 Vilniaus Universitetas

Išgyvenamumo analizė (trukmės modeliai)

Laboratorinis darbas

Darbą atliko:

Vainius Gataveckas, Matas Gaulia, Dovydas Martinkus

Duomenų Mokslas

3 kursas 2 gr.

Vilnius, 2022

**Turinys**

[Naudoti metodai 3](#_Toc102929215)

[Duomenys ir jų šaltiniai 4](#_Toc102929216)

[Tikslas ir uždaviniai 5](#_Toc102929217)

[Atliktos analizės aprašymas 6](#_Toc102929218)

[1. Naudojant R 6](#_Toc102929219)

# Naudoti metodai

Šiame darbe naudotas apibendrintas adityvus modelis su glodniaisiais splainais. Darbas atliktas naudojant R.

Naudoti R paketai:

*tidyverse*

*rsample*

*corrplot*

*car*

*mgcv*

*gratia*

*effect*

# Duomenys ir jų šaltiniai

Telekomunikacijų bendrovės klientų paslaugų atsisakymo (churn) duomenys pagal demografines ir bendrovės naudojamų paslaugų kovariantes.

Duomenų šaltinis – modeldata R paketas.

Originalus šaltinis IBM Watson Analytics churn data. Prieiga per internetą:

<https://ibm.co/2sOvyvy>.

Duomenis sudaro šie kintamieji:

“churn” – ar klientas atsisakė bendrovės paslaugų.

„tenure“ – mėnesių skaičius, kurį laiką klientas naudojosi įmonės paslaugomis (cenzūruota iš dešinės).

„female“ – ar klientas yra moteris.

„senior\_citien“ – ar klientas yra pensijinio amžiaus.

„partner“ – ar klientas turi partnerį (partnerę).

„dependents“ – ar klientas turi išlaikytinių.

„phone\_service“ – ar klientas naudojami bendrovės telefono ryšiu.

„internet\_service“ – ar klientas iš bendrovės užsisakęs internetą.

„monthly\_charges“ – per mėnesį sumokoma suma.

Kitos kovariantės, esančios originaliame duomenų rinkinyje nėra stipriai susijusios su tyrimo uždaviniais, todėl iš anksto pasirinkta jų tyrime nenaudoti.

# Tikslas ir uždaviniai

Tikslas: Sudarius trukmės modelį telekomunikacijų bendrovės paslaugų atsisakymo trukmei įvertinti demografinių požymių įtaką paslaugų atsisakymui.

Uždaviniai:

Demografinių požymių įtakos įvertinimas naudojant tiriamąja duomenų analizę ir Kaplan-Meier išgyvenamumo funkcijas.

Trukmės regresijos modelių, atsižvelgiančių į kitų kovariančių reikšmes sudarymas:

* Cox semiparametrinio modelio sudarymas
* AFT parametrinio modelio sudarymas

Modelių tinkamumo analizė.

# Atliktos analizės aprašymas

1. Naudojant R

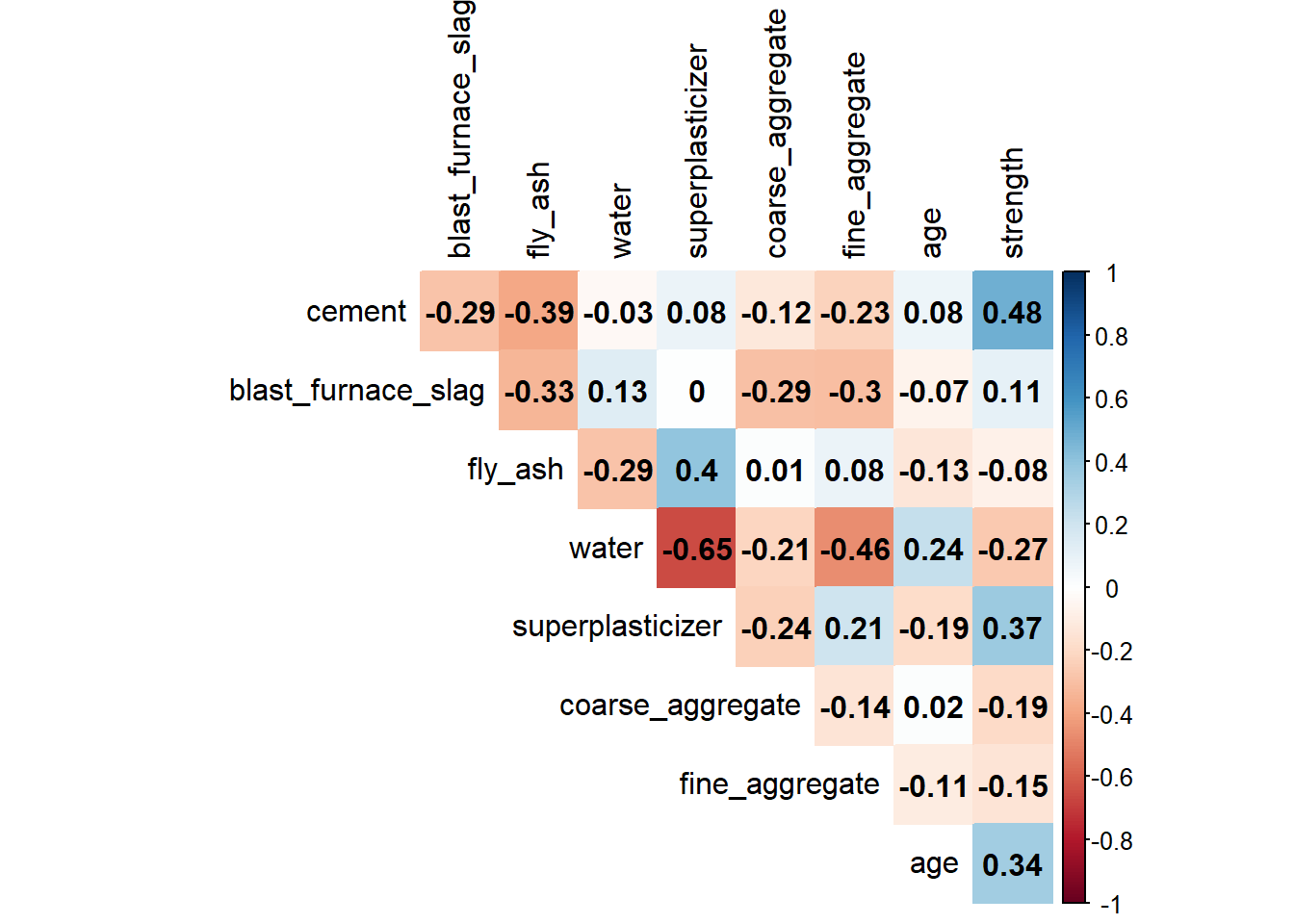
Duomenų aibę apskritai sudaro 1030 stebėjų, tačiau 25 iš jų yra pasikartojantys (pagal visus požymius). Laikyta, kad šitie pasikartojantys stebėjimai neturėtų įtakoti modeliuojamo sąryšio tarp kovariančių ir atsako kintamojo, todėl pasirinkta šiuos stebėjimus pašalinti. Duomenų aibėje nėra praleistų reikšmių. Duomenys padalinti į mokymo ir testavimo aibes naudojant 75-25 santykį.

Nagrinėjant Pirsono koreliacijas tarp betono stiprio ir kovariančių vidutinė arba stipri koreliacija rasta tik su maža dalimi požymių. Kadangi Pirsono koreliacija matuoja tik tiesinį sąryšį tarp kintamųjų, iš prieš tai gauto rezultato pasirinkta laikyti, kad betono stipris gali būti ir netiesinė savo mišinio sudedamųjų dalių funkcija.

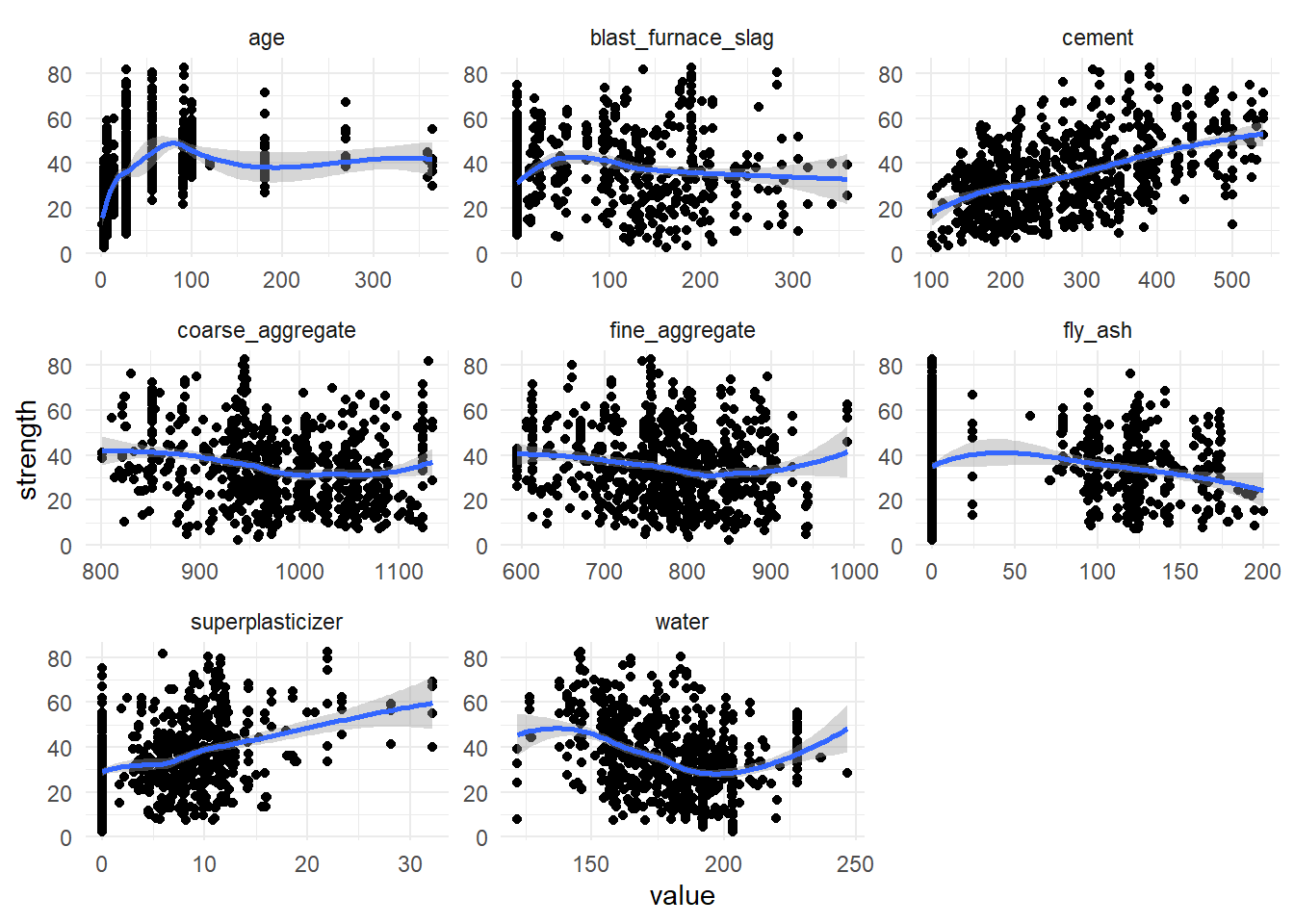
Panašus rezultatas dėl betono stiprio ir kovariančių netiesinio sąryšio gautas kiekvienai kovariantei ir atsakui nubraižius sklaidos diagramas kartu su lokalios regresijos kreive. Šie gauti rezultatai neatsižvelgia į kitų kovariančių reikšmes, todėl jie gali tik sufleruoti apie galimą sąryšį tarp kovariantės ir atsako pilname regresijos modelyje (kuriame atsižvelgiama į kitų kovariančių reikšmes).

# Duomenų šaltinis  
# https://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength  
# (https://www.kaggle.com/datasets/elikplim/concrete-compressive-strength-data-set)  
  
library(tidyverse)  
library(rsample)  
  
concrete <- read\_csv("concrete\_data.csv")  
  
concrete <- concrete %>%  
 unique() %>%  
 rename(strength = concrete\_compressive\_strength)  
  
# Padalijama į mokymo ir testavimo aibes  
set.seed(123)  
concrete\_split <- initial\_split(concrete)  
concrete\_train <- training(concrete\_split)  
concrete\_test <- testing(concrete\_split)

library(corrplot)  
  
correlation\_matrix <- concrete\_train %>%  
 cor()  
  
corrplot(correlation\_matrix, order = "original", method = "color", type = "upper", diag = FALSE, tl.col = "black", addCoef.col = "black")

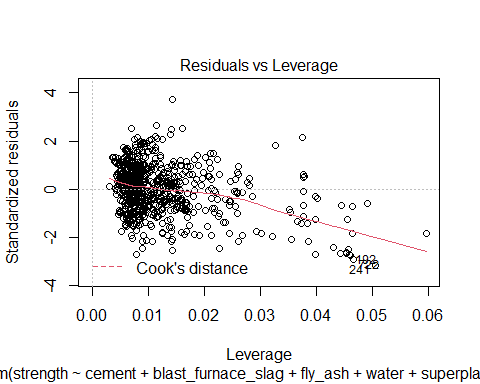
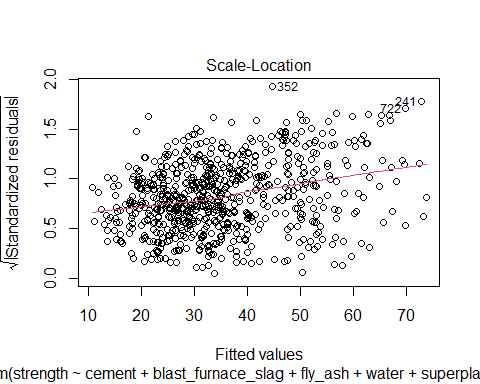
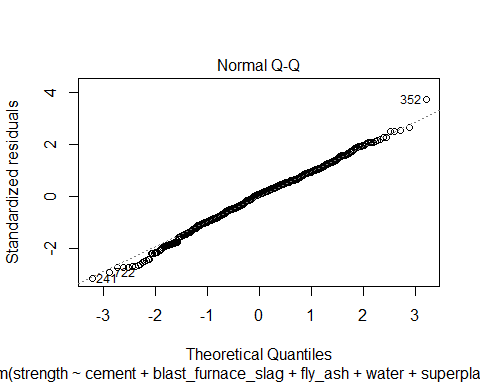
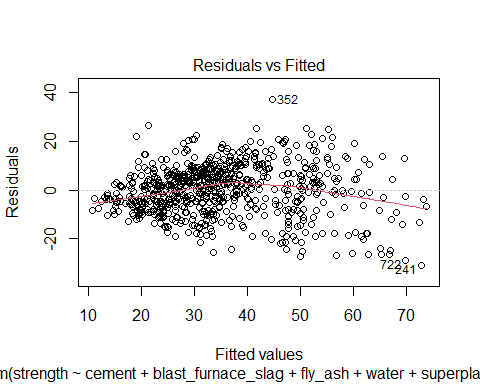


# tarp kovariančių ir atsako matome galimus netiesinis sąryšius  
concrete\_train %>%  
 pivot\_longer(-strength) %>%  
 ggplot(aes(value, strength)) +  
 geom\_point() +  
 facet\_wrap(vars(name), scales = "free") +  
 geom\_smooth() +  
 theme\_minimal()

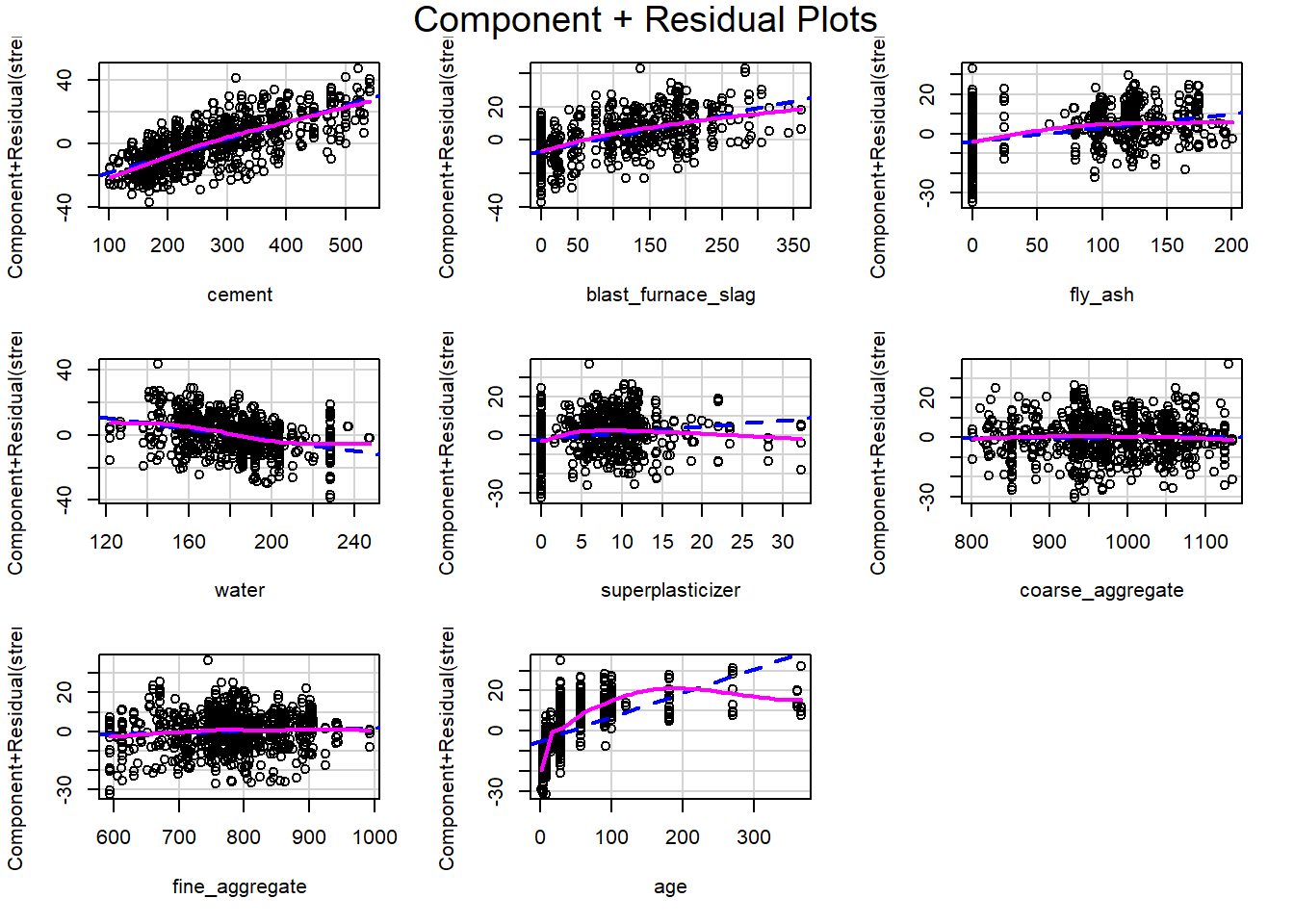


Pirmiausiai sudarytas paprastas tiesinės regresijos modelis. Analizuojant modelio diagnostinius grafikus rastas pakankamai stiprus liekanų heteroskedatiškumas. Pagal dalinių liekanų grafikus (partial residual plot / component residual plot) rasta, kad požymiuose “water” “superplastisizer” “age” turima papildoma informacija, į kurią paprastas tiesinis modelis nesugebėjo atsižvelgti.

# Pirma sudaromas regresijos modelis su vien tiesiniais  
 # kovariančių ir atsako sąryšiais (lyginamasis)  
library(car)  
  
baseline <- lm(strength ~ cement + blast\_furnace\_slag + fly\_ash +  
 water + superplasticizer  
 + coarse\_aggregate + fine\_aggregate + age, data = concrete\_train)  
  
  
plot(baseline)

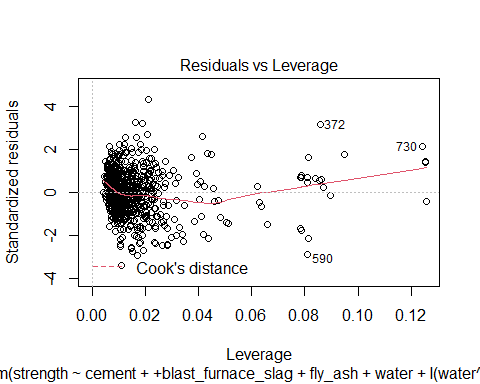
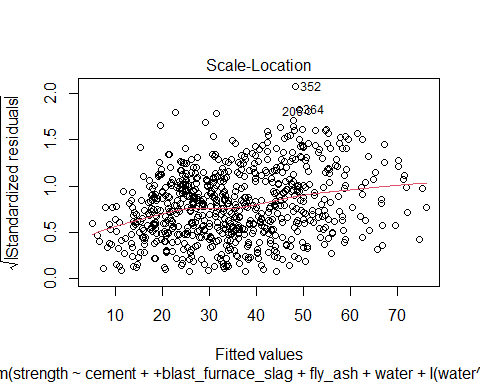
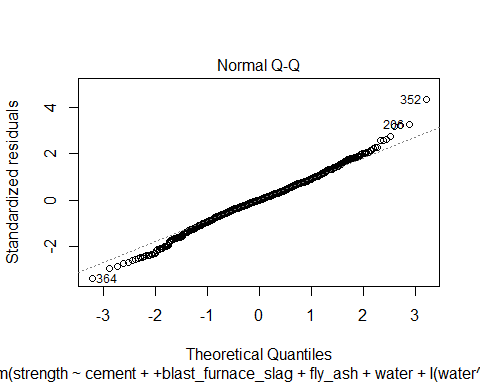
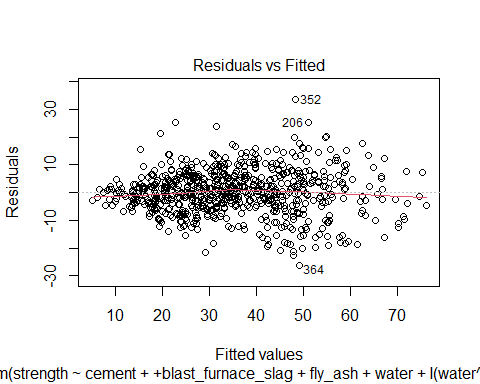


crPlots(baseline)



Siekiant patobulinti gautus rezultatus, sudarytas kitas regresijos modelis, papildomai į modelį įtraukiant prieš tai paminėtų kovariančių antruosius laipsnius (polinominės regresijos modelis). Naudojantis tomis pačiomis diagnostinėmis priemonėmis, laikyta, kad prieš tai rasti modelio nukrypimai nuo teorinio sumažėjo. Naudojant hierarchinių modelių (nested models) palyginimo testą, gautas statistiškai reikšmingas antro modelio skirtumas nuo sudaryto prieš tai (p<0.001).

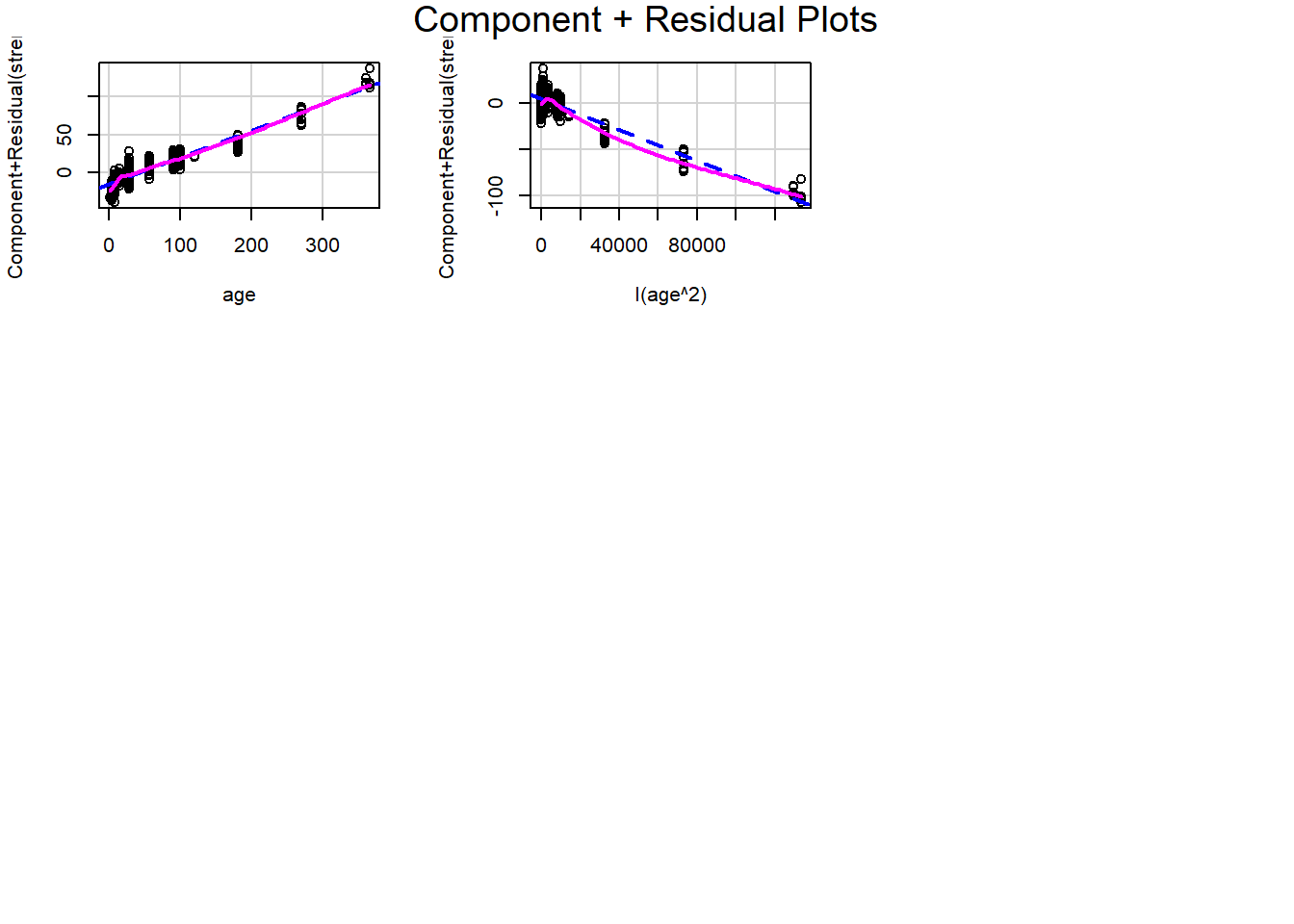
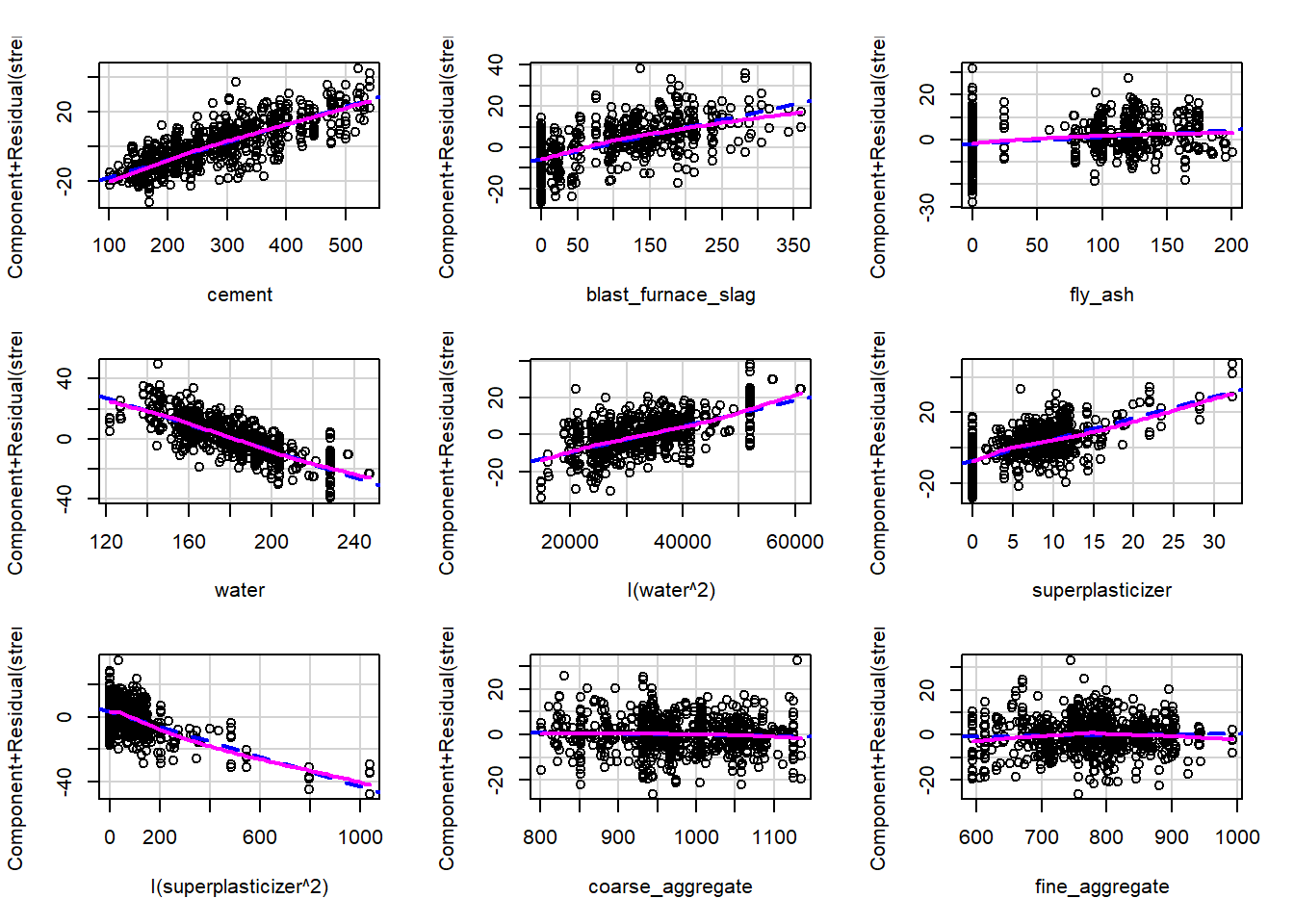
# "naivus" metodas pagerinti modelį pridedant aukštesnius kovariančių laipsnius  
model\_polynomial <- lm(strength ~ cement + blast\_furnace\_slag  
 + fly\_ash + water + I(water^2)  
 + superplasticizer + I(superplasticizer^2)  
 + coarse\_aggregate + fine\_aggregate + age + I(age^2),  
data = concrete\_train  
)  
  
plot(model\_polynomial)



summary(model\_polynomial)

##   
## Call:  
## lm(formula = strength ~ cement + +blast\_furnace\_slag + fly\_ash +   
## water + I(water^2) + superplasticizer + I(superplasticizer^2) +   
## coarse\_aggregate + fine\_aggregate + age + I(age^2), data = concrete\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -26.172 -4.597 0.047 4.843 33.480   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.325e+01 2.519e+01 1.717 0.08639 .   
## cement 1.025e-01 7.386e-03 13.870 < 2e-16 \*\*\*  
## blast\_furnace\_slag 7.631e-02 8.966e-03 8.511 < 2e-16 \*\*\*  
## fly\_ash 3.166e-02 1.147e-02 2.761 0.00591 \*\*   
## water -4.414e-01 1.716e-01 -2.571 0.01032 \*   
## I(water^2) 6.992e-04 4.763e-04 1.468 0.14259   
## superplasticizer 1.209e+00 1.494e-01 8.095 2.35e-15 \*\*\*  
## I(superplasticizer^2) -4.655e-02 5.935e-03 -7.843 1.53e-14 \*\*\*  
## coarse\_aggregate -5.052e-03 8.459e-03 -0.597 0.55051   
## fine\_aggregate 3.445e-03 9.486e-03 0.363 0.71660   
## age 3.527e-01 1.255e-02 28.095 < 2e-16 \*\*\*  
## I(age^2) -8.310e-04 4.081e-05 -20.364 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.825 on 741 degrees of freedom  
## Multiple R-squared: 0.7635, Adjusted R-squared: 0.76   
## F-statistic: 217.5 on 11 and 741 DF, p-value: < 2.2e-16

crPlots(model\_polynomial)



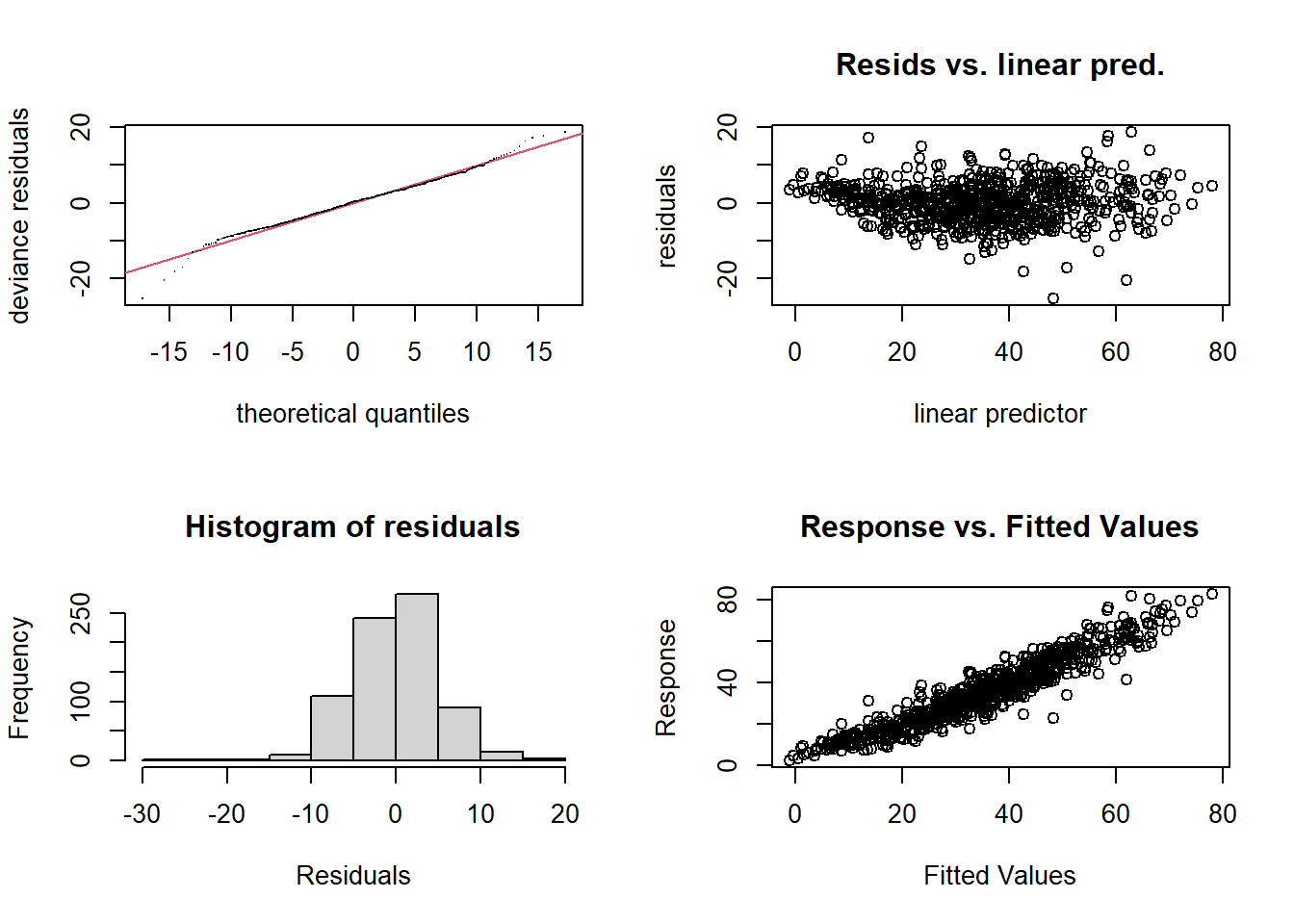
# modelis reikšmingai skiriasi nuo modelio be aukštesnio laipsnio narių  
anova(baseline, model\_polynomial)

## Analysis of Variance Table  
##   
## Model 1: strength ~ cement + blast\_furnace\_slag + fly\_ash + water + superplasticizer +   
## coarse\_aggregate + fine\_aggregate + age  
## Model 2: strength ~ cement + +blast\_furnace\_slag + fly\_ash + water + I(water^2) +   
## superplasticizer + I(superplasticizer^2) + coarse\_aggregate +   
## fine\_aggregate + age + I(age^2)  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 744 75286   
## 2 741 45374 3 29912 162.83 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Siekiant tiksliau modeliuoti betono stiprį negu polinominės regresijos modeliu pasirinkta sudaryti apibendrintą adityvų regresijos modelį, naudojant glodniaisius splainus. Šiam tikslui pasitelkta *mgcv* biblioteka. Naudojant *gam* funkciją iš šios bibliotekos, parametras *λ* parenkamas automatiškai, naudojant generalized cross validation. Imant numatytąjį mazgų skaičių kiekvienai kovariantei, diagnostinuose grafikuose stiprių nukrypimų nuo modelio prielaidų nerasta, tačiau tiek statistiniais testais, tiek naudojantis modelio kovariančių efektų grafikais pastebėta, kad pradinis mazgų skaičius gali būti per mažas tinkamai įvertinti sąryšius tarp kovariančių ir atsako (stipriausiai tai matoma su kovariante „coarse\_aggregate“)

c: 29.5 on 10 and 742 DF, p-value: < 2.2e-16

library(mgcv)  
library(gratia)  
# Alternatyvus modelis: apibendrintas adityvus modelis su glodniaisiais splainais  
  
# lambda parametras parenkamas automatiškai pagal generalized cross-validation  
model\_gam <- gam(strength ~ s(cement) + s(blast\_furnace\_slag) + s(fly\_ash)  
 + s(water) + s(superplasticizer)  
 + s(coarse\_aggregate) + s(fine\_aggregate) + s(age),  
data = concrete\_train,  
select = TRUE  
)  
  
gam.check(model\_gam)

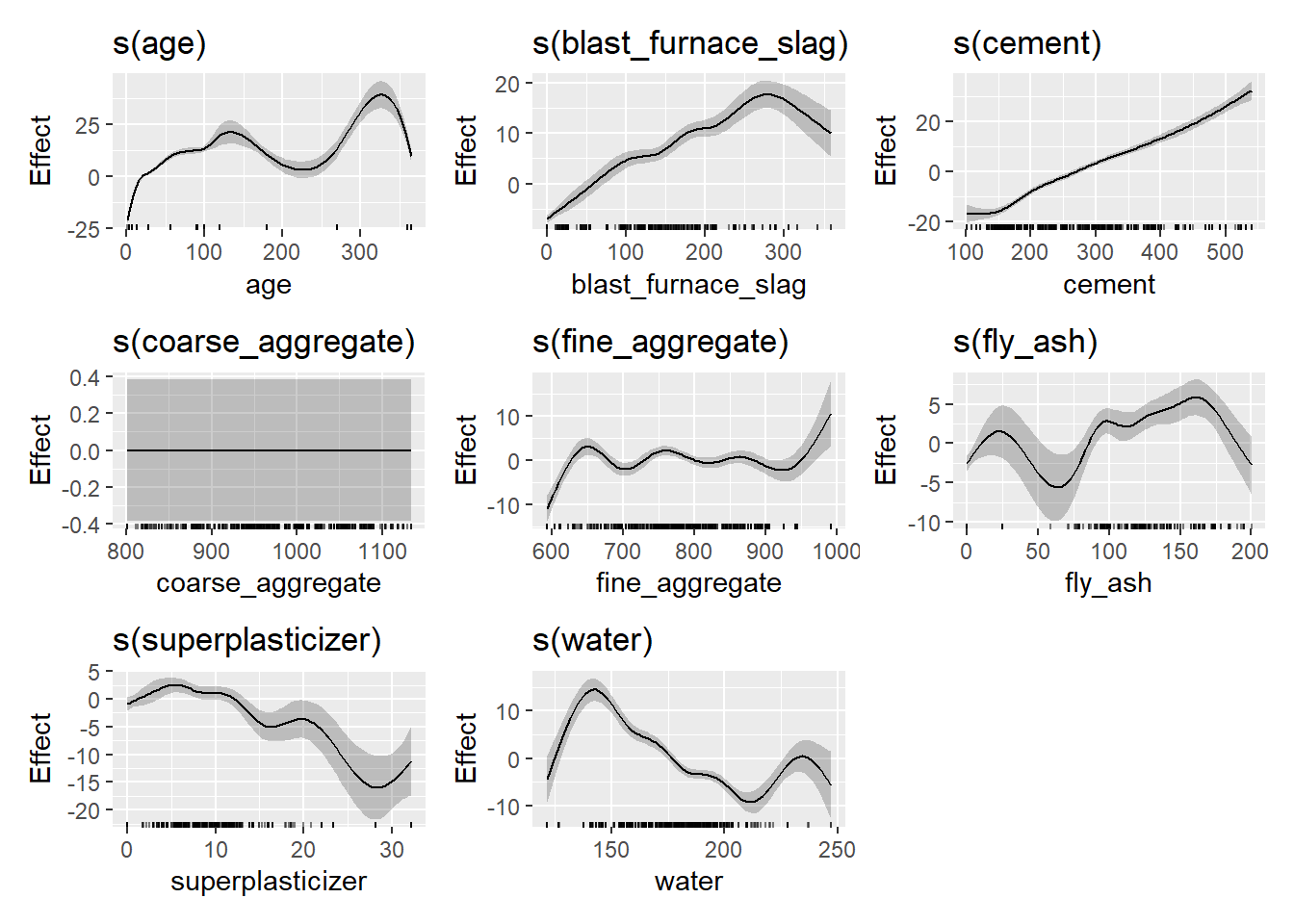


##   
## Method: GCV Optimizer: magic  
## Smoothing parameter selection converged after 128 iterations.  
## The RMS GCV score gradient at convergence was 0.0003962093 .  
## The Hessian was not positive definite.  
## Model rank = 73 / 73   
##   
## Basis dimension (k) checking results. Low p-value (k-index<1) may  
## indicate that k is too low, especially if edf is close to k'.  
##   
## k' edf k-index p-value   
## s(cement) 9.00e+00 6.85e+00 0.91 0.020 \*   
## s(blast\_furnace\_slag) 9.00e+00 6.72e+00 0.96 0.085 .   
## s(fly\_ash) 9.00e+00 7.16e+00 1.01 0.655   
## s(water) 9.00e+00 8.38e+00 0.94 0.005 \*\*   
## s(superplasticizer) 9.00e+00 7.85e+00 0.93 0.030 \*   
## s(coarse\_aggregate) 9.00e+00 2.33e-07 0.86 <2e-16 \*\*\*  
## s(fine\_aggregate) 9.00e+00 8.51e+00 0.87 <2e-16 \*\*\*  
## s(age) 9.00e+00 7.99e+00 1.06 0.950   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(model\_gam)

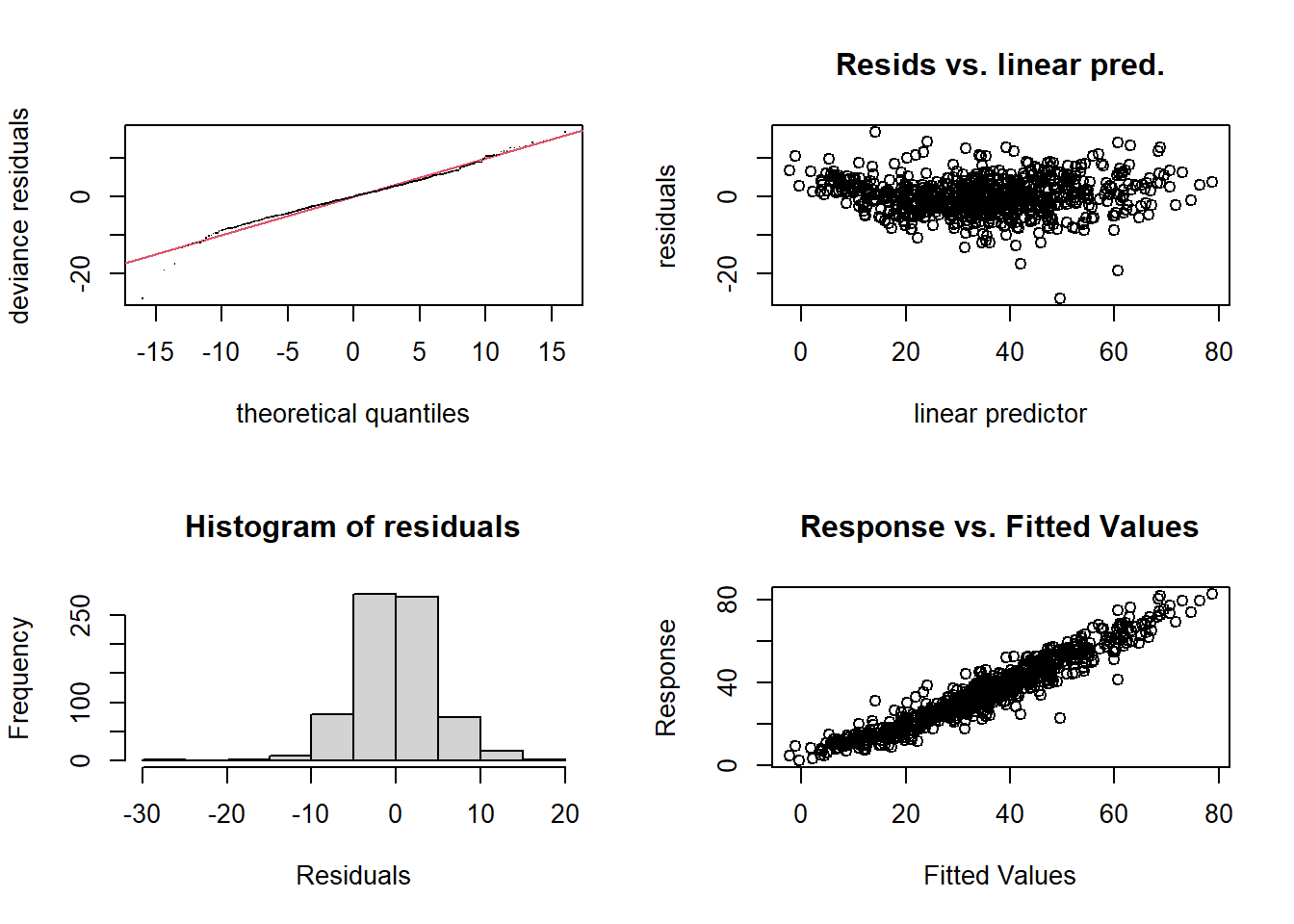
##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## strength ~ s(cement) + s(blast\_furnace\_slag) + s(fly\_ash) + s(water) +   
## s(superplasticizer) + s(coarse\_aggregate) + s(fine\_aggregate) +   
## s(age)  
##   
## Parametric coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 34.7507 0.1955 177.7 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Approximate significance of smooth terms:  
## edf Ref.df F p-value   
## s(cement) 6.853e+00 9 100.841 <2e-16 \*\*\*  
## s(blast\_furnace\_slag) 6.723e+00 9 50.595 <2e-16 \*\*\*  
## s(fly\_ash) 7.162e+00 9 6.798 <2e-16 \*\*\*  
## s(water) 8.383e+00 9 29.612 <2e-16 \*\*\*  
## s(superplasticizer) 7.851e+00 9 7.709 <2e-16 \*\*\*  
## s(coarse\_aggregate) 2.332e-07 9 0.000 0.958   
## s(fine\_aggregate) 8.514e+00 9 9.957 <2e-16 \*\*\*  
## s(age) 7.986e+00 9 266.120 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## R-sq.(adj) = 0.887 Deviance explained = 89.5%  
## GCV = 31.031 Scale est. = 28.787 n = 753

draw(model\_gam)



Padidinus mazgų skaičių ir naudojant visas prieš tai aprašytas diagnostikas nerasta stiprių nukrypimų nuo teorinio modelio. Gautame modelyje visos kovariantės reikšmingos, taip pat gautas statistiškai reikšmingas skirtumas nuo modelio, naudojančio mažesnį (numatytąjį) mazgų skaičių (p<0.001). Siekiant palyginti prieš tai sudarytą daug paprastenį polinominės regresijos ir gautą glodniųjų splainų modelį, abiems iš jų nubraižyti kovariančių efektų grafikai. Matoma, kad apibendrinto adityvaus modelio atveju gauti daug sudėtingesni sąryšiai tarp kovariančių ir atsako.

# Padidinamas mazgų skaičius  
model\_gam2 <- gam(strength ~ s(cement) + s(blast\_furnace\_slag, k = 20) + s(fly\_ash, k = 20)  
 + s(water, k = 20) + s(superplasticizer, k = 20)  
 + s(coarse\_aggregate, k = 20) + s(fine\_aggregate, k = 20) + s(age, k = 10),  
data = concrete\_train,  
select = TRUE  
)  
  
gam.check(model\_gam2)

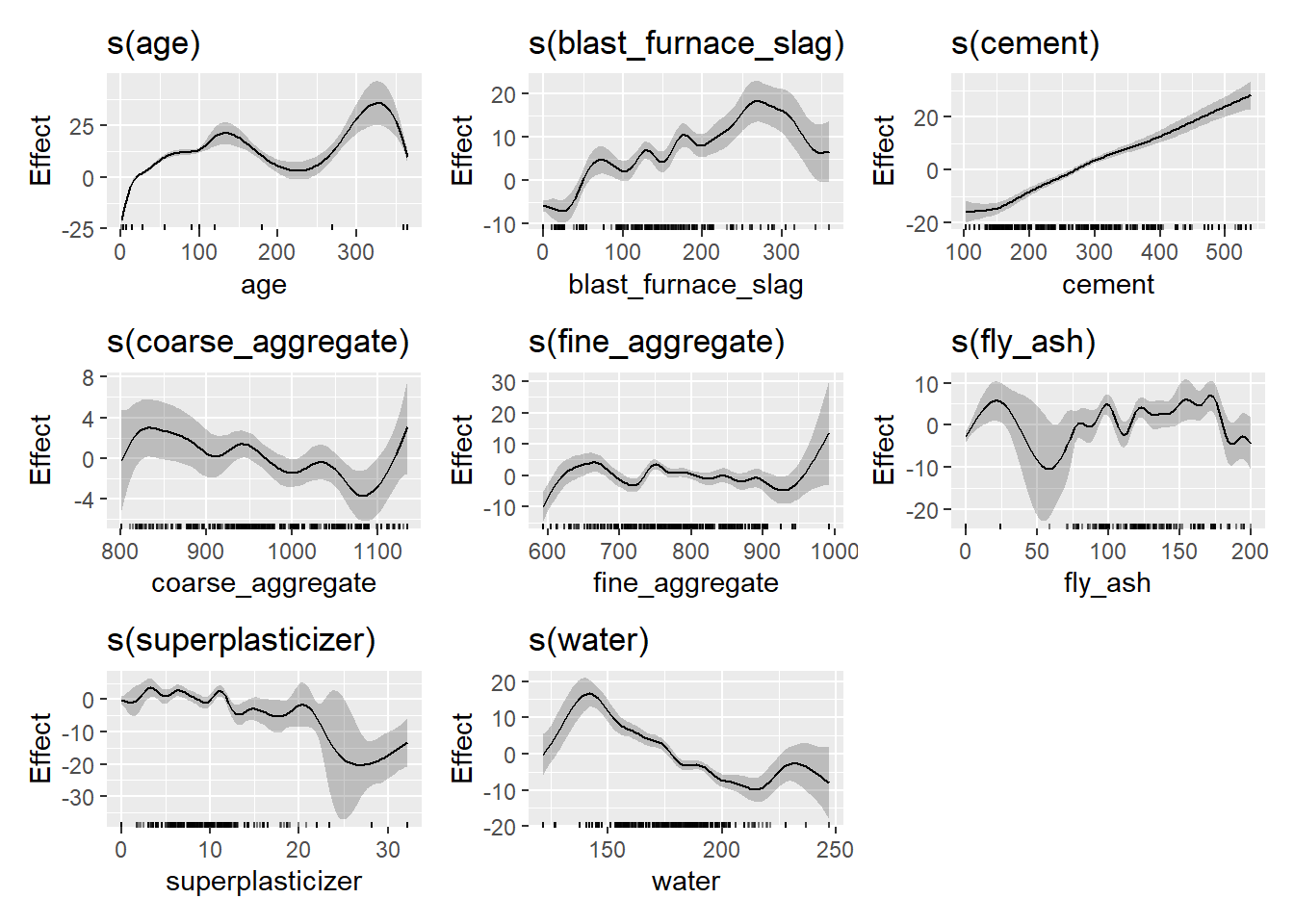


##   
## Method: GCV Optimizer: magic  
## Smoothing parameter selection converged after 82 iterations.  
## The RMS GCV score gradient at convergence was 0.0004353583 .  
## The Hessian was not positive definite.  
## Model rank = 133 / 133   
##   
## Basis dimension (k) checking results. Low p-value (k-index<1) may  
## indicate that k is too low, especially if edf is close to k'.  
##   
## k' edf k-index p-value   
## s(cement) 9.00 6.19 0.99 0.375   
## s(blast\_furnace\_slag) 19.00 13.96 1.03 0.845   
## s(fly\_ash) 19.00 15.09 1.05 0.940   
## s(water) 19.00 12.16 1.01 0.635   
## s(superplasticizer) 19.00 16.00 0.97 0.220   
## s(coarse\_aggregate) 19.00 9.28 0.96 0.105   
## s(fine\_aggregate) 19.00 15.01 0.95 0.055 .  
## s(age) 9.00 8.18 1.07 0.990   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(model\_gam2)

##   
## Family: gaussian   
## Link function: identity   
##   
## Formula:  
## strength ~ s(cement) + s(blast\_furnace\_slag, k = 20) + s(fly\_ash,   
## k = 20) + s(water, k = 20) + s(superplasticizer, k = 20) +   
## s(coarse\_aggregate, k = 20) + s(fine\_aggregate, k = 20) +   
## s(age, k = 10)  
##   
## Parametric coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 34.7507 0.1821 190.9 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Approximate significance of smooth terms:  
## edf Ref.df F p-value   
## s(cement) 6.188 9 23.149 < 2e-16 \*\*\*  
## s(blast\_furnace\_slag) 13.959 19 7.039 < 2e-16 \*\*\*  
## s(fly\_ash) 15.091 19 3.670 < 2e-16 \*\*\*  
## s(water) 12.163 19 8.275 < 2e-16 \*\*\*  
## s(superplasticizer) 16.005 19 3.923 < 2e-16 \*\*\*  
## s(coarse\_aggregate) 9.276 19 1.151 0.00268 \*\*   
## s(fine\_aggregate) 15.014 19 5.826 < 2e-16 \*\*\*  
## s(age) 8.177 9 298.256 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## R-sq.(adj) = 0.902 Deviance explained = 91.5%  
## GCV = 28.651 Scale est. = 24.965 n = 753

draw(model\_gam2)



# modelis statistiškai reikšmingai skiriasi nuo modelio su mažesniu mazgų skaičiumi  
anova(model\_gam, model\_gam2,test="F")

## Analysis of Deviance Table  
##   
## Model 1: strength ~ s(cement) + s(blast\_furnace\_slag) + s(fly\_ash) + s(water) +   
## s(superplasticizer) + s(coarse\_aggregate) + s(fine\_aggregate) +   
## s(age)  
## Model 2: strength ~ s(cement) + s(blast\_furnace\_slag, k = 20) + s(fly\_ash,   
## k = 20) + s(water, k = 20) + s(superplasticizer, k = 20) +   
## s(coarse\_aggregate, k = 20) + s(fine\_aggregate, k = 20) +   
## s(age, k = 10)  
## Resid. Df Resid. Dev Df Deviance F Pr(>F)   
## 1 694.30 20108   
## 2 643.57 16380 50.732 3728.2 2.9437 3.48e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

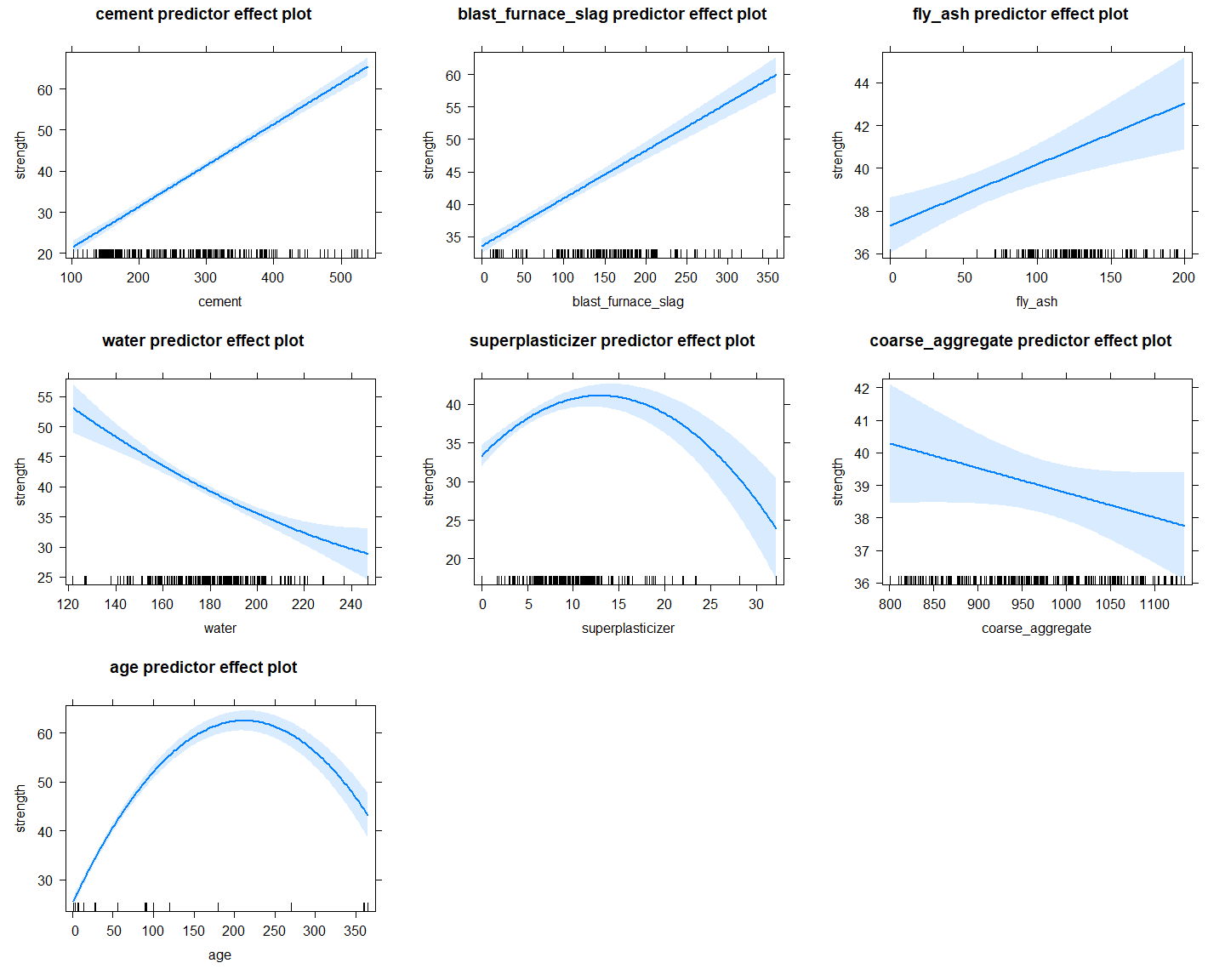
# modelio palyginimas su prieš tai sudarytu polinominės regresijos modeliu  
library(effects)  
AIC(model\_polynomial)

## [1] 5247.325

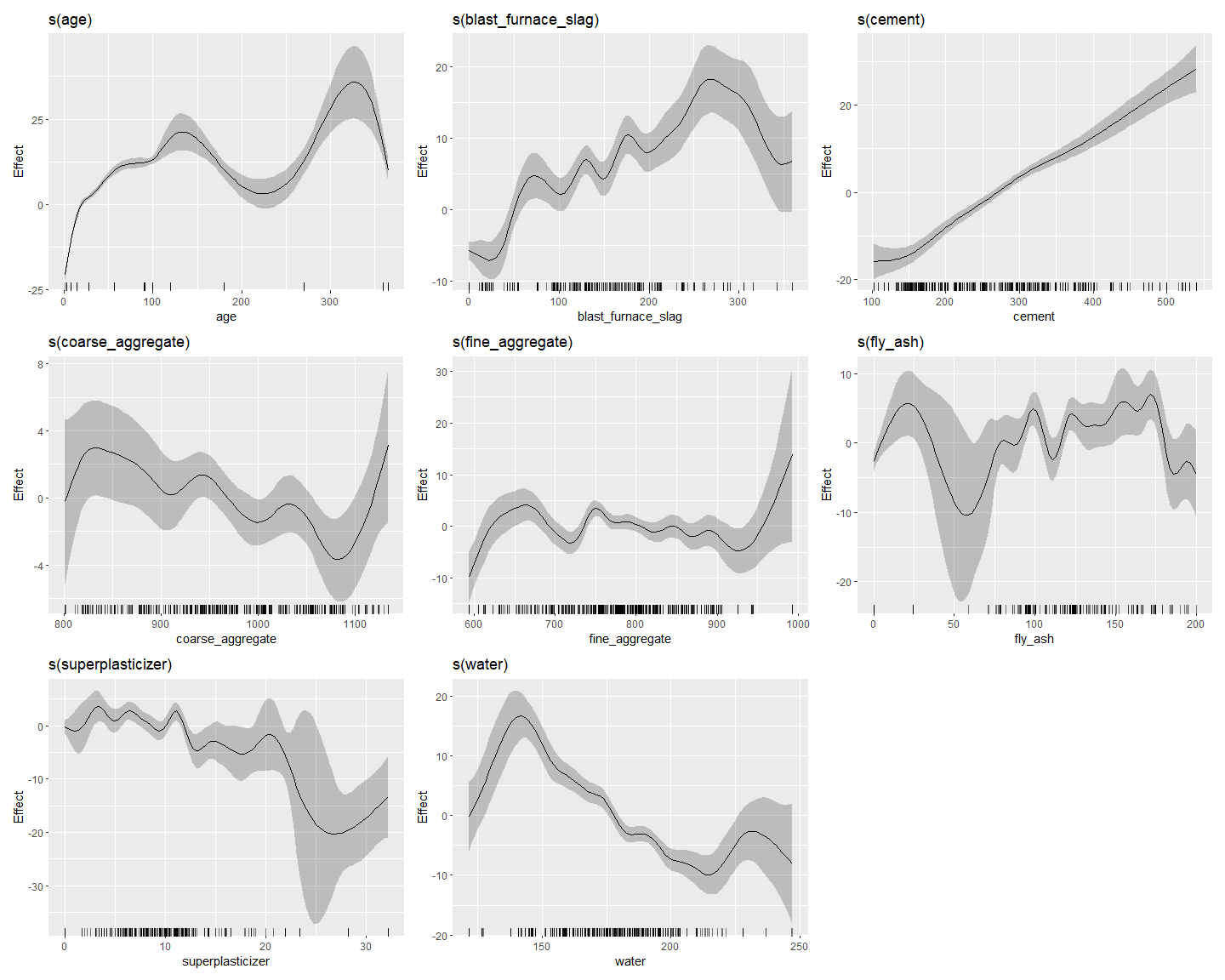
AIC(model\_gam2)

## [1] 4651.725

plot(predictorEffects(model\_polynomial))



draw(model\_gam2)



Kaip pastebėta naudojantis praėjusiais grafikais, glodniųjų splainų modeliu gauti daug sudėtingesni sąryšiai negu gauti naudojantis polinominės regresijos modeliu. Nors sudarytame modelyje naudoti glodnieji splainai, kurie dėl glodinimo baudos apsaugo nuo per didelio modelio prisitaikymo mokymo duomenims, pasirinkta panaudoti testavimo aibę patikrinti, ar modelis tikrai nepersimokė ir tuo pačiu metu įvertinti visų sudarytų modelių prognozavimo kokybę. Gauta, kad tiek pagal vidutinę absoliučią paklaidą (Mean Absolute Error), tiek matuojant vidutinę kvadratinę paklaidą ( (Root) Mean Square Error), glodniųjų splainų modeliu gauti žymiai geresni rezultatai už prieš tai sudarytus modelius (paklaidos daugiau nei dvigubai mažesnės lyginant su paprastu tiesiniu modeliu, apie 50% mažesnės lyginant su polinominiu modeliu). Grafiškai modelių prognozės palygintos naudojant prognozuotų ir tikrų reikšmių sklaidos diagramas. Matoma, kad glodniųjų splainų modeliu gauta daug mažesnė sklaida aplink prognozuotos ir tikros reikšmės lygybės tiesę lyginant su kitais modeliais.

# Nors naudoti glodnieji splainai, naudojant testavimo aibę patikrinama ar kažkur nebuvo padaryta klaidų ir  
# modelis tikrai nepersimokė  
library(yardstick)  
  
concrete\_test <- concrete\_test %>%  
 mutate(predicted\_baseline = predict(baseline, concrete\_test),  
 predicted\_polynomial = predict(model\_polynomial,concrete\_test),  
 predicted\_gam = predict(model\_gam2,concrete\_test))  
  
set <- metric\_set(rmse,mae)  
set(concrete\_test, strength, predicted\_baseline)

## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 11.1   
## 2 mae standard 8.79

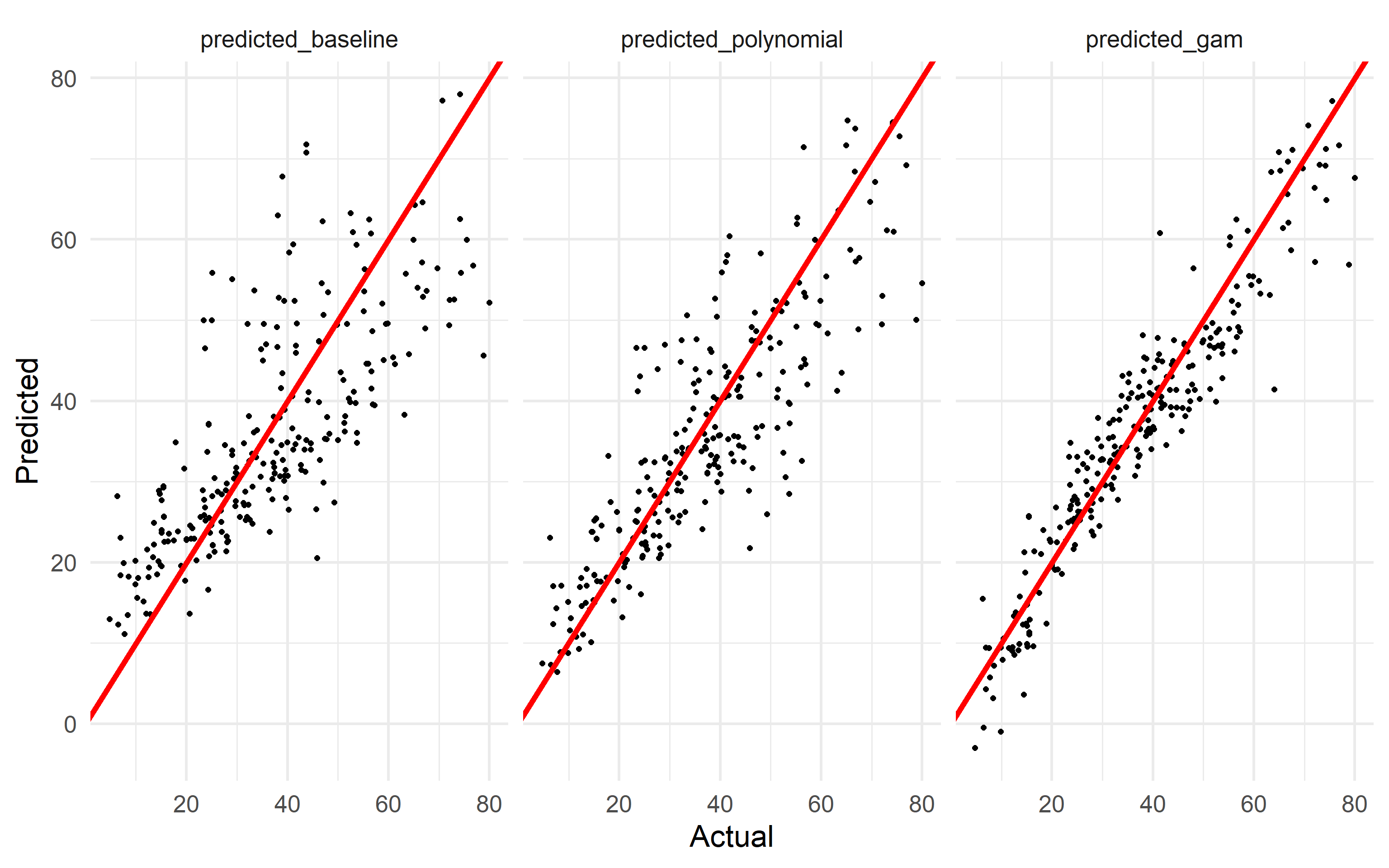
set(concrete\_test, strength, predicted\_polynomial)

## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 9.08  
## 2 mae standard 6.76

set(concrete\_test, strength, predicted\_gam)

## # A tibble: 2 x 3  
## .metric .estimator .estimate  
## <chr> <chr> <dbl>  
## 1 rmse standard 5.47  
## 2 mae standard 4.28

# tiek pagal MAE, tiek pagal RMSE naudojant gam modelį gaunami geresni rezultatai  
  
  
# Modelių prognozių palyginimas  
concrete\_test %>%  
 pivot\_longer(c(predicted\_gam, predicted\_polynomial,predicted\_baseline)) %>%  
 ggplot(aes(strength, value)) +  
 geom\_point() +  
 facet\_wrap(vars(name)) +  
 geom\_abline(color = "red", size = 1.25) +  
 labs(x = "Actual", y = "Predicted") +  
 theme\_minimal()



Išvados:

Atlikus pirminę duomenų aibės analizę rasta, kad betono stipris gali netiesiškai priklausyti nuo jo mišinį sudarančių medžiagų, todėl laikyta, kad paprastas tiesinės regresijos modelis gali būti netinkamas prognozuoti betono stiprį.

Siekiant turėti palyginamąjį modelį sudarytas paprastas tiesinės regresijos modelis. Kaip vienas iš būdų pagerinti modelį buvo pasirinkta polinominė regresija. Į modelį buvo įtraukti antrieji laipsniai tų kintamųjų, kuriuose pagal tiesinės regresijos diagnostinius grafikus rasta neišnaudotos informacijos. Pasitelkiant grafines modelių diagnostikas ir statistinius testus laikyta, kad antru modeliu gaunami geresni rezultatai.

Siekiant dar labiau pagerinti rezultatus (ir parodyti, kad aukštesnių laipsnių įtraukimas į paprasta tiesinį modelį nėra pakankamas tikslioms prognozėms gauti) sudarytas apibendrintas adityvus regresijos modelis, naudojantis glodniuosius splainus. Pirmiausia naudotas numatytasis mazgų skaičius, tačiau naudojant modelio diagnostikas pasirinkta šį skaičių padidinti. Parinkus didesnį mazgų skaičių modelio diagnostikose beveik nerasta nukrypimų. Visos modelyje naudotos kovariantės modelyje statistiškai reikšmingos.

Testavimo aibė panaudota įvertinti visų šių modelių gebėjimą prognozuoti reikšmes. Kaip ir tikėtasi, geriausi rezultatai gauti naudojant glodniųjų splainų modelį (paklaidos pagal dvi skirtingas matavimo metrikas daugiau nei dvigubai mažesnės už tiesinį ir apie 50% mažesnės už polinominį modelį). Dėl šios priežasties daroma išvada, kad sudarytas glodniųjų splainų apibendrintas adityvus tinkamas prognozuoti betono stiprį, be to šiuo modeliu gaunami ryškiai geresni rezultatai negu naudojant paprastesnį metodą netiesiniams sąryšiams tarp atsako ir kovariančių modeliuoti, kuriuo paprastas tiesinės regresijos modelis papildomas kovariančių aukštesniais laipsniais.