Visualizing Results of Individual Metrics

Ishana Rana

2025-07-10

1) Loading the datasets

```
bertscore <- read.csv("C:\\Users\\babus\\OneDrive\\Documents\\uni
uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-
ASR\\data\\eval_results\\BERTScore_scores.csv", skip=1)
str(bertscore)
```

```
## 'data.frame':    6 obs. of  3 variables:
##  $ File         : chr  "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...
##  $ with_punct   : num  84.7 88 84 91.1 94 ...
##  $ without_punct: num  86 89.3 86 93.7 95.8 ...
```

```
head(bertscore)
```

```
##      File with_punct without_punct
## 1 EN2009c      84.66         85.98
## 2 EN2009d      87.98         89.34
## 3 ES2016a      84.01         86.02
## 4 ES2016b      91.10         93.74
## 5 ES2016c      93.98         95.83
## 6 ES2016d      89.66         90.12
```

```
summary(bertscore)
```

```
##      File             with_punct    without_punct
##  Length:6           Min.   :84.01   Min.   :85.98
##  Class :character   1st Qu.:85.49   1st Qu.:86.85
##  Mode  :character   Median :88.82   Median :89.73
##                     Mean   :88.56   Mean   :90.17
##                     3rd Qu.:90.74   3rd Qu.:92.83
##                     Max.   :93.98   Max.   :95.83
```

```
bleu <- read.csv("C:\\Users\\babus\\OneDrive\\Documents\\uni
uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-
ASR\\data\\eval_results\\BLEU_scores.csv")
str(bleu)
```

```
## 'data.frame':    6 obs. of  3 variables:
##  $ File         : chr  "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...
##  $ with_punct   : num  38.8 36.9 43.6 54.3 50 ...
##  $ without_punct: num  38.1 38.5 49.1 59.4 55.1 ...
```

```r
head(bleu)
```

```
##      File with_punct without_punct
## 1 EN2009c      38.77         38.14
## 2 EN2009d      36.93         38.54
## 3 ES2016a      43.61         49.07
## 4 ES2016b      54.26         59.40
## 5 ES2016c      49.96         55.11
## 6 ES2016d      30.65         31.81
```

```r
summary(bleu)
```

```
##      File             with_punct    without_punct
##  Length:6           Min.   :30.65   Min.   :31.81
##  Class :character   1st Qu.:37.39   1st Qu.:38.24
##  Mode  :character   Median :41.19   Median :43.80
##                     Mean   :42.36   Mean   :45.34
##                     3rd Qu.:48.37   3rd Qu.:53.60
##                     Max.   :54.26   Max.   :59.40
```

```r
# splitting dataset as it contains scores of ROUGE-1 and ROUGE-L
lines <- readLines("C:\\Users\\babus\\OneDrive\\Documents\\uni
uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-
ASR\\data\\eval_results\\ROUGE_scores.csv")
split_index <- grep("ROUGE-L", lines)

rouge1_lines <- lines[2:(split_index - 1)]
rougel_lines <- lines[(split_index + 1):length(lines)]

rouge1 <- read.csv(text = rouge1_lines)
rougel <- read.csv(text = rougel_lines)

str(rouge1)
```

```
## 'data.frame':    6 obs. of  3 variables:
##  $ File         : chr  "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...
##  $ with_punct   : num  74.9 72.9 80.1 85 83 ...
##  $ without_punct: num  74.8 73 80 84.9 82.8 ...
```

```r
head(rouge1)
```

```
##      File with_punct without_punct
## 1 EN2009c      74.92         74.76
## 2 EN2009d      72.91         72.97
## 3 ES2016a      80.14         80.04
## 4 ES2016b      84.99         84.93
## 5 ES2016c      83.00         82.82
## 6 ES2016d      70.03         69.89
```

```r
summary(rouge1)
```

```
##      File             with_punct    without_punct
##  Length:6           Min.   :70.03   Min.   :69.89
##  Class :character   1st Qu.:73.41   1st Qu.:73.42
##  Mode  :character   Median :77.53   Median :77.40
##                     Mean   :77.67   Mean   :77.57
##                     3rd Qu.:82.28   3rd Qu.:82.12
##                     Max.   :84.99   Max.   :84.93
```

```
str(rougel)
```

```
## 'data.frame':    6 obs. of  3 variables:
##  $ File         : chr  "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...
##  $ with_punct   : num  39.4 35.1 39.4 37.2 46.9 ...
##  $ without_punct: num  39.2 35.1 39.3 37.2 46.8 ...
```

```
head(rougel)
```

```
##      File with_punct without_punct
## 1 EN2009c      39.35         39.22
## 2 EN2009d      35.10         35.10
## 3 ES2016a      39.37         39.28
## 4 ES2016b      37.24         37.22
## 5 ES2016c      46.86         46.83
## 6 ES2016d      31.62         31.30
```

```
summary(rougel)
```

```
##      File             with_punct    without_punct
##  Length:6           Min.   :31.62   Min.   :31.30
##  Class :character   1st Qu.:35.63   1st Qu.:35.63
##  Mode  :character   Median :38.30   Median :38.22
##                     Mean   :38.26   Mean   :38.16
##                     3rd Qu.:39.37   3rd Qu.:39.27
##                     Max.   :46.86   Max.   :46.83
```

```
wer <- read.csv("C:\\Users\\babus\\OneDrive\\Documents\\uni
uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-
ASR\\data\\eval_results\\WER_scores.csv", skip=1)
str(wer)
```

```
## 'data.frame':    6 obs. of  3 variables:
##  $ File         : chr  "EN2009c" "EN2009d" "ES2016a" "ES2016b" ...
##  $ with_punct   : num  89.1 89.9 88.7 82.9 77 ...
##  $ without_punct: num  86.2 87.5 86.4 78.3 71.6 ...
```

```
head(wer)
```

```
##      File with_punct without_punct
## 1 EN2009c      89.12         86.16
## 2 EN2009d      89.94         87.55
## 3 ES2016a      88.73         86.38
## 4 ES2016b      82.87         78.33
```

```
## 5 ES2016c      77.02         71.61
## 6 ES2016d      86.95         83.21
```

```
summary(wer)
```

```
##      File             with_punct    without_punct
##  Length:6           Min.   :77.02   Min.   :71.61
##  Class :character   1st Qu.:83.89   1st Qu.:79.55
##  Mode  :character   Median :87.84   Median :84.69
##                     Mean   :85.77   Mean   :82.21
##                     3rd Qu.:89.02   3rd Qu.:86.33
##                     Max.   :89.94   Max.   :87.55
```

2) Individual Analysis of each metric

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
library(tidyr)
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
```

```
#install.packages("patchwork")
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.4.3
```

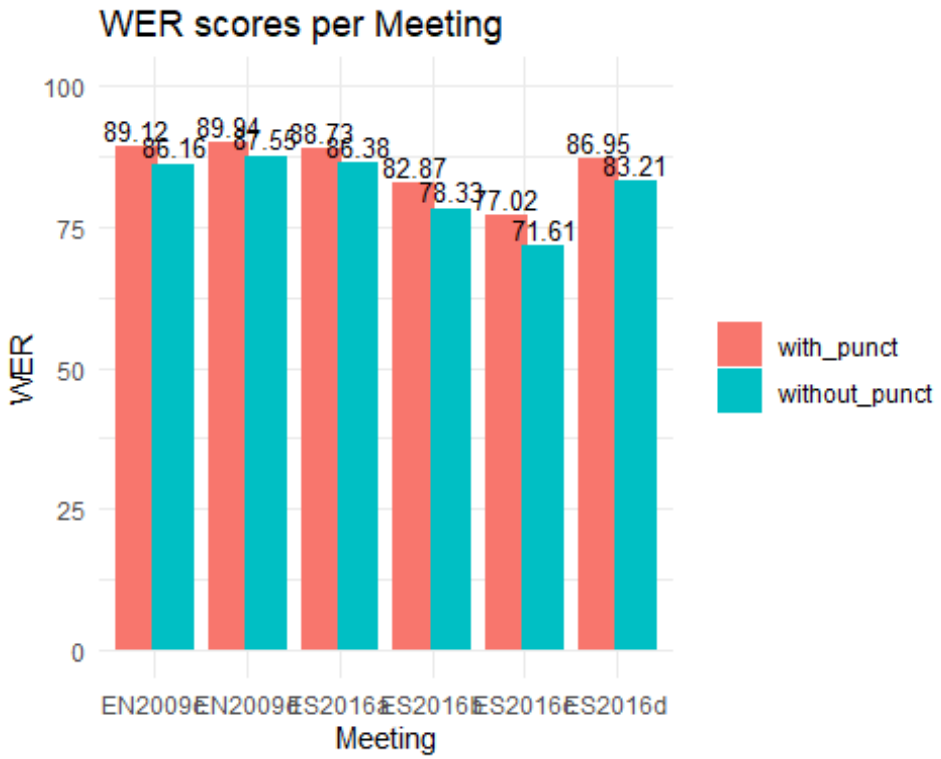
```
# specifying directory where the created plots shall be saved at
plot_dir = "C:\\Users\\babus\\OneDrive\\Documents\\uni
uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-
ASR\\data\\eval_results\\Plots"
dir.create(plot_dir, showWarnings = FALSE)
```

Grouped Bar Chart

```
# WER
wer_long <- pivot_longer(
  wer,
  cols = c(with_punct, without_punct),
  names_to = "Condition",
  values_to = "Score"
)

plot_wer <- ggplot(wer_long, aes(x = File, y = Score, fill = Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  geom_text(
    aes(label = sprintf("%.2f", Score)),
    position = position_dodge(width = 0.8),
    vjust = -0.3,
    size = 3.5
  ) +
```

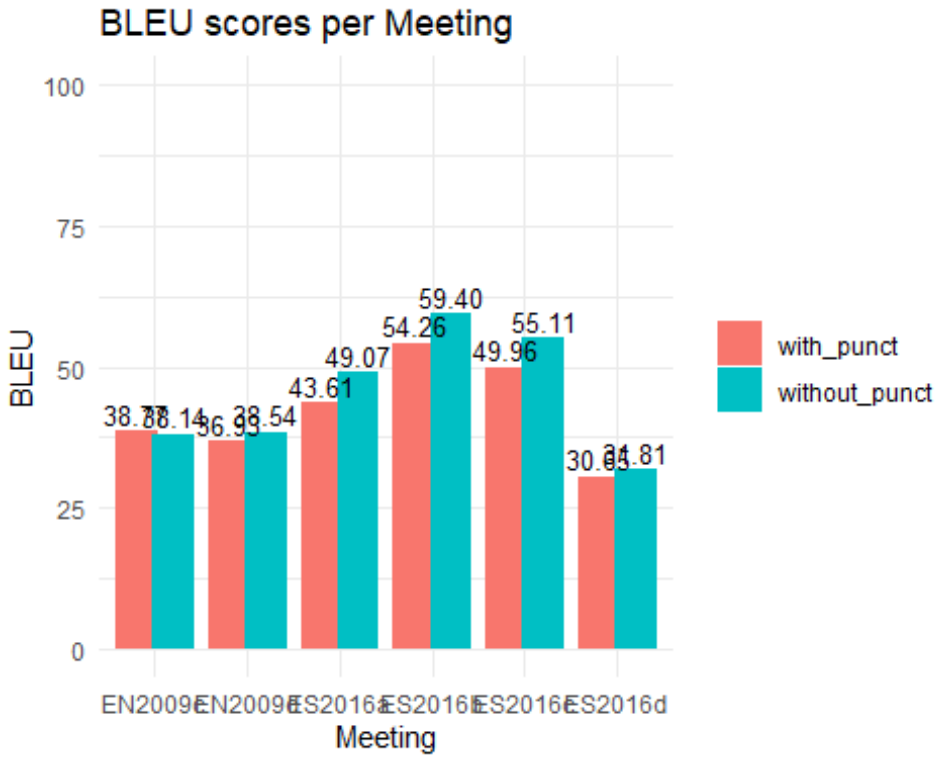
```
  labs(
    title = "WER scores per Meeting",
    y = "WER",
    x = "Meeting"
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())
ggsave(file.path(plot_dir, "wer_barchart.png"), plot_wer, width = 6, height =
4, dpi = 300)
plot_wer
```




```
# BLEU
bleu_long <- pivot_longer(
  bleu,
  cols = c(with_punct, without_punct),
  names_to = "Condition",
  values_to = "Score"
)

plot_bleu<- ggplot(bleu_long, aes(x = File, y = Score, fill = Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  geom_text(
    aes(label = sprintf("%.2f", Score)),
    position = position_dodge(width = 0.8),
    vjust = -0.3,
    size = 3.5
```

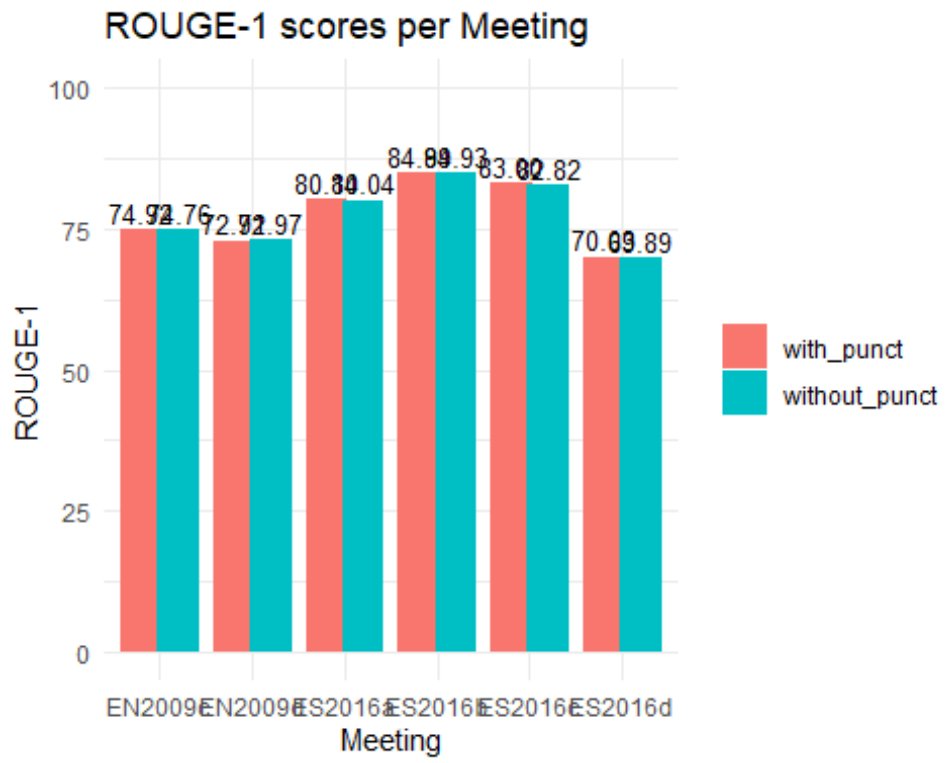
```
  ) +
  labs(
    title = "BLEU scores per Meeting",
    y = "BLEU",
    x = "Meeting"
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())
ggsave(file.path(plot_dir, "bleue_barchart.png"), plot_bleu, width = 6,
height = 4, dpi = 300)
plot_bleu
```




```
# ROUGE-1
rouge_long_1 <- pivot_longer(
  rouge1,
  cols = c(with_punct, without_punct),
  names_to = "Condition",
  values_to = "Score"
)

plot_rouge1 <- ggplot(rouge_long_1, aes(x = File, y = Score, fill =
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  geom_text(
    aes(label = sprintf("%.2f", Score)),
    position = position_dodge(width = 0.8),
```

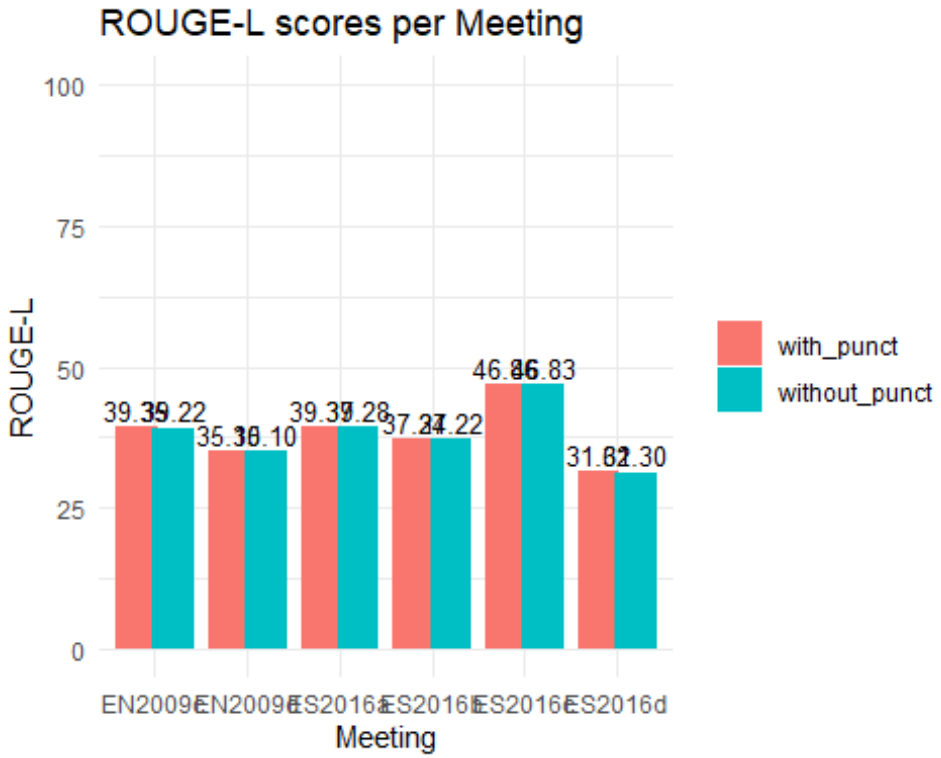
```
    vjust = -0.3,
    size = 3.5
  ) +
  labs(
    title = "ROUGE-1 scores per Meeting",
    y = "ROUGE-1",
    x = "Meeting"
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())
ggsave(file.path(plot_dir, "rouge1_barchart.png"), plot_rouge1, width = 6,
height = 4, dpi = 300)
plot_rouge1
```




```
# ROUGE-L
rouge_long_l <- pivot_longer(
  rougel,
  cols = c(with_punct, without_punct),
  names_to = "Condition",
  values_to = "Score"
)

plot_rougel <- ggplot(rouge_long_l, aes(x = File, y = Score, fill =
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  geom_text(
```

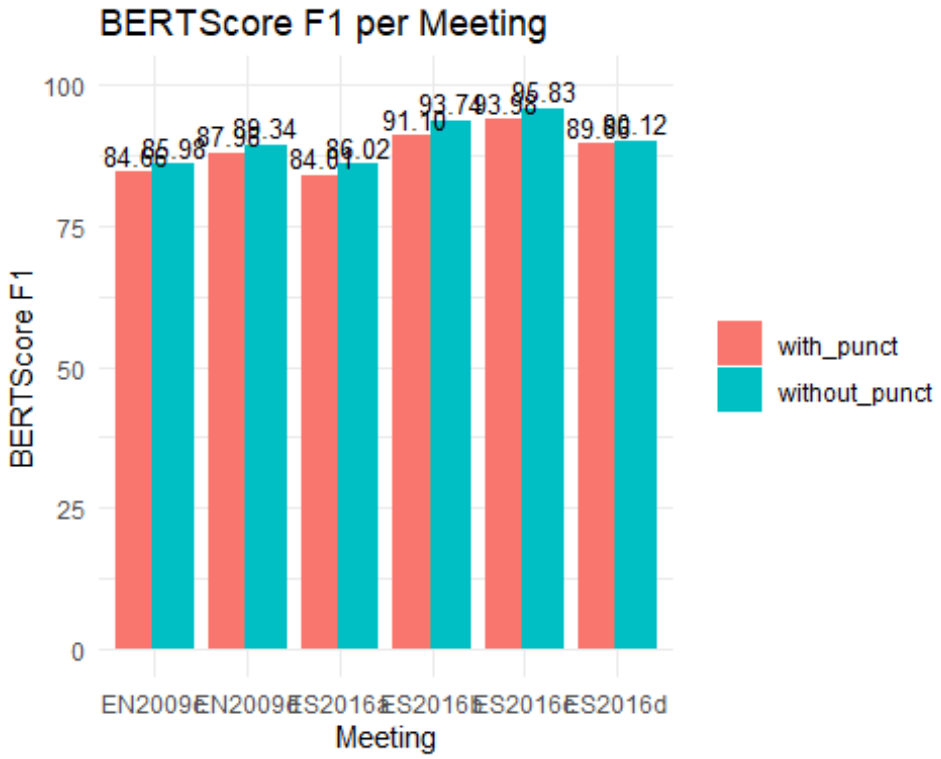
```
    aes(label = sprintf("%.2f", Score)),
    position = position_dodge(width = 0.8),
    vjust = -0.3,
    size = 3.5
  ) +
  labs(
    title = "ROUGE-L scores per Meeting",
    y = "ROUGE-L",
    x = "Meeting"
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())
ggsave(file.path(plot_dir, "rougel_barchart.png"), plot_rougel, width = 6,
height = 4, dpi = 300)
plot_rougel
```




```
# BERTScore
bertscore_long <- pivot_longer(
  bertscore,
  cols = c(with_punct, without_punct),
  names_to = "Condition",
  values_to = "Score"
)

# Plot with value labels
plot_bertscore <- ggplot(bertscore_long, aes(x = File, y = Score, fill =
```

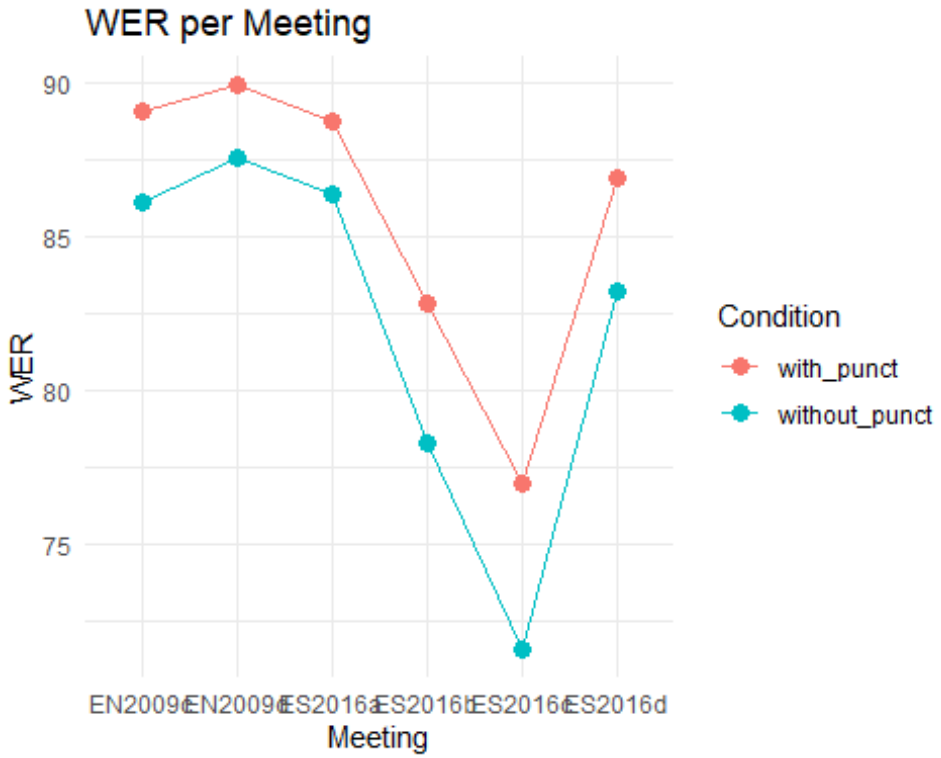
```
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  geom_text(
    aes(label = sprintf("%.2f", Score)),
    position = position_dodge(width = 0.8),
    vjust = -0.3,
    size = 3.5
  ) +
  labs(
    title = "BERTScore F1 per Meeting",
    y = "BERTScore F1",
    x = "Meeting"
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())
ggsave(file.path(plot_dir, "bertscore_barchart.png"), plot_bertscore, width =
6, height = 4, dpi = 300)
plot_bertscore
```



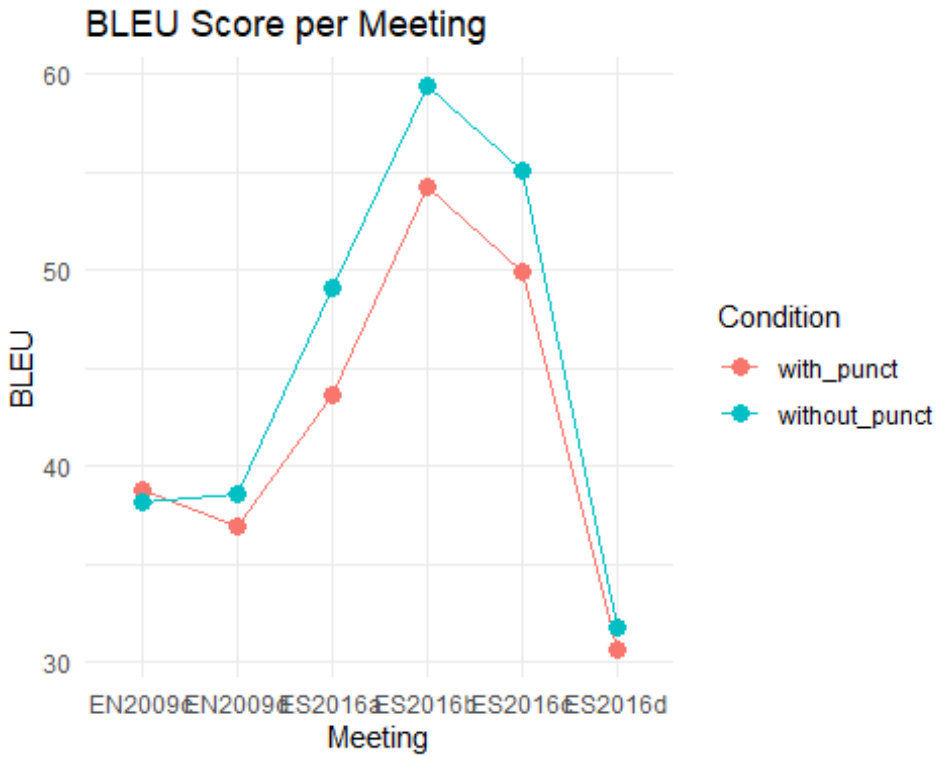
Line Plots

```
line_wer <- ggplot(wer_long, aes(x = File, y = Score, group = Condition,
color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
```

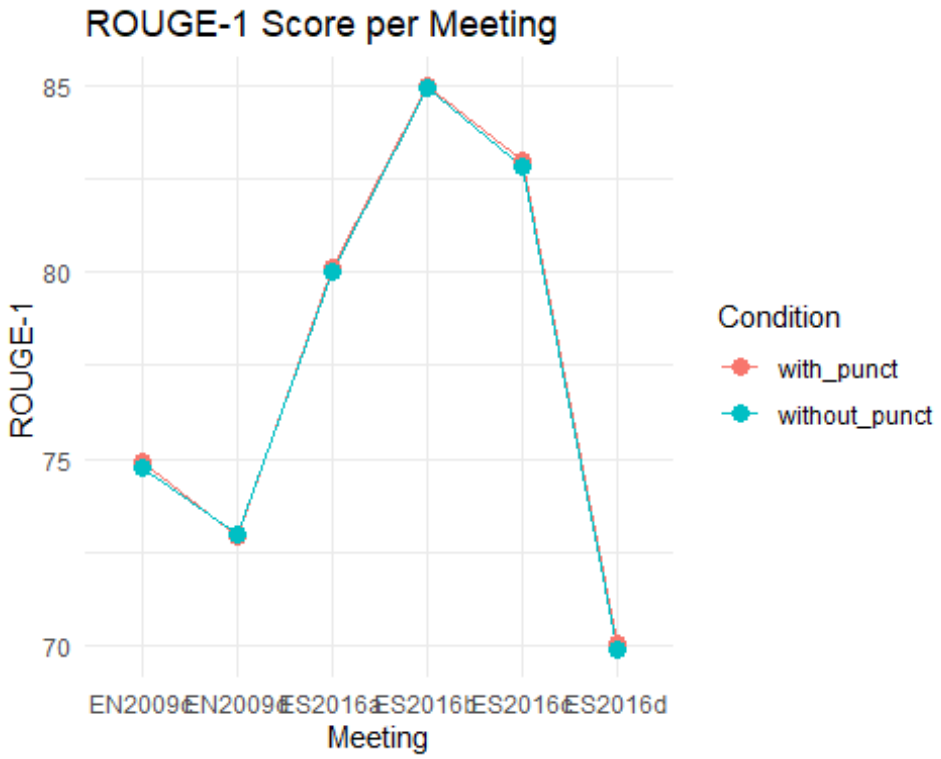
```
  labs(title = "WER per Meeting", y = "WER", x = "Meeting") +
  theme_minimal()
ggsave(file.path(plot_dir, "wer_lineplot.png"), line_wer, width = 6, height =
4, dpi = 300)
line_wer
```



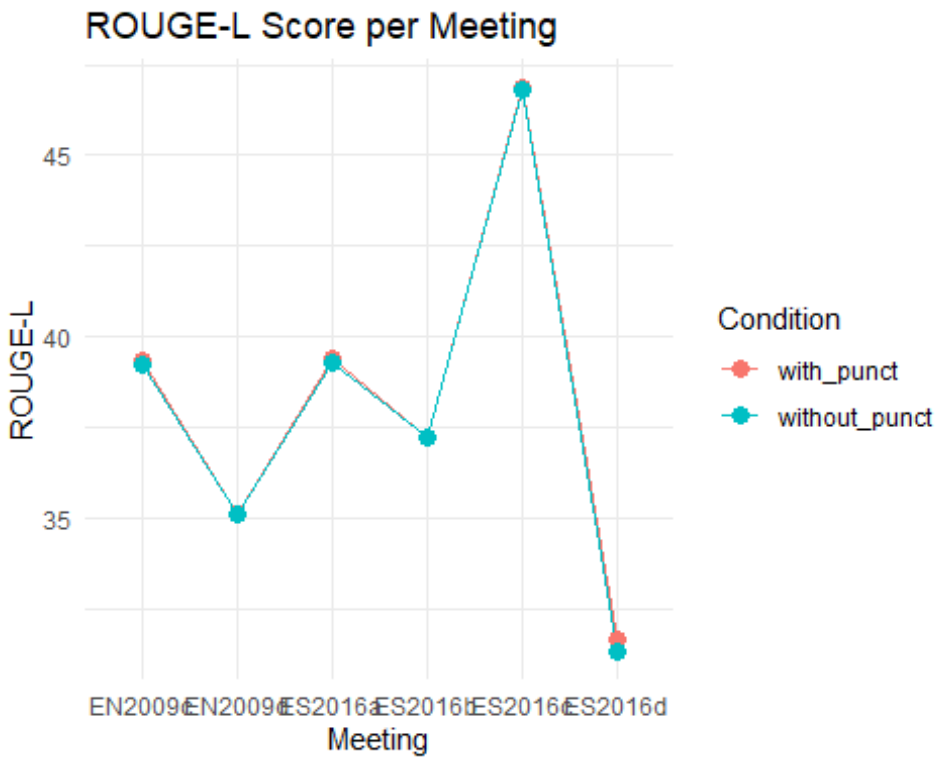

```
line_bleu <- ggplot(bleu_long, aes(x = File, y = Score, group = Condition,
color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "BLEU Score per Meeting", y = "BLEU", x = "Meeting") +
  theme_minimal()
ggsave(file.path(plot_dir, "bleu_lineplot.png"), line_bleu, width = 6, height
= 4, dpi = 300)
line_bleu
```



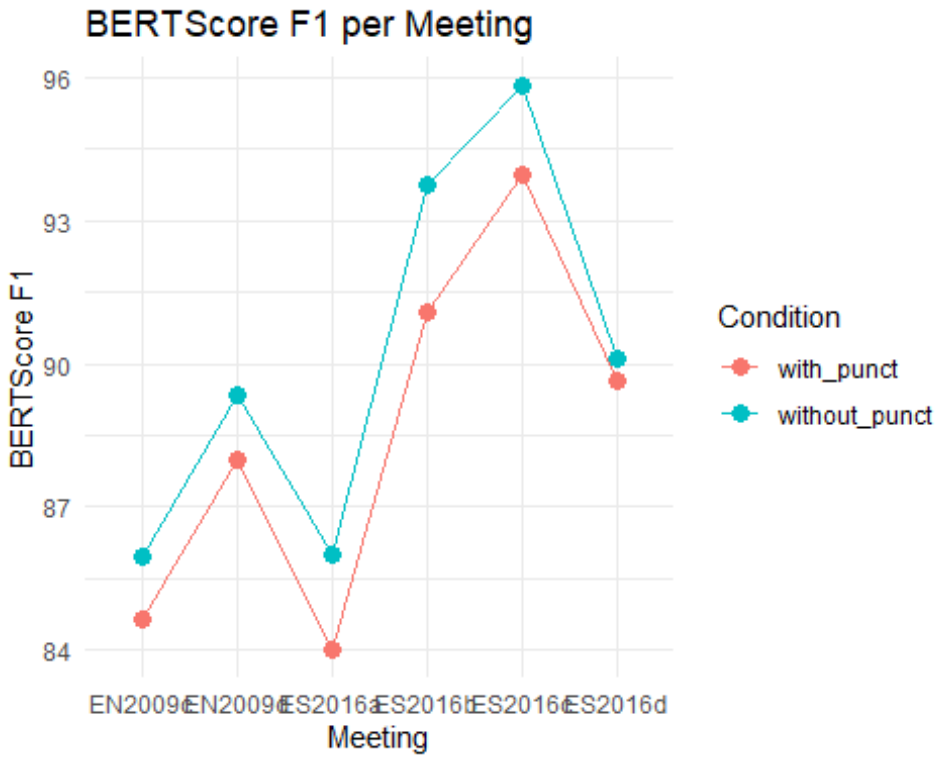

```
line_rouge1 <- ggplot(rouge_long_1, aes(x = File, y = Score, group =
Condition, color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "ROUGE-1 Score per Meeting", y = "ROUGE-1", x = "Meeting") +
  theme_minimal()
ggsave(file.path(plot_dir, "rouge1_lineplot.png"), line_rouge1, width = 6,
height = 4, dpi = 300)
line_rouge1
```




```
line_rougel <- ggplot(rouge_long_l, aes(x = File, y = Score, group =
Condition, color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "ROUGE-L Score per Meeting", y = "ROUGE-L", x = "Meeting") +
  theme_minimal()
ggsave(file.path(plot_dir, "rougel_lineplot.png"), line_rougel, width = 6,
height = 4, dpi = 300)
line_rougel
```




```
line_bertscore <- ggplot(bertscore_long, aes(x = File, y = Score, group =
Condition, color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "BERTScore F1 per Meeting", y = "BERTScore F1", x = "Meeting")
+
  theme_minimal()
ggsave(file.path(plot_dir, "bertscore_lineplot.png"), line_bertscore, width =
6, height = 4, dpi = 300)
line_bertscore
```



Metrics in 1 image

```r
library(stringr)

wer_long_short <- wer_long %>%
  mutate(File = paste0(str_sub(File, 2, 2), "-", str_sub(File, -1, -1)))

bleu_long_short <- bleu_long %>%
  mutate(File = paste0(str_sub(File, 2, 2), "-", str_sub(File, -1, -1)))

rouge_long_1_short <- rouge_long_1 %>%
  mutate(File = paste0(str_sub(File, 2, 2), "-", str_sub(File, -1, -1)))

rouge_long_l_short <- rouge_long_l %>%
  mutate(File = paste0(str_sub(File, 2, 2), "-", str_sub(File, -1, -1)))

bertscore_long_short <- bertscore_long %>%
  mutate(File = paste0(str_sub(File, 2, 2), "-", str_sub(File, -1, -1)))

# WER

p_wer <- ggplot(wer_long_short, aes(x = File, y = Score, group = Condition,
color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "WER per Meeting", y = "WER", x = "") +
```

```r
  theme_minimal()

# BLEU
p_bleu <- ggplot(bleu_long_short, aes(x = File, y = Score, group = Condition,
color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "BLEU per Meeting", y = "BLEU", x = "") +
  theme_minimal()

# ROUGE-1
p_rouge1 <- ggplot(rouge_long_1_short, aes(x = File, y = Score, group =
Condition, color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "ROUGE-1 per Meeting", y = "ROUGE-1", x = "") +
  theme_minimal()

# ROUGE-L
p_rougel <- ggplot(rouge_long_l_short, aes(x = File, y = Score, group =
Condition, color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "ROUGE-L per Meeting", y = "ROUGE-l", x = "") +
  theme_minimal()

p_bertscore <- ggplot(bertscore_long_short, aes(x = File, y = Score, group =
Condition, color = Condition)) +
  geom_line() +
  geom_point(size = 3) +
  labs(title = "BERTScore per Meeting", y = "BERTScore", x = "") +
  theme_minimal()

plot_grid_4 <- (p_wer | p_bleu) /
               (p_rouge1 | p_rougel)
ggsave(file.path(plot_dir, "superficial_metrics.png"), plot_grid_4, width =
10, height = 6, dpi = 300)

plot_grid_5 <- (p_wer | p_bleu) /
               (p_rouge1 | p_rougel) /
               (p_bertscore | plot_spacer())
ggsave(file.path(plot_dir, "all_metrics.png"), plot_grid_5, width = 10,
height = 6, dpi = 300)

plot_wer_s <- ggplot(wer_long_short, aes(x = File, y = Score, fill =
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(
    title = "WER scores per Meeting",
    y = "WER",
```

```
    x = ""
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())

plot_bleu_s<- ggplot(bleu_long_short, aes(x = File, y = Score, fill =
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(
    title = "BLEU scores per Meeting",
    y = "BLEU",
    x = ""
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())

plot_rouge1_s <- ggplot(rouge_long_1_short, aes(x = File, y = Score, fill =
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(
    title = "ROUGE-1 scores per Meeting",
    y = "ROUGE-1",
    x = "Meeting"
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())

plot_rougel_s <- ggplot(rouge_long_l_short, aes(x = File, y = Score, fill =
Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(
    title = "ROUGE-L scores per Meeting",
    y = "ROUGE-L",
    x = ""
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())

plot_bertscore_s <- ggplot(bertscore_long_short, aes(x = File, y = Score,
fill = Condition)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8)) +
  labs(
    title = "BERTScore F1 per Meeting",
    y = "BERTScore F1",
    x = ""
```

```r
  ) +
  ylim(0, 100) +  # ensures all labels fit
  theme_minimal() +
  theme(legend.title = element_blank())

plot_bar_4 <- (plot_wer_s | plot_bleu_s) /
               (plot_rouge1_s | plot_rougel_s)
ggsave(file.path(plot_dir, "bar_superficial_metrics.png"), plot_bar_4, width
= 10, height = 6, dpi = 300)

plot_bar_5 <- (plot_wer_s | plot_bleu_s) /
               (plot_rouge1_s | plot_rougel_s) /
               (plot_bertscore_s | plot_spacer())
ggsave(file.path(plot_dir, "bar_all_metrics.png"), plot_bar_5, width = 10,
height = 6, dpi = 300)
```