Regresinės analizės projektinis darbas

Vainius Gataveckas, Matas Gaulia, Dovydas Martinkus

### Cleaning the data

library(tidyverse)  
library(janitor)  
library(countrycode)  
  
  
pop\_natural <- read\_csv("natural-population-growth.csv") %>%  
 filter(Year == 2017) %>%  
 dplyr::select(1, 4) %>%  
 set\_names(c("country", "natural\_growth")) %>%  
 mutate(country = countryname(country))  
  
pop\_total <- read\_csv("population-growth-rates.csv") %>%  
 filter(Year == 2017) %>%  
 dplyr::select(1, 4) %>%  
 set\_names(c("country", "total\_growth")) %>%  
mutate(country = countryname(country))  
  
  
country\_stats <- read\_csv("country\_profile\_variables.csv") %>%  
 clean\_names() %>%  
 dplyr::select(-c(2, 3, 4, 5, 6, 7)) %>%  
 mutate(country = countryname(country))  
  
happiness <- read\_csv("2017.csv") %>%  
 clean\_names() %>%  
 dplyr::select(-c(2), -starts\_with("whisker"), -c("dystopia\_residual", "happiness\_score", "family")) %>%  
 mutate(country = countryname(country))

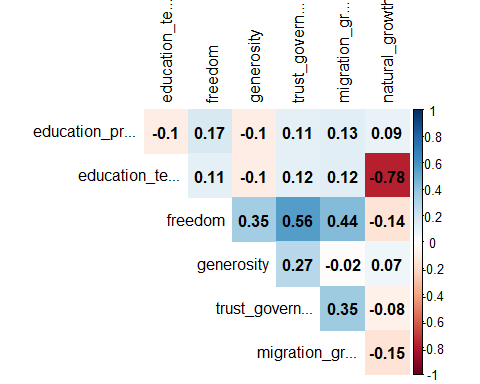
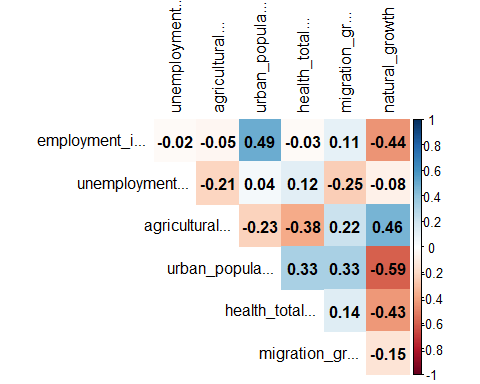
x <- reduce(list(pop\_natural, pop\_total, country\_stats, happiness), inner\_join, by = "country")  
  
  
x <- x %>%  
 dplyr::select(  
 -starts\_with("gdp"),  
 -starts\_with("labour"),  
 -starts\_with("international"),  
 -starts\_with("balance"),  
 -starts\_with("population"),  
 -starts\_with("fertility"),  
 -starts\_with("net"),  
 -starts\_with("energy\_prod"),  
 -starts\_with("forest"),  
 -starts\_with("threatened"),  
 -starts\_with("seats"),  
 -starts\_with("urban\_population\_growth"),  
 -starts\_with("refugees"),  
 -starts\_with("infant"),  
 -starts\_with("life\_expectancy"),  
 -starts\_with("co2"),  
 -starts\_with("economy"),  
 -starts\_with("education\_government"),  
 -starts\_with("energy"),  
 -health\_physicians\_per\_1000\_pop,  
 -individuals\_using\_the\_internet\_per\_100\_inhabitants,  
 -mobile\_cellular\_subscriptions\_per\_100\_inhabitants\_40,  
 -pop\_using\_improved\_sanitation\_facilities\_urban\_rural\_percent  
 ) %>%  
 mutate(across(everything(), ~ replace(., . %in% c("...", "-99", ".../..."), NA))) %>%  
 mutate(across(starts\_with("education"), ~ str\_split(., "/") %>% map(~ mean(as.numeric(.)))))  
  
  
pop <- x$pop\_using\_improved\_drinking\_water\_urban\_rural\_percent  
f1 <- possibly(~`[[`(.x,1),1)  
x$pop\_using\_improved\_drinking\_water\_urban <- pop %>% str\_split("/") %>% map(f1)  
f2 <- possibly(~`[[`(.x,2),1)  
x$pop\_using\_improved\_drinking\_water\_rural <- pop %>% str\_split("/") %>% map(f2)  
  
  
x <- x %>%  
 dplyr::select(-pop\_using\_improved\_drinking\_water\_urban\_rural\_percent) %>%  
 mutate(across(-country, as.numeric)) %>%  
 mutate(migration\_growth = total\_growth - natural\_growth) %>%  
 drop\_na() %>%  
 dplyr::select(-total\_growth)

library(rsample)  
  
set.seed(123)  
  
x <- x %>% mutate(  
 category = factor(case\_when(  
 migration\_growth >= 0 & natural\_growth >= 0 ~ 0, # "P migation, P natural",  
 migration\_growth >= 0 & natural\_growth < 0 ~ 1, # "P migation, N natural",  
 migration\_growth < 0 & natural\_growth >= 0 ~ 2, # "N migation, P natural",  
 TRUE ~ 3 # "N migration, N natural"  
 ))  
)   
  
  
train\_test\_split <- initial\_split(x,prop = 0.8)  
train <- training(train\_test\_split)  
test <- testing(train\_test\_split)  
  
country\_train <- train$country  
country\_test <- test$country  
  
train <- train %>% dplyr::select(-country)

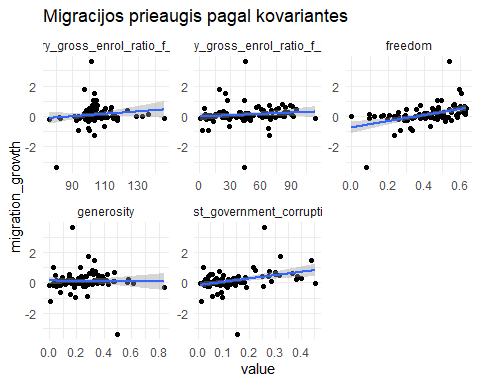
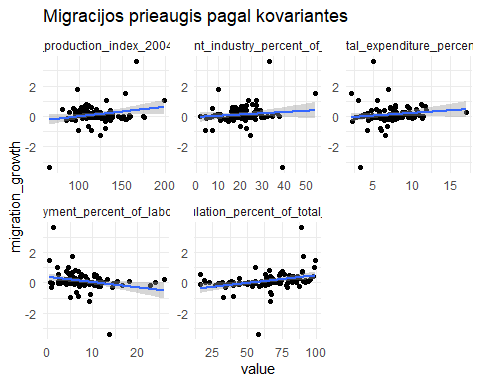
library(recipes)  
  
correlated\_recipe <- recipe(natural\_growth ~ ., data=train) %>%  
 add\_role(migration\_growth, new\_role = "outcome") %>%  
 add\_role(category, new\_role = "outcome") %>%   
 step\_corr(all\_numeric\_predictors(), threshold = 0.7) %>%  
 step\_nzv(all\_numeric\_predictors())  
  
  
correlated\_recipe <- prep(correlated\_recipe, training = train)  
  
train <- bake(correlated\_recipe, NULL)

### Regression models

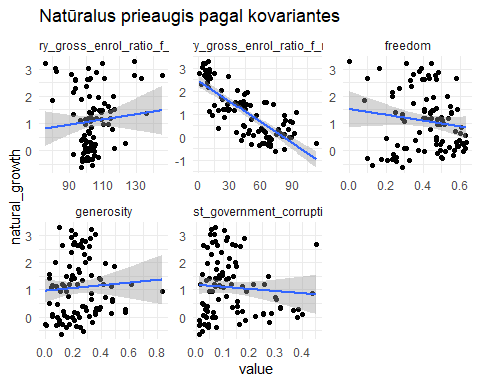
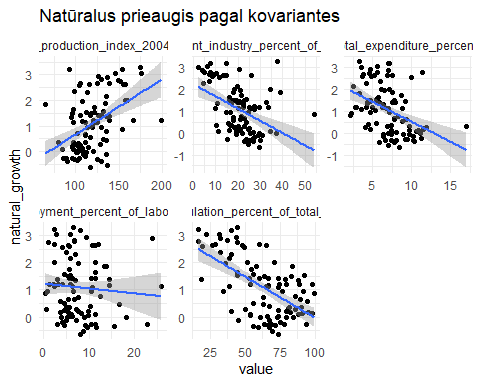
library(corrplot)  
  
  
regression\_train <- train %>% dplyr::select(-category)  
  
correlation <- function(name, name2) {  
 correlation\_matrix <- regression\_train %>%  
 dplyr::select(1:5, {{ name }}, {{ name2 }}) %>%  
 set\_names(., str\_trunc(names(.), 15)) %>%  
 cor()  
  
 corrplot(correlation\_matrix, order = "original", method = "color", type = "upper", diag = FALSE, tl.col = "black", addCoef.col = "black")  
  
  
  
 correlation\_matrix <- regression\_train %>%  
 dplyr::select(6:length(regression\_train), {{ name }}, {{ name2 }}) %>%  
 set\_names(., str\_trunc(names(.), 15)) %>%  
 cor()  
  
 corrplot(correlation\_matrix, order = "original", method = "color", type = "upper", diag = FALSE, tl.col = "black", addCoef.col = "black")  
}  
  
  
correlation(migration\_growth, natural\_growth)



scatterplot <- function(name, name2, main) {  
 a<- regression\_train %>%  
 dplyr::select(1:5, {{ name }}, -{{ name2 }}) %>%  
 pivot\_longer(-{{ name }}) %>%  
 ggplot(aes(x = value, y = {{ name }})) +  
 facet\_wrap(vars(name), scales = "free") +  
 geom\_point() +  
 geom\_smooth(method = "lm") +  
 theme\_minimal() + labs(title=main)  
  
  
 b <- regression\_train %>%  
 dplyr::select(6:length(regression\_train), {{ name }}, -{{ name2 }}) %>%  
 pivot\_longer(-{{ name }}) %>%  
 ggplot(aes(x = value, y = {{ name }})) +  
 facet\_wrap(vars(name), scales = "free") +  
 geom\_point() +  
 geom\_smooth(method = "lm") +  
 theme\_minimal() + labs(title=main)  
   
 plot(a)  
   
 plot(b)  
}  
  
  
scatterplot(migration\_growth, natural\_growth,"Migracijos prieaugis pagal kovariantes")



scatterplot(natural\_growth, migration\_growth,"Natūralus prieaugis pagal kovariantes")

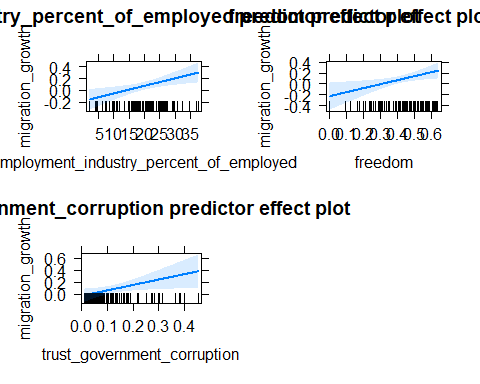
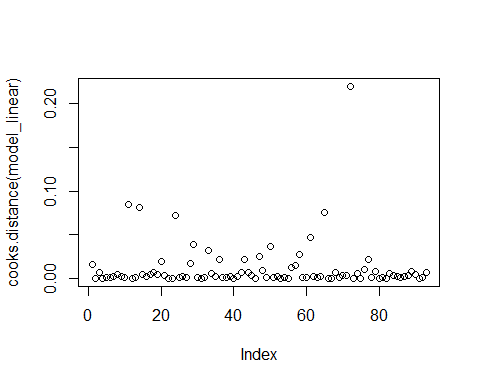
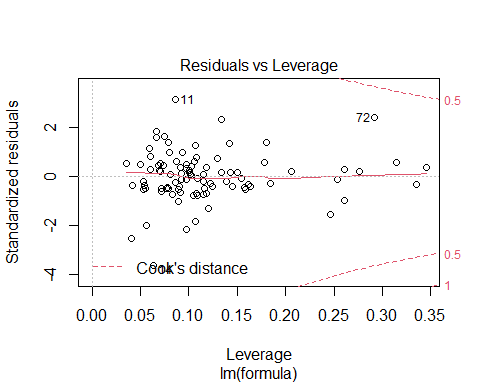
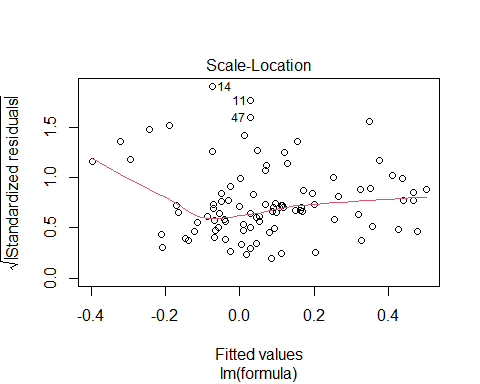
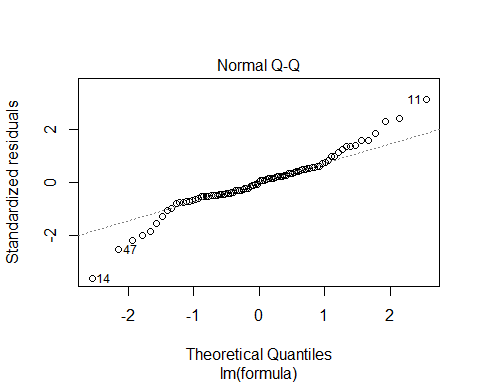
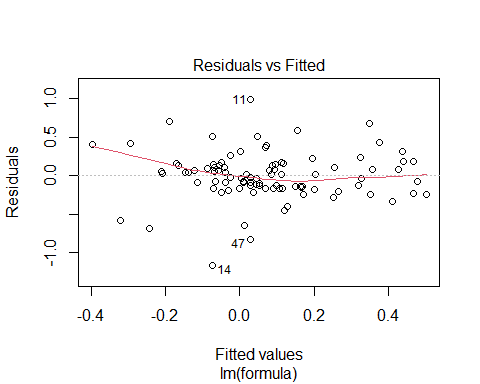
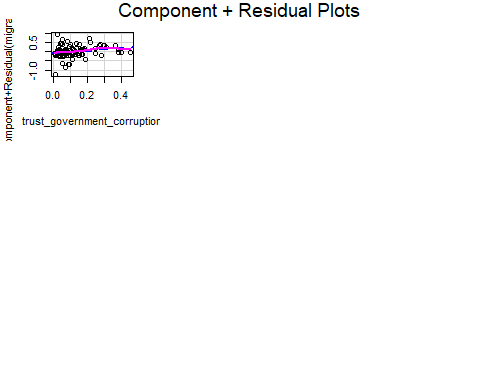
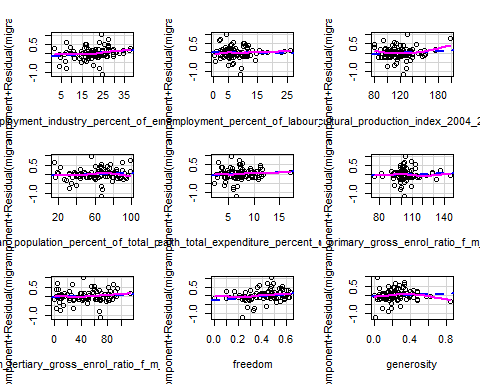


library(car)  
library(effects)  
library(lm.beta)  
  
  
linear\_fit <- function(formula) {  
 model\_linear <- lm(formula, data=data)  
  
 crPlots(model\_linear)  
 plot(model\_linear)  
 plot(cooks.distance(model\_linear))  
   
  
 model\_linear <- MASS::stepAIC(model\_linear, direction = "both",trace=0)  
  
 plot(predictorEffects(model\_linear))  
 print(summary(model\_linear))  
  
  
 stand\_coeffs <- lm.beta(model\_linear)  
  
 coeff\_plot <- tibble(x = names(stand\_coeffs$standardized.coefficients), y = stand\_coeffs$standardized.coefficients) %>%  
 cbind(confint(stand\_coeffs)) %>%  
 set\_names(c("variable", "coeff", "low", "high")) %>%  
 ggplot(aes(variable, coeff)) +  
 geom\_pointrange(aes(ymin = low, ymax = high), color = "blue") +  
 scale\_x\_discrete() +  
 coord\_flip() +  
 theme\_minimal() + labs(x="Kovariantė",y="Standartizuotos koeficientų reikšmės")  
   
 plot(coeff\_plot)  
  
 model\_linear  
}

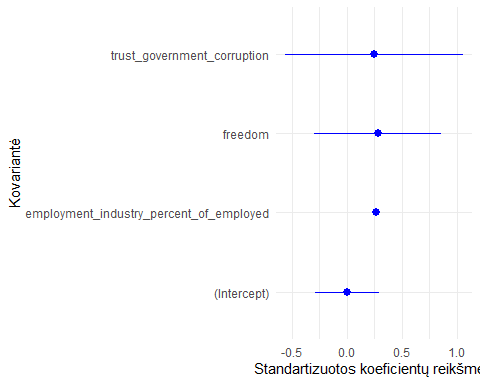
print("Tiesinės regresijos modelis migracijos prieaugiui")

## [1] "Tiesinės regresijos modelis migracijos prieaugiui"

outlier\_indices <- which(abs(regression\_train$migration\_growth) %in% (regression\_train$migration\_growth %>%  
 abs %>% sort(.,decreasing = TRUE) %>% `[`(1:4)))  
data <- regression\_train %>% dplyr::select(-natural\_growth) %>% slice(-outlier\_indices)  
model\_linear\_migration <- linear\_fit(migration\_growth ~ .)



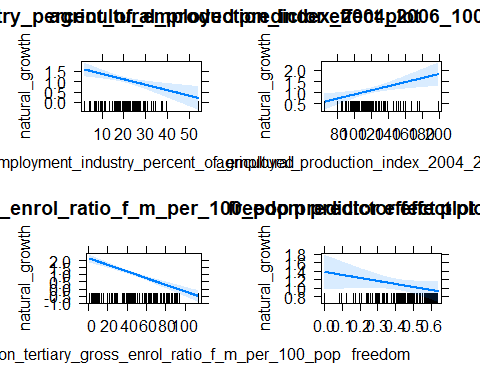
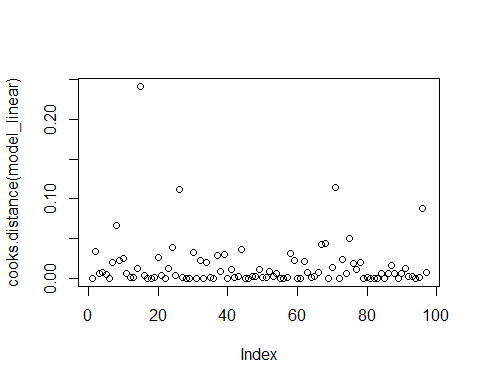
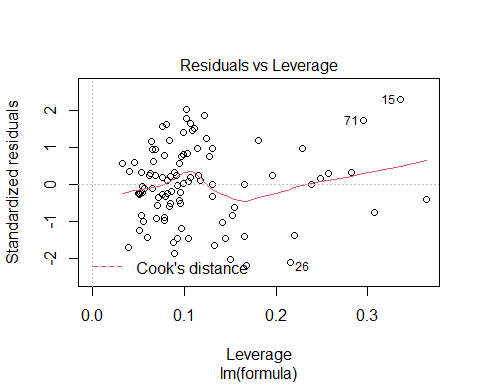
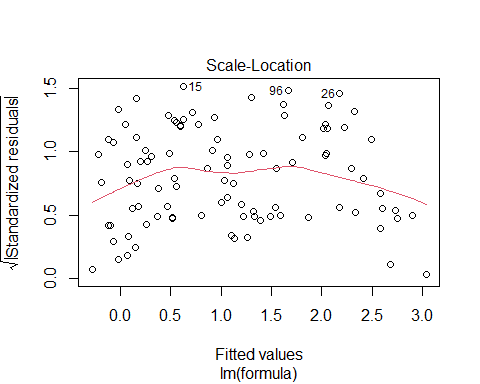
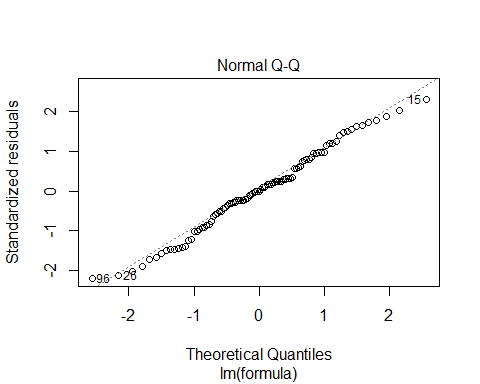
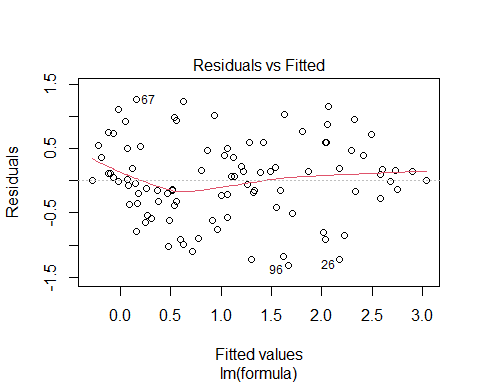
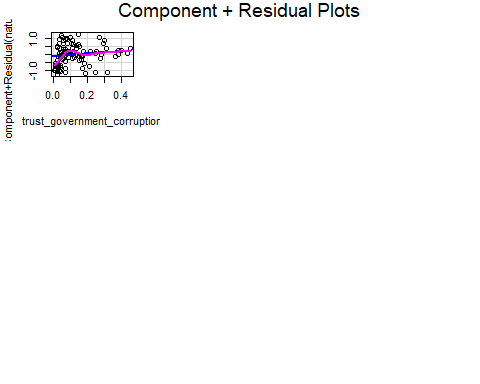
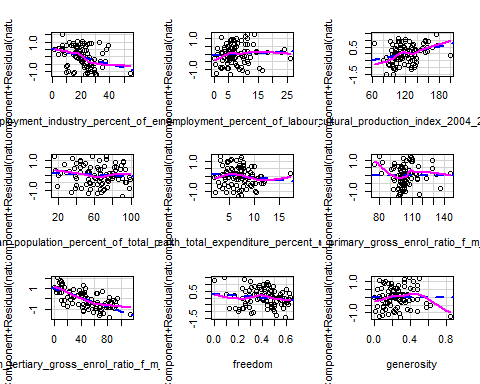
##   
## Call:  
## lm(formula = migration\_growth ~ employment\_industry\_percent\_of\_employed +   
## freedom + trust\_government\_corruption, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.13656 -0.15860 0.00083 0.15623 0.94305   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.59866 0.14753 -4.058 0.000106  
## employment\_industry\_percent\_of\_employed 0.01279 0.00449 2.848 0.005457  
## freedom 0.74118 0.29276 2.532 0.013105  
## trust\_government\_corruption 0.91015 0.40864 2.227 0.028451  
##   
## (Intercept) \*\*\*  
## employment\_industry\_percent\_of\_employed \*\*   
## freedom \*   
## trust\_government\_corruption \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.324 on 89 degrees of freedom  
## Multiple R-squared: 0.247, Adjusted R-squared: 0.2216   
## F-statistic: 9.733 on 3 and 89 DF, p-value: 1.28e-05



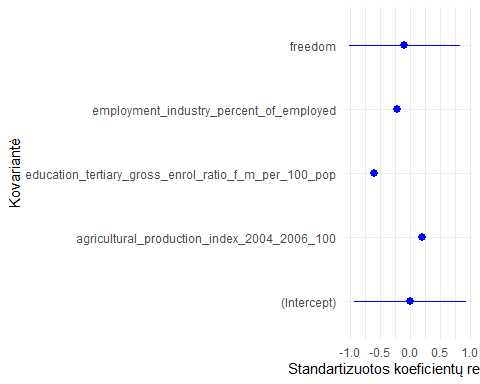
print("Tiesinės regresijos modelis natūraliam prieaugiui")

## [1] "Tiesinės regresijos modelis natūraliam prieaugiui"

data <- regression\_train %>% dplyr::select(-migration\_growth)  
model\_linear\_natural <- linear\_fit(natural\_growth ~ .)



##   
## Call:  
## lm(formula = natural\_growth ~ employment\_industry\_percent\_of\_employed +   
## agricultural\_production\_index\_2004\_2006\_100 + education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
## freedom, data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29210 -0.41415 0.04378 0.36362 1.48870   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 1.900127 0.475964  
## employment\_industry\_percent\_of\_employed -0.026640 0.008437  
## agricultural\_production\_index\_2004\_2006\_100 0.009170 0.003157  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.023327 0.002852  
## freedom -0.706028 0.468160  
## t value Pr(>|t|)   
## (Intercept) 3.992 0.000132 \*\*\*  
## employment\_industry\_percent\_of\_employed -3.158 0.002151 \*\*   
## agricultural\_production\_index\_2004\_2006\_100 2.904 0.004607 \*\*   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -8.180 1.52e-12 \*\*\*  
## freedom -1.508 0.134958   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6421 on 92 degrees of freedom  
## Multiple R-squared: 0.6663, Adjusted R-squared: 0.6518   
## F-statistic: 45.92 on 4 and 92 DF, p-value: < 2.2e-16



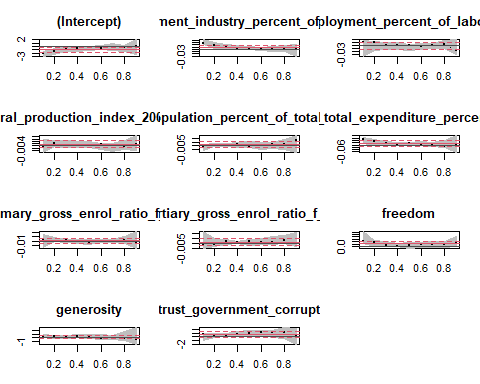
library(quantreg)  
  
quantile\_fit <- function() {  
 model\_quantile <- rq(formula, data = data, tau = tau)  
  
 print(summary(model\_quantile, se = "boot"))  
 plot(summary(model\_quantile))  
 print(anova(model\_quantile, test = "Wald", joint = TRUE))  
  
 model\_quantile  
}

print("Kvantilių regresija migracijos prieaugiui")

## [1] "Kvantilių regresija migracijos prieaugiui"

data <- regression\_train %>% dplyr::select(-natural\_growth) %>% slice(-outlier\_indices)  
tau <- seq(0.1,0.9,0.1)  
formula <- migration\_growth ~ .  
  
model\_quantile\_migration <- quantile\_fit()

##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.1  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -1.97977 0.90227   
## employment\_industry\_percent\_of\_employed 0.02605 0.01207   
## unemployment\_percent\_of\_labour\_force 0.01170 0.01252   
## agricultural\_production\_index\_2004\_2006\_100 -0.00040 0.00398   
## urban\_population\_percent\_of\_total\_population -0.00027 0.00457   
## health\_total\_expenditure\_percent\_of\_gdp 0.04745 0.03094   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00366 0.00698   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00200 0.00494   
## freedom 0.96911 0.75716   
## generosity 0.37700 0.50482   
## trust\_government\_corruption 0.68231 0.63346   
## t value Pr(>|t|)  
## (Intercept) -2.19422 0.03105  
## employment\_industry\_percent\_of\_employed 2.15928 0.03375  
## unemployment\_percent\_of\_labour\_force 0.93451 0.35278  
## agricultural\_production\_index\_2004\_2006\_100 -0.10131 0.91955  
## urban\_population\_percent\_of\_total\_population -0.06008 0.95224  
## health\_total\_expenditure\_percent\_of\_gdp 1.53338 0.12903  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.52469 0.60122  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.40471 0.68674  
## freedom 1.27991 0.20418  
## generosity 0.74680 0.45732  
## trust\_government\_corruption 1.07711 0.28459  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.2  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -1.44516 0.72374   
## employment\_industry\_percent\_of\_employed 0.01645 0.00964   
## unemployment\_percent\_of\_labour\_force 0.00953 0.00956   
## agricultural\_production\_index\_2004\_2006\_100 0.00175 0.00237   
## urban\_population\_percent\_of\_total\_population 0.00055 0.00256   
## health\_total\_expenditure\_percent\_of\_gdp 0.03359 0.01678   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00010 0.00438   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00095 0.00267   
## freedom 0.83431 0.49915   
## generosity 0.31753 0.33972   
## trust\_government\_corruption 0.41345 0.41754   
## t value Pr(>|t|)  
## (Intercept) -1.99678 0.04917  
## employment\_industry\_percent\_of\_employed 1.70615 0.09177  
## unemployment\_percent\_of\_labour\_force 0.99600 0.32218  
## agricultural\_production\_index\_2004\_2006\_100 0.74127 0.46065  
## urban\_population\_percent\_of\_total\_population 0.21494 0.83035  
## health\_total\_expenditure\_percent\_of\_gdp 2.00105 0.04869  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.02210 0.98242  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.35688 0.72210  
## freedom 1.67145 0.09844  
## generosity 0.93470 0.35269  
## trust\_government\_corruption 0.99020 0.32499  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.3  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.82193 0.53511   
## employment\_industry\_percent\_of\_employed 0.00713 0.00780   
## unemployment\_percent\_of\_labour\_force -0.00008 0.00869   
## agricultural\_production\_index\_2004\_2006\_100 0.00080 0.00200   
## urban\_population\_percent\_of\_total\_population 0.00093 0.00251   
## health\_total\_expenditure\_percent\_of\_gdp 0.01587 0.01265   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00081 0.00243   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00149 0.00177   
## freedom 0.27256 0.33727   
## generosity 0.27022 0.24381   
## trust\_government\_corruption 0.34574 0.45371   
## t value Pr(>|t|)  
## (Intercept) -1.53600 0.12839  
## employment\_industry\_percent\_of\_employed 0.91456 0.36310  
## unemployment\_percent\_of\_labour\_force -0.00971 0.99228  
## agricultural\_production\_index\_2004\_2006\_100 0.40039 0.68991  
## urban\_population\_percent\_of\_total\_population 0.37119 0.71146  
## health\_total\_expenditure\_percent\_of\_gdp 1.25452 0.21322  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.33184 0.74085  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.84424 0.40099  
## freedom 0.80814 0.42135  
## generosity 1.10830 0.27097  
## trust\_government\_corruption 0.76203 0.44823  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.4  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.57636 0.56840   
## employment\_industry\_percent\_of\_employed 0.00564 0.00692   
## unemployment\_percent\_of\_labour\_force -0.00007 0.00858   
## agricultural\_production\_index\_2004\_2006\_100 0.00031 0.00231   
## urban\_population\_percent\_of\_total\_population 0.00131 0.00254   
## health\_total\_expenditure\_percent\_of\_gdp 0.00893 0.01326   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00056 0.00282   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00059 0.00162   
## freedom 0.06948 0.30149   
## generosity 0.33949 0.28474   
## trust\_government\_corruption 0.60781 0.50568   
## t value Pr(>|t|)  
## (Intercept) -1.01401 0.31356  
## employment\_industry\_percent\_of\_employed 0.81462 0.41765  
## unemployment\_percent\_of\_labour\_force -0.00769 0.99388  
## agricultural\_production\_index\_2004\_2006\_100 0.13352 0.89411  
## urban\_population\_percent\_of\_total\_population 0.51569 0.60746  
## health\_total\_expenditure\_percent\_of\_gdp 0.67364 0.50243  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.19915 0.84263  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.36680 0.71471  
## freedom 0.23047 0.81830  
## generosity 1.19228 0.23659  
## trust\_government\_corruption 1.20197 0.23284  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.5  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.31291 0.57125   
## employment\_industry\_percent\_of\_employed 0.00508 0.00678   
## unemployment\_percent\_of\_labour\_force -0.00010 0.00687   
## agricultural\_production\_index\_2004\_2006\_100 0.00061 0.00243   
## urban\_population\_percent\_of\_total\_population -0.00049 0.00317   
## health\_total\_expenditure\_percent\_of\_gdp 0.01038 0.01590   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00180 0.00350   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00202 0.00220   
## freedom 0.21212 0.39054   
## generosity 0.16229 0.30505   
## trust\_government\_corruption 1.07603 0.57945   
## t value Pr(>|t|)  
## (Intercept) -0.54777 0.58533  
## employment\_industry\_percent\_of\_employed 0.74943 0.45574  
## unemployment\_percent\_of\_labour\_force -0.01401 0.98886  
## agricultural\_production\_index\_2004\_2006\_100 0.24949 0.80361  
## urban\_population\_percent\_of\_total\_population -0.15578 0.87659  
## health\_total\_expenditure\_percent\_of\_gdp 0.65323 0.51544  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.51309 0.60927  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.91653 0.36208  
## freedom 0.54315 0.58850  
## generosity 0.53202 0.59615  
## trust\_government\_corruption 1.85698 0.06690  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.6  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.50698 0.58915   
## employment\_industry\_percent\_of\_employed 0.00571 0.00628   
## unemployment\_percent\_of\_labour\_force -0.00321 0.00852   
## agricultural\_production\_index\_2004\_2006\_100 0.00093 0.00247   
## urban\_population\_percent\_of\_total\_population 0.00119 0.00295   
## health\_total\_expenditure\_percent\_of\_gdp 0.00551 0.01651   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00003 0.00332   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00128 0.00256   
## freedom 0.31027 0.39715   
## generosity -0.06600 0.36750   
## trust\_government\_corruption 1.33704 0.61688   
## t value Pr(>|t|)  
## (Intercept) -0.86052 0.39201  
## employment\_industry\_percent\_of\_employed 0.90842 0.36632  
## unemployment\_percent\_of\_labour\_force -0.37655 0.70748  
## agricultural\_production\_index\_2004\_2006\_100 0.37506 0.70859  
## urban\_population\_percent\_of\_total\_population 0.40441 0.68697  
## health\_total\_expenditure\_percent\_of\_gdp 0.33361 0.73952  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00929 0.99261  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.50108 0.61766  
## freedom 0.78124 0.43691  
## generosity -0.17960 0.85791  
## trust\_government\_corruption 2.16742 0.03310  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.7  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.30792 0.58541   
## employment\_industry\_percent\_of\_employed 0.00515 0.00806   
## unemployment\_percent\_of\_labour\_force 0.00206 0.00820   
## agricultural\_production\_index\_2004\_2006\_100 -0.00001 0.00279   
## urban\_population\_percent\_of\_total\_population 0.00158 0.00397   
## health\_total\_expenditure\_percent\_of\_gdp 0.00527 0.01947   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00039 0.00357   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00040 0.00292   
## freedom 0.32363 0.43927   
## generosity -0.13481 0.42459   
## trust\_government\_corruption 1.29229 0.67253   
## t value Pr(>|t|)  
## (Intercept) -0.52600 0.60031  
## employment\_industry\_percent\_of\_employed 0.63963 0.52420  
## unemployment\_percent\_of\_labour\_force 0.25159 0.80199  
## agricultural\_production\_index\_2004\_2006\_100 -0.00534 0.99576  
## urban\_population\_percent\_of\_total\_population 0.39694 0.69245  
## health\_total\_expenditure\_percent\_of\_gdp 0.27091 0.78714  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.10967 0.91294  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.13816 0.89045  
## freedom 0.73674 0.46338  
## generosity -0.31752 0.75166  
## trust\_government\_corruption 1.92154 0.05814  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.8  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.33370 0.72616   
## employment\_industry\_percent\_of\_employed -0.00085 0.00838   
## unemployment\_percent\_of\_labour\_force 0.00359 0.01181   
## agricultural\_production\_index\_2004\_2006\_100 -0.00053 0.00320   
## urban\_population\_percent\_of\_total\_population 0.00202 0.00529   
## health\_total\_expenditure\_percent\_of\_gdp -0.01676 0.02555   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00263 0.00480   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00286 0.00336   
## freedom 0.48868 0.55980   
## generosity -0.26734 0.60380   
## trust\_government\_corruption 1.10049 0.81349   
## t value Pr(>|t|)  
## (Intercept) -0.45953 0.64707  
## employment\_industry\_percent\_of\_employed -0.10168 0.91926  
## unemployment\_percent\_of\_labour\_force 0.30414 0.76179  
## agricultural\_production\_index\_2004\_2006\_100 -0.16695 0.86782  
## urban\_population\_percent\_of\_total\_population 0.38079 0.70434  
## health\_total\_expenditure\_percent\_of\_gdp -0.65594 0.51370  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.54913 0.58441  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.85237 0.39649  
## freedom 0.87297 0.38523  
## generosity -0.44277 0.65910  
## trust\_government\_corruption 1.35280 0.17984  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.9  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.10133 1.00813   
## employment\_industry\_percent\_of\_employed -0.00090 0.00947   
## unemployment\_percent\_of\_labour\_force -0.01768 0.01645   
## agricultural\_production\_index\_2004\_2006\_100 0.00134 0.00352   
## urban\_population\_percent\_of\_total\_population 0.00759 0.00540   
## health\_total\_expenditure\_percent\_of\_gdp 0.00203 0.03004   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00229 0.00764   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00018 0.00375   
## freedom 0.85962 0.70606   
## generosity -0.35119 0.90622   
## trust\_government\_corruption 0.25762 0.94724   
## t value Pr(>|t|)  
## (Intercept) -0.10051 0.92018  
## employment\_industry\_percent\_of\_employed -0.09492 0.92461  
## unemployment\_percent\_of\_labour\_force -1.07466 0.28568  
## agricultural\_production\_index\_2004\_2006\_100 0.37927 0.70547  
## urban\_population\_percent\_of\_total\_population 1.40671 0.16329  
## health\_total\_expenditure\_percent\_of\_gdp 0.06747 0.94637  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.29987 0.76503  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.04904 0.96101  
## freedom 1.21748 0.22691  
## generosity -0.38753 0.69937  
## trust\_government\_corruption 0.27197 0.78633



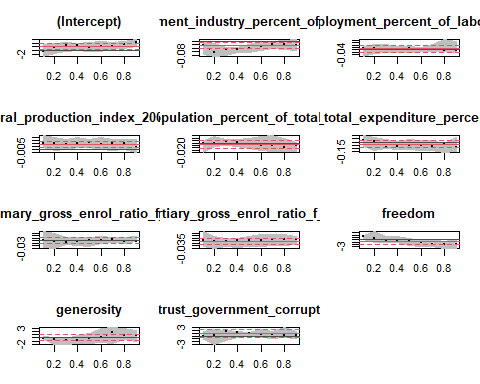
## Quantile Regression Analysis of Deviance Table  
##   
## Model: migration\_growth ~ employment\_industry\_percent\_of\_employed + unemployment\_percent\_of\_labour\_force + agricultural\_production\_index\_2004\_2006\_100 + urban\_population\_percent\_of\_total\_population + health\_total\_expenditure\_percent\_of\_gdp + education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + freedom + generosity + trust\_government\_corruption  
## Joint Test of Equality of Slopes: tau in { 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 }  
##   
## Df Resid Df F value Pr(>F)  
## 1 80 757 0.8318 0.8495

print("Kvantilių regresija natūraliam prieaugiui")

## [1] "Kvantilių regresija natūraliam prieaugiui"

data <- regression\_train %>% dplyr::select(-migration\_growth)  
tau <- seq(0.1,0.9,0.1)  
formula <- natural\_growth ~ .  
  
model\_quantile\_migration <- quantile\_fit()

##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.1  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) -0.42120 1.93842   
## employment\_industry\_percent\_of\_employed -0.02665 0.02579   
## unemployment\_percent\_of\_labour\_force 0.01321 0.02721   
## agricultural\_production\_index\_2004\_2006\_100 0.01077 0.00613   
## urban\_population\_percent\_of\_total\_population -0.00122 0.00949   
## health\_total\_expenditure\_percent\_of\_gdp 0.02983 0.06051   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00152 0.01523   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01904 0.00720   
## freedom 1.39068 1.05332   
## generosity -0.27885 0.72989   
## trust\_government\_corruption -0.28614 1.40898   
## t value Pr(>|t|)  
## (Intercept) -0.21729 0.82850  
## employment\_industry\_percent\_of\_employed -1.03361 0.30422  
## unemployment\_percent\_of\_labour\_force 0.48548 0.62857  
## agricultural\_production\_index\_2004\_2006\_100 1.75561 0.08272  
## urban\_population\_percent\_of\_total\_population -0.12885 0.89778  
## health\_total\_expenditure\_percent\_of\_gdp 0.49295 0.62331  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.09996 0.92061  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -2.64570 0.00969  
## freedom 1.32029 0.19024  
## generosity -0.38205 0.70337  
## trust\_government\_corruption -0.20308 0.83955  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.2  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 1.68145 1.77738   
## employment\_industry\_percent\_of\_employed -0.06124 0.02171   
## unemployment\_percent\_of\_labour\_force 0.00899 0.02277   
## agricultural\_production\_index\_2004\_2006\_100 0.01031 0.00659   
## urban\_population\_percent\_of\_total\_population -0.00178 0.00785   
## health\_total\_expenditure\_percent\_of\_gdp 0.01368 0.06180   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00350 0.01390   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.02107 0.00750   
## freedom 0.75076 1.17501   
## generosity -0.64089 0.81195   
## trust\_government\_corruption 0.35721 1.26013   
## t value Pr(>|t|)  
## (Intercept) 0.94603 0.34679  
## employment\_industry\_percent\_of\_employed -2.82111 0.00594  
## unemployment\_percent\_of\_labour\_force 0.39483 0.69394  
## agricultural\_production\_index\_2004\_2006\_100 1.56533 0.12118  
## urban\_population\_percent\_of\_total\_population -0.22684 0.82109  
## health\_total\_expenditure\_percent\_of\_gdp 0.22145 0.82527  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.25180 0.80179  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -2.80816 0.00616  
## freedom 0.63894 0.52456  
## generosity -0.78933 0.43209  
## trust\_government\_corruption 0.28347 0.77750  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.3  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 2.52828 1.64688   
## employment\_industry\_percent\_of\_employed -0.04821 0.01813   
## unemployment\_percent\_of\_labour\_force 0.00704 0.02186   
## agricultural\_production\_index\_2004\_2006\_100 0.00941 0.00567   
## urban\_population\_percent\_of\_total\_population 0.00282 0.00699   
## health\_total\_expenditure\_percent\_of\_gdp -0.08034 0.06018   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00526 0.01286   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.02103 0.00639   
## freedom -0.07925 1.16353   
## generosity -0.79698 0.75777   
## trust\_government\_corruption 1.73131 1.29610   
## t value Pr(>|t|)  
## (Intercept) 1.53519 0.12841  
## employment\_industry\_percent\_of\_employed -2.65968 0.00933  
## unemployment\_percent\_of\_labour\_force 0.32209 0.74816  
## agricultural\_production\_index\_2004\_2006\_100 1.65938 0.10068  
## urban\_population\_percent\_of\_total\_population 0.40244 0.68836  
## health\_total\_expenditure\_percent\_of\_gdp -1.33495 0.18542  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.40920 0.68341  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -3.29161 0.00145  
## freedom -0.06811 0.94585  
## generosity -1.05173 0.29587  
## trust\_government\_corruption 1.33579 0.18514  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.4  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 2.54139 1.51997   
## employment\_industry\_percent\_of\_employed -0.04577 0.01538   
## unemployment\_percent\_of\_labour\_force 0.00159 0.02049   
## agricultural\_production\_index\_2004\_2006\_100 0.00996 0.00623   
## urban\_population\_percent\_of\_total\_population 0.00128 0.00687   
## health\_total\_expenditure\_percent\_of\_gdp -0.07234 0.05775   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00404 0.01126   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01986 0.00626   
## freedom -0.25945 1.20921   
## generosity -0.88157 1.01058   
## trust\_government\_corruption 1.42596 1.04033   
## t value Pr(>|t|)  
## (Intercept) 1.67200 0.09816  
## employment\_industry\_percent\_of\_employed -2.97487 0.00380  
## unemployment\_percent\_of\_labour\_force 0.07765 0.93829  
## agricultural\_production\_index\_2004\_2006\_100 1.59724 0.11388  
## urban\_population\_percent\_of\_total\_population 0.18702 0.85209  
## health\_total\_expenditure\_percent\_of\_gdp -1.25257 0.21376  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.35868 0.72071  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -3.17096 0.00211  
## freedom -0.21456 0.83062  
## generosity -0.87234 0.38545  
## trust\_government\_corruption 1.37068 0.17404  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.5  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 1.60165 1.53934   
## employment\_industry\_percent\_of\_employed -0.03852 0.01733   
## unemployment\_percent\_of\_labour\_force 0.00790 0.02425   
## agricultural\_production\_index\_2004\_2006\_100 0.01169 0.00596   
## urban\_population\_percent\_of\_total\_population -0.00193 0.00741   
## health\_total\_expenditure\_percent\_of\_gdp -0.04795 0.05238   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00247 0.01184   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01954 0.00642   
## freedom -0.42602 1.09216   
## generosity -0.23533 1.19019   
## trust\_government\_corruption 0.90956 0.85438   
## t value Pr(>|t|)  
## (Intercept) 1.04048 0.30104  
## employment\_industry\_percent\_of\_employed -2.22251 0.02887  
## unemployment\_percent\_of\_labour\_force 0.32554 0.74556  
## agricultural\_production\_index\_2004\_2006\_100 1.96369 0.05280  
## urban\_population\_percent\_of\_total\_population -0.25981 0.79563  
## health\_total\_expenditure\_percent\_of\_gdp -0.91535 0.36257  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.20858 0.83527  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -3.04432 0.00309  
## freedom -0.39007 0.69745  
## generosity -0.19772 0.84373  
## trust\_government\_corruption 1.06459 0.29004  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.6  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 2.44890 1.57908   
## employment\_industry\_percent\_of\_employed -0.03048 0.01895   
## unemployment\_percent\_of\_labour\_force 0.00573 0.01908   
## agricultural\_production\_index\_2004\_2006\_100 0.00955 0.00612   
## urban\_population\_percent\_of\_total\_population -0.00482 0.00867   
## health\_total\_expenditure\_percent\_of\_gdp -0.04561 0.05396   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00009 0.01059   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01851 0.00659   
## freedom -0.93331 1.04947   
## generosity 0.02964 1.43427   
## trust\_government\_corruption 0.56498 1.02814   
## t value Pr(>|t|)  
## (Intercept) 1.55084 0.12461  
## employment\_industry\_percent\_of\_employed -1.60868 0.11135  
## unemployment\_percent\_of\_labour\_force 0.30044 0.76457  
## agricultural\_production\_index\_2004\_2006\_100 1.56042 0.12233  
## urban\_population\_percent\_of\_total\_population -0.55630 0.57945  
## health\_total\_expenditure\_percent\_of\_gdp -0.84540 0.40023  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00863 0.99314  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -2.81017 0.00613  
## freedom -0.88932 0.37631  
## generosity 0.02067 0.98356  
## trust\_government\_corruption 0.54952 0.58407  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.7  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 2.25692 1.56131   
## employment\_industry\_percent\_of\_employed -0.02120 0.01728   
## unemployment\_percent\_of\_labour\_force 0.00136 0.01798   
## agricultural\_production\_index\_2004\_2006\_100 0.01055 0.00656   
## urban\_population\_percent\_of\_total\_population -0.00528 0.00848   
## health\_total\_expenditure\_percent\_of\_gdp -0.06905 0.05209   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.00299 0.01070   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01698 0.00680   
## freedom -1.99306 1.01761   
## generosity 1.72148 1.37841   
## trust\_government\_corruption 0.51851 0.98507   
## t value Pr(>|t|)  
## (Intercept) 1.44552 0.15194  
## employment\_industry\_percent\_of\_employed -1.22652 0.22335  
## unemployment\_percent\_of\_labour\_force 0.07556 0.93995  
## agricultural\_production\_index\_2004\_2006\_100 1.60813 0.11147  
## urban\_population\_percent\_of\_total\_population -0.62276 0.53509  
## health\_total\_expenditure\_percent\_of\_gdp -1.32563 0.18847  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.27931 0.78068  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -2.49789 0.01440  
## freedom -1.95858 0.05340  
## generosity 1.24889 0.21509  
## trust\_government\_corruption 0.52637 0.59998  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.8  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 3.51817 1.27208   
## employment\_industry\_percent\_of\_employed -0.01913 0.01792   
## unemployment\_percent\_of\_labour\_force -0.00572 0.02489   
## agricultural\_production\_index\_2004\_2006\_100 0.00997 0.00587   
## urban\_population\_percent\_of\_total\_population -0.00870 0.00842   
## health\_total\_expenditure\_percent\_of\_gdp -0.04272 0.05163   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00499 0.00875   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01648 0.00706   
## freedom -1.81640 1.04281   
## generosity 0.78080 1.18301   
## trust\_government\_corruption 0.32198 1.16655   
## t value Pr(>|t|)  
## (Intercept) 2.76568 0.00695  
## employment\_industry\_percent\_of\_employed -1.06735 0.28880  
## unemployment\_percent\_of\_labour\_force -0.22994 0.81869  
## agricultural\_production\_index\_2004\_2006\_100 1.69901 0.09293  
## urban\_population\_percent\_of\_total\_population -1.03305 0.30448  
## health\_total\_expenditure\_percent\_of\_gdp -0.82732 0.41035  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.57055 0.56979  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -2.33439 0.02191  
## freedom -1.74184 0.08511  
## generosity 0.66001 0.51101  
## trust\_government\_corruption 0.27601 0.78321  
##   
## Call: rq(formula = formula, tau = tau, data = data)  
##   
## tau: [1] 0.9  
##   
## Coefficients:  
## Value Std. Error  
## (Intercept) 4.66105 1.43052   
## employment\_industry\_percent\_of\_employed -0.02358 0.01870   
## unemployment\_percent\_of\_labour\_force -0.00163 0.02782   
## agricultural\_production\_index\_2004\_2006\_100 0.00310 0.00545   
## urban\_population\_percent\_of\_total\_population -0.00424 0.00823   
## health\_total\_expenditure\_percent\_of\_gdp -0.07877 0.05170   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.00377 0.00853   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.01734 0.00628   
## freedom -1.87499 1.10334   
## generosity 0.59091 1.07216   
## trust\_government\_corruption 0.21857 1.15676   
## t value Pr(>|t|)  
## (Intercept) 3.25829 0.00161  
## employment\_industry\_percent\_of\_employed -1.26125 0.21063  
## unemployment\_percent\_of\_labour\_force -0.05866 0.95336  
## agricultural\_production\_index\_2004\_2006\_100 0.56818 0.57139  
## urban\_population\_percent\_of\_total\_population -0.51528 0.60768  
## health\_total\_expenditure\_percent\_of\_gdp -1.52374 0.13124  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.44259 0.65917  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -2.76400 0.00698  
## freedom -1.69937 0.09286  
## generosity 0.55114 0.58297  
## trust\_government\_corruption 0.18895 0.85058



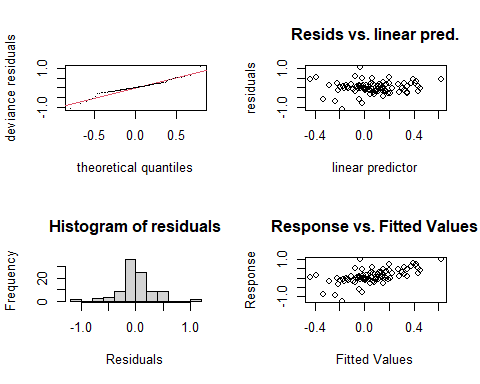
## Quantile Regression Analysis of Deviance Table  
##   
## Model: natural\_growth ~ employment\_industry\_percent\_of\_employed + unemployment\_percent\_of\_labour\_force + agricultural\_production\_index\_2004\_2006\_100 + urban\_population\_percent\_of\_total\_population + health\_total\_expenditure\_percent\_of\_gdp + education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + freedom + generosity + trust\_government\_corruption  
## Joint Test of Equality of Slopes: tau in { 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 }  
##   
## Df Resid Df F value Pr(>F)   
## 1 80 793 2.2304 2.768e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library(mgcv)  
library(gratia)  
  
fit\_gam <- function(formula, data) {  
 model\_gam <- gam(formula, data = data, select = TRUE)  
 gam.check(model\_gam)  
 summary(model\_gam)  
 draw(model\_gam)  
 k.check(model\_gam)  
 model\_gam  
}

print("GAM migracijos prieaugiui")

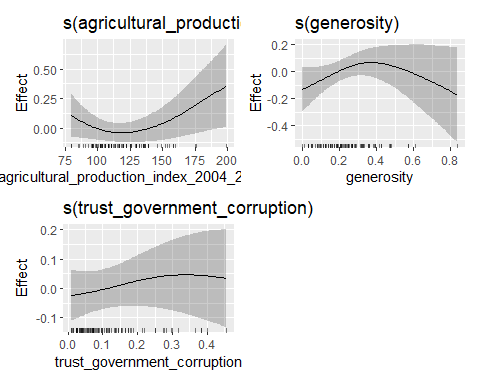
## [1] "GAM migracijos prieaugiui"

data <- regression\_train %>% dplyr::select(-natural\_growth) %>% slice(-outlier\_indices)  
  
# model\_gam\_migration <- fit\_gam(migration\_growth ~ employment\_industry\_percent\_of\_employed +  
# s(unemployment\_percent\_of\_labour\_force,k=15) +   
# s(agricultural\_production\_index\_2004\_2006\_100,k=15) +   
# urban\_population\_percent\_of\_total\_population +  
# health\_total\_expenditure\_percent\_of\_gdp +   
# education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
# education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
# freedom +   
# s(generosity,k=15) +  
# s(trust\_government\_corruption,k=15), data)  
#   
#   
# draw(model\_gam\_migration)  
# k.check(model\_gam\_migration)  
# summary(model\_gam\_migration)  
  
  
model\_gam\_migration <- fit\_gam(migration\_growth ~ employment\_industry\_percent\_of\_employed +  
 s(agricultural\_production\_index\_2004\_2006\_100,k=15) +   
 urban\_population\_percent\_of\_total\_population +  
 health\_total\_expenditure\_percent\_of\_gdp +   
 education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
 education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
 freedom +   
 s(generosity,k=15) +  
 s(trust\_government\_corruption,k=15), data)



##   
## Method: GCV Optimizer: magic  
## Smoothing parameter selection converged after 15 iterations.  
## The RMS GCV score gradient at convergence was 4.587417e-08 .  
## The Hessian was positive definite.  
## Model rank = 49 / 49   
##   
## Basis dimension (k) checking results. Low p-value (k-index<1) may  
## indicate that k is too low, especially if edf is close to k'.  
##   
## k' edf k-index p-value  
## s(agricultural\_production\_index\_2004\_2006\_100) 14.000 1.797 1.26 0.98  
## s(generosity) 14.000 1.580 0.88 0.12  
## s(trust\_government\_corruption) 14.000 0.495 1.03 0.61

draw(model\_gam\_migration)



k.check(model\_gam\_migration)

## k' edf k-index p-value  
## s(agricultural\_production\_index\_2004\_2006\_100) 14 1.7972446 1.2597192 0.9925  
## s(generosity) 14 1.5804495 0.8750931 0.1000  
## s(trust\_government\_corruption) 14 0.4952306 1.0329735 0.5800

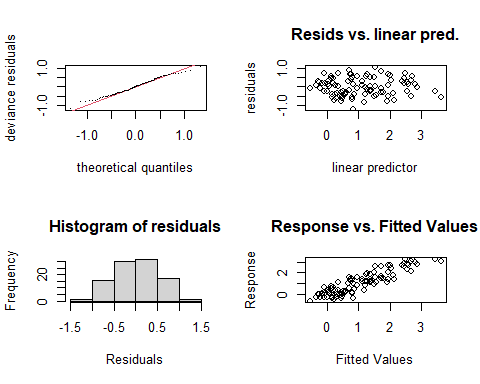
summary(model\_gam\_migration)

##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## migration\_growth ~ employment\_industry\_percent\_of\_employed +   
## s(agricultural\_production\_index\_2004\_2006\_100, k = 15) +   
## urban\_population\_percent\_of\_total\_population + health\_total\_expenditure\_percent\_of\_gdp +   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
## freedom + s(generosity, k = 15) + s(trust\_government\_corruption,   
## k = 15)  
##   
## Parametric coefficients:  
## Estimate Std. Error  
## (Intercept) -0.872661 0.374593  
## employment\_industry\_percent\_of\_employed 0.009393 0.005496  
## urban\_population\_percent\_of\_total\_population 0.001737 0.002512  
## health\_total\_expenditure\_percent\_of\_gdp 0.006767 0.016331  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.001940 0.003343  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.001423 0.001907  
## freedom 0.802932 0.289687  
## t value Pr(>|t|)   
## (Intercept) -2.330 0.0223 \*   
## employment\_industry\_percent\_of\_employed 1.709 0.0912 .   
## urban\_population\_percent\_of\_total\_population 0.691 0.4914   
## health\_total\_expenditure\_percent\_of\_gdp 0.414 0.6797   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.580 0.5632   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop 0.746 0.4577   
## freedom 2.772 0.0069 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Approximate significance of smooth terms:  
## edf Ref.df F p-value   
## s(agricultural\_production\_index\_2004\_2006\_100) 1.7972 14 0.323 0.0689 .  
## s(generosity) 1.5804 14 0.273 0.0805 .  
## s(trust\_government\_corruption) 0.4952 13 0.064 0.1808   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## R-sq.(adj) = 0.275 Deviance explained = 35.3%  
## GCV = 0.11067 Scale est. = 0.097731 n = 93

print("GAM natūraliam prieaugiui")

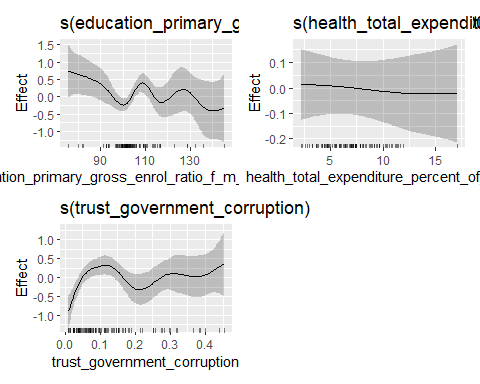
## [1] "GAM natūraliam prieaugiui"

# model\_gam\_natural <- fit\_gam(natural\_growth ~ employment\_industry\_percent\_of\_employed +  
# s(unemployment\_percent\_of\_labour\_force) +   
# agricultural\_production\_index\_2004\_2006\_100 +   
# urban\_population\_percent\_of\_total\_population +  
# s(health\_total\_expenditure\_percent\_of\_gdp) +   
# s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) +   
# education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
# freedom +   
# s(generosity) +  
# s(trust\_government\_corruption), regression\_train %>% dplyr::select(-migration\_growth))  
#   
#   
# draw(model\_gam\_natural)  
# k.check(model\_gam\_natural)  
# summary(model\_gam\_natural)  
  
  
model\_gam\_natural <- fit\_gam(natural\_growth ~ employment\_industry\_percent\_of\_employed +   
 agricultural\_production\_index\_2004\_2006\_100 +   
 urban\_population\_percent\_of\_total\_population +  
 s(health\_total\_expenditure\_percent\_of\_gdp) +   
 s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) +   
 education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
 freedom +   
 s(trust\_government\_corruption), regression\_train %>% dplyr::select(-migration\_growth))



##   
## Method: GCV Optimizer: magic  
## Smoothing parameter selection converged after 22 iterations.  
## The RMS GCV score gradient at convergence was 3.986037e-08 .  
## The Hessian was positive definite.  
## Model rank = 33 / 33   
##   
## Basis dimension (k) checking results. Low p-value (k-index<1) may  
## indicate that k is too low, especially if edf is close to k'.  
##   
## k' edf k-index  
## s(health\_total\_expenditure\_percent\_of\_gdp) 9.00 0.15 0.93  
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) 9.00 6.82 1.07  
## s(trust\_government\_corruption) 9.00 6.17 1.06  
## p-value  
## s(health\_total\_expenditure\_percent\_of\_gdp) 0.24  
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) 0.72  
## s(trust\_government\_corruption) 0.64

draw(model\_gam\_natural)



k.check(model\_gam\_natural)

## k' edf k-index  
## s(health\_total\_expenditure\_percent\_of\_gdp) 9 0.1502352 0.9280672  
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) 9 6.8163434 1.0676728  
## s(trust\_government\_corruption) 9 6.1733842 1.0570000  
## p-value  
## s(health\_total\_expenditure\_percent\_of\_gdp) 0.2200  
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) 0.7200  
## s(trust\_government\_corruption) 0.6275

summary(model\_gam\_natural)

##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## natural\_growth ~ employment\_industry\_percent\_of\_employed + agricultural\_production\_index\_2004\_2006\_100 +   
## urban\_population\_percent\_of\_total\_population + s(health\_total\_expenditure\_percent\_of\_gdp) +   
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) +   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + freedom +   
## s(trust\_government\_corruption)  
##   
## Parametric coefficients:  
## Estimate Std. Error  
## (Intercept) 1.664170 0.459710  
## employment\_industry\_percent\_of\_employed -0.019568 0.008166  
## agricultural\_production\_index\_2004\_2006\_100 0.010253 0.002860  
## urban\_population\_percent\_of\_total\_population -0.006251 0.004304  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -0.015914 0.003159  
## freedom -0.664638 0.498042  
## t value Pr(>|t|)   
## (Intercept) 3.620 0.000522 \*\*\*  
## employment\_industry\_percent\_of\_employed -2.396 0.018955 \*   
## agricultural\_production\_index\_2004\_2006\_100 3.585 0.000585 \*\*\*  
## urban\_population\_percent\_of\_total\_population -1.452 0.150403   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop -5.037 2.99e-06 \*\*\*  
## freedom -1.335 0.185929   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Approximate significance of smooth terms:  
## edf Ref.df F  
## s(health\_total\_expenditure\_percent\_of\_gdp) 0.1502 9 0.019  
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) 6.8163 9 2.314  
## s(trust\_government\_corruption) 6.1734 9 2.959  
## p-value   
## s(health\_total\_expenditure\_percent\_of\_gdp) 0.293332   
## s(education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop) 0.004824 \*\*   
## s(trust\_government\_corruption) 0.000346 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## R-sq.(adj) = 0.768 Deviance explained = 81.2%  
## GCV = 0.34189 Scale est. = 0.27443 n = 97

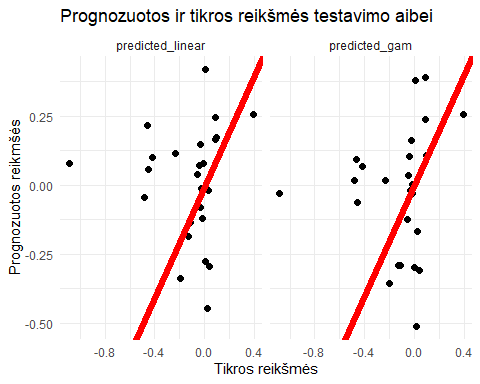
library(yardstick)  
  
  
regression\_eval <- function(column, model\_linear, model\_gam) {  
 print(AIC(model\_linear))  
 print(AIC(model\_gam))  
  
  
 regression\_test <- test %>%  
 mutate(  
 predicted\_linear = predict(model\_linear, test),  
 predicted\_gam = predict(model\_gam, test)  
 )  
  
 set <- metric\_set(rmse, mae)  
   
 print("Tiesinis modelis")  
 print(set(regression\_test, {{ column }}, predicted\_linear))  
 print("GAM modelis")  
 print(set(regression\_test, {{ column }}, predicted\_gam))  
  
 regression\_test %>%  
 pivot\_longer(c(predicted\_gam, predicted\_linear)) %>%  
 mutate(name = factor(name, levels = c("predicted\_linear", "predicted\_gam"))) %>%  
 ggplot(aes({{ column }}, value)) +  
 geom\_point(size = 2) +  
 facet\_wrap(vars(name)) +  
 geom\_abline(color = "red", size = 2.25) +  
 labs(x = "Tikros reikšmės", y = "Prognozuotos reikmšės",  
 title = "Prognozuotos ir tikros reikšmės testavimo aibei") +  
 theme\_minimal()  
}

print("Regresija migracijos prieaugiui")

## [1] "Regresija migracijos prieaugiui"

regression\_eval(migration\_growth, model\_linear\_migration, model\_gam\_migration)

## [1] 60.18487  
## [1] 59.83029  
## [1] "Tiesinis modelis"  
## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 0.368  
## 2 mae standard 0.260  
## [1] "GAM modelis"  
## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 0.351  
## 2 mae standard 0.264

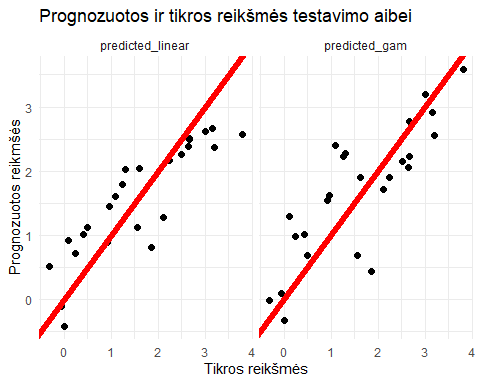


print("Regresija natūraliam prieaugiui")

## [1] "Regresija natūraliam prieaugiui"

regression\_eval(natural\_growth, model\_linear\_natural, model\_gam\_natural)

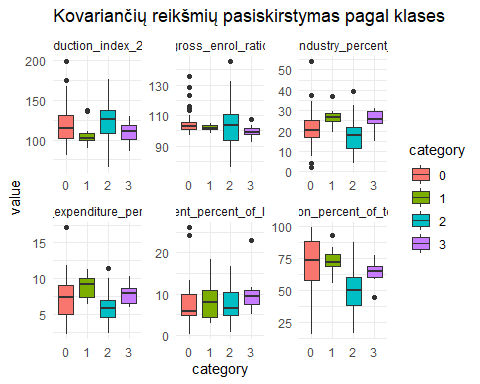
## [1] 196.1966  
## [1] 168.8067  
## [1] "Tiesinis modelis"  
## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 0.593  
## 2 mae standard 0.509  
## [1] "GAM modelis"  
## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 0.676  
## 2 mae standard 0.567



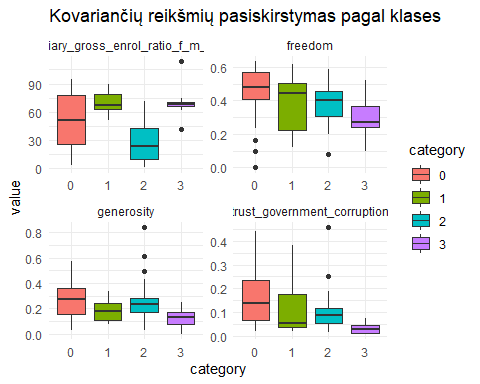
### Classification model

classification\_train <- train %>% dplyr::select(-migration\_growth,-natural\_growth)

classification\_train %>%  
 dplyr::select(1:6, category) %>%  
 pivot\_longer(-category) %>%  
 ggplot(aes(x = category, y = value, fill = category)) +  
 facet\_wrap(vars(name), scales = "free") +  
 geom\_boxplot() +  
 theme\_minimal() + labs(title="Kovariančių reikšmių pasiskirstymas pagal klases")



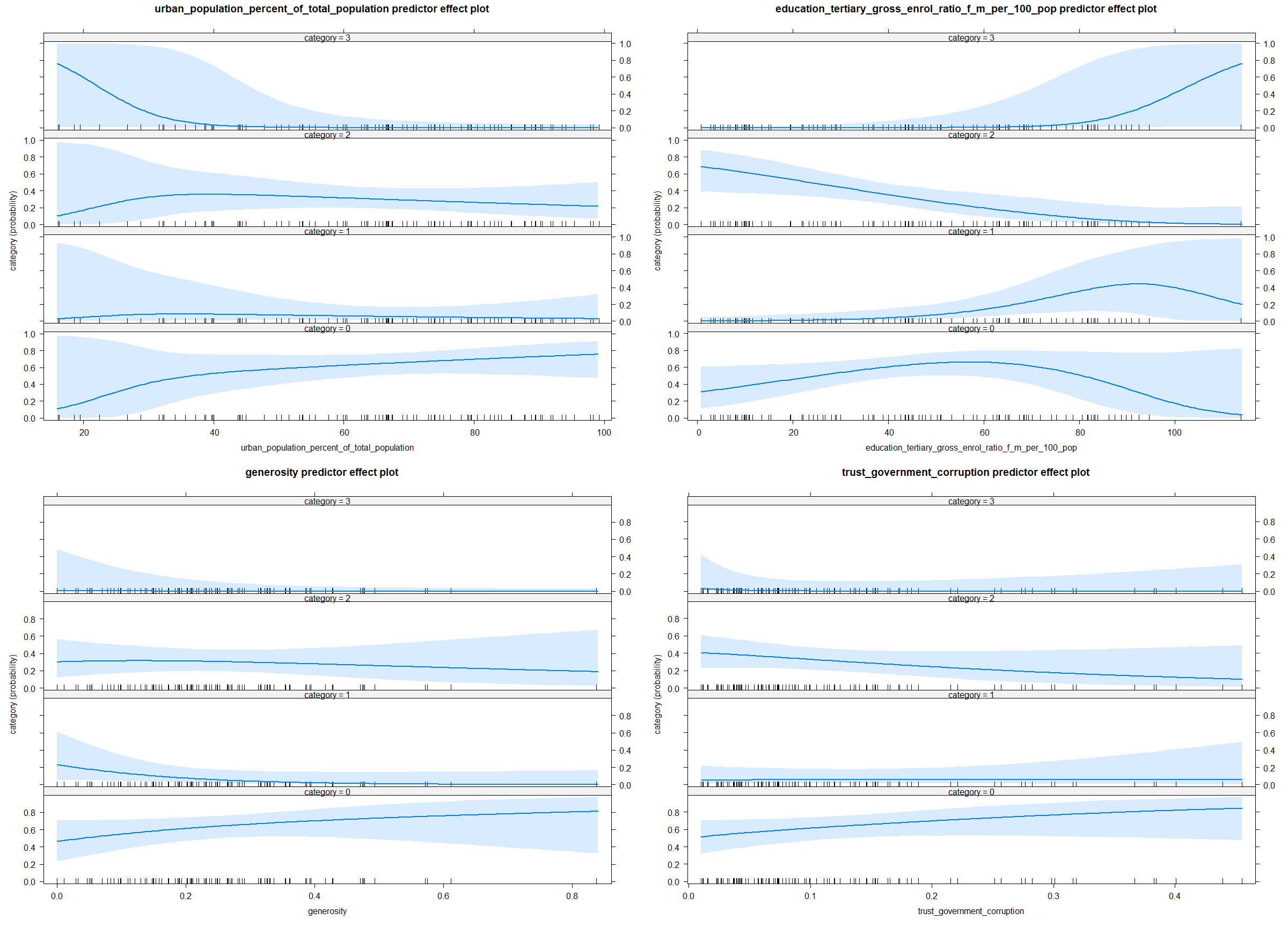
classification\_train %>%  
 dplyr::select(7:length(classification\_train), category) %>%  
 pivot\_longer(-category) %>%  
 ggplot(aes(x = category, y = value, fill = category)) +  
 facet\_wrap(vars(name), scales = "free") +  
 geom\_boxplot() +  
 theme\_minimal() + labs(title="Kovariančių reikšmių pasiskirstymas pagal klases")



summary(model\_logistic)

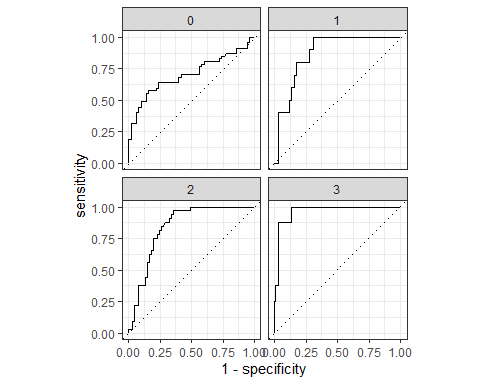
## Call:  
## nnet::multinom(formula = category ~ urban\_population\_percent\_of\_total\_population +   
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + generosity +   
## trust\_government\_corruption, data = classification\_train,   
## trace = FALSE)  
##   
## Coefficients:  
## (Intercept) urban\_population\_percent\_of\_total\_population  
## 1 -1.839977 -0.02292679  
## 2 2.578573 -0.01467582  
## 3 5.840115 -0.19858260  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop generosity  
## 1 0.05915418 -6.961649  
## 2 -0.03401824 -1.223110  
## 3 0.15242683 -10.951204  
## trust\_government\_corruption  
## 1 -0.8558754  
## 2 -4.2532490  
## 3 -39.8926077  
##   
## Std. Errors:  
## (Intercept) urban\_population\_percent\_of\_total\_population  
## 1 2.137231 0.03566644  
## 2 1.017057 0.01748405  
## 3 3.858561 0.09723409  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop generosity  
## 1 0.02581411 3.790026  
## 2 0.01560751 1.809865  
## 3 0.06443421 6.041646  
## trust\_government\_corruption  
## 1 4.031785  
## 2 3.078319  
## 3 22.929942  
##   
## Residual Deviance: 154.4864   
## AIC: 184.4864

plot(predictorEffects(model\_logistic))



eval\_classification <- function(model, classification\_train) {  
 df\_pred\_truth <- tibble(  
 predicted = factor(predict(model, classification\_train)),  
 truth = classification\_train$category  
 ) %>% cbind(as.data.frame(model$fitted.values))  
  
  
 classification\_metrics <- metric\_set(accuracy, f\_meas)  
  
 print("Maišos matrica")  
 print(conf\_mat(df\_pred\_truth,  
 truth = truth,  
 estimate = predicted  
 ))  
  
 print("Modelio kokybės metrikos")  
 print(classification\_metrics(df\_pred\_truth,  
 truth = truth,  
 estimate = predicted  
 ))  
  
 print(roc\_auc(df\_pred\_truth, truth = truth, c("0", "1", "2", "3"), estimator = "macro"))  
  
 roc\_curve(df\_pred\_truth, truth = truth, c("0", "1", "2", "3")) %>%  
 autoplot()  
}  
  
eval\_classification(model\_logistic, classification\_train)

## [1] "Maišos matrica"  
## Truth  
## Prediction 0 1 2 3  
## 0 30 8 12 1  
## 1 3 1 0 0  
## 2 11 0 19 0  
## 3 3 1 1 7  
## [1] "Modelio kokybės metrikos"  
## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 accuracy multiclass 0.588  
## 2 f\_meas macro 0.517  
## # A tibble: 1 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 roc\_auc macro 0.844



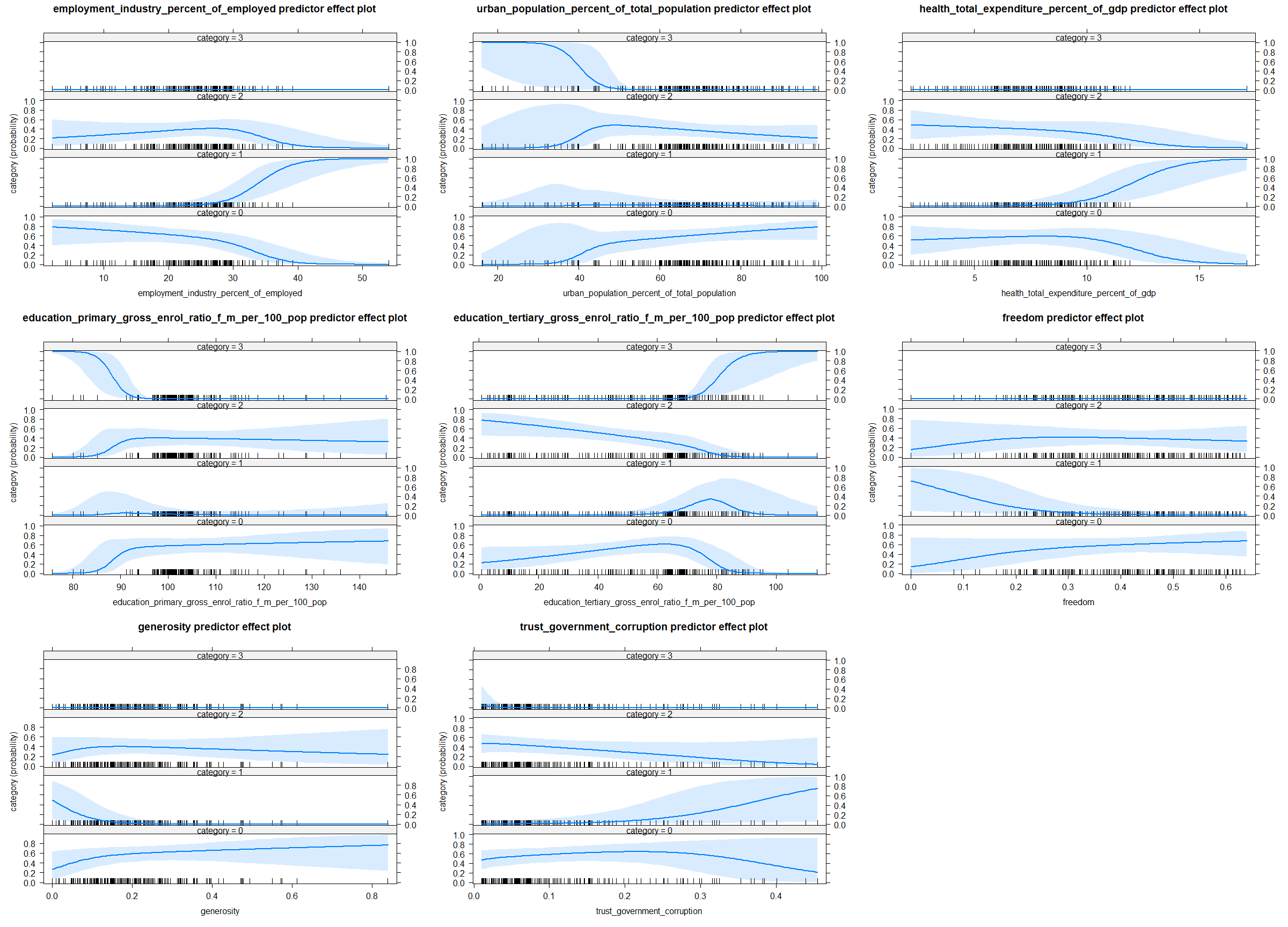
library(themis)  
  
smote\_recipe <- recipe(category ~ .,  
 data = classification\_train) %>%  
 step\_smote(category)  
  
  
smote\_recipe <- prep(smote\_recipe, training = classification\_train)  
  
classification\_train2 <- bake(smote\_recipe, NULL)

model\_logistic2 <- nnet::multinom(category ~ ., data = classification\_train2, trace = FALSE)  
  
model\_logistic2 <- stats::step(model\_logistic2, direction = "both")

summary(model\_logistic2)

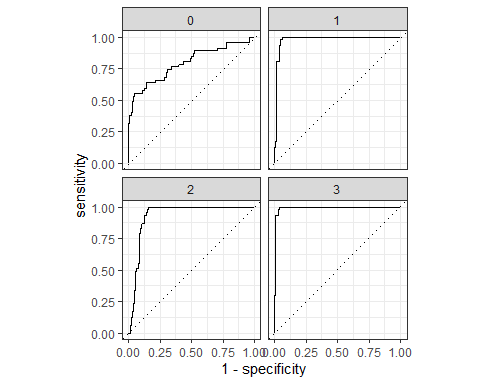
## Call:  
## nnet::multinom(formula = category ~ employment\_industry\_percent\_of\_employed +   
## urban\_population\_percent\_of\_total\_population + health\_total\_expenditure\_percent\_of\_gdp +   
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop + education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop +   
## freedom + generosity + trust\_government\_corruption, data = classification\_train2,   
## trace = FALSE)  
##   
## Coefficients:  
## (Intercept) employment\_industry\_percent\_of\_employed  
## 1 -6.295933 0.39787698  
## 2 4.595093 0.04300743  
## 3 74.492451 -0.07552054  
## urban\_population\_percent\_of\_total\_population  
## 1 -0.04776432  
## 2 -0.02732569  
## 3 -0.39809329  
## health\_total\_expenditure\_percent\_of\_gdp  
## 1 0.87038067  
## 2 -0.06712819  
## 3 -0.00655054  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop  
## 1 -0.102712767  
## 2 -0.008330562  
## 3 -0.653960075  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop freedom generosity  
## 1 0.16003676 -13.289007 -20.283564  
## 2 -0.03058926 -1.290685 -1.207539  
## 3 0.38449203 8.981803 -24.912800  
## trust\_government\_corruption  
## 1 13.905642  
## 2 -4.067271  
## 3 -73.941779  
##   
## Std. Errors:  
## (Intercept) employment\_industry\_percent\_of\_employed  
## 1 10.049979 0.10592725  
## 2 3.140942 0.03994047  
## 3 9.246259 0.10612713  
## urban\_population\_percent\_of\_total\_population  
## 1 0.04656784  
## 2 0.01938735  
## 3 0.10839517  
## health\_total\_expenditure\_percent\_of\_gdp  
## 1 0.2342012  
## 2 0.1363112  
## 3 0.3032902  
## education\_primary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop  
## 1 0.08337851  
## 2 0.02667869  
## 3 0.09284426  
## education\_tertiary\_gross\_enrol\_ratio\_f\_m\_per\_100\_pop freedom generosity  
## 1 0.04193313 4.998406 4.781439  
## 2 0.01562352 2.451155 1.900708  
## 3 0.06801832 5.319684 7.748705  
## trust\_government\_corruption  
## 1 6.534729  
## 2 3.062060  
## 3 5.658675  
##   
## Residual Deviance: 209.9142   
## AIC: 263.9142

plot(predictorEffects(model\_logistic2))



eval\_classification(model\_logistic2, classification\_train2)

## [1] "Maišos matrica"  
## Truth  
## Prediction 0 1 2 3  
## 0 26 0 9 0  
## 1 7 46 0 0  
## 2 12 0 37 0  
## 3 2 1 1 47  
## [1] "Modelio kokybės metrikos"  
## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 accuracy multiclass 0.830  
## 2 f\_meas macro 0.821  
## # A tibble: 1 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 roc\_auc macro 0.925



class\_predictions <- function() {  
 tibble(  
 migration\_growth = predict(model\_gam\_migration, test),  
 natural\_growth = predict(model\_gam\_natural, test)  
 ) %>%  
 mutate(category = factor(case\_when(  
 migration\_growth >= 0 & natural\_growth >= 0 ~ 0, # "P migration, P natural",  
 migration\_growth >= 0 & natural\_growth < 0 ~ 1, # "P migration, N natural",  
 migration\_growth < 0 & natural\_growth >= 0 ~ 2, # "N migration, P natural",  
 TRUE ~ 3  
 ))) %>%  
 pull(category) %>%  
 factor(levels = c(0, 1, 2, 3))  
}  
  
  
df\_pred\_truth <- tibble(  
 predicted\_1 =  
 factor(predict(model\_logistic, test),levels=c(0,1,2,3)),  
 predicted\_2 =  
 factor(predict(model\_logistic2, test),levels=c(0,1,2,3)),  
 predicted\_3 = class\_predictions(),  
 truth = test$category  
)  
  
  
classification\_metrics <- metric\_set(accuracy, f\_meas)  
  
  
print("Pradinis modelis")

## [1] "Pradinis modelis"

conf\_mat(df\_pred\_truth,  
 truth = truth,  
 estimate = predicted\_1  
)

## Truth  
## Prediction 0 1 2 3  
## 0 2 0 3 0  
## 1 0 1 0 0  
## 2 5 0 10 0  
## 3 0 1 2 1

classification\_metrics(df\_pred\_truth, truth, estimate = predicted\_1)

## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 accuracy multiclass 0.56   
## 2 f\_meas macro 0.517

print("SMOTE modelis")

## [1] "SMOTE modelis"

conf\_mat(df\_pred\_truth,  
 truth = truth,  
 estimate = predicted\_2  
)

## Truth  
## Prediction 0 1 2 3  
## 0 2 0 2 0  
## 1 0 1 0 0  
## 2 4 0 11 0  
## 3 1 1 2 1

classification\_metrics(df\_pred\_truth, truth, estimate = predicted\_2)

## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 accuracy multiclass 0.6   
## 2 f\_meas macro 0.524

print("Naudojant du regresijos modelius")

## [1] "Naudojant du regresijos modelius"

conf\_mat(df\_pred\_truth,  
 truth = truth,  
 estimate = predicted\_3  
)

## Truth  
## Prediction 0 1 2 3  
## 0 3 1 7 0  
## 1 0 1 0 1  
## 2 4 0 8 0  
## 3 0 0 0 0

classification\_metrics(df\_pred\_truth, truth, estimate = predicted\_3)

## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 accuracy multiclass 0.48   
## 2 f\_meas macro 0.475