

T412006-17320268-assignment2

17320268

2022-11-02

Set-up

```
rm(list = ls())
library(tidyverse)
library(knitr)
library(ggmosaic)
setwd("F:/MAPEC/Applied methods/assignment 2")
data_qog <- read_csv("2022-10-31-T412006-assignment2-data.csv")
names(data_qog)
```

```
## [1] "ccode"      "cname"      "year"
## [4] "ht_region"  "wdi_area"   "wdi_pop"
## [7] "wdi_popden" "br_dem"     "br_elect"
## [10] "chga_hinst" "ht_regtype" "p_polity2"
## [13] "wdi_gnicon2010" "wdi_gnicapcon2010" "wdi_gdpcapcon2010"
## [16] "wdi_lifexp"  "wdi_litrade" "undp_hdi"
## [19] "wdi_expmil"  "wdi_internet"
```

Q1

```
dim(data_qog)
```

```
## [1] 12156    20
```

```
ncol(data_qog)
```

```
## [1] 20
```

```
nrow(data_qog)
```

```
## [1] 12156
```

Q2

```
names(data_qog)
```

```
## [1] "ccode"          "cname"          "year"
## [4] "ht_region"      "wdi_area"       "wdi_pop"
## [7] "wdi_popden"     "br_dem"         "br_elect"
## [10] "chga_hinst"     "ht_regtype"     "p_polity2"
## [13] "wdi_gnicon2010" "wdi_gnicapcon2010" "wdi_gdpcapcon2010"
## [16] "wdi_lifexp"     "wdi_litrاد"     "undp_hdi"
## [19] "wdi_expmil"     "wdi_internet"
```

Q3

Variable name	Short description	Information
ccode	Country code	Country code with the ISO-3166-1 standard
cname	Country name	Country name
year	Year	
ht_region	Region of the Country	Tenfold politico-geographic classification of world region from 1 to 10
wdi_area	Land area (sq.km)	Country's total area excluding area under inland water bodies, exclusive economic zones and claims to continental shelf
wdi_pop	Total population	All residents regardless of legal status or citizenship (midyear estimates)
wdi_popden	Population density	People per sq. km of land area. Midyear population divided by land area in square kilometers
br_dem	Is the country a democracy	Dummy variable indicator of democracy based on minimalist definition (if there is free and fair election, peaceful turnover of officers)
br_elect	Typology of political institutions	Alternative democracy indicator capturing degree of multi-party competition.
chga_hinst	Regime Institutions	Six-fold classification of political regimes
ht_regtype	Regime Type	Qualitative variable representing the political regime of each country based on 26 levels
p_polity2	Revised Combined Polity Score	Ordinal variable: range from -10 (strongly autocratic) to +10 (strongly democratic)
wdi_gnicon2010	GNI (constant 2010 US dollar)	Gross national income: sum of value added by all resident producers plus any product taxes less subsidies
wdi_gnicapcon2010	GNI per capita (constant 2010 US dollar)	GNI divided by midyear population
wdi_gdpcapcon2010	GDP per capita (constant 2010 US dollar)	Gross domestic product divided by midyear population. GDP is the sum of gross value added by residents producers plus product taxes minus subsidies
wdi_lifexp	Life expectancy at birth, total (years)	Number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life
wdi_litrاد	Literacy rate, adult total (% of people ages 15 and above)	Percentage of the population above 15 who can understand, read and write a short simple statement
undp_hdi	Human Development Index	Summary measure of average achievement in health, knowledge and decent standards of living

Variable name	Short description	Information
wdi_exp	Military expenditure (% of GDP)	All current and capital expenditures on the armed forces
wdi_internet	Individuals using the Internet (% of population)	Internet users who have used Internet in the last 3 months

Q4

```
data_qog %>%
  select(year) %>%
  summarise(
    minium_year = min(year),
    max_year = max(year)
  )
```

```
## # A tibble: 1 x 2
##   minium_year max_year
##         <dbl>   <dbl>
## 1         1960     2020
```

Q5

```
sort(unique(data_qog$name))
```

```
## [1] "Afghanistan"
## [2] "Albania"
## [3] "Algeria"
## [4] "Andorra"
## [5] "Angola"
## [6] "Antigua and Barbuda"
## [7] "Argentina"
## [8] "Armenia"
## [9] "Australia"
## [10] "Austria"
## [11] "Azerbaijan"
## [12] "Bahamas (the)"
## [13] "Bahrain"
## [14] "Bangladesh"
## [15] "Barbados"
## [16] "Belarus"
## [17] "Belgium"
## [18] "Belize"
## [19] "Benin"
## [20] "Bhutan"
## [21] "Bolivia (Plurinational State of)"
## [22] "Bosnia and Herzegovina"
## [23] "Botswana"
```

[24] "Brazil"
[25] "Brunei Darussalam"
[26] "Bulgaria"
[27] "Burkina Faso"
[28] "Burundi"
[29] "Cabo Verde"
[30] "Cambodia"
[31] "Cameroon"
[32] "Canada"
[33] "Central African Republic (the)"
[34] "Chad"
[35] "Chile"
[36] "China"
[37] "Colombia"
[38] "Comoros (the)"
[39] "Congo (the Democratic Republic of the)"
[40] "Congo (the)"
[41] "Costa Rica"
[42] "Côte d'Ivoire"
[43] "Croatia"
[44] "Cuba"
[45] "Cyprus"
[46] "Czechia"
[47] "Czechoslovakia"
[48] "Denmark"
[49] "Djibouti"
[50] "Dominica"
[51] "Dominican Republic (the)"
[52] "Ecuador"
[53] "Egypt"
[54] "El Salvador"
[55] "Equatorial Guinea"
[56] "Eritrea"
[57] "Estonia"
[58] "Eswatini"
[59] "Ethiopia"
[60] "Fiji"
[61] "Finland"
[62] "France"
[63] "Gabon"
[64] "Gambia (the)"
[65] "Georgia"
[66] "German Democratic Republic"
[67] "Germany"
[68] "Ghana"
[69] "Greece"
[70] "Grenada"
[71] "Guatemala"
[72] "Guinea"
[73] "Guinea-Bissau"
[74] "Guyana"
[75] "Haiti"
[76] "Honduras"
[77] "Hungary"

[78] "Iceland"
 ## [79] "India"
 ## [80] "Indonesia"
 ## [81] "Iran (Islamic Republic of)"
 ## [82] "Iraq"
 ## [83] "Ireland"
 ## [84] "Israel"
 ## [85] "Italy"
 ## [86] "Jamaica"
 ## [87] "Japan"
 ## [88] "Jordan"
 ## [89] "Kazakhstan"
 ## [90] "Kenya"
 ## [91] "Kiribati"
 ## [92] "Korea (the Democratic People's Republic of)"
 ## [93] "Korea (the Republic of)"
 ## [94] "Kuwait"
 ## [95] "Kyrgyzstan"
 ## [96] "Lao People's Democratic Republic (the)"
 ## [97] "Latvia"
 ## [98] "Lebanon"
 ## [99] "Lesotho"
 ## [100] "Liberia"
 ## [101] "Libya"
 ## [102] "Liechtenstein"
 ## [103] "Lithuania"
 ## [104] "Luxembourg"
 ## [105] "Madagascar"
 ## [106] "Malawi"
 ## [107] "Malaysia"
 ## [108] "Maldives"
 ## [109] "Mali"
 ## [110] "Malta"
 ## [111] "Marshall Islands"
 ## [112] "Mauritania"
 ## [113] "Mauritius"
 ## [114] "Mexico"
 ## [115] "Micronesia (Federated States of)"
 ## [116] "Moldova (the Republic of)"
 ## [117] "Monaco"
 ## [118] "Mongolia"
 ## [119] "Montenegro"
 ## [120] "Morocco"
 ## [121] "Mozambique"
 ## [122] "Myanmar"
 ## [123] "Namibia"
 ## [124] "Nauru"
 ## [125] "Nepal"
 ## [126] "Netherlands (the)"
 ## [127] "New Zealand"
 ## [128] "Nicaragua"
 ## [129] "Niger (the)"
 ## [130] "Nigeria"
 ## [131] "North Macedonia"

[132] "Norway"
[133] "Oman"
[134] "Pakistan"
[135] "Palau"
[136] "Panama"
[137] "Papua New Guinea"
[138] "Paraguay"
[139] "Peru"
[140] "Philippines (the)"
[141] "Poland"
[142] "Portugal"
[143] "Qatar"
[144] "Romania"
[145] "Russian Federation (the)"
[146] "Rwanda"
[147] "Saint Kitts and Nevis"
[148] "Saint Lucia"
[149] "Saint Vincent and the Grenadines"
[150] "Samoa"
[151] "San Marino"
[152] "Sao Tome and Principe"
[153] "Saudi Arabia"
[154] "Senegal"
[155] "Serbia"
[156] "Serbia and Montenegro"
[157] "Seychelles"
[158] "Sierra Leone"
[159] "Singapore"
[160] "Slovakia"
[161] "Slovenia"
[162] "Solomon Islands"
[163] "Somalia"
[164] "South Africa"
[165] "South Sudan"
[166] "Spain"
[167] "Sri Lanka"
[168] "Sudan"
[169] "Sudan (the)"
[170] "Suriname"
[171] "Sweden"
[172] "Switzerland"
[173] "Syrian Arab Republic (the)"
[174] "Taiwan (Province of China)"
[175] "Tajikistan"
[176] "Tanzania, the United Republic of"
[177] "Thailand"
[178] "Tibet"
[179] "Timor-Leste"
[180] "Togo"
[181] "Tonga"
[182] "Trinidad and Tobago"
[183] "Tunisia"
[184] "Turkey"
[185] "Turkmenistan"

```
## [186] "Tuvalu"
## [187] "Uganda"
## [188] "Ukraine"
## [189] "United Arab Emirates (the)"
## [190] "United Kingdom of Great Britain and Northern Ireland (the)"
## [191] "United States of America (the)"
## [192] "Uruguay"
## [193] "USSR"
## [194] "Uzbekistan"
## [195] "Vanuatu"
## [196] "Venezuela (Bolivarian Republic of)"
## [197] "Viet Nam"
## [198] "Vietnam, North"
## [199] "Vietnam, South"
## [200] "Yemen"
## [201] "Yemen Democratic"
## [202] "Yugoslavia"
## [203] "Zambia"
## [204] "Zimbabwe"
```

Data transformations

Q6

```
str(data_qog$ht_region)
```

```
##  num [1:12156] 8 8 8 8 8 8 8 8 8 8 ...
```

Variable `ht_region` is stored as a numerical variable. This does not make sense, because the variable is a geographical classification of world region from 1 to 10 each number representing a region, it should hence be stored as a nominal variable.

```
data_qog$ht_region <- factor(data_qog$ht_region, levels = c(1:10))
unique(data_qog$ht_region)
```

```
## [1] 8 1 <NA> 3 5 4 10 2 9 7 6
## Levels: 1 2 3 4 5 6 7 8 9 10
```

```
head(data_qog)
```

```
## # A tibble: 6 x 20
##   ccode cname      year ht_region wdi_area wdi_pop wdi_popden br_dem br_elect
##   <dbl> <chr>      <dbl> <fct>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     4 Afghanistan 1960 8          NA 8996967    NA         0         2
## 2     4 Afghanistan 1961 8        652860 9169406   14.0        0         2
## 3     4 Afghanistan 1962 8        652860 9351442   14.3        0         2
## 4     4 Afghanistan 1963 8        652860 9543200   14.6        0         2
## 5     4 Afghanistan 1964 8        652860 9744772   14.9        0         2
## 6     4 Afghanistan 1965 8        652860 9956318   15.3        0         2
```

```
## # ... with 11 more variables: chga_hinst <dbl>, ht_regtype <dbl>,
## #   p_polity2 <dbl>, wdi_gnicon2010 <dbl>, wdi_gnicapcon2010 <dbl>,
## #   wdi_gdpcapcon2010 <dbl>, wdi_lifexp <dbl>, wdi_litrad <dbl>,
## #   undp_hdi <dbl>, wdi_expmil <dbl>, wdi_internet <dbl>
```

Q7

```
data_cntr_reg <- data_qog %>%
  filter(year == 2015) %>%
  select(cname, ht_region)
head(data_cntr_reg)
```

```
## # A tibble: 6 x 2
##   cname          ht_region
##   <chr>         <fct>
## 1 Afghanistan      8
## 2 Albania           1
## 3 Algeria           3
## 4 Andorra           5
## 5 Angola            4
## 6 Antigua and Barbuda 10
```

Q8

```
data_cntr_reg %>%
  arrange(ht_region, cname) %>%
  kable()
```

cname	ht_region
Albania	1
Armenia	1
Azerbaijan	1
Belarus	1
Bosnia and Herzegovina	1
Bulgaria	1
Croatia	1
Czechia	1
Czechoslovakia	1
Estonia	1
Georgia	1
Hungary	1
Kazakhstan	1
Kyrgyzstan	1
Latvia	1
Lithuania	1
Moldova (the Republic of)	1
Montenegro	1
North Macedonia	1

cname	ht_region
Poland	1
Romania	1
Russian Federation (the)	1
Serbia	1
Serbia and Montenegro	1
Slovakia	1
Slovenia	1
Tajikistan	1
Turkmenistan	1
Ukraine	1
USSR	1
Uzbekistan	1
Argentina	2
Bolivia (Plurinational State of)	2
Brazil	2
Chile	2
Colombia	2
Costa Rica	2
Cuba	2
Dominican Republic (the)	2
Ecuador	2
El Salvador	2
Guatemala	2
Haiti	2
Honduras	2
Mexico	2
Nicaragua	2
Panama	2
Paraguay	2
Peru	2
Uruguay	2
Venezuela (Bolivarian Republic of)	2
Algeria	3
Bahrain	3
Cyprus	3
Egypt	3
Iran (Islamic Republic of)	3
Iraq	3
Israel	3
Jordan	3
Kuwait	3
Lebanon	3
Libya	3
Morocco	3
Oman	3
Qatar	3
Saudi Arabia	3
Syrian Arab Republic (the)	3
Tunisia	3
Turkey	3
United Arab Emirates (the)	3
Yemen	3

cname	ht_region
Angola	4
Benin	4
Botswana	4
Burkina Faso	4
Burundi	4
Cabo Verde	4
Cameroon	4
Central African Republic (the)	4
Chad	4
Comoros (the)	4
Congo (the Democratic Republic of the)	4
Congo (the)	4
Côte d'Ivoire	4
Djibouti	4
Equatorial Guinea	4
Eritrea	4
Eswatini	4
Ethiopia	4
Gabon	4
Gambia (the)	4
Ghana	4
Guinea	4
Guinea-Bissau	4
Kenya	4
Lesotho	4
Liberia	4
Madagascar	4
Malawi	4
Mali	4
Mauritania	4
Mauritius	4
Mozambique	4
Namibia	4
Niger (the)	4
Nigeria	4
Rwanda	4
Sao Tome and Principe	4
Senegal	4
Seychelles	4
Sierra Leone	4
Somalia	4
South Africa	4
South Sudan	4
Sudan (the)	4
Tanzania, the United Republic of	4
Togo	4
Uganda	4
Zambia	4
Zimbabwe	4
Andorra	5
Australia	5
Austria	5

cname	ht_region
Belgium	5
Canada	5
Denmark	5
Finland	5
France	5
Germany	5
Greece	5
Iceland	5
Ireland	5
Italy	5
Liechtenstein	5
Luxembourg	5
Malta	5
Monaco	5
Netherlands (the)	5
New Zealand	5
Norway	5
Portugal	5
San Marino	5
Spain	5
Sweden	5
Switzerland	5
United Kingdom of Great Britain and Northern Ireland (the)	5
United States of America (the)	5
China	6
Japan	6
Korea (the Democratic People's Republic of)	6
Korea (the Republic of)	6
Mongolia	6
Taiwan (Province of China)	6
Brunei Darussalam	7
Cambodia	7
Indonesia	7
Lao People's Democratic Republic (the)	7
Malaysia	7
Myanmar	7
Philippines (the)	7
Singapore	7
Thailand	7
Timor-Leste	7
Viet Nam	7
Afghanistan	8
Bangladesh	8
Bhutan	8
India	8
Maldives	8
Nepal	8
Pakistan	8
Sri Lanka	8
Tibet	8
Fiji	9
Kiribati	9

cname	ht_region
Marshall Islands	9
Micronesia (Federated States of)	9
Nauru	9
Palau	9
Papua New Guinea	9
Samoa	9
Solomon Islands	9
Tonga	9
Tuvalu	9
Vanuatu	9
Antigua and Barbuda	10
Bahamas (the)	10
Barbados	10
Belize	10
Dominica	10
Grenada	10
Guyana	10
Jamaica	10
Saint Kitts and Nevis	10
Saint Lucia	10
Saint Vincent and the Grenadines	10
Suriname	10
Trinidad and Tobago	10

Q9

```
data_qog %>%
  filter(ht_region == 5) %>%
  distinct(cname)
```

```
## # A tibble: 27 x 1
##   cname
##   <chr>
## 1 Andorra
## 2 Australia
## 3 Austria
## 4 Belgium
## 5 Canada
## 6 Denmark
## 7 Finland
## 8 France
## 9 Germany
## 10 Greece
## # ... with 17 more rows
```

Q10

```

data_cntr_reg %>%
  mutate(
    my_region = ifelse(ht_region %in% c(2, 10), "Americas",
                      ifelse(cname %in% c("Canada", "United States of America (the)"), "Americas",
                            ifelse(ht_region %in% c(6, 7, 8) | cname %in% c("Tajikistan", "Turkmenistan",
                                ifelse(ht_region %in% c(1, 5), "Europe",
                                      ifelse(ht_region == 3 | cname %in% c("Turkey", "Israel", "Cyprus"),
                                            ifelse(ht_region == 4, "Sub-Saharan Africa",
                                                  ifelse(ht_region == 9 | cname %in% c("Australia", "New Zealand"), "Oceania")
                                )
                            )
                      )
  ) -> data_cntr_reg

data_cntr_reg %>%
  arrange(ht_region, cname) %>%
  kable()

```

cname	ht_region	my_region
Albania	1	Europe
Armenia	1	Europe
Azerbaijan	1	Europe
Belarus	1	Europe
Bosnia and Herzegovina	1	Europe
Bulgaria	1	Europe
Croatia	1	Europe
Czechia	1	Europe
Czechoslovakia	1	Europe
Estonia	1	Europe
Georgia	1	Europe
Hungary	1	Europe
Kazakhstan	1	Asia
Kyrgyzstan	1	Asia
Latvia	1	Europe
Lithuania	1	Europe
Moldova (the Republic of)	1	Europe
Montenegro	1	Europe
North Macedonia	1	Europe
Poland	1	Europe
Romania	1	Europe
Russian Federation (the)	1	Europe
Serbia	1	Europe
Serbia and Montenegro	1	Europe
Slovakia	1	Europe
Slovenia	1	Europe
Tajikistan	1	Asia
Turkmenistan	1	Asia
Ukraine	1	Europe
USSR	1	Europe
Uzbekistan	1	Asia
Argentina	2	Americas
Bolivia (Plurinational State of)	2	Americas
Brazil	2	Americas
Chile	2	Americas
Colombia	2	Americas

cname	ht_region	my_region
Costa Rica	2	Americas
Cuba	2	Americas
Dominican Republic (the)	2	Americas
Ecuador	2	Americas
El Salvador	2	Americas
Guatemala	2	Americas
Haiti	2	Americas
Honduras	2	Americas
Mexico	2	Americas
Nicaragua	2	Americas
Panama	2	Americas
Paraguay	2	Americas
Peru	2	Americas
Uruguay	2	Americas
Venezuela (Bolivarian Republic of)	2	Americas
Algeria	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Bahrain	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Cyprus	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Egypt	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Iran (Islamic Republic of)	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Iraq	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Israel	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Jordan	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Kuwait	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Lebanon	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Libya	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Morocco	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Oman	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Qatar	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Saudi Arabia	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Syrian Arab Republic (the)	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Tunisia	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Turkey	3	North Africa & Middle East (including Israel, Turkey & Cyprus)

cname	ht_region	my_region
United Arab Emirates (the)	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Yemen	3	North Africa & Middle East (including Israel, Turkey & Cyprus)
Angola	4	Sub-Saharan Africa
Benin	4	Sub-Saharan Africa
Botswana	4	Sub-Saharan Africa
Burkina Faso	4	Sub-Saharan Africa
Burundi	4	Sub-Saharan Africa
Cabo Verde	4	Sub-Saharan Africa
Cameroon	4	Sub-Saharan Africa
Central African Republic (the)	4	Sub-Saharan Africa
Chad	4	Sub-Saharan Africa
Comoros (the)	4	Sub-Saharan Africa
Congo (the Democratic Republic of the)	4	Sub-Saharan Africa
Congo (the)	4	Sub-Saharan Africa
Côte d'Ivoire	4	Sub-Saharan Africa
Djibouti	4	Sub-Saharan Africa
Equatorial Guinea	4	Sub-Saharan Africa
Eritrea	4	Sub-Saharan Africa
Eswatini	4	Sub-Saharan Africa
Ethiopia	4	Sub-Saharan Africa
Gabon	4	Sub-Saharan Africa
Gambia (the)	4	Sub-Saharan Africa
Ghana	4	Sub-Saharan Africa
Guinea	4	Sub-Saharan Africa
Guinea-Bissau	4	Sub-Saharan Africa
Kenya	4	Sub-Saharan Africa
Lesotho	4	Sub-Saharan Africa
Liberia	4	Sub-Saharan Africa
Madagascar	4	Sub-Saharan Africa
Malawi	4	Sub-Saharan Africa
Mali	4	Sub-Saharan Africa
Mauritania	4	Sub-Saharan Africa
Mauritius	4	Sub-Saharan Africa
Mozambique	4	Sub-Saharan Africa
Namibia	4	Sub-Saharan Africa
Niger (the)	4	Sub-Saharan Africa
Nigeria	4	Sub-Saharan Africa
Rwanda	4	Sub-Saharan Africa
Sao Tome and Principe	4	Sub-Saharan Africa
Senegal	4	Sub-Saharan Africa
Seychelles	4	Sub-Saharan Africa
Sierra Leone	4	Sub-Saharan Africa
Somalia	4	Sub-Saharan Africa
South Africa	4	Sub-Saharan Africa
South Sudan	4	Sub-Saharan Africa
Sudan (the)	4	Sub-Saharan Africa
Tanzania, the United Republic of	4	Sub-Saharan Africa
Togo	4	Sub-Saharan Africa
Uganda	4	Sub-Saharan Africa
Zambia	4	Sub-Saharan Africa

cname	ht_region	my_region
Zimbabwe	4	Sub-Saharan Africa
Andorra	5	Europe
Australia	5	Europe
Austria	5	Europe
Belgium	5	Europe
Canada	5	Americas
Denmark	5	Europe
Finland	5	Europe
France	5	Europe
Germany	5	Europe
Greece	5	Europe
Iceland	5	Europe
Ireland	5	Europe
Italy	5	Europe
Liechtenstein	5	Europe
Luxembourg	5	Europe
Malta	5	Europe
Monaco	5	Europe
Netherlands (the)	5	Europe
New Zealand	5	Europe
Norway	5	Europe
Portugal	5	Europe
San Marino	5	Europe
Spain	5	Europe
Sweden	5	Europe
Switzerland	5	Europe
United Kingdom of Great Britain and Northern Ireland (the)	5	Europe
United States of America (the)	5	Americas
China	6	Asia
Japan	6	Asia
Korea (the Democratic People's Republic of)	6	Asia
Korea (the Republic of)	6	Asia
Mongolia	6	Asia
Taiwan (Province of China)	6	Asia
Brunei Darussalam	7	Asia
Cambodia	7	Asia
Indonesia	7	Asia
Lao People's Democratic Republic (the)	7	Asia
Malaysia	7	Asia
Myanmar	7	Asia
Philippines (the)	7	Asia
Singapore	7	Asia
Thailand	7	Asia
Timor-Leste	7	Asia
Viet Nam	7	Asia
Afghanistan	8	Asia
Bangladesh	8	Asia
Bhutan	8	Asia
India	8	Asia
Maldives	8	Asia
Nepal	8	Asia

cname	ht_region	my_region
Pakistan	8	Asia
Sri Lanka	8	Asia
Tibet	8	Asia
Fiji	9	The Pacific (including Australia & New-Zealand
Kiribati	9	The Pacific (including Australia & New-Zealand
Marshall Islands	9	The Pacific (including Australia & New-Zealand
Micronesia (Federated States of)	9	The Pacific (including Australia & New-Zealand
Nauru	9	The Pacific (including Australia & New-Zealand
Palau	9	The Pacific (including Australia & New-Zealand
Papua New Guinea	9	The Pacific (including Australia & New-Zealand
Samoa	9	The Pacific (including Australia & New-Zealand
Solomon Islands	9	The Pacific (including Australia & New-Zealand
Tonga	9	The Pacific (including Australia & New-Zealand
Tuvalu	9	The Pacific (including Australia & New-Zealand
Vanuatu	9	The Pacific (including Australia & New-Zealand
Antigua and Barbuda	10	Americas
Bahamas (the)	10	Americas
Barbados	10	Americas
Belize	10	Americas
Dominica	10	Americas
Grenada	10	Americas
Guyana	10	Americas
Jamaica	10	Americas
Saint Kitts and Nevis	10	Americas
Saint Lucia	10	Americas
Saint Vincent and the Grenadines	10	Americas
Suriname	10	Americas
Trinidad and Tobago	10	Americas

Q11

```
data_qog <- data_qog %>%
  mutate(
    my_region = ifelse(ht_region %in% c(2, 10), "Americas",
                      ifelse(cname %in% c("Canada", "United States of America (the)"), "Americas",
                            ifelse(ht_region %in% c(6, 7, 8) | cname %in% c("Tajikistan", "Turkmenistan", "Uzbekistan"), "Asia",
                                  ifelse(ht_region %in% c(1, 5), "Europe",
                                        ifelse(ht_region == 3 | cname %in% c("Turkey", "Israel", "Cyprus"), "Middle East",
                                              ifelse(ht_region == 4, "Sub-Saharan Africa",
                                                    ifelse(ht_region == 9 | cname %in% c("Australia", "New Zealand"), "The Pacific", "Other"))))))))
  )
```

Q12

```
my_gnicon2010 <- data_qog %>%
  select(cname, wdi_pop, wdi_gnicon2010, wdi_gnicapcon2010) %>%
  mutate(my_gnicapcon = wdi_gnicon2010/wdi_pop)
head(my_gnicon2010)
```

```
## # A tibble: 6 x 5
##   cname      wdi_pop wdi_gnicon2010 wdi_gnicapcon2010 my_gnicapcon
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 Afghanistan 8996967      NA          NA          NA
## 2 Afghanistan 9169406      NA          NA          NA
## 3 Afghanistan 9351442      NA          NA          NA
## 4 Afghanistan 9543200      NA          NA          NA
## 5 Afghanistan 9744772      NA          NA          NA
## 6 Afghanistan 9956318      NA          NA          NA
```

Q13

```
data_gnipc <-
  my_gnicon2010 %>%
  mutate(check = ifelse(my_gnicapcon == wdi_gnicapcon2010, 0, 1))
```

Q14

```
data_gnipc %>%
  count(check) %>%
  mutate(percentage = n/sum(n))
```

```
## # A tibble: 3 x 3
##   check      n percentage
##   <dbl> <int>      <dbl>
## 1     0   219    0.0180
## 2     1  5117    0.421
## 3    NA  6820    0.561
```

Q15

It does not look like, when check is equal to one, that the values are different. The reason is perhaps because of how the values are rounded: observations with check = 0 are rounded at the fourth decimal and observations with check = 1 are rounded at the third decimal.

```
data_gnipc <- data_gnipc %>%
  mutate(check2 = ifelse(round(my_gnicapcon, digits = 3) == round(wdi_gnicapcon2010, digits = 3), 0, 1))
```

```
data_gnipc %>%
  count(check2)
```

```
## # A tibble: 3 x 2
##   check2      n
##   <dbl> <int>
## 1     0  5155
## 2     1   181
## 3    NA  6820
```

Describing variation

Q16

```
data_qog %>%
  summarize(mean = mean(wdi_internet, na.rm = TRUE),
            median = median(wdi_internet, na.rm = TRUE),
            max = max(wdi_internet, na.rm = TRUE),
            min = min(wdi_internet, na.rm = TRUE),
            IQR = IQR(wdi_internet, na.rm = TRUE))
```

```
## # A tibble: 1 x 5
##   mean median  max  min  IQR
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  21.8   5.79  100    0  37.9
```

```
data_qog %>%
  filter(wdi_internet == 100)
```

```
## # A tibble: 1 x 21
##   ccode cname      year ht_region wdi_area wdi_pop wdi_popden br_dem br_elect
##   <dbl> <chr>      <dbl> <fct>      <dbl>   <dbl>   <dbl> <dbl>   <dbl>
## 1   784 United Arab~ 2020 3         71020 9890400    139.     0       0
## # ... with 12 more variables: chga_hinst <dbl>, ht_regtype <dbl>,
## #   p_polity2 <dbl>, wdi_gnicon2010 <dbl>, wdi_gnicapcon2010 <dbl>,
## #   wdi_gdpcapcon2010 <dbl>, wdi_lifexp <dbl>, wdi_litradd <dbl>,
## #   undp_hdi <dbl>, wdi_expmil <dbl>, wdi_internet <dbl>, my_region <chr>
```

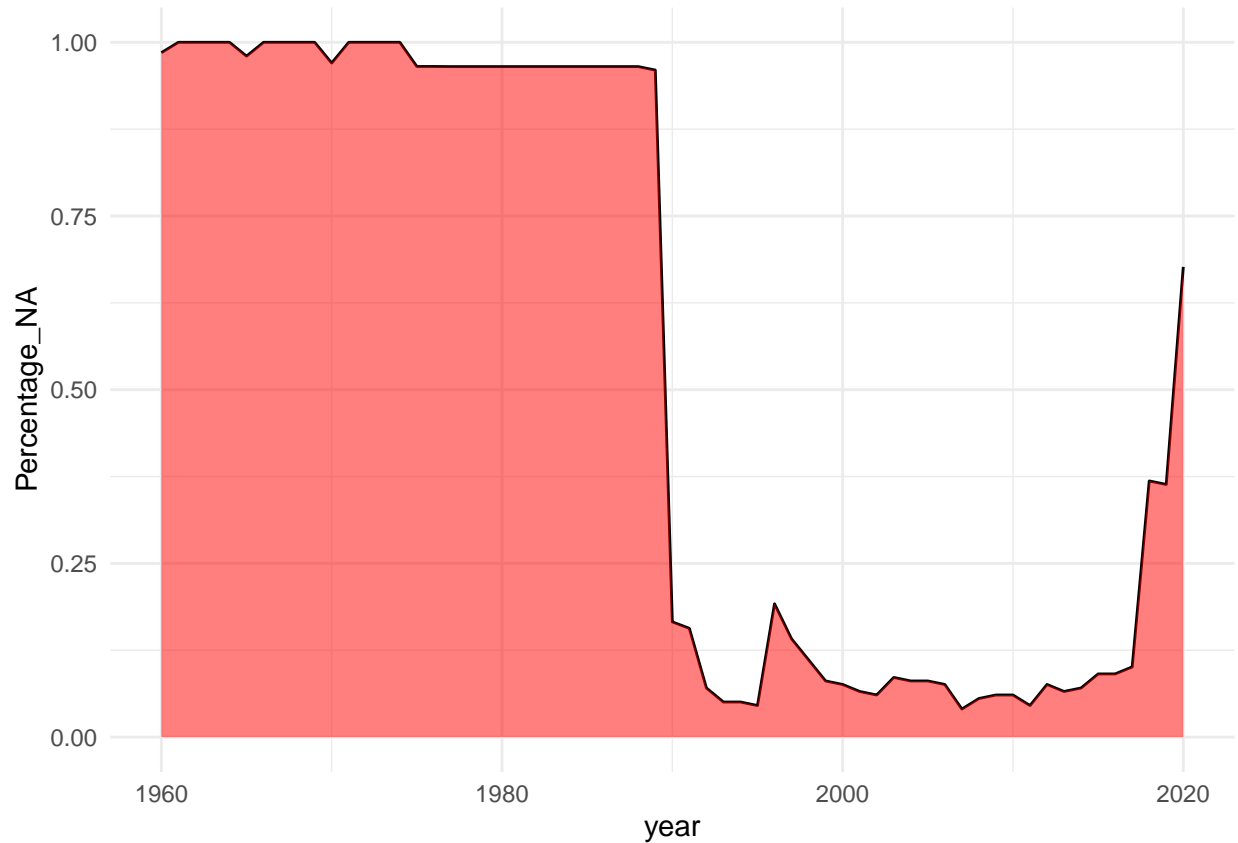
```
data_qog %>%
  filter(wdi_internet == 0 & cname == "Afghanistan" & year == 1990)
```

```
## # A tibble: 1 x 21
##   ccode cname      year ht_region wdi_area wdi_pop wdi_popden br_dem br_elect
##   <dbl> <chr>      <dbl> <fct>      <dbl>   <dbl>   <dbl> <dbl>   <dbl>
## 1     4 Afghanistan 1990 8         652860 12412311    19.0     0       2
## # ... with 12 more variables: chga_hinst <dbl>, ht_regtype <dbl>,
## #   p_polity2 <dbl>, wdi_gnicon2010 <dbl>, wdi_gnicapcon2010 <dbl>,
## #   wdi_gdpcapcon2010 <dbl>, wdi_lifexp <dbl>, wdi_litradd <dbl>,
## #   undp_hdi <dbl>, wdi_expmil <dbl>, wdi_internet <dbl>, my_region <chr>
```

Q17

```
data_qog %>%
  group_by(year) %>%
  count(Na = (is.na(wdi_internet))) %>%
  mutate(Percentage_NA = n/sum(n)) %>%
  ungroup() %>%
  filter(Na == TRUE) %>%
```

```
ggplot(aes(x = year, y = Percentage_NA))+
  geom_line()+
  geom_area(fill = "red", alpha = 0.5)+
  theme_minimal()
```



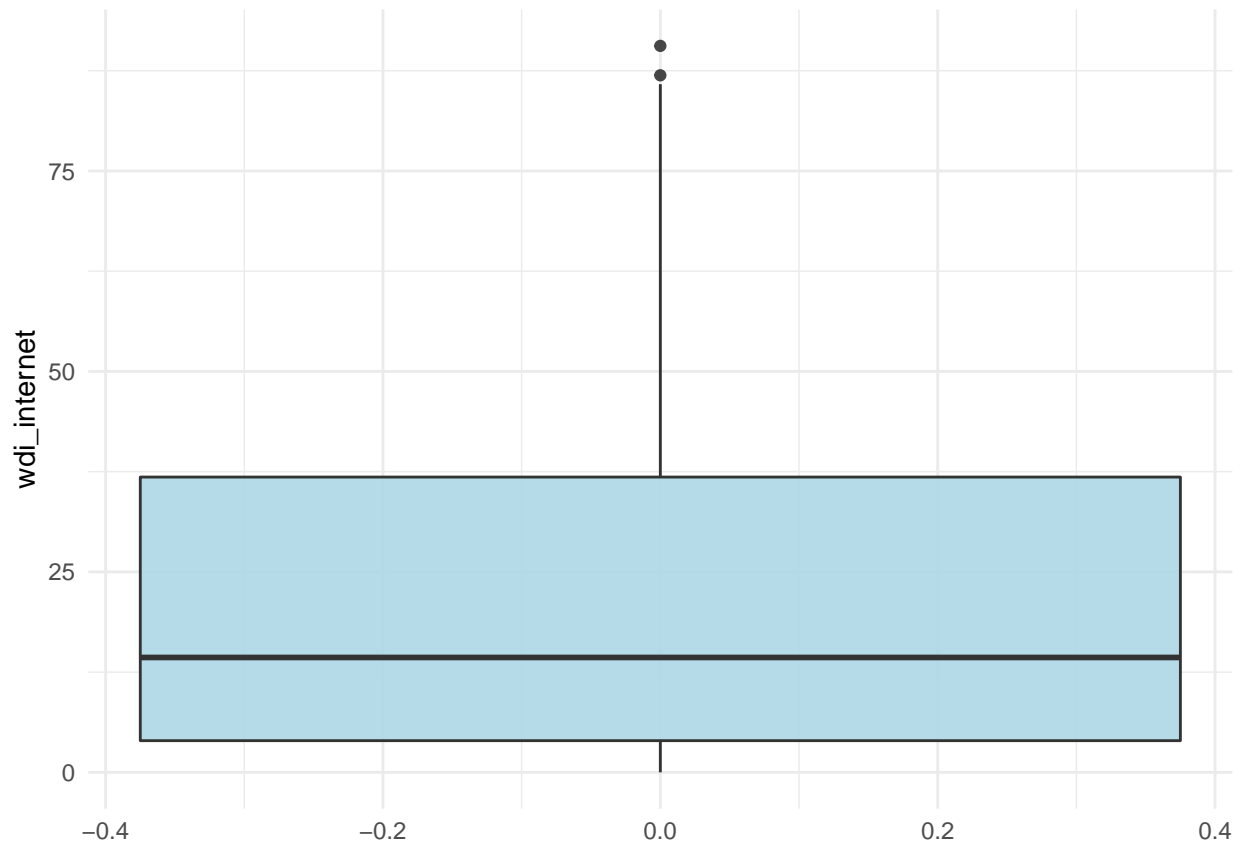
```
data_qog %>%
  group_by(year) %>%
  count(Na = (is.na(wdi_internet))) %>%
  mutate(Percentage_NA = n/sum(n)) %>%
  ungroup() -> na_analysis_internet

na_analysis_internet %>%
  filter(Na == TRUE) %>%
  summarize(best_coverage = min(Percentage_NA),
            best_coverage_year = year[which(Percentage_NA == min(Percentage_NA))])
```

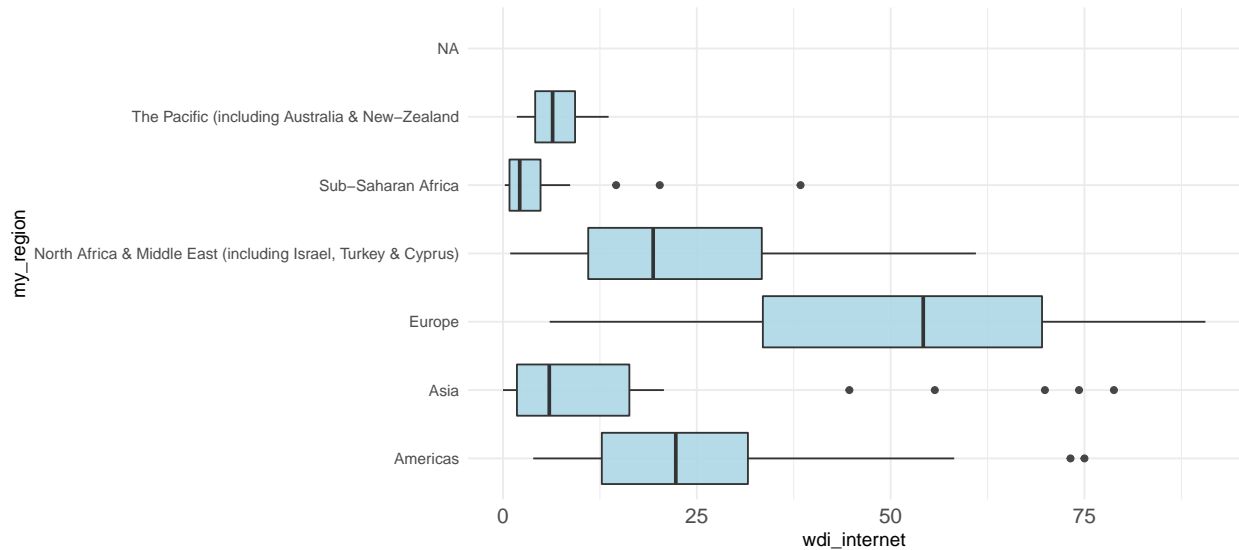
```
## # A tibble: 1 x 2
##   best_coverage best_coverage_year
##         <dbl>         <dbl>
## 1         0.0404         2007
```

Q18

```
data_qog %>%  
  filter(year == 2007) %>%  
  ggplot()+  
  aes(y = wdi_internet)+  
  geom_boxplot(fill = "lightblue", alpha = 0.9)+  
  theme_minimal()
```



```
data_qog %>%  
  filter(year == 2007) %>%  
  ggplot()+  
  aes(y = my_region, x = wdi_internet)+  
  geom_boxplot(fill = "lightblue", alpha = 0.9)+  
  theme_minimal()+  
  theme(axis.text.x = element_text(size = 12))
```



Q19

Analysis of variable life expectancy across countries and time (wdi_lifexp) We want: mean, median, range, variance, sd, IQR and percentage of NA for all dataset

```
data_qog %>%
  summarize(
    mean_life_exp = mean(wdi_lifexp, na.rm = TRUE),
    median_life_exp = median(wdi_lifexp, na.rm = TRUE),
    variance_life_exp = sd(wdi_lifexp, na.rm = TRUE)^2,
    standard_deviation_life_exp = sd(wdi_lifexp, na.rm = TRUE),
    IQR_life_exp = IQR(wdi_lifexp, na.rm = TRUE),
    min_life_exp = min(wdi_lifexp, na.rm = TRUE),
    max_life_exp = max(wdi_lifexp, na.rm = TRUE)
  )
```

```
## # A tibble: 1 x 7
##   mean_life_exp median_life_exp variance_life_exp standard_deviation_life_exp IQR_life_exp
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1      64.3         67.5         131.         11.4         17.3
## # ... with 2 more variables: min_life_exp <dbl>, max_life_exp <dbl>
```

Percentage of NAs

```
data_qog %>%
  select(wdi_lifexp) %>%
  count(is.na(wdi_lifexp)) %>%
  mutate(percentage_NAs = n/sum(n))
```

```
## # A tibble: 2 x 3
##   'is.na(wdi_lifexp)'      n percentage_NAs
##   <lgl>         <int>         <dbl>
## 1 FALSE         9460         0.778
## 2 TRUE          2696         0.222
```

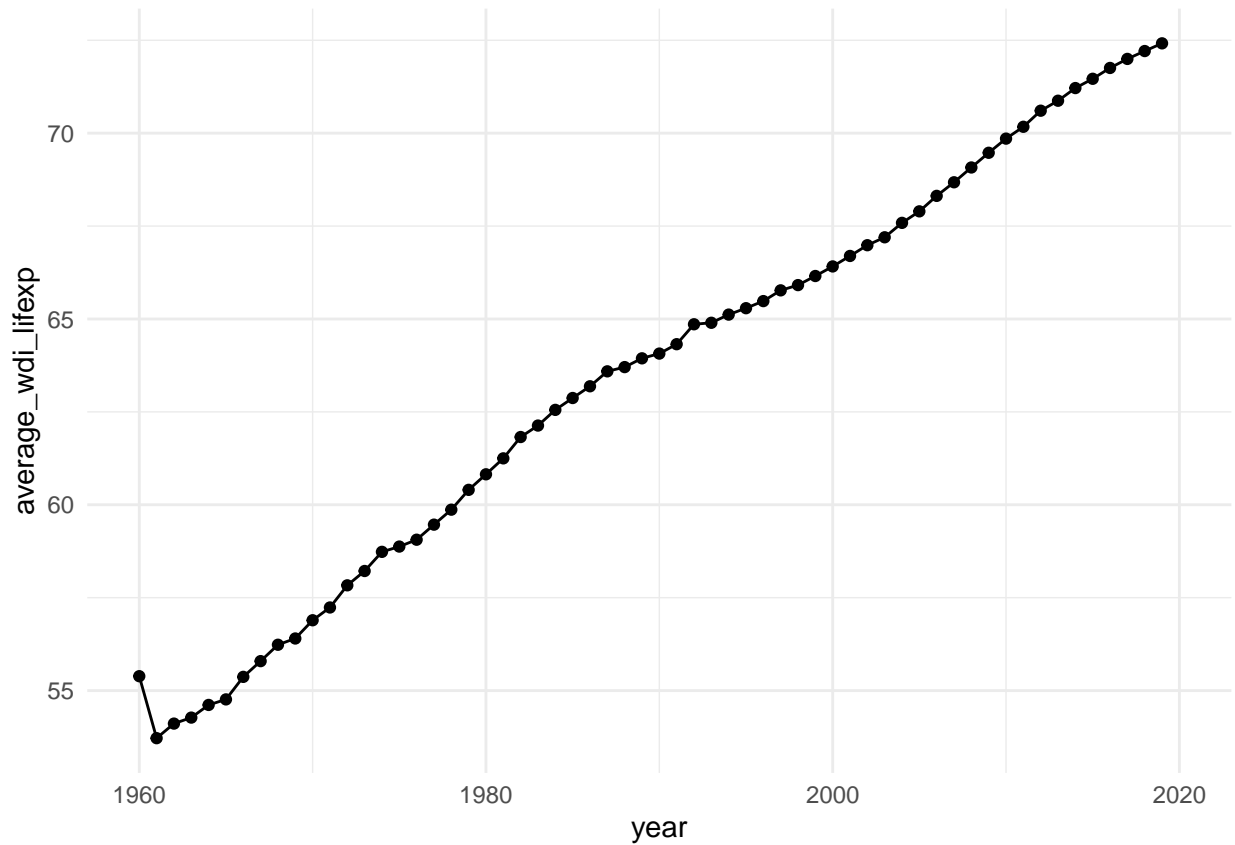
Q20

```
data_qog %>%
  group_by(year) %>%
  summarize(average_wdi_lifexp = mean(wdi_lifexp, na.rm = TRUE)) %>%
  ungroup() -> Q20_tibble
Q20_tibble
```

```
## # A tibble: 61 x 2
##   year average_wdi_lifexp
##   <dbl>         <dbl>
## 1 1960          55.4
## 2 1961          53.7
## 3 1962          54.1
## 4 1963          54.3
## 5 1964          54.6
## 6 1965          54.8
## 7 1966          55.4
## 8 1967          55.8
## 9 1968          56.2
## 10 1969         56.4
## # ... with 51 more rows
```

Plotting average wdi_lifexp with year

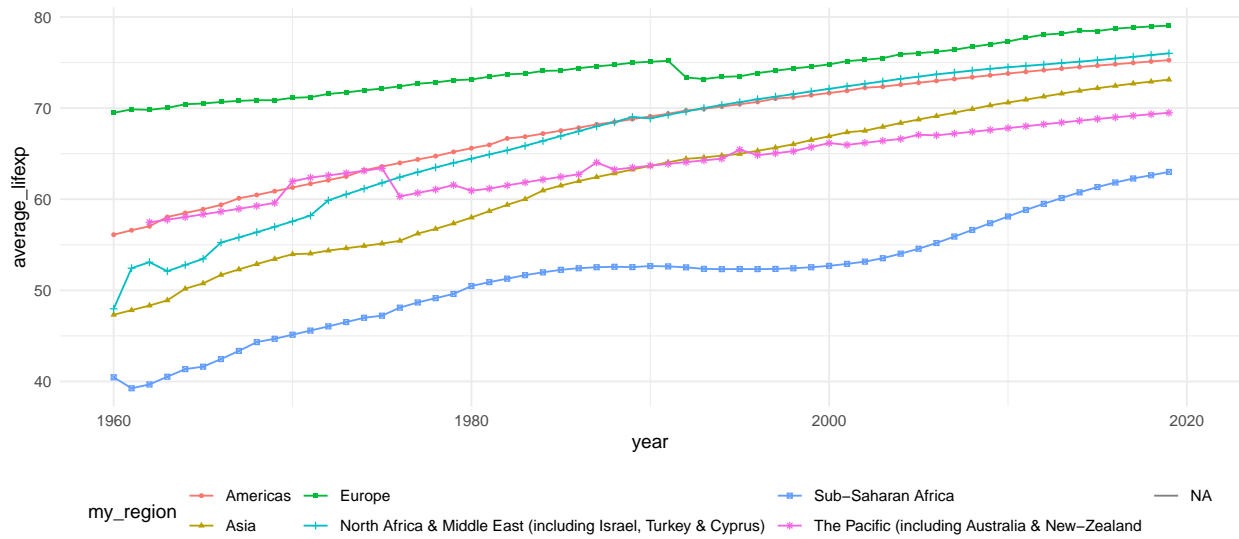
```
data_qog %>%
  group_by(year) %>%
  summarize(average_wdi_lifexp = mean(wdi_lifexp, na.rm = TRUE)) %>%
  ungroup() %>%
  ggplot()+
  aes(x = year, y = average_wdi_lifexp)+
  geom_point()+
  geom_line()+
  theme_minimal()
```



Q21

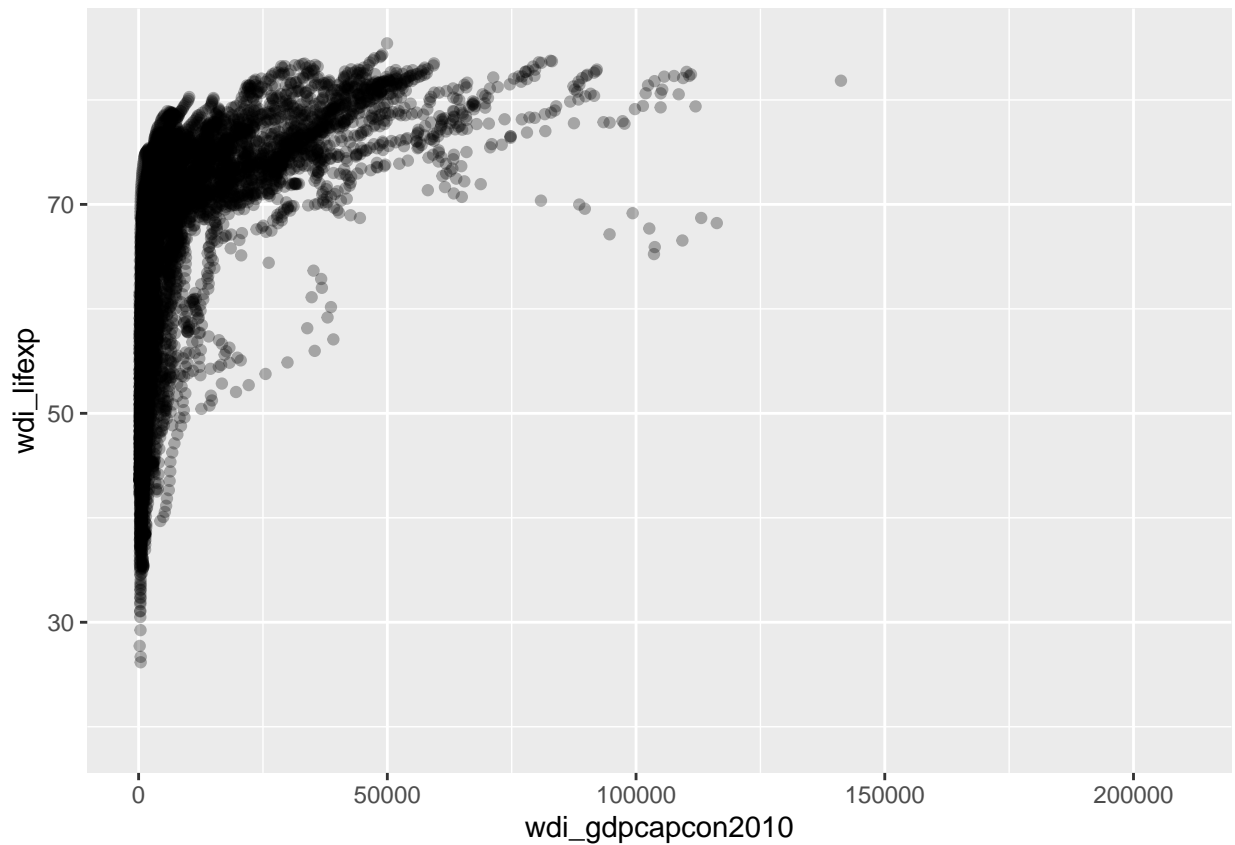
```
data_qog %>%
  group_by(my_region, year) %>%
  summarize(average_lifexp = mean(wdi_lifexp, na.rm = TRUE)) %>%
  ungroup() -> Q21_tibble

data_qog %>%
  group_by(my_region, year) %>%
  summarize(average_lifexp = mean(wdi_lifexp, na.rm = TRUE)) %>%
  ungroup() %>%
  ggplot()+
  aes(x = year, y = average_lifexp, color = my_region, shape = my_region)+
  geom_line()+
  geom_point(size = 1)+
  theme_minimal()+
  theme(legend.position = "bottom")
```

Q22

```
data_qog %>%
  ggplot()+
  aes(x = wdi_gdpcapcon2010, y = wdi_lifexp)+
  geom_point(alpha = 0.3)
```



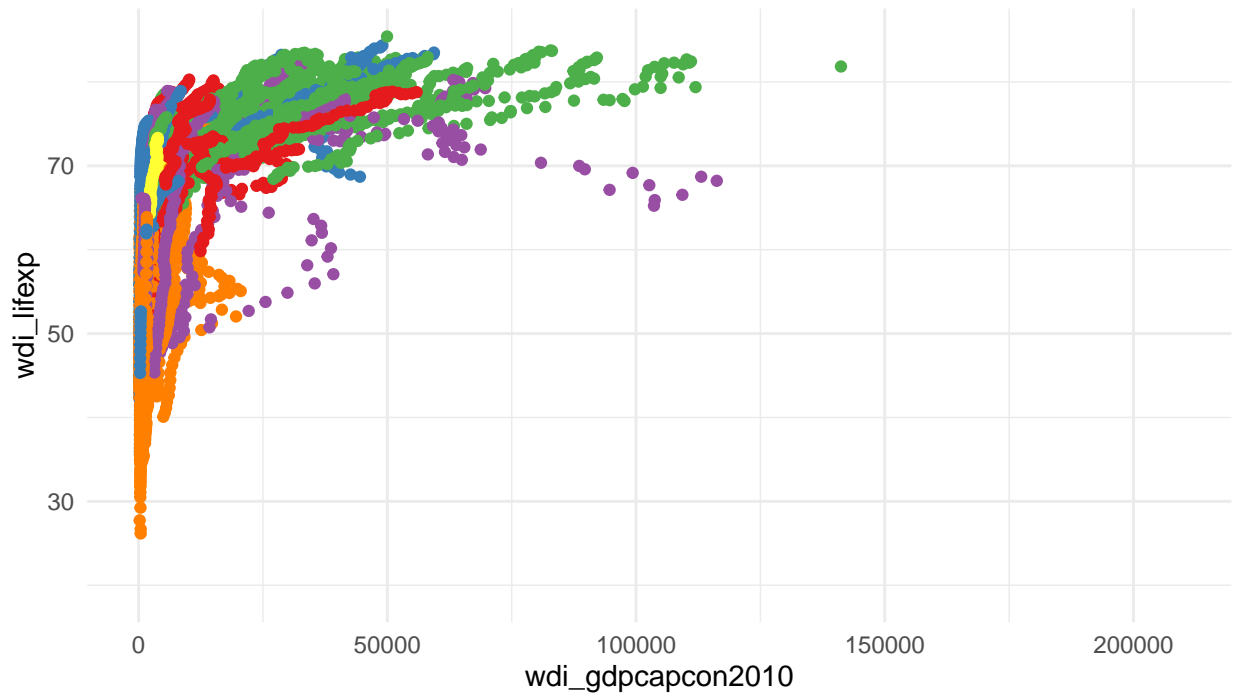
```
cor(data_qog$wdi_gdpcapcon2010, data_qog$wdi_lifexp, use = "complete.obs")
```

```
## [1] 0.5636858
```

There seems to be a positive correlation between GDP per cap and life expectancy.

Q23

```
data_qog %>%  
  ggplot()+  
  aes(x = wdi_gdpcapcon2010, y = wdi_lifexp, color = my_region)+  
  geom_point(alpha = 1)+  
  theme_minimal()+  
  theme(legend.position = "bottom")+  
  scale_color_brewer(palette = "Set1")
```



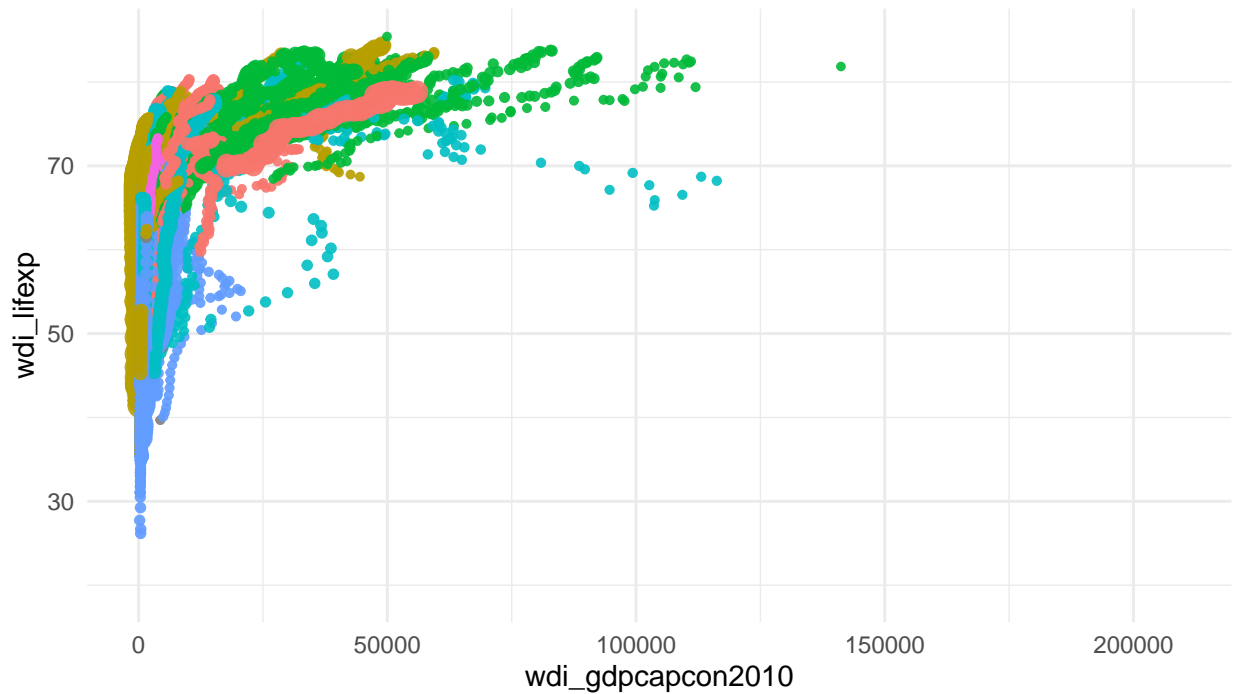
Americas ● Europe ● Sub-Saharan Africa
 Asia ● North Africa & Middle East (including Israel, Turkey & Cyprus) ● The Pacific (including Australia)

- A region with low GDP per cap and low life expectancy is Sub-Saharan Africa
- Regions with high GDP per cap and high life expectancy are Europe and America
- A region for which life expectancy tend to be relatively lower than their GDP per cap is North Africa

North Africa and Middle East is an interesting case whose observations stand out in the graph above. Around 50000 gdp per cap, life expectancy seems to even decrease with gdp. Thus, I am going to include several variable to see any pattern.

I am going first to include the human development index in the graph:

```
data_qog %>%
  ggplot()+
  aes(x = wdi_gdpcapcon2010, y = wdi_lifexp, color = my_region, size = wdi_pop)+
  geom_point(alpha = 0.9)+
  theme_minimal()+
  theme(legend.position = "bottom")
```



Sub-Saharan Africa
 North Africa & Middle East (including Israel, Turkey & Cyprus)
 The Pacific (including Australia & New Zealand)

Q24

```
typeof(data_qog$br_dem)
```

```
## [1] "double"
```

```
typeof(data_qog$p_polity2)
```

```
## [1] "double"
```

br_dem (is the country a democracy) is stored as a double, it is however a dummy variable (nominal variable) taking value 0 if country is not democratic and 1 if country is a democracy. Thus, it does not make sense to compute its mean, I should make a table with the frequencies or relative frequencies or do a barplot.

p_polity2 is also stored as a double. It also does not make sense because p_polity2 (revised combined polity score) is an ordinal variable ranging from -10 (strongly autocratic) to +10 (strongly democratic). To give information about this variable, I would do a table and a barplot (sorted according to the range).

Q25

I first transform those two variables into factors

```
data_qog$br_dem <- factor(data_qog$br_dem, levels = c(0,1), labels = c("Not democratic", "democratic"))
data_qog$p_polity2 <- factor(data_qog$p_polity2, levels = (-10:10))
```

I then do some univariate and bivariate analysis of the variables through tables

```
data_qog %>%
  count(p_polity2) %>%
  kable(caption = "Is the country democratic", align = "c")
```

Table 4: Is the country democratic

p_polity2	n
-10	287
-9	606
-8	328
-7	1311
-6	419
-5	190
-4	231
-3	240
-2	148
-1	177
0	224
1	114
2	95
3	123
4	252
5	334
6	469
7	386
8	593
9	516
10	1577
NA	3536

```
data_qog %>%
  count(br_dem) %>%
  kable(caption = "Is the country democratic", align = "c")
```

Table 5: Is the country democratic

br_dem	n
Not democratic	5140
democratic	4915
NA	2101

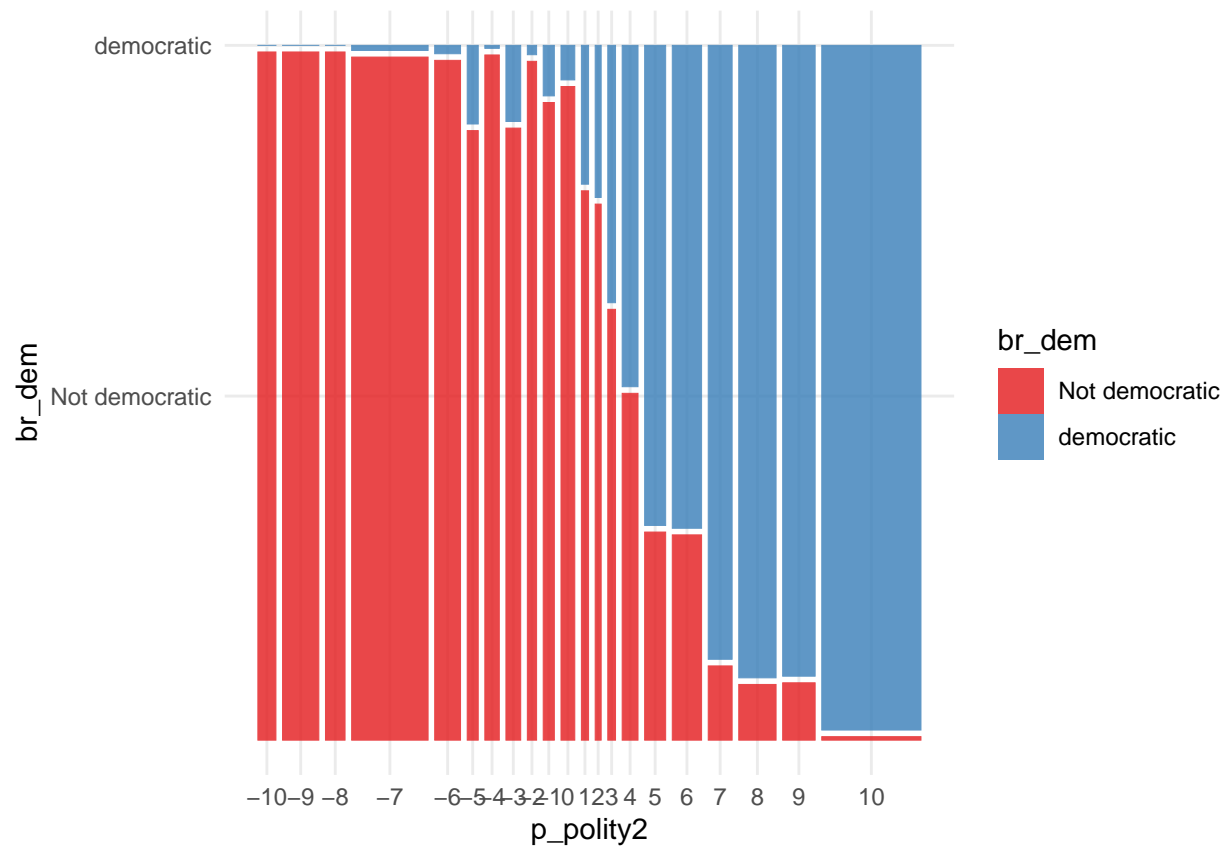
```
kable(table(data_qog$p_polity2, data_qog$br_dem), caption = "Revised combined polity score according to
```

Table 6: Revised combined polity score according to democracy or not

	Not democratic	democratic
-10	287	0
-9	577	0
-8	311	0
-7	1206	9
-6	406	5
-5	155	20
-4	229	1
-3	209	26
-2	145	2
-1	164	13
0	208	11
1	91	23
2	74	21
3	77	46
4	127	125
5	101	233
6	139	327
7	41	339
8	48	545
9	43	473
10	8	1569

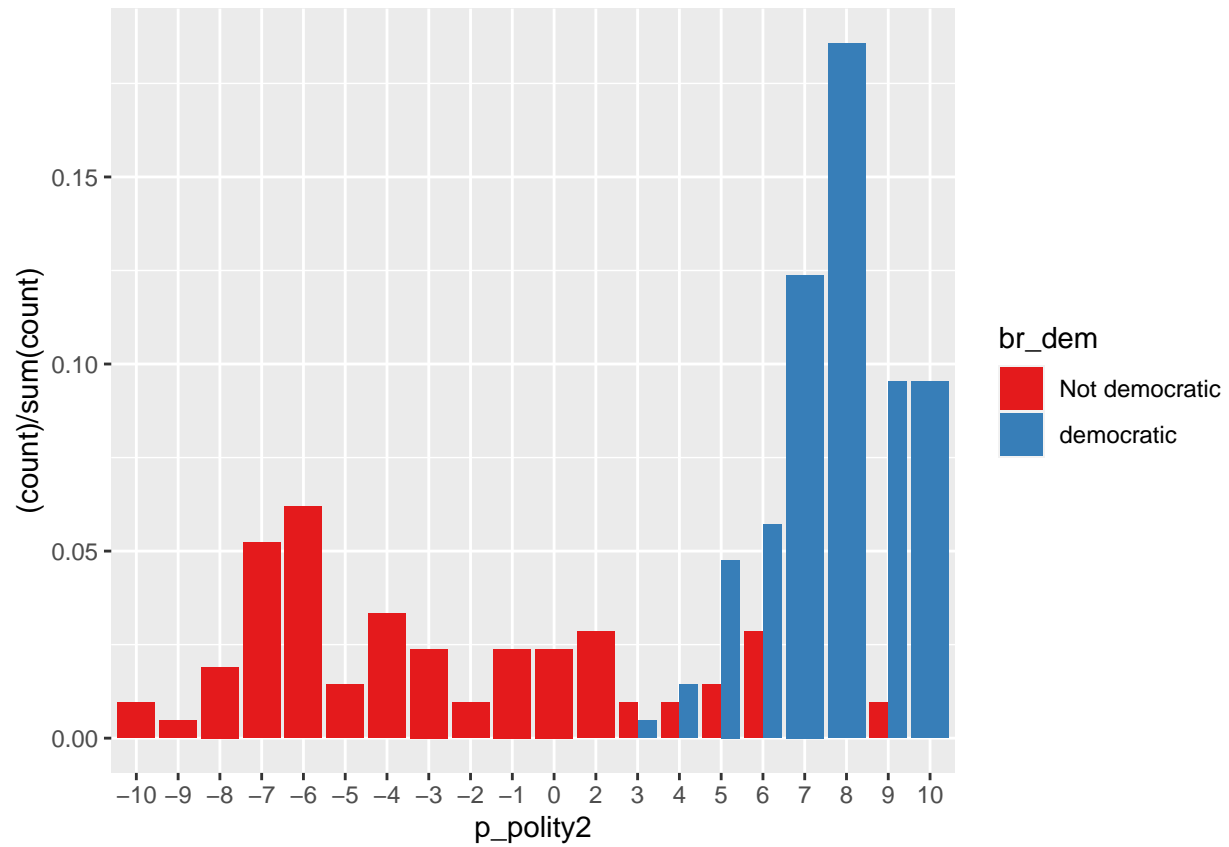
I choose then to do a mosaic plot of p_polity according to br_dem to represent the association of the two variable visually

```
data_qog %>% ggplot()+
  geom_mosaic(aes(x = product(br_dem, p_polity2), fill = br_dem), na.rm = TRUE)+
  theme_minimal()+
  scale_fill_brewer(palette = "Set1")
```



Another possibility is to do a stacked or dodge barplot of p_polity and br_dem

```
data_qog %>%
  drop_na() %>%
  ggplot()+
  aes(x = p_polity2, y = (..count..)/sum(..count..), fill = br_dem)+
  geom_bar(position = "dodge")+
  scale_fill_brewer(palette = "Set1")
```



```
data_qog %>%
  drop_na() %>%
  ggplot()+
  aes(x = p_polity2, y = (..count..)/sum(..count..), fill = br_dem)+
  geom_bar()+
  scale_fill_brewer(palette = "Set1")
```