Exercise

LSCI 253M Plotting exercise

This documents contains plots for the data provided on the Canvas website.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df <- read_csv("data/example_data_for_visualization.csv")
```

```
## Rows: 59091 Columns: 84
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (17): stop, filename, word, label, gender, syll, type, wordinit, poa, hy...
## dbl (67): subj, trial, vot, date, session, start, end, vdur, word_int, f0_1,...
##
## 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.
```

```
head(df)
```

```
## # A tibble: 6 x 84
##     subj stop  filename     trial word   label   vot   date gender session syll
##    <dbl> <chr> <chr>        <dbl> <chr>  <chr> <dbl>  <dbl> <chr>    <dbl> <chr>
## 1 111138 B     111138_2009~     1 BENEF~ B      9.00 2.01e7 F            1 more
## 2 111138 B     111138_2009~   104 BETTER B      6.00 2.01e7 F            1 two
## 3 111138 B     111138_2009~   107 BIG    B      6.00 2.01e7 F            1 one
## 4 111138 B     111138_2009~   110 BEEN   B      5.00 2.01e7 F            1 one
## 5 111138 B     111138_2009~   114 BEEN   B      4.00 2.01e7 F            1 one
## 6 111138 B     111138_2009~   121 BUSH   B     15.0  2.01e7 F            1 one
## # ... with 73 more variables: type <chr>, wordinit <chr>, poa <chr>,
## #   start <dbl>, end <dbl>, hyp_stop <chr>, vowel <chr>, vdur <dbl>,
## #   word_int <dbl>, prec1 <chr>, prec2 <chr>, follow1 <chr>, follow2 <chr>,
## #   pos <chr>, f0_1 <dbl>, f0_2 <dbl>, f0_3 <dbl>, f0_4 <dbl>, f0_5 <dbl>,
## #   f0_6 <dbl>, f0_7 <dbl>, f0_8 <dbl>, f0_9 <dbl>, f0_10 <dbl>, w_start <dbl>,
## #   w_end <dbl>, wdur <dbl>, sent_start <dbl>, sent_end <dbl>, nWords <dbl>,
## #   spk_rate <dbl>, usef0 <dbl>, stop_start <dbl>, stop_end <dbl>, ...
```

Plots

```
# reorder labels to form pairs
df$label <- factor(df$label, levels = c("P", "B", "T", "D","K","G"))
# plot
ggplot(df,aes(x=label,y=cog,fill=label)) +
  geom_boxplot(notch=TRUE,show.legend = FALSE) +
  ylab("CoG (Hz)") +
  xlab("word-initial stop") +
  theme_classic()
```

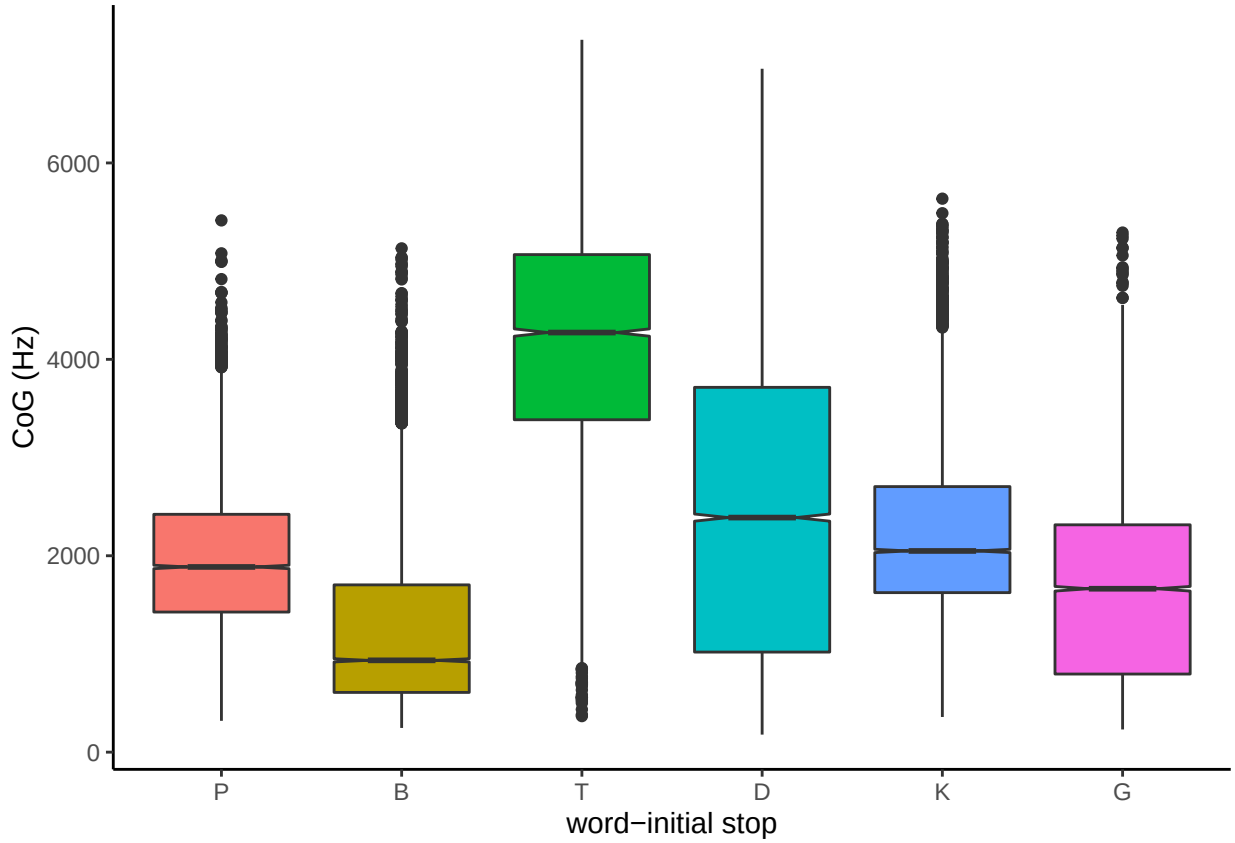

Figure 1: Center of Gravity (CoG) for voiceless/voiced pairs of word-initial stops. We observe that voiceless stops (/p/,/t/,and /k/) have longer CoGs than their voiced counterparts (/b/,/d/, and /g/).

```
# reorder labels to form pairs
df$pos <- factor(df$pos, levels = c("utt_init", "utt_mid", "prepause", "postpause","utt_final"))
ggplot(df,aes(x=vdur,fill=vowel)) +
  geom_density(alpha = 0.6) +
  ylab("vowel duration (ms)") +
  theme_classic() +
  xlim(0,500)
```

```
df$pos <- factor(df$pos, levels = c("utt_init", "utt_mid", "prepause", "postpause","utt_final"))
ggplot(df,aes(x=trial,y=usef0,color=gender)) +
  geom_smooth() +
  xlab("Trial number") +
  ylab("f0 (Hz)") +
```

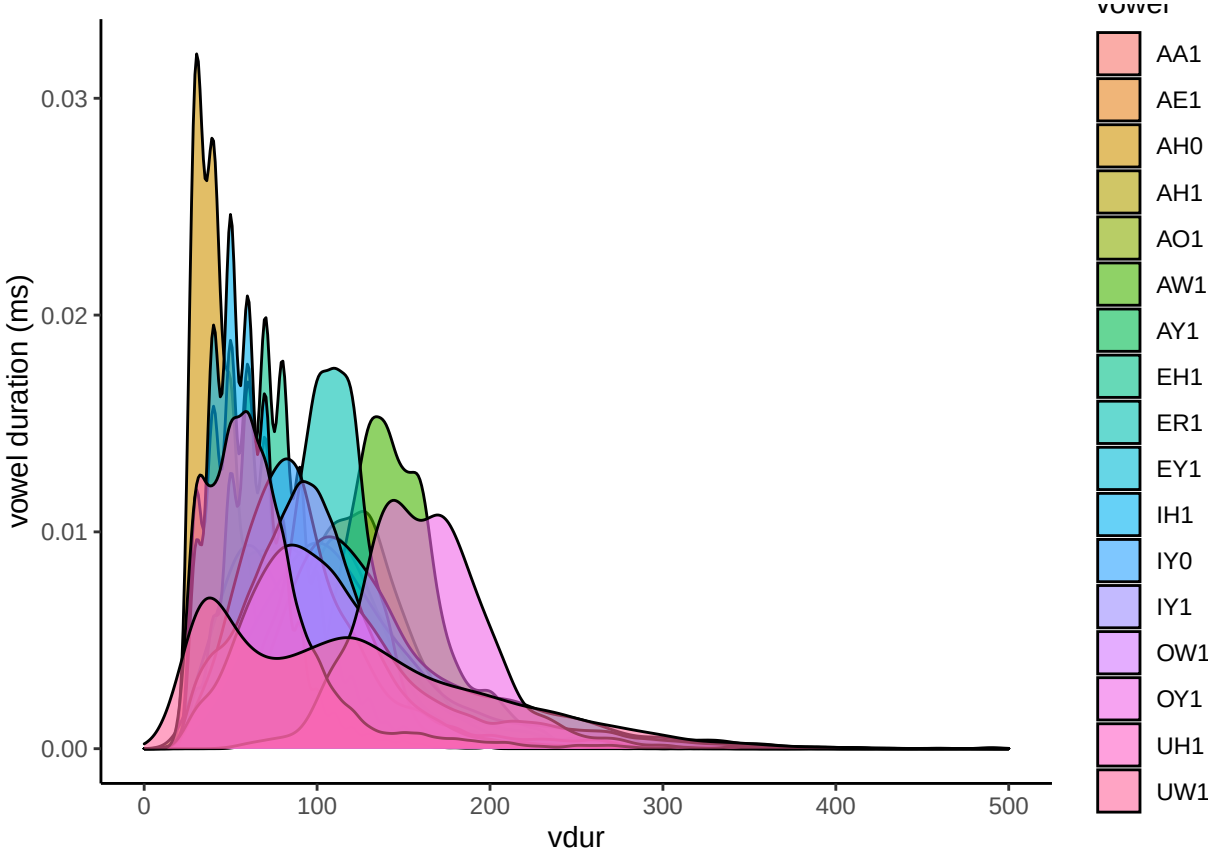

Figure 2: Distribution of vowel durations

```
  theme_classic()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

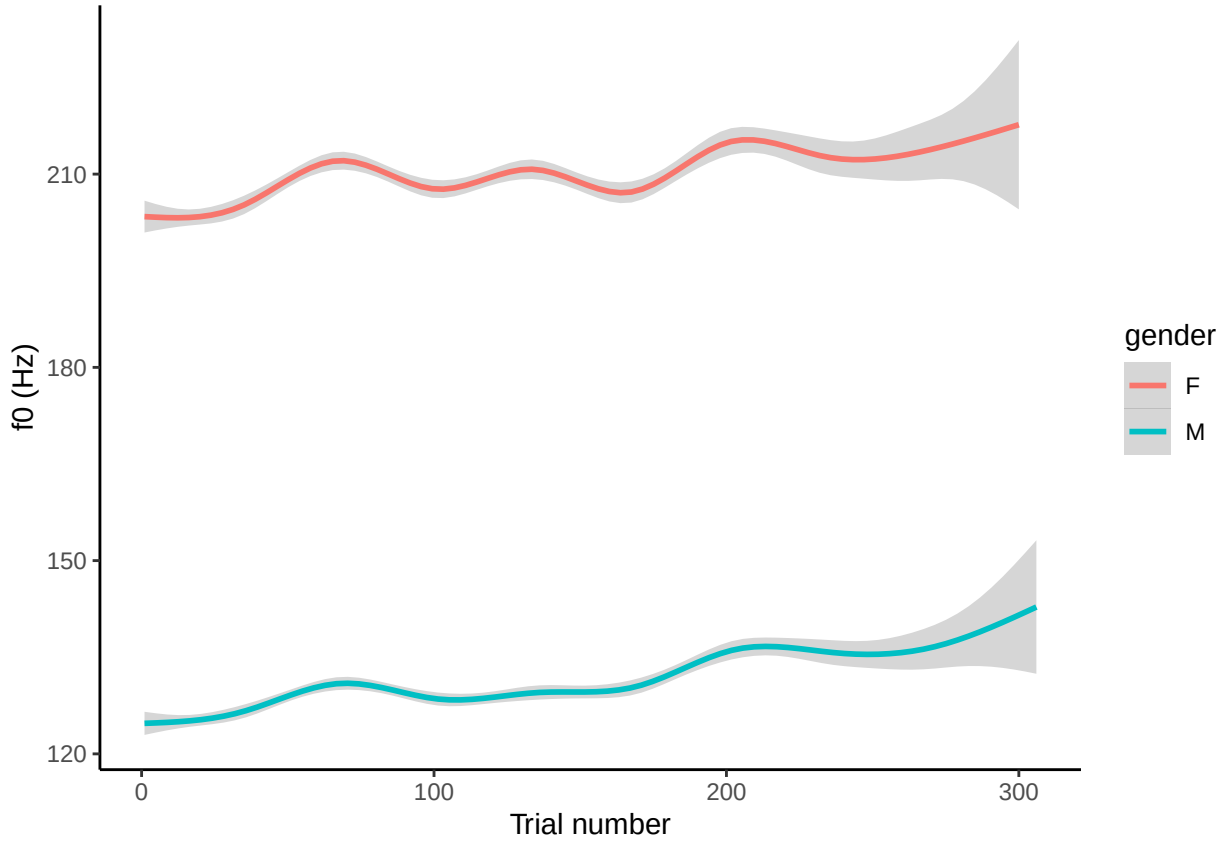

Figure 3: Fundamental frequency (f0) across trials for different utterance positions