Results

2025-08-06

```r
library(readr)
```

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:readr':
##
##     col_factor
```

```r
plot_accuracy <- function(filename, title_text, subtitle_text=NULL) {
  # relative file path
  file_path <- file.path("..", "results", "accuracy_tables", filename)

  # load respective table
  accuracy_data <- read.csv(file_path)
  accuracy_data <- accuracy_data %>%
    mutate(
      method = factor(method, levels = unique(method)),
      method = recode(method, "top5_llm" = "embed top5\nLLM")
    )

  # plot
  ggplot(accuracy_data, aes(x = method, y = accuracy * 100)) +
    geom_bar(stat = "identity", width = 0.7, fill = "#DDDDDD", color = "black") +
    geom_text(aes(label = sprintf("%.1f%%", accuracy * 100)),
```

```r
          vjust = -0.5, size = 5) +
    scale_y_continuous(
      labels = percent_format(scale = 1),
      breaks = seq(0, 100, by = 10),
      limits = c(0, 100),
      expand = c(0, 0)
    ) +
    labs(
      title = title_text,
      subtitle = subtitle_text,
      x = "Method",
      y = "Accuracy"
    ) +
    theme_classic(base_size = 14, base_family = "Helvetica") +
    theme(
      axis.title.x = element_text(size = 18),
      axis.title.y = element_text(size = 18),
      axis.text.x = element_text(size = 18, angle = 0, hjust = 0.5),
      axis.text.y = element_text(size = 14),
      plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
      plot.subtitle = element_text(size = 13, hjust = 0.5),
      legend.position = "none"
    )
}
```
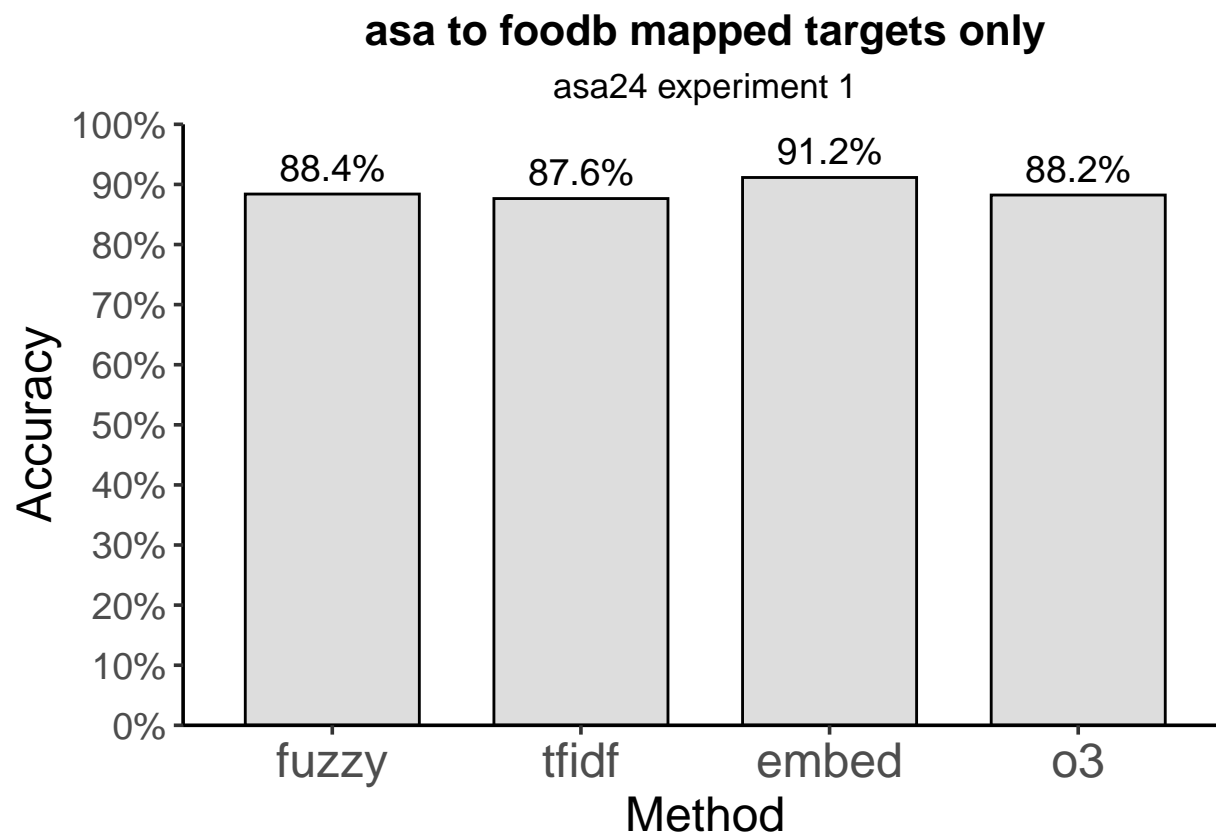
ASA Experiment 1

```r
plot_accuracy(
  filename = "asa24_experiment_1_accuracy.csv",
  title_text = "asa to foodb mapped targets only",
  subtitle_text = "asa24 experiment 1"
)
```

asa to foodb mapped targets only
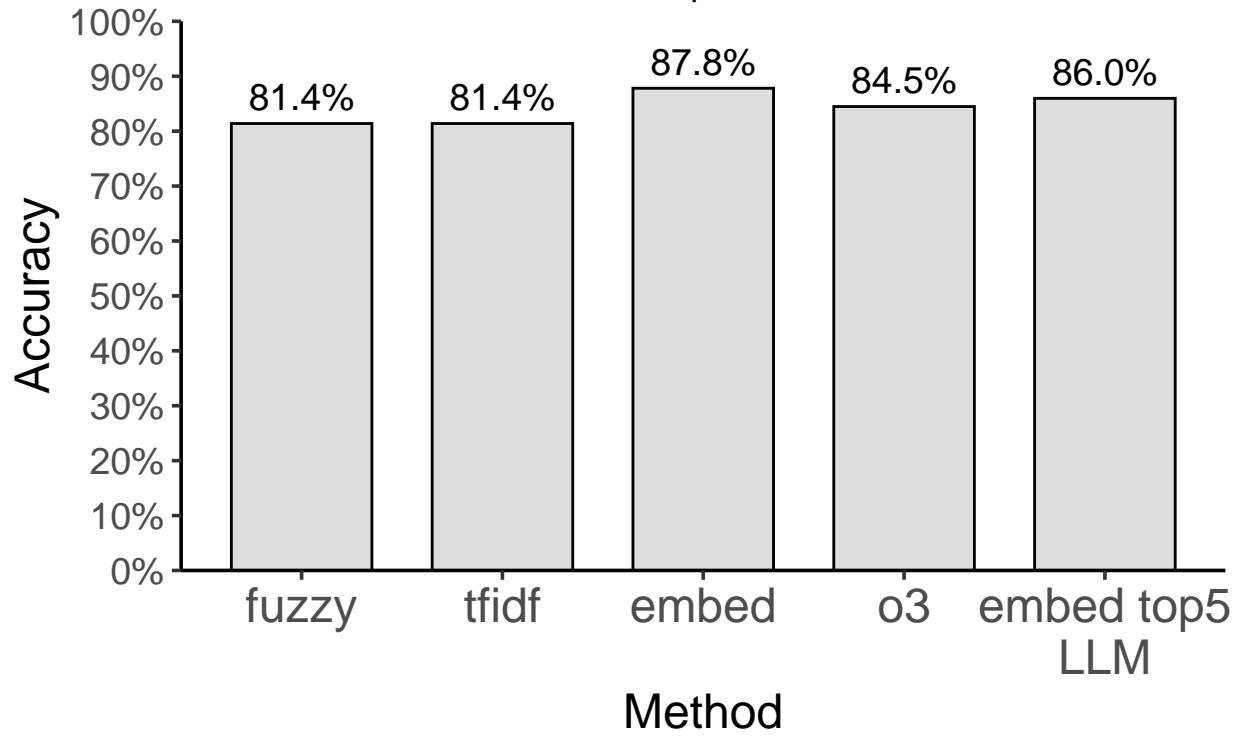asa24 experiment 1

ASA Experiment 2

```
plot_accuracy(
  filename = "asa24_experiment_2_accuracy.csv",
  title_text = "ASA24 Mapping to FooDB",
  subtitle_text = "asa24 experiment 2"
)
```
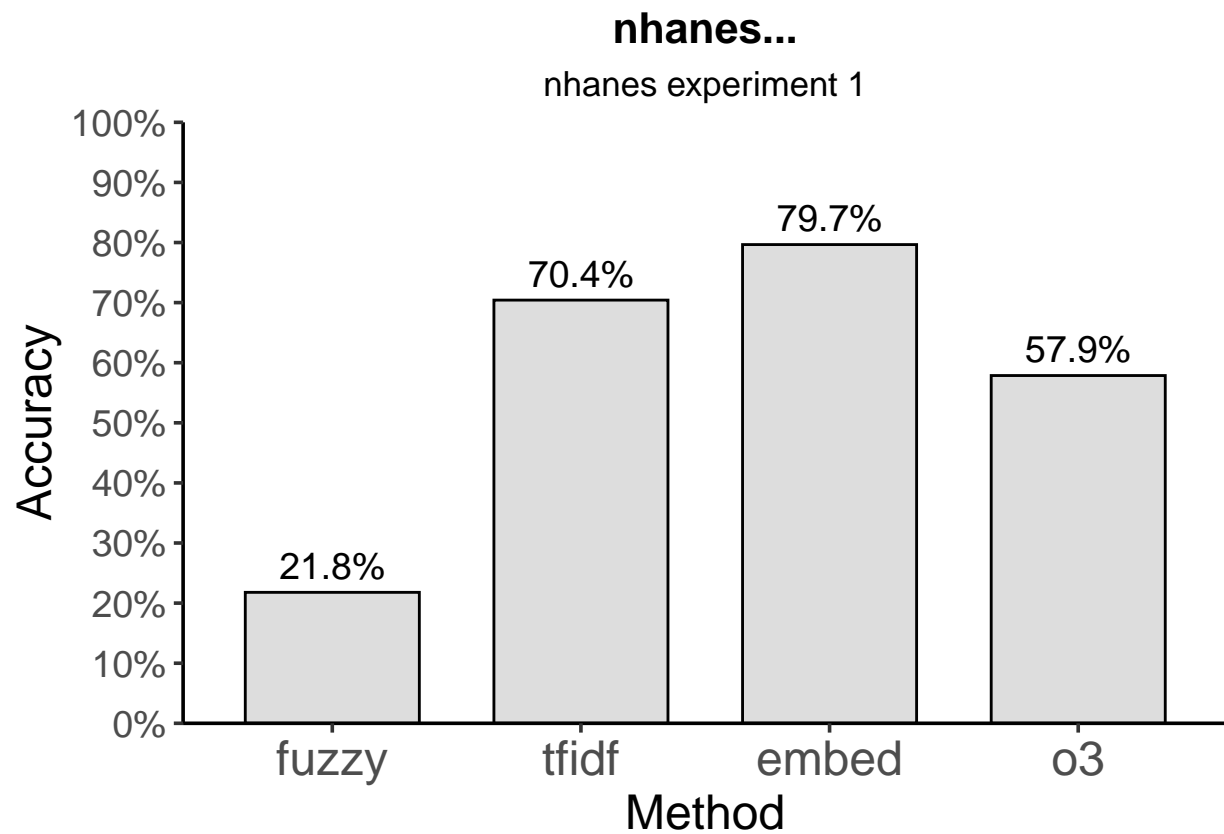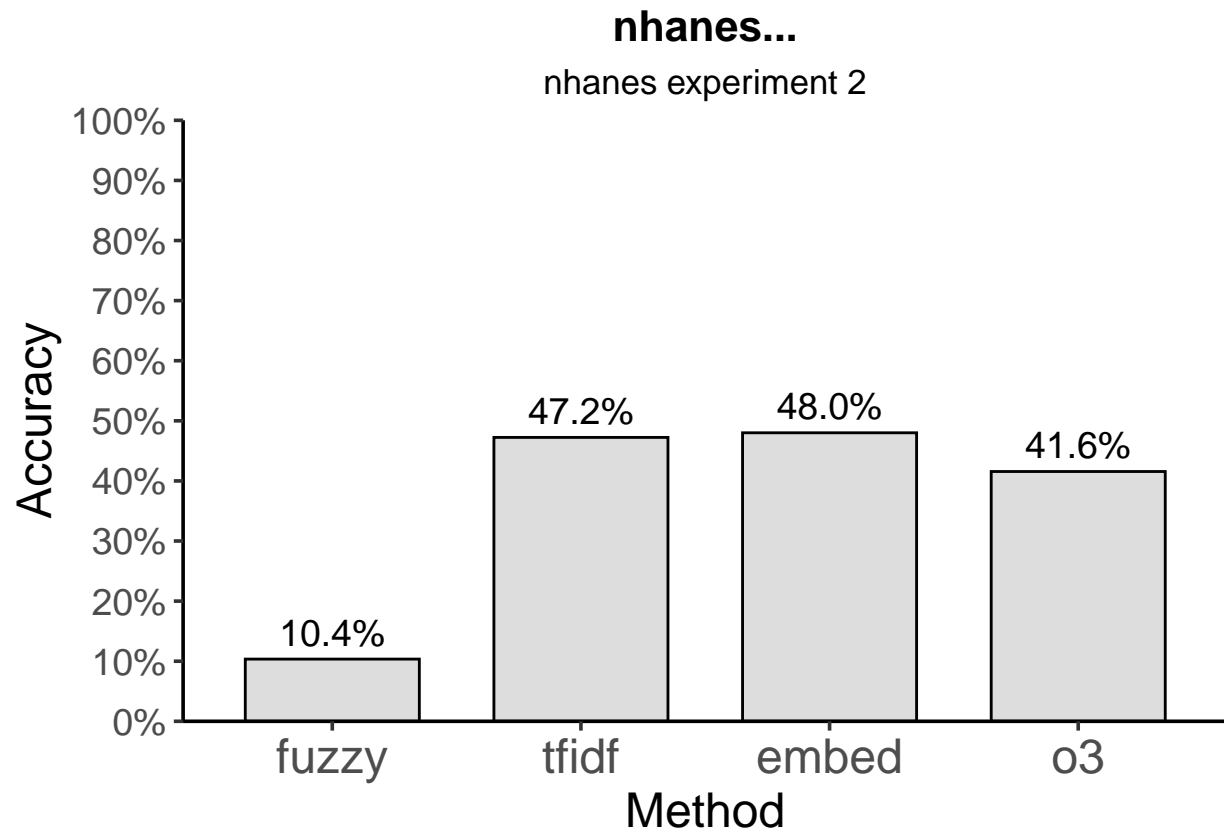
ASA24 Mapping to FooDB

asa24 experiment 2

NHANES Experiment 1

```
plot_accuracy(
  filename = "nhanes_experiment_1_accuracy.csv",
  title_text = "nhanes...",
  subtitle_text = "nhanes experiment 1"
)
```

nhanes...
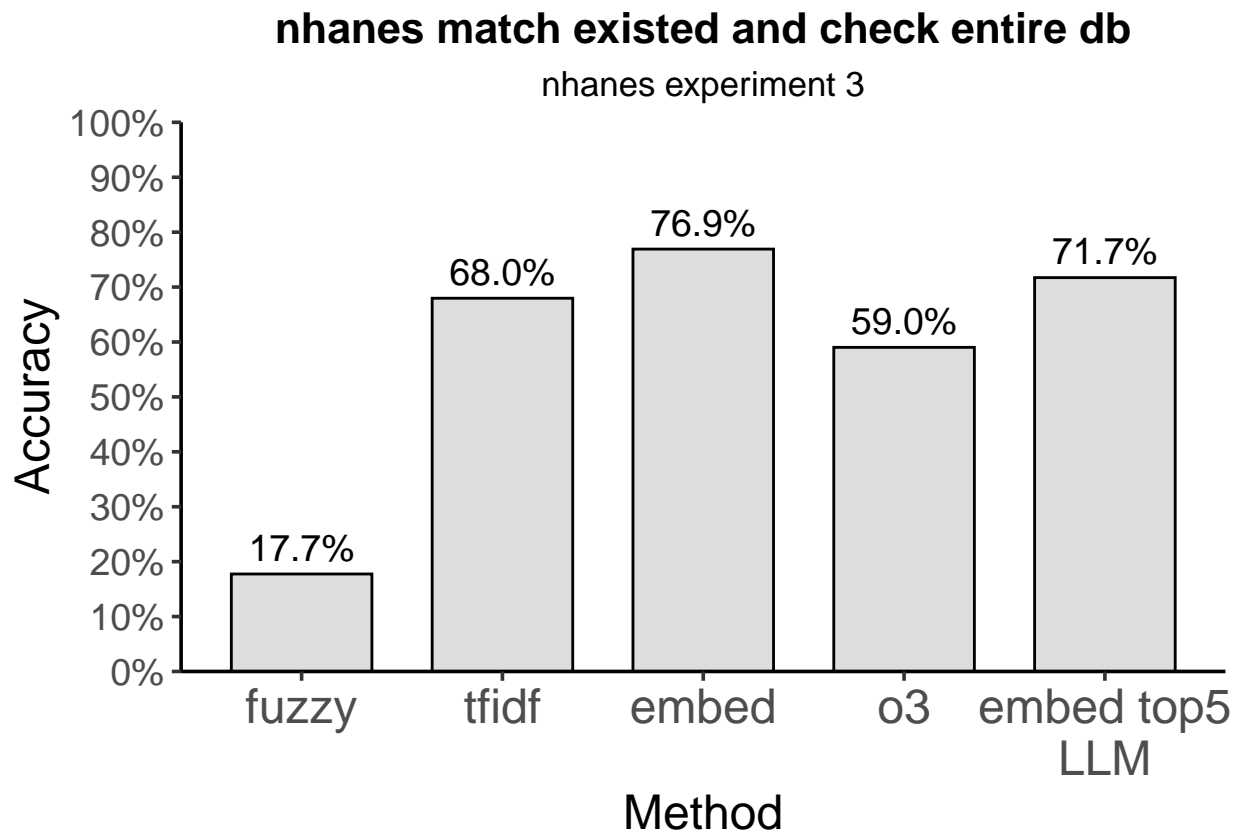
nhanes experiment 1

NHANES Experiment 2

```
plot_accuracy(
  filename = "nhanes_experiment_2_accuracy.csv",
  title_text = "nhanes...",
  subtitle_text = "nhanes experiment 2"
)
```

nhanes...

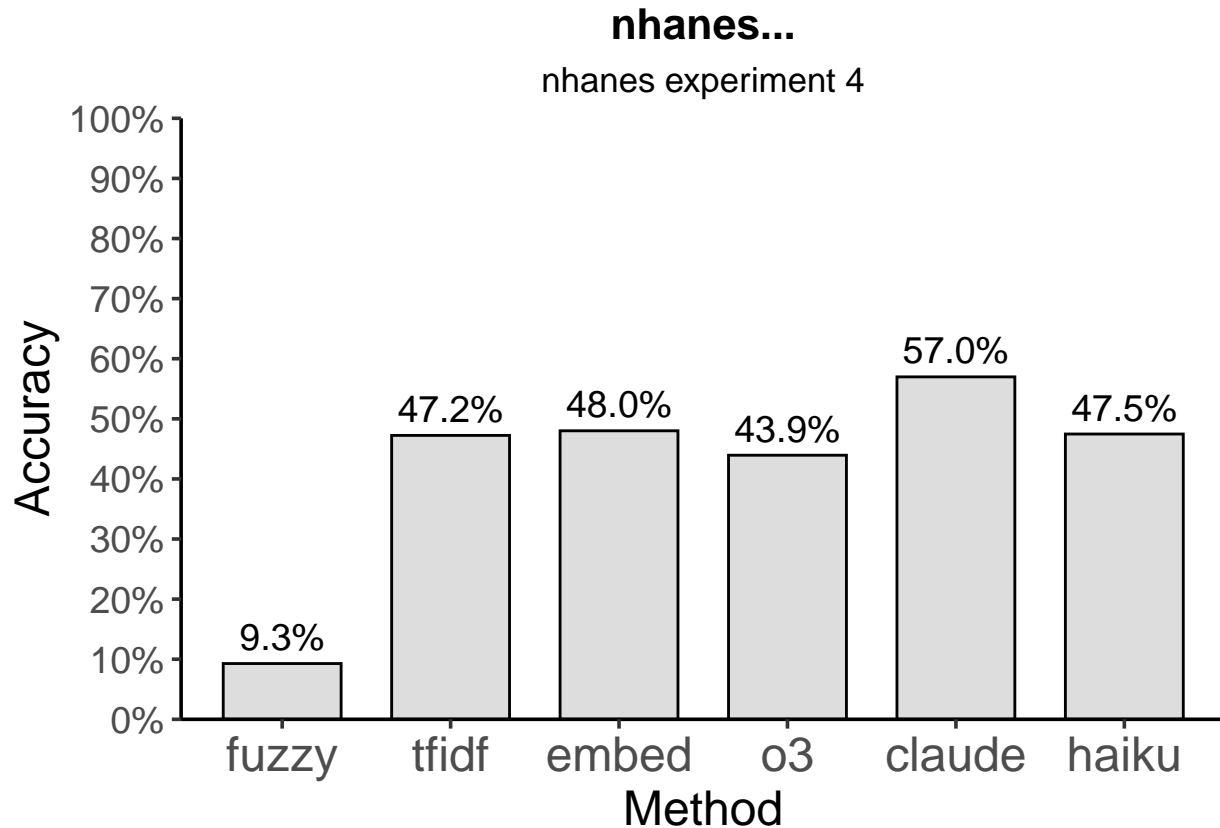nhanes experiment 2

NHANES Experiment 3

```
plot_accuracy(
  filename = "nhanes_experiment_3_accuracy.csv",
  title_text = "nhanes match existed and check entire db",
  subtitle_text = "nhanes experiment 3"
)
```

nhanes match existed and check entire db

nhanes experiment 3

ASA Experiment 4

```
plot_accuracy(
  filename = "nhanes_experiment_4_accuracy.csv",
  title_text = "nhanes...",
  subtitle_text = "nhanes experiment 4"
)
```

nhanes...

nhanes experiment 4

overlayed histogram of

```r
library(ggplot2)
library(readr)
library(dplyr)

asa24_path <- file.path("..", "results", "csv_files", "asa24_experiment_2.csv")
nhanes_path <- file.path("..", "results", "csv_files", "nhanes_experiment_4.csv")

asa24 <- read_csv(asa24_path) %>%
  filter(!is.na(score_tfidf), score_tfidf >= 0, score_tfidf <= 1) %>%
  mutate(Source = "ASA24 to FooDB")
```

```
## Rows: 1198 Columns: 11
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr (7): input_desc, target_desc, match_fuzzy, match_tfidf, match_embed, mat...
## dbl (4): index, score_fuzzy, score_tfidf, score_embed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
nhanes <- read_csv(nhanes_path) %>%
  filter(!is.na(score_tfidf), score_tfidf >= 0, score_tfidf <= 1) %>%
  mutate(Source = "AMPM to Davis Food Glycopedia")
```
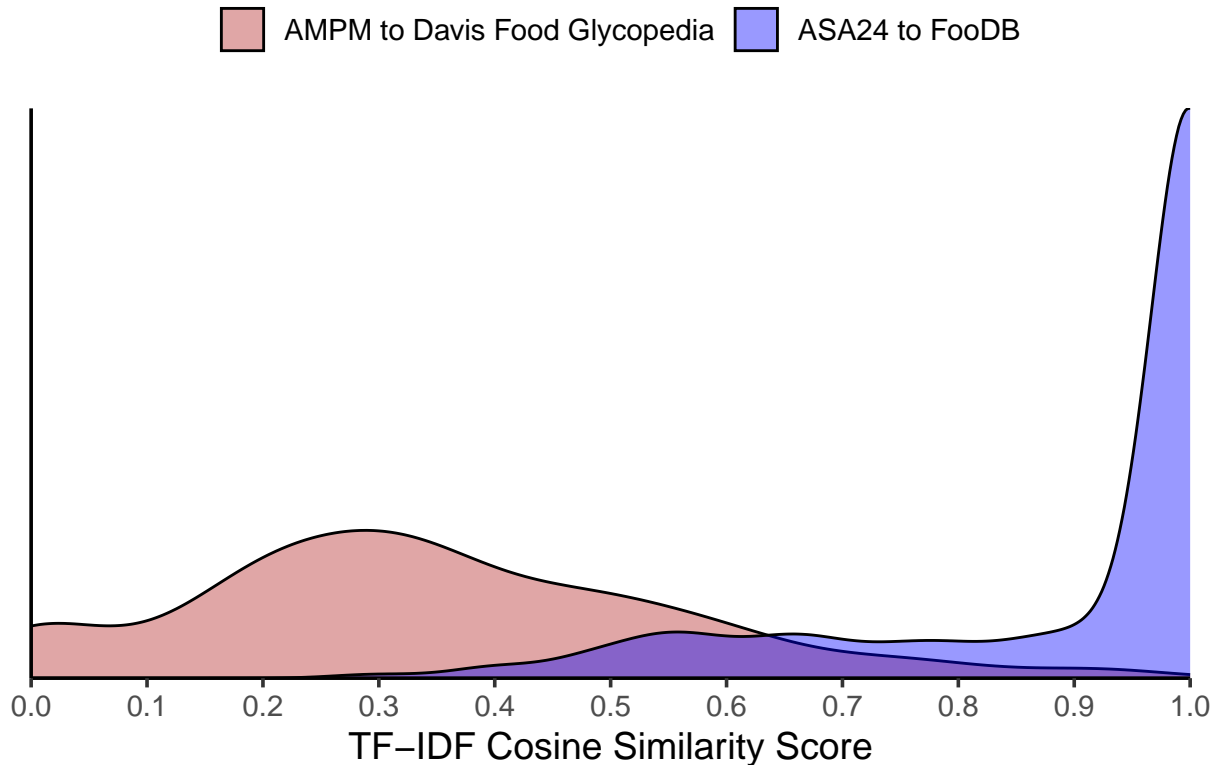
```
## Rows: 1304 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (8): input_desc, target_desc, match_fuzzy, match_tfidf, match_embed, mat...
## dbl (5): label, index, score_fuzzy, score_tfidf, score_embed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
combined <- bind_rows(asa24, nhanes)

# geom_density
ggplot(combined, aes(x = score_tfidf, fill = Source)) +
  geom_density(alpha = 0.4, adjust = 1.25) +
  scale_fill_manual(
    name = "",
    values = c(
      "ASA24 to FooDB" = "blue",
      "AMPM to Davis Food Glycopedia" = "firebrick"
    )
  ) +
  scale_x_continuous(limits = c(0, 1), breaks = seq(0, 1, 0.1), expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  labs(
    title = "Overlayed Similarity Score Density Plot",
    x = "TF-IDF Cosine Similarity Score",
    y = NULL
  ) +
  theme_classic(base_size = 14) +
  theme(
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank(),
    legend.position = "top",
    plot.margin = margin(5, 15, 5, 15)
  )
```

Overlayed Similarity Score Density Plot


```r
library(ggplot2)
library(readr)
library(dplyr)

asa24_path <- file.path("..", "results", "csv_files", "asa24_experiment_2.csv")
nhanes_path <- file.path("..", "results", "csv_files", "nhanes_experiment_4.csv")

asa24 <- read_csv(asa24_path) %>%
  filter(!is.na(score_embed), score_embed >= 0, score_embed <= 1) %>%
  mutate(Source = "ASA24 to FooDB")
```

```
## Rows: 1198 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (7): input_desc, target_desc, match_fuzzy, match_tfidf, match_embed, mat...
## dbl (4): index, score_fuzzy, score_tfidf, score_embed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
nhanes <- read_csv(nhanes_path) %>%
  filter(!is.na(score_embed), score_embed >= 0, score_embed <= 1) %>%
  mutate(Source = "AMPM to Davis Food Glycopedia")
```

```
## Rows: 1304 Columns: 13
```

```
## -- Column specification ---------------------------------------------------------
## Delimiter: ","
## chr (8): input_desc, target_desc, match_fuzzy, match_tfidf, match_embed, mat...
## dbl (5): label, index, score_fuzzy, score_tfidf, score_embed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
combined <- bind_rows(asa24, nhanes)

# geom_density
ggplot(combined, aes(x = score_embed, fill = Source)) +
  geom_density(alpha = 0.4, adjust = 1.25) +
  scale_fill_manual(
    name = "",
    values = c(
      "ASA24 to FooDB" = "blue",
      "AMPM to Davis Food Glycopedia" = "firebrick"
    )
  ) +
  scale_x_continuous(limits = c(0, 1), breaks = seq(0, 1, 0.1), expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  labs(
    title = "Overlayed Similarity Score Density Plot",
    x = "gte-large Cosine Similarity Score",
    y = NULL
  ) +
  theme_classic(base_size = 14) +
  theme(
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank(),
    legend.position = "top",
    plot.margin = margin(5, 15, 5, 15)
  )
```

Overlayed Similarity Score Density Plot

gte−large Cosine Similarity Score

Legend: AMPM to Davis Food Glycopedia, ASA24 to FooDB