

Interactions, 1

```
library(ISLR)
Auto$foreign=ifelse(Auto$origin==1,0,1)
fit.linear <- lm(mpg ~ horsepower + weight + foreign + year, data=Auto)
summary(fit.linear)
```

```
Call:
lm(formula = mpg ~ horsepower + weight + foreign + year, data = Auto)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-9.4008 -2.1335 -0.0628  1.8284 13.6148
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.758e+01  4.143e+00  -4.242 2.77e-05 ***
horsepower  -7.541e-03  9.192e-03  -0.820  0.413
weight       -5.608e-03  4.338e-04 -12.928 < 2e-16 ***
foreign       2.116e+00  4.370e-01  4.841 1.87e-06 ***
year          7.596e-01  5.073e-02  14.975 < 2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.335 on 387 degrees of freedom
Multiple R-squared:  0.8193,    Adjusted R-squared:  0.8174
F-statistic: 438.6 on 4 and 387 DF,  p-value: < 2.2e-16
```

```
fit.inter <- lm(mpg ~ horsepower + weight + foreign + year +
foreign*weight, data=Auto)
```

```
summary(fit.inter)
```

Call:

```
lm(formula = mpg ~ horsepower + weight + foreign + year + foreign *  
    weight, data = Auto)
```

Residuals:

```
      Min       1Q   Median       3Q      Max  
-9.7988 -2.0742 -0.1083  1.6348 12.8584
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) -2.179e+01  4.102e+00  -5.313 1.83e-07 ***  
horsepower  -1.190e-02  8.953e-03  -1.330  0.184  
weight       -4.876e-03  4.446e-04 -10.969 < 2e-16 ***  
foreign       1.091e+01  1.784e+00   6.114 2.38e-09 ***  
year          7.896e-01  4.953e-02  15.942 < 2e-16 ***  
weight:foreign -3.544e-03  6.987e-04  -5.073 6.09e-07 ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 3.234 on 386 degrees of freedom

Multiple R-squared: 0.8306, Adjusted R-squared: 0.8284

F-statistic: 378.4 on 5 and 386 DF, p-value: < 2.2e-16

Interactions, 2

```
fit.inter.2 <- lm(mpg ~ horsepower + weight + foreign + year +  
foreign*weight + foreign*year, data=Auto)  
summary(fit.inter.2)
```

Call:

```
lm(formula = mpg ~ horsepower + weight + foreign + year + foreign *  
weight + foreign * year, data = Auto)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.8995	-1.9335	-0.0871	1.5693	12.0042

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-7.263e+00	5.182e+00	-1.402	0.16184	
horsepower	-1.531e-02	8.780e-03	-1.743	0.08205	.
weight	-5.067e-03	4.364e-04	-11.611	< 2e-16	***
foreign	-2.163e+01	7.562e+00	-2.861	0.00446	**
year	6.113e-01	6.299e-02	9.704	< 2e-16	***
weight:foreign	-3.346e-03	6.839e-04	-4.892	1.47e-06	***
foreign:year	4.167e-01	9.424e-02	4.422	1.28e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.158 on 385 degrees of freedom

Multiple R-squared: 0.8388, Adjusted R-squared: 0.8362

F-statistic: 333.8 on 6 and 385 DF, p-value: < 2.2e-16

```
anova(fit.linear, fit.inter, fit.inter.2)
```

Analysis of Variance Table

Model 1: mpg ~ horsepower + weight + foreign + year

Model 2: mpg ~ horsepower + weight + foreign + year + foreign * weight

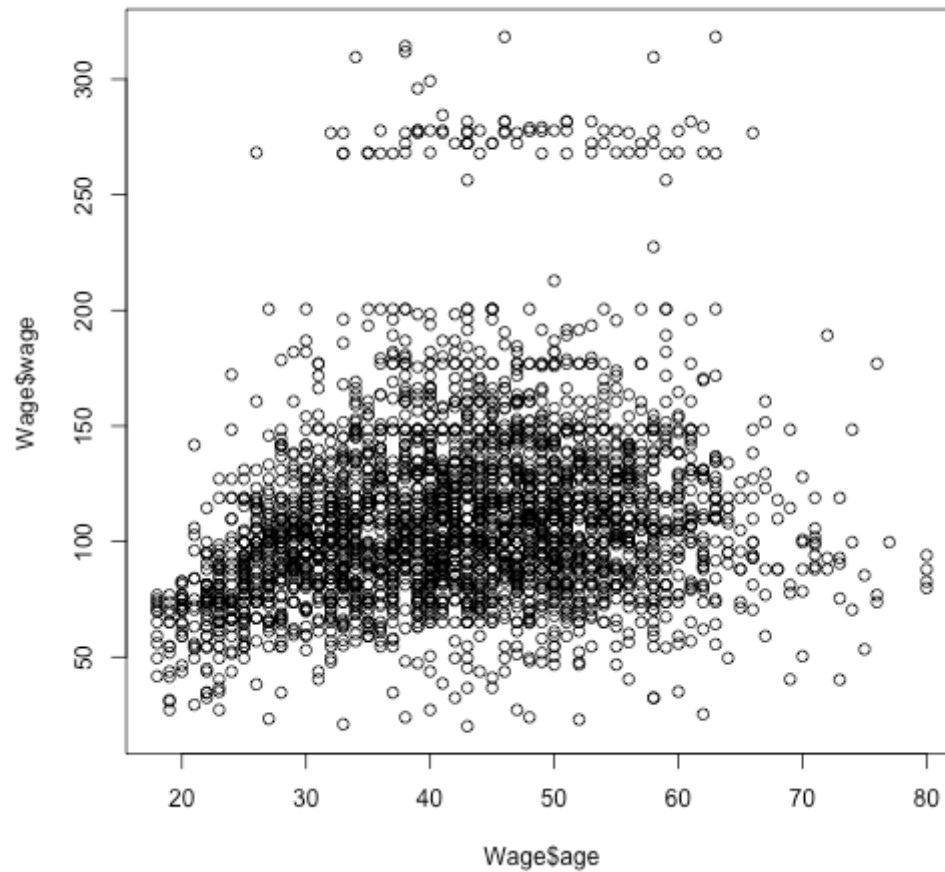
Model 3: mpg ~ horsepower + weight + foreign + year + foreign * weight +
foreign * year

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	387	4304.9					
2	386	4035.8	1	269.10	26.975	3.348e-07	***
3	385	3840.8	1	195.05	19.552	1.275e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Polynomials, 1

```
library(ISLR)  
plot(Wage$age, Wage$wage)
```



```
fit.poly1=lm(wage~age,data=Wage)
summary(fit.poly1)
```

Call:

```
lm(formula = wage ~ age, data = Wage)
```

Residuals:

Min	1Q	Median	3Q	Max
-100.265	-25.115	-6.063	16.601	205.748

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	81.70474	2.84624	28.71	<2e-16 ***
age	0.70728	0.06475	10.92	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.93 on 2998 degrees of freedom

Multiple R-squared: 0.03827, Adjusted R-squared: 0.03795

F-statistic: 119.3 on 1 and 2998 DF, p-value: < 2.2e-16

```
fit.poly2=lm(wage~poly(age,2),data=Wage)
summary(fit.poly2)
```

Call:

```
lm(formula = wage ~ poly(age, 2), data = Wage)
```

Residuals:

Min	1Q	Median	3Q	Max
-99.126	-24.309	-5.017	15.494	205.621

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
----------	------------	---------	----------

```

(Intercept)    111.7036      0.7302   152.99   <2e-16 ***
poly(age, 2)1  447.0679     39.9926    11.18   <2e-16 ***
poly(age, 2)2 -478.3158     39.9926   -11.96   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.99 on 2997 degrees of freedom
Multiple R-squared:  0.08209,    Adjusted R-squared:  0.08147
F-statistic:   134 on 2 and 2997 DF,  p-value: < 2.2e-16

```

```

fit.poly4=lm(wage~poly(age,4),data=Wage)
summary(fit.poly4)

```

```

Call:
lm(formula = wage ~ poly(age, 4), data = Wage)

Residuals:
    Min       1Q   Median       3Q      Max
-98.707 -24.626  -4.993   15.217  203.693

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    111.7036     0.7287  153.283 < 2e-16 ***
poly(age, 4)1  447.0679     39.9148   11.201 < 2e-16 ***
poly(age, 4)2 -478.3158     39.9148  -11.983 < 2e-16 ***
poly(age, 4)3  125.5217     39.9148    3.145  0.00168 **
poly(age, 4)4  -77.9112     39.9148   -1.952  0.05104 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.91 on 2995 degrees of freedom
Multiple R-squared:  0.08626,    Adjusted R-squared:  0.08504
F-statistic:   70.69 on 4 and 2995 DF,  p-value: < 2.2e-16

```



```
anova(fit.poly1, fit.poly2, fit.poly4)
```

Analysis of Variance Table

Model 1: wage ~ age

Model 2: wage ~ poly(age, 2)

Model 3: wage ~ poly(age, 4)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	2998	5022216					
2	2997	4793430	1	228786	143.6025	< 2.2e-16	***
3	2995	4771604	2	21826	6.8497	0.001076	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Polynomials, 2

```
fit.poly4=lm(wage~poly(age,4),data=Wage)
summary(fit.poly4)
```

```
Call:
lm(formula = wage ~ poly(age, 4), data = Wage)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-98.707 -24.626  -4.993  15.217 203.693
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    111.7036     0.7287  153.283 < 2e-16 ***
poly(age, 4)1   447.0679    39.9148   11.201 < 2e-16 ***
poly(age, 4)2  -478.3158    39.9148  -11.983 < 2e-16 ***
poly(age, 4)3   125.5217    39.9148    3.145  0.00168 **
poly(age, 4)4  -77.9112    39.9148   -1.952  0.05104 .
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 39.91 on 2995 degrees of freedom
Multiple R-squared:  0.08626,    Adjusted R-squared:  0.08504
F-statistic: 70.69 on 4 and 2995 DF,  p-value: < 2.2e-16
```

```
fit.poly4.I=lm(wage~age+I(age^2)+I(age^3)+I(age^4),data=Wage)
summary(fit.poly4.I)
```

Call:

```
lm(formula = wage ~ age + I(age^2) + I(age^3) + I(age^4), data = Wage)
```

Residuals:

Min	1Q	Median	3Q	Max
-98.707	-24.626	-4.993	15.217	203.693

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.842e+02	6.004e+01	-3.067	0.002180	**
age	2.125e+01	5.887e+00	3.609	0.000312	***
I(age^2)	-5.639e-01	2.061e-01	-2.736	0.006261	**
I(age^3)	6.811e-03	3.066e-03	2.221	0.026398	*
I(age^4)	-3.204e-05	1.641e-05	-1.952	0.051039	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.91 on 2995 degrees of freedom

Multiple R-squared: 0.08626, Adjusted R-squared: 0.08504

F-statistic: 70.69 on 4 and 2995 DF, p-value: < 2.2e-16

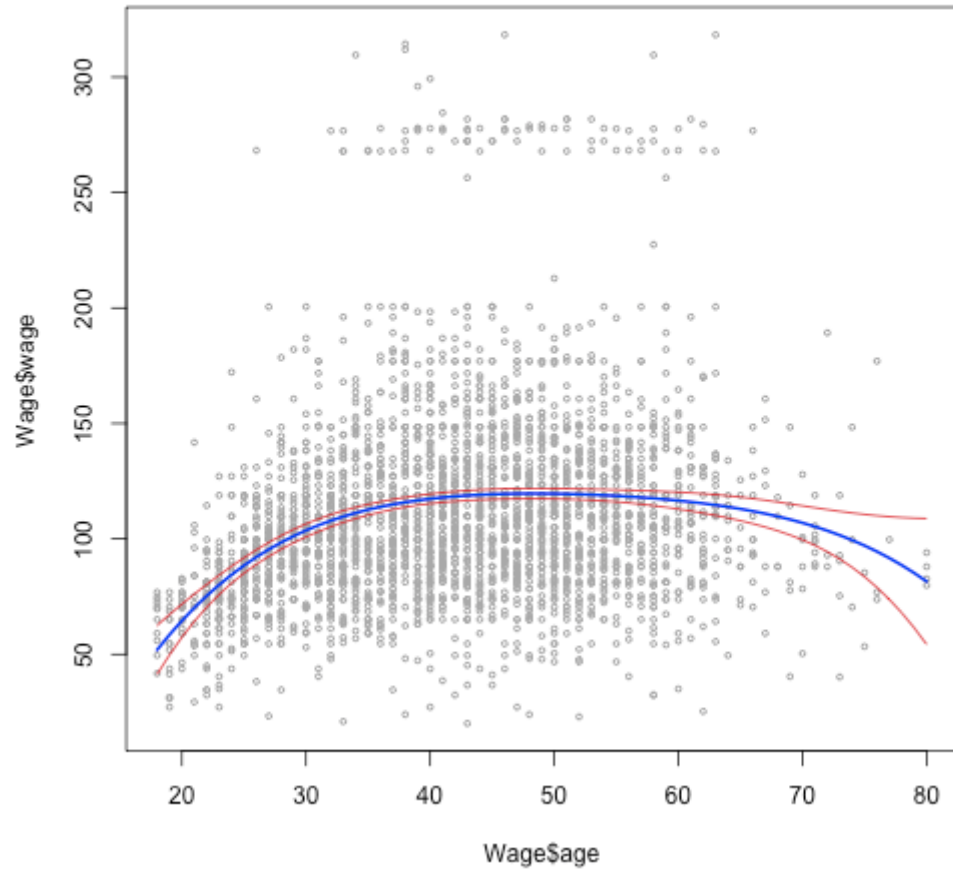
Polynomials, 3

```
agelims=range(Wage$age)
plot(Wage$age,Wage$wage,xlim=agelims,cex=.5,col="darkgrey")
title("Degree-4 Polynomial",outer=T)

age.grid=seq(from=agelims[1],to=agelims[2])
preds=predict(fit.poly4, newdata=list(age=age.grid), se=TRUE)

# Plot estimate + confidence bounds
lines(age.grid,preds$fit,lwd=2, col="blue")
lines(age.grid,preds$fit-2*preds$se.fit,lwd=1, col="red")
lines(age.grid,preds$fit+2*preds$se.fit,lwd=1, col="red")
```

Degree-4 Polynomial



Step Functions

```
plot(Wage$age,Wage$wage,xlim=agelims,cex=.5,col="darkgrey")
title("Step Regression of Wage on Age",outer=T)

# Fit model
fit.step=lm(wage~cut(age,4),data=Wage)
summary(fit.step) # Take a look
```

Call:

```
lm(formula = wage ~ cut(age, 4), data = Wage)
```

Residuals:

Min	1Q	Median	3Q	Max
-98.126	-24.803	-6.177	16.493	200.519

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	94.158	1.476	63.790	<2e-16 ***
cut(age, 4)(33.5,49]	24.053	1.829	13.148	<2e-16 ***
cut(age, 4)(49,64.5]	23.665	2.068	11.443	<2e-16 ***
cut(age, 4)(64.5,80.1]	7.641	4.987	1.532	0.126

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.42 on 2996 degrees of freedom

Multiple R-squared: 0.0625, Adjusted R-squared: 0.06156

F-statistic: 66.58 on 3 and 2996 DF, p-value: < 2.2e-16


```
      49      50      51      52      53      54      55      56
4.763992 4.763992 4.763992 4.763992 4.763992 4.763992 4.763992 4.763992
      57      58      59      60      61      62      63
4.763992 4.763992 4.763992 4.763992 4.763992 4.763992 4.763992
```

```
$df
```

```
[1] 2996
```

```
$residual.scale
```

```
[1] 40.42381
```

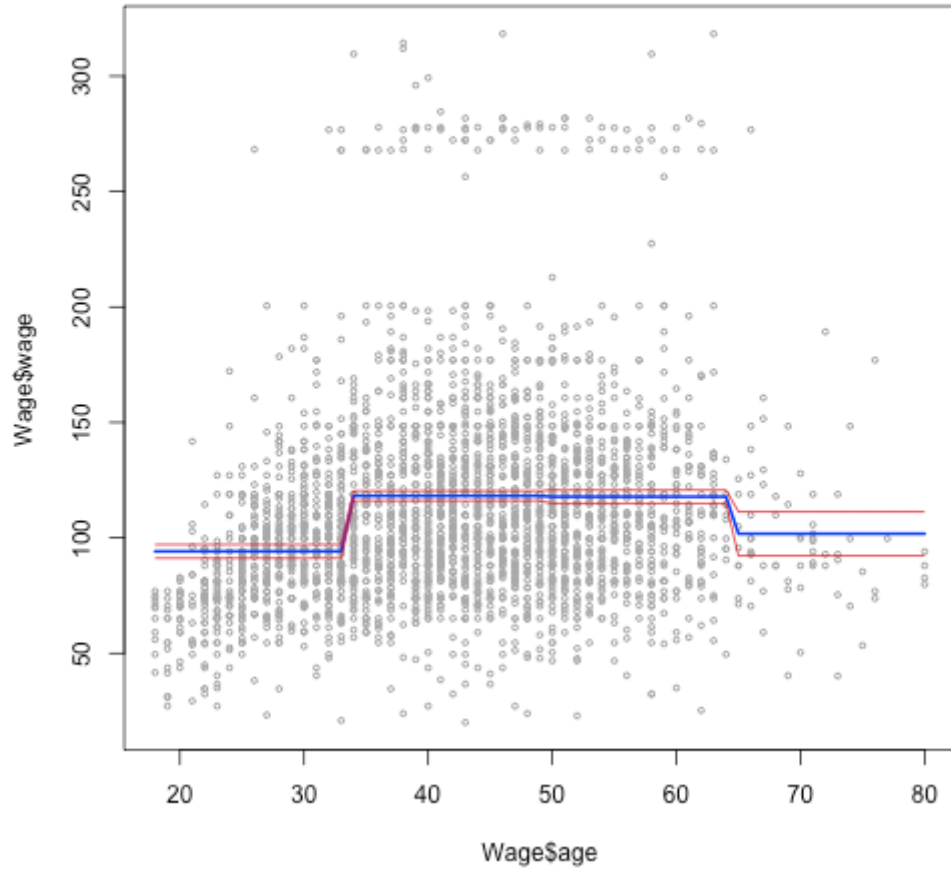
```
# Plot estimate + confidence bounds
```

```
lines(age.grid,preds$fit,lwd=2,col="blue")
```

```
lines(age.grid,preds$fit-2*preds$se.fit,lwd=1,col="red")
```

```
lines(age.grid,preds$fit+2*preds$se.fit,lwd=1,col="red")
```


Step Regression of Wage on Age



Piecewise Linear

```
plot(Wage$age,Wage$wage,xlim=agelims,cex=.5,col="darkgrey")
title("Piecewise Linear Regression of Wage on Age",outer=T)

# Fit model
fit.piecewise=lm(wage~age+
                 I((age-25)*(age>25))+
                 I((age-40)*(age>40))+
                 I((age-60)*(age>60)),
                 data=Wage)
summary(fit.piecewise)
```

Call:

```
lm(formula = wage ~ age + I((age - 25) * (age > 25)) + I((age -
  40) * (age > 40)) + I((age - 60) * (age > 60)), data = Wage)
```

Residuals:

Min	1Q	Median	3Q	Max
-99.795	-24.686	-4.856	15.344	204.671

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-42.4688	23.2695	-1.825	0.06809	.
age	5.3779	0.9739	5.522	3.64e-08	***
I((age - 25) * (age > 25))	-3.4977	1.0800	-3.239	0.00121	**
I((age - 40) * (age > 40))	-1.9800	0.2970	-6.667	3.10e-11	***
I((age - 60) * (age > 60))	-1.4041	0.5424	-2.589	0.00968	**

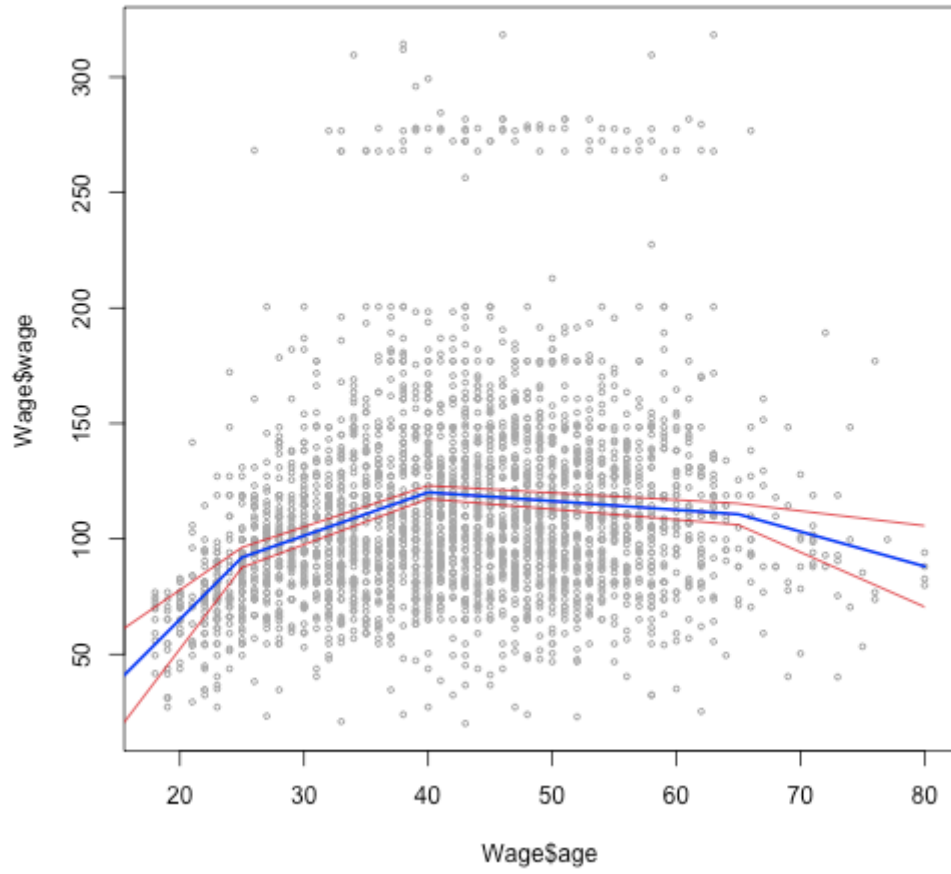
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.91 on 2995 degrees of freedom
Multiple R-squared: 0.08665, Adjusted R-squared: 0.08543
F-statistic: 71.03 on 4 and 2995 DF, p-value: < 2.2e-16

```
# Find piecewise function values
age.knots=c(0,25,40,65,80)
preds=predict(fit.piecewise,list(age=age.knots), se=T)

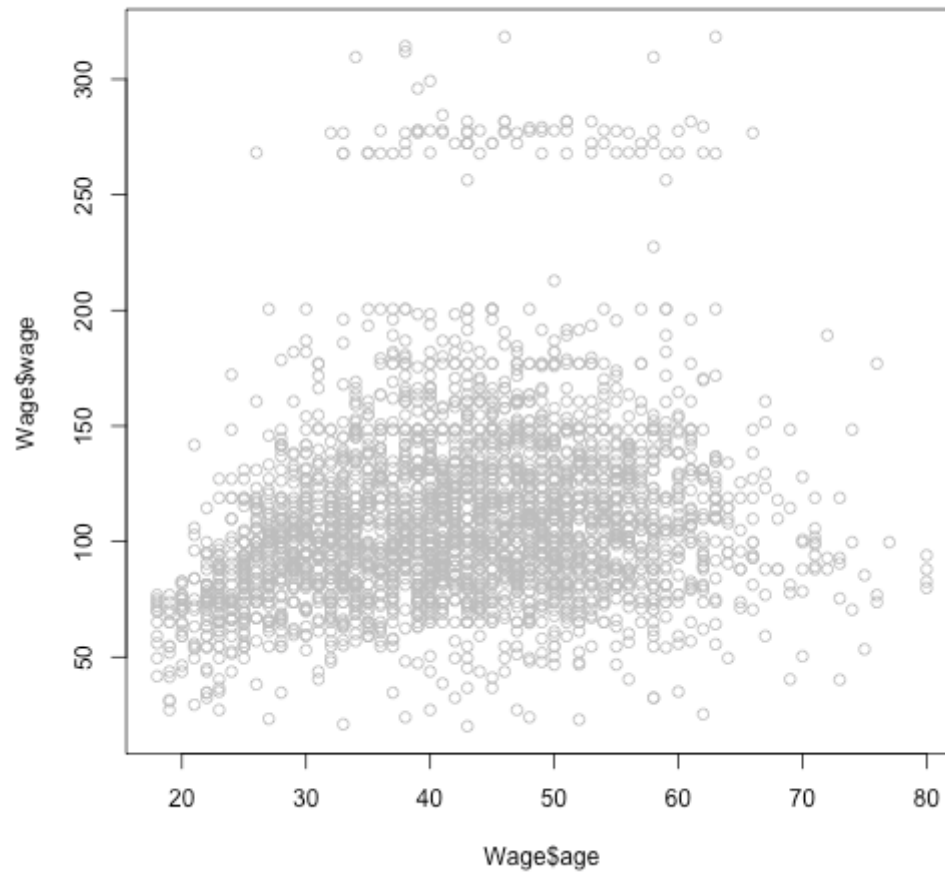
# Plot estimate + confidence bounds
lines(age.knots,preds$fit,col="blue",lwd=2)
lines(age.knots,preds$fit-2*preds$se.fit,col="red",lwd=1)
lines(age.knots,preds$fit+2*preds$se.fit,col="red",lwd=1)
```

Piecewise Linear Regression of wage on Age



Spline Functions 1

```
library(splines)  
plot(Wage$age, Wage$wage,col="gray")
```



```

# Fit model
# Here we specify knots at age = 25, 40, 60
fit.x1=lm(wage~bs(age,knots=c(25,40,60), degree=1),data=Wage)
# Or, evenly spaced knots
# fit.x1.df=lm(wage~bs(age,df=4,degree=1),data=Wage)
summary(fit.x1)

```

Call:

```

lm(formula = wage ~ bs(age, knots = c(25, 40, 60), degree = 1),
    data = Wage)

```

Residuals:

```

      Min       1Q   Median       3Q      Max
-99.795 -24.686  -4.856   15.344  204.671

```

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	54.333	5.957	9.120
bs(age, knots = c(25, 40, 60), degree = 1)1	37.645	6.817	5.522
bs(age, knots = c(25, 40, 60), degree = 1)2	65.847	6.019	10.940
bs(age, knots = c(25, 40, 60), degree = 1)3	63.850	6.319	10.104
bs(age, knots = c(25, 40, 60), degree = 1)4	33.772	10.580	3.192

	Pr(> t)
(Intercept)	< 2e-16 ***
bs(age, knots = c(25, 40, 60), degree = 1)1	3.64e-08 ***
bs(age, knots = c(25, 40, 60), degree = 1)2	< 2e-16 ***
bs(age, knots = c(25, 40, 60), degree = 1)3	< 2e-16 ***
bs(age, knots = c(25, 40, 60), degree = 1)4	0.00143 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39.91 on 2995 degrees of freedom

Multiple R-squared: 0.08665, Adjusted R-squared: 0.08543

F-statistic: 71.03 on 4 and 2995 DF, p-value: < 2.2e-16

Spline Functions 2

```
# plot(Wage$age, Wage$wage,col="gray")

fit.x1=lm(wage~bs(age,knots=c(25,40,60), degree=1),data=Wage)
preds1=predict(fit.x1,newdata=list(age=age.grid),se=T)
# lines(age.knots,preds$fit.x1,col="blue",lwd=2)

fit.x2=lm(wage~bs(age,knots=c(25,40,60), degree=2),data=Wage)
preds2=predict(fit.x2,newdata=list(age=age.grid),se=T)
# lines(age.knots,preds$fit.x2,col="red",lwd=2)

fit.x3=lm(wage~bs(age,knots=c(25,40,60), degree=3),data=Wage)
preds3=predict(fit.x3,newdata=list(age=age.grid),se=T)
# lines(age.knots,preds$fit.x3,col="green",lwd=2)

fit.ns=lm(wage~ns(age,df=4),data=Wage)
preds.ns=predict(fit.ns,newdata=list(age=age.grid),se=T)
# lines(age.knots,preds$fit.ns,col="black",lwd=2)
```