

Statistik 2 – Tutorate

Sitzung 8-10: Multivariate Regression

Lea Elina Hofer, Samuel Rauh, Sebastian Senn, Fynn Siefart, Marco Giesselmann

Lernziele dieser Sitzung



Multivariate Regression

Du kannst in R eine multivariate Regression durchführen



Erstellung von Regressionstabellen

Du kannst in R Regressionstabellen herstellen und anschaulich aufbereiten



Visualisierung von Regressionsergebnissen

Du kannst Ergebnisse von multivariaten Regressionen visualisieren

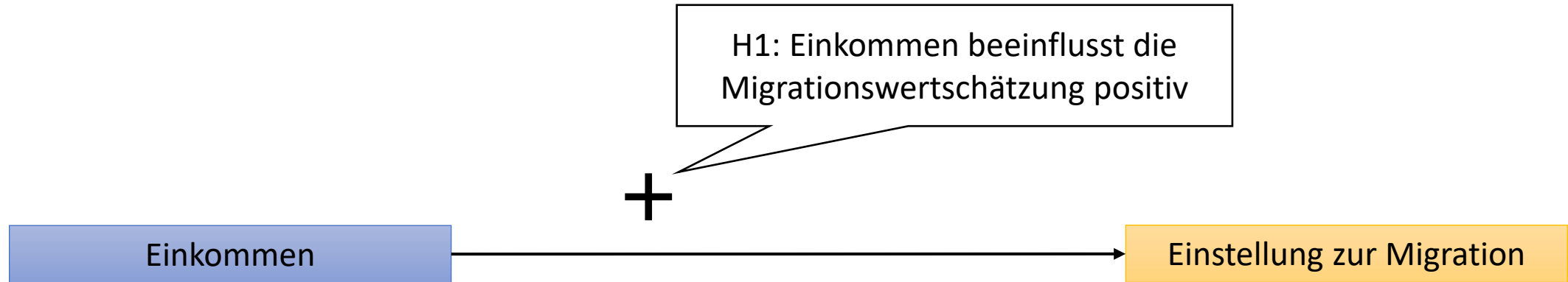
Was für einen Einfluss erwartet ihr? Formuliert eine entsprechende Hypothese

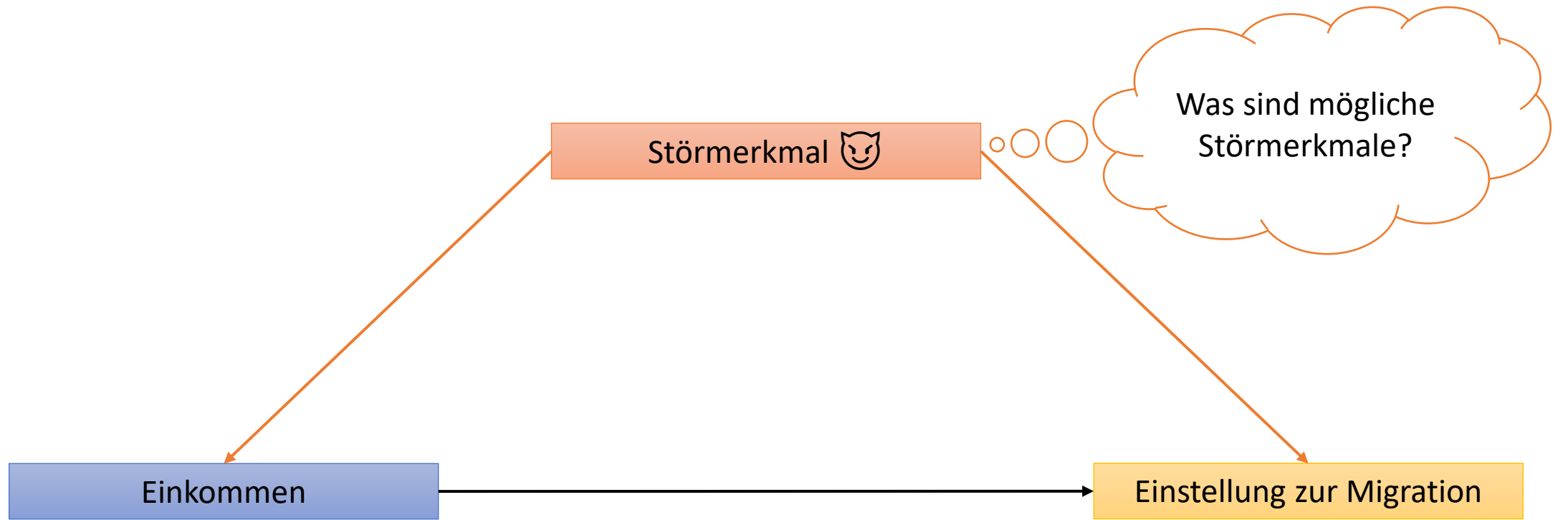
?

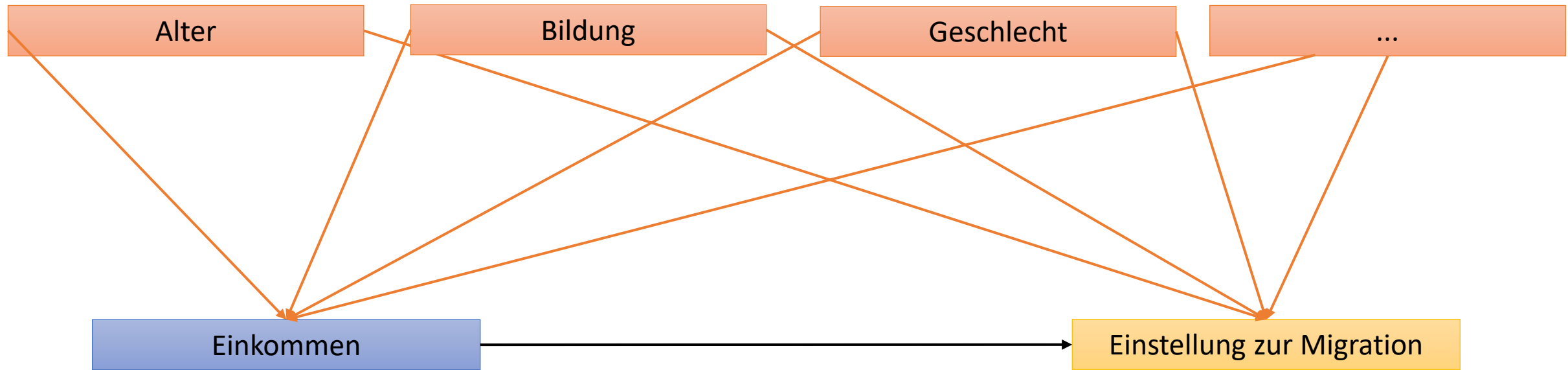
Einkommen



Einstellung zur Migration







Warum integrieren wir Störmerkmale in unsere Analyse?

Wir integrieren Störmerkmale in unsere Analyse, um den bereinigten Zusammenhang zwischen UV und AV zu ermitteln und damit eine gültigere Antwort auf die kausale Hypothese bieten zu können

Achtung 1: Ein Störmerkmal zeichnet aus, dass es einen gut begründbaren Einfluss sowohl auf die UV als auch die AV aufspannt. Nur wenn *beide* Bedingungen gegeben sind handelt es sich um ein Störmerkmal!

Achtung 2: Legitimiert wird die Drittvariablenintegration durch theoretische Begründungen und ggf. externe empirische Verweise - nicht durch die tatsächliche Auswirkung auf den Koeffizienten der UV in eurer empirischen Anwendung

1 Data Management – Inspektion und Selektion

Wir messen *Einkommen* und *Migrationswertschätzung* in unserer Hypothese mit den Variablen **hinctnta** und **imueclt**. Als potenzielle Störmerkmale haben wir die *Bildung* (**eduyrs**), das *Alter* (**agea**) und das *Geschlecht* (**gndr**) identifiziert. Zudem beschränken wir unsere Stichprobe auf die Schweiz. Wir verschaffen uns einen Überblick über alle Variablen und reduzieren unseren Datensatz entsprechend.

```
ess8_CH <- filter(ess8, cntry == "CH")
ess8_CH_ss <- select(ess8_CH, idno, imueclt,
                    hinctnta, eduyrs, agea, gndr)
```

```
look_for(ess8_CH, "hinctnta")
look_for(ess8_CH, "imueclt")
look_for(ess8_CH, "eduyrs")
look_for(ess8_CH, "agea")
look_for(ess8_CH, "gndr")
```

```
summary(ess8_CH_ss)
```

Schau dir an, wie im ESS das Einkommen mit der Variable **hinctnta** gemessen wird. Was fällt dir auf?

Das Einkommen wird im ESS in Dezilen gemessen. (1. Dezil=tiefstes Einkommen, 10. Dezil=höchstes Einkommen)

Informationen zu den genauen Dezilgrenzen findest du auf der ESS-Webseite

1 Data Management – Inspektion und Selektion

Wie ist die Variable **gndr** im ESS kodiert?

Die Variable **gndr** ist im ESS als Variable mit den Werten 1 für Männer und 2 für Frauen hinterlegt. So wird sie wie eine metrische Variable behandelt! Zusätzlich sind Antwortverweigerungen als «No answer» und nicht direkt als NA kodiert (kommen aber in unserer Teilstichprobe gar nicht vor).

gndr
Gender
1
2
1
2
1
1
2
2
1
1
1



gndr
Gender
Male
Female
Male
Female
Male
Male
Female
Female
Male
Male
Male

```
> look_for(ess8_CH, "gndr")
pos variable label col_type missing values
254 gndr Gender dbl+1bl 0 [1] Male
[2] Female
[NA(d)] No answer
```

Weil Sie kategorial ist, faktorisieren wir diese Variable:

```
ess8_CH_ss$gndr <- as_factor(ess8_CH_ss$gndr)
```


Warum eigentlich?

Vor der Regressionsanalyse ist es sinnvoll, einen Überblick zu unseren Variablen und insbesondere deren univariaten Eigenschaften zu erstellen: Die **Stichprobenstatistik**

Mit dem **table1()**-Befehl können wir automatisch eine Tabelle zur Stichprobenstatistik erstellen. Diese kann dann kopiert und in Word weiterbearbeitet werden.

```
table1(~imueclt + hinctnta + eduyrs + agea + gndr, data = ess8_CH_ss)
```

Vor der Regressionsanalyse ist es sinnvoll, einen Überblick über die Daten zu erhalten: Die St

WICHTIG: Bei unseren Stichprobenstatistiken beziehen wir uns immer auf den kompletten Datensatz (also noch mit NAs)

Diese können dann weiterbearbe

```
table1(~imueclt + hinctnta + eduyrs + agea + gndr,
```

Superpraktisch: **table1()** gibt bei kategorialen Variablen keine Mittelwertstatistiken an, sondern Anteilstatistiken

Obwohl diese Kategorie unbelegt ist, bleibt sie im Hintergrund weiter bestehen und wird angezeigt!

	Overall (N=1525)
Country's cultural life undermined or enriched by immigrants	
Mean (SD)	6.07 (2.26)
Median [Min, Max]	6.00 [0, 10.0]
Missing	14 (0.9%)
Household's total net income, all sources	
Mean (SD)	5.47 (2.57)
Median [Min, Max]	5.00 [1.00, 10.0]
Missing	274 (18.0%)
Years of full-time education completed	
Mean (SD)	11.3 (3.50)
Median [Min, Max]	10.0 [0, 26.0]
Missing	3 (0.2%)
Age of respondent, calculated	
Mean (SD)	47.8 (18.8)
Median [Min, Max]	48.0 [15.0, 94.0]
Missing	6 (0.4%)
Gender	
Male	788 (51.7%)
Female	737 (48.3%)
No answer	0 (0%)

Problem: Im Zuge des Faktorisierens per `as_factor()` legt R auch für solche Werte eine Kategorie an, die gar nicht (mehr) in der Variable vorkommen, aber durch ein Label mit ihr verbunden sind.

Lösung: Mit dem Befehl `fct_drop()` aus dem Package `forcats` können wir solche Phantomkategorien entfernen.

```
library(forcats)
ess8_CH_ss$gndr <- fct_drop(ess8_CH_ss$gndr)
table1(~imueclt + hinctnta + eduyrs + agea + gndr, data = ess8_CH_ss)
```

	Overall (N=1525)
Country's cultural life undermined or enriched by immigrants	
Mean (SD)	6.07 (2.26)
Median [Min, Max]	6.00 [0, 10.0]
Missing	14 (0.9%)
Household's total net income, all sources	
Mean (SD)	5.47 (2.57)
Median [Min, Max]	5.00 [1.00, 10.0]
Missing	274 (18.0%)
Years of full-time education completed	
Mean (SD)	11.3 (3.50)
Median [Min, Max]	10.0 [0, 26.0]
Missing	3 (0.2%)
Age of respondent, calculated	
Mean (SD)	47.8 (18.8)
Median [Min, Max]	48.0 [15.0, 94.0]
Missing	6 (0.4%)
Gender	
Male	788 (51.7%)
Female	737 (48.3%)

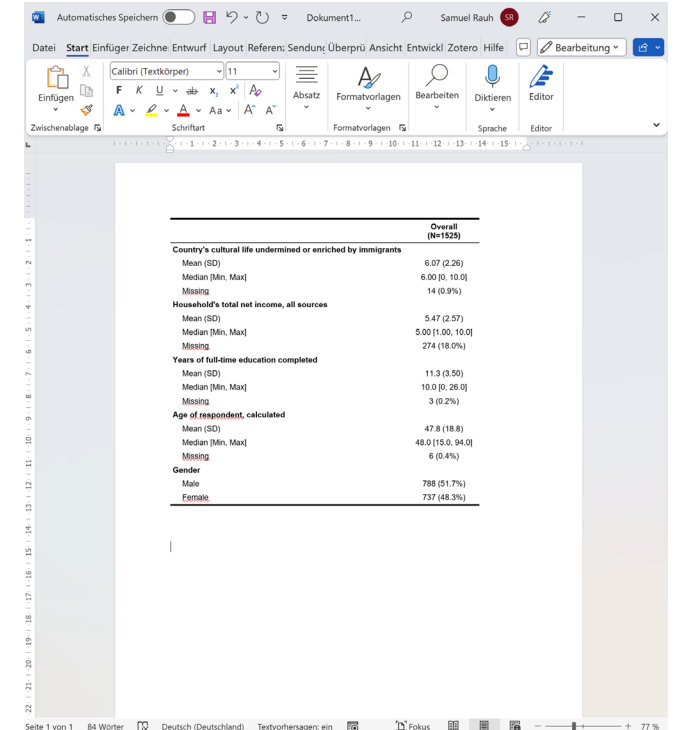
2

Stichprobenstatistik

Die mit table1() generierte Tabelle kann nun kopiert und in Word oder PowerPoint eingefügt und dort weiterbearbeitet werden

	Overall (N=1525)
Country's cultural life undermined or enriched by immigrants	
Mean (SD)	6.07 (2.26)
Median [Min, Max]	6.00 [0, 10.0]
Missing	14 (0.9%)
Household's total net income, all sources	
Mean (SD)	5.47 (2.57)
Median [Min, Max]	5.00 [1.00, 10.0]
Missing	274 (18.0%)
Years of full-time education completed	
Mean (SD)	11.3 (3.50)
Median [Min, Max]	10.0 [0, 26.0]
Missing	3 (0.2%)
Age of respondent, calculated	
Mean (SD)	47.8 (18.8)
Median [Min, Max]	48.0 [15.0, 94.0]
Missing	6 (0.4%)
Gender	
Male	788 (51.7%)
Female	737 (48.3%)

	Overall (N=1525)
Country's cultural life undermined or enriched by immigrants	
Mean (SD)	6.07 (2.26)
Median [Min, Max]	6.00 [0, 10.0]
Missing	14 (0.9%)
Household's total net income, all sources	
Mean (SD)	5.47 (2.57)
Median [Min, Max]	5.00 [1.00, 10.0]
Missing	274 (18.0%)
Years of full-time education completed	
Mean (SD)	11.3 (3.50)
Median [Min, Max]	10.0 [0, 26.0]
Missing	3 (0.2%)
Age of respondent, calculated	
Mean (SD)	47.8 (18.8)
Median [Min, Max]	48.0 [15.0, 94.0]
Missing	6 (0.4%)
Gender	
Male	788 (51.7%)
Female	737 (48.3%)



2

Stichprobenharmonisierung

Vor der Regressionsanalyse werden üblicherweise alle Personen, die in irgendeiner analyserelevanten Variable ein **NA** tragen, aus der Stichprobe entfernt. Damit wird sichergestellt, dass für alle Modelle einer Analyse exakt dieselbe Stichprobe verwendet wird

```
ess8_noNA <- na.omit(ess8_CH_ss)
```

Mit dem **na.omit()** Befehl werden alle Reihen, welche mindestens einen NA Wert enthalten aus dem Datensatz entfernt.

Wichtig: Diese Operation erst *nach* der Stichprobenstatistik durchführen, da sonst keine valides Bild der Nonresponse-Problematik angeboten wird!

3

Bivariate Regression

Berechnen wir zuerst eine bivariate Regression:

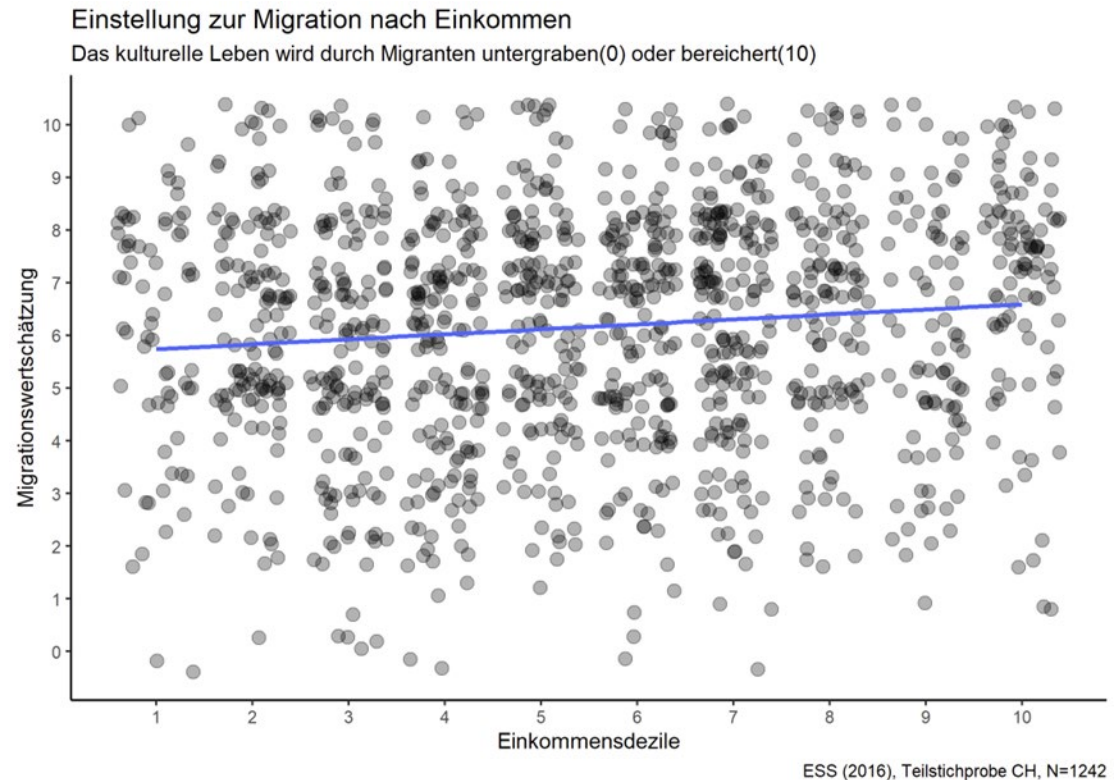
```
bi_model <- lm(imueclt ~ hinctnta, data = ess8_noNA)
summary(bi_model)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5.64297	0.14781	38.177	< 2e-16	***
hinctnta	0.09451	0.02441	3.872	0.000114	***

Interpretation des p-Wertes?

- Die Wahrscheinlichkeit für den erhaltenen Regressionskoeffizienten, wenn es in Wirklichkeit keinen Zusammenhang zwischen Einkommen und Migrationseinstellung gibt, ist kleiner als 1%
- Achtung: Für die Bewertung der kausalen Hypothese verwenden wir aufgrund der Störproblematik dieses unbereinigte Modell nicht mehr.



4

Trivariate Regression

Für eine bivariate Regression benutzen wir folgenden Code

➤ `lm(AV~UV, data=Datensatz)`

Mit "+" können wir nun zusätzliche unabhängige Variablen integrieren

➤ `lm(AV~UV1+UV2+UV3+..., data=Datensatz)`

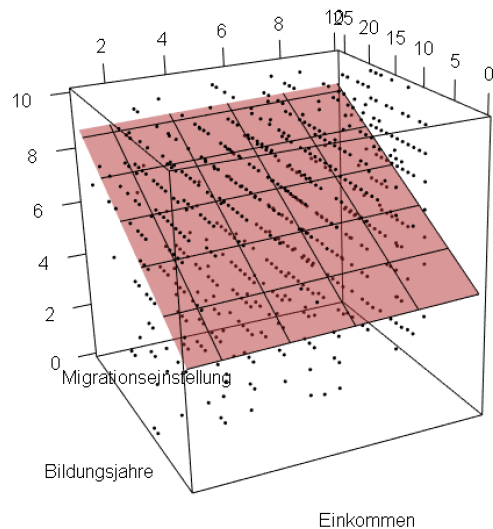
... zum Beispiel für eine trivariate Regression mit der zusätzlichen UV Bildung

```
tri_model <- lm(imueclt ~ hinctnta + eduyrs, data = ess8_noNA)
```

4

Trivariate Regression

```
tri_model <- lm(imueclt ~ hinctnta + eduysr, data = ess8_noNA)
summary(tri_model)
```



Unsere drei Variablen spannen nun einen dreidimensionalen Raum auf. Die Regressionskoeffizienten entsprechen dabei den Neigungsgradienten der bestmöglichen Ebene.

```
call:
lm(formula = imueclt ~ hinctnta + eduysr, data = ess8.CH.noNA)

Residuals:
    Min       1Q   Median       3Q      Max
-6.7692 -1.5479  0.2602  1.4930  4.9479

Coefficients:
            Estimate std. Error t value Pr(>|t|)
(Intercept)  3.93456    0.21889   17.975  <2e-16 ***
hinctnta     0.02937    0.02429    1.209    0.227
eduysr       0.18114    0.01767   10.254  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.125 on 1239 degrees of freedom
Multiple R-squared:  0.08923,    Adjusted R-squared:  0.08776
F-statistic: 60.69 on 2 and 1239 DF,  p-value: < 2.2e-16
```

Wie kann der Regressionskoeffizient interpretiert werden?

*Unter Kontrolle der Bildung steigt die Migrationswertschätzung pro Einkommensdezilstufe im Schnitt **nur noch** um 0.03 Skalenpunkte.*

4

Trivariate Regression

```
> summary(bi_model)
```

```
Call:
lm(formula = imueclt ~ hinctnta, data = ess8_noNA)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-6.3045 -1.3991  0.4119  1.6955  4.2625
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.64297    0.14781  38.177 < 2e-16 ***
hinctnta     0.09451    0.02441   3.872 0.000114 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.213 on 1240 degrees of freedom
Multiple R-squared:  0.01194, Adjusted R-squared:  0.01115
F-statistic: 14.99 on 1 and 1240 DF, p-value: 0.0001137
```

```
> summary(tri_model)
```

```
Call:
lm(formula = imueclt ~ hinctnta + eduysr, data = ess8_noNA)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-6.7692 -1.5479  0.2602  1.4930  4.9479
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.93456    0.21889  17.975 <2e-16 ***
hinctnta     0.02937    0.02429   1.209  0.227
eduysr       0.18114    0.01767  10.254 <2e-16 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.125 on 1239 degrees of freedom
Multiple R-squared:  0.08923, Adjusted R-squared:  0.08776
F-statistic: 60.69 on 2 and 1239 DF, p-value: < 2.2e-16
```

Vergleiche das bivariate und das trivariate Modell. Was fällt auf?

Der Regressionskoeffizient wird unter Kontrolle der Bildung deutlich kleiner und ist zudem nicht mehr statistisch signifikant.

Wir haben eine Scheinkausalität aufgedeckt: Der Zusammenhang zwischen Einkommen und Migrationseinstellung wird durch einen Bildungseffekt dominiert!

5

Regressionstabellen

Der «lm»-Standardoutput (ebenso die Darstellung über «summary()») ist nicht publikationswürdig. **stargazer()** hilft uns bei der akuraten Darstellung der Regressionsergebnisse.

```
stargazer(bi_model, tri_model, type = "text")
```

Erstelle eine tabellarische Übersicht zu Regressionsergebnissen

Binde zwei Regressionsergebnisse in die Tabelle ein: Die erste Spalte der Tabelle soll die unter «bi_model», die zweite Spalte die unter «tri_model» gespeicherten Koeffizienten auflisten

... für projektinterne Kommunikation ok, aber immer noch nicht publikationswürdig.

Dependent variable:			
	imueclt		
	(1)		(2)
hinctnta	0.095*** (0.024)		0.029 (0.024)
eduyrs			0.181*** (0.018)
Constant	5.643*** (0.148)		3.935*** (0.219)
Observations	1,242		1,242
R2	0.012		0.089
Adjusted R2	0.011		0.088
Residual std. Error	2.213 (df = 1240)		2.125 (df = 1239)
F Statistic	14.990*** (df = 1; 1240)		60.694*** (df = 2; 1239)

Note: *p<0.1; **p<0.05; ***p<0.01

Wir passen den **stargazer()**-Befehl nun so an, dass wir eine publikationswürdige Tabelle annähern

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
          type = "text")
```

Basisoutput des **stargazer()**-Befehls

```
=====
                                Dependent variable:
                                -----
                                imueclt
                                (1)                (2)
-----
hinctnta                        0.095***          0.029
                                (0.024)          (0.024)

eduys                            0.181***
                                (0.018)

Constant                        5.643***          3.935***
                                (0.148)          (0.219)

-----
Observations                    1,242          1,242
R2                              0.012          0.089
Adjusted R2                    0.011          0.088
Residual Std. Error    2.213 (df = 1240)    2.125 (df = 1239)
F Statistic             14.990*** (df = 1; 1240) 60.694*** (df = 2; 1239)
=====
Note:                            *p<0.1; **p<0.05; ***p<0.01
> |
```

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
          type = "text",
          dep.var.caption = "",
          dep.var.labels = "")
```

Überschrift “dependent variable“ und Name der AV wird aus der Tabelle entfernt

	(1)	(2)
hinctnta	0.095*** (0.024)	0.029 (0.024)
edyrs		0.181*** (0.018)
Constant	5.643*** (0.148)	3.935*** (0.219)
Observations	1,242	1,242
R2	0.012	0.089
Adjusted R2	0.011	0.088
Residual std. Error	2.213 (df = 1240)	2.125 (df = 1239)
F statistic	14.990*** (df = 1; 1240)	60.694*** (df = 2; 1239)

Note: *p<0.1; **p<0.05; ***p<0.01

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
          type = "text",
          dep.var.caption = "",
          dep.var.labels = "",
          column.labels = c("Bivariates
Modell", "Nettomodell 1"))
```

Benennung der Spalten zu „Bivariates Modell“
und „Nettomodell 1“.

	Bivariates Modell (1)	Nettomodell 1 (2)
hinctnta	0.095*** (0.024)	0.029 (0.024)
eduysr		0.181*** (0.018)
Constant	5.643*** (0.148)	3.935*** (0.219)
Observations	1,242	1,242
R ²	0.012	0.089
Adjusted R ²	0.011	0.088
Residual Std. Error	2.213 (df = 1240)	2.125 (df = 1239)
F Statistic	14.990*** (df = 1; 1240)	60.694*** (df = 2; 1239)

Note: *p<0.1; **p<0.05; ***p<0.01

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
          type = "text",
          dep.var.caption = "",
          dep.var.labels = "",
          column.labels = c("Bivariates Modell",
                             "Nettomodell 1"),
          single.row = T,
          omit.stat = c("f", "ser", "adj.rsq"))
```

Der Standardfehler wird nun auf der gleichen Zeile wie der Koeffizient abgebildet und alle überflüssigen Statistiken werden aus der Tabelle entfernt.

	Bivariates Modell (1)	Nettomodell 1 (2)
hinctnta	0.095*** (0.024)	0.029 (0.024)
eduyrs		0.181*** (0.018)
Constant	5.643*** (0.148)	3.935*** (0.219)
Observations	1,242	1,242
R2	0.012	0.089
Note:	*p<0.1; **p<0.05; ***p<0.01	

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
          type = "text",
          dep.var.caption = "",
          dep.var.labels = "",
          column.labels = c("Bivariates
Modell", "Nettomodell 1"),
          single.row = T,
          omit.stat = c("f", "ser", "adj.rsq"),
          digits = 2, digits.extra = 5)
```

Begrenzung auf zwei Nachkommastellen, mit der Möglichkeit maximal 5 Nachkommastellen anzuzeigen, bis eine substantielle Stelle vorliegt.

In den Sozialwissenschaften verwenden wir in der Regel zwei bis drei Nachkommastellen. Zudem sollten unsere Zahlen immer mindestens eine substantielle Stelle (ungleich Null) haben.

	Bivariates Modell (1)	Nettomodell 1 (2)
hinctnta	0.09*** (0.02)	0.03 (0.02)
eduyrs		0.18*** (0.02)
Constant	5.64*** (0.15)	3.93*** (0.22)
Observations	1,242	1,242
R2	0.01	0.09
Note:	*p<0.1; **p<0.05; ***p<0.01	

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
  type = "text",
  dep.var.caption = "",
  dep.var.labels = "",
  column.labels = c("Bivariates
Modell", "Nettomodell 1")
  single.row = T,
  omit.stat = c("f", "ser", "adj.rsq"),
  digits = 2, digits.extra = 5,
  star.cutoffs = c(0.05, 0.01, 0.001))
```

Anpassung der Signifikanzlevel.

Entsprechend aktueller sozialwissenschaftlicher Konventionen kennzeichnet ein Sternchen das Unterschreiten von $p=0.05$ (und nicht, wie stargazer-Default, von $p=0.1$)

	Bivariates Modell		Nettomodell 1	
	(1)		(2)	
hinctnta	0.09***	(0.02)	0.03	(0.02)
eduyrs			0.18***	(0.02)
Constant	5.64***	(0.15)	3.93***	(0.22)
Observations	1,242		1,242	
R2	0.01		0.09	

Note: * $p<0.05$; ** $p<0.01$; *** $p<0.001$

5

Regressionstabellen

```
stargazer(bi_model, tri_model,
  type = "text",
  dep.var.caption = "",
  dep.var.labels = "",
  column.labels = c("Bivariates
Modell", "Nettomodell 1"),
  single.row = T,
  omit.stat = c("f", "ser", "adj.rsq"),
  digits = 2, digits.extra = 5,
  star.cutoffs = c(0.05, 0.01, 0.001),
  notes = "Daten: ESS(2016),
Teilstichprobe CH. Standardfehler in Klammern",
  title = "Determinanten der
Migrationswertschätzung")
```

Fussnote und Titel zur Tabelle
hinzufügen.

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
hinctnta	0.09*** (0.02)	0.03 (0.02)
edyrs		0.18*** (0.02)
Constant	5.64*** (0.15)	3.93*** (0.22)
Observations	1,242	1,242
R2	0.01	0.09

Note:

*p<0.05; **p<0.01; ***p<0.001
Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

Diese Tabelle sieht nun schon ziemlich gut,
wenn auch nicht publikationsfähig aus

Die weitere Bearbeitung nehmen wir
ausserhalb von R vor. Dazu müssen wir die
Tabelle zunächst exportieren.

```
stargazer(bi_model, tri_model,
          type = "html",
          dep.var.caption = "",
          dep.var.labels = "",
          column.labels = c("Bivariates
Modell", "Nettomodell 1"),
          single.row = T,
          omit.stat = c("f", "ser", "adj.rsq"),
          digits = 2, digits.extra = 5,
          star.cutoffs = c(0.05, 0.01, 0.001),
          notes = "Daten: ESS(2016),
Teilstichprobe CH. Standardfehler in Klammern",
          title = "Determinanten der
Migrationswertschätzung",
          out = "reg_table.doc")
```

Mit dem Befehl **type = "html"** können wir unsere Tabelle als HTML-Format spezifizieren. Dies stellt sicher, dass Ränder und Linien der Tabelle richtig formatiert werden.

Indem, wir im **stargazer()**-Befehl **out =** spezifizieren, können wir angeben unter welchem Namen unser File gespeichert werden soll und zudem auch den Dateityp bestimmen (.doc, .html, .txt etc.).

5

Regressionstabelle

Wie finde ich nun mein abgespeichertes Dokument?

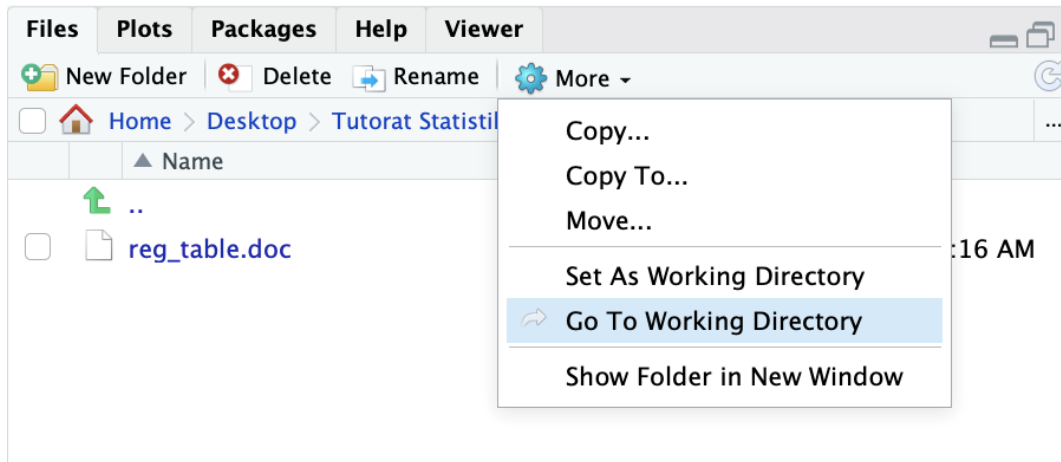
Drei Verschiedene Möglichkeiten:

1) Über die working directory den Speicherort auf dem Computer finden

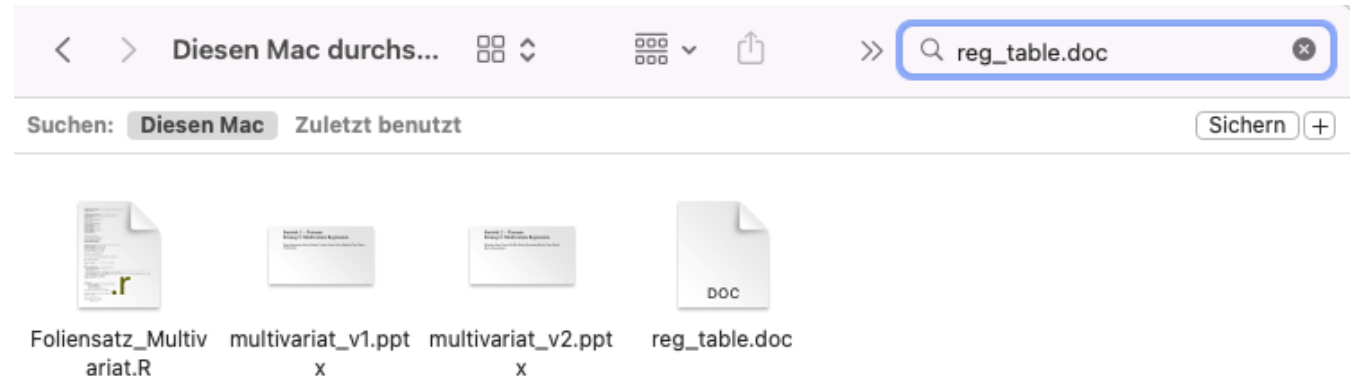
> `getwd()`

[1] `"/Users/sebastian/Desktop/Tutorat Statistik2/Regressionstabellen"`

2) In R-Studio direkt zur working directory



3) Computer nach dem Dokumentnamen durchsuchen



reg_table.doc

Suchen

Samuel Raul

Entwurf Layout Referenzen Sendungen Überprüfen Ansicht Entwicklertools Zotero Hilfe

Standard Kein Leerraum Überschrift Überschrift 2 Titel

Schriftart Absatz Formatvorlagen Bearbeiten

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
<u>hinctnta</u>	0.09*** (0.02)	0.03 (0.02)
<u>eduyrs</u>		0.18*** (0.02)
Constant	5.64*** (0.15)	3.93*** (0.22)
<u>Observations</u>	1,242	1,242
R ²	0.01	0.09

Note: *p<0.05; **p<0.01; ***p<0.001

Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

reg_table.doc

Suchen

Samuel Raul

Entwurf Layout Referenzen Sendungen Überprüfen Ansicht Entwicklertools Zotero Hilfe

Standard Kein Leerraum Überschrift Überschrift 2 Titel

Schriftart Absatz Formatvorlagen Bearbeiten

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Constant	5.64*** (0.15)	3.93*** (0.22)
<u>Observations</u>	1,242	1,242
R ²	0.01	0.09

Note: *p<0.05; **p<0.01; ***p<0.001

Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

Auf Word können wir die Tabelle dann weiterbearbeiten und zum Beispiel die Zeilen bzw. Variablen umbenennen, den p-Wert Schlüssel ergänzen (... wird in einigen Formaten manchmal unvollständig ausgegeben) oder die Schriftart ändern.

...Sinn und Zweck der tabellarischen Darstellung: Einfache, vergleichende Darstellung der wichtigen Regressionsparameter:

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Constant	5.64*** (0.15)	3.93*** (0.22)
<u>Observations</u>	1,242	1,242
R ²	0.01	0.09

Note:

*p<0.05; **p<0.01; ***p<0.001

Daten: [ESS\(2016\)](#), Teilstichprobe CH. Standardfehler in Klammern

- **Unter Kontrolle der Bildung steigt die Migrationswertschätzung pro Dezilstufe im Schnitt um 0.03 Skalenpunkte**
- **Unter Kontrolle der Bildung wird kein statistisch signifikanter Einkommenseffekt auf die Migrationseinstellung ausgewiesen.**
- **Der unbereinigte Einkommenschätzer («Bruttomodell») transportiert also zu grossem Anteil einen Bildungseffekt!**

Vergleichen wir nun die Vorhersagekraft der zwei Modelle. Wieso unterscheidet sich diese?

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Constant	5.64*** (0.15)	3.93*** (0.22)
<u>Observations</u>	1,242	1,242
R ²	0.01	0.09

Note:

*p<0.05; **p<0.01; ***p<0.001

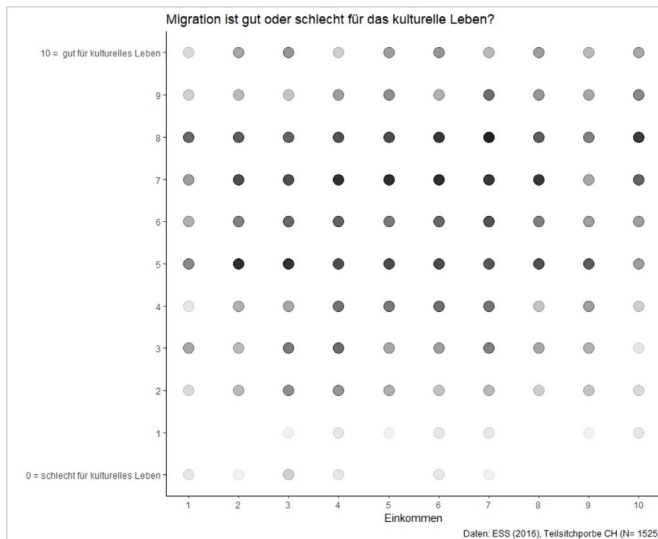
Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

Durch die Integration zusätzlicher Variablen kann mehr Varianz der abhängigen Variable erklärt (bzw. die Vorhersagekraft des Modells verbessert) werden. Indiziert wird dies durch das erhöhte R².

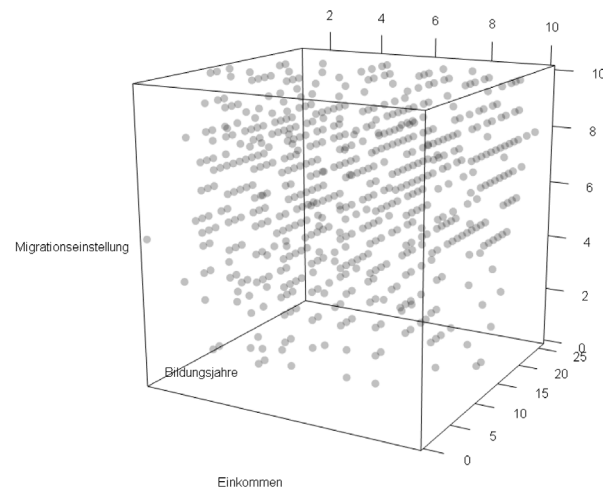
6

Multiple Regression

Wir haben in unserer Regressionsanalyse nun das Störmerkmal *Bildung* kontrolliert. Aus den Vorüberlegungen wissen wir, dass das Alter ebenfalls ein Störfaktor ist. Wir müssen also unser Modell um eine weitere Variable erweitern.



Bivariate Regression



Trivariate Regression

Quadrivariate Regression

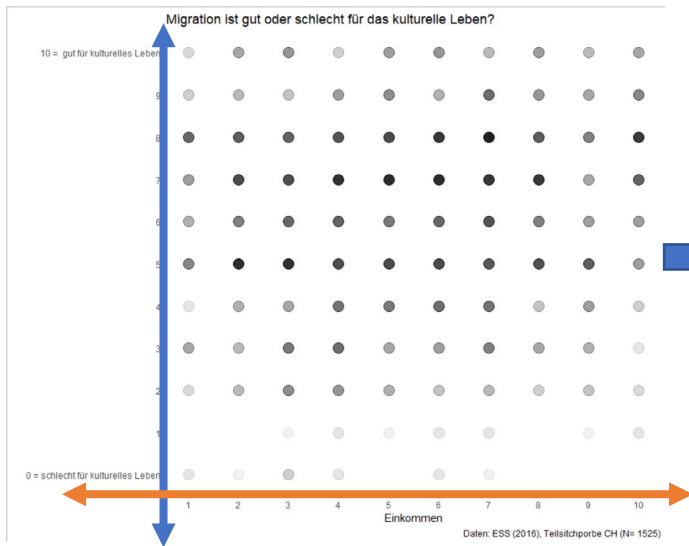
...

Multiple Regression

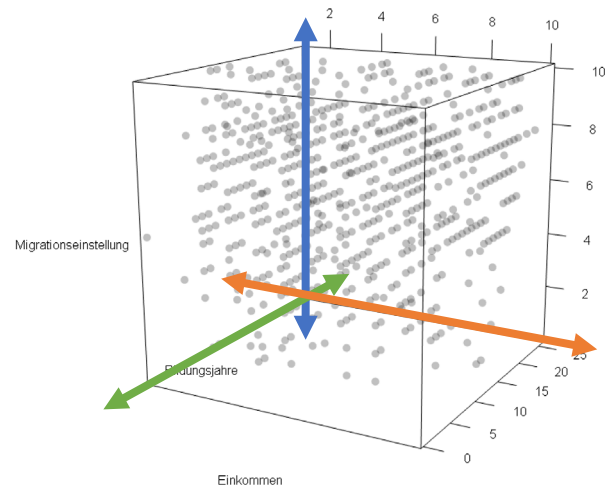
6

Multiple Regression

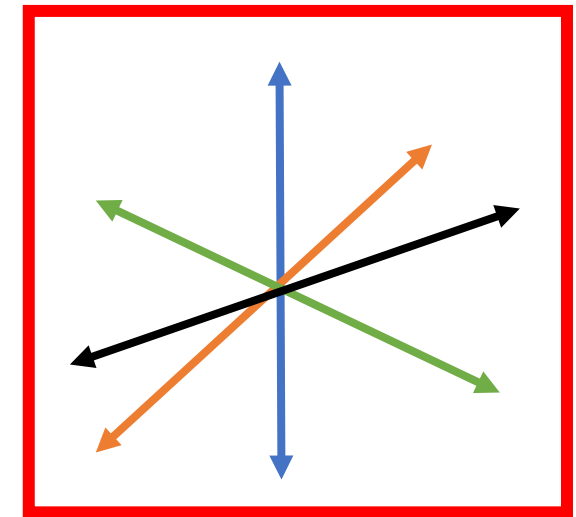
Mit jeder weiteren Variable erweitern wir unseren Streuraum um eine zusätzliche Dimension, sprich unsere Punkte verteilen sich in eine weitere (uns allerdings nicht ersichtliche) Richtung. In unseren Fall hat dann jeder Punkt eine Einkommens-, Bildungs-, Alter-, und Migrationskoordinate.



Bivariate Regression



Trivariate Regression



Quadrivariate Regression

...

Multiple Regression

6

Multiple Regression

Einfache Umsetzung mit R:

```
multi_model <- lm(imueclt ~ hinctnta + eduysr + agea, data = ess8_noNA)
summary(multi_model)
```

Call:

```
lm(formula = imueclt ~ hinctnta + eduysr + agea, data = ess8_noNA)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.7041	-1.5242	0.2662	1.4653	4.9025

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.695148	0.302824	12.202	<2e-16 ***
hinctnta	0.033515	0.024557	1.365	0.173
eduysr	0.183239	0.017759	10.318	<2e-16 ***
agea	0.003946	0.003449	1.144	0.253

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Unter Kontrolle *der Bildung und des Alters* steigt die Migrationswertschätzung mit jedem Dezilssprung des Einkommens um 0.03 Skaleneinheiten.

*Migrationwertschätzung = 3.69 + 0.03 * Einkommen + 0.18 * Bildungsjahre + 0.04 * Alter*

6

Multiple Regression

Integration der Ergebnisse in den stargazer...

```
stargazer(bi_model, tri_model, multi_modell1
  type = "text",
  dep.var.caption = "",
  dep.var.labels = "",
  column.labels = c("Bivariates Modell",
"Nettomodell 1", "Nettomodell 2"),
  single.row = T,
  omit.stat = c("f", "ser", "adj.rsq"),
  digits = 2, digits.extra = 5,
  star.cutoffs = c(0.05, 0.01, 0.001),
  notes = "Daten: ESS(2016),
Teilstichprobe CH. Standardfehler in Klammern",
  title = "Determinanten der
Migrationswertschätzung")
```

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)	Nettomodell 2 (3)
hinctnta	0.09*** (0.02)	0.03 (0.02)	0.03 (0.02)
eduyrs		0.18*** (0.02)	0.18*** (0.02)
agea			0.004 (0.003)
Constant	5.64*** (0.15)	3.93*** (0.22)	3.70*** (0.30)
Observations	1,242	1,242	1,242
R2	0.01	0.09	0.09

Note: *p<0.05; **p<0.01; ***p<0.001

Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

Interpretation der Veränderung zwischen Nettomodell 1 und Nettomodell 2?

Der Koeffizient des Einkommens ist *in diesem Fall* robust gegenüber der Alterskontrolle

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Alter		0.004 (0.003)
Constant	5.64*** (0.15)	3.70*** (0.30)
Observations	1,242	1,242
R ²	0.01	0.09

Note: *p<0.05; **p<0.01; ***p<0.001

Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

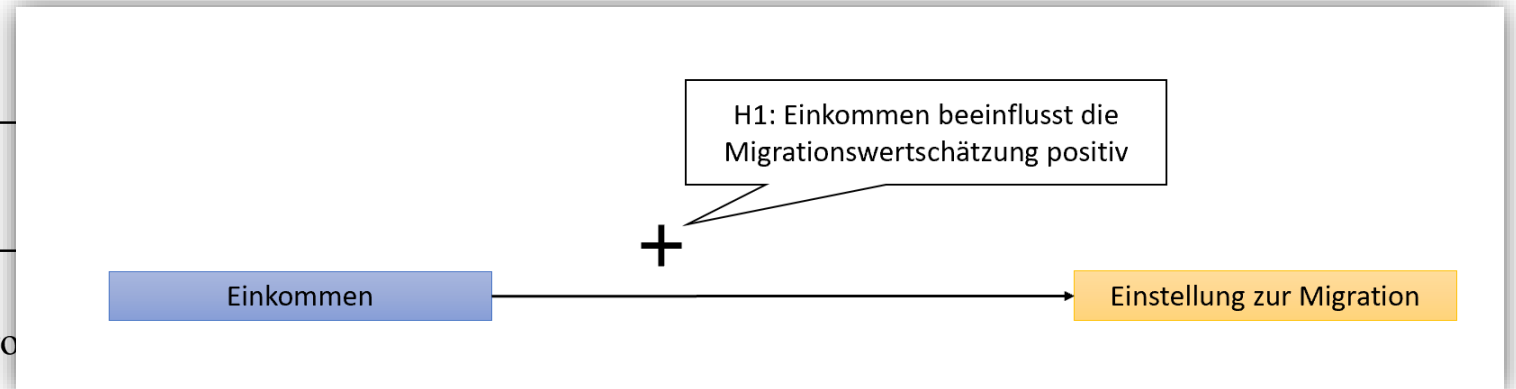
In der publizierten
Regressionstabelle wird meist nur
noch das bivariate Modell und das
„vollständige“ Nettomodell
dargestellt

Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Alter		
Constant	5.64*** (0.15)	
Observations	1,242	
R ²	0.01	

Note:

Daten: ESS(2016), Teilstichpro



- Wie bewerten wir vor dem Hintergrund der Ergebnisse die Hypothese?
- Welches Analyseergebnis nutzen wir zur Hypothesenbewertung?

Hypothese = Kausale Formulierung! → Relevanter Koeffizient = Bereinigter Koeffizient!

„Unter Kontrolle des Bildungs- und des Alterseffektes ist der Einkommenskoeffizient nicht statistisch signifikant von 0 verschieden. Unser Kausalhypothese H1 ist somit *nicht* gestützt!“

Übung

Untersucht nun die die beiden Variablen *Position auf der Links-Rechts-Skala* (Irscale, UV) und Lebenszufriedenheit (stflife, AV).

Bildet zunächst eine sinnvolle Hypothese.

Übung

Untersucht nun die die beiden Variablen *Position auf der Links-Rechts-Skala* (Irscale, UV) und Lebenszufriedenheit (stflife, AV).

H1: Die politische Position beeinflusst die Lebenszufriedenheit.

1. Identifiziert theoretisch fundiert Störmerkmale (metrisch)
2. Berechnet den Koeffizienten der bivariaten Regression
3. Berechnet den Koeffizienten der multivariaten Regression
4. Stellt beide Analysen in einer Tabelle dar
5. Vergleicht und interpretiert die Koeffizienten



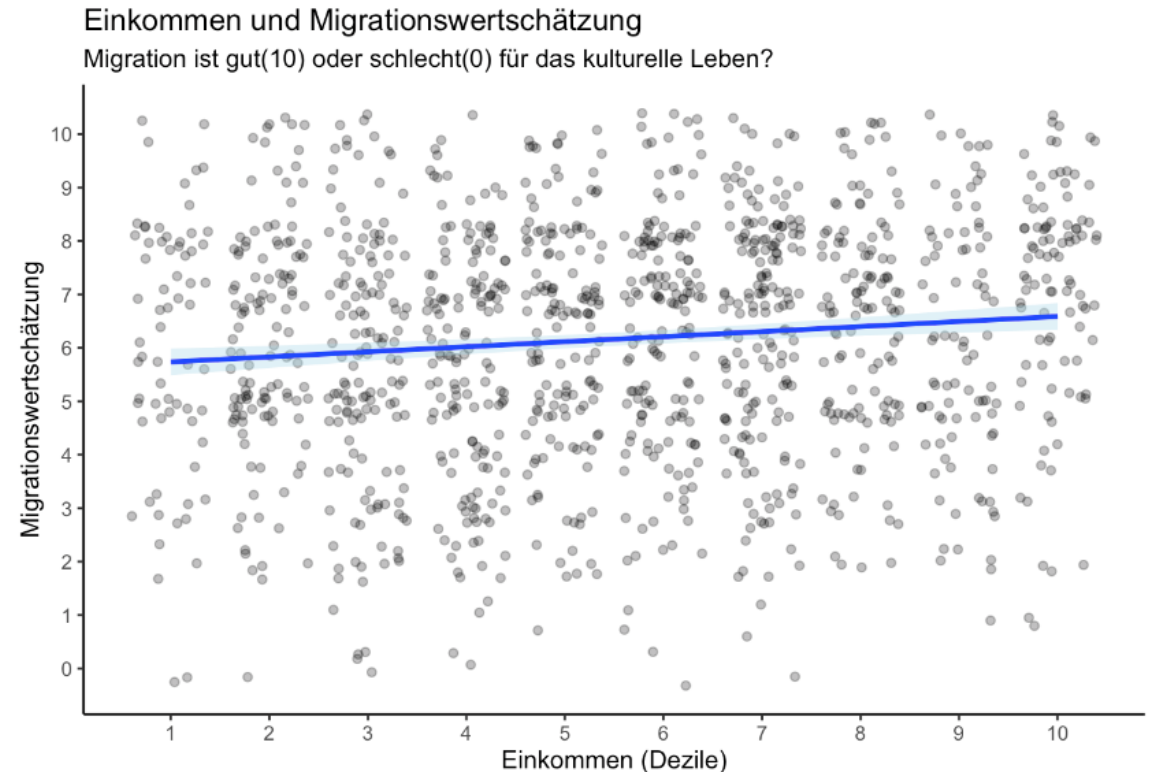
Visualisierung von Regressionsergebnissen

7

Visualisierung

Ihr könnt bereits bivariate Regressionsergebnisse ploten...

```
ggplot(ess8_noNA, aes(hinctnta, imueclt))+
  geom_jitter(alpha=0.3)+
  theme_classic()+
  scale_y_continuous(breaks = seq(0,10))+
  scale_x_continuous(breaks = seq(1,10))+
  geom_smooth(method = "lm", se = T, level = 0.95, color = "blue", fill="lightblue")+
  labs(x = "Einkommen (Dezile)", y = "Migrationswertschätzung",
       title = "Einkommen und Migrationswertschätzung",
       caption = "ESS (2016), Teilstichprobe CH, N=1242",
       subtitle = "Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?")
```



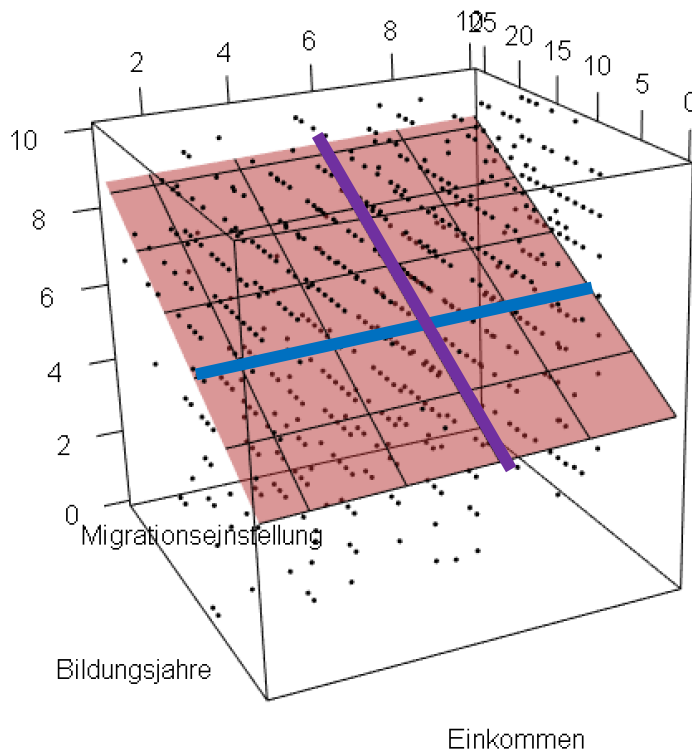
ESS (2016), Teilstichprobe CH, N=1242

...ggplot ist allerdings nicht darauf ausgerichtet, Ergebnisse einer multiplen Regression akkurat zu visualisieren.

7

Visualisierung

Die Ergebnisse einer trivariaten Regression können anhand der bestmöglichen Ebene im 3D-Streuraum visualisiert werden. Die Neigungsgradienten dieser Ebene entsprechen dabei unseren Regressionskoeffizienten.



Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Konstante	5.64*** (0.15)	3.93*** (0.22)
Observations	1,242	1,242
R ²	0.01	0.09

Note:

*p<0.05; **p<0.01; ***p<0.001

Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

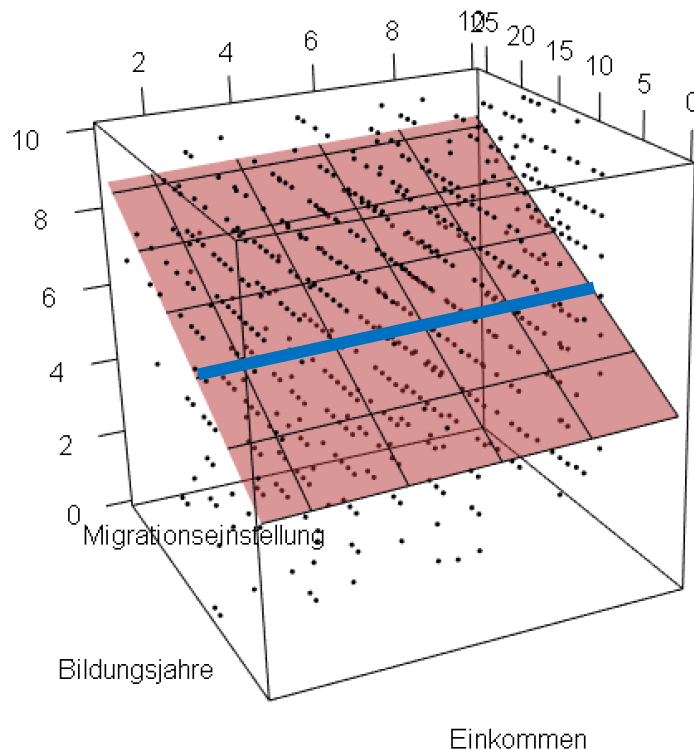
3D-Visualisierungen sind zur Kommunikation von Regressionsergebnissen in der Wissenschaft aber eher unüblich.

Dies liegt unter anderem auch daran, dass wir meistens mehr als 1 Drittvariable berücksichtigen.

7

Visualisierung

Die Ergebnisse einer trivariaten Regression können anhand der bestmöglichen Ebene im 3D-Streuraum visualisiert werden. Die Neigungsgradienten dieser Ebene entsprechen dabei unseren Regressionskoeffizienten.



Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell 1 (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Konstante	5.64*** (0.15)	3.93*** (0.22)
Observations	1,242	1,242
R ²	0.01	0.09

Note:

*p<0.05; **p<0.01; ***p<0.001

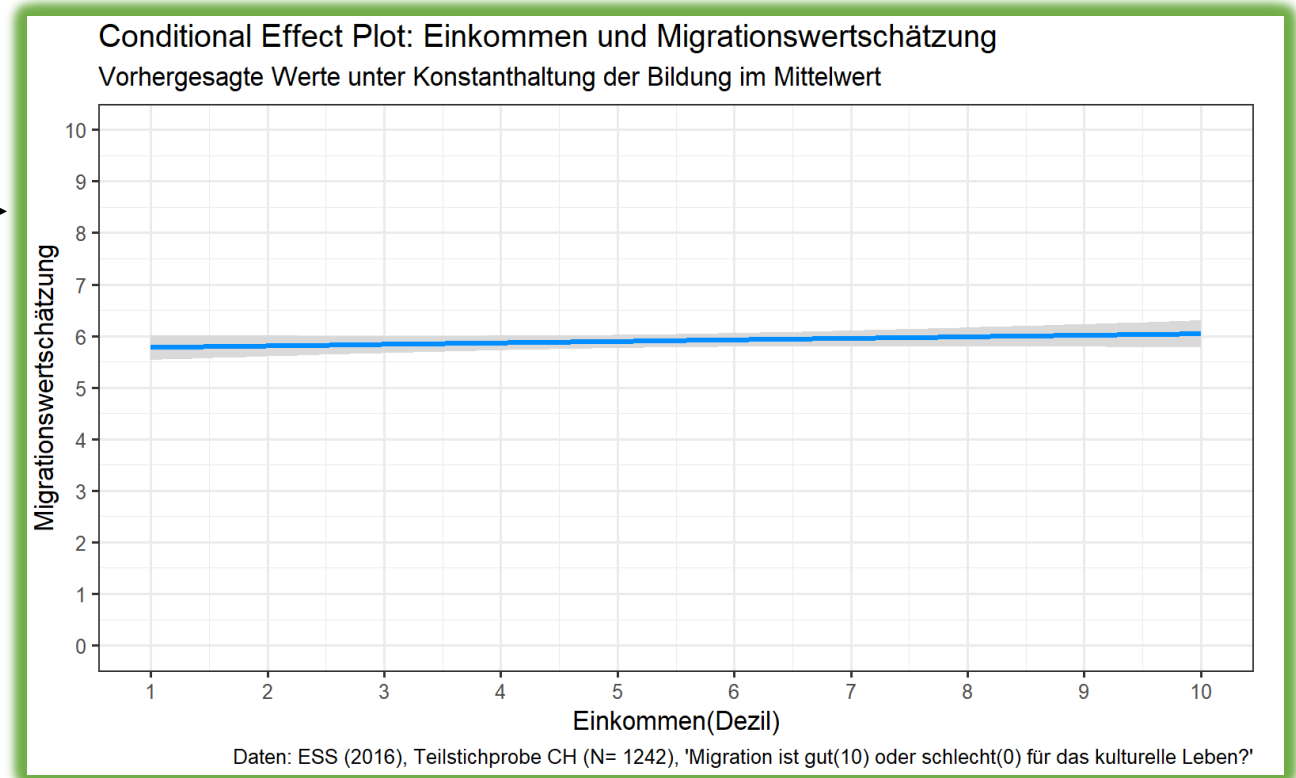
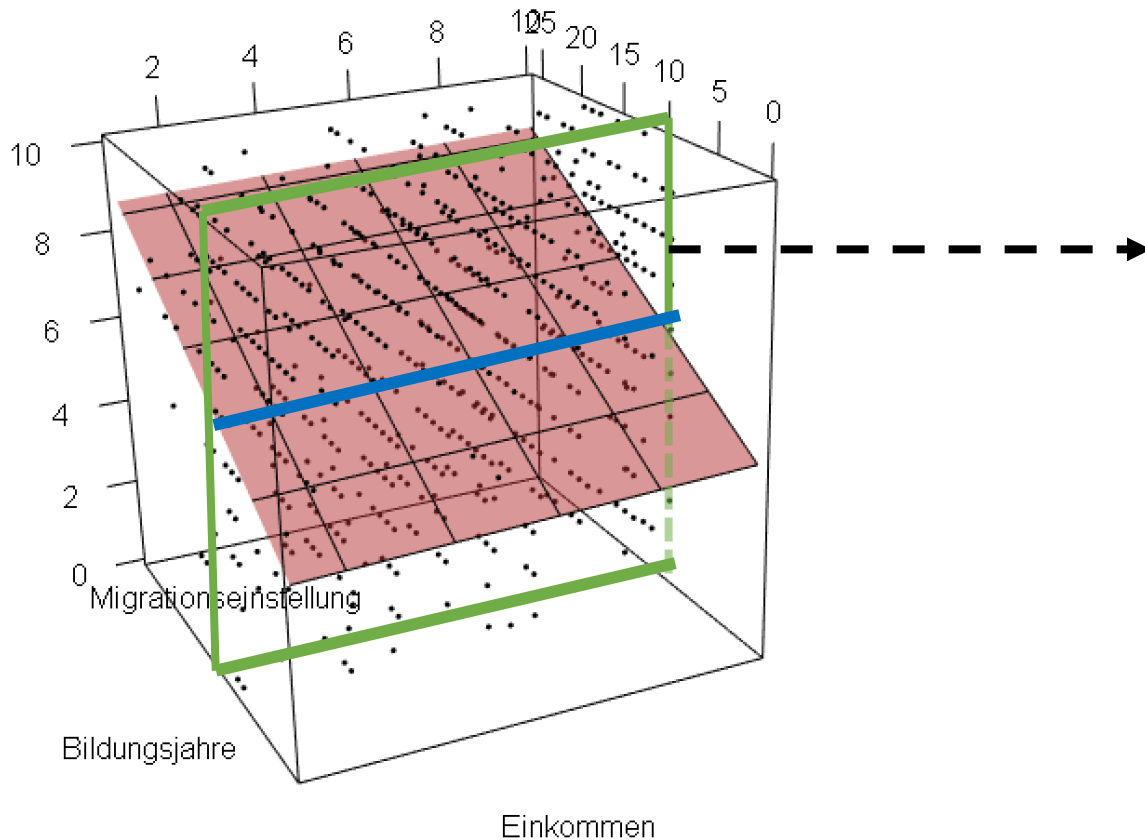
Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

...zudem ist unser Visualisierungsinteresse entsprechend der formulierten Hypothese in der Regel ausschliesslich auf den Einfluss der zentralen unabhängigen Variable gerichtet

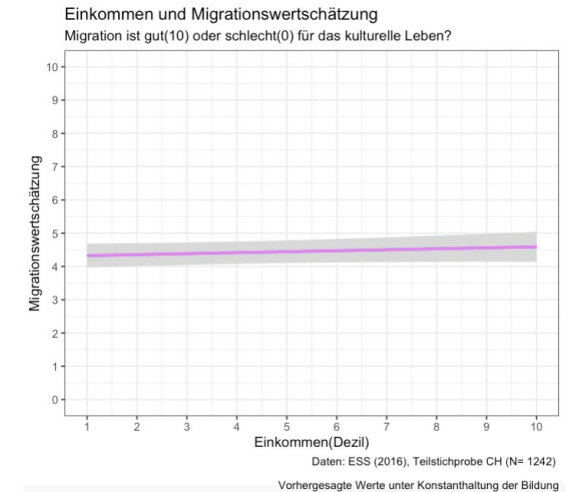
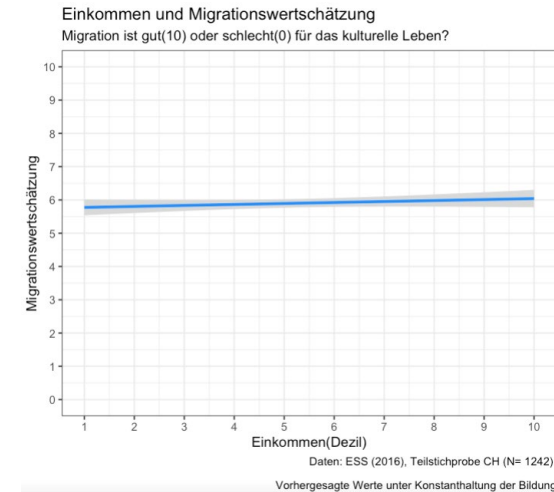
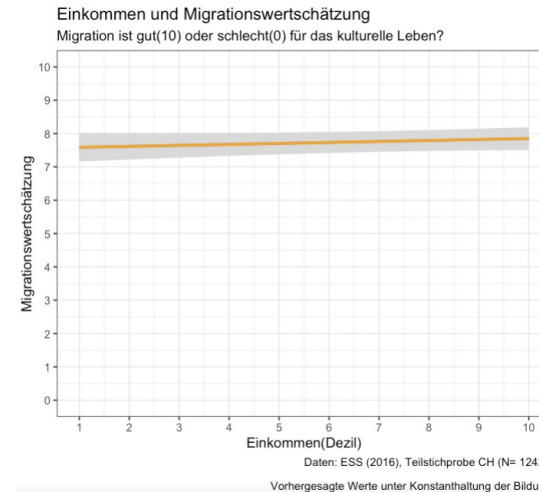
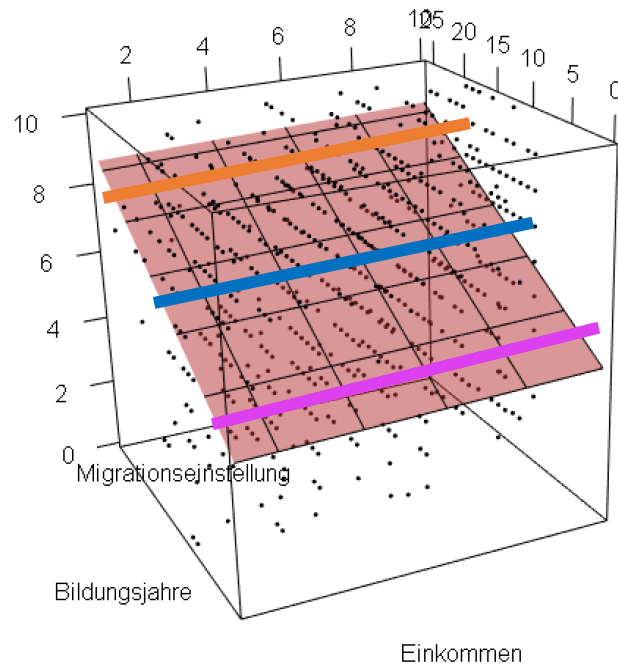
7

Visualisierung

Also visualisieren wir eine einzelne 2-dimensionale Scheibe aus dem Streuraum. Diese Scheibe stellt die Neigung der Regressionsebene in der x/y-Perspektive dar, verdeutlicht also den Effekt von x auf y unter Kontrolle von z.



Welche Scheibe aus dem Streuraum soll nun visualisiert werden?

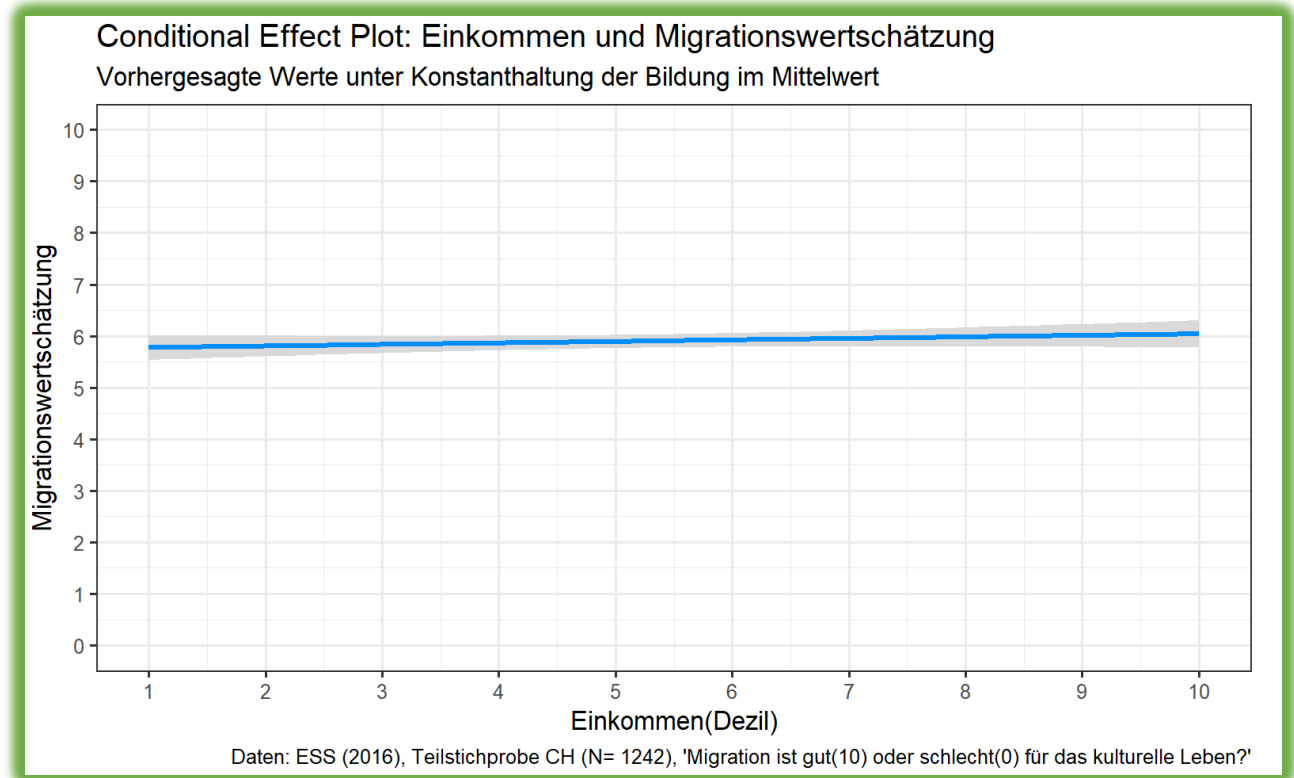
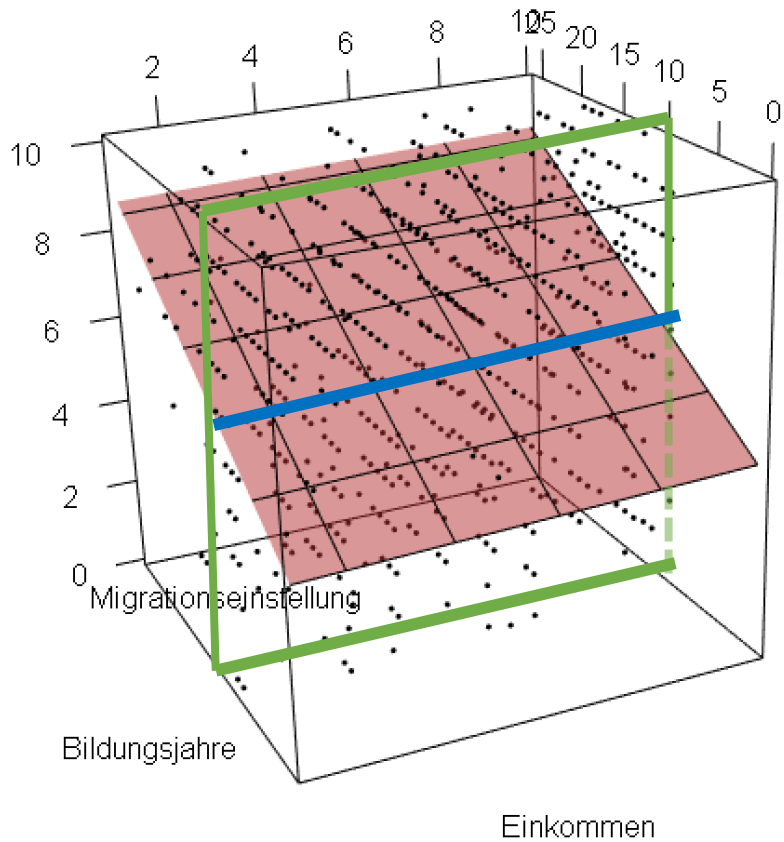


Es spielt keine Rolle, wo wir den Schnitt ansetzen. Der Neigungsgradient der Ebene ist bei allen möglichen Schnittpunkten identisch – alle «seitlichen» Schnitte durch die Ebene repräsentieren den bereinigten Effekt des Einkommens auf die Migrationswertschätzung.

7

Visualisierung

Gemäss Konvention visualisieren wir die Scheibe jeweils für die Mittelwerte der Drittvariablen. Dies ist auch der Default für den **Conditional Effect Plot**.



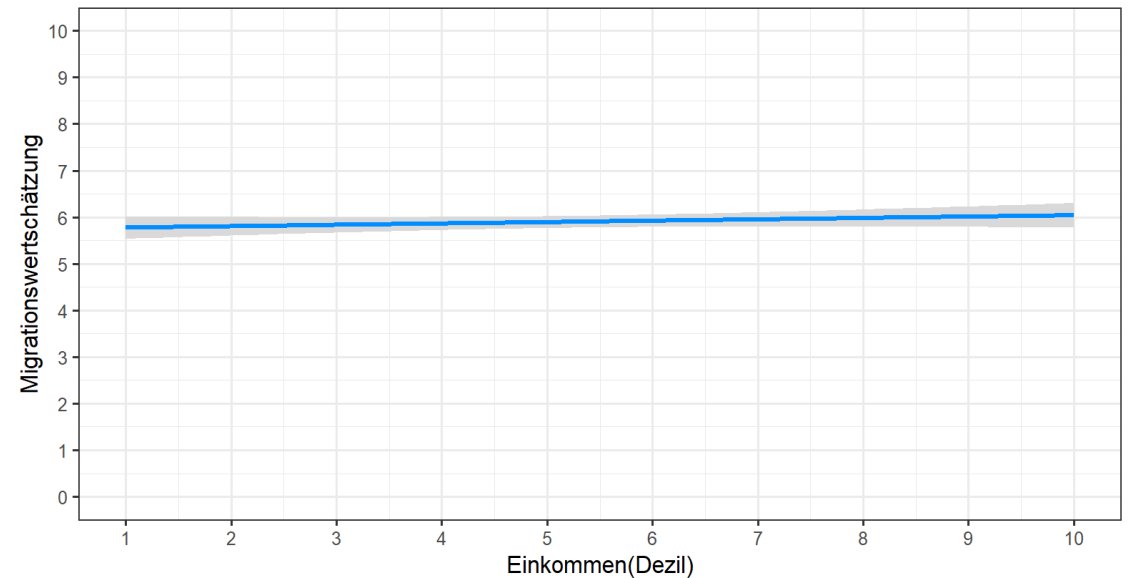
7

Visualisierung: Conditional Effect Plot

Um *Conditional Effect Plots* zu erstellen, verwenden wir die Funktion `visreg::visreg()`

```
tri_plot <- visreg(tri_model,
  xvar = "hinctnta",
  gg = T,
  partial = F,
  rug = F) +
  scale_x_continuous(breaks = seq(from = 1, to = 10, by = 1))+
  scale_y_continuous(breaks = seq(from = 0, to = 10, by = 1))+
  coord_cartesian(ylim = c(0, 10))+
  labs(x = "Einkommen(Dezil)",
  y = "Migrationswertschätzung",
  title = "Conditional Effect Plot: Einkommen und Migrationswertschätzung",
  subtitle = "Vorhergesagte Werte unter Konstanthaltung der Bildung im Mittelwert",
  caption = "Daten: ESS (2016), Teilstichprobe CH (N= 1242),
  'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'") +
  theme_bw()
```

Conditional Effect Plot: Einkommen und Migrationswertschätzung
Vorhergesagte Werte unter Konstanthaltung der Bildung im Mittelwert



Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

```
tri_plot <- visreg(tri_model,  
  xvar = "hinctnta",  
  gg = T,  
  partial = F,  
  rug = F)
```

Regressionsergebnis, das visualisiert werden soll

Variable, deren Effekt abgebildet werden soll

Plot soll ein ggplot-Objekt sein

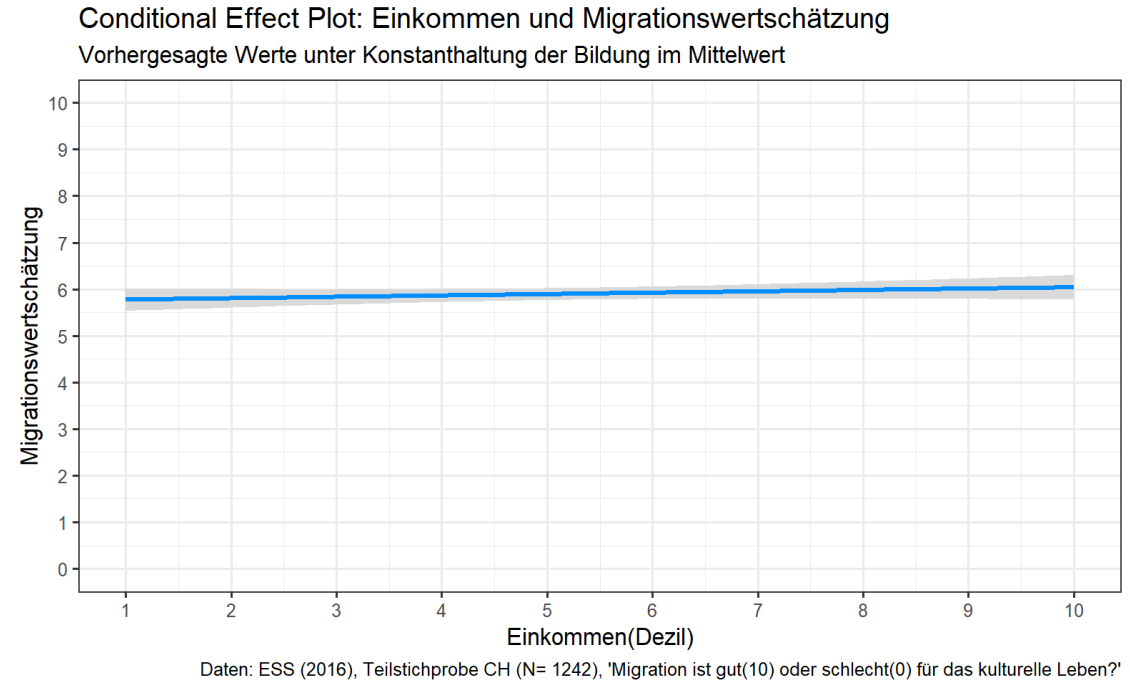
Punktewolke nicht mit abbilden

7

Visualisierung: Conditional Effect Plot

Um *Conditional Effect Plots* zu erstellen, verwenden wir die Funktion `visreg::visreg()`

```
tri_plot <- visreg(tri_model,
  xvar = "hinctnta",
  gg = T,
  partial = F,
  rug = F) +
  scale_x_continuous(breaks = seq(from = 1, to = 10, by = 1))+
  scale_y_continuous(breaks = seq(from = 0, to = 10, by = 1))+
  coord_cartesian(ylim = c(0, 10))+
  labs(x = "Einkommen(Dezil)",
  y = "Migrationswertschätzung",
  title = "Conditional Effect Plot: Einkommen und Migrationswertschätzung",
  subtitle = "Vorhergesagte Werte unter Konstanthaltung der Bildung im Mittelwert",
  caption = "Daten: ESS (2016), Teilstichprobe CH (N= 1242),
  'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'") +
  theme_bw()
```



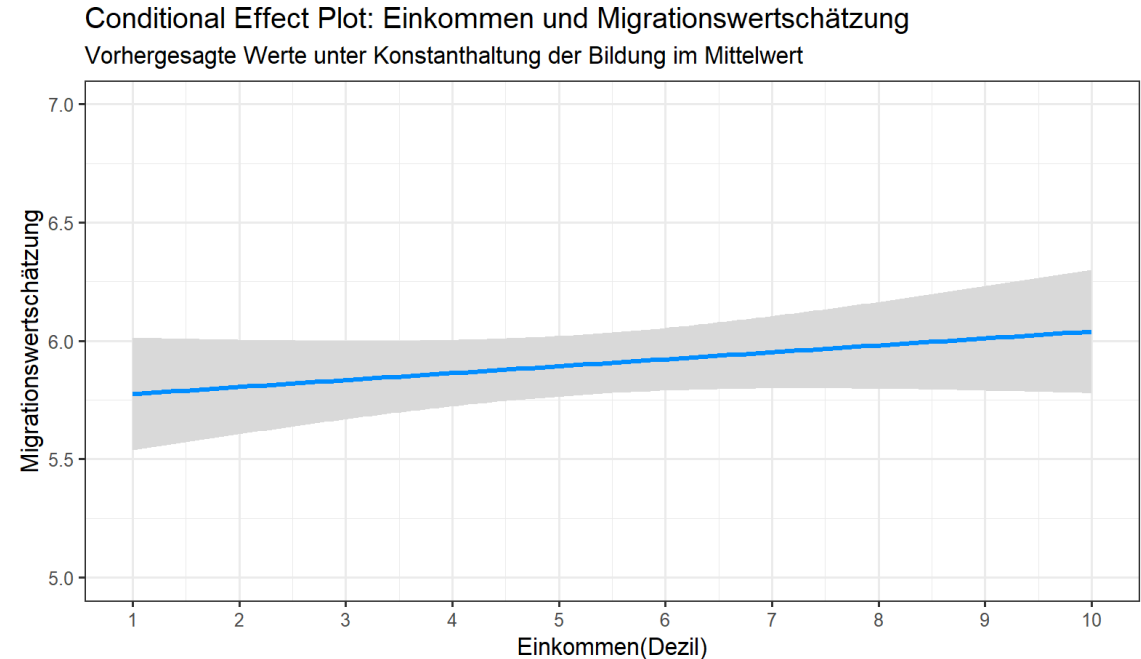
Wir haben nun den bereinigten Effekt des Einkommens auf die Migrationswertschätzung visualisiert

7

Visualisierung: Conditional Effect Plot

Unsere (!) Konvention zur «sauberen» Darstellung der Effektgrösse: Die Y-Achse umfasst grundsätzlich etwa eine Standardabweichung der AV

```
> mean(ess8_noNA$imueclt, na.rm = T)
[1] 6.161031
> sd(ess8_noNA$imueclt, na.rm = T)
[1] 2.225374
```



Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

```
tri_plot<- visreg(tri_model,
  xvar = "hinctnta",
  gg = T,
  partial = F,
  rug = F) +
  scale_x_continuous(breaks = seq(from = 1, to = 10, by = 1))+
  scale_y_continuous(breaks = seq(from = 0, to = 10, by = 0.5))+
  coord_cartesian(ylim = c(5, 7))+
  labs(x = "Einkommen(Dezil)",
  y = "Migrationswertschätzung",
  title = "Conditional Effect Plot: Einkommen und Migrationswertschätzung",
  subtitle = "Vorhergesagte Werte unter Konstanthaltung der Bildung im Mittelwert",
  caption = "Daten: ESS (2016), Teilstichprobe CH (N= 1242),
  'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'") +
  theme_bw()
```

Die Geradenneigung wird damit in Referenz zu einem «typischen» Unterschied gesetzt und ist somit sinnvoll eingeordnet

Achtung: Beschriftungsdichte der y-Achse ggf. anpassen

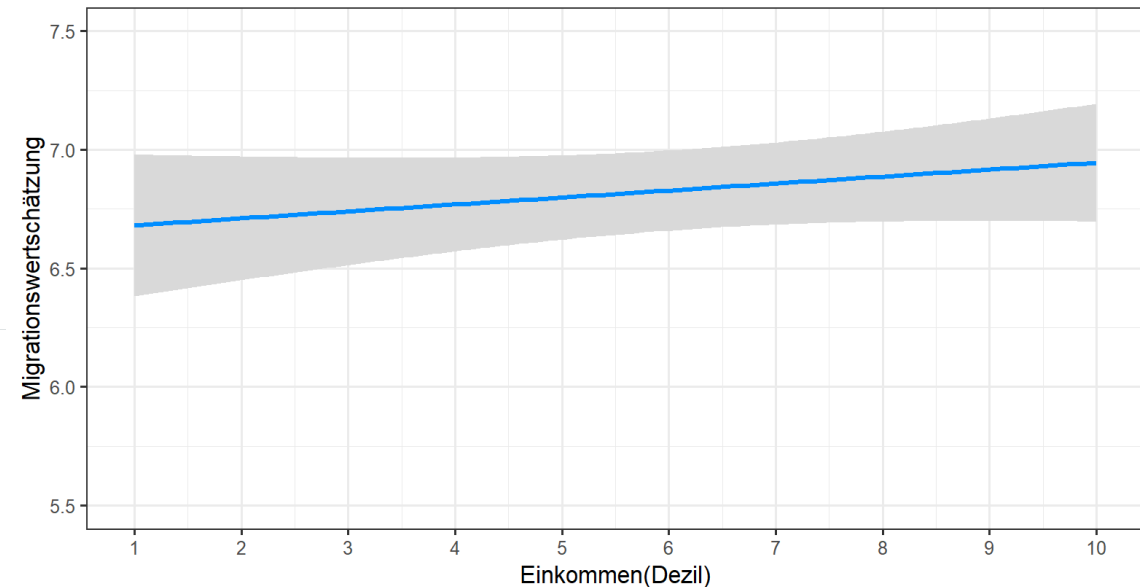
7

Visualisierung: Conditional Effect Plot

Mit dem Teilbefehl **cond()** können wir den Scheibenschnitt bei einem alternativen z-Wert ansetzen (siehe z.B. Folie 56)

```
tri_plot2<- visreg(tri_model,
  cond = list(edu yrs=15),
  xvar = "hinctnta",
  gg = T,
  partial = F,
  rug = F) +
  scale_x_continuous(breaks = seq(from = 1, to = 10, by = 1))+
  scale_y_continuous(breaks = seq(from = 0, to = 10, by = 0.5))+
  coord_cartesian(ylim = c(5.5, 7.5))+
  labs(x = "Einkommen(Dezil)",
    y = "Migrationswertschätzung",
    title = "Conditional Effect Plot: Einkommen und Migrationswertschätzung",
    subtitle = "Vorhergesagte Werte für eine Person mit 12 Bildungsjahren",
    caption = "Daten: ESS (2016), Teilstichprobe CH (N= 1242),
    'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'") +
  theme_bw()
```

Conditional Effect Plot: Einkommen und Migrationswertschätzung
Vorhergesagte Werte für eine Person mit 12 Bildungsjahren



Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

...Hintergrund: Für die Kommentierung eines Effectplots sind konkrete ganzzahlige Werte manchmal etwas dankbarer als «schiefe» Mittelwerte

7 Visualisierung bei mehr als drei Variablen

Auch bei mehr als einer Kontrollvariablen nutzen wir den Conditional Effect Plot zur Visualisierung. Visualisiert wird hier der bereinigte Effekt des Einkommens auf die Migrationswertschätzung unter Konstanthaltung der Bildung *und* des Alters.

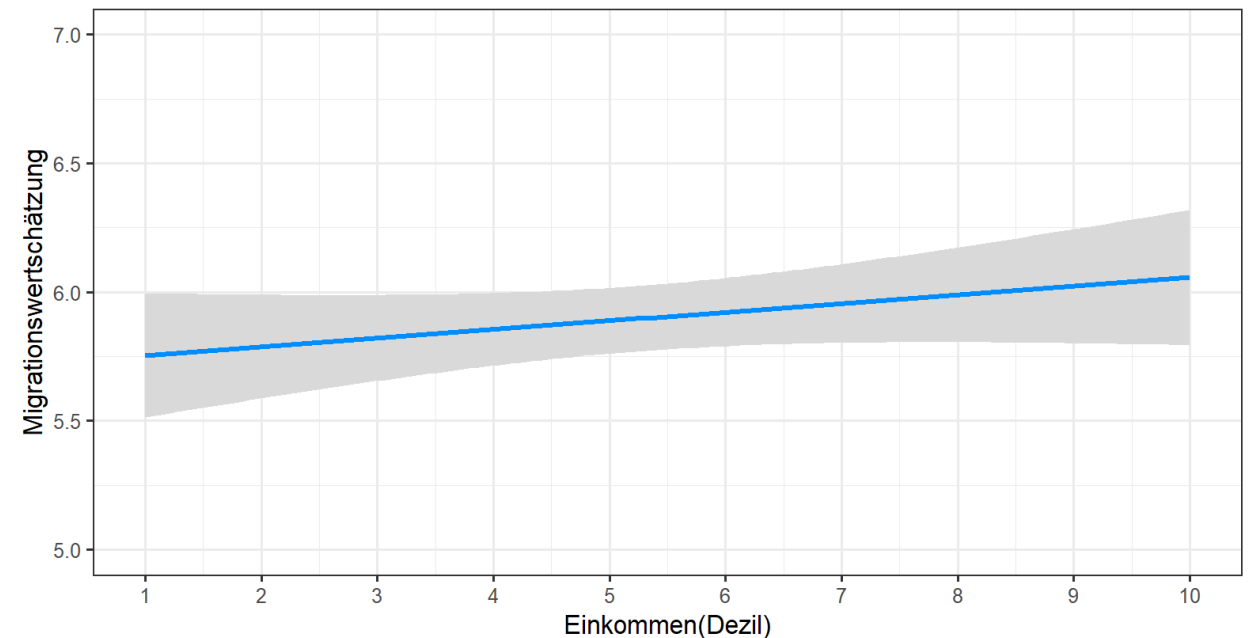
Determinanten der Migrationswertschätzung

	Bivariates Modell (1)	Nettomodell (2)
Einkommen	0.09*** (0.02)	0.03 (0.02)
Bildungsjahre		0.18*** (0.02)
Alter		0.004 (0.003)
Constant	5.64*** (0.15)	3.70*** (0.30)
Observations	1,242	1,242
R ²	0.01	0.09

Note: *p<0.05; **p<0.01; ***p<0.001

Daten: ESS(2016), Teilstichprobe CH. Standardfehler in Klammern

Conditional Effect Plot: Einkommen und Migrationswertschätzung
Vorhergesagte Werte unter Konstanthaltung von Bildung und Alter im Mittelwert



Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

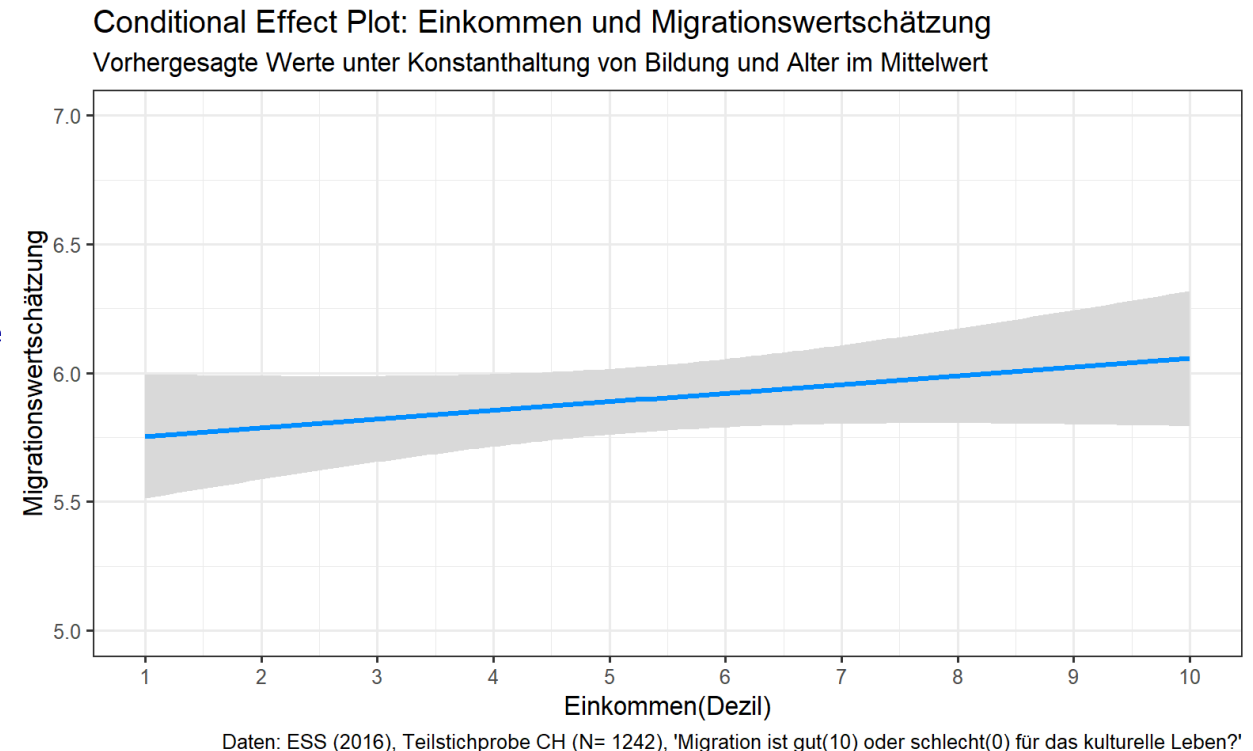
7 Visualisierung bei mehr als drei Variablen

Auch bei mehr als einer Kontrollvariablen nutzen wir den Conditional Effect Plot zur Visualisierung. Visualisiert wird hier der bereinigte Effekt des Einkommens auf die Migrationswertschätzung unter Konstanthaltung der Bildung *und* des Alters.

```
multi_plot<- visreg(multi_model,
                    xvar = "income",
                    gg = T,
                    partial = F,
                    rug = F) +
  scale_x_continuous(breaks = seq(from = 1, to = 10, by = 1))+
  scale_y_continuous(breaks = seq(from = 0, to = 10, by = 0.5))+
  coord_cartesian(ylim = c(5, 7))+
  labs(x = "Einkommen(Dezil)",
       y = "Migrationswertschätzung",
       title = "Conditional Effect Plot: Einkommen und Migrationswertschätzung",
       subtitle = "Vorhergesagte Werte unter Konstanthaltung von Bildung und Alte",
       caption = "Daten: ESS (2016), Teilstichprobe CH (N= 1242),
       'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'") +
  theme_bw()
```

...natürlich muss der Ergebnisbezug in *visreg* entsprechend modifiziert werden

...auch hier nimmt R den 2D-Schnitt durch den Regressionskörper automatisch bei den jeweiligen Mittelwerten unserer Drittvariablen vor

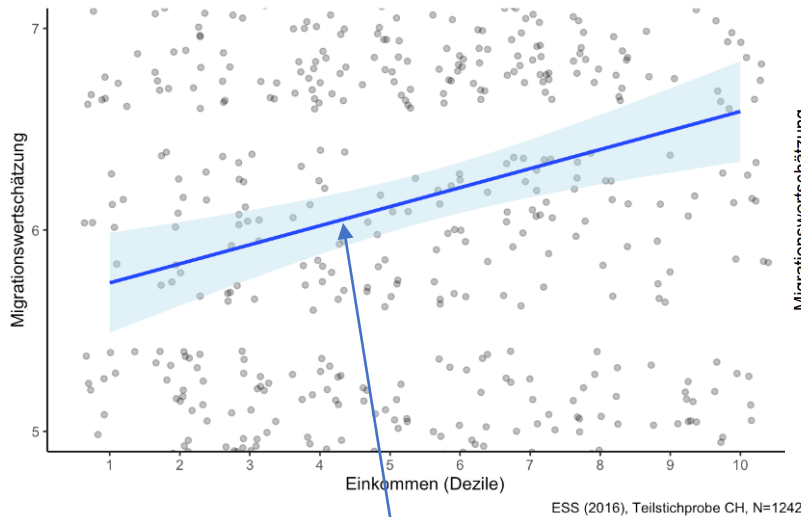


Multivariate Visualisierung und Regressionsgleichung

In allen Szenarien – bivariat, trivariat, quadrivariat,... - greift der abgebildete Ebenenausschnitt immer den Koeffizienten der UV aus der jeweiligen Regressionsgleichung auf

Beispiel Bivariat

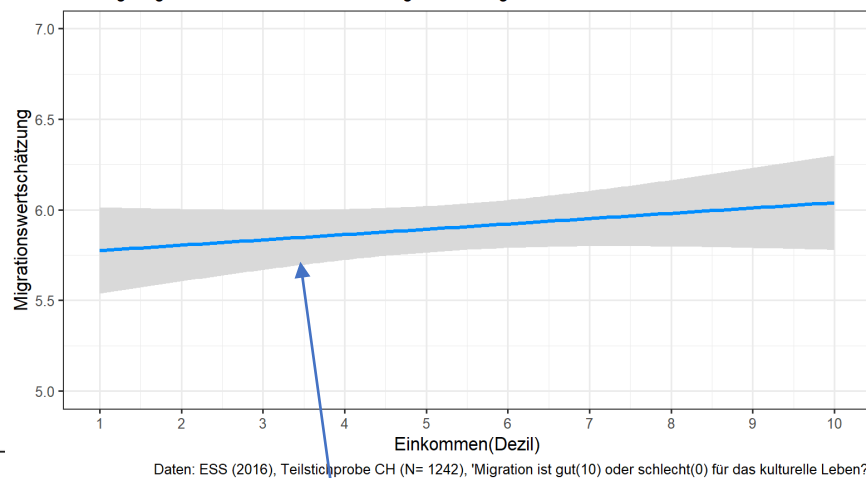
Einkommen und Migrationswertschätzung
Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?



$$y = 5.64 + 0.09 * \text{Einkommen}$$

Beispiel Trivariat

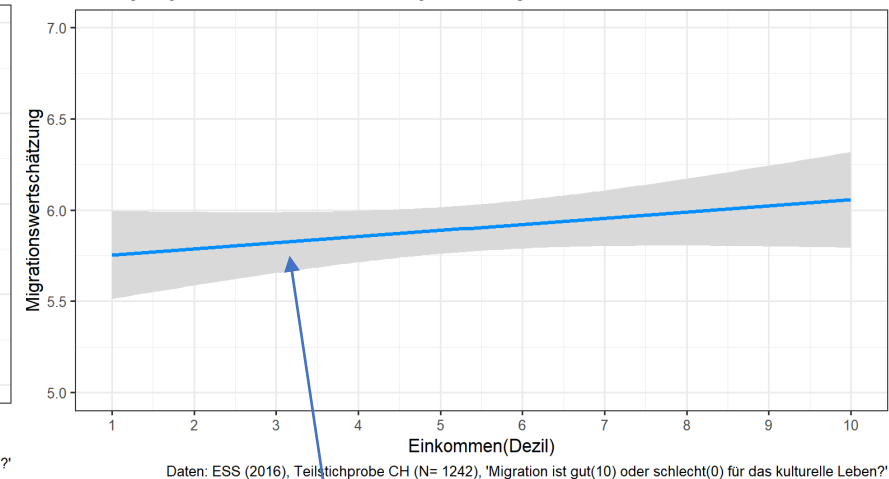
Conditional Effect Plot: Einkommen und Migrationswertschätzung
Vorhergesagte Werte unter Konstanthaltung der Bildung im Mittelwert



$$y = 3.93 + 0.03 * \text{Einkommen} + 0.18 * \text{Bildung}$$

Beispiel Quadrivariat

Conditional Effect Plot: Einkommen und Migrationswertschätzung
Vorhergesagte Werte unter Konstanthaltung von Bildung und Alter im Mittelwert



$$y = 3.70 + 0.03 * \text{Einkommen} + 0.18 * \text{Bildung} + 0.004 * \text{Alter}$$

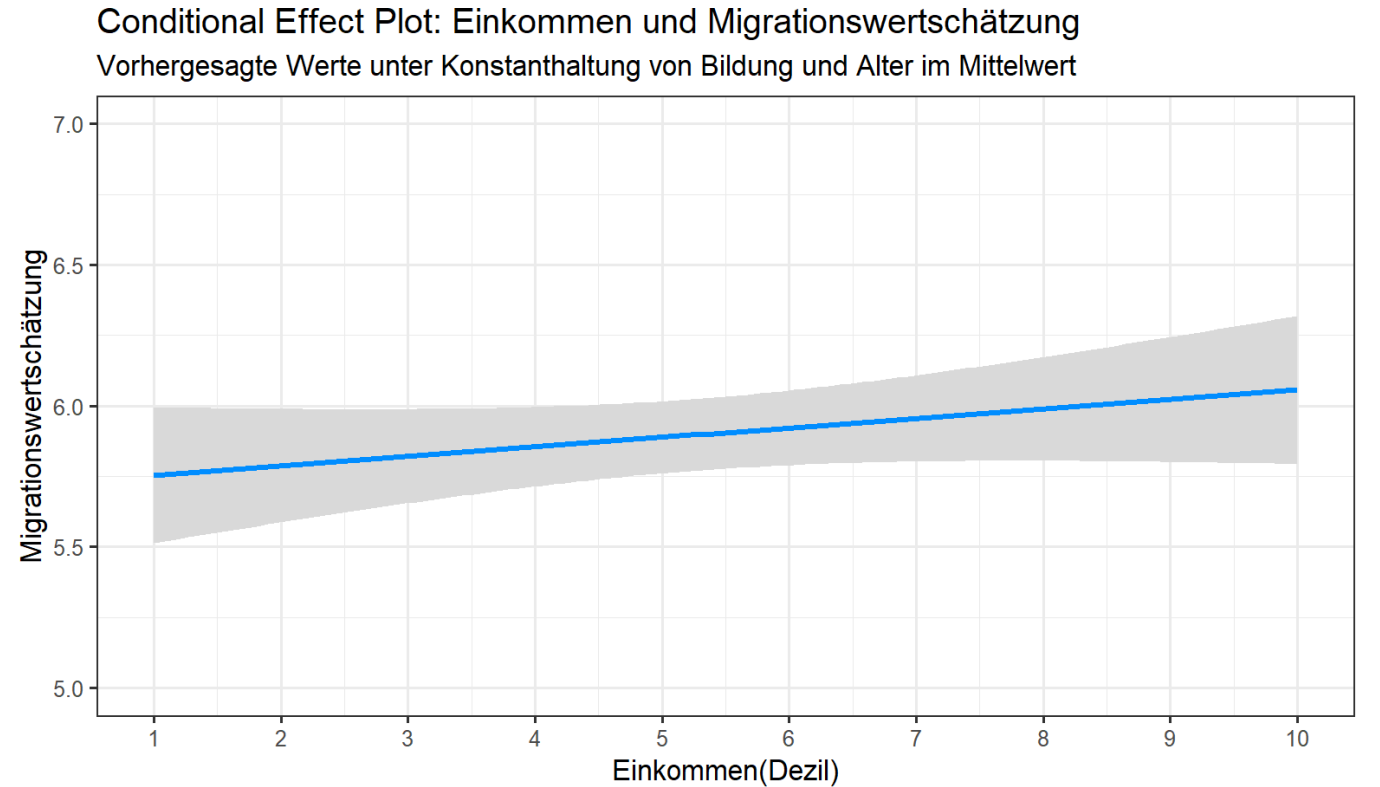
7

Anpassen von Plots – finale Details

Unser Plot ist so schon fast publikationswürdig aber noch nicht perfekt formatiert. Probleme?

Zu kleine Beschriftungen und zu geringe Auflösung!

Aufgabe: Versuche herauszufinden, wie du die Grösse von Beschriftungen ändern kannst



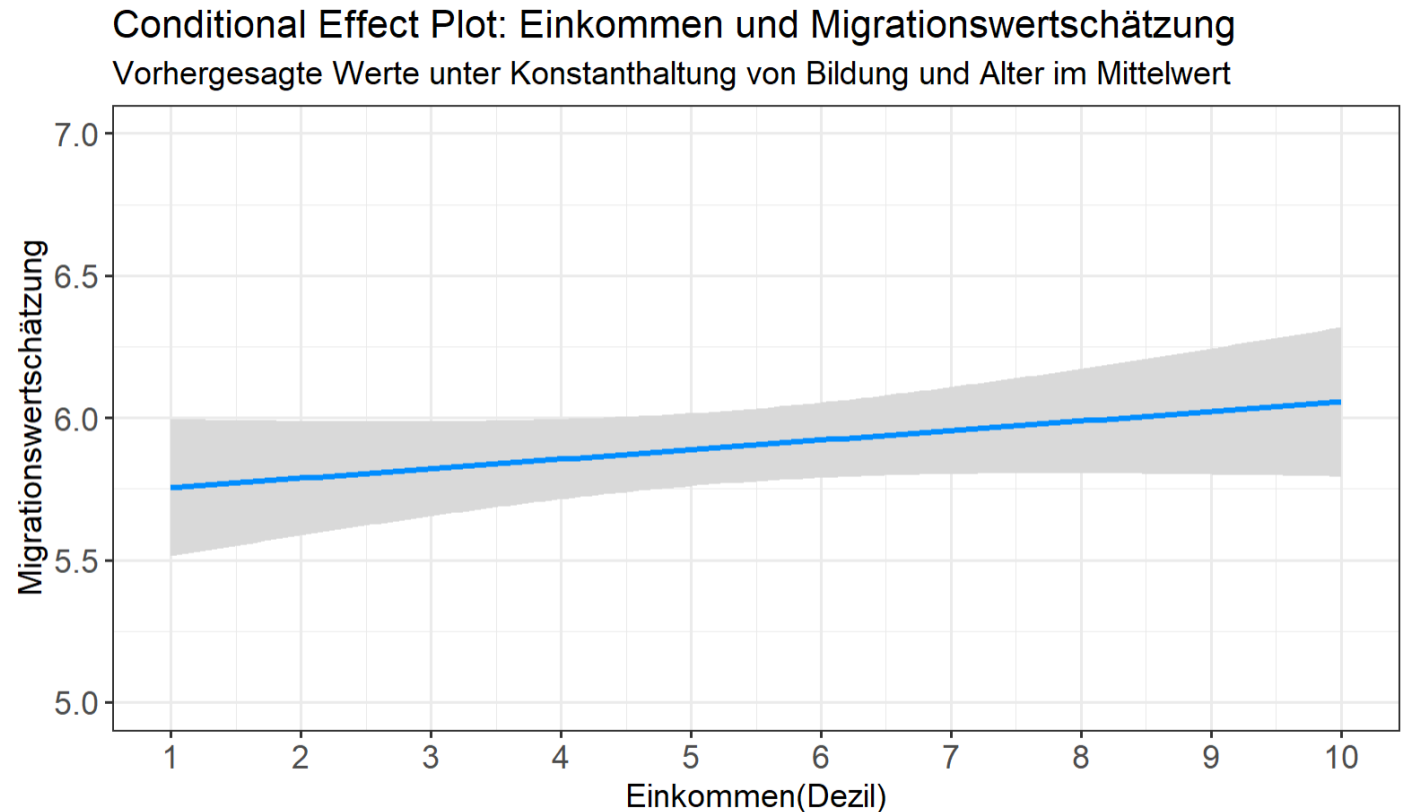
Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

7

Anpassen von Plots – finale Details

Die Grösse von Beschriftungen kann über den Teilbefehl **theme()** angepasst werden.

```
multi_plot +  
  theme(title = element_text(size = 12),  
        axis.text = element_text(size = 12))
```

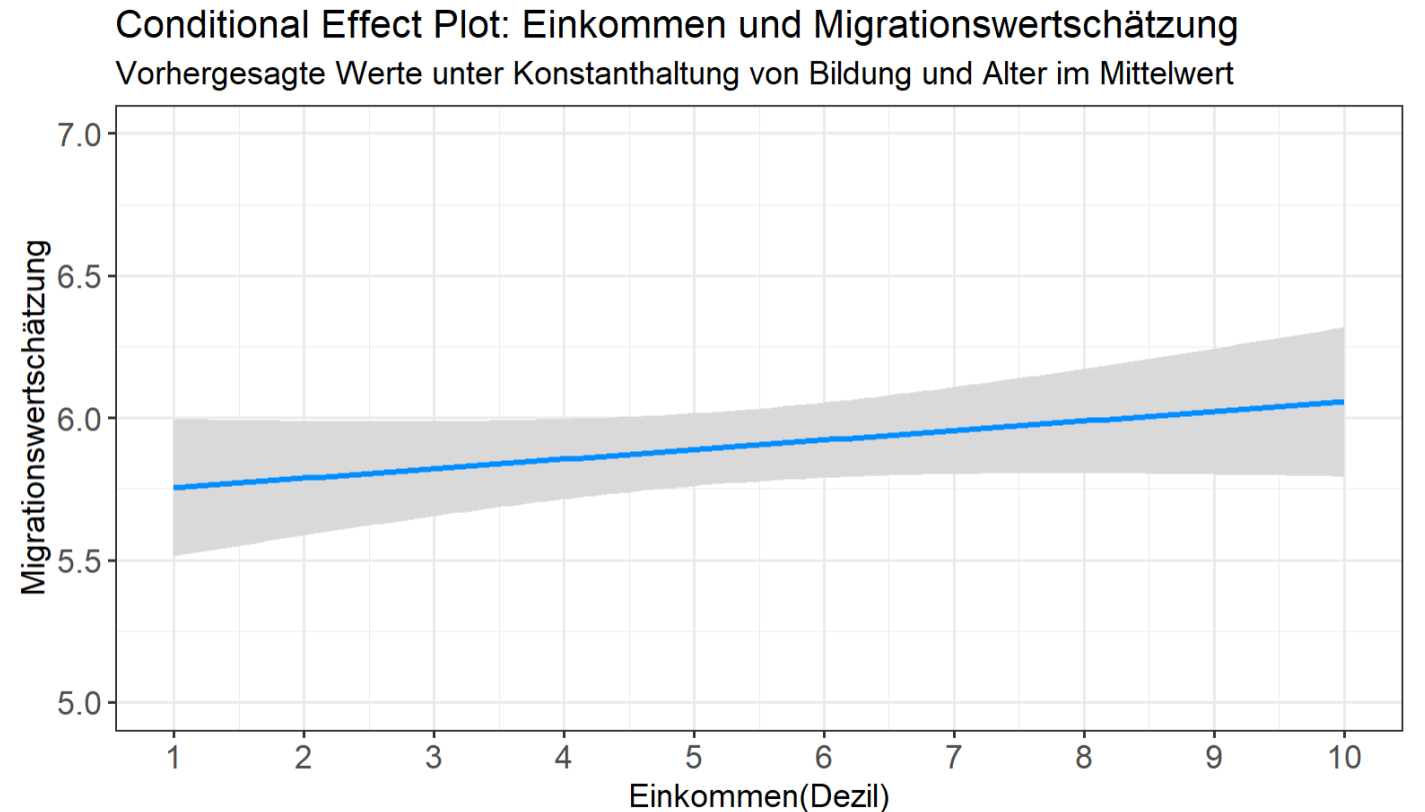


Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

7

Anpassen von Plots – finale Details

Um die Auflösung zu erhöhen, müssen wir unseren Plot über das Skript speichern

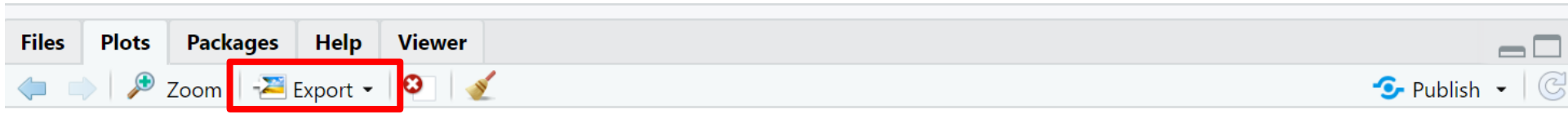


Daten: ESS (2016), Teilstichprobe CH (N= 1242), 'Migration ist gut(10) oder schlecht(0) für das kulturelle Leben?'

7

Bilder Speichern

Bis jetzt haben wir Plots über die Export Funktion abgespeichert



Alternativ können wir aber auch über unsere Syntax die Plots in unsere working directory abspeichern. Dies hat unter anderem den Vorteil, dass wir (und andere Personen) die abgespeicherte Abbildung eins zu eins zu einem späteren Zeitpunkt reproduzieren können. Zusätzlich können wir hier die Auflösung anpassen

```
jpeg(filename = "plot.jpeg", width = 42, height = 30, units = "cm", res = 1200)  
multi_plot  
dev.off()
```

1. Spezifizieren des Datentyps, des Namens, der Grösse und der Auflösung

2. Spezifizieren welcher Plot gespeichert werden soll

3. Speichervorgang mit dev.off() abschliessen.

Bonus – 3D Plots

Disclaimer: Im folgenden Abschnitt wird das Herstellen von 3D-Plots erklärt. An dieser Stelle wird aber nochmals betont, dass 3D-Plots in der wissenschaftliche Kommunikation nicht gebräuchlich sind und **nicht in eure Arbeiten eingebunden werden sollten**. Allerdings eignen sich 3D-Plots gut für didaktische Zwecke, sowie auch die Vermittlung ausserhalb der Wissenschaft

Wir verwenden das Package rgl, um 3D-Plots anzufertigen.

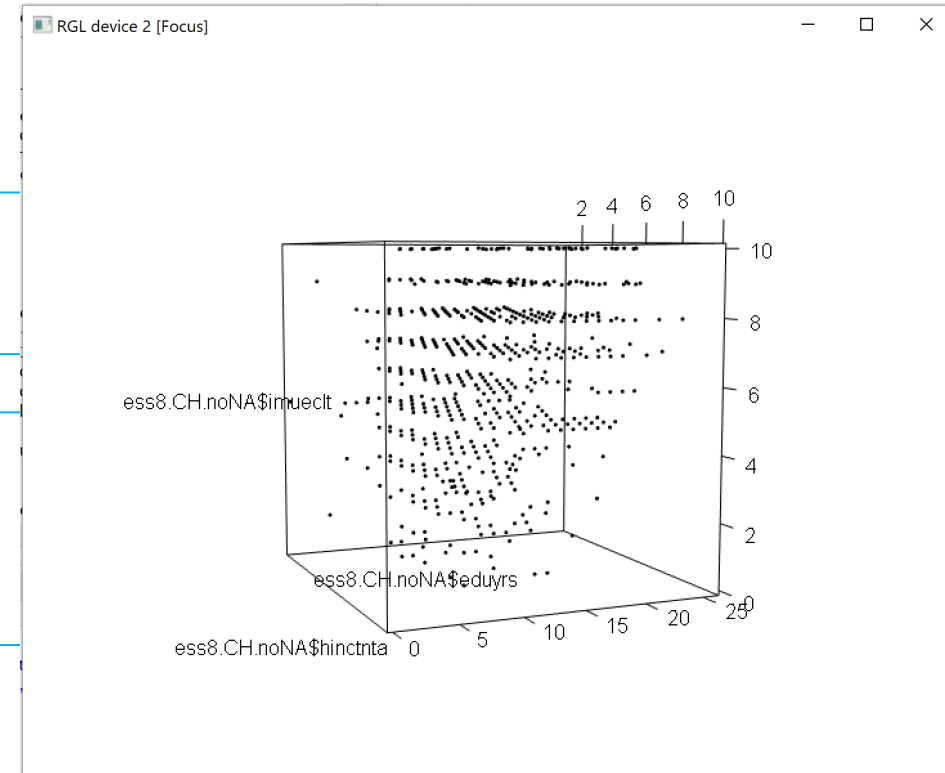
```
library(rgl)
```

Wir brauchen dann den Befehl plot3d, um einen Scatterplot zu machen.

```
plot3d(x=ess8_noNA$hinctnta,  
       z=ess8_noNA$imueclt,  
       y=ess8_noNA$eduyrs)
```

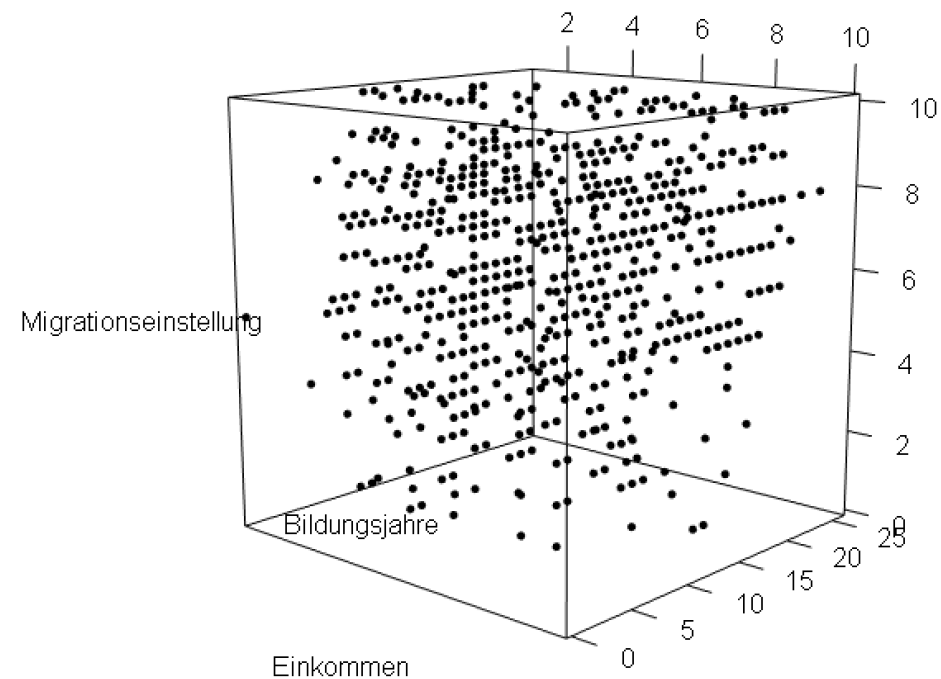
Der plot3d Befehl hat keinen data Teilbefehl, also müssen wir die Achsen immer in folgender Form angeben data\$variable.

Wenn wir denn Befehl ausführe öffnet Rstudio ein neues Fenster mit unserem Scatterplot. Auf diesem Fenster kannst du die Scatterbox, durch «click and drag» drehen.



Wir können mit, xlab, ylab und zlab die Achsenbeschriftungen anpassen und mit size die Grösse der Punkte verändern.

```
plot3d(x=ess8_noNA$hinctnta,  
       z=ess8_noNA$imueclt,  
       y=ess8_noNA$eduyrs,  
       xlab = "Einkommen",  
       ylab = "Bildungsjahre",  
       zlab = "Migrationseinstellung")
```

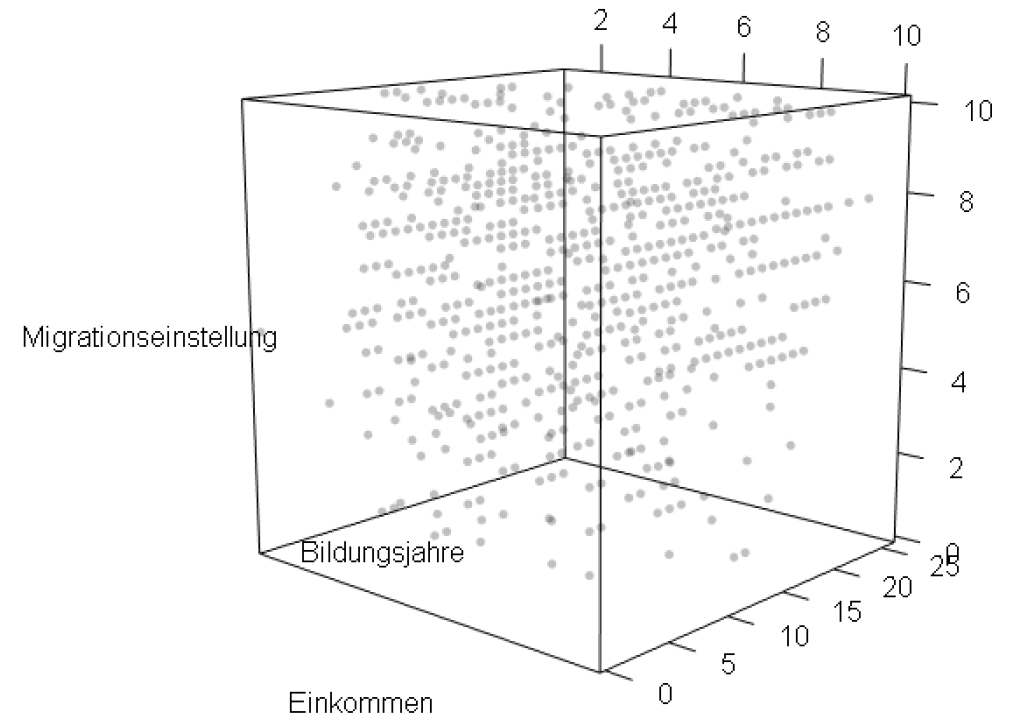


Wir können mit, xlab, ylab und zlab die Achsenbeschriftungen anpassen und mit size die Grösse der Punkte verändern.

```
plot3d(x=ess8_noNA$hinctnta,  
       z=ess8_noNA$imueclt,  
       y=ess8_noNA$eduyrs,  
       xlab = "Einkommen",  
       ylab = "Bildungsjahre",  
       zlab = "Migrationseinstellung")
```

Mit Alpha können wir die Dursichtigkeit der Punkte verändern (0=Ganz durchsichtig, 1=Nicht durchsichtig).

```
plot3d(x=ess8_noNA$hinctnta,  
       z=ess8_noNA$imueclt,  
       y=ess8_noNA$eduyrs,  
       xlab = "Einkommen",  
       ylab = "Bildungsjahre",  
       zlab = "Migrationseinstellung",  
       size =5,  
       alpha = 0.25)
```



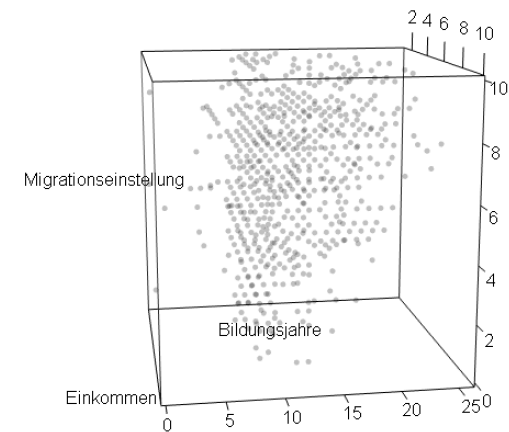
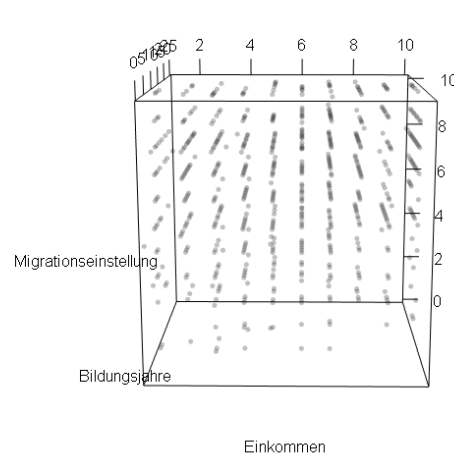
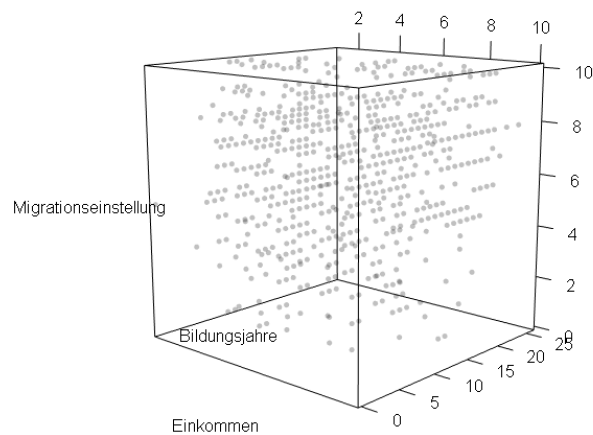
Wir können den Plot mit der Funktion `play3d` automatisch rotieren lassen.

```
play3d(spin3d( axis = c(0, 0, 1), rpm = 2), duration = 30)
```

Mit `spin3d()` geben wir an, dass wir das Ganze rotieren möchten.

Über `axis` geben wir die Rotationsachse an.

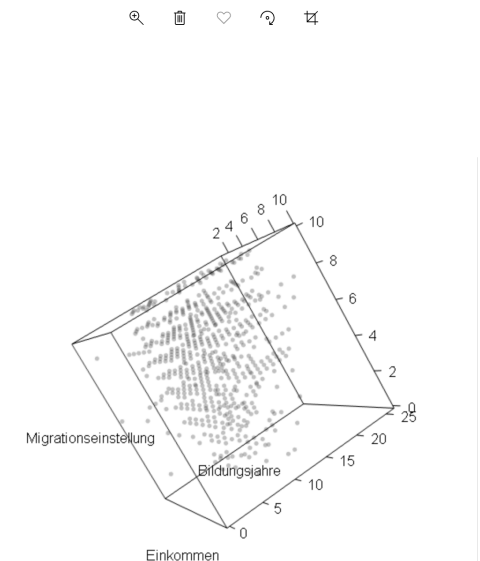
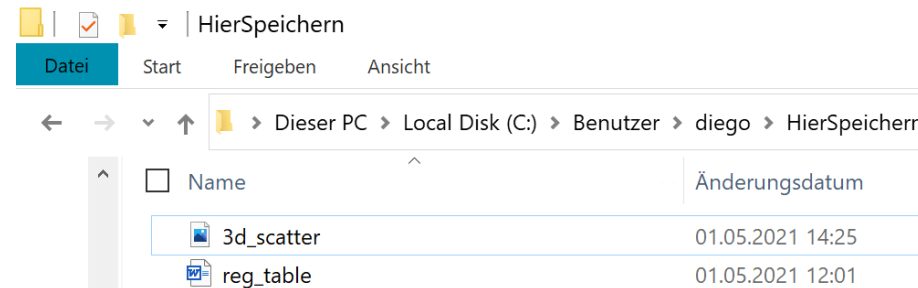
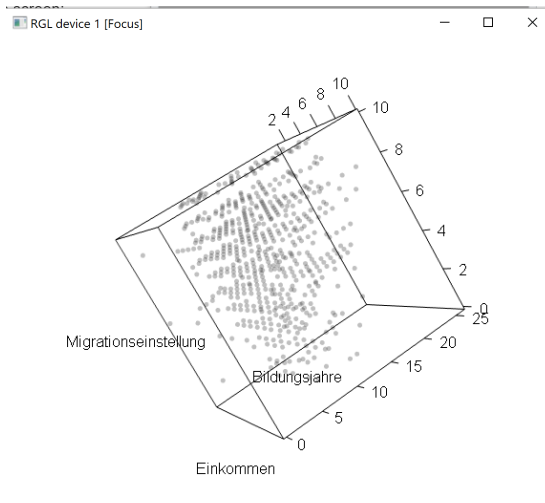
`rpm` steht für rotations per minutes.



Wir können unseren Plot mit dem Befehl `rgl.snapshot()` speichern. Dabei wird jeweils ein Bild vom Plot, wie wir ihn gerade auf dem separaten Fenster sehen gemacht.

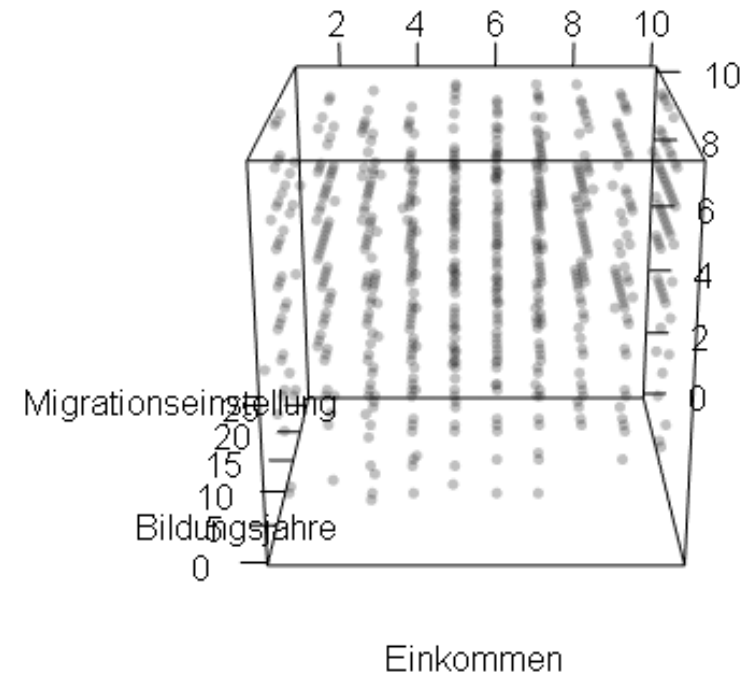
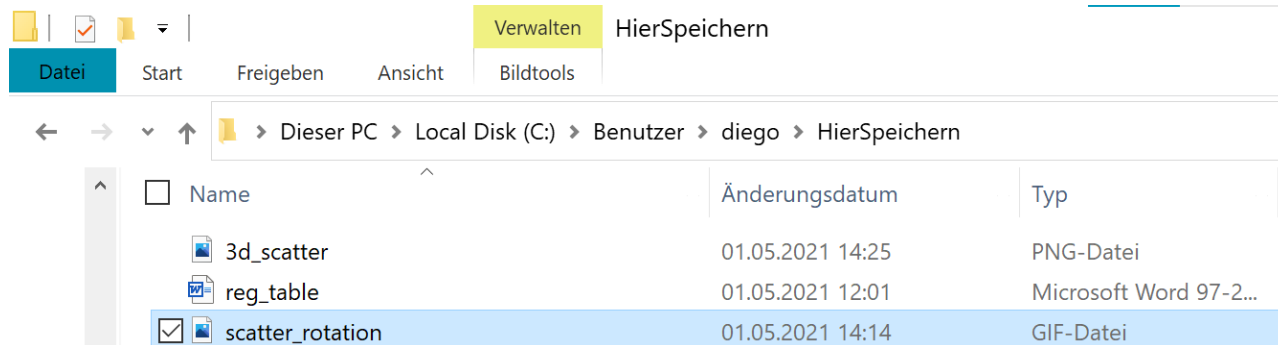
```
rgl.snapshot(filename="3d_scatter.png", fmt = "png")
```

Das Bild wird wieder in unserer working directory gespeichert.



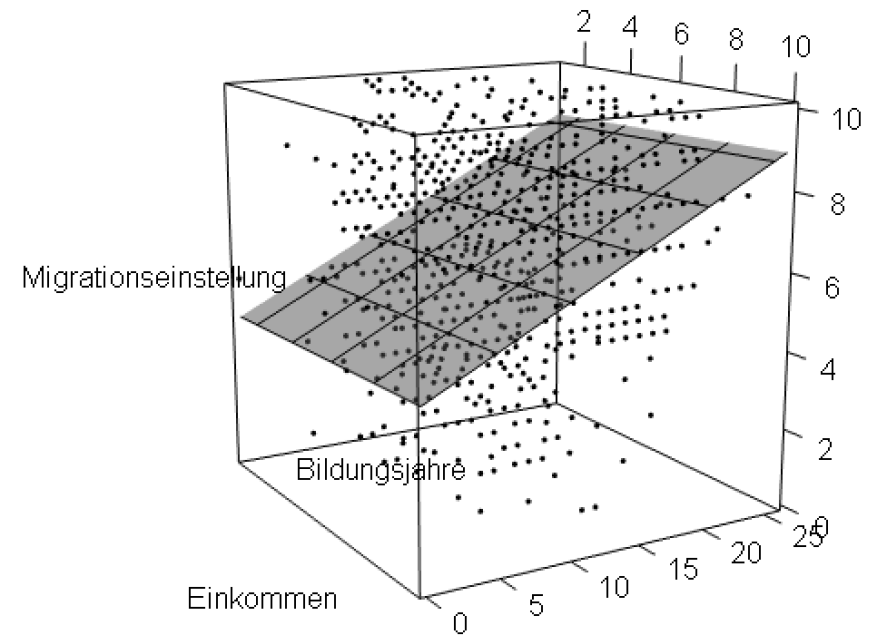
Wir können auch einen gif unseres Plots abspeichern. Hierfür verwenden wir den Befehl `movie3d`. Das gif wird wieder in unserer working directory abgespeichert.

```
movie3d(movie = "scatter_rotation",  
        spin3d( axis = c(0, 0, 1), rpm = 2),  
        duration = 30,  
        dir = getwd(),  
        type = "gif",  
        clean = T)
```



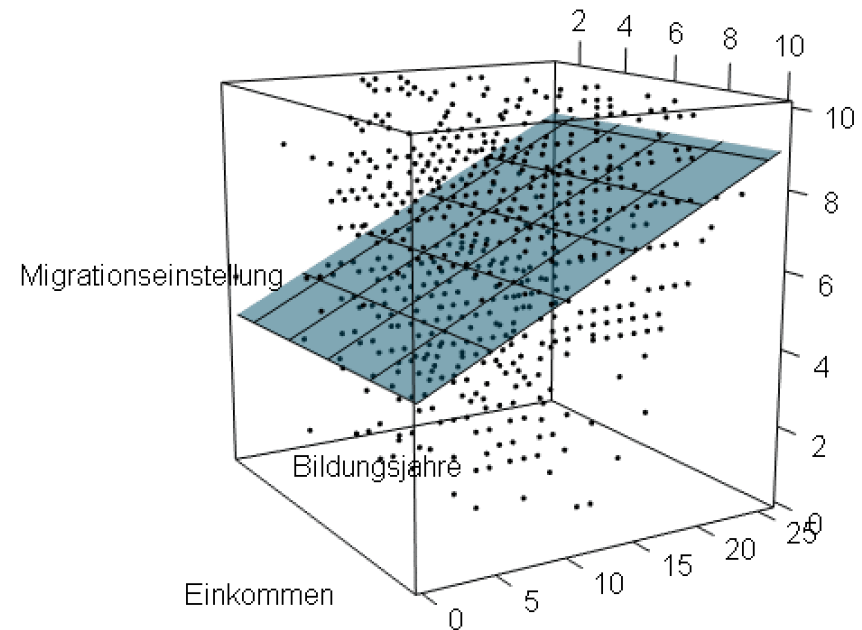
Um die Regressionsebene abzubilden spezifizieren wir im plot3d Befehl, das Modell und geben die Achsen nicht mehr an.

```
plot3d(lm(imueclt ~ hinctnta + eduysr, data = ess8_noNA),  
       xlab = "Einkommen",  
       ylab="Bildungsjahre",  
       zlab = "Migrationseinstellung")
```



Mit `plane.col` können wir die Farbe der Regressionsebene verändern.

```
plot3d(lm(imueclt~hinctnta+eduyrs, data=ess8_noNA),  
       xlab = "Einkommen",  
       ylab="Bildungsjahre",  
       zlab = "Migrationseinstellung",  
       plane.col = "deepskyblue")
```



Das ganze können wir wieder als gif abspeichern.

```
plot3d(lm(imueclt~hinctnta+eduyrs, data=ess8_noNA),  
       xlab = "Einkommen",  
       ylab = "Bildungsjahre",  
       zlab = "Migrationseinstellung",  
       plane.col = "deepskyblue")
```

```
movie3d(movie = "reg_rotation",  
        spin3d( axis = c(0, 0, 1), rpm = 2),  
        duration = 30,  
        dir = getwd(),  
        type = "gif",  
        clean = T)
```

