

# Examples of `eval_methods()` and `plot_eval()` Functions

Joel Beck

```
devtools::load_all()
library(patchwork)

uk_cqr <- readr::read_rds(here::here("data_results", "uk_cqr.rds"))
uk_cqr_qsa <- readr::read_rds(here::here("data_results", "uk_cqr_qsa_uniform.rds"))
uk_cqr_qsa_ensemble <- readr::read_rds(here::here("data_results", "uk_cqr_qsa_uniform_ensemble.rds"))

hub_1 <- readr::read_rds(here::here("data_results", "hub_cqr_1.rds"))
hub_2 <- readr::read_rds(here::here("data_results", "hub_cqr_2.rds"))
hub_combined <- dplyr::bind_rows(hub_1, hub_2)
```

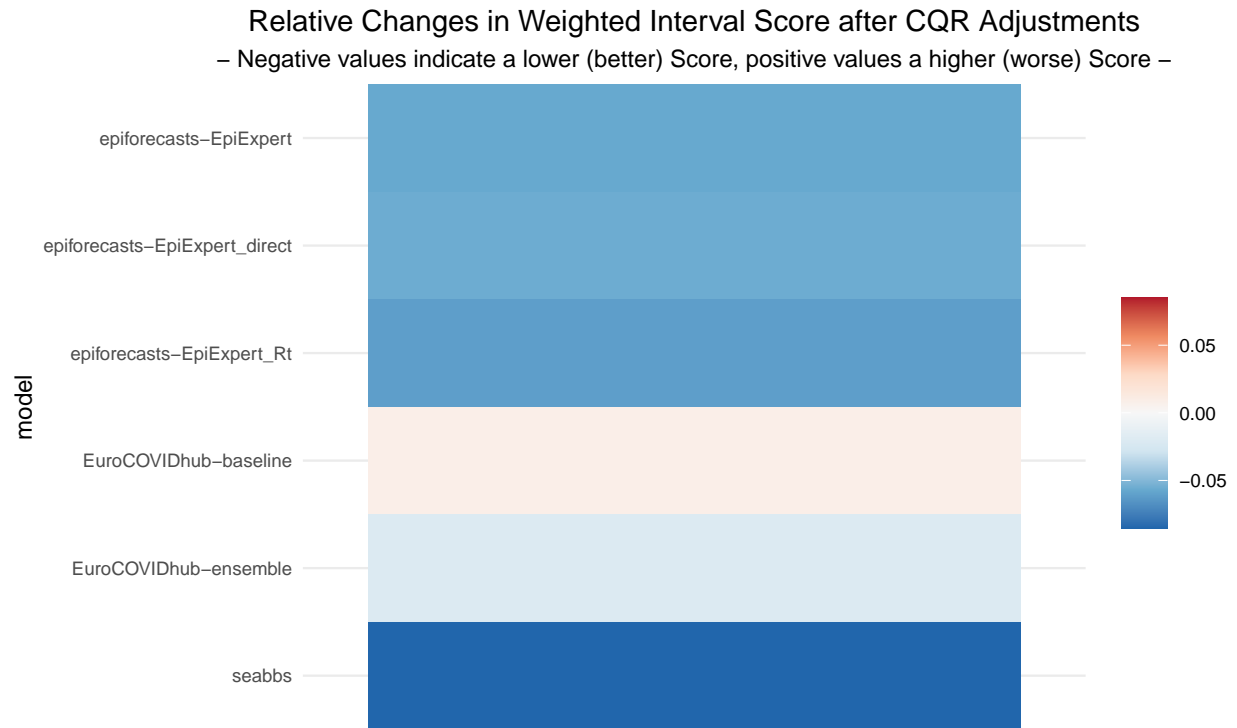
## UK Data

### CQR + Single Category

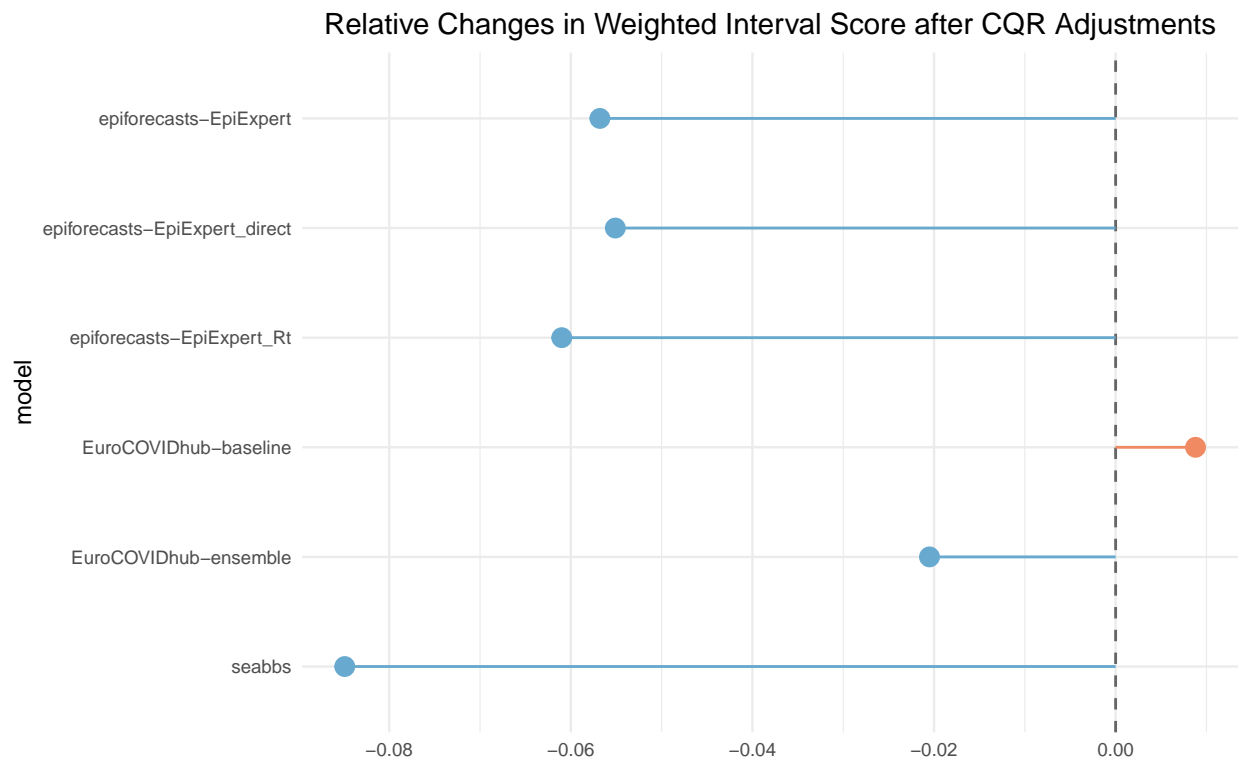
```
df_eval <- eval_methods(uk_cqr, summarise_by = "model")
eval_methods(uk_cqr, summarise_by = "target_type")
## # A tibble: 2 x 2
##   target_type relative_change
##   <chr>          <dbl>
## 1 Cases          -0.054
## 2 Deaths        -0.0721
eval_methods(uk_cqr, summarise_by = "horizon")
## # A tibble: 4 x 2
##   horizon relative_change
##   <dbl>          <dbl>
## 1     1          0.0006
## 2     2         -0.0042
## 3     3         -0.0623
## 4     4         -0.0782
eval_methods(uk_cqr, summarise_by = "quantile")
## # A tibble: 23 x 2
##   quantile relative_change
##   <dbl>          <dbl>
## 1    0.01         -0.401
## 2    0.025        -0.311
## 3    0.05         -0.207
## 4    0.1          -0.148
## 5    0.15         -0.0661
## 6    0.2          -0.0195
## 7    0.25         -0.0034
## 8    0.3          -0.0025
## 9    0.35         -0.0026
```

```
## 10      0.4      0.0038
## # ... with 13 more rows
```

```
plot_eval(df_eval)
```

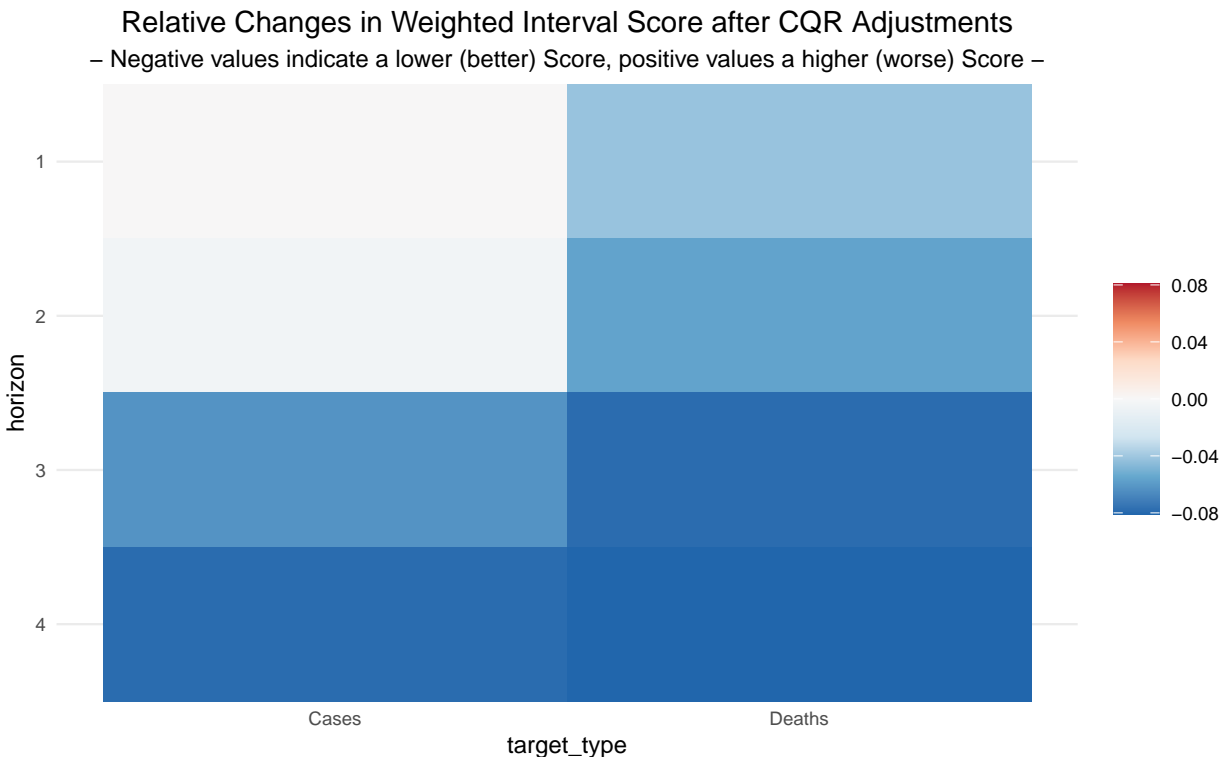


```
plot_eval(df_eval, heatmap = FALSE)
```



## CQR + 2 Categories

```
eval_methods(uk_cqr, summarise_by = c("model", "target_type"))
## # A tibble: 6 x 3
##   model                Cases Deaths
##   <chr>                <dbl>  <dbl>
## 1 epiforecasts-EpiExpert -0.0568 -0.0106
## 2 epiforecasts-EpiExpert_direct -0.0551 -0.0131
## 3 epiforecasts-EpiExpert_Rt -0.0609 -0.0976
## 4 EuroCOVIDhub-baseline    0.0094 -0.129
## 5 EuroCOVIDhub-ensemble -0.0205 -0.014
## 6 seabbs                 -0.0849 -0.0893
eval_methods(uk_cqr, summarise_by = c("model", "horizon"))
## # A tibble: 6 x 5
##   model                `4`      `3`      `2`      `1`
##   <chr>                <dbl>  <dbl>  <dbl>  <dbl>
## 1 epiforecasts-EpiExpert -0.0883 -0.0634  0.0109  0.0039
## 2 epiforecasts-EpiExpert_direct -0.0795 -0.0584 -0.0026 -0.0038
## 3 epiforecasts-EpiExpert_Rt -0.109  -0.0596  0.0153  0.0153
## 4 EuroCOVIDhub-baseline    0.0647  0.0019 -0.0731  0.024
## 5 EuroCOVIDhub-ensemble -0.0301 -0.0289  0.0049  0.0002
## 6 seabbs                 -0.109  -0.102  -0.0093 -0.0294
df_eval <- eval_methods(uk_cqr, summarise_by = c("horizon", "target_type"))
plot_eval(df_eval)
```

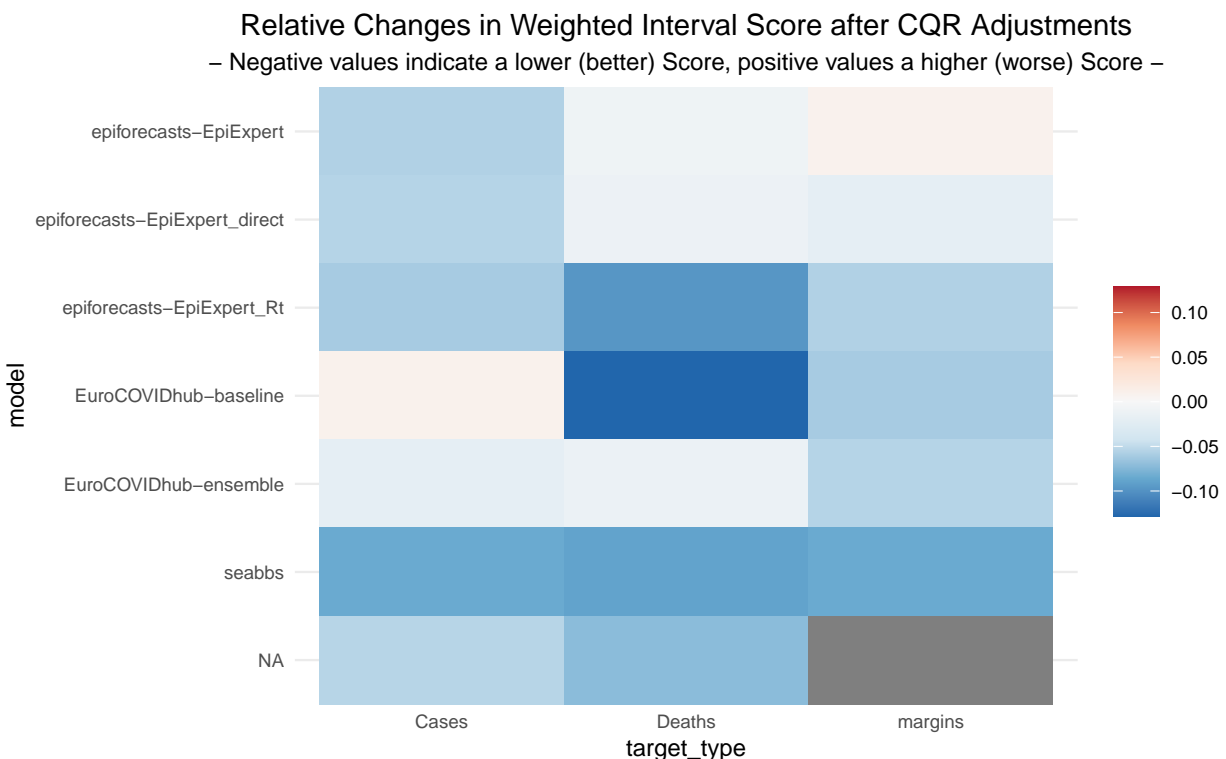


## Add marginal relative changes

```
# not really informative, since margins are dominated by category with largest
# values, here "Cases"
# => these margins are almost identical to one row / column of table

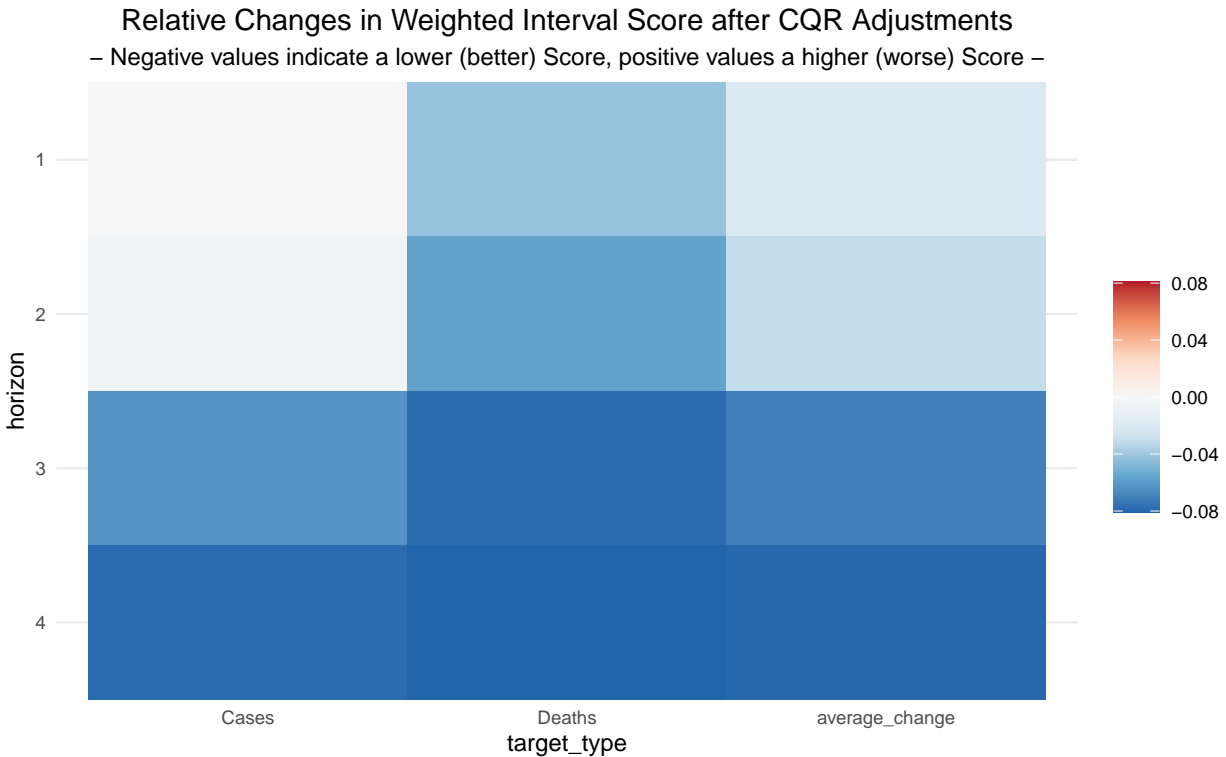
# sorted by increasing relative changes
eval_methods(uk_cqr, summarise_by = "model")
## # A tibble: 6 x 2
##   model                      relative_change
##   <chr>                      <dbl>
## 1 epiforecasts-EpiExpert      -0.0568
## 2 epiforecasts-EpiExpert_direct -0.0551
## 3 epiforecasts-EpiExpert_Rt   -0.061
## 4 EuroCOVIDhub-baseline       0.0088
## 5 EuroCOVIDhub-ensemble      -0.0205
## 6 seabbs                     -0.0849
eval_methods(uk_cqr, summarise_by = "target_type")
## # A tibble: 2 x 2
##   target_type relative_change
##   <chr>          <dbl>
## 1 Cases         -0.054
## 2 Deaths       -0.0721

# not sorted by relative changes
df_eval <- eval_methods(
  uk_cqr,
  summarise_by = c("model", "target_type"), margins = TRUE
)
plot_eval(df_eval)
```

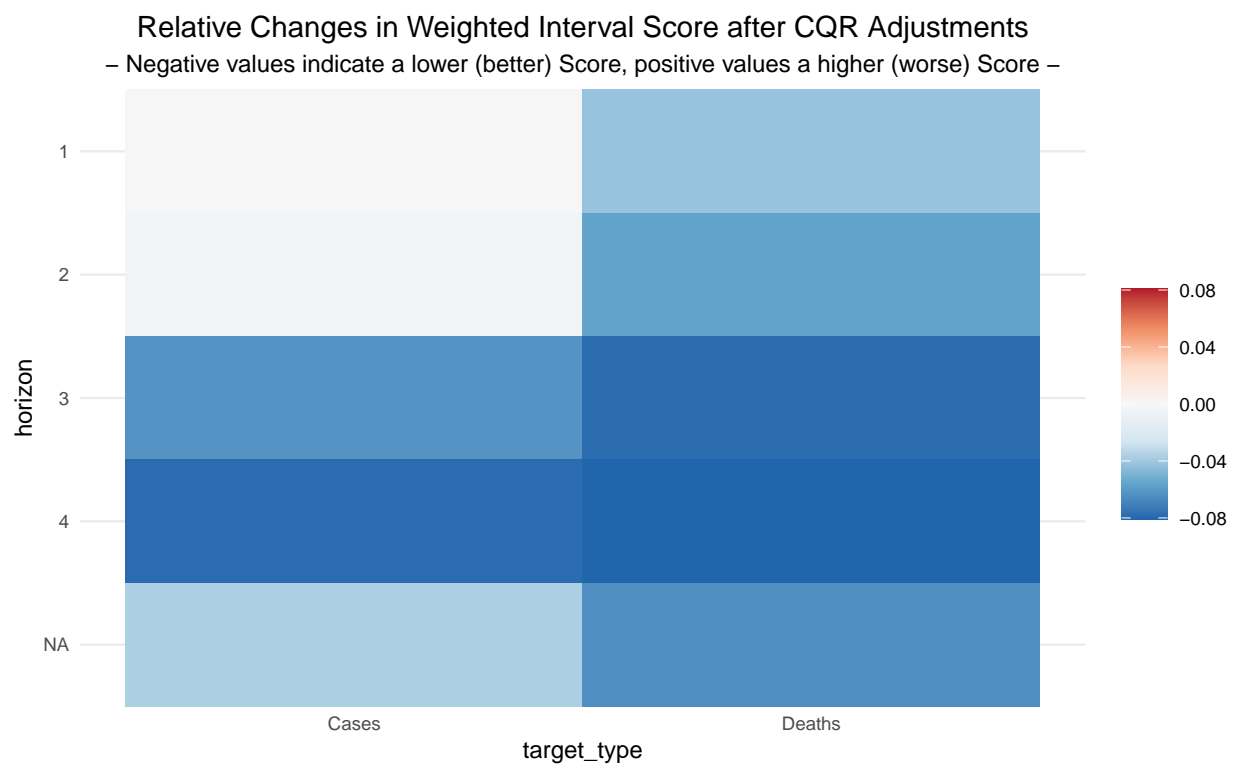


## Add 'average' (geometric mean) relative changes of rows and columns

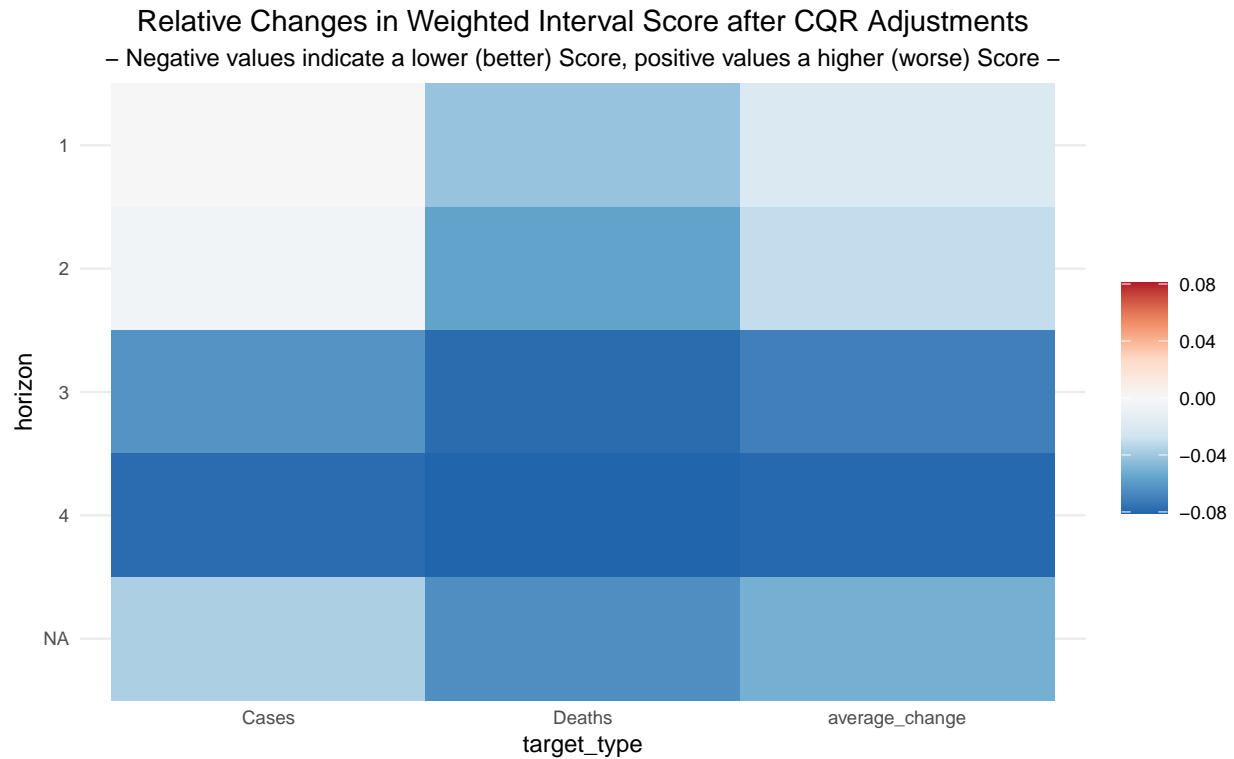
```
df_eval <- eval_methods(  
  uk_cqr,  
  summarise_by = c("horizon", "target_type"), row_averages = TRUE  
)  
plot_eval(df_eval)
```



```
df_eval <- eval_methods(  
  uk_cqr,  
  summarise_by = c("horizon", "target_type"), col_averages = TRUE  
)  
plot_eval(df_eval)
```



```
df_eval <- eval_methods(
  uk_cqr,
  summarise_by = c("horizon", "target_type"), row_averages = TRUE,
  col_averages = TRUE
)
plot_eval(df_eval)
```



## Multiple Methods

### CQR + QSA Uniform

```
eval_methods(uk_cqr_qsa, summarise_by = "model")
## # A tibble: 6 x 3
##   model                cqr qsa_uniform
##   <chr>                <dbl>    <dbl>
## 1 epiforecasts-EpiExpert -0.0568 -0.122
## 2 epiforecasts-EpiExpert_direct -0.0551 -0.124
## 3 epiforecasts-EpiExpert_Rt -0.061 -0.102
## 4 EuroCOVIDhub-baseline 0.0088 0.0286
## 5 EuroCOVIDhub-ensemble -0.0205 -0.0343
## 6 seabbs -0.0849 -0.144
eval_methods(uk_cqr_qsa, summarise_by = "target_type")
## # A tibble: 2 x 3
##   target_type    cqr qsa_uniform
##   <chr>        <dbl>    <dbl>
## 1 Cases      -0.054 -0.1
## 2 Deaths   -0.0721 -0.0775
eval_methods(uk_cqr_qsa, summarise_by = "horizon")
## # A tibble: 4 x 3
##   horizon    cqr qsa_uniform
##   <dbl>    <dbl>    <dbl>
## 1      1 0.0006 0.0312
## 2      2 -0.0042 -0.0472
## 3      3 -0.0623 -0.128
## 4      4 -0.0782 -0.123
eval_methods(uk_cqr_qsa, summarise_by = "quantile")
```

```
## # A tibble: 23 x 3
##   quantile      cqr qsa_uniform
##   <dbl>    <dbl>    <dbl>
## 1  0.01 -0.401    -0.612
## 2  0.025 -0.311    -0.469
## 3  0.05 -0.207    -0.345
## 4  0.1 -0.148    -0.216
## 5  0.15 -0.0661   -0.133
## 6  0.2 -0.0195   -0.0784
## 7  0.25 -0.0034   -0.0422
## 8  0.3 -0.0025   -0.0226
## 9  0.35 -0.0026   -0.0131
## 10 0.4  0.0038    -0.006
## # ... with 13 more rows
```

```
df_eval <- eval_methods(uk_cqr_qsa, summarise_by = "model")
p1 <- plot_eval(df_eval, base_size = 8) + ggplot2::labs(y = NULL)

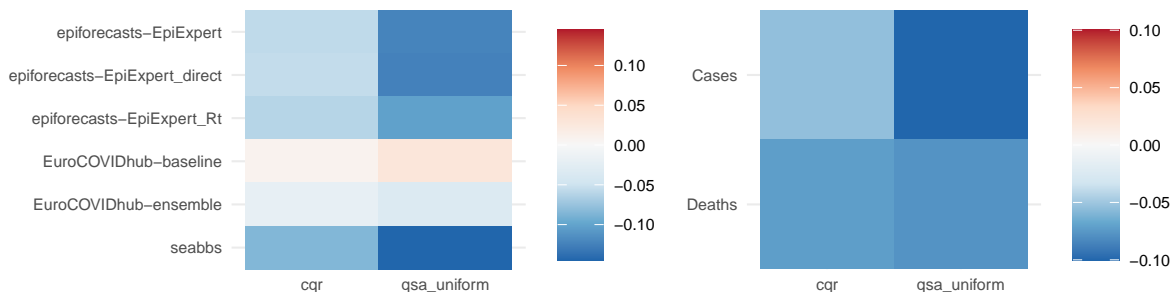
df_eval <- eval_methods(uk_cqr_qsa, summarise_by = "target_type")
p2 <- plot_eval(df_eval, base_size = 8) + ggplot2::labs(y = NULL)

df_eval <- eval_methods(uk_cqr_qsa, summarise_by = "horizon")
p3 <- plot_eval(df_eval, base_size = 8)

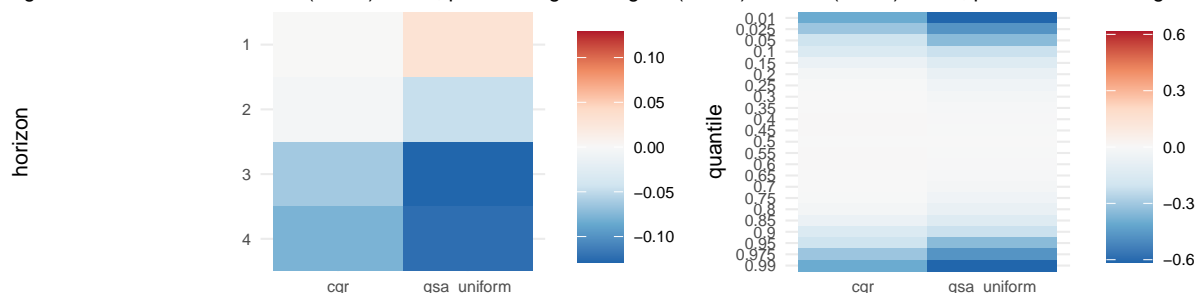
df_eval <- eval_methods(uk_cqr_qsa, summarise_by = "quantile")
p4 <- plot_eval(df_eval, base_size = 8)

(p1 + p2) / (p3 + p4)
```

Relative Changes in Weighted Interval Score after CQR Adjustment  
 - Negative values indicate a lower (better) Score, positive values a higher (worse) Score



Relative Changes in Weighted Interval Score after CQR Adjustment  
 - Negative values indicate a lower (better) Score, positive values a higher (worse) Score





## CQR + QSA Uniform + Ensemble

```
eval_methods(uk_cqr_qsa_ensemble, summarise_by = "model")
## # A tibble: 6 x 4
##   model                                cqr qsa_uniform ensemble
##   <chr>                                <dbl>      <dbl>      <dbl>
## 1 epiforecasts-EpiExpert             -0.0568    -0.122    -0.130
## 2 epiforecasts-EpiExpert_direct      -0.0551    -0.124    -0.137
## 3 epiforecasts-EpiExpert_Rt          -0.061     -0.102    -0.114
## 4 EuroCOVIDhub-baseline               0.0088     0.0286     0.0031
## 5 EuroCOVIDhub-ensemble              -0.0205    -0.0343   -0.0393
## 6 seabbs                             -0.0849    -0.144    -0.174
eval_methods(uk_cqr_qsa_ensemble, summarise_by = "target_type")
## # A tibble: 2 x 4
##   target_type      cqr qsa_uniform ensemble
##   <chr>            <dbl>      <dbl>      <dbl>
## 1 Cases          -0.054     -0.1     -0.116
## 2 Deaths        -0.0721    -0.0775   -0.0679
eval_methods(uk_cqr_qsa_ensemble, summarise_by = "horizon")
## # A tibble: 4 x 4
##   horizon      cqr qsa_uniform ensemble
##   <dbl>      <dbl>      <dbl>      <dbl>
## 1      1  0.0006     0.0312     0.0262
## 2      2 -0.0042    -0.0472    -0.0371
## 3      3 -0.0623    -0.128     -0.139
## 4      4 -0.0782    -0.123     -0.155
eval_methods(uk_cqr_qsa_ensemble, summarise_by = "quantile")
## # A tibble: 23 x 4
##   quantile      cqr qsa_uniform ensemble
##   <dbl>      <dbl>      <dbl>      <dbl>
## 1  0.01 -0.401     -0.612    -0.556
## 2  0.025 -0.311     -0.469    -0.476
## 3  0.05 -0.207     -0.345    -0.348
## 4  0.1 -0.148     -0.216    -0.227
## 5  0.15 -0.0661    -0.133    -0.132
## 6  0.2 -0.0195    -0.0784   -0.0889
## 7  0.25 -0.0034    -0.0422   -0.0806
## 8  0.3 -0.0025    -0.0226   -0.0649
## 9  0.35 -0.0026    -0.0131   -0.049
## 10 0.4  0.0038    -0.006    -0.0256
## # ... with 13 more rows
```

```
df_eval <- eval_methods(uk_cqr_qsa_ensemble, summarise_by = "model")
p1 <- plot_eval(df_eval, base_size = 8) + ggplot2::labs(y = NULL)

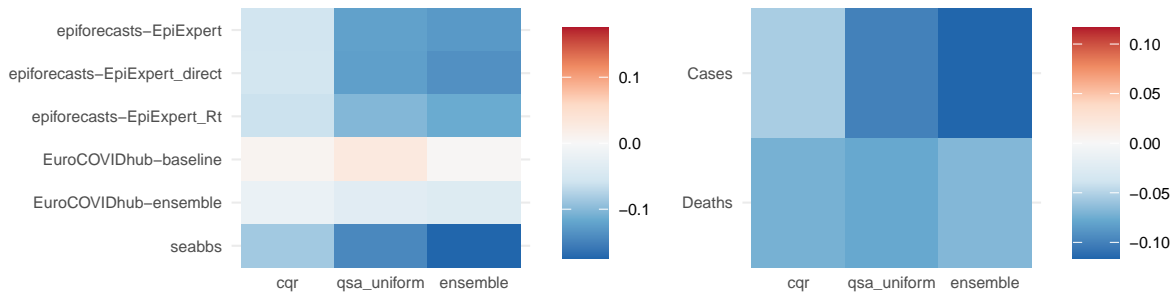
df_eval <- eval_methods(uk_cqr_qsa_ensemble, summarise_by = "target_type")
p2 <- plot_eval(df_eval, base_size = 8) + ggplot2::labs(y = NULL)

df_eval <- eval_methods(uk_cqr_qsa_ensemble, summarise_by = "horizon")
p3 <- plot_eval(df_eval, base_size = 8)

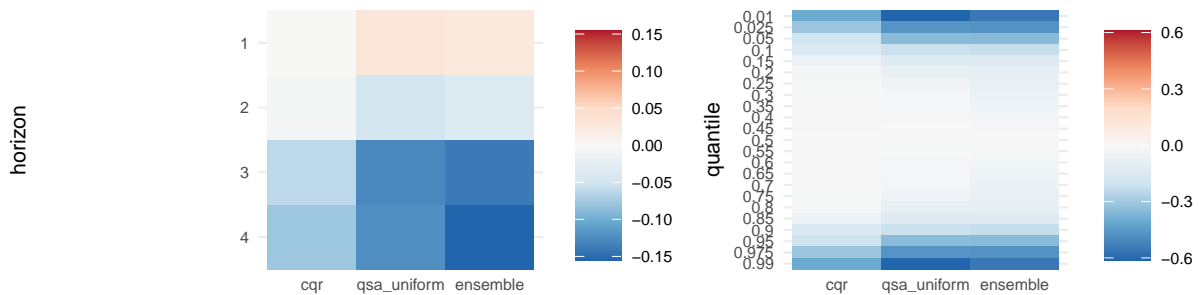
df_eval <- eval_methods(uk_cqr_qsa_ensemble, summarise_by = "quantile")
p4 <- plot_eval(df_eval, base_size = 8)
```

(p1 + p2) / (p3 + p4)

Relative Changes in Weighted Interval Score after CQR Adjustment  
 - Negative values indicate a lower (better) Score, positive values a higher (worse) Score



Relative Changes in Weighted Interval Score after CQR Adjustment  
 - Negative values indicate a lower (better) Score, positive values a higher (worse) Score



## European Forecast Hub Data

```
# data set very large, choose e.g. one model to avoid time limit error
hub_combined <- hub_combined |>
  dplyr::filter(model == "EuroCOVIDhub-ensemble")
```

### CQR + Single Category

```
df_eval <- eval_methods(hub_combined, summarise_by = "location_name")
p1 <- plot_eval(df_eval, heatmap = FALSE, base_size = 8) + ggplot2::labs(y = NULL)

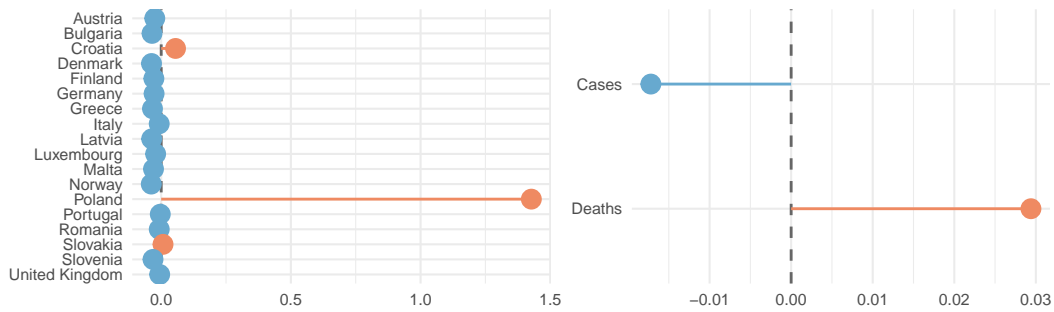
df_eval <- eval_methods(hub_combined, summarise_by = "target_type")
p2 <- plot_eval(df_eval, heatmap = FALSE, base_size = 8) + ggplot2::labs(y = NULL)

df_eval <- eval_methods(hub_combined, summarise_by = "horizon")
p3 <- plot_eval(df_eval, heatmap = FALSE, base_size = 8)

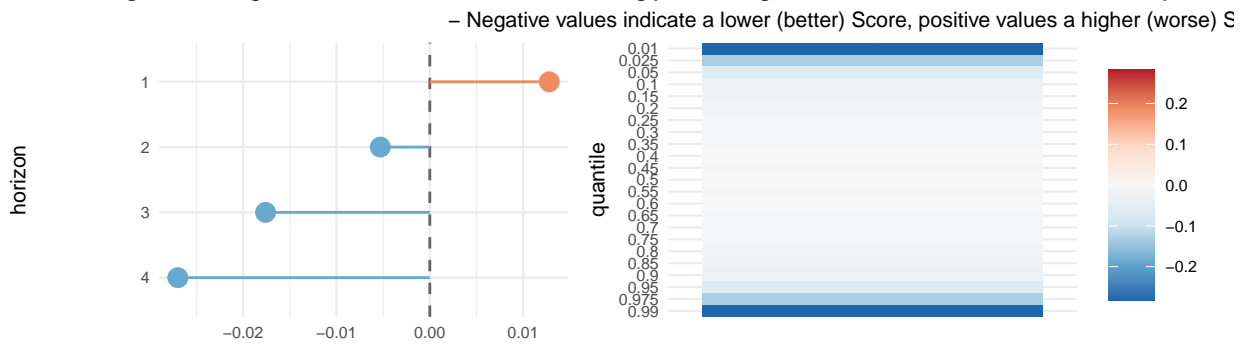
df_eval <- eval_methods(hub_combined, summarise_by = "quantile")
p4 <- plot_eval(df_eval, base_size = 8)

(p1 + p2) / (p3 + p4)
```

Relative Changes in Weighted Interval Score after CQR Adjustment



Relative Changes in Weighted Interval Score after CQR Adjustment



## CQR + 2 Categories

```
df_eval <- eval_methods(
  hub_combined,
  summarise_by = c("horizon", "target_type")
)
plot_eval(df_eval) + ggplot2::labs(x = NULL)
```

