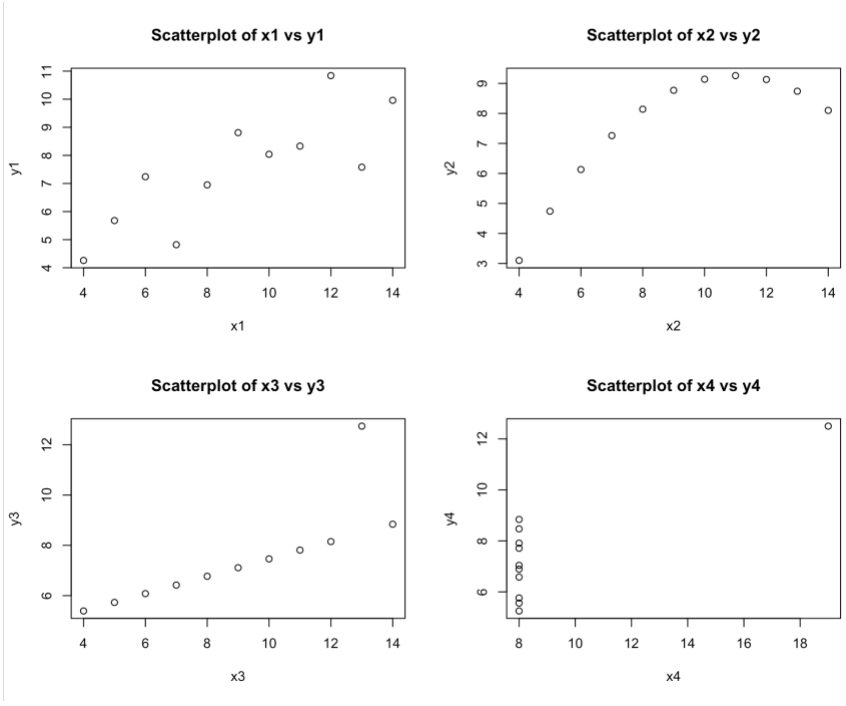
1)

a)

	Mean:	Median:	Standard Deviation:
X_1 =	9.00	9.00	3.32
X_2 =	9.00	9.00	3.32
X_3 =	9.00	9.00	3.32
X_4 =	9.00	8.00	3.32
Y_1 =	7.50	7.58	2.03
Y_2 =	7.50	8.14	2.03
Y_3 =	7.50	7.11	2.03
Y_4 =	7.50	7.04	2.03

b) The data sets seem to have all be very similar for each $X_{1\ldots n}$ and for the $Y_{1\ldots n}$ values. They all seem to have a common mean ranging very little and the same STDEV for all. The median for the Y values seems to be more varied though.

c)

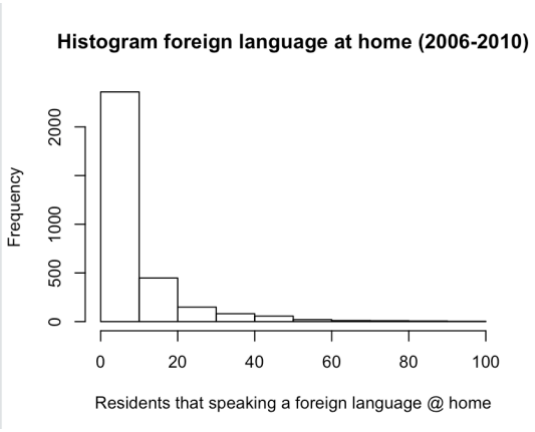


d) All the data set looks very different when you put them in the plots. All have different relationships as x increases, especially the X_4 by Y_4 graph. While most graphs have a slightly positive correlation, they still experience in a different curvature which could all lead to different interpretations.

e) Go off more than just mean, median, data and look at the actual numbers. Because will they all seem pretty similar the graph shows another story about the relationship.

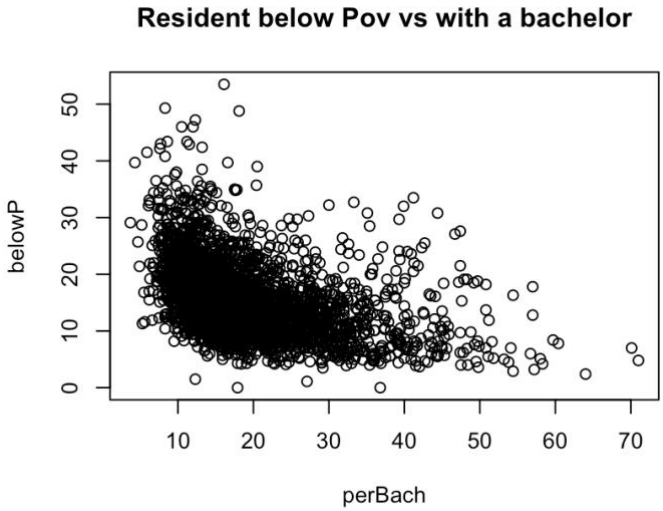
2)

a)



b) median per-county amount of federal spending in 2009 = 214994

c)



Comments: I see a lot of data points around the region of less people with a bachelor and with less people below the poverty line. Also low amount of bachelor holders correlate to a high amount of poverty and vice versa.

d) 3.627%

3)

a) $\frac{Sum\ of\ 24\ houses}{24} = 500{,}000$

$Sum\ of\ 24\ houses = 12{,}000{,}000$

$New\ mean\ house\ price = \frac{12{,}000{,}000+700{,}000}{25} = 508{,}000$

b) $\sigma_{new} = \sqrt{\frac{(n+1)(v^2+sum(x^2)) - (v+sum(x))^2}{(n+1)^2}}$ // sum = sum of 25 elements, n = # of houses

$$= \sqrt{\frac{(n+1)\left(v^2 + n(\sigma^2 + \mu^2)\right) - \left(v + sum(x)\right)^2}{(n+1)^2}}$$
$$= 49{,}623$$

c) No conclusion can be made about the median. Because with the current information we have we cannot tell were 700,000 will fit in the data and we need its position to determine the median.

4)

a)

$$\bar{y} = \frac{(x_1 + c)k + \cdots + (x_n + c)k}{n}$$
$$\bar{y} = \frac{k\left((x_1 + c) + \cdots + (x_n + c)\right)}{n}$$
$$\bar{y} = \frac{k\left((x_1 + \cdots + x_n) + nc\right)}{n}$$
$$\bar{y} = \frac{k\left((x_1 + \cdots + x_n)\right)}{n} + \frac{knc}{c}$$
$$\bar{y} = k\bar{x} + kc$$

$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$
$$S_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\,(x_i + c)k - (k\bar{x} + c)\,)^2$$
$$S_y^2 = k^2\,\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
$$S_y^2 = k^2 S_x^2$$

b)

$$\bar{Z} = \frac{Z_1 + \cdots + Z_n}{n}$$
$$= \frac{\frac{X_1 - \bar{X}}{S_X} + \cdots + \frac{X_n - \bar{X}}{S_X}}{n}$$
$$= \frac{(X_1 + \cdots + X_n) - n\bar{X}}{nS_X}$$
$$= \frac{\bar{X}}{S_X} - \frac{\bar{X}}{S_X} = 0$$
$$\bar{Z} = 0$$

$$S_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$$
$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_X}\right)^2$$
$$= \frac{S_X}{S_X} = 1$$
$$S_z^2 = 1$$

RCODE:

```
x1 = Quartet$x1
x2 = Quartet$x2
x3 = Quartet$x3
x4 = Quartet$x4

y1 = Quartet$y1
y2 = Quartet$y2
y3 = Quartet$y3
y4 = Quartet$y4

meanX1 = mean(x1)
meanX2 = mean(x2)
meanX3 = mean(x3)
meanX4 = mean(x4)
meanY1 = mean(y1)
meanY2 = mean(y2)
meanY3 = mean(y3)
meanY4 = mean(y4)

medianX1 = median(x1)
medianX2 = median(x2)
medianX3 = median(x3)
medianX4 = median(x4)

medianY1 = median(y1)
medianY2 = median(y2)
medianY3 = median(y3)
medianY4 = median(y4)

sdX1 = sd(x1)
sdX2 = sd(x2)
sdX3 = sd(x3)
sdX4 = sd(x4)

sdY1 = sd(y1)
sdY2 = sd(y2)
```

```
sdY3 = sd(y3)
sdY4 = sd(y4)

par(mfrow=c(2,2))
plot(x1,y1, main="Scatterplot of x1 vs y1")
plot(x2,y2, main="Scatterplot of x2 vs y2")
plot(x3,y3, main="Scatterplot of x3 vs y3")
plot(x4,y4, main="Scatterplot of x4 vs y4")

hist(CountyData$foreign_spoken_at_home, main = "Histogram foreign language at home
(2006-2010)" ,  xlab="Residents that speaking a foreign language @ home")
fedS = CountyData$fed_spending
medianFedSpend = median(fedS[which(!is.na(fedS))])
belowP = CountyData$poverty
perBach = CountyData$bachelors
plot(perBach,belowP, main = "Resident below Pov vs with a bachelor")

isGoodNumber <- function(X)
{
  if (X>30) return(TRUE) else return(FALSE)
}
underage = CountyData$age_under_18
over30 = Filter( isGoodNumber, underage)
length(over30);
len = length (over30) / length(underage)
```