MidtermTest

Matous Dzivjak

Arrests

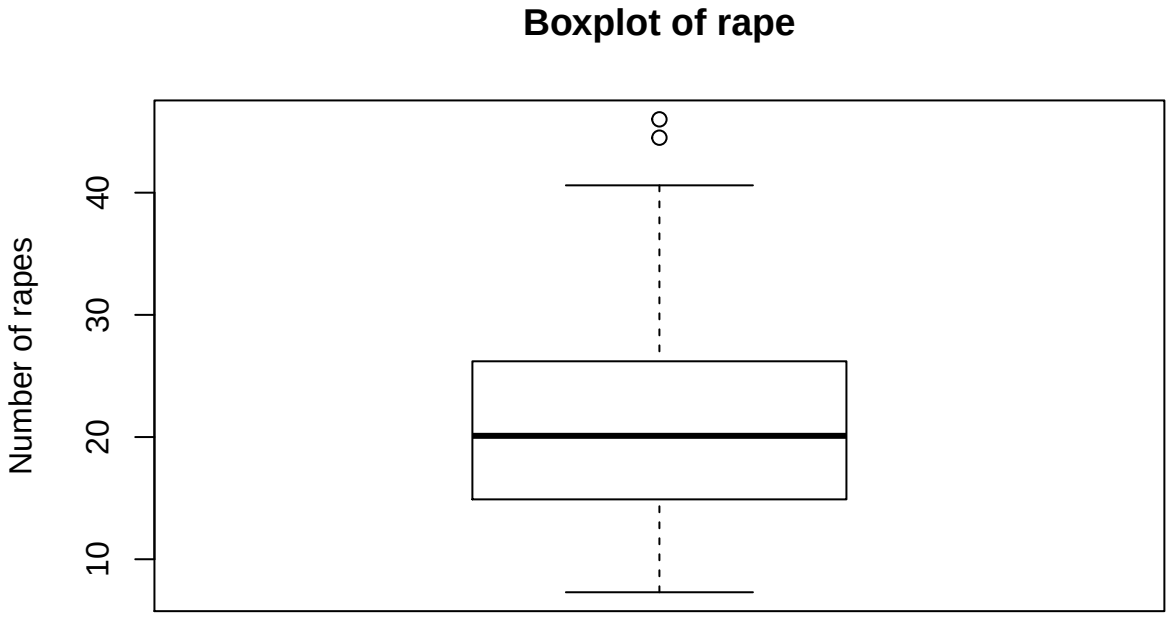
```
advertising <-  read.csv('datasets/Advertising.csv', header = TRUE)
arrests <-  read.csv('datasets/USArrests.csv', header = TRUE)
advertising <- as.data.frame(advertising)
summary(arrests)
```

```
##           X          Murder          Assault         UrbanPop
##  Alabama   : 1   Min.   : 0.800   Min.   : 45.0   Min.   :32.00
##  Alaska    : 1   1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50
##  Arizona   : 1   Median : 7.250   Median :159.0   Median :66.00
##  Arkansas  : 1   Mean   : 7.788   Mean   :170.8   Mean   :65.54
##  California: 1   3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75
##  Colorado  : 1   Max.   :17.400   Max.   :337.0   Max.   :91.00
##  (Other)   :44
##       Rape
##  Min.   : 7.30
##  1st Qu.:15.07
##  Median :20.10
##  Mean   :21.23
##  3rd Qu.:26.18
##  Max.   :46.00
##
```

```
attach(arrests)
quantile(Murder, probs = c(0.05), names=TRUE)
```

```
##    5%
## 2.145
```

```
boxplot(Rape, main = "Boxplot of rape", xlab = "All states", ylab = "Number of rapes")
```



In boxplot the box has center on median (half of the scores are greater and half lower) and lower and upper bounds where 50% of the scores lay (e.g. 50% of the scores are in the box). The lines connected to the box with vertical lines are upper quartile (seventy-five percent of the scores fall below the upper quartile) and lower quartile (twenty-five percent of scores fall below the lower quartile). Separate points are outliers

Advertising

```
library(ggplot2)
attach(advertising)
```

```
## The following object is masked from arrests:
##
##     X
```

```
summary(advertising)
```

```
##        X                TV              radio          newspaper
##  Min.   :  1.00   Min.   :  0.70   Min.   : 0.000   Min.   :  0.30
##  1st Qu.: 50.75   1st Qu.: 74.38   1st Qu.: 9.975   1st Qu.: 12.75
##  Median :100.50   Median :149.75   Median :22.900   Median : 25.75
##  Mean   :100.50   Mean   :147.04   Mean   :23.264   Mean   : 30.55
##  3rd Qu.:150.25   3rd Qu.:218.82   3rd Qu.:36.525   3rd Qu.: 45.10
##  Max.   :200.00   Max.   :296.40   Max.   :49.600   Max.   :114.00
##      sales
##  Min.   : 1.60
##  1st Qu.:10.38
##  Median :12.90
##  Mean   :14.02
##  3rd Qu.:17.40
```

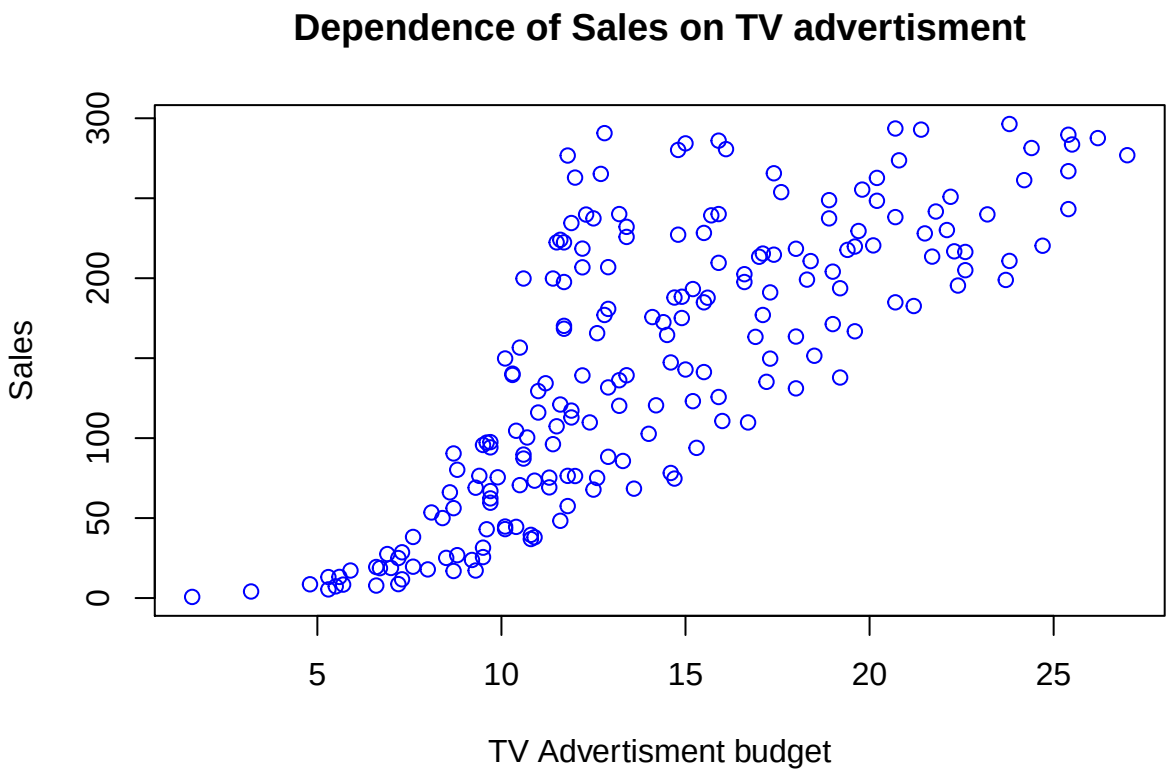
```
##  Max.   :27.00
```

TV sales

```
TVmodel <- lm(sales ~ TV)
summary(TVmodel)
```

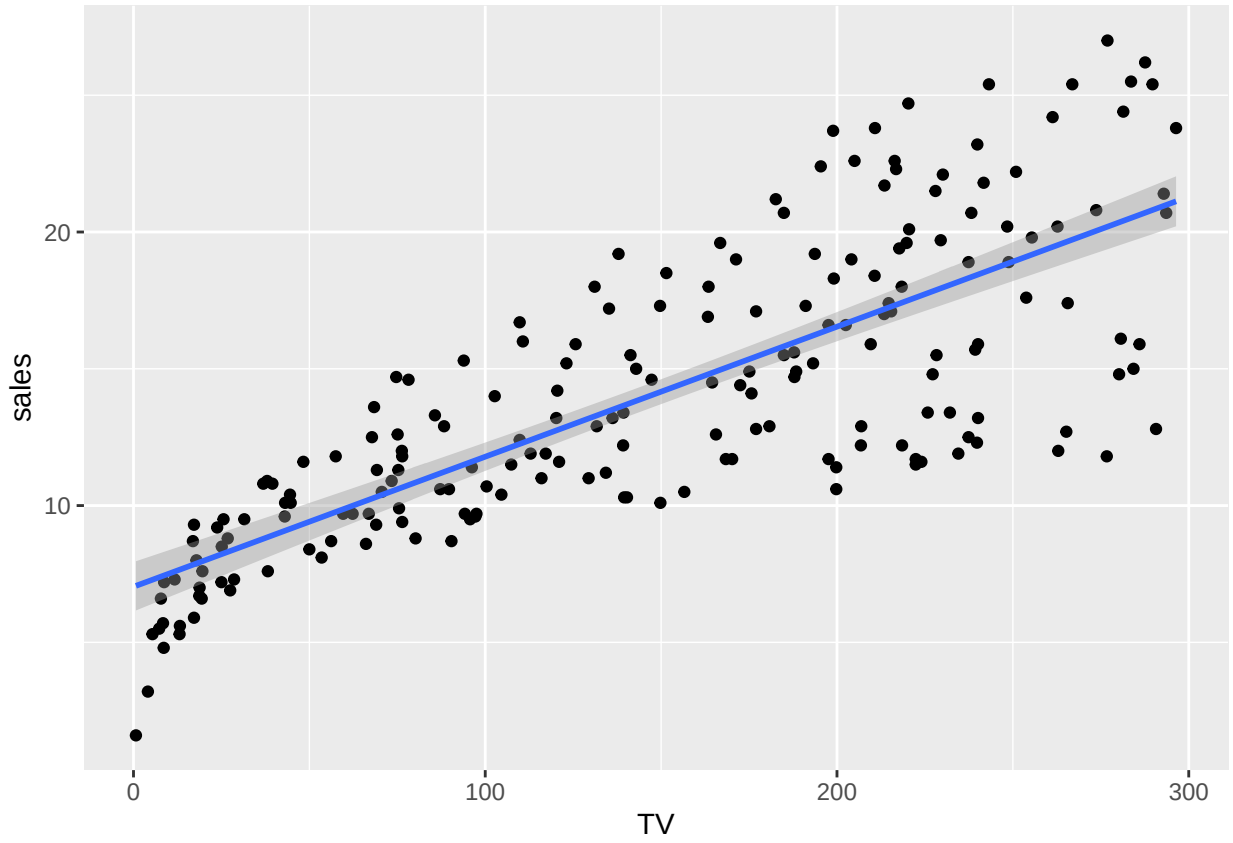
```
##
## Call:
## lm(formula = sales ~ TV)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.3860 -1.9545 -0.1913  2.0671  7.2124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.032594   0.457843   15.36   <2e-16 ***
## TV          0.047537   0.002691   17.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.259 on 198 degrees of freedom
## Multiple R-squared:  0.6119, Adjusted R-squared:  0.6099
## F-statistic: 312.1 on 1 and 198 DF,  p-value: < 2.2e-16
```

```
plot(sales, TV, col = "blue", main = "Dependence of Sales on TV advertisment", xlab = "TV Advertisment b
```




```
ggplot(data = advertising) +
  geom_point(mapping = aes(x = TV, y = sales)) +
```

```
  geom_smooth(mapping = aes(x = TV, y = sales), method=lm)
```

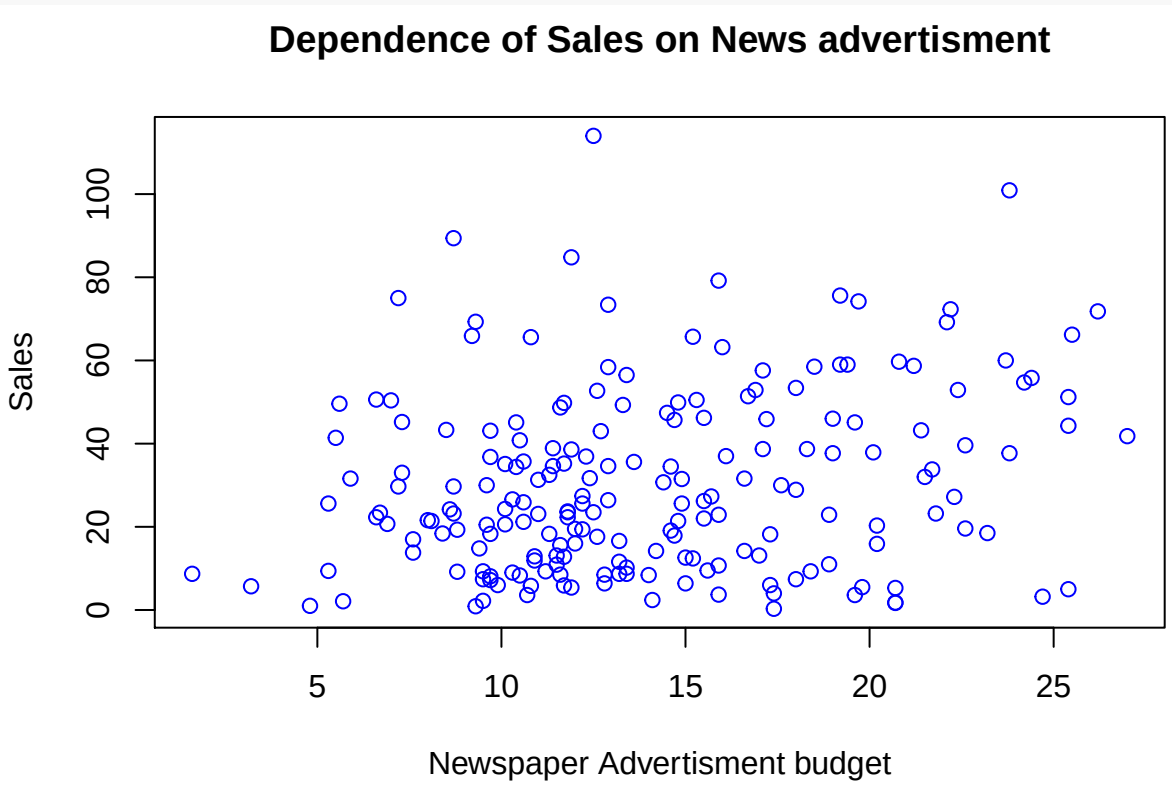


Newspaper sales

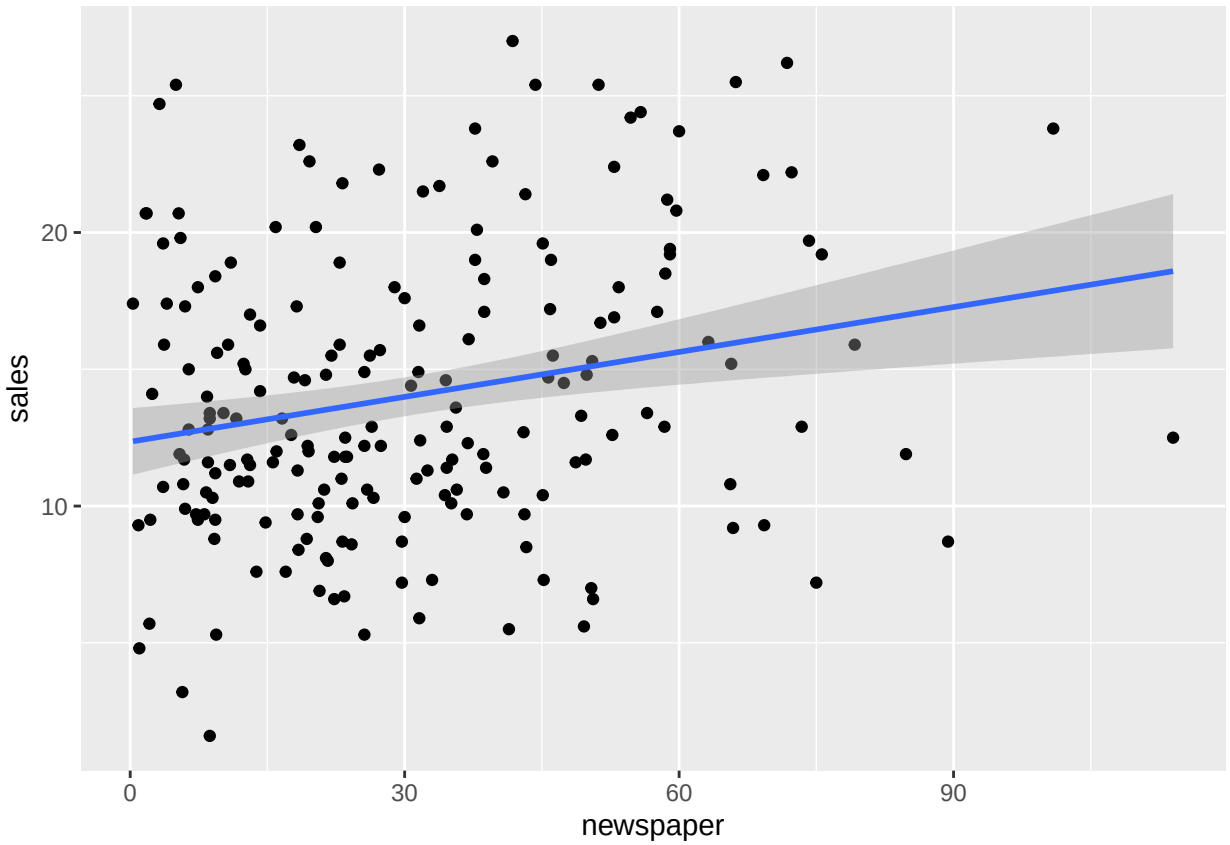
```
newsmodel <- lm(sales ~ newspaper)
summary(newsmodel)
```

```
##
## Call:
## lm(formula = sales ~ newspaper)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2272  -3.3873  -0.8392   3.5059  12.7751
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.35141    0.62142   19.88  < 2e-16 ***
## newspaper    0.05469    0.01658    3.30  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.092 on 198 degrees of freedom
## Multiple R-squared:  0.05212,    Adjusted R-squared:  0.04733
## F-statistic: 10.89 on 1 and 198 DF,  p-value: 0.001148
```

```r
plot(sales, newspaper, col = "blue", main = "Dependence of Sales on News advertisment", xlab = "Newspap
```




```r
ggplot(data = advertising) +
  geom_point(mapping = aes(x = newspaper, y = sales)) +
  geom_smooth(mapping = aes(x = newspaper, y = sales), method=lm)
```



Predict sales

```
newdata <- data.frame(TV = c(50))
predict(TVmodel, newdata)
```

```
##        1
## 9.409426
```

```
newdata <- data.frame(newspaper = c(50))
predict(newsmodel, newdata)
```

```
##        1
## 15.08606
```

We expect greater increase for newspaper because for one unit of advertisment we have bigger change in sales as cen be seen in the coefficient from summary of the model.