# French Elite analysis

*Gary Sztajnman*

*February 22, 2016*

In this homework, we are going to study the "FRENCH FINANCIAL ELITE" dataset downloaded from http://moreno.ss.uci.edu/. This dataset is composed of three 28 by 28 matrices on the influence, the belonging of an elite and the friendship of 28 top French leaders in Finance. We also have information about their background such as religion, education or age…

## Hypothesis

I want to study the characteristics of powerfull member of an elite. At first, I wanted to see the impact of higher education on the influence of leaders in an elite group. However, education level may not be a differentiable factor as all leaders when through inpressive school (in particular in France where people are considered based on the school they went to. Another factor that is more easy to anticipate is the level of religion defined as the belief in god and the frequency of praying.

We want to test the relation between religion and position of power in an elite group. In particular, we hypothesis that the more powerful a leader is the less he prays and believe in god.

## Explanation

In France, the concept of "laïcité" (secularism) is particularly important in higher education school and in high business circle and elite people come from these school. Hence, even if they may come from religious family, they will probably get away from religion after and during their higher education. Furthermore, the financial elite is busy and we could think that don't have to much time to pray and think about religion and god. Finally, financial leaders tends to prove their skills by their understanding of financial markets and business and not by showing their faith in god.

So I would think that French financial leaders are less linked to religion when they are more influencial.

Let's use the dataset to analyze that

## Retrieveing and cleaning the data

```
### Libraries----
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:data.table':
##
##     between, last
##
## The following objects are masked from 'package:stats':
```

```
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(plyr)
```

```
## -------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```
library(igraph)
```

```
##
## Attaching package: 'igraph'
##
## The following objects are masked from 'package:dplyr':
##
##     %>%, as_data_frame, groups, union
##
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
##
## The following object is masked from 'package:base':
##
##     union
```

```
### import data----
netw <- read.csv("http://moreno.ss.uci.edu/ffe.dat", header=T, sep=";", skip= 36)
netw = setDT(netw)[, tstrsplit(DATA., ' ')]
netw = netw[,V1:=NULL]

#splitting the network data to advice/friendship/report
influence <- as.matrix(slice(netw, 1:28))
elite <- as.matrix(slice(netw, 29:56))
friend <- as.matrix(slice(netw, 57:84))
```

```r
#importing attributes data
setwd("/Users/garyair/Desktop/Dropbox/Columbia/SNA/Labs/SNA-EgoNetwork-1/French elite/")
attributes <- read.csv("attributes.csv", header=TRUE)

#Clean missing data in attributes
attributes$igyear <- NULL
attributes$birthplace <- NULL
attributes$polyyear <- NULL
attributes$enayear <- NULL
attributes$zipcode <- NULL


#We decide to restrict our analysis to 3 major topics: age, education and religion. Hence, we delete ot
attributes$fathers.lev <- NULL
attributes$masons <- NULL
attributes$socialreg <- NULL
attributes$eliteprom <- NULL
attributes$prestige <- NULL
attributes$clubs <- NULL
attributes$topboards <- NULL
attributes$inspec.gen <- NULL
attributes$cabinet <- NULL
attributes$finance.min <- NULL
attributes$party <- NULL
attributes$elitevote <- NULL


#Cleaning and polishing data
names(attributes)
```

```
## [1] "birthdate"   "sciencepoly" "polytech"    "university"  "normal.sch"
## [6] "ena"         "religion"
```

```r
#"sciencepoly" "polytech"    "university"  "normal.sch"  "ena" "religion"
attributes$sciencepoly[attributes$sciencepoly == 2] <- FALSE
attributes$sciencepoly[attributes$sciencepoly == 1] <- TRUE

attributes$polytech[attributes$polytech == 2] <- FALSE
attributes$polytech[attributes$polytech == 1] <- TRUE

attributes$university[attributes$university == 2] <- FALSE
attributes$university[attributes$university == 1] <- TRUE

attributes$normal.sch[attributes$normal.sch == 2] <- FALSE
attributes$normal.sch[attributes$normal.sch == 1] <- TRUE

attributes$ena[attributes$ena == 2] <- FALSE
attributes$ena[attributes$ena == 1] <- TRUE
```

## Variables

For this study, 3 types of variables have been selected. 1."birthdate"
2."sciencepoly", "polytech", "university", "normal.sch", "ena"
3. "religion"

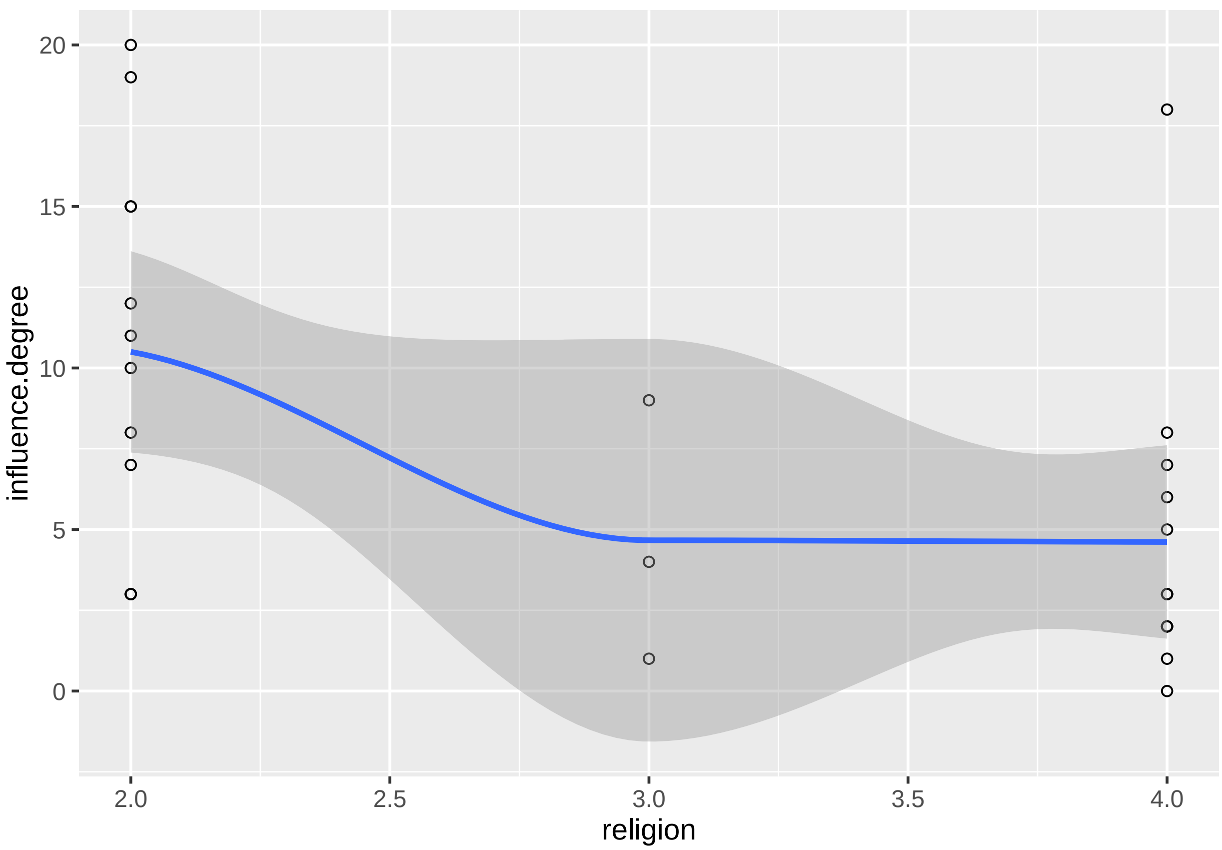
Birthdate is an independent variable, it gives an idea of which elite member is part of which generation Education variables are binary coded (true or false) Religion is a class variable coded from 2 to 4 (2 is low religion level and 4 is high religion level)

To this attribute, we add the ego network measure: transity and network size for elite and influence

Influence network: "WHO INFLUENCES YOU"

```
influenceg=graph.adjacency(influence,mode="directed",weighted=NULL)

### calculate ego network influence size ###
attributes$influence.degree <- degree(influenceg)
ggplot(attributes, aes(x= religion, y =influence.degree )) +
    geom_point(shape=1) +    # Use hollow circles
    geom_smooth()
```

```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : pseudoinverse used at 1.99

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : neighborhood radius 2.01

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : There are other near singularities as well. 4.0401

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : pseudoinverse used at 1.99

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : neighborhood radius 2.01

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : reciprocal condition number 0

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : There are other near singularities as well. 4.0401
```




```
### calculate transitivity of elite ego network ###
attributes$Influence.transitivity =transitivity(influenceg, type="local")
attributes$Influence.transitivity[attributes$Influence.transitivity == "NA"] <- mean(attributes$Influenc

ggplot(attributes, aes(x= religion, y =Influence.transitivity )) +
    geom_point(shape=1) +    # Use hollow circles
    geom_smooth()
```

```
## Warning: Removed 3 rows containing non-finite values (stat_smooth).
```

```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : pseudoinverse used at 1.99
```

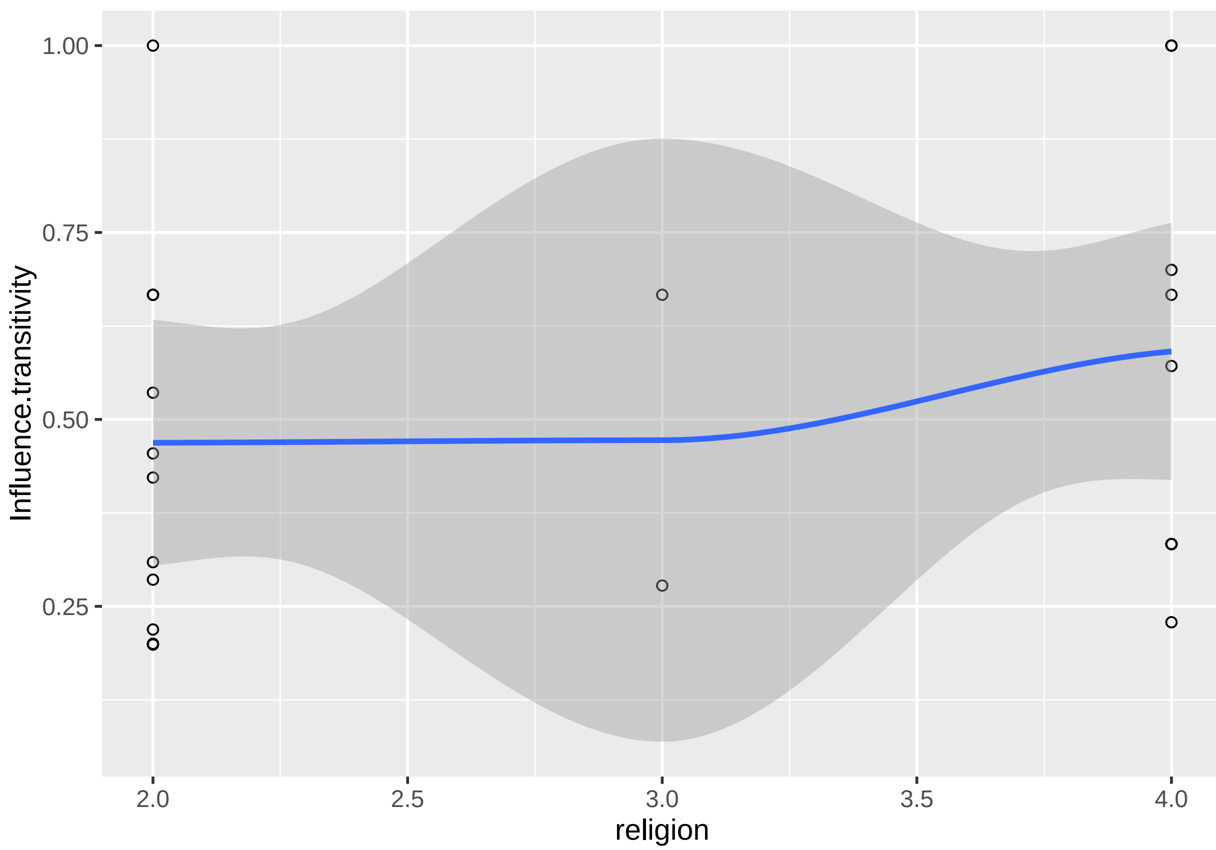
```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : neighborhood radius 2.01
```

```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : reciprocal condition number 1.509e-16
```

```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : There are other near singularities as well. 4.0401
```

```
## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : pseudoinverse used at 1.99
```

```
## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : neighborhood radius 2.01

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : reciprocal condition number 1.509e-16

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : There are other near singularities as well. 4.0401

## Warning: Removed 3 rows containing missing values (geom_point).
```

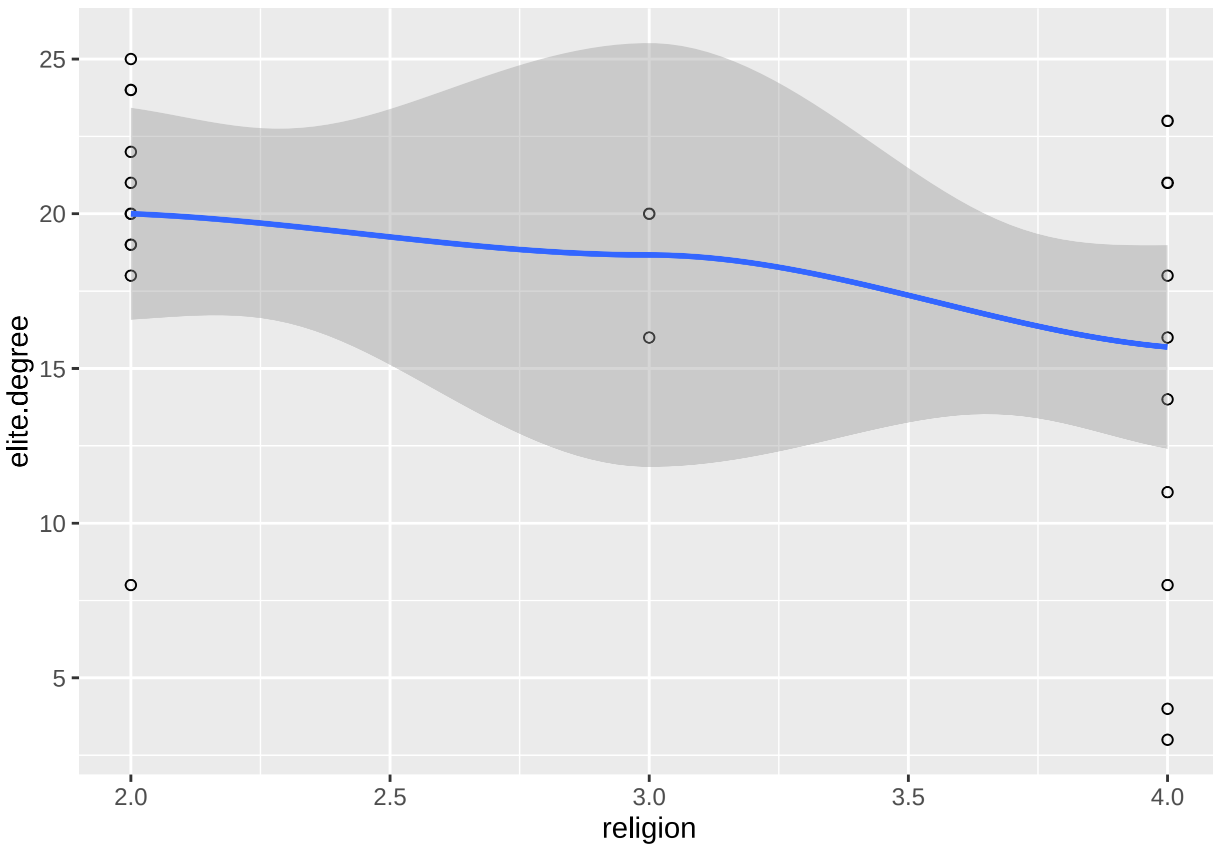



Elite network: "WHO IS IN THE ELITE"

```
eliteg=graph.adjacency(elite,mode="undirected",weighted=NULL)
### calculate ego network elite size ###
attributes$elite.degree <- degree(eliteg)
ggplot(attributes, aes(x= religion, y =elite.degree )) +
    geom_point(shape=1) +    # Use hollow circles
    geom_smooth()
```

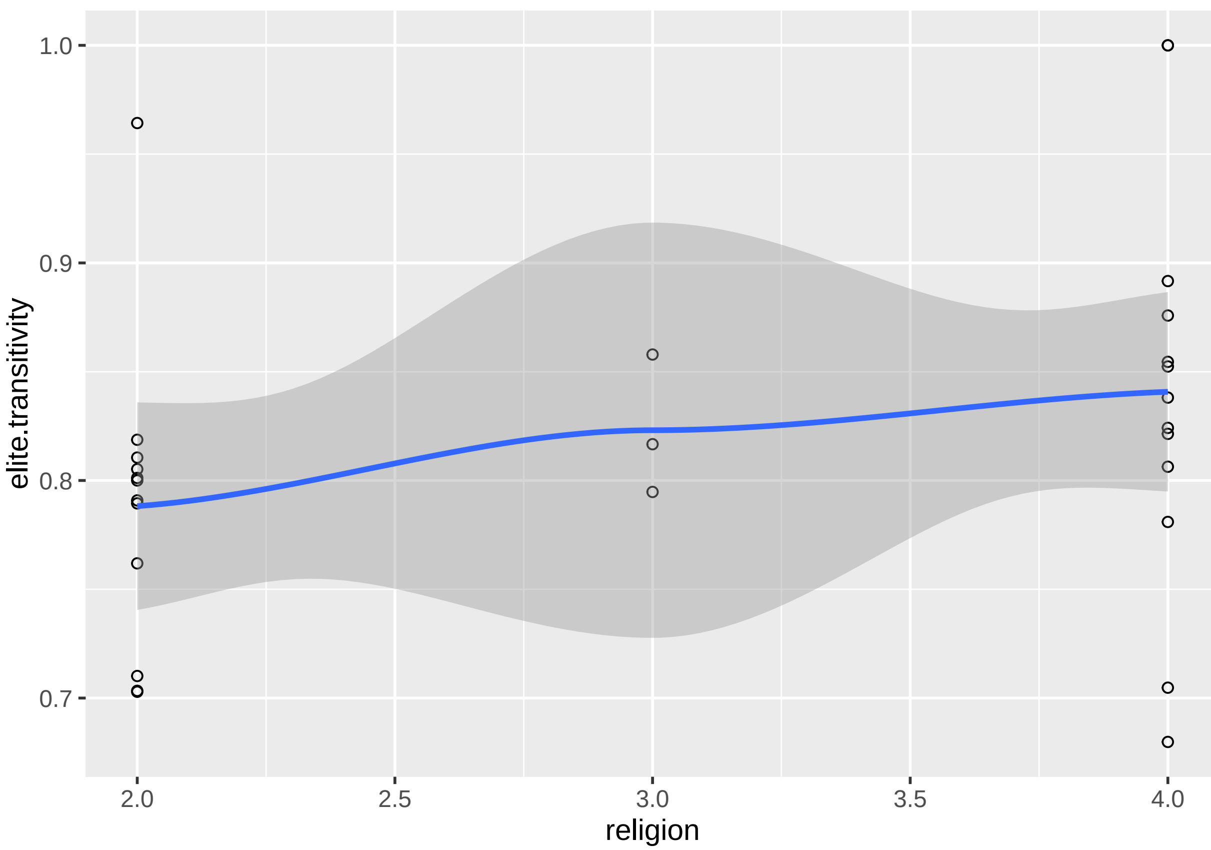
```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : pseudoinverse used at 1.99

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : neighborhood radius 2.01
```

```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : There are other near singularities as well. 4.0401

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : pseudoinverse used at 1.99

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : neighborhood radius 2.01

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : reciprocal condition number 0

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : There are other near singularities as well. 4.0401
```




```
### calculate transitivity of elite ego network ###
attributes$elite.transitivity =transitivity(eliteg, type="local")
attributes$elite.transitivity[attributes$elite.transitivity == "NaN"] <- mean(attributes$elite.transiti

ggplot(attributes, aes(x= religion, y =elite.transitivity )) +
    geom_point(shape=1) +     # Use hollow circles
    geom_smooth()
```

```
## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : pseudoinverse used at 1.99

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : neighborhood radius 2.01

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square,
## normalize, : There are other near singularities as well. 4.0401

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : pseudoinverse used at 1.99

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : neighborhood radius 2.01

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : reciprocal condition number 0

## Warning in predLoess(y, x, newx, s, weights, pars$robust, pars$span, pars
## $degree, : There are other near singularities as well. 4.0401
```



## Initial results

If we look at the elite network and the influence network, we see in both case a clear decreasing factor between religion and the place in the elite. Elite member tend to have a more influential place when they spend less time to pray. If we just plot the average amount of excitement people claim as a function of the average educational attainment of their social circle, we see that as one's social circle is more educated, the higher excitement ego's report.

However, it's interesting to see that transitivity is positively correlated with religion in this dataset. Religious people may be less influential but more connected to their religious peers.

We will now use regression model to further study these ideas.

## Model 1

We first start by studying the impact on influence:

```r
summary(lm(influence.degree~ religion, attributes))
```

```
##
## Call:
## lm(formula = influence.degree ~ religion, data = attributes)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.177 -3.195 -1.318  2.038 13.682
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   16.037      3.323   4.826 5.32e-05 ***
## religion      -2.930      1.045  -2.803  0.00944 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.222 on 26 degrees of freedom
## Multiple R-squared:  0.2321, Adjusted R-squared:  0.2025
## F-statistic: 7.857 on 1 and 26 DF,  p-value: 0.00944
```

```r
summary(lm(Influence.transitivity~ religion, attributes))
```

```
##
## Call:
## lm(formula = Influence.transitivity ~ religion, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.35735 -0.25277 -0.01468  0.20232  0.53565
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.34259    0.17503   1.957   0.0626 .
## religion     0.06088    0.05626   1.082   0.2904
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2696 on 23 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.04845,    Adjusted R-squared:  0.007078
## F-statistic: 1.171 on 1 and 23 DF,  p-value: 0.2904
```

The regression on influence degree validates the idea of negative correlation between religion and influence with a p-value lower than 5% and a beta estimate of approximately -3. Clearly we note that more religion means less influence in our dataset.

However, we can not confirm our hypothesis on religion and the transitivity of influence because the p-value is too high aroung 30%.

## Model 2

To be sure of our first idea on education we will the education variables in our model:

```
summary(lm(influence.degree ~ ena + sciencepoly + polytech + university + normal.sch, attributes))
```

```
##
## Call:
## lm(formula = influence.degree ~ ena + sciencepoly + polytech +
##     university + normal.sch, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4191  -3.3605  -0.0012   2.0421  13.8895
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.142      3.112   2.616   0.0158 *
## ena            3.238      3.078   1.052   0.3042
## sciencepoly   -2.130      3.428  -0.621   0.5408
## polytech       2.277      3.349   0.680   0.5037
## university    -3.140      2.763  -1.136   0.2680
## normal.sch    -9.380      7.287  -1.287   0.2114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.867 on 22 degrees of freedom
## Multiple R-squared:  0.1799, Adjusted R-squared:  -0.00646
## F-statistic: 0.9653 on 5 and 22 DF,  p-value: 0.4602
```

```
summary(lm(Influence.transitivity~ ena + sciencepoly + polytech + university + normal.sch, attributes))
```

```
##
## Call:
## lm(formula = Influence.transitivity ~ ena + sciencepoly + polytech +
##     university + normal.sch, data = attributes)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37512 -0.12043 -0.02078  0.09154  0.42488
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.44285    0.13795   3.210  0.00461 **
## ena         -0.07977    0.13661  -0.584  0.56616
## sciencepoly  0.03248    0.14740   0.220  0.82797
## polytech    -0.13368    0.14850  -0.900  0.37930
## university   0.17957    0.13112   1.369  0.18683
## normal.sch   0.63692    0.31908   1.996  0.06046 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2519 on 19 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.3135, Adjusted R-squared:  0.1328
## F-statistic: 1.735 on 5 and 19 DF,  p-value: 0.1749
```

As expected, we note that we cannot infer any correlation between education and influence because p-values are too high. The dataset is too small to detects the impact of these binary variables on the influence levels.

# Model 3

To confirm our analysis, we will reproduce our linear regression on the elite data. This will help us determine if being part of the elite is linked to being influential.

```r
summary(lm(elite.degree~ religion, attributes))
```

```
##
## Call:
## lm(formula = elite.degree ~ religion, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7768  -1.8162   0.5658   4.1584   7.2232
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   24.406      3.598   6.784 3.36e-07 ***
## religion      -2.157      1.132  -1.906   0.0677 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.654 on 26 degrees of freedom
## Multiple R-squared:  0.1226, Adjusted R-squared:  0.08889
## F-statistic: 3.634 on 1 and 26 DF,  p-value: 0.06771
```

```r
summary(lm(elite.transitivity~ religion, attributes))
```

```
##
## Call:
## lm(formula = elite.transitivity ~ religion, data = attributes)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.161814 -0.029281  0.001468  0.023404  0.175117
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.73668    0.05011  14.703 4.11e-14 ***
## religion     0.02624    0.01576   1.665    0.108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07875 on 26 degrees of freedom
## Multiple R-squared:  0.09636,    Adjusted R-squared:  0.06161
## F-statistic: 2.773 on 1 and 26 DF,  p-value: 0.1079
```

```r
summary(lm(elite.degree ~ ena + sciencepoly + polytech + university + normal.sch, attributes))
```

```
##
## Call:
## lm(formula = elite.degree ~ ena + sciencepoly + polytech + university +
##     normal.sch, data = attributes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.9375  -0.7657   1.1611   2.9221   7.0625
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.3136     3.0398   6.354 2.15e-06 ***
## ena           0.7643     3.0065   0.254   0.8017
## sciencepoly  -0.3870     3.3484  -0.116   0.9090
## polytech     -2.3761     3.2714  -0.726   0.4753
## university   -0.8519     2.6984  -0.316   0.7552
## normal.sch  -16.0779     7.1182  -2.259   0.0342 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.731 on 22 degrees of freedom
## Multiple R-squared:  0.2374, Adjusted R-squared:  0.06405
## F-statistic:  1.37 on 5 and 22 DF,  p-value: 0.2737
```

```r
summary(lm(elite.transitivity~ ena + sciencepoly + polytech + university + normal.sch, attributes))
```

```
##
## Call:
```

```
## lm(formula = elite.transitivity ~ ena + sciencepoly + polytech +
##     university + normal.sch, data = attributes)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.137550 -0.022852 -0.000037  0.026040  0.163041
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.77923    0.04068  19.153 3.29e-15 ***
## ena         -0.01581    0.04024  -0.393   0.6982
## sciencepoly  0.01763    0.04482   0.393   0.6978
## polytech     0.06122    0.04378   1.398   0.1760
## university   0.02020    0.03612   0.559   0.5817
## normal.sch   0.23658    0.09527   2.483   0.0211 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0767 on 22 degrees of freedom
## Multiple R-squared:  0.2746, Adjusted R-squared:  0.1097
## F-statistic: 1.666 on 5 and 22 DF,  p-value: 0.1848
```

The results are similar than with the influence network. We have relatively similar level of p-value and the only thing that we can infer is that religion is negatively correlated with being part of the elite.

# Conclusion

The answer to the main problem seems to be clear. In the case of the French financial elite, one's relation to god does influence its position in his network. We can propose some interpretation about the role of religion in elite networks but more data and more analysis would help us confirm our ideas.