#Durbin-Watson Autocorrelation Testing

#For reference on what this statistic is and how to interpret it, here's a good explanation from Investopedia; https://www.investopedia.com/terms/d/durbin-watson-statistic.asp)
#The gist is that data without autocorrelation generally has a DW statistic between 1.5 and 2.5, and more than 2.5 implies autocorrelation (and could imply that at least one person took the survey more than once with similar responses.)

#First, create a vector that has a count of the survey responses in it, and turn it into a numeric variable instead of an integer

```
> x1 <- 1:56
> x2 <- as.numeric(x1)
```

#Now we can do the actual testing to see if we have autocorrelation present in our combined datasets…

#DW Testing of one of the Q7 questions to major type, with the response order vector added

```
> DWANOVA1 <- aov(Q7.2N ~ RankedQ2NV1 + x2)
> library(lmtest)
Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

        as.Date, as.Date.numeric

> dwtest(DWANOVA1, alternative = 'less')

    Durbin-Watson test

data:  DWANOVA1
DW = 1.7539, p-value = 0.8576
alternative hypothesis: true autocorrelation is less than 0

> dwtest(DWANOVA1, alternative = 'greater')

    Durbin-Watson test

data:  DWANOVA1
DW = 1.7539, p-value = 0.1424
alternative hypothesis: true autocorrelation is greater than 0

> library(car)
```

```
Loading required package: carData
> durbinWatsonTest(DWANOVA1)
lag Autocorrelation D-W Statistic p-value
   1       0.1158979       1.753928   0.304
 Alternative hypothesis: rho != 0
```

#We can also do some Durbin-Watson analysis on the data where business majors and social sciences majors are separated

```
>BizANOVA <- aov(Q7.2N ~ RankedQ2NV2 + x2)
>dwtest(BizDWANOVA, alternative = 'less')
Durbin-Watson test

data:  BizDWANOVA
DW = 1.6949, p-value = 0.9022
alternative hypothesis: true autocorrelation is less than 0

>dwtest(BizANOVA, alternative = 'greater')
Durbin-Watson test

data:  BizDWANOVA
DW = 1.6949, p-value = 0.09785
alternative hypothesis: true autocorrelation is greater than 0

> durbinWatsonTest(BizDWANOVA)
 lag Autocorrelation D-W Statistic p-value
   1       0.1450712       1.694943   0.188
 Alternative hypothesis: rho != 0
```

#The second version of the Question 2 variable also has a Durbin-Watson statistic between 1.5 and 1.7, once again showing that we do not appear to have evidence for auto-correlation

#More Durbin-Watson testing done on Q3 and Q4 specifically

```
> DWANOVA2 <- aov(Q4N ~ Q3N + x2)

> dwtest(DWANOVA2, alternative = 'less')

    Durbin-Watson test

data:  DWANOVA2
DW = 1.9344, p-value = 0.6519
alternative hypothesis: true autocorrelation is less than 0
```

```
> dwtest(DWANOVA2, alternative = 'greater')

    Durbin-Watson test

data:  DWANOVA2
DW = 1.9344, p-value = 0.3481
alternative hypothesis: true autocorrelation is greater than 0

> durbinWatsonTest(DWANOVA2)
 lag Autocorrelation D-W Statistic p-value
   1       0.0004090474          1.934434        0.62
 Alternative hypothesis: rho != 0
```

#It appears that we have somewhat higher DW test statistics between 1.9 and 2.0. However, the test statistics clearly aren't above 2.5, so there's still no strong evidence for autocorrelation. Thus, I would conclude that there *probably* weren't any repeat survey takers.