipid_features .y. group1 group2 n1 n2 statistic df p
##   <chr>          <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>
## 1 CAR 10:1      values Dementia Healthy 495 495 -2.71 988. 6.84e-3
## 2 CAR 18:1      values Dementia Healthy 495 495 -1.30 988. 1.94e-1
## 3 CAR 18:2      values Dementia Healthy 495 495 -3.60 987. 3.4 e-4
## 4 CE 14:0       values Dementia Healthy 495 495 0.757 987. 4.49e-1
## 5 CE 15:0       values Dementia Healthy 495 495 1.17 928. 2.43e-1
## 6 CE 16:1       values Dementia Healthy 495 495 5.84 988. 7.02e-9
## 7 CE 17:1       values Dementia Healthy 495 495 2.60 987. 9.45e-3
## 8 CE 18:2       values Dementia Healthy 495 495 0.546 983. 5.85e-1
## 9 CE 18:3       values Dementia Healthy 495 495 0.622 988. 5.34e-1
```

```
## 10 CE 19:1          values Dementia Healthy    495    495        1.68    949.    9.34e-2
## # i 211 more rows
## # i 2 more variables: p.adj <dbl>, p.adj.signif <chr>
```

Let's plot the p-value distribution:

```
ggplot(data = stat.test_mmse, aes(x = p.adj)) + geom_histogram(binwidth = 0.05,
  fill = "#B0E2FF", color = "black", bins = 15) + labs(x = "Adjusted p-value",
  y = "Frequency") + theme(axis.title.x = element_text(size = 15),
  axis.title.y = element_text(size = 15), legend.text = element_text(size = 15),
  legend.title = element_text(size = 15)) + theme(axis.text.x = element_text(size = 15),
  axis.text.y = element_text(size = 15))
```



Based on the distribution of p-value values, it can be seen that the patient groups are significantly different from each other.

Let's look separately at lipids significantly different between patients with and without dementia

```
stat.test_mmse_significant <- stat.test_mmse[stat.test_mmse$p.adj <=
  0.05, ]
print(stat.test_mmse_significant)
```

```
## # A tibble: 63 x 11
##   lipid_features .y. group1 group2 n1 n2 statistic df p
##   <chr>         <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>
## 1 CAR 10:1     values Dementia Healthy 495 495 -2.71 988. 6.84e-3
## 2 CAR 18:2     values Dementia Healthy 495 495 -3.60 987. 3.4 e-4
## 3 CE 16:1      values Dementia Healthy 495 495 5.84 988. 7.02e-9
## 4 CE 17:1      values Dementia Healthy 495 495 2.60 987. 9.45e-3
## 5 CE 19:5      values Dementia Healthy 495 495 3.56 987. 3.95e-4
## 6 CE 20:5      values Dementia Healthy 495 495 -4.33 987. 1.62e-5
## 7 CE 21:1      values Dementia Healthy 495 495 3.23 964. 1.29e-3
## 8 CE 22:6      values Dementia Healthy 495 495 -2.81 986. 5.06e-3
## 9 Cer 34:1     values Dementia Healthy 495 495 4.45 984. 9.61e-6
## 10 LPC 14:0    values Dementia Healthy 495 495 -2.65 982. 8.27e-3
## # i 53 more rows
## # i 2 more variables: p.adj <dbl>, p.adj.signif <chr>
```

Let's parse the data on lipids:

```
stat.test_mmse$lipid_features <- gsub("LPC 0-(\\d+):(\\d+)",
  "LPC-0 \\1:\\2", stat.test_mmse$lipid_features)
stat.test_mmse_sep <- separate(stat.test_mmse, lipid_features,
  into = c("Class", "Other"), sep = " ", remove = FALSE)
stat.test_mmse_sep <- separate(stat.test_mmse_sep, Other, into = c("Chain_length",
  "Double_bounds"), sep = ":")

stat.test_mmse_significant$lipid_features <- gsub("LPC 0-(\\d+):(\\d+)",
  "LPC-0 \\1:\\2", stat.test_mmse_significant$lipid_features)
stat.test_mmse_signif_sep <- separate(stat.test_mmse_significant,
  lipid_features, into = c("Class", "Other"), sep = " ", remove = FALSE)
stat.test_mmse_signif_sep <- separate(stat.test_mmse_signif_sep,
  Other, into = c("Chain_length", "Double_bounds"), sep = ":")
```

Let's count the percentage of significant lipids by class:

```
stat.test_mmse_all_lipids <- as.data.frame(table(stat.test_mmse_sep$Class))
stat.test_mmse_all_lipid_signif <- as.data.frame(table(stat.test_mmse_signif_sep$Class))
lipids_stat.test_mmse <- merge(stat.test_mmse_all_lipids, stat.test_mmse_all_lipid_signif,
  by = "Var1")
lipids_stat.test_mmse$percentage <- (lipids_stat.test_mmse$Freq.y/lipids_stat.test_mmse$Freq.x) *
  100
colnames(lipids_stat.test_mmse)[1] <- "Class"
```

Let's perform enrichment analysis:

```
results_enrichment <- data.frame(Class = character(), p_value = numeric(),
  adjusted_p_value = numeric(), stringsAsFactors = FALSE)

lipid_classes <- unique(stat.test_mmse_signif_sep$Class)
```

```

for (lipid_class in lipid_classes) {
  q <- sum(stat.test_mmse_signif_sep$Class == lipid_class)
  m <- nrow(stat.test_mmse_signif_sep)
  n <- nrow(stat.test_mmse_sep) - m
  k <- sum(stat.test_mmse_sep$Class == lipid_class)

  p_value <- phyper(q - 1, m, n, k, lower.tail = FALSE, log.p = FALSE)

  adjusted_p_value <- p.adjust(p_value, method = "bonferroni")

  results_enrichment <- rbind(results_enrichment, data.frame(Class = lipid_class,
    p_value = p_value, adjusted_p_value = adjusted_p_value))
}

print(results_enrichment)

```

```

##      Class      p_value adjusted_p_value
## 1     CAR 0.1962578115      0.1962578115
## 2      CE 0.2291220505      0.2291220505
## 3     Cer 0.9679279659      0.9679279659
## 4     LPC 0.5630020508      0.5630020508
## 5      PC 0.0670781464      0.0670781464
## 6   PC-O 0.6395617745      0.6395617745
## 7      PE 0.3649453253      0.3649453253
## 8   PE-P 0.0004496393      0.0004496393
## 9      PI 0.2850678733      0.2850678733
## 10     SM 0.9010157712      0.9010157712
## 11     TG 0.9920125029      0.9920125029

```

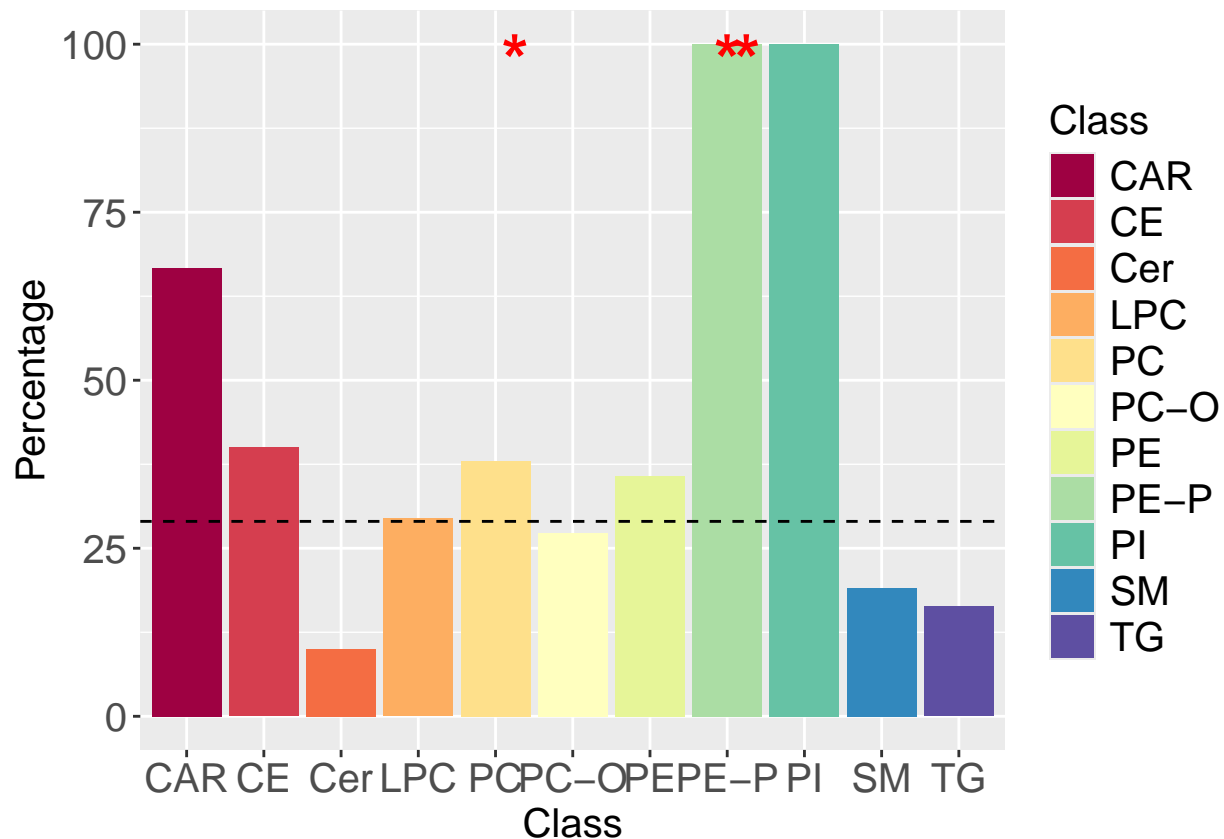
Plot the results:

```

significant_results <- results_enrichment[results_enrichment$adjusted_p_value <
  0.05, ]
significant_results2 <- results_enrichment[results_enrichment$adjusted_p_value <
  0.1, ]

ggplot(lipids_stat.test_mmse, aes(x = Class, y = percentage,
  fill = Class)) + geom_bar(stat = "identity") + scale_fill_brewer(palette = "Spectral") +
  labs(x = "Class", y = "Percentage") + theme(axis.title.x = element_text(size = 15),
  axis.title.y = element_text(size = 15), legend.text = element_text(size = 15),
  legend.title = element_text(size = 15), axis.text.x = element_text(size = 15),
  axis.text.y = element_text(size = 15)) + geom_text(data = significant_results,
  aes(x = Class, y = 0, label = paste0("*")), vjust = -12.7,
  size = 9, color = "red", fontface = "bold") + geom_text(data = significant_results2,
  aes(x = Class, y = 0, label = paste0(" *")), vjust = -12.7,
  size = 9, color = "red", fontface = "bold") + geom_hline(yintercept = 29,
  linetype = "dashed", color = "black")

```

It can be seen that lipids differing significantly between the patient groups belong to different classes. In addition, compared to the expected lipid levels, there is enrichment of such lipid classes as PC and PE-P.

Let's explore differences in lipid content between groups:

Prepare the data:

```
mmse_dementia <- mmse_long[mmse_long$mmse_dementia == "Dementia",
]
mmse_nodementia <- mmse_long[mmse_long$mmse_dementia == "Healthy",
]

mmse_dementia_mean <- mmse_dementia %>%
  group_by(lipid_features) %>%
  summarise(mean_dementia = mean(values))

mmse_nodementia_mean <- mmse_nodementia %>%
  group_by(lipid_features) %>%
  summarise(mean_nodementia = mean(values))

mean_mmse <- merge(mmse_dementia_mean, mmse_nodementia_mean,
  by = "lipid_features")
```

```

mean_mmse_filtered <- mean_mmse %>%
  filter(lipid_features %in% stat.test_mmse_significant$lipid_features)

mean_mmse_filtered$difference <- mean_mmse_filtered$mean_dementia -
  mean_mmse_filtered$mean_nodementia

mean_mmse_filtered$Sign <- ifelse(mean_mmse_filtered$difference >
  0, "Positive", "Negative")

mean_mmse_filtered$Count <- 1

mean_mmse_filtered$Value <- ifelse(mean_mmse_filtered$Sign ==
  "Negative", -mean_mmse_filtered$Count, mean_mmse_filtered$Count)

mean_mmse_filtered$lipid_features <- gsub("LPC 0-(\\d+):(\\d+)",
  "LPC-0 \\1:\\2", mean_mmse_filtered$lipid_features)
mean_mmse_filtered <- separate(mean_mmse_filtered, lipid_features,
  into = c("Class", "Other"), sep = " ", remove = FALSE)
mean_mmse_filtered <- separate(mean_mmse_filtered, Other, into = c("Chain_length",
  "Double_bounds"), sep = ":")

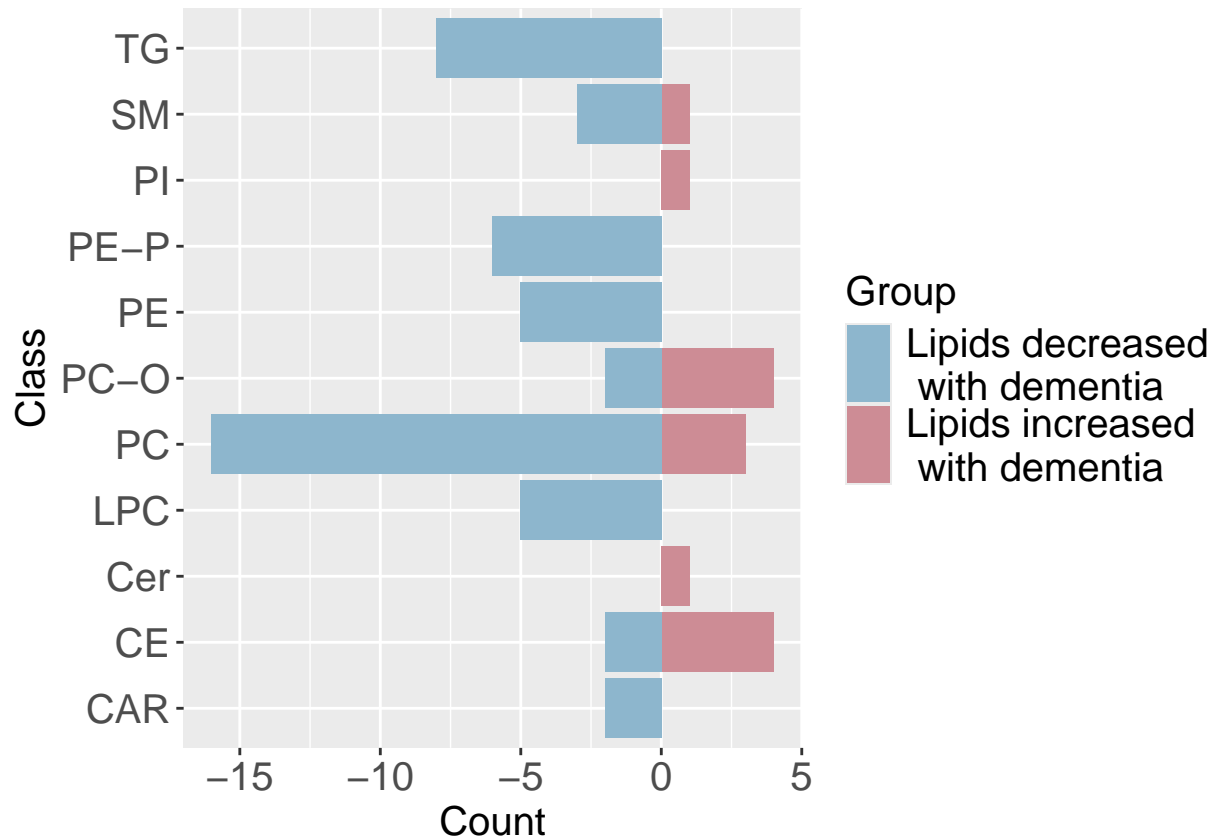
```

Plot the results:

```

ggplot(mean_mmse_filtered, aes(x = Class, y = Value, fill = Sign)) +
  geom_bar(stat = "identity") + scale_fill_manual(values = c(Positive = "#CD8C95",
  Negative = "#8DB6CD"), labels = c(Positive = "Lipids increased \n with dementia",
  Negative = "Lipids decreased \n with dementia")) + coord_flip() +
  labs(x = "Class", y = "Count", fill = "Group") + theme(axis.title.x = element_text(size = 15),
  axis.title.y = element_text(size = 15), legend.text = element_text(size = 15),
  legend.title = element_text(size = 15)) + theme(axis.text.x = element_text(size = 15),
  axis.text.y = element_text(size = 15))

```



We can see that the majority of lipids are reduced in patients with dementia.

Let's look at the number of double bonds in significant lipids and their differences in the patient groups:

Count the fold change:

```
best_mmse_long <- pivot_longer(best_mmse, cols = 10:230, names_to = "lipid_features",
  values_to = "values")
worse_mmse_long <- pivot_longer(worse_mmse, cols = 10:230, names_to = "lipid_features",
  values_to = "values")

best_mmse_mean <- best_mmse_long %>%
  group_by(lipid_features) %>%
  summarise(mean_mmse_high = mean(values))

worse_mmse_mean <- worse_mmse_long %>%
  group_by(lipid_features) %>%
  summarise(mean_mmse_low = mean(values))

mean_mmse <- merge(best_mmse_mean, worse_mmse_mean, by = "lipid_features")
mean_mmse$FC_mmse <- mean_mmse$mean_mmse_low - mean_mmse$mean_mmse_high
```

Plot the results:

```
unique_classes <- unique(stat.test_mmse_sep$Class)
stat.test_mmse_sep$lipid_features <- gsub("LPC 0-(\\d+):(\\d+)",
  "LPC-0 \\1:\\2", stat.test_mmse_sep$lipid_features)
mean_mmse$lipid_features <- gsub("LPC 0-(\\d+):(\\d+)", "LPC-0 \\1:\\2",
  mean_mmse$lipid_features)

for (current_class in unique_classes) {

  stat.test_mmse_lipid <- stat.test_mmse_sep[stat.test_mmse_sep$Class ==
    current_class, ]

  merged_bounds_FC <- merge(mean_mmse, stat.test_mmse_lipid,
    by = "lipid_features")

  if (current_class %in% c("CAR", "CE", "LPC", "LPE", "LPC-0")) {
    merged_bounds_FC$bounds_correct <- round(as.numeric(as.character(merged_bounds_FC$Double_bounds.
      1)
  } else if (current_class %in% c("SM", "Cer", "PC", "PE", "DAG",
    "PC-0", "PE-P")) {
    merged_bounds_FC$bounds_correct <- round(as.numeric(as.character(merged_bounds_FC$Double_bounds.
      1)
  } else if (current_class == "TG") {
    merged_bounds_FC$bounds_correct <- round(as.numeric(as.character(merged_bounds_FC$Double_bounds.
      1)
  } else if (current_class == "PI") {
    merged_bounds_FC$bounds_correct <- round(as.numeric(as.character(merged_bounds_FC$Double_bounds.
      2)
  } else {

    next
  }

  merged_bounds_FC$bounds_correct <- as.character(merged_bounds_FC$bounds_correct)

  if (nrow(merged_bounds_FC) >= 2) {

    model <- lm(FC_mmse ~ bounds_correct, data = merged_bounds_FC)

    summary_result <- summary(model)

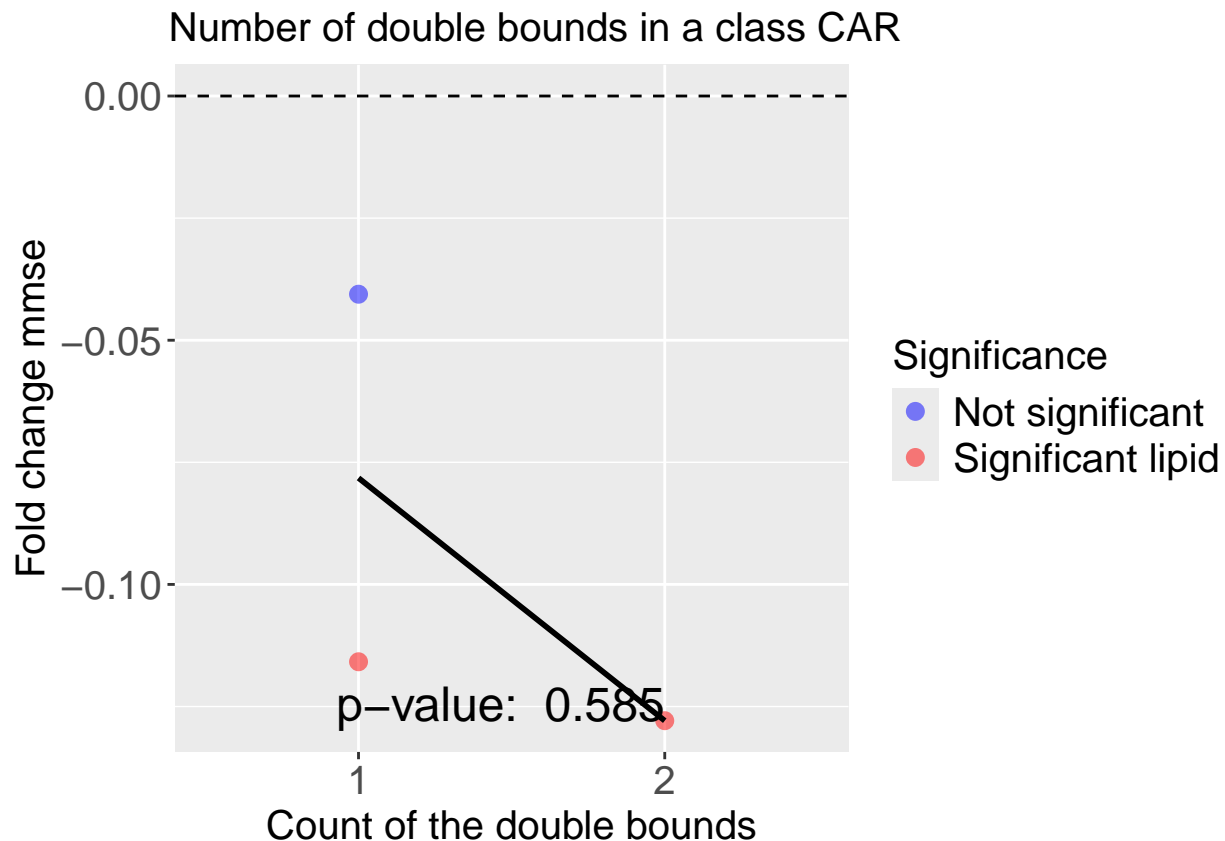
    f <- summary_result$fstatistic
    p_value <- pf(f[1], f[2], f[3], lower.tail = F)

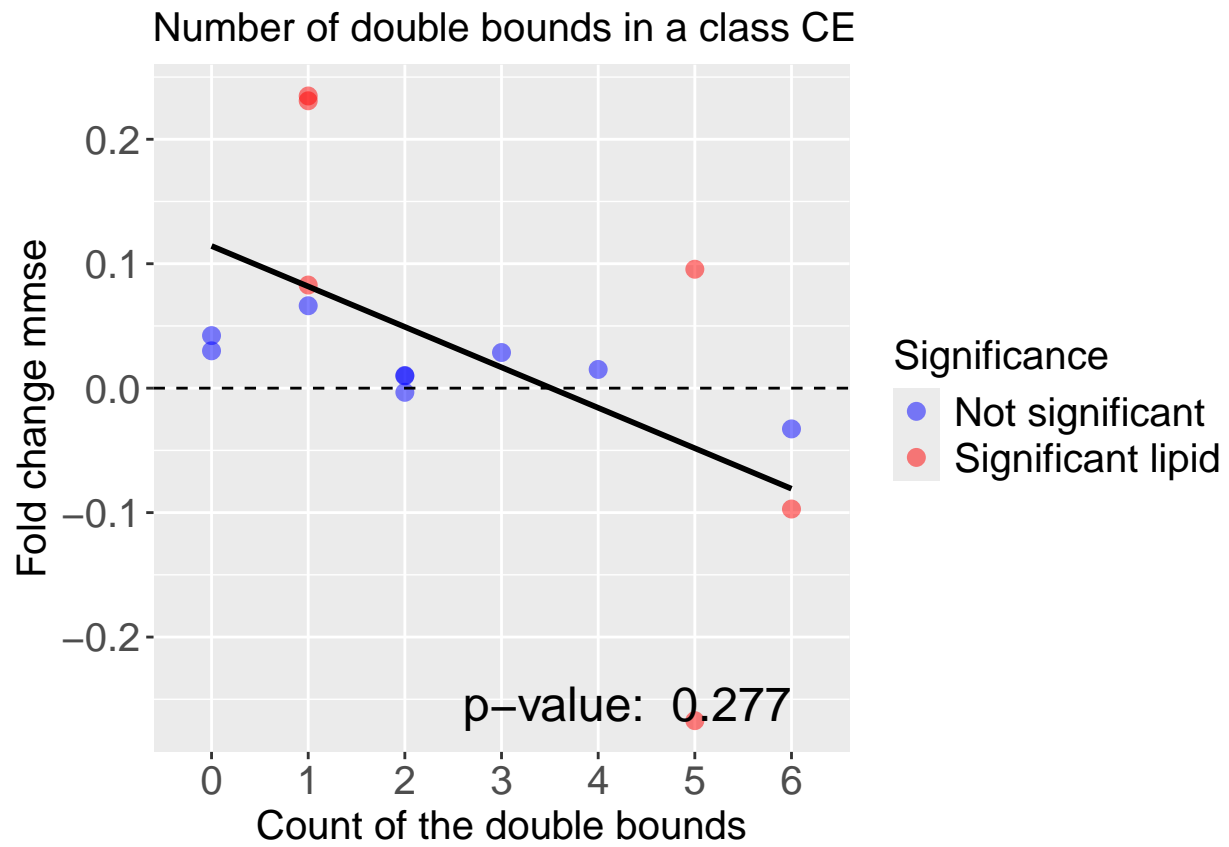
    plot <- ggplot(merged_bounds_FC, aes(x = bounds_correct,
      y = FC_mmse, color = ifelse(lipid_features %in% stat.test_mmse_signif_sep$lipid_features,
        "TRUE", "FALSE"))) + geom_point(shape = 16, size = 3,
      alpha = 0.5) + labs(x = "Count of the double bounds",
      y = "Fold change mmse", title = paste("Number of double bounds in a class",
        current_class)) + geom_vline(xintercept = 0,
      linetype = "dashed", color = "black") + geom_hline(yintercept = 0,
```

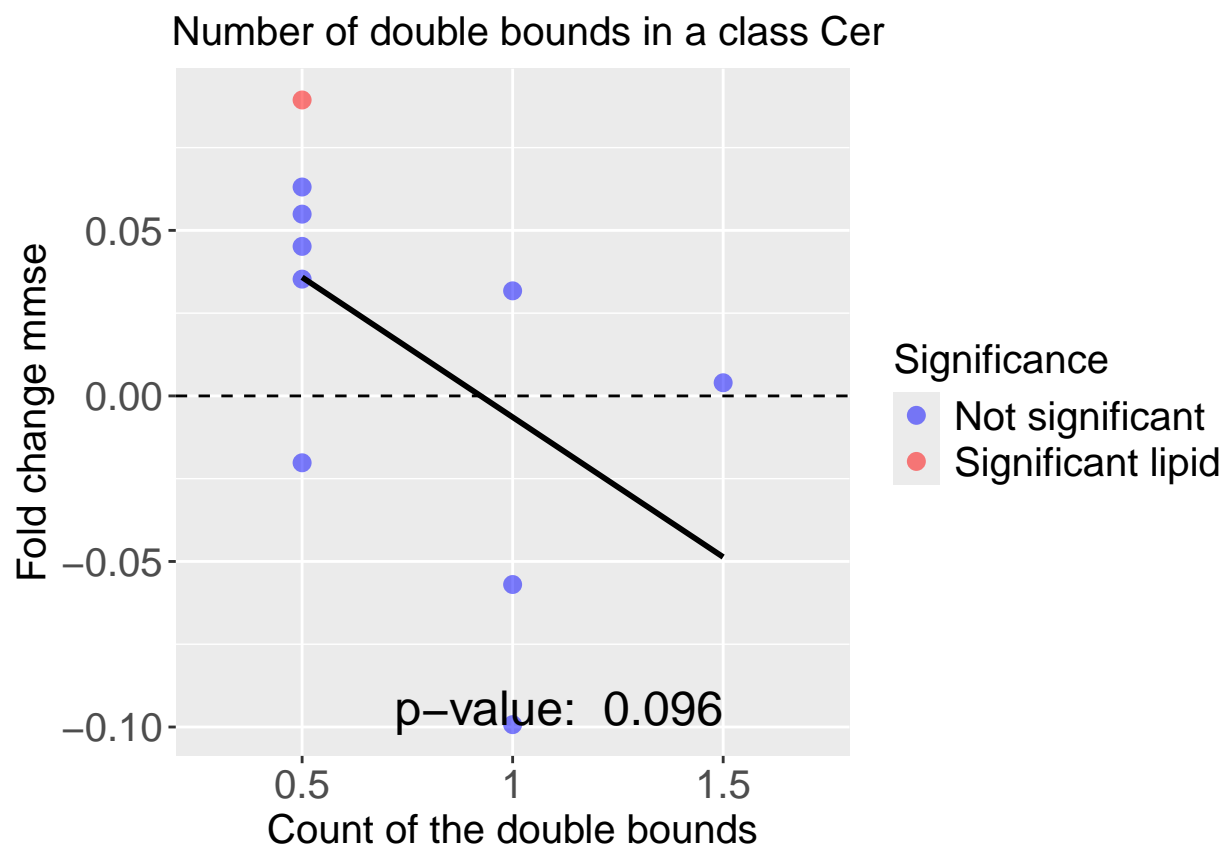
```

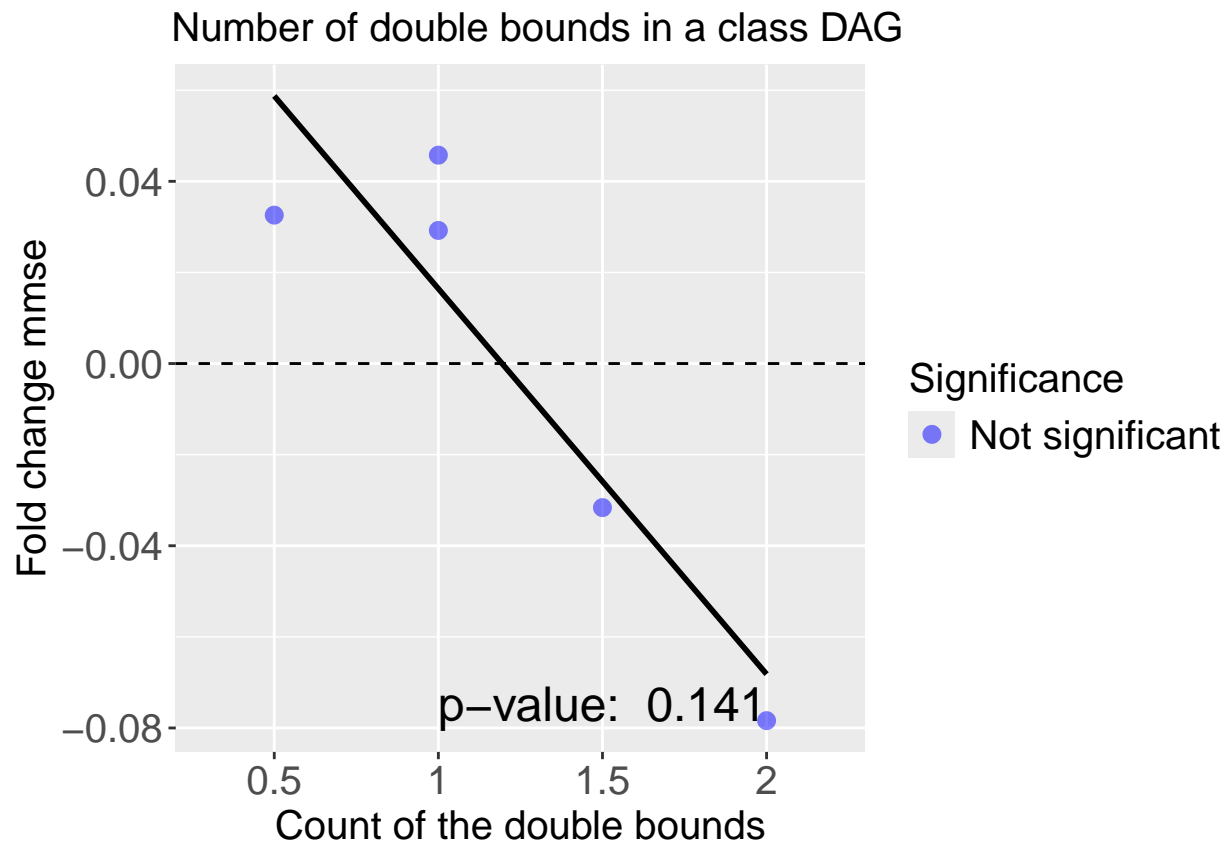
linetype = "dashed", color = "black") + scale_color_manual(values = c(`TRUE` = "red",
`FALSE` = "blue"), name = "Significance", labels = if (all(merged_bounds_FC$lipid_features %
stat.test_mmse_signif_sep$lipid_features))
c("Significant lipid", "Significant lipid") else c("Not significant", "Significant lipid"))
size = 15), axis.title.x = element_text(size = 15),
axis.title.y = element_text(size = 15), legend.text = element_text(size = 15),
legend.title = element_text(size = 15), axis.text.x = element_text(size = 15),
axis.text.y = element_text(size = 15)) + geom_smooth(aes(group = 1),
method = "lm", se = FALSE, color = "black", show.legend = FALSE) +
stat_regline_equation(label.x = "left", label.y = "bottom",
show.legend = FALSE) + annotate("text", x = max(merged_bounds_FC$bounds_correct),
y = min(merged_bounds_FC$FC_mmse), label = paste("p-value: ",
formatC(p_value, digits = 3, format = "f")),
hjust = 1, vjust = 0, size = 6.5, color = "black")
print(plot)
}
}

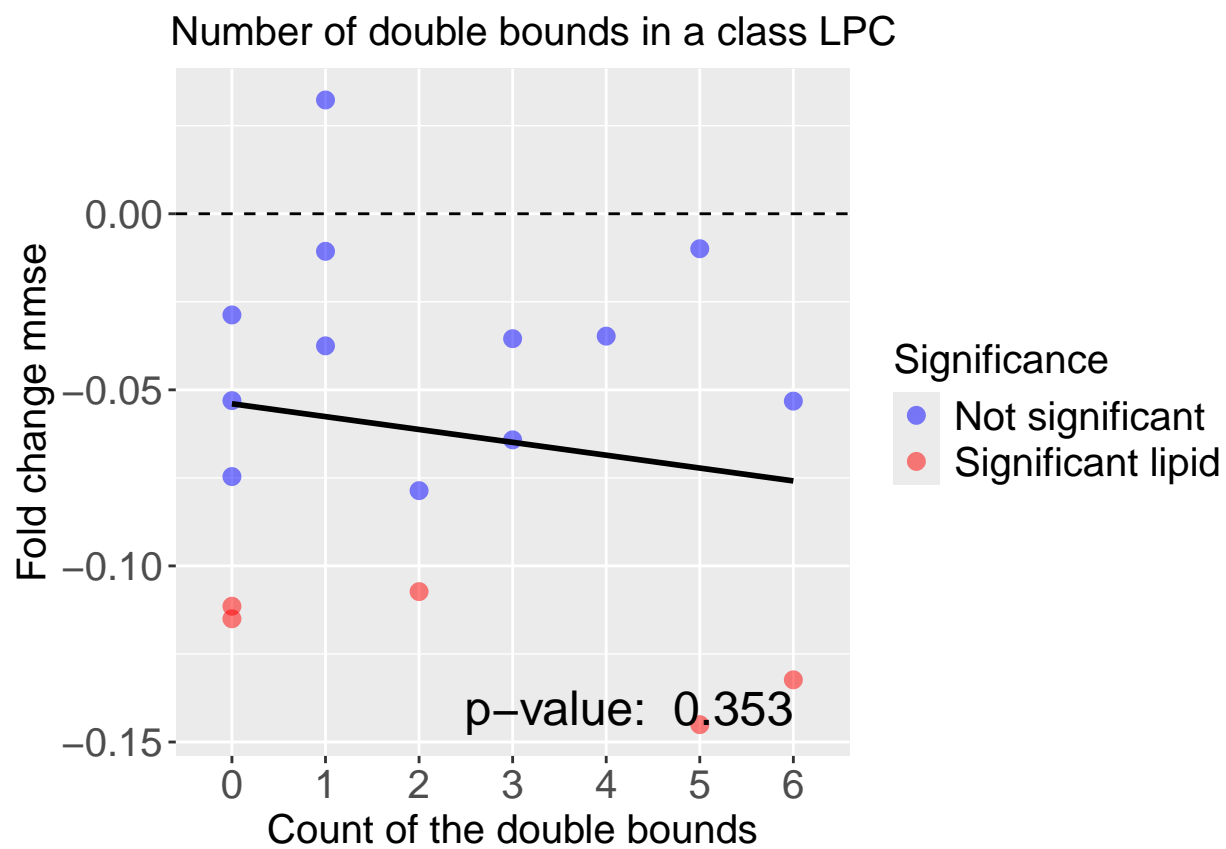
```

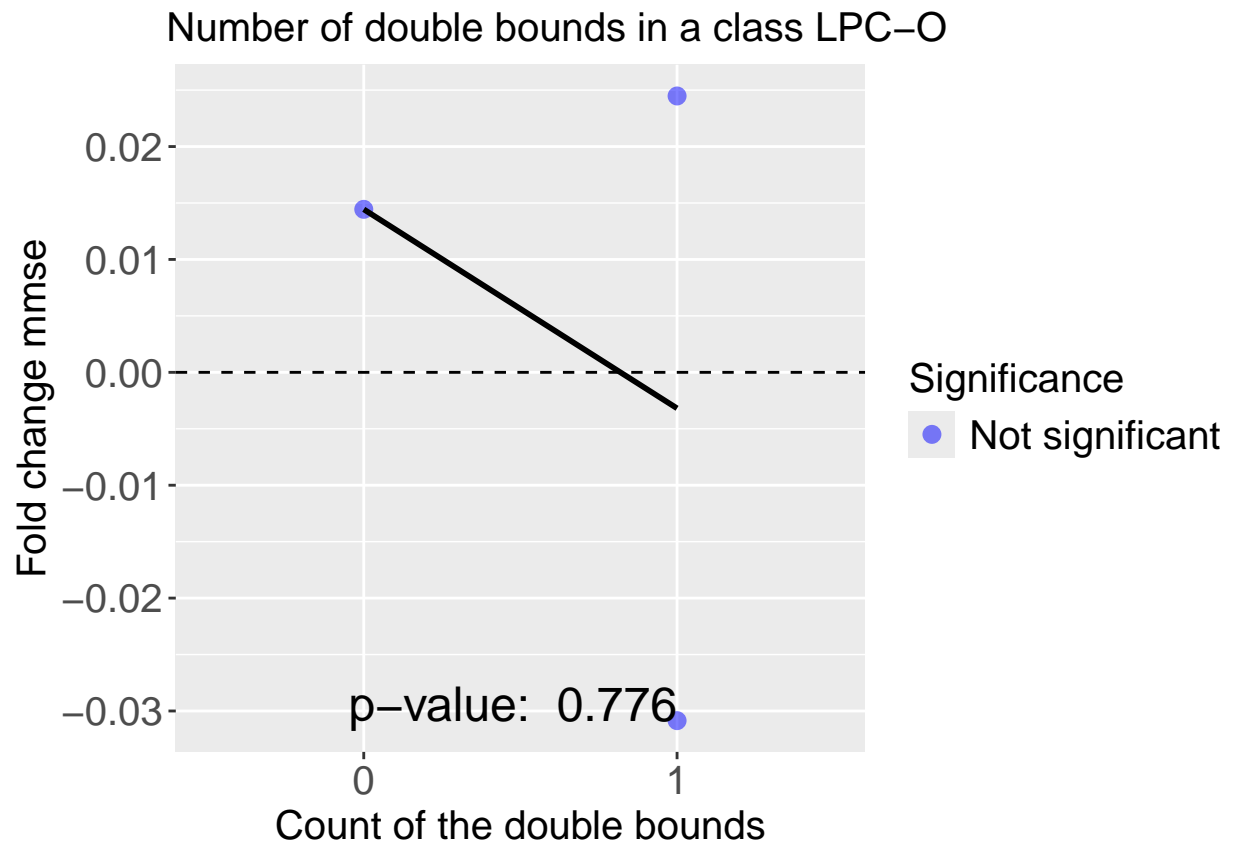


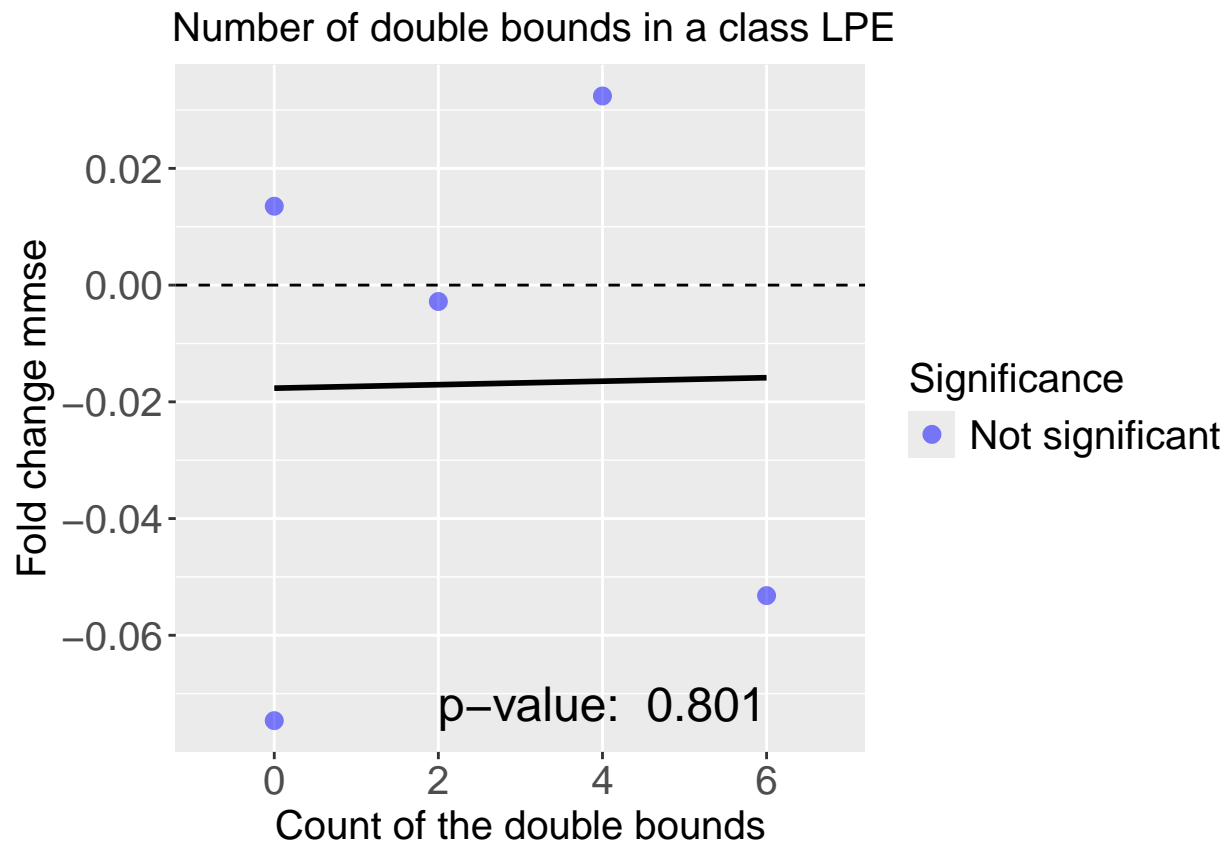


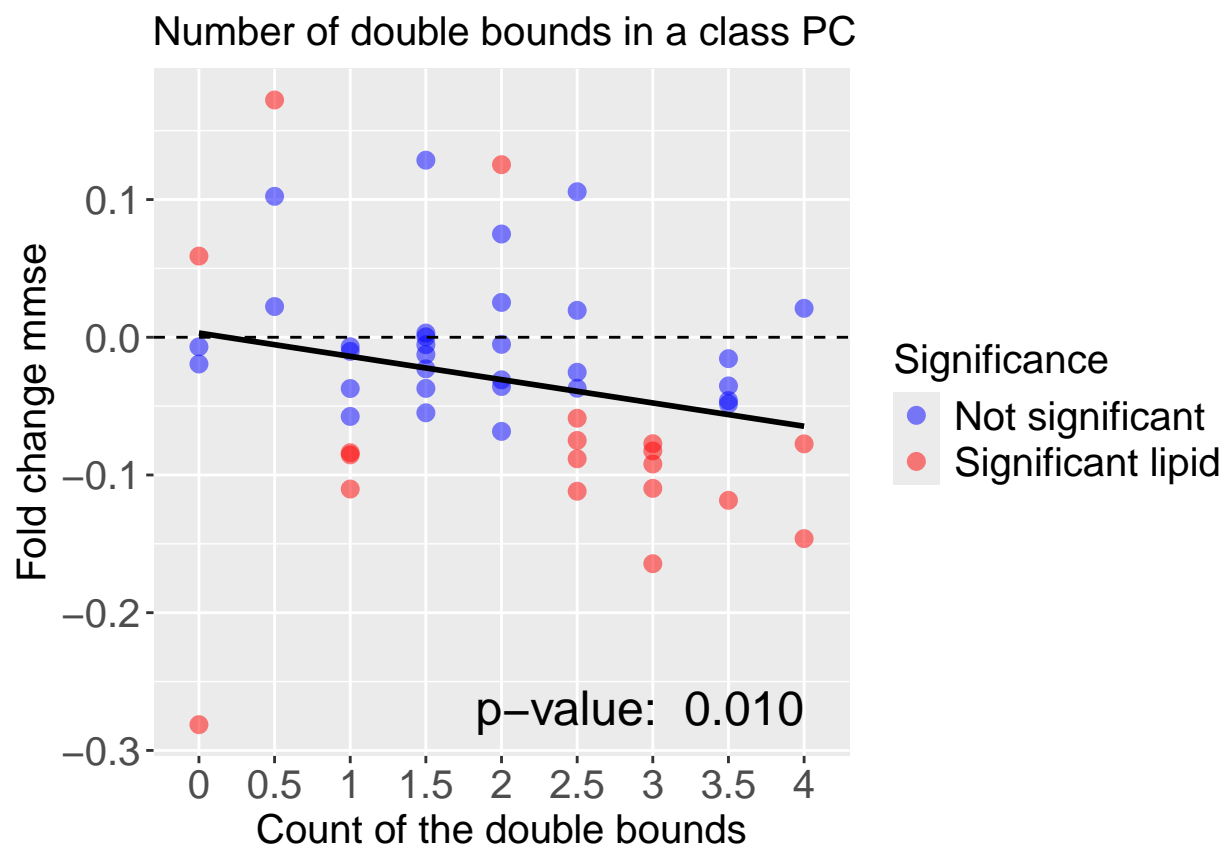


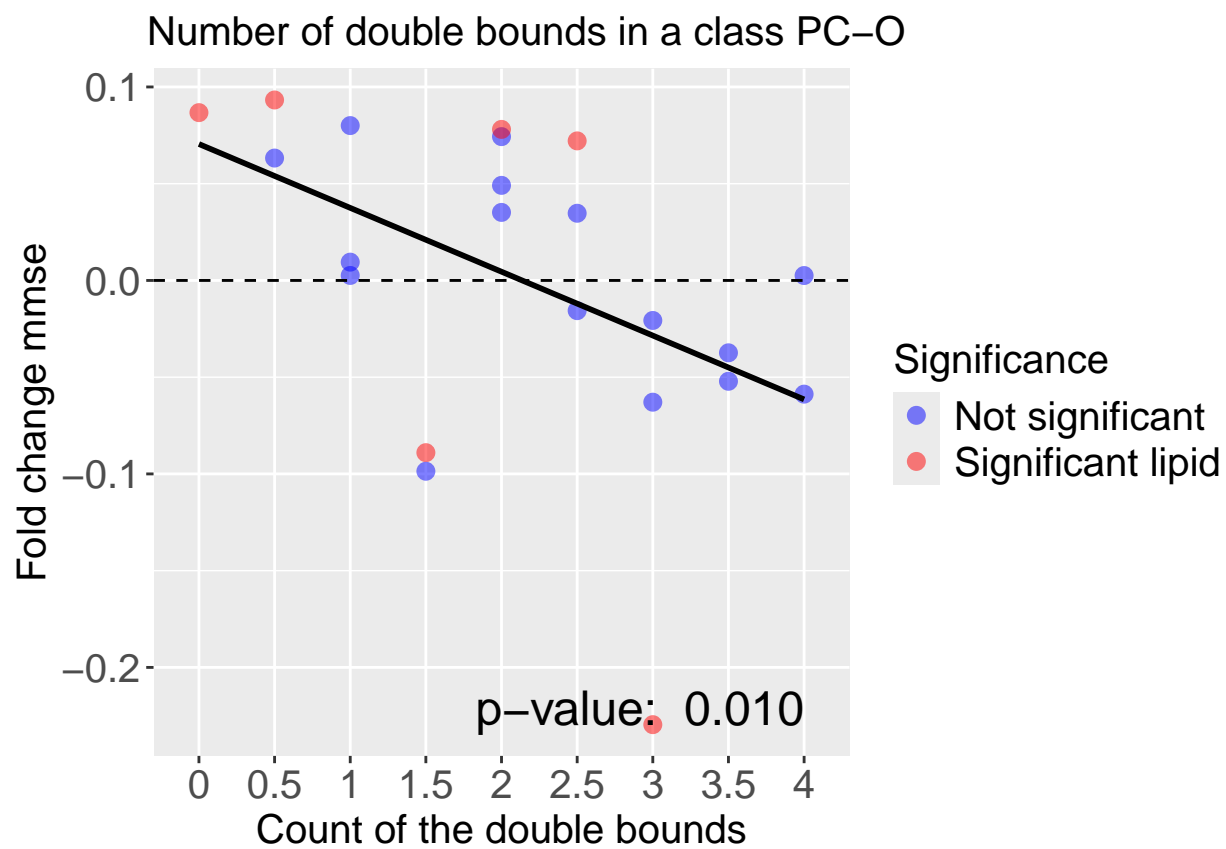


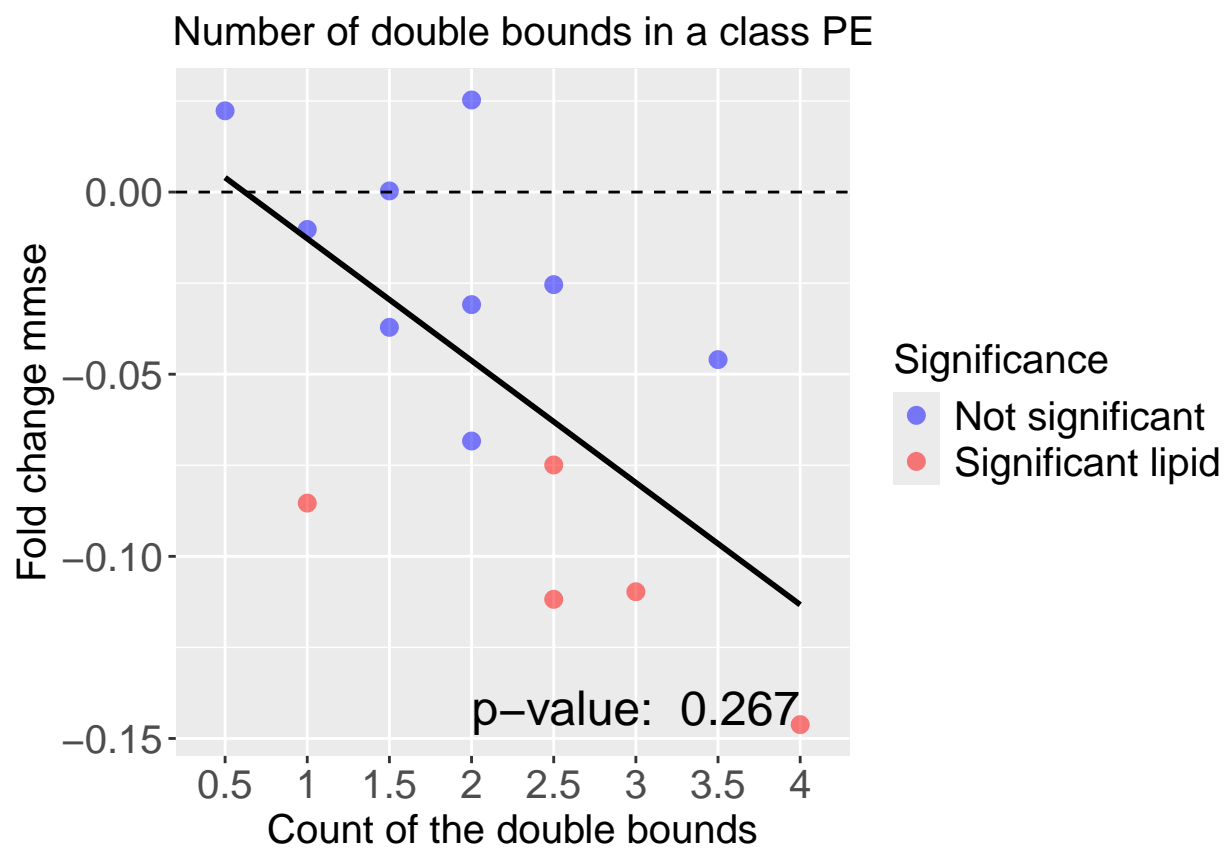


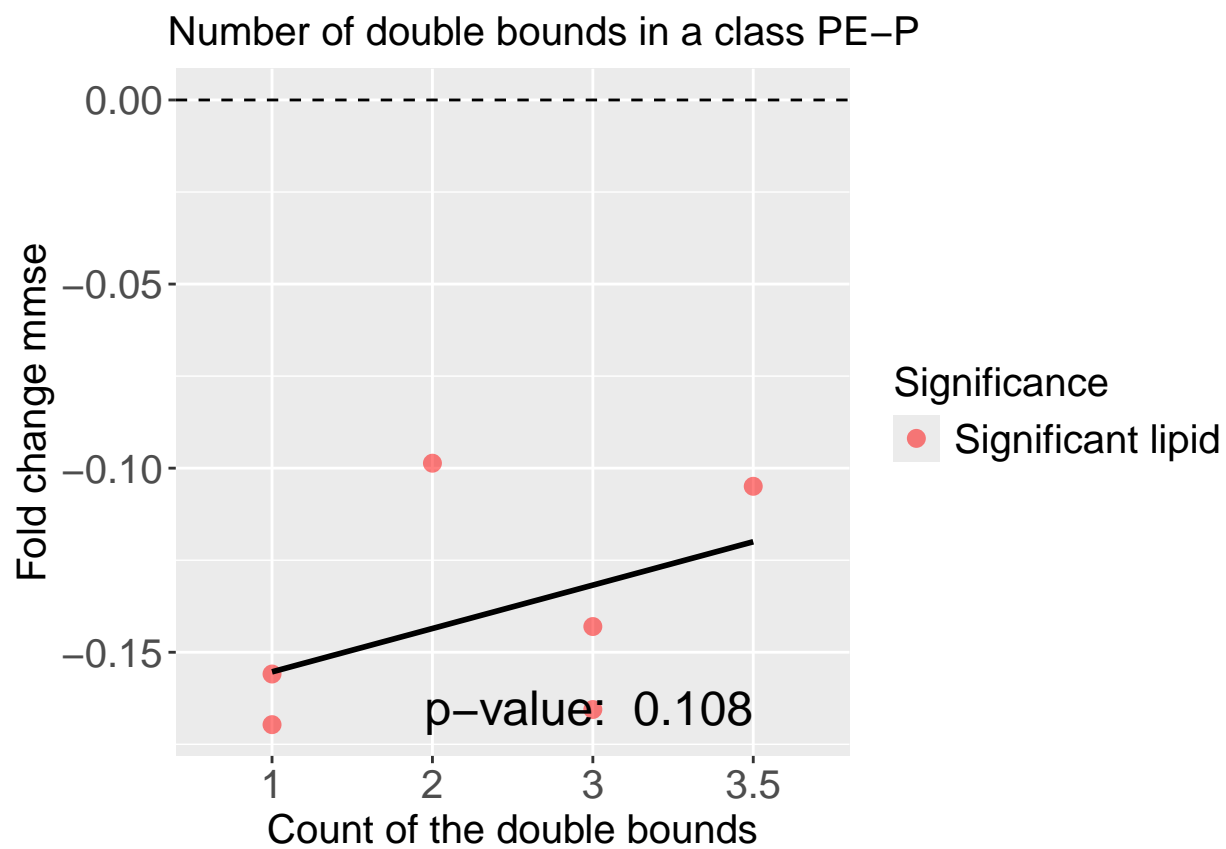


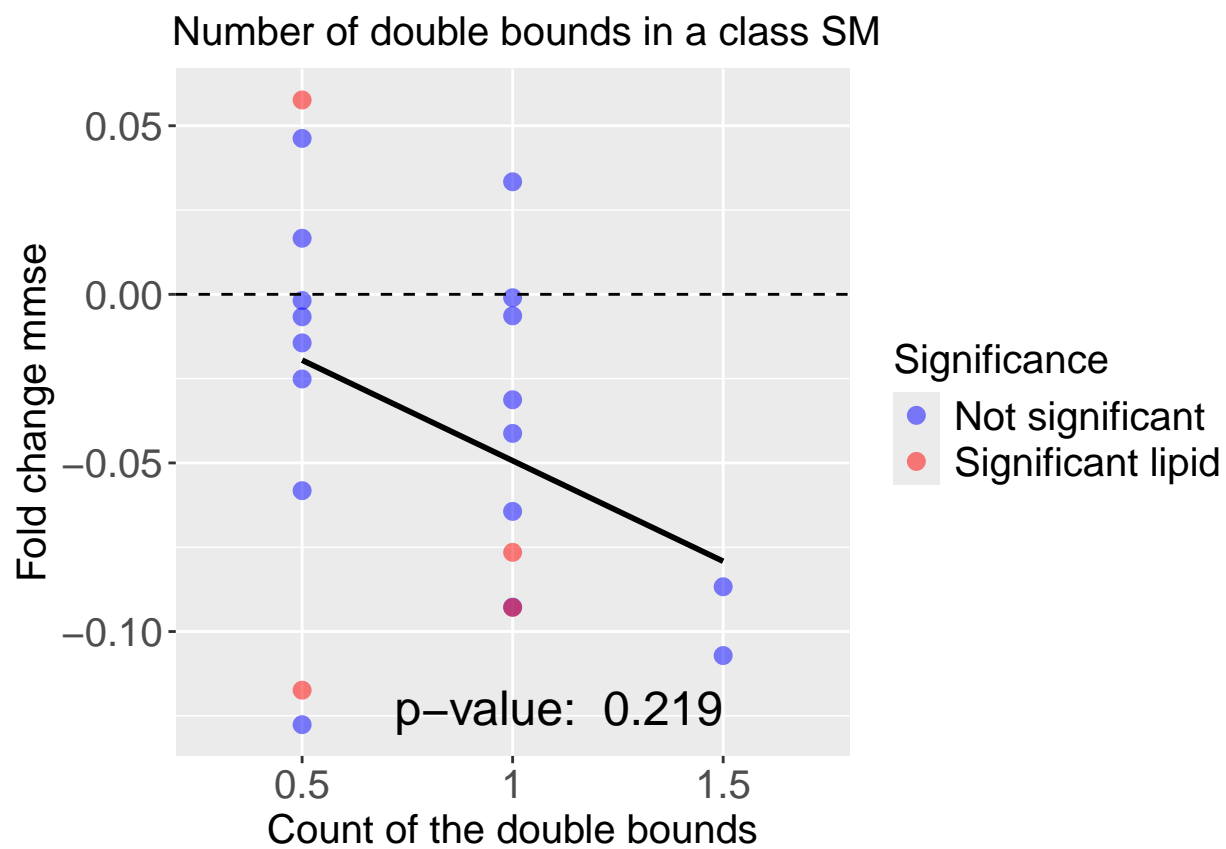














Based on these plots, we can conclude that the level of polyunsaturated lipids is reduced among patients with dementia

Let's look at the number of double bonds as well as chain lengths in the significant lipids and their differences in the patient groups:

```
unique_classes <- unique(stat.test_mmse_sep$Class)
for (current_class in unique_classes) {

  stat.test_mmse_lipid <- stat.test_mmse_sep[stat.test_mmse_sep$Class ==
    current_class, ]
  merged_bounds_FC <- merge(mean_mmse, stat.test_mmse_lipid,
    by = "lipid_features")

  stat.test_mmse_lipid_signif <- stat.test_mmse_signif_sep[stat.test_mmse_signif_sep$Class ==
    current_class, ]
  merged_bounds_FC_signif <- merge(mean_mmse, stat.test_mmse_lipid_signif,
    by = "lipid_features")

  data1 <- merged_bounds_FC
  data2 <- merged_bounds_FC_signif

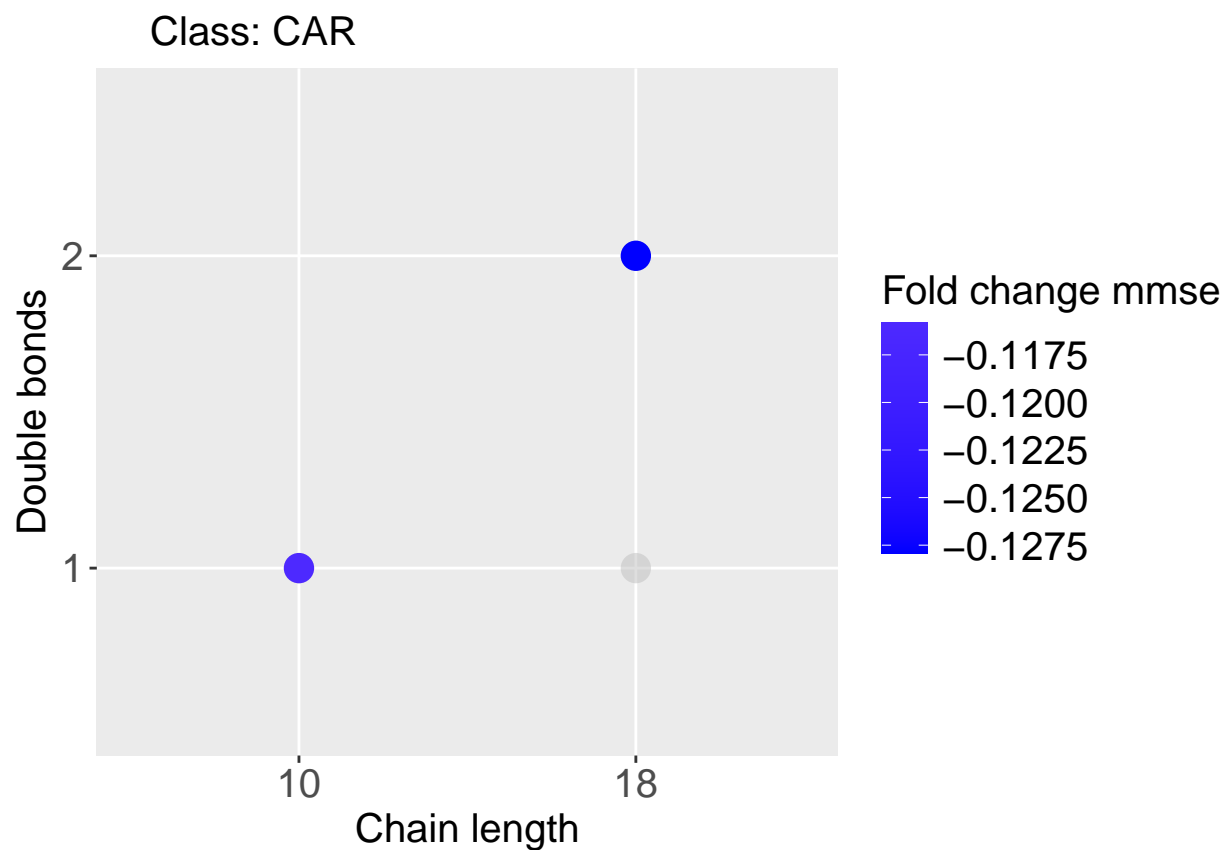
  d2 <- ggplot() + geom_point(data = data1, aes(x = Chain_length,
    y = Double_bounds), color = "grey", pch = 16, cex = 5,
```

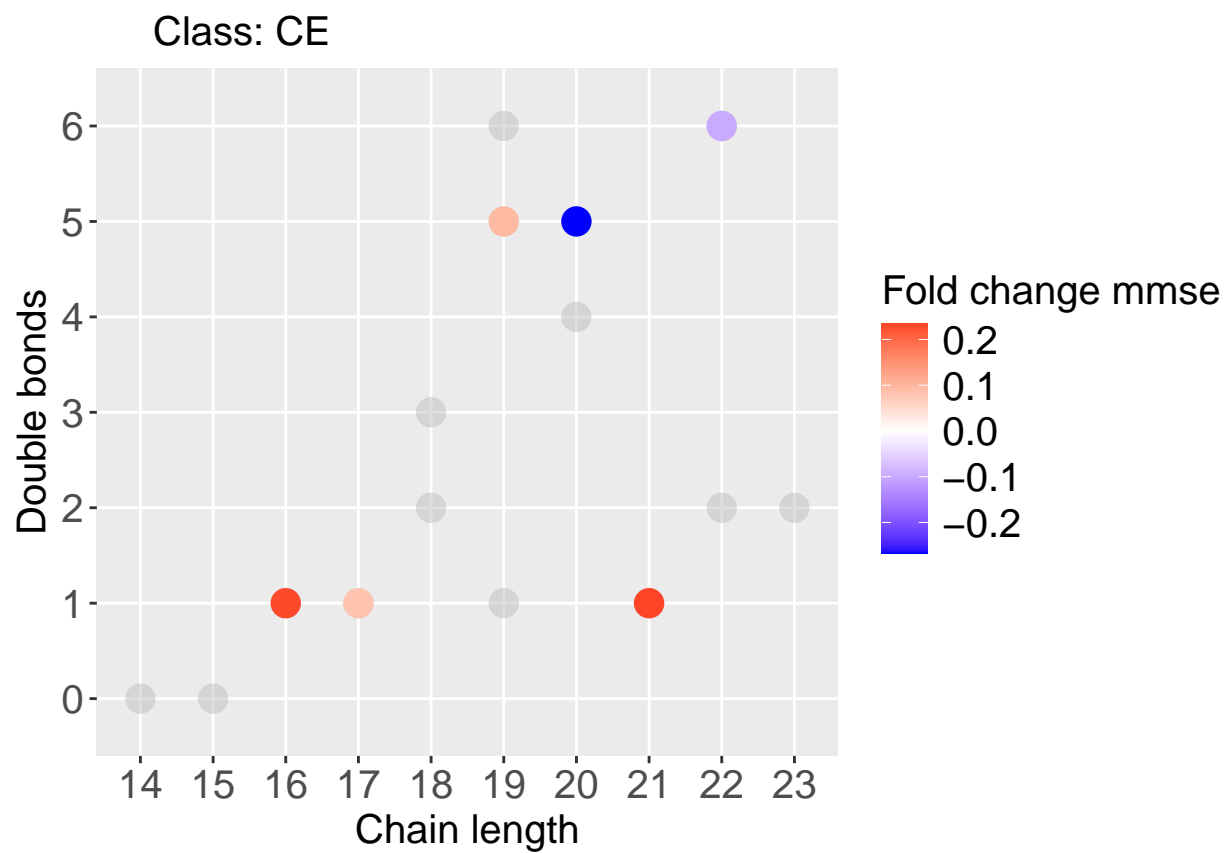
```

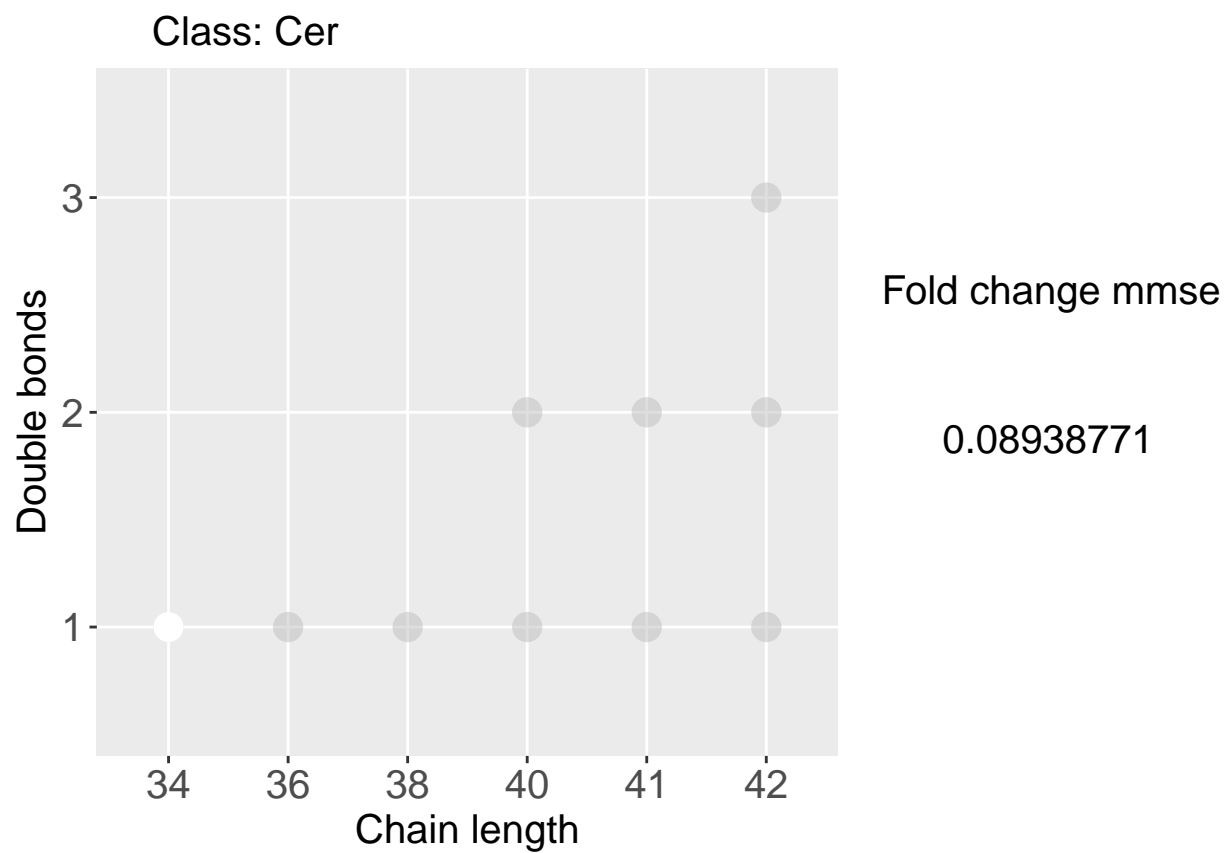
alpha = 0.5) + geom_point(data = data2, aes(x = Chain_length,
y = Double_bounds, color = FC_mmse), pch = 16, cex = 5,
alpha = 1) + labs(x = "Chain length", y = "Double bonds",
color = "FC_mmse") + ggtitle(paste("Class:", current_class)) +
scale_color_gradient2(low = "blue", mid = "white", high = "red",
midpoint = 0, name = "Fold change mmse") + theme(plot.title = element_text(hjust = 0.1,
size = 15), axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15),
legend.text = element_text(size = 15), legend.title = element_text(size = 15),
axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))

print(d2)
}

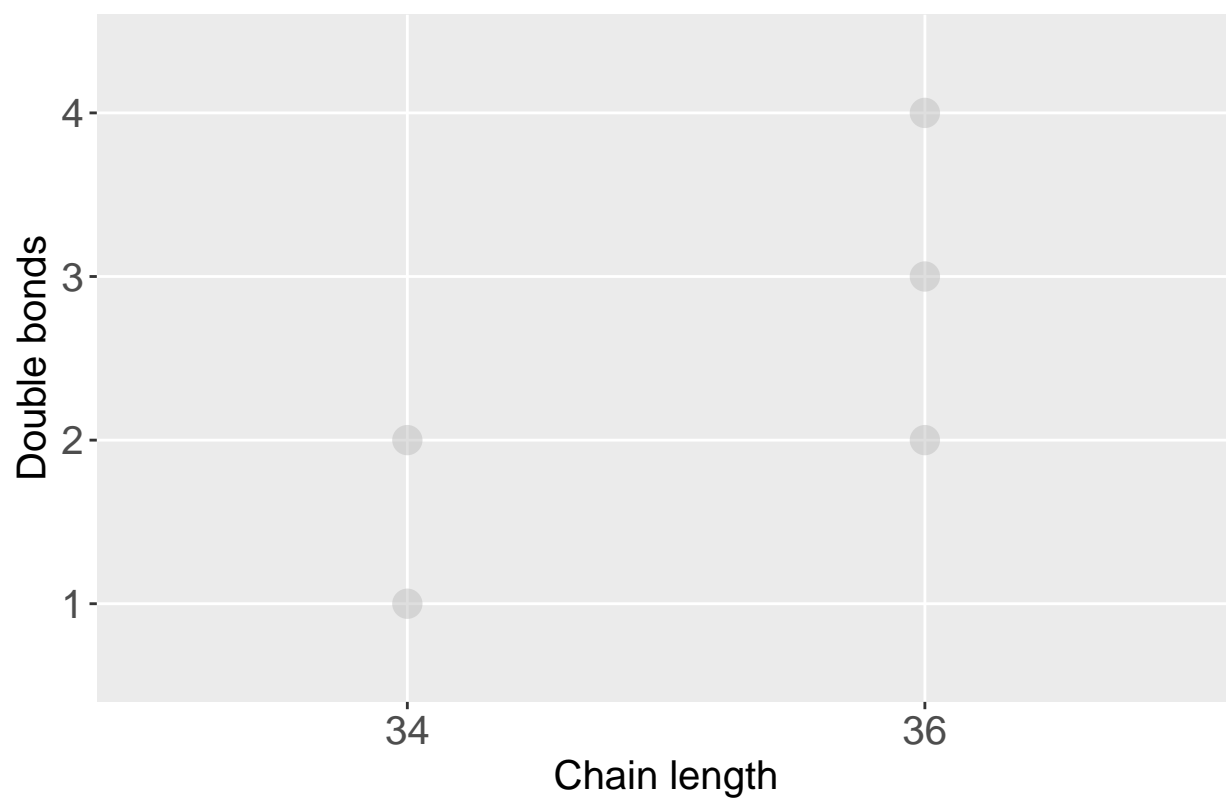
```

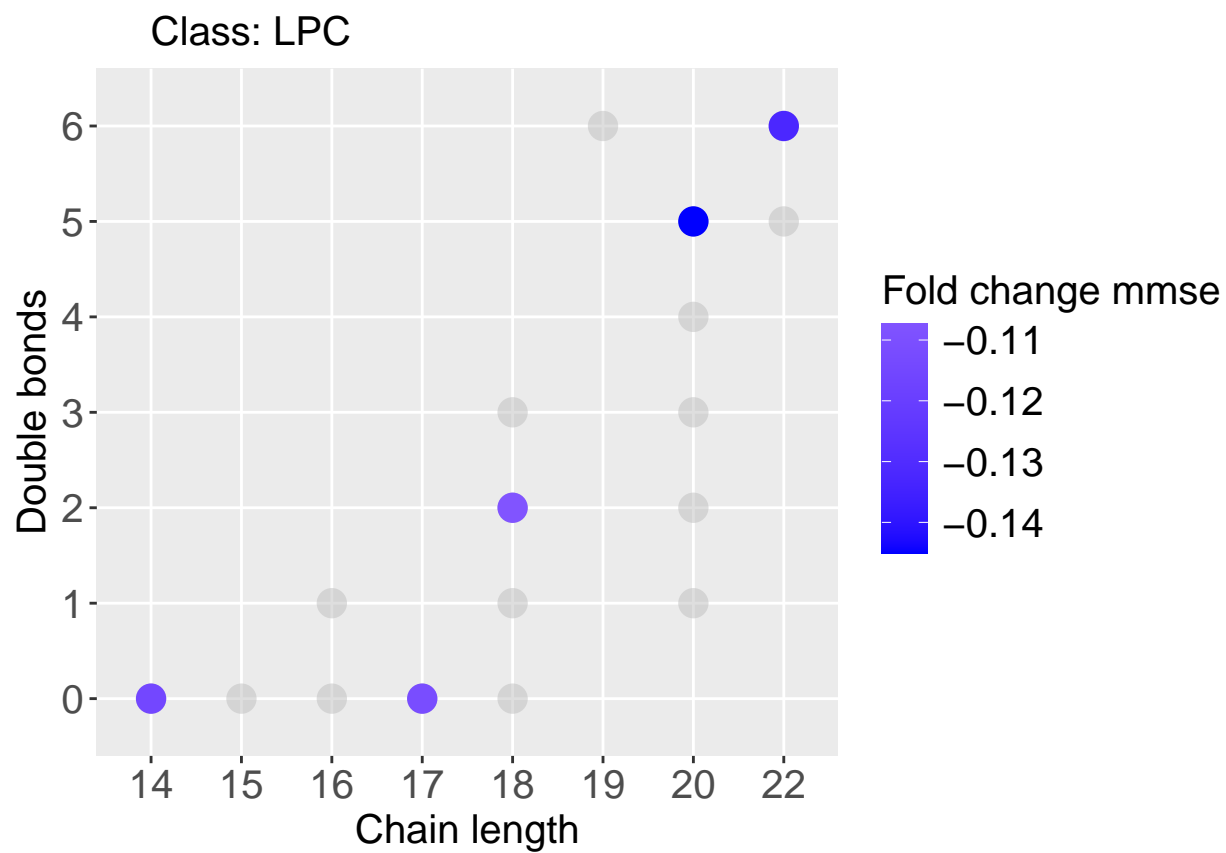


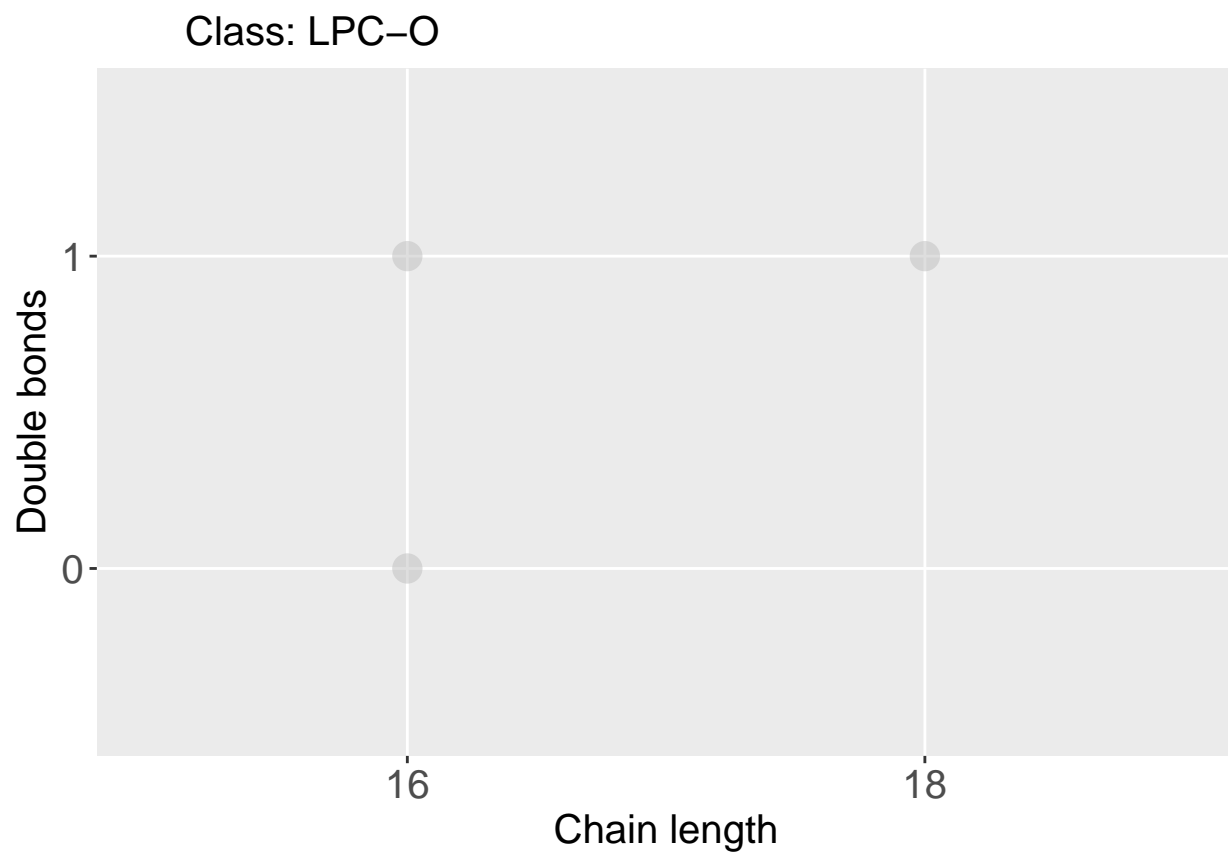




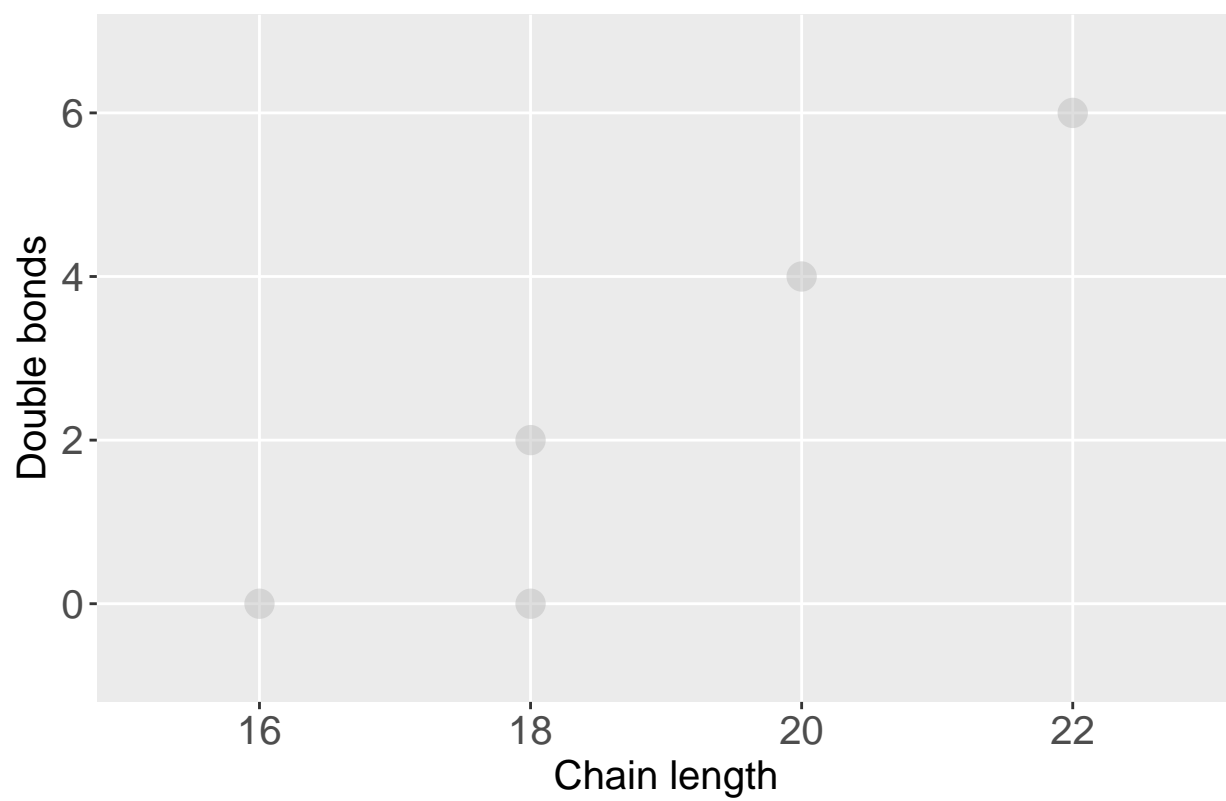
Class: DAG

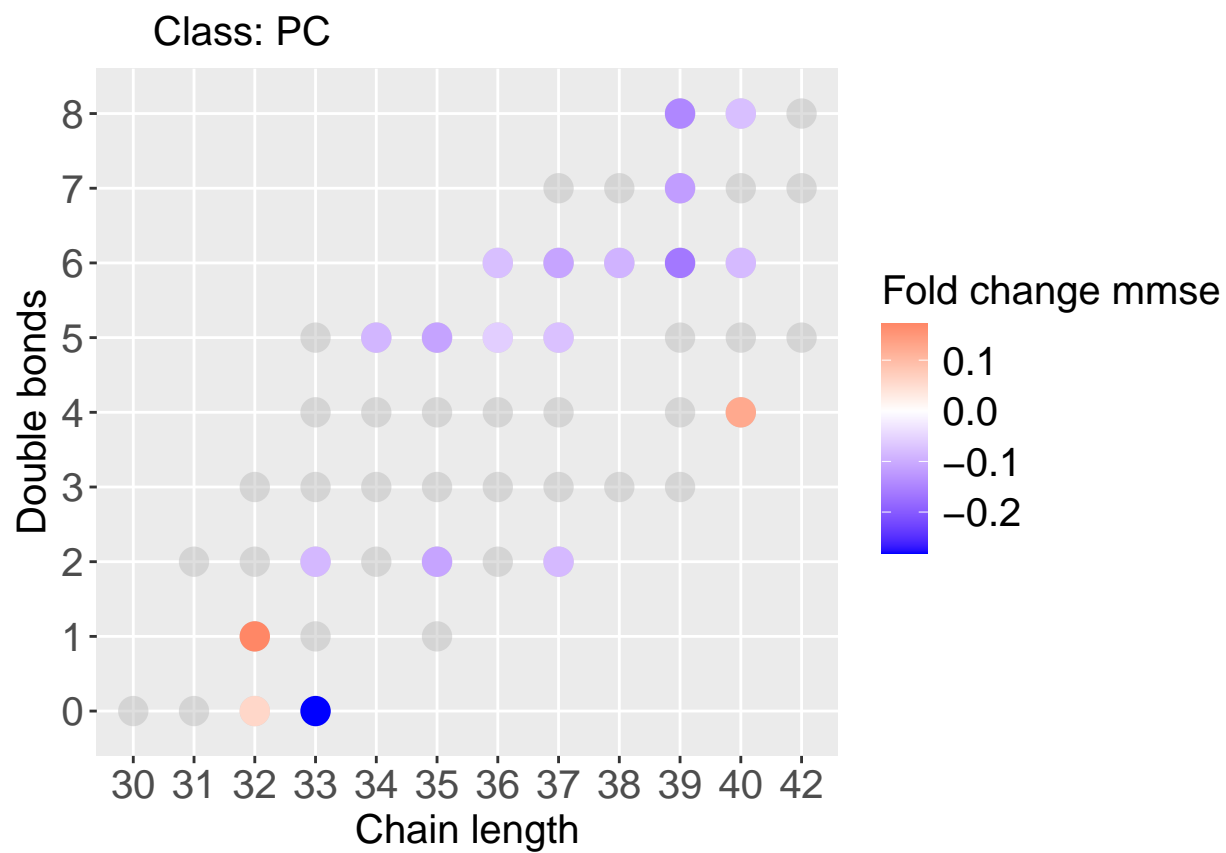


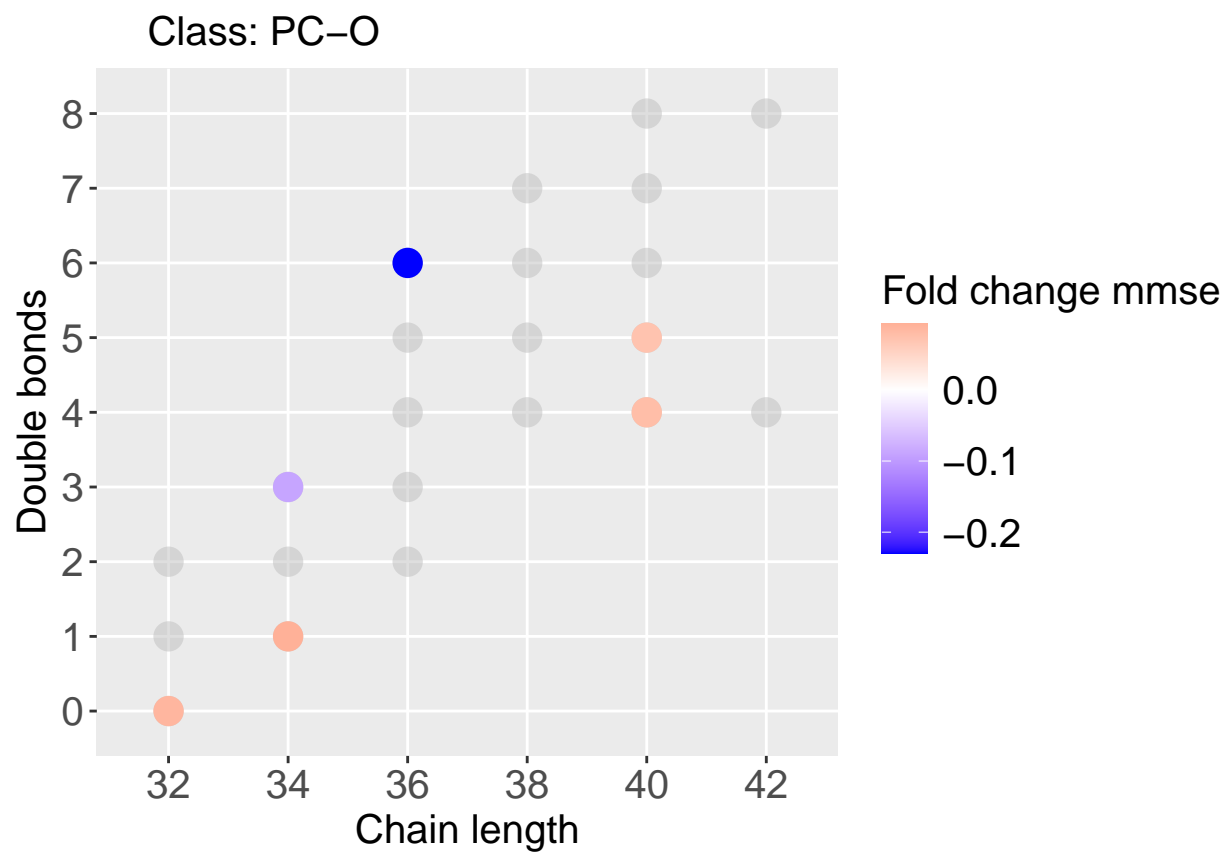


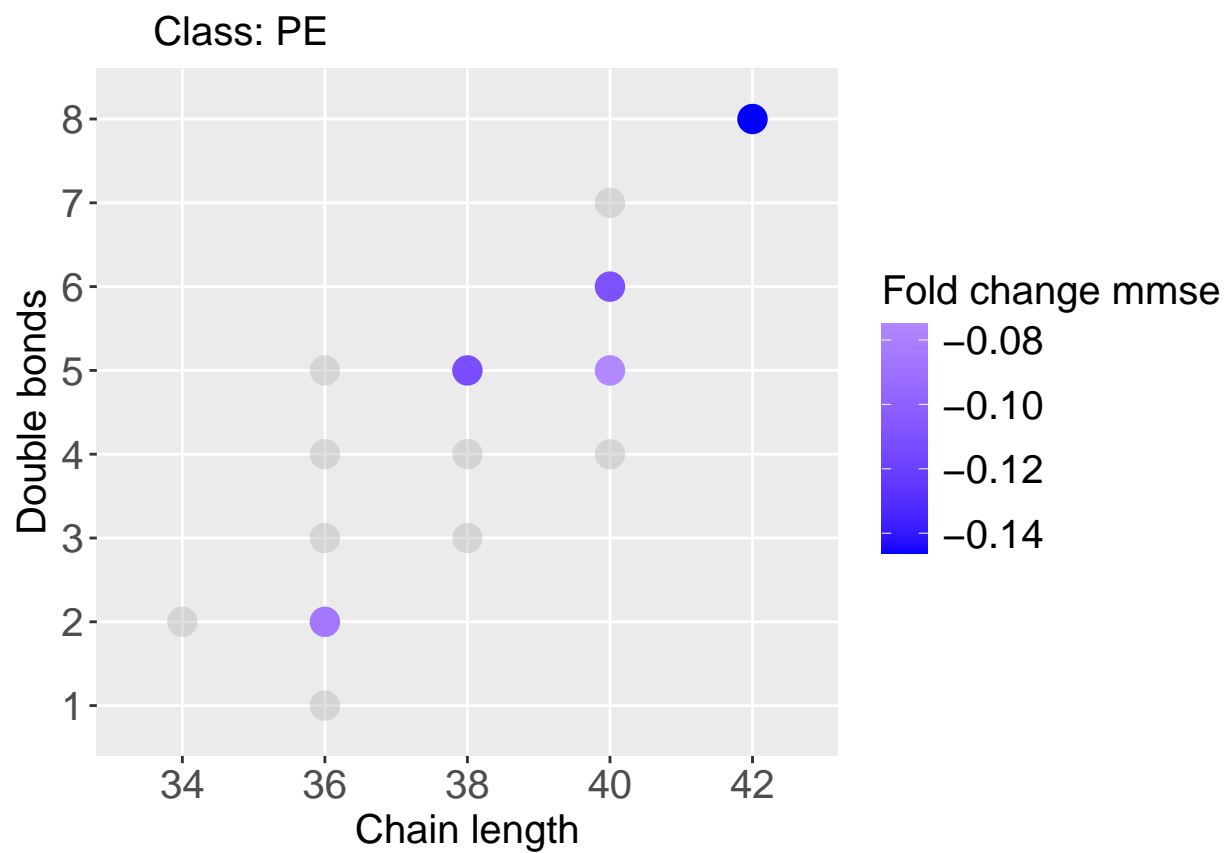


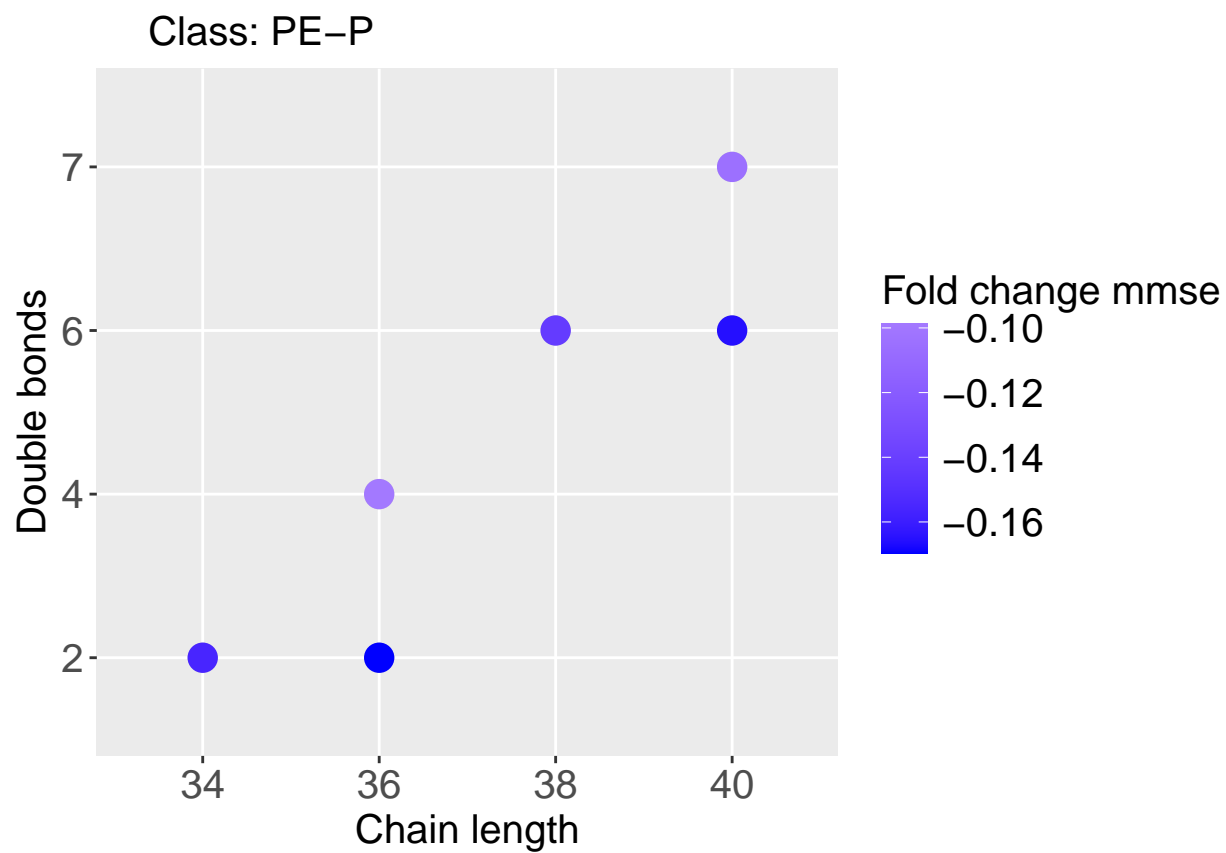
Class: LPE

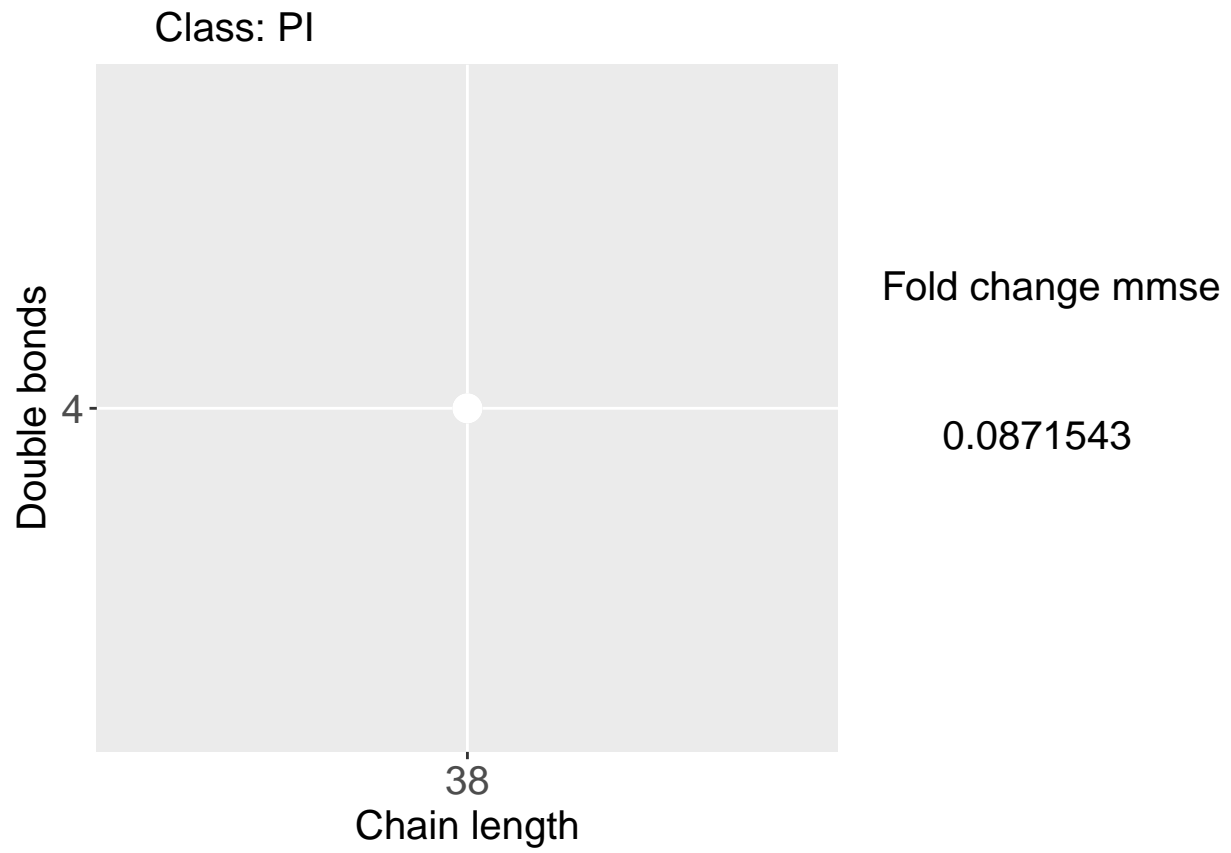


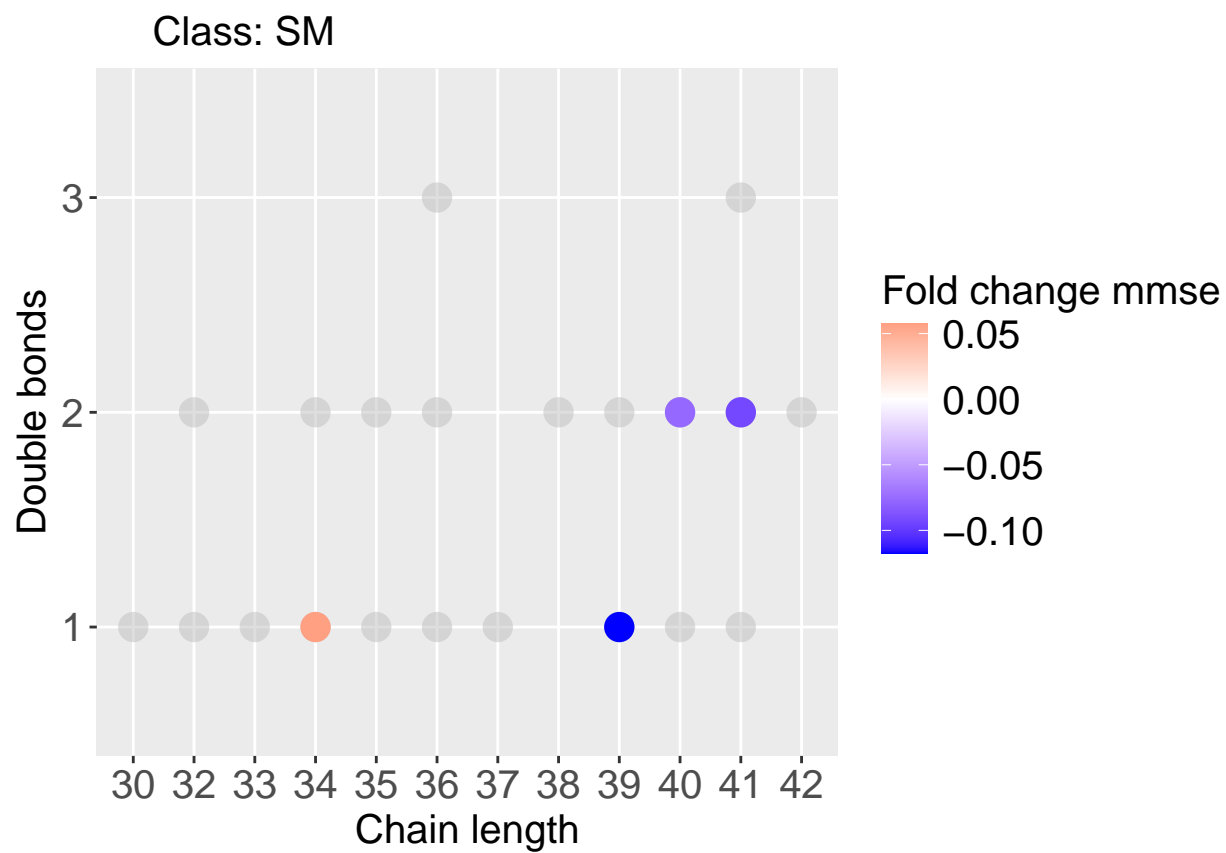


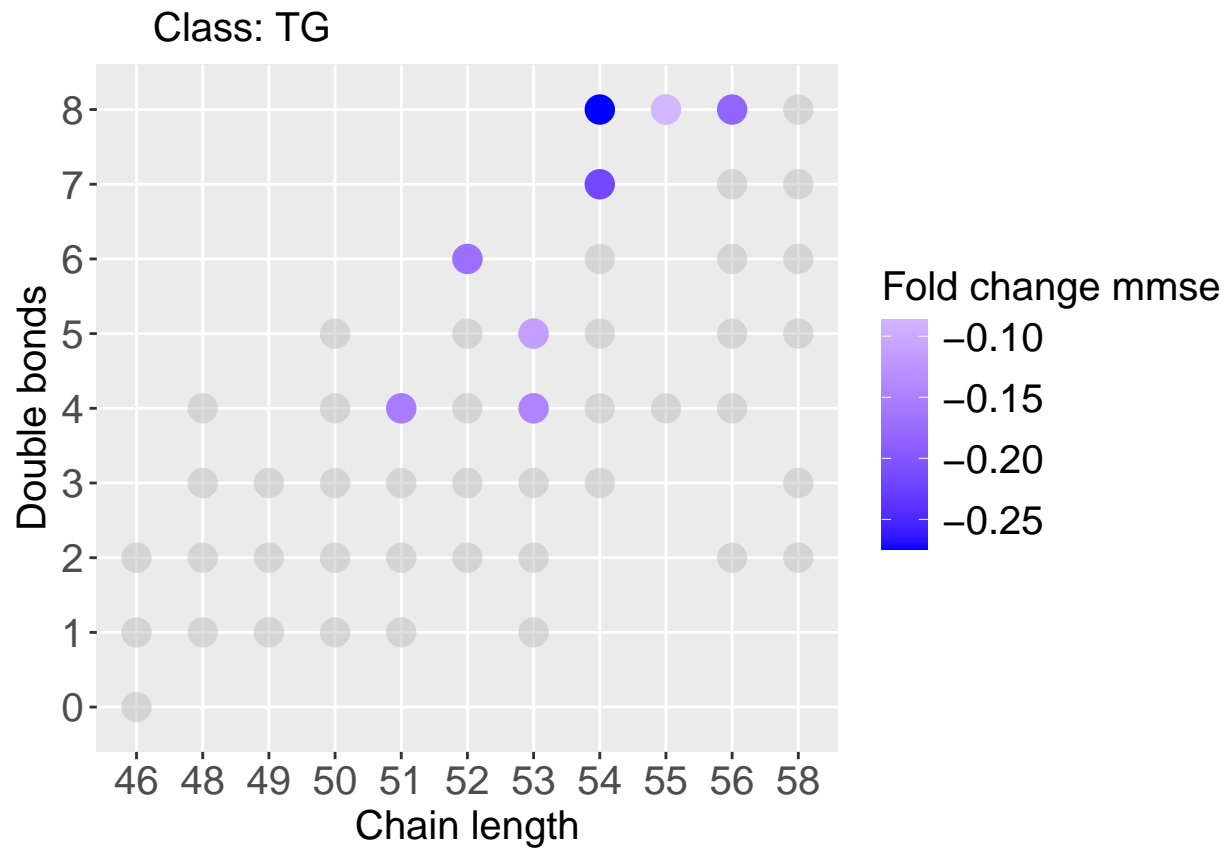












So we can also say that the number of long chain lipids is also reduced in patients with dementia.