Model Exercises

Ai Yukino

HW link

Import packages

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(modelr)
```

Exercise 1

Familiarize yourself with the `heights` data set provided with the `modelr` package.

```
data(heights)
heights
```

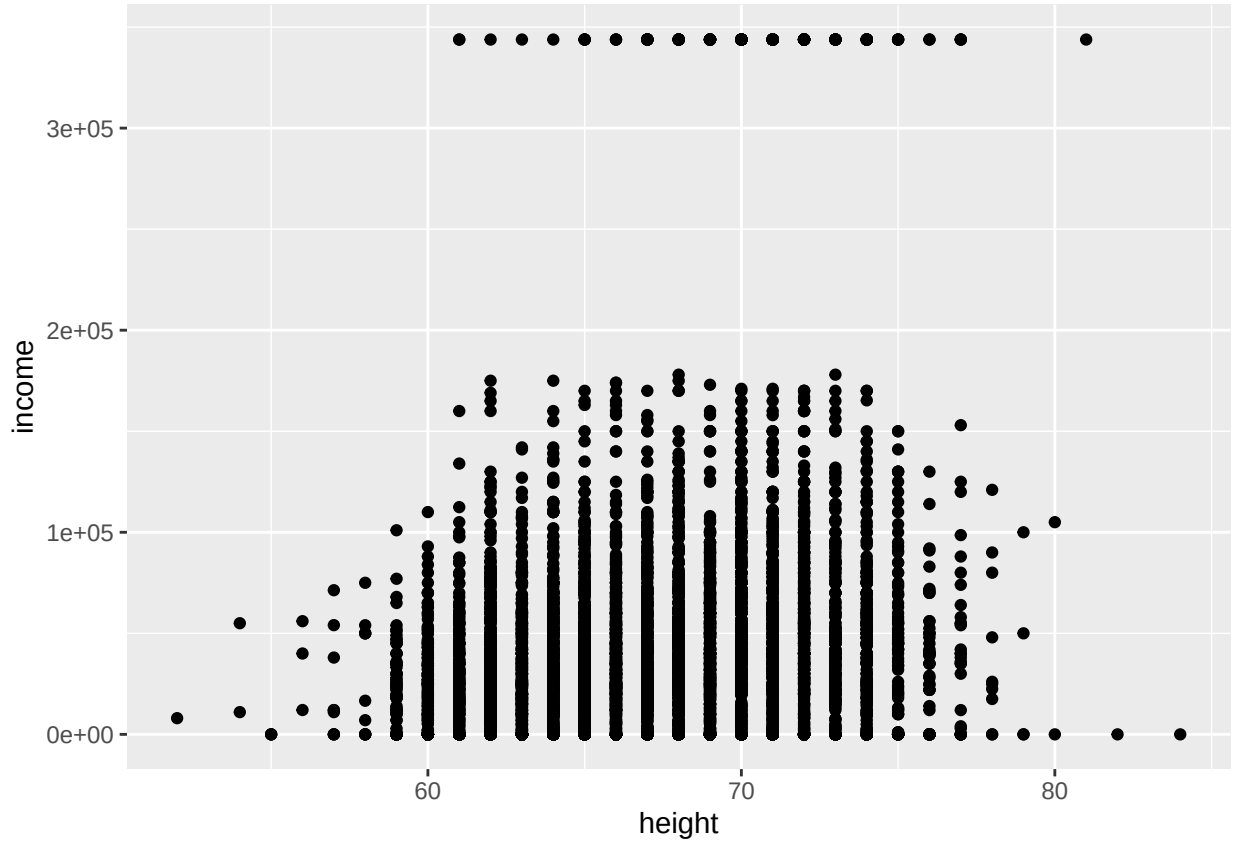
```
## # A tibble: 7,006 x 8
##    income height weight   age marital  sex    education  afqt
##     <int>  <dbl>  <int> <int> <fct>    <fct>      <int> <dbl>
##  1  19000     60    155    53 married  female        13  6.84
##  2  35000     70    156    51 married  female        10 49.4
##  3 105000     65    195    52 married  male          16 99.4
##  4  40000     63    197    54 married  female        14 44.0
##  5  75000     66    190    49 married  male          14 59.7
##  6 102000     68    200    49 divorced female        18 98.8
##  7      0     74    225    48 married  male          16 82.3
##  8  70000     64    160    54 divorced female        12 50.3
##  9  60000     69    162    55 divorced male          12 89.7
## 10 150000     69    194    54 divorced male          13 96.0
## # ... with 6,996 more rows
```

```
# ?heights
```

Exercise 2

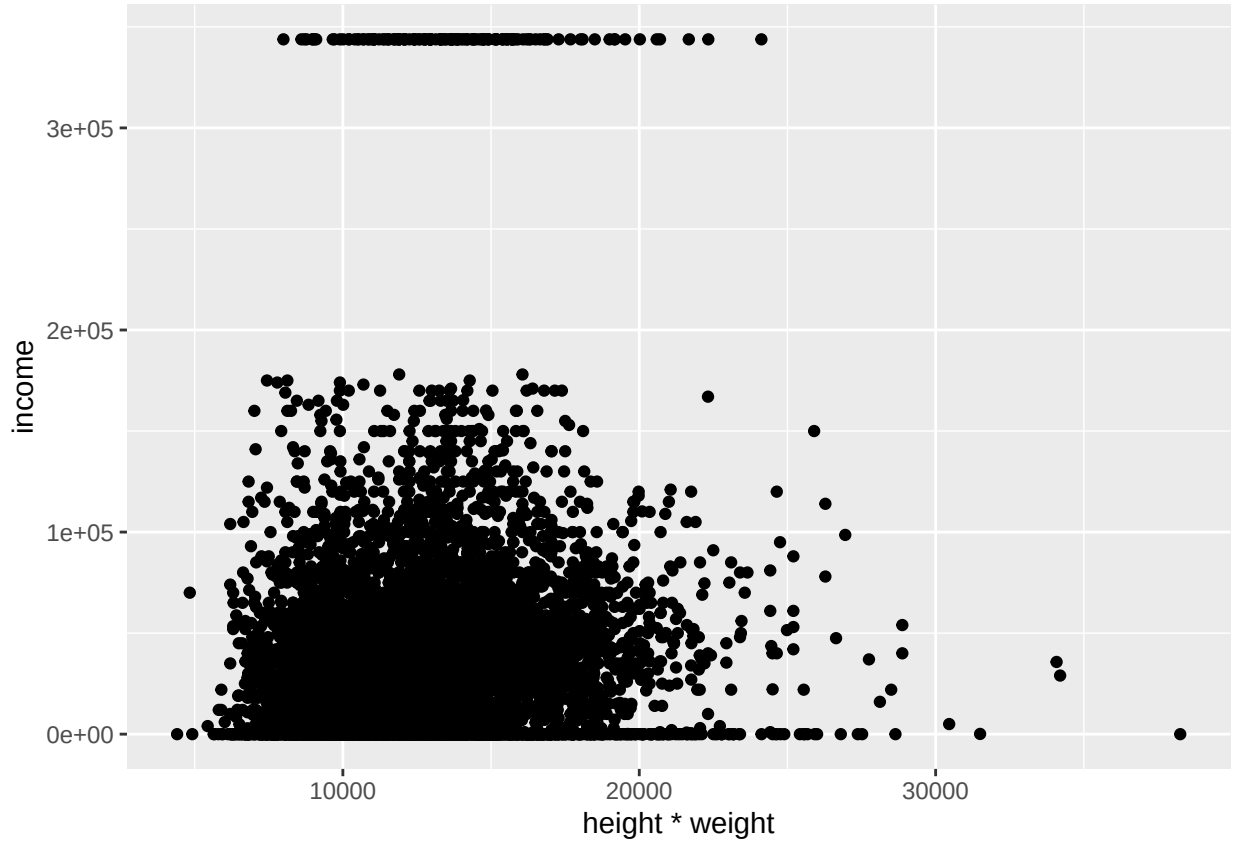
Create a list of formulas for modeling income with: - `height` - `height` · `weight` - linear combination of all variables

```
ggplot(data = heights, aes(y = income, x = height)) +
  geom_point()
```




```
ggplot(data = heights, aes(y = income, x = height * weight)) +
  geom_point()
```

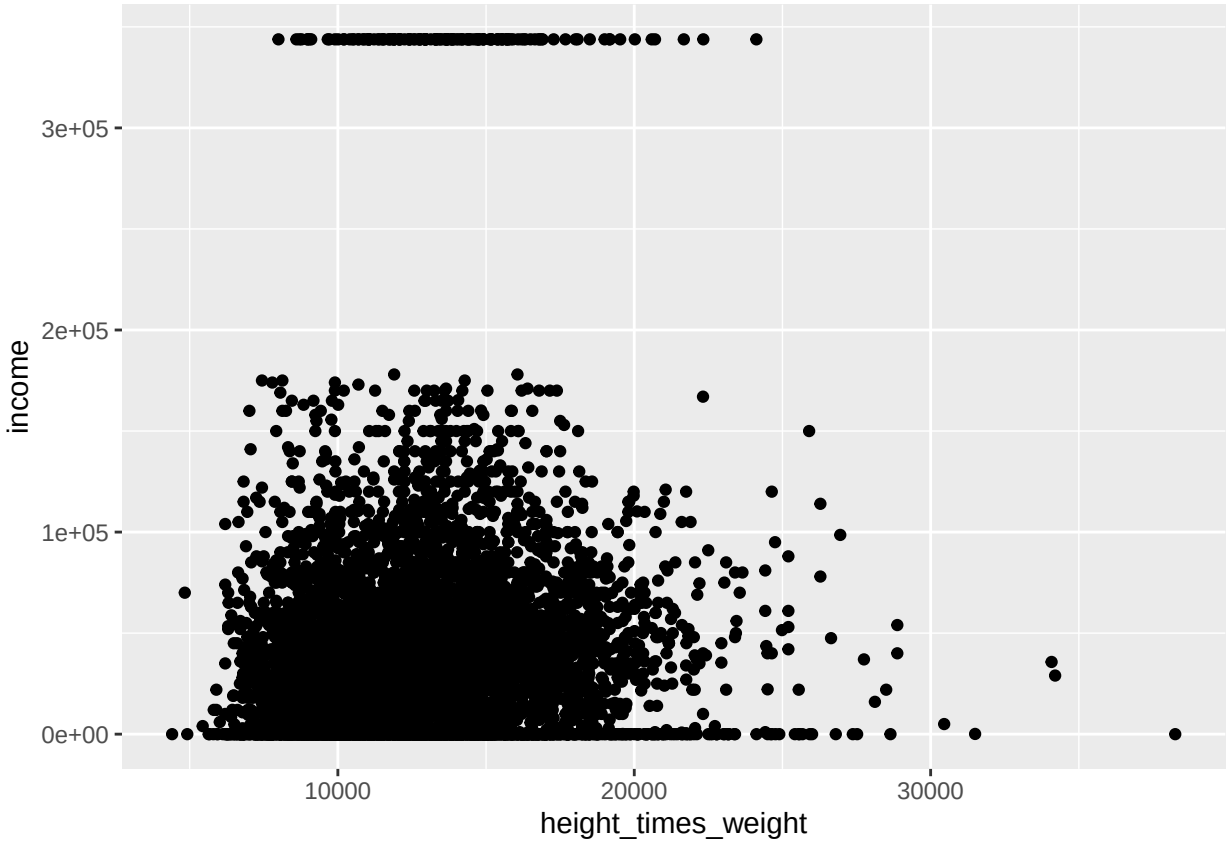
```
## Warning: Removed 95 rows containing missing values (geom_point).
```




```
tb <- heights
tb$height_times_weight = tb$height * tb$weight

ggplot(data = tb, aes(y = income, x = height_times_weight)) +
  geom_point()
```

```
## Warning: Removed 95 rows containing missing values (geom_point).
```


3e+05
2e+05
1e+05
0e+00
income
10000
20000
30000
height_times_weight