Dissert-final-exp

Gabriel Robaina

2024-03-21

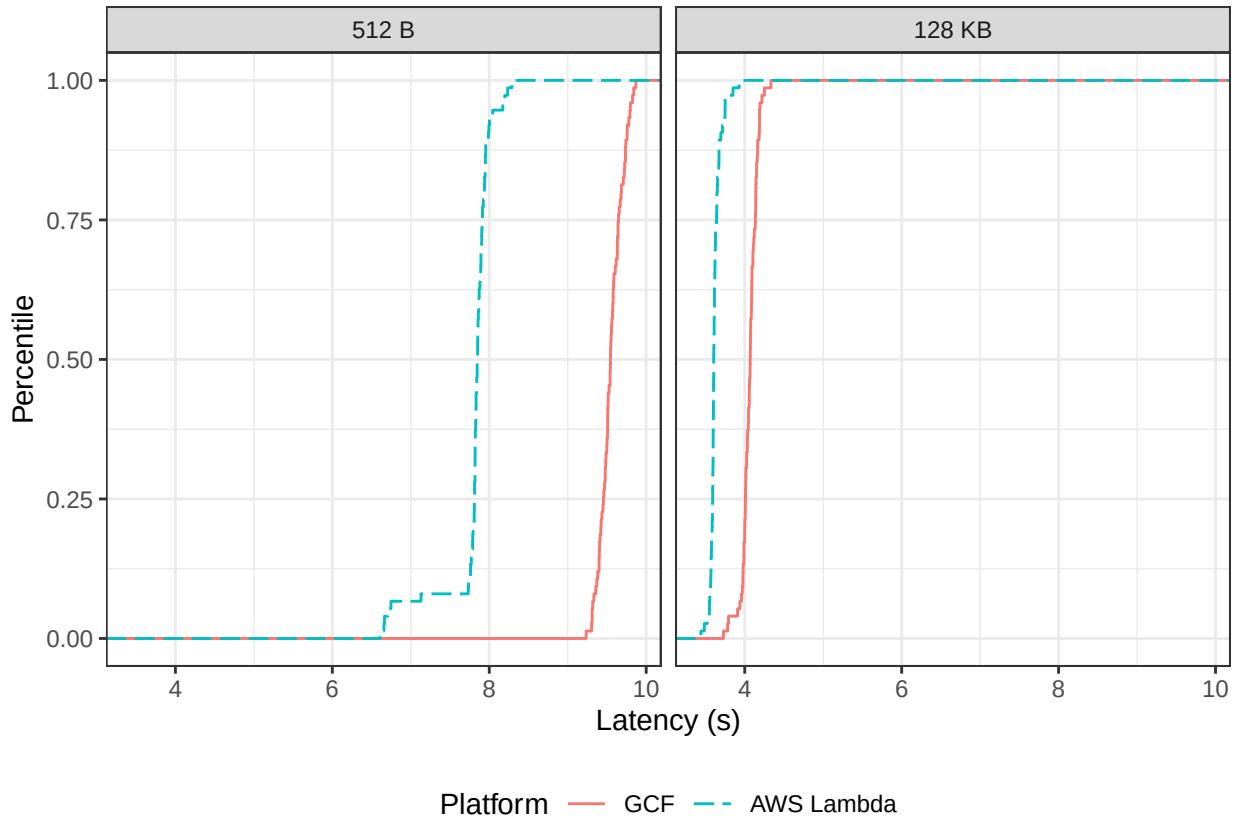
Dissert experiments

Write operations on large files

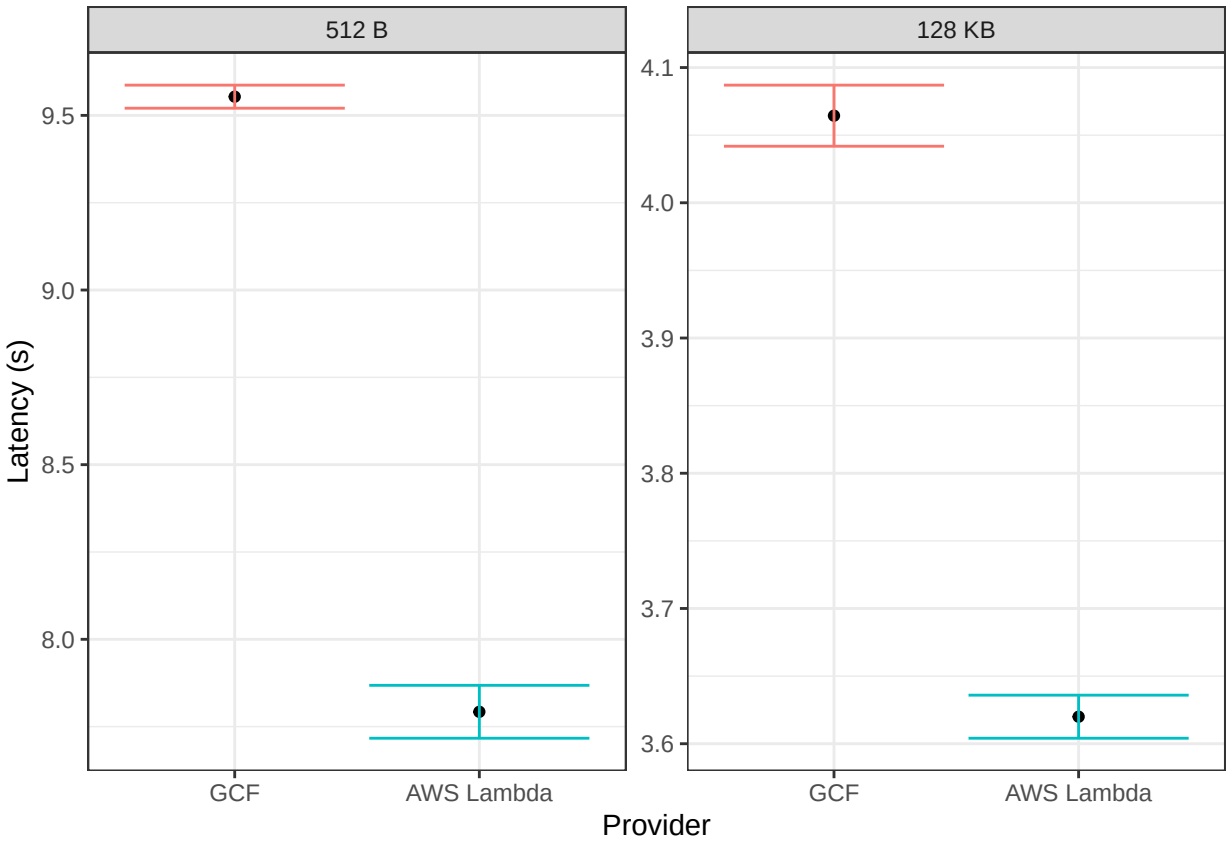
```
## [1] "ECDF of write latency for a 1 GB file"
```

```
## `summarise()` has grouped output by 'system_name'. You can override using the
## `.groups` argument.
```

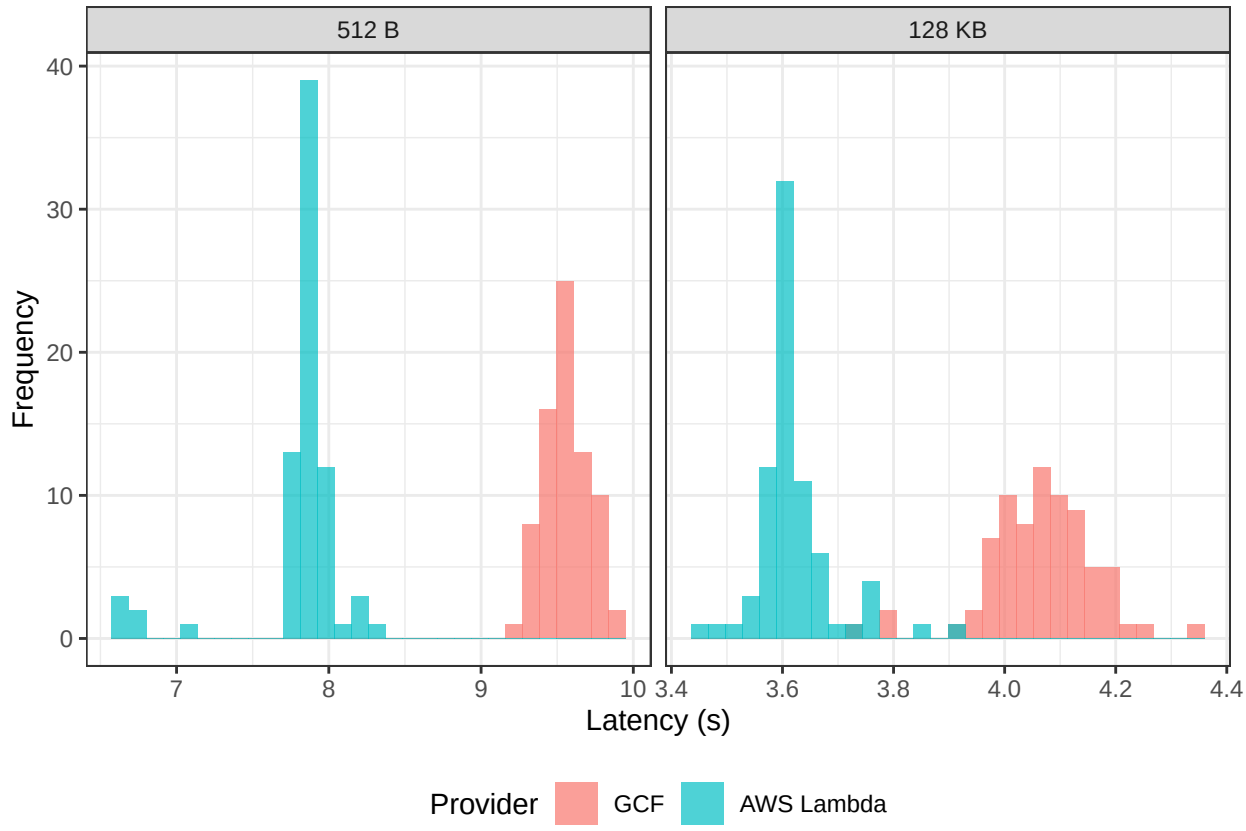
```
## # A tibble: 4 x 3
## # Groups:   system_name [2]
##   system_name io_size_bytes median_latency
##   <chr>             <int64>          <dbl>
## 1 GCF_DD                512           9.54
## 2 GCF_DD             128000           4.07
## 3 LAMBDA_DD             512           7.85
## 4 LAMBDA_DD          128000           3.60
```




```
##    provider io_size    lower    upper     mean
## 1 LAMBDA_DD     512 7.716750 7.868251 7.792501
## 2 LAMBDA_DD  128000 3.604048 3.635907 3.619977
## 3    GCF_DD     512 9.520569 9.586562 9.553566
## 4    GCF_DD  128000 4.041809 4.086982 4.064395
```

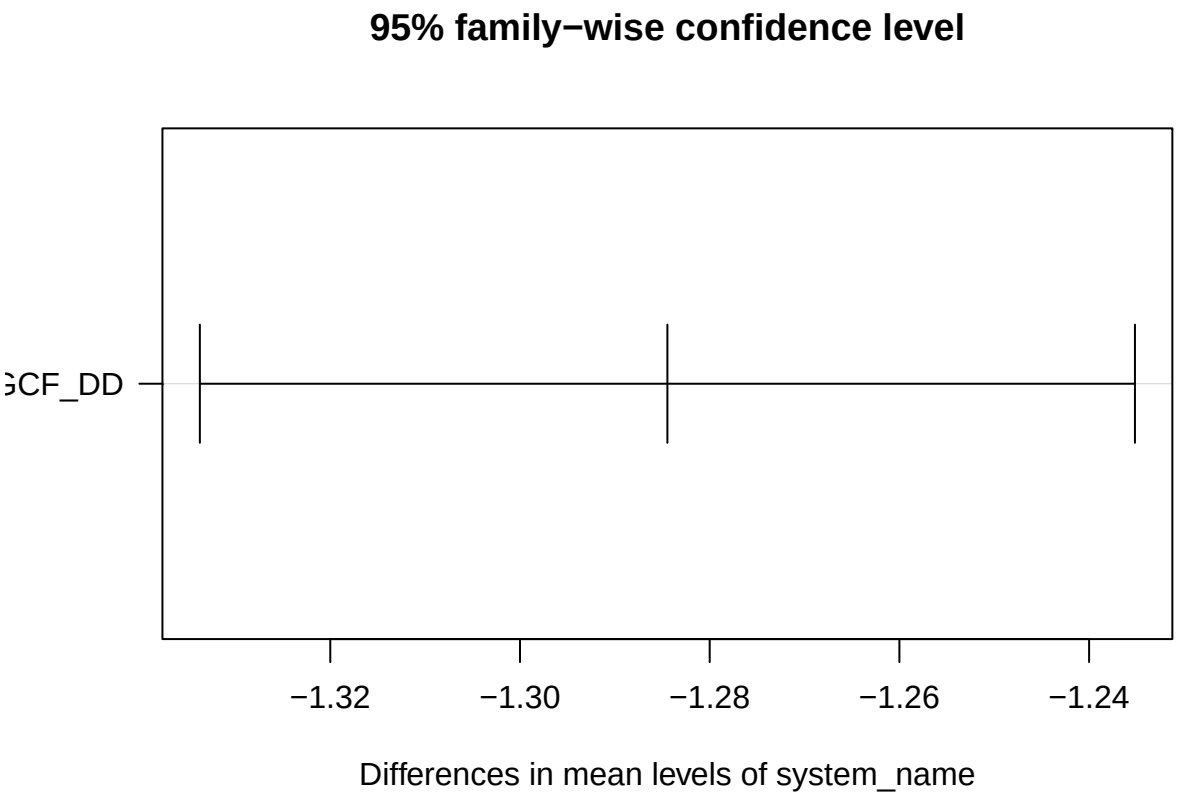


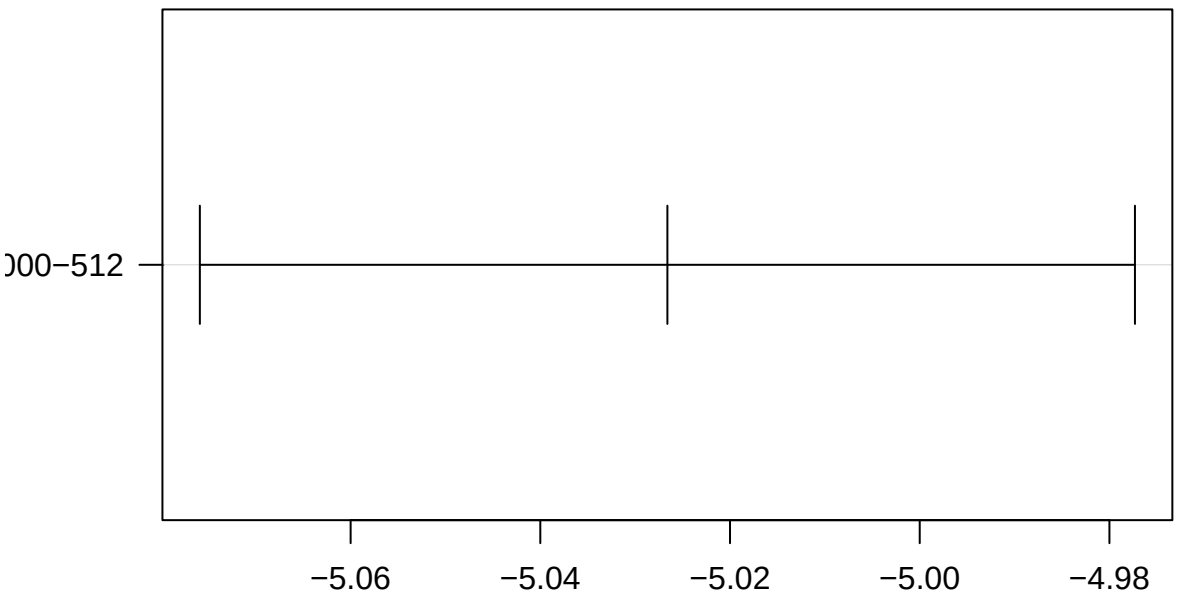

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

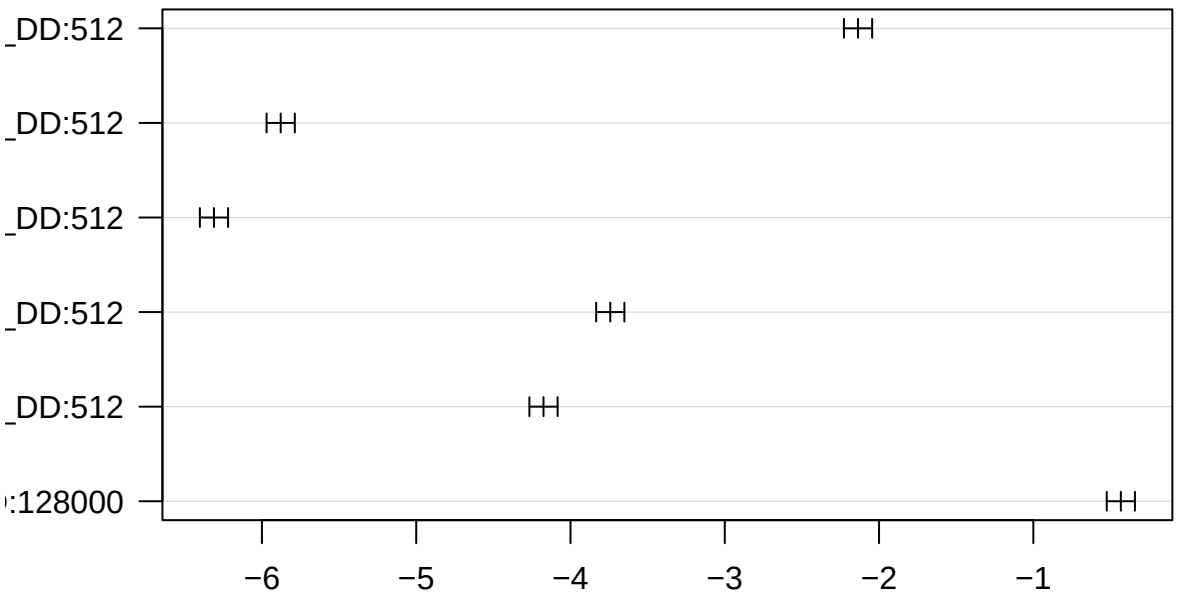



```
## [1] "1 GB file for WRITE operations"
##                                Df Sum Sq Mean Sq F value Pr(>F)
## system_name                     1    247     247    2619 <2e-16 ***
## io_size_bytes_fac               1   3790    3790   40109 <2e-16 ***
## system_name:io_size_bytes_fac   1    109     109    1152 <2e-16 ***
## Residuals                     596     56       0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = latency_seconds ~ system_name * io_size_bytes_fac, data = res)
##
## $system_name
##                        diff       lwr       upr p adj
## LAMBDA_DD-GCF_DD -1.284458 -1.333751 -1.235165     0
##
## $io_size_bytes_fac
##                 diff       lwr      upr p adj
## 128000-512 -5.026602 -5.075895 -4.97731     0
##
## $`system_name:io_size_bytes_fac`
##                                       diff        lwr        upr p adj
## LAMBDA_DD:512-GCF_DD:512       -2.1361541 -2.2275998 -2.0447085     0
## GCF_DD:128000-GCF_DD:512       -5.8782985 -5.9697442 -5.7868529     0
## LAMBDA_DD:128000-GCF_DD:512    -6.3110603 -6.4025059 -6.2196146     0
```

```
## GCF_DD:128000-LAMBDA_DD:512    -3.7421444 -3.8335900 -3.6506988     0
## LAMBDA_DD:128000-LAMBDA_DD:512 -4.1749061 -4.2663518 -4.0834605     0
## LAMBDA_DD:128000-GCF_DD:128000 -0.4327617 -0.5242074 -0.3413161     0
```



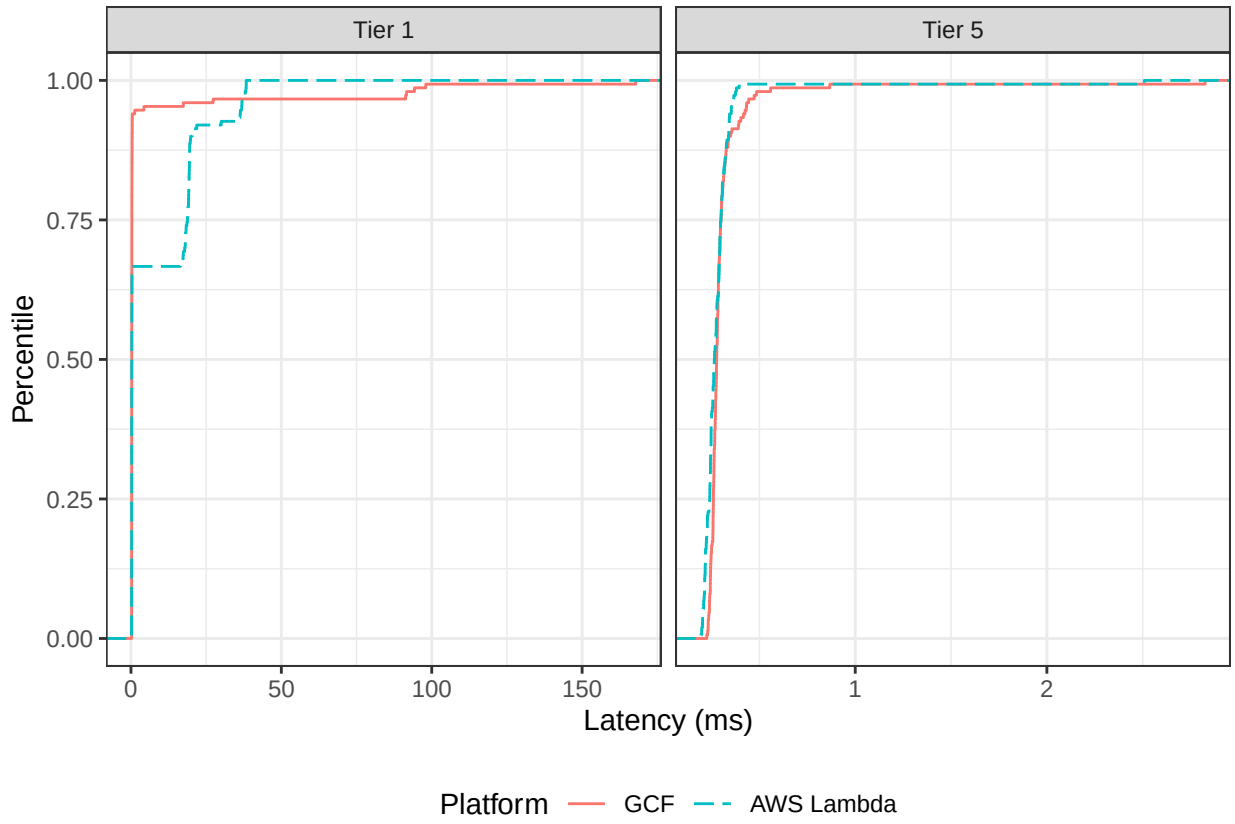

95% family−wise confidence level
000−512
−5.06
−5.04
−5.02
−5.00
−4.98
Differences in mean levels of io_size_bytes_fac



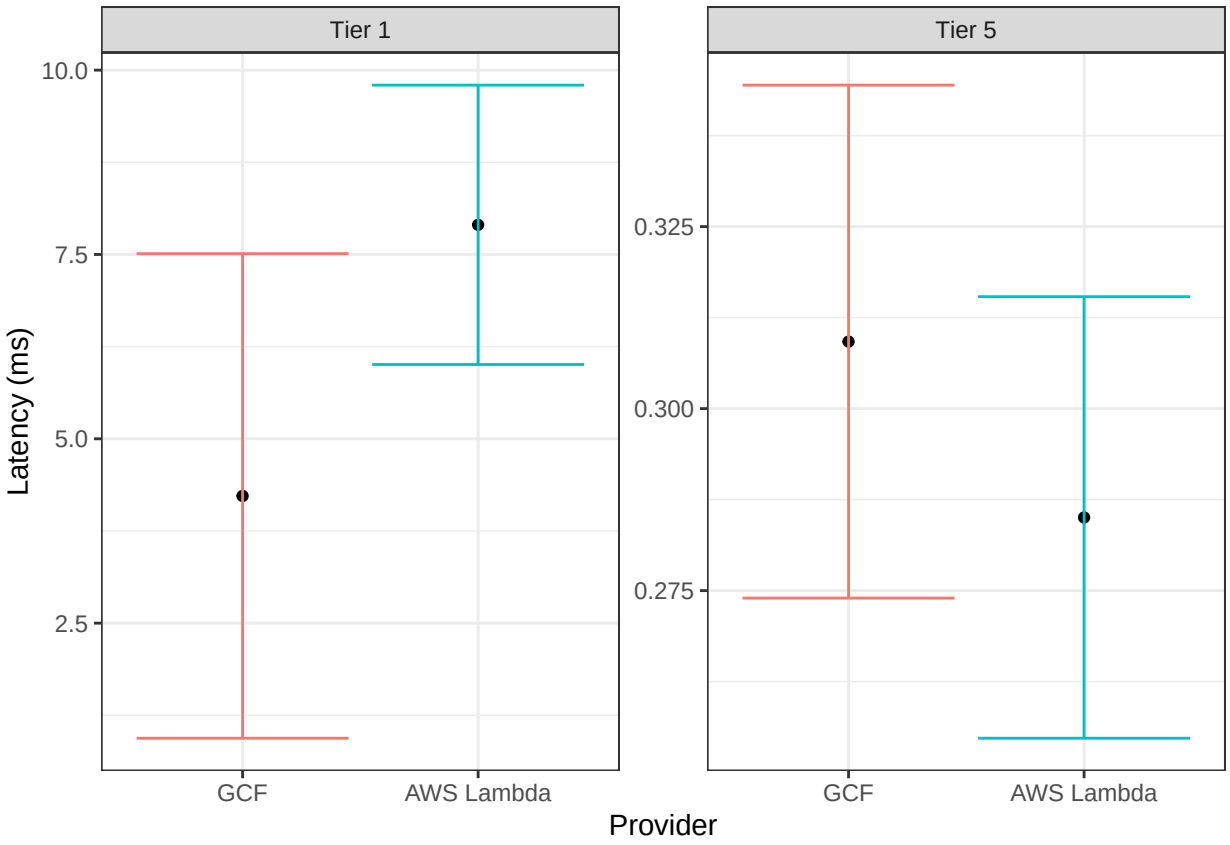

```
## [1] "Sum of squares"
##                                Sum Sq
## system_name                      5.89
## io_size_bytes_fac               90.18
## system_name:io_size_bytes_fac    2.59
## Residuals                        1.34
## [1] "F_crit calculation"
##   factor_df residual_df   F_crit
## 1         1         596 3.857108
```

Write operations on small files

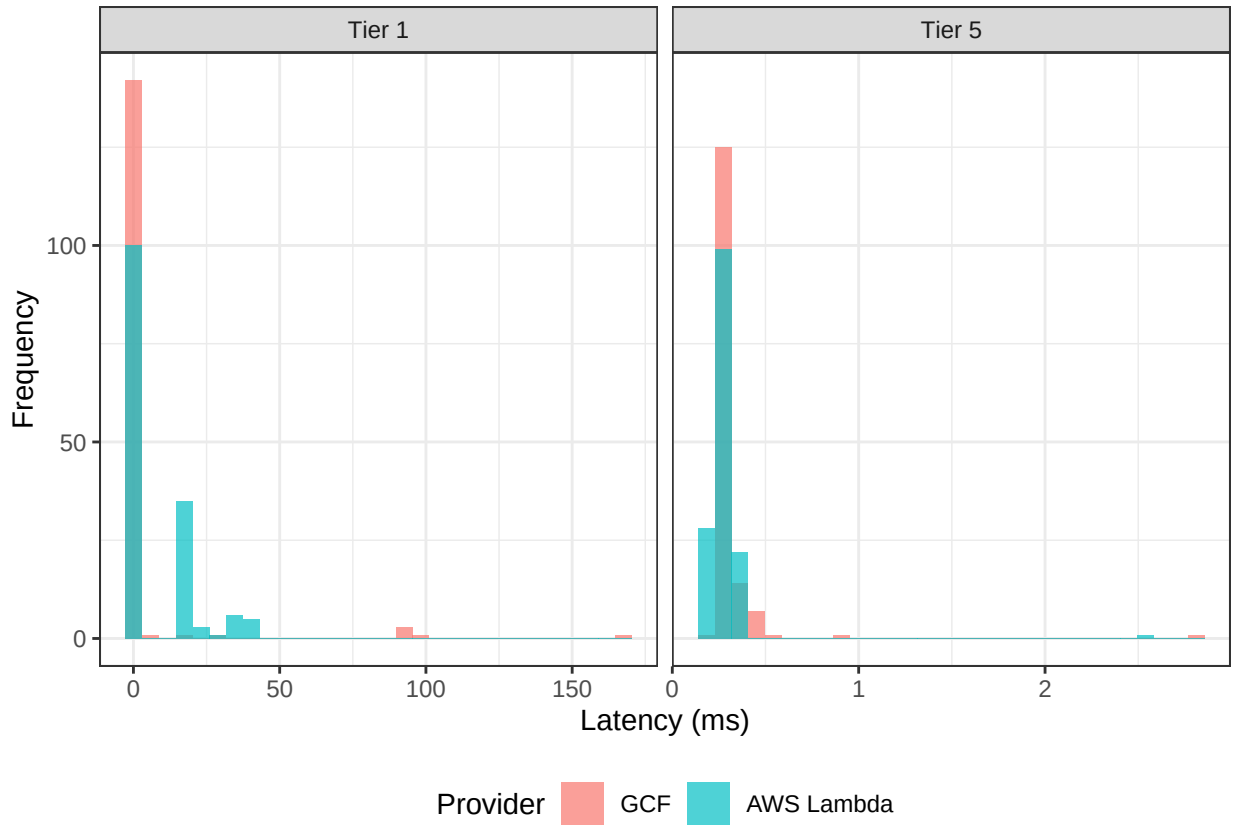
```
## [1] "ECDF of write latency for a 10 KB file and 512 B I/O size"

## `summarise()` has grouped output by 'system_name'. You can override using the
## `.groups` argument.

## # A tibble: 4 x 3
## # Groups:   system_name [2]
##   system_name resource_tier median_latency
##   <chr>       <chr>                 <dbl>
## 1 GCF_DD      TIER_1                0.274
## 2 GCF_DD      TIER_5                0.276
## 3 LAMBDA_DD   TIER_1                0.264
## 4 LAMBDA_DD   TIER_5                0.265
```




```
##    provider resource_tier     lower     upper      mean
## 1 LAMBDA_DD        TIER_1 6.0075037 9.7961710 7.9018374
## 2 LAMBDA_DD        TIER_5 0.2547181 0.3153782 0.2850482
## 3    GCF_DD        TIER_1 0.9390005 7.5109078 4.2249541
## 4    GCF_DD        TIER_5 0.2739734 0.3444359 0.3092047
```

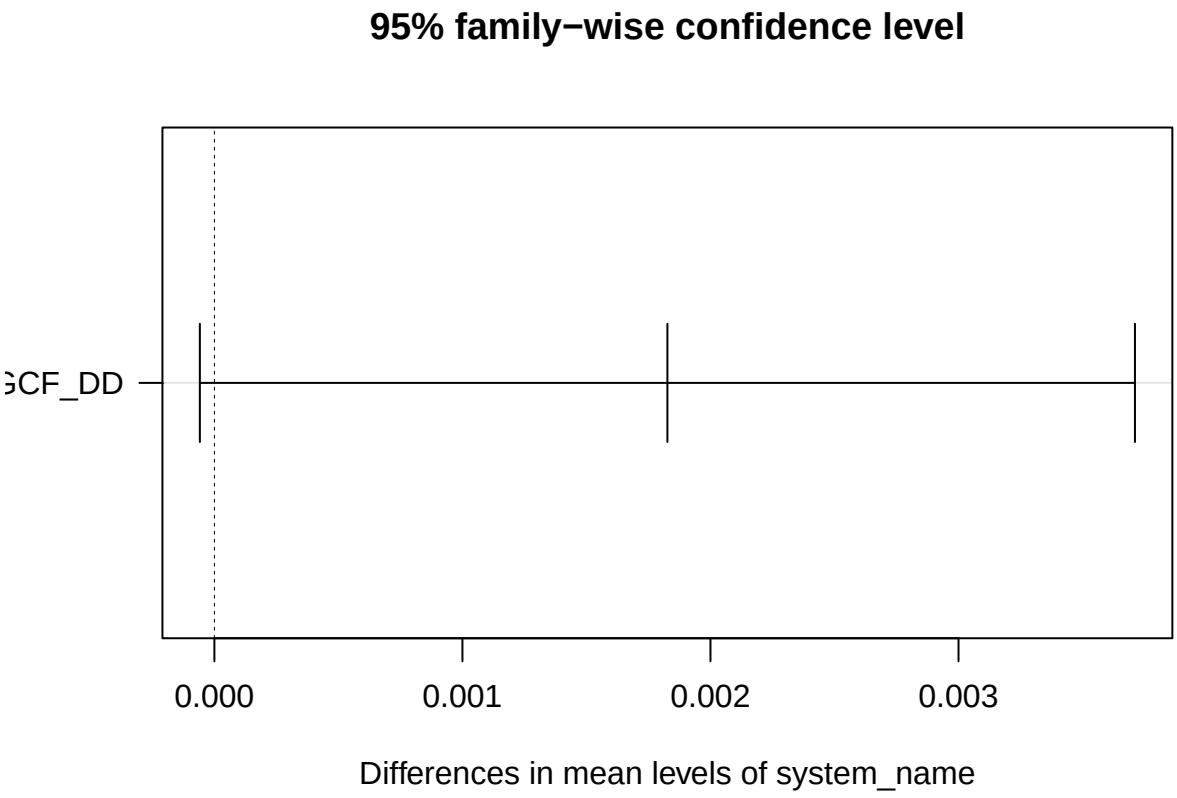


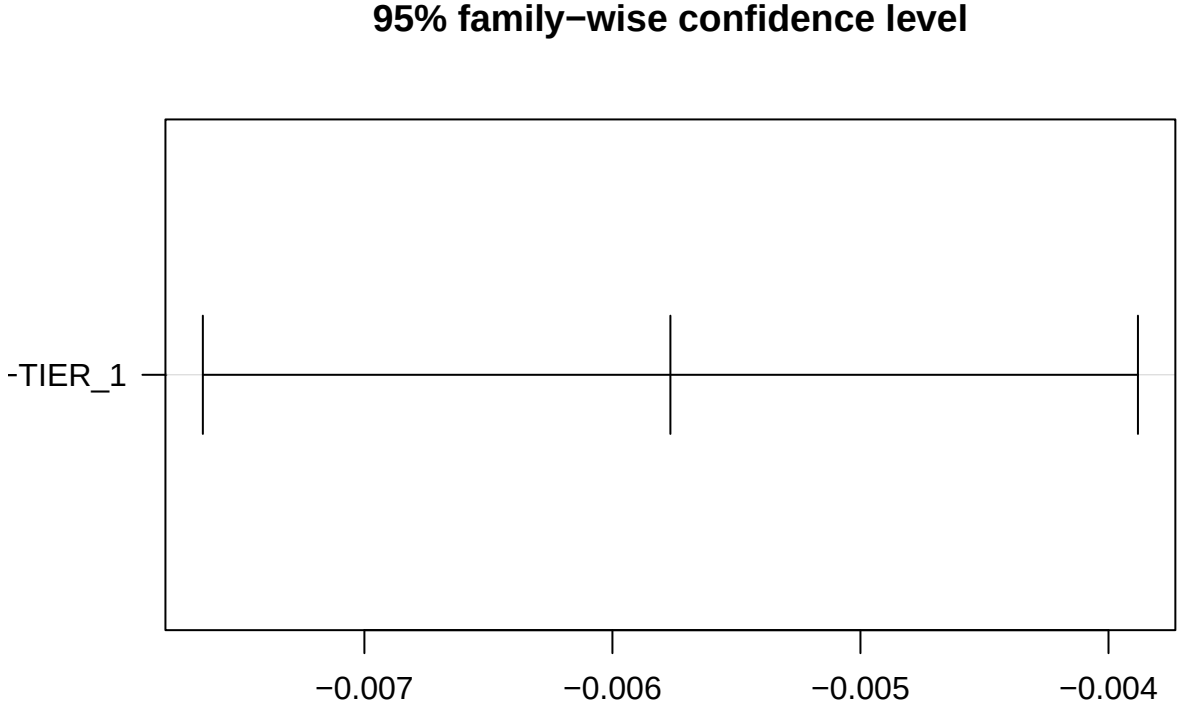

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

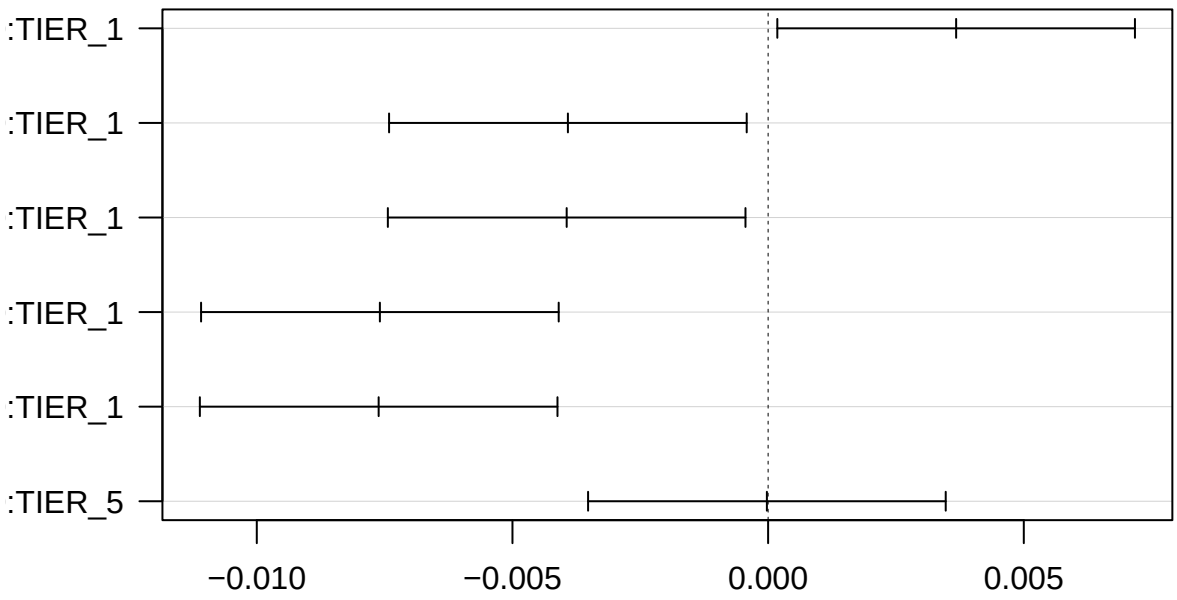



```
## [1] "10 KB file for WRITE operations"
##                            Df  Sum Sq  Mean Sq F value   Pr(>F)
## system_name                 1 0.00050 0.000500   3.621   0.0575 .
## resource_tier               1 0.00499 0.004987  36.093 3.27e-09 ***
## system_name:resource_tier   1 0.00051 0.000514   3.717   0.0543 .
## Residuals                 596 0.08236 0.000138
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = latency_seconds ~ system_name * resource_tier, data = res)
##
## $system_name
##                          diff           lwr         upr     p adj
## LAMBDA_DD-GCF_DD 0.001826363 -5.864643e-05 0.003711373 0.0575418
##
## $resource_tier
##                       diff          lwr         upr p adj
## TIER_5-TIER_1 -0.005766269 -0.007651279 -0.00388126     0
##
## $`system_name:resource_tier`
##                                            diff           lwr           upr
## LAMBDA_DD:TIER_1-GCF_DD:TIER_1     3.676883e-03  0.0001799041  0.0071738623
## GCF_DD:TIER_5-GCF_DD:TIER_1       -3.915749e-03 -0.0074127285 -0.0004187704
## LAMBDA_DD:TIER_5-GCF_DD:TIER_1    -3.939906e-03 -0.0074368851 -0.0004429269
```

```
## GCF_DD:TIER_5-LAMBDA_DD:TIER_1    -7.592633e-03 -0.0110896118 -0.0040956536
## LAMBDA_DD:TIER_5-LAMBDA_DD:TIER_1 -7.616789e-03 -0.0111137683 -0.0041198101
## LAMBDA_DD:TIER_5-GCF_DD:TIER_5    -2.415652e-05 -0.0035211356  0.0034728226
##                                       p adj
## LAMBDA_DD:TIER_1-GCF_DD:TIER_1    0.0349446
## GCF_DD:TIER_5-GCF_DD:TIER_1       0.0210927
## LAMBDA_DD:TIER_5-GCF_DD:TIER_1    0.0200063
## GCF_DD:TIER_5-LAMBDA_DD:TIER_1    0.0000002
## LAMBDA_DD:TIER_5-LAMBDA_DD:TIER_1 0.0000002
## LAMBDA_DD:TIER_5-GCF_DD:TIER_5    0.9999980
```




95% family−wise confidence level
-TIER_1
−0.007
−0.006
−0.005
−0.004
Differences in mean levels of resource_tier



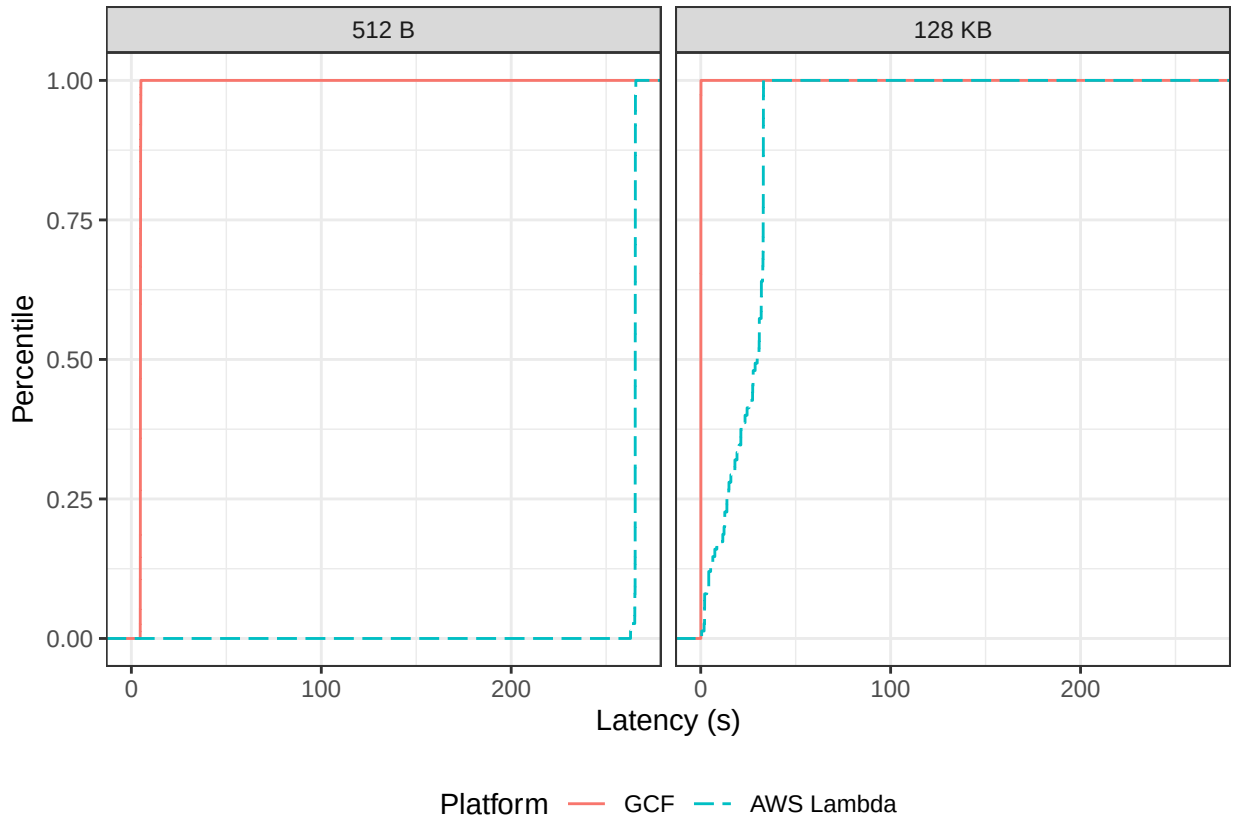

```
## [1] "Sum of squares"
##                            Sum Sq
## system_name                  0.57
## resource_tier                5.64
## system_name:resource_tier    0.58
## Residuals                   93.21
## [1] "F_crit calculation"
##   factor_df residual_df   F_crit
## 1         1         596 3.857108
```

Read operations on large files

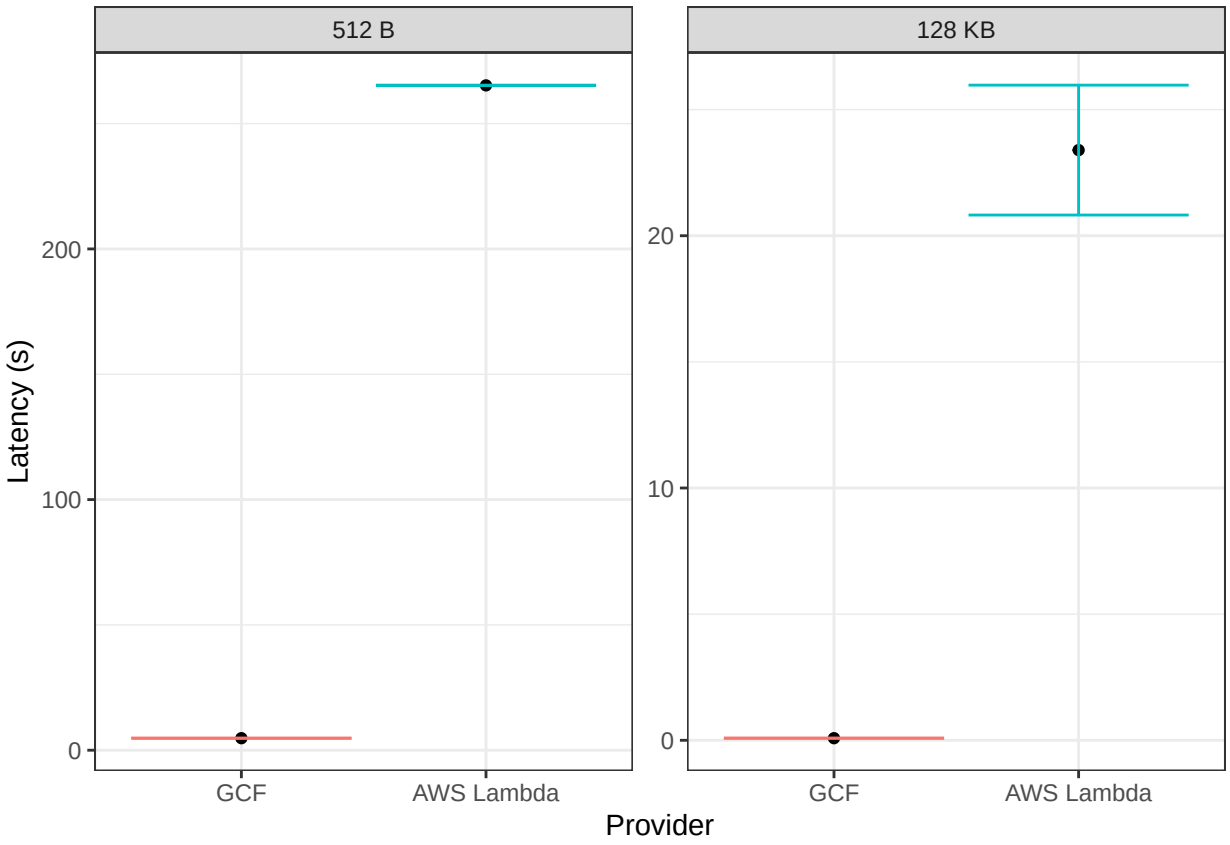
```
## [1] "ECDF of read latency for a 1 GB file"
```

```
## ‘summarise()‘ has grouped output by ’system_name’. You can override using the
## ‘.groups‘ argument.
```

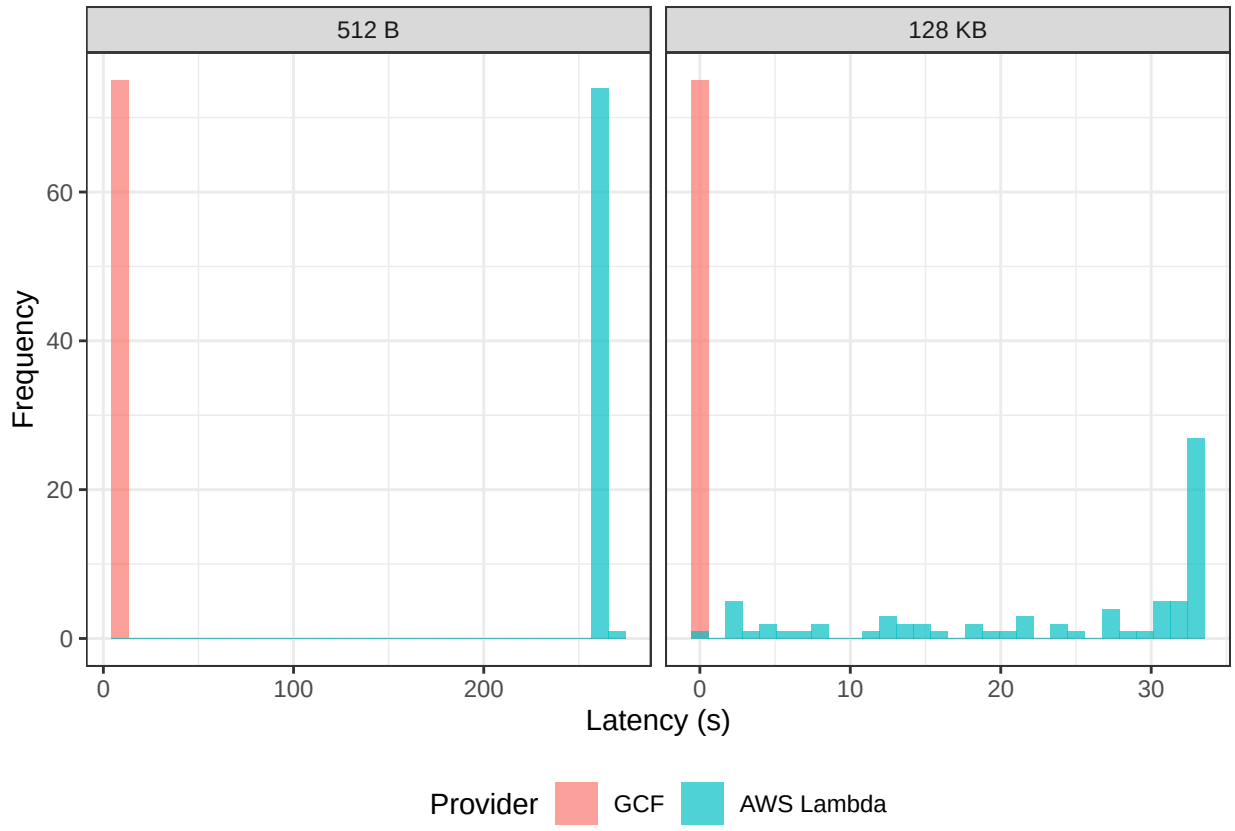
```
## # A tibble: 4 x 3
## # Groups:   system_name [2]
##   system_name io_size_bytes median_latency
##   <chr>             <int64>          <dbl>
## 1 GCF_DD                512         4.79
## 2 GCF_DD             128000         0.0815
## 3 LAMBDA_DD             512       265.
## 4 LAMBDA_DD          128000        29.8
```




```
##    provider io_size        lower        upper         mean
## 1 LAMBDA_DD     512 265.16117135 265.34426865 265.25272000
## 2 LAMBDA_DD  128000  20.82039683  25.96912930  23.39476307
## 3    GCF_DD     512   4.77805686   4.81451941   4.79628813
## 4    GCF_DD  128000   0.08060797   0.08479676   0.08270236
```

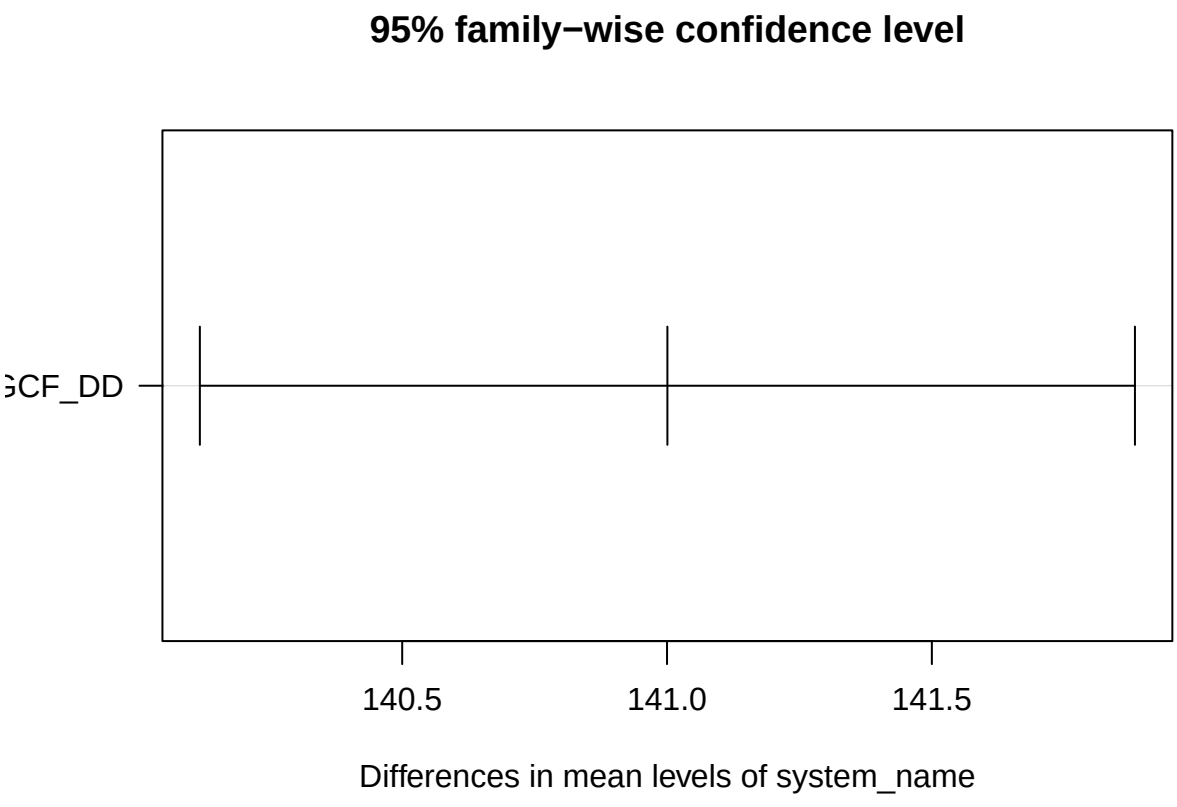


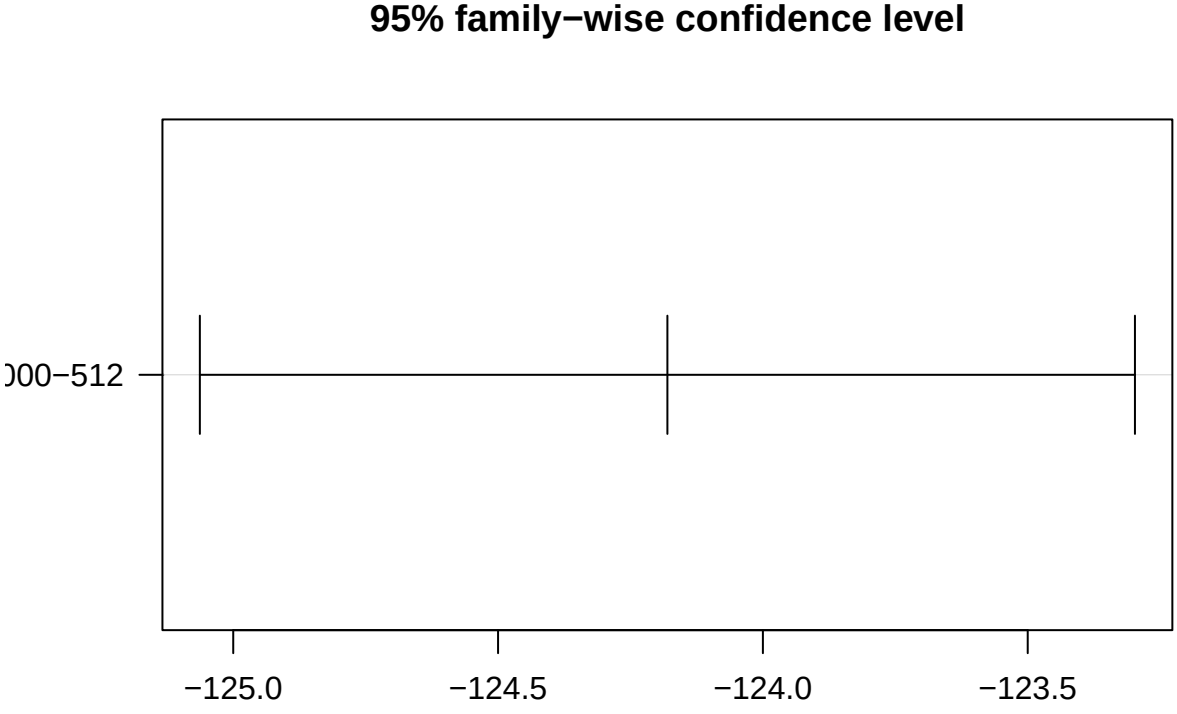

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

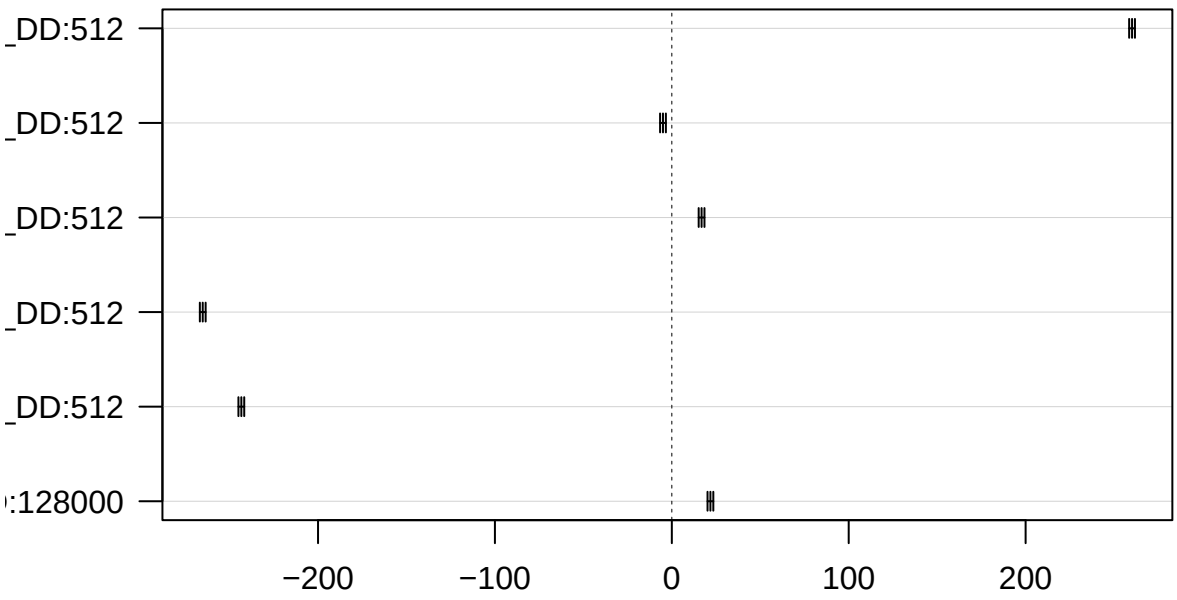



```
## [1] "1 GB file for READ operations"
##                                Df  Sum Sq Mean Sq F value Pr(>F)
## system_name                     1 2982180 2982180   98403 <2e-16 ***
## io_size_bytes_fac               1 2313112 2313112   76326 <2e-16 ***
## system_name:io_size_bytes_fac   1 2130624 2130624   70304 <2e-16 ***
## Residuals                     596   18062      30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = latency_seconds ~ system_name * io_size_bytes_fac, data = res)
##
## $system_name
##                       diff      lwr      upr p adj
## LAMBDA_DD-GCF_DD 141.0007 140.1179 141.8835     0
##
## $io_size_bytes_fac
##                  diff       lwr       upr p adj
## 128000-512 -124.1803 -125.0631 -123.2975     0
##
## $`system_name:io_size_bytes_fac`
##                                         diff         lwr         upr p adj
## LAMBDA_DD:512-GCF_DD:512          260.181915  258.544244  261.819587     0
## GCF_DD:128000-GCF_DD:512           -4.999086   -6.636758   -3.361414     0
## LAMBDA_DD:128000-GCF_DD:512        16.820425   15.182754   18.458097     0
```

```
## GCF_DD:128000-LAMBDA_DD:512    -265.181001 -266.818673 -263.543330     0
## LAMBDA_DD:128000-LAMBDA_DD:512 -243.361490 -244.999162 -241.723819     0
## LAMBDA_DD:128000-GCF_DD:128000   21.819511   20.181839   23.457183     0
```




95% family−wise confidence level
000−512
−125.0
−124.5
−124.0
−123.5
Differences in mean levels of io_size_bytes_fac



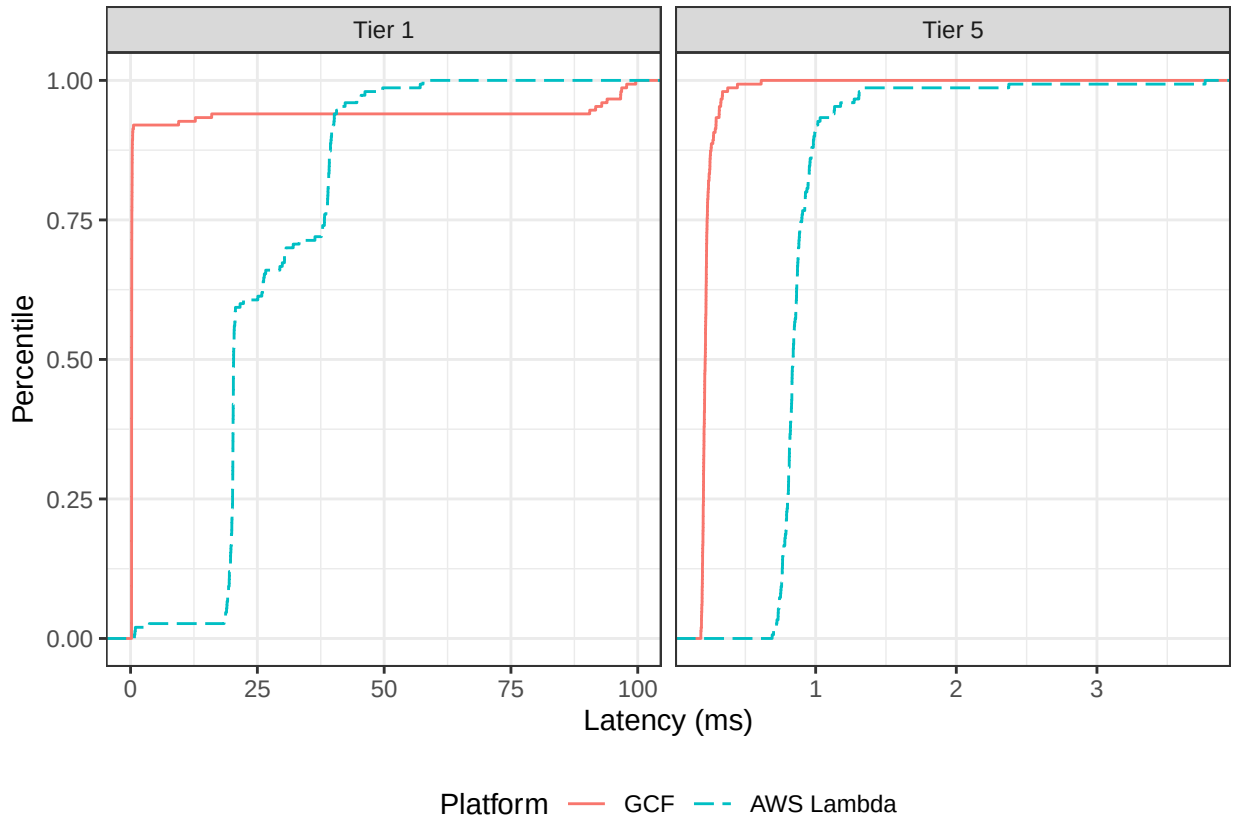

```
## [1] "Sum of squares"
##                                Sum Sq
## system_name                     40.06
## io_size_bytes_fac               31.07
## system_name:io_size_bytes_fac   28.62
## Residuals                        0.24
## [1] "F_crit calculation"
##   factor_df residual_df   F_crit
## 1         1         596 3.857108
```

Read operations on small files

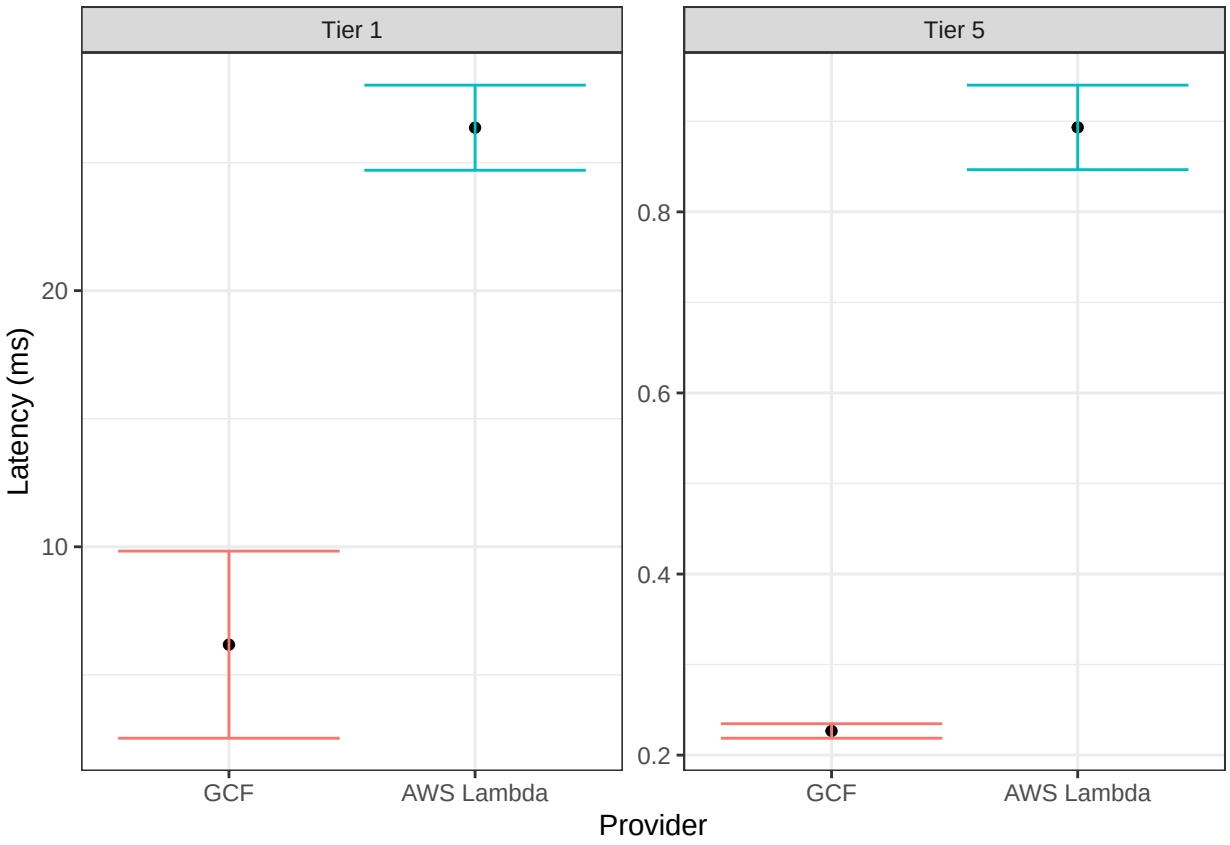
```
## [1] "ECDF of read latency for a 10 KB file and 512 B I/O size"
```

```
## `summarise()` has grouped output by 'system_name'. You can override using the
## `.groups` argument.
```

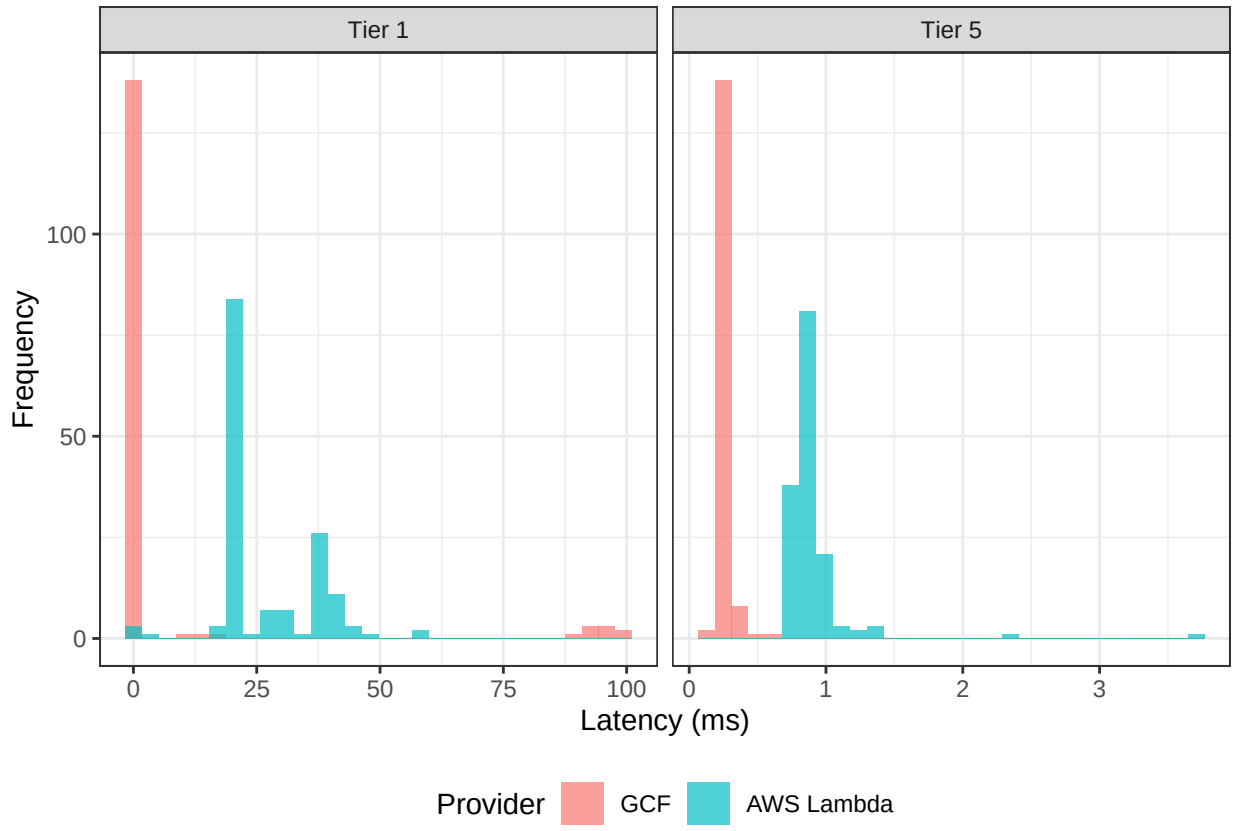
```
## # A tibble: 4 x 3
## # Groups:   system_name [2]
##   system_name resource_tier median_latency
##   <chr>       <chr>                 <dbl>
## 1 GCF_DD      TIER_1                0.211
## 2 GCF_DD      TIER_5                0.214
## 3 LAMBDA_DD   TIER_1                20.3
## 4 LAMBDA_DD   TIER_5                0.839
```




```
##    provider resource_tier      lower      upper       mean
## 1 LAMBDA_DD        TIER_1 24.7018439 28.0245315 26.3631877
## 2 LAMBDA_DD        TIER_5  0.8466467  0.9399647  0.8933057
## 3    GCF_DD        TIER_1  2.5244415  9.8312064  6.1778239
## 4    GCF_DD        TIER_5  0.2186401  0.2346863  0.2266632
```

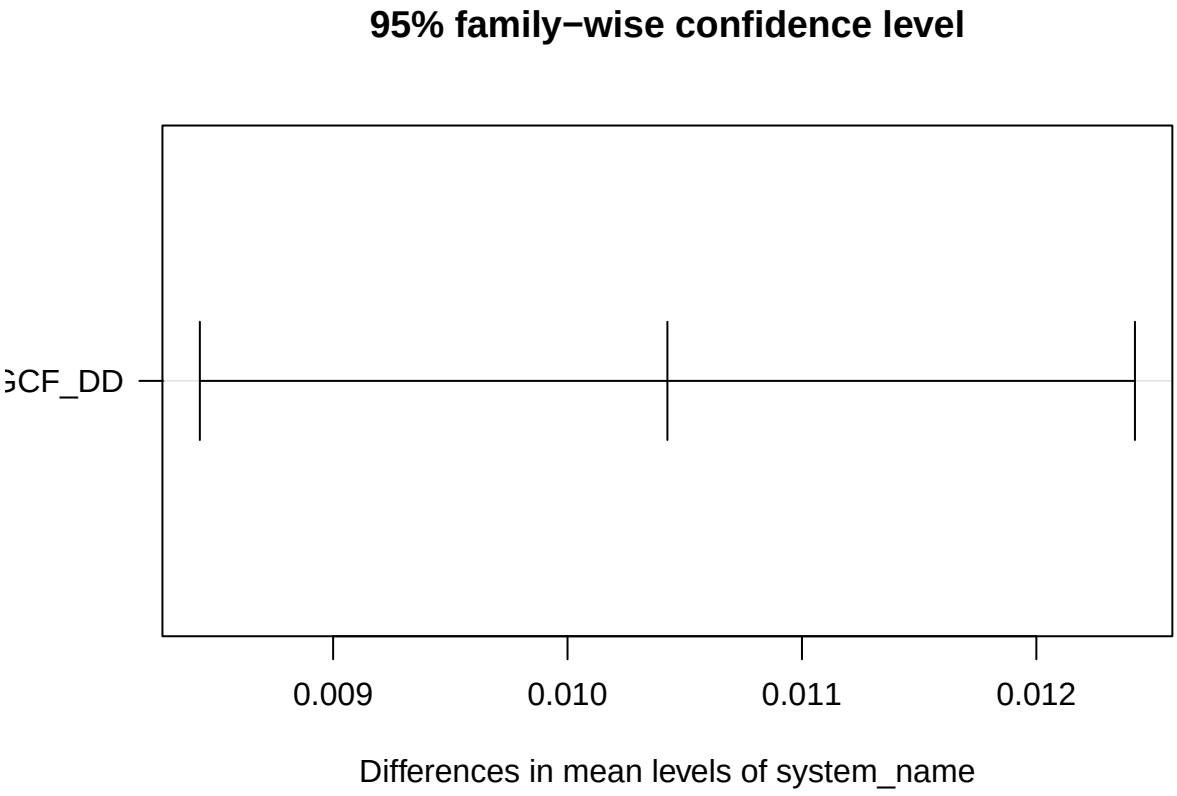


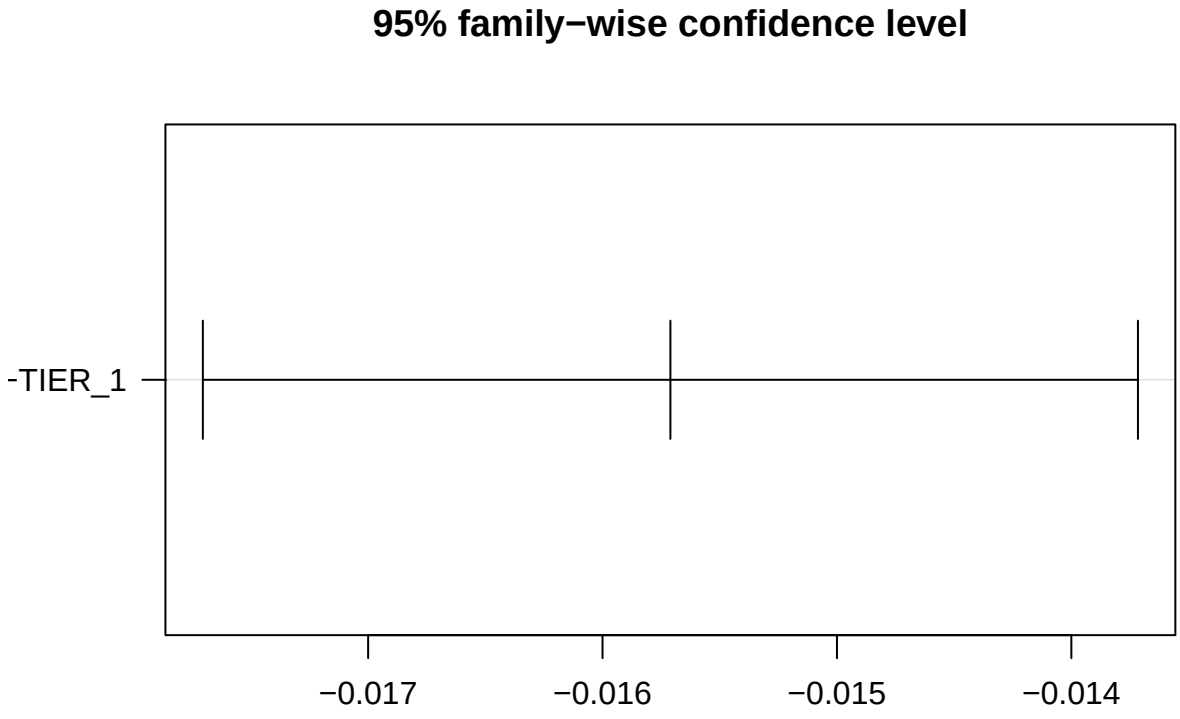

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

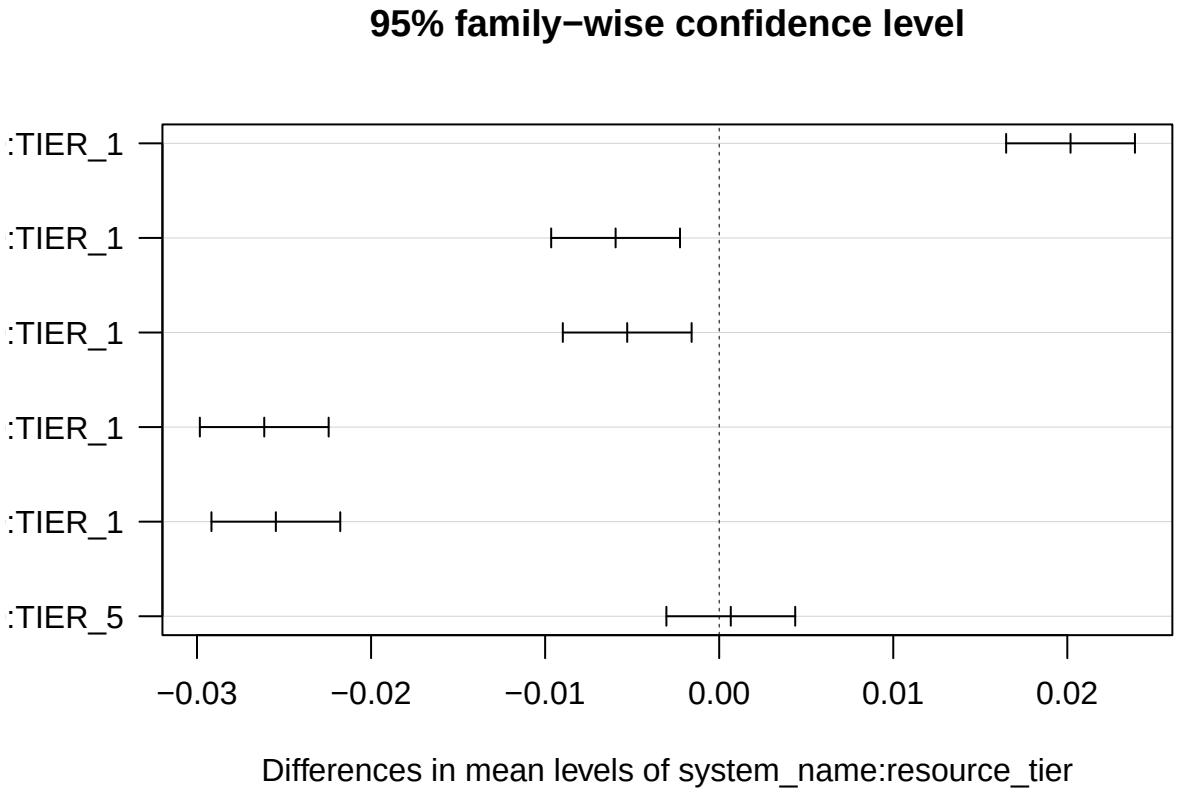



```
## [1] "10 KB file for READ operations"
##                            Df  Sum Sq Mean Sq F value Pr(>F)
## system_name                 1 0.01631 0.01631  105.39 <2e-16 ***
## resource_tier               1 0.03702 0.03702  239.30 <2e-16 ***
## system_name:resource_tier   1 0.01429 0.01429   92.34 <2e-16 ***
## Residuals                 596 0.09221 0.00015
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = latency_seconds ~ system_name * resource_tier, data = res)
##
## $system_name
##                       diff        lwr        upr p adj
## LAMBDA_DD-GCF_DD 0.010426 0.00843142 0.01242059     0
##
## $resource_tier
##                       diff        lwr         upr p adj
## TIER_5-TIER_1 -0.01571052 -0.0177051 -0.01371594     0
##
## $`system_name:resource_tier`
##                                             diff          lwr          upr
## LAMBDA_DD:TIER_1-GCF_DD:TIER_1     0.0201853638  0.016485109  0.023885619
## GCF_DD:TIER_5-GCF_DD:TIER_1       -0.0059511607 -0.009651416 -0.002250906
## LAMBDA_DD:TIER_5-GCF_DD:TIER_1    -0.0052845182 -0.008984773 -0.001584263
```

```
## GCF_DD:TIER_5-LAMBDA_DD:TIER_1    -0.0261365245 -0.029836780 -0.022436269
## LAMBDA_DD:TIER_5-LAMBDA_DD:TIER_1 -0.0254698820 -0.029170137 -0.021769627
## LAMBDA_DD:TIER_5-GCF_DD:TIER_5     0.0006666425 -0.003033613  0.004366898
##                                       p adj
## LAMBDA_DD:TIER_1-GCF_DD:TIER_1    0.0000000
## GCF_DD:TIER_5-GCF_DD:TIER_1       0.0002287
## LAMBDA_DD:TIER_5-GCF_DD:TIER_1    0.0014521
## GCF_DD:TIER_5-LAMBDA_DD:TIER_1    0.0000000
## LAMBDA_DD:TIER_5-LAMBDA_DD:TIER_1 0.0000000
## LAMBDA_DD:TIER_5-GCF_DD:TIER_5    0.9668203
```




95% family−wise confidence level
-TIER_1
−0.017
−0.016
−0.015
−0.014
Differences in mean levels of resource_tier



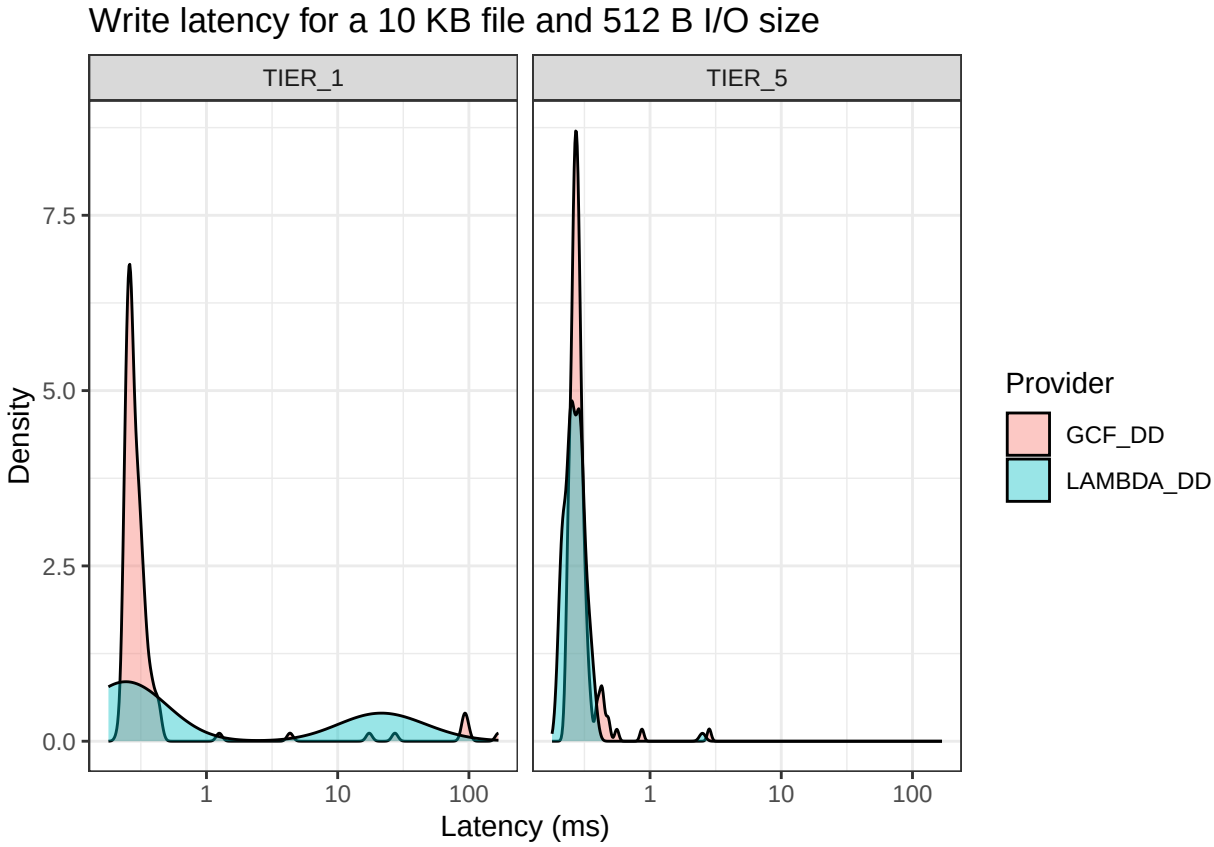

```
## [1] "Sum of squares"
##                            Sum Sq
## system_name                 10.20
## resource_tier               23.16
## system_name:resource_tier    8.94
## Residuals                   57.69
## [1] "F_crit calculation"
##   factor_df residual_df   F_crit
## 1         1         596 3.857108
```

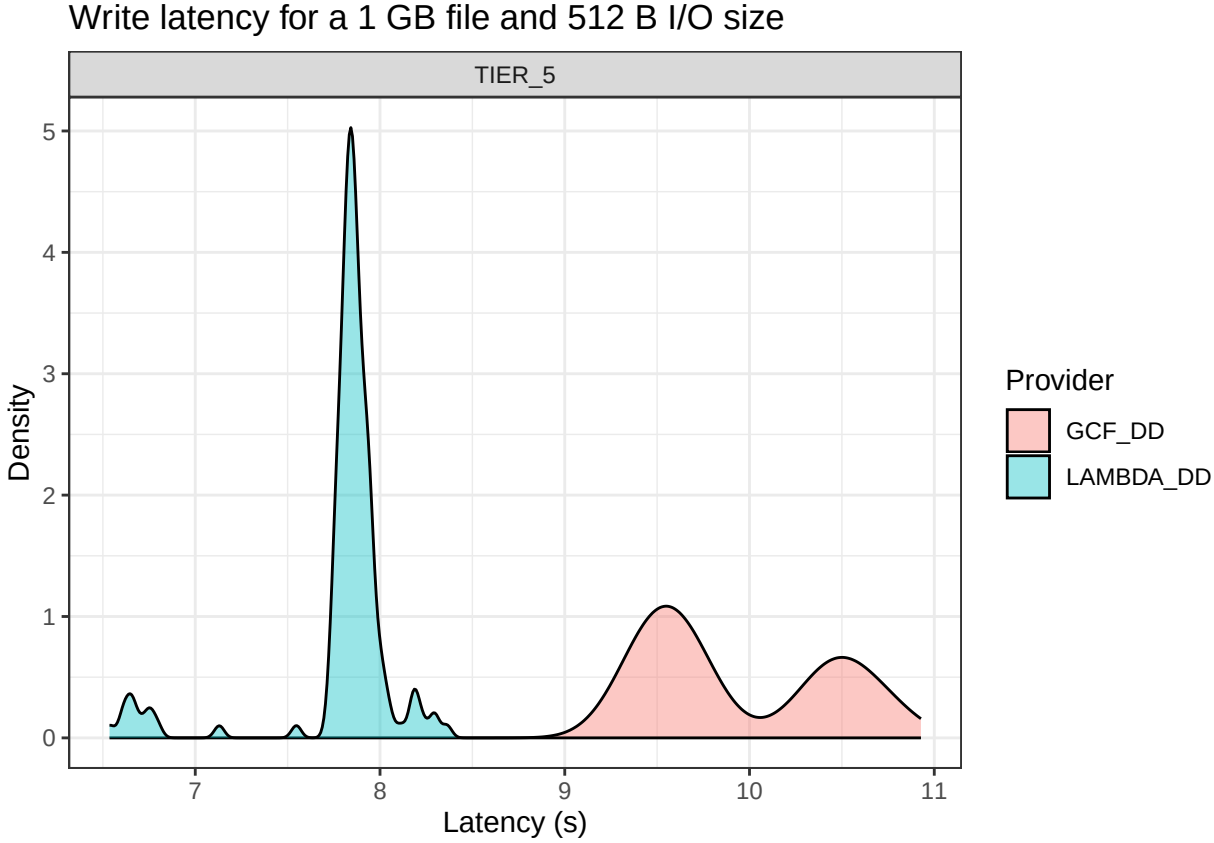
Other plots

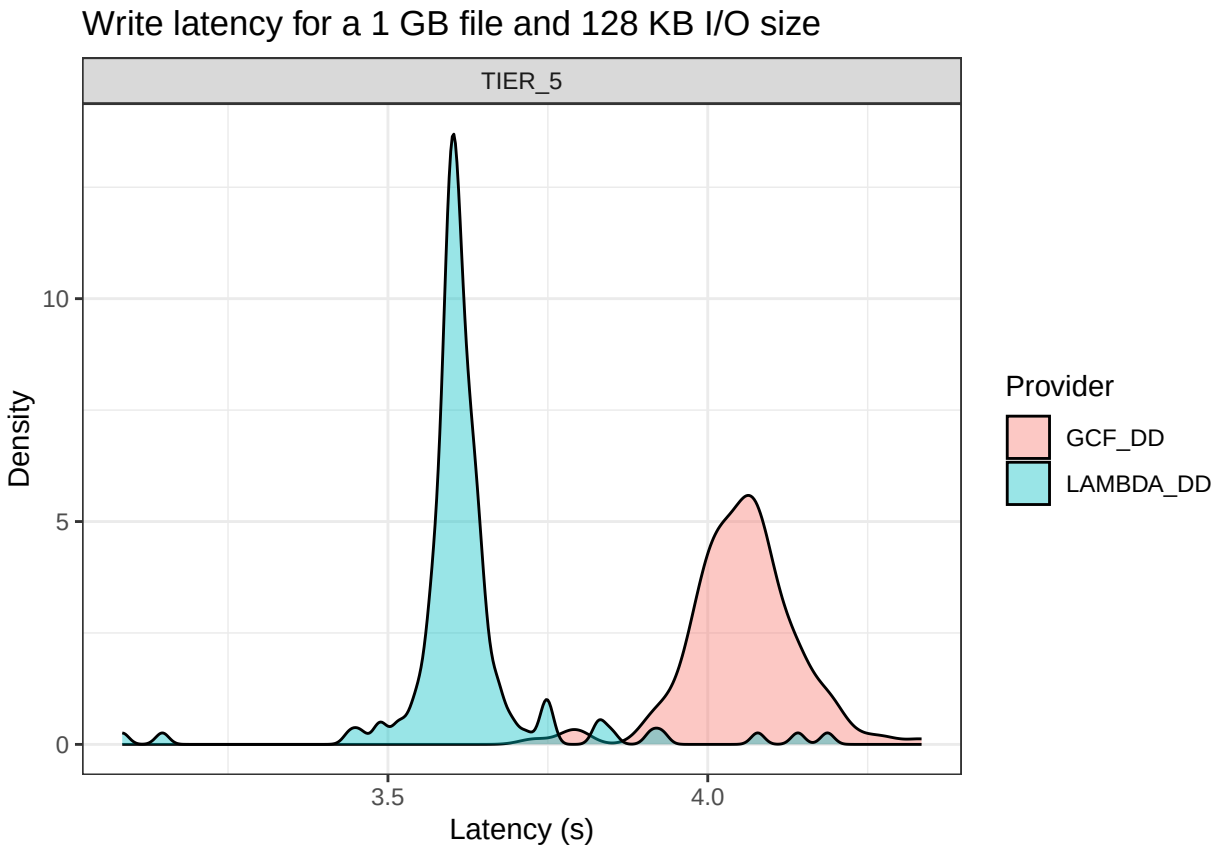
The following are density plots for experiments ran between 02/25/2024 and 03/20/2024.

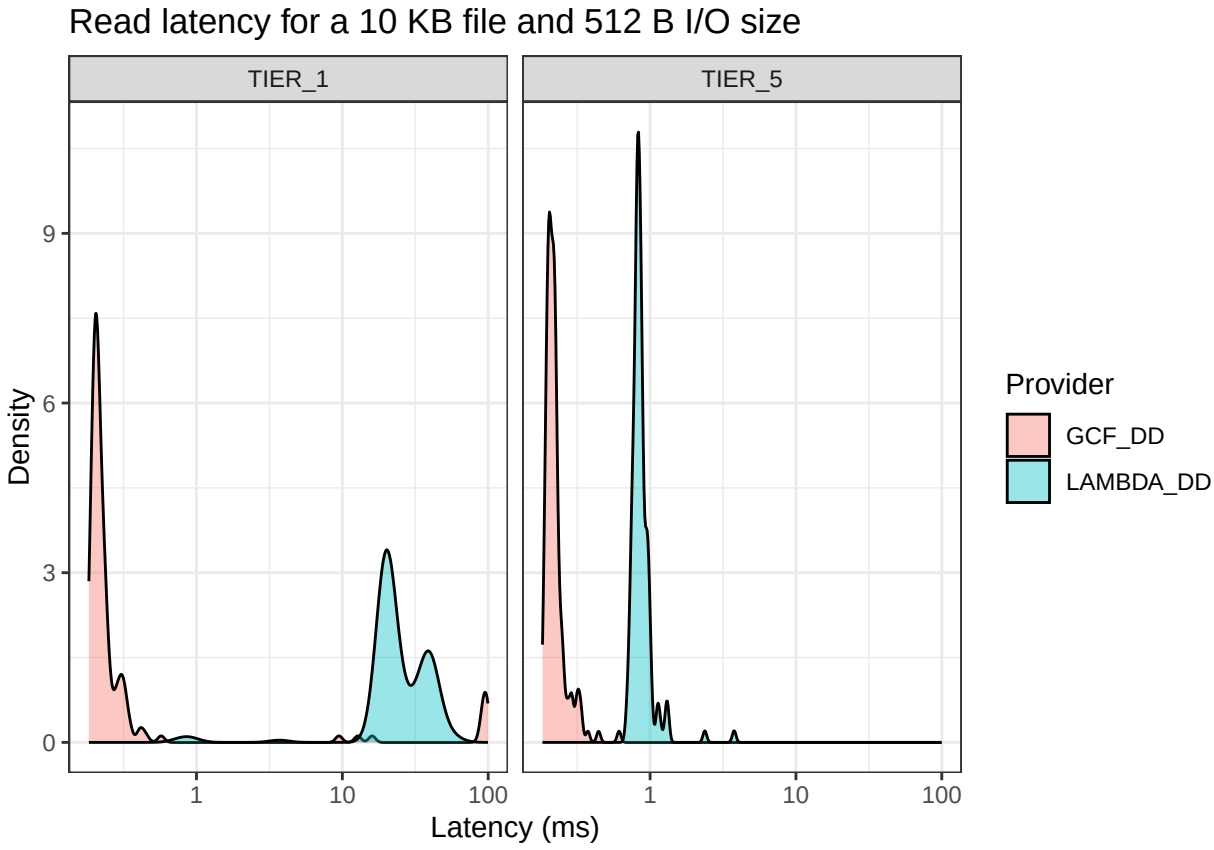
On the first density plot, it is limited between 0 and 1. Its hiding a mode for Lambda_DD on Tier 1 that happens after 25 and 50 ms.

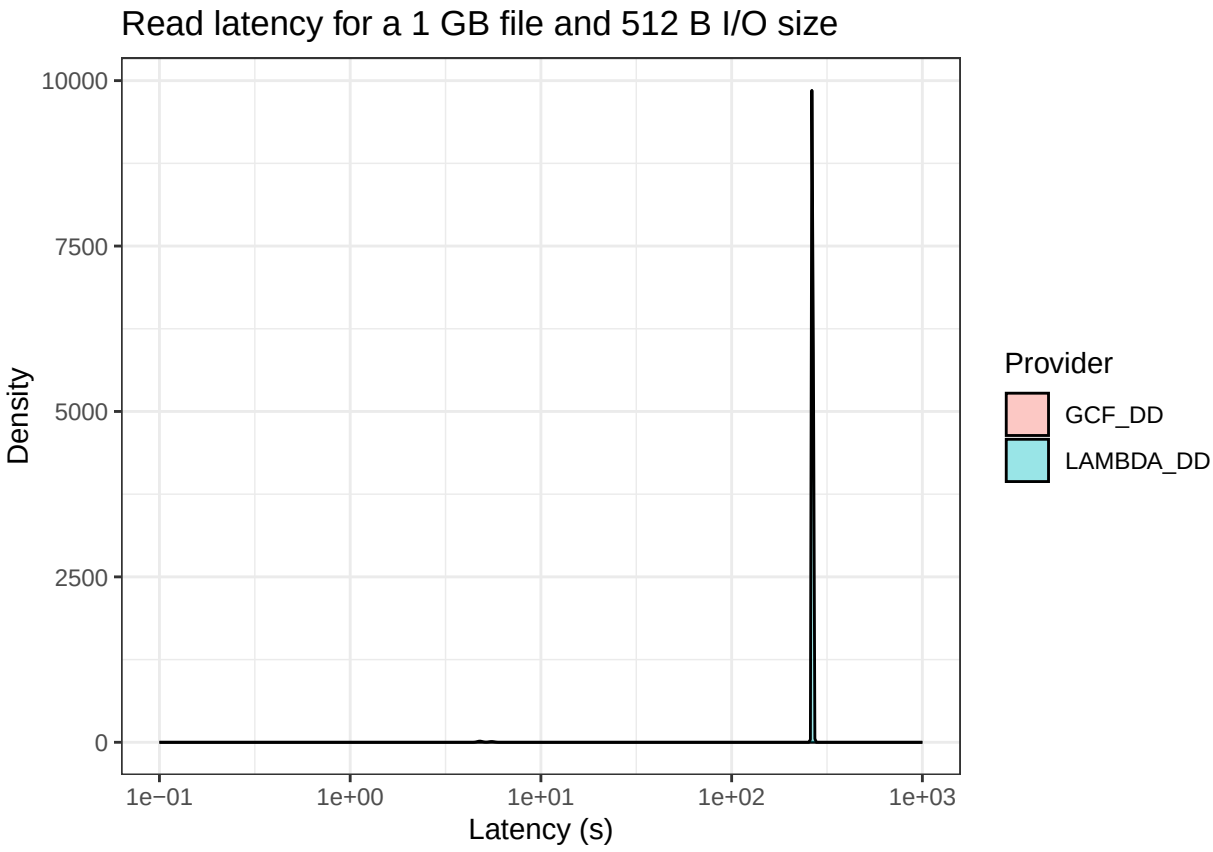
Density plots

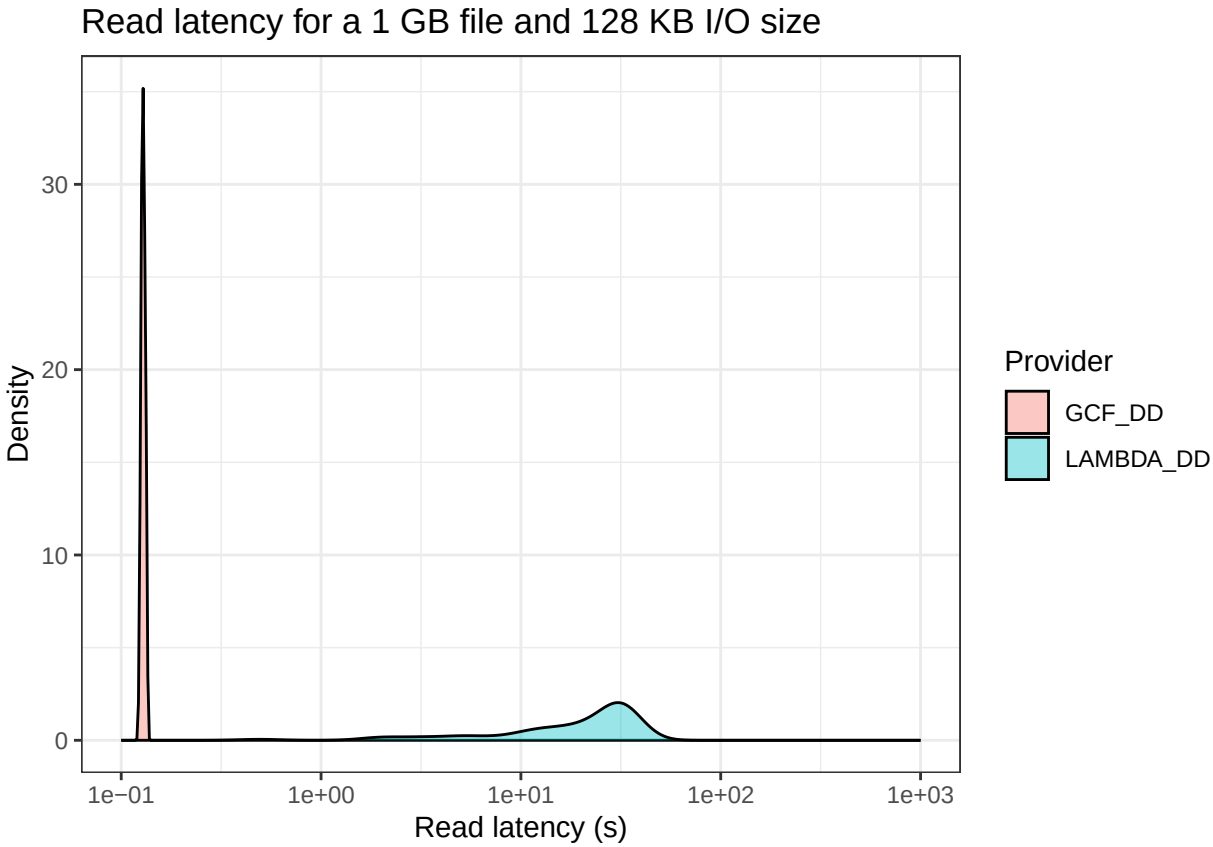




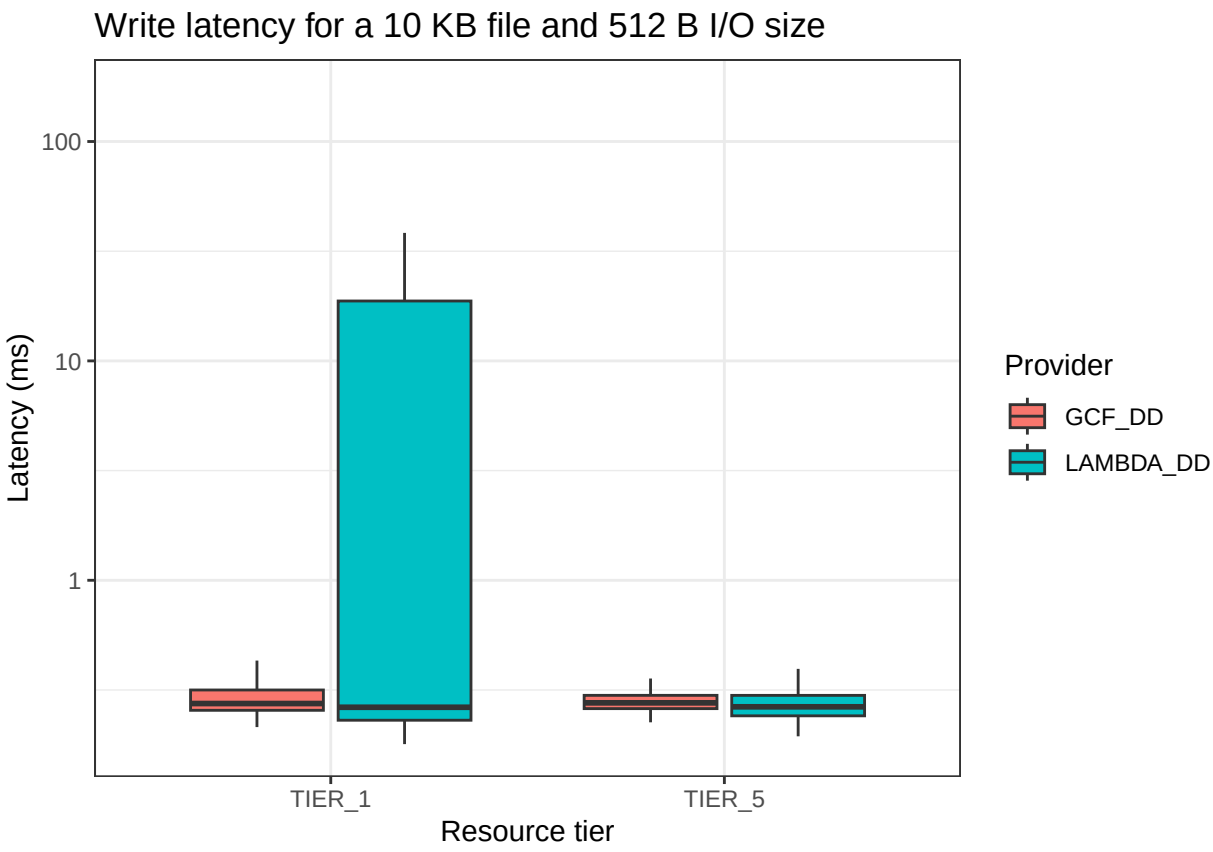

Write latency for a 1 GB file and 128 KB I/O size
TIER_5
Density
10
5
0
3.5
4.0
Latency (s)
Provider
GCF_DD
LAMBDA_DD

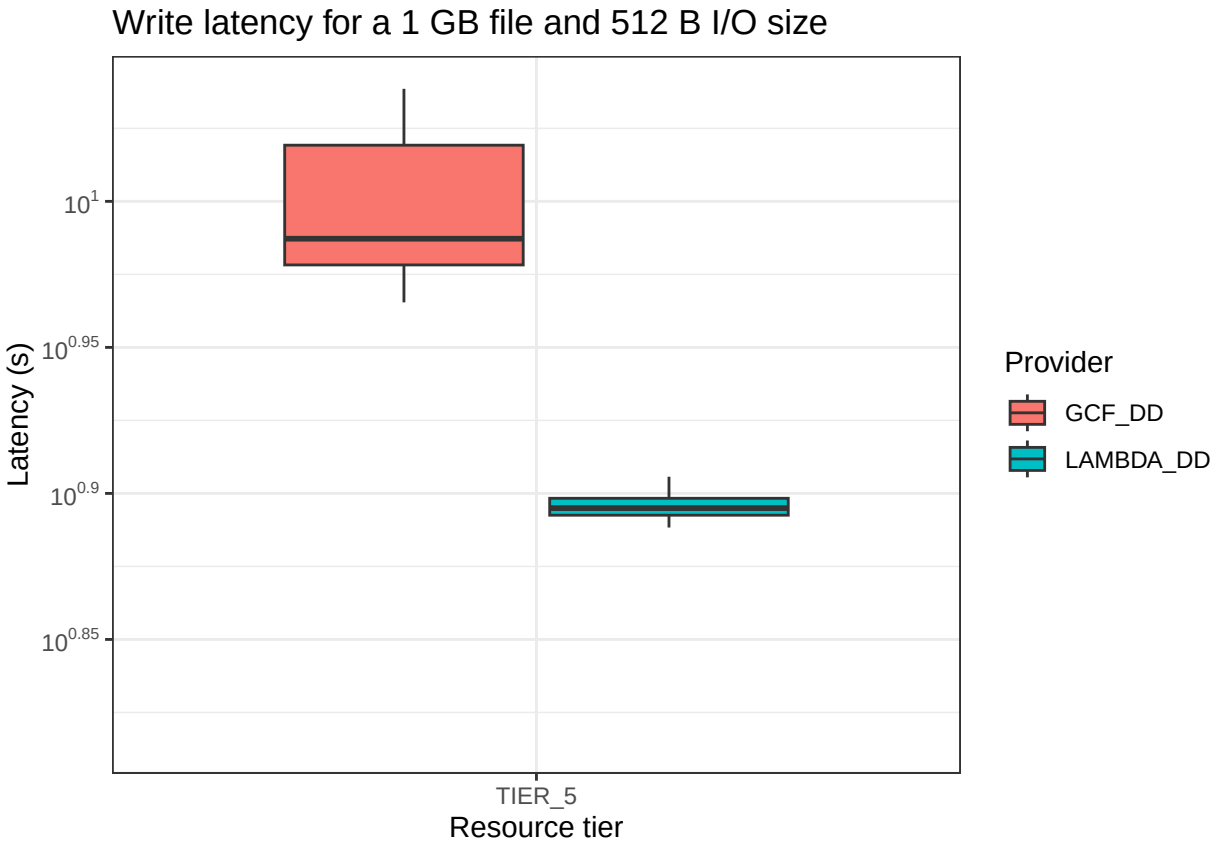

Read latency for a 10 KB file and 512 B I/O size
TIER_1
TIER_5
Density
9
6
3
0
1
10
100
Latency (ms)
Provider
GCF_DD
LAMBDA_DD

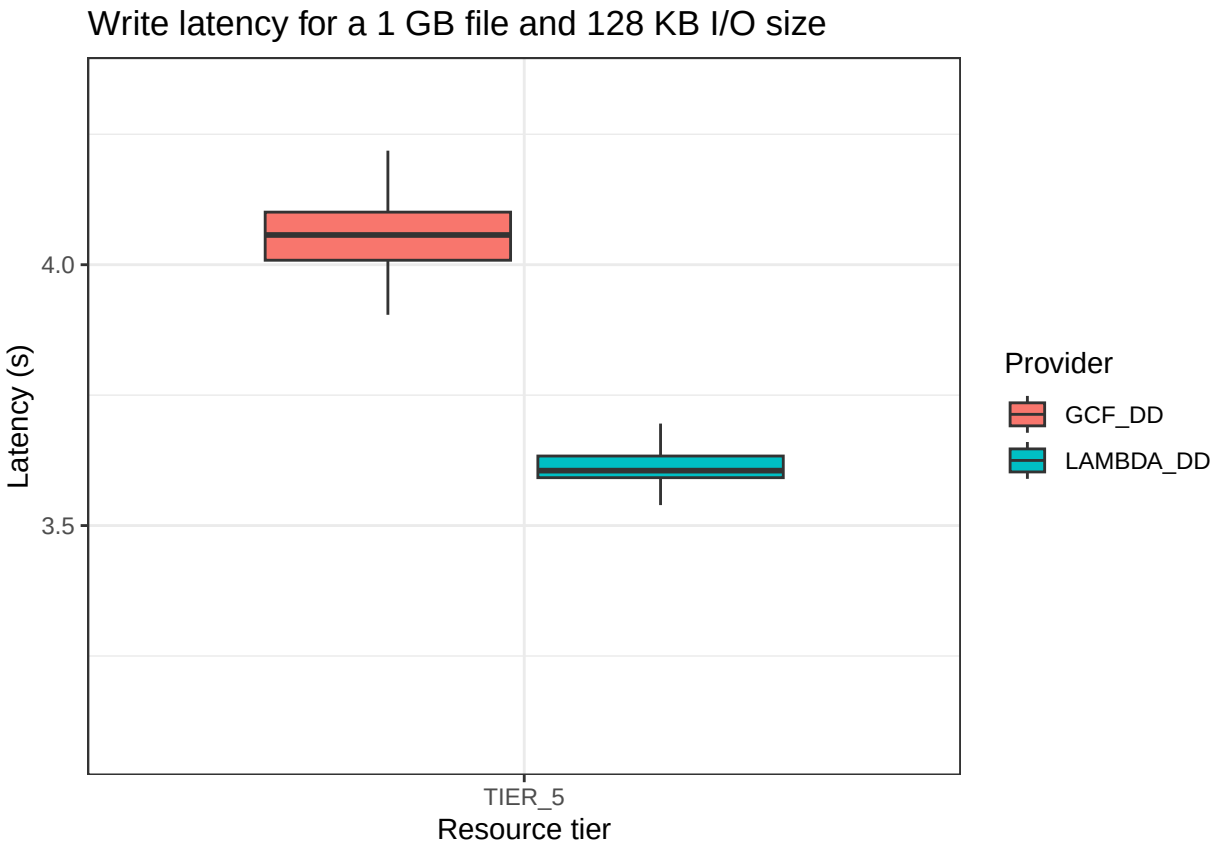

Read latency for a 1 GB file and 512 B I/O size
Density
10000
7500
5000
2500
0
1e−01
1e+00
1e+01
1e+02
1e+03
Latency (s)
Provider
GCF_DD
LAMBDA_DD

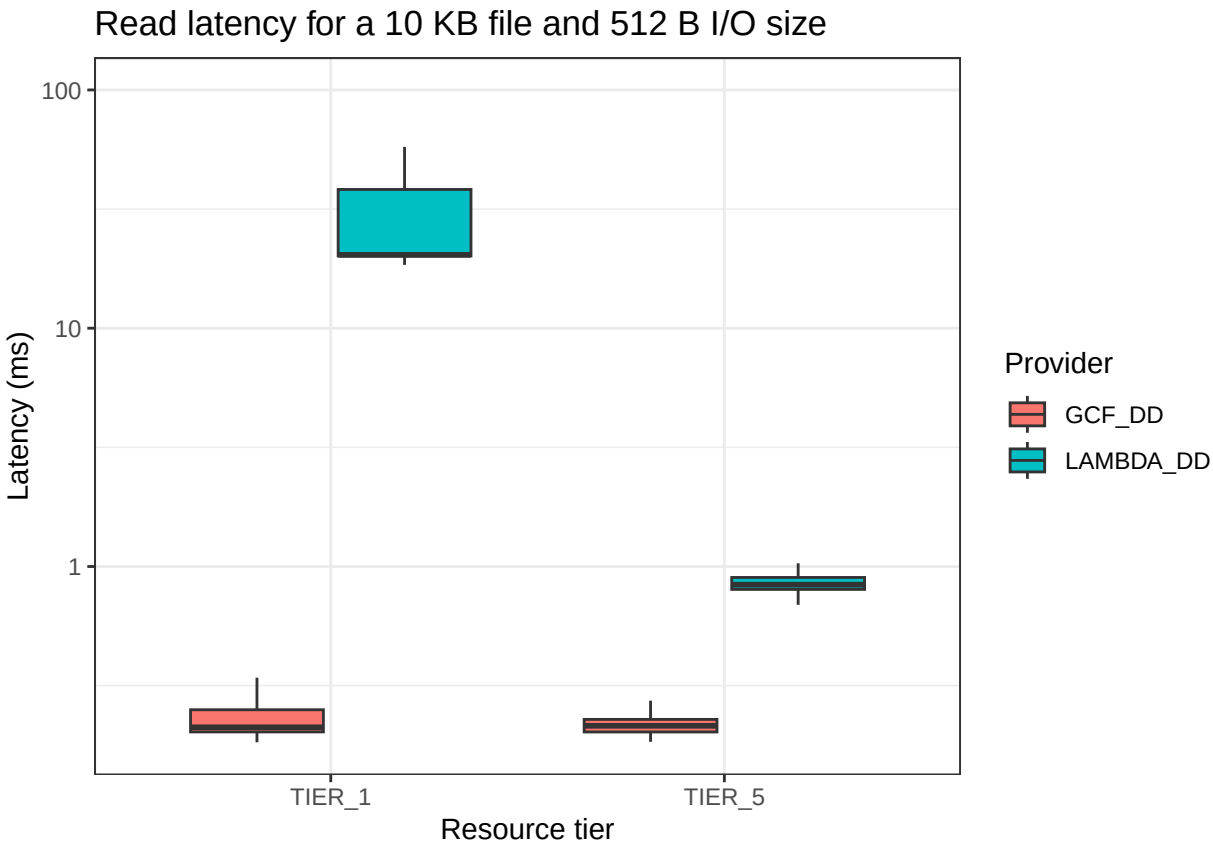

Read latency for a 1 GB file and 128 KB I/O size
Density
30
20
10
0
1e−01
1e+00
1e+01
1e+02
1e+03
Read latency (s)
Provider
GCF_DD
LAMBDA_DD

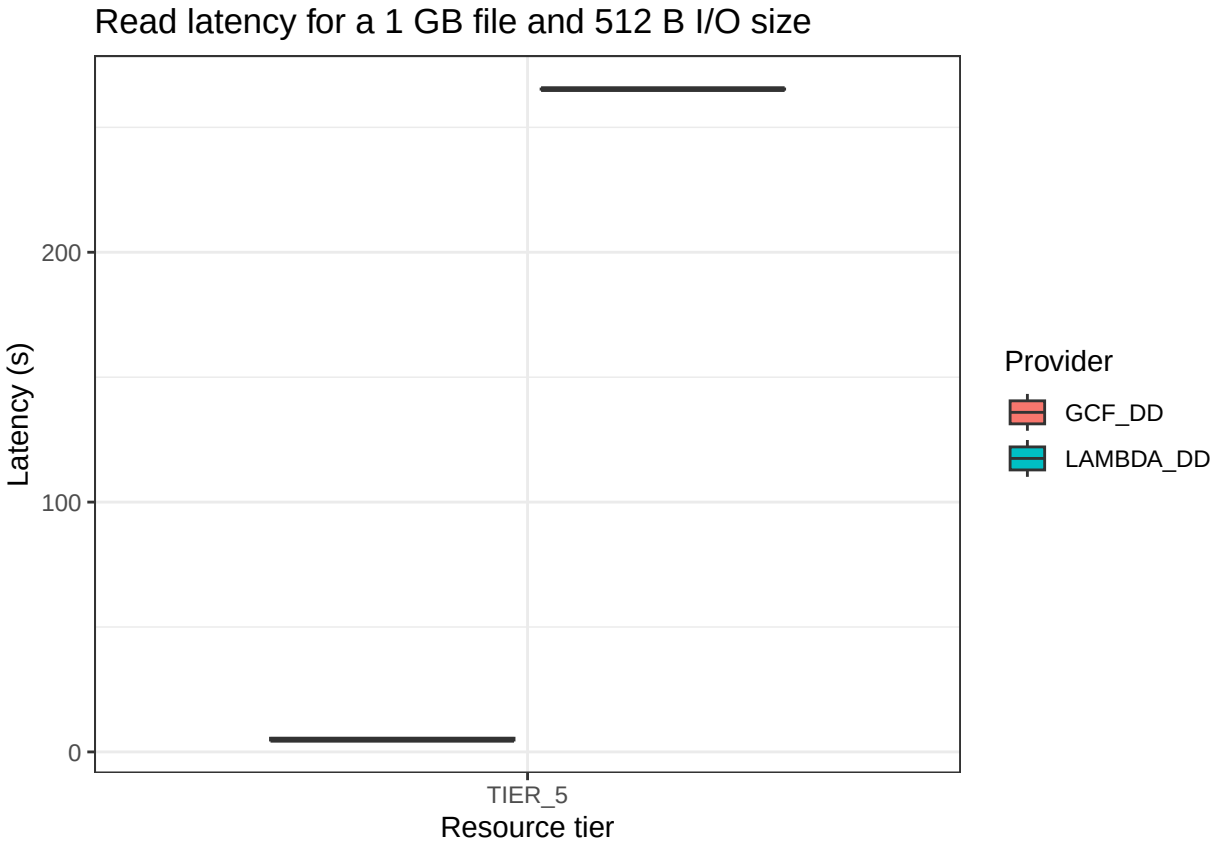
Box plots

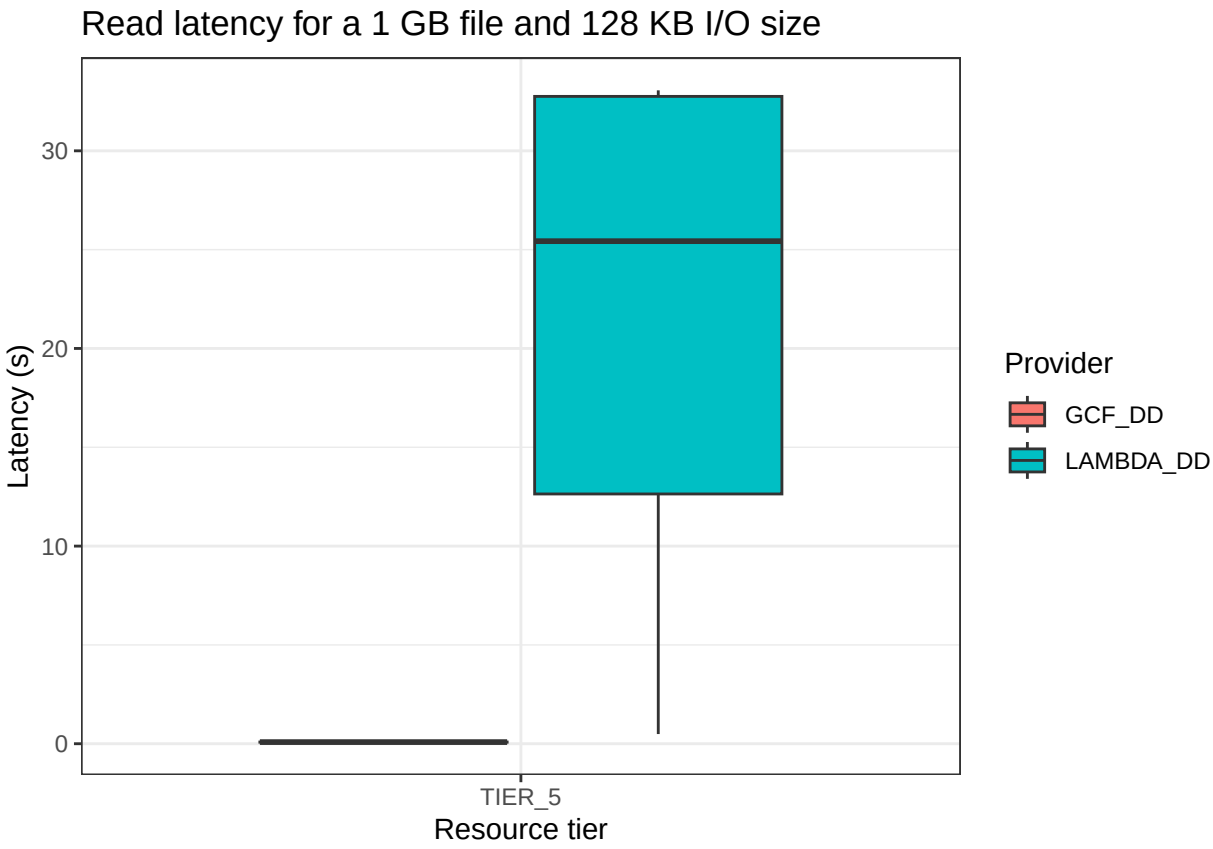

Write latency for a 10 KB file and 512 B I/O size
Latency (ms)
100
10
1
TIER_1
TIER_5
Resource tier
Provider
GCF_DD
LAMBDA_DD


Write latency for a 1 GB file and 512 B I/O size
Latency (s)
10^{1}
$10^{0.95}$
$10^{0.9}$
$10^{0.85}$
TIER_5
Resource tier
Provider
GCF_DD
LAMBDA_DD


Write latency for a 1 GB file and 128 KB I/O size
Latency (s)
4.0
3.5
TIER_5
Resource tier
Provider
GCF_DD
LAMBDA_DD


Read latency for a 10 KB file and 512 B I/O size
100
10
1
Latency (ms)
TIER_1
TIER_5
Resource tier
Provider
GCF_DD
LAMBDA_DD


Read latency for a 1 GB file and 512 B I/O size
Latency (s)
200
100
0
TIER_5
Resource tier
Provider
GCF_DD
LAMBDA_DD


Read latency for a 1 GB file and 128 KB I/O size
Latency (s)
30
20
10
0
TIER_5
Resource tier
Provider
GCF_DD
LAMBDA_DD