

1

```
package 'ggsci' successfully unpacked and MD5 sums checked
package 'cowplot' successfully unpacked and MD5 sums checked
package 'ggsignif' successfully unpacked and MD5 sums checked
package 'gridExtra' successfully unpacked and MD5 sums checked
package 'polynom' successfully unpacked and MD5 sums checked
package 'rstatix' successfully unpacked and MD5 sums checked
package 'ggpubr' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Me\AppData\Local\Temp\Rtmp6liuqt\downloaded_packages
> |
```

2

```
> library(ggplot2)
Warning message:
package 'ggplot2' was built under R version 4.2.3
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

  filter, lag

The following objects are masked from 'package:base':

  intersect, setdiff, setequal, union

Warning message:
package 'dplyr' was built under R version 4.2.3
> library(broom)
Warning message:
package 'broom' was built under R version 4.2.3
> library(ggpubr)
Warning message:
package 'ggpubr' was built under R version 4.2.3
```

3

Import Dataset

Name:

Input File: 

```
"["biking", "smoking", "heart.disease"
"1", 30.8012457056902, 10.8966080237878, 11.7694227799599
"2", 65.129215165507, 2.21956317569129, 2.85408147801992
"3", 1.95966453105211, 17.5883305108873, 17.1778034757926
"4", 44.8001956157386, 2.80255887506064, 6.81664690867892
"5", 69.4284536754712, 15.9745045953896, 4.06222352243168
"6", 54.4036255460232, 29.3331755213439, 9.55004599703223
"7", 49.0561619559303, 9.06084577110596, 7.62450698041603
"8", 4.78460420155898, 12.8350208034972, 15.8546544341453
"9", 65.7307882993482, 11.9912972834427, 3.06746173004763
"10", 35.2574489386752, 23.2776834309334, 12.0984843702583
"11", 51.8255674857646, 14.4351183790714, 6.43024815108675
"12", 52.9361971635371, 25.0748685661238, 8.60827207541993
"13", 48.7674784869887, 11.0232709712582, 6.72252380072458
"14", 26.1668007872067, 6.64574954903219, 10.5978071174237
"15", 10.5530749158934, 5.99050630908459, 14.0794783264219
"16", 47.163716411218, 14.0978371595265, 8.74484530505964
"17", 61.6852562502027, 16.8408167293528, 5.44334202544307
"18", 33.9443941223435, 5.75859518698417, 9.16230641709494]
```

Encoding:

Heading: ☒ Yes ☐ No

Row names:

Separator:

Decimal:

Quote:

Comment:

na.strings:

☐ Strings as factors

Data Frame:

x	biking	smoking	heart.disease
1	30.801246	10.896608	11.769423
2	65.129215	2.219563	2.854081
3	1.959665	17.588331	17.177803
4	44.800196	2.802559	6.816647
5	69.428454	15.974505	4.062224
6	54.403626	29.333176	9.550046
7	49.056162	9.060846	7.624507
8	4.784604	12.835021	15.854654
9	65.730788	11.991297	3.067462
10	35.257449	23.277683	12.098484
11	51.825567	14.435118	6.430248
12	52.936197	25.074869	8.608272
13	48.767478	11.023271	6.722524
14	26.166801	6.645750	10.597807
15	10.553075	5.990506	14.079478
16	47.163716	14.097837	8.744845
17	61.685256	16.840817	5.443342
18	33.944394	5.758595	9.162306

```

      x      biking      smoking
Min.   : 1.0      Min.   : 1.119   Min.   : 0.5259
1st Qu.:125.2    1st Qu.:20.205   1st Qu.: 8.2798
Median :249.5    Median :35.824   Median :15.8146
Mean   :249.5    Mean   :37.788   Mean   :15.4350
3rd Qu.:373.8    3rd Qu.:57.853   3rd Qu.:22.5689
Max.   :498.0    Max.   :74.907   Max.   :29.9467
heart.disease
Min.   : 0.5519
1st Qu.: 6.5137
Median :10.3853
Mean   :10.1745
3rd Qu.:13.7240
Max.   :20.4535
> |

```

```
> summary(lm(heart.disease ~ biking, data = heart.data))
```

```
Call:
```

```
lm(formula = heart.disease ~ biking, data = heart.data)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-4.028	-1.206	-0.004	1.151	3.643

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	17.697884	0.146780	120.57	<2e-16 ***
biking	-0.199091	0.003378	-58.94	<2e-16 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.618 on 496 degrees of freedom
```

```
Multiple R-squared:  0.8751,    Adjusted R-squared:  0.8748
```

```
F-statistic: 3474 on 1 and 496 DF,  p-value: < 2.2e-16
```

```
> summary(lm(heart.disease ~ smoking, data = heart.data))
```

```
Call:
```

```
lm(formula = heart.disease ~ smoking, data = heart.data)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-8.7065	-3.7069	0.5007	3.6597	8.5434

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	7.54311	0.41251	18.286	< 2e-16 ***
smoking	0.17048	0.02355	7.239	1.73e-12 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 4.352 on 496 degrees of freedom
```

```
Multiple R-squared:  0.09556,    Adjusted R-squared:  0.09374
```

```
F-statistic: 52.41 on 1 and 496 DF,  p-value: 1.729e-12
```

```
> summary(heart.disease.lm)

Call:
lm(formula = heart.disease ~ smoking + biking, data = heart.data)

Residuals:
    Min       1Q   Median       3Q      Max
-2.1789 -0.4463  0.0362  0.4422  1.9331

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.984658   0.080137  186.99  <2e-16 ***
smoking      0.178334   0.003539   50.39  <2e-16 ***
biking     -0.200133   0.001366 -146.53  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1










Residual standard error: 0.654 on 495 degrees of freedom
Multiple R-squared:  0.9796,    Adjusted R-squared:  0.9795
F-statistic: 1.19e+04 on 2 and 495 DF,  p-value: < 2.2e-16
```

P values are small enough to be confident that that we can reject the null hypothesis that they are not significant.

5

A

```
> plotting.data<-expand.grid(
+   biking = seq(min(heart.data$biking), max(heart.data$biking), length.out=30),
+   smoking=c(min(heart.data$smoking), mean(heart.data$smoking), max(heart.data$smoking)))
> plotting.data
   biking smoking
1  1.119154  0.52585
2  3.663566  0.52585
3  6.207979  0.52585
4  8.752391  0.52585
5 11.296803  0.52585
6 13.841216  0.52585
7 16.385628  0.52585
8 18.930040  0.52585
9 21.474453  0.52585
10 24.018865  0.52585
11 26.563277  0.52585
12 29.107689  0.52585
13 31.652102  0.52585
14 34.196514  0.52585
15 36.740926  0.52585
```

Environment	History	Connections	Tutorial
   Import Dataset ▾  162 MiB ▾ 			
R ▾  Global Environment ▾			
Data			
 heart.data	498 obs. of 4 variables		
 heart.disease.lm	List of 12		
 plotting.data	90 obs. of 2 variables		

B

```
> predictions<-predict(heart.disease.lm,plotting.data)
> predictions
```

1	2	3	4	5	6
14.85445514	14.34523415	13.83601315	13.32679215	12.81757115	12.30835015
7	8	9	10	11	12
11.79912915	11.28990815	10.78068715	10.27146615	9.76224516	9.25302416
13	14	15	16	17	18
8.74380316	8.23458216	7.72536116	7.21614016	6.70691916	6.19769816
19	20	21	22	23	24
5.68847716	5.17925616	4.67003517	4.16081417	3.65159317	3.14237217
25	26	27	28	29	30
2.63315117	2.12393017	1.61470917	1.10548817	0.59626717	0.08704618
31	32	33	34	35	36
17.51326835	17.00404736	16.49482636	15.98560536	15.47638436	14.96716336
37	38	39	40	41	42
14.45794236	13.94872136	13.43950036	12.93027936	12.42105837	11.91183737
43	44	45	46	47	48
11.40261637	10.89339537	10.38417437	9.87495337	9.36573237	8.85651137
49	50	51	52	53	54
8.24730037	7.82806037	7.32884038	6.81962738	6.31040638	5.80118538

```
> plotting.data
```

	biking	smoking	v3
1	1.119154	0.52585	14.85445514
2	3.663566	0.52585	14.34523415
3	6.207979	0.52585	13.83601315
4	8.752391	0.52585	13.32679215
5	11.296803	0.52585	12.81757115
6	13.841216	0.52585	12.30835015
7	16.385628	0.52585	11.79912915
8	18.930040	0.52585	11.28990815
9	21.474453	0.52585	10.78068715
10	24.018865	0.52585	10.27146615
11	26.563277	0.52585	9.76224516

C

```
> plotting.data[2]<-round(plotting.data[2], digit=2)
> plotting.data
```

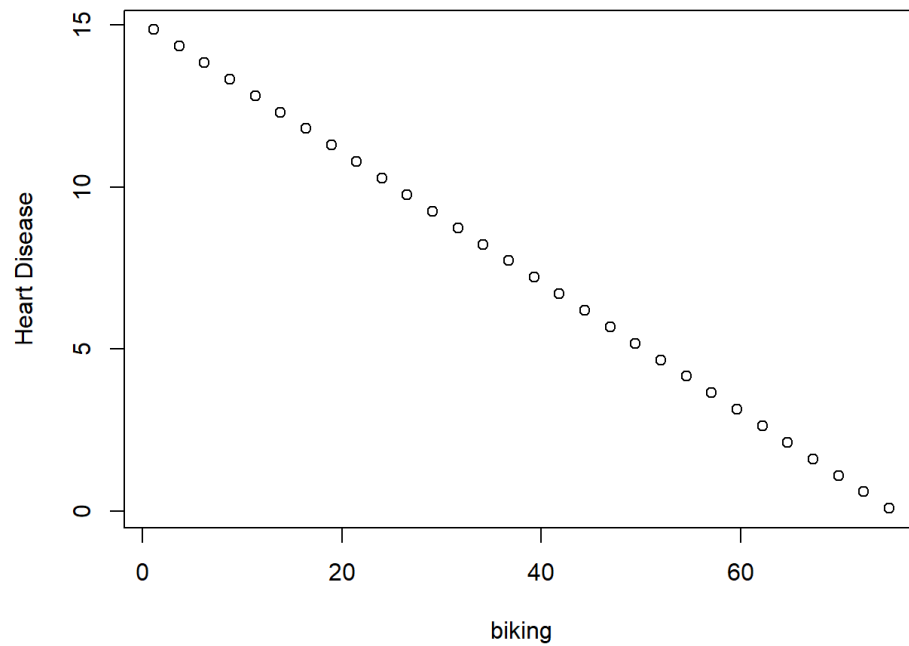
	biking	smoking	v3
1	1.119154	0.53	14.85445514
2	3.663566	0.53	14.34523415
3	6.207979	0.53	13.83601315
4	8.752391	0.53	13.32679215
5	11.296803	0.53	12.81757115
6	13.841216	0.53	12.30835015

D

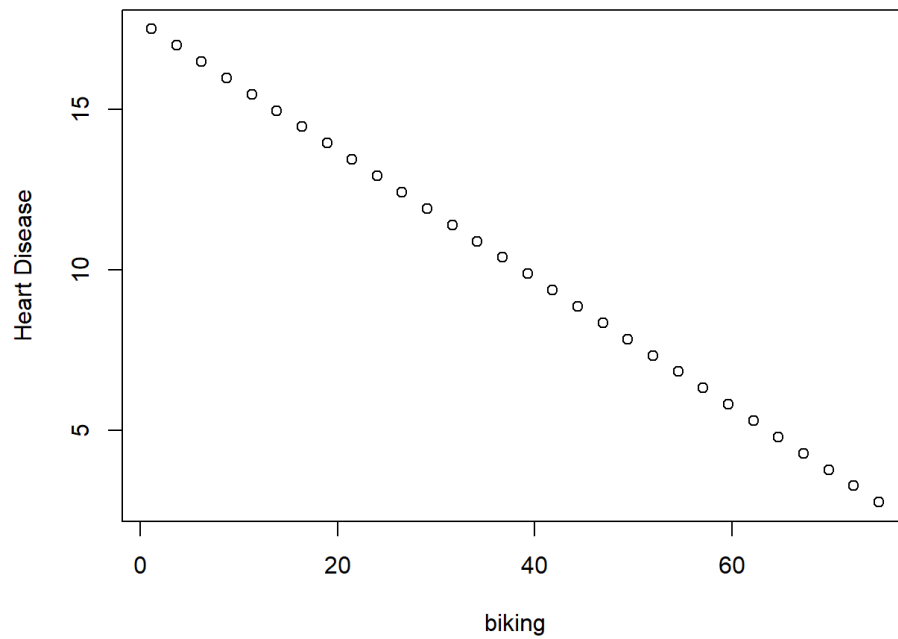
```
> smokingfactor<-as.factor(plotting.data[2])
```

E

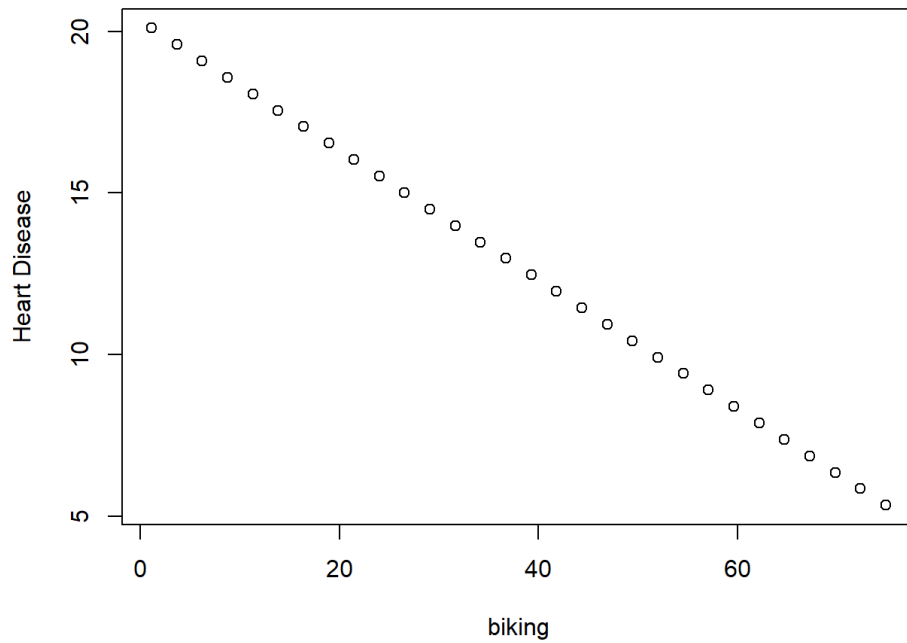
**Rates of heart disease in relation to biking in light smokers**



**Rates of heart disease in relation to biking in average smokers**

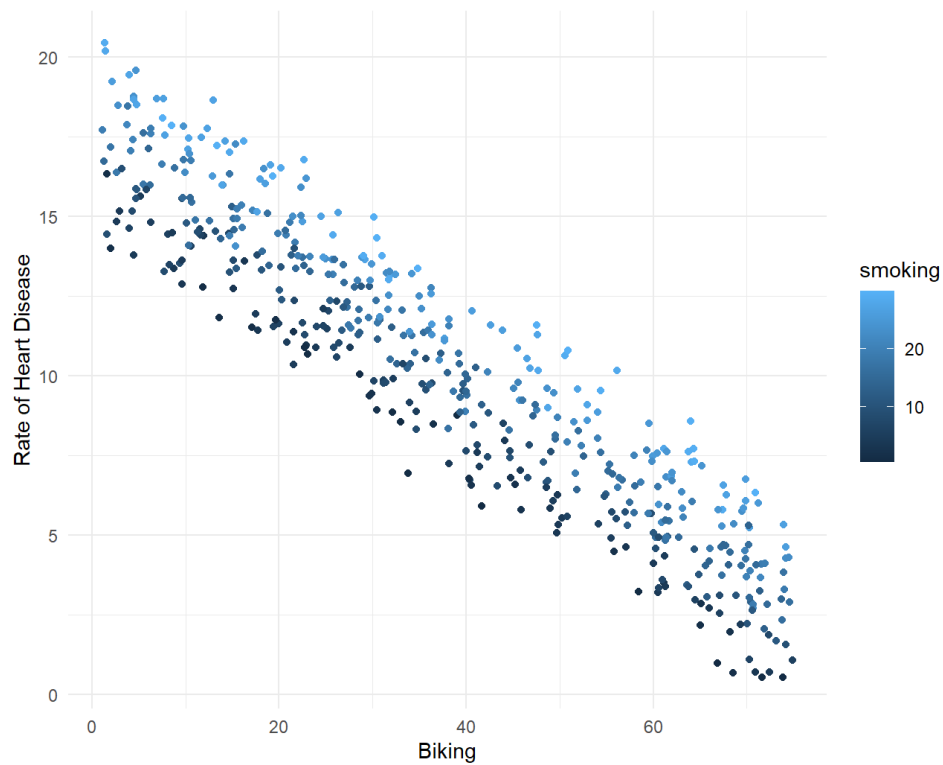


**Rates of heart disease in relation to biking in heavy smokers**



Smoking in general, especially when the subject does not ride a bike very much, seems to noticeably make them more likely to get heart disease.

6



```
> ggplot(heart.data, aes(x = biking, y = heart.disease)) +
+   geom_point(aes(color = smoking)) + geom_smooth(method = "lm", se = FALSE, co
lor = "red") +
+   labs(x = "Biking", y = "Rate of Heart Disease", size = "Heart Disease") +
+   theme_minimal()
```

7

