BodyFat

2023-07-18

National Health and Nutrition Examination Survey 2017-2018

The NHANES includes DEXA scan of participants age 8-59, which allows researchers to calculate the body fat percentage of participants.

A large amount of other data is also collected. Sampling weights are used to collect the data, so a simple arithmetic mean will not work. Instead, weighted means and weighted quantiles must be used to calculate any means and quantiles (in this case, percentiles). In this case `Hmisc::wtd.quantile` calculates weighted quantiles.

Waist/Height Ratio is a commonly used way to calculate a person's health, since calcalating a person's body fat percentage is difficult and hard to do. Even a DEXA scan is not perfectly accurate. Waist/Height Ratio may be a better alternative to BMI in many cases.

A Waist/Height Ratio of under 0.5 is optimal, but the majority of the population is above that, confirming that there is indeed an obesity crisis. As expected, Waist/Height Ratio, BMI, and Body Fat Percentage are all highly correlated with each other.

Women and men store fat differently, so body fat percentage that is healthy is different. For women, body fat percentage should be under 33% and for men, body fat percentage should be under 21%.

For the last section, I use a linear model to calculate the estimated body fat percentage by BMI.

```
library(haven)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## 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
```

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```r
res = read_xpt("C:/Users/willi/Downloads/DXX_J.xpt")
demo = read_xpt("C:/Users/willi/Downloads/DEMO_J.xpt")
body = read_xpt("C:/Users/willi/Downloads/BMX_J.xpt")

total = demo %>%
  left_join(body, by = "SEQN") %>%
  left_join(res, by = "SEQN")

bf = total %>%
  mutate(Race = recode(RIDRETH3,
                       `1` = "Mexican American",
                       `2` = "Other Hispanic",
                       `3` = "Non-Hispanic White",
                       `4` = "Non-Hispanic Black",
                       `6` = "Non-Hispanic Asian",
                       `7` = "Other Race")) %>%
  mutate(Edu = recode(DMDEDUC2,
                      `1` =     "Less than 12th grade",
                      `2` =     "Less than 12th grade",
                      `3` = "High school graduate/GED or equivalent",
                      `4` = "Some college or AA degree",
                      `5` = "College graduate or above",
                      `6` = "Other",
                      `7` = "Other",
                      .default = "Other")) %>%
  mutate(Sex = recode(RIAGENDR,
                      `1` = "Male",
                      `2` = "Female")) %>%
  filter(!is.na(DXDTOPF)) %>%
  filter(RIDAGEYR >= 25 & RIDAGEYR <= 59) %>%
  group_by(RIDAGEYR, Sex) %>%
  summarize(BodyFat05 = Hmisc::wtd.quantile(DXDTOPF, WTMEC2YR, probs = 0.05, normwt = TRUE),
            BodyFat10 = Hmisc::wtd.quantile(DXDTOPF, WTMEC2YR, probs = 0.1, normwt = TRUE),
            BodyFat25 = Hmisc::wtd.quantile(DXDTOPF, WTMEC2YR, probs = 0.25, normwt = TRUE),
            BodyFat50 = Hmisc::wtd.quantile(DXDTOPF, WTMEC2YR, probs = 0.5, normwt = TRUE))
```

```
## 'summarise()' has grouped output by 'RIDAGEYR'. You can override using the
## '.groups' argument.
```

```r
bf = bf %>%
  pivot_longer(cols = 3:6, names_prefix = "BodyFat", names_to = "Body Fat Percentile", values_to = "Per

ggplot(bf, aes(x = RIDAGEYR, y = Percent, color = Sex, linetype = `Body Fat Percentile`, group = interac
  theme_bw() + xlab("Age") + ylab("Body Fat Percentage")+
  labs(fill="Percentile and Sex") +
  ggtitle("Body Fat Percentile by Age and Sex") +
  stat_smooth(se = FALSE, span = 0.65, formula = y ~ x, method = "loess")
```

Body Fat Percentile by Age and Sex
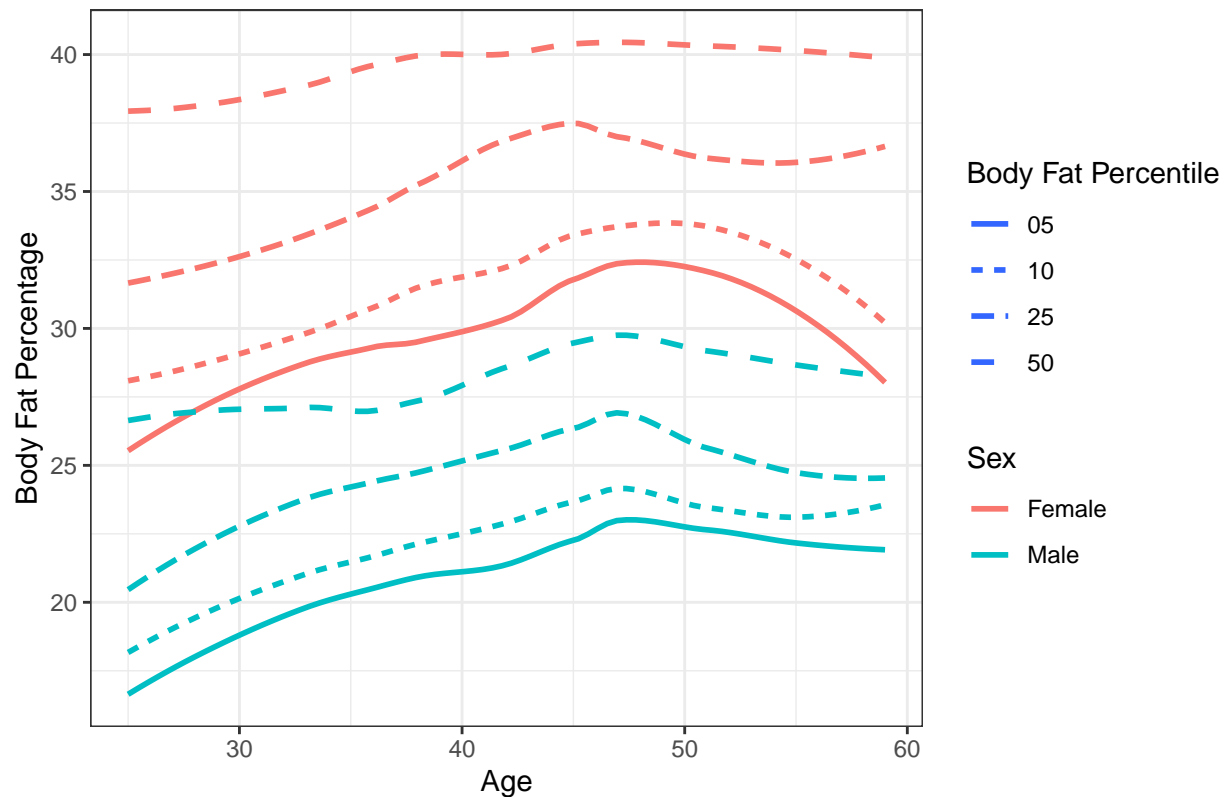

```r
ratio = total %>%
  mutate(Race = recode(RIDRETH3,
                        `1` = "Mexican American",
                        `2` = "Other Hispanic",
                        `3` = "Non-Hispanic White",
                        `4` = "Non-Hispanic Black",
                        `6` = "Non-Hispanic Asian",
                        `7` = "Other Race")) %>%
  mutate(Edu = recode(DMDEDUC2,
                      `1` =      "Less than 12th grade",
                      `2` =      "Less than 12th grade",
                      `3` = "High school graduate/GED or equivalent",
                      `4` = "Some college or AA degree",
                      `5` = "College graduate or above",
                      `6` = "Other",
                      `7` = "Other",
                      .default = "Other")) %>%
  mutate(Sex = recode(RIAGENDR,
                      `1` = "Male",
                      `2` = "Female")) %>%
  mutate(Ratio = BMXWAIST/BMXHT) %>%
  filter(RIDAGEYR >= 25 & RIDAGEYR <= 59) %>%
  group_by(RIDAGEYR, Sex) %>%
  summarize(BodyFat05 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.05, normwt = TRUE),
            BodyFat10 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.1, normwt = TRUE),
            BodyFat25 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.25, normwt = TRUE),
```

```
              BodyFat50 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.5, normwt = TRUE))
```

```
## 'summarise()' has grouped output by 'RIDAGEYR'. You can override using the
## '.groups' argument.
```

```
ratio = ratio %>%
  pivot_longer(cols = 3:6, names_prefix = "BodyFat", names_to = "Waist/Height Percentile", values_to =

ggplot(ratio, aes(x = RIDAGEYR, y = Percent, color = Sex, linetype = `Waist/Height Percentile`, group =
  theme_bw() + xlab("Age") + ylab("Waist/Height Ratio")+
  labs(fill="Percentile and Sex") +
  ggtitle("Waist/Height Percentile by Age and Sex") +
  stat_smooth(se = FALSE, span = 0.65, formula = y ~ x, method = "loess")
```
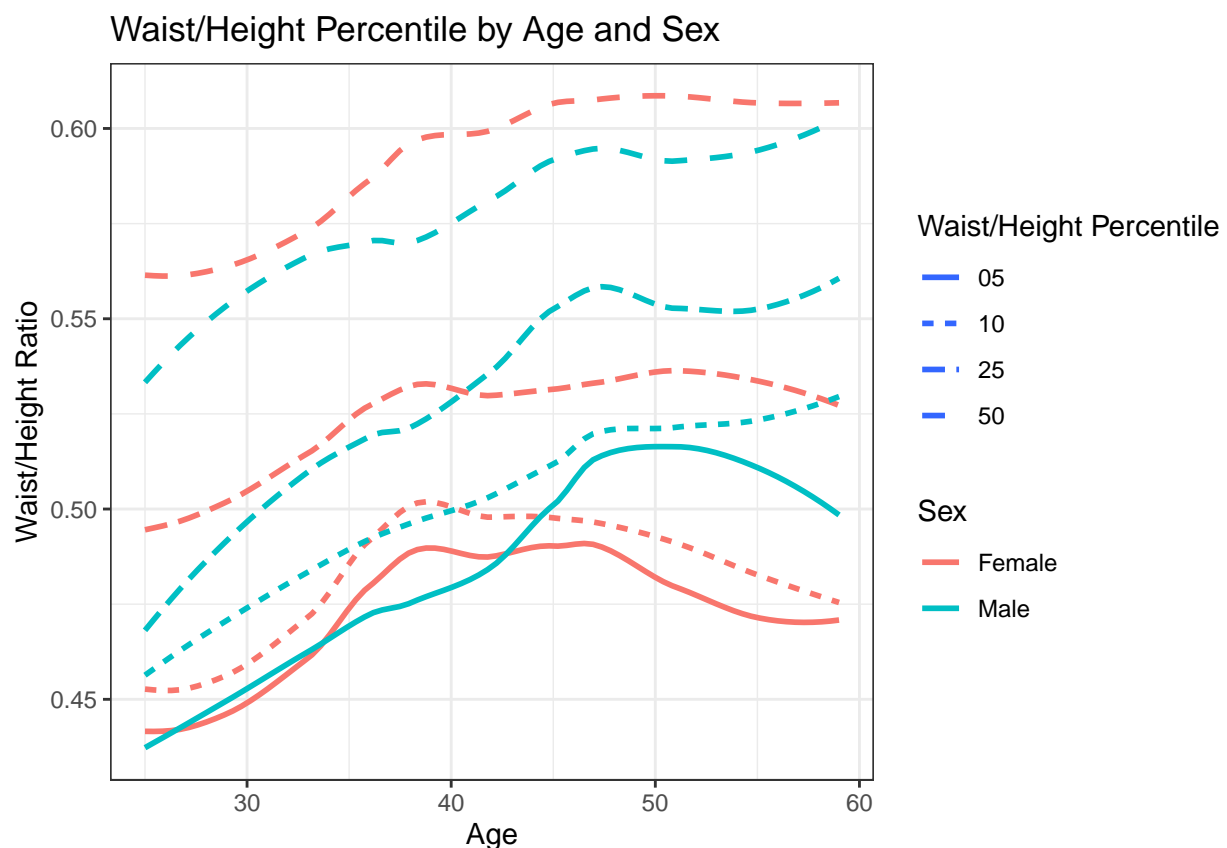


```
ratio = total %>%
  mutate(Race = recode(RIDRETH3,
                       `1` = "Mexican American",
                       `2` = "Other Hispanic",
                       `3` = "Non-Hispanic White",
                       `4` = "Non-Hispanic Black",
                       `6` = "Non-Hispanic Asian",
                       `7` = "Other Race")) %>%
  mutate(Edu = recode(DMDEDUC2,
                       `1` =     "Less than 12th grade",
```

```
                          `2` =       "Less than 12th grade",
                          `3` = "High school graduate/GED or equivalent",
                          `4` = "Some college or AA degree",
                          `5` = "College graduate or above",
                          `6` = "Other",
                          `7` = "Other",
                          .default = "Other")) %>%
  mutate(Sex = recode(RIAGENDR,
                          `1` = "Male",
                          `2` = "Female")) %>%
  mutate(Ratio = BMXWAIST/BMXHIP) %>%
  filter(RIDAGEYR >= 25 & RIDAGEYR <= 59) %>%
  group_by(RIDAGEYR, Sex) %>%
  summarize(BodyFat05 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.05, normwt = TRUE),
            BodyFat10 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.1, normwt = TRUE),
            BodyFat25 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.25, normwt = TRUE),
            BodyFat50 = Hmisc::wtd.quantile(Ratio, WTMEC2YR, probs = 0.5, normwt = TRUE))
```

```
## `summarise()` has grouped output by 'RIDAGEYR'. You can override using the
## `.groups` argument.
```

```
ratio = ratio %>%
  pivot_longer(cols = 3:6, names_prefix = "BodyFat", names_to = "Waist/Hip Percentile", values_to = "Per

ggplot(ratio, aes(x = RIDAGEYR, y = Percent, color = Sex, linetype = `Waist/Hip Percentile`, group = in
  theme_bw() + xlab("Age") + ylab("Waist/Hip Ratio")+
  labs(fill="Percentile and Sex") +
  ggtitle("Waist/Hip Percentile by Age and Sex") +
  stat_smooth(se = FALSE, span = 0.65, formula = y ~ x, method = "loess")
```

Waist/Hip Percentile by Age and Sex
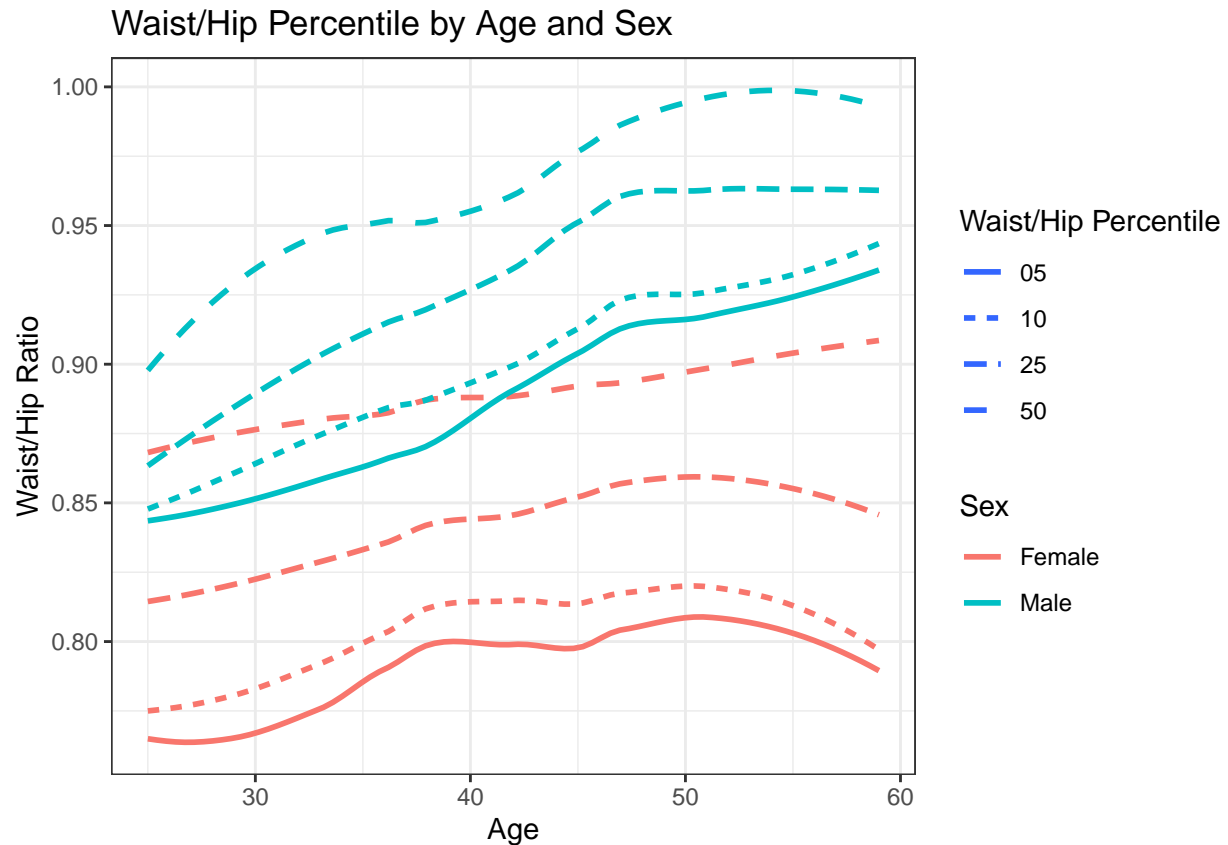
```
total = total %>%
  filter(RIDAGEYR >= 20 & RIDAGEYR <= 59)
gg = lm(DXDTOPF ~ BMXWAIST + I(BMXWAIST^2) + BMXHT + BMXHIP + RIAGENDR + factor(RIDRETH3), data = total)
total$BodyFatPredict = predict(gg, total)
```

```
bf = total %>%
  mutate(Race = recode(RIDRETH3,
                       `1` = "Mexican American",
                       `2` = "Other Hispanic",
                       `3` = "Non-Hispanic White",
                       `4` = "Non-Hispanic Black",
                       `6` = "Non-Hispanic Asian",
                       `7` = "Other Race")) %>%
  mutate(Edu = recode(DMDEDUC2,
                      `1` =     "Less than 12th grade",
                      `2` =     "Less than 12th grade",
                      `3` = "High school graduate/GED or equivalent",
                      `4` = "Some college or AA degree",
                      `5` = "College graduate or above",
                      `6` = "Other",
                      `7` = "Other",
                      .default = "Other")) %>%
  mutate(Sex = recode(RIAGENDR,
                      `1` = "Male",
                      `2` = "Female")) %>%
```

```r
  filter(!is.na(BodyFatPredict)) %>%
  filter(RIDAGEYR >= 25 & RIDAGEYR <= 59) %>%
  group_by(RIDAGEYR, Sex) %>%
  summarize(BodyFat05 = Hmisc::wtd.quantile(BodyFatPredict, WTMEC2YR, probs = 0.05, normwt = TRUE),
            BodyFat10 = Hmisc::wtd.quantile(BodyFatPredict, WTMEC2YR, probs = 0.1, normwt = TRUE),
            BodyFat25 = Hmisc::wtd.quantile(BodyFatPredict, WTMEC2YR, probs = 0.25, normwt = TRUE),
            BodyFat50 = Hmisc::wtd.quantile(BodyFatPredict, WTMEC2YR, probs = 0.5, normwt = TRUE))
```

```
## 'summarise()' has grouped output by 'RIDAGEYR'. You can override using the
## '.groups' argument.
```

```r
below15 = total %>%
  filter(RIAGENDR == 1 & !is.na(DXDTOPF)) %>%
  mutate(Below15 = ifelse(DXDTOPF <= 15, 1, 0)) %>%
  group_by(RIDAGEYR) %>%
  summarize(Below15 = weighted.mean(Below15, WTMEC2YR))

bf = bf %>%
  pivot_longer(cols = 3:6, names_prefix = "BodyFat", names_to = "Body Fat Percentile", values_to = "Perc

ggplot(bf, aes(x = RIDAGEYR, y = Percent, color = Sex, linetype = `Body Fat Percentile`, group = interac
  theme_bw() + xlab("Age") + ylab("Body Fat Percentage")+
  labs(fill="Percentile and Sex") +
  ggtitle("Body Fat Percentile (Estimated with Formula) by Age and Sex") +
  stat_smooth(se = FALSE, span = 0.65, formula = y ~ x, method = "loess")
```
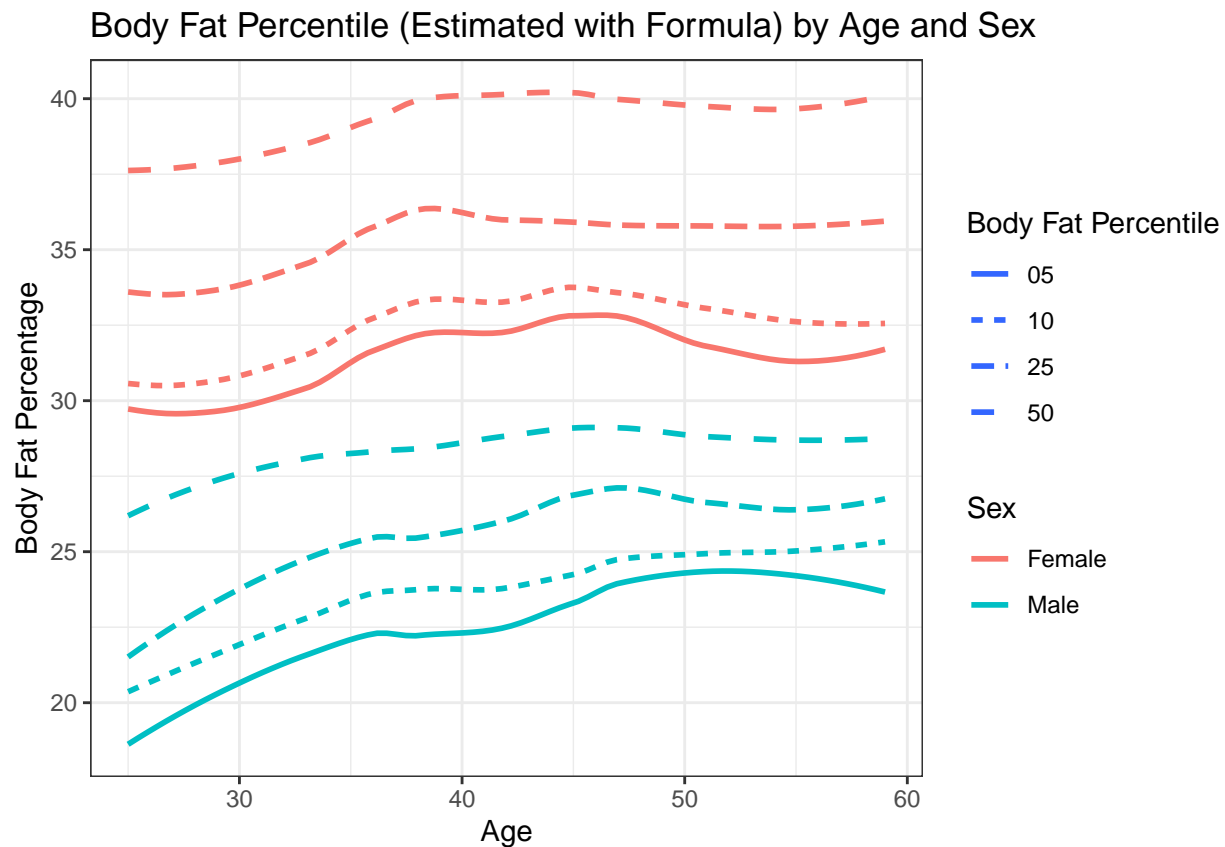
BMI vs Body Fat Percentage from DEXA Scan

```
total = total %>%
  filter(RIDAGEYR >= 20 & RIDAGEYR <= 50)
ggplot(total %>% filter(RIAGENDR == 1), aes(x = DXDTOPF, BMXBMI)) + geom_point() + stat_smooth(se = FALS
```
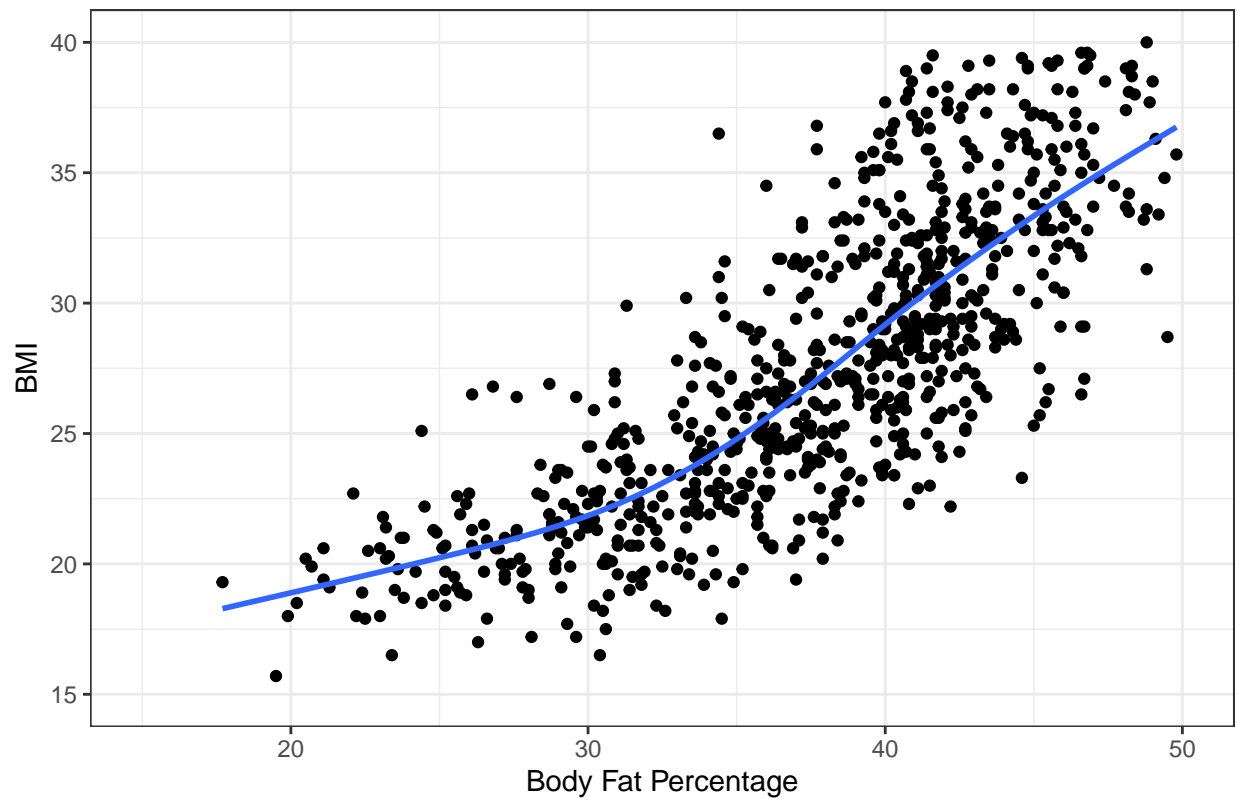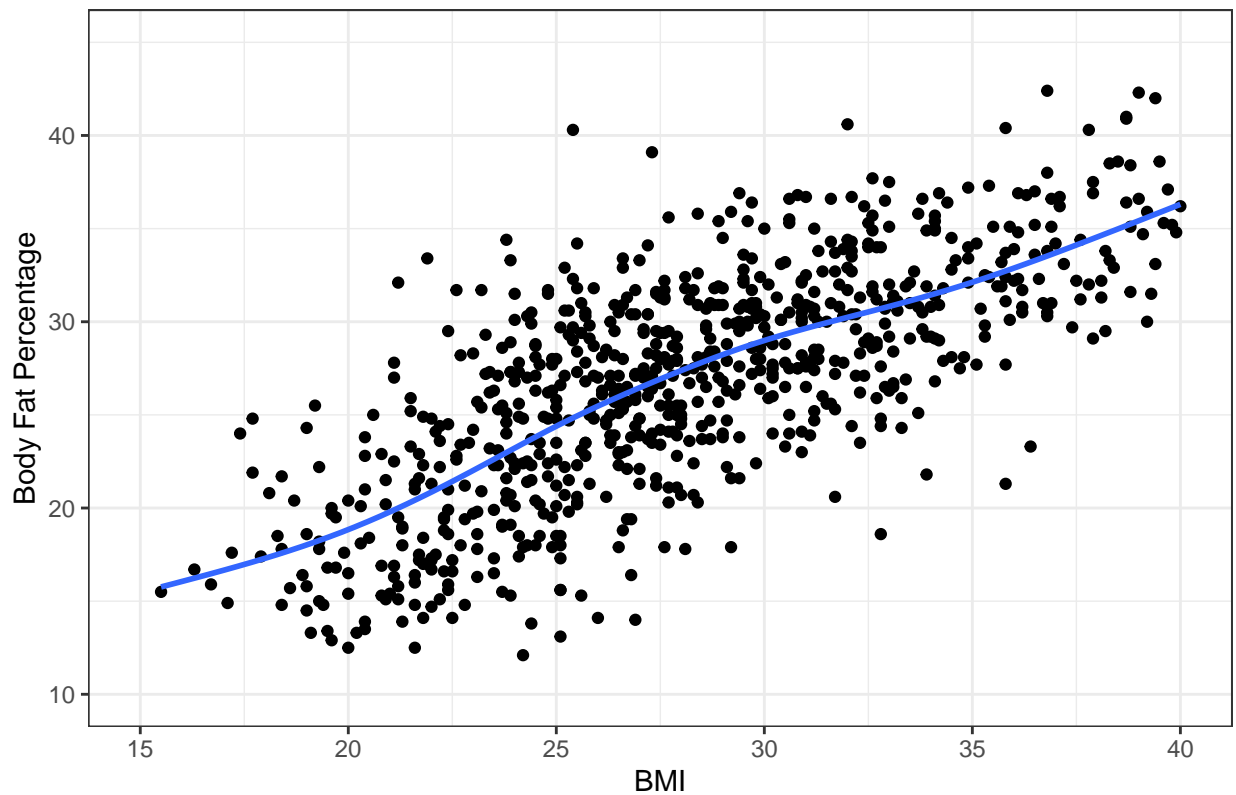
```
## 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 451 rows containing non-finite values ('stat_smooth()').
```
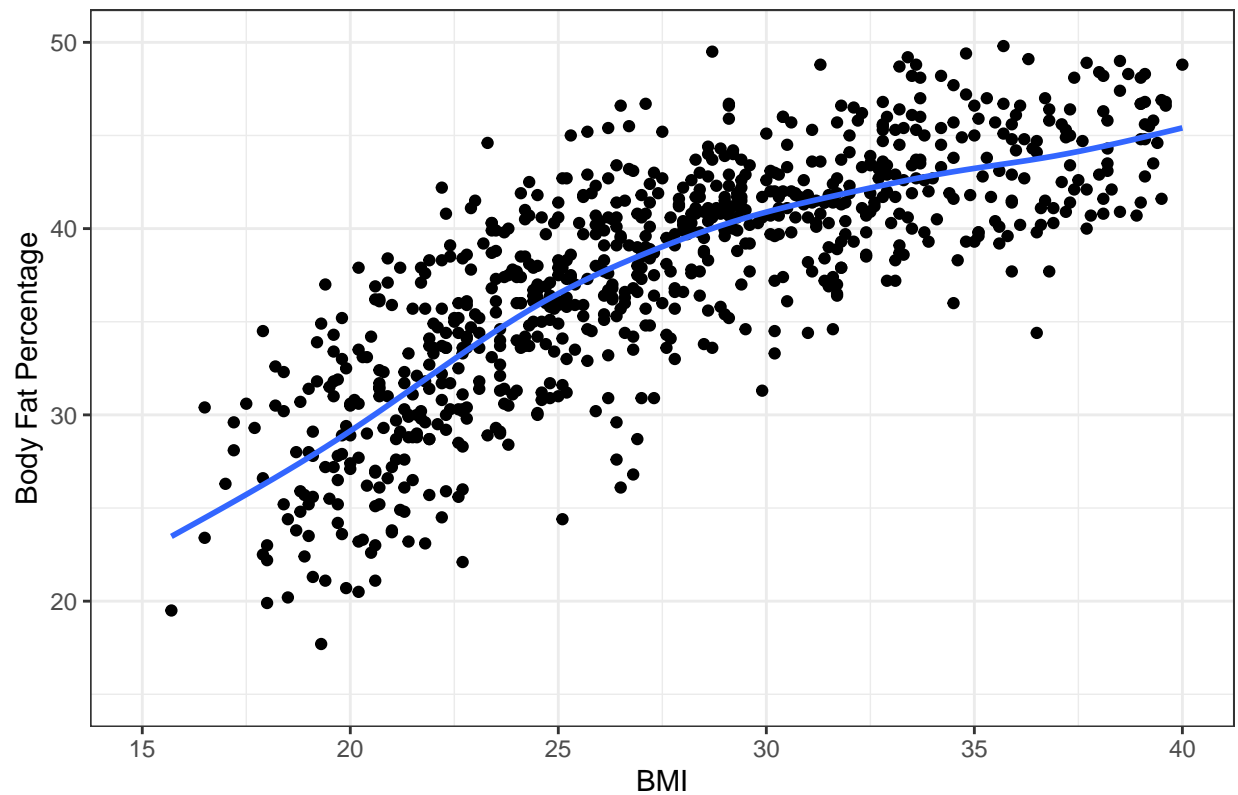
```
## Warning: Removed 451 rows containing missing values ('geom_point()').
```

Body Fat Percentage and BMI for Males Age 20–50
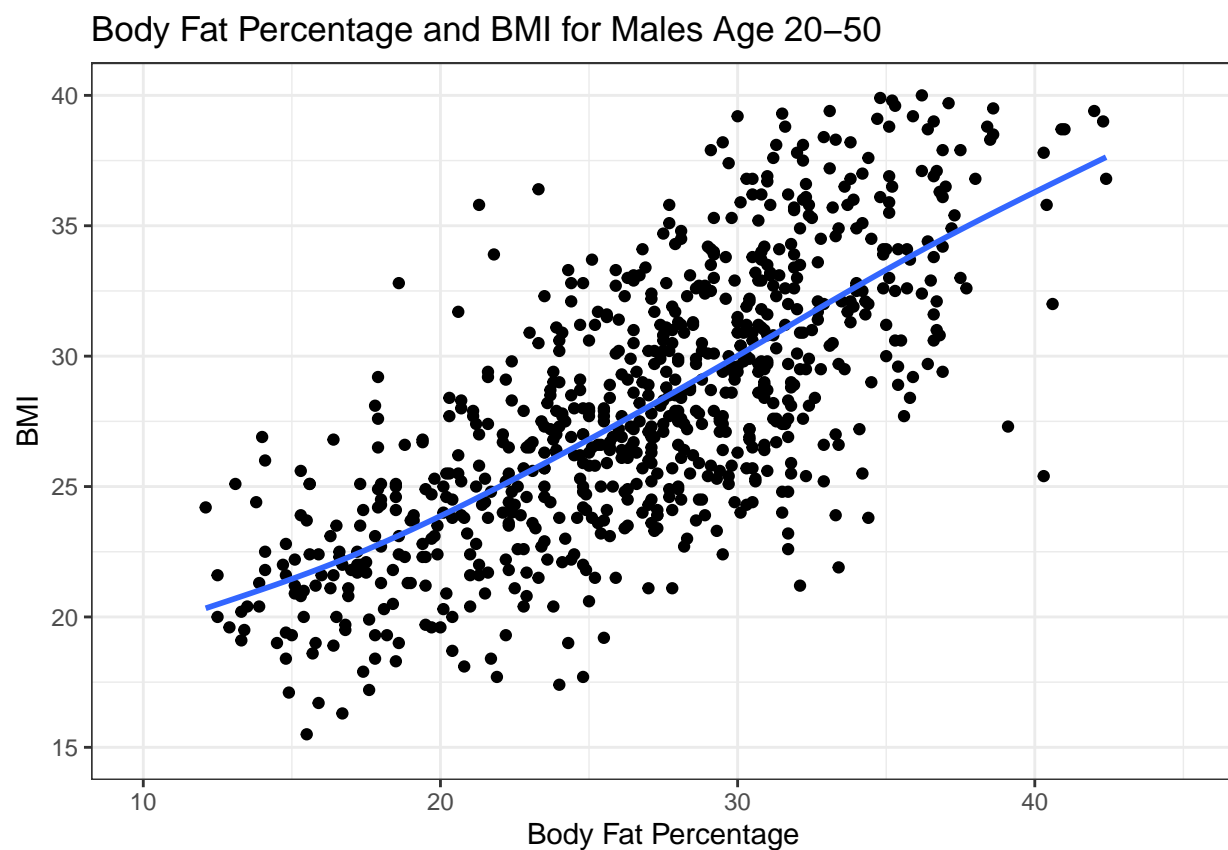

```
ggplot(total %>% filter(RIAGENDR == 1 & BMXHT >= 173 & BMXHT <= 180), aes(x = DXDTOPF, BMXBMI)) + geom_
```

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

```
## Warning: Removed 92 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 92 rows containing missing values ('geom_point()').
```

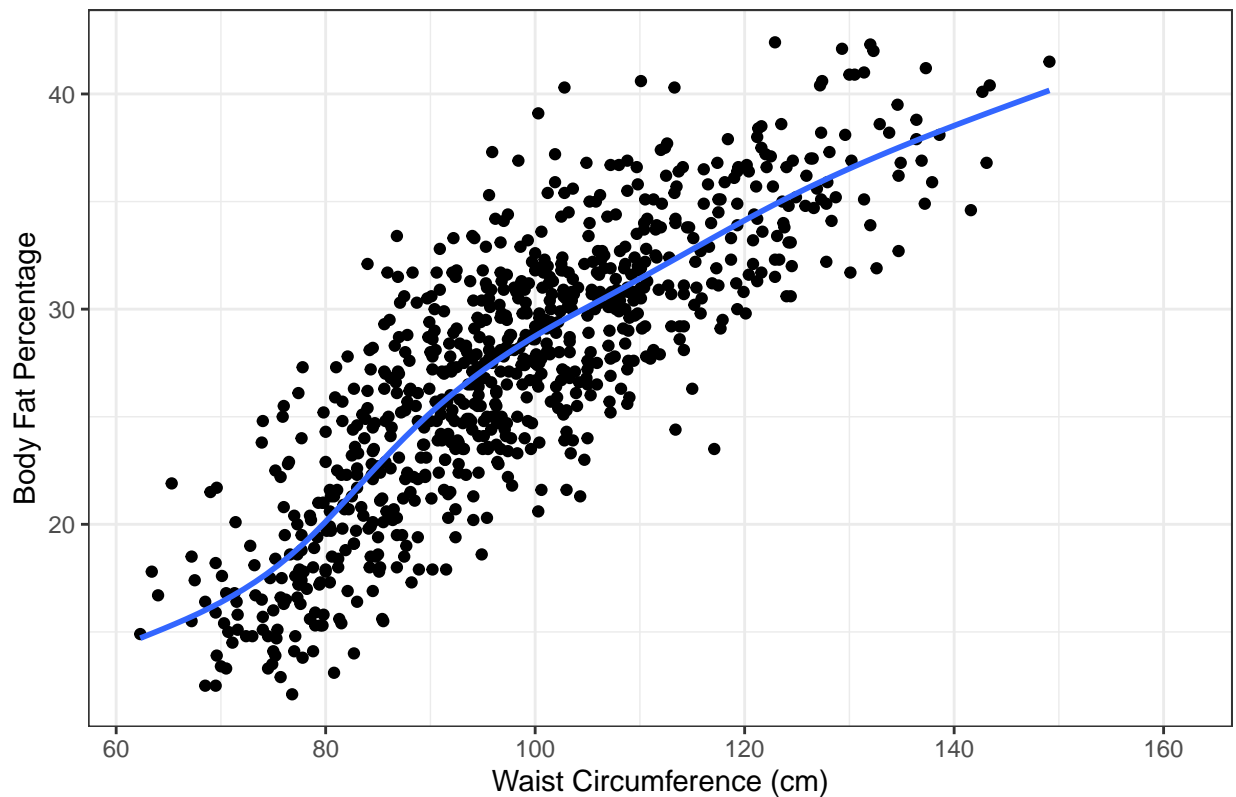## Body Fat Percentage and BMI for Males Age 20–50 (Height 173–180 cm)


```
ggplot(total %>% filter(RIAGENDR == 2), aes(x = DXDTOPF, BMXBMI)) + geom_point() + stat_smooth(se = FALS
```

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

```
## Warning: Removed 564 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 564 rows containing missing values ('geom_point()').
```

Body Fat Percentage and BMI for Females Age 20–50


```r
ggplot(total %>% filter(RIAGENDR == 1), aes(y = DXDTOPF, x = BMXBMI)) + geom_point() + stat_smooth(se =
```

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

## Warning: Removed 451 rows containing non-finite values ('stat_smooth()').

## Warning: Removed 451 rows containing missing values ('geom_point()').
```

BMI and Body Fat Percentage for Males Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 2), aes(y = DXDTOPF, x = BMXBMI)) + geom_point() + stat_smooth(se =
```

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

```
## Warning: Removed 564 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 564 rows containing missing values ('geom_point()').
```

BMI and Body Fat Percentage for Females Age 20–50

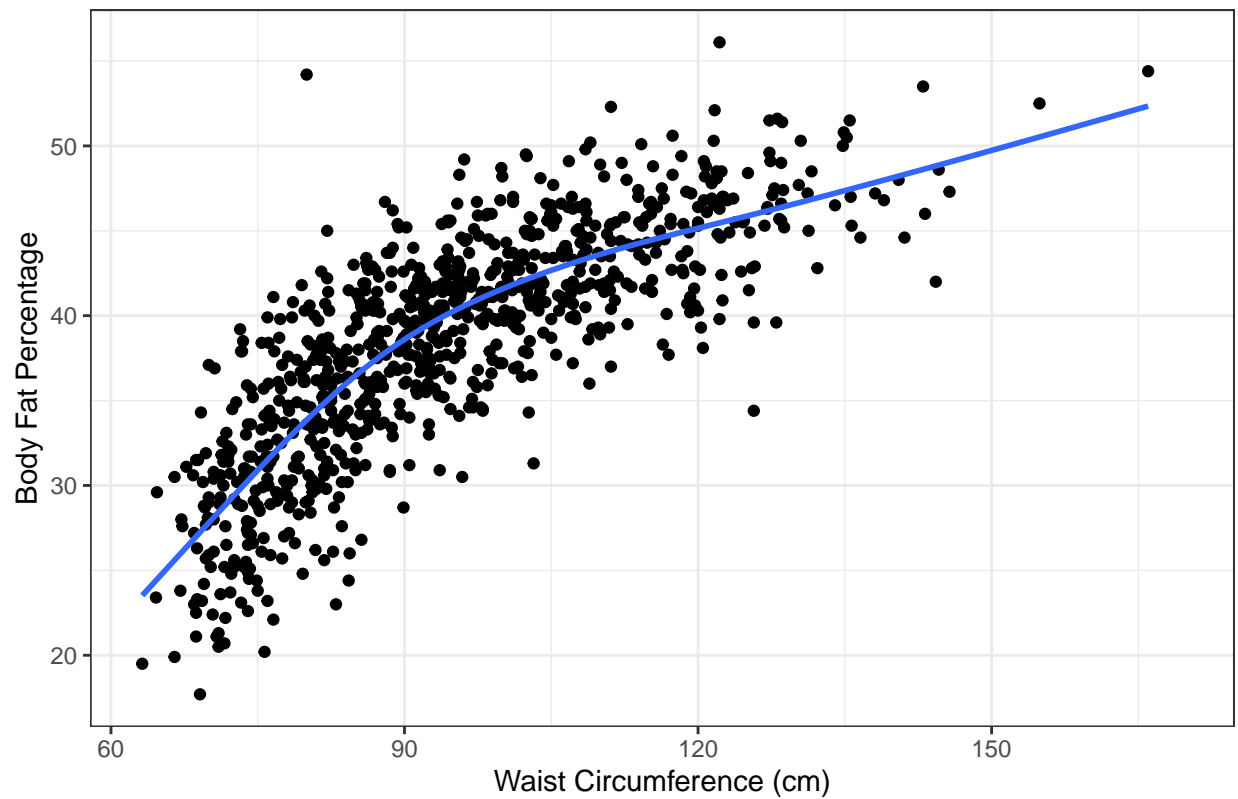```
ggplot(total %>% filter(RIAGENDR == 1), aes(x = BMXWAIST, y = DXDTOPF)) + geom_point() + stat_smooth(se
```

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

## Warning: Removed 409 rows containing non-finite values ('stat_smooth()').

## Warning: Removed 409 rows containing missing values ('geom_point()').

## Waist Circumference and Body Fat Percentage for Males Age 20–50



```
ggplot(total %>% filter(RIAGENDR == 1 & BMXHT >= 173 & BMXHT <= 180), aes(x = BMXWAIST, y = DXDTOPF)) +
```
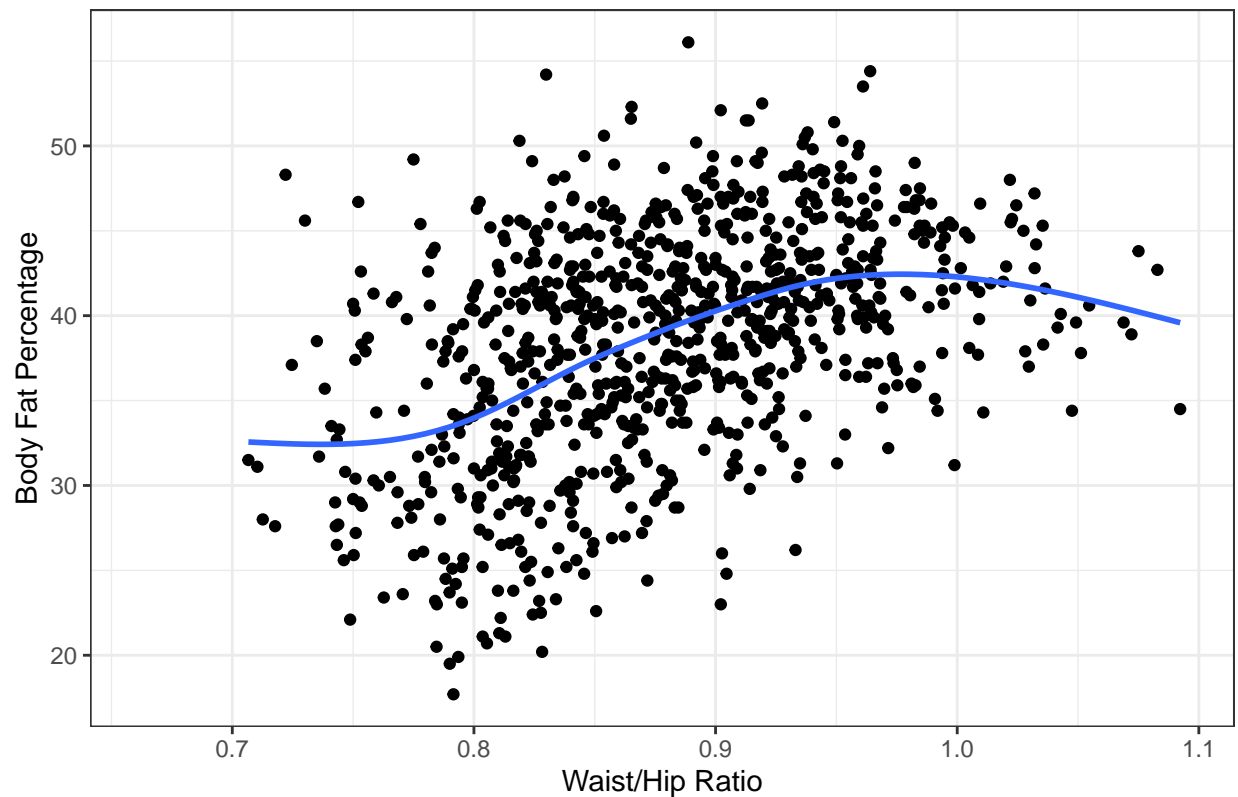
```
'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```

```
## Warning: Removed 79 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 79 rows containing missing values ('geom_point()').
```

Waist Circumference and Body Fat Percentage for Males Age 20–50 (Heigh
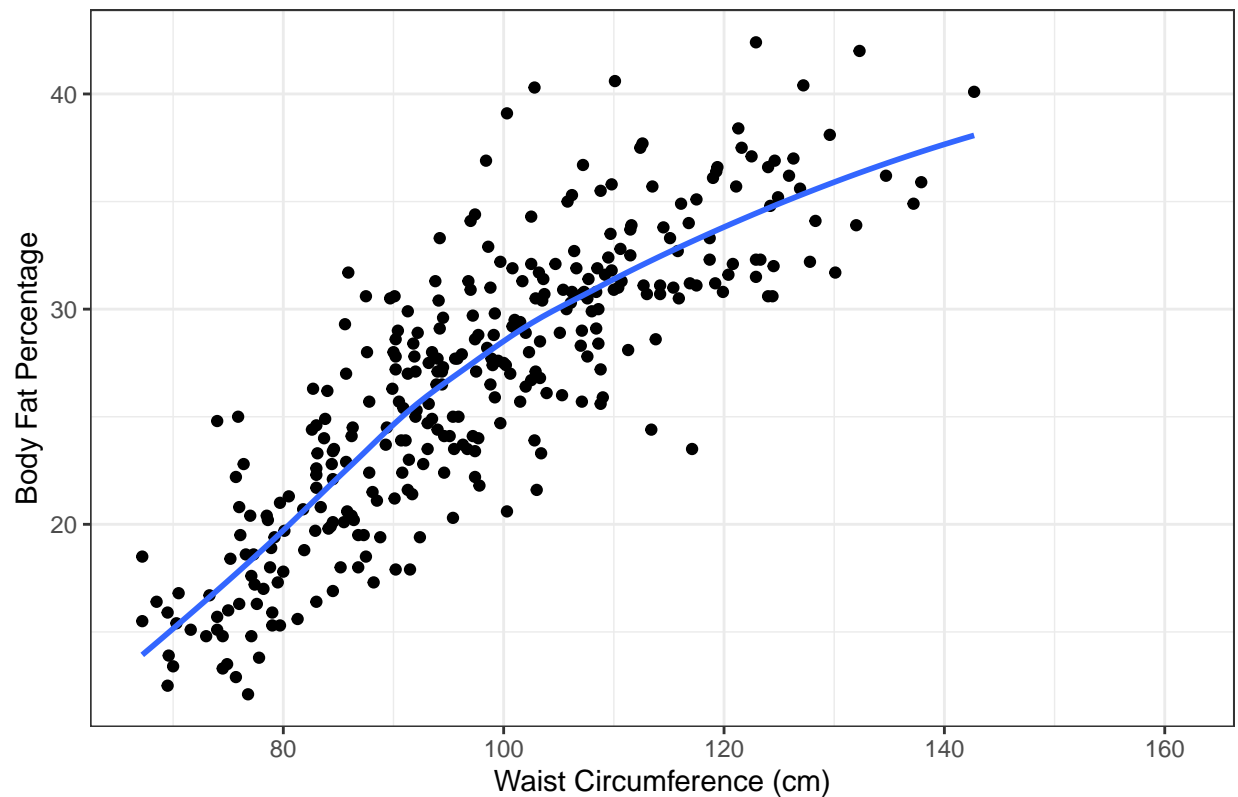

```
ggplot(total %>% filter(RIAGENDR == 2), aes(x = BMXWAIST, y = DXDTOPF)) + geom_point() + stat_smooth(se
```
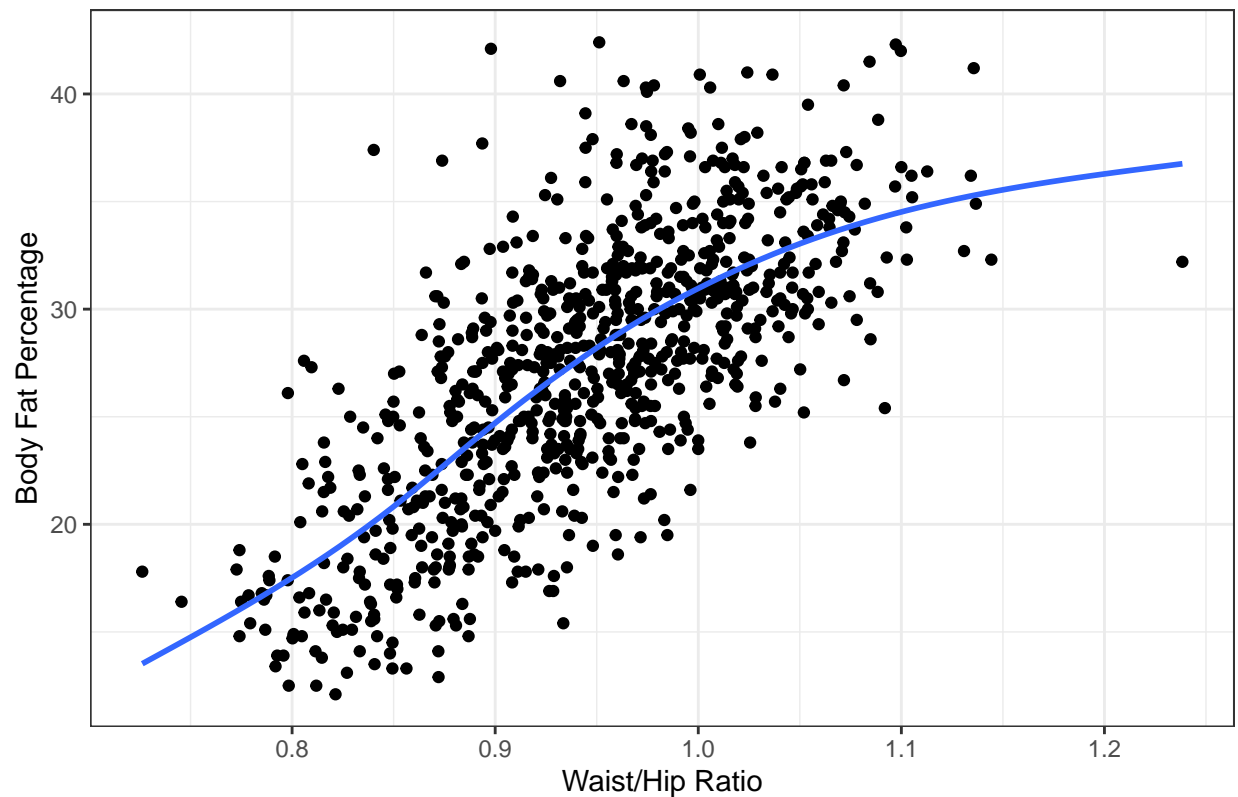
```
## 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 485 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 485 rows containing missing values ('geom_point()').
```

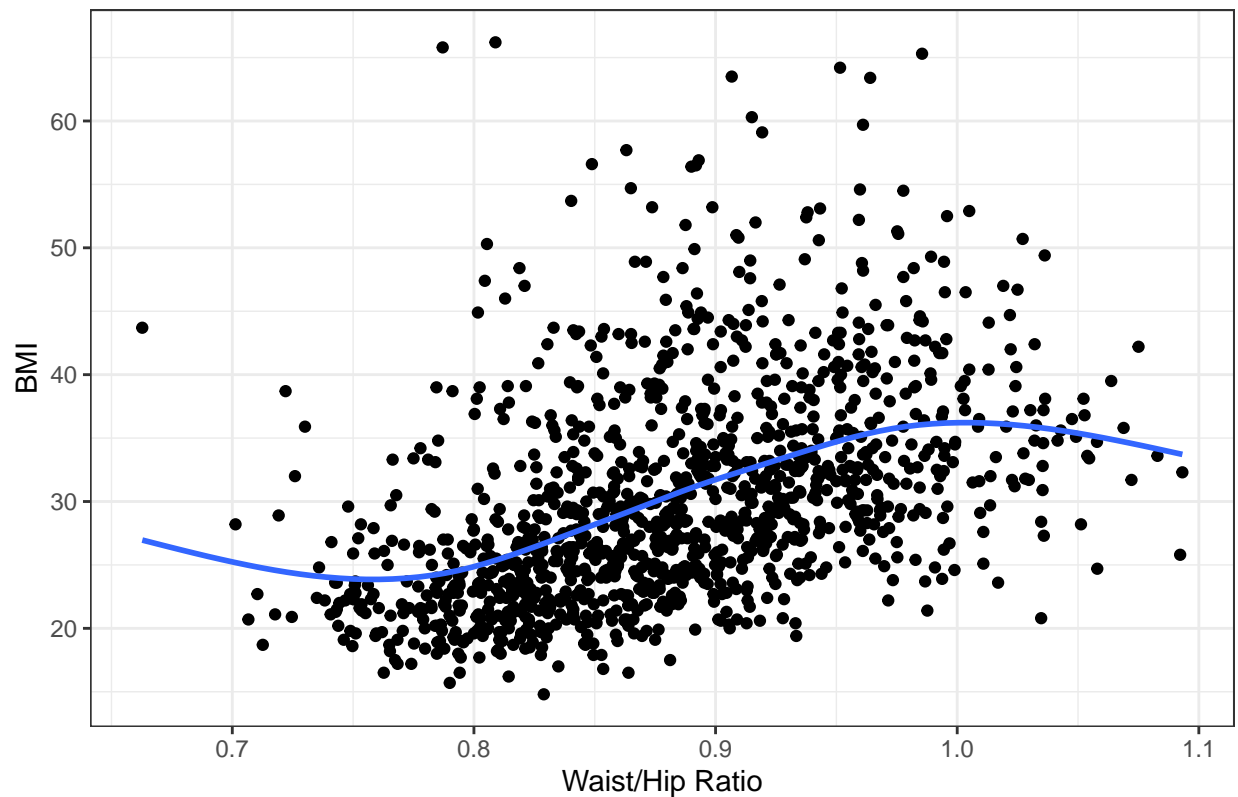# Waist Circumference and Body Fat Percentage for Females Age 20–50


```r
ggplot(total %>% filter(RIAGENDR == 2) %>% mutate(Ratio = BMXWAIST/BMXHIP), aes(x = Ratio, y = DXDTOPF)]
```

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

## Warning: Removed 485 rows containing non-finite values ('stat_smooth()').

## Warning: Removed 485 rows containing missing values ('geom_point()').
```

Waist/Hip Ratio and Body Fat Percentage for Females Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(Ratio = BMXWAIST/BMXHIP), aes(x = Ratio, y = DXDTOPF))
```

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

```
## Warning: Removed 410 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 410 rows containing missing values ('geom_point()').
```

Waist/Hip Ratio and Body Fat Percentage for Males Age 20–50

```
ggplot(total %>% filter(RIAGENDR == 2) %>% mutate(Ratio = BMXWAIST/BMXHIP), aes(x = Ratio, y = BMXBMI))
```

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

## Warning: Removed 139 rows containing non-finite values ('stat_smooth()').

## Warning: Removed 139 rows containing missing values ('geom_point()').
```
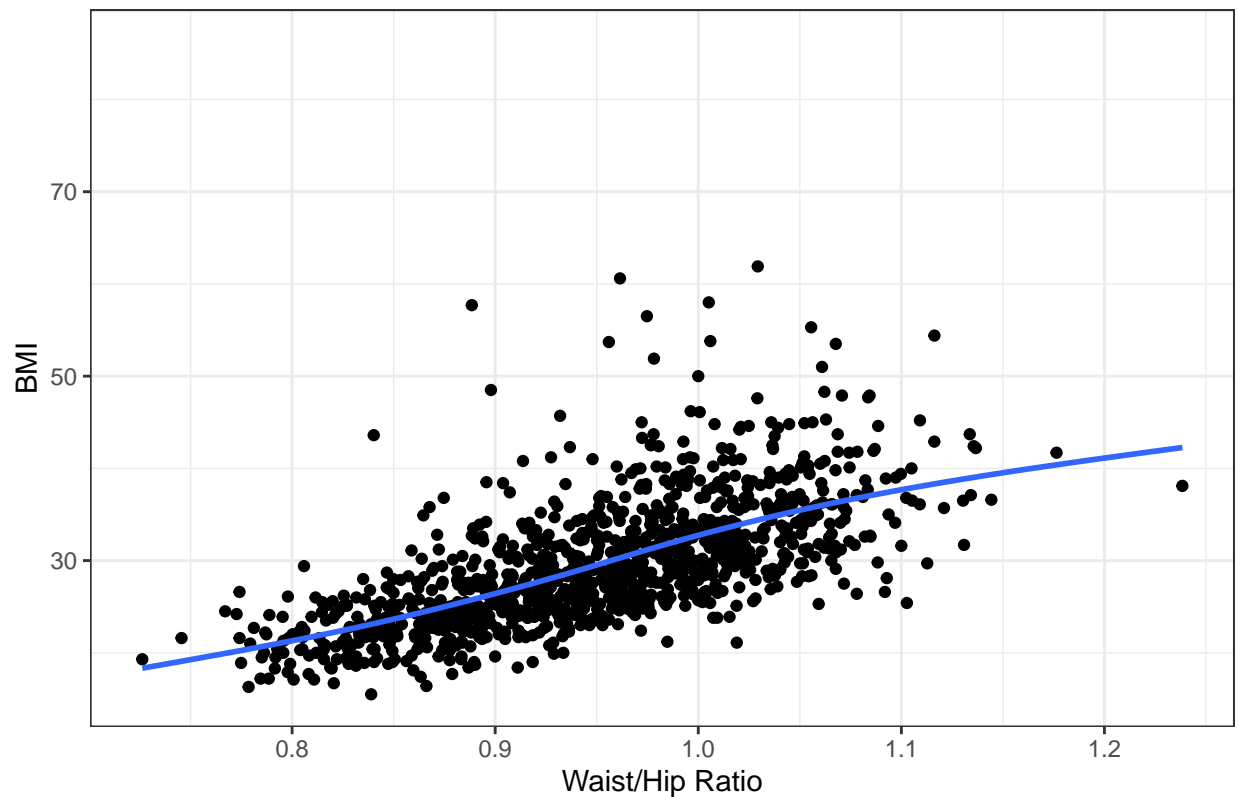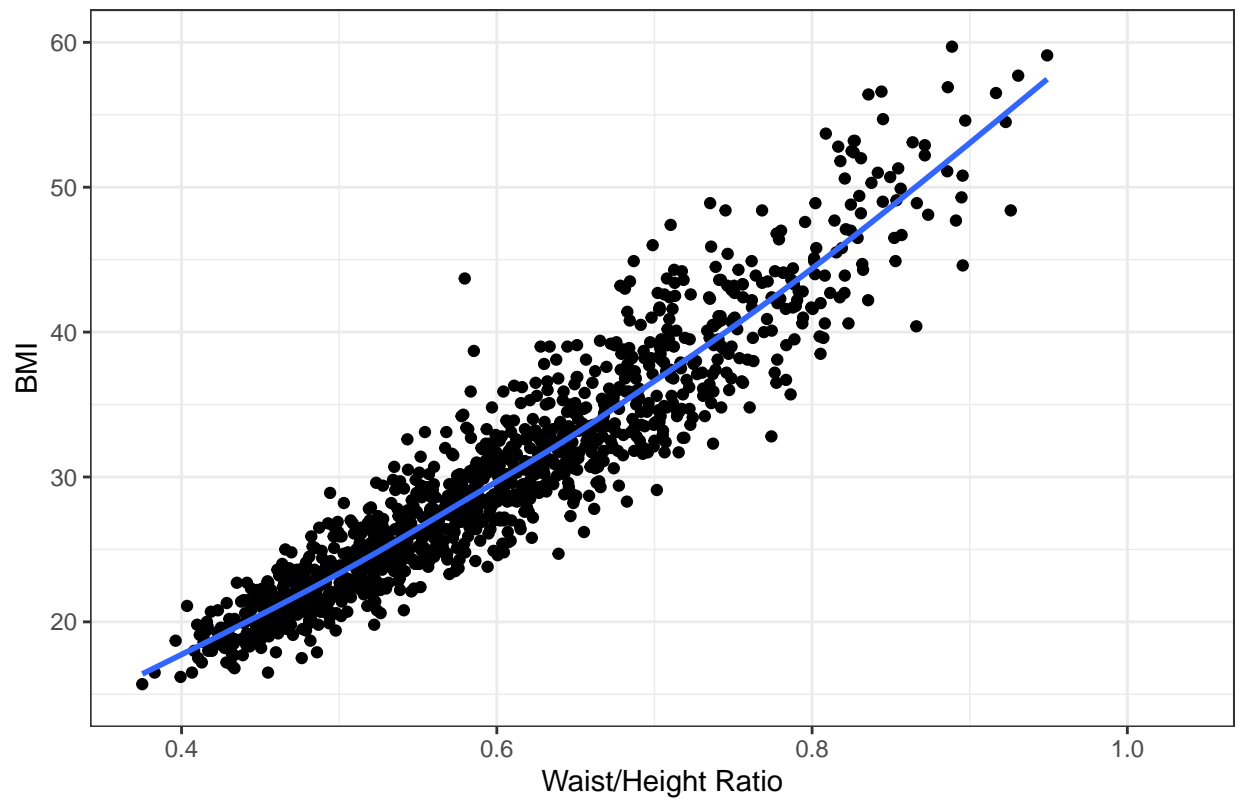
Waist/Hip Ratio and BMI for Females Age 20–50


```r
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(Ratio = BMXWAIST/BMXHIP), aes(x = Ratio, y = BMXBMI))
```

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

## Warning: Removed 128 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 128 rows containing missing values (`geom_point()`).
```

Waist/Hip Ratio and BMI for Males Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 2) %>% mutate(Ratio = BMXWAIST/BMXHT), aes(x = Ratio, y = BMXBMI))
```

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

```
## Warning: Removed 147 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 147 rows containing missing values ('geom_point()').
```
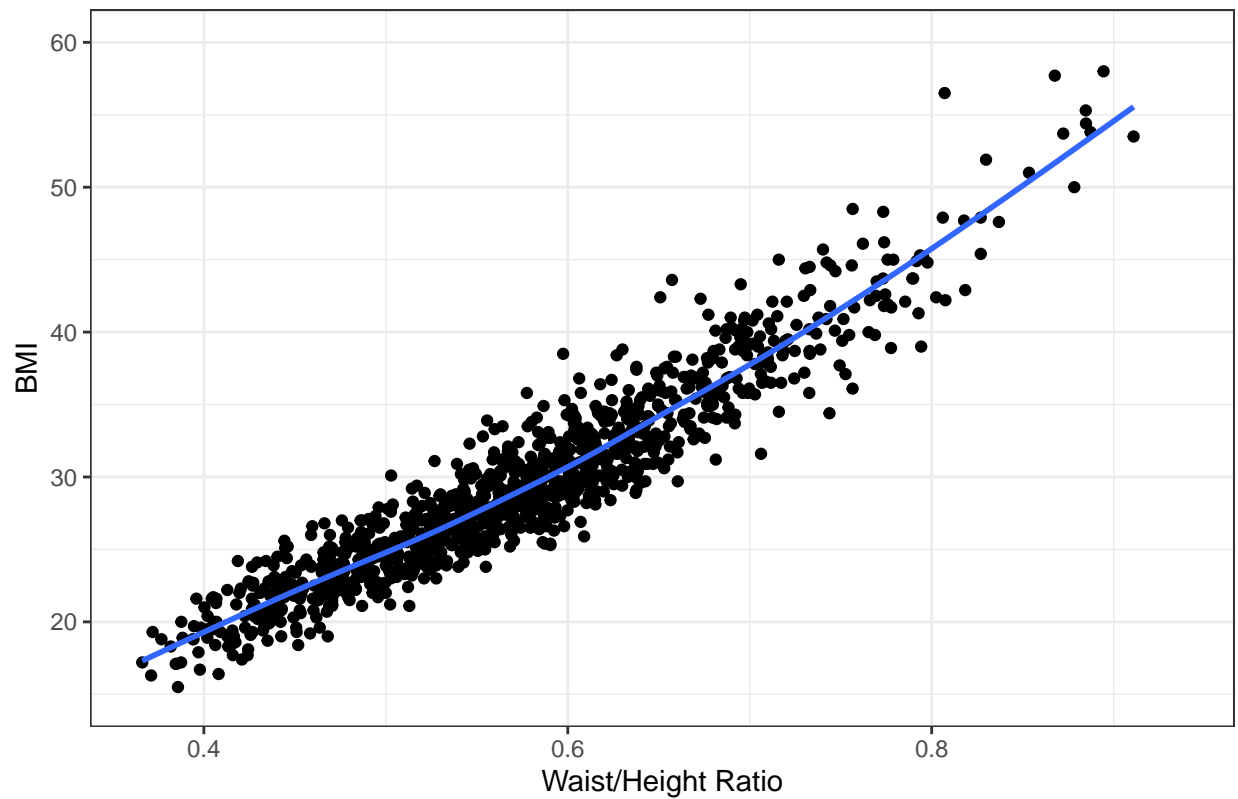
Waist/Height Ratio and BMI for Females Age 20–50
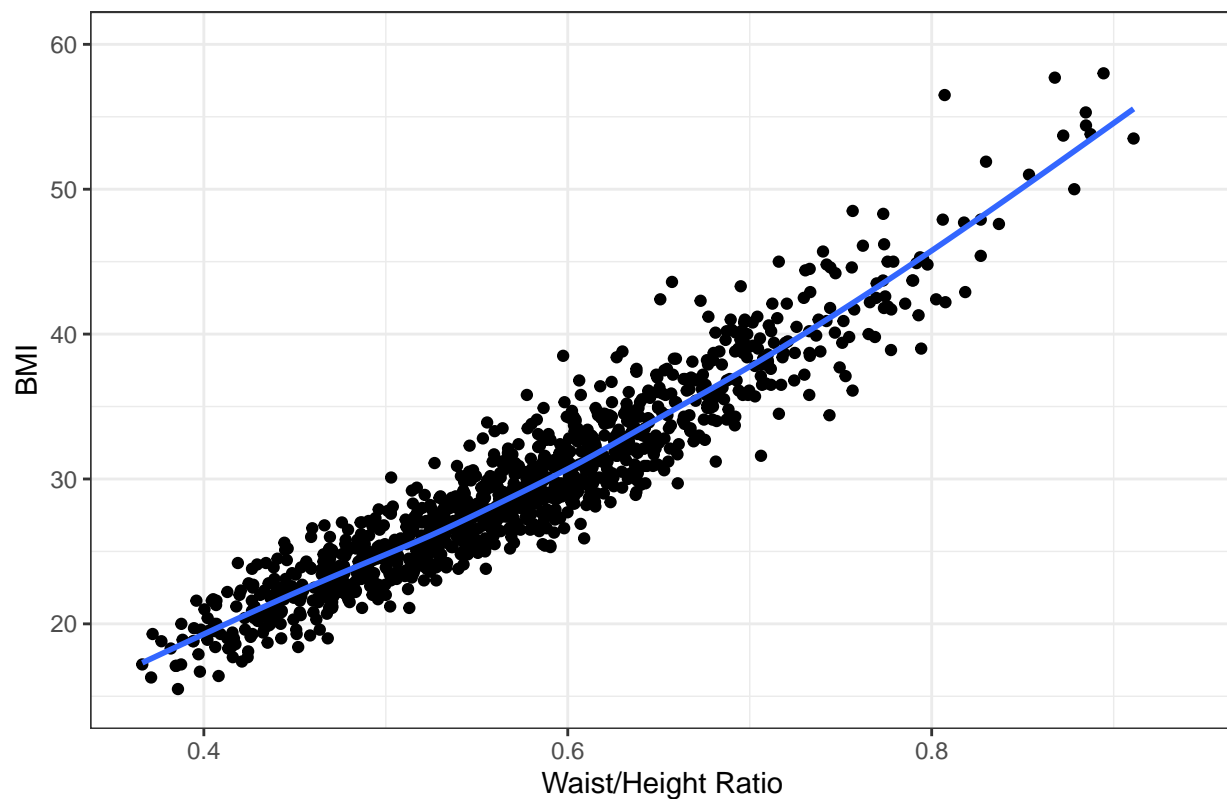


```
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(Ratio = BMXWAIST/BMXHT), aes(x = Ratio, y = BMXBMI))
```

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

```
## Warning: Removed 128 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 128 rows containing missing values ('geom_point()').
```

Waist/Height Ratio and BMI for Males Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(Ratio = BMXWAIST/BMXHT), aes(x = Ratio, y = BMXBMI))
```

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

```
## Warning: Removed 128 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 128 rows containing missing values ('geom_point()').
```

Waist/Height Ratio and BMI for Males Age 20–50


```r
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(Ratio = BMXWAIST/BMXHT, FFMI = BMXBMI * (100 - DXDTOP
```

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

## Warning: Removed 411 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 411 rows containing missing values (`geom_point()`).
```

Fat Free Mass Index and Waist/Height Ratio for Males Age 20–50
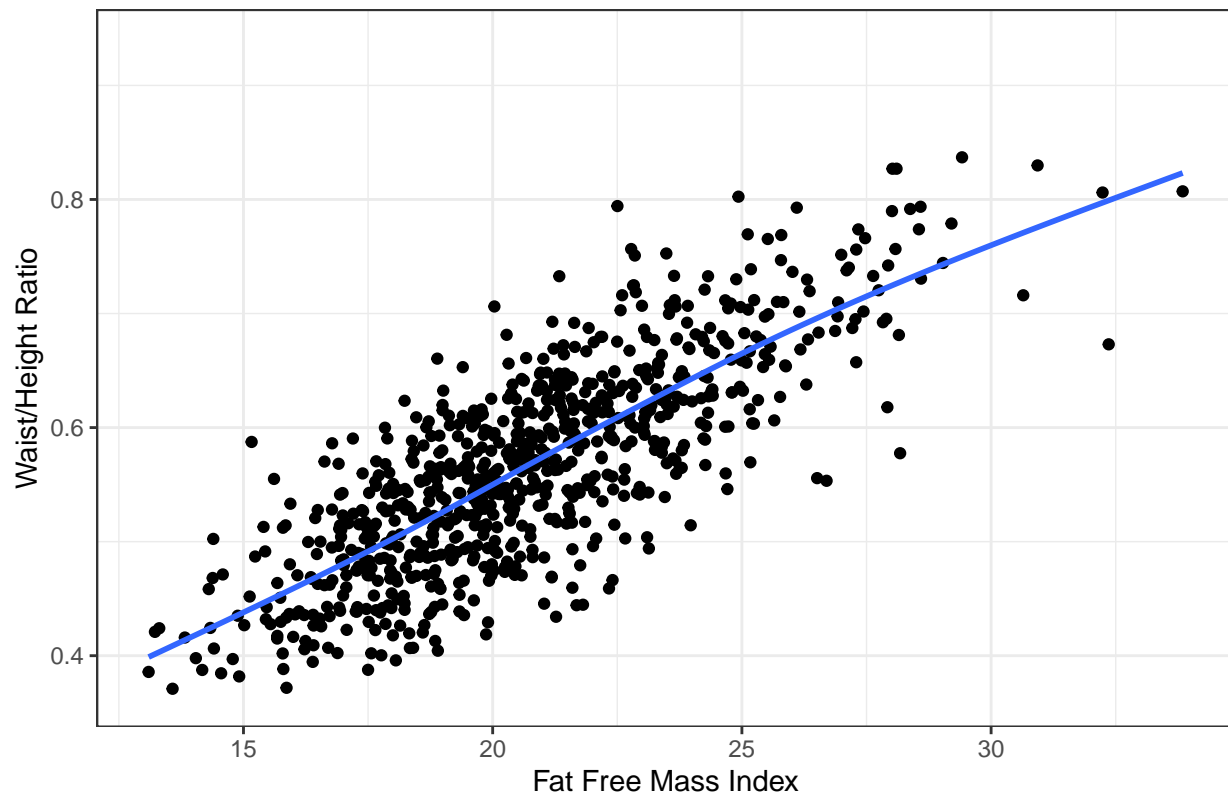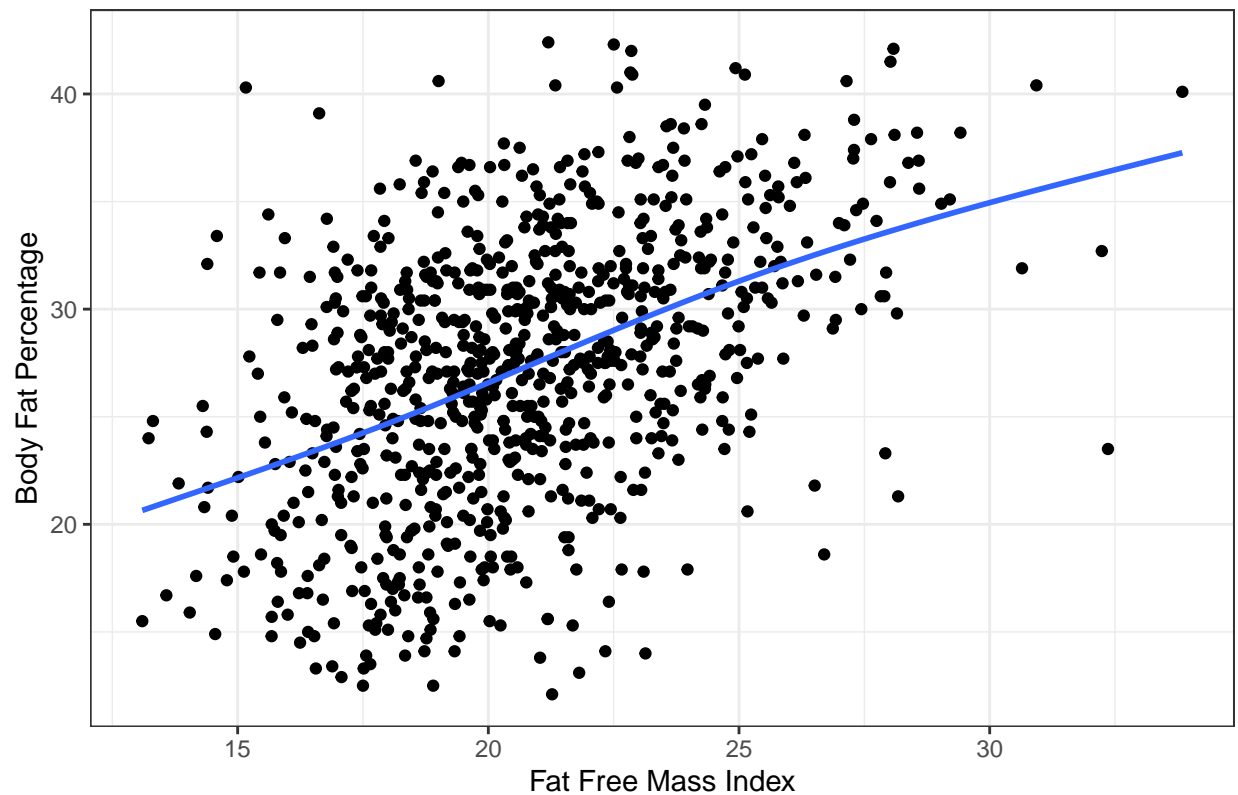


```
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(FFMI = BMXBMI * (100 - DXDTOPF)/100), aes(x = FFMI, y
```

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

```
## Warning: Removed 410 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 410 rows containing missing values ('geom_point()').
```

Fat Free Mass Index and Body Fat Percentage for Males Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 1) %>% mutate(FFMI = BMXBMI * (100 - DXDTOPF)/100), aes(x = FFMI, y
```
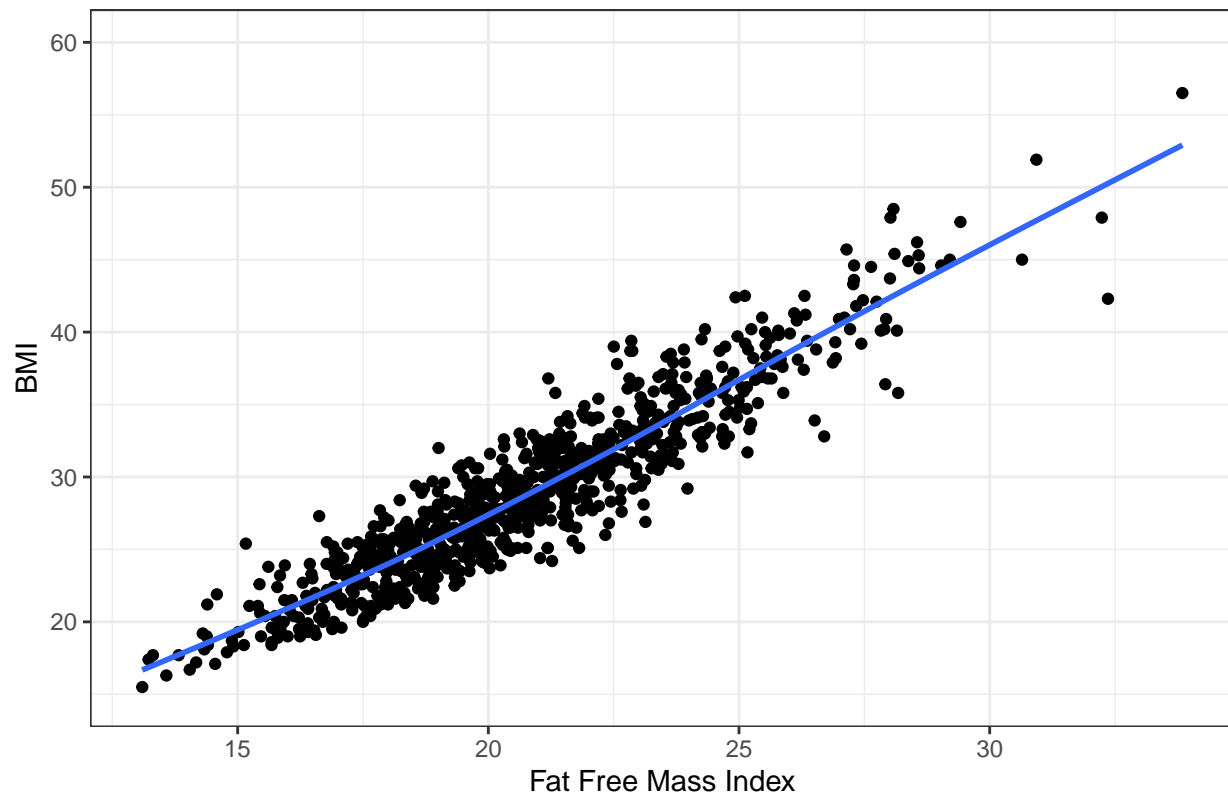
```
## 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 410 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 410 rows containing missing values ('geom_point()').
```

Fat Free Mass Index and BMI for Males Age 20–50


```
get_bf = function(waist, height, hip, race = 6, gender){
  predict(gg, data.frame(BMXWAIST = waist * 2.54, BMXHT = height * 2.54, RIDRETH3 = race, BMXHIP = hip
}
```
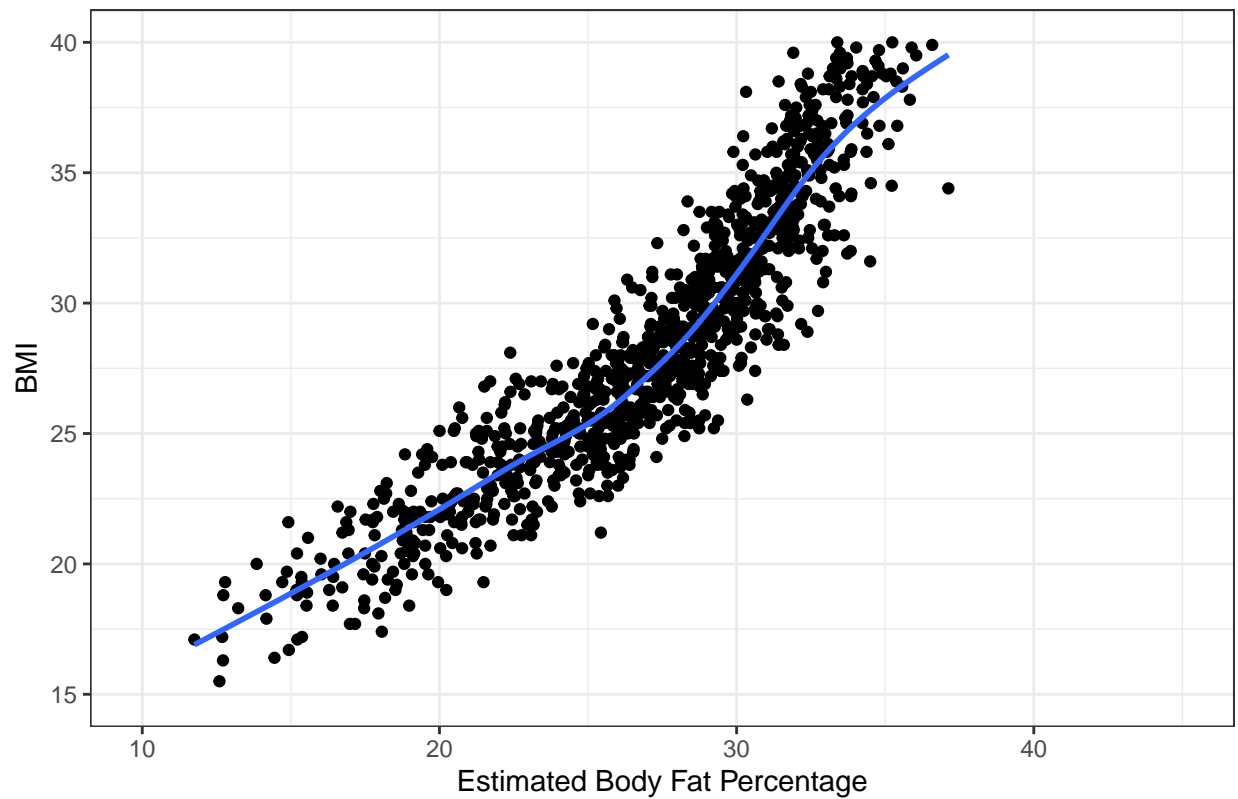
BMI vs Body Fat Percentage Estimated

```
total$DXDTOPF = total$BodyFatPredict
ggplot(total %>% filter(RIAGENDR == 1), aes(x = DXDTOPF, BMXBMI)) + geom_point() + stat_smooth(se = FALS
```

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

```
## Warning: Removed 208 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 208 rows containing missing values ('geom_point()').
```

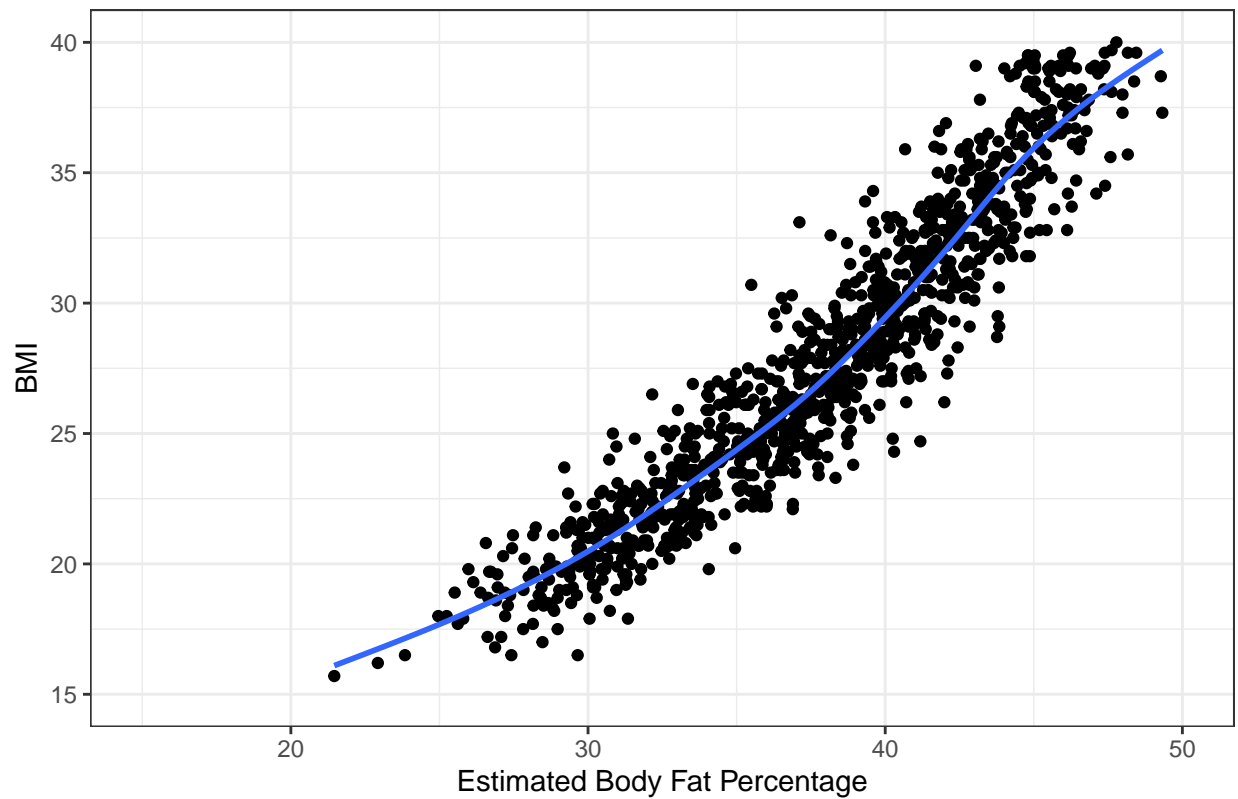## Estimated Body Fat Percentage and BMI for Males Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 2), aes(x = DXDTOPF, BMXBMI)) + geom_point() + stat_smooth(se = FALS
```

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

```
## Warning: Removed 302 rows containing non-finite values ('stat_smooth()').
```
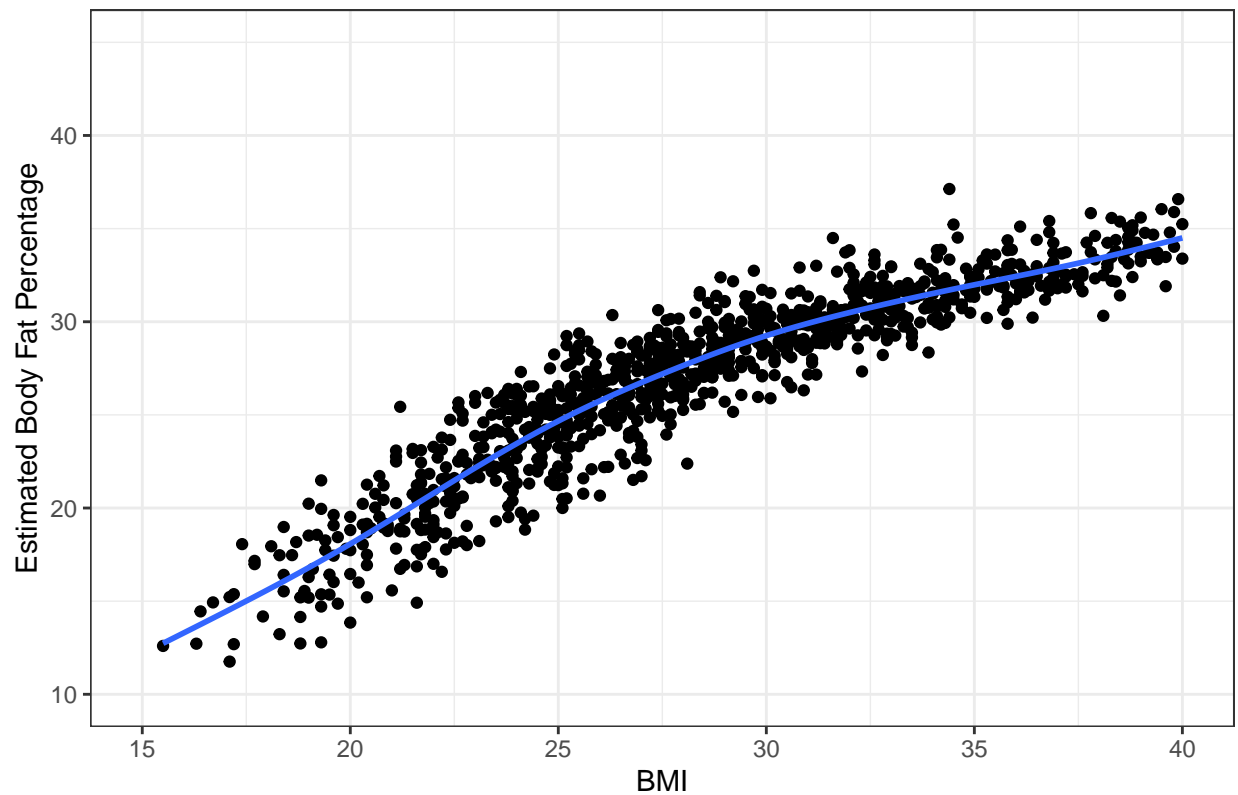
```
## Warning: Removed 302 rows containing missing values ('geom_point()').
```

Estimated Body Fat Percentage and BMI for Females Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 1), aes(y = DXDTOPF, x = BMXBMI)) + geom_point() + stat_smooth(se =
```

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

```
## Warning: Removed 208 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 208 rows containing missing values ('geom_point()').
```

BMI and Estimated Body Fat Percentage for Males Age 20–50


```
ggplot(total %>% filter(RIAGENDR == 2), aes(y = DXDTOPF, x = BMXBMI)) + geom_point() + stat_smooth(se =
```

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

```
## Warning: Removed 302 rows containing non-finite values ('stat_smooth()').
```

```
## Warning: Removed 302 rows containing missing values ('geom_point()').
```

BMI and Estimated Body Fat Percentage for Females Age 20–50