

NLNP Praktikum 4

Robin Baudisch, Merlin Kopfmann, Maximilian Neudert

Inhaltsverzeichnis

A1	2
A2	3
a)	3
b)	8

A1

AIC (Linear):

AIC (n=100)

[1] 297.6641

AIC (n=20)

[1] 61.74052

AIC (Polynomial):

AIC (n=100)

[1] 290.4355

AIC (n=20)

[1] 62.19022

Quadratischer Vorhersagefehler (Linear):

Quadratischer Vorhersagefehler (n=100)

[1] 1.139113

Quadratischer Vorhersagefehler (n=20)

[1] 0.9740106

Quadratischer Vorhersagefehler (Polynomial):

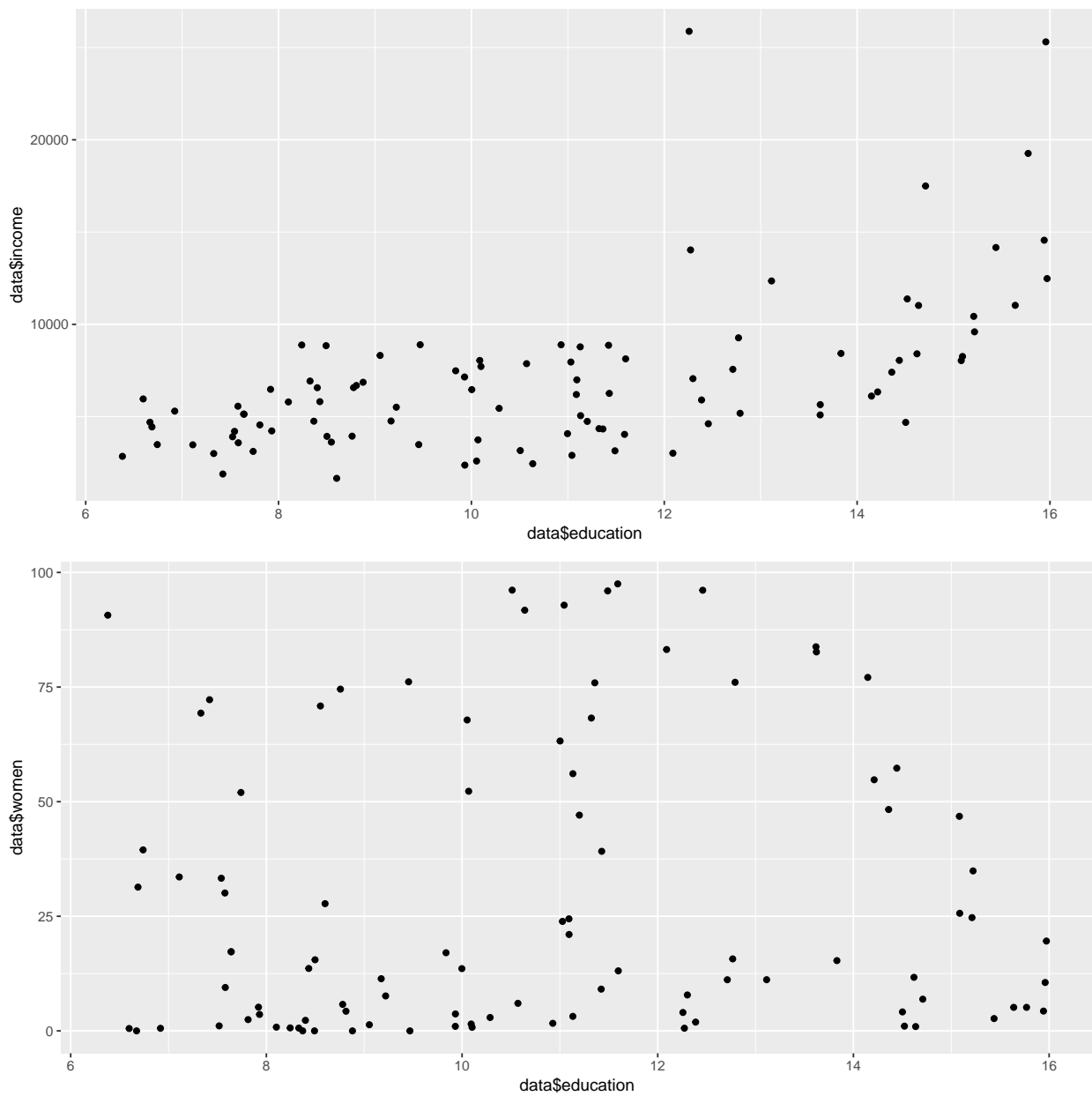
Quadratischer Vorhersagefehler (n=100)

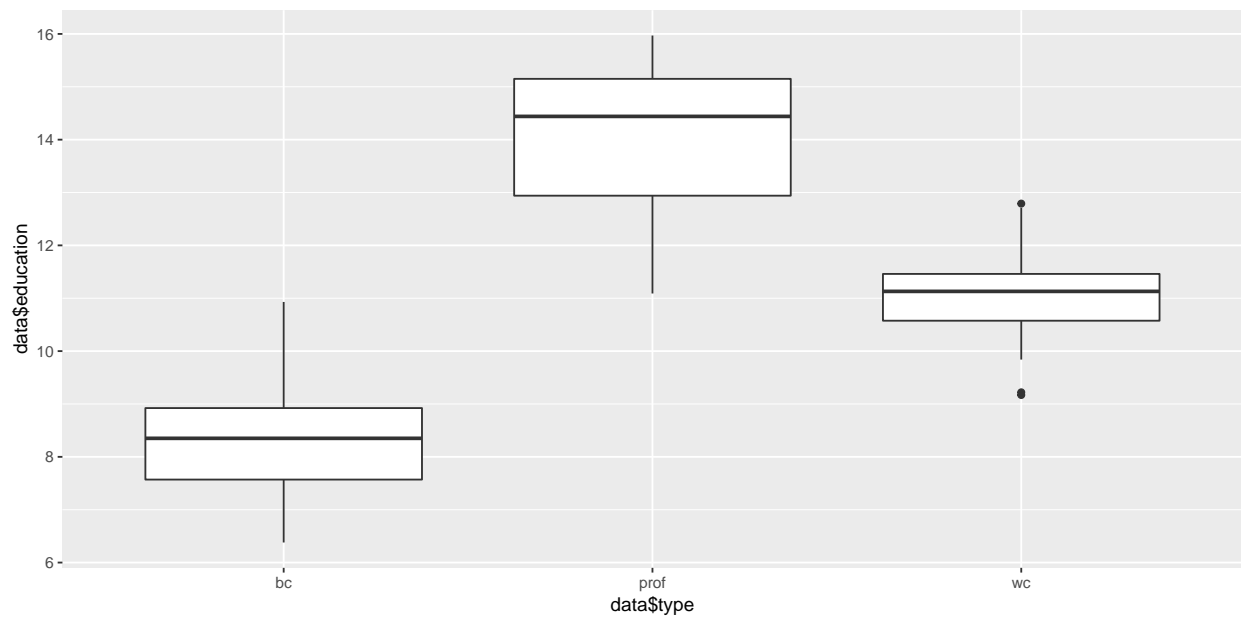
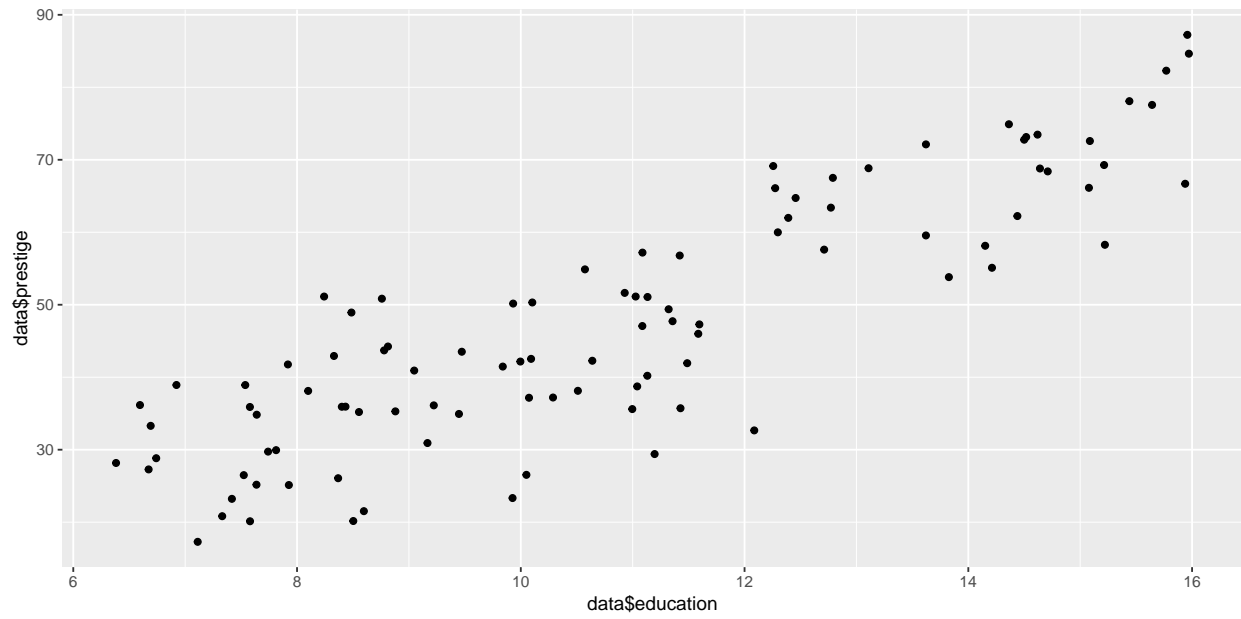
[1] 1.049011

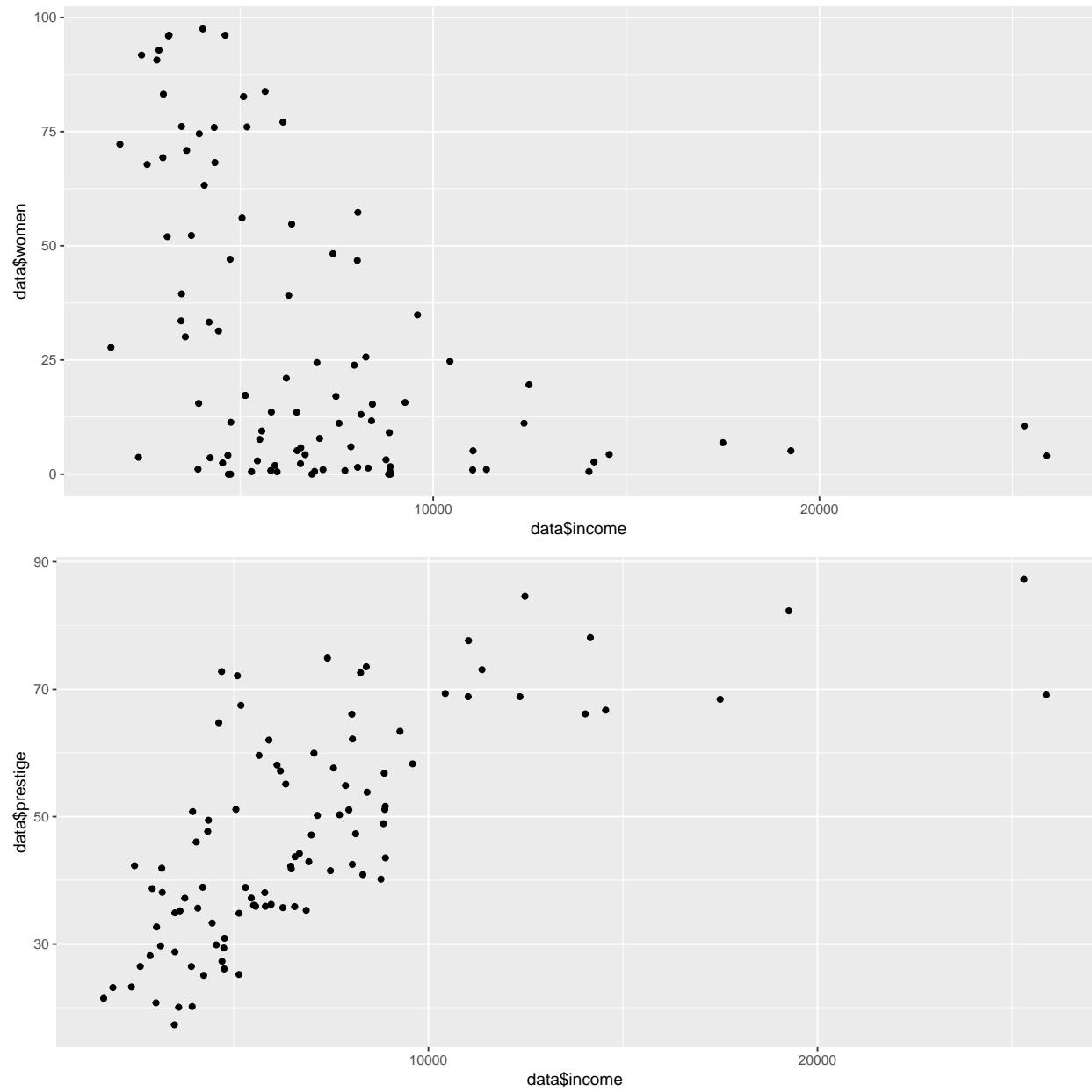
Quadratischer Vorhersagefehler (n=20)

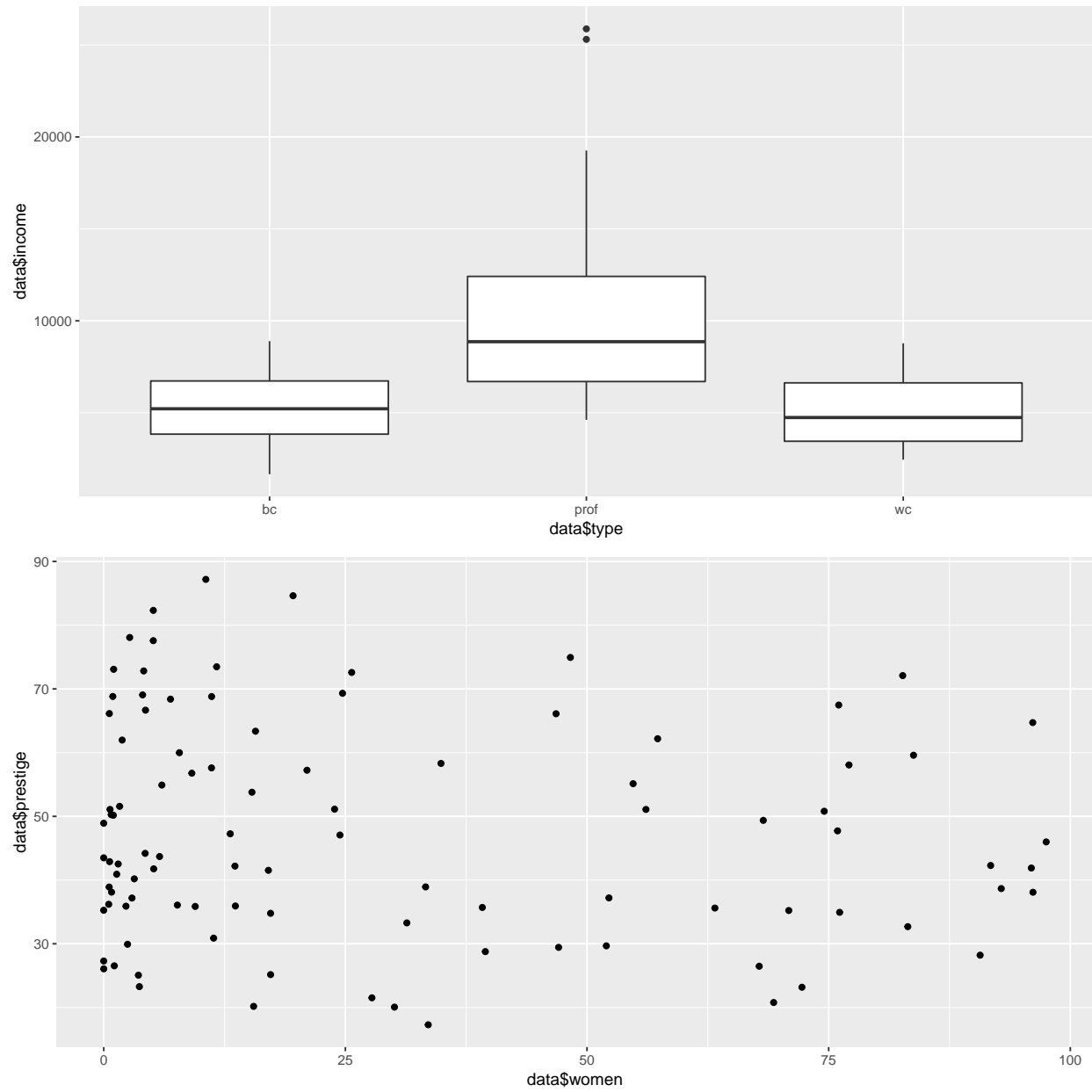
[1] 0.9853232

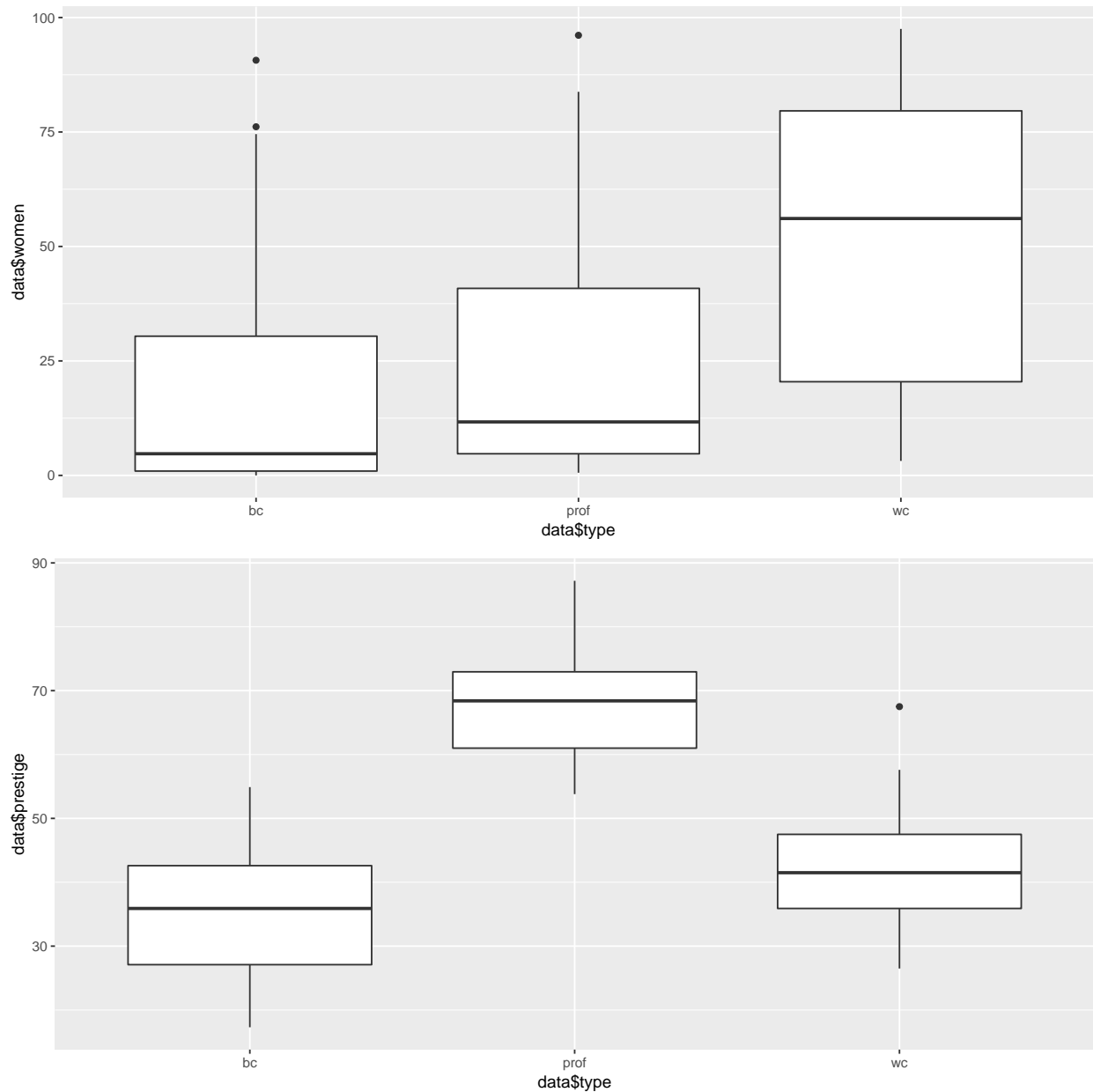
Der durchschnittliche AIC im Falle $n = 20$ ist für das lineare Modell mit $AIC = 61.74$ kleiner als im polynomialen Modell mit $AIC = 62.19$. Im Falle $n = 100$ ist das polynomiale Modell laut dem AIC besser ($AIC = 290.44$) als das lineare Modell ($AIC = 297.66$). Dies stimmt mit den Ergebnissen der geschätzten quadratischen Vorhersagefehlern überein. Bei $n = 20$ ist das lineare Modell leicht besser, im Falle $n = 100$ produziert das polynomiale Modell die genaueren Vorhersagen.

A2**a)**









```

model_list <- list(model1, model2, model3, model4, model5, model6, model7, model8,
  model9)
for (model in 1:length(model_list)) {
  p <- length(model_list[[model]]$coefficients)
  results[model, 1] <- length(model_list[[model]]$coefficients)
  results[model, 2] <- train_error(n_var, model_list[[model]])
  results[model, 3] <- opt_term(n_var, p, model_list[[model]])
  results[model, 4] <- test_error_exp(results[model, 3], results[model, 2])
  results[model, 5] <- AIC(model_list[[model]])
}

results

```

	N_Param	Trainingsfehler	Optimismusterm	Erwarteter_Testfehler
model1	2	146.17611	6.027881	152.20399
model2	2	72.08576	2.972609	75.05837
model3	2	262.83293	10.838471	273.67140
model4	2	285.74478	11.783290	297.52807
model5	3	53.80040	3.327860	57.12826
model6	3	62.76835	3.882578	66.65093
model7	3	64.29518	3.977021	68.27220
model8	4	50.66315	4.178404	54.84155
model9	5	50.59325	5.215799	55.80905
AIC				
model1	772.6235			
model2	703.3419			
model3	830.1208			
model4	838.3117			
model5	676.6695			
model6	691.7781			
model7	694.1334			
model8	672.7814			
model9	674.6461			

- Länge der Ausbildung scheint den größten Einfluss auf die abhängige Variable zu haben, da dass Modell 4 das beste Modell mit nur einer unabhängigen Variable ist.
- Modell 5 bietet aufbauend auf der Variablen Länge der Ausbildung die höchste Genauigkeit, durch Hinzunahme der Variable Einkommen. Frauenanteil (Modell 7) und Berufsklasse (Modell 6) sind dagegen laut AIC schlechtere Modelle.
- Weitere Hinzunahme der unabhängigen Variablen Berufsklasse und Frauenanteil führt zu keiner Verbesserung der Modellgüte.
- Modell 5 scheint das robusteste Modell zu sein, in dem nur signifikante unabhängige Variablen vorkommen. Das Risiko für Overfitting ist klein.

b)

Um AIC-Werte miteinander vergleichen zu können, müssen die Modelle auf den gleichen Daten trainiert werden.