In [2]:
```
install.packages('MASS')
```

```
package 'MASS' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
	C:\Users\hp\AppData\Local\Temp\RtmpuKlUZd\downloaded_packages
```

In [3]:
```
library(MASS)
```

```
Warning message:
"package 'MASS' was built under R version 3.6.3"
```

In [7]:
```
data(Boston)
```

In [8]:
```
Boston
```

crim	zn	indus	chas	nox	rm	age	dis	rad	tax	ptratio	black	lstat	medv
0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
0.02985	0.0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7
0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311	15.2	395.60	12.43	22.9
0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311	15.2	396.90	19.15	27.1
0.21124	12.5	7.87	0	0.524	5.631	100.0	6.0821	5	311	15.2	386.63	29.93	16.5
0.17004	12.5	7.87	0	0.524	6.004	85.9	6.5921	5	311	15.2	386.71	17.10	18.9
0.22489	12.5	7.87	0	0.524	6.377	94.3	6.3467	5	311	15.2	392.52	20.45	15.0
0.11747	12.5	7.87	0	0.524	6.009	82.9	6.2267	5	311	15.2	396.90	13.27	18.9
0.09378	12.5	7.87	0	0.524	5.889	39.0	5.4509	5	311	15.2	390.50	15.71	21.7
0.62976	0.0	8.14	0	0.538	5.949	61.8	4.7075	4	307	21.0	396.90	8.26	20.4
0.63796	0.0	8.14	0	0.538	6.096	84.5	4.4619	4	307	21.0	380.02	10.26	18.2
0.62739	0.0	8.14	0	0.538	5.834	56.5	4.4986	4	307	21.0	395.62	8.47	19.9

In [9]:
```
is.null('Boston')
```

```
FALSE
```

In [12]:
```
summary(Boston)
```

```
     crim                zn              indus            chas        
 Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000  
 1st Qu.: 0.08204   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000  
 Median : 0.25651   Median :  0.00   Median : 9.69   Median :0.00000  
 Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917  
 3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000  
 Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000  
      nox               rm              age              dis         
 Min.   :0.3850   Min.   :3.561   Min.   :  2.90   Min.   : 1.130  
 1st Qu.:0.4490   1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100  
 Median :0.5380   Median :6.208   Median : 77.50   Median : 3.207  
 Mean   :0.5547   Mean   :6.285   Mean   : 68.57   Mean   : 3.795  
 3rd Qu.:0.6240   3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188  
 Max.   :0.8710   Max.   :8.780   Max.   :100.00   Max.   :12.127  
      rad              tax            ptratio          black       
 Min.   : 1.000   Min.   :187.0   Min.   :12.60   Min.   :  0.32  
 1st Qu.: 4.000   1st Qu.:279.0   1st Qu.:17.40   1st Qu.:375.38  
 Median : 5.000   Median :330.0   Median :19.05   Median :391.44  
 Mean   : 9.549   Mean   :408.2   Mean   :18.46   Mean   :356.67  
 3rd Qu.:24.000   3rd Qu.:666.0   3rd Qu.:20.20   3rd Qu.:396.23  
 Max.   :24.000   Max.   :711.0   Max.   :22.00   Max.   :396.90  
     lstat            medv      
 Min.   : 1.73   Min.   : 5.00  
 1st Qu.: 6.95   1st Qu.:17.02  
 Median :11.36   Median :21.20  
 Mean   :12.65   Mean   :22.53  
 3rd Qu.:16.95   3rd Qu.:25.00  
 Max.   :37.97   Max.   :50.00  
```

In [20]:
```
attach(Boston)
```

```
The following object is masked _by_ .GlobalEnv:

    medv
```

In [21]:
```
cor(lstat,medv)
```

```
-0.737662726174015
```

In [22]:
```
cor(medv, Boston)
```

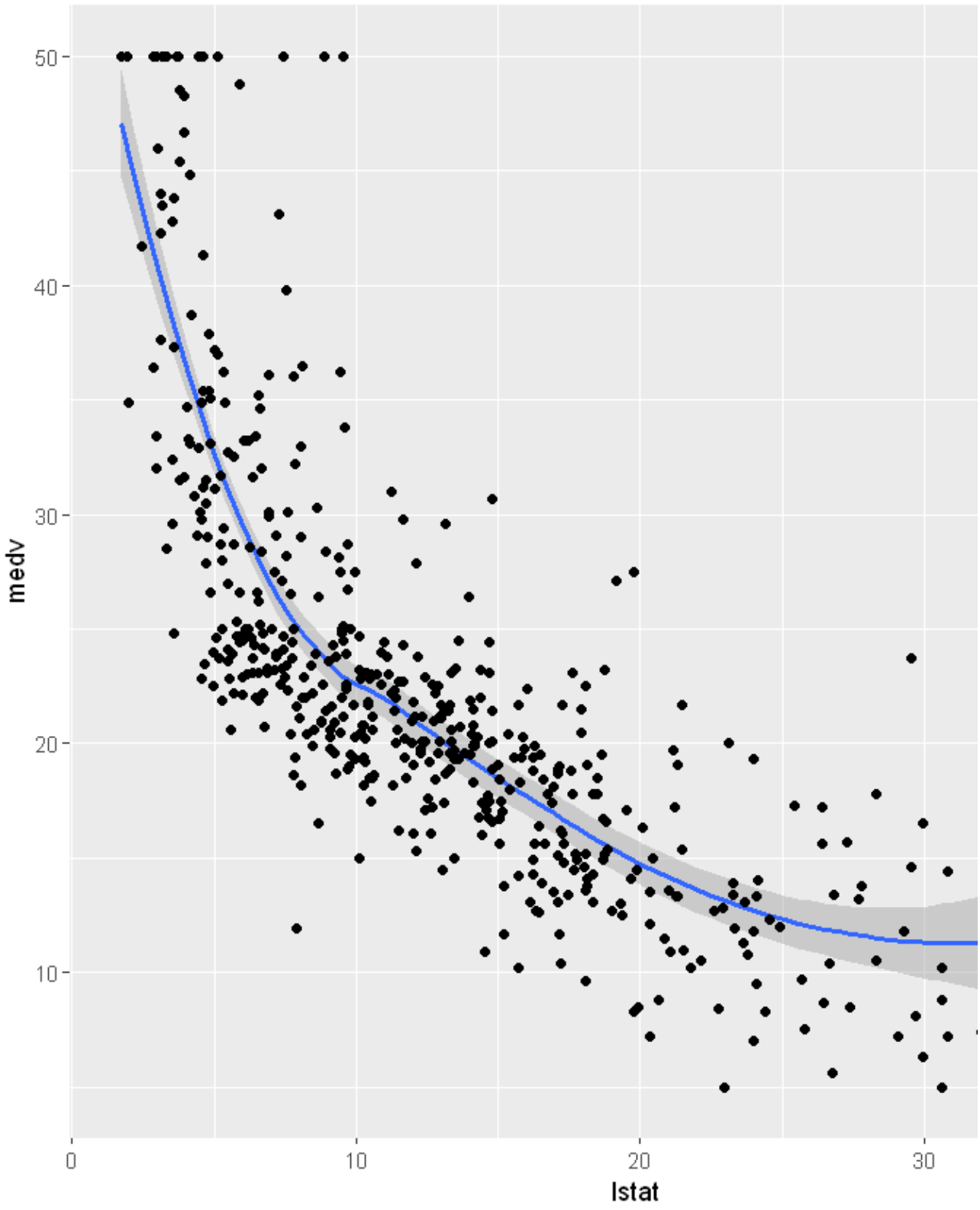
	indus	chas	nox	rm	age	dis	rad	tax	ptratio
4453	-0.4837252	0.1752602	-0.4273208	0.6953599	-0.3769546	0.2499287	-0.3816262	-0.4685359	-0.5077

In [15]:
```
1 library(ggplot2)
```

```
Warning message:
"package 'ggplot2' was built under R version 3.6.3"
```

In [27]:
```
1 ggplot(Boston,aes(x = lstat, y = medv)) + stat_smooth() + geom_point()
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



In [31]:

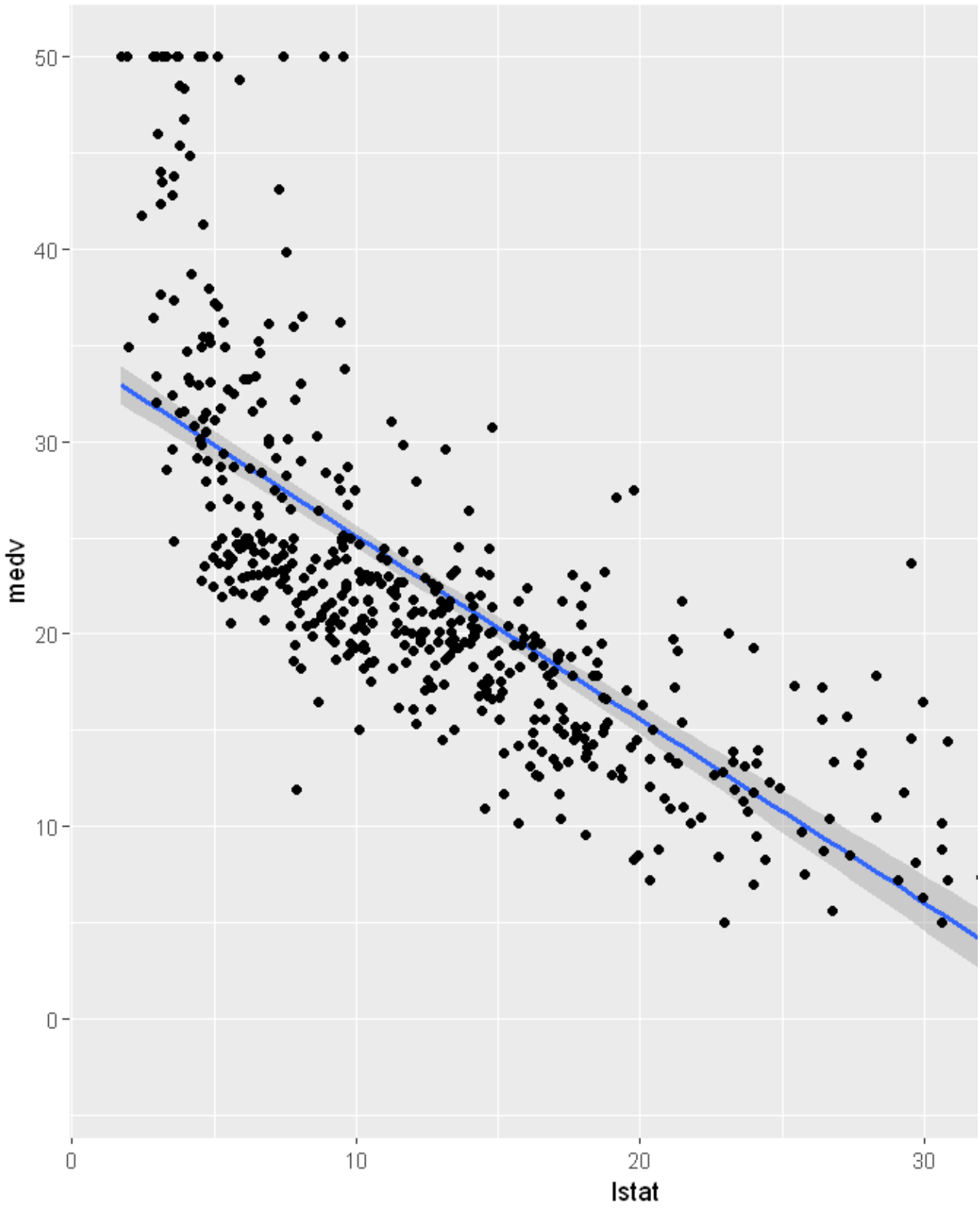
```
model <- lm(medv ~lstat, data= Boston)
model
```

```
Call:
lm(formula = medv ~ lstat, data = Boston)

Coefficients:
(Intercept)        lstat
      34.55        -0.95
```

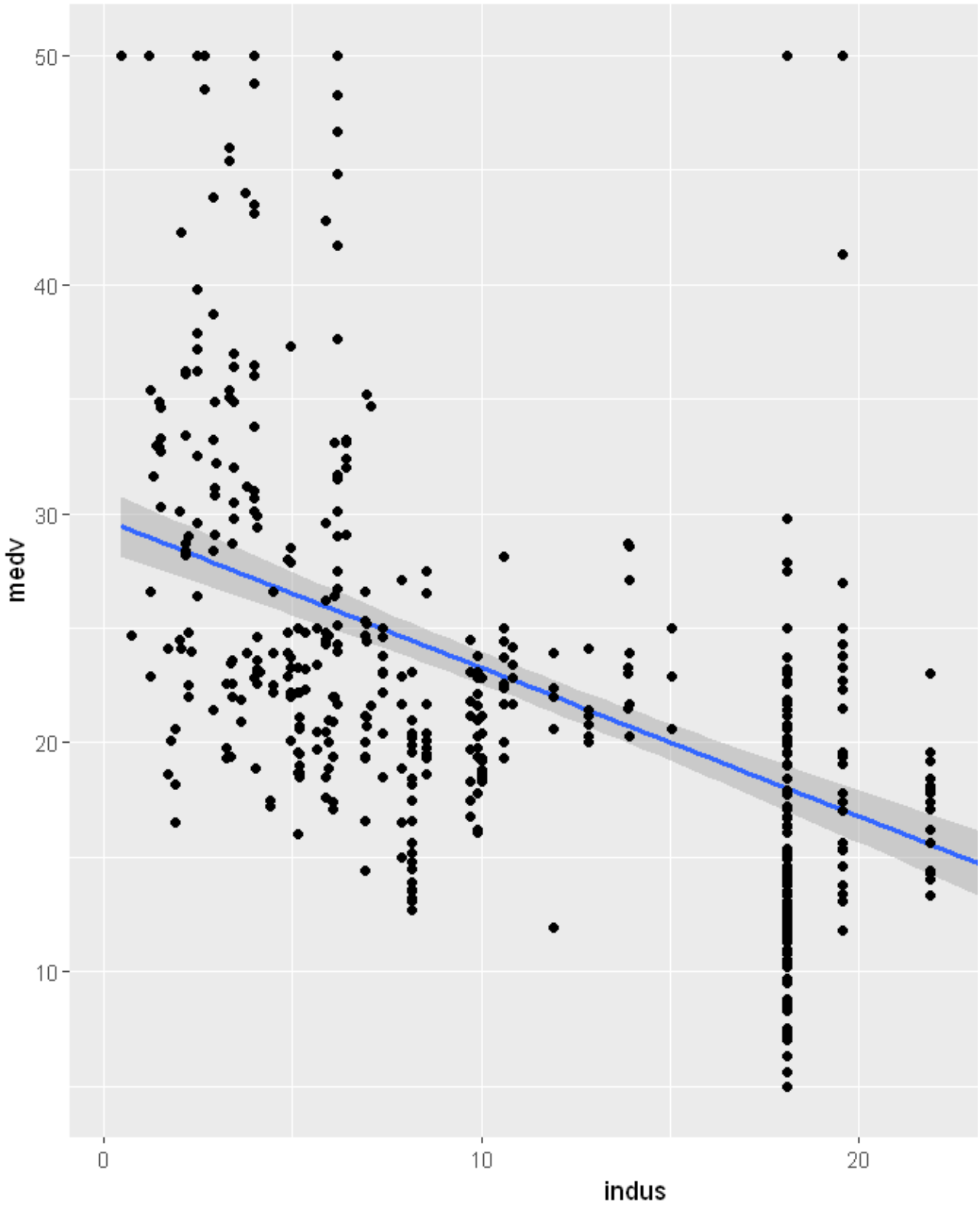
In [32]:
```
ggplot( Boston , aes(lstat, medv)) + stat_smooth(method = lm) + geom_point()
```

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



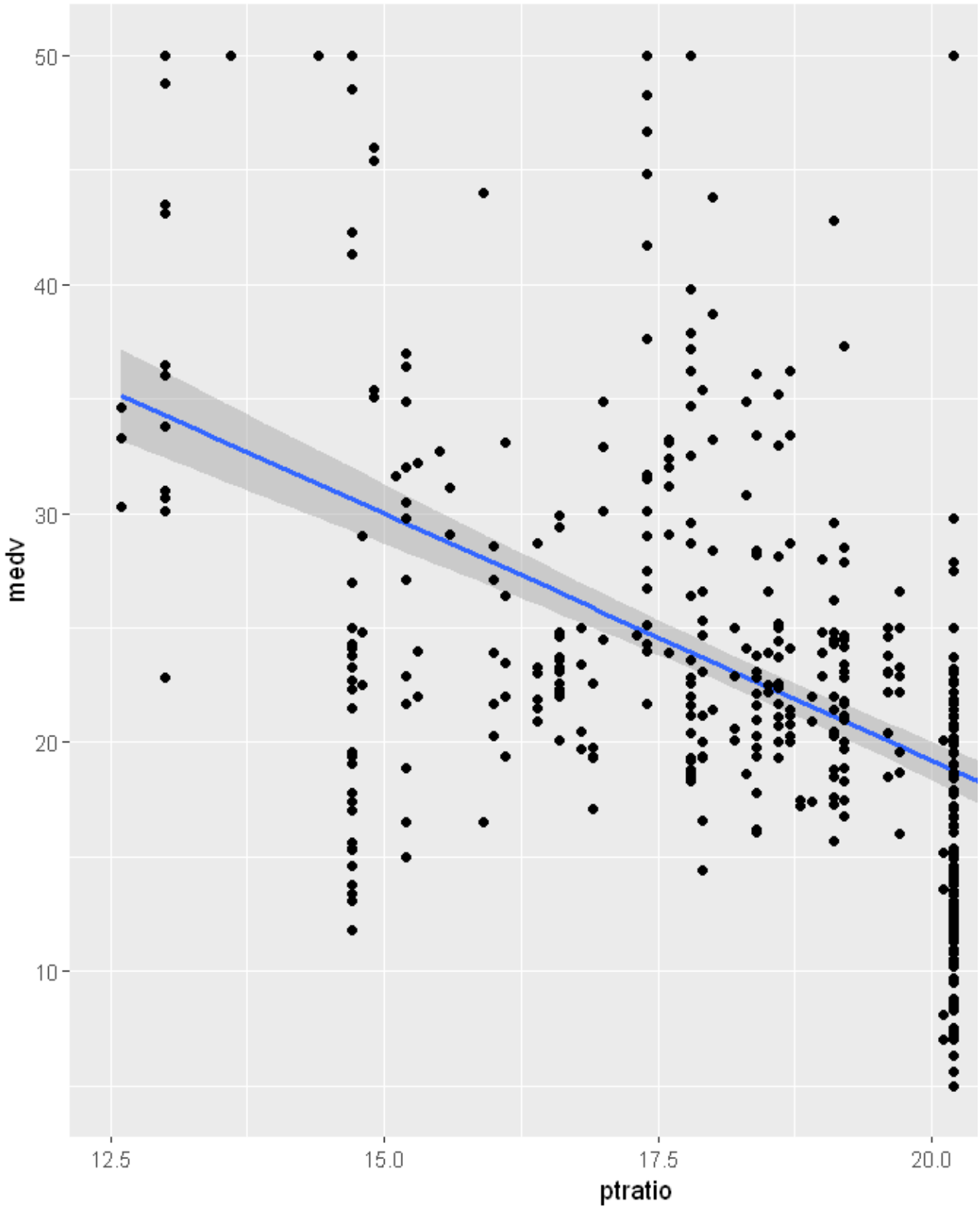
In [38]:
```
ggplot( Boston , aes(indus, medv)) + stat_smooth(method = lm) + geom_point()
```

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



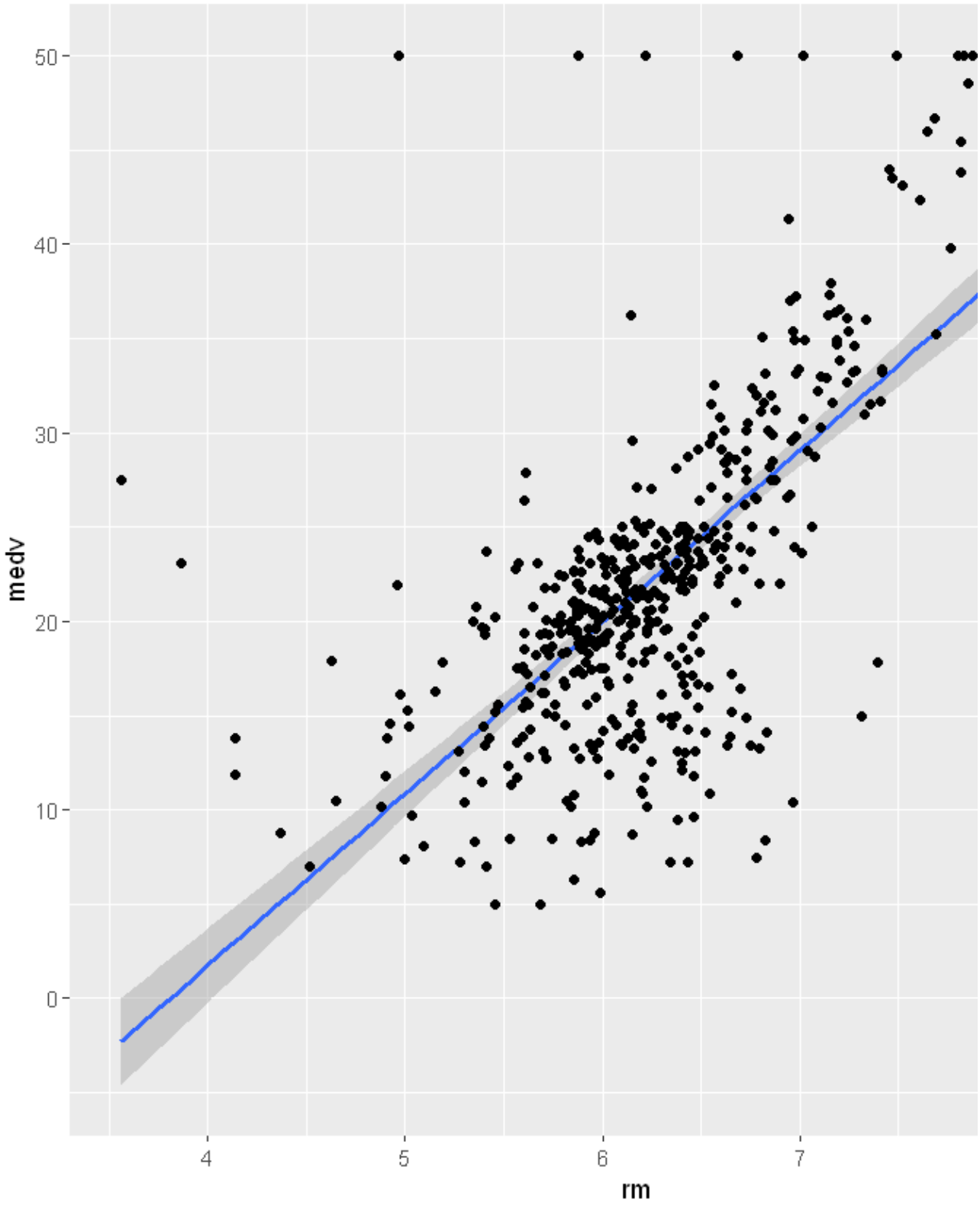
In [39]:
```
ggplot( Boston , aes(ptratio, medv)) + stat_smooth(method = lm) + geom_point
```

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



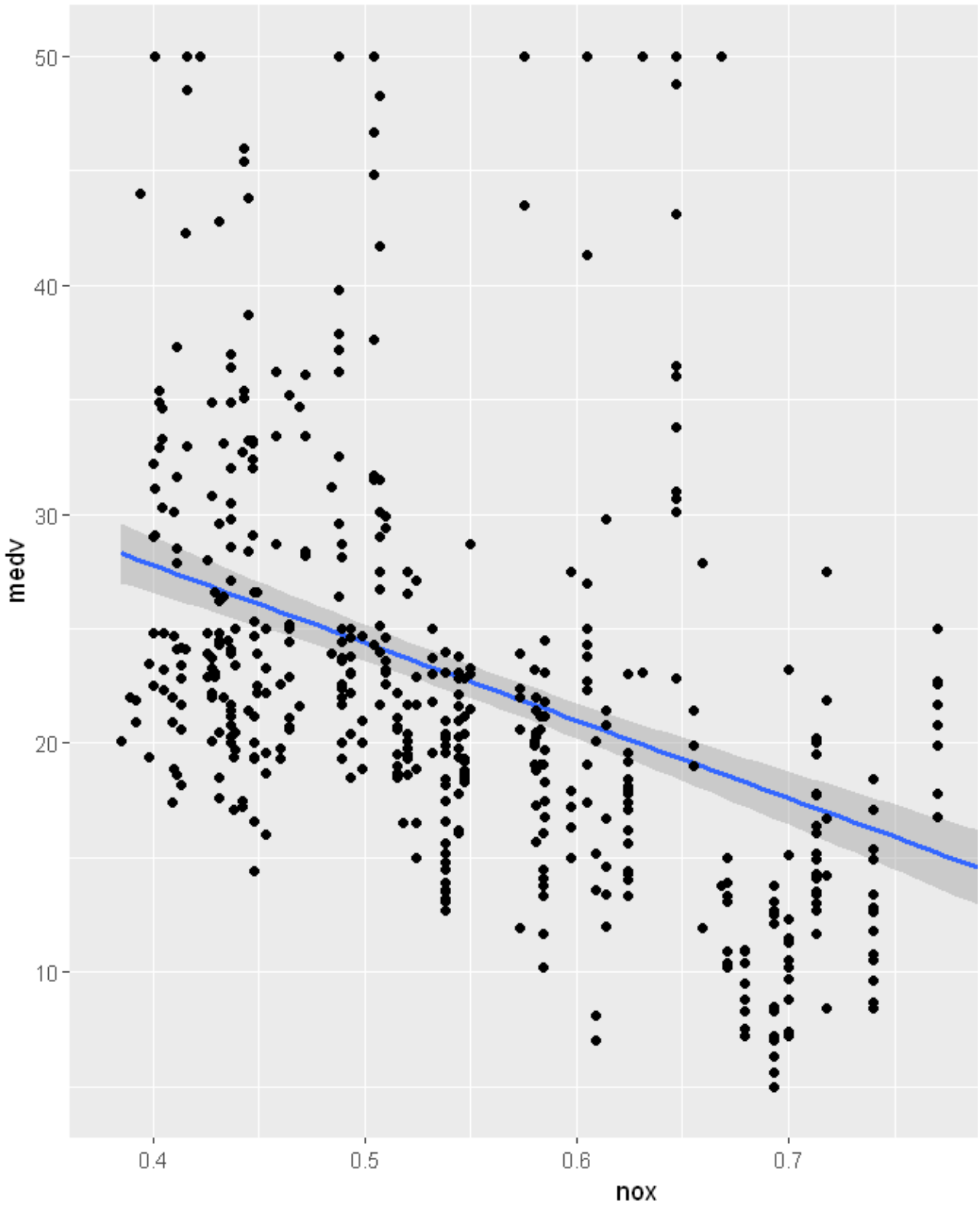
In [40]:
```
1 ggplot( Boston , aes(rm, medv)) + stat_smooth(method = lm) + geom_point()
```

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



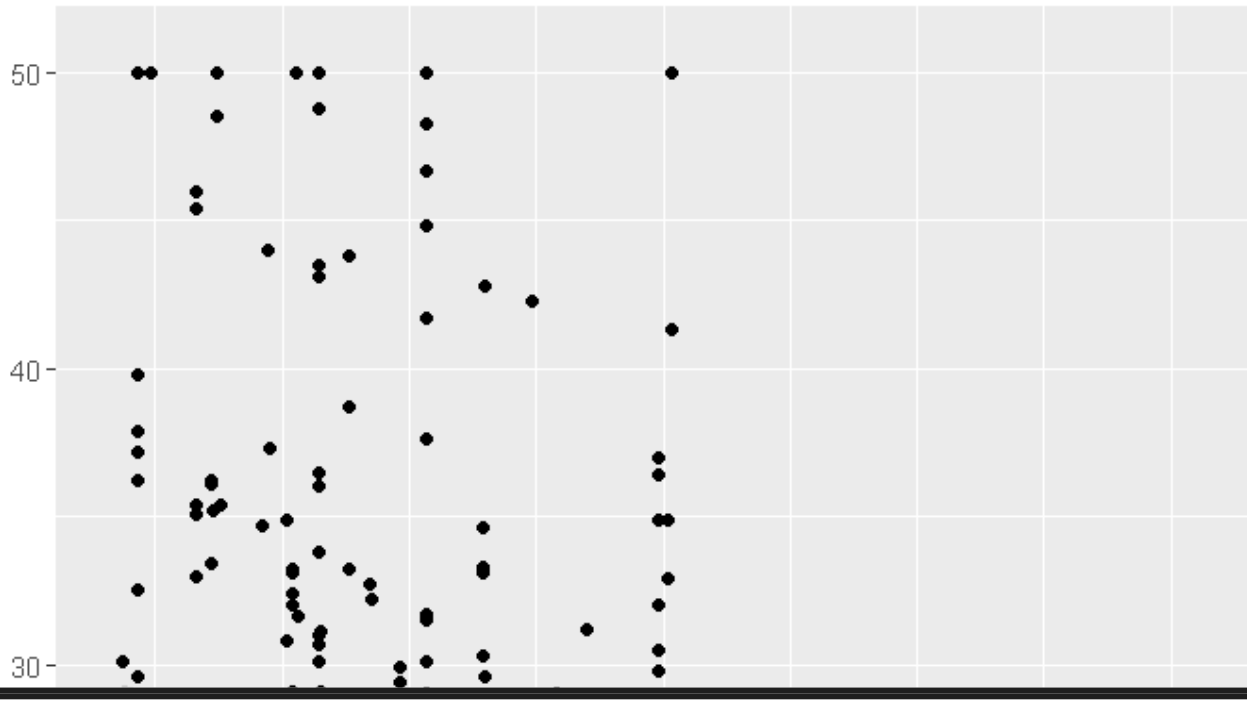
In [41]:
```
1 ggplot( Boston , aes(nox, medv)) + stat_smooth(method = lm) + geom_point()
```

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



In [42]:
```
ggplot( Boston , aes(tax, medv)) + stat_smooth(method = lm) + geom_point()
```

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



In [47]:
```
model2 <- lm(ptratio ~lstat, data= Boston)
model2
```

```
Call:
lm(formula = ptratio ~ lstat, data = Boston)

Coefficients:
(Intercept)        lstat
    17.0207       0.1134
```

In [46]:
```
model3 <- lm(rm ~lstat, data= Boston)
model3
```

```
Call:
lm(formula = rm ~ lstat, data = Boston)

Coefficients:
(Intercept)        lstat
    7.04879     -0.06039
```

In [45]:
```
model4 <- lm(nox ~lstat, data= Boston)
model4
```

```
Call:
lm(formula = nox ~ lstat, data = Boston)

Coefficients:
(Intercept)        lstat
   0.433375     0.009588
```

In [44]:
```
model5 <- lm(indus ~lstat, data= Boston)
model5
```

```
Call:
lm(formula = indus ~ lstat, data = Boston)

Coefficients:
(Intercept)        lstat
     3.7972       0.5801
```

In [43]:
```
model6 <- lm(tax ~lstat, data= Boston)
model6
```

```
Call:
lm(formula = tax ~ lstat, data = Boston)

Coefficients:
(Intercept)        lstat
     245.79        12.84
```

In [33]:
```
summary(model)
```

```
Call:
lm(formula = medv ~ lstat, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.168  -3.990  -1.318   2.034  24.500 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 34.55384    0.56263   61.41   <2e-16 ***
lstat       -0.95005    0.03873  -24.53   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.216 on 504 degrees of freedom
Multiple R-squared:  0.5441,    Adjusted R-squared:  0.5432 
F-statistic: 601.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

In [48]:
```
summary(model2)
```

```
Call:
lm(formula = ptratio ~ lstat, data = Boston)

Residuals:
   Min     1Q Median     3Q    Max 
-5.698 -1.014  0.430  1.415  4.348 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 17.02070    0.18192  93.562   <2e-16 ***
lstat        0.11340    0.01252   9.055   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.01 on 504 degrees of freedom
Multiple R-squared:  0.1399,    Adjusted R-squared:  0.1382 
F-statistic: 81.98 on 1 and 504 DF,  p-value: < 2.2e-16
```

In [49]:
```
summary(model3)
```

```
Call:
lm(formula = rm ~ lstat, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0578 -0.3413 -0.0712  0.2505  2.0507 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.04880    0.05026  140.25   <2e-16 ***
lstat       -0.06039    0.00346  -17.45   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5552 on 504 degrees of freedom
Multiple R-squared:  0.3768,    Adjusted R-squared:  0.3755 
F-statistic: 304.7 on 1 and 504 DF,  p-value: < 2.2e-16
```

In [50]:
```
summary(model4)
```

```
Call:
lm(formula = nox ~ lstat, data = Boston)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.28079 -0.06262 -0.01816  0.04744  0.36677 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.4333754  0.0084703   51.16   <2e-16 ***
lstat       0.0095882  0.0005831   16.44   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09358 on 504 degrees of freedom
Multiple R-squared:  0.3491,    Adjusted R-squared:  0.3478 
F-statistic: 270.4 on 1 and 504 DF,  p-value: < 2.2e-16
```

In [51]:
```
summary(model5)
```

```
Call:
lm(formula = indus ~ lstat, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-14.759  -4.033  -1.223   3.949  16.199

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.79718    0.49549   7.663 9.36e-14 ***
lstat        0.58006    0.03411  17.005  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.474 on 504 degrees of freedom
Multiple R-squared:  0.3646,    Adjusted R-squared:  0.3633
F-statistic: 289.2 on 1 and 504 DF,  p-value: < 2.2e-16
```

In [52]:
```
summary(model6)
```

```
Call:
lm(formula = tax ~ lstat, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-408.35  -90.19  -29.77   91.63  382.21

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 245.7863    12.8133   19.18   <2e-16 ***
lstat        12.8389     0.8821   14.55   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 141.6 on 504 degrees of freedom
Multiple R-squared:  0.2959,    Adjusted R-squared:  0.2945
F-statistic: 211.8 on 1 and 504 DF,  p-value: < 2.2e-16
```

In [34]:
```
confint(model)
```

	2.5 %	97.5 %
(Intercept)	33.448457	35.6592247
lstat	-1.026148	-0.8739505

In [35]:
```
cor(Boston)
```

	crim	zn	indus	chas	nox	rm	age	dis
crim	1.00000000	-0.20046922	0.40658341	-0.055891582	0.42097171	-0.21924670	0.35273425	-0.3
zn	-0.20046922	1.00000000	-0.53382819	-0.042696719	-0.51660371	0.31199059	-0.56953734	0.6
indus	0.40658341	-0.53382819	1.00000000	0.062938027	0.76365145	-0.39167585	0.64477851	-0.7
chas	-0.05589158	-0.04269672	0.06293803	1.000000000	0.09120281	0.09125123	0.08651777	-0.0
nox	0.42097171	-0.51660371	0.76365145	0.091202807	1.00000000	-0.30218819	0.73147010	-0.7
rm	-0.21924670	0.31199059	-0.39167585	0.091251225	-0.30218819	1.00000000	-0.24026493	0.2
age	0.35273425	-0.56953734	0.64477851	0.086517774	0.73147010	-0.24026493	1.00000000	-0.7
dis	-0.37967009	0.66440822	-0.70802699	-0.099175780	-0.76923011	0.20524621	-0.74788054	1.0
rad	0.62550515	-0.31194783	0.59512927	-0.007368241	0.61144056	-0.20984667	0.45602245	-0.4
tax	0.58276431	-0.31456332	0.72076018	-0.035586518	0.66802320	-0.29204783	0.50645559	-0.5
ptratio	0.28994558	-0.39167855	0.38324756	-0.121515174	0.18893268	-0.35550149	0.26151501	-0.2
black	-0.38506394	0.17552032	-0.35697654	0.048788485	-0.38005064	0.12806864	-0.27353398	0.2
lstat	0.45562148	-0.41299457	0.60379972	-0.053929298	0.59087892	-0.61380827	0.60233853	-0.4
medv	-0.38830461	0.36044534	-0.48372516	0.175260177	-0.42732077	0.69535995	-0.37695457	0.2

In [4]:
```
insurance = read.csv("insurance.csv")
insurance
```

age	sex	bmi	children	smoker	region	charges
19	female	27.900	0	yes	southwest	16884.924
18	male	33.770	1	no	southeast	1725.552
28	male	33.000	3	no	southeast	4449.462
33	male	22.705	0	no	northwest	21984.471
32	male	28.880	0	no	northwest	3866.855
31	female	25.740	0	no	southeast	3756.622
46	female	33.440	1	no	southeast	8240.590
37	female	27.740	3	no	northwest	7281.506
37	male	29.830	2	no	northeast	6406.411
60	female	25.840	0	no	northwest	28923.137
25	male	26.220	0	no	northeast	2721.321
62	female	26.290	0	yes	southeast	27808.725
23	male	34.400	0	no	southwest	1826.843
56	female	39.820	0	no	southeast	11090.718
27	male	42.130	0	yes	southeast	39611.758
19	male	24.600	1	no	southwest	1837.237

In [5]:
```
1 is.null('insurance')
```

```
FALSE
```

In [55]:
```
1 attach(insurance)
```

```
The following object is masked from Boston:

    age
```

In [60]:
```
1 cor(age,charges)
```

```
0.299008193330648
```

In [61]:
```
1 str(insurance)
```

```
'data.frame':   1338 obs. of  7 variables:
 $ age     : int  19 18 28 33 32 31 46 37 37 60 ...
 $ sex     : Factor w/ 2 levels "female","male": 1 2 2 2 2 1 1 1 2 1 ...
 $ bmi     : num  27.9 33.8 33 22.7 28.9 ...
 $ children: int  0 1 3 0 0 0 1 3 2 0 ...
 $ smoker  : Factor w/ 2 levels "no","yes": 2 1 1 1 1 1 1 1 1 1 ...
 $ region  : Factor w/ 4 levels "northeast","northwest",..: 4 3 3 2 2 3 3 2 1 2 ...
 $ charges : num  16885 1726 4449 21984 3867 ...
```

In [63]:
```
1 names(insurance)
```

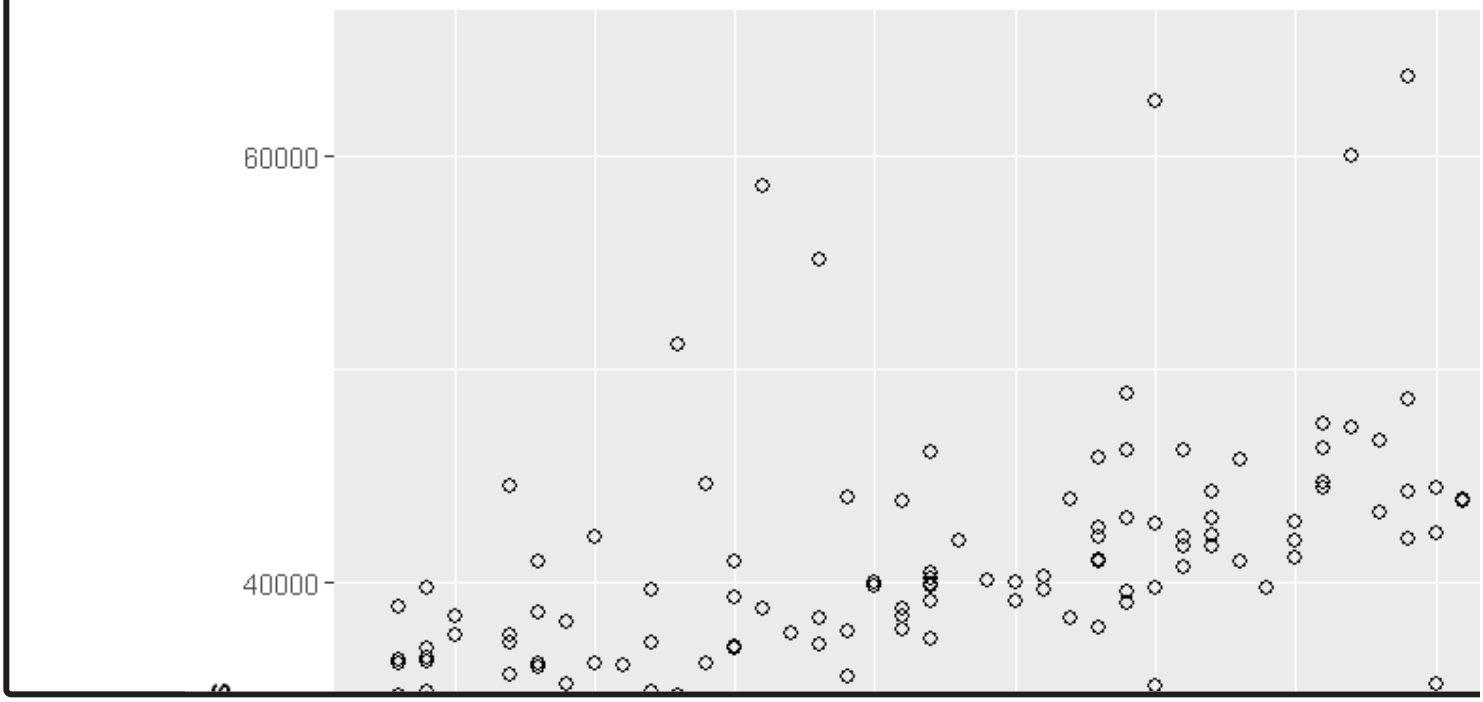
```
'age'   'sex'   'bmi'   'children'   'smoker'   'region'   'charges'
```

In [65]:
```
1 cor(charges,bmi)
```

```
0.198340968833629
```

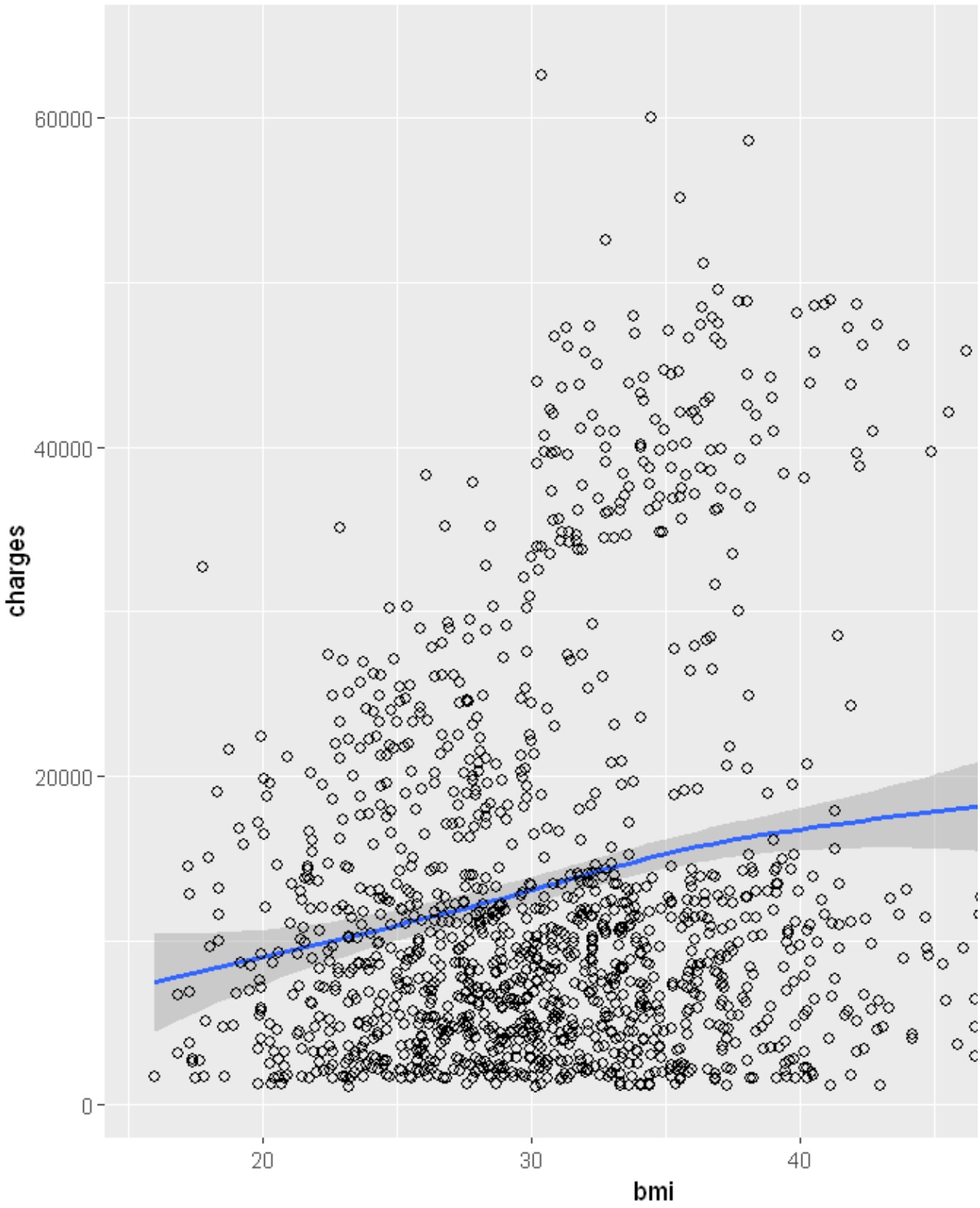
In [68]:
```
1 ggplot(insurance, aes(x = age, y = charges)) + stat_smooth() + geom_point(sha
```

```
`geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



In [73]:
```
1 ggplot(insurance, aes(x = bmi, y = charges)) + stat_smooth() + geom_point(sha
```

```
`geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



In [69]:

```
model7 = lm(age~charges , data = insurance)
model7
```

```
Call:
lm(formula = age ~ charges, data = insurance)

Coefficients:
(Intercept)      charges
  3.460e+01    3.469e-04
```

In [70]:
```
summary(model7)
```

```
Call:
lm(formula = age ~ charges, data = insurance)

Residuals:
     Min       1Q   Median       3Q      Max 
-30.0609 -11.4222   0.1691  11.1759  24.6013 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.460e+01  5.441e-01   63.60   <2e-16 ***
charges     3.469e-04  3.029e-05   11.45   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.41 on 1336 degrees of freedom
Multiple R-squared:  0.08941,   Adjusted R-squared:  0.08872 
F-statistic: 131.2 on 1 and 1336 DF,  p-value: < 2.2e-16
```

In [71]:
```
model8 = lm(bmi~charges , data = insurance)
model8
```

```
Call:
lm(formula = bmi ~ charges, data = insurance)

Coefficients:
(Intercept)      charges  
  2.934e+01    9.988e-05  
```

In [72]:

```
1 summary(model8)
```

```
Call:
lm(formula = bmi ~ charges, data = insurance)

Residuals:
     Min       1Q   Median       3Q      Max
-14.8424  -4.1030  -0.2401   3.8467  23.6758

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.934e+01  2.426e-01 120.956  < 2e-16 ***
charges     9.988e-05  1.350e-05   7.397 2.46e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.979 on 1336 degrees of freedom
Multiple R-squared:  0.03934,   Adjusted R-squared:  0.03862
F-statistic: 54.71 on 1 and 1336 DF,  p-value: 2.459e-13
```