Lab4-Extra.R

stevengan

2022-10-10

```r
source("http://thegrantlab.org/misc/cdc.R")

head(cdc$height)
```

```
## [1] 70 64 60 66 61 64
```

```r
tail(cdc$weight, 20)
```

```
##  [1] 195 210 171 190 180 120 140 200 230 230 195 210 180 165 224 215 200 216 165
## [20] 170
```

```r
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## v purrr   0.3.5
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
#Scatter ploting
ggplot(cdc, aes(x = height, y = weight)) +
  geom_point() +
  labs(x = "Height (inches)", y = "Weight (pounds)") +
  geom_smooth()
```

```
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```
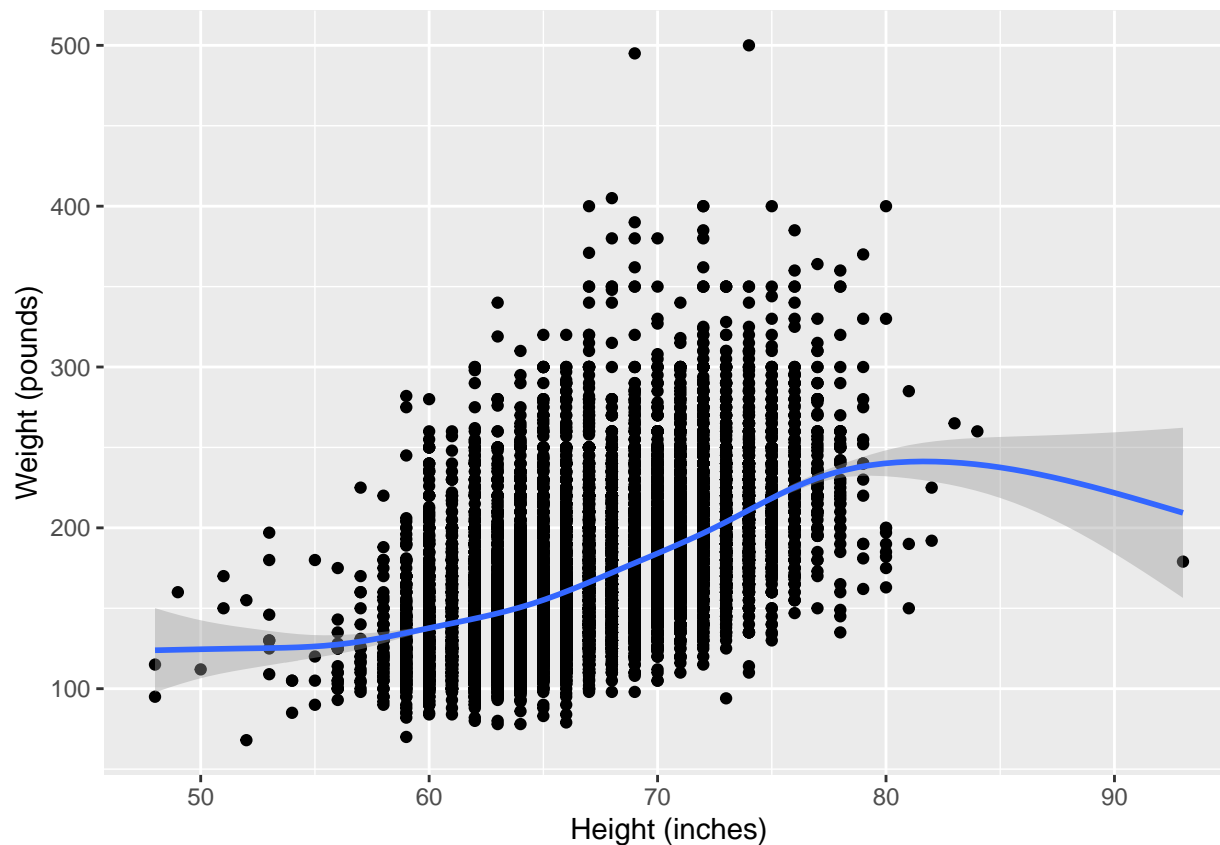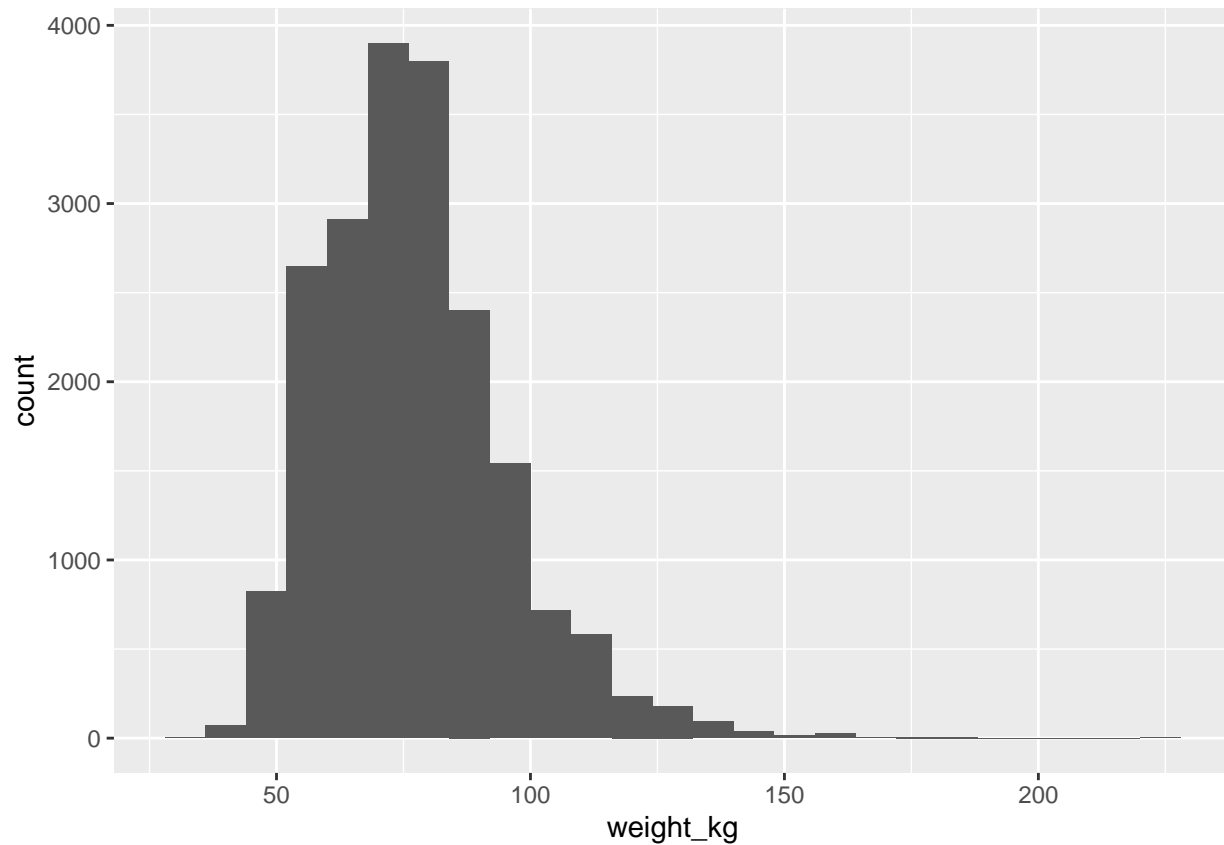
```
cor.test(cdc$height, cdc$weight)
```

```
##
##  Pearson's product-moment correlation
##
## data:  cdc$height and cdc$weight
## t = 94.429, df = 19998, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5456626 0.5648342
## sample estimates:
##       cor
## 0.5553222
```

```
#Metrics transformation, histogram, and BMI
height_m <- cdc$height * 0.0254
weight_kg <- cdc$weight * 0.454

ggplot(cdc, aes(x = weight_kg)) +
  geom_histogram(binwidth = 8)
```
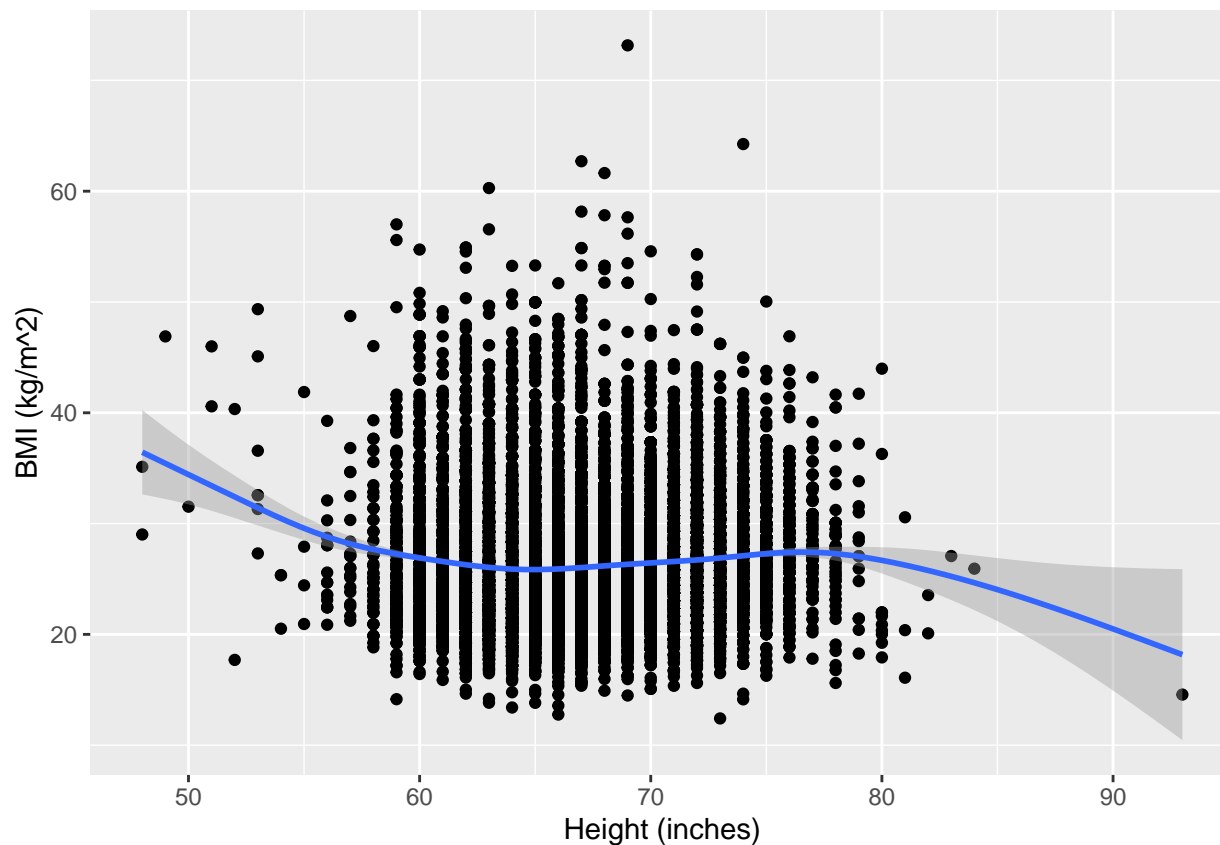
```
bmi = weight_kg / (height_m ^ 2)
cdc2 <- cbind(cdc, height_m, weight_kg, bmi)

ggplot(cdc2, aes(x = height, y = bmi)) + geom_point() +
  labs(x = "Height (inches)", y = "BMI (kg/m^2)") +
  geom_smooth()
```

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

```
cor.test(cdc2$height, cdc2$bmi)
```

```
##
##  Pearson's product-moment correlation
##
## data:  cdc2$height and cdc2$bmi
## t = 4.6008, df = 19998, p-value = 4.235e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.01866616 0.04635524
## sample estimates:
##        cor
## 0.03251694
```

```
#Logical count and subset
head(bmi >= 30, 100)
```

```
##   [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
##  [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
##  [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
##  [61] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [73] FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [85] FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
```

```
## [97]  TRUE FALSE FALSE FALSE
```

```
round(sum(bmi >= 30) / length(bmi) * 100, 1)
```

```
## [1] 19.5
```

```
cdc[567, 6]
```

```
## [1] 160
```

```
cdc[1:10, 6]
```

```
##  [1] 175 125 105 132 150 114 194 170 150 180
```

```
cdc[1:10,]
```

```
##      genhlth exerany hlthplan smoke100 height weight wtdesire age gender
## 1       good       0        1        0     70    175      175  77      m
## 2       good       0        1        1     64    125      115  33      f
## 3       good       1        1        1     60    105      105  49      f
## 4       good       1        1        0     66    132      124  42      f
## 5  very good       0        1        0     61    150      130  55      f
## 6  very good       1        1        0     64    114      114  55      f
## 7  very good       1        1        0     71    194      185  31      m
## 8  very good       0        1        0     67    170      160  45      m
## 9       good       0        1        1     65    150      130  27      f
## 10      good       1        1        0     70    180      170  44      m
```

```
cdc[1:100, c("height", "weight")]
```

```
##     height weight
## 1       70    175
## 2       64    125
## 3       60    105
## 4       66    132
## 5       61    150
## 6       64    114
## 7       71    194
## 8       67    170
## 9       65    150
## 10      70    180
## 11      69    186
## 12      69    168
## 13      66    185
## 14      70    170
## 15      69    170
## 16      73    185
## 17      67    156
## 18      71    185
## 19      75    200
```

```
## 20       67       125
## 21       69       200
## 22       65       160
## 23       73       160
## 24       67       165
## 25       64       105
## 26       68       190
## 27       67       190
## 28       69       160
## 29       61       115
## 30       74       185
## 31       67       166
## 32       71       180
## 33       71       182
## 34       68       185
## 35       64       220
## 36       63       117
## 37       70       160
## 38       69       190
## 39       65       125
## 40       67       160
## 41       65       124
## 42       66       143
## 43       64       118
## 44       73       210
## 45       64       200
## 46       59       145
## 47       68       175
## 48       68       130
## 49       66       112
## 50       61       141
## 51       65       179
## 52       64       135
## 53       62       140
## 54       72       170
## 55       68       150
## 56       69       172
## 57       64       125
## 58       62       168
## 59       64       120
## 60       65       180
## 61       72       186
## 62       66       144
## 63       73       250
## 64       60       160
## 65       74       164
## 66       68       155
## 67       66       140
## 68       64       125
## 69       63       140
## 70       71       190
## 71       65       140
## 72       66       134
## 73       72       217
```

```
## 74       71      215
## 75       63      190
## 76       67      215
## 77       64      112
## 78       62      136
## 79       63      142
## 80       74      210
## 81       68      180
## 82       63      120
## 83       74      225
## 84       74      200
## 85       64      148
## 86       66      143
## 87       60      135
## 88       60      115
## 89       66      193
## 90       64      140
## 91       71      265
## 92       67      166
## 93       64      145
## 94       62      138
## 95       67      137
## 96       74      175
## 97       71      230
## 98       65      140
## 99       74      180
## 100      60      100
```

```r
sum_gender <- sum(as.integer(cdc[bmi >= 30, "gender"]))
nbmi30 <- sum(bmi >= 30)
nmale <- 2 * nbmi30 - sum_gender
nmale
```

```
## [1] 1961
```