

Computer Project 1: Regression

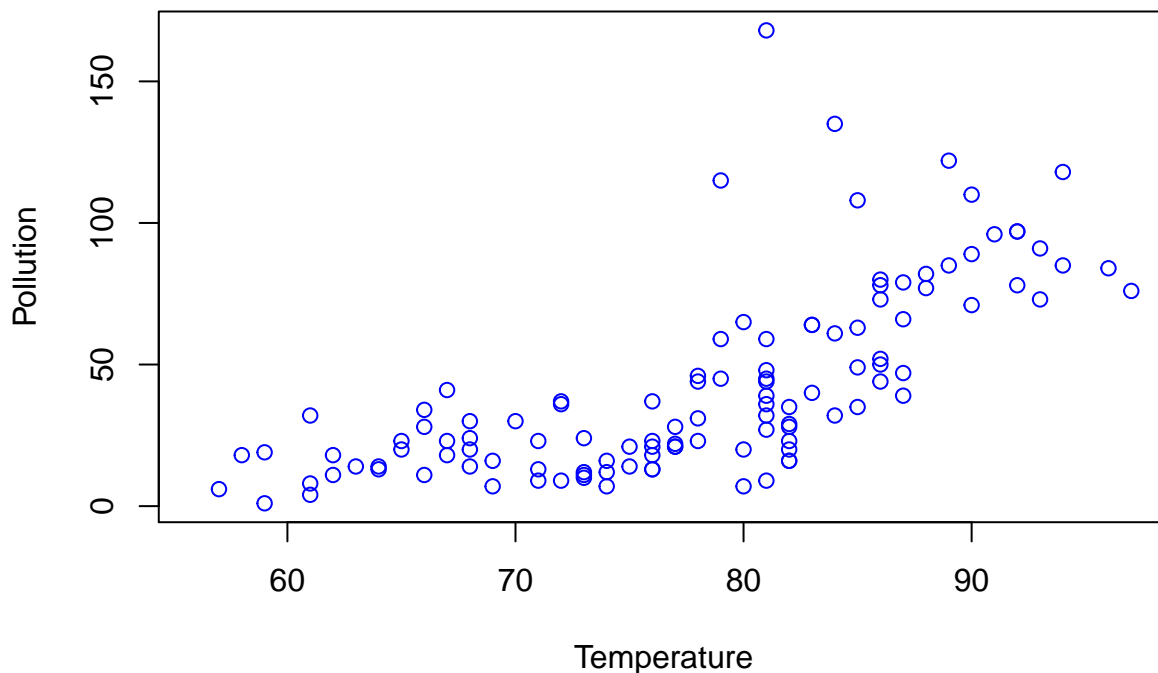
Your Name

2020-06-08

Predicting Pollution Levels from Daily High Temperature

The airquality data set has daily air quality measurements in New York, May to September 1973. Here is a plot of ozone levels (in ppb) as a function of daily high temperature (degrees F).

```
plot(airquality$Temp, airquality$Ozone, col='blue', xlab="Temperature", ylab="Pollution")
```



Based on the graph, would you say the correlation is positive or negative? Is the magnitude close to 1, 0 or something in between?

Correlation

Here is the correlation.

```
cor(airquality$Temp, airquality$Ozone, use='complete.obs')
```

```
## [1] 0.6983603
```

What does the correlation tell you about the association between temperature and ozone? Is it positive or negative? Is it a strong or weak association?

Regression

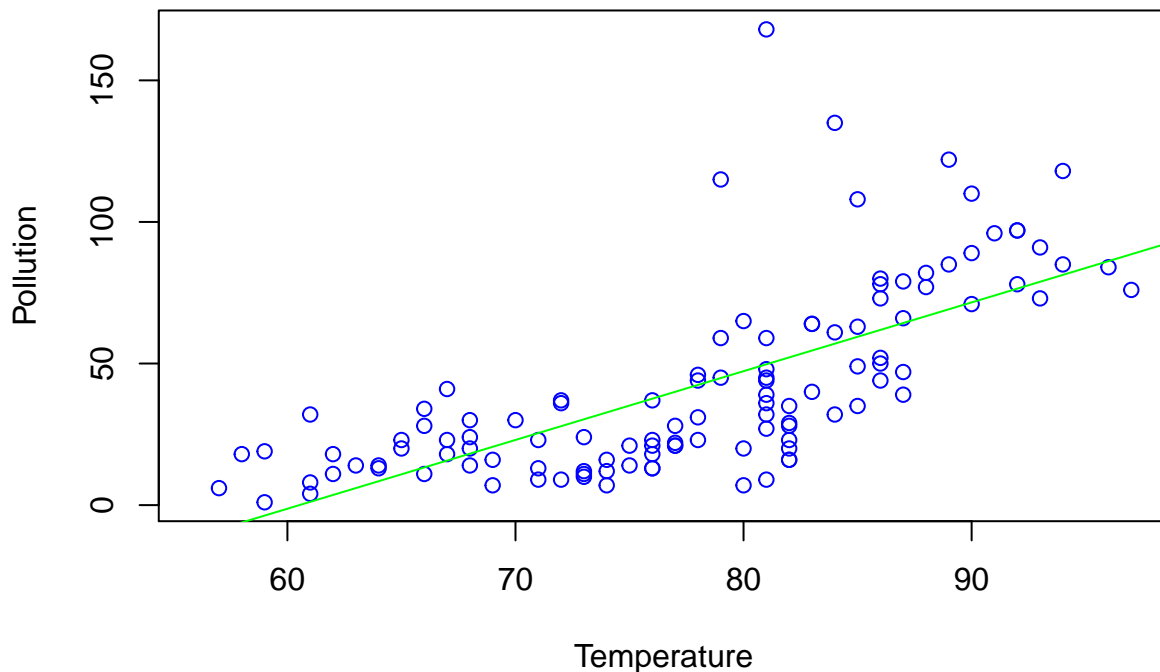
Here is a calculation of the regression line.

```
model <- lm(airquality$Ozone ~ airquality$Temp)
summary(model)

##
## Call:
## lm(formula = airquality$Ozone ~ airquality$Temp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -40.729 -17.409  -0.587   11.306  118.271
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -146.9955     18.2872  -8.038 9.37e-13 ***
## airquality$Temp    2.4287      0.2331   10.418 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.71 on 114 degrees of freedom
## (37 observations deleted due to missingness)
## Multiple R-squared:  0.4877, Adjusted R-squared:  0.4832
## F-statistic: 108.5 on 1 and 114 DF,  p-value: < 2.2e-16
```

Now plot the data with the regression line.

```
plot(airquality$Temp, airquality$Ozone, col='blue', xlab="Temperature", ylab="Pollution")
abline(model, col='green')
```

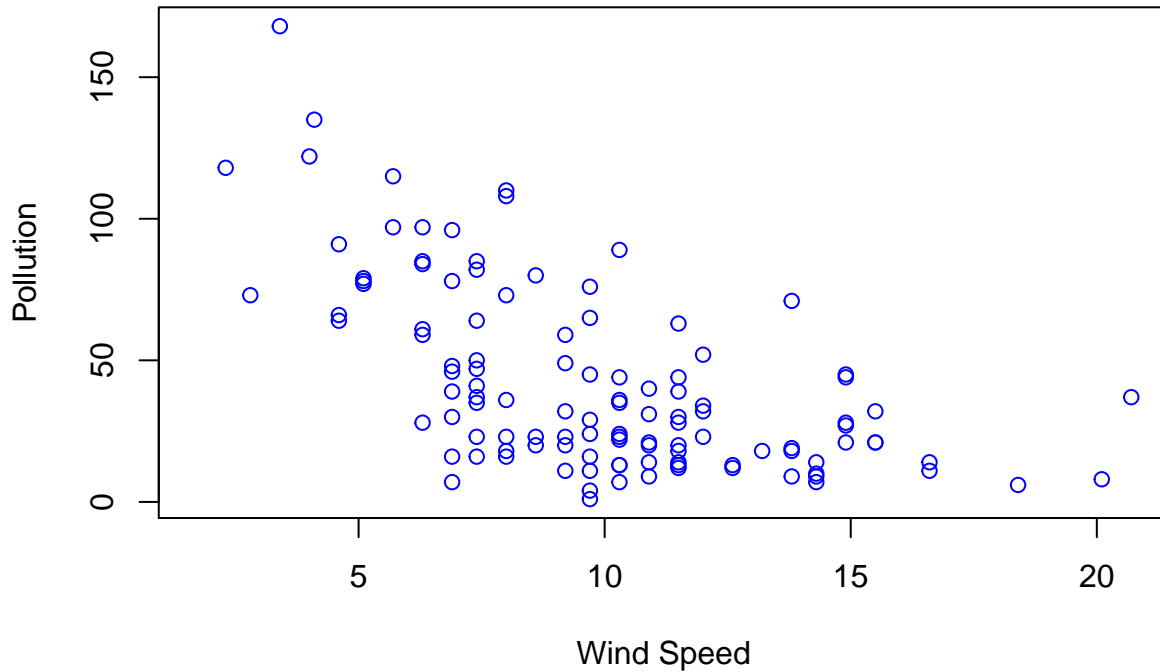


Use the output to write down the equation for the regression line. Does the data seem to fit the line well or would it be better to use a quadratic or some other curve?

Predicting Pollution Levels from Wind Speed

Here is a plot of ozone levels (in ppb) as a function of wind speed (mph).

```
plot(airquality$Wind, airquality$Ozone, col='blue', xlab="Wind Speed", ylab="Pollution")
```



Based on the graph, would you say the correlation is positive or negative? Is the magnitude close to 1, 0 or something in between?

Correlation

Here is the correlation.

```
cor(airquality$Wind, airquality$Ozone, use='complete.obs')
```

```
## [1] -0.6015465
```

What does the correlation tell you about the association between wind speed and ozone? Is it positive or negative? Is it a strong or weak association?

Regression

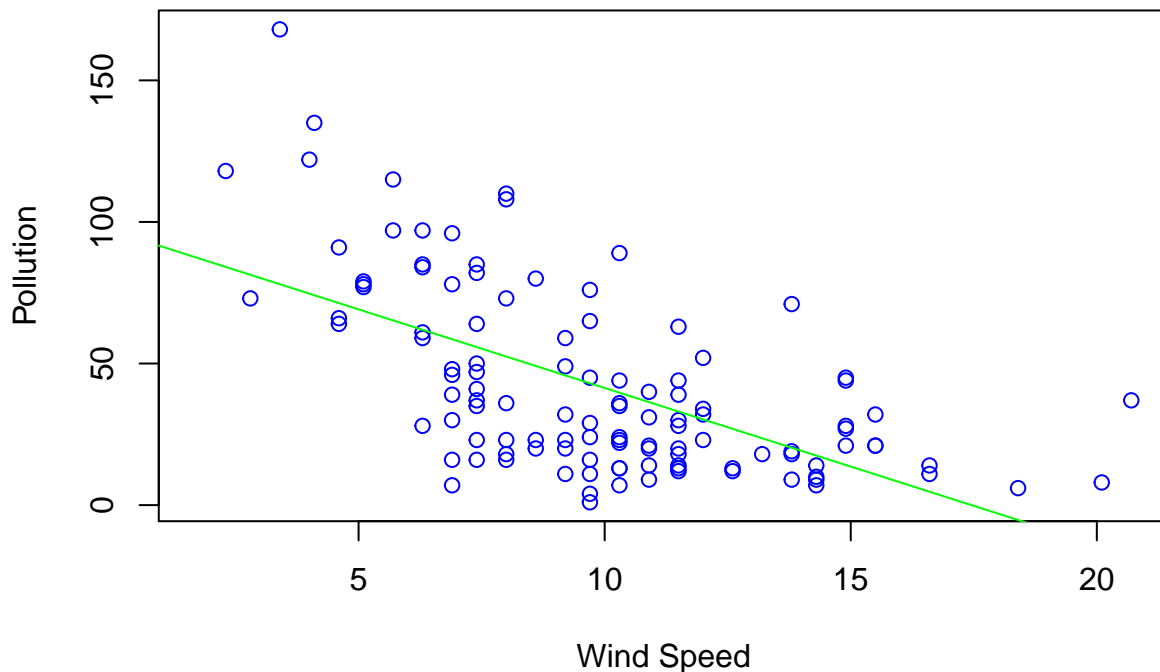
Here is a calculation of the regression line.

```
model <- lm(airquality$Ozone ~ airquality$Wind)
summary(model)

##
## Call:
## lm(formula = airquality$Ozone ~ airquality$Wind)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -51.572 -18.854  -4.868  15.234  90.000
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    96.8729     7.2387   13.38 < 2e-16 ***
## airquality$Wind -5.5509     0.6904   -8.04 9.27e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.47 on 114 degrees of freedom
## (37 observations deleted due to missingness)
## Multiple R-squared:  0.3619, Adjusted R-squared:  0.3563
## F-statistic: 64.64 on 1 and 114 DF,  p-value: 9.272e-13
```

Now plot the data with the regression line.

```
plot(airquality$Wind, airquality$Ozone, col='blue', xlab="Wind Speed", ylab="Pollution")
abline(model, col='green')
```

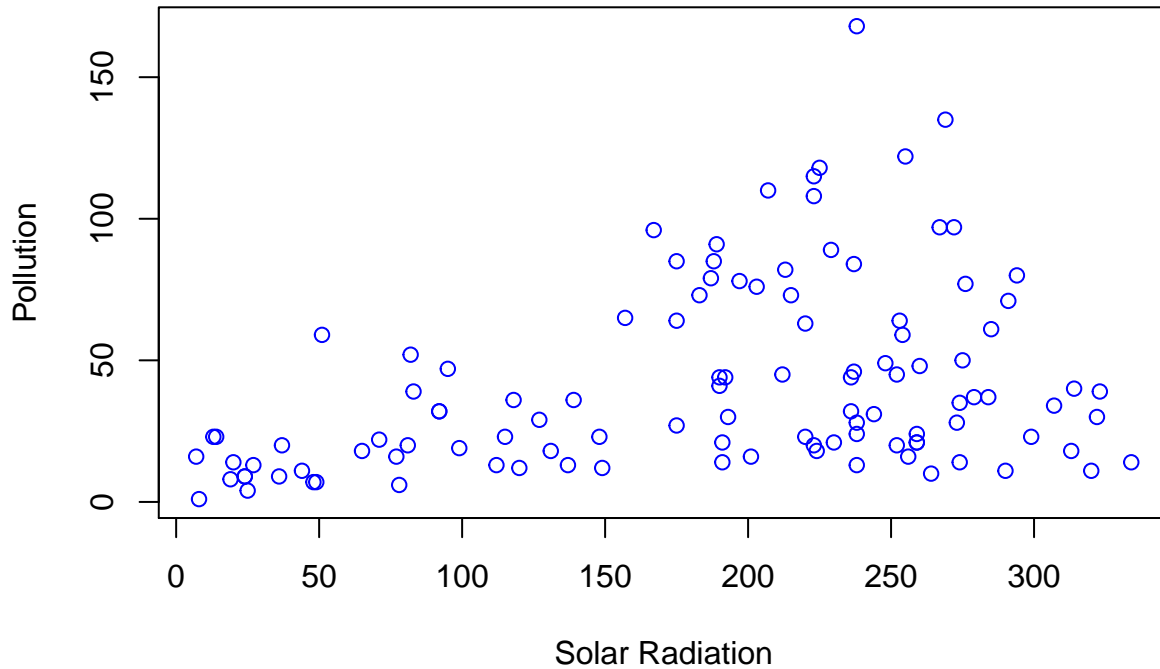


Use the output to write down the equation for the regression line. Does the data seem to fit the line well or would it be better to use a quadratic or some other curve?

Predicting Pollution Levels from Solar Radiation

Here is a plot of ozone levels (in ppb) as a function of solar radiation level (in lang).

```
plot(airquality$Solar.R, airquality$Ozone, col='blue', xlab="Solar Radiation", ylab="Pollution")
```



Based on the graph, would you say the correlation is positive or negative? Is the magnitude close to 1, 0 or something in between?

Correlation

Here is the correlation.

```
cor(airquality$Solar.R, airquality$Ozone, use='complete.obs')
```

```
## [1] 0.3483417
```

What does the correlation tell you about the association between solar radiation and ozone? Is it positive or negative? Is it a strong or weak association?

Regression

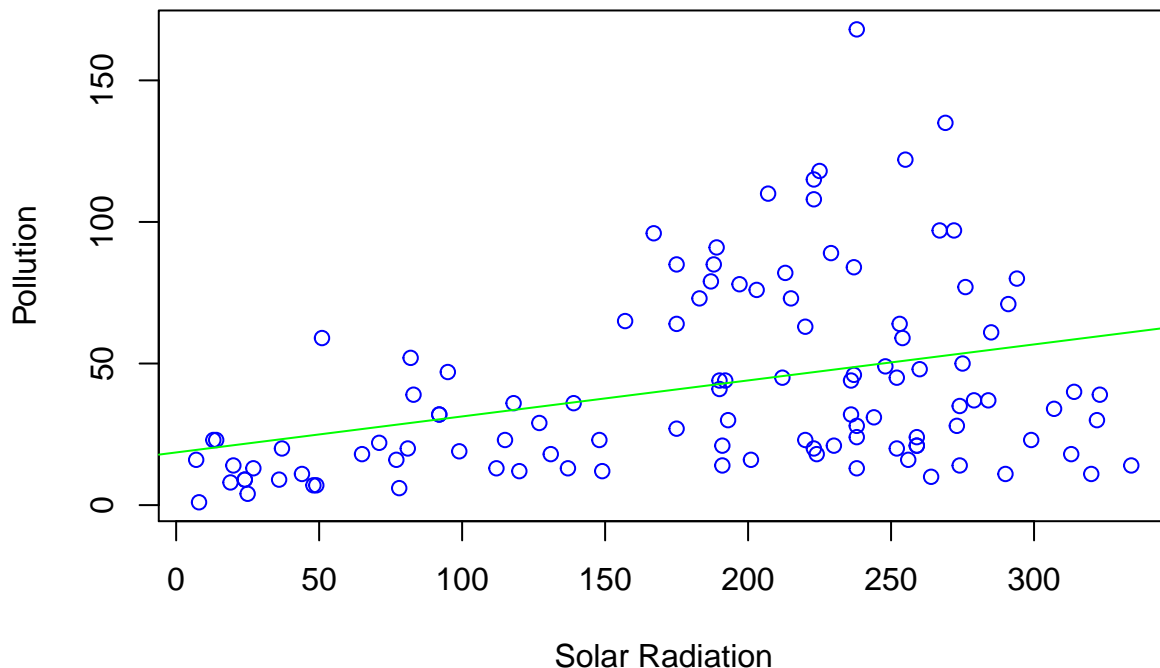
Here is a calculation of the regression line.

```
model <- lm(airquality$Ozone ~ airquality$Solar.R)
summary(model)

##
## Call:
## lm(formula = airquality$Ozone ~ airquality$Solar.R)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -48.292 -21.361  -8.864  16.373 119.136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    18.59873     6.74790   2.756 0.006856 **
## airquality$Solar.R  0.12717     0.03278   3.880 0.000179 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31.33 on 109 degrees of freedom
## (42 observations deleted due to missingness)
## Multiple R-squared:  0.1213, Adjusted R-squared:  0.1133
## F-statistic: 15.05 on 1 and 109 DF,  p-value: 0.0001793
```

Now plot the data with the regression line.

```
plot(airquality$Solar.R, airquality$Ozone, col='blue', xlab="Solar Radiation", ylab="Pollution")
abline(model, col='green')
```



Use the output to write down the equation for the regression line. Does the data seem to fit the line well or would it be better to use a quadratic or some other curve? Which of the three variables would be the most

helpful in predicting ozone levels?
Which the least? Explain your conclusions.