

0.1 Example of R in Colab

```
[1]: # set figure size for plots
options(repr.plot.width=12, repr.plot.height=7)
```

```
[2]: # read our usual world bank data about countries
countries = read.csv("https://fmfi-compbio.github.io/viz/data/World_bank.csv")
```

```
[3]: # show summary of the table
summary(countries)
```

IS03	Country	Region	Income.Group
Length:217	Length:217	Length:217	Length:217
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

Population2000	Population2010	Population2020	Area
Min. :9.638e+03	Min. :1.024e+04	Min. :1.107e+04	Min. : 10
1st Qu.:6.050e+05	1st Qu.:7.055e+05	1st Qu.:7.972e+05	1st Qu.: 11222
Median :5.056e+06	Median :5.769e+06	Median :6.580e+06	Median : 101705
Mean :2.821e+07	Mean :3.201e+07	Mean :3.593e+07	Mean : 623593
3rd Qu.:1.639e+07	3rd Qu.:2.112e+07	3rd Qu.:2.565e+07	3rd Qu.: 478605
Max. :1.263e+09	Max. :1.338e+09	Max. :1.411e+09	Max. :17098250
			NA's :1

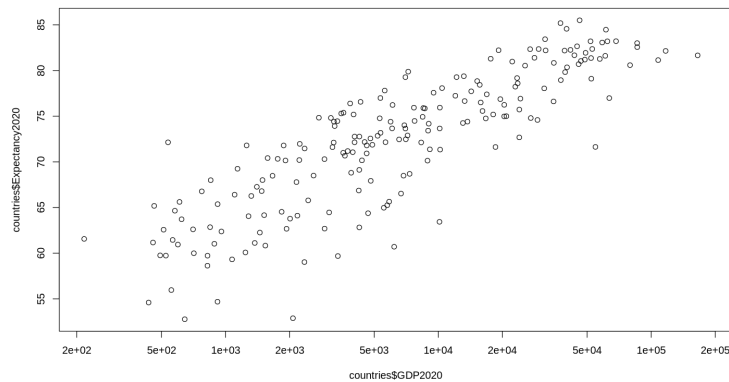
GDP2000	GDP2010	GDP2020	Expectancy2000
Min. : 123.0	Min. : 222.7	Min. : 216.8	Min. :44.52
1st Qu.: 654.6	1st Qu.: 1706.4	1st Qu.: 2262.2	1st Qu.:61.36
Median : 1996.5	Median : 5735.4	Median : 6370.9	Median :70.42
Mean : 8292.9	Mean : 16135.5	Mean : 17547.1	Mean :67.36
3rd Qu.:10721.3	3rd Qu.: 21447.9	3rd Qu.: 22805.3	3rd Qu.:74.50
Max. :81763.8	Max. :161780.7	Max. :182537.4	Max. :81.37
NA's :18	NA's :8	NA's :7	NA's :6

Expectancy2010	Expectancy2020	Fertility2000	Fertility2010
Min. :45.60	Min. :52.78	Min. :0.912	Min. :1.042
1st Qu.:64.48	1st Qu.:66.80	1st Qu.:1.840	1st Qu.:1.765
Median :72.81	Median :72.89	Median :2.660	Median :2.340
Mean :70.58	Mean :72.31	Mean :3.211	Mean :2.891
3rd Qu.:77.58	3rd Qu.:78.04	3rd Qu.:4.395	3rd Qu.:3.827
Max. :83.11	Max. :85.50	Max. :7.732	Max. :7.485
NA's :7	NA's :8	NA's :6	NA's :6

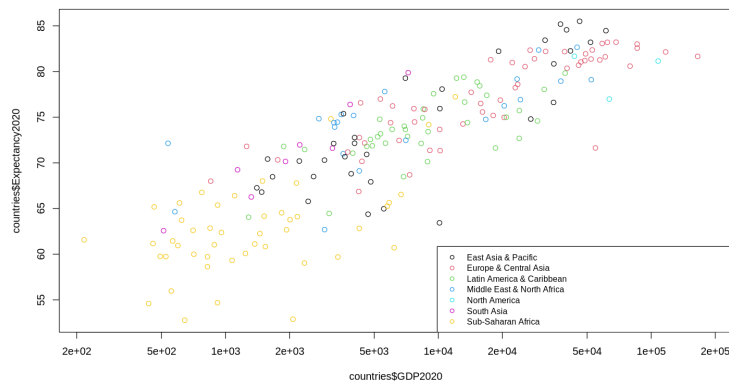
Fertility2020
Min. :0.837
1st Qu.:1.558
Median :2.040
Mean :2.525

3rd Qu.:3.257
Max. :6.892
NA's :6

```
[4]: # simple scatter plot of GDP (in log scale) vs life expectancy  
# using built-in plot function  
plot(countries$GDP2020, countries$Expectancy2020, log='x')
```



```
[5]: # a more complex version of the plot with region used as color  
# region is first converted to "factor", i.e. categorical variable  
region = factor(countries$Region)  
plot(countries$GDP2020, countries$Expectancy2020, col=region, log='x')  
legend('bottomright', legend = levels(region), col=1:nlevels(region), cex = 0.  
      ↪8, pch = 1)
```



```
[6]: # let us also try ggplot2 library  
library(ggplot2)
```

```
[7]: # use ggplot to plot for the same plot, but add region as color
      ggplot(countries, aes(x=GDP2020, y=Expectancy2020, colour = Region)) +
      geom_point() + scale_x_log10()
```

Warning message:

"Removed 15 rows containing missing values or values outside the scale range
(`geom_point()`)."

