# Linguistic Data: Quantitative Analysis and Visualisation

### ANOVA: analysis of variance

*Ilya Schurov, Olga Lyashevskaya, George Moroz, Alla Tambovtseva*

*09 February 2018*

Load data on Icelandic:

```
phono <- read.csv("http://math-info.hse.ru/f/2018-19/ling-data/icelandic.csv")
```
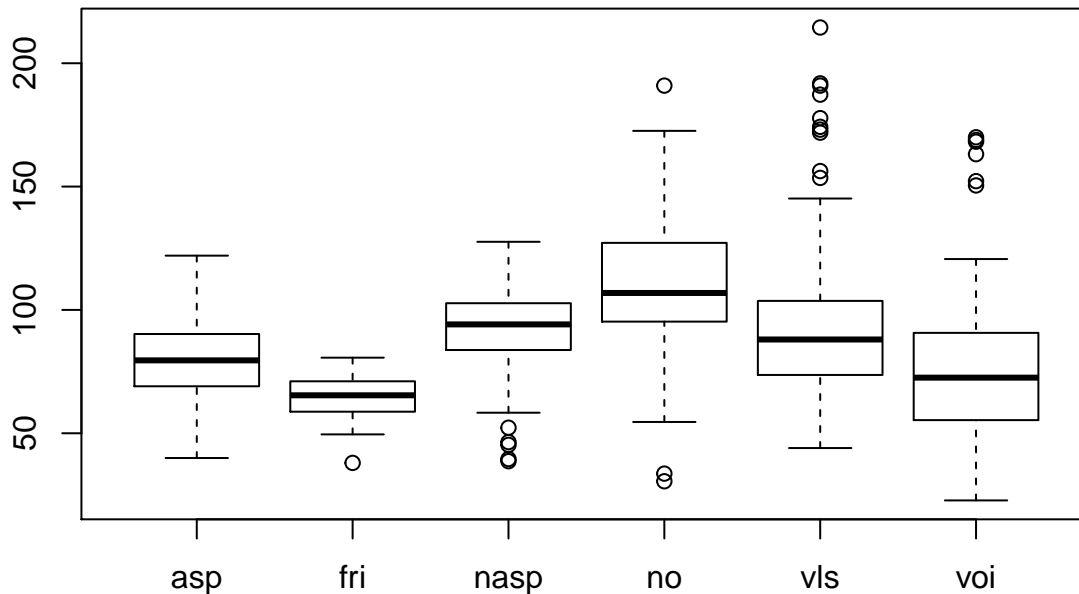
Look at groups of consonants:

```
table(phono$cons1)
```

```
##
##  asp  fri nasp   no  vls  voi
##  304   15  133   94  142  118
```

Create a boxplot for vowel duration for each group of consonants:

```
boxplot(phono$vowel.dur ~ phono$cons1)
```



Perform ANOVA:

```
res <- aov(phono$vowel.dur ~ phono$cons1)
res
```

```
## Call:
##    aov(formula = phono$vowel.dur ~ phono$cons1)
##
## Terms:
##                 phono$cons1 Residuals
## Sum of Squares      96776.3   404073.9
## Deg. of Freedom           5        800
```

1

```
## 
## Residual standard error: 22.47426
## Estimated effects may be unbalanced
```

More informative summary:

```
# H0: there are no difference in population means by groups
summary(res)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)    
## phono$cons1    5  96776   19355   38.32 <2e-16 ***
## Residuals    800 404074     505                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Question:** judging by the output above, can we conclude that average vowel duration differ significantly in different groups of consonants?