19MID0020 Assignment-1

Code

Importing the necessary libraries

Hide

```
library(ggplot2)
```

```
Need help Try Stackoverflow: https://stackoverflow.com/tags/ggplot2
```

Hide

```
library(plotly)
```

```

Attaching package: 'plotly'

The following object is masked from 'package:ggplot2':

    last_plot

The following object is masked from 'package:stats':

    filter

The following object is masked from 'package:graphics':

    layout
```

First 5 records

Hide

```
df = read.csv('Expectancy.csv', header = TRUE, sep =",")
head(df)
```

	X <int>	country <chr>	continent <chr>	year <int>	lifeExp <dbl>	pop <dbl>	gdpPercap <dbl>
1	1	Afghanistan	Asia	1952	28.801	8425333	779.4453
2	2	Afghanistan	Asia	1957	30.332	9240934	820.8530
3	3	Afghanistan	Asia	1962	31.997	10267083	853.1007
4	4	Afghanistan	Asia	1967	34.020	11537966	836.1971
5	5	Afghanistan	Asia	1972	36.088	13079460	739.9811
6	6	Afghanistan	Asia	1977	38.438	14880372	786.1134

6 rows

Last 5 records

Hide

```
tail(df)
```

	X <int>	country <chr>	continent <chr>	year <int>	lifeExp <dbl>	pop <dbl>	gdpPercap <dbl>
1699	1699	Zimbabwe	Africa	1982	60.363	7636524	788.8550
1700	1700	Zimbabwe	Africa	1987	62.351	9216418	706.1573
1701	1701	Zimbabwe	Africa	1992	60.377	10704340	693.4208
1702	1702	Zimbabwe	Africa	1997	46.809	11404948	792.4500
1703	1703	Zimbabwe	Africa	2002	39.989	11926563	672.0386
1704	1704	Zimbabwe	Africa	2007	43.487	12311143	469.7093

6 rows

Data-type of every attribute

Hide

```
str(df)
```

```
'data.frame':   1704 obs. of  7 variables:
 $ X        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ country  : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ continent: chr  "Asia" "Asia" "Asia" "Asia" ...
 $ year     : int  1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
 $ lifeExp  : num  28.8 30.3 32 34 36.1 ...
 $ pop      : num  8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 16317921 22227415 ...
 $ gdpPercap: num  779 821 853 836 740 ...
```

Attributes in the data-set

Hide

```
attr(x = df, which = "names")
```

```
[1] "X"         "country"   "continent" "year"      "lifeExp"   "pop"       "gdpPercap"
```

Hide

```
summary(df)
```

```
       X            country           continent              year         lifeExp          pop             gdpP
ercap       
 Min.   :   1.0   Length:1704        Length:1704        Min.   :1952   Min.   :23.60   Min.   :6.001e+04   Min.
   :241.2  
 1st Qu.: 426.8   Class :character   Class :character   1st Qu.:1966   1st Qu.:48.20   1st Qu.:2.794e+06   1st Q
u.:1202.1  
 Median : 852.5   Mode  :character   Mode  :character   Median :1980   Median :60.71   Median :7.024e+06   Median
 :3531.8  
 Mean   : 852.5                                        Mean   :1980   Mean   :59.47   Mean   :2.960e+07   Mean
   :7215.3  
 3rd Qu.:1278.2                                        3rd Qu.:1993   3rd Qu.:70.85   3rd Qu.:1.959e+07   3rd Q
u.:9325.5  
 Max.   :1704.0                                        Max.   :2007   Max.   :82.60   Max.   :1.319e+09   Max.
   :113523.1  
```

Checking for NULL values

Hide

```
lapply(df,function(x) {
  length(which(is.na(x)))
})
```

```
$X
[1] 0

$country
[1] 0

$continent
[1] 0

$year
[1] 0

$lifeExp
[1] 0

$pop
[1] 0

$gdpPercap
[1] 0
```

1. Choose a country and plot the life expectancy by year (both scatter and line plot)

Hide

```
df_india = subset(df,country=="India")
df_india
```

	X <int>	country <chr>	continent <chr>	year <int>	lifeExp <dbl>	pop <dbl>	gdpPercap <dbl>
697	697	India	Asia	1952	37.373	372000000	546.5657
698	698	India	Asia	1957	40.249	409000000	590.0620
699	699	India	Asia	1962	43.605	454000000	658.3472
700	700	India	Asia	1967	47.193	506000000	700.7706
701	701	India	Asia	1972	50.651	567000000	724.0325
702	702	India	Asia	1977	54.208	634000000	813.3373
703	703	India	Asia	1982	56.596	708000000	855.7235
704	704	India	Asia	1987	58.553	788000000	976.5127
705	705	India	Asia	1992	60.223	872000000	1164.4068
706	706	India	Asia	1997	61.765	959000000	1458.8174

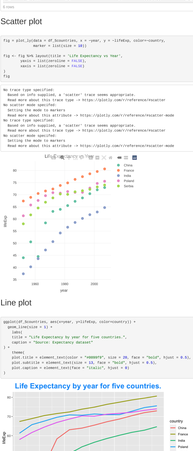
1-10 of 12 rows

Previous 1 2 Next

Scatter Plot

Hide

```
ggplot(df_india, aes(x=year, y=lifeExp)) +
  geom_point(size = 3, color = "#000ED") +
  labs(
    title = "Life Expectancy vs Year",
    subtitle = "LifeExp - Expected number of years remaining for an individual at any given year ",
    caption = "Source: Expectancy dataset"
  ) +
  theme(
    plot.title = element_text(color = "#000ED", size = 20, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 13, face = "bold", hjust = 0.5),
    plot.caption = element_text(face = "italic", hjust = 0)
  )
```



Line Plot

Hide

```
fig = plot_ly(data = df_india, x = ~year, y = ~lifeExp,
              type = 'scatter', mode = 'lines',
              marker = list(size = 20,
              color = 'rgba(255, 182, 193, .9)',
              line = list(color = 'rgba(152, 0, 0, .8)',width = 2)))

fig <- fig %>% layout(title = 'Life Expectancy vs Year',
         xaxis = list(zeroline = FALSE),
         xaxis = list(zeroline = FALSE)
)

fig
```

```
A marker object has been specified, but markers is not in the mode
Adding markers to the mode...
A marker object has been specified, but markers is not in the mode
Adding markers to the mode...
```

2. Plot the life expectancy by year for five countries. (both scatter and line plot)

Hide

```
## Take out the five countries
df_bycountries=df[df$country %in% c('China', 'India', 'France', 'Poland', 'Serbia'), ]
head(df_bycountries)
```

	X <int>	country <chr>	continent <chr>	year <int>	lifeExp <dbl>	pop <dbl>	gdpPercap <dbl>
289	289	China	Asia	1952	44.00000	556263527	400.4486
290	290	China	Asia	1957	50.54896	637408000	575.9870
291	291	China	Asia	1962	44.50136	665770000	487.6740
292	292	China	Asia	1967	58.38112	754550000	612.7057
293	293	China	Asia	1972	63.11888	862030000	676.9001
294	294	China	Asia	1977	63.96736	943000000	741.2375

6 rows

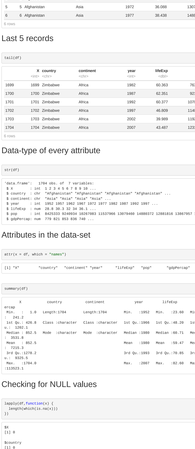
Scatter plot

Hide

```
fig = plot_ly(data = df_bycountries, x = ~year, y = ~lifeExp, color=~country,
              marker = list(size = 15))

fig <- fig %>% layout(title = 'Life Expectancy vs Year',
         xaxis = list(zeroline = FALSE),
         xaxis = list(zeroline = FALSE)
)
fig
```

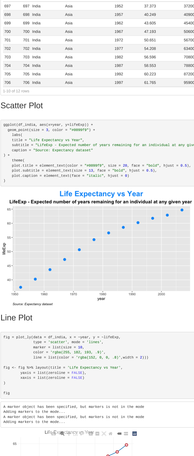
```
No trace type specified:
  Based on info supplied, a 'scatter' trace seems appropriate.
  Read more about this trace type -> https://plotly.com/r/reference/#scatter
No scatter mode specified:
  Setting the mode to markers
  Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode
No trace type specified:
  Based on info supplied, a 'scatter' trace seems appropriate.
  Read more about this trace type -> https://plotly.com/r/reference/#scatter
No scatter mode specified:
  Setting the mode to markers
  Read more about this attribute -> https://plotly.com/r/reference/#scatter-mode
```



Line plot

Hide

```
ggplot(df_bycountries, aes(x=year, y=lifeExp, color=country)) +
  geom_line(size = 2) +
  labs(
    title = "Life Expectancy by year for five countries.",
    caption = "Source: Expectancy dataset"
  ) +
  theme(
    plot.title = element_text(color = "#000ED", size = 20, face = "bold", hjust = 0.5),
    plot.subtitle = element_text(size = 13, face = "bold", hjust = 0.5),
    plot.caption = element_text(face = "italic", hjust = 0)
  )
```



#3. Plot the life expectancy of all the countries in Europe for any particular year.

Hide

```
## Take out European Continent
df_europe2007 = subset(df,continent =="Europe"&year==2007)
head(df_europe2007)
```

	X <int>	country <chr>	continent <chr>	year <int>	lifeExp <dbl>	pop <dbl>	gdpPercap <dbl>
24	24	Albania	Europe	2007	76.423	3600523	5937.030
84	84	Austria	Europe	2007	79.829	8199783	36126.493
120	120	Belgium	Europe	2007	79.441	10392226	33692.605
156	156	Bosnia and Herzegovina	Europe	2007	74.852	4552198	7446.299
192	192	Bulgaria	Europe	2007	73.005	7322858	10680.793
364	364	Croatia	Europe	2007	75.748	4493312	14619.223

6 rows

Bar Plot

Hide

```
fig = plot_ly(data = df_europe2007, x = ~country, y = ~lifeExp,
              type = 'bar', textposition = 'auto')

fig <- fig %>% layout(title = 'Life Expectancy in European Continent in 2007 by year',
         xaxis = list(zeroline = FALSE),
         xaxis = list(zeroline = FALSE)
)
fig
```

```
No trace type specified:
  Based on info supplied, a 'bar' trace seems appropriate.
  Read more about this trace type -> https://plotly.com/r/reference/#bar
No trace type specified:
  Based on info supplied, a 'bar' trace seems appropriate.
  Read more about this trace type -> https://plotly.com/r/reference/#bar
```