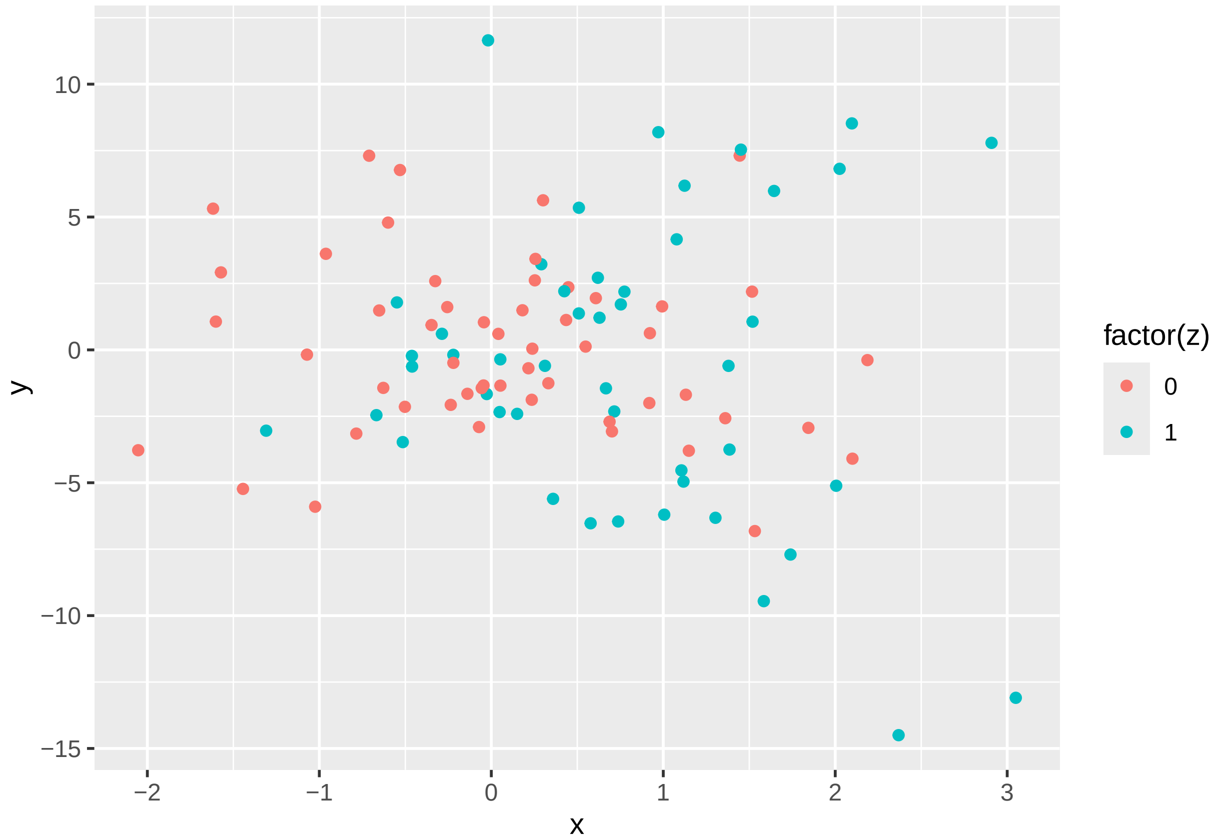
# ROS_CHP10

Michael Rowlands

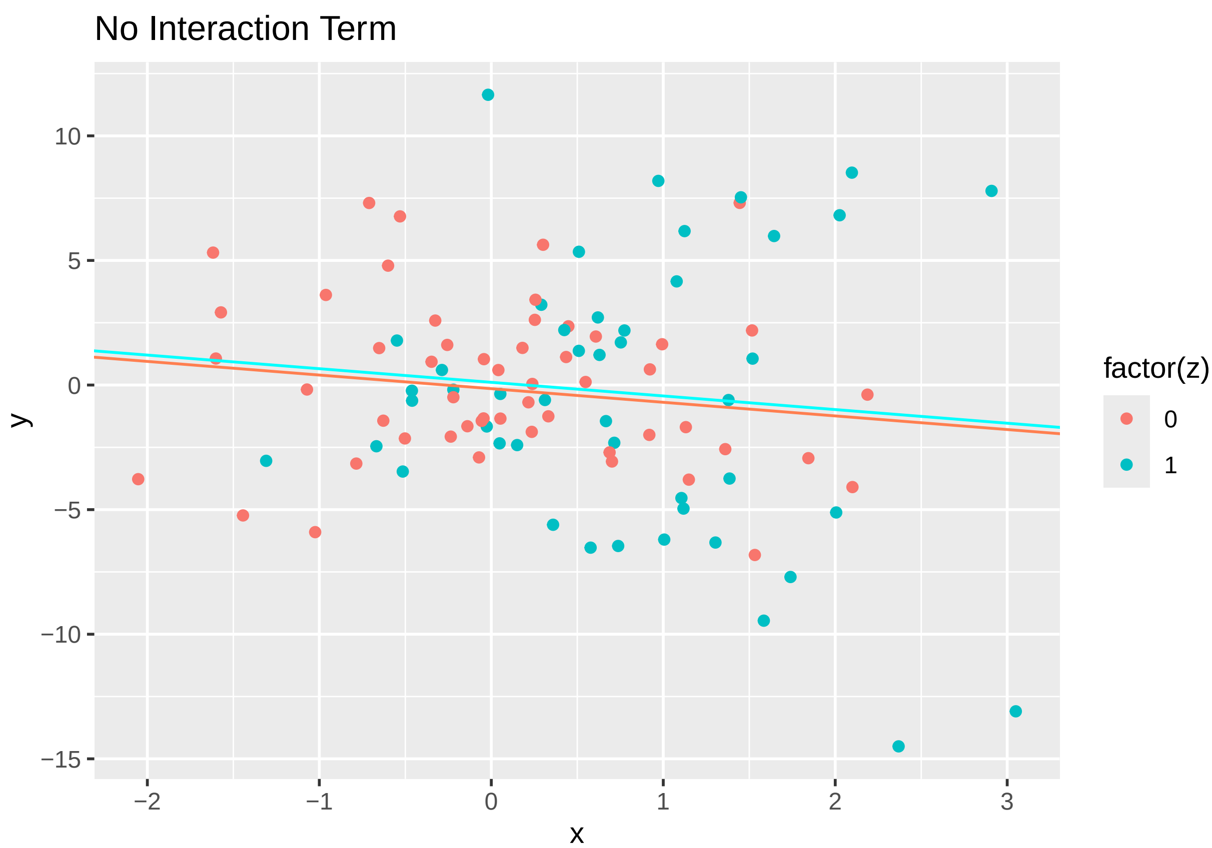
2024-04-18

## Question 1

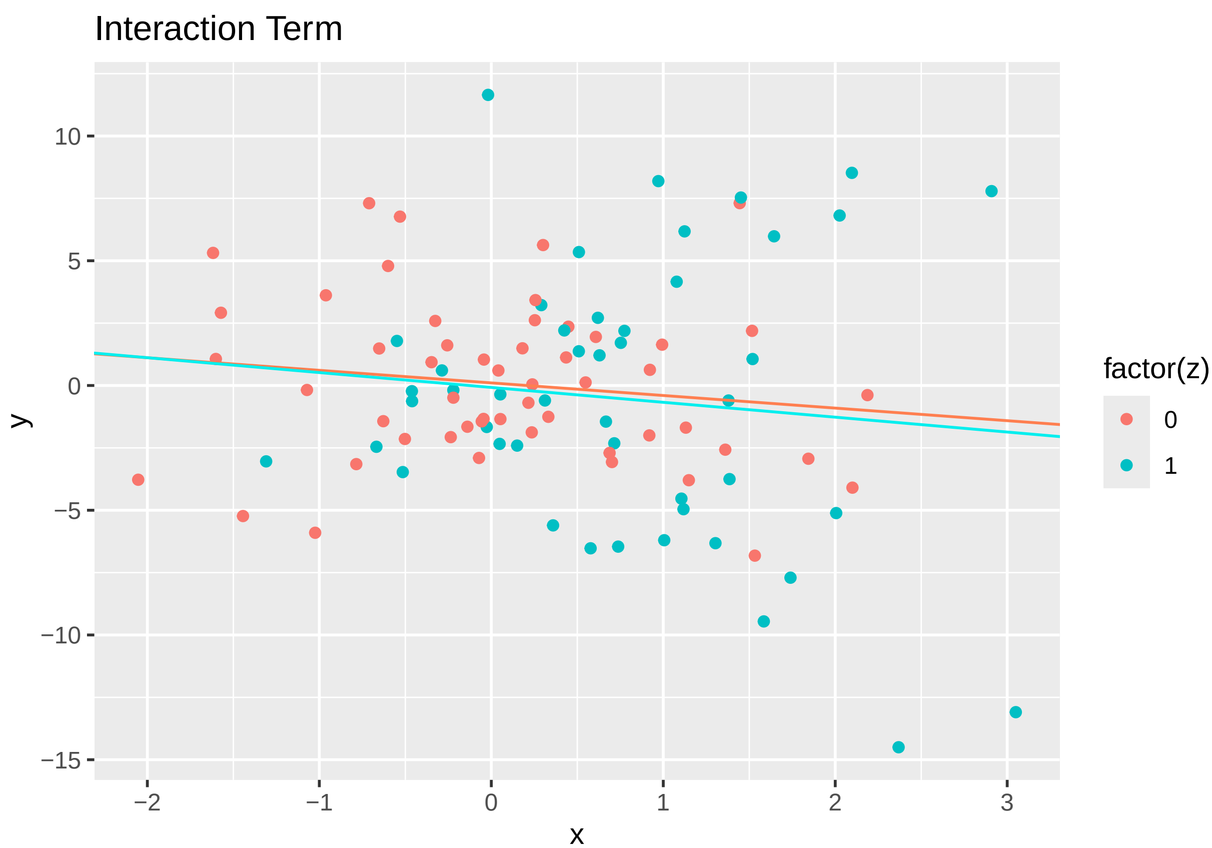
```
z <- rbinom(n=100,size=1,prob=.5)
x <- rnorm(n=100, mean=z, sd=1)
b <- c(1,2,-1,-2)

design <- matrix(c(x,z,x*z), ncol=3)

y = rowSums(design*b) + rnorm(100,0,3)

data <- data.frame(x,z,y)

data %>% ggplot(aes(x,y)) + geom_point(aes(color=factor(z)))
```




```
## [1] "===============No Interaction Term==============="

## stan_glm
##  family:       gaussian [identity]
##  formula:      y ~ x + z
##  observations: 100
##  predictors:   3
## ------
##             Median MAD_SD
## (Intercept)  0.1    0.6
## x           -0.5    0.5
## z           -0.3    1.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 4.5    0.3
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```




```
## [1] "===============With Interaction Term==============="

## stan_glm
##  family:       gaussian [identity]
##  formula:      y ~ x + z + z:x
##  observations: 100
##  predictors:   4
## ------
##             Median MAD_SD
## (Intercept)  0.1    0.6
## x           -0.5    0.6
## z           -0.2    1.1
## x:z         -0.1    1.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 4.5    0.3
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```




# Question 3

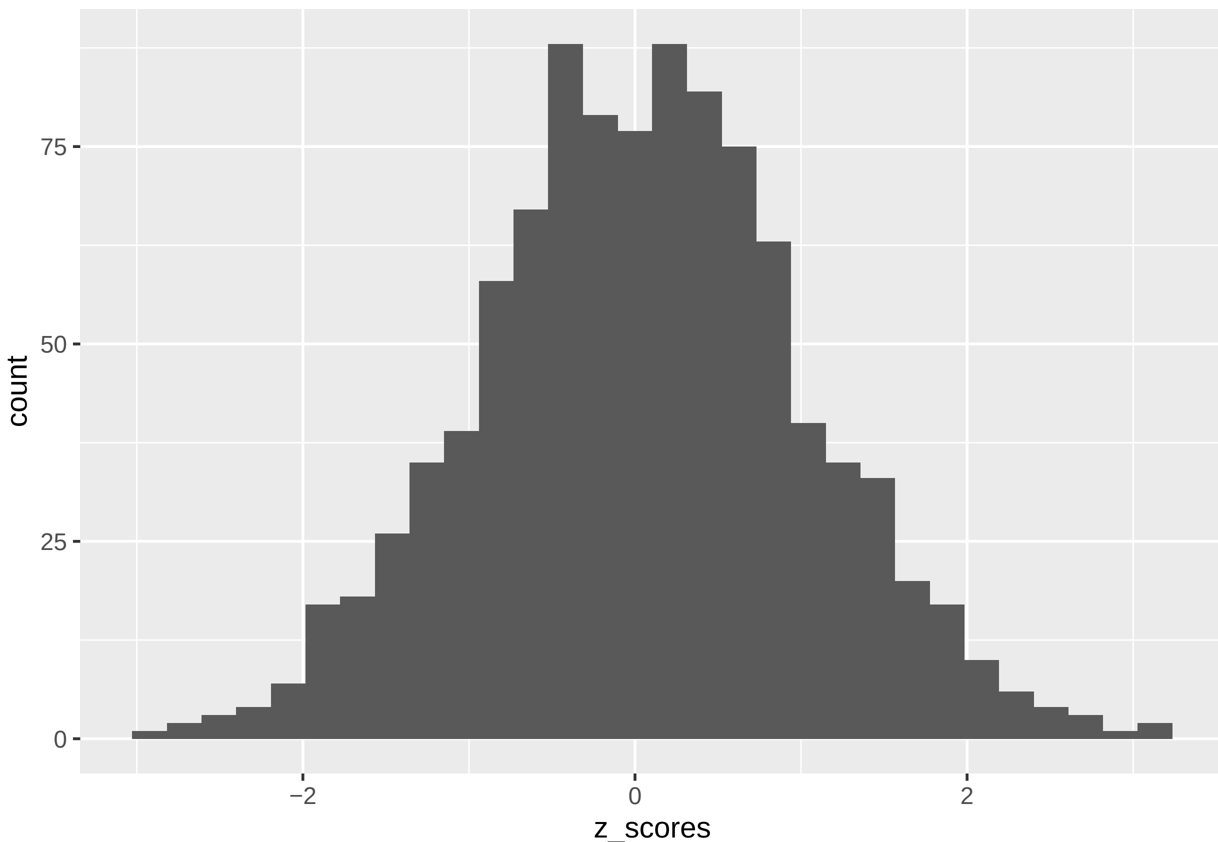
```
var1 <- rnorm(1000,0,1)
var2<- rnorm(1000,0,1)
data_3 <- data.frame(var1,var2)

fit_3 <- stan_glm(var1 ~ var2, prior = NULL, prior_intercept = NULL, prior_aux = NULL, data=data_3, ref
print(fit_3)
```

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      var1 ~ var2
##  observations: 1000
##  predictors:   2
## ------
##             Median MAD_SD
## (Intercept)  0.0    0.0
## var2        -0.1    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 1.0    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

# Question 4

```
z_scores <- c()

for (i in 1:1000){
  var1 <- rnorm(1000,0,1)
  var2 <- rnorm(1000,0,1)
  data_temp <- data.frame(var1, var2)

  fit_temp <- stan_glm(var1 ~ var2, prior = NULL, prior_intercept = NULL, prior_aux = NULL, data=data_te
  slope <- coef(fit_temp)[2]
  mad <- se(fit_temp)[2]
  z_scores <- c(z_scores, slope/mad)

}

ggplot() + aes(z_scores) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

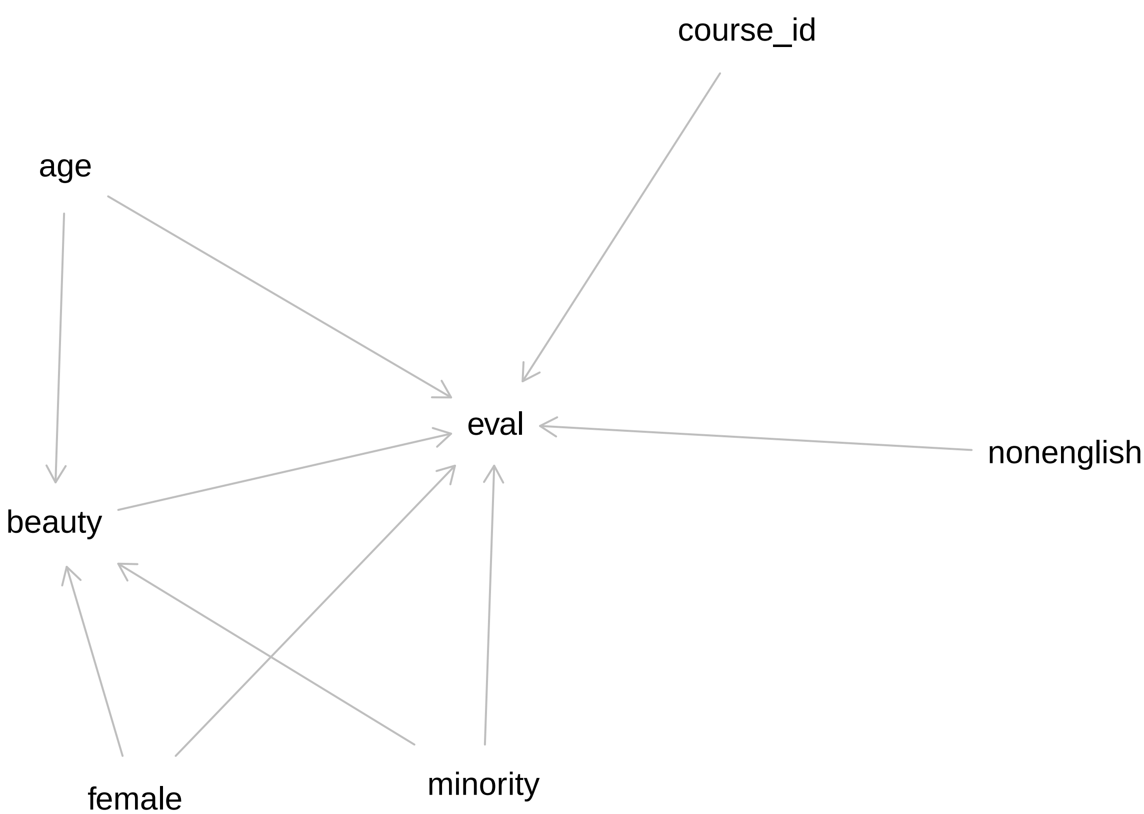



The proportion of significant results is 0.042.

# Question 6

We are interested in evaluating the effect of beauty on course evaluations. We will load in the data and propose a DAG which seems to make sense (at least enough for a practice problem) and find the sufficent adjustment set.

```
## 'data.frame':    463 obs. of  8 variables:
##  $ eval      : num  4.3 4.5 3.7 4.3 4.4 4.2 4 3.4 4.5 3.9 ...
##  $ beauty    : num  0.202 -0.826 -0.66 -0.766 1.421 ...
##  $ female    : int  1 0 0 1 1 0 1 1 1 0 ...
##  $ age       : int  36 59 51 40 31 62 33 51 33 47 ...
##  $ minority  : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ nonenglish: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ lower     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ course_id : int  3 0 4 2 0 0 4 0 0 4 ...
```



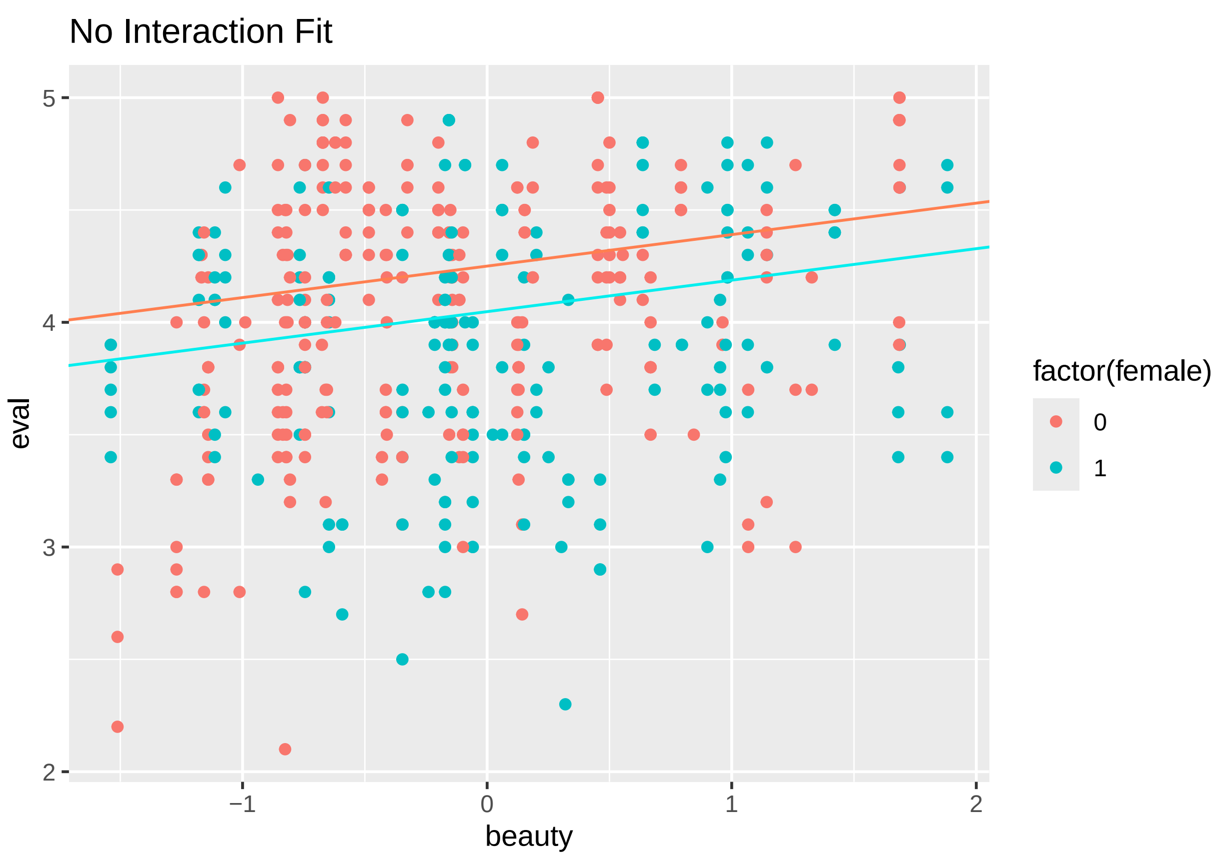

The backdoor criterion would suggest that the sufficient adjustment set to find the direct effect of beauty on evaluation under this DAG is female, age, and minority. To check,

```
print(adjustmentSets(g1, 'beauty', 'eval', effect='direct'))
```

```
## { age, female, minority }
```

We will now fit a model with this adjustment set, but with the functional assumption of no interaction effects.

```
## stan_glm
##  family:       gaussian [identity]
##  formula:      eval ~ beauty + female + age + minority
##  observations: 463
##  predictors:   5
## ------
##             Median MAD_SD
## (Intercept)  4.3    0.1
## beauty       0.1    0.0
## female      -0.2    0.1
## age          0.0    0.0
## minority    -0.1    0.1
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.5    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```




From the model we determine that for a non-minority aged 0 man of average beauty we would expect an evaluation of 4.25. This clearly does not make sense in context since an aged 0 person will not be teaching a college course, so we do not place any stock in its interpretation. More importantly, we determine that every increase in beauty by 1 causes an increase of 0.14 in evaluations under the assumptions listed above.

We will now perform the same analysis with the possibility of interaction effects in our model.

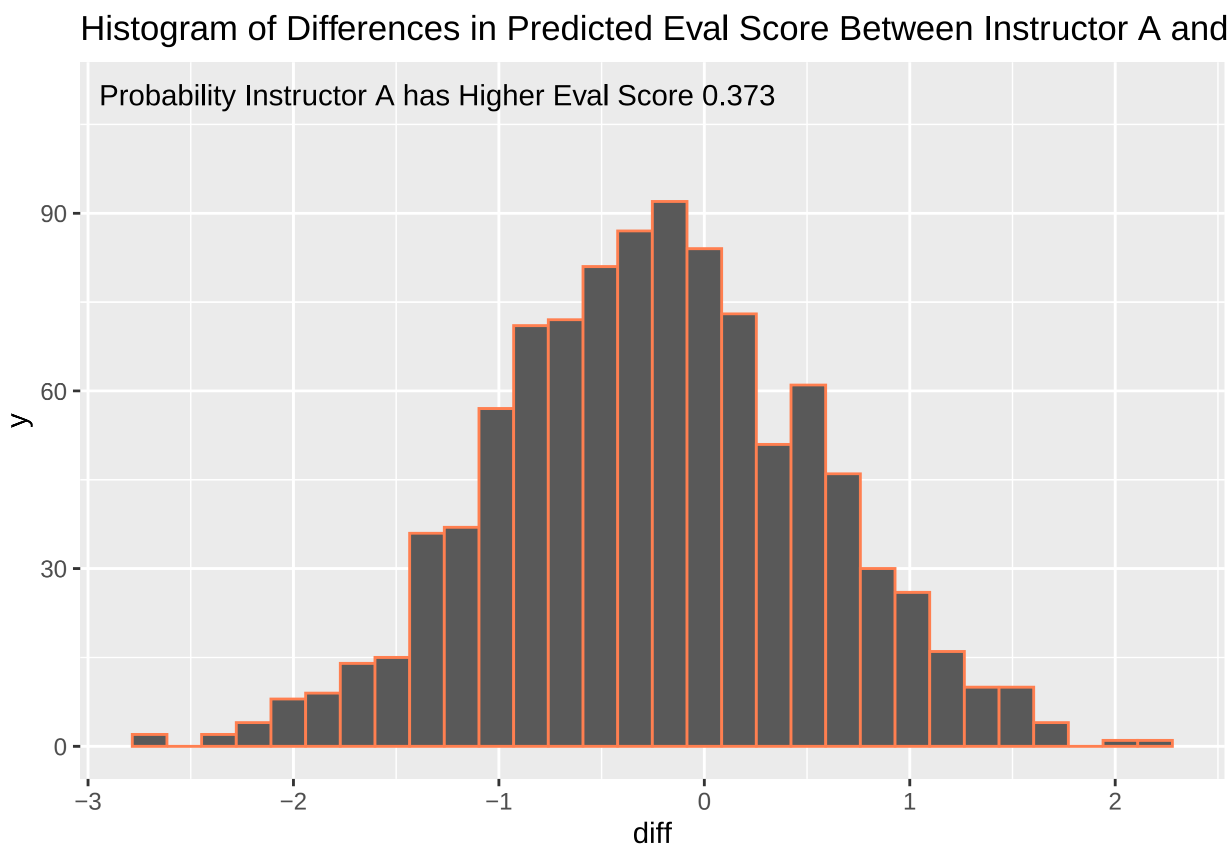
```
## stan_glm
##  family:       gaussian [identity]
##  formula:      eval ~ beauty + age + female + minority + beauty:female + beauty:age
##  observations: 463
##  predictors:   7
## ------
##               Median MAD_SD
## (Intercept)    4.2    0.1
## beauty        -0.4    0.2
## age            0.0    0.0
## female        -0.2    0.1
## minority      -0.1    0.1
## beauty:female  0.0    0.1
## beauty:age     0.0    0.0
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 0.5    0.0
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

# Question 7

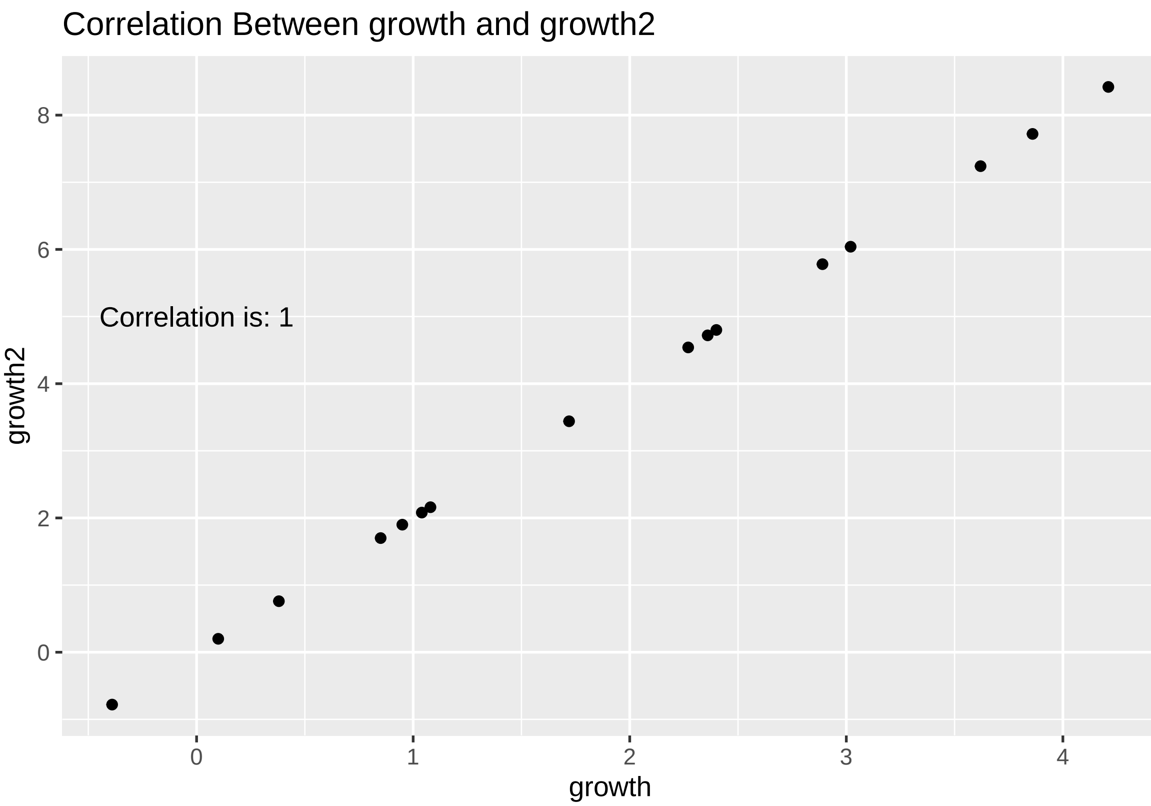
We will use the second model, the model with interaction terms, from question 6 in order to predict the evaluation scores of people with the defined characteristics in the question.

```
inst_a <- data.frame(beauty=-1,age=50, minority=0, female=1)
inst_b <- data.frame(beauty=-0.5,age=60, minority=0, female=0)

a_pred <- posterior_predict(fit_6b, newdata=inst_a, draws=1000)
b_pred <- posterior_predict(fit_6b, newdata=inst_b, draws=1000)

diff <- a_pred - b_pred
prob_a <- sum(diff>0)/1000

ggplot() + aes(diff) + geom_histogram(color='coral') + annotate('text', x=-1.3, y=110, label=paste('Pro
```

```
## ‘stat_bin()‘ using ‘bins = 30‘. Pick better value with ‘binwidth‘.
```


Histogram of Differences in Predicted Eval Score Between Instructor A and
Probability Instructor A has Higher Eval Score 0.373
90
60
30
0
y
−3
−2
−1
0
1
2
diff

## Question 9




A)
B)

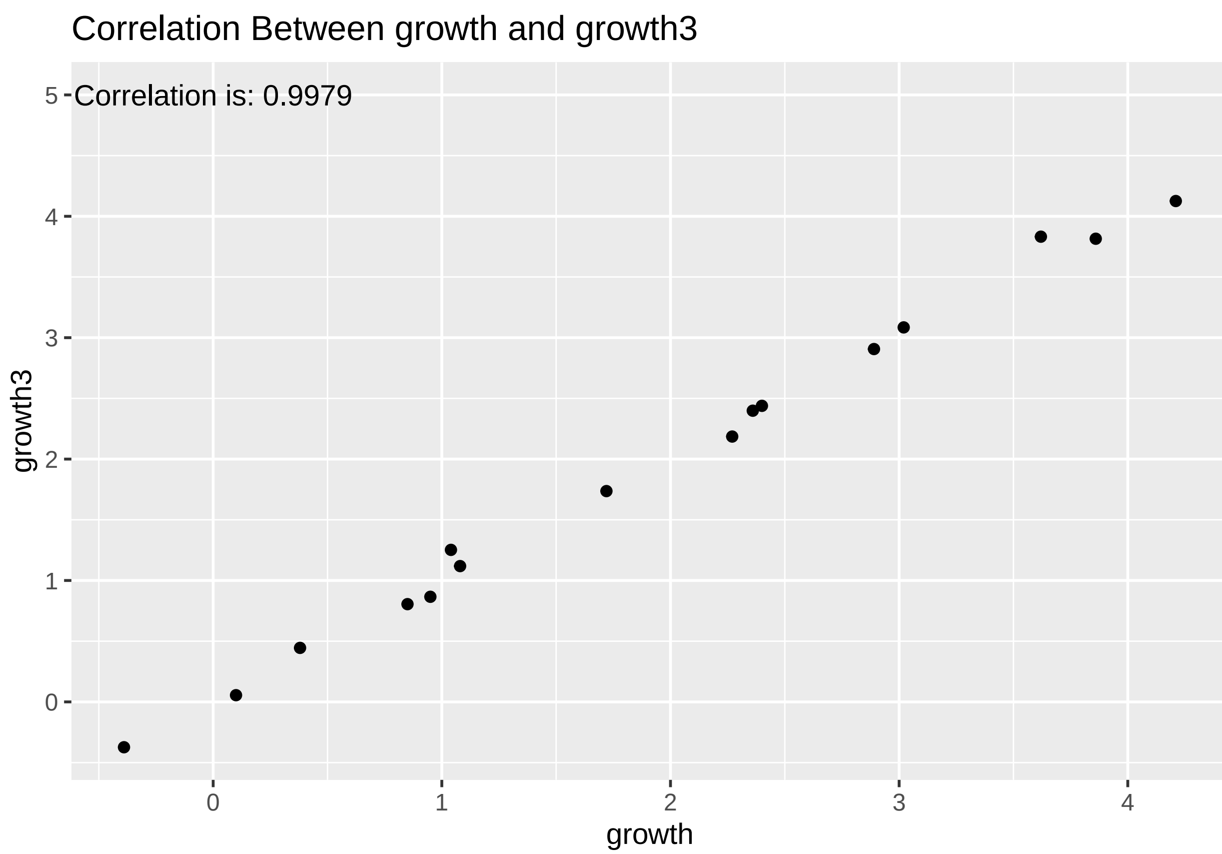
```
## [1] "============Fit From CHP 7=============="

## stan_glm
##  family:       gaussian [identity]
##  formula:      vote ~ growth
##  observations: 16
##  predictors:   2
## ------
##             Median MAD_SD
## (Intercept) 46.3    1.7
## growth       3.0    0.7
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 3.9    0.7
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg

## [1] "============Fit With Collinearity=============="

## stan_glm
```

```
##  family:       gaussian [identity]
##  formula:      vote ~ growth + growth2
##  observations: 16
##  predictors:   3
## ------
##             Median MAD_SD
## (Intercept) 46.3    1.6
## growth       1.3    6.8
## growth2      0.9    3.4
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 3.9    0.7
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

We see that the MAD_SD values for both growth and growth2 are much larger than the growth MAD_SD value in the fit from CHP 7. The earlier model was very sure about the relationship of economic growth and vote percentage in the earlier model, it being positive. The model with collinearity is much less sure; both point estimates for the coefficients on the growth terms are positive, but with the MAD_SD values being very large they could very well be negative from the models understanding.

C)

```
## Warning in hibbs$growth + rnorm(n = length(hibbs), 0, 0.1): longer object length
## is not a multiple of shorter object length
```




```
## [1] "============Fit With Collinearity (but less)============="

## stan_glm
##  family:       gaussian [identity]
##  formula:      vote ~ growth + growth3
##  observations: 16
##  predictors:   3
## ------
##             Median MAD_SD
## (Intercept) 46.2    1.7
## growth       0.8    5.9
## growth3      2.4    5.8
##
## Auxiliary parameter(s):
##       Median MAD_SD
## sigma 3.9    0.8
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

We see similar results from part B.