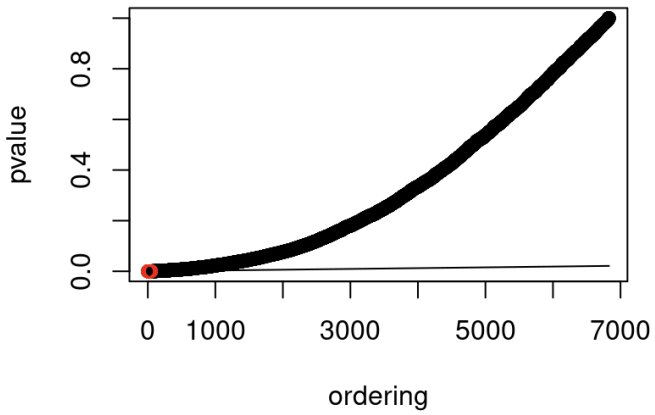
Assignment 6 Code + Analysis

```
library(ISLR)
source("fdr.pck")
library(stats)
data("NCI60")
```

Colon Cancer

Getting the fdr p value for colon cancer, and testing on 2, 1, and 0.05
The fdr value represents the False Discovery Rate

```
colon <- fdr(apply(COL,2,myt),.2,ind=F)
colon
interesting_colon <- colon$interesting

fdr(apply(COL,2,myt),.1,ind=F)
fdr(apply(COL,2,myt),.05,ind=F)
```



Using quartile value 0.2

```
> colon
$interesting
 [1]  256  257  286 5856 5854  251 5950 2640 1867  262 1832 5838 4100 5060 5983  247  181  143  248 5637
[21]  254  243  249  242 5230 4499 4160 4098  134 4550 6248  261 5512 3956  216  349  253 2888 6664 2632
[41] 1930 4384 2551 5956 5859 4394 5535  237

$ind
[1] FALSE
```

Using quartile value 0.1:

```
> fdr(apply(COL,2,myt),.1,ind=F)
$interesting
 [1]  256  257  286 5856 5854  251 5950 2640 1867  262 1832 5838 4100 5060 5983  247  181  143  248 5637
[21]  254  243  249

$ind
[1] FALSE
```

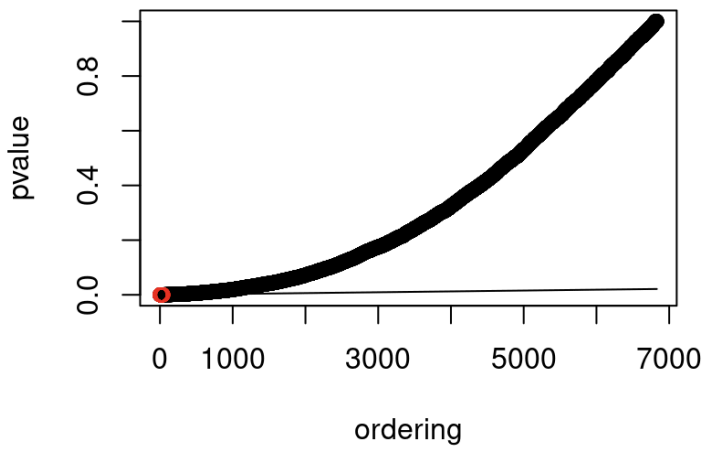
Using quartile value 0.05:

```
> fdr(apply(COL,2,myt),.05,ind=F)
$interesting
[1]  256  257  286 5856 5854  251 5950

$ind
[1] FALSE
```

Melanoma Cancer

Getting the fdr p value for melanoma cancer, and testing on 2, 1, and 0.05

COL1<- (NCI60$data[NCI60$labs=="MELANOMA",])
fdr(apply(COL1,2,myt),.2,ind=F)
$interesting
fdr(apply(COL1,2,myt),.1,ind=F)
fdr(apply(COL1,2,myt),.05,ind=F)



Using quartile value 0.2

```
> melanoma <- fdr(apply(COL1,2,myt),.2,ind=F)
> melanoma
$interesting
 [1] 4347  247  196 4355 2961 4384 4349 5586 1446 4256 4348 4279 4111 5984 2024 4327 4316 4302 4288 4110
[21] 2233 1445 4332  234 5467 6389 4517 4255 4289 2332 4277 6399 5141 4039 3894 4280  287 1743 5320  264
[41] 4088

$ind
[1] FALSE
```

Using quartile value 0.1

```
> fdr(apply(COL1,2,myt),.1,ind=F)
$interesting
 [1] 4347  247  196 4355 2961 4384 4349 5586 1446 4256 4348 4279 4111 5984 2024 4327

$ind
[1] FALSE
```

Using quartile value 0.05

```
> fdr(apply(COL1,2,myt),.05,ind=F)
$interesting
[1] 4347

$ind
[1] FALSE
```

CNS

Getting the fdr p value for cNS cancer, and testing on 2, 1, and 0.05

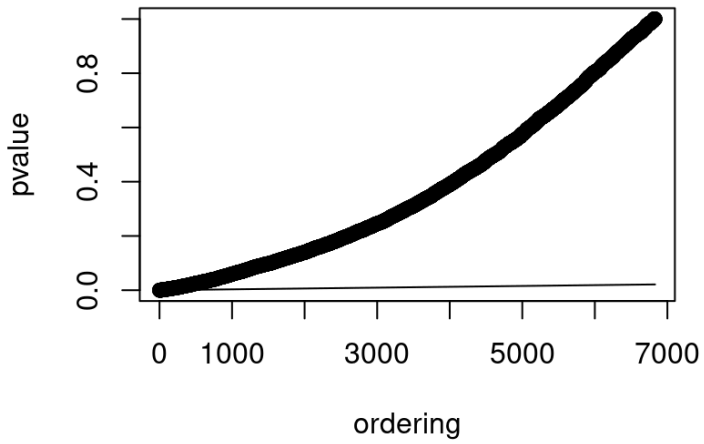
COL2<- (NCI60$data[NCI60$labs=="CNS",])

fdr(apply(COL2,2,myt),.2,ind=F)

$interesting

fdr(apply(COL2,2,myt),.1,ind=F)

fdr(apply(COL2,2,myt),.05,ind=F)



Using quartile 0.2 it is clear that CNS does not print any interesting values.

```
> cns <- fdr(apply(COL2,2,myt),.2,ind=F)
> cns
```

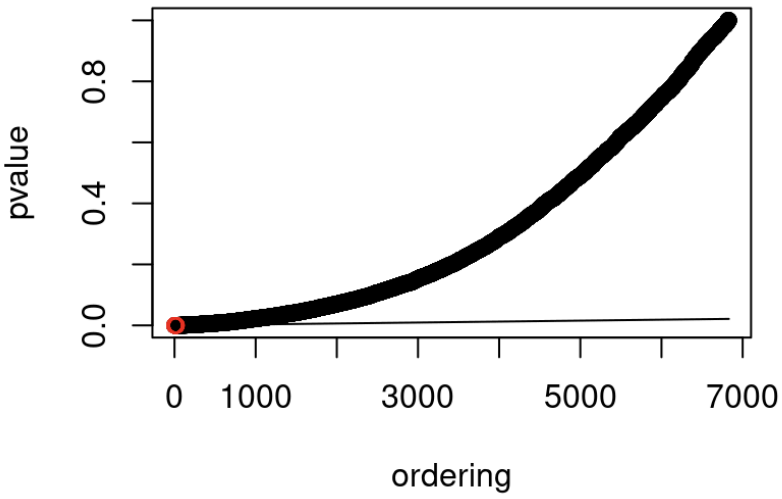
```
> fdr(apply(COL2,2,myt),.1,ind=F)
$q.5
[1] 0.0001014657

$q.95
[1] 0.0004384534

$q.99
[1] 0.0006739304

$q.999
[1] 0.001010725
```

```
> fdr(apply(COL2,2,myt),.05,ind=F)
$q.5
[1] 0.0001014657

$q.95
[1] 0.0004384534

$q.99
[1] 0.0006739304

$q.999
[1] 0.001010725
```

Leukemia

Getting the fdr p value for leukemia cancer, and testing on 2, 1, and 0.05

```
COL3<- (NCI60$data[NCI60$labs=="LEUKEMIA",])
 fdr(apply(COL3,2,myt),.2,ind=F)
 $interesting

 fdr(apply(COL3,2,myt),.1,ind=F)
 fdr(apply(COL3,2,myt),.05,ind=F)
```




```
> fdr(apply(COL3,2,myt),.2,ind=F)
$interesting
 [1] 3933 5872 5868 2170 4244 2080 2320 5878 6016 2068  304 2081 5934 1343 6023 4067 2163 2822 6014 1693
[21] 4243 2215 1915 4554 2079

$ind
[1] FALSE
```

```
> fdr(apply(COL3,2,myt),.1,ind=F)
$q.5
[1] 0.0001014657

$q.95
[1] 0.0004384534

$q.99
[1] 0.0006739304

$q.999
[1] 0.001010725
```

```
> fdr(apply(COL3,2,myt),.05,ind=F)
$q.5
[1] 0.0001014657

$q.95
[1] 0.0004384534

$q.99
[1] 0.0006739304

$q.999
[1] 0.001010725
```

Final Findings

```
# common genes between colon and melanoma genes
commonMELANOMA <- intersect(interesting_colon, interesting_melanoma)

# common genes between colon and CNS genes
commonCNS <- intersect(interesting_colon,interesting_cns)

# common genes between colon and leukemia genes
commonLEUKEMIA <- intersect(interesting_colon,interesting_leukemia)

# common genes between melanoma and CNS
commonMELANOMA_CNS <- intersect(interesting_melanoma,interesting_cns)

# common genes between melanoma and leukemia
commonMELANOMA_LEUKEMIA <-
intersect(interesting_melanoma,interesting_leukemia)

# common genes between CNS and leukemia
commonCNS_LEUKEMIA <- intersect(interesting_cns,interesting_leukemia)

# printing the genes that are common between the corresponding cancers
print(commonMELANOMA)
print(commonCNS)
print(commonLEUKEMIA)
print(commonMELANOMA_CNS)
print(commonMELANOMA_LEUKEMIA)
print(commonCNS_LEUKEMIA)
```

```
> print(commonMELANOMA)
[1]  247 4384
> print(commonCNS)
NULL
> print(commonLEUKEMIA)
integer(0)
> print(commonMELANOMA_CNS)
NULL
> print(commonMELANOMA_LEUKEMIA)
integer(0)
> print(commonCNS_LEUKEMIA)
NULL
```

- commonMELANOMA is the intersection between the interesting colon and melanoma genes. Since it is returning 247 and 4384, we know that there are those two genes in common between the two datasets
- commonCNS is the intersection between colon and CNS genes. Since the output is NULL, it is an indication that there is an absence of data in the interesting genes between the two cancers.
- commonLEUKEMIA is the intersection between the interesting colon and leukemia genes. Since it returns integer(0), we can say that there are NO interesting genes in common between the two cancers
- commonMELANOMA_CNS is the intersection between the interesting melanoma and CNS genes. Since the output is integer(0), we can say that there are NO interesting genes in common between the two cancers
- commonMELANOMA_LEUKEMIA is the intersection between the interesting genes of melanoma and leukemia. Since the output is integer(0), we can say that there are NO interesting genes in common between the two cancers.
- commonCNS_LEUKEMIA is the intersection between the interesting genes of CNS and leukemia. Since the output is NULL, it is an indication that there is an absence of data in the interesting genes between the two cancers.
- Using these, we can observe that the CNS cancer combined with any other cancer does not return any interesting genes in common.

We began with Colon cancer, and continued with Melanoma, CNS, and Leukemia. Using the NCI60 dataset, and the fdr.pck, we used the fdr function. This function returned $interesting values for colon cancer, melanoma, and leukemia and returned $q values for CNS. Relevant output from the fdr function was set to a false discovery rate of 0.2.

Based on the output that was printed when checking which interesting genes have an intersection, there are only two common interesting genes, which were between colon cancer and melanoma. The genes that we found to have an intersection were 247 and 4384. There are no common interesting genes between the following pairs of cancers: Colon and CNS, Colon and Leukemia, Melanoma and CNS, Melanoma and Leukemia, and CNS and Leukemia. So, the only two cancers that share some common interesting genes are colon and melanoma, with the two genes being 247 and 4384. For the other pairs of cancers that we checked, the intersect function returned an empty vector (integer(0) or NULL), which means that there are no common interesting genes between those cancer types based on the provided data and analysis. We can say those pairs of cancers are statistically independent and have no overlap or correlating genes. CNS does not have any interesting genes in the interesting vector that it returns, therefore, when we perform the intersection test on any of the other cancers with CNS, they return null, as the CNS interesting vector itself is null. For the other intersections,

we see that they are values of common gene returns or a clear indication of no common genes present.