Homework #6

Eric Pettengill

```
library(tidyverse)
library(ISLR)
library(splines)
library(broom)
library(ggformula)
library(patchwork)
```

(1) Cubic spline with 3 knots at 25th, 50th, and 75th percentiles.

```
splines.fit <- lm(mpg ~ bs(horsepower, df = 6), data = Auto)

glance(splines.fit)
```

```
##   r.squared adj.r.squared    sigma statistic      p.value df    logLik
## 1 0.7020142     0.6973703 4.293674   151.168 5.646848e-98  7 -1123.892
##        AIC      BIC deviance df.residual
## 1 2263.785 2295.555 7097.721         385
```

(2) Natural spline with 3 knots at 25th, 50th, and 75th percentile

```
natural.splines.fit <- lm(mpg ~ ns(horsepower, df = 4), data = Auto)

glance(natural.splines.fit)
```

```
##   r.squared adj.r.squared    sigma statistic      p.value df    logLik
## 1 0.6956801     0.6925346 4.327842   221.172 1.434613e-98  5 -1128.015
##       AIC      BIC deviance df.residual
## 1 2268.03 2291.857 7248.594         387
```

(3) Smoothing spline with lambda chosen by cross validation

```
smooth.fit <- smooth.spline(Auto$horsepower, Auto$mpg, cv = FALSE)

glance(smooth.fit)
```

```
##         df       lambda  cv.crit pen.crit     crit      spar     lambda.1
## 1 42.14987 5.101567e-07 18.64132 1117.657 18.64132 0.2648644 5.101567e-07
```

(4) Local regression with span = 0.75

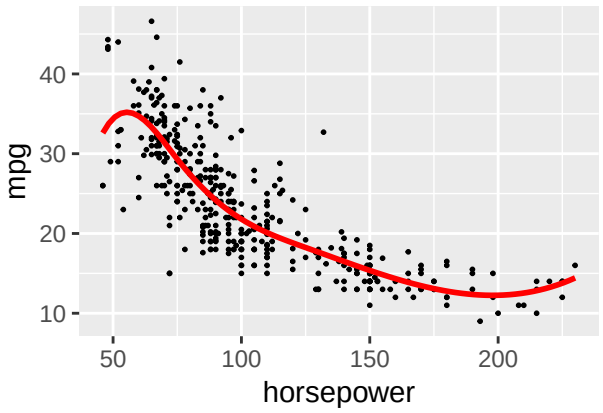
```
local.fit <- loess(mpg ~ horsepower, data = Auto, span = 0.75)

summary(local.fit)
```

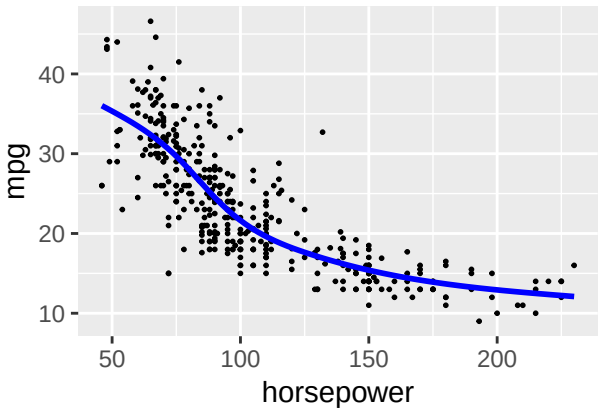
```
## Call:
## loess(formula = mpg ~ horsepower, data = Auto, span = 0.75)
##
## Number of Observations: 392
## Equivalent Number of Parameters: 4.97
## Residual Standard Error: 4.32
## Trace of smoother matrix: 5.43  (exact)
##
## Control settings:
##   span     :  0.75
##   degree   :  2
##   family   :  gaussian
##   surface  :  interpolate      cell = 0.2
##   normalize:  TRUE
##  parametric:  FALSE
## drop.square:  FALSE
```

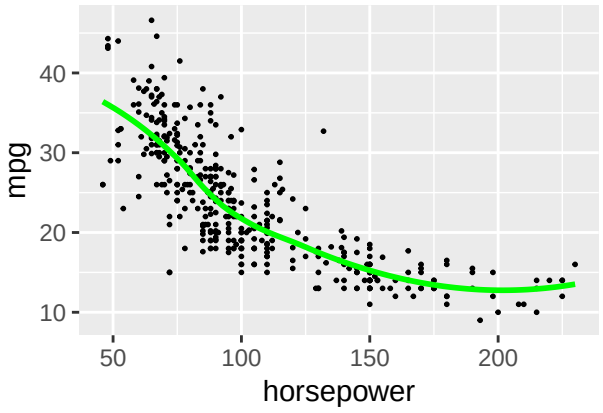
Plot

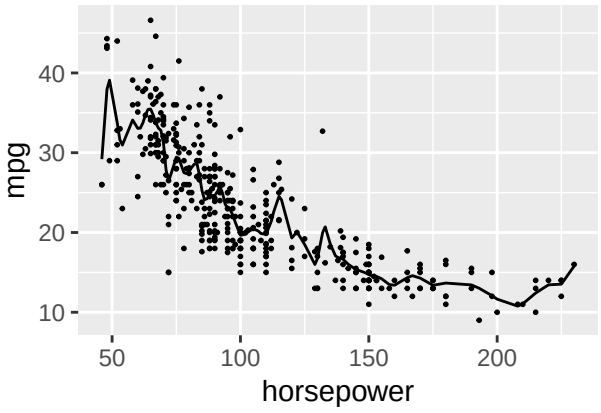
Each methods fit is plotted below. The cubic spline in red, natural spline in blue, local regression in green, and the smoothing spline in black.

```
ns <- ggplot(Auto, aes(x = horsepower, y = mpg)) +
  geom_point(size = 1/3) +
  geom_smooth(method = lm, formula = y ~ ns(x, df = 4), color = "blue", se = FALSE)

cs <- ggplot(Auto, aes(x = horsepower, y = mpg)) +
  geom_point(size = 1/3) +
  geom_smooth(method = lm, formula = y ~ bs(x, df = 6), color = "red", se = FALSE)

loc <- ggplot(Auto, aes(x = horsepower, y = mpg)) +
  geom_point(size = 1/3) +
  geom_smooth(method = loess, span = 0.75, se = FALSE, color = "green")

ss <- ggplot(Auto, aes(x = horsepower, y = mpg)) +
  geom_point(size = 1/3) +
  geom_spline(df = smooth.fit$df)

(cs + ns)/(loc + ss)
```


mpg
40
30
20
10
50
100
150
200
horsepower


mpg
40
30
20
10
50
100
150
200
horsepower


mpg
40
30
20
10
50
100
150
200
horsepower


mpg
40
30
20
10
50
100
150
200
horsepower