

# Analysing the contribution of mmse and sppb scales to changes in lipid profile

Upload the libraries:

```
library(psych)
library(ggplot2)
library(tidyr)
library(patchwork)
library(rstatix)
```

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 evaluate the correlation between scales and demographic factors:

Let's convert the data to the required type:

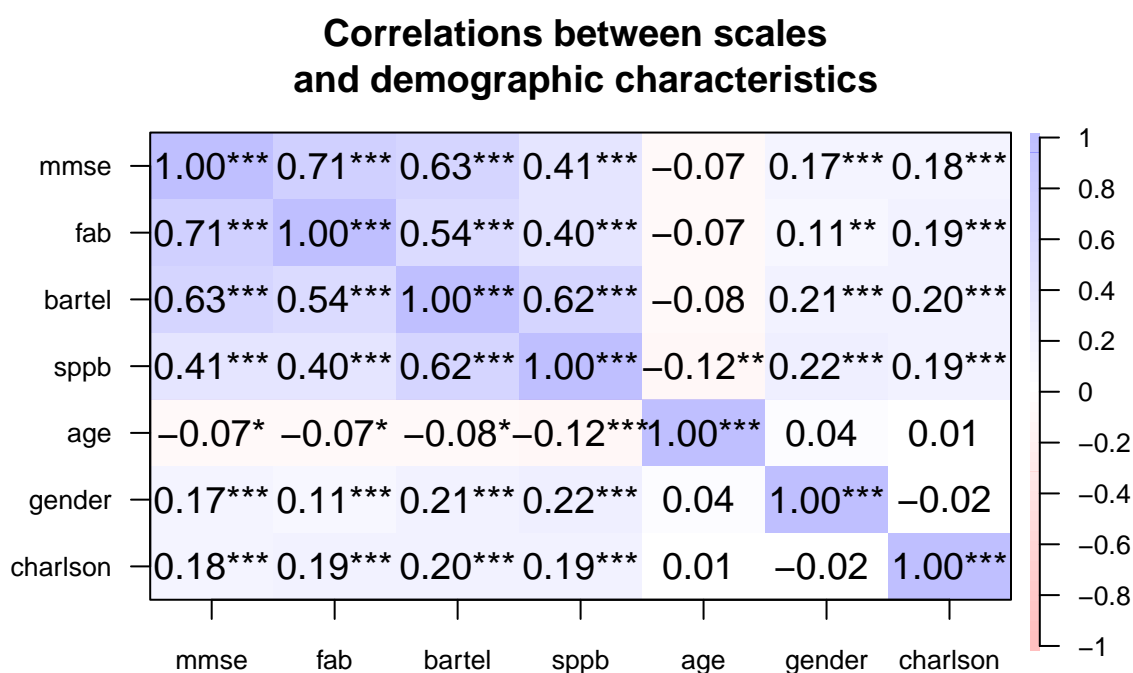
```
columns_to_convert <- which(names(scales_lipids) != "sex" & names(scales_lipids) !=
  "MS ID")
scales_lipids[, columns_to_convert] <- lapply(scales_lipids[,
  columns_to_convert], as.numeric)
```

Let's convert gender data to a numeric type:

```
names(scales_lipids)[names(scales_lipids) == "sex"] <- "gender"
names(scales_lipids)[names(scales_lipids) == "(-1)Charlson"] <- "charlson"
scales_lipids$gender <- ifelse(scales_lipids$gender == "f", 0,
  1)
```

Let's draw a correlation plot:

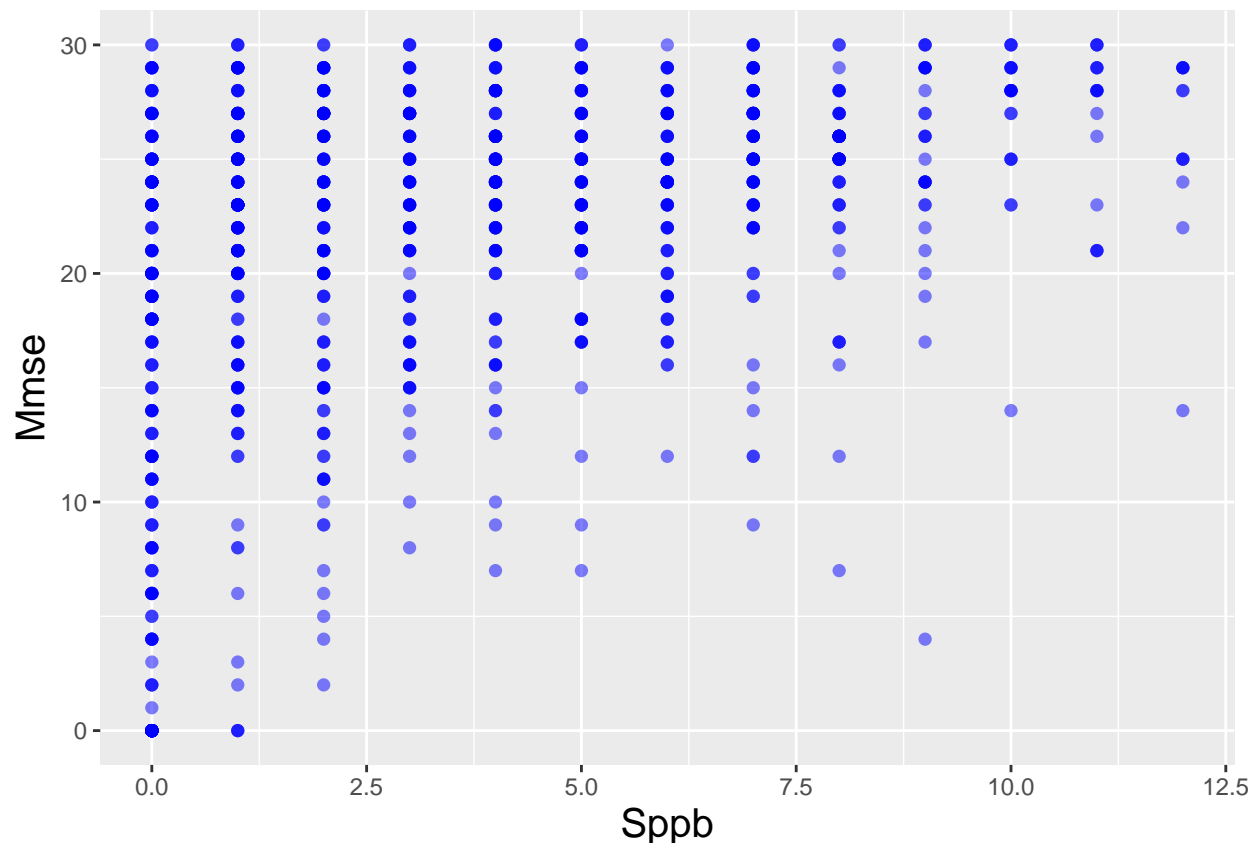
```
corPlot(scales_lipids[, c("mmse", "fab", "bartel", "sppb", "age",
  "gender", "charlson")], cex = 1.2, stars = TRUE, alpha = 0.25,
  cex.axis = 0.8, main = "Correlations between scales \n and demographic characteristics",
  cex.main = 1)
```



Scales are highly correlated with each other. In particular, the mmse scale is highly correlated with the sppb scale. Therefore, let us try to determine the contribution of the physical state of patients according to the sppb scale and the contribution of the cognitive state according to the mmse scale to the apparent changes in the lipid profile.

Let's look at the dependence of the mmse scale on the sppb scale:

```
ggplot(scales_lipids, aes(x = sppb)) + geom_point(aes(y = mmse),
  shape = 16, size = 2, color = adjustcolor("Blue"), alpha = 0.5) +
  labs(x = "Sppb", y = "Mmse") + 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))
```



The resulting graph shows that among patients without dementia (high values according to the mmse scale) we can distinguish two groups of patients with good and poor physical condition according to the sppb scale. Thus, we can remove the influence of mmse on the results and evaluate only the contribution of the sppb scale.

However, in patients among patients with dementia, we cannot distinguish two groups with poor and good physical condition because patients with dementia cannot pass the sppb scale test.

Therefore, in this case, to try to remove the influence of the sppb scale on the resulting differences in the lipid profile when dividing the groups according to the mmse scale we can try to divide the group of patients with poor physical condition into patients with and without dementia.

Let's remove the influence of the mmse scale and look at the differences in lipid profile when comparing the groups using the sppb scale

Let's remove samples with dementia on mmse:

```
mmse_morethen24 <- scales_lipids[scales_lipids$mmse >= 24, ]
```

Let's create a new dataset with the 100 worse values according to sppb:

```
n_rows_sppb <- 103
sorted_sppb <- mmse_morethen24[order(mmse_morethen24$sppb), ]
worse_sppb <- sorted_sppb[1:n_rows_sppb, ]
worse_sppb$sppb_asthenia <- rep("Asthenia", 103)
```

Let's create a new dataset with the 100 best values:

```
sorted_sppb_best <- mmse_morethen24[order(mmse_morethen24$sppb,
  decreasing = TRUE), ]
best_sppb <- sorted_sppb_best[1:n_rows_sppb, ]
best_sppb$sppb_asthenia <- rep("Healthy", 103)
```

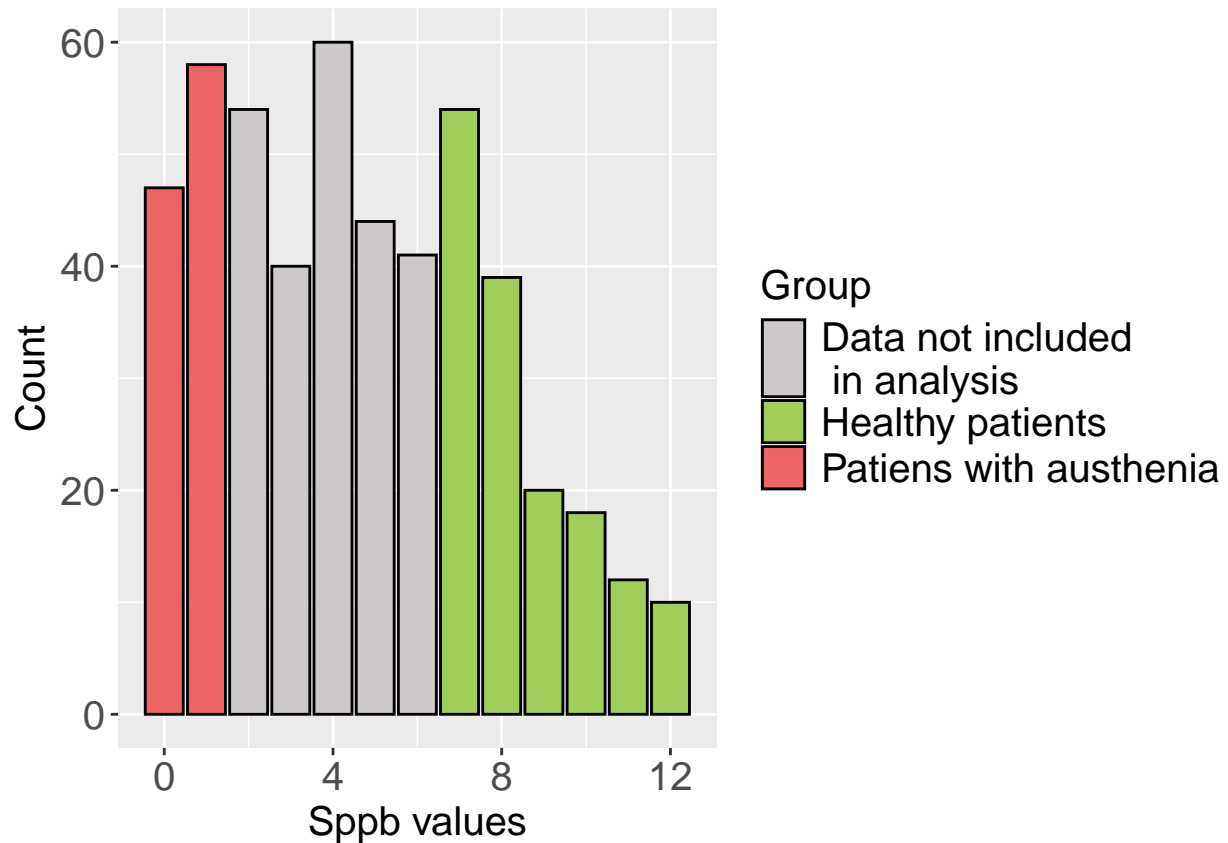
We combine both dataframes into one:

```
sppb_no_mmse <- rbind(worse_sppb, best_sppb)

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

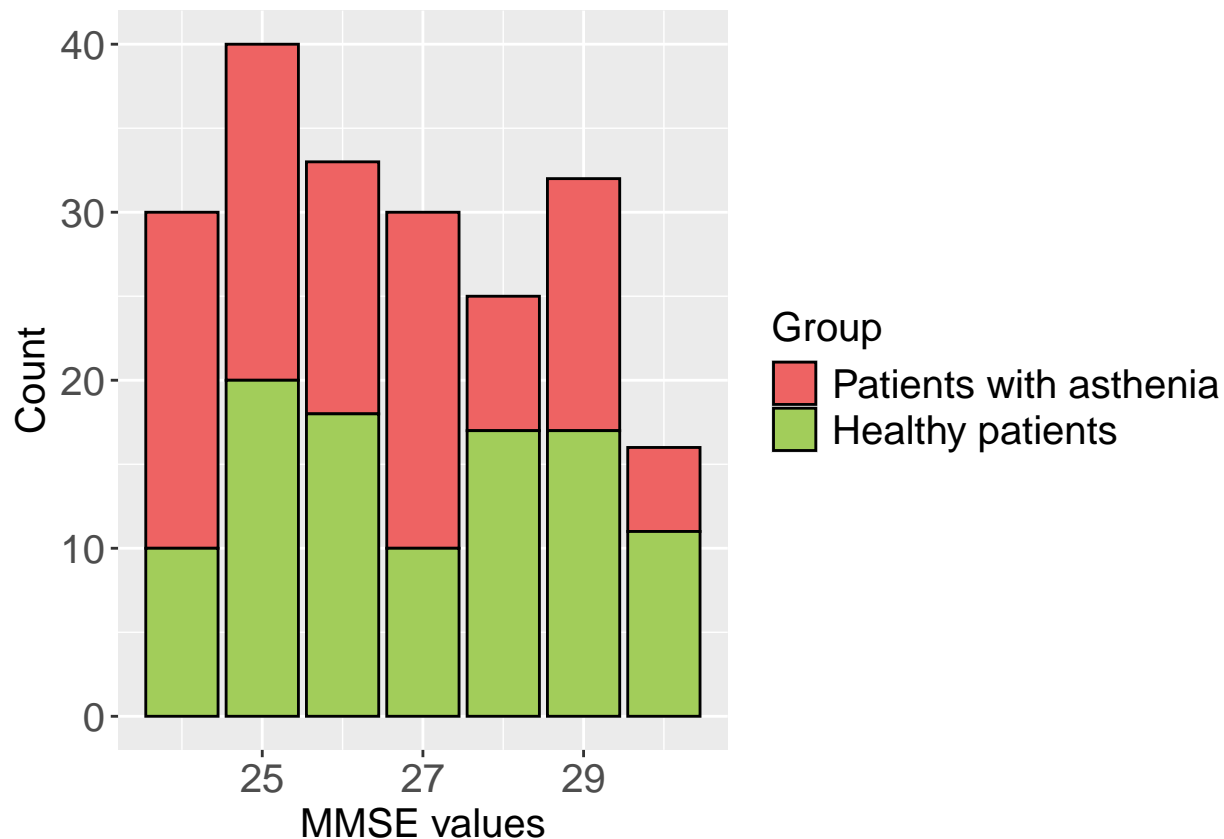
Distribution of sppb in data:

```
ggplot(mmse_morethen24, aes(x = sppb)) + geom_histogram(aes(fill = ifelse(sppb %in%
  best_sppb$sppb, "Healthy patients", ifelse(sppb %in% worse_sppb$sppb,
  "Patients with asthenia", "Data not included \n in analysis"))),
  color = "black", bins = 20, stat = "count") + scale_fill_manual(values = c("#CDC9C9",
  "darkolivegreen3", "#EE6363")) + labs(x = "Sppb 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))
```



Distribution of mmse in data:

```
ggplot(sppb_no_mmse, aes(x = mmse, fill = sppb_asthenia)) + geom_histogram(color = "black",
  bins = 20, stat = "count") + scale_fill_manual(values = c("#EE6363",
  "darkolivegreen3"), labels = c("Patients with asthenia",
  "Healthy patients")) + 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), axis.text.x = element_text(size = 15),
  axis.text.y = element_text(size = 15))
```



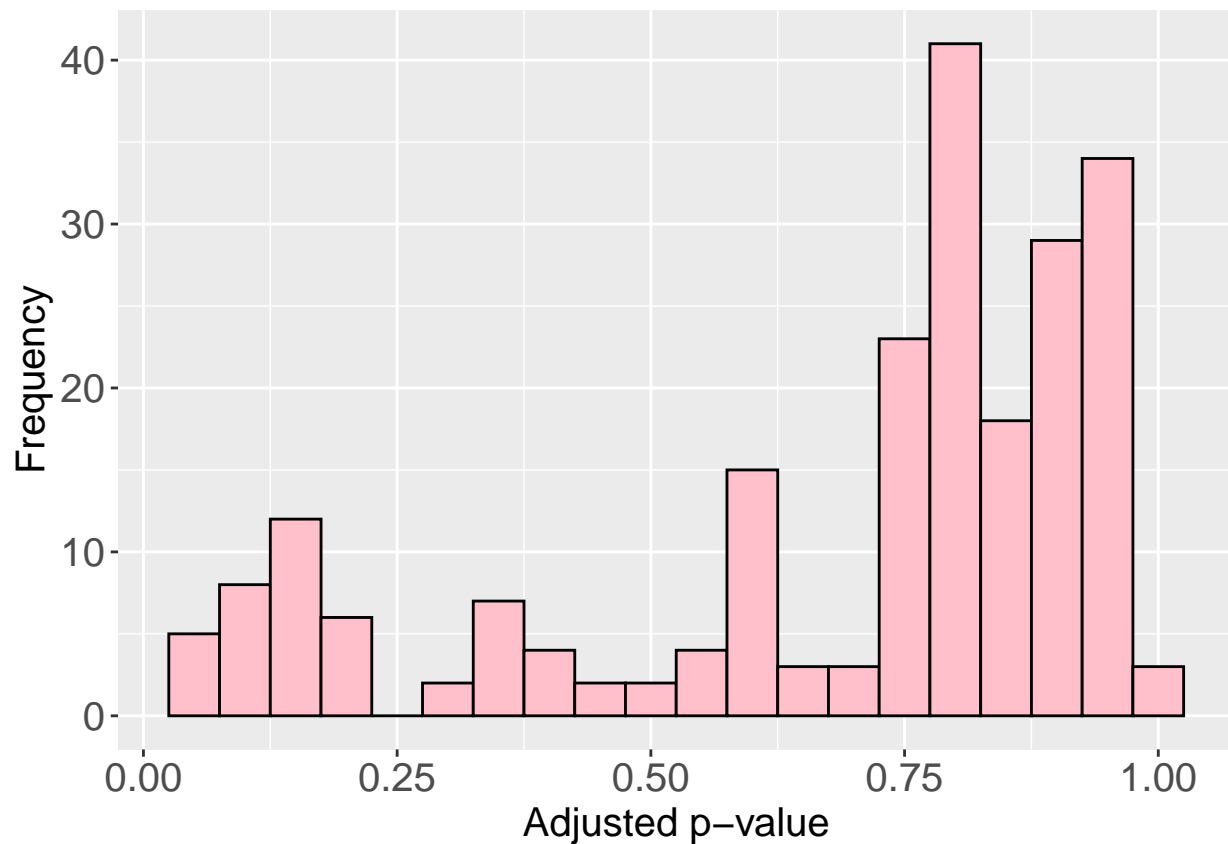
Let's conduct a t-test with the Benjamini-Hochberg correction

```
stat.test_sppb_no_mmse <- sppb_no_mmse_long %>%
  group_by(lipid_features) %>%
  t_test(values ~ sppb_asthenia) %>%
  adjust_pvalue(method = "BH") %>%
  add_significance()
print(stat.test_sppb_no_mmse)
```

```
## # A tibble: 221 x 11
##   lipid_features .y. group1 group2 n1 n2 statistic df p p.adj
##   <chr>          <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl> <dbl>
## 1 CAR 10:1      valu~ Asthe~ Healt~ 103 103 -1.11 201. 2.69e-1 0.728
## 2 CAR 18:1      valu~ Asthe~ Healt~ 103 103 0.797 194. 4.27e-1 0.802
## 3 CAR 18:2      valu~ Asthe~ Healt~ 103 103 -1.55 204. 1.23e-1 0.544
## 4 CE 14:0       valu~ Asthe~ Healt~ 103 103 -0.104 194. 9.17e-1 0.939
## 5 CE 15:0       valu~ Asthe~ Healt~ 103 103 -0.628 184. 5.31e-1 0.838
## 6 CE 16:1       valu~ Asthe~ Healt~ 103 103 3.46 197. 6.52e-4 0.0532
## 7 CE 17:1       valu~ Asthe~ Healt~ 103 103 0.750 199. 4.54e-1 0.822
## 8 CE 18:2       valu~ Asthe~ Healt~ 103 103 0.203 202. 8.39e-1 0.929
## 9 CE 18:3       valu~ Asthe~ Healt~ 103 103 0.660 204. 5.1 e-1 0.824
## 10 CE 19:1      valu~ Asthe~ Healt~ 103 103 0.00345 191. 9.97e-1 0.997
## # i 211 more rows
## # i 1 more variable: p.adj.signif <chr>
```

Let's build the p-value distribution:

```
ggplot(data = stat.test_sppb_no_mmse, aes(x = p.adj)) + geom_histogram(binwidth = 0.05,
  fill = "Pink", 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)) +
  theme(axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))
```



Let's look at the types of lipids for which significant differences emerged:

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

```
## # A tibble: 0 x 11
## # i 11 variables: lipid_features <chr>, .y. <chr>, group1 <chr>, group2 <chr>,
## #   n1 <int>, n2 <int>, statistic <dbl>, df <dbl>, p <dbl>, p.adj <dbl>,
## #   p.adj.signif <chr>
```

We see that without the contribution of mmse the groups of patients according to the sppb scale do not differ from each other in lipid profile.

Let's remove the influence of the sppb scale and look at the differences in lipid profile when comparing groups using the mmse scale

Let's remove healthy samples according to sppb scale:

```
sppb_lessthen7 <- scales_lipids[scales_lipids$sppb <= 7, ]
sppb_lessthen7 <- sppb_lessthen7[sppb_lessthen7$sppb > 0, ]
```

Let's create a new dataset with the 103 worse values according to mmse:

```
n_rows <- 103
sorted_mmse <- sppb_lessthen7[order(sppb_lessthen7$mmse), ]
worse_mmse <- sorted_mmse[1:n_rows, ]
worse_mmse$mmse_dementia <- rep("Dementia", 103)
```

Let's create a new dataset with the 103 best values according to mmse:

```
sorted_mmse_best <- sppb_lessthen7[order(sppb_lessthen7$mmse,
  decreasing = TRUE), ]
best_mmse <- sorted_mmse_best[1:n_rows, ]
best_mmse$mmse_dementia <- rep("Healthy", 103)
```

We combine both dataframes into one:

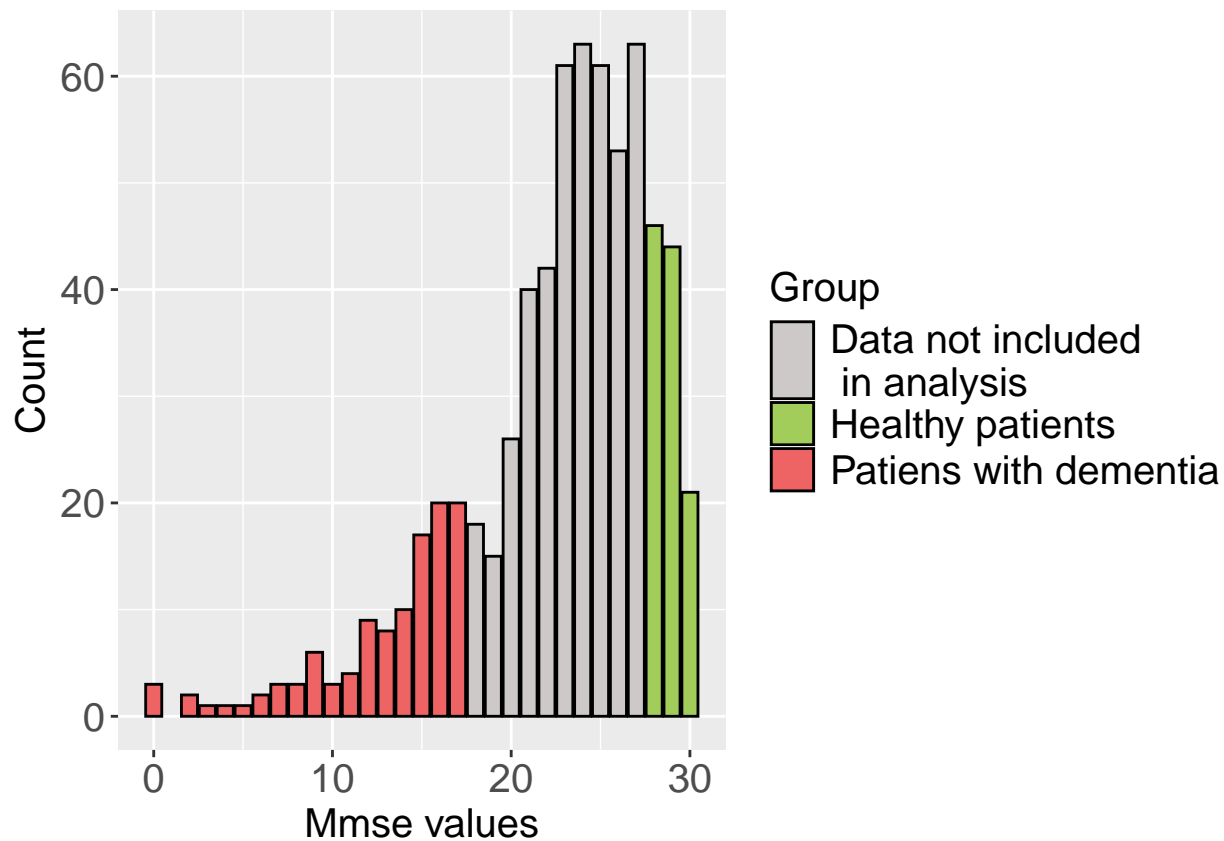
```
mmse_no_sppb <- rbind(worse_mmse, best_mmse)

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

Distribution of mmse in data:

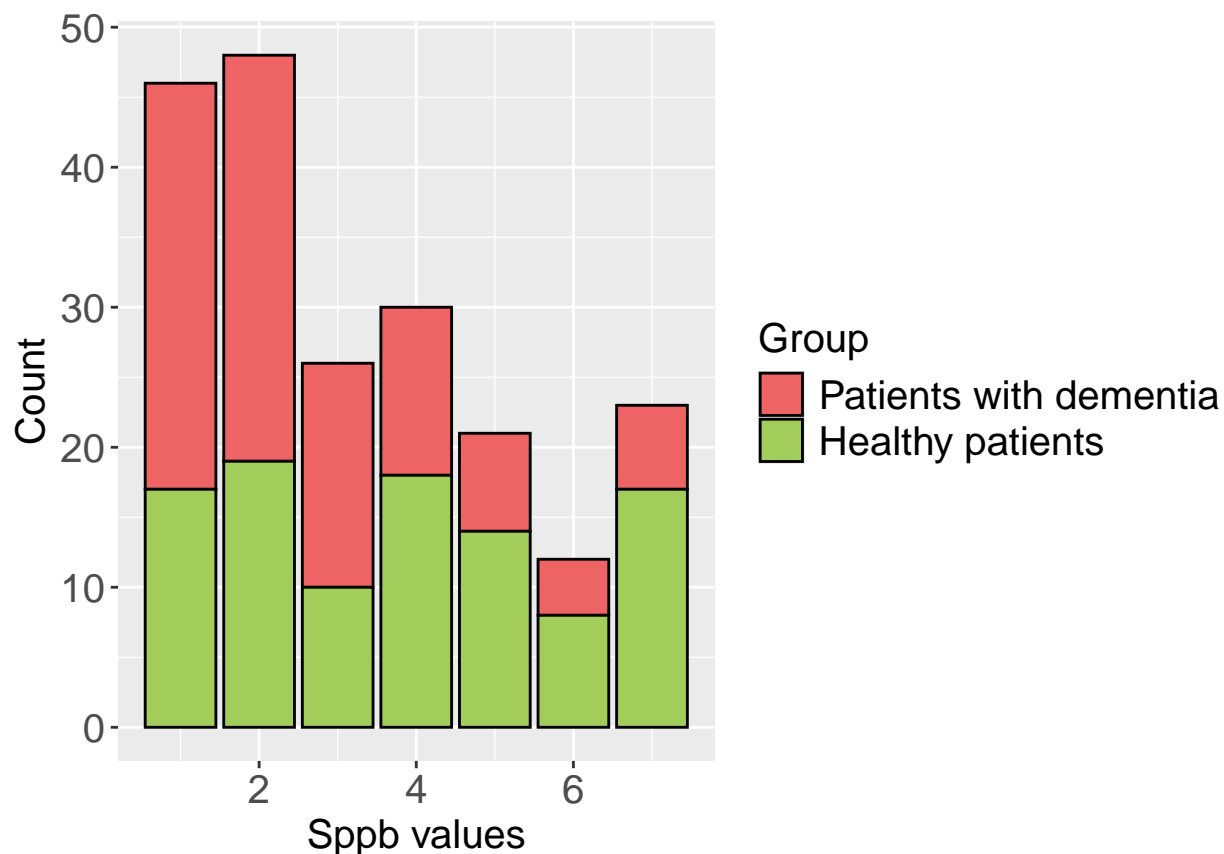
```
ggplot(sppb_lessthen7, aes(x = mmse)) + geom_histogram(aes(fill = ifelse(mmse %in%
  best_mmse$mmse, "Healthy patients", ifelse(mmse %in% worse_mmse$mmse,
  "Patients with dementia", "Data not included \n in analysis"))),
  color = "black", bins = 20, stat = "count") + scale_fill_manual(values = c("#CDC9C9",
  "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))
```





Distribution of sppb in data:

```
ggplot(mmse_no_sppb, aes(x = sppb, fill = mmse_dementia)) + geom_histogram(color = "black",
  bins = 20, stat = "count") + scale_fill_manual(values = c("#EE6363",
  "darkolivegreen3"), labels = c("Patients with dementia",
  "Healthy patients")) + labs(x = "Sppb 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), axis.text.x = element_text(size = 15),
  axis.text.y = element_text(size = 15))
```



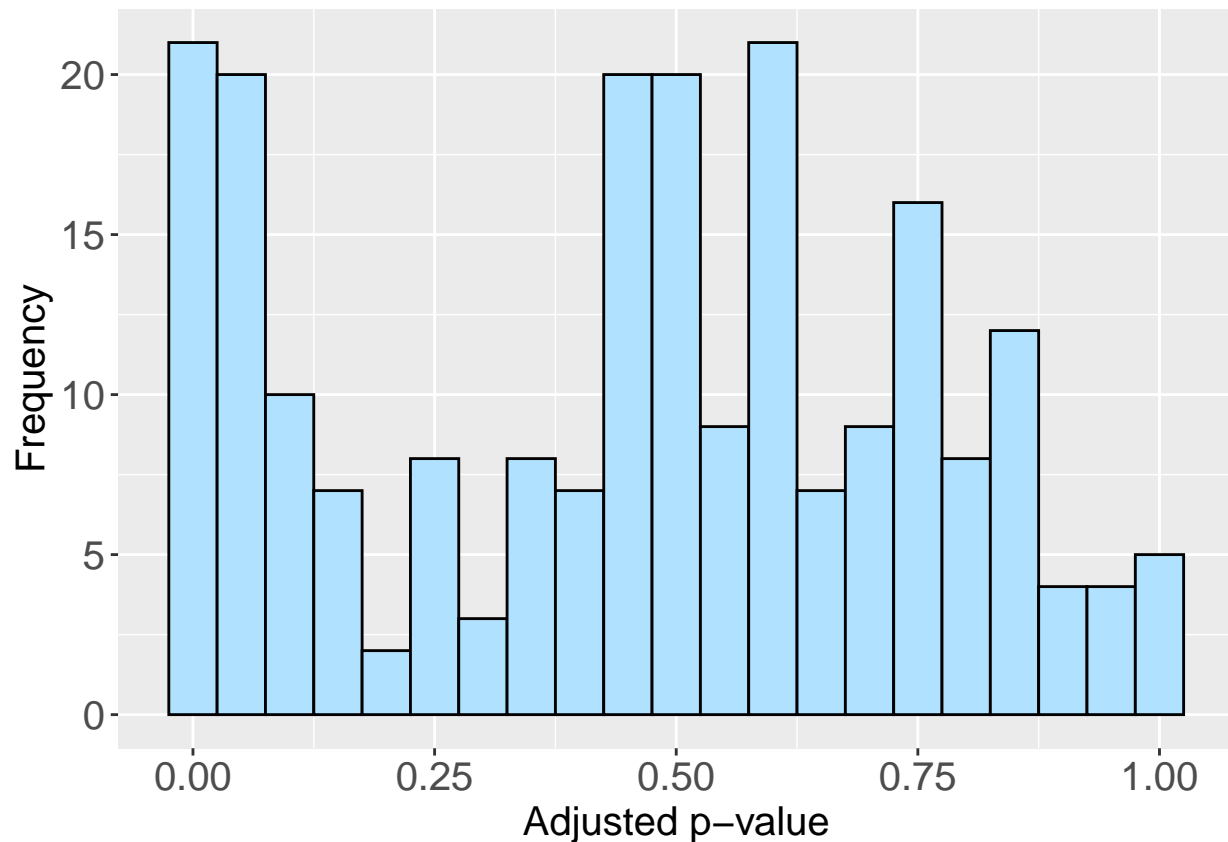
Let's conduct a t-test with the Benjamini-Hochberg correction:

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

```
## # 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  103  103   -2.24  204.  0.0263
## 2 CAR 18:1      values Dementia Healthy  103  103   -2.02  200.  0.0447
## 3 CAR 18:2      values Dementia Healthy  103  103   -3.23  201.  0.00146
## 4 CE 14:0       values Dementia Healthy  103  103   -0.150 200.  0.881
## 5 CE 15:0       values Dementia Healthy  103  103   -0.534 203.  0.594
## 6 CE 16:1       values Dementia Healthy  103  103    3.77  195.  0.000219
## 7 CE 17:1       values Dementia Healthy  103  103    2.61  199.  0.00965
## 8 CE 18:2       values Dementia Healthy  103  103   -0.637 200.  0.525
## 9 CE 18:3       values Dementia Healthy  103  103    1.29  204.  0.2
## 10 CE 19:1      values Dementia Healthy  103  103    0.469 154.  0.64
## # i 211 more rows
## # i 2 more variables: p.adj <dbl>, p.adj.signif <chr>
```

Let's build the p-value distribution:

```
ggplot(data = stat.test_mmse_no_sppb, 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)) +
  theme(axis.text.x = element_text(size = 15), axis.text.y = element_text(size = 15))
```



Let's look at the types of lipids for which significant differences emerged:

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

```
## # A tibble: 37 x 11
##   lipid_features .y. group1 group2 n1 n2 statistic df p
##   <chr>          <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>
## 1 CAR 18:2      values Dementia Healthy 103 103 -3.23 201. 0.00146
## 2 CE 16:1       values Dementia Healthy 103 103 3.77 195. 0.000219
## 3 CE 20:5       values Dementia Healthy 103 103 -4.35 196. 0.0000217
## 4 Cer 34:1      values Dementia Healthy 103 103 2.83 202. 0.00511
## 5 PC 32:1       values Dementia Healthy 103 103 3.42 196. 0.000762
```

```
## 6 PC 33:0      values Dementia Healthy 103 103 -3.46 127. 0.000747
## 7 PC 33:2      values Dementia Healthy 103 103 -2.92 204. 0.00386
## 8 PC 35:2      values Dementia Healthy 103 103 -3.13 204. 0.002
## 9 PC 35:5      values Dementia Healthy 103 103 -3.40 204. 0.000821
## 10 PC 36:5     values Dementia Healthy 103 103 -2.94 183. 0.00372
## # i 27 more rows
## # i 2 more variables: p.adj <dbl>, p.adj.signif <chr>
```

Thus, changes in the lipid profile when comparing patient groups according to mmse are not due to the influence of the sppb scale.