Mark Gallo-Social Circle Analysis GR5062-Lab1

Code ▾

1. Develop a hypothesis about how some ego-circle measure (e.g., degree/size, density, diversity, average-level of alters, homophily, structural holes, or brokerage) may be related to some other variable of interest.

For this lab I wish to focus on how one's social circle characterisitcs effect one's use of the internet. My points of interest are the average religious association of one's social circle, the average age of those within the social circle, and the average education of those within the social circle. I expect that as the average age of those within one's social circle increases, the amount of time spent on science websites will (on average) increase. I expect that as one's social circle becomes more educated that the amount of time spent on science websites will (on average) increase.The measure of average religion is a categorical variable that has been coded to allow for analysis to be conducted. I expect a variation to exist between the average time spent on science websites, given a social circles average religious idenity, but I am not sure what to expect. The variable was included because I am curious to see its effect, however I would like to suggest that as a social circle has (on average) more Christians in it, that this will inspire a decrease in the time spent on science websites.

2. Explain why you think these two variables should be related.

My belief that these variables will have a relationship of some sort stems from a potential reality. As one becomes older they are more likley to have more years of education. If someone has increased amounts of education, they may have been exposed to a diverse sample of material and literature. If they have been exposed to such material, then, with time, one may be able to deduce that scientific literacy can be useful, both for continued success in the classroom and in the professional world. Factors like these led me to believe that as somone gets older and/or gets more education, that the time spent on science websites will increase. I included religion as well because there are those who will leverage religion as a means of disassociating with the sciences. For this reason it would be interested to test whether different religious denominations inspire less or more time spent on science based websites.

3.Tell me about your variables. What is your dependent variable? What are your independent variables? How are they coded? How are they recoded? How are the calculated, if appropriate?

Dependent Variable:

(Sci30)772. In the past 30 days, how often have you visited a web site for: s. Science?

No recoding was necessary as the responces were given greater values as the time spent on science websites increased

Independent Social Variables:

[(averel) = rowMeans of (relig1-relig5)] 323. What is [NAME]'s religious preference? Is it Protestant, Catholic, Jewish, some other religion, or no religion? Do note that the sample contains reponses primarily from those who identify as Protestant or Catholic.

*[(aveage) = rowMeans of (age1-age5)] 322. How old is [NAME]?*

[(aveeduc) = rowMeans of (educ1-educ5)] 321. This card(BH) lists general levels of education. As far as you know, what is [NAME]'s highest level of education?

4.Present your initial results from your first few models. What do they indicated about your hypothesis?

Each of the plots provided below is meant to show the relationship between each of the previously discussed independent and dependent variables.
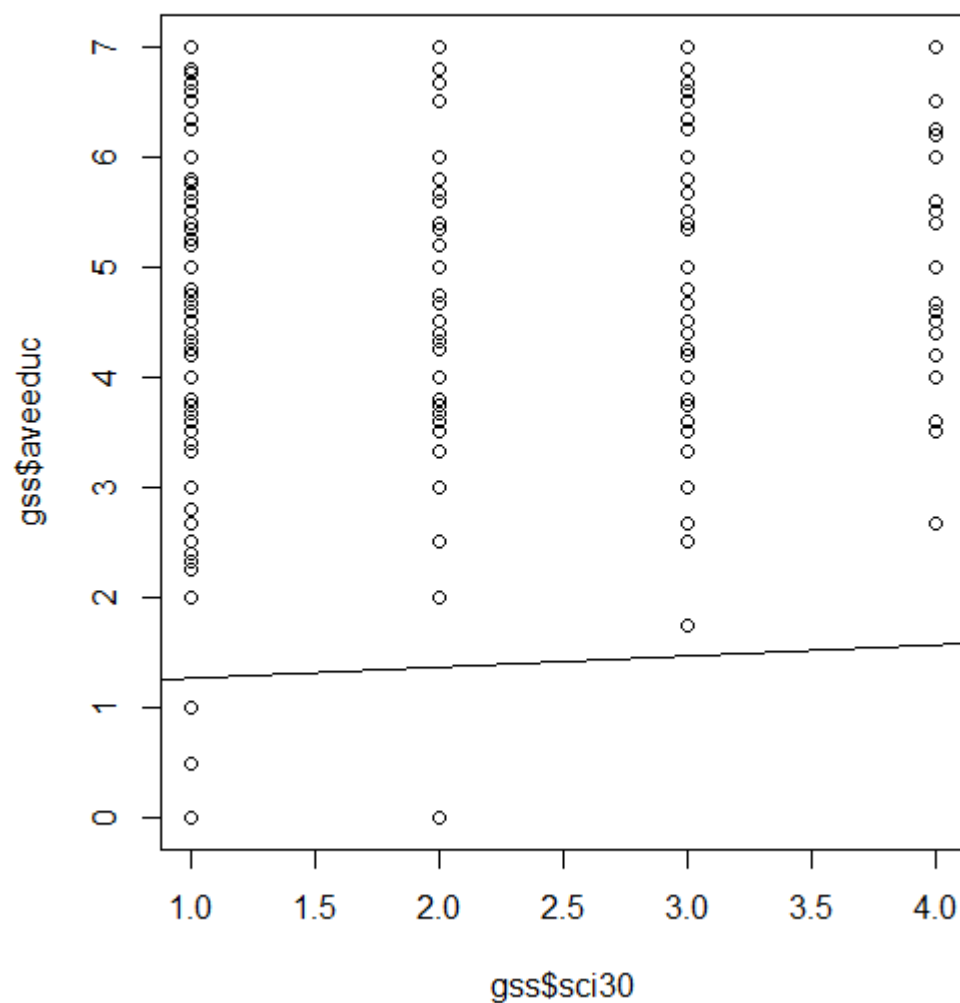
Avg Education and Science Websites: This appears to suggest that one will likely state a greater time spent on science websites as one's social circle becomes increasinly more educated.
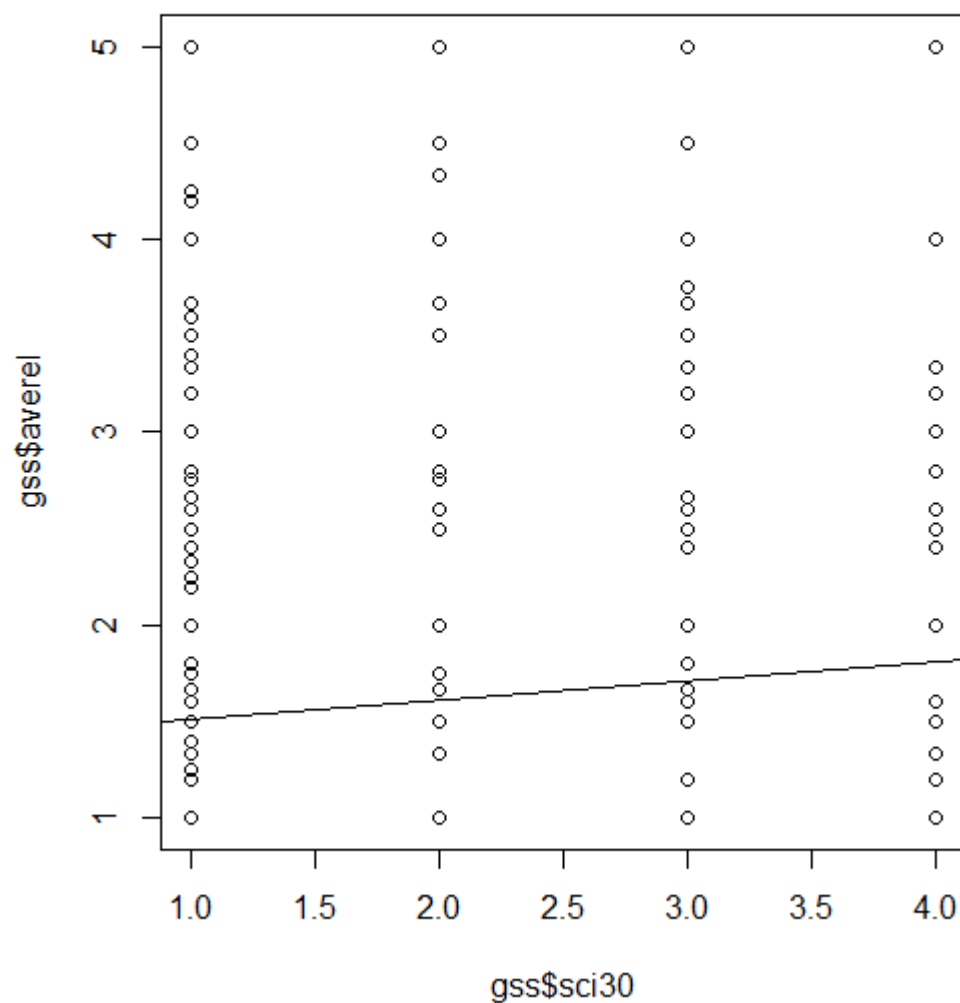
Avg Religion and Science Websites: This appears to suggest that as stated time spent on science websites increases an association with social networks that are Catholic is increasingly likely.
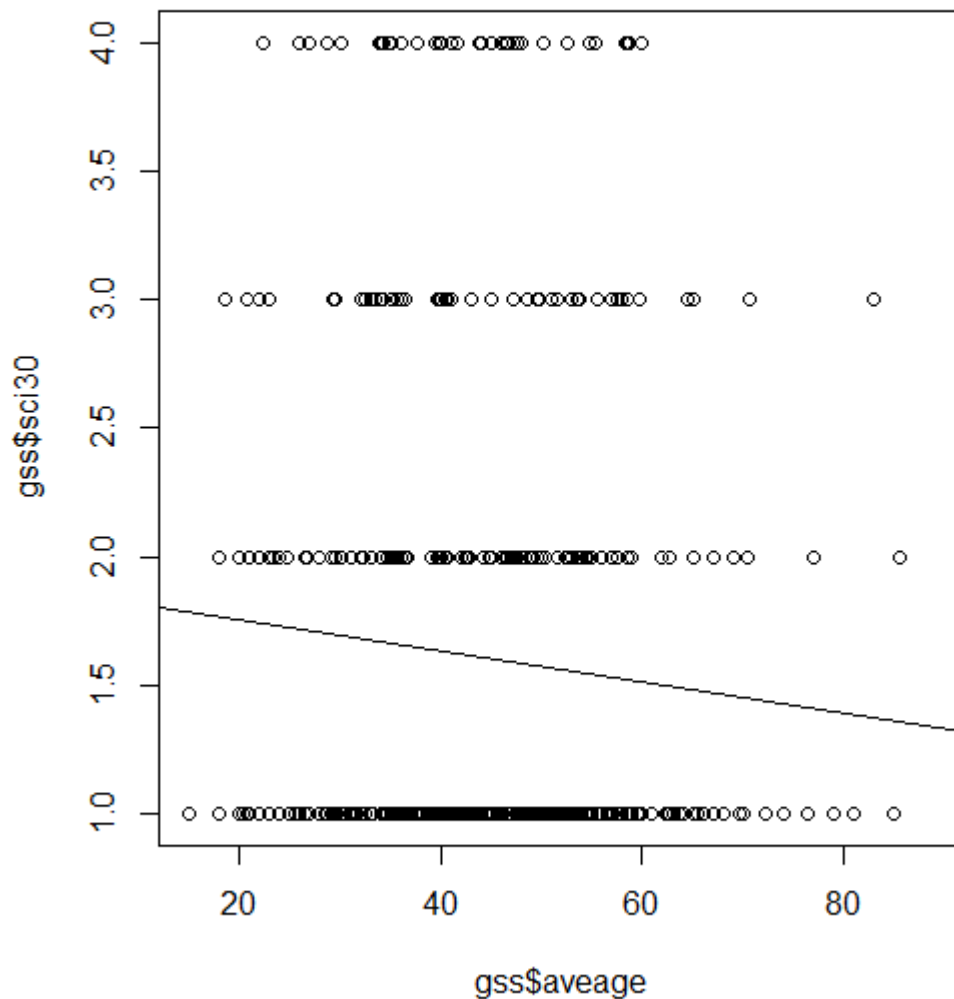
Avg Age and Science Websites: This appears to suggest that as one's social circle ages it one will be likely to report having spent less time on a science website in the past 30 days

Regressions

> My interpretation will be in relation to the final model, as it includes all the variables, it has the greatest explanatory power, and the coefficients from the previous models are not significantly different from the final

In model three, one can see that, for a given social circle, a category increase in religion (effectivley changing religions away from protestant) generates an increase in stated time spent on science websites in the past 30 days of 0.093 points, net of the degree of ego's network, ranging from 1-4.The variable was found to be statistically significant, as a t-value > 2 suggests that we can reject the null hypothesis, that no relation exists, at the 95% confidence level

One can see that, for a given social circle, a one year increase in age generates a decrease in stated time spent on science websites in the past 30 days of 0.0068 points, net of the degree of ego's network, ranging from 14-91.The variable was found to be statistically significant, as the t-value is very near 2 which suggests that we can reject the null hypothesis, that no relation exists, at just below the 95% confidence level

One can see that, for a given social circle, a one category increase increase in education generates an increase in stated time spent on science websites in the past 30 days of 0.114 points, net of the degree of ego's network, ranging from 0-7.The variable was found to be highly statistically significant, as a t-value equal to 4 suggests that we can reject the null hypothesis, that no relation exists, at the 99% confidence level

Hide

```
summary(lm(sci30 ~ averel, data=gss))
```

```
Call:
lm(formula = sci30 ~ averel, data = gss)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9075 -0.6092 -0.5097  0.4240  2.4903

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.41026    0.08615  16.370  < 2e-16 ***
averel       0.09946    0.03562   2.792  0.00544 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9567 on 511 degrees of freedom
  (5516 observations deleted due to missingness)
Multiple R-squared:  0.01502,   Adjusted R-squared:  0.0131
F-statistic: 7.795 on 1 and 511 DF,  p-value: 0.005436
```

Hide

```
summary(lm(sci30 ~ averel+aveage, data=gss))
```

```
Call:
lm(formula = sci30 ~ averel + aveage, data = gss)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0035 -0.6144 -0.4885  0.4299  2.5784

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.690100   0.182582   9.257  < 2e-16 ***
averel       0.097577   0.035897   2.718  0.00679 **
aveage      -0.006232   0.003473  -1.794  0.07339 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9545 on 507 degrees of freedom
  (5519 observations deleted due to missingness)
Multiple R-squared:  0.02219,   Adjusted R-squared:  0.01834
F-statistic: 5.754 on 2 and 507 DF,  p-value: 0.003382
```

Hide

```
summary(lm(sci30 ~ averel + aveage + aveeduc, data=gss))
```

```
Call:
lm(formula = sci30 ~ averel + aveage + aveeduc, data = gss)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1583 -0.6692 -0.4361  0.4098  2.6959

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.213408   0.215333   5.635 2.92e-08 ***
averel       0.094473   0.035555   2.657  0.00813 **
aveage      -0.006860   0.003434  -1.998  0.04628 *
aveeduc      0.113573   0.028390   4.000 7.27e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9417 on 504 degrees of freedom
  (5521 observations deleted due to missingness)
Multiple R-squared:  0.05241,   Adjusted R-squared:  0.04677
F-statistic: 9.293 on 3 and 504 DF,  p-value: 5.447e-06
```

5. Consider alternate specifications of your variables (i.e., recodings of various kinds). Consider interactions among your variables.

Standard Deviation of Avg Age

Hide

```
ages = c("age1", "age2", "age3", "age4", "age5")
sub <- gss[, ages]
sub2=transform(sub, SD=apply(sub,1, sd, na.rm = TRUE))
colnames(sub2)[6] <- "sdage"
gss <- data.frame(gss, sub2[,"sdage"])
colnames(gss)
colnames(gss)[4226] <- "sdage"
```

Standard Deviation of Avg Education

Hide

```
educs = c("educ1", "educ2", "educ3", "educ4", "educ5")
sub <- gss[, educs]
sub2=transform(sub, SD=apply(sub,1, sd, na.rm = TRUE))
colnames(sub2)[6] <- "sdeduc"
gss <- data.frame(gss, sub2[,"sdeduc"])
colnames(gss)
colnames(gss)[4227] <- "sdeduc"
```

Regressions

For my alternate variables I chose to include some additional variables, as well as the standard deviation or "diversity" of a social circles age and education. It appears that having a social cirlce of diverse ages will inspire an increase in the amount of stated time spent on science websites by 0.125 points, this variable does lack statistical significance however, as the t-value is below two. We can also see that as the additional variables were added to the model many of the social circle variables lost their signficance. This sugeests that factors outside of one's social circle may be able to better explain why someone would state having spent a certain amount of time on science websites.

Hide

```
summary(lm(sci30 ~ averel + aveage + aveeduc + as.factor(race)+ childs + sex + sdage+ sdeduc, da
ta=gss))
```

```
Call:
lm(formula = sci30 ~ averel + aveage + aveeduc + as.factor(race) +
    childs + sex + sdage + sdeduc, data = gss)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6353 -0.6317 -0.3104  0.4253  2.7834

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.323638   0.341905   6.796 4.24e-11 ***
averel             0.036918   0.044141   0.836   0.4035
aveage            -0.006678   0.004603  -1.451   0.1477
aveeduc            0.073223   0.036947   1.982   0.0482 *
as.factor(race)2  -0.248488   0.166695  -1.491   0.1369
as.factor(race)3   0.532411   0.231986   2.295   0.0223 *
childs            -0.084480   0.037256  -2.268   0.0239 *
sex               -0.445562   0.099120  -4.495 9.28e-06 ***
sdage              0.012527   0.007348   1.705   0.0890 .
sdeduc            -0.078610   0.062211  -1.264   0.2072
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9239 on 375 degrees of freedom
  (5644 observations deleted due to missingness)
Multiple R-squared:   0.13, Adjusted R-squared:  0.1092
F-statistic: 6.229 on 9 and 375 DF,  p-value: 3.297e-08
```

6. *And give your best conclusion as to whether your initial hypothesis held up -and if not, why not.*

My conclusions have been drawn throughout this lab, but to summarize, my expectations were right in some places and wrong in others. I expected age to have a positive effect, but it had a near 0 negative effect. I expected education to have a positive effect, this was supported by my findings. I was unsure how religion would opperate but it seems that as a social circle realizes a unit increase in religion (ie: changing religions from Protestant to Catholic) the more invested in science websites one will state they are.

Data Index

SCI30   R USE SCIENCE SITE IN PAST 30 DAYS

Description of the Variable

772. In the past 30 days, how often have you visited a web site for: s. Science?

Percent	N	Value	Label
61.5	1,434	1	NEVER
19.8	462	2	1-2 TIMES
9.9	231	3	3-5 TIMES
8.8	205	4	MORE THAN 5 TIMES
	57,083	0	IAP
	3	8	DK
	181	9	NA
100.0	59,599		Total

Properties

Data type: numeric
Missing-data codes: 0,8,9
Mean: 1.66
Std Dev: .97
Record/column: 1/3264

AGE1   AGE OF PERSON #1

Description of the Variable

322. How old is [NAME]?

Percent	N	Value	Label
0.1	3	14	
0.1	3	15	
0.2	4	16	
0.3	7	17	
0.5	13	18	
0.5	12	19	
0.8	20	20	
0.8	20	21	
1.5	36	22	
1.2	29	23	
1.8	44	24	
2.2	53	25	
2.0	49	26	
1.9	47	27	
1.9	47	28	
1.6	40	29	
3.0	74	30	
2.0	48	31	

	2.0	46	31
	2.4	59	**32**

EDUC1   EDUCATION LEVEL OF PERSON #1

Description of the Variable

321. This card (BH) lists general levels of education. As far as you know, what is [NAME]'s highest level of education?

Percent	N	Value	Label
1.0	25	0	1-6 YEARS
4.8	116	1	7-9 YEARS
11.4	275	2	10-12 YEARS
31.6	760	3	H.S. GRAD
16.6	400	4	SOME COLLEGE
5.6	134	5	ASSO. DEGREE
16.7	401	6	BACH. DEGREE
12.3	297	7	GRAD OR PROF
	57,085	-1	IAP
	42	8	DK
	64	9	NA
100.0	59,599		**Total**

Properties

Data type: numeric
Missing-data codes: -1,8,9
Mean: 4.03
Std Dev: 1.79
Record/columns: 1/1229-1230

RELIG1   RELIGIOUS PREFERENCE OF PERSON #1

Description of the Variable

323. What is [NAME]'s religious preference? Is it Protestant, Catholic, Jewish, some other religion, or no religion?

Percent	N	Value	Label
57.0	1,307	1	PROTESTANT
26.8	615	2	CATHOLIC
2.7	63	3	JEWISH
8.8	202	4	NONE
4.7	107	5	OTHER
	57,085	0	IAP
	156	8	DK
	64	9	NA
100.0	59,599		Total

Properties

Data type:	numeric
Missing-data codes:	0,8,9
Mean:	1.77
Std Dev:	1.15
Record/column:	1/1249