Pilot Analysis

Preprocess data

Read in the results data.

```
# TODO: read in the method and dataset meta information and perform joins.
df <-
  # read in all columns as strings because of infinite values in table
  readr::read_tsv(
    "output/pilot/joint_embedding/output.extract_scores.output.tsv",
    col_types = c(.default = "c")
  ) %>%
  mutate(
    # manually convert to floats afterwards
    value = as.numeric(value),
    # extract meta info
    dataset_loader = gsub("_.*", "", dataset_id),
    dataset_group = gsub(".*_", "", dataset_id),
    method_group = gsub("_.*", "", method_id)
  ) %>%
  spread(metric_id, value) %>%
  mutate(
    format_check = (correct_format + finished) / 2
  )
write_tsv(df %>% gather(metric_id, value, -starts_with("dataset_"), -starts_with("method_")), "results/pilot_joi
```

Calculate mean score per method. Infinite values are replaced by the highest value in the results.

```
replace_inf <- function(x) ifelse(is.infinite(x), max(x[is.finite(x)]), x)
summ <- df %>%
  gather(metric_id, value, -starts_with("dataset_"), -starts_with("method_")) %>%
  mutate(value = replace_inf(value)) %>% # replace infinite values with the max
  group_by(method_id, metric_id) %>%
  summarise(
    mean = mean(value),
    sd = sd(value),
    var = var(value)
  )
```

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

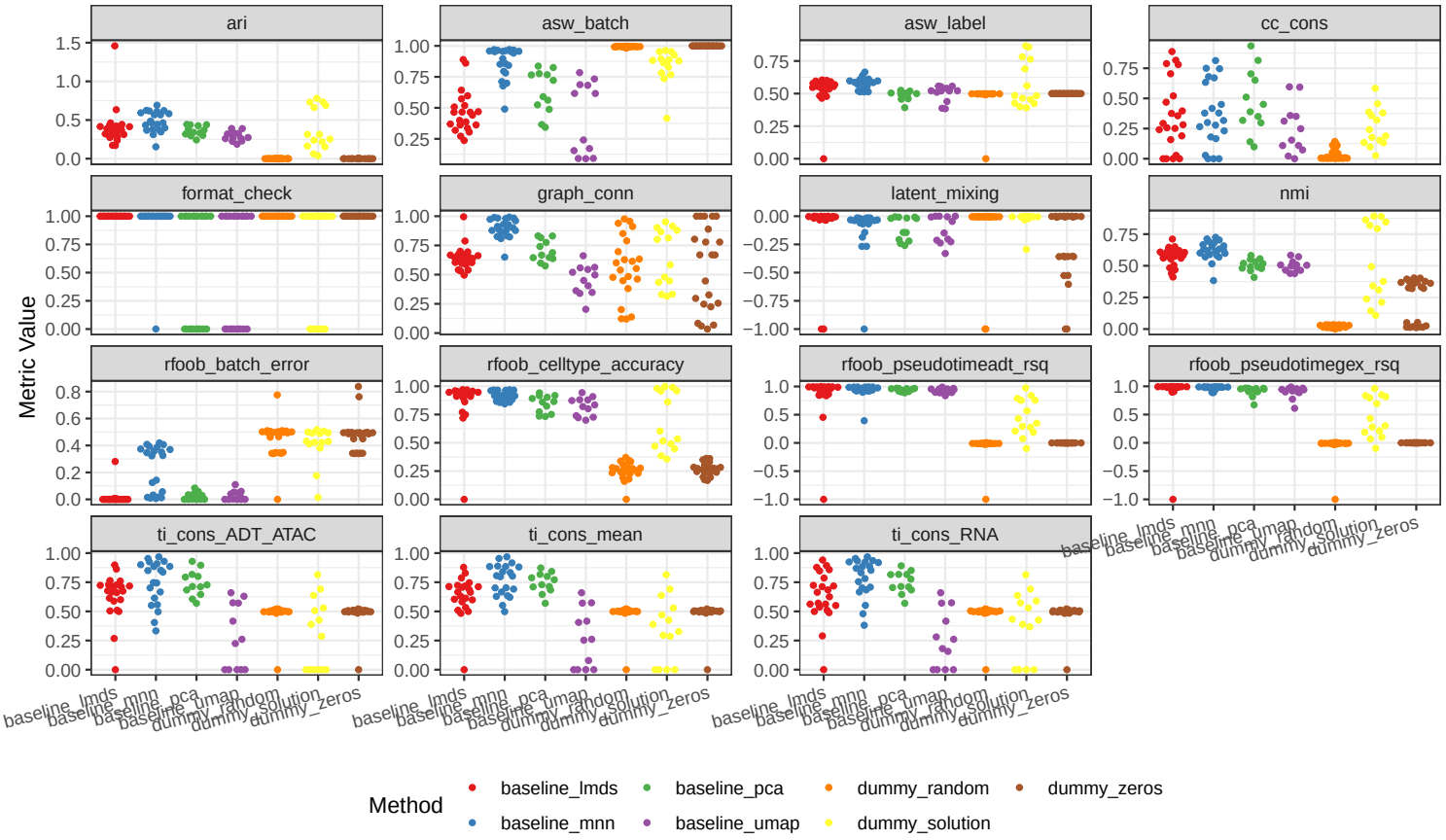
```
write_tsv(summ, "results/pilot_joint_embedding_summary.tsv")
```

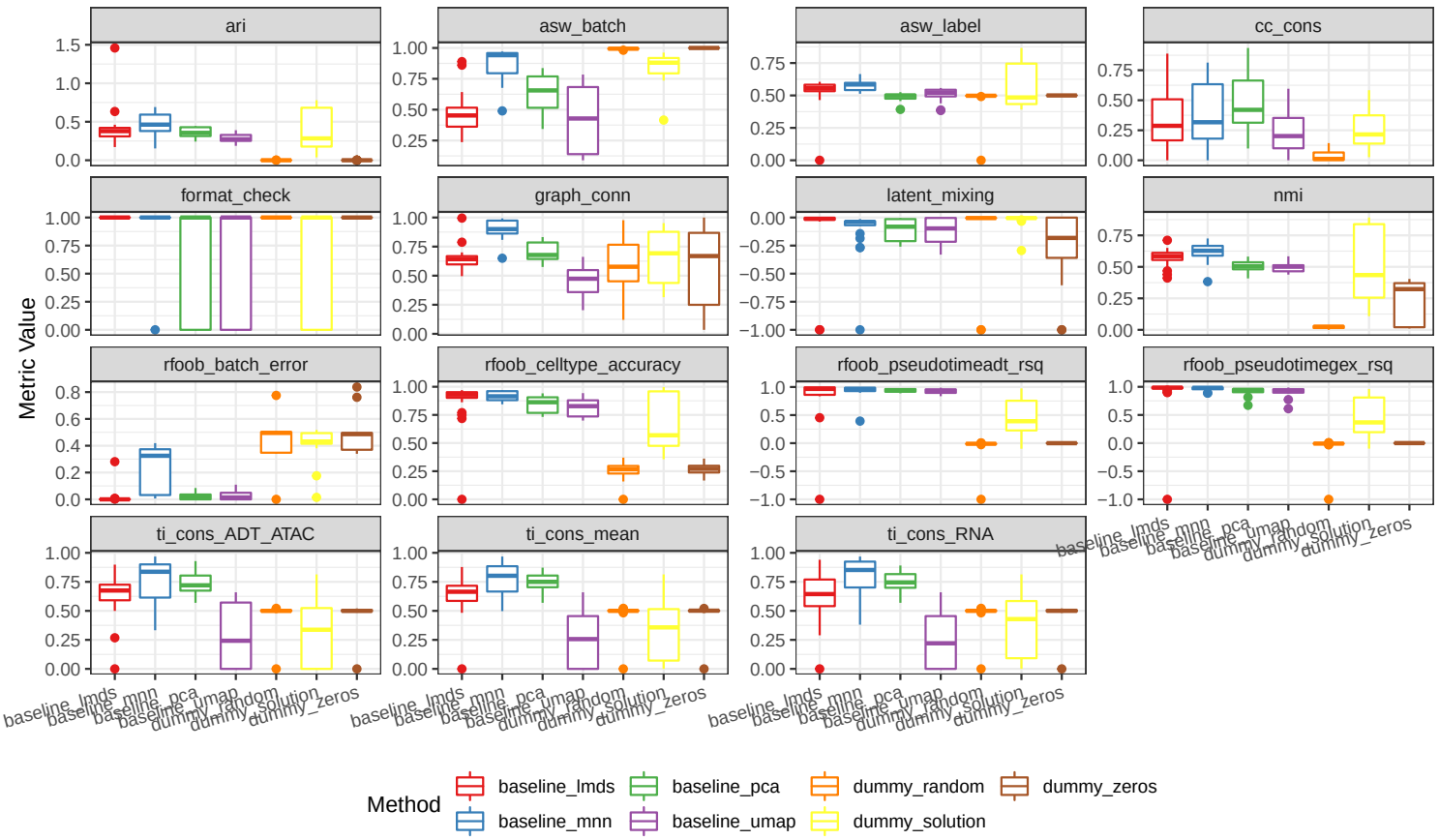
Visualise results

Colour by method.

```
## Warning: Removed 22 rows containing missing values (position_quasirandom).
```

```
## Warning: Removed 22 rows containing non-finite values (stat_boxplot).
```





Colour by dataset source.

```
## Warning: Removed 22 rows containing missing values (position_quasirandom).
```

```
## Warning: Removed 22 rows containing non-finite values (stat_boxplot).
```