

STA5075: Practical 4

Jessica Stow (STWJES003@myuct.ac.za)

2025-02-05

Produce a beautiful plot in a document:

- Gapminder: Plot life expectancy against GDP, don't attach, colour by continent.
- Does a log-transformation help to bring out the information more clearly?
- R Markdown document (html, word, pdf)
- Figure caption
- Axis labels and sizing. Improve visually. Legend.
- Improve size and placement.

```
library(gapminder)
```

```
str(gapminder)
```

```
## tibble [1,704 x 6] (S3: tbl_df/tbl/data.frame)
## $ country   : Factor w/ 142 levels "Afghanistan",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ continent: Factor w/ 5 levels "Africa","Americas",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ year      : int [1:1704] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
## $ lifeExp   : num [1:1704] 28.8 30.3 32 34 36.1 ...
## $ pop       : int [1:1704] 8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 163
## $ gdpPercap: num [1:1704] 779 821 853 836 740 ...
```

```
nlevels(gapminder$continent) # 5 levels
```

```
## [1] 5
```

```
levels(gapminder$continent)
```

```
## [1] "Africa" "Americas" "Asia" "Europe" "Oceania"
```

```
summary(gapminder$gdpPercap)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##    241.2   1202.1   3531.8   7215.3   9325.5  113523.1
```

```
summary(gapminder$lifeExp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    23.60  48.20   60.71   59.47  70.85   82.60
```

```
# Plot life expectancy against GDP
```

```
plot(seq(min(gapminder$gdpPercap), max(gapminder$gdpPercap), length = 1000),
     seq(min(gapminder$lifeExp), max(gapminder$lifeExp), length = 1000),
     type = "n",
     xlab = "GDP per capita in US $",
     ylab = "Life expectancy (years)",
     main = "Relationship between life expectancy (years) and GDP per capita (US $)")
```

```

# Add the points
# AFRICA
africa <- gapminder[gapminder$continent == "Africa",]
points(africa$gdpPerCap, africa$lifeExp, col = "red")

# AMERICAS
americas <- gapminder[gapminder$continent == "Americas",]
points(americas$gdpPerCap, americas$lifeExp, col = "blue")

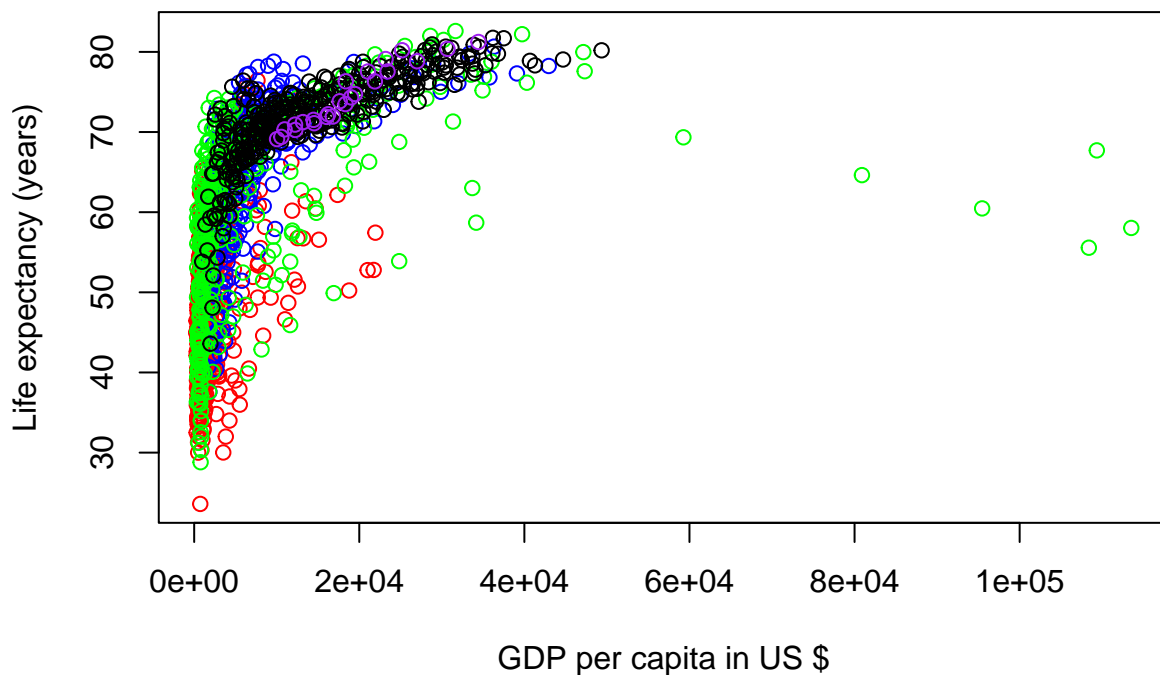
# ASIA
asia <- gapminder[gapminder$continent == "Asia",]
points(asia$gdpPerCap, asia$lifeExp, col = "green")

# EUROPE
europe <- gapminder[gapminder$continent == "Europe",]
points(europe$gdpPerCap, europe$lifeExp, col = "black")

# OCEANIA
oceania <- gapminder[gapminder$continent == "Oceania",]
points(oceania$gdpPerCap, oceania$lifeExp, col = "purple")

```

Relationship between life expectancy (years) and GDP per capita (US



```

# Plot life expectancy against log of GDP

plot(seq(min(log(gapminder$gdpPercap)), max(log(gapminder$gdpPercap)), length = 1000),
     seq(min(gapminder$lifeExp), max(gapminder$lifeExp), length = 1000),
     type = "n",
     cex.lab = 1.2, # axis label size
     cex.axis = 0.8,
     xlab = "GDP per capita in US $ (Log transformed)",
     ylab = "Life expectancy (years)",
     main = "Relationship between life expectancy (years) and GDP per capita (US $)")

# Add the points LOG TRANSFORMED
# AFRICA
africa <- gapminder[gapminder$continent == "Africa",]
points(log(africa$gdpPercap), africa$lifeExp, col = "red")

# AMERICAS
americas <- gapminder[gapminder$continent == "Americas",]
points(log(americas$gdpPercap), americas$lifeExp, col = "blue")

# ASIA
asia <- gapminder[gapminder$continent == "Asia",]
points(log(asia$gdpPercap), asia$lifeExp, col = "green")

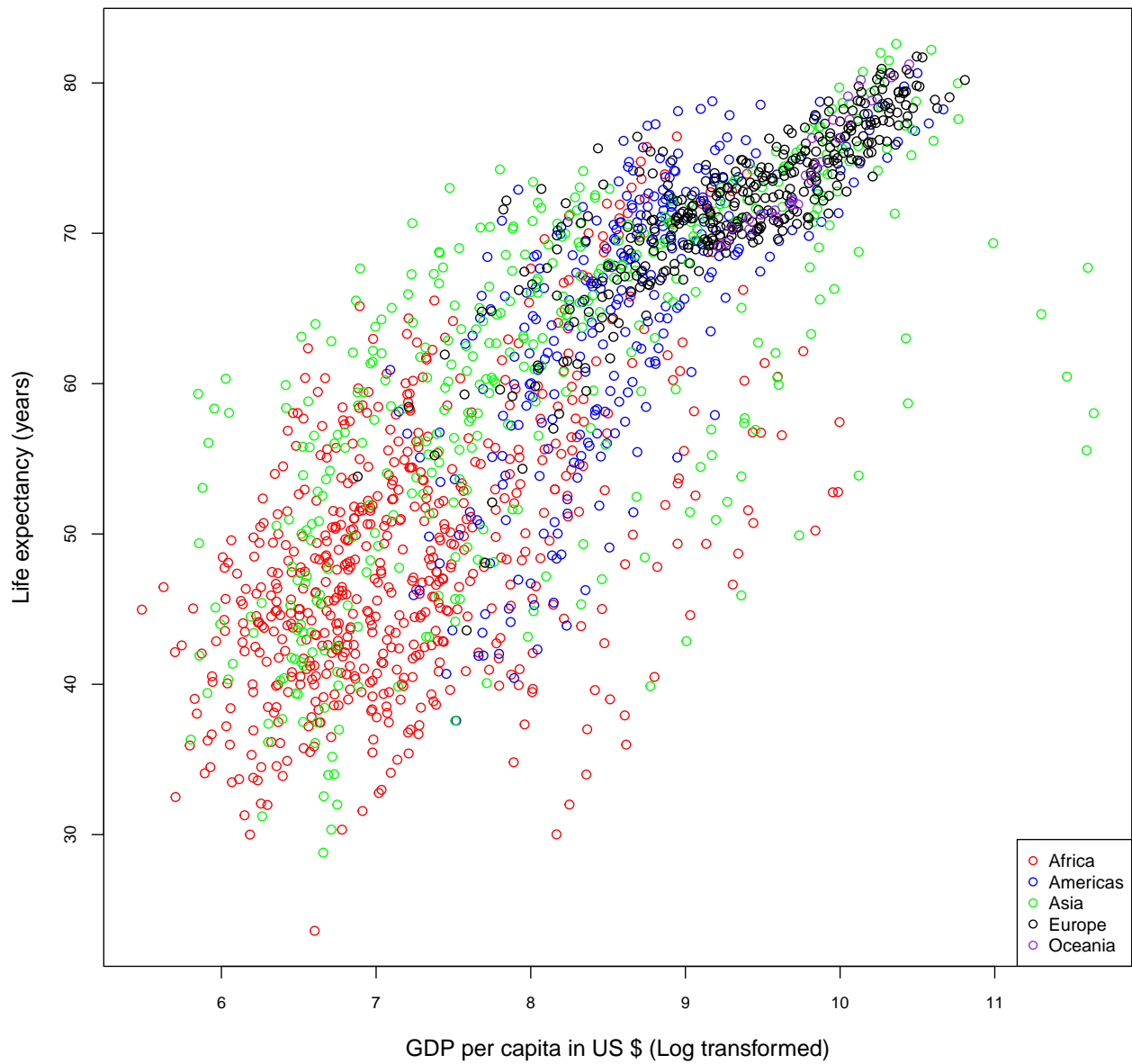
# EUROPE
europe <- gapminder[gapminder$continent == "Europe",]
points(log(europe$gdpPercap), europe$lifeExp, col = "black")

# OCEANIA
oceania <- gapminder[gapminder$continent == "Oceania",]
points(log(oceania$gdpPercap), oceania$lifeExp, col = "purple")

# Legend
legend("bottomright",
      legend = c("Africa", "Americas", "Asia", "Europe", "Oceania"),
      col = c("red", "blue", "green", "black", "purple"),
      pch = 1,
      cex = 0.9)

```

Relationship between life expectancy (years) and GDP per capita (US \$)



- How many different countries occur in this data set?

```
length(unique(gapminder$country)) # 142 countries
```

```
## [1] 142
```

- How many African countries?

```
africa <- gapminder[gapminder$continent == "Africa",]
length(unique(africa$country)) # 52 African countries
```

```
## [1] 52
```

- Which countries have the lowest and highest life expectancy, respectively? In which years? (There are several observations/years per country).

```
lowest <- min(gapminder$lifeExp)
gapminder[gapminder$lifeExp == lowest,] # Rwanda 1992, life expectancy of 23.599 years
```

```
## # A tibble: 1 x 6
##   country continent  year lifeExp      pop gdpPercap
##   <fct>    <fct>    <int>   <dbl>   <int>    <dbl>
## 1 Rwanda  Africa      1992    23.6 7290203     737.
```

```
highest <- max(gapminder$lifeExp)
gapminder[gapminder$lifeExp == highest,] # Japan 2007, life expectancy of 82.603 years
```

```
## # A tibble: 1 x 6
##   country continent  year lifeExp      pop gdpPercap
##   <fct>    <fct>    <int>   <dbl>   <int>    <dbl>
## 1 Japan   Asia        2007    82.6 127467972   31656.
```