ISLR_Advertising

RohitDhankar

28/05/2020

Source of the code and the data is the Book = http://faculty.marshall.usc.edu/gareth-james/ISL/

```
library(readr)
library(ggplot2)
ad_data <- read_csv("Advertising.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   TV = col_double(),
##   Radio = col_double(),
##   Newspaper = col_double(),
##   Sales = col_double()
## )
```

```
attach(ad_data);View(ad_data);names(ad_data);head(ad_data)
```

```
## [1] "X1"        "TV"        "Radio"     "Newspaper" "Sales"
```

```
## # A tibble: 6 x 5
##      X1    TV Radio Newspaper Sales
##   <dbl> <dbl> <dbl>     <dbl> <dbl>
## 1     1 230.   37.8      69.2  22.1
## 2     2  44.5  39.3      45.1  10.4
## 3     3  17.2  45.9      69.3   9.3
## 4     4 152.   41.3      58.5  18.5
## 5     5 181.   10.8      58.4  12.9
## 6     6   8.7  48.9      75     7.2
```

```
# TV , Radio , Newspapoer Budgets are predictors, independent variables, or just - features.
```

```
simple_lm_fit <- lm(Sales ~ TV )
init_multiple_linear_m <- lm(Sales ~ TV + Radio + Newspaper)
typeof(init_multiple_linear_m) # list
```

```
## [1] "list"
```

```
class(init_multiple_linear_m) # lm - Linear Model
```

```
## [1] "lm"
```

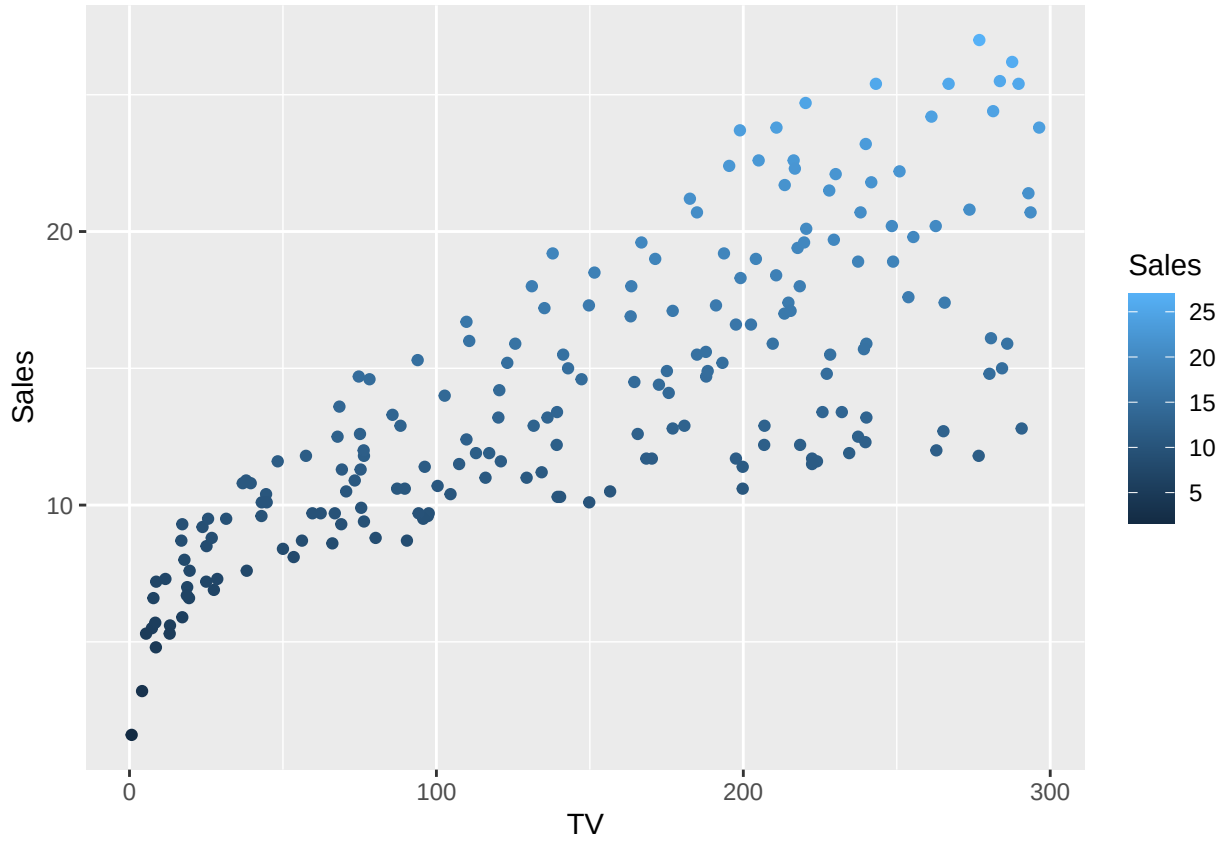
```
summary(init_multiple_linear_m)
```

```
##
## Call:
```

```
## lm(formula = Sales ~ TV + Radio + Newspaper)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8277 -0.8908  0.2418  1.1893  2.8292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.938889   0.311908   9.422   <2e-16 ***
## TV           0.045765   0.001395  32.809   <2e-16 ***
## Radio        0.188530   0.008611  21.893   <2e-16 ***
## Newspaper   -0.001037   0.005871  -0.177     0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.686 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

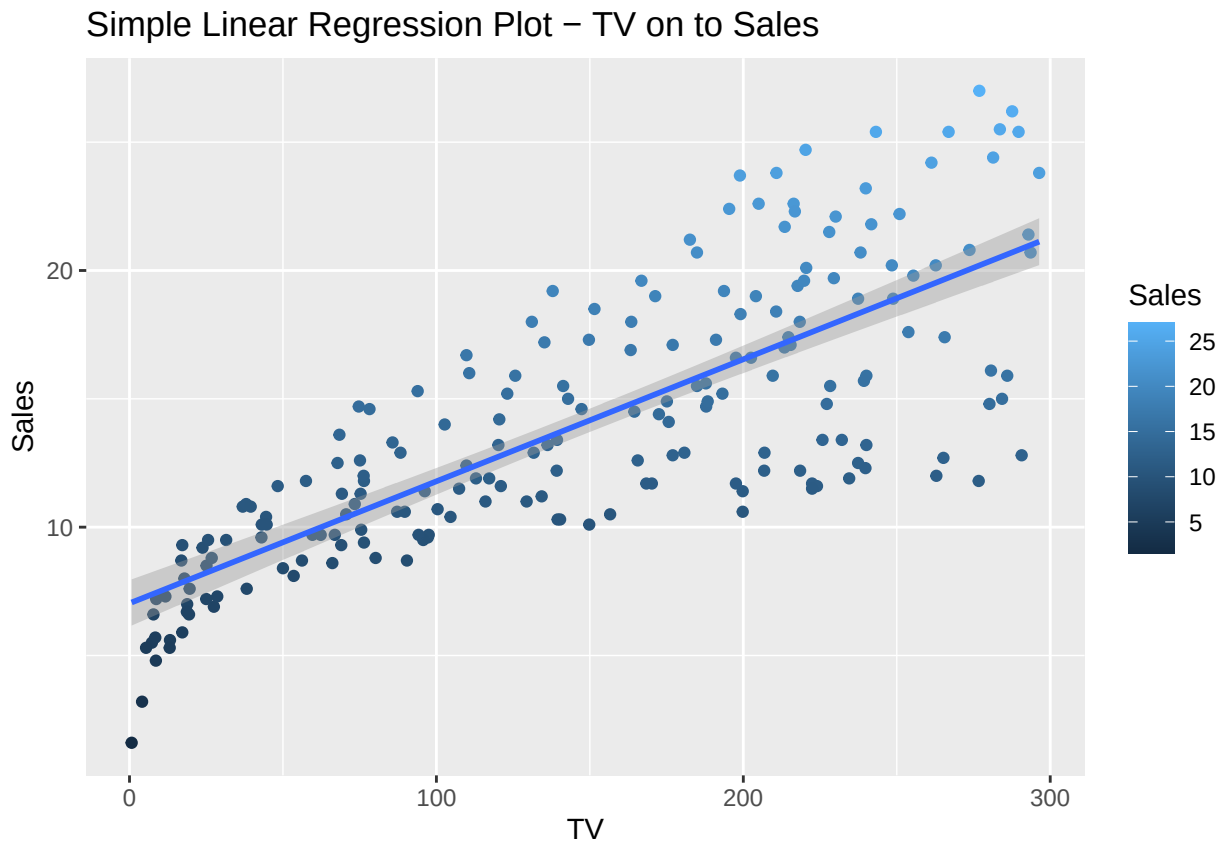
```
# the - Adjusted R-squared:  0.8956 -- 89.56% Variability in Sales can be
# explained by the linear relationship between - TV + Radio regressed on to Sales
```

```
# Seen below a Scatter plot of Sales -
# its Similar to the - lm_plot_sales_tv , besides the Regression Line
ggplot(ad_data,aes(y=Sales,x=TV , color = Sales))+geom_point()
```



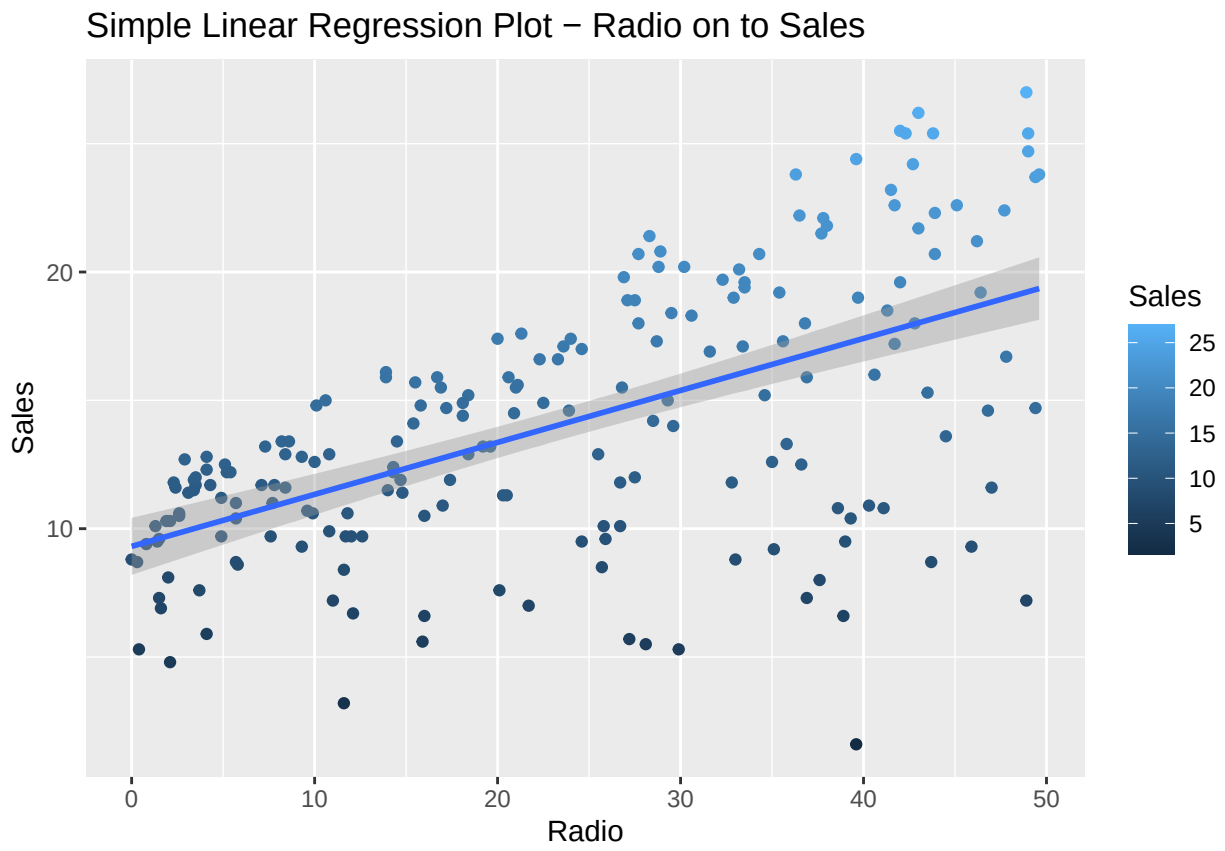
```
#
lm_plot_sales_tv <- ggplot(ad_data, aes(TV, Sales , color = Sales)) + geom_point() + geom_smooth(method
print(lm_plot_sales_tv + ggtitle("Simple Linear Regression Plot - TV on to Sales "))
```

```
## `geom_smooth()` using formula 'y ~ x'
```



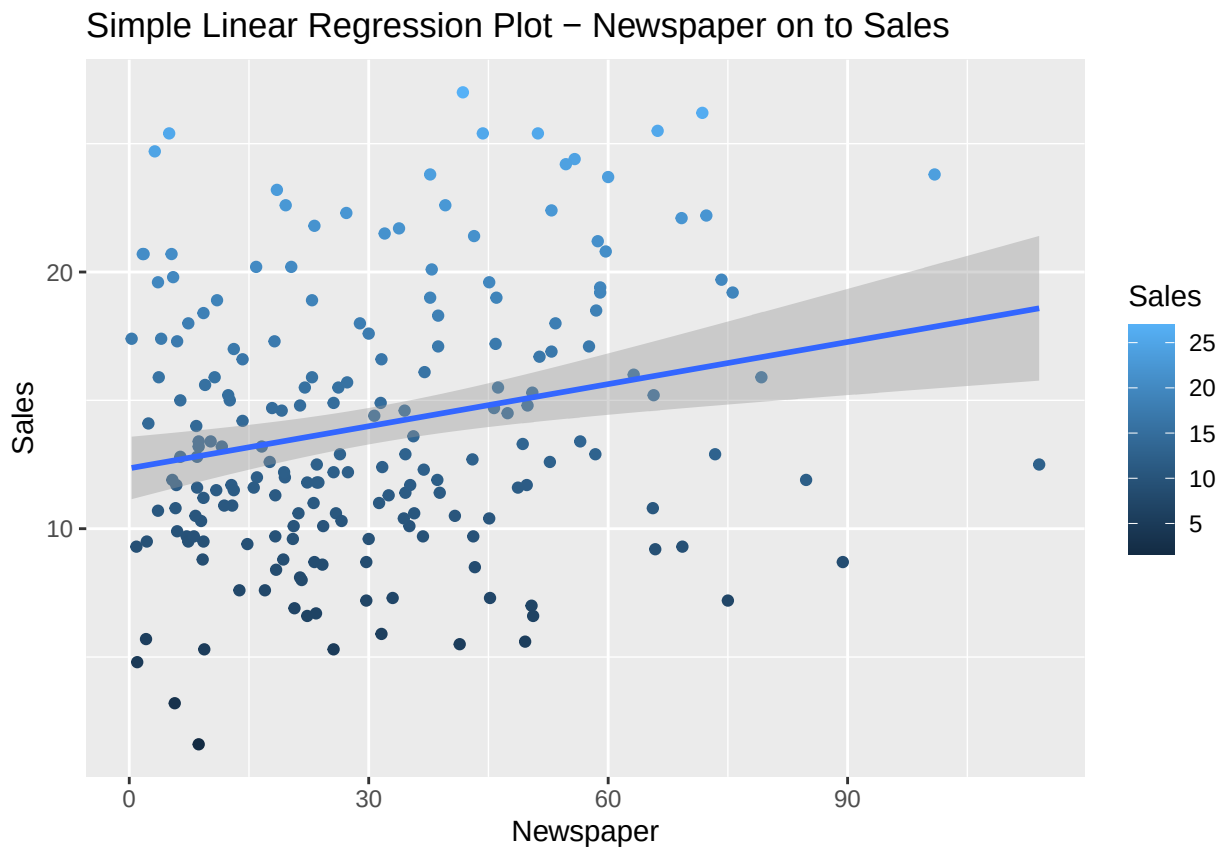

```
# up here -- color = Sales -- is just a notional usage , as we would rather use it with a factor variab
#
lm_plot_sales_radio <- ggplot(ad_data, aes(Radio, Sales , color = Sales)) + geom_point() + geom_smooth(
print(lm_plot_sales_radio + ggtitle("Simple Linear Regression Plot - Radio on to Sales "))
```

```
## `geom_smooth()` using formula 'y ~ x'
```



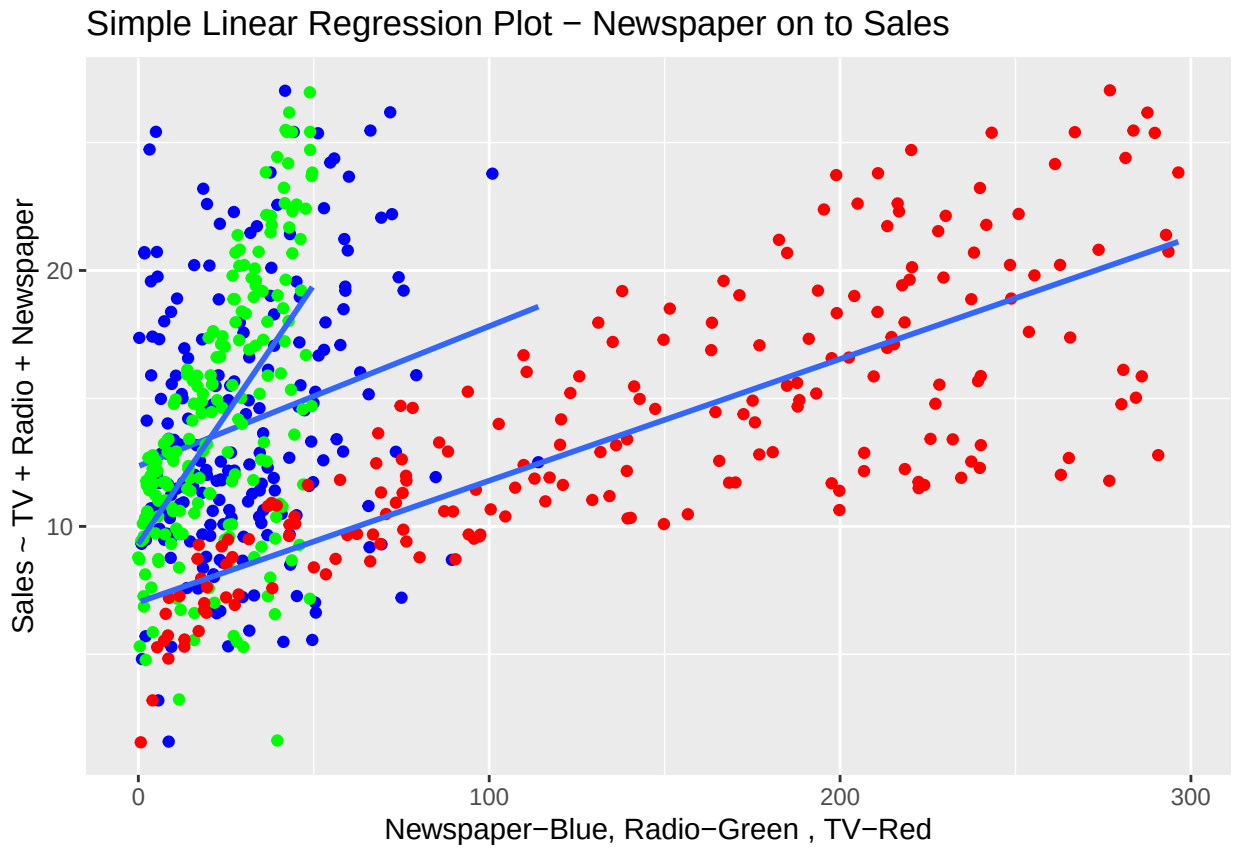

```
#
lm_plot_sales_newsp <- ggplot(ad_data, aes(Newspaper, Sales , color = Sales)) + geom_point() + geom_smoo
print(lm_plot_sales_newsp + ggtitle("Simple Linear Regression Plot - Newspaper on to Sales "))
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
#
mlr_plot_sales_all <- ggplot(ad_data) +
  geom_jitter(aes(Newspaper,Sales), colour="blue") + geom_smooth(aes(Newspaper,Sales), method=lm, se=FA
  geom_jitter(aes(Radio,Sales), colour="green") + geom_smooth(aes(Radio,Sales), method=lm, se=FALSE) +
  geom_jitter(aes(TV,Sales), colour="red") + geom_smooth(aes(TV,Sales), method=lm, se=FALSE) +
  labs(x = "Newspaper-Blue, Radio-Green , TV-Red", y = "Sales ~ TV + Radio + Newspaper")
print(mlr_plot_sales_all + ggtitle("Simple Linear Regression Plot - Newspaper on to Sales "))
```

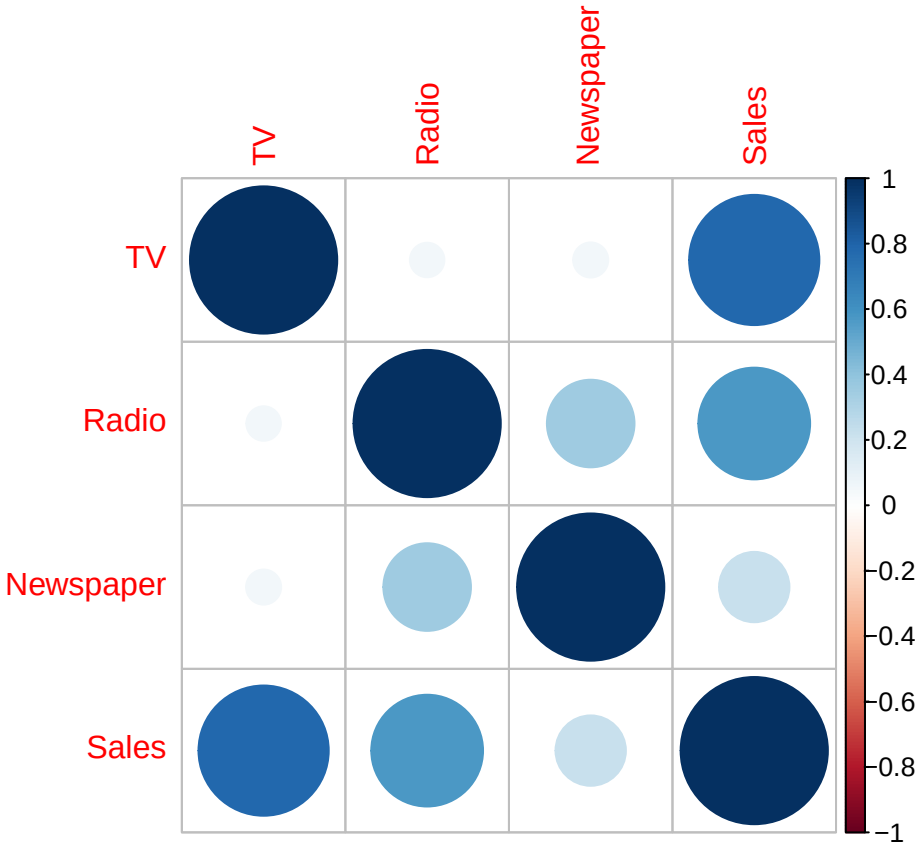
```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



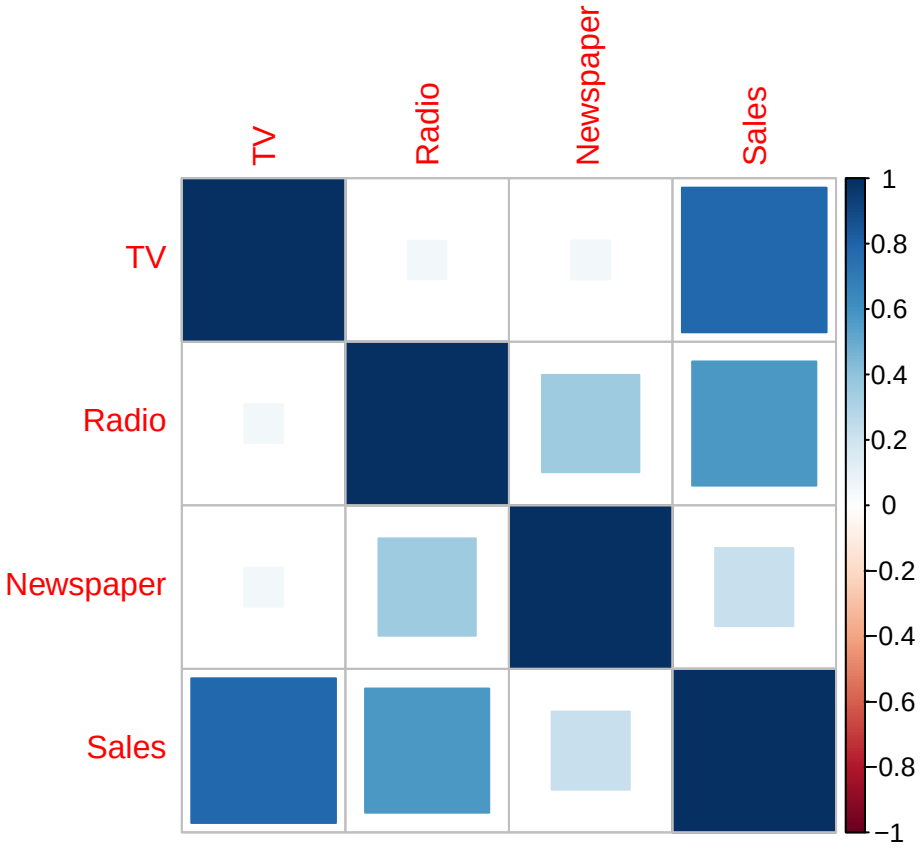

```
#
cor_df <- subset(ad_data, select = c(TV,Radio,Newspaper,Sales))
corr_df_sales <- cor(cor_df, method = c("pearson", "kendall", "spearman"))
#
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corrplot(corr_df_sales, method = "circle")
```




```
#
corrplot(corr_df_sales, method = "square")
```


TV
Radio
Newspaper
Sales
TV
Radio
Newspaper
Sales
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1

```
#
```