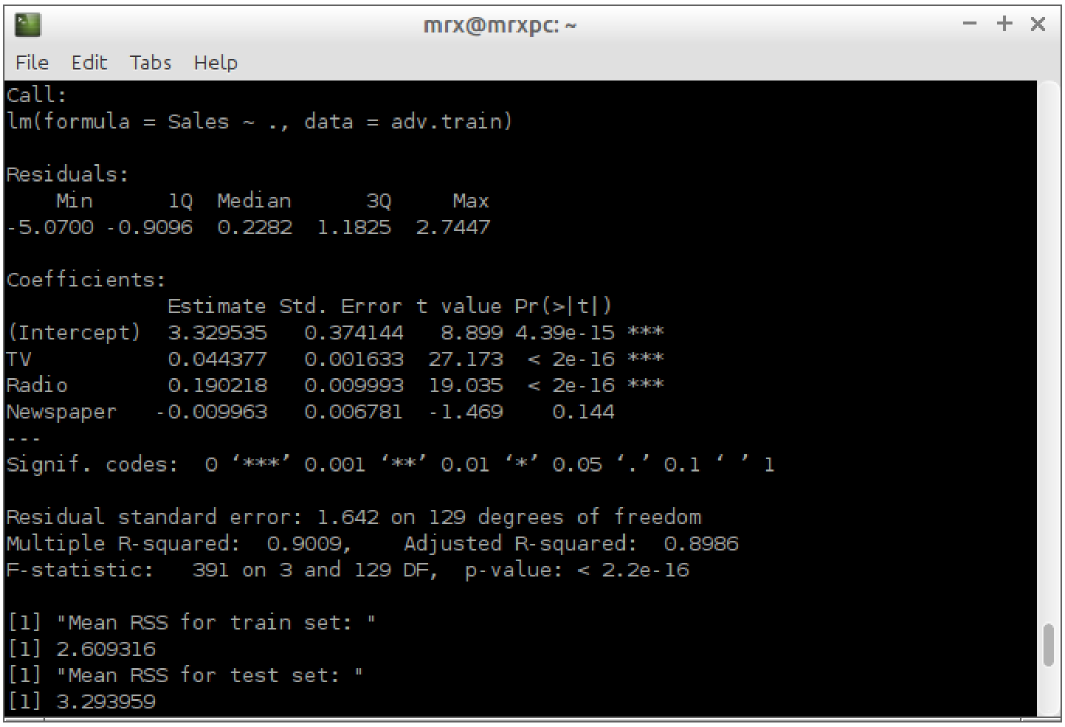
Results analysis for homework 1
Marat Khabibullin

Let's look at the program output shown in Figure 1. One can see (according to p-values) there is a depedency between Sales volume and predictors under consideration and Newspaper predictor is the least significant (because of the high p-value).
Considering mean RSS, we have higher value for the test set and it could be explained by the fact that the model has been created using training data and best fits for this data set.

Scatterplots are shown in Figure 2 and 3 and one can notice clear linear dependance.



```
mrx@mrxpc: ~
File  Edit  Tabs  Help
Call:
lm(formula = Sales ~ ., data = adv.train)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.0700 -0.9096  0.2282  1.1825  2.7447 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.329535   0.374144   8.899 4.39e-15 ***
TV           0.044377   0.001633  27.173  < 2e-16 ***
Radio        0.190218   0.009993  19.035  < 2e-16 ***
Newspaper   -0.009963   0.006781  -1.469    0.144    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.642 on 129 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.8986 
F-statistic:   391 on 3 and 129 DF,  p-value: < 2.2e-16

[1] "Mean RSS for train set: "
[1] 2.609316
[1] "Mean RSS for test set: "
[1] 3.293959
```

Figure 1: Model with all predictors

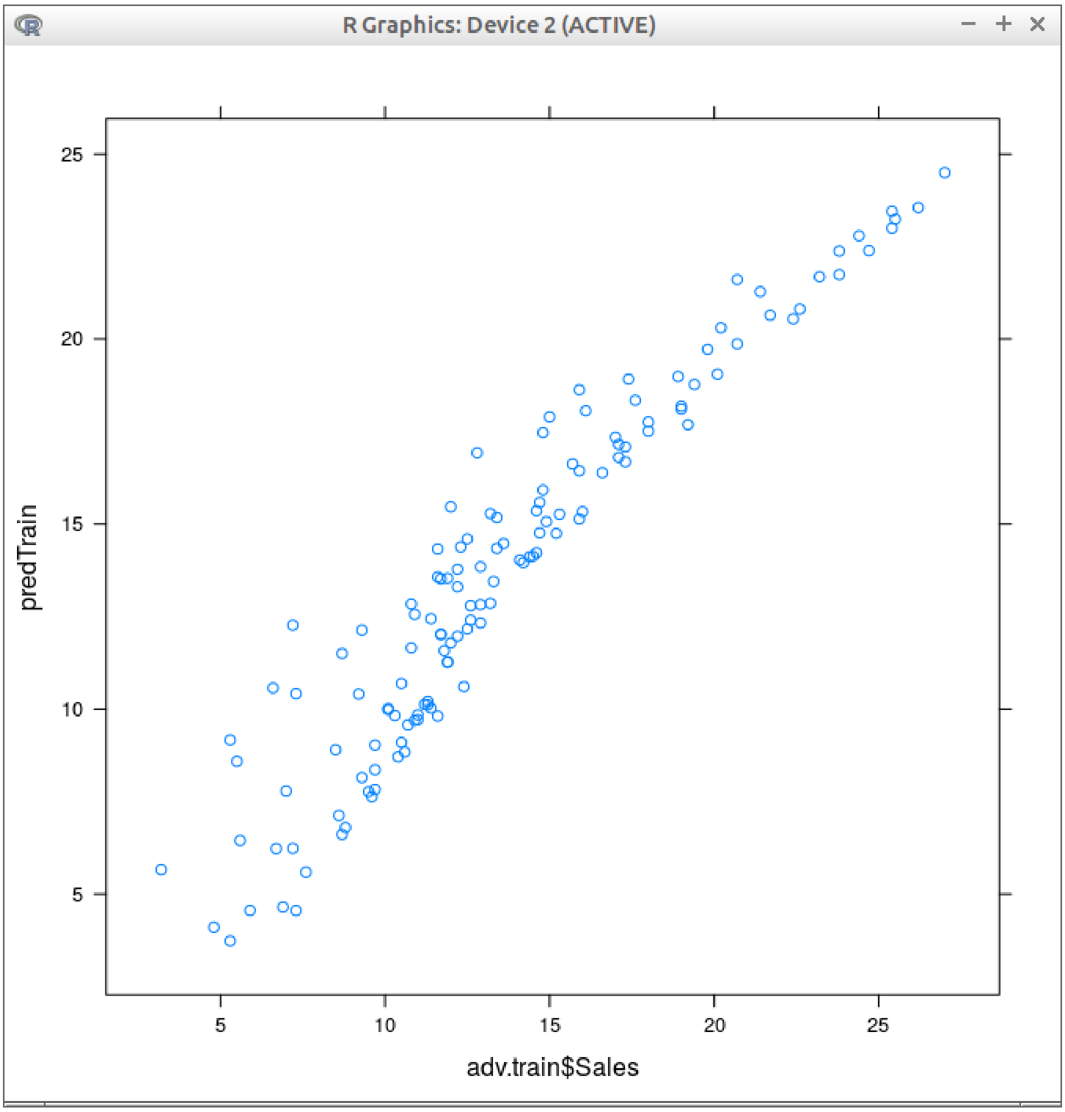


Figure 2: Scatterplot for training set (all predictors)

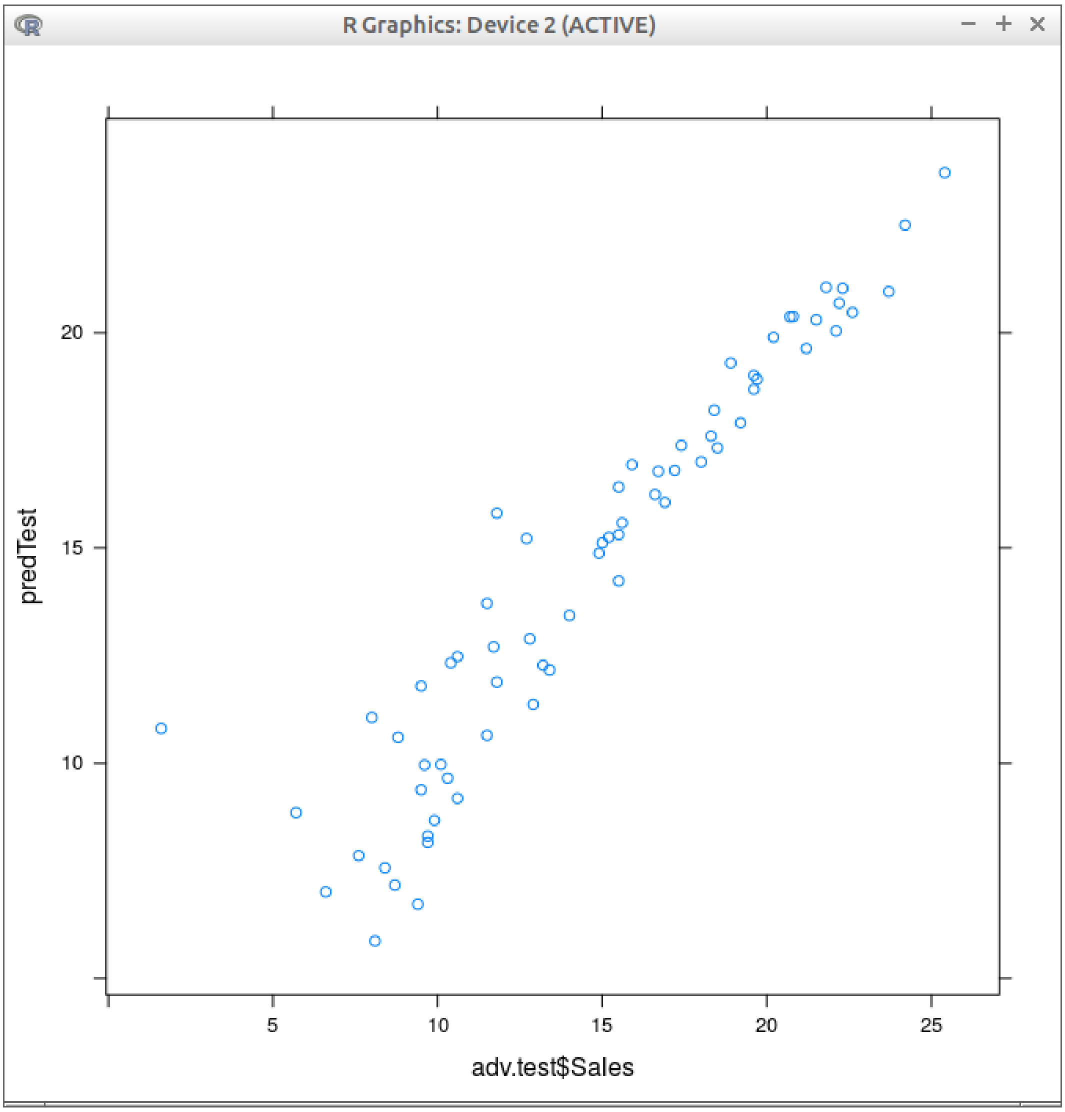


Figure 3: Scatterplot for test set (all predictors)

Now let's consider program output for model with Newspapers predictor removed (Fig. 4). One may notice in comparison with the previous reults (see Fig. 1) F-statistics value has grown significantly, so we can conclude model has become better (better discribes real data).
Moreover, rss values are almost the same as in the model with all predictors. It proves the Nespaper predictor's insignificance.
Scatterplots in Figures 5 and 6 still show linear behaviour.

```
mrx@mrxpc: ~
File  Edit  Tabs  Help

Call:
lm(formula = Sales ~ TV + Radio, data = adv.train)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.3423 -0.9770  0.2671  1.2716  2.7108 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.103137   0.342460   9.061 1.68e-15 ***
TV          0.044558   0.001636  27.240  < 2e-16 ***
Radio       0.185103   0.009409  19.674  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.65 on 130 degrees of freedom
Multiple R-squared:  0.8993,	Adjusted R-squared:  0.8977 
F-statistic: 580.2 on 2 and 130 DF,  p-value: < 2.2e-16

[1] "Mean RSS for train set without Newspaper: "
[1] 2.652986
[1] "Mean RSS for test set without Newspaper: "
[1] 3.097485
```

Figure 4: Model with Newspapers predictor removed

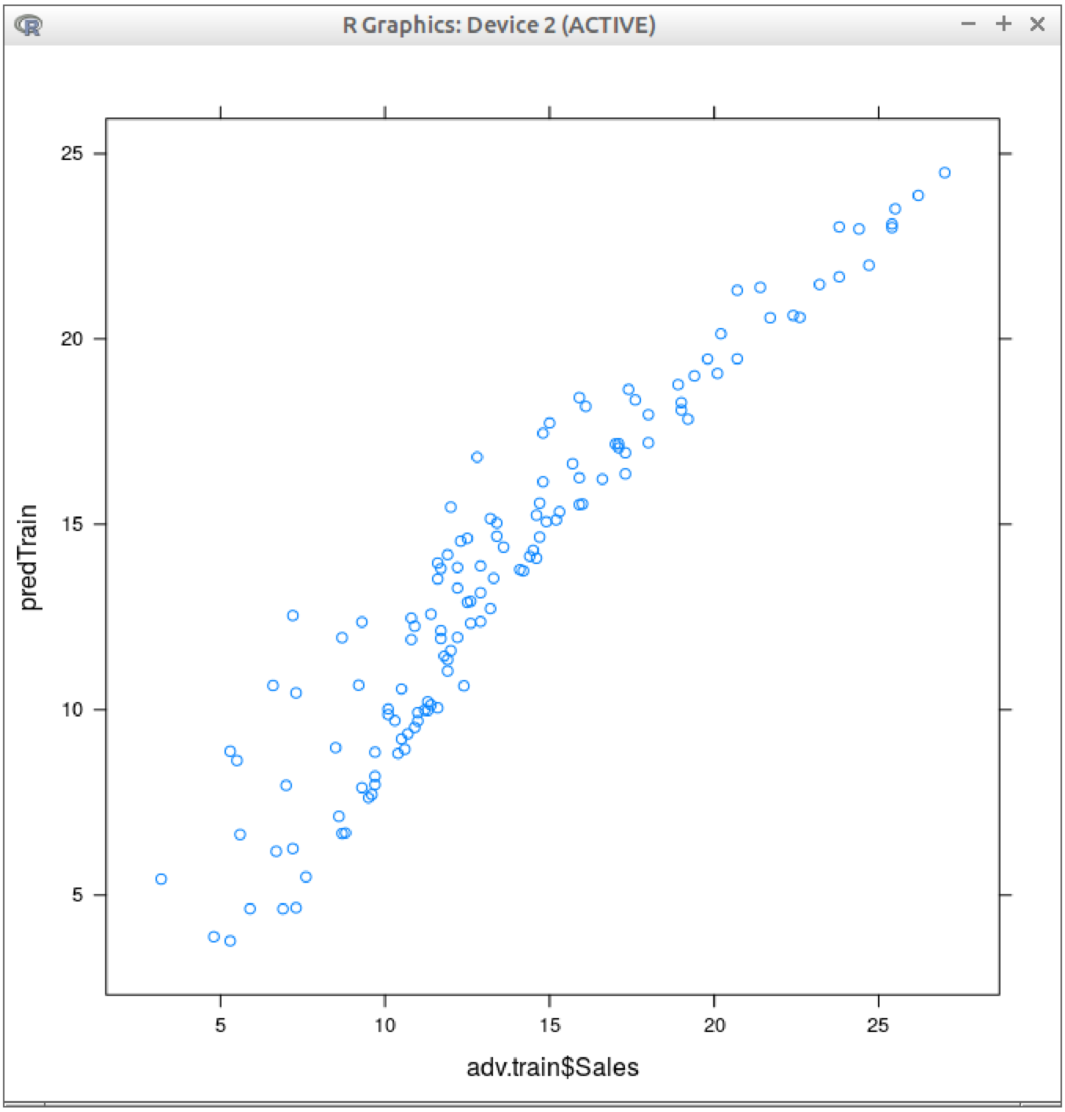


Figure 5: Scatterplot for training set (Newspapers predictor removed)

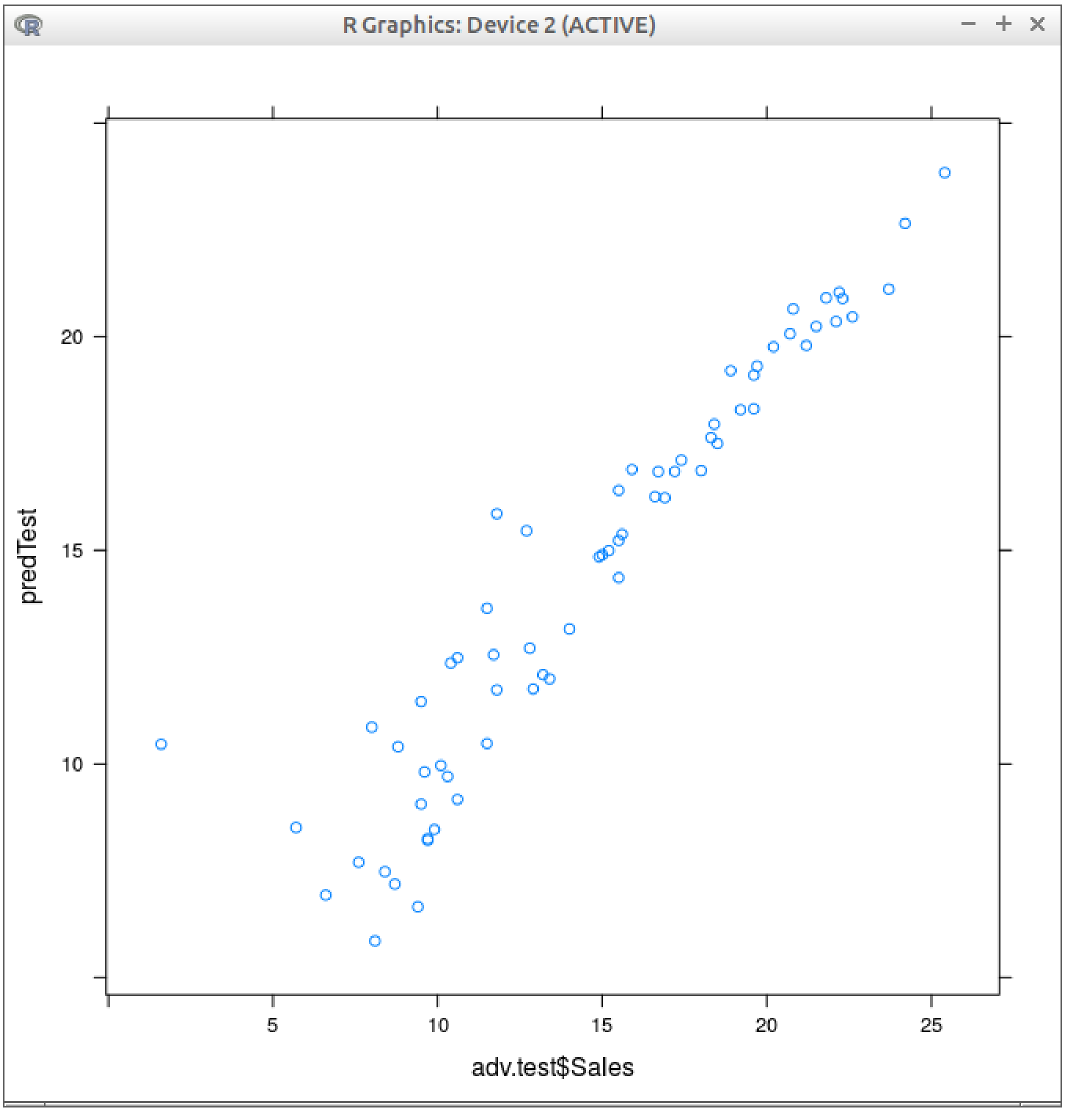


Figure 6: Scatterplot for test set (Newspapers predictor removed)

Next program output (Fig. 7) shows the model with TV predictor removed. F-statistic value has become very low that is corresponds to the fact the TV predictor is significant in our model in general. Moreover, removing significant predictor we have increased rss values for both training and test data sets. Scatterplots in Figures 8 and 9 also reflect the fact of removing significant predictor - the dependencies are not linear anymore, predicted data badly corresspponts to actual one.

```
mrx@mrxpc: ~
File  Edit  Tabs  Help

Call:
lm(formula = Sales ~ Radio + Newspaper, data = adv.train)

Residuals:
     Min       1Q   Median       3Q      Max
-10.9301  -2.3101   0.5021   2.8412   9.2003

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.91755    0.73606  13.474  < 2e-16 ***
Radio        0.20453    0.02578   7.935 8.53e-13 ***
Newspaper   -0.02385    0.01746  -1.366    0.174
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.242 on 130 degrees of freedom
Multiple R-squared:  0.3338,    Adjusted R-squared:  0.3235
F-statistic: 32.56 on 2 and 130 DF,  p-value: 3.437e-12

[1] "Mean RSS for train set without TV: "
[1] 17.5443
[1] "Mean RSS for test set without TV: "
[1] 20.37341
```

Figure 7: Model with Tv predictor removed

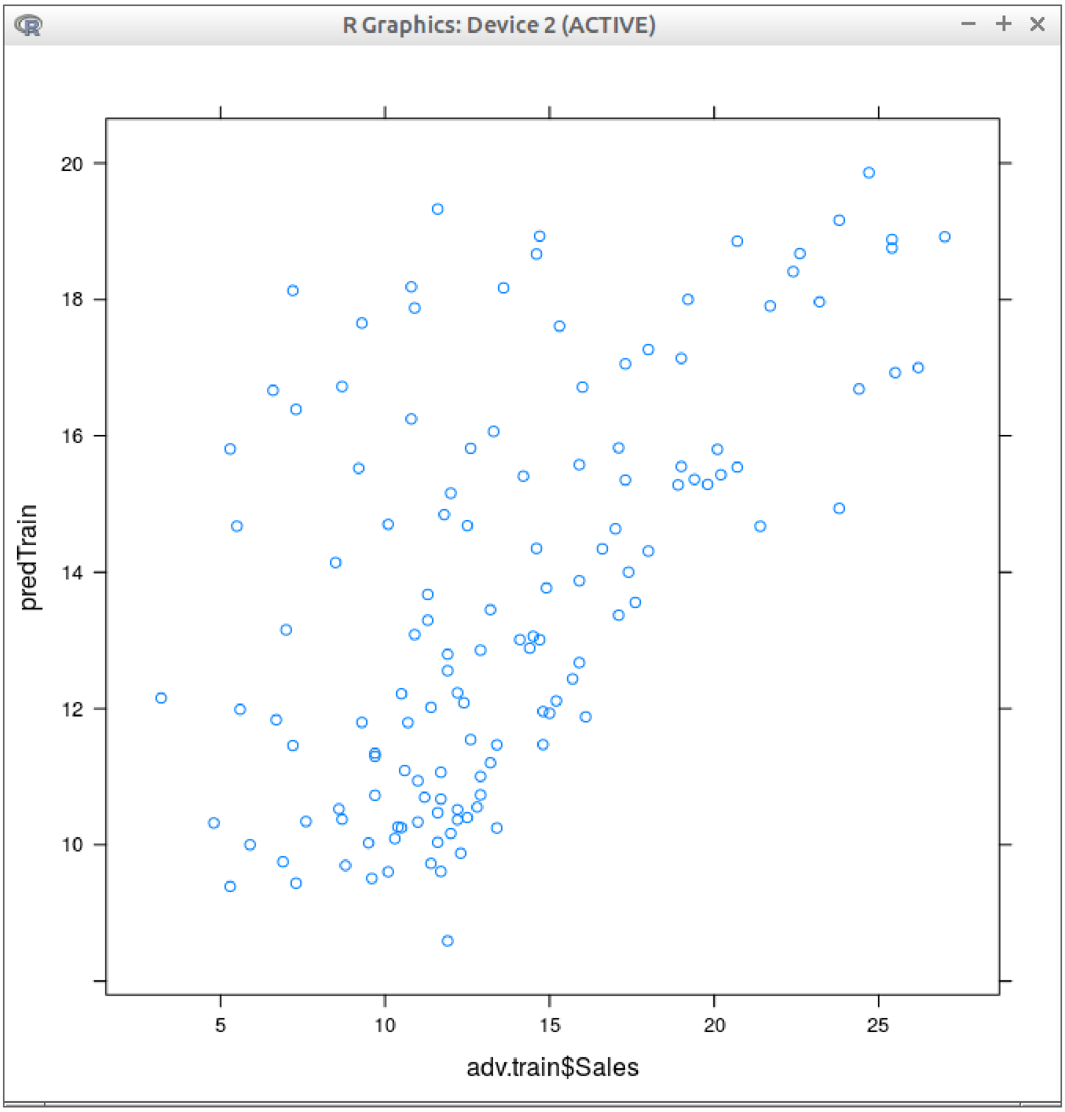


Figure 8: Scatterplot for training set (Tv predictor removed)

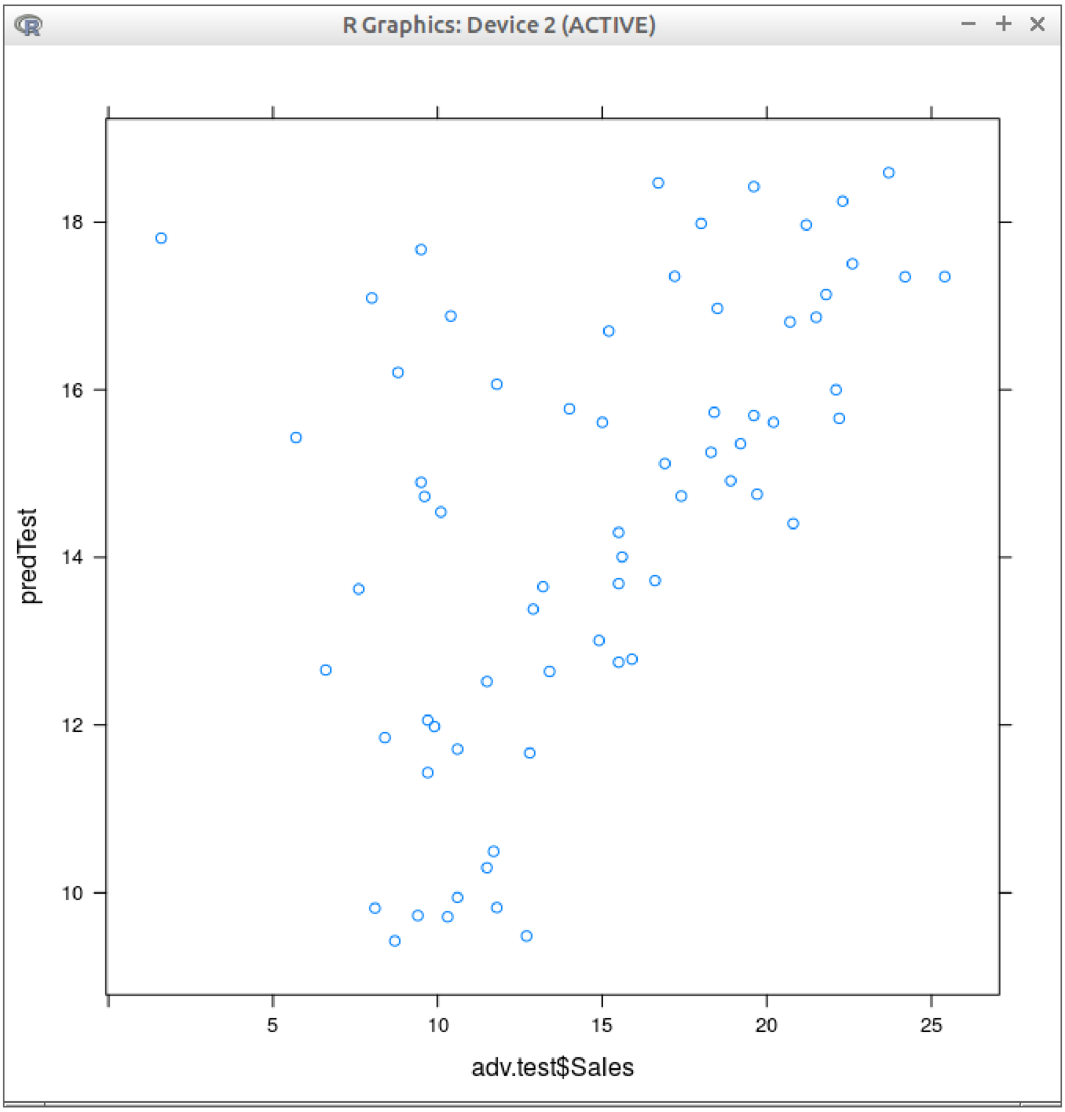


Figure 9: Scatterplot for test set (Tv predictor removed)

The last model to consider is one with all predictor removed (Fig. 10). As we can see rss values have drastically increased in comparison with the initial model values. Also, scatterplots in Figures 11 and 12 show predicted Sales volumes are constant and equals to calculated Intercept value (see Fig. 10). It clearly represents independency of model predictions and input data making such model useless.

```
mrx@mrxpc: ~
File  Edit  Tabs  Help

Call:
lm(formula = Sales ~ 1, data = adv.train)

Residuals:
    Min      1Q  Median      3Q     Max
-10.466  -3.166  -1.166   2.934  13.334

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.6662     0.4472   30.56   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.157 on 132 degrees of freedom

[1] "Mean RSS for train set with intercept only: "
[1] 26.33308
[1] "Mean RSS for test set with intercept only: "
[1] 28.97199
>
```

Figure 10: Model with all predictors removed

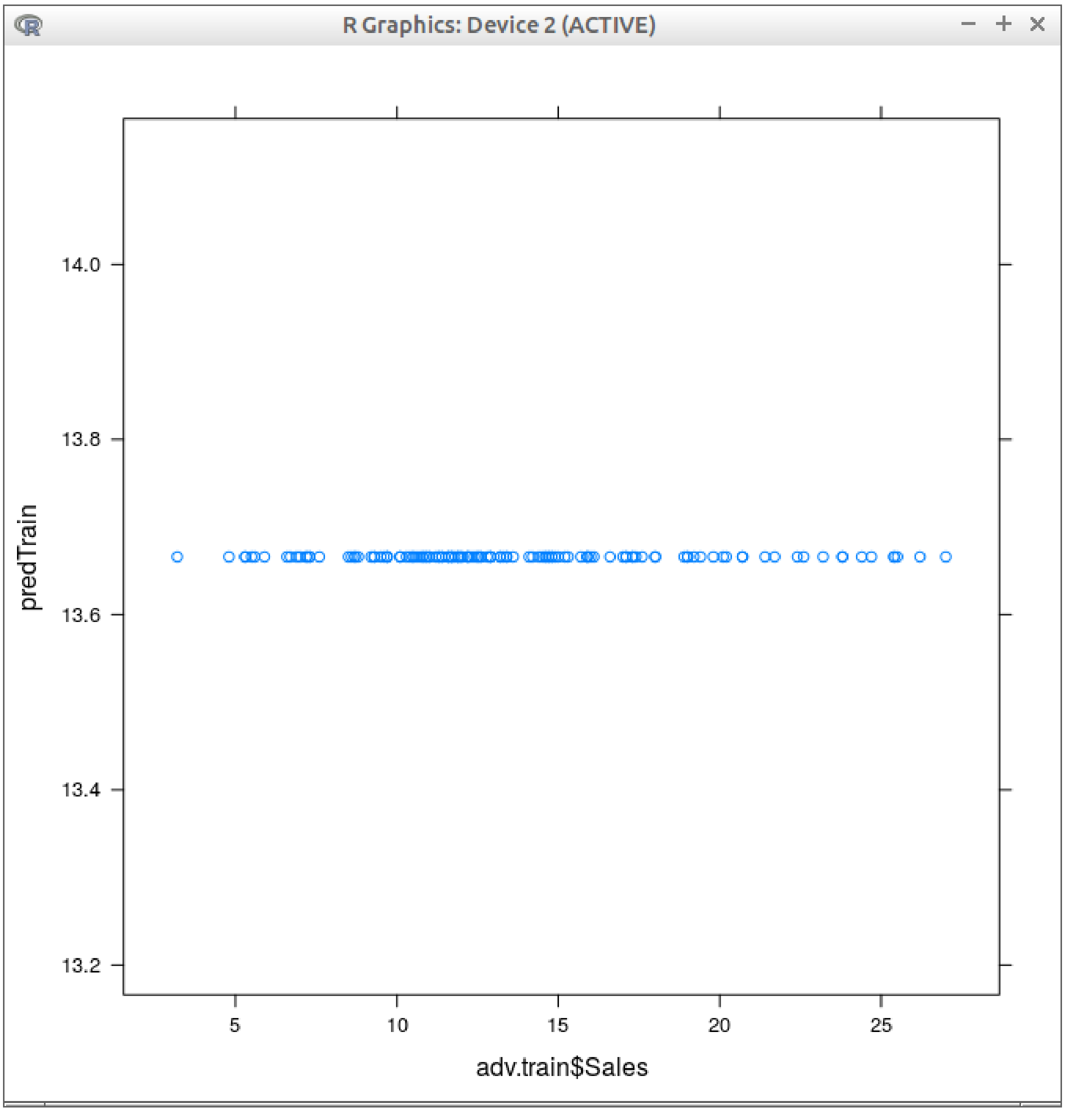


Figure 11: Scatterplot for training set (all predictors removed)

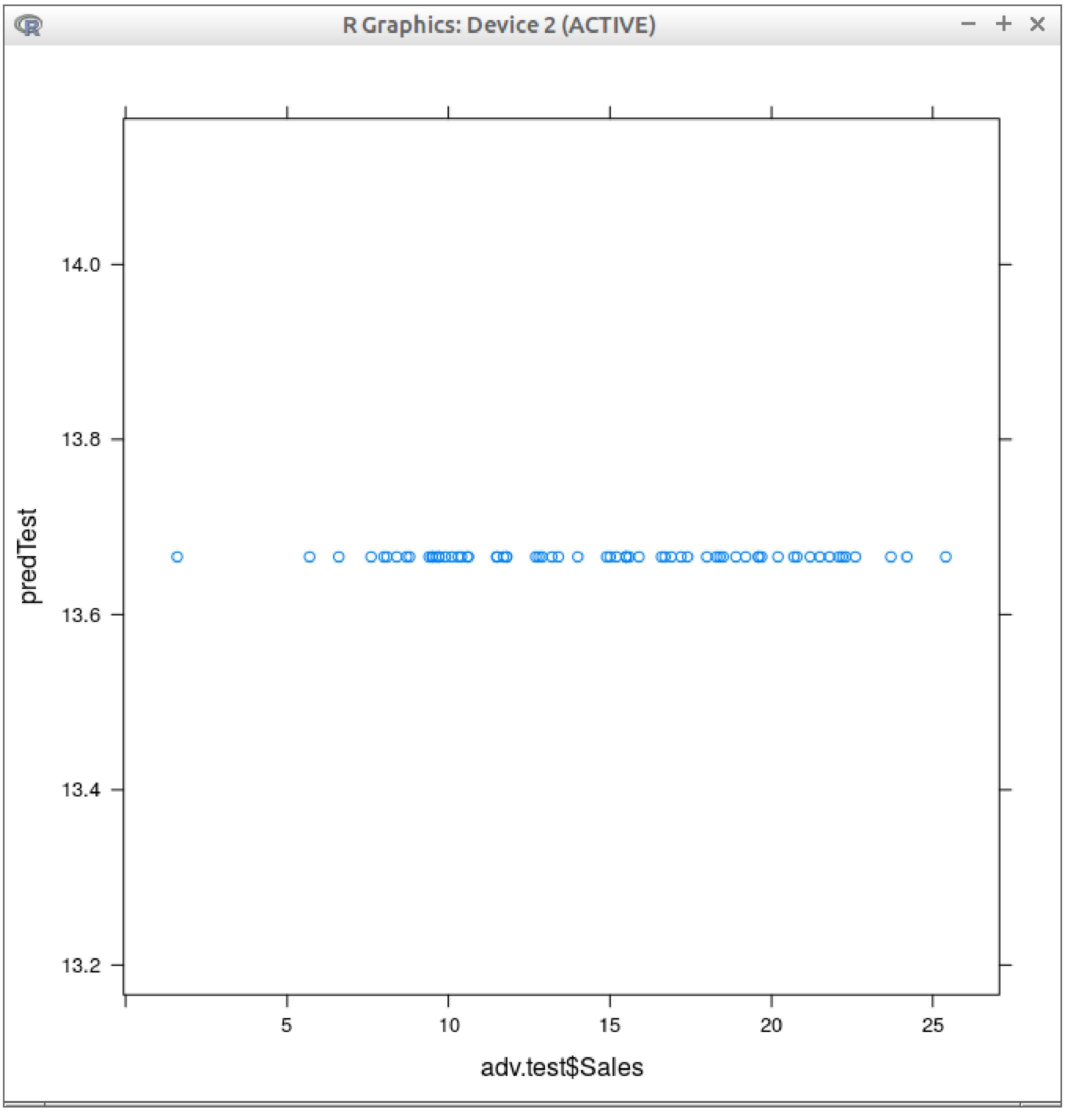


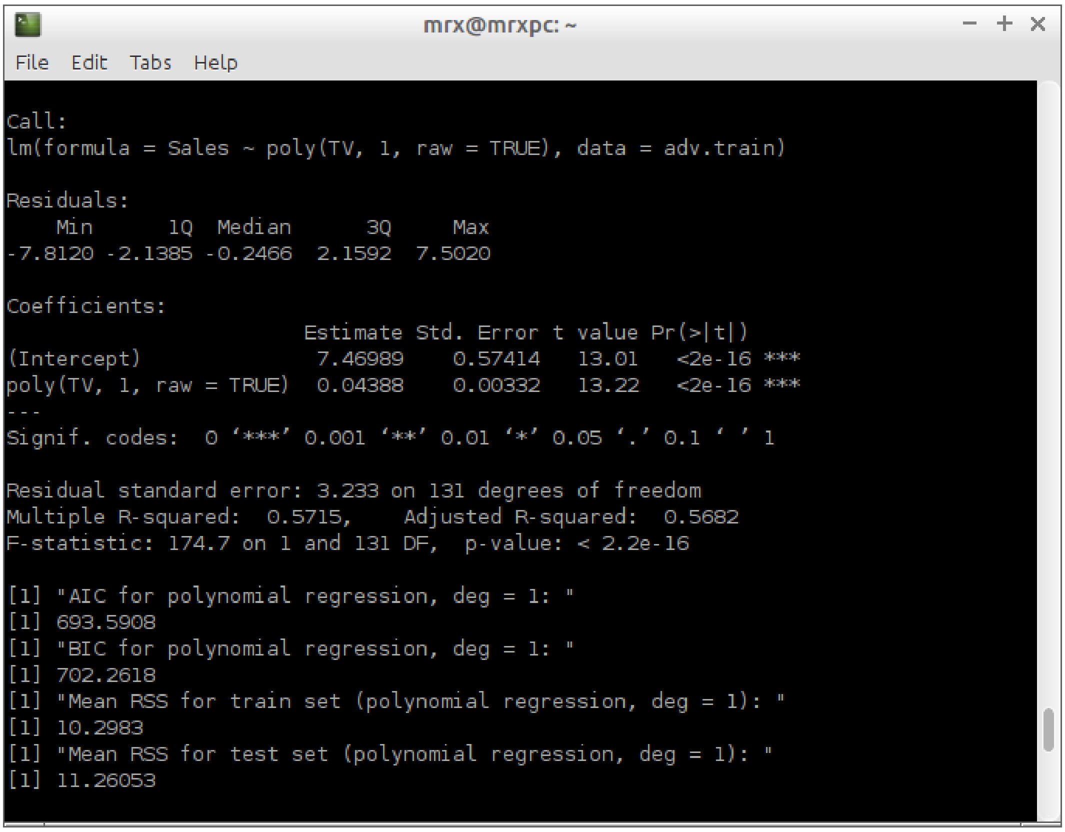
Figure 12: Scatterplot for test set (all predictors removed)

In conclusion it is important to note that on different program runs different result are observed. For example, one program run can give higher rss value for test data set in comparison to training data set. This fact could be explained by randomness of choosing what part of all the input data will be the test one and what will be the training one.

Additional task: polynomial regression

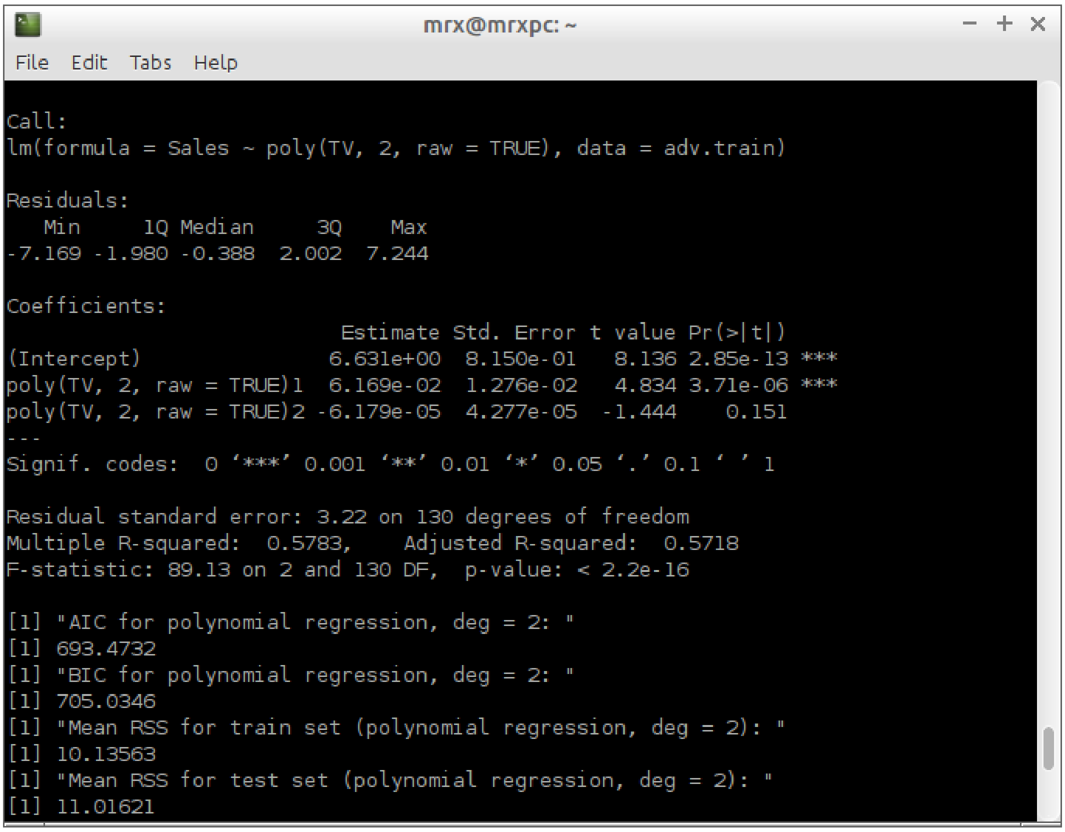
Let's look at the Figures 13, 14, 15 and 16 representing polynomial regression models with degrees 1, 2, 3 and 4 respectively.
First of all, F-statistics are significantly lower in comparison with one in linear model with Newspapers predictor removed (of course, current and previous results are obtained from different program runs, but still avarage value of F-statistics for linear model with Newspapers predictor removed between runs is close to 500). Moreovere, as the degree is increasing the F-statistics value is decreasing. Also, while degree is increasing predictors become more and more insignificant having predictor in the power of 1 the most significant among the others.
To study models' suitability AIC and BIC values are also shown. As one can notice, both grow slowly as polynomial degree increases having average value approximately equals to 700. It is important to note AIC and BIC values for linear model with Newspapers predictor removed are approximately 530.
Mean RSS values for training and test data are pretty high and remain almost unchanged as polynomial degree increases.
Figures 17 - 20 and 21 - 24 show scatterplots for training and test data respectively. As degree is increasing the scatterplots are almost unchanged and resemble scatterblot for degree = 1. It prooves the fact predictors in powers higher than 1 are least significant and don't affect the model much.

All the obtained results show the polynomial models are less suitable for the data set under study then linear ones and the higher the degree the worst model describes our data. So, speaking about using polynomial regression for this task the best choice is 2nd degree polynomial.



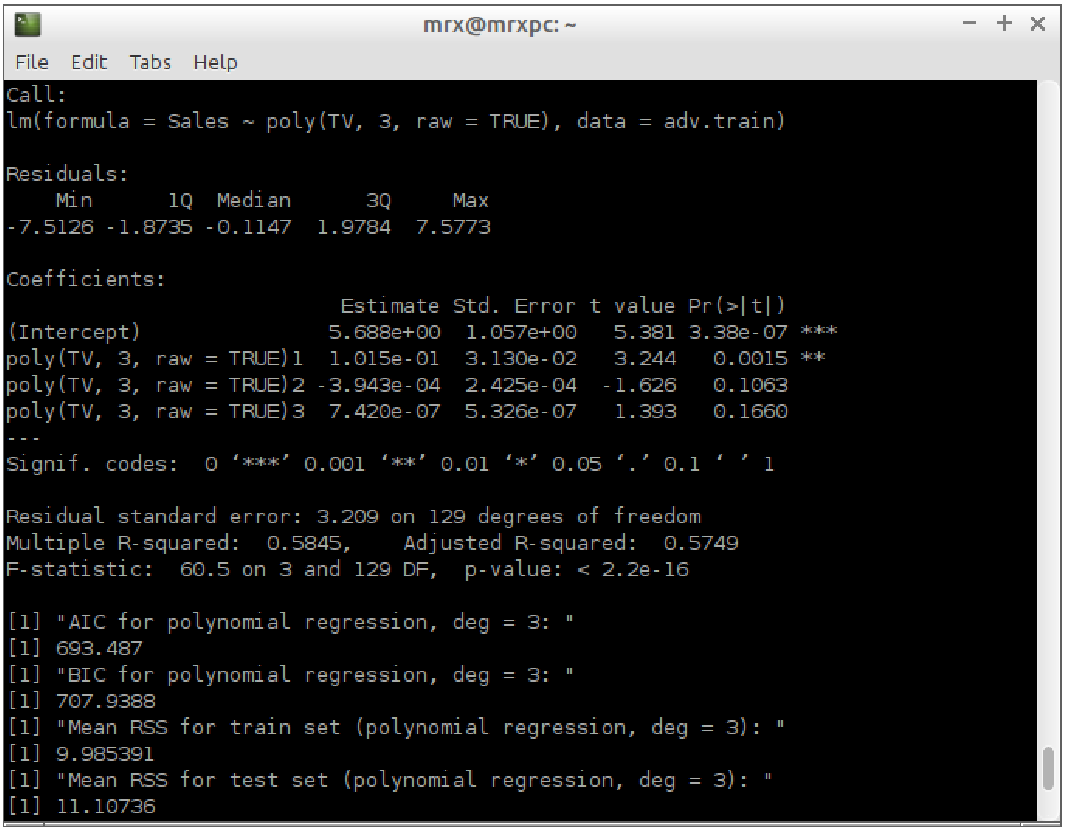

```
Call:
lm(formula = Sales ~ poly(TV, 1, raw = TRUE), data = adv.train)

Residuals:
    Min      1Q  Median      3Q     Max
-7.8120 -2.1385 -0.2466  2.1592  7.5020

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              7.46989    0.57414   13.01   <2e-16 ***
poly(TV, 1, raw = TRUE)  0.04388    0.00332   13.22   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.233 on 131 degrees of freedom
Multiple R-squared:  0.5715,    Adjusted R-squared:  0.5682
F-statistic: 174.7 on 1 and 131 DF,  p-value: < 2.2e-16

[1] "AIC for polynomial regression, deg = 1: "
[1] 693.5908
[1] "BIC for polynomial regression, deg = 1: "
[1] 702.2618
[1] "Mean RSS for train set (polynomial regression, deg = 1): "
[1] 10.2983
[1] "Mean RSS for test set (polynomial regression, deg = 1): "
[1] 11.26053
```

Figure 13: Polynomial regression with deg = 1



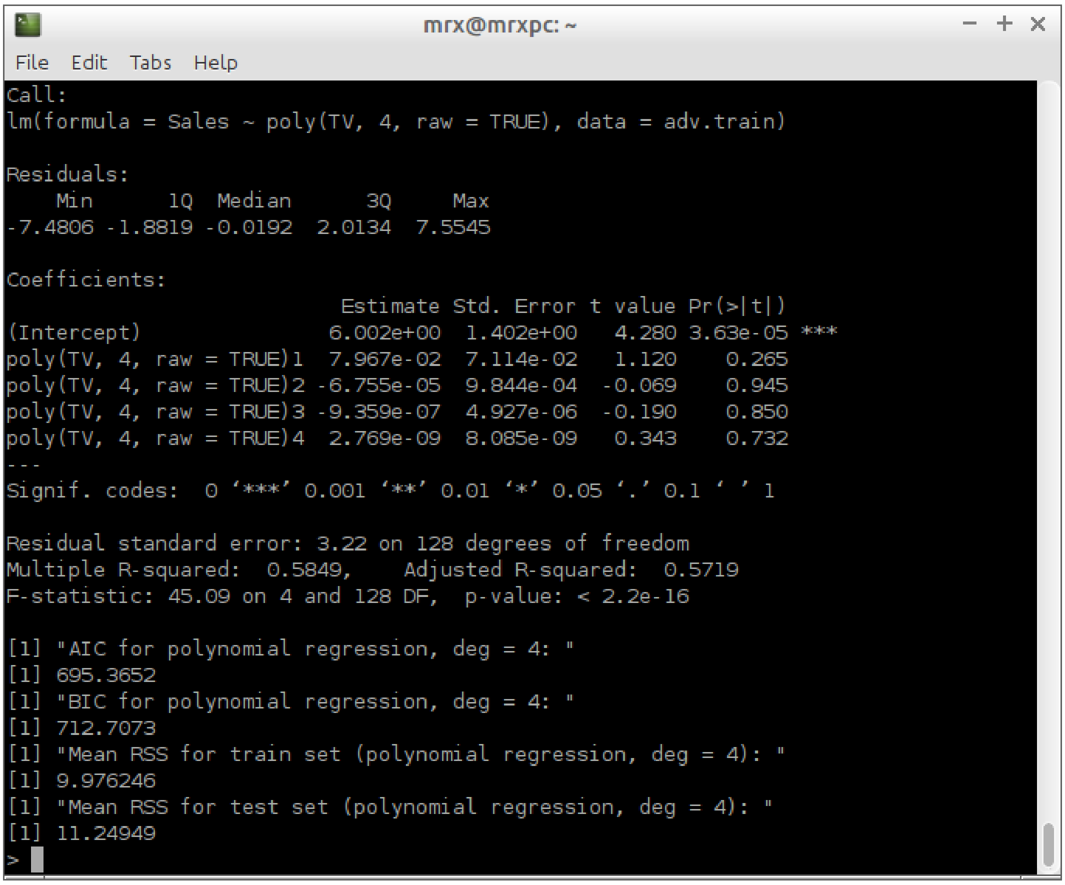

```
Call:
lm(formula = Sales ~ poly(TV, 2, raw = TRUE), data = adv.train)

Residuals:
   Min     1Q Median     3Q    Max
-7.169 -1.980 -0.388  2.002  7.244

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               6.631e+00  8.150e-01   8.136 2.85e-13 ***
poly(TV, 2, raw = TRUE)1  6.169e-02  1.276e-02   4.834 3.71e-06 ***
poly(TV, 2, raw = TRUE)2 -6.179e-05  4.277e-05  -1.444    0.151
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.22 on 130 degrees of freedom
Multiple R-squared:  0.5783,    Adjusted R-squared:  0.5718
F-statistic: 89.13 on 2 and 130 DF,  p-value: < 2.2e-16

[1] "AIC for polynomial regression, deg = 2: "
[1] 693.4732
[1] "BIC for polynomial regression, deg = 2: "
[1] 705.0346
[1] "Mean RSS for train set (polynomial regression, deg = 2): "
[1] 10.13563
[1] "Mean RSS for test set (polynomial regression, deg = 2): "
[1] 11.01621
```

Figure 14: Polynomial regression with deg = 2



```
Call:
lm(formula = Sales ~ poly(TV, 3, raw = TRUE), data = adv.train)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.5126 -1.8735 -0.1147  1.9784  7.5773 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              5.688e+00  1.057e+00   5.381 3.38e-07 ***
poly(TV, 3, raw = TRUE)1 1.015e-01  3.130e-02   3.244   0.0015 ** 
poly(TV, 3, raw = TRUE)2 -3.943e-04  2.425e-04  -1.626   0.1063    
poly(TV, 3, raw = TRUE)3 7.420e-07  5.326e-07   1.393   0.1660    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.209 on 129 degrees of freedom
Multiple R-squared:  0.5845,	Adjusted R-squared:  0.5749 
F-statistic:  60.5 on 3 and 129 DF,  p-value: < 2.2e-16

[1] "AIC for polynomial regression, deg = 3: "
[1] 693.487
[1] "BIC for polynomial regression, deg = 3: "
[1] 707.9388
[1] "Mean RSS for train set (polynomial regression, deg = 3): "
[1] 9.985391
[1] "Mean RSS for test set (polynomial regression, deg = 3): "
[1] 11.10736
```

Figure 15: Polynomial regression with deg = 3




```
Call:
lm(formula = Sales ~ poly(TV, 4, raw = TRUE), data = adv.train)

Residuals:
    Min      1Q  Median      3Q     Max
-7.4806 -1.8819 -0.0192  2.0134  7.5545

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               6.002e+00  1.402e+00   4.280 3.63e-05 ***
poly(TV, 4, raw = TRUE)1  7.967e-02  7.114e-02   1.120    0.265
poly(TV, 4, raw = TRUE)2 -6.755e-05  9.844e-04  -0.069    0.945
poly(TV, 4, raw = TRUE)3 -9.359e-07  4.927e-06  -0.190    0.850
poly(TV, 4, raw = TRUE)4  2.769e-09  8.085e-09   0.343    0.732
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.22 on 128 degrees of freedom
Multiple R-squared:  0.5849,    Adjusted R-squared:  0.5719
F-statistic: 45.09 on 4 and 128 DF,  p-value: < 2.2e-16

[1] "AIC for polynomial regression, deg = 4: "
[1] 695.3652
[1] "BIC for polynomial regression, deg = 4: "
[1] 712.7073
[1] "Mean RSS for train set (polynomial regression, deg = 4): "
[1] 9.976246
[1] "Mean RSS for test set (polynomial regression, deg = 4): "
[1] 11.24949
>
```

Figure 16: Polynomial regression with deg = 4

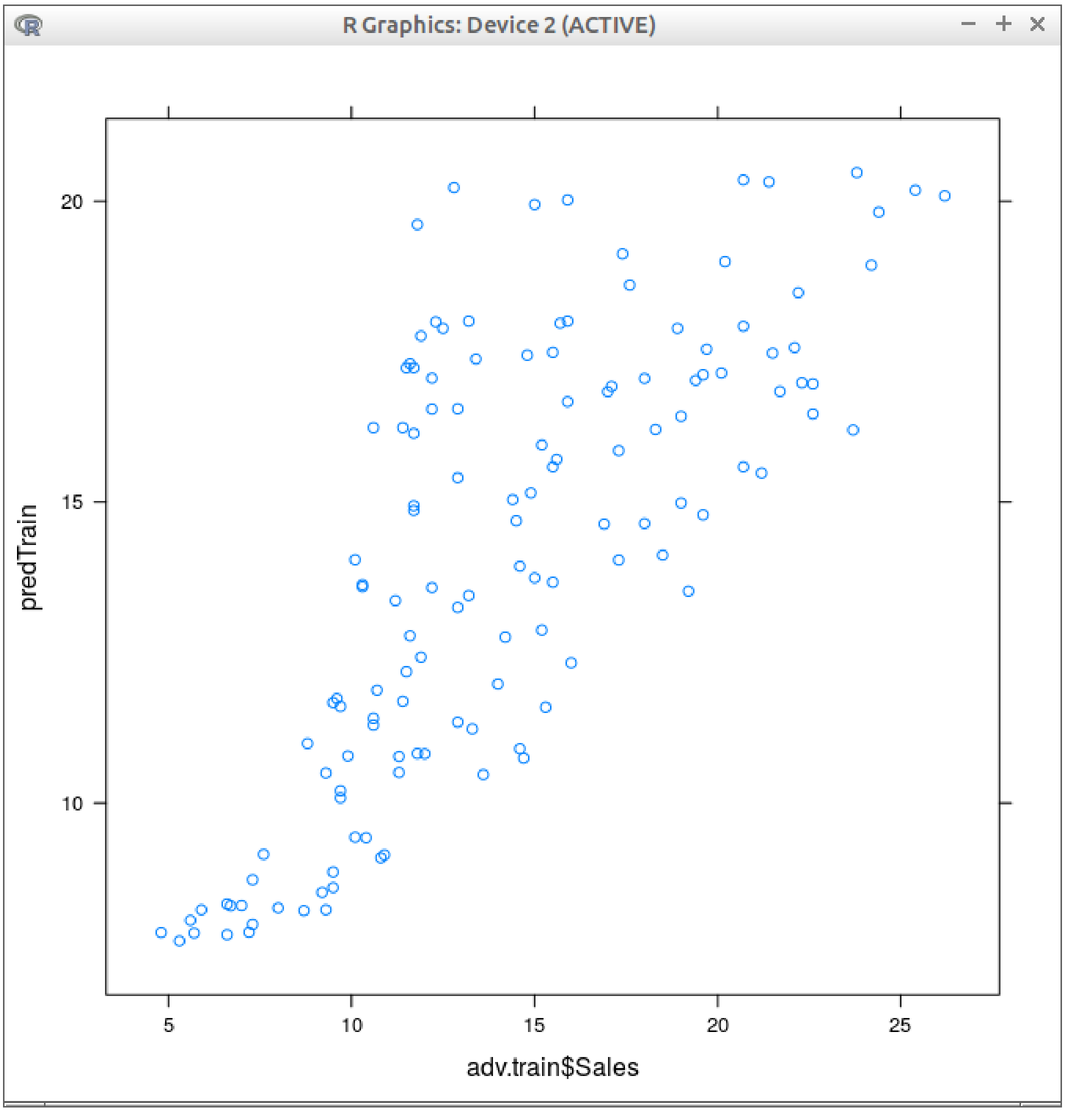


Figure 17: Scatterplot for training set (polynomial regression with deg = 1)

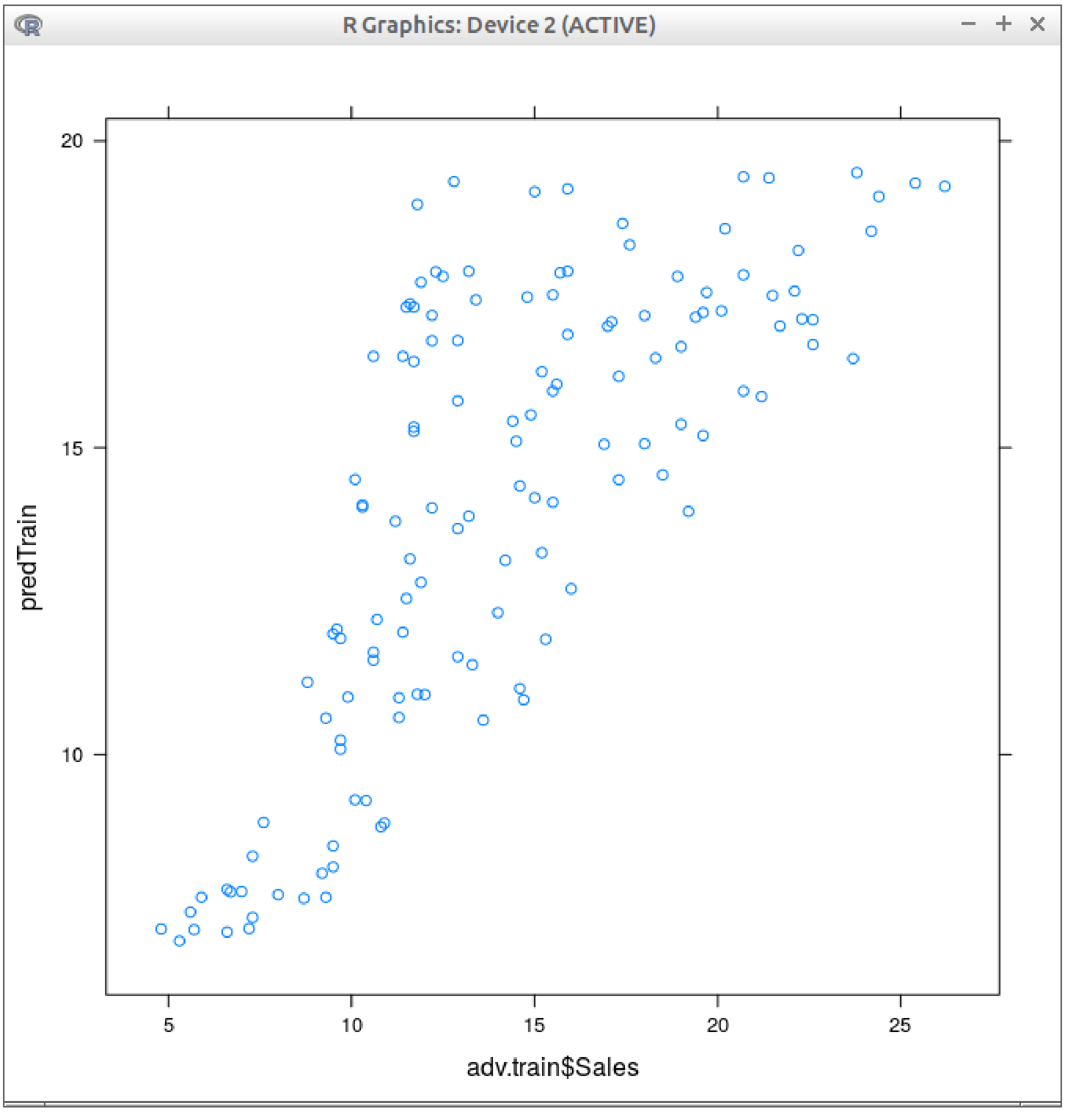


Figure 18: Scatterplot for training set (polynomial regression with deg = 2)

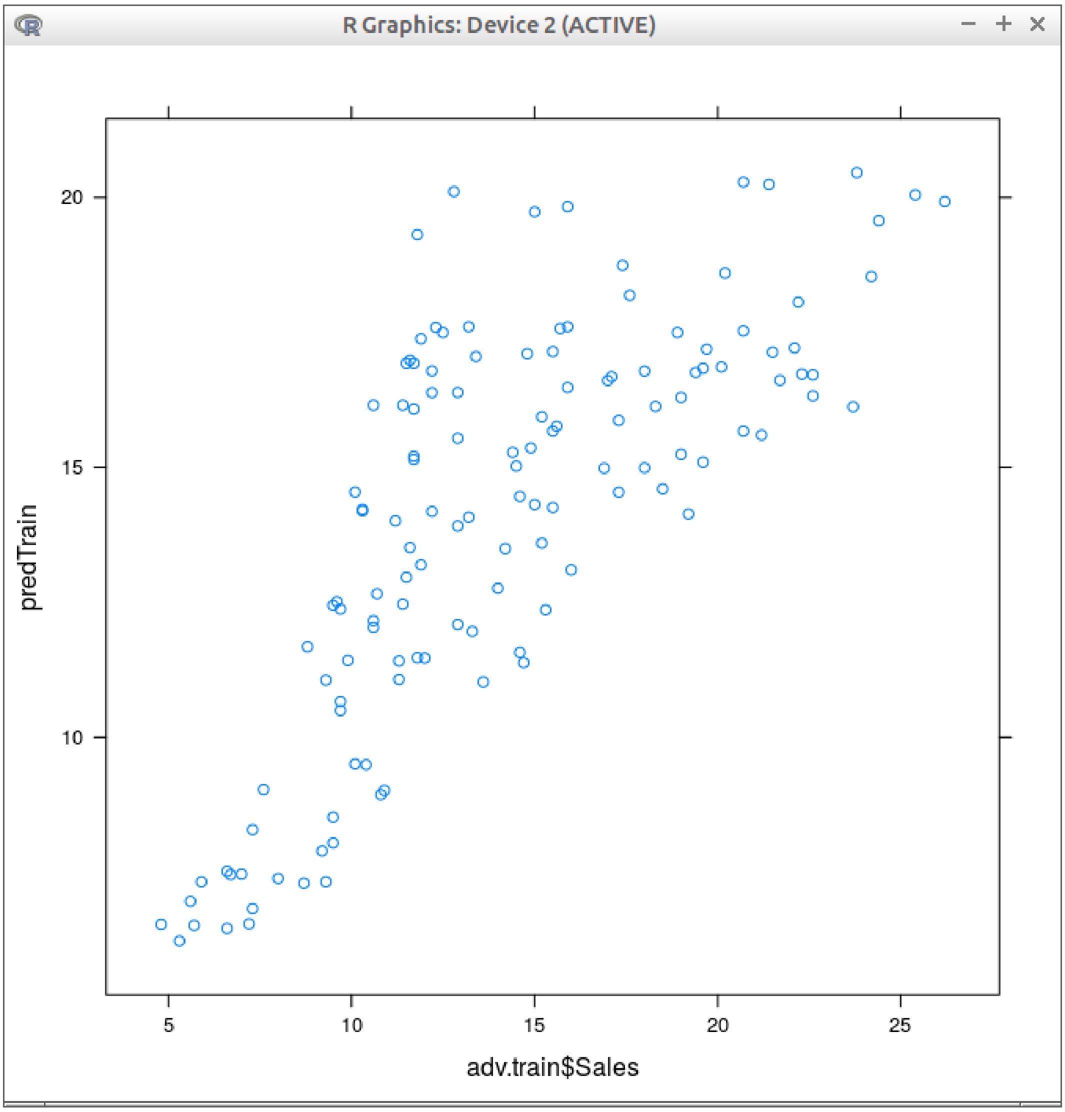


Figure 19: Scatterplot for training set (polynomial regression with deg = 3)

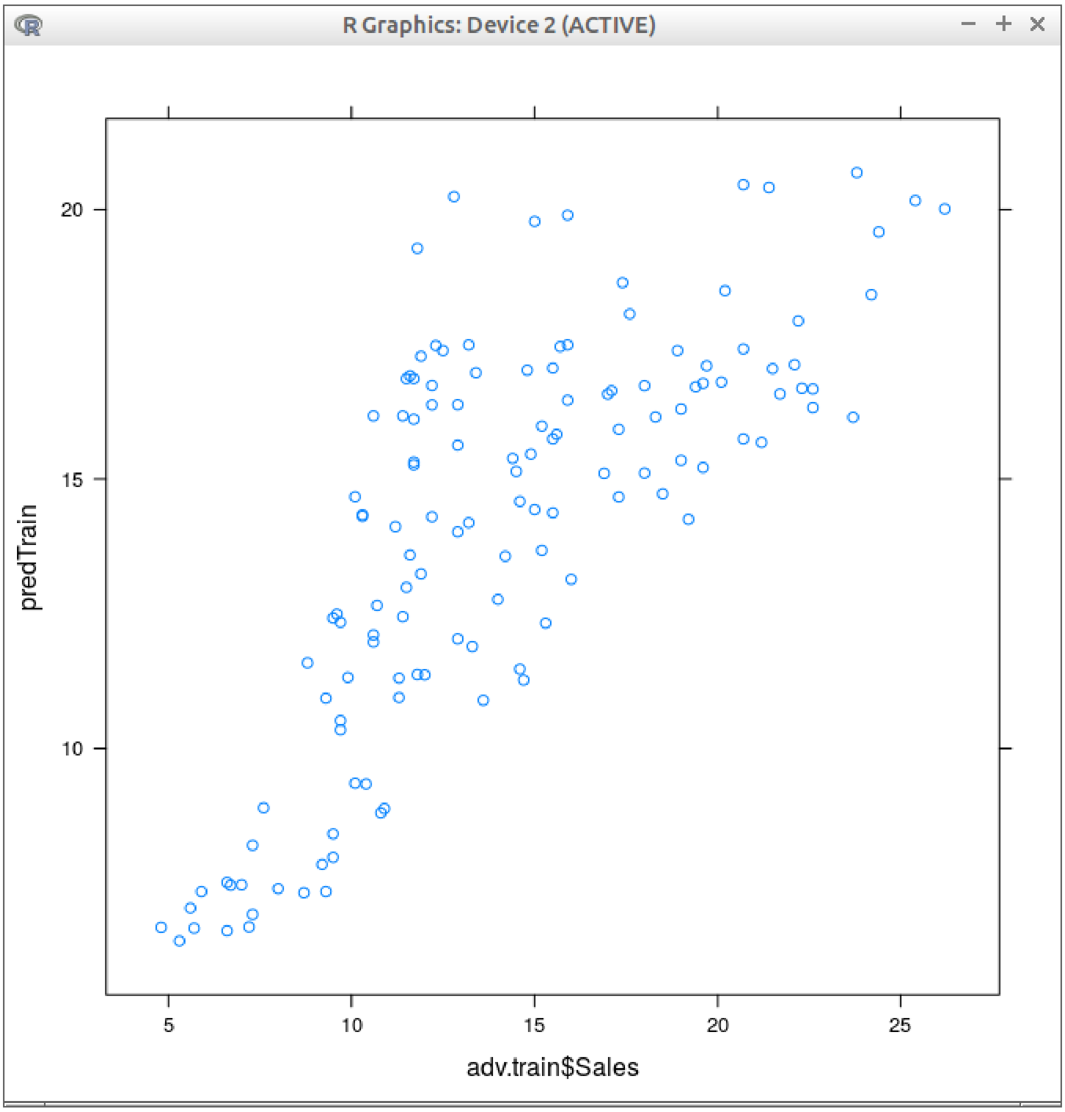


Figure 20: Scatterplot for training set (polynomial regression with deg = 4)

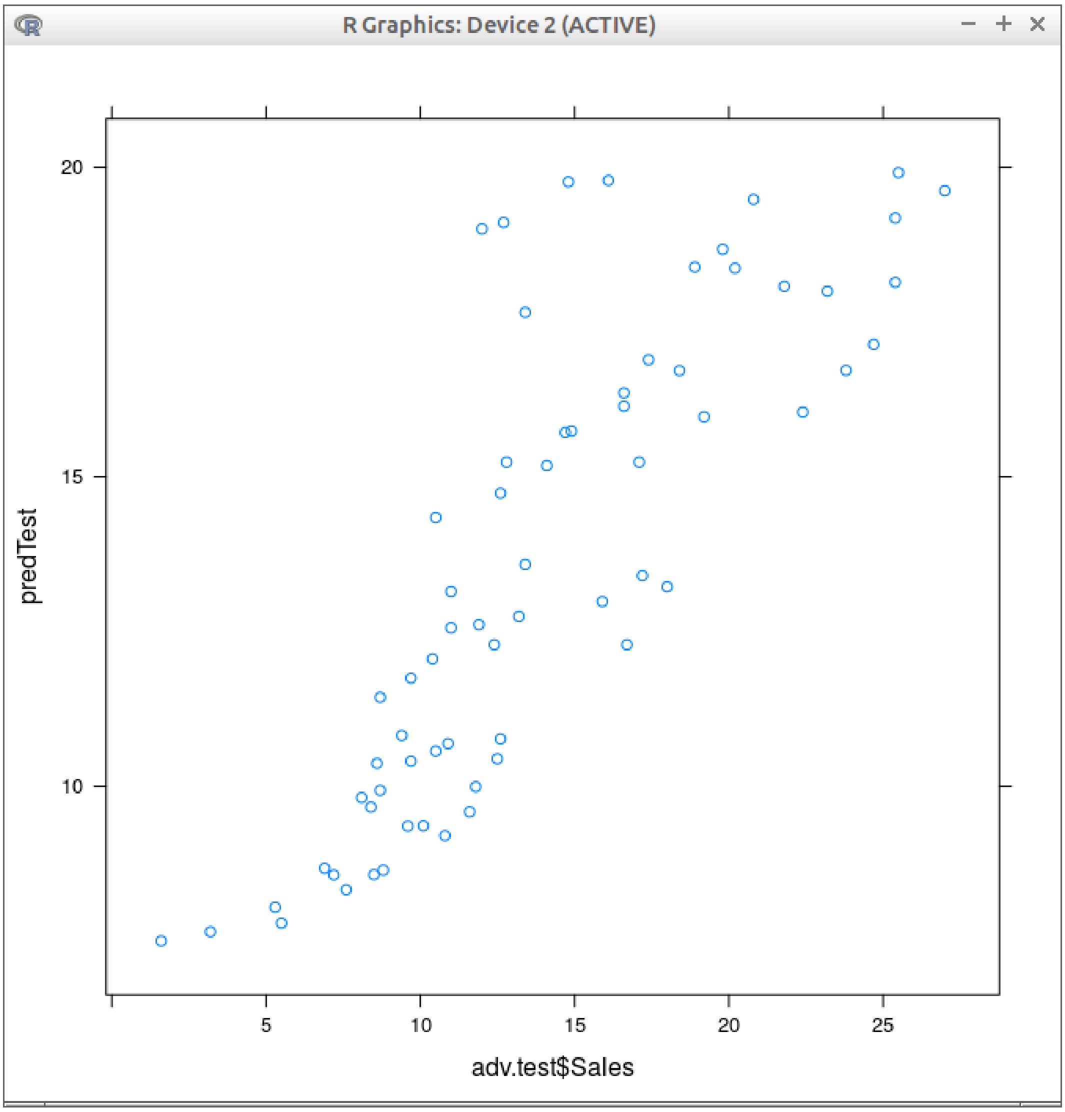


Figure 21: Scatterplot for test set (polynomial regression with deg = 1)

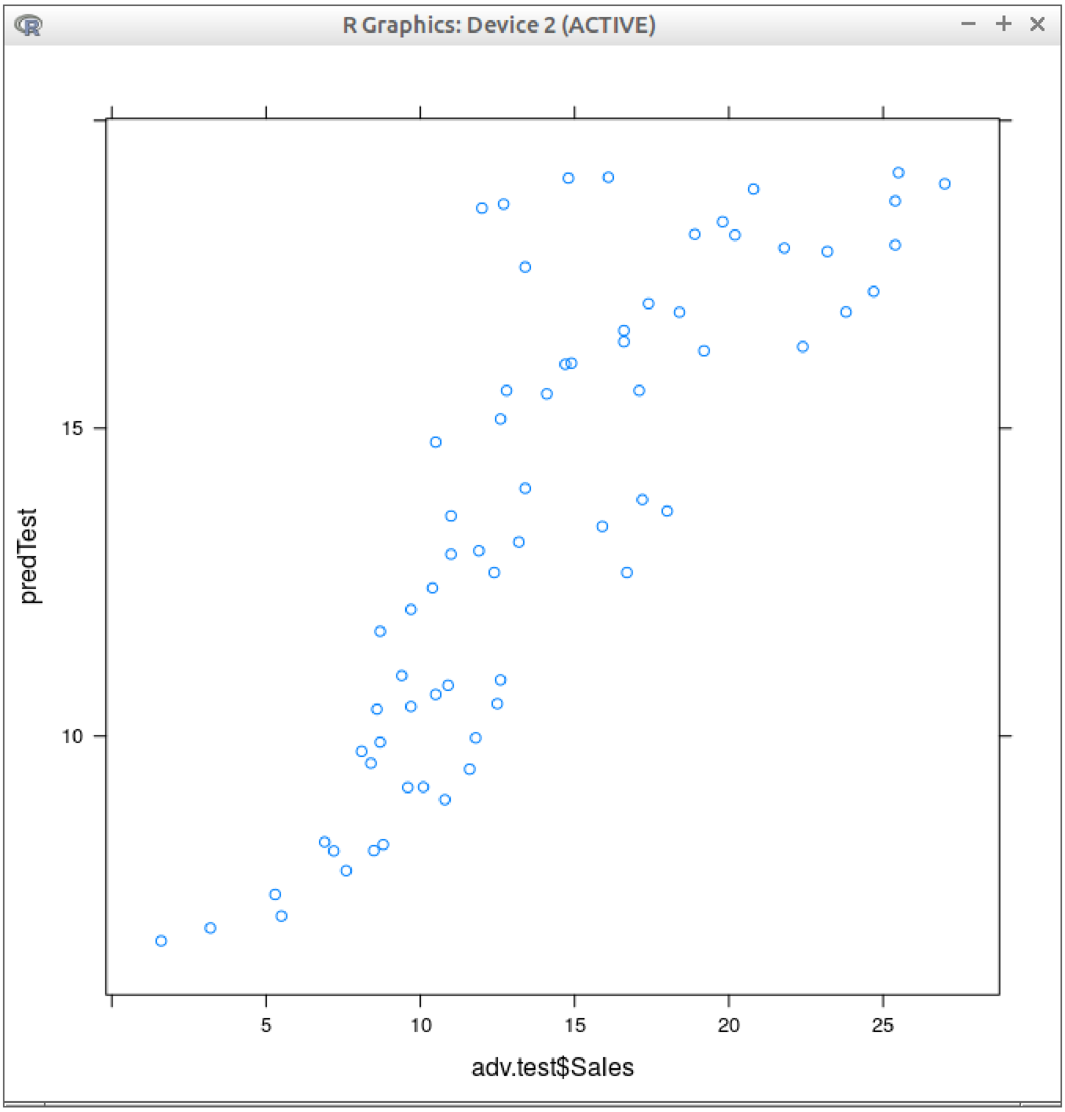


Figure 22: Scatterplot for test set (polynomial regression with deg = 2)

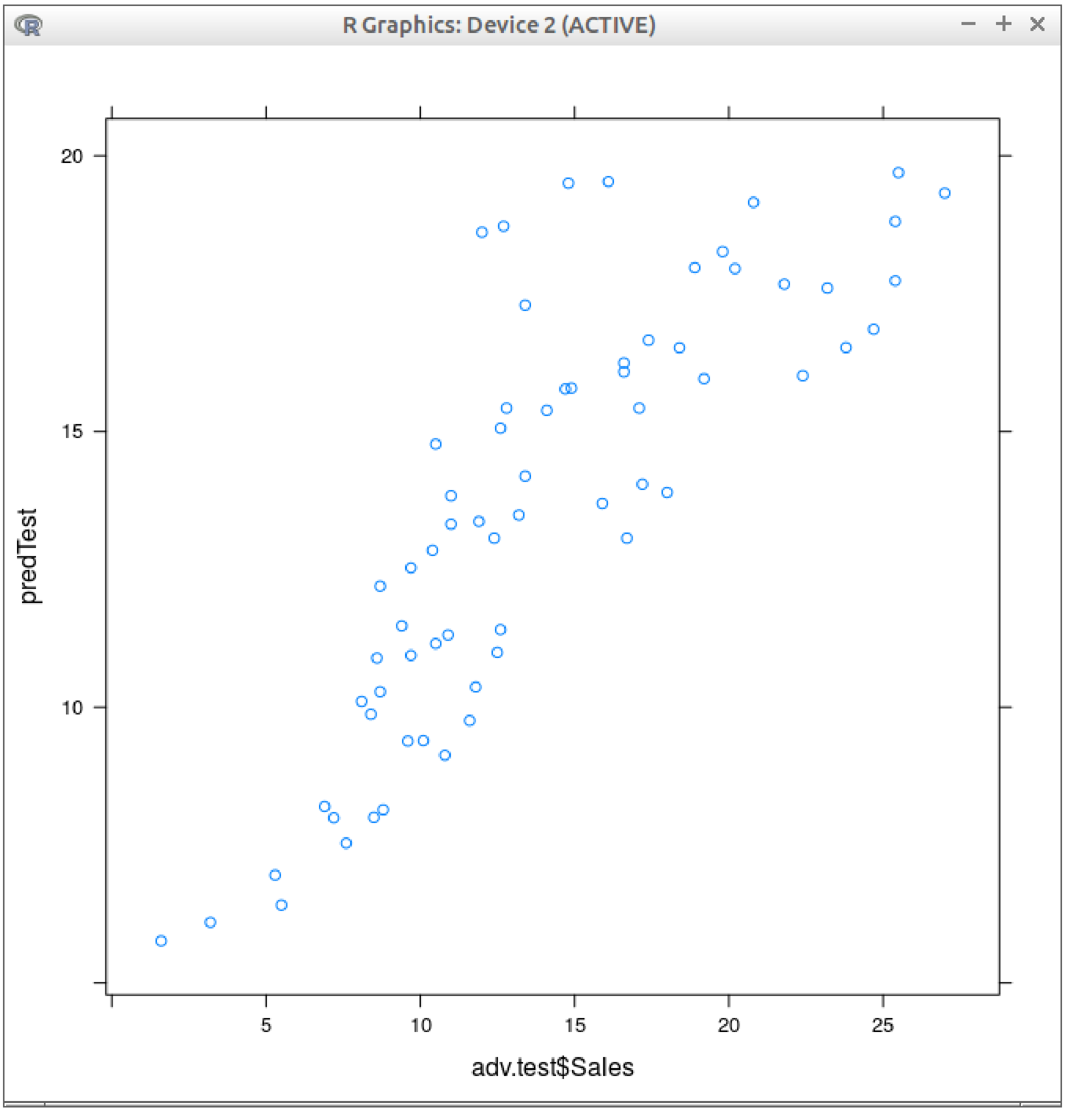


Figure 23: Scatterplot for test set (polynomial regression with deg = 3)

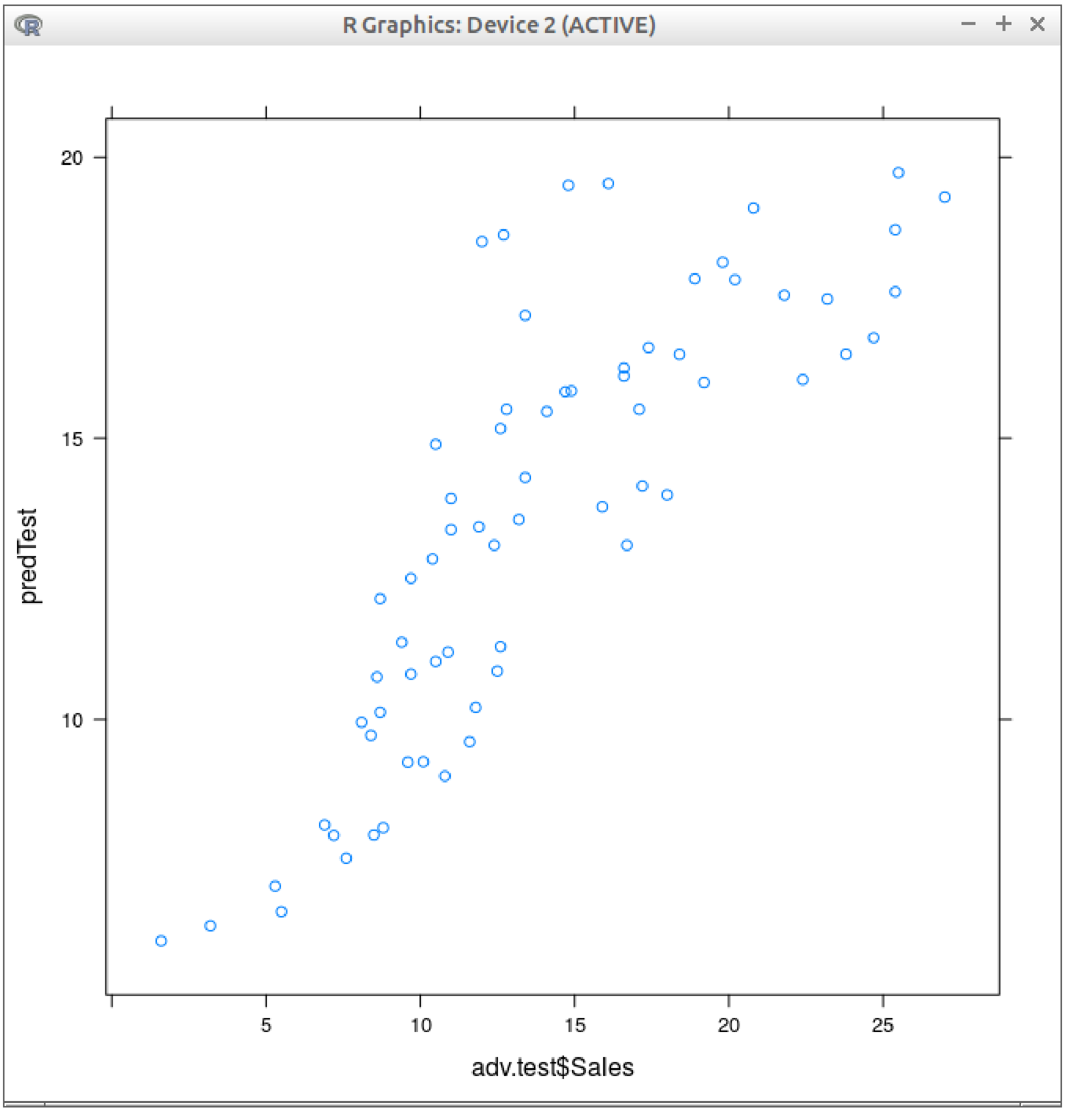


Figure 24: Scatterplot for test set (polynomial regression with deg = 4)