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 ...

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

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

# 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 ...

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

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

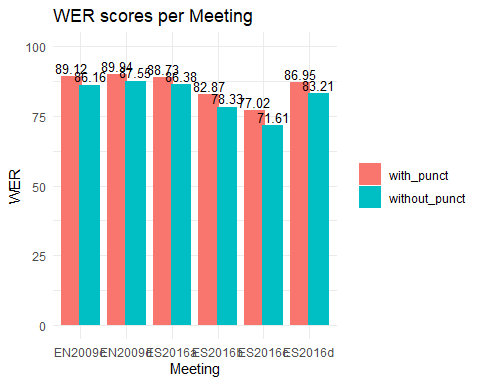
#install.packages("patchwork")  
library(patchwork)

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

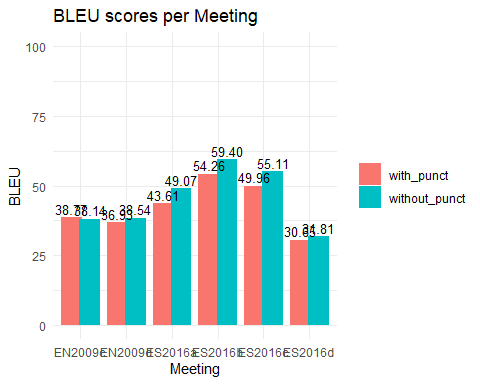
# specifying directory where the created plots shall be saved at  
plot\_dir = "C:\\Users\\babus\\OneDrive\\Documents\\uni uzh\\FS25\\conversational speech processing\\mypaper\\Beyond-WER-in-ASR\\data\\eval\_results\\Plots"  
dir.create(plot\_dir, showWarnings = FALSE)

Grouped Bar Chart

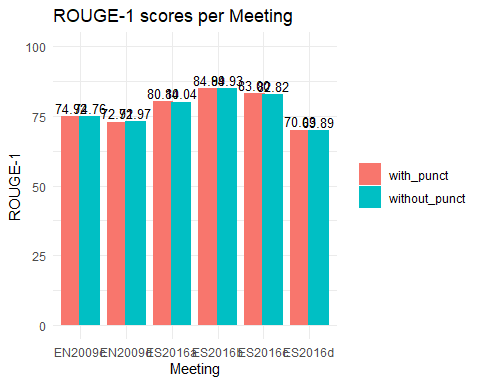
# WER  
wer\_long <- pivot\_longer(  
 wer,  
 cols = c(with\_punct, without\_punct),  
 names\_to = "Condition",  
 values\_to = "Score"  
)  
  
plot\_wer <- ggplot(wer\_long, aes(x = File, y = Score, fill = Condition)) +  
 geom\_bar(stat = "identity", position = position\_dodge(width = 0.8)) +  
 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



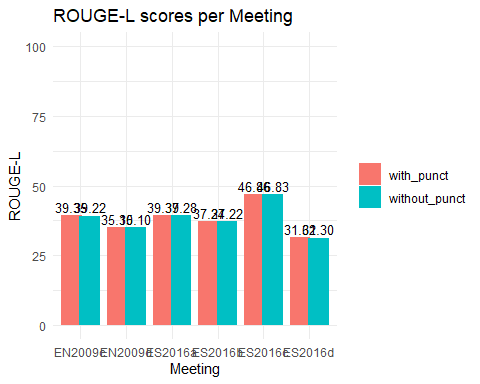
# BLEU  
bleu\_long <- pivot\_longer(  
 bleu,  
 cols = c(with\_punct, without\_punct),  
 names\_to = "Condition",  
 values\_to = "Score"  
)  
  
plot\_bleu<- ggplot(bleu\_long, aes(x = File, y = Score, fill = Condition)) +  
 geom\_bar(stat = "identity", position = position\_dodge(width = 0.8)) +  
 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



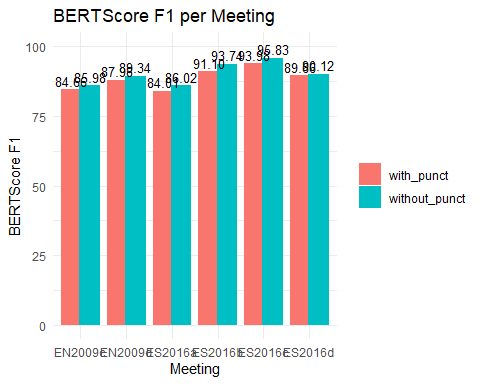
# ROUGE-1  
rouge\_long\_1 <- pivot\_longer(  
 rouge1,  
 cols = c(with\_punct, without\_punct),  
 names\_to = "Condition",  
 values\_to = "Score"  
)  
  
plot\_rouge1 <- ggplot(rouge\_long\_1, aes(x = File, y = Score, fill = Condition)) +  
 geom\_bar(stat = "identity", position = position\_dodge(width = 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

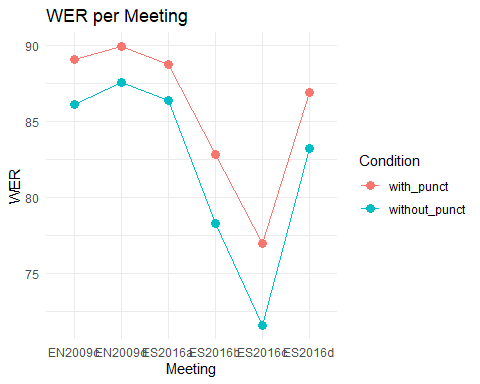


# BERTScore  
bertscore\_long <- pivot\_longer(  
 bertscore,  
 cols = c(with\_punct, without\_punct),  
 names\_to = "Condition",  
 values\_to = "Score"  
)  
  
# Plot with value labels  
plot\_bertscore <- ggplot(bertscore\_long, aes(x = File, y = Score, fill = Condition)) +  
 geom\_bar(stat = "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

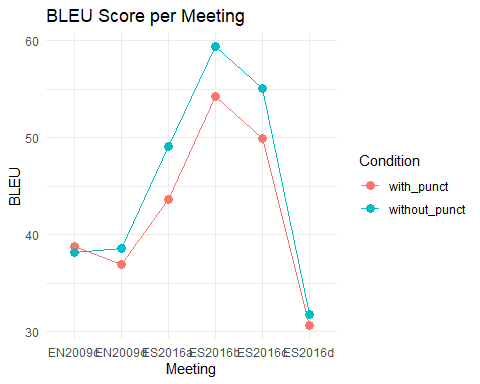


Line Plots

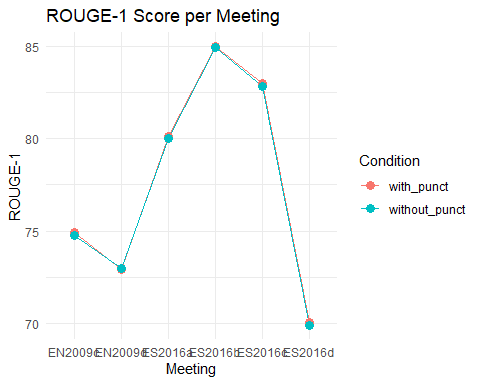
line\_wer <- ggplot(wer\_long, aes(x = File, y = Score, group = Condition, color = Condition)) +  
 geom\_line() +  
 geom\_point(size = 3) +  
 labs(title = "WER per Meeting", y = "WER", x = "Meeting") +  
 theme\_minimal()  
ggsave(file.path(plot\_dir, "wer\_lineplot.png"), line\_wer, width = 6, height = 4, dpi = 300)  
line\_wer



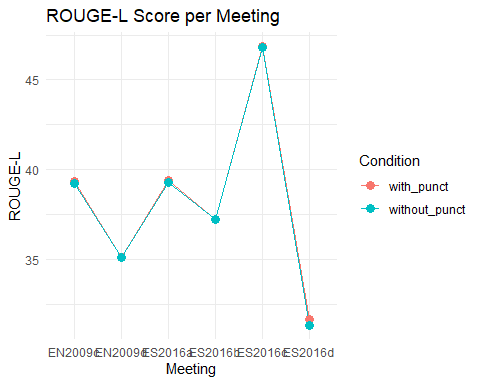
line\_bleu <- ggplot(bleu\_long, aes(x = File, y = Score, group = Condition, color = Condition)) +  
 geom\_line() +  
 geom\_point(size = 3) +  
 labs(title = "BLEU Score per Meeting", y = "BLEU", x = "Meeting") +  
 theme\_minimal()  
ggsave(file.path(plot\_dir, "bleu\_lineplot.png"), line\_bleu, width = 6, height = 4, dpi = 300)  
line\_bleu



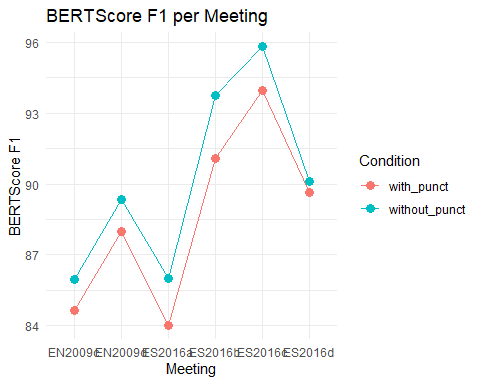
line\_rouge1 <- ggplot(rouge\_long\_1, aes(x = File, y = Score, group = Condition, color = Condition)) +  
 geom\_line() +  
 geom\_point(size = 3) +  
 labs(title = "ROUGE-1 Score per Meeting", y = "ROUGE-1", x = "Meeting") +  
 theme\_minimal()  
ggsave(file.path(plot\_dir, "rouge1\_lineplot.png"), 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

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 = "") +  
 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 = ""  
 ) +  
 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)