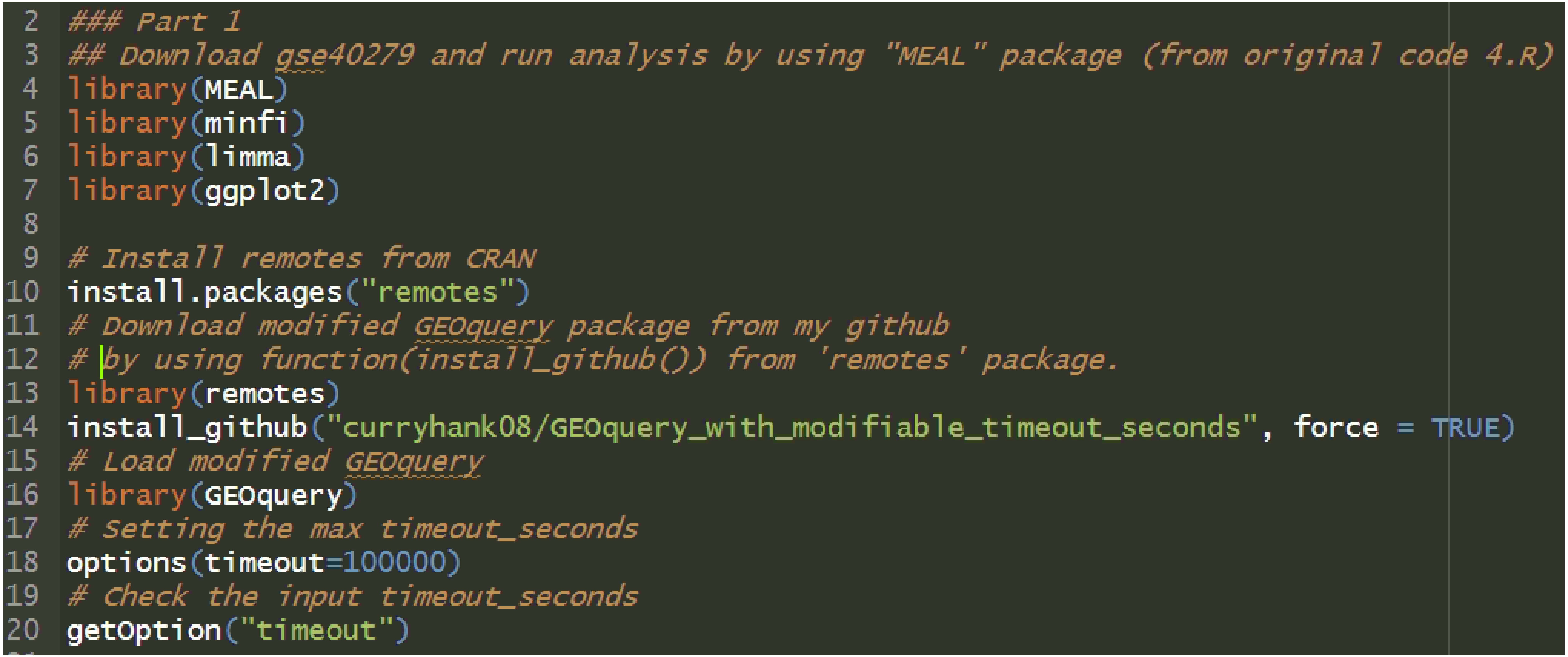
Analysis of gse40279 by MEAL and limma

姚博瀚　楊豐宇

這次做了什麼：

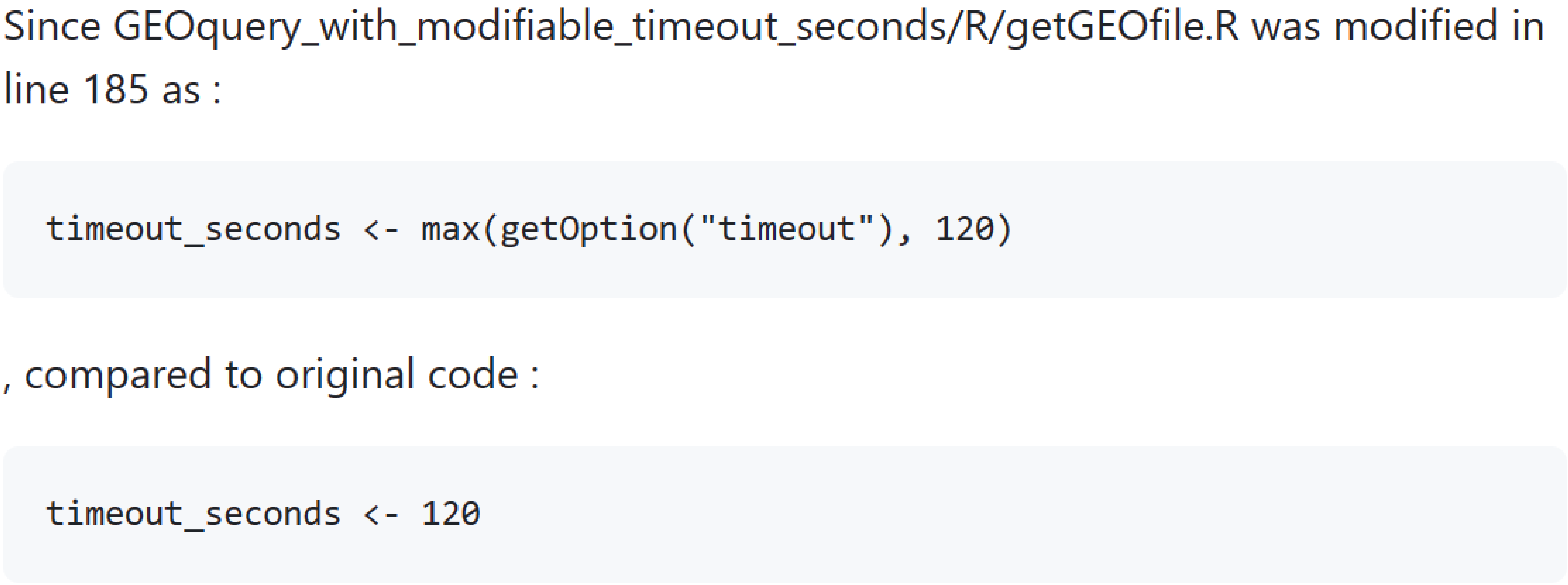
- 修改GEOquery套件
- 年齡分層
- QQplot
- beta-value distribution of a specific probe
- number of significant genes
- 修正limma design部分

修改GEOquery套件



```r
### Part 1
## Download gse40279 and run analysis by using "MEAL" package (from original code 4.R)
library(MEAL)
library(minfi)
library(limma)
library(ggplot2)

# Install remotes from CRAN
install.packages("remotes")
# Download modified GEOquery package from my github
# by using function(install_github()) from 'remotes' package.
library(remotes)
install_github("curryhank08/GEOquery_with_modifiable_timeout_seconds", force = TRUE)
# Load modified GEOquery
library(GEOquery)
# Setting the max timeout_seconds
options(timeout=100000)
# Check the input timeout_seconds
getOption("timeout")
```

修改GEOquery套件



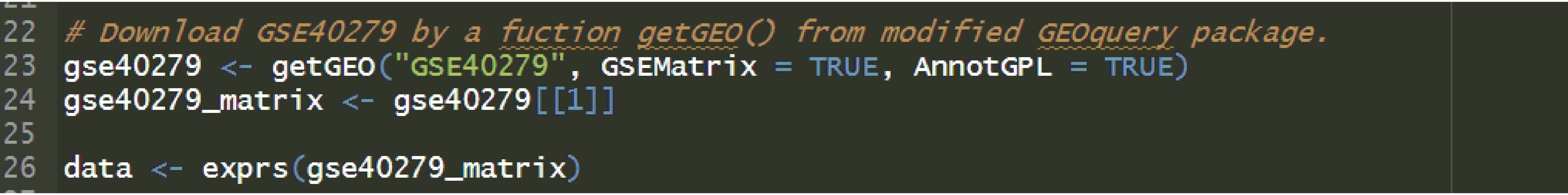
Since GEOquery_with_modifiable_timeout_seconds/R/getGEOfile.R was modified in line 185 as :

```
timeout_seconds <- max(getOption("timeout"), 120)
```

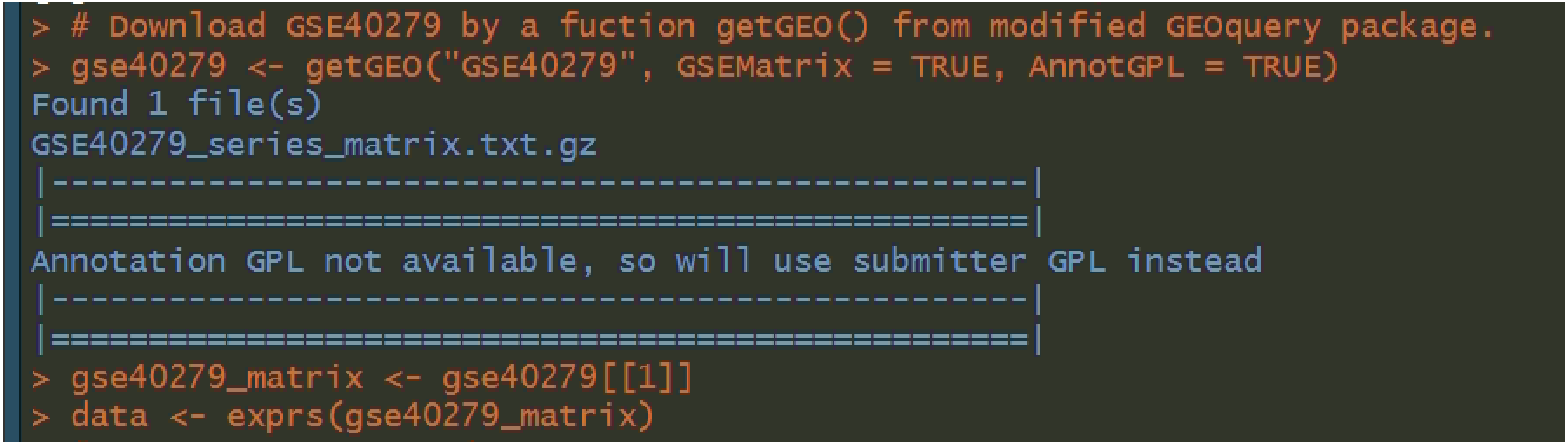
, compared to original code :

```
timeout_seconds <- 120
```

修改GEOquery套件

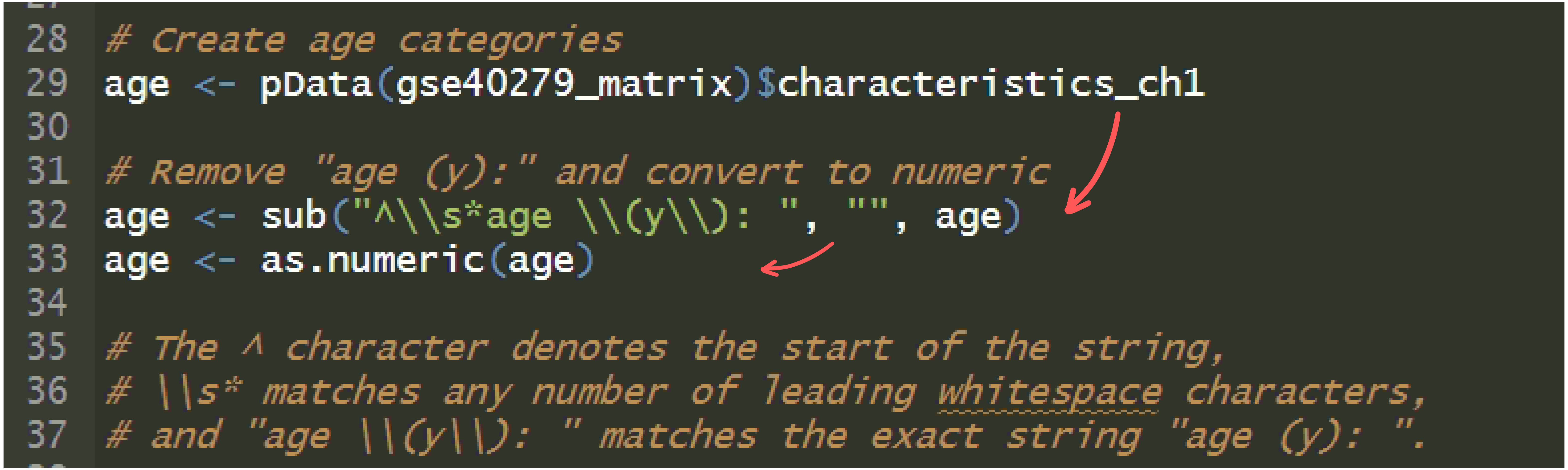


```r
# Download GSE40279 by a fuction getGEO() from modified GEOquery package.
gse40279 <- getGEO("GSE40279", GSEMatrix = TRUE, AnnotGPL = TRUE)
gse40279_matrix <- gse40279[[1]]

data <- exprs(gse40279_matrix)
```

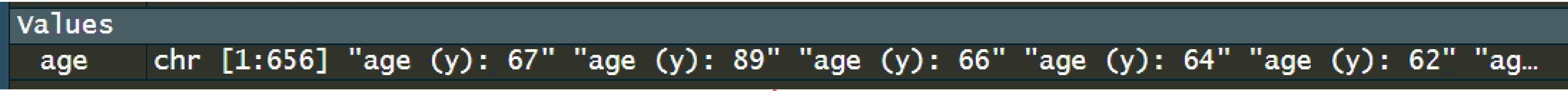


```
> # Download GSE40279 by a fuction getGEO() from modified GEOquery package.
> gse40279 <- getGEO("GSE40279", GSEMatrix = TRUE, AnnotGPL = TRUE)
Found 1 file(s)
GSE40279_series_matrix.txt.gz
|--------------------------------------------------|
|==================================================|
Annotation GPL not available, so will use submitter GPL instead
|--------------------------------------------------|
|==================================================|
> gse40279_matrix <- gse40279[[1]]
> data <- exprs(gse40279_matrix)
```

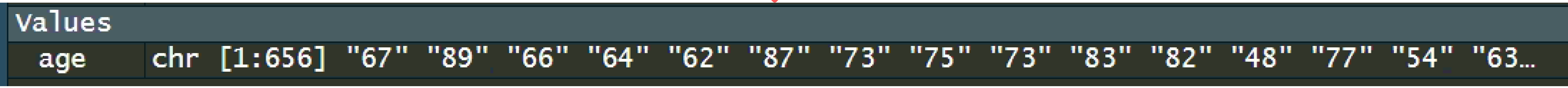
年齡分層



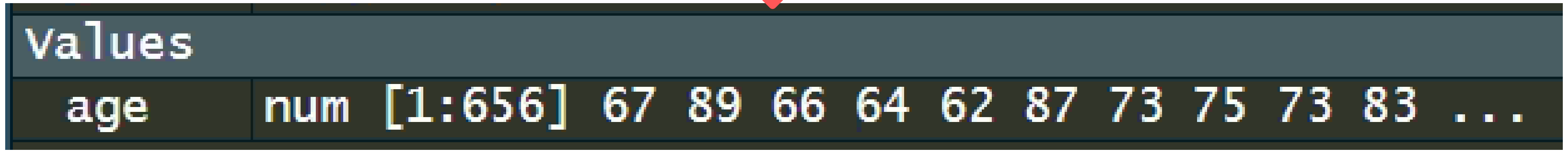
```r
# Create age categories
age <- pData(gse40279_matrix)$characteristics_ch1

# Remove "age (y):" and convert to numeric
age <- sub("^\\s*age \\(y\\): ", "", age)
age <- as.numeric(age)

# The ^ character denotes the start of the string,
# \\s* matches any number of leading whitespace characters,
# and "age \\(y\\): " matches the exact string "age (y): ".
```



```
Values
  age    chr [1:656] "age (y): 67" "age (y): 89" "age (y): 66" "age (y): 64" "age (y): 62" "ag…
```

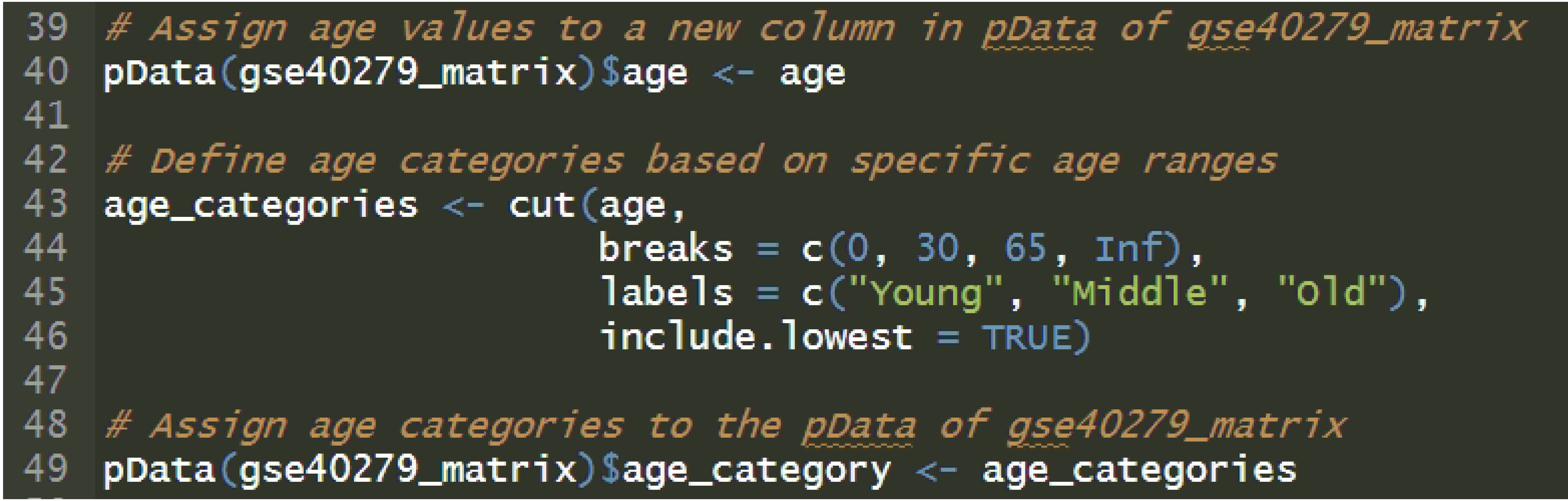


```
Values
  age    chr [1:656] "67" "89" "66" "64" "62" "87" "73" "75" "73" "83" "82" "48" "77" "54" "63…
```

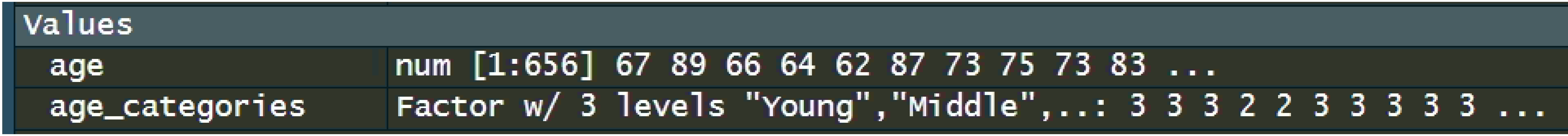


```
Values
  age    num [1:656] 67 89 66 64 62 87 73 75 73 83 ...
```

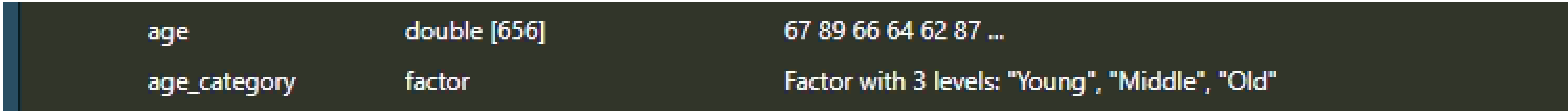
年齡分層



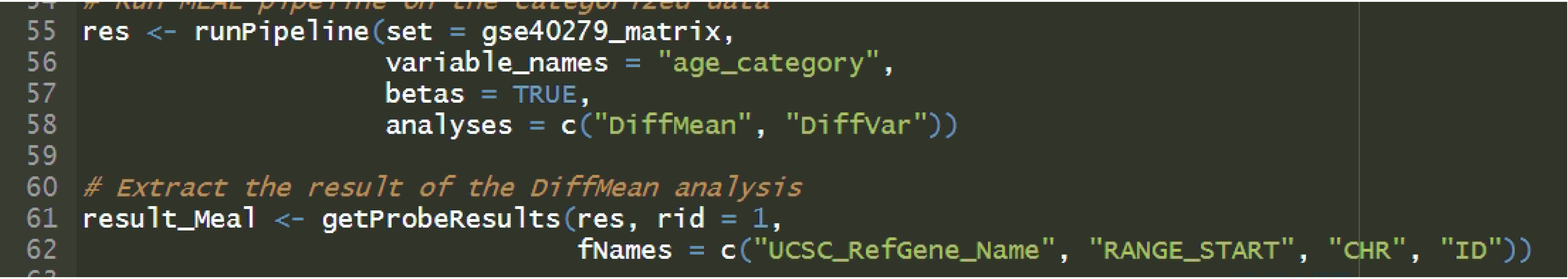
```r
# Assign age values to a new column in pData of gse40279_matrix
pData(gse40279_matrix)$age <- age

# Define age categories based on specific age ranges
age_categories <- cut(age,
                      breaks = c(0, 30, 65, Inf),
                      labels = c("Young", "Middle", "Old"),
                      include.lowest = TRUE)

# Assign age categories to the pData of gse40279_matrix
pData(gse40279_matrix)$age_category <- age_categories
```



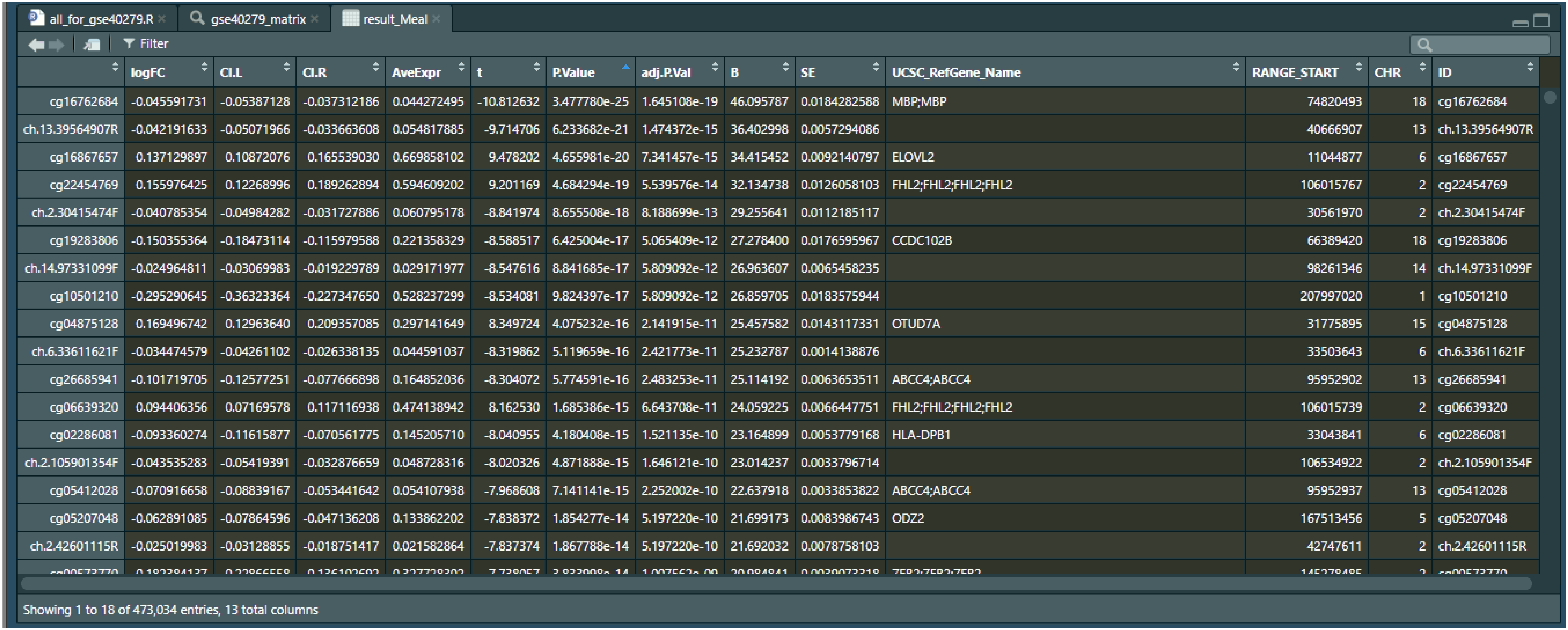
```
Values
  age               num [1:656] 67 89 66 64 62 87 73 75 73 83 ...
  age_categories    Factor w/ 3 levels "Young","Middle",..: 3 3 3 2 2 3 3 3 3 3 ...
```



```
age             double [656]    67 89 66 64 62 87 ...
age_category    factor          Factor with 3 levels: "Young", "Middle", "Old"
```

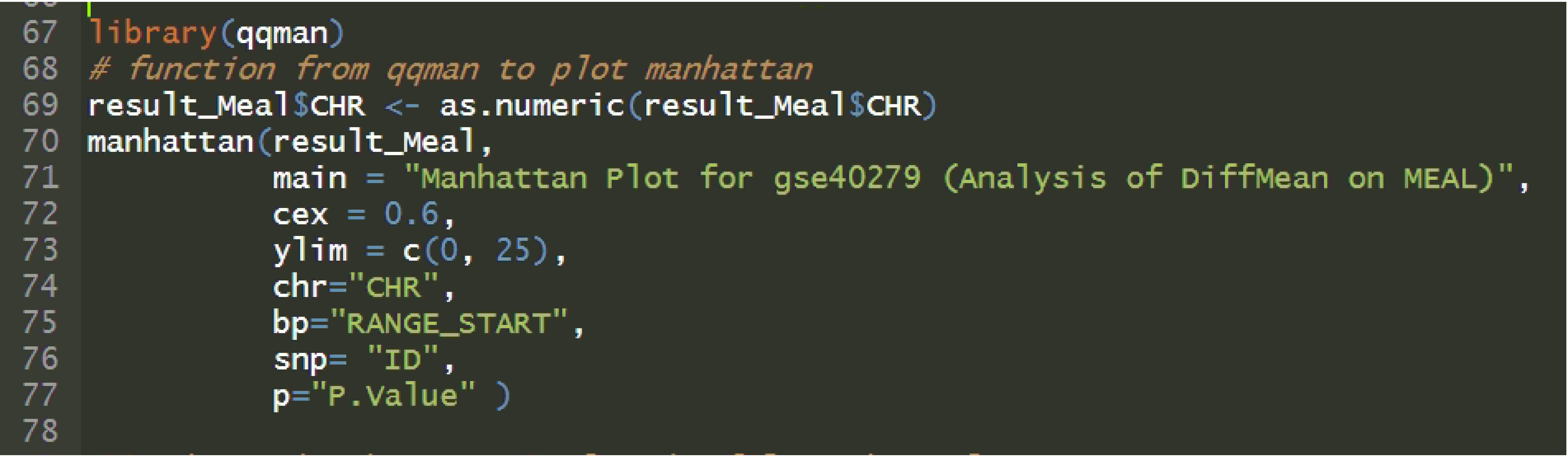


```r
# Run MEAL pipeline on the categorized data
res <- runPipeline(set = gse40279_matrix,
                   variable_names = "age_category",
                   betas = TRUE,
                   analyses = c("DiffMean", "DiffVar"))

# Extract the result of the DiffMean analysis
result_Meal <- getProbeResults(res, rid = 1,
                               fNames = c("UCSC_RefGene_Name", "RANGE_START", "CHR", "ID"))
```



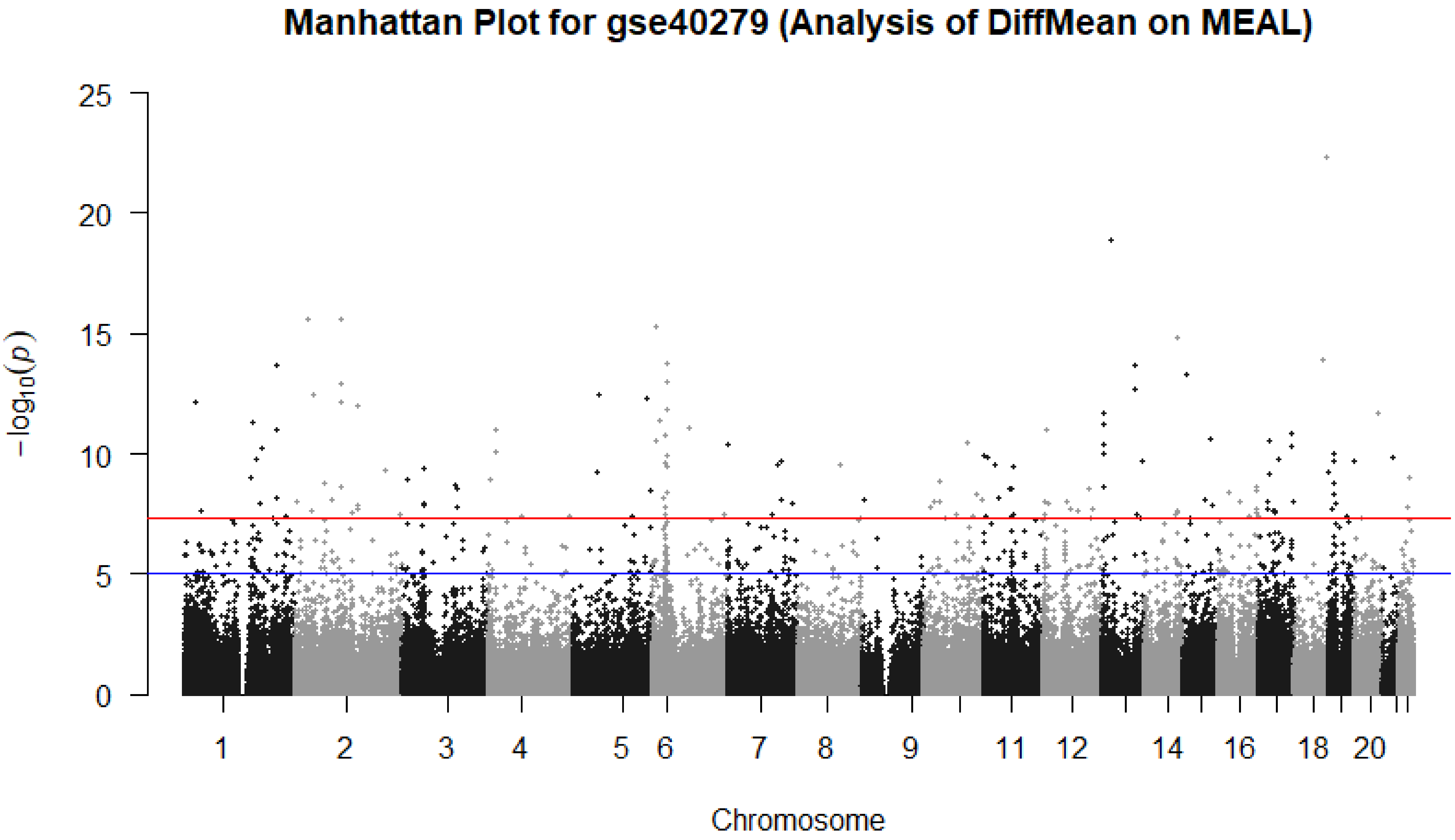
	logFC	CI.L	CI.R	AveExpr	t	P.Value	adj.P.Val	B	SE	UCSC_RefGene_Name	RANGE_START	CHR	ID
cg16762684	-0.045591731	-0.05387128	-0.037312186	0.044272495	-10.812632	3.477780e-25	1.645108e-19	46.095787	0.0184282588	MBP;MBP	74820493	18	cg16762684
ch.13.39564907R	-0.042191633	-0.05071966	-0.033663608	0.054817885	-9.714706	6.233682e-21	1.474372e-15	36.402998	0.0057294086		40666907	13	ch.13.39564907R
cg16867657	0.137129897	0.10872076	0.165539030	0.669858102	9.478202	4.655981e-20	7.341457e-15	34.415452	0.0092140797	ELOVL2	11044877	6	cg16867657
cg22454769	0.155976425	0.12268996	0.189262894	0.594609202	9.201169	4.684294e-19	5.539576e-14	32.134738	0.0126058103	FHL2;FHL2;FHL2;FHL2	106015767	2	cg22454769
ch.2.30415474F	-0.040785354	-0.04984282	-0.031727886	0.060795178	-8.841974	8.655508e-18	8.188699e-13	29.255641	0.0112185117		30561970	2	ch.2.30415474F
cg19283806	-0.150355364	-0.18473114	-0.115979588	0.221358329	-8.588517	6.425004e-17	5.065409e-12	27.278400	0.0176595967	CCDC102B	66389420	18	cg19283806
ch.14.97331099F	-0.024964811	-0.03069983	-0.019229789	0.029171977	-8.547616	8.841685e-17	5.809092e-12	26.963607	0.0065458235		98261346	14	ch.14.97331099F
cg10501210	-0.295290645	-0.36323364	-0.227347650	0.528237299	-8.534081	9.824397e-17	5.809092e-12	26.859705	0.0183575944		207997020	1	cg10501210
cg04875128	0.169496742	0.12963640	0.209357085	0.297141649	8.349724	4.075232e-16	2.141915e-11	25.457582	0.0143117331	OTUD7A	31775895	15	cg04875128
ch.6.33611621F	-0.034474579	-0.04261102	-0.026338135	0.044591037	-8.319862	5.119659e-16	2.421773e-11	25.232787	0.0014138876		33503643	6	ch.6.33611621F
cg26685941	-0.101719705	-0.12577251	-0.077666898	0.164852036	-8.304072	5.774591e-16	2.483253e-11	25.114192	0.0063653511	ABCC4;ABCC4	95952902	13	cg26685941
cg06639320	0.094406356	0.07169578	0.117116938	0.474138942	8.162530	1.685386e-15	6.643708e-11	24.059225	0.0066447751	FHL2;FHL2;FHL2;FHL2	106015739	2	cg06639320
cg02286081	-0.093360274	-0.11615877	-0.070561775	0.145205710	-8.040955	4.180408e-15	1.521135e-10	23.164899	0.0053779168	HLA-DPB1	33043841	6	cg02286081
ch.2.105901354F	-0.043535283	-0.05419391	-0.032876659	0.048728316	-8.020326	4.871888e-15	1.646121e-10	23.014237	0.0033796714		106534922	2	ch.2.105901354F
cg05412028	-0.070916658	-0.08839167	-0.053441642	0.054107938	-7.968608	7.141141e-15	2.252002e-10	22.637918	0.0033853822	ABCC4;ABCC4	95952937	13	cg05412028
cg05207048	-0.062891085	-0.07864596	-0.047136208	0.133862202	-7.838372	1.854277e-14	5.197220e-10	21.699173	0.0083986743	ODZ2	167513456	5	cg05207048
ch.2.42601115R	-0.025019983	-0.03128855	-0.018751417	0.021582864	-7.837374	1.867788e-14	5.197220e-10	21.692032	0.0078758103		42747611	2	ch.2.42601115R

Showing 1 to 18 of 473,034 entries, 13 total columns

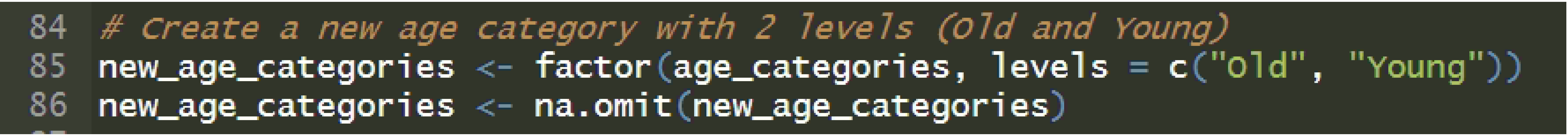


```r
library(qqman)
# function from qqman to plot manhattan
result_Meal$CHR <- as.numeric(result_Meal$CHR)
manhattan(result_Meal,
          main = "Manhattan Plot for gse40279 (Analysis of DiffMean on MEAL)",
          cex = 0.6,
          ylim = c(0, 25),
          chr="CHR",
          bp="RANGE_START",
          snp= "ID",
          p="P.Value" )

```

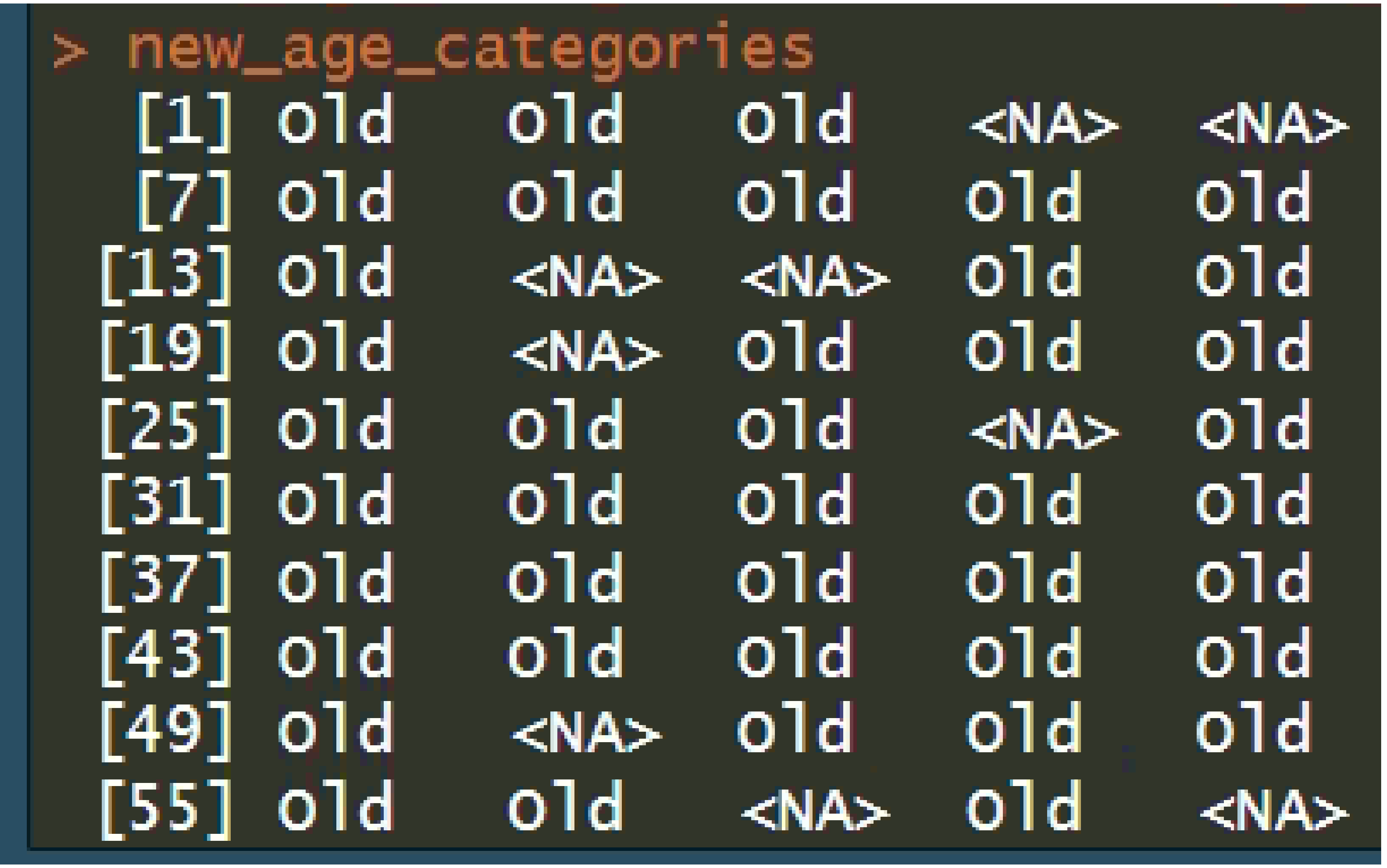
MEAL



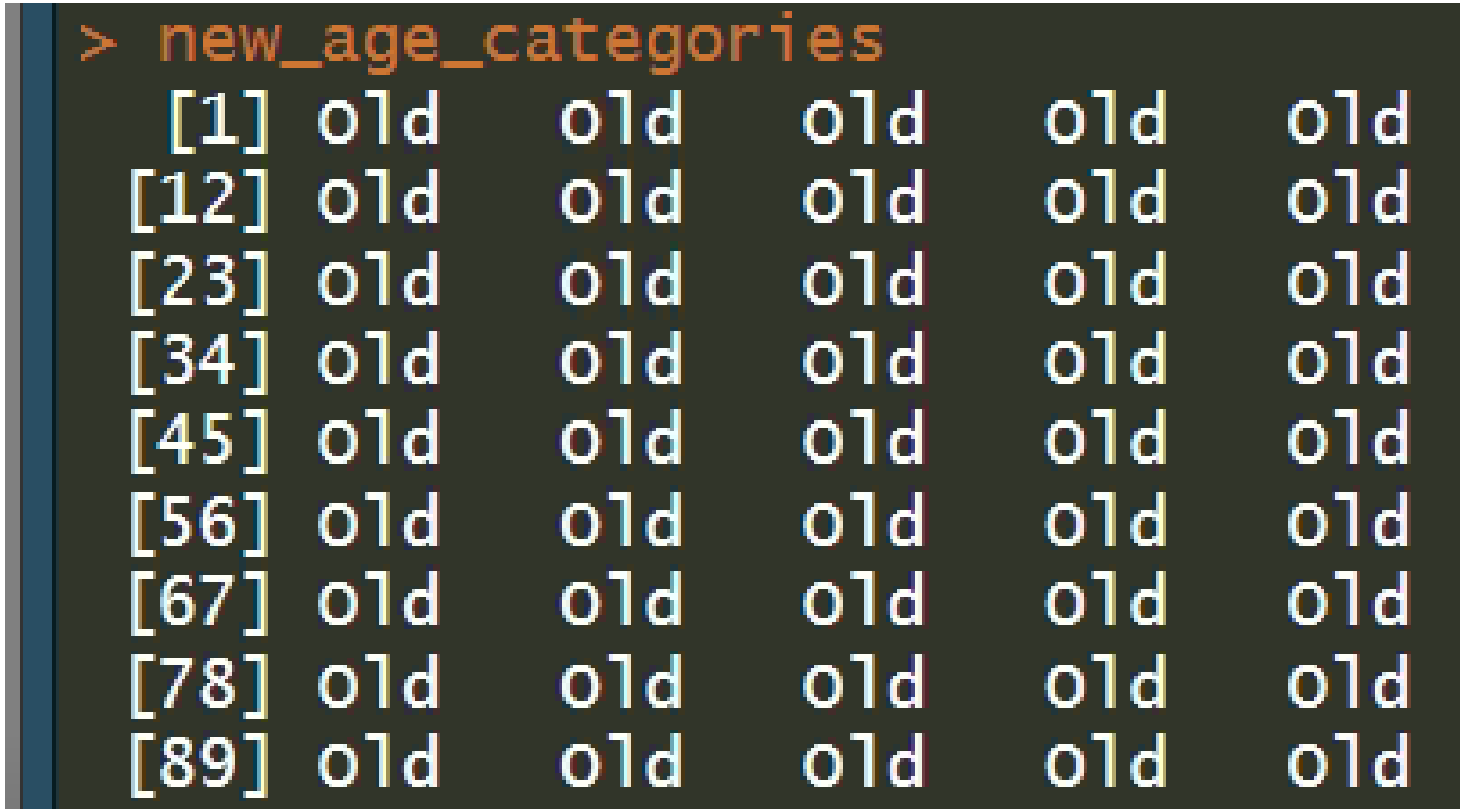
Manhattan Plot for gse40279 (Analysis of DiffMean on MEAL)
$-\log_{10}(p)$
25
20
15
10
5
0
Chromosome
1
2
3
4
5
6
7
8
9
11
12
14
16
18
20



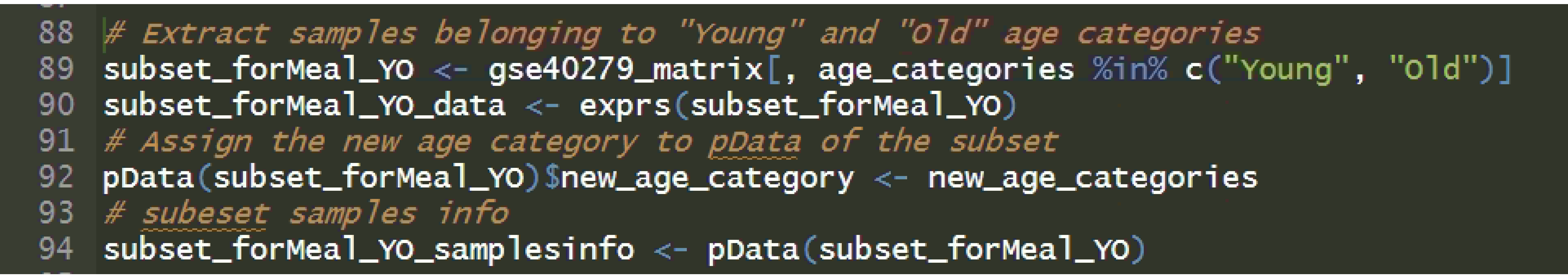
```r
# Create a new age category with 2 levels (Old and Young)
new_age_categories <- factor(age_categories, levels = c("Old", "Young"))
new_age_categories <- na.omit(new_age_categories)
```



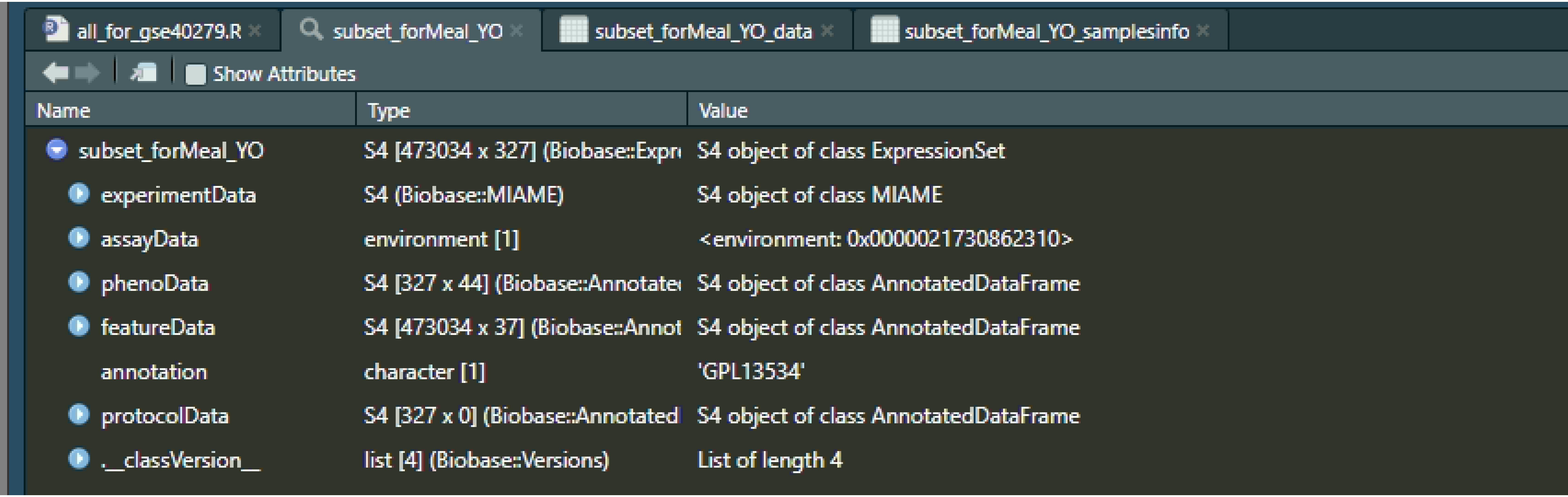
```
> new_age_categories
  [1] Old   Old   Old   <NA>  <NA>
  [7] Old   Old   Old   Old   Old
 [13] Old   <NA>  <NA>  Old   Old
 [19] Old   <NA>  Old   Old   Old
 [25] Old   Old   Old   <NA>  Old
 [31] Old   Old   Old   Old   Old
 [37] Old   Old   Old   Old   Old
 [43] Old   Old   Old   Old   Old
 [49] Old   <NA>  Old   Old   Old
 [55] Old   Old   <NA>  Old   <NA>
```



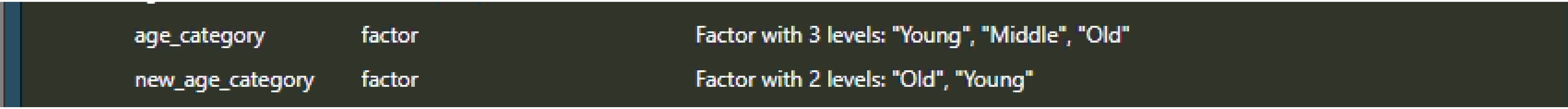
```
> new_age_categories
  [1] Old   Old   Old   Old   Old
 [12] Old   Old   Old   Old   Old
 [23] Old   Old   Old   Old   Old
 [34] Old   Old   Old   Old   Old
 [45] Old   Old   Old   Old   Old
 [56] Old   Old   Old   Old   Old
 [67] Old   Old   Old   Old   Old
 [78] Old   Old   Old   Old   Old
 [89] Old   Old   Old   Old   Old
```



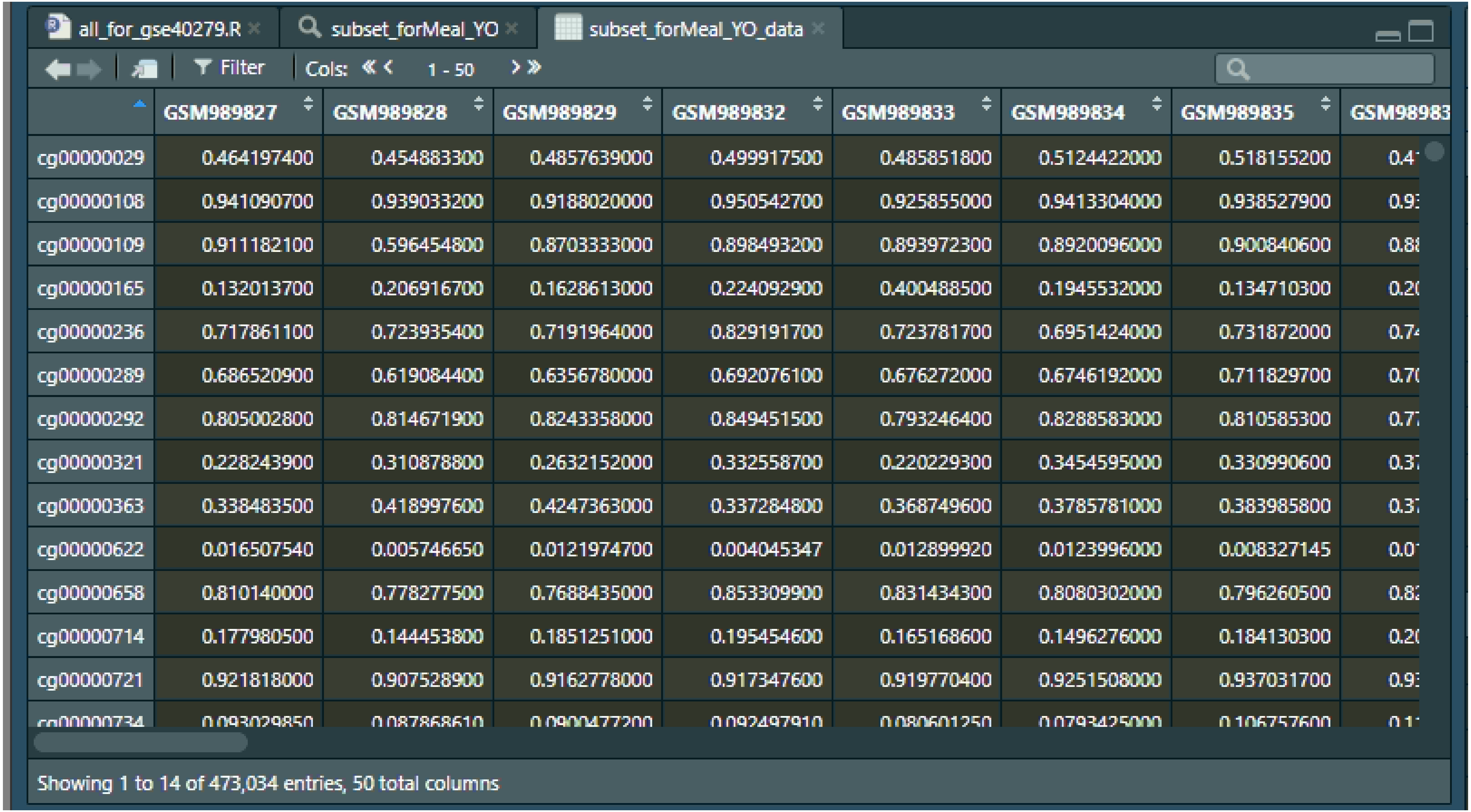
```r
# Extract samples belonging to "Young" and "Old" age categories
subset_forMeal_YO <- gse40279_matrix[, age_categories %in% c("Young", "Old")]
subset_forMeal_YO_data <- exprs(subset_forMeal_YO)
# Assign the new age category to pData of the subset
pData(subset_forMeal_YO)$new_age_category <- new_age_categories
# subeset samples info
subset_forMeal_YO_samplesinfo <- pData(subset_forMeal_YO)
```



Name	Type	Value
subset_forMeal_YO	S4 [473034 x 327] (Biobase::Expr	S4 object of class ExpressionSet
experimentData	S4 (Biobase::MIAME)	S4 object of class MIAME
assayData	environment [1]	<environment: 0x0000021730862310>
phenoData	S4 [327 x 44] (Biobase::Annotate	S4 object of class AnnotatedDataFrame
featureData	S4 [473034 x 37] (Biobase::Annot	S4 object of class AnnotatedDataFrame
annotation	character [1]	'GPL13534'
protocolData	S4 [327 x 0] (Biobase::Annotated	S4 object of class AnnotatedDataFrame
.__classVersion__	list [4] (Biobase::Versions)	List of length 4

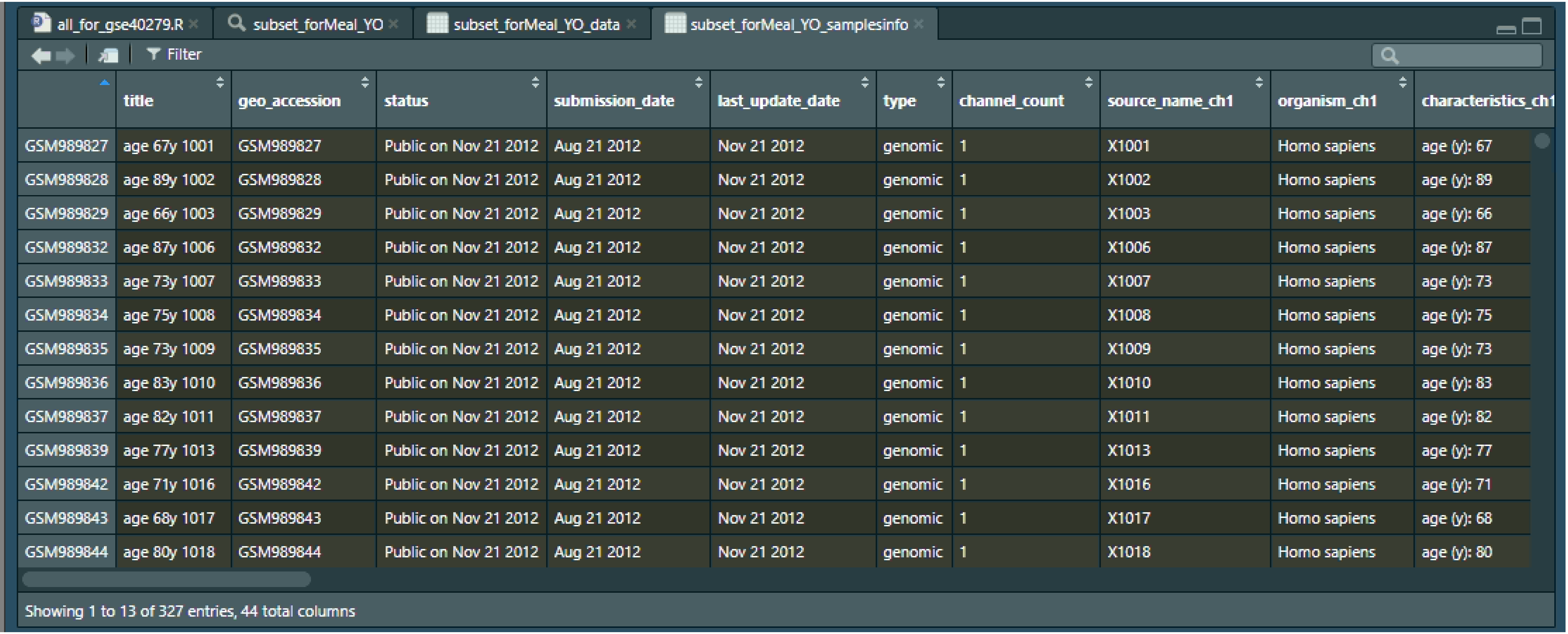


```
age_category        factor        Factor with 3 levels: "Young", "Middle", "Old"
new_age_category    factor        Factor with 2 levels: "Old", "Young"
```



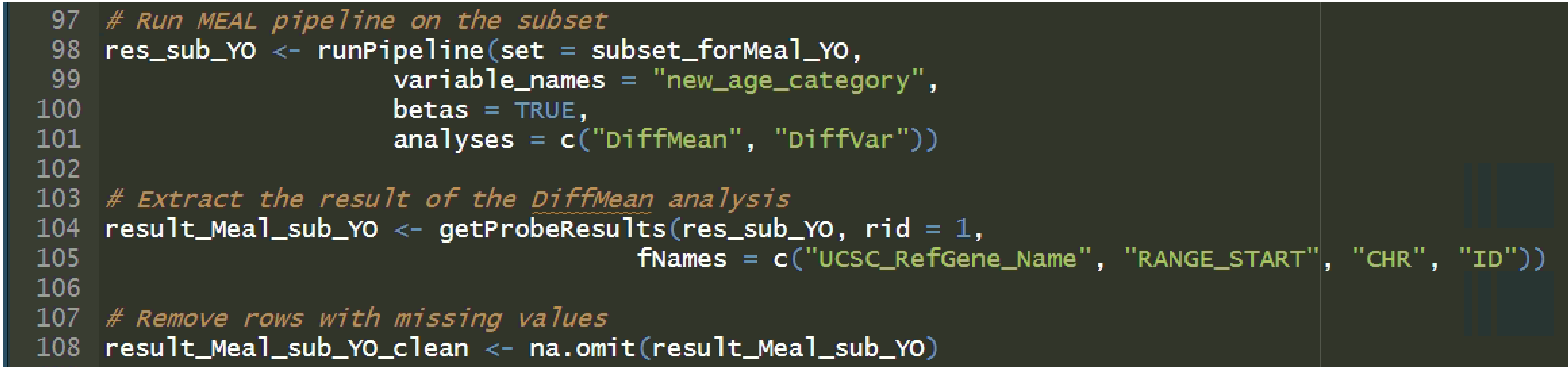
	GSM989827	GSM989828	GSM989829	GSM989832	GSM989833	GSM989834	GSM989835	GSM98983
cg00000029	0.464197400	0.454883300	0.4857639000	0.499917500	0.485851800	0.5124422000	0.518155200	0.4
cg00000108	0.941090700	0.939033200	0.9188020000	0.950542700	0.925855000	0.9413304000	0.938527900	0.93
cg00000109	0.911182100	0.596454800	0.8703333000	0.898493200	0.893972300	0.8920096000	0.900840600	0.88
cg00000165	0.132013700	0.206916700	0.1628613000	0.224092900	0.400488500	0.1945532000	0.134710300	0.20
cg00000236	0.717861100	0.723935400	0.7191964000	0.829191700	0.723781700	0.6951424000	0.731872000	0.74
cg00000289	0.686520900	0.619084400	0.6356780000	0.692076100	0.676272000	0.6746192000	0.711829700	0.70
cg00000292	0.805002800	0.814671900	0.8243358000	0.849451500	0.793246400	0.8288583000	0.810585300	0.77
cg00000321	0.228243900	0.310878800	0.2632152000	0.332558700	0.220229300	0.3454595000	0.330990600	0.37
cg00000363	0.338483500	0.418997600	0.4247363000	0.337284800	0.368749600	0.3785781000	0.383985800	0.37
cg00000622	0.016507540	0.005746650	0.0121974700	0.004045347	0.012899920	0.0123996000	0.008327145	0.01
cg00000658	0.810140000	0.778277500	0.7688435000	0.853309900	0.831434300	0.8080302000	0.796260500	0.82
cg00000714	0.177980500	0.144453800	0.1851251000	0.195454600	0.165168600	0.1496276000	0.184130300	0.20
cg00000721	0.921818000	0.907528900	0.9162778000	0.917347600	0.919770400	0.9251508000	0.937031700	0.93
cg00000734	0.093029850	0.087868610	0.0900477200	0.092497910	0.080601250	0.0793425000	0.106757600	0.11

Showing 1 to 14 of 473,034 entries, 50 total columns

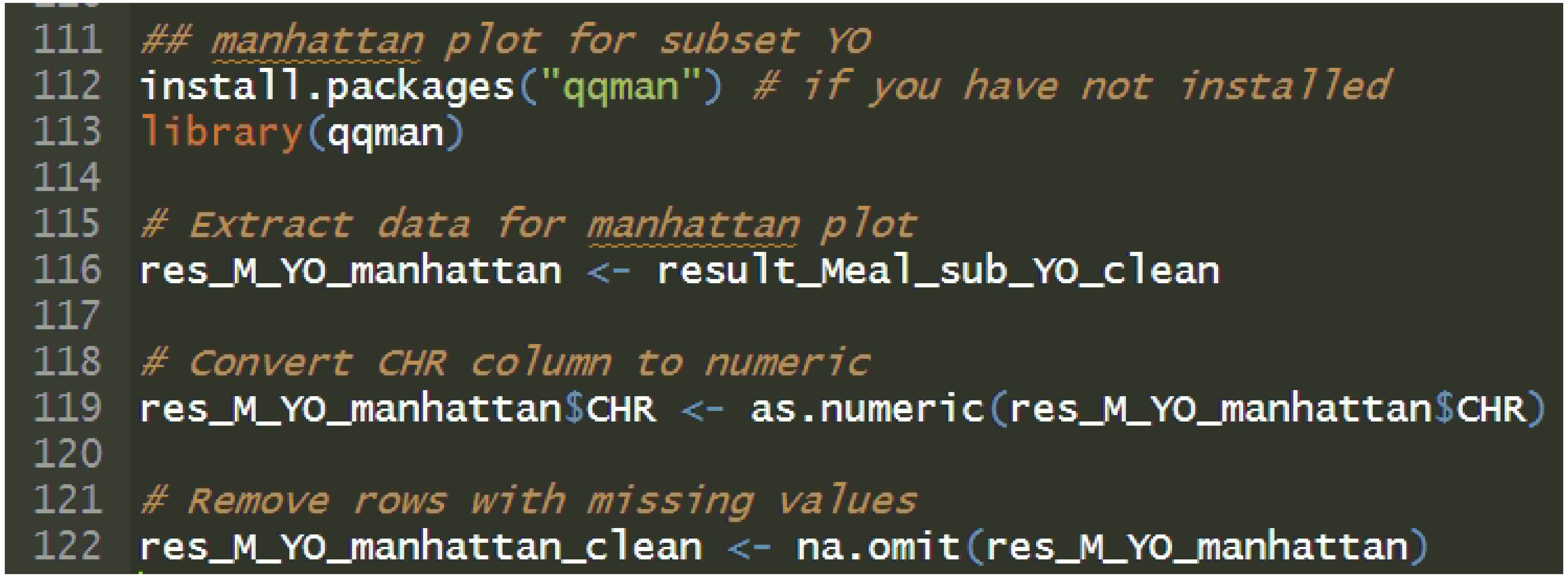


	title	geo_accession	status	submission_date	last_update_date	type	channel_count	source_name_ch1	organism_ch1	characteristics_ch1
GSM989827	age 67y 1001	GSM989827	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1001	Homo sapiens	age (y): 67
GSM989828	age 89y 1002	GSM989828	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1002	Homo sapiens	age (y): 89
GSM989829	age 66y 1003	GSM989829	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1003	Homo sapiens	age (y): 66
GSM989832	age 87y 1006	GSM989832	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1006	Homo sapiens	age (y): 87
GSM989833	age 73y 1007	GSM989833	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1007	Homo sapiens	age (y): 73
GSM989834	age 75y 1008	GSM989834	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1008	Homo sapiens	age (y): 75
GSM989835	age 73y 1009	GSM989835	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1009	Homo sapiens	age (y): 73
GSM989836	age 83y 1010	GSM989836	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1010	Homo sapiens	age (y): 83
GSM989837	age 82y 1011	GSM989837	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1011	Homo sapiens	age (y): 82
GSM989839	age 77y 1013	GSM989839	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1013	Homo sapiens	age (y): 77
GSM989842	age 71y 1016	GSM989842	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1016	Homo sapiens	age (y): 71
GSM989843	age 68y 1017	GSM989843	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1017	Homo sapiens	age (y): 68
GSM989844	age 80y 1018	GSM989844	Public on Nov 21 2012	Aug 21 2012	Nov 21 2012	genomic	1	X1018	Homo sapiens	age (y): 80

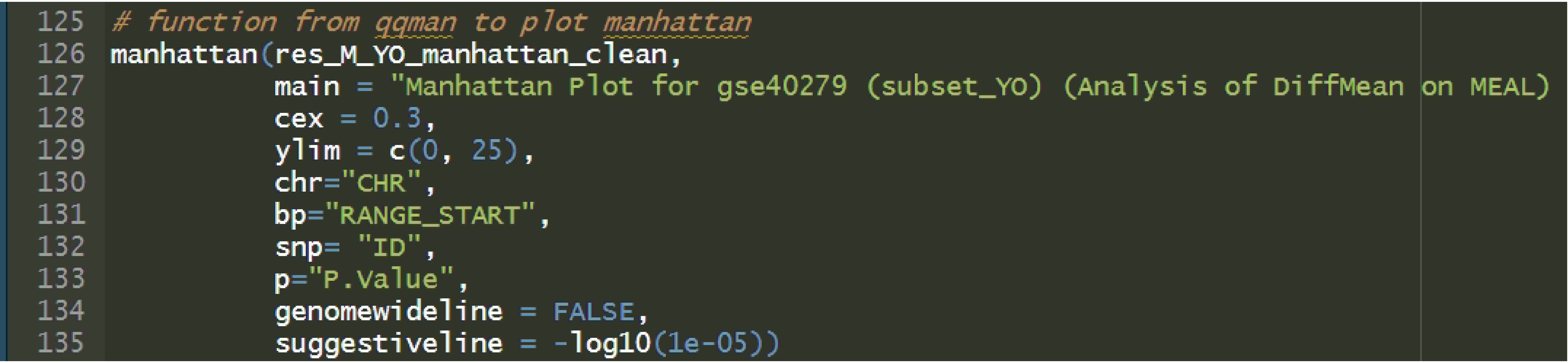
Showing 1 to 13 of 327 entries, 44 total columns



```r
# Run MEAL pipeline on the subset
res_sub_YO <- runPipeline(set = subset_forMeal_YO,
                   variable_names = "new_age_category",
                   betas = TRUE,
                   analyses = c("DiffMean", "DiffVar"))

# Extract the result of the DiffMean analysis
result_Meal_sub_YO <- getProbeResults(res_sub_YO, rid = 1,
                                   fNames = c("UCSC_RefGene_Name", "RANGE_START", "CHR", "ID"))

# Remove rows with missing values
result_Meal_sub_YO_clean <- na.omit(result_Meal_sub_YO)
```



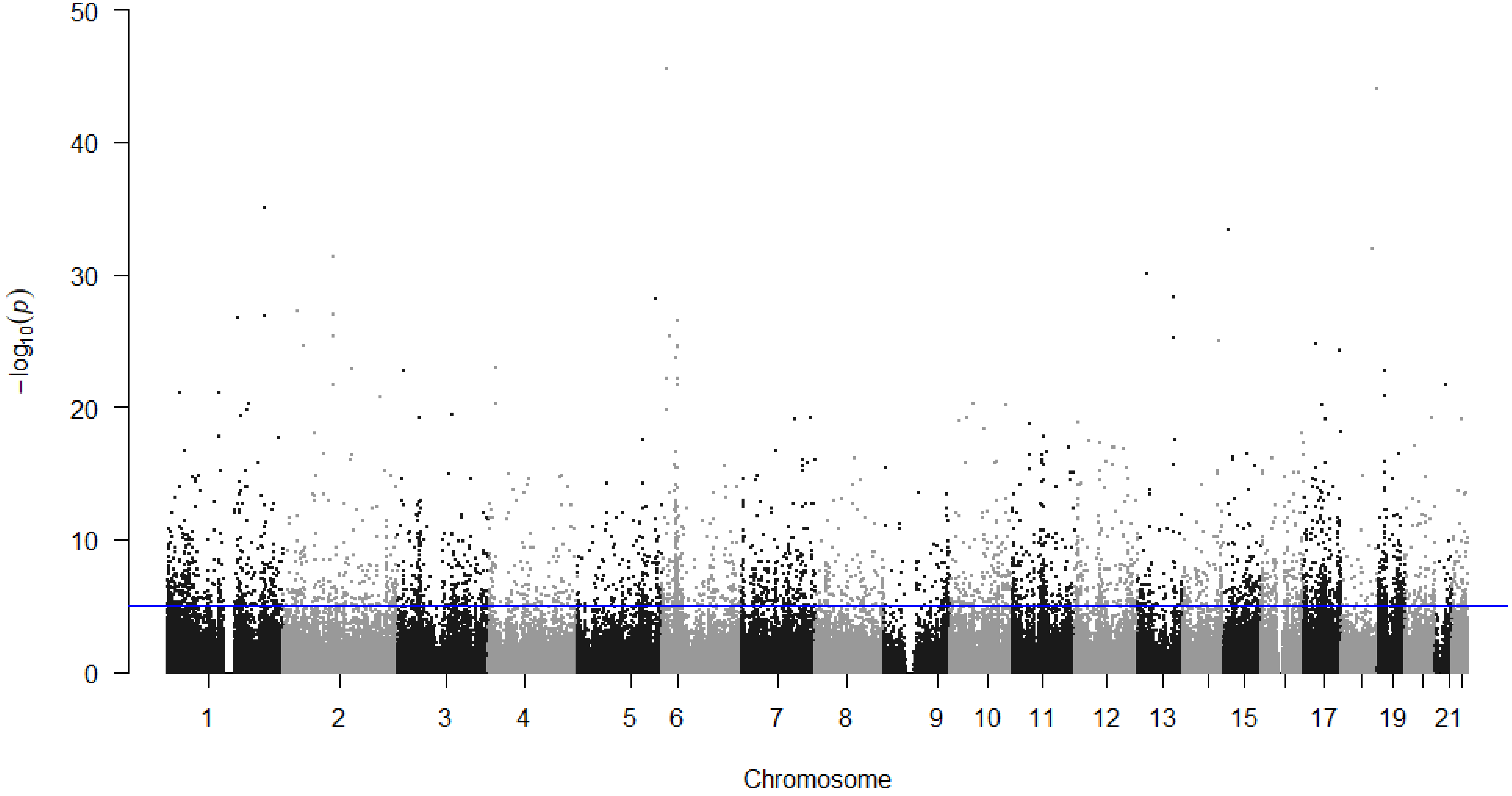
```r
## manhattan plot for subset YO
install.packages("qqman") # if you have not installed
library(qqman)

# Extract data for manhattan plot
res_M_YO_manhattan <- result_Meal_sub_YO_clean

# Convert CHR column to numeric
res_M_YO_manhattan$CHR <- as.numeric(res_M_YO_manhattan$CHR)

# Remove rows with missing values
res_M_YO_manhattan_clean <- na.omit(res_M_YO_manhattan)
```

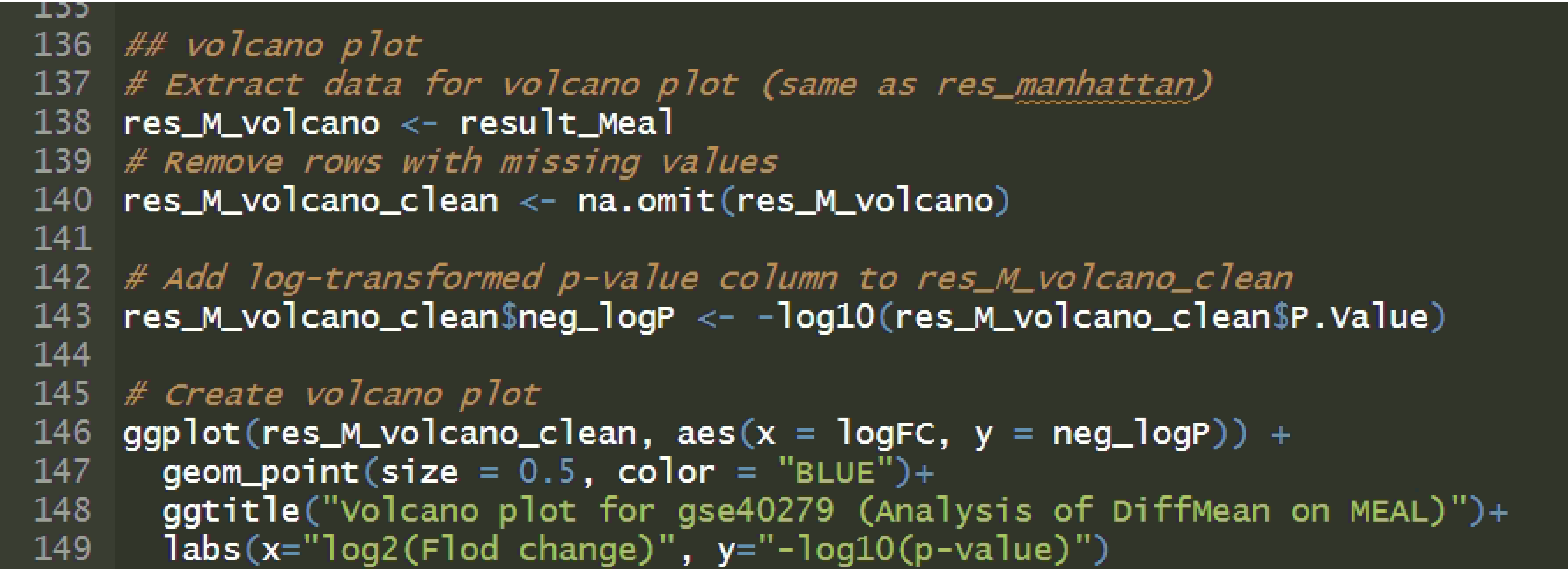


```r
# function from qqman to plot manhattan
manhattan(res_M_YO_manhattan_clean,
          main = "Manhattan Plot for gse40279 (subset_YO) (Analysis of DiffMean on MEAL)
          cex = 0.3,
          ylim = c(0, 25),
          chr="CHR",
          bp="RANGE_START",
          snp= "ID",
          p="P.Value",
          genomewideline = FALSE,
          suggestiveline = -log10(1e-05))
```

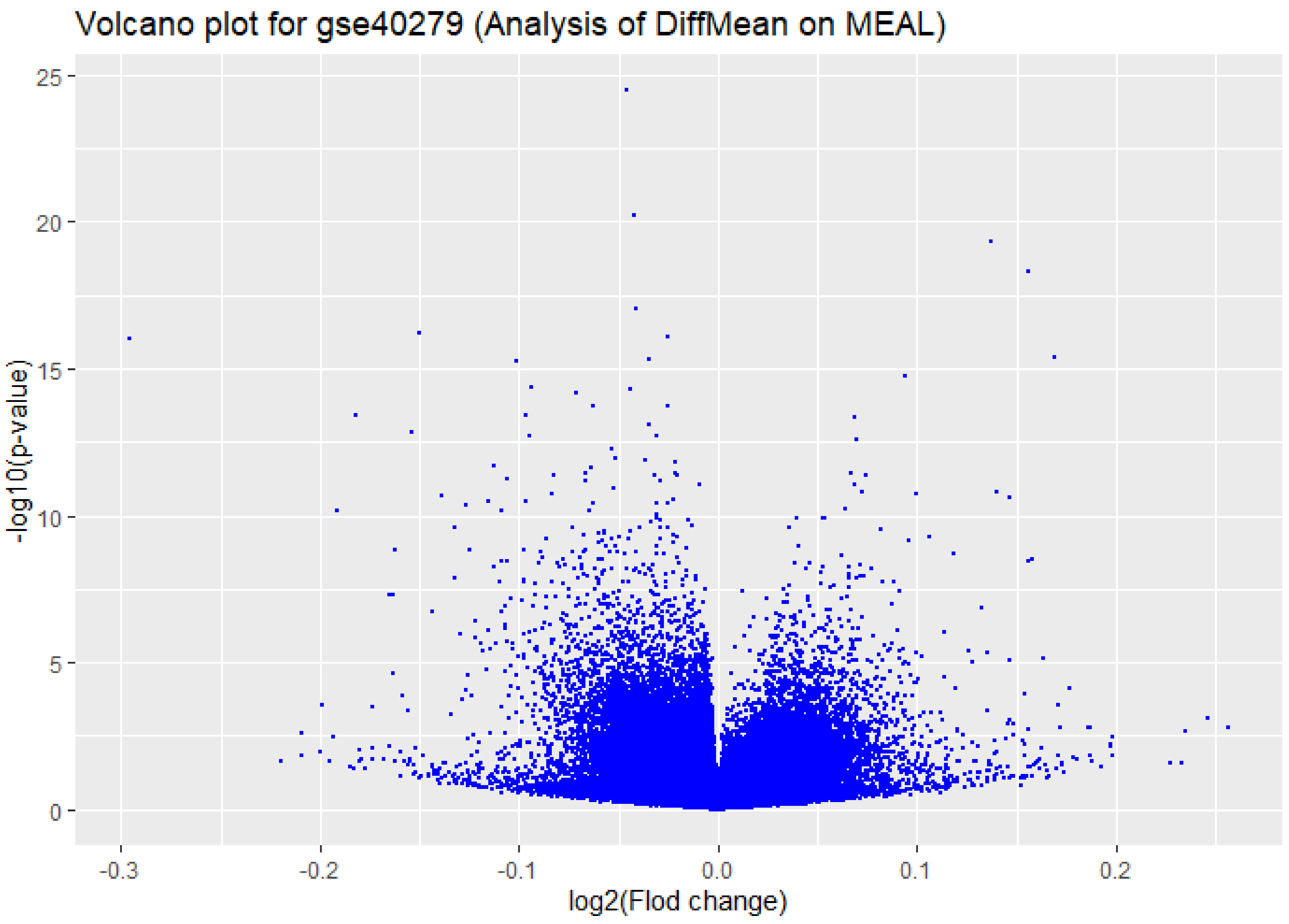
MEAL

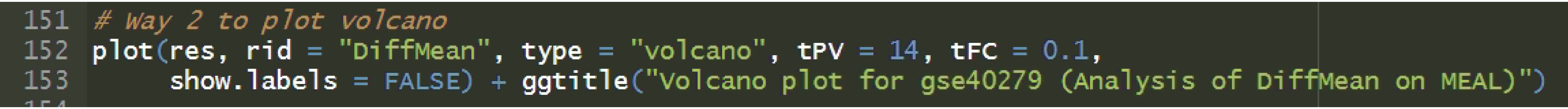
Manhattan Plot for gse40279 (subset_YO) (Analysis of DiffMean on MEAL)



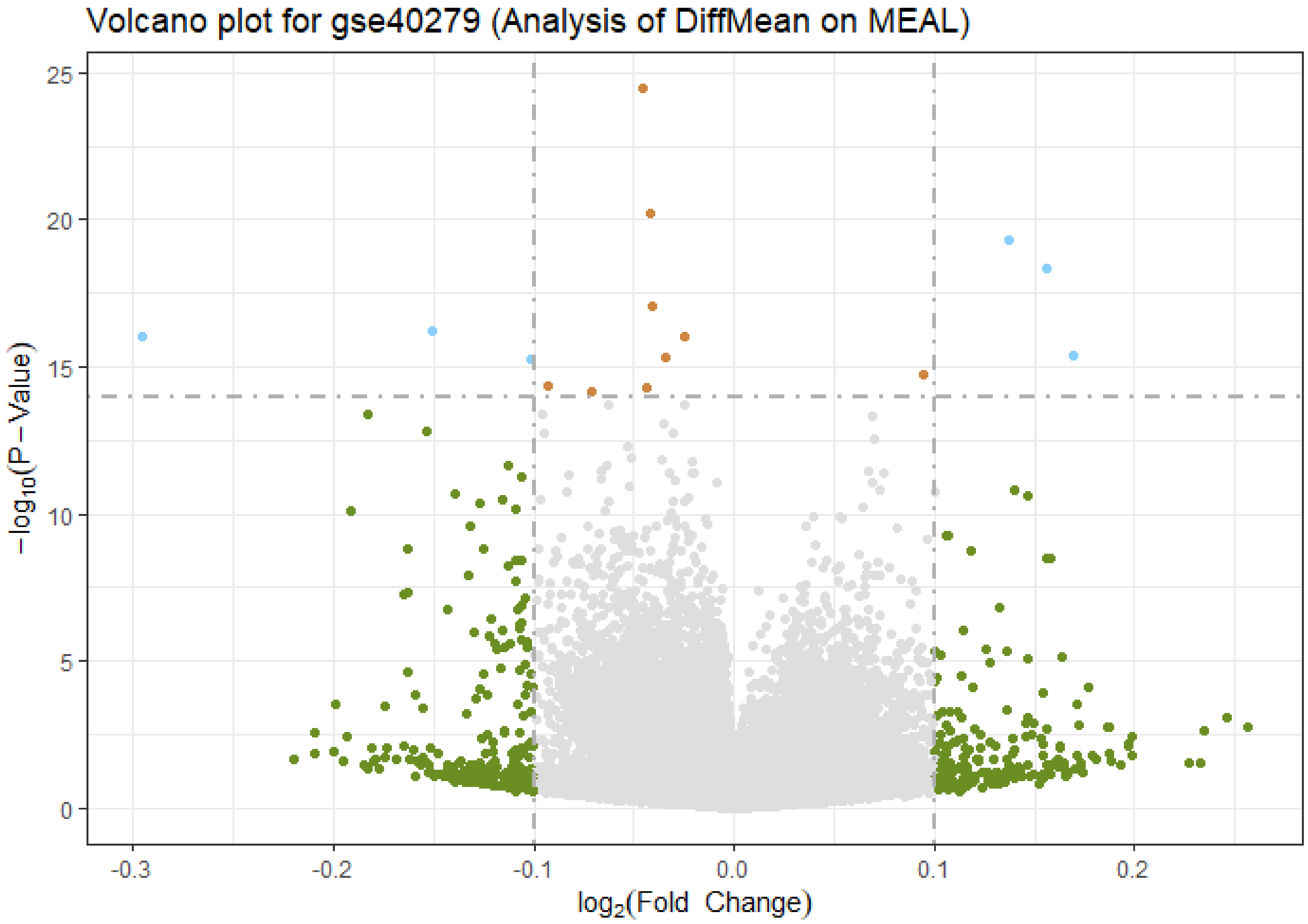


```r

## volcano plot
# Extract data for volcano plot (same as res_manhattan)
res_M_volcano <- result_Meal
# Remove rows with missing values
res_M_volcano_clean <- na.omit(res_M_volcano)

# Add log-transformed p-value column to res_M_volcano_clean
res_M_volcano_clean$neg_logP <- -log10(res_M_volcano_clean$P.Value)

# Create volcano plot
ggplot(res_M_volcano_clean, aes(x = logFC, y = neg_logP)) +
  geom_point(size = 0.5, color = "BLUE")+
  ggtitle("Volcano plot for gse40279 (Analysis of DiffMean on MEAL)")+
  labs(x="log2(Flod change)", y="-log10(p-value)")
```

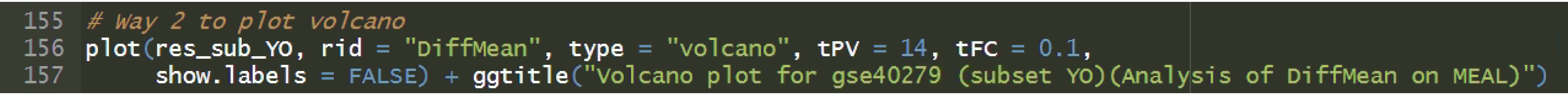

Volcano plot for gse40279 (Analysis of DiffMean on MEAL)
-log10(p-value)
25
20
15
10
5
0
-0.3
-0.2
-0.1
0.0
0.1
0.2
log2(Flod change)



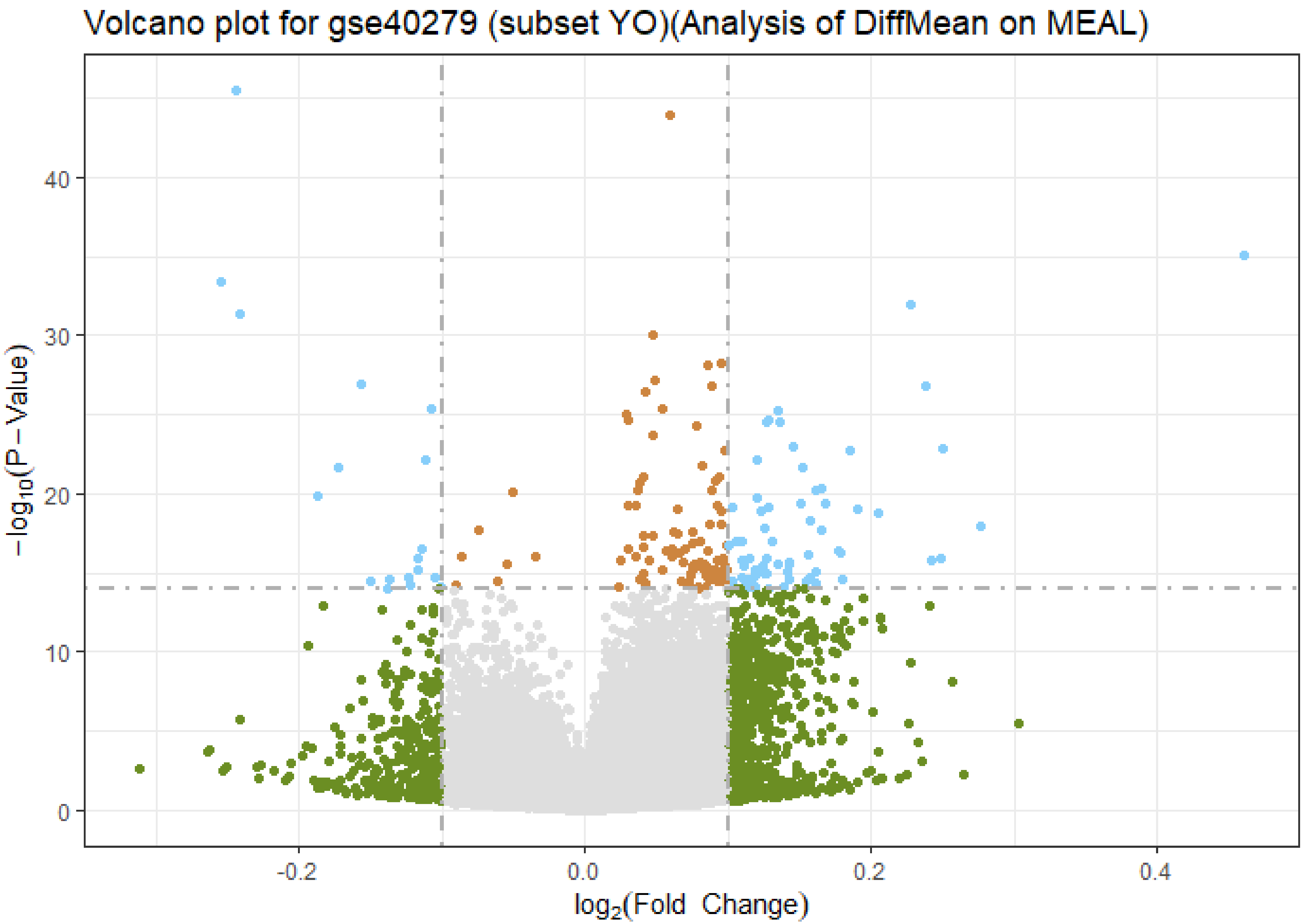
```r
# Way 2 to plot volcano
plot(res, rid = "DiffMean", type = "volcano", tPV = 14, tFC = 0.1,
     show.labels = FALSE) + ggtitle("Volcano plot for gse40279 (Analysis of DiffMean on MEAL)")
```



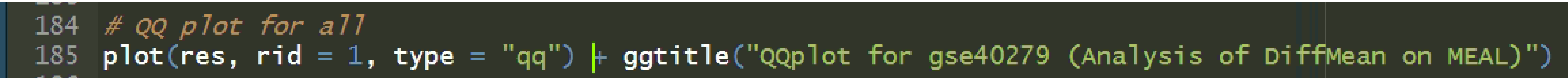
Volcano plot for gse40279 (Analysis of DiffMean on MEAL)



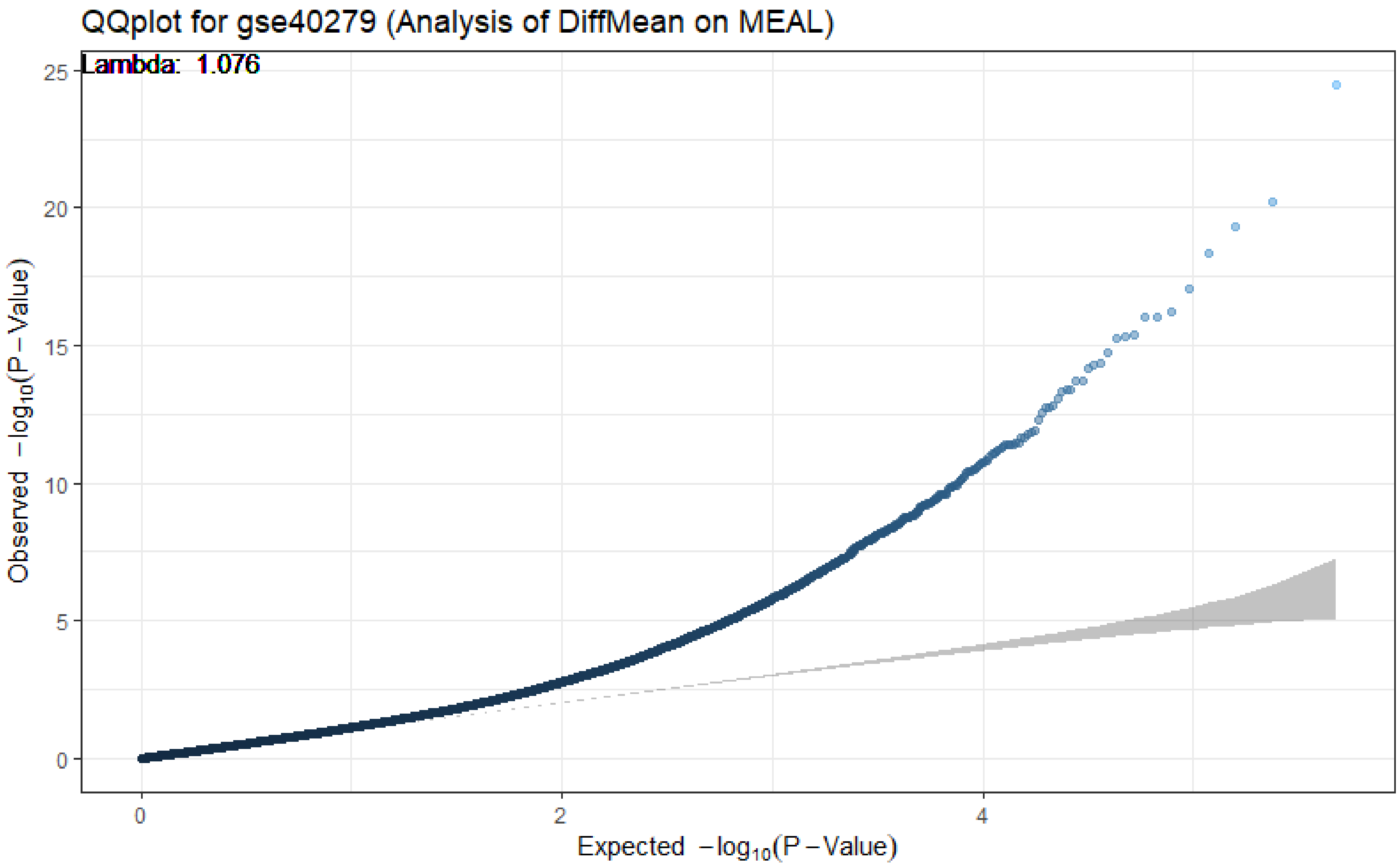
```
# Way 2 to plot volcano
plot(res_sub_YO, rid = "DiffMean", type = "volcano", tPV = 14, tFC = 0.1,
     show.labels = FALSE) + ggtitle("Volcano plot for gse40279 (subset YO)(Analysis of DiffMean on MEAL)")
```



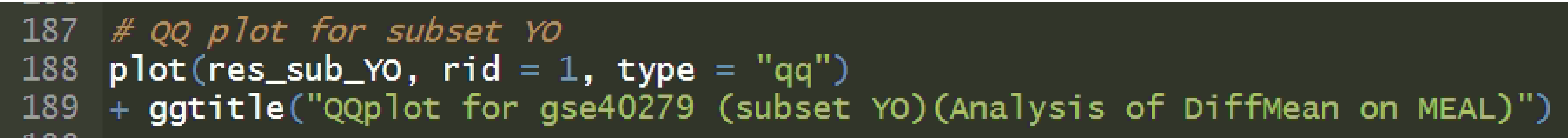
QQplot



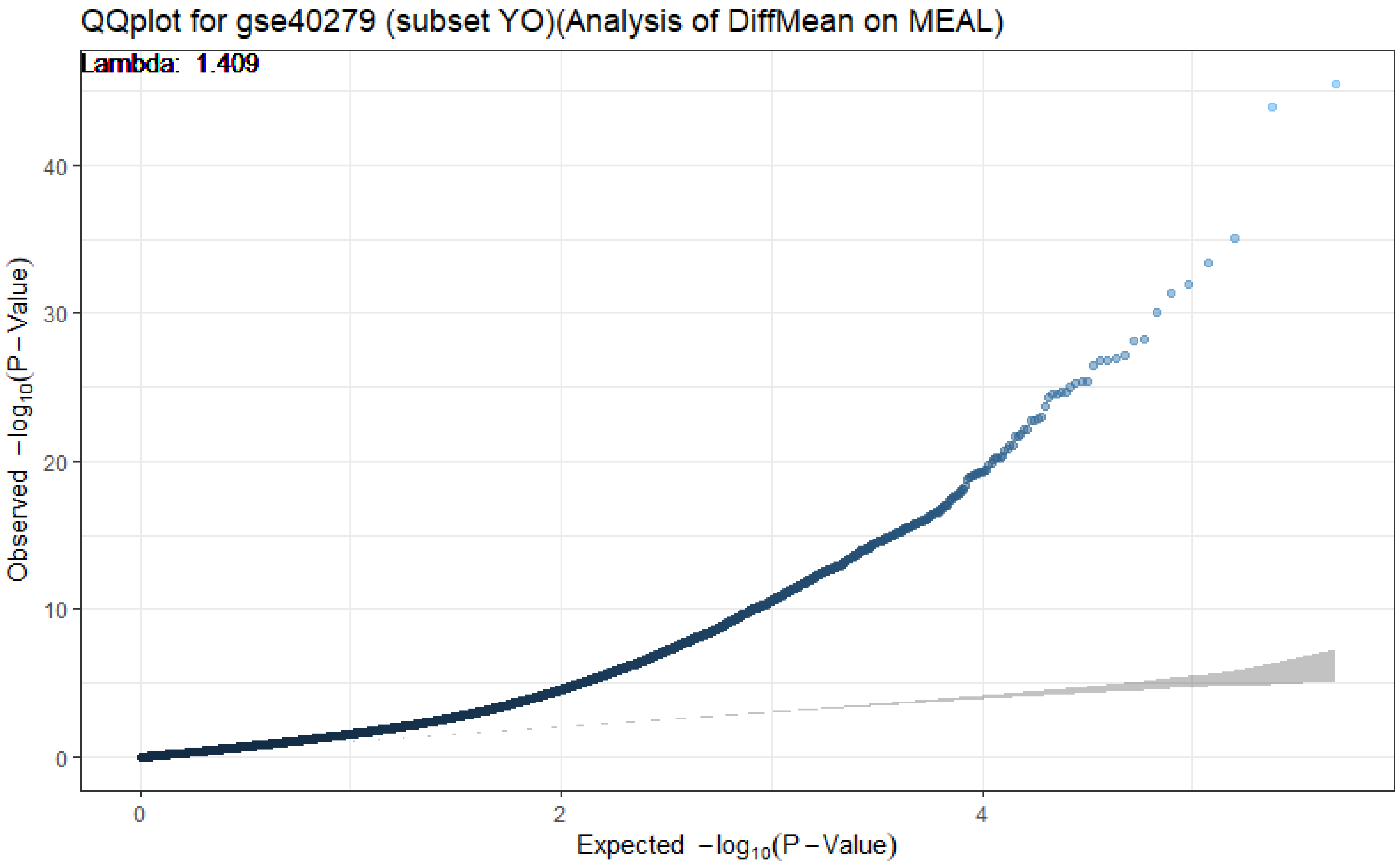
```
# QQ plot for all
plot(res, rid = 1, type = "qq") + ggtitle("QQplot for gse40279 (Analysis of DiffMean on MEAL)")
```



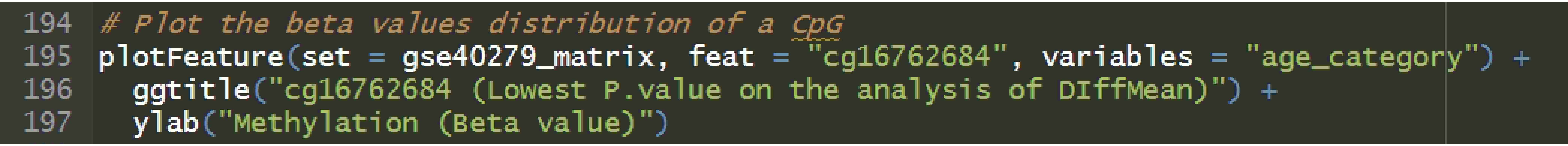
QQplot



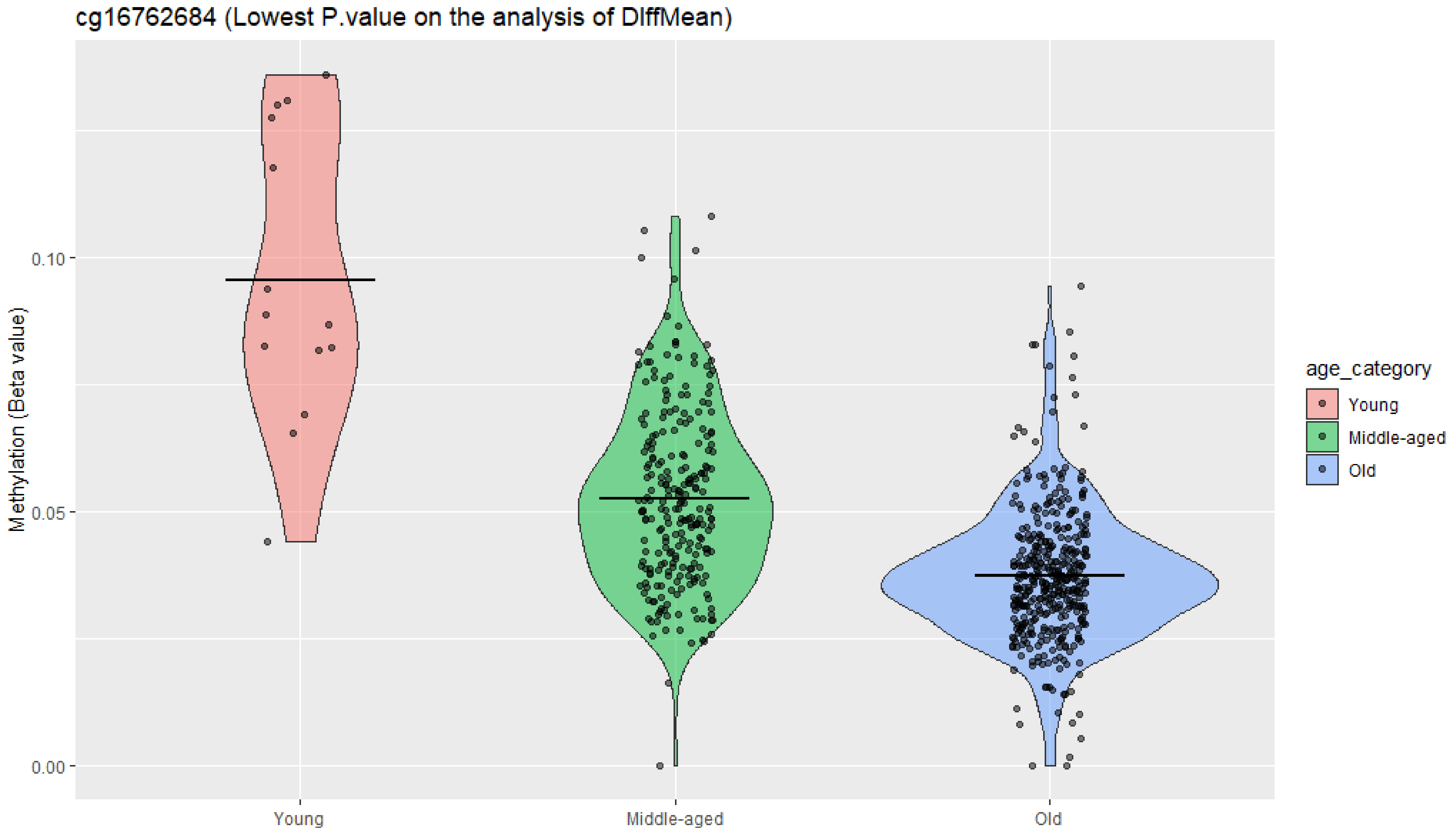
```
# QQ plot for subset YO
plot(res_sub_YO, rid = 1, type = "qq")
+ ggtitle("QQplot for gse40279 (subset YO)(Analysis of DiffMean on MEAL)")
```



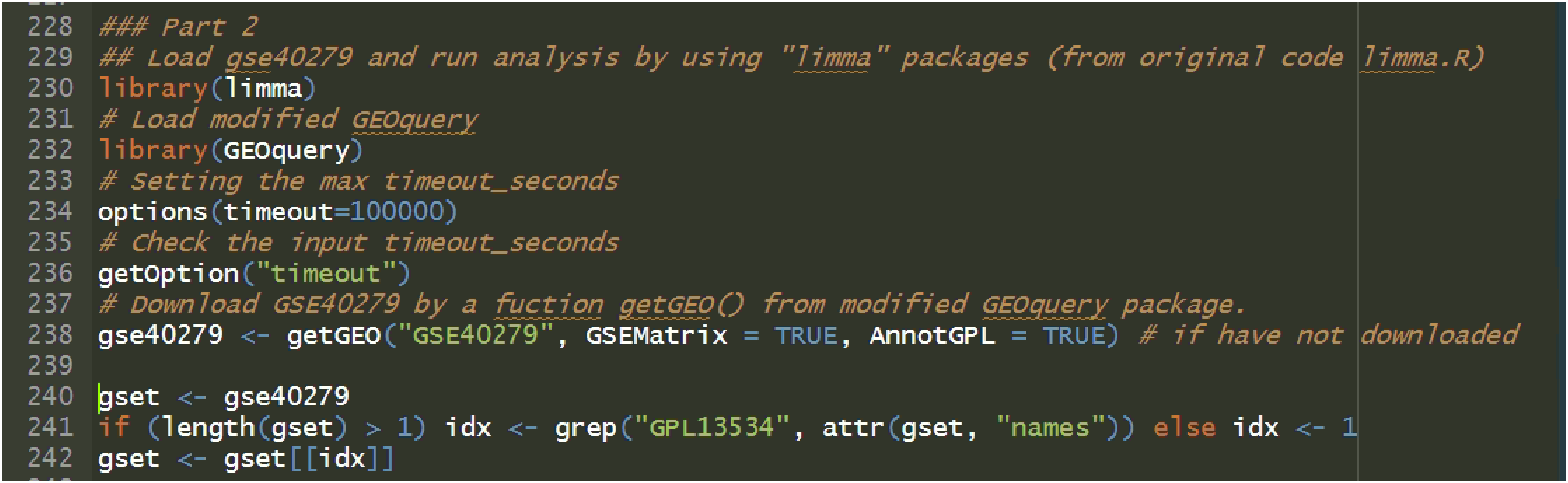
beta-value distribution of a specific probe



```r
# Plot the beta values distribution of a CpG
plotFeature(set = gse40279_matrix, feat = "cg16762684", variables = "age_category") +
  ggtitle("cg16762684 (Lowest P.value on the analysis of DIffMean)") +
  ylab("Methylation (Beta value)")
```

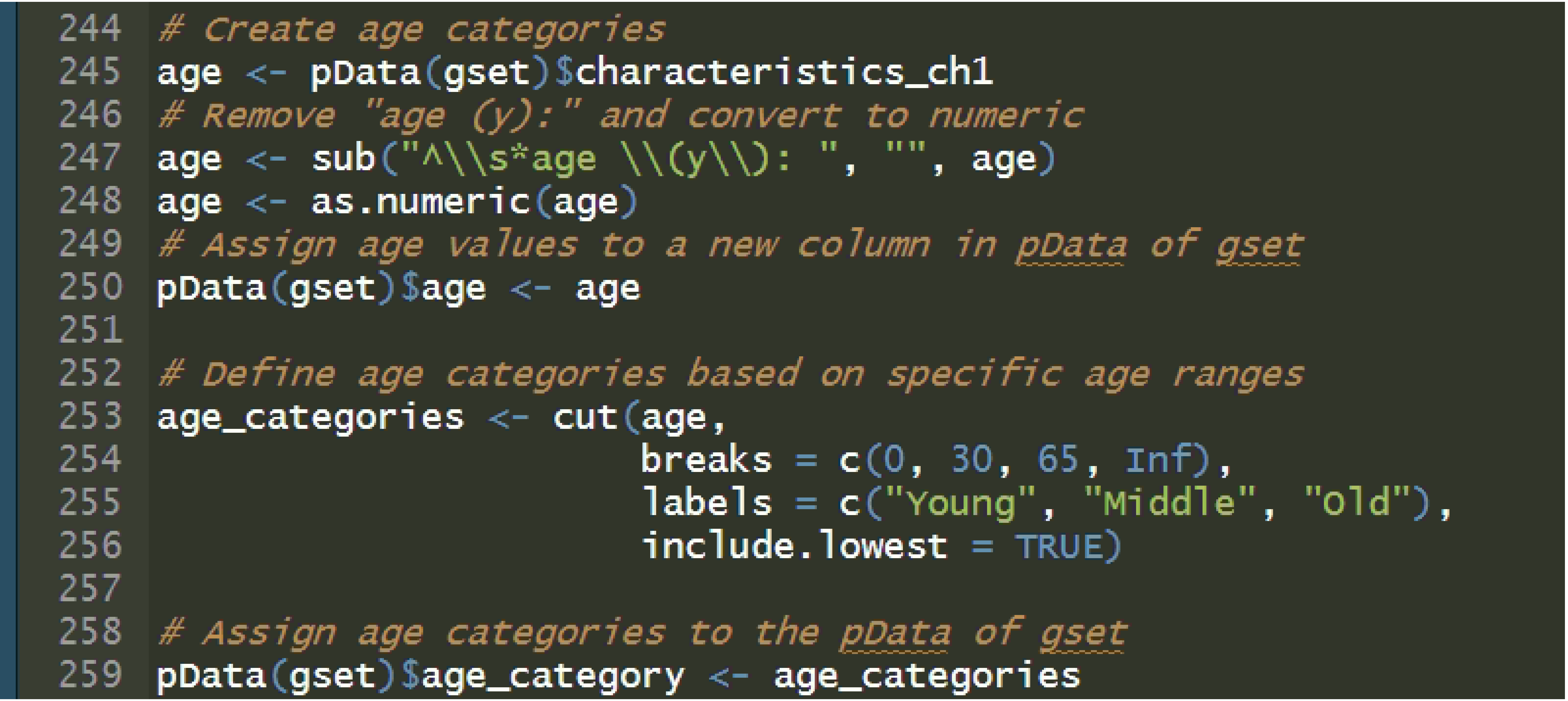


limma



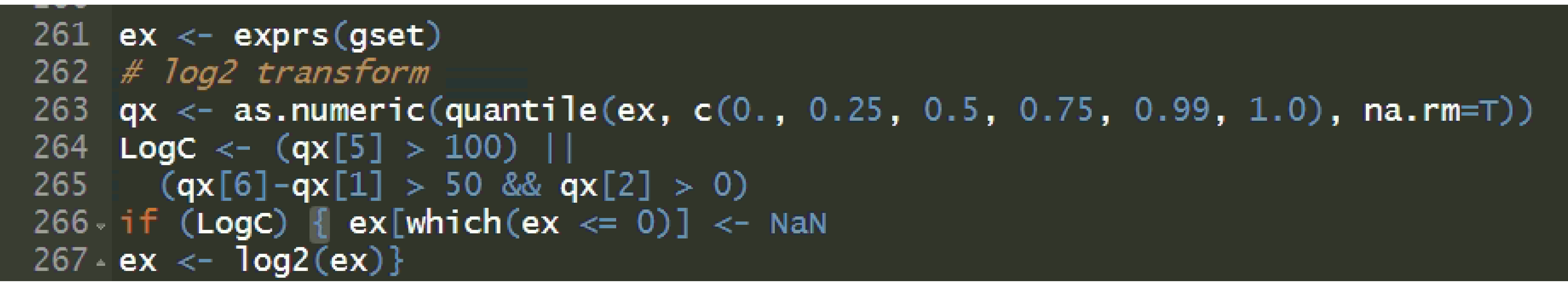
```r
### Part 2
## Load gse40279 and run analysis by using "limma" packages (from original code limma.R)
library(limma)
# Load modified GEOquery
library(GEOquery)
# Setting the max timeout_seconds
options(timeout=100000)
# Check the input timeout_seconds
getOption("timeout")
# Download GSE40279 by a fuction getGEO() from modified GEOquery package.
gse40279 <- getGEO("GSE40279", GSEMatrix = TRUE, AnnotGPL = TRUE) # if have not downloaded

gset <- gse40279
if (length(gset) > 1) idx <- grep("GPL13534", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]
```

limma



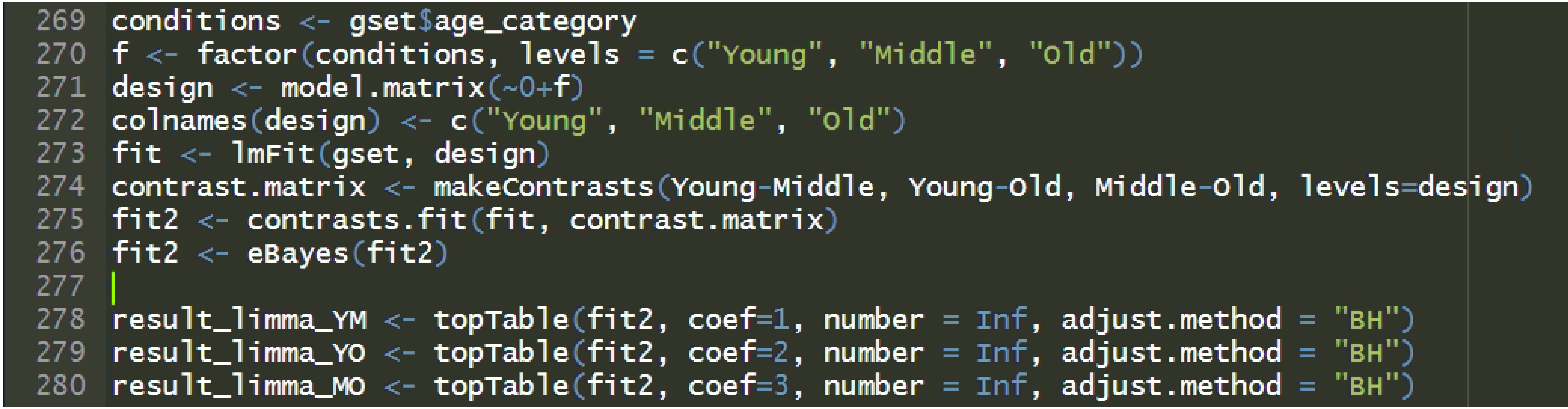
```r
# Create age categories
age <- pData(gset)$characteristics_ch1
# Remove "age (y):" and convert to numeric
age <- sub("^\\s*age \\(y\\): ", "", age)
age <- as.numeric(age)
# Assign age values to a new column in pData of gset
pData(gset)$age <- age

# Define age categories based on specific age ranges
age_categories <- cut(age,
                      breaks = c(0, 30, 65, Inf),
                      labels = c("Young", "Middle", "Old"),
                      include.lowest = TRUE)

# Assign age categories to the pData of gset
pData(gset)$age_category <- age_categories
```

limma



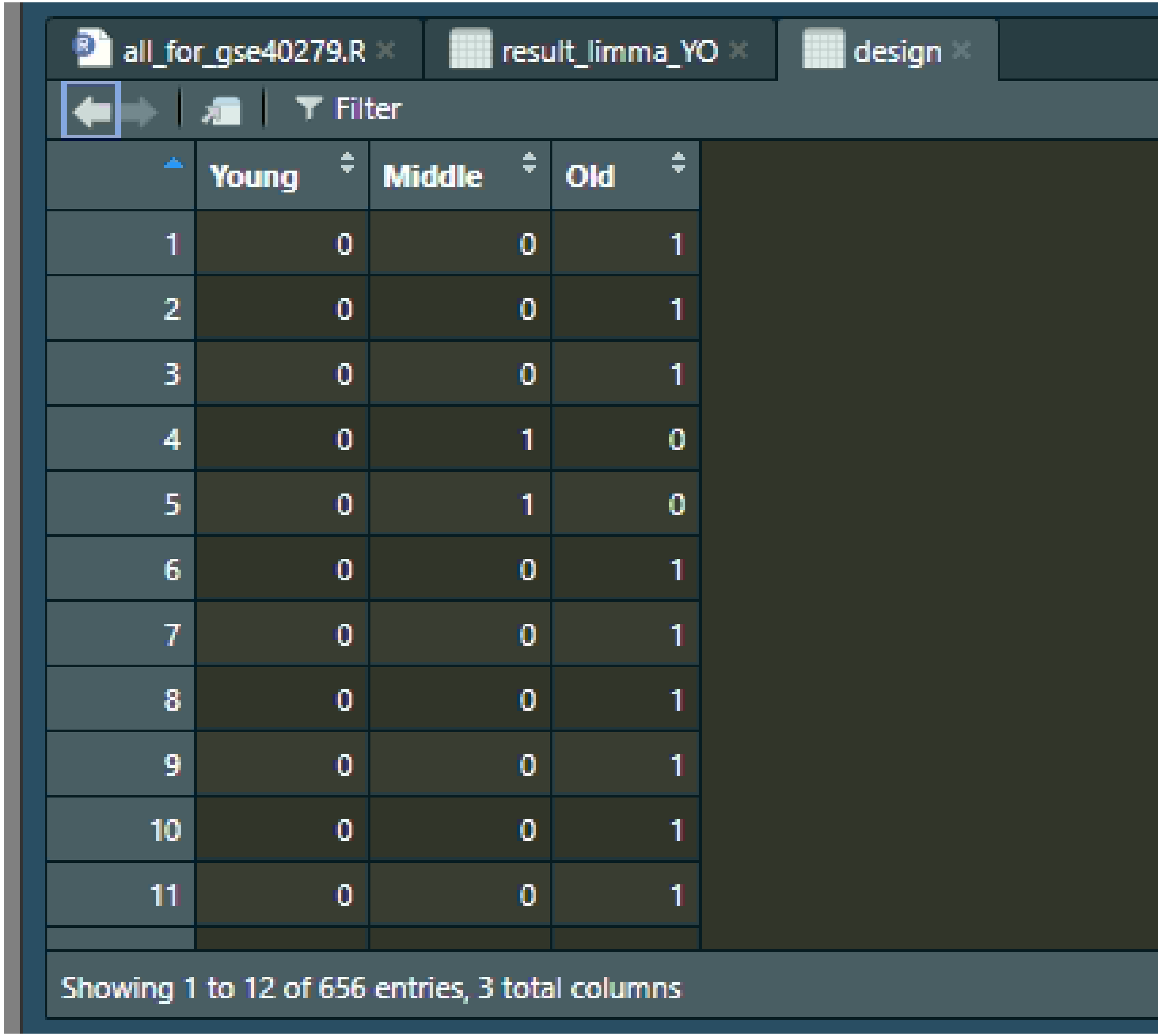
```
ex <- exprs(gset)
# log2 transform
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
  (qx[6]-qx[1] > 50 && qx[2] > 0)
if (LogC) { ex[which(ex <= 0)] <- NaN
ex <- log2(ex)}
```

修正limma design部分



```r
conditions <- gset$age_category
f <- factor(conditions, levels = c("Young", "Middle", "Old"))
design <- model.matrix(~0+f)
colnames(design) <- c("Young", "Middle", "Old")
fit <- lmFit(gset, design)
contrast.matrix <- makeContrasts(Young-Middle, Young-Old, Middle-Old, levels=design)
fit2 <- contrasts.fit(fit, contrast.matrix)
fit2 <- eBayes(fit2)

result_limma_YM <- topTable(fit2, coef=1, number = Inf, adjust.method = "BH")
result_limma_YO <- topTable(fit2, coef=2, number = Inf, adjust.method = "BH")
result_limma_MO <- topTable(fit2, coef=3, number = Inf, adjust.method = "BH")
```

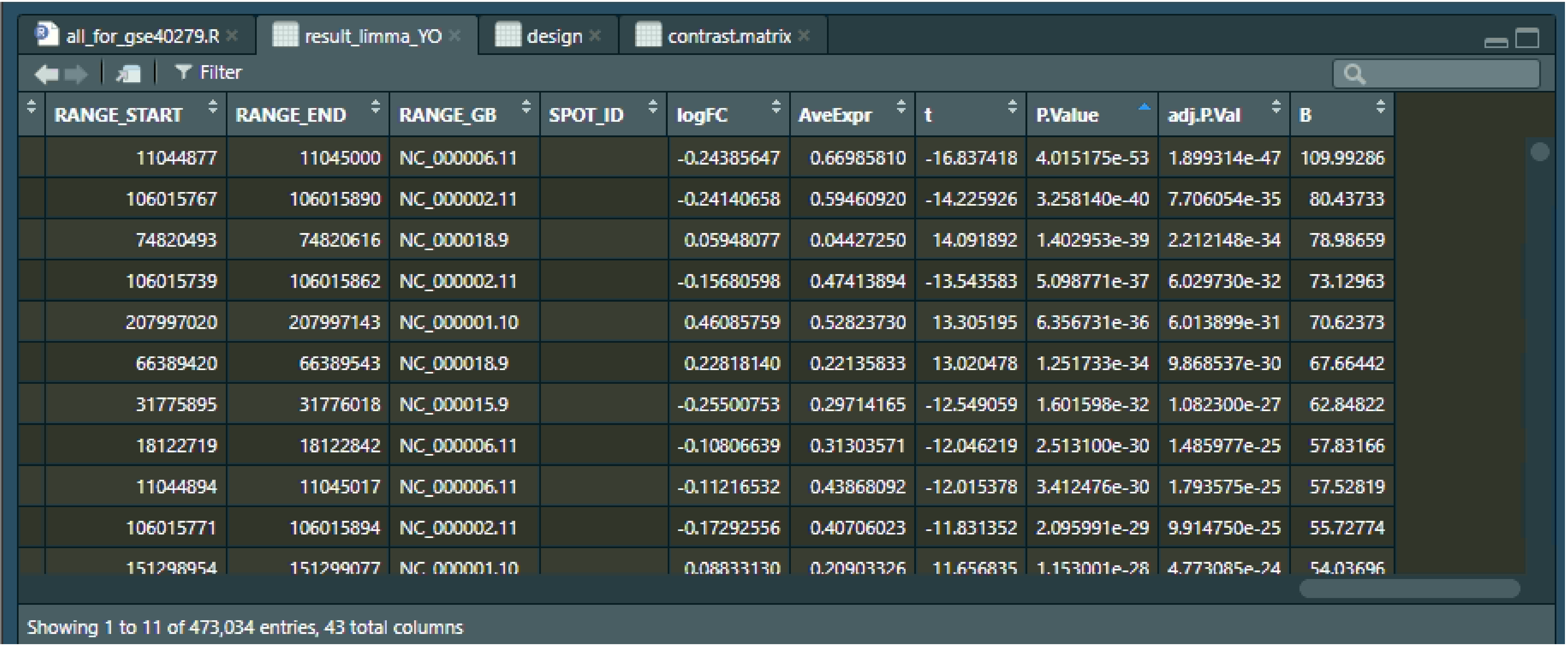
limma



	Young	Middle	Old
1	0	0	1
2	0	0	1
3	0	0	1
4	0	1	0
5	0	1	0
6	0	0	1
7	0	0	1
8	0	0	1
9	0	0	1
10	0	0	1
11	0	0	1

Showing 1 to 12 of 656 entries, 3 total columns

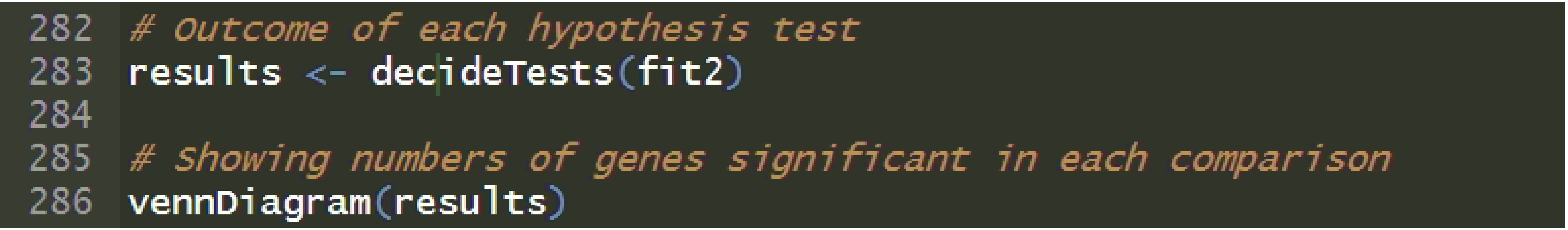
limma



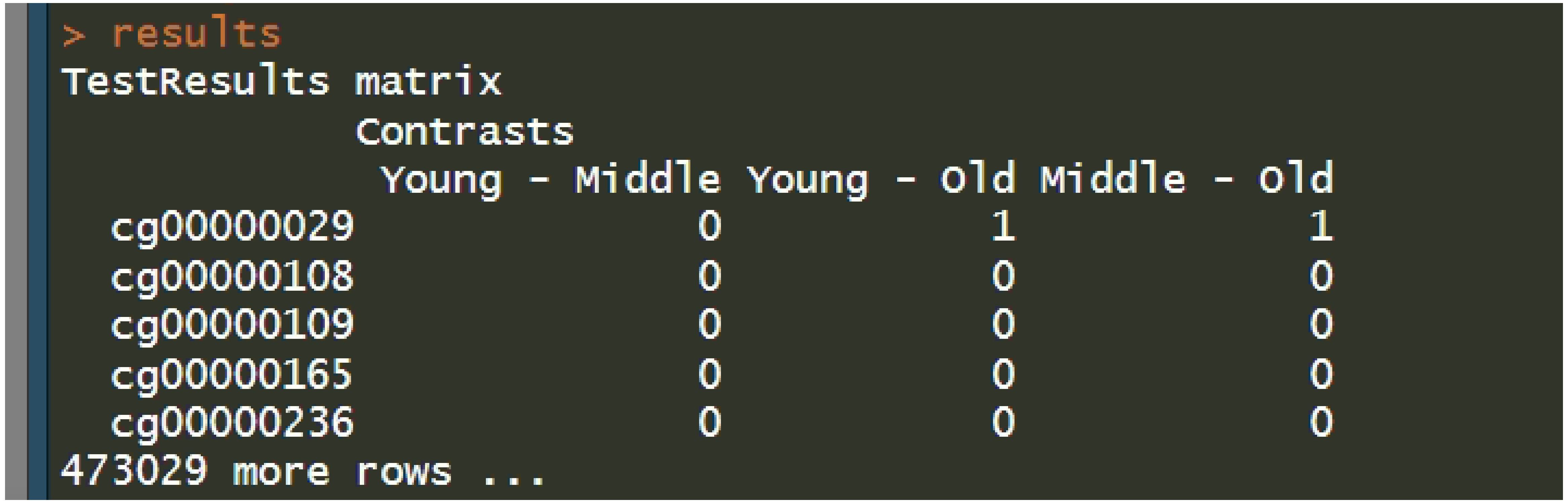
	RANGE_START	RANGE_END	RANGE_GB	SPOT_ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
	11044877	11045000	NC_000006.11		-0.24385647	0.66985810	-16.837418	4.015175e-53	1.899314e-47	109.99286
	106015767	106015890	NC_000002.11		-0.24140658	0.59460920	-14.225926	3.258140e-40	7.706054e-35	80.43733
	74820493	74820616	NC_000018.9		0.05948077	0.04427250	14.091892	1.402953e-39	2.212148e-34	78.98659
	106015739	106015862	NC_000002.11		-0.15680598	0.47413894	-13.543583	5.098771e-37	6.029730e-32	73.12963
	207997020	207997143	NC_000001.10		0.46085759	0.52823730	13.305195	6.356731e-36	6.013899e-31	70.62373
	66389420	66389543	NC_000018.9		0.22818140	0.22135833	13.020478	1.251733e-34	9.868537e-30	67.66442
	31775895	31776018	NC_000015.9		-0.25500753	0.29714165	-12.549059	1.601598e-32	1.082300e-27	62.84822
	18122719	18122842	NC_000006.11		-0.10806639	0.31303571	-12.046219	2.513100e-30	1.485977e-25	57.83166
	11044894	11045017	NC_000006.11		-0.11216532	0.43868092	-12.015378	3.412476e-30	1.793575e-25	57.52819
	106015771	106015894	NC_000002.11		-0.17292556	0.40706023	-11.831352	2.095991e-29	9.914750e-25	55.72774
	151298954	151299077	NC_000001.10		0.08833130	0.20903326	11.656835	1.153001e-28	4.773085e-24	54.03696

Showing 1 to 11 of 473,034 entries, 43 total columns

limma

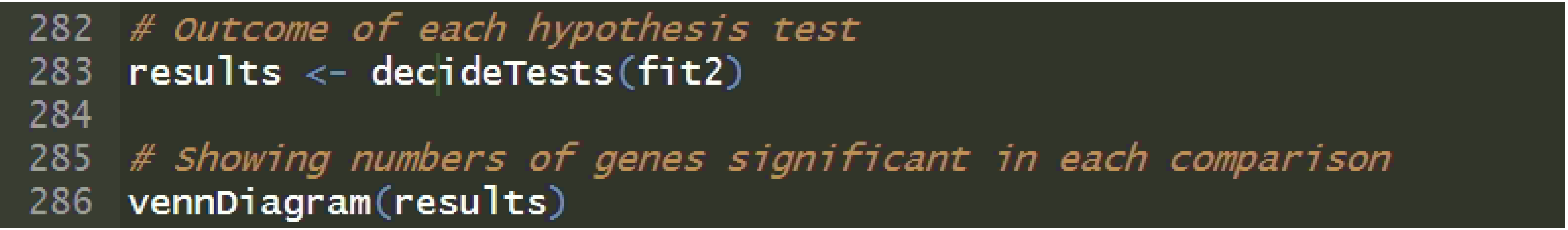


```r
# Outcome of each hypothesis test
results <- decideTests(fit2)

# Showing numbers of genes significant in each comparison
vennDiagram(results)
```

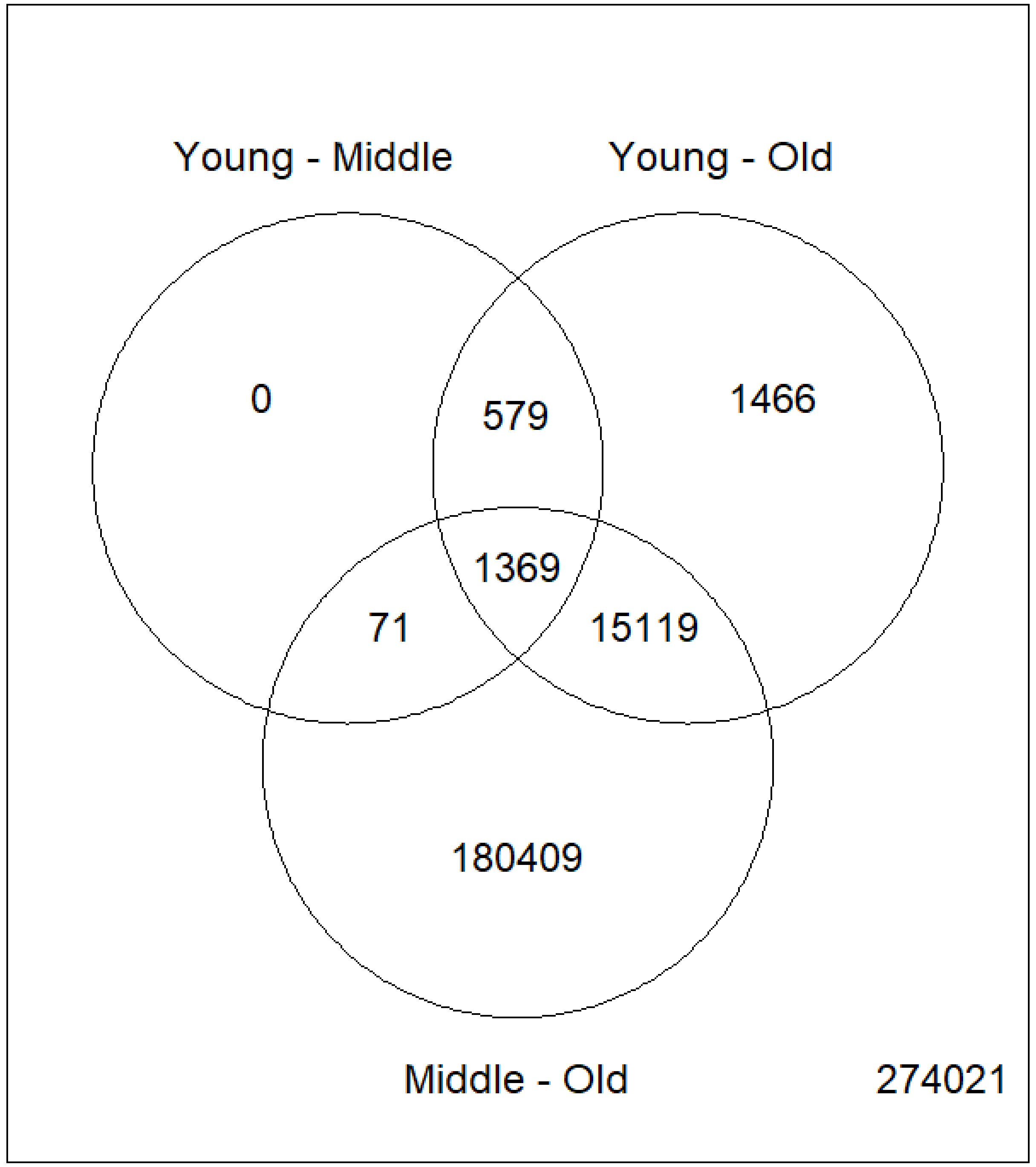


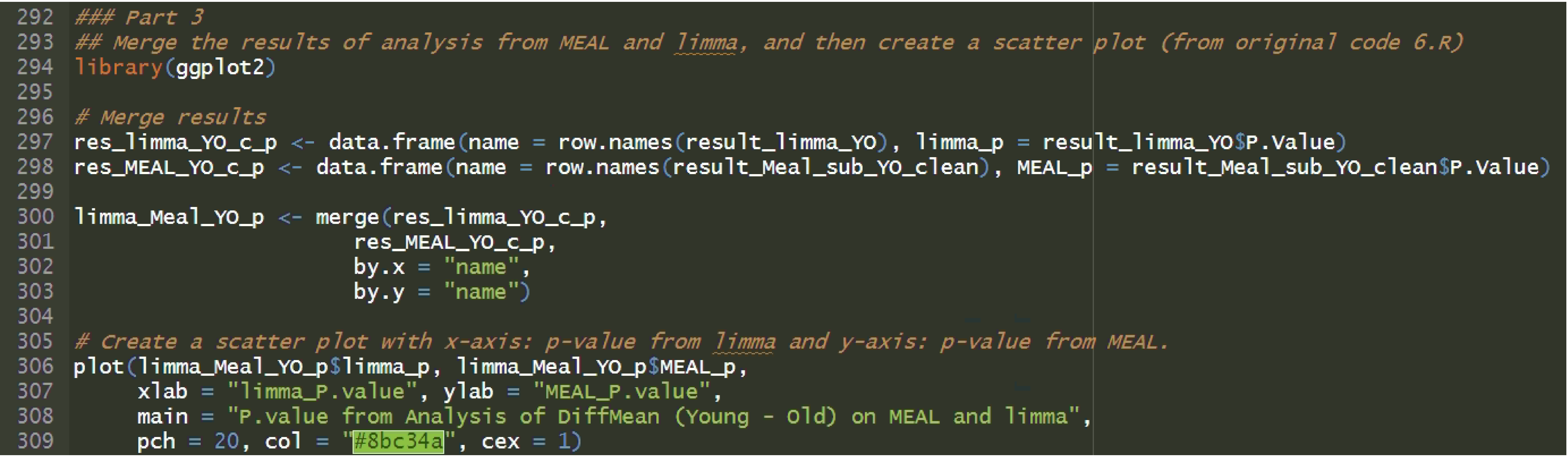
```
> results
TestResults matrix
            Contrasts
             Young - Middle Young - Old Middle - Old
  cg00000029              0           1            1
  cg00000108              0           0            0
  cg00000109              0           0            0
  cg00000165              0           0            0
  cg00000236              0           0            0
473029 more rows ...
```

number of significant genes by limma

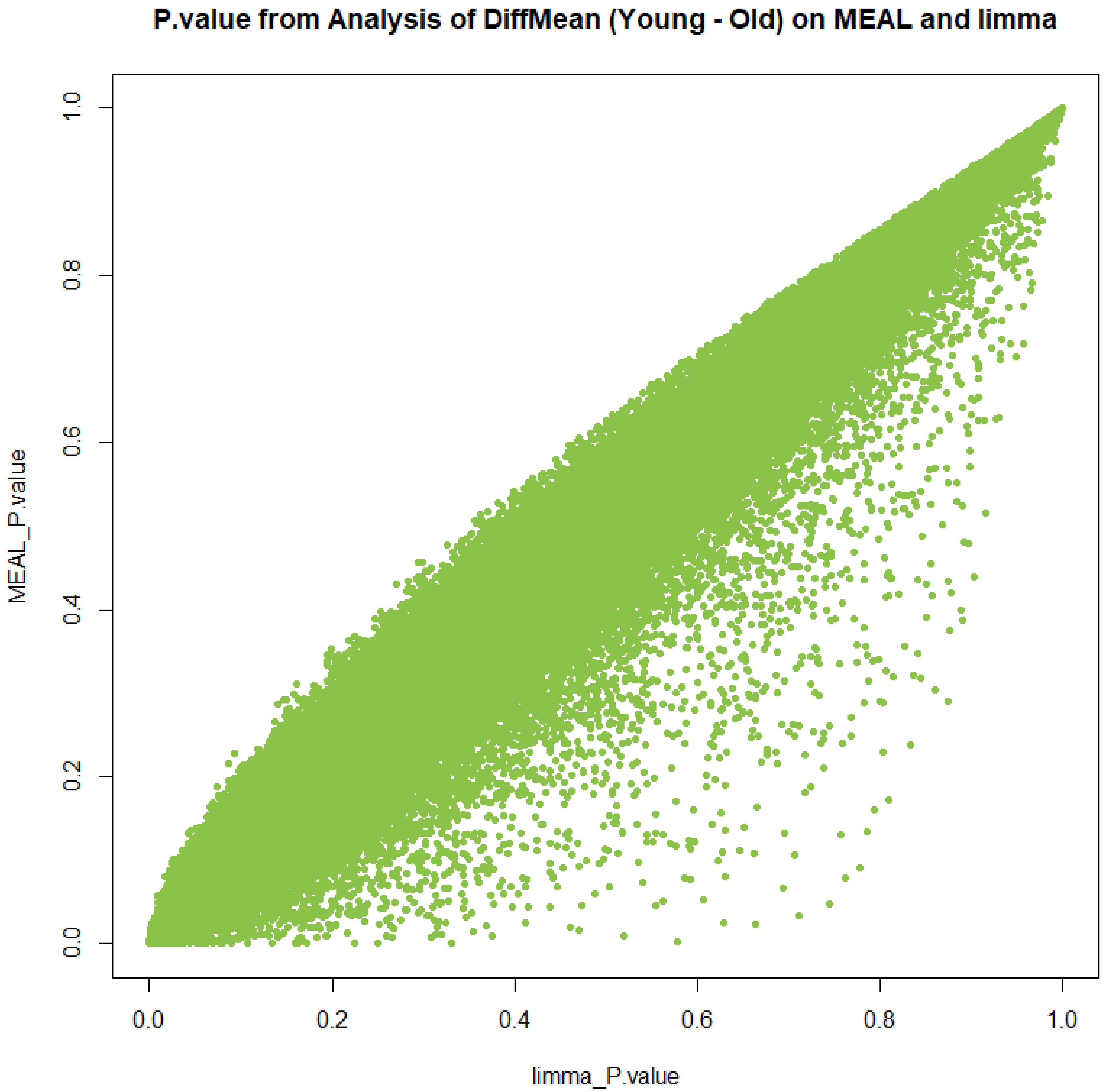


```
# Outcome of each hypothesis test
results <- decideTests(fit2)

# Showing numbers of genes significant in each comparison
vennDiagram(results)
```



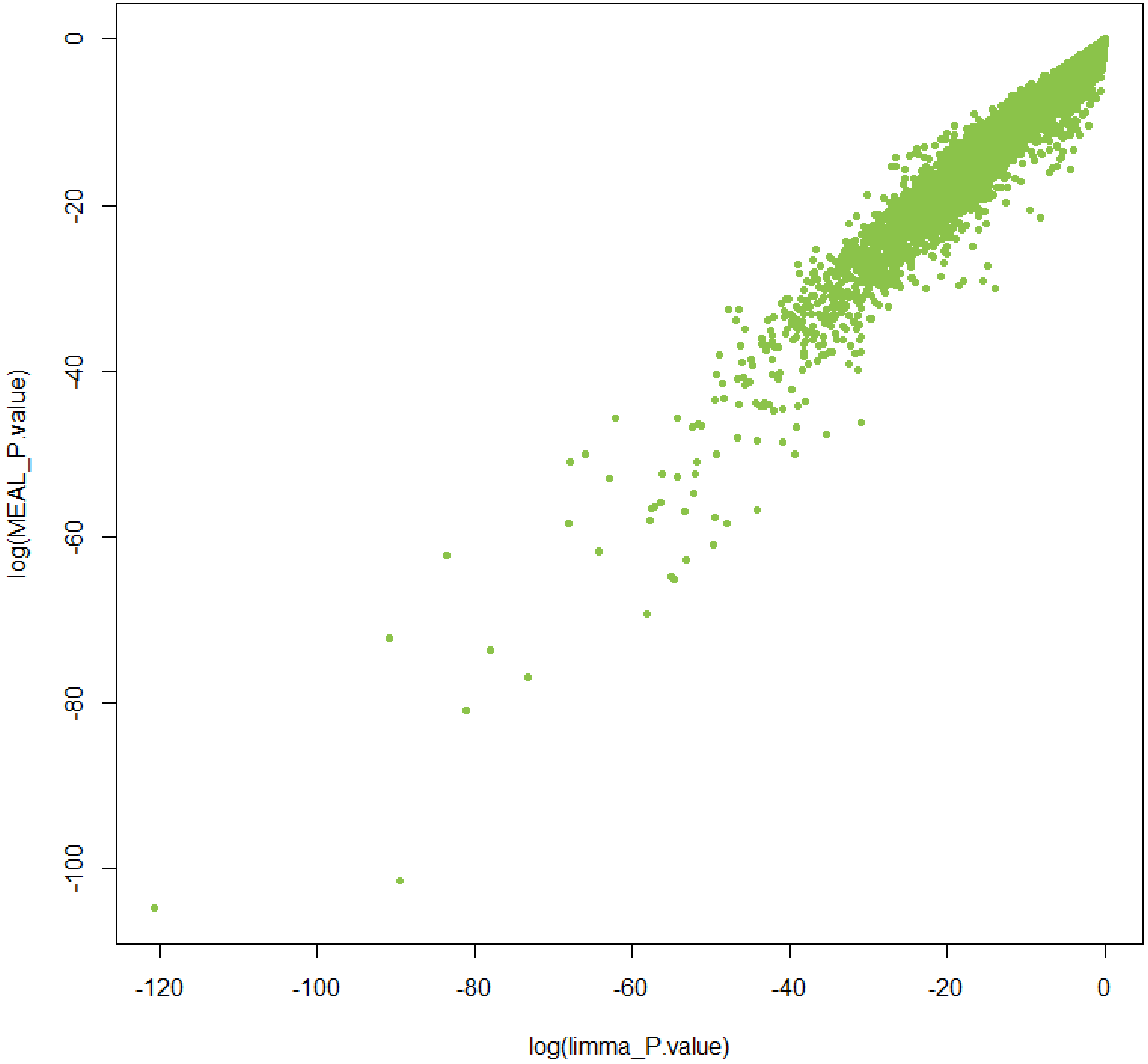


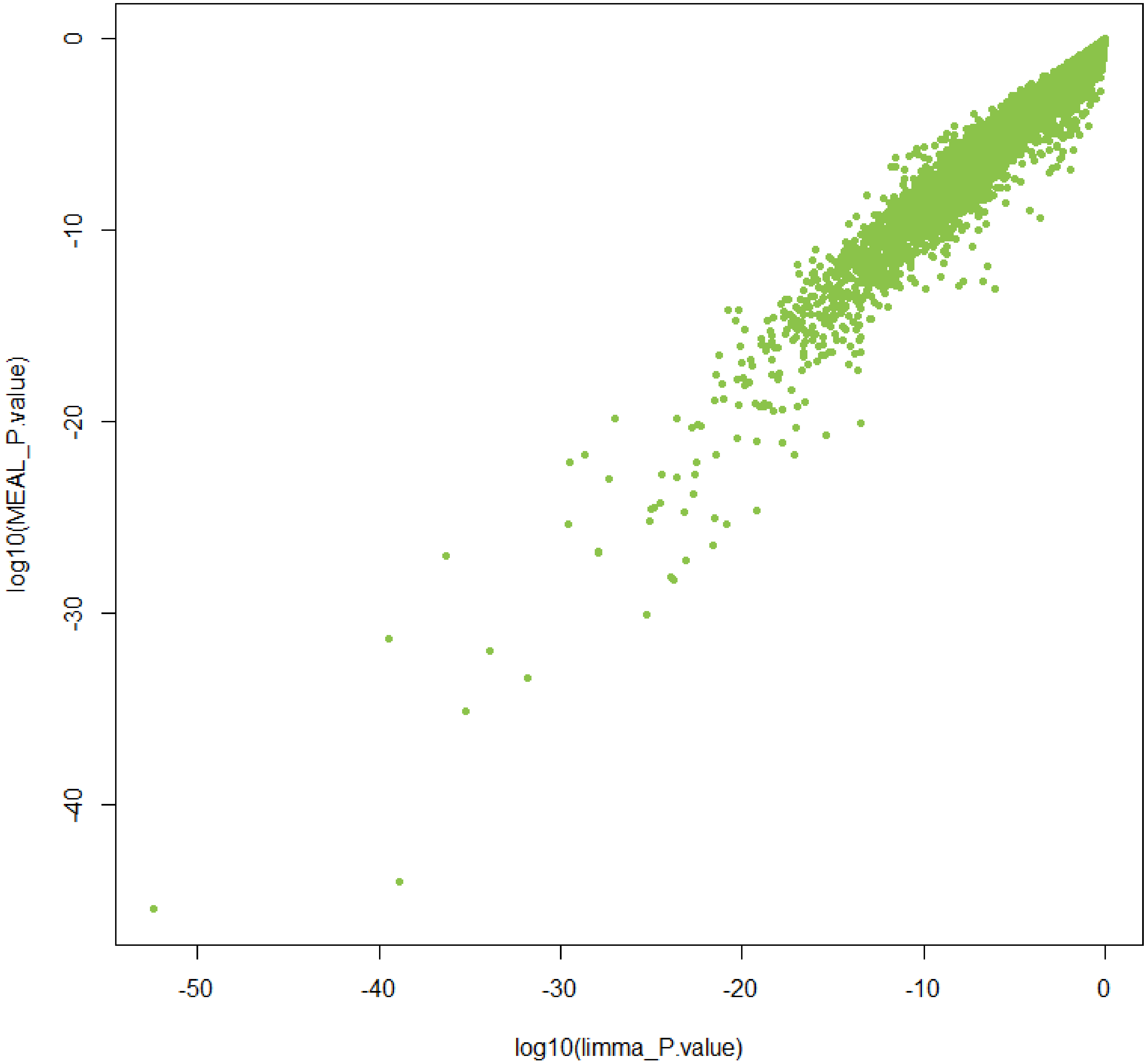
```r
### Part 3
## Merge the results of analysis from MEAL and limma, and then create a scatter plot (from original code 6.R)
library(ggplot2)

# Merge results
res_limma_YO_c_p <- data.frame(name = row.names(result_limma_YO), limma_p = result_limma_YO$P.Value)
res_MEAL_YO_c_p <- data.frame(name = row.names(result_Meal_sub_YO_clean), MEAL_p = result_Meal_sub_YO_clean$P.Value)

limma_Meal_YO_p <- merge(res_limma_YO_c_p,
                      res_MEAL_YO_c_p,
                      by.x = "name",
                      by.y = "name")

# Create a scatter plot with x-axis: p-value from limma and y-axis: p-value from MEAL.
plot(limma_Meal_YO_p$limma_p, limma_Meal_YO_p$MEAL_p,
     xlab = "limma_P.value", ylab = "MEAL_P.value",
     main = "P.value from Analysis of DiffMean (Young - Old) on MEAL and limma",
     pch = 20, col = "#8bc34a", cex = 1)
```



P.value from Analysis of DiffMean (Young - Old) on MEAL and limma

P.value from Analysis of DiffMean (Young - Old) on MEAL and limma





P.value from Analysis of DiffMean (Young - Old) on MEAL and limma