In [50]:
```
library(MASS)
library(lmtest)
library(gridExtra)
library(tidyverse)
library(readxl)
```

In [7]:
```
library(ggcorrplot)
```

In [8]:
```
getwd()
list.files()
```

'/Users/xuren/pydev/p1_udj_assignments/s11_forecasting'

'S11 UDJ SALES.xls' 's11_sales.ipynb'

In [9]:
```
sales_tibble = read_excel("S11 UDJ SALES.xls")
```

In [10]:
```
head(sales_tibble)
```

A tibble: 6 × 9

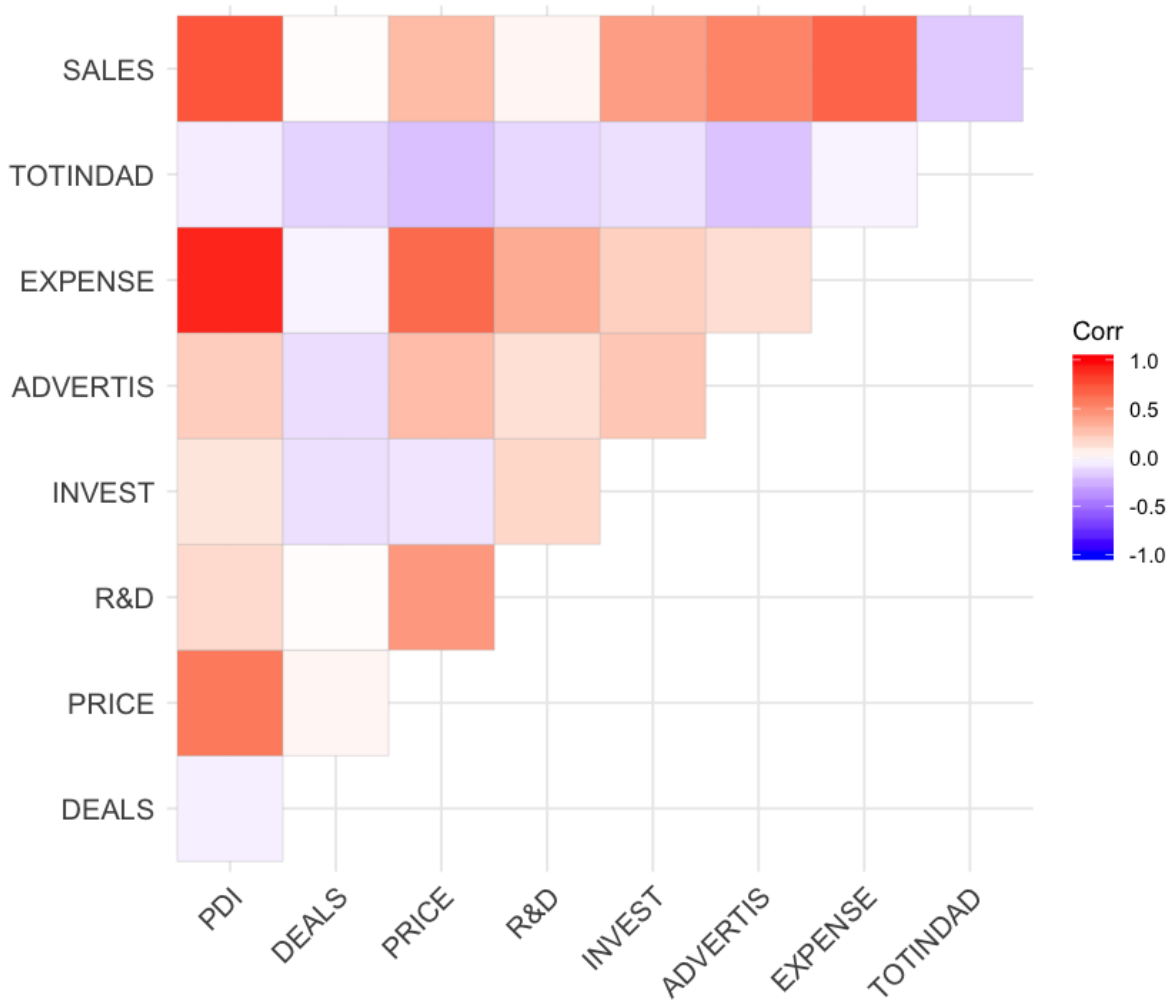
PDI	DEALS	PRICE	R&D	INVEST	ADVERTIS	EXPENSE	TOTINDAD	SALES
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
398	138	56.20	12.11	49.89	76.86	228.80	98.20	5540.39
369	118	59.04	9.33	16.59	88.81	177.45	224.95	5439.04
268	129	56.72	28.75	89.18	51.30	166.40	263.03	4290.00
484	111	57.86	12.89	106.73	39.65	258.05	320.93	5502.34
394	146	59.11	13.38	142.55	51.65	209.30	406.99	4871.77
332	140	60.11	11.09	61.28	20.55	180.05	247.00	4708.08

In [11]:
```
cor(sales_tibble %>% dplyr::select(-SALES))
```

A matrix: 8 × 8 of type dbl

	PDI	DEALS	PRICE	R&D	INVEST	ADVERTIS	EXPI
PDI	1.00000000	-0.051975827	0.58178908	0.162751135	0.10655595	0.2057018	0.9017
DEALS	-0.05197583	1.000000000	0.04420068	0.008085699	-0.09995092	-0.1095744	-0.040
PRICE	0.58178908	0.044200685	1.00000000	0.437099447	-0.08937882	0.2763723	0.638
R&D	0.16275114	0.008085699	0.43709945	1.000000000	0.16733956	0.1268537	0.362
INVEST	0.10655595	-0.099950919	-0.08937882	0.167339564	1.00000000	0.2380239	0.195
ADVERTIS	0.20570182	-0.109574424	0.27637231	0.126853716	0.23802387	1.0000000	0.143
EXPENSE	0.90171491	-0.040904276	0.63869001	0.362759847	0.19561952	0.1439814	1.000
TOTINDAD	-0.05611052	-0.152765420	-0.21352408	-0.129457379	-0.10084968	-0.1986011	-0.040

In [12]:
```
ggcorrplot(cor(sales_tibble), type="upper")
```



In [13]:
```
model1 <- lm(SALES ~ ., data = sales_tibble)
```

In [14]:
```
summary(model1)
```

```
Call:
lm(formula = SALES ~ ., data = sales_tibble)

Residuals:
   Min     1Q Median     3Q    Max 
-385.9 -118.0  -13.0  138.2  462.4 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3027.6336   572.3599   5.290 9.36e-06 ***
PDI            3.3723     1.4137   2.385   0.0234 *  
DEALS          4.6953     3.0103   1.560   0.1290    
PRICE        -18.1112     7.4927  -2.417   0.0217 *  
`R&D`         -9.9033     6.0021  -1.650   0.1090    
INVEST         1.6895     0.7157   2.361   0.0247 *  
ADVERTIS       8.2907     1.6445   5.042 1.91e-05 ***
EXPENSE        4.4434     2.5076   1.772   0.0862 .  
TOTINDAD      -0.4427     0.3646  -1.214   0.2339    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 234.6 on 31 degrees of freedom
Multiple R-squared:  0.8338,    Adjusted R-squared:  0.791 
F-statistic: 19.44 on 8 and 31 DF,  p-value: 4.437e-10
```

In [15]:
```
step.model <- stepAIC(model1, direction = "both",
                      trace = FALSE)
summary(step.model)
```

```
Call:
lm(formula = SALES ~ PDI + DEALS + PRICE + `R&D` + INVEST + ADVERTIS + 
    EXPENSE, data = sales_tibble)

Residuals:
    Min      1Q  Median      3Q     Max 
-466.83 -138.75    4.12  118.98  417.08 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2732.9641   522.1804   5.234 1.01e-05 ***
PDI            3.4996     1.4202   2.464   0.0193 *  
DEALS          5.3063     2.9898   1.775   0.0855 .  
PRICE        -16.2325     7.3852  -2.198   0.0353 *  
`R&D`         -9.5020     6.0372  -1.574   0.1253    
INVEST         1.8193     0.7129   2.552   0.0157 *  
ADVERTIS       8.4837     1.6489   5.145 1.30e-05 ***
EXPENSE        4.0191     2.5015   1.607   0.1179    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 236.4 on 32 degrees of freedom
Multiple R-squared:  0.8259,    Adjusted R-squared:  0.7879 
F-statistic: 21.69 on 7 and 32 DF,  p-value: 1.812e-10
```

In [19]:
```
# ?stepAIC
```

In [20]:
```
final.model <- lm(SALES ~ PDI + PRICE + ADVERTIS + INVEST, data=sales_tibble)
```

In [21]:
```
summary(final.model)
```

```
Call:
lm(formula = SALES ~ PDI + PRICE + ADVERTIS + INVEST, data = sales_tibble)

Residuals:
    Min      1Q  Median      3Q     Max 
-571.71 -138.10   16.65  190.58  455.81 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3333.3884   363.2868   9.176 7.66e-11 ***
PDI            5.5343     0.7056   7.844 3.22e-09 ***
PRICE        -15.3901     6.5801  -2.339  0.02518 *  
ADVERTIS       7.5453     1.6519   4.568 5.88e-05 ***
INVEST         1.9211     0.6879   2.793  0.00842 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 248.7 on 35 degrees of freedom
Multiple R-squared:  0.7892,    Adjusted R-squared:  0.7651 
F-statistic: 32.76 on 4 and 35 DF,  p-value: 2.176e-11
```

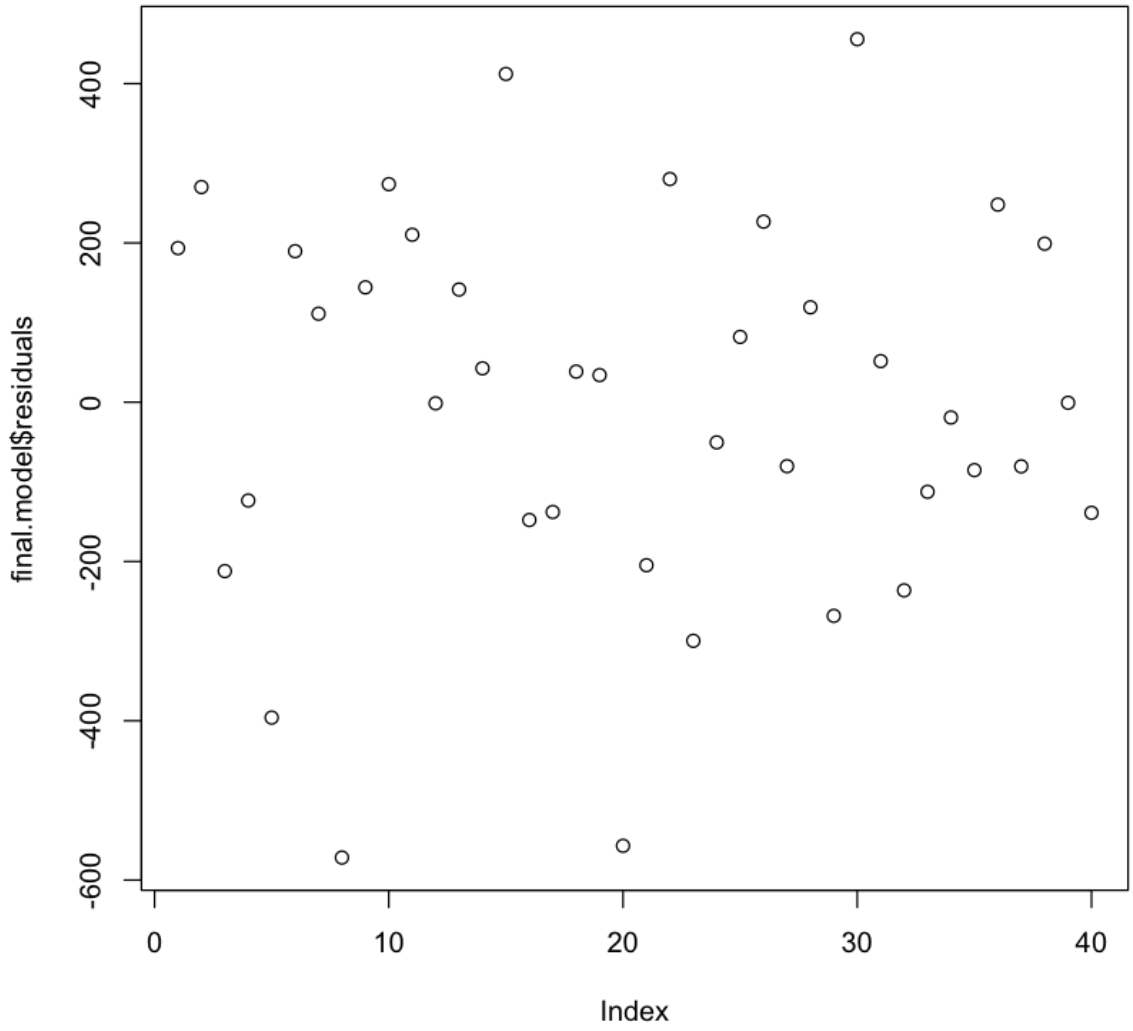
In [ ]:
```

```

In [22]:
```
final.model$residuals
```

1	193.482955103847
2	270.142481480832
3	-212.061562254153
4	-123.405856405797
5	-396.005373164189
6	189.610711808098
7	111.093253475075
8	-571.714499394502
9	144.234478379027
10	273.768203403839
11	210.282836640206
12	-1.39599268870773
13	141.517770970126
14	42.5378192248762
15	412.130946614613
16	-147.884253976745
17	-137.846844843227
18	38.4711286979419
19	34.0144479685638
20	-556.952779260608
21	-204.67736459852
22	280.208426696856
23	-299.616681963429
24	-50.4365603766034
25	81.9734260855476
26	226.747524619609
27	-80.3240474703674
28	119.14843034955
29	-268.237432438695
30	455.814682025909
31	51.5936621005788
32	-236.230804849129
33	-112.491499560325
34	-19.2001703998441
35	-85.2775708596983
36	248.107383309242
37	-80.5578291120477
38	199.02167802253
39	-0.720546114289327
40	-138.864577245988

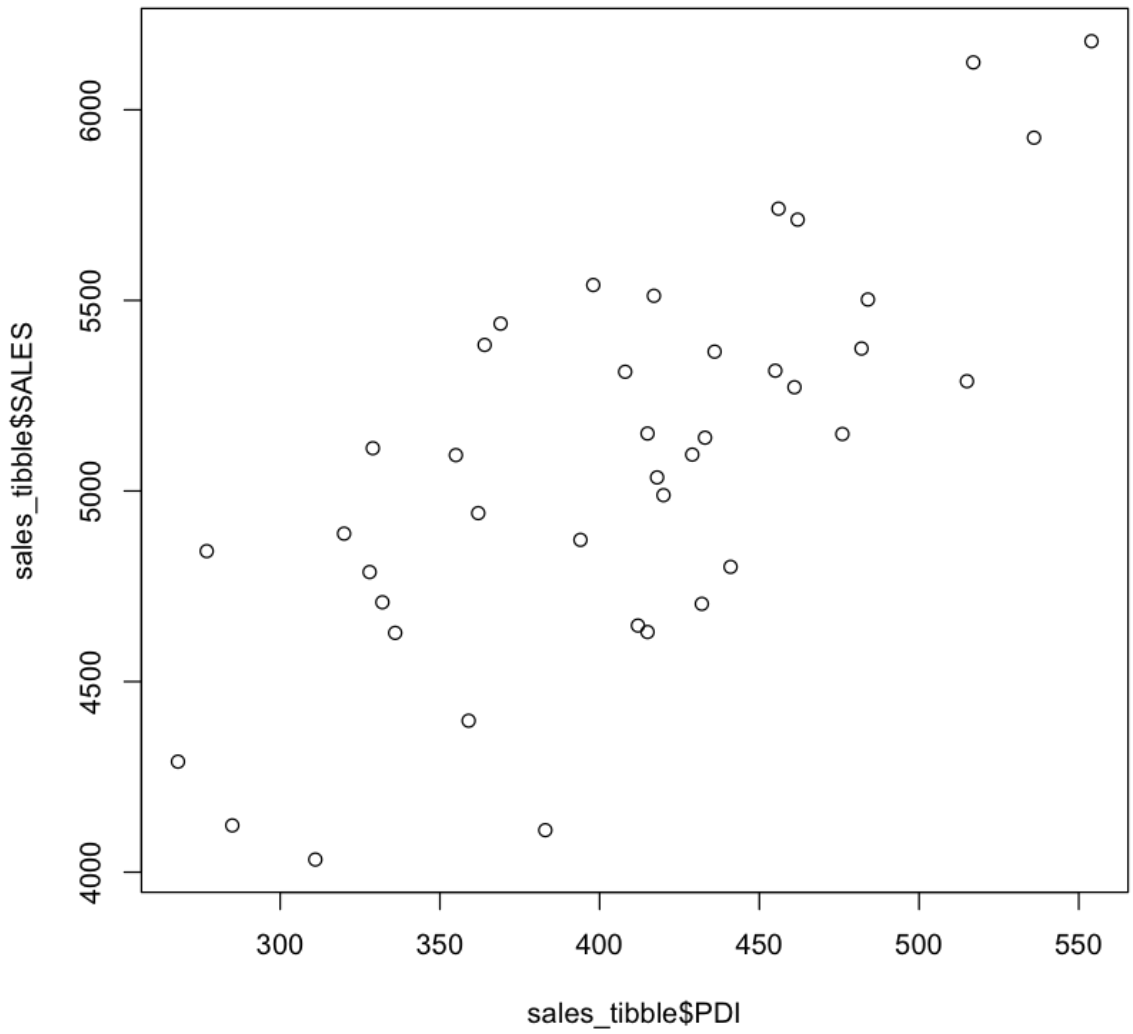
In [23]:
```
plot(final.model$residuals)
```

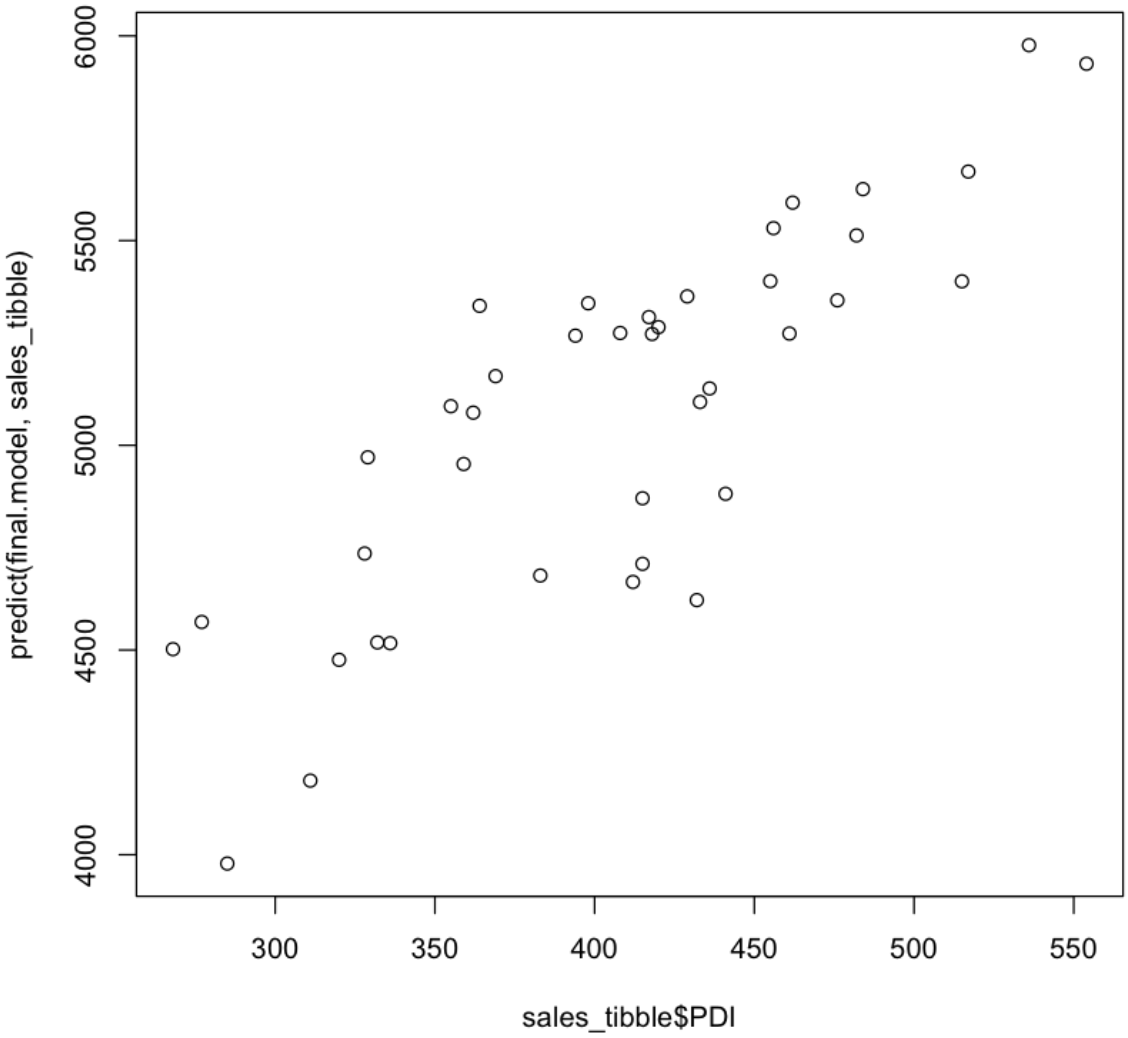


In [24]:
```
predict(final.model, sales_tibble)
```

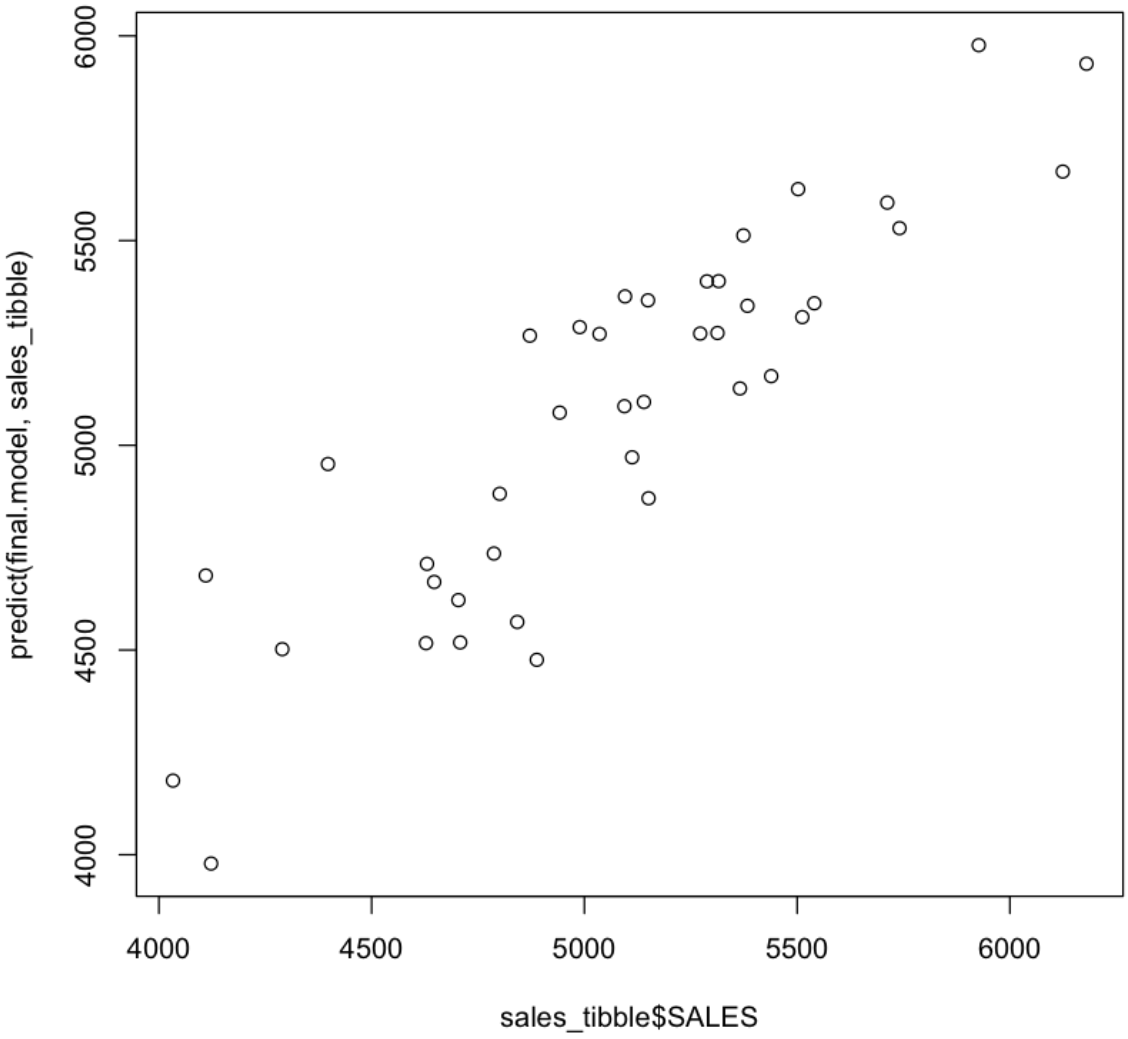
1	5346.90704489616
2	5168.89751851917
3	4502.06156225416
4	5625.7458564058
5	5267.77537316419
6	4518.4692881919
7	4516.71674652493
8	4681.95449939451
9	3978.45552162097
10	4568.48179659616
11	5530.3671633598
12	5095.49599268871
13	4970.70222902988
14	5340.66218077513
15	4476.03905338539
16	4181.01425397675
17	5079.80684484323
18	5274.32887130206
19	5105.85555203144
20	4954.31277926061
21	5354.14736459852
22	4870.62157330315
23	5288.63668196343
24	5977.29656037661
25	4621.90657391445
26	5138.84247538039
27	4710.41404747037
28	5592.71156965045
29	5363.7174324387
30	5668.55531797409
31	4735.74633789942
32	5271.85080484913
33	5400.50149956033
34	4666.21017039985
35	5400.9075708597
36	5931.95261669076
37	4881.52782911205
38	5313.10832197747
39	5272.93054611429
40	5512.63457724599

In [25]:
```
plot(x=sales_tibble$PDI, y=sales_tibble$SALES)
plot(x=sales_tibble$PDI, y=predict(final.model, sales_tibble))
```


sales_tibble$SALES
4000
4500
5000
5500
6000
300
350
400
450
500
550
sales_tibble$PDI


predict(final.model, sales_tibble)
6000
5500
5000
4500
4000
300
350
400
450
500
550
sales_tibble$PDI

In [26]:
```
plot(sales_tibble$SALES, predict(final.model, sales_tibble))
```



In [27]:
```
dwtest(final.model)
```

```
        Durbin-Watson test

data:  final.model
DW = 2.222, p-value = 0.6875
alternative hypothesis: true autocorrelation is greater than 0
```

In [28]:
```
dw_stat <- dwtest(final.model)
```

In [29]:
```
str(dw_stat)
```

```
List of 5
 $ statistic  : Named num 2.22
  ..- attr(*, "names")= chr "DW"
 $ method     : chr "Durbin-Watson test"
 $ alternative: chr "true autocorrelation is greater than 0"
 $ p.value    : num 0.687
 $ data.name  : chr "final.model"
 - attr(*, "class")= chr "htest"
```

In [30]:
```
summary(dw_stat)
```

```
            Length Class  Mode
statistic   1      -none- numeric
method      1      -none- character
alternative 1      -none- character
p.value     1      -none- numeric
data.name   1      -none- character
```

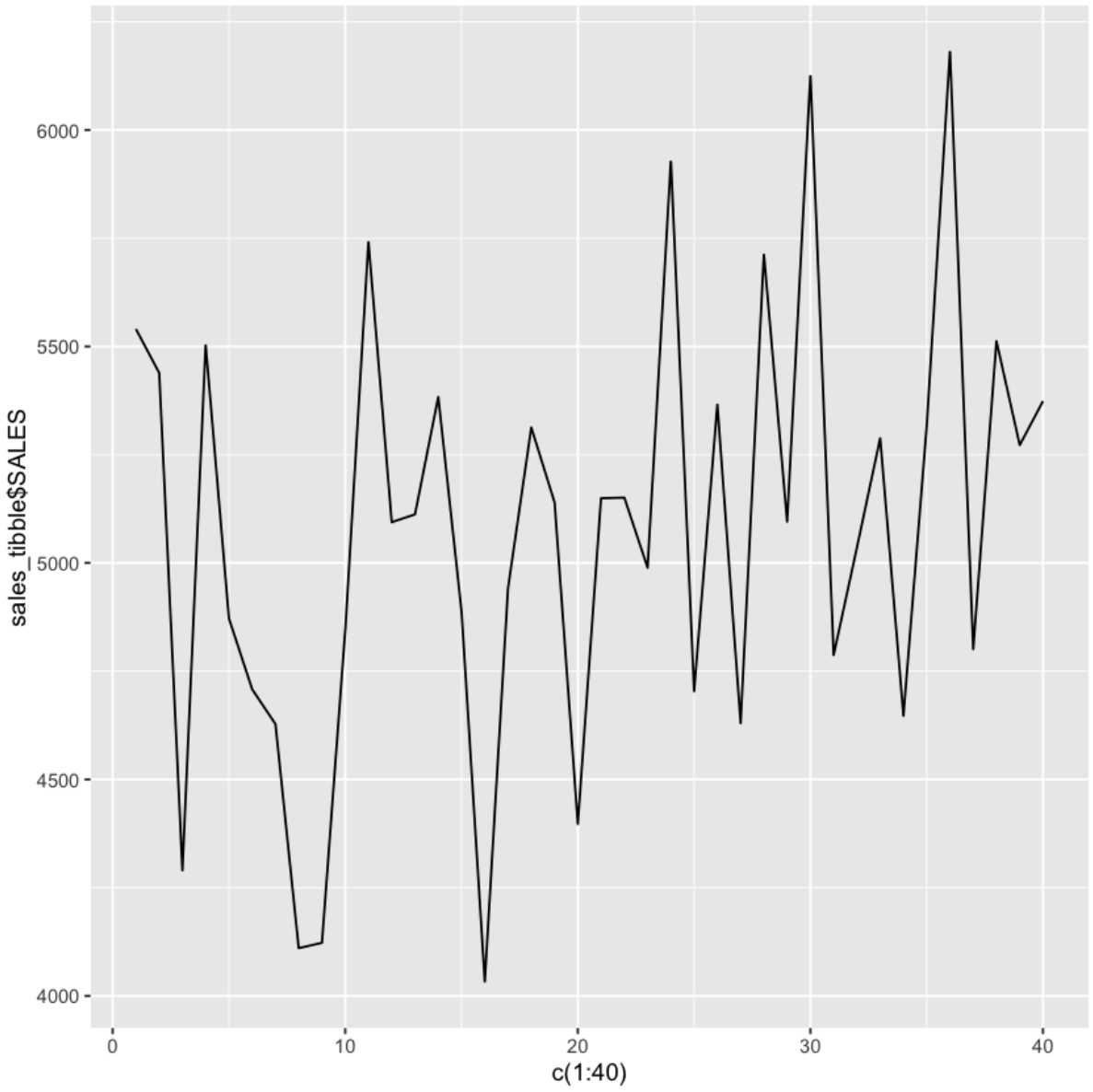
In [33]:
```
length(c(1971:1991))
```

```
21
```

In [36]:
```
nrow(sales_tibble)
```

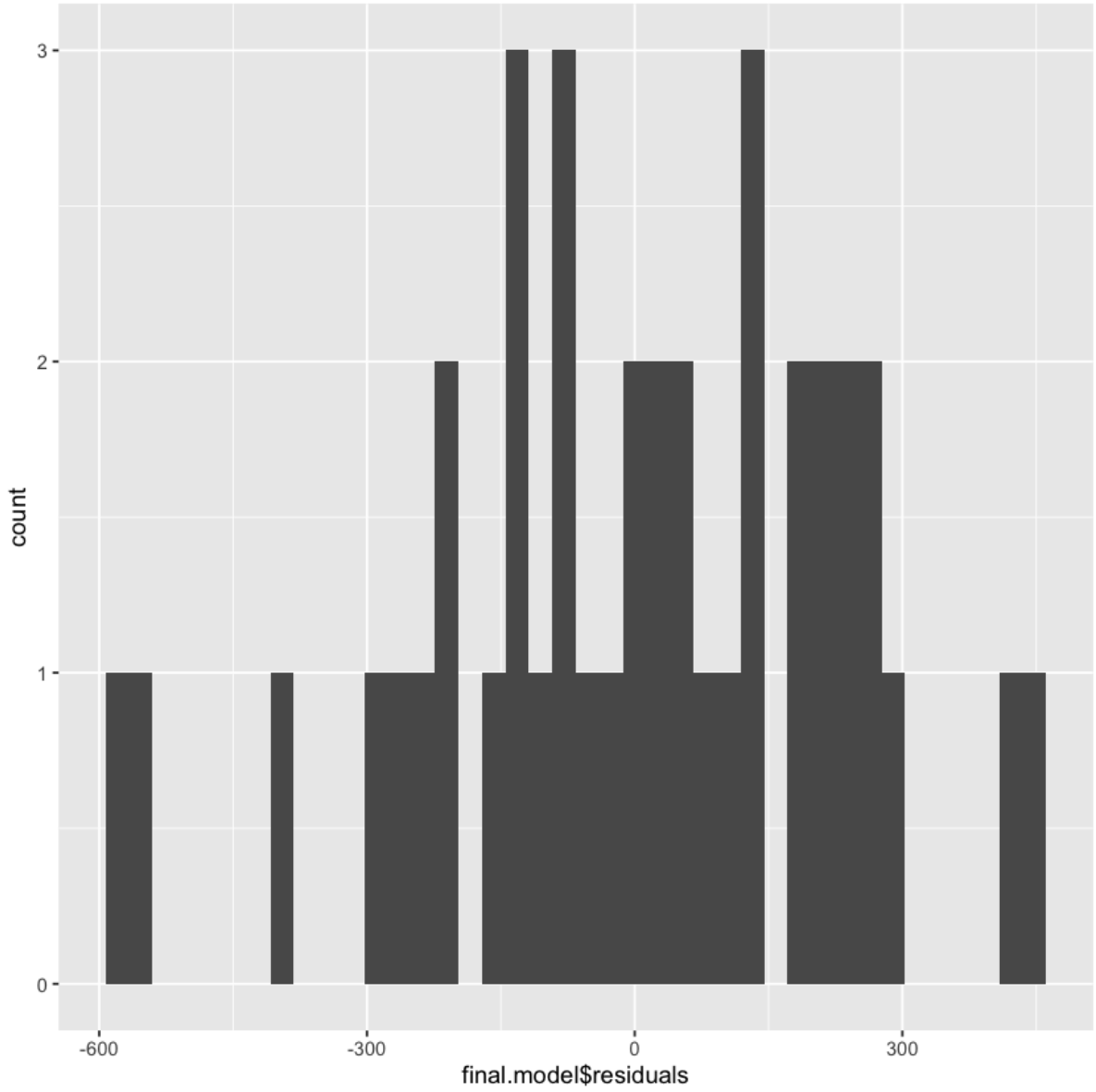
```
40
```

In [40]:
```
ggplot(data=sales_tibble, aes(x=c(1:40), y=sales_tibble$SALES)) + geom_line()
```



Histograms of Residuals against Normal KDE

In [45]:
```
df <- data.frame(PF = 10*rnorm(1000))
ggplot(data=sales_tibble, aes(final.model$residuals)) +
geom_histogram(bins=40)
```

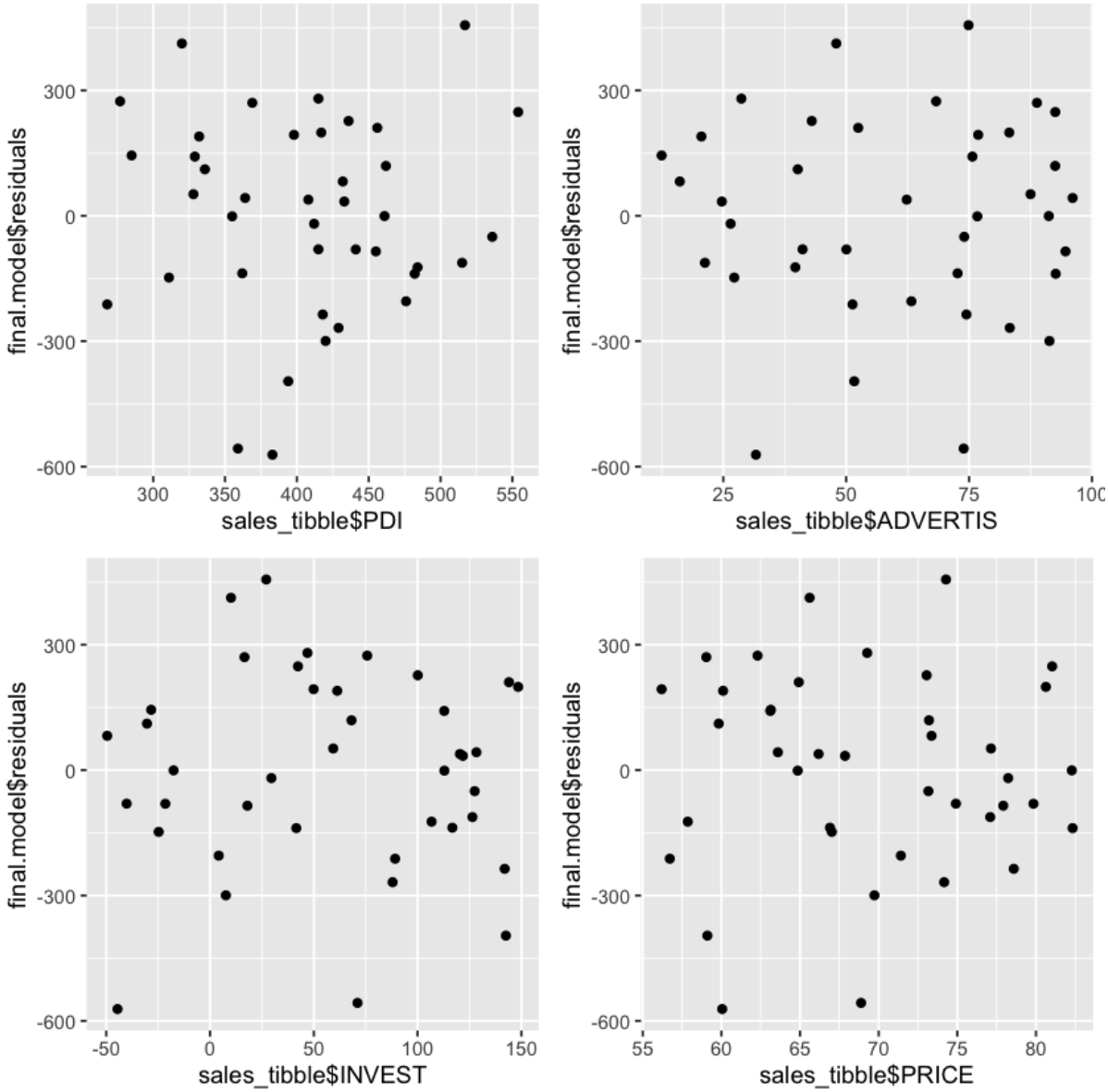


Plot residuals against:

- PDI
- Advertising
- Investment
- Price

In [52]:
```
g1 <- ggplot(data=sales_tibble, aes(x=sales_tibble$PDI, y=final.model$residuals)) + geom_point()
g2 <- ggplot(data=sales_tibble, aes(x=sales_tibble$ADVERTIS, y=final.model$residuals)) + geom_point()
g3 <- ggplot(data=sales_tibble, aes(x=sales_tibble$INVEST, y=final.model$residuals)) + geom_point()
g4 <- ggplot(data=sales_tibble, aes(x=sales_tibble$PRICE, y=final.model$residuals)) + geom_point()
```

In [53]:
```
grid.arrange(g1, g2, g3, g4)
```



In [ ]:
```

```