final_project_knitattempt2

jenny

2025-05-14

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#Import my HINTS6 data
hints6_public <- read_sav("//Users/jennymai/Desktop/SOC542/Final Project/hints6_public.sav")
```

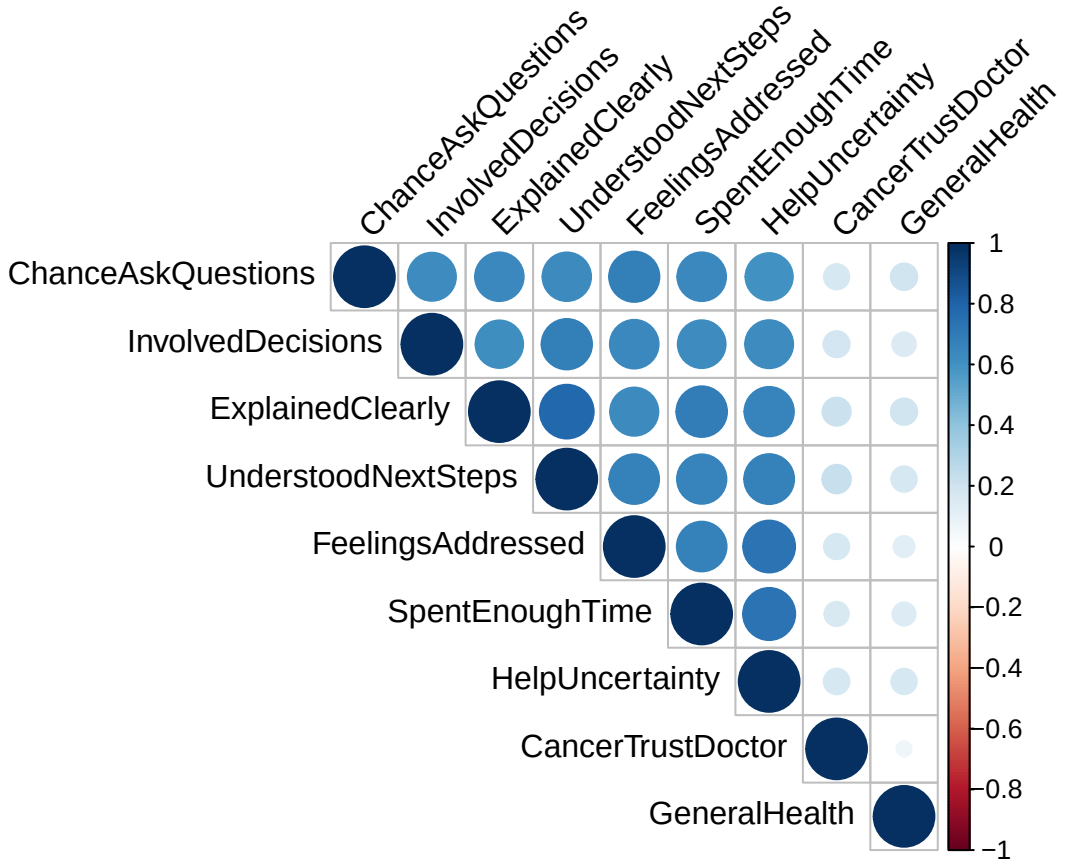
Data Cleaning

```
#Bivariate correlation
library(corrplot)

vars <- cancer_data_reverse %>%
  dplyr::select(ChanceAskQuestions, InvolvedDecisions, ExplainedClearly, UnderstoodNextSteps, FeelingsA
cor_matrix <- cor(vars, use = "pairwise.complete.obs")
round(cor_matrix, 2)
```

```
##                     ChanceAskQuestions InvolvedDecisions ExplainedClearly
## ChanceAskQuestions                1.00              0.63             0.64
## InvolvedDecisions                 0.63              1.00             0.61
## ExplainedClearly                  0.64              0.61             1.00
## UnderstoodNextSteps               0.64              0.69             0.78
## FeelingsAddressed                 0.69              0.64             0.63
## SpentEnoughTime                   0.64              0.62             0.70
## HelpUncertainty                   0.61              0.63             0.66
## CancerTrustDoctor                 0.17              0.19             0.22
## GeneralHealth                     0.19              0.15             0.19
##                     UnderstoodNextSteps FeelingsAddressed SpentEnoughTime
## ChanceAskQuestions                 0.64              0.69            0.64
## InvolvedDecisions                  0.69              0.64            0.62
## ExplainedClearly                   0.78              0.63            0.70
## UnderstoodNextSteps                1.00              0.67            0.66
## FeelingsAddressed                  0.67              1.00            0.67
## SpentEnoughTime                    0.66              0.67            1.00
## HelpUncertainty                    0.68              0.73            0.74
## CancerTrustDoctor                  0.22              0.17            0.16
```

```
## GeneralHealth                    0.18              0.12           0.15
##                     HelpUncertainty CancerTrustDoctor GeneralHealth
## ChanceAskQuestions             0.61              0.17          0.19
## InvolvedDecisions              0.63              0.19          0.15
## ExplainedClearly               0.66              0.22          0.19
## UnderstoodNextSteps            0.68              0.22          0.18
## FeelingsAddressed              0.73              0.17          0.12
## SpentEnoughTime                0.74              0.16          0.15
## HelpUncertainty                1.00              0.18          0.18
## CancerTrustDoctor              0.18              1.00          0.06
## GeneralHealth                  0.18              0.06          1.00
```

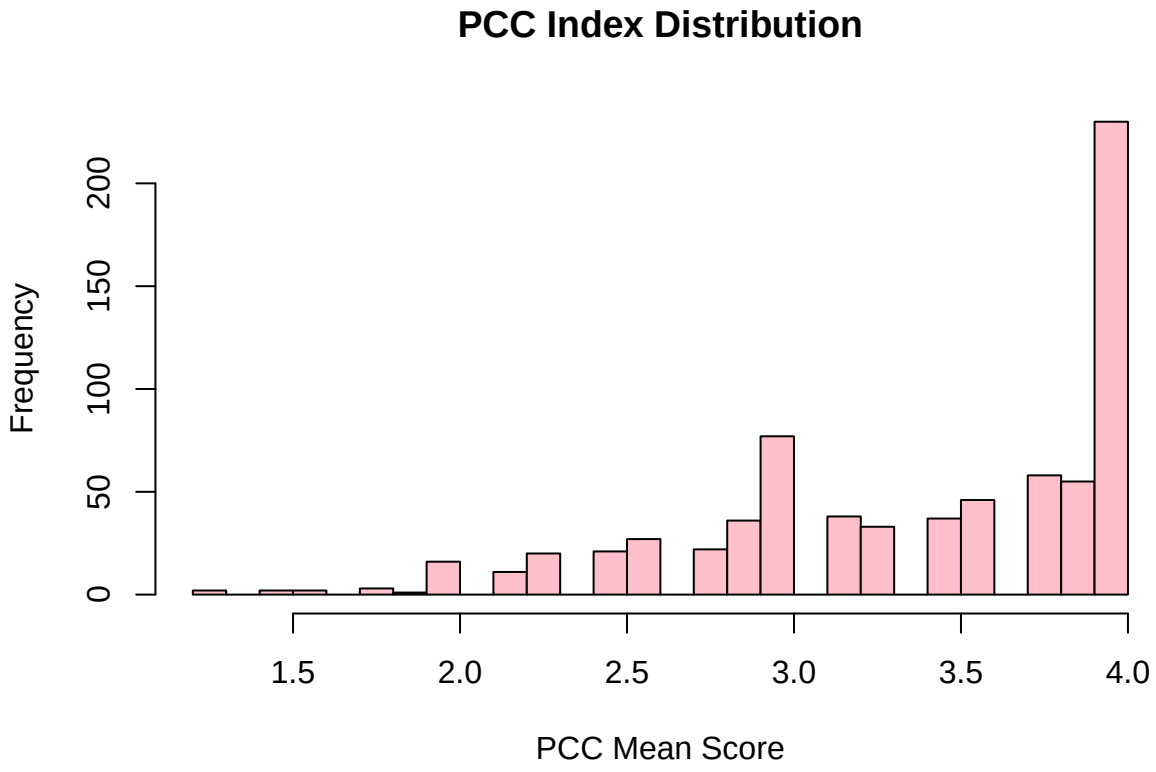
```r
corrplot(cor_matrix, method = "circle", type = "upper",
         tl.col = "black", tl.srt = 45)
```




```r
#IV - Create PCC index (make mean score of every observation)
cancer_data_reverse <- cancer_data_reverse %>%
  mutate(
    pcc_index_mean = rowMeans(dplyr::select(., ChanceAskQuestions, InvolvedDecisions, ExplainedClearly,
  )

#Descriptives for the PCC index (M = 3.389)
summary(cancer_data_reverse$pcc_index_mean)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.286   3.000   3.571   3.389   4.000   4.000
```

```r
#Histogram of the means from the PCC index -- skewed to the right
hist(cancer_data_reverse$pcc_index_mean,
     main = "PCC Index Distribution",
     xlab = "PCC Mean Score", col = "pink", breaks = 20)
```




```r
#Cronbach's alpha (a = 0.88)
library(psych)
alpha(vars)
```

```
##
## Reliability analysis
## Call: alpha(x = vars)
##
##   raw_alpha std.alpha G6(smc) average_r S/N    ase mean   sd median_r
##       0.88      0.88     0.9      0.46 7.6 0.0065  3.4 0.53     0.62
##
##     95% confidence boundaries
##          lower alpha upper
## Feldt     0.86  0.88  0.89
## Duhachek  0.86  0.88  0.89
##
##  Reliability if an item is dropped:
##                     raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r
## ChanceAskQuestions       0.85      0.86    0.88      0.44 6.2   0.0078 0.068
## InvolvedDecisions        0.85      0.86    0.88      0.44 6.3   0.0078 0.068
```

```
## ExplainedClearly         0.85      0.86    0.87      0.43 6.0   0.0078 0.066
## UnderstoodNextSteps      0.85      0.86    0.87      0.43 6.0   0.0080 0.064
## FeelingsAddressed        0.85      0.86    0.87      0.43 6.1   0.0081 0.064
## SpentEnoughTime          0.85      0.86    0.87      0.43 6.1   0.0081 0.064
## HelpUncertainty          0.85      0.86    0.87      0.43 6.1   0.0083 0.065
## CancerTrustDoctor        0.89      0.90    0.91      0.54 9.4   0.0063 0.050
## GeneralHealth            0.91      0.91    0.91      0.55 9.6   0.0044 0.046
##                       med.r
## ChanceAskQuestions     0.62
## InvolvedDecisions      0.62
## ExplainedClearly       0.61
## UnderstoodNextSteps    0.61
## FeelingsAddressed      0.61
## SpentEnoughTime        0.61
## HelpUncertainty        0.62
## CancerTrustDoctor      0.64
## GeneralHealth          0.64
##
##  Item statistics
##                         n raw.r std.r r.cor r.drop mean   sd
## ChanceAskQuestions    737  0.80  0.80  0.78   0.74  3.5 0.65
## InvolvedDecisions     737  0.79  0.80  0.77   0.72  3.4 0.73
## ExplainedClearly      737  0.83  0.84  0.83   0.78  3.6 0.60
## UnderstoodNextSteps   737  0.84  0.85  0.85   0.79  3.5 0.68
## FeelingsAddressed     737  0.83  0.82  0.81   0.76  3.3 0.80
## SpentEnoughTime       737  0.83  0.82  0.81   0.76  3.3 0.81
## HelpUncertainty       737  0.84  0.83  0.83   0.78  3.2 0.87
## CancerTrustDoctor     737  0.32  0.37  0.23   0.22  3.8 0.48
## GeneralHealth         737  0.39  0.34  0.21   0.19  3.2 0.97
##
## Non missing response frequency for each item
##                           1    2    3    4    5 miss
## ChanceAskQuestions     0.01 0.07 0.32 0.61 0.00    0
## InvolvedDecisions      0.01 0.10 0.33 0.55 0.00    0
## ExplainedClearly       0.00 0.06 0.32 0.62 0.00    0
## UnderstoodNextSteps    0.01 0.08 0.31 0.60 0.00    0
## FeelingsAddressed      0.02 0.16 0.35 0.47 0.00    0
## SpentEnoughTime        0.03 0.14 0.36 0.47 0.00    0
## HelpUncertainty        0.04 0.18 0.34 0.44 0.00    0
## CancerTrustDoctor      0.00 0.03 0.17 0.80 0.00    0
## GeneralHealth          0.04 0.20 0.37 0.31 0.08    0
```

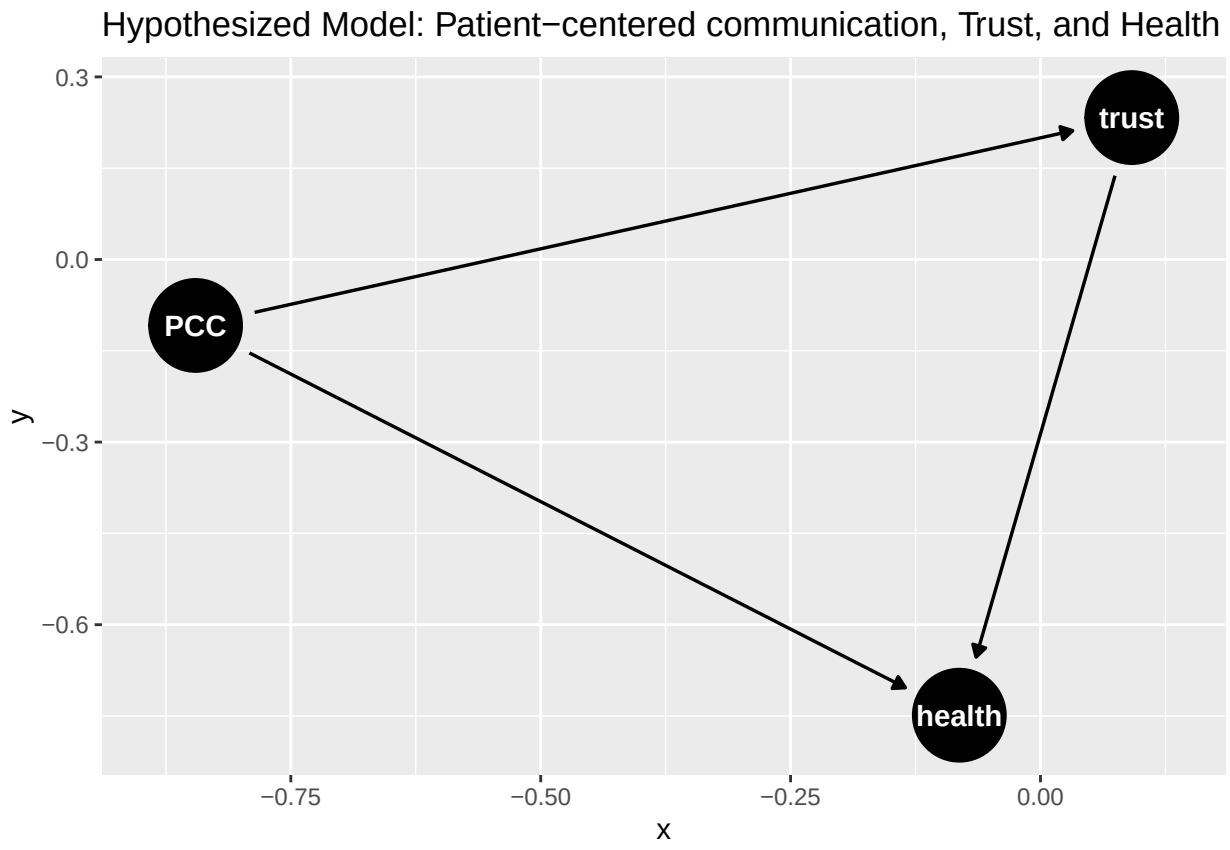
#Descriptive of demographics

```
#New data frame with ONLY the PCC index variable, trust, general health, and control variables (n = 740
cancer_d_reverse_index <- cancer_data_reverse %>%
  dplyr::select(pcc_index_mean, CancerTrustDoctor, GeneralHealth, Age, RaceEthn, BirthGender, Education

#Descriptive
datasummary_skim(dplyr::select(cancer_d_reverse_index,
                               pcc_index_mean, CancerTrustDoctor, GeneralHealth,
                               Age, RaceEthn, BirthGender, Education))
```

	Unique	Missing Pct.	Mean	SD	Min	Median	Max
pcc_index_mean	20	0	3.4	0.6	1.3	3.6	4.0
CancerTrustDoctor	4	0	3.8	0.5	1.0	4.0	4.0
GeneralHealth	5	0	3.2	1.0	1.0	3.0	5.0
R1. What is your age?	67	0	68.0	12.2	19.0	69.0	98.0
		N	%				
RaceEthn	Hispanic	60	8.1				
	Non-Hispanic White	566	76.8				
	Black or African American	76	10.3				
	American Indian or Alaska Native	2	0.3				
	Asian	15	2.0				
	Native Hawaiian or other Pacific Islander	0	0.0				
	Multiple Races	18	2.4				
BirthGender	Female	431	58.5				
	Male	306	41.5				
Education	College graduate	180	24.4				
	High school or less	155	21.0				
	Postgraduate	171	23.2				
	Some college	176	23.9				
	Vocational school	55	7.5				

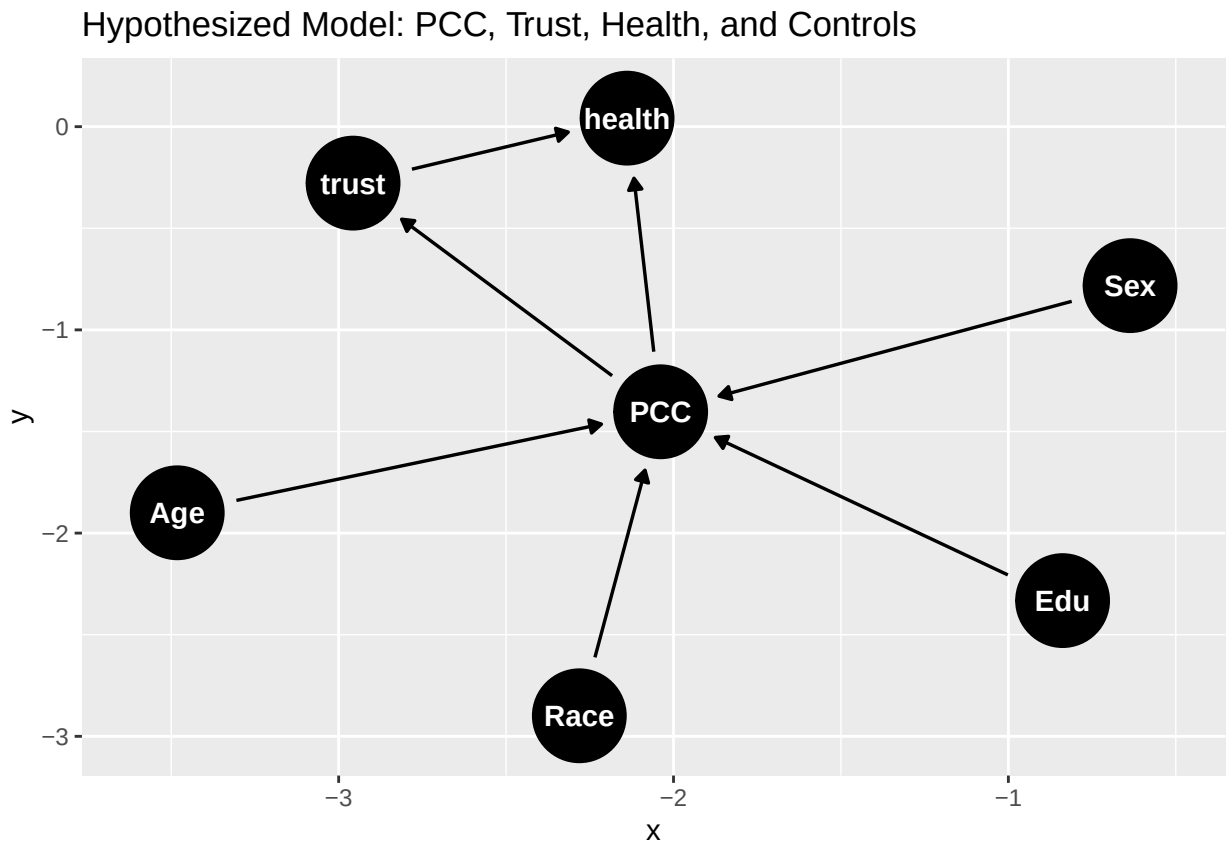
DAG

```
library(dagitty)
library(ggdag)
dag_model <- dagitty('dag {
  PCC -> trust
  trust -> health
  PCC -> health
}')

ggdag(dag_model, text = TRUE) +
  ggtitle("Hypothesized Model: Patient-centered communication, Trust, and Health")
```

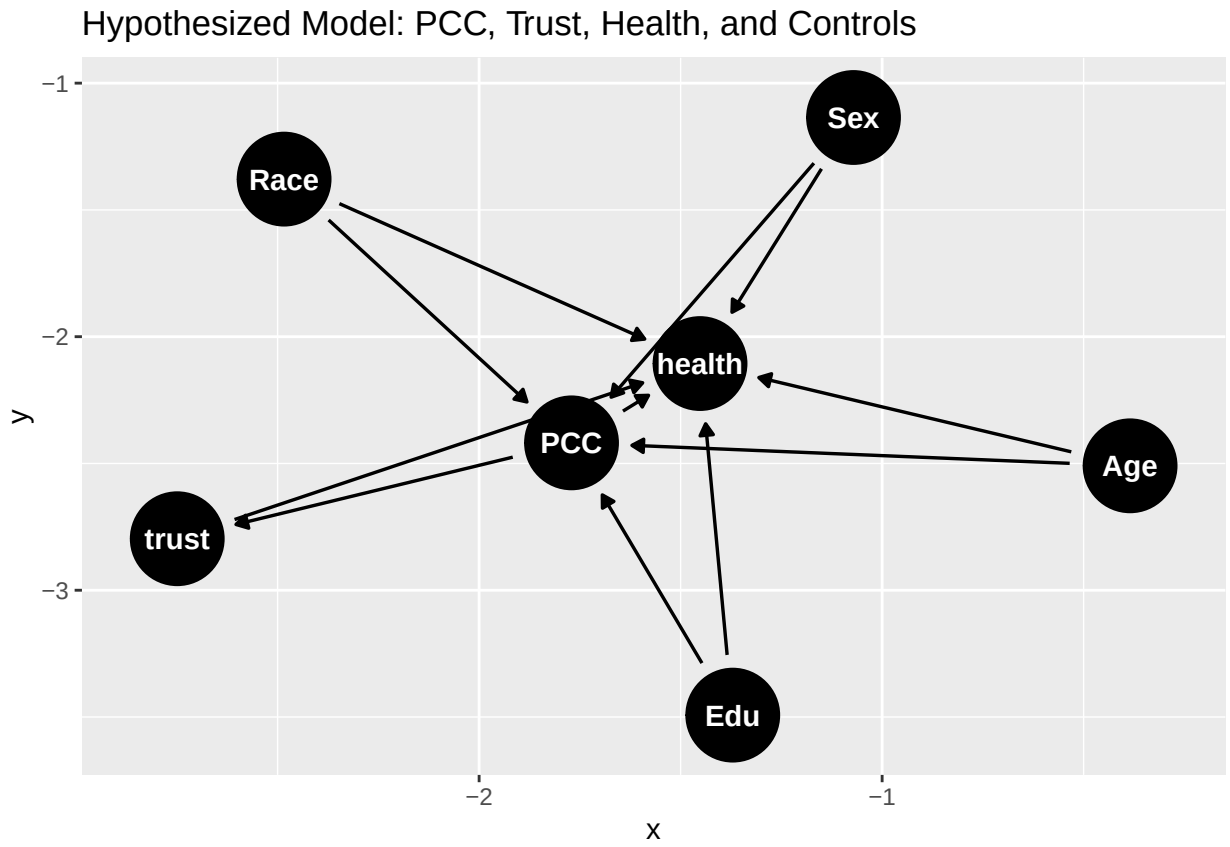



```
#DAG 2 with control variables
dag_model <- dagitty('dag {
  PCC -> trust
  trust -> health
  PCC -> health
  Age -> PCC
  Sex -> PCC
  Edu -> PCC
  Race -> PCC
}')
```

```
ggdag(dag_model, text = TRUE) +
  ggtitle("Hypothesized Model: PCC, Trust, Health, and Controls")
```




```
#DAG 3 with control variables
dag_model3 <- dagitty('dag {
  PCC -> trust
  trust -> health
  PCC -> health
  Age -> PCC
  Sex -> PCC
  Edu -> PCC
  Race -> PCC
  Age -> health
  Sex -> health
  Edu -> health
  Race -> health
}')

ggdag(dag_model3, text = TRUE) +
  ggtitle("Hypothesized Model: PCC, Trust, Health, and Controls")
```

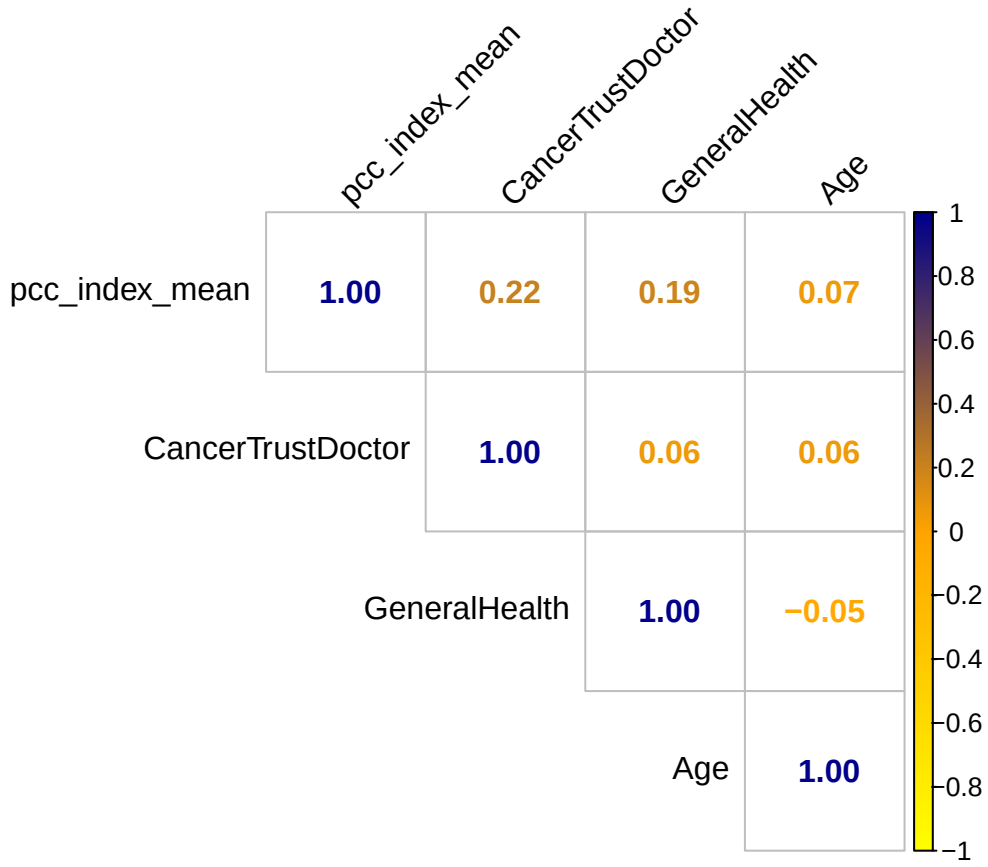


Correlation with PCC index

```
#NEW correlation matrix with the main variables after PCC becomes an index
vars2 <- cancer_d_reverse_index %>%
  dplyr::select(pcc_index_mean, CancerTrustDoctor, GeneralHealth, Age)
cor_matrix2 <- cor(vars2, use = "pairwise.complete.obs")
round(cor_matrix2, 2)
```

```
##                   pcc_index_mean CancerTrustDoctor GeneralHealth   Age
## pcc_index_mean              1.00              0.22          0.19  0.07
## CancerTrustDoctor           0.22              1.00          0.06  0.06
## GeneralHealth               0.19              0.06          1.00 -0.05
## Age                         0.07              0.06         -0.05  1.00
```

```
corrplot(cor_matrix2,
         method = "number",
         type = "upper",
         tl.col = "black",
         tl.srt = 45,
         number.cex = 1,
         col = colorRampPalette(c("yellow", "orange", "darkblue"))(200),
         number.font = 2)
```



OLS

```
#PCC variable (IV) will be treated as a continuous variable. GeneralHealth (DV) is a continuous variabl

#H1: PCC -> TRUST / OLS1 (with no controls)
ols1 <- lm(CancerTrustDoctor ~ pcc_index_mean, data = cancer_d_reverse_index)
summary(ols1)
```

```
##
## Call:
## lm(formula = CancerTrustDoctor ~ pcc_index_mean, data = cancer_d_reverse_index)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8796  0.1204  0.1449  0.2674  0.5858
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.19371    0.09668   33.03  < 2e-16 ***
## pcc_index_mean  0.17147    0.02806    6.11 1.62e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4728 on 735 degrees of freedom
## Multiple R-squared:  0.04834,    Adjusted R-squared:  0.04704
## F-statistic: 37.33 on 1 and 735 DF,  p-value: 1.616e-09
```

```
modelsummary(
  lm(CancerTrustDoctor ~ pcc_index_mean, data = cancer_d_reverse_index),
  statistic = "std.error",
  stars = TRUE,
  fmt = 2,
  title = "PCC Predicting Trust",
  output = "huxtable"
)
```

Table 1: PCC Predicting Trust

	(1)
(Intercept)	3.19***
	(0.10)
pcc_index_mean	0.17***
	(0.03)
Num.Obs.	737
R2	0.048
R2 Adj.	0.047
AIC	991.2
BIC	1005.0
Log.Lik.	-492.605
F	37.332
RMSE	0.47
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

```
#OLS1_1 = PCC -> trust (controlling for demographics)
#Dummy-code my control variables
cancer_d_reverse_index <- cancer_d_reverse_index %>%
  mutate(
    race_white = ifelse(RaceEthn == "Non-Hispanic White", 1,
                        ifelse(is.na(RaceEthn), NA, 0)),
    education_college = ifelse(Education %in% c("College graduate", "Postgraduate"), 1,
                               ifelse(is.na(Education), NA, 0))
  )

#age, gender, education doesn't do much to explain the relationship between PCC -> trust. Being White k
```

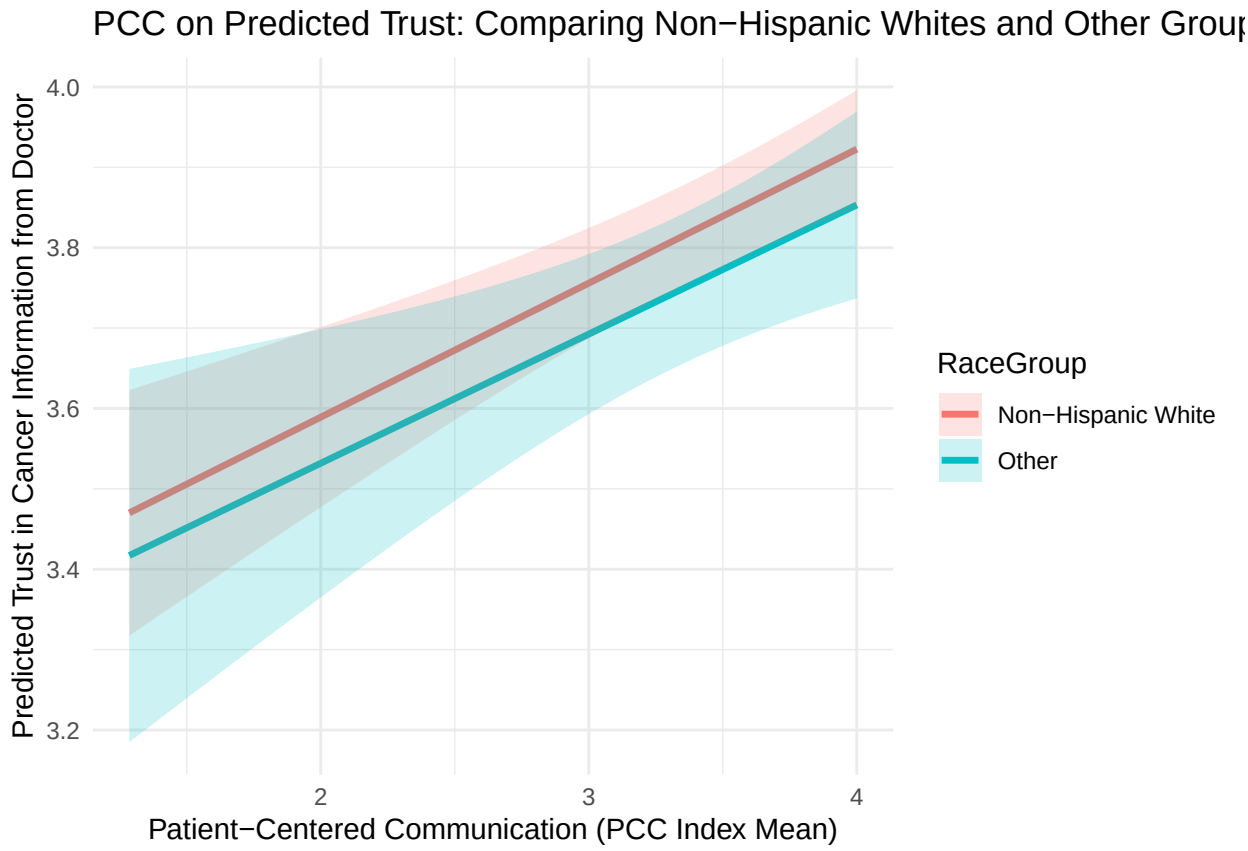
```r
ols1_1 <- lm(CancerTrustDoctor ~ pcc_index_mean + Age + race_white + BirthGender + education_college, da
summary(ols1_1)
```

```
##
## Call:
## lm(formula = CancerTrustDoctor ~ pcc_index_mean + Age + race_white +
##     BirthGender + education_college, data = cancer_d_reverse_index)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.93086  0.07469  0.15514  0.25385  0.56681
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.011819   0.134580  22.379  < 2e-16 ***
## pcc_index_mean    0.165003   0.028111   5.870 6.62e-09 ***
## Age               0.001850   0.001463   1.265    0.206
## race_white        0.065408   0.041938   1.560    0.119
## BirthGenderMale   0.001351   0.036118   0.037    0.970
## education_college 0.057208   0.035388   1.617    0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4716 on 731 degrees of freedom
## Multiple R-squared:  0.05828,    Adjusted R-squared:  0.05184
## F-statistic: 9.048 on 5 and 731 DF,  p-value: 2.332e-08
```

```r
modelsummary(
  lm(CancerTrustDoctor ~ pcc_index_mean + Age + race_white + BirthGender + education_college, data = ca
  statistic = "std.error",
  stars = TRUE,
  fmt = 2,
  title = "PCC Predicting Trust",
  output = "huxtable"
)
```

```r
#given that race sort of helps to explain the relationship between PCC and trust, I am trying to compar
ols_race <- lm(CancerTrustDoctor ~ pcc_index_mean * race_white + Age + education_college + BirthGender,
               data = cancer_d_reverse_index)

pred_data <- expand.grid(
  pcc_index_mean = seq(min(cancer_d_reverse_index$pcc_index_mean, na.rm = TRUE),
                       max(cancer_d_reverse_index$pcc_index_mean, na.rm = TRUE),
                       length.out = 100),
  race_white = c(0, 1),
  Age = mean(cancer_d_reverse_index$Age, na.rm = TRUE),
  education_college = 1,
  BirthGender = "Male"
)

ci <- predict(ols_race, newdata = pred_data, interval = "confidence")
pred_data$predicted <- ci[, "fit"]
```

```r
pred_data$ci_lower <- ci[, "lwr"]
pred_data$ci_upper <- ci[, "upr"]
pred_data$RaceGroup <- ifelse(pred_data$race_white == 1, "Non-Hispanic White", "Other")

ggplot(pred_data, aes(x = pcc_index_mean, y = predicted, color = RaceGroup, fill = RaceGroup)) +
  geom_line(size = 1.2) +
  geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), alpha = 0.2, color = NA) +
  labs(
    x = "Patient-Centered Communication (PCC Index Mean)",
    y = "Predicted Trust in Cancer Information from Doctor",
    title = "PCC on Predicted Trust: Comparing Non-Hispanic Whites and Other Groups"
  ) +
  theme_minimal()
```




```r
#PCC variable (IV) will be treated as a continuous variable. GeneralHealth (DV) is a continuous variabl

#H2: PCC -> HEALTH / OLS2
#OLS2 = pcc -> general health (with controls)
ols2 <- lm(GeneralHealth ~ pcc_index_mean + Age + race_white + BirthGender + education_college, data = c
summary(ols2)
```

```
##
## Call:
## lm(formula = GeneralHealth ~ pcc_index_mean + Age + race_white +
##     BirthGender + education_college, data = cancer_d_reverse_index)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.57603 -0.63329  0.01497  0.67853  2.23078
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.118970   0.262619   8.069 2.92e-15 ***
## pcc_index_mean     0.288249   0.054855   5.255 1.95e-07 ***
## Age               -0.003931   0.002854  -1.377  0.16889
## race_white         0.227614   0.081838   2.781  0.00555 **
## BirthGenderMale   -0.045475   0.070481  -0.645  0.51900
## education_college  0.459956   0.069056   6.661 5.36e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9202 on 731 degrees of freedom
## Multiple R-squared:   0.11,  Adjusted R-squared:  0.1039
## F-statistic: 18.07 on 5 and 731 DF,  p-value: < 2.2e-16
```

```
modelsummary(
  lm(GeneralHealth ~ pcc_index_mean + Age + race_white + BirthGender + education_college, data = cancer
  statistic = "std.error",
  stars = TRUE,
  fmt = 2,
  title = "PCC Predicting General Health",
  output = "huxtable"
)
```

```
#OLS4 PCC + trust -> health
ols4 <- lm(GeneralHealth ~ pcc_index_mean + CancerTrustDoctor + Age + race_white + BirthGender + educati
summary(ols4)
```

```
##
## Call:
## lm(formula = GeneralHealth ~ pcc_index_mean + CancerTrustDoctor +
##     Age + race_white + BirthGender + education_college, data = cancer_d_reverse_index)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.57622 -0.63352  0.01921  0.67701  2.23394
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.106048   0.341146   6.173 1.11e-09 ***
## pcc_index_mean     0.287541   0.056171   5.119 3.93e-07 ***
## CancerTrustDoctor  0.004290   0.072224   0.059  0.95265
## Age               -0.003939   0.002859  -1.377  0.16879
## race_white         0.227334   0.082030   2.771  0.00572 **
## BirthGenderMale   -0.045481   0.070529  -0.645  0.51923
## education_college  0.459711   0.069226   6.641 6.10e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.9208 on 730 degrees of freedom
## Multiple R-squared:   0.11,  Adjusted R-squared:  0.1027
## F-statistic: 15.04 on 6 and 730 DF,  p-value: 2.875e-16
```

```
modelsummary(
  lm(GeneralHealth ~ pcc_index_mean + CancerTrustDoctor + Age + race_white + BirthGender + education_col
  statistic = "std.error",
  stars = TRUE,
  fmt = 2,
  title = "PCC and Trust Predicting General Health with Controls",
  output = "huxtable"
)
```

```
#OLS5 PCC * trust -> health
ols5 <- lm(GeneralHealth ~ pcc_index_mean + pcc_index_mean*CancerTrustDoctor + Age + race_white + BirthG
summary(ols5)
```

```
##
## Call:
## lm(formula = GeneralHealth ~ pcc_index_mean + pcc_index_mean *
##     CancerTrustDoctor + Age + race_white + BirthGender + education_college,
##     data = cancer_d_reverse_index)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.56459 -0.62625  0.01157  0.67978  2.21754
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1.435957   1.246602   1.152  0.24974
## pcc_index_mean                    0.503734   0.390888   1.289  0.19791
## CancerTrustDoctor                 0.186915   0.334658   0.559  0.57666
## Age                              -0.003985   0.002862  -1.393  0.16417
## race_white                        0.228449   0.082092   2.783  0.00553 **
## BirthGenderMale                  -0.045725   0.070564  -0.648  0.51720
## education_college                 0.457195   0.069405   6.587 8.58e-11 ***
## pcc_index_mean:CancerTrustDoctor -0.058197   0.104130  -0.559  0.57641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9213 on 729 degrees of freedom
## Multiple R-squared:  0.1104, Adjusted R-squared:  0.1018
## F-statistic: 12.92 on 7 and 729 DF,  p-value: 1.014e-15
```

```
modelsummary(
  lm(GeneralHealth ~ pcc_index_mean + pcc_index_mean*CancerTrustDoctor + Age + race_white + BirthGender
  statistic = "std.error",
  stars = TRUE,
  fmt = 2,
  title = "Interaction of PCC x trust on Health with Controls",
  output = "huxtable"
)
```

Table 2: PCC Predicting Trust

	(1)
(Intercept)	3.01***
	(0.13)
pcc_index_mean	0.17***
	(0.03)
Age	0.00
	(0.00)
race_white	0.07
	(0.04)
BirthGenderMale	0.00
	(0.04)
education_college	0.06
	(0.04)
Num.Obs.	737
R2	0.058
R2 Adj.	0.052
AIC	991.5
BIC	1023.7
Log.Lik.	-488.734
F	9.048
RMSE	0.47
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Table 3: PCC Predicting General Health

	(1)
(Intercept)	2.12***
	(0.26)
pcc_index_mean	0.29***
	(0.05)
Age	-0.00
	(0.00)
race_white	0.23**
	(0.08)
BirthGenderMale	-0.05
	(0.07)
education_college	0.46***
	(0.07)
Num.Obs.	737
R2	0.110
R2 Adj.	0.104
AIC	1976.9
BIC	2009.1
Log.Lik.	-981.450
F	18.066
RMSE	0.92
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Table 4: PCC and Trust Predicting General Health with Controls

	(1)
(Intercept)	2.11***
	(0.34)
pcc_index_mean	0.29***
	(0.06)
CancerTrustDoctor	0.00
	(0.07)
Age	-0.00
	(0.00)
race_white	0.23**
	(0.08)
BirthGenderMale	-0.05
	(0.07)
education_college	0.46***
	(0.07)
Num.Obs.	737
R2	0.110
R2 Adj.	0.103
AIC	1978.9
BIC	2015.7
Log.Lik.	-981.448
F	15.035
RMSE	0.92
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Table 5: Interaction of PCC x trust on Health with Controls

	(1)
(Intercept)	1.44
	(1.25)
pcc_index_mean	0.50
	(0.39)
CancerTrustDoctor	0.19
	(0.33)
Age	-0.00
	(0.00)
race_white	0.23**
	(0.08)
BirthGenderMale	-0.05
	(0.07)
education_college	0.46***
	(0.07)
pcc_index_mean × CancerTrustDoctor	-0.06
	(0.10)
Num.Obs.	737
R2	0.110
R2 Adj.	0.102
AIC	1980.6
BIC	2022.0
Log.Lik.	-981.290
F	12.920
RMSE	0.92
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	