Acemoglu cikk reprodukálása

Granát M., Papp L., Szabó D.

2021. október 27.

Tartalomjegyzék

Bevezetés 1

Táblázat 1 2

GDP adat összehasonlítás . . . 3

Világban adat . . . 3

Penn World Table adat . . . 3

Foglakoztatottra jutó kibocsátás . . . 5

Average protection against expropriation risk . . . 6

Constrain on executive in 1900 . . . 6

Táblázat 2 7

Táblázat 3 11

Táblázat 4 12

Táblázat 5 14

Táblázat 6 16

Táblázat 7 20

Sachs geography . . . 21

Bevezetés

Összehasonlításként az összes csatolt adatot letöltöttem Acemoglu honalapjáról, hogy ellenőrizni tudjuk az általunk gyűjtöttek helyességét. Sajnos már az ide feltöltött adatok sem minden esetben korrektek. Az adatfájlok mellé feltöltött .do fájlban az alábbi komment olvasható például: „I am not sure what values of extmort4 were used to calculate quantiles, perhaps from an earlier version of the data that I do not have access to. Thus, I cannot exactly reproduce Table 1, but the summary statistics that emerge using quantiles of extmort4 in the final dataset are similar."

Annak belátására, hogy jó adattal dolgozunk-e, az Acemoglu által feltöltött adatállománnyal és az általunk összegyűjtött adatokkal is reprodukálom a táblázatokat.

Az `ace_data1` tartalmazza az Acemoglu oldalára Table 2 pont alatt feltöltött DTA fájl adatait.

```
list.files(str_c(WD, "/Acemoglu-data"), full.names = TRUE) %>%
  # read all data from authors website
  walk(~ assign(x = str_c("ace_data", str_remove_all(., "\\D")), # new name
               value = haven::read_dta(.),
               envir = .GlobalEnv)
  )
```

A `basecp` vektor tartalmazza a felhasználni kívánt országok 3 betűs jelölését (a 64 elemű minta).

```
baseco <- ace_data1 %>%
  filter(!is.na(baseco)) %>%
  pull(shortnam) %>%
  as.character()
```

Táblázat 1

Átlagok:

```
ace_data1 %>%
  select(-shortnam) %>%
  {
    bind_rows(
      mutate(., n = ifelse(is.na(baseco), "base", "outer")) %>%
        group_by(n) %>%
        summarise_all(mean, na.rm = TRUE) %>%
        ungroup() %>%
        filter(n == "base"),
      summarise_all(., mean, na.rm = TRUE) %>%
        mutate(n = "world")
    )
  } %>%
  pivot_longer(-1) %>%
  pivot_wider(names_from = n)
```

name	base	world
euro1900	38.6174757	30.102410
avexpr	7.4541958	6.988548
logpgp95	8.4622103	8.304196
cons1	4.0625000	3.630435
democ00a	0.1290323	1.122222
cons00a	1.1944444	1.854167
extmort4	148.9463330	214.964893
logem4	4.4935445	4.611108
loghjypl	-1.5011873	-1.709099
baseco	NaN	1.000000

Kvantilisek:

```
ace_data1 %>%
  filter(!is.na(baseco)) %>%
  mutate(g = cut(x = extmort4, breaks = quantile(extmort4, probs = (0:4)*.25), include.lowest = T, F))
  group_by(g) %>%
  summarize_if(is.numeric, mean, na.rm = TRUE)
```

g	euro1900	avexpr	logpgp95	cons1	democ00a	cons00a	extmort4	logem4	loghjypl	baseco
1	32.6562502	7.743316	8.873877	4.333333	4.133333	4.066667	31.92353	3.224033	-1.075929	1
2	23.3333340	6.363636	8.327826	2.733333	2.285714	2.800000	73.31400	4.293698	-1.496654	1

g	euro1900	avexpr	logpgp95	cons1	democ00a	cons00a	extmort4	logem4	loghjypl	baseco
3	8.1705882	5.973262	7.777815	3.125000	0.187500	1.125000	139.75647	4.919492	-2.195082	1
4	0.5333333	5.890909	7.199134	3.428571	0.000000	1.000000	781.33600	6.346973	-3.025752	1

Ez alapján a változók mögötti értékek[1]:

```
NiceName <- function(x) {
  # code names and their meaning
  case_when(
    x == "logpgp95" ~ "Log GDP per capita (PPP) in 1995",
    x == "loghjypl" ~ "Log output pper worker in 1988",
    x == "avexpr" ~ "Average protection against expropriation risk",
    x == "cons1" ~ "Constrain on executive in 1900",
    x == "democ00a" ~ "Democracy in 1900",
    x == "euro1900" ~ "Eurpean settlements in 1900",
    x == "logem4" ~ "Log European settler mortaility",
    TRUE ~ x
  )
}
```

GDP adat összehasonlítás

Az összehasonlítás kivitelezéséhez minden változót azonos kóddal mentek, mint Acemoglu, de utána írom a forrás rövidítését.

Világban adat

A Világ Bank API segítségével azonnali hozzáférés biztosít az adattábláihoz, így csak annak kódjára van szükség.

```
logpgp95_wb <- wb(indicator = "NY.GDP.PCAP.PP.CD") %>%
  filter(date == 1995) %>%
  transmute(geo = iso3c, logpgp95 = log(value))
```

Penn World Table adat

```
pwt100 <- readxl::read_excel(str_c(WD, "/data/pwt100.xlsx"), sheet = "Data")
```

A megbeszélteknek megfelelően az `rgdpe` változót használjuk fel.

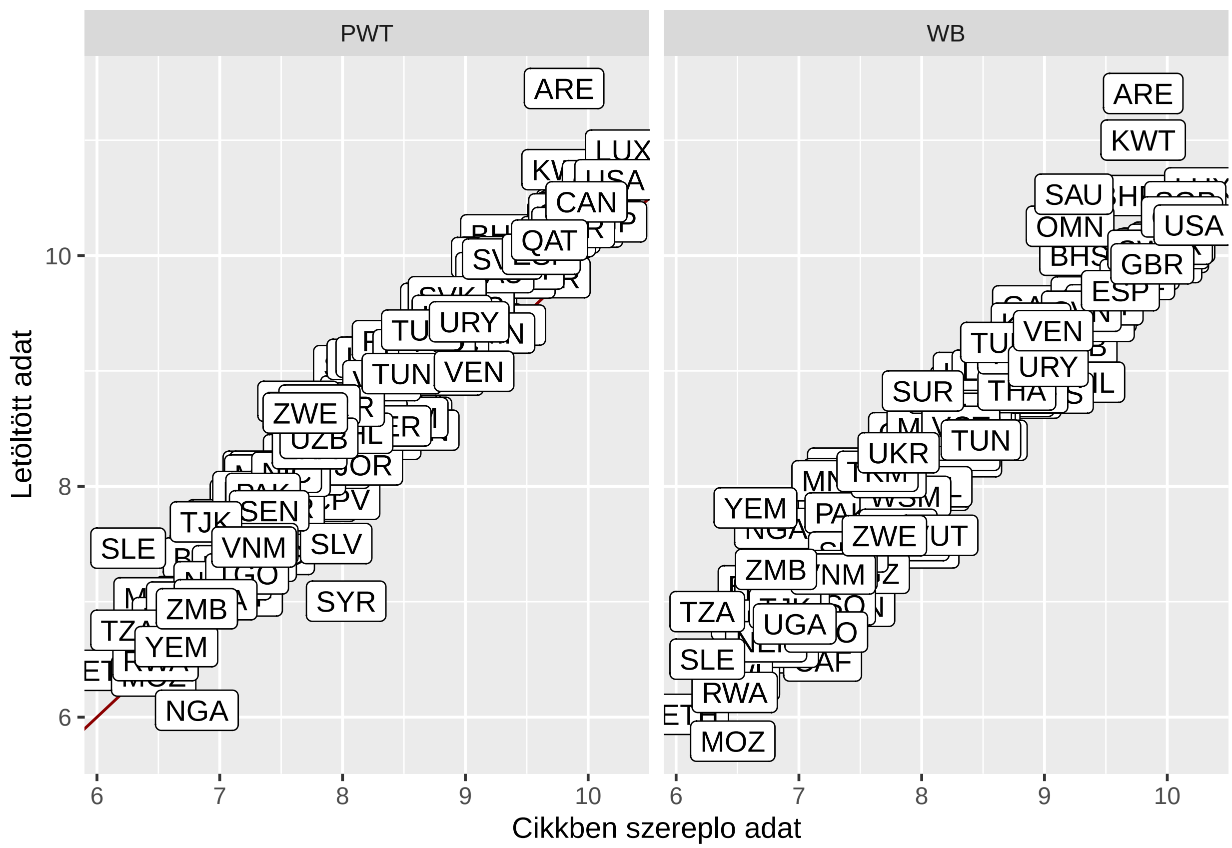
```
logpgp95_pwt <- pwt100 %>%
  filter(year == 1995) %>%
  transmute(geo = countrycode, logpgp95_pwt = log(rgdpe/pop))
# rgde: Expenditure-side real GDP at chained PPPs (in mil. 2017US$)
```

```
logpgp95_checkdf <- list(
  logpgp95_wb %>%
    set_names("shortnam", "logpgp95_wb"),
  logpgp95_pwt %>%
    set_names("shortnam", "logpgp95_pwt"),
```

[1]Egy függvénybe beteszem a kódnak megfelelő neveket, így az ábrákon már ezzel könnyen tudunk kódok helyett változó neveket feltüntetni

```
  ace_data1
) %>%
  reduce(full_join) %>%
  select(shortnam, starts_with("logpgp95")) %>%
  pivot_longer(-c(shortnam, logpgp95)) %>%
  na.omit() %>%
  mutate(
    name = str_remove_all(name, "logpgp95_"),
    name = str_to_upper(name)
  )
```

```
ggplot(logpgp95_checkdf) +
  aes(logpgp95, value, label = shortnam) +
  facet_wrap(~ name) +
  geom_abline(color = "red4") +
  geom_label() +
  labs(x = "Cikkben szereplő adat", y = "Letöltött adat")
```



1. ábra. A Világ Bank, a Penn World Table és a cikkből származó GDP/fő adatok összehasonlítása

```
logpgp95_checkdf %>%
  group_by(name) %>%
  summarise(
    pearson_cor = cor(value, logpgp95, method = "pearson"),
    spearman_cor = cor(value, logpgp95, method = "spearman"),
    rel_diff = mean(value - logpgp95 / logpgp95),
```

```
    n = n()
  )
```

name	pearson_cor	spearman_cor	rel_diff	n
PWT	0.9607875	0.9653931	7.734534	155
WB	0.9519478	0.9603487	7.473285	156

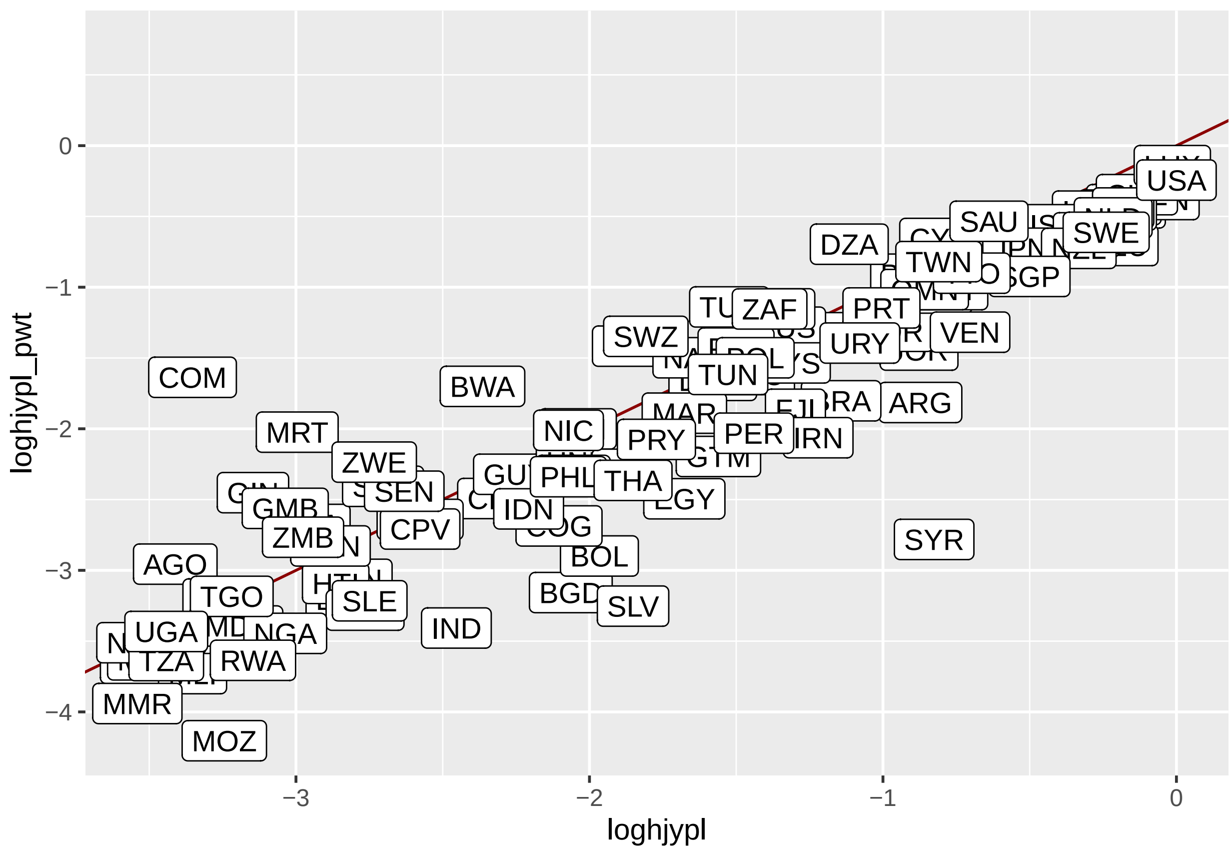
A korrelációk és az eltérések alapján számomra nem teljesen egyértelmű, hogy melyik forrás GDP adatát érdemesebb használni, de az ábra alapján a Világ Bank adata közelebb van ahhoz, amit a cikkben látunk.

Foglakoztatottra jutó kibocsátás

Ezt a változót is a Penn World Tableből emeltük ki.

```
loghjypl_pwt <- pwt100 %>%
  filter(year == 1988) %>%
  transmute(geo = countrycode, loghjypl_pwt = log(rgdpe/emp/1e5))
```

```
loghjypl_pwt %>%
  right_join(ace_data1, by = c("geo" = "shortnam")) %>%
  ggplot() +
  aes(loghjypl, loghjypl_pwt, label = geo) +
  geom_abline(color = "red4") +
  geom_label()
```



```
logYL <- readxl::read_excel(str_c(WD, "/data/logYL.xlsx"))
```

```
logYL %>%
  select(shortnam = Cod, logYL:logL) %>%
  pivot_longer(-1) %>%
  left_join(ace_data1) %>%
  head()
```

shortnam	name	value	euro1900	avexpr	logpgp95	cons1	democ00a	cons00a	extmort4	logem4	loghjypl	baseco
DZA	logYL	9.35998	13	6.5	8.389359	2	0	1	78.2	4.35927	-1.114742	1
DZA	logKL	10.30679	13	6.5	8.389359	2	0	1	78.2	4.35927	-1.114742	1
DZA	alogKY	0.47341	13	6.5	8.389359	2	0	1	78.2	4.35927	-1.114742	1
DZA	School	2.39000	13	6.5	8.389359	2	0	1	78.2	4.35927	-1.114742	1
DZA	logHL	0.32026	13	6.5	8.389359	2	0	1	78.2	4.35927	-1.114742	1
DZA	logA	8.56631	13	6.5	8.389359	2	0	1	78.2	4.35927	-1.114742	1

Ezt az adatot végül nem emlékszem miért nem használtuk fel.

Average protection against expropriation risk

Ezt a cikkből emeljük át.

Constrain on executive in 1900

```
p5v2018 <- readxl::read_excel(str_c(WD, "/data/p5v2018d.xls"))
```

Az `scode` félrevezető változó volt, a `ccode` iso kódra való átváltásával volt érdemes elvégezni a párosítást.

```
p5v_filtered <- p5v2018 %>%
  transmute(
    country = countrycode::countrycode(ccode, "cown", "iso3c"),
    time = map2(byear, eyear, ~ seq(from = .x, to = .y)),
    exconst, democ) %>%
  unnest(time) %>%
  arrange(time) %>%
  mutate_at(-(1:2), ~ ifelse(. < 0, NA, .)) %>%
  group_by(country) %>%
  group_modify(~ fill(., everything(), .direction = "up")) %>%
  ungroup() %>%
  filter(!duplicated(str_c(country, time)))
```

```
exconst <- p5v_filtered %>%
  select(-democ) %>%
  filter(time %in% c(1900, 1970, 1990)) %>%
  group_by(time) %>%
  group_modify(~ left_join(tibble(country = baseco), .)) %>%
  replace_na(list(exconst = 1)) %>%
```

```
  ungroup() %>%
  pivot_wider(names_from = time, values_from = exconst, names_prefix = "exconst_")
```

```
democ <- p5v_filtered %>%
  arrange(time) %>%
  filter(!duplicated(country) | time == 1900) %>%
  pivot_longer(exconst:democ) %>%
  mutate(
    time = ifelse(time == 1900, "00", "01"),
    name = str_c(name, "_", time)
  ) %>%
  select(-time) %>%
  pivot_wider(names_from = name, values_from = value)
```

```
ind_time <- p5v_filtered %>%
  group_by(country) %>%
  slice_min(time, n = 1) %>%
  ungroup() %>%
  transmute(country, ind_time = time)
```

```
tibble(baseco) %>%
  anti_join(tibble(baseco = unique(p5v_filtered$country))) %>%
  pull(baseco) %>%
  {cat(str_c("Az alábbi országok azonban nem érhetőek el a táblázatban: ", str_c(., collapse = ", ")))}
```

```
## Az alábbi országok azonban nem érhetőek el a táblázatban: BHS, HKG, MLT, ZAR
```

Táblázat 2

```
tibble(model_range = c("Whole world", "Base sample")) %>%
  slice(rep(1:n(), times = 4)) %>%
  mutate(
    model = str_c("Model ", row_number()),
    explanatory = c(
      "avexpr",
      "avexpr",
      "avexpr + lat_abst",
      "africa + lat_abst + avexpr + other + asia",
      "avexpr + lat_abst",
      "africa + lat_abst + avexpr + other + asia",
      "avexpr",
      "avexpr"
    ),
    dependent = c(rep("logpgp95", 6), rep("loghjypl", 2)),
    formula = str_c(dependent, " ~ ", explanatory),
    data = list(ace_data2),
    var_names = map(data, names),
    var_names = map2(var_names, formula, ~ keep(.x, function(x) str_detect(.y, x))),
    data = map2(data, model_range, ~ switch(.y, "Base sample" =  filter(.x, !is.na(baseco)), .x)),
    data = map2(data, var_names, select),
    data = map(data, na.omit),
    fit = map2(formula, data, lm),
    map_df(fit, ~ broom::glance(.)["r.squared"]),
```

```
    coef = map(fit, broom::tidy)
  ) %>%
  unnest(coef) %>%
  transmute(model, model_range, r2 = scales::percent(r.squared), term,
            est = str_c(round(estimate, 2), " (", round(std.error, 2), ")")) %>%
  filter(term != "(Intercept)") %>%
  pivot_wider(names_from = term, values_from = est) %>%
  pivot_longer(-1) %>%
  pivot_wider(names_from = model) %>%
  mutate_all(~ ifelse(is.na(.), "", .)) %>%
  knitr::kable(caption = "Acemoglu adatai alapján készült becslésünk a táblázat 2-re.")
```

5. táblázat: Acemoglu adatai alapján készült becslésünk a táblázat 2-re.

name	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
model_range	Whole world	Base sample	Whole world	Base sample	Whole world	Base sample	Whole world	Base sample
r2	61.1%	54.0%	62.3%	71.4%	62.3%	71.4%	55.4%	48.6%
avexpr	0.53 (0.04)	0.52 (0.06)	0.46 (0.06)	0.4 (0.06)	0.46 (0.06)	0.4 (0.06)	0.45 (0.04)	0.46 (0.06)
lat_abst			0.87 (0.49)	0.88 (0.63)	0.87 (0.49)	0.88 (0.63)		
africa				-0.88 (0.17)		-0.88 (0.17)		
other				0.11 (0.38)		0.11 (0.38)		
asia				-0.58 (0.23)		-0.58 (0.23)		

```
laporta_qgov <- readxl::read_excel("data/laporta_qgov.xls",
                                   sheet = "data")
```

```
laporta_qgov %>%
  transmute(country = str_to_upper(flopcode), lat_abst, legor_fr, legor_uk) %>%
  right_join(tibble(country = baseco))
```

country	lat_abst	legor_fr	legor_uk
AGO	0.1366667	1	0
ARG	0.3777778	1	0
AUS	0.3000000	0	1
BFA	0.1444445	1	0
BGD	0.2666667	0	1
BHS	0.2683333	0	1
BOL	0.1888889	1	0
BRA	0.1111111	1	0
CAN	0.6666667	0	1
CHL	0.3333333	1	0
CIV	0.0888889	1	0
CMR	0.0666667	1	0
COG	0.0111111	1	0
COL	0.0444444	1	0

country	lat_abst	legor_fr	legor_uk
CRI	0.1111111	1	0
DOM	0.2111111	1	0
DZA	0.3111111	1	0
ECU	0.0222222	1	0
EGY	0.3000000	1	0
ETH	0.0888889	1	0
GAB	0.0111111	1	0
GHA	0.0888889	0	1
GIN	0.1222222	1	0
GMB	0.1475556	0	1
GTM	0.1700000	1	0
GUY	0.0555556	0	1
HKG	0.2461111	0	1
HND	0.1666667	1	0
HTI	0.2111111	1	0
IDN	0.0555556	1	0
IND	0.2222222	0	1
JAM	0.2016667	0	1
KEN	0.0111111	0	1
LKA	0.0777778	0	1
MAR	0.3555556	1	0
MDG	0.2222222	1	0
MEX	0.2555556	1	0
MLI	0.1888889	1	0
MLT	0.3944444	1	0
MYS	0.0255556	0	1
NER	0.1777778	1	0
NGA	0.1111111	0	1
NIC	0.1444445	1	0
NZL	0.4555556	0	1
PAK	0.3333333	0	1
PAN	0.1000000	1	0
PER	0.1111111	1	0
PRY	0.2555556	1	0
SDN	0.1666667	0	1
SEN	0.1555556	1	0
SGP	0.0135556	0	1
SLE	0.0922222	0	1
SLV	0.1500000	1	0
TGO	0.0888889	1	0
TTO	0.1222222	0	1
TUN	0.3777778	1	0
TZA	0.0666667	0	1
UGA	0.0111111	0	1
URY	0.3666667	1	0
USA	0.4222222	0	1
VEN	0.0888889	1	0
VNM	0.1777778	0	0
ZAF	0.3222222	0	1
ZAR	0.0000000	1	0

Mind a 64 országra megvan benne a kellő adat (szélességi fog, francia/angol jogrendszer).

A kontinens dummy-kat az R `{countrycode}` csomagjával készítettük.

```
check_df_2 <- ace_data2 %>%
  transmute(country = shortnam, avexpr) %>%
  full_join(logpgp95_wb, by = c("country" = "geo")) %>%
  full_join(loghjypl_pwt, by = c("country" = "geo")) %>%
  full_join(
    laporta_qgov %>%
      transmute(country = str_to_upper(flopcode), lat_abst) %>%
      mutate(
        continent = countrycode::countrycode(country, "iso3c", "continent")
      ) %>%
      transmute(
        country, lat_abst, continent = ifelse(continent %in% c("Americas", "Asia", "Africa"), continent
      )
  ) %>%
  transmute(
    shortnam = country,
    africa = ifelse(continent == "Africa", 1, 0),
    lat_abst,
    avexpr,
    logpgp95 = logpgp95,
    other = ifelse(continent == "other", 1, 0),
    asia = ifelse(continent == "Asia", 1, 0),
    loghjypl = loghjypl_pwt,
    baseco = ifelse(shortnam %in% baseco, 1, NA)
  ) %>%
  filter(shortnam %in% ace_data2$shortnam)
```

```
tibble(model_range = c("Whole world", "Base sample")) %>%
  slice(rep(1:n(), times = 4)) %>%
  mutate(
    model = str_c("Model ", row_number()),
    explanatory = c(
      "avexpr",
      "avexpr",
      "avexpr + lat_abst",
      "africa + lat_abst + avexpr + other + asia",
      "avexpr + lat_abst",
      "africa + lat_abst + avexpr + other + asia",
      "avexpr",
      "avexpr"
    ),
    dependent = c(rep("logpgp95", 6), rep("loghjypl", 2)),
    formula = str_c(dependent, " ~ ", explanatory),
    data = list(check_df_2),
    var_names = map(data, names),
    var_names = map2(var_names, formula, ~ keep(.x, function(x) str_detect(.y, x))),
    data = map2(data, model_range, ~ switch(.y, "Base sample" =  filter(.x, !is.na(baseco)), .x)),
    data = map2(data, var_names, select),
    data = map(data, na.omit),
    fit = map2(formula, data, lm),
    map_df(fit, ~ broom::glance(.)["r.squared"]),
    coef = map(fit, broom::tidy)
  ) %>%
```

```
  unnest(coef) %>%
  transmute(model, model_range, r2 = scales::percent(r.squared), term,
            est = str_c(round(estimate, 2), " (", round(std.error, 2), ")")) %>%
  filter(term != "(Intercept)") %>%
  pivot_wider(names_from = term, values_from = est) %>%
  pivot_longer(-1) %>%
  pivot_wider(names_from = model) %>%
  mutate_all(~ ifelse(is.na(.), "", .)) %>%
  knitr::kable(caption = "Táblázat 2 az általunk összerakott adatokkal futtatva.")
```

7. táblázat: Táblázat 2 az általunk összerakott adatokkal futtatva.

name	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
model_range	Whole world	Base sample	Whole world	Base sample	Whole world	Base sample	Whole world	Base sample
r2	46.49%	47.34%	48.96%	65.75%	48.96%	65.75%	43.47%	48.25%
avexpr	0.49 (0.05)	0.52 (0.07)	0.39 (0.06)	0.43 (0.07)	0.39 (0.06)	0.43 (0.07)	0.41 (0.04)	0.48 (0.06)
lat_abst			1.34 (0.59)	0.7 (0.78)	1.34 (0.59)	0.7 (0.78)		
africa				-0.96 (0.19)		-0.96 (0.19)		
other				-0.08 (0.44)		-0.08 (0.44)		
asia				-0.57 (0.26)		-0.57 (0.26)		

Táblázat 3

```
logem4 <- ace_data7 %>%
  select(shortnam, logem4)
```

A halálozási adatot Acemoglutól vesszük át, csakúgy, mint az európaiak jelenlétét 1900-ban, illetve a kizsákmányolást.

```
check_df_3 <- ace_data8 %>%
  select(country = shortnam, euro1900, logem4, avexpr) %>%
  left_join(ind_time) %>%
  left_join(democ[c("country", "exconst_00", "exconst_01", "democ_00")]) %>%
  replace_na(list(exconst_00 = 1, democ_00 = 1)) %>%
  left_join(transmute(laporta_qgov, country = str_to_upper(flopcode), lat_abst))
```

```
check_df_3 %>%
  pivot_longer(c(exconst_00, democ_00, exconst_01, euro1900, logem4)) %>%
  group_by(name) %>%
  # select(avexpr, value) %>%
  nest() %>%
  crossing(lat = c(T, F)) %>%
  mutate(
    f = "avexpr ~ value",
    f = ifelse(lat, str_c(f, " + lat_abst"), f),
    f = ifelse(name == "exconst_01", str_c(f, " + ind_time"), f)
```

```
  ) %>%
  mutate(
    fit = map2(f, data, ~ lm(formula = .x, data = .y)),
    map_df(fit, ~ broom::tidy(.)[2, c("estimate", "std.error")])
  )%>%
  select(-data, -fit) %>%
  knitr::kable(caption = "Táblázat 3 az általunk összerakott adatokkal futtatva.")
```

8. táblázat: Táblázat 3 az általunk összerakott adatokkal futtatva.

name	lat	f	estimate	std.error
democ_00	FALSE	avexpr ~ value	0.2852108	0.0655908
democ_00	TRUE	avexpr ~ value + lat_abst	0.1449548	0.0577480
euro1900	FALSE	avexpr ~ value	0.0277188	0.0030886
euro1900	TRUE	avexpr ~ value + lat_abst	0.0195029	0.0043813
exconst_00	FALSE	avexpr ~ value	0.3740866	0.0719134
exconst_00	TRUE	avexpr ~ value + lat_abst	0.2051687	0.0652248
exconst_01	FALSE	avexpr ~ value + ind_time	0.0786423	0.0756017
exconst_01	TRUE	avexpr ~ value + lat_abst + ind_time	0.0010908	0.0646732
logem4	FALSE	avexpr ~ value	-0.6467746	0.1152147
logem4	TRUE	avexpr ~ value + lat_abst	-0.4668231	0.1313965

Az OLS becslések reprodukálásából az látszik, hogy elfogadhatóak az adataink.

Táblázat 4

4. táblázat C paneljének reprodukálása (csak az OLS becslések).

A cikkből tudjuk, hogy kik a neo-európaik:

```
neo_europe <- c("AUS", "CAN", "NZL", "USA")
```

```
tibble(model = c("simple", "wo_neo_europes", "wo_africa", "continent_dummy")) %>%
  mutate(model = fct_inorder(model)) %>%
  crossing(lat = c(F, T)) %>%
  bind_rows(tibble(model = "loghj", lat = F)) %>%
  mutate(
    formula = ifelse(model == "loghj", "loghjypl ~ avexpr", "logpgp95 ~ avexpr"),
    formula = ifelse(lat, str_c(formula, " + lat_abst"), formula),
    data = list(ace_data4),
    formula = ifelse(model == "continent_dummy", str_c(formula, " + africa + asia"), formula),
    data = map2(data, model, function(d, m) {
      if (m == "wo_africa") {
        return(filter(d, africa == 0))
      }

      if (m == "wo_neo_europes") {
        return(filter(d, !(shortnam %in% neo_europe)))
      }
      d
    }),
    ols = map2(formula, data, lm),
    map_df(ols, ~ select(filter(broom::tidy(.), term == "avexpr"), ols_term = estimate, ols_se = std.er
```

```
  )%>%
  select(-data, -ols) %>%
  knitr::kable(caption = "Táblázat 4 az Acemoglu által közétett adatokkal futtatva.")
```

9. táblázat: Táblázat 4 az Acemoglu által közétett adatokkal futtatva.

model	lat	formula	ols_term	ols_se
simple	FALSE	logpgp95 ~ avexpr	0.5318713	0.0406212
simple	TRUE	logpgp95 ~ avexpr + lat_abst	0.4634816	0.0554959
wo_neo_europes	FALSE	logpgp95 ~ avexpr	0.5254061	0.0431302
wo_neo_europes	TRUE	logpgp95 ~ avexpr + lat_abst	0.4532272	0.0582558
wo_africa	FALSE	logpgp95 ~ avexpr	0.4516027	0.0409728
wo_africa	TRUE	logpgp95 ~ avexpr + lat_abst	0.4694565	0.0556584
continent_dummy	FALSE	logpgp95 ~ avexpr + africa + asia	0.4151938	0.0402418
continent_dummy	TRUE	logpgp95 ~ avexpr + lat_abst + africa + asia	0.3941342	0.0502543
loghj	FALSE	loghjypl ~ avexpr	0.4462022	0.0388804

A kód a 4. táblázathoz megfelelő, Acemoglu adataival jól adja vissza a cikkben közölt számokat.

```
check_df_4 <- ace_data2 %>%
  transmute(country = shortnam, avexpr) %>%
  full_join(logpgp95_wb, by = c("country" = "geo")) %>%
  full_join(loghjypl_pwt, by = c("country" = "geo")) %>%
  full_join(
    laporta_qgov %>%
      transmute(country = str_to_upper(flopcode), lat_abst) %>%
      mutate(
        continent = countrycode::countrycode(country, "iso3c", "continent")
      ) %>%
      transmute(
        country, lat_abst, continent = ifelse(continent %in% c("Americas", "Asia", "Africa"), continent
      )
  ) %>%
  transmute(
    shortnam = country,
    africa = ifelse(continent == "Africa", 1, 0),
    lat_abst,
    avexpr,
    logpgp95 = logpgp95,
    other = ifelse(continent == "other", 1, 0),
    asia = ifelse(continent == "Asia", 1, 0),
    loghjypl = loghjypl_pwt,
    baseco = ifelse(shortnam %in% baseco, 1, NA),
    rich4 = ifelse(shortnam %in% neo_europe, 1, 0)
  ) %>%
  right_join(logem4)
```

```
tibble(model = c("simple", "wo_neo_europes", "wo_africa", "continent_dummy")) %>%
  mutate(model = fct_inorder(model)) %>%
  crossing(lat = c(F, T)) %>%
  bind_rows(tibble(model = "loghj", lat = F)) %>%
  mutate(
```

```
    formula = ifelse(model == "loghj", "loghjypl ~ avexpr", "logpgp95 ~ avexpr"),
    formula = ifelse(lat, str_c(formula, " + lat_abst"), formula),
    data = list(check_df_4),
    formula = ifelse(model == "continent_dummy", str_c(formula, " + africa + asia"), formula),
    data = map2(data, model, function(d, m) {
      if (m == "wo_africa") {
        return(filter(d, africa == 0))
      }

      if (m == "wo_neo_europes") {
        return(filter(d, !(shortnam %in% neo_europe)))
      }
      d
    }),
    ols = map2(formula, data, lm),
    map_df(ols, ~ select(filter(broom::tidy(.), term == "avexpr"), ols_term = estimate, ols_se = std.er
  ) %>%
  select(-data, -ols) %>%
  knitr::kable(caption = "Táblázat 4 az általunk összerakott adatokkal futtatva.")
```

10. táblázat: Táblázat 4 az általunk összerakott adatokkal futtatva.

model	lat	formula	ols_term	ols_se
simple	FALSE	logpgp95 ~ avexpr	0.4863571	0.0502057
simple	TRUE	logpgp95 ~ avexpr + lat_abst	0.3945373	0.0637200
wo_neo_europes	FALSE	logpgp95 ~ avexpr	0.4828435	0.0525107
wo_neo_europes	TRUE	logpgp95 ~ avexpr + lat_abst	0.3870106	0.0663707
wo_africa	FALSE	logpgp95 ~ avexpr	0.3765598	0.0503862
wo_africa	TRUE	logpgp95 ~ avexpr + lat_abst	0.3518110	0.0641558
continent_dummy	FALSE	logpgp95 ~ avexpr + africa + asia	0.3675238	0.0485631
continent_dummy	TRUE	logpgp95 ~ avexpr + lat_abst + africa + asia	0.3278276	0.0574870
loghj	FALSE	loghjypl ~ avexpr	0.4091236	0.0444821

Mindegiyk modell esetében .1 értékkel alacsonyabb az avexpr együtthatója.

Táblázat 5

5. táblázat C paneljének reprodukálása (csak az OLS becslések).

```
tibble(model = 1:9) %>%
  mutate(
    data = list(filter(ace_data5, baseco == 1)),
    data = map2(data, model, function(d, m) {
      if (m %in% 3:4) {
        return(filter(d, f_brit == 1))
      }
      return(d)
    }),
    formula = "logpgp95 ~ avexpr",
    formula = ifelse(model %in% c(2, 4, 6, 8, 9), str_c(formula, " + lat_abst"), formula),
    formula = ifelse(model %in% 1:2, str_c(formula, " + f_brit + f_french"), formula),
    formula = ifelse(model %in% 5:6, str_c(formula, " + sjlofr"), formula),
```

```
    formula = ifelse(model == 9, str_c(formula, " + f_french + sjlofr"), formula),
    ols = map2(formula, data, lm),
    map_df(ols, ~ select(filter(broom::tidy(.), term == "avexpr"), ols_term = estimate, ols_se = std.er
  ) %>%
  select(-data, -ols) %>%
  knitr::kable(caption = "Táblázat 5 az Acemoglu által közétett adatokkal futtatva.")
```

11. táblázat: Táblázat 5 az Acemoglu által közétett adatokkal futtatva.

model	formula	ols_term	ols_se
1	logpgp95 ~ avexpr + f_brit + f_french	0.5276210	0.0652136
2	logpgp95 ~ avexpr + lat_abst + f_brit + f_french	0.4652465	0.0666210
3	logpgp95 ~ avexpr	0.6057617	0.0930560
4	logpgp95 ~ avexpr + lat_abst	0.5489083	0.1081558
5	logpgp95 ~ avexpr + sjlofr	0.5626974	0.0636217
6	logpgp95 ~ avexpr + lat_abst + sjlofr	0.5084451	0.0666312
7	logpgp95 ~ avexpr	0.5221070	0.0611850
8	logpgp95 ~ avexpr + lat_abst	0.4678871	0.0641642
9	logpgp95 ~ avexpr + lat_abst + f_french + sjlofr	0.4817401	0.0650243

A 4. modellnél nem tudni mi történik. A cikk alapján számunkra nem derül ki, hogy miben különbözik a 3. modelltől. Próbáltuk, hogy hátha a lattitude változó szerepeltetésének elfelejtése okozta a gondot, de nem jutottunk jobb eredményre.

```
colonial_dummies <- readxl::read_excel("data/colonial dummies.xlsx")
```

```
ace_data5 %>%
  select(shortnam, f_brit) %>%
  left_join(colonial_dummies, by = c("shortnam" = "Code")) %>%
  replace_na(list(British = 0)) %>%
  count(f_brit, British)
```

f_brit	British	n
0	?	1
0	0	109
0	1	2
1	?	2
1	0	25
1	1	23
NA	0	1

A Wikipédiáról legyűjtött kolónia dummy változók közel sem egyeznek meg a cikkben publikálttal, így ezeket is Acemoglutól kell átvennünk.

```
check_df_5 <- laporta_qgov %>%
  transmute(shortnam = str_to_upper(flopcode), catho80, muslim80, no_cpm80, sjlofr = legor_fr, lat_abst)
  right_join(select(ace_data5, shortnam, f_brit, f_french, avexpr)) %>%
  left_join(logem4) %>%
  left_join(logpgp95_wb, by = c("shortnam" = "geo")) %>%
  filter(shortnam %in% baseco)
```

```
tibble(model = 1:9) %>%
  mutate(
    data = list(check_df_5),
    data = map2(data, model, function(d, m) {
      if (m %in% 3:4) {
        return(filter(d, f_brit == 1))
      }
      return(d)
    }),
    formula = "logpgp95 ~ avexpr",
    formula = ifelse(model %in% c(2, 4, 6, 8, 9), str_c(formula, " + lat_abst"), formula),
    formula = ifelse(model %in% 1:2, str_c(formula, " + f_brit + f_french"), formula),
    formula = ifelse(model %in% 5:6, str_c(formula, " + sjlofr"), formula),
    formula = ifelse(model == 9, str_c(formula, " + f_french + sjlofr"), formula),
    ols = map2(formula, data, lm),
    map_df(ols, ~ select(filter(broom::tidy(.), term == "avexpr"),
                         ols_term = estimate, ols_se = std.error))
  )%>%
  select(-data, -ols) %>%
  knitr::kable(caption = "Táblázat 5 az Acemoglu által közétett adatokkal futtatva.")
```

13. táblázat: Táblázat 5 az Acemoglu által közétett adatokkal futtatva.

model	formula	ols_term	ols_se
1	logpgp95 ~ avexpr + f_brit + f_french	0.5066329	0.0760667
2	logpgp95 ~ avexpr + lat_abst + f_brit + f_french	0.4707181	0.0770829
3	logpgp95 ~ avexpr	0.5875402	0.1038970
4	logpgp95 ~ avexpr + lat_abst	0.5541073	0.1160129
5	logpgp95 ~ avexpr + sjlofr	0.5515663	0.0741962
6	logpgp95 ~ avexpr + lat_abst + sjlofr	0.5204580	0.0763341
7	logpgp95 ~ avexpr	0.5245166	0.0714193
8	logpgp95 ~ avexpr + lat_abst	0.4942784	0.0730202
9	logpgp95 ~ avexpr + lat_abst + f_french + sjlofr	0.4858679	0.0760322

Az előzőekben leírt 4. modell problémáján túl a 6. modell is némileg eltérő eredményt ad a mi adatainkkal.

Táblázat 6

```
laporta_qgov %>%
  transmute(shortnam = str_to_upper(flopcode), avelf) %>%
  right_join(ace_data6, by = "shortnam") %>%
  select(1:3) %>%
  knitr::kable()
```

shortnam	avelf.x	avelf.y
AFG	0.4484412	0.4484412
AGO	0.7727550	0.7727550
ARE	NA	NA
ARG	0.1769318	0.1769318

shortnam	avelf.x	avelf.y
ARM	NA	NA
AUS	0.1127971	0.1127971
AUT	0.0332123	0.0332123
AZE	NA	NA
BDI	0.0133333	0.0133333
BEL	0.3638344	0.3638344
BEN	0.6831192	0.6831192
BFA	0.5467181	0.5467181
BGD	0.0000000	0.0000000
BGR	0.1157407	0.1157407
BHR	NA	NA
BHS	0.0000000	0.0000000
BIH	NA	NA
BLR	NA	NA
BLZ	0.4090909	0.4090909
BOL	0.5994118	0.5994118
BRA	0.0557809	0.0557809
BRB	0.0733333	0.0733333
BTN	0.4375000	0.4375000
BWA	0.3775000	0.3775000
CAF	0.7855555	0.7855555
CAN	0.3761876	0.3761876
CHE	0.3075980	0.3075980
CHL	0.0506250	0.0506250
CHN	0.2333264	0.2333264
CIV	0.8565299	0.8565299
CMR	0.8520455	0.8520455
COG	0.6692753	0.6692753
COL	0.0557960	0.0557960
COM	1.0000000	1.0000000
CPV	0.3750000	0.3750000
CRI	0.0532143	0.0532143
CZE	NA	NA
DEU	0.0437553	0.0437553
DJI	0.7142857	0.7142857
DNK	0.0275000	0.0275000
DOM	0.0108306	0.0108306
DZA	0.2937240	0.2937240
ECU	0.3253571	0.3253571
EGY	0.0230618	0.0230618
ERI	NA	NA
ESP	0.2745161	0.2745161
EST	NA	NA
ETH	0.6770825	0.6770825
FIN	0.1050000	0.1050000
FJI	0.8000000	0.8000000
FRA	0.1454839	0.1454839
GAB	0.7966667	0.7966667
GBR	0.1062846	0.1062846
GEO	NA	NA
GHA	0.7060708	0.7060708
GIN	0.7598300	0.7598300

shortnam	avelf.x	avelf.y
GMB	0.7803704	0.7803704
GNB	0.8500000	0.8500000
GRC	0.0777778	0.0777778
GTM	0.4766667	0.4766667
GUY	0.2377778	0.2377778
HKG	0.2368212	0.2368212
HND	0.0973593	0.0973593
HRV	NA	NA
HTI	0.0644048	0.0644048
HUN	0.0650794	0.0650794
IDN	0.6905679	0.7421975
IND	0.7421975	0.6905679
IRL	0.0903571	0.0903571
IRN	NA	NA
IRQ	NA	NA
ISL	0.1000000	0.1000000
ISR	0.3270557	0.3270557
ITA	0.0388682	0.0388682
JAM	0.0125000	0.0125000
JOR	0.0297194	0.0297194
JPN	0.0098803	0.0098803
KAZ	NA	NA
KEN	0.8269649	0.8269649
KGZ	NA	NA
KOR	0.0000000	0.0000000
KWT	NA	NA
LAO	0.2500000	0.2500000
LBR	0.8030553	0.8030553
LBY	0.1214057	0.1214057
LKA	0.3257469	0.3257469
LSO	0.2097619	0.2097619
LTU	NA	NA
LUX	0.2166667	0.2166667
LVA	NA	NA
MAR	0.3479719	0.3479719
MDA	NA	NA
MDG	0.0627486	0.0627486
MEX	0.1740741	0.1740741
MKD	NA	NA
MLI	0.8085691	0.8085691
MLT	0.1033333	0.1033333
MMR	0.3839685	0.3839685
MNG	0.0737439	0.0737439
MOZ	0.7863327	0.7863327
MRT	0.2700000	0.2700000
MUS	0.7084848	0.7084848
MWI	0.6224373	0.6224373
MYS	0.6103704	0.6103704
NAM	0.7283334	0.7283334
NER	0.7329143	0.7329143
NGA	0.8567090	0.8567090
NIC	0.0991796	0.0991796

shortnam	avelf.x	avelf.y
NLD	0.0634354	0.0634354
NOR	0.0698837	0.0698837
NPL	0.4500000	0.4500000
NZL	0.1475699	0.1475699
OMN	NA	NA
PAK	0.6215599	0.6215599
PAN	0.1908333	0.1908333
PER	0.4315909	0.4315909
PHL	0.7238207	0.7238207
PNG	0.8027273	0.8027273
POL	0.0390492	0.0390492
PRK	NA	NA
PRT	0.0025000	0.0025000
PRY	0.4111111	0.4111111
QAT	NA	NA
ROM	0.1220442	0.1220442
RUS	NA	NA
RWA	0.0608621	0.0608621
SAU	NA	NA
SDN	0.5121800	0.5121800
SEN	0.7788682	0.7788682
SGP	0.3215287	0.3215287
SLE	0.8129566	0.8129566
SLV	0.0514286	0.0514286
SOM	0.0791160	0.0791160
STP	0.0000000	0.0000000
SUR	0.7500000	0.7500000
SVK	NA	NA
SVN	NA	NA
SWE	0.0650368	0.0650368
SWZ	0.0000000	0.0000000
SYR	0.0947920	0.0947920
TCD	0.6662003	0.6662003
TGO	0.7284868	0.7284868
THA	0.3569488	0.3569488
TJK	NA	NA
TKM	NA	NA
TTO	0.2312698	0.2312698
TUN	0.0702522	0.0702522
TUR	0.1635684	0.1635684
TZA	0.8902469	0.8902469
UGA	0.8357907	0.8357907
UKR	NA	NA
URY	0.0666667	0.0666667
USA	0.2090314	0.2090314
UZB	NA	NA
VEN	0.0525000	0.0525000
VNM	0.1176471	0.1176471
YEM	0.0122000	0.0122000
YUG	0.2777778	0.2777778
ZAF	0.8309818	0.8309818
ZAR	0.8722529	0.8722529

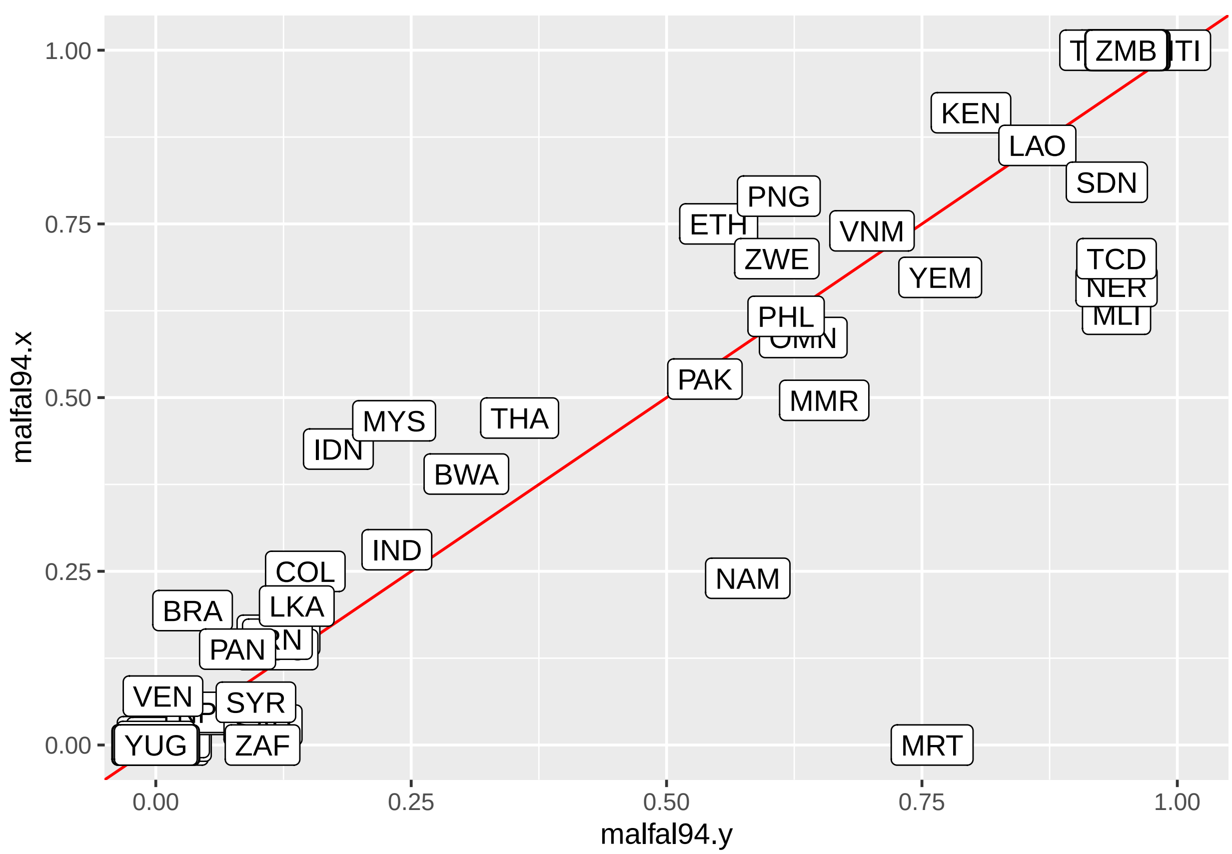
shortnam	avelf.x	avelf.y
ZMB	0.8293977	0.8293977
ZWE	0.5986364	0.5986364
YUG	0.2777778	0.2777778
TWN	0.2550563	NA

A két cikkben lévő adatok az átlagos fregmentációra vonatkozóan megegyeznek (nem lepődtünk meg, ezt hivatkozta).

Táblázat 7

```
malfal94 <- haven::read_dta(str_c(WD, "/data/gallup_malaria94.dta"))
```

```
malfal94 %>%
  select(geo = wbcode, malfal94) %>%
  left_join(
    ace_data7, by = c("geo" = "shortnam")
  ) %>%
  select(geo, starts_with("mal")) %>%
  ggplot(aes(malfal94.y, malfal94.x, label = geo)) +
  geom_abline(color = "red") +
  geom_label()
```



Sachs geography

Sajnos a változókat sehogy nem sikerült beolvasni. Az alábbi kód kiszedi a számokat, de a táblázat hiányai miatt a sorokat egyesével kell a megfelelő helyre tolni.

```
list.files(str_c(WD, "/data/mc"), full.names = T) %>%
  map(~ tesseract::ocr(., engine = tesseract::tesseract("eng"))) %>% # read from image
  map(function(imag) {
    imag %>%
      str_split(pattern = "\\\n") %>%
      map(~ str_split(., pattern = " ")) %>%
      map(enframe) %>%
      .[[1]] %>%
      mutate(
        value = map(value, str_remove_all, "«"),
        value = map(value, str_remove_all, "[*]"),
        value = map(value, str_remove_all, "©"),
        numbers = map(value, parse_double),
        nonna = map(numbers, ~ which(!is.na(.))),
        min_nonna = map_dbl(nonna, min),
      ) %>%
      filter(is.finite(min_nonna)) %>%
      mutate(
        country = map2_chr(value, min_nonna, ~ str_c(.x[1:(.y-1)], collapse = " ")),
        # numbers = map2(numbers, min_nonna, ~ .x[.y:length(.x)]),
        numbers = map(value, ~ keep(., function(u) str_detect(u, "\\d") | str_detect(u, "x") | str_dete
        numbers = map(numbers, as.numeric)
      )
  }
  ) %>%
  bind_rows() %>%
  select(country, numbers) %>%
  unnest(numbers) %>%
  group_by(country) %>%
  group_modify(~ mutate(.x, i = row_number())) %>%
  ungroup() %>%
  pivot_wider(names_from = i, values_from = numbers) %>%
  head()
```

country	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
'Afghanistan	603	6.3969	[illegible]	[illegible]	[illegible]	16.4	0.0000	0.366667	-6.552970	1.000	0.0000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
'Cameroon	2116	7.654	[illegible]	1.00	5473	74.4	5.63479	-24.433330	NA	0.066667	-3.96009	280	1	2.00000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
'Saudi Arabia	9910	9.2017	[illegible]	[illegible]	[illegible]	[illegible]	0.122	[illegible]	[illegible]	[illegible]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
'Sweden	18540	0.827	[illegible]	0.00	7855	54.0	24.000	[illegible]	[illegible]	-6.907760	0.0000	[illegible]	18540	0.827	[illegible]	[illegible]	78.554	NA	24	0.301	[illegible]	[illegible]	[illegible]	[illegible]	NA
'Switzerland	25860	11.16	[illegible]	0.00	7865	57.0	59.000	0.0000	[illegible]	-6.907760	0.0000	[illegible]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Albania	1210	7.098	[illegible]	0.00	7275	30.0	15.700	0.85278	[illegible]	-4.942750	0.0000	[illegible]	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA