```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------- tidyverse 1.2.1 --

## v ggplot2 3.1.0       v purrr   0.3.2
## v tibble  2.1.1       v dplyr   0.8.0.1
## v tidyr   0.8.3       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0

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

```
library(tidymodels)
```

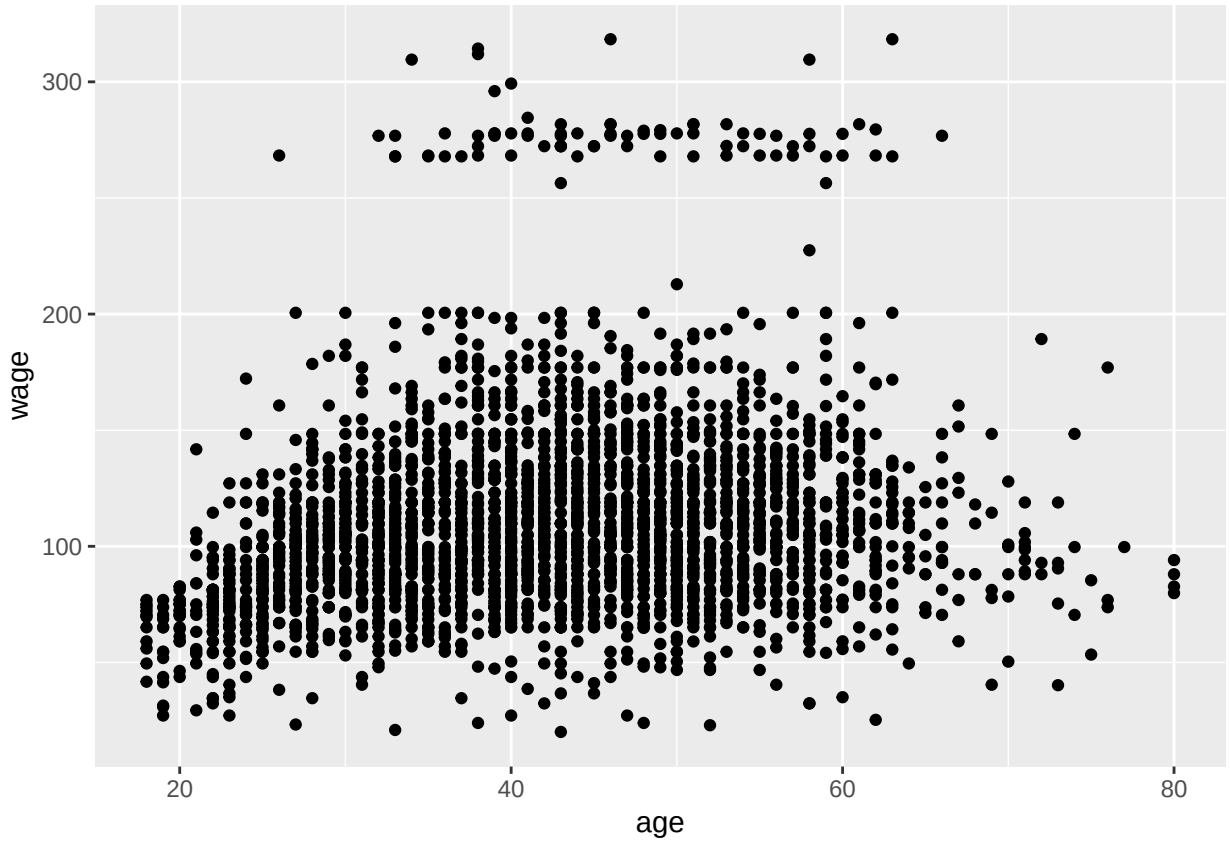
```
## -- Attaching packages ---------------------------- tidymodels 0.0.2 --

## v broom     0.5.1     v recipes   0.1.4
## v dials     0.0.2     v rsample   0.0.4
## v infer     0.4.0     v yardstick 0.0.3
## v parsnip   0.0.1

## -- Conflicts ------------------------------- tidymodels_conflicts() --
## x scales::discard() masks purrr::discard()
## x dplyr::filter()   masks stats::filter()
## x recipes::fixed()  masks stringr::fixed()
## x dplyr::lag()      masks stats::lag()
## x yardstick::spec() masks readr::spec()
## x recipes::step()   masks stats::step()
```

```
library(splines)
library(gam)
```

```
## Loading required package: foreach

##
## Attaching package: 'foreach'

## The following objects are masked from 'package:purrr':
##
##     accumulate, when

## Loaded gam 1.16
```

```
library(ISLR)
```

```
ggplot(Wage) + geom_point(aes(age, wage))
```



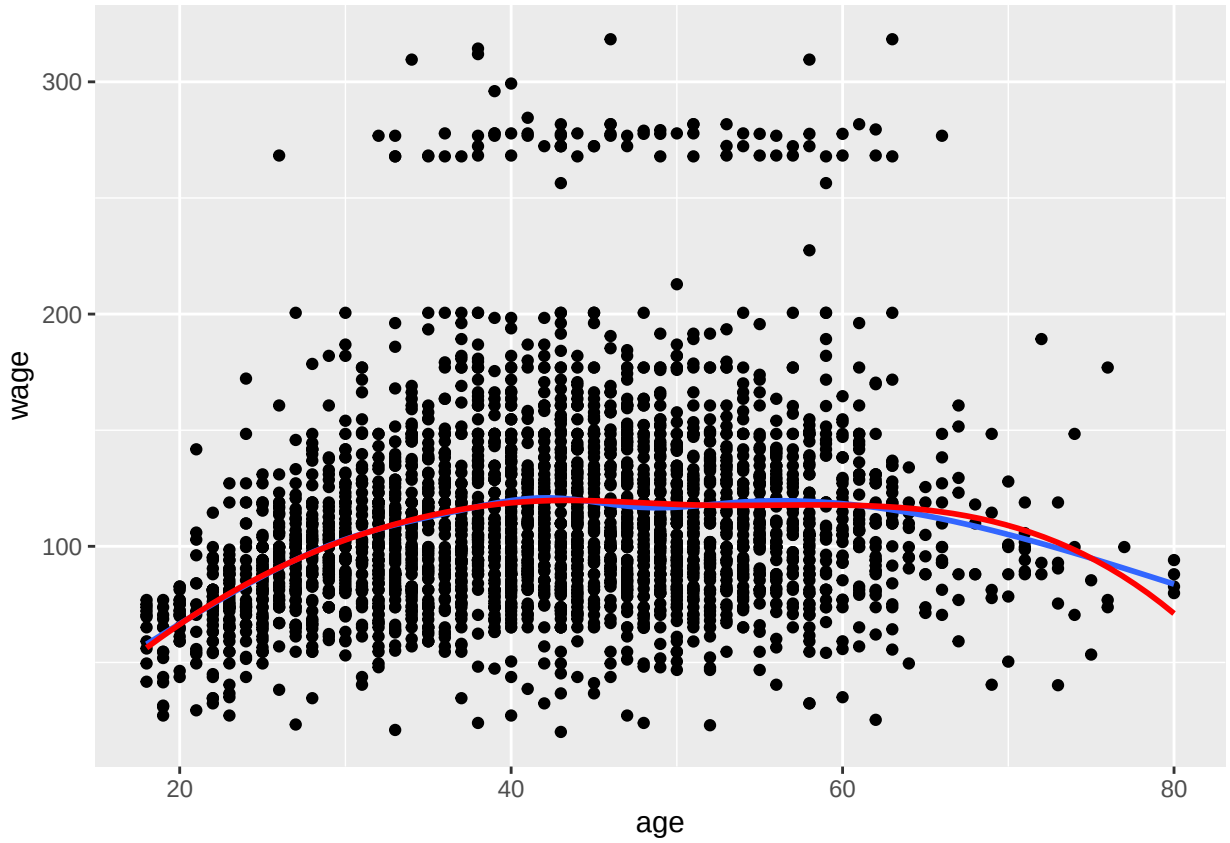
Cubic spline

```
fit_bs <- lm(wage ~ bs(age, knots = c(30, 45, 60)), data = Wage)
fit_bs2 <- lm(wage ~ bs(age, df = 10), data = Wage)
```

Natural cubic spline

```
fit_ns <- lm(wage ~ ns(age, knots = c(30, 45, 60)), data = Wage)
fit_ns2 <- lm(wage ~ ns(age, df = 10), data = Wage)
```

```
ggplot(Wage, aes(age, wage)) +
    geom_point() +
    geom_smooth(method = "lm", formula = y ~ ns(x, df = 6), se = FALSE) +
    geom_smooth(method = "lm", formula = y ~ bs(x, df = 6), color = "red", se = FALSE)
```



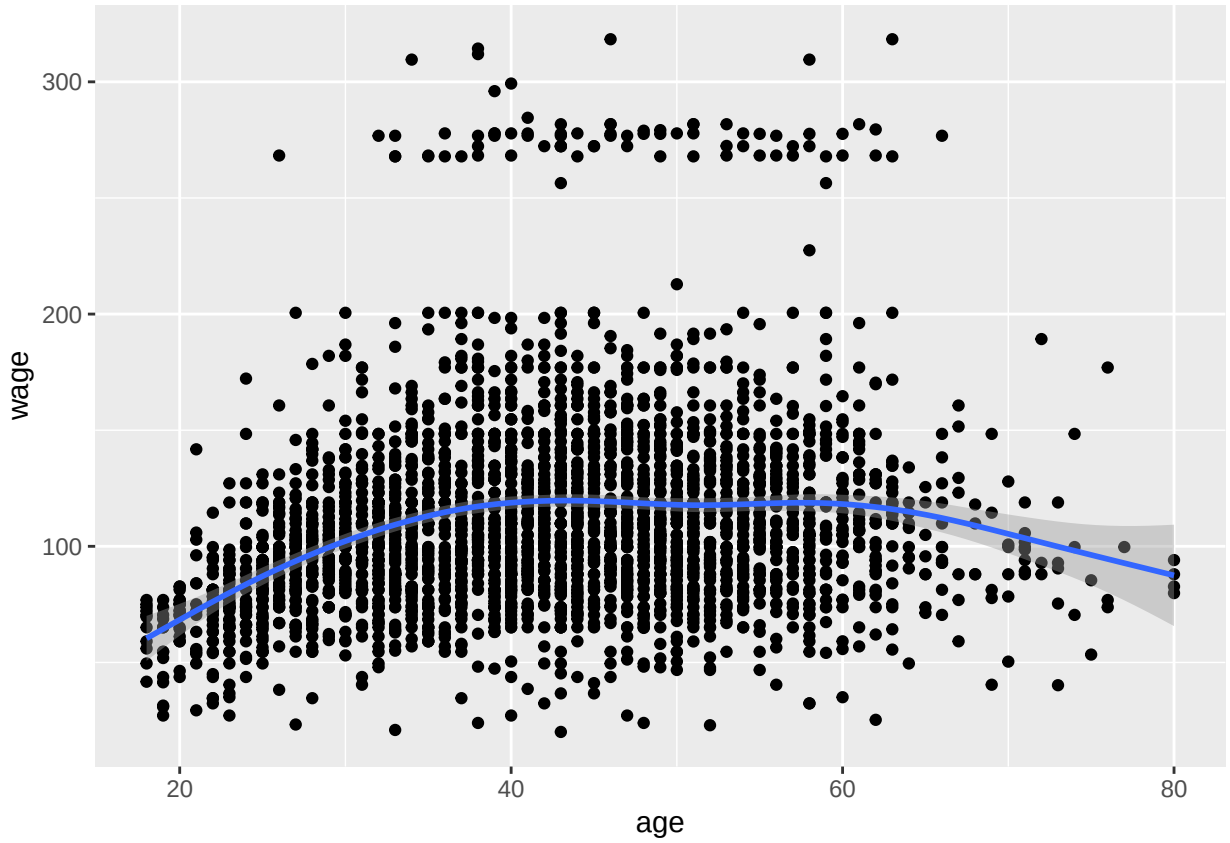
Smoothing spline

```
fit_smooth <- smooth.spline(Wage$age, Wage$wage, cv = TRUE)
```

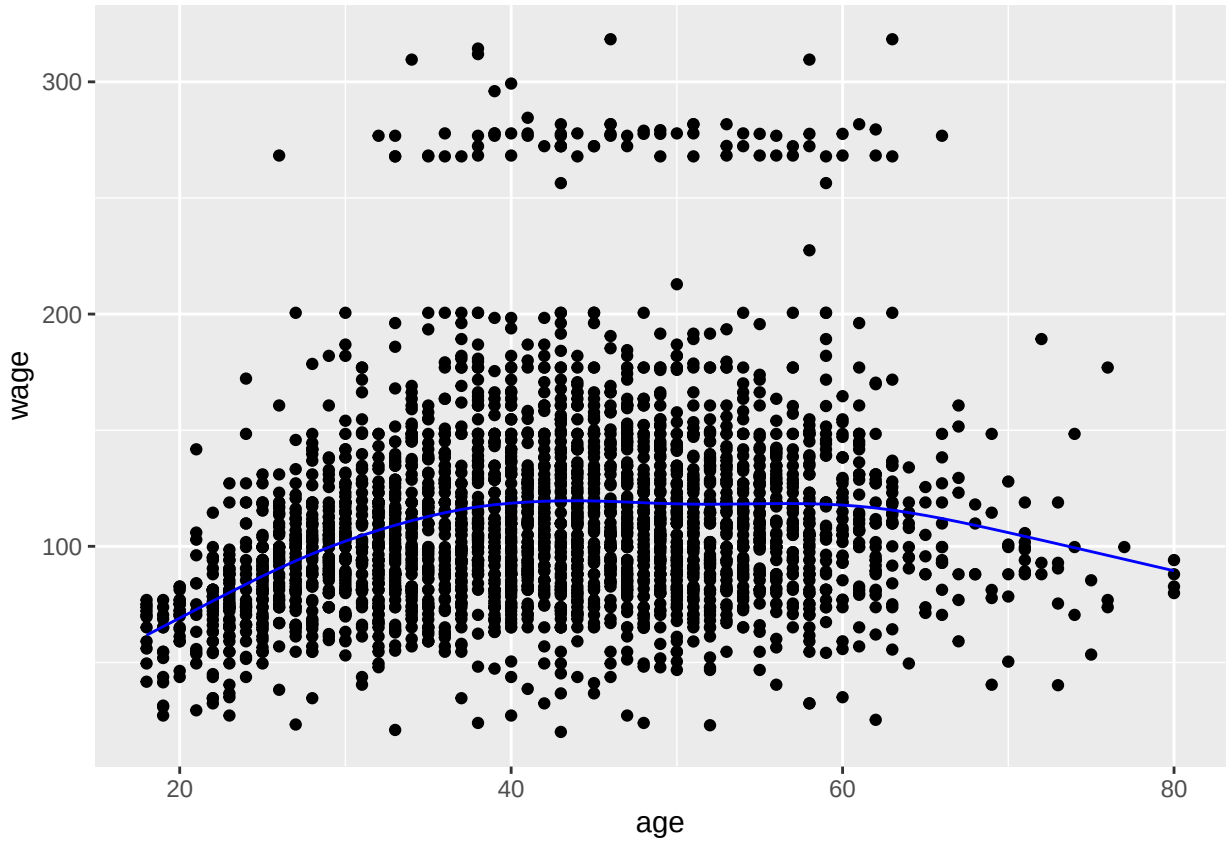
```
## Warning in smooth.spline(Wage$age, Wage$wage, cv = TRUE): cross-validation
## with non-unique 'x' values seems doubtful
```

```
ggplot(Wage, aes(age, wage)) +
    geom_point() +
    geom_smooth()
```

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

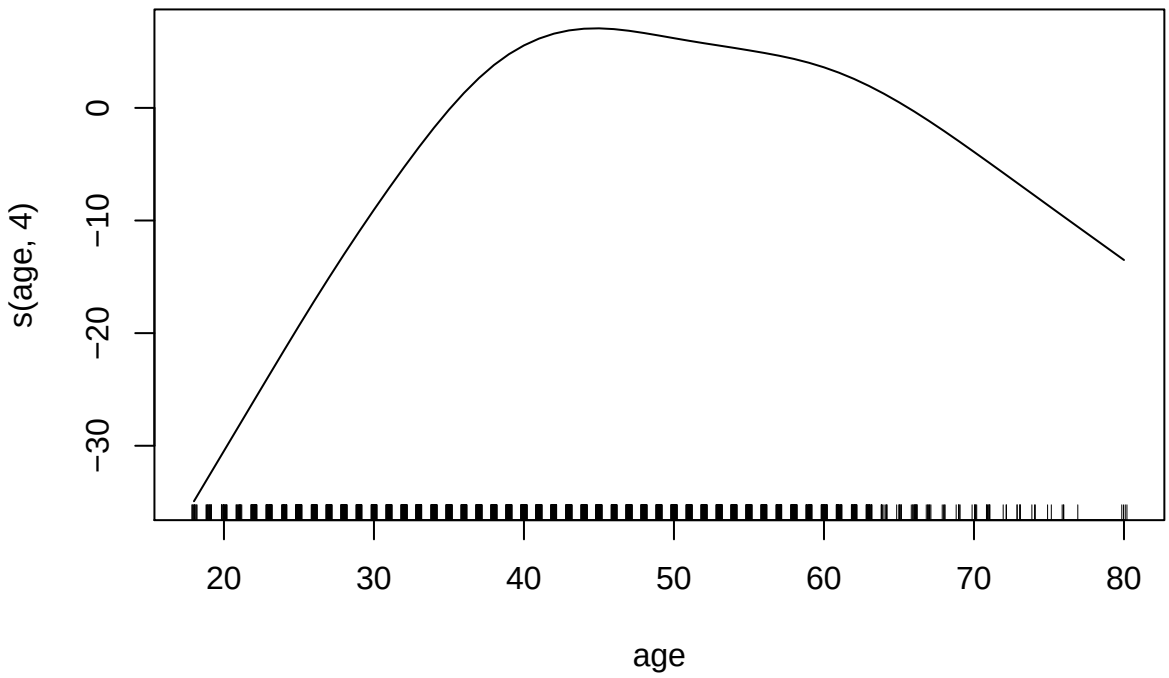


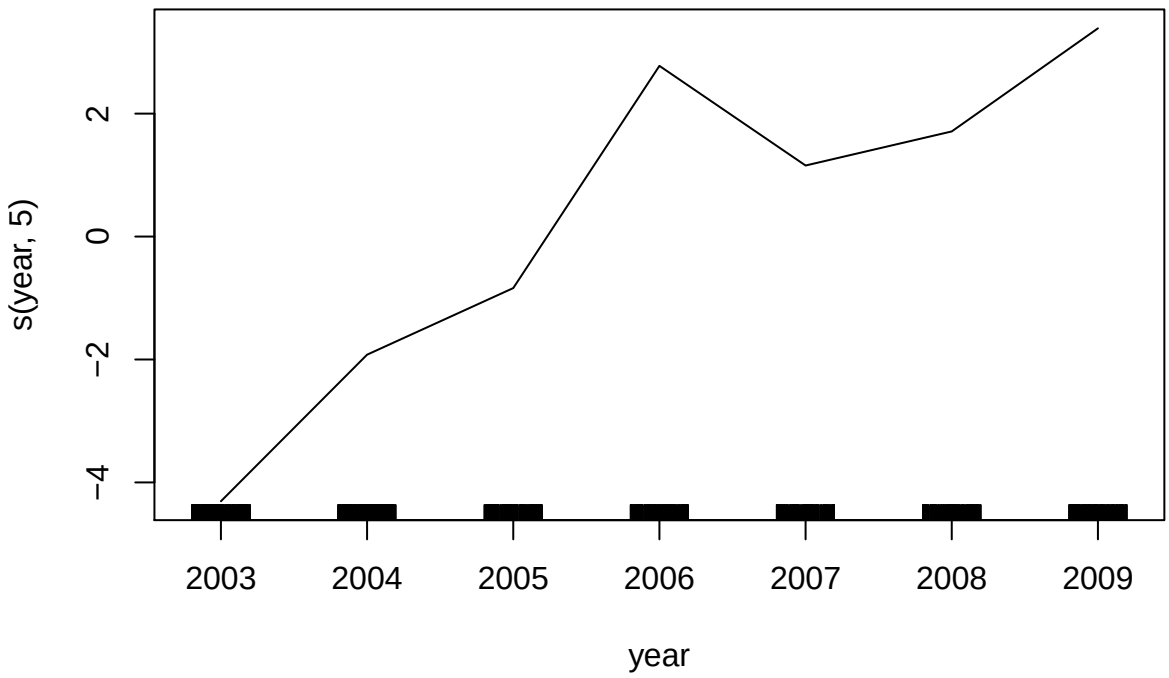

```
# with a specific df
fit_smooth <- smooth.spline(Wage$age, Wage$wage, df = 6)
Wage2 <- Wage %>% mutate(pred = predict(fit_smooth, age)$y)
ggplot(Wage2, aes(age, wage)) +
    geom_point() +
    geom_line(aes(age, pred), color = "blue")
```

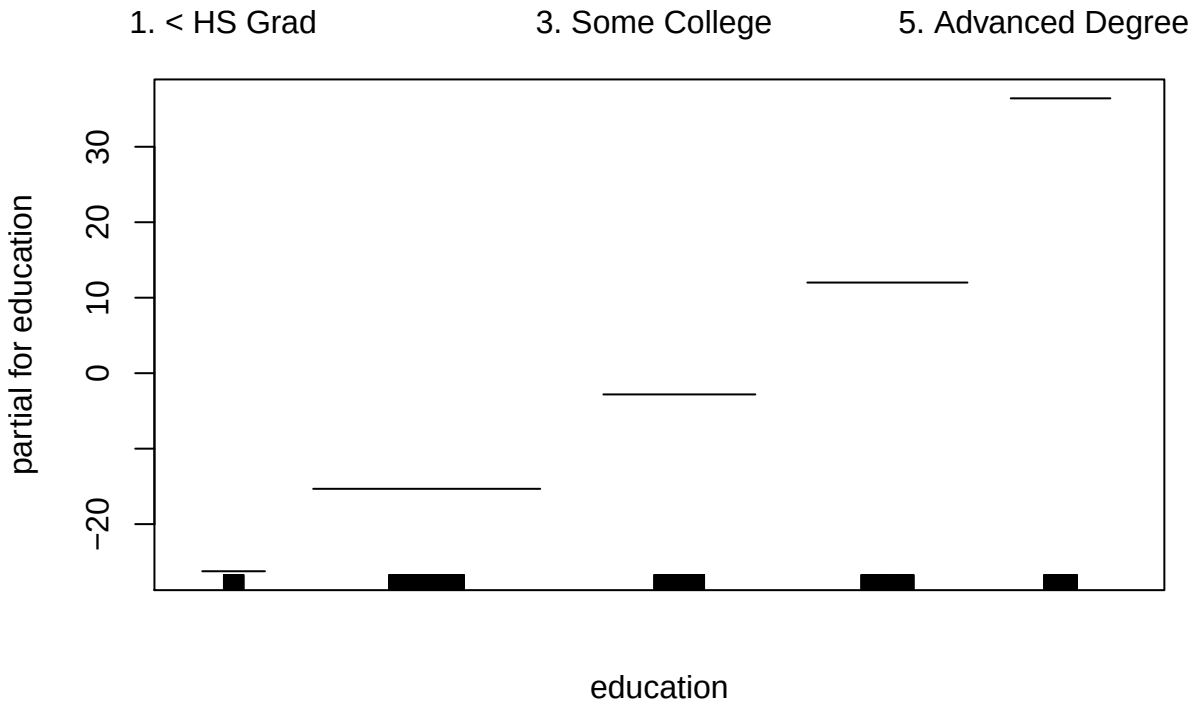


GAM

```
fit_gam <- gam(wage ~ s(age, 4) + s(year, 5) + education, data = Wage)
plot(fit_gam)
```


s(age, 4)
0
−10
−20
−30
20
30
40
50
60
70
80
age


s(year, 5)
2
0
−2
−4
2003
2004
2005
2006
2007
2008
2009
year


1. < HS Grad
3. Some College
5. Advanced Degree
partial for education
−20
0
10
20
30
education