Cluster analysis

Upload the libraries:

```r
library(gplots)
library(ggplot2)
library(RColorBrewer)
library(patchwork)
library(dplyr)
library(psych)
library(tidyr)
library(rstatix)
library(ggpubr)
```

Upload the data:

```r
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:

```r
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)
```

Let's perform k-mean clustering of samples according to scales:

The main purpose of clustering is to identify groups with different levels of successful ageing according to the available scales.

Prepare the data:

```
scales_data <- scales_lipids[, c(2:5, 9), drop = FALSE]
scales_data_scale <- scale(scales_data)
df_scales <- dist(scales_data_scale)
df_scales <- as.matrix(df_scales)
```

Perform the clusterization with **7 clusters:**

```
set.seed(123)
km_7 <- kmeans(df_scales, centers = 7, nstart = 100)
```
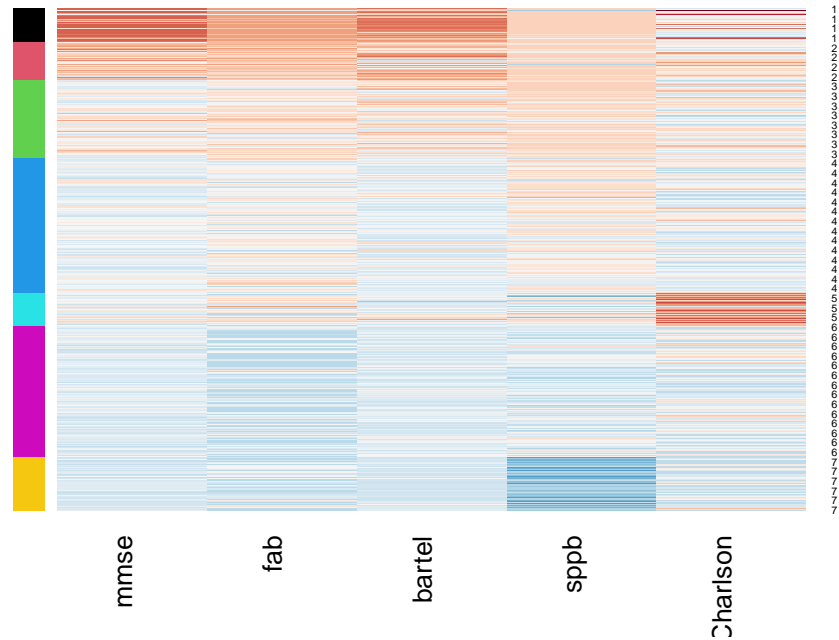
Change the clusters names:

```
new_cluster_numbers <- c(2, 4, 5, 3, 1, 7, 6)
km_7$cluster <- recode(km_7$cluster, !!!setNames(new_cluster_numbers,
    1:7))
new_cluster_counts <- table(km_7$cluster)
```

Create the heatmap with the results of clusterization:

```
cluster_order7 <- c(1, 2, 3, 4, 5, 6, 7)
cluster_order <- order(factor(km_7$cluster, levels = cluster_order7))
cluster_labels7 <- paste("Cluster", unique(km_7$cluster), sep = " ")
annotation_row_7 <- as.character(factor(km_7$cluster[cluster_order]))

heatmap.2(scales_data_scale[cluster_order, ], trace = "none",
    col = colorRampPalette(brewer.pal(11, "RdBu"))(100), scale = "none",
    dendrogram = "row", Rowv = FALSE, Colv = FALSE, margins = c(5,
        5), RowSideColors = annotation_row_7, labCol = colnames(scales_data_scale),
    labRow = km_7$cluster[cluster_order], cexCol = 1.2)
```

In this case, on the heat map, the redder the area on the graph, the more samples there are with low scores on one scale or another, i.e. the more patients with less successful ageing there are.

Let's look at data distribution in clusters:

Let's parse the data by clusters:

```r
cluster_labels <- km_7$cluster
scales_lipids_with_clusters <- cbind(scales_lipids, Cluster = cluster_labels)
names(scales_lipids_with_clusters)[names(scales_lipids_with_clusters) ==
    "(-1)Charlson"] <- "charlson"

df_cl1 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
    1, ]
df_cl2 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
    2, ]
df_cl3 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
    3, ]
df_cl4 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
    4, ]
df_cl5 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
    5, ]
df_cl6 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
```

```
      6, ]
df_cl7 <- scales_lipids_with_clusters[scales_lipids_with_clusters$Cluster ==
      7, ]
```

Let's build the distribution of data in clusters by age:

```
c_cc1 <- ggplot(df_cl1, aes(x = age)) + geom_histogram(fill = "#FFD700",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 1") + theme(plot.title = element_text(hjust = 0.5,
    size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc2 <- ggplot(df_cl2, aes(x = age)) + geom_histogram(fill = "#B4EEB4",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 2") + theme(plot.title = element_text(hjust = 0.5,
    size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 10), axis.text.y = element_text(size = 12))

c_cc3 <- ggplot(df_cl3, aes(x = age)) + geom_histogram(fill = "#FFA500",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 3") + theme(plot.title = element_text(hjust = 0.5,
    size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc4 <- ggplot(df_cl4, aes(x = age)) + geom_histogram(fill = "#EEA2AD",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 4") + theme(plot.title = element_text(hjust = 0.5,
    size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc5 <- ggplot(df_cl5, aes(x = age)) + geom_histogram(fill = "#87CEFF",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 5") + theme(plot.title = element_text(hjust = 0.5,
    size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc6 <- ggplot(df_cl6, aes(x = age)) + geom_histogram(fill = "purple",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 6") + theme(plot.title = element_text(hjust = 0.5,
    size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
    legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc7 <- ggplot(df_cl7, aes(x = age)) + geom_histogram(fill = "red",
    color = "black", bins = 20) + labs(x = "Age", y = "Count",
    title = "Age distribution \n in cluster 7") + theme(plot.title = element_text(hjust = 0.5,
```
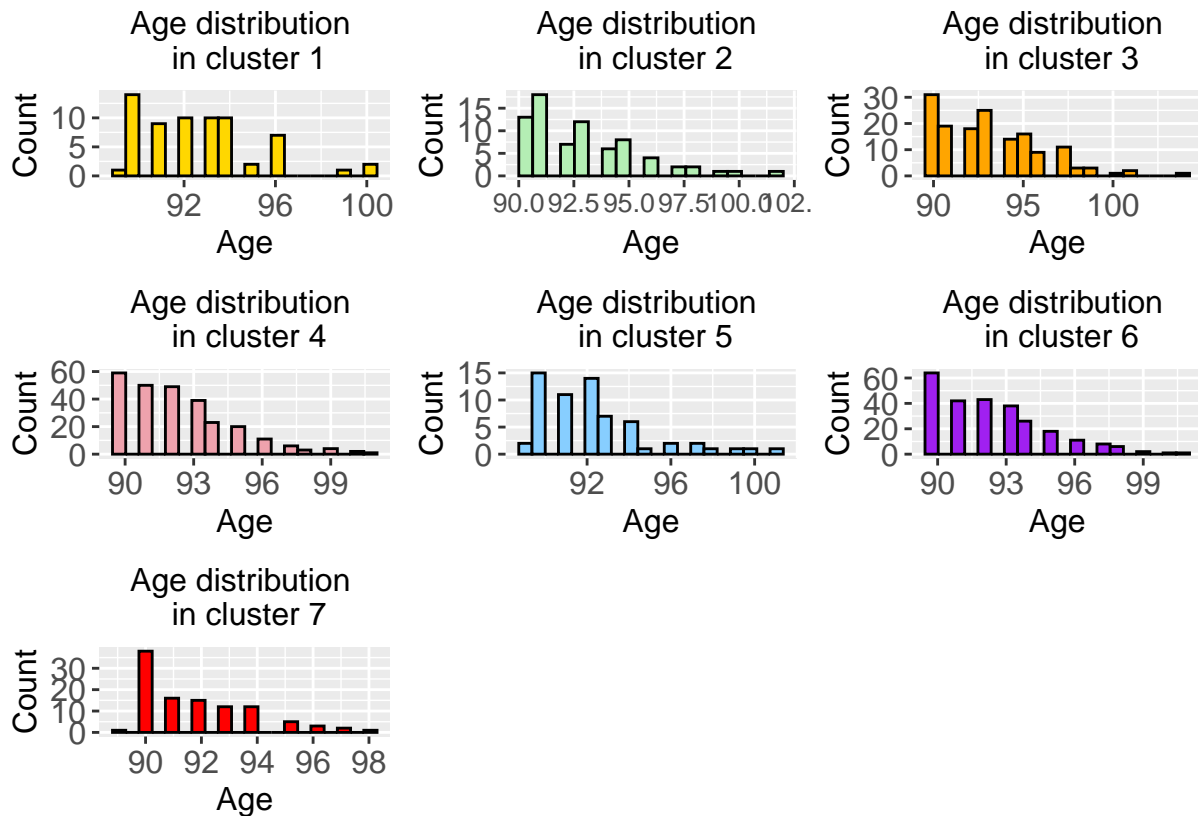
```
        size = 12), axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
        theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc1 + c_cc2 + c_cc3 + c_cc4 + c_cc5 + c_cc6 + c_cc7
```


The age distributions in the clusters look similar.

Let's build the distribution of data in clusters by gender:

```
c_cc8 <- ggplot(df_cl1, aes(x = factor(sex))) + geom_histogram(fill = "#FFD700",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 1") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc9 <- ggplot(df_cl2, aes(x = factor(sex))) + geom_histogram(fill = "#B4EEB4",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 2") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
```

```r
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc10 <- ggplot(df_cl3, aes(x = factor(sex))) + geom_histogram(fill = "#FFA500",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 3") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc11 <- ggplot(df_cl4, aes(x = factor(sex))) + geom_histogram(fill = "#EEA2AD",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 4") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc12 <- ggplot(df_cl5, aes(x = factor(sex))) + geom_histogram(fill = "#87CEFF",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 5") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc13 <- ggplot(df_cl6, aes(x = factor(sex))) + geom_histogram(fill = "purple",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 6") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc14 <- ggplot(df_cl7, aes(x = factor(sex))) + geom_histogram(fill = "red",
    color = "black", bins = 20, stat = "count") + labs(x = "Gender",
    y = "Count", title = "Gender distribution \n in cl 7") +
    theme(plot.title = element_text(hjust = 0.5, size = 12),
        axis.title.x = element_text(size = 12), axis.title.y = element_text(size = 12),
        legend.text = element_text(size = 12), legend.title = element_text(size = 12)) +
    theme(axis.text.x = element_text(size = 12), axis.text.y = element_text(size = 12))

c_cc8 + c_cc9 + c_cc10 + c_cc11 + c_cc12 + c_cc13 + c_cc14
```
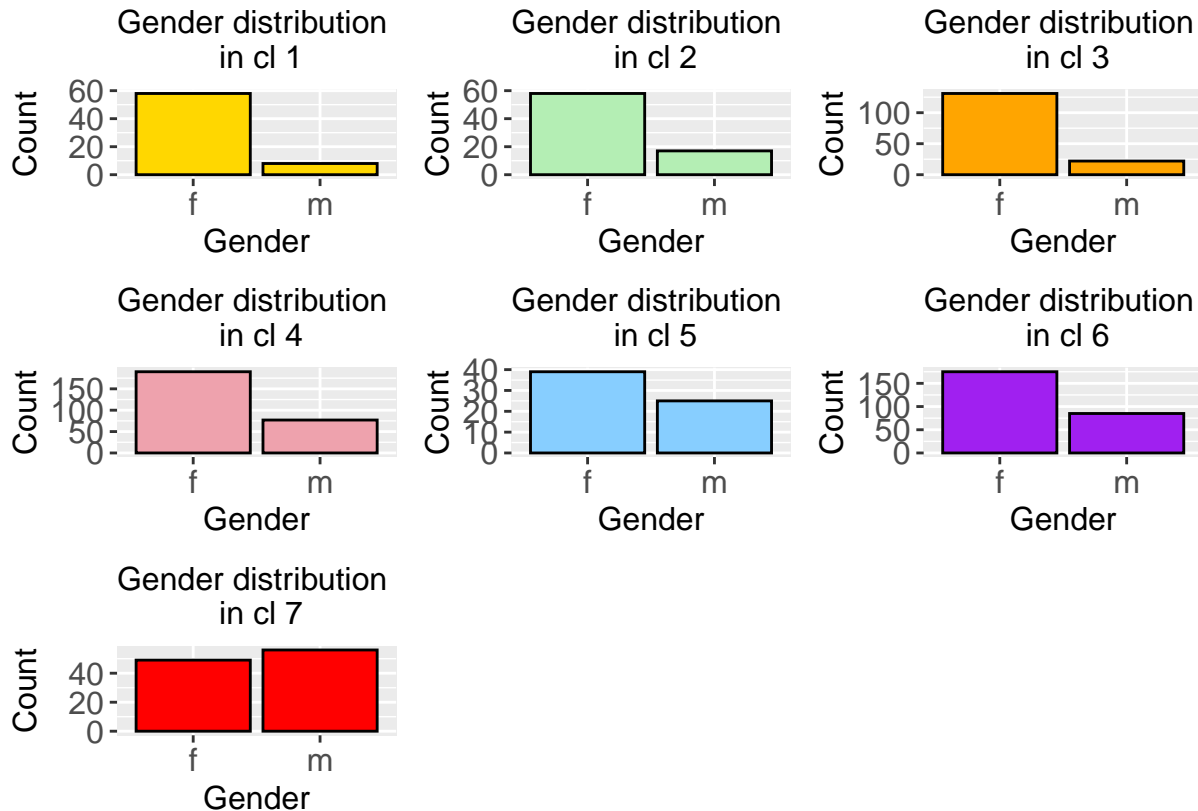
Gender distribution in cl 1

Gender distribution in cl 2

Gender distribution in cl 3

Gender distribution in cl 4

Gender distribution in cl 5

Gender distribution in cl 6

Gender distribution in cl 7

In general, all clusters have about the same ratio of males to females, but clusters 7 and 5 have many male samples.

Comparison of lipid profiles of cluster samples with each other:

```
scales_lipids_with_clusters_long <- pivot_longer(scales_lipids_with_clusters,
    cols = 10:230, names_to = "lipid_features", values_to = "values")


allclusters <- c(1, 2, 3, 4, 5, 6, 7)

test_results <- list()
test_results_significant <- list()

clusters_lipids_signif <- matrix(0, nrow = 7, ncol = 7)
colnames(clusters_lipids_signif) <- c("cluster 1", "cluster 2",
    "cluster 3", "cluster 4", "cluster 5", "cluster 6", "cluster 7")
rownames(clusters_lipids_signif) <- c("cluster 1", "cluster 2",
    "cluster 3", "cluster 4", "cluster 5", "cluster 6", "cluster 7")

for (clusternumber in allclusters) {

    current_cluster <- filter(scales_lipids_with_clusters_long,
        Cluster == clusternumber)
```

```
    for (comparison_cluster in allclusters) {

        if (comparison_cluster != clusternumber) {

            comparison_data_cluster <- filter(scales_lipids_with_clusters_long,
                Cluster == comparison_cluster)

            comparison_data <- rbind(current_cluster, comparison_data_cluster)

            test_result <- comparison_data %>%
                group_by(lipid_features) %>%
                t_test(values ~ Cluster) %>%
                adjust_pvalue(method = "BH") %>%
                add_significance()

            test_results[[paste("stat.test_cl", clusternumber,
                "_", comparison_cluster, sep = "")]] <- test_result

            test_result_significant <- test_result[test_result$p.adj <=
                0.1, ]

            test_results_significant[[paste("stat.test_cl", clusternumber,
                "_", comparison_cluster, sep = "")]] <- test_result_significant

            clusters_lipids_signif[clusternumber, comparison_cluster] <- nrow(test_result_significant)
        }
    }
}
```
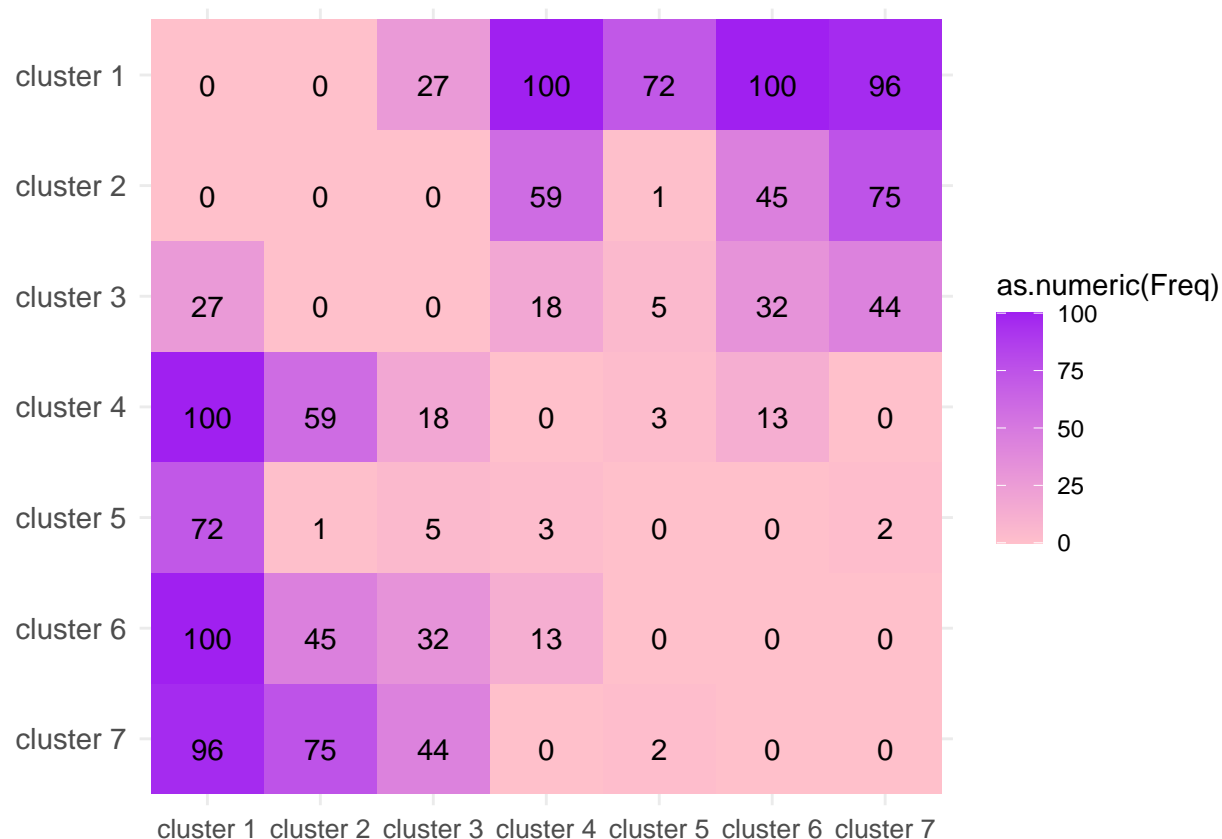
Let's look at the number of significant lipids resulting from the comparison of
the lipid profiles of the clusters:

```
clusters_lipids_significant <- clusters_lipids_signif[c(1, 2,
    3, 4, 5, 6, 7), c(7, 6, 5, 4, 3, 2, 1)]
gr <- colorRampPalette(c("pink", "purple"))
df <- as.data.frame(clusters_lipids_significant)
df <- as.data.frame(as.table(clusters_lipids_significant))

ggplot(df, aes(x = Var1, y = Var2, fill = as.numeric(Freq))) +
    geom_tile() + geom_text(aes(label = as.numeric(Freq)), vjust = 1,
    size = 4) + scale_fill_gradient(low = "pink", high = "purple") +
    theme_minimal() + theme(axis.text.x = element_text(size = 10),
    axis.text.y = element_text(size = 11)) + theme(axis.title.x = element_blank(),
    axis.title.y = element_blank())
```

	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7
cluster 1	0	0	27	100	72	100	96
cluster 2	0	0	0	59	1	45	75
cluster 3	27	0	0	18	5	32	44
cluster 4	100	59	18	0	3	13	0
cluster 5	72	1	5	3	0	0	2
cluster 6	100	45	32	13	0	0	0
cluster 7	96	75	44	0	2	0	0

as.numeric(Freq): 100, 75, 50, 25, 0

Thus, in the case of ageing, we see a complex structure of different groups with different levels of successful ageing, differing in their lipid profile.

Let's see how much on average the lipid profile of the clusters correlate with each other:

Count z - scores:

```
scale_lipids_clust <- scale(scales_lipids_with_clusters[, 10:230])
scale_lipids_clust <- as.data.frame(scale_lipids_clust)
scale_lipids_clust$MS.ID <- scales_lipids_with_clusters$MS.ID
scale_lipids_clust$Cluster <- scales_lipids_with_clusters$Cluster
scale_lipids_clust_long <- pivot_longer(scale_lipids_clust, cols = 1:221,
    names_to = "lipid_features", values_to = "values")
```

Count the mean:

```
allclusters <- c(1, 2, 3, 4, 5, 6, 7)

for (clusternumber in allclusters) {
```

```r
    scale_cluster <- scale_lipids_clust_long[scale_lipids_clust_long$Cluster ==
        clusternumber, ]

    cluster_mean <- scale_cluster %>%
        group_by(lipid_features) %>%
        summarise(mean = mean(values))

    assign(paste("mean_cluster_", clusternumber, sep = ""), cluster_mean)
}

mean_cluster_1 <- mean_cluster_1 %>%
    rename(mean_1 = mean)

mean_all_cluster <- mean_cluster_1
mean_all_cluster$mean_2 <- mean_cluster_2$mean
mean_all_cluster$mean_3 <- mean_cluster_3$mean
mean_all_cluster$mean_4 <- mean_cluster_4$mean
mean_all_cluster$mean_5 <- mean_cluster_5$mean
mean_all_cluster$mean_6 <- mean_cluster_6$mean
mean_all_cluster$mean_7 <- mean_cluster_7$mean
```
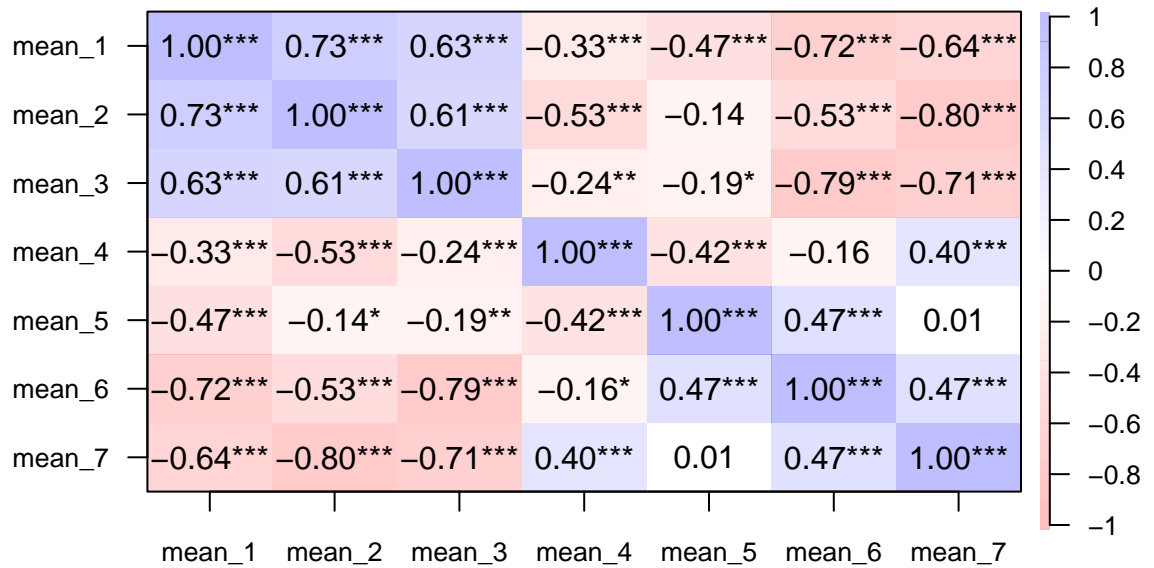
Plot the results:

```r
corPlot(mean_all_cluster[, c("mean_1", "mean_2", "mean_3", "mean_4",
    "mean_5", "mean_6", "mean_7")], cex = 1, stars = TRUE, alpha = 0.25,
    cex.axis = 0.8, main = "Correlation of mean z-scores")
```

Correlation of mean z–scores

	mean_1	mean_2	mean_3	mean_4	mean_5	mean_6	mean_7
mean_1	1.00***	0.73***	0.63***	−0.33***	−0.47***	−0.72***	−0.64***
mean_2	0.73***	1.00***	0.61***	−0.53***	−0.14	−0.53***	−0.80***
mean_3	0.63***	0.61***	1.00***	−0.24**	−0.19*	−0.79***	−0.71***
mean_4	−0.33***	−0.53***	−0.24***	1.00***	−0.42***	−0.16	0.40***
mean_5	−0.47***	−0.14*	−0.19**	−0.42***	1.00***	0.47***	0.01
mean_6	−0.72***	−0.53***	−0.79***	−0.16*	0.47***	1.00***	0.47***
mean_7	−0.64***	−0.80***	−0.71***	0.40***	0.01	0.47***	1.00***

Clusters 1, 2 and 3 as well as 4, 5, 6 and 7 were the most similar to each other in their lipid profile.

Let's take a closer look at the differences in lipid profile between clusters 1 and 6, 3 and 6:

Parse the data:

```
stat.test_cl1_cl6_signif <- test_results_significant$stat.test_cl1_6
stat.test_cl3_cl6_signif <- test_results_significant$stat.test_cl3_6
stat.test_cl1_cl3_signif <- test_results_significant$stat.test_cl1_3

cl1_long <- scales_lipids_with_clusters_long[scales_lipids_with_clusters_long$Cluster ==
    1, ]

cl3_long <- scales_lipids_with_clusters_long[scales_lipids_with_clusters_long$Cluster ==
    3, ]

cl6_long <- scales_lipids_with_clusters_long[scales_lipids_with_clusters_long$Cluster ==
    6, ]
```

Count the mean:

```r
cl1_mean <- cl1_long %>%
    group_by(lipid_features) %>%
    summarise(mean_cl1 = mean(values))

cl3_mean <- cl3_long %>%
    group_by(lipid_features) %>%
    summarise(mean_cl3 = mean(values))

cl6_mean <- cl6_long %>%
    group_by(lipid_features) %>%
    summarise(mean_cl6 = mean(values))
```

Count the fold change:

```r
mean_1_6 <- merge(cl1_mean, cl6_mean, by = "lipid_features")
mean_1_6$FC_1_6 <- mean_1_6$mean_cl1 - mean_1_6$mean_cl6

mean_3_6 <- merge(cl3_mean, cl6_mean, by = "lipid_features")
mean_3_6$FC_3_6 <- mean_3_6$mean_cl3 - mean_3_6$mean_cl6

FC_16_36 <- merge(mean_1_6, mean_3_6, by = "lipid_features")
```

Merging dataframes into one:

```r
FC_16_36_signif1_6 <- merge(FC_16_36, stat.test_cl1_cl6_signif,
    by = "lipid_features")
FC_16_36_signif3_6 <- merge(FC_16_36, stat.test_cl3_cl6_signif,
    by = "lipid_features")
data16_36 = merge(stat.test_cl1_cl6_signif, stat.test_cl3_cl6_signif,
    by = "lipid_features")

FC_16_36_signif_both <- merge(FC_16_36, data16_36, by = "lipid_features")
```

Plot the results:

```r
data1 <- FC_16_36_signif1_6
data2 <- FC_16_36_signif3_6
data3 <- FC_16_36_signif_both


ggplot() + geom_point(data = data1, aes(x = FC_1_6, y = FC_3_6,
    color = "Significant lipids \n comparing clusters 1 and 6"),
    pch = 16, cex = 3, alpha = 0.5) + geom_point(data = data2,
    aes(x = FC_1_6, y = FC_3_6, color = "Significant lipids \n comparing clusters 3 and 6"),
    pch = 16, cex = 3, alpha = 0.5) + geom_point(data = data3,
```

```
    aes(x = FC_1_6, y = FC_3_6, color = "Intersection"), pch = 16,
    cex = 3, alpha = 0.5) + labs(x = "Fold change between cl1 and cl6",
    y = "Fold change between \n cl3 and cl6", color = "Group") +
    geom_vline(xintercept = 0, linetype = "dashed", color = "black") +
    geom_hline(yintercept = 0, linetype = "dashed", color = "black") +
    geom_abline(slope = 1, intercept = 0, color = "black") +
    scale_color_manual(values = c("#FF9999", "blue", "#8FBC8F")) +
    guides(color = guide_legend(title = NULL)) + theme_minimal() +
    theme(legend.position = "right", legend.text = element_text(size = 15),
        axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15),
        axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15))
```
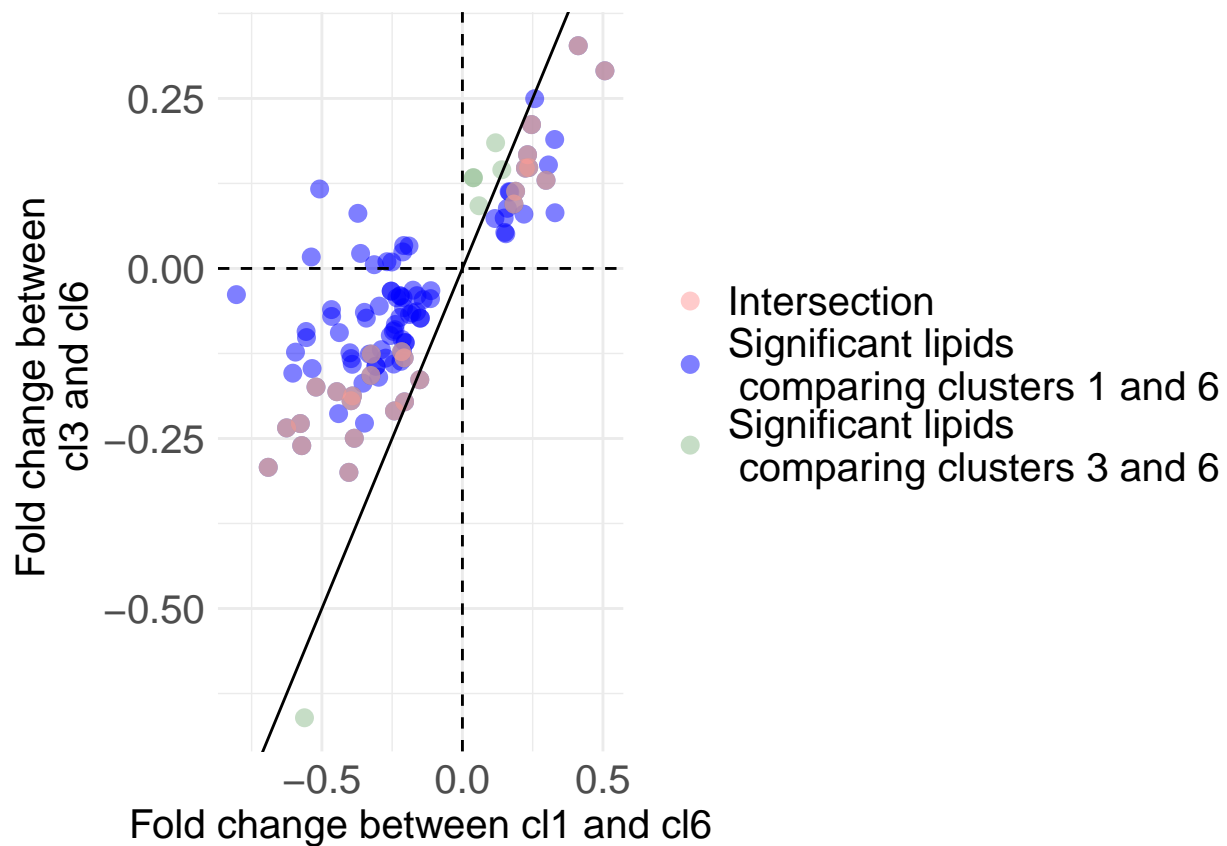

Let's do a correlation test:

```
merged_FC_16_36 <- rbind(FC_16_36_signif1_6, FC_16_36_signif3_6)
unique_FC_16_36 <- merged_FC_16_36[!duplicated(merged_FC_16_36$lipid_features),
    ]

cor_test_result_36_16 <- cor.test(unique_FC_16_36$FC_1_6, unique_FC_16_36$FC_3_6)
cor_test_result_36_16
```

```
##
##  Pearson's product-moment correlation
```

```
## 
## data:  unique_FC_16_36$FC_1_6 and unique_FC_16_36$FC_3_6
## t = 12.554, df = 104, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6871571 0.8422284
## sample estimates:
##       cor
## 0.7761726
```

Plot the histigrams for scales:

```
mmse_charlson <- ggplot() + geom_histogram(data = df_cl1, aes(x = mmse,
    fill = "Cluster 1"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl3, aes(x = mmse,
    fill = "Cluster 3"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl6, aes(x = mmse,
    fill = "Cluster 6"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + scale_fill_manual(values = c(`Cluster 1` = "#8FBC8F",
    `Cluster 3` = "blue", `Cluster 6` = "#FF9999")) + labs(fill = "Groups") +
    theme_minimal() + theme(plot.title = element_text(hjust = 0.5,
    size = 10), axis.title.x = element_text(size = 10), axis.title.y = element_text(size = 10),
    legend.text = element_text(size = 6), legend.title = element_text(size = 6),
    legend.position = "right") + theme(axis.text.x = element_text(size = 8),
    axis.text.y = element_text(size = 8))

fab_charlson <- ggplot() + geom_histogram(data = df_cl1, aes(x = fab,
    fill = "Cluster 1"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl3, aes(x = fab,
    fill = "Cluster 3"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl6, aes(x = fab,
    fill = "Cluster 6"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + scale_fill_manual(values = c(`Cluster 1` = "#8FBC8F",
    `Cluster 3` = "blue", `Cluster 6` = "#FF9999")) + labs(fill = "Groups") +
    theme_minimal() + theme(plot.title = element_text(hjust = 0.5,
    size = 11), axis.title.x = element_text(size = 11), axis.title.y = element_text(size = 11),
    legend.text = element_text(size = 7), legend.title = element_text(size = 7),
    legend.position = "right") + theme(axis.text.x = element_text(size = 9),
    axis.text.y = element_text(size = 9))

bartel_charlson <- ggplot() + geom_histogram(data = df_cl1, aes(x = bartel,
    fill = "Cluster 1"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl3, aes(x = bartel,
    fill = "Cluster 3"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl6, aes(x = bartel,
    fill = "Cluster 6"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + scale_fill_manual(values = c(`Cluster 1` = "#8FBC8F",
    `Cluster 3` = "blue", `Cluster 6` = "#FF9999")) + labs(fill = "Groups") +
    theme_minimal() + theme(plot.title = element_text(hjust = 0.5,
    size = 11), axis.title.x = element_text(size = 11), axis.title.y = element_text(size = 11),
    legend.text = element_text(size = 7), legend.title = element_text(size = 7),
    legend.position = "right") + theme(axis.text.x = element_text(size = 9),
```

```r
    axis.text.y = element_text(size = 9))

sppb_charlson <- ggplot() + geom_histogram(data = df_cl1, aes(x = sppb,
    fill = "Cluster 1"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl3, aes(x = sppb,
    fill = "Cluster 3"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl6, aes(x = sppb,
    fill = "Cluster 6"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + scale_fill_manual(values = c(`Cluster 1` = "#8FBC8F",
    `Cluster 3` = "blue", `Cluster 6` = "#FF9999")) + labs(fill = "Groups") +
    theme_minimal() + theme(plot.title = element_text(hjust = 0.5,
    size = 11), axis.title.x = element_text(size = 11), axis.title.y = element_text(size = 11),
    legend.text = element_text(size = 7), legend.title = element_text(size = 7),
    legend.position = "right") + theme(axis.text.x = element_text(size = 9),
    axis.text.y = element_text(size = 9))

charlson <- ggplot() + geom_histogram(data = df_cl1, aes(x = charlson,
    fill = "Cluster 1"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl3, aes(x = charlson,
    fill = "Cluster 3"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + geom_histogram(data = df_cl6, aes(x = charlson,
    fill = "Cluster 6"), alpha = 0.6, bins = 20, color = "black",
    stat = "count") + scale_fill_manual(values = c(`Cluster 1` = "#8FBC8F",
    `Cluster 3` = "blue", `Cluster 6` = "#FF9999")) + labs(fill = "Groups") +
    theme_minimal() + theme(plot.title = element_text(hjust = 0.5,
    size = 11), axis.title.x = element_text(size = 11), axis.title.y = element_text(size = 11),
    legend.text = element_text(size = 7), legend.title = element_text(size = 7),
    legend.position = "right") + theme(axis.text.x = element_text(size = 9),
    axis.text.y = element_text(size = 9))

mmse_charlson + fab_charlson + bartel_charlson + sppb_charlson +
    charlson
```
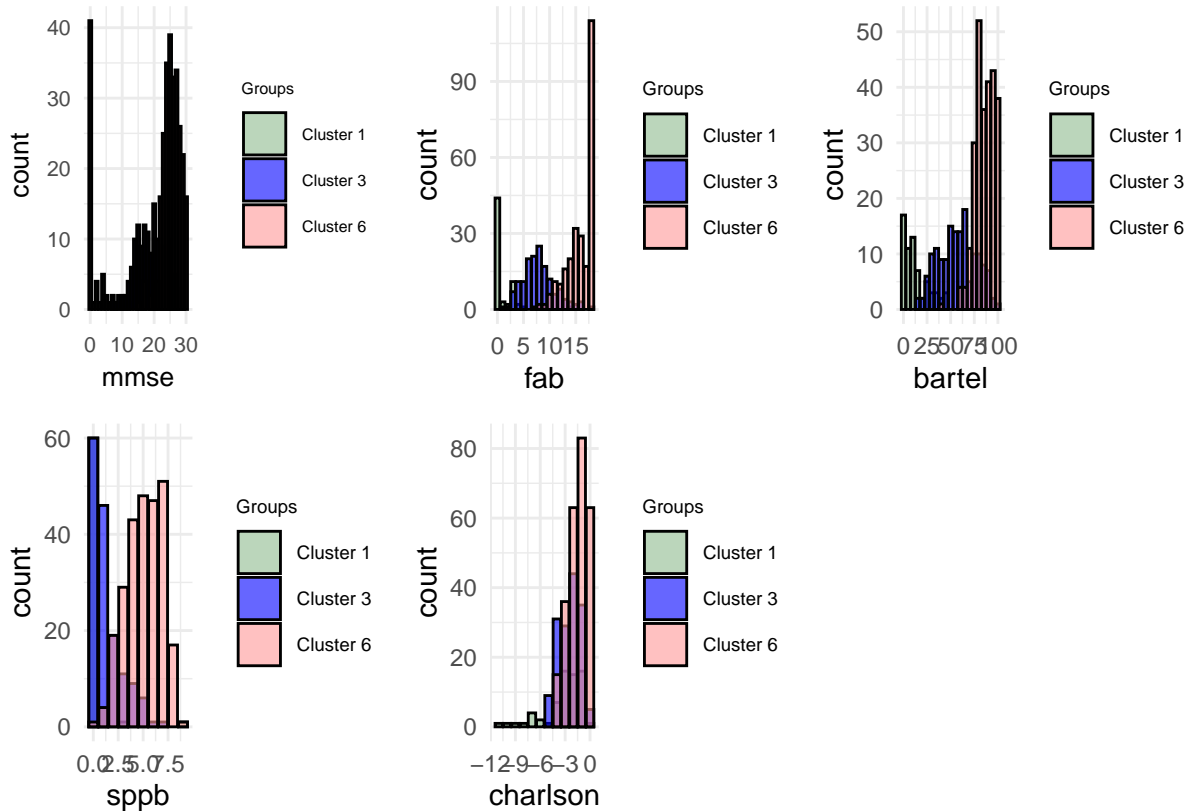
Thus, most of the significant lipids obtained by comparing clusters 3 and 6 is the same as those obtained by comparing clusters 1 and 6.

Let's look at the significant lipids obtained by comparing clusters 1 and 6:

Let's parse the data on lipids:

```r
stat.test_cl1_cl6 <- test_results$stat.test_cl1_6

stat.test_cl1_cl6$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", stat.test_cl1_cl6$lipid_features)
stat.test_cl1_cl6_sep <- separate(stat.test_cl1_cl6, lipid_features,
    into = c("Class", "Other"), sep = " ", remove = FALSE)
stat.test_cl1_cl6_sep <- separate(stat.test_cl1_cl6_sep, Other,
    into = c("Chain_length", "Double_bounds"), sep = ":")

stat.test_cl1_cl6_signif$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", stat.test_cl1_cl6_signif$lipid_features)
stat.test_cl1_cl6_signif_sep <- separate(stat.test_cl1_cl6_signif,
    lipid_features, into = c("Class", "Other"), sep = " ", remove = FALSE)
stat.test_cl1_cl6_signif_sep <- separate(stat.test_cl1_cl6_signif_sep,
    Other, into = c("Chain_length", "Double_bounds"), sep = ":")
```

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

```
stat.test_cl1_cl6_all_lipids <- as.data.frame(table(stat.test_cl1_cl6_sep$Class))
stat.test_cl1_cl6_all_lipid_signif <- as.data.frame(table(stat.test_cl1_cl6_signif_sep$Class))
lipids_stat.test_cl1_cl6 <- merge(stat.test_cl1_cl6_all_lipids,
    stat.test_cl1_cl6_all_lipid_signif, by = "Var1")
lipids_stat.test_cl1_cl6$percentage <- (lipids_stat.test_cl1_cl6$Freq.y/lipids_stat.test_cl1_cl6$Freq.x)
    100
colnames(lipids_stat.test_cl1_cl6)[1] <- "Class"
```

Let's perform enrichment analysis:

```
results_enrichment_cl1_cl6 <- data.frame(Class = character(),
    p_value = numeric(), adjusted_p_value = numeric(), stringsAsFactors = FALSE)

lipid_classes <- unique(stat.test_cl1_cl6_signif_sep$Class)

for (lipid_class in lipid_classes) {
    q <- sum(stat.test_cl1_cl6_signif_sep$Class == lipid_class)
    m <- nrow(stat.test_cl1_cl6_signif_sep)
    n <- nrow(stat.test_cl1_cl6_sep) - m
    k <- sum(stat.test_cl1_cl6_sep$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_cl1_cl6 <- rbind(results_enrichment_cl1_cl6,
        data.frame(Class = lipid_class, p_value = p_value, adjusted_p_value = adjusted_p_value))
}

print(results_enrichment_cl1_cl6)
```

```
##      Class      p_value adjusted_p_value
## 1      CAR 0.091117585      0.091117585
## 2       CE 0.753480109      0.753480109
## 3      Cer 0.744434176      0.744434176
## 4      DAG 0.750250026      0.750250026
## 5      LPC 0.339476728      0.339476728
## 6      LPE 0.952640605      0.952640605
## 7       PC 0.989954221      0.989954221
## 8     PC-O 0.400964809      0.400964809
## 9       PE 0.944949832      0.944949832
## 10    PE-P 0.256565662      0.256565662
## 11      SM 0.821506620      0.821506620
## 12      TG 0.001192339      0.001192339
```

Plot the results:

```
significant_results_cl1_cl6 <- results_enrichment_cl1_cl6[results_enrichment_cl1_cl6$adjusted_p_value <
    0.05, ]
significant_results2_cl1_cl6 <- results_enrichment_cl1_cl6[results_enrichment_cl1_cl6$adjusted_p_value
    0.1, ]

base_palette <- brewer.pal(11, "Spectral")

extended_palette <- c(base_palette, "pink")

ggplot(lipids_stat.test_cl1_cl6, aes(x = Class, y = percentage,
    fill = Class)) + geom_bar(stat = "identity") + scale_fill_manual(values = extended_palette) +
    labs(x = "Class", y = "Percentage") + theme(axis.title.x = element_text(size = 13),
    axis.title.y = element_text(size = 13), legend.text = element_text(size = 13),
    legend.title = element_text(size = 13), axis.text.x = element_text(size = 13),
    axis.text.y = element_text(size = 13)) + geom_text(data = significant_results_cl1_cl6,
    aes(x = Class, y = 0, label = paste0("*")), vjust = -12.9,
    size = 9, color = "red", fontface = "bold") + geom_text(data = significant_results2_cl1_cl6,
    aes(x = Class, y = 0, label = paste0("  *")), vjust = -12.9,
    size = 9, color = "red", fontface = "bold") + geom_hline(yintercept = 45,
    linetype = "dashed", color = "black")
```
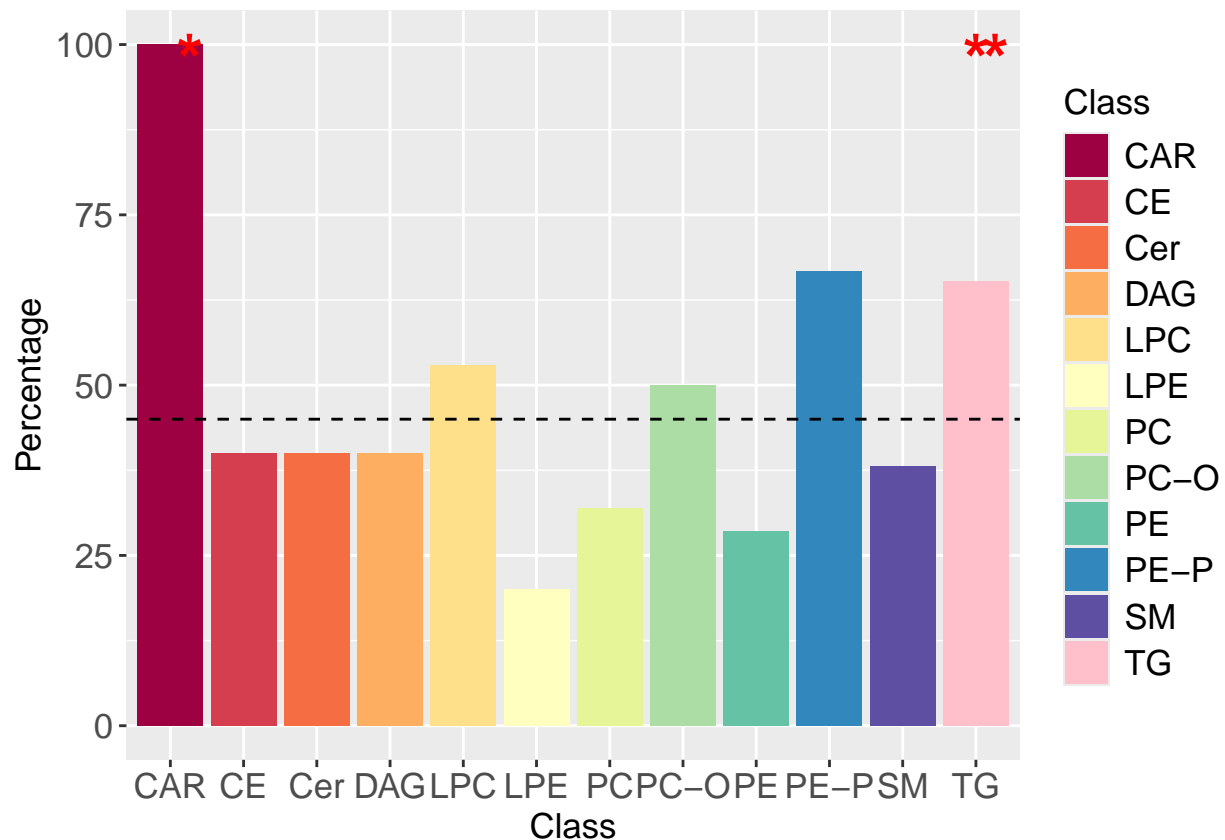
It can be seen that lipids differing significantly between the clusters 1 and 6 belong to different classes. In addition, compared to the expected lipid levels, there is enrichment of such lipid classes as CAR and TG.

Let's explore differences in lipid content between clusters:

Prepare the data:

```
mean_1_6_signif <- mean_1_6 %>%
    filter(lipid_features %in% stat.test_cl1_cl6_signif$lipid_features)

mean_1_6_signif$Sign <- ifelse(mean_1_6_signif$FC_1_6 > 0, "Positive",
    "Negative")

mean_1_6_signif$Count <- 1

mean_1_6_signif$Value <- ifelse(mean_1_6_signif$Sign == "Negative",
    -mean_1_6_signif$Count, mean_1_6_signif$Count)

mean_1_6_signif$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", mean_1_6_signif$lipid_features)
mean_1_6_signif <- separate(mean_1_6_signif, lipid_features,
    into = c("Class", "Other"), sep = " ", remove = FALSE)
mean_1_6_signif <- separate(mean_1_6_signif, Other, into = c("Chain_length",
    "Double_bounds"), sep = ":")
```

Plot the results:

```
ggplot(mean_1_6_signif, 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 in cluster 1",
    Negative = "Lipids decreased \n in cluster 1")) + 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))
```
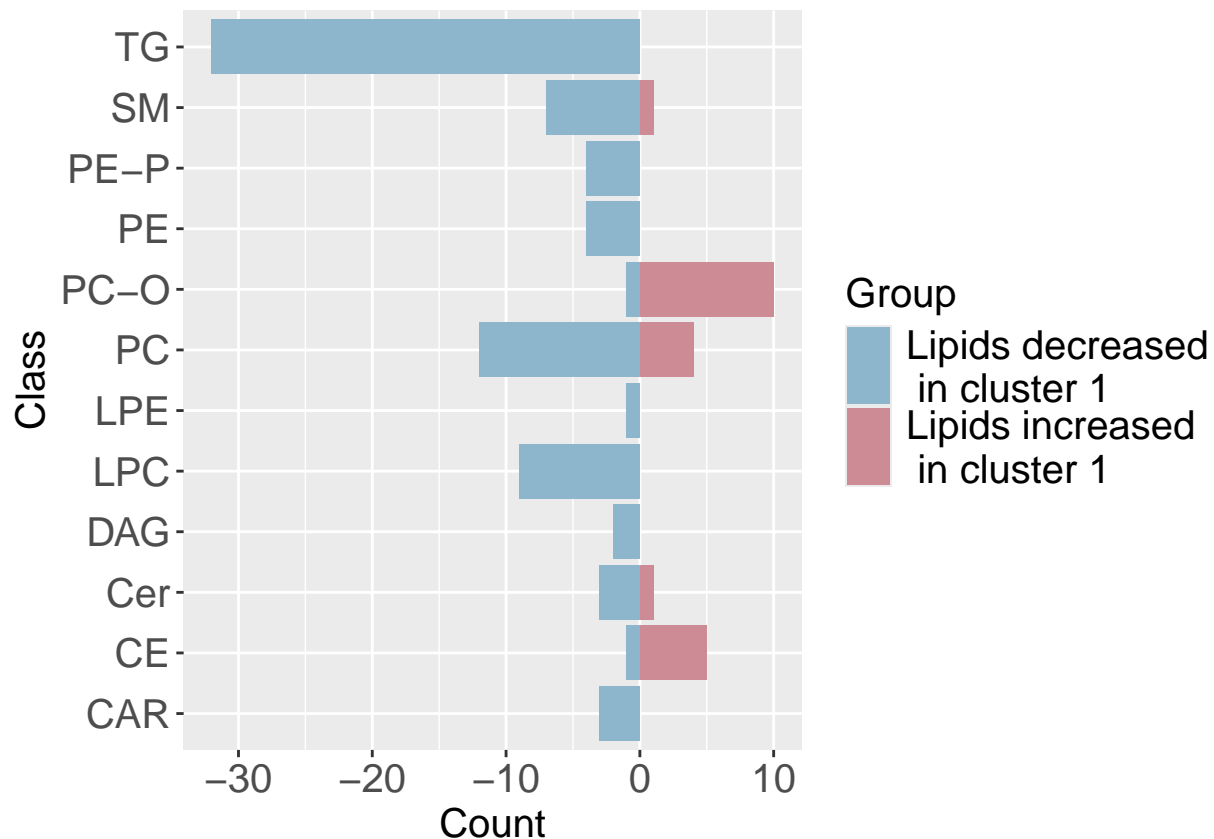
We can see that the majority of lipids are reduced in patients from less successfully aging cluster 1.

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

```r
unique_classes <- unique(stat.test_cl1_cl6_sep$Class)
stat.test_cl1_cl6_sep$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", stat.test_cl1_cl6_sep$lipid_features)
mean_1_6$lipid_features <- gsub("LPC O-(\\d+):(\\d+)", "LPC-O \\1:\\2",
    mean_1_6$lipid_features)


for (current_class in unique_classes) {

    stat.test_lipid <- stat.test_cl1_cl6_sep[stat.test_cl1_cl6_sep$Class ==
        current_class, ]

    merged_bounds_FC <- merge(mean_1_6, stat.test_lipid, by = "lipid_features")

    if (current_class %in% c("CAR", "CE", "LPC", "LPE", "LPC-O")) {
        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-O", "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_1_6 ~ 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_1_6, color = ifelse(lipid_features %in% stat.test_cl1_cl6_signif_sep$lipid_features,
                "TRUE", "FALSE"))) + geom_point(shape = 16, size = 3,
            alpha = 0.5) + labs(x = "Count of the double bounds",
            y = "Fold between cl1 and cl6", title = paste("Number of double bounds in a class",
                current_class)) + geom_vline(xintercept = 0,
            linetype = "dashed", color = "black") + geom_hline(yintercept = 0,
            linetype = "dashed", color = "black") + scale_color_manual(values = c(`TRUE` = "red",
            `FALSE` = "blue"), name = "Significance", labels = if (all(merged_bounds_FC$lipid_features
            stat.test_cl1_cl6_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_1_6), label = paste("p-value: ",
                formatC(p_value, digits = 3, format = "f")),
            hjust = 1, vjust = 0, size = 6.5, color = "black")
        print(plot)

    }
}
```
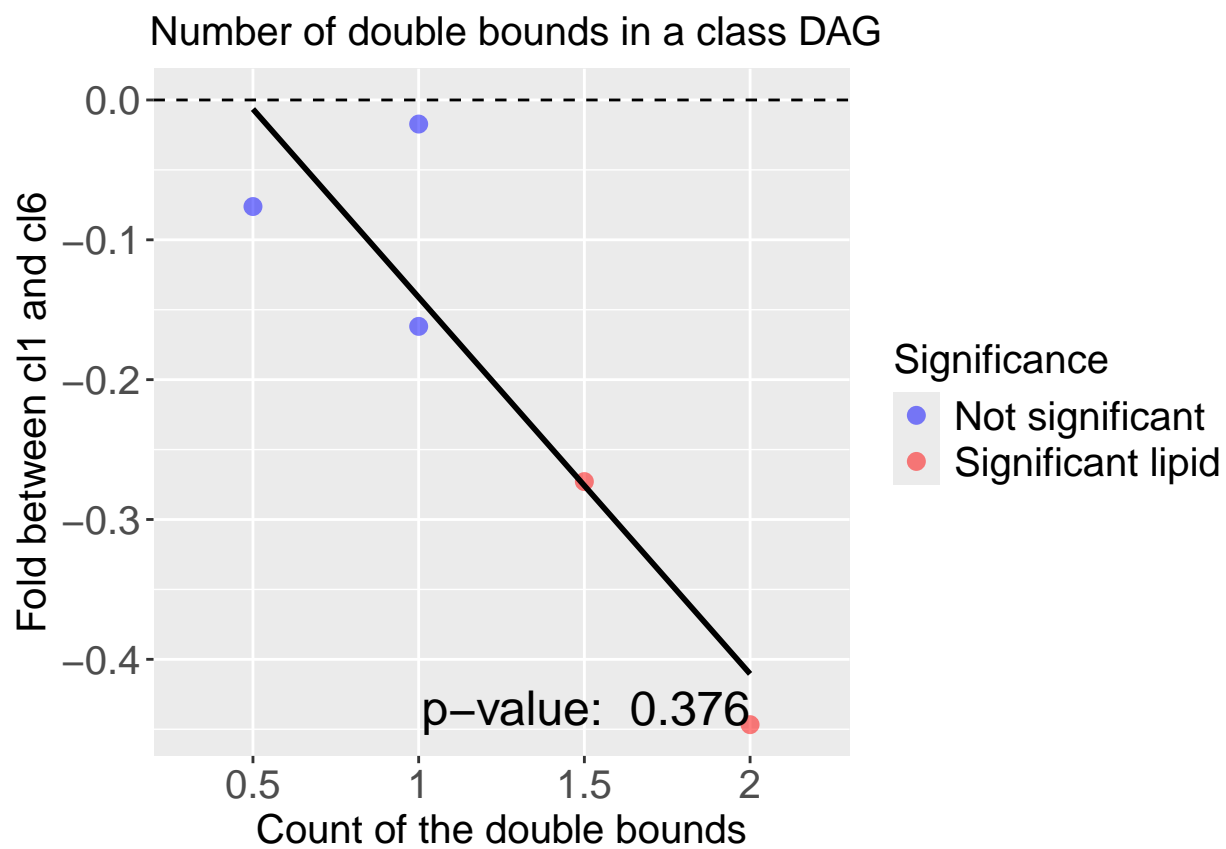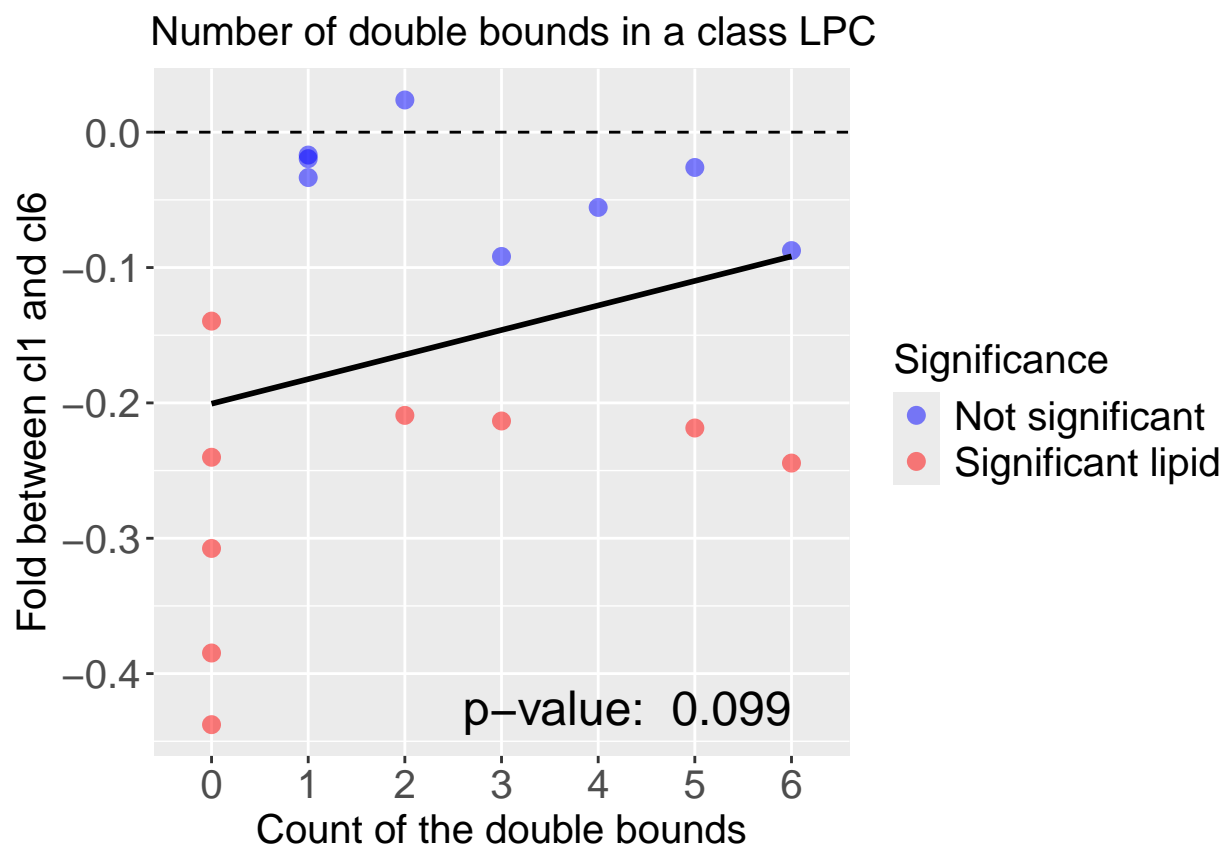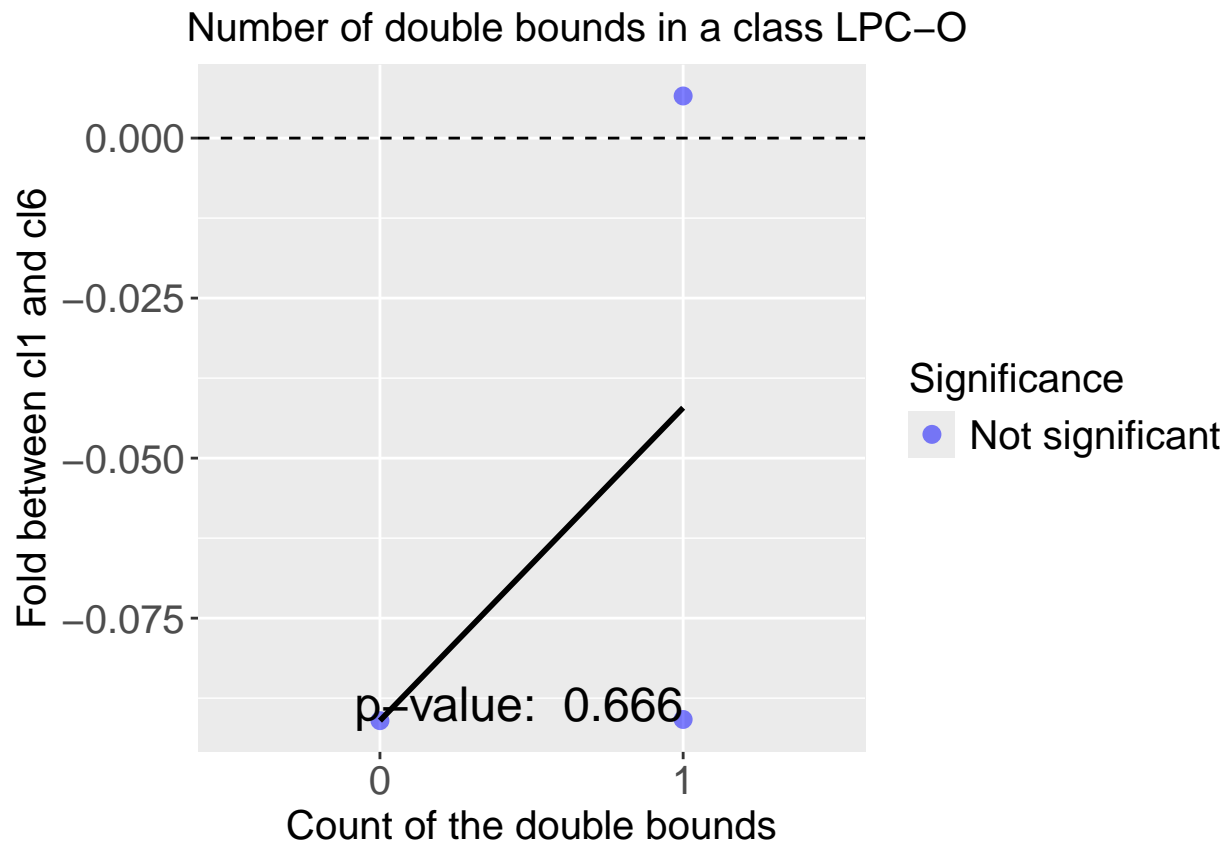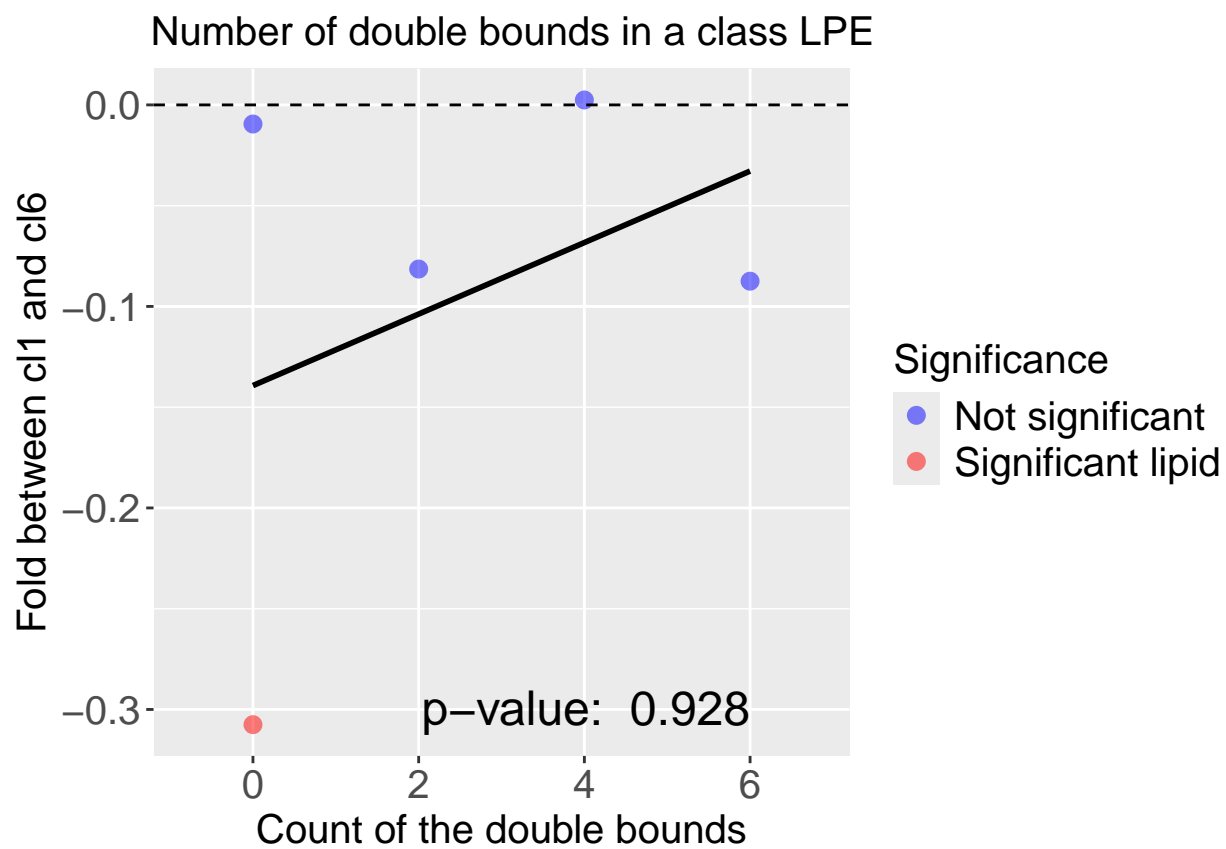
Number of double bounds in a class CAR

Fold between cl1 and cl6

Count of the double bounds

Significance

● Significant lipid

p−value: 0.229

Number of double bounds in a class CE

Fold between cl1 and cl6 / Count of the double bounds

p–value: 0.486

Significance
- Not significant
- Significant lipid

Number of double bounds in a class Cer

Number of double bounds in a class DAG

Fold between cl1 and cl6

Count of the double bounds

Significance
● Not significant
● Significant lipid

p−value: 0.376

Number of double bounds in a class LPC

Significance
- Not significant
- Significant lipid

p–value: 0.099

Fold between cl1 and cl6

Count of the double bounds

Number of double bounds in a class LPC−O

Number of double bounds in a class LPE

Fold between cl1 and cl6

Count of the double bounds

Significance
- Not significant
- Significant lipid

p−value: 0.928

Number of double bounds in a class PC

p-value: 0.028

Number of double bounds in a class PC–O

Fold between cl1 and cl6

Count of the double bounds

Significance
- Not significant
- Significant lipid

p–value: 0.026

Number of double bounds in a class PE

Number of double bounds in a class PE−P

Number of double bounds in a class SM

p-value: 0.636

Number of double bounds in a class TG

Based on these plots, we can conclude that the level of polyunsaturated lipids is reduced among patients from less successfully aging cluster 1.

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

```
unique_classes <- unique(stat.test_cl1_cl6_sep$Class)

for (current_class in unique_classes) {

    stat.test_lipid <- stat.test_cl1_cl6_sep[stat.test_cl1_cl6_sep$Class ==
        current_class, ]
    merged_bounds_FC <- merge(mean_1_6, stat.test_lipid, by = "lipid_features")

    stat.test_lipid_signif <- stat.test_cl1_cl6_signif_sep[stat.test_cl1_cl6_signif_sep$Class ==
        current_class, ]
    merged_bounds_FC_signif <- merge(mean_1_6, stat.test_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_1_6), pch = 16, cex = 5,
    alpha = 1) + labs(x = "Chain length", y = "Double bonds",
    color = "FC_1_6") + ggtitle(paste("Class:", current_class)) +
    scale_color_gradient2(low = "blue", mid = "white", high = "red",
        midpoint = 0, name = "Fold change \n between cl1 and cl6 ") +
    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: CE

Class: Cer

Double bonds

Chain length

Fold change
between cl1 and cl6

0.1
0.0
−0.1
−0.2
−0.3

Class: LPC

Class: LPC–O

Double bonds

Chain length

Class: LPE

Fold change
between cl1 and cl6

−0.3075334

Double bonds

Chain length

Class: PC

Class: PC−O

Class: PE

Class: PE−P

Class: PI
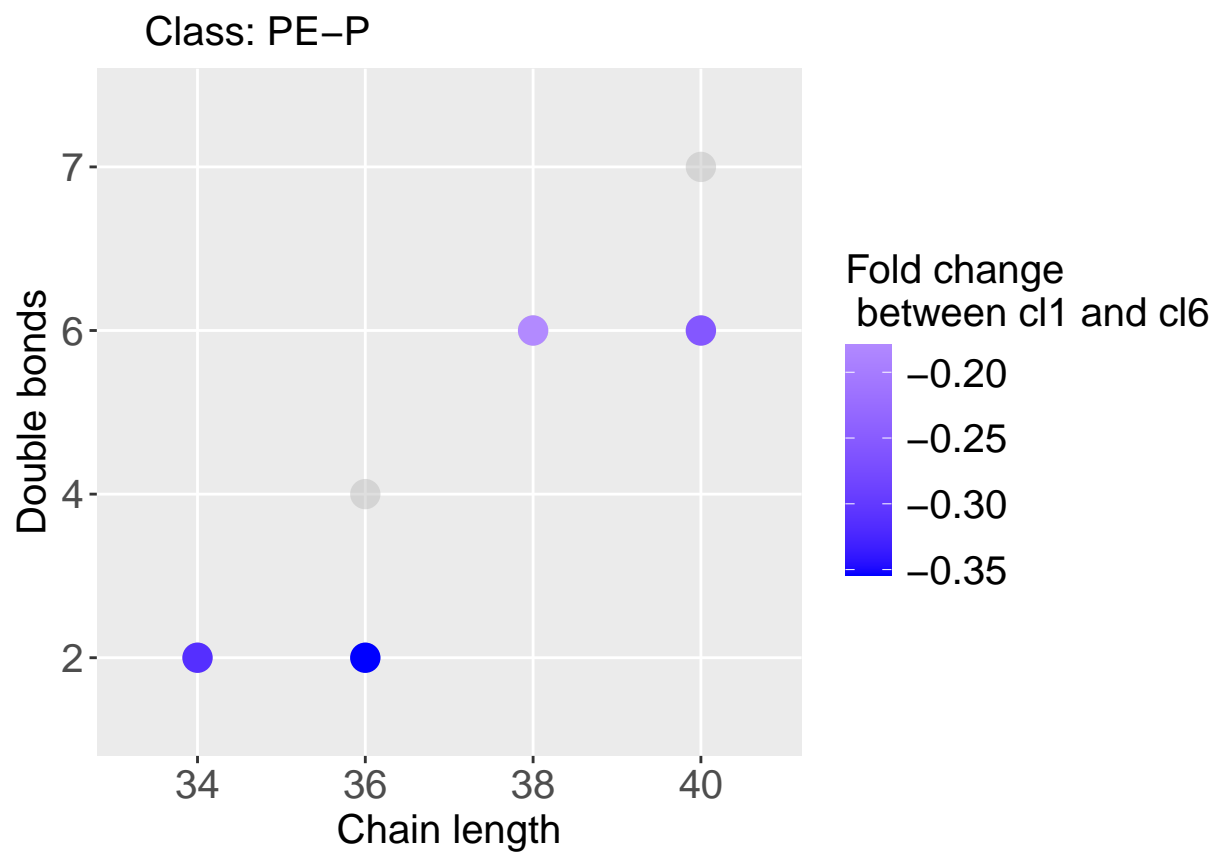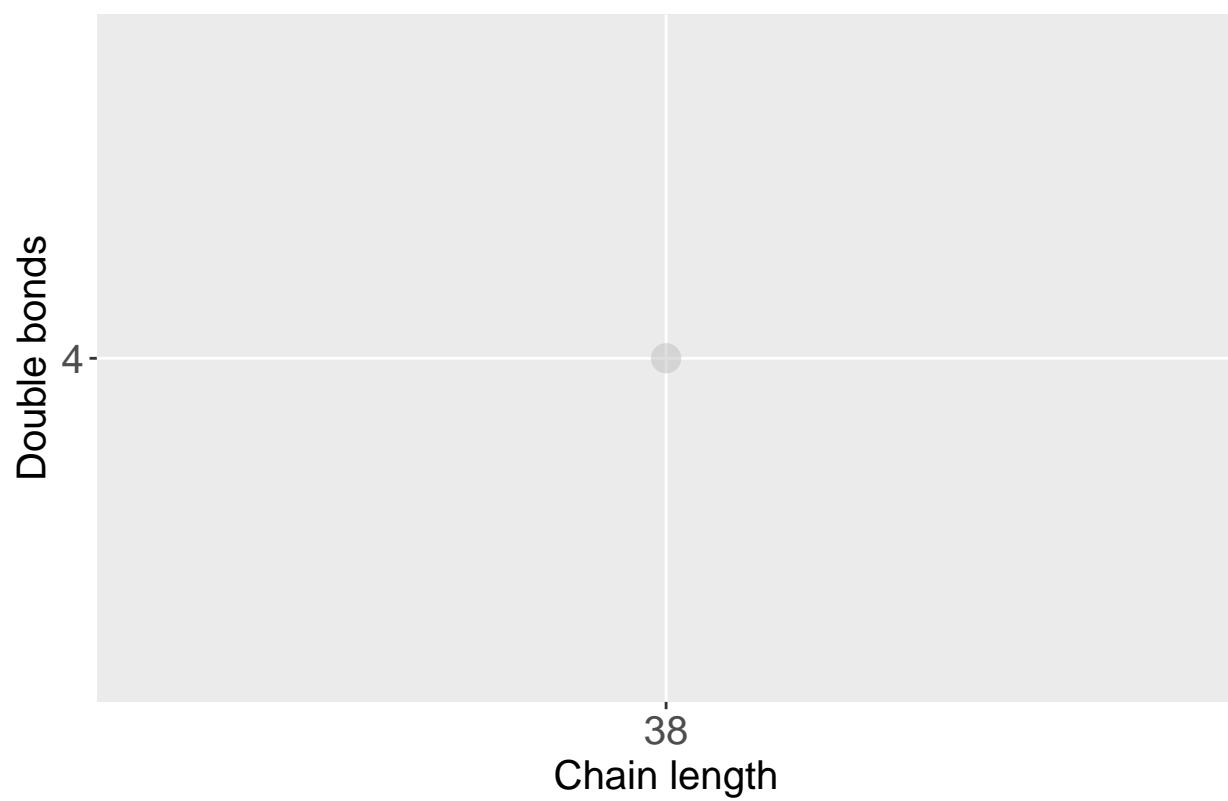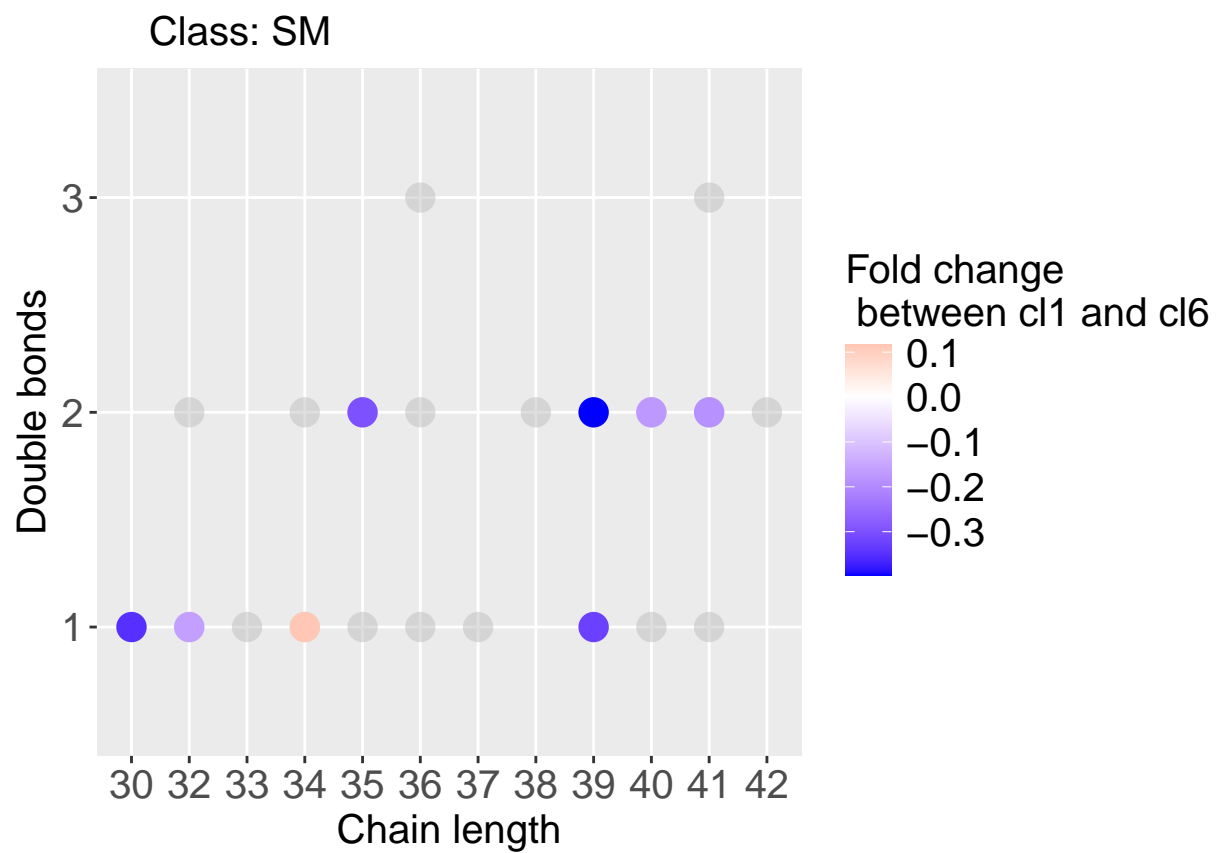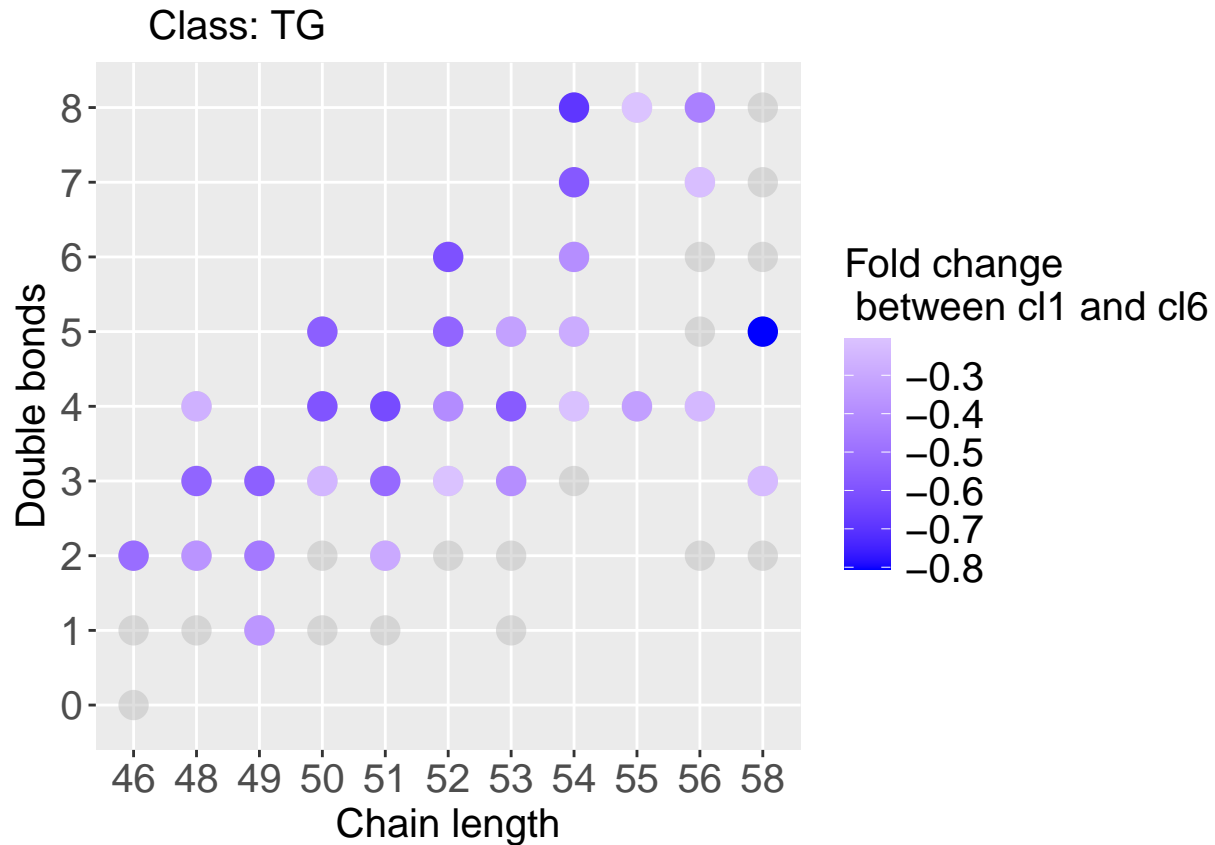
Double bonds

4 -

38

Chain length

Class: SM

Class: TG
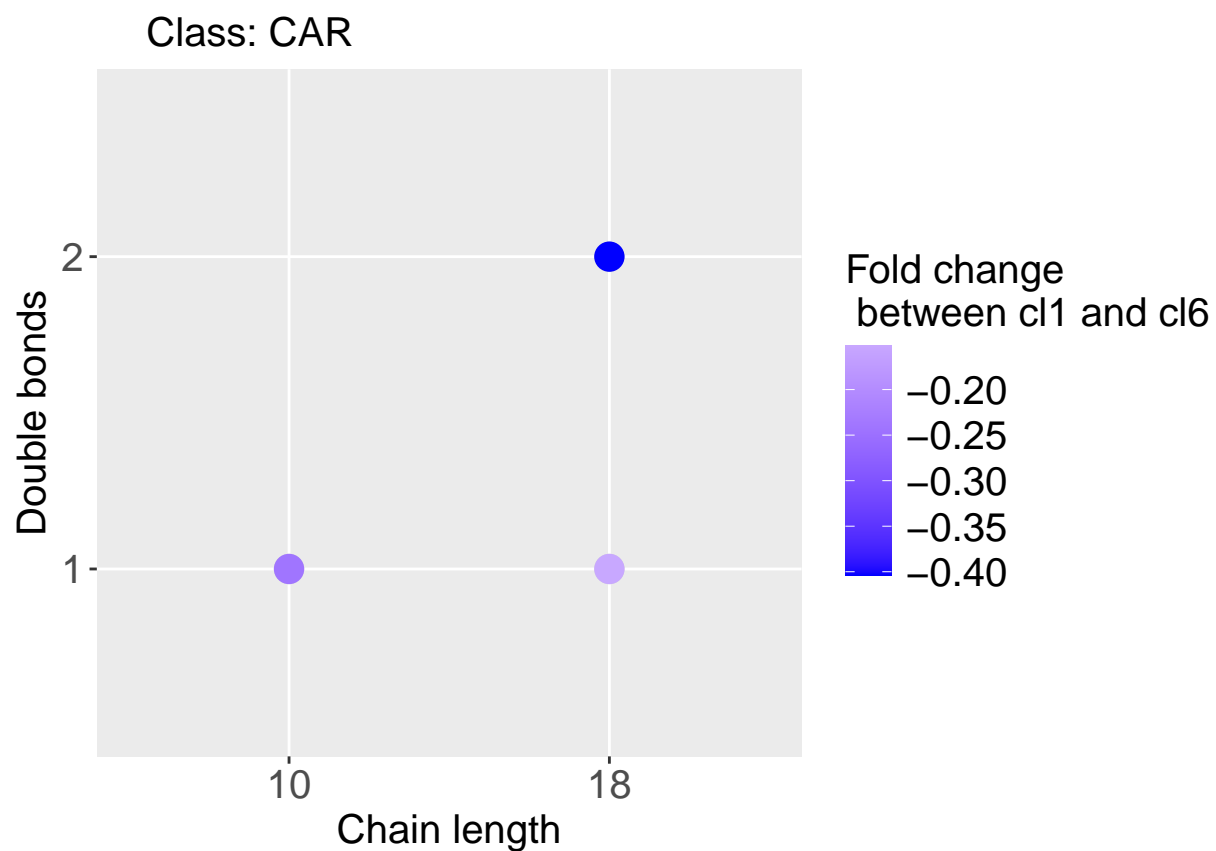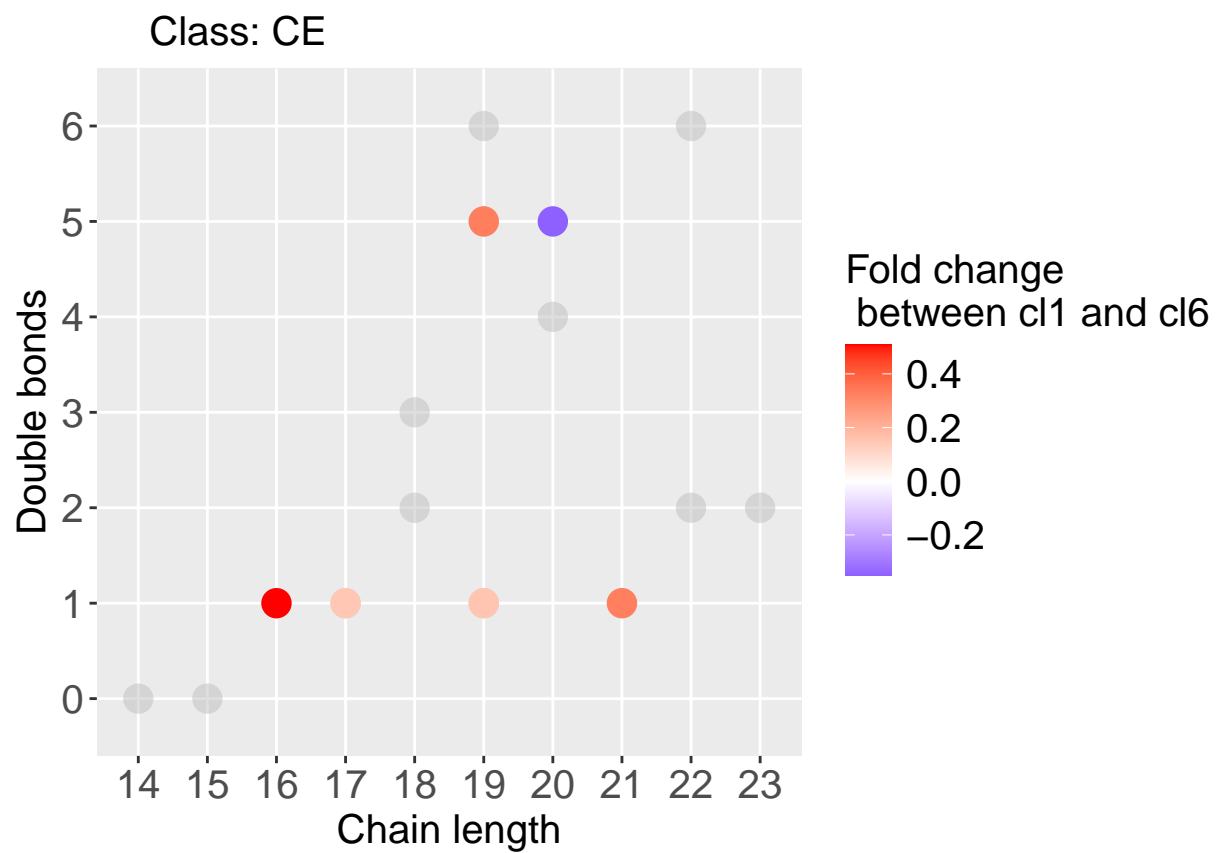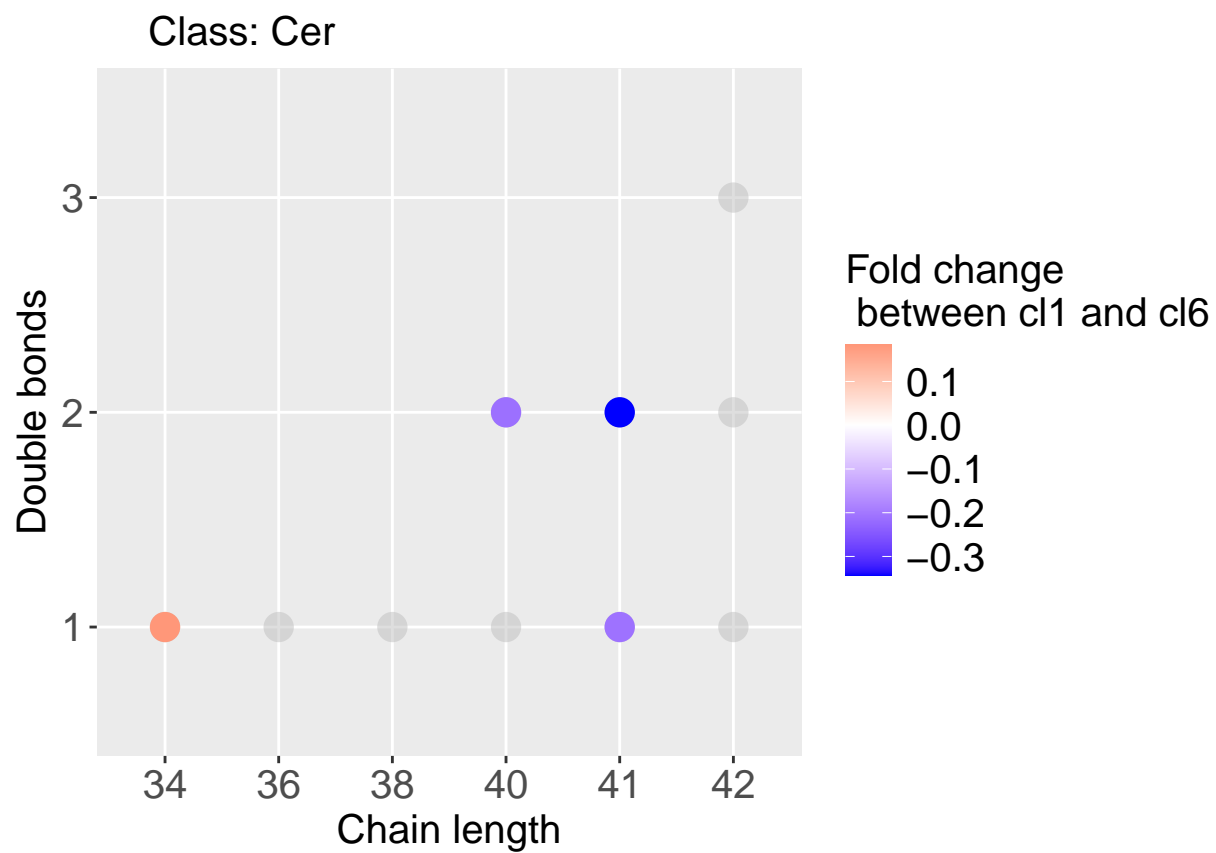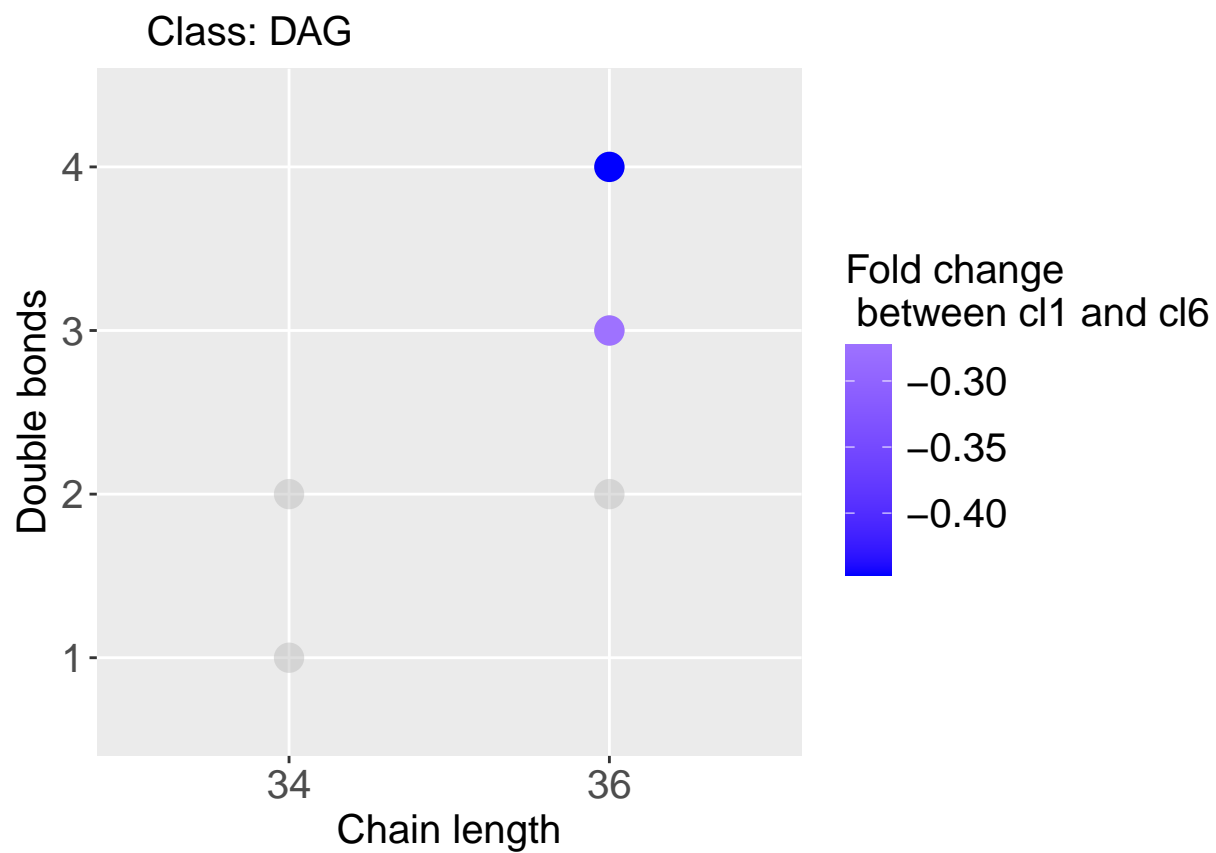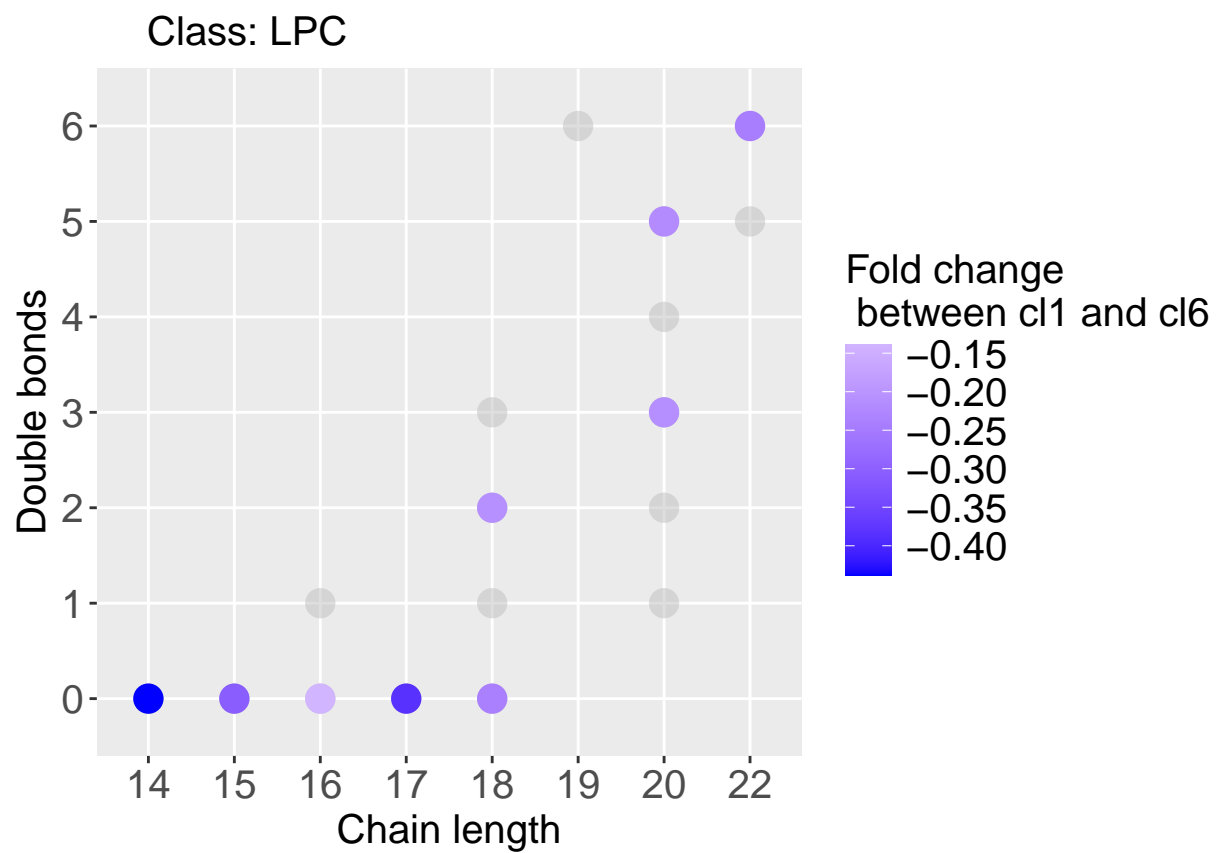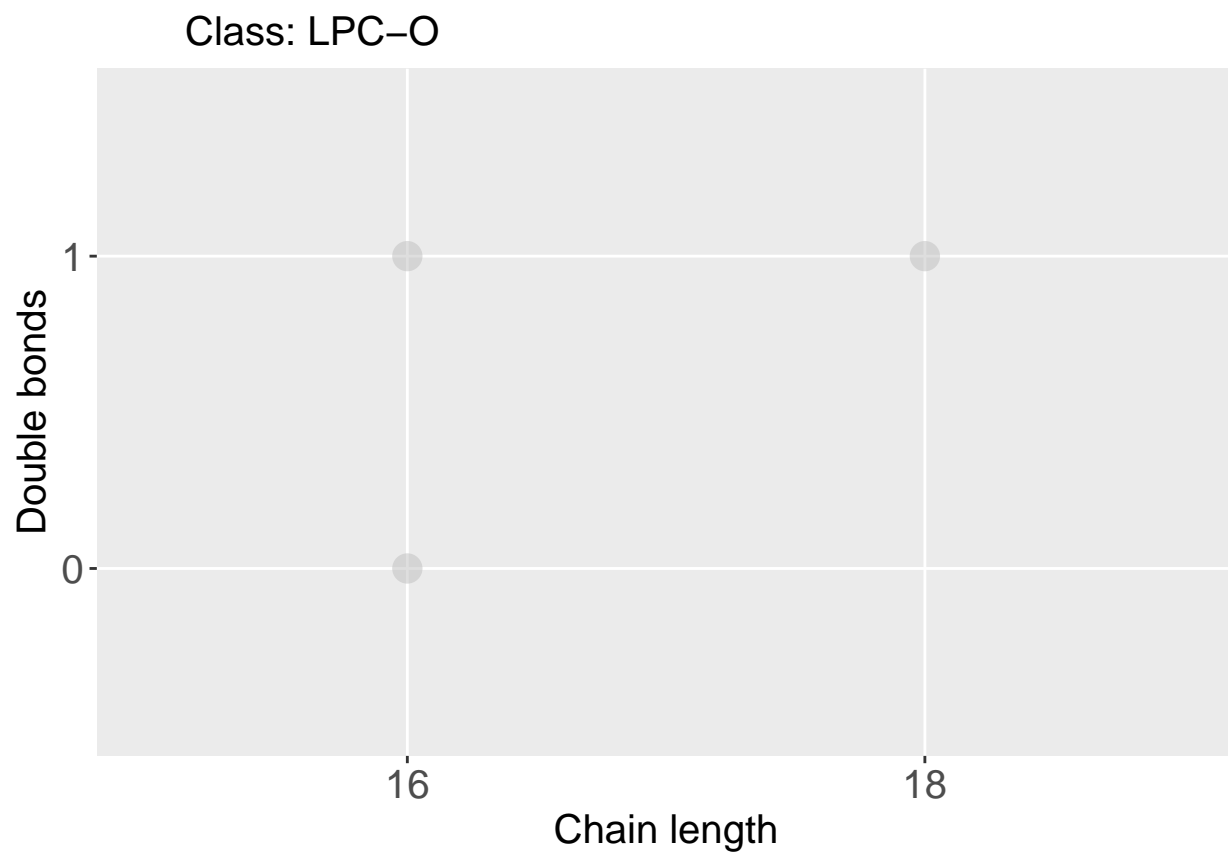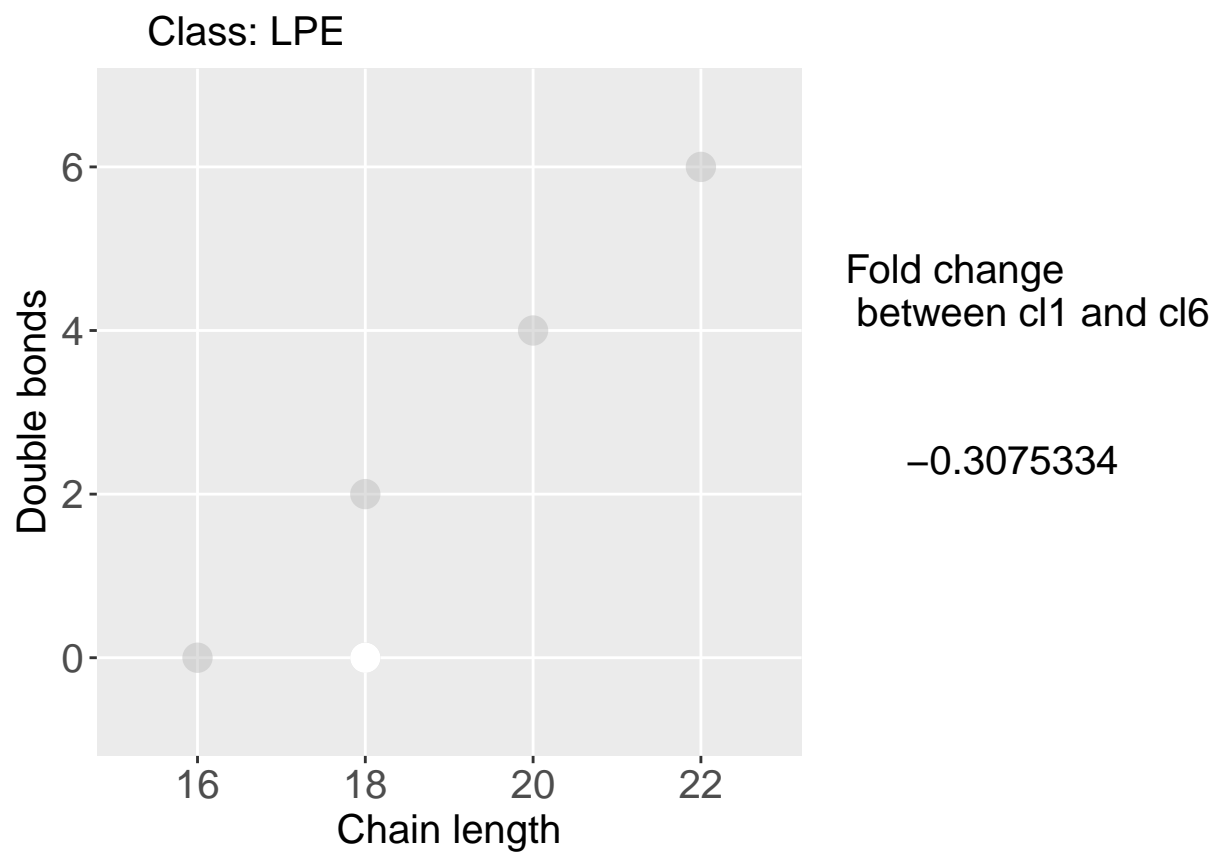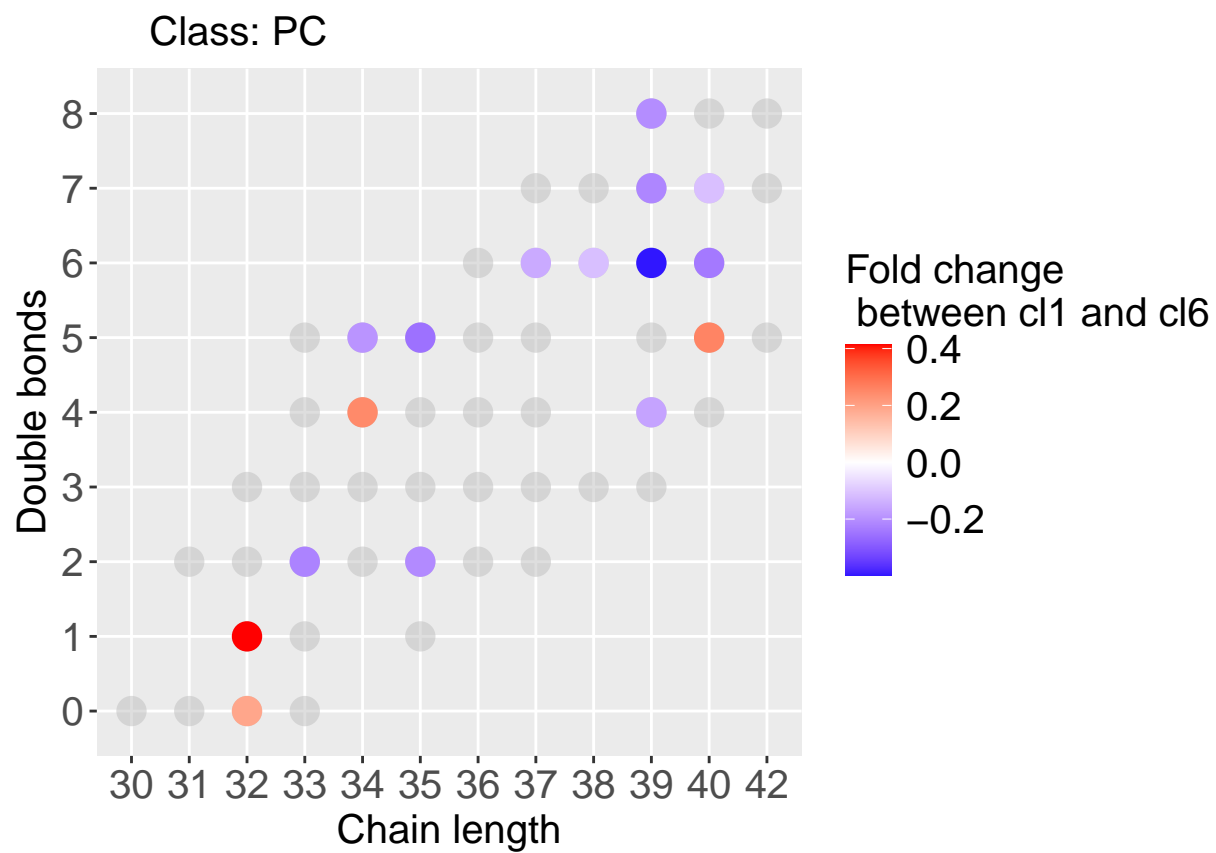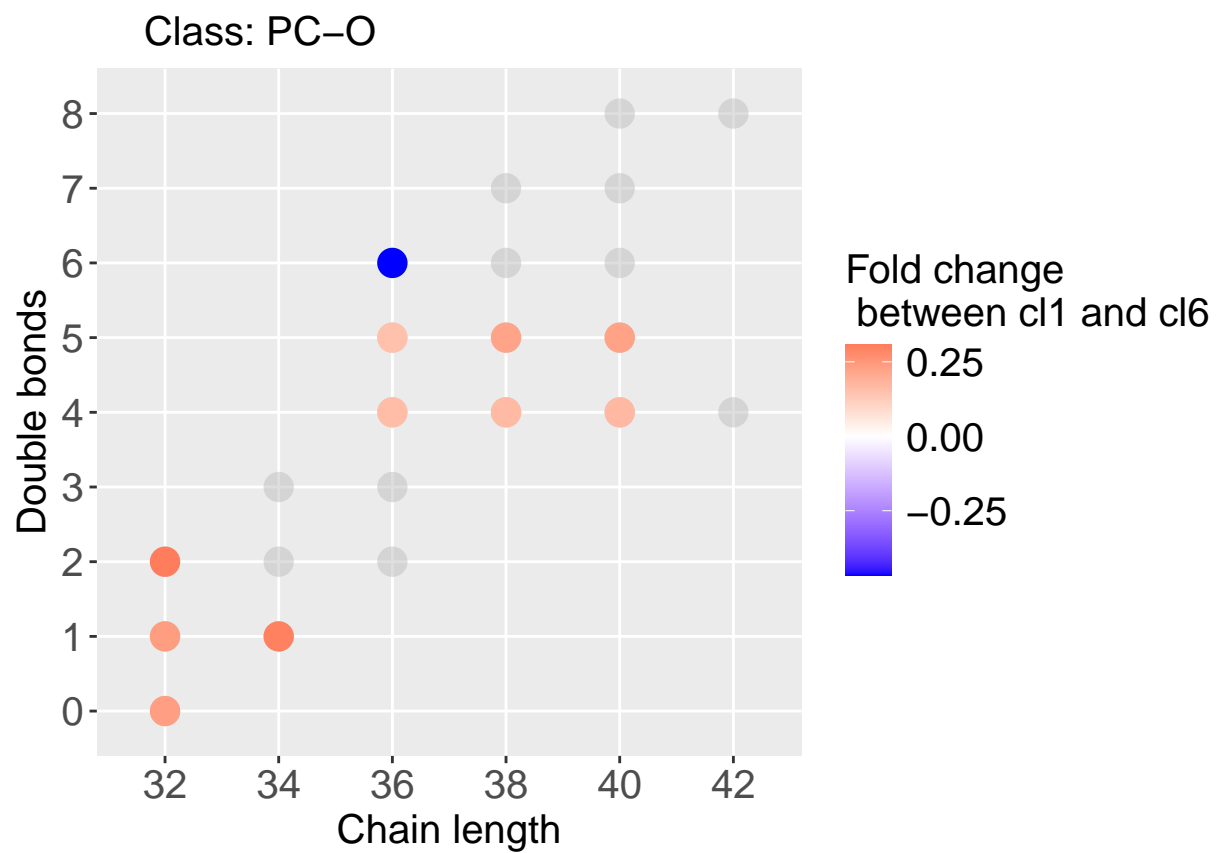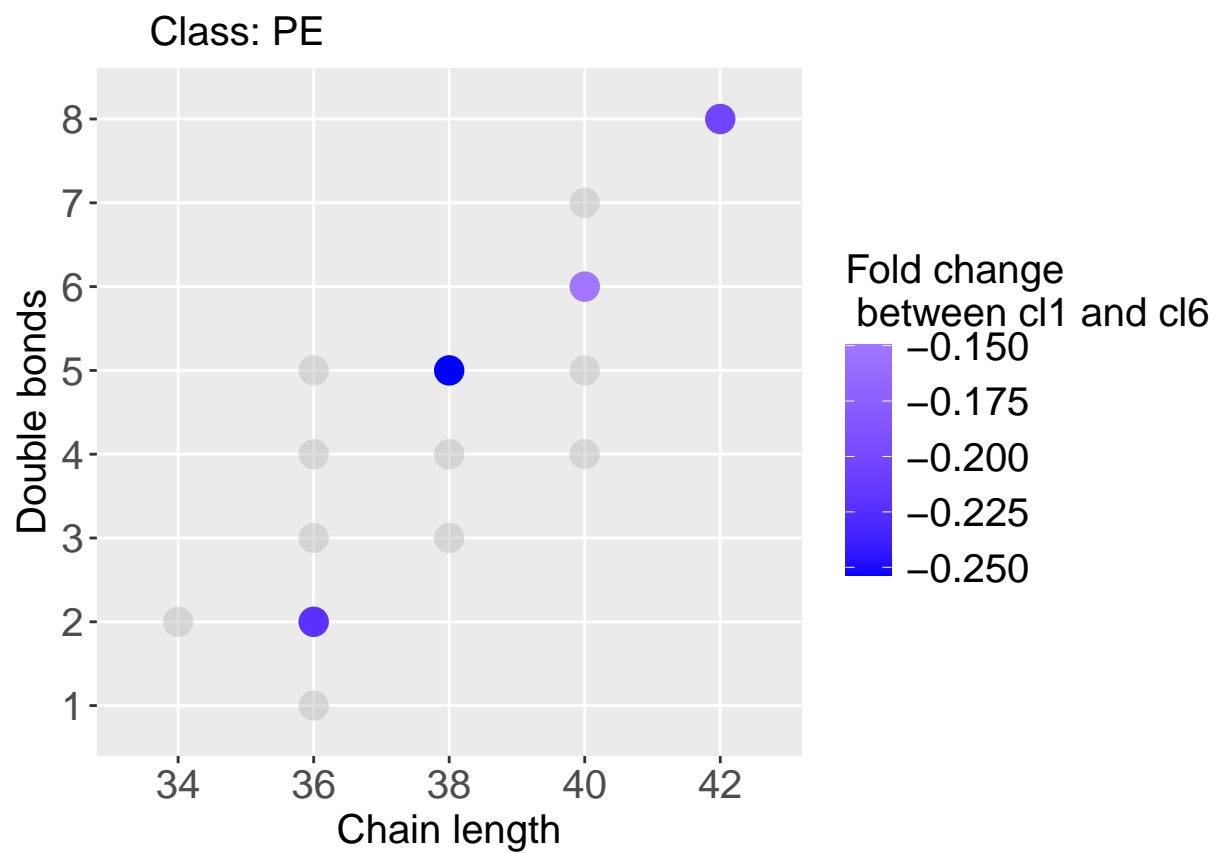
So we can also say that the number of long chain lipids is also reduced in patients from less successfully aging cluster 1.

Let's take a closer look at the differences in lipid profile between clusters 1 and 6, 4 and 6:

Parse the data:

```
stat.test_cl1_cl6_signif <- test_results_significant$stat.test_cl1_6
stat.test_cl4_cl6_signif <- test_results_significant$stat.test_cl4_6

cl1_long <- scales_lipids_with_clusters_long[scales_lipids_with_clusters_long$Cluster ==
    1, ]

cl4_long <- scales_lipids_with_clusters_long[scales_lipids_with_clusters_long$Cluster ==
    4, ]

cl6_long <- scales_lipids_with_clusters_long[scales_lipids_with_clusters_long$Cluster ==
    6, ]
```

Count the mean:

```
cl1_mean <- cl1_long %>%
    group_by(lipid_features) %>%
    summarise(mean_cl1 = mean(values))

cl4_mean <- cl4_long %>%
    group_by(lipid_features) %>%
    summarise(mean_cl4 = mean(values))

cl6_mean <- cl6_long %>%
    group_by(lipid_features) %>%
    summarise(mean_cl6 = mean(values))
```

Count the fold change:

```
mean_1_6 <- merge(cl1_mean, cl6_mean, by = "lipid_features")
mean_1_6$FC_1_6 <- mean_1_6$mean_cl1 - mean_1_6$mean_cl6

mean_4_6 <- merge(cl4_mean, cl6_mean, by = "lipid_features")
mean_4_6$FC_4_6 <- mean_4_6$mean_cl4 - mean_4_6$mean_cl6

FC_16_46 <- merge(mean_1_6, mean_4_6, by = "lipid_features")
```

Merging dataframes into one:

```
FC_16_46_signif1_6 <- merge(FC_16_46, stat.test_cl1_cl6_signif,
    by = "lipid_features")
FC_16_46_signif4_6 <- merge(FC_16_46, stat.test_cl4_cl6_signif,
    by = "lipid_features")
data16_46 = merge(stat.test_cl1_cl6_signif, stat.test_cl4_cl6_signif,
    by = "lipid_features")

FC_16_46_signif_both <- merge(FC_16_46, data16_46, by = "lipid_features")
```

Plot the results:

```
data1 <- FC_16_46_signif1_6
data2 <- FC_16_46_signif4_6
data3 <- FC_16_46_signif_both


ggplot() + geom_point(data = data1, aes(x = FC_1_6, y = FC_4_6,
    color = "Significant lipids \n comparing clusters 1 and 6"),
    pch = 16, cex = 3, alpha = 0.5) + geom_point(data = data2,
    aes(x = FC_1_6, y = FC_4_6, color = "Significant lipids \n comparing clusters 4 and 6"),
    pch = 16, cex = 3, alpha = 0.5) + geom_point(data = data3,
```

```r
    aes(x = FC_1_6, y = FC_4_6, color = "Intersection"), pch = 16,
    cex = 3, alpha = 0.5) + labs(x = "Fold change between cl1 and cl6",
    y = "Fold change between \n cl4 and cl6", color = "Group") +
    geom_vline(xintercept = 0, linetype = "dashed", color = "black") +
    geom_hline(yintercept = 0, linetype = "dashed", color = "black") +
    geom_abline(slope = 1, intercept = 0, color = "black") +
    scale_color_manual(values = c("#FF9999", "blue", "#8FBC8F")) +
    guides(color = guide_legend(title = NULL)) + theme_minimal() +
    theme(legend.position = "right", legend.text = element_text(size = 15),
        axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15),
        axis.title.x = element_text(size = 15), axis.title.y = element_text(size = 15))
```
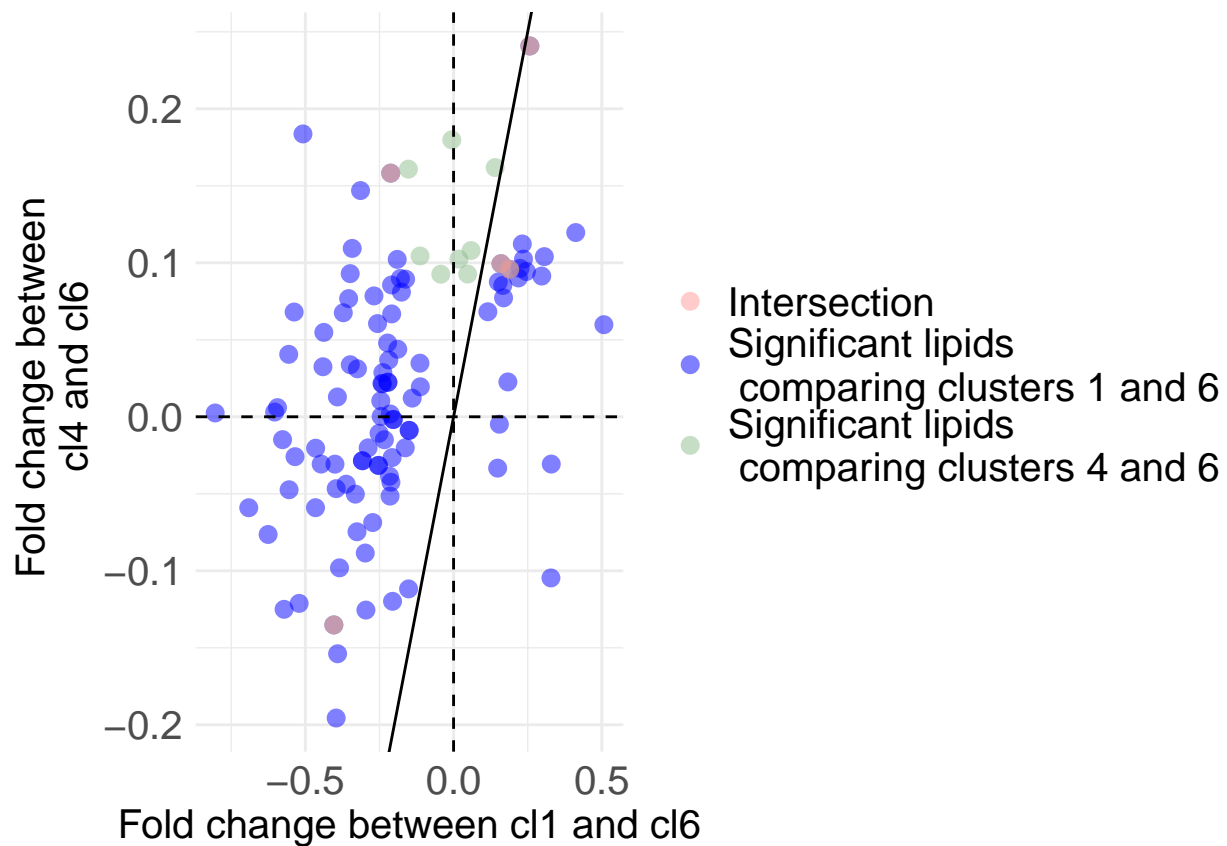

Let's do a correlation test:

```r
merged_FC_16_46 <- rbind(FC_16_46_signif1_6, FC_16_46_signif4_6)
unique_FC_16_46 <- merged_FC_16_46[!duplicated(merged_FC_16_46$lipid_features),
    ]

cor_test_result_46_16 <- cor.test(unique_FC_16_46$FC_1_6, unique_FC_16_46$FC_4_6)
cor_test_result_46_16
```

```
##
##  Pearson's product-moment correlation
##
## data:  unique_FC_16_46$FC_1_6 and unique_FC_16_46$FC_4_6
## t = 5.046, df = 106, p-value = 1.879e-06
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2739045 0.5807714
## sample estimates:
##       cor
## 0.4400985
```

In this case, the significant lipids that resulted when comparing clusters 4 and 6 were not the same as when comparing clusters 1 and 6.

Let's look at the significant lipids that resulted from comparing clusters 4 and 6:

Let's parse the data on lipids:

```
stat.test_cl4_cl6 <- test_results$stat.test_cl4_6

stat.test_cl4_cl6$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", stat.test_cl4_cl6$lipid_features)
stat.test_cl4_cl6_sep <- separate(stat.test_cl4_cl6, lipid_features,
    into = c("Class", "Other"), sep = " ", remove = FALSE)
stat.test_cl4_cl6_sep <- separate(stat.test_cl4_cl6_sep, Other,
    into = c("Chain_length", "Double_bounds"), sep = ":")

stat.test_cl4_cl6_signif$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", stat.test_cl4_cl6_signif$lipid_features)
stat.test_cl4_cl6_signif_sep <- separate(stat.test_cl4_cl6_signif,
    lipid_features, into = c("Class", "Other"), sep = " ", remove = FALSE)
stat.test_cl4_cl6_signif_sep <- separate(stat.test_cl4_cl6_signif_sep,
    Other, into = c("Chain_length", "Double_bounds"), sep = ":")
```

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

```
stat.test_cl4_cl6_all_lipids <- as.data.frame(table(stat.test_cl4_cl6_sep$Class))
stat.test_cl4_cl6_all_lipid_signif <- as.data.frame(table(stat.test_cl4_cl6_signif_sep$Class))
lipids_stat.test_cl4_cl6 <- merge(stat.test_cl4_cl6_all_lipids,
    stat.test_cl4_cl6_all_lipid_signif, by = "Var1")
lipids_stat.test_cl4_cl6$percentage <- (lipids_stat.test_cl4_cl6$Freq.y/lipids_stat.test_cl4_cl6$Freq.x
    100
colnames(lipids_stat.test_cl4_cl6)[1] <- "Class"
```

Let's perform enrichment analysis:

```
results_enrichment_cl4_cl6 <- data.frame(Class = character(),
    p_value = numeric(), adjusted_p_value = numeric(), stringsAsFactors = FALSE)

lipid_classes <- unique(stat.test_cl4_cl6_signif_sep$Class)
```

```
for (lipid_class in lipid_classes) {
    q <- sum(stat.test_cl4_cl6_signif_sep$Class == lipid_class)
    m <- nrow(stat.test_cl4_cl6_signif_sep)
    n <- nrow(stat.test_cl4_cl6_sep) - m
    k <- sum(stat.test_cl4_cl6_sep$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_cl4_cl6 <- rbind(results_enrichment_cl4_cl6,
        data.frame(Class = lipid_class, p_value = p_value, adjusted_p_value = adjusted_p_value))
}

print(results_enrichment_cl4_cl6)
```

```
##   Class    p_value adjusted_p_value
## 1   CAR 0.16700608       0.16700608
## 2   Cer 0.01512226       0.01512226
## 3    PC 0.04696711       0.04696711
## 4  PC-O 0.37861835       0.37861835
## 5  PE-P 0.30792698       0.30792698
```

Plot the results:

```
significant_results_cl4_cl6 <- results_enrichment_cl4_cl6[results_enrichment_cl4_cl6$adjusted_p_value <
    0.05, ]
significant_results2_cl4_cl6 <- results_enrichment_cl4_cl6[results_enrichment_cl4_cl6$adjusted_p_value
    0.1, ]

#         'Spectral'  7
base_palette <- brewer.pal(11, "Spectral")

#
extended_palette <- c(base_palette, "pink")  #

ggplot(lipids_stat.test_cl4_cl6, aes(x = Class, y = percentage,
    fill = Class)) + geom_bar(stat = "identity") + scale_fill_manual(values = extended_palette) +
    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_cl4_cl6,
    aes(x = Class, y = 0, label = paste0("*")), vjust = -11.6,
    size = 9, color = "red", fontface = "bold") + geom_text(data = significant_results2_cl4_cl6,
    aes(x = Class, y = 0, label = paste0("  *")), vjust = -11.6,
    size = 9, color = "red", fontface = "bold") + geom_hline(yintercept = 6,
    linetype = "dashed", color = "black")
```
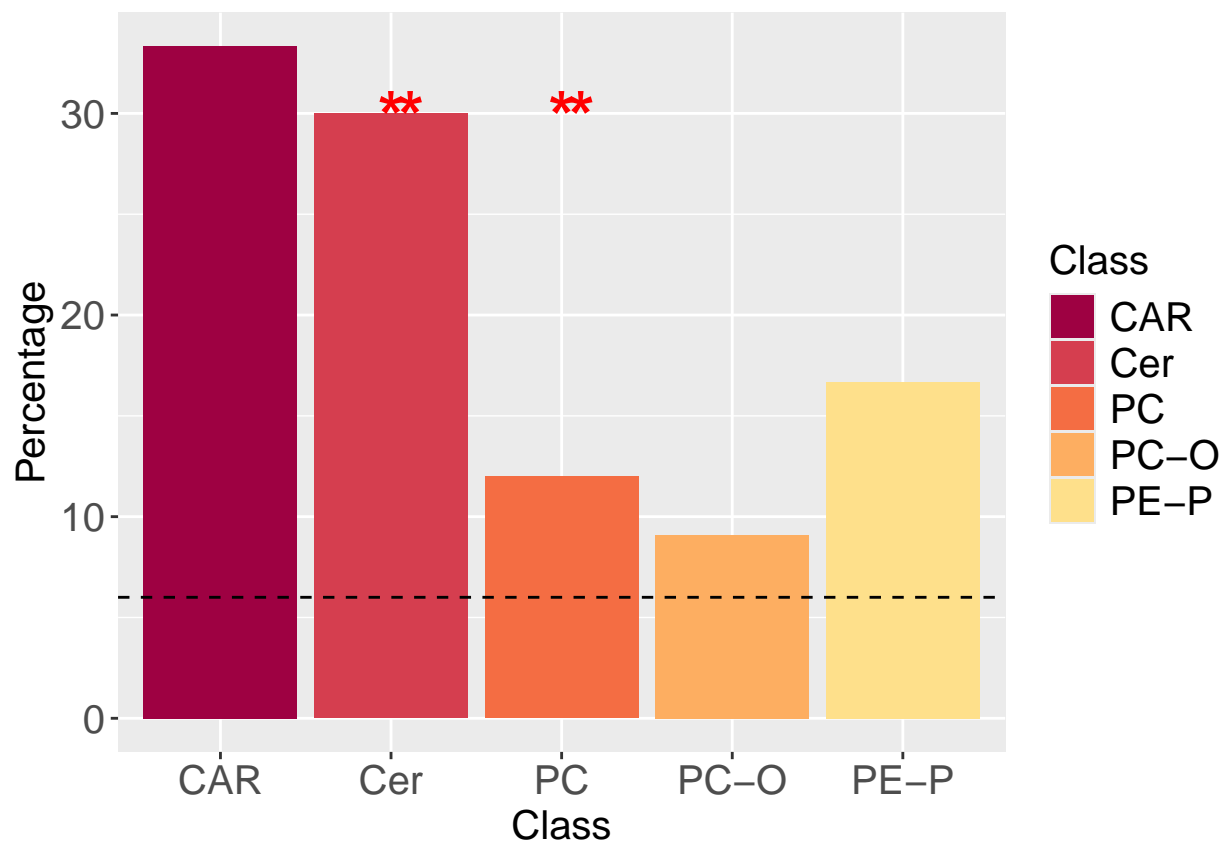
It can be seen that lipids differing significantly between the clusters 4 and 6 belong to different classes. In addition, compared to the expected lipid levels, there is enrichment of such lipid classes as Cer and PC.

Let's explore differences in lipid content between cluster groups:

Prepare the data:

```
mean_4_6_signif <- mean_4_6 %>%
    filter(lipid_features %in% stat.test_cl4_cl6_signif$lipid_features)

mean_4_6_signif$Sign <- ifelse(mean_4_6_signif$FC_4_6 > 0, "Positive",
    "Negative")

mean_4_6_signif$Count <- 1

mean_4_6_signif$Value <- ifelse(mean_4_6_signif$Sign == "Negative",
    -mean_4_6_signif$Count, mean_4_6_signif$Count)

mean_4_6_signif$lipid_features <- gsub("LPC O-(\\d+):(\\d+)",
    "LPC-O \\1:\\2", mean_4_6_signif$lipid_features)
mean_4_6_signif <- separate(mean_4_6_signif, lipid_features,
    into = c("Class", "Other"), sep = " ", remove = FALSE)
```

```
mean_4_6_signif <- separate(mean_4_6_signif, Other, into = c("Chain_length",
    "Double_bounds"), sep = ":")
```

Plot the results:

```
ggplot(mean_4_6_signif, 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 in cluster 4",
    Negative = "Lipids decreased \n in cluster 4")) + 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))
```
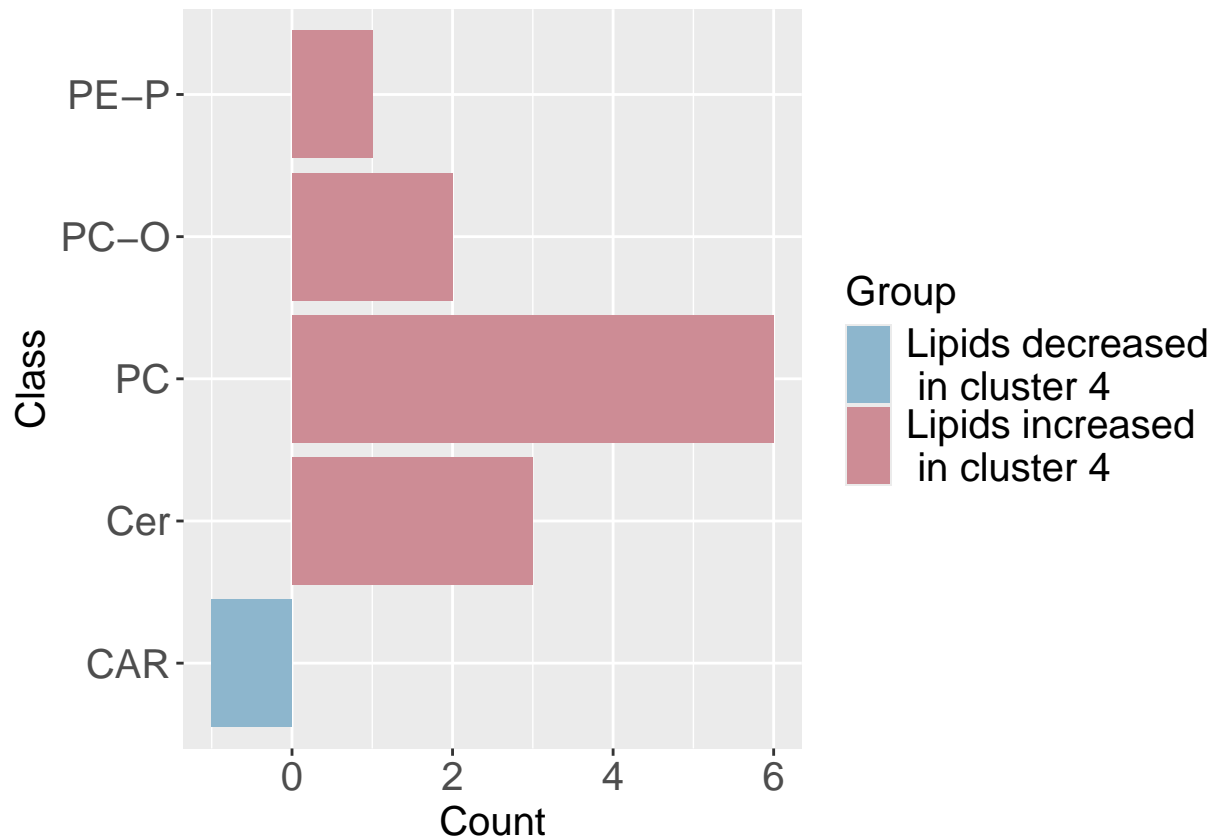

It can be seen that most of the significant lipids are increased in patients from the less successfully ageing cluster 4.