

#Assignment 01 - FML ## By E.R.M.S.C.Ekanayake

#1. download data set from the web #source: <https://www.kaggle.com/datasets/mirichoi0218/insurance>
 (https://www.kaggle.com/datasets/mirichoi0218/insurance) #make sure it is saved as a CSV file

#2.Import the dataset into R

```
setwd("F:/1st sem/ML/Assignment 01/New folder/2")
Data <- read.csv("insurance.csv",header= TRUE,sep=",")
Data
```

##	age	sex	bmi	children	smoker	region	charges
## 1	19	female	27.900	0	yes	southwest	16884.924
## 2	18	male	33.770	1	no	southeast	1725.552
## 3	28	male	33.000	3	no	southeast	4449.462
## 4	33	male	22.705	0	no	northwest	21984.471
## 5	32	male	28.880	0	no	northwest	3866.855
## 6	31	female	25.740	0	no	southeast	3756.622
## 7	46	female	33.440	1	no	southeast	8240.590
## 8	37	female	27.740	3	no	northwest	7281.506
## 9	37	male	29.830	2	no	northeast	6406.411
## 10	60	female	25.840	0	no	northwest	28923.137
## 11	25	male	26.220	0	no	northeast	2721.321
## 12	62	female	26.290	0	yes	southeast	27808.725
## 13	23	male	34.400	0	no	southwest	1826.843
## 14	56	female	39.820	0	no	southeast	11090.718
## 15	27	male	42.130	0	yes	southeast	39611.758
## 16	19	male	24.600	1	no	southwest	1837.237
## 17	52	female	30.780	1	no	northeast	10797.336
## 18	23	male	23.845	0	no	northeast	2395.172
## 19	56	male	40.300	0	no	southwest	10602.385
## 20	30	male	35.300	0	yes	southwest	36837.467
## 21	60	female	36.005	0	no	northeast	13228.847
## 22	30	female	32.400	1	no	southwest	4149.736
## 23	18	male	34.100	0	no	southeast	1137.011
## 24	34	female	31.920	1	yes	northeast	37701.877
## 25	37	male	28.025	2	no	northwest	6203.902
## 26	59	female	27.720	3	no	southeast	14001.134
## 27	63	female	23.085	0	no	northeast	14451.835
## 28	55	female	32.775	2	no	northwest	12268.632
## 29	23	male	17.385	1	no	northwest	2775.192
## 30	31	male	36.300	2	yes	southwest	38711.000
## 31	22	male	35.600	0	yes	southwest	35585.576
## 32	18	female	26.315	0	no	northeast	2198.190
## 33	19	female	28.600	5	no	southwest	4687.797
## 34	63	male	28.310	0	no	northwest	13770.098
## 35	28	male	36.400	1	yes	southwest	51194.559
## 36	19	male	20.425	0	no	northwest	1625.434
## 37	62	female	32.965	3	no	northwest	15612.193
## 38	26	male	20.800	0	no	southwest	2302.300
## 39	35	male	36.670	1	yes	northeast	39774.276
## 40	60	male	39.900	0	yes	southwest	48173.361

#3. Print out descriptive statistics for a selection of quantitative and categorical variables

```
table(Data$sex)
```

```
##
## female    male
##      17      23
```

```
summary(Data$sex)
```

```
##      Length      Class      Mode
##           40 character character
```

```
table(Data$age)
```

```
##
## 18 19 22 23 25 26 27 28 30 31 32 33 34 35 37 46 52 55 56 59 60 62 63
##   3  4  1  3  1  1  1  2  2  2  1  1  1  1  3  1  1  1  2  1  3  2  2
```

```
summary(Data$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    18.00   23.00   31.50   36.90   55.25   63.00
```

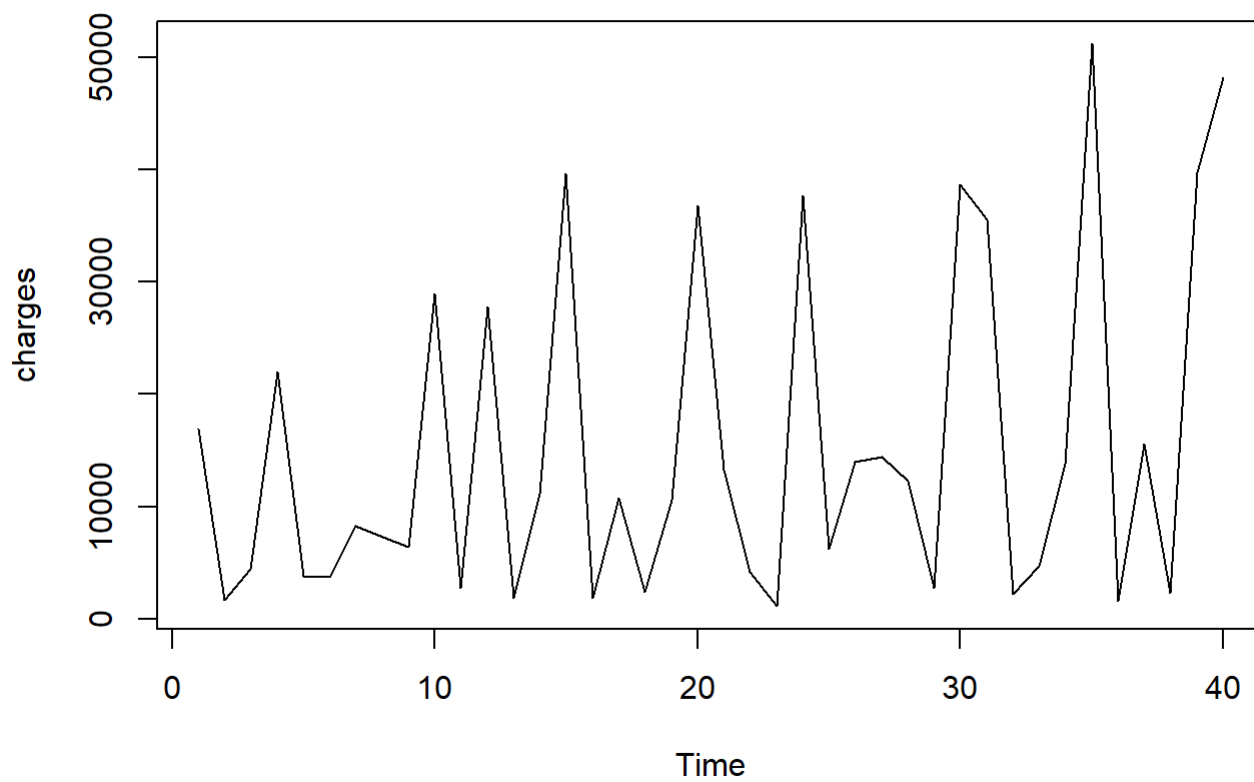
#4. Transform at least one variable. It doesn't matter what the transformation is

```
Data$Dummy_sex <- ifelse(Data$sex == "male",1,0)
Data
```

##	age	sex	bmi	children	smoker	region	charges	Dummy_sex
## 1	19	female	27.900	0	yes	southwest	16884.924	0
## 2	18	male	33.770	1	no	southeast	1725.552	1
## 3	28	male	33.000	3	no	southeast	4449.462	1
## 4	33	male	22.705	0	no	northwest	21984.471	1
## 5	32	male	28.880	0	no	northwest	3866.855	1
## 6	31	female	25.740	0	no	southeast	3756.622	0
## 7	46	female	33.440	1	no	southeast	8240.590	0
## 8	37	female	27.740	3	no	northwest	7281.506	0
## 9	37	male	29.830	2	no	northeast	6406.411	1
## 10	60	female	25.840	0	no	northwest	28923.137	0
## 11	25	male	26.220	0	no	northeast	2721.321	1
## 12	62	female	26.290	0	yes	southeast	27808.725	0
## 13	23	male	34.400	0	no	southwest	1826.843	1
## 14	56	female	39.820	0	no	southeast	11090.718	0
## 15	27	male	42.130	0	yes	southeast	39611.758	1
## 16	19	male	24.600	1	no	southwest	1837.237	1
## 17	52	female	30.780	1	no	northeast	10797.336	0
## 18	23	male	23.845	0	no	northeast	2395.172	1
## 19	56	male	40.300	0	no	southwest	10602.385	1
## 20	30	male	35.300	0	yes	southwest	36837.467	1
## 21	60	female	36.005	0	no	northeast	13228.847	0
## 22	30	female	32.400	1	no	southwest	4149.736	0
## 23	18	male	34.100	0	no	southeast	1137.011	1
## 24	34	female	31.920	1	yes	northeast	37701.877	0
## 25	37	male	28.025	2	no	northwest	6203.902	1
## 26	59	female	27.720	3	no	southeast	14001.134	0
## 27	63	female	23.085	0	no	northeast	14451.835	0
## 28	55	female	32.775	2	no	northwest	12268.632	0
## 29	23	male	17.385	1	no	northwest	2775.192	1
## 30	31	male	36.300	2	yes	southwest	38711.000	1
## 31	22	male	35.600	0	yes	southwest	35585.576	1
## 32	18	female	26.315	0	no	northeast	2198.190	0
## 33	19	female	28.600	5	no	southwest	4687.797	0
## 34	63	male	28.310	0	no	northwest	13770.098	1
## 35	28	male	36.400	1	yes	southwest	51194.559	1
## 36	19	male	20.425	0	no	northwest	1625.434	1
## 37	62	female	32.965	3	no	northwest	15612.193	0
## 38	26	male	20.800	0	no	southwest	2302.300	1
## 39	35	male	36.670	1	yes	northeast	39774.276	1
## 40	60	male	39.900	0	yes	southwest	48173.361	1

#5. Plot at least one quantitative variable, and one scatterplot

```
attach(Data)
plot.ts(charges)
library(ggplot2)
```



```
ggplot(Data,aes(charges,age))+  
  geom_point(color="red")
```

