

1) Data Cleaning

a. Raw data snapshot

	country	year	sex	age	suicides_no	population	suicides/100k	country-year	HDI for year	gdp_for_year	gdp_per_capita	generation
	<chr>	<dbl>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<chr>
1	Albania	1987	male	15-2~	21	312900	6.71	Albania1987	NA	2156624900	796	Generatio~
2	Albania	1987	male	35-5~	16	308000	5.19	Albania1987	NA	2156624900	796	silent
3	Albania	1987	female	15-2~	14	289700	4.83	Albania1987	NA	2156624900	796	Generatio~
4	Albania	1987	male	75+ ~	1	21800	4.59	Albania1987	NA	2156624900	796	G.I. Gene~
5	Albania	1987	male	25-3~	9	274300	3.28	Albania1987	NA	2156624900	796	Boomers
6	Albania	1987	female	75+ ~	1	35600	2.81	Albania1987	NA	2156624900	796	G.I. Gene~
7	Albania	1987	female	35-5~	6	278800	2.15	Albania1987	NA	2156624900	796	silent
8	Albania	1987	female	25-3~	4	257200	1.56	Albania1987	NA	2156624900	796	Boomers
9	Albania	1987	male	55-7~	1	137500	0.73	Albania1987	NA	2156624900	796	G.I. Gene~
10	Albania	1987	female	5-14~	0	311000	0	Albania1987	NA	2156624900	796	Generatio~
11	Albania	1987	female	55-7~	0	144600	0	Albania1987	NA	2156624900	796	G.I. Gene~
12	Albania	1987	male	5-14~	0	338200	0	Albania1987	NA	2156624900	796	Generatio~

Cleaned data snapshot

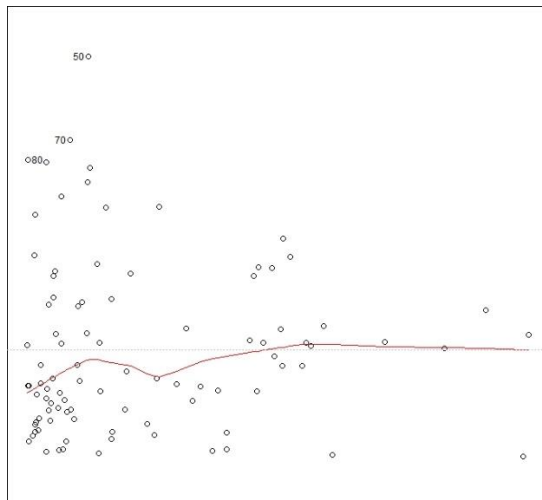
```
> head(data, 12)
```

	country	year	sex	age	suicides_no	population	suicide_100k	gdp_for_year	gdp_per_capita	continent
1	Albania	1987	Male	15-24	21	312900	6.71	2156624900	796	Europe
2	Albania	1987	Male	35-54	16	308000	5.19	2156624900	796	Europe
3	Albania	1987	Female	15-24	14	289700	4.83	2156624900	796	Europe
4	Albania	1987	Male	75+	1	21800	4.59	2156624900	796	Europe
5	Albania	1987	Male	25-34	9	274300	3.28	2156624900	796	Europe
6	Albania	1987	Female	75+	1	35600	2.81	2156624900	796	Europe
7	Albania	1987	Female	35-54	6	278800	2.15	2156624900	796	Europe
8	Albania	1987	Female	25-34	4	257200	1.56	2156624900	796	Europe
9	Albania	1987	Male	55-74	1	137500	0.73	2156624900	796	Europe
10	Albania	1987	Female	5-14	0	311000	0.00	2156624900	796	Europe
11	Albania	1987	Female	55-74	0	144600	0.00	2156624900	796	Europe
12	Albania	1987	Male	5-14	0	338200	0.00	2156624900	796	Europe

- Removed countries with <3 years of data
- Removed year 2016
- Removed certain columns
- Cleaned data value
- Get respective continent from countrycode package
- ...

2) Statistics for linear model

a. Outliers Removal



Homoscedasticity likely to be violated.

b. Influential Points Removal (Cook's Distance $< 4/n$)

c. Linear regression output

```
> # linear regression model
> summary(model2)

Call:
lm(formula = suicides_per_100k ~ gdp_per_capita, data = g10_plotly)

Residuals:
    Min       1Q   Median       3Q      Max
-11.706  -4.880  -1.436   3.169  18.978

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.436e+00  1.080e+00   7.809 1.37e-11 ***
gdp_per_capita 1.180e-04  4.818e-05   2.448  0.0164  *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.036 on 85 degrees of freedom
Multiple R-squared:  0.06588,    Adjusted R-squared:  0.05489
F-statistic: 5.995 on 1 and 85 DF,  p-value: 0.0164
```

d. Normality assumption violated

```
Shapiro-wilk normality test

data:  model2$residuals
W = 0.93409, p-value = 0.0002554
```