Explorative Analysis

Discovering patterns by visualising distriubtions and relationships

Ayyaz Azeem

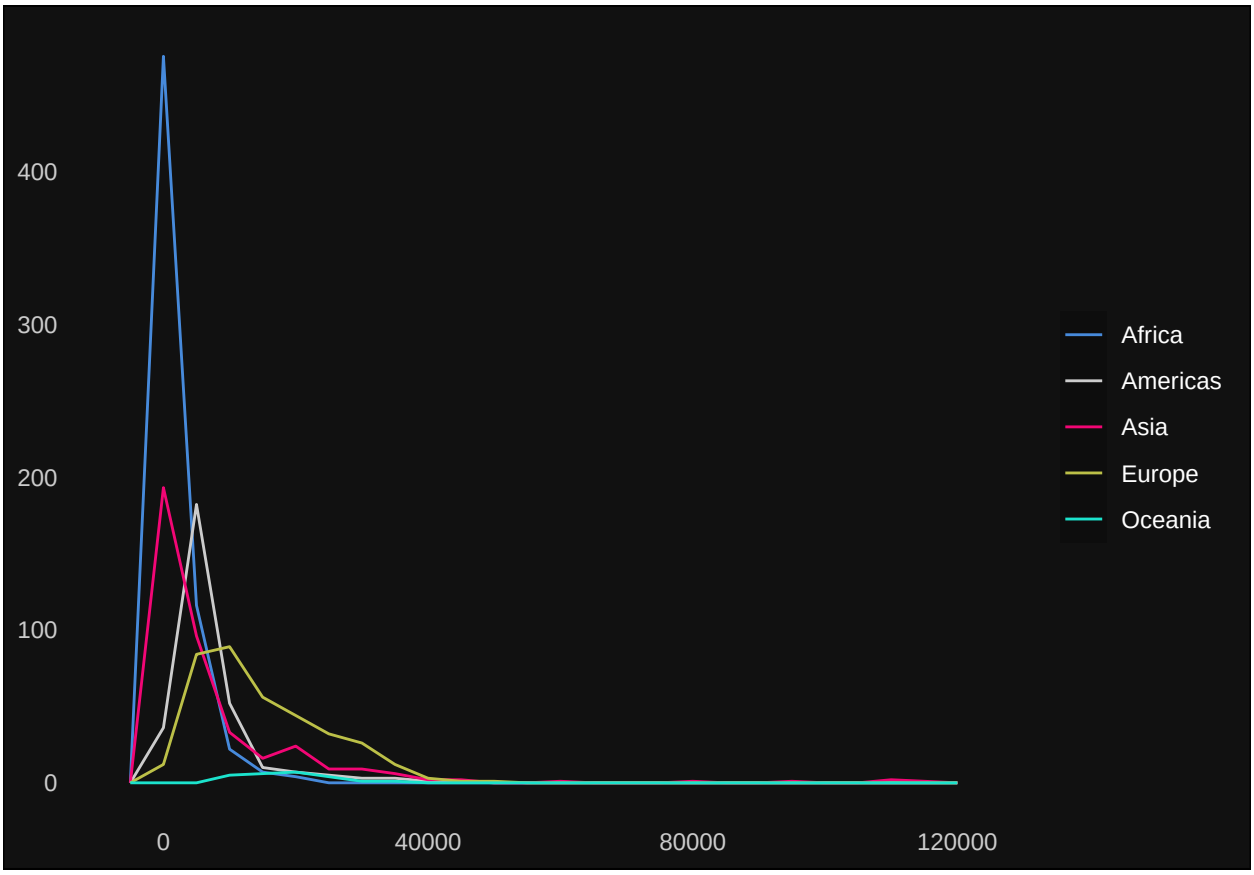
11/15/2020

Task 01

Exercise 01 Now its time to create your own theme. You are free to chose one theme as for example the dark theme and make changes on top. Think about your colors, the style of the axis labels, lines and ticks and the legend.

Exercise 02

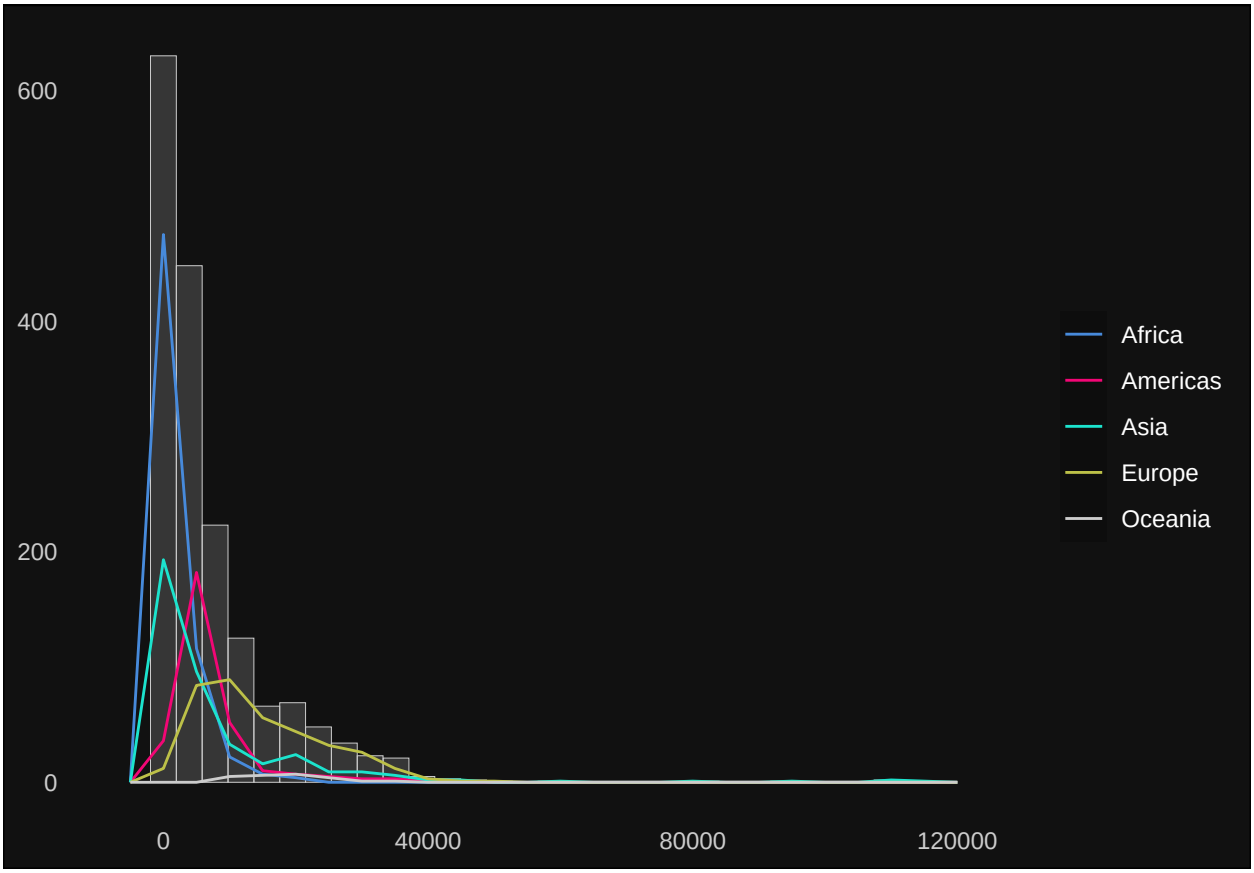
Choose a different variable like for example population or GDP per capita and check in the same way its frequency using a frequency plot.



Exercise 03

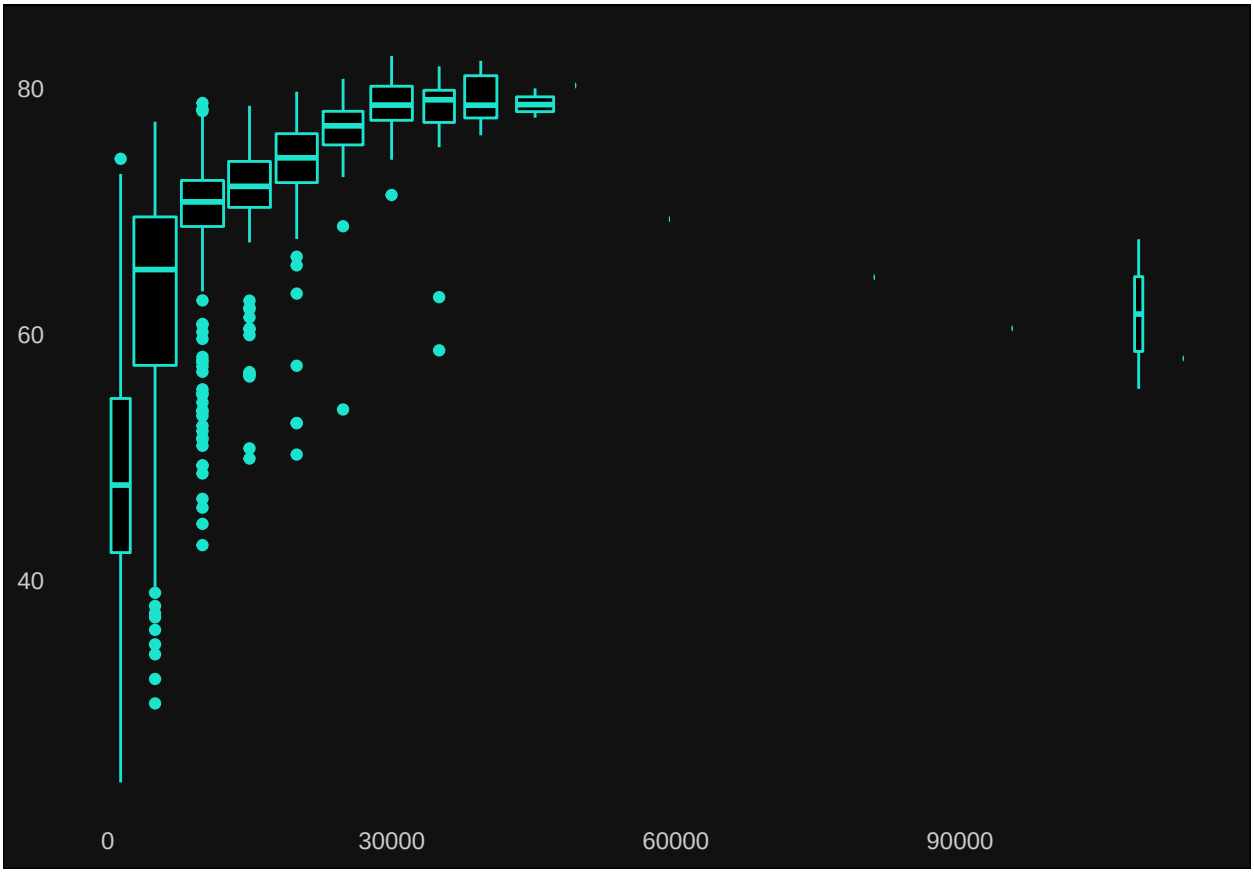
Choose a different variable like for example population or GDP per capita and check in the same way its frequency now using a histogram.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Exercise 04

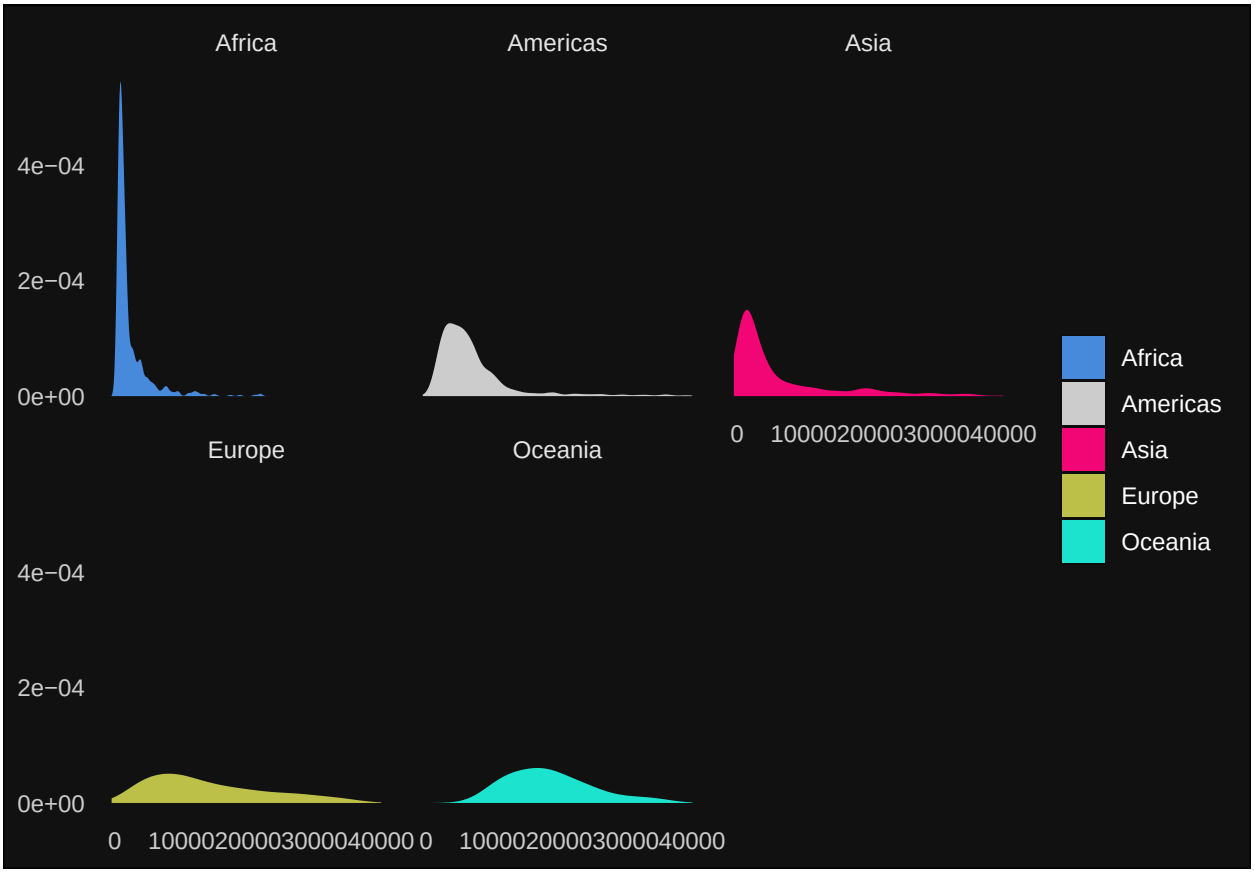
Choose a different variable like for example population or GDP per capita and check in the same way its frequency now using a boxplot.



Exercise 05

Choose a different variable like for example population or GDP per capita and check in the same way its frequency now using a density plot.

```
## Warning: Removed 14 rows containing non-finite values (stat_density).
```

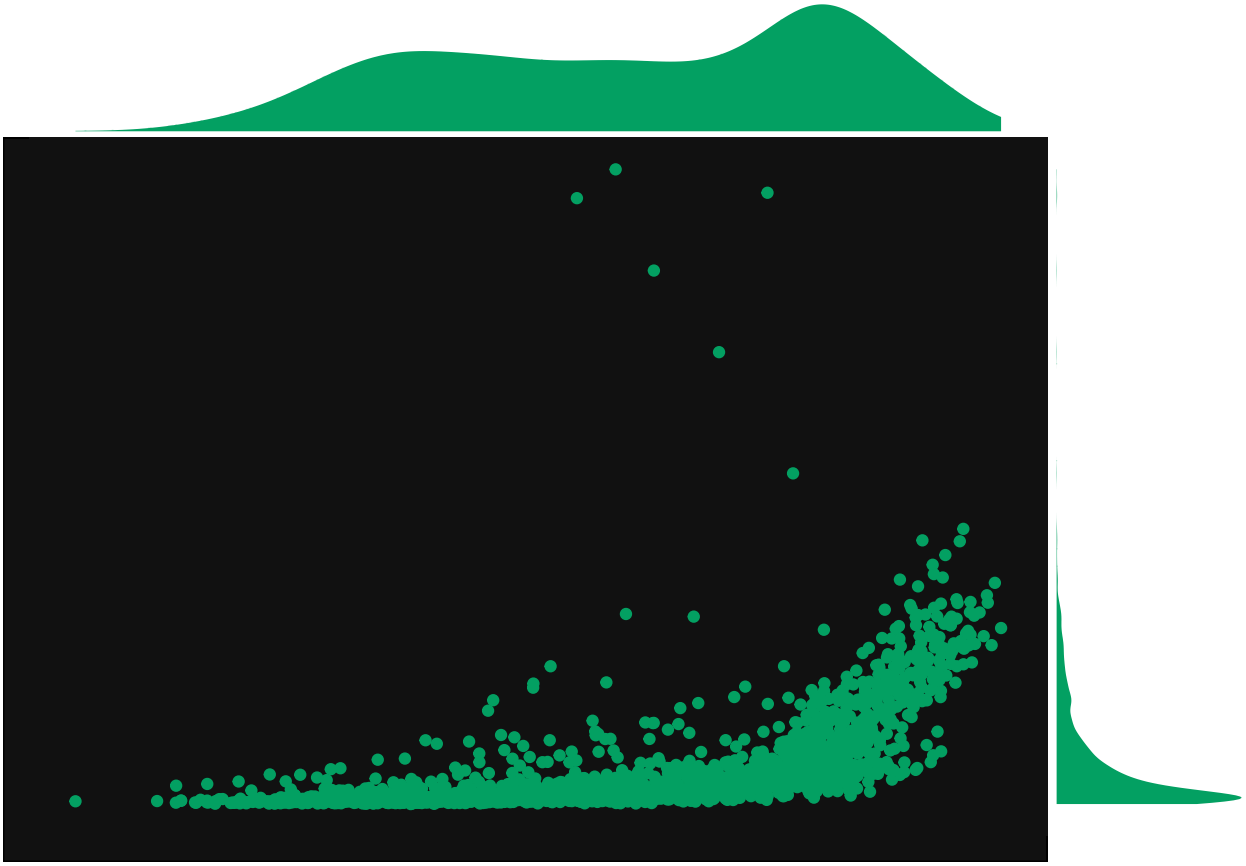


Exercise 06

Chose two variables from he gapminder dataset and check their relationship by using a scatter plot.

Exercise 07

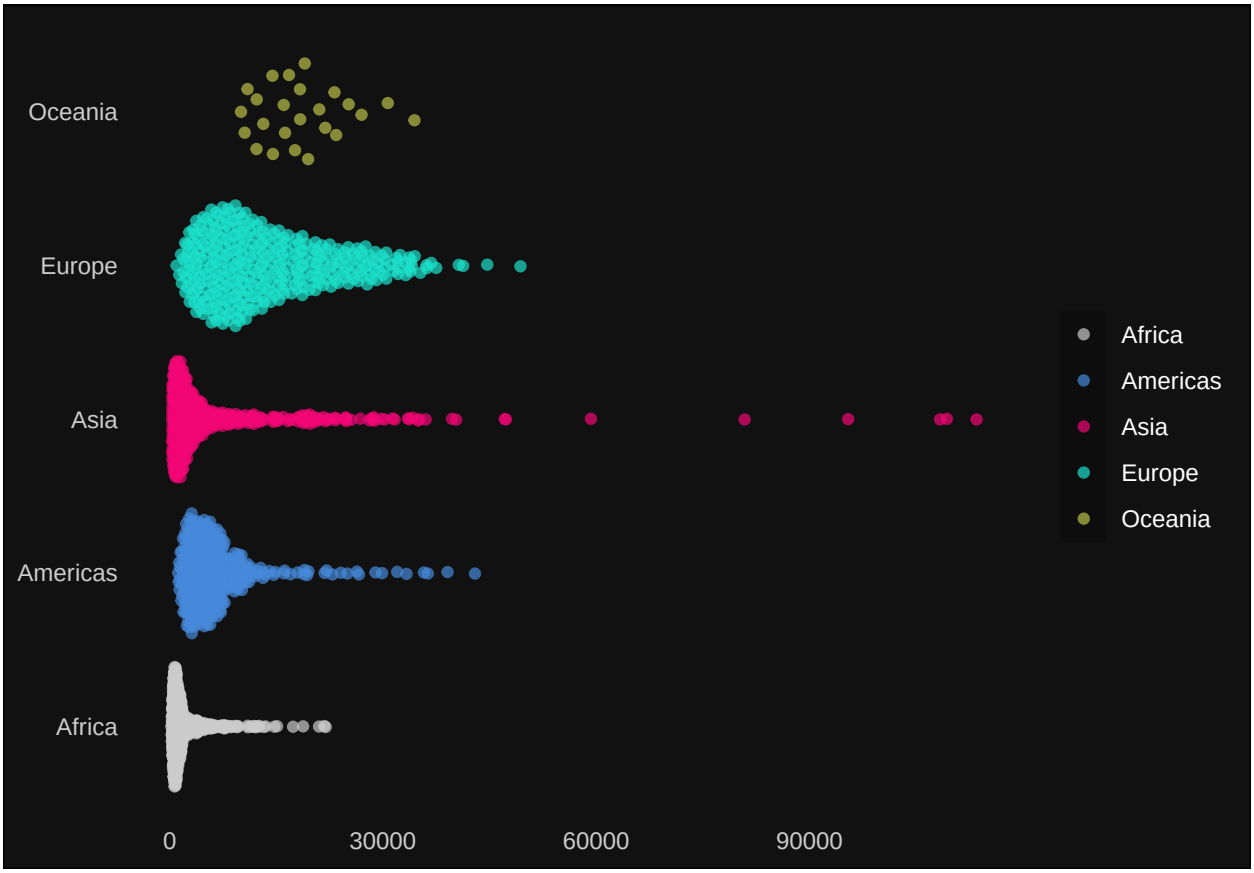
Chose two variables from the gapminder dataset and create in the same way the marginal scatter plots.



Exercise 08

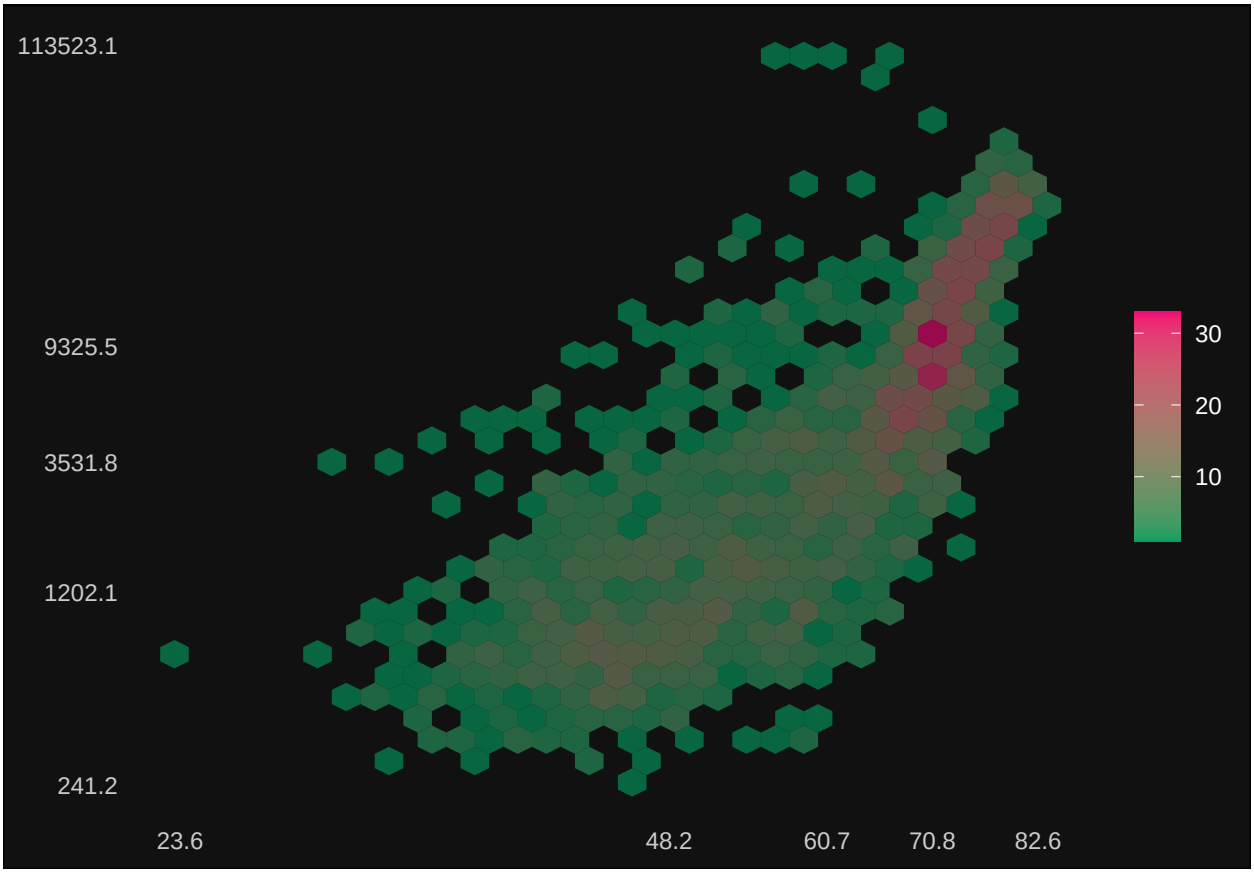
Chose the variables continent and any numeric variable from the gapminder dataset and create the beeswamp plots.

```
## Warning: package 'ggbeeswarm' was built under R version 4.0.3
```



Exercise 09

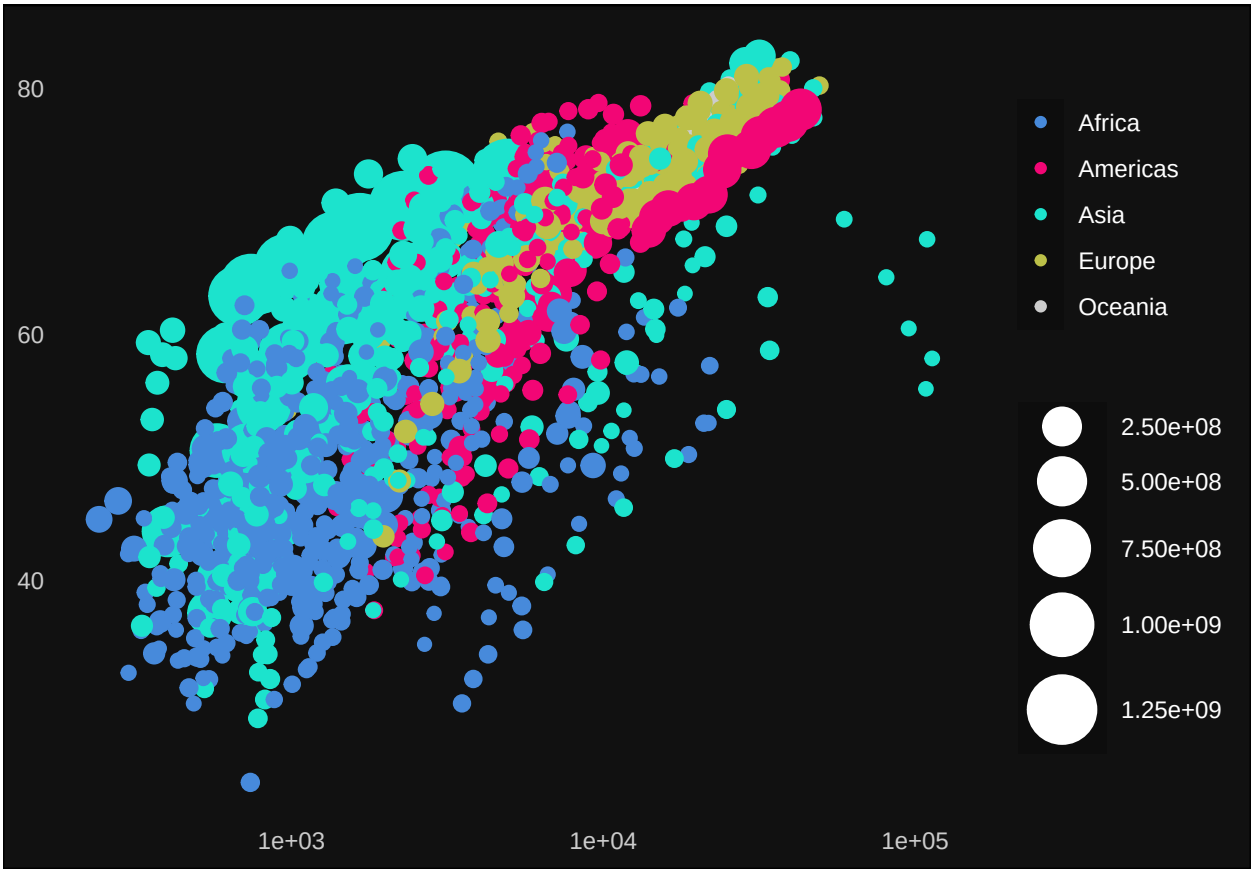
Chose two variables from the gapminder dataset and check their relationship by using the hexagonal binning.



Exercise 10

Use the examples and play around a bit with the options in order to get familiar with scales and axes.

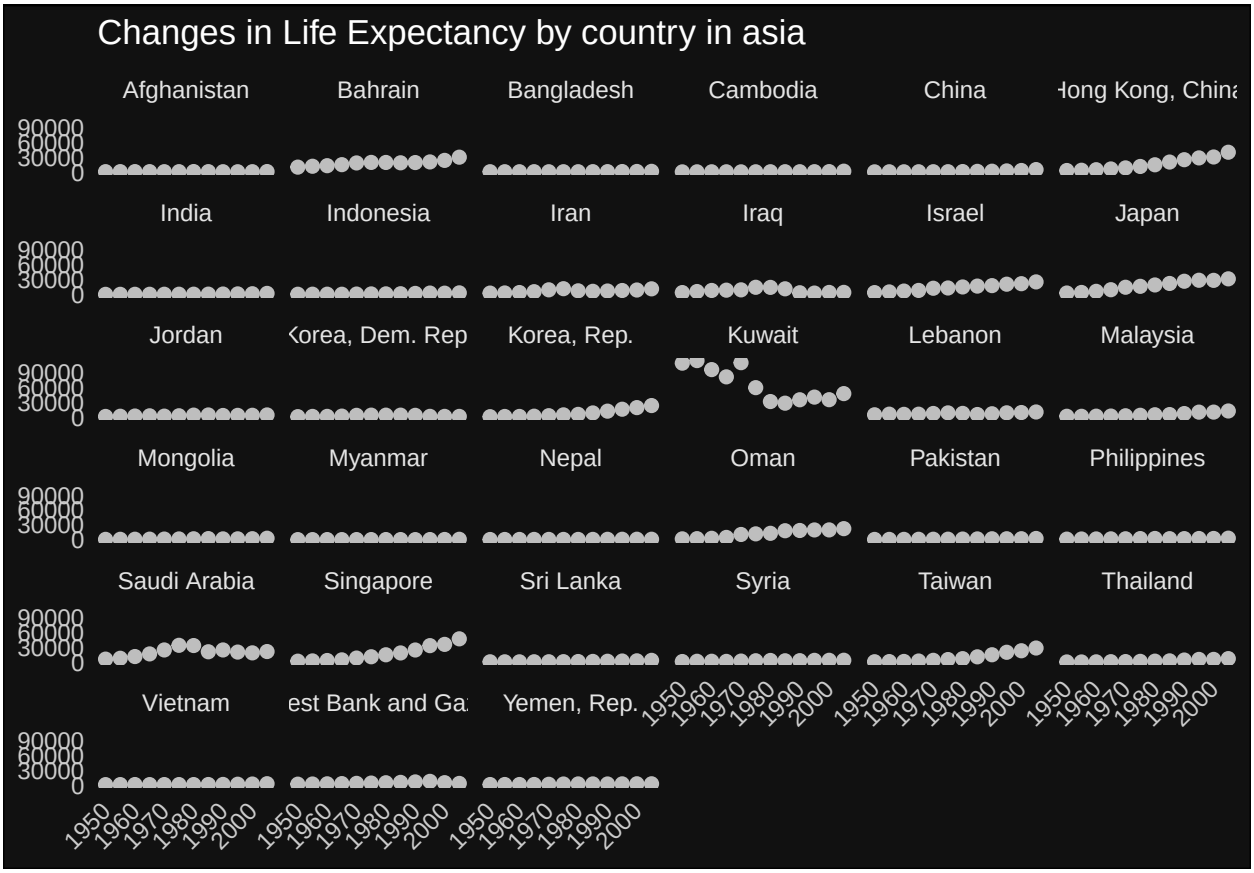
```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```



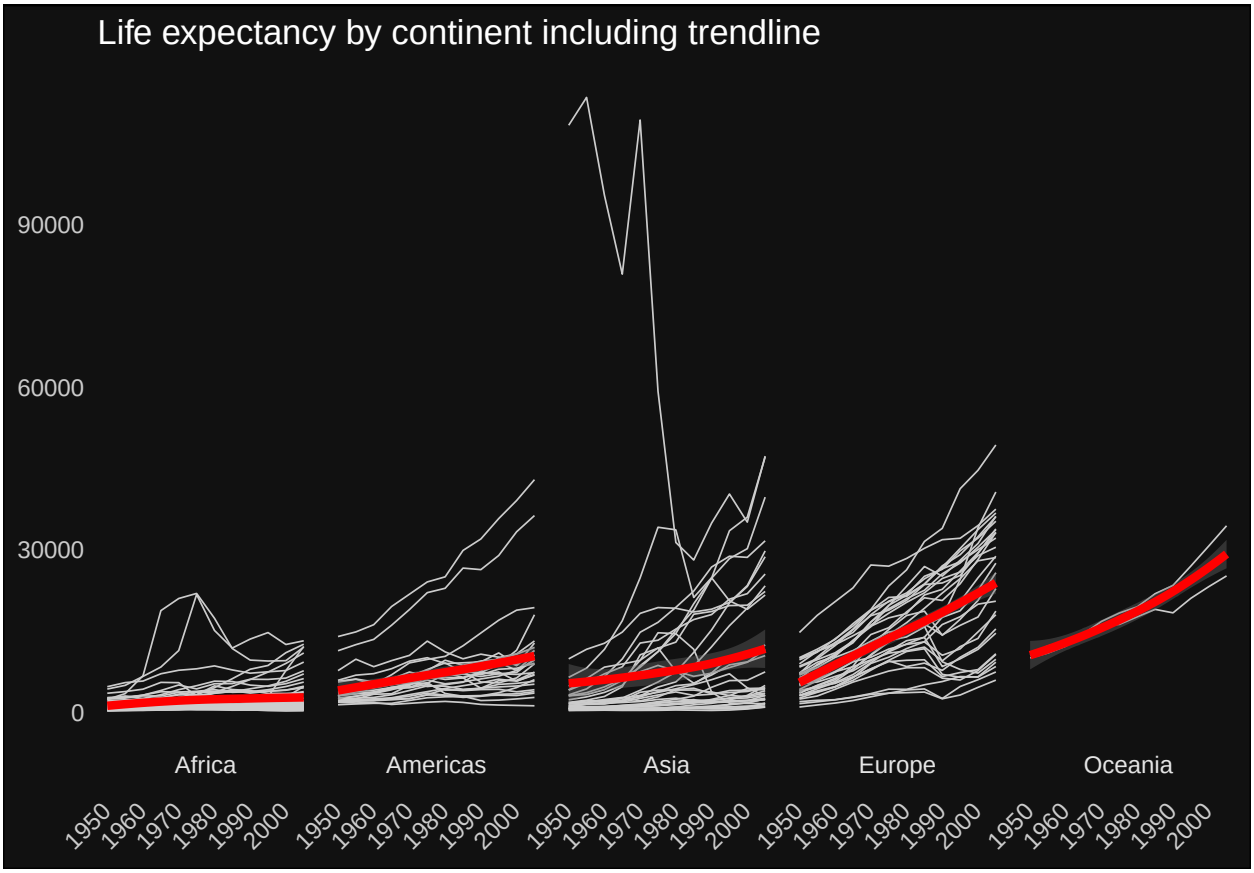
Exercise 11

Now chose a different variable from the gapminder dataset and do the same analysis flow, feel free to check on another continent as well.

Changes in Life Expectancy by country in asia
Afghanistan
Bahrain
Bangladesh
Cambodia
China
Hong Kong, China
India
Indonesia
Iran
Iraq
Israel
Japan
Jordan
Korea, Dem. Rep
Korea, Rep.
Kuwait
Lebanon
Malaysia
Mongolia
Myanmar
Nepal
Oman
Pakistan
Philippines
Saudi Arabia
Singapore
Sri Lanka
Syria
Taiwan
Thailand
Vietnam
est Bank and Ga
Yemen, Rep.
90000
60000
30000
0
1950
1960
1970
1980
1990
2000




```
## `geom_smooth()` using formula 'y ~ x'
```




```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1951.7
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 5.275
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0
```

```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 27.826
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 1951.7
```

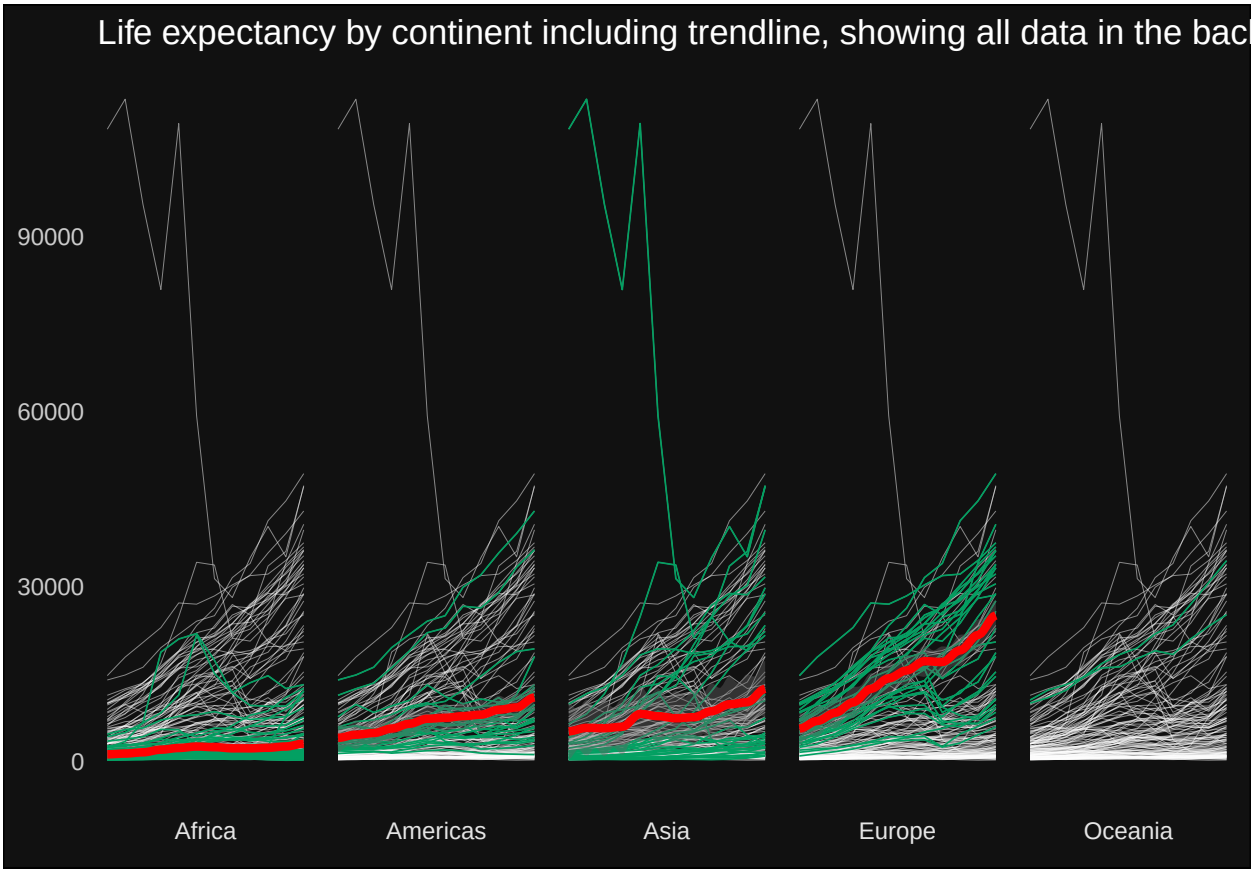
```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 5.275
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 27.826

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1951.7

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 5.275

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 27.826

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 1951.7

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 5.275

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 27.826

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1951.7

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 5.275

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0
```

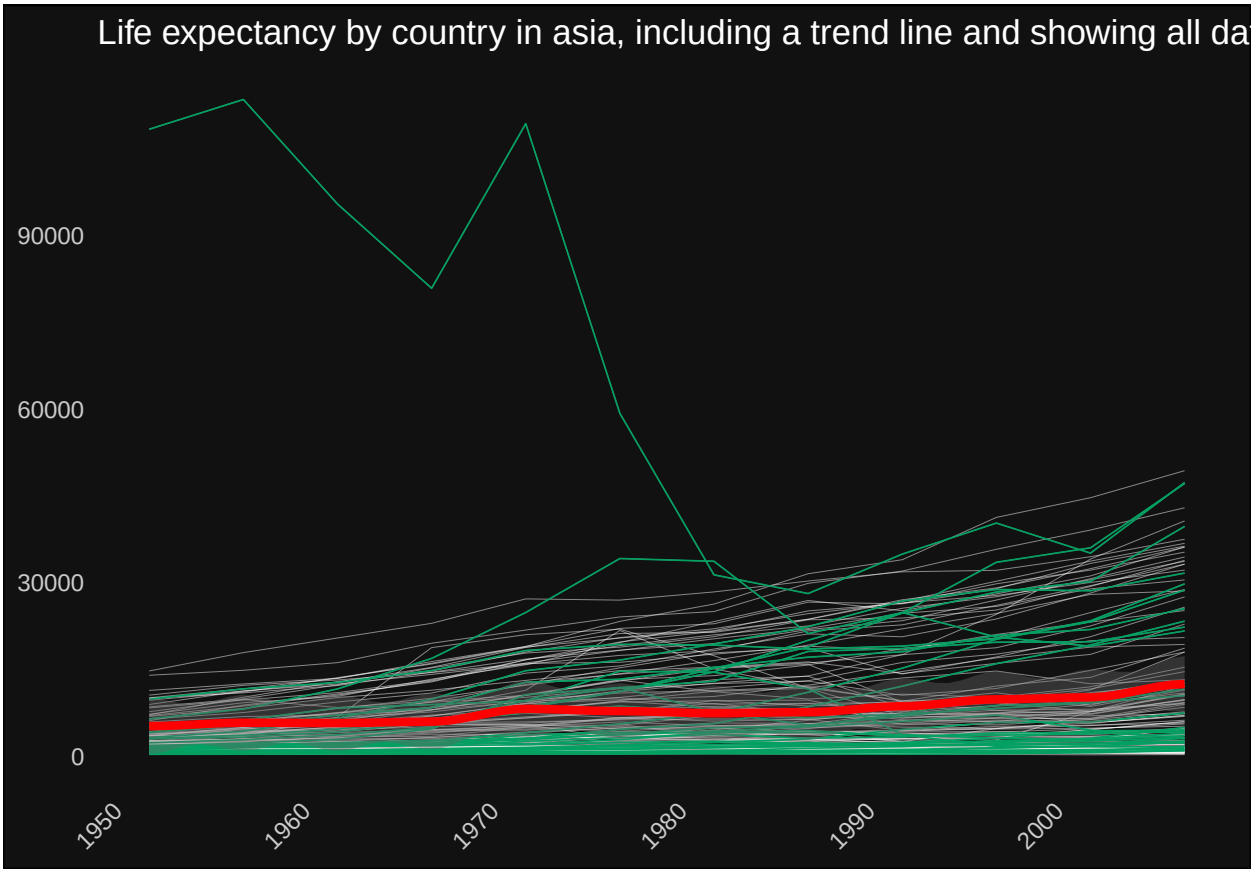
```
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 27.826

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 1951.7

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 5.275

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 27.826

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1951.7

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 5.275

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 27.826

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 1951.7

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 5.275

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0
```

```
## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 27.826

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : at 1951.7

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : radius 0.075625

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1951.7

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 0.275

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 1

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : at 2007.3

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : radius 0.075625

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : all data on boundary of neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 0.075625

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
```

```
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : zero-width neighborhood. make span bigger

## Warning: Computation failed in `stat_smooth()`:
## NA/NaN/Inf in foreign function call (arg 5)
```

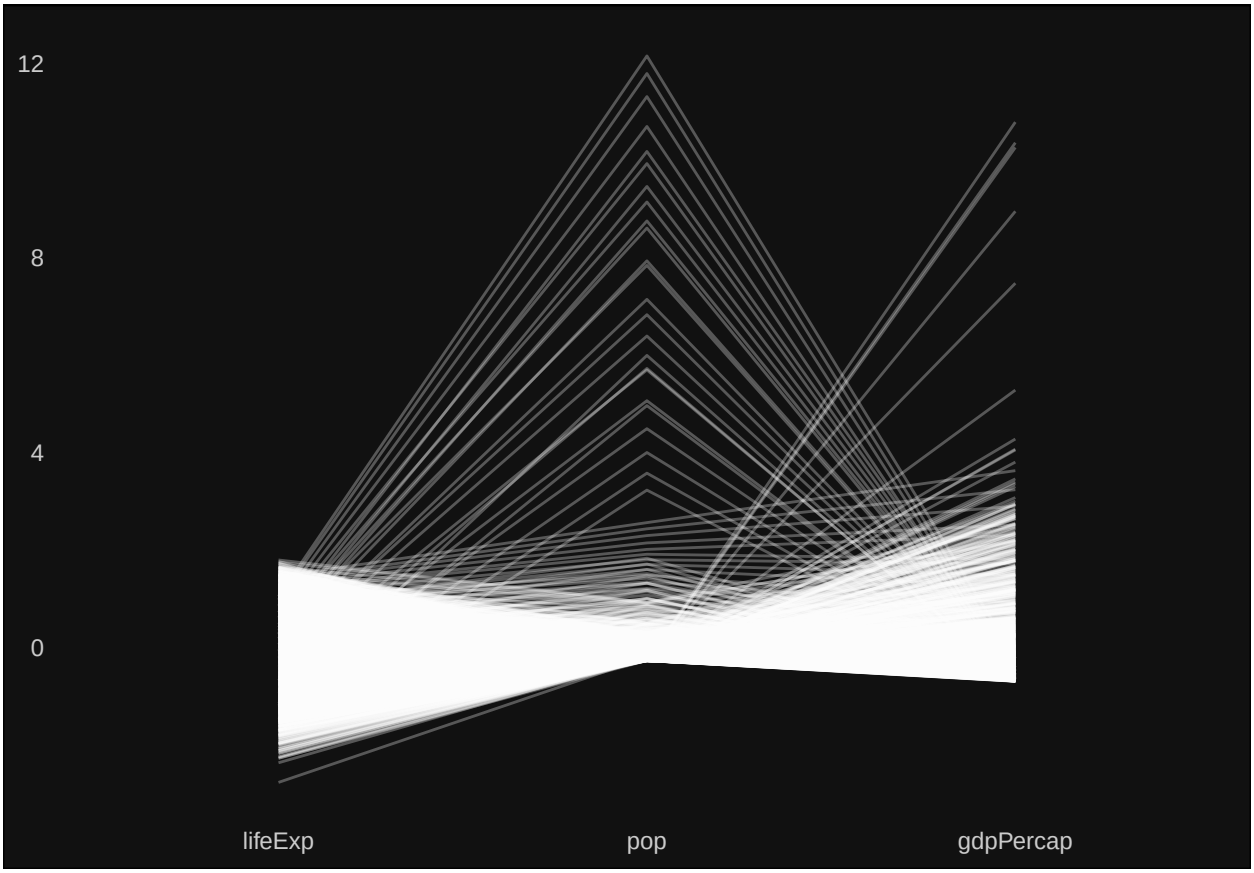


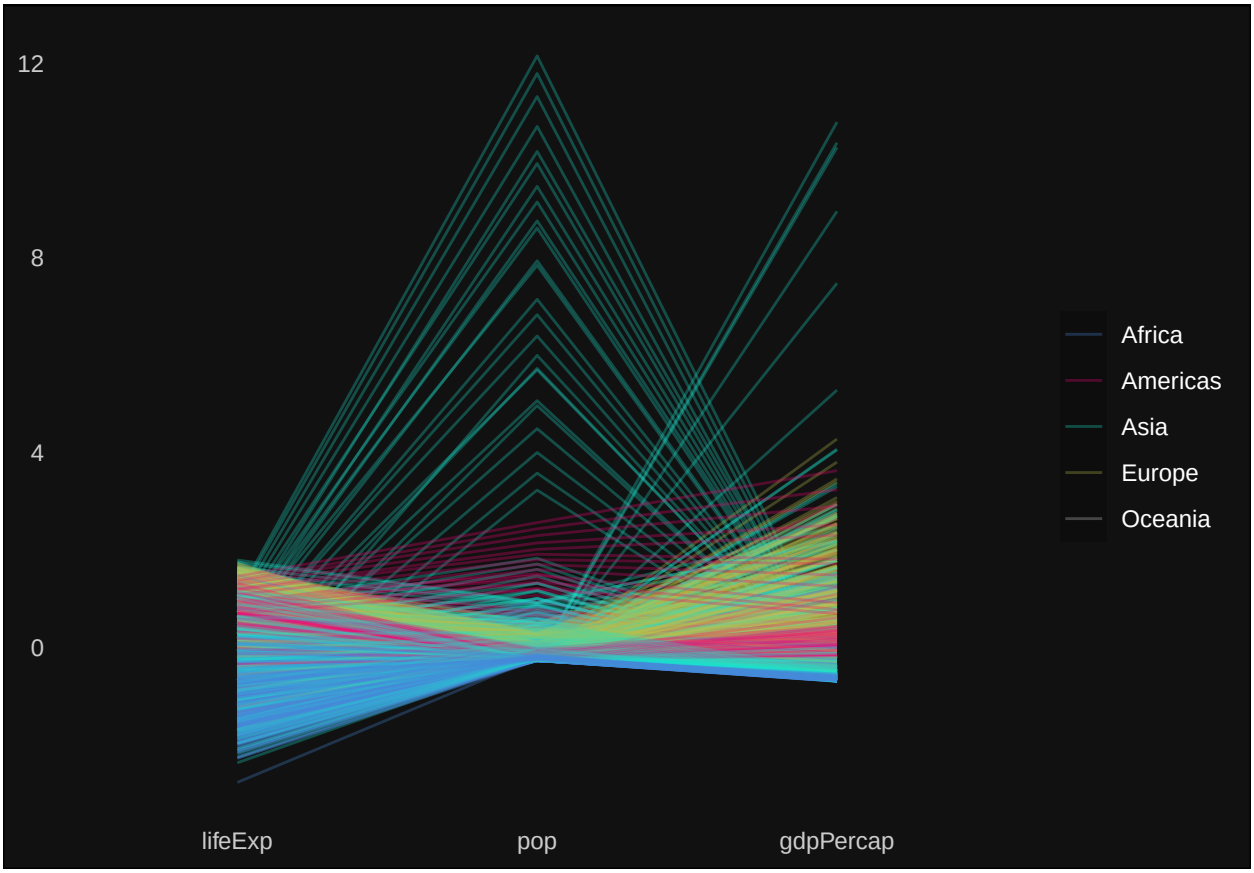
```
## `geom_smooth()` using formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 1951.7

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 5.275

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 27.826

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
## 1951.7

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
## 5.275

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
## number 0

## Warning in predLoess(object$y, object$x, newx = if
## (is.null(newdata)) object$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object)), : There are other near
## singularities as well. 27.826
```

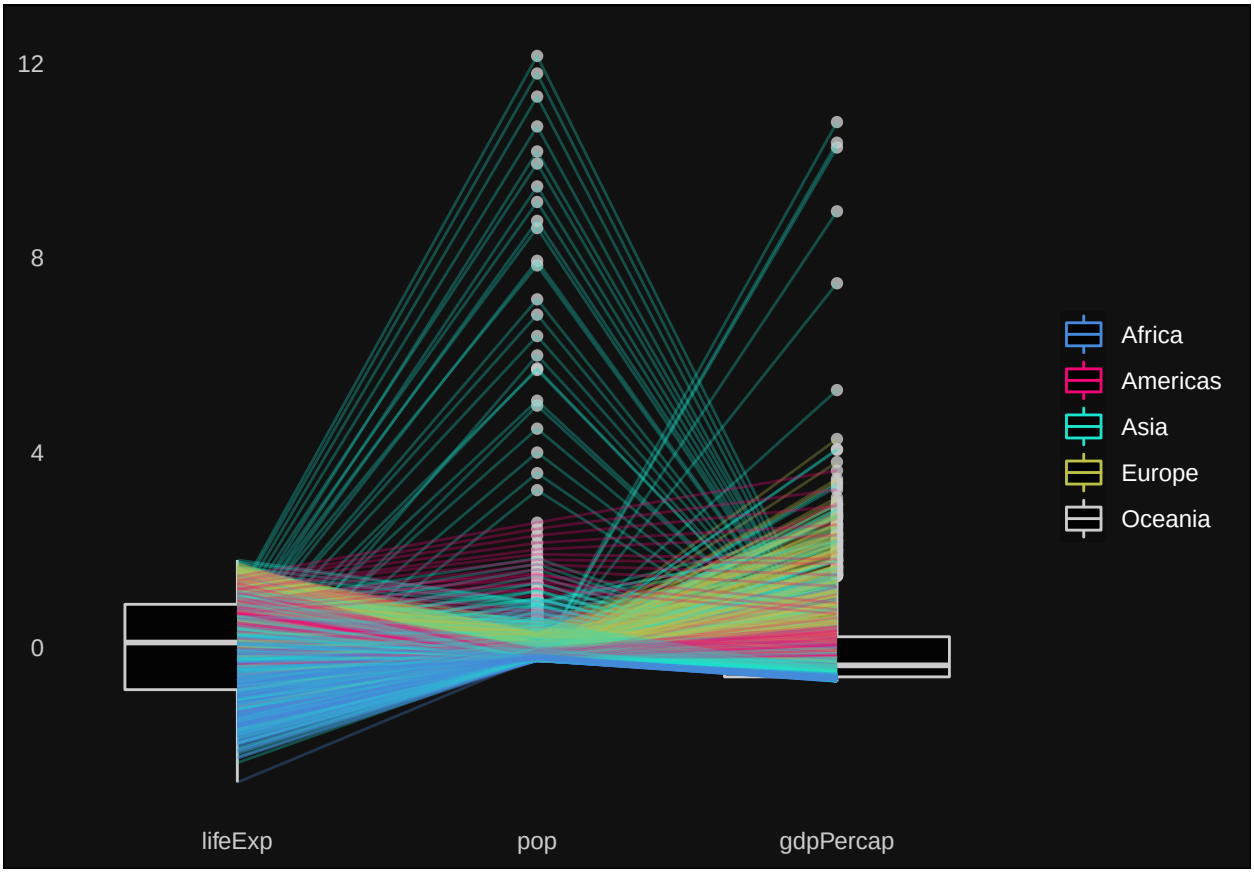


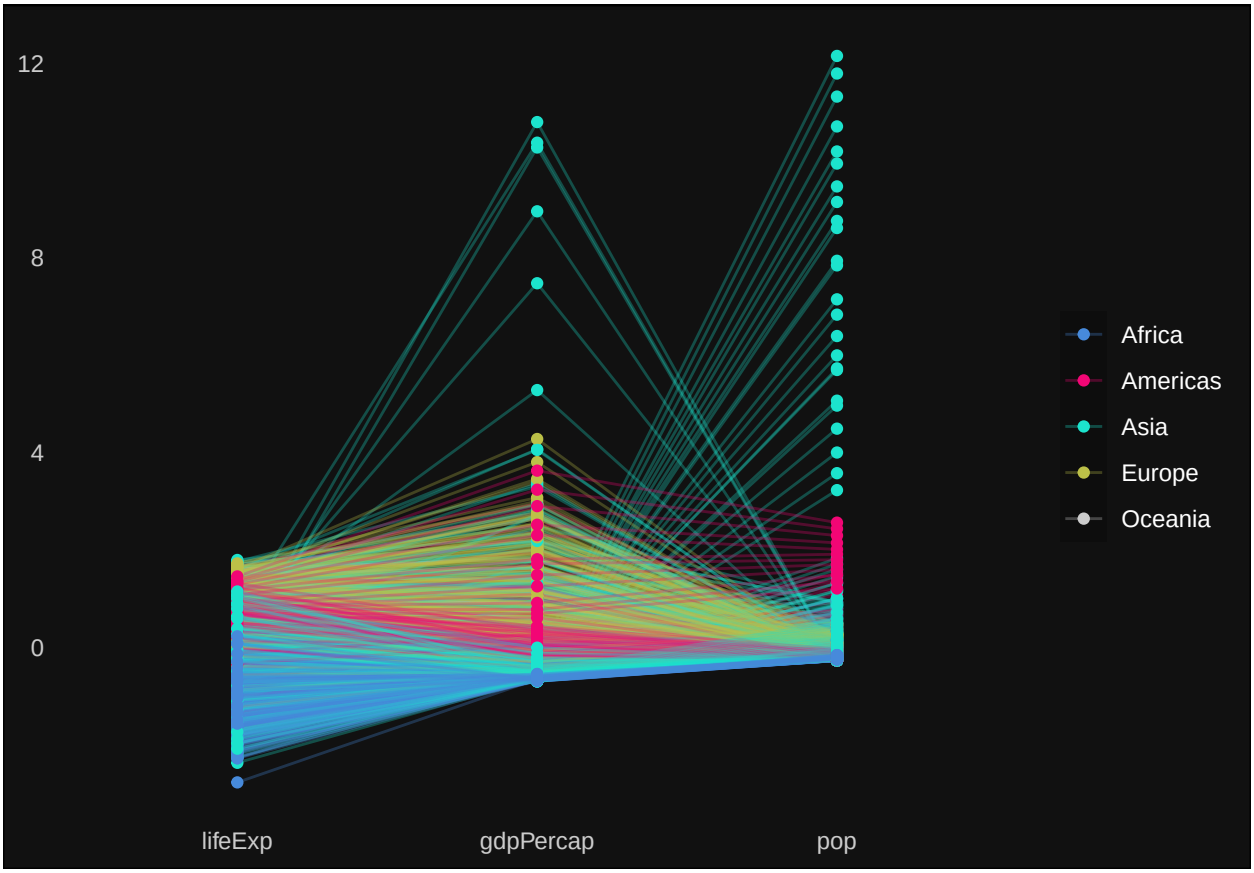
Exercise 12

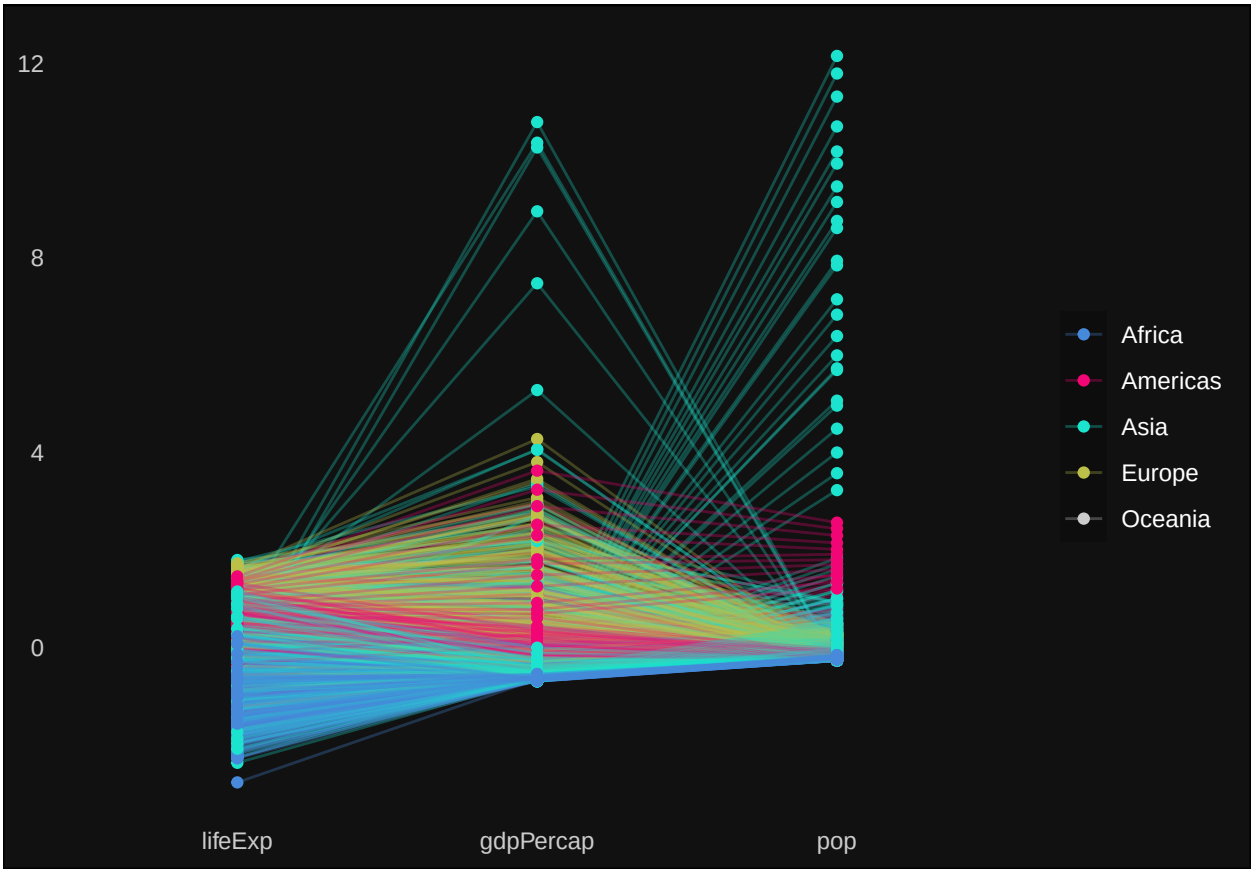
Create different versions of parallel coordinates charts using the gapminder dataset. Also try small multiple.

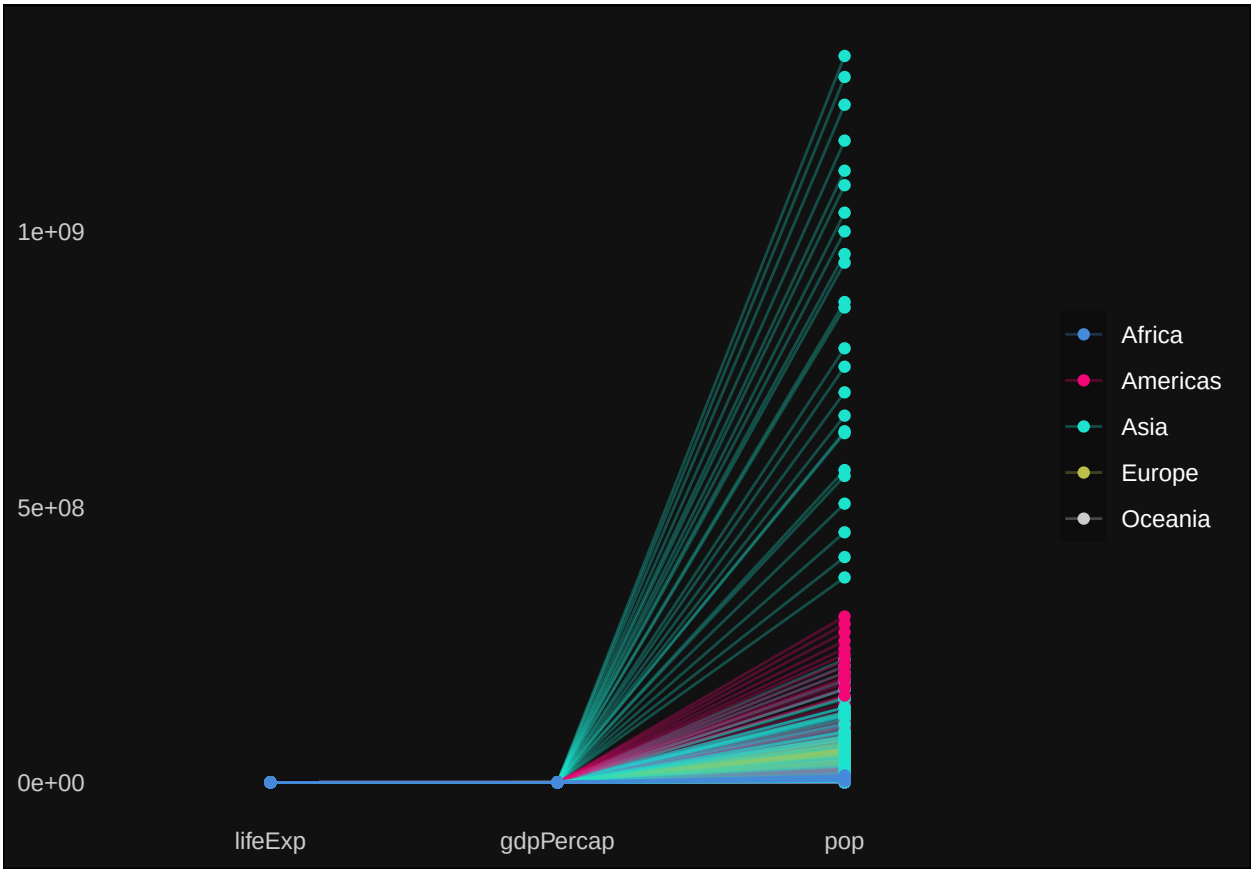

12
8
4
0
lifeExp
pop
gdpPercap

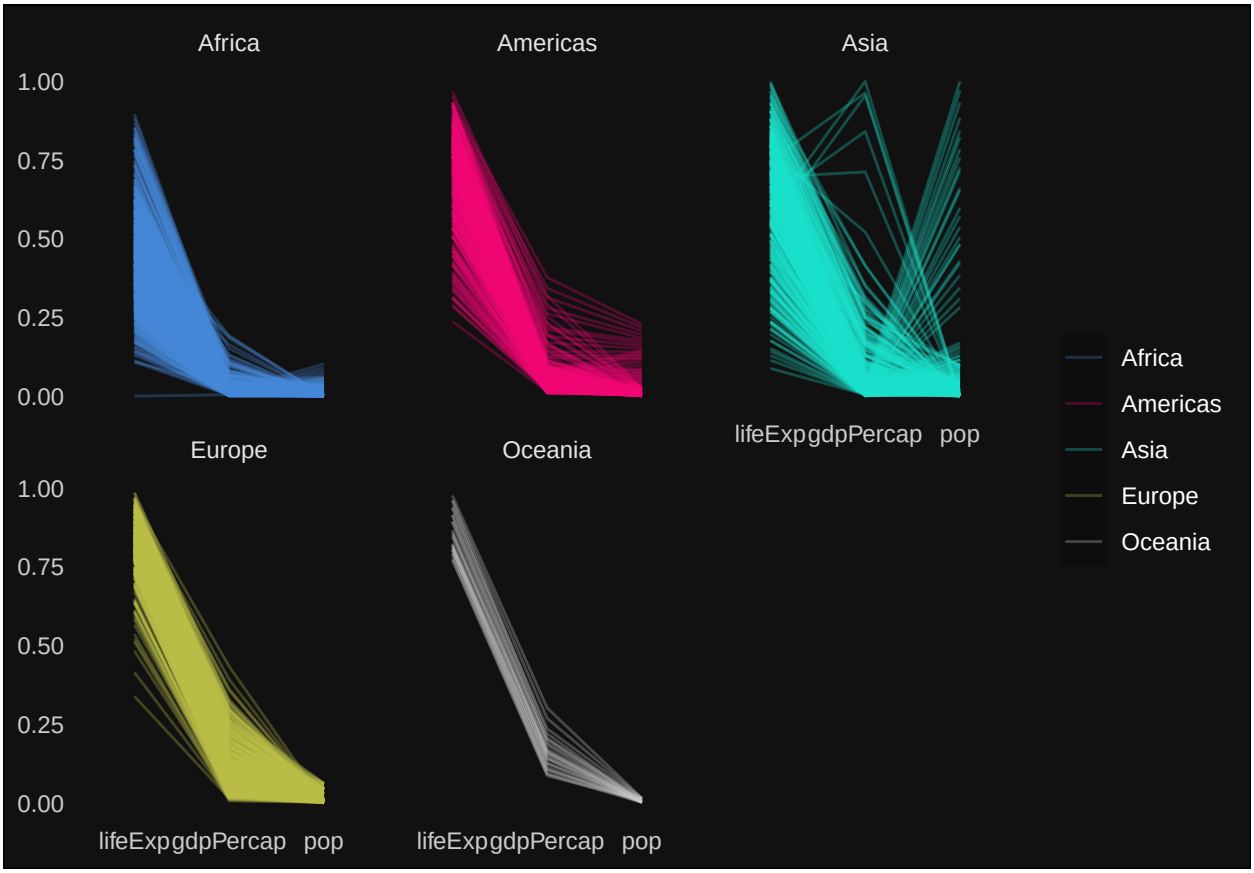

12
8
4
0
lifeExp
pop
gdpPercap
Africa
Americas
Asia
Europe
Oceania

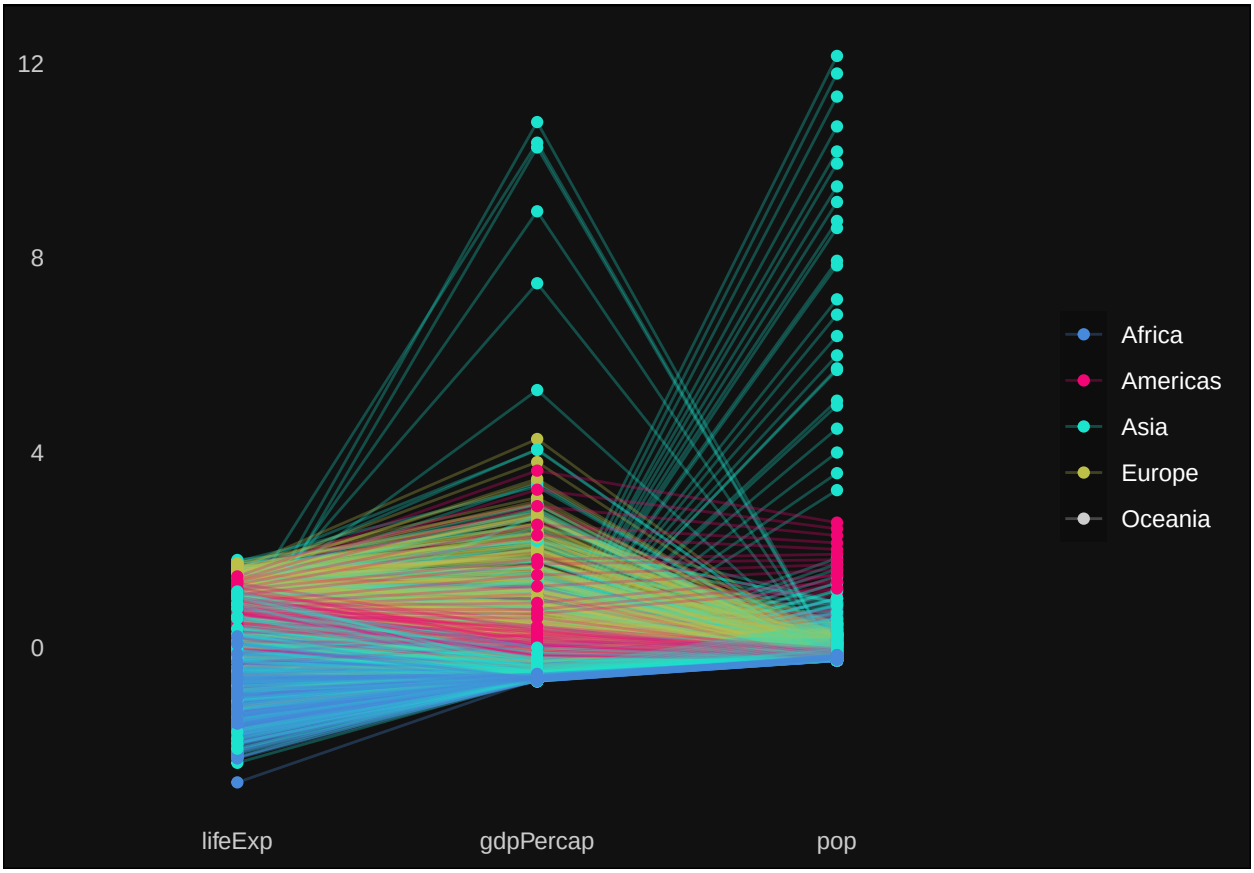

12
8
4
0
lifeExp
pop
gdpPercap
Africa
Americas
Asia
Europe
Oceania

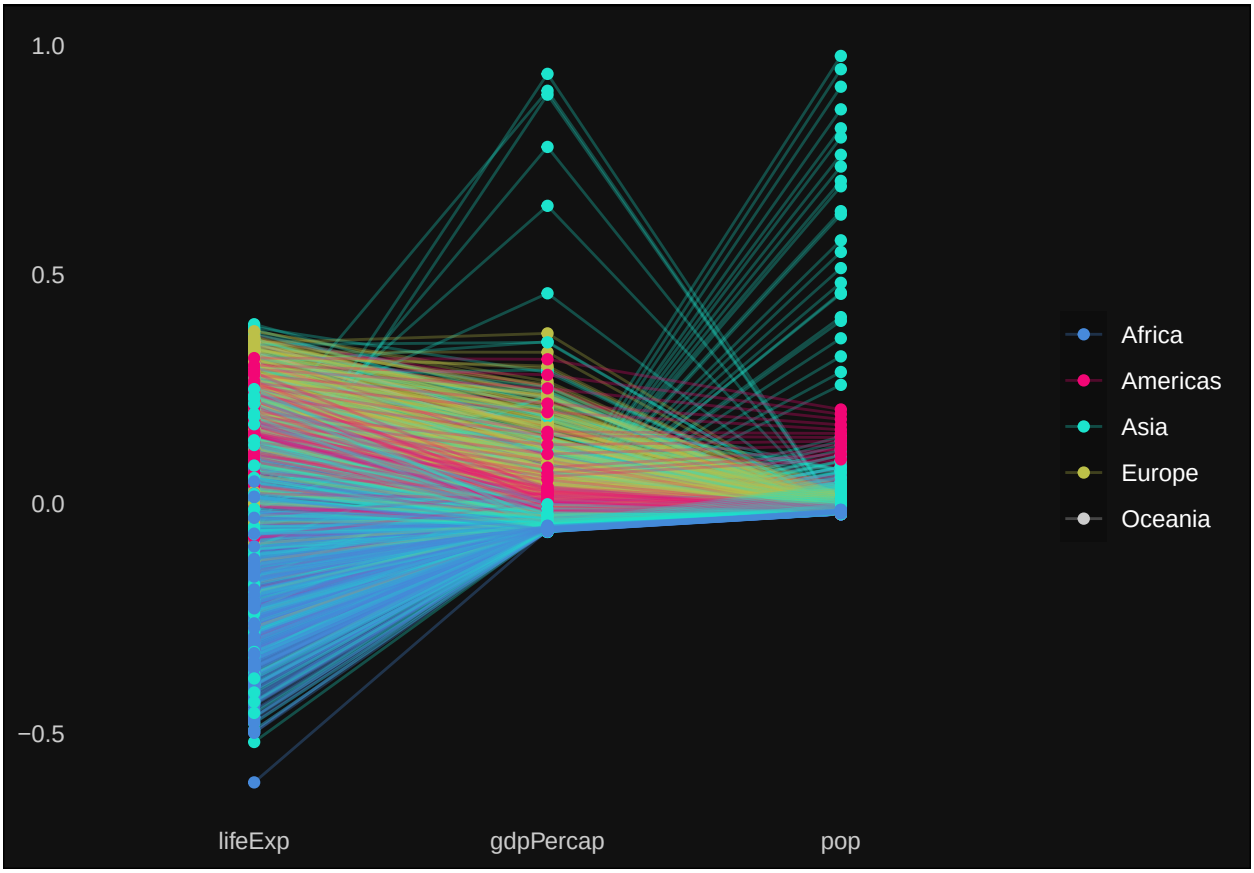

12
8
4
0
lifeExp
gdpPercap
pop
Africa
Americas
Asia
Europe
Oceania


12
8
4
0
lifeExp
gdpPercap
pop
Africa
Americas
Asia
Europe
Oceania


1e+09
5e+08
0e+00
lifeExp
gdpPercap
pop
Africa
Americas
Asia
Europe
Oceania


Africa
Americas
Asia
Europe
Oceania
1.00
0.75
0.50
0.25
0.00
lifeExp
gdpPercap
pop
Africa
Americas
Asia
Europe
Oceania


12
8
4
0
lifeExp
gdpPercap
pop
Africa
Americas
Asia
Europe
Oceania



Exercise 13

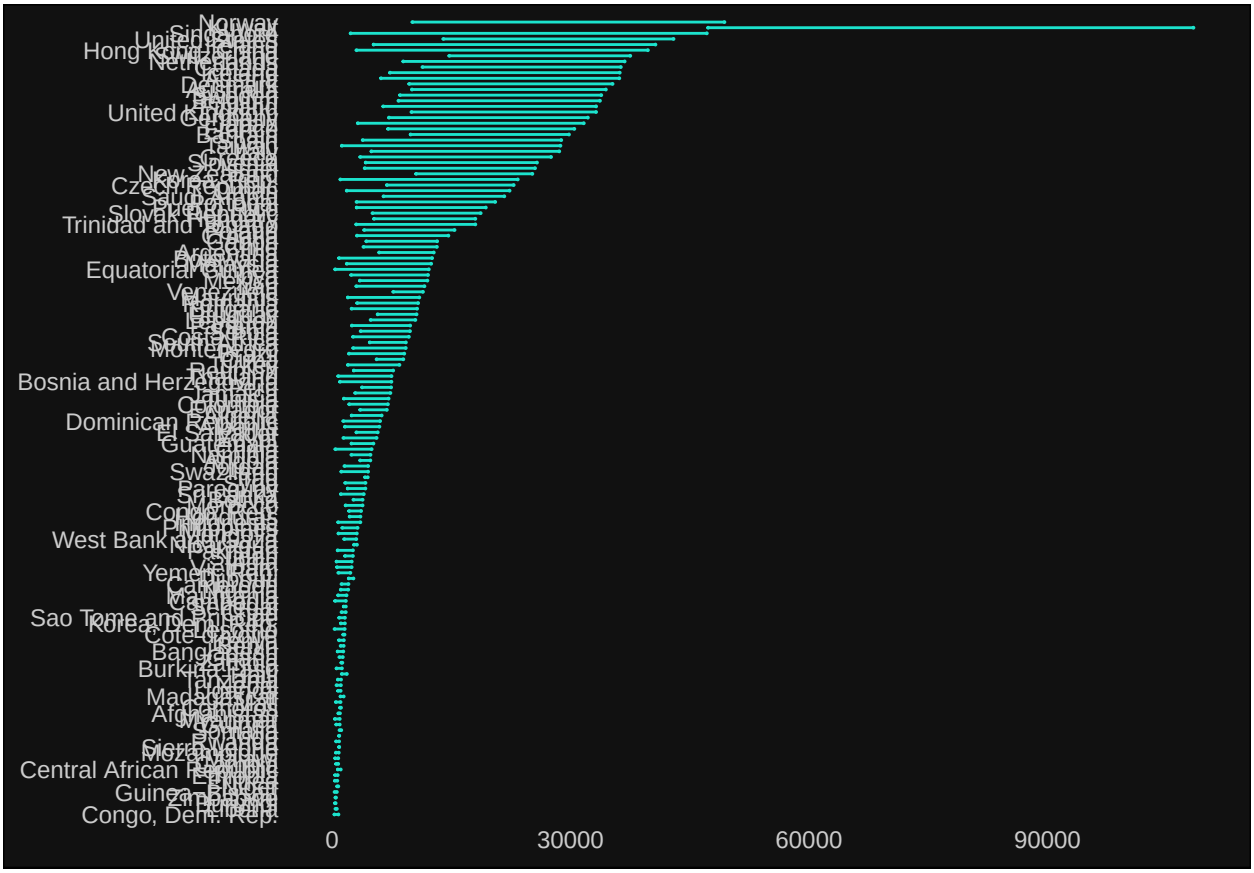
Now choose a different continent than Europe and also sort by 2007 instead of 1952.

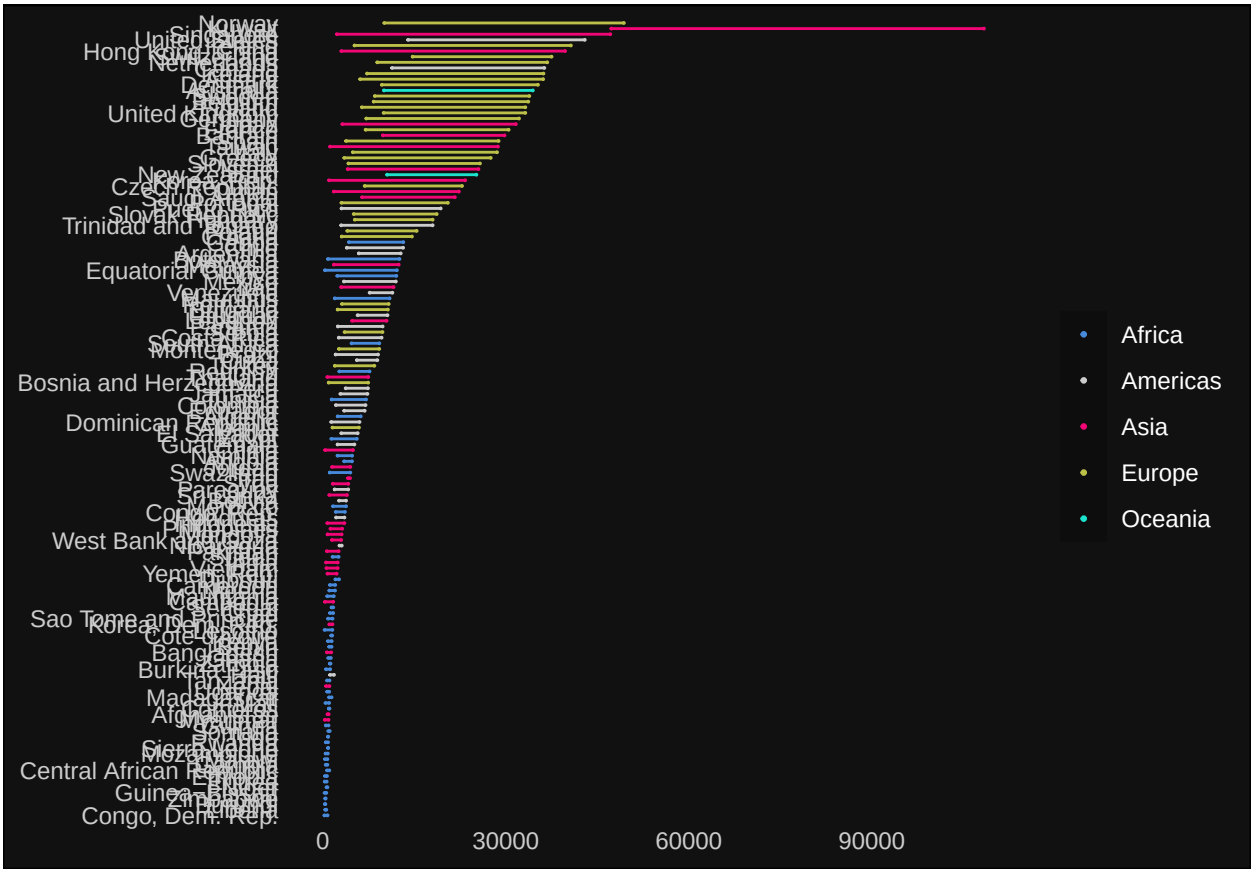
```
## [1] "country"   "continent" "year"      "gdpPercap"
```

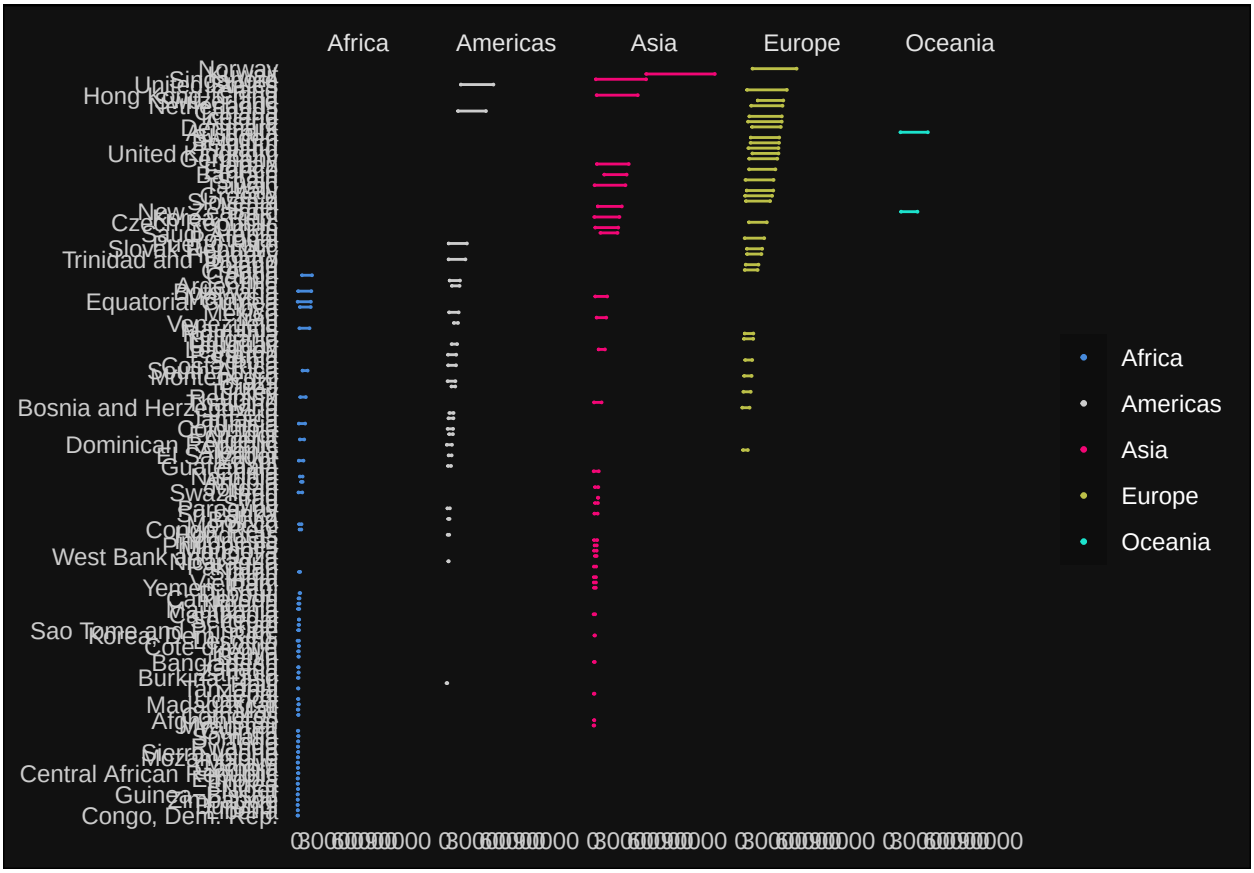
```
## # A tibble: 10 x 4
##    country     continent  year gdpPercap
##    <fct>       <fct>     <int>     <dbl>
##  1 Afghanistan Asia       1952      779.
##  2 Afghanistan Asia       2007      975.
##  3 Albania     Europe     1952     1601.
##  4 Albania     Europe     2007     5937.
##  5 Algeria     Africa     1952     2449.
##  6 Algeria     Africa     2007     6223.
##  7 Angola      Africa     1952     3521.
##  8 Angola      Africa     2007     4797.
##  9 Argentina   Americas   1952     5911.
## 10 Argentina   Americas   2007    12779.
```

```
## tibble [284 x 4] (S3: tbl_df/tbl/data.frame)
##  $ country  : Factor w/ 142 levels "Afghanistan",..: 1 1 2 2 3 3 4 4 5 5 ...
##  $ continent: Factor w/ 5 levels "Africa","Americas",..: 3 3 4 4 1 1 1 1 2 2 ...
##  $ year     : int [1:284] 1952 2007 1952 2007 1952 2007 1952 2007 1952 2007 ...
##  $ gdpPercap: num [1:284] 779 975 1601 5937 2449 ...
```

```
##         country       continent        year        gdpPercap
##  Afghanistan:  2   Africa  :104   Min.   :1952   Min.   :   277.6
##  Albania    :  2   Americas: 50   1st Qu.:1952   1st Qu.:  1106.4
##  Algeria    :  2   Asia    : 66   Median :1980   Median :  3051.4
##  Angola     :  2   Europe  : 60   Mean   :1980   Mean   :  7702.7
##  Argentina  :  2   Oceania :  4   3rd Qu.:2007   3rd Qu.:  8943.2
##  Australia  :  2                  Max.   :2007   Max.   :108382.4
##  (Other)    :272
```

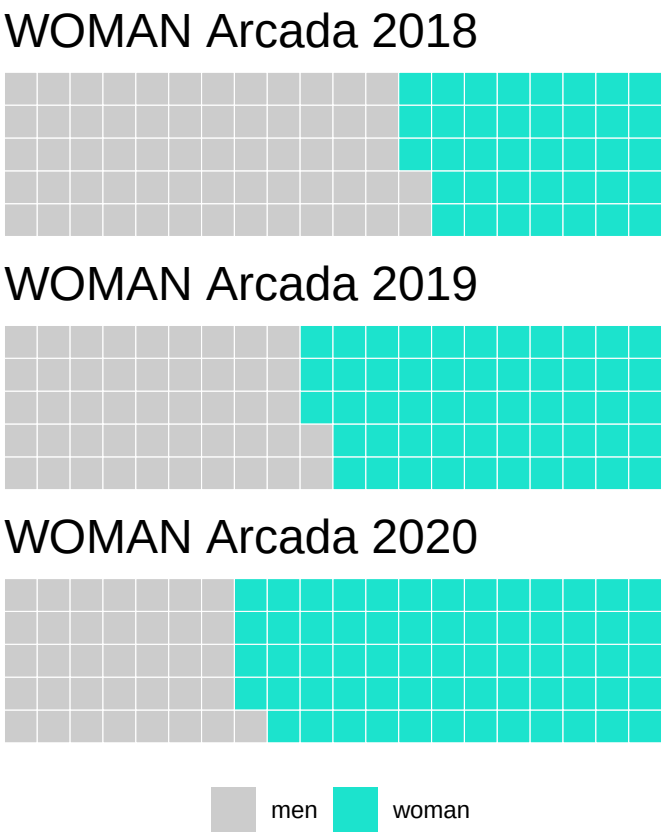



Africa
Americas
Asia
Europe
Oceania
0
30000
60000
90000



Exercise 14

Create your own small multiple waffle chart by using different data and a nwe style.



Exercise 15

Put all the code of the results of the exercise into a markdown file ansd save it as a pdf.

Task2

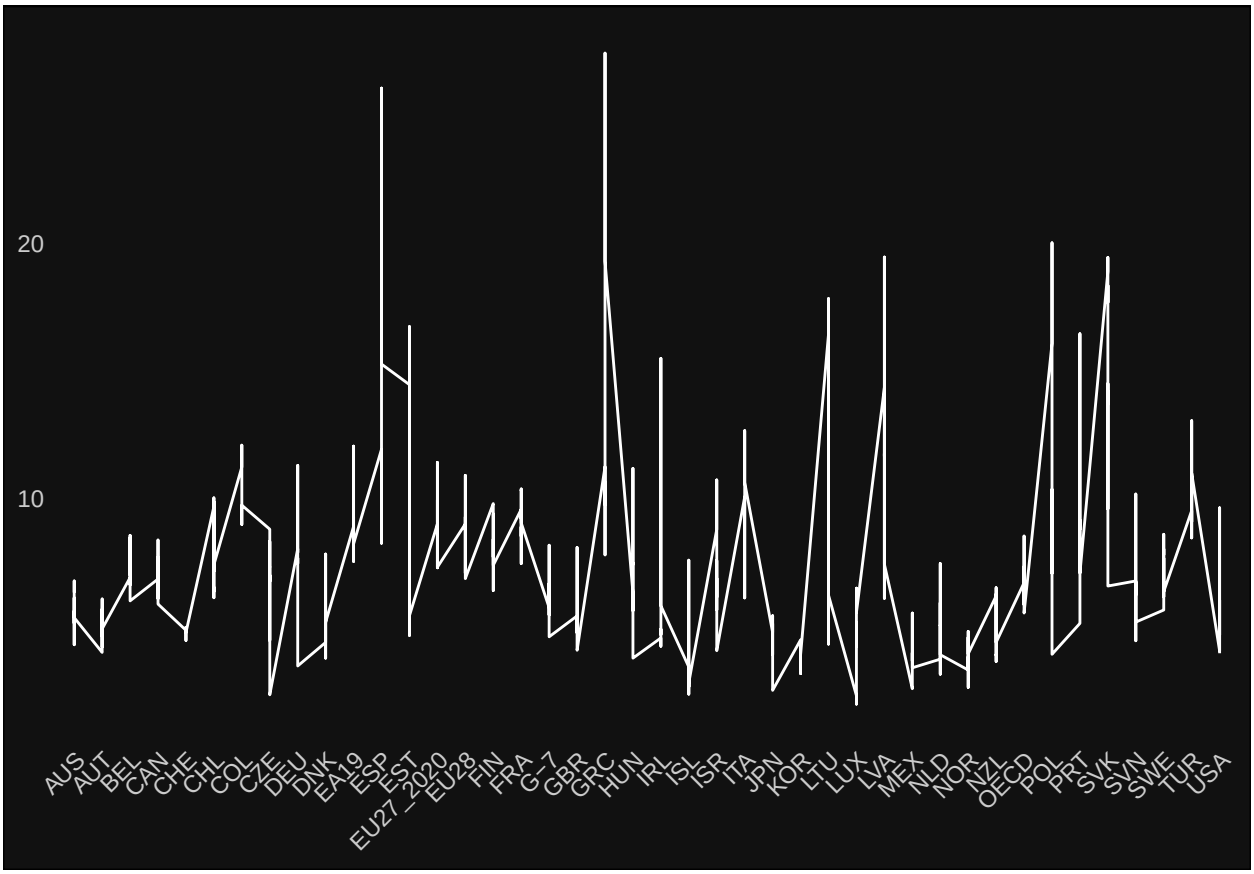
```
## [1] "ï..LOCATION" "INDICATOR"   "SUBJECT"     "MEASURE"     "FREQUENCY"
## [6] "TIME"        "Value"       "Flag.Codes"
```

```
##    ï..LOCATION INDICATOR SUBJECT MEASURE FREQUENCY TIME    Value Flag.Codes
## 1          AUS       HUR     TOT   PC_LF         A 2000 6.285546
## 2          AUS       HUR     TOT   PC_LF         A 2001 6.742173
## 3          AUS       HUR     TOT   PC_LF         A 2002 6.368911
## 4          AUS       HUR     TOT   PC_LF         A 2003 5.928420
## 5          AUS       HUR     TOT   PC_LF         A 2004 5.396734
## 6          AUS       HUR     TOT   PC_LF         A 2005 5.033881
## 7          AUS       HUR     TOT   PC_LF         A 2006 4.785240
## 8          AUS       HUR     TOT   PC_LF         A 2007 4.379151
## 9          AUS       HUR     TOT   PC_LF         A 2008 4.234330
## 10         AUS       HUR     TOT   PC_LF         A 2009 5.560385
```

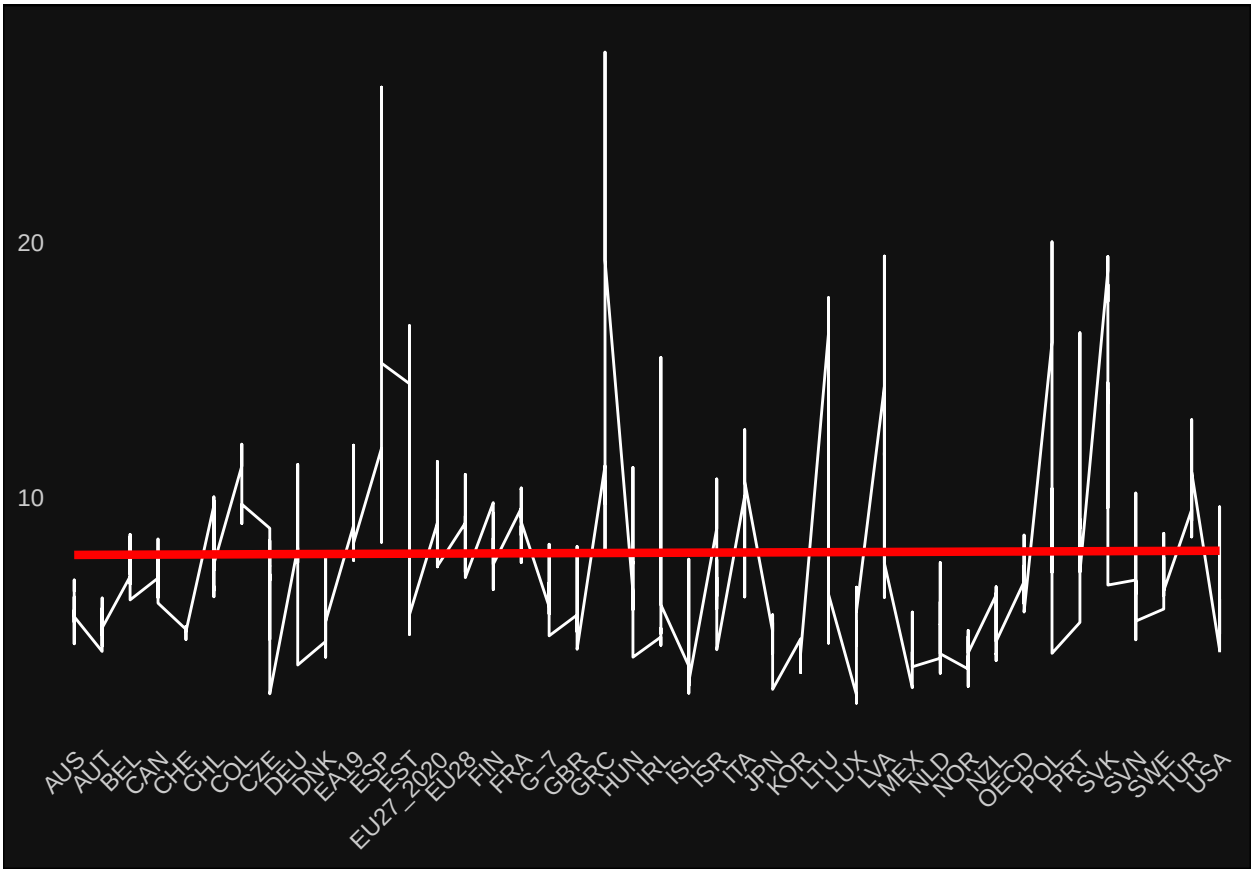
```
## 'data.frame':    772 obs. of  8 variables:
##  $ ï..LOCATION: chr  "AUS" "AUS" "AUS" "AUS" ...
##  $ INDICATOR  : chr  "HUR" "HUR" "HUR" "HUR" ...
```

```
##  $ SUBJECT    : chr  "TOT" "TOT" "TOT" "TOT" ...
##  $ MEASURE    : chr  "PC_LF" "PC_LF" "PC_LF" "PC_LF" ...
##  $ FREQUENCY  : chr  "A" "A" "A" "A" ...
##  $ TIME       : int  2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
##  $ Value      : num  6.29 6.74 6.37 5.93 5.4 ...
##  $ Flag.Codes : chr  "" "" "" "" ...
```

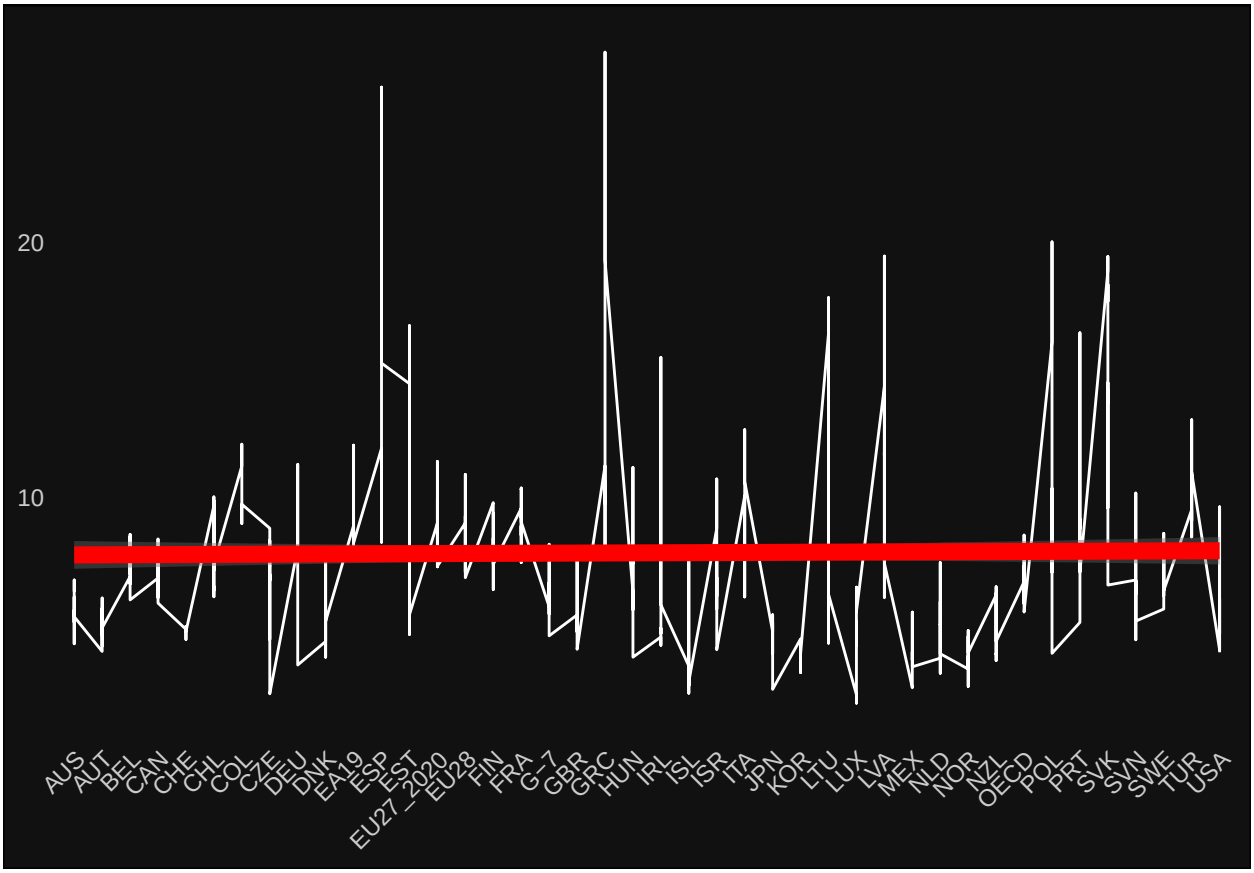
```
##  ï..LOCATION         INDICATOR           SUBJECT            MEASURE
##  Length:772         Length:772         Length:772         Length:772
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##   FREQUENCY             TIME          Value         Flag.Codes
##  Length:772         Min.   :2000   Min.   : 1.900   Length:772
##  Class :character   1st Qu.:2005   1st Qu.: 5.112   Class :character
##  Mode  :character   Median :2009   Median : 7.150   Mode  :character
##                     Mean   :2009   Mean   : 7.810
##                     3rd Qu.:2014   3rd Qu.: 9.267
##                     Max.   :2018   Max.   :27.492
```

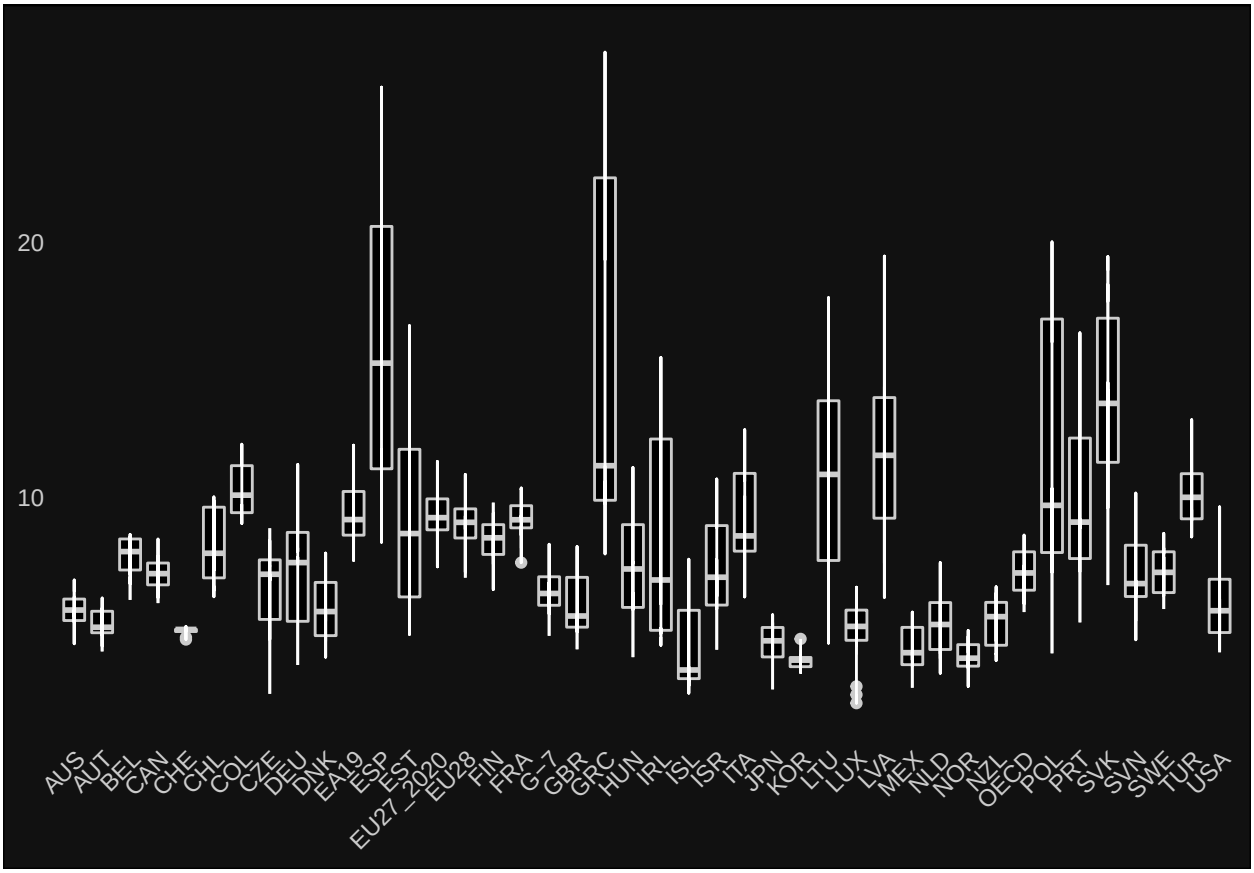


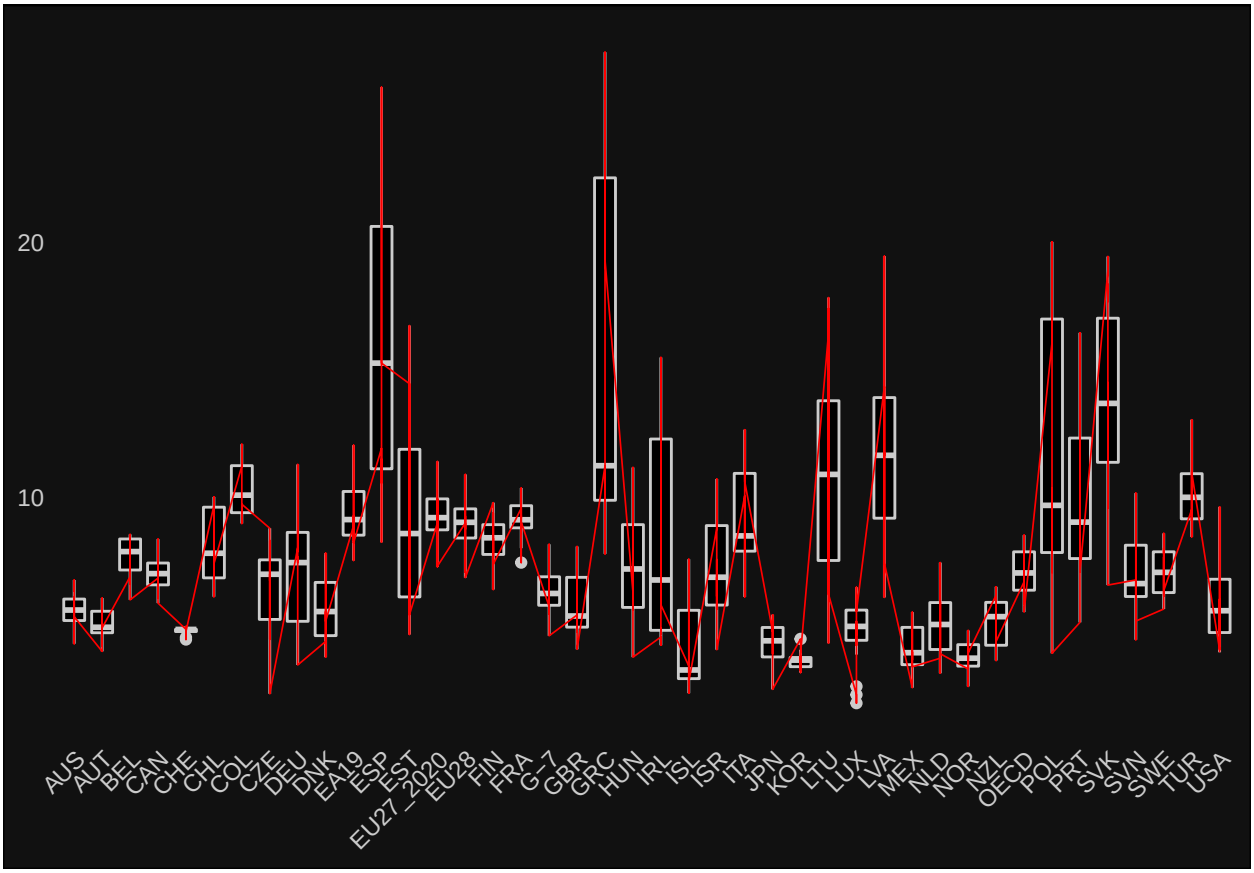

```
## `geom_smooth()` using formula 'y ~ x'
```



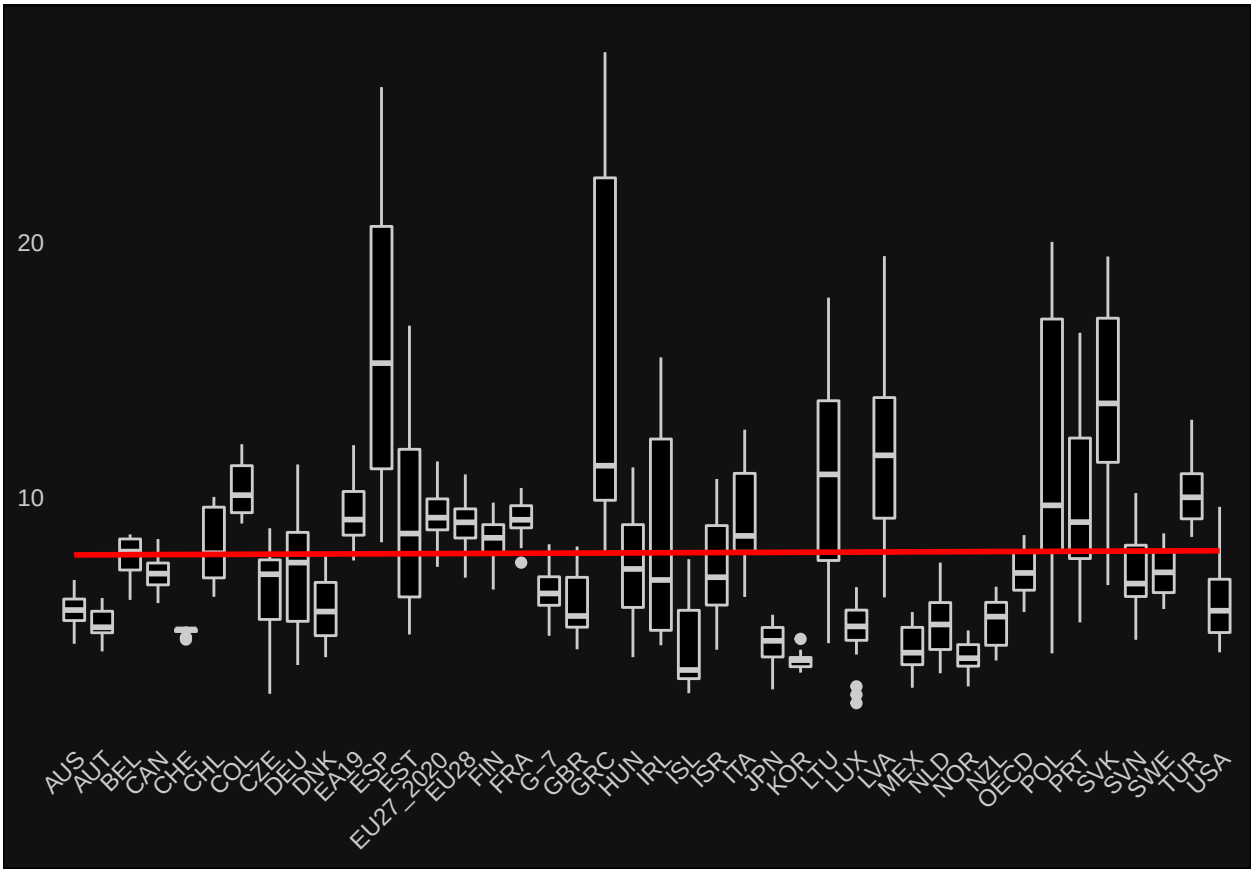

```
## `geom_smooth()` using formula 'y ~ x'
```


20
10
AUS
AUT
BEL
CAN
CHE
CHL
COL
CZE
DEU
DNK
EA19
ESP
EST
EU27_2020
EU28
FIN
FRA
G-7
GBR
GRC
HUN
IRL
ISL
ISR
ITA
JPN
KOR
LTU
LUX
LVA
MEX
NLD
NOR
NZL
OECD
POL
PRT
SVK
SVN
SWE
TUR
USA


20
10
AUS
AUT
BEL
CAN
CHE
CHL
COL
CZE
DEU
DNK
EA19
ESP
EST
EU27_2020
EU28
FIN
FRA
G-7
GBR
GRC
HUN
IRL
ISL
ISR
ITA
JPN
KOR
LTU
LUX
LVA
MEX
NLD
NOR
NZL
OECD
POL
PRT
SVK
SVN
SWE
TUR
USA




```
## `geom_smooth()` using formula 'y ~ x'
```


20
10
AUS
AUT
BEL
CAN
CHE
CHL
COL
CZE
DEU
DNK
EA19
ESP
EST
EU27_2020
EU28
FIN
FRA
G-7
GBR
GRC
HUN
IRL
ISL
ISR
ITA
JPN
KOR
LTU
LUX
LVA
MEX
NLD
NOR
NZL
OECD
POL
PRT
SVK
SVN
SWE
TUR
USA