



The effect of socio-economic indicators on the fertility rates

An empirical analysis of fertility rates among the regions of Europe

Marcell Granát

February 18, 2021

Contents

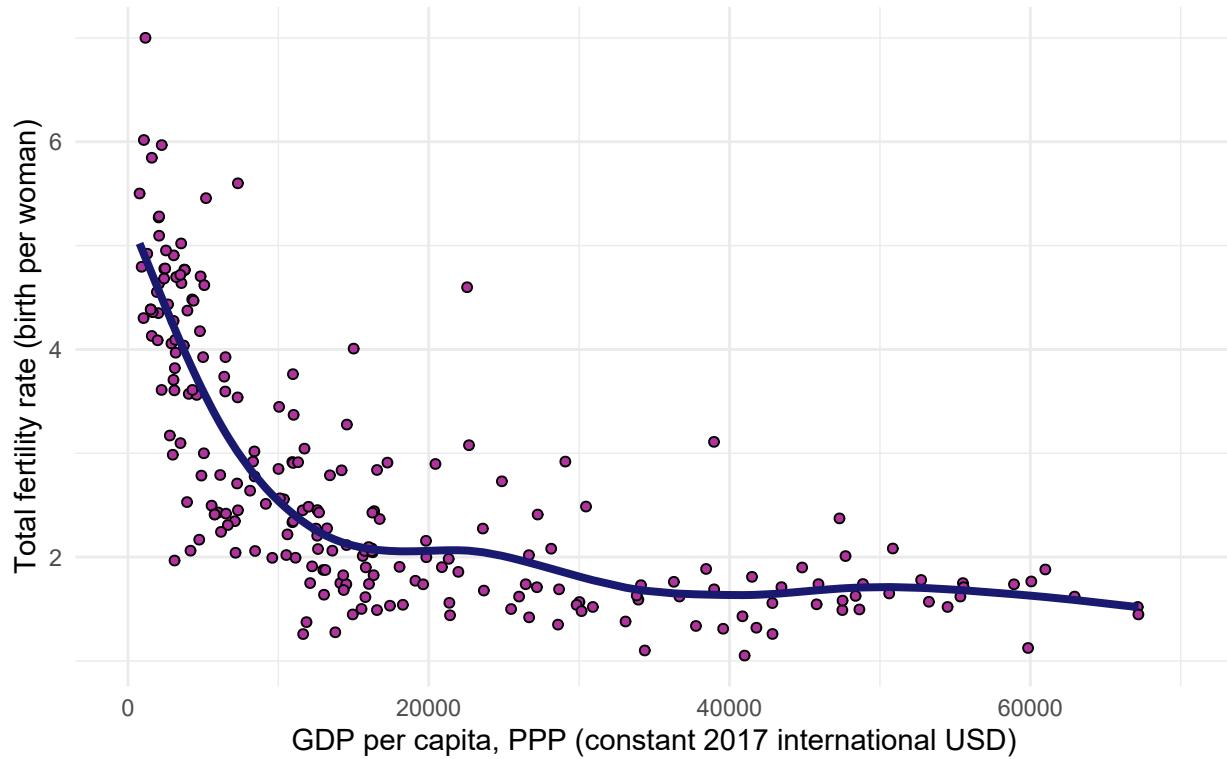
Introduction	3
Data	3
Human development	3
A decent standard of living	4
Long and healthy life	4
Knowledge	5
Explore data	5
Model building	14
Appendix: R codes	16

Abstract

Here is the abstract.

Introduction

Ay alábbi rövid kézirat a brit **Family Expenditure Survey** adataiból kinyert fogyasztási jellemzőket tárgyalja. Az empirikus elemzés egyszerű lineáris regresszió alkalmazásával készül. Az adatok összesen 1519 család jövedelmét és kiadását tartalmazzák, továbbá 6 termékcsoport (alkohol, ruházkodás, nem alkoholtartalmú élelmiszer, fűtés, közlekedés, egyéb) fogyasztáson belüli részarányát. Dolgozatom során én kizárolag az egy gyermekek családokra szűkítem az elemzést¹.



Own editing based on the Figure 5-2. from Kreiszné Hudák (2019).
The trend is drawn via splines.
Source of the data: World Bank.

Figure 1: Seemingly negative effect of gross domestic product on fertility rates based on nation level observations (2017)

Data

Human development

Table 1: Indicators of similarity between the Human Development Indices provided by UNDP and GDI

Indicator	Value
R^2	99.76%
Spearman R^2	99.71%
Mean absolute deviation	0.007
Mean absolute percentage deviation	1.20%

¹Feladatkiírás által előírt megkötés.

A decent standard of living

Table 2: Indicators of similarity between the income component of the Human Development Indices provided by GDL and the estimation based on regional GDP

Indicator	Value
R^2	92.43%
Spearman R^2	94.54%
Mean absolute deviation	0.0474
Mean absolute percentage deviation	6.00%

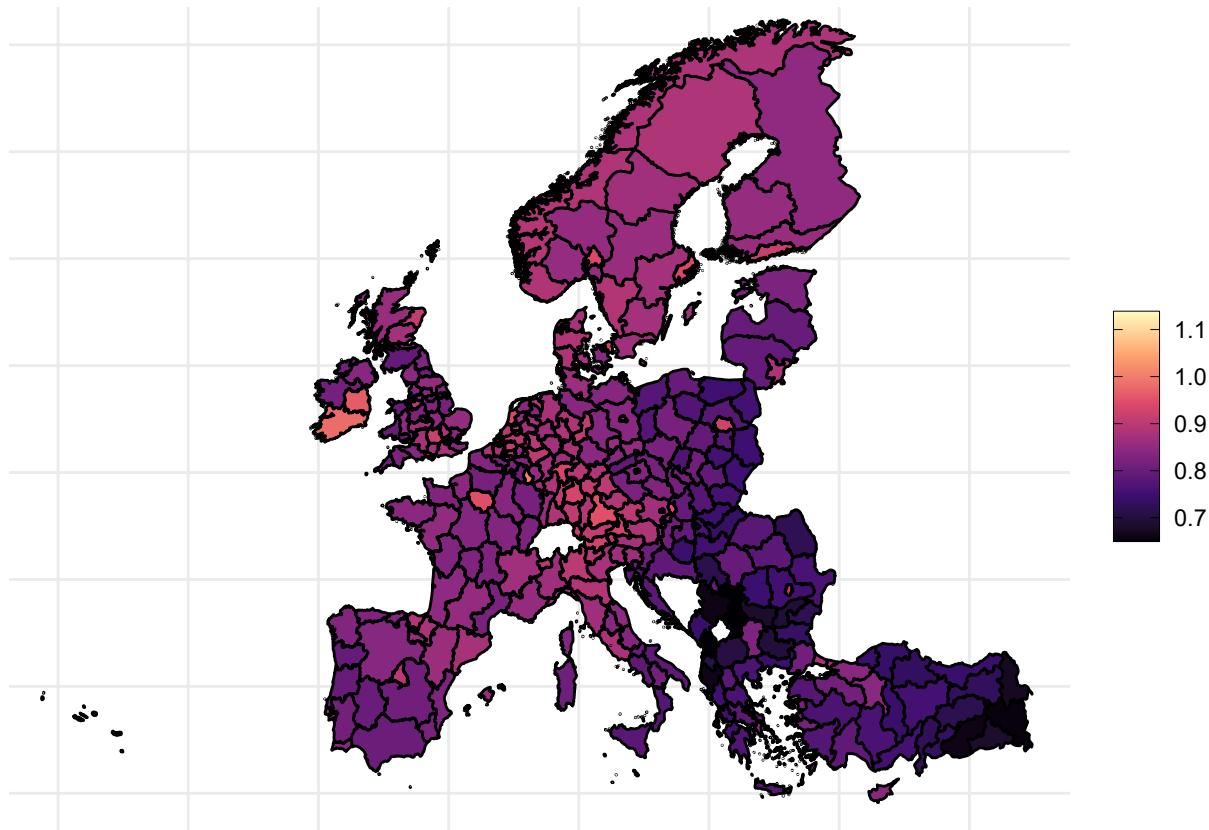


Figure 2: Calculated income index based on regional GDP data

Long and healthy life

Table 3: Indicators of similarity between the health component of the Human Development Indices provided by GDL and the estimation based on regional life expectancy

Indicator	Value
R^2	98.24%

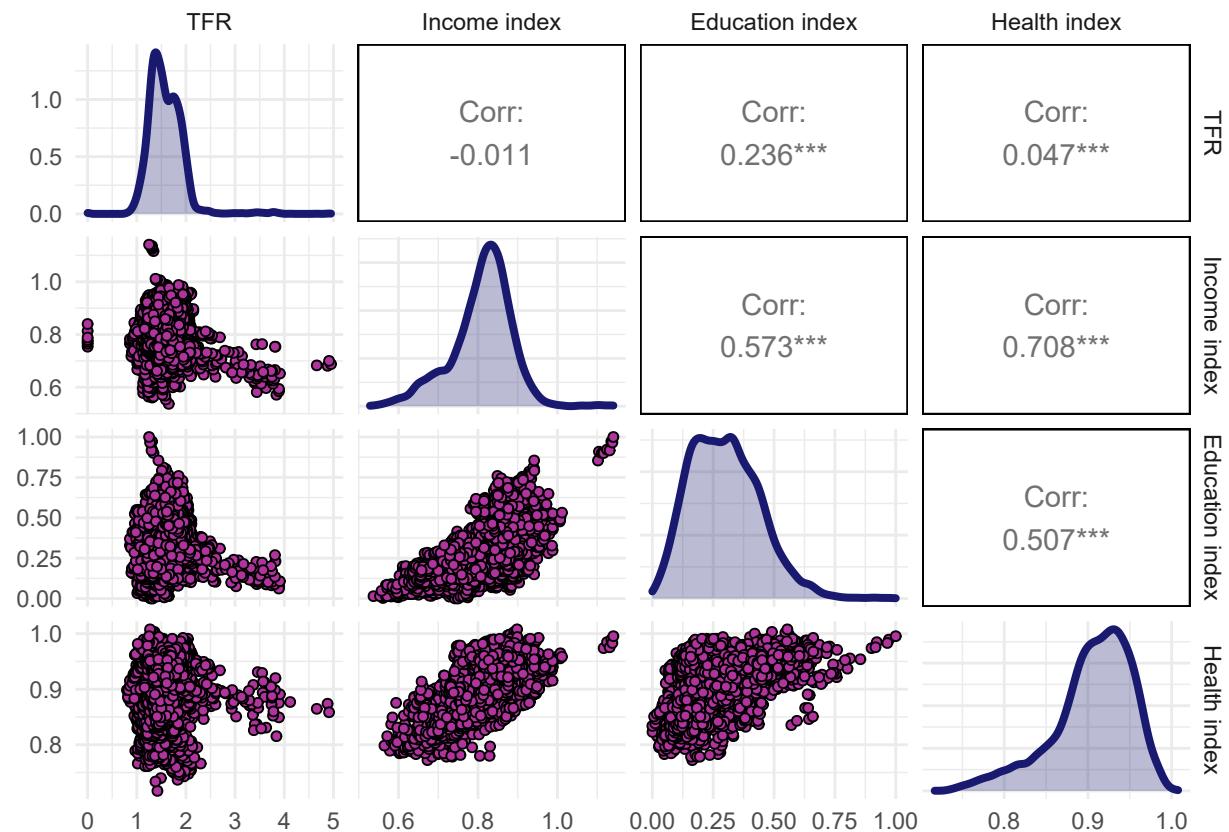
Indicator	Value
Spearman R^2	98.38%
Mean absolute deviation	0.0041
Mean absolute percentage deviation	0.45%

Knowledge

Table 4: Indicators of similarity between the knowledge component of Human Development Indices provided by UNDP and the calculated principal components using educational attainment level

Indicator	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8	Comp 9
R^2	33.41%	23.21%	10.74%	0.04%	0.58%	0.13%	0.00%	0.01%	0.01%
Spearman R^2	21.18%	23.65%	14.09%	0.05%	0.60%	0.46%	13.39%	1.79%	1.81%

Explore data



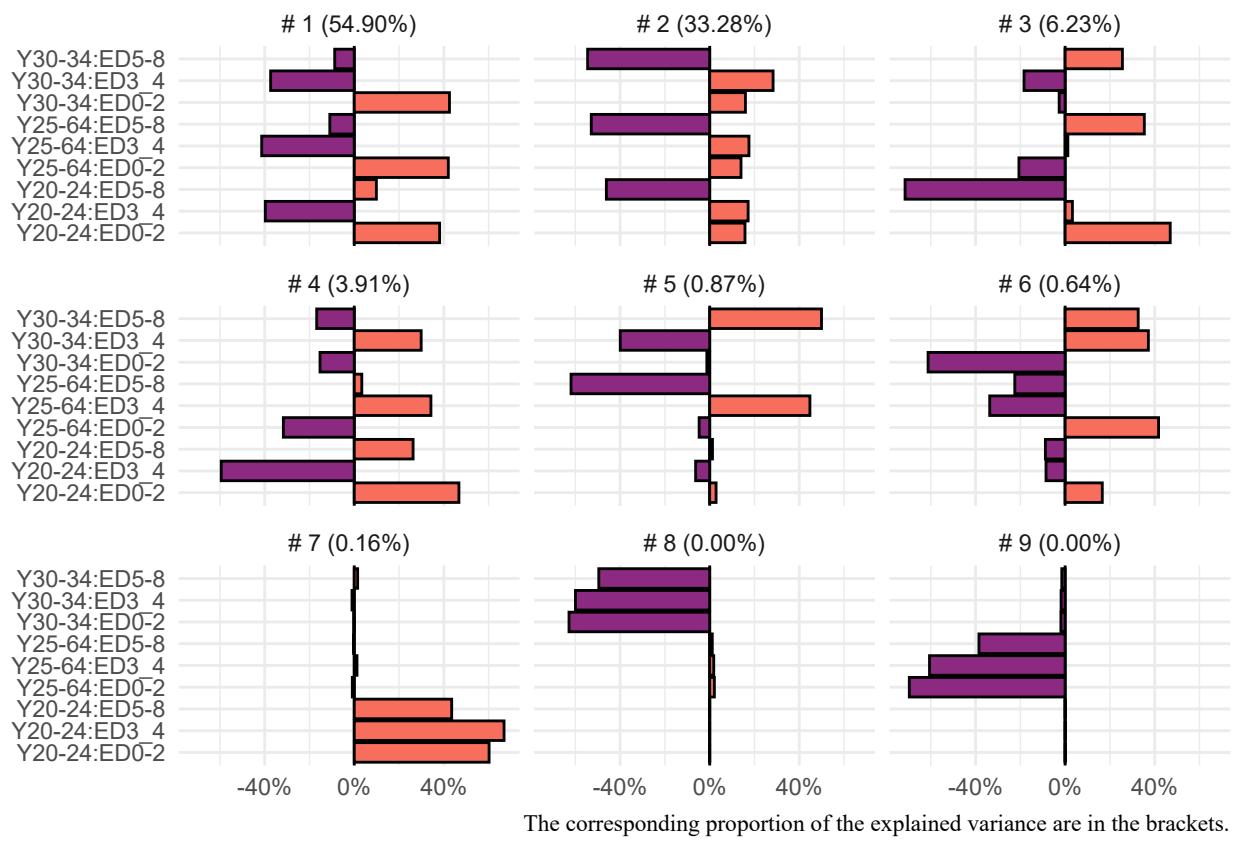
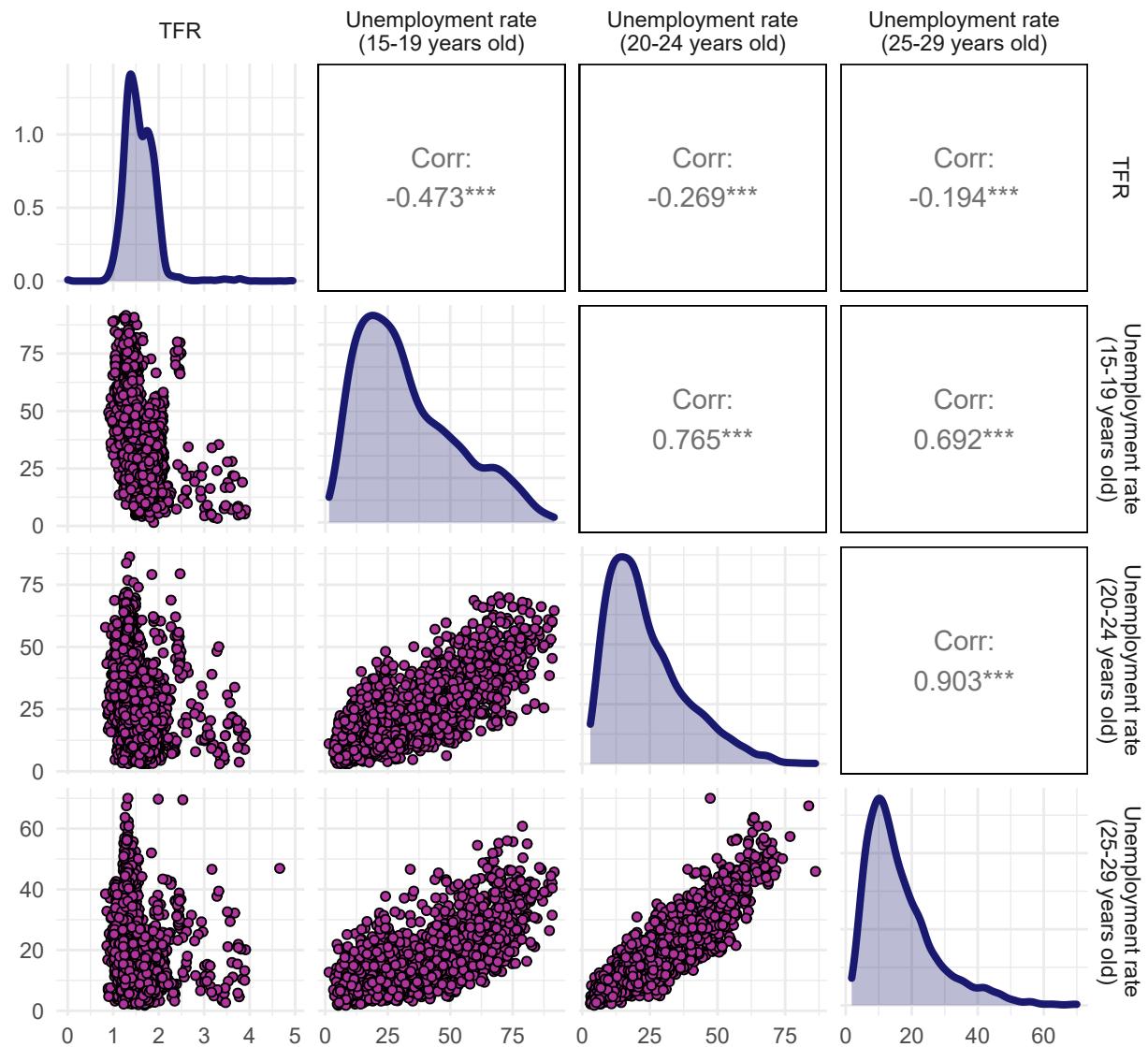
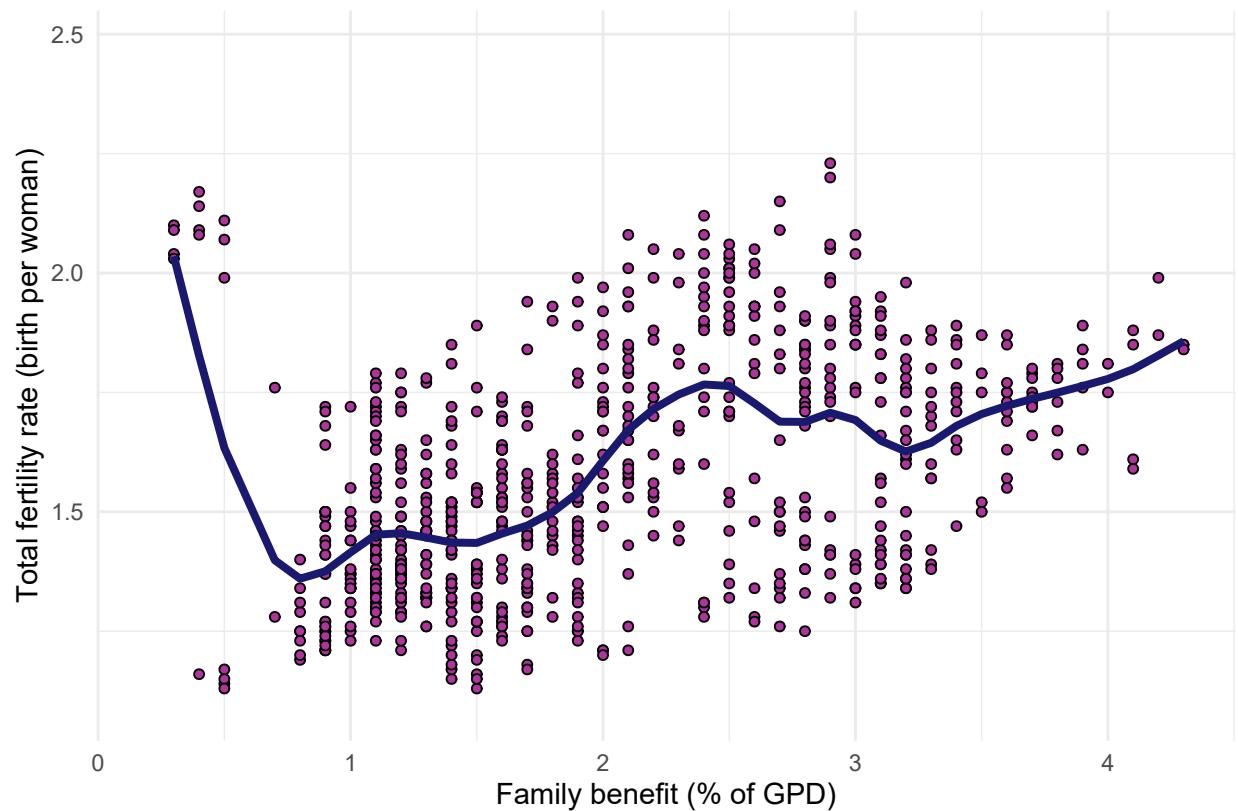
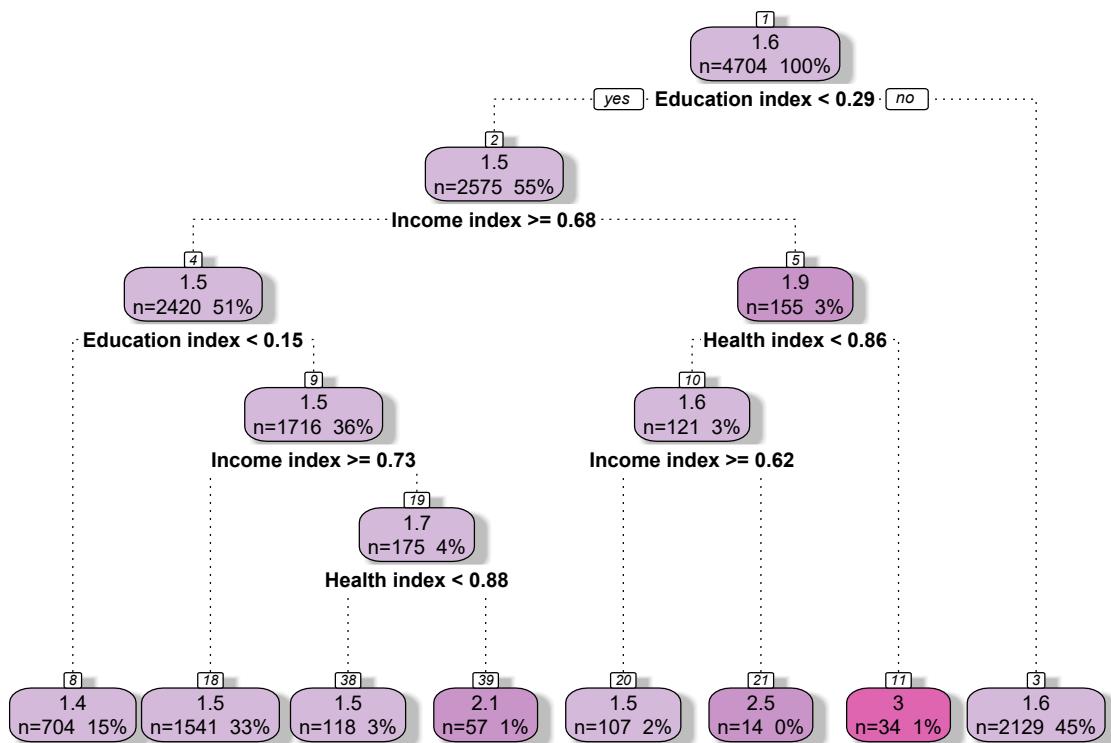
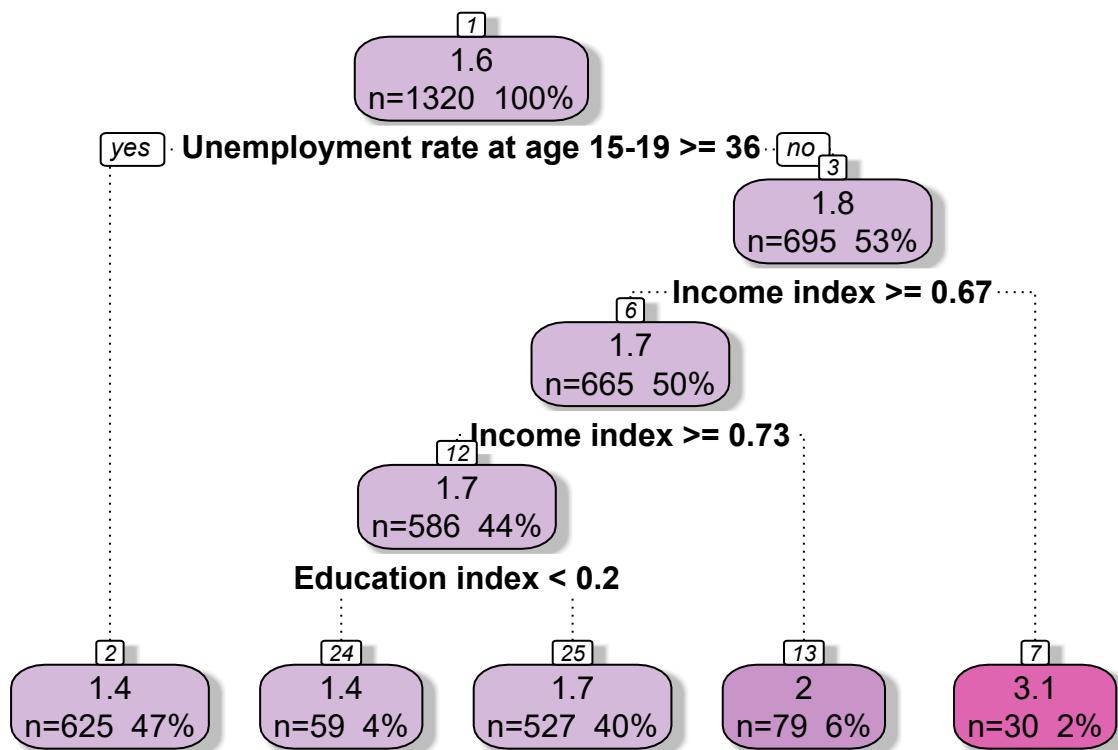


Figure 3: PCAs and the explained variance





Figure 4: Regression tree explaining the TFR ($cp = 0.02$)



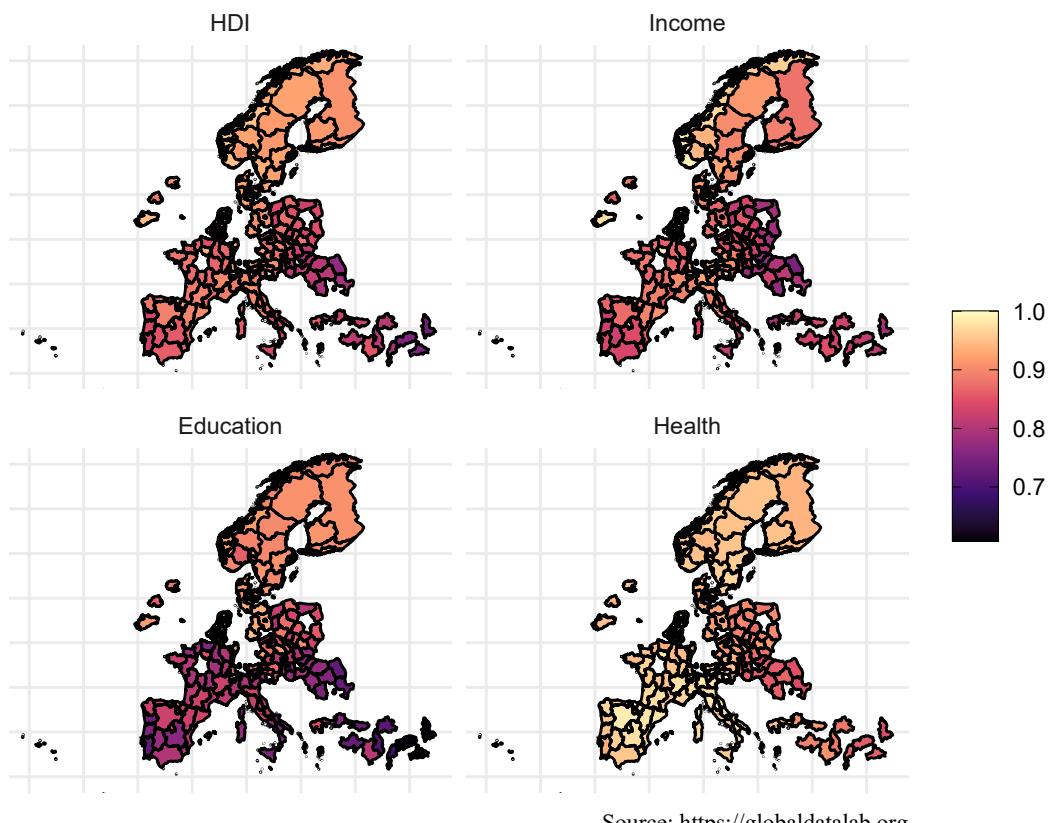


Figure 5: HDI and its components based on the dataset from Global Data Lab (2017)

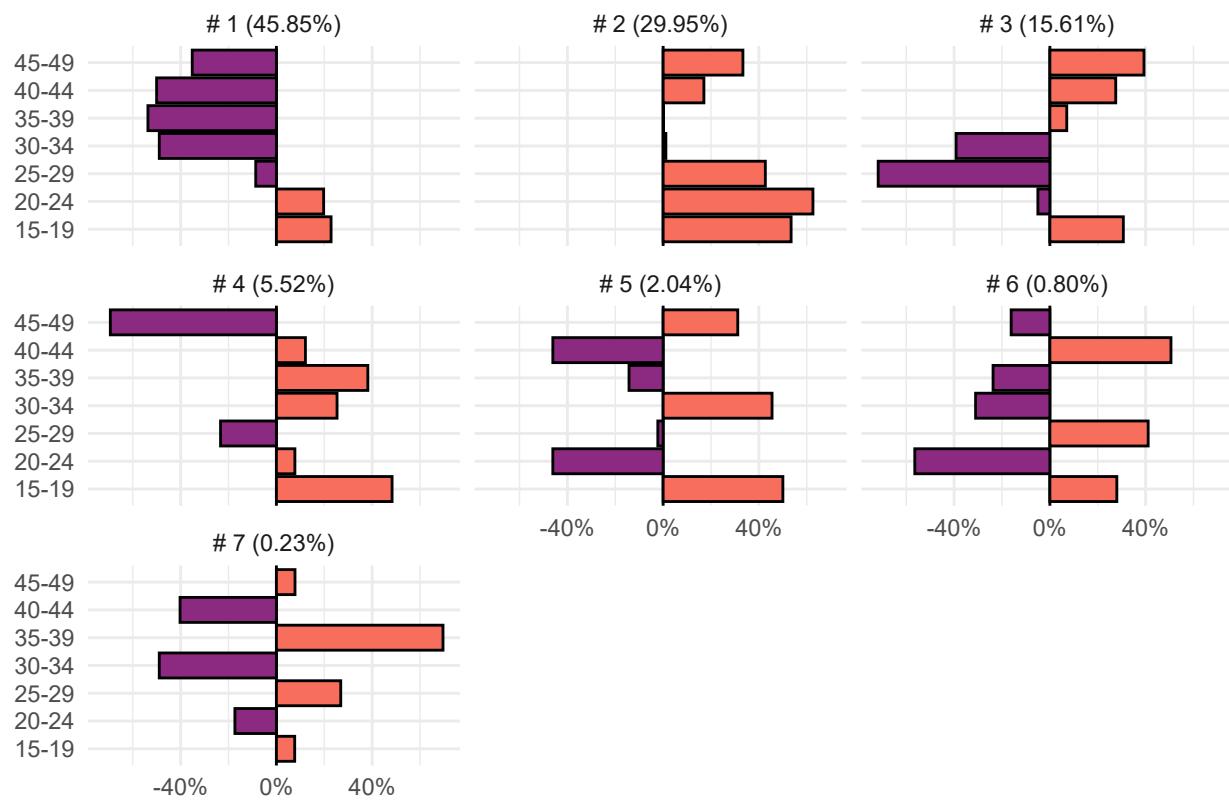


Figure 6: PCAs and the explained variance

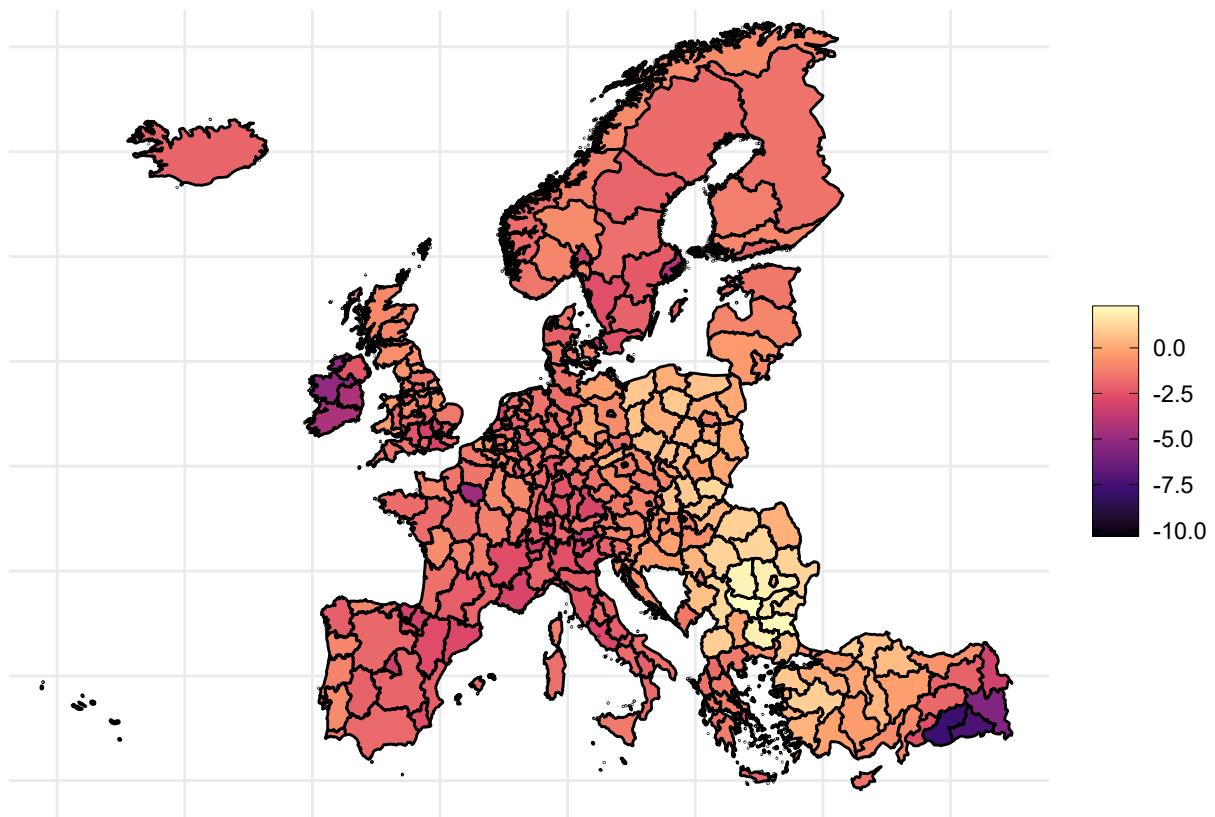


Figure 7: Scores of the 1st PCA component by regions (2017)

Model building

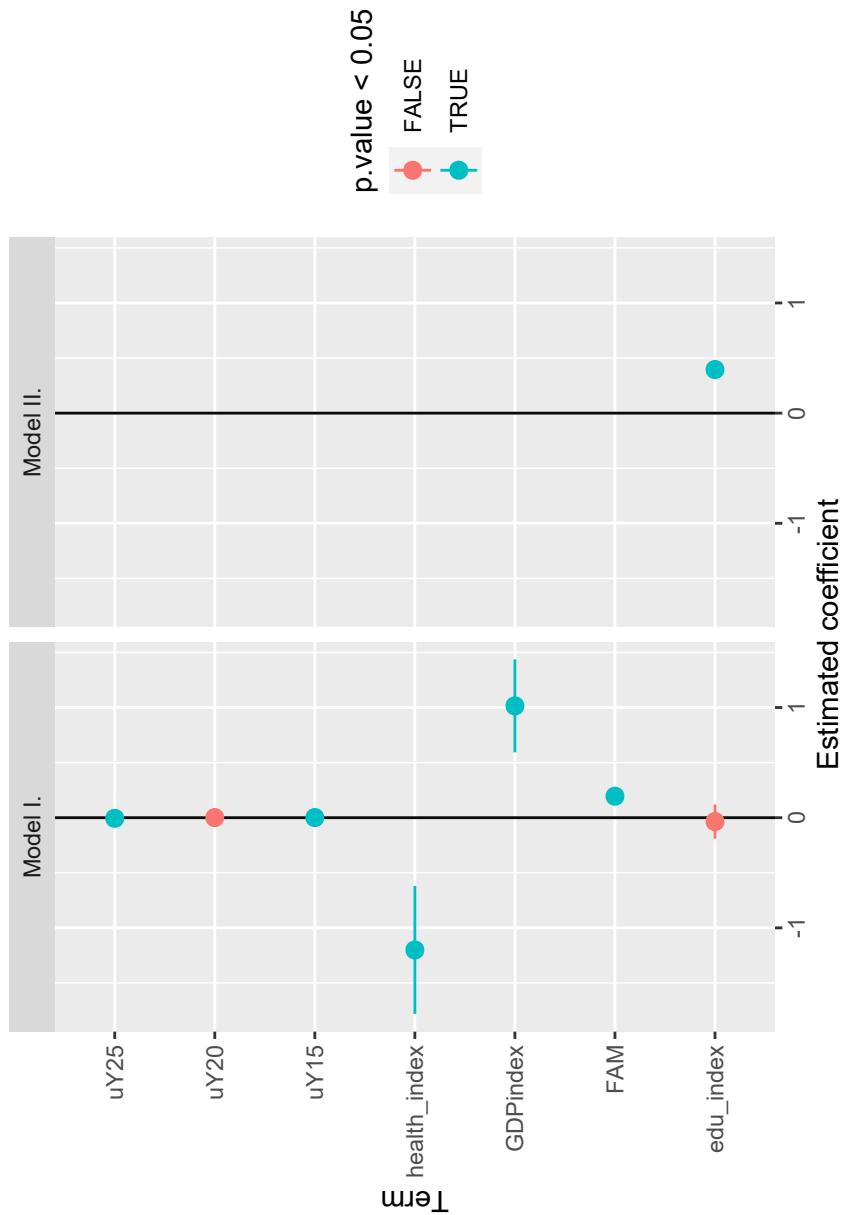


Table 5: Models

Indicator	Model I.	Model II.
pooltest	0.00%	0.00%
phtest	0.00%	7.24%
<i>AdjustedR</i> ²	6.49%	0.29%

References

UNITED NATIONS DEVELOPMENT PROGRAMME (2020), ‘Technical notes: Calculating the human development indices’.

URL: <http://hdr.undp.org/en/content/calculating-indices>

Appendix: R codes

```

1 # Set up -----
2
3 ## Packages =====
4
5 library(tidyverse)
6 library(patchwork)
7 library(knitr)
8 library(broom)
9 library(eurostat)
10
11 ## Gg theme =====
12
13 update_geom_defaults("point", list(fill = "#B1339E",
14                               shape = 21,
15                               color = "black",
16                               size = 1.4))
17 update_geom_defaults("line",
18                      list(color = "midnightblue", size = 1.4))
19
20 update_geom_defaults("smooth", list(color = "red4", size = 1.4))
21
22 update_geom_defaults("density",
23                      list(color = "midnightblue", fill = "midnightblue", alpha = .3, size = 1.4))
24
25 extrafont::loadfonts(device="win")
26
27 theme_set(theme_minimal() + theme(
28   legend.direction = "vertical",
29   # text = element_text(family = "Impact"),
30   plot.caption = element_text(family = "serif")
31 ))
32
33 # https://data.worldbank.org/indicator/SP.DYN.TFRT.IN
34
35 WB_fertility <- read_csv("WB_fertility.csv",
36                           skip = 4)
37
38 # https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD
39
40 WB_GDP <- read_csv("WB_GDP.csv",
41                      skip = 4)
42
43 merge(WB_fertility %>%
44       select('Country Name', '2017') %>%
45       rename(tfr = '2017'),
46       WB_GDP %>%
47       select('Country Name', '2017') %>%
48       rename(GDP = '2017')) %>%
49 ggplot(aes(GDP, tfr)) + geom_point() +
50 ggformula::geom_spline() +
51 scale_x_continuous(limits = c(0, 7e+4)) +
52 labs(y = "Total fertility rate (birth per woman)",
```

```

53     x = "GDP per capita, PPP (constant 2017 international USD)",
54     caption = "Own editing based on the Figure 5-2. from Kreiszné Hudák (2019).
55     The trend is drawn via splines.
56     Source of the data: World Bank."
57 )
58
59 plot_NUTS2 <- function(df, viridis_c = T, ..., all.x = F) {
60   p <- df %>%
61     {merge(eurostat::get_eurostat_geospatial(nuts_level = 2), ., all.x = all.x)} %>%
62     ggplot(aes(fill = values)) +
63     geom_sf(color = "black") +
64     theme(
65       axis.text = element_blank()
66     ) +
67     xlim(c(-30, 44)) +
68     ylim(c(35, 70)) +
69     labs(fill = NULL)
70
71   if (viridis_c) {
72     p <- p + scale_fill_viridis_c(option = "magma", ...,
73                                   guide = guide_colorbar(frame.colour = "black",
74                                              ticks.colour = "black"),
75                                   na.value = "white")
76   }
77   p
78 }
79
80 # Data import -----
81
82 # Eurostat database =====
83
84 f_data <- get_eurostat("demo_r_find2", time_format = "num") %>%
85   select(geo, time, var = indic_de, values)
86
87
88 # Data from Global Data Labor =====
89
90 ### Sub-national data ##########
91
92 # source of csv files: https://globaldatalab.org/
93
94 GDL_import <- function(x) {
95   get_eurostat_geospatial(nuts_level = 2) %>%
96     data.frame() %>%
97     tibble %>%
98     mutate(
99       ISO_Code = countrycode::countrycode(CNTR_CODE, origin = "iso2c", "iso3c"),
100      ISO_Code = ifelse(CNTR_CODE == "UK", "GBR", ISO_Code),
101      ISO_Code = ifelse(CNTR_CODE == "EL", "GRC", ISO_Code),
102    ) %>%
103     select(ISO_Code, NUTS_NAME, geo) %>%
104     merge(read_csv(x), by = "ISO_Code") %>%
105     mutate(

```

```

106     z = stringdist::stringsim(NUTS_NAME, Region)
107   ) %>%
108   arrange(desc(z)) %>%
109   filter(!duplicated(Region)) %>%
110   filter(!duplicated(NUTS_NAME)) %>%
111   filter((z > .5 | Country %in% c("Greece", "Turkey", 'Romania',
112   'Malta', 'Italy')) & NUTS_NAME != "Dresden") %>%
113   select(geo, '1990':'2018') %>%
114   pivot_longer(-1, names_to = "time", values_to = "values") %>%
115   mutate(time = as.numeric(time))
116 }
117
118 GDL_subnat <- GDL_import("GDL-Sub-national-HDI-data.csv") %>% rename(HDI = values) %>%
119   merge(
120     GDL_import("GDL-Educational-index--data.csv") %>% rename(education = values)
121   ) %>%
122   merge(
123     GDL_import("GDL-Health-index-data.csv") %>% rename(health = values)
124   ) %>%
125   merge(
126     GDL_import("GDL-Income-index-data.csv") %>% rename(income = values)
127   )
128
129 ##### National data #####
130
131 GDL_nat <- read_csv("GDL-Sub-national-HDI-data.csv") %>%
132   filter(Level == "National") %>%
133   select(Country, 6:34) %>% pivot_longer(-1, names_to = "time", values_to = "values") %>%
134   mutate(time = as.numeric(time)) %>%
135   na.omit()
136
137 # Data from UNDP =====
138
139 # source: http://hdr.undp.org/
140
141 HDI_UNDP <- read_csv("Human Development Index (HDI).csv",
142   skip = 5) %>%
143   select(!starts_with("X"), - 'HDI Rank') %>%
144   mutate_at(-1,
145     function(x) {as.numeric(ifelse(x == "...", NA, x))}) %>%
146   pivot_longer(-1, names_to = "time", values_to = "values") %>%
147   mutate(time = as.numeric(time)) %>%
148   na.omit()
149
150 merge(GDL_nat %>% rename(GDL = values),
151       HDI_UNDP %>% rename(UNDP = values)) %>%
152 {
153   c(
154     scales::percent(cor(x = .\$GDL, y = .\$UNDP)^2, accuracy = .01),
155     scales::percent(cor(x = .\$GDL, y = .\$UNDP, method = "spearman")^2, accuracy = .01),
156     as.character(format(mean(abs(.\$GDL - .\$UNDP)), digits = 1)),
157     scales::percent(mean(abs(.\$GDL - .\$UNDP) / .\$UNDP), accuracy = .01)
158   )

```

```

159 } %>%
160 {tibble(
161   Indicator = c("$R^2$", "Spearman $R^2$",
162                 "Mean absolute deviation", "Mean absolute percentage deviation"),
163   Value = .
164 )} %>%
165 kable(
166   caption = "Indicators of similarity between the Human Development Indices
167   provided by UNDP and GDL"
168 )
169
170 GDP_index <- get_eurostat("nama_10r_2gdp", time_format = "num") %>%
171   filter(unit == "PPS_HAB") %>% # Purchasing power standard (PPS) per inhabitant
172   select(-unit) %>%
173   rename(GDP = values) %>% merge(
174     get_eurostat("ert_bil_eur_a", time_format = "num") %>% # EUR/USD annual avg exc r
175       filter(currency == "USD" & statinfo == "AVG") %>%
176       select(time, e = values)
177   ) %>%
178   {
179     GDP <- .$.GDP/.$e # mutate to USD
180     mutate(.,
181       GDPindex = (log(GDP) - log(100))/
182         (log(75000) - log(100)))
183   }
184 }
185 merge(GDL_subnat, GDP_index) %>%
186   {
187     c(
188       scales::percent(cor(x = .$.income, y = .$.GDPindex)^2, accuracy = .01),
189       scales::percent(cor(x = .$.income, y = .$.GDPindex, method = "spearman")^2, accuracy = .01),
190       as.character(format(mean(abs(.$.income - .$.GDPindex)), digits = 1, nsmall = 4)),
191       scales::percent(mean(abs(.$.income - .$.GDPindex))/.$.GDPindex, accuracy = .01)
192     )
193   } %>%
194   {tibble(
195     Indicator = c("$R^2$", "Spearman $R^2$",
196                   "Mean absolute deviation", "Mean absolute percentage deviation"),
197     Value = .
198   )} %>%
199 kable(
200   caption = "Indicators of similarity between the income component of the
201   Human Development Indices provided by GDL and the estimation based on regional GDP"
202 )
203 GDP_index %>%
204   filter(time == 2017) %>%
205   select(geo, values = GDPindex) %>%
206   plot_NUTS2
207 health_index <- get_eurostat("demo_r_mlifexp", time_format = "num") %>%
208   filter(age == "Y_LT1" & sex == "T") %>%
209   select(geo, time, le = values) %>%
210   mutate(
211     health_index = (le - 20) / (85 - 20)

```

```

212 )
213 merge(GDL_subnat, health_index) %>%
214 {
215   c(
216     scales::percent(cor(x = .\$health, y = .\$health_index)^2, accuracy = .01),
217     scales::percent(cor(x = .\$health, y = .\$health_index, method = "spearman")^2, accuracy = .01),
218     as.character(format(mean(abs(.\$health - .\$health_index)), digits = 1, nsmall = 4)),
219     scales::percent(mean(abs(.\$health - .\$health_index)/.\$health_index), accuracy = .01)
220   )
221 } %>%
222 {tibble(
223   Indicator = c("$R^2$",
224                 "Spearman $R^2$",
225                 "Mean absolute deviation", "Mean absolute percentage deviation"),
226   Value = .
227 )} %>%
228 kable(
229   caption = "Indicators of similarity between the health component of the
230   Human Development Indices provided by GDL and the estimation based on regional life expectancy"
231 )
232 edu_wide <- get_eurostat("edat_lfse_04", time_format = "num") %>%
233   filter(sex == "T" & !str_detect(isced11, "GEN") &
234         !str_detect(isced11, "VOC") & isced11 != "ED3-8"
235   ) %>%
236   mutate(
237     var = str_c(age, ":", isced11)
238   ) %>%
239   select(geo, time, var, values) %>%
240   pivot_wider(names_from = var, values_from = values) %>%
241   {
242     x <- .
243     names(x) <- letters[1:length(x)]
244     x <- cbind(
245       .[, 1:2],
246       mice::complete(mice::mice(select(x, -a,-b), printFlag = F))
247     )
248     names(x) <- names(.)
249     x
250   }
251 edu_comps <- edu_wide%>%
252   select(-time, -geo) %>%
253   na.omit() %>%
254   {princomp(scale(.))}

255 edu_comp_vars <- edu_comps %>%
256   summary() %>%
257   {.\$sdev^2/sum(.\$sdev^2)} %>%
258   scales::percent(accuracy = .01) %>%
259   {str_c("# ", 1:length(.), " (", ., ")")}

260
261 edu_comps %>%
262   .\$loadings %>%
263   unclass() %>%
264   data.frame() %>%

```

```

265  rownames_to_column() %>%
266  pivot_longer(-1) %>%
267  mutate(
268    name = as.numeric(str_remove(name, 'Comp.')),
269  ) %>%
270  arrange(name) %>%
271  mutate(
272    name = edu_comps %>%
273      summary() %>%
274      {$.sdev^2/sum(.\$sdev^2)} %>%
275      scales::percent(accuracy = .01) %>%
276      {str_c("# ", 1:length(.), " (", ., ")")} %>%
277      .[name]
278  ) %>%
279  ggplot +
280  aes(rownames, value, fill = value < 0) +
281  geom_hline(yintercept = 0) +
282  geom_col(color = 'black') +
283  coord_flip() +
284  scale_fill_viridis_d(guide = F, option = "magma", begin = .4,
285                        end = .7, direction = -1) +
286  scale_y_continuous(labels = scales::percent) +
287  facet_wrap(~name, ncol = 3) +
288  labs(x = NULL, y = NULL, caption =
289        "The corresponding proportion of the explained variance are in the brackets.")
290 edu_comps %>% .\$scores %>%
291   cbind(edu_wide) %>% merge(GDL_subnat) %>%
292   select(3:11, education) %>%
293 {
294   x <- .\$education
295   apply(select(., -education), 2, function(y) {
296     c(
297       scales::percent(cor(x = x, y = y)^2, accuracy = .01),
298       scales::percent(cor(x = x, y = y, method = "spearman")^2, accuracy = .01)
299     )
300   })
301 }
302 } %>%
303 data.frame() %>%
304 mutate(Indicator = c("$R^2$", "Spearman $R^2$")) %>%
305 rename_all(function(x) str_replace(x, "p.", "p ")) %>%
306 select(Indicator, 1:9) %>%
307 kable(
308   caption = "Indicators of similarity between the knowledge component of Human Development Indices
309   provided by UNDP and the calculated principal components using educational attainment level"
310 )
311 edu_index <- edu_comps %>%
312   .\$scores %>%
313   data.frame() %>%
314   select(2) %>%
315   cbind(edu_wide) %>%
316   select(geo, time, edu_index = Comp.2) %>%
317   mutate(
318     edu_index = -edu_index,

```

```

318     edu_index = edu_index + abs(min(edu_index)),
319     edu_index = edu_index/max(edu_index)
320   )
321 dat <- f_data %>%
322   pivot_wider(names_from = var, values_from = values) %>%
323   merge(edu_index, all = T) %>%
324   merge(health_index, all = T) %>%
325   merge(GDP_index, all = T)
326 FAM_df <- get_eurostat("spr_exp_sum", time_format = "num") %>%
327   filter(spdeps == "FAM" & unit == "PC_GDP") %>%
328   rename(FAM = values, country = geo) %>%
329   select(-(spdeps:unit))
330
331 yth_empl_byage <- get_eurostat("yth_empl_110", time_format = "num") %>%
332   filter(unit == "PC" & sex == "F") %>%
333   filter(age %in% c("Y15-19", "Y20-24", "Y25-29")) %>%
334   select(-unit, -sex) %>%
335   pivot_wider(names_from = age, values_from = values) %>%
336   rename(
337     "uY15" = "Y15-19",
338     "uY20" = "Y20-24",
339     "uY25" = "Y25-29"
340   )
341
342 dat <- dat %>%
343   mutate(country = str_sub(geo, end = 2)) %>%
344   merge(FAM_df, all.x = T, all.y = F) %>%
345   merge(yth_empl_byage, all.x = T, all.y = F)
346
347 dat %>%
348   filter(str_length(geo) == 4) %>%
349   select(TOTFERRT, GDPindex, edu_index, health_index) %>%
350   rename(
351     'TFR'= 'TOTFERRT',
352     'Income index' = 'GDPindex',
353     'Health index' = 'health_index',
354     'Education index' = 'edu_index'
355   ) %>%
356   GGally::ggpairs()
357
358 dat %>%
359   filter(str_length(geo) == 4) %>%
360   select(TOTFERRT, uY15, uY20, uY25) %>%
361   set_names("TFR", "Unemployment rate\n(15-19 years old)",
362             "Unemployment rate\n(20-24 years old)",
363             "Unemployment rate\n(25-29 years old)") %>%
364   GGally::ggpairs()
365 dat %>%
366   filter(str_length(geo) == 2) %>%
367   ggplot(aes(FAM, TOTFERRT)) + geom_point() +
368   ggformula::geom_spline() +
369   labs(y = "Total fertility rate (birth per woman)",
370        x = "Family benefit (% of GPD)",
```

```

371     caption = "The trend is drawn via splines."
372   )
373
374 m_part <- dat %>%
375   filter(str_length(geo) == 4) %>%
376   select(TOTFERRT, GDPindex, edu_index, health_index) %>%
377   rename(
378     'Income index' = 'GDPindex',
379     'Health index' = 'health_index',
380     'Education index' = 'edu_index'
381   ) %>%
382   na.omit() %>%
383   rpart::rpart(formula = TOTFERRT ~ ., cp = .02)
384
385 m_part %>% rattle::fancyRpartPlot(palettes = 'PuRd', sub = NULL)
386
387 m_part %>% summary()
388 m_part2 <- dat %>%
389   filter(str_length(geo) == 4) %>%
390   select(TOTFERRT, GDPindex, edu_index, health_index, uY15, uY20, uY25) %>%
391   rename(
392     'Income index' = 'GDPindex',
393     'Health index' = 'health_index',
394     'Education index' = 'edu_index',
395     'Unemployment rate at age 15-19' = 'uY15'
396   ) %>%
397   na.omit() %>%
398   rpart::rpart(formula = TOTFERRT ~ ., cp = .02)
399
400 m_part2 %>% rattle::fancyRpartPlot(palettes = 'PuRd', sub = NULL)
401
402 GDL_subnat %>%
403   filter(time == 2017) %>%
404   select(-time) %>%
405   pivot_longer(-1, names_to = "var", values_to = "values") %>%
406   filter(!is.na(var)) %>%
407   mutate(
408     var = str_to_title(var),
409     var = str_replace(var, "Hdi", "HDI"),
410     var = factor(var,
411                   levels = c("HDI", "Income", "Education", "Health"),
412                   ordered = T)
413   ) %>%
414   plot_NUTS2() + facet_wrap(~ var, ncol = 2) +
415   labs(caption = "Source: https://globaldatalab.org")
416
417 fertility_byage <- get_eurostat("demo_r_frate2", time_format = "num") %>%
418   filter(age != "TOTAL") %>%
419   transmute(age = age, geo = geo, time, values)
420
421 fr_agegroups_wide <- fertility_byage %>%
422   mutate(age = str_remove(age, "Y"), # sorting into categories by 5 years
423         age = str_remove(age, "10-"),

```

```

424     age = str_remove(age, "_GE"),
425     age = as.numeric(age),
426     age_group = paste0((age %/% 5)*5, "-", (age %/% 5 + 1)*5 - 1),
427     age_group = ifelse(age_group == "50-54", "50-", age_group)
428 ) %>%
429 group_by(age_group, time, geo) %>%
430 summarise(values = sum(values)) %>%
431 ungroup() %>%
432 pivot_wider(names_from = "age_group", values_from = "values") %>%
433 select(-"10-14", -"50-") %>%
434 na.omit() %>%
435 arrange(time, geo)

436 fr_age_comps <- fr_agegroups_wide %>%
437   select(-time, -geo) %>%
438   na.omit() %>%
439   {princomp(scale(.))}

440 fr_age_comp_vars <- fr_age_comps %>%
441   summary() %>%
442   {$.sdev^2/sum($.sdev^2)} %>%
443   scales::percent(accuracy = .01) %>%
444   {str_c("# ", 1:length(.), " (", ., ")")}

445 fr_age_comps %>%
446   .$loadings %>%
447   unclass() %>%
448   data.frame() %>%
449   rownames_to_column() %>%
450   pivot_longer(-1) %>%
451   mutate(
452     name = as.numeric(str_remove(name, 'Comp.'))
453   ) %>%
454   arrange(name) %>%
455   mutate(
456     name = fr_age_comps %>%
457       summary() %>%
458       {$.sdev^2/sum($.sdev^2)} %>%
459       scales::percent(accuracy = .01) %>%
460       {str_c("# ", 1:length(.), " (", ., ")")}) %>%
461       .[name]
462   ) %>%
463   ggplot +
464   aes(rowname, value, fill = value < 0) +
465   geom_hline(yintercept = 0) +
466   geom_col(color = 'black') +
467   coord_flip() +
468   scale_fill_viridis_d(guide = F, option = "magma", begin = .4,
469                         end = .7, direction = -1) +
470   scale_y_continuous(labels = scales::percent) +
471   facet_wrap(~name, ncol = 3) +
472   labs(x = NULL, y = NULL, caption =
473         "The corresponding proportion of the explained variance are in the brackets.")

```

```

477 fr_age_comps %>%
478   .$score %>%
479   .[,1] %>%
480   {cbind(values =., na.omit(fr_agegroups_wide))} %>%
481   select(1:3) %>%
482   pivot_wider(names_from = geo, values_from = values) %>%
483   arrange(time) %>%
484   {apply(.[, -1], 2, FUN = function(x) {x <- na.omit(x)
485     (tail(x, 1) - head(x, 1))/length(x)})} %>%
486   {tibble(FID = names(.), values = .)} %>%
487   plot_NUTS2()

488
489 fr_age_comps %>%
490   .$score %>%
491   .[,1] %>%
492   {cbind(values =., na.omit(fr_agegroups_wide))} %>%
493   select(1:3) %>%
494   filter(time == 2017) %>%
495   plot_NUTS2()

496
497 dat_plm <- dat %>%
498   select(
499     geo, time, TOTFERRT, edu_index, health_index, GDPindex, FAM, uY15, uY20, uY25
500   ) %>%
501   filter(str_length(geo) == 4 & !is.na(TOTFERRT))
502 library(plm)

503
504 m_panels <- c("TOTFERRT ~ edu_index + health_index + GDPindex + FAM + uY15 + uY20 + uY25", "TOTFERRT ~ "
505   lapply(function(formula) {
506     pooling <- plm(eval(formula), data = dat_plm, model = "pooling")
507     within <- plm(eval(formula), data = dat_plm, model = "within")
508     random <- plm(eval(formula), data = dat_plm, model = "random")

509
510     # X <- fixef(within) %>%
511     # data.frame() %>%
512     # rownames_to_column() %>%
513     # set_names('geo', 'alpha')
514     #
515     # X <- merge(dat_plm, X) %>%
516     # mutate(TOTFERRT = TOTFERRT + alpha)
517     #
518     # X <- lm(data = dat_plm, formula = eval(paste(formula, "+ geo")))
519     #
520     # X <- lm(data = dat_plm, formula = eval(paste(formula, "+ geo"))) %>%
521     # coef() %>%
522     # data.frame() %>%
523     # rownames_to_column() %>%
524     # set_names('geo', 'alpha') %>%
525     # filter(geo == "(Intercept)" | str_detect(geo, "geo")) %>%
526     # mutate(geo = str_replace(geo, "geo_"))

527
528     # merge(dat_plm, X)
529
530

```

```

531   list(
532     tests = c(
533       pooltest(pooling, within)$p.value,
534       phtest(within, random)$p.value,
535       plm::r.squared(within, dfcor = T)),
536     model = within,
537     OLS = lm(data = dat_plm, formula = eval(paste(formula, "+ geo")))
538   )
539 )
540 m_panels %>%
541   lapply(function(output) {
542     output$model %>% broom::tidy(conf.int = T) %>%
543       rownames_to_column()
544   }) %>%
545   reduce(rbind) %>%
546   mutate(rowname = paste0("Model ", as.roman(cumsum(rowname == 1)), "."))
547   %>%
548   ggplot() +
549     aes(estimate, term, color = p.value < .05) +
550     geom_vline(xintercept = 0, color = "gray4") +
551     geom_point() +
552     geom_pointrange(aes(xmin = conf.low, xmax = conf.high)) +
553     facet_wrap(~rowname, nrow = 1) +
554     theme_grey() +
555     labs(x = "Estimated coefficient", y = "Term")
556
557 m_panels %>%
558   lapply(function(output) {
559     output$tests
560   }) %>%
561   reduce(rbind) %>%
562   t() %>%
563   data.frame() %>%
564   mutate_all(function(x) scales::percent(x, accuracy = .01)) %>%
565   {set_names(., paste0("Model ", as.roman(1:ncol(.))), ".")} %>%
566   mutate(
567     Indicator = c("pooltest", "phtest", "$Adjusted R^2$")
568   ) %>% # TODO ORDER
569   select(Indicator, everything()) %>%
knitr::kable(caption = "Models") # TODO NAME

```