HW5-Trinath Sai Subhash Reddy

Trinath Sai Subhash Reddy Pittala, Uma Maheswara R Meleti, Hemanth Vasireddy

2023-04-19

Abstract

The objective of this report is to analyze the given data set consisting of information about individuals personal and cognitve traits and wanted to verify that individuals having good cognitve and personal triats tend to have more interest in white collar jobs than others.

Introduction

The problem being studied in this report is to understand how people interest in white collar jobs varys with their skill set. This type of analysis can be help financial institutions such a banks that gives loans to individuals who wanted to pursue courses. There is a high chance that people who have ambition for high paying jobs will have better repaying capability. Based on the personal and cognitve triats bankers can see if they are capable of achieving those goals, and can assess the risk by modelling with the previous data with personal and cognitve traits and their ambitions interests. We are assuming that the interest rating given by individuals in the dataset_train is true and genuine.

We found following data for average salary at payscale.com.

Occupation	Average Salary
Doctor	330,365
Lawyer	91,653
Business Executive	81,708
Architect	66,281
Stock Broker	62,793
Actor	62,496
Carpentry	61,007
Artist	59,156
Truck Driver	57,290
Police	56,644
Clergyman	54,869
Mortician	52,513
Social Worker	51,934
Teacher	51,568
Fireman	51,296
Sales Representative	49,884
Landscaper	47,558
Forest Ranger	38,000

We have selected the threshold for high paying jobs to be $65000.

Methods

We use multiple linear regression models with polynomial and interaction terms to predict the interest of the individual in high-paying jobs. We will use the coefficients from these models to determine the strength and direction of the relationships between the variables. We will use R programming language for this analysis, using the "lm" function to build our models. We used ggplot2 for data visualization and ggpairs to find the correlation between the variables.

Data

We have used ggpairs plot to find the correlation between the variables and found the following observations

Impulsiveness has a negative correlation with education, age, vocabulary, reading, sentence comprehension, mathematics, and geometry.

Worry has a positive correlation with stress.

The thrill-seeking trait has a positive correlation with impulsiveness.

Social dominance has a positive correlation with sociability.

Analytical Skills have a positive correlation with vocabulary, reading, Sentence Completion, Mathematics, and Geometry.

Geometry has a positive correlation with vocabulary, reading comprehension, and mathematics.

Mathematics has a positive correlation with vocabulary, reading comprehension, and sentence completion.

Sentence Completion has a positive correlation with vocabulary and reading comprehension.

Reading comprehension has a positive correlation with vocabulary and education.

Vocabulary has a positive correlation with education.

Results

We found that individuals with higher levels of analytical skills are more likely to have an interest in high-paying jobs. This finding is consistent with the idea that analytical skills are important for success in many high-paying occupations such as finance, law, and consulting.

We also found that geometry and mathematics skills are positively related to interest in high-paying jobs. This may be because these skills are important in fields such as engineering, computer science, and data analysis which tend to offer high salaries.

Furthermore, our analysis suggests that impulsiveness, worry, and stress have a negative impact on the interest in high-paying jobs. This finding is consistent with previous research that has shown that individuals with higher levels of impulsiveness and anxiety may have difficulty with long-term planning and decision making, which can limit their opportunities for high-paying jobs.

Finally, our analysis found that age and education are significant predictors of interest in high-paying jobs. Specifically, we found that individuals who are younger and have higher levels of education are more likely to express an interest in high-paying jobs. This finding is consistent with the idea that younger individuals may have higher aspirations and greater willingness to take on risks associated with pursuing high-paying careers. Additionally, individuals with higher levels of education may have greater access to information and resources that enable them to pursue high-paying careers.

Overall, our analysis provides support for the idea that personal and cognitive traits are important factors that contribute to the interest in high-paying jobs. Our findings suggest that analytical skills, geometry,

mathematics, social dominance, and sociability are positively related to interest in high-paying jobs, while impulsiveness, worry, and stress are negatively related to interest in high-paying jobs. Additionally, age and education were found to be significant predictors of interest in high-paying jobs.

Discussion

Our findings can be useful for financial institutions such as banks in predicting the loan repayment capability of individuals based on their personal and cognitive traits. For example, individuals with high cognitive abilities such as analytical skills and mathematics are more likely to have higher-paying jobs, which can lead to better loan repayment capability. Moreover, our findings can be used by organizations in selecting candidates for high-paying jobs based on their personal and cognitive traits.

However, it is important to note that our analysis has limitations. The data set used in this report is fabricated and may not be representative of the entire population. Furthermore, our analysis is based on self-reported interest in high-paying jobs, which may not accurately reflect actual job choices. Additionally, our models are based on linear regression assumptions, which may not hold true in all cases and have terrible r squared values. Future research can explore the use of other models to further validate our findings.

In conclusion, our analysis suggests that personal and cognitive traits are important factors that contribute to the interest in high-paying jobs. Our findings have practical implications for financial institutions and organizations in selecting individuals for high-paying jobs.

```
path = file.path("./interest.sav")
dataset = read_sav(path)
dataset
colnames(dataset)
```

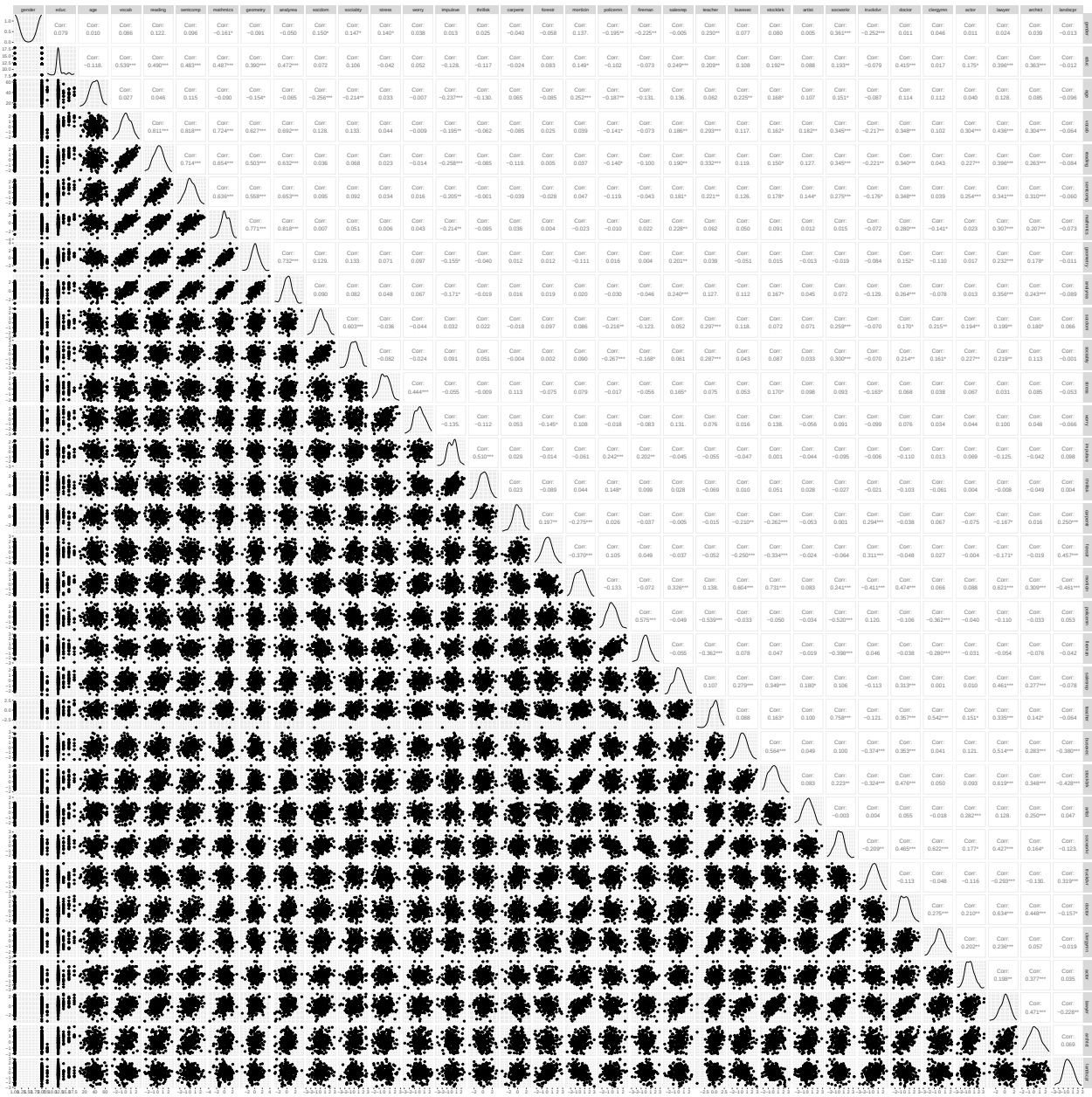
```
summary(dataset)
```

```
# Define number of bins
n_bins <- 5

# Loop over each column in the dataset_train and apply
# binning
for (col_name in names(dataset)) {
    # Skip non-numeric columns
    if (!is.numeric(dataset[[col_name]])) {
        next
    }

    # Create new column with binned values
    dataset <- dataset %>%
        mutate(!!paste0(col_name, "_bin") := ntile(dataset[[col_name]],
            n = n_bins))
}

dataset
```

```
set.seed(123456789)
dataset_train <- dataset[sample(nrow(dataset), 0.8 * nrow(dataset)),
    ]
dataset_test <- dataset[setdiff(1:nrow(dataset), rownames(dataset_train)),
    ]
```

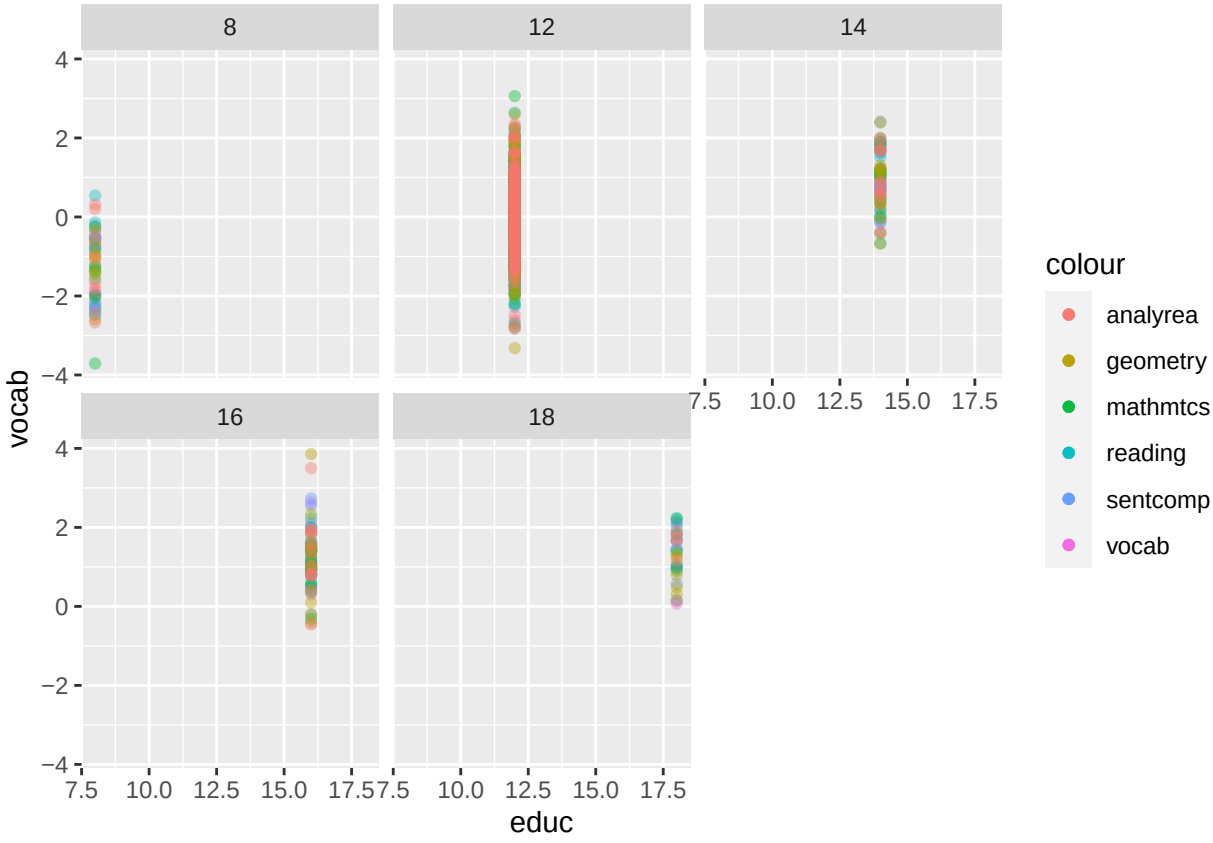
```
dim(dataset_train)
dim(dataset_test)
```

```
ggpairs(dataset_train[, c("gender", "educ", "age", "vocab", "reading",
    "sentcomp", "mathmtcs", "geometry", "analyrea", "socdom",
    "sociabty", "stress", "worry", "impulsve", "thrillsk", "carpentr",
    "forestr", "morticin", "policemn", "fireman", "salesrep",
    "teacher", "busexec", "stockbrk", "artist", "socworkr", "truckdvr",
    "doctor", "clergymn", "actor", "lawyer", "archtct", "landscpr")])
```



```
ggplot() + geom_point(data = dataset_train, aes(x = educ, y = vocab,
    color = "vocab"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = educ, y = reading, color = "reading"), alpha = 0.4) +
```

```
    geom_point(data = dataset_train, aes(x = educ, y = sentcomp,
        color = "sentcomp"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = educ, y = mathmtcs, color = "mathmtcs"), alpha = 0.4) +
    geom_point(data = dataset_train, aes(x = educ, y = geometry,
        color = "geometry"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = educ, y = analyrea, color = "analyrea"), alpha = 0.4) +
    facet_wrap(~educ)
```




```
M_busexec <- lm(busexec ~ as.factor(gender) + educ + age + vocab +
    reading + sentcomp + mathmtcs + geometry + analyrea + socdom +
    sociabty + stress + worry + impulsve + thrillsk, data = dataset_train)
M_busexec_step <- step(M_busexec, direction = "backward", scope = list(lower = ~1,
    upper = M_busexec))
summary(M_busexec_step)
```

```
M2_busexec <- lm(busexec ~ +poly(age, 3, raw = TRUE) * poly(socdom,
    3, raw = TRUE), data = dataset_train)
M2_busexec_step <- step(M2_busexec, direction = "backward", scope = list(lower = ~1,
    upper = M2_busexec))
summary(M2_busexec_step)
```

```
predictions <- predict(M2_busexec_step, newdata = dataset_test)  # make predictions on test data using
rmse <- sqrt(mean((dataset_test$busexec - predictions)^2))  # calculate root mean squared error
r_squared <- cor(dataset_test$busexec, predictions)^2  # calculate R-squared value
rmse
r_squared
```

```
M_doctor <- lm(doctor ~ as.factor(gender) + educ + age + vocab +
    reading + sentcomp + mathmtcs + geometry + analyrea + socdom +
    sociabty + stress + worry + impulsve + thrillsk, data = dataset_train)
M_doctor_step <- step(M_doctor, direction = "backward", scope = list(lower = ~1,
    upper = M_doctor))
summary(M_doctor_step)
```

```
M2_doctor <- lm(doctor ~ +poly(educ, 3, raw = TRUE) * poly(age,
    3, raw = TRUE), data = dataset_train)
M2_doctor_step <- step(M2_doctor, direction = "backward", scope = list(lower = ~1,
    upper = M2_doctor))
summary(M2_doctor_step)
```

```
predictions <- predict(M2_doctor_step, newdata = dataset_test)  # make predictions on test data using f
rmse <- sqrt(mean((dataset_test$doctor - predictions)^2))  # calculate root mean squared error
r_squared <- cor(dataset_test$doctor, predictions)^2  # calculate R-squared value
rmse
r_squared
```

```
M_lawyer <- lm(lawyer ~ as.factor(gender) + educ + age + vocab +
    reading + sentcomp + mathmtcs + geometry + analyrea + socdom +
    sociabty + stress + worry + impulsve + thrillsk, data = dataset_train)
M_lawyer_step <- step(M_lawyer, direction = "backward", scope = list(lower = ~1,
    upper = M_lawyer))
summary(M_lawyer_step)
```

```
M2_lawyer <- lm(lawyer ~ +poly(educ, 3, raw = TRUE) * poly(age,
    3, raw = TRUE) * poly(vocab, 3, raw = TRUE) * poly(worry,
    3, raw = TRUE), data = dataset_train)
M2_lawyer_step <- step(M2_lawyer, direction = "backward", scope = list(lower = ~1,
    upper = M2_lawyer))
summary(M2_lawyer_step)
```

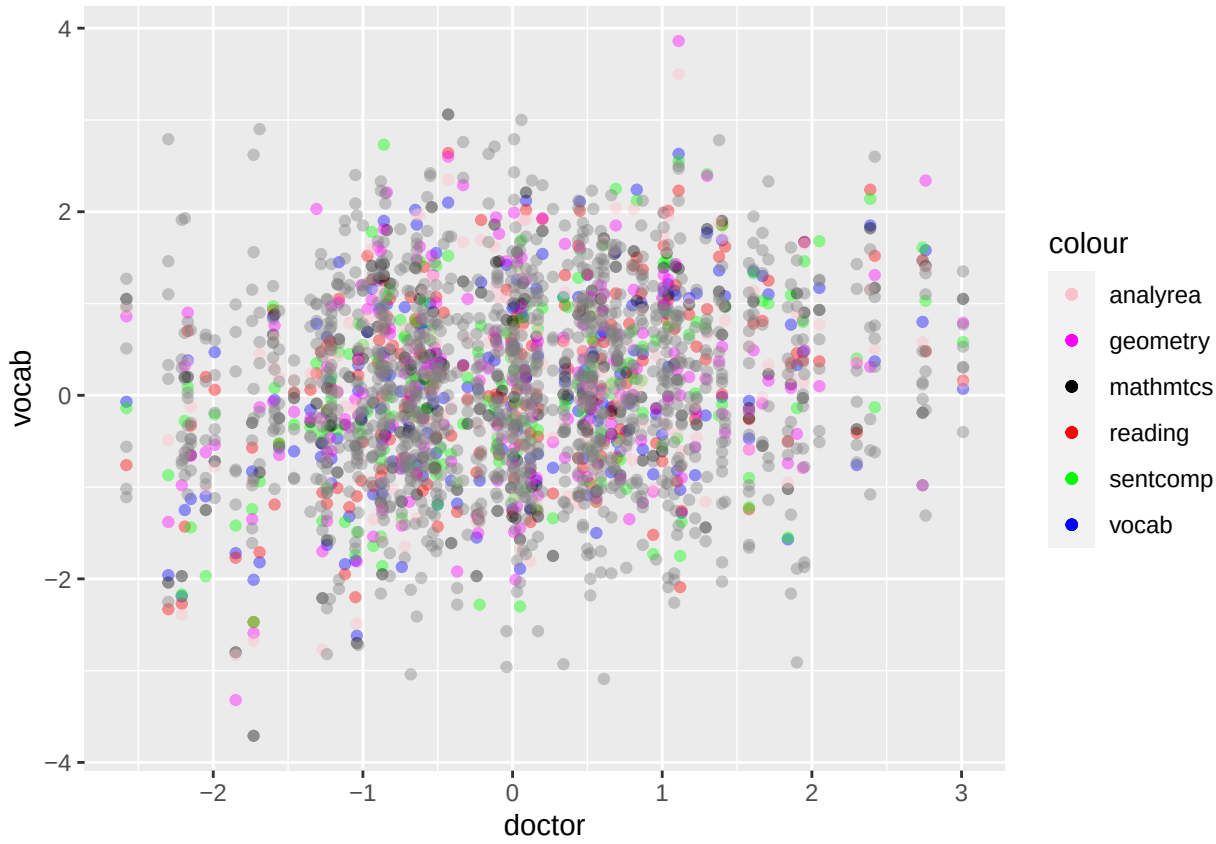
```
predictions <- predict(M2_lawyer_step, newdata = dataset_test)  # make predictions on test data using f
rmse <- sqrt(mean((dataset_test$lawyer - predictions)^2))  # calculate root mean squared error
r_squared <- cor(dataset_test$lawyer, predictions)^2  # calculate R-squared value
rmse
r_squared
```

```
M_archtct <- lm(archtct ~ as.factor(gender) + educ + age + vocab +
    reading + sentcomp + mathmtcs + geometry + analyrea + socdom +
    sociabty + stress + worry + impulsve + thrillsk, data = dataset_train)
M_archtct_step <- step(M_archtct, direction = "backward", scope = list(lower = ~1,
    upper = M_archtct))
summary(M_archtct_step)
```

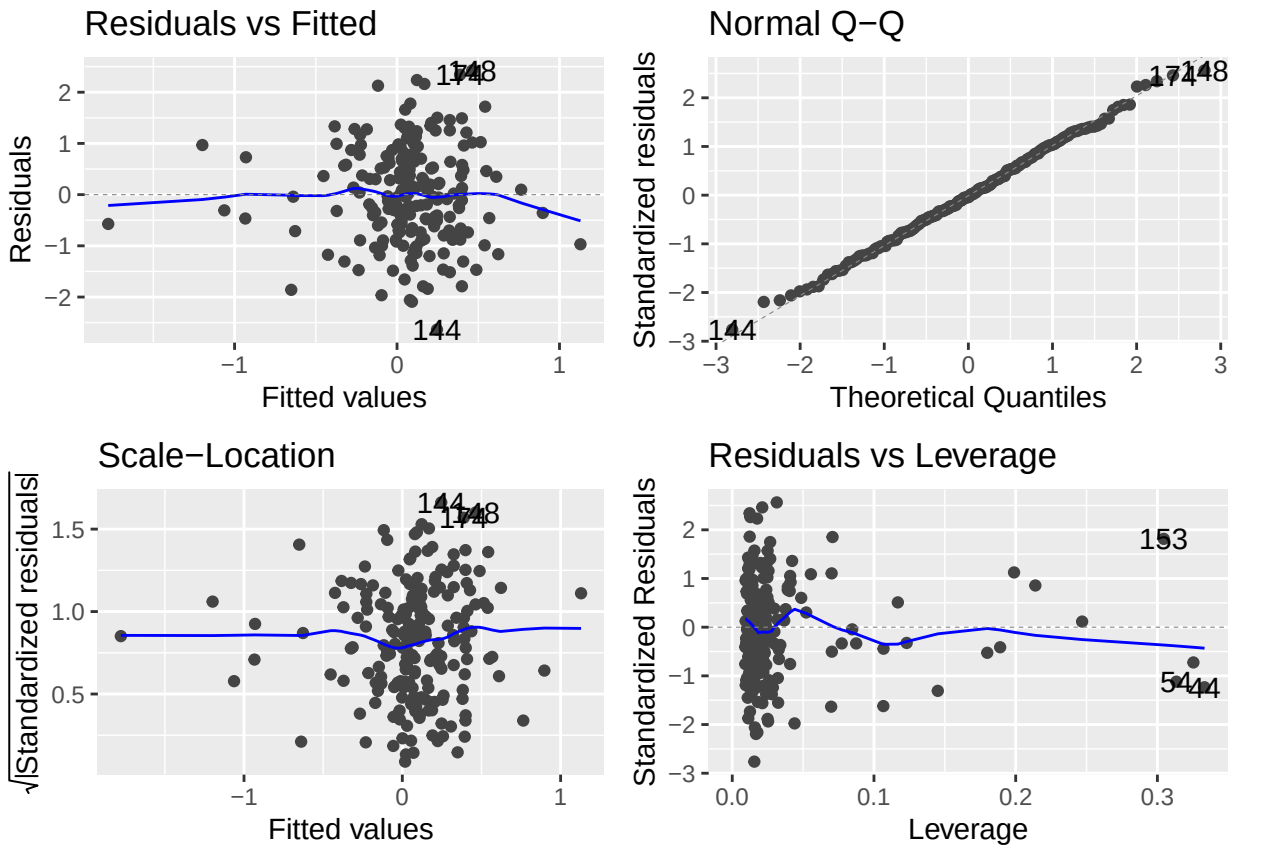
```
M2_archtct <- lm(archtct ~ as.factor(gender) * poly(educ, 3,
    raw = TRUE) * poly(age, 3, raw = TRUE) * poly(socdom, 3,
    raw = TRUE), data = dataset_train)
M2_archtct_step <- step(M2_archtct, direction = "backward", scope = list(lower = ~1,
    upper = M2_archtct))
summary(M2_archtct_step)
```

```
predictions <- predict(M2_archtct_step, newdata = dataset_test)  # make predictions on test data using
rmse <- sqrt(mean((dataset_test$archtct - predictions)^2))  # calculate root mean squared error
r_squared <- cor(dataset_test$archtct, predictions)^2  # calculate R-squared value
rmse
r_squared
```

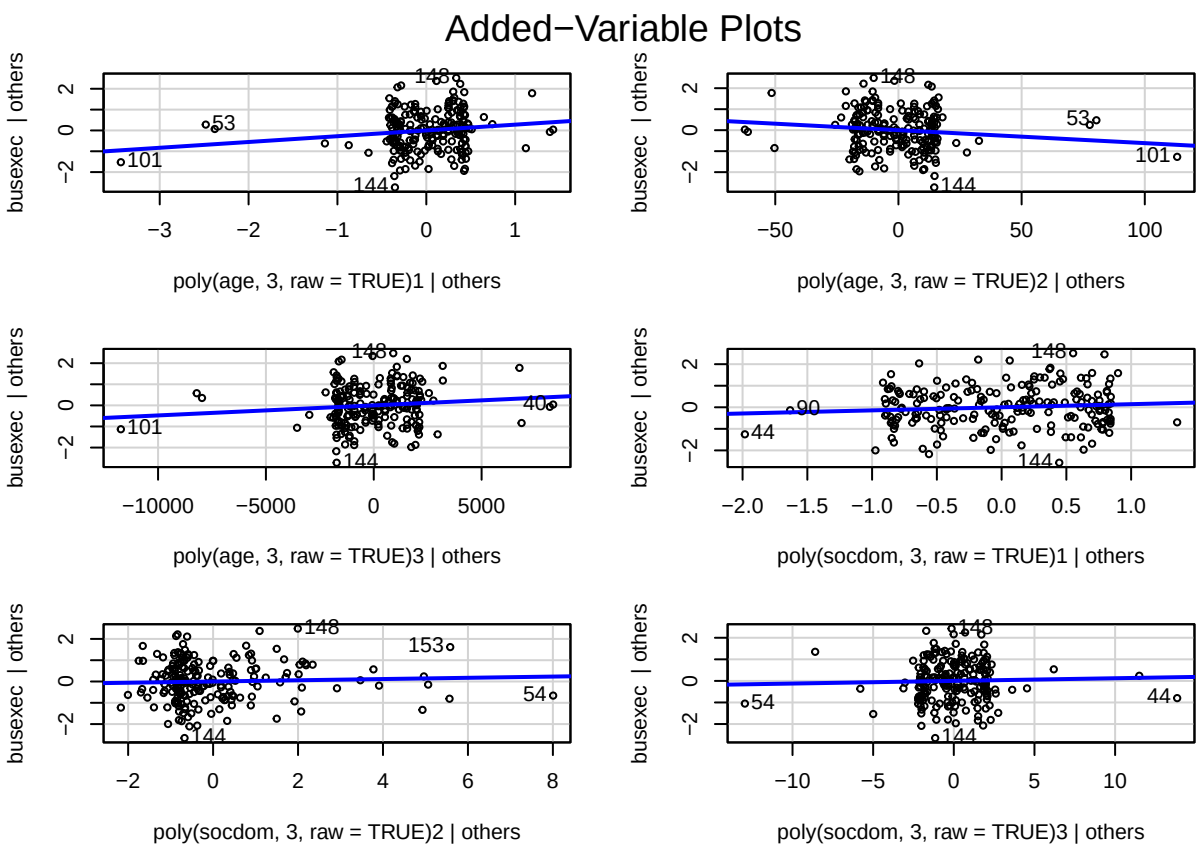
```
ggplot() + geom_point(data = dataset_train, aes(x = doctor, y = vocab,
    color = "vocab"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = doctor, y = reading, color = "reading"), alpha = 0.4) +
    geom_point(data = dataset_train, aes(x = doctor, y = sentcomp,
        color = "sentcomp"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = doctor, y = mathmtcs, color = "mathmtcs"), alpha = 0.4) +
    geom_point(data = dataset_train, aes(x = doctor, y = geometry,
        color = "geometry"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = doctor, y = analyrea, color = "analyrea"), alpha = 0.4) +
    geom_point(data = dataset_train, aes(x = doctor, y = socdom,
        color = "socdom"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = doctor, y = sociabty, color = "sociabty"), alpha = 0.4) +
    geom_point(data = dataset_train, aes(x = doctor, y = stress,
        color = "stress"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = doctor, y = worry, color = "worry"), alpha = 0.4) +
    geom_point(data = dataset_train, aes(x = doctor, y = impulsve,
        color = "impulsve"), alpha = 0.4) + geom_point(data = dataset_train,
    aes(x = doctor, y = thrillsk, color = "thrillsk"), alpha = 0.4) +
    scale_color_manual(values = c(vocab = "blue", reading = "red",
        sentcomp = "green", mathmtcs = "black", geometry = "magenta",
        analyrea = "pink"))
```



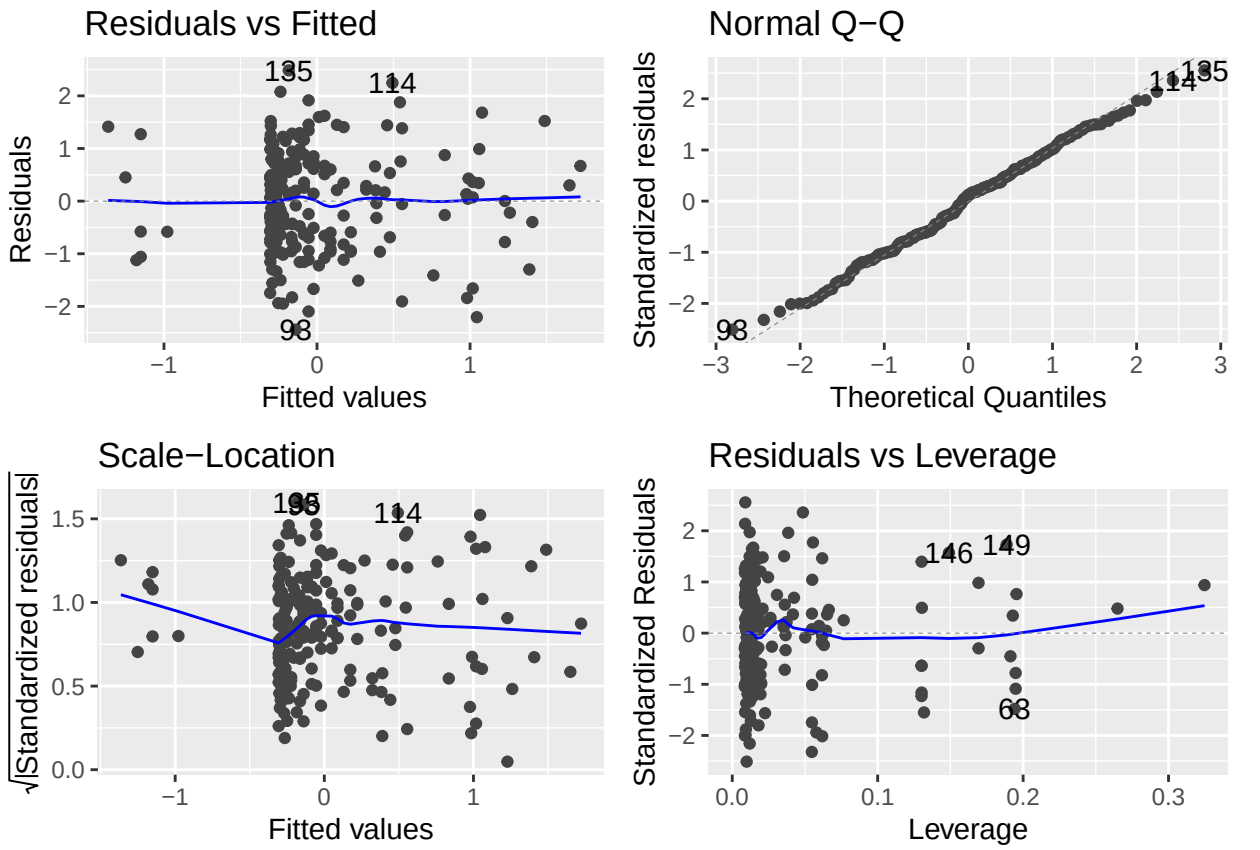

```
autoplot(M2_busexec_step)
```



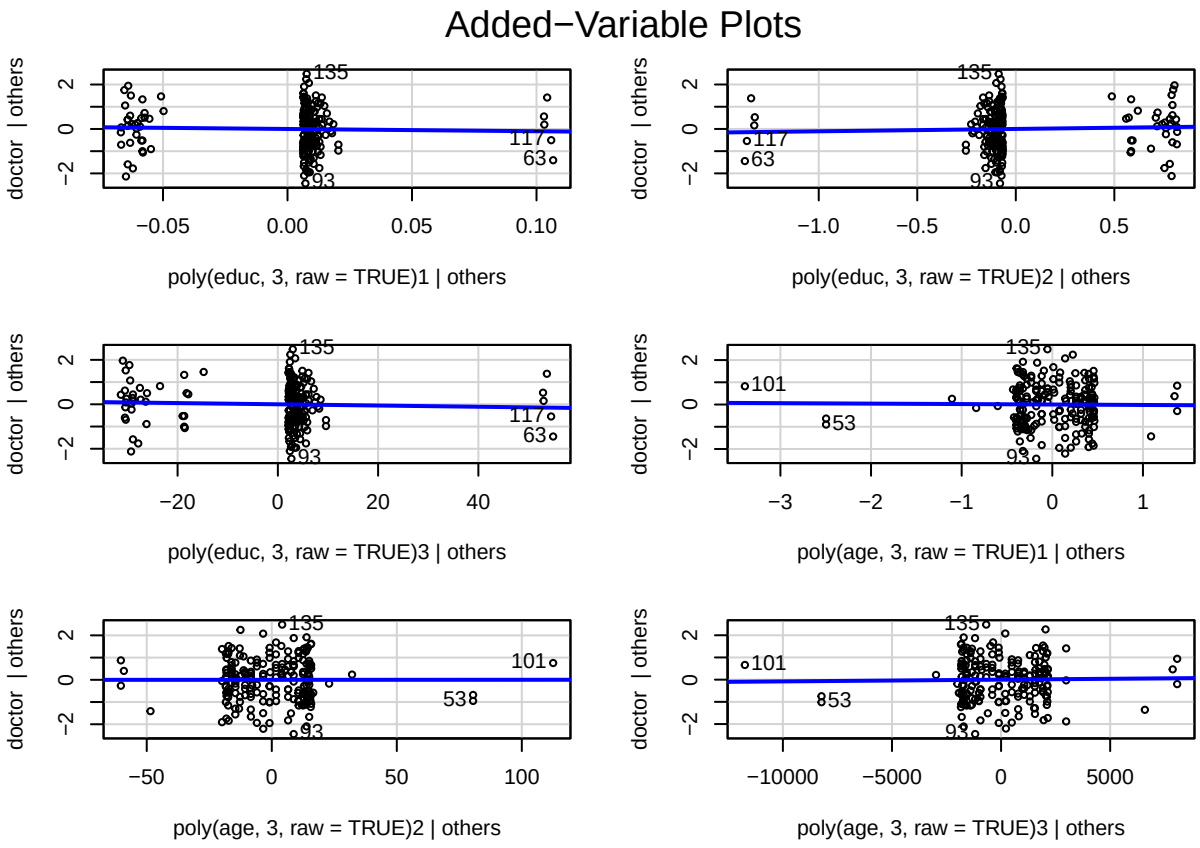

```
avPlots(M2_busexec_step)
```



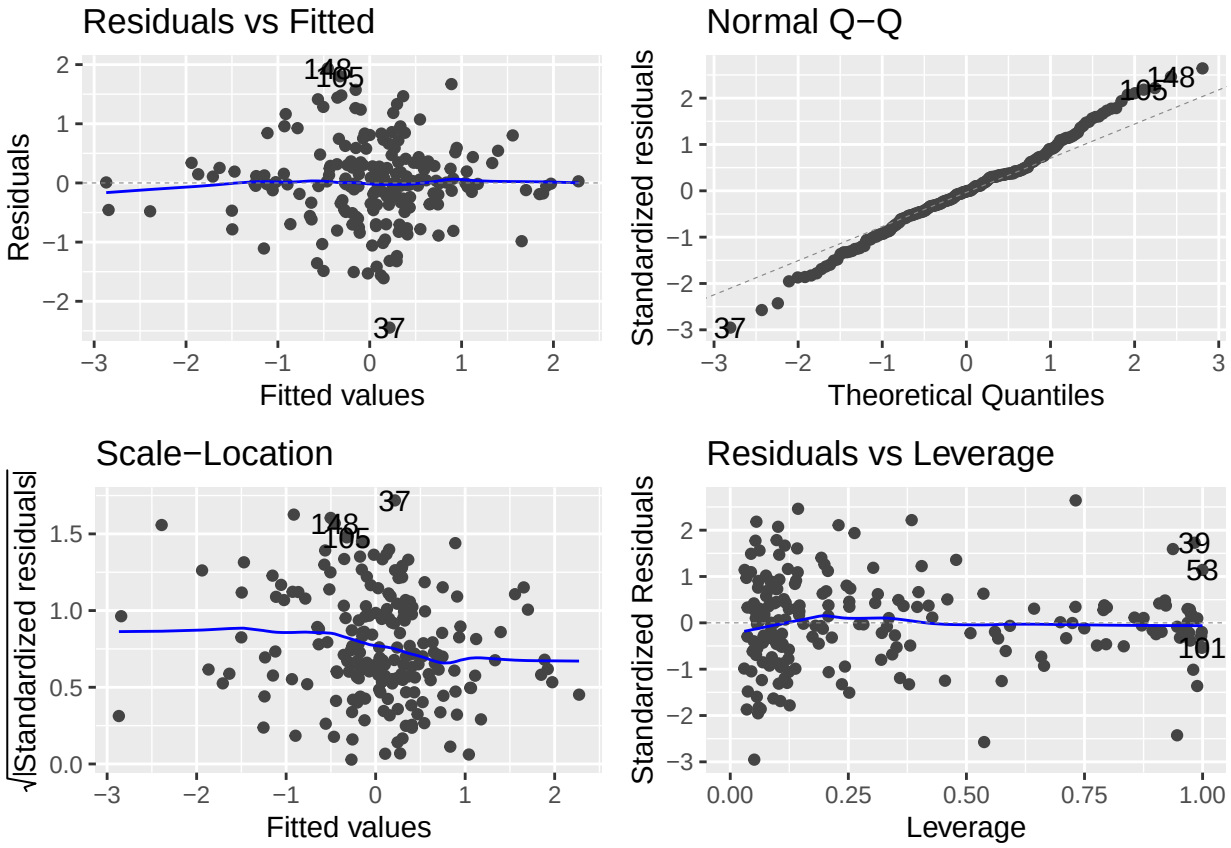

```
autoplot(M2_doctor_step)
```



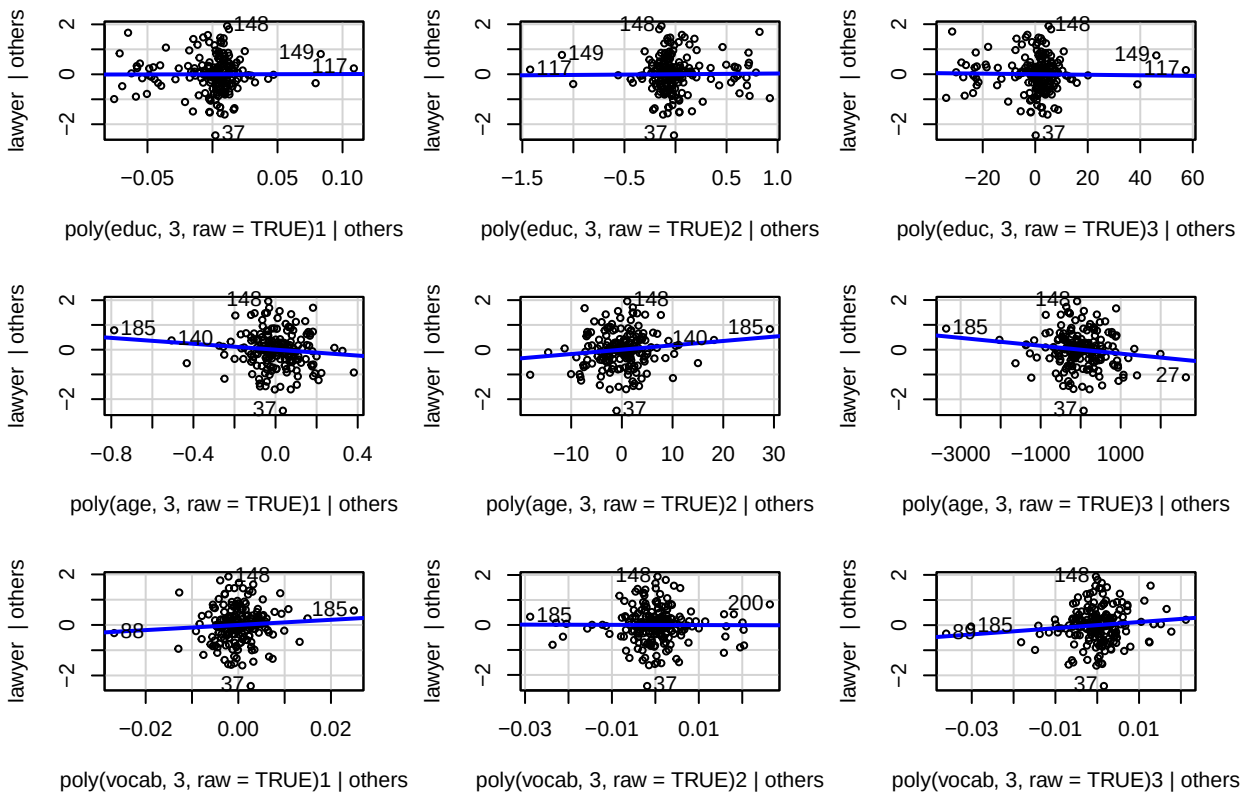

```
avPlots(M2_doctor_step)
```

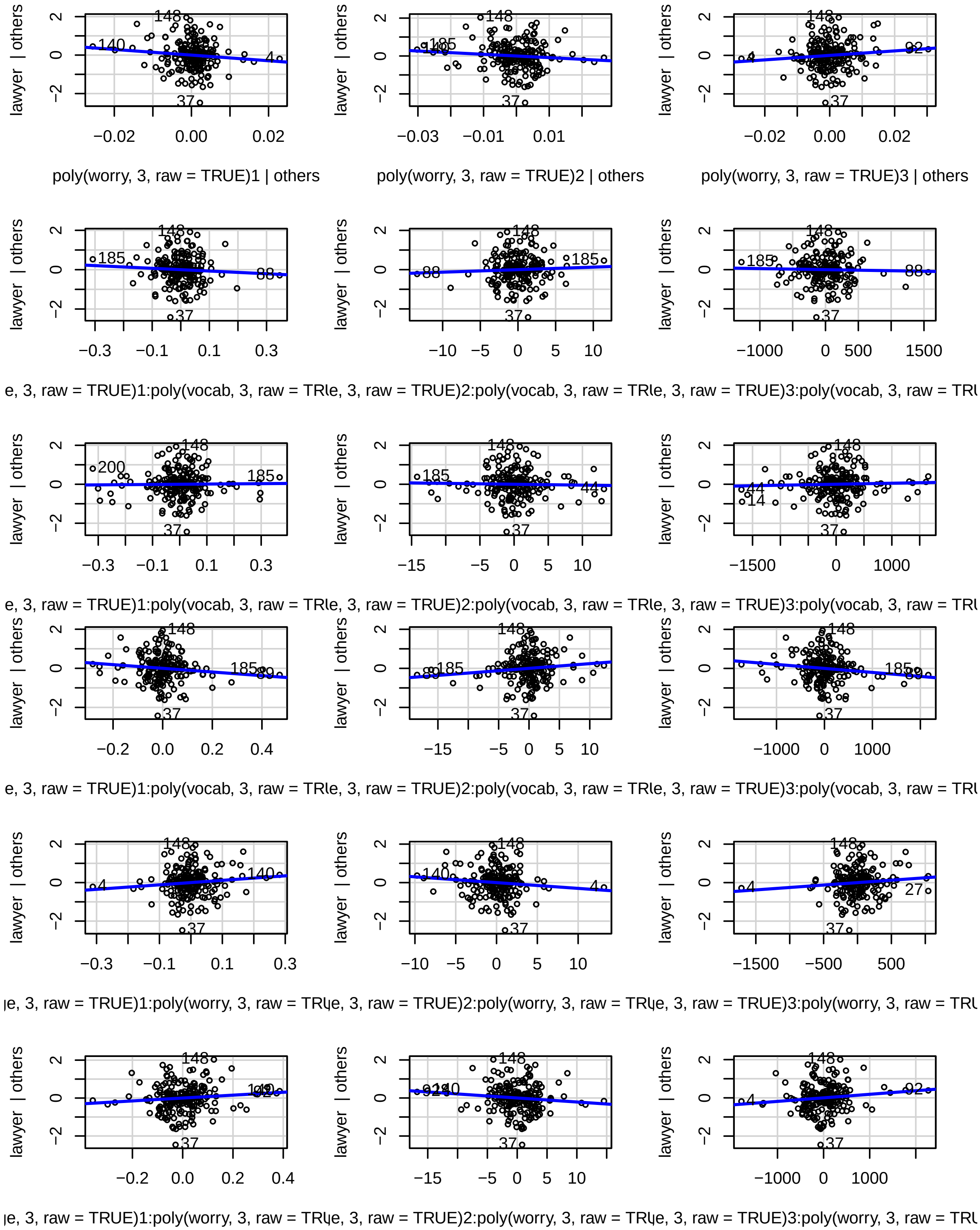


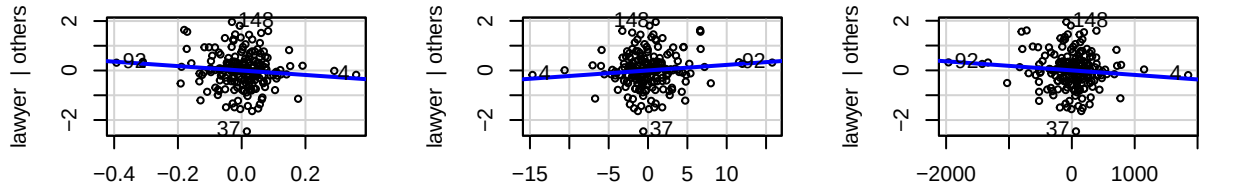

```
autoplot(M2_lawyer_step)
```



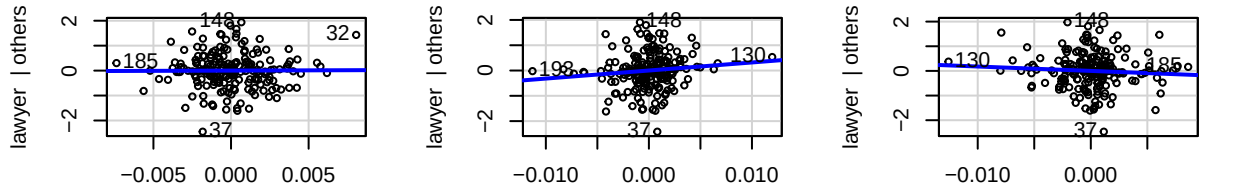

```
avPlots(M2_lawyer_step)
```



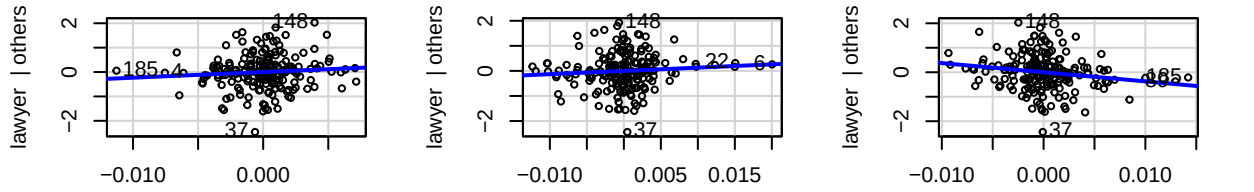

lawyer | others
poly(worry, 3, raw = TRUE)1 | others
poly(worry, 3, raw = TRUE)2 | others
poly(worry, 3, raw = TRUE)3 | others
e, 3, raw = TRUE)1:poly(vocab, 3, raw = TR
le, 3, raw = TRUE)2:poly(vocab, 3, raw = TR
le, 3, raw = TRUE)3:poly(vocab, 3, raw = TR
e, 3, raw = TRUE)1:poly(vocab, 3, raw = TR
le, 3, raw = TRUE)2:poly(vocab, 3, raw = TR
le, 3, raw = TRUE)3:poly(vocab, 3, raw = TR
e, 3, raw = TRUE)1:poly(vocab, 3, raw = TR
le, 3, raw = TRUE)2:poly(vocab, 3, raw = TR
le, 3, raw = TRUE)3:poly(vocab, 3, raw = TR
je, 3, raw = TRUE)1:poly(worry, 3, raw = TR
ye, 3, raw = TRUE)2:poly(worry, 3, raw = TR
ye, 3, raw = TRUE)3:poly(worry, 3, raw = TR
je, 3, raw = TRUE)1:poly(worry, 3, raw = TR
ye, 3, raw = TRUE)2:poly(worry, 3, raw = TR
ye, 3, raw = TRUE)3:poly(worry, 3, raw = TR



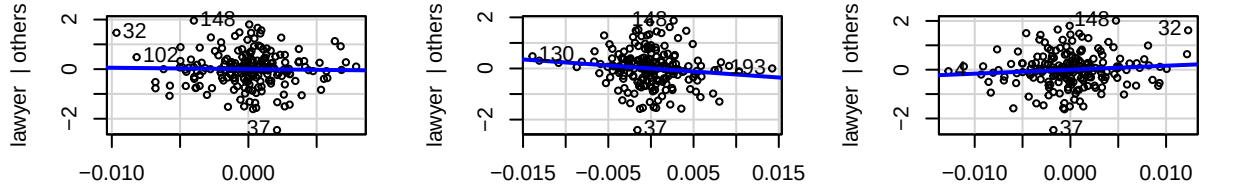
e, 3, raw = TRUE)1:poly(worry, 3, raw = TRUe, 3, raw = TRUE)2:poly(worry, 3, raw = TRUe, 3, raw = TRUE)3:poly(worry, 3, raw = TRU



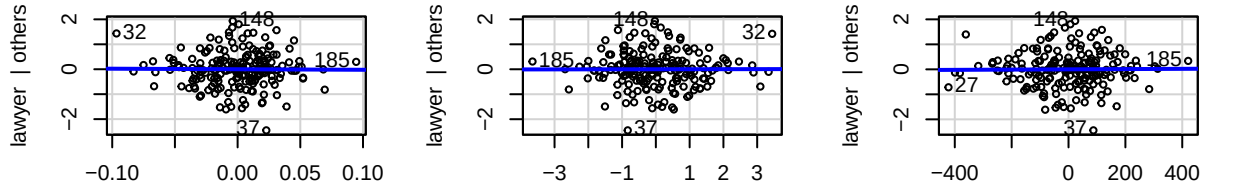
ab, 3, raw = TRUE)1:poly(worry, 3, raw = TRab, 3, raw = TRUE)2:poly(worry, 3, raw = TRab, 3, raw = TRUE)3:poly(worry, 3, raw = TR



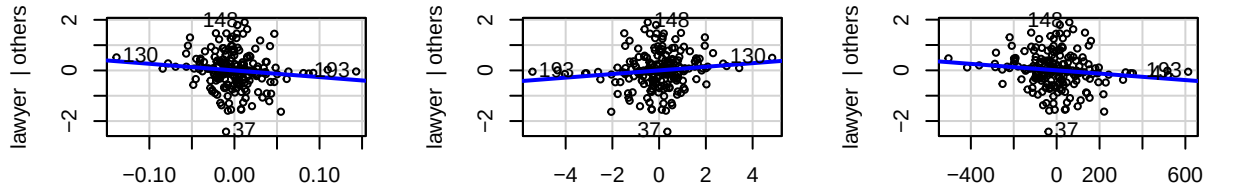
ab, 3, raw = TRUE)1:poly(worry, 3, raw = TRab, 3, raw = TRUE)2:poly(worry, 3, raw = TRab, 3, raw = TRUE)3:poly(worry, 3, raw = TR



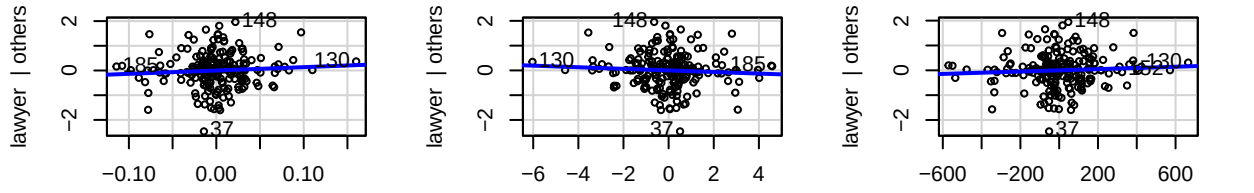
ab, 3, raw = TRUE)1:poly(worry, 3, raw = TRab, 3, raw = TRUE)2:poly(worry, 3, raw = TRab, 3, raw = TRUE)3:poly(worry, 3, raw = TR



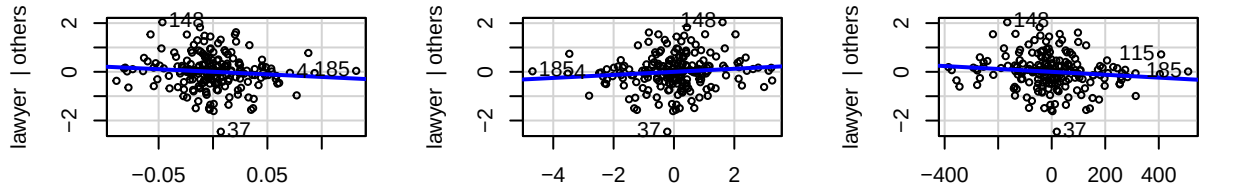
UE)1:poly(vocab, 3, raw = TRUE)1:poly(worrUE)2:poly(vocab, 3, raw = TRUE)1:poly(worrUE)3:poly(vocab, 3, raw = TRUE)1:poly(worr



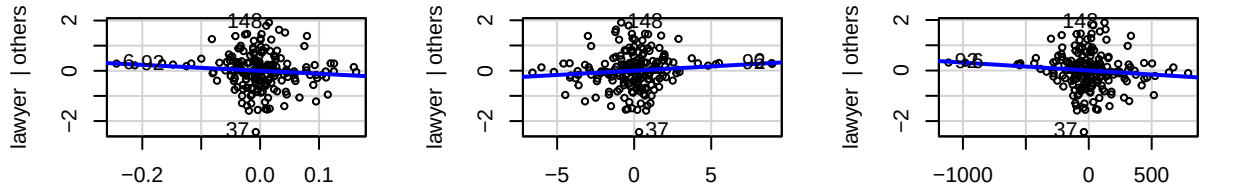
UE)1:poly(vocab, 3, raw = TRUE)2:poly(worrUE)2:poly(vocab, 3, raw = TRUE)2:poly(worrUE)3:poly(vocab, 3, raw = TRUE)2:poly(worr



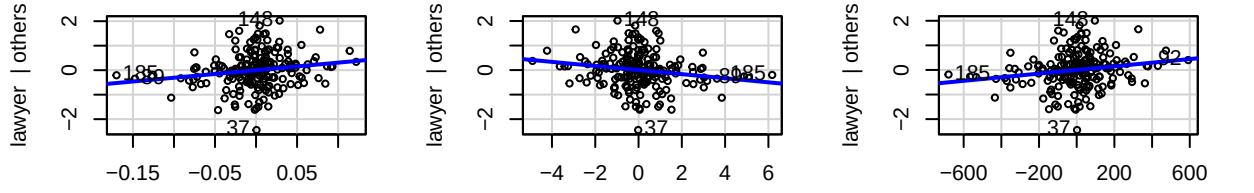
UE)1:poly(vocab, 3, raw = TRUE)3:poly(worr UE)2:poly(vocab, 3, raw = TRUE)3:poly(worr UE)3:poly(vocab, 3, raw = TRUE)3:poly(worr



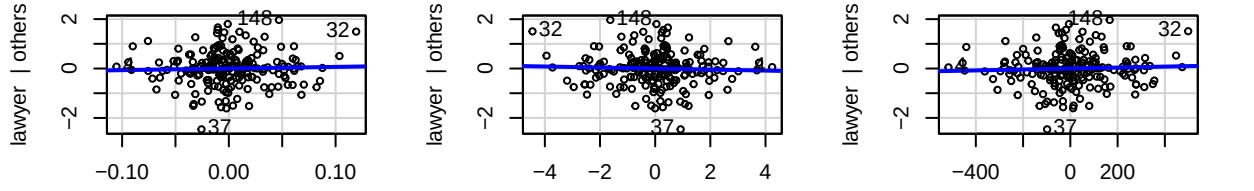
UE)1:poly(vocab, 3, raw = TRUE)1:poly(worr UE)2:poly(vocab, 3, raw = TRUE)1:poly(worr UE)3:poly(vocab, 3, raw = TRUE)1:poly(worr



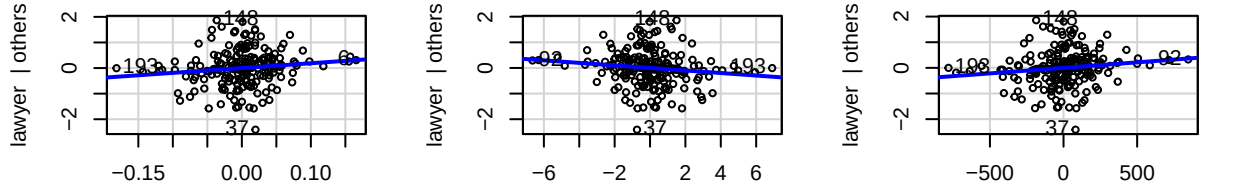
UE)1:poly(vocab, 3, raw = TRUE)2:poly(worr UE)2:poly(vocab, 3, raw = TRUE)2:poly(worr UE)3:poly(vocab, 3, raw = TRUE)2:poly(worr



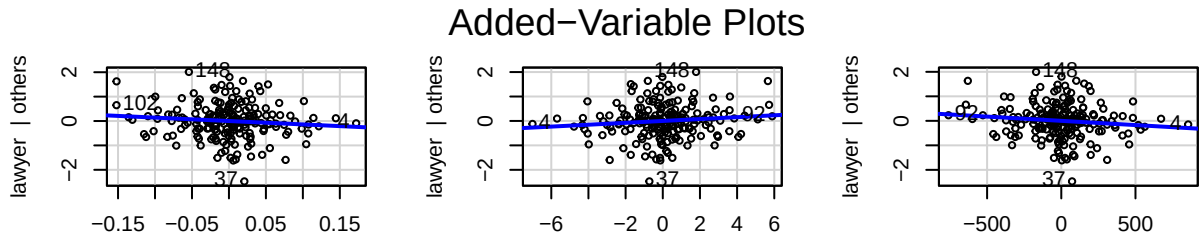
UE)1:poly(vocab, 3, raw = TRUE)3:poly(worr UE)2:poly(vocab, 3, raw = TRUE)3:poly(worr UE)3:poly(vocab, 3, raw = TRUE)3:poly(worr



UE)1:poly(vocab, 3, raw = TRUE)1:poly(worr UE)2:poly(vocab, 3, raw = TRUE)1:poly(worr UE)3:poly(vocab, 3, raw = TRUE)1:poly(worr

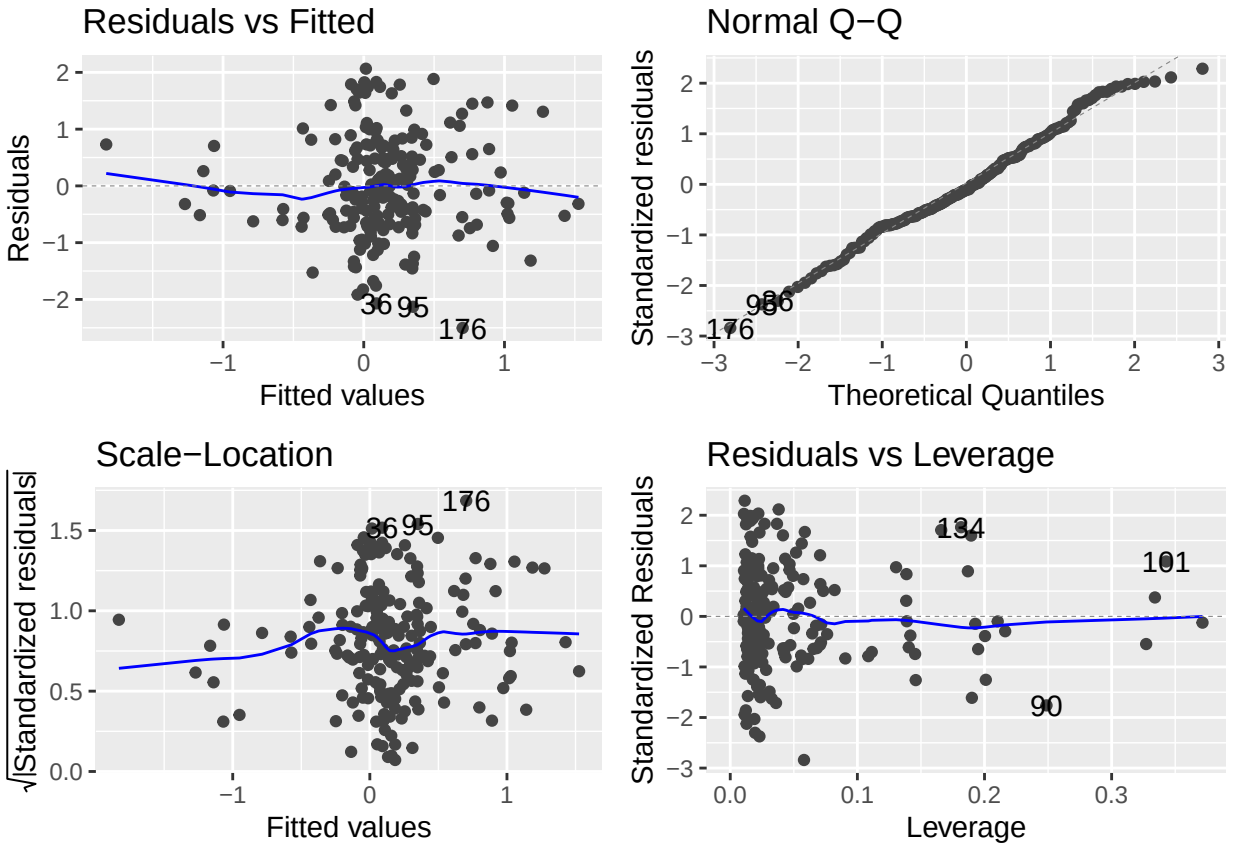


UE)1:poly(vocab, 3, raw = TRUE)2:poly(worr UE)2:poly(vocab, 3, raw = TRUE)2:poly(worr UE)3:poly(vocab, 3, raw = TRUE)2:poly(worr

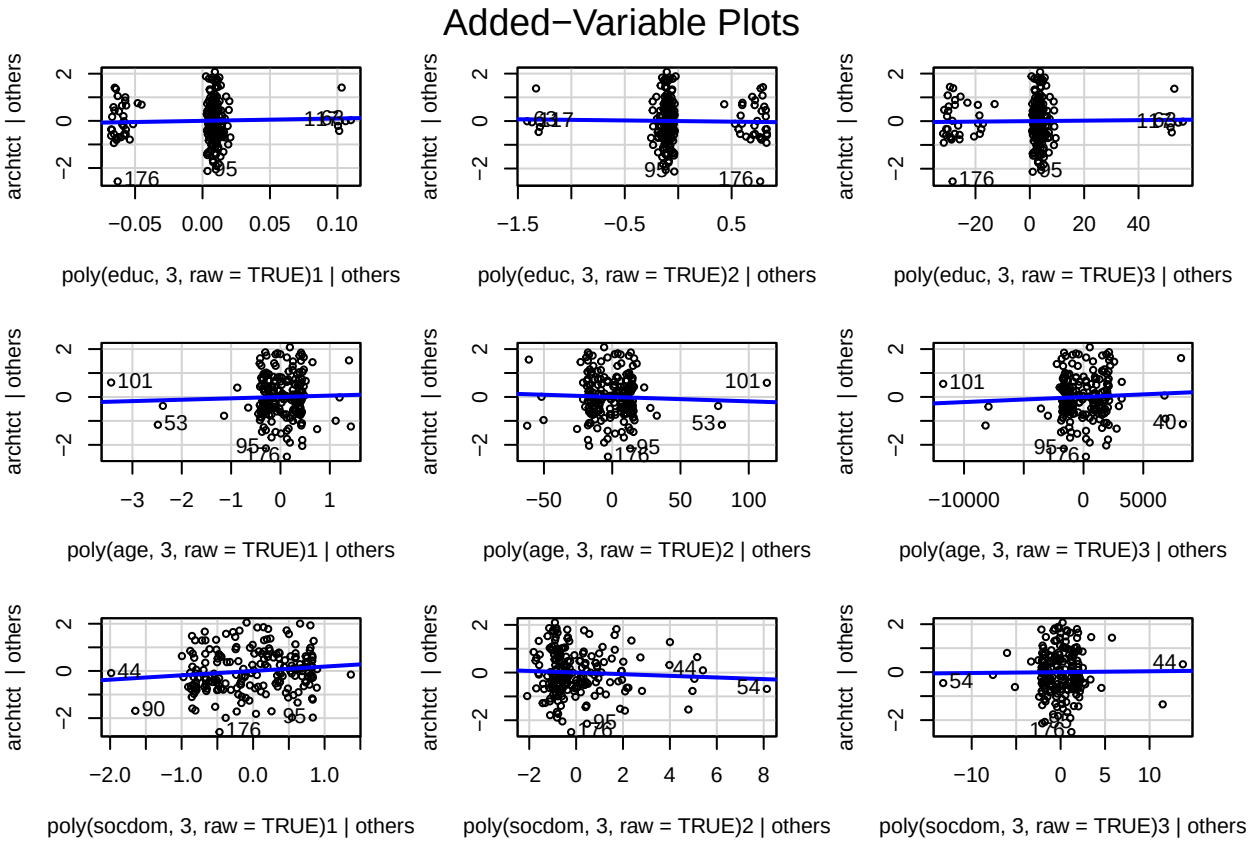


UE)1:poly(vocab, 3, raw = TRUE)3:poly(worrUE)2:poly(vocab, 3, raw = TRUE)3:poly(worrUE)3:poly(vocab, 3, raw = TRUE)3:poly(worr

```
autoplot(M2_archtct_step)
```




```
avPlots(M2_archtct_step)
```


Added−Variable Plots
archtct | others
poly(educ, 3, raw = TRUE)1 | others
poly(educ, 3, raw = TRUE)2 | others
poly(educ, 3, raw = TRUE)3 | others
poly(age, 3, raw = TRUE)1 | others
poly(age, 3, raw = TRUE)2 | others
poly(age, 3, raw = TRUE)3 | others
poly(socdom, 3, raw = TRUE)1 | others
poly(socdom, 3, raw = TRUE)2 | others
poly(socdom, 3, raw = TRUE)3 | others