Curran_Thomas_HW2

Thomas Curran

4/17/2017

Debating with Quanteda:

1) Install the quanteda package:

2) Create DEBATES Object

```
#Trump Before Nomination
summary(lm(log(trump_0_sum$Types) ~log(trump_0_sum$Tokens)))
```

```
##
## Call:
## lm(formula = log(trump_0_sum$Types) ~ log(trump_0_sum$Tokens))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60819 -0.06860  0.02607  0.10580  0.26392
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.16549    0.03684   4.492 1.93e-05 ***
## log(trump_0_sum$Tokens)  0.86022    0.01200  71.664  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1509 on 98 degrees of freedom
## Multiple R-squared:  0.9813, Adjusted R-squared:  0.9811
## F-statistic:  5136 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
#Trump After Nomination
summary(lm(log(trump_1_sum$Types) ~log(trump_1_sum$Tokens)))
```

```
##
## Call:
## lm(formula = log(trump_1_sum$Types) ~ log(trump_1_sum$Tokens))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51042 -0.09591  0.02800  0.13390  0.37368
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -0.06855    0.05740  -1.194    0.235
## log(trump_1_sum$Tokens)  0.89443    0.01501  59.590   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2111 on 98 degrees of freedom
```

```
## Multiple R-squared:  0.9731, Adjusted R-squared:  0.9729
## F-statistic:  3551 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
#Clinton Before Nomination
summary(lm(log(clinton_0_sum$Types) ~log(clinton_0_sum$Tokens)))
```

```
##
## Call:
## lm(formula = log(clinton_0_sum$Types) ~ log(clinton_0_sum$Tokens))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74579 -0.08561  0.00587  0.11926  0.29497
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.05524    0.05401  -1.023    0.309
## log(clinton_0_sum$Tokens)  0.91354    0.01375  66.425   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1856 on 98 degrees of freedom
## Multiple R-squared:  0.9783, Adjusted R-squared:  0.978
## F-statistic:  4412 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
#Clinton After Nomination
summary(lm(log(clinton_1_sum$Types) ~log(clinton_1_sum$Tokens)))
```

```
##
## Call:
## lm(formula = log(clinton_1_sum$Types) ~ log(clinton_1_sum$Tokens))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74579 -0.08561  0.00587  0.11926  0.29497
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.05524    0.05401  -1.023    0.309
## log(clinton_1_sum$Tokens)  0.91354    0.01375  66.425   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1856 on 98 degrees of freedom
## Multiple R-squared:  0.9783, Adjusted R-squared:  0.978
## F-statistic:  4412 on 1 and 98 DF,  p-value: < 2.2e-16
```

```
#Lexical Diversity: Trump before nomination
trump0_dfm<-dfm(corpus_subset(DEBATES, speaker=='trump' & primary==0))
lexdiv_trump0<-textstat_lexdiv(trump0_dfm, measure="TTR")
mean(lexdiv_trump0)
```

```
## [1] 189.5621
```

```
trump1_dfm<-dfm(corpus_subset(DEBATES, speaker=='trump' & primary==1))
lexdiv_trump1<-textstat_lexdiv(trump1_dfm, measure="TTR")
mean(lexdiv_trump1)
```

```
## [1] Inf
```

```
clinton0_dfm<-dfm(corpus_subset(DEBATES, speaker=='clinton' & primary==0))
lexdiv_clinton0<-textstat_lexdiv(clinton0_dfm, measure="TTR")
mean(lexdiv_clinton0)
```

```
## [1] 148.7434
```

```
clinton1_dfm<-dfm(corpus_subset(DEBATES, speaker=='clinton' & primary==1))
lexdiv_clinton0<-textstat_lexdiv(clinton1_dfm, measure="TTR")
mean(lexdiv_clinton0)
```

```
## [1] 292.2885
```