1. First, we want to know if diversification is an effective strategy for venture capital firms.

(A) Create a plot, such as a scatter plot or loess curve, that indicates whether firms with more diversified portfolios have more successful investments or not. Define diversification as the number of industry areas a firm is invested in, divided by its total number of investments.

(B) Run a appropriate regression predicting the number of successful investments as a function of diversification. Is diversification related to having more successful investments?

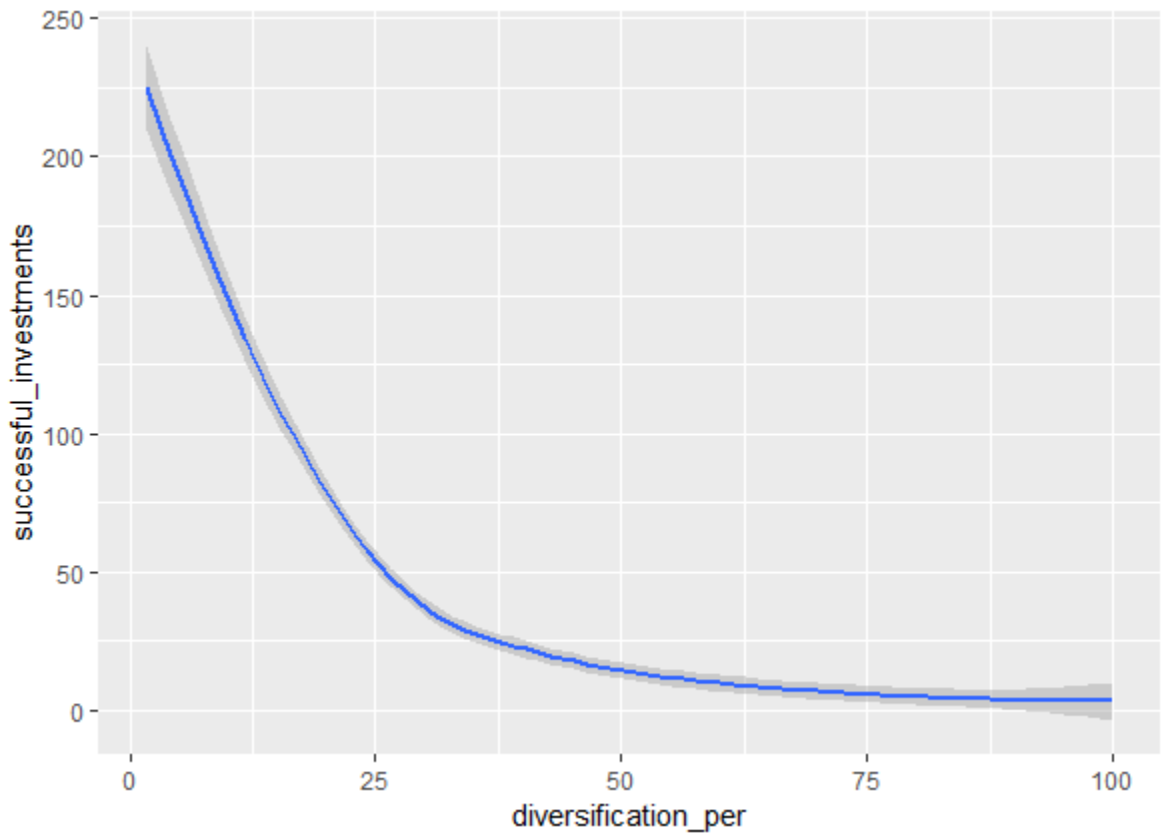
1a)

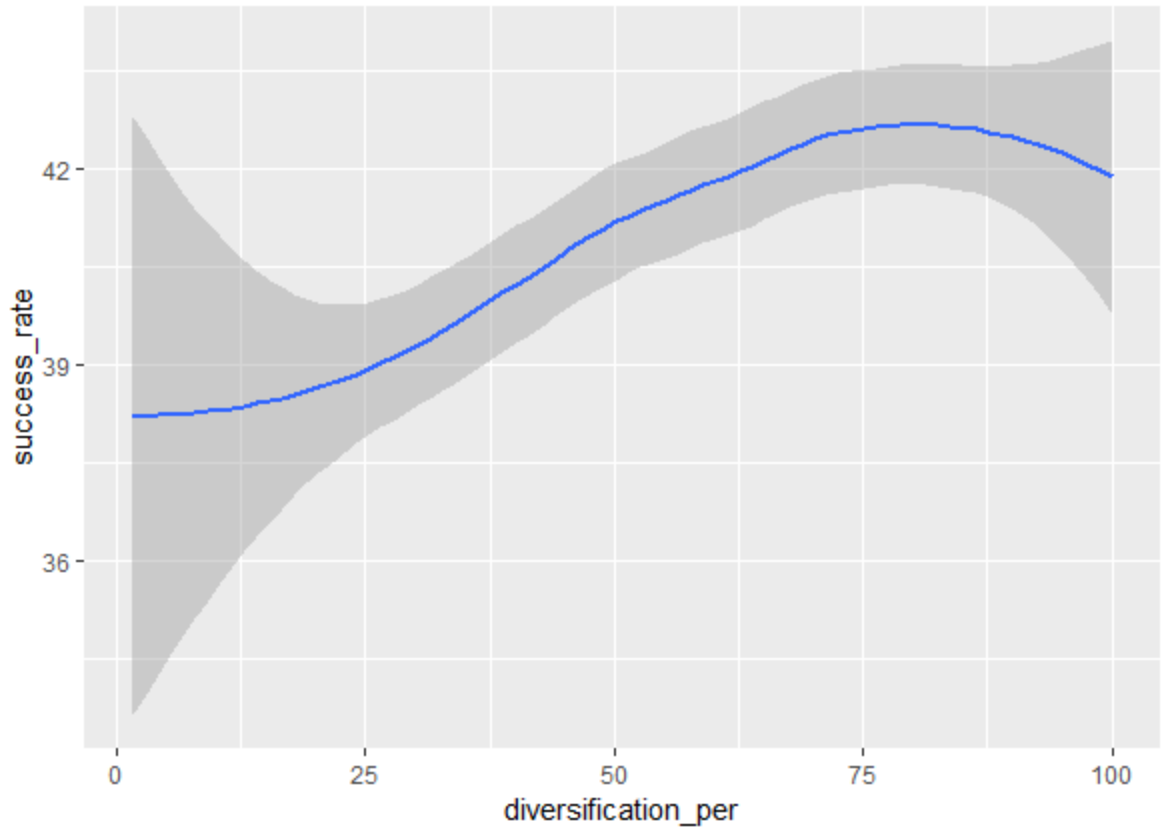
Data was cleaned to make sure results make sense. Investment firms with less than 5 investments have been excluded. Few observations were found where

number of successful investments > total number of investments

Such observations were deleted

A plot was created with success rate too.





Number of successful investments are reduced as diversification increases. But I feel number of investments need to be considered too. The success rate increases as diversification increases upto a point. Beyond this point, success rate starts to decrease again This makes a lot of sense.

1b) The regression results give further proof of our hypothesis

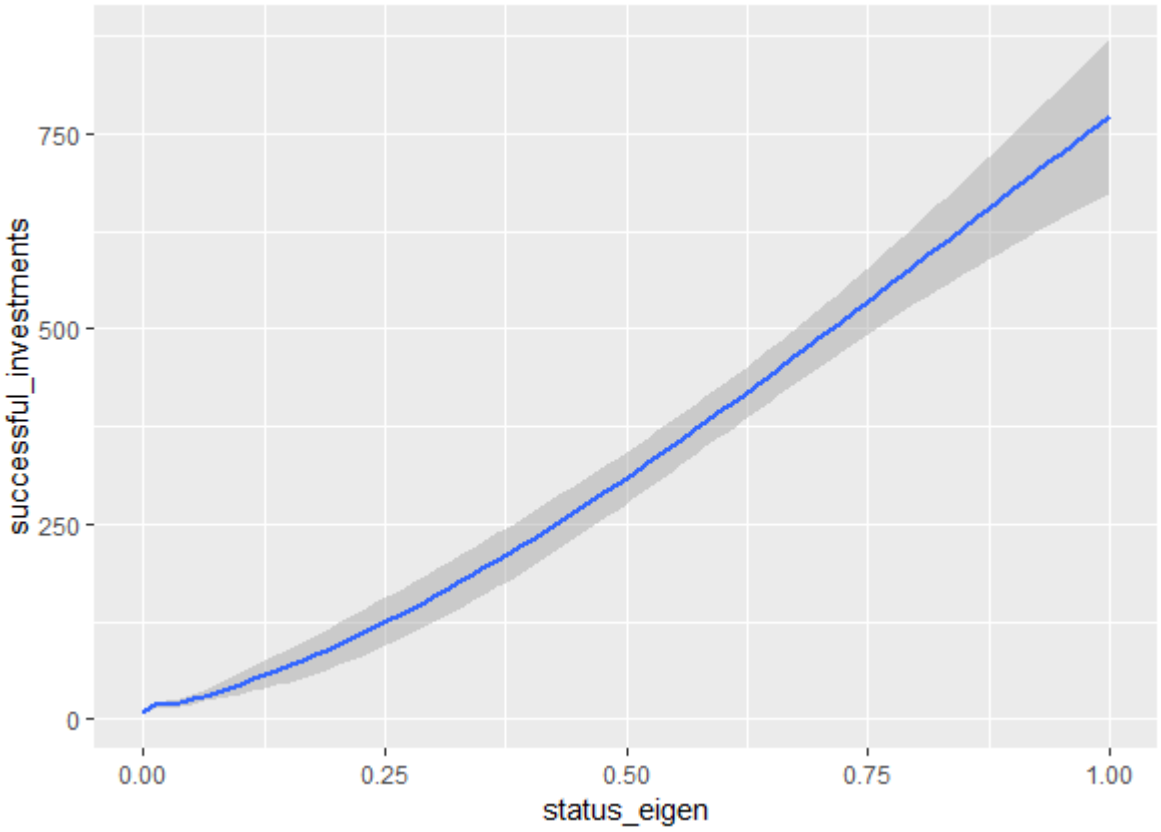
```
model1<-glm(successful_investments ~  diversification_per + number_of_deals, data=all_deals_5)

                          #Estimate Std. Error t value Pr(>|t|)
#(Intercept)             -8.183254   0.515263  -15.88   <2e-16 ***
#diversification_per    0.105583   0.008336   12.67   <2e-16 ***
#number_of_deals     0.497972   0.001140  436.88   <2e-16 ***
#model1<-glm(success_rate ~  diversification_per + number_of_deals, data=all_deals_5)
                          #Estimate Std. Error t value Pr(>|t|)
 # (Intercept)            36.826412   0.746259  49.348  < 2e-16 ***
 #diversification_per    0.071740   0.012073   5.942 2.93e-09 ***
 #number_of_deals     0.007961   0.001651   4.822 1.44e-06 ***
```

2)

(A) Create a plot, such as a scatterplot or loess curve, that indicates whether firms with higher status have more successful investments or not.

(B) Run a appropriate regression predicting the number of successful investments as a function of status. Is status related to having more successful investments?



A)

Graph indicates that firms with higher status generally have more successful investments

B)

model2<-glm(successful_investments ~ status_eigen + number_of_deals, family="poisson", data=status_vs_success)

#Estimate Std. Error t value Pr(>|t|)

#(Intercept) 2.715e+00 2.930e-03 926.5 <2e-16 ***

#status_eigen 4.702e+00 1.321e-02 355.9 <2e-16 ***

#number_of_deals 2.847e-04 1.209e-06 235.5 <2e-16 ***

The positive co-efficients show that number of successful investments is positively dependent on the status

3)

(A) Run a appropriate regression predicting the number of successful investments as a function of the interaction of status and diversification. Is this interaction related to having more successful investments?

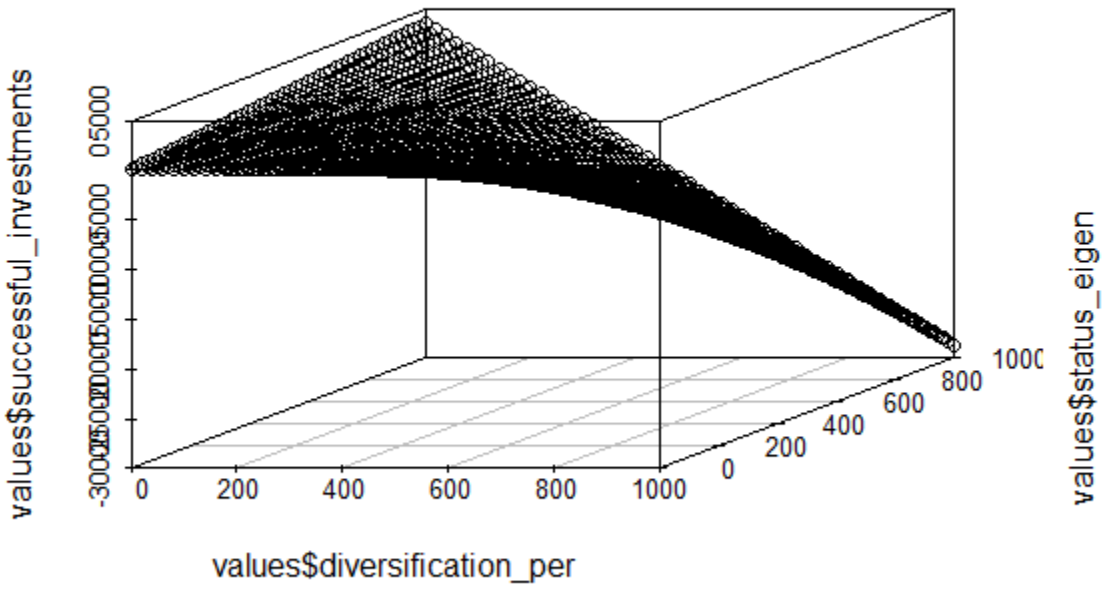
```
model3 = glm(successful_investments ~ diversification_per + status_eigen +
diversification_per*status_eigen, family="poisson",data=status_vs_success)
```

	#Estimate	Std. Error	t value	Pr(>|t|)
#diversification_per	-0.0370097	0.0001534	-241.23	<2e-16 ***
#status_eigen	3.6518521	0.0235325	155.18	<2e-16 ***
# diversification_per:status_eigen	-0.0322797	0.0019138	-16.87	<2e-16 ***

The results show a negative effect. This does not make sense. Maybe number of investments needs to be added as a control. The results need to be looked at more deeply.

B)

Create a plot that illustrates the interaction effect predicted by the regression. We can accomplish this using a 3D scatterplot generated by the fitted values of the model.



4)

```
model4 = multinom(startup_state ~ diversification_per + status_eigen + divers
ification_per*status_eigen, final_merge)
z = summary(model4)$coefficients/summary(model4)$standard.errors
(1 - pnorm(abs(z), 0, 1)) * 2
```

```
                         (Intercept) diversification_per status_eigen
Generating Revenue     0.000000e+00        5.345888e-07   0.00000000
Other                  0.000000e+00        1.097472e-09   0.01644662
Profitable             0.000000e+00        0.000000e+00   0.00000000
Ramp-Up/Clinical Trial 2.692291e-12        6.685355e-05   0.03564259

                                  diversification_per:status_eigen
Generating Revenue                                2.666363e-08
Other                                             5.107026e-15
Profitable                                        0.000000e+00
Ramp-Up/Clinical Trial                            5.107026e-14
```

Could not make sense of the results. Computed logistic regression to see what factors lead to business failing

```
model_out_of_business<-glm(out_of_business ~ status_eigen + diversification_p
er+status_eigen*diversification_per , data = final_merge, family = "binomial"
)
summary(model_out_of_business)
```

```
                                    #Estimate Std. Error  z value Pr(>|z|)
#(Intercept)                      -2.8136656  0.0201501 -139.635   <2e-16 ***
#status_eigen                      0.8514719  0.0522857   16.285   <2e-16 ***
#diversification_per               0.0036134  0.0004141    8.726   <2e-16 ***
#status_eigen:diversification_per -0.0024100  0.0039547   -0.609    0.542
```

This shows that firms which do not have a synergy of diversification and status might lead to failure